

Hi class,

This is a guideline for your final presentation. Each group will have 15 minutes, including the follow-up question session. You might break down the presentation in the following ways:

1. Introduction: You need to introduce your data. It does not mean to show your spreadsheets, but you need to address the original source of data and how you have access. Think very hard about how to summarize your data both in words and slides. You need to motivate the audience to pay attention to your data.
2. Data Overview: How large is your data? Time is limited, but you need to explain what information is available in your data. Your data consists of rows and columns, and each cell contains some information. You need to decide how much you will introduce.
3. Data preparation: Depending on your source, you might need to clean up your data. You need to explain what you would do. If you use Kaggle data, you must explain how your Kaggle data is different from the original source. If the original source is not publicly available, you must know the steps necessary to get it.
4. Six figures: You need to show six figures. Some figures should describe the overview of your data, and the rest of the figures need to highlight something interesting. You can split on your own. Also, you need at least two figures using ggplot and at least one figure using Tableau. For example: 5 ggplot and 1 Tableau are fine, but 1 ggplot and 5 Tableau are not. This is because we spent more time on R in class.
5. Pros/Cons: You must address the potential advantages of your data and make clear what's the limitation when using yours.

You will write a review paper accordingly.