homework-3-submission

Yonatan Grossman

2025-05-28

Here’s the link

https://github.com/Yonatan-Grossman/ENVS-193DS\_homework-03.git

library(tidyverse) #read in packages

── 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 3.5.2 ✔ 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

library(here)

here() starts at C:/Users/16177/193 DS/ES 193DS/git/ENVS-193DS\_homework\_3/ENVS-193DS\_homework-03

library(flextable)

Attaching package: 'flextable'  
  
The following object is masked from 'package:purrr':  
  
 compose

library(janitor)

Attaching package: 'janitor'  
  
The following objects are masked from 'package:stats':  
  
 chisq.test, fisher.test

library(dplyr)  
library(magick)

Linking to ImageMagick 6.9.12.98  
Enabled features: cairo, freetype, fftw, ghostscript, heic, lcms, pango, raw, rsvg, webp  
Disabled features: fontconfig, x11

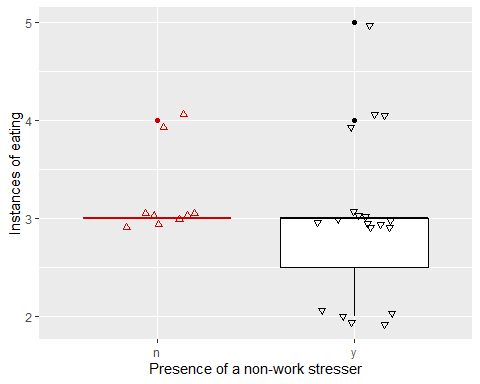
# Problem 1

## a.

I could make side-by-side box and whisker plots comparing the number of times I ate on days where there is and isn’t a non-work stresser. I find that when I’m stressed I tend to work more and spend less time eating so I think the mediam number of times I ate when stressed will be lower, and I think the range will also be lower since the more stressed I am the overall fewer opportunities I give myself to eat.

## b.

personal\_data <- read.csv("C:/Users/16177/193 DS/ES 193DS/git/ENVS-193DS\_homework\_3/ENVS-193DS\_homework-03/data/Personal data - Sheet1.csv") #reading in the personal data  
  
data\_clean <- personal\_data |> #makes a new object called data\_clean which is just the data from personal\_data  
 clean\_names() #cleans the data  
  
  
ggplot(data\_clean,  
 aes(x = presence\_of\_a\_non\_work\_stressor\_y\_n,  
 y = instances\_of\_eating,  
 color = presence\_of\_a\_non\_work\_stressor\_y\_n,  
 shape = presence\_of\_a\_non\_work\_stressor\_y\_n)) +  
 geom\_boxplot(show.legend = FALSE) + #boxplot  
 geom\_jitter(height = 0.1, width = 0.2, show.legend = FALSE) + #jitterplot  
 scale\_color\_manual(values = c("y" = "black", "n" = "red3")) + #custom colors  
 scale\_shape\_manual(values = c("y" = 6, "n" = 2)) + #custom shapes  
 labs (x = "Presence of a non-work stresser", y = "Instances of eating") #axis labels



## c.

**Figure 1: Number of times I eat on days I have a non-work stresser and days I don’t:**  
Comparison of number of times I eat on days I’m stressed for a reason unrelated to work (*n* = 19) and days where I’m not (*n* = 9) The boxplots represent median values, IQR, and spread of observations of number of times I ate food those days. Colors represent the location (black = stressor present, red = no stressor). The triangles represent individual observations.

## d.

personal\_summary <- data\_clean |>   
 group\_by(presence\_of\_a\_non\_work\_stressor\_y\_n) |>   
 summarise(  
 median = median(instances\_of\_eating), #calculating the following  
 n = length(instances\_of\_eating),  
 IQR = IQR(instances\_of\_eating),  
 min = min(instances\_of\_eating),  
 max = max(instances\_of\_eating),  
 .groups = "drop"  
 ) |>   
 rename("non-work stressor?" = presence\_of\_a\_non\_work\_stressor\_y\_n) |> #renaming the titles   
 flextable() |> #make a flextable   
 set\_table\_properties(layout = "autofit")  
personal\_summary #output the flextable

| non-work stressor? | median | n | IQR | min | max |
| --- | --- | --- | --- | --- | --- |
| n | 3 | 9 | 0.0 | 3 | 4 |
| y | 3 | 19 | 0.5 | 2 | 5 |

# Problem 2

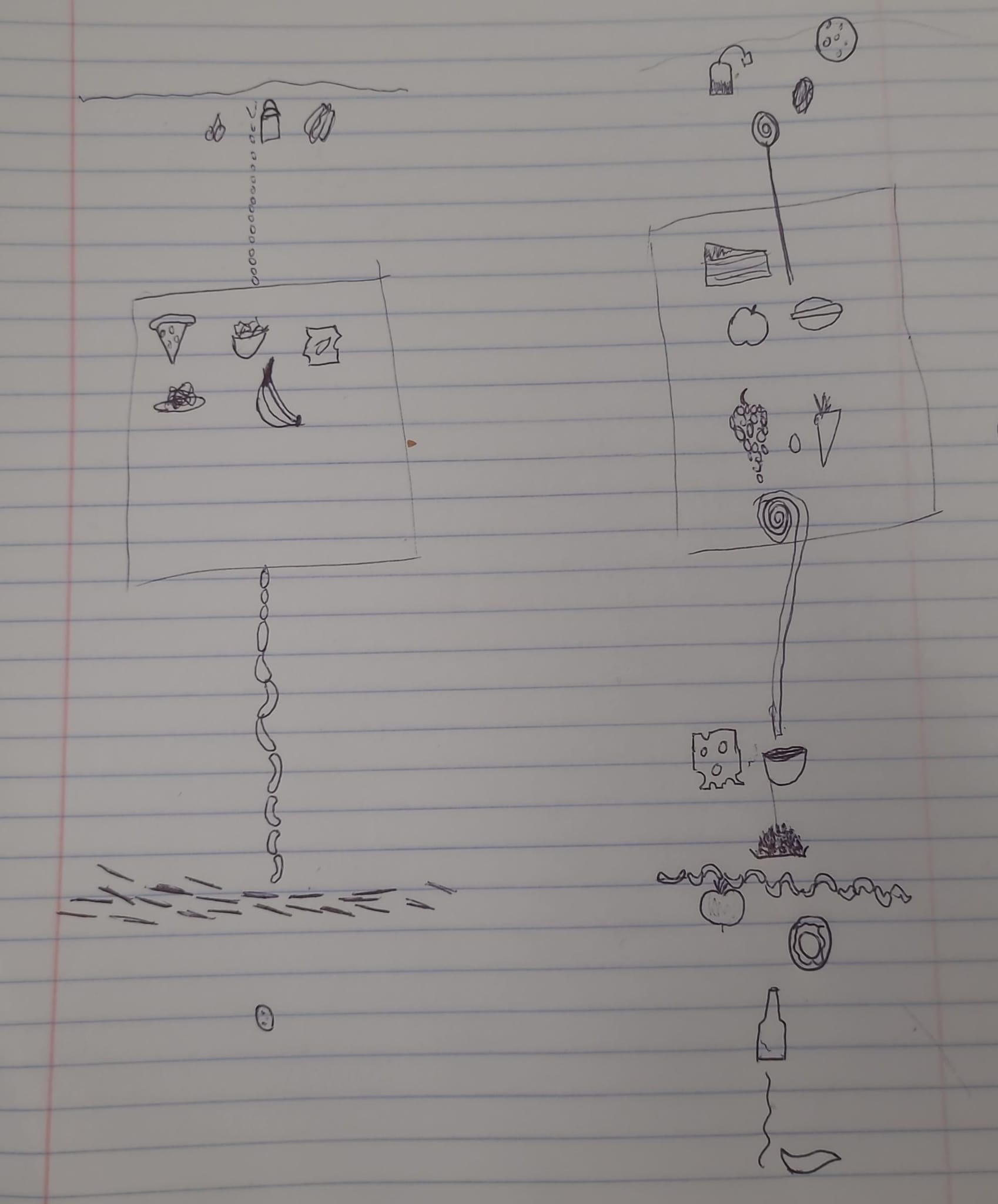
## a.

I think an affective visualization for my personal data could be a a photo or sculpture type thing where food or food wrappers are used as the individual data points for a jitterplot, and then using food boxes and things like uncooked pasta to make the outline for the boxplots. I could use foods of generally lighter colors on a lighter background for the non-stressed side, and then food that’s generally darker for the stressed side, just to give a visual indication of the difference. I really like the idea of using food for this because the dcata points represent times I ate, meaning the data points for this visualization would be represented by the thing they’re represented. I just need to experiment a little bit to find a good way of distinguishing the stressed side from the non-stressed side, and a specific way of showing outliers as distrinct from the jitter data points

## b.

sketch <- image\_read("C:/Users/16177/193 DS/ES 193DS/git/ENVS-193DS\_homework\_3/ENVS-193DS\_homework-03/WhatsApp Image 2025-05-28 at 2.44.03 PM.jpeg")  
  
sketch\_scaled <- image\_scale(sketch, "800")  
  
print(sketch)

# A tibble: 1 × 7  
 format width height colorspace matte filesize density  
 <chr> <int> <int> <chr> <lgl> <int> <chr>   
1 JPEG 1698 2048 sRGB FALSE 138968 72x72



## c.

# Problem 3

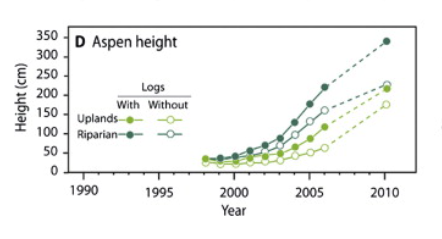
## a.

The statistical test used was a Student’s t-test. The response predictor variable was aspen location (Riparian vs. upland) and the response variable was current annual growth (of the aspen).

## b.

scrnsht <- image\_read("C:/Users/16177/OneDrive/Pictures/Screenshots/Screenshot 2025-04-24 233738.png")  
  
  
print(scrnsht)

# A tibble: 1 × 7  
 format width height colorspace matte filesize density  
 <chr> <int> <int> <chr> <lgl> <int> <chr>   
1 PNG 442 239 sRGB TRUE 72119 38x38



This is the image I included in homework 2, which wasn’t exactly the data visualization of the statistical test, but I’ll critique it for visual clarity since otherwise I would just have to find another figure anyway.

The authors did a pretty good job of responding their data. The x- and y-axis are properly labeled, the 4 trends are visually distinct and intuitively labeled in the legend, and there’s a title. The predictions are visually distinct (dotted line) and the data points are provided. Also units were provided where necessary.

## c.

I think the authors did a pretty good job in regards to visual clutter, the actual data points are the only color on the whole chart (except the legend) which draws your eyes towards it. The essential items are included and the data points are made distinct without going overboard, and the legend is a grid to improve efficiency of space. On the whole there’s not that much data on this chart, but I don’t think you could remove much more without getting rid opf important elements.

## d.

I would recommend they use box and whisker plots with underlying jitterplots to represent the data. The x-axis would be site type, and have “riparian” and “upland” as options. The y-axis would be the current annual growth rate in cm/year. For both site types there would be a box and whisker plot showing the range, IQR, median, and outliers of the data for that site, plus a jitterplot (with height = 0) showing the actual data points themselves. The plot should have colors that are quite distinct from each other, and unique shapes should be chosen for the jitterplots. Horizontal gridlines should be included to make it easier to read the jitterplots. This would allow someone to easily at a glance see the differences in the growth rates between the sites using the boxplots, which would also help them see the range and general spread of the data. The jitterplot would let people still see the underlying data for more robustness.