Project Proposal

Due Friday, October 9, 11:59 PM

Taalin, Anna, Edgar, Jerry

Test

Project Proposal

CodeStone chose to explore the data from the World Happiness Reports of 2015 to 2020, published by the Sustainable Development Network. Considering the concerning state of our world in 2020, including the worsening effects of climate change, a global pandemic, 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 persit, 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.

Generally, we want to explore the following question: which factors correlate most strongly to happiness across the globe? Using this question, we also plan on examining whether these vary by region, and what factors have the strongest correlations. With these results we hope to be able to offer guidance to countries on which aspects of their citizens' life they might want to focus on improving. Our hypothesis is that happiness will be strongly correlated to measures of freedom, life expectancy, and economy (GDP per capita). We also expect happiness to vary by region.

As aforementioned, the data set we'll be using comes from the World Happiness Report (WHR), that being the most available and reliable scaled measurement of happiness. Specifically, we'll be using the 2018-2019 World Happiness Reports, as we'd like to track time-consistent correlations to happiness. Unfortunately, not only do the World Happiness Reports vary in columns and column names, they also vary in the scale of their variables. For example, the range for Freedom in 2020 is about 0.397 - 0.975, while in 2019, it is 0 - 0.631. Therefore, we chose two most recent years, which were exactly the same in both the variables they looked at, and how they evaluated those variables. String data includes country and region. The rest of the values are doubles indicating Happiness Score, GDP per capita, Life Expectancy, Perceptions of Corruption, Generosity, Freedom to Make Life Choices, and Social Support/Family. 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).

https://www.gallup.com/178667/gallup-world-poll-work.aspx

https://worldhappiness.report/faq/

https://www.kaggle.com/mathurinache/world-happiness-report

Helliwell, John F., Richard Layard, and Jeffrey Sachs, eds. 2012. World Happiness Report 2012. New York: UN Sustainable Development Solutions Network.

Gallup World Poll. (2014, October 14). How does the Gallup world poll work? Gallup.com. https://www.gallup.com/178667/gallup-world-poll-work.aspx

```
library(tidyverse)
library(plyr) #package for join command
```

```
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)
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...
```

```
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>
```

```
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)) %>%
  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))

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

```
## # A tibble: 0 x 11
## # ... with 11 variables: Overall_rank <dbl>, Country <chr>, Score <dbl>,
## #   GDP_per_capita <dbl>, Social_support <dbl>, Healthy_life_expectancy <dbl>,
## #   Freedom_to_make_life_choices <dbl>, Generosity <dbl>,
## #   Perceptions_of_corruption <dbl>, year <chr>, Region <chr>
```

https://dnidzgorski.wordpress.com/2017/06/09/r-fix-column-names-spaces/