

# Lab3

Negar

2/5/2022

```
# load your packages here:
library(gapminder)
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()
```

24.2.1 - 1.1

```
gapminder |>
  filter(country %in% c('China', 'Brazil', 'Australia'),
         year >= 1970 & year <= 1979)
```

```
## # 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 Brazil    Americas    1972   59.5  100840058  4986.
## 4 Brazil    Americas    1977   61.5  114313951  6660.
## 5 China     Asia       1972   63.1  862030000   677.
## 6 China     Asia       1977   64.0  943455000   741.
```

24.2.2 - 1.2

```
gapminder |>
  filter(country %in% c('China', 'Brazil', 'Australia'),
         year >= 1970 & year <= 1979) |>
  select(country, gdpPercap)
```

```
## # A tibble: 6 x 2
##   country    gdpPercap
##   <fct>      <dbl>
## 1 Australia  16789.
## 2 Australia  18334.
## 3 Brazil     4986.
## 4 Brazil     6660.
## 5 China      677.
```

```
## 6 China          741.
```

#### 24.2.3 - 1.3

```
new_object <- gapminder |>
  arrange(year) |>
  group_by(country) |>
  mutate(lag_1_lifExp = lag(lifeExp),
         change_in_lifExp = lifeExp - lag_1_lifExp) |>
  filter(change_in_lifExp < 0)
new_object <- new_object[, c("country", "continent", "year", "lifeExp", "pop",
                             "gdpPercap", "change_in_lifExp", "lag_1_lifExp")]
new_object
```

```
## # A tibble: 102 x 8
## # Groups:   country [52]
##   country      continent  year lifeExp      pop gdpPercap change_in_lifExp
##   <fct>         <fct>    <int>  <dbl>    <int>    <dbl>         <dbl>
## 1 China        Asia      1962   44.5  665770000    488.         -6.05
## 2 Cambodia     Asia      1972   40.3   7450606     422.         -5.10
## 3 Czech Republic Europe    1972   70.3   9862158    13108.        -0.0900
## 4 Netherlands Europe    1972   73.8  13329874    18795.        -0.0700
## 5 Slovak Republic Europe    1972   70.4   4593433     9674.        -0.630
## 6 Bulgaria     Europe    1977   70.8   8797022     7612.        -0.0900
## 7 Cambodia     Asia      1977   31.2   6978607     525.         -9.10
## 8 El Salvador  Americas  1977   56.7   4282586     5139.        -1.51
## 9 Poland       Europe    1977   70.7   34621254    9508.        -0.180
## 10 Uganda      Africa    1977   50.4   11457758     844.        -0.666
## # ... with 92 more rows, and 1 more variable: lag_1_lifExp <dbl>
```

#### 24.2.4 - 1.4

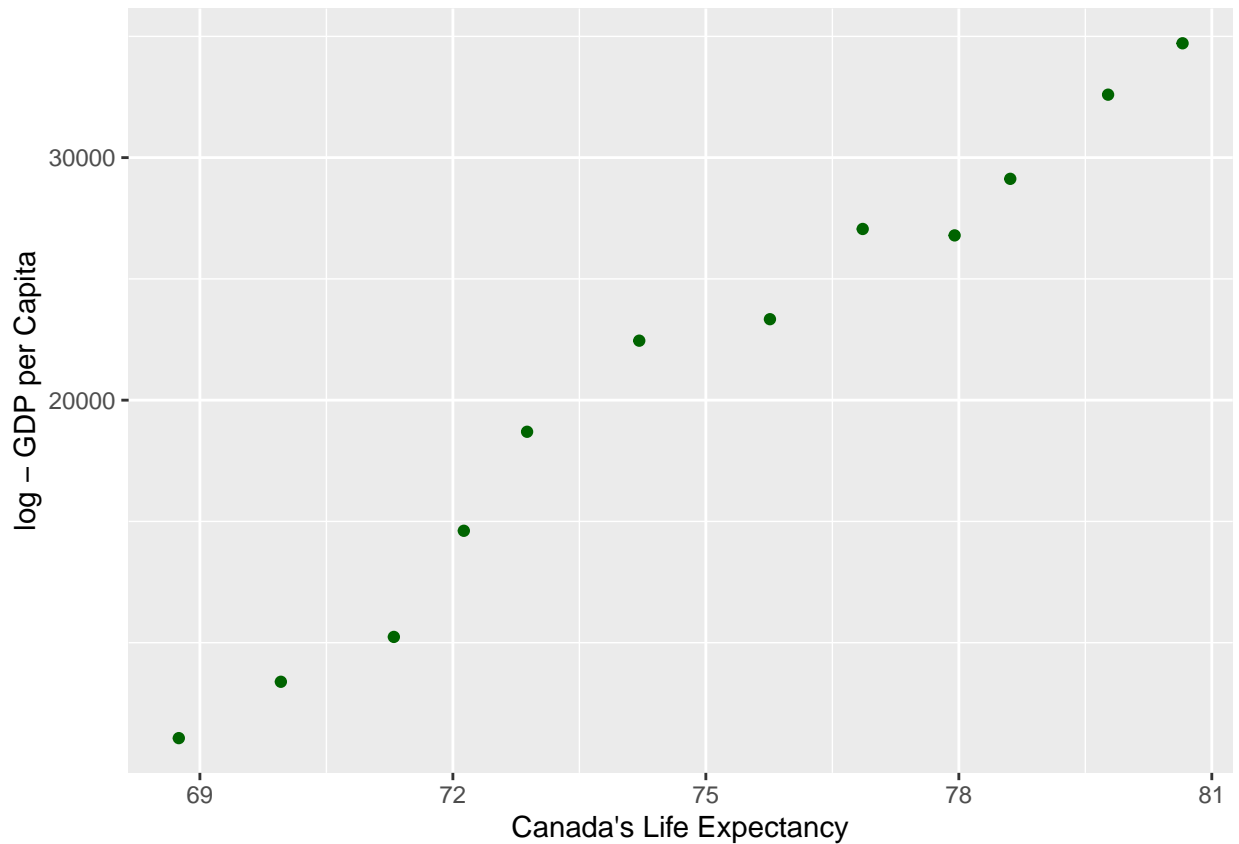
```
gapminder |>
  group_by(country) |>
  summarize(max_gdpPercap = max(gdpPercap))
```

```
## # A tibble: 142 x 2
##   country      max_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
```

#### 24.2.5 - 1.5

```
library(ggplot2)
ggplot(gapminder |>
  filter(country == "Canada")) +
  aes(x = lifeExp, y = gdpPercap) +
```

```
geom_point(color='darkgreen') +
scale_y_log10(name="log - GDP per Capita") +
xlab("Canada's Life Expectancy")
```



24.3.1 - 2.1

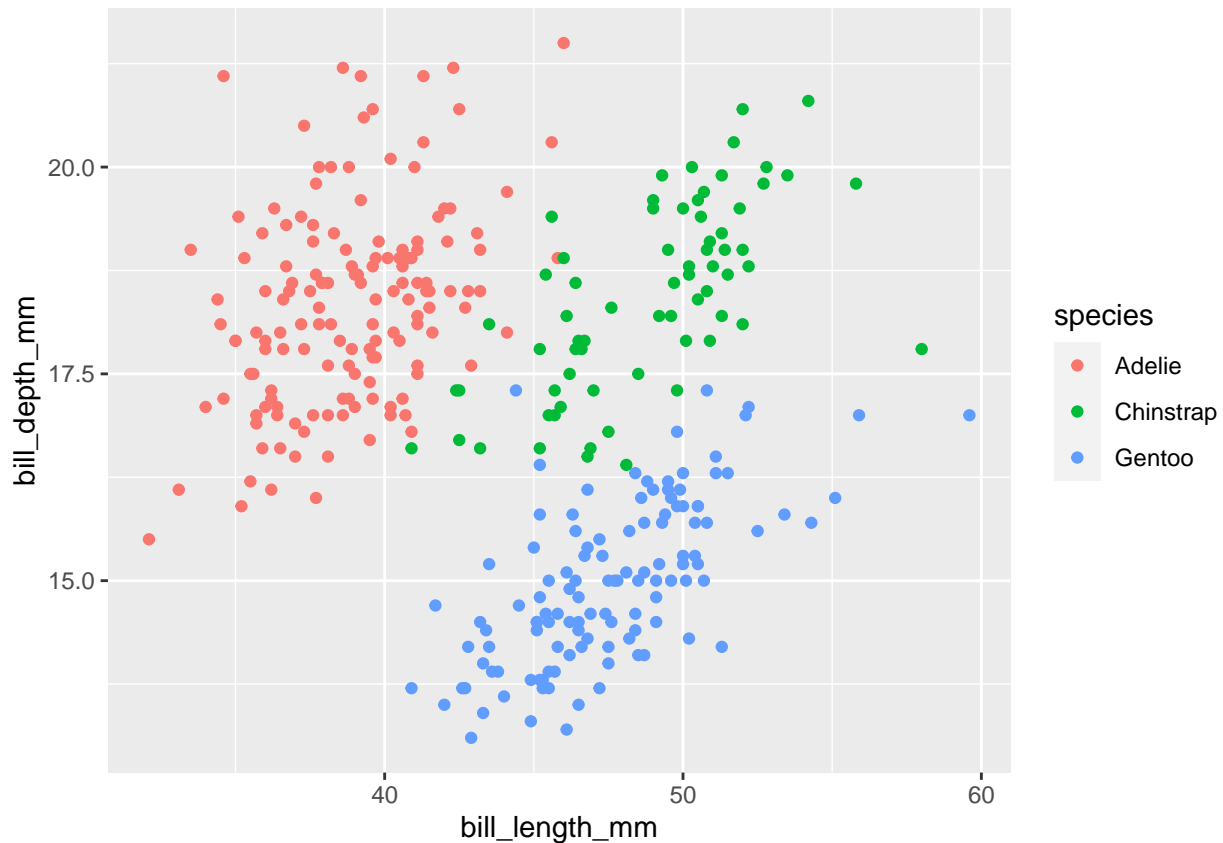
```
library(ggplot2)
library(palmerpenguins)
data <- palmerpenguins::penguins
data <- data %>% drop_na()
stat_desc <- data |>
  group_by(species) |>
  select(bill_length_mm, bill_depth_mm) |>
  summarise(sample_size = n(),
            mean_bill_length_mm = mean(bill_length_mm, na.rm = TRUE),
            mean_bill_depth_mm = mean(bill_depth_mm, na.rm = TRUE),
            range_bill_length_mm = range(bill_length_mm, na.rm = TRUE),
            range_bill_depth_mm = range(bill_depth_mm, na.rm = TRUE),
            std_bill_length_mm = sd(bill_length_mm, na.rm = TRUE),
            std_bill_depth_mm = sd(bill_depth_mm, na.rm = TRUE),
            median_bill_length_mm = median(bill_length_mm,
                                           na.rm = TRUE),
            median_bill_depth_mm = median(bill_depth_mm, na.rm = TRUE))

## Adding missing grouping variables: `species`
## `summarise()` has grouped output by 'species'. You can override using the `.groups` argument.
```

```
# Transpose the dataset, for better displaying
t(stat_desc)
```

```
##           [,1]      [,2]      [,3]      [,4]      [,5]
## species      "Adelie"    "Adelie"    "Chinstrap" "Chinstrap" "Gentoo"
## sample_size   "146"      "146"      " 68"      " 68"      "119"
## mean_bill_length_mm "38.82397" "38.82397" "48.83382" "48.83382" "47.56807"
## mean_bill_depth_mm "18.34726" "18.34726" "18.42059" "18.42059" "14.99664"
## range_bill_length_mm "32.1"      "46.0"      "40.9"      "58.0"      "40.9"
## range_bill_depth_mm "15.5"      "21.5"      "16.4"      "20.8"      "13.1"
## std_bill_length_mm  "2.662597" "2.662597" "3.339256" "3.339256" "3.106116"
## std_bill_depth_mm   "1.219338" "1.219338" "1.135395" "1.135395" "0.985998"
## median_bill_length_mm "38.85"     "38.85"     "49.55"     "49.55"     "47.40"
## median_bill_depth_mm "18.40"     "18.40"     "18.45"     "18.45"     "15.00"
##           [,6]
## species      "Gentoo"
## sample_size   "119"
## mean_bill_length_mm "47.56807"
## mean_bill_depth_mm "14.99664"
## range_bill_length_mm "59.6"
## range_bill_depth_mm "17.3"
## std_bill_length_mm  "3.106116"
## std_bill_depth_mm   "0.985998"
## median_bill_length_mm "47.40"
## median_bill_depth_mm "15.00"
```

```
ggplot(data) +
  aes(x = bill_length_mm, y = bill_depth_mm) +
  geom_point(aes(color=species))
```

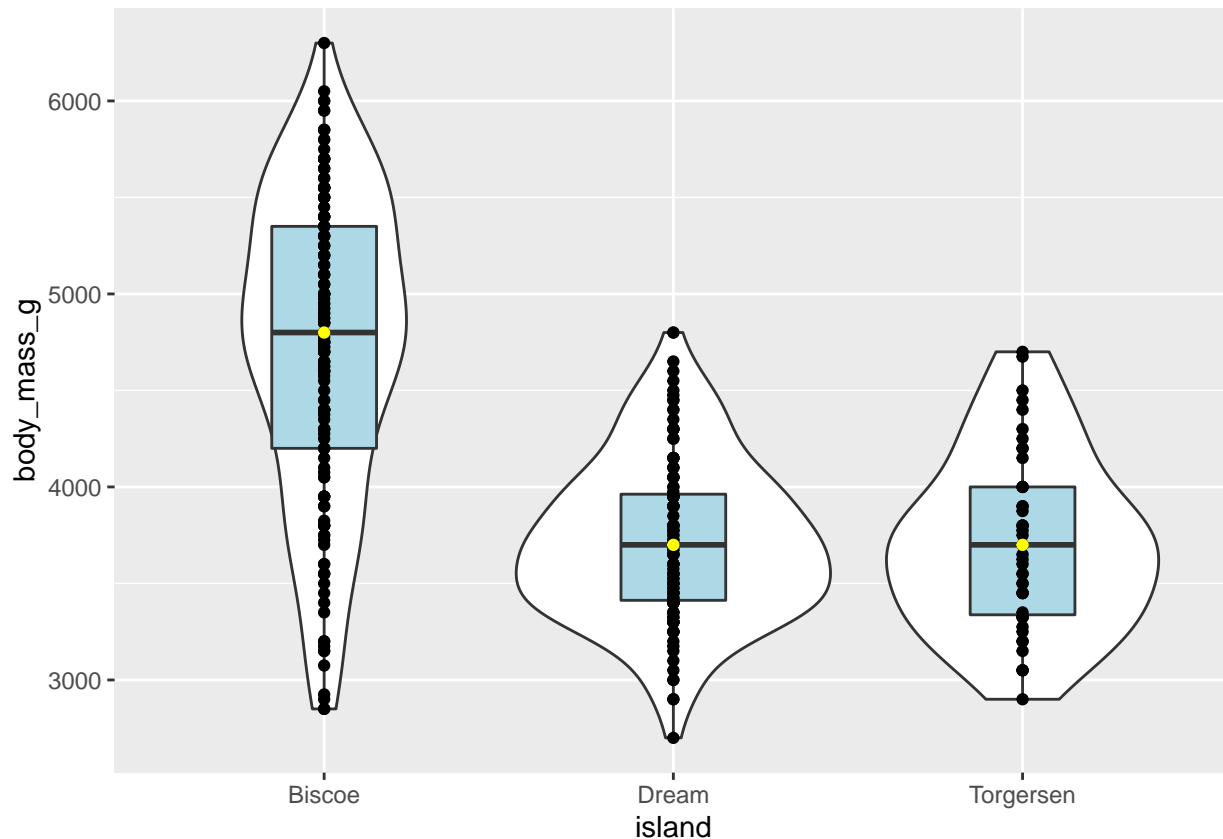


24.3.2 - 2.2

```
data <- palmerpenguins::penguins
data <- data %>% drop_na()
stat_desc_2 <- data |>
  select(island, body_mass_g) |>
  group_by(island) |>
  summarise(sample_size = n(),
            mean_body_mass_g = mean(body_mass_g, na.rm = TRUE),
            std_body_mass_g = sd(body_mass_g, na.rm = TRUE),
            median_body_mass_g = median(body_mass_g, na.rm = TRUE))
# Transpose the dataset, for better displaying
t(stat_desc_2)
```

```
##           [,1]      [,2]      [,3]
## island      "Biscoe"    "Dream"    "Torgersen"
## sample_size "163"      "123"      " 47"
## mean_body_mass_g "4719.172" "3718.902" "3708.511"
## std_body_mass_g "790.8601" "412.9356" "451.8464"
## median_body_mass_g "4800"    "3700"    "3700"
```

```
ggplot(data,
       mapping = aes(x = island, y = body_mass_g)) +
  geom_violin() +
  geom_boxplot(width = .3, fill = "lightblue") +
  geom_point() +
  stat_summary(fun = median, geom = "point", color = "yellow")
```



#### 24.4 Bonus Exercise

No it does not work. If we use the given code, we get half of the result, since “==” sign compare each element one by one is a roll, and if it is True, then display it in result. In this example, as the first and second row is “Afghanistan” and our vector is “c(“Rwanda”, “Afghanistan”)”, we get “FALSE, TRUE” result so then the first row will not display in the result.

```
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 code is as follow:

```
gapminder |>
  filter(country %in% c("Rwanda", "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
```