Data Cleaning

1. Summary metrics of provided data (by table)

2. Initial selection of subdata, initial cleaning

3. Merge, pivot, combinations

4. A few simple EDA visuals

Data Visualization

1. Heatmap (like SF food inspection, looking at metrics by county?) (A)

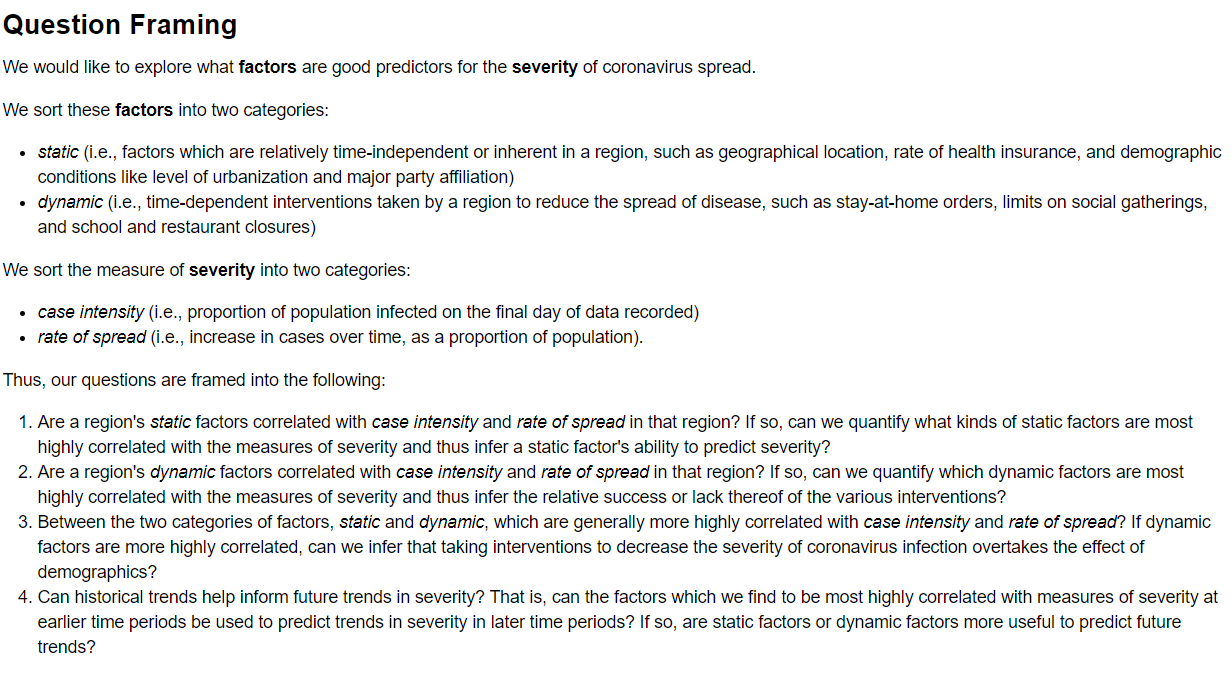
2. Comparisons of demographics (geographic location, urban/rural, state govn't party?) (A)

3. Time series analysis (X)

4. Output variables (all)

5. Intervention vs. rates (M)

Current Question Framing



Old Questions

1. Are there differences in the spread based on region? (areas of country, urban/rural, republican/democrat governor, rate of health insurance)

2. What interventions are most strongly correlated with changing how fast coronavirus spreads in a region? Can we quantify the relative success or lack thereof of the various interventions?

3. What "output" dominates / is most important? (total cases, active cases, deaths, hospitalization...?)

4. Can the above factors be used to predict future trends in data (test positives, deaths, mortality rate, etc.)? What factors dominate - demographics (static), interventions (active)?

5. What kind of model and features is best for modeling the number of coronavirus cases in a region? Can historical trends inform our predictions on the number of coronavirus cases in a given day?

X. Can we quantify how much more or less actionable items (e.g. testing, stay-at-home orders, and business/social gathering regulations), in comparison to non-actionable items (e.g. population, proportion of pre-existing health conditions) affect spread rate, death rate, and recovery rate of coronavirus in a region? Can we quantify the relative success or lack thereof of the various interventions?

X. What kind of model and features is best for modeling the number of coronavirus cases in a region? Can historical trends inform our predictions on the number of coronavirus cases in a given day?

Make sure to answer these prompts in the report:

1. What were two or three of the most interesting features you came across for your particular question?

2. Describe one feature you thought would be useful, but turned out to be ineffective.

3. What challenges did you find with your data? Where did you get stuck?

4. What are some limitations of the analysis that you did? What assumptions did you make that could prove to be incorrect?

5. What ethical dilemmas did you face with this data?

6. What additional data, if available, would strengthen your analysis, or allow you test some other hypothesises?

7. What ethical concerns might you encounter in studying this problem? How might you address those concerns?