

hw3

Homework 3 yay

Git repository: https://github.com/cooperevan/ENVS-193DS_homework-03.git

```
library(tidyverse)
library(here)
library(gt)
library(readxl)
library(janitor)
library(ggthemes)
library(PNWColors)
```

Problem 1: personal data

a. data summarization

The data could be summarized by comparing my observations of the response variable, hours asleep, to the type of workout I did on that day such as legs, push, or pull, to better understand if a different type of workout changes my sleep schedule, and can be completed by creating a boxplot. I can also compare sleep to other variables such as calories eaten that day, workout duration, time of workout, and screen time by creating a scatterplot of the data to compare.

b. Visualization

Using the summary you described in part a, **create a visualization of your data comparing your summarized response variable between groups**. If you are calculating a mean or median, show the underlying data in addition to your summary.

```
STATS_workout_data_Sheet1_2 <- read_csv("~/Downloads/STATS workout data - Sheet1-2.csv")
```

Rows: 33 Columns: 8

-- Column specification -----

Delimiter: ","

chr (5): Day of the week, Duration of Workout, Exercise type, Time of Day of...

dbl (3): Asleep, Screen time, Calories eaten

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.

```
workout_data <- STATS_workout_data_Sheet1_2
screen_data <- workout_data |>
  clean_names() |>
  drop_na() |>
  rename(day = day_of_the_week) |>
  select(day, asleep, screen_time)

workout_clean <- workout_data |>
  clean_names() |>
  rename(day = day_of_the_week, work_time = duration_of_workout) |>
  select(day, asleep, exercise_type)

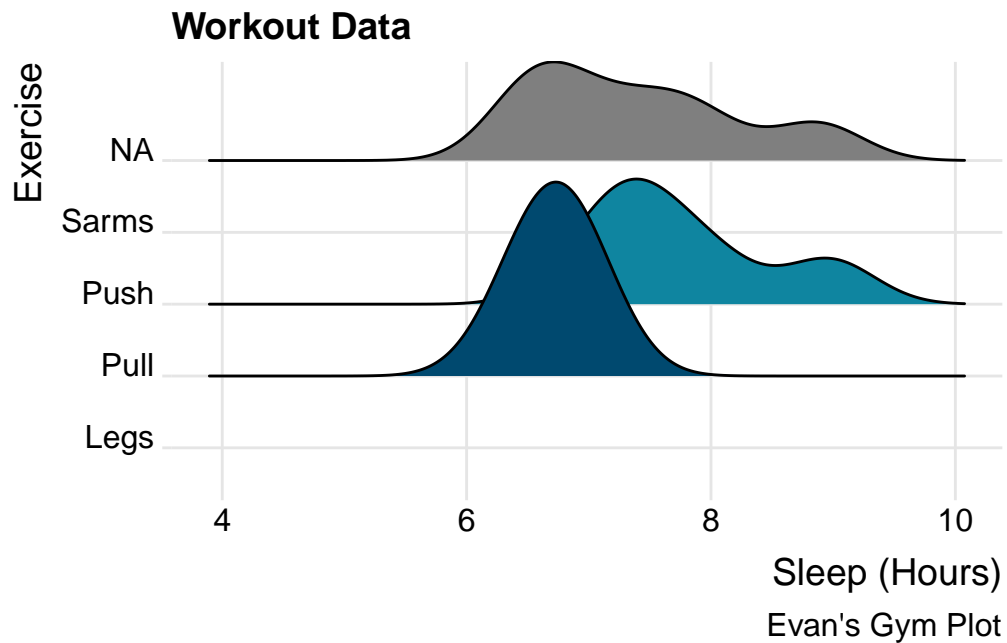
workout_01 <- workout_data |>
  clean_names() |>
  rename(day = day_of_the_week, work_time = duration_of_workout) |>
  select(day, asleep, exercise_type, work_time)

workout_legs <- workout_clean |>
  filter(exercise_type == "Legs")
```

```
#affective visualization
ggplot(workout_clean,
  aes(x = asleep, y = exercise_type, fill = exercise_type)) +
  geom_density_ridges() +
  theme_ridges() +
  theme(legend.position = "none")+
  scale_fill_manual(values = PNWColors::pnw_palette("Bay", 5))+
  labs(x = "Sleep (Hours)",
    y = "Exercise",
    title = "Workout Data",
    caption = "Evan's Gym Plot")
```

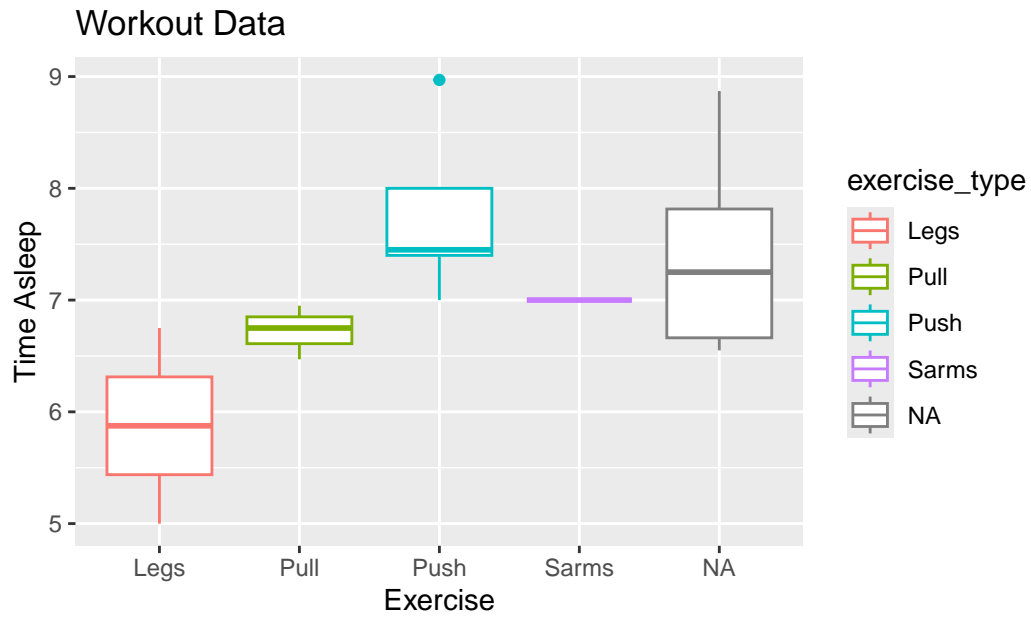
Picking joint bandwidth of 0.369

Warning: Removed 16 rows containing non-finite outside the scale range (``stat_density_ridges()``).



```
ggplot(workout_clean,
  aes(x = exercise_type,
    y = asleep,
    color = exercise_type)) +
  geom_boxplot()+
  labs(x = "Exercise",
    y = "Time Asleep",
    title = "Workout Data",
    caption = "Evan's Gym Plot")
```

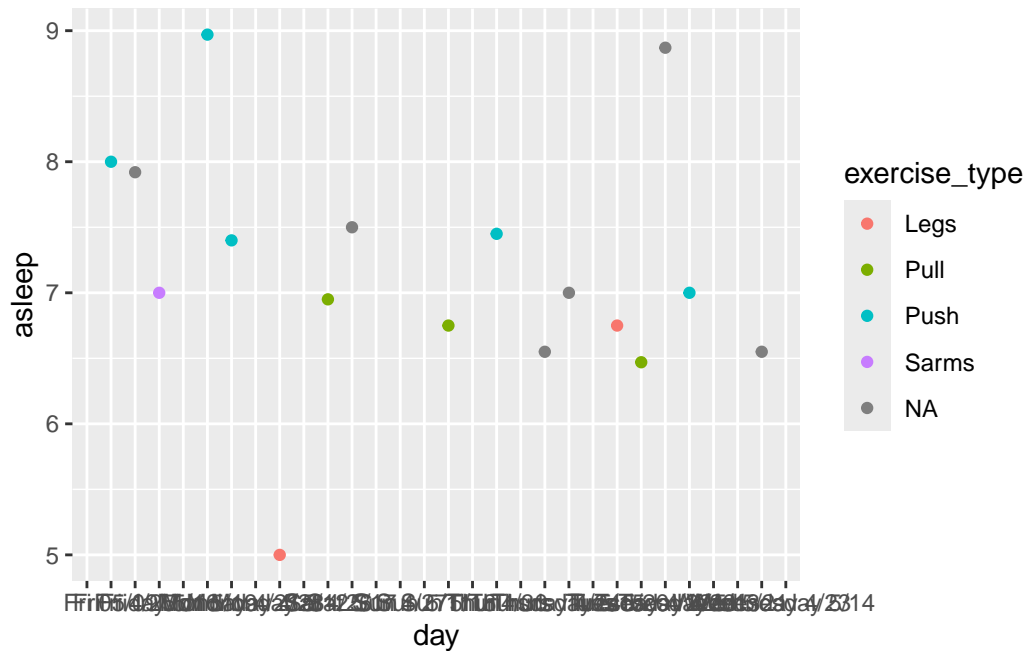
Warning: Removed 16 rows containing non-finite outside the scale range (``stat_boxplot()``).



Evan's Gym Plot

```
ggplot(workout_clean,  
  aes(x = day,  
      y = asleep,  
      color = exercise_type))+  
geom_point()
```

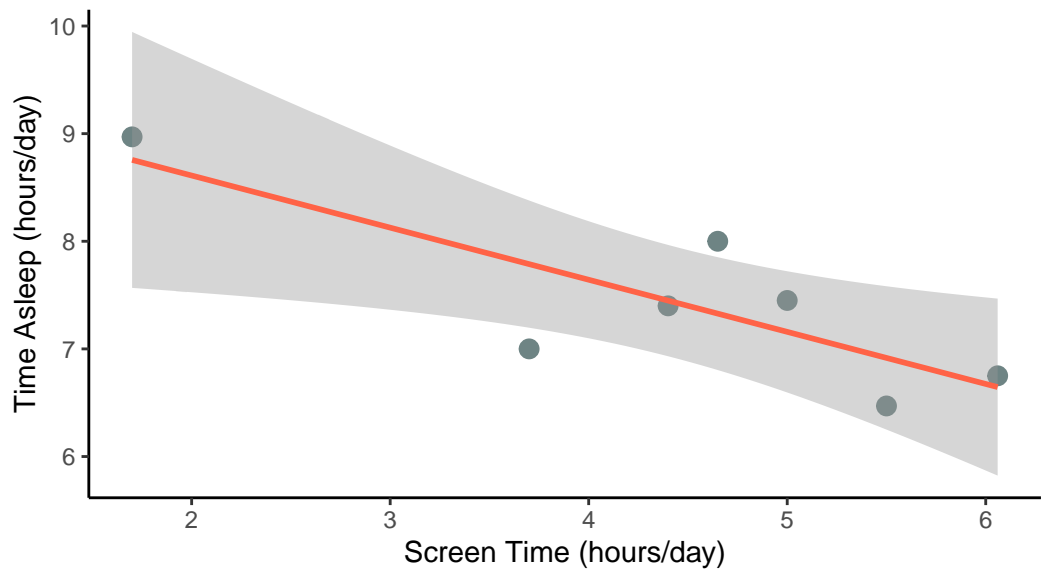
Warning: Removed 16 rows containing missing values or values outside the scale range (`geom_point()`).



```
ggplot(screen_data, aes(x = screen_time,
                        y = asleep)) +
  geom_point(color = "darkslategray",
            size = 3,
            alpha = 0.7) +
  geom_smooth(method = "lm",
            se = TRUE,
            color = "tomato") +
  labs(
    title = "Relationship Between Screen Time and Sleep Time",
    caption = "evans data",
    x = "Screen Time (hours/day)",
    y = "Time Asleep (hours/day)") +
  theme_classic()
```

`geom_smooth()` using formula = 'y ~ x'

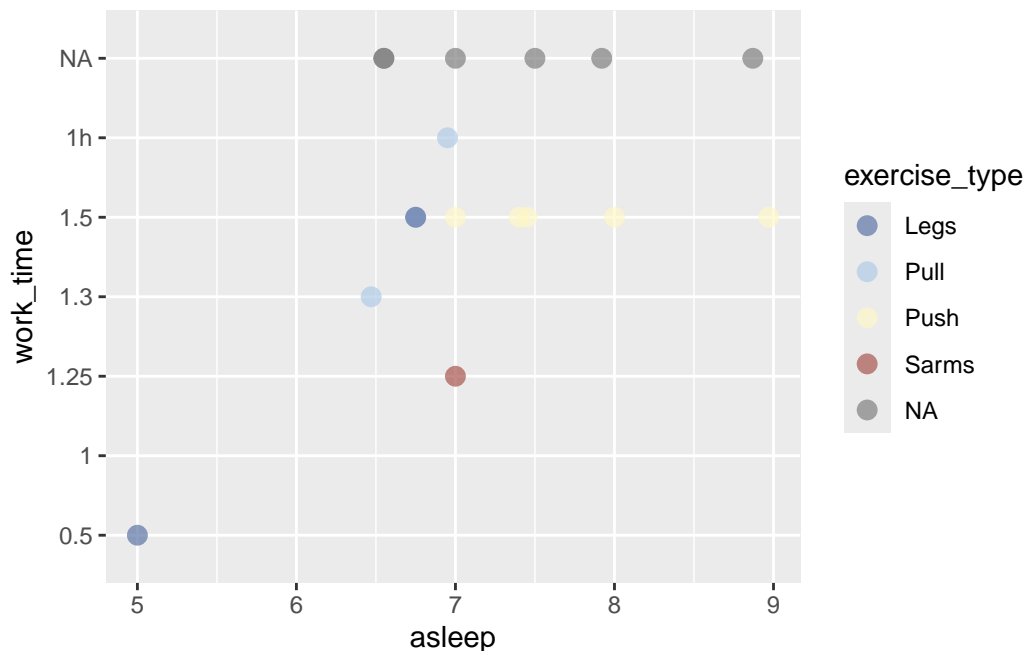
Relationship Between Screen Time and Sleep Time



evans data

```
#plot with x = workout duration, y = sleep, and shape = type of workout
ggplot(workout_01,
       aes(x = asleep, y = work_time, color = exercise_type))+
  geom_point(size = 3,
            alpha = 0.7)+
  theme()+
  scale_color_manual(values = PNWColors::pnw_palette("Shuksan2", 4))
```

Warning: Removed 16 rows containing missing values or values outside the scale range (`geom_point()`).



Use colors in your figure (that are not the default ggplot settings).

Make sure your axis labels are full, readable text (not just your column names).

c. Caption (5 points)

Write a *caption* for your figure.

d. Table presentation (10 points)

Using `gt` or `flextable`, create a table with the same data summary that you describe in part a and visualized in part b. For example, if you described and visualized means, make a table with means. If you need to, round any numbers to one decimal point.

```
workout_data |> gt()
```

Problem 2. Affective visualization (24 points)

In this problem, you will create an **affective visualization** using your personal data in preparation for workshops during weeks 9 and 10.

Day of the week	Duration of Workout	Exercise type	Asleep	Time of Day of Workout	Alcohol?
Monday 4/21	1.5	Push	7.40	Morning	Yes
Tuesday 4/22	1.3	Pull	6.47	Morning	No
Wednesday 4/23	NA	NA	6.55	NA	Yes
Thursday 4/24	NA	NA	6.55	NA	Yes
Fri 4/25	NA	NA	7.92	NA	Yes
Sat 4/26	1h	Pull	6.95	Morning	Yes
Sun 4/27	NA	NA	NA	NA	Yes
Monday 4/28	NA	NA	NA	NA	No
Tuesday 4/29	NA	NA	8.87	NA	Yes
Thur 4/30	1.5	Push	7.45	Morning	Yes
Fri 05/09	1.5	Push	8.00	Afternoon	Yes
Sun 5/11	1.25	Pull	NA	Afternoon	No
Monday 5/12	0.5	Legs	5.00	Afternoon	No
Tuesday 5/13	1.5	Push	7.00	Afternoon	No
Wednesday 5/14	1.5	Pull	NA	Morning	No
Thursday 5/15	NA	NA	7.00	NA	No
Friday 5/16	1.25	Sarms	7.00	Morning	No
Sat 5/17	NA	NA	7.50	NA	No
Sun 5/18	1.5	Pull	6.75	Night	No
Mon 5/19	1.5	Push	8.97	Morning	No
Tues 5/20	1.5	Legs	6.75	Morning	No
Wed 5/21	1	Pull	NA	NA	No
Thur	NA	NA	NA	NA	NA
Fri	NA	NA	NA	NA	NA
Sat	NA	NA	NA	NA	NA
Sun	NA	NA	NA	NA	NA
Mon	NA	NA	NA	NA	NA
Tues	NA	NA	NA	NA	NA
Wed	NA	NA	NA	NA	NA
Thurs	NA	NA	NA	NA	NA
Fri	NA	NA	NA	NA	NA
Sat	NA	NA	NA	NA	NA
Sun	NA	NA	NA	NA	NA

In lecture, we talked about the three vertices of data visualization: 1) exploratory, 2) affective, and 3) communicative. We've done a lot of exploratory and communicative visualization, but have yet to think about affective visualization.

a. Describe in words what an affective visualization could look like for your personal data (3-5 sentences). (2 points)

An affective visualization is a visualization that has provokes emotion in the viewer using creativity and novelty. For my personal data on whether type of workout impacts sleep, a possibility can be to use the website From Data to Viz to find interesting ways to display the data, like using the ggrridge package and creating a ridge plot for the data.

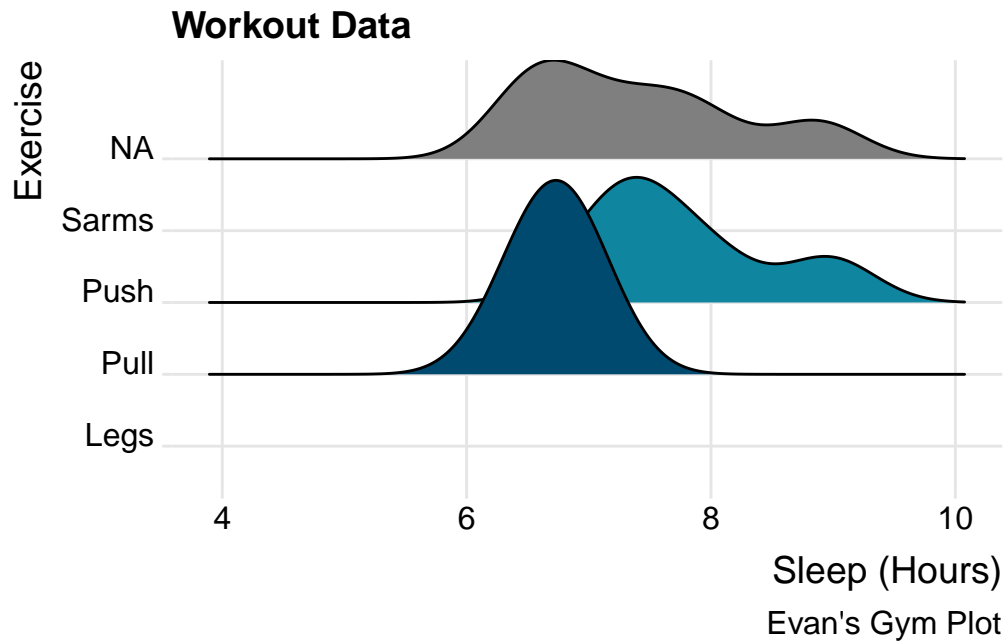
b. Create a sketch (on paper) of your idea.

c. Make a draft of your visualization.

```
ggplot(workout_clean,
       aes(x = asleep, y = exercise_type, fill = exercise_type)) +
  geom_density_ridges() +
  theme_ridges() +
  theme(legend.position = "none")+
  scale_fill_manual(values = PNWColors::pnw_palette("Bay", 5))+
  labs(x = "Sleep (Hours)",
       y = "Exercise",
       title = "Workout Data",
       caption = "Evan's Gym Plot")
```

Picking joint bandwidth of 0.369

Warning: Removed 16 rows containing non-finite outside the scale range (``stat_density_ridges()``).



d. Write an artist statement. (8 points)

An artist statement gives the audience context to understand your work. For each of the following points, write 1-2 sentences to address:

- the content of your piece (what are you showing?)
I'm showing the relationship between my sleep and type of workout
- the influences (what did techniques/artists/etc. did you find influential in creating your work?)
- the form of your work (written code, watercolor, oil painting, etc.)
- your process (how did you create your work?)

Problem 3. Statistical critique

a. Revisit and summarize (6 points)

What are the statistical tests the authors are using to address their main research question? (Note: you have already written about this in homework 2! Find that text and provide it again here!)

Insert the figure or table you described in Homework 2 here.

b. Visual clarity (10 points)

In 1-3 sentences, answer the question that best fits your paper.

If you inserted a figure in Part a: How clearly did the authors *visually* represent their statistics in figures? For example, are the x- and y-axes in a logical position? Do they show summary statistics (means and SE, for example) and/or model predictions, and if so, do they show the underlying data?

If you inserted a table in Part b: How clearly does the table represent the data underlying tests?

If you have neither: Critique another figure or table in the text for visual clarity (and insert a screenshot of that figure/table in your submission).