Assignment 2

Deconstruct, Reconstruct Web Report

Priya Ningappa Madabal (s3959738)

Assessment declaration checklist

Please carefully read the statements below and check each box if you agree with the declaration. If you do not check all boxes, your assignment will not be marked. If you make a false declaration on any of these points, you may be investigated for academic misconduct. Students found to have breached academic integrity may receive official warnings and/or serious academic penalties. Please read more about academic integrity here (https://www.rmit.edu.au/students/student-essentials/assessment-and-exams/academic-integrity). If you are unsure about any of these points or feel your assessment might breach academic integrity, please contact your course coordinator for support. It is important that you DO NOT submit any assessment until you can complete the declaration truthfully.

By checking the boxes below, I declare the following:

- ☑ I have not impersonated, or allowed myself to be impersonated by, any person for the purposes of this assessment
- ☑ This assessment is my original work and no part of it has been copied from any other source except where due acknowledgement is made.
- ☑ No part of this assessment has been written for me by any other person except where such collaboration has been authorised by the lecturer/teacher concerned.
- ☑ Where this work is being submitted for individual assessment, I declare that it is my original work and that no part has been contributed by, produced by or in conjunction with another student.
- ☑ I give permission for my assessment response to be reproduced, communicated compared and archived for the purposes of detecting plagiarism.
- ☑ I give permission for a copy of my assessment to be retained by the university for review and comparison, including review by external examiners.

I understand that:

- ☑ Plagiarism is the presentation of the work, idea or creation of another person as though it is your own. It is a form of cheating and is a very serious academic offence that may lead to exclusion from the University. Plagiarised material can be drawn from, and presented in, written, graphic and visual form, including electronic data and oral presentations. Plagiarism occurs when the origin of the material used is not appropriately cited.
- ☑ Plagiarism includes the act of assisting or allowing another person to plagiarise or to copy my work.

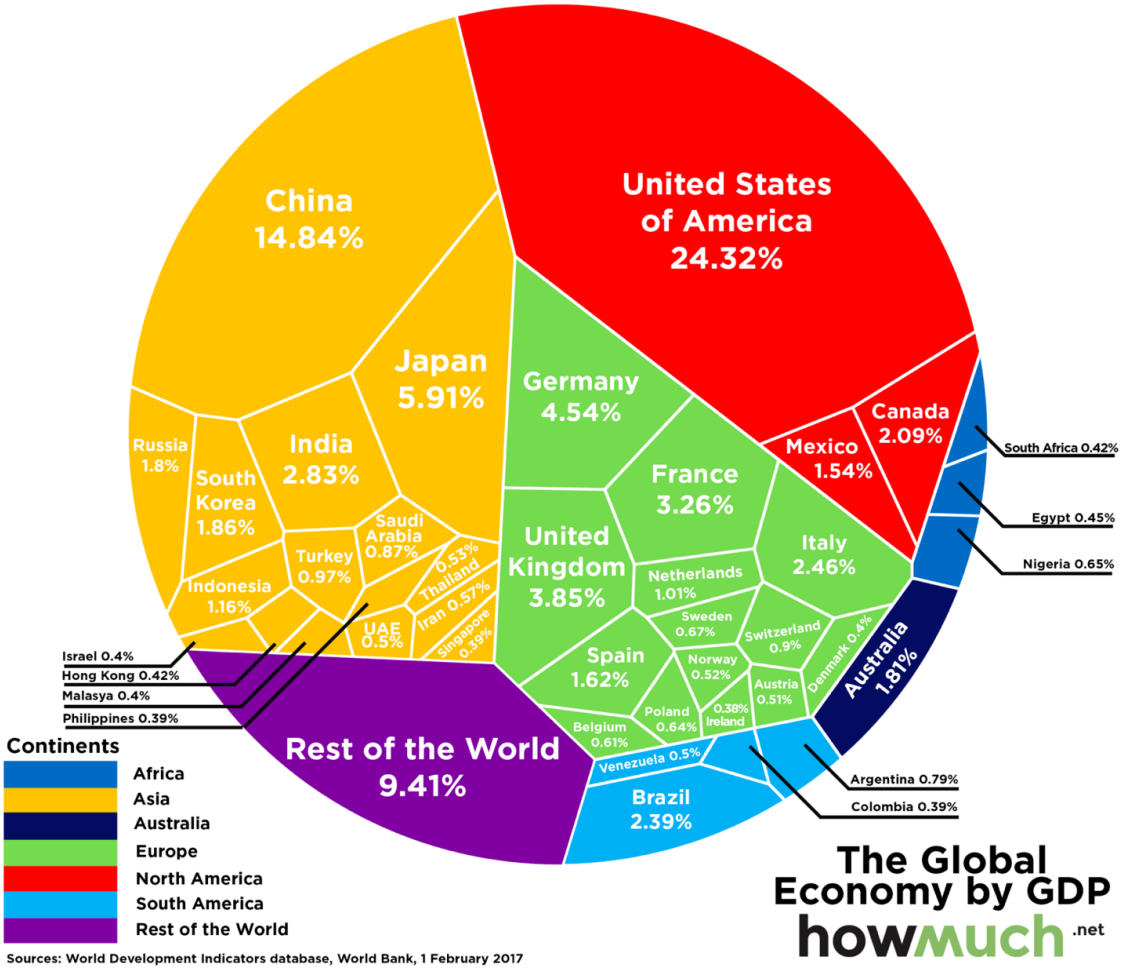
I agree and acknowledge that:

- ☑ I have read and understood the Declaration and Statement of Authorship above.
- ☑ If I do not agree to the Declaration and Statement of Authorship in this context and all boxes are not checked, the assessment outcome is not valid for assessment purposes and will not be included in my final result for this course.

Deconstruct

Original

The original data visualisation selected for the assignment was as follows:



https://foreignpolicy.com/2017/02/24/infographic-heres-how-the-global-gdp-is-divvied-up/ (https://foreignpolicy.com/2017/02/24/infographic-heres-how-the-global-gdp-is-divvied-up/)

The webpage provides information on a breakdown of Global GDP(US million dollars) Distribution for the year 2017. According to recent data released by the World Bank, the global economy is estimated to be around $74.1 trillion. The United States remains the dominant player, contributing nearly a quarter of the world's GDP. Contrary to predictions of China surpassing the U.S., China's share of the global economy is at 14.84%, while the U.S. leads with 24.32%. Japan, once a major economic force, has seen its share decline to 5.91%.

An infographic created by HowMuch.net based on World Bank data from 2015 reveals that the Asian economic bloc represents the largest portion of the global economy at 33.84%, followed by North America at 27.95%, and Europe at 21.37%. Despite the attention on emerging economies, India, Brazil, and Nigeria contribute relatively smaller percentages of 2.83%, 2.39%, and 0.65%, respectively. Notably, the combined contribution of 155 other countries in the “rest of the world” category is roughly equivalent to the gap between the U.S. and China’s shares of the global GDP.

The dataset used for this visualization was obtained from The World Bank.

Source: The World Bank

Link: https://databank.worldbank.org/source/world-development-indicators# (https://databank.worldbank.org/source/world-development-indicators#)

In order to obtain the specific data needed for the visualization, I applied the following filters:

1. Selected “Countries” as the category to filter the dataset by countries.
2. Chose “GDP(current US$)” as the series of interest, focusing on Gross Domestic Product in current US dollars.
3. Narrowed down the data to the year 2017.

By applying these filters, I obtained the relevant dataset in CSV format, which was then used to create the visualization.

Objective and Audience

The objective and audience of the original data visualisation chosen can be summarised as follows:

Objective

The primary objective of the original data visualization was to:

Depict Global GDP Distribution: The visualization aimed to present a clear and easily understandable representation of how the global Gross Domestic Product (GDP) is distributed among various countries or regions.

Explain Economic Disparities: It likely sought to provide insights into the economic disparities and inequalities that exist on a global scale. By visualizing this data, it aimed to communicate how different countries contribute to and benefit from the global economy.

Audience

The intended audience for this data visualization includes:

- General Public: The visualization was designed to be accessible to a broad audience, including individuals who may not have a deep background in economics. It aimed to convey complex economic data in a simple and engaging manner.
- Students and Educators: It could serve as an educational tool for students studying economics, geography, or global affairs. Teachers and educators might use it to illustrate economic concepts in the classroom.

- Policymakers: Policymakers and government officials could find value in this visualization to understand the economic relationships between countries and make informed decisions related to trade, international relations, and economic development.
- Economists and Researchers: Professionals in the field of economics, researchers, and analysts may use this visualization to study global economic trends, trade flows, and economic disparities in more detail.
- Global Affairs Enthusiasts: Individuals interested in global affairs, geopolitics, and international relations may use this visualization to gain insights into the economic aspects of different countries' roles in the world.

In summary, the visualization aimed to make complex economic data accessible and understandable to a wide range of individuals, including the general public, students, policymakers, and professionals in the field of economics and global affairs. It served as a tool for raising awareness and fostering discussions about global economic disparities and inter dependencies.

Critique

The visualisation chosen had the following three main issues:

1. Lack of Proper Labeling: One of the issues with the original data visualization is the absence of proper labeling like title, which can hinder the audience's ability to interpret and understand the information being presented. The issue of countries and their respective percentages of GDP appearing cluttered and unclear in the visualization can significantly affect its effectiveness. In a visualization that involves comparing multiple entities, such as countries, with their corresponding percentages of GDP, it's crucial to ensure that the labels (country names and percentages) are legible and do not overlap. Overlapping labels can make it difficult for viewers to identify and differentiate between the data points.
2. Color Choice and Accessibility: The original data visualization employs a color scheme that may not be optimal for accessibility. The issue of color choice and accessibility is critical in data visualizations to ensure that information is conveyed accurately and inclusively. By selecting color schemes that consider the needs of individuals with color vision deficiencies and testing for accessibility, data visualizations can reach a broader audience and provide a more effective means of communication.
3. Pie Chart Issue for a Large Number of Countries: Using a pie chart to represent a large number of countries can be problematic due to inherent limitations of pie charts. A pie chart typically consists of slices, each representing a portion of the whole (in this case, the GDP percentage of a country). When there are numerous countries to represent, each country becomes a slice, resulting in a cluttered and complex chart.Pie charts are less effective for comparing values across a large number of categories. Viewers may find it challenging to gauge the differences in GDP percentages among numerous countries because the angles and areas of slices are harder to compare than lengths in a bar chart.

Reconstruct

Code

The following code was used to fix the issues identified in the original.

```
# Load necessary libraries

library(ggplot2)
library(dplyr)
library(knitr)
library(countrycode)
```

```
# Import/Read data.

getwd()
```

```
## [1] "/Users/priya/Desktop/Semester 2/Data Viz./Assignment 2"
```

```
GDP_40 <- read.csv("GDP_40.csv")
```

```
# We are using the installed CountryCode package to create a new column called "Co
ntinent" for all the countries.

# To distinguish between North America and South America for only the top 40 count
ries, employed a function to manually assign countries to their respective contine
nts.

map_to_continent <- function(country_name) {
  if (country_name %in% c("Canada", "United States", "Mexico")) {
    return("North America")
  } else if (country_name %in% c("Venezuela", "Brazil", "Argentina", "Colombia"))
{
    return("South America")
  } else if (country_name == "Australia") {
    return("Australia")
  } else {
    return(countrycode::countrycode(country_name, "country.name", "continent"))
  }
}

# Utilizing the mutate() function to merge the dataset with the continent.

GDP_40_with_continent <- GDP_40 %>%
  mutate(Continent = sapply(Country.Name, map_to_continent))
```

```
# Sorting the dataset and selecting variables for analysis.

# Renaming the column to "Country" and "GDP($)."

colnames(GDP_40_with_continent)[colnames(GDP_40_with_continent) == "Country.Name"]
<- "Country"
colnames(GDP_40_with_continent)[colnames(GDP_40_with_continent) == "X2017..YR2017.
"] <- "GDP($)"

# Selecting only required variables (Country, GDP($), and Continent) for the analy
sis.

Country_GDP_40 <- GDP_40_with_continent[, c(1, 5, 6)]
```

```
# Checking NA values and putting NA=0.

null_check <- is.na(Country_GDP_40$`GDP($)`)

print(null_check)
```

```
##   [1]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE
##  [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [61] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [73] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [85] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##  [97] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [109] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [121] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [133] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [145] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [157] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [169] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [181] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [193] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [205] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [217] FALSE
```

```
# Replace NA values in the "GDP($)" column with 0

Country_GDP_40$`GDP($)` <- ifelse(is.na(Country_GDP_40$`GDP($)`), 0, Country_GDP_4
0$`GDP($)`)

# Arrange the data in descending order based on the GDP.

Country_GDP_40 <- Country_GDP_40 %>%
  arrange(desc(Country_GDP_40$`GDP($)`))
```

```
# Calculate the percentage of GDP for the top 40 countries

top_40_countries <- Country_GDP_40[1:40, ]

total_gdp <- sum(Country_GDP_40$`GDP($)`)

top_40_countries$Percentage_GDP <- (top_40_countries$`GDP($)` / total_gdp)


# Format the Percentage_GDP column to display numeric values with 2 decimal places

top_40_countries$Percentage_GDP <- round(top_40_countries$Percentage_GDP * 100, 4)

# Selecting required variables (Country, Continent and Percentage_GDP)
top_40_countries <- top_40_countries[, c(1, 3, 4)]

# Print the top 40 countries

print(top_40_countries)
```

```
##                 Country     Continent Percentage_GDP
## 1         United States North America        24.1820
## 2                 China          Asia        15.2841
## 3                 Japan          Asia         6.1219
## 4               Germany        Europe         4.5824
## 5        United Kingdom        Europe         3.3317
## 6                 India          Asia         3.2919
## 7                France        Europe         3.2220
## 8                Brazil South America         2.5619
## 9                 Italy        Europe         2.4357
## 10               Canada North America         2.0476
## 11          Korea, Rep.          Asia         2.0162
## 12   Russian Federation        Europe         1.9544
## 13            Australia     Australia         1.6469
## 14                Spain        Europe         1.6305
## 15               Mexico North America         1.4388
## 16            Indonesia          Asia         1.2609
## 17              Turkiye          Asia         1.0665
## 18          Netherlands        Europe         1.0353
## 19         Saudi Arabia          Asia         0.8877
## 20          Switzerland        Europe         0.8631
## 21            Argentina South America         0.7991
## 22               Sweden        Europe         0.6717
## 23               Poland        Europe         0.6514
## 24              Belgium        Europe         0.6242
## 25   Iran, Islamic Rep.          Asia         0.6042
## 26             Thailand          Asia         0.5666
## 27              Austria        Europe         0.5181
## 28               Norway        Europe         0.4988
## 29 United Arab Emirates          Asia         0.4848
## 30         South Africa        Africa         0.4736
## 31              Nigeria        Africa         0.4665
## 32               Israel          Asia         0.4448
## 33            Singapore          Asia         0.4262
## 34 Hong Kong SAR, China          Asia         0.4237
## 35             Pakistan          Asia         0.4211
## 36              Ireland        Europe         0.4176
## 37              Denmark        Europe         0.4123
## 38          Philippines          Asia         0.4078
## 39             Malaysia          Asia         0.3962
## 40             Colombia South America         0.3872
```

```
# Creating a new dataframe for the "Rest of countries" with remaining countries.

rest_countries <- Country_GDP_40[-(1:40), ]

# Calculate the percentage of GDP for the rest of the countries.

rest_countries$Percentage_GDP <- (rest_countries$`GDP($)` / total_gdp)

# Format the Percentage_GDP column for the rest of the countries.

rest_countries$Percentage_GDP <- round(rest_countries$Percentage_GDP * 100, 4)

# Calculating the sum of the "GDP($)" values in the rest_countries.

rest_gdp_sum <- sum(rest_countries$`GDP($)`)

# Calculating percentage of GDP for rest of the countries.

rest_perc <- round((rest_gdp_sum / total_gdp) * 100, 4)

# Creating a new data frame named rest_of_world.

rest_of_world <- data.frame(
  Country = "Rest of the World",
  `GDP($)` = rest_gdp_sum,
  Percentage_GDP = rest_perc,
  Continent = "Rest of the World"
)

#Changing the column name to "GDP($).

colnames(rest_of_world)[2] <- "GDP($)"

#Selecting required variables (Country, Continent and Percentage_GDP)

rest_of_world <- rest_of_world[, c(1, 4, 3)]

# Print the rest of the world data
print(rest_of_world)
```

```
##             Country         Continent Percentage_GDP
## 1 Rest of the World Rest of the World         9.0425
```

```
# Combine the top 40 countries and "Rest of the World" data frames

final_data <- dplyr::bind_rows(top_40_countries, rest_of_world)

print(final_data)
```

```
##                 Country         Continent Percentage_GDP
## 1         United States     North America        24.1820
## 2                 China              Asia        15.2841
## 3                 Japan              Asia         6.1219
## 4               Germany            Europe         4.5824
## 5        United Kingdom            Europe         3.3317
## 6                 India              Asia         3.2919
## 7                France            Europe         3.2220
## 8                Brazil     South America         2.5619
## 9                 Italy            Europe         2.4357
## 10               Canada     North America         2.0476
## 11          Korea, Rep.              Asia         2.0162
## 12   Russian Federation            Europe         1.9544
## 13            Australia         Australia         1.6469
## 14                Spain            Europe         1.6305
## 15               Mexico     North America         1.4388
## 16            Indonesia              Asia         1.2609
## 17              Turkiye              Asia         1.0665
## 18          Netherlands            Europe         1.0353
## 19         Saudi Arabia              Asia         0.8877
## 20          Switzerland            Europe         0.8631
## 21            Argentina     South America         0.7991
## 22               Sweden            Europe         0.6717
## 23               Poland            Europe         0.6514
## 24              Belgium            Europe         0.6242
## 25   Iran, Islamic Rep.              Asia         0.6042
## 26             Thailand              Asia         0.5666
## 27              Austria            Europe         0.5181
## 28               Norway            Europe         0.4988
## 29 United Arab Emirates              Asia         0.4848
## 30         South Africa            Africa         0.4736
## 31              Nigeria            Africa         0.4665
## 32               Israel              Asia         0.4448
## 33            Singapore              Asia         0.4262
## 34 Hong Kong SAR, China              Asia         0.4237
## 35             Pakistan              Asia         0.4211
## 36              Ireland            Europe         0.4176
## 37              Denmark            Europe         0.4123
## 38          Philippines              Asia         0.4078
## 39             Malaysia              Asia         0.3962
## 40             Colombia     South America         0.3872
## 41    Rest of the World Rest of the World         9.0425
```

```
# Horizontal stacked bar chart

plot <- ggplot(final_data, aes(x = reorder(Country, Percentage_GDP), y = Percentag
e_GDP, fill = Continent)) +
  geom_bar(stat = "identity", color = "black")  +
  geom_text(aes(label = paste0(round(Percentage_GDP, 2), "%")), hjust = -0.1, size
= 3, color = "black") +
  theme_minimal() +
  labs(x = "Country", y = "Percentage of Global GDP (US million dollars)", fill =
"Continent") +
  scale_fill_brewer(palette = "Set2") +
  theme(plot.title = element_text(hjust = 0.5, vjust = 0.5, face = "bold"),
    legend.text = element_text(face = "bold", family = "Arial", size = 12, margin
= margin(b = 6)),
    legend.title = element_text(face = "bold", family = "Arial", size = 12),
    legend.key.height = unit(1, "cm"))+
  coord_flip() +
  ggtitle("Top 40 Countries by Percentage of Global GDP") +
  ylim(0, 30)
```

Reconstruction

The following plot fixes the main issues in the original.

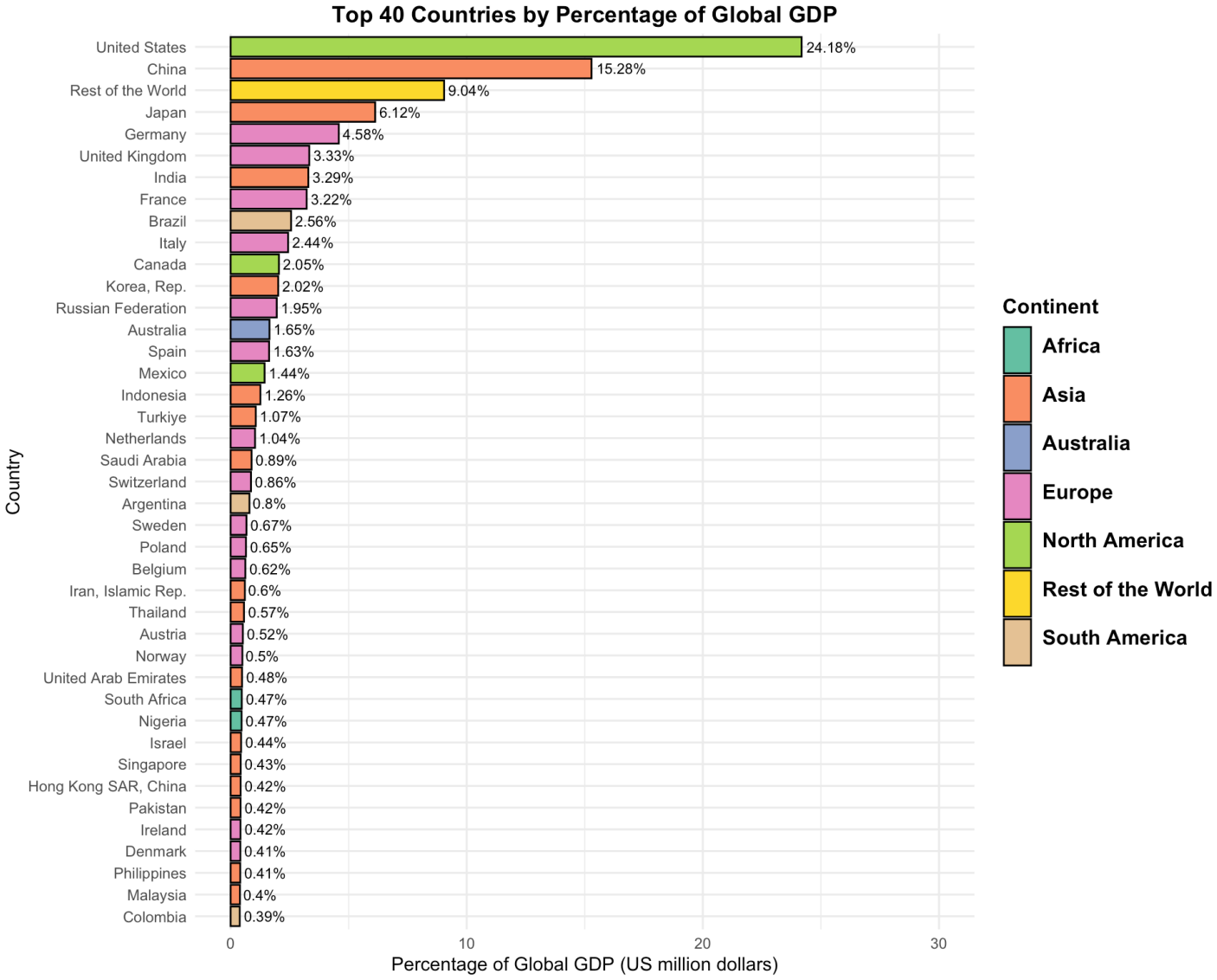


Fig: Top 40 countries GDP(%) in US million dollars

References

The reference to the original data visualisation choose, the data source(s) used for the reconstruction and any other sources used for this assignment are as follows:

- Main Source: Robbie Gramer, “Infographic: Here’s How the Global GDP Is Divvied Up”. Retrieved FEBRUARY 24, 2017 from foreignpolicy website: https://foreignpolicy.com/2017/02/24/infographic-heres-how-the-global-gdp-is-divvied-up/ (https://foreignpolicy.com/2017/02/24/infographic-heres-how-the-global-gdp-is-divvied-up/)
- Dataset source: The World Bank, “The World Bank - World Development Indicators”. Retrieved for the year 2017 from the world bark website: ’https://databank.worldbank.org/source/world-development-indicators# (https://databank.worldbank.org/source/world-development-indicators#)
- Graph references:

1. Prof. James Baglin, RMIT University, Data Visualisation and Communication (2350). Retrieved from Syllabus/Modules website https://rmit.instructure.com/courses/107036/assignments/syllabus (https://rmit.instructure.com/courses/107036/assignments/syllabus)
2. Jory Catalpa, Kyle Zrenchik, Yunxi Yang, University of Minnesota, “How to Make a Stacked Bar Chart in R Using ggplot2”. Retrieved from rstudio-pubs-statistic.s3.amazonaws website: http://rstudio-pubs-static.s3.amazonaws.com/3256_bb10db1440724dac8fa40da5e658ada5.html (http://rstudio-pubs-static.s3.amazonaws.com/3256_bb10db1440724dac8fa40da5e658ada5.html)
3. R*Basics(@rbasics9031), “RBasics: Creating Stacked, Grouped, and Horizontal Bar Charts in R”. Retrieved JUNE 16, 2018, from YouTube channel https://www.youtube.com/watch?v=CfiMyrbyJcc (https://www.youtube.com/watch?v=CfiMyrbyJcc)
4. Nicholas Tierney, “RMarkdown for Scientists - 10. Customising your figures”. Retrieved 09 SEPTEMBER 2020 from rmd4sci.njtierney website https://rmd4sci.njtierney.com/customising-your-figures.html (https://rmd4sci.njtierney.com/customising-your-figures.html)

- Country code package: Vincent Arel-Bundock [aut, cre] (https://orcid.org/0000-0003-2042-7063 (https://orcid.org/0000-0003-2042-7063)), CJ Yetman [ctb] (https://orcid.org/0000-0001-5099-9500 (https://orcid.org/0000-0001-5099-9500)), Nils Enevoldsen [ctb] (https://orcid.org/0000-0001-7195-4117 (https://orcid.org/0000-0001-7195-4117)), Samuel Meichtry [ctb] (https://orcid.org/0000-0003-2165-791X (https://orcid.org/0000-0003-2165-791X)), “Package ’countrycode”. Retrieved MAY 30, 2023, from cran.r-project.org website: https://cran.r-project.org/web/packages/countrycode/countrycode.pdf (https://cran.r-project.org/web/packages/countrycode/countrycode.pdf) ,
- dplyr::bind_rows() function: Zach, “How to Use bind_rows and bind_cols in dplyr (With Examples)”, Retrieved AUGUST 27, 2021, from statology.org website: https://www.statology.org/dplyr-bind_rows-bind_cols/ (https://www.statology.org/dplyr-bind_rows-bind_cols/)