

Midterm #2: Tidyverse Skills Check

INSERT STUDENT NAME HERE

Learning Objectives:

Students will demonstrate their ability to translate questions into code using **tidyverse** packages and verbs.

- **dplyr:** `filter()`, `mutate()`, `group_by()`, `summarise()`, `count()`, ect
- **ggplot2:** `ggplot()`, `aes()`, `geom_bar`, `geom_col`, `geom_histogram`, `geom_boxplot`

```
library(tidyverse)
```

Step 0: Library tidyverse

Step 1: Load the Data (AllTrails App)

AllTrails is a fitness and travel mobile app used in outdoor recreational activities. AllTrails is commonly used for outdoor activities such as hiking, mountain biking, climbing and snow sports. The service allows users to access a database of trail maps, which includes crowdsourced reviews and images.

Citations:

- <https://www.alltrails.com/>
- <https://en.wikipedia.org/wiki/AllTrails>

Data Source: These data come from Kaggle

- <https://www.kaggle.com/datasets/planejane/national-park-trails>

```
allTrails <- read_csv("https://raw.githubusercontent.com/kitadasmalley/DATA151/main/Data/AllTrails%20da
```

Step 2: Look at the data What variables are available to work with?

```
#str(allTrails)
```

For this assessment you will use the following variables:

- **area_name:** Name of National Park
- **state_name:** Name of State where trail is located
- **length:** Length of hike in meters
- **elevation_gain:** Elevation gain in meters
- **route_type:** Out and back, Loop, or Point to Point
- **avg_rating:** Star rating (out of 5)

Step 3: Calculate the Grade (5 points)

Average grade can be calculated by taking the total elevation gain of the trail, dividing by the total distance, multiplied by 100 to equal a percent grade. The elevation and distance must be in the same units.

Citation:

- <https://venturacountytrails.org/TrailMaps/TrailSymbols.htm#:~:text=Average%20grade%20can%20be%20calculated,as>

TASK: Add a column to the data frame calculate the average grade for each trail.

Please do NOT create a new data frame, you can overwrite the data frame name and add on a new column. You will be using this in subsequent parts.

```
#allTrails<-allTrails%>%  
## INSERT CODE HERE ##
```

Step 4: Distribution of Grade (8 points)

A. Graphic - Histogram (3 points)

TASK: Make a histogram showing the distribution trail grades. Don't forget to title your graph.

```
## INSERT CODE HERE ##
```

B. Insight (5 points)

TASK: Comment on the shape of this histogram

```
## INSERT HERE ##
```

Step 5: Convert Units (5 points)

The AllTrails website reports trail length and elevation gain in miles and feet, respectively.

See example here: <https://www.alltrails.com/trail/us/washington/sentinell-peak-via-grey-wolf-deer-loop>

- 1 meter = 0.000621371 miles
- 1 meter = 3.28084 feet

TASK: Convert the values for length and elevation to miles and feet, respectively.

Please do NOT create a new data frame, you can overwrite the data frame name and add on a new column. You will be using this in subsequent parts.

```
## INSERT CODE HERE ##
```

Step 6: Compare Route Types (4 points)

Create a side-by-side box plot to compare the distributions of elevation gain in feet across the three route types (loop, out and back, point to point). Please fill the boxes with color for each route type.

```
## INSERT CODE HERE ##
```

Step 7: Finding Family Friendly Hikes (10 points)

Our family likes to hike together! We have three children (ages 2, 7, and 11) so we have some limitations.

TASK: Find the best hike that fits ALL of the following conditions:

- Is in Oregon
- Is is a loop
- Is less than 3 miles
- Is "easy" (has a grade less than 5)

Note: Best is defined by the highest average rating

```
## INSERT CODE HERE ##
```

Respond in a full sentence, with which hike is the best family friendly hike in an Oregon National Park:

```
## INSERT ANSWER HERE ##
```

Step 8: Trails within National Parks (within States) (5 points)

TASK: Find the number of trails and the average star rating within each National Park.

- Which national park has the most trails?
- Which national park has the highest rated trails?

Create a new data frame to accomplish this so that we can use this in the following step.

Hint: You can group by two variables. Please do this to keep the State in which each National Park is located.

```
## INSERT CODE HERE ##
```

Step 9: Most National Parks (5 points)

TASK: Find the State with the most National Parks.

Hint: You should use the data frame you created in the previous step.

```
## INSERT CODE HERE ##
```

Step 10: Join Regions (8 points)

States are grouped by regions with the following data frame:

```
stateRegion<-data.frame(state_name=state.name,  
                        region=state.region)
```

```
head(stateRegion)
```

```
##  state_name region  
## 1    Alabama  South  
## 2     Alaska   West  
## 3    Arizona   West  
## 4   Arkansas  South  
## 5 California   West  
## 6   Colorado   West
```

A: Join (2 points)

TASK: Add a column for region to the data frame

- You can do this with the data frame you generated in the previous step (9)
- OR you can use the full complete data frame

Hint: Use `left_join`

```
## INSERT CODE HERE ##
```

B: National Parks by Region (3 points)

Find the number of National Parks in each region. Create a new data frame to accomplish this and use it in the next step.

```
## INSERT CODE HERE ##
```

C: Bar Graph (3 points)

Create a bar graph for the number of National Parks per region.

```
## INSERT CODE HERE ##
```

EXTRA CREDIT: MAKE A MAP! (10 points)

Make a map of the average trail star rating by state