# R intro – dual workshops february 2019

**Content / themes covered**

* What is R?
* The RStudio environment
* The R language – defining objects
* Types
* Classes
* Types of objects (vectors, lists)
* Functions: forstå funktioner og opret funktioner
* Den indbyggede hjælpefunktion
* While & for loops
* If-statements
* Dataframes: Opret, subset, funktioner
* Factors
* Indlæs filer – forskellige formater
* Basale visualiseringer (nævn EDA og table data workshop)
* Tekstdata: indlæsning og basale funktioner (demo til text mining workshop)
* Eksporter data
* Stack exchange (og internettet generelt)
* Programmeringshovedpiner: NA, NULL, empty

**Structure**

* Inspiration fra “hit the ground running”
  + Noget, der inkorporer næsten alle elementer, som det er meningen, at de to workshops skal dække
* Fra bunden: RStudio og R sproget
* Resten i rækkefølge som ovenstående…
* Slut af med tekstdata, webscraping-eksempel (API)

**Exercises**

* Definer objekter og tjek klasse/type
* Skriv en funktion
* Brug et for loop
* Brug et if-statement (integrer egen funktion)
* Indlæs en fil som dataframe
* Tilføj række til dataframe
* Udled info fra dataframe
* Udfør basal visualisering

**Packages**

Blot nok til at illustrere, hvad pakker er for noget, og til at vise nogle ting

**To-do inden dag 2**

* Upload til CALDISS\_commons
* Installer pakker på server: haven, ggplot2, tm, stringr, wordcloud, pdftools

**Dag 2**

* Exhibition script
* Working directory (changing it manually)
* Pakker / libraries
  + install.packages(‘haven’)
  + install.packages(‘ggplot2’)
  + library(haven)
  + library(ggplo2)
  + library(stringr)
* haven
  + load data
  + explore
    - summary
    - head / tail
    - unique
    - table
    - mean, min, max
    - print\_labels
* exercise: creating age variable
* ggplot2
  + geom\_histogram
  + geom\_point
    - jitter
* if-conditions
  + Subsetting data
  + Error control
* Functions
  + Creating functions
  + Using if in functions
  + Print
* List
  + Creating a list
  + Subsetting a list
* For loops
  + Combined with if
* ggplot2:
  + geom\_smooth
  + facetwrap
* text, stringr, tm
  + grep, grepl
  + gsub
  + paste
  + str\_split
  + str\_to\_lower
  + str\_to\_upper

**Exercises**

* 1 (efter haven)
  + Loading the data
  + Creating age variable
  + Are the participants in the UK older than the ES?
* 2 (efter ggplot2)
  + Create histogram (age)
  + Create scatterplot (INSTtrst and age)
* 3 (efter if og function)
  + Create object storing the average height in the data
  + Create two subsets: 1 containing only DK respondents, 1 with GB
  + Create a function that can be used to check if the mean height of a country is above or below the mean for the whole dataset
  + Check your function with your two subsets (DK and GB) – is the mean above or below the data mean?
* 4 (efter list og for)
  + Create a list containing a subset for each country
  + Create a histogram for weight for each country using your new list and a for loop
  + (sum up by showing facet\_wrap)
* 5 (efter tekst)
  + Using what you have learned, explore what party seems to have the voters with the most trust in people (ppltrst)