GOT DATA?

Lab 2: Exploring Data with Graphical Displays and Numerical Summaries

INTRODUCTION

Even if you are not planning to be a statistician, it is likely that sometime in the future you will use data to make a decision or to simply learn more about a population or subject of interest. The first step in exploring data is to make a graphical display. A graphical display helps to identify overall patterns and detect potentially unusual observations. Graphical displays along with numerical summaries are fundamental steps to analyzing a data set.

Table 2.1 lists common graphical displays and numerical summaries for different variable types.

Table 2.1: *Graphical Displays and Numerical Summaries by Variable type*

Variable Type	Graphical Display	Numerical Summary
Categorical	Pie Chart or Bar Graph	Frequencies, proportions (or percentages)
Quantitative	Dot plot, Stem and Leaf Plot, Histogram, or Boxplot	Mean, median, standard deviation, range, mode, quartiles
One categorical variable and one quantitative variable	Side-by-side Boxplots or Side-by-side Stem and leaf Plots	Mean, median, standard deviation, range, mode, quartiles for each level of the category variable
Two categorical variables	Side-by-side Bar Graphs or Side-by-side Pie Charts	Frequencies, proportions, conditional proportions
Two quantitative variables	Scatterplot	Mean, median, standard deviation, range, mode, quartiles for each variable, correlation coefficient

SETTING

In order to learn more about fellow students in your course, a class survey is administered. Each question of the survey represents a different variable. Your instructor will collect results from the survey and summarize them in a spreadsheet in advance. The variables are displayed in the columns of the spreadsheet and the observations or individual student responses are displayed in the rows of the spreadsheet. Graphical displays and numerical summaries will be used to draw conclusions from the survey results.

Class Survey:

1. **Gender:** What is your gender? F for female and M for male.
2. **Commute:** About how long (in minutes) does it take you to get from your residence to your first class of the week using your typical mode of transportation?
3. **Eat:** How much money did you spend the last time you went out to eat?
4. **Residence:** Do you live on or off campus?
5. **Pets:** How many pets do you and/or your family have?
6. **Work:** How many hours per week do you work during the semester?
7. **Text:** About how many text messages do you send per day?
8. **Clothing:** How much did you spend on the last article of clothing that you purchased?

MATERIALS

Statistical Software

METHODS

For each of the eight questions, the survey asked for values of a variable. Use Table 2.2 to classify each of the variables as categorical (C), discrete quantitative (DQ), or continuous quantitative (CQ).

Your lab instructor will provide access to the data set collected for your class. In addition, the lab instructor will show you how to create graphs and find numerical summaries using statistical software. Afterwards, you will explore the data in your group by completing the tasks in the Discussion Question section. Spend time becoming familiar with the software because we will use the functions that you learn today in almost all of the remaining labs.

Reality Check

An article in the *The New York Times* titled "For Today's Graduate, Just One Word: Statistics" describes data as "the raw material of knowledge" (Lohr, 2009). Erik Brynjolfsson, director of the Massachusetts Institute of Technology's Center for Digital Business is quoted: "We're rapidly entering a world where everything can be monitored and measured. But the big problem is going to be the ability of humans to use, analyze, and make sense of the data." The article tells the story of Carrie Grimes, an "Internet-age" statistician at Google who analyzes data to improve its search engine. She majored in anthropology and archaeology at Harvard but enjoyed using data analysis in her research. Fortunately for her, statisticians are "increasingly in demand and—even cool."

Name _____ Section _____

Turn in the Results and Discussion Question sections.

Results

Table 2.2: *Variable classification*							
Gender	Commute	Eat	Residence	Pets	Work	Text	Clothing

Discussion Questions

1. Choose a **categorical** variable and make an appropriate graphical display. Give your graph an appropriate title and copy it to a Microsoft Word© document. Beneath the graph, write at least one observation about the graph. Use software to calculate numerical summaries for the variable. Copy and paste these values into the Microsoft Word© document.

2. Choose a **quantitative** variable and make an appropriate graphical display. Give your graph an appropriate title and copy it to the Microsoft Word© document containing the previous graph. Beneath the graph, write at least one observation about the graph. Are there any outliers? Use software to calculate a numerical summary for the variable. Copy and paste these values into your Microsoft Word© document.

3. Choose **two categorical variables** and make an appropriate graphical display. Give your graph an appropriate title and copy it to the same Microsoft Word© document as the previous graphs. Beneath the graph, write at least one observation about the graph. Also, use software to calculate numerical summaries for the variables. Copy and paste these values into your Microsoft Word© document.

4. Choose one **quantitative and one categorical variable** and make an appropriate graphical display. Give your graph an appropriate title and copy it to the same Microsoft Word© document as the previous graphs. Beneath the graph, write at least one observation about the graph. Also, use software to calculate numerical summaries for the variables. Copy and paste these values into your Microsoft Word© document.

Print and turn in the Microsoft Word© document that your group created.

REFERENCES

Lohr, S. (2009, August 5). For today's graduate, just one word: statistics. *The New York Times*. Retrieved from http://www.nytimes.com/2009/08/06/technology/06stats.html