hw03_YongzhengParkerLi

Yongzheng Parker Li

This repository is the homepage of *Homework 03* of the course STAT545A, taught by Vincenzo Coia at the University of British Columbia (UBC). The detailed requirements of this assignment could be found here. The STAT545A course page is here. My own participation repository is here.

Bring Rectangular Data In

This section installs the related packages for this assignment: working with the gapminder data we explored in class.

```
library(gapminder)
library(tidyverse)
```

Task 1 (TK1): Get the Maximum and Minimum of GDP per Capita for All Continents

Utilizing group_by and summarise functions, I divide countires into different contients and cauculate the correspoding maximum and minimum of GDP per capit for all continents. I create the straightfoward bar figures to show the results.

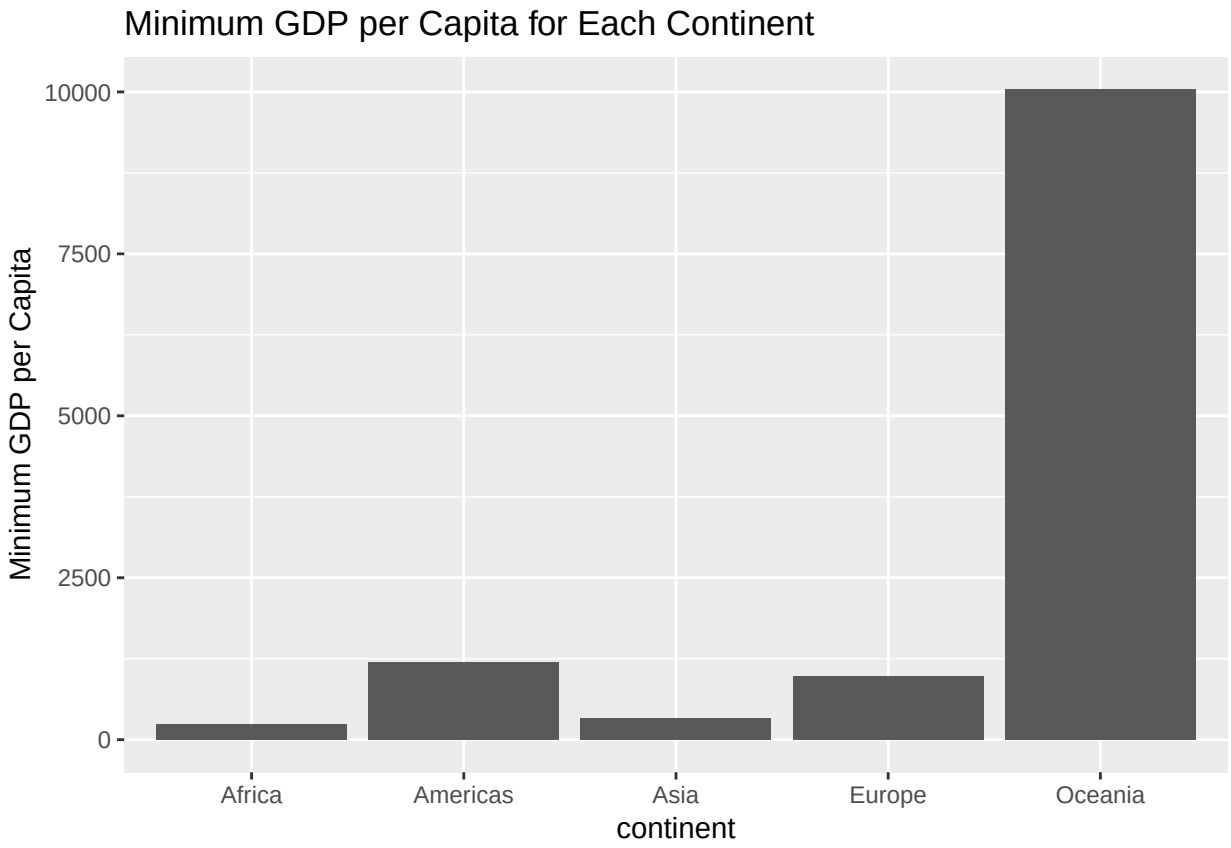
```
tk1 <- gapminder %>%
  group_by(continent) %>%
  summarise(min = min(gdpPercap), max = max(gdpPercap))
knitr::kable(tk1)
```

continent	min	max
Africa	241.1659	21951.21
Americas	1201.6372	42951.65
Asia	331.0000	113523.13
Europe	973.5332	49357.19
Oceania	10039.5956	34435.37

Some basic observations from the table and graphs below: Africa has the worst GDP per capital situation. Oceania has the best. Asia has a highest gdp per capita (max) but there is huge gap.

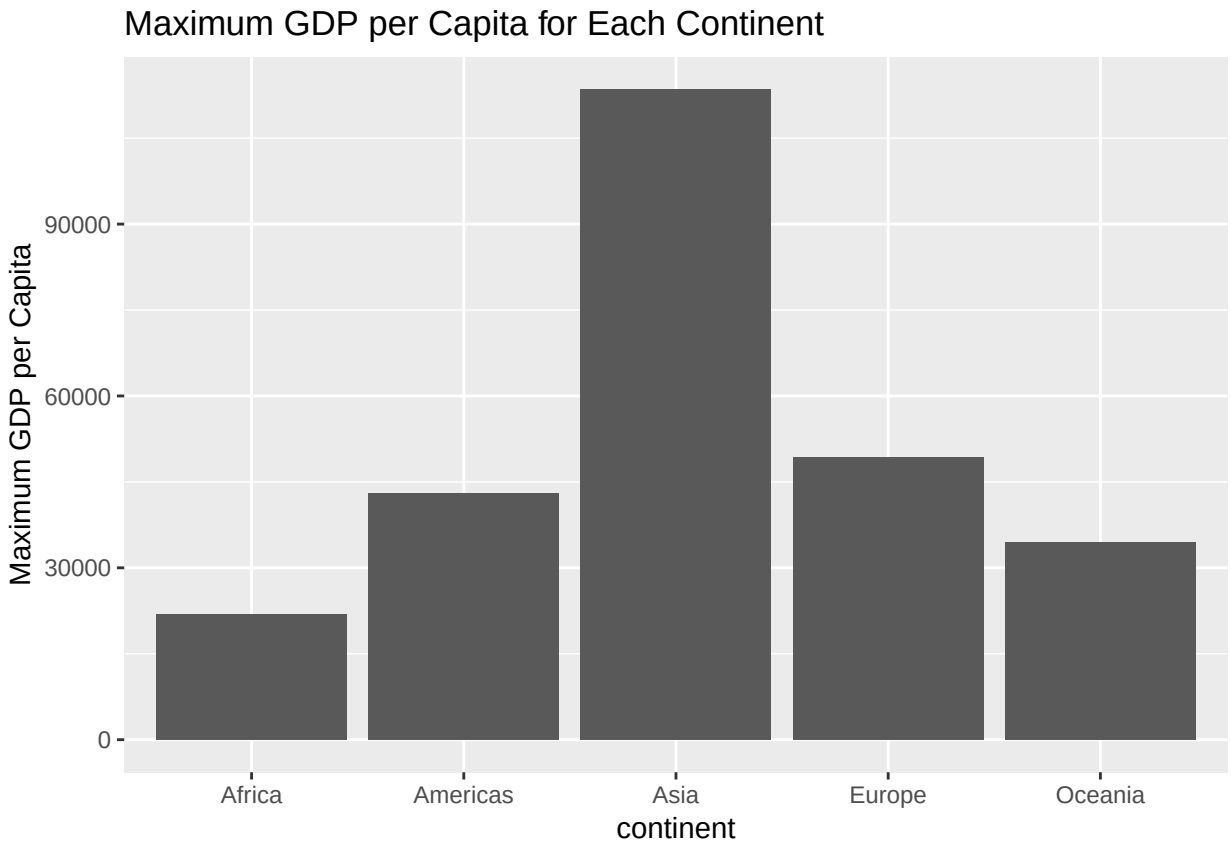
Here is the bar graph for Minimum GDP per capita for Each continent.

```
tk1_min <- ggplot(tk1, aes(continent, min)) +
  geom_bar(stat = "identity") +
  ggtitle("Minimum GDP per Capita for Each Continent") +
  ylab("Minimum GDP per Capita")
tk1_min
```



Here comes the bar graph for Maximum GDP per capita for Each continent.

```
tk1_max <- ggplot(tk1, aes(continent, max)) +
  geom_bar(stat = "identity") +
  ggtitle("Maximum GDP per Capita for Each Continent") +
  ylab("Maximum GDP per Capita")
tk1_max
```



Task 2 (TK2): Spread of GDP per Capita within the Continents

This is an "add-on" of the first task. Mean and standard deviation are cauculated to show the spread. Boxplots and spread figure is presented.

```
tk2 <- gapminder %>%
  group_by(continent) %>%
  summarize(
  min_GDPper = min(gdpPercap),
  max_GDPper = max(gdpPercap),
  mean_GDPper = mean(gdpPercap),
  sd_GDPper = sd(gdpPercap)
  )
knitr::kable(tk2)
```

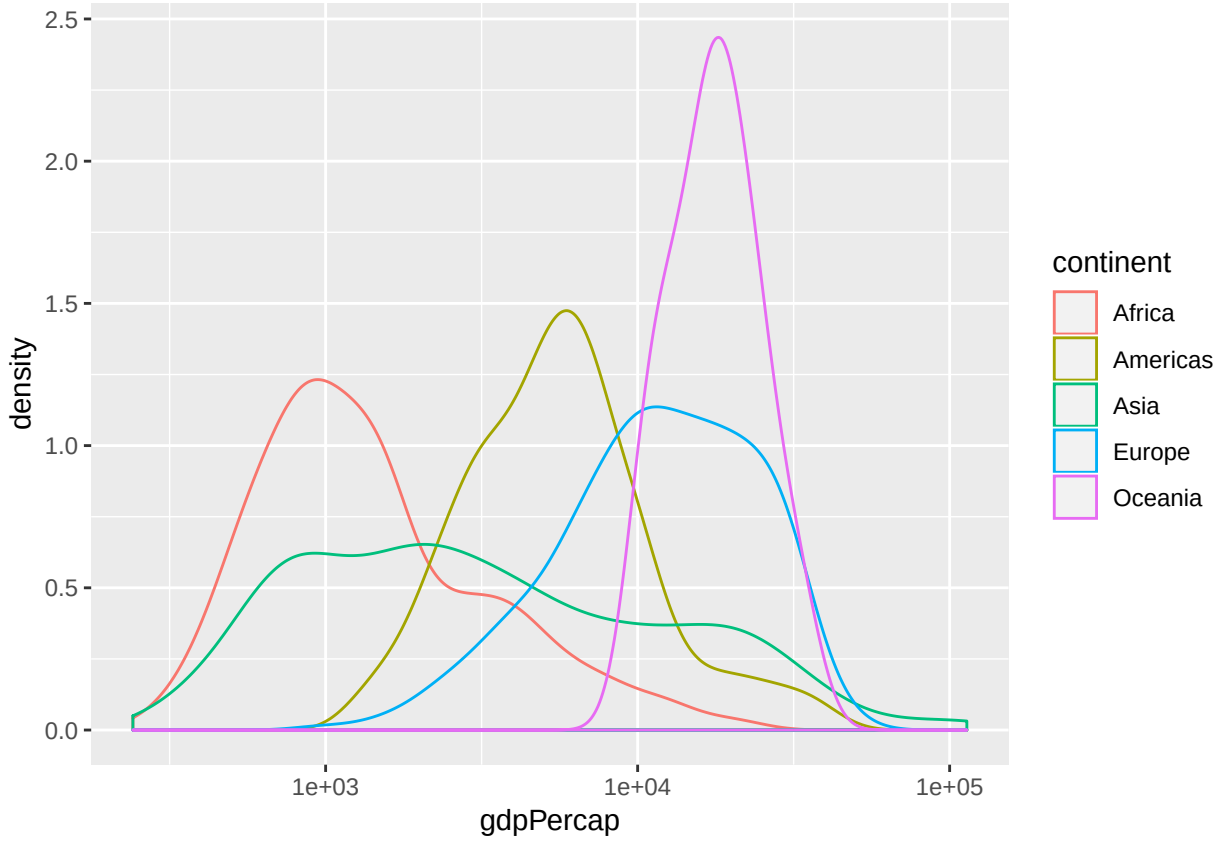
continent	min_GDPper	max_GDPper	mean_GDPper	sd_GDPper
Africa	241.1659	21951.21	2193.755	2827.930
Americas	1201.6372	42951.65	7136.110	6396.764
Asia	331.0000	113523.13	7902.150	14045.373
Europe	973.5332	49357.19	14469.476	9355.213
Oceania	10039.5956	34435.37	18621.609	6358.983

Some basic observations from the table and figures below (except for the ones in task 1): tail matters! From the spread, we know all the continets are not normally distributed. Some countries within each continet are

making the spread skewed to one direction.

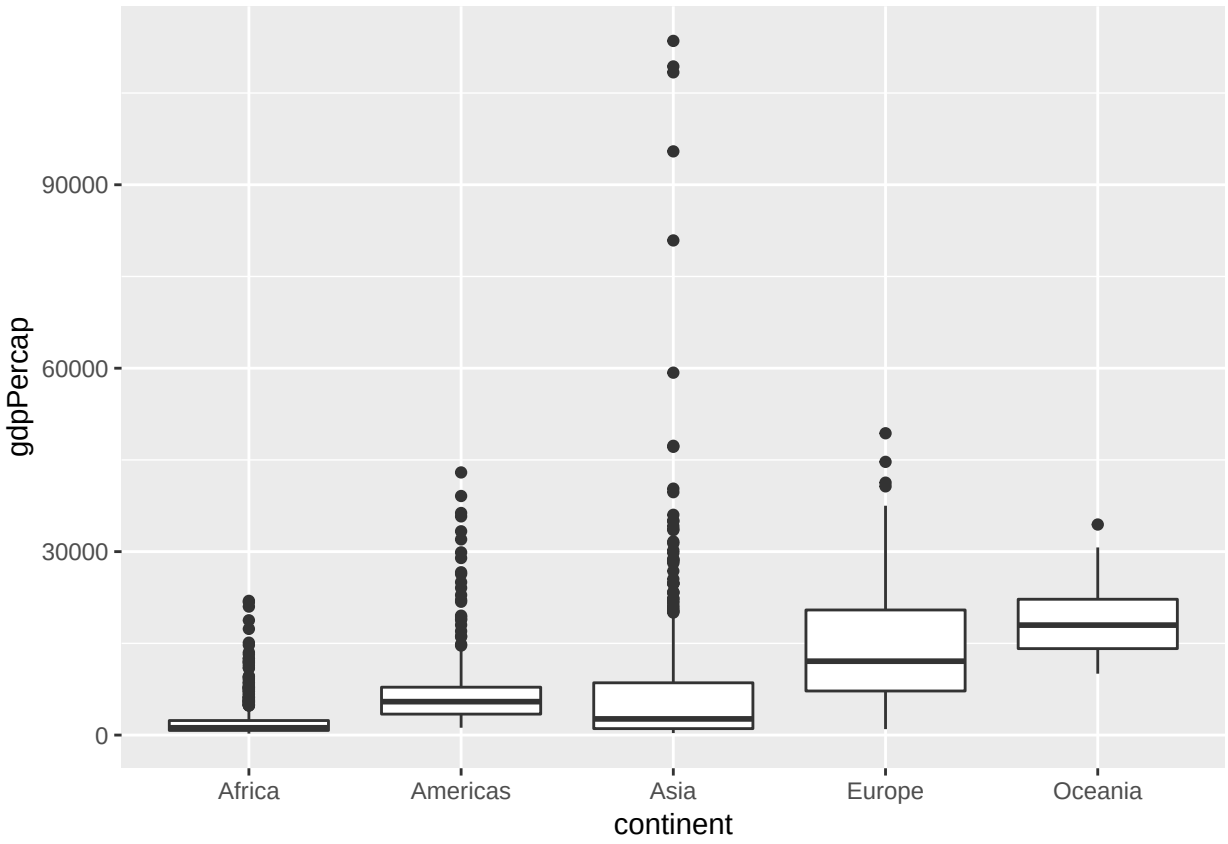
Two graphs are demonstrated here to show the spread. One is the density graph, the other is the boxplot.

```
gapminder %>%
  ggplot(aes(x=gdpPercap, color=continent)) +
  scale_x_log10() +
  geom_density()
```



Here is the boxplot. Personally I prefer the boxplot because it is more straightforward to digest the info. However, density plot is nicer visually.

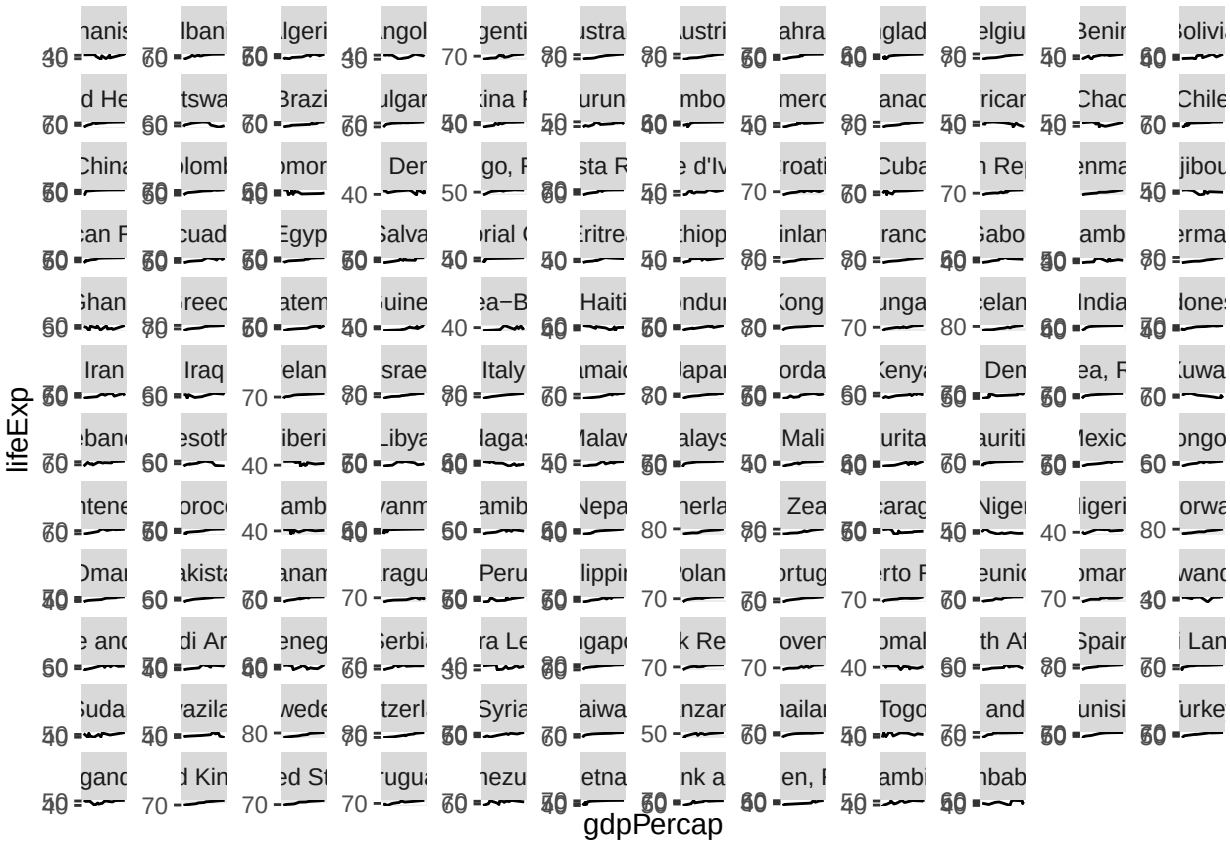
```
gapminder %>%
  ggplot(aes(x=continent, y=gdpPercap)) +
  geom_boxplot()
```



Task 3 Interesting Stories

After a quick analysis of GDP per capita (taks 1 and 2), now I move to show the relationship between GDP per capita. My ultimate goal is to creat a multilevel model (level one: country; level two: continent. longtinidual data). However, due to the course knowledge process, for this part, I only show the each-country-within relationship graph.

```
gapminder %>%
  group_by(continent) %>%
  ggplot(aes(x=gdpPercap, y=lifeExp))+
  scale_y_log10()+
  facet_wrap(~country, scales = "free") +
  geom_line()+
  scale_x_continuous(breaks = NULL)
```



It seems high gdp per capita is related to high life length. However, we need a more spohiscated model to cauculate it. Something like this (won't work, just a basic-need package and more data):

```
tk3<-lme(gdpPercap~1+lifeExp+pop,
                    random = list(~1+lifeExp|country, ~1+lifeExp|continent),
                    data=gapminder, na.action=na.omit, method="ML", control=list(opt="optim"))

summary(tk3)
```