Assignment 2 Q1

Ian Murphy

2019-09-18

Assignment 2 - Question 1

1.1

Let's import the gapminder dataset, dplyr, and tidyverse.

```
library(gapminder)
library(tidyverse)
suppressPackageStartupMessages(library(tidyverse))
```

Let's use the filter() function to extract 3 countries in the 1970s. The three countries I will use are Canada, France, and Germany.

```
new_gap <- gapminder %>%
            filter(country == "Canada" |
                   country == "France" |
                   country == "Germany") %>%
            filter(1969 < year & year <= 1980)

new_gap
```

```
## # A tibble: 6 x 6
##   country continent  year lifeExp      pop gdpPercap
##   <fct>   <fct>     <int>   <dbl>    <int>     <dbl>
## 1 Canada  Americas   1972    72.9 22284500    18971.
## 2 Canada  Americas   1977    74.2 23796400    22091.
## 3 France  Europe     1972    72.4 51732000    16107.
## 4 France  Europe     1977    73.8 53165019    18293.
## 5 Germany Europe     1972    71   78717088    18016.
## 6 Germany Europe     1977    72.5 78160773    20513.
```

As you can see, we've extracted 3 countries during the 1970s, called "reduced_gap".

1.2

Let's use pipe operator to select country and gdpPercap from the "reduced_gap" in 1.1.

```
new_gap %>%
  select(country, gdpPercap)
```

```
## # A tibble: 6 x 2
##   country gdpPercap
##   <fct>       <dbl>
## 1 Canada     18971.
## 2 Canada     22091.
## 3 France     16107.
```

```
## 4 France     18293.
## 5 Germany    18016.
## 6 Germany    20513.
```

1.3

Now we will filter gapminder to all entries that have experienced a drop in life expectancy, between 2007 and 1952.

```
gapminder %>%
  select(country, year, lifeExp) %>%
  group_by(country) %>%
  mutate(lagged_lifeExp = lag(lifeExp, 11)) %>%
  mutate(change_in_LE = lifeExp - lagged_lifeExp) %>%
  filter(change_in_LE < 0)
```

```
## # A tibble: 2 x 5
## # Groups:   country [2]
##   country    year lifeExp lagged_lifeExp change_in_LE
##   <fct>     <int>   <dbl>          <dbl>        <dbl>
## 1 Swaziland  2007    39.6           41.4        -1.79
## 2 Zimbabwe   2007    43.5           48.5        -4.96
```

After running all of this code, we have discovered that Swaziland and Zimbabwe have experienced an overall life expectancy decrease from 1952 to 2007.

1.4

We will filter gapminder so that it only shows the max GDP Per Capita of each country. We will group_by country, then we will make use of the slice() function, which will only keep the max value.

```
gapminder %>%
  select(country, gdpPercap) %>%
  group_by(country) %>%
  slice(which.max(gdpPercap))
```

```
## # A tibble: 142 x 2
## # Groups:   country [142]
##    country     gdpPercap
##    <fct>           <dbl>
##  1 Afghanistan      978.
##  2 Albania         5937.
##  3 Algeria         6223.
##  4 Angola          5523.
##  5 Argentina      12779.
##  6 Australia      34435.
##  7 Austria        36126.
##  8 Bahrain        29796.
##  9 Bangladesh      1391.
## 10 Belgium        33693.
## # ... with 132 more rows
```

This produces the correct table.

1.5

Produce a scatterplot of Canada's life expectancy vs. GDP per Cap, using ggplot2. That is to say, gdpPercap is on x axis, and lifeExp is on y axis. This shows us that as gdpPercap increases, so does Canadian lifeExp.

```
gapminder %>%
  filter(country == "Canada") %>%
  ggplot(aes(log(gdpPercap), lifeExp)) + geom_point(colour="blue")
```