

Seattle Public Utilities (SPU) has provided data from repeated surveys: one relating to service ratings and the other customer environmental behavior. SPU provides garbage, recycling, yard/food waste, water, sewer and storm water services to Seattle residents. One of SPU's goals is to provide equal levels of service quality regardless of gender, race, age, or housing class. Further, SPU services have a direct impact on environmental quality. As a public agency SPU wants to minimize the impact of their customers' behavior on the environment with respect to the services they do or could provide.

Task: EDA Technical Appendix and graph

You will be provided with one of two data sets from SPU, one with customer ratings of SPU services or one with ratings of customer environmental behavior. Both data sets contain some demographic data as well, but it is limited. Data visualization is a key component of Exploratory Data Analysis (EDA), along with examining descriptive statistics of individual variables and relationships between variables. You are to perform EDA on one of the data sets as specified below, trying to address one of the primary concerns mentioned above. You are essentially doing the Technical Appendix portion of an analysis, that is, you do not need to write up a report or make publication quality graphics. However, you do need to document your steps and your findings as you perform your EDA. One thing to note is that the data sets have many NA observations so you will need to use the command "na.rm = TRUE" for many of your calculations or visualizations. You will submit your R code file to the class web page on canvas at the end of class, so make sure to put your and your partners' names at the top of the R code. Please work in groups of 2 or 3.

Data: data sets are on the Canvas web page, a variable key is not provided but you should be able to figure out what each variable represents.

If you are sitting on the left side of the room download: 2_EBI_Data.csv

If you are sitting on the right side of the room download: 2_RS_Data.csv

Software: You should use R because Excel does not deal with NA values and there are several 1000 observations so Excel could be cumbersome without knowing proper short-cuts.

Technical Appendix and graph to be submitted to the link on Canvas at end of class

1. Technical Appendix is basically your R code. It is important for your work to be clear and reproducible. Provide a clear description of what you have done by documenting the steps of your analysis in your code. This often includes your data source, the steps you took and "code" with clear and complete documentation (i.e. write out why/what you are doing next to calculations). You should describe what question you are trying to answer when working on a code chunk and what you find: the question can be posed at the top of the code chunk and the findings at the bottom.
2. Graph. Select the graph that you feel best displays the most important finding of your EDA. The main title of the graph should indicate what the graph shows and the subtitle should contain the names of people in your group.
 - R basic: (... , main = "Older customers get worse service", sub = "Jeb, Jab, and Jib")
 - ggplot2: ggtitle("Older customers get worse service\n sub = "Jeb, Jab, and Jib")