

Workshop R Markdown

Martine Jansen

mm.jansen@fontys.nl



- Reproducible documents
- HTML-pages
 - Pdf
 - Nicely formatted HTML/pdf
 - Word
- Integration of text, analyses, figures, comments, ... in 1 document

R Markdown

from  Studio®

I immediately knew the [Case Shiller Home Price Index visualization](#) would be perfect for reuse. This is a bit of a hack of `rCharts` and should not be considered best practices, but it is a demo package. In the spirit of this d

- Generalize the d3 code a l
- Build in `R` with `rCharts` tr

A Pandoc Markdown Article Starter and Template

Steven V. Miller *Clemson University*

Headings

As I men
supplied

This style provides first and second-level headings (that is, # and ##), demonstrated in the next section. You may get unexpected output if you try to use ### and smaller headings.

1. # IN HIS LATER BOOKS², Tufte starts each section with a bit of vertical space, a non-
2. # indented paragraph, and sets the first few words of the sentence in small caps. To
3. # accomplish this using this style, call the `newthought()` function in `tufte` in an *inline R*
4. r expression ``r`` as demonstrated at the beginning of this paragraph.³
5. r
6.
7. d
8. m
9. m
10.
11.

Figures

Margin Figures

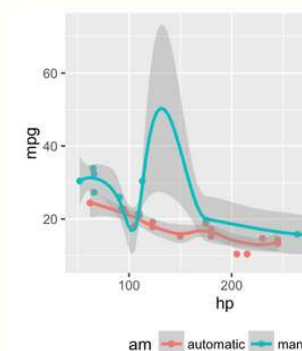
	Hello		World	
car	mpg	cyl	disp	hp
Mazda RX4	21.0	6	160	110
Mazda RX4 Wag	21.0	6	160	110
Datsun 710	22.8	4	108	93
Hornet 4 Drive	21.4	6	258	110
Hornet Sportabout	18.7	8	360	175

Tufte's work. To place figures in the margin `gin = TRUE`. For example:

› provide a figure caption. You can adjust the `nd fig.height` chunk options. These are `y` scaled down to fit within the handout margin.

² [Beautiful Evidence](#)

³ Note you should not assume `tufte` has your R session. You should either `library(tufte)` in your R Markdown document before you `newthought()`, or use `tufte::newthought()`.



<https://rmarkdown.rstudio.com/gallery.html>

https://cran.r-project.org/web/packages/kableExtra/vignettes/awesome_table_in_html.html

2019-02-21

Interactive
reproducible
documents

- HTML-widgets
- Shiny
- Dashboards
- Presentations
- Books, websites,

...

R Markdown

from  Studio®

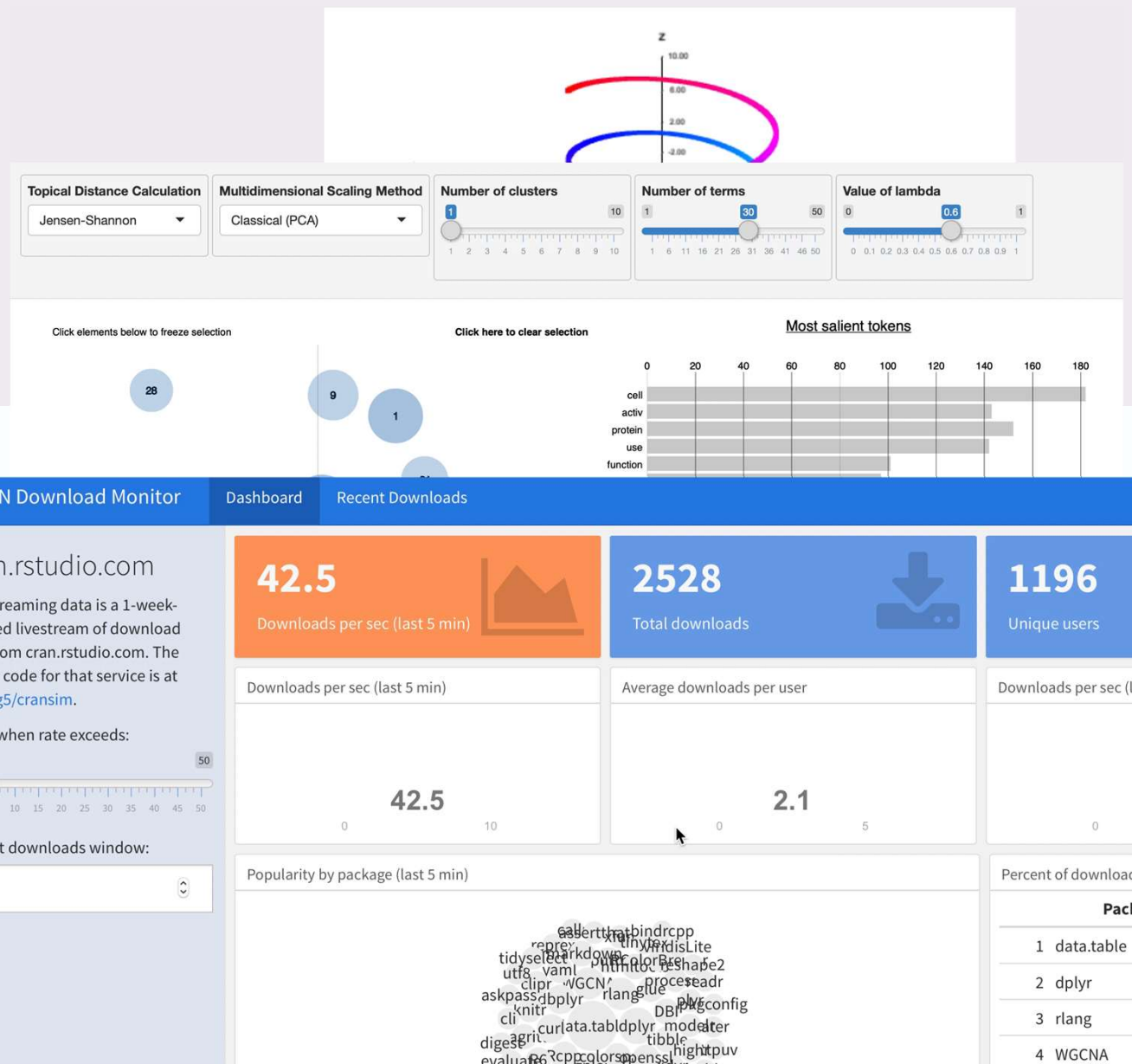
<https://www.htmlwidgets.org>

<https://www.rstudio.com/products/shiny/shiny-user-showcase/>

<https://gallery.shinyapps.io/cran-gauge/>

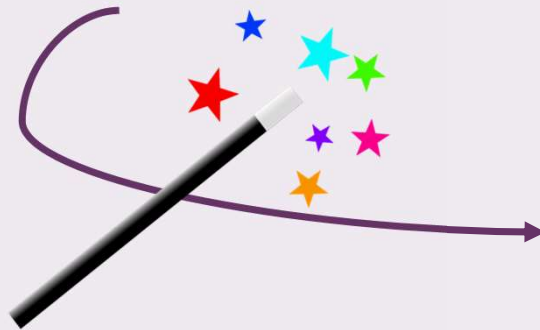
2019-02-21

Fonty



What WE will do:

IncomingInternati							
Bestand	Start	Invoegen	Pagina-indeling	Formules	Gegevens	Controleren	
	A	B	C	D	E	F	G
1	Program	2006	2007	2008	2009	2010	2011
2	Research University - Bachelor	6.670	7.684	9.057	10.320	11.231	12.371
3	Research University - Master	6.042	7.252	8.647	10.349	12.180	13.712
4	University of Applied Sciences - Bachelor	17.961	19.075	20.432	22.445	24.015	25.027
5	University of Applied Sciences - Master	909	983	1.011	1.022	1.075	1.089
6							



Incoming Degree Students

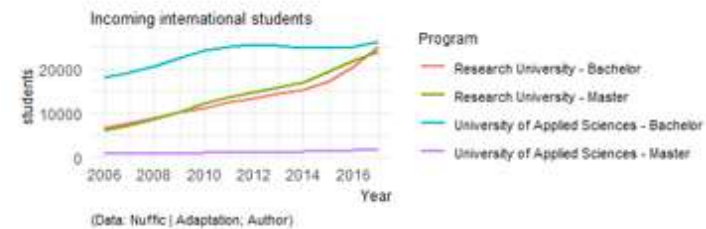
Martine Jansen

30 januari 2019

1. Introduction

In 2006, 31582 international students were enrolled in a higher education institution in the Netherlands. In 2017 there were 76908, a growth of 144% over a period of 11 years.

1.1 A figure

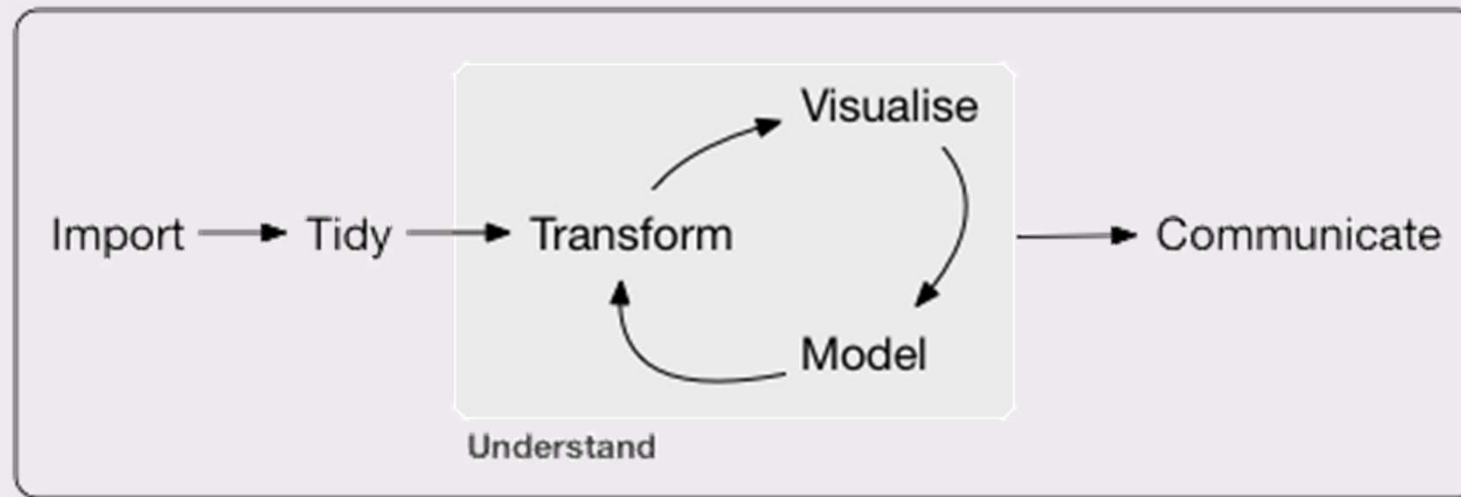


1.2 A table

Choice of Program, per year

Program	2006	2017
Research University - Bachelor	21%	32%
Research University - Master	19%	31%
University of Applied Sciences - Bachelor	57%	34%
University of Applied Sciences - Master	3%	3%

How we will do this:



From: R for Data Science
Garrett Golemund
Hadley Wickham
<http://r4ds.had.co.nz/>

Program



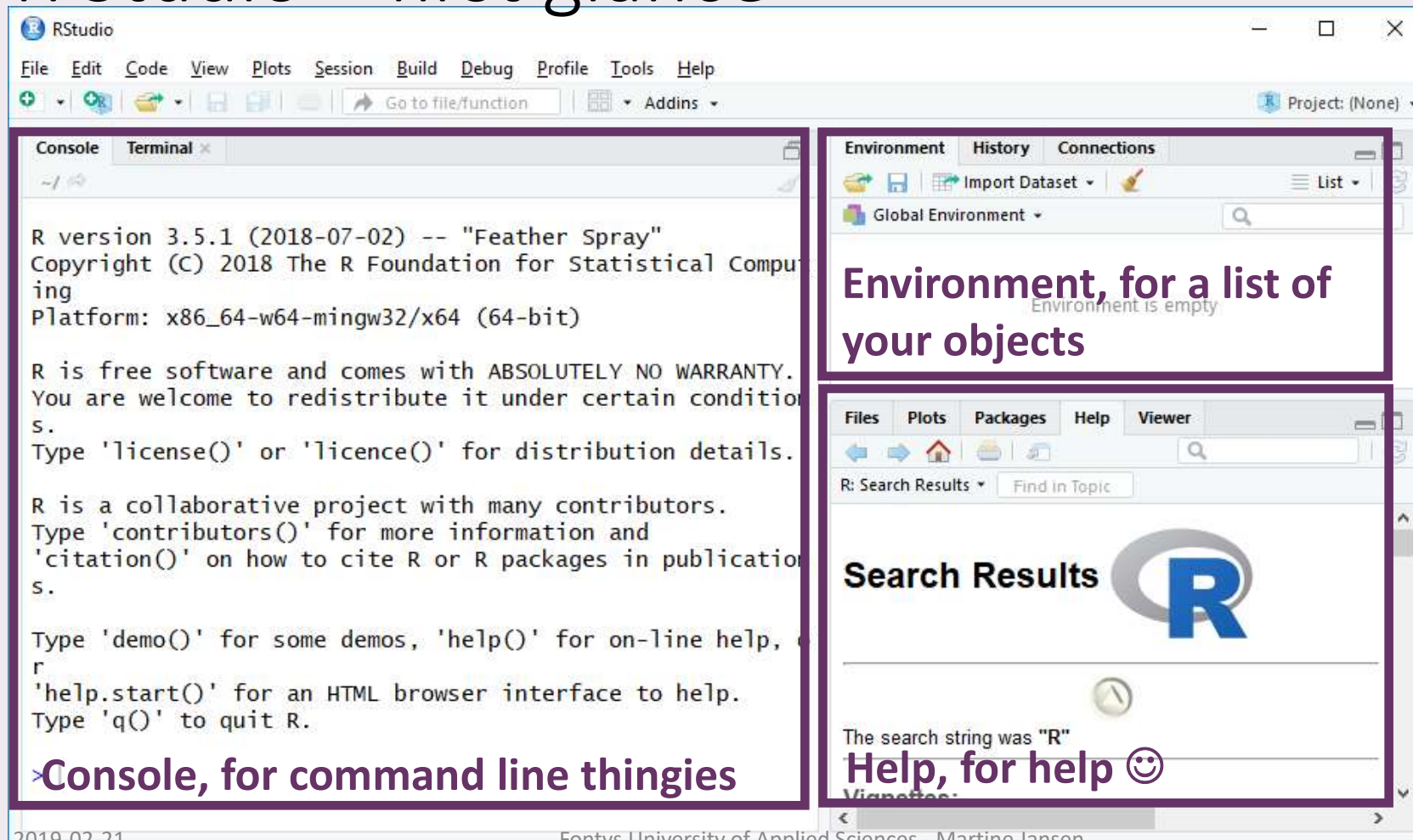
<https://www.rstudio.com/>



<https://www.r-project.org/>

R Studio – first glance

(Open R Studio)



Some command line R, in the Console

<https://nl.freepik.com/>



```
> 2 + 3  
[1] 5
```

After the “>” type in R commands, hit enter
Output is shown, it has only [1] line, and is 5

```
> w <- 2 + 3
```

“<-” for assignment, give w the value of 2 + 3, no output!
But look at tab “Environment”:

Environment	History	Connections
Import Dataset		
Global Environment		
Values		
w	5	

```
> w + 4  
[1] 9
```

Output of w + 4 is 9, object w unchanged

```
> w <- w + 4  
> |
```

No output,
object w changed

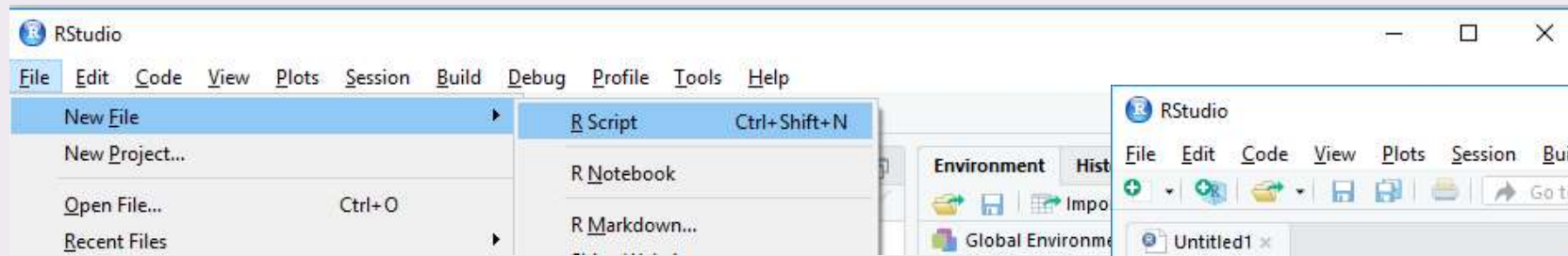
Environment	History	Connections
Import Dataset		
Global Environment		
Values		
w	9	



Sweep the
global
environment

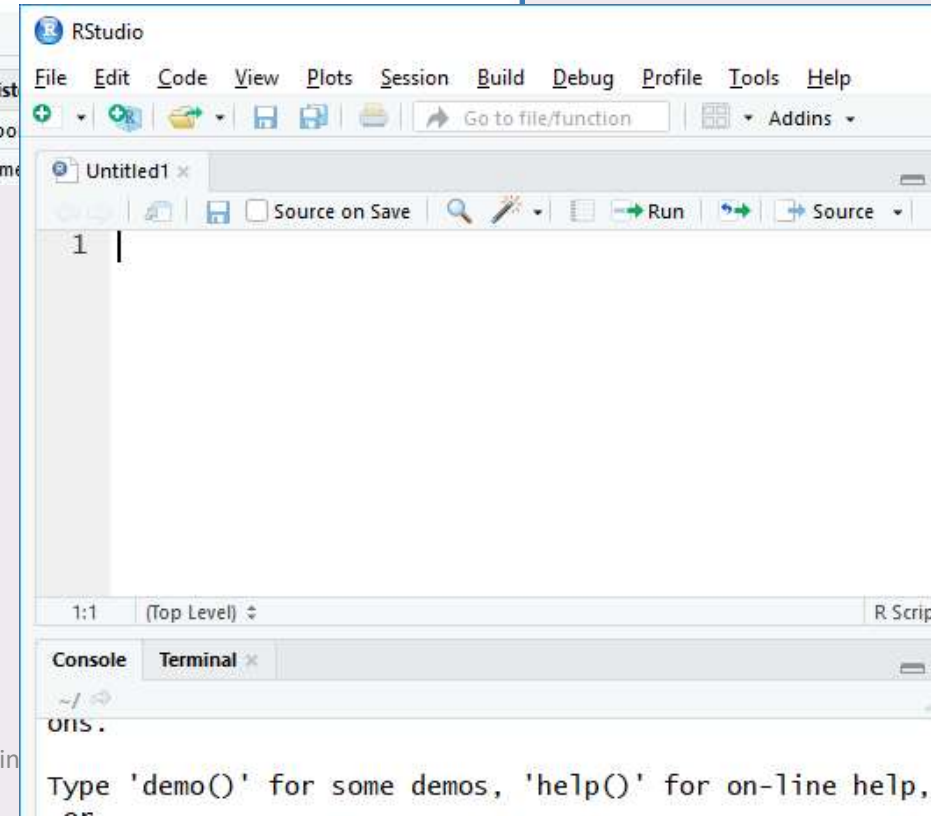
Keep all useful commands in a Script

File / New File / R Script



New Untitled Script File →

Console now below →



Add some code and run some code

A screenshot of the RStudio application window. The main editor shows a script with four lines of R code:

```
1 p <- 50
2 s <- "The number of participants is"
3 paste(s,p)
4
```

 The Environment pane on the right is empty, displaying "Environment is empty". The Console pane at the bottom shows a prompt ">". The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The toolbar contains icons for saving, running, and other functions. A purple callout box is overlaid on the bottom right of the script editor, containing instructions on how to run the code. Another purple callout box is overlaid on the bottom left of the script editor, containing an instruction to enter the code.

• enter the code above

- select the code to run, or (in case it is only one line), place cursor on line
- hit Run, or CTRL-ENTER

Some code ran

The screenshot displays the RStudio environment with the following components:

- Source Editor:** Contains the following R code:

```
1 p <- 50
2 s <- "The number of participants is"
3 paste(s,p)
4
```
- Environment Pane:** Shows the 'Global Environment' with the following values:

Variable	Value
p	50
s	"The number of participan..."
- Console:** Shows the execution output:

```
> p <- 50
> s <- "The number of participants is"
> paste(s,p)
[1] "The number of participants is 50"
>
```
- Search Results Pane:** Displays the R logo and the text "The search string was 'R'".

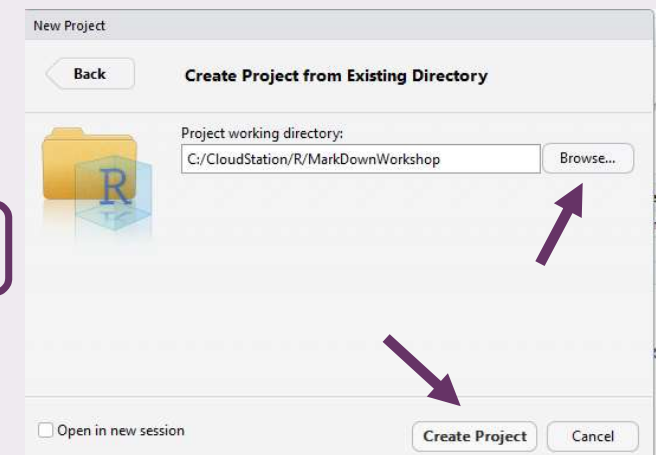
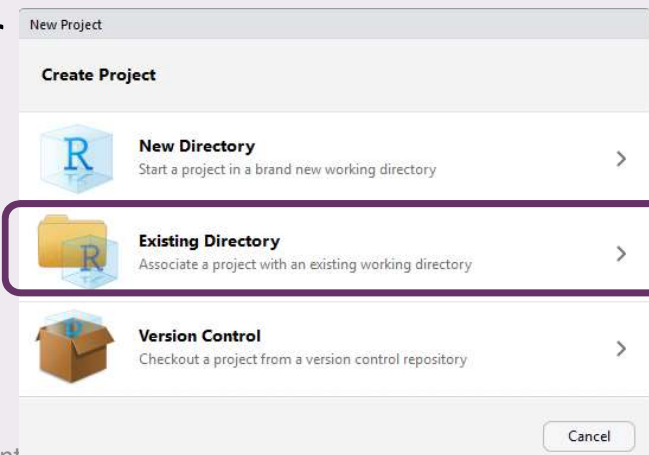
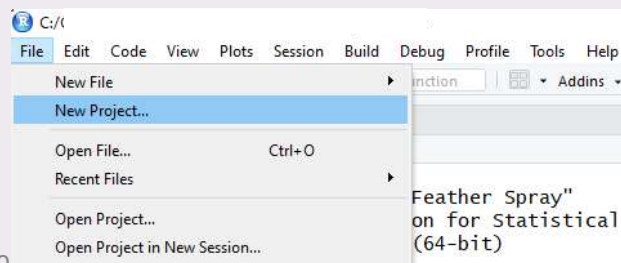
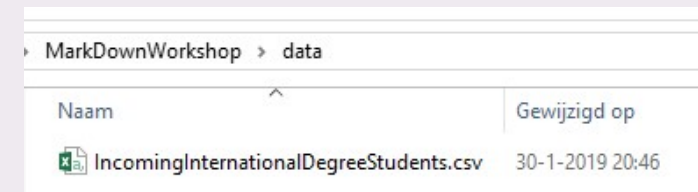
objects in
Environment

code and output
in Console

R project, keeping it together



- In Windows Explorer, make a folder “MarkDownWorkshop”
- In this folder make a subfolder “data”
 - the received csv goes here
- Open Rstudio, make this folder an R project:
 - File / New Project ...
 - Existing Directory
 - Browse to the folder
 - Create Project

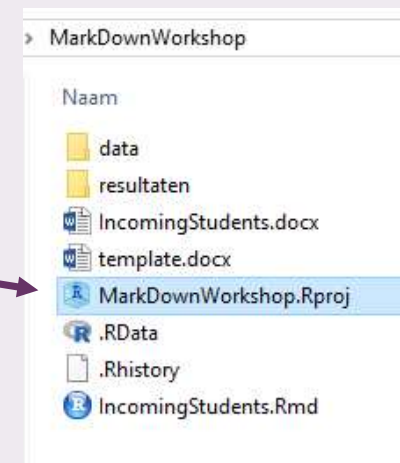


Just so that you know the proper way to open an existing R Project:



Close R Studio,
no need to save anything

Open your R Project again,
by doubleclicking the .Rproj
file in the project folder

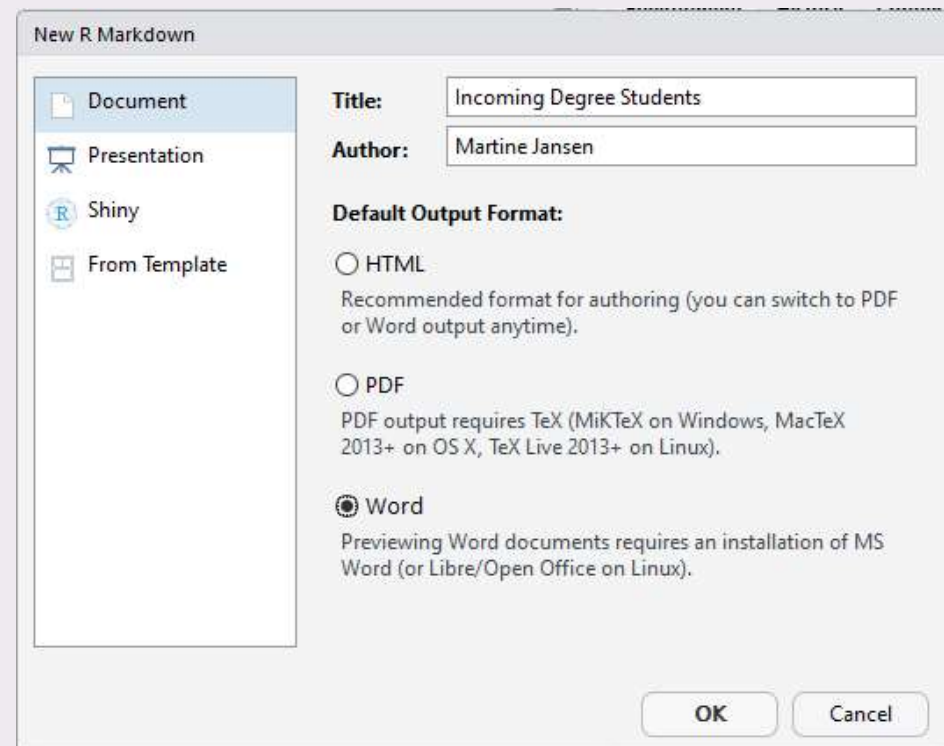
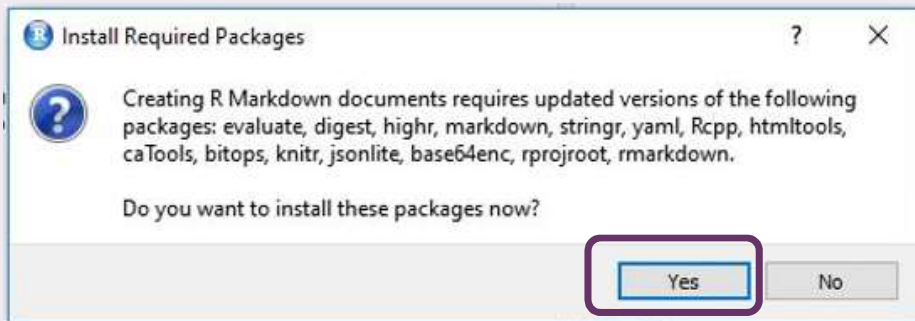
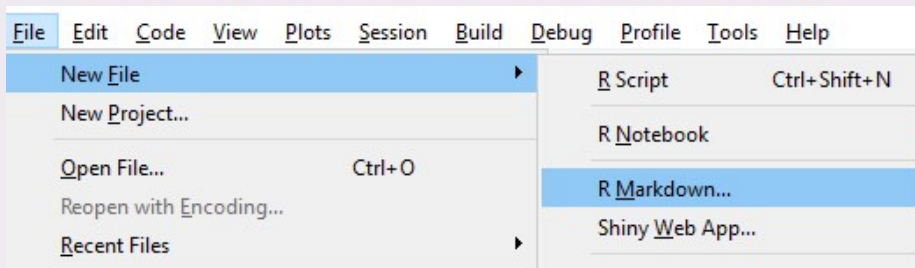


Project folder

Add an R Markdown file, for the script



- File / New File / R Markdown
- Fill in title and author
- Choose Default Output Format Word
- OK



Second glance

The screenshot shows the RStudio interface with the following components:

- Editor:** Contains a new R Markdown document titled "IncomingStudents.Rmd" with the following boilerplate code:

```
1 ---
2 title: "Incoming Degree Students"
3 author: "Martine Jansen"
4 date: "30 januari 2019"
5 output: word_document
```
- Environment:** Shows "Global Environment" and "Environment is empty".
- Files:** Shows the project directory "C:\CloudStation\R\MarkDownWorkshop" with the following files:

Name	Size	Modified
..		
.Rhistory	56 B	Jan 31, 2019, 7:53 A
IncomingStudents.docx	12.8 KB	Jan 30, 2019, 10:31
IncomingStudents.Rmd	891 B	Jan 30, 2019, 10:31
MarkDownWorkshop.Rproj	218 B	Feb 1, 2019, 2:04 PM
- Console:** Shows the R prompt and the following text:

```
r on-line help, or
'help.start()' for an HTML browser interf
ace to help.
Type 'q()' to quit R.
```

An untitled doc with boiler plate R Markdown code

> Console, now in the bottom

2019-02-21 Fontys University of Applied Sciences - Martine Jansen


```
1 ---
2 title: "Incoming Degree Students"
3 author: "Martine Jansen"
4 date: "30 januari 2019"
5 output: word_document
6 ---
7
8 {r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10
11
12 ## R Markdown
13
14 This is an R Markdown document. Markdown is a simple formatting syntax
15 for authoring HTML, PDF, and MS Word documents. For more details on
16 using R Markdown see <http://rmarkdown.rstudio.com>.
17
18 When you click the Knit button a document will be generated that
19 includes both content as well as the output of any embedded R code
20 chunks within the document. You can embed an R code chunk like this:
21
22 {r cars}
23 summary(cars)
24
25
26 ## Including Plots
27
28 You can also embed plots, for example:
29
30 {r pressure, echo=FALSE}
31 plot(pressure)
```

Doc is still untitled 😊

YAML header:

YAML ain't Markup Language

R Chunk named **setup** : R code

: Header 2

plain text, an URL and **bold**

unnamed R Chunk : R code

R Chunk named
pressure: R code

The anatomy of an R Chunk

An R Chunk starts with:

- three backticks ```
- curly brace open {
- lower case r

given name

Chunk options,
this one means we do not
wish to see the R code in the
output

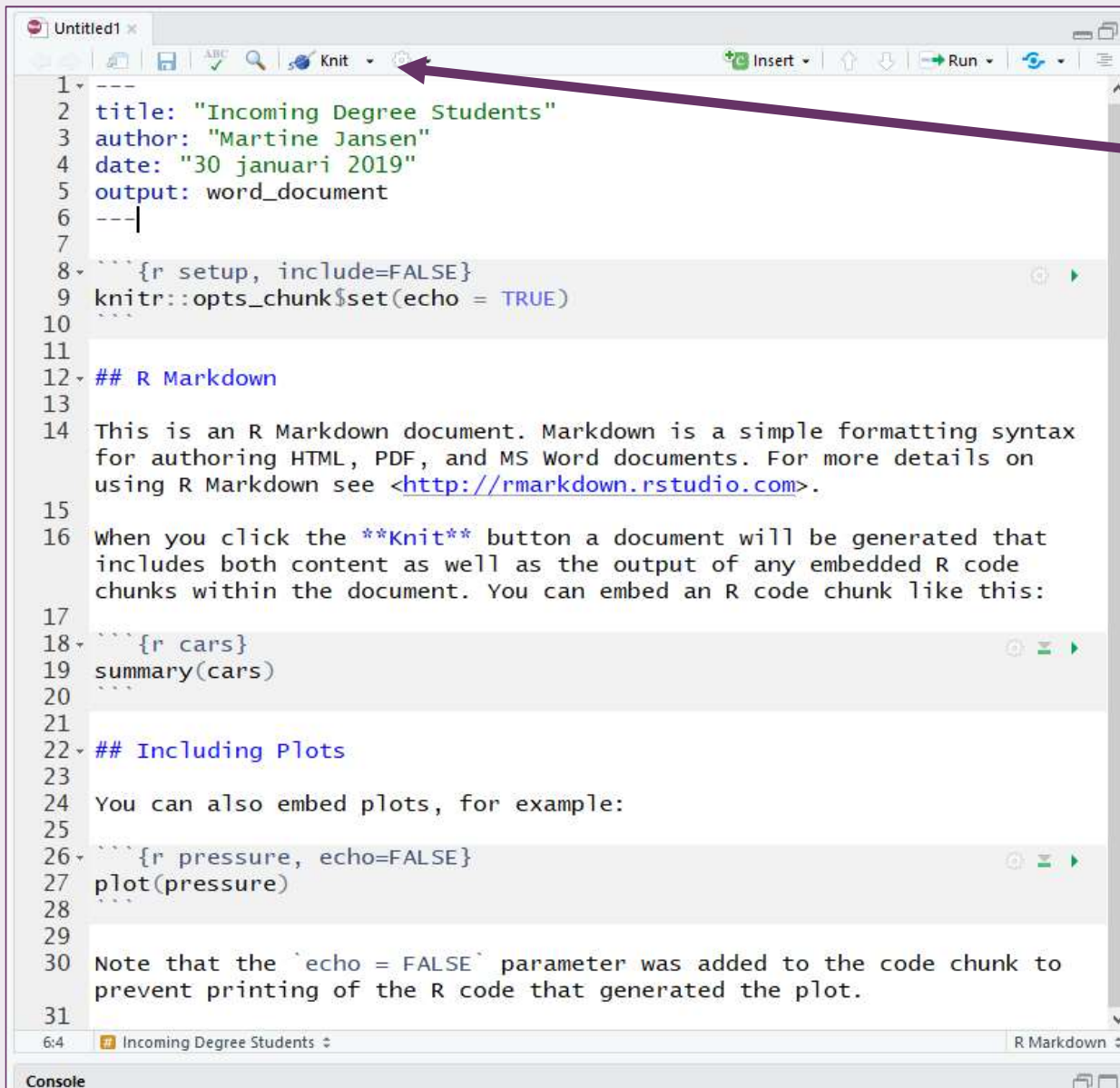
the closing
curly brace }

```
```{r pressure, echo=FALSE}  
plot(pressure)
```
```

R code

An R Chunk closes with:

- three backticks ```



```
1 ---
2 title: "Incoming Degree Students"
3 author: "Martine Jansen"
4 date: "30 januari 2019"
5 output: word_document
6 ---
7
8 {r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10
11 ## R Markdown
12
13 This is an R Markdown document. Markdown is a simple formatting syntax
14 for authoring HTML, PDF, and MS Word documents. For more details on
15 using R Markdown see <http://rmarkdown.rstudio.com>.
16
17 When you click the Knit button a document will be generated that
18 includes both content as well as the output of any embedded R code
19 chunks within the document. You can embed an R code chunk like this:
20
21 {r cars}
22 summary(cars)
23
24 ## Including Plots
25
26 You can also embed plots, for example:
27
28 {r pressure, echo=FALSE}
29 plot(pressure)
30
31 Note that the `echo = FALSE` parameter was added to the code chunk to
32 prevent printing of the R code that generated the plot.
```

6:4 Incoming Degree Students R Markdown

Console

Click Knit

When asked,
save the file as
"IncomingStudents"
(in the project folder)

Wait a bit, and then:



Bestand
Hulpmiddelen
Beeld
IncomingStudents.docx [Alleen-lezen] (Compatibiliteitsmodus) - Word

Incoming Degree Students

Martine Jansen
30 januari 2019

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

| ## | speed | dist |
|-------------|-------|----------------|
| ## Min. | : 4.0 | Min. : 2.00 |
| ## 1st Qu.: | 12.0 | 1st Qu.: 26.00 |
| ## Median : | 15.0 | Median : 36.00 |
| ## Mean : | 15.4 | Mean : 42.98 |
| ## 3rd Qu.: | 19.0 | 3rd Qu.: 56.00 |
| ## Max. | :25.0 | Max. :120.00 |

Including Plots

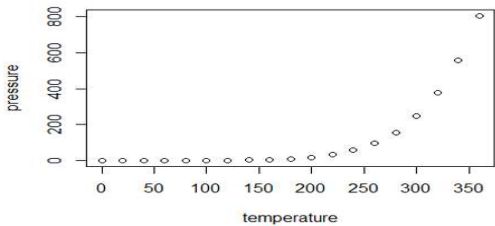
You can also embed plots, for example:

```
## speed dist
```

| ## | speed | dist |
|-------------|-------|----------------|
| ## Min. | : 4.0 | Min. : 2.00 |
| ## 1st Qu.: | 12.0 | 1st Qu.: 26.00 |
| ## Median : | 15.0 | Median : 36.00 |
| ## Mean : | 15.4 | Mean : 42.98 |
| ## 3rd Qu.: | 19.0 | 3rd Qu.: 56.00 |
| ## Max. | :25.0 | Max. :120.00 |


Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Einde van document

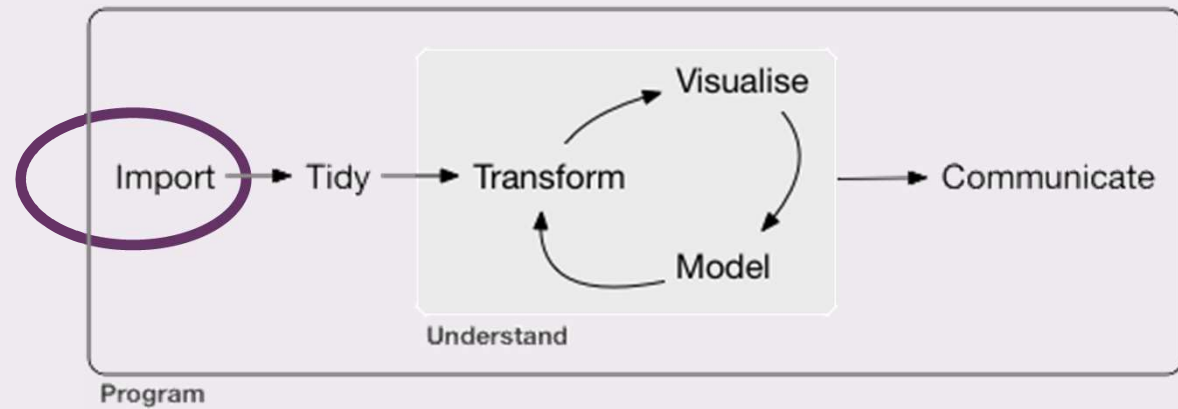


Scherf 1-2 van 2

100 %

Yay, it is a word document!
Unfortunately so far only title,
author and date is ok.
The rest is still boiler plate.
We'll fix that now.
Close this word doc.

Import



- `read.csv()`
 - `read.csv2()`
 - `read_csv()`
 - `read_csv2()`
 - `read_xlsx()`
 - ...
- } **base R**
- } **readR, part of core tidyverse**
- } **readxl, part of tidyverse**

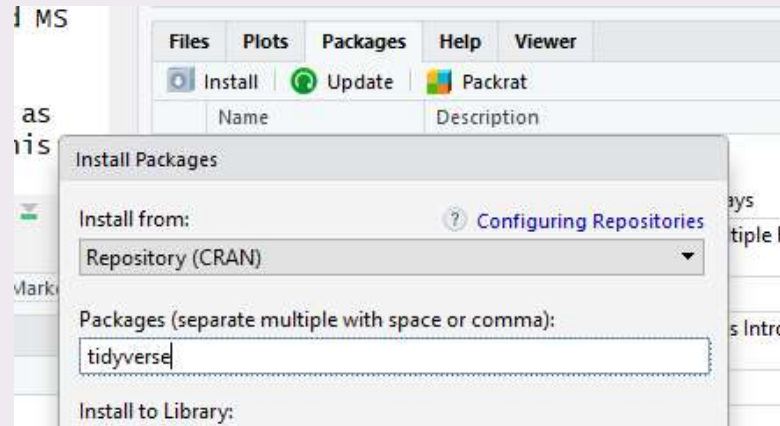
Install package



Only need to do so ONCE,
it is like buying a book



Either through the tab Packages:



Or by typing in the tab Console `install.packages("tidyverse")`



<https://www.tidyverse.org/>

Delete most of the content



```
1 ---
2 title: "Incoming Degree Students"
3 author: "Martine Jansen"
4 date: "30 januari 2019"
5 output: word_document
6 ---
7
8 {r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10
11
12 ## R Markdown
13
14 This is an R Markdown document. Markdown is a simple formatting syntax
15 for authoring HTML, PDF, and MS Word documents. For more details on
16 using R Markdown see <http://rmarkdown.rstudio.com>.
17
18 When you click the Knit button a document will be generated that
19 includes both content as well as the output of any embedded R code
20 chunks within the document. You can embed an R code chunk like this:
21
22 {r cars}
23 summary(cars)
24
25 ## Including Plots
26
27 You can also embed plots, for example:
28
29 {r pressure, echo=FALSE}
30 plot(pressure)
31
32 Note that the `echo = FALSE` parameter was added to the code chunk to
33 prevent printing of the R code that generated the plot.
```

Use the tidyverse, read the data & Knit



```
IncomingStudents.Rmd* x
1 ---
2 title: "Incoming Degree Students"
3 author: "Martine Jansen"
4 date: "30 januari 2019"
5 output: word_document
6 ---
7
8 ```{r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10
11 library(tidyverse)
12
13
14
15 ```{r readData}
16 # read file IncomingInternationalDegreeStudents.csv
17 location_of_data <- file.path("data",
18                               "IncomingInternationalDegreeStudents.csv")
19 dStudents <- read_csv2(location_of_data)
20
21
22
23
24
```

Add this r Chunk

Load the desired package every time, it is like using the book



within an R Chunk is a **comment**



The result

Not what we
aimed for....

We do not want
to see all this
programming
code and
messages....

Close this doc again

Incoming Degree Students

Martine Jansen

30 januari 2019

```
# read file IncomingInternationalDegreeStudents.csv

location_of_data <- file.path("data",
                              "IncomingInternationalDegreeStudents.csv")
dStudents <- read_csv2(location_of_data)

## Using ',' as decimal and '.' as grouping mark. Use read_delim() for more
## control.

## Parsed with column specification:
## cols(
##   Program = col_character(),
##   `2006` = col_number(),
##   `2007` = col_number(),
##   `2008` = col_number(),
##   `2009` = col_number(),
##   `2010` = col_number(),
##   `2011` = col_number(),
##   `2012` = col_number(),
##   `2013` = col_number(),
##   `2014` = col_number(),
##   `2015` = col_number(),
##   `2016` = col_number(),
##   `2017` = col_number()
## )
```

Adjustment in **setup** & Knit again



```
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = FALSE, warning = FALSE, message = FALSE)

library(tidyverse)

```
```

| Incoming Degree Students

Martine Jansen

30 januari 2019

Very calm and quiet indeed.

And nothing to be seen.

We didn't tell R Markdown to show us anything 😊

Some content



```
9
10 {r setup, include=FALSE}
11 knitr::opts_chunk$set(echo = FALSE, warning = FALSE)
12
13 library(tidyverse)
14
15
16
17 # Introduction
18
19 The data
20
21 {r}
22 dStudents
23
24
```

Header 1

Introduction

The data

```
## # A tibble: 4 x 13
##   Program `2006` `2007` `2008` `2009` `2010` `2011` `2012` `2013` `2014`
##   <chr>    <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Resear~    6670   7684   9057  10320  11231  12371  13248  14304  15357
## 2 Resear~    6042   7252   8647  10349  12180  13712  14595  15846  16914
## 3 Univer~   17961  19075  20432  22445  24015  25027  25343  25215  24579
## 4 Univer~    909    983   1011   1022   1075   1089   1125   1239   1452
## # ... with 3 more variables: `2015` <dbl>, `2016` <dbl>, `2017` <dbl>
```

Incoming Degree Students

Martine Jansen

30 januari 2019

The ugly, though
informative way
to display a table

A less ugly way: `kable()`, part of the knitr package



coming up



A less ugly way of displaying a table

- Install the package knitr
- Load the package knitr
- Change the code:



← How?

← Where do you put this code?

```
# Introduction
```

```
The data
```

```
...{r}
dStudents %>%
  kable()
...
```

The %>% is the “pipe” operator,
from the magrittr package.



(part of the tidyverse,
is already installed)

The thing before the %>%
serves as the first argument for
the function after the %>%

- Knit the document and see the change
- Look at the help info on `kable()` how to add a caption
- Give the table a caption, Knit and see the change




```

1 ---
2 title: "Incoming Degree Students"
3 author: "Martine Jansen"
4 date: "30 januari 2019"
5 output: word_document
6 ---
7
8 ```{r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = FALSE, warning = F
10
11 library(tidyverse)
12 library(knitr)
13
14 ```
15
16 ```{r readData}
17
18 # read file IncomingInternationalDegreeStudents
19
20 location_of_data <- file.path("data",
21                               "IncomingInternat
22 dStudents <- read_csv2(location_of_data)
23
24 ```
25
26 # Introduction
27
28 The data
29
30 ```{r}
31 dStudents %>%
32   kable(caption = "The data")
33 ```
34

```

Better looking, but data is not tidy!

Incoming Degree Students

Martine Jansen

30 januari 2019

Introduction

The data

The data

| Program | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 |
|---|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Research University - Bachelor | 6670 | 7684 | 9057 | 10320 | 11231 | 12371 | 13248 | 14304 | 15357 | 17111 | 20359 | 24814 |
| Research University - Master | 6042 | 7252 | 8647 | 10349 | 12180 | 13712 | 14595 | 15846 | 16914 | 19393 | 21902 | 23858 |
| University of Applied Sciences - Bachelor | 17961 | 19075 | 20432 | 22445 | 24015 | 25027 | 25343 | 25215 | 24579 | 24536 | 24813 | 26088 |
| University of Applied Sciences - Master | 909 | 983 | 1011 | 1022 | 1075 | 1089 | 1125 | 1239 | 1452 | 1545 | 1724 | 2148 |

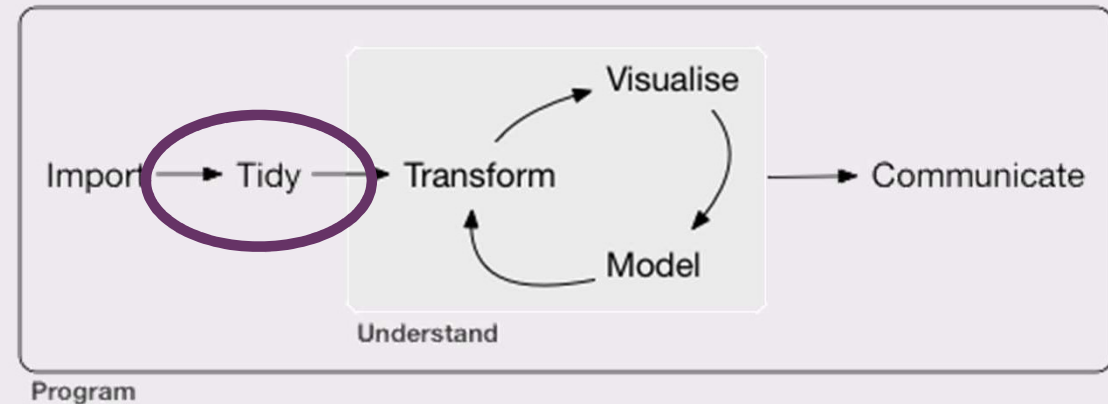
Tidy

Tidy datasets are easy to

- manipulate
- model
- visualise,

and have a specific structure:

- each variable is a column
- each observation is a row,
- each type of observational unit is a table.



| Program | 2006 | 2007 | 2008 | 2009 | 2010 |
|--------------------------------|-------|-------|-------|-------|-------|
| Research University - Bachelor | 6670 | 7684 | 9057 | 10320 | 11230 |
| Research University - Master | 6042 | 7252 | 8647 | 10349 | 12180 |
| University of Applied Sciences | 17961 | 19075 | 20432 | 22445 | 24010 |



| Program | Year | n |
|--------------------------------|------|------|
| Research University - Bachelor | 2006 | 6670 |
| ... | ... | ... |

Tidy dStudents

the chunk now has a name: **readData**



```
'''{r readData}

# read file IncomingInternationalDegreeStudents.csv

location_of_data <- file.path("data",
                              "IncomingInternationalDegreeStudents.csv")
dStudents <- read_csv2(location_of_data) %>%
  # the data is not in tidy format,
  # reshape wide to long using tidyr::gather
  # the first column (Program) is ok, no gathering necessary
  gather(key = "Year", value = "n", - Program)

'''
```



Name of the new column
collecting all **the names**
of the old columns

Name of the new column
collecting all **the values**
of the old columns

Gather all columns,
minus the column
Program

Intermezzo: a local View

meanwhile in
the console ...

```
> View(dStudents)
```



while RStudio is busy
see the green line,

select this code & CTRL+Enter

objects in the local
environment

dStudents data
has 2 char vars
and 1 num var

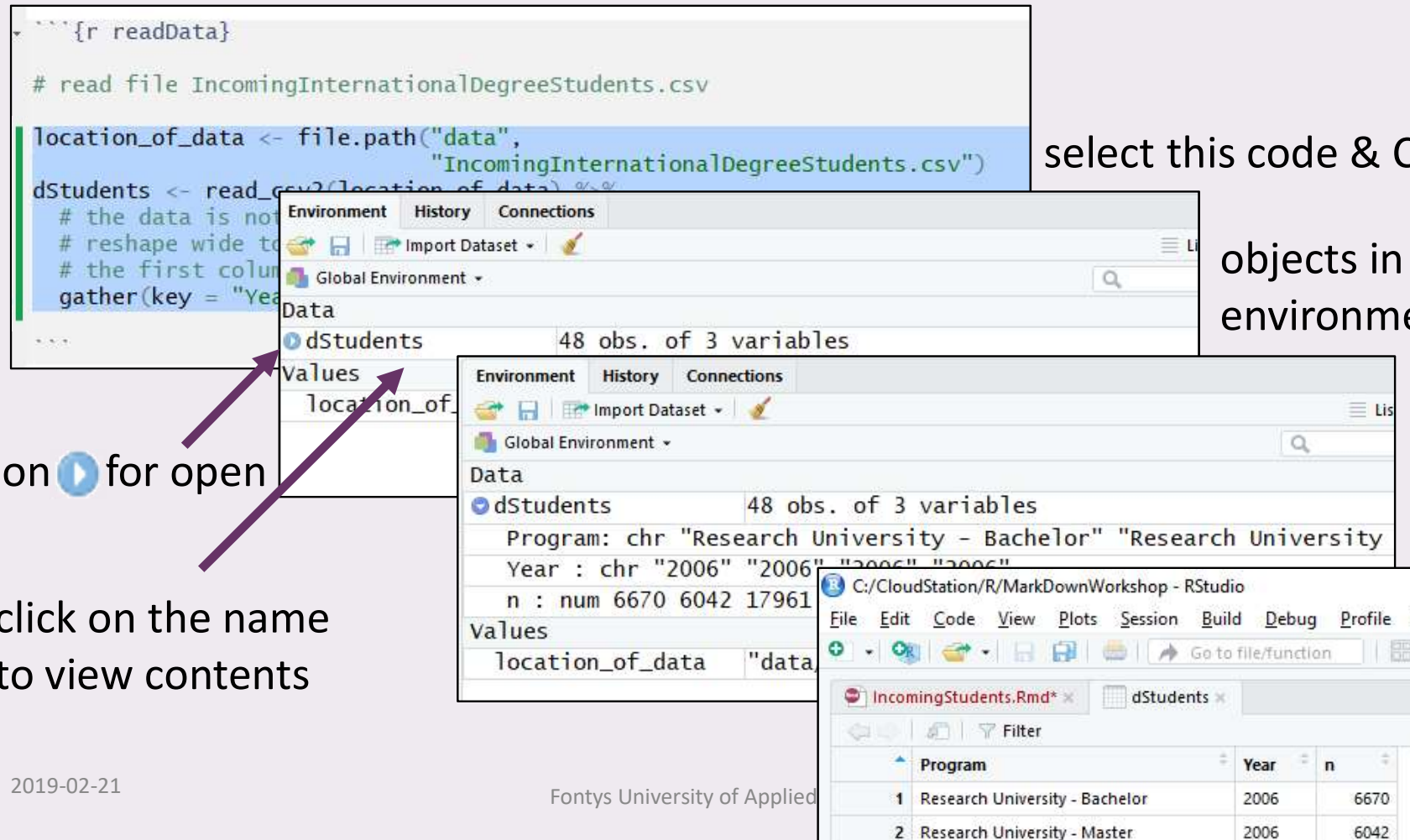
click on  for open

click on the name
to view contents

2019-02-21

Fontys University of Applied

a new tab
with the
contents of
dStudents



The screenshot shows the RStudio interface. The top pane contains R code for reading a CSV file. The middle pane shows the Environment window with 'dStudents' selected. The bottom pane shows the Data viewer with a table of data.

```
{r readData}  
  
# read file IncomingInternationalDegreeStudents.csv  
  
location_of_data <- file.path("data",  
                              "IncomingInternationalDegreeStudents.csv")  
dStudents <- read_csv2(location_of_data) %>%  
# the data is not  
# reshape wide to  
# the first column  
gather(key = "Year",
```

Environment History Connections
Global Environment
Data
dStudents 48 obs. of 3 variables
Values
location_of_data

Environment History Connections
Global Environment
Data
dStudents 48 obs. of 3 variables
Program: chr "Research University - Bachelor" "Research University"
Year : chr "2006" "2006" "2006" "2006"
n : num 6670 6042 17961
Values
location_of_data "data"

C:/CloudStation/R/MarkDownWorkshop - RStudio
File Edit Code View Plots Session Build Debug Profile I
Go to file/function
IncomingStudents.Rmd* x dStudents x
Filter
Program Year n
1 Research University - Bachelor 2006 6670
2 Research University - Master 2006 6042

See the result, knit the Rmd



Bestand Hulpmiddelen Beeld IncomingStudents.docx [Alleen-lezen] [Compatibiliteitsmodus] - Word

Incoming Degree Students

Martine Jansen
30 januari 2019

Introduction

The data

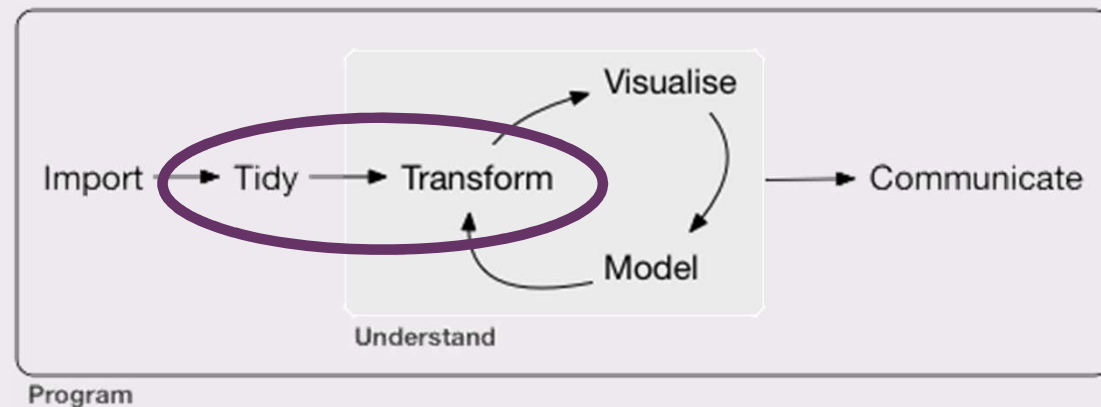
The data

| Program | Year | n |
|---|------|-------|
| Research University - Bachelor | 2006 | 6670 |
| Research University - Master | 2006 | 6042 |
| University of Applied Sciences - Bachelor | 2006 | 17961 |
| University of Applied Sciences - Master | 2006 | 909 |
| Research University - Bachelor | 2007 | 7684 |
| Research University - Master | 2007 | 7252 |
| University of Applied Sciences - Bachelor | 2007 | 19075 |
| University of Applied Sciences - Master | 2007 | 983 |
| Research University - Bachelor | 2008 | 9057 |
| Research University - Master | 2008 | 8647 |
| University of Applied Sciences - Bachelor | 2008 | 20432 |
| University of Applied Sciences - Master | 2008 | 1011 |
| Research University - Bachelor | 2009 | 10320 |
| Research University - Master | 2009 | 10349 |
| University of Applied Sciences - Bachelor | 2009 | 22445 |
| University of Applied Sciences - Master | 2009 | 1022 |
| Research University - Bachelor | 2010 | 11231 |
| Research University - Master | 2010 | 12180 |
| University of Applied Sciences - Bachelor | 2010 | 24015 |
| University of Applied Sciences - Master | 2010 | 1075 |
| Research University - Bachelor | 2011 | 12371 |
| Research University - Master | 2011 | 13712 |
| University of Applied Sciences - Bachelor | 2011 | 25027 |
| University of Applied Sciences - Master | 2011 | 1089 |
| Research University - Bachelor | 2012 | 13248 |
| Research University - Master | 2012 | 14595 |
| University of Applied Sciences - Bachelor | 2012 | 25343 |
| University of Applied Sciences - Master | 2012 | 1125 |
| Research University - Bachelor | 2013 | 14304 |
| Research University - Master | 2013 | 15846 |
| University of Applied Sciences - Bachelor | 2013 | 25215 |
| University of Applied Sciences - Master | 2013 | 1239 |
| Research University - Bachelor | 2014 | 15357 |
| Research University - Master | 2014 | 16914 |
| University of Applied Sciences - Bachelor | 2014 | 24579 |
| University of Applied Sciences - Master | 2014 | 1452 |
| Research University - Bachelor | 2015 | 17111 |
| Research University - Master | 2015 | 19393 |
| University of Applied Sciences - Bachelor | 2015 | 24536 |
| University of Applied Sciences - Master | 2015 | 1545 |
| Research University - Bachelor | 2016 | 20359 |
| Research University - Master | 2016 | 21902 |
| University of Applied Sciences - Bachelor | 2016 | 24813 |
| University of Applied Sciences - Master | 2016 | 1724 |
| Research University - Bachelor | 2017 | 24814 |
| Research University - Master | 2017 | 23858 |
| University of Applied Sciences - Bachelor | 2017 | 26088 |
| University of Applied Sciences - Master | 2017 | 2148 |

Einde van document

Data is tidy, but
this is not the
desired content

Transform



`select()`

choose columns

`filter()`

choose rows

`mutate()`

make or change columns

`arrange()`

order rows

`summarise()`

summarise rows

`group_by()`

make subgroups from rows, usually combined with 1 of the above

dExample

| FirstName | LastName | Age |
|-----------|----------|-----|
| Abe | Jans | 30 |
| Bert | Jans | 28 |
| Bert | Bertens | 40 |

dExample %>%

select(FirstName)

| FirstName |
|-----------|
| Abe |
| Bert |
| Bert |

dExample %>%

filter(FirstName == "Bert")

| FirstName | LastName | Age |
|-----------|----------|-----|
| Bert | Jans | 28 |
| Bert | Bertens | 40 |

dExample %>%

mutate(First_and_LastName = paste(FirstName, LastName))

| FirstName | LastName | Age | First_and_LastName |
|-----------|----------|-----|--------------------|
| Abe | Jans | 30 | Abe Jans |
| Bert | Jans | 28 | Bert Jans |
| Bert | Bertens | 40 | Bert Bertens |

dExample

| FirstName | LastName | Age |
|-----------|----------|-----|
| Abe | Jans | 30 |
| Bert | Jans | 28 |
| Bert | Bertens | 40 |

```
dExample %>%  
  select(Age) %>%  
  arrange(Age)
```

| Age |
|-----|
| 28 |
| 30 |
| 40 |

dExample %>%

```
  summarise(max_Age = max(Age),  
            min_Age = min(Age))
```

| max_Age | min_Age |
|---------|---------|
| 40 | 28 |

dExample %>%

```
  group_by(FirstName) %>%  
  summarise(size = n(),  
            max_Age = max(Age),  
            min_Age = min(Age))
```

| FirstName | size | max_Age | min_Age |
|-----------|------|---------|---------|
| Abe | 1 | 30 | 30 |
| Bert | 2 | 40 | 28 |



A calculated table

Goal: an R Chunk that calculates:

Choice of Program, per year

| Program | 2006 | 2017 |
|---|------|------|
| Research University - Bachelor | 21% | 32% |
| Research University - Master | 19% | 31% |
| University of Applied Sciences - Bachelor | 57% | 34% |
| University of Applied Sciences - Master | 3% | 3% |

Useful hints:

- build a pipeline
- start with the data `dStudents %>%`
- only use the rows with 2006 or 2017
- calculate per year the total
- then calculate the part
- format it like a percentage
- display with the caption

paste

%>%

kable

Year %in% c("2006", "2017")

spread

mutate

round

select

group_by

filter

A solution

```
```{r}
dStudents %>%
 filter(Year %in% c(2006, 2017)) %>%
 group_by(Year) %>%
 mutate(Total_per_year = sum(n),
 part = n / Total_per_year,
 part = paste0(round(100*part,0), "%")) %>%
 select(Program, Year, part) %>%
 spread(Year, part) %>%
 # caption in kable means title
 kable(caption = "Choice of Program, per year",
 align = "lrr")
...
```
```

Some base R and some variables

`data$ColumnName`

to access the column with `ColumnName` from your data

```
# The first year in the data  
min_year <- min(dStudents$Year)
```

1. In a new R Chunk named **calculations**, calculate:
 - a variable containing the minimum year
 - a variable containing the maximum year
2. Place this Chunk after the **readData** Chunk
3. Adjust the code in the Chunk with the table, so it will use the new variables
4. Knit and observe the word-doc looks the same



A solution

```
```{r calculations}  

The first year in the data
min_year <- min(dStudents$Year)

the most recent year in the data
max_year <- max(dStudents$Year)

```
```

```
```{r}  
dStudents %>%
 filter(Year %in% c(min_year, max_year)) %>%
 group_by(Year) %>%
 mutate(Total_per_year = sum(n),
```

# A problem



In the Chunk **calculations**, try:

```
diff_years <- max_year - min_year
```

Select all the code in the Chunk and let it run locally ( CTRL+enter)

```
{r calculations}

The first year in the data
min_year <- min(dStudents$Year)

the most recent year in the data
max_year <- max(dStudents$Year)

diff_years <- max_year - min_year
|

##

Error in max_year - min_year : non-numeric argument to binary operator
```

Character minus Character  
is not defined

Data	
dStudents	48 obs. of
Program	chr "Research Univ
Year	chr "2006" "2006" "2
n	num 6670 6042 17961 909
Values	
location_of_data	"data/Inco
max_year	"2017"
min_year	"2006"

An error!  
Why?



# Solution to this problem

`convert` If TRUE will automatically run `type.convert()` on the key column. This is useful if the column names are actually numeric, integer, or logical.

```
##{r readData}

read file IncomingInternationalDegreeStudents.csv

location_of_data <- file.path("data",
 "IncomingInternationalDegreeStudents.csv")
dStudents <- read_csv2(location_of_data) %>%
 # the data is not in tidy format,
 # reshape wide to long using tidyr::gather
 # the first column (Program) is ok, no gathering necessary
 gather(key = "Year", value = "n", - Program, convert = TRUE)

##
```

Adjust the code in the **readData** Chunk  
Let all the code run locally  
Observe that the error message is gone



Environment	History	Connections
Import Dataset ▾		
Global Environment ▾		
Data		
dStudents	48 obs. of 3	
Program	chr "Research Univer	
Year	int 2006 2006 2006 200	
n	num 6678 6642 17961 909 7	
Values		
diff_years	11L	
location_of_data	"data/Incomi	
max_year	2017L	
min_year	2006L	

# Inline R code

Use an R variable in the text: ``r min_year``

Goal: Backtick r variable backtick



## Introduction

In 2006, 31582 international students were enrolled in a higher education institution in the Netherlands. In 2017 there were 76908, a growth of 144% over a period of 11 years.

In the R Chunk **calculations**:

- calculate variables for all underlined text  
(hint: calculate a dataframe with totals per year)
- add the text with the variables in the body of the RMarkdown

```

{r calculations}

The first year in the data
min_year <- min(dStudents$Year)

the most recent year in the data
max_year <- max(dStudents$Year)

diff_years <- max_year - min_year

totals per year, only for the min and the max year
dTotalPerYear <- dStudents %>%
 group_by(Year) %>%
 summarise(Total = sum(n)) %>%
 filter(Year %in% c(min_year, max_year))

base R for calculating the total in max_year
total_max_year <- dTotalPerYear$Total[dTotalPerYear$Year == max_year] %>%
 # the counts were numeric, we want an integer
 as.integer()

another way for the min_year
total_min_year <- dTotalPerYear %>%
 filter(Year == min_year) %>%
 select(Total) %>%
 # now still a dataset, we want only the value
 as.integer()

the growth
growth_between <- (total_max_year - total_min_year) / total_min_year
and now formatted as percentage
growth_between <- sprintf("%.0f%%", 100* growth_between)

...

```

## # Introduction

In `r min\_year`, `r total\_min\_year` international students were enrolled in a higher education institution in the Netherlands. In `r max\_year` there were `r total\_max\_year`, a growth of `r growth\_between` over a period of `r diff\_years` years.

## Introduction

In 2006, 31582 international students were enrolled in a higher education institution in the Netherlands. In 2017 there were 76908, a growth of 144% over a period of 11 years.

*Choice of Program, per year*

Program	2006	2017
Research University - Bachelor	21%	32%
Research University - Master	19%	31%
University of Applied Sciences - Bachelor	57%	34%
University of Applied Sciences - Master	3%	3%

# The knitted doc as start of template



## Incoming Degree Students

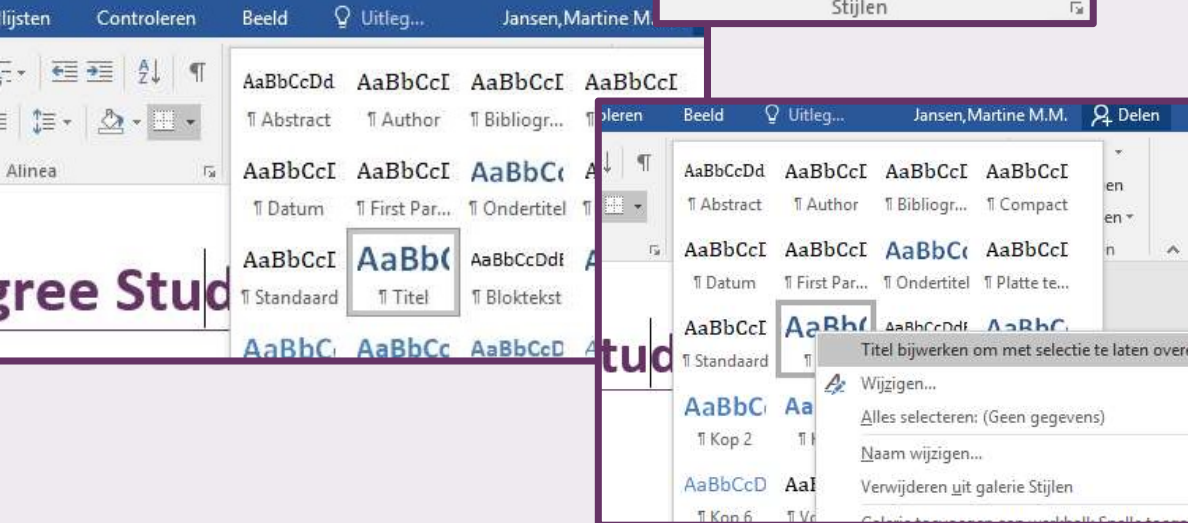
Martine Jansen

30 januari 2019

AaBbCcDd AaBbCcI AaBbCcI

¶ Abstract ¶ Author ¶ Bibliogr...

Stijlen



Adjust the styles in this doc:

- Select some text and format it
- Place cursor on formatted text
- Go to “Styles” menu, click open
- Search for the selected style
- Right click, choose “Update ... to match selection

When done formatting,  
save doc in the project folder as  
template.docx

# Add a reference\_docx to the Rmd



```

title: "Incoming Degree Students"
author: "Martine Jansen"
date: "30 januari 2019"
output:
 word_document:
 reference_docx: template.docx

```

Knitting will display the doc with the styles in your template

important:

- indent 2 spaces
- 2nd line 2 more

## Incoming Degree Students

Martine Jansen

30 januari 2019

### 1. Introduction

In 2006, 31582 international students were enrolled in a higher education institution in the Netherlands. a growth of 144% over a period of 11 years.

*Choice of Program, per year*



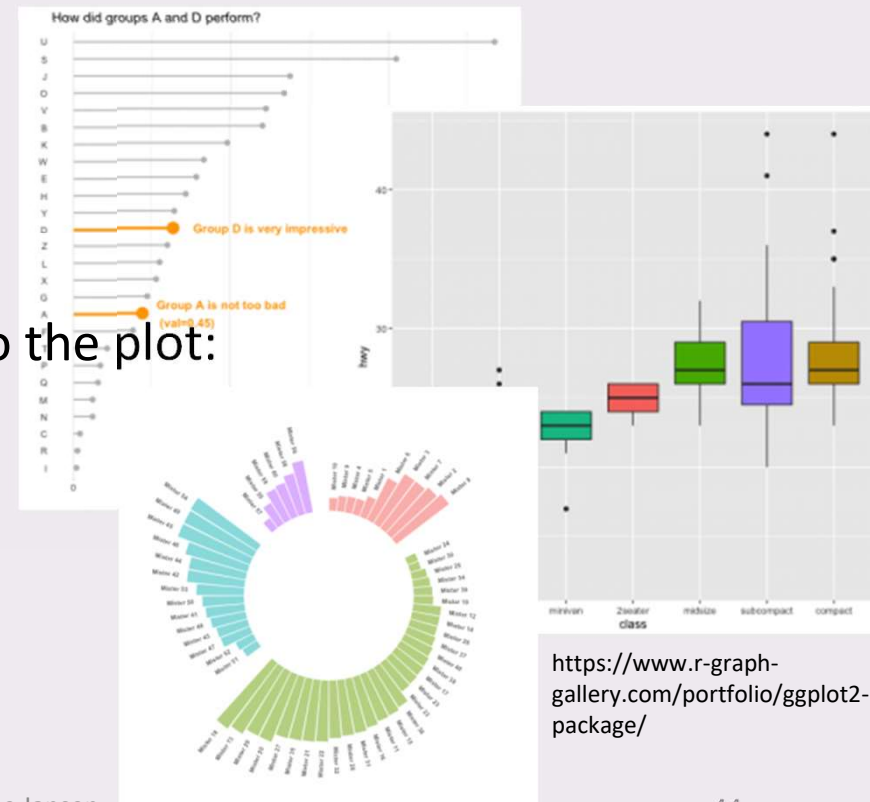
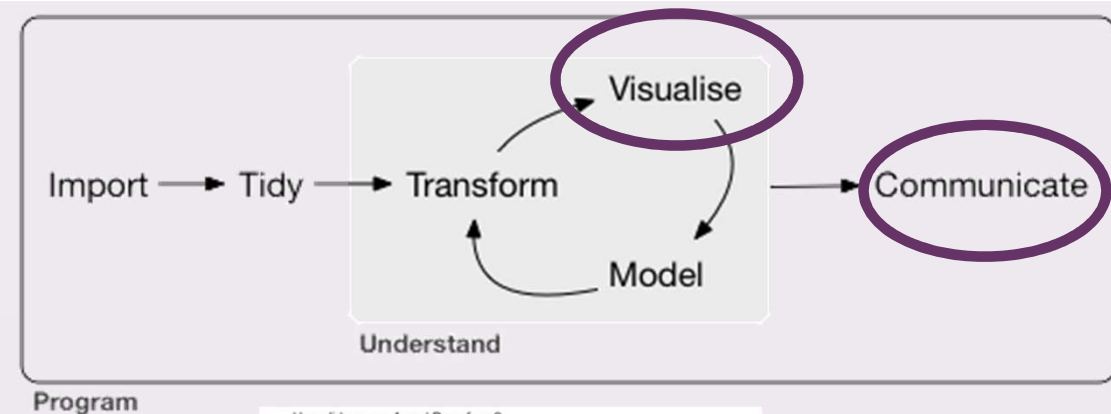
# ggplot

ggplot2 is a system for declaratively creating graphics, based on The Grammar of Graphics.

<https://ggplot2.tidyverse.org/>

A ggplot has

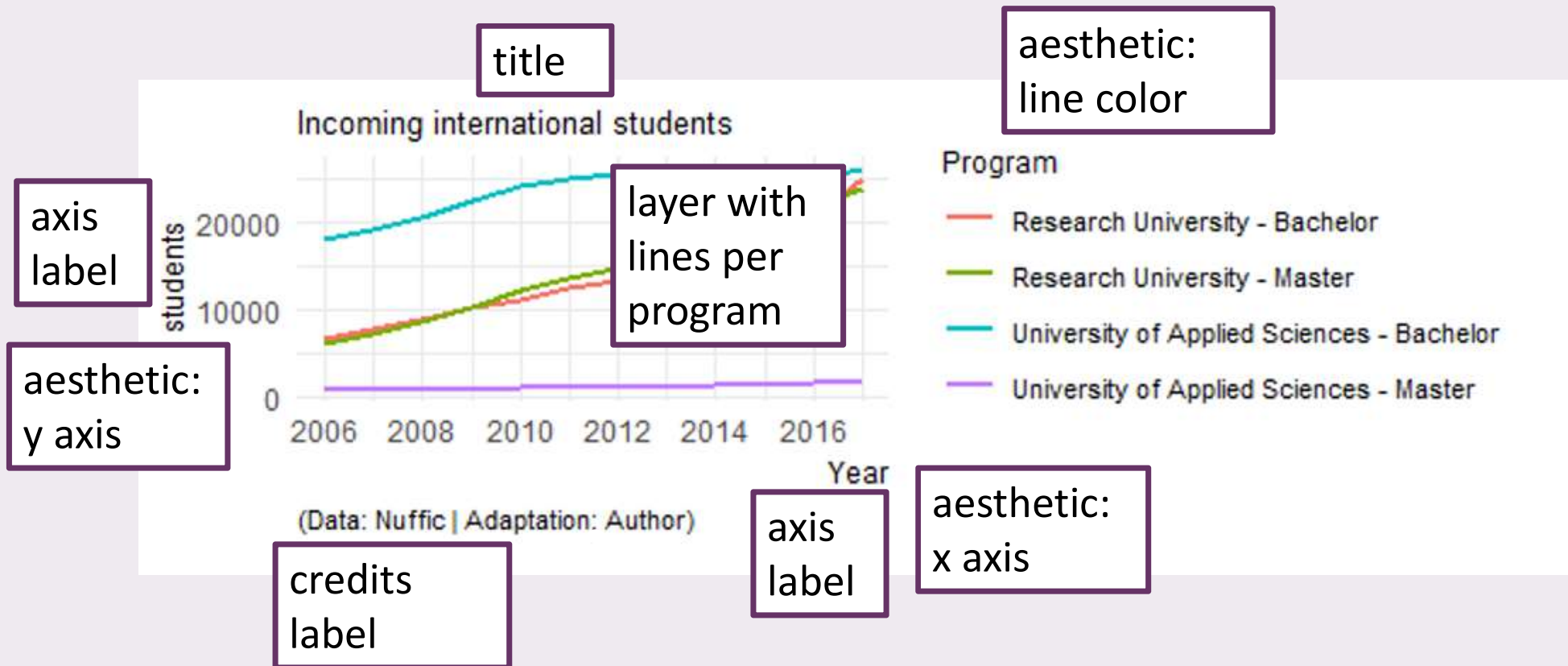
- data
- mapping: how variables of the data are mapped to the plot:
  - x and y axis
  - color or size of lines and points
  - alpha, for transparency settings
- one or more layers, with the actual plot parts
- a lot of possible formatting options



<https://www.r-graph-gallery.com/portfolio/ggplot2-package/>



# A ggplot with its parts



# A ggplot for the document

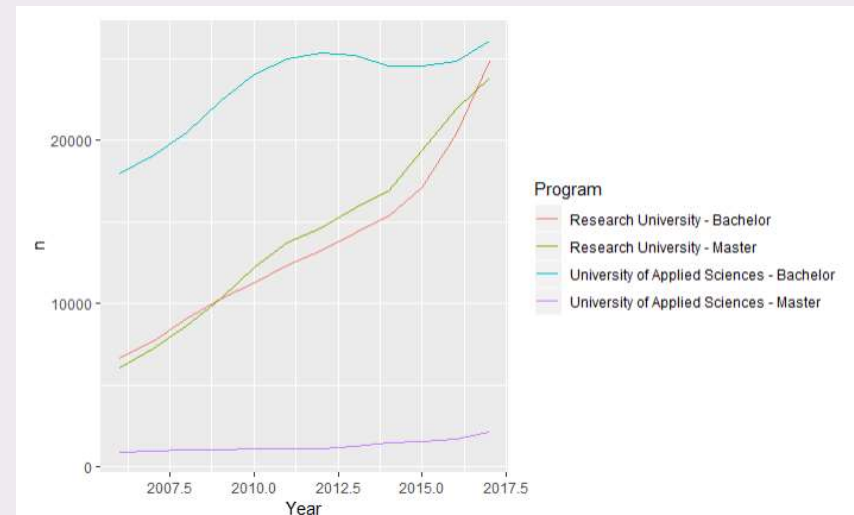
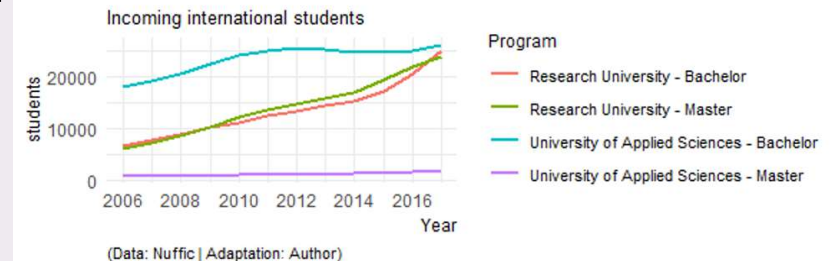
- Add a new R Chunk, name it **plotperprogram**
- Add the code below, and fill in the spaces

```
dStudents %>%
 ggplot(aes(x = Year,
 y = n,
 color = Program)) +
 geom_line()
```

- Let the code run locally and see the result, right below the Chunk



Reminder:



# Some improvements on the plot

```
dStudents %>%
 ggplot(aes(x = Year, y = n, color = Program)) +
 geom_line(size = 1) +
 # better values on the x axis:
 scale_x_continuous(breaks = seq(from = min_year, to = max_year, by = 2)) +
 # add titles and better labels
 labs(y = "students",
 title = "Incoming international students",
 # caption in ggplot is used for source & credits| information
 caption = "(Data: Nuffic | Adaptation; Author)") +
 # another theme
 theme_minimal() +
 # caption on the left, label x axis on the right
 # and set some font sizes
 theme(plot.caption = element_text(hjust = 0, size = 8),
 plot.title = element_text(size = 10),
 axis.title = element_text(size = 9),
 axis.title.x = element_text(hjust = 1),
 legend.title = element_text(size = 9),
 legend.text = element_text(size = 8))
```



Add this code  
per line,  
and after each  
added line,  
CTRL+enter it so  
you can see the  
changes.

Save yourself some  
typing, and do not  
type the comment  
lines 😊

# The last tweaks

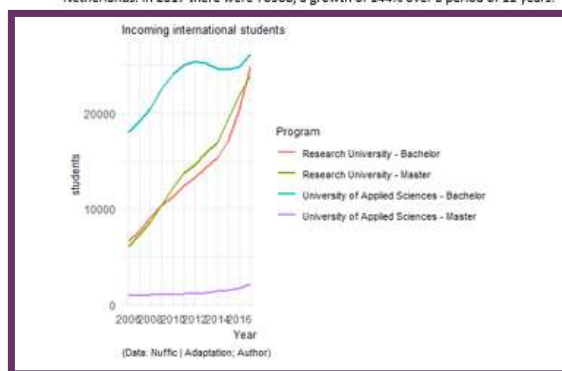


## Incoming Degree Students

Martine Jansen  
30 januari 2019

### 1. Introduction

In 2006, 31582 international students were enrolled in a higher education institution in the Netherlands. In 2017 there were 76908, a growth of 144% over a period of 11 years.



Choice of Program, per year

Program	2006	2017
Research University - Bachelor	21%	32%
Research University - Master	19%	31%
University of Applied Sciences - Bachelor	57%	34%
University of Applied Sciences - Master	3%	3%

The plot has not a nice aspect ratio  
And there are no headers 2 yet

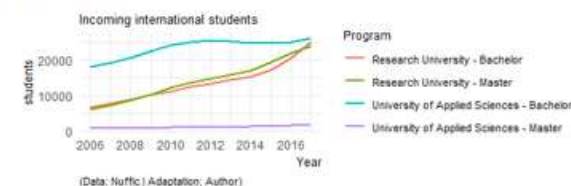
## Incoming Degree Students

Martine Jansen  
30 januari 2019

### 1. Introduction

In 2006, 31582 international students were enrolled in a higher education institution in the Netherlands. In 2017 there were 76908, a growth of 144% over a period of 11 years.

#### 1.1 A figure



#### 1.2 A table

Choice of Program, per year

Program	2006	2017
Research University - Bachelor	21%	32%
Research University - Master	19%	31%
University of Applied Sciences - Bachelor	57%	34%
University of Applied Sciences - Master	3%	3%

Header 2 is not formatted yet.  
Format it, save the document as template.docx.  
Knit again, see the difference.

Save the final word doc in a folder results

```
A figure
```

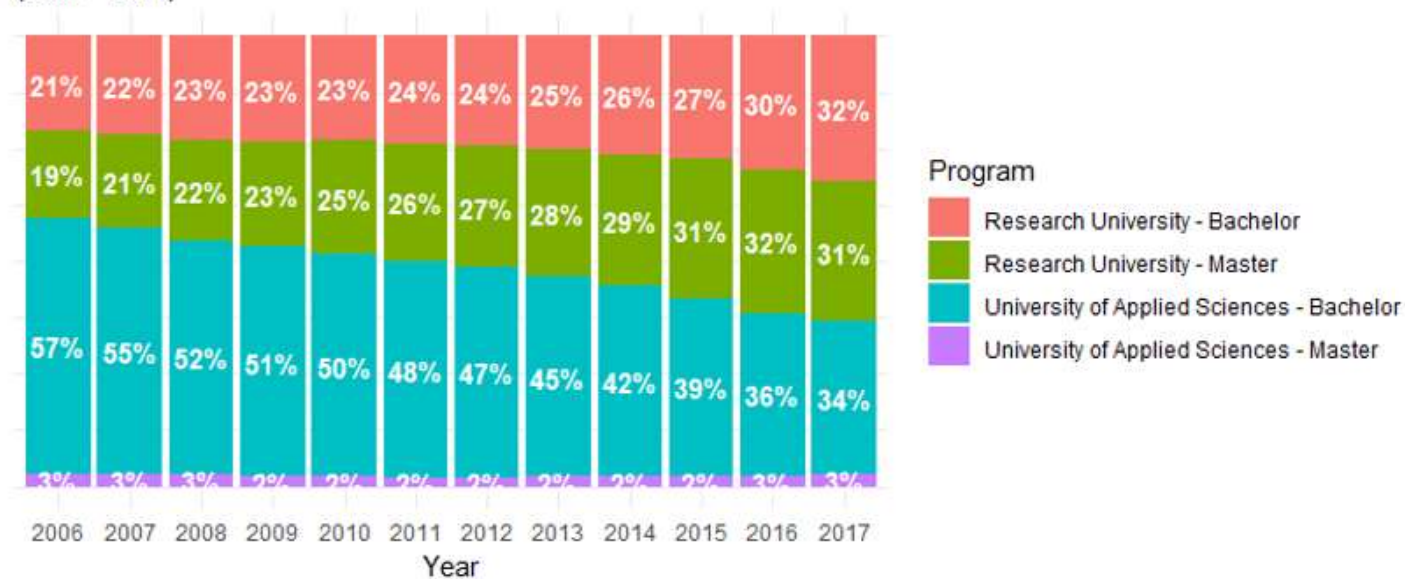
```
{r plotperprogram, fig.height = 2, fig.width = 6}
```

# In case there is any time left



## 1.3 Extra

Relative more students in NL for Research University  
(2006 - 2017)




(Data: Nuffic | Adaptation; Author)

Reconstruct  
this figure in  
the R  
Markdown  
document

Hint:

1. Calculate the percentages per year
2. Make the barplot
3. Add the extra info

This completes this workshop

This is only a start t

More:

R for Data Science  
Grolemund/ Wickham

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

Lessons R Markdown  
Rstudio

<https://rmarkdown.rstudio.com/lesson-1.html>

R Markdown: The Definitive Guide  
Xie/Allaire/Wickham

<https://bookdown.org/yihui/rmarkdown/>

R Graphics Cookbook  
Chang

<http://www.cookbook-r.com/Graphs/>

**Thank you,  
and have lots of  
fun with R and  
R Markdown**