R course, session 1

1. introduction:

Introduce instructors.

Introduce course structure.

Students introduce themselves. (lab, research interest, why do you take this course).

2. Brief introduction of R:

R is a statistical computing language. Very powerful for computational biologists.

Strength:

1) it has many statistical models integrated (function, data type).

2) it has very powerful graphical capabilities. (it provides a fully programmable graphical language, ggplot2).

3) it is free and open source software, allows anyone to use and modify.

4) it is very useful for biologists. (bioconductor > 1000 packages for biological data analysis interact with sequencing data).

Weakness:

1) Not very good at working with data of very complicated structure (designing a games).

2) Not very good when dealing with very large data set. (R has to read entire dataset into memory before working on it).

Distributed computation framework: Hadoop, spark

3. Download:

R-3.4.3:

<https://mirrors.sorengard.com/cran>

Windows: base -> “Download R 3.4.3 for Windows”

Mac:

Check mac OSX version: Apple menu -> “About This Mac”

R-3.4.3.pkg for El Capitan (OSX 10.11) and higher

R-3.3.3.pkg for Mavericks (OSX 10.9) and higher

R-3.2.1.pkg for Snow Lepard (OSX 10.6-10.8)

Rstudio:

4. introduce Rstudio (graphical user interface).

5. Teaching.