PHP 1510 Homework 2

Mariel Pacada

10/17/2019

Question 1

How many unique countries are represented per continent?

Below, we see a tibble that shows how many unique countries are in each continent.

```
by_continent <- gapminder %>%
                  select(continent, country) %>%
                  group_by(continent) %>%
                  summarize(countryCount = n_distinct(country))

by_continent
```

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

Question 2

Which European nation had the lowest GDP per capita in 1997?

Albania had the lowest GDP per capita in 1997 Europe.

```
by_GDP <- gapminder %>%
            filter(continent == "Europe", year == 1997) %>%
            select(country, gdpPercap) %>%
            arrange(gdpPercap)

head(by_GDP)
```

```
## # A tibble: 6 x 2
##   country                gdpPercap
##   <fct>                      <dbl>
## 1 Albania                    3193.
## 2 Bosnia and Herzegovina     4766.
## 3 Bulgaria                   5970.
## 4 Montenegro                 6466.
## 5 Turkey                     6601.
## 6 Romania                    7347.
```

Question 3

According to the data available, what was the average life expectancy across each continent in the 1980s?

Below, we see a tibble that shows the average life expectancy in each continent in the 1980s.

```
continent_life_exp <- gapminder %>%
                        filter(year > 1979 & year < 1990) %>%
                        group_by(continent) %>%
                        summarize(avgLifeExp = mean(lifeExp))

continent_life_exp
```

```
## # A tibble: 5 x 2
##   continent avgLifeExp
##   <fct>          <dbl>
## 1 Africa          52.5
## 2 Americas        67.2
## 3 Asia            63.7
## 4 Europe          73.2
## 5 Oceania         74.8
```

Question 4

What 5 countries have the highest total GDP over all years combined?

Kuwait, Switzerland, Norway, the United States, and Canada are the five countries with the highest GDP over all years in this dataset.

```
total_gdp <- gapminder %>%
               select(country, gdpPercap) %>%
               group_by(country) %>%
               summarize(totalGdp = sum(gdpPercap)) %>%
               arrange(desc(totalGdp))

head(total_gdp, 5)
```

```
## # A tibble: 5 x 2
##   country       totalGdp
##   <fct>            <dbl>
## 1 Kuwait         783995.
## 2 Switzerland    324892.
## 3 Norway         320968.
## 4 United States  315134.
## 5 Canada         268929.
```

Question 5

What countries and years had life expectancies of at least 80 years?

Below, we can see the 22 instances (for 13 distinct countries) where life expectancy was at least 80 years.

```
life_exp_eighty <- gapminder %>%
                     select(country, lifeExp, year) %>%
                     filter(lifeExp >= 80)

life_exp_eighty
```

```
## # A tibble: 22 x 3
##    country          lifeExp  year
##    <fct>              <dbl> <int>
##  1 Australia           80.4  2002
##  2 Australia           81.2  2007
##  3 Canada              80.7  2007
##  4 France              80.7  2007
##  5 Hong Kong, China    80    1997
##  6 Hong Kong, China    81.5  2002
##  7 Hong Kong, China    82.2  2007
##  8 Iceland             80.5  2002
##  9 Iceland             81.8  2007
## 10 Israel              80.7  2007
## 11 Italy               80.2  2002
## 12 Italy               80.5  2007
## 13 Japan               80.7  1997
## 14 Japan               82    2002
## 15 Japan               82.6  2007
## 16 New Zealand         80.2  2007
## 17 Norway              80.2  2007
## 18 Spain               80.9  2007
## 19 Sweden              80.0  2002
## 20 Sweden              80.9  2007
## 21 Switzerland         80.6  2002
## 22 Switzerland         81.7  2007
```

Question 6

Which three countries have had the most consistent population estimates (i.e. lowest standard deviation) across the years of available data?

Sao Tome and Principe, Iceland, and Montenegro are the three countries with the most consistent population estimates over the years of available data.

```
lowest_sd <- gapminder %>%
               select(country, pop) %>%
               group_by(country) %>%
               summarize(popStdev = sd(pop)) %>%
               arrange(popStdev)

head(lowest_sd, 3)
```

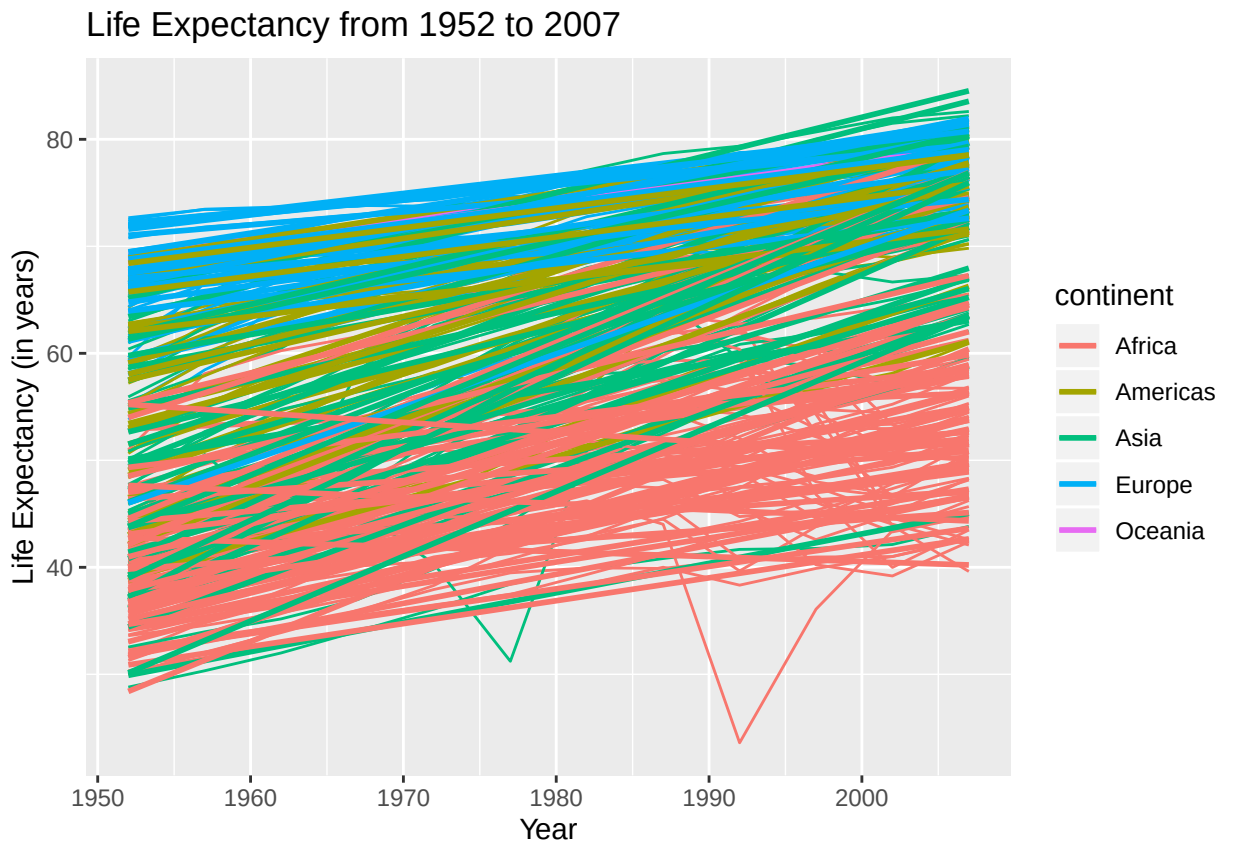
```
## # A tibble: 3 x 2
##   country                popStdev
##   <fct>                     <dbl>
## 1 Sao Tome and Principe   45906.
## 2 Iceland                 48542.
## 3 Montenegro              99738.
```

Question 7 and 8

Create a plot about life expectancy.

The plot below shows the progression of each country's life expectancy over time. Although not all **countries** have followed a consistent upward trend, the plot shows an overall upward trend for life expectancy across all **continents**. While it seems all the continents have experienced a steady increase (and at a considerably similar rate), their starting (1952) and ending (2007) values are distinct from each other. Africa has had the lowest life expectancy over the years, while Oceania and Europe lead the others.

```
ggplot(gapminder, aes(x = year, y = lifeExp, group = country, color = continent)) +
  geom_line() + geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Life Expectancy from 1952 to 2007", x = "Year",
       y = "Life Expectancy (in years)")
```



Question 9 and 10

Create boxplots of life expectancy by continent.

The boxplots below visualize the spread of average life expectancy for each continent. Similarly to their trends over time, Oceania and Europe have very similar distributions and lead the rest in terms of median life expectancy. (Note, however, that Europe has several outliers unlike Oceania.) Africa has the lowest median life expectancy, and Asia has the largest standard deviation.

```
ggplot(gapminder, aes(x = continent, y = lifeExp)) + geom_boxplot() +
  geom_jitter(alpha = 0.04) + labs(title = "Life Expectancy by Continent (1952 - 2007)",
                                   x = "Continent", y = "Life Expectancy (in years)")
```