Lab 3

Otobo Ujah

2/15/2022

Ex 1.0: Loading libraries

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(gapminder)
```

Ex 1.1 : Three countries in 1970's

```
gapminder |>
  filter(country == "Nigeria"| country == "Ghana"| country == "Senegal") |>
  filter(year >= 1970 & year <=1979)
```

```
## # A tibble: 6 x 6
##   country continent  year lifeExp      pop gdpPercap
##   <fct>   <fct>     <int>   <dbl>    <int>     <dbl>
## 1 Ghana   Africa     1972    49.9  9354120     1178.
## 2 Ghana   Africa     1977    51.8 10538093      993.
## 3 Nigeria Africa     1972    42.8 53740085     1698.
## 4 Nigeria Africa     1977    44.5 62209173     1982.
## 5 Senegal Africa     1972    45.8  4588696     1598.
## 6 Senegal Africa     1977    48.9  5260855     1562.
```

Ex 1.2 : Three countries in 1970's with respective GDP Per Capita

```
gapminder |> filter(country == "Nigeria"| country == "Ghana"| country == "Senegal") |>
  filter(year >= 1970 & year <=1979) |>
  select(country, gdpPercap)
```

```
## # A tibble: 6 x 2
##   country gdpPercap
##   <fct>       <dbl>
## 1 Ghana       1178.
## 2 Ghana        993.
## 3 Nigeria     1698.
## 4 Nigeria     1982.
## 5 Senegal     1598.
## 6 Senegal     1562.
```

Ex 1.3: Changes in life expectancy by country

```
lifechanged <-  gapminder|>
  arrange(country) |>
  mutate(lifeChange = lifeExp - lag(lifeExp))  |> filter(lifeChange < 0)
```

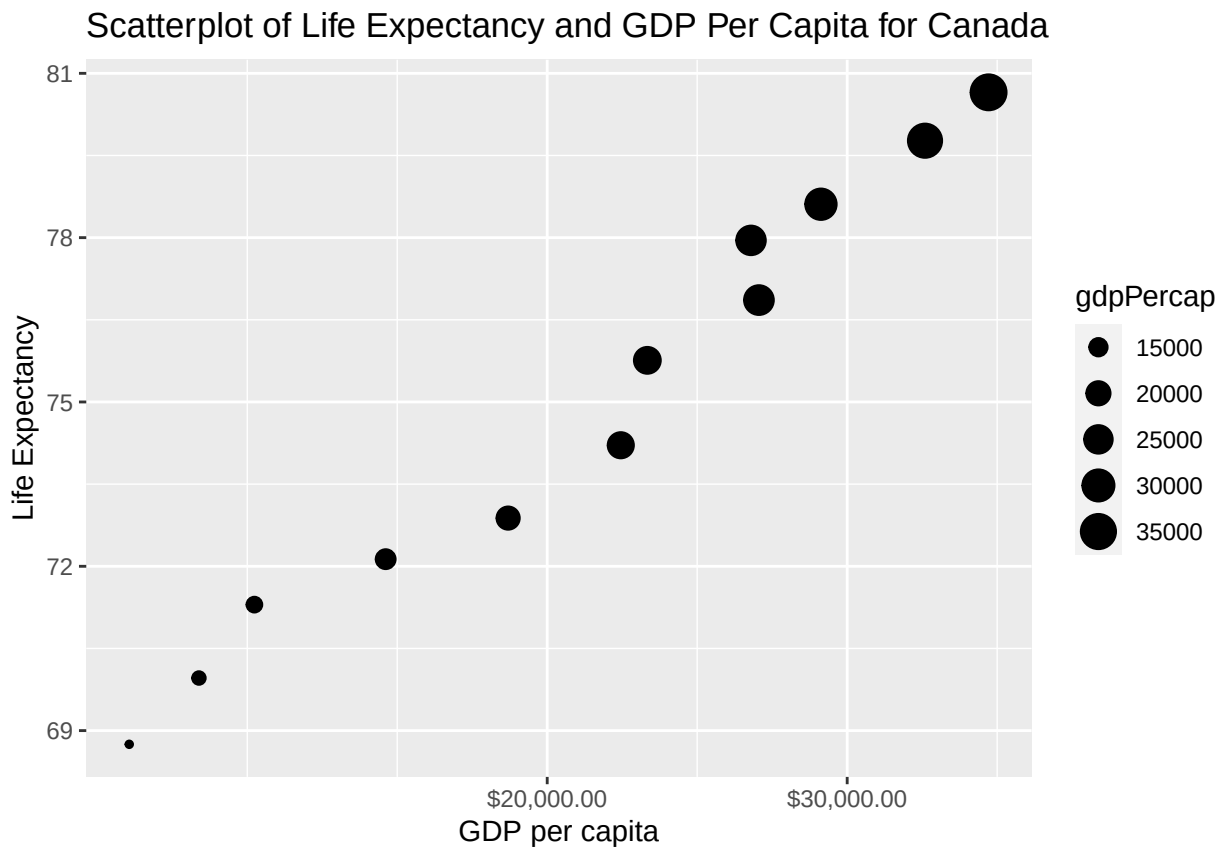
Ex 1.4: Maximum GDP per capita experienced by each country

```
gapminder |>
  arrange(country) |>
  group_by(country) |>
  filter(gdpPercap==max(gdpPercap))
```

```
## # A tibble: 142 x 6
## # Groups:   country [142]
##    country     continent  year lifeExp       pop gdpPercap
##    <fct>       <fct>     <int>   <dbl>     <int>     <dbl>
##  1 Afghanistan Asia       1982    39.9  12881816      978.
##  2 Albania     Europe     2007    76.4   3600523     5937.
##  3 Algeria     Africa     2007    72.3  33333216     6223.
##  4 Angola      Africa     1967    36.0   5247469     5523.
##  5 Argentina   Americas   2007    75.3  40301927    12779.
##  6 Australia   Oceania    2007    81.2  20434176    34435.
##  7 Austria     Europe     2007    79.8   8199783    36126.
##  8 Bahrain     Asia       2007    75.6    708573    29796.
##  9 Bangladesh  Asia       2007    64.1 150448339     1391.
## 10 Belgium     Europe     2007    79.4  10392226    33693.
## # ... with 132 more rows
```

Ex 1.5 Scatterplot of Canada's life expectance vs GDP Per Capita

```
gapminder |>filter(country=="Canada") |>
  ggplot(aes(y = lifeExp, x = gdpPercap, size = gdpPercap)) +
  geom_point() +
  scale_x_continuous(
    name = "GDP per capita",
    trans = "log10",
    labels = scales::dollar_format()) +
  labs(y = "Life Expectancy",
       title = "Scatterplot of Life Expectancy and GDP Per Capita for Canada")
```



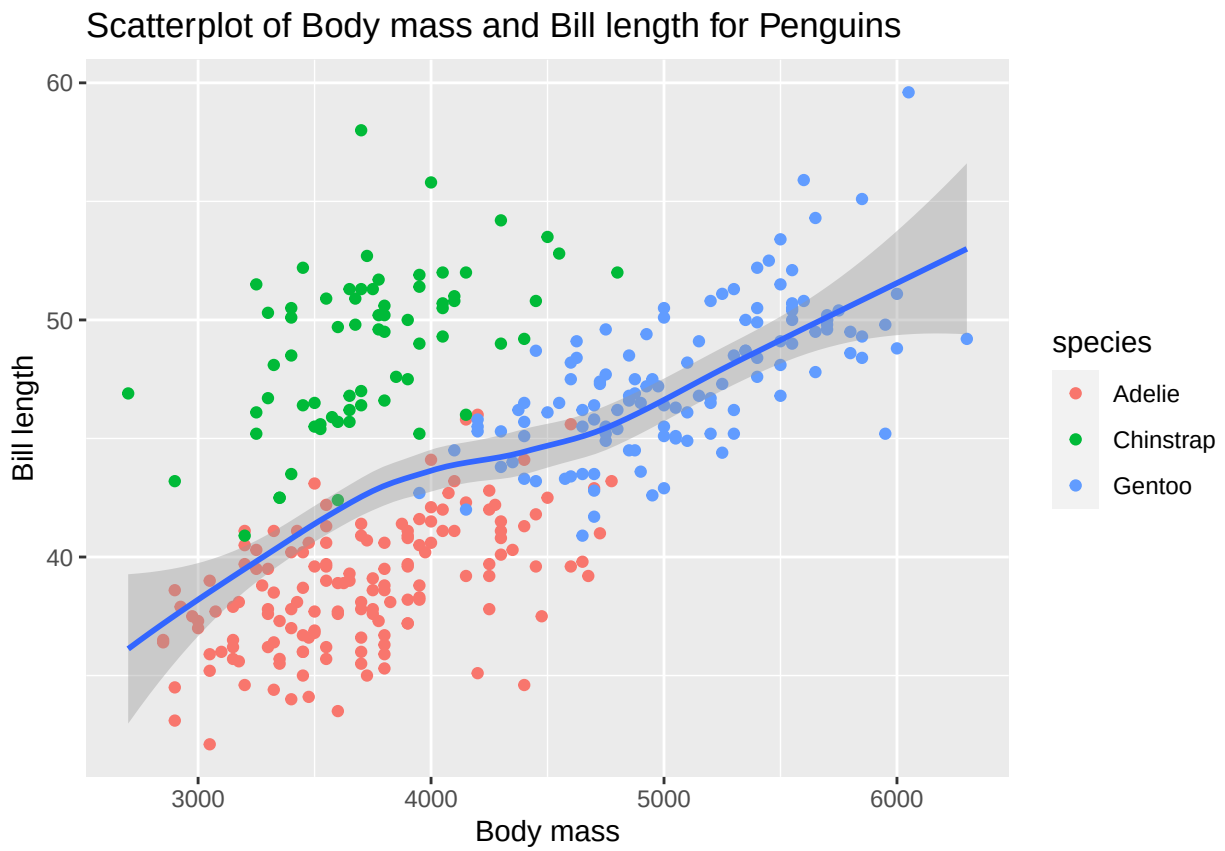
Ex 2.1 Exploring Palmerpenguins using dplyr and ggplot

```
library(palmerpenguins)

penguins |>
  group_by(species) |>
  summarise(bill_length_mean = mean(bill_length_mm, na.rm = T),
            body_mass_mean = mean(body_mass_g, na.rm = T),
            bill_length_sd = sd(bill_length_mm, na.rm = T),
            body_mass_sd = sd(body_mass_g, na.rm = T))
```

```
## # A tibble: 3 x 5
##   species   bill_length_mean body_mass_mean bill_length_sd body_mass_sd
##   <fct>                <dbl>          <dbl>          <dbl>        <dbl>
## 1 Adelie                38.8          3701.           2.66         459.
## 2 Chinstrap             48.8          3733.           3.34         384.
## 3 Gentoo                47.5          5076.           3.08         504.
```

```
penguins |> ggplot(aes( x =  body_mass_g, y=  bill_length_mm)) +
  geom_point(aes(color =species))+
  geom_smooth() +
  labs(x = "Body mass",
       y = "Bill length",
       title = "Scatterplot of Body mass and Bill length for Penguins")
```



Ex 2.2: Exploring Categorical and Continous variable

```
penguins |>
  group_by(species) |>summarise(n = n())
```

```
## # A tibble: 3 x 2
##   species       n
##   <fct>     <int>
## 1 Adelie      152
## 2 Chinstrap    68
## 3 Gentoo      124
```

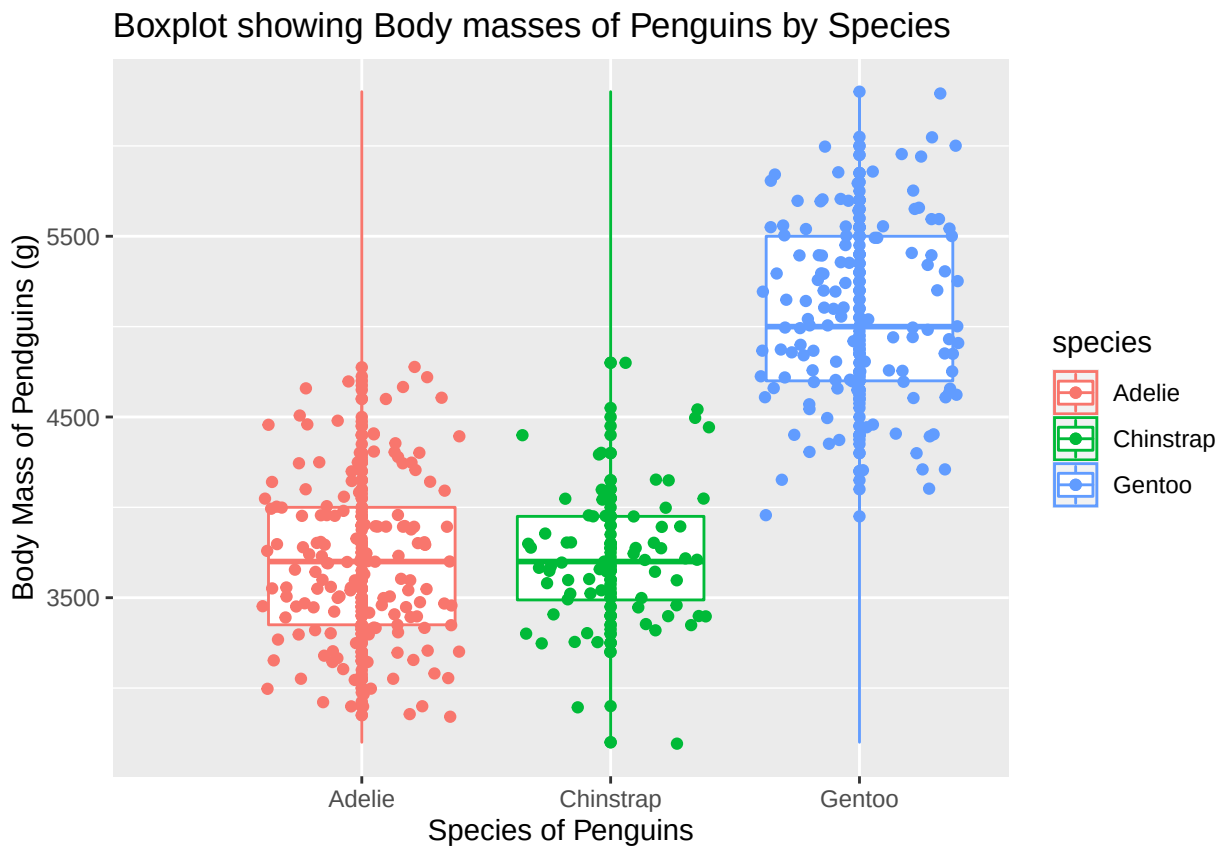
```
penguins |>
  group_by(species) |> ggplot(aes(x = species, y = body_mass_g, color = species)) +
  geom_boxplot() +
  geom_point()+
  geom_jitter() +
  geom_density()+
  labs( x = "Species of Penguins",
        y = "Body Mass of Pendguins (g)",
        title = "Boxplot showing Body masses of Penguins by Species")
```

```
## Warning: Removed 2 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 2 rows containing non-finite values (stat_density).
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```



24.4 Bonus Exercise

This code **filter(gapminder, country == c("Rwanda", "Afghanistan"))**‘ not not return the desired out. The assumption/rationale for this code was to subaset the data set to include countries containing only Rwanda and Afghanistan. However when run, the code gives a different output. It shows only a total of 12 observations (6 observations for Rwanda and 6 for Afghanistan) as shown below

```
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.
```

The correct syntax that would return the desired output is **filter(gapminder, country == "Rwanda"| country == "Afghanistan")**. By running this code, the correct output is displayed as shown below

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

```
## # 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
```