

Project Presentation, STAT 651

Due: Sunday, December 1 by 5:00 PM

Instructions:

- For this project, select a data set of interest and apply methods learned in class to create an engaging data visualization. Use packages discussed in lecture, such as `ggplot2`, `plotly`, `lubridate`, and `leaflet`. Visualizations that are interactive, or involve date-time or geospatial data are encouraged.
- You may work in a group of 2-3 students, or individually.
- Submit your presentation as a video recording to a shared folder on Google Drive by **5:00 PM on Sunday, December 1**. The recording can be made using Zoom, or similar software, and should be about 5 minutes long.
- For your presentation, prepare 3-5 slides, including a title slide, a data description slide, and slides displaying your data visualization(s). For interactive plots, you may use a screenshot for the slides, and then refer to the rendered HTML file during the recording.
- Additionally, by **Sunday, December 1** each group should submit the following two files to Canvas: (1) presentation slides in PDF format, and (2) Quarto document with R code rendered to HTML or PDF format.
- During the last week of class I will host Zoom sessions during the scheduled time (M/W 8-9:30) to play the recordings submitted by each group. You are required to attend on Zoom to support your classmates.

Grading: Specific expectations are provided below.

- The title slide includes a title for your project and the names of all group members.
- The source of the data set is provided, and relevant variables are listed and described.
- The selected data visualization(s) illustrates important aspects of the data set.
- The data visualization is well-formatted, with appropriate labels and scales.
- Each group member contributes to the project.
- Your presentation is not exceedingly long (under 10 minutes, please).
- R code is provided in a Quarto document.

Projects that meet these expectations will receive an A. Projects with minor flaws, that mostly address the above expectations, will receive an A-. Projects that fail to meet several expectations in significant ways will receive a B or C. Projects that are incomplete, plagiarized, or demonstrate little interest or effort will not receive a passing grade.

Data Sources:

Here are some potential sources for data sets. You do not need to limit yourself to these. However, **do not reuse a data set that has already been used in lecture or homework.**

- Tidy Tuesdays: <https://github.com/rfordatascience/tidytuesday>
- Kaggle: <https://www.kaggle.com/datasets>
- FiveThirtyEight: <https://data.fivethirtyeight.com/>
R package: `library(fivethirtyeight)`
- OpenIntro: <https://www.openintro.org/data/>
R package: `library(openintro)`
- UCI Machine Learning Repository:
<https://archive.ics.uci.edu/>
- DataSF: <https://datasf.org/opendata/>
- Awesome Public Datasets:
<https://github.com/awesomedata/awesome-public-datasets>
- Google data set search: <https://datasetsearch.research.google.com/>

To get a list of the data sets in an R package run the command `data(package = "name")`. For example, run the following command to get a list of data sets in the `fivethirtyeight` package:

```
data(package = "fivethirtyeight")
```