Stat 133, Fall 2014
Homework 2: Graphics with Traffic Data
Due on bSpace by Thursday, Sept 18, 11:55pm
Directions: Turn in a plain text file containing the `R` commands you use to solve the following problems. The file should be named LastnameFirstnameHW2 with either a .r or .R suffix. The file should run and produce the plots. You should also include all the plots generated as a single .pdf file. In many cases there is more than one answer that will work and some of the answers are somewhat subjective, just answer as best you can. Put your name in a comment at the top of the file, and add comments throughout to indicate problem numbers. Some problems involve functions we have not covered in class; for these you will need to use `help`, `help.search`, and/or the internet.

Traffic flow on I-80

The data available for this assignment can be found on the web at
`http://www.stat.berkeley.edu/users/nolan/stat133/data/flow-occ-table.txt`
Do *not* download this file to your computer. You will read it directly from the web into `R`.

These data have been collected by loop detectors at one particular location of eastbound Interstate 80 in Sacramento. There are six columns and 1740 rows in the data set. The rows correspond to successive 5 minute intervals for consecutive days in March 2003, where the data values in a row report the flow (number of cars) and occupancy (the proportion of time there was a car over the loop) in each of three lanes on the freeway. Lane 1 is the leftmost lane, lane 2 is in the center, and lane 3 is the rightmost. The first row of data corresponds to the first 5 minutes after midnight on the first day, and the last row corresponds to the five minute interval beginning at 00:55 on the last day. The original data are from the Freeway Performance Measurement System (PEMS) website: `http://pems.dot.ca.gov`.

1. **Read the data directly from the web into `R` using readLines,read.table, read.csv, or scan. Do not download the data to your computer as a txt file before reading it into R. Explain in one sentence the reason for choosing the function that you did for reading the data into R.**

2. **Using the `summary` function, explore the distributions for the flow in each of the three lanes. Which lanes have the most and least flow (or are they pretty similar)? (Add the answer as a comment to your file.)**

3. **Make a scatterplot of Flow2 against Flow3. To the plot add a red line with slope 1.5 and intercept 0. Is the statement, "The flow in lane 2 is typically about 50% higher than in lane 3," an accurate description of the relationship you found? (Add the answer as a comment to your file. You do not need to turn in the plot, just the code and your comment.)**

4. **Create a new vector called `fracday` that gives the fraction of a day at which each observation interval (described above) begins. For example, the first element of your vector, corresponding to the first observation, should be 0 for midnight, and the second element should be equal to `5/(24*60)`, since the second interval began five minutes later and there are 24×60 minutes in a day. Note that the dataset contains more than one day, so you need to figure out how to have the vector "reset" to 0 at midnight of each day.**

5. **Make a scatterplot showing the relationship between flow and time of day, with `fracday` on the x-axis and total flow (summing all three lanes) on the y-axis. To make the points a bit easier to see, change the plotting character to something different than the default and also change the size of the character. Include a comment in your file describing the pattern you see. You do not need to turn in the plot, just the code and your comment.**

6. **Consider the relationship between flow and occupancy. Plot flow against occupancy for lane 1. In one sentence, describe the shape of the points. Relate the shape you see to a property of traffic. Again, you don't need to hand in the plot, just the code used to make the plot and a comment about why the plot looks the way it does.**

7. **Make scatterplots in the same way as for #5, but one for each day that has complete data for that day. Make all the plots with the same vertical scale, and give each one an x axis label which tells what day it is (ie Day 1, Day 2, etc). For each day, make the color of the points different. Based on these scatterplots, which days do you think are on the weekend, and why? Also, look at the late evenings. Two of the days have a different pattern than the others. What do you think might have happened on those nights?**