General Directions for Analyzing Project Data

This is a general guide to help you get started in analyzing your data.

Work with your group members. Refer to your friendly neighborhood statistics instructor if you have any questions/problems.

Some steps may not apply in your particular case.

Preliminaries

- Keep a Word document as a Journal of your progress
- Load the data into SPSS

    - Start up SPSS
    - Use the File -> Read Text Data wizard
    - Pay attention to steps 2 and 4

- Set up the data properties

    - Data -> Define Variable properties
    - Check that "measure" is appropriate (Nominal/Ordinal/Scale)
    - Add appropriate Label to the variable (top right)
    - For categorical variables, add value labels next to each value
    - Look for values that shouldn't be different (e.g. "yes" and "Yes" as separate categories)

- Save the data file often

Early Explorations

- Goal: To **own** your data. You should get a grasp for the basic behavior of your data.
- Compute appropriate numerical summaries and graphs for each individual variable

    - Categorical: Frequency tables, bargraphs, pie charts
    - Scalar: Mean, Std.dev, quartiles, skewness, outliers, histogram, boxplot

- Summarize conclusions in journal
- Don't worry about finalizing graphs/tables. This is "exploratory": It's for you to learn what is going on with your data.
- You may want to also save your viewer file
- Consider interactions between your variables

    - Categorical - Categorical: 100% stacked bar charts
    - Categorical - Scalar: boxplots, or split the file and compute numerical summaries
    - Scalar - Scalar: scatterplot, "loess" line, linear regression if appropriate

- Decide on variable transformations

  - **–** Should a scalar variable be "grouped up" to form a categorical variable?
  - **–** Should some categories in a categorical variable be consolidated to create more balanced groupings?
  - **–** etc

Moving further

- Start looking at the questions you considered in your proposal's "expectations" section.
- For each question

  - **–** Prepare graphs and numerical summaries that would help answer it.
  - **–** Consider the effect of other variables to the interaction you observe (for example might a GPA effect based on gender be an effect based on year behind the scenes, because you happened to have too many male seniors in your survey?)
  - **–** Consider the effect of group/sample sizes on your conclusions. Do you have too few people in a group to draw conclusions about that group? Do you have outliers that might affect what you are seeing?

- Keep asking questions and digging deeper
- When you are satisfied, start putting together two things:

  - **–** Word document
    - ∗ Brief introduction and conclusions sections
    - ∗ Sections for each question/topic you wish to address
    - ∗ Graphs/tables to back up your results
    - ∗ Make sure to customize the graphs/tables, make them **presentable**
    - ∗ Graphs are best customized in SPSS first, before transfering.
  - **–** Powerpoint presentation
    - ∗ Edit the graphs (they probably require different editing than the document ones)
    - ∗ Graphs should be readable from a distance. Test them!
    - ∗ Avoid long text