Country Happiness Exploration

Will Koehrsen

April 8, 2017

Introduction

I decided to explore country happiness rankings from 2016. This dataset is able for download from Kaggle. There are plenty of other interesting datasets available from Kaggle and I would highly recommended checking them out if one is in search of data to visualize.

Variables in Data

Country, Region, Happiness Rank, Happiness Score, Lower Confidence Interval, Upper Confidence Interval,
Economy (GDP per Capita), Family, Health (Life Expectancy), Freedom, Trust (Government Corruption), Generosity, Dystopia Residual

I wanted to explore what makes a country happy. I already had some hypotheses, and I thought this would be a great way to test them. I started off with a simple scatterplot of Happiness Score vs Freedom with the points sized by generosity. I then found the correesponding correlation coefficient between happiness score and freedom. The results show a slight positive correlation but does not imply freedom causes happiness.

I wanted to see which regions had the highest happiness scores on average. Here is a simple bar plot.

```
ggplot(aes(x = Region, y= Happiness.Score),  data = df) + geom_bar(stat= 'summary', fun.y='mean', color='black', fill='orange') +
  theme(axis.text.x = element_text(angle = 60, hjust=1)) +
  labs(x = 'Region', y = 'Mean Happiness') + scale_y_continuous(breaks=seq(0,8,1)) + labs(title='Mean Happiness Score by Region')
```


```
regions <- group_by(df, Region)
summarize(regions, mean(Happiness.Score))
```

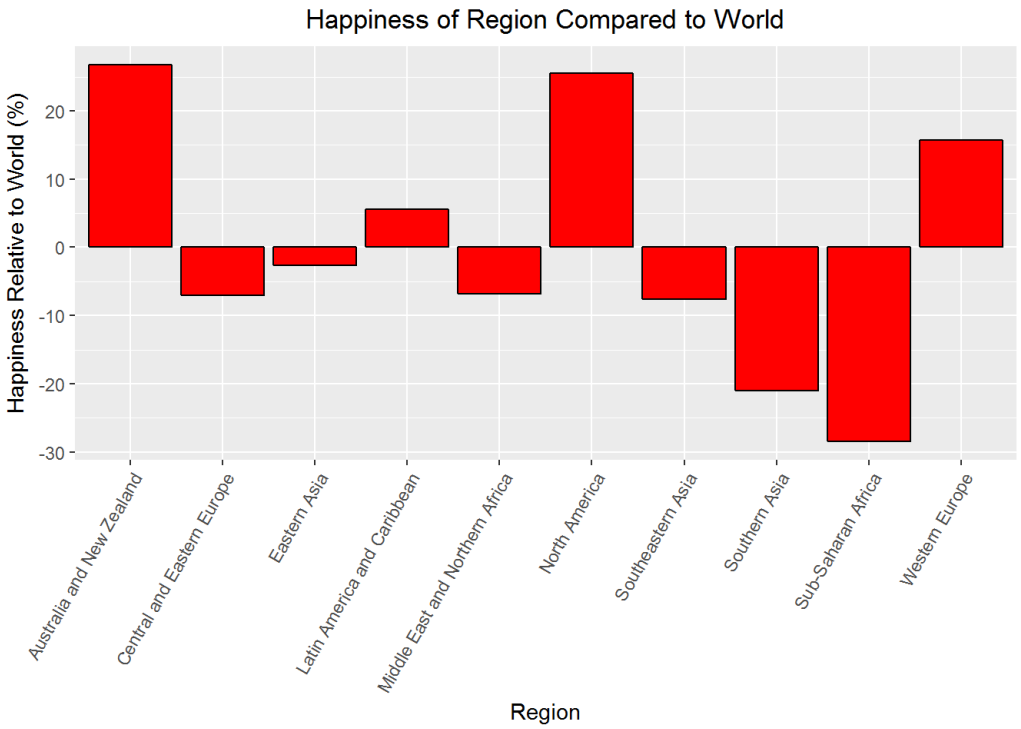
```
## # A tibble: 10 × 2
##                            Region `mean(Happiness.Score)`
##                            <fctr>                   <dbl>
## 1        Australia and New Zealand                7.323500
## 2       Central and Eastern Europe                5.370690
## 3                     Eastern Asia                5.624167
## 4      Latin America and Caribbean                6.101750
## 5  Middle East and Northern Africa                5.386053
## 6                     North America                7.254000
## 7                Southeastern Asia                5.338889
## 8                     Southern Asia                4.563286
## 9                Sub-Saharan Africa                4.136421
## 10                   Western Europe                6.685667
```

It appears as those Australia and New Zealand narrowly edges out North America for greatest happiness. The average among all of the regions is fairly high however. I will now graph the variation from the average happiness. Expressed as (mean happiness score of region - mean happiness score overall) / (mean happiness score overall)

```
regions <- group_by(df, Region)
regions <- summarize(regions,
                     mean_happiness = mean(Happiness.Score),
                     median_happiness = median(Happiness.Score),
                     n = n())

regions <- transform(regions, relative_happiness = 100 * (mean_happiness - mean(mean_happiness)) / mean(mean_h
appiness))

#Another bar plot with the relative happiness
ggplot(aes(x = Region, y = relative_happiness), data = regions) + geom_bar(stat='identity', color= 'black', fi
ll = 'red')  +
  theme(axis.text.x = element_text(angle = 60, hjust=1)) + labs(y='Happiness Relative to World (%)', x = 'Regi
on', title='Happiness of Region Compared to World')
```



Those are intriguing results.

However, one of the limitations of this data is that the sample size in region is relatively small.

```
regions[, c("Region", "n")]
```

```
##                            Region  n
## 1         Australia and New Zealand  2
## 2        Central and Eastern Europe 29
## 3                      Eastern Asia  6
## 4       Latin America and Caribbean 24
## 5  Middle East and Northern Africa 19
## 6                      North America  2
## 7                  Southeastern Asia  9
## 8                      Southern Asia  7
## 9                 Sub-Saharan Africa 38
## 10                    Western Europe 21
```

It is difficult to draw conclusions based on those small sample sizes. Let's return to the full data to examine a few more correlations.

```
# Print out correlations for every pair of variables

library(Hmisc)
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Loading required package: Formula
```

```
##
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:dplyr':
##
##     combine, src, summarize
```

```
## The following objects are masked from 'package:base':
##
##     format.pval, round.POSIXt, trunc.POSIXt, units
```

```
ccs <- as.matrix(df[, c(4, 7:13)])
rcorr(ccs, type='pearson')
```

```
##                               Happiness.Score Economy..GDP.per.Capita.
## Happiness.Score                          1.00                     0.79
## Economy..GDP.per.Capita.                 0.79                     1.00
## Family                                   0.74                     0.67
## Health..Life.Expectancy.                 0.77                     0.84
## Freedom                                  0.57                     0.36
## Trust..Government.Corruption.            0.40                     0.29
## Generosity                               0.16                    -0.03
## Dystopia.Residual                        0.54                     0.07
##                               Family Health..Life.Expectancy. Freedom
## Happiness.Score                 0.74                    0.77    0.57
## Economy..GDP.per.Capita.        0.67                    0.84    0.36
## Family                          1.00                    0.59    0.45
## Health..Life.Expectancy.        0.59                    1.00    0.34
## Freedom                         0.45                    0.34    1.00
## Trust..Government.Corruption.   0.21                    0.25    0.50
## Generosity                      0.09                    0.08    0.36
## Dystopia.Residual               0.12                    0.10    0.09
##                               Trust..Government.Corruption. Generosity
## Happiness.Score                                        0.40       0.16
## Economy..GDP.per.Capita.                               0.29      -0.03
## Family                                                 0.21       0.09
## Health..Life.Expectancy.                               0.25       0.08
## Freedom                                                0.50       0.36
## Trust..Government.Corruption.                          1.00       0.31
## Generosity                                             0.31       1.00
## Dystopia.Residual                                      0.00      -0.13
##                               Dystopia.Residual
## Happiness.Score                            0.54
## Economy..GDP.per.Capita.                   0.07
## Family                                     0.12
## Health..Life.Expectancy.                   0.10
## Freedom                                    0.09
## Trust..Government.Corruption.              0.00
## Generosity                                -0.13
## Dystopia.Residual                          1.00
##
## n= 157
##
##
## P
##                               Happiness.Score Economy..GDP.per.Capita.
## Happiness.Score                               0.0000
## Economy..GDP.per.Capita.      0.0000
## Family                        0.0000          0.0000
## Health..Life.Expectancy.      0.0000          0.0000
## Freedom                       0.0000          0.0000
## Trust..Government.Corruption. 0.0000          0.0002
## Generosity                    0.0498          0.7509
## Dystopia.Residual             0.0000          0.3931
##                               Family Health..Life.Expectancy. Freedom
## Happiness.Score               0.0000 0.0000                  0.0000
## Economy..GDP.per.Capita.      0.0000 0.0000                  0.0000
## Family                               0.0000                  0.0000
## Health..Life.Expectancy.      0.0000                         0.0000
## Freedom                       0.0000 0.0000
```

```
## Trust..Government.Corruption. 0.0072 0.0016                   0.0000
## Generosity                    0.2643 0.3442                   0.0000
## Dystopia.Residual             0.1355 0.2088                   0.2537
##                               Trust..Government.Corruption. Generosity
## Happiness.Score               0.0000                        0.0498
## Economy..GDP.per.Capita.      0.0002                        0.7509
## Family                        0.0072                        0.2643
## Health..Life.Expectancy.      0.0016                        0.3442
## Freedom                       0.0000                        0.0000
## Trust..Government.Corruption.                               0.0000
## Generosity                    0.0000
## Dystopia.Residual             0.9712                        0.0968
##                               Dystopia.Residual
## Happiness.Score               0.0000
## Economy..GDP.per.Capita.      0.3931
## Family                        0.1355
## Health..Life.Expectancy.      0.2088
## Freedom                       0.2537
## Trust..Government.Corruption. 0.9712
## Generosity                    0.0968
## Dystopia.Residual
```

Well that is certainly interesting. Happiness is most strongly correlated with GDP per capita and least strongly correlated with generosity. It is also strongly correlated with life expectancy and family. I wonder what the top ten happiest countries and the bottom 10 sadest countries are.

```
df <- arrange(df, Happiness.Score)
head(df, 10)
```

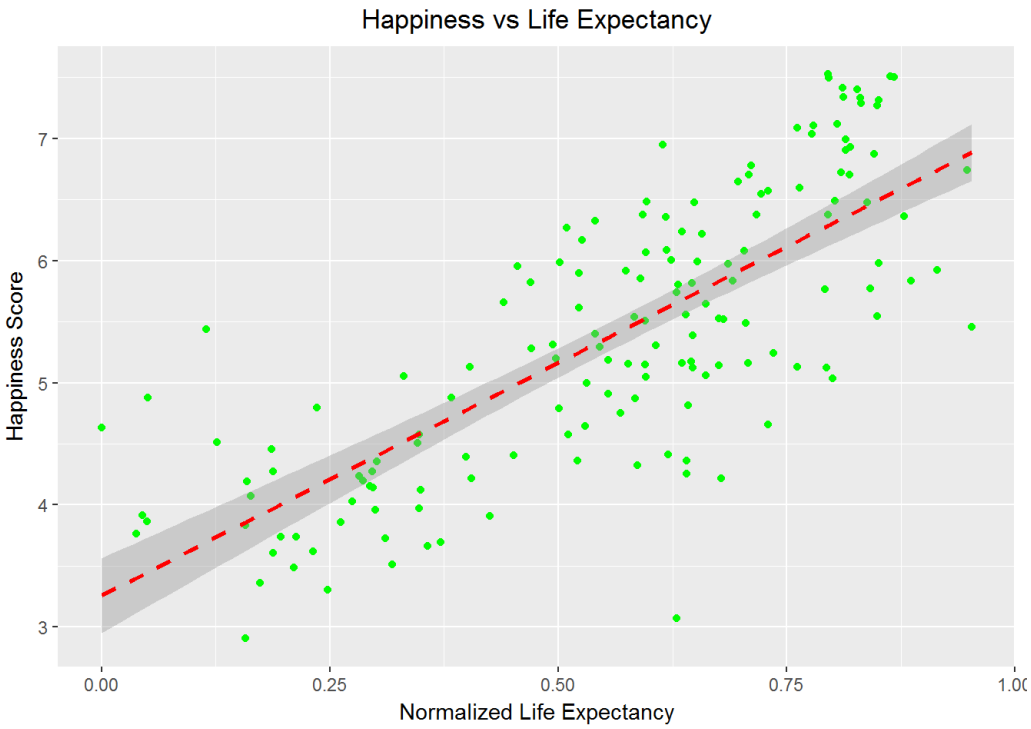
```
##        Country                          Region Happiness.Rank
## 1      Burundi              Sub-Saharan Africa            157
## 2        Syria Middle East and Northern Africa            156
## 3         Togo              Sub-Saharan Africa            155
## 4  Afghanistan                   Southern Asia            154
## 5        Benin              Sub-Saharan Africa            153
## 6       Rwanda              Sub-Saharan Africa            152
## 7       Guinea              Sub-Saharan Africa            151
## 8      Liberia              Sub-Saharan Africa            150
## 9     Tanzania              Sub-Saharan Africa            149
## 10  Madagascar              Sub-Saharan Africa            148
##    Happiness.Score Lower.Confidence.Interval Upper.Confidence.Interval
## 1            2.905                    2.732                     3.078
## 2            3.069                    2.936                     3.202
## 3            3.303                    3.192                     3.414
## 4            3.360                    3.288                     3.432
## 5            3.484                    3.404                     3.564
## 6            3.515                    3.444                     3.586
## 7            3.607                    3.533                     3.681
## 8            3.622                    3.463                     3.781
## 9            3.666                    3.561                     3.771
## 10           3.695                    3.621                     3.769
##    Economy..GDP.per.Capita.  Family Health..Life.Expectancy. Freedom
## 1                   0.06831 0.23442                  0.15747 0.04320
## 2                   0.74719 0.14866                  0.62994 0.06912
## 3                   0.28123 0.00000                  0.24811 0.34678
## 4                   0.38227 0.11037                  0.17344 0.16430
## 5                   0.39499 0.10419                  0.21028 0.39747
## 6                   0.32846 0.61586                  0.31865 0.54320
## 7                   0.22415 0.31090                  0.18829 0.30953
## 8                   0.10706 0.50353                  0.23165 0.25748
## 9                   0.47155 0.77623                  0.35700 0.31760
## 10                  0.27954 0.46115                  0.37109 0.13684
##    Trust..Government.Corruption. Generosity Dystopia.Residual
## 1                        0.09419    0.20290           2.10404
## 2                        0.17233    0.48397           0.81789
## 3                        0.11587    0.17517           2.13540
## 4                        0.07112    0.31268           2.14558
## 5                        0.06681    0.20180           2.10812
## 6                        0.50521    0.23552           0.96819
## 7                        0.11920    0.29914           2.15604
## 8                        0.04852    0.24063           2.23284
## 9                        0.05099    0.31472           1.37769
## 10                       0.07506    0.22040           2.15075
```

```
tail(df, 10)
```

```
##         Country                    Region Happiness.Rank Happiness.Score
## 148      Sweden            Western Europe             10           7.291
## 149   Australia Australia and New Zealand              9           7.313
## 150 New Zealand Australia and New Zealand              8           7.334
## 151 Netherlands            Western Europe              7           7.339
## 152      Canada             North America              6           7.404
## 153     Finland            Western Europe              5           7.413
## 154      Norway            Western Europe              4           7.498
## 155     Iceland            Western Europe              3           7.501
## 156 Switzerland            Western Europe              2           7.509
## 157     Denmark            Western Europe              1           7.526
##     Lower.Confidence.Interval Upper.Confidence.Interval
## 148                     7.227                     7.355
## 149                     7.241                     7.385
## 150                     7.264                     7.404
## 151                     7.284                     7.394
## 152                     7.335                     7.473
## 153                     7.351                     7.475
## 154                     7.421                     7.575
## 155                     7.333                     7.669
## 156                     7.428                     7.590
## 157                     7.460                     7.592
##     Economy..GDP.per.Capita.  Family Health..Life.Expectancy. Freedom
## 148                  1.45181 1.08764                  0.83121 0.58218
## 149                  1.44443 1.10476                  0.85120 0.56837
## 150                  1.36066 1.17278                  0.83096 0.58147
## 151                  1.46468 1.02912                  0.81231 0.55211
## 152                  1.44015 1.09610                  0.82760 0.57370
## 153                  1.40598 1.13464                  0.81091 0.57104
## 154                  1.57744 1.12690                  0.79579 0.59609
## 155                  1.42666 1.18326                  0.86733 0.56624
## 156                  1.52733 1.14524                  0.86303 0.58557
## 157                  1.44178 1.16374                  0.79504 0.57941
##     Trust..Government.Corruption. Generosity Dystopia.Residual
## 148                       0.40867    0.38254           2.54734
## 149                       0.32331    0.47407           2.54650
## 150                       0.41904    0.49401           2.47553
## 151                       0.29927    0.47416           2.70749
## 152                       0.31329    0.44834           2.70485
## 153                       0.41004    0.25492           2.82596
## 154                       0.35776    0.37895           2.66465
## 155                       0.14975    0.47678           2.83137
## 156                       0.41203    0.28083           2.69463
## 157                       0.44453    0.36171           2.73939
```

It appears that the Scandanavian countries tend to be the happiest in the world while the sub-saharan African countries have the lowest happiness. I would like to see how this changes over the years, but the only other data set was from 2015 so there would likely not be noticeable trends in the data. I will finish up by graphing a few of the strongest correlations.

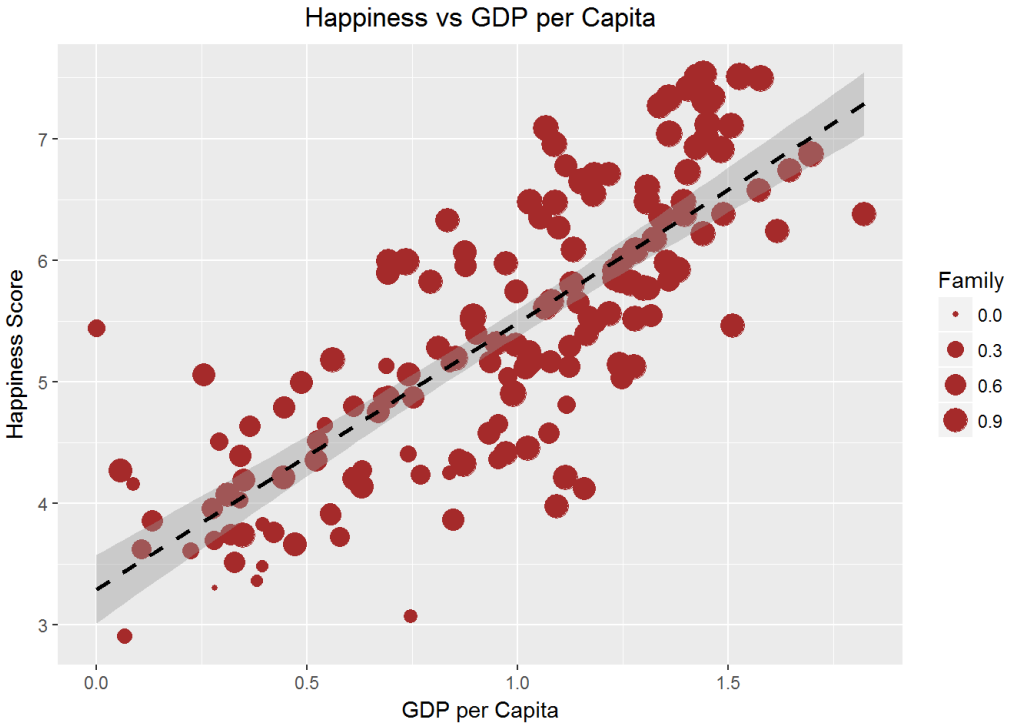
```
# Happiness vs life expectancy
ggplot(aes(x = Health..Life.Expectancy., y = Happiness.Score), data = df) + geom_point(color='green') +
  labs(x = 'Normalized Life Expectancy', y= 'Happiness Score', title='Happiness vs Life Expectancy') + geom_sm
ooth(method='lm', color= 'red', linetype=2)
```



A pretty strong correlation there.

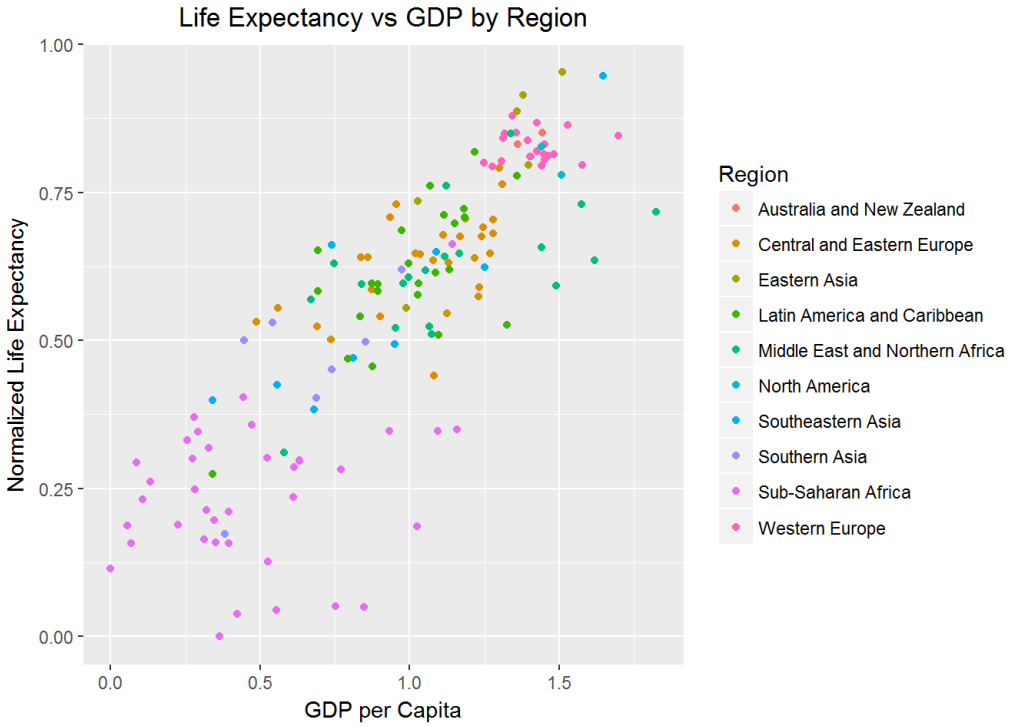
(The second highest). Now for the highest correlation. I will size the bubble by family

```
ggplot(aes(x = Economy..GDP.per.Capita., y = Happiness.Score), data = df) + geom_point(aes(size = Family), col
or = 'brown') +
  labs(x = 'GDP per Capita', y= 'Happiness Score', title='Happiness vs GDP per Capita') + geom_smooth(method =
'lm', color='black', linetype = 2)
```



I want to do one final plot unrelated to Happiness Score. I'll look at the strong relationship between GDP and life expectancy. The points can be colored by region.

```
ggplot(aes(x = Economy..GDP.per.Capita. , y = Health..Life.Expectancy.), data =df) + geom_point(aes(color=Regi
on)) +
  labs(x='GDP per Capita', y= 'Normalized Life Expectancy', title = 'Life Expectancy vs GDP by Region')
```

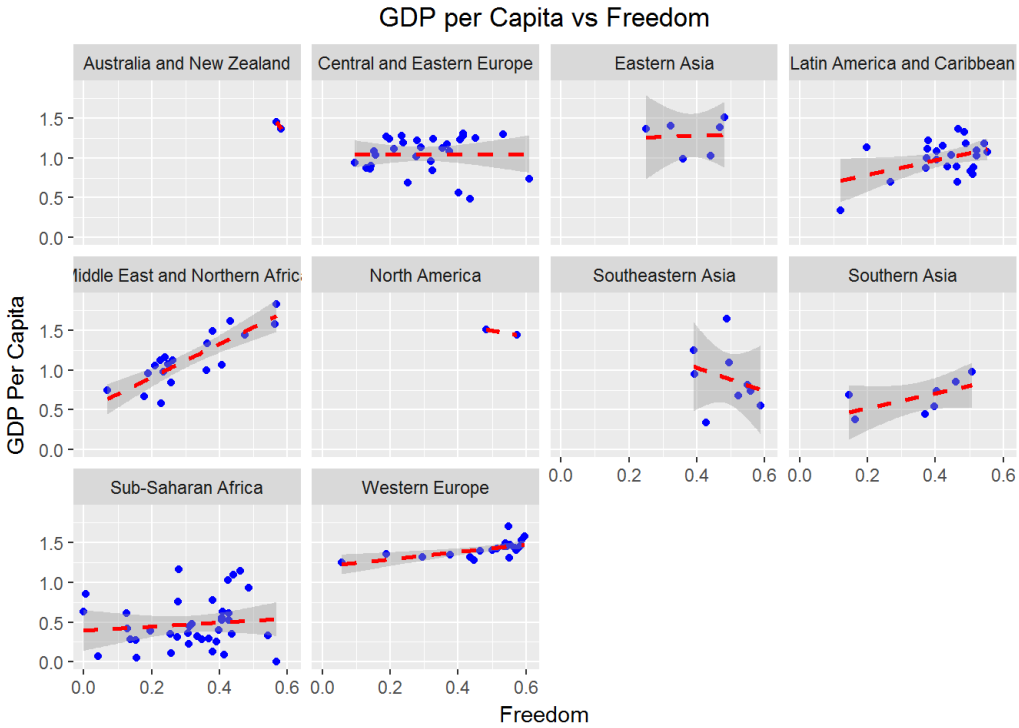


That is fairly revealing. I think I want to construct another plot, this time faceted by each region. I think I will go with GDP vs Freedom and also Dystopia vs Trust in Government

```
ggplot(aes(x = Freedom, y = Economy..GDP.per.Capita.), data = df) + facet_wrap(~Region) + geom_point(color='blue') +
  labs(x='Freedom', y='GDP Per Capita', title = 'GDP per Capita vs Freedom') + geom_smooth(method='lm', color='red', linetype=2)
```

```
## Warning in qt((1 - level)/2, df): NaNs produced

## Warning in qt((1 - level)/2, df): NaNs produced
```




```
with(df, cor.test(Freedom,Economy..GDP.per.Capita. ))
```

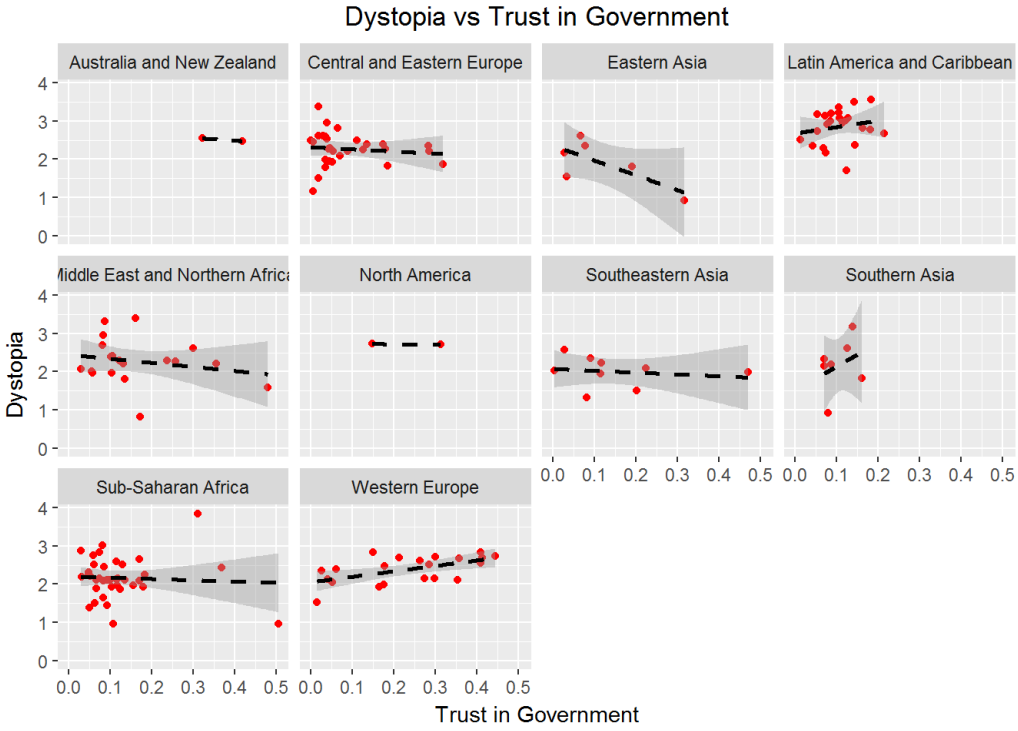
```
##
##  Pearson's product-moment correlation
##
## data:  Freedom and Economy..GDP.per.Capita.
## t = 4.8391, df = 155, p-value = 3.124e-06
```

```
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2180166 0.4910549
## sample estimates:
##       cor
## 0.3622828
```

```
ggplot(aes(x = Trust..Government.Corruption., y = Dystopia.Residual), data = df) + facet_wrap(~Region) +  geom
_point(color= 'red') +
  labs(x = 'Trust in Government' , y = 'Dystopia', title= 'Dystopia vs Trust in Government') + geom_smooth(met
hod='lm', color='black', linetype = 2)
```

```
## Warning in qt((1 - level)/2, df): NaNs produced

## Warning in qt((1 - level)/2, df): NaNs produced
```



It appears that there is no relationship between trust in government and dystopia, as indicated by the correlation coefficient. Moreover, there is a slight positive correlation between freedom and GDP per capita.

Conclusions

The most strongly correlated factor with happiness is the economy as measured in GDP followed by health as measured in life expectancy. The most correlated variables at all are economy and life expectancy which is not unexpected. I would think that a more robust economy would lead to better health incomes and unsurprisingly, happier citizens. What was surprising, is that happiness is not very strongly related to generosity which is what I would have expected. Again, I need to do some more research into the metrics used in this data as I do not understand all of the values and what is represented. The happiest regions are Australia and New Zealand and North America while the most unhappy regions are sub-saharan Africa and southern asia. The happiest countries in the world are Denmark, Switzerland, Iceland, and Norway. I am not surprised to see the Scandanavian countries, with their impressive public support systems, top the list. There are still many relationships and connections that can be drawn from this data. Perhaps it could be useful in terms of directing aid money to where it will be most effective. Knowing the underlying relationships between factors could inform better decisions when it comes to helping others. At the end of the day, data analysis is about finding the relationships and then using them to create better systems. Here is an additional research for those who are still curious (World Happiness Report)[http://worldhappiness.report/ed/2017/]