

W03-1: Correlation

This worksheet helps you in identifying correlations between variables in a data frame. After completing this worksheet you should know how to compute a simple correlation between two variables (in general two columns of a data frame).

Things you need for this worksheet

- R — the interpreter can be installed on any operation system. For Linux, you should use the r-cran packages supplied for your Linux distribution. If you use Ubuntu, [this](#) is one of many starting points. If you use windows, you could install R from the official [CRAN](#) web page.
- R Studio — we recommend to use R Studio for (interactive) programming with R. You can download R Studio from the official [web page](#).
- your script and data from [W02-1: Reading CSV files](#)

Learning log assignments

🚨 First things first: the following analysis is build on top of your script from [W02-1](#). Please copy your script “W02-1.R”, rename the copy to “W03-1.R” and use it for the programming tasks of this worksheet.

After reading a data set for analysis and a first overview of the data set, one could start with looking into some dependencies between the individual variables. Since some types of correlation analysis require linear dependencies, it is always a good idea to visualize the data first.

Let's have a closer look on animal activity and vegetation coverage.

😞 Please visualize the relationship between animal activity and vegetation using the `plot()` function.

Look at the visualization and try to interpret the relationship.

😞 In order to get a quantitative indicator for the relationship, please compute an appropriate correlation using the `cor.test()` function.

How does the correlation result fit to your interpretation of the visual relationship? Is the correlation result reliable?

While a correlation analysis between animal activity and vegetation cover might be always worth a shot, a more structured approach in searching for correlations might be more appropriate if you analyze a data set. So what about getting not only one but all correlations.

😞 Please compute the correlation of all individual combinations of two variables using the `cor()` function. Check the help of the function on how to handle NA values which are missing values (i.e. animal activity is only available on 50 of the plots and hence is missing on all others).

Have a look at the result - confusing, right? So let's visualize it before we look for the correlations.

😞 Please visualize the pairwise correlations using the `corrplot()` function.

Have a look at the visualization - now we are talking! Please interpret the correlations and come up with theoretically guided explanations for two or three of the observed correlations.

From: <http://bisfogo.environmentalinformatics-marburg.de/> - **BIS-Fogo**

Permanent link: <http://bisfogo.environmentalinformatics-marburg.de/doku.php?id=en:learning:schools:s01:worksheets:ba-ws-03-1>

Last update: **2015/09/22 15:22**

