Hiroshi Suzuki

2023-03-02

2

Contents

AI

2019

2019 9

2020   Data Analysis for Researchers   20

PDF ePub

PDF   ePub

- PDF
- ePub

Chapter 1

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

APPENDIX **??**

          variable   object

7

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
  ```

  –

---

[1]1 2

- （6 ）

- (D-DRIVE²)2017 ～

-
  - 20 :2019 1 8
  - http://www.mext.go.jp/b_menu/shingi/chousa/koutou/095/gaiyou/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.)
- :2014-2015[3], (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/

---

[3]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/opendata/

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

- :

  1.
  2.
  3.

  – :https://www.data.go.jp/list-of-database/

  – :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

- Phython: 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: ht tps://www.gnu.org/philosophy/ fsfs/rms-essays.pdf
- Richard Stallman TEDxGeneva 2014: https://youtu.be/Ag1AKIl_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.or
g/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   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

AI

B.1

AI　[1]　2019　6　11　　AI　2019　　　　　AI
(　　　　　　)　　　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

---
[1] 5　20　22　　11　IT　EXPO
[2] 22000.6　　　(　)
[3]
[4] 42012.8

B.2

AI 2019 AI

AI AI

AI (Artificial Intelligence( ) AI

2019 AI AI AI

([11])

(Decision Science) (Empirical
Study) (,) ( )

Google,
Amazon, Netflix (Recommendation System)

[5]

AI
)

AI
AI AI Artificial General Intelligence (AGI) AI

AI

B.3

B.3.1

(IoT,
Big Data)

(Public Data)

[6]

(R,
Python) edX, Coursera, JMOOC MOOCs

---

[5] ([7, 9])
[6]20

'Skills and Self - ID Top Factors'
[13]   ([8]
)   GAFA

B.3.2

                                     TA
            TA

2025   5   AI

                    R   Free

             (   )
       AI   Respect

B.3.3

                        (   )
               (   )

B.3.4

1. Broad Band Internet Access   Broad Band   Internet Access

2. RStudio (for R), Jupyter (for python)   Data Science   Interactive   Security

3.   Team   and   Collaborative   Teaching

4.   MOOCs OCW   Interactive

5.   Fact   Base,
   Evidence Base   AI   Black Box
                  Fake News Hate Speach   Data Sciece   r

Learning Goals

- A. Data Science: What? and Why?
  - Data Science ?   AI ? Data Scientists?
  - Data Science   ? -
- B. Business   Data Science

- C. Data Science   :Data   Data
- D. Data Analysis

B.4   References

[1]   AI (AI   )

[2] AI   2019   AI 2019   6   11

[3]   AI

[4]

[5]

[6]   ~ Moocs   (:   )

[7] AUTOMATE THIS: how algorithm came to rule our world by Christopher Steiner WikipediA

[8] Doing Data Science by Rachel Schutt and Cathy O'Neil Publisher's Site

[9] Weapon's of Math Destruction by Cathy O'Neil WikipediA, Review: Scientific American   AI

[10] edX Professional Certificate in Data Science, edX Data Science Ethics

[11] Factfulness by Hans Rosling, et. al. Gapminder, WikipediA

[12] How I learned to understand the world by Hans Rosling with Fanny Haergestam Book Information

[13] Analyzing the Analyzers (O'Reilly) by Harlan Harris, Sean Murphy, and Marck Vais- man Publisher's Site

B.5

AI   2020 3

HP: http://www.mi.u-tokyo.ac.jp/consortium/mc_literacy.html

B.5.1

-
  –
-
-           AI
  –   AI

-    AI
  -                               AI
  -
-
  -
  -   MOOCs
    Online
  -                               OCW,   MOOCs
- P.5   AI   (   )
  -                                     1500
  -
- P.9   AI
  -
  -
- P.10   AI
  -                               AI
- P.11   (~)
  -
- P.12 2.
  -
- P.13 3.   AI
  -
- P.19   AI
  - 1.   AI
  - 2.   Excel Google Spread Sheet   RStuio.cloud
    RStudio Server
- P.20 1.
  -                                TA
  -                                 10
- P.21 2.
  -      Exercise
  -
- P.23 3.
  -
  - MOOCs   Exercise
- P.24 4.   (   )
  - MOOCs
- P.27 2.
  -
  -
  -
  -

B.5.2

-
-
-
-
- Visualization

B.5.3

- OCW MOOCs
-
-
-

B.5.4

- naive utilitarianism AI
-
-

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 System)

　　　([1, 2])

　　　　　　　　　　　　AI 　　　　　　　　Buzz　Word　　AI
　　　( )　　　　　　　　　　　　　　　　　　　　　AI
　AI

　　　　　　　　　　　　　　　　　　　　　　　　　　　　(Exporatory　Data　Analysis)
　　　　　　　　　　　　　　　(Data Visualization)

　　　AI 　　　　　　　　AI
　　　　　　　([1],[4][2])

27

([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_US6%5B
Gapminder   ])   ([6])

(   ) e-Stat (https:
//www.e-stat.go.jp)   [3]

C.3   ?

AI
[4]

[5]   Free–Open–Online[6]
(Empirical)

Excel   R python   [7]   (Reproducible)
(Literate Programming)
python   R

R   RStudio IDE[8]
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

---

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_rosli
ng_the_best_stats_you_ve_ever_seen?language=ja

---

[9]Data 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 D   2022 11 2   R
                    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 D 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

F   Data Analysis for Researchers D

G.1

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

Appendix H

Chat GPT

Chat GPT   DeepL

1. First Chat
2. Digital Health
3.
4. Big Mac Index
5. map data of WIR2022   WIR2022
6. Miscellanous
7. Data Science Education

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.

Chat GPT It would be helpful if you could provide the specific codes that you would like me to correct. Without knowing what the code is supposed to do or what errors it is producing, it is difficult for me to provide a useful correction. Could you please provide more context or detail about the problem you are trying to solve?

[HS] Sorry, the code is: y<-"DA4R":print y

Chat GPT To correct the code, you can remove the quotation marks around "DA4R" and change the assignment operator <- to the more standard =. The corrected code would be:

```scss
Copy code


y = DA4R


print(y)
```

This code will assign the value DA4R to the variable y, and then print the value of y to the console.

Note that in R, it is generally recommended to use the assignment operator <- rather than =, as it can be easier to read and understand. However, = is also a valid assignment operator in R, so either one can be used depending on your preference.

[HS] How can we insert line break without sending the message in chat got?

Chat GPT There are a few ways to insert a line break in a message without sending it in a chat or messaging app:

1. Use the "Enter" or "Return" key on your keyboard. This will create a new line in the message, but will not send it until you press the "Send" button or key.
2. Use a line break character. Depending on the app or platform you are using, you may be able to insert a line break by typing a special character, such as \\n or \\r. For example, you could type Hello\nWorld to create a line break between the two words.
3. Use a line break code. Some apps and platforms support the use of special codes to insert line breaks. For example, in HTML, you can use the code to create a line break.
4. Use a formatting option. Many chat and messaging apps have formatting options that allow you to create a line break by applying a specific format to your text. For example, you may be able to use the "Bullets" or "Numbering" options to create a line break.

It's important to note that the specific method you use to insert a line break may depend on the app or platform you are using, as well as the capabilities of the device or computer you are using.

[HS] Let me try.

Chat GPT Sure, go ahead and try using one of the methods I described to insert a line break in your message. If you have any questions or run into any issues, feel free to ask for additional help.

[HS] I entered the line and then used return key. Then automatically the message was sent.

Chat GPT It sounds like the chat or messaging app you are using is set to automatically send messages when the "Enter" or "Return" key is pressed. This is a common feature in many apps and can be convenient for quickly sending messages, but it can also lead to accidental sends if you are not careful.

To avoid this issue, you can try one of the following options:

1. Use a different key to create a line break. Depending on the app or platform you are using, you may be able to use a different key to create a line break without sending the message. For example, you might be able to use the "Shift + Enter" or "Ctrl + Enter" keys to create a line break.
2. Change the app's settings. Some apps allow you to change the way the "Enter" or "Return" key works, such as by requiring you to use a different key or combination of keys to send messages. You may be able to find this option in the app's settings or preferences menu.
3. Use a different app. If the app you are using does not have a way to change the way the "Enter" or "Return" key works, you may want to consider using a different app that does offer this option. There are many chat and messaging apps available, and some of them may offer more control over when messages are sent.

I hope this helps! Let me know if you have any other questions or need further assistance.

[HS] OK. Shift + Enter worked

[HS] next try is Ctrl + Enter

Chat GPT Glad to hear that the "Shift + Enter" combination worked for you! It's good to know that you have a few different options for creating line breaks in your messages without sending them.

Regarding the "Ctrl + Enter" combination, this is another option that can be used to create a line break in some apps and platforms. However, it is important to note that the specific key combinations that are supported for creating line breaks may vary depending on the app or platform you are using.

If the "Ctrl + Enter" combination does not work for you, you may want to check the app's documentation or help resources to see if it supports this or any other

key combinations for creating line breaks. You could also try searching online for tips or tricks specific to the app or platform you are using.

I hope this helps! Let me know if you have any other questions or need further assistance.

[HS] Ctrl+Enter did not work. Thanks anyhow.

Chat GPT I'm sorry that the "Ctrl + Enter" combination did not work for you. It's possible that this key combination is not supported by the app or platform you are using, or that it has been disabled or overridden by another setting.

If you are unable to find a way to create line breaks using the keys on your keyboard, you may want to try using one of the other methods I mentioned earlier, such as using a line break character or code, or using a formatting option in the app.

For example, if the app supports it, you might be able to use the `\\n` character to create a line break, like this:

```
Copy code


Hello\nWorld
```

Or, you might be able to use a formatting option such as "Bullets" or "Numbering" to create a line break.

If you are still having trouble, you could try looking for help resources or documentation for the app or platform you are using, or consider trying a different app or platform that offers more control over line breaks.

I hope this helps! Let me know if you have any other questions or need further assistance.

[HS] Can you respond to messages not in English, say German, Chinese, Japanese, Korean, Bahasa Indonesia, Greek, Swahili, Hebrew? Since I am learning these languages.

Chat GPT As an AI language model, I am able to understand and respond to text input in a variety of languages, including German, Chinese, Japanese, Korean, Bahasa Indonesia, Greek, Swahili, and Hebrew. However, my ability to understand and produce text in these languages may vary depending on the complexity and clarity of the input, as well as my own capabilities and limitations as a language model.

I am trained on a large dataset of texts in a variety of languages, and I use this training to generate responses to user input. However, I do not have the same level of understanding and fluency in all languages, and my responses may not always be perfect or fully accurate.

If you are learning a new language and would like to practice with me, I would be happy to try to understand and respond to your input. Just keep in mind

that I may not always be able to produce a perfectly accurate or fluent response, particularly if the input is complex or difficult to understand.

I hope this helps! Let me know if you have any other questions or need further assistance.

[HS]

Chat GPT I'm sorry, but I'm unable to understand your input. Could you please provide more context or clarify your question?

[HS]

Chat GPT I'm sorry, but I'm still unable to understand your input. Could you please provide more context or clarify your question in a way that I can understand? Alternatively, you can try asking your question in English, and I will do my best to help you.

[HS]

Chat GPT

[HS] Would you tell me the importance of data science education for all students?

Chat GPT Data science education can be important for all students for a variety of reasons. Here are a few key benefits of data science education:

1. Data literacy: Data science education can help students develop important skills in data literacy, which is the ability to understand and use data to make informed decisions. In today's world, data is increasingly being used to drive decisions in all kinds of organizations, and being able to understand and work with data is becoming an increasingly important skill.

2. Problem-solving: Data science education can also help students develop problem-solving skills. Data science involves using data and analytical tools to identify patterns and trends, and to solve complex problems. These skills can be useful in a wide range of fields and industries.

3. Career opportunities: Data science is a growing field, and there is increasing demand for professionals with data science skills. Data science education can help students prepare for careers in data science or related fields, and can open up a wide range of career opportunities.

4. Real-world application: Data science education can also help students understand how data is used in the real world, and how it can be applied to solve problems and make decisions in a variety of contexts. This can

help students develop a more holistic understanding of the role that data plays in society.

Overall, data science education can be an important part of a well-rounded education, and can help students develop important skills and knowledge that can be useful in a variety of fields and industries.

[HS] Can we teach data science without mathematics?

Chat GPT It is generally challenging to teach data science without some level of mathematics. This is because data science often involves working with mathematical concepts and techniques, such as statistical analysis, probability theory, and machine learning algorithms.

However, the level of mathematics required for data science can vary depending on the specific area of focus within data science and the complexity of the problems being addressed. Some data science applications may require more advanced mathematical skills, while others may be more accessible to those with a basic understanding of math concepts.

Therefore, it may be possible to teach some aspects of data science without a strong foundation in mathematics, but a solid understanding of math is generally helpful for understanding and working with data science concepts and techniques. This is especially true for more advanced topics in data science, such as machine learning and statistical modeling.

Overall, while a strong foundation in mathematics can be beneficial for learning data science, it is not necessarily a requirement for all areas of the field. The specific math skills needed will depend on the specific goals and objectives of the data science education program, as well as the level of complexity of the problems being addressed.

[HS] What do you think about the data science education using world development indicator?

Chat GPT Using world development indicators as part of data science education can be a useful and informative approach. World development indicators are statistical measures that track the progress and status of countries around the world in various areas such as economic development, health, education, and the environment.

Using world development indicators in data science education can help students understand how data can be used to track and understand global trends and patterns, and can provide valuable insights into the challenges and opportunities facing different countries around the world.

In addition, working with world development indicators can help students develop skills in data analysis, visualization, and communication, as they explore and interpret the data. These skills can be valuable in a wide range of career fields and industries.

Overall, incorporating world development indicators into data science education can be a useful and informative way to engage students in the study of data and its applications, and can help them develop important skills and knowledge.

[HS] Would you give such a example? A topic of data science education using WDI?

Chat GPT Certainly! Here is an example of a topic in data science education that could use world development indicators:

Topic: Analyzing and visualizing trends in global economic development

Objective: Using data from the World Bank's World Development Indicators (WDI) database, students will analyze and visualize trends in economic development across countries over time.

Activities:

- Students will access the WDI database and select a set of indicators related to economic development, such as GDP per capita, poverty rate, and unemployment rate.
- Students will use data visualization tools to create charts and graphs showing the trends in these indicators over time for a selection of countries.
- Students will analyze the trends and patterns in the data, and consider possible factors that may be driving the changes.
- Students will write a report or presentation summarizing their findings and discussing the implications of the trends for economic development in the countries studied.

This topic could be adapted to different levels of difficulty and detail depending on the students' backgrounds and interests, and could be expanded to include additional indicators and countries as needed.

I hope this example gives you an idea of how world development indicators could be used in data science education. Let me know if you have any other questions or need further assistance.

H.1.1

[HS]

Chat GPT

10   5

2

[HS]

Chat GPT

ANOVA

2

[HS] FIFA   2022

Chat GPT FIFA   2022

32   8   2

2022

2022   VAR

2022

[HS]

Chat GPT

[HS]   y<-"DA4R":print y

       "DA4R"   <-   =

scss
Copy code

```
y = DA4R

print(y)
```

This code will assign the value DA4R to the variable y, and then print the value of y to the console.

Note that in R, it is generally recommended to use the assignment operator <- rather than =, as it can be easier to read and understand. However, = is also a valid assignment operator in R, so either one can be used depending on your preference.

[HS]

   1.   Enter   Return

2. Use a line break character. Depending on the app or platform you are using, you may be able to insert a line break by typing a special character, such as **\\n** or **\\r**. For example, you could type Hello\nWorld to create a line break between the two words.

3. Use a line break code.   HTML

4.

[HS]

Chat GPT

Enter   Return

1.   Shift + Enter   Ctrl +
   Enter

2.   Enter   Return

3.   Enter   Return

[HS] OK   Shift + Enter

[HS]   Ctrl + Enter

Chat GPT Shift + Enter

Ctrl + Enter

Ctrl + Enter

[HS] Ctrl+Enter

Chat GPT Ctrl + Enter

            ←

```
Copy code
```

```
Hello\nWorld
```

[HS]

Chat GPT AI

[   ]

[HS]

Chat GPT I'm sorry, but I'm unable to understand your input. Could you please provide more context or clarify your question?

[HS]

Chat GPT I'm sorry, but I'm still unable to understand your input. Could you please provide more context or clarify your question in a way that I can understand? Alternatively, you can try asking your question in English, and I will do my best to help you.

[HS]

Chat GPT

[HS]

Chat GPT

    1.

    2.

    3.

    4.

[HS]

Chat GPT

[HS]

Chat GPT

[HS]   WDI

Chat GPT

     (WDI)

- WDI   GDP
-
-
-

H.2 Digital Health -

2023.02.05

[HS] I want to know about the so-called digital health. For example, measuring the condition by the wearable device. Is there any academic research on the epidemiology using such devices?

[ChatGPT] Yes, there have been multiple studies and academic research done in the field of digital health, specifically regarding the use of wearable devices for monitoring health conditions. These studies have focused on a variety of topics such as the accuracy of wearable device readings, the potential for these devices to improve disease management and health outcomes, and the role of wearable devices in population health and epidemiology studies. For example, wearable devices have been used to monitor physical activity levels, heart rate, and sleep patterns, which can provide valuable insights into population health and disease trends. Additionally, wearable devices have been used to track

the spread of infectious diseases and to monitor disease outbreaks, providing valuable information for public health decision-making.

[HS] Would you give a couple of examples with sources?

[ChatGPT] Certainly! Here are a few examples of academic studies related to digital health and wearable devices:

1. "Validation of Wearable Sensors for Continuous Monitoring of Physical Activity and Sleep" - This study, published in the Journal of Medical Internet Research in 2016, aimed to validate the accuracy of wearable devices for monitoring physical activity and sleep. The results showed that wearable devices can provide reliable and valid measurements of physical activity and sleep patterns.
2. "The Potential of Wearable Devices in Chronic Disease Management: A Systematic Review" - This study, published in the Journal of Medical Systems in 2016, reviewed the literature on the use of wearable devices in the management of chronic diseases such as diabetes, heart disease, and respiratory disease. The results showed that wearable devices have the potential to improve disease management and health outcomes in these conditions.
3. "Using Wearable Devices to Track the Spread of Infectious Diseases" - This study, published in the Journal of Medical Internet Research in 2018, investigated the potential of wearable devices for tracking the spread of infectious diseases. The results showed that wearable devices can provide valuable data for disease surveillance and outbreak response.

These are just a few examples, but there are many more studies in the field of digital health and wearable devices. I hope this information is helpful!

[HS] Is there any discussion on regulations and moral standards for such privacy research?

[ChatGPT] Yes, there has been significant discussion and debate surrounding the regulations and ethical considerations for digital health research, particularly with regards to privacy. The use of wearable devices and other digital health technologies generates a large amount of personal data, which raises important privacy concerns.

For example, there are concerns about the security of personal health information and the potential for this information to be used for unauthorized purposes. There are also concerns about the potential for discrimination based on the information collected by wearable devices, such as discrimination in employment or insurance based on health status.

To address these concerns, there are a number of regulations and guidelines that have been developed to ensure the privacy and security of personal health information in digital health research. For example, the General Data Protection Regulation (GDPR) in the European Union and the Health Insurance Portabil-

ity and Accountability Act (HIPAA) in the United States are two examples of privacy regulations that apply to digital health research.

In addition to these regulations, there are also ethical guidelines and codes of conduct for digital health research that outline the responsibilities of researchers and organizations in protecting the privacy and security of personal health information. For example, the World Medical Association has developed the Declaration of Helsinki, which provides ethical guidelines for medical research involving human subjects, and the Council for International Organizations of Medical Sciences has developed the International Ethical Guidelines for Health-Related Research Involving Humans, which provides similar guidelines for international health research.

Overall, the regulations and ethical considerations for digital health research are constantly evolving and it is important for researchers and organizations to stay up to date on the latest developments in this field to ensure that privacy and security are protected.

[HS] Can you give leading institutions planning to do such research worldwide?

[ChatGPT] There are many institutions and organizations around the world that are conducting research in the field of digital health and wearable devices. Here are a few examples of leading institutions in this field:

1. Harvard Medical School - Harvard Medical School is one of the leading institutions for digital health research, with multiple research groups and initiatives dedicated to studying the impact of wearable devices and other digital health technologies on health and wellness.
2. Stanford University - Stanford University has a strong tradition of innovation in digital health, and its School of Medicine has multiple research programs focused on wearable devices and other digital health technologies.
3. MIT Media Lab - The MIT Media Lab is a research lab at the Massachusetts Institute of Technology that focuses on interdisciplinary research at the intersection of technology and health. They have several research groups dedicated to digital health, including the Biomedical Cybernetics group, which focuses on the development of wearable devices and other digital health technologies.
4. Imperial College London - Imperial College London is a leading institution for research in digital health, with multiple research programs and initiatives focused on wearable devices and other digital health technologies.
5. University of California, San Francisco - The University of California, San Francisco is a leading institution for digital health research, with multiple research programs focused on the development and implementation of wearable devices and other digital health technologies.

These are just a few examples, and there are many other institutions and organizations around the world that are conducting research in the field of digital health and wearable devices.

[HS] Do you know institutions or examples conducting research in a local region but in an intensive way to promote health?

[ChatGPT] Yes, there are many institutions and organizations around the world that are conducting research on digital health and wearable devices with a focus on promoting health in a local region. Here are a few examples:

1. Kermanshah University of Medical Sciences, Iran - Kermanshah University of Medical Sciences has established a research center for digital health, which aims to promote health in the local region through the use of wearable devices and other digital health technologies. This center conducts research on a variety of topics related to digital health, including the development of wearable devices for disease management and the evaluation of the impact of digital health interventions on health outcomes.
2. National University of Singapore, Singapore - The National University of Singapore has established a research program in digital health, which aims to promote health in the local region through the use of wearable devices and other digital health technologies. This program conducts research on a variety of topics related to digital health, including the development of wearable devices for disease management and the evaluation of the impact of digital health interventions on health outcomes.
3. University of the Witwatersrand, South Africa - The University of the Witwatersrand has established a research program in digital health, which aims to promote health in the local region through the use of wearable devices and other digital health technologies. This program conducts research on a variety of topics related to digital health, including the development of wearable devices for disease management and the evaluation of the impact of digital health interventions on health outcomes.

These are just a few examples, and there are many other institutions and organizations around the world that are conducting research in the field of digital health and wearable devices with a focus on promoting health in a local region.

[HS] What do you predict the future of epidemiology in connection with this digital health or research using IT devices such as wearable devices monitoring health?

[ChatGPT] The use of wearable devices and other digital health technologies has the potential to revolutionize epidemiology and the study of disease and health outcomes. Here are a few ways that digital health and wearable devices may shape the future of epidemiology:

1. Increased Data Collection: Wearable devices and other digital health technologies generate a large amount of data on individuals' health and behaviors, which can be used to better understand patterns of disease and health outcomes. This increased data collection has the potential to greatly enhance our understanding of the causes of disease and the factors that influence health outcomes.

2. Real-Time Monitoring: Wearable devices allow for real-time monitoring of individuals' health, which has the potential to greatly improve our ability to detect outbreaks of disease and track the spread of illness. This real-time monitoring has the potential to greatly enhance our ability to respond to public health emergencies and reduce the spread of disease.

3. Improved Data Analysis: With the increased data collection made possible by wearable devices and other digital health technologies, there will be a growing need for advanced data analysis methods and tools to make sense of this data. This will require new and innovative approaches to data analysis, including the use of machine learning and artificial intelligence.

4. Personalized Medicine: The data generated by wearable devices and other digital health technologies has the potential to inform the development of personalized medicine, where treatments are tailored to an individual's specific needs and health status. This has the potential to greatly improve health outcomes and reduce the risk of adverse events associated with treatments.

Overall, the future of epidemiology in connection with digital health and wearable devices is very promising, with the potential to greatly enhance our understanding of disease and improve health outcomes. However, it is important to ensure that the privacy and security of personal health information are protected as we move forward with this research.

[HS] Thank you very much. I want to come back to you after reviewing your comments.

[ChatGPT] You're welcome! I'm here to help, so feel free to reach out to me again if you have any additional questions. Have a great day!

H.2.1

[HS]

[ChatGPT]

[HS]

[ChatGPT]

1. "Validation of Wearable Sensors for Continuous Monitoring of Physical Activity and Sleep" - 2016 Journal of Medical Internet Research

2. "The Potential of Wearable Devices in Chronic Disease Management" A Systematic Review" - 2016 Journal of Medical Systems

3. "Using Wearable Devices to Track the Spread of Infectious Diseases" - 2018 Journal of Medical Internet Research

[HS]

[ChatGPT]

                                                            GDPR   HIPAA

[HS]

[ChatGPT]

1.   -   1
2.   -
3. MIT   - MIT   Biomedical
   Cybernetics
4. Imperial College London - Imperial College London
5.   -

[HS]

[ChatGPT]

1.   -
2.   -
3.   -

[HS]   IT

[ChatGPT]

1.   1.

2.

3.   3.

4.

[HS]

[ChatGPT]

www.DeepL.com/Translator

H.3

[HS] 60

Chat GPT 60

[HS]

Chat GPT

[HS]

Chat GPT

   1.
   2.
   3.

   1.
   2.
   3.

[HS]

Chat GPT

   3.

[HS]

Chat GPT   OpenAI

[HS]

Chat GPT

   1.   GDP 50
   2.   GDP 60
   3.   GDP 60

[HS]

Chat GPT

   20   1950   1980

[HS]

Chat GPT Hmm...something seems to have gone wrong. Maybe try me again in a little bit.

Chat GPT

[HS]

Chat GPT

[HS]

Chat GPT   Web

[HS]   Web

Chat GPT   URL

https://www.government.se/government-policy/education/

[HS]

Chat GPT

[HS]

Chat GPT Yes, I am more proficient in English and can respond faster in English. I am also trained to understand and respond in other languages including French, Spanish, German, Italian, Portuguese, Dutch, and Russian.

[HS] Thank you. Then I will get back to you in English next time.

Chat GPT You're welcome! I'm here to help. Don't hesitate to reach out to me in English or any other language I am capable of.

H.4   Big Mac Index

[HS] in R would you teach me how to rewrite data.table .SD in tidyverse

[ChatGPT] Certainly! In data.table, .SD refers to the subset of data for each group of a grouped data.table. In tidyverse, you can use the dplyr package to achieve similar functionality.

Assuming you have loaded the tidyverse package, here is an example of how you can rewrite a data.table code block that uses .SD:

```R
Copy code
library(tidyverse)


# Load example data
data(mtcars)
```

```
mtcars_dt <- as.data.table(mtcars)

# Original data.table code
mtcars_dt[, lapply(.SD, mean), by = cyl]

# Equivalent code using tidyverse
mtcars %>%
  group_by(cyl) %>%
  summarize(across(everything(), mean))
```

In this example, we first load the tidyverse package and convert the mtcars data frame to a data.table. The original data.table code calculates the mean of each column for each group defined by the cyl columng. The equivalent code using tidyverse first groups the data by cyl, then uses summarize() and across() to calculate the mean of every column.

Note that across() takes a column selection helper, such as everything(), to specify which columns to apply the function to. In this case, everything() means "apply the function to every column".

[HS] can you rewrite the following, which is taken from economist big mac indices.

[ChatGPT] Certainly! Here's an example of how you can rewrite the data.table code used to calculate the Big Mac index from The Economist using tidyverse:

```
R
Copy code
library(tidyverse)

# Load data
big_mac_data <- read.csv("big-mac-full-index.csv")

# Original data.table code
big_mac_data[, .(raw_index = mean(local_price / dollar_ex)), by = date]

# Equivalent code using tidyverse
big_mac_data %>%
  group_by(date) %>%
  summarize(raw_index = mean(local_price / dollar_ex))
```

In this example, we first load the tidyverse package and read in the data using read.csv(). The original data.table code calculates the Big Mac index by dividing the local price by the exchange rate for each observation, then takes the mean of these ratios for each date. The equivalent code using tidyverse groups the data by date, then uses summarize() to calculate the mean of the local_price / dollar_ex ratio for each group.

Note that in tidyverse, summarize() is used to calculate summary statistics for

each group, while mutate() is used to create new variables based on existing variables. In this case, we are calculating a summary statistic (mean()) for the local_price / dollar_ex ratio, so we use summarize().

[HS]

```
for(currency in base_currencies) {
    big_mac_index[ ,                               # we don't want a subset, so our first a
        (currency) :=                   # we'll add a new column named for the base set
            dollar_price /              # we divide the dollar price in each row by
                                        # the dollar price on the *base currency*'s row (.S
            .SD[currency_code == currency]$dollar_price -        # that contains only t
            1,                          # one means parity (neither over- nor under-valued]
                                        # to get an over/under-valuation value
        by=date                         # and of course, we'll group these rows by date
    ]
}
```

[ChatGPT] Here's an example of how you can rewrite the data.table code that calculates the Big Mac index for multiple base currencies using tidyverse:

```R
Copy code
library(tidyverse)

# Load data
big_mac_data <- read.csv("big-mac-full-index.csv")

# Define base currencies to loop through
base_currencies <- unique(big_mac_data$currency_code)

# Equivalent code using tidyverse
big_mac_data <- big_mac_data %>%
  split(.$date) %>%
  map_df(~ {     .x %>%
      mutate(across(contains("price"), ~ dollar_price / .x[dollar_price != 0 & currency
      pivot_longer(cols = ends_with("_price"), names_to = "price_type", values_to = "pr
      filter(price_type == "local_price") %>%
      select(-price_type) %>%
      pivot_wider(names_from = "currency_code", values_from = "price_value")
}, .id = "date")

# Clean up column names
names(big_mac_data) <- gsub("\\.", "_", names(big_mac_data))
```

In this example, we first load the tidyverse package and read in the data using read.csv(). The original data.table code calculates the Big Mac index for multiple base currencies by looping through base_currencies and adding a new

column to the big_mac_index data.table for each base currency. The equivalent code using tidyverse uses split() and map_df() to group the data by date, then apply a series of transformations to calculate the Big Mac index for each base currency.

In the map_df() call, we use mutate() and across() to divide the dollar price for each observation by the dollar price for the base currency's row, and subtract 1 to get the over/undervaluation value. We then use pivot_longer() to convert the wide-format data to long format, filter for the local_price values, and use pivot_wider() to convert the data back to wide format with a column for each currency. Finally, we use gsub() to clean up the column names by replacing periods with underscores.

Note that in tidyverse, it is generally recommended to avoid using loops and instead use functions like map() and purrr::walk() to iterate over data frames or lists. In this case, we use split() and map_df() to group the data by date and apply a transformation to each group.

[HS] Give me time to understand what you wrote. I will get back to you later.

[ChatGPT] No problem, take your time! Let me know if you have any questions or if there's anything I can clarify.

H.4.1

[HS] R tidyverse data.table .SD

[ChatGPT] ! data.table .SD data.table Tidyverse dplyr

tidyverse .SD data.table

```R
library(tidyverse)

# Load example data
data(mtcars)
mtcars_dt <- as.data.table(mtcars)

# Original data.table code
mtcars_dt[, lapply(.SD, mean), by = cyl]

# Equivalent code using tidyverse
mtcars %>%
  group_by(cyl) %>%
  summarize(across(everything(), mean))
```

tidyverse mtcars data.table data.table cyl Tidyverse cyl

across() everything() everything()

[HS]

[ChatGPT]   The Economist   data.table   tidyverse

```R
Copy code
library(tidyverse)

# Load data
big_mac_data <- read.csv("big-mac-full-index.csv")

# Define base currencies to loop through
base_currencies <- unique(big_mac_data$currency_code)

# Equivalent code using tidyverse
big_mac_data <- big_mac_data %>%
  split(.$date) %>%
  map_df(~ {    .x %>%
      mutate(across(contains("price"), ~ dollar_price / .x[dollar_price != 0 & currency
      pivot_longer(cols = ends_with("_price"), names_to = "price_type", values_to = "pr
      filter(price_type == "local_price") %>%
      select(-price_type) %>%
      pivot_wider(names_from = "currency_code", values_from = "price_value")
}, .id = "date")

# Clean up column names
names(big_mac_data) <- gsub("\\.", "_", names(big_mac_data))
```

      tidyverse   read.csv()   data.table   base_currencies   big_mac_index
data.table   Tidyverse   split() map_df()   Big
Mac

map_df()   mutate() across()   1   pivot_longer()
                        local_price   pivot_wider()   gsub()

Tidyverse   map() purrr::walk()   split()
  map_df()

[HS]

[ChatGPT]

H.5   map data of WIR2022

[HS] I want to study World Inequality Report 2022

[ChatGPT] I am sorry, World Inequality Report 2022 is not yet published. The
most recent report available is the World Inequality Report 2018 which was

published by the World Inequality Lab. It provides a comprehensive view of global economic inequality, using data on income, wealth, and consumption. You can find it on their website: https://wir2018.wid.world/

[HS] It is already published here: https://wir2022.wid.world/

[ChatGPT] I apologize for the confusion, as per my knowledge cut-off in 2021, the World Inequality Report 2022 was not yet published, However, you are correct that the report is now available on the website you provided. The World Inequality Report 2022 provides an updated view of global economic inequality, including new data and analyses on topics such as the distribution of wealth, the evolution of income inequality, and the impact of taxes and transfers on inequality. You can access the full report, as well as related data and research, on the World Inequality Lab's website.

[HS] Starting from its executive summary, I want to recreate each chart by R first.

[ChatGPT] Recreating the charts from the World Inequality Report 2022 using R is a great way to gain a deeper understanding of the data and analyses presented in the report.

To start, you will need to download the data used in the report. The data used in the World Inequality Report 2022 is available on the World Inequality Database website (https://wid.world/data/) in a variety of formats, including CSV and XLS.

Once you have downloaded the data, you can use R to read in the data and begin analyzing it. There are several libraries in R that can be used for data visualization, such as ggplot2, lattice, and plotly. These libraries provide a wide range of options for creating different types of charts and plots, and are easy to use.

To recreate the charts in the report, you will need to understand the data structure and the methods used in the report. You can also refer the R codebook provided by WID.world, that will help you recreate the charts from the report using R.

Keep in mind that recreating the charts may take some time and practice, but with a good understanding of the data and the R programming language, you should be able to recreate the charts and gain a deeper understanding of the data presented in the World Inequality Report 2022.

[HS] Yes, thank you. I have tried recreating some of them using tidyverse package.

[ChatGPT] Great! The tidyverse package in R is a collection of popular packages for data manipulation, visualization, and modeling, and is a great choice for working with the data from the World Inequality Report 2022. The package includes ggplot2 for data visualization, dplyr for data manipulation, and tidyr for data tidying, which are all useful for recreating the charts from the report.

To recreate the charts from the report, you will need to first clean and reshape the data to match the format used in the report. The dplyr and tidyr packages provide a wide range of functions for cleaning and reshaping data, making it easy to prepare the data for analysis and visualization.

Once you have the data prepared, you can use ggplot2 to create the charts. ggplot2 is a powerful visualization library that provides a wide range of options for creating different types of charts, including line plots, bar charts, scatter plots, and more. The library also allows you to easily customize the appearance of your charts, such as changing colors, font sizes, and axis labels.

You can also use other packages like plotly for interactive visualization, or create map plots using ggmap or leaflet packages.

If you have any specific questions or issues, feel free to ask me again.

[HS] I worked on F3 of the executive summary and the map chart. Do you know the map data used there.

[ChatGPT] The World Inequality Report 2022, as well as the data provided on the World Inequality Database website, use a wide range of data sources to create the charts and analyses presented in the report. The data used in the map chart you mentioned, is provided by the World Inequality Database and is based on a combination of income and consumption data from national and international sources.

The data sources used in the World Inequality Report 2022 varies by country and by type of data. The data sources include national statistical offices, central banks, and other national and international organizations. The report also uses data from household surveys, such as the World Bank's Living Standards Measurement Study and the European Union's Survey of Income and Living Conditions, as well as data from tax records and wealth surveys.

The map data can be found in the World Inequality Database under the "Inequality of Household Income" section, and you can download it in a CSV format.

Please note that the data and the methodology used to create the charts in the report may be complex and may require a significant amount of time and expertise to understand and replicate. The WID team also provides a detailed documentation and R codebook to help with data replication and visualization.

[HS] Good! I will try to find it.

[ChatGPT] I'm glad to hear that. The World Inequality Database provides a wealth of information and data on global economic inequality, and the data used in the map chart you mentioned should be available on the website. You can access the data by navigating to the "Inequality of Household Income" section, and selecting the appropriate country and data type. The data is provided in a CSV format, which can be easily read into R for further analysis and visualization.

I recommend you to also read the documentation provided by the WID team, as it will give you a better understanding of the data and the methods used to create the charts in the report. The documentation includes information on the data sources, the methods used to estimate and analyze the data, and any limitations or uncertainties in the data.

Keep in mind that recreating the charts from the report may take some time and practice, but with a good understanding of the data and the R programming language, you should be able to recreate the charts and gain a deeper understanding of the data presented in the World Inequality Report 2022.

If you have any further questions or issues, please let me know.

[HS] As a matter of fact, I started out using the map_data("world") of the ggplot2 package. However, I needed to change some of the country names to fit WIR data. Sometimes it involves sensitive issues.

[ChatGPT] Yes, it's common to encounter discrepancies in country names and codes when working with data from different sources. The map_data("world") function in the ggplot2 package uses a specific set of country names and codes, which may not match the names and codes used in the World Inequality Report 2022.

To change the country names and codes to match the data from the World Inequality Report 2022, you can use the rename() function from the dplyr package. You can use this function to rename the country names and codes in your data to match the names and codes used in the report.

You can also use left_join() function from dplyr package to join the data from World Inequality Report with the map data, and then use the country code of WIR data to map it with the map data.

It's also important to keep in mind that some country names and codes may be sensitive and may be subject to political or cultural disputes. It is important to be mindful of these issues and to use the country names and codes that are most appropriate and respectful in your analysis.

If you have any specific questions or issues, feel free to ask me again.

[HS] I thought that `rename()` function works only for column names.

[ChatGPT] You are correct that the rename() function in the dplyr package is typically used to rename column names in a data frame. However, it can also be used to rename other elements of a data frame, such as the levels of a factor variable.

In the case of the map data, the country names are typically stored as levels of a factor variable, which can be accessed and modified using the levels() function in R.

You can use mutate() function from dplyr package and using levels(data$country) and data$country <- factor(data$country, levels = levels) to change the levels of the factor variable

containing the country names.

Additionally, you can use stringr package to clean the country names, in case there are some inconsistencies with the names used in the World Inequality Report 2022.

It's also important to keep in mind that some country names may change over time. It's a good idea to double-check that your country names match the latest data from the World Inequality Report.

[HS] I used a combination of mutate and case_when to mend the data one by one.

[ChatGPT] Yes, using a combination of mutate() and case_when() from the dplyr package is a good approach to change specific country names to match the data from the World Inequality Report 2022.

The mutate() function can be used to add a new column to the data frame, and the case_when() function can be used to specify the conditions for renaming the country names.

You can use case_when() inside mutate() function to change the country names one by one. For example, if you want to change the name of country 'A' to 'B', you can use the following code:

```
Copy code
data <- data %>%
mutate(country = case_when(country == "A" ~ "B",
                                  TRUE ~ country))
```

It's also a good idea to double-check that your country names match the latest data from the World Inequality Report 2022.

Please also keep in mind that some country names may change over time, and that some country names may be sensitive and may be subject to political or cultural disputes. It's important to be mindful of these issues and to use the country names that are most appropriate and respectful in your analysis.

[HS] I believe that left_join you suggested is also helpful, but I could not check what needs to be done after I changed some of them using the above method.

[ChatGPT] Yes, using the left_join() function from the dplyr package is another approach to match the country names and codes in your data to those used in the World Inequality Report 2022.

With left_join function, you can join the data from the World Inequality Report with the map data, and then use the country code of WIR data to map it with the map data.

You can join the data from the World Inequality Report 2022 with the map data on a common column, typically the country code, using the following code:

```
Copy code
map_data_with_inequality_data <- left_join(map_data, World_Inequality_Report_data, by = c("Countr
```

After joining the data, you can use the country codes from the World Inequality Report 2022 to map the data to the map data.

Please keep in mind that this approach is useful when the country codes match in both datasets, if not you need to change them accordingly.

It is always a good idea to check the data after joining and to make sure that the country names and codes match correctly.

If you have any other questions or issues, please let me know.

[HS]

```
world_map <- map_data("world")
world_map_wir <- world_map %>%
  mutate(region = case_when(region == "Democratic Republic of the Congo" ~ "DR Congo",
                            region == "Republic of Congo"  ~ "Congo",
                            region == "Russia"  ~ "Russian Federation",
                            region == "South Korea" ~ "Korea",
                            region == "Laos" ~  "Lao PDR",
                            region == "Vietnam" ~ "Viet Nam",
                            TRUE ~ region))
```

[ChatGPT] That is a good approach to change the country names in the map data to match the country names used in the World Inequality Report 2022. Using the mutate() function in combination with the case_when() function allows you to change specific country names to match the data from the report. In the code you provided, you are using the region column from the world_map data and the case_when() function to change the country names to match the names used in the World Inequality Report 2022.

It is a good practice to double check the country names after you have changed them, to ensure that they match correctly. Also, make sure that the country names you are using are the latest and official names of the countries.

If you have any other questions or issues, please let me know.

[HS] I don't know if some countries are still missing. What is the best way to check the missing countries?

[ChatGPT] One way to check for missing countries in your data is to compare the country names in your data with a reference list of country names. You can use a list of official country names from a reputable source, such as the United Nations or the World Bank, to check if all the countries in your data are present.

Another way is to use the setdiff() function from the dplyr package to find the difference between the countries in your data and the reference list of countries. This function returns the elements that are in one set but not in another.

For example, you can use the following code to find the countries in your data that are not in the reference list:

```
Copy code
missing_countries <- setdiff(world_map$region, reference_list$Country)
```

You can also use anti_join() function from dplyr package to see if there are any missing countries

```
Copy code
missing_countries <- anti_join(world_map, reference_list, by = c("region" = "Country"))
```

It's important to note that this approach may not be able to detect all missing countries if the country names are misspelled or if there are other discrepancies in the data.

It's also important to keep in mind that some countries may change their names over time, or merge with other countries. So, it's a good idea to double-check that your country names match the latest data from the World Inequality Report 2022 or the reference list you are using.

Please let me know if you have any other questions or issues.

[HS] Good. Thank you. I will get back to you again later.

[ChatGPT] You're welcome. I'm glad I could help.

Please feel free to reach out to me if you have any further questions or issues. I'll be happy to assist you with your analysis and visualization of the data from the World Inequality Report 2022.

Have a good day! Regenerate response

```
ChatGPT Jan 9 Version. Free Research Preview. Our goal is to make AI systems more natu
```

H.6   WIR2022

[HS]   2022

[ChatGPT]   World Inequality Report 2022   World Inequality Lab   World Inequality Report 2018   ht
tps://wir2018.wid.world/

[HS]   https://wir2022.wid.world/

[ChatGPT] 2021   World Inequality Report 2022   2022

[HS]   R

[ChatGPT]   2022   R

                        World Inequality Report 2022   World Inequality Database   https://wid.world/data/   CSV XLS

R   R ggplot2 lattice plotly

WID.world   R   R

R   World Inequality Report 2022

[HS]   tidyverse

[ChatGPT]   ！R tidyverse   World Inequality Re-
port 2022   ggplot2   dplyr   tidyr

dplyr tidyr

ggplot2

plotly   ggmap   leaflet

[HS]   F3

[ChatGPT]   2022

2022

World Inequality Database Inequality of Household Income   CSV

WID   R

[HS]

[ChatGPT]   World Inequality Database

WID

R   World Inequality Report 2022

[HS]   ggplot2   map_data("world")   WIR

[ChatGPT]   ggplot2   map_data("world")   2022

World Inequality Report 2022   dplyr   rename()

dplyr   left_join()   World Inequality Report   WIR

[HS] rename()'

[ChatGPT] dplyr   rename()

R   levels()

dplyr   mutate()   levels(data*country*)*data*country<-factor(data$country,
levels = levels)

World Inequality Report 2022   stringr

World Inequality Report

[HS] mutate case_when   1

[ChatGPT]   dplyr   mutate() case_when()   World Inequality Re-
port 2022

mutate()   case_when()

mutate()   case_when()   'A' 'B'

```
#

   <-     %>%.
mutate(country = case_when(country == "A" ~ "B",
TRUE ~ country))
```

World Inequality Report 2022

[HS]   left_join

[ChatGPT]   dplyr   left_join()   World   Inequality   Report
2022   1

left_join   World Inequality Report   WIR

World Inequality Report 2022

```
#
map_data_with_inequality_data <- left_join(map_data, World_Inequality_Report_data, by =
```

World Inequality Report 2022

[HS]

```
world_map <- map_data("world")
world_map_wir <- world_map %>%.
mutate(region = case_when(region == "     " ~ "DR  ",
region == "    " ~ "  ",
region == "  " ~ "    ",
region == "South Korea" ~ "  ",
region == "Laos" ~ "Lao PDR",
region == "  " ~ "    ",
TRUE ~    ))
```

[ChatGPT]   2022   mutate() case_when()   world
Inequality Report 2022

[HS]

[ChatGPT]   1

    dplyr   setdiff()

```
#
missing_countries <- setdiff(world_map$region, reference_list$Country)
```

  dplyr   anti_join()

```
#
missing_countries <- anti_join(world_map, reference_list, by = c("region" = "Country")))
```

                                              2022

[HS]

[ChatGPT]

                    World Inequality Report 2022

H.7   Miscellanous

HS] What is the default value of the span in geom_smooth?

[ChatGPT] The default value of the span parameter in the geom_smooth()
function in ggplot2 package of R is 0.75.

The span parameter controls the degree of smoothing for the loess smoother
used to draw the smoothed line. A smaller value of span results in a more
wiggly line that follows the data more closely, while a larger value results in a
smoother line that is less sensitive to individual data points.

If no value is specified for span, then the default value of 0.75 is used. However,
you can adjust the value of span to customize the degree of smoothing in the
plot.

[HS] In R on RStudio under Japanese Window system, changing the repository of CRAN site to Japanese solved package installation problems. Why is it? Does Japanese site provide modified packages?

[ChatGPT] Changing the repository of CRAN site to Japanese may have solved package installation problems because it allows R to access a mirror site that is physically closer to the user and potentially has faster download speeds. This can reduce the likelihood of connection errors and timeouts during package installation.

It is unlikely that the Japanese CRAN mirror provides modified packages, as CRAN itself maintains strict standards for package submissions and updates. The CRAN mirrors simply provide a way for users to access the same set of packages hosted on CRAN, but from a different location.

It is worth noting that using a local CRAN mirror can also improve the availability of certain packages, as some packages may only be available on specific mirrors due to differences in operating systems or hardware architectures. This can be particularly important for users working with specialized software or hardware configurations.

[HS] Is it true that if we backup windows documents folder using OneDrive, then home directory of the user is in OneDrive?

[ChatGPT] No, it is not true that backing up the Windows Documents folder using OneDrive means that the home directory of the user is in OneDrive. OneDrive is a cloud-based file hosting service provided by Microsoft, which can be used to store and sync files across multiple devices.

When you set up OneDrive, you can choose which folders to sync with the cloud, including the Documents folder. This means that any files stored in the Documents folder will be uploaded to OneDrive and can be accessed from other devices where you have signed in to OneDrive.

However, the home directory of the user (often located in "C:\Users") includes more than just the Documents folder. It contains various subfolders such as Desktop, Downloads, Music, Pictures, and Videos, which also contain important user data. By default, only the Documents folder is synced with OneDrive.

So while using OneDrive to backup your Documents folder is a useful way to ensure your important files are stored safely and accessible from anywhere, it does not mean that your entire home directory is stored in OneDrive.

[HS] But when I installed R and R Studio, then by Sys.getenv() I confirmed that the home directory is in One Drive.

[ChatGPT] It's possible that when you installed R and R Studio, the default location for your user's home directory was changed to a OneDrive folder. This could happen if you had previously set up OneDrive to sync your user's home directory.

When you run Sys.getenv() in R or R Studio, it returns a list of environment variables for your current session. One of the variables is "HOME", which specifies the location of your home directory. If your home directory is in a OneDrive folder, the value of this variable will reflect that.

It's worth noting that while you can change the location of your user's home directory, doing so can have unintended consequences and may cause issues with certain applications or system functions. Therefore, it's generally recommended to keep your user's home directory in its default location unless you have a specific reason to move it.

If you're unsure about how to manage your user's home directory or have concerns about its location, it's a good idea to consult with a qualified IT professional or Microsoft support.

[HS] Is there any way, we can change the home directory to be outside of One Drive even if we used it before in OneDrive?

[ChatGPT] Yes, it is possible to change the location of your user's home directory to be outside of OneDrive, even if you have used it before in OneDrive. Here are the general steps:

1. Create a new folder outside of OneDrive where you want your user's home directory to be located.
2. Copy the contents of your current user's home directory from OneDrive to the new folder you created.
3. Open the registry editor by pressing the Windows key + R, typing "regedit" and pressing Enter.
4. Navigate to the following registry key: HKEY_LOCAL_MACHINE\SOFTWARE\Microsoft\Windows NT\CurrentVersion\ProfileList
5. Find the key for your user account, which will be named with a long string of characters ending in ".bak". Note that the ".bak" extension will only be present if there is already a profile folder for your account in the default location.
6. Right-click on the key for your user account and select "Rename". Remove the ".bak" extension from the end of the key name.
7. Double-click on the "ProfileImagePath" value in the right pane and change the path to the new location of your user's home directory.
8. Click "OK" to save the changes.
9. Restart your computer for the changes to take effect.

After completing these steps, your user's home directory will be located outside of OneDrive. However, be aware that changing the location of your user's home directory can cause issues with certain applications or system functions. Therefore, it's generally recommended to keep your user's home directory in its default location unless you have a specific reason to move it.

[HS] You stopped without finishing your message.

[ChatGPT] I apologize for the inconvenience. Here's the complete message:

If you have any important files in your old OneDrive-based home directory, make sure to copy them to the new directory location before making the change. It's also a good idea to create a backup of your system before making any significant changes.

Keep in mind that changing the location of your user's home directory may cause issues with certain applications or system functions. Therefore, it's generally recommended to keep your user's home directory in its default location unless you have a specific reason to move it. If you're unsure about how to manage your user's home directory or have concerns about its location, it's a good idea to consult with a qualified IT professional or Microsoft support.

H.7.1

[HS] geom_smooth span

[ChatGPT] R ggplot2   geom_smooth()   span   0.75

span   Loess   span

span   0.75   span

[HS]   Window   RStudio   R   CRAN

[ChatGPT] CRAN

CRAN   CRAN   CRAN   CRAN

                                CRAN

[HS] OneDrive   Windows   OneDrive

[ChatGPT]   OneDrive   Windows   OneDrive   OneDrive Microsoft

OneDrive   Documents   Documents
OneDrive   OneDrive

                    C: \Users   Documents   Desktop Downloads Music Pictures Videos

    OneDrive   Documents
OneDrive

[HS]   R R Studio   Sys.getenv()   One Drive

[ChatGPT] R R Studio   OneDrive   OneDrive

R   R Studio Sys.getenv()   1 HOME   OneDrive

                                IT   Microsoft

[HS] OneDrive   OneDrive

[ChatGPT]   OneDrive   OneDrive

1. OneDrive

2.   OneDrive

3. Windows +R   regedit   Enter

4.   HKEY_LOCAL_MACHINE

5.   ".bak"   ".bak "

6.   ".bak "

7.   ProfileImagePath

8. OK

9.

              OneDrive

[HS]

[ChatGPT]

 OneDrive

H.8   Data Science Education

[HS] I am designing data science education or curriculum for all students at universities in Japan. All students include non-science major students.

Chat GPT That's a great initiative! Designing a data science curriculum that can be accessible to all students regardless of their major is a challenging but rewarding task. Here are some key considerations that you may find helpful:

1. Foundation in mathematics and statistics: Students should have a strong foundation in mathematics, especially linear algebra, calculus, and probability theory. They should also have an understanding of descriptive and inferential statistics.
2. Programming skills: Familiarity with at least one programming language, such as Python or R, is essential for data science. Students should be able to write code, manipulate data, and visualize results.
3. Data preprocessing and exploration: Data science starts with the data. Students should learn how to preprocess data to make it suitable for analysis, as well as how to explore the data to gain insights and identify patterns.
4. Machine learning: Machine learning is a crucial aspect of data science. Students should learn the fundamentals of supervised and unsupervised learning, including regression, classification, clustering, and dimensionality reduction.

5. Data visualization: Effective data visualization is important for communicating results and insights to stakeholders. Students should learn how to create various types of charts and plots, and how to choose the appropriate visualization for a given situation.
6. Ethics: Data science has the potential to impact society in significant ways. Students should learn about ethical considerations in data science, such as data privacy, bias, and algorithmic fairness.
7. Real-world applications: To make the curriculum more engaging and relevant, it is important to include real-world examples and projects that illustrate the practical applications of data science.

I hope these suggestions are helpful as you design your data science curriculum for all students in Japan. Good luck!

[HS] Yes, I know. What you wrote may be the standard. But the first thing we should pursue is the students' awareness to get motivated. Starting from setting the mathematical foundation, or computer skills cannot motivate students to be involved in it.

Chat GPT I completely agree. Motivating students to be interested in data science is indeed a key factor in ensuring their success in the field. Here are some ideas that you can use to increase student engagement and interest in data science:

1. Highlight the impact of data science: Emphasize the real-world impact that data science can have on a wide range of fields and industries, such as healthcare, finance, transportation, and social media. Showing students how data science is used to solve problems and make decisions can help them see the relevance and importance of the subject.
2. Use case studies and real-world examples: Provide students with real-world case studies and examples that demonstrate the application of data science in various fields. This can help them understand the practical applications of data science and make the subject more relatable and engaging.
3. Encourage hands-on learning: Data science is a practical field, and hands-on learning is an effective way to engage students and help them retain information. Provide students with opportunities to work on projects and data sets, so they can apply the concepts they have learned and see the results for themselves.
4. Foster a community of learners: Create a supportive and collaborative learning environment where students can work together, share their experiences, and ask questions. Encouraging peer-to-peer learning can help students feel more motivated and engaged in the subject.
5. Introduce real-world challenges: Provide students with real-world challenges or problems that they can solve using data science techniques. This can help them see the practical applications of data science and feel motivated to continue learning.

By highlighting the impact of data science, using real-world examples, and encouraging hands-on learning, you can create an engaging and relevant curriculum that will help students develop a passion for data science.

[HS] Thank you. I want students for the next generation to be prepared for the AI world. However, the first thing students prepare is to get used to thinking based on data. That includes challenging their own bias and developing the skill to think using the data.

Chat GPT I understand your goal! Developing critical thinking skills and a data-driven mindset is indeed an important aspect of data science education. Here are some strategies you can use to achieve this goal:

1. Teach data literacy: Emphasize the importance of data literacy, which includes understanding data sources, data types, and how to clean, manipulate, and visualize data. Teach students how to critically evaluate the quality and accuracy of data, and how to determine the best methods for analyzing and interpreting the results.
2. Encourage exploration and experimentation: Encourage students to experiment with different data sets, algorithms, and models. By exploring and experimenting, students can develop their critical thinking skills and gain a deeper understanding of the data and its patterns.
3. Discuss ethical considerations: Data science has the potential to impact society in significant ways, and students should learn about ethical considerations in data science, such as data privacy, bias, and algorithmic fairness. Encourage students to think critically about the ethical implications of their work and to consider the potential impact of their findings on society.
4. Foster a data-driven mindset: Encourage students to adopt a data-driven mindset and to rely on data and evidence to inform their decisions. By developing this mindset, students can learn to challenge their own biases and think critically about the information they are presented with.
5. Provide real-world examples: Provide students with real-world examples and case studies that illustrate the application of data science in various fields. This can help them understand the practical implications of data science and encourage them to think critically about the role that data plays in decision-making.

By incorporating these strategies into your curriculum, you can help students develop critical thinking skills and a data-driven mindset that will serve them well in the AI world.

[HS] When we discuss data science, we tend to think about big data. But another important aspect is that now lots of institutions are providing data as free open public data of the world. So getting skills to analyze such open public data may prepare the next generation to work in the future.

Chat GPT Yes, that's a great point. Working with open public data is an important aspect of data science, and it can provide students with valuable

skills and experience that will be useful in the future. Here are some strategies you can use to incorporate open public data into your curriculum:

1. Highlight the importance of open public data: Explain to students the significance of open public data and how it can be used to address real-world problems. Emphasize the role that data science can play in making sense of large, complex data sets and in providing insights that can inform decision-making.
2. Teach data acquisition and cleaning skills: Teach students how to find, access, and clean open public data. This will help them develop important data acquisition and cleaning skills that are critical to working with large data sets.
3. Provide opportunities to work with open public data: Provide students with opportunities to work with real-world open public data sets and to apply data science techniques to extract insights and make predictions. This will give them hands-on experience with real-world data and help them develop a deeper understanding of the process of data analysis.
4. Encourage collaboration and data sharing: Encourage students to collaborate with one another and to share their findings with the broader data science community. This will help them develop a sense of community and promote the sharing of knowledge and best practices.
5. Promote the use of open source tools: Encourage the use of open source data analysis tools and software, such as R and Python, which are widely used in the data science community. By using these tools, students will have access to a wide range of resources and support and will be better prepared to work in a real-world data science setting.

By incorporating open public data into your curriculum and emphasizing the importance of data acquisition, cleaning, and sharing, you can help students develop the skills they need to work with large data sets and prepare them for the future of data science.

[HS] Yes. I am thinking of Data from World Bank such as World Development Indicators, OECD data, UN data, Eurostat, …. What is your opinion?

Chat GPT That's a great idea! Using data from World Bank, OECD, UN, and Eurostat is a fantastic way to provide students with exposure to real-world data and to help them develop important data analysis skills. These organizations collect and maintain large, high-quality data sets on a variety of topics, including economic indicators, health, education, and environmental sustainability, among others.

By using this data, students can work on meaningful projects and gain a deeper understanding of the issues and challenges facing the world today. They can also develop the skills needed to analyze and interpret large, complex data sets and to communicate their findings effectively.

Additionally, using data from these organizations can help students develop a global perspective and an understanding of the interconnectedness of different

issues and challenges. This is particularly important in the context of data science, as data can provide insights that can inform policy and decision-making on a global scale.

In conclusion, using data from organizations such as the World Bank, OECD, UN, and Eurostat is an excellent way to provide students with exposure to real-world data and to help them develop the skills they need to work in the field of data science.

[HS] I am thinking of two steps. The first level is to look at the data using the dashboards these institutions are providing. The second level is to use R packages to import data and do basic analysis by themselves.

Chat GPT That's a great approach! By starting with the dashboards provided by these institutions, you can give students a high-level overview of the data and help them get a sense of what's available. This can provide a foundation for more in-depth analysis and exploration in the second step, where students can import data using R packages and do basic analysis on their own.

Here's a rough outline of what the two steps might look like:

Step 1: Dashboard exploration

- Introduce students to the dashboards provided by organizations such as World Bank, OECD, UN, and Eurostat
- Encourage students to explore the dashboards and to get a sense of the different types of data that are available
- Have students identify specific data sets that they would like to work with

Step 2: Basic data analysis using R

- Teach students how to import data into R using packages such as tidyverse, rwbdriv, and eurostat
- Teach students basic data analysis techniques, such as data cleaning, visualization, and basic statistical analysis
- Encourage students to apply what they've learned to the data sets they identified in step 1 and to do some basic analysis on their own
- Have students present their findings and discuss any insights they gained from their analysis

By following this two-step approach, you can provide students with a comprehensive introduction to data analysis and help them develop the skills they need to work with real-world data. Additionally, by focusing on open public data, you can help students develop a sense of the importance of data and how it can be used to inform policy and decision-making on a global scale.

[HS] I am familiar with tidyverse and the WDI package and some more. But let us start discussing the first level, using the dashboard, each institute is providing.

Chat GPT Understood! The first level, using the dashboards provided by institutions such as World Bank, OECD, UN, and Eurostat, is a great way to give students a high-level overview of the data and to help them get a sense of what's available. Here are a few suggestions for incorporating this into your curriculum:

1. Start with an introduction to the institutions: Provide a brief overview of the World Bank, OECD, UN, and Eurostat, and explain what types of data they collect and maintain.
2. Show students the dashboards: Walk students through the dashboards provided by each institution and help them understand how to navigate and use them. Encourage students to explore the dashboards on their own and to get a sense of the different types of data that are available.
3. Identify specific data sets: Have students identify specific data sets that they would like to work with. This could be done in a group setting, where students discuss the different data sets they've explored and decide on one to focus on.
4. Discuss the data: Once students have identified specific data sets, have them discuss what they've learned from exploring the data. Encourage students to ask questions and to discuss any insights or patterns they've noticed.
5. Use the data for discussion: Use the data students have explored as a starting point for discussions and debates about the world and the issues it faces. Encourage students to consider how the data can be used to inform policy and decision-making.

By following these steps, you can give students a high-level overview of the data that's available and help them develop an understanding of the types of data that institutions such as the World Bank, OECD, UN, and Eurostat collect and maintain. Additionally, by using the data as a starting point for discussion, you can help students develop critical thinking skills and encourage them to consider how data can be used to inform policy and decision-making.

[HS] Yes, nurturing or developing students' critical thinking skills is the key. I want students to create questions by looking at the data of these institutions. However, it is not easy to let all students be motivated to create and ask questions by just looking at the data. Do you have any suggestions?

Chat GPT Sure! Encouraging students to develop their critical thinking skills and create questions by looking at data can be a challenge, but there are a few strategies you can use to help motivate them:

1. Use real-world examples: Connect the data students are exploring to real-world issues and problems that they may be interested in. This can help to make the data more relevant and engaging for students.
2. Make it interactive: Encourage students to work in small groups and to discuss the data and their questions with one another. This can help to foster a sense of community and to keep students motivated and engaged.

3. Provide opportunities for student-led discovery: Give students the opportunity to take the lead and to guide the discussion. Encourage students to share their questions and insights with the class, and to discuss any patterns or trends they've noticed.
4. Encourage creativity: Encourage students to think creatively and to come up with unique and innovative questions. This can help to foster a sense of ownership and to keep students motivated to continue exploring the data.
5. Offer incentives: Consider offering incentives, such as extra credit or recognition, to students who come up with particularly insightful questions or who make notable contributions to the discussion.

By using these strategies, you can help to motivate students to create and ask questions based on the data they're exploring. Additionally, by encouraging students to take an active role in the discovery process, you can help to foster their critical thinking skills and to give them a deeper understanding of the data and the issues it represents.

[HS] Yes, small group work. I am thinking of the same thing. But what do you think of the role of instructors? In the beginning, there may be many faculty members cooperative with the project, but these instructors are not trained in this direction, and they may not have rich experiences. Since I want all students to get involved in this activity or take such a course, we need a lot of instructors. Do you have any suggestions to instructors or instructor's role?

Chat GPT As an instructor in a data science course, there are several key roles you can play to support students as they work in small groups to explore data:

1. Facilitator: Your role as a facilitator is to help students work together effectively and to ensure that the discussion stays focused and on track. You can help to keep the conversation moving forward and to encourage students to ask questions and share their insights.
2. Resource: As an expert in the field, you can provide students with the resources they need to explore the data and to answer their questions. This might include guidance on using the dashboards and data visualization tools, as well as advice on how to interpret the results of their analyses.
3. Coach: Your role as a coach is to help students develop their critical thinking skills and to encourage them to think deeply about the data. This might involve asking probing questions, helping students to articulate their ideas more clearly, and offering guidance on how to analyze the data effectively.
4. Evaluator: Finally, you can play an important role in evaluating students' work and providing feedback that can help them to improve. This might involve grading group assignments, providing individual feedback on students' questions and insights, and offering guidance on how to refine their analysis techniques.

By taking on these various roles, you can help to ensure that students get the

most out of the course and are able to develop their critical thinking skills and their understanding of data science. Additionally, by working closely with students, you can help to foster a sense of community and to create an environment that is supportive and engaging.

[HS] Very good. Instructors at universities have a background in research. However, the field is narrow. They tend to stay in their own field. But students are not. Do you have any good idea to coach these instructors?

Chat GPT Yes, coaching instructors in data science can be a challenge, especially if they have limited prior experience in the field. However, there are several strategies that you can use to help instructors get up to speed:

1. Provide training and resources: Offer instructors training and resources that can help them to develop their knowledge of data science and data visualization. This might include workshops, online courses, and tutorials, as well as access to books and other resources.
2. Encourage collaboration: Encourage instructors to collaborate with each other and with experts in the field. This can help to broaden their understanding of data science and to provide them with a range of perspectives on how to teach the subject.
3. Emphasize the importance of hands-on learning: Encourage instructors to emphasize the importance of hands-on learning in their courses. This might involve incorporating interactive exercises and projects into their lessons, as well as encouraging students to work with real-world data.
4. Offer support and feedback: Finally, offer ongoing support and feedback to instructors as they develop their skills and incorporate data science into their courses. This might involve providing guidance on how to incorporate new materials into their lessons, as well as feedback on their teaching style and the effectiveness of their lessons.

By taking these steps, you can help instructors to feel more confident in their ability to teach data science, and to develop a deeper understanding of the subject and how to engage students with it. Additionally, by working closely with instructors, you can help to ensure that the course is well-structured, effective, and engaging for students.

[HS] Collaboration among instructors is a challenging topic. But if students get motivated, they may stimulate instructors. So the key is to provide students with good resources. Do you have any suggestions for such resources to start with?

Chat GPT There are many excellent resources available for students to get started with learning data science. Here are some suggestions to consider:

1. Online courses: Websites like Coursera, Udemy, and edX offer a range of data science courses that can help students to develop their skills. These courses can be a great way to get started with learning data science, and they often include interactive exercises and projects that can help students

to practice what they have learned.

2. Data visualization tools: Tools like Tableau, PowerBI, and D3.js can be a great way for students to start exploring data and to develop their data visualization skills. These tools are often user-friendly and can be used to create interactive dashboards and visualizations that help students to understand patterns and relationships in data.

3. Open data: Encourage students to work with open data from sources such as the World Bank, the United Nations, and Eurostat. These datasets can be a great way for students to practice their data analysis skills, and they offer a wealth of information on a range of topics that can be used to generate interesting questions and projects.

4. R packages: R is a powerful programming language that is widely used in data science. Encourage students to start learning R and to explore packages like tidyverse, WDI, and ggplot2 that can be used to perform data analysis and visualization.

5. Books: There are many excellent books available that can help students to develop their skills in data science. Some popular titles include "Data Science from Scratch" by Joel Grus, "R for Data Science" by Hadley Wickham and Garrett Grolemund, and "Data Visualization with ggplot2" by Hadley Wickham.

By providing students with these resources and encouraging them to work with real-world data, you can help to motivate and engage them in learning data science. Additionally, by encouraging students to work in small groups and to collaborate with each other, you can help to create a supportive and engaging learning environment that can help to further develop their skills and understanding of the subject.

[HS] Yes, I looked at data science courses of MOOCs, such as Coursera or edX or so. But these courses are to teach data science skills, and not to motivate students. There are a few good videos but we need something for students to get excited. We need such a curriculum.

Chat GPT In order to motivate students and get them excited about data science, it's important to provide them with a curriculum that is engaging and interactive. Here are some suggestions for how you could design such a curriculum:

1. Problem-based learning: Start with real-world problems that students can relate to, and that they can solve using data science techniques. This could include problems like understanding the relationship between poverty and education, predicting traffic patterns in a city, or analyzing trends in energy consumption.

2. Interactive visualizations: Encourage students to use data visualization tools like Tableau, PowerBI, or D3.js to explore and understand data. These tools can be used to create interactive dashboards and visualizations that help students to see patterns and relationships in data.

3. Group projects: Encourage students to work in small groups on data

science projects. This can help to build a supportive and collaborative learning environment, and can give students the opportunity to learn from each other and to apply what they have learned in a practical setting.
4. Guest speakers: Invite experts in the field of data science to speak to your students. This can help to bring the subject to life and to give students a better understanding of what a career in data science might look like.
5. Hands-on workshops: Provide students with hands-on workshops where they can apply what they have learned in the classroom. For example, you could run a workshop on using R for data analysis, or on using Tableau for data visualization.
6. Student-led discussions: Encourage students to lead discussions and to share their work with the class. This can help to build a supportive and collaborative learning environment, and can give students the opportunity to share their knowledge and ideas with each other.

By incorporating these elements into your curriculum, you can help to create an engaging and interactive learning environment that will motivate and inspire your students to learn more about data science.

[HS] Thank you very much. I will review your responses and get back to you again.

Chat GPT You're welcome! I'm here to help, so please feel free to reach out to me if you have any further questions or concerns. Good luck with your data science education curriculum!

H.8.1

[HS]

Chat GPT

1.

2.   Python R   1

3.

4.

5.

6.

7.

[HS]

Chat GPT

1.

2.

3.

4.

5.

[HS] AI

Chat GPT !

1.

2.

3.

4.

5.

AI

[HS]

Chat GPT

1.

2.

3.

4.

5. R Python

[HS] OECD ……

Chat GPT OECD Eurostat

OECD Eurostat

[HS] 2 2 R

Chat GPT R 2

2

1

• OECD Eurostat

-
-

2: R

- tidyverse rwbdriv eurostat   R
-
-      1
-

 2

[HS]   tidyverse WDI

Chat GPT   OECD   Eurostat

   1.   OECD   Eurostat

   2.

   3.

   4.

   5.

                        OECD   Eurostat

[HS]

Chat GPT

   1.

   2.

   3.

   4.

   5.   5.   :

[HS]

Chat GPT

   1.   :

   2.

   3.

   4.

[HS]

Chat GPT

   1.

   2.

   3.

   4.

Chat GPT

[HS]

Chat GPT

   1.   Coursera Udemy edX   Web

   2.   Tableau PowerBI D3.js

   3.   Eurostat

   4. R   R   R   tidyverse WDI ggplot2

   5.   Joel Grus Data Science from Scratch Hadley Wickham Garrett Grolemund R for Data Science Hadley Wickham Data Visualization with ggplot2

[HS]   Coursera edX   MOOCs

Chat GPT

   1.

   2.   Tableau PowerBI D3.js

   3.

   4.

   5.   R   Tableau

   6.

[HS]

Chat GPT

Appendix I

Bookdown

I.1 About

This is a *sample* book written in **Markdown**. You can use anything that Pandoc's Markdown supports; for example, a math equation $a^2 + b^2 = c^2$.

I.1.1 Usage

Each **bookdown** chapter is an .Rmd file, and each .Rmd file can contain one (and only one) chapter. A chapter *must* start with a first-level heading: `# A good chapter`, and can contain one (and only one) first-level heading.

Use second-level and higher headings within chapters like: `## A short section` or `### An even shorter section`.

The `index.Rmd` file is required, and is also your first book chapter. It will be the homepage when you render the book.

I.1.2 Render book

You can render the HTML version of this example book without changing anything:

1. Find the **Build** pane in the RStudio IDE, and

2. Click on **Build Book**, then select your output format, or select "All formats" if you'd like to use multiple formats from the same book source files.

Or build the book from the R console:

```
bookdown::render_book()
```

To render this example to PDF as a `bookdown::pdf_book`, you'll need to install XeLaTeX. You are recommended to install TinyTeX (which includes XeLaTeX): https://yihui.org/tinytex/.

I.1.3   Preview book

As you work, you may start a local server to live preview this HTML book. This preview will update as you edit the book when you save individual .Rmd files. You can start the server in a work session by using the RStudio add-in "Preview book", or from the R console:

```
bookdown::serve_book()
```

I.2   Hello bookdown

All chapters start with a first-level heading followed by your chapter title, like the line above. There should be only one first-level heading (`#`) per .Rmd file.

I.2.1   A section

All chapter sections start with a second-level (`##`) or higher heading followed by your section title, like the sections above and below here. You can have as many as you want within a chapter.

An unnumbered section

Chapters and sections are numbered by default. To un-number a heading, add a `{.unnumbered}` or the shorter `{-}` at the end of the heading, like in this section.

I.3   Cross-references

Cross-references make it easier for your readers to find and link to elements in your book.

I.3.1   Chapters and sub-chapters

There are two steps to cross-reference any heading:

1. Label the heading: `# Hello world {#nice-label}`.
   - Leave the label off if you like the automated heading generated based on your heading title: for example, `# Hello world` = `# Hello world {#hello-world}`.
   - To label an un-numbered heading, use: `# Hello world {-#nice-label}` or `{# Hello world .unnumbered}`.
2. Next, reference the labeled heading anywhere in the text using `\@ref(nice-label)`; for example, please see Chapter I.3.

- If you prefer text as the link instead of a numbered reference use: any text you want can go here.

I.3.2 Captioned figures and tables

Figures and tables *with captions* can also be cross-referenced from elsewhere in your book using `\@ref(fig:chunk-label)` and `\@ref(tab:chunk-label)`, respectively.

See Figure I.1.

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```
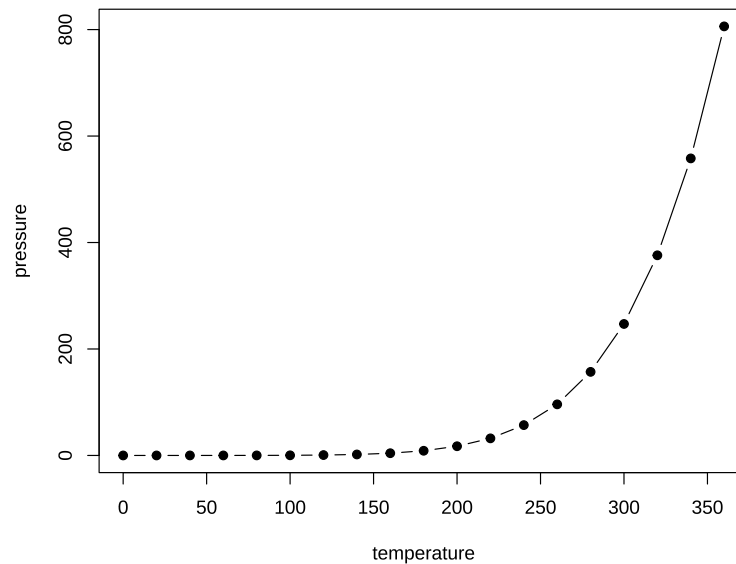

Figure I.1: Here is a nice figure!

Don't miss Table I.1.

```
knitr::kable(
  head(pressure, 10), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

I.4 Parts

You can add parts to organize one or more book chapters together. Parts can be inserted at the top of an .Rmd file, before the first-level chapter heading in that same file.

Table I.1: Here is a nice table!

temperature	pressure
0	0.0002
20	0.0012
40	0.0060
60	0.0300
80	0.0900
100	0.2700
120	0.7500
140	1.8500
160	4.2000
180	8.8000

Add a numbered part: `# (PART) Act one {-}` (followed by `# A chapter`)

Add an unnumbered part: `# (PART\*) Act one {-}` (followed by `# A chapter`)

Add an appendix as a special kind of un-numbered part: `# (APPENDIX) Other stuff {-}` (followed by `# A chapter`). Chapters in an appendix are prepended with letters instead of numbers.

I.5   Footnotes and citations

I.5.1   Footnotes

Footnotes are put inside the square brackets after a caret `^[]`. Like this one [1].

I.5.2   Citations

Reference items in your bibliography file(s) using `@key`.

For example, we are using the **bookdown** package (Xie, 2023) (check out the last code chunk in index.Rmd to see how this citation key was added) in this sample book, which was built on top of R Markdown and **knitr** (Xie, 2015) (this citation was added manually in an external file book.bib). Note that the `.bib` files need to be listed in the index.Rmd with the YAML `bibliography` key.

The `bs4_book` theme makes footnotes appear inline when you click on them. In this example book, we added `csl: chicago-fullnote-bibliography.csl` to the `index.Rmd` YAML, and include the `.csl` file. To download a new style, we recommend: https://www.zotero.org/styles/

---

[1]This is a footnote.

The RStudio Visual Markdown Editor can also make it easier to insert citations: https://rstudio.github.io/visual-markdown-editing/#/citations

I.6 Blocks

I.6.1 Equations

Here is an equation.

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{I.1}$$

You may refer to using `\@ref(eq:binom)`, like see Equation (I.1).

I.6.2 Theorems and proofs

Labeled theorems can be referenced in text using `\@ref(thm:tri)`, for example, check out this smart theorem I.1.

Theorem I.1. *For a right triangle, if c denotes the* length *of the hypotenuse and a and b denote the lengths of the **other** two sides, we have*

$$a^2 + b^2 = c^2$$

Read more here https://bookdown.org/yihui/bookdown/markdown-extensions-by-bookdown.html.

I.6.3 Callout blocks

The `bs4_book` theme also includes special callout blocks, like this `.rmdnote`.

You can use **markdown** inside a block.

```
head(beaver1, n = 5)
#>   day time  temp activ
#> 1 346  840 36.33     0
#> 2 346  850 36.34     0
#> 3 346  900 36.35     0
#> 4 346  910 36.42     0
#> 5 346  920 36.55     0
```

It is up to the user to define the appearance of these blocks for LaTeX output.

You may also use: `.rmdcaution`, `.rmdimportant`, `.rmdtip`, or `.rmdwarning` as the block name.

The R Markdown Cookbook provides more help on how to use custom blocks to design your own callouts: https://bookdown.org/yihui/rmarkdown-cookbook/custom-blocks.html

I.7   Sharing your book

I.7.1   Publishing

HTML books can be published online, see: https://bookdown.org/yihui/bookdown/publishing.html

I.7.2   404 pages

By default, users will be directed to a 404 page if they try to access a webpage that cannot be found. If you'd like to customize your 404 page instead of using the default, you may add either a `_404.Rmd` or `_404.md` file to your project root and use code and/or Markdown syntax.

I.7.3   Metadata for sharing

Bookdown HTML books will provide HTML metadata for social sharing on platforms like Twitter, Facebook, and LinkedIn, using information you provide in the `index.Rmd` YAML. To setup, set the `url` for your book and the path to your `cover-image` file. Your book's `title` and `description` are also used.

This `bs4_book` provides enhanced metadata for social sharing, so that each chapter shared will have a unique description, auto-generated based on the content.

Specify your book's source repository on GitHub as the `repo` in the `_output.yml` file, which allows users to view each chapter's source file or suggest an edit. Read more about the features of this output format here:

https://pkgs.rstudio.com/bookdown/reference/bs4_book.html

Or use:

```
?bookdown::bs4_book
```

Bibliography

Xie, Y. (2015). *Dynamic Documents with R and knitr.* Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition. ISBN 978-1498716963.

Xie, Y. (2023). *bookdown: Authoring Books and Technical Documents with R Markdown.* R package version 0.32.