Part 2 - An Extension Plan

The Course Project

The course project is broken into four parts:

- Part 1 - Common Analysis sets the stage for the subsequent assignments. In A4 you conduct a base analysis. All of the students in the class will conduct the same analysis, but with a slightly different data subset.
- Part 2 - Extension Plan will require you to ask a human centered data science question that extends the work in A4: Common Analysis.
- Part 3 - Presentation will require you to give a modified (shorter) PechaKucha presentation of your completed project.
- Part 4 - Project Repository, creation of a fully documented repository and also requires the submission of a written project report.

Extension Plan

In this assignment you are to plan an extension to the analysis that you performed in Course Project, Part 1. There are many potential ways to build upon or extend what you learned in Part 1. The deliverable for this assignment is a study plan for your proposed work. Your plan will cover a variety of details, including what data you will use, what you will do with the data (e.g., a statistical analysis, train a model), what results you expect, and most importantly, why your project is interesting or important).

Your study plan should cover:

1. **Motivation/problem statement:** Why are you planning to do this analysis? Why is it potentially interesting and useful, from a scientific or practical perspective? What do you hope to learn? Your motivation statement should explain how/why the problem is human centered. That most of the COVID-19 data reflect people as data points does not make a problem immediately human centered.
2. **Research questions and/or hypotheses:** These describe what you hope to discover or determine. Keep in mind this should be related to the Common Analysis that you conducted in Part 1. There are lots of possible questions related to the pandemic, and the Part 1 data set has day-by-day data for your county. For example you might ask:
   - Example research question: How was the local economy influenced by the pandemic?
   - Example hypothesis: Every 10% change in confirmed COVID cases resulted in a 5% decrease in employment for service sector workers in <CITY> within the county.

   The county you were assigned probably has at least one very large metropolitan area. You could use that large metro area as a potential comparator for the whole county.

Alternatively, you could consider potential differences among cities in your county. You could also try to consider whether there were urban/rural differences if you can find geographically structured data. The possibilities are wide-open because the pandemic has been highly datafied.

3. **Data to be used**: What additional data do you plan to use, and why? Summarize what is represented in the dataset, **provide a link to the dataset**, and specify license/terms of use. Highlight any possible ethical considerations to using this dataset. Briefly explain how this data will allow you to answer your question and how it expands on the Common Analysis.
4. **Unknowns and dependencies**: Are there any factors outside of your control that might impact your ability to address the supplementary questions? The purpose of this section is to get you thinking, in a practical sense, about your ability to answer the supplementary research question in the time allotted.
5. **Methodology**: Describe how you plan to investigate this phenomenon. Don't just describe what your analytical methods are (e.g. "ordinary least squares", "student's t-test", "heatmap visualization", or "recurrent neural network"), it's critical to justify why these are appropriate methods for gathering and analyzing your data, or presenting your findings. You are expected to be thorough here: please describe to the best of your ability the entire series of gathering, analysis, and presentation methods you plan to use.
6. **Timeline to completion**: You should list each of the milestones and significant tasks you will need to complete the extended analysis. You might need to: collect data, build a model, test or analyze the model, visualize the results, and document your process. Think of the tasks that you need to complete. You should then provide a timeline that would allow you to complete the analysis on deadline. Keep in mind that you have a presentation (Course Project, Part 3) and final report (Course Project, Part 4) to complete as well. You'll need some time for those activities.

Requirements

Your study plan should include the items listed above as a write up. Our rough estimate is that you should be able to **cover all of these items in 4-5 pages**. Make sure you cite/reference any statistical method or visualization technique that you propose to use in your analysis.