Final Project

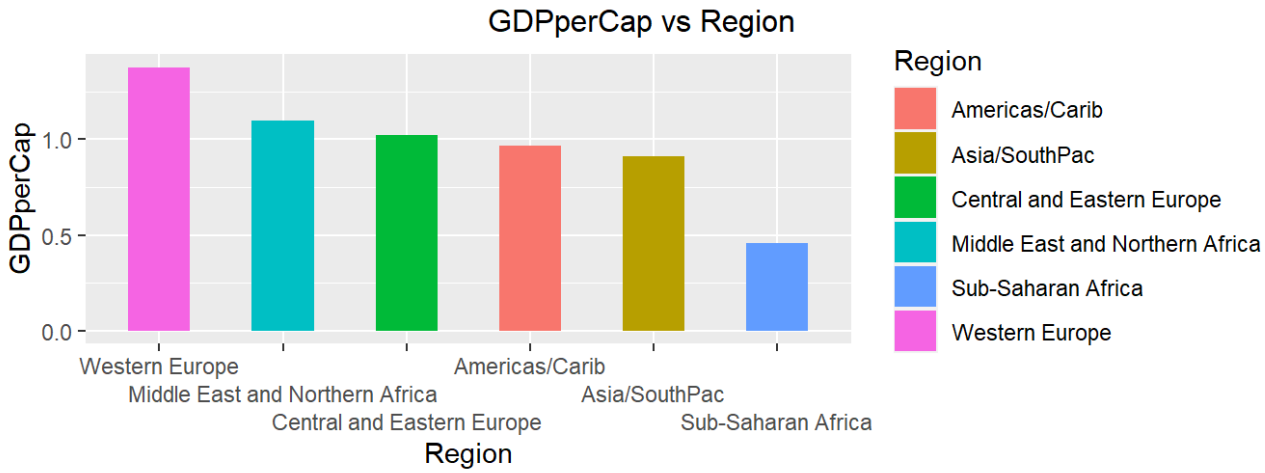
Wassim Wazzi

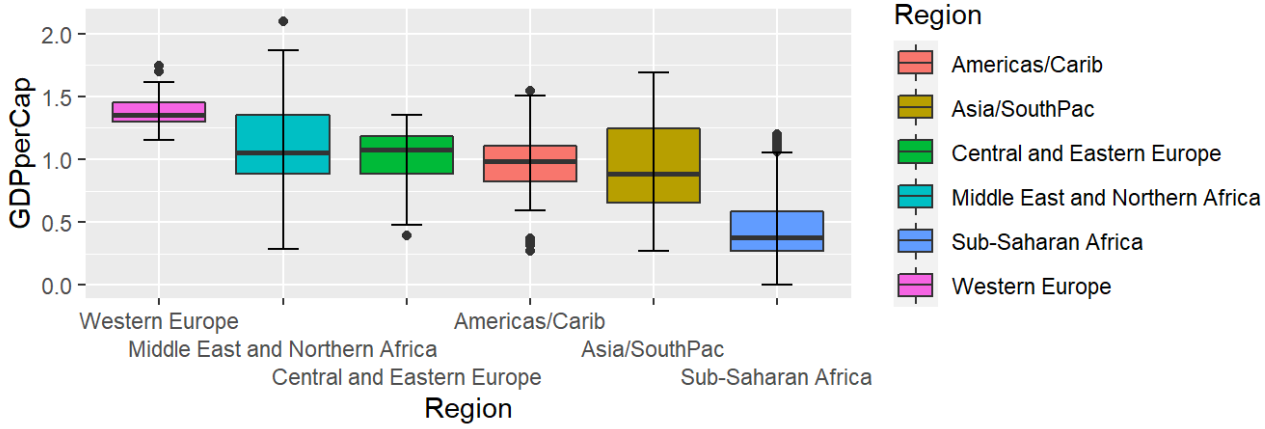
01/12/2020

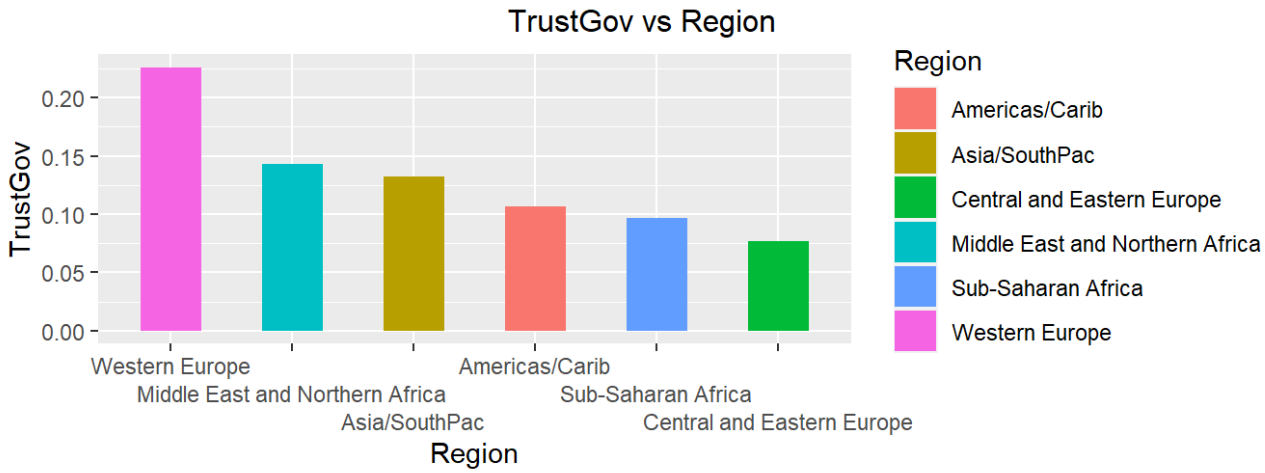
TASK 1:

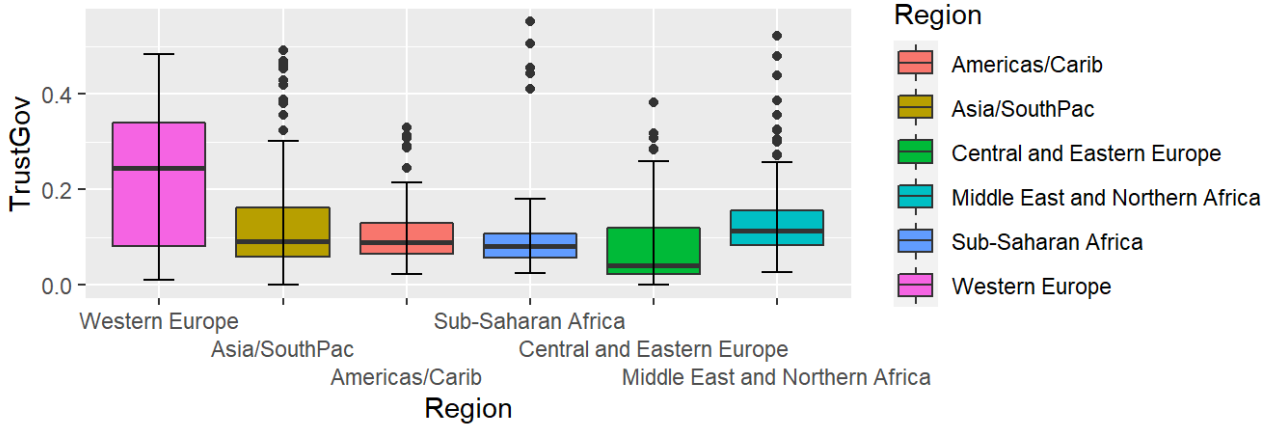
Table summarising the Mean, Median, and Standard Deviation for each characteristic in each Region in 2019

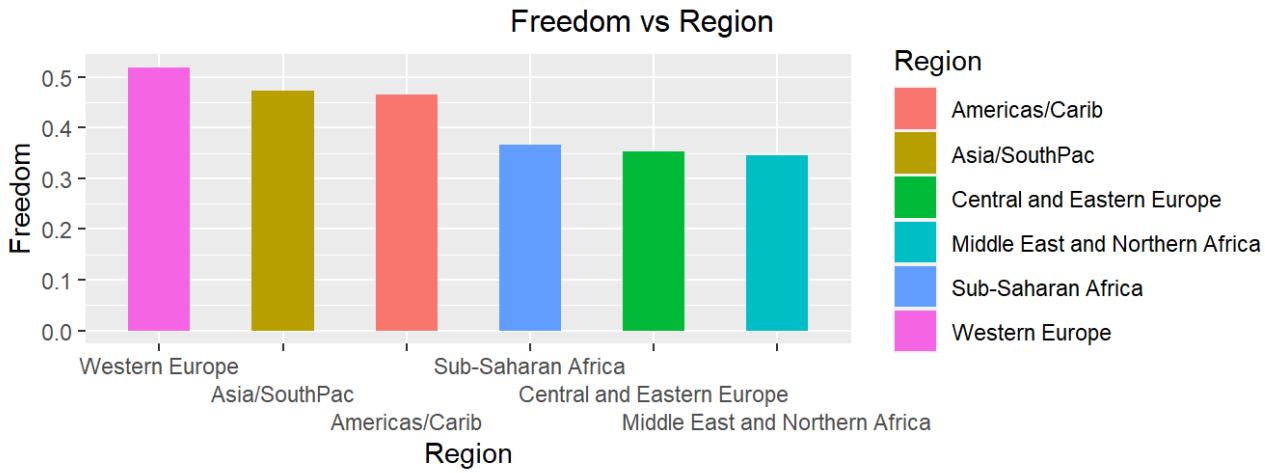
characteristic	Year	Americas/Carib	Asia/SouthPac	Central and Eastern Europe	Middle East and Northern Africa	Sub-Saharan Africa	Western Europe
TrustGov_Sd	2019	0.0623573	0.1197383	0.0820340	0.0970753	0.0779829	0.1360714
TrustGov_Med	2019	0.0889905	0.0922269	0.0399421	0.1141907	0.0812400	0.2456716
TrustGov_Avg	2019	0.1067851	0.1320434	0.0764541	0.1427490	0.0969380	0.2258620
HealthLifeExp_Sd	2019	0.1373746	0.2086569	0.1035949	0.1287947	0.1507297	0.0744276
HealthLifeExp_Med	2019	0.7060781	0.6370000	0.7024793	0.6601500	0.2956377	0.8840000
HealthLifeExp_Avg	2019	0.7040199	0.6629144	0.6960031	0.6645420	0.3020207	0.8934446
Generosity_Sd	2019	0.1076228	0.1652456	0.0882324	0.1115547	0.0735053	0.1313928
Generosity_Med	2019	0.1797800	0.3118500	0.1423600	0.1480000	0.2079485	0.2743550
Generosity_Avg	2019	0.1978436	0.3189631	0.1549616	0.1778799	0.2094616	0.2722297
GDPperCap_Sd	2019	0.2479531	0.3521000	0.2252283	0.3429141	0.2936298	0.1155637
GDPperCap_Med	2019	0.9855000	0.8854164	1.0738488	1.0526600	0.3758465	1.3554690
GDPperCap_Avg	2019	0.9668096	0.9091126	1.0195455	1.0999229	0.4574615	1.3752193
Freedom_Sd	2019	0.1276590	0.1413834	0.1316191	0.1533834	0.1174694	0.1447031
Freedom_Med	2019	0.4896350	0.5080000	0.3503400	0.3120000	0.3810000	0.5651200
Freedom_Avg	2019	0.4660353	0.4729501	0.3528401	0.3456892	0.3664873	0.5183998
Family_Sd	2019	0.2398357	0.3430227	0.2912919	0.3065098	0.2798173	0.2076877
Family_Med	2019	1.2486150	1.1250000	1.2272950	1.0250000	0.8910000	1.4300000
Family_Avg	2019	1.2101941	1.0582244	1.1747695	0.9925362	0.8571959	1.3482706

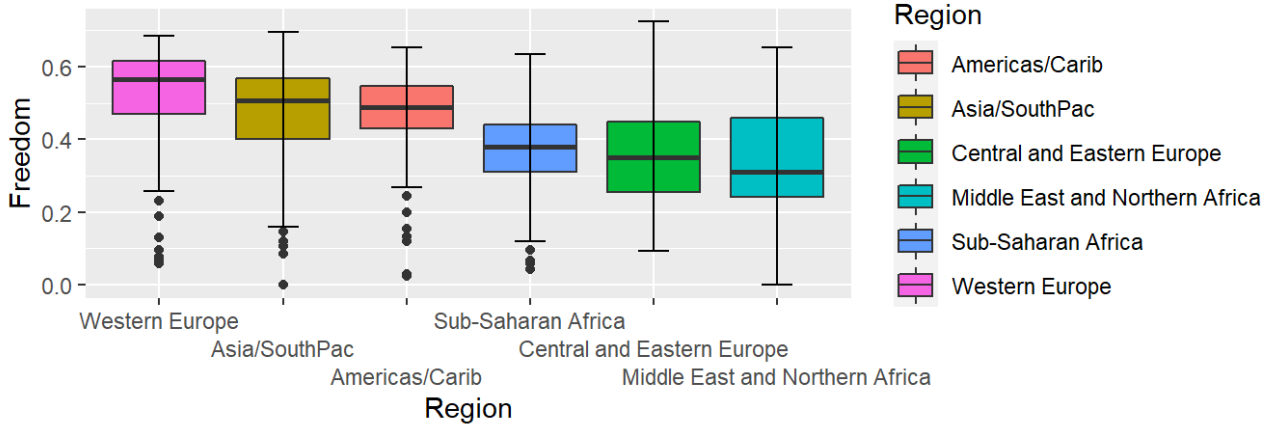

GDPperCap vs Region
GDPperCap
1.0
0.5
0.0
Western Europe
Middle East and Northern Africa
Central and Eastern Europe
Americas/Carib
Asia/SouthPac
Sub-Saharan Africa
Region
Region
Americas/Carib
Asia/SouthPac
Central and Eastern Europe
Middle East and Northern Africa
Sub-Saharan Africa
Western Europe

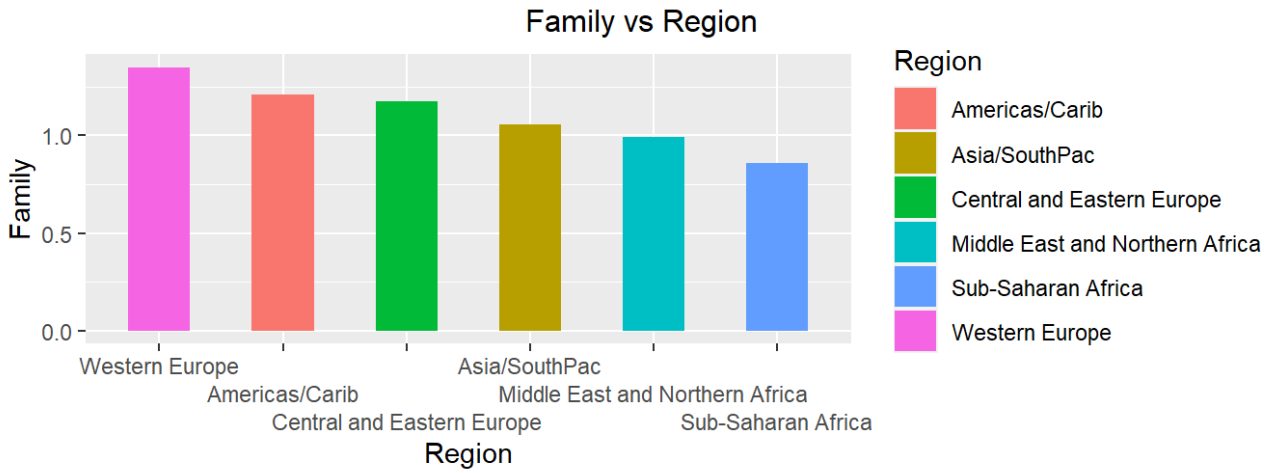

GDPperCap
2.0
1.5
1.0
0.5
0.0
Western Europe
Middle East and Northern Africa
Central and Eastern Europe
Americas/Carib
Asia/SouthPac
Sub-Saharan Africa
Region
Region
Americas/Carib
Asia/SouthPac
Central and Eastern Europe
Middle East and Northern Africa
Sub-Saharan Africa
Western Europe

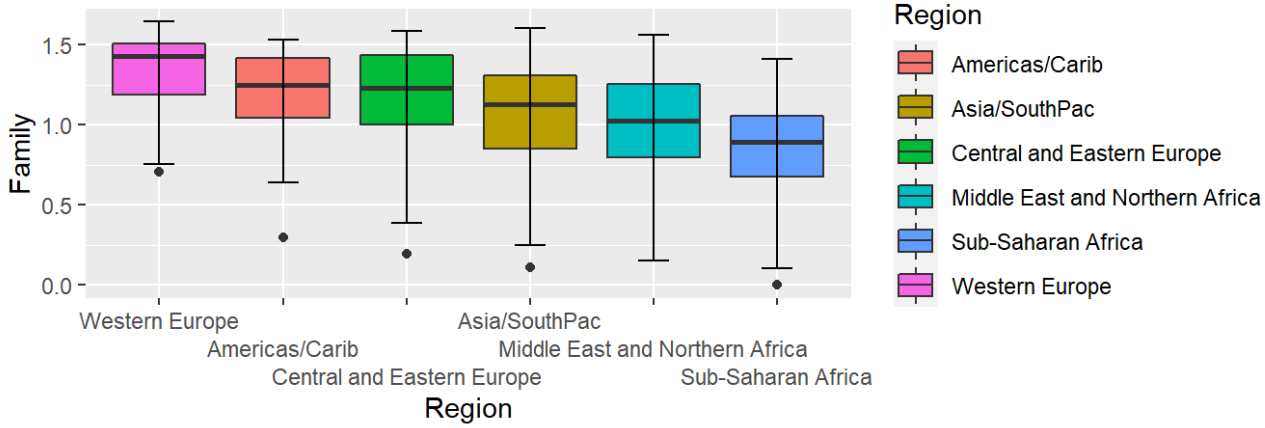

TrustGov vs Region
TrustGov
0.20
0.15
0.10
0.05
0.00
Western Europe
Middle East and Northern Africa
Asia/SouthPac
Americas/Carib
Sub-Saharan Africa
Central and Eastern Europe
Region
Region
Americas/Carib
Asia/SouthPac
Central and Eastern Europe
Middle East and Northern Africa
Sub-Saharan Africa
Western Europe

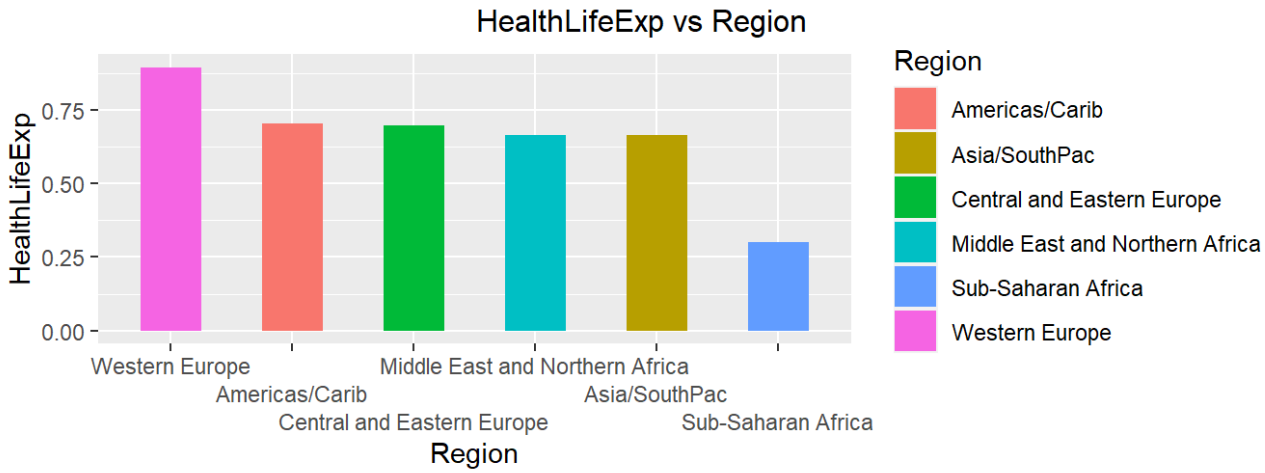

TrustGov
0.4
0.2
0.0
Western Europe
Asia/SouthPac
Americas/Carib
Sub-Saharan Africa
Central and Eastern Europe
Middle East and Northern Africa
Region
Region
Americas/Carib
Asia/SouthPac
Central and Eastern Europe
Middle East and Northern Africa
Sub-Saharan Africa
Western Europe

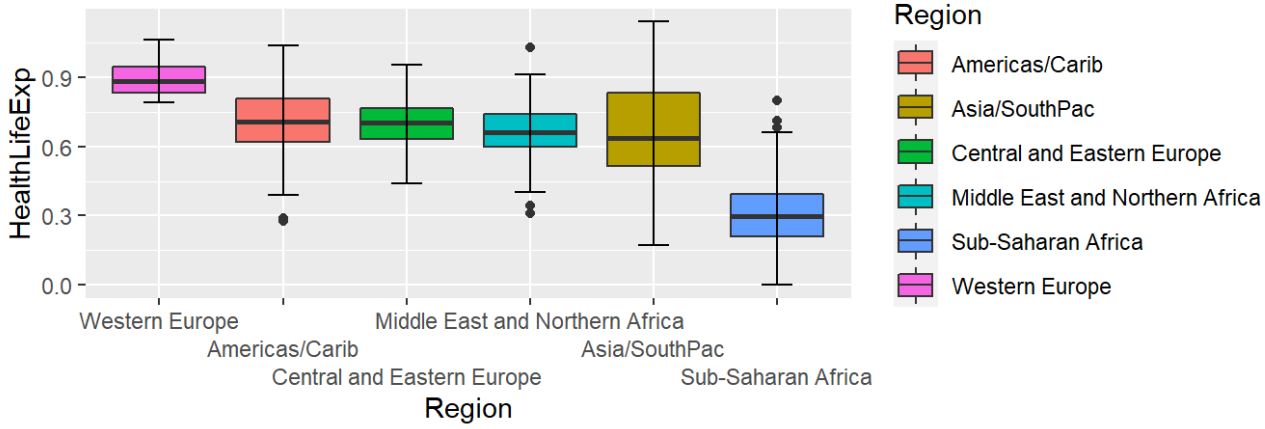

Freedom vs Region
Freedom
0.0
0.1
0.2
0.3
0.4
0.5
Western Europe
Asia/SouthPac
Americas/Carib
Sub-Saharan Africa
Central and Eastern Europe
Middle East and Northern Africa
Region
Region
Americas/Carib
Asia/SouthPac
Central and Eastern Europe
Middle East and Northern Africa
Sub-Saharan Africa
Western Europe

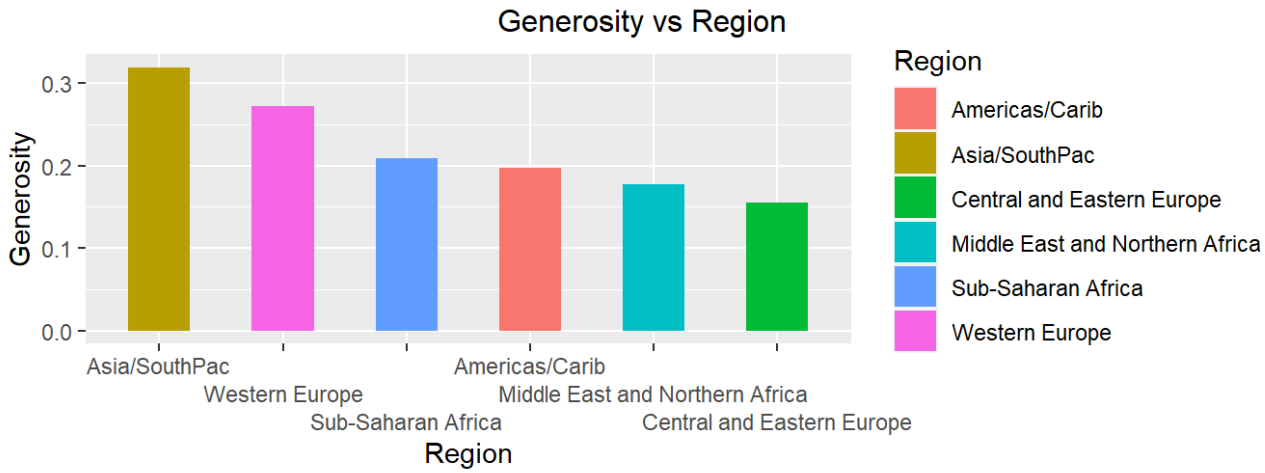

Freedom
0.0
0.2
0.4
0.6
Western Europe
Asia/SouthPac
Americas/Carib
Sub-Saharan Africa
Central and Eastern Europe
Middle East and Northern Africa
Region
Region
Americas/Carib
Asia/SouthPac
Central and Eastern Europe
Middle East and Northern Africa
Sub-Saharan Africa
Western Europe

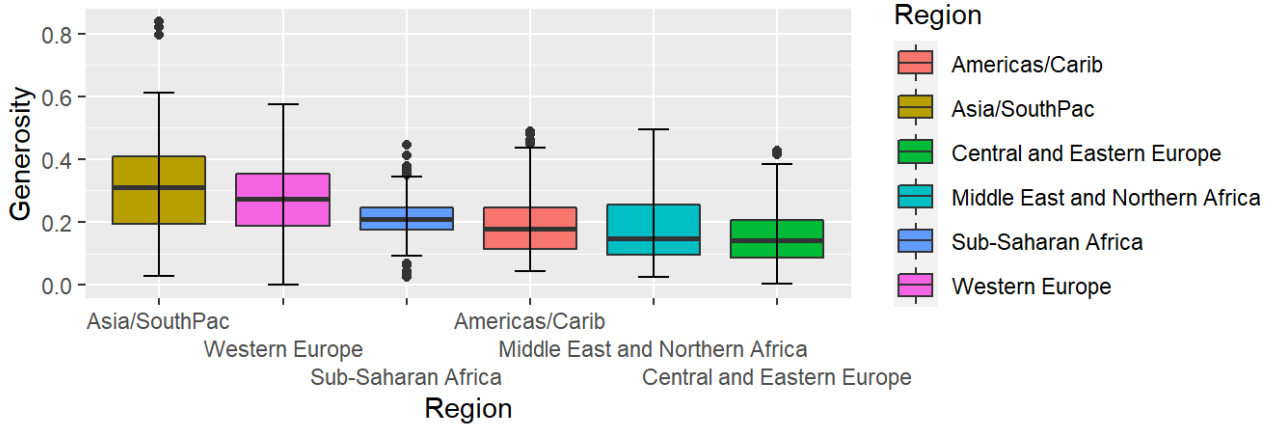

Family vs Region
Family
0.0
0.5
1.0
Western Europe
Americas/Carib
Central and Eastern Europe
Asia/SouthPac
Middle East and Northern Africa
Sub-Saharan Africa
Region
Region
Americas/Carib
Asia/SouthPac
Central and Eastern Europe
Middle East and Northern Africa
Sub-Saharan Africa
Western Europe


Family
0.0
0.5
1.0
1.5
Western Europe
Americas/Carib
Central and Eastern Europe
Asia/SouthPac
Middle East and Northern Africa
Sub-Saharan Africa
Region
Region
Americas/Carib
Asia/SouthPac
Central and Eastern Europe
Middle East and Northern Africa
Sub-Saharan Africa
Western Europe


HealthLifeExp vs Region
HealthLifeExp
0.00
0.25
0.50
0.75
Western Europe
Americas/Carib
Central and Eastern Europe
Middle East and Northern Africa
Asia/SouthPac
Sub-Saharan Africa
Region
Region
Americas/Carib
Asia/SouthPac
Central and Eastern Europe
Middle East and Northern Africa
Sub-Saharan Africa
Western Europe


HealthLifeExp
0.0
0.3
0.6
0.9
Western Europe
Americas/Carib
Central and Eastern Europe
Middle East and Northern Africa
Asia/SouthPac
Sub-Saharan Africa
Region
Region
Americas/Carib
Asia/SouthPac
Central and Eastern Europe
Middle East and Northern Africa
Sub-Saharan Africa
Western Europe


Generosity vs Region
Generosity
0.0
0.1
0.2
0.3
Asia/SouthPac
Western Europe
Sub-Saharan Africa
Americas/Carib
Middle East and Northern Africa
Central and Eastern Europe
Region
Region
Americas/Carib
Asia/SouthPac
Central and Eastern Europe
Middle East and Northern Africa
Sub-Saharan Africa
Western Europe


Generosity
0.0
0.2
0.4
0.6
0.8
Asia/SouthPac
Western Europe
Sub-Saharan Africa
Americas/Carib
Middle East and Northern Africa
Central and Eastern Europe
Region
Region
Americas/Carib
Asia/SouthPac
Central and Eastern Europe
Middle East and Northern Africa
Sub-Saharan Africa
Western Europe

a.

If we take a look at the summary table provided, it is clear that the values for each characteristic differ in each region in terms of average, median and standard deviation, this seems to suggest that there is an association between the scores and the region.

The trust in government is very high in Western Europe(highest average of 0.226) but the spread is very high if we look at the boxplot (25% below 0.1 and 25% above 0.3 and reaching 0.5). The averages are the lowest in Africa and Central Europe with almost 75% of countries at or below 0.1, while it is medium in other regions, so it seems that this characteristic is strongly associated to the region.

The Health Life Excpectancy average is similar in all regions, except for it being high in Western Europe and very low in Africa. Conversely, the standard deviation is low in the former and high in the latter, and also very high in Asia. The median values are similar to the average values in all the regions. A medium strength association is observed between the region and life expectancy, this suggests that medical care and health is not too different in all regions, except for Africa having very poor life expectancy and all countries in Western Europe guaranteeing long lives to its citizens.

In terms of Generosity, it is Asia that leads the way with the highest value, followed by Western Europe and the rest can be seen in the bar plot. Similarly, there is a clear association between the region and generosity levels, given the different cultures and traditions in the regions of the world, this association is not surprising.

Again, we can see an association with the GDPperCap and the region, with Western Europe having the highest score and Africa the lowest. It can be noted too that the deviation is high in the ME/NA and Asian regions if we look at the Sd numbers and the length of the boxplots, while there is very little spread in Western EUrope with most countries having high GDPperCap.

The freedom score is mildly associated with the region, there is a long tail in the boxplots for the top 3 countries, and similarly there is a large spread of values in all the regions which implies that the countries of the same region have different levels of freedom.

The family score seems to have the weakest association with similar values in all regions, except for Africa having the lowest average again. Only the spread seems to differ in the regions (by the boxplot), with median values quite similar

b.

The scores seem to follow a similar trend in all region, with Western Europe topping the score in all categories except for Family, with most countries in Western Europe having high scores as well. Africa, Central Europe and ME/NA have the lowest scores and lots of deviation from the median score, showing inequalities between different countries in the same region. Americas and Asia are in the upper middle.

TASK 2:

A table showing the happiness score for each country in every year

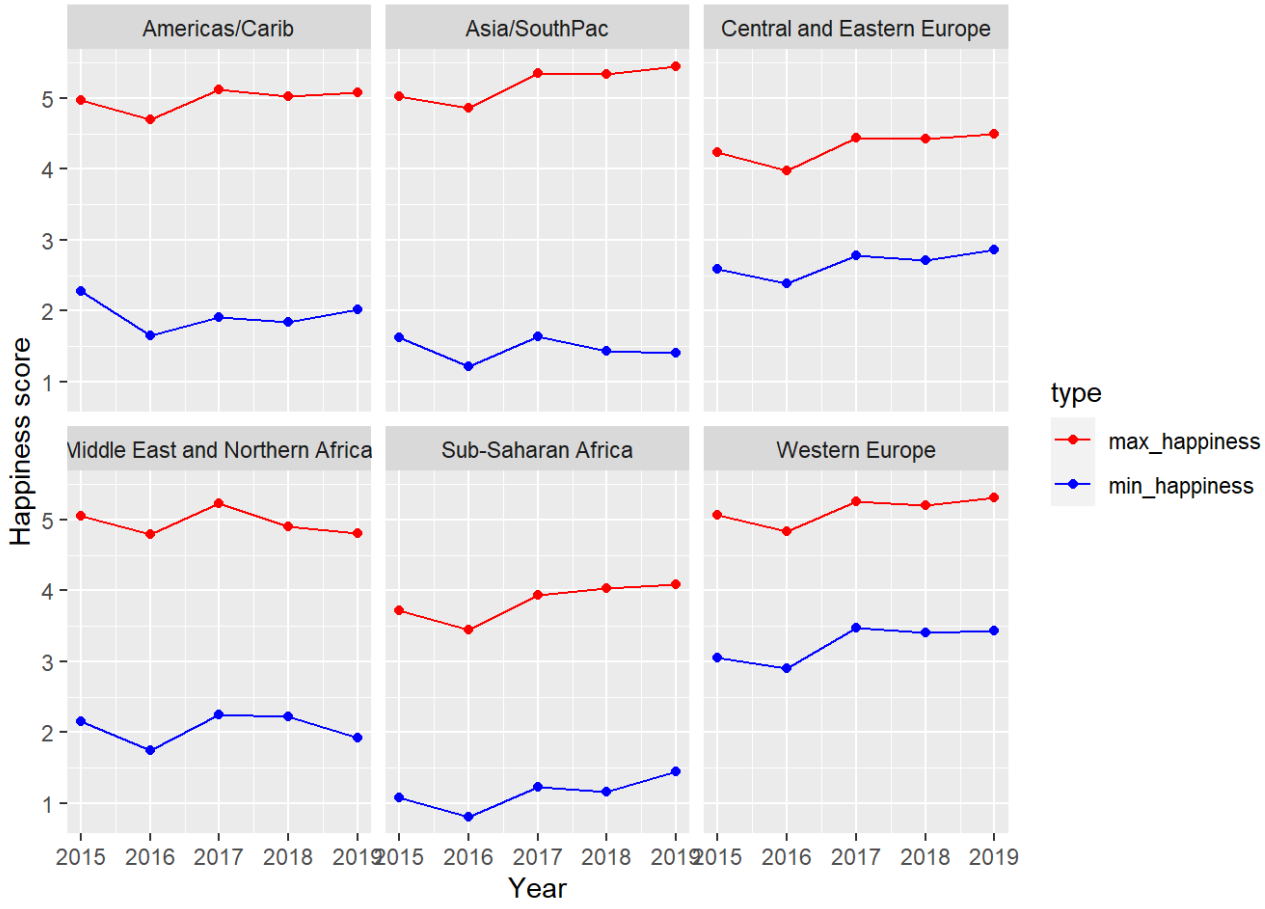
```
## # A tibble: 704 x 3
##    Country      Year Happiness
##    <chr>       <dbl>     <dbl>
##  1 Switzerland  2015      5.07
##  2 Iceland      2015      4.86
##  3 Denmark      2015      5.04
##  4 Norway       2015      5.06
##  5 Canada       2015      4.98
##  6 Finland      2015      4.79
##  7 Netherlands  2015      4.91
##  8 Sweden       2015      4.99
##  9 New Zealand  2015      5.02
## 10 Australia    2015      5.02
## # ... with 694 more rows
```

Table and line plot showing the max and min happiness for each region in each year

Max and Min Happiness in Each Region

Region	Year	max_happiness	min_happiness
Americas/Carib	2015	4.975180	2.276090
Asia/SouthPac	2015	5.021830	1.622450
Central and Eastern Europe	2015	4.231900	2.591240
Middle East and Northern Africa	2015	5.054560	2.153620
Sub-Saharan Africa	2015	3.715630	1.071520
Western Europe	2015	5.069580	3.056480
Americas/Carib	2016	4.699180	1.656580
Asia/SouthPac	2016	4.858920	1.214180
Central and Eastern Europe	2016	3.972480	2.380920
Middle East and Northern Africa	2016	4.792300	1.750630
Sub-Saharan Africa	2016	3.445550	0.800490
Western Europe	2016	4.832930	2.903370
Americas/Carib	2017	5.129123	1.905827
Asia/SouthPac	2017	5.355699	1.642976
Central and Eastern Europe	2017	4.444907	2.785860
Middle East and Northern Africa	2017	5.229065	2.247503
Sub-Saharan Africa	2017	3.931327	1.221511

Region	Year	max_happiness	min_happiness
Western Europe	2017	5.260052	3.478122
Americas/Carib	2018	5.023000	1.839000
Asia/SouthPac	2018	5.337000	1.436000
Central and Eastern Europe	2018	4.425000	2.716000
Middle East and Northern Africa	2018	4.902000	2.218000
Sub-Saharan Africa	2018	4.040000	1.153000
Western Europe	2018	5.211000	3.410000
Americas/Carib	2019	5.086000	2.015000
Asia/SouthPac	2019	5.456000	1.411000
Central and Eastern Europe	2019	4.499000	2.857000
Middle East and Northern Africa	2019	4.810000	1.922000
Sub-Saharan Africa	2019	4.093000	1.449000
Western Europe	2019	5.313000	3.437000



There is little change in the min and max happiness scores over time. We can see that the scores have very similar trends in all regions. First, there is a decrease in the max and min happiness from 2015 to 2016, then an increase from 2016 to 2017, forming a “V” shape in each graph, and later on, a slight decrease or slight increase depending on the region. We can also notice that while the values for the max and min happiness are changing over time, by 2019, the values became very close to the original values in 2015. However, in Africa we see a positive change, contrary to the dip in happiness in the ME/NA region. The trend of the lines appears to be similar across the entire world, as if the variation in happiness levels is something that affects the whole world and not just isolated regions.

Task 3

Top 10 Countries With Highest Happiness Score

Country	Avg_happiness
Singapore	5.161286
Norway	5.134778
New Zealand	5.099459
Switzerland	5.095652
Denmark	5.083876
Australia	5.054115

Country	Avg_happiness
Luxembourg	5.051504
Ireland	5.044687
Sweden	5.023141
Canada	4.982497

Top 10 Countries With Highest Positive Happiness Score Change Between 2015 and 2019

Country	Happiness_change
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

```
# The code I used in the project

##TASK 1:

## Summary table showing the average, median and standard deviation for each characteristic in each
 region in 2019

hp_data[,-1]%>%group_by(Region,Year=2019)%>%summarise_all(list(Avg=~mean(.,na.rm=TRUE),Med=~median
(.,na.rm = TRUE),Sd=~sd(.,na.rm = TRUE)))%>%
  pivot_longer(!c("Region","Year"),names_to = "characteristic")%>%pivot_wider(id_cols = c("character
istic","value","Year"),names_from=Region)%>%arrange(desc(characteristic))%>%kbl()%>%kable_styling()


##table with averages to be used for plotting the data

hp_reg_mean=hp_data[,-1]%>%group_by(Region,Year=2019)%>%summarise_all(~mean(.,na.rm=TRUE))



GDP_bar=ggplot(hp_reg_mean,aes(x=reorder(Region,-GDPperCap),y=GDPperCap,fill=Region))+geom_bar(stat=
"identity",width = 0.5)+  scale_x_discrete(guide = guide_axis(n.dodge=3)) + xlab("Region")

Trust_bar=ggplot(hp_reg_mean,aes(x=reorder(Region,-TrustGov),y=TrustGov,fill=Region))+geom_bar(stat=
"identity",width = 0.5)+  scale_x_discrete(guide = guide_axis(n.dodge=3))+ xlab("Region")

Family_bar=ggplot(hp_reg_mean,aes(x=reorder(Region,-Family),y=Family,fill=Region))+geom_bar(stat="id
entity",width = 0.5)+  scale_x_discrete(guide = guide_axis(n.dodge=3))+ xlab("Region")

Freedom_bar=ggplot(hp_reg_mean,aes(x=reorder(Region,-Freedom),y=Freedom,fill=Region))+geom_bar(stat=
"identity",width = 0.5)+  scale_x_discrete(guide = guide_axis(n.dodge=3))+ xlab("Region")

Health_bar=ggplot(hp_reg_mean,aes(x=reorder(Region,-HealthLifeExp),y=HealthLifeExp,fill=Region))+geo
m_bar(stat="identity",width = 0.5)+  scale_x_discrete(guide = guide_axis(n.dodge=3))+ xlab("Region")

Gen_bar=ggplot(hp_reg_mean,aes(x=reorder(Region,-Generosity),y=Generosity,fill=Region))+geom_bar(sta
t="identity",width = 0.5)+  scale_x_discrete(guide = guide_axis(n.dodge=3))+ xlab("Region")

GDP_box=ggplot(hp_data,aes(x=reorder(Region,-GDPperCap),y=GDPperCap,fill=Region))+geom_boxplot()+sta
t_boxplot(geom="errorbar",width=0.25)+  scale_x_discrete(guide = guide_axis(n.dodge=3))+ xlab("Regio
n")

Trust_box=ggplot(hp_data,aes(x=reorder(Region,-TrustGov),y=TrustGov,fill=Region))+geom_boxplot()+sta
t_boxplot(geom="errorbar",width=0.25)+  scale_x_discrete(guide = guide_axis(n.dodge=3))+ xlab("Regio
n")

Freedom_box=ggplot(hp_data,aes(x=reorder(Region,-Freedom),y=Freedom,fill=Region))+geom_boxplot()+sta
t_boxplot(geom="errorbar",width=0.25)+  scale_x_discrete(guide = guide_axis(n.dodge=3))+ xlab("Regio
n")

Family_box=ggplot(hp_data,aes(x=reorder(Region,-Family),y=Family,fill=Region))+geom_boxplot()+stat_b
oxplot(geom="errorbar",width=0.25)+  scale_x_discrete(guide = guide_axis(n.dodge=3))+ xlab("Region")

Health_box=ggplot(hp_data,aes(x=reorder(Region,-HealthLifeExp),y=HealthLifeExp,fill=Region))+geom_bo
xplot()+stat_boxplot(geom="errorbar",width=0.25)+  scale_x_discrete(guide = guide_axis(n.dodge=3))+
 xlab("Region")
```

```
Gen_box=ggplot(hp_data,aes(x=reorder(Region,-Generosity),y=Generosity,fill=Region))+geom_boxplot()+s
tat_boxplot(geom="errorbar",width=0.25)+  scale_x_discrete(guide = guide_axis(n.dodge=3))+ xlab("Reg
ion")

grid.arrange(GDP_bar,GDP_box,top="GDP vs Region")
cat("\n\n")
grid.arrange(Trust_bar,Trust_box,top="TrustGov vs Region")
cat("\n\n")
grid.arrange(Freedom_bar,Freedom_box,top="Freedom vs Region")
cat("\n\n")
grid.arrange(Family_bar,Family_box,top="Family vs Region")
cat("\n\n")
grid.arrange(Health_bar,Health_box,top= "HealthLifeExp vs Region")
cat("\n\n")
grid.arrange(Gen_bar,Gen_box, top="Generosity vs Region")


##-------------------------------------------------------------------------------------------------
--------------

## TASK 2:

## a
nrows=nrow(hp_data)
Happiness=vector(mode="numeric",length=nrows)
for(i in 1:nrows){
Happiness[i]=sum(hp_data[i,c("GDPperCap","TrustGov","Family","HealthLifeExp","Generosity","Freedom"
)])
}
hp_happy=hp_data%>%mutate(Happiness=Happiness)
hp_happy[is.na(hp_happy)]=0
hp_happy%>%subset(,c("Country","Year","Happiness"))

## b

happy_summ=hp_happy%>%group_by(Region,Year)%>%summarise(max_happiness=max(Happiness,na.rm = TRUE),mi
n_happiness=min(Happiness,na.rm = TRUE))%>%arrange(Year)
happy_summ%>%kbl()%>%kable_styling()

## need to store max and min happiness under same variable in order to plot them in the same graph

happy_summ=happy_summ%>%pivot_longer(c("max_happiness","min_happiness"),names_to="type")

ggplot(happy_summ,aes(x=Year,y=value,color=type))+geom_line()+geom_point(width=0.5)+facet_wrap(~Regi
on)+ylab("Happiness score")+ scale_color_manual(values=c("red","blue"))

##-------------------------------------------------------------------------------------------------
----------

##TASK 3:

##a
hp_happy%>%group_by(Country)%>%summarize(Avg_happiness=mean(Happiness))%>%arrange(desc(Avg_happines
s))%>%slice(1:10)%>%kbl(caption = "Top 10 Countries With Highest Happiness Score ")%>%kable_styling
(full_width = FALSE)
```

```
##b

##function that creates happiness change for any table with columns Country, Year and Happiness
happiness_change=function(table,nyears){
tbl=tibble(Country=character(),Happiness_change=double())
table=table%>%select(c("Country","Year","Happiness"))
countries=unique(table$Country)
c_len=length(countries)
for(i in 1:c_len){
  country=countries[i]
  tmp=table%>%subset((Country==country))
  hp_change=(tmp[nyears,3]-tmp[1,3])[[1]]
  tbl=tbl%>%add_row(Country=country,Happiness_change=hp_change)

}
tbl
}
happiness_change(table=hp_happy,nyears = 5)%>%arrange(desc(Happiness_change))%>%slice(1:10)%>%kbl(ca
ption = "Top 10 Countries With Highest Positive Happiness Score Change Between 2015 and 2019 ")%>%ka
ble_styling(full_width = FALSE)
```