Final Project

Anna MacFarlane, Edgar Salas, Jerry Fu, Taalin RaoShah

10/27/2020

Introduction

CodeStone chose to explore the data from the World Happiness Reports of 2018 and 2019, published by the Sustainable Development Network. According to the WHR, this data is derived from the Gallup World Poll, a systematic telephone survey and in person interviews in over 160 countries whose surveys claim to represent 80%+ of the population (2014). The calls are made via random phone number generation and randomly selecting households (GWP, 2014).

Generally, we want to explore the following question: which factors correlate most strongly to happiness across the globe? We chose to focus on four variables as well as the overall happiness score in our analysis. The four variables chosen were GDP per capita, Perceptions of Corruption, Social Support, and Life Expectancy, selected because we thought they represented economic, political, social, and medical aspects of well being respectively. Furthermore, the data set included 312 cases, each representing the data for a particular country. Guided by our general research question, we also examined the variation of their distribution by region, and which of these factors have the strongest correlations.

Considering the concerning state of our world in 2020, including the worsening effects of climate change, threats to democracy, and much more, we found it topical and insightful to evaluate what contributes to happiness within each nation and across the globe. The first World Happiness Report, published in 2012, presents the report as a means of grappling with the countless contradictions that exist in modern society such as the balance between pursuing economic success versus protecting the environment or the tradeoffs between personal profit and community trust (Helliwell et al., 2012). Eight years later, these paradoxes persist, and the potential solutions are closely linked to definitions of morality, heightening their controversy. Considering the continued debate over such questions, we believe there are grounds for further investigation into trends of happiness over time and the factors that contribute to it.

```r
library(tidyverse)
library(plyr) #package for join command
library(maps) #package for world map

report_2018 <- read_csv("data/2018.csv") %>%
  mutate(year = "2018",
         `Perceptions of corruption` = as.numeric(`Perceptions of corruption`))
report_2019 <- read_csv("data/2019.csv") %>%
  mutate(year = "2019")
country_region <- read_csv("data/2020.csv") %>%
  mutate(year = "2020") %>%
  select(`Country name`, `Regional indicator`)

names(country_region) <- c("Country", "Region")
names(report_2018) <- str_replace_all(names(report_2018), c(" " = "_"))
names(report_2019) <- str_replace_all(names(report_2018), c(" " = "_"))

worldhappiness <- full_join(report_2018, report_2019)
```

```r
worldhappiness <- worldhappiness %>%
  rename(c("Country_or_region" = "Country")) %>%
  left_join(country_region)

glimpse(worldhappiness)
```

```
## Rows: 312
## Columns: 11
## $ Overall_rank            <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,...
## $ Country                 <chr> "Finland", "Norway", "Denmark", "Icela...
## $ Score                   <dbl> 7.632, 7.594, 7.555, 7.495, 7.487, 7.4...
## $ GDP_per_capita          <dbl> 1.305, 1.456, 1.351, 1.343, 1.420, 1.3...
## $ Social_support          <dbl> 1.592, 1.582, 1.590, 1.644, 1.549, 1.4...
## $ Healthy_life_expectancy <dbl> 0.874, 0.861, 0.868, 0.914, 0.927, 0.8...
## $ Freedom_to_make_life_choices <dbl> 0.681, 0.686, 0.683, 0.677, 0.660, 0.6...
## $ Generosity              <dbl> 0.202, 0.286, 0.284, 0.353, 0.256, 0.3...
## $ Perceptions_of_corruption <dbl> 0.393, 0.340, 0.408, 0.138, 0.357, 0.2...
## $ year                    <chr> "2018", "2018", "2018", "2018", "2018"...
## $ Region                  <chr> "Western Europe", "Western Europe", "W...
```

```r
worldhappiness %>%
  filter(is.na(Region))
```

```
## # A tibble: 20 x 11
##    Overall_rank Country Score GDP_per_capita Social_support Healthy_life_ex~
##           <dbl> <chr>   <dbl>          <dbl>          <dbl>            <dbl>
## 1            26 Taiwan   6.44           1.36           1.44            0.857
## 2            32 Qatar    6.37           1.65           1.30            0.748
## 3            38 Trinid~  6.19           1.22           1.49            0.564
## 4            49 Belize   5.96           0.807          1.10            0.474
## 5            58 Northe~  5.84           1.23           1.21            0.909
## 6            76 Hong K~  5.43           1.40           1.29            1.03
## 7            97 Bhutan   5.08           0.796          1.34            0.527
## 8            98 Somalia  4.98           0              0.712           0.115
## 9           137 Sudan    4.14           0.605          1.24            0.312
## 10          142 Angola   3.80           0.73           1.12            0.269
## 11          150 Syria    3.46           0.689          0.382           0.539
## 12           25 Taiwan   6.45           1.37           1.43            0.914
## 13           29 Qatar    6.37           1.68           1.31            0.871
## 14           39 Trinid~  6.19           1.23           1.48            0.713
## 15           64 Northe~  5.72           1.26           1.25            1.04
## 16           76 Hong K~  5.43           1.44           1.28            1.12
## 17           84 North ~  5.27           0.983          1.29            0.838
## 18           95 Bhutan   5.08           0.813          1.32            0.604
## 19          112 Somalia  4.67           0              0.698           0.268
## 20          149 Syria    3.46           0.619          0.378           0.44
## # ... with 5 more variables: Freedom_to_make_life_choices <dbl>,
## #   Generosity <dbl>, Perceptions_of_corruption <dbl>, year <chr>, Region <chr>
```

```r
worldhappiness <- worldhappiness %>%
  mutate(Region = ifelse(Country == "Taiwan", "East Asia", Region)) %>%
  mutate(Region = ifelse(Country == "Qatar", "Middle East and North Africa", Region)) %>%
  mutate(Region = ifelse(Country == "Trinidad & Tobago", "Latin America and Caribbean", Region)) %>%
  mutate(Region = ifelse(Country == "Belize", "Latin America and Caribbean", Region)) %>%
  mutate(Region = ifelse(Country == "Northern Cyprus", "Middle East and North Africa", Region)) %>%
```

```r
  mutate(Region = ifelse(Country == "Hong Kong", "East Asia", Region)) %>%
  mutate(Region = ifelse(Country == "Bhutan", "South Asia", Region)) %>%
  mutate(Region = ifelse(Country == "Somalia", "Middle East and North Africa", Region)) %>%
  mutate(Region = ifelse(Country == "Sudan", "Middle East and North Africa", Region)) %>%
  mutate(Region = ifelse(Country == "Angola", "Middle East and North Africa", Region)) %>%
  mutate(Region = ifelse(Country == "Syria", "Middle East and North Africa", Region)) %>%
  mutate(Region = ifelse(Country == "North Macedonia", "Central and Eastern Europe", Region))
```

Is there a relationship between region of the world and happiness score?

```r
world_map <- map_data("world") %>%
  mutate(region = ifelse(region == "USA", "United States", region)) %>%
  mutate(region = ifelse(region == "Democratic Republic of the Congo", "Congo (Kinshasa)", region))

happiness_score_map <- left_join(worldhappiness, world_map, by = c("Country" = "region"))
ggplot(happiness_score_map, aes(x = long, y = lat, group = group)) +
  geom_polygon(aes(fill = Score), color = "white") +
  labs(title = "World Map", subtitle = "colored by Happiness Score", x = "", y = "") +
  scale_fill_viridis_c(option = "D")
```


https://www.datanovia.com/en/blog/how-to-create-a-map-using-ggplot2/
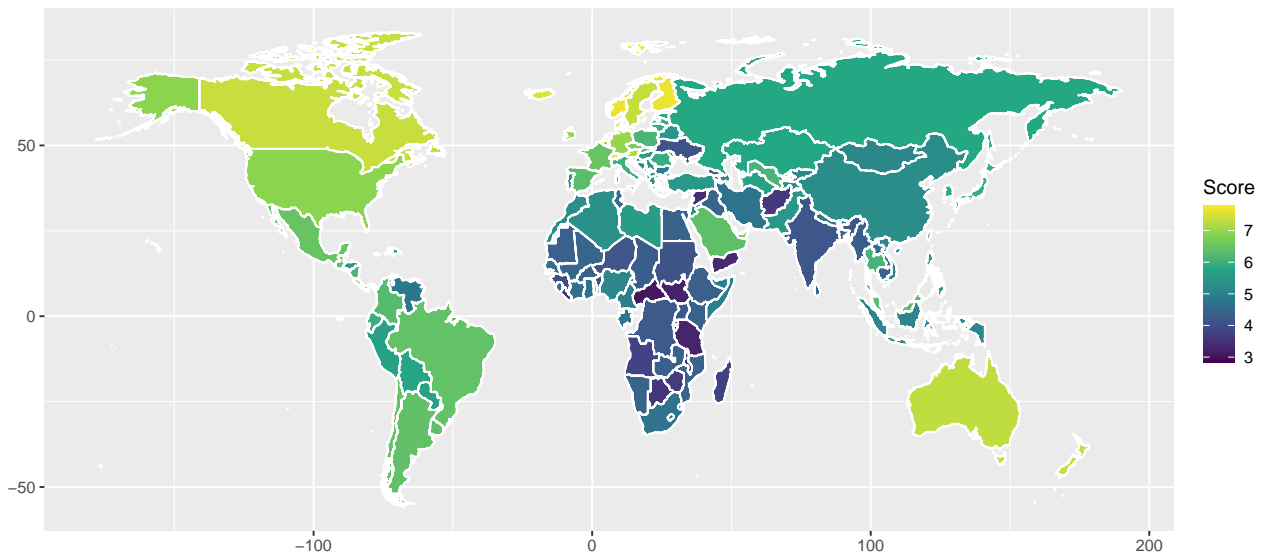
```r
chisq.test(worldhappiness$Score, worldhappiness$Region)
```

```
## Warning in chisq.test(worldhappiness$Score, worldhappiness$Region): Chi-squared
## approximation may be incorrect

##
##  Pearson's Chi-squared test
##
## data:  worldhappiness$Score and worldhappiness$Region
## X-squared = 2729.8, df = 2646, p-value = 0.1253
```

Multiple regression for all considered factors

```r
m_happy <- lm(Score ~ Perceptions_of_corruption + Social_support +
                GDP_per_capita + Healthy_life_expectancy, data = worldhappiness)

m_happy %>%
  tidy() %>%
  select(term, estimate)
```

```
## # A tibble: 5 x 2
##   term                      estimate
##   <chr>                        <dbl>
## 1 (Intercept)                   2.10
## 2 Perceptions_of_corruption     1.95
## 3 Social_support                1.41
## 4 GDP_per_capita                0.827
## 5 Healthy_life_expectancy       0.941
```

```r
glance(m_happy)$r.squared
```

```
## [1] 0.7526699
```

Perceptions of Corruption

```r
# Linear model for relationship between perceptions of corruption score and # overall happiness score
m_corrupt <- lm(Score ~ Perceptions_of_corruption, data = worldhappiness)

m_corrupt %>%
  tidy() %>%
  select(term, estimate)
```

```
## # A tibble: 2 x 2
##   term                      estimate
##   <chr>                        <dbl>
## 1 (Intercept)                   4.87
## 2 Perceptions_of_corruption     4.62
```

```r
# Visualization of linear model
ggplot(data = worldhappiness, aes(x = Perceptions_of_corruption,
                          y = Score)) +
  geom_point(size = 0.8) +
  geom_smooth(method = lm, color = "blue", size = 0.5) +
  labs(title = "Linear model for relationship between perceptions of corruption
and overall happiness score",
      subtitle = "Moderate, positive assocation",
      x = "Perceptions of corruption score",
      y = "Overall happiness score")
```

```
## Warning: Removed 1 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```

Linear model for relationship between perceptions of corruption and overall happiness score

Moderate, positive assocation
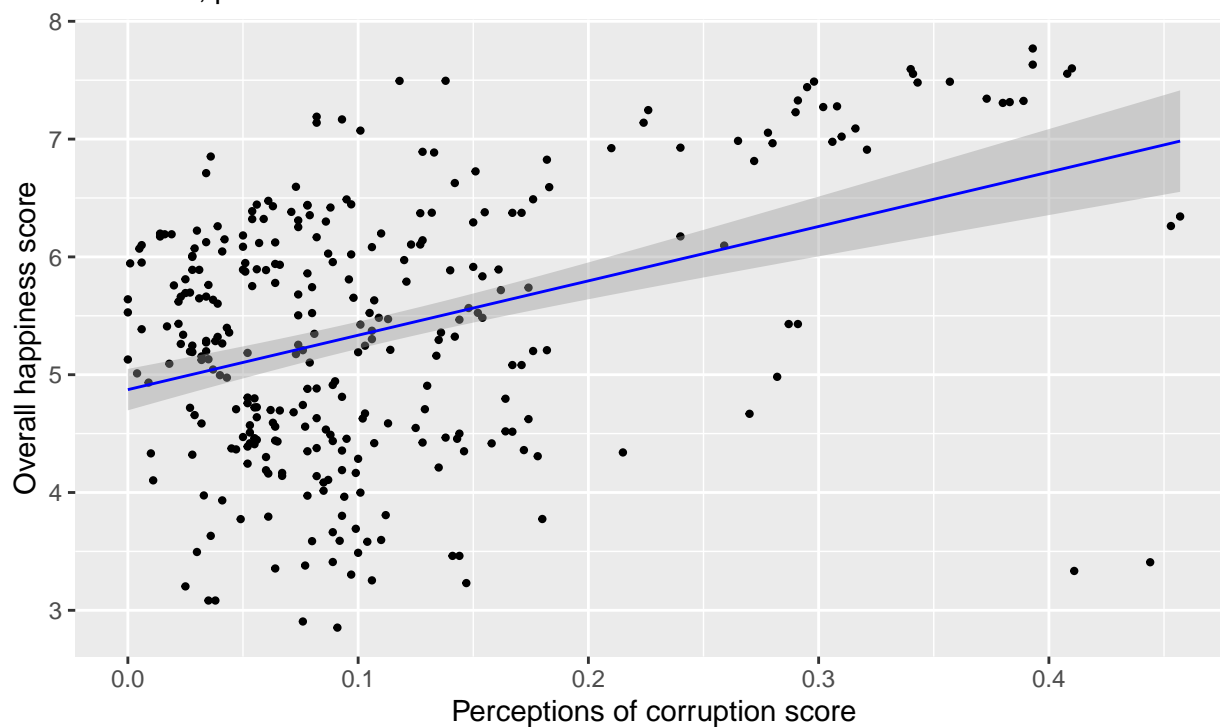

```
#Confidence interval for linear model above
broom::confint_tidy(m_corrupt, conf.level = 0.95) %>%
  slice(2)
```

```
## Warning: confint_tidy is now deprecated and will be removed from a future
## release of broom. Please use the applicable confint method.
```

```
## # A tibble: 1 x 2
##   conf.low conf.high
##      <dbl>     <dbl>
## 1     3.42      5.82
```