ETC1010 - 5510 Project

ETC1010 and ETC5510

2021-04-23

The project is designed to give you experience collecting or finding your own data set, determining the appropriate questions to answer about the data, and planning how to execute analysis of the data. The project involves several parts. **The project represents 15%** of your final grade for both ETC1010 and ETC5510.

1. Locate a suitable data source and determine appropriate questions that could be answered using this data. It cannot be data set from kaggle. It needs to be from an original source. If it is in csv format, there need to be more than one file or multiple sheets. **Challenge yourself to work with data addressing a problem in today's world and have fun with it and with your teammates!**
2. Cleaning of your data, in order to answer your questions. This is the important part to illustrate in your project, because we are expecting you to be able to demonstrate your ability to take a messy data set and organise it for later analysis.
3. Use different visualizations to explore the data.
4. Simple analysis using methods covered in class; exploratory data analysis, numerical and visual summaries of the data, and the application of basic modeling strategies. The focus is on trying to answer some of the questions you posed. You are not expected to answer all, if you have a long lots of questions.
5. Each team must write a report using an Rmd file with the final project and each team member much contribute. **Each team member must choose at least a research question to investigate.**
6. Each team will present their work in a 15minutes presentation follow by questions.

IMPORTANT: If the analysis is completed using software other than R, or not written up using R Markdown, the project should receive a 0 regardless of its content.

IMPORTANT: The Rmd file containing all the project must knit without errors. If the Rmd file does not jnit the project should received a 0 regardless of its content.

IMPORTANT: This project will be conducted collaborative with your team.Individuals that fail to contribute to the group work will get a 0 in this assessment. **Each team member must choose at least a research question to investigate.**

Each team member is expected to participate equally in all aspects of the work, including the writing and oral presentation. Each team will produce a reproducible report written in Rmd.

Due date	Turn in	Points
Milestone 1: 30th April (Week 8)	Get in touch with team members and select data and tentative questions prospective team members	10
Milestone 2: 10th May (Week 10)	Electronic copy of your data, and a page of data description, and the cleaning done.	10
Milestone 3: 24th May (Week 12)	Final version of the Rmd report + **grading (confidentially)** the work of each of the other members of the team	70

Due date	Turn in	Points
Milestone 4: (Presentations Week 12)	Project presentations during class periods. **All students are expected to attend, and points will be deducted for non-attendance.**	10

Marking Guide for the report (must be produced on an Rmd file)

To help you complete the project, below is a rubric to guide you through the different sections that your report must have.

content	description	Excellent (HD)	Very good (D)	Good (C)	Satisfactory (P)	Unsatisfactory (F)
Introduction	Explanation of the problem of interest (10%)	Motivation and explanation of 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.	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	Rationale for data selection. Description of cleaning procedures and analytics (15%)	List of questions being addressed. Key parts of analysis clearly explained. Detailed and concise explanation of data being used and what was observed with a comprehensive explanation of the methods used to tidy and wrangle data including reasons.	A description of the chosen data and includes an informed rationale for its use. Learnings from analysis are clearly articulated.	Description and rationale is soundly presented and demonstrates some learnings from analysis.	Description and rationale is reasonably presented and lists basic learnings from analysis.	Description of chosen data is unclear and/or not shown and demonstrates no or little learning.

content	description	Excellent (HD)	Very good (D)	Good (C)	Satisfactory (P)	Unsatisfactory (F)
Visualisation	Graphical representation of data (15%)	Choice of plots match the required analysis and problem being studied. Appropriate mappings of variables to plot elements. Use of proximity and similarity and cognitive principles in plot design. Neatly labelled axes and legends. Annotations on plot as needed to indicate important features, e.g. outliers labelled. Excellent use of interactive elements, such as mouse over, or animation to help communicate additional information.	Choice of plots match the required analysis and problem being studied. Appropriate mappings of variables to plot elements. Use of proximity and similarity and cognitive principles in plot design. Neatly labelled axes and legends. Annotations on plot as needed to indicate important features, e.g. outliers labelled.	Choice of plots match the required analysis and problem being studied. Appropriate mappings of variables to plot elements. Use of proximity and similarity and cognitive principles in plot design. Neatly labelled axes and legends.	Appropriate mappings of variables to plot elements. Use of proximity and similarity and cognitive principles in plot design. Neatly labelled axes and legends.	Incorrect usage of visualizations.

content	description	Excellent (HD)	Very good (D)	Good (C)	Satisfactory (P)	Unsatisfactory (F)
Analysis components	Components that communicate the analysis being carried out (10%)	Separate components or tabs, that clearly correspond to each of the main questions being addressed. Highly appropriate choice of data plots included, numerical summaries and application of models. Exceptional user interaction elements appropriate for helping to explore the data, matching different aspects of the questions of interest.	Separate components or tabs, that correspond to each of the main questions being addressed. Appropriate choice of data plots and user interaction elements for exploring data.	Separate components or tabs, that correspond to each of the main questions being addressed. Reasonable choice of data plots and user interaction elements.	Separate components or tabs correspond to main questions being addressed, but lacks detail. Rudimentary choice of data plots and user interaction elements.	Separate components or tabs not used and/or main questions being addressed is unclear. Inappropriate plot choice and/or not used. Inappropriate user interaction elements and/or not used or do not allow for exploration of data.
Conclusions	Answers to the research questions and conclusions from the data analysis exercise (10%)	Relevant conclusions explicitly tied to analysis and to context.	Conclusions relevant, but partially correct or partially complete.	Conclusions relevant, but partially correct or partially incomplete.	Conclusions are correct.	Conclusions incorrect or not based on facts.
Expression and grammar	Scholarly, succinct with correct spelling, grammar and punctuation (5%)	Writing style is exceptional, scholarly and succinct that's 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 (5%)	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 fro mother sources without citation.

No late turn-ins accepted