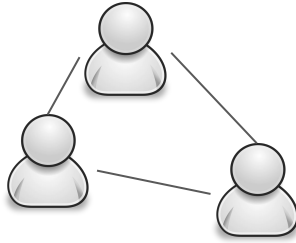


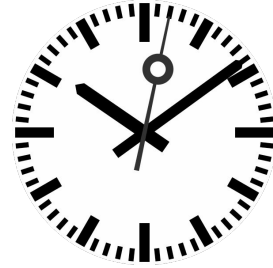
We hebben een
clubke opgericht...

Dirk, Joost, Hans, Damiano & Stijn

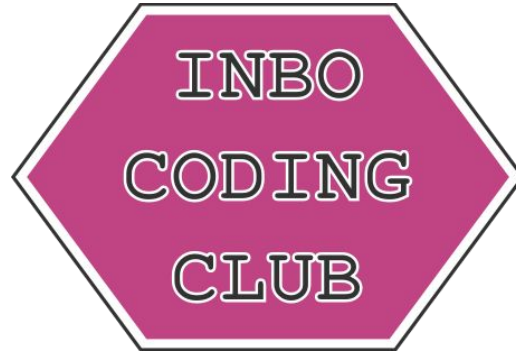
...omdat we zelf willen
bijleren
over code schrijven
(in R)!



a supportive
self-learning
community



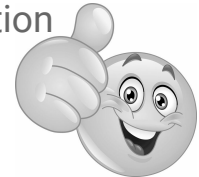
dedicated time to
practice and learn



connect with a
global research
community



turn
fear and anxiety
into motivation



Vandaag

1. Kennismaking van de R noden en in-house talenten
2. Aan de slag met data import

⊙ CHANGED AN HOUR AGO

FREELY

INBO CODING CLUB

Manual of this questionnaire

I feel confident working with the pac.

I feel confident when coding...

When using R, I'm struggling with...

I'm wondering if, using R, I would be

Expand all

Back to top

Go to bottom

INBO CODING CLUB

23 January, 2018

Manual of this questionnaire

For each of the following sections/questions add your answer to the list. In case your answer is already enlisted by someone else, do NOT add the answer as a new item, but provide a * sign before the answer.

As an **example**, suppose the question: *What's your favorite Simpson character?* and the already enlisted answers are:

- * Bart
- * Lisa

and your answer would be *Bart* as well; the result after your edit will be:

- ** Bart

Ga naar de
vragenlijst...

<https://hackmd.io/s/rk6vH53NM>



Click on of these to edit

🕒 CHANGED 5 DAYS AGO

INBO CODING CLUB

23 january, 2018

Manual of this questionnaire

For each of the following sections/questions, add already enlisted by someone else, do NOT add the before the answer.

As an **example**, suppose the question: *What's your enlisted answers are:*

* Bart




DATA
IMPORT

Het



concept

We bepaalden een aantal tussentijdse outputs. Als je zelf een output hebt bereikt, voeg dan een  toe aan je laptopscherm.

Het doel is dat **iedereen**  behaalt

- iemand met meer  dan jij? Vraag hulp!
- iemand met minder  dan jij? Geef hulp!

Dit [cheatsheet](#) kan je zeker helpen!

Data Import :: CHEAT SHEET

R's **tidyverse** is built around **tidy data** stored in **tibbles**, which are enhanced data frames.

The front side of this sheet shows how to read text files into R with **readr**.

The reverse side shows how to create tibbles with **tibble** and to layout tidy data with **tidyr**.

OTHER TYPES OF DATA
Try one of the following packages to import other types of files

- **haven** - SPSS, Stata, and SAS files
- **readxl** - excel files (.xls and .xlsx)
- **DBI** - databases
- **jsonlite** - json
- **xml2** - XML
- **httr** - Web APIs
- **rvest** - HTML (Web Scraping)

Save Data
Save *x*, an R object, to *path*, a file path, as:

Comma Delimited File

Read Tabular Data

These functions share the common arguments:

```
read_(file, col_names = TRUE, col_types = NULL, locale = default_locale(), na = c("", "NA"),  
quoted_na = TRUE, comment = "", trim_ws = TRUE, skip = 0, n_max = Inf, guess_max = min(1000,  
n_max), progress = interactive())
```

Comma Delimited Files
read_csv("file.csv")
To make file.csv run:
write_file(x = "a,b,c\n1,2,3\n4,5,NA", path = "file.csv")

Semi-colon Delimited Files
read_csv2("file2.csv")
write_file(x = "a,b;c\n1,2,3\n4,5,NA", path = "file2.csv")

Files with Any Delimiter
read_delim("file.txt", delim = "|")
write_file(x = "a|b|c\n1|2|3\n4|5|NA", path = "file.txt")

Fixed Width Files
read_fwf("file.fwf", col_positions = c(1, 3, 5))
write_file(x = "a b c\n1 2 3\n4 5 NA", path = "file.fwf")

Tab Delimited Files
read_tsv("file.tsv") Also **read_table()**.
write_file(x = "a\tb\tc\n1\t2\t3\n4\t5\tNA", path = "file.tsv")

USEFUL ARGUMENTS


Data types

readr functions guess the types of each column and convert types when appropriate (but will NOT convert strings to factors automatically).

A message shows the type of each column in the result.

```
## Parsed with column specification:  
## col<tbl>  
##   age = col_integer(), # age is an integer  
##   sex = col_character(), # sex is a character  
##   earn = col_double() # earn is a double (numeric)
```

1. Use **problems()** to diagnose problems
`x <- read_csv("file.csv"); problems(x)`
2. Use a **col_** function to guide parsing
 - **col_guess()** - the default
 - **col_character()**
 - **col_double()**, **col_euro_double()**
 - **col_datetime**(format = "") Also
 - **col_date**(format = ""), **col_time**(format = "")



Aan de slag!

Data file om in te lezen: [20180123_rainfall_klemskerke.csv](#)

Download deze file, zodat die lokaal beschikbaar is. Maak een *nieuw project* in Rstudio om aan de slag te gaan.

Doelstellingen (*zie volgende slides*):

1. De dataset ingelezen in R, as is
2. De dataset in R met subset van propere kolomnamen en bruikbare data
`types: datetime, value en quality_code`
3. Een lijn-plot van de gehele tijdsreeks

Meer uitdagingen nodig? Maak dan bv. een bar plot van de jaarlijkse som. En er staan nog een aantal andere data sets te wachten, of voeg gerust je eigen data set toe aan de google drive...



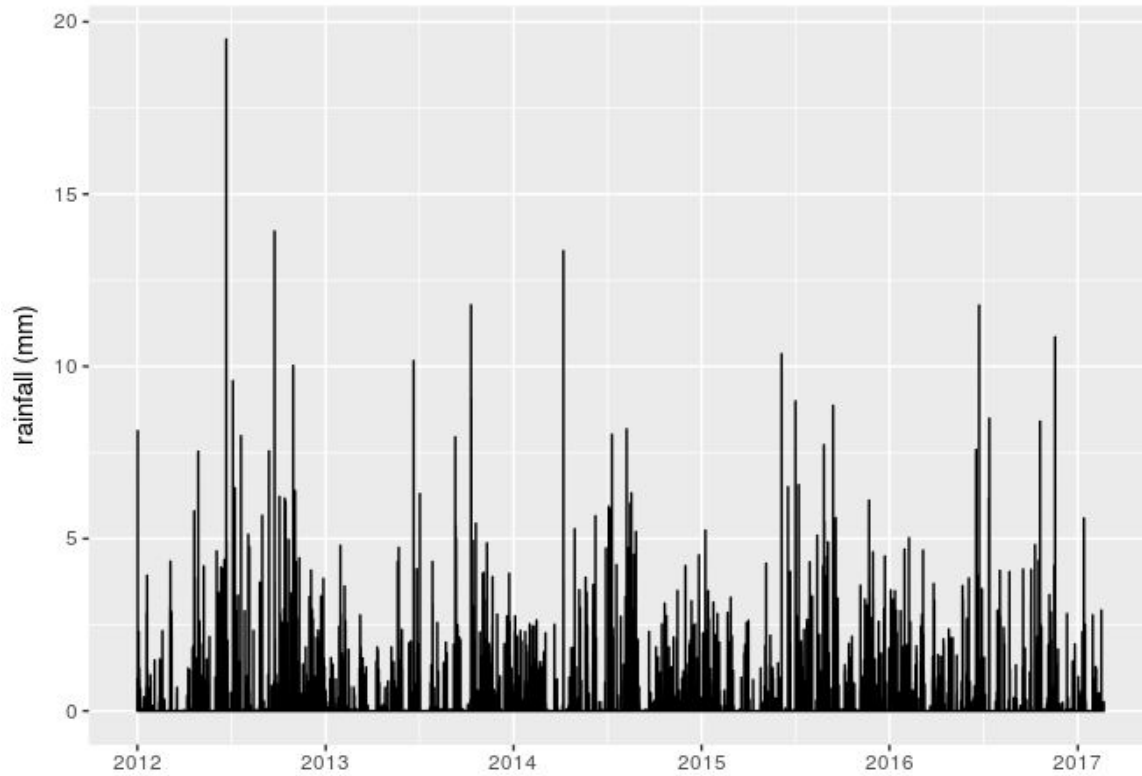
```
> head(klemskerke)
# A tibble: 6 x 5
  `#Timestamp` Value `Quality Code` `Absolute Value` `AV Quality Code`
  <dtm> <chr> <int> <chr> <chr>
1 2011-12-31 23:00:00 0,00 220 <NA> <NA>
2 2012-01-01 00:00:00 0,95 220 <NA> <NA>
3 2012-01-01 01:00:00 0,72 220 <NA> <NA>
4 2012-01-01 02:00:00 0,20 220 <NA> <NA>
5 2012-01-01 03:00:00 0,00 220 <NA> <NA>
6 2012-01-01 04:00:00 0,27 220 <NA> <NA>
```

Remark: just a possible solution, to have a common objective



```
> head(klemskerke)
# A tibble: 6 x 3
  datetime value quality_code
  <dtm>   <dbl>      <int>
1 2011-12-31 23:00:00  0.00      220
2 2012-01-01 00:00:00  0.95      220
3 2012-01-01 01:00:00  0.72      220
4 2012-01-01 02:00:00  0.20      220
5 2012-01-01 03:00:00  0.00      220
6 2012-01-01 04:00:00  0.27      220
```

Remark: just a possible solution, to have a common objective

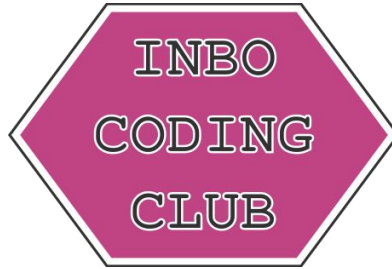


Remark: just a possible solution, to have a common objective

Andere resources voor data import...

- <https://inbo-tutorials.netlify.com/styleguide/r/>
- Heb je grote data files? zie [reading large data files in R](#)
- [Check het verschil](#) tussen tibble en data.frame
- ...





Zaal: Herman Teirlinck - 01.71 - Frans Breziers

Datum: 22/02/2018, van 10:00 tot 12:00

(inschrijving open vanaf 01/02/2018)

Backup...

The screenshot shows the 'Open Data Portaal' website. The header includes the logo 'gent: zoveel stad' and the text 'Open Data Portaal' on the left, and 'Aanmelden' with a user icon on the right. A navigation menu contains 'Home', 'Data', 'Nieuws & blog', 'Apps', 'Contact', and 'Devzone'. Below the menu is a search bar with a dropdown menu set to 'Alle' and the text 'Zoeken naar apps, data, ideeën,' followed by a search icon. The breadcrumb trail reads 'Stad.gent > Open Data Portaal > Data > Natuurlijke Groei'. The main content area features the title 'Natuurlijke Groei' and a subtitle 'Natuurlijke bevolkingsaan groei per wijk 1999-2009'. Below this, it says 'Aangepast 06-03-2017'. There are three metadata items: 'Categorie' (tag icon) 'Stad, beleid en bestuur', 'Data kwaliteit' (star icon) 'Geen kwaliteitsgarantie', and 'Databron' (link icon) 'Datatank Gent'. On the right side, there are two subscription buttons: 'Abonneer op deze dataset' with the text 'U krijgt deze E-mails op uw Mijn Gent E-mail adres bij datawijzigingen.', and 'Abonneer op structuurwijzigingen' with the text 'U krijgt deze E-mails op uw Mijn Gent E-mail adres bij structuurwijzigingen.'

gent: zoveel stad Open Data Portaal Aanmelden

Home Data Nieuws & blog Apps Contact Devzone

Alle Zoeken naar apps, data, ideeën,

Stad.gent > Open Data Portaal > Data > Natuurlijke Groei

Natuurlijke Groei

Natuurlijke bevolkingsaan groei per wijk 1999-2009

Aangepast 06-03-2017

- Categorie**
Stad, beleid en bestuur
- Data kwaliteit**
Geen kwaliteitsgarantie
- Databron**
Datatank Gent

Abonneer op deze dataset
U krijgt deze E-mails op uw Mijn Gent E-mail adres bij datawijzigingen.

Abonneer op datawijzigingen

Abonneer op structuurwijzigingen
U krijgt deze E-mails op uw Mijn Gent E-mail adres bij structuurwijzigingen.

Backup...

Check the open data set [20180123_gent_groeiperwijk.csv](#)

This is NOT a *tidy* (see [cheatsheet](#)) data set... We can make this a tidy data set to have the full power of ggplot and others...

By doing so, you'll be able to use:

- gather
- separate
- parse_integer

all explained in the [cheatsheet](#)!

Why not creating a [heatmap](#)?

```
> tidy_bevolking
# A tibble: 275 x 3
  wijk      year growth
  <chr> <int> <int>
1 Binnenstad 1999 -36
2 Bloemekenswijk 1999 12
3 Brugse Poort - Rooigem 1999 85
4 Dampoort 1999 107
5 Drongen 1999 3
6 Elisabethbegijnhof - Papegaai 1999 -4
7 Gentbrugge 1999 4
8 Kanaaldorpen en -zone 1999 5
9 Ledeborg 1999 -4
10 Macharius - Heirnis 1999 47
# ... with 265 more rows
```

