RStudio and R Resources

Happy Scientist Seminar

Emil Hvitfeldt

2019-1-22

Motivation

We will talk about what you are able to do in RStudio with R and where to find information and rescources to do it great.



- Data analysis scripts
- Interactive web applications
- Documents
- Reports
- Graphs
- more

- Syntax highlighting
- code completion
- smart indentation
- Integrated R help
- data viewer
- Version control with Git

R File Edit Code View Plots Session Build Debug Profile Tools Help		emilhans 🕞 🕘
🔍 💿 🗸 🦓 🧁 🖌 🔒 🚔 🖌 Go to file/function		Project: (None) -
Untitled1* ×	_	Environment History Connections
🔄 🖓 🖉 🔚 🖓 🔍 🖋 Knit 🗸 🔅 🗸 👘 🔞 Insert 🗸 🖓 🖓	in • 🧐 • 🗏	area and a second
13 14 When you click the **Knit** button a document will be generated that includes both content as output of any embedded R code chunks within the document. You can embed an R code chunk like t 15 16 * ```{r cars} 17 library(gapminder) 18 summary(gapminder)	well as the chis: ② ≚ ▶	<pre>summary(gapminder) ggplot(gapminder, aes(pop, lifeExp)) + geom_point() library(ggplot2) ggplot(gapminder, aes(pop, lifeExp)) + geom_point()</pre>
19 country continent year lifeExp pop Afghanistan: 12 Africa :624 Min. :1952 Min. :23.60 Min. :6.001e+04 Albania : 12 Americas:300 1st Qu.:1966 1st Qu.:48.20 1st Qu.:2.794e+06 Algeria : 12 Europe:360 Median :19.47 Median: :0.71 Median: :0.960e+07 Argentina : 12 Oceania: :24 3rd Qu.:1993 3rd Qu.:70.85 3rd Qu.:1.959e+07 Australia : 12 Max. :2007 Max. :82.60 Max. :1.319e+09	.el ≪ ×	<pre>mtcars ggplot(mtcars, aes(mpg, disp, color = cyl)) + geom_point() ggplot(mtcars, aes(mpg, disp, color = factor(cyl))) + geom_point() ggplot(mtcars, aes(mpg, disp, color = factor(cyl))) + geom_point()</pre>
		Files Plots Packages Help Viewer
Min. : 241.2 1st Qu.: 1202.1 Median : 3531.8 Mean : 7215.3 3rd Qu.: 9325.5 Max. :113523.1		400-
29:15 🙆 Chunk 3: pressure 🗢	R Markdown ‡	
Console Terminal ×		
~/ 🖄		
3rd Qu.: 9325.5 Max :113523 1		300 - factor(cyl)
<pre>> ggplot(gapminder, aes(pop, lifeExp)) + + geom_point() Error in ggplot(gapminder, aes(pop, lifeExp)) : could not find function "ggplot" > library(ggplot2) > > ggplot(gapminder, aes(pop, lifeExp)) + + geom_point() > mtcars > ggplot(mtcars, aes(mpg, disp, color = cyl)) + + geom_point() > ggplot(mtcars, aes(mpg, disp, color = factor(cyl))) +</pre>		
<pre>+ geom_point() > gaplot(mtcars, aes(mpg, disp, color = factor(cvl))) +</pre>		
<pre>> geom_point() > </pre>		10 15 20 25 30 35 mpg









Packages

Idea -> Code -> Package -> Share -> Happiness

If you want to do something, chance has it someone have worked on it before

CRAN (The Comprehensive R Archive Network)



https://cran.r-project.org/

https://cran.rproject.org/web/packages/available_packages_by_name.html

Available CRAN Packages By Name

<u>A B C D E F G H I J K L M N O P Q R S T U V W X Y Z</u>

<u>A3</u>	Accurate, Adaptable, and Accessible Error Metrics for Predictive Models	
<u>abbyyR</u>	Access to Abbyy Optical Character Recognition (OCR) API	
<u>abc</u>	Tools for Approximate Bayesian Computation (ABC)	
abc.data	Data Only: Tools for Approximate Bayesian Computation (ABC)	
<u>ABC.RAP</u>	Array Based CpG Region Analysis Pipeline	
<u>ABCanalysis</u>	Computed ABC Analysis	
abcdeFBA	ABCDE_FBA: A-Biologist-Can-Do-Everything of Flux Balance Analysis with this package	
<u>ABCoptim</u>	Implementation of Artificial Bee Colony (ABC) Optimization	
ABCp2	Approximate Bayesian Computational Model for Estimating P2	
abcrf	Approximate Bayesian Computation via Random Forests	
abctools	Tools for ABC Analyses	
abd	The Analysis of Biological Data	
abe	Augmented Backward Elimination	
<u>abf2</u>	Load Gap-Free Axon ABF2 Files	
ABHgenotypeR	Easy Visualization of ABH Genotypes	
abind	Combine Multidimensional Arrays	
<u>abjutils</u>	Useful Tools for Jurimetrical Analysis Used by the Brazilian Jurimetrics Association	
<u>abn</u>	Modelling Multivariate Data with Additive Bayesian Networks	
abnormality	Measure a Subject's Abnormality with Respect to a Reference Population	
abodOutlier	Angle-Based Outlier Detection	
<u>ABPS</u>	The Abnormal Blood Profile Score to Detect Blood Doping	
AbsFilterGSEA	Improved False Positive Control of Gene-Permuting GSEA with Absolute Filtering	
AbSim	Time Resolved Simulations of Antibody Repertoires	
<u>abundant</u>	High-Dimensional Principal Fitted Components and Abundant Regression	
Ac3net	Inferring Directional Conservative Causal Core Gene Networks	
ACA	Abrupt Change-Point or Aberration Detection in Point Series	11
acc	Exploring Accelerometer Data	
accoloromotry	Functions for Processing Accelerometer Data	

https://cran.r-project.org/web/views/

CRAN Task Views

CRAN task views aim to provide some guidance which packages on CRAN are relevant for tasks related to a certain topic. They give a brief overview of the included packages and can be automatically installed using the <u>ctv</u> package. The views are intended to have a sharp focus so that it is sufficiently clear which packages should be included (or excluded) - and they are *not* meant to endorse the "best" packages for a given task.

- To automatically install the views, the <u>ctv</u> package needs to be installed, e.g., via install.packages("ctv") and then the views can be installed via install.views or update.views (where the latter only installs those packages are not installed and up-to-date), e.g., ctv::install.views("Econometrics") ctv::update.views("Econometrics")
- The task views are maintained by volunteers. You can help them by suggesting packages that should be included in their task views. The contact e-mail addresses are listed on the individual task view pages.
- For general concerns regarding task views contact the <u>ctv</u> package maintainer.

Topics

<u>Bayesian</u>	Bayesian Inference
ChemPhys	Chemometrics and Computational Physics
<u>ClinicalTrials</u>	Clinical Trial Design, Monitoring, and Analysis
Cluster	Cluster Analysis & Finite Mixture Models
Databases	Databases with R
DifferentialEquations	Differential Equations
Distributions	Probability Distributions
Econometrics	Econometrics
Environmetrics	Analysis of Ecological and Environmental Data
ExperimentalDesign	Design of Experiments (DoE) & Analysis of Experimental Data
ExtremeValue	Extreme Value Analysis
Finance	Empirical Finance
FunctionalData	Functional Data Analysis
Genetics	Statistical Genetics
<u>Graphics</u>	Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization
HighPerformanceComputing	High-Performance and Parallel Computing with R
<u>Hydrology</u>	Hydrological Data and Modeling
MachineLearning	Machine Learning & Statistical Learning
MedicalImaging	Medical Image Analysis
<u>MetaAnalysis</u>	Meta-Analysis
<u>MissingData</u>	Missing Data
<u>ModelDeployment</u>	Model Deployment with R
Multivariate	Multivariate Statistics
NaturalLanguageProcessing	Natural Language Processing
NumericalMathematics	Numerical Mathematics
OfficialStatistics	Official Statistics & Survey Methodology
<u>Optimization</u>	Optimization and Mathematical Programming
Pharmacokinetics	Analysis of Pharmacokinetic Data

https://www.bioconductor.org/

https://www.bioconductor.org/packages/release/BiocViews.html#___So

R Markdown



- Make slides for presentations (HTML5, LaTeX Beamer, or PowerPoint).
- Build interactive applications based on Shiny.
- Write journal articles.
- Much more.













_ _ _

YAML (optional)

_ _ _

YAML: YAML Ain't Markup Language

_ _ _

YAML (optional)

The R in R Markdown

Chunks of code surrounded by

_ _ _

YAML (optional)

Chunks of code surrounded by

Text with simple text formating.

The R in R Markdown

The Markdown in R Markdown

```
The slope of the regression is `r round(b[1], digits = 1)` in the output.
```

```
---
author: "Emil Hvitfeldt"
date: "1/20/2019"
output: html_document
```

We built a **linear** regression model.

```
```{r}
fit <- lm(mpg ~ disp, data = mtcars)
b <- coef(fit)
plot(mpg ~ disp, mtcars)</pre>
```

The slope of the regression is `r round(b[1], digits = 1)` in the output.

### What we get

#### **Amazing Document**

Emil Hvitfeldt 1/20/2019

We built a **linear** regression model.

fit <- lm(mpg ~ disp, data = mtcars)
b <- coef(fit)
plot(mpg ~ disp, mtcars)</pre>



The slope of the regression is 29.6 in the output.

```
The slope of the regression is `r round(b[1], digits = 1)` in the output.
```

```

author: "Emil Hvitfeldt"
date: "1/20/2019"
output: pdf_document

We built a **linear** regression model.
```

```
```{r}
fit <- lm(mpg ~ disp, data = mtcars)
b <- coef(fit)
plot(mpg ~ disp, mtcars)
````</pre>
```

The slope of the regression is `r round(b[1], digits = 1)` in the output.

### What we get

Amazing Document Emil Hvitfeldt 1/20/2019

We built a **linear** regression model.

fit <- lm(mpg - disp, data = mtcars)
b <- coef(fit)
plot(mpg - disp, mtcars)</pre>



The slope of the regression is 29.6 in the output.

## **Build-in Formats**

- github\_document
- html\_document
- latex\_document
- md\_document
- odt\_document
- pdf\_document
- rtf\_document
- word\_document

- ioslides\_presentation
- powerpoint\_presentation
- slidy\_presentation
- beamer\_presentation

- xaringan

- xaringan

- flexdashboard

- xaringan
- flexdashboard
- learnr

- xaringan
- flexdashboard
- learnr
- rticles

### xaringan

[ʃæ.ˈriŋ.gæn]

build passing

An R package for creating slideshows with remark.js through R Markdown. The package name **xaringan** comes from Sharingan, a dōjutsu in Naruto with two abilities: the "Eye of Insight" and the "Eye of Hypnotism". A presentation ninja should have these basic abilities, and I think remark.js may help you acquire these abilities, even if you are not a member of the Uchiha clan.



Please see the full documentation as a presentation here (中文版在此). The remark.js website

provides a quick introduction to the underlying syntax upon which **xaringan** builds. If you prefer reading a book, **xaringan** is also documented in the R Markdown book (Chapter 7). You can use **remotes** to install the package:

https://github.com/yihui/xaringan

### flexdashboard: Easy interactive dashboards for R

- Use R Markdown to publish a group of related data visualizations as a dashboard.
- Support for a wide variety of components including htmlwidgets; base, lattice, and grid graphics; tabular data; gauges and value boxes; and text annotations.
- Flexible and easy to specify row and column-based layouts. Components are intelligently re-sized to fill the browser and adapted for display on mobile devices.
- Storyboard layouts for presenting sequences of visualizations and related commentary.



• Optionally use Shiny to drive visualizations dynamically.

Learn more about flexdashboard: http://rmarkdown.rstudio.com/flexdashboard/



### learnr: Interactive tutorials for R

The **learnr** package makes it easy to turn any R Markdown document into an interactive tutorial. Tutorials consist of content along with interactive components for checking and reinforcing understanding. Tutorials can include any or all of the following:

- 1. Narrative, figures, illustrations, and equations.
- 2. Videos (supported services include YouTube and Vimeo).
- 3. Code exercises (R code chunks that users can edit and execute directly).
- 4. Quiz questions.
- 5. Interactive Shiny components.

You can find documentation on using the learnr package here: https://rstudio.github.com/learnr/

### Welcome

Visualisation is an important tool for generating insights, but your data won't always arrive ready to visualize. Often you'll need to filter unwanted observations from your data or create new variables and summaries to visualize. In this tutorial, you will learn how to filter your data, including:

- How to use filter() to extract observations that pass a logical test
- How to write logical comparisons in R
- How to combine logical comparisons with Boolean operators
- How to handle missing values within comparisons

The readings in this tutorial follow *R* for Data Science (http://r4ds.had.co.nz/), section 5.2.

Continue

The rticles package provides a suite of custom R Markdown LaTeX formats and templates for various formats, including:

- ACM articles
- ACS articles
- AEA journal submissions
- AMS articles
- Biometrics articles
- Bulletin de l'AMQ journal submissions
- CTeX documents
- Elsevier journal submissions
- IEEE Transaction journal submissions
- JSS articles
- MDPI journal submissions
- Monthly Notices of the Royal Astronomical Society articles
- NNRAS journal submissions
- PeerJ articles
- Royal Society Open Science journal submissions
- Sage journal submissions
- Springer journal submissions

## Shiny

Combines the computational power of R with the interactivity of the modern web.

Powerful web framework.



https://gallery.shinyapps.io/050-kmeans-example/

### Iris k-means clustering

Sepal.Length	•
Y Variable	
Sepal.Width	•
Cluster count	
3	-

Kmeans example	server.R	ui.R	<b>1</b> show with app	]	
(http://www.rstudio.com/)	function(:	input, output, session) {	3	7/	52
	# Combin	ne the selected variables into a new data fr	ame J	/ /	JZ

### https://gallery.shinyapps.io/082-word-cloud/

### Word Cloud

AM	id Sum	mer	Nigh	t's Dr	ean
Cha	nge				
	_				
Minim	um Fre	quer	icy:		
Minim	15	quen			5
Minim 1	15 11	21	31	41	5
Minim	15 11 11	21 21	31 r of <sup>1</sup>	41 Words	5 S:
Minim 1 1 Maxin 1	15 11 11 100	21 Imbe	31 r of <sup>1</sup>	41 Words	5) S: 300

Word cloud	global.R	server.R	ui.R	<b>1</b> show with app	
<pre>cfereshteh@rstudio.com&gt; (http://www.rstudio.com/)</pre>	# Text of # A Mid Su	the books ımmer Night	downloaded from: 's Dream:	38 /	52
A simple word cloud generator, based on	# http://	/www.gutenb	erg.org/cache/epub/2242/pg2242.tx	st	

### https://gallery.shinyapps.io/086-bus-dashboard/

#### Show

- Northbound
- Southbound
- Eastbound
- Westbound

Note: a route number can have several different trips, each with a different path. Only the most commonlyused path will be displayed on the map.

Zoom to fit buses



Source data updates every 30 seconds.



### https://apps.garrickadenbuie.com/rstudioconf-2019/

<b>40</b> Tweets Today		<b>39</b> Tweeters Today	8	<b>2.16</b> Tweets/hr Today	X
56,00 rstudio::conf Like	6 <b>(</b> )	<b>4,120</b> rstudio::conf Tweets		<b>10,597</b> #rstats Tweets	R
Tweet Volume	Tweets by Hour of D	ay			
		Most Liked in	12 Hours		40 / 52

http://www.dataseries.org/

Choose variable, e.g., GDP

## **TIDIED UP**

Switzerland's data series in one place

# Where do I go for more information?

- Cheatsheets
- Books
- rweekly.org
- Forums
- USCbiostats/software-dev

## Have a question? Google it



https://www.google.com/



About 9,150,000 results (0.46 seconds)

#### Parsnip - Wikipedia

#### https://en.wikipedia.org/wiki/Parsnip 🔻



The parsnip (Pastinaca sativa) is a root vegetable closely related to the carrot and parsley. It is a biennial plant usually grown as an annual. Its long, tuberous ... Description · Taxonomy · Nutrients · Cultivation

#### Roasted Parsnips - How to Cook Parsnips - Parsnip Recipes https://www.fifteenspatulas.com > Recipes > Vegetable Sides 💌



Apr 14, 2018 - Roasted Parsnips are simple to make and incredibly delicious as a vegetable side to any dinner. They have a natural sweetness that ...

#### What is Parsnip Good For? - Mercola Food Facts - Dr. Mercola https://foodfacts.mercola.com > Vegetables 💌



Jan 9, 2017 - But what about the other vegetables that are not often given the same attention, though they are equally healthy? The parsnip is one example ...



Parsnips Vegetable

The parsnip is a root vegetable closely related to the carrot and parsley. It is a biennial plant usually grown as an annual. Its long, tuberous root has cream-colored skin and flesh, and left in the ground to mature, it becomes sweeter in flavor after winter frosts. Wikipedia

Nutrition Facts Parsnips 🔻

44 / 52

<

### https://rseek.org/



Created and maintained by <u>Sasha Goodman</u>. Serving the R community since 2007. Version 2.0.

Download and Install R

### https://rseek.org/



parsnip									<b>X</b>	
All	Github	Package	Documentation	R-project	Blog	Source	RStudio	Twitter	Popular Package	е
About 389,	000 results	(0.32 second	ds)					Sor	t by: Relevance	Ŧ

#### CRAN - Package parsnip

#### https://cran.r-project.org/package=parsnip

Nov 12, 2018 ... parsnip: A Common API to Modeling and Analysis Functions. A common ... Author: Max Kuhn [aut, cre], Davis Vaughan [aut], RStudio [cph]. Labeled Package R-project

#### parsnip - Tidyverse

#### https://www.tidyverse.org/articles/2018/11/parsnip-0-0-1/



The parsnip package is now on CRAN. It is designed to solve a specific problem related to model fitting in R, the interface. Many functions have different ...

#### Examples of non-linear optimization with dials and parsnip help ...

https://community.rstudio.com/t/examples-of...parsnip.../19603



Dec 9, 2018 ... RStudio Community · Examples of non-linear optimization with dials and parsnip help · Machine Learning and Modeling · tidymodels ... Labeled Issues RStudio

46 / 52

#### R Weekly 2018-49 parsnip, Validation | RWeekly.org - Blogs to ...

#### https://rweekly.org/2018-49.html



Dec 3, 2018 ... R Weekly 2018-49 parsnip, Validation ... R in the Real World ... How to work with strings in base R - An overview of 20+ methods for daily ... Labeled <u>Blog</u>

## Cheatsheets



https://www.rstudio.com/resources/cheatsheets/

## Books

### https://bookdown.org/

bookdown

Home About Archive Tags Authors Log in

**9 A O** 

### BOOKDOWN

#### Write HTML, PDF, ePub, and Kindle books with R Markdown

The **bookdown** package is an <u>open-source R package</u> that facilitates writing books and long-form articles/reports with R Markdown. Features include:

- Generate printer-ready books and ebooks from R Markdown documents.
- A markup language easier to learn than LaTeX, and to write elements such as section headers, lists, quotes, figures, tables, and citations.
- Multiple choices of output formats: PDF, LaTeX, HTML, EPUB, and Word.
- Possibility of including dynamic graphics and interactive applications (HTML widgets and Shiny apps).
- Support a wide range of languages: R, C/C++, Python, Fortran, Julia, Shell scripts, and SQL, etc.
- LaTeX equations, theorems, and proofs work for all output formats.
- Can be published to GitHub, bookdown.org, and any web servers.
- Integrated with the RStudio IDE.
- One-click publishing to https://bookdown.org.

Below is a list of featured books. For a full list, please see the <u>archive</u> page. For the full documentation of the **bookdown** package, please see the free <u>online</u> <u>book</u> *bookdown: Authoring Books and Technical Documents with R Markdown*.



## **Books**

#### **R** Programming for Data Science

by Roger D. Peng

#### 2016-12-22



The R programming language has become the de facto programming language for data science. Its flexibility, power, sophistication, and expressiveness have made it an invaluable tool for data scientists around the world. This book is about the fundamentals of R programming. You will get started with the basics of the language, learn how to manipulate datasets, how to write functions, and how to debug and optimize code. With the

C Star

63

9

fundamentals provided in this book, you will have a solid foundation on which to build your data science toolbox. *Read more*  $\rightarrow$ 

#### Efficient R programming

by Colin Gillespie, Robin Lovelace

#### 2016-11-30



Colin Gillespie & Robin Lovelac

Efficient R Programming is about increasing the amount of work you can do with R in a given amount of time. It's about both computational and programmer efficiency. [...] This is the online version of the O'Reilly book: Efficient R programming. Pull requests and general comments are welcome. Colin Gillespie is Senior lecturer (Associate professor) at Newcastle University, UK. His research interests are high performance statistical

🖓 Star

452

computing and Bayesian statistics. He is regularly employed as a consultant by Jumping Rivers and has been teaching R since 2005 at a variety of levels, ranging ... *Read more*  $\rightarrow$ 

10

### https://www.rweekly.org/

#### R Weekly 2019-03 RStudio Conf

#### RWeekly.org Live Mail Feed Conf About All Draft Submit Night

#### Live

- hrbrthemes 0.6.0 on CRAN + Other In-Development Package News (rud.is)
- **ventstudies Event Study Analysis** (cran.r-project.org)
- **Q** CluMix Clustering and Visualization of Mixed-Type Data (cran.r-project.org)

More

#### R Weekly 2019-03 RStudio Conf 2019

#### 21 Jan 2019 🔟 🛉 🎔

<b>T</b> type to filter	
Release Date: 2019-01-21	

#### Highlight

• RStudio Conf 2019 Slides - Karl Broman • Contains workshop materials and presentation slides for RStudio Conf

(github.com)

- The Unreasonable Effectiveness of Public Work (dropbox.com)
- Our Colour of Magic: The open sourcery of fantastic R packages ( docs.google.com )

#### Insights

• You did a sentiment analysis with tidytext but you forgot to do dependency parsing to answer WHY is something positive/negative ( bnosac.be )

50 / 52

2019 Highlight Insights R in the Real World R in Organizations R in Academia Resources New Packages Updated Packages Tutorials R Project Updates Upcoming Events in 3 Months Jobs Call for Participation

Quotes of the Week



https://community.rstudio.com/



https://stackoverflow.com/

### https://github.com/USCbiostats/software-dev

### **Software Development Standards**

This project's main contents is located in the project's Wiki:

### **Coding Standards**

- 1. Coding Standards
- 2. Software Thinking
- 3. Development Workflow
- 4. Misc

We do have some direct guidelines developed as issue templates here.

### **Bioghost Server**

- 1. Introduction
- 2. Setup
- 3. Cheat Sheet

#### HPC in R

- Parallel computing in R
- parallel
- iterators+foreach
- RcppArmadillo + OpenMP