Homework 1: Exploratory Analysis

The primary goal of this assignment is to have you begin to "read" a data set using the graphical and numerical tools we have been discussing in class. You will do this by 1) applying and interpreting a small extension to two-variable mosaic plots, 2) analyzing your own data set, something not presented in class, and 3) thinking creatively about what can be inferred from the large data set compiled by the UCLA Registrar and explained in lecture 3.

1. In class, we discussed two-variable mosaic plots and illustrated their use in visually assessing the association between pairs of variables. During Lecture 2, we hinted at the fact that the logic they followed to display a two-way table of data could be extended to three- or four- or k-way tables. In that same lecture, we showed how you could use R to make such a table. Using the `mosaicplot` function, create a display of a 3-way table using the BRFSS data and describe what it means. Be very explicit about what the different rectangles mean and the logic followed to produce them.

2. In the last five years, we have seen the rise of the open data movement. Organizations are increasingly making their data available on the web for the general public to examine. Sites like `infochimps.com` aggregate a large number of data sets for the public to analyze. For this question, identify a data set of your own, something on a topic that interests you, and tell a story based on those data using some of the tools you have learned in class. Feel free to contact your TA or the instructor if you have questions about how best to load the data you've found into R or the software package you've chosen to work in. We are looking for a handful of graphics with some discussion of what you've found analyzing the data.

3. At the end of lecture 3, we introduced a data set consisting of all the schedules for all the students enrolled at UCLA in Winter of 2011. In their current format (each row referring to a student in a class), the data tell a very restricted story. We would like to see what these data tell us about life on campus. This will almost certainly require reformatting the data, creating a new data table with a different "unit of observation" (as opposed to the current "student in a

class"). Depending on the language you've chosen to do your analysis, this recoding might be easy or hard. If you have picked R, the instructor and TA can help you write a function to reformat the data – But you will have to specify exactly what data you want in each row and how you would propose creating it from the current data set. Then, tell the story of life on campus using the new data. What have you learned? Express this in terms introduced in class.