BS & IS 6500:601:801 Lab 1

We will be using data from the Grey Code Corporation (GCC). Here is a description of the corporation and the data:

Grey Code Corporation is a media and marketing company involved in magazine and book publishing and television broadcasting. GCC's portfolio of home and family magazines have been a long running strength, but the company has expanded to become a provider of a spectrum of services (market research, communications planning, Website advertising, etc.) that can enhance its clients' brands.

GCC's relational database contains over a terabyte of data encompassing 75 million customers. GCC uses the data in its database to develop campaigns for new customer acquisition, customer reactivation, and identification of cross-selling opportunities for products. For example, GCC will generate separate version of a monthly issue of a magazine that will differ only by the advertisements the magazines contain. They will mail a subscribing customer the version with the print ads that the GCC database has determined will most interest the customer.

A problem facing GCC is how to boost the customer response rate to the renewal offers that it mails to its magazine subscribers. The industry response rate is about 2%, but GCC has historically had a higher response rate. GCC's director of database marketing, Chris Grey, wants to ensure that GCC maintains its place as one of the top achievers in targeted marketing. In one effort directed at maintaining this position, GCC is currently considering the development of a targeted marketing strategy for a hobby-based magazine. The file *GCC* contains 38 columns (each corresponding to a distinct variable) and over 40,000 rows (each corresponding to a distinct customer). The *Description* worksheet contains a description of each variable in the file GCC. The *Data* worksheet contains the data to be used to train, validate, and test a classification method. The *NewDataToPredict* worksheet contains data on a new set of former subscribers to the hobby-based magazine whom GCC would like to classify as likely or not likely to respond to a targeted renewal offer.

Instructions for Lab 1:

Use the *GCC* data located in the Cases folder to explore and clean up the data on the *Data* worksheet. Use the **Guide for Descriptive Data Analysis** from folder #5, and the **Correlation Basics** from folder #7 on Springboard to help you. Review the JMP tutorials also via the Springboard links (Course Docs → JMP Support).

These are the steps:

1. Data Cleaning
    a. Check for any missing data.
    b. Consider recoding some variables.
    c. Consider transforming some variables.
    d. Consider deleting some variables.* You might want to use the Hide or Exclude feature instead of eliminating some variables. I say this since you may want to use those variables in future models.
2. Generate graphs and tables.
    a. Explore relationships between variables (correlations).

3. Generate some descriptive statistics.
    a. Check for outliers, measures of center, measures of dispersion, etc.
4. Document the processes used in 1-3 above. Upload your word document to the Dropbox by the due date. You can provide as much detail as possible as you will refer back to this document later. **Remember that you are not doing any analyses such as hypotheses tests or models! You are simply preparing the data for future analyses and submitting documentation of the steps you took to clean up the data and examine the variables.**