Biostatistics

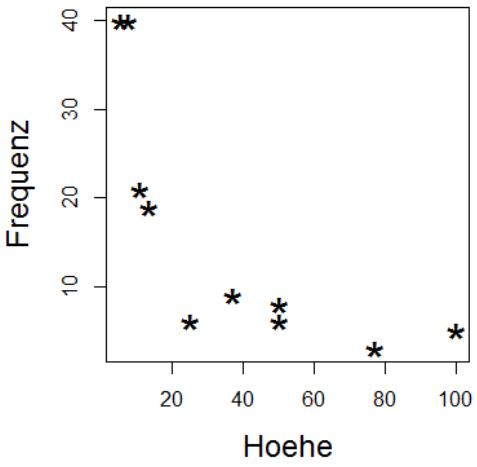
In-class-exercise week 8
Topic: Correlation & Simple Regression

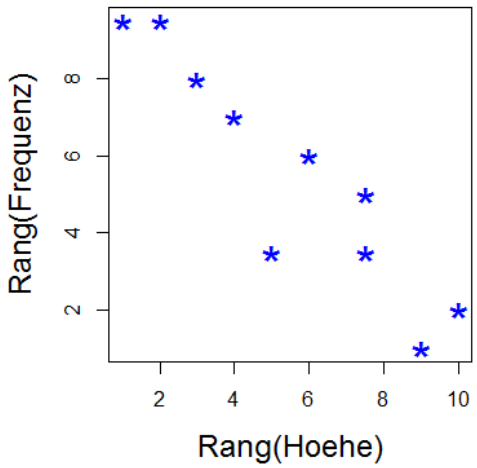
Problem 1) In a science article (Science, 164 (1969), p.1513) a study was presented that investigated the association between the hight of a waterfall and the frequence of vibrations in the ground.

Name	h: Hoehe	f: Frequenz	Rang(f)	Rang(h)
Lower.Yellowstone	100	5	2	10
Yosemite	77	3	1	9
Canadian.Niagara	50	6	3.5	7.5
American.Niagara	50	8	5	7.5
Upper.Yellowstone	37	9	6	6
Lower.Gullfoss	25	6	3.5	5
Firehole	13.3	19	7	4
Godafoss	10.9	21	8	3
Upper.Gullfoss	7.7	40	9.5	2
Fort.Greeley	5.2	40	9.5	1

The following graph shows a scatterplot of frequence vs hight.

a) Describe the form of the association.
The association is monotonically decreasing. The relationship is not linear but looks more like an exponential decay.





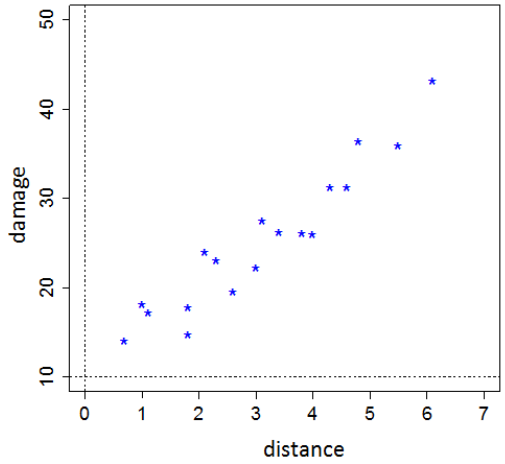
b) Is the Pearson correlation here appropriate? Give reasons.

No, the Pearson correlation is not appropriate, since the relationship is not linear.

c) Determine in the table the missing ranks and visualize the relationships as scatter plot.

Problem 2

The damage of cells due to inflammation depend on the distance to the source (data made up).



We have used R to modeled the data with a linear regression::

```
Call:
lm(formula = damage ~ dist, data = fire)

Residuals:
    Min      1Q  Median      3Q     Max
-4.1689 -2.1322 -0.1006  2.1287  3.5850

Coefficients:
            Estimate    Std. Error  t value    Pr(>|t|)
(Intercept)   10.292      1.408      7.312    1.75e-06 ***
dist           4.820      0.406     11.874    2.40e-09 ***
---
Residual standard error: 2.637 on 16 degrees of freedom
```

- What is the estimate for the intercept, what is the estimate for the slope?

estimated intercept =10.9
estimated slope=4.8

- Write down the estimated linear relationship describing how the damage depend on the distance.

$\hat{y} = 10.9 + 4.8 \cdot x$ with y=damage, x=distance

- Draw the corresponding line in the scatterplot.
- What damage does your model predict for the distance 5? Mark the prediction in the plot.

$\hat{y}(5) = 10.9 + 4.8 \cdot 5 = 34.5$

- What does "16 degrees of freedom" tell you?

Df=#data-#parameters(in lin model). Since we estimate 2 parameters (intercept and slope) and have df=16, we have 18 observations.