Data Visualization in R

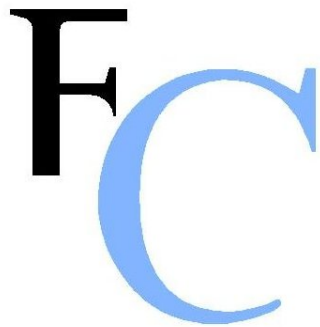
Solutions to Hands On Exercises

L. Torgo

ltorgo@fc.up.pt

Departamento de Ciência de Computadores / Faculdade de Ciências
Universidade do Porto

Mar, 2021



FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Hands on Data Visualization - the Algae data set

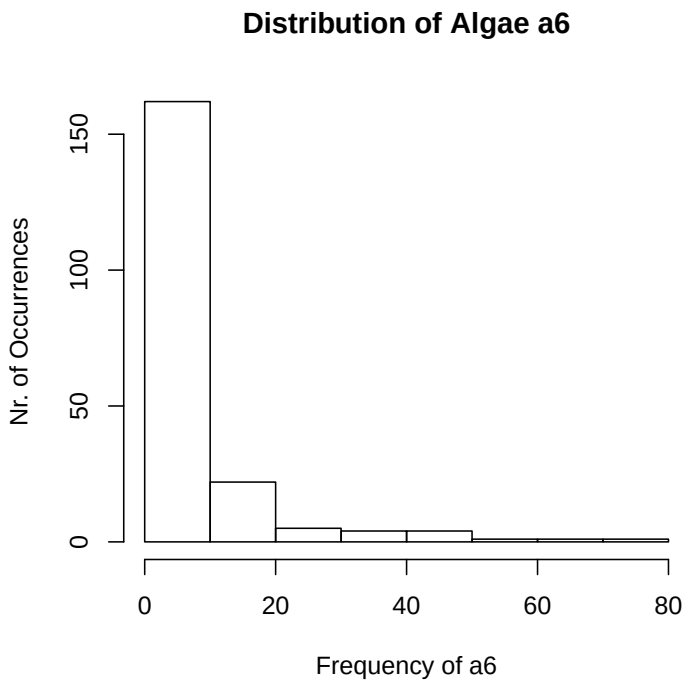
Using the Algae data set from package `DMwR` answer to the following questions:

1. Create a graph that you find adequate to show the distribution of the values of algae `a6` solution
2. Show the distribution of the values of `size` solution
3. Check visually if it is plausible to consider that `oPO4` follows a normal distribution solution

Solution of Exercise 1

- Create a graph that you find adequate to show the distribution of the values of algae `a6`

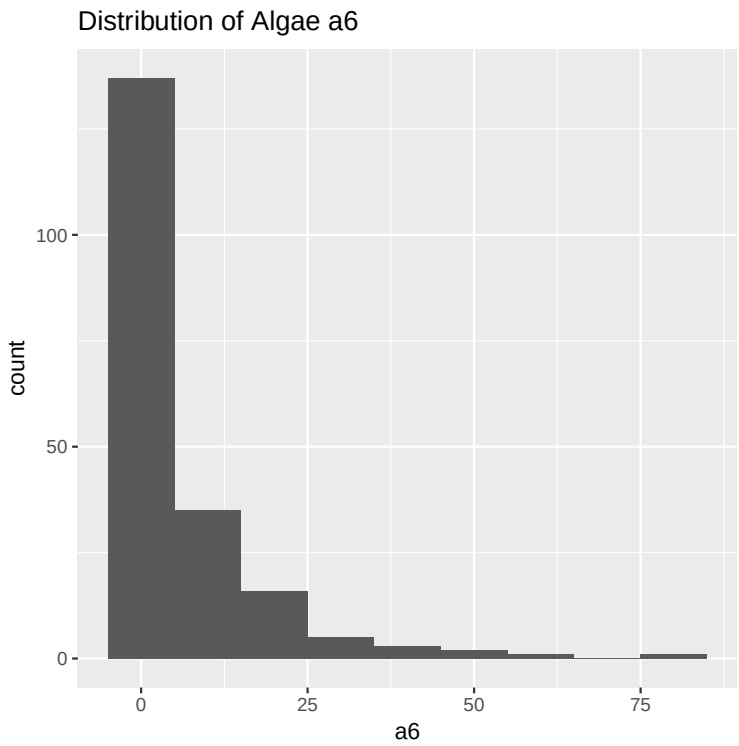
```
hist(algae$a6,main="Distribution of Algae a6",xlab="Frequency of a6",
     ylab="Nr. of Occurrences")
```



Solution of Exercise 1 with ggplot2

- Create a graph that you find adequate to show the distribution of the values of algae a6

```
ggplot(algae,aes(x=a6)) + geom_histogram(binwidth=10) +
    ggtitle("Distribution of Algae a6")
```

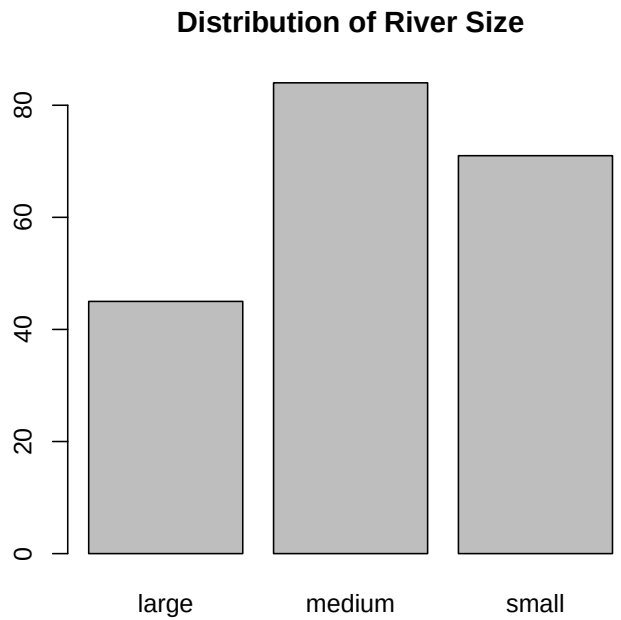


Go Back

Solution to Exercise 2

- Show the distribution of the values of `size`

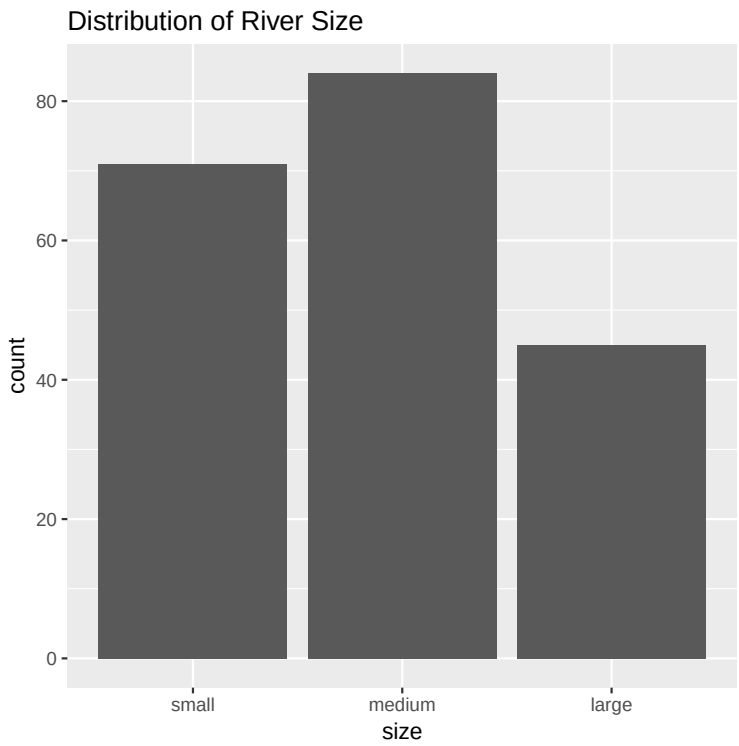
```
barplot(table(algae$size), main = "Distribution of River Size")
```



Solution to Exercise 2 with ggplot2

- Show the distribution of the values of `size`

```
algae$size <- factor(algae$size,levels=c("small", "medium", "large"))
ggplot(algae,aes(x=size)) + geom_bar() +
    ggtitle("Distribution of River Size")
```

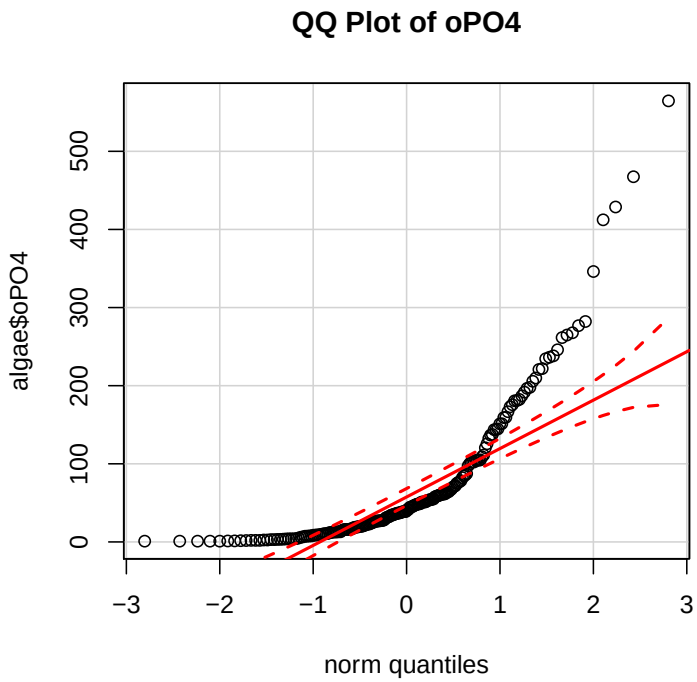


Go Back

Solutions to Exercise 3

- Check visually if it is plausible to consider that `oPO4` follows a normal distribution

```
library(car)
qqPlot(algae$oPO4,main = "QQ Plot of oPO4")
```



Hands on Data Visualization - Algae data set

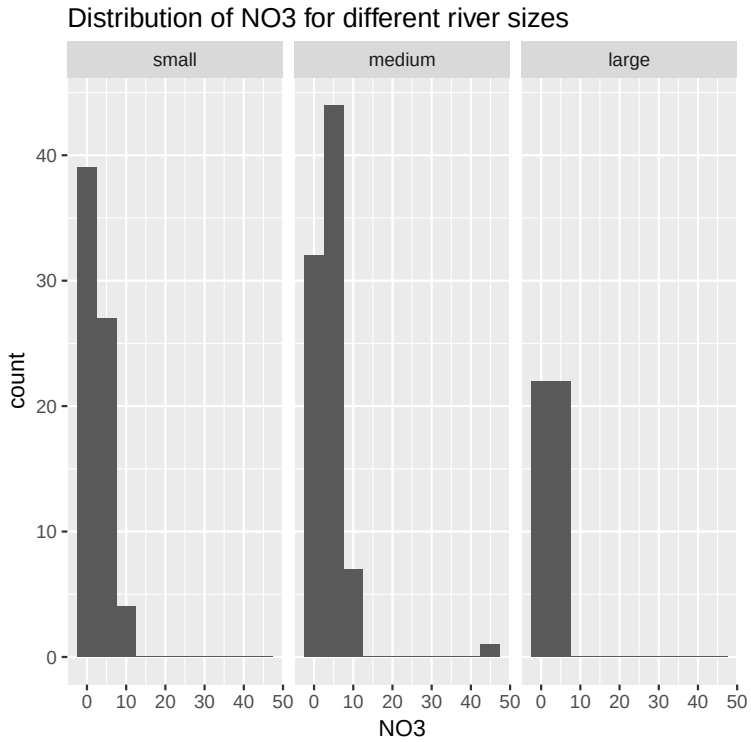
Using the Algae data set from package `DMwR` answer to the following questions:

1. Produce a graph that allows you to understand how the values of NO3 are distributed across the sizes of river solution
2. Try to understand (using a graph) if the distribution of algae a1 varies with the speed of the river solution

Solutions to Exercise 1

- Produce a graph that allows you to understand how the values of NO3 are distributed across the sizes of river.

```
algae$size <- factor(algae$size,levels=c("small", "medium", "large"))
ggplot(algae,aes(x=NO3)) + geom_histogram(binwidth=5) + facet_wrap(~size)
    ggtitle("Distribution of NO3 for different river sizes")
```



Solutions to Exercise 2

- Try to understand (using a graph) if the distribution of algae a1 varies with the speed of the river

```
algae$speed <- factor(algae$speed,levels=c("low", "medium", "high"))
ggplot(algae,aes(x=speed,y=a1)) + geom_boxplot() +
    ggtitle("Distribution of a1 for different River Speeds")
```