Project Ideas

Covid-19 – Nytimes US Covid

* Covid State Count Dataset
* Covid County Dataset

Explanatory Data Analysis - Micah

* Plot of cumulative cases by state
* Visualziing spatial distribution of cases through time

Analysis

* Auto Correlation between Cases and Deaths
* Linear Correlation between States Cases/Deaths and Demographics
  + Population
  + Density
* Looking at case rate correlation with key holidays/events
* Web Scraping of Twitter

**Project Outline**

**Name –** Exploring Temporal Trends in NYTimes US State and County Covid 19 Data

**Introduction**

* **C**ovid 19 cases was a novel respitory virus that quickly spread across US in 2020
* Cases exponentially grew in different states at different times
  + Early spring peak for NY
  + Summer Rise for South (Florida, Texas, Arizona)
  + Fall Rise for West (Ca)
* New Death rate lagged behind cases by a few weeks
  + Incubation time for virus
  + Takes time for people to get sick, go to hospital, die
* ***Can we develop a model to predict future deaths based on current case numbers?***

**Methodology**

* **Exploratory Data Analysis - Micah**
  + Visualizing timeseries for average daily cases and deaths by state
  + Visuals and statistics for monthly trends in cases by states
    - Box plot of average daily cases per month by state
    - Box plot of average daily deaths per month
  + Animated Map of cases across states
    - Shows when different regions were at their peak

**Algorithm**

* Developing Lag-1 Autogressive model to predict daily deaths as a function of cases
  + What is the lag between the two? (Time Lagged Cross Correlation (TLCC) & Windowed TLCC)
  + Is it consistent between states? And across time?
  + Hierarchical Clustering[[1]](#footnote-0) (might need county level, but for demo just do in state data)
  + Causality inference of vaccination? (get vaccination data?) Probably too early but can still do.
  + SEIR modeling for estimate and projection?[[2]](#footnote-1)
* ~~Webscraping Twitter (probably not necessary )~~
  + ~~Correlation between collective attention (number of hashtags) on COVID and the cases in each state?~~

**Conclusions**

* Does death forecast model agree with what we would assume based on what we know about the disease?
* What are the implications for public policy?

**Discussion**

**Appendix –**

* Figures, Tables, Codes

1. https://towardsdatascience.com/how-to-apply-hierarchical-clustering-to-time-series-a5fe2a7d8447 [↑](#footnote-ref-0)
2. https://covid19-projections.com/about/#about-the-model [↑](#footnote-ref-1)