Stat 133, Spring '08
Homework 1: R Introduction
Due 6 pm Tuesday, 5 Feb

Traffic flow on highways in California

The data available for this lab assignment can be found on the web at

`www.stat.berkeley.edu/users/nolan/stat133/data/flow-occ-table.txt`

These data have been collected by loop detectors at one particular location of eastbound Interstate 80 in Sacramento, California. There are six columns and 1740 rows in the data set. The rows correspond to successive five minute intervals from March 14 to 20, 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 original data are from the Freeway Performance Measurement System (PEMS) website

`http://pems.eecs.berkeley.edu/Public/`

Your tasks for this assignment are to:

- Read the data directly from the web into R using one of the R functions **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.

- Explore the data visually to answer the following questions[1]:

  1. Examine the relationships of the flows in the three lanes using boxplots. 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? Explain in one sentence.

  2. Consider the flows in the three lanes using time series plots. Which lane typically serves the most traffic? Is the statement, "When one lane is congested, the others are too," an accurate description of the relationship you found?

---

[1]These questions are derived from those supplied by John Rice, Dept. Statistics, U.C. Berkeley

3. 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 propoerty of traffic. Is this relationship born out in the other two lanes. Explain any similarities and differences that you find.

Consider using the R functions **plot**, **boxplot**, **density**, **hist**, **pairs**, and **points**. Also consider using some of the arguments to the plotting functions to improve the appearance of your plot, such as **ylim**, **xlab**, **main**, **log**, **type**, **col**, and **lwd**.

Submit for grading your

- Paragraph of findings of 200-300 words.

- One plot for each of the above questions.

- The commands used to conduct your analysis.

Your submission should be contained in a **single** `pdf` file. If for example, the report is composed in Word as a `doc` file that contains the written paragraph, code, and plots, then you will need to print it to a `pdf` file before submitting it.

Submit this HW via **bspace** by 6pm Feb 5, and turn in a hard copy in class on Thursday, Feb 7.