Stat 133, Fall '06
Homework 2: Data Structures in R
Due Friday, 15 Sep

Manipulating Data in R

Our analysis of data is often made easier by having it in a suitable structure. This lab gives you practice with reshaping, reformatting, and labeling data. The functions **as.factor**, **as.numeric** etc. and **[ ]** and **[[ ]]** may prove very useful in this lab as well as functions from the apply family, **tapply**, **lapply**, **apply**, **sapply**.

1. Load into R the R dataset found on the web at

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

   Make the following adjustments to the **chip** data frame and call the resulting data frame **intel**:

   - Use the **Name** column for row names, and drop this column from the data frame.
   - Turn **speed** into a factor.
   - For those rows which have a value "GHz" as the **speed**, scale up the value in **Clock** by 1000.
   - Create a new column called **RelativeSpeed** that rescales all values in **Clock** by the first value of clock speed.

   Be sure to write your code in a general way. For example, do not depend on the fact that there are 10 rows in the data frame, and you know that the last two rows have values of "GHz" in the **speed** variable. Instead, search in the **speed** variable for those elements with particular values, and use that information to adjust **Clock**.

2. Daily precipitation has been collected for 56 weather stations in the Colorado Front Range and made available to you as R objects at the following two rda files

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

   http://www.stat.berkeley.edu/users/nolan/
                stat133/data/FrontRange2.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.

   - Explore the difference between the structures of these two sets of R objects. Describe the difference in two-three sentences. Answer the following questions using either objects, and explain why you chose the object you did.
   - Check that the number of rainfall recordings that you have for each station matches the number provided in the variable **FR**.
   - Find the maximum annual rainfall for each year and each station.
   - Find the 99th percentile of the maximum values for each station. Provide one plot that well represents the distribution of these percentiles (recall the kinds of plots from the first week of class).

Upload 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) by 6 p.m. on the due date.