

Look ahead

Statistics with R
Basel R Bootcamp



April 2019

Hello Statistician!

In 2 days, 8 sessions, and 16 hours you have come a long way.

	Sat, 6 Apr	Sun, 7 Apr
0900	Welcome	Recap
0930	Intro to Statistics +Practical	Mixed-models +Practical
	R for Statistics +Interactive	
1200	<i>Lunch</i>	<i>Lunch</i>
1300	Linear models I +Practical	Robust Statistics +Practical
1500		
1530	Linear models II +Practical	New Statistics +Practical
1730		Look ahead
1800	Wrapup	<i>Apero</i>



from nomaly.io



from artsandculture.google.com

"All models are wrong, but some are useful."

George E. P. Box

1919-2013

"An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem."

John W. Tukey

1915-2000

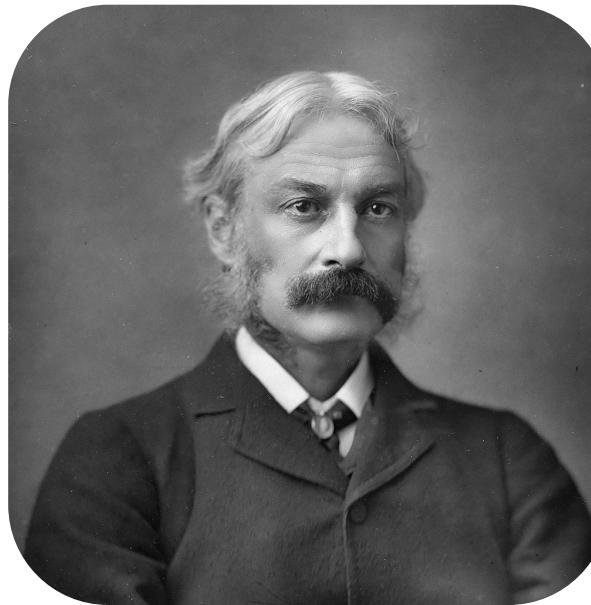


from hamhigh.co.uk

"To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of."

Ronald A. Fisher

1890-1962



from wikiquote.org

"[He] uses statistics like a drunken man uses a lamp post, more for support than illumination."

Andrew Lang

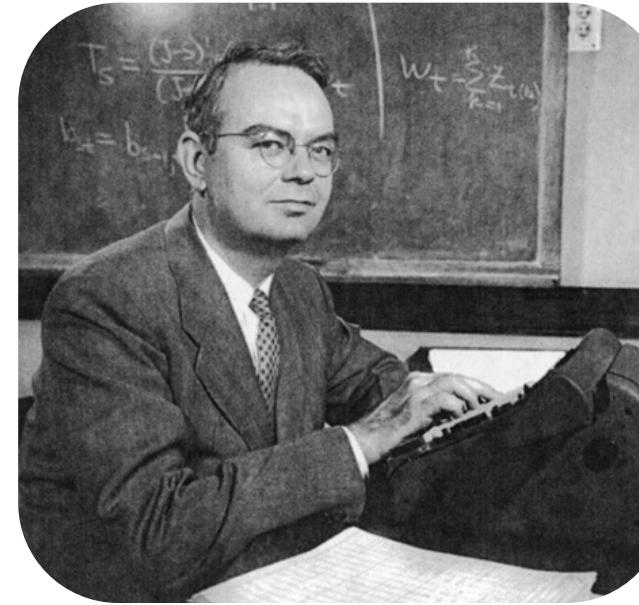
1844-1912



from [independent.co.uk](https://www.independent.co.uk)

"Statistical thinking will one day be as necessary a qualification for efficient citizenship as the ability to read and write."

H. G. Wells
1866-1946



from americanscientist.org

"It's easy to lie with statistics; it is easier to lie without them."

C. Frederick Mosteller
1916-2006

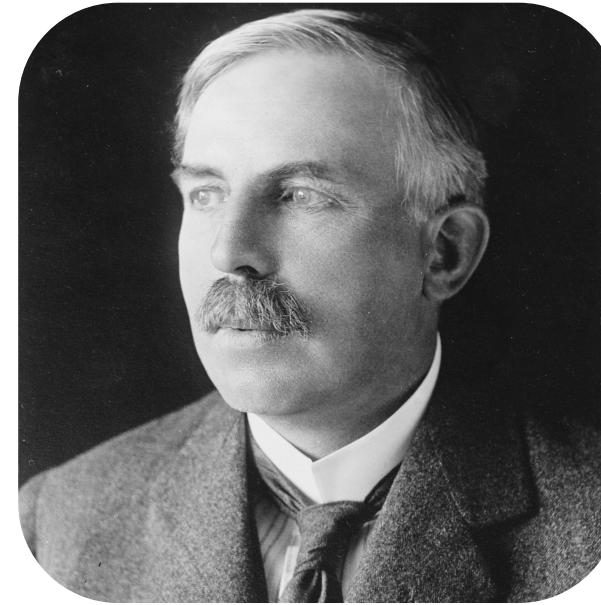


from bigear.org

"The subjectivist (i.e. Bayesian) states his judgements, whereas the objectivist sweeps them under the carpet by calling assumptions knowledge, and he basks in the glorious objectivity of science."

I. J. Good

1916-2009



from [wikipedia.org](https://en.wikipedia.org)

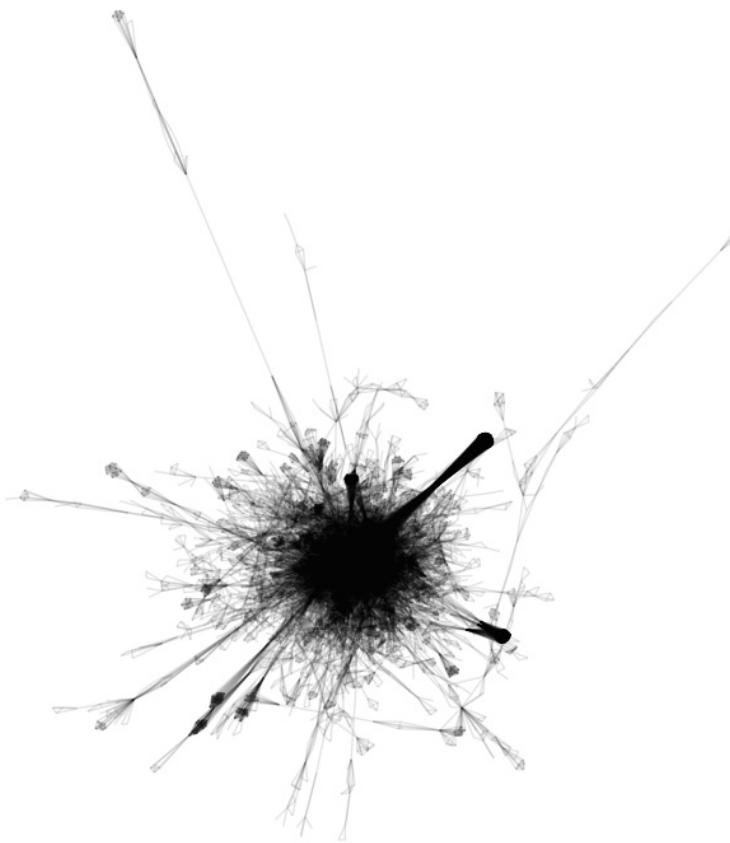
"If your experiment needs statistics, you ought to have done a better experiment."

Ernest Rutherford

1871-1937

Stuff we didn't cover

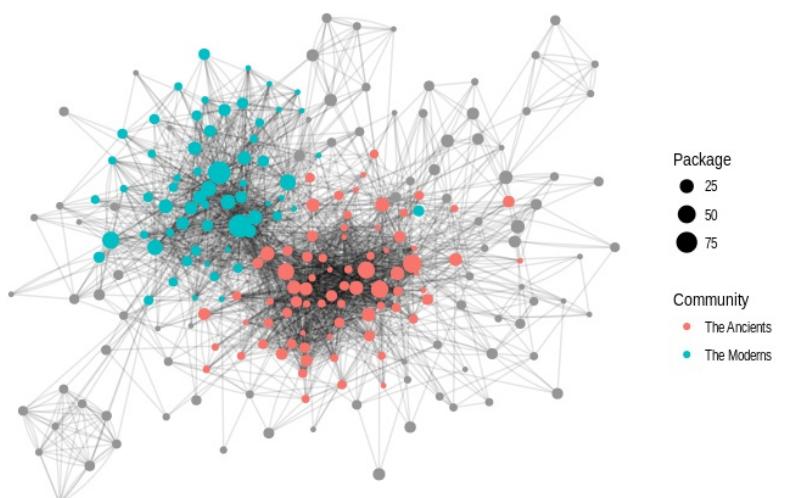
1. Networks
2. Tidyverse
3. Machine learning
4. Text analysis
5. Rcpp
6. Forms



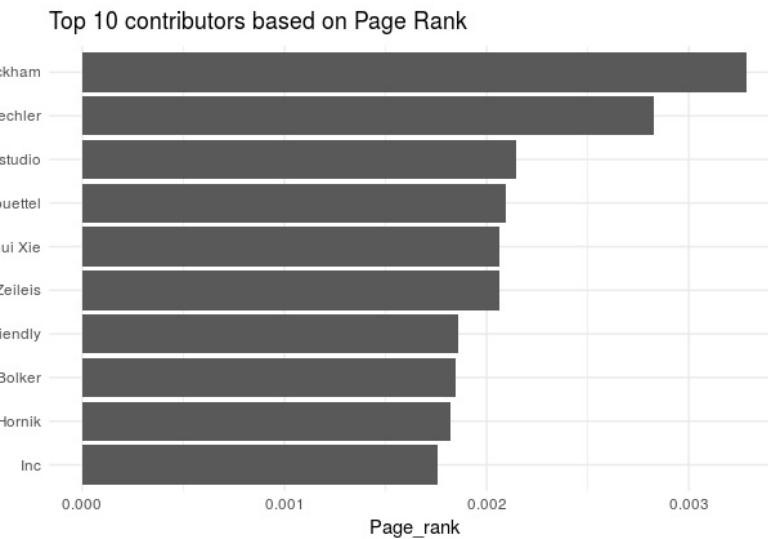
source R-bloggers.com

Networks

A **social graph** of package Co-authors using **tidyverse** plus **ggraph**, an extension for **ggplot2** for graphs (aka networks) and **igraph**, an extremely powerful library for network analysis. Find the code and additional explanations [here](#).



source [R-bloggers.com](#)



source [R-bloggers.com](#)

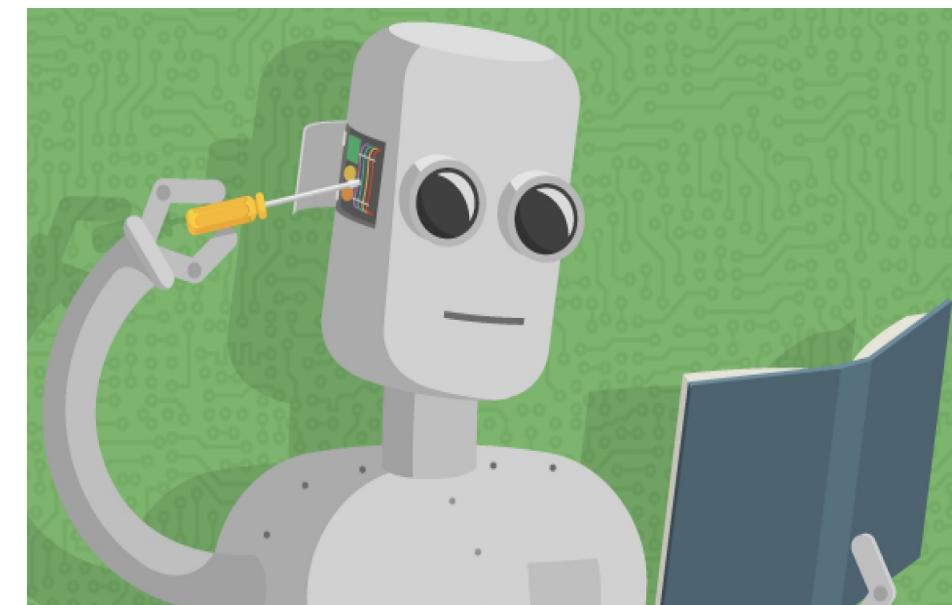
Tidyverse

Package	Description
<code>ggplot2</code>	Amazing package for plotting.
<code>dplyr</code>	Amazing package for data wrangling
<code>tidyverse</code>	Helper functions to get your data tidy
<code>readr</code>	Powerful package for reading in data.
<code>purrr</code>	Package for functional programming in R.
<code>tibble</code>	Package providing the <code>tibble</code> class
<code>stringr</code>	Amazing package for string manipulation
<code>forcats</code>	Package to solve common problems with factors

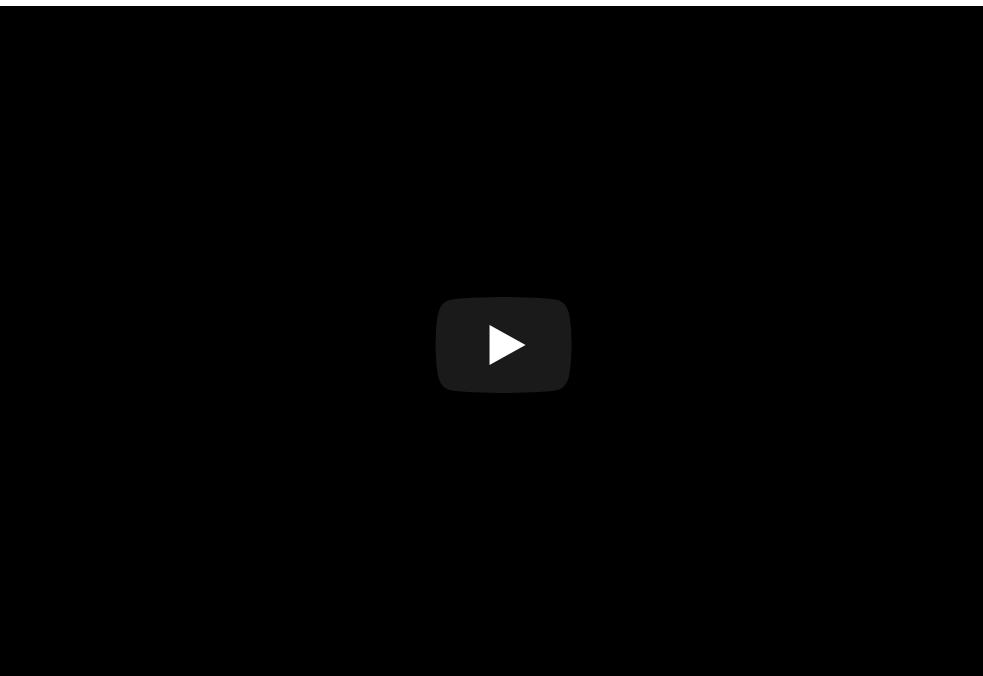


Machine learning

Package	Description
<code>caret</code>	Umbrella package for diverse machine learning algorithms.
<code>mlr</code> , <code>e1071</code> , etc.	Other umbrella packages.
<code>randomForest</code> , <code>rpart</code> , <code>FFTrees</code>	Decision trees.
<code>cluster</code> , <code>fastcluster</code> , <code>cstab</code> , etc.	Cluster analysis.
<code>forecast</code> , <code>mgm</code> , <code>timeSeries</code> , etc.	Time series models.
<code>tensorflow</code>	Time series models.



Text analysis



Sundar Pichai @ Google IO, May 2018

Package	Description
<code>tm, tidytext</code>	General text analysis packages
<code>stringr, stringi</code>	String operations and regular expressions.
<code>rvest, XML</code>	Scraping content of the internet
<code>text2vec</code>	Vector representation of words.
<code>SentimentAnalysis</code>	Sentiment analysis.
<code>twitteR, streamR, jsonlite</code>	Streaming and parsing tweets.
<code>Rfacebook</code>	Access to Facebook API.

Rcpp

By now one of the most referenced R packages is Rcpp - **R's interface to C++**. With often relatively little effort due to **Rcpp sugar**, Rcpp can provide vast speed improvements, which many packages today rely on Rcpp in the background for **swift code execution**. Rcpp becomes particularly powerful, when supplemented with BH, which makes available a collection **free, peer-reviewed C++ libraries**, and **RcppArmadillo**, which available the high-performance **Armadillo** library for linear algebra methods.



source classic105.com

```
//include <Rcpp.h>
using namespace Rcpp;
//The cppFunction will automatically add this.

//Or, prefix Rcpp objects with the Rcpp namespace e.g.:
Rcpp::NumericVector xx(10);

//Matrix of 4 rows & 5 columns (filled with 0)
NumericMatrix xx(4, 5);

//Fill with value
int xsize = xx.nrow() * xx.ncol();
for (int i = 0; i < xsize; i++) {
    xx[i] = 7;
}
//Same as above, using STL fill
std::fill(xx.begin(), xx.end(), 8);

//Assign this value to single element
//(1st row, 2nd col)
xx(0,1) = 4;

//Reference the second column
//Changes propagate to xx (same applies for Row)
NumericMatrix::Column zzcol = xx( _, 1);
zzcol = zzcol * 2;

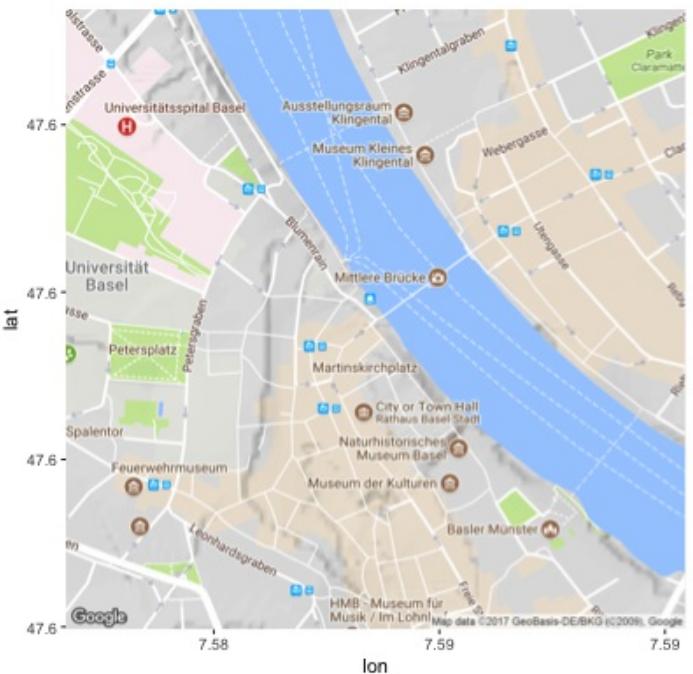
//Copy the second column into new object
NumericVector zz1 = xx( _, 1);
//Copy the submatrix (top left 3x3) into new object
NumericMatrix zz2 = xx( Range(0,2),
Range(0,2));
```

Quick Reference Guide

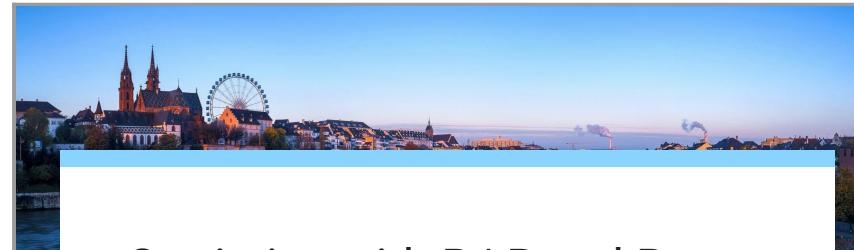
Google Forms & Maps

New packages also allow you to interact with [Google Maps](#) and [Google Forms](#). Use `ggmap` to access Google Maps and `googlesheets` to access Google Forms.

```
library(ggmap)  
ggmap(get_map(c(7.588576, 47.559601), zoom=16))
```



Bootcamp Follow-up Survey Link



Statistics with R | Basel R Bootcamp Follow-up questionnaire

Please be so kind to take a few minutes and provide us with feedback for your recent bootcamp. In the first part you will have the chance to tell us what you think about the workshop in general. In the second part you can indicate which sessions you liked/disliked and why.

* Erforderlich

How did you hear about the bootcamp? *

- Google / Google Ad
- LinkedIn

13 / 19

How to continue



Books

Here is a very incomplete series of good books. They are ordered by complexity, beginning with user-friendly books on **learning statistics** in R and ending with books focusing on the more **advanced topics of the R language**.



Websites

The web is a great place to learn about R.

[Google](#) or [Rseek](#), which is a wrapper around google to maximize hits related to R. However, most of the time Google works just fine. Just be sure to add `r` to the search query.

[R-bloggers](#) is a website on which R users inform each other on the newest developments. See, e.g., Nathaniel's [entry](#).

[Stackoverflow](#) is a website on which R users exchange problems and solutions to problems. Try post something yourself. You will be amazed by the turnaround.



Support & Consulting



Dr. Dirk Wulff

dirkwulff.org
github.com/dwulff
cstab,
[mousetraps](https://mousetraps.net), [memnet](https://memnet.net) [choicepp](https://choicepp.net)



Dr. Nathaniel Phillips

nathanielpPhillips.com
github.com/ndphillips
yarr,
FFTrees



Markus Steiner

github.com/mdsteiner
ShinyPsych,
EFAdiff

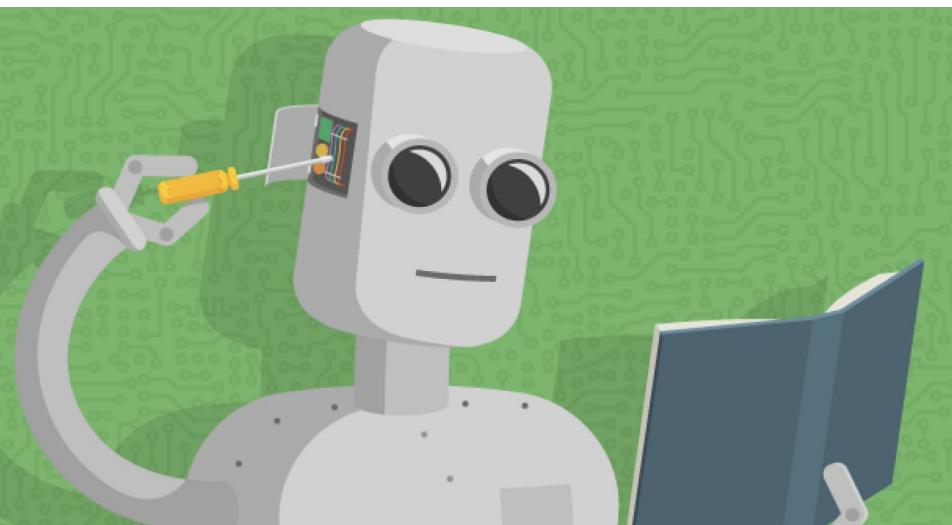


Dr. Michael Schulte-M.

schulte-mecklenbeck.com
github.com/schultem

Next Bootcamps

Applied Machine Learning with R



Basel, May 11-12, 2019

R for Data Science



Basel, June 11-12, 2019

One more thing...

19 / 19



www.therbootcamp.com

Statistics with R | April 2019