STAT 545A Assignment 03: dplyr/ggplot2 Part II

```
library(tidyverse)
library(ggplot2)
library(gapminder)
library(gridExtra)
library(grid)
library(spelling)
knitr::opts_chunk$set(echo = TRUE)
```

Introduction:

In this assignment, I used the gapminder dataset which contains country data from 1952 to 2007 (12 years) for 142 countries in 5 continents. Specifically, I focused on the variables:

- GDP per capita (US$, inflation-adjusted) from the World Bank (WB) and
- Life Expectancy (years), defined as the average number of years a newborn child would live if current mortality patterns were to stay the same.

I chose the tasks 2, 3 and 5

Task Option 2

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

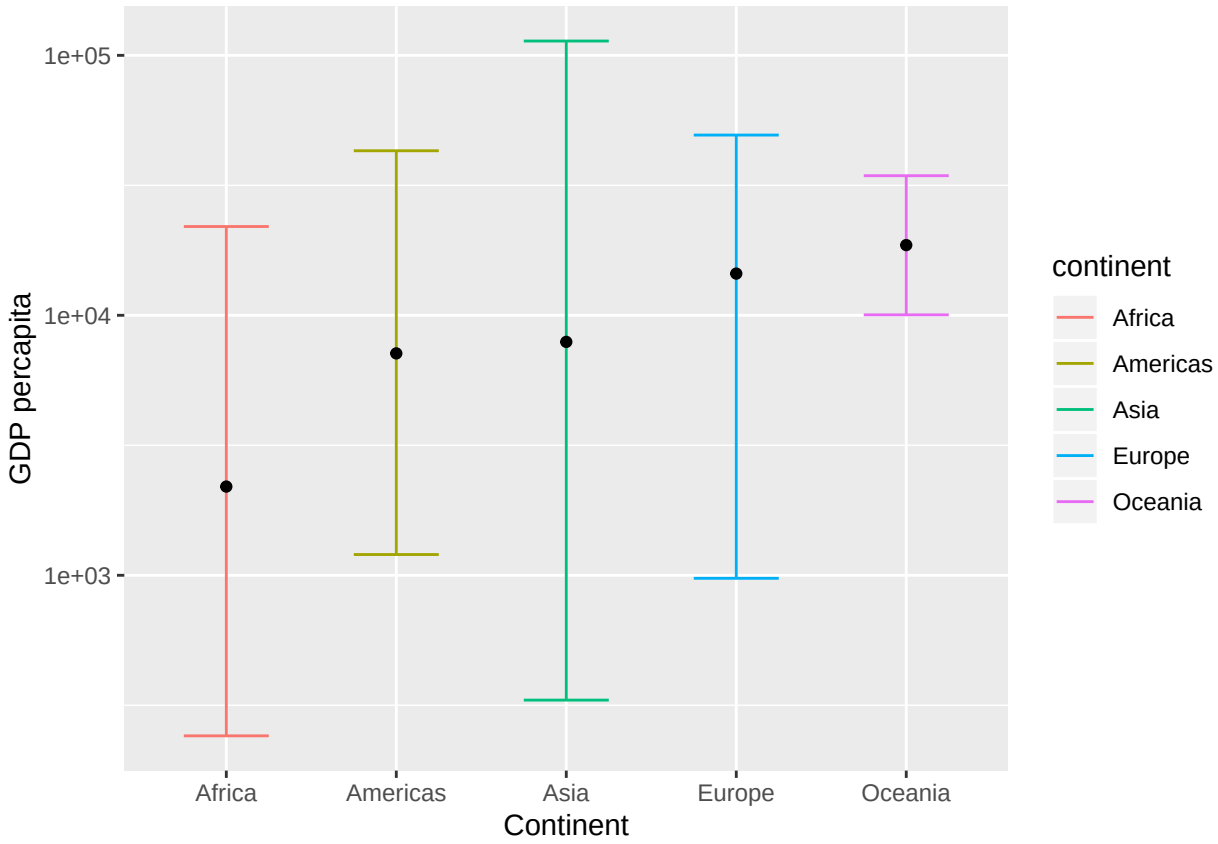
In the table below, I summarized the minimum and maximum GDP percapita observed in the database by continent. Africa shows the lowest GDP per capita and Asia the highest.

```
rangeGdpP <- gapminder %>%
  group_by(continent) %>%
  summarize(min_gdpPercap = min(gdpPercap),
            max_gdpPercap = max(gdpPercap),
            mean_gdpPercap = mean(gdpPercap))
rangeGdpP
```

```
## # A tibble: 5 x 4
##   continent min_gdpPercap max_gdpPercap mean_gdpPercap
##   <fct>             <dbl>         <dbl>          <dbl>
## 1 Africa             241.         21951.          2194.
## 2 Americas          1202.         42952.          7136.
## 3 Asia               331         113523.          7902.
## 4 Europe             974.         49357.         14469.
## 5 Oceania          10040.         34435.         18622.
```

In the figure below, I create a graph that contains the range for each continent. I included also the average GDP per capita as well. Asia is the continent that shows a wider range of values.

```
ggplot(rangeGdpP) +
  geom_errorbar(aes(x=continent, ymin=min_gdpPercap, ymax=max_gdpPercap, color=continent, width=.5))  +
  geom_point(aes(x=continent, y=mean_gdpPercap)) +
  scale_y_log10() +
  ylab("GDP percapita") + xlab("Continent")
```



Task Option 3

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

The table below depicts the variation within each continent about the GDP per capita. As seen in the previous task, Asia presents a higher variation with a standard deviation considerably higher than the rest of the continents.

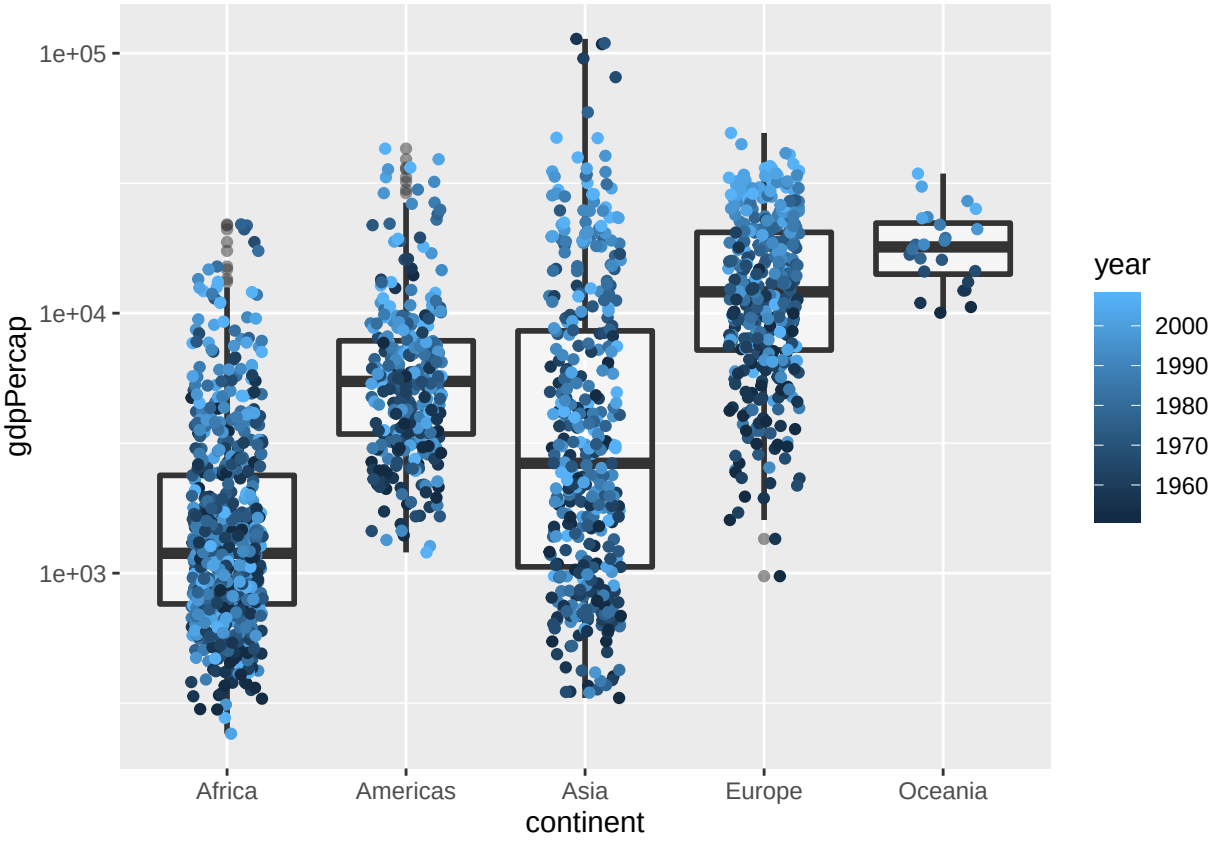
```
gapminder %>%
group_by(continent) %>%
summarize(mean=mean(gdpPercap),
          min=min(gdpPercap),
          max=max(gdpPercap),
          std=sd(gdpPercap),
          q25=quantile(gdpPercap,0.25),
          q50=quantile(gdpPercap, 0.5),
          q75=quantile(gdpPercap, 0.75))
```

```
## # A tibble: 5 x 8
##   continent   mean    min     max    std    q25    q50    q75
##   <fct>      <dbl>  <dbl>   <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
## 1 Africa     2194.   241.  21951.  2828.   761.  1192.  2377.
## 2 Americas   7136.  1202.  42952.  6397.  3428.  5466.  7830.
## 3 Asia       7902.   331  113523. 14045.  1057.  2647.  8549.
## 4 Europe    14469.   974.  49357.  9355.  7213. 12082. 20461.
## 5 Oceania   18622. 10040.  34435.  6359. 14142. 17983. 22214.
```

The figure below shows the spread of the GDP per capita by continent. The box shows the standard deviation

around the average GDP per capita. In the individual country's GDP data points is possible to see a slight increasing trend in time. This can be observed clearly in Oceania.

```
ggplot(gapminder, aes(x=continent, y=gdpPercap)) +
  geom_boxplot(alpha=0.5, size=1, shape=1,)  + geom_jitter(width = 0.2, aes(colour = year)) +
  scale_y_log10()
```



Task Option 5

How is life expectancy changing over time on different continents?

To answer this question I simply calculated the continent yearly average and median Life Expectancy.

```
gapminder %>%
    group_by(continent, year) %>%
    summarise(median_lifeExp=median(lifeExp),
              mean_lifeExp=mean(lifeExp))
```

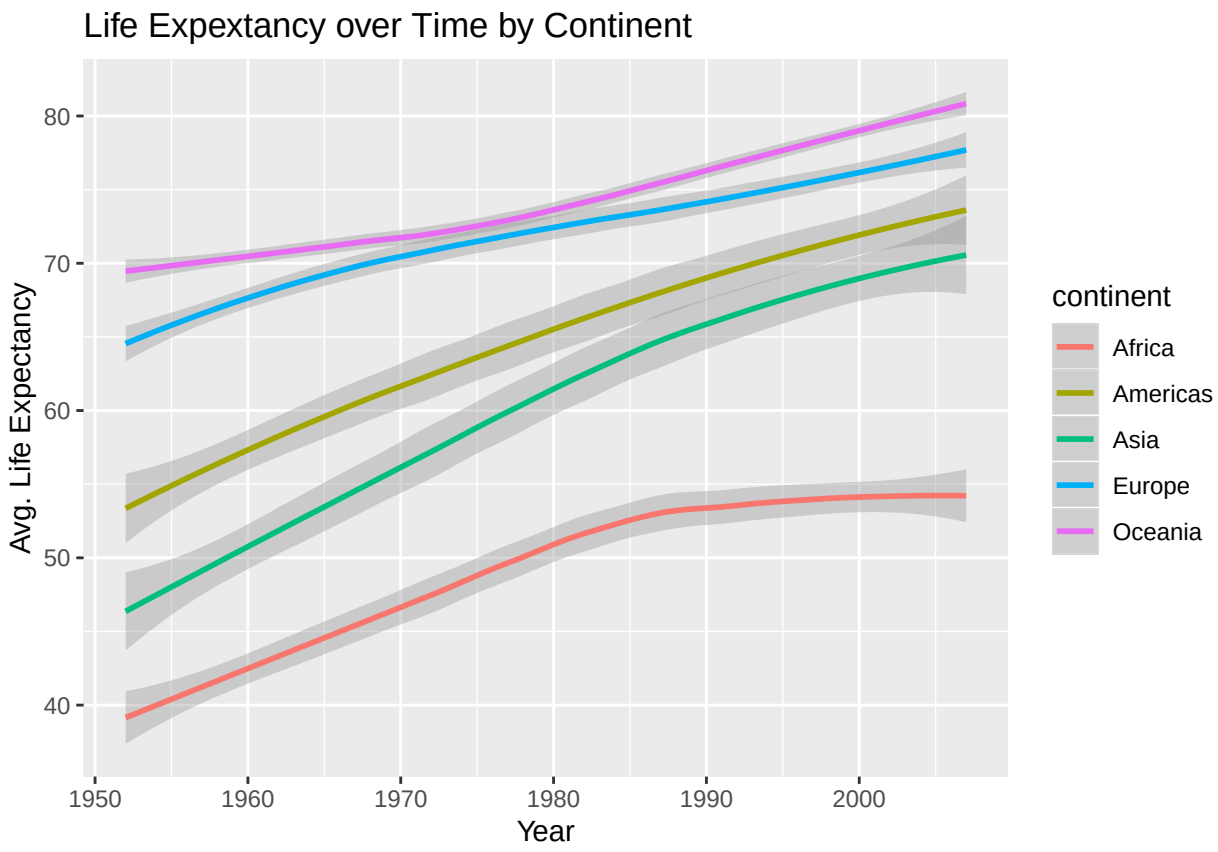
```
## # A tibble: 60 x 4
## # Groups:   continent [5]
##    continent  year median_lifeExp mean_lifeExp
##    <fct>     <int>          <dbl>        <dbl>
##  1 Africa     1952           38.8         39.1
##  2 Africa     1957           40.6         41.3
##  3 Africa     1962           42.6         43.3
##  4 Africa     1967           44.7         45.3
##  5 Africa     1972           47.0         47.5
##  6 Africa     1977           49.3         49.6
```

```
##  7 Africa     1982           50.8         51.6
##  8 Africa     1987           51.6         53.3
##  9 Africa     1992           52.4         53.6
## 10 Africa     1997           52.8         53.6
## # ... with 50 more rows
```

In the graph below, I show how average life expectancy is growing in time. Each continent shows a positive trend.

```
ggplot(gapminder,
       aes(year,lifeExp, colour = continent)) +
  geom_smooth() + ggtitle("Life Expextancy over Time by Continent") +
  xlab("Year") + ylab("Avg. Life Expectancy")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Optional Exercise

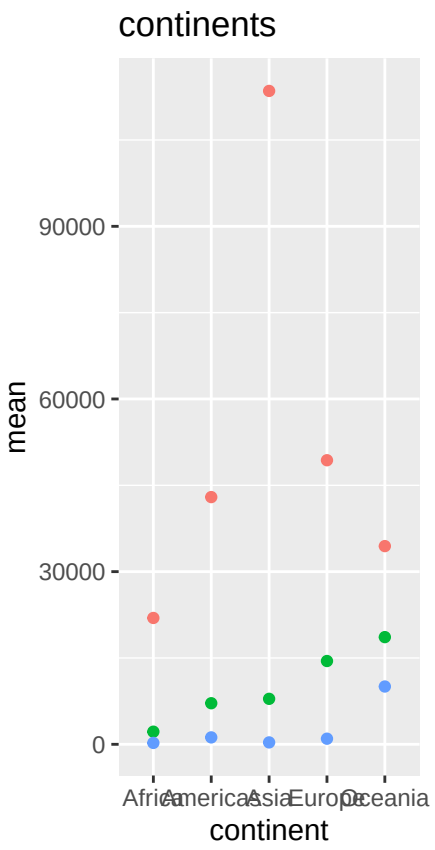
Table and figure side-by-side. In the table, I only show the average GDP to fit both figures side by side.

```
ex <- gapminder %>%
group_by(continent) %>%
summarize(mean=mean(gdpPercap),
          min=min(gdpPercap),
          max=max(gdpPercap))

exg <- ex %>%
  ggplot(aes(continent)) +
```

```
  geom_point(aes(y = mean, colour = "mean")) +
  geom_point(aes(y = min, colour = "min")) +
  geom_point(aes(y = max, colour = "max")) +
  ggtitle("continents")

  grid.arrange( exg, tableGrob(ex[1:4, 1:2]), nrow = 1, widths = 3:1)
```



	continent	mean
1	Africa	2193.755
2	Americas	7136.110
3	Asia	7902.150
4	Europe	14469.476