Assignment 2

Contents

Set Up

Load the necessary packages for this assignment:

```
library(gapminder)
library(tidyverse)
library(knitr)
```

Exercise 1: Basic `dplyr`

Exercise 1.1

Use `filter()` to subset the `gapminder` data to three countries of your choice in the 1970s.

```
filter(gapminder,
       year >= 1970 &
         year < 1980,
       country == "Canada" |
         country == "China" |
         country == "Japan") %>%
  kable()
```

country	continent	year	lifeExp	pop	gdpPercap
Canada	Americas	1972	72.88000	22284500	18970.5709
Canada	Americas	1977	74.21000	23796400	22090.8831
China	Asia	1972	63.11888	862030000	676.9001
China	Asia	1977	63.96736	943455000	741.2375
Japan	Asia	1972	73.42000	107188273	14778.7864
Japan	Asia	1977	75.38000	113872473	16610.3770

Exercise 1.2

Use the pipe operator `%>%` to select `country` and `gdpPercap` from your filtered dataset in 1.1.

```
filter(gapminder,
       year >= 1970 &
         year < 1980,
       country == "Canada" |
         country == "China" |
         country == "Japan") %>%
  select("country", "gdpPercap") %>%
  kable()
```

country	gdpPercap
Canada	18970.5709
Canada	22090.8831
China	676.9001
China	741.2375
Japan	14778.7864
Japan	16610.3770

Exercise 1.3

Filter gapminder to all entries that have experienced a drop in life expectancy. Be sure to include a new variable that's the increase in life expectancy in your tibble. Hint: you might find the `lag()` or `diff()` functions useful.

```
gapminder %>%
  arrange(year) %>%
  group_by(country) %>%
  mutate(lifeExpInc = lifeExp - lag(lifeExp)) %>%
  filter(lifeExpInc < 0) %>%
  kable()
```

country	continent	year	lifeExp	pop	gdpPercap	lifeExpInc
China	Asia	1962	44.50136	665770000	487.6740	-6.0476
Cambodia	Asia	1972	40.31700	7450606	421.6240	-5.0980
Czech Republic	Europe	1972	70.29000	9862158	13108.4536	-0.0900
Netherlands	Europe	1972	73.75000	13329874	18794.7457	-0.0700
Slovak Republic	Europe	1972	70.35000	4593433	9674.1676	-0.6300
Bulgaria	Europe	1977	70.81000	8797022	7612.2404	-0.0900
Cambodia	Asia	1977	31.22000	6978607	524.9722	-9.0970
El Salvador	Americas	1977	56.69600	4282586	5138.9224	-1.5110
Poland	Europe	1977	70.67000	34621254	9508.1415	-0.1800
Uganda	Africa	1977	50.35000	11457758	843.7331	-0.6660
Congo, Dem. Rep.	Africa	1982	47.78400	30646495	673.7478	-0.0200
Croatia	Europe	1982	70.46000	4413368	13221.8218	-0.1800
Denmark	Europe	1982	74.63000	5117810	21688.0405	-0.0600
El Salvador	Americas	1982	56.60400	4474873	4098.3442	-0.0920
Eritrea	Africa	1982	43.89000	2637297	524.8758	-0.6450
Hungary	Europe	1982	69.39000	10705535	12545.9907	-0.5600
Serbia	Europe	1982	70.16200	9032824	15181.0927	-0.1380
Uganda	Africa	1982	49.84900	12939400	682.2662	-0.5010
Angola	Africa	1987	39.90600	7874230	2430.2083	-0.0360
Congo, Dem. Rep.	Africa	1987	47.41200	35481645	672.7748	-0.3720
Norway	Europe	1987	75.89000	4186147	31540.9748	-0.0800
Poland	Europe	1987	70.98000	37740710	9082.3512	-0.3400

country	continent	year	lifeExp	pop	gdpPercap	lifeExpInc
Romania	Europe	1987	69.53000	22686371	9696.2733	-0.1300
Rwanda	Africa	1987	44.02000	6349365	847.9912	-2.1980
Zambia	Africa	1987	50.82100	7272406	1213.3151	-1.0000
Albania	Europe	1992	71.58100	3326498	2497.4379	-0.4190
Botswana	Africa	1992	62.74500	1342614	7954.1116	-0.8770
Bulgaria	Europe	1992	71.19000	8658506	6302.6234	-0.1500
Burundi	Africa	1992	44.73600	5809236	631.6999	-3.4750
Cameroon	Africa	1992	54.31400	12467171	1793.1633	-0.6710
Central African Republic	Africa	1992	49.39600	3265124	747.9055	-1.0890
Congo, Dem. Rep.	Africa	1992	45.54800	41672143	457.7192	-1.8640
Congo, Rep.	Africa	1992	56.43300	2409073	4016.2395	-1.0370
Cote d'Ivoire	Africa	1992	52.04400	12772596	1648.0738	-2.6110
Hungary	Europe	1992	69.17000	10348684	10535.6285	-0.4100
Iraq	Asia	1992	59.46100	17861905	3745.6407	-5.5830
Jamaica	Americas	1992	71.76600	2378618	7404.9237	-0.0040
Kenya	Africa	1992	59.28500	25020539	1341.9217	-0.0540
Korea, Dem. Rep.	Asia	1992	69.97800	20711375	3726.0635	-0.6690
Liberia	Africa	1992	40.80200	1912974	636.6229	-5.2250
Puerto Rico	Americas	1992	73.91100	3585176	14641.5871	-0.7190
Romania	Europe	1992	69.36000	22797027	6598.4099	-0.1700
Rwanda	Africa	1992	23.59900	7290203	737.0686	-20.4210
Sierra Leone	Africa	1992	38.33300	4260884	1068.6963	-1.6730
Somalia	Africa	1992	39.65800	6099799	926.9603	-4.8430
Tanzania	Africa	1992	50.44000	26605473	825.6825	-1.0950
Uganda	Africa	1992	48.82500	18252190	644.1708	-2.6840
Zambia	Africa	1992	46.10000	8381163	1210.8846	-4.7210
Zimbabwe	Africa	1992	60.37700	10704340	693.4208	-1.9740
Botswana	Africa	1997	52.55600	1536536	8647.1423	-10.1890
Bulgaria	Europe	1997	70.32000	8066057	5970.3888	-0.8700
Cameroon	Africa	1997	52.19900	14195809	1694.3375	-2.1150
Central African Republic	Africa	1997	46.06600	3696513	740.5063	-3.3300
Chad	Africa	1997	51.57300	7562011	1004.9614	-0.1510
Congo, Dem. Rep.	Africa	1997	42.58700	47798986	312.1884	-2.9610
Congo, Rep.	Africa	1997	52.96200	2800947	3484.1644	-3.4710
Cote d'Ivoire	Africa	1997	47.99100	14625967	1786.2654	-4.0530
Gabon	Africa	1997	60.46100	1126189	14722.8419	-0.9050
Iraq	Asia	1997	58.81100	20775703	3076.2398	-0.6500
Kenya	Africa	1997	54.40700	28263827	1360.4850	-4.8780
Korea, Dem. Rep.	Asia	1997	67.72700	21585105	1690.7568	-2.2510
Lesotho	Africa	1997	55.55800	1982823	1186.1480	-4.1270
Malawi	Africa	1997	47.49500	10419991	692.2758	-1.9250
Namibia	Africa	1997	58.90900	1774766	3899.5243	-3.0900
Nigeria	Africa	1997	47.46400	106207839	1624.9413	-0.0080
South Africa	Africa	1997	60.23600	42835005	7479.1882	-1.6520
Swaziland	Africa	1997	54.28900	1054486	3876.7685	-4.1850
Tanzania	Africa	1997	48.46600	30686889	789.1862	-1.9740
Trinidad and Tobago	Americas	1997	69.46500	1138101	8792.5731	-0.3970
Uganda	Africa	1997	44.57800	21210254	816.5591	-4.2470
Zambia	Africa	1997	40.23800	9417789	1071.3538	-5.8620
Zimbabwe	Africa	1997	46.80900	11404948	792.4500	-13.5680
Benin	Africa	2002	54.40600	7026113	1372.8779	-0.3710
Botswana	Africa	2002	46.63400	1630347	11003.6051	-5.9220

country	continent	year	lifeExp	pop	gdpPercap	lifeExpInc
Cameroon	Africa	2002	49.85600	15929988	1934.0114	-2.3430
Central African Republic	Africa	2002	43.30800	4048013	738.6906	-2.7580
Chad	Africa	2002	50.52500	8835739	1156.1819	-1.0480
Cote d'Ivoire	Africa	2002	46.83200	16252726	1648.8008	-1.1590
Gabon	Africa	2002	56.76100	1299304	12521.7139	-3.7000
Ghana	Africa	2002	58.45300	20550751	1111.9846	-0.1030
Iraq	Asia	2002	57.04600	24001816	4390.7173	-1.7650
Jamaica	Americas	2002	72.04700	2664659	6994.7749	-0.2150
Kenya	Africa	2002	50.99200	31386842	1287.5147	-3.4150
Korea, Dem. Rep.	Asia	2002	66.66200	22215365	1646.7582	-1.0650
Lesotho	Africa	2002	44.59300	2046772	1275.1846	-10.9650
Malawi	Africa	2002	45.00900	11824495	665.4231	-2.4860
Montenegro	Europe	2002	73.98100	720230	6557.1943	-1.4640
Mozambique	Africa	2002	44.02600	18473780	633.6179	-2.3180
Myanmar	Asia	2002	59.90800	45598081	611.0000	-0.4200
Namibia	Africa	2002	51.47900	1972153	4072.3248	-7.4300
Nigeria	Africa	2002	46.60800	119901274	1615.2864	-0.8560
South Africa	Africa	2002	53.36500	44433622	7710.9464	-6.8710
Swaziland	Africa	2002	43.86900	1130269	4128.1169	-10.4200
Togo	Africa	2002	57.56100	4977378	886.2206	-0.8290
Trinidad and Tobago	Americas	2002	68.97600	1101832	11460.6002	-0.4890
Zambia	Africa	2002	39.19300	10595811	1071.6139	-1.0450
Zimbabwe	Africa	2002	39.98900	11926563	672.0386	-6.8200
Gabon	Africa	2007	56.73500	1454867	13206.4845	-0.0260
Lesotho	Africa	2007	42.59200	2012649	1569.3314	-2.0010
Mozambique	Africa	2007	42.08200	19951656	823.6856	-1.9440
South Africa	Africa	2007	49.33900	43997828	9269.6578	-4.0260
Swaziland	Africa	2007	39.61300	1133066	4513.4806	-4.2560

Exercise 1.4

Filter gapminder so that it shows the max GDP per capita experienced by each country. Hint: you might find the `max()` function useful here.

```
gapminder %>%
    group_by(country) %>%
    summarize(maxGDP = max(gdpPercap)) %>%
    kable()
```

country	maxGDP
Afghanistan	978.0114
Albania	5937.0295
Algeria	6223.3675
Angola	5522.7764
Argentina	12779.3796
Australia	34435.3674
Austria	36126.4927
Bahrain	29796.0483
Bangladesh	1391.2538
Belgium	33692.6051
Benin	1441.2849
Bolivia	3822.1371

country	maxGDP
Bosnia and Herzegovina	7446.2988
Botswana	12569.8518
Brazil	9065.8008
Bulgaria	10680.7928
Burkina Faso	1217.0330
Burundi	631.6999
Cambodia	1713.7787
Cameroon	2602.6642
Canada	36319.2350
Central African Republic	1193.0688
Chad	1704.0637
Chile	13171.6388
China	4959.1149
Colombia	7006.5804
Comoros	1937.5777
Congo, Dem. Rep.	905.8602
Congo, Rep.	4879.5075
Costa Rica	9645.0614
Cote d'Ivoire	2602.7102
Croatia	14619.2227
Cuba	8948.1029
Czech Republic	22833.3085
Denmark	35278.4187
Djibouti	3694.2124
Dominican Republic	6025.3748
Ecuador	7429.4559
Egypt	5581.1810
El Salvador	5728.3535
Equatorial Guinea	12154.0897
Eritrea	913.4708
Ethiopia	690.8056
Finland	33207.0844
France	30470.0167
Gabon	21745.5733
Gambia	884.7553
Germany	32170.3744
Ghana	1327.6089
Greece	27538.4119
Guatemala	5186.0500
Guinea	945.5836
Guinea-Bissau	838.1240
Haiti	2011.1595
Honduras	3548.3308
Hong Kong, China	39724.9787
Hungary	18008.9444
Iceland	36180.7892
India	2452.2104
Indonesia	3540.6516
Iran	11888.5951
Iraq	14688.2351
Ireland	40675.9964
Israel	25523.2771

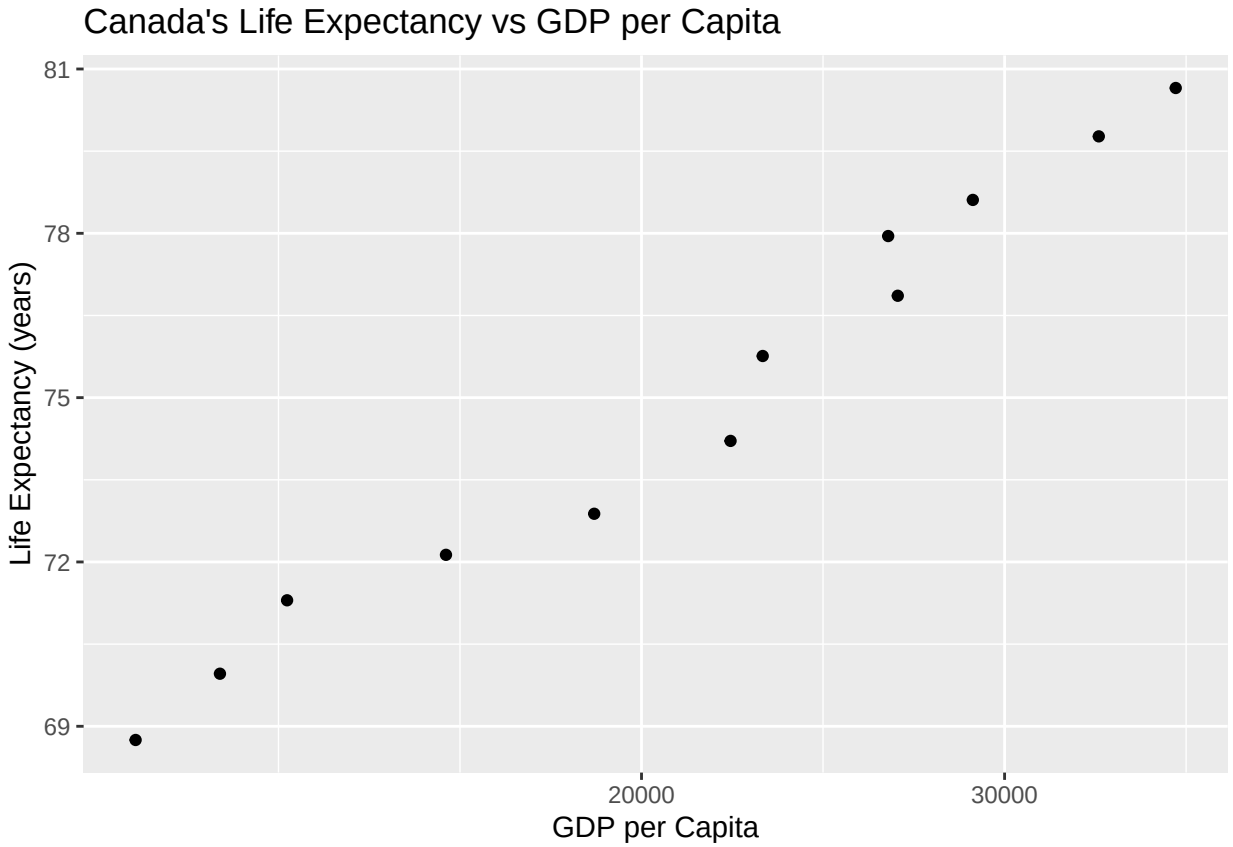
country	maxGDP
Italy	28569.7197
Jamaica	7433.8893
Japan	31656.0681
Jordan	4519.4612
Kenya	1463.2493
Korea, Dem. Rep.	4106.5253
Korea, Rep.	23348.1397
Kuwait	113523.1329
Lebanon	10461.0587
Lesotho	1569.3314
Liberia	803.0055
Libya	21951.2118
Madagascar	1748.5630
Malawi	759.3499
Malaysia	12451.6558
Mali	1042.5816
Mauritania	1803.1515
Mauritius	10956.9911
Mexico	11977.5750
Mongolia	3095.7723
Montenegro	11732.5102
Morocco	3820.1752
Mozambique	823.6856
Myanmar	944.0000
Namibia	4811.0604
Nepal	1091.3598
Netherlands	36797.9333
New Zealand	25185.0091
Nicaragua	5486.3711
Niger	1054.3849
Nigeria	2013.9773
Norway	49357.1902
Oman	22316.1929
Pakistan	2605.9476
Panama	9809.1856
Paraguay	4258.5036
Peru	7408.9056
Philippines	3190.4810
Poland	15389.9247
Portugal	20509.6478
Puerto Rico	19328.7090
Reunion	7670.1226
Romania	10808.4756
Rwanda	881.5706
Sao Tome and Principe	1890.2181
Saudi Arabia	34167.7626
Senegal	1712.4721
Serbia	15870.8785
Sierra Leone	1465.0108
Singapore	47143.1796
Slovak Republic	18678.3144
Slovenia	25768.2576

country	maxGDP
Somalia	1450.9925
South Africa	9269.6578
Spain	28821.0637
Sri Lanka	3970.0954
Sudan	2602.3950
Swaziland	4513.4806
Sweden	33859.7484
Switzerland	37506.4191
Syria	4184.5481
Taiwan	28718.2768
Tanzania	1107.4822
Thailand	7458.3963
Togo	1649.6602
Trinidad and Tobago	18008.5092
Tunisia	7092.9230
Turkey	8458.2764
Uganda	1056.3801
United Kingdom	33203.2613
United States	42951.6531
Uruguay	10611.4630
Venezuela	13143.9510
Vietnam	2441.5764
West Bank and Gaza	7110.6676
Yemen, Rep.	2280.7699
Zambia	1777.0773
Zimbabwe	799.3622

Exercise 1.5

Produce a scatterplot of Canada's life expectancy vs. GDP per capita using `ggplot2`, without defining a new variable. That is, after filtering the gapminder data set, pipe it directly into the `ggplot()` function. Ensure GDP per capita is on a log scale.

```
gapminder %>%
  filter(country == "Canada") %>%
  select(lifeExp, gdpPercap) %>%
  ggplot(aes(gdpPercap, lifeExp)) +
  geom_point() +
  scale_x_log10() +
  ggtitle("Canada's Life Expectancy vs GDP per Capita") +
  ylab("Life Expectancy (years)") +
  xlab("GDP per Capita")
```



Exercise 2: Explore individual variables with `dplyr`

Pick one categorical variable and one quantitative variable to explore. Answer the following questions in whichever way you think is appropriate, using `dplyr`:

What are possible values (or range, whichever is appropriate) of each variable?

- Categorical variable: `country`
- Possible values for the categorical variable are the categories itself. The range would be however many categories there are.
- There are 142 countries in this dataset.

```
gapminder %>%
  select(country) %>%
  unique() %>%
  nrow()
```

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

- Quantitative variable: `lifeExp`
- Possible values are any positive numbers (realistically <= 100). In the case of `lifeExp`, a range of 0 to the maximum life expectancy in the dataset (rounded to a whole number) would be an appropriate range.
- The highest life expectancy is 83 years in this dataset.

```
gapminder %>%
  select(lifeExp) %>%
  max() %>%
  round()
```

```
## [1] 83
```

What values are typical? What's the spread? What's the distribution?

Etc., tailored to the variable at hand. Feel free to use summary stats, tables, figures.

Typical Values

- A typical value for life expectancy is anywhere from 0 to 100 (realistically), but in this dataset the minimum is 23.60 years and the maximum is 82.60 years.
- A typical value for country is any of the 142 countries included in this dataset.

```
# Find the number of countries in this dataset:
gapminder %>%
  select(country) %>%
  unique() %>%
  nrow()
```

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

```
summary(gapminder) %>%
    kable()
```

country	continent	year	lifeExp	pop	gdpPercap
Afghanistan: 12	Africa :624	Min. :1952	Min. :23.60	Min. :6.001e+04	Min. : 241.2
Albania : 12	Americas:300	1st Qu.:1966	1st Qu.:48.20	1st Qu.:2.794e+06	1st Qu.: 1202.1
Algeria : 12	Asia :396	Median :1980	Median :60.71	Median :7.024e+06	Median : 3531.8
Angola : 12	Europe :360	Mean :1980	Mean :59.47	Mean :2.960e+07	Mean : 7215.3
Argentina : 12	Oceania : 24	3rd Qu.:1993	3rd Qu.:70.85	3rd Qu.:1.959e+07	3rd Qu.: 9325.5
Australia : 12	NA	Max. :2007	Max. :82.60	Max. :1.319e+09	Max. :113523.1
(Other) :1632	NA	NA	NA	NA	NA

- The interquartile range is 22.6475 years.

Spread: Range

- The range is defined as the difference between the highest and lowest values.
- The lowest life expectancy is 23.599 years, and the highest life expectancy is 82.603 years.
- The range is 59.004 years.

```
range <- gapminder %>%
    select(lifeExp) %>%
        range()

range[2] - range[1]
```

```
## [1] 59.004
```

Spread: Interquartile Range

- The first quartile is 48.20 years, the second quartile (the median) is 60.71 years, and the third quartile is 70.85 years.
- The interquartile range is the third minus the first quartile (which can also be calculated using the function `IQR()`):

```r
select(gapminder, lifeExp) %>%
    summary() %>%
    kable()
```

lifeExp
Min. :23.60
1st Qu.:48.20
Median :60.71
Mean :59.47
3rd Qu.:70.85
Max. :82.60

```r
# Finding the quartiles
(thirdq <- summary(gapminder$lifeExp)["3rd Qu."])
```

```
## 3rd Qu.
## 70.8455
```

```r
(firstq <- summary(gapminder$lifeExp)["1st Qu."])
```

```
## 1st Qu.
##  48.198
```

```r
# Using subtraction:
unname(thirdq - firstq)
```

```
## [1] 22.6475
```

```r
# Using IQR:
IQR(gapminder$lifeExp)
```

```
## [1] 22.6475
```

Spread: Variance

- The variance can be calculated using the function `var()`.
- The variance in life expectancy is 166.8517 years.

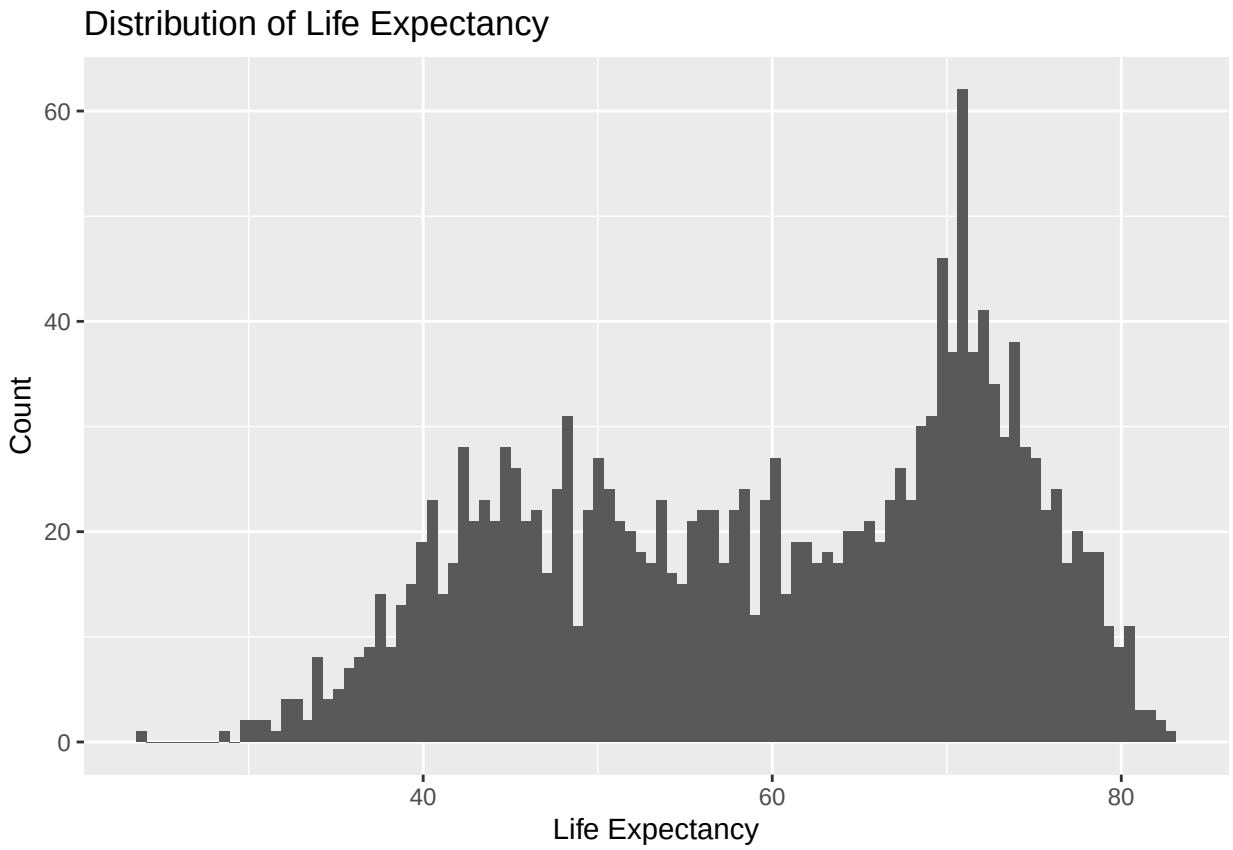
```r
gapminder %>%
    select(lifeExp) %>%
    var()
```

```
##          lifeExp
## lifeExp 166.8517
```

Distribution

- The distribution of life expectancy can be visualized using a histogram.

```r
ggplot(gapminder, aes(lifeExp)) +
  geom_histogram(bins=100) +
  ggtitle("Distribution of Life Expectancy") +
  xlab("Life Expectancy") +
  ylab("Count")
```



- The distribution of `country` can be determined using `table()`.
- In this case, there are 12 data points (i.e. rows) per country in this dataset.

```
gapminder %>%
  count(country) %>%
  kable()
```

country	n
Afghanistan	12
Albania	12
Algeria	12
Angola	12
Argentina	12
Australia	12
Austria	12
Bahrain	12
Bangladesh	12
Belgium	12
Benin	12
Bolivia	12
Bosnia and Herzegovina	12
Botswana	12
Brazil	12
Bulgaria	12
Burkina Faso	12
Burundi	12
Cambodia	12

country	n
Cameroon	12
Canada	12
Central African Republic	12
Chad	12
Chile	12
China	12
Colombia	12
Comoros	12
Congo, Dem. Rep.	12
Congo, Rep.	12
Costa Rica	12
Cote d'Ivoire	12
Croatia	12
Cuba	12
Czech Republic	12
Denmark	12
Djibouti	12
Dominican Republic	12
Ecuador	12
Egypt	12
El Salvador	12
Equatorial Guinea	12
Eritrea	12
Ethiopia	12
Finland	12
France	12
Gabon	12
Gambia	12
Germany	12
Ghana	12
Greece	12
Guatemala	12
Guinea	12
Guinea-Bissau	12
Haiti	12
Honduras	12
Hong Kong, China	12
Hungary	12
Iceland	12
India	12
Indonesia	12
Iran	12
Iraq	12
Ireland	12
Israel	12
Italy	12
Jamaica	12
Japan	12
Jordan	12
Kenya	12
Korea, Dem. Rep.	12
Korea, Rep.	12

country	n
Kuwait	12
Lebanon	12
Lesotho	12
Liberia	12
Libya	12
Madagascar	12
Malawi	12
Malaysia	12
Mali	12
Mauritania	12
Mauritius	12
Mexico	12
Mongolia	12
Montenegro	12
Morocco	12
Mozambique	12
Myanmar	12
Namibia	12
Nepal	12
Netherlands	12
New Zealand	12
Nicaragua	12
Niger	12
Nigeria	12
Norway	12
Oman	12
Pakistan	12
Panama	12
Paraguay	12
Peru	12
Philippines	12
Poland	12
Portugal	12
Puerto Rico	12
Reunion	12
Romania	12
Rwanda	12
Sao Tome and Principe	12
Saudi Arabia	12
Senegal	12
Serbia	12
Sierra Leone	12
Singapore	12
Slovak Republic	12
Slovenia	12
Somalia	12
South Africa	12
Spain	12
Sri Lanka	12
Sudan	12
Swaziland	12
Sweden	12

country	n
Switzerland	12
Syria	12
Taiwan	12
Tanzania	12
Thailand	12
Togo	12
Trinidad and Tobago	12
Tunisia	12
Turkey	12
Uganda	12
United Kingdom	12
United States	12
Uruguay	12
Venezuela	12
Vietnam	12
West Bank and Gaza	12
Yemen, Rep.	12
Zambia	12
Zimbabwe	12

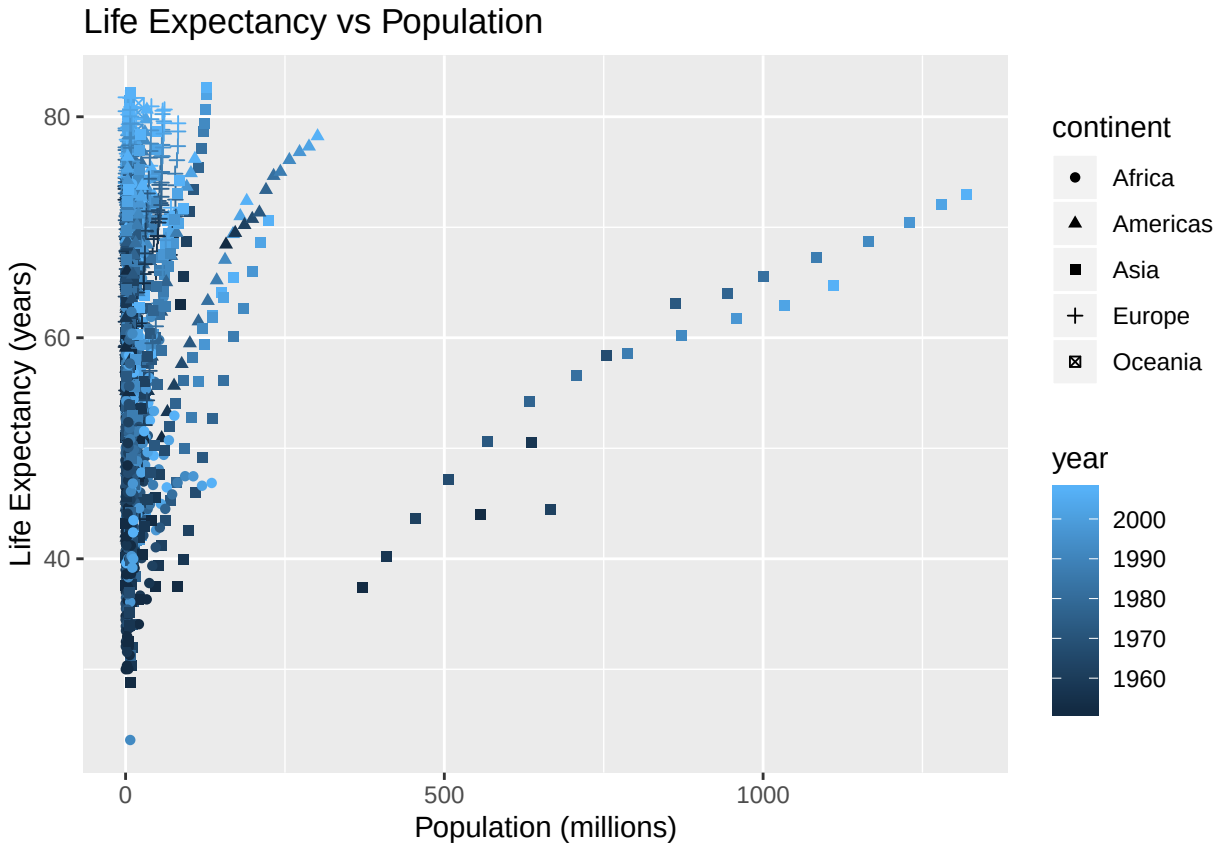
Exercise 3: Explore various plot types

Make two plots that have some value to them. That is, plots that someone might actually consider making for an analysis. Just don't make the same plots we made in class – feel free to use a data set from the `datasets` R package if you wish. You don't have to use all the data in every plot! It's fine to filter down to one country or a small handful of countries.

1. A scatterplot of two quantitative variables:

- The plot below shows that in Asia, the life expectancy has increased as the population increased, whereas most other continents have a steady increase in life expectancy while the population remains relatively the same.

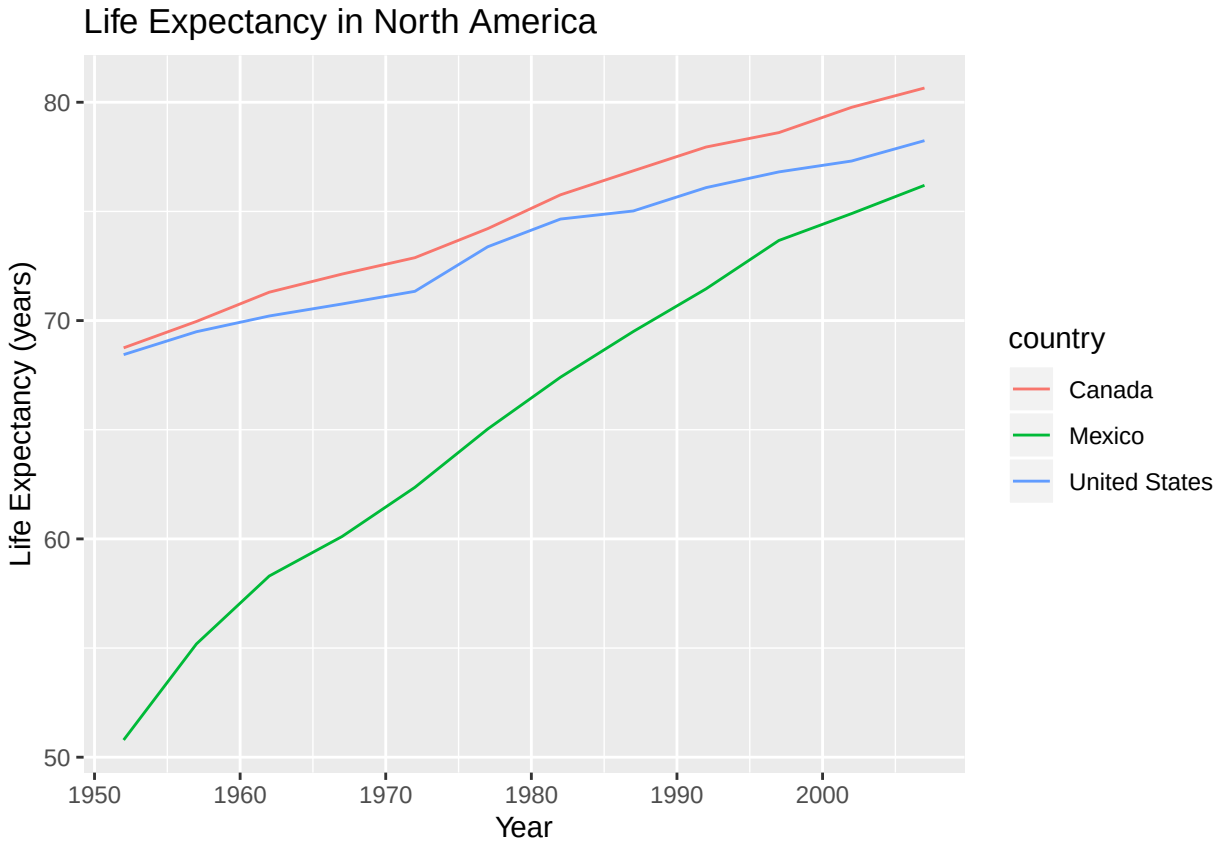
```
gapminder %>%
    mutate(million=pop/(10**6)) %>%
    ggplot(aes(x=million,y=lifeExp)) +
    geom_point(aes(shape = continent, colour = year)) +
    ggtitle("Life Expectancy vs Population") +
    xlab("Population (millions)") +
    ylab("Life Expectancy (years)")
```



2. One other plot besides a scatterplot:

- The plot below shows that in North America, the life expectancy has been increasing since 1952, where Mexico, in particular has experienced the greatest increase in life expectancy (a much steeper slope in the graph). Canada has a higher life expectancy than the United States, the rate of increase is very similar.

```
gapminder %>%
    filter(country == "Canada" | country == "United States" | country == "Mexico") %>%
    ggplot(aes(x = year, y = lifeExp)) +
    geom_line(aes(colour = country)) +
    ggtitle("Life Expectancy in North America") +
    xlab("Year") +
    ylab("Life Expectancy (years)")
```



Recycling (Optional)

Evaluate this code and describe the result. Presumably the analyst's intent was to get the data for Rwanda and Afghanistan. Did they succeed? Why or why not? If not, what is the correct way to do this?

```
filter(gapminder, country == c("Rwanda", "Afghanistan"))
```

```
# The Analyst's Way:
filter(gapminder, country == c("Rwanda", "Afghanistan")) %>%
    nrow()
```

```
## [1] 12
```

```
# The Correct Way:
filter(gapminder, country == "Rwanda" | country == "Afghanistan") %>%
    nrow()
```

```
## [1] 24
```

```
nrow(gapminder)
```

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

- As depicted above, the analyst's way is missing half the data (since it has half the number of rows than the correct method).
- The analyst did NOT succeed, because he or she used the vector in the filtering condition. This vector happens to be of length 2, whereas the gapminder dataset has 1704 rows. So during comparison, each country of each row is compared to this vector of length 2, on repeat.
- This process is known as 'recycling', where the vector of length 2 is actually recycled– meaning that in essence, the 1704 rows are being compared to 852 sequential copies of this vector of length 2.

- To demonstrate, I've subsampled the dataset to the first 6 rows using `head()`, and filtered using the analyst's way. As you can see, only 3 of the 6 rows have been filtered out (as observed above, where 50% of the data is missing).

```
head(gapminder) %>%
    filter(country == c("Rwanda", "Afghanistan")) %>%
    kable()
```

country	continent	year	lifeExp	pop	gdpPercap
Afghanistan	Asia	1957	30.332	9240934	820.8530
Afghanistan	Asia	1967	34.020	11537966	836.1971
Afghanistan	Asia	1977	38.438	14880372	786.1134

- This is due to the fact that for this reduced subsample, these are the exact comparisons being made, where only the TRUE ones are being filtered out as matching the criteria.

```
table(
    c(
        "Afghanistan" == "Rwanda",
        "Afghanistan" == "Afghanistan",
        "Afghanistan" == "Rwanda",
        "Afghanistan" == "Afghanistan",
        "Afghanistan" == "Rwanda",
        "Afghanistan" == "Afghanistan"
    )
) %>%
    kable(col.names= c("Boolean", "Frequency"))
```

Boolean	Frequency
FALSE	3
TRUE	3

- This is a result of an error in logic. The analyst most likely assumed that the comparison being made at each row of this reduce subsample would be:

```
table(
    c(
        "Afghanistan" == c("Rwanda", "Afghanistan"),
        "Afghanistan" == c("Rwanda", "Afghanistan"),
        "Afghanistan" == c("Rwanda", "Afghanistan"),
        "Afghanistan" == c("Rwanda", "Afghanistan"),
        "Afghanistan" == c("Rwanda", "Afghanistan"),
        "Afghanistan" == c("Rwanda", "Afghanistan")
    )
) %>%
    kable(col.names= c("Boolean", "Frequency"))
```

Boolean	Frequency
FALSE	6
TRUE	6

- The analyst did not consider the recycling that R does when comparing vectors are of different lengths.

Tibble display

Present numerical tables in a more attractive form using `knitr::kable()` for small tibbles (say, up to 10 rows), and `DT::datatable()` for larger tibbles.

- see above exercises as well
- `DT::datatable()` not used as this is a `LaTeX` output, not `html`.

```
kable(head(gapminder))
```

country	continent	year	lifeExp	pop	gdpPercap
Afghanistan	Asia	1952	28.801	8425333	779.4453
Afghanistan	Asia	1957	30.332	9240934	820.8530
Afghanistan	Asia	1962	31.997	10267083	853.1007
Afghanistan	Asia	1967	34.020	11537966	836.1971
Afghanistan	Asia	1972	36.088	13079460	739.9811
Afghanistan	Asia	1977	38.438	14880372	786.1134