

Data 227 - Autumn 2023 HW 2
Data Visualization and Communication - Trimble
Due: Friday October 13, 2023 11:59pm

Homework 1, Hurricane tracks

1. The `storms.csv` file contains a subset of the NOAA Atlantic hurricane database best track data. The data includes the positions and attributes of named North Atlantic tropical storms from 1975-2020, measured every six hours during the lifetime of a storm.
 - (a) Check the storage type of each variable and state whether it matches your expectations from reading the dataset documentation (i.e., are categorical variables saved as strings? are numerical variables saved as integers or floating points?). (1 pt)
 - (b) Let's take a closer look at a single storm. Choose a storm that has no missing values, and save a subset of the data belonging to that storm in a new dataframe. (Tropical storms live for a few weeks. Don't treat storms with the same name in different years together.) (2 pts)
 - (c) Create a scatterplot comparing your storm's maximum sustained windspeed (in knots) to its air pressure at the storm's center (in millibars) for all six-hour measurements. Use different visual channels to also include information about the storm's status and Saffir-Simpson storm category. How many attributes are displayed in your graph, and what channels are you using to encode them? (note, hurricane/tropical storm status and wind speed, though they may be spread out into multiple columns, measure the same thing.) (2 pts)
 - (d) Describe the relationship between wind and pressure. Do the storms' status and category seem to have any impact on this relationship? You may use external sources to answer second half of this question provided that you cite them. (2 pts)
 - (e) Don't forget that each storm was tracked over time and space. Create a scatterplot with latitude on the x-axis and longitude on the y-axis to recreate the storm's path. In addition, encode information from `tropicalstorm_force_diameter` and `hurricane_force_diameter`—pay special attention to what Tufte has to say in this week's reading. For an example, view my plot below. (3 pts)
2. NOAA, the US federal government agency responsible for weather forecasting, maintains bouys in Lake Michigan with sensors that measure illumination, air temperature, water temperature, dewpoint, and other meteorological variables.

Station 45198, for instance, is near downtown Chicago and has data starting in 2021 to present (though some fields are missing for some dates).

https://www.ndbc.noaa.gov/station_history.php?station=45198

Examine the data, find some fact about it, and design a data-rich visualization that communicates this fact.

If you find the three years of history at this location limiting, you can obtain and plot data from other weather stations.

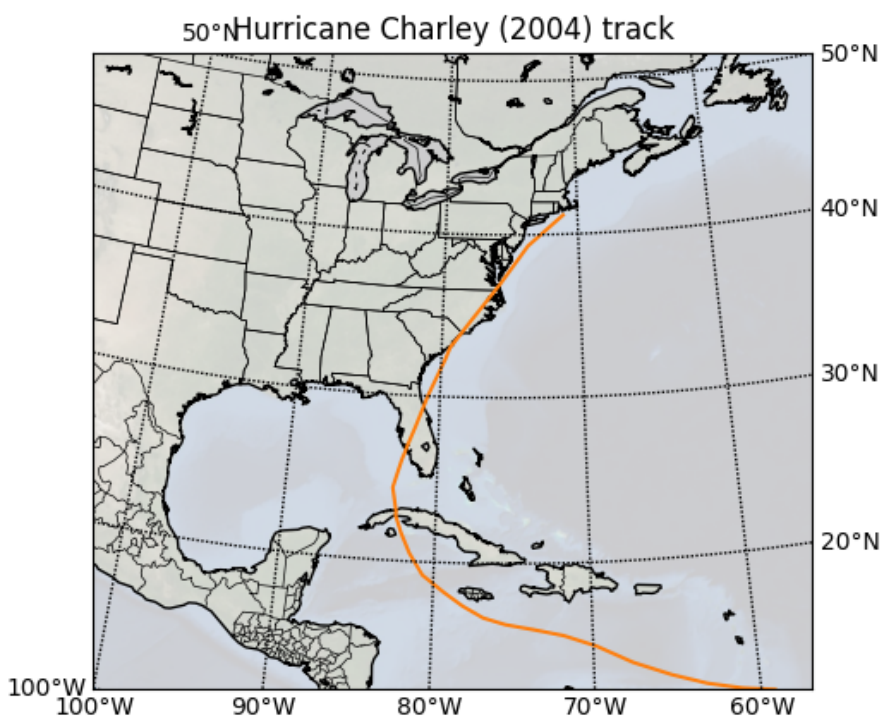


Figure 1: Sample code to produce this map is provided