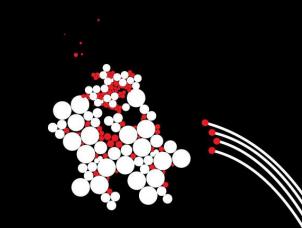
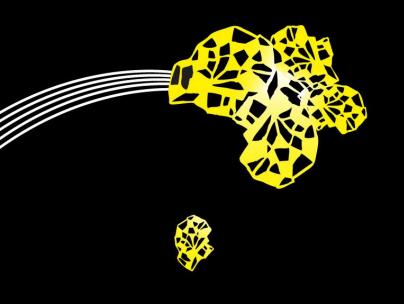
UNIVERSITY OF TWENTE.



INTRODUCTION INTO WORKING WITH R

SESSION 1 – VERSION 17/11/2019

BENJAMIN ZIEPERT





INTRODUCTION INTO WORKING WITH R SESSION 1

Lecturers: Benjamin Ziepert Authors: Benjamin Ziepert & Dr. Elze G. Ufkes

The course will:

- Teach you the basics of R
- Practice an advanced data-analyses that can't be done with SPSS
- Enable you to further study R on your own

The course will not:

Enable you to do all statistical analysis in R

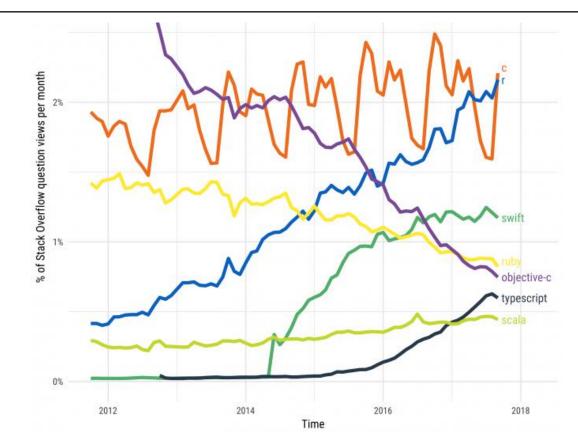


Open Source

- Powerful and flexible
- The standard for data science

Programming becomes more important in the workplace and as teachers we want to prepare you for that reality.

WHY R? R GROWTH



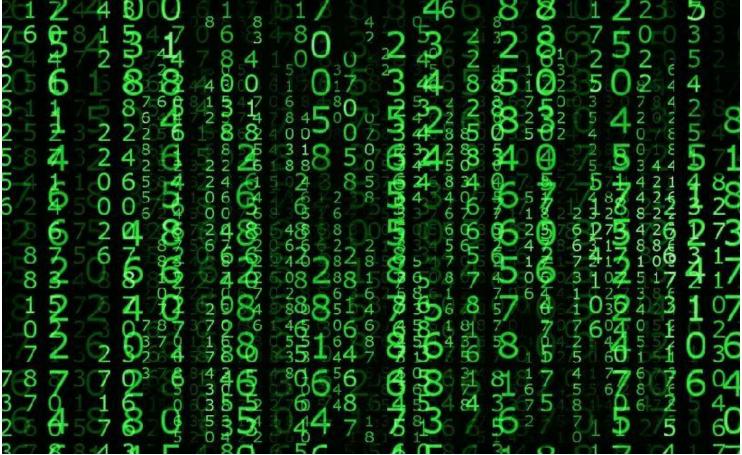
Source: stackoverflow.blog

WHY R? COMPANIES USING R



Source: listendata.com





MAKE AN INVESTMENT

"Learning to code is empowering and can hugely improve a researcher's career prospects. But it does require an investment"



ANTICIPATE HURDLES IN THE BEGINNING

"Typos, for example, bring work to a standstill, she says. They didn't put a space and the script won't run; they put two dashes and the script won't run."



PLAN CODING TIME WITH PEERS

"... people [should] pick a language that's popular with their colleagues and work initially in four-hour blocks, which he says provide enough time to work through hurdles and get a sense of progress."



SEEK HELP FROM THE START

Perhaps the biggest barrier is insecurity ... "Many people think, I'll just figure it out on my own first. I'm not good enough yet to ask questions'," she says. Instead, they should seek help from others to gain more skills.





PLANNING



- 1. Learn the benefits
- 2. Getting up to speed with the basics of R
 - Create figures
 - Run analysis
 - Basic R coding knowledge
- 3. Getting introduced to the extensive possibilities of R
 - Completing a R-project wherein you challenge yourself

PLANNING OVERVIEW

3 Lectures

- Introduction into R
- Statistical analysis
- Analyzing social media content

2 Self-study assignment's using DataCamp

Reading

- R is for Revolution (Culpepper & Aguinis, 2010)
- Scientific computing: Code alert (Baker, 2017)

PLANNING OVERVIEW

Passing requirements

- Attendance of all sessions
- Complete DataCamp assignments with at least 8000 XP (Self-study)
- Complete R script assignment with statistical analysis (Session 2)
- Complete Twitter analysis and present results (Session 3)

PLANNING TODAY

- Introduction in R
- Graphics
- Statistical analysis
- Preparing next lecture

R BASICS SOFTWARE

R

- Core software
- https://cloud.r-project.org

RStudio

- Integrated development environment (IDE) for R
- https://www.rstudio.com

R BASICS

Let's have a look at the software.

✓ Please open RStudio now.



	RStudio		
🔍 📲 🖕 🗐 😬 🍺 Gu to The Function 💿 🔛 🔸 Addine 🔸			🕓 Project. (None) =
• Unotified 1 ×		Environment History	
🗇 🗇 👩 😝 🔿 Source or Save 💁 者 + 🕄 🔸	🕞 Run 📑 🕞 Source + 🖻	🔐 🔒 📑 Import Dataset - 🍯	= Lat - 😨
1		Global Environment +	(9,
		Endramont is en	pty
		Hiles Plats Packages Help Viewer	
1:1 (Top Level) :	K Senat 3		
Constant (Construction) of Version 2-2-2 constructions for Startistical Corputing Copyright (C) 2018 The & Readering for Startistical Corputing Plantform: Side_add-construction.4 (C-14):0 Readering - Side_add-construction.4 (C-14):0 Readering - Side_add-construction.5 (Side_add-constructions) Type "NiconseQ" or "NiconseQ" for distribution details.			
Netural language support but running in an English locale K is a collaborative project with many contributors. Type 'contributors' for new Information call 'cluster()' on how to clus R or R packages in publications.			
Type 'demo()' for some demos, 'help()' for on-line help, or 'ne(p,storr()' for on HTML browser interface to help. Type 'q()' to quit R.			
>			

R BASICS RSTUDIO



•••	RStudio	
🔍 • 🚰 • 🔒 🔝 🚔 🏕 Go to file/function 👘 🔛 • Addins •		🕓 Project: (None) 🖛
Untitled1 ×	-3	Environment History
🗢 🗇 🔊 🔒 🗍 Source on Save 💁 🖉 • 🗐 🔹	🔿 Run 📑 🕞 Source 🔹 🚍	確 🔒 📑 Import Dataset - 🍯
1		Global Environment - Q.
	Contrat	
	Script	
		· · · ·
		Files Plots Packages Help Viewer
		(1 1) 2 Zoom Z Export - 9 4
		Output
1:1 (Top Level) +	R Script +	
Console ~/Downloads/ 🥪 N version 3.5.2 (2010-10-51) Suicere Punckin Patch		
Copyright (C) 2016 The R Foundation for Statistical Computing	Console	
Platform: x86_64-apple-darwin13.4.8 (64-bit)		
R is free software and comes with ABSOLUTELY NO WARRANTY.		
You are welcome to redistribute it under certain conditions. Type 'license()' or 'licence()' for distribution details.		
Noticel leaves comment but comming in an Earlich leaves		
Natural language support but running in an English locale		
R is a collaborative project with many contributors. Type 'contributors()' for more information and		
'citationO' on how to cite R or R packages in publications.		
Type 'demo()' for some demos, 'help()' for on-line help, or		
'help.start()' for an HTML browser interface to help.		
Type 'q()' to quit R.		
>		

R BASICS

- Run line or selection: [Cmd] / [Ctrl] + [Enter]
 - Code will be transferred to the console and runs there
- Document your code well with comments
 - Characters that come after # are skipped
- Be precise, punctuation and capitalization is important
 - DataBase ≠ database

Save	🗐 📄 📄 Source on Save
	± + 1
	(Top Level) \$
	le ~/ ≈

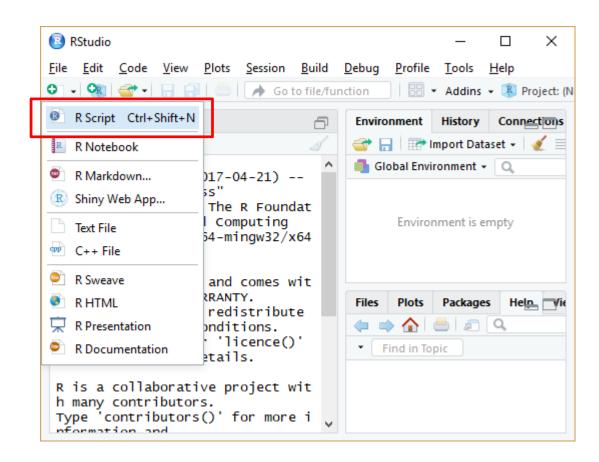
R BASICS

- ✓ Go now to <u>benjaminziepert.com/teaching</u>
- $\checkmark\,$ Download all files and save them in one folder
- \checkmark Open Session 1 \rightarrow Handout R basics: statistical graphs and analysis

R BASICS CREATE SCRIPT FILE

✓ Open R Studio
✓ Create R Script
✓ Save R Script

Tip: save all files in one location



R BASICS

1 INSTALLING AND ACTIVATING PACKAGES

- Packages add functionality to R
- Use install.packages()
 - For instance: install.packages("tidyverse")
 - You only have to install the package once
 - When asked, decline to install from source package or to compile a package.
 - Installation doesn't work? Check the FAQ.

✓ Copy the text from the gray box in the handout to your R file and then run the line with [Cmd] / [Ctrl] + [Enter].

R BASICS 1 INSTALLING AND ACTIVATING PACKAGES

RStudio Menu alternative

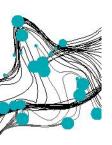
Profile	Tools	Help					
🗄 👻 Addin	Ins	stall Packages					
	Cł	eck for Package	Updates				
	Version Control						

Install Packages								
Install from:	⑦ Configuring Repositories							
Repository (CRAN)	•							
Packages (separate multiple	with space or comma):							
tidyvers								
tidyverse prary:								
C:/R/Library [Default]	•							
Install dependencies								
	Install Cancel							

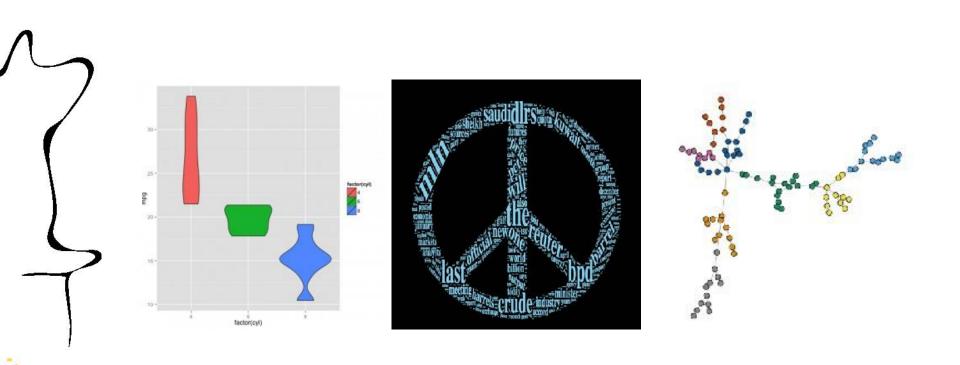
R BASICS 1 INSTALLING AND ACTIVATING PACKAGES

Activate the package using **library()**

You have to do this every time / session you want to use the package



GRAPHICS CREATING (YOUR FIRST?) R VISUALIZATION



Source: r-graph-gallery.com

GRAPHICS

2.1 OPEN THE DATA FRAME MPG

✓ Run library("ggplot2")

✓ Run mpg to open the data frame

mpg is a data set for the fuel economy data from 1999 and 2008 for 38 popular car models

		manufacturer	mod	lel	displ	year	cyl	trans	drv	cty	hwy	fl	class
		<chr></chr>	<ch< td=""><td>ir></td><td><dbl></dbl></td><td><int></int></td><td><int></int></td><td><chr></chr></td><td>< chr ></td><td><int></int></td><td><int></int></td><td>< chr ></td><td>< chr ></td></ch<>	ir>	<dbl></dbl>	<int></int>	<int></int>	<chr></chr>	< chr >	<int></int>	<int></int>	< chr >	< chr >
	1	audi	a4		1.80	1999	4	auto(15)	f	18	29	р	compac
	2	audi	a4		1.80	1999	4	manual(m5)	f	21	29	р	compac
	3	audi	a4		2.00	2008	4	manual(m6)	f	20	31	p	compac
	4	audi	a4		2.00	2008	4	auto(av)	f	21	30	р	compac
	5	audi	a4		2.80	1999	6	auto(15)	f	16	26	р	compac
	6	audi	a4		2.80	1999	6	manual(m5)	f	18	26	р	compac
	7	audi	a4		3.10	2008	6	auto(av)	f	18	27	p	compac
	8	audi	a4	quattro	1.80	1999	4	manual(m5)	4	18	26	p	compac
	9	audi	a4	quattro	1.80	1999	4	auto(15)	4	16	25	p	compac
1	.0	audi	a4	quattro	2.00	2008	4	manual(m6)	4	20	28	p	compac

GRAPHICS HOW TO CREATE A VISUALIZATION?

How can we visualize this data?

- For instance, what is the frequency of engine sizes?
- \rightarrow We use the graphics package ggplot2

ggplot2 was installed with tidy verse packages and is used for graphics.

GRAPHICS 2.2 HISTOGRAM

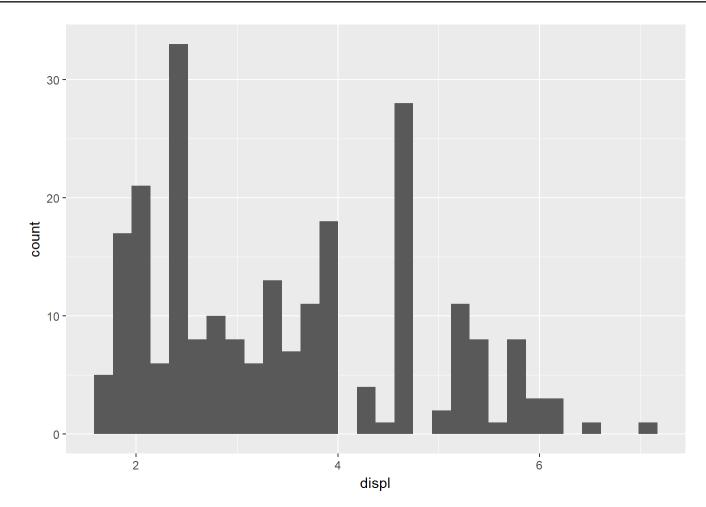
Creates coordinate system based on a data frame

Adds a layer of some geometric object

ggplot(data = mpg) + aes(displ) + geom_histogram()

Specifies mapping of variables in the data frame onto aesthetic attributes

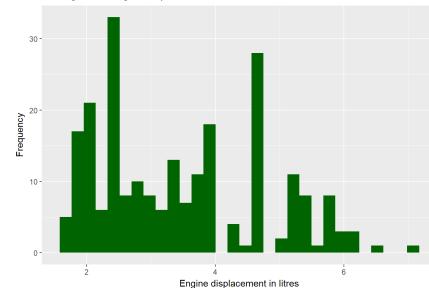




GRAPHICS 2.3 UPDATE LABELS AND COLOR

```
ggplot(data = mpg) +
aes(x = displ) +
geom_histogram(fill = 'darkgreen') +
labs(title = "Histrogram of engine displacement",
        x = "Engine displacement in litres",
        y = "Frequency")
```

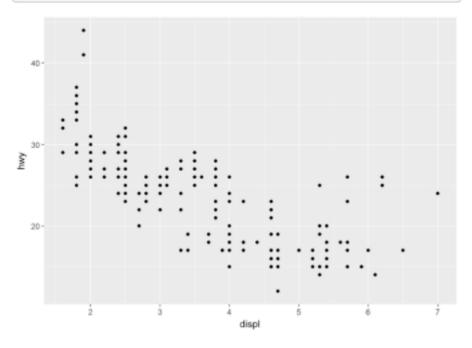
geom_histrogram() is now
filled with an color and labels
(labs()) are added.



Histrogram of engine displacement

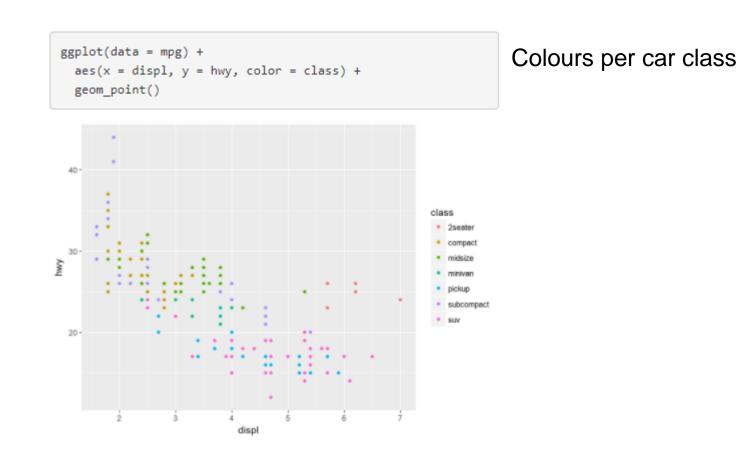
GRAPHICS 2.4 CREATE A SCATTER DOT

```
ggplot(data = mpg) +
aes(x = displ, y = hwy) +
geom_point()
```



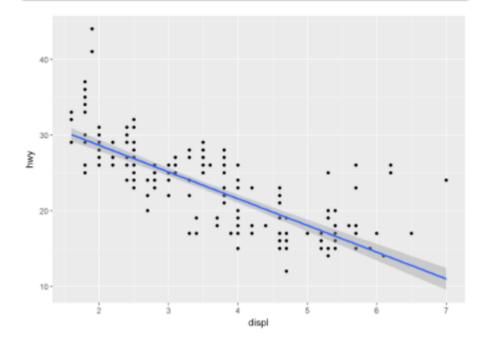
geom_histrogram() is now
replaced with geom_point()
and we added hwy to the
variables.

GRAPHICS 2.5 ADDING MORE AESTHETIC MAPPINGS



GRAPHICS 2.6 ADDING REGRESSION LINE

```
ggplot(data = mpg) +
aes(x = displ, y = hwy) +
geom_point() +
geom_smooth(method=lm)
```



geom_smooth(method=lm)

What does this graph tell us?

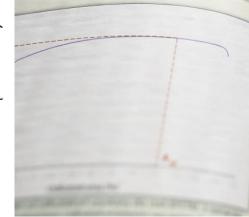
You can find more info about graphics at

- <u>http://www.sthda.com/englis</u>
 <u>h/wiki/ggplot2-essentials</u>
- <u>http://www.r-graph-gallery.com</u>

STATISTICS DESCRIPTIVE, CORRELATION & LINEAR

do mile monovanego Mozna powiedzieć, że portusza się w zbiorze

there expanses many fills of the second seco



c) Rys. 21 Histogramy przedsta z polskich hut w ramach n EPSTAL o średnic żebrowanych EPSTAL o śr anica plastyczności (m a) R. - granica plastyczności wytrzymałość na rozci b) R_m - wytrzymałość na ro - wydłużenie przy mak c) A - wydłużenie przy maksy -odkształcenie PN-EN 10080/2005), który wymaga | wstępne badania wyrobów, inspeksiç do dlug od producenta prowadzenia wewnętrz- cję zakładu, ciągły nadzór oraz - co jakoici, co nej kontroli produkcji, wykonywania wyróżnia system 1+ od pozostałych wykarapt prees ningo unuprheisjarych badań - badania sondażowe próbek pobraprofick polymersch w exhlution area aych w zakładzie, w obrocie lub na

STATISTICS 3.1 DESCRIPTIVE STATISTICS

summary(mpg)

##	manufacturer	model	displ	year
##	Length:234	Length:234	Min. :1.600	Min. :1999
##	Class :characte	er Class :characte	er 1st Qu.:2.400	1st Qu.:1999
##	Mode :characte	er Mode :characte	er Median :3.300	Median :2004
##			Mean :3.472	Mean :2004
##			3rd Qu.:4.600	3rd Qu.:2008
##			Max. :7.000	Max. :2008
##	cyl	trans	drv	cty
##	Min. :4.000	Length:234	Length:234	Min. : 9.00
##	1st Qu.:4.000	Class :character	Class :character	1st Qu.:14.00
##	Median :6.000	Mode :character	Mode :character	Median :17.00
##	Mean :5.889			Mean :16.86
##	3rd Qu.:8.000			3rd Qu.:19.00
##	Max. :8.000			Max. :35.00
##	hwy	fl	class	
##	Min. :12.00	Length:234	Length:234	
##	1st Qu.:18.00	Class :character	Class :character	
##	Median :24.00	Mode :character	Mode :character	
##	Mean :23.44			
##	3rd Qu.:27.00			
##	Max. :44.00			

STATISTICS LINEAR STATISTICS

3.3 Independent T-Test

t.test(x, y)

3.4 One Way Anova

- aov(y ~ x, data =
 mydata)
- 3.5 Multiple Linear regression
- Im(y ~ x1 + x2 + x3, data = mydata)

Formula

- $\bullet \ y = x_1 + \dots + x_k$
- $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon$

More statistics:

- <u>https://www.statmethod</u>
 <u>s.net/stats/index.html</u>
- Discovering Statistics
 Using R by Andy Field.





- At home: DataCamp assignment
- Now: Check R and RStudio installation

NEXT LECTURE SELF-STUDY ASSIGNMENTS



Complete the 3 assignments before the day of the next lecture:

- 1. Introduction to R (4 hours)
 - Whole course
- 2. Importing data (2 hours)
 - Only do the chapter "Importing data from statistical software packages" in the course "Importing Data in R (Part 2)"
- 3. Bring at least one question for the Q&A next lecture

To pass the DataCamp assignments your XP must stay above 7000.

 Therefore, try to understand what you do before clicking on hint or show solution.

NEXT LECTURE

PREPARING AND CHECKING INSTALLATION

- ✓ Make sure R, RStudio and Rtools (windows only) are up to date.
- Please install or update the following packages: "tidyverse", "ggplot2", "Hmisc", "twitteR", "tm", "wordcloud", "psych", "devtools" and "gplots".
- ✓ Update all packages
- ✓ Open "S01F03 Test Package Installation.R" and call me.

ADDITIONAL INFORMATION

Check the R Studio Cheat sheets: <u>Base R</u>, <u>R Studio</u> & <u>more</u> ...

Statistics

- https://www.statmethods.net/stats/index.html
- Discovering Statistics Using R by Andy Field.

Graphics

- <u>http://www.sthda.com/english/wiki/ggplot2-essentials</u>
- <u>http://www.r-graph-gallery.com</u>