

Please complete each problem and submit a PDF with your solutions. If you use code to solve the problems, you should include it in your submission.

Problem 1 (10 points)

- A. The `storms.csv` file contains a subset of the NOAA Atlantic hurricane database best track data. The data includes the positions and attributes of storms from 1975-2020, measured every six hours during the lifetime of a storm. 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)

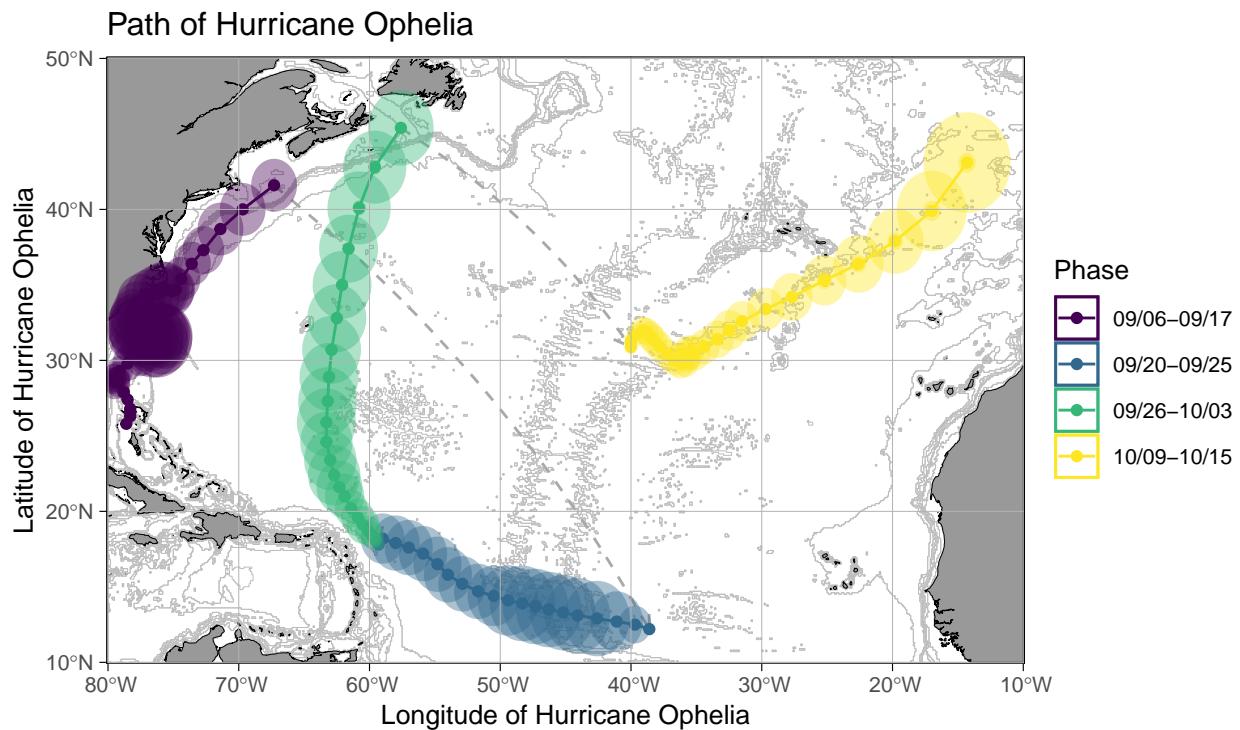
R Code

```
storms <- read.csv("https://vincentarelbundock.github.io/Rdatasets/csv/dplyr/storms.csv")
```

Python Code

```
import pandas as pd
storms = pd.read_csv("https://vincentarelbundock.github.io/Rdatasets/csv/dplyr/storms.csv")
```

- B. There are quite a few datapoints in this dataset—to be exact, 11,859. 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. (2 pts)
- C. Create a scatterplot comparing your storm's maximum sustained wind speed (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? (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)



Problem 2 (10 points)

For this assignment, you will create a static visualization of the City of Chicago Salary database for a recent year. You can download a recent database from the Chicago City Data Portal: <https://data.cityofchicago.org/Administration-Finance/Current-Employee-Names-Salaries-and-Position-Title/xzfq-xp2w>.

It should have approximately 30,000 rows and looks like this:

R

##	Name	Job.Titles	Department		
## 1	AARON, JEFFERY M	SERGEANT	POLICE		
## 2	AARON, KARINA	POLICE OFFICER (ASSIGNED AS DETECTIVE)	POLICE		
## 3	ABAD JR, VICENTE M	CIVIL ENGINEER IV	WATER MGMNT		
## 4	ABARCA, EMMANUEL	CONCRETE LABORER	TRANSPORTN		
## 5	ABARCA, FRANCES J	POLICE OFFICER	POLICE		
## 6	ABBATACOLA, ROBERT J	ELECTRICAL MECHANIC	AVIATION		
##	Full.or.Part.Time	Salary.or.Hourly	Typical.Hours	Annual.Salary	Hourly.Rate
## 1	F	Salary	NA	122568	NA
## 2	F	Salary	NA	110796	NA
## 3	F	Salary	NA	122112	NA
## 4	F	Hourly	40	NA	45.9

## 5	F	Salary	NA	86730	NA
## 6	F	Hourly	40	NA	51.0

Python

```
##           Name ... Hourly.Rate
## 0  AARON,  JEFFERY M ...      NaN
## 1      AARON,  KARINA ...      NaN
## 2 ABAD JR,  VICENTE M ...      NaN
## 3  ABARCA,  EMMANUEL ...    45.9
## 4  ABARCA,  FRANCES J ...      NaN
##
## [5 rows x 8 columns]
```

Examine the data, find a fact that is contained within the data, and design a visualization that communicates that fact. Include not more than one paragraph to go with the visualization, like a figure caption, and include one to three paragraphs explaining your question and your graphical design, and any code you used to produce the visualization. The figure caption should describe the origin of the dataset.

You do not have to use any specific tools to produce the visualization (you could even draw it by hand), but you need to find something interesting and display it effectively.