Exploring Gapminder Dataset (HW3 - STAT 545)

Phuong (Sam) Can

12/09/2019

Contents

Let's first load the `gapminder` dataset.

```
knitr::opts_chunk$set(echo = TRUE)
library(tidyverse)
library(tsibble)
library(gapminder)
gapminder
```

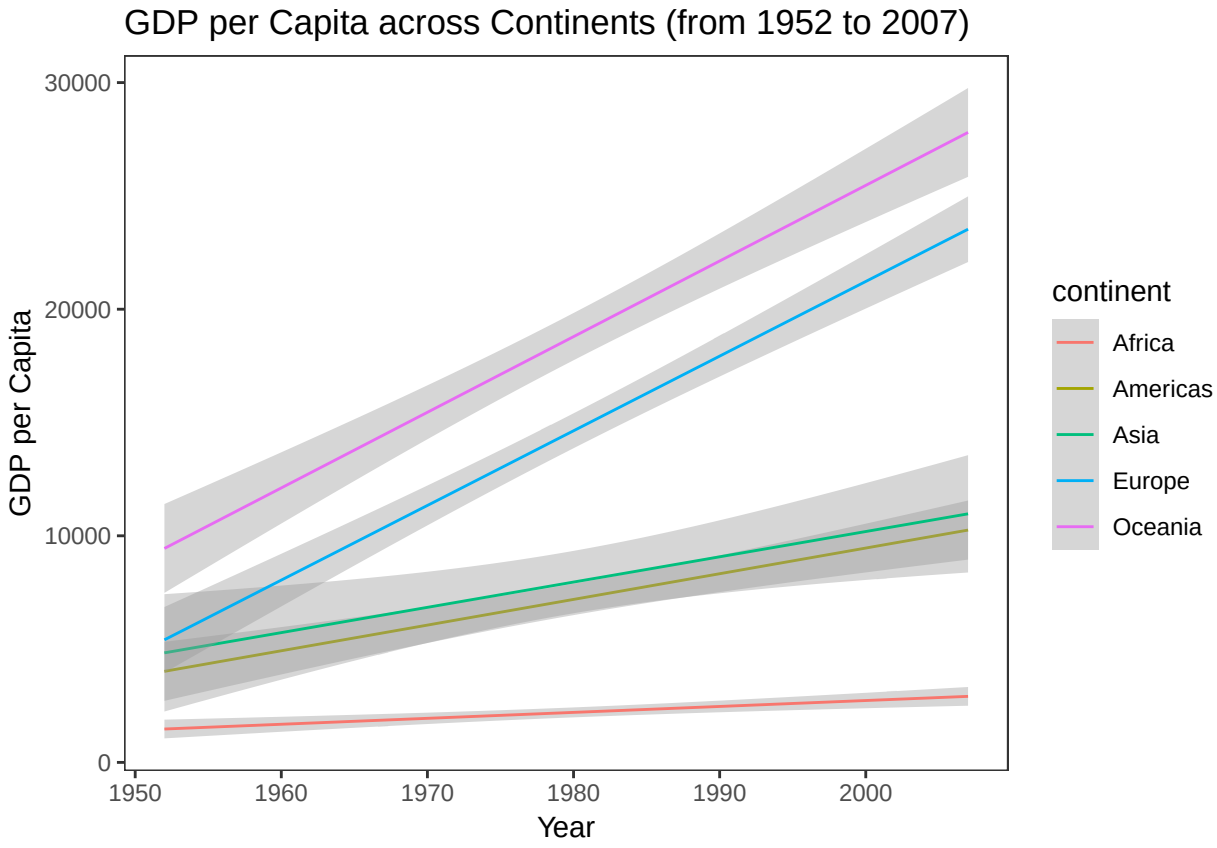
```
## # A tibble: 1,704 x 6
##    country     continent  year lifeExp      pop gdpPercap
##    <fct>       <fct>     <int>   <dbl>    <int>     <dbl>
##  1 Afghanistan Asia       1952    28.8  8425333      779.
##  2 Afghanistan Asia       1957    30.3  9240934      821.
##  3 Afghanistan Asia       1962    32.0 10267083      853.
##  4 Afghanistan Asia       1967    34.0 11537966      836.
##  5 Afghanistan Asia       1972    36.1 13079460      740.
##  6 Afghanistan Asia       1977    38.4 14880372      786.
##  7 Afghanistan Asia       1982    39.9 12881816      978.
##  8 Afghanistan Asia       1987    40.8 13867957      852.
##  9 Afghanistan Asia       1992    41.7 16317921      649.
## 10 Afghanistan Asia       1997    41.8 22227415      635.
## # ... with 1,694 more rows
```

Then, let us look at the nuanced stories that the dataset `gapminder` entails.

GDP per Capita

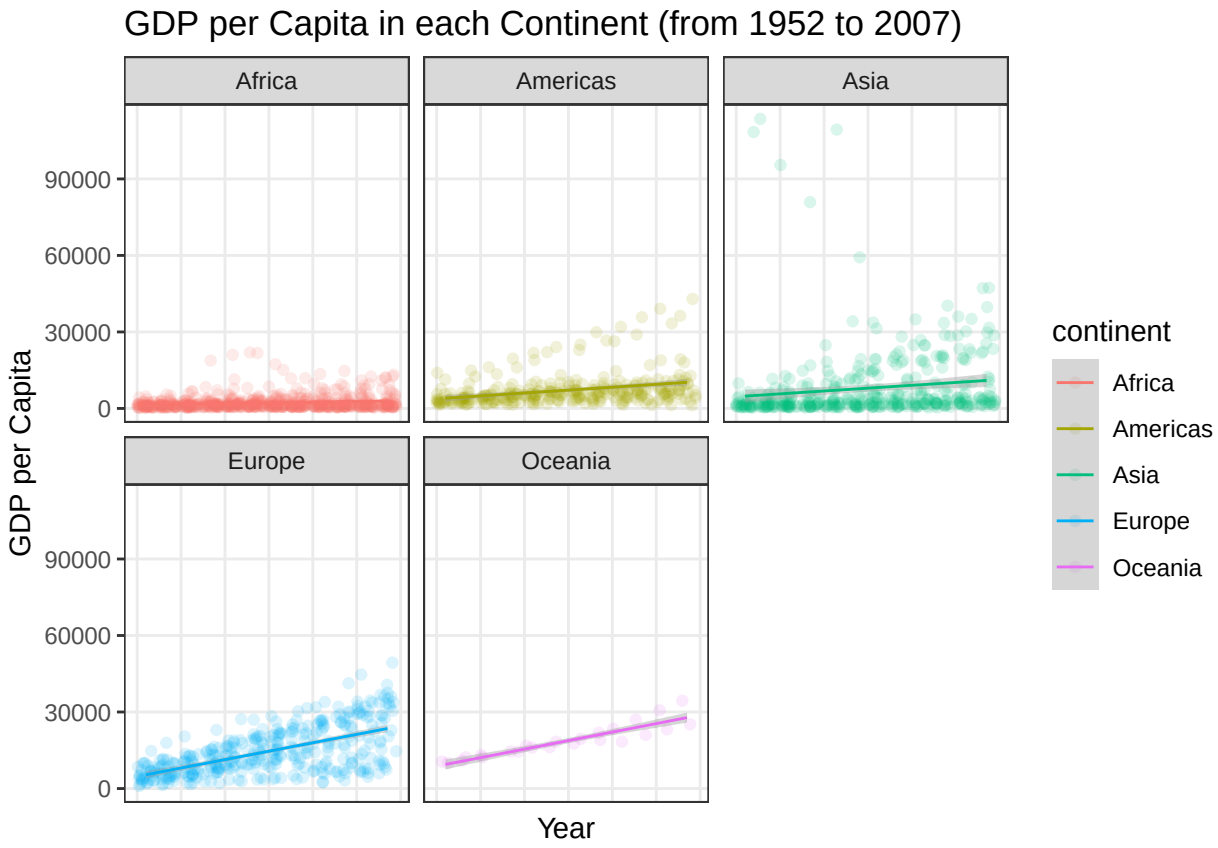
First, the dataset offers insights into how the GDP per capita changed within each country and each continent overtime. We will look at five regression lines in which `gdpPercap` is regressed on `year` to see the overall trend of change in GDP per capita for each continent from 1952 to 2007.

```
ggplot(gapminder, aes(x=year, y=gdpPercap, color = continent)) +
  geom_smooth(method='lm',formula=y~x, size = 0.5) +
  theme_bw() +
  theme(panel.grid = element_blank()) +
  labs(title="GDP per Capita across Continents (from 1952 to 2007)", x="Year", y="GDP per Capita")
```



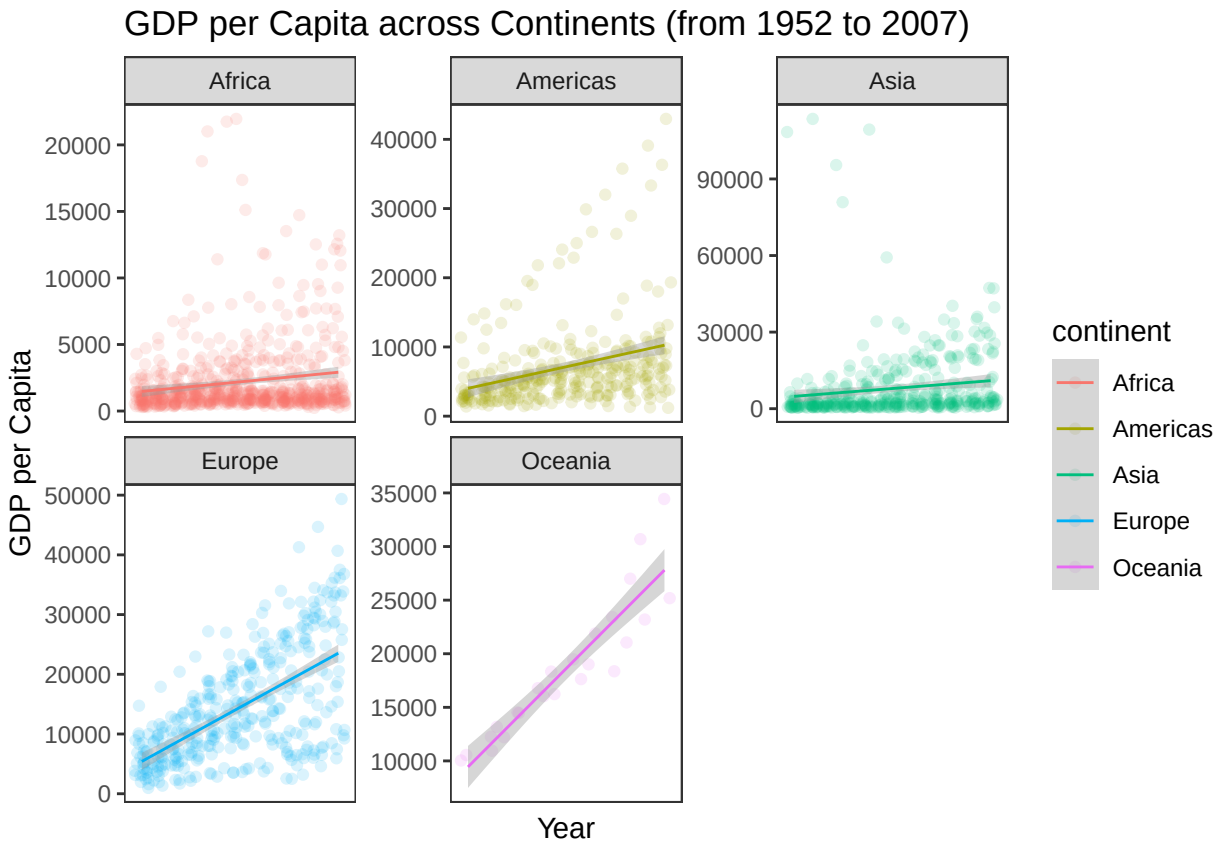
As we can see in the graph, Oceania had the steepest increase in GDP per capita among the continents from 1952 to 2007, followed by Europe, Asia, Americas, and lastly, Africa. Specifically, on average, GDP per capita in Oceania increased by 19512.1 over the time period, while there was only a slight increase of 1836.46 in GDP per capita in Africa. This varying increases can be attributable to the differing industrialization periods in each continent.

```
ggplot(gapminder, aes(x=year, y=gdpPercap, color = continent)) +
  geom_jitter(alpha = 0.15) +
  geom_smooth(method='lm',formula=y~x, size = 0.5) +
  facet_wrap(~ continent) +
  theme_bw() +
  theme(panel.grid.minor = element_blank(), axis.text.x = element_blank(), axis.ticks.x = element_blank
  labs(title="GDP per Capita in each Continent (from 1952 to 2007)", x="Year", y="GDP per Capita")
```



We can examine the increase per continent by faceting data into individual continents. This scatterplot with jittering also informs us about how spread the data are within each continent.

```
ggplot(gapminder, aes(x=year, y=gdpPercap, color = continent)) +
  geom_jitter(alpha = 0.15) +
  geom_smooth(method='lm',formula=y~x, size = 0.5) +
  facet_wrap(~ continent, scales = "free_y") +
  theme_bw() +
  theme(panel.grid = element_blank(), axis.text.x = element_blank(), axis.ticks.x = element_blank()) +
  labs(title="GDP per Capita across Continents (from 1952 to 2007)", x="Year", y="GDP per Capita")
```



When we put the data for each continent in free scales, or let the limits for GDP per capita vary across individual continents' graphs, we can examine more closely the overall increase for each continent within each continent's own range of GDP per capita data.

Life Expectancy

Let's look at what continent has the biggest increase in life expectancy overtime.

```
gap_inc <- gapminder %>%
  group_by(country) %>%
  mutate(lifeExp_inc = diff(lifeExp, lag = 11)) %>%
  group_by(continent) %>%
  mutate(n_countries = n_distinct(country)) %>%
  mutate(lifeExp_inc = mean(lifeExp_inc))

ggplot(gap_inc, aes(continent, lifeExp_inc)) +
  geom_bar(stat = "sum") +
  scale_fill_grey() +
  theme(legend.position = "none", panel.grid = element_blank()) +
  theme_bw() +
  labs(title="Increase in Life Expectancy across Continents (from 1952 to 2007)", x="Continent", y="Inc
```

Asia had the biggest jump in life expectancy, followed by Americas, Africa, Europe, and Oceania. Specifically, life expectancy in Asia increased by 24.41 years, while life expectancy in Oceania, the continent where living standards and life expectancy were already high, only increased by 11.46 years.

The Story behind China

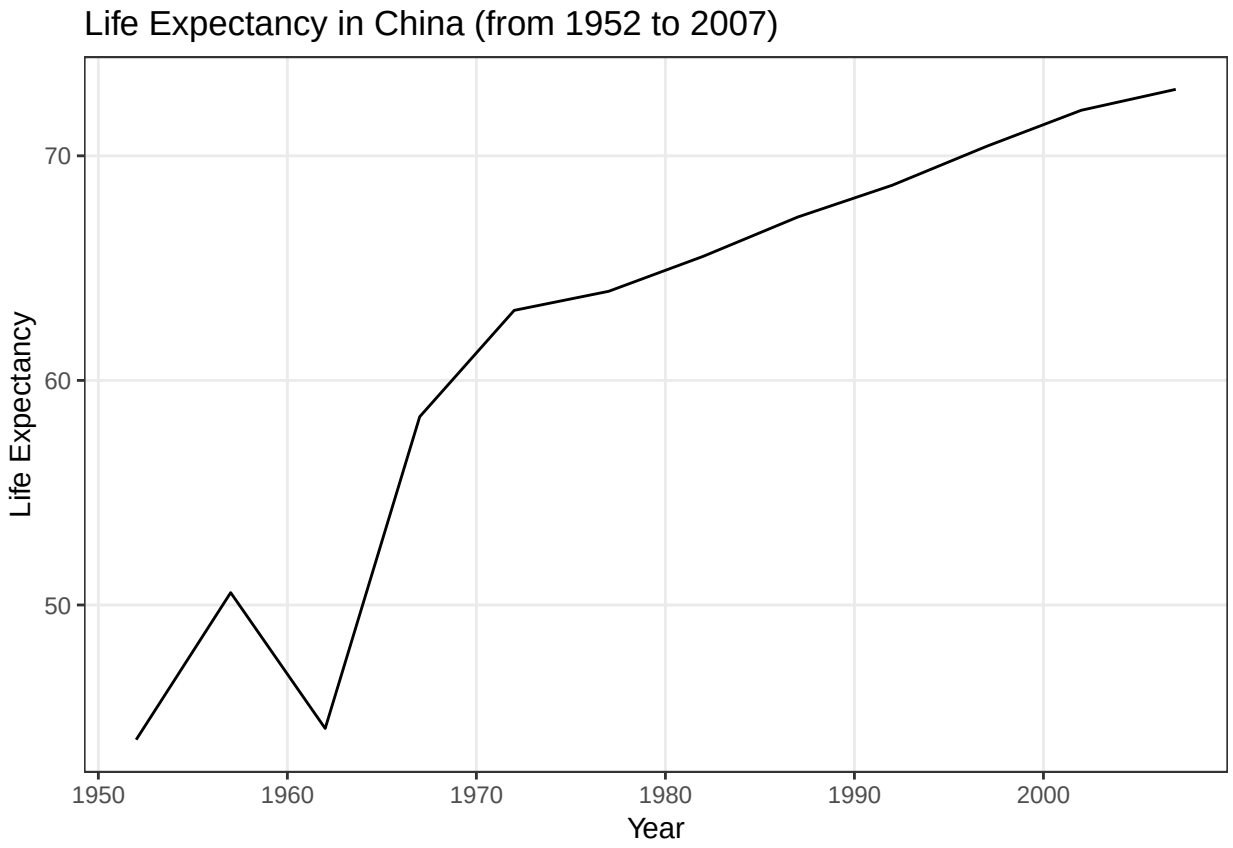
From 1952 to 2007, China has gone through historical events that directly affected many of its social outcomes such as life expectancy and GDP per capita. Let's look at them more closely.

Social variables' data in China are presented below:

```
gap_china <- gapminder %>%
  filter(country == "China")
gap_china
```

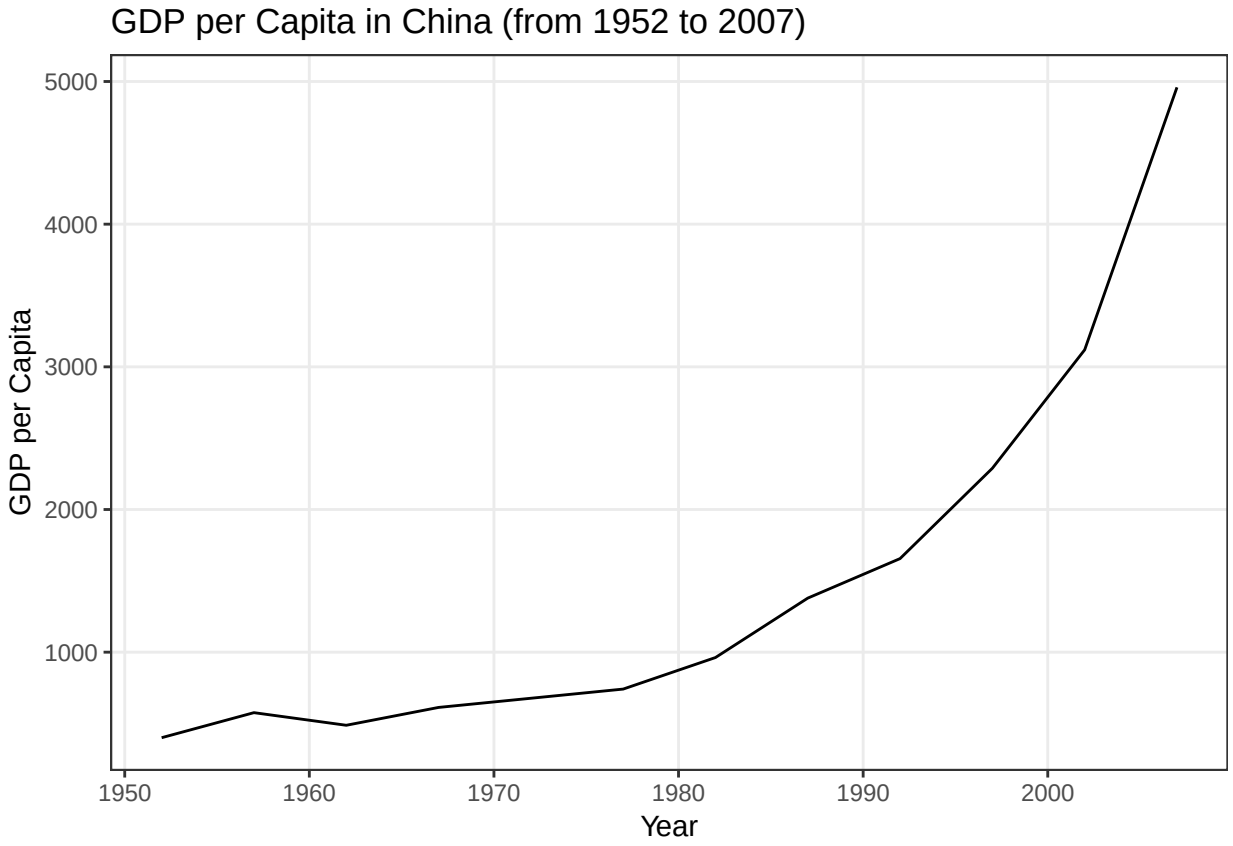
```
## # A tibble: 12 x 6
##    country continent  year lifeExp        pop gdpPercap
##    <fct>   <fct>     <int>   <dbl>      <int>     <dbl>
##  1 China   Asia       1952    44    556263527      400.
##  2 China   Asia       1957    50.5  637408000      576.
##  3 China   Asia       1962    44.5  665770000      488.
##  4 China   Asia       1967    58.4  754550000      613.
##  5 China   Asia       1972    63.1  862030000      677.
##  6 China   Asia       1977    64.0  943455000      741.
##  7 China   Asia       1982    65.5 1000281000      962.
##  8 China   Asia       1987    67.3 1084035000     1379.
##  9 China   Asia       1992    68.7 1164970000     1656.
## 10 China   Asia       1997    70.4 1230075000     2289.
## 11 China   Asia       2002    72.0 1280400000     3119.
## 12 China   Asia       2007    73.0 1318683096     4959.
```

```
ggplot(gap_china, aes(year, lifeExp)) +
  geom_line() +
  labs(title="Life Expectancy in China (from 1952 to 2007)", x="Year", y="Life Expectancy") +
  theme_bw() +
  theme(panel.grid.minor = element_blank())
```



Life expectancy in China didn't simply follow a monotonic increase. In fact, there was a sharp plunge near 1958-1963, right when the Chinese government launched the five-year economic plan "Great Leap Forward" that caused economic breakdown and millions of deaths for starvation. However, life expectancy quickly rose afterwards thanks to progressive economic policies and improvement in living standards.

```
ggplot(gap_china, aes(year, gdpPercap)) +
  geom_line() +
  labs(title="GDP per Capita in China (from 1952 to 2007)", x="Year", y="GDP per Capita") +
  theme_bw() +
  theme(panel.grid.minor = element_blank())
```



China obviously experienced a huge spike in GDP per capita throughout the last 50 years! From a low-income country with 400.45 GDP per capita, it became an upper-middle-income economy with 4959.11 GDP per capita. This spike happened thanks to the *Open-door policy* that opened the country to foreign investment, market economy, and thriving private sector.