**STAT 40001/MA59800 Statistical Computing Fall 2017**

**Lab-5**

1) The *Duncan* data frame has 45 rows and 4 columns. Data on the prestige and other characteristics of 45 U. S. occupations in 1950. The data is in the library car*.*

a) Access the data.  
install.packages("car")  
library("car")  
data(Duncan)

b) Print first five observations of the data set.  
> head(Duncan,5)

type income education prestige

accountant prof 62 86 82

pilot prof 72 76 83

architect prof 75 92 90

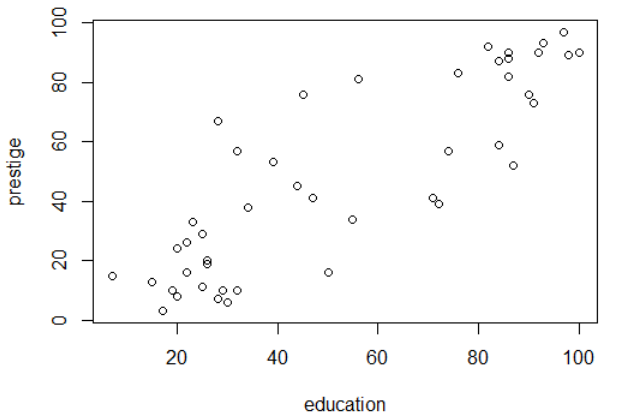
author prof 55 90 76

chemist prof 64 86 90

c) Use scatterplot to display the prestige scores according to the education level.  
> education = Duncan$education

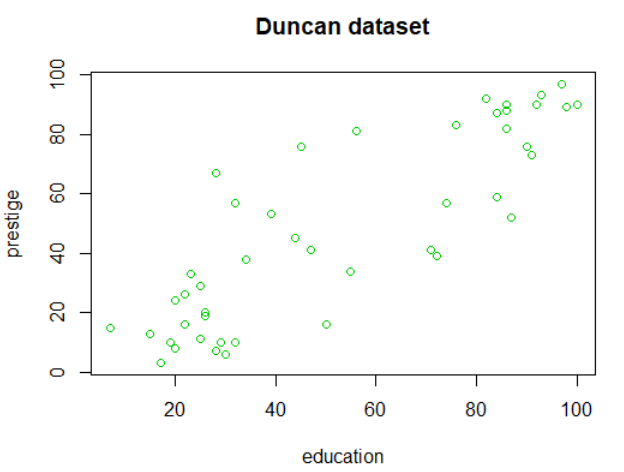
> prestige = Duncan$prestige

> plot(education,prestige)



d) Change the color, title, labels etc. and save it.

> plot(education,prestige,col='3',main='Duncan dataset',xlab = 'education',ylab = 'prestige')



2) The Davis data in the car package contains data on the measured and reported heights and weights of 200 men and women engaged in regular exercise.

a) Access the data.

> data(Davis)

b) A few of the data values are missing and are marked as “NA”. Clean the data by deleting the missing values.

> newdata = na.omit(Davis)

c) How many individuals do you have with complete information?

> nrow(newdata)

[1] 181

3) Link below provides a list of datasets related to economics (Data are from: principles of Econometrics)

<http://www.principlesofeconometrics.com/poe4/poe4stata.htm>

1. Import dataset entitled “savings” in R.

> data = read.dta("C:\\Users\\wu1114\\Desktop\\savings.dta");

> head(data,5)

savings income avgincome

1 2.412 83.830 65.917

2 2.473 68.147 64.553

3 4.594 84.205 71.658

4 3.893 84.016 64.584

5 3.816 52.269 60.696

1. What is the dimension of the data?

> dim(data)

[1] 50 3

1. Draw a histogram of the data related to the income. Please make sure to change the color, provide the title, labels etc.

> hist(data$income,col=c(1,2,3,4,5,6),xlab="income",ylab='value',main = 'Income Distribution')

