Homework 02: Exploring Data

Brandon Z. Lu

Question 1

```
library(dplyr) # install.packages("dplyr")
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2) # install.packages("ggplot2")
library(scales) # install.packages("scales")
library(tidyr) # install.packages("tidyr")
library(haven) # install.packages("haven")
library(fredr) # Loading the database from https://cran.r-project.org/web/packages/fredr/vignettes/fred
#install.packages("fredr")
fredr_set_key("bdd1cf84197911e0b8c4346fdd943c50")
# usethis::edit_r_environ()
# FRED_API_KEY=bdd1cf84197911e0b8c4346fdd943c50
```

```
# https://fred.stlouisfed.org/series/RKNANPINA666NRUG
capital_stock <- fredr(
  series_id = "RKNANPINA666NRUG",
  observation_start = as.Date("1970-01-01"),
  observation_end = as.Date("2019-12-31")
)

head(capital_stock)
```

```
## # A tibble: 6 x 5
##   date       series_id           value realtime_start realtime_end
##   <date>     <chr>               <dbl> <date>         <date>
## 1 1970-01-01 RKNANPINA666NRUG 2007558. 2025-03-06     2025-03-06
## 2 1971-01-01 RKNANPINA666NRUG 2125584. 2025-03-06     2025-03-06
## 3 1972-01-01 RKNANPINA666NRUG 2246936. 2025-03-06     2025-03-06
## 4 1973-01-01 RKNANPINA666NRUG 2356271. 2025-03-06     2025-03-06
## 5 1974-01-01 RKNANPINA666NRUG 2469048. 2025-03-06     2025-03-06
## 6 1975-01-01 RKNANPINA666NRUG 2572254. 2025-03-06     2025-03-06
```

```
# https://fred.stlouisfed.org/series/EMPENGINA148NRUG
persons_engaged <- fredr(
```

```r
  series_id = "EMPENGINA148NRUG",
  observation_start = as.Date("1970-01-01"),
  observation_end = as.Date("2019-12-31")
)

#https://fred.stlouisfed.org/series/AVHWPEINA065NRUG
avg_hours_worked <- fredr(
  series_id = "AVHWPEINA065NRUG",
  observation_start = as.Date("1970-01-01"),
  observation_end = as.Date("2019-12-31")
)

head(persons_engaged)
```

```
## # A tibble: 6 x 5
##   date       series_id        value realtime_start realtime_end
##   <date>     <chr>            <dbl> <date>         <date>
## 1 1970-01-01 EMPENGINA148NRUG  195. 2025-03-06     2025-03-06
## 2 1971-01-01 EMPENGINA148NRUG  201. 2025-03-06     2025-03-06
## 3 1972-01-01 EMPENGINA148NRUG  208. 2025-03-06     2025-03-06
## 4 1973-01-01 EMPENGINA148NRUG  214. 2025-03-06     2025-03-06
## 5 1974-01-01 EMPENGINA148NRUG  221. 2025-03-06     2025-03-06
## 6 1975-01-01 EMPENGINA148NRUG  227. 2025-03-06     2025-03-06
```

```r
head(avg_hours_worked)
```

```
## # A tibble: 6 x 5
##   date       series_id        value realtime_start realtime_end
##   <date>     <chr>            <dbl> <date>         <date>
## 1 1970-01-01 AVHWPEINA065NRUG 2077. 2025-03-06     2025-03-06
## 2 1971-01-01 AVHWPEINA065NRUG 2077. 2025-03-06     2025-03-06
## 3 1972-01-01 AVHWPEINA065NRUG 2077. 2025-03-06     2025-03-06
## 4 1973-01-01 AVHWPEINA065NRUG 2075. 2025-03-06     2025-03-06
## 5 1974-01-01 AVHWPEINA065NRUG 2072. 2025-03-06     2025-03-06
## 6 1975-01-01 AVHWPEINA065NRUG 2070. 2025-03-06     2025-03-06
```

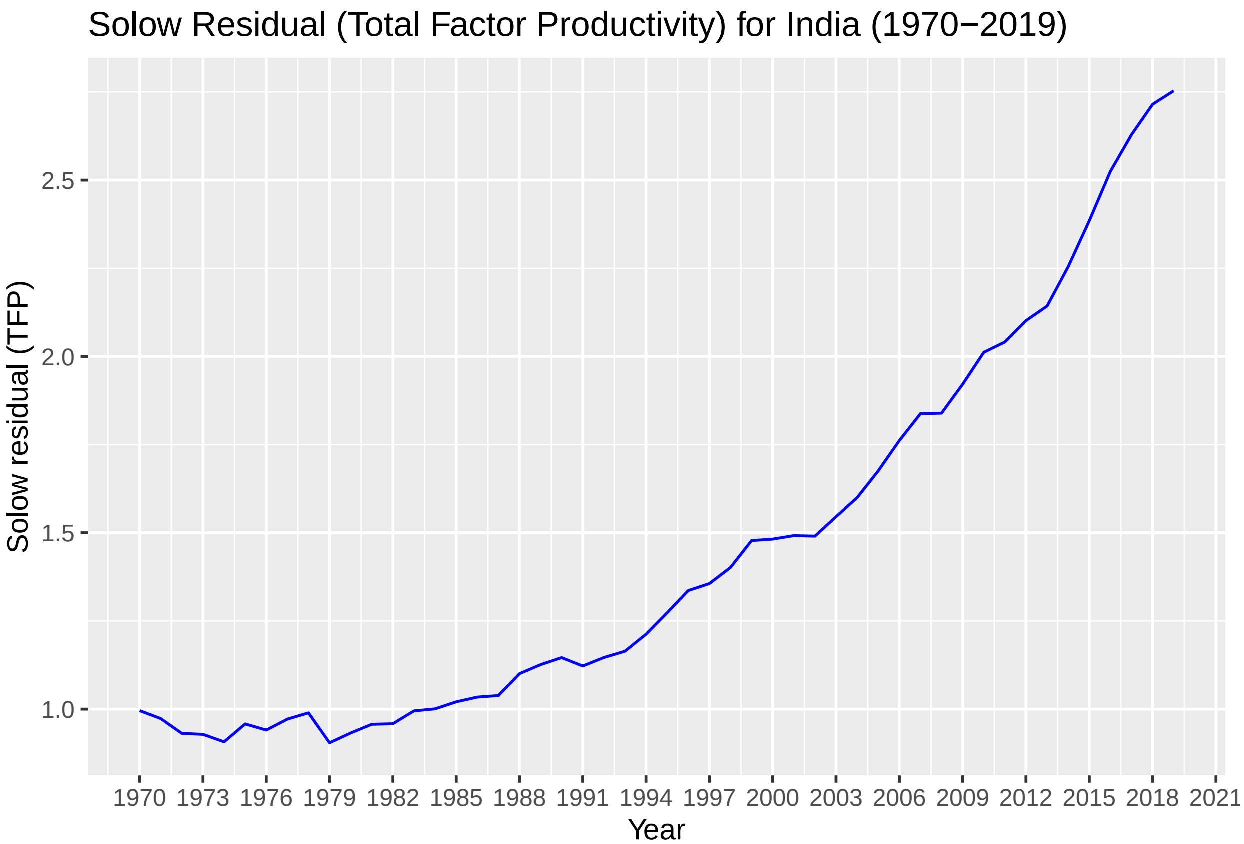
```r
# https://fred.stlouisfed.org/series/RGDPNAINA666NRUG
real_gdp <- fredr(
  series_id = "RGDPNAINA666NRUG",
  observation_start = as.Date("1970-01-01"),
  observation_end = as.Date("2019-12-31")
)

head(real_gdp)
```

```
## # A tibble: 6 x 5
##   date       series_id          value realtime_start realtime_end
##   <date>     <chr>              <dbl> <date>         <date>
## 1 1970-01-01 RGDPNAINA666NRUG 684889. 2025-03-06     2025-03-06
## 2 1971-01-01 RGDPNAINA666NRUG 696141. 2025-03-06     2025-03-06
## 3 1972-01-01 RGDPNAINA666NRUG 692289. 2025-03-06     2025-03-06
## 4 1973-01-01 RGDPNAINA666NRUG 715104. 2025-03-06     2025-03-06
## 5 1974-01-01 RGDPNAINA666NRUG 723580. 2025-03-06     2025-03-06
## 6 1975-01-01 RGDPNAINA666NRUG 789787. 2025-03-06     2025-03-06
```

```r
# Rename 'value' columns for clarity
real_gdp <- real_gdp %>% rename(gdp = value)
capital_stock <- capital_stock %>% rename(capital = value)
persons_engaged <- persons_engaged %>% rename(persons = value)
avg_hours_worked <- avg_hours_worked %>% rename(hours = value)

# Merge datasets
merged_data_india <- real_gdp %>%
  inner_join(capital_stock, by = "date") %>%
  inner_join(persons_engaged, by = "date") %>%
  inner_join(avg_hours_worked, by = "date")

merged_data_india <- merged_data_india %>%
  mutate(labor = persons * hours)

alpha <- 0.33

merged_data_india <- merged_data_india %>%
  mutate(
    tfp = gdp / (capital^alpha * labor^(1 - alpha))
  )
```

Assuming a Cobb-Douglas production function with α= 0.33, find and plot the Solow residual of the Indian economy. Are there any patterns that catch your attention in that plot?

```r
ggplot(merged_data_india, aes(x = date, y = tfp)) +
  geom_line(color = "blue") +
  labs(
    title = "Solow Residual (Total Factor Productivity) for India (1970-2019)",
    x = "Year",
    y = "Solow residual (TFP)"
  ) +
  scale_x_date(date_breaks = "3 years", date_labels = "%Y")
```




```
# theme_minimal()
```

1. Gradual increase before 1990: the slope rate is low and the growth is slow.
2. Fast sustained increase after 1991: might be correlated to the economic reform leading to an increase technology adoption, innovation, which increase the TFP (Singer, 1978).
3. Some fluctuations around 2007-2010: the stagnation of growth of TFP is likely due to the global financial crisis.

Question 02

```
Maddison_data <- read_dta("maddison2023_web.dta")
# MPD version 2023: Bolt, Jutta and Jan Luiten van Zanden (2024). Maddison style estimates of the evolu
```

Geary-Khamis dollar definition

Geary-Khamis dollar is a type of dollar that has the same amount of PPP as US Dollars in the US at a given point of time (Bolt, 2024). For example, people use the 1990 and 2016 USD as the standard of reference. The PPP as the basis of comparison is key here because it standardizes the currency by examining the same bundle of goods in different countries.

Plot GDP per capita

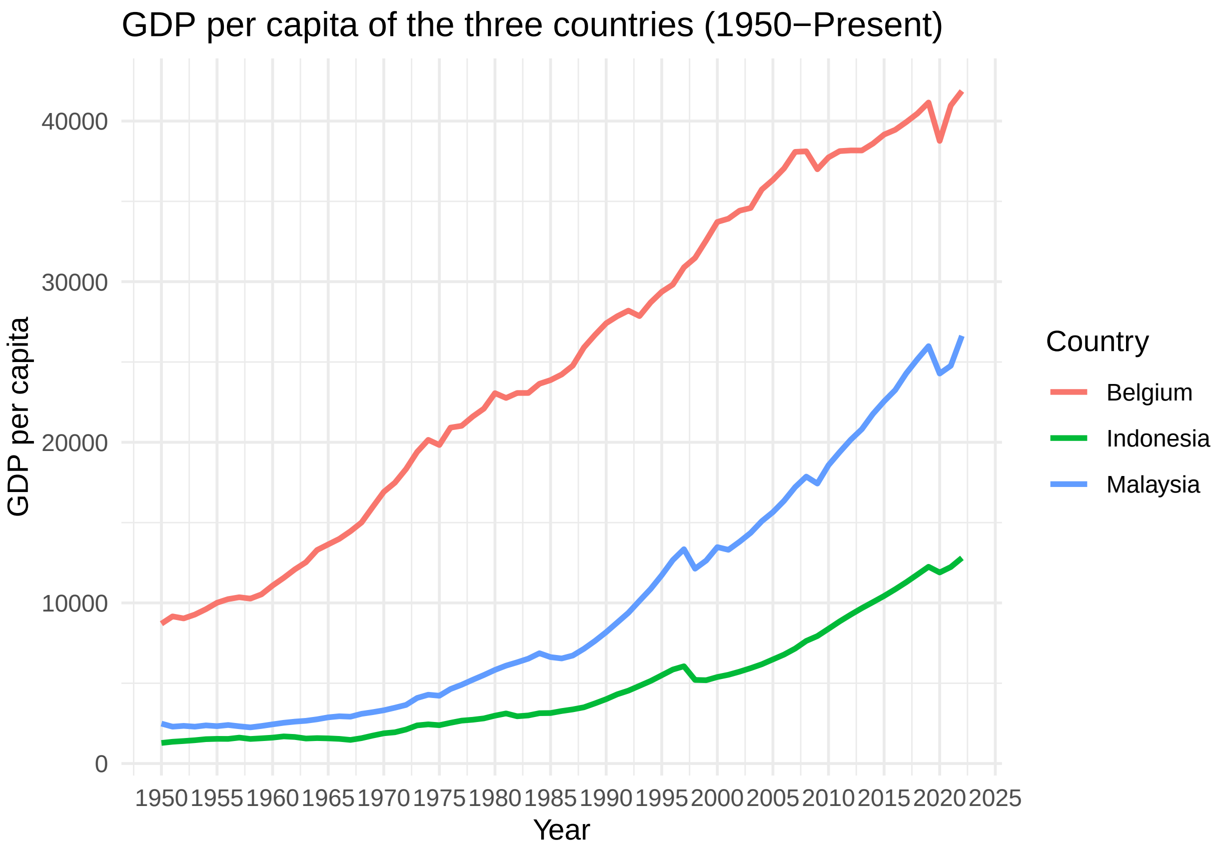
Plot GDP per capita for Belgium, Malaysia and Indonesia since 1950 (use just one graph for the three countries to be able to compare them).

```r
head(Maddison_data)
```

```
## # A tibble: 6 x 6
##   countrycode country     region                       year gdppc   pop
##   <chr>       <chr>       <chr>                       <dbl> <dbl> <dbl>
## 1 AFG         Afghanistan South and South East Asia     1    NA    NA
## 2 AFG         Afghanistan South and South East Asia   730    NA    NA
## 3 AFG         Afghanistan South and South East Asia  1000    NA    NA
## 4 AFG         Afghanistan South and South East Asia  1090    NA    NA
## 5 AFG         Afghanistan South and South East Asia  1150    NA    NA
## 6 AFG         Afghanistan South and South East Asia  1252    NA    NA
```

```r
# Filter the data for Belgium, Malaysia, and Indonesia since 1950
filtered_Maddison_data <- Maddison_data %>%
  filter(country %in% c("Belgium", "Malaysia", "Indonesia"), year >= 1950)

# Plot GDP per capita for the three countries
ggplot(filtered_Maddison_data, aes(x = year, y = gdppc, color = country)) +
  geom_line(size = 1) +
  labs(title = "GDP per capita of the three countries (1950-Present)",
       x = "Year",
       y = "GDP per capita",
       color = "Country") +
  scale_x_continuous(breaks = seq(1950, 2025, by = 5)) +
  theme_minimal()
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



Catch-up growth

Assume that GDP per capita in all three countries will continue to grow at the average growth rate they have experienced during the last 10 years of available data. How long will it take for Malaysia and Indonesia to catch up with Belgium in terms of GDP per capita? Detail the steps you take and discuss your results.

```
# Filter the data for the last 10 years
last_10_years <- Maddison_data %>%
  filter(year >= 2022-9) %>%
  filter(country %in% c("Belgium", "Malaysia", "Indonesia"))

# Compute the average annual growth rate (avg_growth_rate) for each country
growth_rates <- last_10_years %>%
  group_by(country) %>%
  summarize(
    gdp_initial = first(gdppc),
    gdp_final = last(gdppc),
    avg_growth_rate = (gdp_final / gdp_initial)^(1/10) - 1
  )

# Get the most recent GDP per capita for each country
latest_gdp <- Maddison_data %>%
  filter(year == max(year), country %in% c("Belgium", "Malaysia", "Indonesia")) %>%
  select(country, gdppc)

# Merge growth rates with latest GDP data
growth_rates <- left_join(growth_rates, latest_gdp, by = "country")
```

```r
# Belgium's GDP per capita (target)
gdp_belgium <- growth_rates %>% filter(country == "Belgium") %>% pull(gdppc)

# Function:
# Input: gdp_start, gdp_target, gdp_growth_rate
# Output: catchup_time
catchup_time <- function(gdp_start, growth_rate, gdp_target) {
  log(gdp_target / gdp_start) / log(1 + growth_rate)
}

# Compute the catch-up time required
func_catchup_years <- growth_rates %>%
  filter(country %in% c("Malaysia", "Indonesia")) %>%
  mutate(years_to_catch_up = catchup_time(gdppc, avg_growth_rate, gdp_belgium))

print(func_catchup_years)
```

```
## # A tibble: 2 x 6
##   country   gdp_initial gdp_final avg_growth_rate  gdppc years_to_catch_up
##   <chr>           <dbl>     <dbl>           <dbl>  <dbl>             <dbl>
## 1 Indonesia       9676.    12802.          0.0284 12802.              42.3
## 2 Malaysia       20819.    26629.          0.0249 26629.              18.4
```

Therefore, although the year-on-year growth rates for Indonesia and Malasia are similar, because Indonesia's initial GDP per capita was so low at $9676 compared to $20819 in Malaysia, the catch-up growth for Indonesia (42.34 years) will be much longer than that of Malaysia (18.39 years).

References

Bolt, J. (2024). Maddison-style estimates of the evolution of the world economy: A new 2023 update. Journal of Economic Surveys, 39(2), 631-671. https://doi.org/10.1111/joes.12618

Singer, H. W. (1978). Foreign trade regimes and economic development: India : Jagdish N. bhagwati and T.N. srinivasan, (national bureau of economic research, new york, 1975-distributed by columbia university press) pp. xxiv + 261, $15.00. Elsevier B.V. https://doi.org/10.1016/0022-1996(78)90011-9