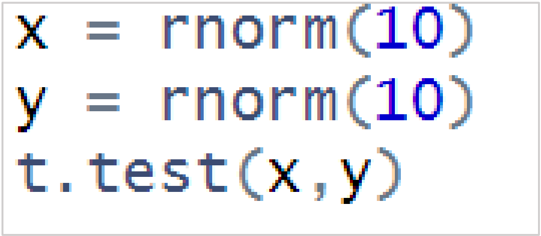
R-intro:
The very introductionary introduction

Sine Zambach 2021

Researcher at CBS and author of ”Kvinde kend din kode”

Ph.d. in AI

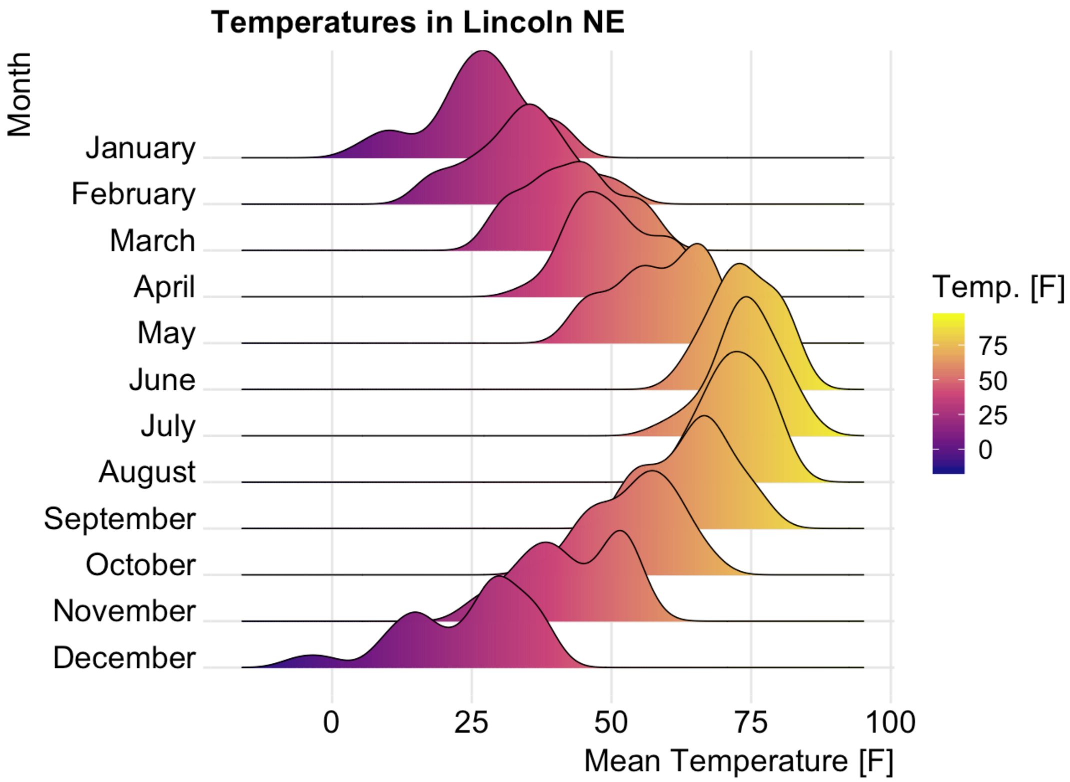
Why R?



```
x = rnorm(10)
y = rnorm(10)
t.test(x,y)
```

- Great stat-program
- Open source
- Excel can do up to regression

- R vs python:
  - If you do stat – it is elegant
  - First learning steps are normally easier
  - Visualizations!
  - Rstudio – a really nice environment



Plan

After session 1, you will be able to run R, open files and do some simple data-wrangling.

- Today
  - Input file formats
  - Calculate stuff in R
  - Data Frames

After session 2, you will be able to wrangle even more data and have the basics for modelling and statistical tests for you to go on with more sophisticated R-work.

- Next time
  - A bit more on data frames
  - Introduction to modelling (linear models, t-test, fisher's test)
  - Visualization in R (R is great for visualization! We try out histogram, barplot, scatter plot, dot-and-whiskers plot)

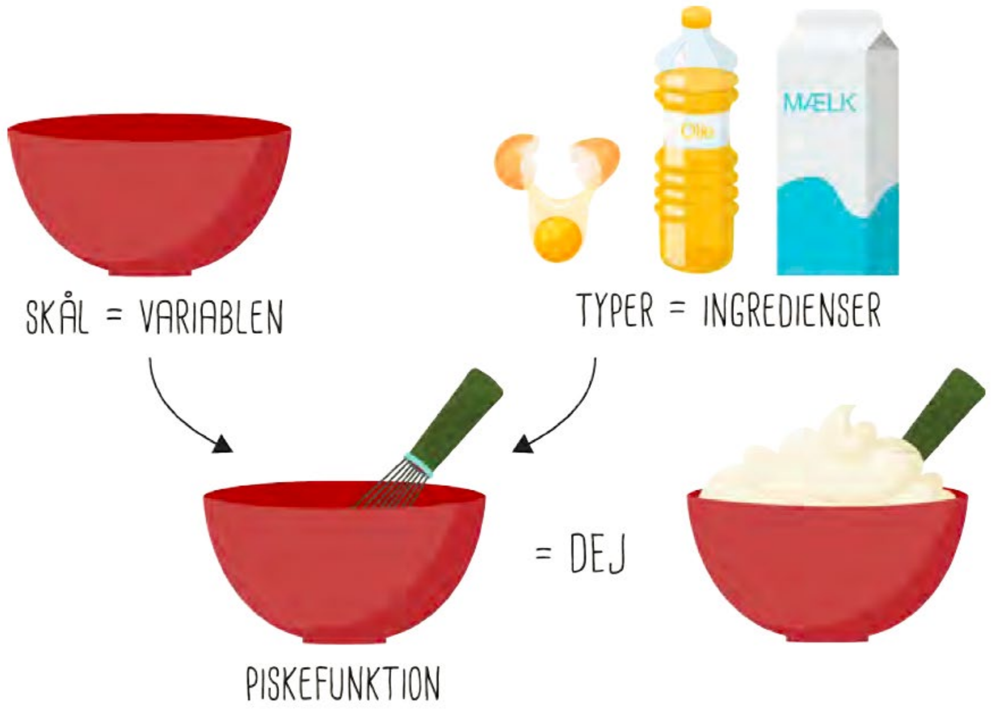
Get started

- Watch the two videos
- I'll demonstrate how to use Rstudio, load in a file and how to run code

- Use other resources
  - R for data science: https://r4ds.had.co.nz/
  - Danish: Gahners R-guide: http://erikgahner.dk/Rguide/
  - Cheat sheets: e.g.
    - https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf
    - https://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf

Data syntax

- Data structures
- Data formats
- Functions



Variable, datatyper og funktioner

Data syntax – think of grammar

Variables

- Atomic variables `x <- 3`
- Vectors `v <- c(3,5,7,9)`
- Data frames
  ```
  df <- matrix(c(1,3,5,2,”h”,”w”),
                       nrow = 2)
  ```

Data types

- Data can be structured as *numeric, characters, factors*
- If a wrong format: `v <- as.character(v)`
  - Or `as.numeric()` or `as.factor()`

Calculating stuff within R

Change format of variable

- Add, minus, multiplication, etc (+,-,*,/)
- Square root: srt()
- Logarithm: log(), log10(), log2()
- Potens: x^2

Questions-session

Exercise

- Try to create
  - a single variable
  - a vector
  - a dataframe (or read in a tabular file)

- Change data type of variable (using `as.character()`, `as.numeric(), etc)`
- Do some calculations

Functions

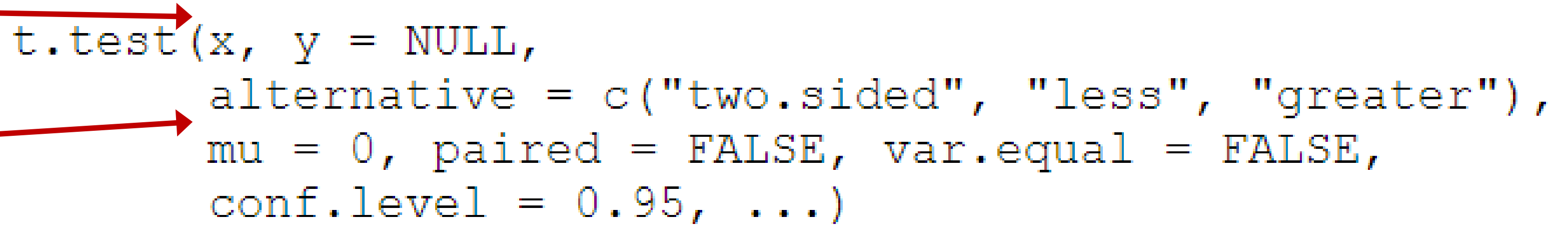
- A predefined operation that does something to your data
- You can also built your own function (not today…) or borrow one from the internet

Examples

- as.numeric() is a function
- c() is a function
- read.excel() Is a function

Functions have arguments

- The data
- Settings



```
t.test(x, y = NULL,
       alternative = c("two.sided", "less", "greater"),
       mu = 0, paired = FALSE, var.equal = FALSE,
       conf.level = 0.95, ...)
```

Questions-session

Exercise

Looking at the function t.test()

```
x = rnorm(10)
y = rnorm(10)
t.test(x,y)
```

- Try to add `5` to `x`
- Try to do it one-sided (e.g. `"less",` meaning H_0 is that y is smaller than x, and `"greater“`, the other way around)

A bit more on data frames

- `df <- matrix(c(1,3,5,2,”h”,”w”), nrow = 2)`
- Read files – needs a library: `library(readxl)`
- Inspect dataframe:
  - `View()`
  - `str()`
  - `summary()`

Accessing data frames

- df[rows,columns]
- df$columns[row]

- The same:
  - `df[,1]`
  - `df[,”X1”]`
  - `df$X1`

Questions-session

After session 2, you will be able to wrangle even more data and have the basics for modelling and statistical tests for you to go on with more sophisticated R-work.

- Next time
    - A bit more on data frames
    - Introduction to modelling (linear models, t-test, fisher's test)
    - Visualization in R (R is great for visualization! We try out histogram, barplot, scatter plot, dot-and-whiskers plot)