

# Lab-08

Write your name here

3/17/23

## Preface

The goal of this assignment is to help you gain more familiarity with using **ggplot** to visualize distributions. In this lab we provide less scaffolding and more open-ended questions. As always, please come to office hours and reach out to your teaching staff if you have any questions.

## Data

We will work with data on NYC Yellow cab trip data in January 2021 from [TCL](#). We have seen these data in our Prelim 1. Start by importing these data and assigning them to a name.

```
trips <- read_csv("taxi-trips.csv")
```

1. We'll start by making some simple visualizations of the distribution of the trip distance. First, make a histogram. Make the first bar start right at zero, customize the bins or binwidth to suit your tastes, and use a named color to delineate between bars.

**2. Make a density plot of the distribution of the trip distance.**

3. Use the data frame `speeds` to plot the density of trip speeds, faceting by the hour during which the trip began. Use `geom_density`'s `fill` argument to fill the densities with a named color or hex code of your choice. Arrange the facets so they are in 6 rows and 4 columns. What, if any, conclusions can you draw from the resulting plot?

```
speeds <- trips |>
  mutate(duration = dropoff_datetime - pickup_datetime,
         hour = hour(pickup_datetime)) |>
  filter(duration > 0 & trip_distance > 0) |>
  select(pickup_datetime, dropoff_datetime, hour, trip_distance, duration) |>
  mutate(speed = trip_distance/(as.numeric(duration)/60/60)) |>
  filter(speed < 55) # filter out 55+ miles per hour
```

4. Make an overlapping density plot of the distribution of speeds. Fill by hour, treating hour as a categorical variable by encoding it as a factor (see `?as_factor()`). Adjust the transparency to make it more readable.

5. Reproduce the plot above as a ridgeline plot instead of an overlapping density plot. Fill all the densities with a single named color of your choice, and use `color` to make the density lines themselves white. What, if any, conclusions can you draw from the resulting plot?

**6. Comparing your plots from questions 3, 4, and 5, which one do you think is the most effective visualization? Why?**