ANALYZING POKÉMON

Setting up a project

Open RStudio, press the small arrow in the upper right corner to create a new project and use a fitting name (or via File -> New Project...). Put your data in the folder you created for the project. Think about what could be a useful file-structure.

R Markdown

Create a new R markdown file. Click yes if you get pop-up windows to install required packages. In R Markdown you can combine code, results and explanatory text into a reproducible, dynamic report. You can use it to create slideshows, pdfs, html documents, Word files and more.
Take a look at the R Markdown cheat sheet (click).
Type along with the code in this exercise. At some points you will have to come up with your own code. Using google is allowed and recommended.

Dataset ## Import First, try to import your data. Create a code block and use the following function. (Run a line by pressing cntr/cmnd + enter or execute a full block by pressing cntr/cmnd + shift + enter)

```r
filelocation <- "Data_Pokemon.csv" #Enter the location of the data of your computer here.
fdata <- read.csv2(filelocation)
```

Data impression The function creates a data frame from the file in your memory. Take a look at your data, two useful functions here are dim() and head().

1. What does the data "look" like, what are the dimensions of the table you created?

2. Did you read it correctly? If not, what should you change? (Hint: use ?read.csv2 and take a look at the raw data in notepad. Hint: the parameter sep)
3. What "class" is each column?

4. Is the data a matrix or a data frame? Why? Is you are unsure, review the matrix and data frame links in the practical description.

You have read the file correctly and can continue when you get the following result:

```r
dim(fdata)
```

```
## [1] 721   8
```

#Distribution of the data We'll focus on the "Attack" and "Defense" columns of the table.
Have a (very brief) look at the distribution of these two variables.

1. Does there seem to be a normal distribution of both variables? Hint: use hist(). Hint: Remember what a normal distribution is.

2. Is there a transformation you can do to get a normal distribution?

We continu with the untransformed data.

#Data visualization

We will use the regular plot() function of R.

1. Make a scatterplot with "Defense" on the x-axis and "Attack" on the y-axis using the plot() function.

- Add a main header, an x and y axis label (hint: main, xlab, ylab)
- choose nice a plot character and size (hint: pch, cex)
- choose a nice color (hint: col)

2. There are hunderds of Pokémon, but not so many Types. How many types are there? Hint: levels(). Advanced Hint: length(levels())

3. What is the most and least common Type? Hint: table(). Advanced Hint: data.frame(table()).

4. Describe the shape of the cloud underneath the plot, does there seem to be a trend? Are there any outliers?

5. Calculate the correlation coefficient between Defense and Attack. Hint: cor(). Advanced Hint: cor.test()

6. Color the points of one of the Types (choose yourself) a different color Hint: points(). Hint: fdata$Defense[fdata$Type=="Water"]

7. Add three other Types with another color. Hint: points()

8. Add a legend. Hint: legend()

9. "Knit" your document. Does it still work?

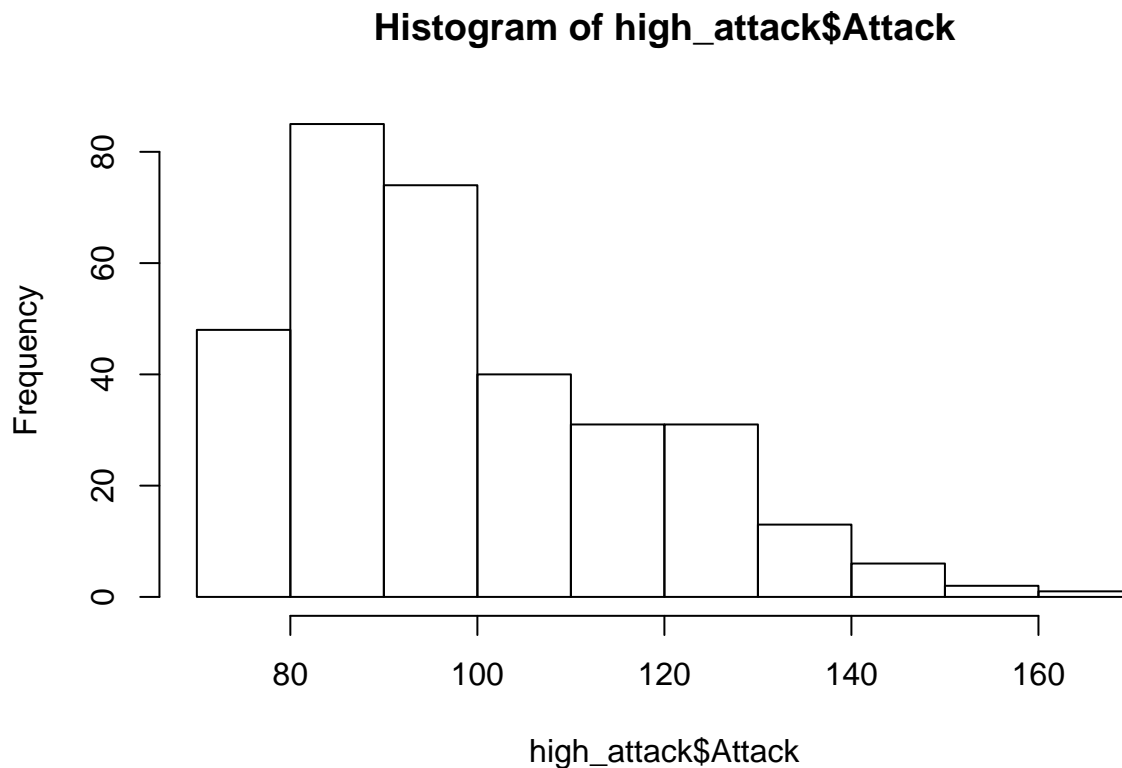Your final result shoul look something like this.
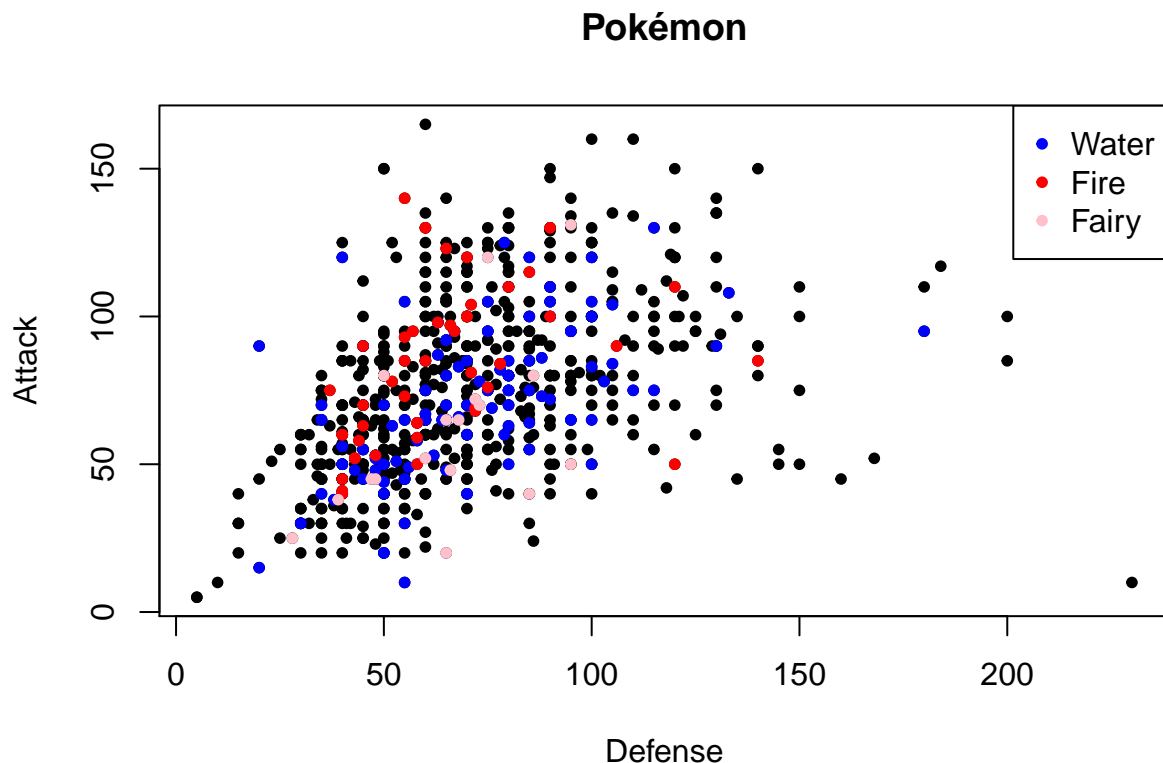
#Statistical test ##Subset We want to select part of our data: specific pokemon (rows), which have a certain condition. We can use which() to subset data in R, as folows:

```
high_attack <- fdata[which(fdata$Attack > 75), ]
dim(high_attack)
```

```
## [1] 331    8
```

```
hist(high_attack$Attack)
```

Histogram of high_attack$Attack

1. Subset the data based on the type of pokemon. We want to compare "Water" pokemon to "Fire" pokemon, so we want two subsets: water and fire.

##Statistical test Now we want to compare "Attack" points between "Water" and "Fire" Pokemon. Let's assume the data does not follow a normal distribution.
1. Which statistical test should you use? Hint: wilcox.test() Hint: compare t.test() 2. Is there a significant difference between "Water" and "Fire" Pokemon?
3. Make a boxplot of Attack points between "Water" and "Fire" Pokemon. Hint: boxplot(). Add x and y labels.
4. Knit your script.

Now compare in the same way Sp.Attack between Fighting and Electric Pokemon.

#EXTRA for the ambitious
Could you write a function to compare two variables and make a boxplot? Try to make this function in a

different .R file and source it in your R Markdown file.
Hint 1: take a look at R for Dummies on how to source.
Hint 2: check the function function()