

Untitled

Homework 3

Part 1. Setup

```
# read in libraries
library(tidyverse)
```

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

```
library(here)
```

here() starts at /Users/tanveersingh/github/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(readxl)
```

```
# read in personal data
```

```
mydata <- read_csv("Personal Data Project - Sheet1.csv")
```

Rows: 13 Columns: 9

-- Column specification -----

Delimiter: ","

chr (4): Date, Time of day, Location, Light

dbl (4): Number of pages, Total time of reading, Distractions #, Comprehension

time (1): Time started

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.

Part 2. Problems

a.

To summarize the data and compare a response variable between categories, I could calculate the pages per minute for each session to compare reading effectiveness across different locations. This comparison would be informative because different environments might offer varying levels of comfort or light, which could impact my focus, and consequently, how well I understood the material.

b.

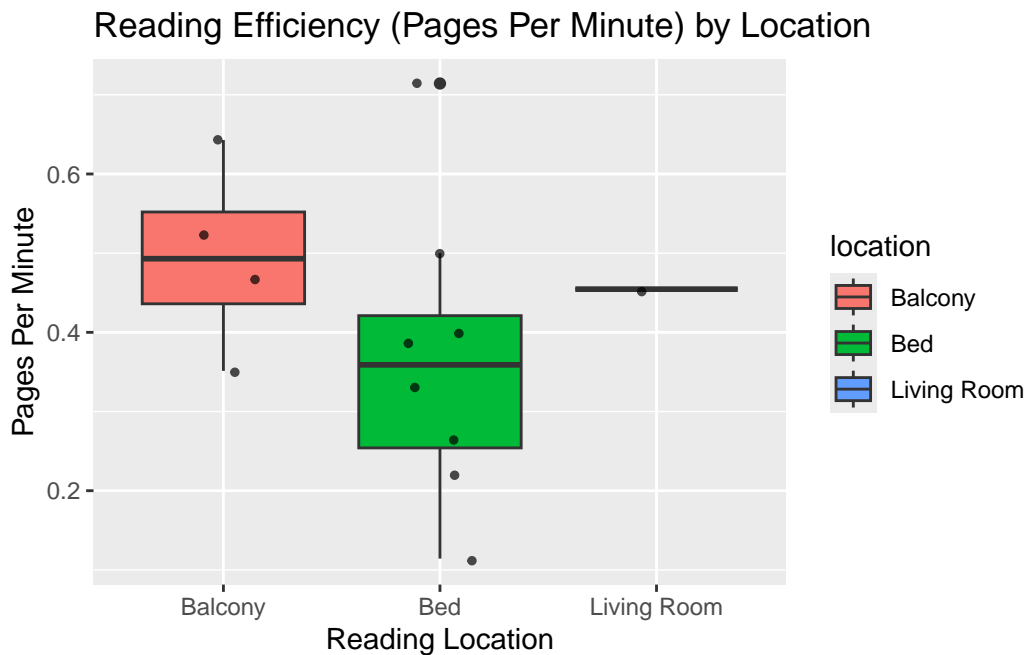
```

df <- clean_names(mydata) |> # clean column names
# adding pages_per_minute column
mutate(pages_per_minute = ifelse(total_time_of_reading > 0, number_of_pages / total_time_of_reading, 0))

mean_pages_per_minute <- df |>
  group_by(location) |>
  summarise(Mean_PPM = mean(pages_per_minute)) |>
  rename("Location" = location)

ggplot(df, aes(x = location,
               y = pages_per_minute)) +
  geom_boxplot(aes(fill = location)) +
  geom_jitter(width = 0.2, size = 1, color = "black", alpha = 0.7) +
  labs(
    x = "Reading Location",
    y = "Pages Per Minute",
    title = "Reading Efficiency (Pages Per Minute) by Location"
  )

```



c.

Data collected from reading sessions between May 12-26, 2025. Efficiency measured as pages read per minute

d.

```
table1 <- flextable(mean_pages_per_minute) |>
  colformat_double(j = "Mean_PPM", digits = 1) |>
  autofit() |>
  theme_booktabs() |>
  align(align = "center", part = "all")
```

table1

Location	Mean_PPM
Balcony	0.5
Bed	0.4
Living Room	0.5