Lab 12 Homework

Sameeksha Deshatty

2025-02-24

Instructions

Answer the following questions and complete the exercises in RMarkdown. Please embed all of your code and push your final work to your repository. Your final lab report should be organized, clean, and run free from errors. Remember, you must remove the # for the included code chunks to run. Be sure to add your name to the author header above. For any included plots, make sure they are clearly labeled. You are free to use any plot type that you feel best communicates the results of your analysis.

Make sure to use the formatting conventions of RMarkdown to make your report neat and clean!

Load the libraries

```
library(tidyverse)
library(janitor)
library(naniar)
```

```
options(scipen = 999)
```

Resources

The idea for this assignment came from Rebecca Barter's ggplot tutorial so if you get stuck this is a good place to have a look.

Gapminder

For this assignment, we are going to use the dataset gapminder. Gapminder includes information about economics, population, and life expectancy from countries all over the world. You will need to install it before use.

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

1. Use the function(s) of your choice to get an idea of the overall structure of the data frame, including its dimensions, column names, variable classes, etc. As part of this, determine how NA's are treated in the data.

```r
str(gapminder)
```

```
## tibble [1,704 x 6] (S3: tbl_df/tbl/data.frame)
##  $ country  : Factor w/ 142 levels "Afghanistan",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ continent: Factor w/ 5 levels "Africa","Americas",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ year     : int [1:1704] 1952 1957 1962 1967 1972 1977 1982 1987 1992 1997 ...
##  $ lifeExp  : num [1:1704] 28.8 30.3 32 34 36.1 ...
##  $ pop      : int [1:1704] 8425333 9240934 10267083 11537966 13079460 14880372 12881816 13867957 163
##  $ gdpPercap: num [1:1704] 779 821 853 836 740 ...
```

```r
head(gapminder)
```

```
## # A tibble: 6 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.
```

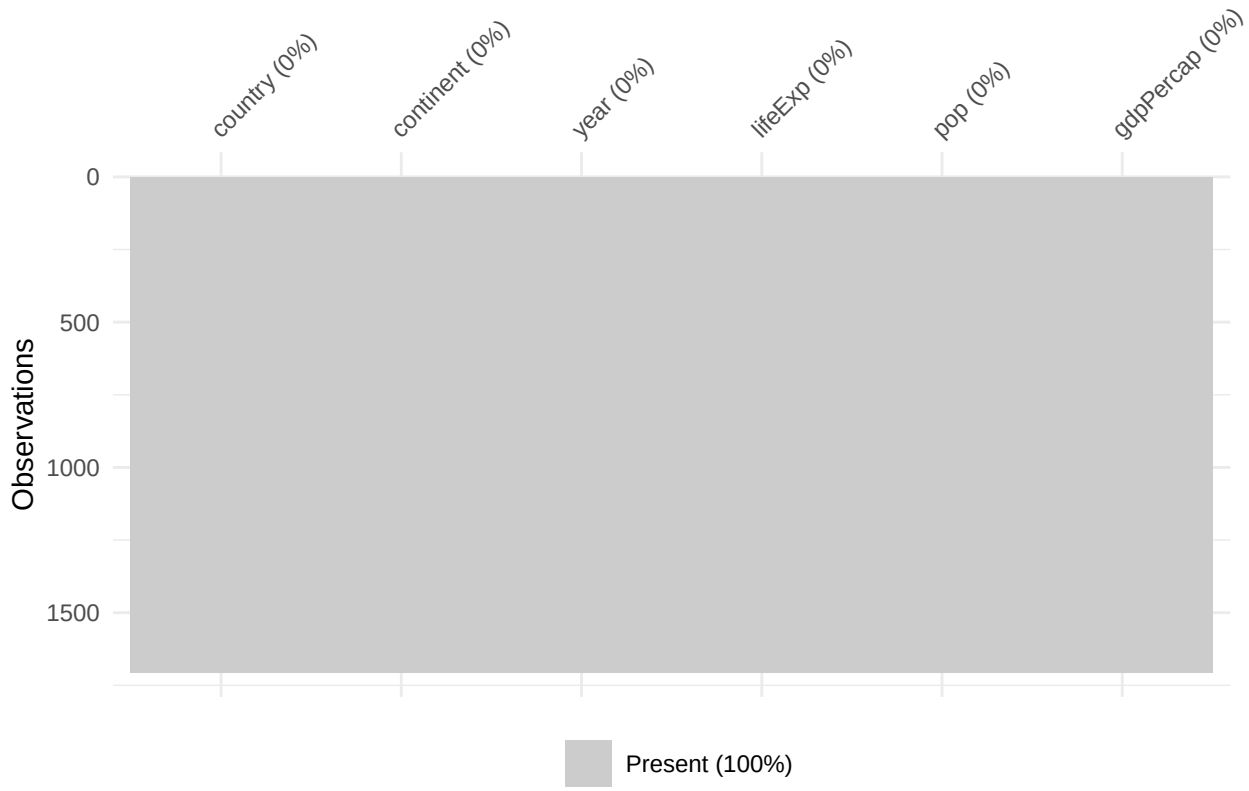
```r
summary(gapminder)
```

```
##         country        continent        year         lifeExp
##  Afghanistan:  12   Africa  :624   Min.   :1952   Min.   :23.60
##  Albania    :  12   Americas:300   1st Qu.:1966   1st Qu.:48.20
##  Algeria    :  12   Asia    :396   Median :1980   Median :60.71
##  Angola     :  12   Europe  :360   Mean   :1980   Mean   :59.47
##  Argentina  :  12   Oceania : 24   3rd Qu.:1993   3rd Qu.:70.85
##  Australia  :  12                  Max.   :2007   Max.   :82.60
##  (Other)    :1632
##       pop              gdpPercap
##  Min.   :     60011   Min.   :   241.2
##  1st Qu.:   2793664   1st Qu.:  1202.1
##  Median :   7023596   Median :  3531.8
##  Mean   :  29601212   Mean   :  7215.3
##  3rd Qu.:  19585222   3rd Qu.:  9325.5
##  Max.   :1318683096   Max.   :113523.1
##
```

```r
sum(is.na(gapminder))
```

```
## [1] 0
```

```r
vis_miss(gapminder)
```



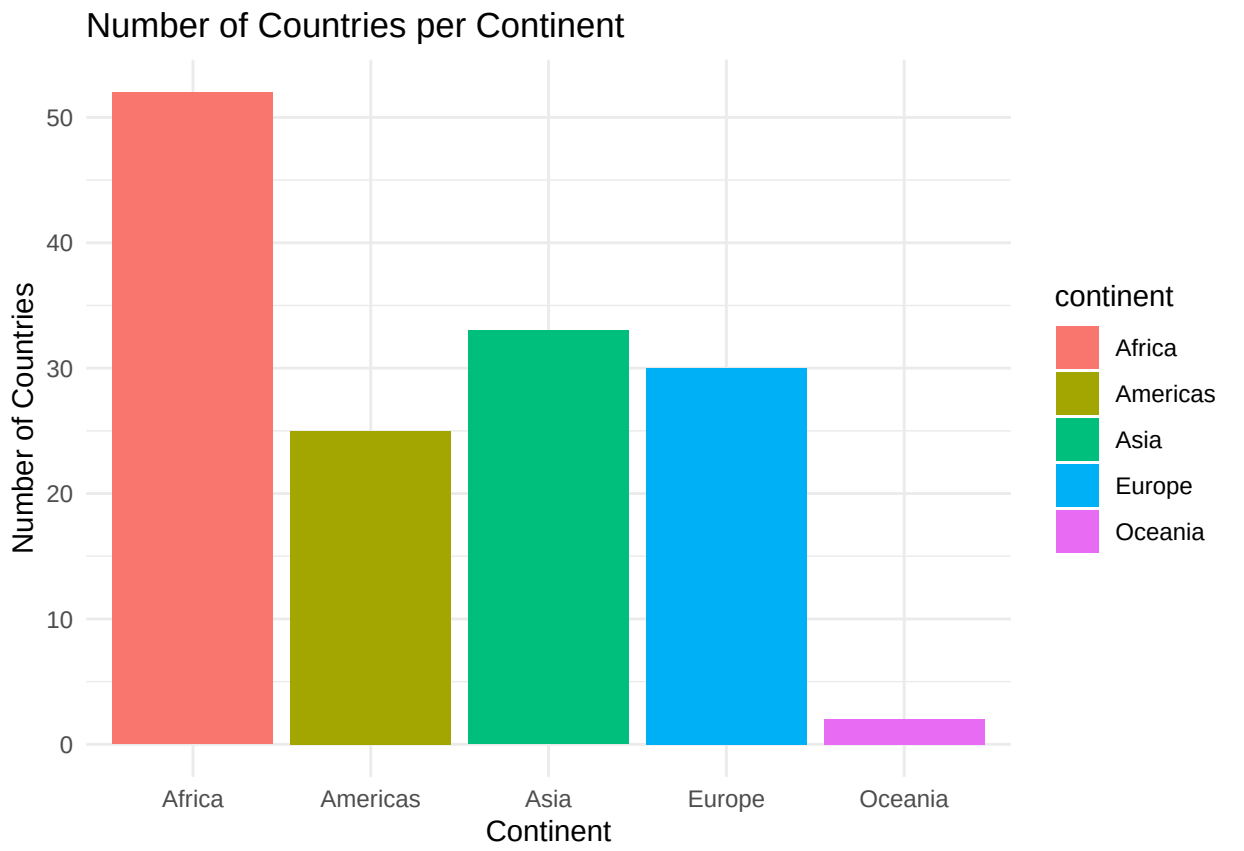
2. How many countries are represented in this dataset? Make a table and a plot that shows the number of countries by continent.

```
country_counts <- gapminder %>%
  group_by(continent) %>%
  summarise(num_countries = n_distinct(country))

country_counts
```

```
## # A tibble: 5 x 2
##   continent num_countries
##   <fct>             <int>
## 1 Africa               52
## 2 Americas             25
## 3 Asia                 33
## 4 Europe               30
## 5 Oceania               2
```

```
ggplot(country_counts, aes(x = continent, y = num_countries, fill = continent)) +
  geom_bar(stat = "identity") +
  labs(title = "Number of Countries per Continent", x = "Continent", y = "Number of Countries") +
  theme_minimal()
```



3. Which country has the largest population growth since 1952? Show this as a table.

```
pop_growth <- gapminder %>%
  group_by(country) %>%
  summarise(pop_growth = max(pop) - min(pop)) %>%
  arrange(desc(pop_growth))

pop_growth[1, ]
```

```
## # A tibble: 1 x 2
##   country pop_growth
##   <fct>        <int>
## 1 China    762419569
```

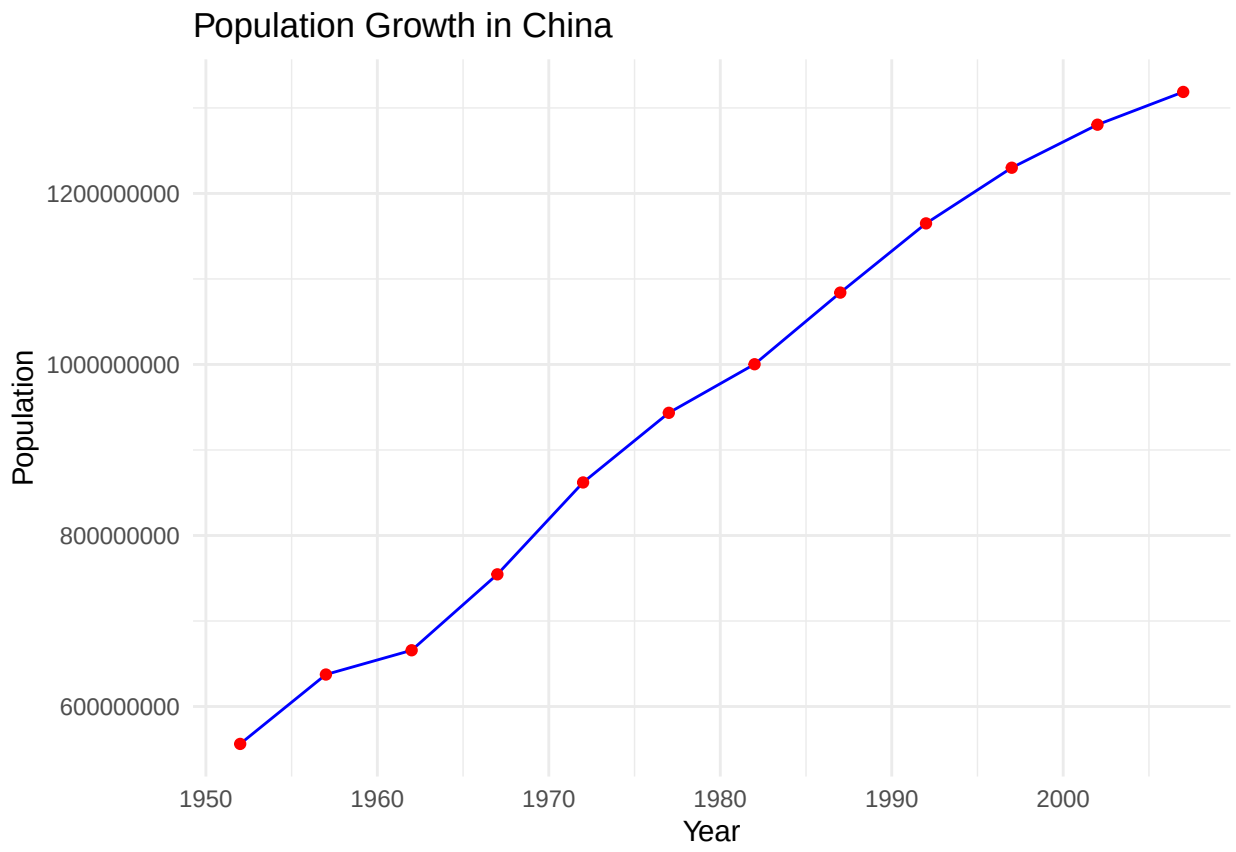
China has the largest population growth since 1970.

4. Make a plot that shows population growth for the country you found in question #3. This plot should show the change over time.

```
top_country <- pop_growth$country[1]

gapminder %>%
  filter(country == top_country) %>%
  ggplot(aes(x = year, y = pop)) +
```

```
  geom_line(color = "blue") +
  geom_point(color = "red") +
  labs(title = paste("Population Growth in", top_country),
       x = "Year", y = "Population") +
  theme_minimal()
```



5. How has global life expectancy changed between 1952 and 2007? Show the min, mean, and max for all countries in the dataset. Show this as a table.

```
life_expectancy_summary <- gapminder %>%
  group_by(year) %>%
  summarise(
    min_lifeExp = min(lifeExp),
    mean_lifeExp = mean(lifeExp),
    max_lifeExp = max(lifeExp)
  )
life_expectancy_summary
```

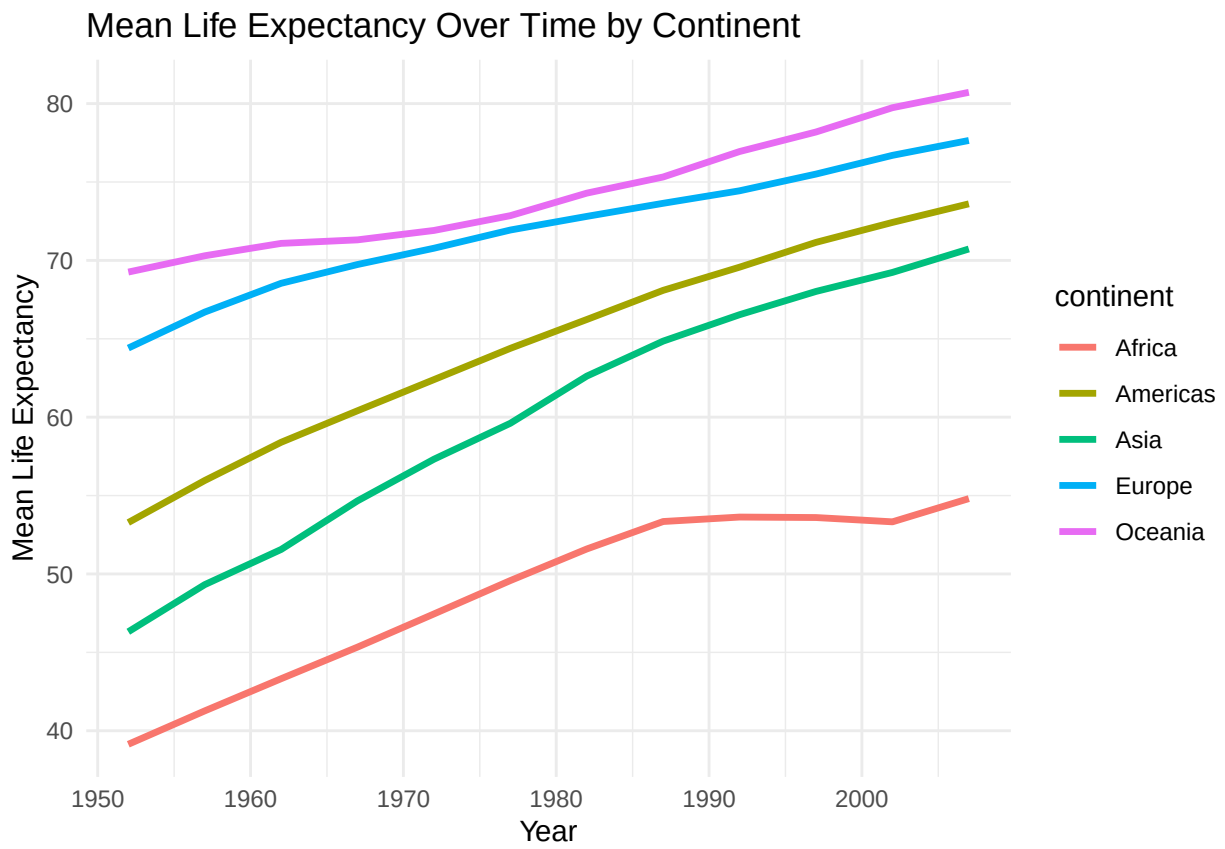
```
## # A tibble: 12 x 4
##     year min_lifeExp mean_lifeExp max_lifeExp
##    <int>       <dbl>        <dbl>       <dbl>
##  1  1952        28.8         49.1        72.7
##  2  1957        30.3         51.5        73.5
##  3  1962        32.0         53.6        73.7
```

```
##  4  1967        34.0         55.7         74.2
##  5  1972        35.4         57.6         74.7
##  6  1977        31.2         59.6         76.1
##  7  1982        38.4         61.5         77.1
##  8  1987        39.9         63.2         78.7
##  9  1992        23.6         64.2         79.4
## 10  1997        36.1         65.0         80.7
## 11  2002        39.2         65.7         82
## 12  2007        39.6         67.0         82.6
```

6. Make a plot that shows how mean life expectancy has changed over time for each continent. What is your interpretation of what happened in Africa between 1987 and 2002?

```r
gapminder %>%
  group_by(year, continent) %>%
  summarise(mean_lifeExp = mean(lifeExp), .groups = "drop") %>%
  ggplot(aes(x = year, y = mean_lifeExp, color = continent)) +
  geom_line(size = 1.2) +
  labs(title = "Mean Life Expectancy Over Time by Continent",
       x = "Year", y = "Mean Life Expectancy") +
  theme_minimal()
```

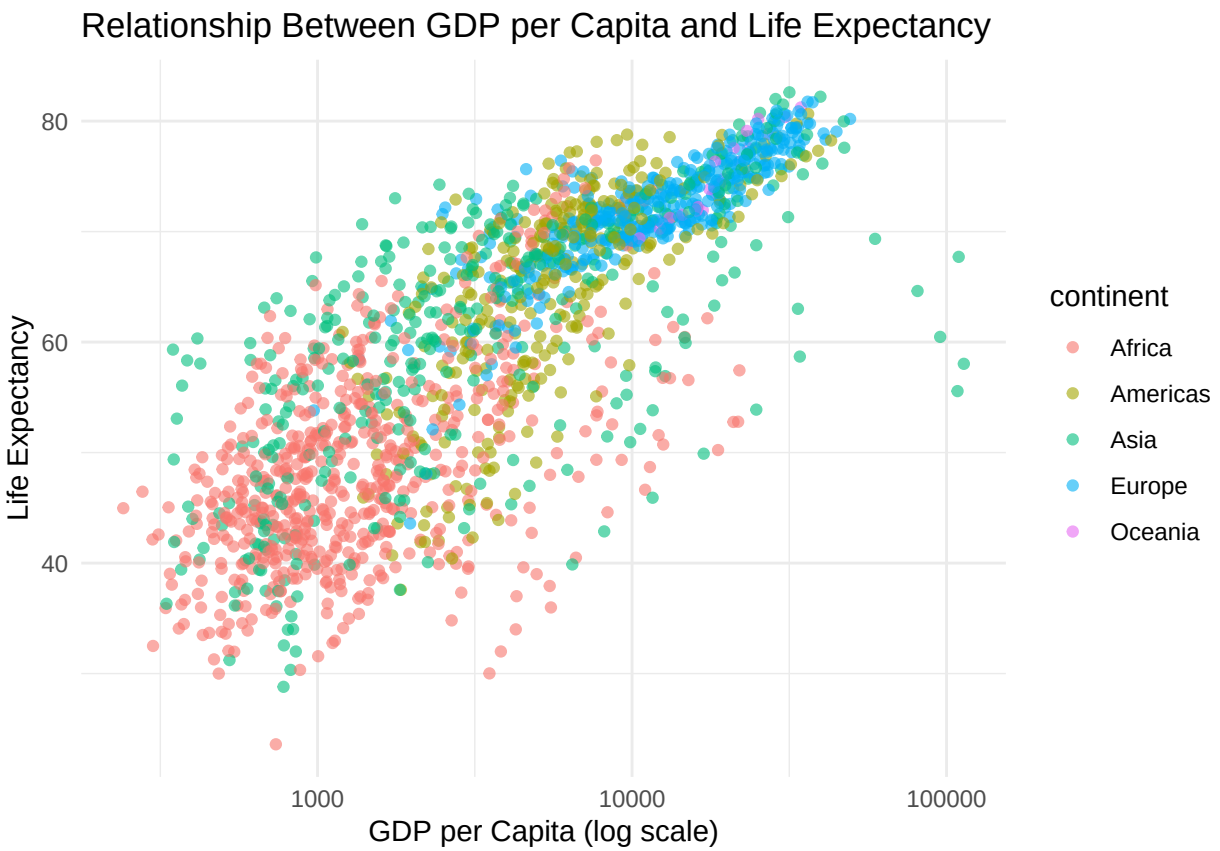
```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```



My interpretation is that between 1987 and 2002, the mean life expectancy in Africa held steady at around 54 years and was not increasing. Most likely due to adverse conditions and a lack of technological health improvement as oppposed to the rest of the world.

7. We are interested in the relationship between per capita GDP and life expectancy; i.e. does having more money help you live longer? Show this as a plot.

```
gapminder %>%
  ggplot(aes(x = gdpPercap, y = lifeExp, color = continent)) +
  geom_point(alpha = 0.6) +
  scale_x_log10() +  # Use log scale for better visualization
  labs(title = "Relationship Between GDP per Capita and Life Expectancy",
       x = "GDP per Capita (log scale)", y = "Life Expectancy") +
  theme_minimal()
```



Yes, the relationship between GDP per capita and life expectancy is positive, meaning a higher GDP per capita does lead to a higher life expectancy. This is likely to advanced technology and access to better healthcare that comes with having money.

8. Which five countries have had the highest GDP per capita growth over the years represented in this dataset? Show this as a table.

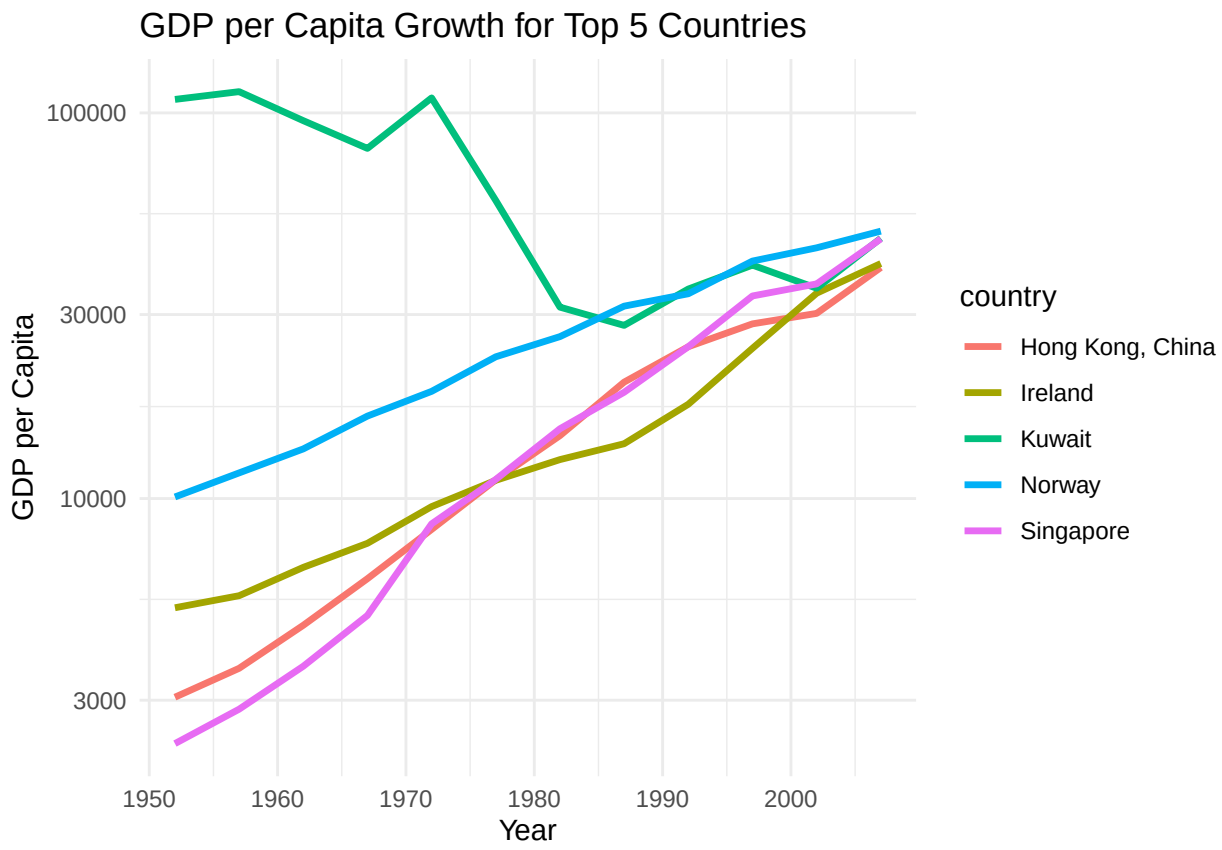
```
gdp_growth <- gapminder %>%
  group_by(country) %>%
  summarise(gdp_growth = max(gdpPercap) - min(gdpPercap)) %>%
  arrange(desc(gdp_growth))

top_5_gdp_countries <- gdp_growth %>% slice(1:5)
top_5_gdp_countries
```

```
## # A tibble: 5 x 2
##   country          gdp_growth
##   <fct>                 <dbl>
## 1 Kuwait               85405.
## 2 Singapore            44828.
## 3 Norway               39262.
## 4 Hong Kong, China     36671.
## 5 Ireland              35466.
```

9. How does per capita GDP growth compare between these same five countries? Show this as a plot.

```
gapminder %>%
  filter(country %in% top_5_gdp_countries$country) %>%
  ggplot(aes(x = year, y = gdpPercap, color = country)) +
  geom_line(size = 1.2) +
  labs(title = "GDP per Capita Growth for Top 5 Countries",
       x = "Year", y = "GDP per Capita") +
  scale_y_log10() +  # Use log scale
  theme_minimal()
```



10. Do one analysis of your choice that includes a table and plot as outputs.

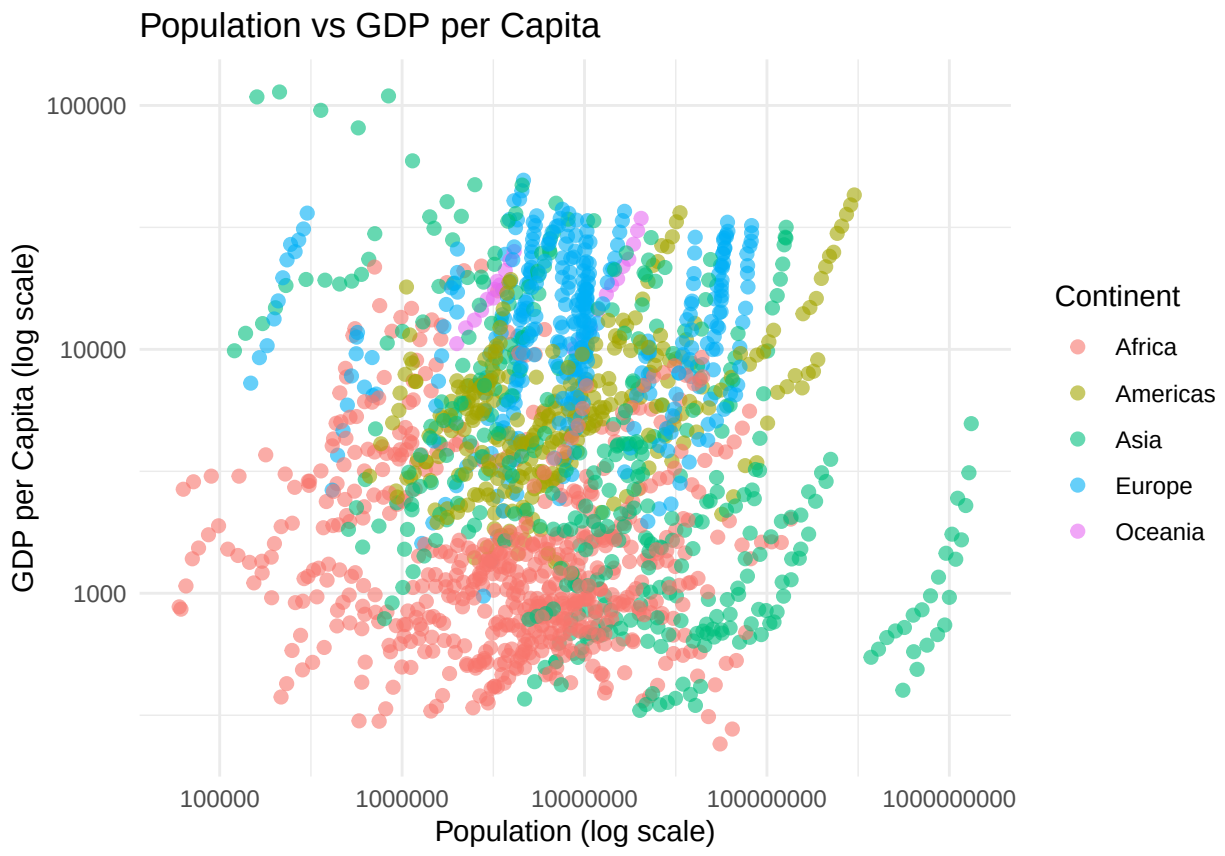
```
pop_gdp_table <- gapminder %>%
  group_by(country) %>%
  summarise(
    `Latest Population (millions)` = max(pop) / 1e6,
    `Latest GDP per Capita (USD)` = max(gdpPercap),
    `Total GDP (Billion USD)` = (max(pop) * max(gdpPercap)) / 1e9
  ) %>%
  arrange(desc(`Total GDP (Billion USD)`))

pop_gdp_table
```

```
## # A tibble: 142 x 4
##    country  Latest Population (m~1 Latest GDP per Capit~2 Total GDP (Billion U~3
##    <fct>                     <dbl>                  <dbl>                  <dbl>
##  1 United ~                  301.                  42952.                  12934.
```

```
##  2 China                1319.              4959.              6540.
##  3 Japan                 127.             31656.              4035.
##  4 India                1110.              2452.              2723.
##  5 Germany                82.4            32170.              2651.
##  6 United ~               60.8            33203.              2018.
##  7 France                 61.1            30470.              1861.
##  8 Brazil                190.              9066.              1723.
##  9 Italy                  58.1            28570.              1661.
## 10 Mexico                109.             11978.              1302.
## # i 132 more rows
## # i abbreviated names: 1: 'Latest Population (millions)',
## #   2: 'Latest GDP per Capita (USD)', 3: 'Total GDP (Billion USD)'
```

```r
gapminder %>%
  ggplot(aes(x = pop, y = gdpPercap, color = continent)) +
  geom_point(alpha = 0.6, size = 2) +
  scale_x_log10() +
  scale_y_log10() +
  labs(
    title = "Population vs GDP per Capita",
    x = "Population (log scale)",
    y = "GDP per Capita (log scale)",
    color = "Continent"
  ) +
  theme_minimal()
```



Knit and Upload

Please knit your work as a .pdf or .html file and upload to Canvas. Homework is due before the start of the next lab. No late work is accepted. Make sure to use the formatting conventions of RMarkdown to make your report neat and clean!