Stat 133, Fall '05
Homework 1: R Introduction
Due Tuesday, 13 Sep

Traffic flow on highways in California

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

`http://www.stat.berkeley.edu/users/rice/UCLA/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**. That is, 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 one of the following questions[1]:

    1. Examine the relationships of the flows in the three lanes. Are statements of the form, "The flow in lane 2 is typically about 50% higher than in lane 3," accurate descriptions of the relationships you found? If not, how would you verbally summarize the nature of the relationships?,
    2. Compare the flows in the three lanes. Which lane typically serves the most traffic? Make plots to support or refute the statement, "When one lane is congested, the others are too."
    3. Flow can be regarded as a measure of the throughput of the system. How does this throughput depend on congestion? Consider the following conjecture: "When there are very few cars on the road, flow is small and so is congestion. Adding a few more cars may increase congestion but not enough so that velocity is decreased, so flow will also increase. Now beyond some point, increasing occupancy (congestion) will decrease velocity, but since there will then be more cars in total, flow will still continue to increase." Does this seem plausible to you?

    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**.

- Write one paragraph (maximum 200 words) that describes your findings. Do not write about the code that you used to find your results.

The question that you are to address depends on the last letter in your stat account name.

[1]These questions have been supplied by John Rice, Dept. Statistics, U.C. Berkeley

Letter	Question
a - h	1
i - p	2
q - z	3

Submit for grading your

- Paragraph of findings.
- Two plots that best explain your findings.
- 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 sending it to the GSI.

Email this file as an attachment to Joel (s133@stat.berkeley.edu), with your name and stat user account in the subject of the email message and on the report, by 10 p.m. on the due date. Make certain to save a copy of your email submission.