

Final Project

Kathryn Ellestad

2025-04-28

Libraries

```
library(readxl)
library(ggplot2)
library(tidyverse)
```

```
## Warning: package 'tidyr' 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      1.0.2      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

```
if (knitr::is_latex_output()) {  
  # PDF output: light theme  
  theme_set(  
    theme_minimal(base_size = 14, base_family = "Georgia") +  
    theme(  
      plot.title = element_text(face = "bold", color = "black", hjust = 0.5),  
      plot.subtitle = element_text(color = "black", hjust = 0.5),  
      axis.title.x = element_text(color = "black", margin = margin(t = 10)),  
      axis.title.y = element_text(color = "black", margin = margin(r = 10)),  
      axis.text.x = element_text(color = "black"),  
      axis.text.y = element_text(color = "black"),  
      panel.background = element_rect(fill = "white", color = NA),  
      plot.background = element_rect(fill = "white", color = NA),  
      panel.grid.major.x = element_line(color = "grey70"),  
      panel.grid.major.y = element_line(color = "grey70"),  
      panel.grid.minor = element_blank(),  
      legend.position = "none"  
    )  
  )  
} else {  
  # HTML output: dark theme  
  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")  
  
#used https://ggplot2.tidyverse.org/reference/ggtheme.html to learn how to change the theme.
```

Data

```
SleepandReadData <- read_excel("C:/Users/kathr/OneDrive/School/Data211/Final Project/Data 211 project T  
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

$H_0 : \mu_e = \mu_p$

$H_a : \mu_e \neq \mu_p$

```
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:

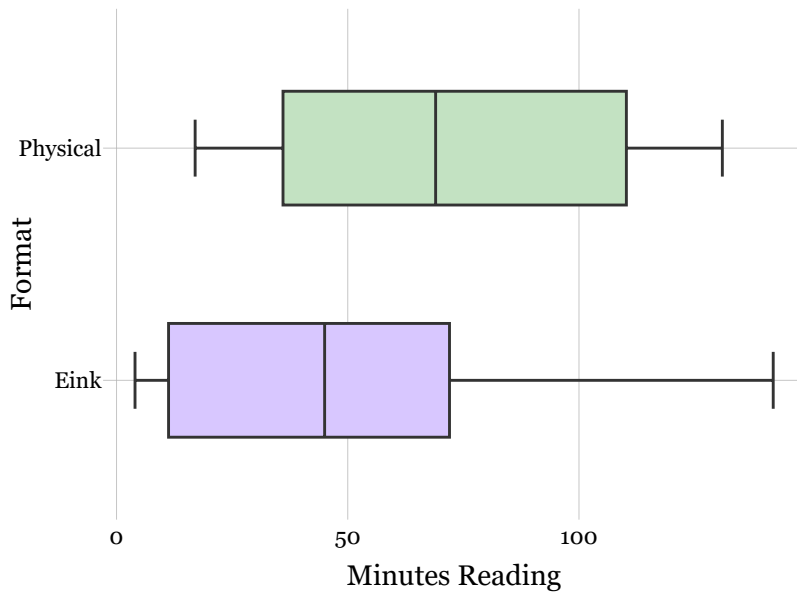
p-value = .4004 > α = .05, therefore with a 5% significance level, we fail to reject H_0 . There is not sufficient evidence to support the claim that the format used significantly affects the amount of minutes spent reading.

```
Boxplot <- ggplot(SleepandReadData, aes(x = Format, y = MinutesReading, fill = Format)) +
  geom_boxplot(width = 0.6,
    outlier.shape = 21,
    outlier.fill = "white",
    outlier.color = "black",
    outlier.size = 2) +
  scale_fill_manual(values = pastel_colors) +
  labs(
    title = "Reading Time by Format",
    y = "Minutes Reading",
    x = "Format"
  ) +
  coord_flip()

ggplotly(Boxplot)
```

```
## file:///C:/Users/kathr/AppData/Local/Temp/RtmpsBabTL/file6b8473011991/widget6b84353d33d3.html screen
```

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:

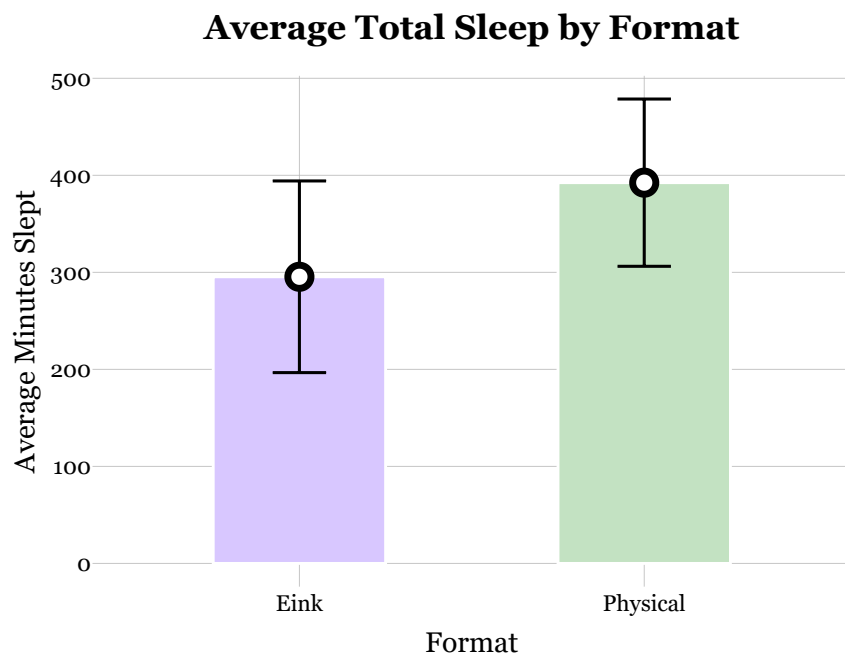
p-value = .09576 > α = .05, therefore with a 5% significance level, we fail to reject H_0 . There is not sufficient evidence to support the claim that the format chosen greatly affects the amount of sleep I get each night.

```
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/kathr/AppData/Local/Temp/RtmpsBabTL/file6b84728e2b6e/widget6b8458f41c1c.html screen
```



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 = -0.60455, df = 8.2952, p-value = 0.5616
## alternative hypothesis: true difference in means between group Eink and group Physical is not equal
## 95 percent confidence interval:
## -1.6088694 0.9372314
## sample estimates:
## mean in group Eink mean in group Physical
## 1.542957 1.878776
```

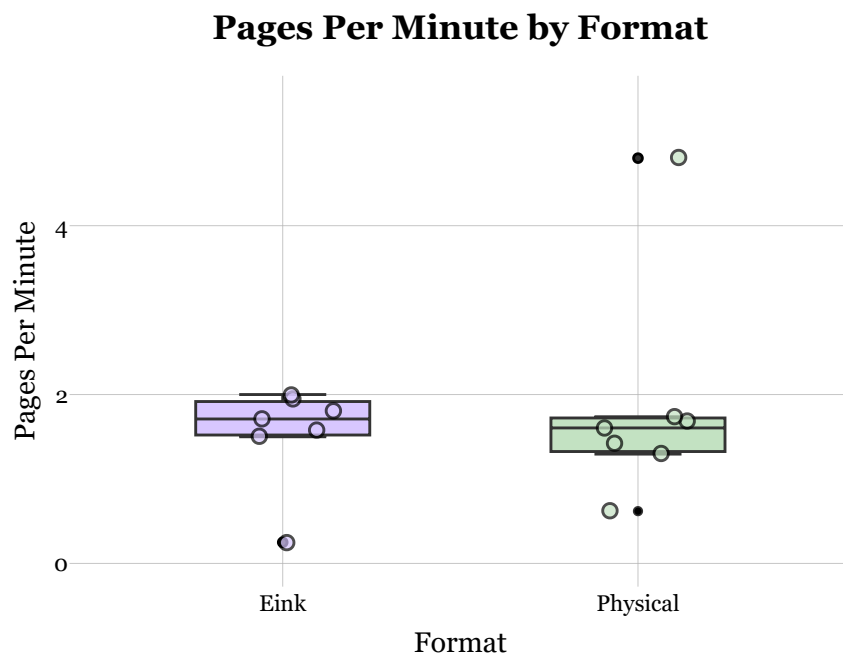
Conclusion:

p-value = .562 > $\alpha = .05$, 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 my reading speed.

```
PagesBoxPlot <- ggplot(SleepandReadData, aes(x = Format, y = PagesPerMinute, fill = Format)) +
  geom_boxplot(width = 0.5,
    outlier.shape = 21,
    outlier.fill = "white",
    outlier.color = "black",
    outlier.size = 2) +
  geom_jitter(width = 0.15,
    size = 2.5,
    color = "black",
    alpha = 0.7) +
  scale_fill_manual(values = pastel_colors) +
  labs(
    title = "Pages Per Minute by Format",
    y = "Pages Per Minute",
    x = "Format"
  ) +
  ylim(0, 5.5)

ggplotly(PagesBoxPlot)
```

```
## file:///C:/Users/kathr/AppData/Local/Temp/RtmpsBabTL/file6b84499573d7/widget6b84cbc793e.html screens
```

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 > 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 how long it takes to fall asleep.

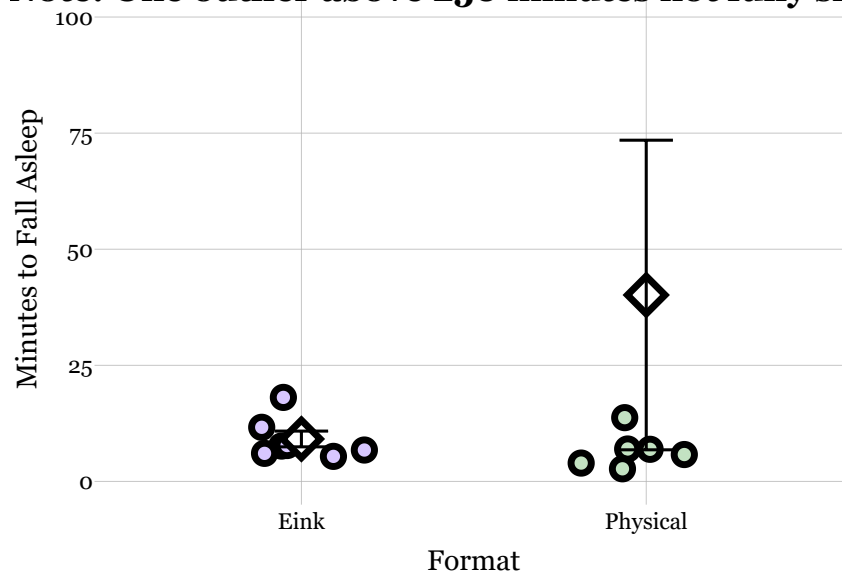
```
TimetoSleepPlot <- ggplot(SleepandReadData, aes(x = Format, y = MinutesToSleep, fill = Format)) +
  geom_jitter(width = 0.2,
    size = 3.5,
    shape = 21,
    color = "black",
    stroke = 1.2) +
  stat_summary(fun = mean,
    geom = "point",
    shape = 23,
    size = 5,
    fill = "white",
    color = "black",
    stroke = 1.2) +
  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 outlier above 250 minutes not fully shown)",
    y = "Minutes to Fall Asleep",
    x = "Format"
  ) +
  coord_cartesian(ylim = c(0, 100))
```

```
ggplotly(TimetoSleepPlot)
```

```
## file:///C:/Users/kathr\AppData/Local/Temp/RtmpsBabTL/file6b8454715a21/widget6b8467055dad.html screen
```

Minutes to Fall Asleep by Format

Note: One outlier above 250 minutes not fully shown



Additional Graphs

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

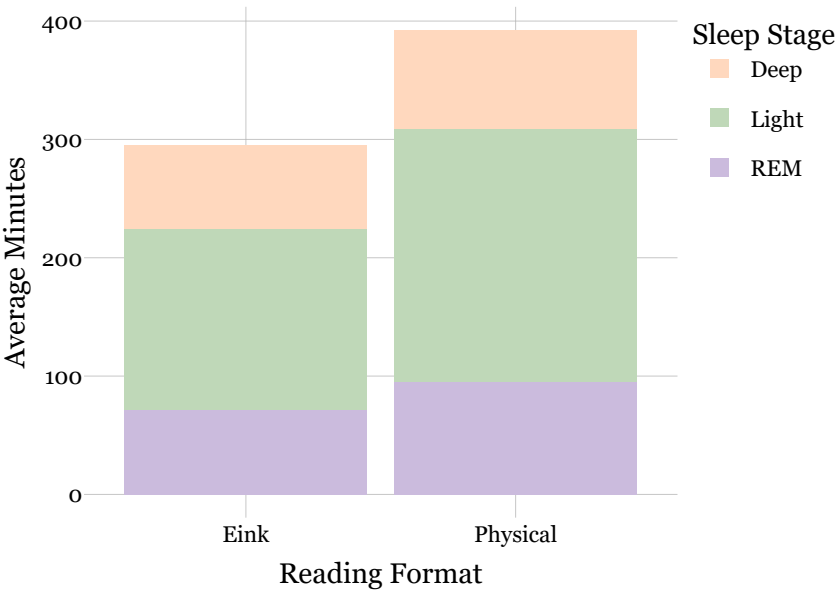
```
SleepStageAvg <- SleepandReadData %>%
  group_by(Format) %>%
  summarise(
    REM = mean(REM),
    Light = mean(Light),
    Deep = mean(Deep)
  ) %>%
  pivot_longer(cols = c(REM, Light, Deep),
               names_to = "Stage",
               values_to = "AvgMinutes")

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") +
  theme(legend.position="right")

ggplotly(StackedBars)
```

file:///C:/Users/kathr/AppData/Local/Temp/RtmpsBabTL/file6b843da96434/widget6b84309d1e73.html screen

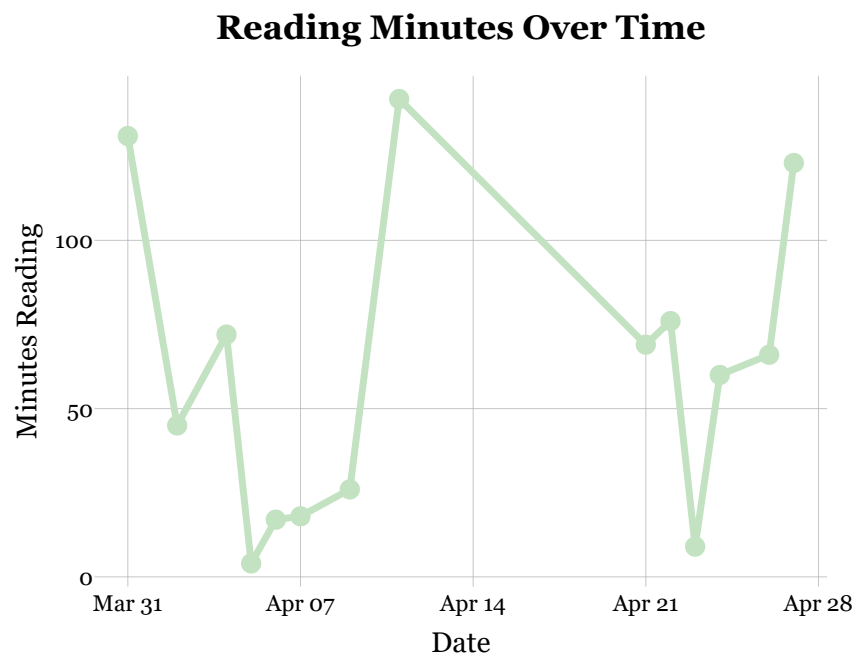
Average Minutes in Sleep Stages by Format



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 Minutes Over Time",  
        x = "Date",  
        y = "Minutes Reading")  
  
ggplotly(Lineplot)
```

file:///C:/Users/kathr/AppData/Local/Temp/RtmpsBabTL/file6b845d3f11f4/widget6b844b1b802.html screens



Bar chart showing average pages read by day of week

```
DayOfWeekSummary <- SleepandReadData %>%
  mutate(DayOfWeek = factor(DayOfWeek,
                             levels = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday")))
  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", "#A3C94A"))
  labs(title = "Average Pages Read by Day of Week",
       y = "Average Pages",
       x = "Day of Week")

ggplotly(DayOfWeekPlot)
```

file:///C:/Users/kathr/AppData/Local/Temp/RtmpsBabTL/file6b8442bb300e/widget6b84e204bf4.html screens

