HW3-Data Analysis

Aylin Mumcular

27 09 2019

install.packages("gapminder") install.packages("tidyverse") install.packages("dplyr") install.packages("ggrepel") install.packages("ggpubr")

Produce:

- A tibble, using dplyr as your data manipulation tool
- An accompanying plot of data from the tibble, using ggplot2 as your visualization tool
- Some dialogue about what your tables/figures show (doesn't have to be much)

1. Get the maximum and minimum of GDP per capita for all continents.

```
#Tibble
gapminder %>%
  group_by(continent) %>%
  summarize(maxGDPpercap = max(gdpPercap),
            minGDPpercap = min(gdpPercap)) %>%
              arrange(maxGDPpercap)
```

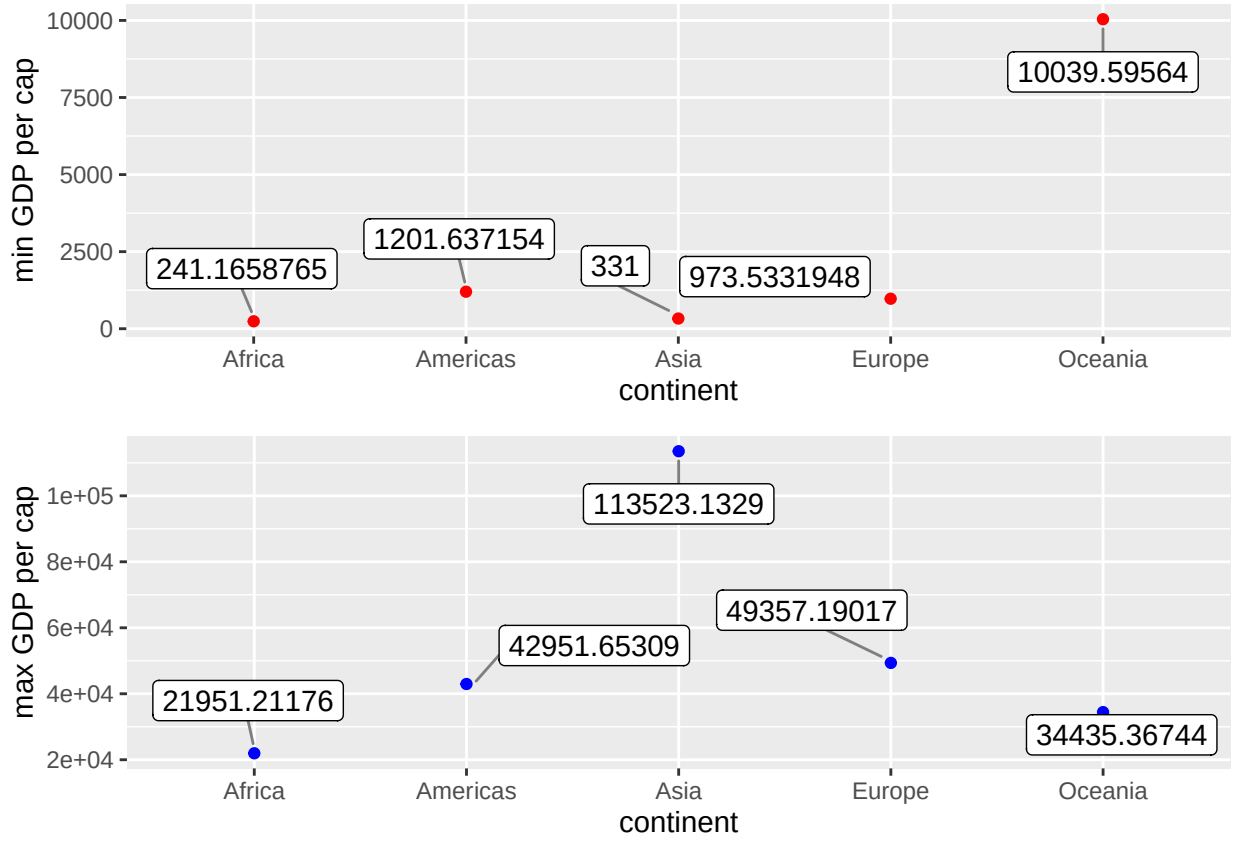
```
## # A tibble: 5 x 3
##   continent maxGDPpercap minGDPpercap
##   <fct>            <dbl>        <dbl>
## 1 Africa          21951.         241.
## 2 Oceania         34435.       10040.
## 3 Americas        42952.        1202.
## 4 Europe          49357.         974.
## 5 Asia           113523.         331
```

```
#Plot
p1 <- gapminder %>%
  group_by(continent) %>%
  summarize(maxGDPpercap = max(gdpPercap),
            minGDPpercap = min(gdpPercap)) %>%
             ggplot(aes(continent, minGDPpercap)) +
               geom_point(colour = "red") +
                  geom_label_repel(aes(label = minGDPpercap),
  #To add values as labels
  #(source: https://stackoverflow.com/questions/15624656/label-points-in-geom-point)
                   box.padding   = 0.35,
                   point.padding = 0.5,
                   segment.color = 'grey50') +
                   ylab("min GDP per cap")

p2 <- gapminder %>%
  group_by(continent) %>%
  summarize(maxGDPpercap = max(gdpPercap),
            minGDPpercap = min(gdpPercap)) %>%
             ggplot(aes(continent, maxGDPpercap)) +
               geom_point(colour = "blue") +
                  geom_label_repel(aes(label = maxGDPpercap),
                   box.padding   = 0.35,
```

```
                  point.padding = 0.5,
                  segment.color = 'grey50') +
                  ylab("max GDP per cap")

ggarrange(p1, p2, nrow = 2, ncol = 1) #Combine two plots
```



Minimum GDP per capita order of continents from highest to lowest is Oceania, Americas, Europe, Asia, and Africa whereas maximum GDP per capita order of continents from highest to lowest is Asia, Europe, Americas, Oceania, and Africa. Africa is the lowest GDP per capita continent according to both maximum and minimum GDP per capita measures. Asia has the highest gap between countries in terms of GDP per Capita.

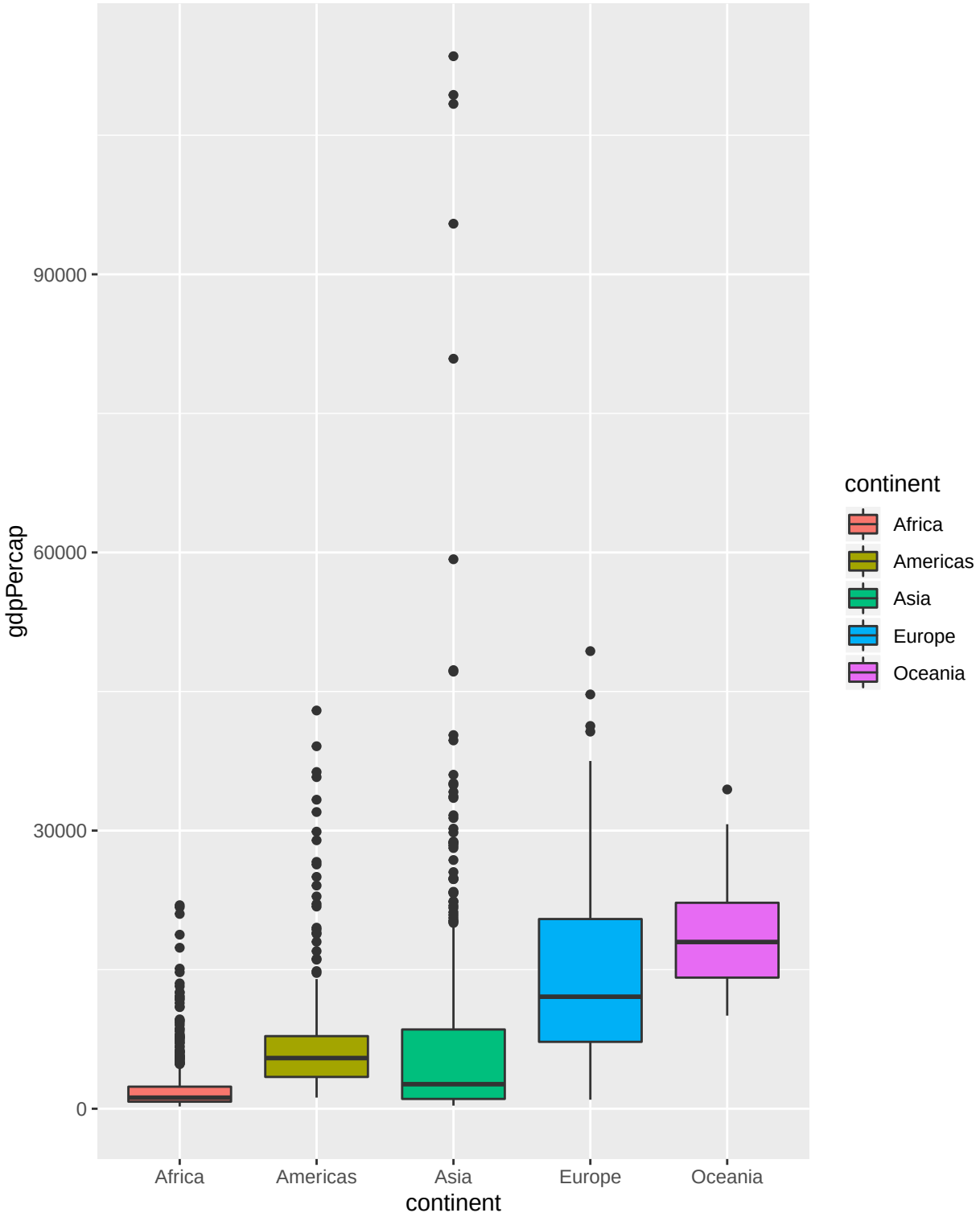
2. Look at the spread of GDP per capita within the continents.

```
#Tibble
gapminder %>%
  group_by(continent) %>%
  summarize(maxGDPpercap = max(gdpPercap),
            minGDPpercap = min(gdpPercap),
            rangeGDPpercar = max(gdpPercap)-min(gdpPercap),
            stdGDPpercap = sd(gdpPercap),
            meanGDPpercap = mean(gdpPercap),
            medianGDPpercap = median(gdpPercap),
            firstquantile = quantile(gdpPercap, 0.25),
            thirdquantile = quantile(gdpPercap, 0.75))
```

```
## # A tibble: 5 x 9
```

```
##   continent maxGDPpercap minGDPpercap rangeGDPpercar stdGDPpercap
##   <fct>            <dbl>        <dbl>          <dbl>        <dbl>
## 1 Africa          21951.         241.         21710.        2828.
## 2 Americas        42952.        1202.         41750.        6397.
## 3 Asia           113523.         331         113192.       14045.
## 4 Europe          49357.         974.         48384.        9355.
## 5 Oceania         34435.       10040.         24396.        6359.
## # ... with 4 more variables: meanGDPpercap <dbl>, medianGDPpercap <dbl>,
## #   firstquantile <dbl>, thirdquantile <dbl>
```

```r
#Plot
gapminder %>%
  ggplot(aes(continent, gdpPercap, fill = continent)) +
  geom_boxplot()
```



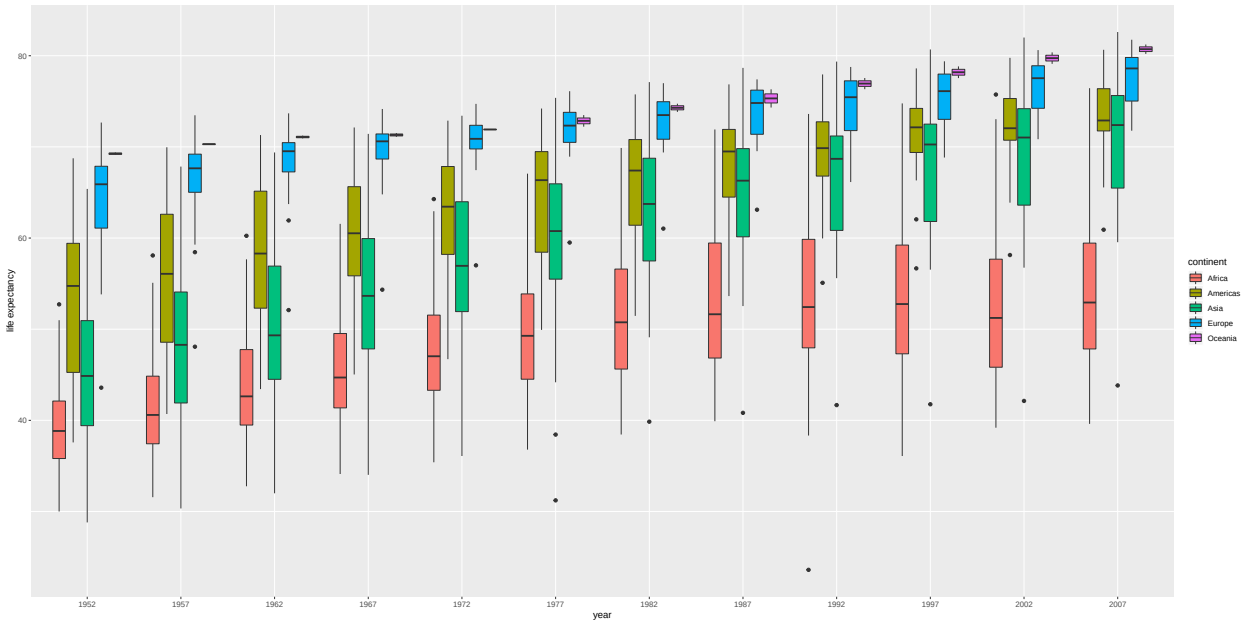
Asia has the highest range and standard deviation in terms of GDP per capita while Africa has the lowest. Europe has the highest interquantile range. Oceania has the highest mean and median GDP per capita, followed by Europe.

3. How is life expectancy changing over time on different continents?

```r
#Tibble
gapminder %>%
  group_by(continent, year) %>%
    summarize(maxlifeExp = max(lifeExp),
              minlifeExp = min(lifeExp),
              meanlifeExp = mean(lifeExp),
              medianlifeExp = median(lifeExp))
```

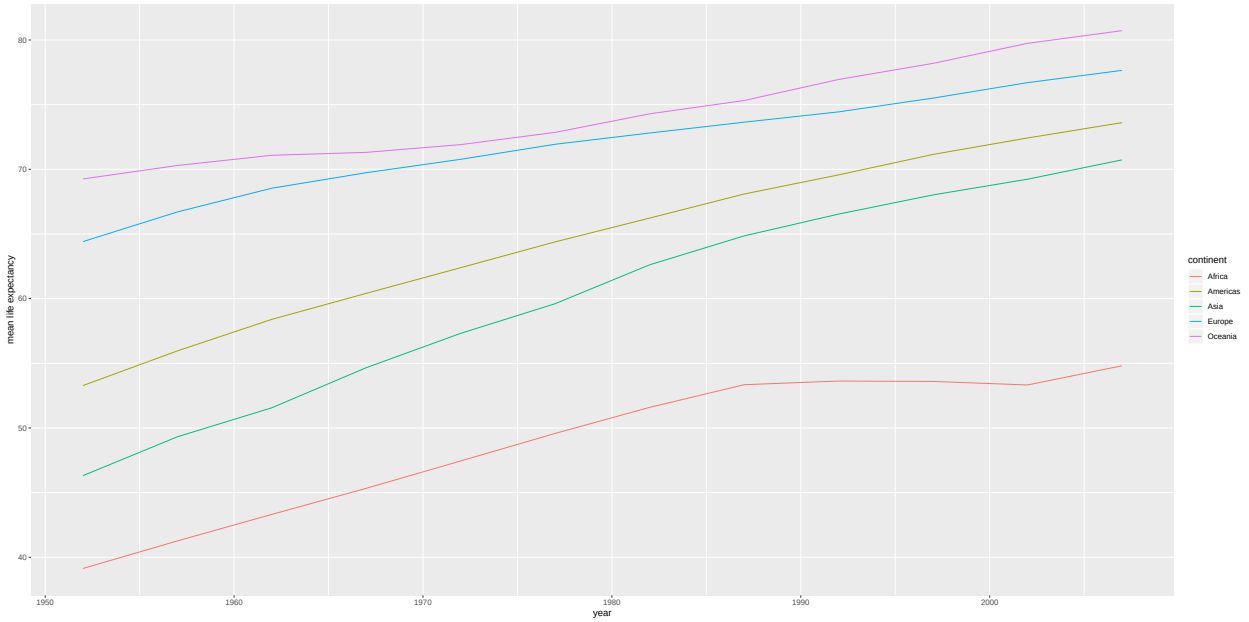
```
## # A tibble: 60 x 6
## # Groups:   continent [5]
##    continent  year maxlifeExp minlifeExp meanlifeExp medianlifeExp
##    <fct>     <int>      <dbl>      <dbl>       <dbl>          <dbl>
##  1 Africa     1952       52.7       30          39.1           38.8
##  2 Africa     1957       58.1       31.6        41.3           40.6
##  3 Africa     1962       60.2       32.8        43.3           42.6
##  4 Africa     1967       61.6       34.1        45.3           44.7
##  5 Africa     1972       64.3       35.4        47.5           47.0
##  6 Africa     1977       67.1       36.8        49.6           49.3
##  7 Africa     1982       69.9       38.4        51.6           50.8
##  8 Africa     1987       71.9       39.9        53.3           51.6
##  9 Africa     1992       73.6       23.6        53.6           52.4
## 10 Africa     1997       74.8       36.1        53.6           52.8
## # ... with 50 more rows
```

```r
#Plot
gapminder %>%
  mutate(year = factor(year)) %>%
  #Convert year from numeric to factor to print multiple boxplots within each year
    ggplot(aes(year, lifeExp, fill = continent)) +
    geom_boxplot() +
    ylab("life expectancy")
```




```r
#Plot 2
gapminder %>%
```

```
  group_by(continent, year) %>%
    summarize(meanlifeExp = mean(lifeExp)) %>%
      ggplot(aes(year, meanlifeExp, group = continent, colour = continent)) +
      geom_line() +
      ylab("mean life expectancy")
```



In general, there is an increasing trend in the life expectancy of each continent over time. Oceania and Africa have experienced some drop in life expectancy in the past.