

Hiroshi Suzuki

2023-02-28

Contents

AI

2019

2019 9

2020

Data Analysis for Researchers

20

PDF ePub

PDF ePub

- PDF
- ePub

Chapter 1

R python
dio IDE integrated development environment, R R Stu-
Cloud Reproducibility Posit
ming R Markdown Literate Program-
Bookdown R Markdown bookdown
R

APPENDIX ??

variable object

Appendix A

MOOCs

2019 9 17

A.1 YouTube

PDF

A.2

A.2.1

A.2.1.1

A.2.1.1.1

- 5 :2016 1 22 ¹
– <https://www8.cao.go.jp/cstp/kihonkeikaku/index5.html>
 - :2016 12 21 -
 1. ()
 - 2.
 3.
<https://www8.cao.go.jp/cstp/kihonkeikaku/index5.html>
-

¹1 2

- (6)
 - (D-DRIVE²)2017 ~
 - - 20 :2019 1 8
 - http://www.mext.go.jp/b_menu/shingi/chousa/koutou/095/gaiyoku/1412367.htm
 - : 2019 3 26
 - https://www.meti.go.jp/shingikai/economy/risukei_jinzai/20190326_report.html
 - ()
 - AI :2019 3 29 <https://www8.cao.go.jp/cstp/aigensoku.pdf>
 - AI 2019 AI 2019 6 11 <https://www.kantei.go.jp/jp/singi/tougou-innovation/pdf/aisenryaku2019.pdf>
1. 1: (50 /) AI [MOOC]
 2. 2: (100 /) AI
 3. 3: (AI)

A.2.1.2

<http://www.mi.u-tokyo.ac.jp/consortium/index.html>

-
-

A.2.1.3

- 2019

A.2.2

A.2.2.1 IR

IR IR

A.2.2.2

2019 3 MOOCs

²Doctoral program for Data-Related InnoVation Expert

A.2.2.3 : (Data Analysis for Researchers)

- () 3
- :Rotary Peace Fellow, The Project for Human Resource Development Scholarship (JDS) 4 10-25
- :
- :2 (70 × 2 × 10) 1 1
- : (R Markdown etc.)
- :2014-2015³, (2016), 2017

A.2.2.4

1. Introduction to R, Open Data and Free Software
2. Basic R Objects and Commands
2. Data Frame Manipulation
3. Linear Regression and Graphics
4. Dynamic Documents Using Rmarkdown
6. Statistical analysis with R II
5. Statistical analysis with R III
6. Statistical analysis with R IV
7. Guest Lecture and preparation for presentations
8. Final presentations

A.2.2.5

R Studio (PC) R R Markdown

- Base R
 - cars:
 - iris:
- package MASS
- WDI: World Bank Development indicators for R


```
library(WDI)
#GDP (current US$)
gdp <- WDI(country = c("US", "JP", "CN", "KR"),
  indicator = "NY.GDP.MKTP.CD",
  start = 1960, end = 2017)
```

 - wbstats
- Quandl package: <https://www.quandl.com/tools/r>
- Google Trends: <https://trends.google.co.jp/>
- Yahoo Finance: <https://finance.yahoo.com/quote/DATA/>

³2015

(Computing for Researchers)

A.2.3

A.2.3.1

 ? ?

- :
-
-
-
-
-
-
- ()

AI

A.2.3.1.1

-
- ()

A.2.3.2 ()

-
-
-
-
- ()

—
—
—
— () ? ?
— () AI
— ()
—

-

A.2.3.3

? Liberal Arts?

-
-
-
-

A.2.3.4**Teaching to Learning**

- *
- - () Community of Inquiry (CoI), - Communication of Facts

A.2.3.5**Students with Various Backgrounds**

- :
- ()

A.2.3.6 Literacy**Expand your horizon!**

- :
—
— Q & A ()
—
- :
— I, II, A, B - :
—
- :
—
—

A.2.3.7 Resources**IT / Cloud**

- (Open/Public Data)
- (Online/Cloud)
- (Free and online/cloud system)

A.2.3.8 :

- 1.
- 2.
3. http://www.soumu.go.jp/menu_seisaku/ictseisaku/ictriyou/ope
ndata/

A.2.3.9 World Bank: Open Data Defined

The term “Open Data” has a very precise meaning. Data or content is open if anyone is free to use, re-use or redistribute it, subject at most to measures that

preserve provenance and openness.

1. The data must be **legally open**, which means they must be placed in the public domain or under liberal terms of use with minimal restrictions.
2. The data must be **technically open**, which means they must be published in electronic formats that are machine readable and non-proprietary, so that anyone can access and use the data using common, freely available software tools. Data must also be publicly available and accessible on a public server, without password or firewall restrictions. To make Open Data easier to find, most organizations create and manage Open Data catalogs.

A.2.3.10

List of Open Data Catalogue

- [:https://www.data.go.jp](https://www.data.go.jp)
- :
 - 1.
 - 2.
 - 3.
 - [:https://www.data.go.jp/list-of-database/](https://www.data.go.jp/list-of-database/)
 - [:https://www.jma.go.jp/jma/menu/menureport.html](https://www.jma.go.jp/jma/menu/menureport.html)
- U.S. Government's Open Data: <https://www.data.gov>
- EU Open Data Portal: <http://data.europa.eu/euodp/en/home>
- UK Open Data: <https://data.gov.uk>
- World Bank: New Ways of Looking at Poverty
 - Open Data: <https://data.worldbank.org>
 - World Development Indicators: <http://datatopics.worldbank.org/world-development-indicators/>
- UN Data: <http://data.un.org>
- WHO Data: <https://www.who.int/gho/en/>
- Google Public Data: :7 :136 <https://www.google.com/publicdata/directory>
- Open Knowledge Foundation: <https://okfn.org>
 - Global Open Data Index: <https://index.okfn.org>

A.2.3.10.1 Free Software, Online Access R

- R Project for Statistical Computing: <https://www.r-project.org>
- R Studio: <https://www.rstudio.com>
- R Studio Cloud: <https://rstudio.cloud>

A.2.3.10.2 Python

- Python: <https://www.python.org>
- Anaconda: <https://www.anaconda.com>
- Jupyter Notebook Cloud: Binder, Kaggle Kernels, Google Collaborate, CoCalc, PaizaCloud, etc.

A.2.3.10.3 Free Software

- Free Software, Free Society: Selected Essays of Richard M. Stallman: <https://www.gnu.org/philosophy/fsfs/rms-essays.pdf>
- Richard Stallman TEDxGeneva 2014: https://youtu.be/Ag1AKII_2GM

A.2.3.10.4 Online Learning Source**List of Online Help and Mini Courses**

- Online
 - TutorialPoint: <https://www.tutorialspoint.com/>
 - DataCamp: <https://www.datacamp.com/home>
 - Code Academy: <https://www.codecademy.com>
 - RStudio Premier: <https://rstudio.cloud/learn/primers>
- User Community

A.2.3.10.5 MOOCs

- OED: MOOC n. massive open online course, an educational course made available to a large number of people via the internet.
- First MOOC: 2008 by Dave Cormier, Connectivism and Connective Knowledge (CCK08)
- MIT OpenCourseWare 2002: ()OER Stanford U Model: UC Berkeley:
- MOOC : 2012 Stanford U. MIT

A.2.4 Massive Open Online Courses (MOOCs)**Moocs****A.2.4.1 MOOCs ****

-
-
-
- (datasets) () *

A.2.4.2 MOOCs

-
-
-
-

A.2.4.3 Moocs

A.2.4.3.1 Coursera Stanford U.

- Stanford U.: Machine Learning <https://www.coursera.org/learn/machine-learning>
- (:2,513,476)
- Johns Hopkins U.: Data Science, 10 courses [, R] <https://www.coursera.org/specializations/jhu-data-science>
- U. of Michigan: Applied Data Science, 5 courses [, python] <https://www.coursera.org/specializations/data-science-python>
- 100% U. of Illinois, U. of Michigan, U. of Colorado,

A.2.4.3.2 edX MIT Harvard U.

- Harvard U.: Data Science, 9 courses [, R]
- MIT: Statistics and Data Science, 5 courses [Graduate Level]
- <https://www.edx.org/micromasters/mitx-statistics-and-data-science>
- Microsoft, IBM, UC SanDiego

A.2.4.4 Professional Certificate in Data Science

A.2.4.4.1 : HarvardX, through edX

- URL: <https://online-learning.harvard.edu/series/professional-certificate-data-science>
- Book: <https://rafalab.github.io/dsbook/> (R Markdown Document)

A.2.4.4.2 R Data Camp Assessment

1. Data Science: R Basics; R
2. Data Science: Visualization;
3. Data Science: Probability;
4. Data Science: Inference and Modeling;
5. Data Science: Productivity Tools; Unix, Git, GitHub, R Markdown

- 6. Data Science: Wrangling;
- 7. Data Science: Linear Regression;
- 8. Data Science: Machine Learning;
- 9. Data Science: Capstone

A.2.4.4.3 Professional Certificate in Data Science

8 weeks 8 weeks 8 weeks 8 weeks 8 weeks 8 weeks 8 weeks 8 weeks 2 weeks
 Required R Packages for Examples: tidyverse, dslabs: <https://cran.r-project.org/web/packages/dslabs/dslabs.pdf>

-
- 2010 FBI
- Gapminder: Almost nobody knows the basic global facts! (Gapminder Test)
 - TED (Hans Rosling) <https://www.gapminder.org>
 - Health and income outcomes for 184 countries from 1960 to 2016
 - Country, Year, Infant deaths per 1000, Life expectancy in years, Average of children per woman, Country population, GDP, Continent, Geographical region
- Brexit
- 2016
- UC Berkeley
-
-

A.2.4.4.4 JMOOC MOOC

(1)

- Week 1
 -
 - - gacco:
- Week 2
 -
 - :
- Week 3
 - Excel, R, Python
- Week 4
 -
 - -
 -

A.2.4.4.5 JMOOC: [https:// www. jmooc. jp](https://www.jmooc.jp) 10 8

- II: ()

A.2.5

A.2.5.1

-
-
-
-
-
-

A.2.5.2 R Markdown Or jupyter notebook for python

- Reproducible Research:
- Literate Programming: Code, Script * html, doc, presentation, pdf (using TeX)
-

A.2.5.3

-
-
-
- Empirical ([,])
-

Quote: On Listening to Lectures, by Plutarch The correct analogy for the mind is not a vessel that needs filling, but wood that needs igniting - no more - and then it motivates one towards originality and instills the desire for truth. (<https://quoteinvestigator.com/2013/03/28/mind-fire/>)

()

A.2.6

A.2.6.1

Quote: Apple co-founders Steve Jobs and Steve Wozniak didn't have degrees when they launched what has become one of the most valuable companies in the world. And now Apple CEO Tim Cook is

spreading the word that would-be programmers really don't need the endorsement of a university to be able to create something of commercial value, such as an app for the Apple App Store. <https://www.zdnet.com/article/apple-ceo-tim-cook-you-dont-need-a-degree-to-code-mobile-apps/>

A.2.6.2

AI

Appendix B

MOOCs

3 12 (II) covid-19

B.1

(AI [1] 2019 6 11 AI 2019 AI
2025
1. (50 /) AI [MOOC]
2. (100 /) AI
3. (AI)
6 25% 1500 AI (100
/) 1 6
AI [3] 2000 2 GP 3
2012 4 [4] STEAM
1, 3 (AI)
AI AI

¹5 20 22 11 IT EXPO
²22000.6 ()
³3
⁴42012.8

B.2

AI AI 2019 AI

AI AI (Artificial Intelligence()) AI

2019 AI AI AI

AI

([11])

(,) (Decision Science) (Empirical

Study) ()

AI

(Recommendation System) Google, Amazon, Netflix

5

AI AI

([12])

AI

AI AI Artificial Gen-

eral Intelligence (AGI) AI

AI

B.3

B.3.1

5 ([7, 9])

(IoT,
Big Data)

(Public Data)

6

(R, Python)

edX, Coursera, JMOOC

MOOCs

‘Skills and Self - ID Top Factors’

[13]

([8])

GAFA

B.3.2

TA

TA

2025

5

AI

()

R

Free

()

AI

Respect

B.3.3

()

()

- [8] Doing Data Science by Rachel Schutt and Cathy O’Neil
- [9] Weapon’s of Math Destruction by Cathy O’Neil AI
- [10] edX Professional Certificate in Data Science: <https://online-learning.harvard.edu/series/professional-certificate-data-science> edX Data Science Ethics: <https://www.edx.org/course/data-science-ethics>
- [11] Factfulness by Hans Rosling, et. al.
- [12] How I learned to understand the world by Hans Rosling with Fanny Haerges-tam
- [13] Analyzing the Analyzers (O’Reilly) by Harlan Harris, Sean Murphy, and Marck Vais- man

Appendix C

1
2020 3 2020 6

C.1 ? AI?

(Data Science) 10
AI(Artificial Intelligence)
(Decision Making Science)
2 7 (Training Data, 2 7)
2 7 () ()
Google, Amazon (Recommendation Sys-
tem)
([1, 2])
AI Buzz Word AI
() AI
AI
(Ex-
poratory Data Analysis)
(Data Visualization)
AI AI
([1],[4]²)

¹2019 3 31 ()College of Liberal Arts, International Christian University 2019
⁴ ¹
² “Weapons of Math Destruction” . (Weapons of Mass Destruction))

([8])

C.2

– 10 [5] ? Gapminder
 (<https://www.gapminder.org>) 13 10
 (bias) TED ([9]) Gapminder
 2017 Google Google Pub-
 lic Data (https://www.google.com/publicdata/directory?hl=en_US&5B
 Gapminder]) ([6])
 () e-Stat (<https://www.e-stat.go.jp>)³

C.3 ?

AI
 4
 5 Free-Open-Online⁶
 (Empirical)
 Excel R python⁷ (Reproducible)
 (Literate Programming)
 python R
 R RStudio IDE⁸
 RStudio.cloud
 RStudio.cloud <http://RStudio.cloud> Sign Up Google
 Project New Project RStudio
 Tools Install Packages ... swirl Console library(swirl)
 Enter() swirl() Enter R
<https://foods4all.github.io>
 RStudio.cloud Learn Cheat Sheets() Learn
 Primer() The Basic() RStudio.cloud Log
 Out

³ (specialized agencies)

⁴Gapminder Tools World Bank <https://data.worldbank.org>

⁵ AI : <https://dashboard.e-stat.go.jp>

⁶

⁷

⁷ Free Software Foundation <https://www.fsf.org/about/>

⁸Integrated Development Environment /

2 1 R ([3]) tidyverse
 R
 2 ([2])
 MOOCs (Massive Open Online Courses) JMOOC(<https://www.jmooc.jp>)
 HarvardX Data Science [7]
 9 2 9 (Audit) R
 (Script) Google
 MOOCs
<https://foods4all.github.io> Free-Open-Online

C.4

10

()

- [1] Christopher Steiner 2013.
- [2] Rachel Schutt, Cathy O’Neil 2014.
- [3] R Hadley Wickham 2017. “R for Data Science” (<https://r4ds.had.co.nz>) Jeffrey B. Arnold (<https://jrnold.github.io/r4ds-exercise-solutions/>)
- [4] AI Cathy O’Neil 2018 .
- [5] Factfulness()– 10 Hans Rosling BP 2019.
- [6] Hans Rosling 2019.
- [7] :edX Professional Certificate in Data Science: <https://online-learning.harvard.edu/series/professional-certificate-data-science>
- [8] :edX Data Science Ethics: <https://www.edx.org/course/data-science-ethics>
- [9] Hans Rosling TED Talk(): https://www.ted.com/talks/hans_rosling_the_best_stats_you_ve_ever_seen?language=ja

⁹Data Science: R Basics, Visualization
¹⁰ 2

Appendix D

Data Analysis for Researchers

D.1

International Christian University QALL401
2020

D.2

2021 Bookdown 2022 2021 Book-
down
bookdown R Markdown R
 2022 2023 12

- Data Analysis for Researchers 2021
- Data Analysis for Researchers 2022

Appendix E

Introduction to R

E.1 70 x2

Data Analysis for Researchers ?? 2022 11 2 R 12 Data Analysis for Researchers

Data Analysis for Researchers Introduction

E.2

RMarkdown RNotebook [Code] RMarkdown
RStudio
• RNotebook

Appendix F

F.1

Data Analysis for Researchers ??					2023	2	20	22	24	70
R					Web	Google			Colab	R
Windows	R	R Studio			Google	Colab				
Public	Data				World	Development			Indicator	
Template										
Windows					R	R Studio			Posit	
Cloud	RStudio	Cloud	20		PC	Posit	Cloud			

F.2

- R Studio R Slide + Vido
 - R Markdown
- R RNotebook
- RNotebook
 - WDI
 -

Appendix G

Data Science for All

searchers ??

??

Data Analysis for Re-

G.1

- - Data Science for All -
 - Let's enjoy data science together!
- 2019 2020

Appendix H

Chat GPT

Chat GPT

DeepL

- First Chat
- Digital Health
-
- Big Mac Index

H.1 First Chat -

[HS] Explain the difference between equality and fairness by an example story

Chat GPT Equality refers to the idea that everyone should be treated the same, regardless of their personal characteristics or circumstances. Fairness, on the other hand, refers to the idea that people should be treated justly and in a way that is reasonable and appropriate given the specific situation.

Here is an example story that illustrates the difference between equality and fairness:

There are two children, Anna and Billy, who are both competing in a race at school. Anna is a fast runner, while Billy is slower. The teacher decides to give both children a head start in the race, with Anna starting 10 meters ahead of the starting line and Billy starting 5 meters ahead.

In this situation, equality would dictate that both children should start at the same point, regardless of their running abilities. However, fairness would dictate that the teacher's decision to give each child a different head start is appropriate, as it takes into account the fact that Anna is a faster runner and gives Billy a chance to catch up.

While equality aims to treat everyone the same, fairness recognizes that different people may have different needs or abilities and seeks to treat them in a way

that is fair and just.

[HS] What is the difference between a differentiable function and a continuous function in reals

Chat GPT A function is said to be differentiable at a point if it is possible to define a derivative of the function at that point. The derivative of a function at a point is a measure of how the function is changing at that point. A function is differentiable if it is possible to define a derivative at every point in its domain.

On the other hand, a function is continuous at a point if it is possible to draw the graph of the function without lifting your pencil from the paper. This means that there are no sudden jumps or breaks in the graph of the function at that point. A function is continuous if it is continuous at every point in its domain.

So, in summary, a differentiable function is one for which the derivative can be defined at every point, while a continuous function is one that does not have any breaks or jumps in its graph.

Both differentiability and continuity are important concepts in calculus and are often used to describe the behavior of functions. However, they are not the same thing and a function can be either one or both. For example, a function can be differentiable but not continuous, or continuous but not differentiable.

[HS] what is the difference between data science and statistical analysis? explain by examples

Chat GPT Data science and statistical analysis are related fields that involve the collection, processing, and analysis of data. However, they have some important differences.

Data science is a broad field that encompasses a wide range of techniques and methods for working with data. It includes statistical analysis, but it also includes other methods such as machine learning, data visualization, and data engineering. Data scientists use a variety of tools and techniques to analyze data, with the goal of extracting insights, making predictions, and solving problems.

For example, a data scientist might use machine learning algorithms to analyze a large dataset in order to predict customer behavior, or they might use data visualization techniques to create interactive graphics that help to communicate the results of an analysis.

Statistical analysis, on the other hand, is a more specific field that focuses on the use of statistical methods to analyze data. Statistical analysts use statistical techniques such as regression analysis, hypothesis testing, and analysis of variance (ANOVA) to understand relationships between variables and make predictions.

For example, a statistical analyst might use regression analysis to understand the relationship between a company's sales and its advertising budget, or they might

use hypothesis testing to determine whether there is a significant difference between the means of two groups.

In summary, data science is a broad field that encompasses a wide range of techniques and methods for working with data, while statistical analysis is a more specific field that focuses on the use of statistical methods to analyze data.

[HS] Write a short essay on the FIFA World Cup 2022 like Makuranosoushi of Seishonagon in Japanese

Chat GPT The FIFA World Cup 2022 is a highly anticipated international soccer tournament that will be held in Qatar. This will be the first time that the World Cup has been held in the Middle East, and it is expected to be a memorable and exciting event.

The tournament will feature 32 teams from around the world, who will compete in a series of matches to determine the ultimate champion. The teams will be divided into eight groups, with the top two teams from each group advancing to the knockout round.

The World Cup 2022 is sure to be filled with drama and excitement, as the world's best soccer players come together to compete on the biggest stage. The tournament will feature many talented and skilled players, including stars such as Lionel Messi, Cristiano Ronaldo, and Neymar Jr.

The World Cup 2022 will also be notable for its innovative and cutting-edge technology, as the tournament will feature the use of video assistant referees (VAR) and goal-line technology to ensure fair and accurate play.

Overall, the World Cup 2022 is sure to be a memorable and exciting event, and soccer fans around the world are eagerly anticipating its arrival.

[HS] What is Makuranosoushi of Seishonagon?

Chat GPT Makura no Soushi (), also known as The Pillow Book, is a Japanese literature classic written by Sei Shonagon during the Heian period (794-1185). It is a collection of personal observations, anecdotes, and reflections written in a diary-like format.

The Pillow Book covers a wide range of topics, including the beauty of nature, the customs and manners of the court, and the author's own personal experiences and thoughts. It is known for its wit, charm, and insight, and is considered a masterpiece of Japanese literature.

Sei Shonagon was a court lady and a contemporary of Murasaki Shikibu, the author of The Tale of Genji. Both The Pillow Book and The Tale of Genji are considered important works of literature from the Heian period and are widely studied and admired in Japan and around the world.

[HS] In R, correct the following codes.