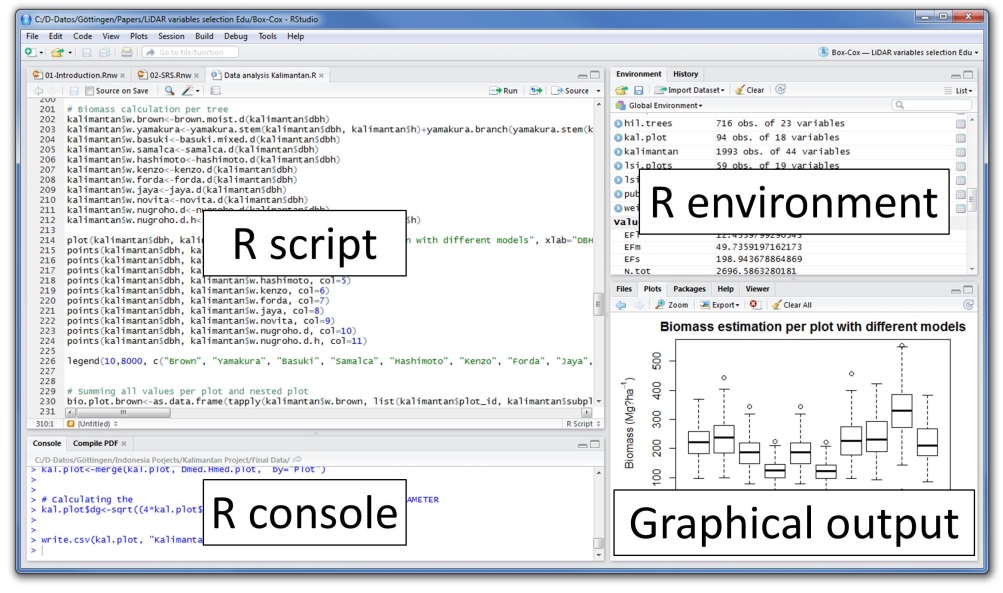
**Assignment 1 – Descriptive Statistics and Descriptive Spatial Statistics**

**Part 1 – Descriptive Statistics in R**

**Tool used: R studio.** I recommend the use of R studio, because it provides a good interface and supports tasks like installation of packages: 

**Dataset used**: USArrests from package “datasets”. To get access to the data, first install and then load the package: installation can be done in the packages tab of R studio (lower right window); once you installed the package, you can load it with the command *library(datasets).* (Check which packages are loaded with *search()*. To list the data coming with the package datasets enter *data()*). Metadata about the dataset with *help(USArrests)*. You may want to put the dataset in a new variable for not altering the original dataset (remember that the assignment of data to a variable work with an arrow <-).

1. **Histograms and density plots**
2. Paste the following code into R and explain what the single commands are producing. Consult the help pages of R for further information on the single commands:

x <- rnorm(100)

y <- seq(-4,4,length.out=200)

hist(x,freq=F,ylim=c(0,0.5))

lines(density(x),col="red",lwd=2)

lines(y,dnorm(y),col="blue",lwd=2)

1. Alter the code of exercise 2a) such that you create a density histogram with an overlaid density graph of one variable in USArrests. Explore what the parameter *breaks* does when plotting a histogram.
2. **Exploring Plots**
3. For the data USArrests prepare boxplots for comparing the distributions of variables in the USArrests dataset. Does it make sense to plot several boxplots in one figure using the command similar to boxplot(USArrests[,1], USArrests[,2], col=4) or for the whole dataframe with boxplot(USArrests)? Would including the scale command help boxplot(scale(USArrests))? What does scale() do? Chose colors for your plots, select meaningful axis labels, a title etc.

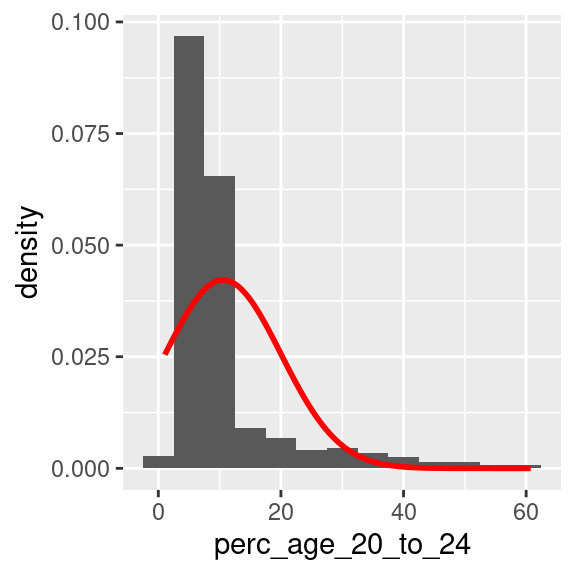
Comment: You can also look at distributions of values in a stem plot: e.g., stem(USArrests[,1])

1. Go to <http://www.statmethods.net/graphs/bar.html> for an overview of “simple graphics” in R. Do a dot chart for one variable of USArrests. Then do a dotchart after ordering the values for one variable for USArrests using the command order(). Chose colors for your plots, select meaningful axis labels, a title etc.
2. **Bivariate Analysis**
3. Calculate correlation coefficients for the USArrests dataset and do a scatter plot using the commands *cor(…)* and *pairs(…)* . Which variables have the highest correlation? Is the correlation positive or negative?

**4) Descriptive Statistics with different functions**

Please go to section 6.5 of Stefan de Sabbata’s book on introducing R granolarr: <https://sdesabbata.github.io/granolarr/practicals/bookdown/exploratory-data-analysis.html>. This section of Stefano’s book contains a histogram of a variable overlaid with a normal distribution. Please try to reproduce the diagram shown with the USArrests data. Working with Stefano’s instructions might require a quick read of introductory sections of the granolarr book.

Here the specific graphic that I am referring to:



**Assignment Due date and submission**

Please summarize the answers in a short report including plots and submit it via Blackboard until **April 16, 2021.**