ENVS 193DS Homework 1

Eliana Shandalov

2025-04-12

Homework set up

```
library(tidyverse)

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
✔ dplyr     1.1.4     ✔ readr     2.1.5
✔ forcats   1.0.0     ✔ stringr   1.5.1
✔ ggplot2   3.5.1     ✔ tibble    3.2.1
✔ lubridate 1.9.4     ✔ tidyr     1.3.1
✔ purrr     1.0.4
── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

glaciers <- read_csv("glacial_volume_loss_copy.csv")

Rows: 43 Columns: 11
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
dbl (11): Year, Europe, Arctic, Alaska, Asia, North_America, South_America, ...

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.

# read in data (glacial_volume_loss_copy.csv) and saved as object called `glaciers`
```

Problem 1. Measures of central tendency and data spread

a.

I collected continuous data because each observation is the length of an individual salamander, and they are not integer values; they are decimals.

b.

```
salamander_lengths <- c(4.6, 4.4, 6.2, 5.2, 3.7, 6.0, 3.9, 4.6, 2.7)
# storing salamanders as an object, input lengths in parentheses

mean(salamander_lengths)

[1] 4.588889

# calculating the mean length of salamanders using listed values
```

Sample mean: 4.6 cm

c.

```
salamander_lengths |> # use object 'salamander_lengths' AND THEN
  var(salamander_lengths) # calculate the variance of salamander lengths

[1] 1.228611
```

Sample variance: 1.2 cm

d.

```
  sd(salamander_lengths)

[1] 1.108427

# calculate the standard deviation of salamander lengths
```

Sample standard deviation: 1.1 cm

Problem 2. Visualizing data

a.

Both files have the data organized into columns, but the data in glacial_volume_loss.csv file has full column names as opposed to abbreviations, and other information including units listed above the table.

b.

Who

Who collected and processed the data? Mark Dyurgerov for the University of Colorado created the data and it was edited by W. Meier and R. Armstrong.

What

What are the data about? The data is about mountain and sub polar glaciers because it includes measurements about them from 1945 to 2003.

Why

Why was the data collected? The data was collected to track how glaciers were changing, in term sof cumulative volume change, annual sea level rise, and cumulative rise in sea level between 1945 and 2003 based on the parameters that were studied.

Where

Where was the data collected? The data was collected in Europe, the Arctic, Alaska, Asia, North America, South America, and the Antarctic.

When

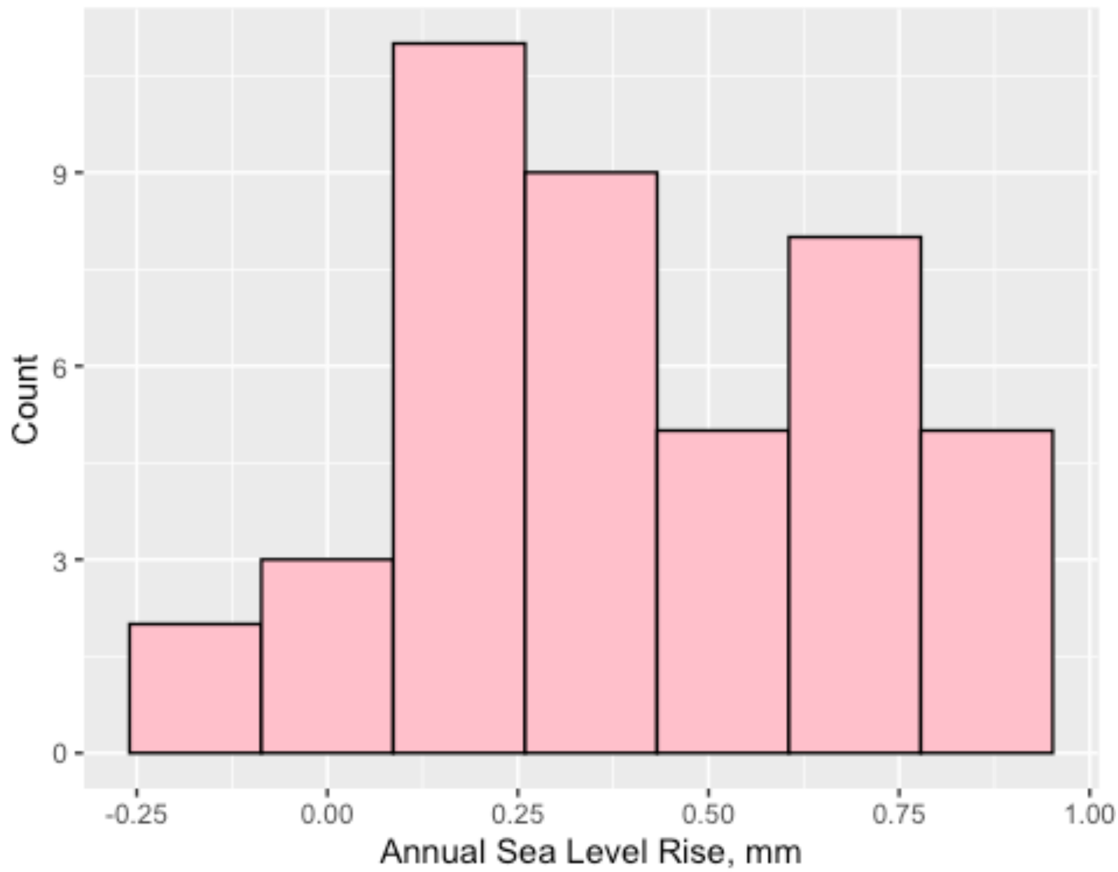
When were the data collected? The data were collected from 1945-2003.

How

How much do the data cost? It seems like the data are free because we are able to access it.

c.

```
ggplot(data = glaciers, # use 'glaciers' data
       aes(x = Annu_sea_rise, # aesthetics: x-axis based on values from
'Annu_sea_rise' column
           fill = type)) + # fill the histogram based on type
  geom_histogram(bins = 7, # 7 bins based on Rice Rule, 43 data points
                 color = "black", # outline color of bins will be black
                 fill = "pink") + # fill color of bins will be pink
labs(x = "Annual Sea Level Rise, mm", # labeling x-axis
     y = "Count") # labeling y-axis
```

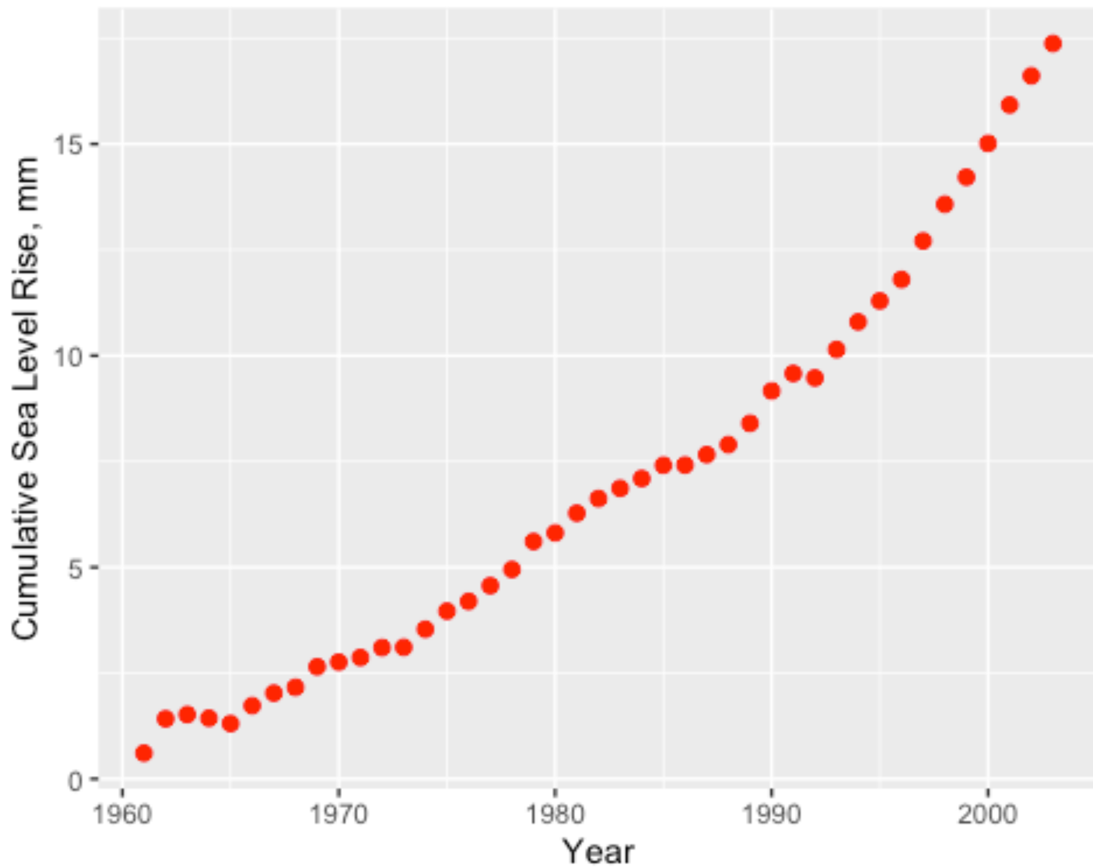



```
# creating the histogram
```

d.

```
ggplot(data = glaciers, # use 'glaciers' data
       aes(x = Year, # aesthetics, x-axis will be from 'Year' column
           y = Cumu_sea_rise)) + # y- axis will be from 'Cumu_sea_rise'
column
   geom_point(size = 2, color = "red") + # adding points of that are size 2

labs(x = "Year", # x-axis retitled to 'Year'
     y = "Cumulative Sea Level Rise, mm") # y-axis retitled to 'Cumulative
Sea Level Rise, mm'
```




```
# Creating the scatterplot
```

Part 3. Personal data

Idea 2.

a.

Do I walk to and from campus slower when I talk on the phone?

b.

My response variable is the time it takes me to walk to or from campus, starting/ending at my house and the beginning of the IV tunnel, which I will measure in minutes.

c.

My predictor variable is on the phone, yes/no.

d.

I would record the date and time of day.

e.

I will record temperature, if it's raining/not, what kind of shoes I'm wearing, and how many minutes before class I leave the house.

f.

Time frame in minutes: continuous Date: mm/dd/yyyy: continuous Time: hh:mm, 24 hour time: continuous Temperature: degrees Farenheit: continuous Raining/not: yes/no: binary categorical Type of shoes: converse, sneakers, flip flops, boots: categorical Minutes before class I leave: minutes, continuous On the phone: yes/no: binary categorical

g.

I will get the time of walk from the stopwatch on my phone, date and time on my home screen of phone, temperature from weather app, raining/not from surroundings and weather app, type of shoes from my surroundings, and minutes before class I leave from home from any clock.

h.

I would take down necessary data before I leave the house, and the reset of it after I arrive at the edge of the IV tunnel.

i.

Date	Time I leave the house	Temperature (degrees F)	Raining: yes/no	Type of shoes	Time before class I leave (min)	Time it takes me to walk to tunnel (min)	On the phone
4/10/2025	14:46	66	No	Flip flops	44	29	No
4/14/2025	11:52	65	No	Sneakers	38	23	No
4/15/2025	9:15	63	No	Converse	45	32	Yes

Idea 1.

a.

Do I go to bed earlier when I have less classes?

b.

My response variable is the time I go to bed, measured on a 24 hour clock.

c.

My predictor variable is how many hours of class I have in a day.

d.

I would record the date and the time I go to bed.

e.

I should measure the time I wake up, how early my class the next day starts, what time I had dinner, and if I go to the gym or not.

f.

Date: mm/dd/yyyy: continuous What time I go to bed: 24 hour clock, hh:mm: continuous How many hours of class I have that day, hh: continuous Time I wake up: 24 hour clock, hh:mm: continuous How early my class the next day is: 24 hour clock, hh:mm: continuous What time I had dinner: 24 hour clock, hh:mm: continuous If I go to the gym: yes/no, binary categorical

g.

I will use a clock for most measurements/my phone, otherwise I will use my surroundings.

h.

I would record the data before I go to bed.

i.

Date	Time I go to bed	Time I wake up	Time I had dinner	Time class next day is	Hours of class I have (hours)	If I go to gym
4/10/2025	0:30	9:00	17:00	12:30	3	yes
4/14/2025	1:20	9:30	18:00	10:00	6	no
4/15/2025	23:50	8:00	17:30	12:30	3	yes

Problem 4. Setting up statistical critique

Paper 1

Douaik, Ahmed, et al. 2011. “Statistical Methods for the Analysis of Soil Spatial and Temporal Variability.” Principles, Application and Assessment in Soil Science.

Paper 2

Schuh, Amy, et al. 2025. “Modeling natural coinfection in a bat reservoir shows modulation of Marburg virus shedding and spillover potential.” PLoS Pathogens 21:3.

Paper 3

Goldberg, Tony, et al. 2008. “Forest Fragmentation as Cause of Bacterial Transmission among Nonhuman Primates, Humans, and Livestock, Uganda.” Emerging Infectious Diseases 14:9.