Lab 5

Alexander Hernandez

09/08/22

1) The Duncan df has 45 wors and 4 cols of data about US occupations in 1950.

a) Access the data (library 'car')

```
library(car)
```

```
## Loading required package: carData
```

```
data(Duncan)
```

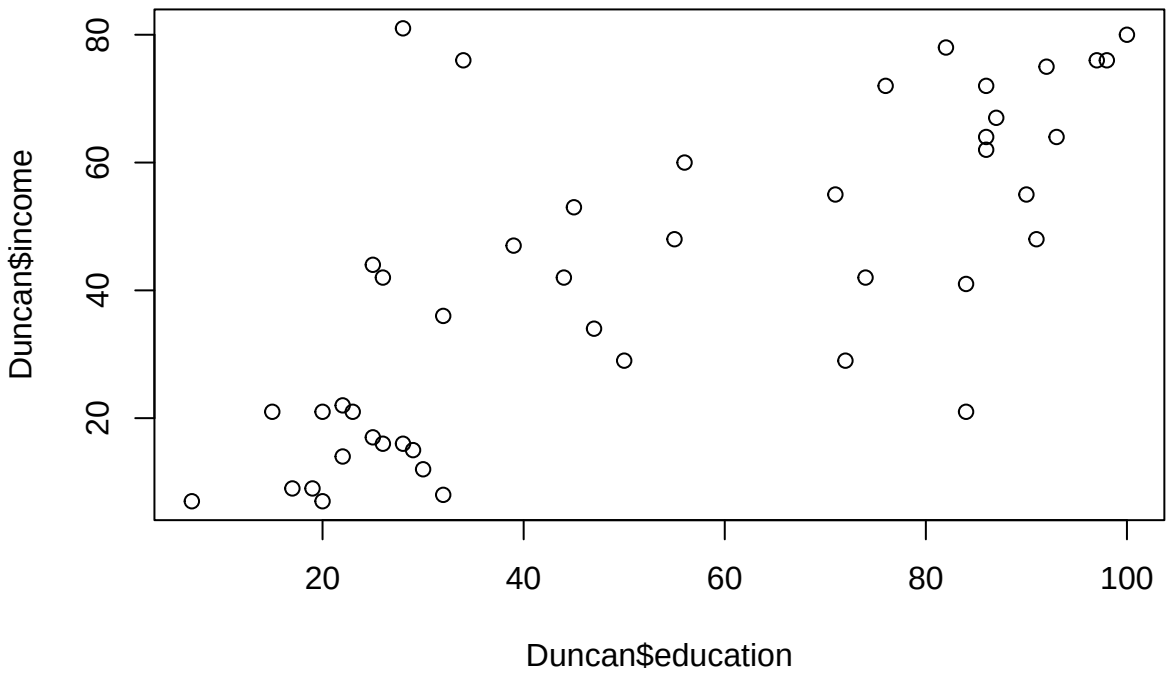
b) Print the first five obs. of the data set

```
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
```

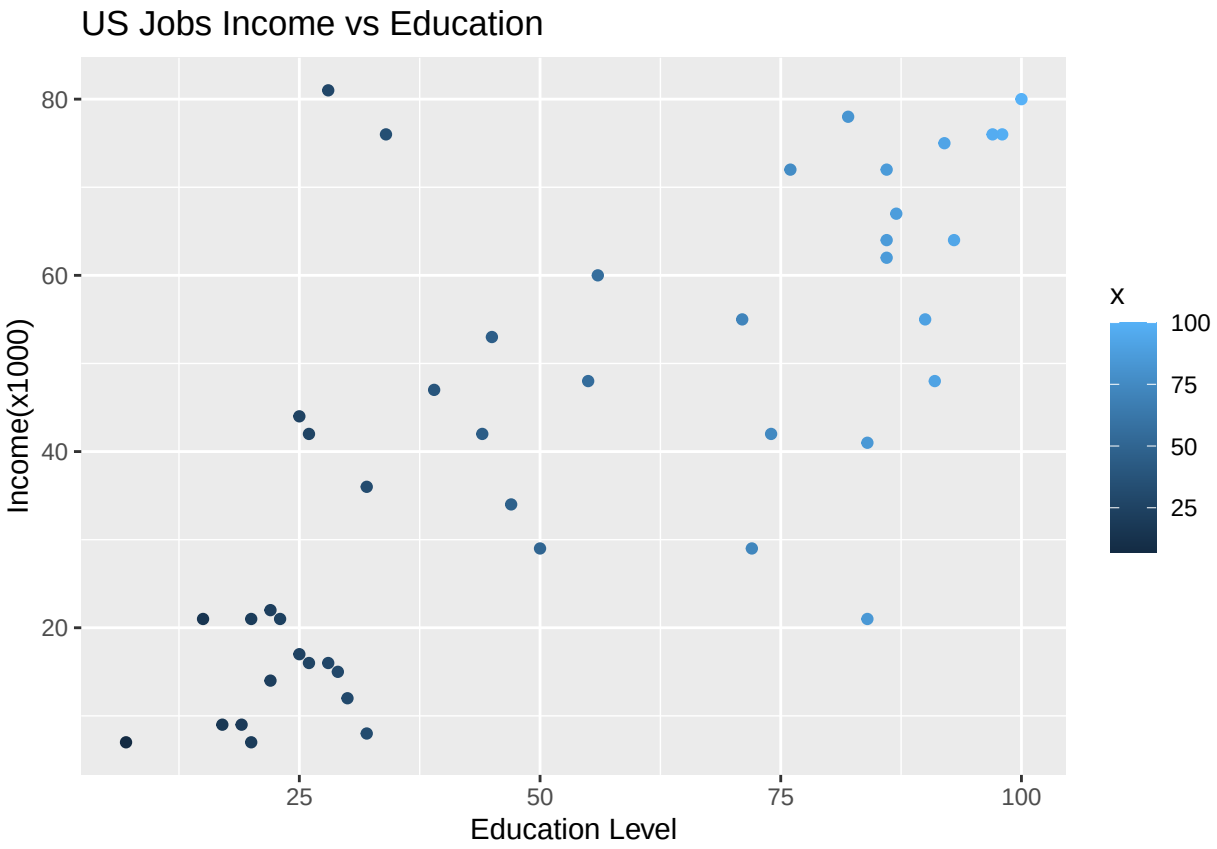
c) Use scatterplot to display the prestige scores according to education level

```
library(ggplot2)
plot(Duncan$education, Duncan$income)
```



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

```
df1 = data.frame(x= Duncan$education, y = Duncan$income)
attach(df1)
ggplot(df1, aes(x, y)) +
  geom_point(aes(color=x)) +
  labs(title="US Jobs Income vs Education", x="Education Level", y="Income(x1000)")
```



2) The Davis data contains measured and reported height+weight of 200 men and women who exercise

a) Access the data (library 'car')

```
library(car)
data(Davis)
```

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

```
NewDavis = na.omit(Davis)
```

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

```
nrow(NewDavis)
```

```
## [1] 181
```

3) Access the 'Elections' from 'mdsr' package and extract variable names

```
library(mdsr)
data(Elections)
names(Elections)
```

```
##  [1] "Ward"
##  [2] "Precinct"
##  [3] "Registered Voters at 7am"
##  [4] "Voters Registering at Polls"
##  [5] "Voters Registering by Absentee"
##  [6] "Total Registrations"
##  [7] "Voters at Polls"
##  [8] "Absentee Voters"
##  [9] "Total Ballots Cast"
## [10] "Total Turnout"
## [11] "Percentage Absentee"
## [12] "% Registered to Total (Election Day)"
## [13] "Spoiled Ballots"
```

4) Use the link to obtain economic related datasets

http://www.principlesofeconometrics.com/poe4/poe4stata.htm http://www.principlesofeconometrics.com /poe4/data/stata/savings.dta ## a) Access the dataset 'savings'

```
library(readstata13)
savings_file = file.choose()
savings = read.dta13(savings_file)
```

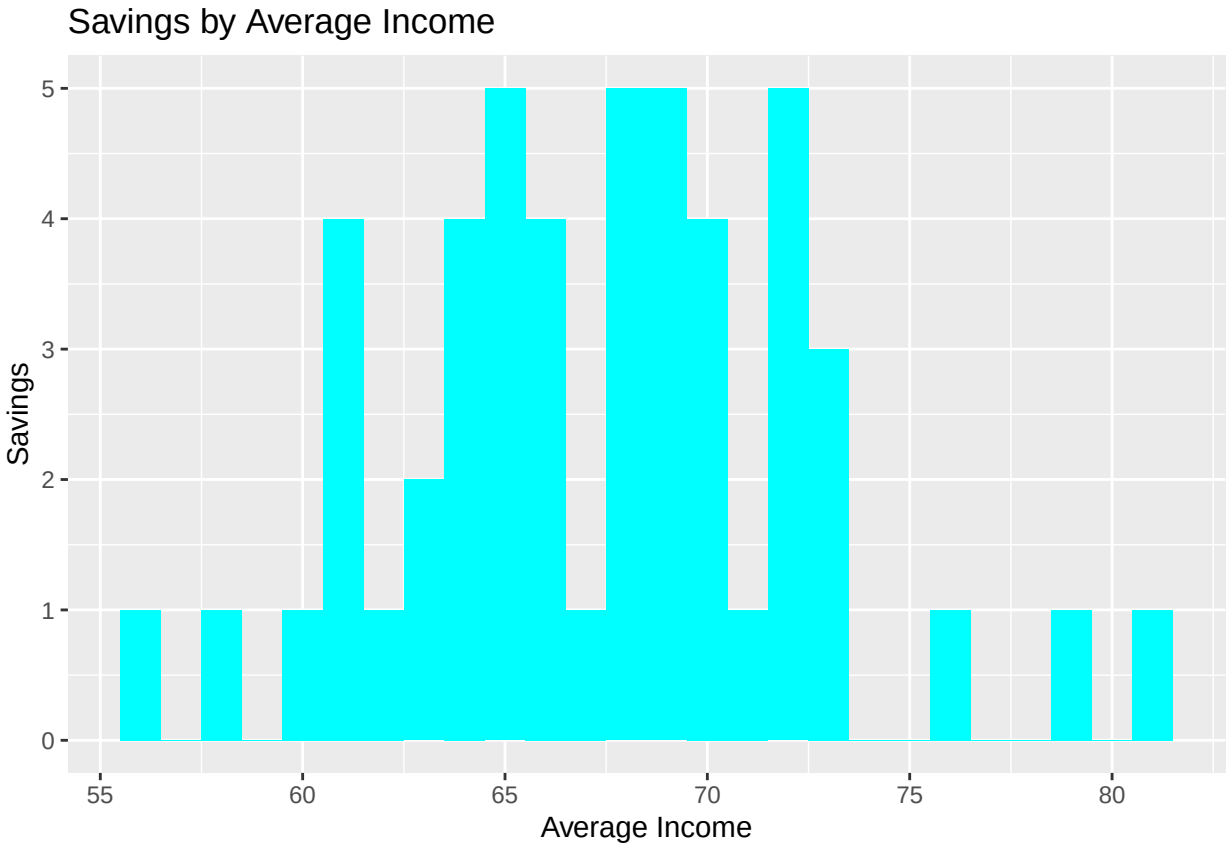
b) What are the dimensions of the data?

```
dim(savings)
```

```
## [1] 50  3
```

c) Draw a histogram of the data related to the income. Change color, title, labels, etcs.

```
library(ggplot2)
df2 = data.frame(x= savings$avgincome, y = savings$savings)
ggplot(df2, aes(x)) +
  geom_histogram(binwidth=1, fill="cyan") +
  labs(title="Savings by Average Income", x="Average Income", y="Savings")
```


Savings by Average Income
Savings
Average Income
0
1
2
3
4
5
55
60
65
70
75
80