

# Programming and Analyzing in R 1

Wouter van Atteveldt <http://vanatteveldt.com>

May 2016

# Introduction

- Who are you? :-)
- What are you working on?
- What do you think R will do for you?
- What experience do you have with
  - R
  - Other statistical packages
  - Programming (inc. SPSS syntax, Stata .do)
- Are you taking this course for credit?

# Course components

- Lectures
- Hands-on sessions
- For credits:
  - Assignments

# Course overview

## ① Today

- Getting started
- Your data in R
- Lab: Assignment 1

## ② Tuesday

- Merging and transforming data
- Classical statistics and visualization
- Lab: Assignment 2

## ③ Wednesday

- Advanced statistics and programming

## ④ Thursday

- Texts and networks

# What is R?

- Open source
- Multi-platform
- Text/Console
- Community driven
  - Packages

# Cathedral vs. Bazaar



# R vs SPSS

- Syntax mode only
- Full featured programming language
- Tons of extensions

# The R ecosystem

- R
- RStudio
- Packages



# Interactive sessions

- Slides: <http://vanatteveldt.com/>
- Handouts and data:  
<http://github.com/vanatteveldxt/learningr>
- Please interrupt me!

# Using R

- R as a calculator
- Reading in data
- Data frames vs. vectors
- Simple plots

# First hands-on session

- `1_r_calculator.md`
- <http://www.openintro.org/stat/labs.php>, "Intro to R and RStudio"

# Reading/writing data

- R: `save`, `load(file)`
- CSV: `read.csv`, `read.csv2`, `write.csv`
- SPSS: + `read.spss(to.data.frame=T)`
  - `library(foreign)`

# Inspecting data

- Inspecting a data frame
  - `names`
  - `summary`
  - `head` and `tail`
  - RStudio: `View`
- Inspecting a column (vector)
  - `summary`
  - `length`
  - `mean`, `sum`, etc

# Data manipulation: subsets

- `data[rows, columns]`
- Leave out either argument:
  - `data[rows, ]`
  - `data[, columns]`
- Use index, names, or calculation
  - `data[5, ]`
  - `data[, c(1,2,3)]`
  - `data[, c("x", "y")]`
  - `data[, -1]`
  - `data[data$x > 3, ]`

# Data manipulation: calculated columns

- Assignment to part of a data frame
- `mydata$c = mydata$a + mydata$b`
- `mydata$c[mydata$a > 3] =`  
`mydata$c[mydata$a > 3] + 1`

# Good coding practice

- Code is written once, read many times
- Some good practices:
  - ① Descriptive variable names
  - ② Use comments to explain decisions
  - ③ Clean up code when you're done
  - ④ Write self-contained scripts
    - Start with reading data, end with analysis
    - Test by clearing all variables, running script



# Hands on 1b

- `2_playing.md`
- `http://www.openintro.org/stat/labs.php,  
"Introduction to data"`

# Assignment 1

The goal of the first substantive assignment is to get started on using R for your own research.

- Choose a topic or problem that you are currently working on or that you previously solved using a different statistics package.
- Write down what the data is; what question you want to “ask” the data (hypotheses, research questions, etc); and what analyses you will need to get the answers.
- Write an R script to read your data into R. Explore the shape and contents of your data. Is there an internal structure? How many data frames do you have? What data types do the columns have?
- Make a number of tables to show your data and explain what they mean.