Introduction to R

1. Why R?

"R is really important to the point that its hard to overvalue it. It allows statisticians to do very intricate and complicated analyses without knowing the blood and guts of computing systems." — Daryl Pregibon, a research scientist at Google

Advantages:

- Open source (SAS and Matlab is not, if you consider SAS as a statistical software)
- Abundant statistical packages
- Simplicity (Don't need to worry too much about variable type, memory allocation, loops. Very easy for beginners.)
- Data visualization (ggplot2, googleVis, rChart…)
- Nice IDE (Rstudio, Eclipse)
- Downside: it could be Slow (Need to be smart while coding)

2. Rstudio

Download link Or, download the preview version download link

- Short intro of Rstudio
    - Control R version
    - vim editor
    - Change theme
    - Version control with Git or SVN
- Create a project
- Create R script
- Save workspace and changed script
- Create other scripts (LaTex, R markdown, Rnw)

Useful shortcut for writing R script:

- Ctrl+Space or Tab, give you more information of the function.
- Tab can also give you the details of the argument.
- Ctrl+Shift+c comment out one line or a block of highlighted code.
- Ctrl+Shift+c remove the "#" on a block of commented code.
- While executing code, environment tab shows the objects in your current environment
- Executing the code: Ctrl+R or Ctrl+Enter or Source (useful for editing function: Source on save)
- Debugging: options(error=browser)

A demo of R markdown:

- Insert title "====="
- Insert section "#", subsection "##", subsubsection "###", keep going
- Insert bullet points "- "
- Insert list "1. "
- Insert list within list: Tab+"1. "
- Insert inline equation α
- Insert display equations

$$Y \sim Normal(\mu, \sigma^2)$$
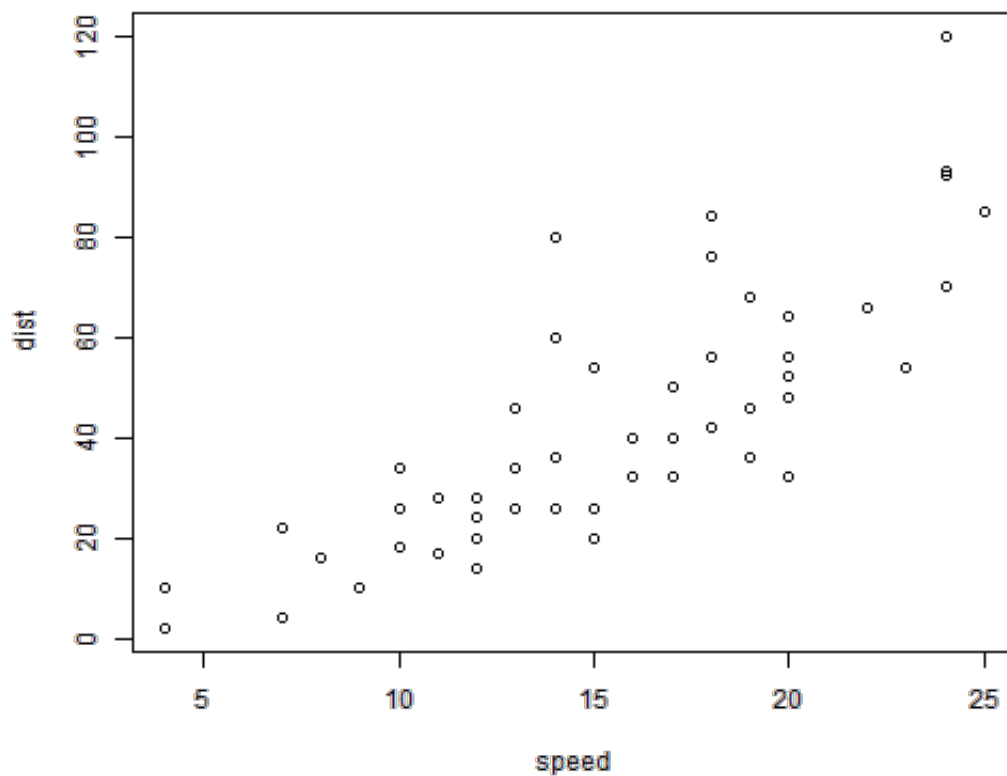
- Insert code `foo <- NULL`
- Insert and execute R code

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2
##  1st Qu.:12.0   1st Qu.: 26
##  Median :15.0   Median : 36
##  Mean   :15.4   Mean   : 43
##  3rd Qu.:19.0   3rd Qu.: 56
##  Max.   :25.0   Max.   :120
```

- Insert plot

```
plot(cars)
```

3. R Naming convention

Be consistent!
In all the 4411 packages on CRAN, there is a total of 339032 parameter names and 76176 function names.

- all lower case

  `searchpaths` ...

- period separated

  `as.numeric`, `read.table` ...

- underscore separated

  `package_version` ...

- lower camel case (suggested)

  `colSums`, `sessionInfo` ...

- Upper camel case

  `Vectorize`, `NextMethod` …