STAT401, Spring 2021: **Group R-Project**

March 13, 2021

This document will outline the guidelines for the Group R-Project. Please read carefully and reach out to me if you have any questions.

1 Project Outline and Datasets

You can work with different datasets that can be found in Kaggle, for example you can analyse datasets for the 2019 Presidential Elections.
The suggested dataset is the COVID-19 dataset curated by John Hopkins University:JHU COVID-19 dataset.

The goal of this project would be to get experience in handling a current real-world dataset and using some of the inferential techniques learnt in this class to derive observations and conclusions.

The scope of this project is broad, and here are some general guidelines:

1. Understand the dataset/datasets. Note that there are different datasets in the repository. You would want to understand the different variables that are involved, identify which ones are categorical/quantitative, and how the variables interact with each other.

2. You want to come with different questions/hypothesis that you want to ask/test. Please keep all your methods within the scope of this course.
   Note that this is often the most difficult step, you want to think about your questions very carefully. It can be a good idea to have multiple questions that are considered.

3. Some of you who took Stat400 with me last Fall must already have experience handling this dataset. Perphaps you can bring that experience to your group.

4. Use the data to derive answers and/or insights into the questions that you asked. We have multiple inferential tools at our disposal.

2 Project Groups

I have already assigned you into groups of 6, based on your section. Each group must meet with me at least two times before the end of the semester, appropriate logistics and meeting times must be planned in advance.
Each group will have an assigned group leader. If you are interested in being a group leader please reach out me.

3   Resources

Please note that you have multiple resources available.

1. I will be available to answer questions, if you need any help, consider coming to office hours. I am happy to talk about project ideas before or after class.

2. This dataset is very well studied, feel free to checkout any examples online. Please keep in mind that copy-pasting ideas is plagiarism. It is important (to the extent possible) that you provide relevant citations/credits for things you use from the internet.

3. You are allowed to modify, complement, and supplement the data as needed. Let your imagination guide you!

4   Grading

This project will comprise of 12% of the total course grade and will be graded out of a total of 240 points.
There will be two required submissions of the project, grading criteria for it will be as follows:

1. **Preliminary Submission:**
   You will need to submit a basic outline of you project idea. This submission will be graded out of 80 points.

2. **Final Submission:** This will be graded out of 160 points.
   You will need to submit three items as part of the final submission:

   (a) **Project Notebook**: Both the .Rmd and .html files. I will provide details about this as the final due date approaches.

   (b) **Project Participation and Contribution document:** This will be a .doc file containing the contributions of each of the group members.

   (c) **Project Presentation:** Each group will have to record a video presentation at most 90secs long: either submit the video as part of Canvas submission or provide a link to your uploaded video on Youtube/Vimeo (latter is preferable). I will provide more details about this later in the semester.
   We will have presentations on the last day of classes, where each presenting group will get 90secs to make the presentation. We might not be able to have all groups to present, so will have a sign-up sheet, and depending on how many groups sign-up, we might (in the spirit of the course) use a random process to get a sample!

   (d) **Group Participation:** Each group has a group page on Canvas. It is recommended that you use that to participate. If you prefer, it is possible to create groups in Piazza too, please talk to your TA about it. Depending on your participation and feedback, I might request to meet with your group to get a sense of how your are progressing with the project.

   I will provide more details about grading later in the course. **You only need to make one submission per group.**

5   Important Notes

The goal of this project is to **study a real-world dataset in a collaborative setting within the scope and abilities explored in this class.** This is **NOT** a competition between the different groups.
To this end please understand the following:

1. It is important every group member contributes to the project. The group leader must inform me if there are problems organizing the group.

2. The ideas and questions you seek to answer do not have to be complicated. Please make sure you are sticking to the scope of this class.

3. Even though there is a component of using R, this is **NOT** just a coding project. Please leverage the diverse individual skills of all your group members in this collaborative setup.

4. Everyone in the group should feel comfortable voicing their ideas. Please make sure that you create an inclusive and non-threatening environment when having group discussions.

5. This is not a competition, so don't spend too much time judging your peers ideas, there is room for many questions and different answers. This is not to say that you cannot have disagreements, just make sure you keep them civil.

6. Any form of misbehaviour will not be tolerated. Please inform me right away if there are any issues with members of your group.