Stat 133, Fall '05
Homework 2: Data Structures in R
Due Tuesday, 20 Sep

Getting Data in the Right Shape

Our analysis of data is often made easier by having it in a suitable structure. This lab will give you practice with reshaping, reformatting, and labeling data. The **rep**, **c**, and **seq** functions as well as **as.factor**, **as.numeric** etc. and **[ ]** and **[[ ]]** may prove very useful in this lab.

There are three parts to this assignment. You are to complete each part.

1. Use **read.csv** to read in the data used in the last assignment from the web at

   ```
   http://www.stat.berkeley.edu/users/rice/
                   UCLA/flow-occ-table.txt
   ```

   As you know, the object will be a six column data frame. Convert this data frame into a three column data frame, with the following columns,

   - Flow – contains the traffic flow for each of the three lanes stacked on top of each other, Flow1, Flow2, Flow3, i.e. the vector will be of length 3×1740.

   - Occ – contains the occupancy for each of the three lanes stacked on top of each other, Occ1, Occ2, Occ2.

   - Lane – a factor with levels "slow", "middle" and "fast", corresponding to lanes 3, 2, and 1, respectively, where the value in a row of the Lane vector in the new data frame corresponds to the lane of traffic in which the flow and occupancy measurements were taken.

   Write your code in a general way so that it will run correctly regardless of the number of rows in the data frame.

2. Use **read.table** to read in the data found on the web at

```
http://www.stat.berkeley.edu/users/nolan/
              stat133/data/Chip.txt
```

Be sure to

- Use the Name column for row names
- Use the first row for column names
- Eliminate the column called "speed".
- For those rows which have GHz as the speed, scale up the value in Clock by 1000.

3. Daily precipitation has been collected for 5 (a subset of the 56 available) weather stations in the Colorado Front Range and made available to you as an R object at the following

```
http://www.stat.berkeley.edu/users/nolan/
              stat133/data/FrontRange.rda
```

In addition to the day and amount of daily rainfall recorded, for each weather station we also have the station's latitude, longitude, elevation, the number of days the station was in operation, and the mean total summer precipitation at the station.

- Use R functions such as **class**, **length**, **length**, **dim**, **names**, and **as.numeric**, etc. to determine the exact structure of the R object FR. Include with your code, a few sentences that describe this structure, e.g. FR is a list of ___ elements, where the first element named ___ is a ___....
- Check that the number of rainfall recordings that you have for each station matches the number provided in FR.
- Find the average rainfall for each station.
- Find the average annual rainfall at each station.

Email this file as a **plain text** (include your code and answers in the same text document – you may want to use the comment symbol # before your answers) 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.