Exercise 1

Considering the following table, represent graphically the two variables, compute the coefficient of correlation among the two variables and provide an interpretation of the results (both manually and with R):

firm id	1	2	3	4	5	6
wage (X)	45	30	84	63	62	61
n. employees (Y)	14	16	46	32	22	21

Exercise 2 with R

From R, load the dataset USAarrest. Hint: run the following codes:

```
library(datasets)
library(help = "datasets")

data(USArrests)
str(USArrests)

# represent Murder and UrbanPop with a graphical represent,
# from a graphical representation, is it possible to see if a relationship exists?
# compute coefficient of correlation among the two variables.
```

Exercise 6 with R

From R, load the dataset Arthritis. Hint: run the following codes:

```
## Load vcd package
library(vcd)

## Load Arthritis dataset (data frame)
data(Arthritis)

str(Arthritis)
summary(Arthritis)

#see the variables included in the data
#prepare a two table with variables Sex and Improved
#compute chi2 and Cramer-V and provide an interpretation of the results obtained.
```

Solutions

Exercise 1

A) Scatter plot
B) Correlation

Stipendio x	n. Dip y	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
45	14	-12.5	-11.17	139.58	156.25	124.69
30	16	-27.5	-9.17	252.08	756.25	84.03
84	46	26.5	20.83	552.08	702.25	434.03
63	32	5.5	6.83	37.58	30.25	46.69
62	22	4.5	-3.17	-14.25	20.25	10.03
61	21	3.5	-4.17	-14.58	12.25	17.36
345	151	0	0	952.50	1677.5	716.83

$$\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1677.5}{6} = 279.58 \rightarrow \sigma_x = 16,72$$

$$\sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{716.83}{6} = 119.47 \rightarrow \sigma_y = 10,93$$

$$\sigma_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{952.50}{6} = 158,75$$

$$\rho_{xy} = \frac{158,75}{16,72 * 10,93} = 0,8686$$

$$\overline{stipendi} = \frac{1}{n}\sum_{1=1}^{n} i = \frac{45 + 30 + \cdots + 61}{6} = 57,5$$

$$\overline{dipendenti} = \frac{1}{n}\sum_{1=1}^{n} i = \frac{14 + 16 + \cdots +}{6} = 25,17$$

poi le varianze e la covarianza

$$\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad \sigma_Y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 \qquad \sigma_{XY} = \sum_{i=1}^{n}\frac{1}{n}(x_i - \bar{x})(y_i - \bar{y})$$