Final Project

Kathryn Ellestad

2025-04-28

Libraries

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.4.3
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.4.3
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.4.3
```

```
## Warning: package 'purrr' was built under R version 4.4.3
```

```
## Warning: package 'dplyr' was built under R version 4.4.3
```

```
## Warning: package 'stringr' was built under R version 4.4.3
```

```
## Warning: package 'lubridate' was built under R version 4.4.3
```

```
## -- 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 lubridate 1.9.4     v tibble    3.2.1
## v purrr     1.0.4     v tidyr     1.3.1
## -- 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 errors
```

```
library(plotly)
```

```
## Warning: package 'plotly' was built under R version 4.4.3
```

```
##
## Attaching package: 'plotly'
##
## The following object is masked from 'package:ggplot2':
##
##      last_plot
##
## The following object is masked from 'package:stats':
##
##      filter
##
## The following object is masked from 'package:graphics':
##
##      layout
```

Aesthetics

```r
theme_set(
  theme_minimal(base_size = 14, base_family = "Georgia") +
    theme(
      plot.title = element_text(face = "bold", color = "#f8f8f8", hjust = 0.5),
      plot.subtitle = element_text(color = "#cccccc", hjust = 0.5),
      axis.title.x = element_text(color = "#f8f8f8", margin = margin(t = 10)),
      axis.title.y = element_text(color = "#f8f8f8", margin = margin(r = 10)),
      axis.text.x = element_text(color = "#f8f8f8"),
      axis.text.y = element_text(color = "#f8f8f8"),
      panel.background = element_rect(fill = "#222222", color = NA),
      plot.background = element_rect(fill = "#222222", color = NA),
      panel.grid.major.x = element_line(color = "grey30"),
      panel.grid.major.y = element_line(color = "grey30"),
      panel.grid.minor = element_blank(),
      legend.position = "none"
    )
)

pastel_colors <- c("Physical" = "#C3E2C2", "Eink" = "#D8C7FF")
```

Data

```r
SleepandReadData <- read_excel("C:/Users/kathryne/OneDrive/School/Data211/Final Project/Reading and Sle
SleepandReadData <- SleepandReadData %>%
  mutate(Format = str_to_title(Format))
```

Hypothesis 1: Does the format I use significantly affect the minutes spent reading?

Hypothesis

μ_e = Average minutes spent reading eink

μ_p = Average minutes spent reading physical

$H_0 : \mu_e = \mu_p$

$H_a : \mu_e \neq \mu_p$

```r
t.test(MinutesReading ~ Format, data = SleepandReadData, alternative = "two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  MinutesReading by Format
## t = -0.8722, df = 11.854, p-value = 0.4004
## alternative hypothesis: true difference in means between group Eink and group Physical is not equal
## 95 percent confidence interval:
##  -75.03211  32.17497
## sample estimates:
##     mean in group Eink mean in group Physical
##               50.57143               72.00000
```

Conclusion:

p-value $= .4004 > \alpha = .05$, therefore with a 5% significance level, we fail to reject H_0. There is not sufficient evidence to support the claim that the format used significantly affects the amount of minutes spent reading.

```r
ColChart <- ggplot(SleepandReadData, aes(x = Format, y = MinutesReading, fill = Format)) +
  stat_summary(fun = "mean", geom = "col", width = 0.6) +
  scale_fill_manual(values = pastel_colors) +
  labs(title = "Average Reading Time by Format",
       y = "Minutes Reading",
       x = "Format") +
  coord_flip()

ggplotly(ColChart)
```

```
## file:///C:\Users\kathryne\AppData\Local\Temp\Rtmp67jrss\file6dc02c2157cb\widget6dc04fb92ee1.html scre
```

Average Reading Time by Format
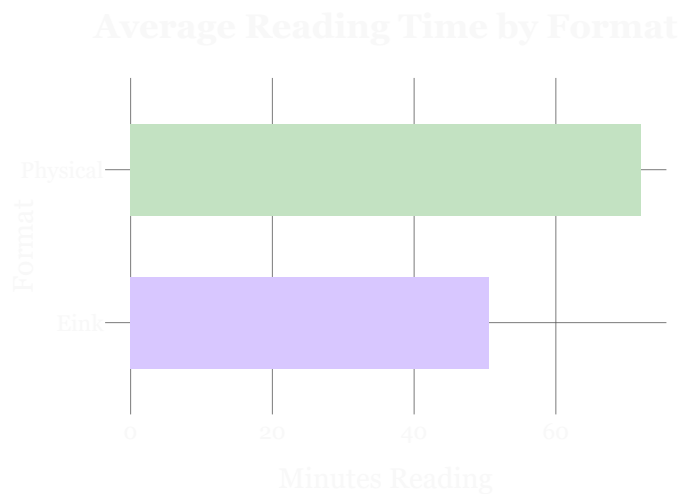
Hypothesis 2: Is the amount of sleep I get different based on the format I read before bed?

Hypothesis

μ_e = Average sleep minutes after eink

μ_p = Average sleep minutes after physical

$H_0 : \mu_e = \mu_p$

$H_a : \mu_e \neq \mu_p$

```
t.test(TotalTimeAsleep ~ Format, data = SleepandReadData, alternative = "two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  TotalTimeAsleep by Format
## t = -1.8106, df = 11.784, p-value = 0.09576
## alternative hypothesis: true difference in means between group Eink and group Physical is not equal
## 95 percent confidence interval:
##  -213.96426   19.96426
## sample estimates:
##     mean in group Eink mean in group Physical
##               295.4286               392.4286
```

Conclusion:

p-value $= .09576 > \alpha = .05$, therefore with a 5% significance level, we fail to reject H_0. There is not sufficient evidence to support the claim that the format chosen greatly affects the amount of sleep I get each night.

```
SummaryTable <- SleepandReadData %>%
  group_by(Format) %>%
  summarise(
    ave = mean(TotalTimeAsleep, na.rm = TRUE),
    se = sd(TotalTimeAsleep, na.rm = TRUE) / sqrt(n()),
    n = n(),
    tstar = qt(1 - 0.05/2, df = n() - 1)
  )

SleepvsFormat <- ggplot(SummaryTable, aes(x = Format, y = ave, fill = Format)) +
  geom_col(width = 0.5, color = "white") +
  geom_errorbar(aes(ymin = ave - tstar * se, ymax = ave + tstar * se), width = 0.2, linewidth = 0.8) +
  geom_point(shape = 21, size = 4, color = "black", fill = "white", stroke = 1.2) +
  scale_fill_manual(values = pastel_colors) +
  labs(title = "Average Total Sleep by Format",y="Average Minutes Slept")

ggplotly(SleepvsFormat)
```

```
## file:///C:\Users\kathryne\AppData\Local\Temp\Rtmp67jrss\file6dc03137d5b\widget6dc01b7655c5.html scre
```

Average Total Sleep by Format
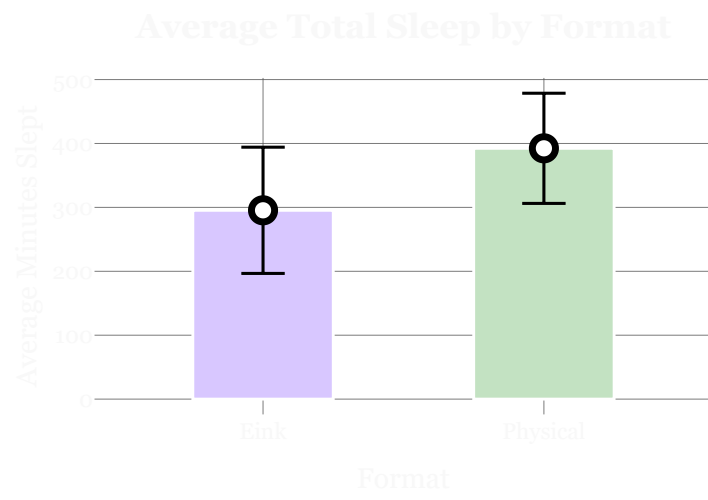
Hypothesis 3: Does the format that I use affect the speed at which I read?

Hypothesis

μ_e = Average pages per minute with eink.

μ_p = Average pages per minute with physical

$H_0 : \mu_e = \mu_p$

$H_a : \mu_e \neq \mu_p$

```
t.test(PagesPerMinute ~ Format, data = SleepandReadData, alternative = "two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  PagesPerMinute by Format
## t = -0.60393, df = 8.2984, p-value = 0.562
## alternative hypothesis: true difference in means between group Eink and group Physical is not equal
## 95 percent confidence interval:
##  -1.608408  0.937481
## sample estimates:
##      mean in group Eink mean in group Physical
##                1.543333               1.878797
```

Conclusion:

p-value = .562 > α = .05, therefore with a 5% significance level, we fail to reject H_0. There is not sufficient evidence to support the claim that the format used greatly affects reading speed.

```
Barchart <- SleepandReadData %>%
  group_by(Format) %>%
  summarise(AvgPages = mean(Pages)) %>%
  ggplot(aes(x = Format, y = AvgPages, fill = Format)) +
  geom_col() +
  scale_fill_manual(values = pastel_colors) +
  labs(title = "Average Pages Read by Format",
       y = "Average Pages",
       x = "Format")

ggplotly(Barchart)
```

```
## file:///C:\Users\kathryne\AppData\Local\Temp\Rtmp67jrss\file6dc061b65400\widget6dc044665498.html scr
```

Average Pages Read by Format
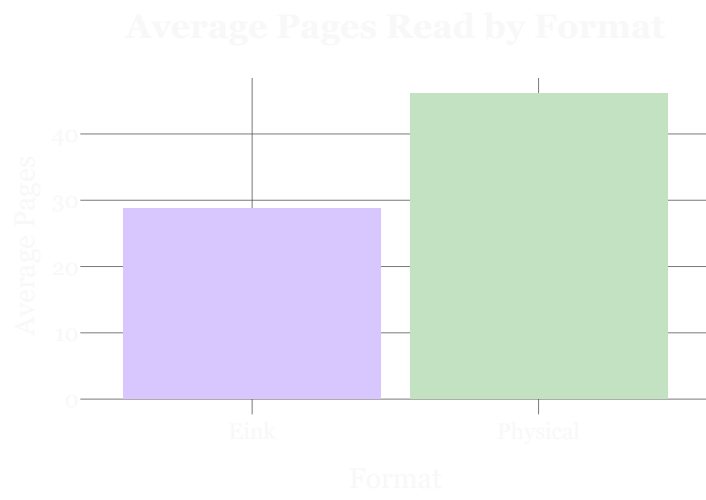
Hypothesis 4: Does the format that I use affect how quickly I fall asleep?

Hypothesis

μ_e = Average minutes to fall asleep after eink

μ_p = Average minutes to fall asleep after physical

$H_0 : \mu_e = \mu_p$

$H_a : \mu_e \neq \mu_p$

```
t.test(MinutesToSleep ~ Format, data = SleepandReadData, alternative = "two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  MinutesToSleep by Format
## t = -0.92872, df = 6.0311, p-value = 0.3887
## alternative hypothesis: true difference in means between group Eink and group Physical is not equal
## 95 percent confidence interval:
##  -112.57453    50.57453
## sample estimates:
##     mean in group Eink mean in group Physical
##               9.142857              40.142857
```

Conclusion:

p-value $= .3887 > \alpha = .05$, therefore with a 5% significance level, we fail to reject H_0. There is not sufficient evidence to support the claim that the format used greatly affects how long it takes to fall asleep.

```
TimetoSleepPlot <- ggplot(SleepandReadData, aes(x = Format, y = MinutesToSleep, fill = Format)) +
  geom_jitter(width = 0.2, size = 3.5, shape = 21, color = "black", stroke = 1.2) +
  stat_summary(fun = mean, geom = "point", shape = 23, size = 5, fill = "white", color = "black", stroke
  stat_summary(fun.data = mean_se, geom = "errorbar", width = 0.2, linewidth = 1) +
  scale_fill_manual(values = pastel_colors) +
  labs(title = "Minutes to Fall Asleep by Format.\n (Note: one extreme outlier clipped for clarity)",y=
  coord_cartesian(ylim = c(0, 100))

ggplotly(TimetoSleepPlot)
```

```
## file:///C:\Users\kathryne\AppData\Local\Temp\Rtmp67jrss\file6dc04e57724f\widget6dc072a0d2d.html scre
```
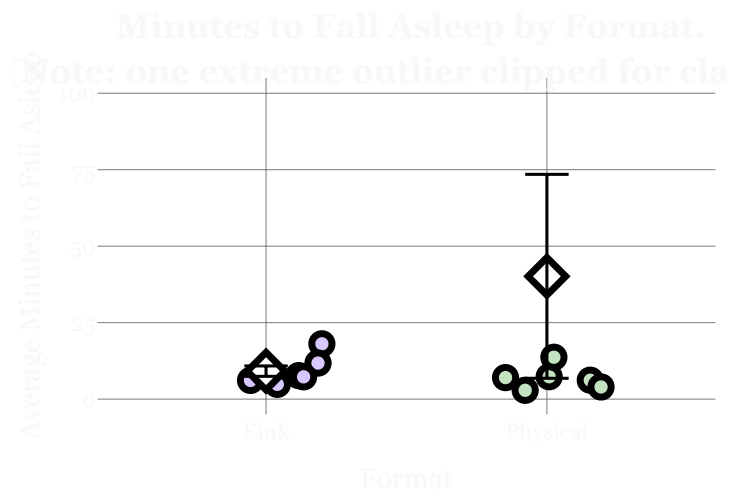
Graphs

Scatter plot of amount of sleep and minutes read

```r
Scatterplot <- ggplot(SleepandReadData, aes(x = MinutesReading, y = TotalTimeAsleep, color = Format)) +
  geom_point(size = 4, alpha = 0.8) +
  scale_color_manual(values = pastel_colors) +
  labs(title = "Total Time Asleep vs Reading Time",
       x = "Minutes Reading",
       y = "Minutes Asleep")

ggplotly(Scatterplot)
```

```
## file:///C:\Users\kathryne\AppData\Local\Temp\Rtmp67jrss\file6dc06bb73393\widget6dc04ba170c5.html scr
```

Minutes Asleep

Minutes Reading
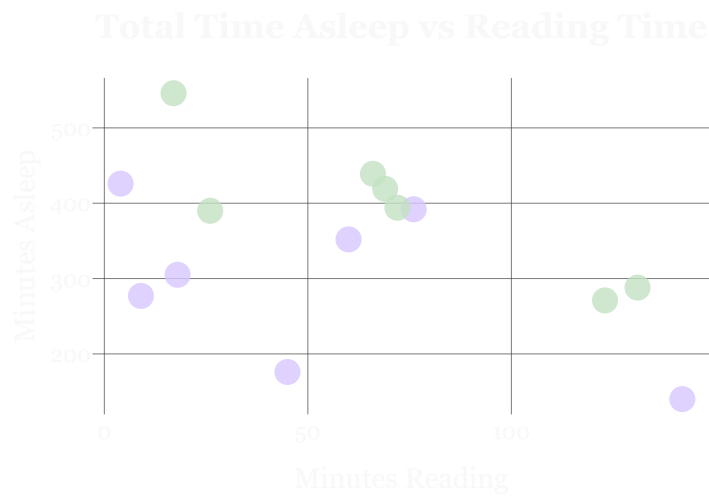
Stacked/clustered bar chart of average sleep stage minutes by format

```r
SleepStageAvg <- SleepandReadData %>%
  group_by(Format) %>%
  summarise(
    REM = mean(REM),
    Light = mean(Light),
    Deep = mean(Deep)
  ) %>%
  pivot_longer(cols = c(REM, Light, Deep), names_to = "Stage", values_to = "AvgMinutes")

StackedBars <- ggplot(SleepStageAvg, aes(x = Format, y = AvgMinutes, fill = Stage)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = c("REM" = "#CBBBDD", "Light" = "#BFD8B8", "Deep" = "#FFD8BE")) +
  labs(title = "Average Minutes in Sleep Stages by Format",
       x = "Reading Format",
       y = "Average Minutes",
       fill = "Sleep Stage")

ggplotly(StackedBars)
```

```
## file:///C:\Users\kathryne\AppData\Local\Temp\Rtmp67jrss\file6dc050aa2771\widget6dc030e4364.html scre
```

Average Minutes in Sleep Stages by Format

Average Minutes in Sleep Stages by Format

Reading Format
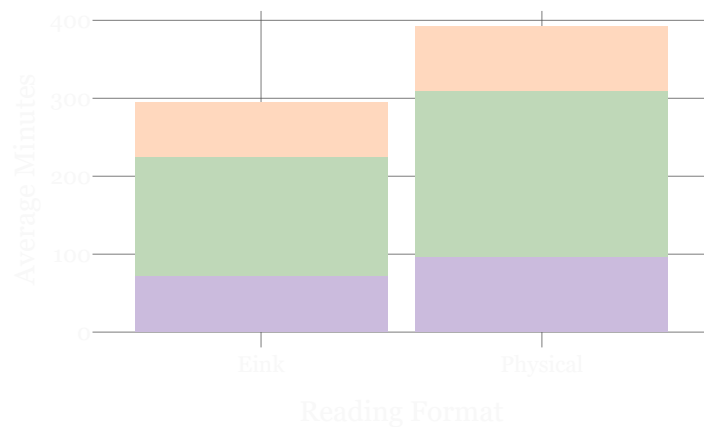
Line chart showing reading time by date

```
Lineplot <- ggplot(SleepandReadData, aes(x = Date, y = MinutesReading)) +
  geom_line(linewidth = 1.2, color = "#c3e2c2") +
  geom_point(size = 3, color = "#c3e2c2") +
  labs(title = "Reading Time Over Two Weeks",
       x = "Date",
       y = "Minutes Reading")

ggplotly(Lineplot)
```

```
## file:///C:\Users\kathryne\AppData\Local\Temp\Rtmp67jrss\file6dc015d03e74\widget6dc0dbf768.html scree
```

Reading Time Over Two Weeks
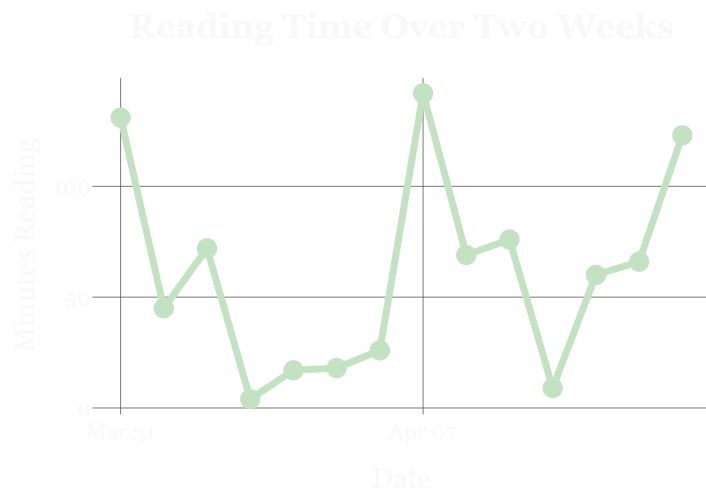
Bar chart showing average pages read by day of week

```
DayOfWeekSummary <- SleepandReadData %>%
  mutate(DayOfWeek = factor(DayofWeek, levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday"
  group_by(DayOfWeek) %>%
  summarise(AvgPages = mean(Pages, na.rm = TRUE))

DayofWeekPlot <- ggplot(DayOfWeekSummary, aes(x = DayOfWeek, y = AvgPages, fill = DayOfWeek)) +
  geom_col() +
  scale_fill_manual(values = c("#C3E2C2", "#D8C7FF", "#FFD8BE", "#BFD8B8", "#FBE4C9", "#CBBBDD", "#A3C9A
  labs(title = "Average Pages Read by Day of Week",
       y = "Average Pages",
       x = "Day of Week")

ggplotly(DayofWeekPlot)
```

```
## file:///C:\Users\kathryne\AppData\Local\Temp\Rtmp67jrss\file6dc076cd7bc7\widget6dc02de7338d.html scr
```

Average Pages Read by Day of Week