

Homework 5

Sally Innis(34393119)

Exercise 1

The value of the “here” package? - Using the here::here package overrides the default behaviour of .R files that make the default working directory the top level of the project file. The here package lets you specify what level of the project file you are working in. A;sp, Here() let’s you see where you are working. The package allows for easier file sharing between users because it specifies file locations therefore helps reproducibility.

```
library(here)
```

```
## here() starts at /Users/sallyinnis/Documents/Term 1 2019/R Coding Course/stat545-hw-sallyinnis
here()
```

```
## [1] "/Users/sallyinnis/Documents/Term 1 2019/R Coding Course/stat545-hw-sallyinnis"
```

Exercise 2

Choose one dataset to explore by: I have chosen Gapminder as the dataset and country as the factor
Then explore the effects by 1) Comparing the results of arrange on the original and re-leveled factor and 2) Plotting a figure of before/after re-leveling the factor.

Exploration of Factor:

```
## Factor w/ 142 levels "Afghanistan",...: 1 1 1 1 1 1 1 1 1 1 ...
```

```
## [1] 1704
```

```
## [1] 142
```

```
## # A tibble: 142 x 1
```

```
##   value
```

```
##   <chr>
```

```
## 1 Afghanistan
```

```
## 2 Albania
```

```
## 3 Algeria
```

```
## 4 Angola
```

```
## 5 Argentina
```

```
## 6 Australia
```

```
## 7 Austria
```

```
## 8 Bahrain
```

```
## 9 Bangladesh
```

```
## 10 Belgium
```

```
## # ... with 132 more rows
```

Remove Observations associated with Oceania:

```
## [1] 142
```

```
## [1] 1680
```

```
## [1] 140
```

```
## # A tibble: 140 x 1
```

```
##   value
```

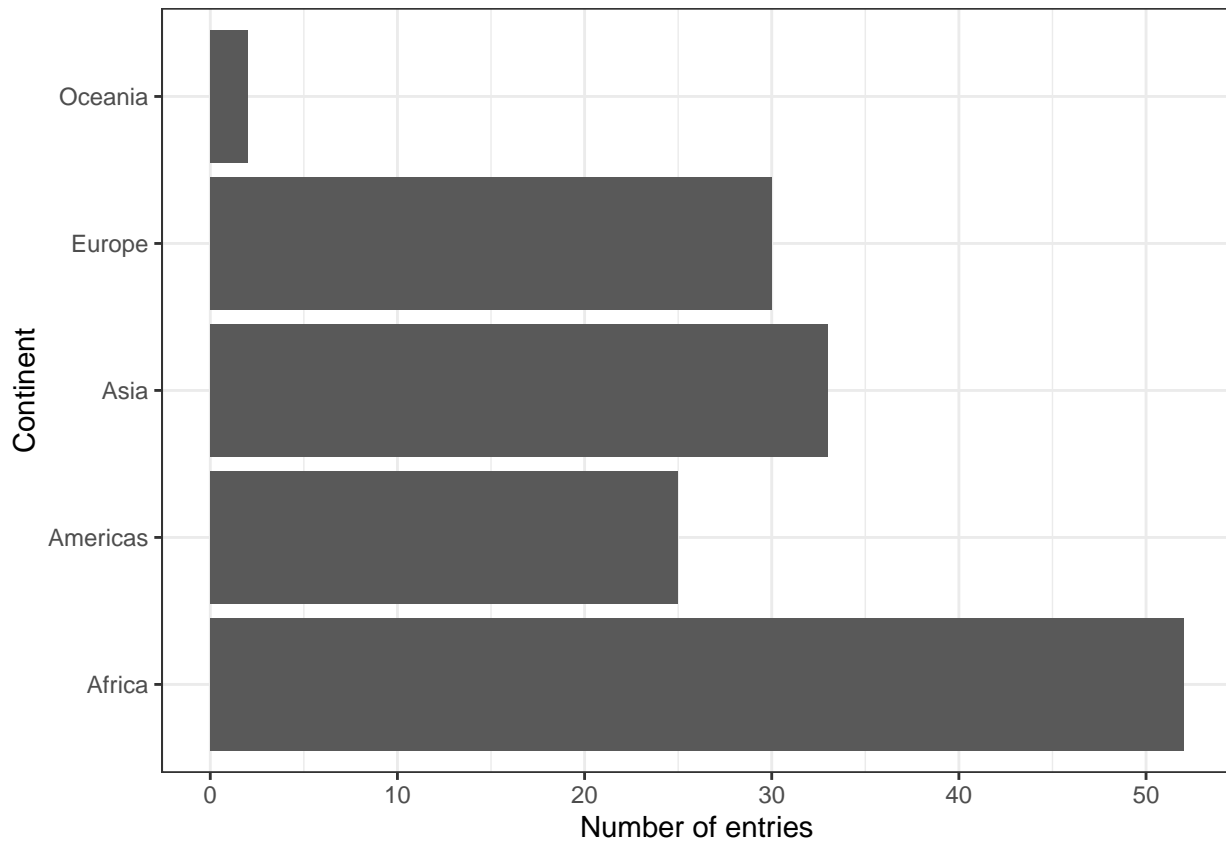
```
##      <chr>
## 1 Afghanistan
## 2 Albania
## 3 Algeria
## 4 Angola
## 5 Argentina
## 6 Austria
## 7 Bahrain
## 8 Bangladesh
## 9 Belgium
## 10 Benin
## # ... with 130 more rows
```

There are 2 less countries included in the evaluation of gapminder countries due to the dropping of Oceania as a continent.

2.Reorder levels based on knowledge from data.

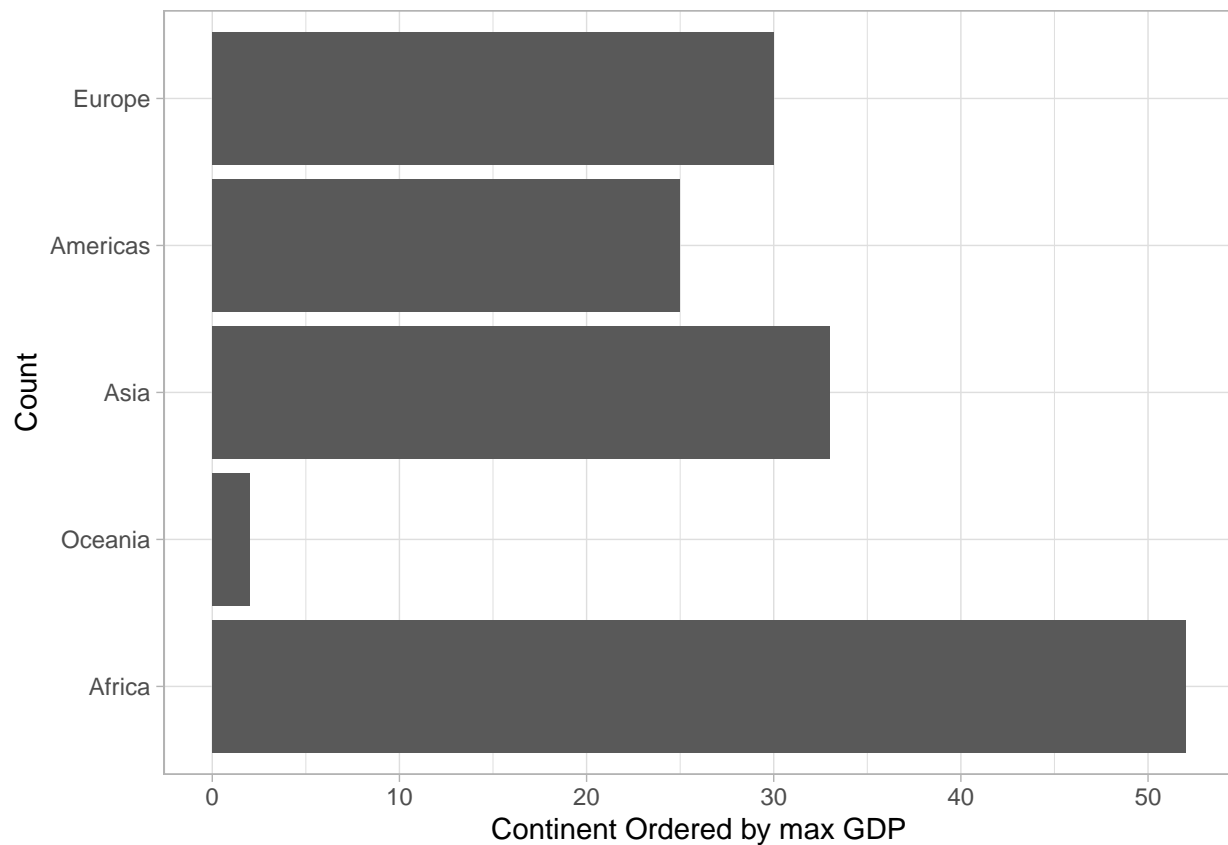
Below is the unordered plot of gapminder # of continents in 2002, as seen the plot has listed the continents in alphabetical order.

```
gapminder%>%
  filter(year == 2002) %>%
  ggplot() +
  geom_bar(aes(continent)) +
  coord_flip() +
  theme_bw() +
  ylab("Number of entries") + xlab("Continent")
```



I have chosen to reorder the gapminder dataset to display the continents in the order of maximum GDP in 2002 with Europe having the highest GDP at the top of the graph and Africa with the lowest GDP at the bottom.

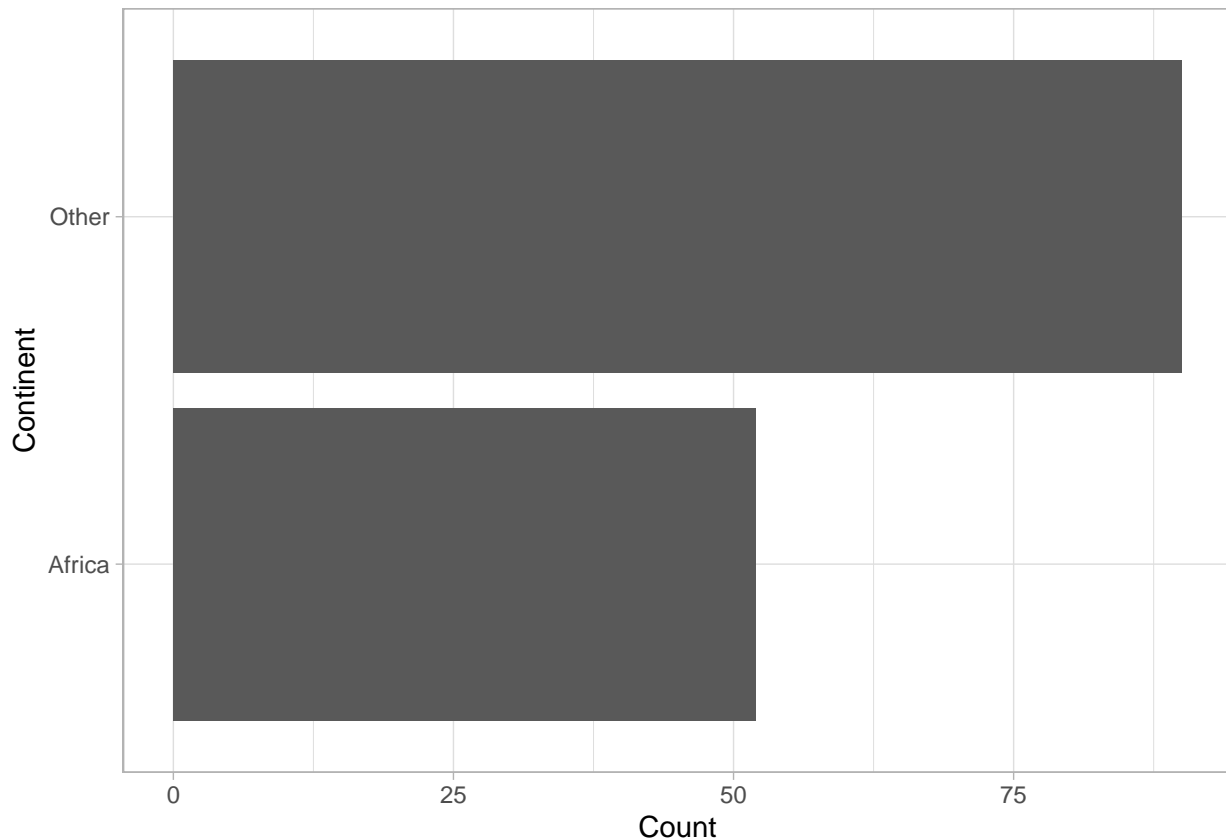
```
gapminder %>%  
  filter(year == 2002) %>%  
  ggplot(aes(fct_reorder(continent, gdpPercap, max))) +  
  geom_bar() +  
  coord_flip() +  
  xlab("Count") + ylab("Continent Ordered by max GDP") +  
  theme_light()
```



Exploration of forcats package:

`fct_lump` “lumps” all other continents together except for the numbers specified by `n`.

```
gapminder %>%  
  filter(year == 2002) %>%  
  ggplot(aes(fct_lump(continent, n = 1))) +  
  geom_bar() +  
  coord_flip() +  
  xlab("Continent") + ylab("Count") +  
  theme_light()
```



Exercise 3: File input/output

1) Export a grouped dataset to .csv

```
as_tibble(groupedgapminder <- gapminder %>%
  group_by(country) %>%
  summarize(Min_lifeExp = min(lifeExp), Min_gdpPercap = min(gdpPercap)))
```

```
## # A tibble: 142 x 3
##   country      Min_lifeExp Min_gdpPercap
##   <fct>         <dbl>         <dbl>
## 1 Afghanistan    28.8           635.
## 2 Albania         55.2          1601.
## 3 Algeria         43.1          2449.
## 4 Angola          30.0          2277.
## 5 Argentina       62.5          5911.
## 6 Australia       69.1         10040.
## 7 Austria         66.8          6137.
## 8 Bahrain         50.9          9867.
## 9 Bangladesh      37.5           630.
## 10 Belgium         68           8343.
## # ... with 132 more rows
```

```
write_csv(groupedgapminder, here::here("Homework5", "groupedgapminder.csv"))
```

2) With the imported data, play around with factor levels and use factors to order your data with one of your factors (i.e. non-alphabetically). For the I/O method(s) you chose, comment on whether or not your newly created file survived the round trip of writing to file then reading

back in.

I ordered countries by minimum life expectancy. The file "survived the round trip of writing to file and reading back in.

```
readgapminder <- read_csv(here::here("Homework5", "groupedgapminder.csv"))
```

```
## Parsed with column specification:
## cols(
##   country = col_character(),
##   Min_lifeExp = col_double(),
##   Min_gdpPercap = col_double()
## )
```

```
readgapminder %>%
  arrange(fct_reorder(country, Min_lifeExp, min))
```

```
## # A tibble: 142 x 3
##   country      Min_lifeExp Min_gdpPercap
##   <chr>          <dbl>         <dbl>
## 1 Rwanda          23.6           493.
## 2 Afghanistan     28.8           635.
## 3 Gambia           30            485.
## 4 Angola           30.0          2277.
## 5 Sierra Leone    30.3           575.
## 6 Cambodia         31.2           368.
## 7 Mozambique       31.3           390.
## 8 Burkina Faso     32.0           543.
## 9 Guinea-Bissau    32.5           300.
## 10 Yemen, Rep.     32.5           782.
## # ... with 132 more rows
```

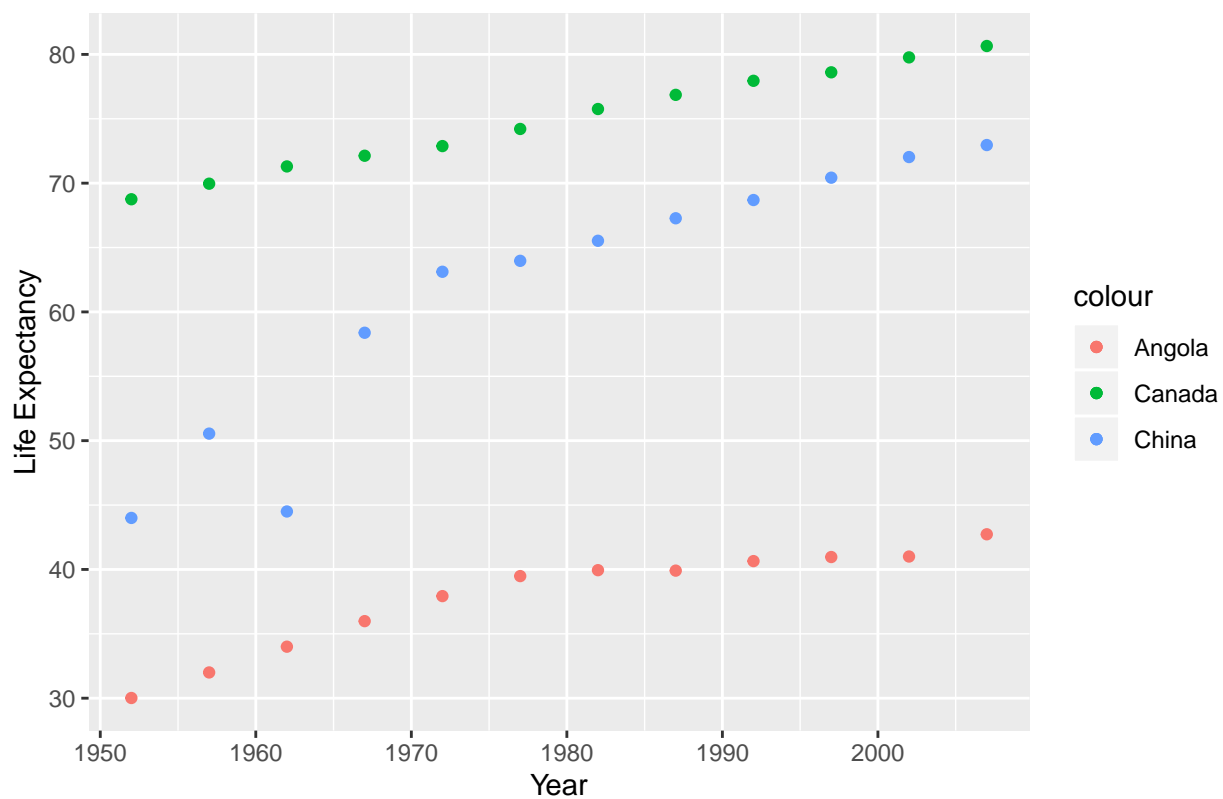
Exercise 4

For this exercise I took my original scatter plot from homework 4 and I tried to plot this as an percentage area plot. It was a suggested graph from the data to viz site. After I did this it was very clear to me that this graph did not fit the message the graphics should display. I think my original plot did a better job, however, I think the best change I could make to the graph is to simply change the graph from a scatter plot to a line graph. Also my legend is mislabeled as colour instead of country. I also added titles.

```
lifeExpdata = gapminder %>%
  filter(country == "Canada" | country == "China" | country == "Angola") %>%
  pivot_wider(id_cols = c(year),
    names_from = country,
    values_from = lifeExp)

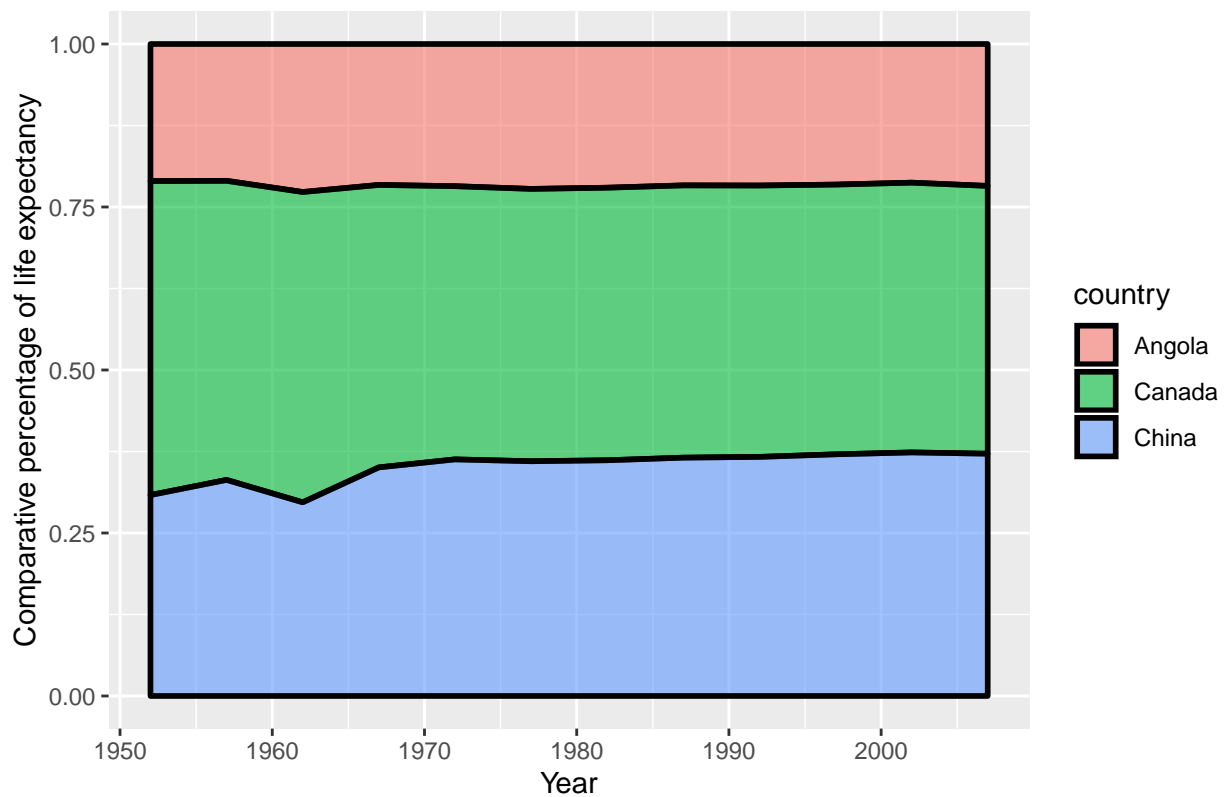
lifeExpdata %>%
  ggplot(aes(year, Angola, colour = "Angola")) +
  geom_point() +
  geom_point(aes(year, Canada, colour = "Canada")) +
  geom_point(aes(year, China, colour = "China")) +
  xlab('Year') +
  ylab('Life Expectancy') + #add title
  ggtitle("Original Scatter Plot from Homework 4")
```

Original Scatter Plot from Homework 4



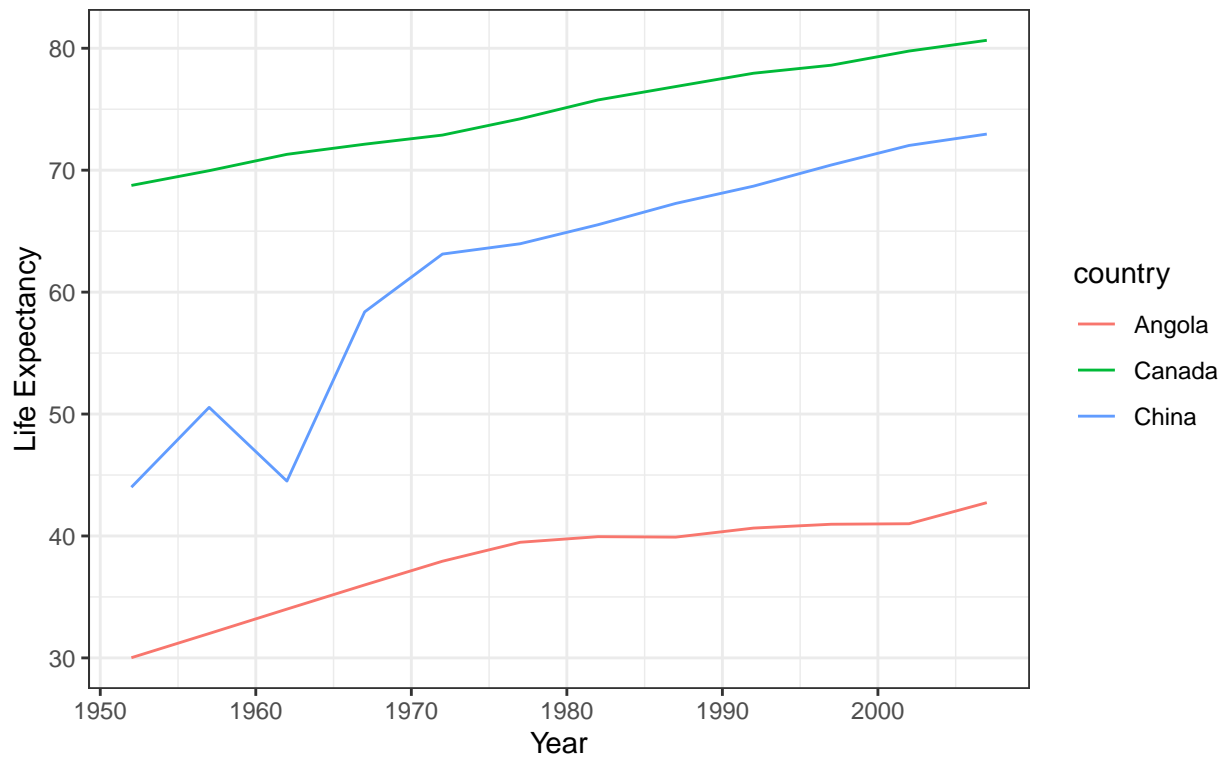
```
(newgap = gapminder %>%
  filter(country == "Canada" | country == "China" | country == "Angola") %>%
  group_by(year, country) %>%
  summarise(n = sum(lifeExp)) %>%
  mutate(percentage = n/sum(n)) %>%
  ggplot(aes(year, percentage, fill = country)) +
  geom_area(alpha=0.6 , size=1, colour="black") +
  ggtitle("Updated graph from Homework 4 using percent area calculations") +
  ylab('Comparative percentage of life expectancy') +
  xlab('Year'))
```

Updated graph from Homework 4 using percent area calculations



```
(updatedgapminder = gapminder %>%
  filter(country == "Canada" | country == "China" | country == "Angola") %>%
  ggplot(aes(year, lifeExp, colour = country)) +
  geom_line() +
  xlab('Year') +
  ylab('Life Expectancy') +
  ggtitle('Updated Homework 4 graph:
    Life Expectancy over time in Angola, Canada and China') +
  theme_bw())
```


Updated Homework 4 graph:
Life Expectancy over time in Angola, Canada and China



Exercise 5

Save the last plot displayed.

```
updatedgapminder = gapminder %>%  
  filter(country == "Canada" | country == "China" | country == "Angola") %>%  
  ggplot(aes(year, lifeExp, colour = country)) +  
    geom_line() +  
    xlab('Year') +  
    ylab('Life Expectancy') +  
    ggtitle('Updated Homework 4 graph:  
            Life Expectancy over time in Angola, Canada and China') +  
    theme_bw()  
ggsave("updatedgapminderplot.png")
```

Saving 6.5 x 4.5 in image

Updated Homework 4 graph:
Life Expectancy over time in Angola, Canada and China

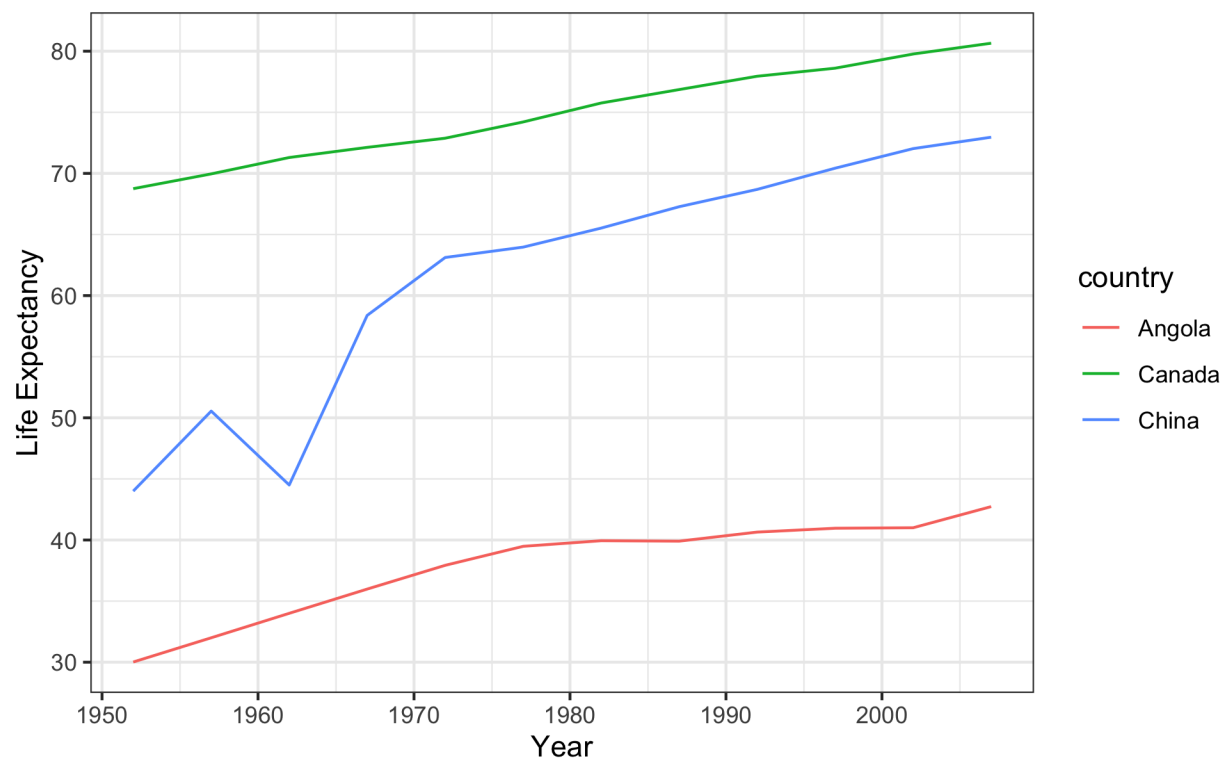


Figure 1: Loaded and Embedded Exercise 5 file