

Data 227 - Autumn 2023 HW 5
Data Visualization and Communication - Trimble
Due: Saturday November 4, 2023 11:59pm

1. Stacked bar charts + color

- (a) A new file, names.csv, contains the "gender" of the hurricane name (by convention, the World Meteorological Society determines a list of storm names, alternating names traditionally used for males and names traditionally used for females every calendar year—see Tropical Cyclone Naming History and Retired Names). Combine storms.csv with the new file so that you have a dataframe with the name, "gender", and year of each storm. You should have about 500 distinct storms in the new dataframe (remember, some names were used in more than one year). (1 pt)
- (b) With your new dataframe, create a stacked bar chart showing the number of storms in each year. The "stacked" categories should be the gender of the storm name. Use an appropriate color scheme for the gender categories based on the type of data you are displaying, and explain why you selected the scheme. Include a caption discussing your graph. (2 pts)
- (c) Create a scatterplot showing the tracks of each storm: longitude on the x-axis, latitude on the y-axis for the duration of each storm. Use an appropriate color scheme for the average windspeed based on the type of data you are displaying, and explain why you selected the scheme. Include a caption discussing your graph. (2 pts)
- (d) Using the graph from Part C, highlight the hurricane you tracked in Homework 2 with a different color. Use an appropriate color scheme to highlight your hurricane, and explain why you selected that scheme. (1 pt)
- (e) At some point in your career as a data scientist, you may want to create and use a custom color palette. In your data visualization module/package of choice (altair, ggplot2, matplotlib, plotnine, or seaborn), create a custom six color palette based on something you enjoy—e.g., a movie, musical album, sports team, TV show, etc. For example, consider this palette based on colors worn in the 2021 Presidential Inauguration. You may want to use a color picker such as <https://imagecolorpicker.com/en>. Apply the new color scheme to your graph in Part C. (2 pts) (Many of these custom color schemes will look awful, it's true.)

2. New type of visualization

Using a dataset of your choosing, create a distinct type of data visualization that you have not made before. This means you may not make a bar chart, histogram, boxplots, or scatterplot—consider instead types such as a directed network graph, choropleth map, marimeko graph, Sankey charts, "wheat" histogram plots, or others. You might look at data you have used in the past and try to present a different view of it.

Include a figure caption and just one paragraph discussing your findings and the graphical design. The figure caption should describe the origin of the dataset.

Separately attach any code you used to produce the visualization.

You do not have to use any specific tools to produce the visualization but you need to find something interesting and display it effectively.