Assignment 2

Linda Dumalo

24/09/2019

Exercise 1: Basic dplyr

Exercise 1.2

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

```
gapminder %>%
  filter(country == "Australia" | country == "Canada"| country == "Germany") %>%
  filter(between(year, 1970, 1980))
```

```
## # A tibble: 6 x 6
##   country   continent  year lifeExp      pop gdpPercap
##   <fct>     <fct>     <int>   <dbl>    <int>     <dbl>
## 1 Australia Oceania    1972    71.9 13177000    16789.
## 2 Australia Oceania    1977    73.5 14074100    18334.
## 3 Canada    Americas   1972    72.9 22284500    18971.
## 4 Canada    Americas   1977    74.2 23796400    22091.
## 5 Germany   Europe     1972    71   78717088    18016.
## 6 Germany   Europe     1977    72.5 78160773    20513.
```

Exercise 1.3

Filter gapminder to all entries that have experienced a drop in life expectancy.

delta_lifeExp is the change in life exptancy from 3 years prior to the the value in the year column.

```
gapminder %>%
  mutate(lifeExp, lag_lifeExp = lag(lifeExp)) %>%
  mutate(lifeExp, delta_lifeExp = lifeExp - lag_lifeExp) %>%
  select(country, year, delta_lifeExp) %>%
  drop_na(delta_lifeExp)
```

```
## # A tibble: 1,703 x 3
##    country      year delta_lifeExp
##    <fct>       <int>         <dbl>
##  1 Afghanistan  1957        1.53
##  2 Afghanistan  1962        1.66
##  3 Afghanistan  1967        2.02
##  4 Afghanistan  1972        2.07
##  5 Afghanistan  1977        2.35
##  6 Afghanistan  1982        1.42
##  7 Afghanistan  1987        0.968
##  8 Afghanistan  1992        0.852
##  9 Afghanistan  1997        0.0890
## 10 Afghanistan  2002        0.366
## # ... with 1,693 more rows
```

The data was then sorted to display the countries that experienced the largest drop in life expectancy from 2002 to 2007 at the top of the table.

```
gapminder %>%
  mutate(lifeExp, lag_lifeExp = lag(lifeExp)) %>%
  mutate(lifeExp, delta_lifeExp = lifeExp - lag_lifeExp) %>%
  filter(delta_lifeExp < 0) %>%
  arrange(-desc(delta_lifeExp)) %>%
  select(country, year, delta_lifeExp) %>%
  drop_na(delta_lifeExp)
```

```
## # A tibble: 221 x 3
##    country                   year delta_lifeExp
##    <fct>                    <int>         <dbl>
##  1 Central African Republic  1952         -45.2
##  2 Somalia                   1952         -44.9
##  3 India                     1952         -44.4
##  4 Sierra Leone              1952         -43.7
##  5 Gabon                     1952         -43.7
##  6 Djibouti                  1952         -43.5
##  7 Oman                      1952         -42.6
##  8 Angola                    1952         -42.3
##  9 Benin                     1952         -41.2
## 10 Burkina Faso              1952         -41.0
## # ... with 211 more rows
```

```
gapminder %>%
  mutate(lifeExp, lag_lifeExp = lag(lifeExp)) %>%
  mutate(lifeExp, delta_lifeExp = lifeExp - lag_lifeExp) %>%
   filter(year == 2007) %>%
  arrange(-desc(delta_lifeExp)) %>%
  select(country,delta_lifeExp) %>%
  drop_na(delta_lifeExp) %>%
  rename(delta_lifeExp_2002_2007 = delta_lifeExp)
```

```
## # A tibble: 142 x 2
##    country      delta_lifeExp_2002_2007
##    <fct>                          <dbl>
##  1 Swaziland                     -4.26
##  2 South Africa                  -4.03
##  3 Lesotho                       -2.00
##  4 Mozambique                    -1.94
##  5 Gabon                         -0.026
##  6 Chad                           0.126
##  7 Nigeria                        0.251
##  8 Italy                          0.306
##  9 Jamaica                        0.520
## 10 Montenegro                     0.562
## # ... with 132 more rows
```

Exercise 1.4

Filter gapminder so that it shows the max GDP per capita experienced by each country.

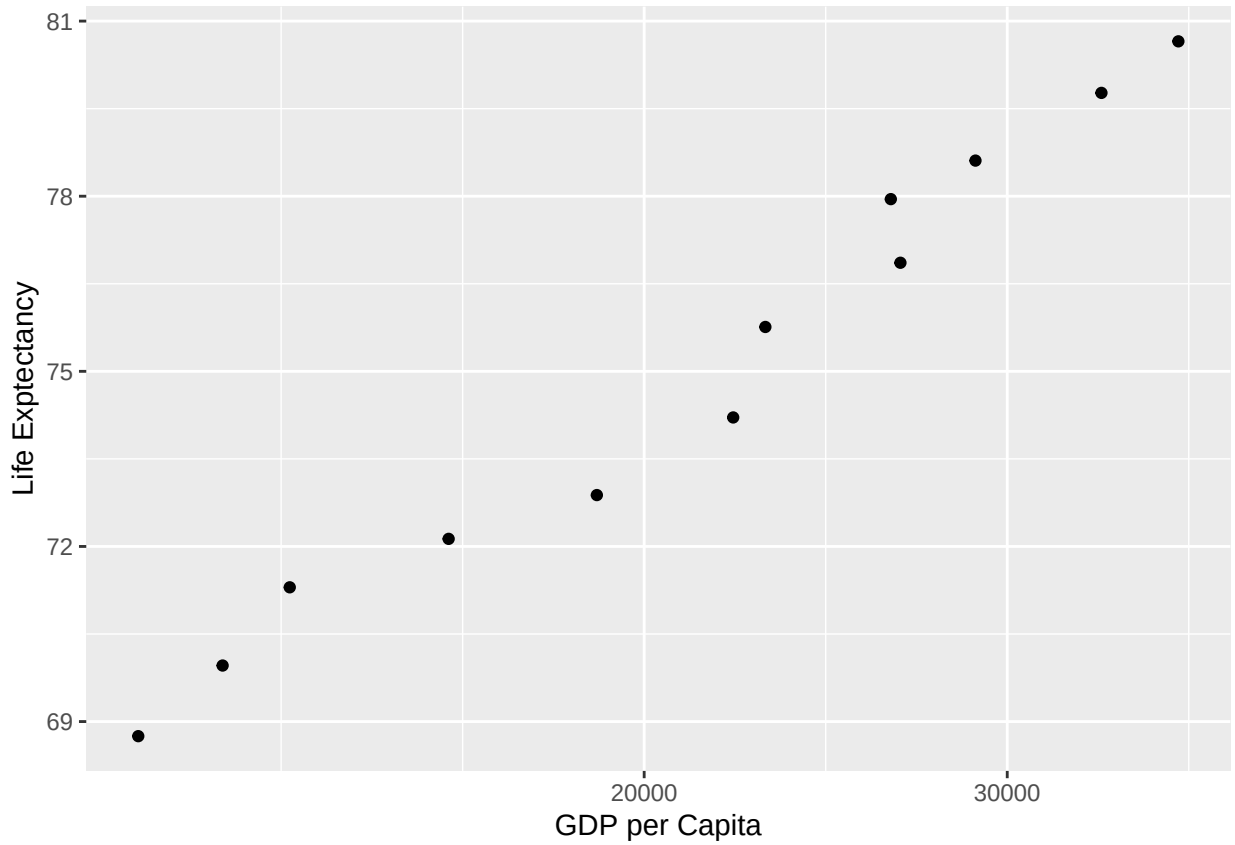
```r
gapminder %>%
  group_by(country) %>%
  filter(gdpPercap == max(gdpPercap)) %>%
  select(country, year, gdpPercap) %>%
  rename(Max_GDPperCap = gdpPercap)
```

```
## # A tibble: 142 x 3
## # Groups:   country [142]
##    country      year Max_GDPperCap
##    <fct>       <int>         <dbl>
##  1 Afghanistan  1982          978.
##  2 Albania      2007         5937.
##  3 Algeria      2007         6223.
##  4 Angola       1967         5523.
##  5 Argentina    2007        12779.
##  6 Australia    2007        34435.
##  7 Austria      2007        36126.
##  8 Bahrain      2007        29796.
##  9 Bangladesh   2007         1391.
## 10 Belgium      2007        33693.
## # ... with 132 more rows
```

Exercise 1.5

Produce a scatterplot of Canada's life expectancy vs. GDP per capita using ggplot()

```r
gapminder %>%
  filter(country == "Canada") %>%
  ggplot(aes(gdpPercap, lifeExp)) +
  geom_point() +
  scale_x_log10() +
  xlab("GDP per Capita") +
  ylab("Life Exptectancy")
```




```
  # theme_bw was not working!
  # wanted to add labels corresonding to the year for each point
```

Exercise 2

Categorical Variable Exploration

Exploring the GDP per capita (gdpPercap) for the countries in Asia for the year 2007. The GDP per capita is stated in USD.

Maximum GDP Per Capita

```
Maxgdp <- gapminder %>%  # was, trying to show the higest and the lowest gdppercap in 1 table, realize
  filter(continent == 'Asia') %>%
  filter(year == 2007) %>%
  summarize(max_gdpPercap = max(gdpPercap), country = country[gdpPercap == max_gdpPercap]) %>%
  select(country, max_gdpPercap) %>%
  top_n(1)
```

```
## Selecting by max_gdpPercap
```

Minimum GDP Per Capita

```
Mingdp <- gapminder %>%  # was, trying to show the higest and the lowest gdppercap
  select(country, continent, year, gdpPercap) %>%
  filter(continent == 'Asia') %>%
  filter(year == 2007) %>%
  summarize(min_gdpPercap = min(gdpPercap), country = country[gdpPercap == min_gdpPercap]) %>%
  select(country, min_gdpPercap) %>%
  top_n(-1)
```

```
## Selecting by min_gdpPercap
```

Minimum and Maximum GDP per capita in 2007 in Asia

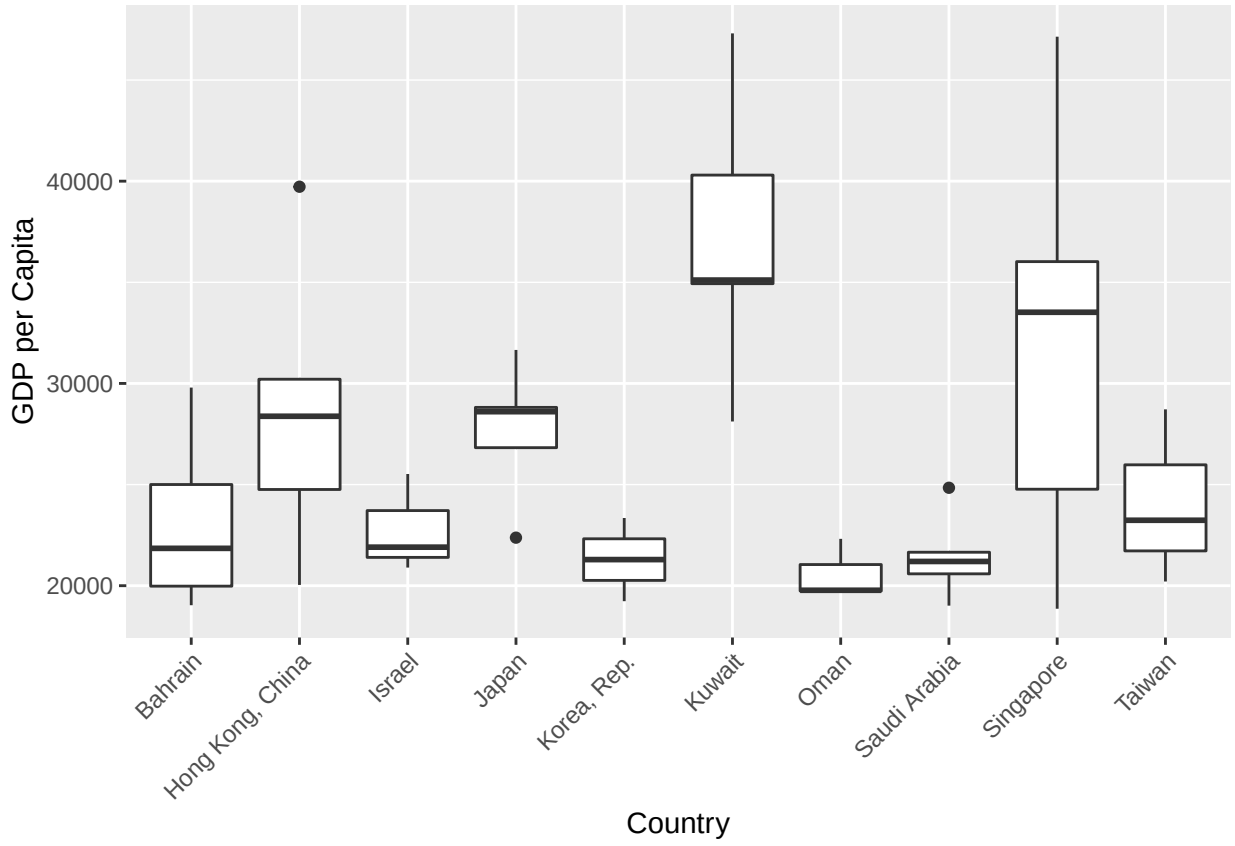
Minimum GDP per capitat in 2007 was **Myanmar** at $ 944 USD and the maximum GDP per capita was **Kuwait** with $ 47307.

Ideally, would have plotted these into a bar graph or something

Exercise 3: Explore various plot type

Instead of focusing on the year 2007, the GDP per capita in the countries of Asia will be analyzed from 1987 to 2007 For the countries with the top 10 GDP

```
gapminder %>%
  filter(continent == 'Asia') %>%
  filter(between(year,1987,2007)) %>%
  group_by(country) %>%
  group_by(gdpPercap) %>%
  arrange(desc(gdpPercap)) %>%
  ungroup() %>%
  slice(1:40) %>% # trying to narrow down to the countries with the top 10 highest GDP per Capita
  ggplot(aes(country, gdpPercap)) +
  geom_boxplot() +
  xlab("Country") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  # theme(axis.text.x = element_text(angle = 45, hjust = 1))
  ylab("GDP per Capita")
```




```
  # not pretty...
```

Exploring a Dfferent dataset

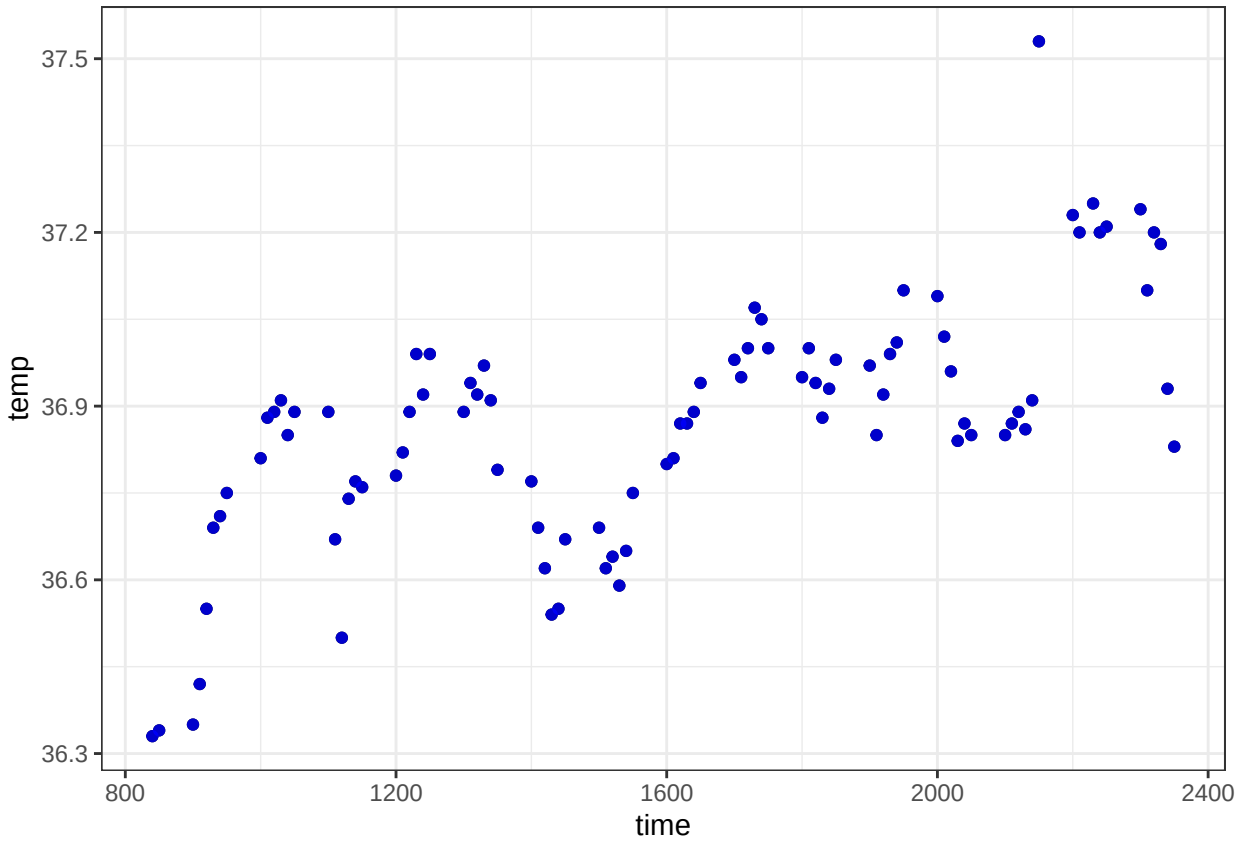
Source for description of data: https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/beavers.html

Description of Dataframe

Reynolds (1994) describes a small part of a study of the long-term temperature dynamics of beaver *Castor canadensis* in north-central Wisconsin. Body temperature was measured by telemetry every 10 minutes for females. The data used corresponds to 1 day in December, starting at 8:40 am (840) to 23:50 (2350). Temperatures are in degree celsius.

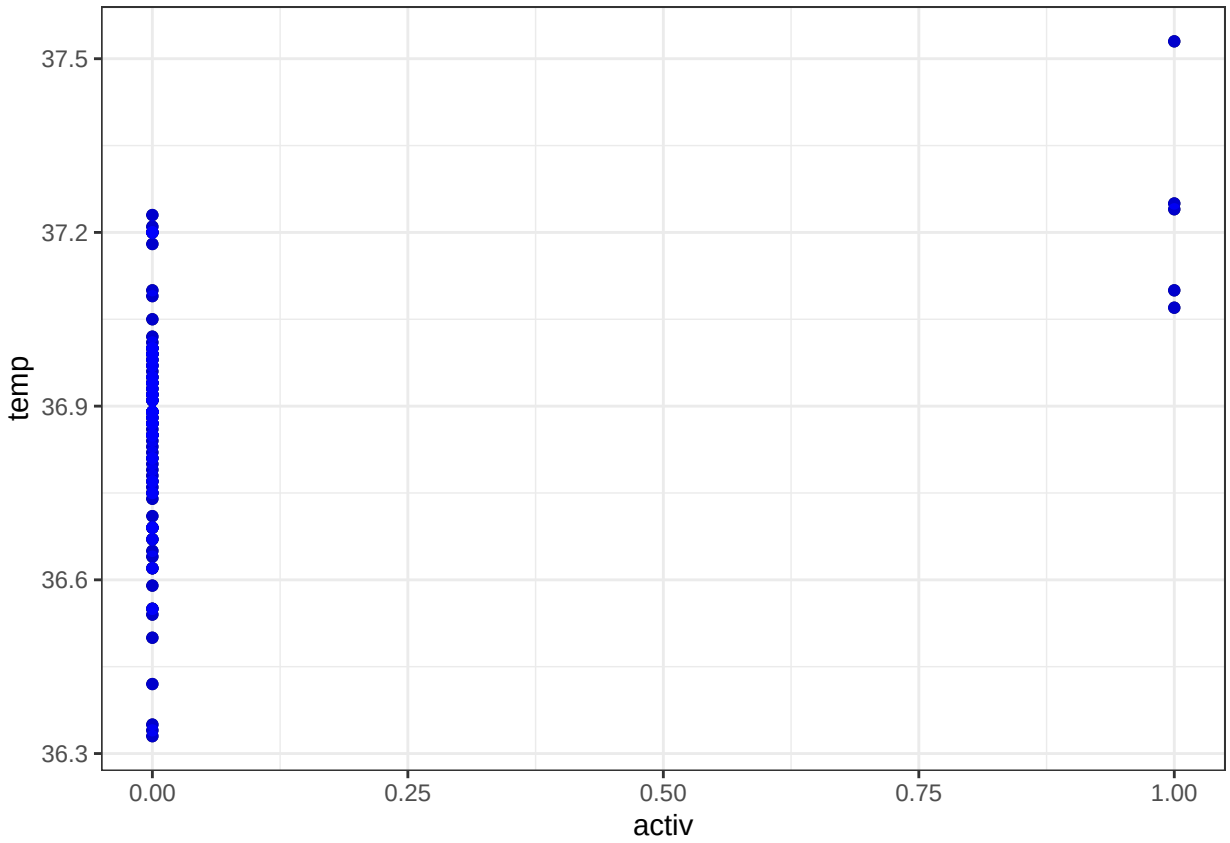
```
beaverday <- data.frame(beaver1) %>%
  slice(1:91)
```

```
ggplot(beaverday, aes(time, temp)) +
  geom_point(aes(time, temp)) +
  geom_point(colour = "blue", alpha = 0.8) +
  theme_bw()
```



Scatter plot to show the relationship between inactivity (0) and activity (1).

```
ggplot(beaverday, aes(activ, temp)) +
  geom_point(aes(activ, temp)) +
  geom_point(colour = "blue", alpha = 0.8) +
  theme_bw()
```



Bonus

Original code:

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

```
## # A tibble: 12 x 6
##    country     continent  year lifeExp      pop gdpPercap
##    <fct>       <fct>     <int>   <dbl>    <int>     <dbl>
##  1 Afghanistan Asia       1957    30.3  9240934      821.
##  2 Afghanistan Asia       1967    34.0 11537966      836.
##  3 Afghanistan Asia       1977    38.4 14880372      786.
##  4 Afghanistan Asia       1987    40.8 13867957      852.
##  5 Afghanistan Asia       1997    41.8 22227415      635.
##  6 Afghanistan Asia       2007    43.8 31889923      975.
##  7 Rwanda      Africa     1952    40    2534927      493.
##  8 Rwanda      Africa     1962    43    3051242      597.
##  9 Rwanda      Africa     1972    44.6  3992121      591.
## 10 Rwanda      Africa     1982    46.2  5507565      882.
## 11 Rwanda      Africa     1992    23.6  7290203      737.
## 12 Rwanda      Africa     2002    43.4  7852401      786.
```

Presumably the analyst's intent was to get the data for Rwanda and Afghanistan however it appears that there are some rows missing.

```r
gapminder %>%
    filter(country == "Afghanistan" | country == "Rwanda")
```

```
## # A tibble: 24 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.
##  7 Afghanistan Asia       1982    39.9 12881816      978.
##  8 Afghanistan Asia       1987    40.8 13867957      852.
##  9 Afghanistan Asia       1992    41.7 16317921      649.
## 10 Afghanistan Asia       1997    41.8 22227415      635.
## # ... with 14 more rows
```