Data Analytics 1-24-19

install.packgages(“packagename”) #To install a package

library(packagename) #To call a function in a package, you must load it first

getwd()

Setwd(“C:/users/abcd/efgh”)

mydata = read.table(file = “allFound.csv”, header=T, sep= “,”)

help(“read.table”) #showing explanation of read.table

R Objects

-Data Frame : Values stored in rows and columns, Different columns may have different data modes.

Some useful functions

-str(obj) > summarize the structure of an object

-head(data), tail(data) > output the top/bottom rows

head(mydata, 10) #showing first to tenth row

tail(mydata, 10) #showing last to (last-10)th row

-class(obj) > get the object type, such as data frame

-names(data) > get the column names

-dim(data) > get the number of rows and columns

**Retrieve Data from Data Frame**

1.Retrieve your data by column index

mydata[,1]:retrieve the 1st column

or

Data$grade:retrieve column by column name

2.How to select specific variables

myvars = c[“v1”, “v2”, “v3”]

3.How to exclude specific variables

newdata = mydata[c(-3,-5)] #exclude 3rd and 5th variable #**minus(-)** means excluding.

4.How to select specific rows

newdata = subset[mydata, age >= 20 & sex ==”male”, select=c(ID, Weight)]

#select all rows that have male students with age no less than 20, We keep the ID and Weight columns

(53:00~)

nal = mydata$Nationality

**Install.packages(‘plyr’)**

library(plyr)

cf = count(nal)

we can get x, frequencies (cf)

label = cf$x

freq = cf$freq

crf = table(data$Nationality)/nrow(mydata) <<divided by how many rows in mydata

we can get (crf)

pie(crf)

barplot(crf)

barplot(freq, names.arg=label)

help(“barplot”)

grade = mydate$Grade

summary(grade) #we can get Min, q1,q2,q3, Mean, Max.

**Install.packages(‘psych’)**

library(psych)

describe(grade) #we can get more information than summary(grade) function.

hist(grade)

boxplot(subset, col=rainbow(ncol(subset)))