ETC5521 Assignment 1 Instructions

Di Cook & Emi Tanaka

2020-08-11

The assignment is designed to develop your experience in planning an exploratory data analysis, determining the appropriate questions to answer about the data, and taking the initial steps for the analysis. The assignment involves several parts. The assignment represents 20% of your final grade for ETC5521.

1. Get in touch and chat with your assignment 1 buddy. The information on this is available in the *first sheet* here (https://docs.google.com/spreadsheets/d/1rH4LjF56yMTUzzMuxiaqKFVZdPXrYEbCewnu0jbobrA/edit?usp=sharing). Only your Monash username is shown here, so add the rest of the email address to get in touch with your buddy. If you have problems one of the instructors or tutors can help out.
2. With your buddy, choose a data set from the list from the *second sheet* here (https://docs.google.com/spreadsheets/d/1rH4LjF56yMTUzzMuxiaqKFVZdPXrYEbCewnu0jbobrA/edit?usp=sharing) and **write your team name in the third column** against the data set you picked. If another team has already chosen it, you will need to choose another. You may only select one dataset.
3. Accept the GitHub Classroom Assignment here (https://classroom.github.com/g/TLS4kzRE) using a GitHub Classroom compatible web browser (e.g. Chrome or Firefox; Safari gets stuck when accepting the assignment so do not use Safari). This should generate a public GitHub repository that can be found at https://github.com/etc5521-2020 (https://github.com/etc5521-2020). **You (or your buddy) will need to create the team when accepting the assignment (it should match your team name in Step 2) and your buddy will need to join that team.** Your GitHub assignment 1 repo should contain the file `README.md`, `index.Rmd`, `etc5521-assignment1.Rproj` and `.gitignore`.
4. Write your data description under the header “Data description” in `index.Rmd`. A data description should tell the reader the source of the data, and detail the structure of the data, including variables, time frame of collection, collection methods.
5. Brainstorm what questions would be possible to answer with this data (and only this data). Articulate the primary question, and then three secondary questions.
6. Finish your report, which should be structured as follows
    - Introduction and motivation: Give the bigger picture of the data, and inspire the reader to learn more about the problem by reading your analysis.
    - Data description, details given in 4.
    - Analysis and Findings, which includes at least one plot or numerical summary for each of your questions, that helps the reader arrive at an answer. You should also write paragraphs describing the methods, summaries and findings.
    - References
7. Provide peer feedback on assigned report. You will be randomly assigned another team’s report, and be expected to reproduce their analysis, and provide constructive input on the analysis. The purpose of peer feedback is to learn how other analysts approach problems, and to help improve other analysts analyses.

Deadlines:

Due date	Turn in	Points
Step 1, 2, 3: 14th Aug	We will check that this has been done on Github Classroom	1
Step 4, 5: 21st Aug	We will check that all have contributed these pieces to github repo	1

Due date	Turn in	Points
Step 6: 3pm 28th Aug	Final report available on repo	15
Step 7: 4th Sep	Peer feedback provided as an issue on your assigned group's repo	3

Marking guide for assignment

To help you complete the assignment, below is a rubric to guide you to what we are expecting in your final report:

content	description	Excellent (HD)	Very good (D)	Good (C)	Satisfactory (P)	Unsatisfactory (F)
Introduction and motivation	Explanation of the problem of interest (20%)	Motivation and explanation of the problem of interest to communicate the scenario. Outline encouraging exploration of other sections. Data sources explained, including limitations that might affect possible analysis and conclusions.	Explanation of problem of interest is VERY clear and provides information about the scenario. Details of data sources are provided, with limitations affecting analysis.	Explanation of problem of interest is clear and provides information about the scenario.	Explanation of problem of interest is rudimentary and lacks detail.	Explanation of the problem is unclear and/or not shown. There is no explanation of the problem to be solved.

content	description	Excellent (HD)	Very good (D)	Good (C)	Satisfactory (P)	Unsatisfactory (F)
Data description	Description of data, any cleaning procedures and transformation (20%)	List of questions to be addressed by the analysis. Description of the variables, as organised into tidy form. Detailed and concise explanation of data being used and what was observed from the original source with a comprehensive overview of the methods used to tidy and wrangle data including reasons. Original data source, and Tidy Tuesday site cited.	List of questions to be addressed by the analysis. Description of the variables, as organised into tidy form. Original data source, and Tidy Tuesday site cited.	Questions, and data description are soundly presented and demonstrates ability to translate tidy data form into appropriate questions.	Questions, and data description is reasonably presented and lists basic understanding of translating tidy data form into usable questions.	Description of data is unclear and/or not shown and demonstrates no or little understanding. Questions relate to methods to be used rather than the data.

content	description	Excellent (HD)	Very good (D)	Good (C)	Satisfactory (P)	Unsatisfactory (F)
Exploratory Data Analysis	Graphical representation of data (40%)	Choice of plots match the questions being answered and problem being studied. Appropriate mappings of variables to plot elements. Use of proximity and similarity and other cognitive principles in plot design. Neatly labelled axes and legends. Annotations on plot as needed to indicate important features, e.g. outliers labelled. Some interactive plot elements, like mouse over labelling. Explanations of what is learned from each of the plots is provided, and plots have appropriate captions.	Choice of plots match the questions being answered and problem being studied. Appropriate mappings of variables to plot elements. Use of proximity and similarity and other cognitive principles in plot design. Some interactive plot elements, like mouse over labelling. Explanations of what is learned from each of the plots is provided, and plots have appropriate captions.	Choice of plots match the questions being answered and problem being studied. Plots not optimally designed, but rationale for choices is clear. No interactive elements. Explanations of what is learned from each of the plots is provided, and plots have appropriate captions.	Choice of plots match the questions being answered and problem being studied. Plots not optimally designed. No interactive elements. Explanations of what is learned from each of the plots is provided, and plots have appropriate captions.	Plots don't match questions. Poor plot design, no rationale. No explanations on what is learned, or captions.
Expression and grammar	Scholarly, succinct with correct spelling, grammar and punctuation (10%)	Writing style is exceptional, scholarly and succinct that is free from spelling, grammar and punctuation errors.	Writing style is scholarly, free from spelling, grammar and punctuation errors.	Writing style is scholarly, but wordy. Free from spelling, grammar and punctuation errors.	Writing is scholarly and wordy. Contains some grammatical, punctuation and spelling errors.	Writing is unscholarly. Many grammatical, punctuation and spelling errors.

content	description	Excellent (HD)	Very good (D)	Good (C)	Satisfactory (P)	Unsatisfactory (F)
References	Application of accurate and consistent APA 6th style (10%)	The appropriate referencing style has been used consistently, with no errors.Includes citations for software used, and data sources.	The appropriate referencing style has been used consistently, with very few errors.	The appropriate referencing style has been used consistently, with few errors.	The appropriate referencing style has been used much of the time, but attention needs to be given to reducing the number of errors.	Material used from other sources without citation.

No late turn-ins accepted