

## Data Description & Descriptive Statistics

Your task for this project is to describe a set of data of your choice and to perform some basic statistical analyses. The report should include:

- The source of your data and a description all relevant variables.
- The data (or a subset of the data.) (This might be a copy from the actual source, a handwritten sheet, or (preferably) a printout from the computer.)
- Appropriate Summary Statistics (These should be calculated in R. It is not sufficient to merely provide output with no explanation or interpretation. You do not need to explain definitions of the statistics, but you should indicate what the statistics say about your data. Try to examine relationships as well as describing individual distributions.) You might think of making an appropriate table of some kind.
- Graphical Displays (These should effectively summarize your data and point out any interesting features. These must also be interpreted. You do not need a picture or table for every variable.)
- A comment on anything of interest that occurred in doing the project. Were the data approximately what you expected or did some of the results surprise you? How did the sampling go? Do you think you got a representative sample of your population?

## Notes:

- Your data should include at least 10 variables, with at least 4 independent quantitative and at least 4 independent categorical variables. (Label/ID/name does not count as a categorical variable because we can't summarize it.)
- You should have at least 30 independent cases / observations.
- Because we will be doing hypothesis testing, you need to indicate what population your data describes. If it is a census, then maybe it is representative of an even larger population? (For example, a census of state information from 1999 might be somewhat representative of 2000?) Also, discuss the limitations of describing a larger population.
  - Instead of doing a census of one year, you might try doing a stratified random sample... For each of the years between 1990 - 2000, take a random sample of 5 states.
- This report does not need to be written up as a paper, but it should not be simply computer output. Make sure that everything you turn in is annotated.
- Do not be tempted to turn in everything you do. Only turn in the interesting parts of the analysis. One of the hardest parts of being a consultant is figuring out what to tell the researcher.
- Computer output that is attached and not described will be ignored.
- Ideas for possible projects and data sources are available on my web site.
- Please type and double space your paper so that it's easier for me to read/grade.