

Final Project

Xinyi Wang

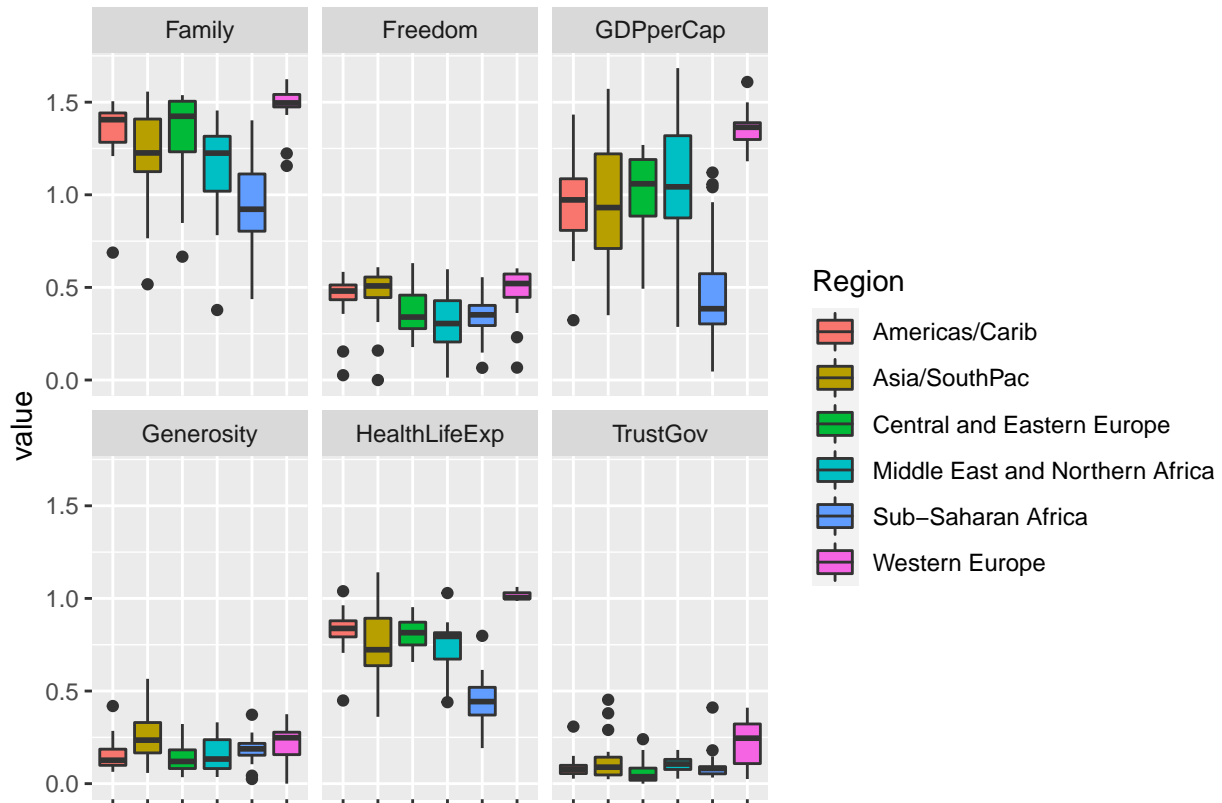
07/12/2020

```
library(tidyverse)
library(dplyr)
library(knitr)
happiness_data <-
  read_csv("D:\\Courses\\MATH 208\\Final Project\\Project_Happiness_data.csv"))
```

```
data_2019 <- happiness_data %>% filter(Year == 2019)
#(a)
new_data_2019 <- data_2019 %>%
  select( c("Region", "GDPperCap", "TrustGov", "Family", "Freedom",
            "HealthLifeExp", "Generosity")) %>%
  pivot_longer(cols = c("GDPperCap", "TrustGov",
                        "Family", "Freedom",
                        "HealthLifeExp", "Generosity"),
               names_to = "Scores")
new_data_2019
```

```
## # A tibble: 846 x 3
##   Region      Scores      value
##   <chr>      <chr>      <dbl>
## 1 Western Europe GDPperCap    1.34
## 2 Western Europe TrustGov     0.393
## 3 Western Europe Family       1.59
## 4 Western Europe Freedom     0.596
## 5 Western Europe HealthLifeExp 0.986
## 6 Western Europe Generosity    0.153
## 7 Western Europe GDPperCap    1.38
## 8 Western Europe TrustGov     0.41
## 9 Western Europe Family       1.57
## 10 Western Europe Freedom     0.592
## # ... with 836 more rows
```

```
new_data_2019 %>% ggplot(aes(x=Region, fill = Region, y = value)) +
  geom_boxplot() + facet_wrap(~Scores) + labs(x="") +
  theme(axis.text.x = element_blank())
```



Task1

(a) The numerical summary and the plots are provided as above.

In my opinion the boxplot is the most informative plot in showing the relationship between the multiple quantitative data and one qualitative data. In order to get the boxplot, we need to get the pivot_longer version of the data.

(b) The association between score and region varies amongst the scores. Based on the locations of boxes in each score, where the 25th to the 75th percentile belong, different scores show the different trends among regions. The median of each score in each region also shows the different association.

```
total_happiness <- happiness_data %>% group_by(Country,Year)%>%
  mutate(TotalScore = sum(GDPperCap,TrustGov,Family,
    Freedom,HealthLifeExp,Generosity))%>% ungroup()%>%
  summarise(Country,Year,Region,TotalScore)
total_happiness
```

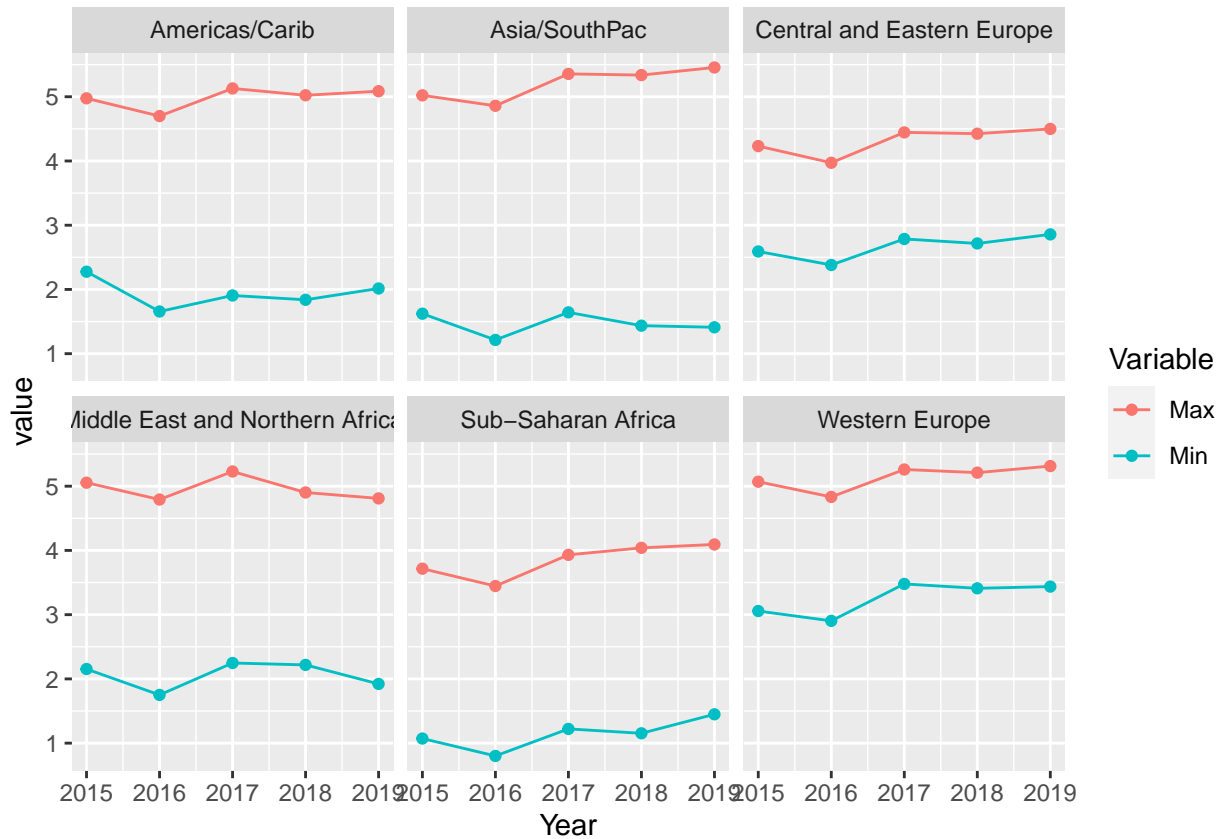
```
## # A tibble: 705 x 4
##   Country      Year Region      TotalScore
##   <chr>      <dbl> <chr>      <dbl>
## 1 Switzerland 2015 Western Europe 5.07
## 2 Iceland     2015 Western Europe 4.86
## 3 Denmark     2015 Western Europe 5.04
## 4 Norway      2015 Western Europe 5.06
## 5 Canada      2015 Americas/Carib 4.98
```

```
## 6 Finland      2015 Western Europe      4.79
## 7 Netherlands  2015 Western Europe      4.91
## 8 Sweden       2015 Western Europe      4.99
## 9 New Zealand  2015 Asia/SouthPac       5.02
## 10 Australia   2015 Asia/SouthPac       5.02
## # ... with 695 more rows
```

```
extreme <- total_happiness %>% group_by(Region,Year)%>%
  mutate(Max = max(TotalScore,na.rm = TRUE),
         Min = min(TotalScore,na.rm = TRUE))%>%
  ungroup()%>%summarise(Year,Region,Max,Min) %>% unique()
extreme
```

```
## # A tibble: 30 x 4
##   Year Region      Max    Min
##   <dbl> <chr>    <dbl> <dbl>
## 1  2015 Western Europe    5.07  3.06
## 2  2015 Americas/Carib    4.98  2.28
## 3  2015 Asia/SouthPac     5.02  1.62
## 4  2015 Middle East and Northern Africa 5.05  2.15
## 5  2015 Central and Eastern Europe    4.23  2.59
## 6  2015 Sub-Saharan Africa    3.72  1.07
## 7  2016 Western Europe    4.83  2.90
## 8  2016 Americas/Carib    4.70  1.66
## 9  2016 Asia/SouthPac     4.86  1.21
## 10 2016 Middle East and Northern Africa 4.79  1.75
## # ... with 20 more rows
```

```
extreme_for_plot <- extreme %>%
  pivot_longer(cols=c(Max, Min), names_to = "Variable")
extreme_for_plot %>% group_by(Region) %>%
  ggplot(aes(x=Year,y=value,col= Variable))+
  geom_line()+geom_point()+facet_wrap(~Region)
```



Task2

(b) The minimum and maximum happiness scores doesn't change much over time. In each region, the maximum and minimum values all change around a value in a small range. And all values decrease in Year 2016, and increase after Year 2016.

Task3 (a)

```
average_by_country<-total_happiness %>% group_by(Country) %>%
  mutate(Average = mean(TotalScore))%>% ungroup()%>%
  summarise(Country,Average)%>%arrange(desc(Average))%>%
  unique()
average_by_country%>% head(10)%>%kable()
```

Country	Average
Singapore	5.161286
Norway	5.134778
New Zealand	5.099459
Switzerland	5.095652
Denmark	5.083876
Australia	5.054115
Luxembourg	5.051504
Ireland	5.044687
Sweden	5.023141
Canada	4.982497

(b)

```
tbl_for_diff <- total_happiness %>%  
  filter(Year == 2015 | Year == 2019)%>%  
  pivot_wider(id_cols=Country,names_from=Year,  
              values_from=TotalScore)  
diff_happiness <- tbl_for_diff %>%  
  mutate(Diff = pull(tbl_for_diff,"2019") -  
           pull(tbl_for_diff,"2015")) %>%  
  summarise(Country,Diff)%>%arrange(desc(Diff))  
diff_happiness %>% head(10) %>% kable()
```

Country	Diff
Bangladesh	0.87266
Myanmar	0.82002
Kosovo	0.80622
Croatia	0.76760
Pakistan	0.64243
Montenegro	0.60988
Nepal	0.60395
India	0.60145
Serbia	0.60092
Bosnia and Herzegovina	0.59800