CE202 R Lesson - Import Excel and Linear Regression

Yiğit Çetinel
Spring 2017

How to import Excel file into R

1. Use "Import Dataset" under Environment tab
2. Click "From Excel"
3. Click "Browse" and select "europe.xlsx" from where you locate it
4. Change name as "europe" and click "Import"

View the data:

```
View(europe)
```

Check the column names:

```
names(europe)
```

```
## [1] "Name"                          "Area (km²)"
## [3] "Population"                    "Population density (per km²)"
```

Second one and fourth one look complicated, so we can change the column names to simplify:

```
names(europe)[2] = "Area"
names(europe)[4] = "Density"
```

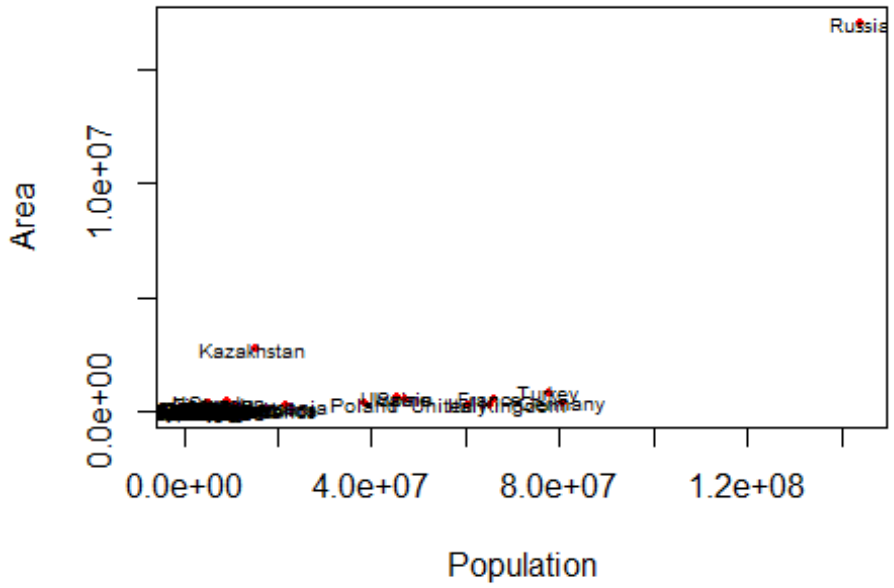
Check the head of the data now:

```
head(europe)
```

```
##         Name   Area Population Density
## 1    Albania  28748    2831741    98.5
## 2    Andorra    468      68403   146.2
## 3    Armenia  29743    3060631   101.5
## 4    Austria  83858    8169929    97.4
## 5 Azerbaijan  86600    9165000   105.8
## 6    Belarus 207560    9458000    45.6
```

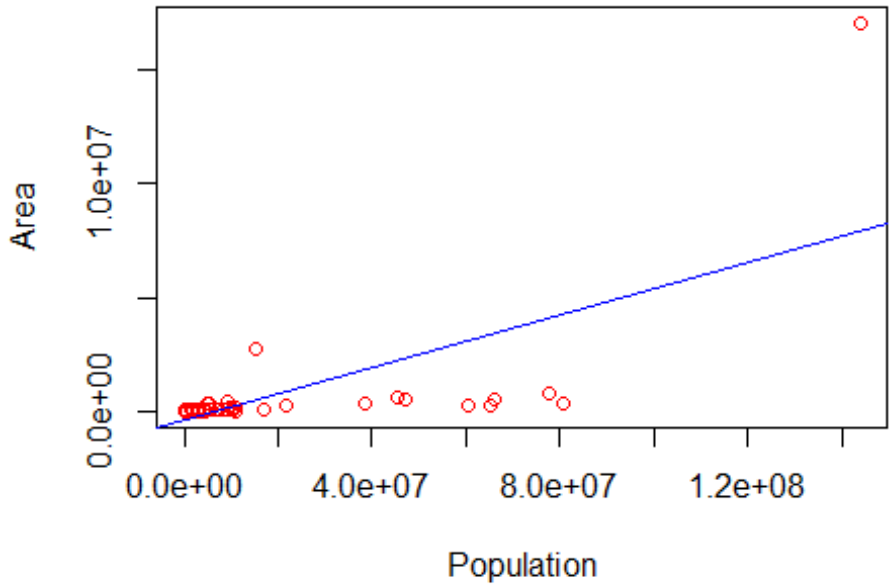
Let us plot Area vs Population of the Europe data:

```
plot(Area ~ Population, europe, col="red")
text(Area ~ Population, europe, labels=Name)
```



Linear regression using Area and Population:

```
model = lm(Area ~ Population, europe)
plot(Area ~ Population, europe, col="red")
abline(model, col="blue")
```



The model is not good because of the last data (Russia). Let us try again without Russia.

```
europeWithoutRussia = subset(europe, Name!="Russia")
model = lm(Area ~ Population, europeWithoutRussia)
plot(Area ~ Population, europeWithoutRussia, col="red",pch=16,cex=.6)
text(Area ~ Population, europeWithoutRussia, labels=Name,cex=.6)
abline(model, col="blue")
```