STATS 105 Basic R Commands

- *Importing and accessing data*

  - Importing data with no header row
    ```
    data1 <- read.table("http://www.stat.ucla.edu/~hqxu/stat105/data/EX06_24.txt")
    ```

  - Importing data with header row
    ```
    data2 <- read.table("http://www.stat.ucla.edu/~hqxu/stat105/data/EX11_04.txt",
    header=T)
    ```

  - View variable names
    `names(data1)` (returns: `[1] "V1"`; if no variable names specified, R assigns `V1`, `V2`, etc.)
    `names(data2)` (returns: `[1] "price" "tax"`)

  - View data
    ```
    data1
    data2
    ```
    `data2$price` (returns vector of just the `price` variable)

  - Attaching variable names for later reference
    `attach(data2)`
    `price` (can now just type variable name to return vector of just that variable)

  - Creating a variable from scratch
    `var1 <- c(1,6,3,7,2)` (creates a variable called "var1" with 5 observations)

- *Basic summary statistics*

  - Number of observations
    `length(price)` (must use variable name – and not dataset name – in this command)

  - Mean
    `mean(data2)` (returns the mean for each variable in the data)
    `mean(price)` (returns the mean for the variable `price` only if variables have first been attached;
        will compute statistics on one variable at a time more often)

  - Median
    `median(price)`

  - Percentile
    `quantile(price,0.15)` (returns the 15th percentile of the data, for example)

  - Sample variance
    `var(price)`

  - Standard deviation
    `sd(price)`
    `sqrt(var(price))` (returns the same value as the `sd` command)

  - Minimum, maximum and range
    `min(price)`
    `max(price)`
    `r <- min(price)-max(price)` (assigns a variable named `r` to be the range of the data)

  - Five-number summary and mean
    `summary(price)`

- ***Visual Data Displays***

  - Stem-and-leaf diagram
    ```
    stem(price)
    ```

  - Histogram
    ```
    hist(price, main="TITLE", xlab="X-AXIS LABEL", ylab="Y-AXIS LABEL")
    ```
    (titles optional)
    ```
    hist(price, breaks=seq(20,48,by=4), xlim=c(20,48), ylim=c(0,10))
    ```
    (customizes number of bins and scales of axes)

  - Box plot
    ```
    boxplot(price, main="Boxplot for Price", ylab="Price")
    ```

  - Scatter plot
    ```
    plot(price, tax)
    ```
    (`price` on x-axis, `tax` on y-axis)

  - Normal probability plot
    ```
    qqnorm(price,datax=TRUE)
    qqline(price,datax=TRUE)
    ```

- ***Linear Regression***

  - Simple linear regression
    ```
    fit <- lm(tax ~ price)
    ```
    (`price` is x-variable, `tax` is y-variable; doesn't have to be called `fit`)

  - Show regression output
    ```
    summary(fit)
    ```

  - Plot regression
    ```
    plot(price, tax)
    ```
    (`price` on x-axis, `tax` on y-axis)
    ```
    abline(fit)
    ```

- ***Saving work***

  - Text File
    - The easiest way to save your work is to write and edit or copy and paste all your commands in a text file. Then you can copy and paste them back into R when you need them.

  - Save all commands used during an R session
    - First, set your working directory to where you would like to save the file
      - On PC: File >> Change dir or Change Directory
      - On Mac: Misc >> Change Working Directory (or Apple+D)
    - Then save your history
      ```
      savehistory (file ="history .log")
      ```

- ***Installing R and More R Help***

  - Install the latest version of R from http://cran.r-project.org or http://cran.stat.ucla.edu
  - R reference card: http://www.stat.ucla.edu/~hqxu/stat105/R/R-refcard.pdf