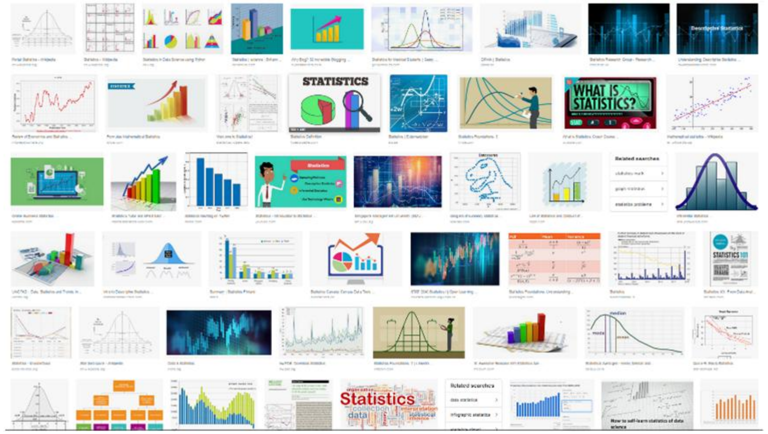
Engaging Everyone with Open Data Science

Kimmo Vehkalahti
Centre for Social Data Science
University of Helsinki, Finland

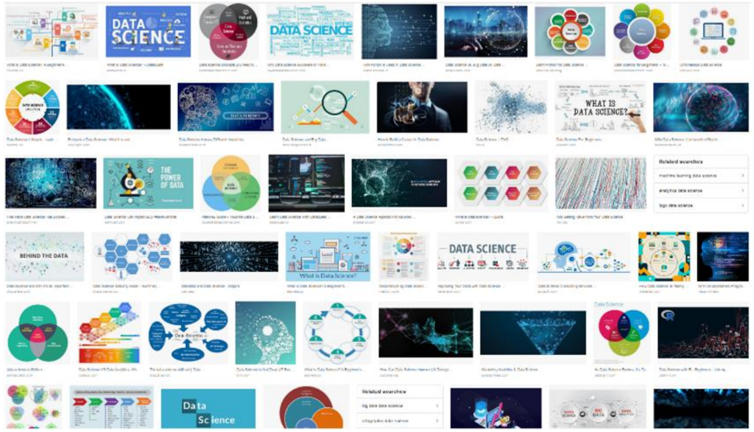
IASE 2019 Satellite Conference *"Decision Making Based on Data"* 13-16 Aug 2019 in Kuala Lumpur, Malaysia

Outline

1. Introduction
2. Open Data Science
3. Results
4. Conclusion
5. References



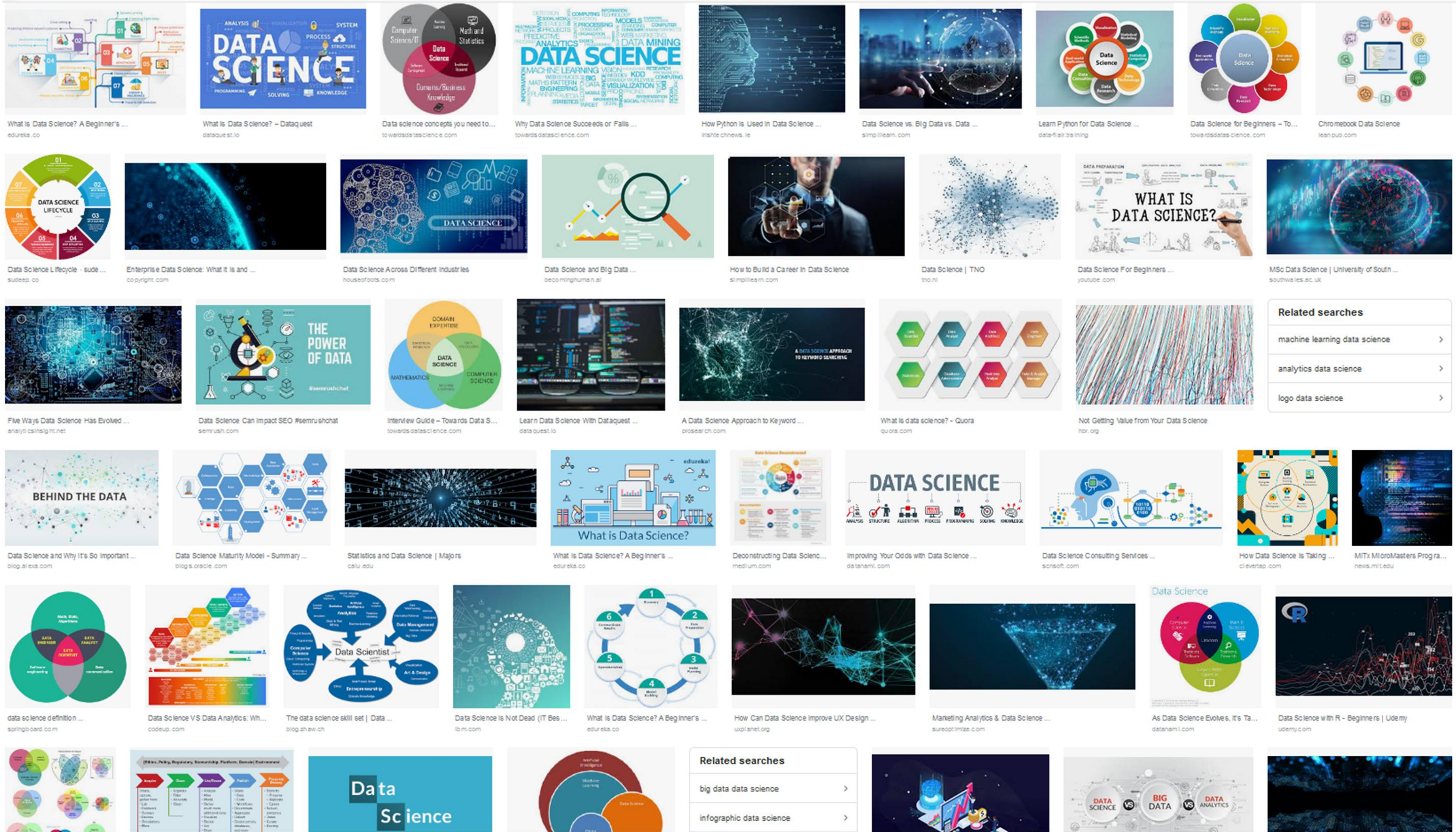
Statistics (Google search)



Data Science (Google search)

Open Data Science (Google search)


Portal: Statistics - Wikipedia
en.wikipedia.org
Statistics - Wikipedia
en.wikipedia.org
Statistics in Data Science using Python
edx.org
Statistics | science | Britann...
britannica.com
Why Blog? 52 Incredible Blogging ...
expresswriters.com
Statistics for Medical Students | Geeky ...
geekymedics.com
DPMA | Statistics
dpma.de
Statistics Research Group - Research ...
cardiff.ac.uk
Descriptive Statistics
Understanding Descriptive Statistics ...
towardsdatascience.com
Review of Economics and Statistics ...
mitpressjournals.org
STATISTICS
Formulas Mathematical Statistics
byjus.com
Welcome to Statistics!
statistics.rutgers.edu
STATISTICS
680 × 440
Statistics Definition
investopedia.com
+2w
Statistics | Externwebben
slu.se
Statistics Foundations: 3
lynda.com
WHAT IS STATISTICS?
STAT
What Is Statistics: Crash Course ...
youtube.com
Mathematical statistics - Wikipedia
en.wikipedia.org
Online Business Statistics ...
wpforms.com
Statistics Tutor and SPSS tutor ...
maths-statistics-tutor.com
statistics hashtag on Twitter
twitter.com
Statistics
Sampling Methods
Descriptive Statistics
Inferential Statistics
Use Technology Wisely
Statistics - Introduction to Statistics ...
youtube.com
Singapore Management University (SMU ...
smu.edu.sg
Datasaurus
dangers of summary statistics ...
verdazo.com
Law of Statistics and Distrust of ...
toppr.com
Related searches
statistics math
graph statistics
statistics problems
Inferential Statistics ...
analyticsvidhya.com
UNCTAD | Data, Statistics and Trends in ...
unctad.org
Intro to Descriptive Statistics ...
towardsdatascience.com
Summary | Statistics Finland
stat.fi
Statistics Canada: Census Data Tools ...
huronshores.ca
STAT*2040 Statistics I | Open Learning ...
courses.opened.uoguelph.ca
Pdf
Mean
Variance
Statistics Foundations: Understanding ...
pluralsight.com
Statistics
suomenpankki.fi
STATISTICS 101
Statistics 101: From Data Anal...
amazon.com
Statistics - MoodleDocs
docs.moodle.org
Standard score - Wikipedia
en.wikipedia.org
Data & Statistics
irena.org
wwPDB: Download Statistics
wwpdb.org
Statistics Foundations: 2 | LinkedIn ...
linkedin.com
10 Awesome Reasons Why Statistics Are ...
medium.com
median
mode
mean
Statistical Averages - Mean, Median and ...
data36.com
Simple Regression
Quick-R: Basic Statistics
statmethods.net
BIOLOGY LETTERS
Statistics
collection
data
interpretation
statistical inference
Related searches
data statistics
infographic statistics
statistics clipart
How to self-learn statistics of data science



What Is Data Science? A Beginner's ...
What Is Data Science? – Dataquest
Data science concepts you need to...
Why Data Science Succeeds or Fails ...
How Python Is Used In Data Science ...
Data Science vs. Big Data vs. Data ...
Learn Python for Data Science ...
Data Science for Beginners – To...
Chromebook Data Science
Data Science Lifecycle - sude ...
Enterprise Data Science: What It Is and ...
Data Science Across Different Industries
Data Science and Big Data ...
How to Build a Career in Data Science
Data Science | TNO
Data Science For Beginners ...
MSc Data Science | University of South ...
Five Ways Data Science Has Evolved ...
Data Science Can Impact SEO #semrushchat
Interview Guide – Towards Data S...
Learn Data Science With Dataquest ...
A Data Science Approach to Keyword ...
What is data science? - Quora
Not Getting Value from Your Data Science
Related searches
machine learning data science
analytics data science
logo data science
Data Science and Why It's So Important ...
Data Science Maturity Model - Summary ...
Statistics and Data Science | Majors
What Is Data Science? A Beginner's ...
Deconstructing Data Scienc...
Improving Your Odds with Data Science ...
Data Science Consulting Services ...
How Data Science Is Taking ...
MITx MicroMasters Progra...
data science definition ...
Data Science VS Data Analytics: Wh...
The data science skill set | Data ...
Data Science Is Not Dead (IT Bes...
What Is Data Science? A Beginner's ...
How Can Data Science Improve UX Design ...
Marketing Analytics & Data Science ...
As Data Science Evolves, It's Ta...
Data Science with R - Beginners | Udemy
Related searches
big data data science
infographic data science


Open Data Science - YouTube
youtube.com
Open Data Science
ods.ai
Open Data Science Jobs Board - Jobs ...
jobs.opendatascience.com
Dubai Data Science (Dubai, U.A.E.) | Meetup
meetup.com
Open Data Science
ods.ai
Open Data Science Machine Learning ...
datascienceinsights.in
Introduction to Open Data Science ...
courses.helsinki.fi
Data Science Hands-on with Open Source ...
cognitiveclass.ai
Open Data Science at SIO
open-data-science-at-sio.github.io
Open Data Science | Domino Data Lab
dominodatalab.com
Anthony Goldbloom, CEO of Kaggle ...
medium.com
Corporate blog Open Data Science / Habr
m.habr.com
Open data science for marine managem...
ohi-science.org
Open Data Science in the AI Era ...
slideshare.net
Karthik Ram – Electronic Resources ...
electroniclibrarian.org
Open Data Science: beyond traditional ...
slideshare.net
Related searches
use of data science
why data science is important
data science for good
Open Data Science · GitHub
github.com
Collaboration and workflow in open data ...
lombardidatahub.com
ODSC Grant Award | Open ...
odsc.com
Open Data Science Conference - Europe ...
startup-calendar.com
Open Data Science Confere...
linkedin.com
Open Data Science о Slack, xgboost и GPU
pvsm.ru
The Journey to Open Data Science
kdnuggets.com
Data Science on Hadoop in the Enterprise
slideshare.net
Open data science for marine management ...
oceanhealthindex.org
less time using open data science ...
nature.com
Open Data Science Conference | Cente...
cdait.gatech.edu
Open Data Science Conference - Deep ...
odsc.com
Last batch of notebooks for Think Stats ...
pinterest.com
Open data science for marine management
ohi-science.org
Open Data Science Conference West ...
community.cadence.com
St. Petersburg Open Data Science Meetup ...
meetup.com
Open Data Science Conference ...
thedatalist.com
Open Data Science Conference (ODSC ...
linkedin.com
San Francisco – Open Data Science ...
smrfoundation.org
Open Data Science and Machine Learning ...
vision.cloudera.com
Breaking Data Science Open
know.continuum.io
R: Getting Started with Data S...
pinterest.com
How To Get Started With Data Lakes ...
medium.com
SML 310 cracks open data science for ...
csml.princeton.edu
Domino Data Science Platform | D...
dominodatalab.com
Related searches
i love data science
https://www.google.com/search?q="open+data+science"&client=firefox-b-e&source=lnms&tbm=isch&sa=X&ved=0ahUKEwi-3e-G_objAhWk5KYKHQ5bBzQQ_AUIECgB&biw=2520&bih=1404&dpr=0.67#

1. Introduction

- The ongoing “Data revolution” sets more requirements for the students and researchers on all fields of science.
- One could say (without exaggerating too much) that

  We should all be data scientists.

- The term “data science” is a good synonym to statistics.
- “Statistics” vs “Data science” is also a question of brand/image.

1. Introduction

According to Wikipedia:

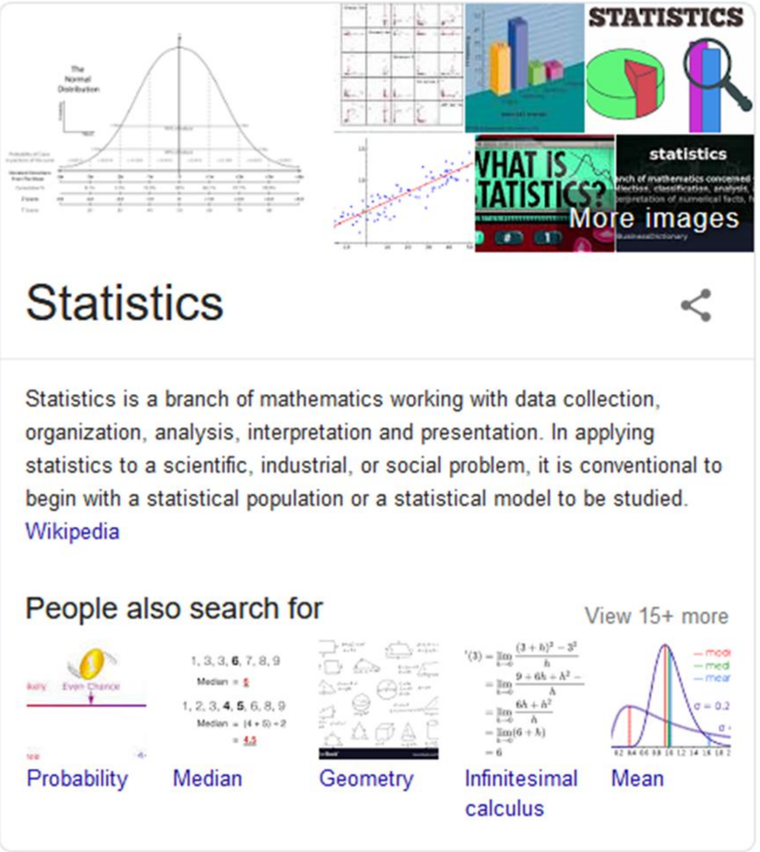
- *"Statistics is a branch of mathematics working with data collection, organization, analysis, interpretation and presentation."*

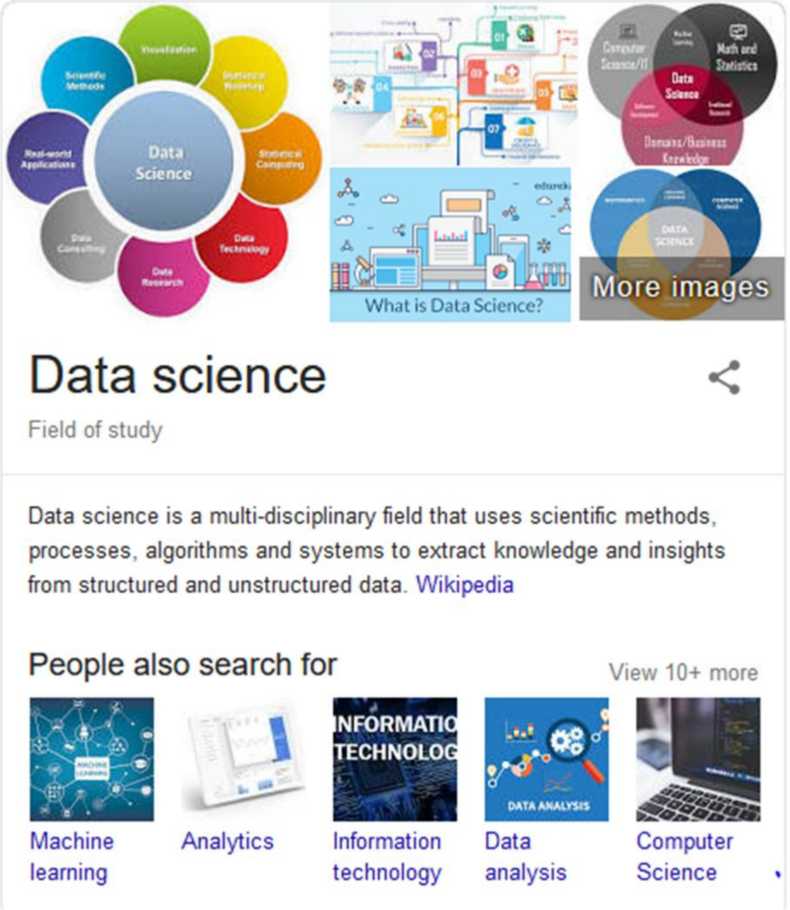
while

- *"Data science is a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data."*

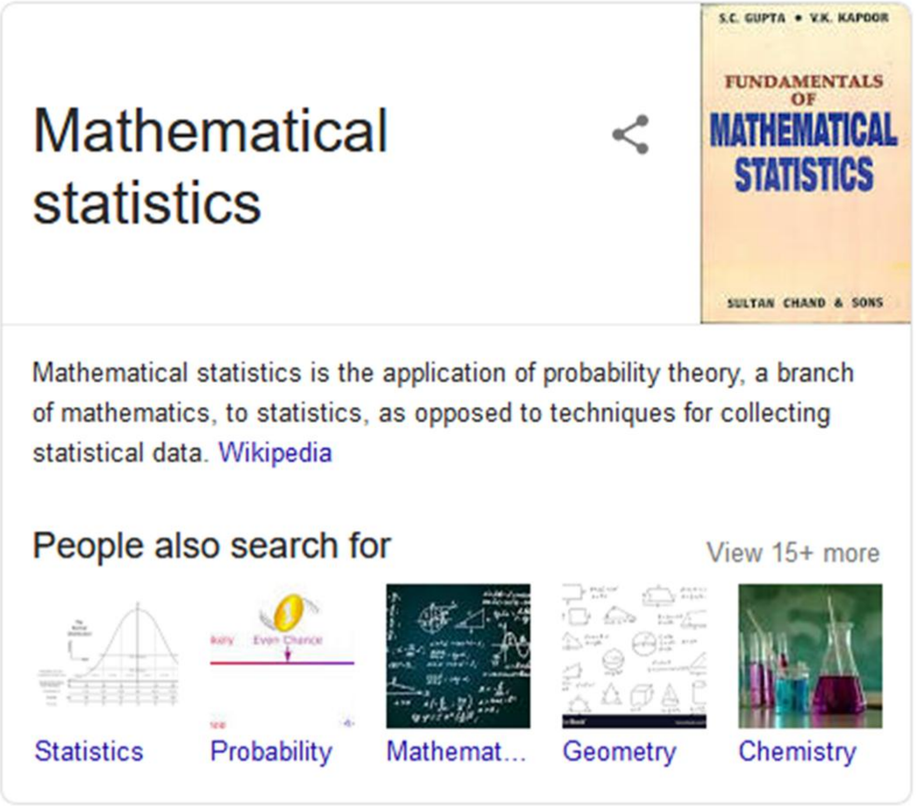
Which one of these definitions sounds more interesting? Perhaps some combination of these would better describe our field?

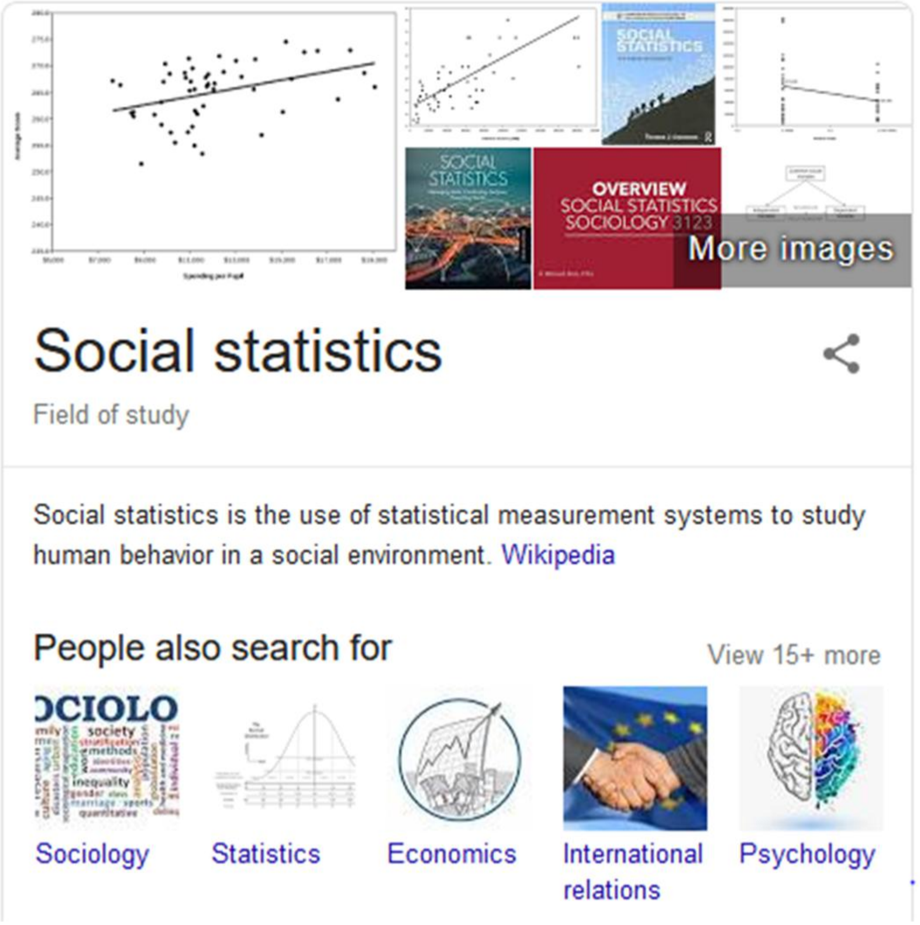
1. Introduction





1. Introduction





1. Introduction

- Teaching of statistics should focus more on practical data science.
- Special emphasis needed on data wrangling:
  - Preparing the data for the analyses
  - Looking at the data via simple but clever visualizations
- Important overall learning goals:
  - Principles and practices of open science and reproducible research
  - Statistical and algorithmic thinking, sharing of code and data
- State-of-the-art tools like RStudio and R Markdown freely available!
- Thus: many reasons why I like to use the term *Open Data Science.*

2. Open Data Science

- New course established to respond to the serious need around:
- 

- Primary target: Doctoral students of social sciences and humanities
- Suitable for ”anyone” (master’s / bachelor’s / exchange / post docs)
- So far, 100+ participants every time (organized 3 times 2017-2018)

2. Open Data Science

General learning objectives of the course were stated as follows:

- *”After completing this course you will understand the principles and advantages of using open research tools with open data and understand the possibilities of reproducible research.”*

and

- *“You will know how to use RStudio, R Markdown, and GitHub for these tasks, and know how to learn more of these open software tools. You will also know how to apply certain statistical methods of data science, that is, data-driven statistics.”*

2. Open Data Science

Seven weeks of study online, w/ one optional computer class / week:

1. Start me up!
2. Regression and model validation
3. Logistic regression
4. Clustering and classification
5. Dimensionality reduction techniques
6. Analysis of longitudinal data


Introduction to Open Data Science

CONTENTS

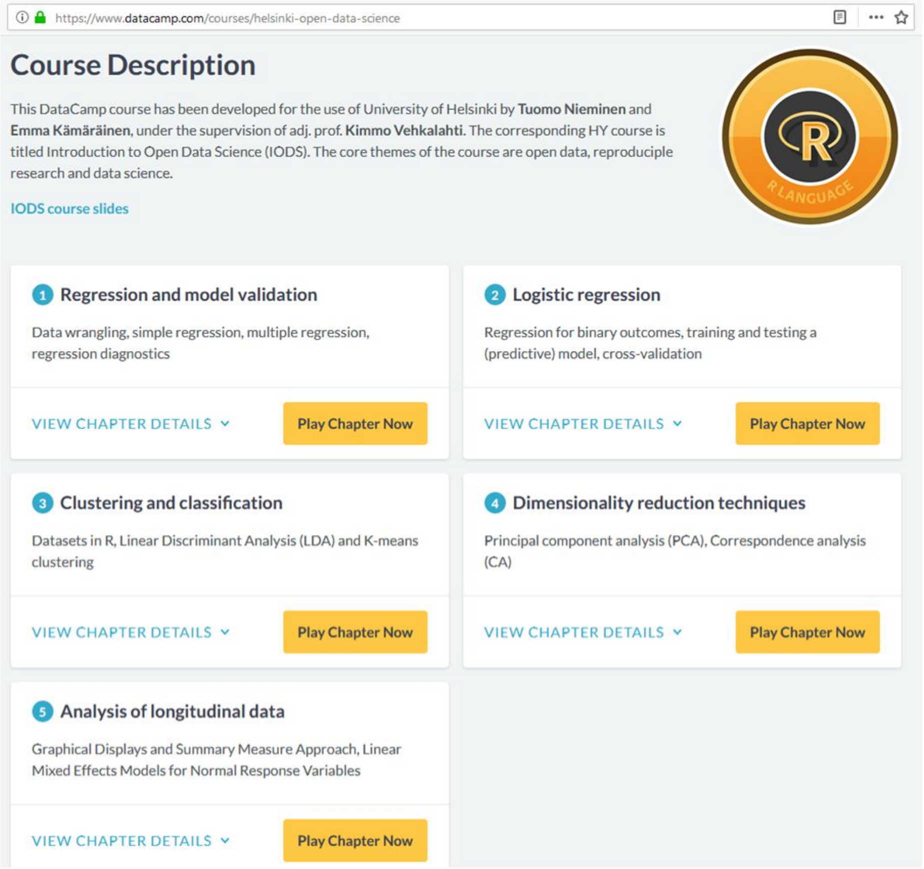
Welcome to the course!
1. Start me up!
2. Regression and model validation
3. Logistic regression
4. Clustering and classification
5. Dimensionality reduction techniques
6. Analysis of longitudinal data
7. Some books for your curiosity
8. Deadlines, forums, FAQ

Far from a traditional, systematic statistics course! A mixture of

- Statistical modeling (LM, GLM etc.)
- Data analysis (PCA, MCA etc.)

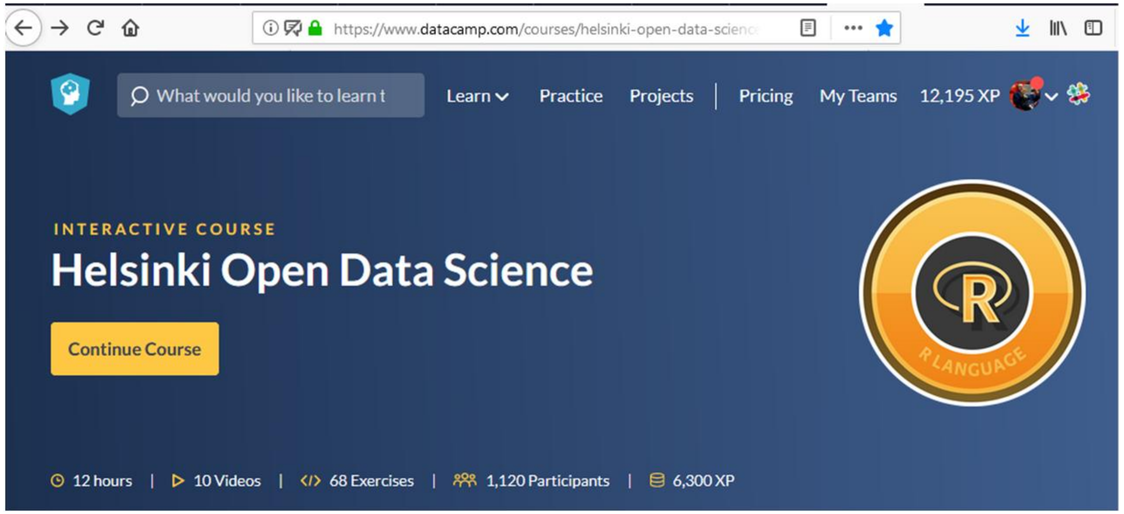
2. Open Data Science

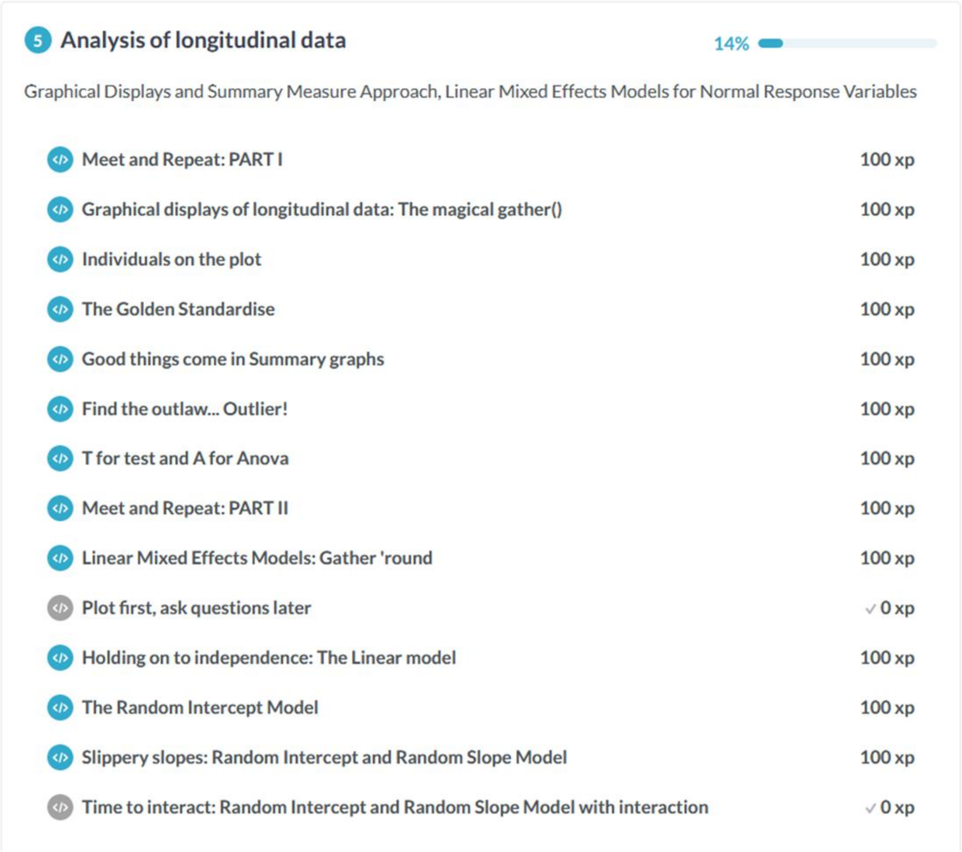
- Dedicated free DataCamp course ”Helsinki Open Data Science” supports learning the R skills
- Easy, interactive way to explore and learn the weekly R tricks to be used
- R code can then be copied to Rstudio
- www.datacamp.com



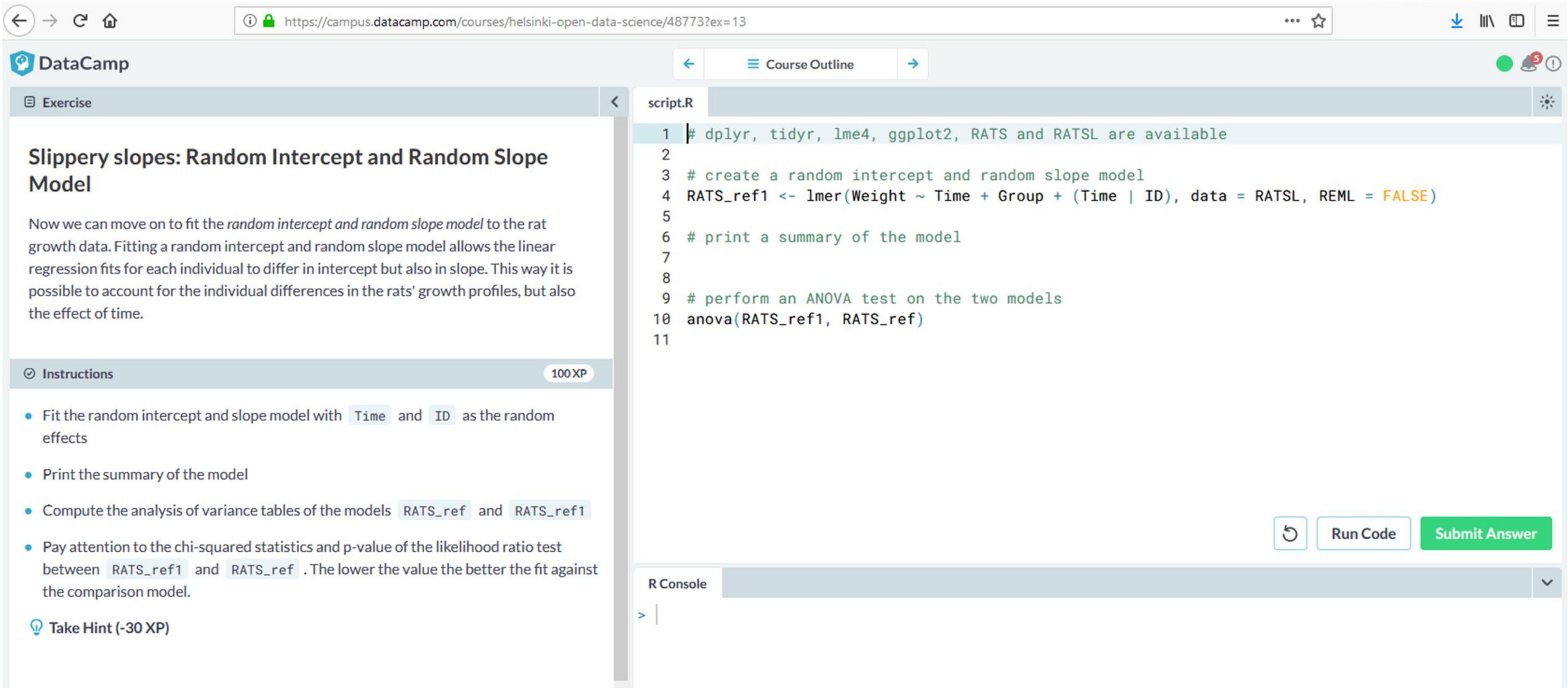
2. Open Data Science

Some views of the DataCamp platform:





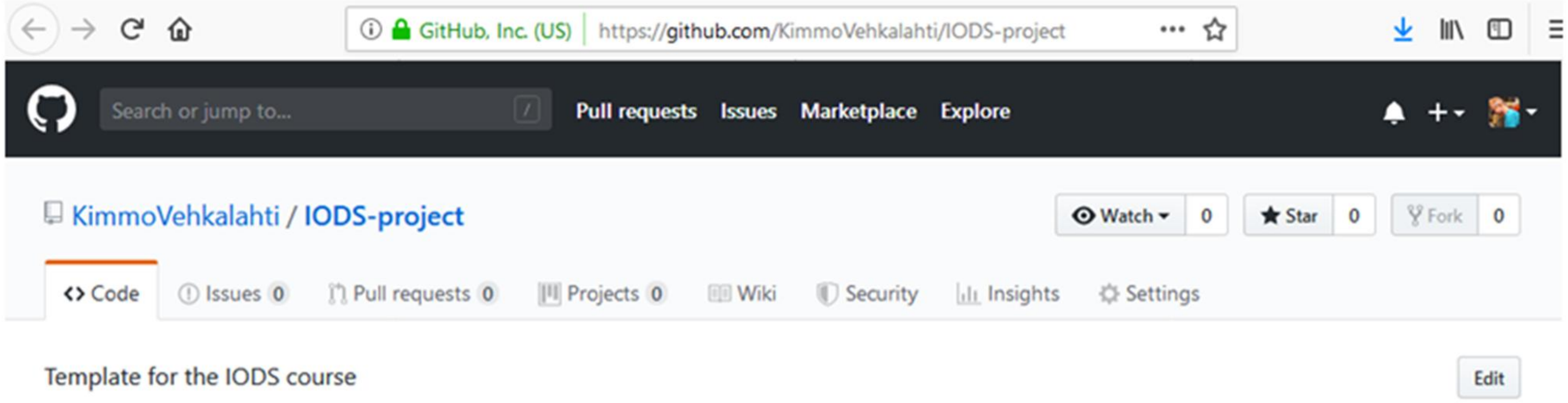
2. Open Data Science



Kimmo Vehkalahti, University of Helsinki, Finland: Engaging Everyone with Open Data Science
IASE 2019 Satellite Conference *"Decision Making Based on Data"* 13-16 Aug 2019 in Kuala Lumpur, Malaysia

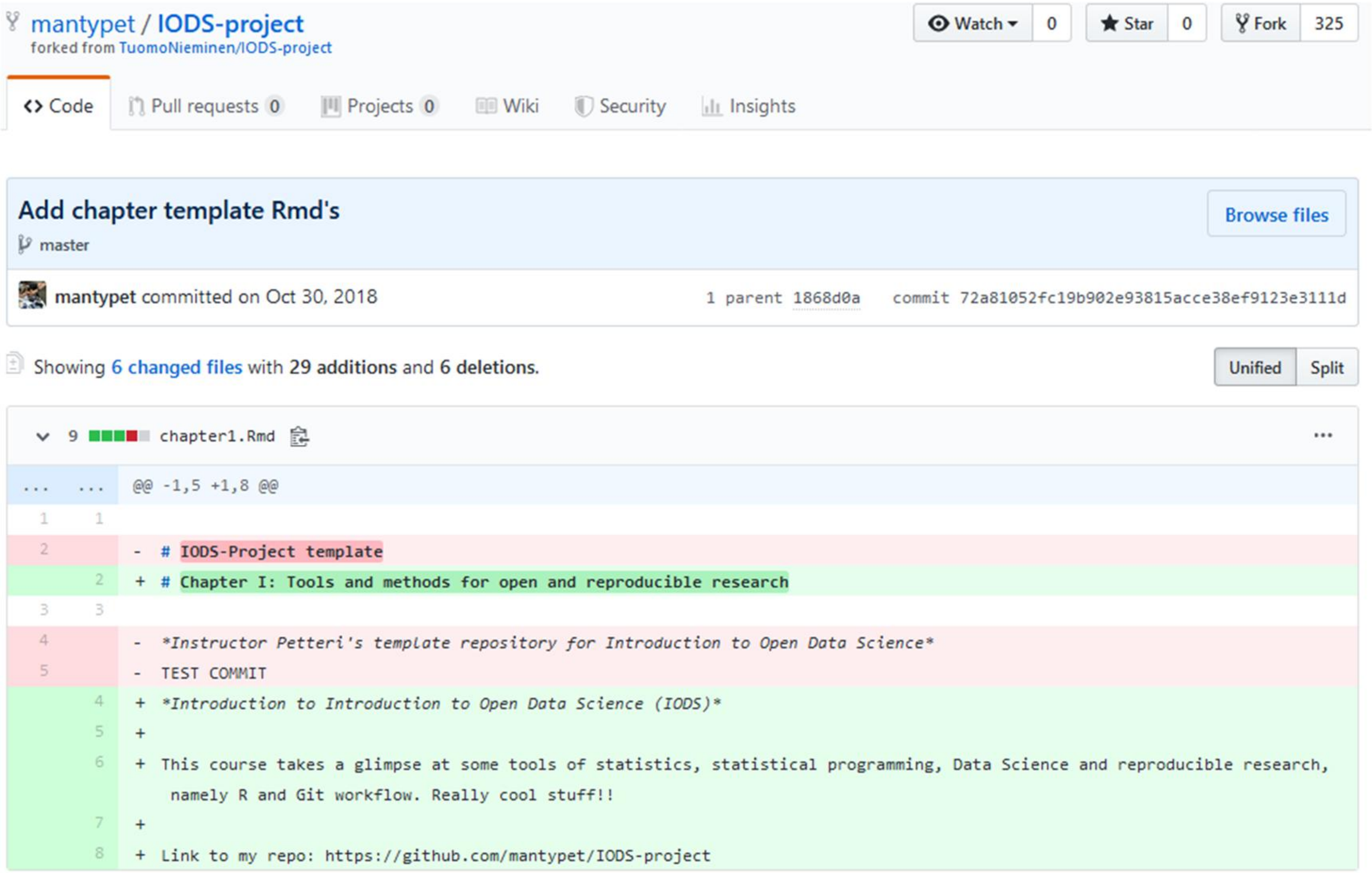
2. Open Data Science

- Weekly assignments consist of data wrangling and analysis exercises.
- They are practiced on DataCamp and then completed with RStudio.
- All the students’ weekly reports are saved and shared on GitHub, using ready-made templates downloaded on the first week.

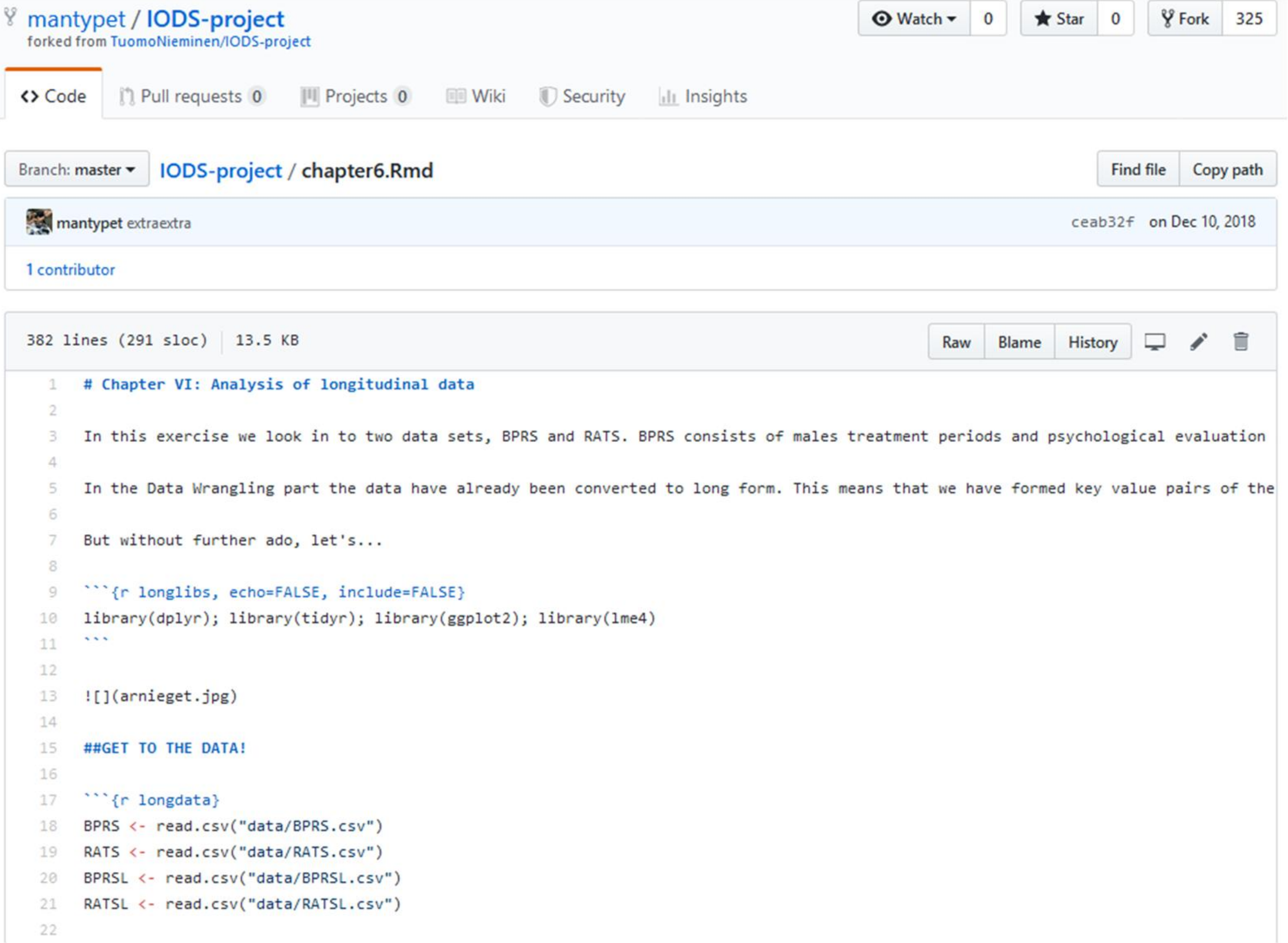


- R Markdown is used and the reports are *knitted* into HTML files.

2. Open Data Science



2. Open Data Science


mantypet / IODS-project
forked from TuomoNieminen/IODS-project
Watch 0 Star 0 Fork 325
Code Pull requests 0 Projects 0 Wiki Security Insights
Branch: master IODS-project / chapter6.Rmd
Find file Copy path
mantypet extraextra ceab32f on Dec 10, 2018
1 contributor
382 lines (291 sloc) 13.5 KB
Raw Blame History

````
# Chapter VI: Analysis of longitudinal data

In this exercise we look in to two data sets, BPRS and RATS. BPRS consists of males treatment periods and psychological evaluation

In the Data Wrangling part the data have already been converted to long form. This means that we have formed key value pairs of the

But without further ado, let's...

```{r longlibs, echo=FALSE, include=FALSE}
library(dplyr); library(tidyr); library(ggplot2); library(lme4)
```


![](arnieget.jpg)

##GET TO THE DATA!

```{r longdata}
BPRS <- read.csv("data/BPRS.csv")
RATS <- read.csv("data/RATS.csv")
BPRSL <- read.csv("data/BPRSL.csv")
RATSL <- read.csv("data/RATSL.csv")
```
````

2. Open Data Science

https://mantypet.github.io/IODS-project/#chapter-vi-analysis-of-longitudinal-data

IODS course project

Petteri Mäntymaa

Last updated 2018-12-10 17:47:51

(Petteri is one of my previous & precious Teaching Assistants and very important technical developers of the IODS course.)

/mantypet.github.io/IODS-project/#chapter-vi-analysis-of-longitudinal-data

Chapter VI: Analysis of longitudinal data

In this exercise we look in to two data sets, BPRS and RATS. BPRS consists of males treatment periods and psychological evaluation scores between treatment groups. RATS is about the growth of rats in different growth profile groups.

In the Data Wrangling part the data have already been converted to long form. This means that we have formed key value pairs of the variable under interest and the variable indicating different measurement times. To simplify, we get to have the time variable as, oh well – *a variable*, hence can take it in to account in our investigation and analysis!

But without further ado, let's...

GET TO THE DATA!

HIDE

```
BPRS <- read.csv("data/BPRS.csv")
RATS <- read.csv("data/RATS.csv")
BPRSL <- read.csv("data/BPRSL.csv")
RATSL <- read.csv("data/RATSL.csv")

BPRS$treatment <- factor(BPRS$treatment)
BPRS$subject <- factor(BPRS$subject)
RATS$ID <- factor(RATS$ID)
RATS$Group <- factor(RATS$Group)

BPRSL$treatment <- factor(BPRSL$treatment)
BPRSL$subject <- factor(BPRSL$subject)
RATSL$ID <- factor(RATSL$ID)
RATSL$Group <- factor(RATSL$Group)
```

Summaries and graphical inspections of RATS

Let's implement the analyses of BPRS to the RATS data and check out the wrangled RATSL.

CODE

2. Open Data Science

- Weekly peer-reviews of 3 reports for 6 weeks.
- Grading with a scale from 0 (fail) to 5 (excellent).
- Teachers check the integrity of the peer-reviews.
- Course grade is completely based on peer-reviews.

The grades of the IODS 3.0 are briefly summarized below.

GRADE	f	%	
1	5	6.0	*****
2	7	8.3	*******
3	6	7.1	******
4	23	27.4	***********************
5	43	51.2	***

3. Results

- The course has been a HUGE success story, so we are quite happy!
- Some excerpts from anonymous student feedback: (BOLD ADDED)

”I really enjoyed this course, to be honest this is the best course that I had in Helsinki. Combining both DataCamp and Rstudio exercise was amazing idea, it helped me alot. Even though I have been using R since couple of years but during this course I learned more sophisticated ways of programming.”

3. Results

- More excerpts from anonymous student feedback: (BOLD ADDED)

”The course was really interesting and hands-on approach worked well. Datacamp exercises were well organised. Need for this kind of applied statistical (data science) courses where you’re needed to clean your dataset and then use correct statistical methods is in high demand. You can get a feel that you’re learning something actually useful for real life. Learning Github has been really huge benefit.”

3. Results

- More excerpts from anonymous student feedback: (BOLD ADDED)

”First of all I want to thank you all about this course which has been the funniest and most interesting ever. This was my first touch to R, GitHub and Slack. I never thought that I would get this excited about something, but I did. I noticed that the R environment is an endless world and its not as difficult as I thought at first. I will definitely continue to learn codes and statistics.”

4. Conclusion

- There is a huge need for more (and more) data scientists.
- Teaching of statistics should focus more on data science, with a special emphasis on data wrangling.
- The statistics curriculum should be updated and the term “data science” used as a synonym to statistics.
- Our new course gives an excellent example of how to engage students to learn skills of reproducible, open data science.

5. References *(cited in the paper)*

- Greenacre, M. & Blasius, J., eds. (2006). *Multiple Correspondence Analysis and Related Methods.* Boca Raton, Florida: Chapman and Hall/CRC.
- Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences.* London: SAGE.
- R Core Team (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. *https://www.R-project.org/*
- Vehkalahti, K. & Everitt, B. S. (2019). *Multivariate Analysis for the Behavioral Sciences*, Second edition. Boca Raton, Florida: Chapman and Hall/CRC. *https://github.com/KimmoVehkalahti/MABS*
- Xie, Y., Allaire, J.J. & Grolemund, G. (2018). *R Markdown: The Definitive Guide.* Boca Raton, FL: Chapman and Hall/CRC.

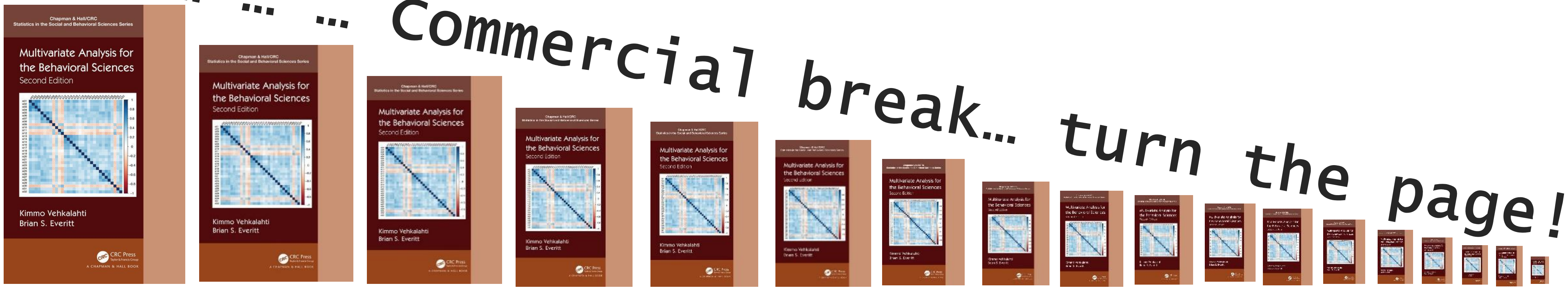
IASE 2019

Engaging Everyone with Open Data Science

62nd ISI World Statistics Congress
ISI 2019 Kuala Lumpur
18 - 23 August 2019

Thank you for your attention!

… … … Commercial break… turn the page! ☺



… … … Thanks! J Just a few slides from my invited talk in the 27th IWMS (Shanghai, China) in June 2019, entitled:

Multivariate Analysis for Data Scientists

Kimmo Vehkalahti

University of Helsinki, Finland

IWMS-2019: The 27th International Workshop on Matrices and Statistics

6-9 June 2019 | Shanghai, China

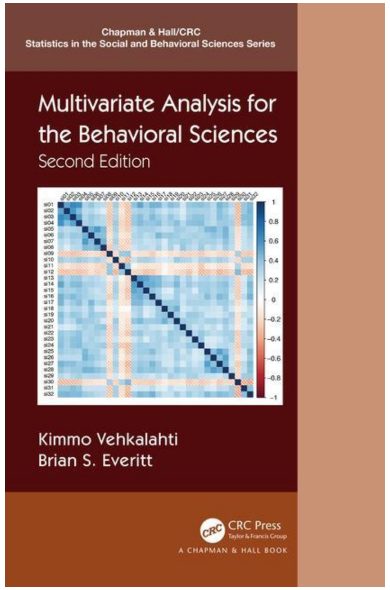
Kimmo Vehkalahti & Brian S. Everitt:

Multivariate Analysis for the Behavioral Sciences

Second Edition

Chapman and Hall/CRC Press, 2019

Table of Contents



Let us close with an example of MDS from Chapter 14:

- A view of the book (pp. 278-279):
  Data (dissimilarity matrix), analysis, figure, and interpretations

- A view of the same example from material freely available online on GitHub:
  Analysis and figure with R Markdown

https://www.crcpress.com/Multivariate-Analysis-for-the-Behavioral-Sciences-Second-Edi

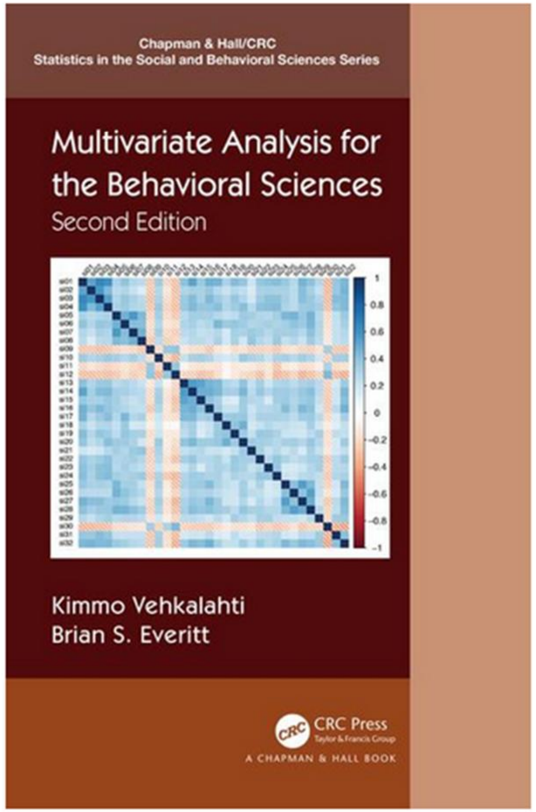
Finland ▾   Wish List   My Account   Contact Us

CRC Press
Taylor & Francis Group

SHOPPING CART

About Us ▾   Resources ▾   Textbooks ▾   Featured Authors   Search or Browse by Subject

Home / Statistics / Statistics for the Social and Behavioral Sciences / Multivariate Analysis for the Behavioral Sciences, Second Edition



Multivariate Analysis for the Behavioral Sciences, Second Edition

2nd Edition

Kimmo Vehkalahti, Brian S. Everitt

Hardback £49.99	eBook £44.99	eBook Rental from £25.00

CRC Press
Published January 8, 2019
Textbook - 415 Pages - 133 B/W Illustrations
ISBN 9780815385158 - CAT# K339858
Series: Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences

For Instructors   Request Inspection Copy

Select Format: Hardback

Quantity: 1

GBP £49.99

Add to Cart

Add to Wish List

FREE Standard Shipping!

Summary

Description

Table of Contents

Multivariate Analysis for the Behavioral Sciences, Second Edition is designed to show how a variety of statistical methods can be used to analyse data collected by psychologists and other behavioral scientists. Assuming some familiarity with

Instructors

We provide complimentary e-inspection copies of primary

TABLE 14.7
Proximity Matrix of Ten Remarkable Classical Music Composers Selected and Compared by Olli Mustonen

	Bac	Hay	Moz	Bee	Sch	Bra	Sib	Deb	Bar	Šos
Bach	0									
Haydn	50	0								
Mozart	30	10	0							
Beethoven	20	15	20	0						
Schubert	40	30	25	10	0					
Brahms	40	70	40	20	15	0				
Sibelius	40	90	70	25	60	20	0			
Debussy	50	50	50	80	50	70	35	0		
Bartók	30	80	80	60	70	70	35	15	0	
Šostakovitš	30	40	50	40	60	70	20	40	20	0

14.2.4 Mapping Composers of Classical Music

Our final example of the use of classical scaling involves data on composers of classical music, taken with permission from Mustonen (1996, 156–159) and Mustonen (1995, 167–170). Seppo Mustonen (a Finnish professor of Statistics) asked his son Olli Mustonen (a Finnish pianist, conductor, and composer) to select ten remarkable composers from different era of classical music and compare those composers with each other intuitively based on their entire production and style. Olli Mustonen made his comparisons using a scale from 0 to 100 in a way that the more he considered the composers to differ, the higher the score he gave. After about half an hour's reflection, he presented the proximity matrix given in Table 14.7, where the selected composers appear roughly in chronological order. We can see that the scale was applied with intervals of five units, and that the greatest difference was 90 units, occurring between Sibelius and Haydn.

Applying classical scaling to the data in Table 14.7 leads to four negative eigenvalues for the matrix **B** (see Technical Section 14.1) and so the dissimilarity matrix shown there is clearly non-Euclidean. Here we will look at the fit criteria described in Technical Section 14.1 as a guide to the number of dimensions needed to adequately represent the dissimilarity values in Table 14.7. For the one-dimensional solution we obtain the values

$$P_1^{(1)} = 0.35 \text{ and } P_1^{(2)} = 0.58$$

while for the two-dimensional solution, the values obtained are

$$P_2^{(1)} = 0.58 \text{ and } P_2^{(2)} = 0.83$$

which would seem to suggest two dimensions (although the first one does not approach 0.8 before eight dimensions). Also both the alternative criteria (the

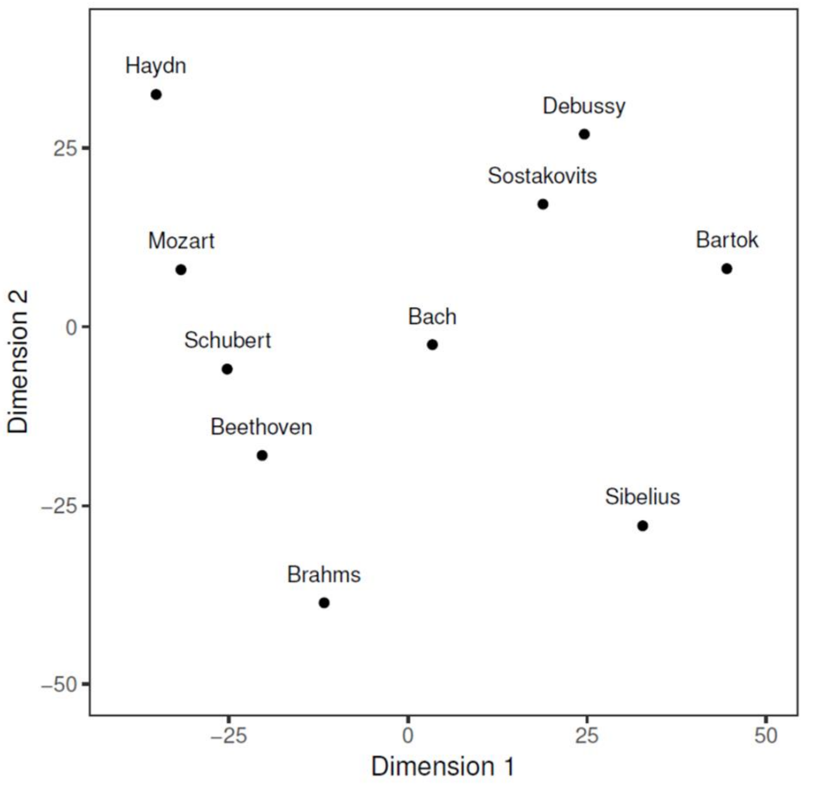


FIGURE 14.4
Resulting map from classical scaling of the classical composers.

trace and the magnitude) support the conclusion, so we shall proceed with two dimensions, following the original lines of interpretations of Mustonen (1996).

The resulting map of composers is shown in Figure 14.4. The first dimension (from left to right) appears to be related to time, with one significant exception: the "timeless" Bach is placed in the middle. The second dimension (from top to bottom) can be interpreted as a transition from "light" music to "heavy" music. Indeed, the *Viennese Classics* (Haydn, Mozart, Schubert, and Beethoven) form a logical chain, accompanied by Brahms, who, together with Sibelius, is located in the "heavyweight division". The modern composers (Debussy, Šostakovitš, and Bartók) seem to form a cluster of their own, and it is perfectly understandable that, of these composers, it is Šostakovitš who gets settled nearest to Bach. A rather lonely Sibelius is placed at a considerable distance from all other composers.

Let us close with an example of MDS from Chapter 14:

- A view of the book (pp. 278-279):
  Data (dissimilarity matrix), analysis, figure, and interpretations

- A view of the same example from material freely available online on GitHub:
  Analysis and figure with R Markdown

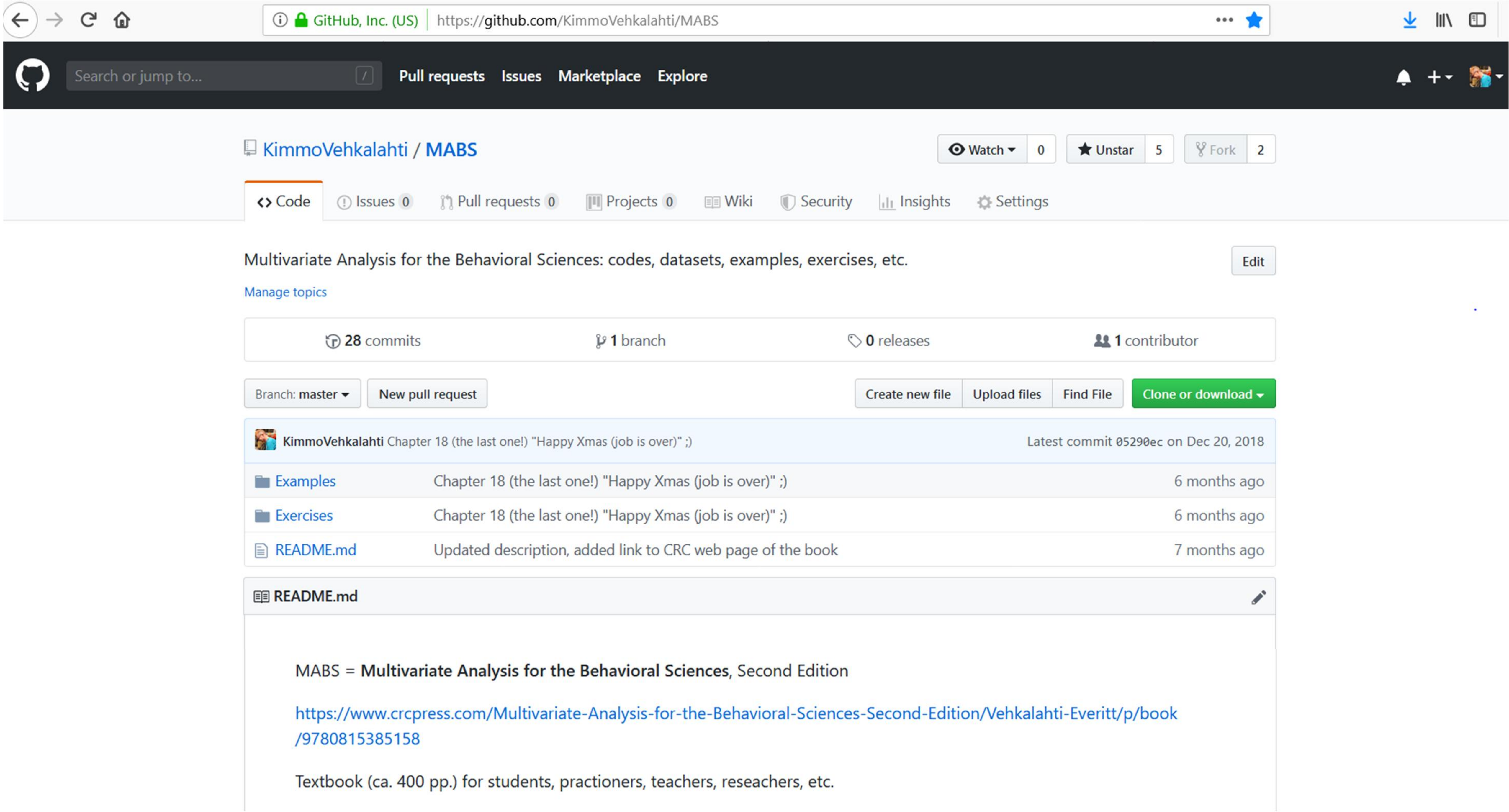

GitHub, Inc. (US)
https://github.com/KimmoVehkalahti/MABS
Search or jump to...
Pull requests
Issues
Marketplace
Explore
KimmoVehkalahti / MABS
Watch 0
Unstar 5
Fork 2
Code
Issues 0
Pull requests 0
Projects 0
Wiki
Security
Insights
Settings
Multivariate Analysis for the Behavioral Sciences: codes, datasets, examples, exercises, etc.
Edit
Manage topics
28 commits
1 branch
0 releases
1 contributor
Branch: master
New pull request
Create new file
Upload files
Find File
Clone or download
KimmoVehkalahti Chapter 18 (the last one!) "Happy Xmas (job is over)" ;)
Latest commit 05290ec on Dec 20, 2018
Examples
Chapter 18 (the last one!) "Happy Xmas (job is over)" ;)
6 months ago
Exercises
Chapter 18 (the last one!) "Happy Xmas (job is over)" ;)
6 months ago
README.md
Updated description, added link to CRC web page of the book
7 months ago
README.md
MABS = Multivariate Analysis for the Behavioral Sciences, Second Edition
https://www.crcpress.com/Multivariate-Analysis-for-the-Behavioral-Sciences-Second-Edition/Vehkalahti-Everitt/p/book/9780815385158
Textbook (ca. 400 pp.) for students, practioners, teachers, reseachers, etc.

Table 14.7: Proximity Matrix of Ten Remarkable Classical Music Composers Selected and Compared by Olli Mustonen

```
composers <- c("Bach", "Haydn", "Mozart", "Beethoven", "Schubert", "Brahms",
               "Sibelius", "Debussy", "Bartok", "Sostakovits")
OMD <- matrix(
c( 0,  50,  30,  20,  40,  40,  40,  50,  30,  30,
  50,   0,  10,  15,  30,  70,  90,  50,  80,  40,
  30,  10,   0,  20,  25,  40,  70,  50,  80,  50,
  20,  15,  20,   0,  10,  20,  25,  80,  60,  40,
  40,  30,  25,  10,   0,  15,  60,  50,  70,  60,
  40,  70,  40,  20,  15,   0,  20,  70,  70,  70,
  40,  90,  70,  25,  60,  20,   0,  35,  35,  20,
  50,  50,  50,  80,  50,  70,  35,   0,  15,  40,
  30,  80,  80,  60,  70,  70,  35,  15,   0,  20,
  30,  40,  50,  40,  60,  70,  20,  40,  20,   0
 ), nrow = 10, ncol = 10, byrow = TRUE, dimnames = list(composers, composers))

n <- dim(OMD)[1]
OMDS <- cmdscale(d = OMD, k = n-1, eig = TRUE, list. = TRUE)
```

```
## Warning in cmdscale(d = OMD, k = n - 1, eig = TRUE, list. = TRUE): only 5
## of the first 9 eigenvalues are > 0
```

```
as.matrix(format(OMDS$eig, scientific = FALSE, justify = "right", nsmall = 0L, digits = 0))
```

```
##       [,1]
##  [1,] " 7459"
##  [2,] " 4830"
##  [3,] " 2288"
##  [4,] "  752"
##  [5,] "  514"
##  [6,] "    0"
##  [7,] " -661"
##  [8,] " -906"
##  [9,] " -937"
## [10,] "-2912"
```

```
pk1 <- cumsum(abs(OMDS$eig))/sum(abs(OMDS$eig))
pk2 <- cumsum(OMDS$eig^2)/sum(OMDS$eig^2)
pk1
```

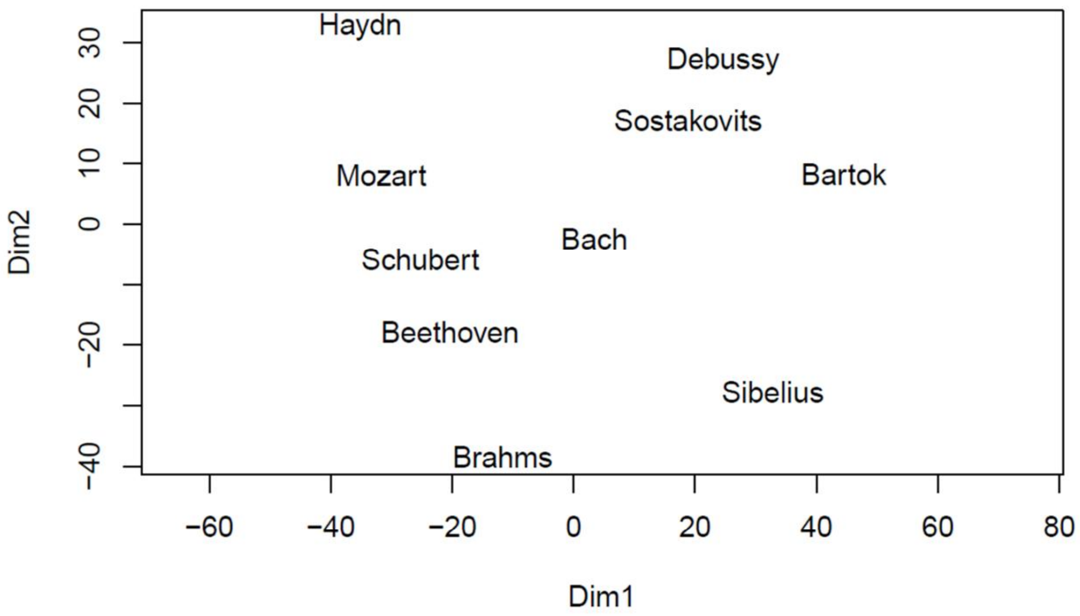
```
##  [1] 0.35 0.58 0.69 0.72 0.75 0.75 0.78 0.82 0.86 1.00
```

```
pk2
```

```
##  [1] 0.58 0.83 0.88 0.89 0.89 0.89 0.89 0.90 0.91 1.00
```

Figure 14.4

```
OwMDS <- wcmdscale(d = OMD, k = n-1, eig = TRUE)
plot(OwMDS, cex = 1.0)
```



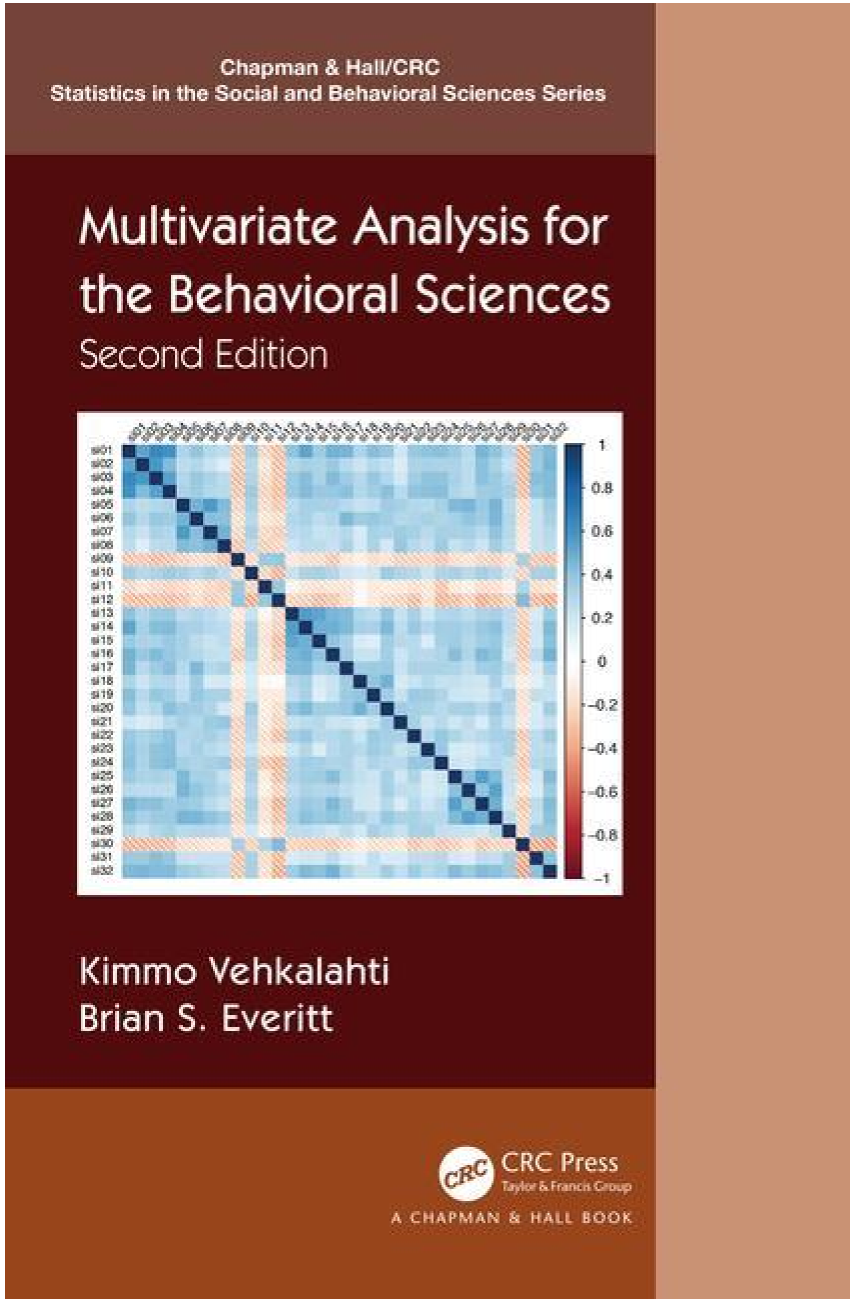

```
Owcscal <- as.data.frame(scores(OwMDS$points[, 1:2]))

library(ggplot2)

p1 <- ggplot(Owcscal, aes(x = Dim1, y = Dim2))
p2 <- p1 + geom_point() + geom_text(aes(label = composers),
                                    position = position_nudge(y = +4), size=4)
p3 <- p2 + scale_x_continuous(name = "Dimension 1", limits = c(-40, +50))
p4 <- p3 + scale_y_continuous(name = "Dimension 2", limits = c(-50, +40))
p5 <- p4 + theme_bw()
p6 <- p5 + theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank())
p7 <- p6 + coord_fixed(ratio = 1)
p7
```

谢谢

Thanks!


Chapman & Hall/CRC
Statistics in the Social and Behavioral Sciences Series
Multivariate Analysis for the Behavioral Sciences
Second Edition
Kimmo Vehkalahti
Brian S. Everitt
CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK