# week12 exercise

Xiang Li

2023/12/30

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(brolgar)
```

## Exercise 1

```
setwd("/Users/macbookpro/Desktop/learning materials/Statistical Computing with R")
data_df = read.csv("data/irish_polls.csv")
data_df[data_df == "Not Available"] = NA
to_decimal = function(x) {
    return(sub("%", "", x))
}
data_df[, 10:21] = lapply(data_df[, 10:21], to_decimal)
data_df[, 10:21] = as.numeric(unlist(data_df[, 10:21]))/100
colnames(data_df)[c(11:12, 17)] = c("Fianna.Fail", "Sinn.Fein", "Aontu")
```

### 2

```
order_data_df = mutate(data_df, Fieldwork.End = as.Date(Fieldwork.End, format = "%Y-%m-%d"))
order_data_df = arrange(order_data_df, desc(Fieldwork.End))
slice_data_df = head(data_df, 10)
```

### 3

```
mean_df = as.data.frame(t(summarize(slice_data_df, across("Fine.Gael":"Other", mean))))
colnames(mean_df) = "ave"
mean_df
```

```
##                                    ave
## Fine.Gael                        0.253
## Fianna.Fail                      0.170
## Sinn.Fein                        0.307
## Labour.Party                     0.045
## Solidarity.People.Before.Profit  0.028
## Social.Democrats                 0.040
## Green.Party                      0.044
## Aontu                            0.022
## Renua.Ireland                       NA
## Independent.Alliance                NA
## Independents                        NA
## Other                               NA
```

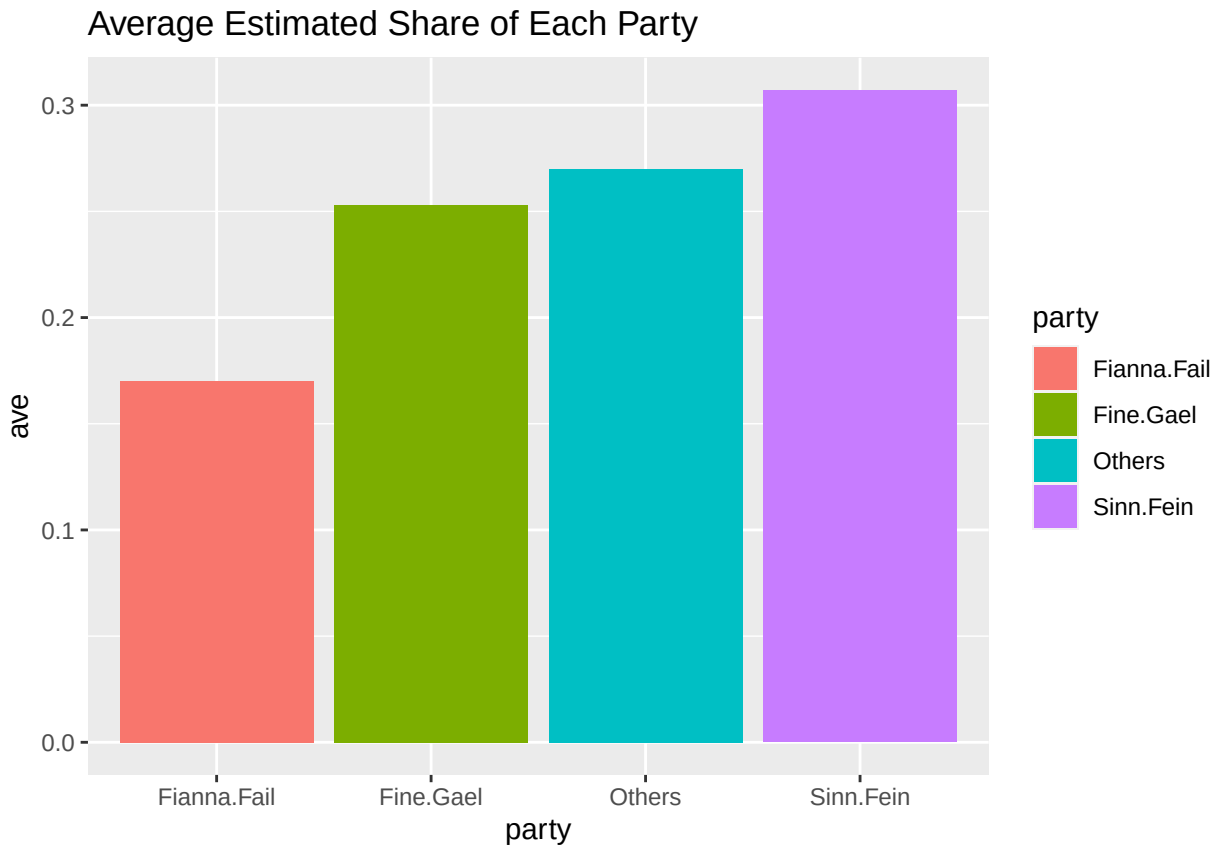
### 4

```
mean_df1 = filter(mean_df, ave > 0.06)
new_row = data.frame(ave = 1 - sum(mean_df1$ave), row.names = "Others")
mean_df1 = rbind(mean_df1, new_row)
mean_df1
```

```
##               ave
## Fine.Gael   0.253
## Fianna.Fail 0.170
## Sinn.Fein   0.307
## Others      0.270
```

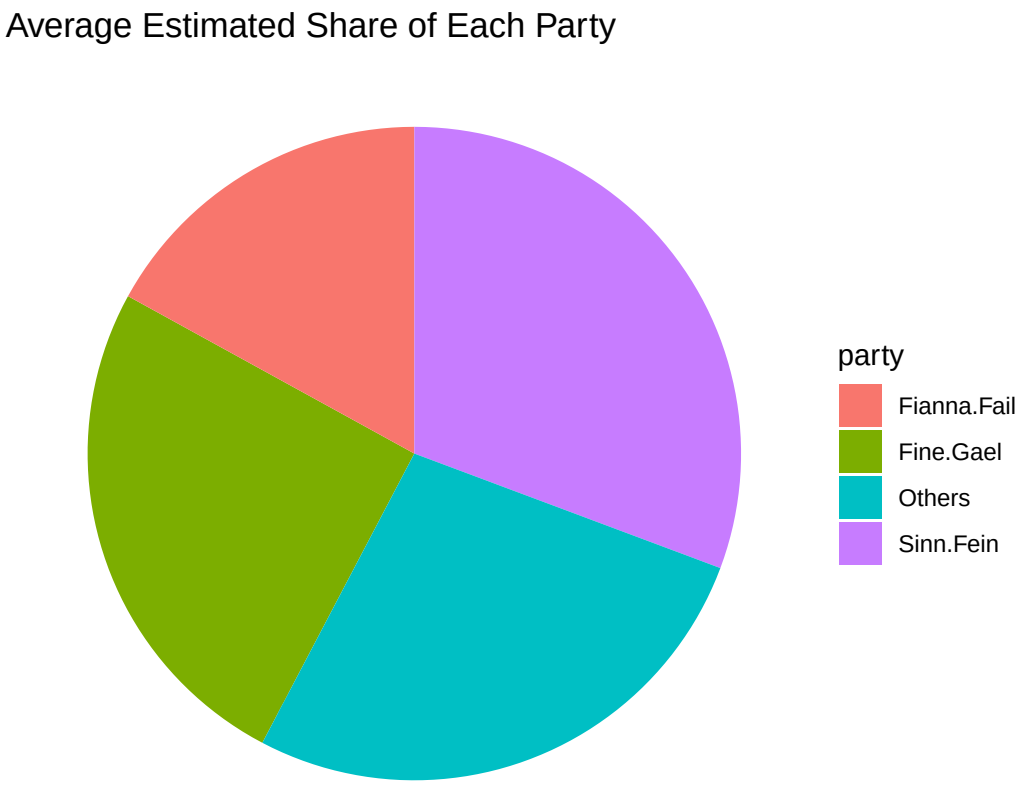
### 5

```
mean_df1$party = rownames(mean_df1)
ggplot(mean_df1, aes(x = party, y = ave, fill = party)) + geom_bar(stat = "identity") +
    ggtitle("Average Estimated Share of Each Party")
```




## 6

```
mean_df1$party = rownames(mean_df1)
ggplot(mean_df1, aes(x = "", y = ave, fill = party)) + geom_bar(stat = "identity") +
    coord_polar("y") + ggtitle("Average Estimated Share of Each Party") + theme_void()
```




## Exercise 2

### 3

```
data_df = read.csv("data/human-rights-index-vs-gdp-per-capita.csv")
```

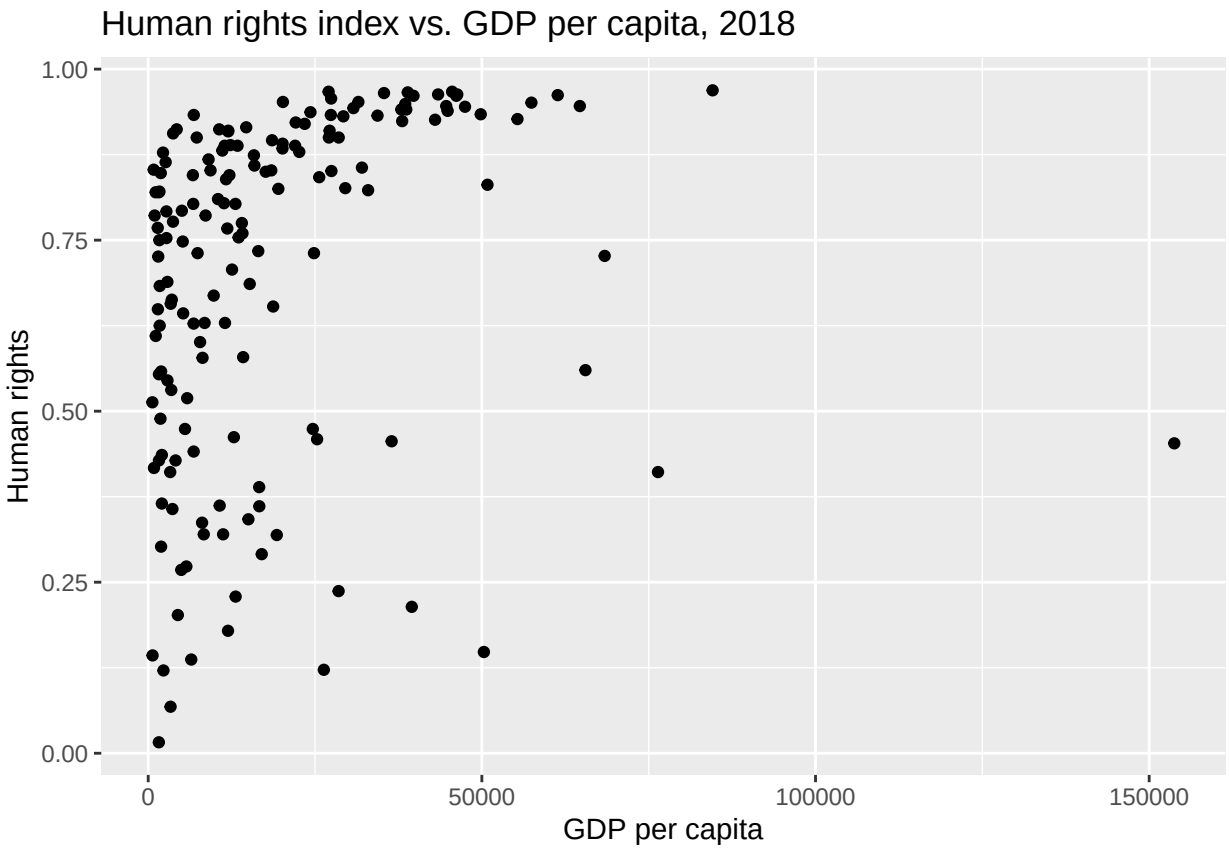
### 4

```
data2018_df = filter(data_df, Year == 2018)
data2018_df = rename(data2018_df, HRI = civ_libs_vdem_owid, GDPpc = GDP.per.capita,
    Population = Population..historical.estimates.)
```

### 5

```
ggplot(data2018_df, aes(x = GDPpc, y = HRI)) + geom_point() + ggtitle("Human rights index vs. GDP per ca
    labs(x = "GDP per capita", y = "Human rights")
```

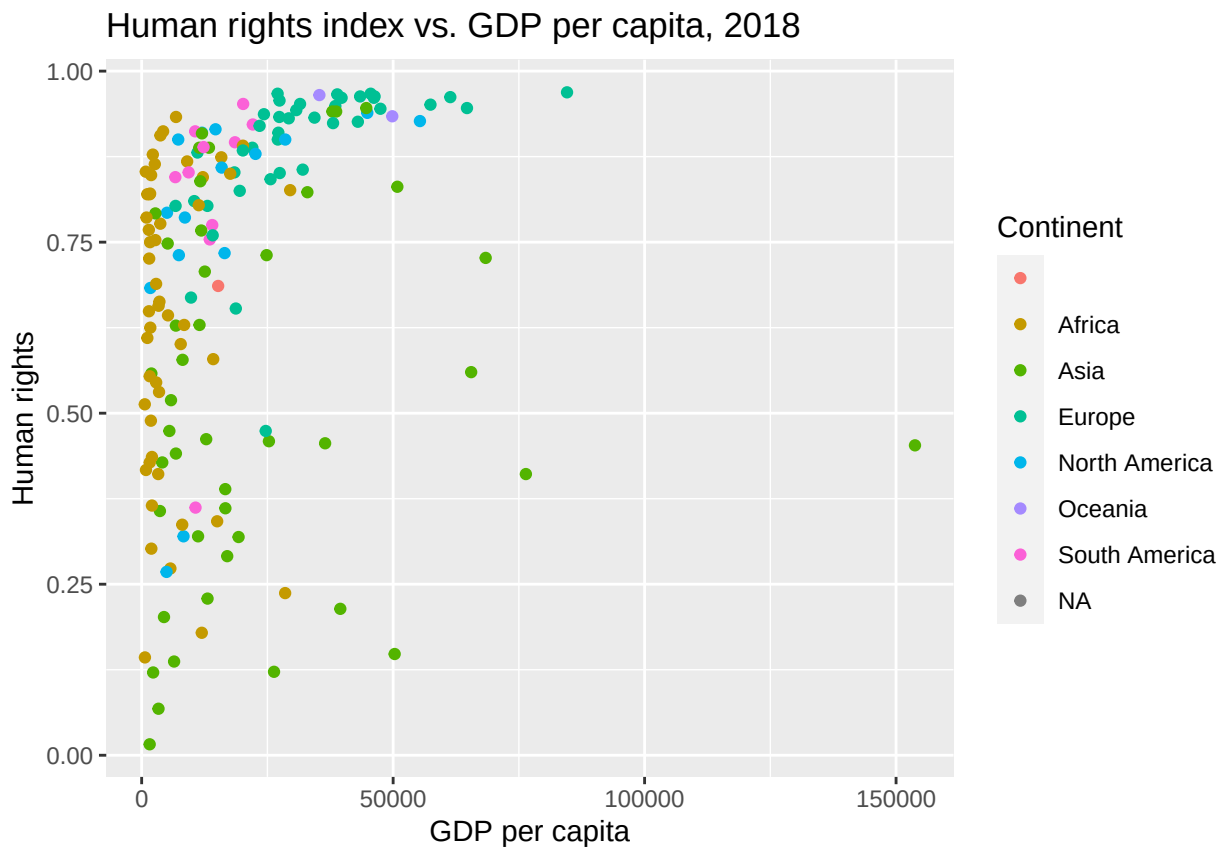
```
## Warning: Removed 108 rows containing missing values (‘geom_point()‘).
```




## 6

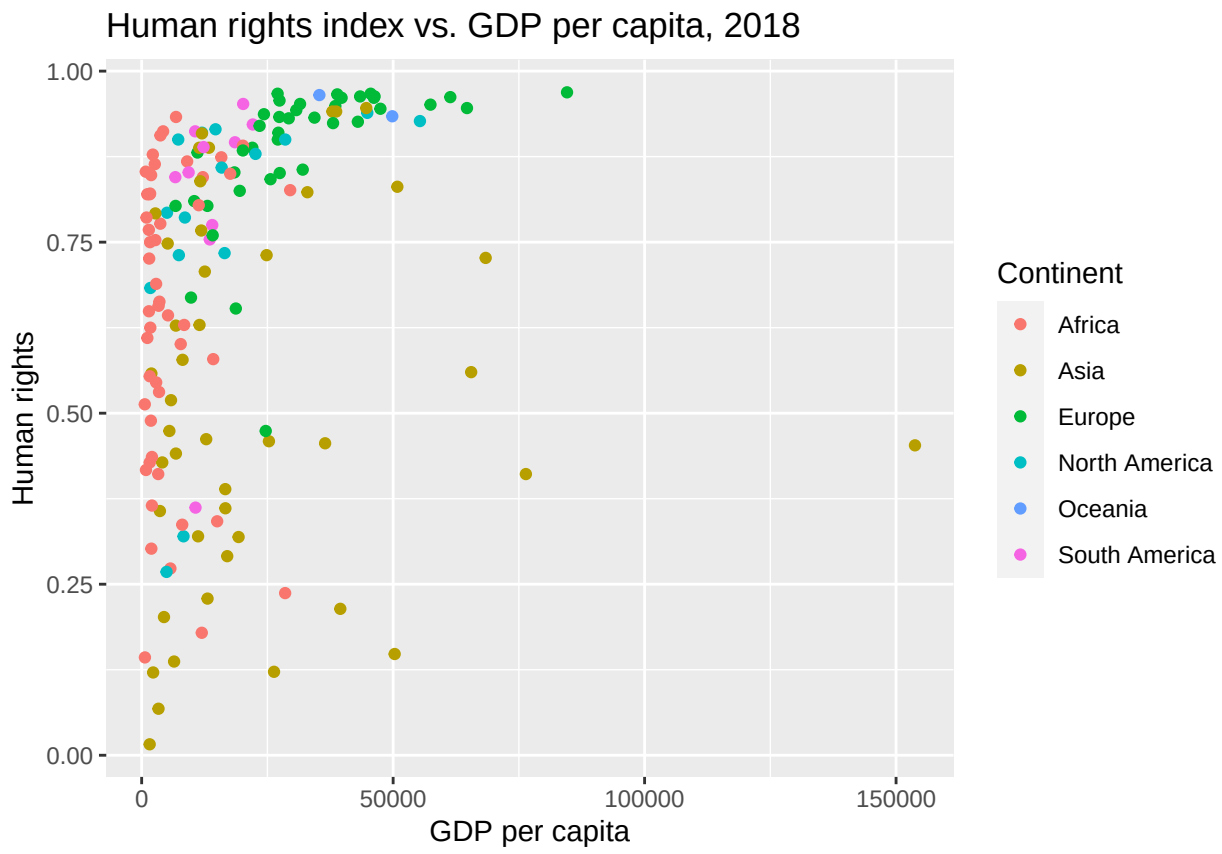
```
data2015_df = select(filter(data_df, Year == 2015), Entity, Continent)
data2018_df = select(data2018_df, !Continent)
data2018_df = merge(data2018_df, data2015_df, by = "Entity", all.x = TRUE)
ggplot(data2018_df, aes(x = GDPpc, y = HRI, color = Continent)) + geom_point() +
    ggtitle("Human rights index vs. GDP per capita, 2018") + labs(x = "GDP per capita",
    y = "Human rights")
```

```
## Warning: Removed 108 rows containing missing values (‘geom_point()‘).
```



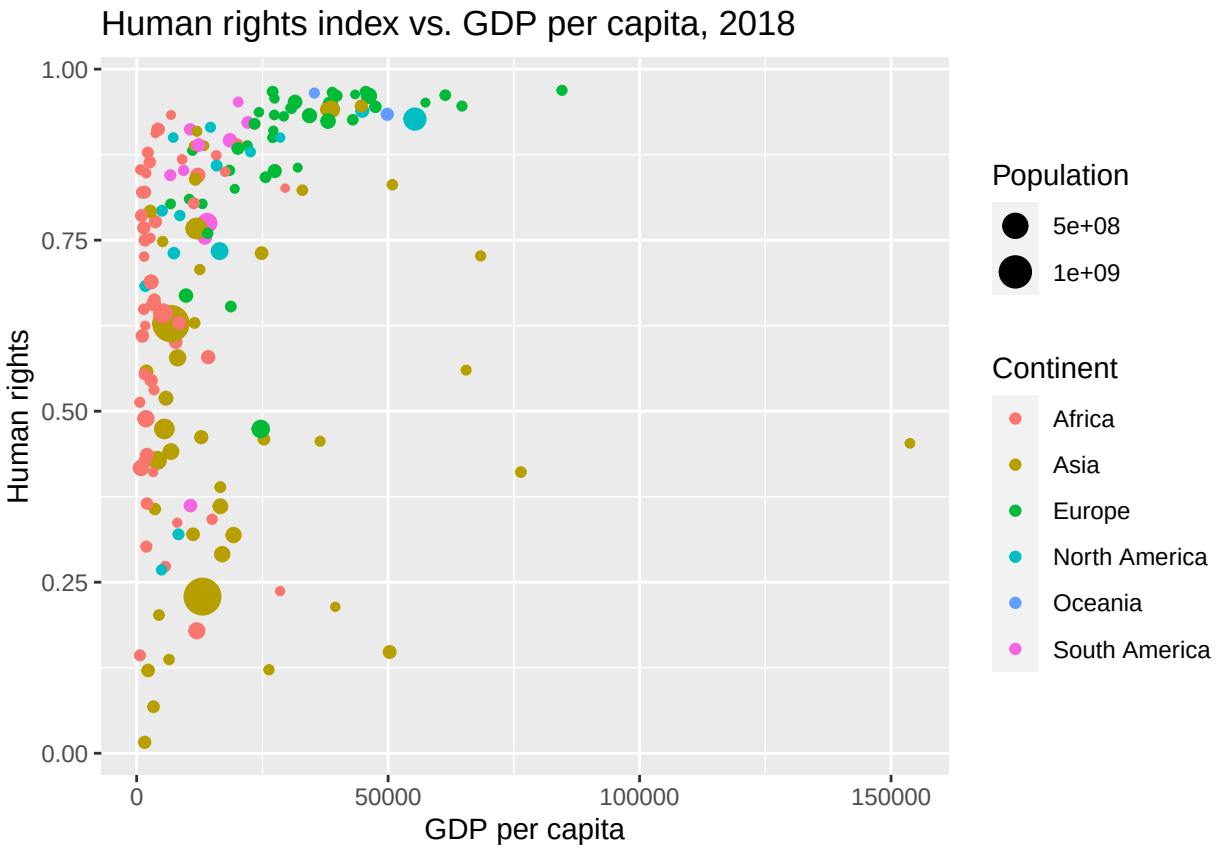

**7**

```
data2018_df = filter(data2018_df, !(is.na(GDPpc) | is.na(HRI) | Entity == "World"))
ggplot(data2018_df, aes(x = GDPpc, y = HRI, color = Continent)) + geom_point() +
    ggtitle("Human rights index vs. GDP per capita, 2018") + labs(x = "GDP per capita",
    y = "Human rights")
```




8

```
ggplot(data2018_df, aes(x = GDPpc, y = HRI, color = Continent, size = Population)) +
    geom_point() + ggtitle("Human rights index vs. GDP per capita, 2018") + labs(x = "GDP per capita",
    y = "Human rights")
```




## Exercise 3

```
heights = as.data.frame(heights)
```

### 1

```
heights1980 = filter(heights, year == 1980)
dplyr::count(heights1980, continent)
```

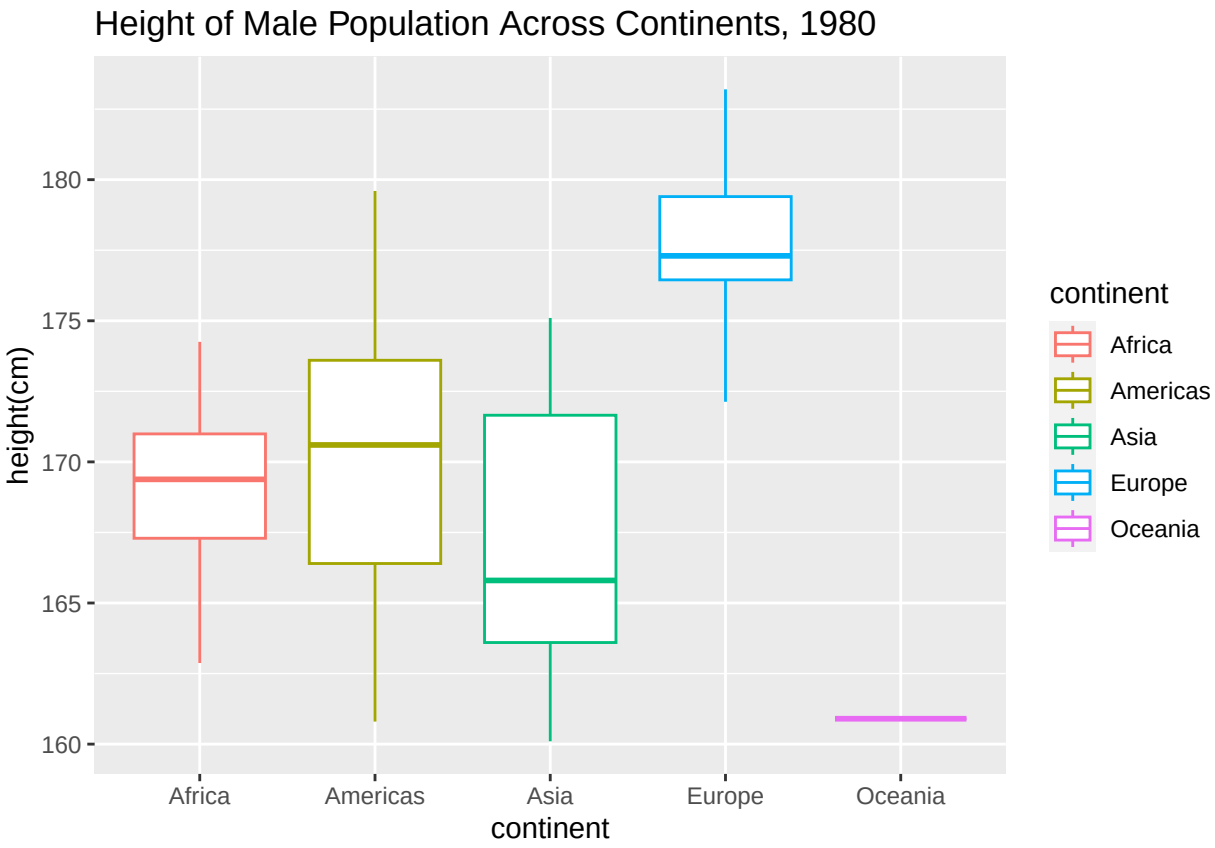
```
##   continent  n
## 1    Africa 34
## 2  Americas 17
## 3      Asia 21
## 4    Europe 19
## 5   Oceania  1
```

### 2

```
mean_h1980 = summarize(group_by(heights1980, continent), mean = mean(height_cm))
```

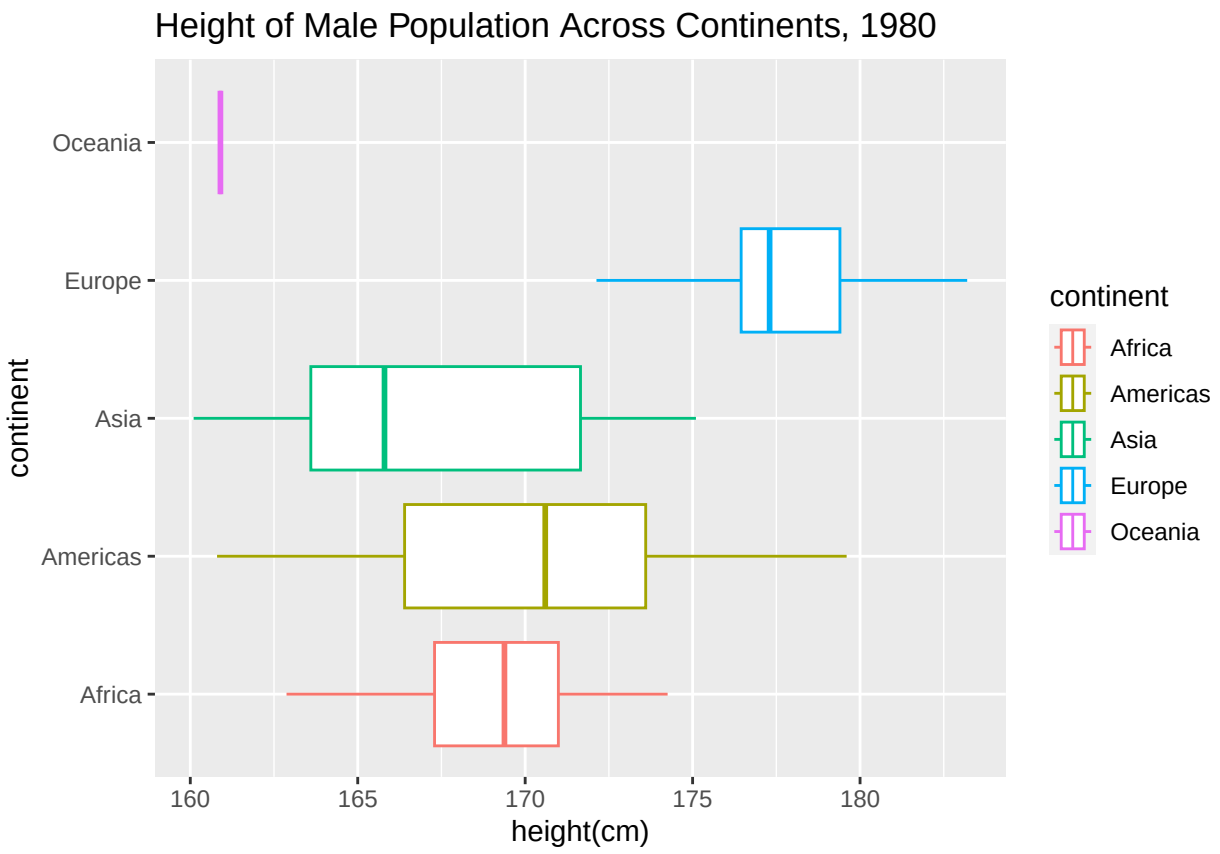
Vertical boxplot:

```
vb = ggplot(heights1980, aes(x = continent, y = height_cm, color = continent)) +
    geom_boxplot() + ggtitle("Height of Male Population Across Continents, 1980") +
    labs(y = "height(cm)")
vb
```



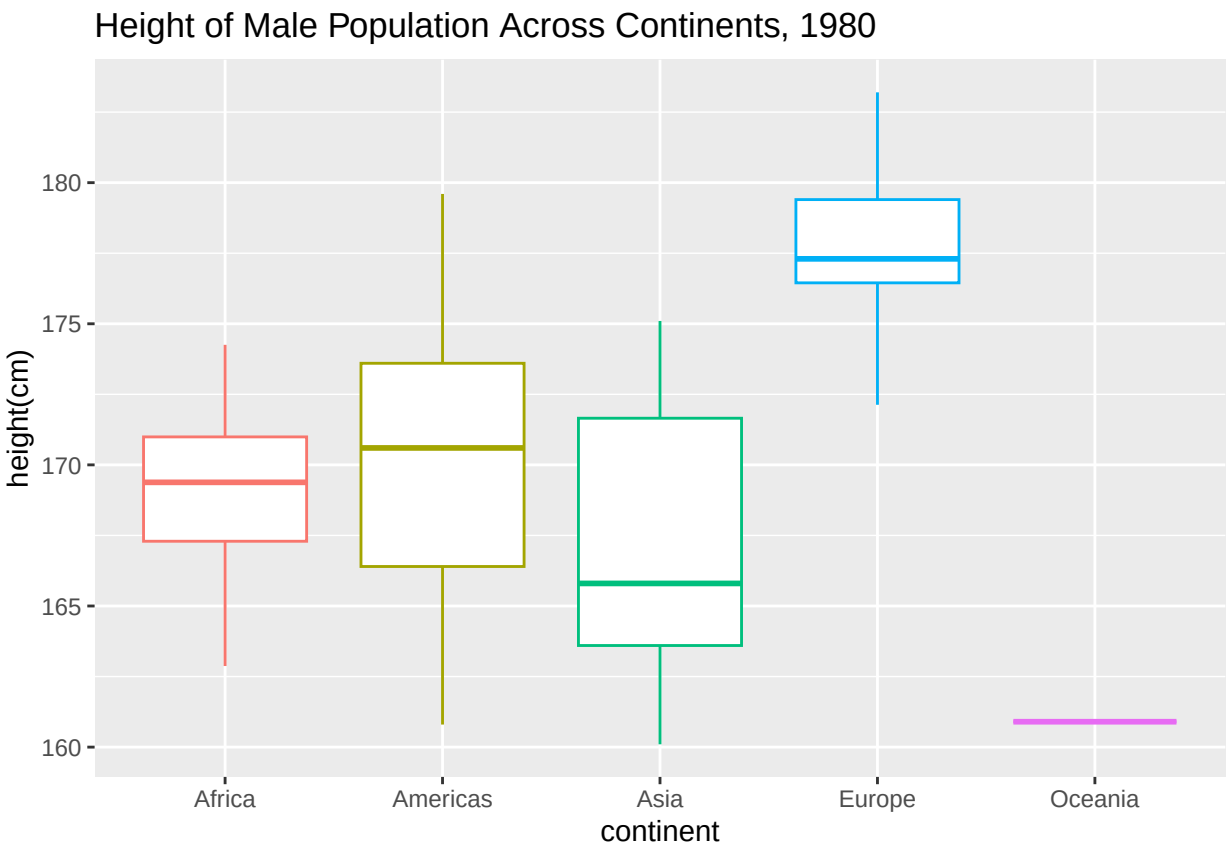

Horizontal boxplot:

```
hb = ggplot(heights1980, aes(x = height_cm, y = continent, color = continent)) +
    geom_boxplot() + ggtitle("Height of Male Population Across Continents, 1980") +
    labs(x = "height(cm)")
hb
```



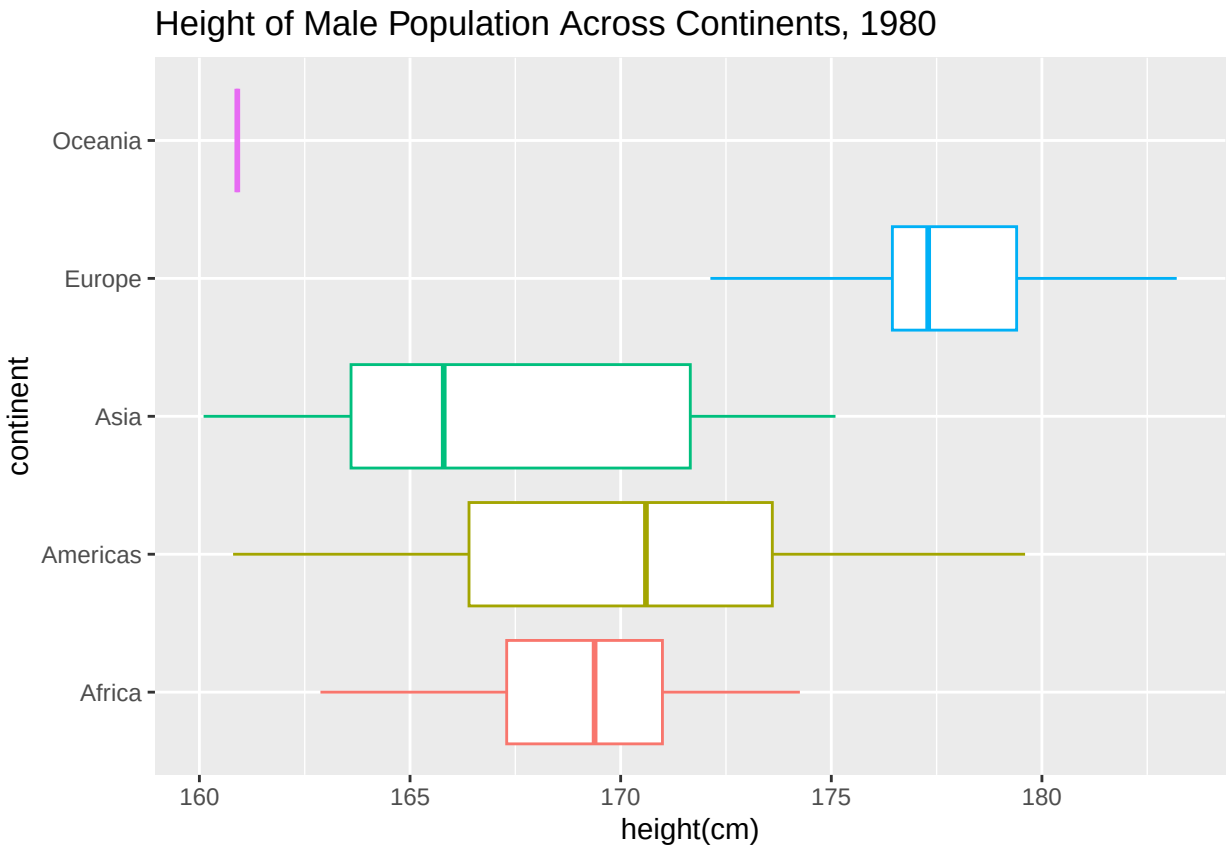

Vertical boxplot without legend:

```
vb + theme(legend.position = "none")
```




Horizontal boxplot without legend:

```
hb + theme(legend.position = "none")
```


Height of Male Population Across Continents, 1980
continent
Oceania
Europe
Asia
Americas
Africa
160
165
170
175
180
height(cm)