Project 1: Happiness and Health

Hannah LeBlanc

3/17/2021

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com (http://rmarkdown.rstudio.com).

Part 1: Tidy

I titled my project Happiness and Health and chose these datasets because I hope to one day become a health care professional myself and value the role access to health care plays in the happiness of a population. The datasets being used for this project are known as "health_systems" and "world_happiness" from the website/database kaggle. The health_systems dataset describes the health systems of several countries including observations pertaining to the completeness of birth and death, health expenditures and the ratio of phsyicians to every 1000 patients, for example. THe world_happiness dataset involves the answers to poll questions that involve traits associated with happiness (and unhappiness) and an overall rating of happiness on the Cantril ladder. Observations include rankings of generosity and freedom as well corruption and negative affect. I was able to bridge the two datasets because of the rankings of health observations found in the world happiness report. Overall, I expect to find a very positive relationship between health and happiness and I hope to use information like health expenditures to support my prediction.

Both of my datasets are considered tidy data because variables make up only the columns. In my tidying step, I chose to remove one column of the health_systems dataset known as "Province_State" as more often then not there was no value observed in this column.

```
#remove the column known as Province_State from the health_systems dataset
health_systems <- health_systems %>% select(-c(Province_State))
```

Part 2: Join/Merge

In order to join my datasets, I used an inner join function, which keeps only the rows that have matches between the datasets and drops rows without a match. I am most interested in the comparison of the happiness ratings of countries and the health care systems that are in place in the countries observed. For this reason, I only wanted to keep the observations for countries that were represented in both the world_happiness and health_systems dataset. With the health_systems dataset having more countries listed (210 versus 156), I knew there would be some countries that would not have observations in the world_happiness dataset. To validate my assumption, I joined the datasets expecting drop in the number of countries included in my combined dataset.

Sure enough, the joined dataset known as health_happiness includes observations for only **148** countries. I used the full_join function to determine how many total cases there were if I kept all rows from both datasets in order to determine how many cases were dropped. This full_join resulted in a dataset called health_happiness_full which has **229** observations. The difference between my two datasets is 81, meaning **81** rows were dropped.

```
#apply inner join on datasets world_happiness and health_systems, matching the observations with the key variable
#equates the two variables Country_Region and Country (region) from the two datasets as they represent the same o
bservation just use different variable names
health_happiness <- world_happiness %>%
  inner_join(health_systems, by=c("Country (region)" = "Country_Region")) %>%
  rename('HealthyLifeExpectancy' = 'Healthy life\r\nexpectancy') %>%
  rename('Country_Region' = 'Country (region)')

#observe the inner join result
health_happiness_inner_preview <- health_happiness %>%
    select(c('Country_Region','Ladder','Freedom','World_Bank_Name','Health_exp_pct_GDP_2016'))
hhi_rows <- nrow(health_happiness)
hhi_cols <- ncol(health_happiness)
kable(head(health_happiness_inner_preview,5), caption=str_interp("Health Happiness Inner Join Preview (Actual siz
e: ${hhi_rows} x ${hhi_cols})"))
```

Health Happiness Inner Join Preview (Actual size: 148 x 23)

Country_Region	Ladder	Freedom	World_Bank_Name	Health_exp_pct_GDP_2016
Finland	1	5	Finland	9.5
Denmark	2	6	Denmark	10.4
Denmark	2	6	Faroe Islands	NaN
Denmark	2	6	Greenland	NaN
Norway	3	3	Norway	10.5

```
#apply full join on datasets world_happiness and health_systems, matching the observations with the key variable
health_happiness_full <- world_happiness %>%
  full_join(health_systems,by= c("Country (region)" = "Country_Region"))

#observe the full join result
health_happiness_full_preview <- health_happiness_full %>%
    select(c('Country (region)','Ladder','Freedom','World_Bank_Name','Health_exp_pct_GDP_2016'))
hhf_rows <- nrow(health_happiness_full)
hhf_cols <- ncol(health_happiness_full)
kable(tail(health_happiness_full_preview,5), caption=str_interp("Health Happiness Full Join Preview (Actual size:
${hhf_rows} x ${hhf_cols})"))
```

Health Happiness Full Join Preview (Actual size: 229 x 23)

Country (region)	Ladder	Freedom	World_Bank_Name	Health_exp_pct_GDP_2016
US	NA	NA	United States	17.100000000000001
NA	NA	NA	Vanuatu	3.7
US	NA	NA	Virgin Islands (U.S.)	NaN
NA	NA	NA	West Bank and Gaza	NaN
NA	NA	NA	Yemen, Rep.	5.6

Part 3: Summary Statistics

Core diplyr Functions

1. Filter - filter countries with best healthy life expectancy
2. Select - select freedom and corruption ratings by country
3. Arrange - order countries by most expensive health care compared to GDP
4. GroupBy/Mutate/Summarize
    - mutate country name to create column based on first letter
    - group by first letter of country name
    - analyze average freedom based on first letter of country name

Here I mutated the dataset in order to create a column based on the first letter of the country in which the data was taken. It also puts the observations in order from the smallest ladder value to the largest.

```
# Example using diplyr.filter
healthy_life_expectancy <- health_happiness %>%
  filter(HealthyLifeExpectancy <= 5) %>%
  select('Country_Region','HealthyLifeExpectancy')
kable(head(healthy_life_expectancy, 5))
```

Country_Region	HealthyLifeExpectancy
Switzerland	4
France	5
France	5
France	5
France	5

```
#filters out the HealthyLifeExpectancy values that are less than or equal to 5
#selects for the columns Country_Region by HealthyLifeExpectancy
```

It is important to note that the four France values are not the same province, I deleted the province column for simplicity.

```
# Example using diplyr.select
freedom_and_corruption <- health_happiness %>%
  select('Country_Region','Freedom','Corruption')
kable(head(freedom_and_corruption, 5))
```

Country_Region	Freedom	Corruption
Finland	5	4
Denmark	6	3
Denmark	6	3
Denmark	6	3
Norway	3	8

```
#selects for the columns Country_Region by Freedom and Corruption to show the
```

The same thing goes for this table, the multiple Denmark observations come from the multiple provinces considered in the dataset.

```
# Example using diplyr.arrange
health_gdp <- health_happiness %>%
  filter(Health_exp_pct_GDP_2016 != NaN) %>%
  arrange(desc(Health_exp_pct_GDP_2016)) %>%
  select('Country_Region','Health_exp_pct_GDP_2016')
kable(head(health_gdp, 5))
```

Country_Region	Health_exp_pct_GDP_2016
Armenia	9.9
United Kingdom	9.8000000000000007
Liberia	9.6
Finland	9.5
Zimbabwe	9.4

```
#filters out NA values and arranges in descending order
#selects the Country_Region by the Health_exp_pct_GDP_2016
```

With this code, I filtered out the NA values and arranged the observations in descending order. Only the top 5 values are shown in the table above.

```
# Example using groupby, summarize, mutate
freedom <- health_happiness %>%
  mutate(country_initial = substr(Country_Region,1,1)) %>%
  group_by(country_initial) %>%
    summarize(mean_freedom_rtg = mean(Freedom))
kable(head(freedom, 5))
```

country_initial	mean_freedom_rtg
A	89.00000
B	83.84615
C	75.18750
D	15.25000
E	79.20000

```
#mutates the Country_Region data to give the country initial
#groups the data by country_initial and summarizes the data (calculates means)
```

Here I used the mutate function in order to represent the countries by an assigned initial. I then grouped the data by the country initial and summarized it by calculating the means.

Summarize Numeric Columns

All numeric column summaries

	Result
Ladder_mean	73.00000
SD of Ladder_mean	74.23649
Positive affect_mean	72.56463
Negative affect_mean	75.38095
Social support_mean	73.23810
Freedom_mean	73.40816
Corruption_mean	71.15108
Generosity_mean	75.91837
LogOfGDP_mean	71.19863
HealthyLifeExpectancy_mean	69.47973
Ladder_sd	45.77496
SD of Ladder_sd	45.95789

	Result
Positive affect_sd	44.20145
Negative affect_sd	44.81260
Social support_sd	46.36046
Freedom_sd	45.27719
Corruption_sd	44.79857
Generosity_sd	45.14748
LogOfGDP_sd	43.69749
HealthyLifeExpectancy_sd	43.35967

Here I calculated the mean and standard deviation values for the variables included in the dataset. The mean is a good way to measure the average/middle values of a dataset given that it is normally distributed. The standard deviation represents a good measure of spread of the distribution from the mean. The means and standard deviations of the data seem to be pretty similar across variables, which was super interesting.

Summaries grouped by country initial

country_initial	Ladder_mean	SD of Ladder_mean	Positive affect_mean	Negative affect_mean	Social support_mean	Freedom_mean	Corruption_mean	Generosity_mean	LogOf
A	77.87500	55.75000	94.25000	83.25000	86.25000	89.00000	71.50000	94.25000	
B	87.23077	76.00000	93.61538	86.84615	83.07692	83.84615	NA	91.23077	
C	84.81250	93.81250	64.31250	90.93750	99.18750	75.18750	NA	91.50000	
D	20.75000	48.50000	34.50000	38.75000	16.75000	15.25000	15.25000	39.75000	
E	82.20000	72.20000	66.00000	80.20000	80.60000	79.20000	65.00000	108.60000	
F	19.40000	16.00000	53.00000	54.80000	26.00000	56.20000	17.60000	63.80000	
G	85.62500	101.62500	78.75000	90.00000	106.75000	91.37500	70.75000	94.37500	
H	89.33333	99.33333	80.33333	74.33333	93.66667	109.66667	89.00000	57.00000	
I	68.00000	61.62500	75.12500	93.75000	70.25000	75.12500	71.12500	31.50000	
J	71.66667	90.66667	78.66667	61.66667	55.33333	67.00000	NA	109.66667	
K	79.50000	75.50000	71.75000	38.00000	64.00000	59.25000	NA	40.25000	
L	74.00000	68.28571	94.57143	79.14286	69.57143	86.71429	75.85714	79.00000	
M	84.50000	74.75000	78.83333	73.50000	80.75000	88.91667	92.50000	90.50000	
N	53.11111	74.00000	49.44444	67.22222	56.88889	50.44444	45.33333	54.00000	
P	57.28571	85.42857	55.57143	82.00000	63.42857	49.28571	94.14286	93.28571	
Q	29.00000	86.00000	NA	NA	NA	NA	NA	NA	
R	89.33333	67.33333	76.66667	57.66667	90.00000	61.66667	91.66667	97.66667	
S	66.53846	62.38462	68.00000	63.38462	61.69231	67.46154	NA	62.53846	
T	97.66667	88.66667	89.33333	89.50000	90.66667	91.66667	88.16667	78.83333	
U	47.11111	59.22222	55.77778	55.33333	36.00000	58.55556	NA	31.11111	
V	101.00000	84.00000	99.00000	81.00000	56.50000	84.00000	98.00000	118.00000	
Z	142.00000	134.00000	73.50000	81.00000	112.50000	84.50000	66.00000	97.00000	

In the above table, the means and standard deviations calculated in the above section were placed in columns along with their corresponding country (initial).

Part 4: Make Visualizations

Correlation Matrix Heatmap

```
corr_matrix <- health_happiness %>%
  select(Ladder,'Positive affect','Negative affect',Freedom,Corruption,Generosity) %>%
  cor(use = "p")

melted_cormat <- melt(corr_matrix)

ggheatmap <- ggplot(melted_cormat, aes(Var2, Var1, fill = value))+
  geom_tile(color = "white") +
  scale_fill_gradient2(low = "black", high = "#cc5a2a", mid = "white",
    midpoint = 0, limit = c(-1,1), space = "Lab",
    name="Pearson\nCorrelation") +
  geom_text(aes(Var2, Var1, label = sprintf("%0.3f", round(value, digits = 3))), color = "black", size = 4) +
  theme(
    axis.title.x = element_blank(),
    axis.title.y = element_blank(),
    panel.grid.major = element_blank(),
    panel.border = element_blank(),
    panel.background = element_blank(),
    axis.ticks = element_blank()
   )

print(ggheatmap)
```

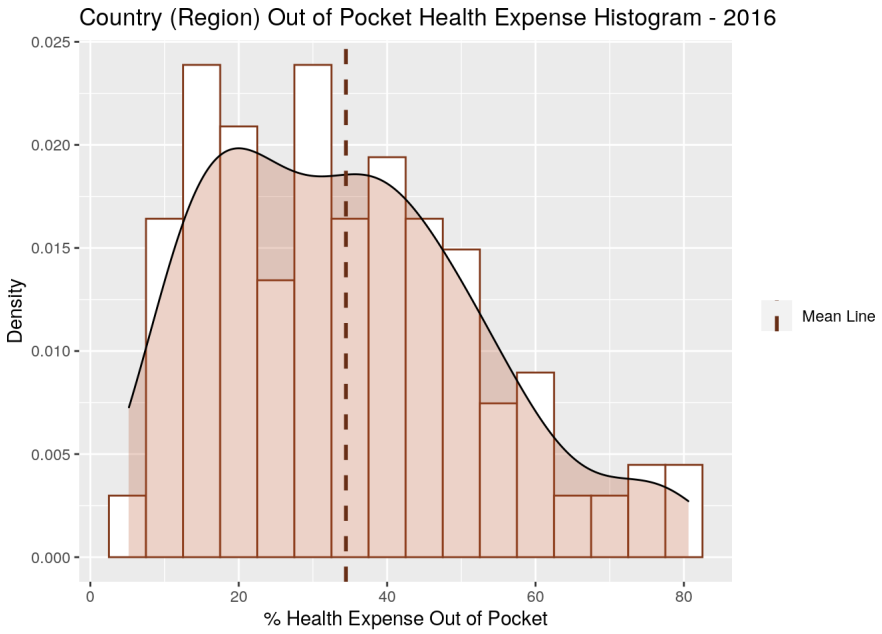
Heatmap Reference (http://www.sthda.com/english/wiki/ggplot2-quick-correlation-matrix-heatmap-r-software-and-data-visualization)

The heatmap above gives the correlation between the variables in the health systems dataset. It was interesting to me that the greatest positive correlation was between Freedom and Positive. As a future health care professional, I value the answer to the question "what makes a person happy?" and I believe it has a lot to do with an individual's health. It will be important to keep in mind the weight of each of these variables in my future profession.

Histogram with Density Plot

```
# Histogram with density plot
tidydf <- health_happiness %>%
  filter(Health_exp_out_of_pocket_pct_2016 != NaN) %>%
  mutate(Health_exp_out_of_pocket_pct_2016 = as.numeric(Health_exp_out_of_pocket_pct_2016))

#hist(tidydf$Health_exp_out_of_pocket_pct_2016)
ggplot(tidydf, aes(x=Health_exp_out_of_pocket_pct_2016)) +
  geom_histogram(aes(y=..density..), binwidth=5, color="#80381a", fill="white") +
  geom_density(alpha=.25, fill="#b34e24") +
  geom_vline(aes(xintercept=mean(Health_exp_out_of_pocket_pct_2016), color="Mean Line", label="Mean"), linetype=
"dashed", size=1) +
  labs(
    title="Country (Region) Out of Pocket Health Expense Histogram - 2016",
    x="% Health Expense Out of Pocket",
    y = "Density"
  ) +
  theme(legend.title = element_blank()) +
  scale_color_manual(values = c("#662d15"))
```

```
## Warning: Ignoring unknown aesthetics: label
```

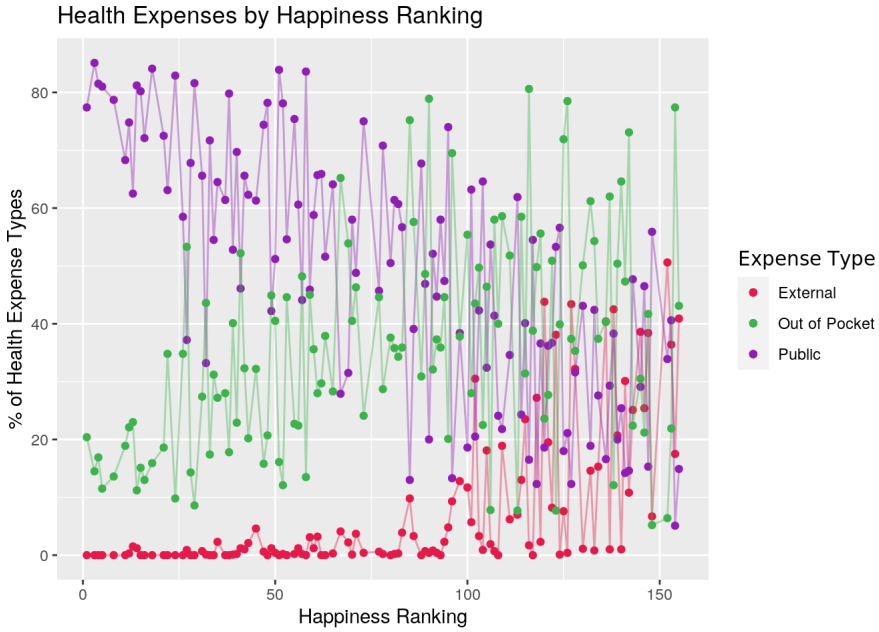


This histogram represents the distribution of % health expenses out of pocket in the countries included in the datasets. It is right skewed with a mean value of about 35. Histogram Reference (http://www.sthda.com/english/wiki/ggplot2-histogram-plot-quick-start-guide-r-software-and-data-visualization)

Third Plot

```
# Plot Health Expense Breakdown versus Happiness Rank (Ladder)
ladder_data <- health_happiness %>%
  mutate_all(type.convert) %>%
  filter(!is.nan(Health_exp_public_pct_2016)) %>%
  filter(!is.nan(Health_exp_out_of_pocket_pct_2016)) %>%
  filter(!is.nan(External_health_exp_pct_2016)) %>%
  group_by(Ladder) %>%
    summarise(
      external_exp = min(External_health_exp_pct_2016),
      public_exp = min(Health_exp_public_pct_2016),
      outofpocket_exp = min(Health_exp_out_of_pocket_pct_2016)
    )

# Plot triple line graph
line_alpha <- 0.4
ggplot(ladder_data, aes(x=Ladder)) +
  geom_line(aes(y = external_exp), color = "#e6194B", alpha = line_alpha) +
  geom_point(aes(y = external_exp, color = "External")) +
  geom_line(aes(y = public_exp), color = "#911eb4", alpha = line_alpha) +
  geom_point(aes(y = public_exp, color = "Public", labels="Public")) +
  geom_line(aes(y = outofpocket_exp), color = "#3cb44b", alpha = line_alpha) +
  geom_point(aes(y = outofpocket_exp, color = "Out of Pocket", labels="Out of Pocket")) +
  labs(title="Health Expenses by Happiness Ranking",x="Happiness Ranking", y = "% of Health Expense Types", colou
r="Expense Type") +
  theme(legend.title=element_text("Legend"), legend.box  ="vertical", legend.position = "right") +
  scale_color_manual(values = c("#e6194B","#3cb44b","#911eb4"))
```

```
## Warning: Ignoring unknown aesthetics: labels

## Warning: Ignoring unknown aesthetics: labels
```



This plot was one that I found incredibly interesting. You can see the about 150 ranked countries involved in the dataset and their corresponding external, out of pocket and public expense types. The smaller the number on the x axis (Happiness Ranking), the higher it was ranked. The external expense type sees very little change, however the bottom ranked 50 or so countries see a bit of an increase in external expenses. Most notable was the change that could be seen around the rank 75. The highest rank half of the countries (1-75) saw higher public health expenses over out of pocket expenses. This changes for the lower ranked countries (76 to about 150) in which the out of pocket expenses climb and become much larger than the public expenses seen around rank 150. To me this means that happiness ranking does depend on the type of expenses individuals experience in their health systems. People are happier when they are paying less out of pocket for their health care. This would be interesting to further study and see if this pattern can be replicated.

K-Means Clustering

```
numerics <- health_happiness %>% distinct(Country_Region, .keep_all = TRUE)
numerics <- numerics %>%
  select(Country_Region, Freedom, Corruption, Health_exp_out_of_pocket_pct_2016, 'Physicians_per_1000_2009-18') %
>%
  mutate_all(type.convert) %>%
  filter(!is.na(Freedom)) %>%
  filter(!is.na(Corruption)) %>%
  filter(!is.na(Health_exp_out_of_pocket_pct_2016)) %>%
  filter(!is.na(`Physicians_per_1000_2009-18`)) %>%
  as.data.frame()


rownames(numerics) <- numerics$Country_Region
numerics <- subset(numerics, select = -c(Country_Region) )
print(numerics)
```

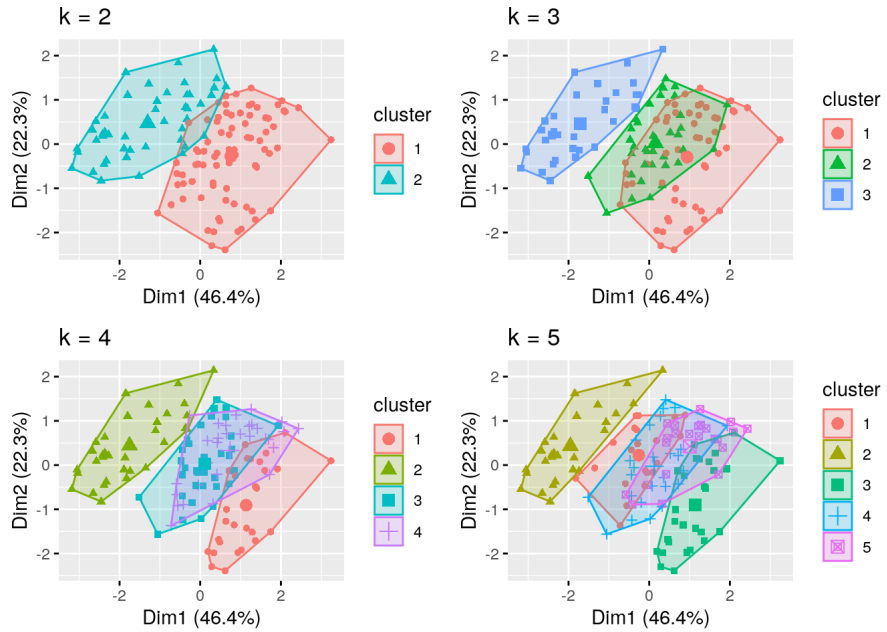
```
##                          Freedom Corruption Health_exp_out_of_pocket_pct_2016
## Finland                        5          4                                20.4
## Denmark                        6          3                                13.7
## Norway                         3          8                                14.5
## Iceland                        7         45                                16.9
## Netherlands                   19         12                                11.5
## Switzerland                   11          7                                29.6
## Sweden                        10          6                                15.2
## New Zealand                    8          5                                13.6
## Canada                         9         11                                14.6
## Austria                       26         19                                18.9
## Australia                     17         13                                18.9
## Costa Rica                    16         58                                22.1
## Israel                        93         74                                23.0
## Luxembourg                    28          9                                11.2
## Ireland                       33         10                                13.0
## Germany                       44         17                                12.4
## Belgium                       53         20                                15.9
## Malta                         12         32                                34.8
## Mexico                        71         87                                40.4
## France                        69         21                                 9.8
## Chile                         98         99                                34.8
## Guatemala                     25         82                                53.3
## Spain                         95         78                                23.8
## Panama                        32        104                                27.4
## Brazil                        84         71                                43.6
## Uruguay                       30         33                                17.4
## Singapore                     20          1                                31.2
## El Salvador                   74         85                                27.2
## Italy                        132        128                                23.1
## Slovakia                     108        142                                17.8
## Trinidad and Tobago           51        141                                40.1
## Poland                        52        108                                22.9
## Uzbekistan                     1         18                                52.2
## Lithuania                    122        113                                32.3
## Colombia                      56        124                                20.2
## Slovenia                      13         97                                12.0
## Nicaragua                     70         43                                32.2
## Argentina                     54        109                                15.8
## Romania                       57        146                                20.7
## Cyprus                        81        115                                44.9
## Ecuador                       42         68                                40.5
## Thailand                      18        131                                12.1
## Latvia                       126         92                                44.6
## Estonia                       45         30                                22.7
## Jamaica                       49        130                                22.4
## Mauritius                     40         96                                48.2
## Japan                         64         39                                13.5
## Honduras                      39         79                                45.0
## Kazakhstan                    80         57                                35.6
## Bolivia                       35         91                                28.0
## Hungary                      138        140                                29.7
## Paraguay                      34         76                                37.9
## Peru                          61        132                                28.3
## Portugal                      37        135                                27.8
## Pakistan                     114         55                                65.2
## Russia                       107        127                                40.5
## Philippines                   15         49                                53.9
## Serbia                       124        118                                40.5
## Moldova                      128        148                                46.3
## Montenegro                   139         77                                24.1
## Croatia                      118        139                                15.4
## Dominican Republic            43         52                                44.6
## Bosnia and Herzegovina       137        145                                28.7
## Turkey                       140         50                                16.5
## Malaysia                      36        137                                37.6
## Belarus                      131         37                                35.8
## Greece                       150        123                                34.3
## Mongolia                     112        119                                35.9
## Nigeria                       75        114                                75.2
## Kyrgyzstan                    38        138                                57.6
## Algeria                      149         46                                30.9
## Morocco                       76         84                                48.6
## Azerbaijan                   101         22                                78.9
## Lebanon                      136        133                                32.1
## Indonesia                     48        129                                37.3
## Vietnam                       23         86                                44.6
## Bhutan                        59         25                                20.1
## Cameroon                      90        120                                69.5
## Bulgaria                     115        147                                48.0
## Ghana                         91        117                                37.8
```

```
## Nepal                          67         65                          55.4
## Benin                         103         75                          43.5
## Congo (Brazzaville)            92         60                          49.7
## Gabon                         119        103                          22.5
## Laos                           22         27                          46.4
## South Africa                   85        102                           7.8
## Albania                        87        134                          58.0
## Cambodia                        2         94                          58.6
## Senegal                       121         88                          51.8
## Namibia                        82         98                           7.7
## Niger                         111         51                          58.5
## Burkina Faso                  127         47                          31.4
## Armenia                       123         93                          80.6
## Iran                          117         44                          38.8
## Guinea                        109         70                          49.8
## Georgia                       104         28                          55.6
## Gambia                         89         26                          23.6
## Kenya                          72        105                          27.7
## Mauritania                    151         67                          50.9
## Mozambique                     46         40                           7.7
## Tunisia                       143        101                          39.9
## Bangladesh                     27         36                          71.9
## Iraq                          130         66                          78.5
## Congo (Kinshasa)              125        106                          37.4
## Mali                          110        107                          35.3
## Sri Lanka                      55        111                          50.1
## Chad                          142         80                          61.2
## Ukraine                       141        143                          54.3
## Ethiopia                      106         53                          37.4
## Uganda                         99         95                          40.3
## Egypt                         129         89                          62.0
## Zambia                         73         69                          12.1
## Togo                          120         72                          50.4
## India                          41         73                          64.6
## Liberia                        94        126                          47.3
## Comoros                       148         81                          73.1
## Madagascar                    146        116                          22.4
## Burundi                       135         23                          30.5
## Zimbabwe                       96         63                          21.2
## Haiti                         152         48                          41.7
## Botswana                       60         54                           5.2
## Rwanda                         21          2                           6.4
## Tanzania                       78         34                          21.9
## Afghanistan                   155        136                          77.4
## Central African Republic      133        122                          43.1
##                         Physicians_per_1000_2009-18
## Finland                                         3.8
## Denmark                                         4.5
## Norway                                          4.6
## Iceland                                         4.0
## Netherlands                                     3.5
## Switzerland                                     4.2
## Sweden                                          5.4
## New Zealand                                     3.0
## Canada                                          2.6
## Austria                                         5.1
## Australia                                       3.6
## Costa Rica                                      1.1
## Israel                                          3.2
## Luxembourg                                      3.0
## Ireland                                         3.1
## Germany                                         4.2
## Belgium                                         3.3
## Malta                                           3.8
## Mexico                                          2.2
## France                                          3.2
## Chile                                           1.1
## Guatemala                                       0.4
## Spain                                           4.1
## Panama                                          1.6
## Brazil                                          2.1
## Uruguay                                         5.0
## Singapore                                       2.3
## El Salvador                                     1.6
## Italy                                           4.1
## Slovakia                                        2.5
## Trinidad and Tobago                             2.7
## Poland                                          2.4
## Uzbekistan                                      2.4
## Lithuania                                       4.3
## Colombia                                        2.1
```

```
## Slovenia                          3.0
## Nicaragua                         1.0
## Argentina                         4.0
## Romania                           2.3
## Cyprus                            2.0
## Ecuador                           2.1
## Thailand                          0.8
## Latvia                            3.2
## Estonia                           3.5
## Jamaica                           1.3
## Mauritius                         2.0
## Japan                             2.4
## Honduras                          0.3
## Kazakhstan                        3.3
## Bolivia                           1.6
## Hungary                           3.2
## Paraguay                          1.4
## Peru                              1.3
## Portugal                          3.3
## Pakistan                          1.0
## Russia                            4.0
## Philippines                       1.3
## Serbia                            3.1
## Moldova                           3.2
## Montenegro                        2.3
## Croatia                           3.0
## Dominican Republic                1.6
## Bosnia and Herzegovina            2.0
## Turkey                            1.8
## Malaysia                          1.5
## Belarus                           4.1
## Greece                            4.6
## Mongolia                          2.9
## Nigeria                           0.4
## Kyrgyzstan                        1.9
## Algeria                           1.8
## Morocco                           0.7
## Azerbaijan                        3.4
## Lebanon                           2.3
## Indonesia                         0.4
## Vietnam                           0.8
## Bhutan                            0.4
## Cameroon                          0.1
## Bulgaria                          4.0
## Ghana                             0.2
## Nepal                             0.7
## Benin                             0.2
## Congo (Brazzaville)               0.1
## Gabon                             0.4
## Laos                              0.5
## South Africa                      0.9
## Albania                           1.2
## Cambodia                          0.2
## Senegal                           0.1
## Namibia                           0.4
## Niger                             0.1
## Burkina Faso                      0.1
## Armenia                           2.9
## Iran                              1.1
## Guinea                            0.1
## Georgia                           5.1
## Gambia                            0.1
## Kenya                             0.2
## Mauritania                        0.2
## Mozambique                        0.1
## Tunisia                           1.3
## Bangladesh                        0.5
## Iraq                              0.8
## Congo (Kinshasa)                  0.1
## Mali                              0.1
## Sri Lanka                         1.0
## Chad                              0.0
## Ukraine                           3.0
## Ethiopia                          0.1
## Uganda                            0.1
## Egypt                             0.8
## Zambia                            0.1
## Togo                              0.0
## India                             0.8
## Liberia                           0.0
## Comoros                           0.2
```

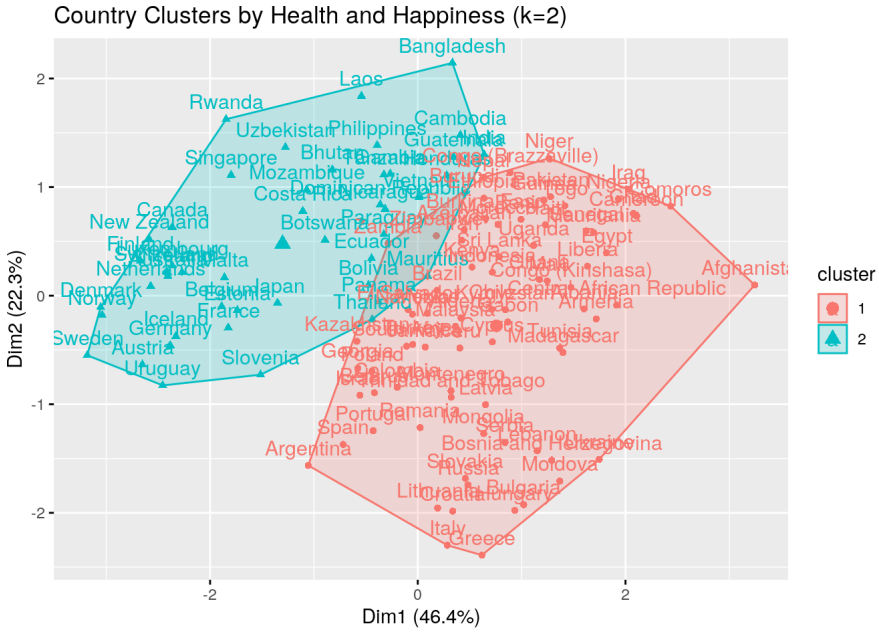
```
## Madagascar                          0.2
## Burundi                             0.1
## Zimbabwe                            0.1
## Haiti                               0.2
## Botswana                            0.4
## Rwanda                              0.1
## Tanzania                            0.0
## Afghanistan                         0.3
## Central African Republic            0.1
```

```
k2 <- kmeans(numerics, centers = 2, nstart = 50)
k3 <- kmeans(numerics, centers = 3, nstart = 50)
k4 <- kmeans(numerics, centers = 4, nstart = 50)
k5 <- kmeans(numerics, centers = 5, nstart = 50)

p2 <- fviz_cluster(k2, geom = "point", data = numerics) + ggtitle("k = 2")
p3 <- fviz_cluster(k3, geom = "point",  data = numerics) + ggtitle("k = 3")
p4 <- fviz_cluster(k4, geom = "point",  data = numerics) + ggtitle("k = 4")
p5 <- fviz_cluster(k5, geom = "point",  data = numerics) + ggtitle("k = 5")

grid.arrange(p2, p3, p4, p5, nrow = 2)
```



Based on the above clusters, it seems as though the two cluster graph shows the optimal cluster number, as the two clusters have the least overlap. If I were to do this again, I may restric the number of countries involved because the clusters are overwheliming as there are so many data points invovled.

```
fviz_cluster(k2, data=numerics) +
  labs(title="Country Clusters by Health and Happiness (k=2)")
```


Country Clusters by Health and Happiness (k=2)
Dim2 (22.3%)
Dim1 (46.4%)
cluster
1
2
Bangladesh
Laos
Rwanda
Cambodia
Uzbekistan
Philippines
Singapore
Mozambique
Canada
New Zealand
Botswana
Ecuador
Mauritius
Bolivia
Brazil
Japan
Latvia
France
Iceland
Germany
Sweden
Austria
Uruguay
Slovenia
Denmark
Norway
Niger
Iraq
Comoros
Cameroon
Uganda
Egypt
Liberia
Sri Lanka
Kenya
Congo (Kinshasa)
Afghanistan
Central African Republic
Algeria
Madagascar
Tunisia
Malaysia
Gabon
Latvia
Trinidad and Tobago
Romania
Mongolia
Portugal
Spain
Serbia
Lebanon
Bosnia and Herzegovina
Argentina
Slovakia
Russia
Moldova
Bulgaria
Croatia
Hungary
Italy
Greece