homework\_03

Sanjana Sujeet

2025-05-28

library(tidyverse)  
library(janitor)  
library(here)  
library(gt)  
library(dplyr)  
library(ggplot2)

**GitHub Repository:** <https://github.com/sanjanasujeet/ENVS-193DS_workshop-07>

## a. Data Summarizing

My data tracked hours I read, pages I read, and social hours per day as my response variables and stress level as my predictor variable. I don’t think I will analyze all these variables but the main thing that I want to look at is how stress levels affect my reading.

I will compare the number of hours I read per day to my quantified daily stress level, which I calculated based on the number of upcoming assignments and tests, as well as my perceived stress. I also want to calculate my reading pace by dividing pages read by minutes reading and compare this to my stress level to see if stress influences the pace that I read.

To summarize my data, I could calculate the mean number of pages I read per day across different stress levels ( low, medium, high stress). I could also compare the average reading pace between these groups to see if stress impacts not just how much I read, but how fast I read. Comparing means within stress levels is informative because it is a simple way to see if there is a difference between how different stress levels affect me. I may read less efficiently when I feel overwhelmed, or I may read more when trying to avoid stress through a calming activity.

## b. Visualization and c. Caption

## cleaning data  
data<- read.csv("data/springtracking2.csv") #read in data  
data\_clean<-clean\_names(data) #cleaning column names  
data\_clean <- data\_clean |>   
 rename(hours = hours\_of\_activity) #shortens name cuz i think i will be using it often  
data\_clean <- data\_clean |>   
 mutate(pace = pages\_read / hours) #creates new column pace which is pages/min  
data\_clean <- data\_clean %>%  
 mutate(pace = replace(pace, is.nan(pace), 0))# removes NaN and makes them 0 instead(this was a calculation issue)  
data\_clean <- data\_clean %>%  
 mutate(stress\_level = case\_when(  
 quantified\_stress >= 0 & quantified\_stress <= 9 ~ "low",  
 quantified\_stress >= 10 & quantified\_stress <= 19 ~ "medium",  
 quantified\_stress >= 20 ~ "high")) # creates new column called stress\_level and takes quantified stress levels and makes 3 levels out of it  
  
data\_clean <- data\_clean %>%  
 mutate(stress\_level = factor(stress\_level, levels = c("low", "medium", "high"))) # orders the levels

#graphing  
  
ggplot(data\_clean %>%   
 filter(!is.na(quantified\_stress), !is.na(pages\_read)),  
 aes(x = quantified\_stress, y = pages\_read)) +  
 geom\_point(color = "cyan4", alpha = 0.7, size = 2) +  
 geom\_smooth(method = "lm", color = "royalblue", se = TRUE) +  
 labs(  
 title = "Pages Read vs. Quantified Stress",  
 x = "Quantified Stress",  
 y = "Pages Read"  
 ) +  
 theme\_minimal()

ggplot(data\_clean |> filter(!is.na(stress\_level), !is.na(pages\_read)),  
 aes(x = stress\_level, y = pages\_read, fill = stress\_level)) +  
 geom\_boxplot() +  
 scale\_fill\_manual(values = c("low" = "lightblue", "medium" = "cadetblue", "high" = "cyan4")) +  
 labs(  
 title = "Pages Read by Stress Level",  
 x = "Stress Level",  
 y = "Pages Read"  
 ) +  
 theme\_minimal() +  
 theme(legend.position = "none")

ggplot(data\_clean %>%   
 filter(!is.na(quantified\_stress), !is.na(pace), pace != 0),  
 aes(x = quantified\_stress, y = pages\_read)) +  
 geom\_point(color = "salmon", alpha = 0.7, size = 2) +  
 geom\_smooth(method = "lm", color = "darkred", se = TRUE) +  
 labs(  
 title = "Pace vs. Quantified Stress (Excluding no reading days)",  
 x = "Quantified Stress",  
 y = "Pace(pages per hour)"  
 ) +  
 theme\_minimal()

ggplot(data\_clean |> filter(!is.na(stress\_level), !is.na(pace)),  
 aes(x = stress\_level, y = pace, fill = stress\_level)) +  
 geom\_boxplot() +  
 scale\_fill\_manual(values = c("low" = "rosybrown1", "medium" = "salmon", "high" = "indianred4")) +  
 labs(  
 title = "Pages Read by Stress Level (Excluding no reading days)",  
 x = "Stress Level",  
 y = "Pace(Pages per Hour)"  
 ) +  
 theme\_minimal() +  
 theme(legend.position = "none")