Impacts on Country Happiness

Importing the first data sets and clean them

The primary dataset for the project's name is "Happy." The CSV comes in as a singular row containing both the Country name and the World Happiness index score associated with that country. Clean Data requires each observation to have an individual row, so the separate command was used. Then that data set was joined with the Democracy index data set. This data set includes an overall democracy index score, as well as countries ratings in various vital variables relating to democracy, such as the functioning of the government, political participation, political regime type, and more.

```
## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 1 rows [76].
```

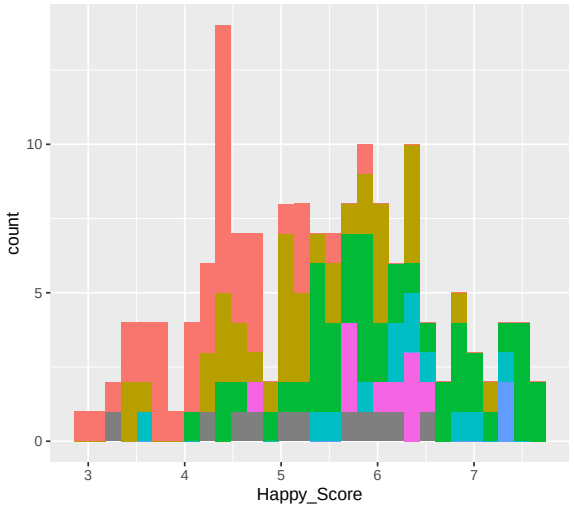
```
## Observations: 156
## Variables: 11
## $ Country                          <chr> "Finland", "Norway", "Denmark"...
## $ Happy_Score                      <dbl> 7.632, 7.594, 7.555, 7.495, 7....
## $ Rank                             <chr> "8", "1", "5", "2", "10", "11"...
## $ Score                            <dbl> 9.14, 9.87, 9.22, 9.58, 9.03, ...
## $ Electoral_Process.and.Pluralism <dbl> 10.00, 10.00, 10.00, 10.00, 9....
## $ Functioning_of_Gouv              <dbl> 8.93, 9.64, 9.29, 9.29, 9.29, ...
## $ Political_Participation          <dbl> 8.33, 10.00, 8.33, 8.89, 7.78,...
## $ Political_Cultyre                <dbl> 8.75, 10.00, 9.38, 10.00, 9.38...
## $ Civil_Liberties                  <dbl> 9.71, 9.71, 9.12, 9.71, 9.12, ...
## $ Regime_type                      <chr> "Full democracy", "Full democr...
## $ Continent                        <chr> "Europe", "Europe", "Europe", ...
```

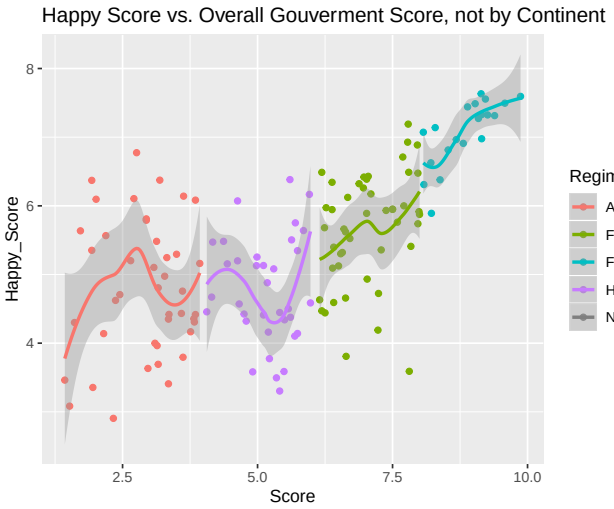
Begin Exploring the Data

Exploring the data, we can see happiness scores available in the countries with democratic information range from about 2.905 to 7.632. At the bottom are many third world Asian and African countries such as Syria and Rwanda, while the top contains many first-world Oceanic and European countries (also Canada) such as Australia and Denmark. An average score is 5.376, and the 50th percentile is 5.37. These values being so close together suggests there is no significant skew of the data. The histogram shows us that the most frequent happiness scores are between 4.333 and 4.5. Next, we can see that the overall democracy index score is highly related to regime type and happiness. That relationship suggested this data is worth exploring. The following plot shows us this relationship holds regardless of continent, and finally which regimes are the most common, and where they are in the world. Full Democracies, for example, are the least common type overall and most prevalent in Europe.

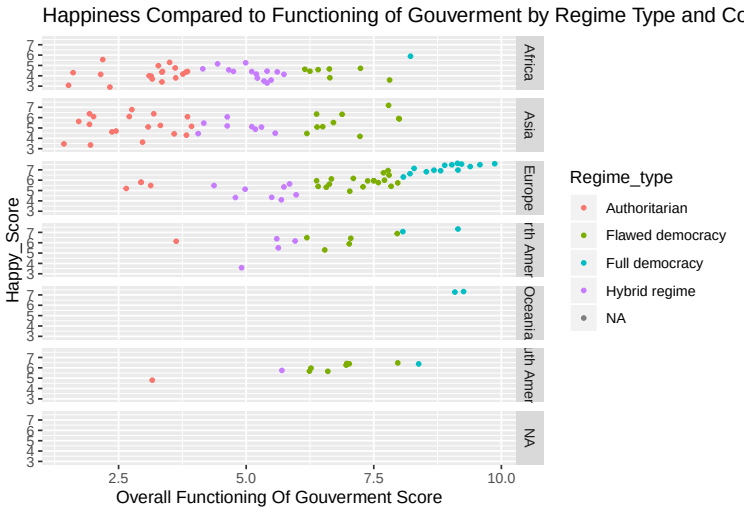
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   2.905   4.452   5.358   5.376   6.170   7.632       1
```

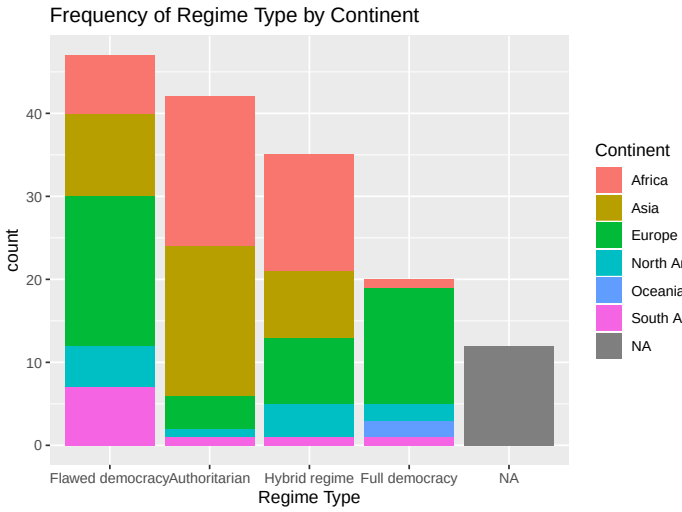
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```






```
## Warning: Ignoring unknown parameters: method, formula
```





Run First Model of the Data Since it began to appear as though the overall democracy index and happiness scores are highly related to the regime type, so to check this theory, I put the variables from this dataset into a regression to estimate the effect regime type had on happiness score. To do this, 0-1 variables were created for each regime type to be put into a regression alongside other democracy index variables.

```
### Create 0-1 Variables for each of the regime types in the dataset
For_Stats_Graph_Join_Numeric3 <- For_Stats_Graph_Join_Numeric1 %>%
  mutate(Is_Full_Democracy = ifelse(Regime_type == "Full democracy", 1, 0 )) %>%
  mutate(Is_Flawed_Democracy = ifelse(Regime_type == "Flawed democracy",1,0)) %>%
  mutate(Is_Hybrid_Regime = ifelse(Regime_type == "Hybrid regime",1,0)) %>%
  mutate(Is_Authoritarian = ifelse(Regime_type == "Authoritarian",1,0)) %>%
  mutate(Is_Missing = ifelse(Regime_type == "NA",1,0))

### Run regression with categorical variables to be interperted as relative to is Hybr
Updated_Regression <- lm(data=For_Stats_Graph_Join_Numeric3, Happy_Score ~ Electoral_Pro
summary(Updated_Regression)
```

```
##
```

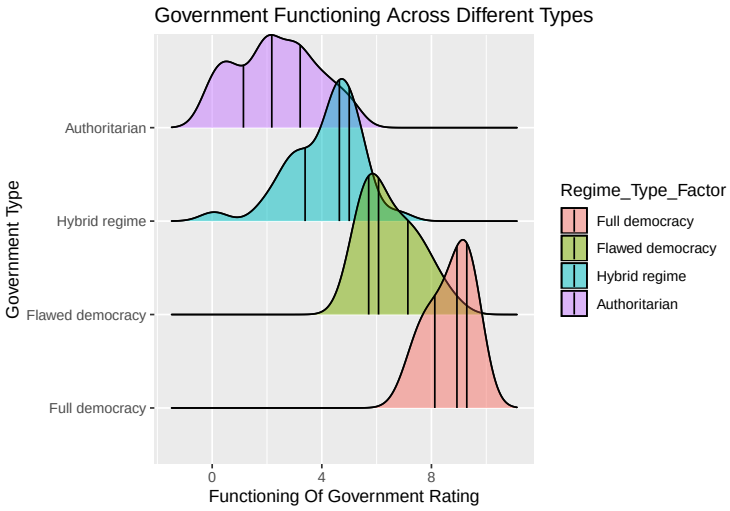
```
## Call:
## lm(formula = Happy_Score ~ Electoral_Process.and.Pluralism +
##     Functioning_of_Gouv + Political_Participation + Political_Cultyre +
##     Civil_Liberties + Is_Full_Democracy + Is_Flawed_Democracy +
##     Is_Authoritarian, data = For_Stats_Graph_Join_Numeric3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.23247 -0.47778  0.04975  0.50468  1.66387
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      4.06085    0.59352   6.842 2.48e-10 ***
## Electoral_Process.and.Pluralism -0.04057    0.06256  -0.649  0.51775
## Functioning_of_Gouv              0.23609    0.05720   4.128 6.38e-05 ***
## Political_Participation         -0.03094    0.06048  -0.512  0.60980
## Political_Cultyre               -0.07183    0.06676  -1.076  0.28389
## Civil_Liberties                  0.09115    0.07985   1.141  0.25569
## Is_Full_Democracy                1.36074    0.45977   2.960  0.00364 **
## Is_Flawed_Democracy              0.34443    0.26866   1.282  0.20204
## Is_Authoritarian                 0.44511    0.32969   1.350  0.17924
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7868 on 135 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.5476, Adjusted R-squared:  0.5208
## F-statistic: 20.43 on 8 and 135 DF,  p-value: < 2.2e-16
```

Plot the relationship

It appears as though regime type is not as important as I may have expected outside of full democracies. Using the regression, it appears that the actual functioning of Government rating is generally more critical than regime type. Full democracy had a statistically significant increase in happiness rating independent of other democracy index indicators. Breaking this down with the ridge plot, the functioning of government rating is highly related to the type of government and appears to be more important at driving happiness scores up or down. While not what was originally expected, this could make sense because many of the countries at the bottom of the functioning of government ratings are in shambles. Think about war-torn countries like Syria and Afganistan, for example. Their government's ability to function properly prevents them from providing basic needs to citizens, and those basic needs likely matters more than political ideals to people more worried about where their next meal will come from than democratic freedom. As for full democracies, most if not all of the most affluent, safe, and happy places on earth are full democracies like Canada, which likely explains why that had a significant co-efficient.

```
## [1] "Authoritarian"    "Flawed democracy" "Full democracy"
## [4] "Hybrid regime"
```

```
## Picking joint bandwidth of 0.493
## Picking joint bandwidth of 0.493
```



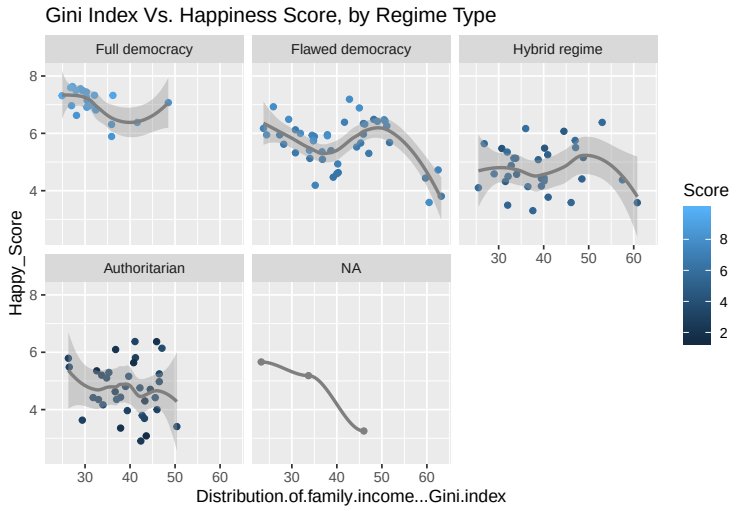
Begin Adding Other datasets To Examine Demographic information Now time to add some additional data sets to look at demographic information. The first data set loaded in was called "GINI" which contains information on various country's Gini Coefficient. This can range from 0-100, where 0 is perfect wealth equality, or each person in a country having the same level of aggregate wealth. On the contrary, 100 would be perfect wealth inequality, where the wealthiest person owned all of the wealth. Next was the public debt dataset, which contains information on public debt as a percentage of GDP by country. Then there are the GDP Composition dataset breaks down what sectors the country's GDP comes from as a % of total GDP. The sectors showed are industry, the agricultural sector, and the service industry. Next, I added GDP per capita, which breaks down the GDP divided by the population for a variety of countries. Then education expenditure as a percentage of GDP was added. Finally, After education expenditure, both youth and overall unemployment percentage. Some code was also added to make the variable names easier to understand.

Begin Exploring all the Demographic Data

The first variable I chose to look at was the Gini co-efficient data set. Intuitively it is easy to imagine how a relatively low Gini coefficient leads to a high overall happiness score because more people get a piece of the countries economic pie. At first glance, the Gini Co-efficient's in full democracies were lower on average, suggesting they had more wealth equality.

```
### Set Regime Type as a factor, then change the levels
for_graph_new <- Next_Master %>%
  transform(Regime_Type_Factor = as.factor(Regime_type))

for_graph_new$Regime_Type_Factor <- factor(for_graph_new$Regime_Type_Factor,
                                           levels = c("Full democracy", "Flawed democracy",
```



Isolate Gini Index

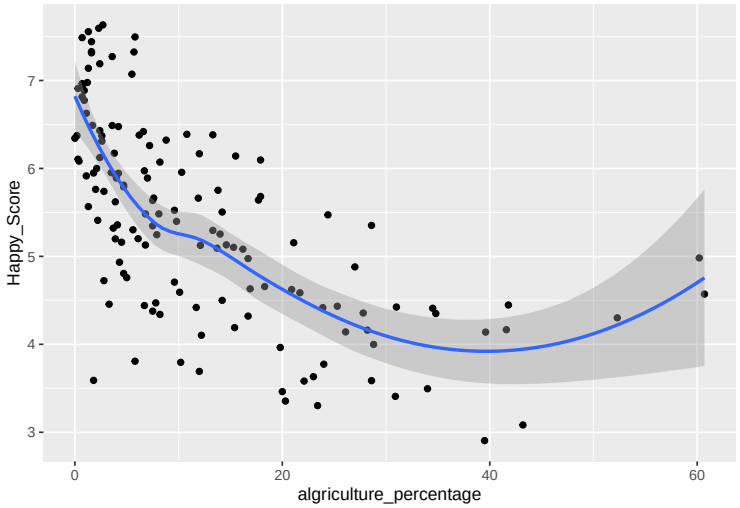
```
Updated_Regression2 <- lm(data=Master_9, Happy_Score ~ Electoral_Process.and.Pluralism +
summary(Updated_Regression2)
```

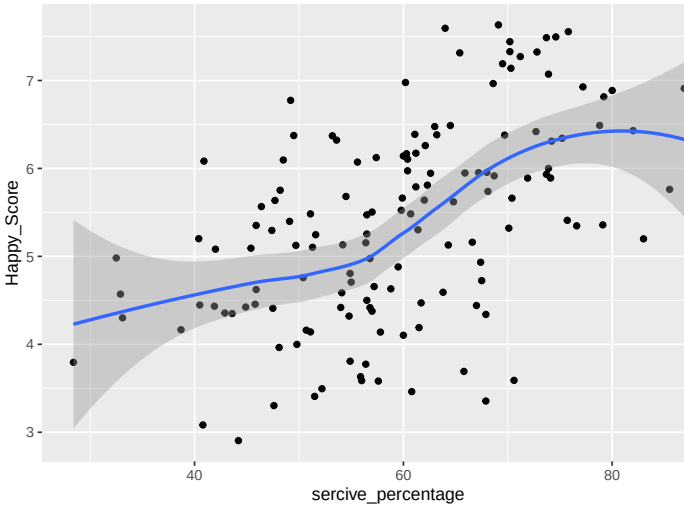
```
##
## Call:
## lm(formula = Happy_Score ~ Electoral_Process.and.Pluralism +
##     Functioning_of_Gouv + Political_Participation + Political_Cultyre +
##     Civil_Liberties + Is_Full_Democracy + Is_Flawed_Democracy +
##     Is_Authoritarian + Gini_Index, data = Master_9)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.82488 -0.50173  0.02418  0.42367  1.71461
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     4.760472   0.760285   6.261 6.03e-09 ***
## Electoral_Process.and.Pluralism -0.014061   0.064758  -0.217 0.828468
## Functioning_of_Gouv              0.245128   0.062524   3.921 0.000147 ***
## Political_Participation         -0.028226   0.062245  -0.453 0.651023
## Political_Cultyre               -0.079120   0.066849  -1.184 0.238904
## Civil_Liberties                  0.042798   0.082569   0.518 0.605174
## Is_Full_Democracy                1.373245   0.473599   2.900 0.004439 **
## Is_Flawed_Democracy              0.447421   0.277418   1.613 0.109392
## Is_Authoritarian                 0.439928   0.361218   1.218 0.225631
## Gini_Index                      -0.017272   0.008503  -2.031 0.044432 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7653 on 121 degrees of freedom
##   (25 observations deleted due to missingness)
## Multiple R-squared:  0.5866, Adjusted R-squared:  0.5558
## F-statistic: 19.07 on 9 and 121 DF,  p-value: < 2.2e-16
```

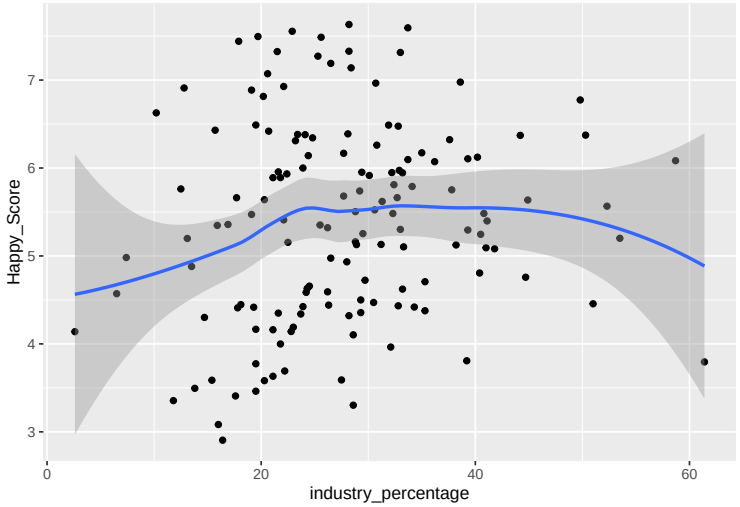
Using a multiple linear regression to account for the Gini Co-efficient's correlation with regime type, it began to look like wealth equality, is an important factor in the happiness equation where an increasing Gini Co-efficient decreased overall happiness, on average. ## Additional Insights

To go along with the Gini Co-efficient data, the variables chosen to be analyzed were Youth Unemployment Percentage and agriculture industry percentage (as % of GDP) because they were the variables with the most robust relationship with happiness score holding the other variables constant.

Note industry percentage was discluded to avoid co-linearity. The 3 percentages usually added up to 100%, so one of the variables had to be discluded to avoid multicollinearity, much like the dummy variable. To decide which one got removed, scatterplots were used to estimate which 2 composition percentages should be used.







The relationship between the agricultural sector is non-linear, so a square term was added to account for the quadratic relationship.

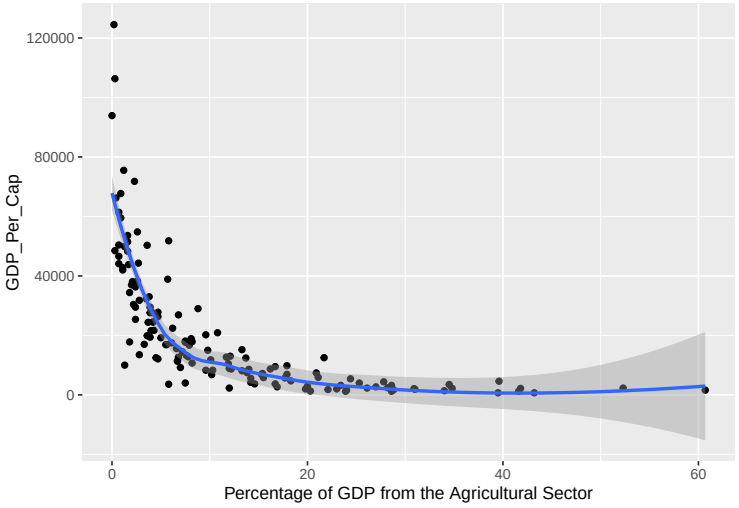
```
##
## Call:
## lm(formula = Happy_Score ~ Functioning_of_Gouv + Political_Participation +
##     Political_Cultyre + Civil_Liberties + Is_Full_Democracy +
##     Is_Authoritarian + Is_Hybrid_Regime + Gini_Index + Youth_Unemployment_Perc +
##     Unemployment_Perc + algriculture_percentage + sercive_percentage +
##     agriculutre_square, data = master_10)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -1.77839 -0.39221  0.01583  0.37337  1.50808
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 6.5922939  0.8998264   7.326 4.51e-11 ***
## Functioning_of_Gouv         0.1099559  0.0579758   1.897 0.060555 .
## Political_Participation    -0.0480172  0.0540774  -0.888 0.376549
## Political_Cultyre          -0.0962870  0.0562694  -1.711 0.089919 .
## Civil_Liberties             0.0478873  0.0619810   0.773 0.441439
## Is_Full_Democracy           0.9675369  0.2639120   3.666 0.000384 ***
## Is_Authoritarian            0.0624961  0.3768657   0.166 0.868600
## Is_Hybrid_Regime           -0.0038167  0.2327460  -0.016 0.986947
## Gini_Index                 -0.0063484  0.0071543  -0.887 0.376863
## Youth_Unemployment_Perc    -0.0233499  0.0060088  -3.886 0.000176 ***
## Unemployment_Perc          -0.0092758  0.0065218  -1.422 0.157827
## algriculture_percentage    -0.0929632  0.0174340  -5.332 5.37e-07 ***
## sercive_percentage          0.0020350  0.0078411   0.260 0.795726
## agriculutre_square          0.0011328  0.0003385   3.346 0.001128 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.606 on 108 degrees of freedom
##   (34 observations deleted due to missingness)
## Multiple R-squared:  0.7404, Adjusted R-squared:  0.7092
## F-statistic:  23.7 on 13 and 108 DF,  p-value: < 2.2e-16
```

Note the co-efficient on Gini coefficient has inched closer to 0, making it less impactful of a variable than in previous models. The estimate on the Gini Index is increasing towards 0 means it was likely the relationship between Gini Index and other indicators like Youth Unemployment rate that drove down the Gini Index estimate.

Examine Agricultural Relationship

As economies have developed worldwide, they generally have moved from the agricultural sector towards the service industry. As a result, the percentage of GDP from the agricultural sector might be a useful proxy for how developed the economy is. Using GDP as a proxy for development, this relationship holds.



The relationship between the percentage of

GDP from the agricultural sector closely mirrors our proxy for development. As the portion of a Nation's GDP coming from the agricultural sector decreases, the overall wealth of that country tends to be high. The relationship suggests demographic information such as which sectors a countries GDP is made up of is essential because they can give a good proxy on how developed the countries economy is.

Overall the demographic and political information included can explain around 70% of the variance in happiness by countries. The most important factors are if the country is a full democracy. If so, we can expect almost a full point increase in happiness relative to a flawed democracy. Then there is the functioning of the government rating. As the functioning of the government increases by 1 point on the democracy index's scale, happiness increases by about 0.1, on average. From the demographic information, youth unemployment percentage is an important variable, where the lower youth unemployment percentage, the better, even independent of its relationship with overall unemployment percentage. Finally, there is which sector the countries GDP comes from. There is a quadratic relationship between the % of a countries GDP coming from the agricultural industry and happiness because as Nations develop, they tend to move away from the agricultural sector.

Technology

The focus on the technology analysis is to determine the impact of technology on perceived happiness. To begin, we will look at the democracy index dataset. This table is only being used for its Country and Continent categorical variables. As this will be joined to another dataset in the future.

```
## # A tibble: 5 x 2
##   Country     Continent
##   <chr>       <chr>
## 1 Norway      Europe
## 2 Iceland     Europe
## 3 Sweden      Europe
## 4 New Zealand Oceania
## 5 Denmark     Europe
```

Next, we will import the happiness score dataset from the World Happiness Report 2018. It has two variables Country (categorical variable) and Happiness score (continuous variable) rated from 1 to 10. A few basic statistics of this dataset include a minimum value of 2.905, a maximum value of 7.632, and a mean of 5.376. Based on the 25th and 75th percentiles it appears most coutries have a happiness score inbetween 4.454 and 6.168.

```
##    Country          Happiness score
##  Length:156         Min.   :2.905
##  Class :character   1st Qu.:4.454
##  Mode  :character   Median :5.378
##                     Mean   :5.376
##                     3rd Qu.:6.168
##                     Max.   :7.632
```

Before we join the datasets together, the possible NA values will want to be noted. I can see a few of the country names have a few subtle differences. I accommodated for this using the stringr library, for example changing "Trinidad & Tobago" to "Trinidad and Tobago". We will now join the datasets together; this will allow countries to have an additional Continent variable.

```
## # A tibble: 5 x 3
##   Country     Happiness_score Continent
```

```
##   <chr>                <dbl> <chr>
## 1 Norway                7.59 Europe
## 2 Iceland               7.50 Europe
## 3 Sweden                7.31 Europe
## 4 New Zealand           7.32 Oceania
## 5 Denmark               7.56 Europe
```

With several NA values for the happiness score column,for the purposes of this study they are replaced with the mean of the happiness score column, to include as many countries as possible. This will allow for statistics from more countries will be included in the analysis.

```
## # A tibble: 5 x 3
##   Country     Happiness_score Continent
##   <chr>                <dbl> <chr>
## 1 Norway                7.59 Europe
## 2 Iceland               7.50 Europe
## 3 Sweden                7.31 Europe
## 4 New Zealand           7.32 Oceania
## 5 Denmark               7.56 Europe
```

After noticing several continents with low numbers of countries included in them, I decided to refactor the continents into four groups. Africa, Asia, Europe, and the Americas, this keeps the count for each continent closer to each other for easier comparison.

```
## # A tibble: 6 x 2
##   Continent         n
##   <chr>         <int>
## 1 Africa           50
## 2 Asia             42
## 3 Europe           45
## 4 North America    14
## 5 Oceania           4
## 6 South America    12
```

```
## # A tibble: 4 x 2
##   Continent     n
##   <fct>     <int>
## 1 Africa       50
## 2 Asia         46
## 3 Europe       45
## 4 Americas     26
```

A plot of the happiness scores by continent shows most countries in Europe have the highest happiness scores on their 50th-75th percentile, as North America has a higher happiness score on their 25th-50th percentile.

The next 3 data sets being used are internet users, phone subscribers, and country population. All the data sets have a qualitative categorical variable in Country. Internet users dataset includes the total amount of internet users in the Country as a quanitative continuous variable. Phone subscribers has the total amount of phone subscriptions in each respective Country. Finally, the population data set has the sum population of each Country. The goal of this data is to examine if phone and internet subscribers lead to a higher happiness rating. Summary statistics show a large maximum and minimum value, with about a 3,000,000 minimum and a 7,800,000,000,000 total subscriptions. This is an extremely large range for cell phone users.

```
##    Country          Total_subscriptions
##  Length:219         Min.   :3.018e+03
##  Class :character   1st Qu.:8.480e+05
##  Mode  :character   Median :6.653e+06
##                     Mean   :7.416e+07
##                     3rd Qu.:2.278e+07
##                     Max.   :7.806e+09
```

All 3 of the data sets are joined together, and new features are added, including the percentage of internet and phone usages based on population. This is to mitigate the extremely large range between countries with low population and countries with high population. At first glance, some Countries have phone users above 100%. This could be due to all businesses having several new phones, landline phones at residential homes, and countless people who have multiple phones for personal/business use. Furthermore, this results in more phone subscriptions than the total population. The summary statistics show a about 49% of the population have internet use.

```
## # A tibble: 5 x 8
##   Country Total_subscript~ Internet_users Population percent_internet
##   <chr>              <dbl>          <dbl>      <dbl>            <dbl>
## 1 Albania          3497950        2016516    3057220             66.0
## 2 Algeria         49873389       17291463   41657488             41.5
## 3 Angola          13323952        2622403   30355880              8.64
## 4 Argent~         61897379       30786889   44694198             68.9
## 5 Armenia          3488524        1891775    3038217             62.3
```

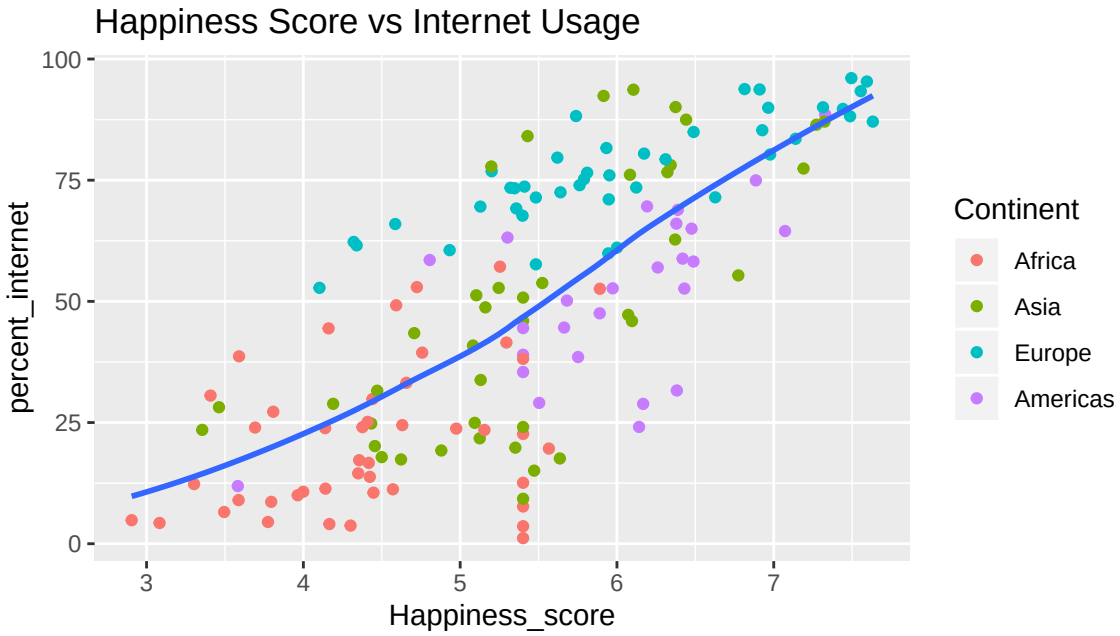
```
## # ... with 3 more variables: percent_phone <dbl>, Happiness_score <dbl>,
## #   Continent <fct>
```

```
##  percent_internet percent_phone     Internet_users
##  Min.   : 1.157   Min.   :  8.475   Min.   :    63084
##  1st Qu.:23.993   1st Qu.: 87.871   1st Qu.:  1297800
##  Median :51.018   Median :113.072   Median :  4368618
##  Mean   :49.095   Mean   :109.409   Mean   : 20939535
##  3rd Qu.:73.631   3rd Qu.:129.563   3rd Qu.: 13929337
##  Max.   :96.055   Max.   :254.256   Max.   :730723960
```

Analyzing internet usage shows that as the internet usage percentage increases, so does happiness score. For example, Europe has the highest average happiness score at 6.06/10 and also the highest population percentage who use the internet at 76.9%. Although the happiness score index is based on several factors, internet usage relates to general quality of life conditions. This includes internet users typically having food, shelter, and a degree of wealth good enough to pay for internet each month. Furthermore, this isn't, including the benefit of access to knowledge and entertainment that comes along with an internet subscription.

```
## # A tibble: 4 x 3
##   Continent average_score percent_internet
##   <fct>             <dbl>            <dbl>
## 1 Africa             4.43             21.5
## 2 Asia               5.46             48.4
## 3 Europe             6.06             76.9
## 4 Americas           5.98             50.9
```

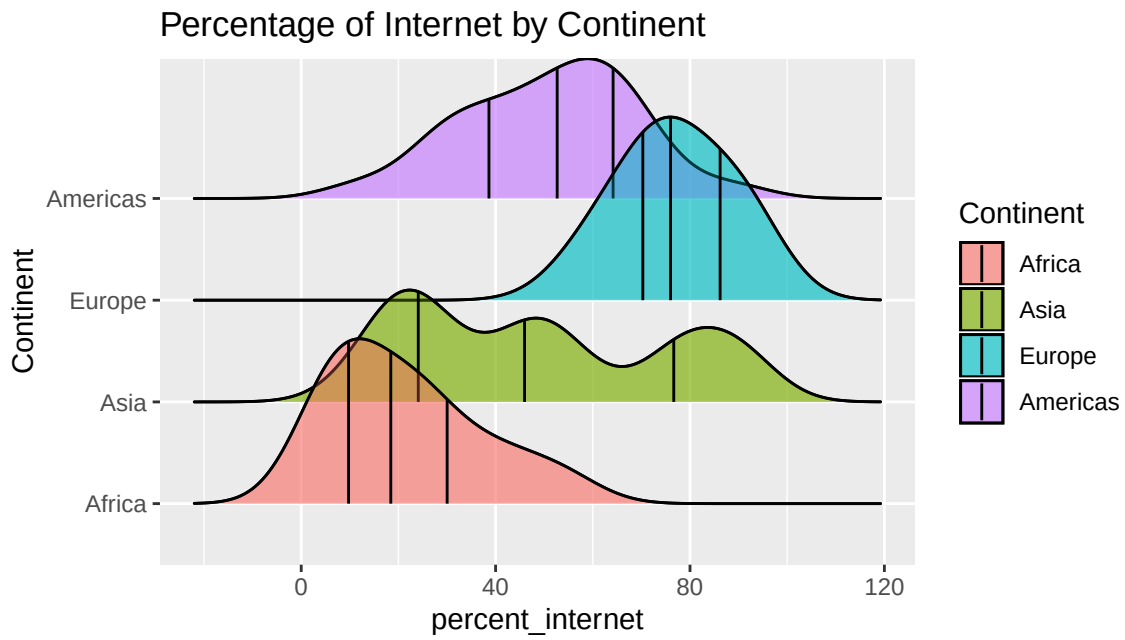
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



This is further reinforced with a Pearson Correlation value of about 0.75. The correlation shows the strong connection between happiness score and interest access. The ridge plot of the data shows Asia seems to have an extremely large variance in terms of internet access. This can be displayed with the highest standard deviation of all of the continents at 26.6. Outlining some countries in Asia have low access to the internet and others extremely high access.

```
## [1] 0.7451462
```

```
## Picking joint bandwidth of 7.72
## Picking joint bandwidth of 7.72
```

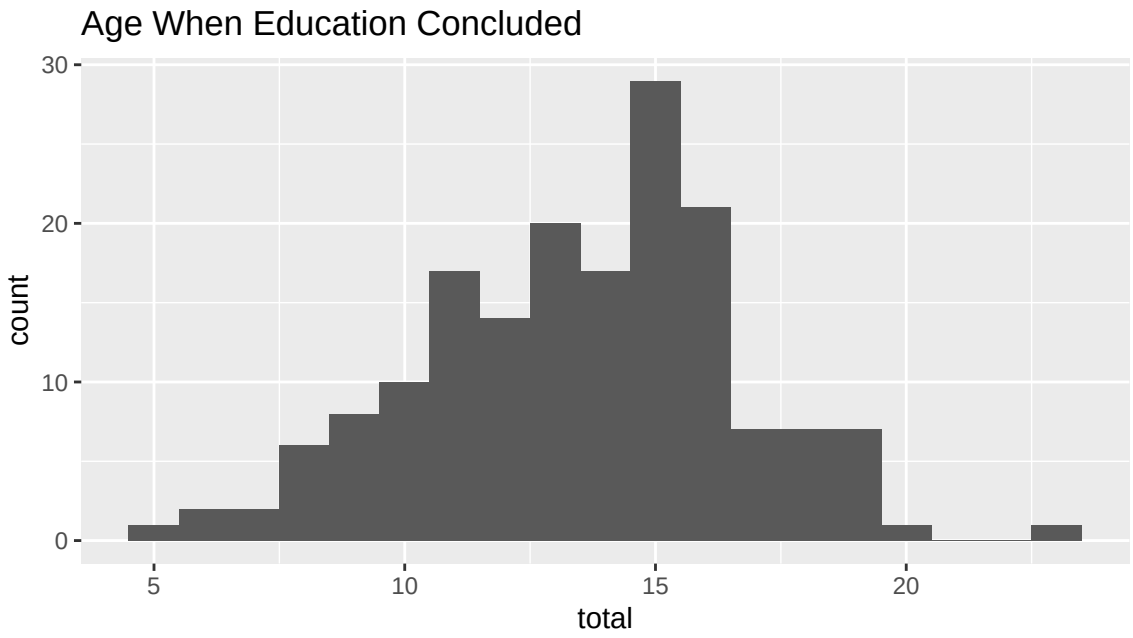



```
## # A tibble: 4 x 2
##   Continent percent_internet_sd
##   <fct>                   <dbl>
## 1 Africa                   15.3
## 2 Asia                     26.6
## 3 Europe                   11.3
## 4 Americas                 17.7
```

Education

Importing the schooling Expectancy data will provide insight into the average age students end their education in each respective Country. The Country is a categorical variable, with total, male and female representing the age of students when their schooling as a continuous variable. Summary statistics show a mean age of about 13.51 years old when students finish their education, which is surprisingly low. I will use this data set combined with the life expectancy data to attempt pattern discovery within life expectancy, education, and, subsequently, happiness score.

Looking at the histogram for the schooling data, with the mean lower than the median, there appears to be a negative skew in the data. Furthermore, a multimodal distribution is present with 3 distinct modes. Zooming in on the outlier on the right side of the data shows Australia with an age of 23 when their education ends, this seems extremely high.

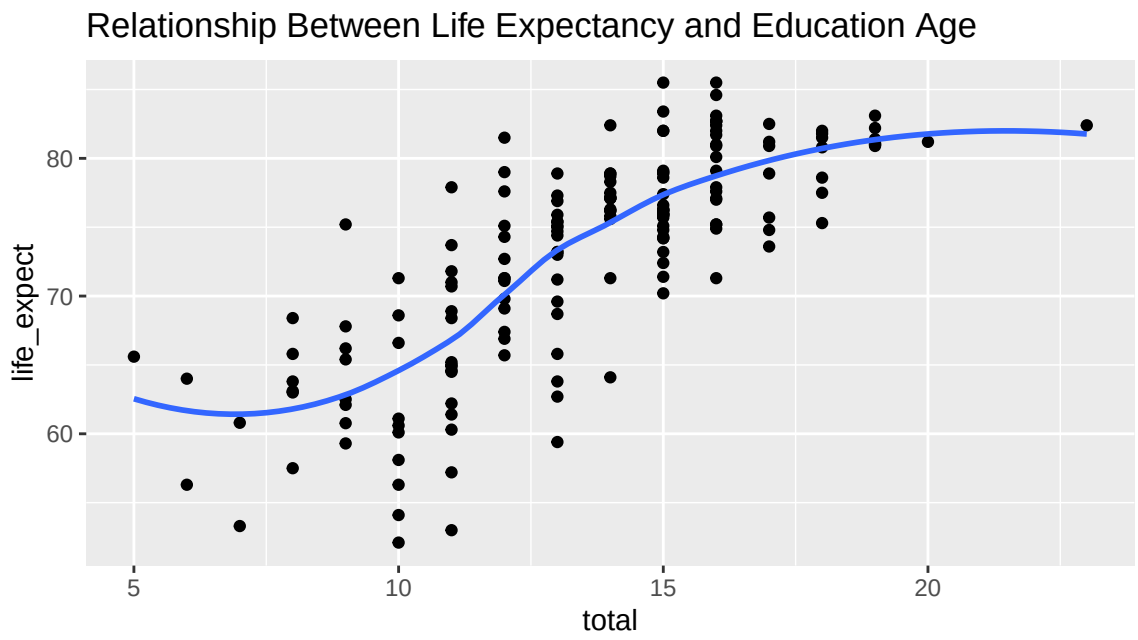



```
##      total           male          female
##  Min.   : 5.00   Min.   : 6.00   Min.   : 5.00
##  1st Qu.:11.00   1st Qu.:11.00   1st Qu.:11.00
##  Median :14.00   Median :14.00   Median :14.00
##  Mean   :13.51   Mean   :13.36   Mean   :13.71
##  3rd Qu.:16.00   3rd Qu.:15.00   3rd Qu.:16.00
##  Max.   :23.00   Max.   :23.00   Max.   :23.00
##                  NA's   :3       NA's   :3
```

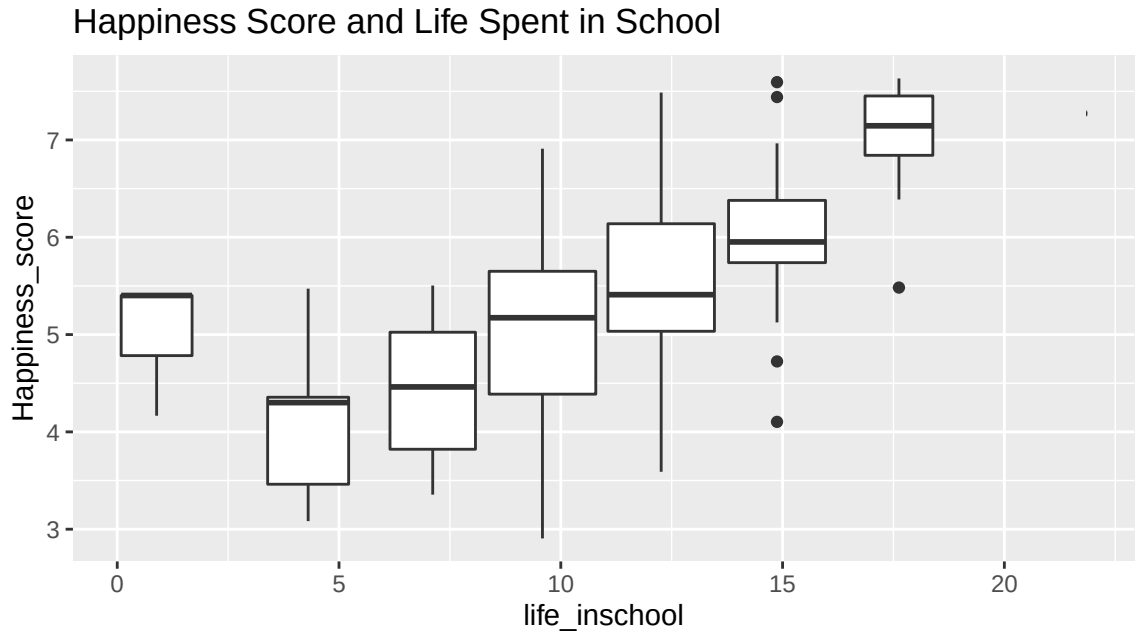
```
## # A tibble: 1 x 4
##   Country   total  male female
##   <chr>     <dbl> <dbl>  <dbl>
## 1 Australia    23    23     23
```

Analyzing life expectancy Country is a qualitative categorical variable, as life expectancy is a quantitative continuous variable representing life expectancy for each country. It appears that the longer the life expectancy increases the education age expectancy also increases.

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

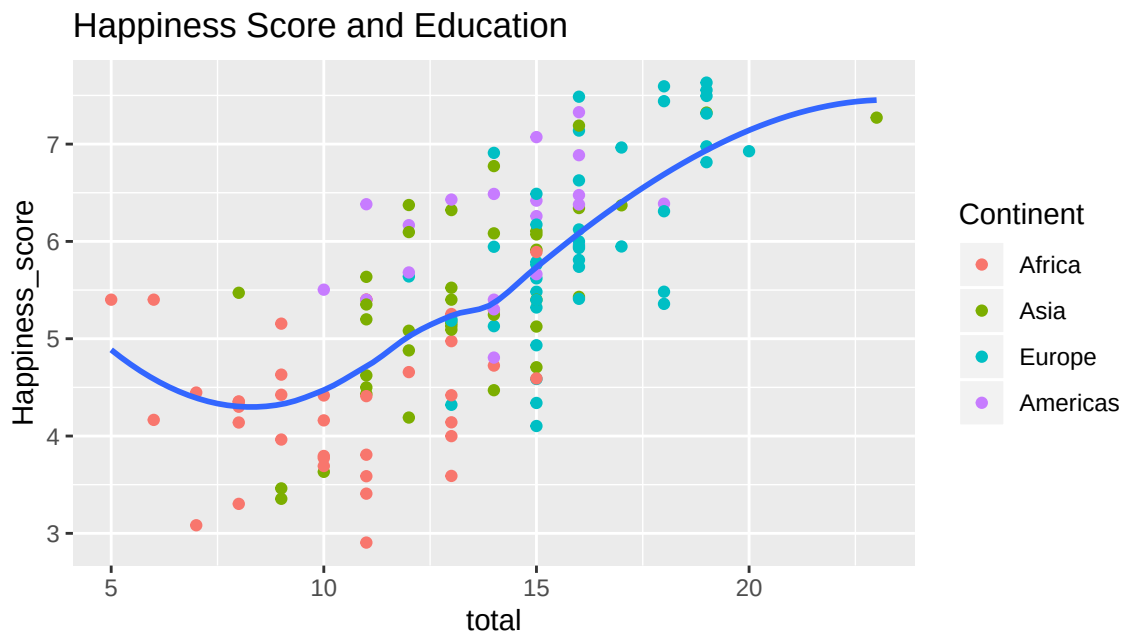


This makes sense, as the ratio between life and education changes, for example, pursuing education until the age of 20 with a life expectancy of 80 is much different than the same situation but with a life expectancy of 45. With a lower life expectancy, potential students might be forced to join the workforce if the head of the house dies at a younger age. This also correlates to happiness scores significantly, as the higher percentage of life spent in school seems to equal higher happiness generally.



Relating back to happiness score shows a higher level of education also correlates to a higher happiness score. A good edcuation typically means a good paying job, allowing for the purchase of needs as well as wants, thus increasing happiness.

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Education and Technology Linear Regression

The linear regression from the education and technology variables greatly reinforces the observations recorded for these sectors. Variables including the percentage of internet users, life expectancy, and ending age of education all are positive influencers on happiness scores. Life expectancy seems to be the positive driving influencer to Happiness score. With each year of life, expectancy comes a 0.052 increase in happiness score. Therefore 10 years of life expectancy increases happiness score by 0.52 which is massive with the main range of 4 to 7 on the happiness score data

```
##
## Call:
## lm(formula = Happiness_score ~ percent_internet + life_expect +
##     total, data = model)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.60080 -0.40809  0.01668  0.54192  1.38983
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.360875   0.830641   0.434 0.664680
## percent_internet 0.014391   0.004318   3.333 0.001121 **
## life_expect      0.052342   0.014309   3.658 0.000368 ***
## total            0.040071   0.033338   1.202 0.231568
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6804 on 130 degrees of freedom
##   (77 observations deleted due to missingness)
## Multiple R-squared:  0.6344, Adjusted R-squared:  0.626
## F-statistic:  75.2 on 3 and 130 DF,  p-value: < 2.2e-16
```

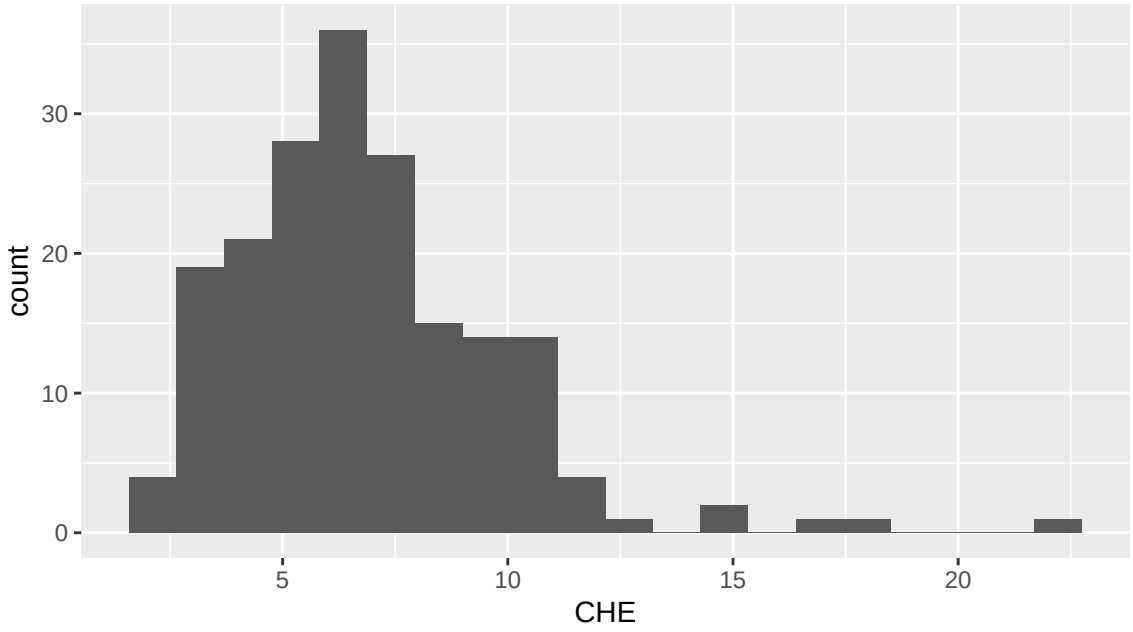
These variables also hold a connection to each other, as each variable is dropped from the regression, the variables remaining further increase the happiness score variable. For example, percent internet usage increase from 0.01 per percent to 0.029. A 10% increase in a countries internet usage leads to a 0.29 point increase in happiness score.

```
##
## Call:
## lm(formula = Happiness_score ~ percent_internet, data = model)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.52914 -0.50895  0.00558  0.58602  1.46685
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       4.00053    0.11843   33.78   <2e-16 ***
## percent_internet  0.02893    0.00210   13.78   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.723 on 152 degrees of freedom
##   (57 observations deleted due to missingness)
## Multiple R-squared:  0.5552, Adjusted R-squared:  0.5523
## F-statistic: 189.8 on 1 and 152 DF,  p-value: < 2.2e-16
```

In conclusion, technology and education play a large role in a countries happiness score. It seems having quality education and access to technology equates to shelter, food, water and entertainment.

Heath

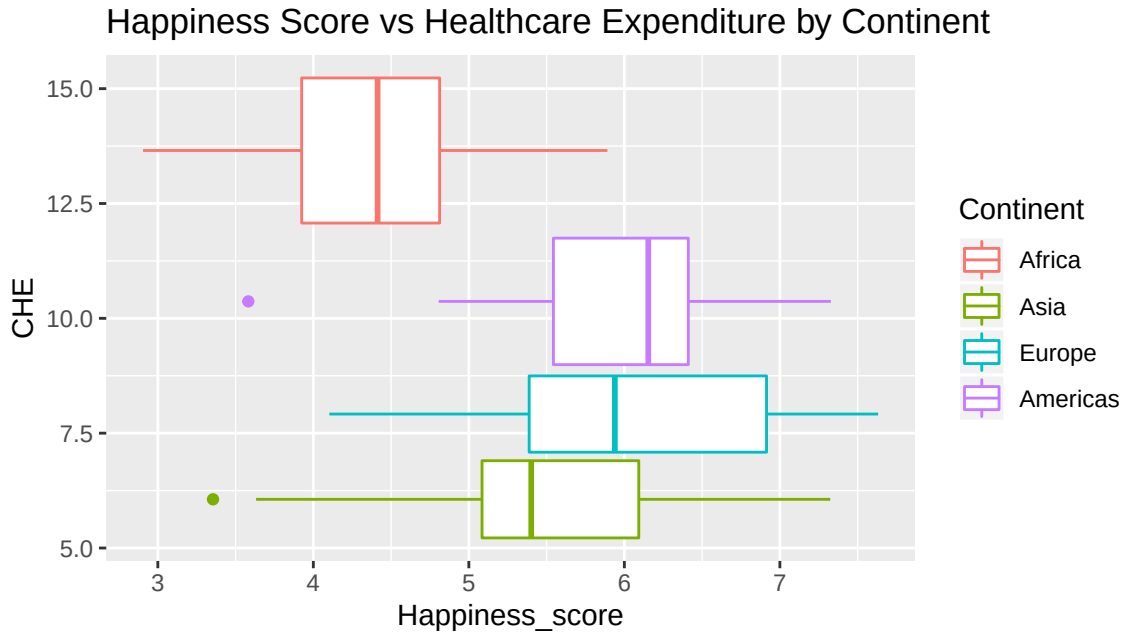
The Health Expenditure data set shows the percentage of GDP that each country spends on health care. It contains the qualitative continuous variable as CHE, representing the percentage that each country spends on healthcare.



```
##    Country                CHE
##  Length:188         Min.   : 2.000
##  Class :character   1st Qu.: 4.800
##  Mode  :character   Median : 6.300
##                     Mean   : 6.816
##                     3rd Qu.: 8.200
##                     Max.   :22.100
```

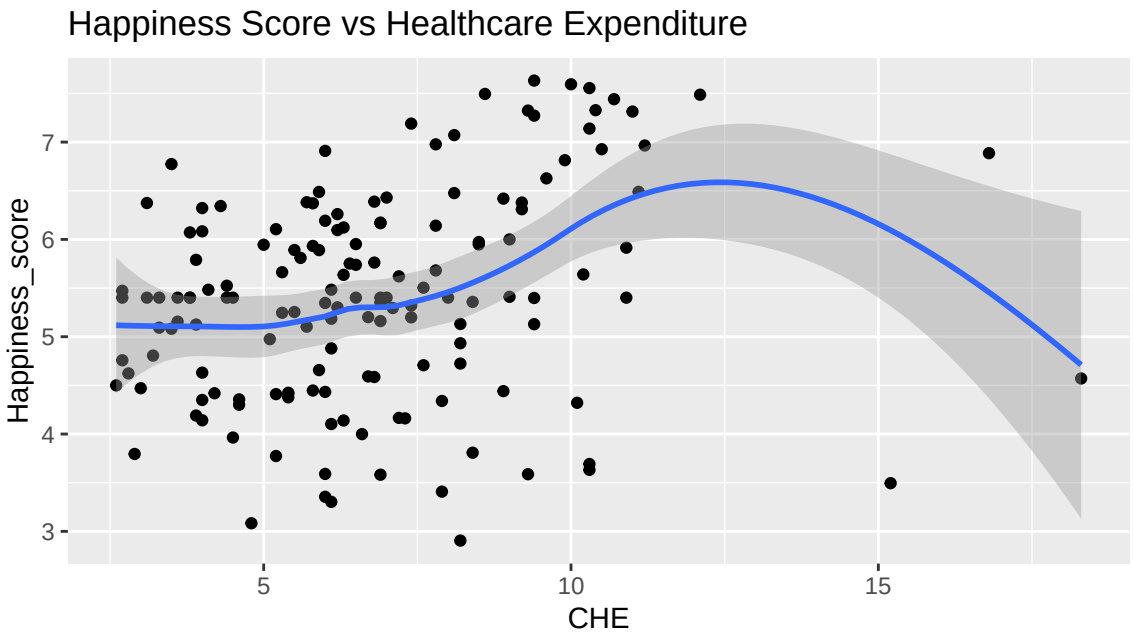
The histogram displayed above shows each countries annual health expenditure (CHE) in relation to their individual GDP totals. On average, countries are spending 6.3% of their GDP on healthcare. The graph was adjusted to filter out an outlier that was recorded as spending over 400% of their GDP on healthcare. It appears most countries spebd 4-8% of their weath on healthcare.

The happiness score data set will be joined to the health expenditure data set to see the relationship between expenditure and Continent.

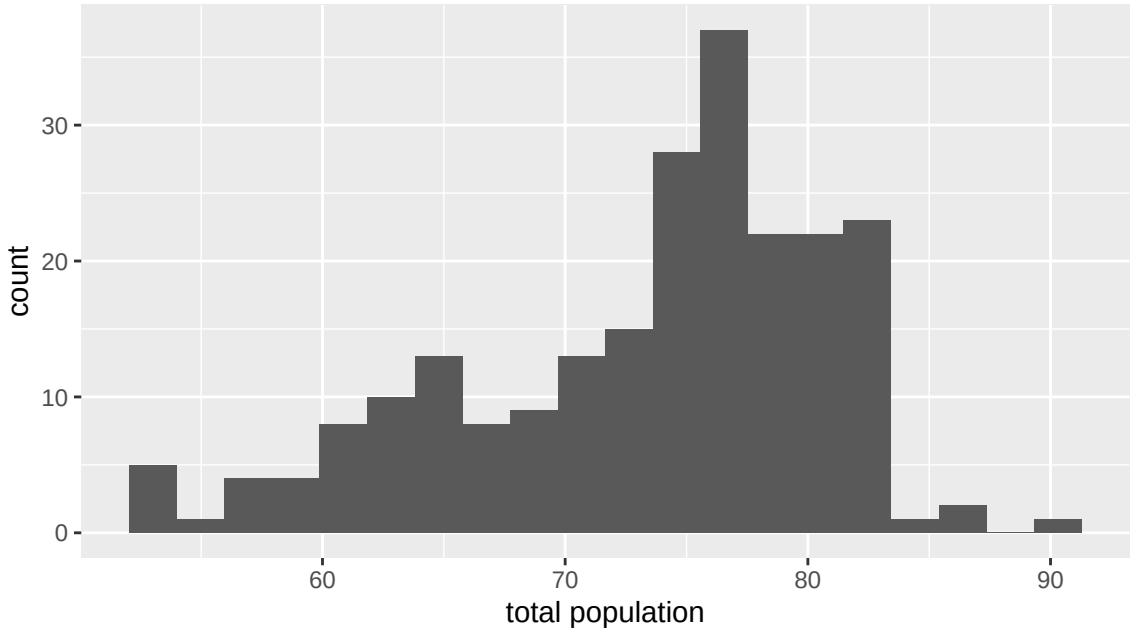


The graph shows a connection between countries happiness scores and their health care expenditure (CHE). In order to group the rather large set of data better, the graph has been adjusted to show continents instead of individual countries. The common return of the graph as seen with Asia, Europe, and the Americas is that the more money being spent on health care, the higher the happiness score. It is clear that Africa spends a higher percent of their GDP on health care, yet their considerably lower on average in the happiness score category. This may be a result of Africa having a lower annual GDP, as even though their CHE percentage may be higher than the others, they're still spending less money and therefore have a low happiness score.

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



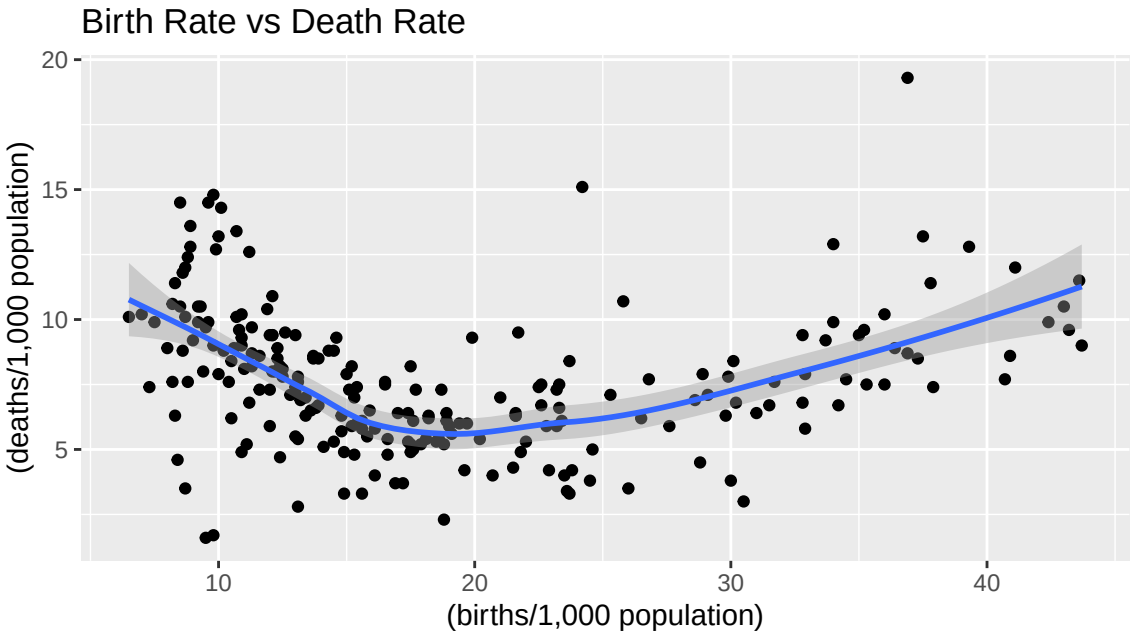
The graph displays the data from each country, as the CHE percentage per country increases, so does their happiness score. There are obvious outliers as there is a decline near the higher end of the CHE scores, this can be a similar case as previously stated with Africa's CHE in relation to their comparably lower GDP. When certain countries are not bringing in a lot of money, the little amounts they may spend on health care contribute to a higher portion of their GDP compared to the countries who have a much larger GDP total.




```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   52.10   68.45   75.20   73.33   78.90   89.40       7
```

The histogram shows the life expectancy of males and females across each country. This data can be used to determine which sex tends to live longer on average, using the 'summary' application its clear that females tend to live longer.

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

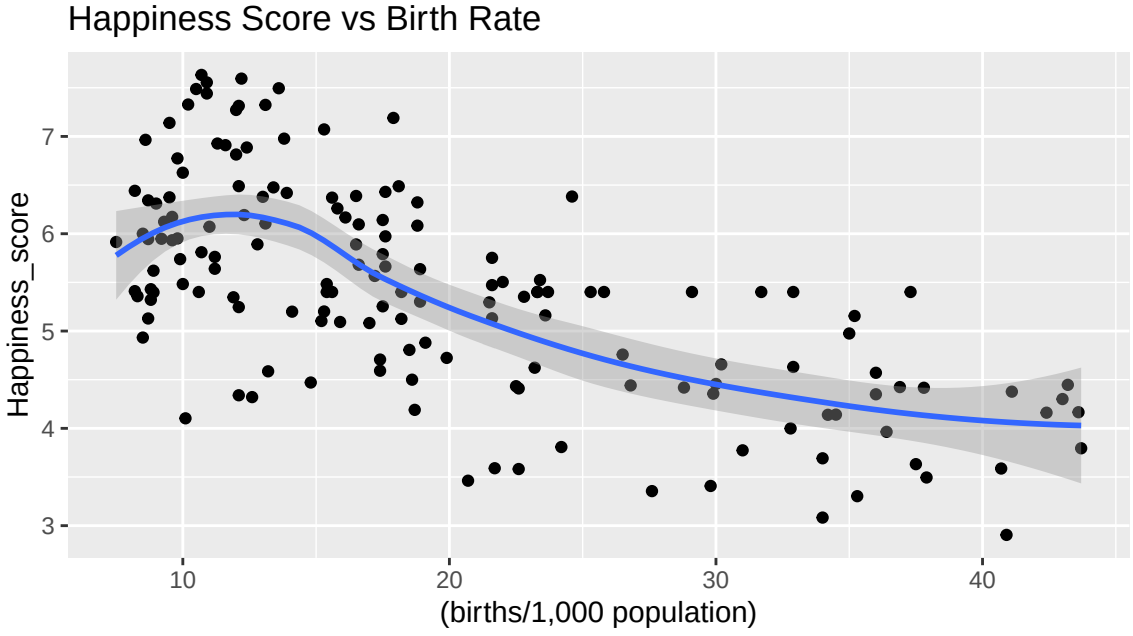


The graph displays the mid-point along the death rate and birth rate datasets. Each set is adjusted to calculate per every 1000 people in order to display a more visible result. The graph shows that even as the birth rates climb from 10 – 40, the death rate remains constant at 10 per every 1000.

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning: Removed 11 rows containing non-finite values (stat_smooth).

## Warning: Removed 11 rows containing missing values (geom_point).
```



The graph displays the connection between the birth rates and happiness score within countries. The data shows that as the birth rates increase, the happiness score decreases. Such results can be attributed to many factors including; the more people within a population requires the healthcare to be spread out more, thus decreasing the potential quality per person. Also, the same concept can be applied to the amount of money that will be required to support larger families as the birthrates increase. If a family has more kids, they are then spreading their resources thin in an attempt to provide for all members.