Overview

This is the first of 3 unit projects that you will complete this semester. All unit projects are group projects. Once established, you will work with the same team of 4 -5 people for the entire semester. Group members are expected to contribute equally to the data analysis and project write-up. You may set up any communications and/or meeting system needed to achieve group cooperation and input.

You will choose one dataset from the approved list on ELMS. You will work with for the same dataset for the entire semester.

Unit Project 1 Instructions

For Unit Project 1, your main task is to write a Briefing Memo to your supervisor, Dr. Random. Dr. Random has asked that you perform a basic data exploration in order to:
1) give him an initial profile of the population from which the data was taken,
2) explain how the data was collected and any possible limitations in the data method, and
3) identify where there might be some potential associations that should be explored with further statistical analysis. You will then conclude with a set of recommendations.

You will use an R Markdown HTML notebook to write your briefing memo. A template for this assignment is provided on ELMS. You will write your analyses in full sentences, with complete answers to the questions that are posed for each part of Format and Analyses. **Your project team will submit both an R Markdown notebook file and the HTML file that is generated from the template. There will be only one submission per team.**

EACH PERSON ON YOUR TEAM IS RESPONSIBLE FOR RUNNING AND WRITING UP ONE OF THE DATA ANALYSES. The name of the responsible person should be listed with the analysis.

Format of Briefing Memo (Using the R Markdown template)

1. Header (Group Responsibility): Create a header for your memo entitled "Briefing Memorandum", followed by "To:", "From:", and "Subject:" lines

2. Background (Group Responsibiity): Provide a brief (one paragraph, please!) background description of 1) your task and 2) the context of your analysis. The context of your analysis MUST include: origin of dataset (if applicable), population of interest, data collection method, size of the data set, and a table of variables/factors to be explored with an explanation of what they mean and what type of variables they are.

3. Analysis (Individual Responsibility): ***Each team member is individually responsible for the written analysis in the team report which corresponds to his/her designated statistical analysis, as explained in Table below (Analyses A1 – A5).*** The written analysis should include any summary numerical and graphical output for the analysis, neatly displayed and labelled. Be sure to respond to the questions given with each analysis in Table A. **Use full sentences.** Remember, you are explaining your results to Dr. Random. *As you discuss your analysis, be sure to interpret what you are finding in the context of your particular data situation.*

4. Recommendations (Group Responsibility): In this section, your team will briefly summarize the most interesting features of your data. Are you seeing any trends, associations, or interesting features? Highlight these and make a recommendation that they be subjected to further statistical analysis. If there isn't anything interesting, say so!

Individual Analyses (Part 3 of Briefing Memo)

Explore/Analyze Your Data!

Table A presents the list of the analyses that will be included in your briefing memo. EACH TEAM MEMBER IS INDIVIDUALLY RESPONSIBLE FOR ONE ANALYSIS. You will decide among yourselves who will do what. Each person on the team will develop the listed R analysis and discussion/interpretation text for their assigned analysis. The R code and discussion/interpretation text will be incorporated into the SINGLE R Markdown file for the entire team , properly annotated and commented where necessary. ***The individual responsible for the code and analysis must be clearly indicated on the template.***

Table A

	Analysis Using R	**Discussion/Interpretation**
A1. **Analysis of One Quantitative Variable**:	For one of the quantitative variables, find the summary statistics (mean, standard deviation, five number summary) and at least one appropriate graphical display of the data.	Discuss any data cleaning or transformation that you had to make to analyze the variable. State and explain the meaning of the measures of center and spread from your analysis. Are there any outliers? Is the distribution symmetric, skewed, or some other shape? How can you tell?
A2. **Analysis of One Categorical Variable**	For one of the categorical variables, create a frequency table AND a relative frequency table (be sure this is labeled, either in comments and/or on the titles of the tables). Produce at least one appropriate graphical display of the variable.	Discuss any data cleaning or transformation that you had to make to analyze the variable. Give the definition of the levels of this variable. Explain in full sentences what your numerical and graphic summaries mean.
A3. **Analysis of One Relationship between Two Categorical Variables**	For two categorical variables, produce a two-way table and discuss any relevant proportions. Produce at least one appropriate graphical display of the relationship between the variables.	Discuss why you would hypothesize a potential association. Discuss any data cleaning or transformation that you had to make to analyze the variables. Give the definition of the levels of this variable. Does there appear to be an association between the two variables? Discuss why your summaries leads you to that

		(potential) conclusion. Describe the association, if one appears to exist.
A4. **Analysis of One Relationship between a Categorical Variable and a Quantitative Variable**:	Produce a table of some summary statistics (quantitative variable) to compare the groups (categorical variable). Include a side-by-side plot and describe it.	Discuss why you would hypothesize a potential association. Does there appear to be an association between the two variables? Discuss why your summaries leads you to that (potential) conclusion. Describe the association, if one appears to exist. .
5. Analysis of One Relationship between Two Quantitative Variables:	For a pair of quantitative variables, create a scatterplot.	Identify which variable is response variable and which is the explanatory variable. Explain/justify your choice! Explain what you are seeing on scatterplot in terms of trend and direction. Comment on the strength of a potential linear association.