

Is Ann Arbor experiencing more extreme weather?

Week 2, Lecture 3

<https://live.ds306.org>

September 8, 2026

Lecture 2 recap

```
library(tidyverse)
theme_set(theme_minimal(base_size = 13, base_family = "sans"))

load("a2weather.RData")
a2weather <- as_tibble(a2weather) |>
  mutate(
    year = as.integer(format(date, "%Y")),
    month = as.integer(format(date, "%m"))
  )
```

Last week we learned:

- `mutate()` to create new columns;
- `filter()` to retain selected rows;
- `select()` to retain selected columns; and
- `|>` to apply several operations in order.

Today we will use the same tools on a new dataset.

What we'll learn today

- Sort observations with `arrange()`.
- Calculate group summaries with `group_by()` and `summarize()`.
- Recognize missing values and missing rows, and how they affect averages.
- Break a question into steps, compare results, and explain the evidence.

ALL
THE NEWS
FIT TO
GOSSIP



THE FACULTY ENQUIRER

HIGHER ED. LOWER STANDARDS. SAME GREAT GOSSIP.

ACADEMIA
NEVER
SLEEPS
(BUT IT DOES
PARK OUTSIDE)

VOL. 17, NO. 28

TUESDAY, AUGUST 26, 2025

\$1.50

TENURE NO PROTECTION AGAINST TIMBER

Area professor's car smashed by falling tree!



A fallen tree brought a sudden end to one professor's peaceful morning, crushing their Subaru in a neighborhood near campus on Monday. No injuries were reported, but sources say the car's grad school days may be over.

SECOND ANGLE:



Another view shows the full extent of the arboreal ambush. "Same car, different wood angle," said a witness.

SQUIRREL DENIES INVOLVEMENT

"I was in a different tree," says local squirrel.

"A neighborhood squirrel, reached for comment, strongly denied any role in Monday's tree-related incident.

"I've got an alibi," the squirrel said, before scampering up a fence and out of reach.

Wildlife experts say there is no evidence squirrels can fell trees this size, but "they do like to take credit for stuff."



PARKING OFFICE STILL ISSUES TICKET



Despite the obvious act of nature, campus parking officials confirmed a citation was issued. "Trees don't void parking rules," said a spokesperson. The professor plans to appeal, noting extenuating circumstances — namely, a tree.

TREE DECLINES COMMENT



The fallen tree, unreachable for comment, remained silent Tuesday. A note was left at the scene, but there was no response. "Maybe it's just leafing it alone," said one neighbor.

SAME CAMPUS. HIGHER DRAMA.

smashed car

Today's questions

1. How can we summarize noisy daily temperatures to see seasonal and longer-term patterns?
2. Have extreme precipitation days become more frequent?

The data

```
a2weather
```

The [Global Historical Climatology Network](#) collects daily measurements from weather stations around the world.

Our table contains observations from one station on the U-M campus: USC00200230.

- Each row describes one day at this station.
- tmax is that day's highest temperature, in Celsius.
- Date is encoded as year, month, day.

Last week's observations

```
a2weather |>  
  filter(date >= "2026-08-30") |>  
  select(date, tmin, tmax, prcp)
```

Q1: Can we see a trend in daily temperatures?

Before looking at the data, what sort of trend do you expect to see in daily temperatures over the years? Be as specific as possible.

Sorting data with `arrange()`

- Suppose we want to find the days with the lowest high temperatures.
- `arrange(tmax)` sorts rows from the smallest daily high to the largest.
- Each day's date and other measurements move together with its temperature.
- Sorting changes the order of the rows; it does not remove any observations.

```
a2weather |>
  select(date, tmin, tmax, prcp) |>
  arrange(tmax)
```

Reverse order with desc()

To find the hottest days, put the largest values first.

```
a2weather |>
  select(date, tmin, tmax, prcp) |>
  arrange(desc(tmax))
```

- desc(tmax) puts the largest temperatures first.
- You can sort by other columns too. Dates run from earliest to latest; desc() reverses that order.

What was the highest temperature ever recorded in Ann Arbor?

According to a2weather, what was the coldest temperature ever recorded in Ann Arbor, and on what date? Write the code you would use, and the answer.

Long-term trends

```
ggplot(a2weather) + geom_point(aes(x = date, y = tmax))
```

Interpretation

Does this plot tell you about long-term trends in daily high temperatures? Why or why not?

What would make the pattern easier to see?

- It's hard to visualize the long-term trend because the daily temperatures fluctuate so much.
- To make the long-term trend easier to see, we can **average** the data so that some of the noise is smoothed out.

Calculating the mean

Using what we already know, we can calculate the average high temperature for any given month or year:

```
weather_2025 <- a2weather |>  
  filter(year == 2025) |>  
  select(date, year, month, tmax)  
  
mean(weather_2025$tmax)
```

Summarizing data with `summarize()`

We can automate the process of calculating the mean by using `summarize()`:

```
a2weather |>
  filter(year == 2025) |>
  summarize(mean_tmax = mean(tmax))
```

- `filter(year == 2025)` keeps just the days in 2025.
- `mean(tmax)` calculates the mean of its temperature column.
- `mean_tmax` = names the result column. You choose this name.

What if we use all the years?

Without the filter, we average daily highs across the entire dataset:

```
a2weather |> summarize(mean_tmax = mean(tmax))
```

Why did we get NA instead of a temperature?

Missing data

Why we got NA

- R has a special value, NA, for representing missing data.
- You can think of NA as “I don’t know.”
- R uses NA to mark **explicitly missing** data.

```
NA + 5    # I don't know + 5 = I don't know  
1 < NA    # Is 1 less than I don't know? I don't know.  
mean(c(20, NA, 30))
```

We know two of these temperatures, but not the third. So we don’t know the mean of all three.

Missing data in the weather record

Some dates have a row, but tmax is NA. Those temperatures are **explicitly missing**.

```
a2weather |>  
  filter(year == 1982, is.na(tmax)) |>  
  select(date, tmax)
```

Calculate the mean of the temperatures we have

The argument `na.rm = TRUE` tells `mean()` to leave out missing values.

```
mean(c(20, NA, 30), na.rm = TRUE)  
summarize(a2weather, mean_tmax = mean(tmax, na.rm = TRUE))
```

- The first calculation averages just two temperatures: 20 and 30.
- The second returns a one-row table with the mean of the recorded highs.

Thinking about missing data

Consider two scenarios:

1. Temperatures are missing because the sensor randomly fails for no reason.
2. Temperatures are missing because the sensor fails more often on very cold days.

How would each scenario affect the mean temperature we calculate?

Grouping and summarizing data

A separate mean for every year

We know how to calculate the mean daily high for 2025:

```
a2weather |>
  filter(year == 2025) |>
  summarize(mean_tmax = mean(tmax, na.rm = TRUE))
```

- To get 2024, we could change the year in `filter()`.
- But we want to compare **every year**, without repeating this code over a hundred times.
- Removing `filter()` gives one mean for the whole record, not one per year.

Tell R which days belong together

We want all the 1891 rows summarized together, then all the 1892 rows, and so on.

```
grouped_years <- a2weather |> group_by(year)
```

- `group_by(year)` marks which rows belong to the same group.
- Unlike `filter(year == 2025)`, it keeps the other years too.
- The daily temperatures are still there; no means have been calculated yet.

Now calculate within each year

```
annual_means <- grouped_years |> summarize(  
  mean_tmax = mean(tmax, na.rm = TRUE)  
)
```

- R calculates `mean( tmax, . . . )` separately for each year's rows.
- The result has one row per year, with its year and mean daily high.
- The calculation is the same as before; grouping changes which rows it uses.

Annual mean daily highs

```
ggplot(annual_means, aes(x = year, y = mean_tmax)) +  
  geom_line() + geom_point() +  
  labs(x = "Year", y = "Mean daily high (C)")
```

Outlying years

An **outlier** is a value that is unusually far from the other values in a dataset.

Do you see any outlying years in the plot above? What do you think could have caused them?

Keeping track of the number of days

```
a2weather |> group_by(year) |> summarize(  
  n_days = n(), mean_tmax = mean(tmax, na.rm = TRUE)  
)
```

Use the table above: what explains the outlier that we saw in the plot?

Implicitly missing dates

```
a2weather |>
  filter(year == 1891) |>
  summarize(first_date = min(date), last_date = max(date))
```

- The series begins on October 1, 1891.
- January through September have **no rows** in the table. Those days are **implicitly missing** when we try to summarize the whole year.
- Thus, in 1891, the first nine months are **implicitly missing**, and only the last three months (which are colder than usual) are represented in the data.

Do some years have a lot of missing data?

- `n()` counts rows, including days with a missing `tmax`.
- `sum(is.na(tmax))` counts explicitly missing temperatures.
- `!` reverses the test, so `sum(!is.na(tmax))` counts recorded temperatures.

```
a2weather |> group_by(year) |>
  summarize(
    n_days = n(),
    n_missing = sum(is.na(tmax)),
    n_complete = sum(!is.na(tmax))
  )
```

Summarizing by decade

```
a2weather |> mutate(decade = year %% 10 * 10) |>  
  group_by(decade) |>  
  summarize(mean_tmax = mean(tmax, na.rm = TRUE))
```

Advanced: moving average

```
a2weather |>
  arrange(date) |>
  mutate(
    moving_avg = slider::slide_mean(
      tmax, before = 4000, after = 4000, na_rm = TRUE
    )
  ) |>
  ggplot(aes(x = date)) +
  geom_line(aes(y = moving_avg))
```

Q2: Are extreme precipitation days becoming more frequent?

What is your plan of attack?

How would you use these daily weather data to decide whether extreme precipitation has become more frequent?

Think on your own, then compare ideas with your group. Submit a plan with 3–5 steps, in order. What would you measure, compare, and check? You don't need to write R code yet.

Our plan of attack

1. Define an extreme precipitation day and choose two periods.
2. Count those days in each year and check for missing observations.
3. Compare the average yearly counts in the two periods.
4. Use a two-sample test to assess the evidence.
5. Explain the difference, the uncertainty, and what to investigate next.

Step 1: Define what we will count

- prcp is daily precipitation in millimeters, including the water equivalent of snow.
- For this analysis, call a day with **at least 25.4 mm (1 inch)** an extreme precipitation day. (I chose this semi-arbitrarily.)

```
extreme_cutoff <- 25.4
```

What would this definition miss?

Two days each receive 1 inch of precipitation. On one, it falls in 20 minutes; on the other, it is spread over 24 hours. Our rule counts both. What does this measure capture, and what does it miss?

Choose the comparison periods

- The baseline is 1961–1990.
- The recent period is 1996–2025.
- Each spans 30 calendar years. We still need to check how much data we have.

```
comparison_years <- c(1961:1990, 1996:2025)

rain_days <- a2weather |>
  filter(year %in% comparison_years)
```

`%in%` keeps years that appear in our list.

Step 2: Mark the extreme days

```
rain_days <- rain_days |>  
  mutate(extreme = prcp >= extreme_cutoff)  
  
rain_days |> select(date, prcp, extreme)
```

- `extreme` is TRUE or FALSE when precipitation is recorded.
- If `prcp` is missing, `extreme` is NA: we don't know whether it qualifies.

Count days within each year

When we add logical values, R counts TRUE as 1 and FALSE as 0.

```
sum(c(TRUE, FALSE, TRUE))

annual_rain <- rain_days |>
  group_by(year) |>
  summarize(
    recorded_days = sum(!is.na(prcp)),
    extreme_days = sum(extreme, na.rm = TRUE)
  )

annual_rain
```

`na.rm = TRUE` counts the extreme days we observed. It cannot tell us how many occurred on missing days.

Could missing data create an apparent increase?

Suppose an earlier year has 270 recorded days and a recent year has 365. The recent year has more recorded extreme precipitation days. What would you need to check before interpreting that as an increase? Suggest a way to make the comparison fairer.

Keep years with a full set of observations

For our first comparison, use only years with precipitation recorded every day.

```
annual_rain <- annual_rain |>
  mutate(expected_days = if_else(leap_year(year), 366L, 365L))

annual_rain |> filter(recorded_days < expected_days)

complete_rain <- annual_rain |>
  filter(recorded_days == expected_days)
```

`leap_year(year)` checks whether the calendar year has 366 days. This leaves 25 baseline years and 30 recent years.

Label the two periods

```
complete_rain <- complete_rain |>
  mutate(
    period = if_else(year <= 1990,
      "Baseline: 1961-1990", "Recent: 1996-2025")
  )

head(complete_rain)
```

There is still one row per year. We have added its period, not combined the years.

Predict the comparison

Before calculating the period means, predict how the number of extreme precipitation days per year has changed. Give a rough difference in days per year and a reason. What result would surprise you?

Step 3: Compare average yearly counts

```
period_summary <- complete_rain |>
  group_by(period) |>
  summarize(
    n_years = n(),
    mean_extreme_days = mean(extreme_days),
    sd_extreme_days = sd(extreme_days)
  )

period_summary
```

`n()` now counts years. `sd()` describes how much the yearly counts vary within each period.

Calculate the difference

```
baseline_years <- complete_rain |>
  filter(period == "Baseline: 1961-1990")

recent_years <- complete_rain |>
  filter(period == "Recent: 1996-2025")

observed_difference <-
  mean(recent_years$extreme_days) - mean(baseline_years$extreme_days)

observed_difference
```

The result is a difference in **extreme precipitation days per year**.

What would this change mean?

Use the two means and their difference to describe the change in ordinary language. Would this difference matter to someone responsible for flooded streets or basements? What else would they need to know?

Look at the individual years

```
ggplot(complete_rain, aes(x = year, y = extreme_days, color = period)) +  
  geom_point() +  
  labs(  
    x = "Year", y = "Days with at least 1 inch of precipitation",  
    color = "Period"  
  )
```

Does the plot change your interpretation?

Does looking at the plot reveal anything that simply comparing the means hid? Point to something in the plot that strengthens or weakens your confidence in the comparison, and explain why.

Compare the two samples

```
comparison_test <- t.test(  
  recent_years$extreme_days,  
  baseline_years$extreme_days,  
  alternative = "greater"  
)  
  
comparison_test
```

The recent years are the first input, so "greater" asks whether their mean exceeds the baseline mean.

Step 5: Explain what we found

Write two or three sentences for someone who hasn't seen our code. Explain what changed, by how much, and how convincing you find the evidence. Include one limitation. Specify our definition of an extreme day, the location, and the periods; don't just report that a p-value is small.

What should we investigate next?

Someone reads our result and says, “Storms are becoming more destructive in Ann Arbor.” What additional evidence would you need to evaluate that claim? Suggest one concrete next analysis or source of data, and explain what it would help us learn.

What we still need to check

- Dropping incomplete years can change which years our comparison represents.
- The test treats years as independent. Correlation between successive years could affect the uncertainty calculation.
- Results could depend on the threshold. Check other defensible thresholds, and report them rather than selecting whichever gives the smallest p-value.
- Daily precipitation totals do not measure wind, flood damage, or what caused a particular storm. One station does not establish a global pattern.

Exit question

Suppose we wanted to study extreme heat instead. How would you define an extreme day, and which parts of today's analysis could you reuse? Describe one choice or complication you would need to reconsider.

Week 2, Lecture 3

Closed September 8, 2026 at 3:47 PM

15 class questions · 999 anonymous responses · 0 student questions

The following slides preserve what students submitted during class. No names or login information are included.

Before looking at the data, what sort of trend do you expect to see in daily temperatures over the years? Be as specific as possible.

90 anonymous responses

General warming trend

48

Responses expecting average daily temperatures to increase over years (attributed to climate change/global warming) without specifying magnitude.

Warming with seasonal/periodic cycle

9

Answers noting the seasonal oscillation (warmer summers, colder winters) combined with an overall upward trend.

Increasing extremes/greater variability

8

Predictions that temperatures will become more extreme or show larger ranges (higher highs, lower lows) or more outliers over time.

Other responses

8

Responses the model could not place reliably.

Response themes for question 1

Slight/quantified warming expectations**7**

Students giving specific magnitudes or rates of warming (e.g., degrees per year/decade or totals over decades).

Unclear/affirmative short replies**4**

Very brief or non-descriptive answers that do not state a clear trend.

Linking temperature to precipitation or urbanization**3**

Students mentioning associated changes like increased precipitation, decreased rainfall, or urban heat effects alongside temperature trends.

Stable or no long-term change**3**

Responses expecting little or no change in average temperatures over years (stagnation or stability).

Student responses

1. Daily temperatures should increase over years due to global warming.
2. Due to climate change, I would expect daily temperatures to rise.
3. Due to global warming, I would expect the daily temperatures to gradually increase over the years.
4. during winter time, the temperature goes down and during summer time, the temperature goes up! overall, the years go by the temperature has been rising during the same time of the year
5. I believe that daily temperatures have increased over the years
6. I expect daily temperatures to rise overall across the years, as urbanization is known to cause a rise in local temperature, combined with the effects of climate change.
7. i expect daily temps to increase over the years on average
8. I expect that daily temperatures are going to be increasing over the years due to global warming. I think there will also be a greater variety in daily temperature in recent years as well.
9. I expect that temperatures have increased over the years. Due to global warming and other "man-made" factors, I think that the temp has increased slowly over decades/centuries

Student responses

1. I expect that temperatures will increase on average.
2. I expect the daily temperature to vary greatly WITHIN a single year and over the years I expect it to slightly rise (due to global warming)
3. I expect to see daily temperatures over the years rising over the years as a result of global warming. I would expect an increase of about 2 degrees over the past 20 years.
4. I expect to see daily temps rising over the years due to global warming
5. I expect to see the daily temperatures rise slightly over the years, but for annual trends like seasonal changes to remain the same.
6. I expect to see the temperatures rising in general because of global warming.
7. I hypothesize the trend will be that the temperature has been slowly increasing over the years due to climate change
8. I intend to see daily temperature slightly rise over the years due to global warming and the feedback loops involved.
9. I predict that daily temperatures have increased over the years on average, due to climate change and global warming.

Student responses

1. i think daily temp will increase over the years beacuse climate change
2. I think over the years, the daily temperatures in Ann Arbor have increased throughout the summer
3. I think that due to global warming / climate change, the daily temperatures over the years will be on average greater than they used to be.
4. I think that temperatures will have risen over the year because of global warming and other factors
5. I think there will be an increase in average temprature as he years go n.
6. I would expect a trend of increase in daily tempartures on par with that of global warming trends.
7. I would expect daily temperatures to rise over the years because of climate change and more CO2 trapping heat in the atmosphere
8. I would expect higher and higher temperatures due to global warming.
9. I would expect temperatures to increase over the years

Student responses

1. I would expect that the average temperature would increase over the years due to global warming
2. I would expect the average daily temperature to increase per year, due to global warming.
3. I would expect the daily temp to increase over time
4. I would expect the daily temperatures over the years to increase due to global warming and climate change.
5. I would expect to see a slight rise in daily temperatures over the past five years. By around .25 degrees celsius.
6. I would expect to see an increase in average daily temperatures over the years by a degree or two
7. I would expect to see higher temperatures, on average, throughout the year. Maybe there would also be an increase of precipitation as well
8. I would expect to see the average max temp to rise due to global warming
9. I would expect to see the daily temperature averages to remain similar for the season, but gradually getting hotter as the years progress

Student responses

1. I'd expect it to slowly increase year after year due to global warming. otherwise it looks like an oscillating curve
2. If we see a lot more data then we might be able to see a pattern where the temperatures are rising.
3. On average due to global warming, average daily temps throughout the year will increase.
4. On average, getting warmer because of global warming.
5. Over the years I would expect to see a steady rise in daily temperatures due to global warming
6. The average is probably increasing over time
7. We expect an upward overall trend in the average daily temperature due to global warming.
8. We expect yearly temperature variations (colder in winter, warmer in summer). Maybe general warming trend.
9. We might expect to see a small increase in daily temperature that mirrors increases in industrial activity due to the effects of climate change.

Student responses

1. We shold be expecting a gradual increase in the daily temperatures over the years around the same time of the year.
2. We should see a slight rasing temp over years, since the global warming.
3. We would expect to see a rise in average daily temps

Student responses

1. among each year, temps will be up during months of summer, dip down during winter (so more of an oscillating trend in tempertaure) but overall i expect a general increase in temp.
2. Daily temperatures tend to go up until the hottest summer and go down until the coldest winter.
3. I expect the daily temperature to rise from around February to August since the seson is changing from winter to summer. However, after August, I think the daily temperature will drop back as the season changes from summer to winter back.
4. I think that we will have higher temperatures in the summer and the fall, and then towards the winter and spring we are going to have colder temperature, so I think it is going to get more and more away from the original mean because of outside variables like global warming and such
5. I would definitely say the trendfollows th season , hot and
6. I would expect to see a general rise in temperatures across seasons; that is, summers getting hotter over time.

Student responses

1. Temperature should decrease during the winter months and increase during the summer months. Over the course of a few years, temperatur extremes should become more common.
2. temps go up until early sep and then down later
3. Tend to see warmer temps in spring and summer, cooler temps In fall and winter. Weather correlates with season

Student responses

1. a slight increase in temperature throughout the years as well as more outlier days with either very cold or very warm weather
2. Across years, I feel as if the yearly minimum would gradually get lower, and the yearly maximum would gradually get higher. I would attribute this to climate change getting progressively worse as time has gone on.
3. As time passes by, temperatures in summer would become hotter and temperatures in winter would become colder. Overall, I expect the temperatures to become more and more extreme.
4. daily temperature should fluctuate among days, but should follow a roughly periodic change where winter has lower and summer has higher temperature.
5. I expect daily temperatures to have become more extreme over the years, with a larger range due to climate change. I expect generally that temperatures have become warmer as well.
6. I expect to see a slight average rise overall and more extreme temperatures in terms of heat and cold. I as well expect to see November to be cooler in the past as opposed to now.

Student responses

1. i think they will become more drastic with the changes in climate chnage. wolder winters and hotter summers
2. Lower mins, higher maximums, and more extreme percipitation.

Student responses

1. Expect tempature to go up very slowly over the years due to global warming
2. I think daily temperatures are rising over the year because of global warming/climate change. I also have noticed that the "start" of winter is getting pushed back later and later. For example, the first snows are arriving later and later.
3. I wold exepect the daily average temperature to be rising in the 2000s due to global warming.
4. it continuously decreasing
5. Its going to be warmer during the summer and colder temperature during the winter.
6. Possibly that average temperatures are rising by some percentage.
7. summer high, winter low; average temperature going higher each year; 4 seasons temperature pattern
8. We can only see the temp is decreasing but cant say will decreaseing in fulture year

Student responses

1. An average increase in temperature of maybe 2 degrees C.
2. For each given day in a year, I'd expect to see maybe a one-degree increase every 5-10 years.
3. I expect a slow warming trend, maybe 1 to 2°C since 1891, with nighttime lows rising faster than daytime highs. It'll probably be hard to see day to day though, since seasonal swings are way bigger than the trend.
4. I expect that daily temperature would increase by a fractional amount, less than 1 degree, every year with significant variance. This is because of global warming.
5. I think the temperature has risen obviously, but more specifically maybe like one degree each year.
6. I'm expecting the daily temperature in terms of maximum and minimum to increase by a very minute amount over the years, maybe seeing an average of .3 or so celsius over the roughly 40-year period and a decrease in overall rainfall per year.
7. Slightly increasing year over year

Student responses

1. I expect the low temperature and high temperature averages would have dropped and risen by about a degree respectively since temperature takes a lot of time to change but also because of global warming
2. I expect to observe a larger or smaller or stable number between the highest daily temperature and the lowest
3. There is a little bit of a trend, it goes up then down
4. Yes

Student responses

1. an increase in tempature due to global warming. the increase in temp may lead to increase in rain as we are surrounded by many large sources of water
2. I expect that the precipitation to get higher and I also expect the temperature to increase as well
3. The tempeature will be higher, because of the green house effect; and the precipitation will normally keep the same

Student responses

1. i expect stagnation
2. I think the temperatures will be about the same for each day throughout the years. Even if it is not exactly the same, I think the same months across years will have the same averages. warming trend.
3. it is gonna keep stable in each season

According to `a2weather`, what was the coldest temperature ever recorded in Ann Arbor, and on what date? Write the code you would use, and the answer.

92 anonymous responses

Correct answer with arrange code**35**

States the coldest temperature was -30°C on 1994-01-19 and shows or references using `arrange(tmin)` (or equivalent `select+arrange`) on `a2weather`.

Correct answer without explicit code**24**

Gives -30°C on 1994-01-19 as the answer but does not include explicit code.

Mentions a different numeric value or unit confusion**22**

Gives a temperature or unit that conflicts with the majority (different year, different temp, Fahrenheit vs Celsius, or other numeric disagreements).

Other responses**11**

Responses the model could not place reliably.

Student responses

1. -30
2. -30 C on 1/19/1994
3. -30 C on 1/19/1994. You use the arrange code to sort them.
4. -30 C on 1994-01-19
5. -30 degrees celsius on January 19, 1994.
6. -30 degrees, arrange(tmin)
7. -30 on 1994-01-19
8. -30 on Jan 19 1994
9. -30 on January 19, 1994.

Student responses

1. -30, 1994/01/19
2. -30. degrees Celsius on January 19th, 1994
3. -30C on 1994
4. 1994-01-19 -30 arrange(tmin)
5. 1994-01-19 -30 degrees C
6. 1994-01-19 and it was -30 arrange(tmin)
7. 1994-01-19 arrange(tmin)
8. 1994-01-19 it was -30 C.
9. 1994-01-19 its -30

Student responses

1. 1994-01-19 there was a -30 degree day
2. 1994-01-19 was the coldest temperature ever recorded,being -30 degrees celcius. arrange(tmin)
3. 1994-01-19 with -30
4. 1994-01-19, with a temperature of -30 degrees C
5. a2weather |> select(date, tmin, tmax, prcp) |> arrange(tmin)
6. a2weather |> select(date, tmin, tmax, prcp) |> arrange(tmin)
7. a2weather |> select(date, tmin, tmax, prcp) |> arrange(tmin) # date tmin tmax prcp # <date> <dbl> <dbl> <dbl> # 1 1994-01-19 -30 -20.6 0.3
8. a2weather |> select(date, tmin, tmax, prcp) |> arrange(tmin) 1994-1-19 with -30 degee Celsius
9. a2weather |> select(date, tmin, tmax, prcp) |> arrange(tmin) on January 19th, 1994, the minimum temp of the day wasy -30 degrees celsius.

Student responses

1. `a2weather |> select(date, tmin, tmax) |> arrange(tmin)`
2. `a2weather |> select(date, tmin) |> arrange(tmin)` The coldest temperature ever recorded is -30 degrees Celsius on 1994-01-19.
3. Jan 19, 1994 at -30C
4. Jan 19th 1994, -30 deg C. `a2weather |> select(date, tmin, tmax, prcp) |> arrange(tmin)`
5. on 1994-01-19 where the tmin is -30 degree celsius
6. The coldest day in Ann Arbor was Jan 19 1994 with the temperature being -30 C
7. The coldest temp is -30, the date is 1994-01-19, and the code is `a2weather |> select(date, tmin, tmax, prcp) |> arrange(tmin)`
8. use `arrange(tmin)` the coldest temp is -30 celsius on January 19th, 1994

Student responses

1. -20.6 degrees on January 19, 1994
2. -30
3. -30 degrees Celcius and it was 1994
4. -30 degrees celsius
5. -30 degrees celsius on January 19th 1994
6. -30 degrees celsius, in 1994 Jan 19
7. -30.6 on Jan 1994
8. 1 1994-01-19 -30 -20.6 0.3
9. 1/19/1994 and it was -30 degrees celsius

Student responses

1. 1/19/1994 in Ann Arbor, it was -30 degrees.
2. 1934-07-24 20°
3. 1994-01-19 arrange(desc(tmin))
4. 1994/1/19 the coldest date recorded (-30 celcius)
5. a2weather |> select(date, tmin, tmax, prcp) |> arrange((tmin)) January 19th, 1994. The temperature was -30 degrees F. I looked at the first row of the outputtedtibble.
6. a2weather |> select(date, tmin, tmax, prcp) |> arrange(desc(tmax))
7. a2weather |> select(date, tmin, tmax, prcp) |> arrange(tmax) 1994-1-30 -30
8. a2weather |> select(date, tmin, tmax, prcp) |> arrange(tmin)
9. According to the previous slide, we selected the temperature and then we arranged by tmax which gave us the lowest temperature recorded which was -30 degrees fhrenheit on Jan 19th 1994

Student responses

1. It was January 19th, 1994 with a high of -20 C and low of -30 C.
2. January 9th, 1994, with a temperature of -30 degrees. A2weather |> # When a command is being pulled out select(date, tmin, tmax, prcp) arrange(tmax)
3. The coldest temperature ever recorded in Ann Arbor was -30 degrees Celsius on January 19th, 1919
4. The coldest temperature ever recorded was -30 degrees celsius
5. The coldest temperature ever recorded was -30 degrees celsius. This was on 1994-01-19. I would do arrange(tmax)
6. The lowest is 30

Student responses

1. -20.6 C lowest maximum temp on jan
2. -20.6 C on january 19th 1994
3. -23
4. -30.6 degrees c in jan 19 1994
5. 1891-10-01 was the date and the temperature was -30 degrees celcius
6. 1934-06-28 19.4.
7. 1936-07-08
8. 1993-01-19 was the coldest day with -30 degrees C
9. 1994-01-14, it reached -30F

Student responses

1. 1994-01-19 at -20.6C
2. 1994/1/19. 20.6 degrees c
3. 30, 1994-1-19
4. 40.6 degrees celsius
5. 7 24
6. `a2weather |> arrange(desc(tmax))` -30c on 1994-01-19
7. `a2weather |> select(date, tmin, tmax, prcp) |> arrange(desc(tmax))` is the code, and the low was -30 degrees celcius
8. `a2weather |> select(date, tmin, tmax, prcp) |> arrange(tmax)` and then the first rows tmin which is 30 is the coldest ever tempnegative
9. `a2weather |> select(date, tmin, tmax, prcp) |> arrange(tmin)` ddnt see the date but

Student responses

1. August 4, 1930. `arrange(desc(tmin))`
2. Jan 1, 1994
3. `select (date,tmin) |> desc(tmin)`
4. September 6 60 -30 c

Student responses

1. -30
2. 1934-06-28 19.4
3. 1994-01-19 -30
4. 1994-01-19 -30 January 19, 1994 and it was -30 degrees Farenheit. I used arrange(tmin)
5. `a2weather |> select(date, tmin, tmax, prcp) |> arrange((tmin))`
6. `a2weather |> select(date, tmin, tmax, prcp) |> arrange(desc(tmax))`
7. About 40 Celsius in
8. `arrange(select(a2weather, tmin, date), tmin)`
9. `arrange(tmin)`

Student responses

1. The coldest temperature ever recorded in Ann Arbor was 19.1 degrees celsius on 1934-06-28
2. using arrange and looking at tmin, you can see that the lowest temperature recorded was -30 C on January 19 1994

Does this plot tell you about long-term trends in daily high temperatures? Why or why not?

92 anonymous responses

Too crowded/overplotted to see trends

58

Responses stating the plot is cluttered, dense, or overplotted so trends cannot be discerned.

Some trend info visible despite clutter

9

Responses that say the plot still shows some aspects of long-term behavior (ranges or min/max trends) even if messy.

Daily variability/seasonality hides long-term trend

7

Responses noting day-to-day fluctuations or seasonal patterns make long-term trends hard to detect without smoothing or aggregation.

Other responses

7

Responses the model could not place reliably.

Response themes for question 3

Recommend aggregation or trend lines	6
Responses suggesting averaging by month/year, plotting means, or adding trend lines to reveal long-term patterns.	
Axis/label or data alignment problems	4
Responses pointing out missing axis labels, unclear x-axis (dates), truncated or unlabeled points affecting interpretation.	
Other responses	1

Student responses

1. I feel like no because there are too many points on the plot to actually make an assumption. I think we can remove some of the points to have a clearer plot to then make an assumption.
2. I would say no, As visualized by the graph, the whole range of both y axis of -20 to 40 and the date range, both show a giant blackrectangle, which is super hard to see a potential trend line or slope, so this means that we can't really make a trend analysis with a generic result
3. It does not really show you the general trends because there are so many points you cant really see on a overall trend because its so crowded as well as it doesnt really show a relationship so extremes can make 2 different years look very similar visually
4. It indicates that the lowest recorded tempature has been rising over time but the plot is overcrowded
5. It jus looks awful, it would be extremely difficult to understand this data plot.
6. No because it is too over-saturated and data points cannot be distinguished from each other
7. no because the data is to condensed and hard to read

Student responses

1. No because the data is way too clustered and it is hard to establish a concrete trend. I think establishing some sort of trend line will help visualize a long-term trend
2. no because the plot is overcrowded as well as shows temperatures from different times of the year.
3. No because there are so many data points.
4. No it doesn't because there are so many points that the plot is completely cluttered, so you can't see and trends from it.
5. no its too crowded
6. No the plot is too crowded and you can't clearly see any trendlines. It might be better to colorcode and remove some of the data
7. No too much data its totally black
8. No, because according to the plot we see a massive data set, which might not be necessary when predicting long-term trends. It might be more beneficial to average out the data points so there is a more obvious trend that we can observe.
9. No, because so many values are plotted, it is hard to see and compare the daily high temperatures over time.

Student responses

1. No, because the plot is so dense and convoluted it is incredibly hard to read and draw conclusions from. a better way wouldbe to plot monthly averages or max/mins to have more readable data
2. no, because the range is too wide and overplotted
3. No, because there are so many points plotted that the graph is unreadable. The graph could instead plot yearly trends instead of daily.
4. No, because there are too many points that are crowding the plot which makes it hard to draw any trends. it should maybe be focused on a specific set of dates with a trend line to show how the temperatures have changed over time
5. No, because there is so much data on this plot it is impossible to actually see any trends in the data.
6. No, due to overplotting, because it is impossible to read and you should have an average line to show trends anyway.
7. No, it contains too many plots to be visually inspected and analysed. The graph displays no obvious pattern.
8. No, it does not since there are so many data points it is really hard to look at the trends as it looks like a slop of data on the graph right now

Student responses

1. No, it doesn't. There are too many data points to draw any conclusions.
2. No, it has too many data points and is unreadable, impossible to see any long-term trends
3. No, it is too crowded, because too many data
4. no, it is too vague
5. No, it is very cluttered and hard to read
6. No, it's too crowded to see any specific trends.
7. No, the graph is too dense to say anything substantial about long term trends
8. No, the plot is too convoluted to make any statistical inferences.
9. No, the range is too wide and results in a cluttered view. There are no viewable or understandable trends when it comes to daily temperature

Student responses

1. No, the scatterplot is extremely crowded and since no values are left out or analyzed, it is hard to tell long term trends. Lots of values are truncated and left out.
2. No, there are so many data points that look almost identical as the years continue on. There is no visible trend. A way to maybe see a pattern is to find the line of best fit.
3. No, there are too many data points which makes it very cluttered
4. no, there are too many points that we can not see any trends. it is better to look at a smaller range
5. No, there are way too many points, and none of the points are labeled, so you don't know what each point represents.
6. No, there is no clear visual trend since there are too many points. To improve we could average each year and then create a trend line
7. No, there is too many data points to be able to make a conclusion.
8. No, this plot doesn't show long-term trends in daily high temps. This plot shows every single high temp from the timeline which include 42,000 ish points (doesn't show the data well at all) A better way would be to take averages from each year and plot those

Student responses

1. No, too condensed and doesn't clearly show any reasonable trends
2. No, too many points to see any trends, need to refine the data points.
3. No, we plotted too much data, making it difficult to see what is actually happening
4. No,because there are many instances at highs of every temperature so the data gets clustered and you can't make a real takeaway.
5. No. Because this plot is hard for visualization since it contains too much information.
6. No. There is too much information to see any shape or trends in the data.
7. Not
8. Not really because there are so many values, you can't really seem to get a general pattern. It just looks like a blob with no specific shape.
9. Not really, it's hard to view any trends because there are too many points plotted

Student responses

1. Not really. It's an absurd amount of data condensed into only one visualization, you aren't able to make anything out because of it.
2. Not really. With 49,000 points crammed together the seasonal swing makes one solid black blob, so any slow trend of a degree or two is completely hidden. You'd need to average by year (or plot one season at a time) to actually see it.
3. Not very helpful
4. The plot is too cluttered to determine any long term trends.
5. There is no overall trend line on the plot, so it doesn't really tell us long-term trends.
6. This plot does not tell us much about long term trends because it is much too cluttered. To fix this we could maybe do a more random sample rather than using 40,000 points of data
7. This plot doesnt tell us much since the data points are indistinguishable from each other, making it impossible to identify patterns or even shape
8. This plot is over plotted and doesn't show much in regards to tends.

Student responses

1. It tells a little bit about the temperature as the range of the temperature between the years seems to be very similar
2. Not really but if I had to guess, would say that the temperatures have actually gotten less extreme in ann arbor as they seem to be less scattered or polarized as time has gone on
3. Not really. It can show that temperatures are getting warmer in the tmin and colder in the tmax but still not quite understandable to see any trend
4. They seem pretty consistent the long term trends
5. this plot shows that the long-term trends is not increasing or decreasing
6. Yes because you can use the trends of th maximums and minimums to get a prety good idea of the trends in the temperatures.
7. yes, it shows long-term trends because it shows the total years it is tracking.
8. Yes, this plot tells us the range of long-term trends in daily high temperatures, which is beween -20 and 40 degrees Celsius. However, I would suggest pulling means to allow for plotting less points overall and cleaning up the data.
9. Yes, you can look at the trends of the tmax and the tmin since you can see where the dots end

Student responses

1. It does not. This doesn't normalize for things like season, which just shows huge deviation in the plot.
2. It's hard to read the trend from the plot because the daily temperatures fluctuate too much. Maybe take average or make the data points less.
3. No, because daily data has day-to-day variations which doesn't show long-term trends that may require years of data.
4. No. Daily fluctuation hinders interpreting the overall trend
5. not really because it just shows daily weather data
6. We can not really find the trend from the current plot, because there is no conditions like date ranges.
7. Well at first glance the plot looks very noisy and it is hard to tell exactly what the overall trend is. I expect to see a fluctuating trend, and so maybe one thing that could be done is to remove the markers. Also zooming into the data might help more, focusing in on specific areas or zones.

Student responses

1. no because points are scattered randomly, with nothing suggesting a trend.
2. No, the data is not spread out enough to see any trends.
3. No, the plot just grabs the highest temperature and plots those
4. Not really. the graph plotted every 42000 points which made the graph look very difficult to read.
5. This plot doesn't tell use about long-term trends in daily high temperatures because with every date plotted, it's harder to visualize the general trend
6. Yes
7. yes because it is shown on the plot.

Student responses

1. It does seem to show that there is not much change, but it is too many points to see clearly. adding just the highest temp from each month could cut down on the number of points and still show a trend
2. No since it is very clustered because there are 365 days per year, maybe instead we can do yearly average
3. no, i would create a trend line rather than just look at overall data
4. Not very well, with all individual data points plotted, it obscures the trends, if we took maybe the avg per month rather than for each day, it would help.
5. Nuh uh, there's too much clutter to be able to tell. We need to take averages over a set span of time, or better yet create a line of best fit.
6. There are way too many points, that it is too difficult to determine anything. One suggestion would be to calculate the average daily temperature for each year and see if it trends upwards or downwards, or there could even be no trend at all.

Student responses

1. It does not, because it appears that all the recorded high and low temperatures from every day in a give year seems to be plotted on one x-value. It does not compare exact dates between years.
2. No because you don't know what the x-axis says and there's no dates labeled. It doesn't show any trends over time
3. No, because it deletes the most extreme data and the data left is very stable, no big fluctuation and is arond a same level
4. No. All points cluster together, we cannot see any possible patterns, and we also have missing values

Consider two scenarios: 1. Temperatures are missing because the sensor randomly fails for no reason. 2. Temperatures are missing because the sensor fails more often on very cold days. How would each scenario affect the mean temperature we calculate?

90 anonymous responses

Random missingness leaves mean unbiased

53

Students say random sensor failures do not substantially change the calculated mean (may increase variance or noise but remain roughly unbiased given enough data).

Cold-biased missingness inflates mean

19

Students say failures occurring more on very cold days remove low values and make the computed mean higher than the true mean (systematic upward bias).

Uncertain or terse responses

10

Very brief, unclear, or single-word replies that do not elaborate beyond minimal signal.

Other responses

6

Responses the model could not place reliably.

Response themes for question 4

Other responses

2

Student responses

1. Average temp may not be affected depending on variance 2. Average temp would be much warmer than the real values.
1. If temperature is missing randomly it won't affect as much 2. Colder temperatures missing means that the mean we calculate will be significantly higher than the true mean
1. If temperatures are missing because the sensor is randomly failing, it will not have as great an affect on the mean temperature because it is random 2. If the sensor does not work because it is very cold than the mean will be warped and become higher than it actually was becaue we are getting rid of the bottom numbers
1. If temperatures are missing because the sensor randomly fails for no reason, it may not affect the mean temperature too much, as if it is random, there should still be data for every temperature. 2. If temperatures are missing because it breaks when very cold, the mean temperature will be recorded as higher than it actually is, and the lower temps will e missing

Student responses

1. 1. if we dropped values then it would most likely be okay to drop the values as there is no particular favoritism in the random data 2. this would favor warmer days slightly and make the temperatures average higher than what it actually was potentially raising a high year
2. 1. It won't make a lot of difference since it's totally random. 2. The average temperature would be higher (biased) since the colder temps are not recorded.
3. 1. It wouldn't affect it as much since the data will average out. 2. It would skew the data so we think that it's warmer, when in reality, we just don't have part of the data.
4. 1. It wouldn't effect the results as much as the missing values are random 2. It would skew the data to think it is warmer because it stops working in cold temps
5. 1. no visible change since it random. 2. the mean would start leaning towards hot days due to the lack of cold days
6. 1. Random failure isn't too bad because the data points which would fail would also be random. In the cold weather days our data would skew us towards a more warm temperature on average because the coldest days are forgotten.

Student responses

1. Random failure wouldn't affect the mean temperature. 2. Frequent failure on freezing days would lead to an increase on the mean temperature since there are plenty of missing data on cold days.
1. Should be ok 2. Mean will be higher than actual;
1. Temperatures wouldn't be affected as much if the sensor failed randomly. 2. The mean would be higher if it failed on cold days.
1. Temperatures failing at random shouldn't affect the mean with a large enough data set since it's at random. 2. Temperatures missing on very cold days would likely cause the mean to be slightly higher due to a lack of very cold values.
1. The first scenario would not have much of an effect on the mean temperature because it only fails on random days. 2. The second scenario would increase the mean temperature if the sensor fails on cold days, so more warm days are recorded
1. The mean temperature should not be greatly affected. 2. The mean temperature would be higher than the actual value should be

Student responses

1. The mean temperature will be around the same since the chance that the sensor will miss the hot or cold temperature is the same. 2. The mean temperature will be reported higher than the actual mean temperature since it will miss the low temperatures and only record the warm or high temperatures.
1. The total number of observations is reduced, but the integrity of the data is not affected. 2. The data integrity is flawed because values skewed towards higher
1. This isn't so bad and our missing data shouldn't affect the mean temp as much because the failures are random. 2. This is bad because failure to record the very cold days will result in a mean that is higher than it should be.
1. this would not really affect the mean temperature since it has an equal chance of failing on any given day. 2. our mean temp would be misrepresented as warmer than the true mean because mostly cold values will not be recorded due to the sensor failing.
1. will not affect the mean 2. will make the mean temperature go up

Student responses

1. your mean calculations would not change if it is truly random since with a big enough data set it would all just average out anyways. 2. This would screw up your mean and median calculations because then you would get an inaccurate calculation since you're missing the coldest days
1. This would likely not impact the mean if when the sensor failed was truly random. 2. This would make the mean (if you took out the NA values) larger than it actually is in reality.
- 1) This would not greatly affect the mean because it would only miss a few entries 2) This would cause the mean temperature to be greater than it actually is
- First scenario, would probably not impact the average much. The second scenario would result in the average being higher than it should be, since less cold days are recorded

Student responses

1. For case 1, it would be less of an issue because incomplete data points are randomly distributed across all possible times we can collect temperature data. On the other hand, for case 2, we will need to be careful in handling null values since null values are more likely to not be present in cold days and higher mean temp will result.
2. For scenario 1, the random failing of the sensors wouldn't impact the mean much because it is not recording for random days (as opposed to scenario 2). For scenario 2, the sensor failing would result in less results on cold days showing a higher temp overall
3. if failure is random, then the mean should be essentially accurate. However, if the sensor fails more on extremely cold days, then the calculated mean will overestimate the true mean.
4. if its random, it wont affect the avg temp that much. if its consistently broken during cold temperatures, it will inaccurately skew the data to appear hotter

Student responses

1. If the temperatures are missing because the sensor fails on very cold days, the mean temperature would be higher than it actually is, because the colder days would have missing data. If the sensor randomly fails, the mean temperature would be more accurate because there isn't a specific association with certain temperature range
2. In scenario 1 the missing days are basically a random sample, so dropping them just makes the mean a little noisier but it's still unbiased. In scenario 2 you're systematically throwing out the coldest days, so the mean comes out biased too warm no matter how much data you have.
3. In scenario 1, the mean would be affected less severely because the failures are random, so in a large sample it wouldn't affect the mean as much. In scenario 2, only the lowest values would be excluded, making the mean much higher than it should be.

Student responses

1. In scenario 1, the overall mean temperature would be mostly unaffected, because every day has the same chance for the sensor to fail. In scenario 2, the overall mean temperature data would be higher than it actually is, because the sensor fails to capture the colder days.
2. in scenario one the mean would not be different because the failure is random and hits any day indiscriminantly, in scenario two the mean would be higher than it should because the sensor fails on the cold days so they are left ou of the data.
3. In the first case, the missing temperatures would likely be randomly spread across the range of data due to their procurement being random. However, in the second case, the missing temperatures would give a bias to the data because they would be on the lower end of the data, so the mean temperature would be higher than in actuality.
4. Mean stays roughly unbiased/accurate1. Fails more on cold days: Cold temperatures are missing → calculated mean is too high.

Student responses

1. Scenario 1 shouldn't affect the mean temperature value, Scenario 2 will increase the mean temperature by removing the lowest readings
2. Scenario 2 because the missing temperatures would all be lower and make the mean seem larger than it really is. Scenario 1 would be ok for the mean since it's random.
3. the first one would not affect the mean temperature that much if it is random. but for second scenerio, the mean temperature would appear higher than what's actually the mean since more colder temps would not be recorded
4. The first scenario should not have a strong effect on the mean temperature, because completely random fails would be both low and high values, meaning it shouldn't skew the mean. The second scenario would make the mean temperature incorrectly higher, because the sensor fails would remove many more low data points than higher ones.
5. The first scenario won't affect the result mean much because the missingness is random. The second scenario will make the mean higher because the data are biased, with the lowest values missing.

Student responses

1. the first scenario would cause temperature data to be less accurate because some is missing. the second would cause the average temp to be artificially higher
2. The first scenario would likely have little effect on our data summary and analysis. The second scenario would create a bias that leads us to believe that the average temperature in a given year is higher than it actually is.
3. The first scenario would mean that our mean would be partially correct because the omitted data would be random. In the second scenario, our mean would be higher than what it actually is because the sensor omitted the coldest temperatures.
4. The first scenario would not affect the mean temperature, as if it is truly random, then in the long run the missing values would naturally average out to what the mean temperature is. In the second scenario, colder temperatures are more likely to be missing values, so the mean temperature would be slightly higher than it should be because it is missing cold values more often

Student responses

1. The first scenario would not effect our analysis as it is random chance and does not add bias to our analysis. However, in the second one we would have to consider it as it would be biased towards warmed temperatures and be right skewed.
2. The first scenario would not have much of an impact on the data however the second one would push the averages higher than what they should have been.
3. The first scenerio will have randomness so it might not affect the mean as much as the second scenrio where average will be higher than expected because the low values aere ignored
4. The first shouldn't affect the mean (since the missing values are truly random) but in the second the mean would be much higher, since we are missing super cold values
5. The first situation wouldn't change the mean of the temperatures that much whereas the second scenario could drastically change the mean because if all recorded low temperatures are removed the caverage will be way higher than the actual average

Student responses

1. The mean temperature would remain roughly the same in the first scenario, but in the second, the mean temperature would be higher than its actual value.
2. The second scenario would affect the mean to a greater extent because if we fail to acquire data from very cold days, we are essentially left with data that is inaccurate, whereas in the first scenario, we might be missing data on both cold and warm days, which would less affect the mean.
3. This scenario would impact all temperatures in the same way, therefore not impacting the mean much. This scenario would impact the mean, as it would not include the coldest day. This skews the mean to be hotter.

Student responses

1. 1 would not affect the mean, 2 would inflate the mean
2. 1. If it was truly random then it wouldn't have that much effect on the mean temperature. 2. If it only broke on very cold days then the mean average temperature would be higher.
3. 1. it will affect the overall pattern, considering this could be a random variation 2. it will cause the overall temperature pattern to be higher.
4. 1. The first scenario I would say doesn't necessarily change the mean too much if not at all. This is because this is happening at random and random distribution can affect any data, so I would say that it is no problem, but for the second one I would say that the mean will be skewed higher than it is because a certain values of colder values will be neglected and bias affect the data set.
5. 1. This will randomly affect the mean temperature, as these errors will either increase or decrease actual mean but we wouldn't know which way as it is random. 2. This will increase the mean temperature as it is not recording temperatures on colder days

Student responses

1. This would affect the mean temperature marginally in an unpredictable way. 2. This would increase the mean temperature as there is less low-temperature density.
1. We don't know how the mean would change because it is random failure. 2. The mean temperature would probably be higher because on cold days, it is not tracking the temperature properly
- 1) This scenario could affect the mean temperature by either being too low or too high, however there isn't a clear reason for the failure. 2) This scenario would affect the mean temperature by making the mean too high.
- case 1: probably wouldn't affect the dataset, case 2: would eliminate low values
- First can be ignored, but second one can't because it is a special case extreme one
- For 1. the standard deviation would be greater. For 2. the standard deviation would be lesser and the mean temperature would be higher than it actually is in reality
- If it is missing because the sensor failed on cold days then the average we are calculating would be higher than the true average of all the temperatures

Student responses

1. In the first one you would just miss data randomly, while in the second it would skew the data making your conclusions think there are more hot days.
2. in the first scenario, it would be a random drop of data, which in theory would not affect the mean. In the second scenario it would make our mean seem higher than the tru value actually should be
3. In the first scenario, since it is random, the high temperatures and low temperatures have an equal chance of being rmeoved from thr dataset, so the std would decrease with the same mean. however, if the cold temepratures are removed more often, then the data is going to be skewed to be warmer
4. Scenario 1 would affect the temperature because we would have NAs randomly in the data but Scenario 2 would affect the mean temperature because we would see more NAs show up in colder days
5. Scenario 2 would affect the mean more since it is an extreme value that can pull the mean lower

Student responses

1. The first scenario should cause significant change to the calculated mean across a given time period, since hot and cold readings are equally likely to be removed. The second scenario will result in mean that is higher than the actual mean temp, since colder days are removed
2. The first would not affect the mean a lot. The second would cause the mean to be lower than expected

Student responses

1. 1
2. 1. mean temperature should consider outliers otherwise it would be greatly rffect
3. 1. random missingness 2. cold missingness
4. 1. the mean wouldn't really change because as long as temperatures arent extreme the amount of values will account for this missing one 2. Extreme temps raise or lower the mean so it would hange a little i included
5. 1. The temperatures here are generally irrelevant to the whole dataset, given that they don't signify anything in specific. 2. The fact that incredibly cold days are missing from the dataset and ONLY cold days means that the year's mean will be lower than if they were preserved.
6. For
7. If its 1. then there is likely no significant difference in mean high temp If its 2. then there is a underrepresentation of daily high temps on the coldest days, meaning there is a significant upward skew on mean daily high temp
8. In the scenario with random missing temps, you couldn't really predict the effect on the mean. In the scenario where Is missing on colder days, can think high meN
9. ji

Student responses

1. The second. Because if the sensor randomly fails it won't affect the lower without affecting the higher, it's more justice

Student responses

1. 1. Dependin on the temperature the mean will change but slightly. 2. It would drop the mean temperature since the sensor fails often on very cold days
2. 1. random effect. 2. lead to high mean
3. 1. We can safely ignore the missing values because there is an equal chance 2. do more work,I
4. For the first scenario, it would affect the data because maybe it skews the data since it skips a day. For the second scenario, that would really screw up the winter data because it only fails on certian cold daus
5. the first one means that we are missing random data, the second senario means that we are missing significant portion of our data as we do not record the cold days
6. The first scenario wouldn't really affect anything. The second one may cause the average temperature to be skewed because it is only happening on a specific type of weather day.

Do you see any outlying years in the plot above? What do you think could have caused them?

93 anonymous responses

First-year/early-year outlier Identifies the earliest plotted year (often cited as 1891/1892 or the first data point) as an outlier.	47
Missing or insufficient data for that year Attributes the outlier to missing values or too few samples in the first/early year, making its mean unreliable.	15
Partial-year recording or started in October Explains the early outlier by noting the dataset began late in the year (e.g., October), so averages reflect only colder months.	15
Measurement error or faulty sensors Claims instrument inaccuracies, glitches, or older technology caused the anomalous early value.	13

Response themes for question 5

Other responses

3

Student responses

1. 1891 and several years after that, I think they might be caused by imprecise methods to record or way to collect data.
2. 1891 is an outlier as it's mean tmax is much lower than any other year. Maybe the sensors were faulty
3. 1891 is an outlier I believe based on when they started to collect the data maybe towards the end of the year
4. 1891 is an outlier year with the temperature being very low, this may be due to the first recorded temperature starting that year (year 1) was in the winter so all measurements that year were cold
5. 1891 is an outlying year. Maybe during the first year of data collection, they had a lot more missing data as they were figuring out the process of collection.
6. 1891 looks like an outlier as it sits far below the rest of the values, this may be due to it being the first year and having inaccurate data.
7. 1891 looks to be an outlier because it is so low. I believe this could have been a result of the university starting their data collection partway through the year and therefore not having complete data for 1891

Student responses

1. 1891, or whenever the first year of this dataset began. Perhaps it could have been due to a recording error, or maybe due to large amounts of missing data?
2. I see an outlying year in the first plotted point, which is a much lower value compared to the others. This may be because there were many missing values in the first year, so the average isn't as accurate as the following years that had more values recorded
3. I see one outlier as the very first entry 8.6 deg C, this is typical of R to plot the lowest possible value first. This is a clear outlier in the plot
4. I see one outlier at the very beginning, likely caused by data collection being started in the cold month of October for that year.
5. It looks like there is an outlier at the very beginning of the data. This is because the data started to be recorded in October, so it led to a lower average temperature
6. Outlier was the first year the data was recorded. This was likely caused by the data initially being recorded in October, so it only contained temperatures from Oct-Dec, thus making it look much colder than the other years.

Student responses

1. That first datapoint (year ~1900) is outlying. I suspect that it's because of a low number of samples taken that year, such that they may have just been taken in winter.
2. The first data point, maybe due to certain climate events at the time.
3. The first point is an outlier.
4. The first point of the annual mean plotted is an outlier a number right above 8 while the others are above that. I think that older data might have some things missing, especially when simplified
5. The first recorded year (1891) has an unusually low mean temperature. This could be because not many temperatures were recorded, and the ones that were recorded were on colder days.
6. The first year (1891) is an outlier because it was only recorded starting in october, leading to a lower temperature being reported than what was actually the case
7. The first year (1892) is reported significantly lower than the other years. Could be because of missing data since it was the start when it began recording temperature data.

Student responses

1. The first year appears to be an outlier on the low end, one reason that might have caused this is that we may not have the full year of data, so if we only have the fall and winter months from this year then the average daily temperature would be lower.
2. The first year has a much lower average. That year started tracking in October.
3. the first year in the data is a huge outlier. It could be due to a change in recording technology or maybe they started recording the data not at the start of the year
4. The first year is extremely low, maybe because the recording started late in that year.
5. the first year is probably an outlier. it started collecting data in october which means there is only winter data for that year which makes the average a lot lower than expected
6. The first year measured is an outlier and it could be because the measuring started during the winter so the summer did not boost the average
7. The first year, 1891, seems to b a low outlir. This could be due to an imperfect sensing method, or only recording part of the year.

Student responses

1. The first year, which could have been some time in the 19th century shows the only outlier. This is likely due to a lack of advancement in recording technology.
2. The first year: 1891; i think it's cause this is the first year, and we have a lot of missing data that is hot or days in this year might be cold
3. The mean daily high for the first recorded year seems to be much lower than the other years. This could be due to them starting recording these temperatures later in the year, so the average doesn't include the summer months
4. The obvious one is 1891 at about 8.5°C, way below everything else. That's not a real cold year though, the record starts in October, so that "year" is only fall and winter days and the average gets dragged down.
5. The one outlier year I see is the very first year, 1891. This could have been caused by the date meteorologists actually started recording temperatures within the year. Starting late in the fall would lead to a dataset skewing toward colder temperatures for that year.

Student responses

1. The outlying year is 1891 and the reason for it is that the data collection for that year started in October, meaning mostly colder data was collected
2. The very first point around 1900 at 8 C is very low
3. the very first year has a very low responses thus outlying
4. The very first year seems a lot lower than all the other years It is possible that the data began to be collected after summer, so it didn't have as many highs
5. The year 1891 has a substantially lower average temperature. A possible reason could be that the station was completed halfway of the year so not all 365 days were recorded. Moreover, it was probably built in fall or winter so the hottest days weren't included, thereby lowering the mean temperature of the year.
6. There is an outlier in 1891, which is significantly lower than other years. This is possibly because this dataset started in the fall/winter that year and only recorded days with lower temperature.
7. There is one outlier at the beginning of the plot, which could possibly be caused by inaccurate temperature measurement methods or environmental hazards.

Student responses

1. There seems to be an outlier in the first year plotted. This could be because it started in October, so we don't have the full year of data
2. We see that the first year has an unusually low mean. That is likely because the data collection began on October 1 of that year, and the mean temperature from October 1 to December 31 of any year will be lower than the year-round mean.
3. Yes - I see an outlier in the first year of recording this data in Ann Arbor. Maybe the experiment being new could've caused this?
4. Yes, 1891. Considering it was the first year of weather being recorded I think it is reasonable to assume advancements in technology were made after this first year or the uniiversity found a better way to record the weather.
5. Yes, the first year is extremely low. Maybe the data collection only started towards the end of the year when it wa colder
6. Yes, the year 1891 has an outlier. They could've started the data in the last several months, leading it to seem colder than the rest of the years. It could also have a lot of missing data since it was so long ago.

Student responses

1. Yes, year 1891 is typically low compared to the overall other years. The reason might be that we don't have the whole year data of 1891, so maybe only the winter and autumn data are recorded.
2. Yes. I think that there are outliers in the first year of the data. They have extremely low average values when compared to the rest of the data. I think this is probably because the collection of data did not start on the first of the year but rather later in the year when it was already colder

Student responses

1. 1892, much lower than even the next year, we dropped missing value, and there is more frm
2. An outlier could be the daily temp being 8.5 C before the 1900s. This could have been a particularly cold year, bad recordkeeping, or glitch in temperature reading.
3. The data point in the beginning seems to be too low compared to other data points later on. Likely poor measurement, or even a miscalculation of the Celsius temperature.
4. the first data point, because it was the first year there were probably a lot more missing values, making it less accurate than the actual average
5. The first year is an outlier year. Lots of missing values could have caused this.
6. Too low around 1930
7. Yes because the cold weather in one year is extremely long
8. yes I say an outlier around 1890 and I think maybe the little amount of data collected back in additon to the season might have caused that outlier
9. Yes there are outliers, for example, year 2012 is higher than the neighbours and year 2014 is lower. This may b because extreme weathers or some rare pheonomeon.

Student responses

1. yes there are some outliers. this can just be normal statistical anomalies
2. Yes there is an outlier of around 9 in the late 1800s, and I would assume its because of missing data during the hotter months of that year
3. Yes, I see a couple of outliers in the data. I think they could have been caused by issues with missing data or inaccurate reporting.
4. Yes, the first year. I think what caused this was that it was the first year that data was recorded so the device wasn't accurate
5. Yes, the first year. It could be that the reporting in the early 1800s wasn't as accurate and a lot of data got removed, leaving the average at a way lower spot than the other years.
6. Yes, there is a significant outlier around the beginning of the data circa 1900 or so. It could be caused by a sensor failure or extreme storm.

Student responses

1. 1880 is a outlier. There might not have been the ebst devices to track this information accurately
2. An outlier could be around the year 1890 where temperaures were very low. maybe they couldve started data collection at a later month which messed up the average calculation
3. first year was, because data for it was only partial
4. I see an outlying years at the very beginning of the plot, I think it is because the data collection is not sufficient at that yer
5. I think the first value ar arround year 1910 is na outlier it only has a temperature of 8 . I think it was cuased becaused only certaitn days wwere recorded and I would assume it only recorded wintere temperatures and not sumemr termperatures , only reocrded temrperatures later in the year.
6. One outlyer I see in the plot is the first year. This could have been caused by misrecordings of the temperatures, or misreadings of the recordings.
7. the first entry seems like an outlier--likely aused by the recording starting late in the year

Student responses

1. The first value is very low compared to the rest of the means, I think this could have been caused by measurements starting in the fall/winter leading to lower average temp for the year
2. the first year because the temperature was very low, and i think what caused it is that we only started data collection for that year from october, meaning that all the temperature collected are during winter, leading to colder averages.
3. The first year, because the data collection started in October, so the first year only has 3 months of data versus 12 months of the year.
4. the very first year has a very low mean, this might be because they started tracking the temp at the end of the year in the winter months
5. The very first year recorded is an outlier, likely because the data collection started in the colder months (october), and so that year's average consists of only fall and winter temperatures.
6. Yes, at the very beginning of the plot the tmax is super low. This could have been caused by a really weird temperature day or the temp reader malfunctioning.

Student responses

1. Yes, in the first eyar of data there is a extremely low oulier. this could because the radar tha was used to track temperatures wa gltichy in the first few years
2. Yes, The first year is an outlier year, it appears to have data that is skewed very cold towards 8 celsius, but I think this is because I remember the data only starting in october and traditionally the last three months of the year are very cold so it makes sense that the data is seen as colder

Student responses

1. I do, the beginning point. If I had to guess, it is because the data starts at 0?
2. I see outliers. caused by miscalculation in data
3. Measurement error
4. One, it changes the mean of the graph
5. The first year was abnormally cold, and 1963 was abnormally warm. Conditions like El nino, wildfires, or other weather events could have caused these years.
6. There's an outlier as the first point, where it's a very low value. Possibly sensors weren't as good back then?
7. yes due to
8. Yes I do see some outlying years. I think that data collecting issues among the year causes those drastic outliers
9. yes the first years are major outliers and are extremely cold. possibly poor technology that incorrectly found temperature could be at fault here

Student responses

1. Yes there is an outlying year at some point in the late 1800s and this is probably due to the lack of technology that was available to properly track the temperatures during that time
2. Yes, in 1892, and this could be because of inaccurate data collection or because they only started recording in the winter of that year as it is the first year.
3. yes, the first year has a very low average, it is possible that they started recording temps at the end of the year, and so that avg only includes low winter temps
4. Yes, there is an outlying year near the beginning of the plot. Possibly there was some issue with the measurement of the data.

Student responses

1. Yes
2. yes i think a lot of things caused it
3. yes, it might just be a very cold weather or missing data

Use the table above: what explains the outlier that we saw in the plot?

92 anonymous responses

Partial year / only 92 days recorded	59
Responses noting that 1891 contains only about 92 days of observations (data collection started late, e.g., October) so the year is incomplete.	
Seasonal bias from late start (cold months)	19
Responses explaining the outlier as due to the recorded days being from colder months (fall/winter), lowering the mean.	
Other responses	7
Responses the model could not place reliably.	
Vague or unclear answers	4
Responses that are non-informative, single letters, or vague remarks.	

Response themes for question 6

Missing / incomplete data generally

3

Responses that mention missing observations or incomplete dataset without specifying count or season.

Student responses

1. 1891 full year was not measured on 92 days were. So it is not an accurate average of the calender year
2. 1891 has missing data days, missing data leads to an affected mean that may be inaccurate.
3. 1891 has only a portion of the year recorded, likely in th colder months.
4. 1891 is weird because the measurements started later into the year. typically the yearswould have 365/6 measurements, but this eyar only has 92, meaning it would naturally be lower because of the season they started in
5. 1891 only contains 92 points, which may suggest only partial year coverage, which may have been fall/winter.
6. 1891 only has 92 days instead of a full 365, since the station didn't start recording until October. So that mean is just Oct through Dec, which is why it comes out at 8.42 instead of a normal full-year value.
7. 1891 seems to be missing a huge amount of reported dates. This means that it probably starting reporting after the summer months and has a dissporportionate cold weather due to that.

Student responses

1. 1891 was the outlying year and this was the first year that the data started being tracked. the table shows that the sample size is smaller so the mean highest temperature wouldn't be accurate since it does not account for the entire year
2. in 1891, only 92 days of data are collected instead of 365 days
3. In 1891, only 92 days were recordered. However, the rest of the years, the sample size was arounfd 365 days.
4. In 1891, the average was computed with a significantly smaller demonination
5. In 1891, the number of days in which this average was computer is only 92 days which means the data was cut off that year at a certain point.
6. Incomplete dataset, year 1891 only have 92 days of data, which may consists of only winter or autumn datas.
7. It looks like the outlier is coming from the first year of data and the data started in october, not january. 1891 has 92 days of temps recorded and the rest of the years have 366 days.
8. it was recorded that year only partially
9. less days were recorded in 1891. These days were likely in the fall or winter when it is colder, pushing the mean tmax down.

Student responses

1. lot of missing values in 1891, started recording very late int he year
2. Not a hole year data for 1891
3. number of days in 1891 was icomplete
4. Number of days recorded in 1891 is only 92, which probably represents days in colder months and ignores the summer ones
5. only 92 days
6. Only 92 days had recorded values in 1891 compared to around 365 for the other years, which explains the outlier, as there is a lot of missing data for 1891.
7. only 92 days were captured in that year
8. Only 92 days were recorded in 1891.
9. only 92 days were recorded meaning they started at the end of the year in the wintermonths

Student responses

1. Only 92 days, roughly one season, was recorded in the dataset. The season can be assumed to exclude warmer seasons (perhaps only being the winter or fall), thus creating a lower average.
2. Only the last 92 daily temperatures of 1891 were collected.
3. Rather than getting all of the data for a full year, 1891 only has 92 days worth of data and therefore is not representing that year well.
4. Start date was October.
5. the average was only computed over 92 days instead of the full year
6. the data was recorded in the first year starting in october so the average is from only the last 92 days of the year which are generally colder than the rest of the year.
7. The fewer number of days that have recorded data in 1891 explains the outlier. It started later in the year and has a colder temperature bias.
8. The first 1891 point only has data from 92 out of 365 days
9. The first year does not showcase a full year but rather only represents the last 92 days of 1891.

Student responses

1. The first year had less days for the average, while the rest of the years used all the days in the year.
2. The first year only has 92 recorded temperatures instead of around 365 like the other years.
3. The lack of days data was sampled explains the outlier we saw in the plot, for only 92 out of 365 days had data collected. Those days were in the colder months.
4. The number of days recorded in 1891 is significantly lower than the other years. It could have reported the last 92 days, which will be way lower temperature than the actual mean since it will record only the winter.
5. The number of days that the average was computed is significantly lower than the typical year: only 92 days.
6. The number of days which weather was recorded in 1891 is significantly shorter and indicates a seasonal bias.
7. The number of entries(n) for 1891 is only 92 days, usggsting that only part of the year's weather was recorded
8. The outlier is likely corresponded to only 92 days being reported in 1891, rather than 364-366

Student responses

1. The result of this outlier is likely due to the fact that the sample size is too low, which results in more sample variation in the true estimate.
2. the sample size is too low in 1891 which probably caused a bias towards a lower temp
3. The year 1891 doesn't have enough data only 92 days. That is why there's an outlier in the first point.
4. There are far less entries for 1891's max temperature than the rest of the years.
5. There are less total days in the first year because we started late
6. There are only 92 days that were recorded in 1891, which means only about the last 3 months that has lower temperature were included in this dataset, causing an lower end outlier
7. There are only 92 observations for 1891 which likely doesn't account for warmer weather.
8. There are only 92 reported days for 1891 compared to at least 365 for other years
9. There is a different number of days that the average was computed by. it is much smaller than the normal 366/365 days

Student responses

1. There were fewer days of data collection in 1891.
2. There were only 92 recorded days of temperature, leading to a more volatile standard deviation.
3. They only measured 92 days out of 365, which could've been all in the winter.
4. They started most of the way through the year.
5. We had fewer days for that year, likely the collection did start late, in winter
6. We only have 92 days worth of data for 1891, so we are missing data from a lot of the year causing the average daily temp for this year to be skewed very low
7. We only tracked 92 days in the first year, which were likely the last days of the year and since that's fall and wintertime it's colder on average than the rest of the year, leading to a colder overall temperature and lower max.
8. What explains the outlier in the first year is that the temperatures were not recorded for the full year, only 92 days

Student responses

1. Days that have a high t_{max} aren't recorded in those 92 days.
2. In the first year, the temperatures were only recorded from october onward
3. It shows that only 92 days are recorded hence it deidnt reocrd warm temperatures and have a much lowere mean then the other values.
4. less days and only in a specific part of year where temps differ from other parts that are accounted for in different years
5. Missing observaions from n_{days} . Only the later years of data were recorded.
6. The amount of days is lower and was probably during winter and didn't include summer months.
7. The data taken was only from the winter season
8. The days are only 92, only winter is recorded
9. The first date is in October, so it is missing 9 months of data.

Student responses

1. The first year only had data from the last quarter of the year, missing the warmer summer months
2. The first year only includes the colder winter months.
3. The outlier is that there is only 92 days in the year of 1891 that were recorded as opposed to every other year that has almost all of their days, which shows that it was skewed and only contained the latter months and had a mean that was colder.
4. There weren't as many entries in the first year of the data, may have just been a really cold day or couple of days that was recorded in the first year of the study
5. They are missing a bulk of the days of the year; perhaps they started later into the year.
6. They started recording in October
7. we only had 92 days meaning we collected just the cold months of the year
8. we only started collecting data later in the year so only the colder months are recorded which would lower the average than expected
9. We only started recording this data late in 1891, and so only in winter making the average very low.

Student responses

1. yes, probaly a late start

Student responses

1. fewer days recorded, mor likely for variane
2. In 1891 it seems like measurements started at the end of the year in winter, which means the mean temperature would be much lower
3. it was the mean that explains
4. less days are being recorded
5. that year istarted at the end of the year
6. The data lacks the full year of temperatures for 1891.
7. we only have data during winter

Student responses

1. E because it doesn't have enough data set
2. the number of days that collected data
3. Low readings
4. V

Student responses

1. Missing data
2. Outliers are due to a lot of missing data, missing days
3. the data was not recorded throughout the entire year

How would you use these daily weather data to decide whether extreme precipitation has become more frequent? Think on your own, then compare ideas with your group. Submit a plan with 3–5 steps, in order. What would you measure, compare, and check? You don't need to write R code yet.

90 anonymous responses

Define an extreme-precipitation threshold

46

Set a clear numeric or percentile cutoff (fixed baseline period, percentiles, absolute inches/mm, IQR-based fence) that classifies a day as extreme.

Count extreme days per year and plot trend

12

Create a yearly count or percentage of days exceeding the threshold, summarize by year (or decade), and visualize to detect increasing frequency or trend.

Other responses

12

Responses the model could not place reliably.

Use summary statistics (means, SDs, outliers)

8

Response themes for question 7

Broaden to multiple variables or event types	6
Compare aggregated periods (decades or bins)	3
Other responses	3

Consider other measures (total annual precip, seasonal patterns, tmin/tmax, heat/other extremes) or classify multiple types of extreme events.

Group data by multi-year blocks (decades, 5-year bins) and compare averages or counts between periods to assess change.

Student responses

1. Check each year for missing prcp days so gaps don't look like fewer storms. 2. Pick a cutoff for "extreme," like the wettest 1% of days. 3. Count how many days each year go over that cutoff. 4. Plot those counts by year and see if recent years are higher. 5. Try a different cutoff to make sure the pattern still holds.
1. Count days per station-year and drop station-years with too much missing data, so gaps don't fake a trend. 2. Set the extreme threshold from an early baseline period (e.g. 95th percentile of wet days) and hold it fixed for all years. 3. For each station-year, count days above that threshold, plus total precip and wet days. 4. Regress the annual count on year (Poisson or negative binomial, station as a factor) and report change per decade. 5. Rerun with a different threshold, baseline, and
1. Create a threshold for extreme precipitation (ie 24mm) 2. Identify # of days within a year that meet the threshold 3. Per year make a percentage of days that met threshold 4. Have a graph of years vs percentage of days with extreme weather

Student responses

1. 1. decide a level to check extremeness 2. count how many days exceed the threshold 3. plot over time
2. 1. Define a threshold for extreme precipitation 2. Filter the days that raining amount exceeds this bar 3. Calculate the number of those days in each year 4. Study the trend of this number across years 5. Maybe use some test and calculate p value and compare it to 0.05
3. 1. define extreme precipitation (> some number of inches) 2. count days (per year) where there was extreme precipitation 3. graph points and look for trend
4. 1. Define extreme precipitation 2. Measure all precipitation 3. analyze the data
5. 1. Define what extreme precipitation is, maybe find values that lie within the 95th percentile of all the recorded data. 2. Once the cutoff value is determined go through the data and get the total number of extreme values. 3. Once you have the total number calculate the number of extreme instances every decade or every 20 years.

Student responses

1. Define what extreme precipitation is 2. Filter for only days that have above our threshold 3. Count number of extreme precipitation days for each year 4. sort by year 5. graph
1. Define what extreme precipitation is. 2. Count how many days each year have reached extreme precipitation. 3. Plot the data set
1. Define what extreme weather is a. Perhaps 90th percentile weather over the first 10 years of the data, then apply that to the rest. 2. Apply that to each year, see how many days fall above that line. 3. Check in with client to make sure that our definition and findings are reasonable.
1. Determine what is extreme precipitation for Ann Arbor. Maybe a certain percentile? 2. count the number of days, grouped by year 3. compare the number of days for every year 4. use arrange(desc) to see the years with the most extreme precipitation

Student responses

1. 1. first determine what extreme precipitation is, generate the avg precipitation per year, above avg is extreme. 2. For each year, get the count of how many days have extreme percipitation 3. Plot the counts, if increasing trend can decide that extreme percipitation is becoming more frequent
2. 1. Group by years 2. Determine a threshold above which the precipitation is considered as an extreme 3. Count the number of days that go above the threshold for each year 4. Draw a plot and watch the trend.
3. 1. I woud define what "extreme" means and find a cutoff for extreme precipitation 2. I wound then count the number of extreme precipitation days there are per year to establish a trend 3. I would then compare early years and recent years to see if the extreme precipitation has become more frequent over time
4. 1. I would research and define what number precipitation is considered extreme 2. then I would filter only these values out as well as include dates 3. I would then group by year 4. I would then make a ggplot and look for trends in amount of extreme precipitation days year to year

Student responses

1. 1. Make a table of all recorded days, filtered in order of how much precipitation there was that day 2. Choose a threshold of precipitation that is considered extreme 3. Analyze if there are more "extreme" days in th past 20 years compared to the rest of the recorded data
2. 1. Numerically define 'extreme' precipitation, for example more than two inches of precipitation per day. 2. Make a table with each year as a row, with a column containing the number of 'extreme' precipitation days 3. Make a plot with the year as the x axis and the number of extreme days as the y axis 4. Observe if there is an increasing trend.
3. 1. Select a value that would be the minimum to qualify as extreme precipitation 2. Use the filter to filter the data to days that have extreme precipitation 3. Organize the data by year and use the summarize function to summarize how many days per year have extreme precipitation 4. Plot the data to see whether there is a trend.

Student responses

1. 1. set the threshold for extreme precipitation 2. filter the rows with precipitation value higher than the threshold 3. count how many years in every decade and compare
2. 1. Sort the precipitation of all days 2. Calculate the mean, and quadrants of precipitation. 3. If precipitation $> \text{mean} + 1.5 \cdot \text{IQR}$ it's extreme.
3. 1. We would first have to define what "extreme" means and make it a threshold. 2. Then we would have to compare the precipitation numbers to that threshold to see which days exceeded the threshold 3. Count how many days there are per year with extreme precipitation 4. Then track how the number of extreme precipitation days have increased over the years
4. 1. Every ten year's mean rain level 2. Rain level by years 3. Do a label
5. 1) Decide what level of precipitation is considered "extreme" 2) Records the precipitation for rainy days and gather the amount of days per year that hit the extreme amount 3) Plot our findings on a graph along with the summary statistics

Student responses

1. Average the precipitation levels by year, sort by year ascending, come up with a threshold of what constitutes as extreme precipitation, and then see if it continually goes up by year or if there are any trends
2. Count the days where the precipitation is above a certain threshold for each year. Graph the number over the years.
3. Define an "extreme" amount of precipitation. Then find total per year. Graph the year and compare the years extreme precipitation days and compare between the years.
4. Define what 'extreme' means, find the dates when the rainfall falls into that category, and show the number or percentage of extreme days per year.
5. Define what constitutes an "extreme" precipitation day, Then add a column for each day which signifies if a day is an extreme day, then get the total per year and plot how many extreme days there per year and see if there is a trend there
6. Define what counts as extreme precipitation. Count the number of extreme precipitation days each year. Compare the counts across years. Check whether there is an increasing trend over time.

Student responses

1. define what extreme precipitation is; ex: mm>10, filter down to those rows and then make a ggplot that shows the trend of extreme weather days over time.
2. define what extreme precipitation means >some number. then by year, average out how many appeared. then, plot these averages and see if there are more. then, check for outliers/missing data
3. Define what extreme precipitation means, maybe 20-30mm of precipitation, reduce the dataset to only values that meet this requirement, sort data by year, compare number of extreme precipitation days in each year against each other
4. Figure out a threshold for determining if a day has extreme weather, figure out how many extreme weather days there are per year, and then plot the total number of days to see if its increasing or run tests on it

Student responses

1. First we need to set a numerical baseline for what exactly is extreme precipitation. After that, we can see how many times precipitation exceeds this baseline every month or so, and then data can be aggregated to then be analyzed further for a positive or negative trend. A positive trend would mean more frequent extreme precipitation.
2. First, define what "extreme" means (how much is classified as extreme), then filter the data to show only the days that would be classified as extreme. Then, you can graph the number of extreme precipitation days per year.
3. First, I would decide what level of participation in a day is considered extreme. Then, I would create a variable to discern how many of these days there are each year. Then, I would sum up each year and plot the value
4. I would calculate the mean for each year, calculate the standard deviation, and determine how many days per year are 2 standard deviations away from the mean. I would then compare across years or potentially decades the total count of 'extreme' weather days to see if they are more frequent now than earlier in the data set.

Student responses

1. I would find the number of days over 30 inches of precipitation. I would find the number of days per year and per decade that fit this description.
2. I would first choose a threshold and set that as the extreme value and then compare the precipitation to the extreme value and count the number in each year and then compare the number of extremes between all the years to see a trend
3. I would first determine what extreme precipitation is, maybe more than 6 inches of rain could be extreme. I then would calculate the frequency of days that had more than 6 inches each year. And then I would plot the data to see if the frequency of days that had 6 inches of precipitation each year is rising.
4. if it rains more than 5 inches over the course of 24 hours, i would consider it as extreme precipitation
5. see the precipitation amount of each day, define what is frequent and define what is extreme precipitation; eg frequent = days ≥ 50 ; extreme = amount $\geq 50\text{mm}$

Student responses

1. Start by removing NA rows, compute summary statistics next examine values that are 1.5 times the upper quartile and note these entries as extreme precipitation.
2. the first step is to classify what is considered extreme weather and then count the frequency of those days and how much they have changed throughout the years. I would measure average rainfall in each year to determine extreme weather thresholds by year.
3. The first thing that I would do would be to define what extreme precipitation actually means by looking at different quartiles of data to identify these ranges. I would then work on sorting out my data in order to get averages and determine where they were compared to these extreme values

Student responses

1. 1. Collect data from previous years' weather. 2. Define what extreme and frequent means. How much it differs from the mean. If it is above or below the average by 10C, then it is extreme. 3. Count extreme days per year and plot the trend
2. 1. define extreme precipitation 2. calculate the number of days each year by grouping by year for extreme precipitation 3. Compare it to other years / decades.
3. 1. Define extreme precipitation threshold 2. Create a new column if precipitation > the threshold ture or false 3. Sum the days over that extreme precipitation in days. 4. Plot the days w extreme previpiation aevery year and see if there is an upward trend
4. 1. get precipitation day by day 2. sort them in a year 3. draw a ggplot (line plot) to see the patern
5. 1. measure the daily weather 2. put them in a table 3. compare them 4. measure the data above or below to decide extreme precipitation 5. use the percentage change to decide whether extreme precipitation become mo9re frequent

Student responses

1. Measure the precipitation for all recorded dates 2. Compare missing data and determine reason for missing data 3. Average precipitation over timeline and visualize trend 4. Compare precipitation values to historic averages to classify weather days as "extremeprecipitation" and count those days
2. filter the days to include the ones with precipitation above a certain amount, then group the data by year and count the number of days in each group, then compare between the years.
3. First, check the table and clean it up if there are missing data values. Second, group the data by years, every 5 years, or every decade, and find the average precipitation level. Third, Look at the data and see if there are any trends that would point to more extreme precipitation levels.
4. Flag days with extreme precipitation (after deciding a threshold for "extreme"). Then, count how many extreme precipitation days each year has. Lastly, plot the number of extreme precipitation days for each year to see if there is an upward trend.

Student responses

1. is precipitation in the top 75% category of rain or whatever, group by years, then count the amount of days in each year in which the precipitation is above that threshold.
2. Step 1: Collect the number of days per year there is precipitation. Step 2: Plot the number of days vs. the year. Step 3: Collect the total amount of precipitation per year and plot that against the year.
3. step 1: create a key threshold that means means very extreme precipitation once past that threshold step 2: count the days in a year that pass that threshold step 3: do a percentageduring a year to see the pctgs

Student responses

1. - I think that you could use the daily weather data to decide whether extreme participation has become more frequent by comparing the amount of rain between the years, and also comparing even just daily how much it rained between now and then. - Plan: make table in descending order of top prcp, and then one again but by the day
2. 1. look at the data 2. gather the number of days it rained in a year/amout of rain 3. compare it across years and look at the trends
3. 1. put the data into a table 2. determine what is considered extreme weather, count the frequency 3. compare averages in the teo time periods
4. 1. Take average precipitation for year 2. Define extreme precipitita
5. Define and compare
6. define what extreme is sort by year plot data
7. Define what is considered extreme precipitation, collect data on average precipitation, that analyze whether this is an increaseing trend or not.

Student responses

1. First I would think that you need to consider previous data. If you dont know what is normal/average, so you need to determine that. After you have figured out how often precipitation occurs, then you can start measuring the rain data for the year and then comparing it to the past to see if it deviates. The standard deviations between year might be a good way to see within a year if things have changed and then comparing years standard deviations can also see if there was a difference as well.
2. Set a standard to decide which is considered extreme precepitation or not. Then we can count the number of extreme precipitation days in each year and compare how the numbers changed.
3. simply count for the tmax for each year and compare them
4. We need to settle a temp that is considered extreme first. Then, we need to see how much data we have, and if it is like the high/low temps in the previous data, we need to group average temp by year and analyze a trend in that dot plot. W can then determine from the more condensed plot if there are more extreme precipitation levels over the years

Student responses

1. would define extreme precipitation and compare yearly frequencies, trends, and changes. check data consistency

Student responses

1. Categorize by decade 2. Create a ggplot showing the trend with geom.line on prec 3. check if the slope is increasing in recent decades
1. gather all the date 2. find key datas (mean, average, increasing, stdev) 3. compare all the data 4. check if there is extreme precipitation compared to the past
- find if its a outlier or statistically significant
- Find the mean and remove the outlier and Replace NA values,
- I would begin by finding the average precipitation for each year then calculating the standard deviations of each year.
- I would measure the standard deviations of the maximum and minimum temperatures over the course of each year and see if they grow and decrease over the course of each year.
- Set a normal range to define "outlier", use the lower/upper fences to determine whether if the data is an outlier.

Student responses

1. We need to determine if the average amount of precipitation has increased over time. Filter the dataset by some long unit measure of time (month, year), average each unit by its amount of precipitation, and compile it all into one visualization to establish if precipitation has distinctly and progressively increased.

Student responses

1. - decide what extreme weather is - compare extreme weather year by year - graph extreme weather counts - make sure to consider high heat, rain, tornado, high snow, very cold etc
2. find years with higher standard deviations to see which years have more extreme temperature minimums and maximums. Then find the years with the highest precipitation. Then we would see where those years of interest land relative to the current date.
3. I would use the amount of precipitation in the daily weather data to see if the average water source is growing or dropping. I should check if I am comparing the same thing and check the unit
4. use precipitation, tmin, tmax, see how much general precipitation there is, define a benchmark amount of precipitation to define what 'extreme' is, see how how frequent each year has extreme precipitation
5. We would measure precipitation, mins, and maxes. for precipitation set a benchmark amt for "high" and compare amounts of high precipitation year over year to track changes. for min and max we would see if these values have become more extreme over the years

Student responses

1. We would probably look at a large scale plot to see any trends and if not, keep going and narrowing down the data. Maybe group by season, year, decade or anything similar to that. We should also find the average daily precipitation in those groups as well.

Student responses

1. 1. Categorize different types of "extreme precipitation" 2. Organize the amount of each type for each year 3. See the amounts of each type of extreme precipitation throughout the years.
2. 1. Measure current year's precipitation and past years' precipitation 2. Create graphs, tables, and use statistical techniques to compare and contrast the data between years, then between decades 3. Graph the average precipitation in per years and in per decades to see if there is an observable trend of precipitation change throughout the years.
3. compare precipitation in decades, and compare then and draw a plot

Student responses

1. 1. categorize what extreme precipitation is 2. organize the data by year 3. take averages of precipitation 4. plot and analyze for high precipitioat
2. 1. Filter out the NA values 2. Take the mean percipitation by year 3. Plot this over time.
3. 1.) Select for relevant years and precipitation 2.) Group by year to see aggregation trends for precip. 3.) Filter for data that includes no less than 360 days of data. 4.) Aggreagate mean

Two days each receive 1 inch of precipitation. On one, it falls in 20 minutes; on the other, it is spread over 24 hours. Our rule counts both. What does this measure capture, and what does it miss?

89 anonymous responses

Misses intensity or rate

34

Emphasizes that the rule fails to capture how quickly precipitation fell, i.e., peak intensity or rate (important for flooding/extreme events).

Captures total daily rainfall

30

States the measure records the total precipitation per calendar day (days reach the 1-inch threshold) without distinguishing other features.

Counts different events as same impact

10

Points out that treating a short intense downpour and a day-long drizzle equally conflates events with different impacts (flooding, drains, damage).

Response themes for question 8

Misses duration and timing details	8
Notes the measure ignores how long or when during the day the precipitation occurred and that multi-day timing can matter or cause misclassification across calendar days.	
Other responses	6
Responses the model could not place reliably.	
Other responses	1

Student responses

1. Captures daily frequency. It misses instantaneous frequency.
2. I do not think it matters because both cases are examples of extreme weather. It does miss intensity but that is not necessarily the question.
3. It captures a good measure of precipitation for that day, but one is much quicker/extreme and the other is more normalized. It misses the duration in which it takes to reach extreme precipitation.
4. it captures the amount of rain, but not the rate, people would experience day 1 as 'heavier rain' but we could it as the same amount
5. It captures the total amount of precipitation in a day, but it misses the peak severity of precipitation in a day, which is at least a component in what we think of as extreme precipitation, since more severe precipitation can lead to "freak accidents" and other rarer events.
6. It counts the aggregate dep per day, rather than the peak rainfall rate. It could drizzle all day for a day versus rain hard for 30 minutes. One is obviously extreme weather, while the other isn't.

Student responses

1. it did not capture over what time period it rained. 1 inch over 24 hours is very little rain, while 1 inch over 20 minutes is a lot of rain
2. It does not know the difference, and I am not really sure if it matters too much. I would say if it rains for 24 hours, that is extreme, and if it rains intensely for 20 minutes, that can also be extreme. Our thing just gets daily precipitation total.
3. It doesn't capture the intensity of the rain that day.
4. It measures how heavy/intense the precipitation was. It captures the overall daily amount
5. It measures total amount of precipitation, but it misses the rate of precipitation which we should consider for extreme precipitation classification
6. It misses precipitation per hour or all at once
7. It misses the frequency of that precipitation. 20 minutes of extreme rainfall might be what we consider extreme weather. But we could also classify it being more extreme if it rains for longer overall
8. it misses the intensity of rain over the period where it is raining as it only looks at the overall amount

Student responses

1. It successfully captures the day it reached more than 1 inch of precipitation, but cannot capture the length of the precipitation.
2. its definitely more extreme if it falls in 20 minutes
3. Missing the precipitation intensity, which matters. It only capture the total amount, but precipitation intensity differences would tell whether the whether is really bad or overall mild
4. Taking into account time is also important, if you disregard time then you lose the info that if inch of rain falls in 1 hour compared to a day that is different definitions of extreme
5. The first day rains much more heavier than the next day, it misses the total amount
6. the measure captures the days with 1 inch of precipitation, but it misses the impact
7. The time determines the prec for each unit, and we cannot define without a fixed measurement.
8. They would both capture the extreme rainfall observations given the definition but does not account for intensity of rainfall
9. This basically groups flash flooding events and very rainy days together. The two are definitely very different in terms of how they affect people.

Student responses

1. This measure captures all days that get to that minimum threshold, but you wouldn't categorize it as extreme precipitation if it's a long drizzle.
2. This measure captures the amount of rain but not the length of time it was raining for and the severity of the rain
3. This measure captures total rain amount in a day, but it misses the frequency of rain falling down. The one falls in 20 minutes should be seen as more extreme
4. This measure considers both as extreme weather, but it misses the time interval of that precipitation
5. This measure counts the amount of precipitation. It misses the rate at which the precipitation happens, not taking into account the time change.
6. This measure misses extreme weather events where rain falls very quickly and instead counts days where rainfall is gradual.
7. this measure only captures the amount of precipitation per day but not the speed of precipitation.

Student responses

1. This measure would miss the timeline of the weather event, because the current rule only assesses what was recorded over the entire day. While it would capture days with over one inch, it might be considered more extreme if the one inch happened within only 20 minutes.
2. This would record a day in which it rained a moderate amount over the whole day as the same as a day in which there was a flash storm that poured over a short period of time.
3. Unfortunately, the time it takes for the precipitation to fall isn't really included within the dataset. As such, we will miss the intensity of precipitation across a specific period of time for many days.
4. Well the best way I can think to describe this is like rain density in a way lol, like for example if it rained that much in 20 its like more dense than rain in 24 hours, so we should make something to accomated this maybe calcualte rainin a ratio to help make sure we capture total rain in a day

Student responses

1. It captures only the days that receive 1 inch of precipitation regardless regardless of how long it rained that day. it misses how much rain fell and the rate of how it fell on a particular day
2. It captures the overall amount of rain, but fails to capture the rate at which it falls. More rapid rainfall can leads to more severe damage and other consequences.
3. It captures the total amount of rainfall in a day but doesn't show how extreme the weather was throughout the day. All of the rain falling in 20 minutes is more extreme than spread out throughout the day.
4. It captures the total daily precipitation, but misses the intensity/rate of the precipitation.
5. It captures the total precipitation for each day, but fails to indicate the real-time precipitation rate.
6. It captures the total precipitation for the day but it ignores whether it was over a long period of time or in a short span (like a storm)
7. it captures the total precipitation of a day, it misses the severity of the storm/ the speed that rain was falling

Student responses

1. It captures the total volume of water that fell in a day, but not the intensity. The 20 minute version is way more likely to cause flash flooding and overwhelm storm drains, while an inch spread over 24 hours mostly just soaks in, and daily totals treat them as the same thing. It also misses multi day storms, where a few moderate days in a row add up to real flooding but no single day hits the cutoff.
2. It captures total precipitation in a day, but misses intensity/speed which is important when considering extreme weather
3. It captures total rain volume and not peak rainfall.
4. It would be captured and identified as an extreme cutoff. However, it doesn't take into account the rate at which the rainfall happened, because 1 in over a day is like a drizzle but 1 inch in an hour is heavy downpour.
5. The cutoff measures total rain but misses instantaneous rain intensity. Whether this matters depends on what we mean when we say 'extreme'.
6. The measure captures only total amount of rain in one day. It misses capturing the amount of time in which rain is falling.

Student responses

1. the measure captures the days that hit that min extreme threshold however it does not account for how that threshold is met
2. The measure captures total precipitation over a day, however, it misses the average precipitation/hour or per minute, which is a more accurate measure of precipitation when comparing and contrasting our data for this study.
3. These two instances will be recorded in the same way based on the structure of how our data is presented to us. It captures the total precipitation in one day, but it misses instantaneous intensity.
4. This captures total daily precipitation, but misses the relative intensity of the precipitation within a day.
5. This information captures that there was 1 inch of precipitation on both of those days, however, it misses the length of time in which that rain fell.
6. this measure captures any day >1 inch. it does miss the length of time the precipitation happens in. it only captures quantity, not time it takes.
7. this measure captures both days as having extreme precipitation, it misses the spread per day
8. This measure captures daily precipitations

Student responses

1. this measure captures daily total precipitation but misses the intensity. On the day where it is spread over 24 hours it may not be considered extreme.
2. This measure captures the overall precipitation in one day. It misses the precipitation/time
3. This measure captures the total amount of precipitation in a time period but not the intensity.
4. This measure captures the total amount of rain in a day, but it disregards how fast that rain could fall to qualify as extreme. One inch of precipitation in 20 minutes is arguably more extreme than one inch of rain spread over 24 hours
5. This measure captures the total amount of rainfall (so both would be classified as extreme), but the day where all the rain fell in 20 mins might signify a more intense (or extreme) storm.
6. This measure captures the total amount of rainfall over the course of a whole day, so there could be events of extremely heavy rainfall that only lasts for 15 minutes and is less than an inch that would not be counted and there could be cases with light rainfall over the whole day that is counted

Student responses

1. This measure captures the total rainfall in a day, however it does not capture the time in which the rain falls. Thus it might be prone to miss extreme weather events, like a flooding, which is more likely in the first scenario presented.
2. This measure captures the total rainfall that we are experiencing. However, it is missing how extreme the precipitation actually was. If we are thinking about extreme weather than we most likely want to look at the suddenness and intensity of rainfall in a small period of time where thing like flooding for example may be more likely
3. This measure succeeds in capturing the total daily precipitation, but it misses the intensity with which it falls. One inch in 20 min is much more intense than one inch spread throughout the whole day.

Student responses

1. Capture a extrem rain day but miss a lot of time is not rain
2. it captures precipttion accurately and misses more chance for accuracy
3. It captures that there is precipitation on both days but it fails to capture the pressure of the precipitation
4. the severity of the rain at one point in the day however it does capture the total amount.
5. This measure captures the inch of rain however it doesn't capture the extreme downfall the one day had as opposed to the other. Within a 24 hour period it is much more likely for an inch of rain to ocur.
6. this measure will capture that both are "extreme" precipitation. it will miss the window of time that this happened in, as it is likely more rare for that much rain to fall in 20 minutes than over the course of an ntire day which could be another indicator of extreme weather
7. this measure will capture the total precipitation of both, but it will miss the variety of the precipitation
8. This rule counts the amount of precipitation, but one is only counting from the sky, and the other is putting herslef out there to guess and ampathize

Student responses

1. This would record a day in which is rained slowly over the course of a day and another when alot of rain came at once as the same even though they aren't
2. Total volume may be matter or the time length.

Student responses

1. It captures the days that have 1 inch but misses on whether it actually ended being a lot, like water on the ground or whether it was so slight and evaporated fast if you incorporate all 24 hours. Also timing is left out like if it was only at night.
2. It captures the duration that the precipitation occurs over, but misses to have a consistent duration to measure precipitation trends over
3. It captures the time it takes to receive 1 inch of precipitation which can help establish frequency
4. it measure total rainfall, but not limit time. it is hard to calculate time, because data set do not contain it
5. It's missing the time period of the inch of precipitation. Over 24 hours mean that its 1/24 for an hour which is less precipitation than 1 inch an hour. We need to set the rule for a precipitation for a certain time period.
6. The data captures precipitation per calendar day. So if precipitation does not fall as 1 inch within a calendar day it will be missed.
7. the one over 24 hrs would not be captured since it goes into a diff day
8. time of duration with which the rain fell

Student responses

1. Both will be counted towards this dataset since there is no timeline in the definition, only the total of the day.
2. it capture the amount it misses the speed
3. it would not catch the difference in time between the different days. we could use a rate stat such as cm/min to adjust for that
4. our measure counts daily precipitation, not necessarily how extreme or quick the precipitation was.
5. The definition only says
6. this measure doesn't capture the difference it reads the days as exactly the same

Suppose an earlier year has 270 recorded days and a recent year has 365. The recent year has more recorded extreme precipitation days. What would you need to check before interpreting that as an increase? Suggest a way to make the comparison fairer.

92 anonymous responses

Compare proportions or rates

38

Students say you should compute the fraction, percentage, rate, or ratio of extreme-precipitation days relative to recorded days (e.g., extreme days/total recorded days) instead of using raw counts.

Check missing days and their timing

17

Students emphasize inspecting which days are missing and whether missingness is random or concentrated in particular seasons (which could bias results if missing days coincide with high-precipitation periods).

Other responses

12

Responses the model could not place reliably.

Response themes for question 9

Normalize by subsampling or adjusting counts

7

Students propose making the years comparable by subsampling or scaling: randomly select 270 days from the 365-day year, multiply or scale counts to a common denominator, drop extra days, or otherwise resample to equalize days compared.

Exclude or replace incomplete year

6

Students recommend omitting the year with too many missing days or finding another fully recorded year as a fairer comparison rather than trying to adjust partial data.

Unclear or minimal responses

6

Responses that are brief, uncertain, or not informative about the comparison strategy.

Response themes for question 9

Investigate nearby years or additional data

5

Students suggest checking adjacent years or additional datasets to contextualize whether the observed difference reflects a trend or is due to anomalous/missing data.

Other responses

1

Student responses

1. A year with more recorded days would naturally record more extreme weather events, even if the underlying frequency is the same. Normalizing the count of extreme weather events to the number of days recorded could eliminate this effect.
2. Check the observed probability of heavy rain per day per year instead of just the aggregate number of days per year.
3. check the proportion of extreme weather days over the total days
4. Check the proportion of extremedays and recorded days. Find the largest proportion.
5. Check which values are missing and see if there is a trend. Then, compare what percentage of recorded days had extreme weather.
6. compare proportions not raw counts
7. compare the percentage of extreme precipitation out of total recorded days
8. Consider finding the ratio of extreme days to days recorded and compare those two values. It's also worth considering that these extreme events could've caused weather to not be recorded.
9. do a percentage instead of using pure number, pctgs will account for the missing days

Student responses

1. I need to calculate the ratio of extreme precipitation days and the total number of recorded days for a fairer comparison.
2. I should compare the percentage of days recorded in the year
3. I would check the percentage of extreme precipitation days in both years and compare the percentages.
4. I would multiply the number of recorded extreme precipitation days of the more recent year containing 365 recorded days by $270/365$ in order to properly adjust to the ratio of recorded days.
5. I would need to check the amount of days that precipitation status was known in the year, as well as the extreme precipitation days. A better comparison would be the proportion of extreme precipitation days out of all recorded precipitation days, for example $5/270$ vs. $8/365$.
6. I would need to check the ratio between the years before I interpret this as an increase. In order to make a fair comparison I would compare the percentage of days with extreme between all years

Student responses

1. i would want to check the proportion of the extreme days in each of these years and compare those proportions. since theres almost 100 extra days in the recent year it would make sense for there to be a higher number of extreme days, but checking the proportion would make this comparison more fair
2. Instead of comparing the number of days, it will be better to compare the percenage of extreme precipitation days out of the total recorded days.
3. It could still be proportionally the same even if the number increases because the total number of days recorded also increases
4. Looking at just the raw numbers, this is not a good way to actually interpret it. Instead, we would probably want to look at maybe the percentage of extreme participation days taking place.
5. no, i need to calculate a ratio
6. ratio of extreme weather per day
7. take the average of extreme rain by days recorded but even that wouldnt be perfect because the missing days could have been in rain season.

Student responses

1. The earlier year is missing 95 days, so it had fewer chances to record an extreme day, and I'd check which months are missing since losing summer would matter most. Fairer fix: use extreme days per recorded day instead of raw counts, or drop years below 90% complete.
2. The rate of extrem precipitation in years
3. use the rate/proportion of extreme days
4. We should check the rate at which extreme days to recorded days; this will be fairer
5. We would have to find the percentage or proportion of extreme rain days vs the amount of days recorded for that year.
6. We would need to check the proportion of days that had extreme precipitation before being able to see an increase, because some of the missing days could've had extreme precipitation.
7. We would need to make a proportional ratio so we can compare years. If we take amount of rain days over total rain days recorded, we can then make sure to compare years beause then we are accounting for the missing days byt taking their proportion out
8. We would need to measure the proportion of extreme weather events before interpreting that as an increase

Student responses

1. Yes, we need to find the fraction of extreme days to days recorded. else it doesn't accurate
2. You could calculate a percentage of recorded extreme precipitation days which would give a fair comparison regardless of the sample size that we are looking at.
3. You need to check the amount of extreme weather days compared to the number of total days with a data point. We could make this comparison by obtaining a proportion of days with extreme weather.
4. you need to check the percentage of days recorded that had extreme weather since the years have a different number of recorded days.
5. you should average out number of extreme days based on number of days of that year recorded so that no missing data affects the outcome
6. You would have to look at a percentage of each year to see the true comparison
7. You would need to check the proportion of days (assuming the ones missing are random) that had extreme weather, compared to the amount of recorded days.
8. You would need to look at the percentage of days that were extreme to see if it was higher

Student responses

1. Check if the missing days in the earlier year is random, or all during a certain season.
2. Check the exact numbers of the extreme days. Maybe compare the percentage of extreme days divided by total days. Also need to check whether a whole season is missing, may cause biased data.
3. check whether the moe extreme precipitation days happen in the extra 95 days.
4. I would first check if the missing value are relatively uniformly distributed in time period. If it's not and concentrated in, say, several months, we shouldn't use this year. If it is, we should calculate the ratio of extreme vs days recorded
5. I would need to check if the missing days are related to the weather itself and caused an inability to measure precipitation.
6. It might be that the year missing days was missing extreme weather days, not just random days. We would like to check the randomness of the missing.
7. See if more of the missing values are present in days where we would expect higher precipitation. In that case, the measurement would be biased and we would need to handle the null values carefully.

Student responses

1. We would need to make sure that those missing 95 days are not missing because they are extreme weather days. You can check this by seeing what days are missing and make an educated guess if it is in high rainfall seasons
2. what days were missing
3. would need to check if missing days are related to them having extreme precipitation, or if by randomness. either way, an adjustment of some sort would be necessary to correct it,
4. You should check if there is any trend to which days are not recorded in the earlier year (for example, if there was no data recorded in the summer) as that can affect the data. To make the comparison fairer, you could discard some days from the later year when summarizing.
5. You should check if there is anything that caused there to only be 270 recorded days or if it is just completely random, because it is possible that the missing data is related to the precipitation and that would skew our results
6. You would need to check what days during the year the 95 missing days were. We could take a random sample to make the comparison fairer.

Student responses

1. You would need to check what the recorded days actually were for the earlier year. It's possible that there are seasonal issues, or if a season with lots of rainfall was ignored.
2. You would need to check which days in the earlier year in which precipitation data wasn't available to determine if the missing data was related to extreme precipitation or not.
3. you would need to check why the year is missing data, If it is completely random then thats ok but if its not random and maybe if theres too much rain it doesnt track then there is a problem with your data
4. You would need to see the remainter 95 days for the earlier year or other earlie years to see if thre is an actuald iffERENCE

Student responses

1. check for the average extreme weather days in every month
2. Check the average, the scale (units) of measurement, and see if they align with previous days. The sudden increase might have something to do with the increase in recoded days of the year.
3. I dont think it currently a fair comparison because the sample size of the earlier year is significantly less than the sample sixe of the recent year. This can most definently have skewed the results for the more recent year to have more recoreded ectreme precipitation
4. I would check the ratio of extreme weather days to total days, and also why the data was not recorded over the missing days on the 270 year, it could have been missed due to extreme weather
5. I would need to check if the missing days from the earlier year were at random; otherwise, if they might be missing precisely because of extreme precipitation, then the calculated number of extreme days in that year will be an underestimate. A fairer way to make the comparison might look at the days that were captured in both years.

Student responses

1. I would need to check whether the averages are skewed because of that missing data, or whether there truly is an increase in precipitation.
2. It would not be fair as there are 100 more days in which extreme rain could occur. To make it more fair we could look at the average number of days recorded that there was extreme rain
3. Percentage of days recorded where extreme rain occurred.
4. The recent year having more recorded extreme precipitation days shouldn't necessarily be interpreted as an increase, because with an earlier year having 270 recorded days, the recorded extreme precipitation days may be in line with the number of recorded days.
5. The season of the days recorded for the year with 270 and then whether this messes with the estimate
6. Use %
7. You would need to check the data to see if all of the values from the short year were that much higher than from the full year.

Student responses

1. check the extra days if more precipitation is recorded there, and to make it fairer, you could make it so the same amount of days are compared between both years
2. I would make sure to check that each year's extreme days is on the same scale so that it is a fair comparison, so I would multiply the smaller days by $365 / x$ recorded days so that they are all out of 365 days.
3. Its definitely going to skew the data a little bit. I think in order to make the data more fair you should randomly sample 270 of the 365 days for the full year and then compare the two.
4. To make it fairer we should select 270 randomly from the 365 day year so that we can actually check if its because of the large data or really if this year has more precipitation days
5. We can try to addd 95 days to year 1 or remove 95 days from year 2 based on their on data level, so that we can sure hat the number can represent the data on that day
6. You could normalize the data by randomly selecting 270 days from the 365 days.

Student responses

1. you would need to check the number of days, you could filter the other year to only have the days the previous oe is missing or you could randomly select months to compare

Student responses

1. I would either omit that year from the comparison group or I would research into what those missing days the rainfall may of been like but that is a long process and usually not worth our time
2. Need to check whether the recorded days are complete and the same across the years we want to compared, one way is to only select the complete recorded years.
3. no this would not be a good indication. If we could look at other years around it that may not hav this missing data or just exclude this year if its baised to a particular time frame that may increase or decrease rain amount
4. This is not a good indication, I dont tthink we can conclude that there is more missing weather as we have like around 90 msisind days. So always check how many dayda recorded, hoose another earlier year for this ocmparison to make it fairer.
5. You should try to reconcile the earlier year to 365 days and find data for missing dates, if you can't do that, it's not a fair comparison and you should find another year that has a full data frame.

Student responses

1. You would absolutelt need a check. I think its best to standardise the years to have a minimum number of recroded days, such as 320.

Student responses

1. Before checking the interpretation as a good increase, we should consider where people are from
2. I'm not sure.
3. It tell us extreme case
4. Missing data
5. no, more days in the new one
6. y

Student responses

1. Probably check what happened in the unusually low year and what happened in the high year maybe an event occurred to mess the data. Once you figure out what happened can pick around the data to get a more accurate.
2. There is nothing we can check but we can compare by decade or block to get a better picture.
3. You could check the years around it. You could be missing days that have extreme data or not, but you don't know.
4. you need to check the years around that year as you could be missing extreme days but don't actually know
5. You would want to do some of your own research to check another dataset and track the rainfall in the earlier year. If this does not work, you can also check the years around it and see how those compare to the earlier year.

Before calculating the period means, predict how the number of extreme precipitation days per year has changed. Give a rough difference in days per year and a reason. What result would surprise you?

91 anonymous responses

General increase stated without amount

23

Students assert an increase in extreme precipitation but give no numeric estimate or give vague reasoning about climate change.

Predict small increase (1–3 days)

18

Students predict a slight rise in extreme precipitation days (about 1–3 days per year) citing warming/greater moisture or general increased extremes; surprise would be a decrease or no change.

Predict moderate increase (3–7 days)

15

Students estimate a moderate increase (around 3–7 days per year), often attributing it to climate change and expecting decreases to be surprising.

Response themes for question 10

Expect little or no change / data too noisy**10**

Students think there will be minimal change or that noise/missing data prevents detecting a change; they would be surprised by a large difference.

Predict larger increase ($\approx 10+$ days)**9**

Students expect a substantial rise (around ten or more days per year) due to global warming or accelerating trends; some specify large numbers (10–20+).

Unclear or non-answer**9**

Responses that are empty, uninterpretable, or simply say they don't know.

Predict decrease or fewer extremes**4**

Students predict fewer extreme precipitation days (decrease) or expect decrease after data cleaning; they would be surprised by an increase.

Response themes for question 10

Other responses

3

Responses the model could not place reliably.

Student responses

1. Average days per year will increase. And there is no way to tell what the change will be.
2. Because it seems like all the missing days are in the baseline period, I would expect that to play a role in comparing the period means, because that means the baseline has less data to collect.
3. Due to climate change I would expect there to be a slight increase in the number of extreme precipitation days per year in the more recent period.
4. Extreme days become more frequent. I could be surprised if it is dropping.
5. Extreme precipitation days have increased. I predict it is still a little smaller all data
6. Get higher? Based on the data
7. I believe the number of extreme precipitation has The The difference will be pretty big. I would be surprised if it is similar.
8. I predict it has slightly increased to one day per year, maybe 2-3 more days and due to climate change. Any other reason might surprise me
9. I think it would create more extreme precipitation particularly around Ann Arbor as the many water sources would evaporate and fall at a faster rate

Student responses

1. I think precipitation has also had a greater increasing trend recently than before, given that climate change has caused more precipitaton in the midwest area and less in the western US
2. i think the number of extreme precipitation days per year has increased because of climate change
3. i think there will be more extreme precipitation days
4. I would be surprised if there was no change or a very extremem change. Climate change is meant to be something that happes over very long periods of time so a very large change would be surprising. 20+ avergage increase
5. I would expect a rise in the number of extreme precipitation days per year due to global warming.
6. I would expect about 0.75 extra extreme days in the last 30 years
7. I would expect that the number of extreme precipitation days to increase with time. It would surprise me if it had decreased or remained the same. It will be a relatively small increase increasing in magnititude.

Student responses

1. I would make the prediction right now that current days have had more days of extreme precipitation per year. My observation lives in global trends of warming causing unbalance and more rain
2. I would say that the recent years have seen a rise in extreme precipitation that is accelerating at a faster rate than back in the 1960s. This is because of global warming
3. I would see a rise in the number of extreme weathers
4. It has increased because the weather is becoming extreme, the day of extreme weather should also increase. If the days stay stable, I would be surprised with in 5 days change
5. It will get more higher frequency
6. the number of extreme precipitation days happens more often in colder days
7. the reason is because of global warming, so it will increase

Student responses

1. A slight increase in number of extreme precipitation days per year
2. I feel that with global warming, the amount of extreme precipitation days has gone up. I think it would be a very small difference though, something like two days a decade.
3. I predict a slight increase in extreme precipitation days due to climate change creating more extreme weather, although the change is likely not dramatic and under 3 days due to the large period of measure for both periods. I would be surprised if the result was very large.
4. I predict that the newer period has more frequent extreme precipitation days. Maybe a difference in 2-3 days, and it would surprise me if the standard period had more.
5. I predict that the number of extreme precipitation days per year has potentially increased over time. I think the difference will only be one or two days. I would be surprised if we got a huge difference.
6. I predict that the number of extreme precipitation days per year will increase in the second period by around 1-2 days.

Student responses

1. I suppose there will be more precipitation days per year as time goes by. Probably 1 more extreme precipitation days each year. A decrease in extreme precipitation days per year would astonish me.
2. I think it may have by around 2 days, would be surprised if higher
3. I think it will be small increase
4. I think it would be a slow change of the number of extreme precipitation days, meaning that over 4-5 years the number of days would increase by 1. It would be surprising if there was no change
5. I think it would be slightly more by maybe 2-3 days since it has gotten warmer, more rain and less snow.
6. I would expect a slight increase of extreme precipitation days per year in the 2nd period.
7. I would expect extreme precipitation days to trend up over the years very slowly. About 1-2 more every year. A big jump would surprise me.
8. I would expect the number of extreme precipitation days to increase by one to two due to climate change over the years, a decrease in extreme weather would surprise me

Student responses

1. I would predict that the number will stay the same or increase by 1-3 days because I dont know of any change that would lead to an increase but I do feel like over time we have had an increase is rain
2. I'd guess about 1 more extreme day per year in the recent period, maybe going from around 4 to 5, since warmer air holds more moisture so heavy rain events should get a bit more common. It'd surprise me if recent years came out clearly lower, or if the jump were huge like 3+ days, since the year to year swings alone are that big.
3. Maybe the number will grow from period 1 to period 2 by 1 day? Because global warming makes weather more extreme. If the difference is less, I would be surprised.
4. Predict that more recent years has 1-2 days of extreme precipitation days per year than previous years. The number is the same would surprise me.

Student responses

1. I expect an increase in extreme days in recent years due to climate change. I expect it to be 5 more days.
2. i predict an increase of 5 days/yr because of climate change leading to more extreme weather espeially snow fall. i ould b surprised by little to no change
3. I predict it has increased by a few days a year, around 5 is my guess. It would surprise me if there was a decrease in extreme days.
4. I predict that the current period will have a higher number of extreme precipitation days per year by 4 days, i'd be surprised if it was lower in recent times.
5. I predict that the number of extreme precipitation days per year has increased by maybe about 5-10 days between periods. I would say this is because the gradual average temperature increase leads to increased storm activity.
6. I predict the number of extreme precipitation das per year has increaased by about 3 as a result of more extreme weather brought by climate change.
7. I think the new period will have around 3 or 4 more extreme days per year on average

Student responses

1. I think there are probably 5 more days per year with extreme precipitation. A result of o change would definitely surprise me.
2. I would expect the number of extreme weather events to increase in the recent times, likely between 3-8,due to the effects of climate change.
3. I would predict the number of extreme precipitation days per year to change over the years. A rough difference could be 4 to 5 days per year due to climate change and given the amount of extreme days we've seen already
4. I would roughly expect there to be about 3-5 more extreme weather days this year, as global warming leads to extreme temperatyre shifts.I would be surprised if there were less extreme days this period compared to less
5. It has probably increased by about 4 days per year. A result that would surprise was if it decreased.
6. The number has seemingly increased. There will be a small increase of about 3 days
7. the number of etreme days per year will have slight increase, about 3-5 days. it would be shocking if 6-10

Student responses

1. We might expect the number of extreme precipitation days per year to be higher in the recent than baseline period. One estimate might be about 3 more extreme days per year. I would be surprised if the baseline period had significantly more extreme precipitation days than the recent period.

Student responses

1. Can't measure the change because th data is too noisy; need to simplify the date ranges. Increase would be very large
2. difference will vary little.
3. I expect that the change in extreme precipitation days per year will not be able to be seen because the data is noisy.
4. I predict roughly the same number of extreme days per year. A result that would surprise me is a major difference either way.
5. I predict the number of extreme precipitation days per year to not have changed as much after fixing the years that had missing data and smaller sample sizes.
6. I think that there would be little change in extreme precipitation per year. A large differnece would surprise me
7. I would expect the difference between the yars to not be too massive because each year has simila trends unless there is too much noise, etc
8. I would predict not a significant change as I cannot come up with a reason for a change. It would surprise me if there a big change
9. Not a big change

Student responses

1. There is no change because the data is too noisy. I cannot measure the change.

Student responses

1. 10 days
2. I predict that the number of extreme precipitation days per year has increased due to climate change, maybe by 20 days. A result greater than a 100 day difference would surprise me.
3. i would estimate that there are probably around 10-15 more extreme weather events now than in the past.
4. i would expect more rain more in more recent years with about 10 extreme days a year
5. I would expect that there is more precipitation in the modern events. I would estimate around 10-15 more extreme weather days on average from the modern data set due to global warming events.
6. I would expect the number of extreme precipitation days per year to increase over time. I would guess that the days per year with extreme precipitation would increase by roughly 5 from the baseline years because of global warming. I don't think global warming has been significant enough to cause more of a change. It would surprise me if there was no change.
7. I would guess that there is an increase of about 10 days per year due to global warming

Student responses

1. I would say the number of extreme precipitation days has increased by a small margin. These days seem more often because of global warming. A decrease would surprise me. The margin would be about 10 days.
2. The number of extreme precipitation days increases very little

Student responses

1. a big differernce would be suprising
2. don't know
3. I would assume that the data there would be a difference of about one day per year more from the first to the last year o acerage that out for per yaer becuae of global warming there is more rain
4. I would expect there to be more extreme precipitation days per year because our instruments have gotten better and there will be less days with no data.
5. I would predict a slight increase (because climate change results in more extreme weather). I think it would be 10 days more. I would be supried if there was a large decrease >30 days
6. I would predict it woudl increase maybe 10 days due to global warming. I would be supried if there is no change.
7. I'd guess around 5 more per year just because it feels like the tornados/other weather events cause those.
8. L
9. maybe like 7 more days of rain because of climate change

Student responses

1. have less extreme dat., due to fliter
2. I guess the precipitation goes down for the second period, considering the less day for raining. maybe 10 days
3. I would expect the amount of extreme rain events to show a slight decrease. Probably around 2-3 days per year due to warmer tempatures and drier summers.
4. The number of extreme precipitation days per year should have decreased because now we have taken out the data with incomplete records in a year. Thus mathematically there should be a change of around 10 days.

Student responses

1. I predict that during the summer period the extreme precipitation days are a lot more than winter season.
2. Increase ~2 days; decrease surprises.
3. It would surprise me that it has decreased.

Use the two means and their difference to describe the change in ordinary language. Would this difference matter to someone responsible for flooded streets or basements? What else would they need to know?

0 anonymous responses

No responses were submitted.

Does looking at the plot reveal anything that simply comparing the means hid? Point to something in the plot that strengthens or weakens your confidence in the comparison, and explain why.

0 anonymous responses

No responses were submitted.

Write two or three sentences for someone who hasn't seen our code. Explain what changed, by how much, and how convincing you find the evidence. Include one limitation. Specify our definition of an extreme day, the location, and the periods; don't just report that a p-value is small.

88 anonymous responses

Increase of ~1–1.5 extreme days

24

Students report a quantitative increase of roughly 1 to 1.5 extreme precipitation days per year between the baseline and recent periods (often ~1961–1990 vs ~1996–2025/1990s–present) in Ann Arbor.

Statistically significant evidence (p-value)

23

Students emphasize that a t-test produced a small p-value (commonly ~0.018 or <0.05), interpreting this as convincing evidence of an increase between the baseline and recent periods for Ann Arbor.

Ambiguous or minimal detail responses

15

Short, vague, or code-focused comments that lack clear quantitative change, periods, or limitations.

Response themes for question 13

Definition: ≥ 1 inch extreme day**8**

Students explicitly define an extreme day as at least one inch (or 1+ inch) of precipitation and note Ann Arbor as the location and baseline vs recent periods.

Other responses**7**

Responses the model could not place reliably.

Missing data and excluded years as limitation**5**

Students point to missing observations, dropped years, or incomplete records (especially earlier decades) as a limitation that could affect results.

Increase ~2 days or larger**4**

Students state a larger increase (around 2 days per year) in extreme precipitation days in Ann Arbor and cite a one-sided test or general increase.

Response themes for question 13

Other responses

2

Student responses

1. Ann Arbor went from about 5 inch-plus rain days a year to about 6.5, so roughly 1.5 extra soaking days, or 28% more. That matters to someone handling flooded basements since drains get maxed out more often, but they'd also want to know how fast the rain falls, whether wet days are clustering back to back, and whether 1.5 is big compared to the year to year swing of about 2.3.
2. Between the two sequential ranges of 30 years of weather data, there was an increase of the average of about 1.5 days of extreme weather per year. This was proved to be statistically significant because the chance that any given sample of 30 years that would have a difference in the average like this is below 0.05.
3. Extreme precipitation increased by about 2 days per year between the earlier and recent periods. This suggests extreme days became more frequent. At the U-M Ann Arbor station, extreme precipitation (≥ 25.4 mm) increased by about 2 days per year between the earlier and recent periods. The one-sided test provides evidence of an increase, though missing observations are a limitation.

Student responses

1. Given that an extreme day is defined as a day with 1 or more inches of precipitation, we saw that in Ann Arbor between a baseline period of 1961-1990 and recent period of 1996-2025, on average, the number of extreme days went up by about 1.5 days per year. Based on the t-test, we concluded that this was a statistically significant difference, meaning that we can say with high confidence that the number of extreme days went up. I find the evidence reasonably convincing.
2. Over baseline and recent years, events of extreme precipitation has become more common by about 1.5 events a year. The evidence is statistically significant by the p-value. Our definition of extreme day is over an inch of rain.
3. Over the course of 1961-2025, there is strong evidence to suggest that the frequency of extreme precipitation days, days in which there is 1 or more inches of precipitation, have increased in Ann Arbor. This was found by comparing the period of 1961-1990 to 1996-2025.

Student responses

1. The amount of extremem precipitation days from 1960-1990 is lower than the amount from 1996-2026 by around 1.5 days. We used a 2-sided t-test and comparing means to find this difference and see that the newer time period has a higher mean. The evidence is significant because we observed a very small p value in our t test
2. The average number of extreme weather days per year, where there was more than an inch of rain in Ann Arbor, Michigan, US, is greater now than it was in the period of 1961-1990. Our p-value of 0.018 is statistically significant and rejects the hypothesis that there is no difference in the mean number of days of extreme weather per year.
3. The average of extreme days per year has increase by 1.5 from the first period to the second period. The p-value was 0.018, which indicates that this is convincing evidence of an overall increase in extreme precipitation dats per year. This is measuring days with at least one inch of precipitation in Ann arbor over the period 1961-1990 and 1996-2025

Student responses

1. The average of the extreme days of precipitation per year has increased by 1.5 from the first to second period. The p-value was 0.018 which indicates this is convincing evidence of an overall increase in extreme weather days per year.
2. The mean of extreme rainfall is higher in recent years than in the past. There are about 1.5 more extreme rain events in recent years than before. Using a t-test we were able to conclude that the findings were accurate.
3. The more recent years showed more extreme weather as around 1.5 days were added in the more recent years. The evidence did how to be significant because it was less than 0.05. Maybe one limitation is that there are mutiple oher factors and data missing. An extrme day was classified as 1 inch of precipitation, location is ann arbor, and the period was up until 1960 and then from 1965-current

Student responses

1. The number of days with extreme precipitation, counted as an inch or more of rain, has increased by about 1.5 per year on average between 1960-1990 and 1996-2025. The evidence shows a 95% confidence that there is a positive change in extreme precipitation years in these two groups
2. The number of days with extreme rainfall in modern days was about 1.5 days higher than the baseline days, this led us to the hypothesis that as time has moved on there is greater amount of extreme days. We had a p value of 0.018 which means that our numbers could happen coincidentally at about a 2 percent rate which is rare so our findings are statistically significant.
3. There is a statistically significant difference between the mean number of extreme precipitation days between the baseline period and the new period. The p-value less than 0.05 gives us evidence to argue that there are more extreme days in the new time period vs the baseline period. An extreme day would be one with more than 1 inch of precipitation in a calendar day. There was a 1.5 day increase in the new time period vs. baseline period.

Student responses

1. We defined a day of extreme precipitation as having more than 1 inch of rainfall. We then defined 2 periods and found the average number of extreme precipitation days per year in each group. We then ran a test to see if these means are truly different where we found the p-value is 0.018 and since this is less than 0.05 we have statistically significant evidence that the number of extreme precipitation days has increased. One limitation is that we were missing some data.
2. We defined an extreme day as a day with greater than one inch of rain. The data was sourced from Ann Arbor in two periods of baseline years and recent years. Our p-value provides statistical evidence to suggest that the amount of days with extreme precipitation has increased in the recent period.

Student responses

1. We found strong evidence that the frequency of extreme precipitation has increased by about 1.5 days in recent years. We defined an extreme day as over 1 inch of rain a day, in Ann Arbor, and we had ~1960-1990 as our 'baseline' period, and ~1995-2025 as our comparison for the increase. A limitation of this method is that we arbitrarily defined extreme precipitation as one inch, with no real evidence that an inch of precipitation is extreme.
2. We found strong that from 1996-2025, there were on average 1.5 days per year more of extreme precipitation (more than 1 inch of rain), as compared to 1960-1990.
3. We found that, on average, there are 1.44 more days with more than one inch of rainfall after 1996 than in the years from 1961-1990. This increase is significant, but we had to drop some years with missing data.

Student responses

1. We looked at data from a baseline period, and data from recent years to see if the amount of extreme precipitation has changed. We found that, on average, recent years have about 1.45 more days of extreme precipitation than in the past. The data showed that our findings are very unlikely to be from random chance, suggesting that the amount of extreme precipitation days Ann Arbor receives per year has increased. One limitation in our study is that it doesn't take into account storm intensity.
2. We measured the average day per year with extreme precipitation over 1 in between two periods and compared them. Recent precipitation increased by about 1.4 days than earlier with a p-value of $0.018 < 0.05$. There's a significant increase in days w/ extreme precipitation at a level of 0.05

Student responses

1. We measured the difference in average extreme rain days per year (extreme = ≥ 1 inch of rain in a day), and compared 1960-1990 to 1990-2026 to see if on average they were different. We found an average difference of one day per year of more extreme rain in recent years, and the evidence was conclusive enough to have statistical evidence of different means. A limitation is our definition of extreme rain, which could be faulty in some cases.
2. What changed was the amount of days per year we received more than one inch of rain when comparing a baseline set of reference years to a collection of more recent years. We receive now, on average, one and a half days more of extreme weather rain now than 4 decades ago. The evidence is pretty convincing due to a large collection of data sampling here in Ann Arbor and trends elsewhere around the world, along with a statistically significant p-value. Data was not collected at times.

Student responses

1. Among the population of dates where there is rainfall daily, precipitation has increased heavily over recent years. We did so by running a two-sample t-test that produced a p-value of 0.018, showing enough statistical significance that the two samples have different values in rainfall. There are still limitations to this process, omitting days where there is only rain is harmful.
2. An extreme day is a day with over an inch of rainfall, the location is Ann Arbor Michigan. The p value is less than 0.05, meaning there is strong evidence that the amount of extreme weather days have increased over time. One limitation is the absence of data in the 1960s.
3. I'm 95% confident the true mean lies within the given confidence interval. The two-sided T test also provided evidence at an alpha level greater than 0.05 that there is significant evidence to reject our hypothesis that there was no statistical significance between the two means. One limitation is that the first period had a lower sample size.

Student responses

1. Our definition of an extreme day was any day that had at least an inch of rain here in Ann Arbor, and the periods were from 1961-1990 and 1995-2025. According to the t-test, there was a p-value of less than 0.05, meaning the evidence is statistically significant, where we can say that the true mean in average days with extreme rain has increased from the baseline period to the current period. One limitation is that the baseline years were missing a couple years of data.
2. p-value = 0.018, therefore, alternative hypothesis: true difference in means is greater than 0 is true. This means that the extreme raining days per year is actually higher in the recent period 1996-2025 than the period 1961-1990.
3. p-value = 0.018, which means we can reject the null hypothesis, we can be 95% confident that the alternative hypothesis is true
4. since the p-value is smaller than 0.05, we can say there is a difference between baseline years and recent years extreme precipitation days.

Student responses

1. The amount of extreme precipitation increased, on average, between the period of 1961-1990 and 1991-2025. We defined extreme precipitation by having more than one inch of rain in a day in Ann Arbor. Since our p-value is less than 5%, we have statistically significant evidence that the average days of "extreme precipitation" increased.
2. the average number of heavy rainfall days has increased between 0.3 and 2.3, extrame weather is any day more than a inch.
3. The days of extreme precipitation changed by about 1 day in between the years 1961-1990 and 1996-2025. I find the evidence to be semi convincing but there were some missing dates in our dataset that could change our answer. Our definition of an extreme day was one that recieved more than 1 inch of rain on that given day in Ann Arbor/near the university. We can conclude that there was a change because our p-value was greater than 0.01 which means we can reject the null hypothesis.

Student responses

1. the numer of days of extreme weather in recent years was about 1.5 days higher than the baseline. we created the hypothesis that as time has moved forward, more extreme rain days are happening. we ran t test and had p-value of 0.018 which gives us signifcant evidnece
2. The recent years have witnessed an increase in extreme precipitation days than the baseline years by around 1.14 extreme days per year. A two-sample t-test was conducted and the resulting p-value indicates that we are confident to support the result. However, the limitation lies in the fact that we didn't collect all days with some having missing data.
3. The test investigated whether the mean number of extreme rainfall events was higher in the recent period than in the baseline period by a significant margin. The test produced a p-value of 0.018, which indicates that the difference in rainfall event between the recent period and baseline period is significant.

Student responses

1. There has been more extreme precipitation days in recent years compared to the baseline of 1960-1990. When we compared the yearly mean of extreme precipitation days, the difference is significantly different. A limitation we have is that we are unable to measure the intensity of said precipitation because our definition of extreme precipitation is over an inch of rain over 24 hours in Ann Arbor.
2. There is statistically significant evidence to show a difference in average precipitation means over the years. A value of ~ 1.4 shows this and according to our t-test/p-value, there is enough evidence to show this difference is significant.
3. There is strong evidence to suggest that the mean number of days with extreme precipitation has increased from the baseline period to the present period, in Ann Arbor, Michigan. We defined a day with extreme precipitation as receiving an inch or more of precipitation.
4. We can conclude that the recent period, which is 1996-2025, has more extreme precipitation than the baseline one, considering the small p-value. In this case, this result and difference are significant.

Student responses

1. We compared the difference in means in recent_years (1991-present) and baseline_years (before 1991). We are 95% confident that the true difference in means between recent_years and baseline_years is between 0.09 and 2.795. We found that the amount of extreme precipitation has increased between these 2 time periods.
2. We compared the extreme days (temperature-wise) from recent days to that of baseline days (years!). Because $p\text{-value} < 0.05$, there is convincing evidence that there is a difference between the mean extreme temps between the diff types of years
3. We find the evidence convincing that the amount of extreme weather days has grown between the two periods. This is because the p-value was low
4. We measured the amount of extreme weather events in modern times (past 30 years) vs further in the past (1960-1990), to see if these events are more common today. After comparing the weather from 1960 to 1990, and 1995-2026, we concluded that there is a statistically significant increase in extreme weather events ($p = 0.03$)

Student responses

1. We were able to use a threshold to define the exceeding precipitation days to compare the precipitation days per year so that it would be easier to compare recent periods to a baseline period. Then we saw that there was a slight 1.5 inch increase in the extreme precipitation days and measured how statistically significant that number was. the t-test showed that we got a p-value of 0.036 which is less than 5%, meaning that it is statistically significant and the change in extreme precipitation
2. What we analyzed is if the frequency of extreme precipitation is greater in recent years than an earlier period as the baseline. From our analysis, we can say that there is a fairly significant amount of evidence that there is not only a difference in the frequency of extreme precipitation, but that more recent years have resulted in more frequent extreme precipitation. This was evidenced by a t-test, but one limitation was that we only included years where all non-null data points were included

Student responses

1. crated a baseline period, a recent period with ab 30 years for each, compared each avg number of extreme precipitation, found it was not significant difference
2. Extreme days are days that has certain value greater or less than our pre-determined thresholds. We can see how temperatures change by sorting using `arrange()` and remove the rows with NA terms and compare.
3. I would say that the t-test confirmed that there is a statistically significant difference between the amount of extreme precipitation days in recent years than the baseline, and that the average number of days per year with extreme precipitaiton in recent years is greater by at least 1 day than the baseline years.
4. In order to extract value from the data to see a trend and if our input is vivable, set a amount foran extreme day and check the location so see if there are othr factors at play, period gives us the intensity.

Student responses

1. It seems that the average temperature has increased and the number of extreme participation days has also increased. This is all over many many decades, so that helps with showing a true growth or decay in data. The evidence is convincing because the p-values for any/all tests were less than 0.05 which tells us our evidence is convincing enough to say there has been a change.
2. our code changed a lot
3. percentage changed
4. So we started with some irregularities in our data and we wanted to see if they were significant in determining temperature differences
5. this code helps us to see the data about extreme rainfall whether of the past year in a statistical view, by showing consistent statistical evidence
6. We discovered that the average temperature has risen by about 2 degrees since we first started tracking it 1891. We discovered that the daily average extreme weather which we defined as over an inch of rain has also increased since we first started tracking it.
7. We group up the different time zone and have the mean of each year, and then count the extreme day to see will the extreme day increase

Student responses

1. We make 2 groups of 29 years for each, and there we find out that there is a increase trend by compare these 2 groups with stats methods by about 1 degree cel in An Arbor. However, we dropped some na, may slightly change the result we get.
2. We set a threshold for what the definition of an extreme precipitation day was, we then computed the count of those extreme days, accouting for missing data in 1960s, split by recent and baseline, and determined signifxcance of resul
3. We took the proportion of days in each year that had extreme rainfall to see if there were any trends. Then we ran a test that saw if it differed from the mean or increased in a meaningful way, looking at the chance that the trends we saw was just due to random chance.

Student responses

1. we went through all historical rainfall data. derived that an extreme day was a day that has had over an inch of rain in a short amount of time. then we calculated the proportion of those days to the days recorded to calculate the fraction of extreme days per year (of recorded data points). We then calculated the mean difference between the two halves of our years (older data 1960-s80s and newer data 1980-2000s) to see weather or not this proportion changed relative to the current date. we discover

Student responses

1. Based on the definition of extreme precipitation being at least 1 inch of rain in a day and we saw that there was a 1.5 percent increase with a p-val that suggests a statistically significant finding. This meant that according to the data we've put into our calculation not including missing values or years will missing days ew have noteced an increasing trend of extreme precipitation.
2. Because our p-value from the t-test was <0.05 , there is evidence to suggest that there is more frequent extreme precipitation frequency from 1996-2025 than 1961-1990. We defined extreme precipitation as the average rainfall in a year greater than an inch
3. It seems that the average number of extreme rainfall days has increased between the two periods. Somewhere between 0.3 and 2.3 days. Extreme rainfall is any day where it rained more than an inch, in Ann Arbor, which the old period being approximately between 1940 and 1990, and the new period being between 1990 and 2025.

Student responses

1. our definition of an extreme day is one where precipitation > 1 in in Ann Arbor, MI over 1960-1990 and 1996-2025. we found that the average extreme precipitation days per year has increased by around 1 day per year over this time period. I dont find this super convincing as that is a negligible amount and could be due to random chance alone.
2. The amount of extreme precipitation days in full years from 2 periods, one from late 1900s to more modern period has increased by around 1. this is convincing, strong evidence that there is an increase from back then to now. our definition of extreme days was greater than 1 inch in a day. A limitation is that it only considers amount, not intensity. Also is in ann arbor
3. We checked if the number of extreme meaning over one inch of precipitation days increased since the last period which went up to 1990, while the other one thats recent is from then to now. The mean number of days per year with extreme precipitation has increased, and the evidence is reasonably convincing based on the p value.

Student responses

1. We constructed a t test in the end which is used I believe to repeated look at samples of data and determine if the standard deviations in the data are major greatly from the mean, and I think we find that it is very small the data and that potentially there is not a trend but I think the limitation is that you really cant see what the data is dicussing in the t test, it doest really give you a value for how the precipitation rain is. Our definition of an extreme day was 25 in celcius and we
2. We saw that the average number of days that had extreme participation has increased in recent years when compared to a baseline period stretching from the 60s to the 90s. An extreme day is any day where there was more than 1 inch of rain fall. The location is for Ann Arbor. The two periods are the reference which is 60s-90s and the recent which is 90s-present. We can say with some certainty that there is a difference in the mean number of extreme weather days now compared to our reference period

Student responses

1. the amount of extreme precipitation increased on average between the two time periods. an extreme day is one that crosses a certain threshold, it was at one location
2. The data was about precipitation trends, and the data showed it differed from 3-6 inches per year. an extreme day would be a p value over 3 in.
3. There is significant evidence that the extreme weather condition got worse over the years as more recent years see more extreme weather
4. WE campoared the days with extreme rain as we defined being greater then one inch in a time period bertwwen 1960 and 1990 and 1996 and 2026. we then compared the means using a test to seeif the means were differnt and to see if the differnce is significant and not just do to random values. when we ran the test we did find a differnce and therfor there is most liklly a significant statistical difference between the two means
5. We found that there is a statistically signifigant increase in the number of extreme rainfall days from the 1960-1990s to 1995-2025, where extreme rainfall is > 1 inch per day. This is a conclusion just for ann arbor michigan

Student responses

1. We measured the difference in extreme precipitation levels over the years to justify whether there was an increase. We can conclude that we have significant statistical evidence to infer a difference between past precipitation levels and precipitation levels in recent years, with the latter being higher.
2. We measured the extreme weather participation and compared a set of baseline years and recent years and tried to figure out whether the number of extreme weather participation has increased across the recent years. The p-value was a very small number indicating a solid trend

Student responses

1. precipitation changed by a small amount, but the evidence is convincing. one limitation is that it's only in Ann Arbor and some data is not there
2. We considered the outliers and we excluded them, we also considered days with missing data, not full 365 days year. It changed a lot and after the changes I find the evidence more convincing. Extreme day is day with precipitation that is higher than what we chose. The location is Michigan Ann Arbor. Periods are divided into to from the 1830s until today.
3. We decided not to only count the days but also make it relevant to the days that data was recorded that year. Then we conducted a t-test to see how accurate the data might be. One limitation might be that the definition of extreme day doesn't give any timeline.
4. We found evidence that the frequency of extreme precipitation has increased over the years. Since we have accounted for missing data and other outliers, our findings are decently strong. One limitation is we don't know the intensity or timing of the precipitation.

Student responses

1. We have filtered out years without complete data and used two time peirods for earlier year and later years. to compare. I think the change is pretty confident as it is over the 0.5 years freshsholdls but because we are msising some values we are not sure.

Student responses

1. There has been an increase in the number of extreme weather days in Ann Arbor since the 1910s. We can be pretty sure that this isn't just chance and that it's actually a general overall increase that we are seeing. Essentially, we see 1.5 more extreme weather days on average compared to someone from like 1932 for example.
2. there is some significant evidence to say we have more extreme days now
3. We discovered that the average daily weather has risen by about 2 degrees since we first started tracking in 1891. We discovered that the amount of extreme precipitation days which we defined as over an inch of rain has risen by about 1.4 days since we first started tracking
4. We measured the change in weather in Ann Arbor over the course of time. We have found that precipitation has modestly increased year by year throughout Ann Arbor's history.

Someone reads our result and says, "Storms are becoming more destructive in Ann Arbor." What additional evidence would you need to evaluate that claim? Suggest one concrete next analysis or source of data, and explain what it would help us learn.

0 anonymous responses

No responses were submitted.

Suppose we wanted to study extreme heat instead. How would you define an extreme day, and which parts of today's analysis could you reuse? Describe one choice or complication you would need to reconsider.

0 anonymous responses

No responses were submitted.