World Happiness Report Analysis

Jessie Geng

Family, Generosity, and Freedom by Region in 2019

To assess each of the scores by region in 2019, we first look at the distribution of the scores across region. From Figure 1, we see that they vary greatly depending on the region. A common pattern we see is Western Europe consistently ranking high and Sub-Saharan Africa ranking low. Some interesting variables to look at are government trust and generosity. For government trust, we see that every region has a very low median score except for Western Europe, but they also have the largest range of scores by far. This is in contrast to all the other variables where Western Europe often has the smallest range of values.

Another interesting variable is generosity, where each region's distribution looks fairly similar to the others. Asia and the South Pacific and Western Europe do have relatively higher medians, and the range of responses in Asia is much greater, but this is the one of the only variables where there isn't one region who is either much obviously higher or lower than the others.

```
# plots of 6 variables by region
gdp = happiness2019 %>% ggplot(aes(x=Region, y=GDPperCap, color=Region)) + geom_boxplot() +
  labs(title="GDP Per Capita") +
  theme(axis.text.x=element_blank(), plot.title=element_text(hjust=0.5))

gov = happiness2019 %>% ggplot(aes(x=Region, y=TrustGov, color=Region)) + geom_boxplot() +
  labs(title="Government Trust") +
  theme(axis.text.x=element_blank(), plot.title=element_text(hjust=0.5))

fam = happiness2019 %>% ggplot(aes(x=Region, y=Family, color=Region)) + geom_boxplot() +
  labs(title="Family") +
  theme(axis.text.x=element_blank(), plot.title=element_text(hjust=0.5))

free = happiness2019 %>% ggplot(aes(x=Region, y=Freedom, color=Region)) + geom_boxplot() +
  labs(title="Freedom") +
  theme(axis.text.x=element_blank(), plot.title=element_text(hjust=0.5))

health = happiness2019 %>% ggplot(aes(x=Region, y=HealthLifeExp, color=Region)) + geom_boxplot() +
  labs(title="Health") +
  theme(axis.text.x=element_blank(), plot.title=element_text(hjust=0.5))

gen = happiness2019 %>% ggplot(aes(x=Region, y=Generosity, color=Region)) + geom_boxplot() +
  labs(title="Generosity") +
  theme(axis.text.x=element_blank(), plot.title=element_text(hjust=0.5))

# plot all 6 in a grid
ggarrange(gdp, gov, fam, free, health, gen, ncol=3, nrow=2, common.legend=TRUE, legend="bottom")
```

Table 1 contains the mean scores by region, where we can confirm that Western Europe has some of the highest mean scores across all the 6 variables and that Sub-Saharan Africa has some of the lowest scores across the 6 variables.

When we look at the distribution of scores across region, we see that while the absolute values of each score by region differs, the position of each region relative to the other regions remains fairly consistent. In other words, we
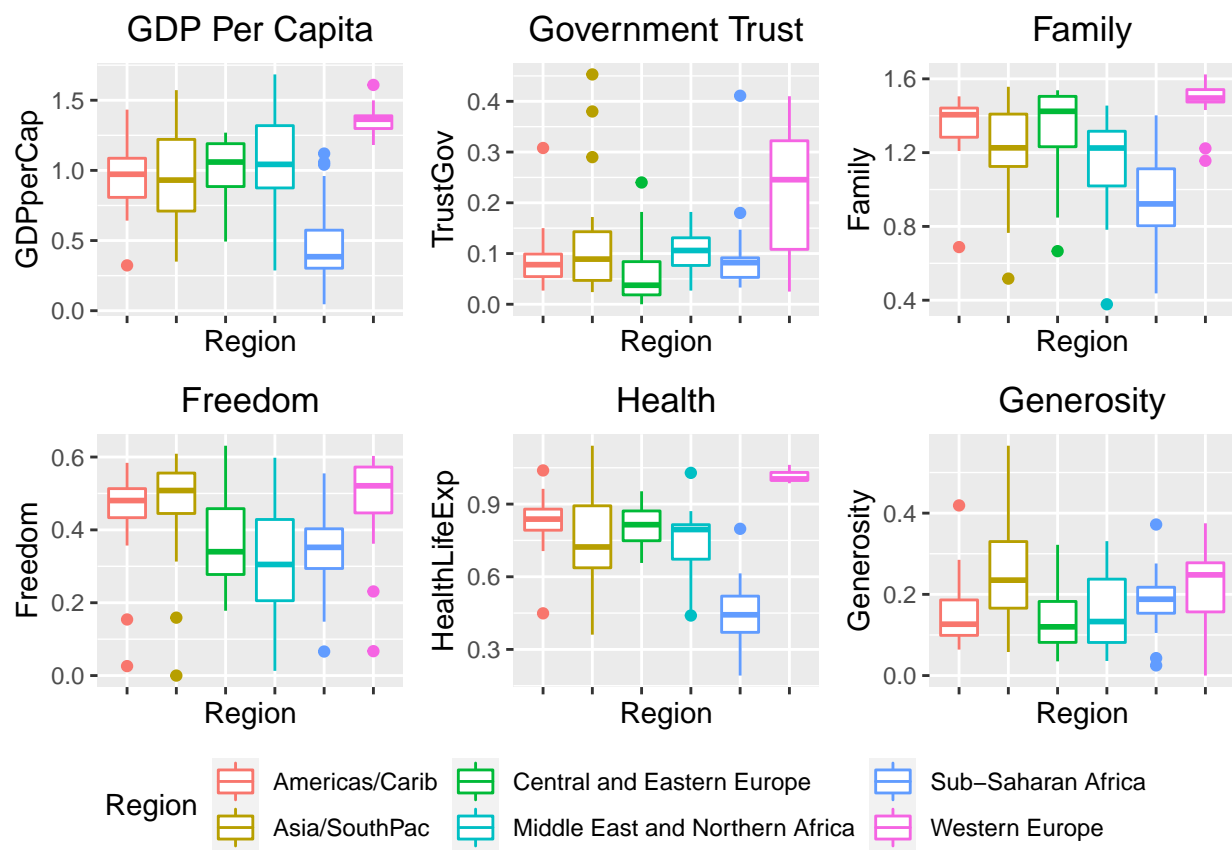
Figure 1: Distribution of Scores by Region

have that the 6 scores seem to show a similar pattern, with Sub-Saharan Africa ranking the lowest and Western Europe ranking the highest, with the other regions scattered somewhere in between.

However, when we look at each specific region across the six scores, we see a different picture. Most of the regions varied differently on the different scores, which we can see in Table 2.

```
# mean scores by region
happiness2019 %>% group_by(Region) %>% summarise_at(vars(GDPperCap:Generosity), mean) %>%
  kable(format="pipe", row.names=FALSE, digits=4,
        caption="Mean Scores by Region", align=rep("c", 8))
```

Table 1: Mean Scores by Region

Region	GDPperCap	TrustGov	Family	Freedom	HealthLifeExp	Generosity
Americas/Carib	0.9535	0.0879	1.3463	0.4458	0.8295	0.1553
Asia/SouthPac	0.9256	0.1240	1.2140	0.4588	0.7725	0.2520
Central and Eastern Europe	1.0224	0.0628	1.3404	0.3580	0.8085	0.1412
Middle East and Northern Africa	1.0591	0.1050	1.1487	0.3179	0.7511	0.1535
Sub-Saharan Africa	0.4662	0.0876	0.9381	0.3422	0.4408	0.1883
Western Europe	1.3620	0.2212	1.4839	0.4826	1.0138	0.2210

```
# variance of scores by region
happiness2019 %>% group_by(Region) %>% summarise_at(vars(GDPperCap:Generosity), var) %>%
  kable(format="pipe", row.names=FALSE, digits=4,
        caption="Variance of the Scores by Region", align=rep("c", 8))
```

Table 2: Variance of the Scores by Region

Region	GDPperCap	TrustGov	Family	Freedom	HealthLifeExp	Generosity
Americas/Carib	0.0583	0.0035	0.0296	0.0164	0.0125	0.0077
Asia/SouthPac	0.1096	0.0137	0.0703	0.0222	0.0436	0.0163
Central and Eastern Europe	0.0467	0.0042	0.0471	0.0142	0.0053	0.0065
Middle East and Northern Africa	0.1207	0.0016	0.0690	0.0278	0.0206	0.0076
Sub-Saharan Africa	0.0795	0.0047	0.0568	0.0103	0.0166	0.0045
Western Europe	0.0098	0.0171	0.0126	0.0182	0.0006	0.0102

Predicted Happiness Score

We now move to looking at each country's happiness score, which we compute for each country from 2015 to 2019 by summing across the 6 scores for that country in each of those years.

```
# country happiness scores
happiness$Happiness <- apply(happiness[, 3:8], 1, sum, na.rm=TRUE)
```

Now we compute the minimum and maximum happiness scores for each region in each year, which we will then plot.

```
# min and max happiness scores by region per year
happiness_min_max <- happiness %>% group_by(Region, Year) %>%
  summarize(Min=min(Happiness), Max=max(Happiness))
```

We see the ranges of happiness scores for each region in Figure 2, where we notice while that they differ across regions, they don't appear to change much within each region across the years. Each region's minimum and

maximum happiness scores remains fairly constant from 2015 to 2019. More notably, they don't appear to be increasing over time, which is what one may initially think to be the case.

```
# plot of region min/max happiness scores over time
happiness_min_max %>% ggplot(aes(x=Year, group=Region, col=Region)) +
  geom_linerange(aes(ymin=Min, ymax=Max)) +
  facet_wrap(~Region) + labs(x="Year", y="Happiness Score") +
  theme(legend.title=element_text(size=14), legend.text=element_text(size=12))
```
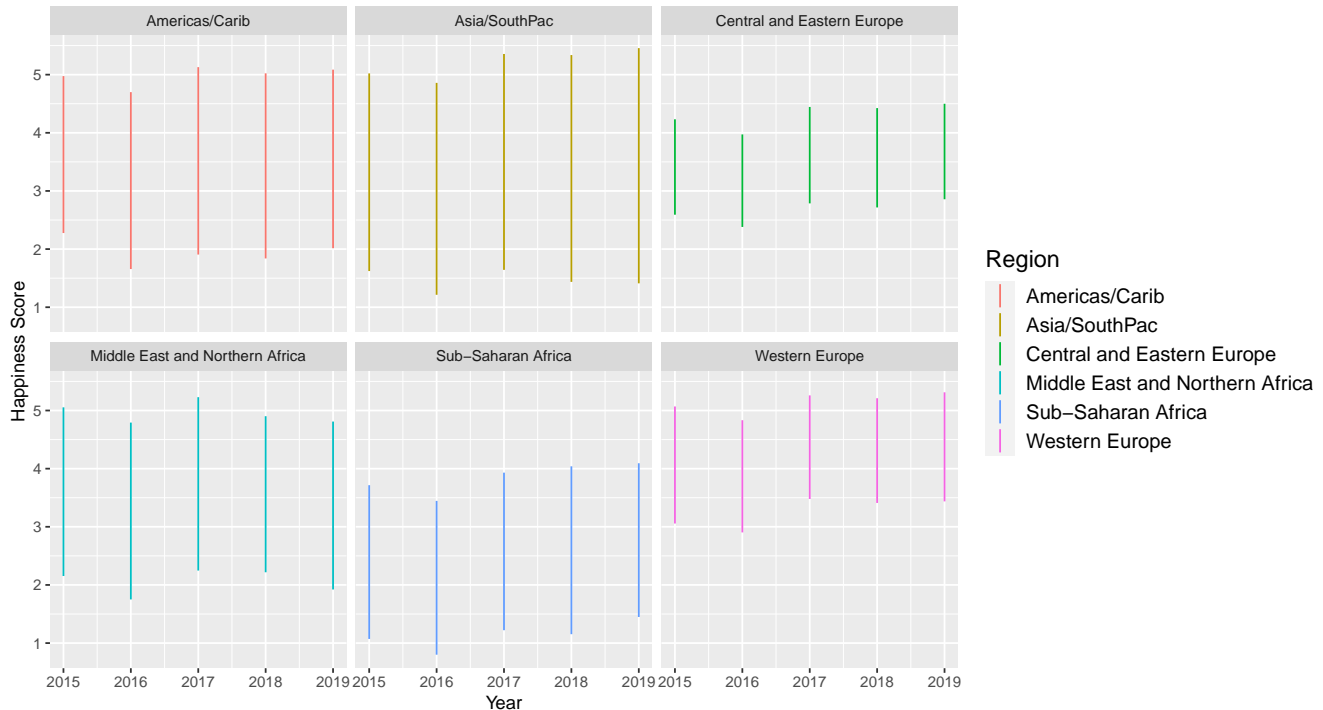

Figure 2: Minimum and Maximum Happiness Scores by Region

Out of curiosity, we now look at some other associations, namely those between each of the 6 variables and the earlier computed happiness score. These plots can be found in Figure 3, where we plot each of the 6 variables against the happiness score. Unsurprisingly, we see that GDP per capita and health + life expectancy (which I would imagine are also highly correlated with one another) have an almost perfectly positive relationship with the happiness score. Family and freedom also appear to be positively associated with happiness, which is a somewhat more surprising result. On the other hand, government trust and generosity do not appear to be linearly related to happiness.

```
# plot of each score against happiness
gdp_hap = happiness %>% ggplot(aes(x=GDPperCap, y=Happiness, color=Region)) + geom_point()

gov_hap = happiness %>% ggplot(aes(x=TrustGov, y=Happiness, color=Region)) + geom_point()

fam_hap = happiness %>% ggplot(aes(x=Family, y=Happiness, color=Region)) + geom_point()

free_hap = happiness %>% ggplot(aes(x=Freedom, y=Happiness, color=Region)) + geom_point()

health_hap = happiness %>% ggplot(aes(x=HealthLifeExp, y=Happiness, color=Region)) + geom_point()

gen_hap = happiness %>% ggplot(aes(x=Generosity, y=Happiness, color=Region)) + geom_point()
```

```
# plot all 6 in a grid
ggarrange(gdp_hap, gov_hap, fam_hap, free_hap, health_hap, gen_hap, ncol=3, nrow=2,
          common.legend=TRUE, legend="bottom")
```
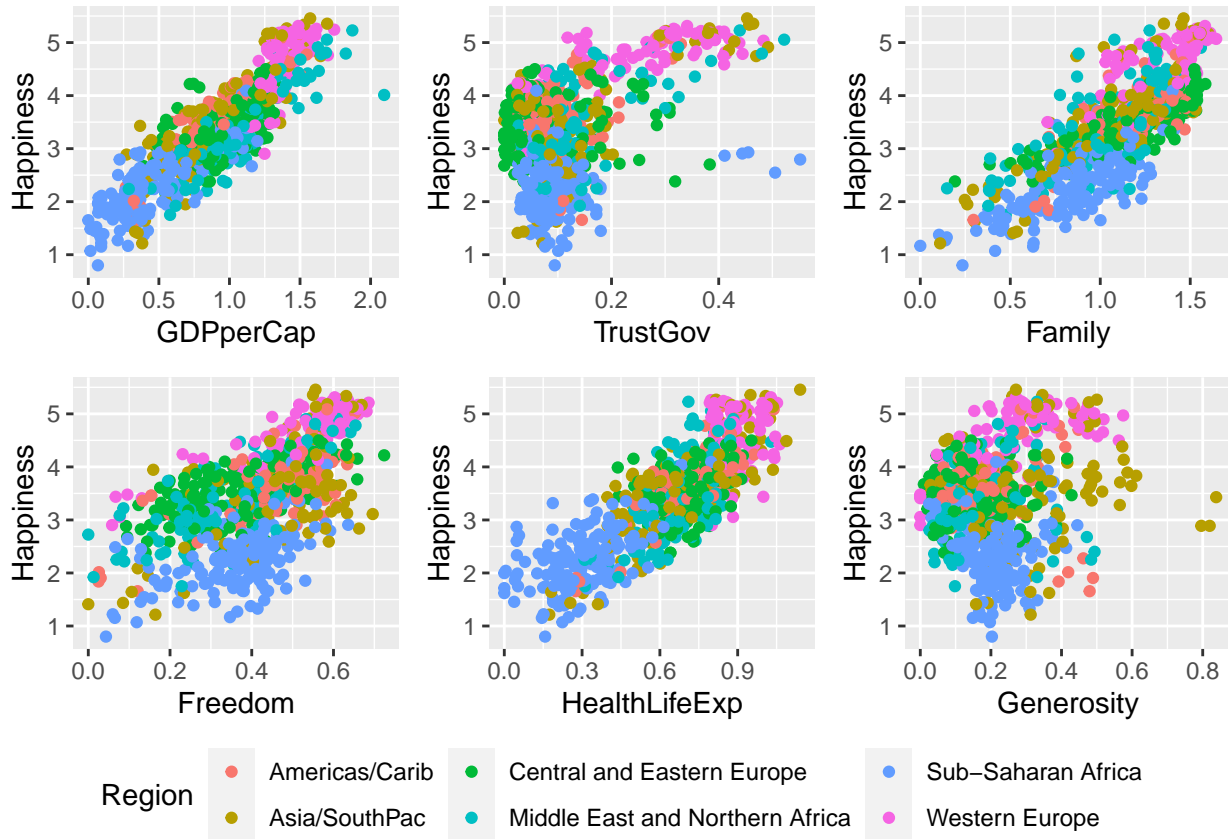

Figure 3: Relationship Between Each Score and Happiness

Happiest Countries Over Time

Our analyses now take us to comparing each country's happiness scores. We first examine the top 10 countries who had the highest average happiness scores over the 5 years. These results can be found in Table 4, where we see that Singapore tops the list with an average happiness score of 5.16. Looking at the regions for the top 10, we see that they're mainly located in Asia or the South Pacific and Western Europe, with the only entry from the Americas or the Caribbean being Canada.

```
# highest average happiness scores
happiness %>% group_by(Country, Region) %>% summarize(`Average Happiness`=mean(Happiness)) %>%
  arrange(desc(`Average Happiness`)) %>% head(n=10) %>%
  kable(format="pipe", row.names=FALSE, digits=4,
        caption="Top 10 Highest Average Happiness Scores", align=rep("c", 3))
```

Table 3: Top 10 Highest Average Happiness Scores

Country	Region	Average Happiness
Singapore	Asia/SouthPac	5.1613

Country	Region	Average Happiness
Norway	Western Europe	5.1348
New Zealand	Asia/SouthPac	5.0995
Switzerland	Western Europe	5.0957
Denmark	Western Europe	5.0839
Australia	Asia/SouthPac	5.0541
Luxembourg	Western Europe	5.0515
Ireland	Western Europe	5.0447
Sweden	Western Europe	5.0231
Canada	Americas/Carib	4.9825

Table 5 contains the top 10 countries who experienced the largest positive change in their happiness scores from 2015 to 2019. Many of these countries have an almost 1 point difference in their happiness scores between 2015 and 2019 and we see that all these countries are located in either Asia and the South Pacific or Central and Eastern Europe.

```
# largest positive change in happiness scores
happiness %>% filter(Year==2015 | Year==2019) %>% group_by(Country, Region) %>%
  summarize(Change=diff(Happiness)) %>%
  arrange(desc(Change)) %>% head(n=10) %>%
  kable(format="pipe", row.names=FALSE, digits=4,
        caption="Top 10 Largest Positive Change in Happiness Scores", align=rep("c", 2))
```

Table 4: Top 10 Largest Positive Change in Happiness Scores

Country	Region	Change
Bangladesh	Asia/SouthPac	0.8727
Myanmar	Asia/SouthPac	0.8200
Kosovo	Central and Eastern Europe	0.8062
Croatia	Central and Eastern Europe	0.7676
Pakistan	Asia/SouthPac	0.6424
Montenegro	Central and Eastern Europe	0.6099
Nepal	Asia/SouthPac	0.6039
India	Asia/SouthPac	0.6014
Serbia	Central and Eastern Europe	0.6009
Bosnia and Herzegovina	Central and Eastern Europe	0.5980