

[home](#)[vacatures](#)[opleidingen](#)[interim](#)[contact](#)

## opleidingen



### R BASICS

locatie: amersfoort | 31 mei, 7 juni, 14 juni en 21 juni 2017  
Hieronder vallen de volgende onderdelen:

- Data preparatie en manipulatie
- Data visualisatie
- Predictive modeling



### Data Driven Marketing

locatie: amersfoort | 2017



knowledge in the **ATTIC.**  
powered by jansen en van der zee

### contact

Onze marketing intelligence opleidingen zijn samengesteld op basis van de vraag uit de markt. Staat er echter geen passende opleiding voor je bij maar heb je wel een specifieke behoefte? Neem dan [contact](#) op met jansen en van der zee.

Wij bespreken graag de mogelijkheden

**R** **BASICS**  
Longhow Lam

# Agenda

- Avond sessie 1
  - Achtergrond en Overzicht R
  - De RStudio omgeving
  - R projecten, Data types & structures
- Avond sessie 2
  - Data preparatie en manipulatie
  - De tidyverse packages
- Avond sessie 3
  - Data visualisatie
  - Legacy plot functies
  - ggplot2
  - Interactive graphs
- Avond sessie 4
  - Machine learning
  - Predictive modeling in R

Laten we vast het volgende downloaden

<https://github.com/longhowlam/inleidingR>

# Iets over mijzelf

- MSc Mathematics (1995) Vrije Universiteit Amsterdam (drs. wiskunde)
- MTD Applied Statistics (1997) Technical University Delft (twee jarige AIO toegepaste statistiek)
- ✓ Veel ervaring met SAS (Base / Stat / Guide/ Miner / VA / VS)
- ✓ Veel ervaring met R
- ✓ Verschillende functies als data scientist
  - RTL -- Data scientist
  - ABNAMRO -- Risk modeler
  - Business&Decision -- Quantitative consultant
  - Experian -- data miner

Ben nu ZZP'er, als je een nogelijk project hebt:  
Contacteer me gewoon. Ik kom zeker een kop koffie drinken

Follow me:  @longhowlam



LinkedIn posts



Wordpress blog

# Achtergrond / Overzicht R

# Achtergrond / overzicht van R

1978 Ontstaan van de S programmeertaal (AT&T Bell Labs)

1985 StatSci maakte daar commercieel product S-PLUS van

1995 Ross Ihaka and Robert Gentleman at the University of Auckland □ R

“The R foundation” gevestigd aan de *Wirtschaftsuniversität Wien*  
*bestaande uit*

- *Board*
- *Core members (25 leden die de sourcecode onderhouden)*
- *Donors, supporting members and benefactors: Waaronder SHELL.*

R valt onder GNU general public license, onder deze voorwaarden:

- Binaries van R vrij te downloaden en her-verspreiden
- Complete source code is vrij te downloaden, modificeren en her-distribueren

# Achtergrond / overzicht van R



- 2015 opgericht
- Aantal bedrijven die R ondersteunen
- Sponseren projecten die "R verder helpen"

## Platinum Members



## Gold Members



## Silver Members



# De verschillende R software componenten

# Een aantal verschillende R software componenten



R Core / Base



R Studio (server)

R Shiny (server)

R Packages



Scalable  
Machine learning

SparkR/ SparklyR

SQL Server R Services

Meer dan 10000 packages op CRAN  
zeer handige packages zijn:

- dplyr data preparatie / wrangling
- stringr manipulatie van strings
- lubridate voor datum manipulatie
- httr, xml werken met web APIs
- rodbc connectie naar databases
- ggplot2 grafische plot functies
- plotly grafische plot functies
- arules market basket analyse
- arulesVis mba visualisaties

Zie verder ook: [deze link](#)



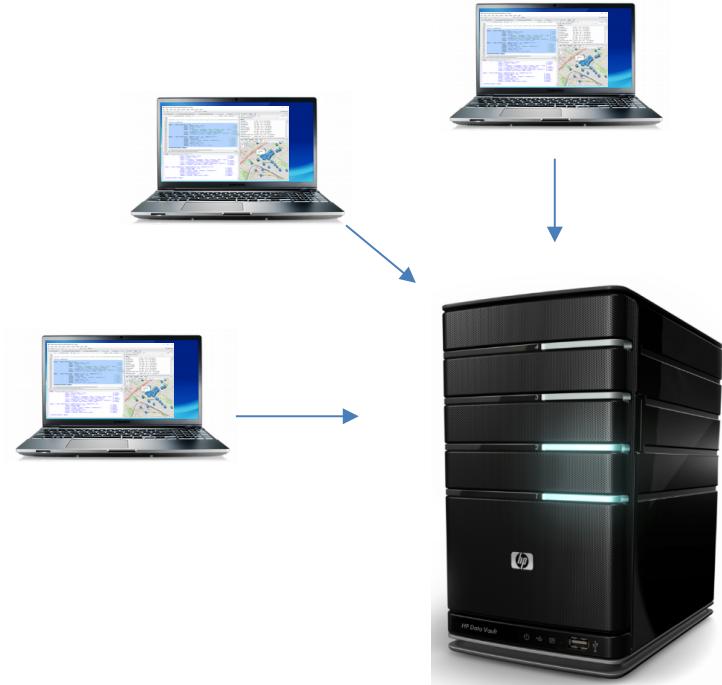
Microsoft®  
SQL Server®

# Architecture R

Desktop /  
laptop



Single  
Server



Spark distributed compute cluster



# Waarom R gebruiken?

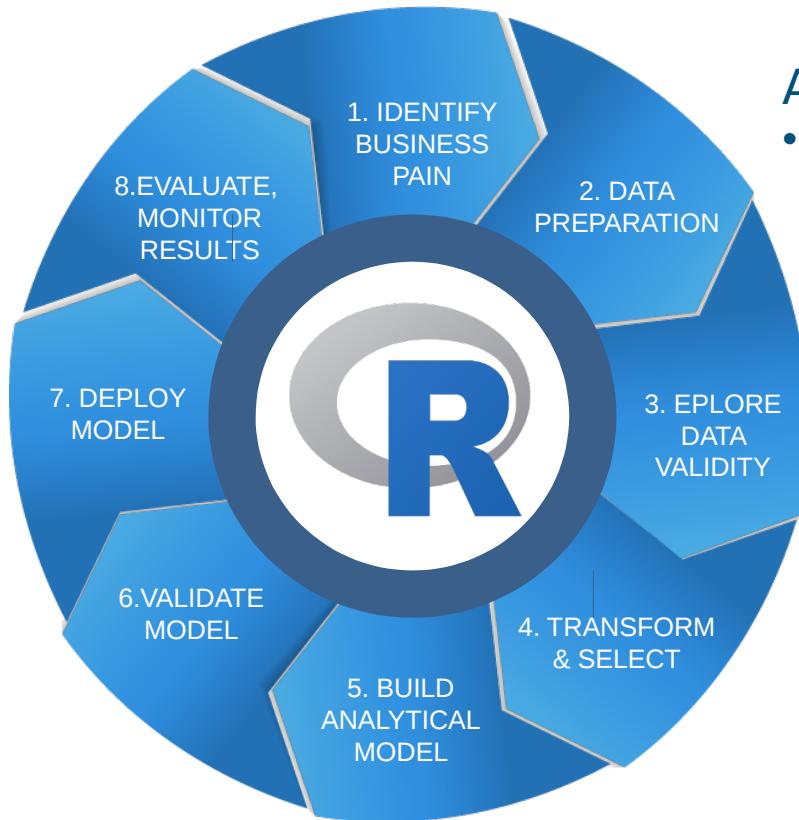
- ✓ Populariteit enorm toegenomen
  - Veel Universiteiten gebruiken het in het onderwijs
  - De tool van de nieuwe generatie data scientist
- ✓ Open source, geen licentie aanschaf
- ✓ Support via community (hele grote, stack overflow, blogs, online!!) of derde partijen.
- ✓ Heel veel actieve ontwikkelingen
  - Community
  - Bedrijven
- Je moet na denken!!!



Er is een leercurve..... maar die is te overwinnen!!!!

# The analytics life cycle

Out of-scope



Avond sessie 1&2

- Tidyverse, dplyr, stringr, etc.

Avond sessie 3

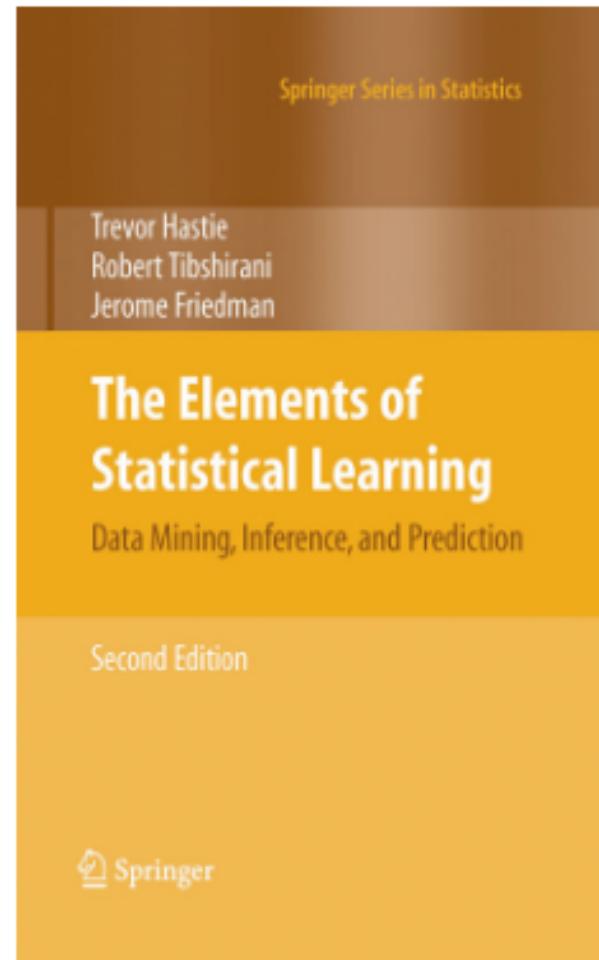
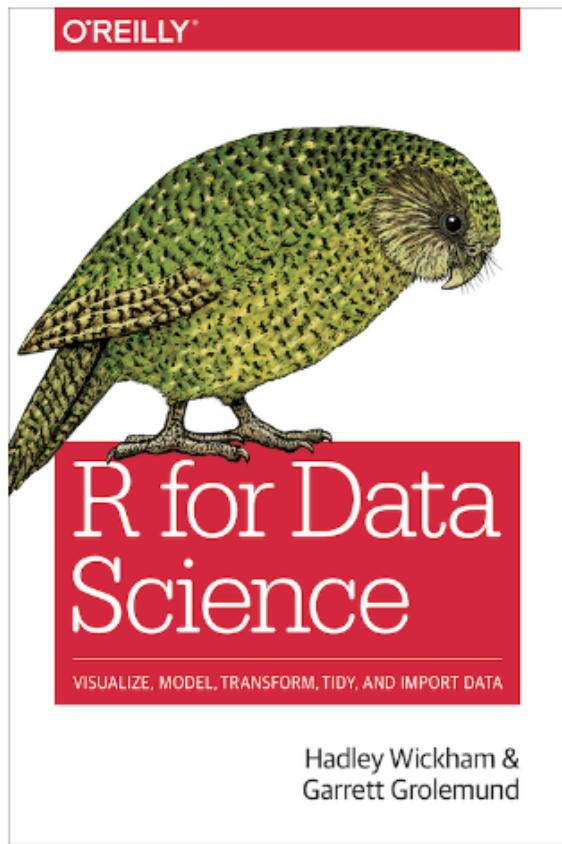
- Ggplot2
- Plotly
- Interactive plots

Avond sessie 4

- Lm, glm, h2o, glmnet, ranger, etc.

# Twee boeken die ik aanraad

<http://r4ds.had.co.nz/>



[http://statweb.stanford.edu/~tibs/ElemStatLearn/printings/ESLII\\_print10.pdf](http://statweb.stanford.edu/~tibs/ElemStatLearn/printings/ESLII_print10.pdf)



data  
iku

