

# Data Visualization in R

## Solutions to Hands On Exercises

L. Torgo

ltorgo@dal.ca

Faculty of Computer Science / Institute for Big Data Analytics  
Dalhousie University

May, 2021



# Hands on Data Visualization - the Algae data set

Using the Algae data set from package `DMwR2` (extra package you need to install) answer to the following questions:

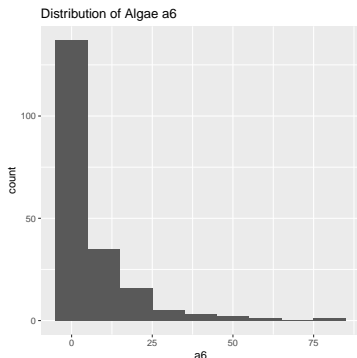
- 1 Create a graph that you find adequate to show the distribution of the values of algae `a6` solution
- 2 Show the distribution of the values of `size` solution
- 3 Check visually if it is plausible to consider that `oPO4` follows a normal distribution solution



# Solution of Exercise 1

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

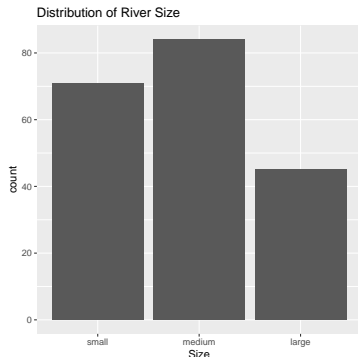
```
ggplot(data = algae, mapping = aes(x=a6)) +  
  geom_histogram(binwidth=10) +  
  ggtitle("Distribution of Algae a6")
```



## Solution to Exercise 2

- Show the distribution of the values of `size`

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



## Solutions to Exercise 3

- Check visually if it is plausible to consider that  $\text{oPO}_4$  follows a normal distribution

```
ggplot(algae, aes(sample=oPO4)) +  
  stat_qq() + stat_qq_line()
```

