

STAT400, Fall2020: **Final R-Project**

November 8, 2020

This document will outline the guidelines for the final project Please read carefully and reach out to me if you have any questions.

1 Project Outline

As part of this project, you will analyse the COVID-19 dataset that was introduced in class. The dataset can be downloaded from the following link: [JHU COVID-19 dataset](#). I have also shared it in the project folder in Canvas.

The goal of this project would be to get experience in handling a current real-world dataset.

You want to consider the following 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 find appropriate visualizations for the data. This is typically called Exploratory Data Analysis (EDA).
3. You want to come with different questions/hypothesis that you want to ask/test. Note that the questions and methods that you use must be within the scope of what we have studied in this class.
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.
4. Use the data to derive answers and/or insights into the questions that you asked. Make sure to work within the scope of the tools learnt in this class.

2 Project Groups

Here are some guidelines for the groups:

1. There are total of 19 groups and each group has about 15 students.
2. I have assigned a group leader for each group. Please reach out to me if you are the assigned leader, and you don't want to take that responsibility.

3. The role of the leader is to primarily be a point of contact for your mentors (more below) and me. You will also have to keep track of certain spreadsheets involving interactions and ideas from your group members.
4. Each group has its own page in Canvas, it is important that you engage in discussions and announcements there. Your active involvement in the project will be considered for the final grade.

3 Resources

Please note that you have multiple resources available.

1. I have spoken with the Teaching Assistants and they have agreed to be mentors for the project. Following are the assignments of your mentors:
 - (a) Groups 1-6 : Jialin Li
 - (b) Groups 7-12: Maia Karpovich
 - (c) Groups 13-19: Al Fahad.
2. You can reach out to your mentors with specific questions, cross check your ideas, and get feedback on your methods. Please make sure you take advantage of this resource.
3. I will be available to answer questions, if you need any help, consider coming to office hours. You can also email me your questions.
4. I will try to be available 10 mins before lecture, and you can always come by then and bounce ideas and questions.
5. 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.

4 Grading

This project will comprise of 12% of the total course grade and is part of the Final Exam component.

The project will be graded on the following criteria:

1. the involvement of individual group members on the canvas group page. Making announcement, having discussions, uploading files, responding to each other with feedback etc.
2. you will have a shareable spreadsheet where the group members will have to update their personal ideas/suggestions and contributions.
3. there will be a 1.5min presentation in front of the entire class, during lecture (tentatively) on December 10th, 2020.
4. it is important that you adhere to the project guidelines describe in the sections: Project Outline and Important Notes.
5. the group will have to upload a single .html report on Canvas.

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. To this end please understand the following:

1. 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.
2. 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.
3. 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.
4. 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.
5. Any form of misbehaviour will not be tolerated. Please inform me right away if there are any issues with members of your group.