

Case Study Guidelines

Due: 09.45am, Thursday, May 3rd, 2018

General Details

- You are required to complete this case study in groups of 2; you can choose your own groups. If you are unable to find a partner, contact me and I will find you one. If you do not contact me I will assume you have a partner.
- If you are unable to complete this case in a group of two (either you are alone or your group exceeds two people), you must come see me to be granted an exception, otherwise this is not permitted.
- You must submit a paper copy of the case study at 09.45am, Thursday, May 3rd, 2018 in class. You are also required to upload your case study to Canvas in both `pdf` and `Rmd` format. Late submissions are not permitted and will receive a grade of 0.
- As the case study is heavily focused on graphics, the deliverable should be printed in color, double-sided and stapled in the top left corner.
- The case will be competitively graded on a scale of 0 – 100. The best case study (or case studies) will receive a grade of 100, and all other case studies will be receive a lower grade, based on how they compare to the best case(es).

About the Case Study

The body of the case study should be a mix of text and graphs. The Case Study should include the following sections (in this order):

1. Executive Summary
2. Introduction to the Data
3. Exploratory Analysis
4. Conclusions and Future Analysis

The **Executive Summary** should be no more than 250 words explaining, with extreme concision, your data and findings.

In the **Conclusions and Future Analysis** section, discuss what data was missing from the data set. What additional data would you have liked to have during the analysis? What other questions could you answer if you had this data?

- All graphs generated in the case study must be `ggplot` graphs, and the final document should be an `RMarkdown` (`Rmd`) file. Graphs should take up no more than half a page.
- Be creative and use `ggplot`'s layers to create interesting and insightful graphs that captivate the reader.
- Be sure to clean the data before you analyze it. If you have messy data with `NAs`, incorrect data, etc., how did you deal with it? Row-wise deletions, zeroing-out data, imputation? Explain and justify your approach.
- The Case Study is expected to be 5-10 pages in length when printed. It should contain a mix of `ggplot` graphs and text explaining the graphs, the specific data used for those graphs, and the conclusions you arrive at based on those graphs.
- Be sure to set `ECHO = T` in all code chunks which generate graphs, so the reader can see the `ggplot` code in the body of the case study.

- You should not have graphs in your case study that are not referenced in the text, i.e., no orphaned graphs.
- You are not required to run any inferential statistical tests on the data, e.g., confirming that a significant statistical relationship exists between one variable and another. You are however recommended to run a descriptive statistical analysis, i.e., means, medians, quartiles, etc.

In the body of the report be sure to care for the following: clear purpose and organization, high-quality visuals, appropriate grammar and use of language.

About Case Study Data

There are a few loose guidelines about the data you should follow.

1. The data you use should be sourced from a reputable location and should be real, i.e, not fabricated.
2. You are required to provide a `url` to the source for your data.
3. The data should ideally be at 10,000 observations and 10 variables, resulting in a minimal amount of 100,000 entries. This is a guideline, not a rule. I will ultimately confirm with you if your data is sufficient substantial to support a case study.
4. Be passionate about your data. Like sports? Use sports data. Interested in global statistics? Visit the UN website. Like weather patterns? Get weather data. There are no rules as to the type of data you use. I strongly recommend you choose a data set that personally interests you so that you might be naturally motivated to take a deep dive and extract significant insights.

Remember

- Style matters. A nice, clean deliverable in a legible font, well-spaced lines, graphs that are not cut off with axis-labels in legible fonts, etc., is important.
- Be succinct in your writing but not deficient. If you can explain something in 5 words instead of 10, do it. If you can submit a 2-page case study analysis and say exactly what someone who submits a 10-page case analysis does, you will be rewarded with a significantly stronger grade.
- The English language is important. Pay attention to grammar and syntax. Do not write colloquially: you are not tweeting, you are writing a formal case study. Poor writing will bring lower your grade.

Grading Scheme

Executive Summary	15%
Quality of Data	15%
Body of Report (Text)	20%
Body of Report (Graphs)	50%