Data: Weather, Citibike, MTA Turnstiles for subways, (Uber/Lyft/Taxi)

## Story/Hypothesis

* We want to examine the relationship between weather and different modes of public transportation
  + More specifically…
  + Good weather: longer Citi Bike rides, more Citi Bike users, less taxi and subway usage
  + Bad weather: inverse of ^

## Data Cleaning & Aggregation

* Weather: datetime, temp, precip, snow, windgust, cloudcover, visibility, conditions, icon
* Citi Bike: starttime, stoptime, start station coordinates, end station coordinates
* Subway: trip duration, station, date, time, entries, exits
* Yellow & Green Taxi:
  + <https://data.cityofnewyork.us/Transportation/2021-Yellow-Taxi-Trip-Data/m6nq-qud6>
  + <https://data.cityofnewyork.us/Transportation/2021-Green-Taxi-Trip-Data/djnb-wcxt>
  + <https://data.cityofnewyork.us/Transportation/2020-Green-Taxi-Trip-Data/pkmi-4kfn>
  + <https://data.cityofnewyork.us/Transportation/2020-Yellow-Taxi-Trip-Data/kxp8-n2sj>

1. Download weather data for past 10 years
   1. \* Look into Uber/Lyft/Taxi data
2. Clean up data:
   1. Weather: datetime match format with other data sets
   2. Subway: add coordinates, aggregate for each day
   3. CitiBike: aggregate daily rides for each start and end station
3. Combine data frames and look for patterns
   1. Group by weather condition
   2. Group by temperature
   3. Group by precipitation/snow %
4. Create visuals
5. Prediction

## Visualization

* Heat map over time of turnstiles being used in locations (consider what time interval we should use)
* Network graph of Citibike data

## Prediction

* How many Citibikes are used/taxis hailed/subway swipes depending on the weather

## Presentation

* Shots on the subway and riding a citibike in different weather conditions (can add effects for sunshine, rain, snow, etc)
* Graphic animation of a route we are taking from one location to another