



# Cheat sheet

by Henrik von Wehrden

Main basics to find your way into R, serving as a starting point. Continue to use until you know all commands by heart, and add ad lib

Start R, check for relics, locate harddrive locations, pre-format data properly get install & load libraries you need, make sure that your data is in the right format.

## Setup your R environment

<code>install.packages</code>	install new libraries
<code>library</code>	load a library that is installed
<code>getwd</code>	locate your working directory
<code>setwd</code>	set your the working directory
<code>dir()</code>	check directory for loaded files
<code>objects()</code>	check workspace for objects
<code>rm(list=ls())</code>	clean the workspace

## Data types

integer  
numeric  
ordinal  
factor  
character  
logical

## Conditions

<code>x&lt;-1</code>	x gets 1
<code>x==y</code>	x equals y
<code>x&gt;y</code>	x larger y
<code>x!=y</code>	x unequal y

## Get help

`?help`  
`help.search`

Read your data in, consider checking for outliers or data consistency, clean data.

## Loading data

<code>read.table</code>	Load txt table
<code>read.csv</code>	Load a csv file
<code>read_excel</code>	Loads Excel file
<code>attach</code>	Attach data
<code>detach</code>	Detach data

## Subselection of data

<code>x[lines, columns]</code>	Where to subselect
<code>x[, c(2:3)]</code>	Select column 2 & 3
<code>x[c(1, 4), ]</code>	Select rows 1 & 4
<code>x[x==5]</code>	Elements that equal 5
<code>x[x&gt;3]</code>	Elements larger than 3

First inspect the data inside out. Check the structure, overall summary, the first few lines, data distribution for continuous data and counts for factors. Based on the data and your hypothesis, use the appropriate test or model.

## Data inspection

<code>dim</code>	Dimension of dataframe
<code>str</code>	Structure of the dataset
<code>summary</code>	Summary of data
<code>head</code>	Inspect top rows of table
<code>length</code>	length of a vector
<code>hist</code>	Visualize data distribution
<code>table</code>	Get count of factor levels

## Simple tests

<code>chisq.test</code>	structure of the dataset
<code>var.test</code>	summary of main info
<code>t.test</code>	inspect top rows of table

## Analysis of variance

<code>table(factor)</code>	Check factor counts
<code>aov</code>	Make balanced Anova
<code>boxplot</code>	Make a boxplot
<code>TukeyHSD</code>	Perform posthoc test
<code>pairwise.t.test</code>	Compare all levels

## Simple calculations

<code>min</code>	Minimum value
<code>max</code>	Maximum value
<code>mean</code>	Mean value
<code>median</code>	Median value
<code>var</code>	Variance
<code>sd</code>	Standard deviation
<code>round</code>	Round values
<code>log</code>	Natural log
<code>exp</code>	Exponential

## Correlation & Regression

<code>cor</code>	correlation coefficient
<code>cor.test</code>	Correlation test
<code>model&lt;-lm(x~y)</code>	Make a linear model
<code>hist(resid(model))</code>	Check residuals
<code>plot(x~y)</code>	Make a x-y plot
<code>abline(model)</code>	Plot regression line

Make sure to flip the data upside down, visualize the data, make sure to contextualize the model and to be critical of your potential mistakes. Knowledge is not experience.