Happiness

Noah Lee

2023-06-25

World Happiness Report data for 2015

Dataset source: https://www.kaggle.com/datasets/unsdsn/world-happiness Dataset includes information gathered for the World Happiness Report survey for the year 2015.

Load the data

```
whr2015 <- read.csv("2015.csv")
```

Prepare the data 1. Rename the columns by replacing periods(.) with an underscore(_) to be able to work with sqldf functions. 2. Remove the varaible "Happiness_Rank" as the dataframe's index can be used as the same.

```
#Rename the columns
colnames(whr2015)[3] = "Happiness_Rank"
colnames(whr2015)[4] = "Happiness_Score"
colnames(whr2015)[5] = "Standard_Error"
colnames(whr2015)[6] = "GDP"
colnames(whr2015)[8] = "Life_Exp"
colnames(whr2015)[10] = "Trust"
colnames(whr2015)[12] = "Dystopia_Res"
#Remove Happiness_Rank as it is not needed
whr2015 <- whr2015[,-3]

head(whr2015) #a preview of the data
```

```
##         Country         Region Happiness_Score Standard_Error     GDP  Family
## 1 Switzerland Western Europe           7.587         0.03411 1.39651 1.34951
## 2     Iceland Western Europe           7.561         0.04884 1.30232 1.40223
## 3     Denmark Western Europe           7.527         0.03328 1.32548 1.36058
## 4      Norway Western Europe           7.522         0.03880 1.45900 1.33095
## 5      Canada  North America           7.427         0.03553 1.32629 1.32261
## 6     Finland Western Europe           7.406         0.03140 1.29025 1.31826
##   Life_Exp Freedom   Trust Generosity Dystopia_Res
## 1  0.94143 0.66557 0.41978    0.29678      2.51738
## 2  0.94784 0.62877 0.14145    0.43630      2.70201
## 3  0.87464 0.64938 0.48357    0.34139      2.49204
## 4  0.88521 0.66973 0.36503    0.34699      2.46531
## 5  0.90563 0.63297 0.32957    0.45811      2.45176
## 6  0.88911 0.64169 0.41372    0.23351      2.61955
```

Overview of the data

A look into the top and bottom 10 countries, along with where the United States ranks.

```
#Top and bottom 10 countries and their happiness scores
head(whr2015, n=10)
```

```
##          Country                       Region Happiness_Score Standard_Error     GDP
## 1    Switzerland              Western Europe           7.587        0.03411 1.39651
## 2        Iceland              Western Europe           7.561        0.04884 1.30232
## 3        Denmark              Western Europe           7.527        0.03328 1.32548
## 4         Norway              Western Europe           7.522        0.03880 1.45900
## 5         Canada               North America           7.427        0.03553 1.32629
## 6        Finland              Western Europe           7.406        0.03140 1.29025
## 7    Netherlands              Western Europe           7.378        0.02799 1.32944
## 8         Sweden              Western Europe           7.364        0.03157 1.33171
## 9    New Zealand Australia and New Zealand           7.286        0.03371 1.25018
## 10     Australia Australia and New Zealand           7.284        0.04083 1.33358
##       Family Life_Exp Freedom   Trust Generosity Dystopia_Res
## 1    1.34951  0.94143 0.66557 0.41978    0.29678      2.51738
## 2    1.40223  0.94784 0.62877 0.14145    0.43630      2.70201
## 3    1.36058  0.87464 0.64938 0.48357    0.34139      2.49204
## 4    1.33095  0.88521 0.66973 0.36503    0.34699      2.46531
## 5    1.32261  0.90563 0.63297 0.32957    0.45811      2.45176
## 6    1.31826  0.88911 0.64169 0.41372    0.23351      2.61955
## 7    1.28017  0.89284 0.61576 0.31814    0.47610      2.46570
## 8    1.28907  0.91087 0.65980 0.43844    0.36262      2.37119
## 9    1.31967  0.90837 0.63938 0.42922    0.47501      2.26425
## 10   1.30923  0.93156 0.65124 0.35637    0.43562      2.26646
```

```
tail(whr2015, n=10)
```

```
##           Country                              Region Happiness_Score Standard_Error
## 149          Chad               Sub-Saharan Africa           3.667        0.03830
## 150        Guinea               Sub-Saharan Africa           3.656        0.03590
## 151   Ivory Coast               Sub-Saharan Africa           3.655        0.05141
## 152  Burkina Faso               Sub-Saharan Africa           3.587        0.04324
## 153   Afghanistan                     Southern Asia           3.575        0.03084
## 154        Rwanda               Sub-Saharan Africa           3.465        0.03464
## 155         Benin               Sub-Saharan Africa           3.340        0.03656
## 156         Syria Middle East and Northern Africa           3.006        0.05015
## 157       Burundi               Sub-Saharan Africa           2.905        0.08658
## 158          Togo               Sub-Saharan Africa           2.839        0.06727
##          GDP  Family Life_Exp Freedom   Trust Generosity Dystopia_Res
## 149  0.34193 0.76062  0.15010 0.23501 0.05269    0.18386      1.94296
## 150  0.17417 0.46475  0.24009 0.37725 0.12139    0.28657      1.99172
## 151  0.46534 0.77115  0.15185 0.46866 0.17922    0.20165      1.41723
## 152  0.25812 0.85188  0.27125 0.39493 0.12832    0.21747      1.46494
## 153  0.31982 0.30285  0.30335 0.23414 0.09719    0.36510      1.95210
## 154  0.22208 0.77370  0.42864 0.59201 0.55191    0.22628      0.67042
## 155  0.28665 0.35386  0.31910 0.48450 0.08010    0.18260      1.63328
## 156  0.66320 0.47489  0.72193 0.15684 0.18906    0.47179      0.32858
## 157  0.01530 0.41587  0.22396 0.11850 0.10062    0.19727      1.83302
## 158  0.20868 0.13995  0.28443 0.36453 0.10731    0.16681      1.56726
```

```
#Where the United States stands
whr2015[whr2015$Country == 'United States',]
```

```
##            Country        Region Happiness_Score Standard_Error     GDP  Family
## 15 United States North America           7.119         0.03839 1.39451 1.24711
##    Life_Exp Freedom  Trust Generosity Dystopia_Res
## 15  0.86179 0.54604 0.1589    0.40105      2.51011
```
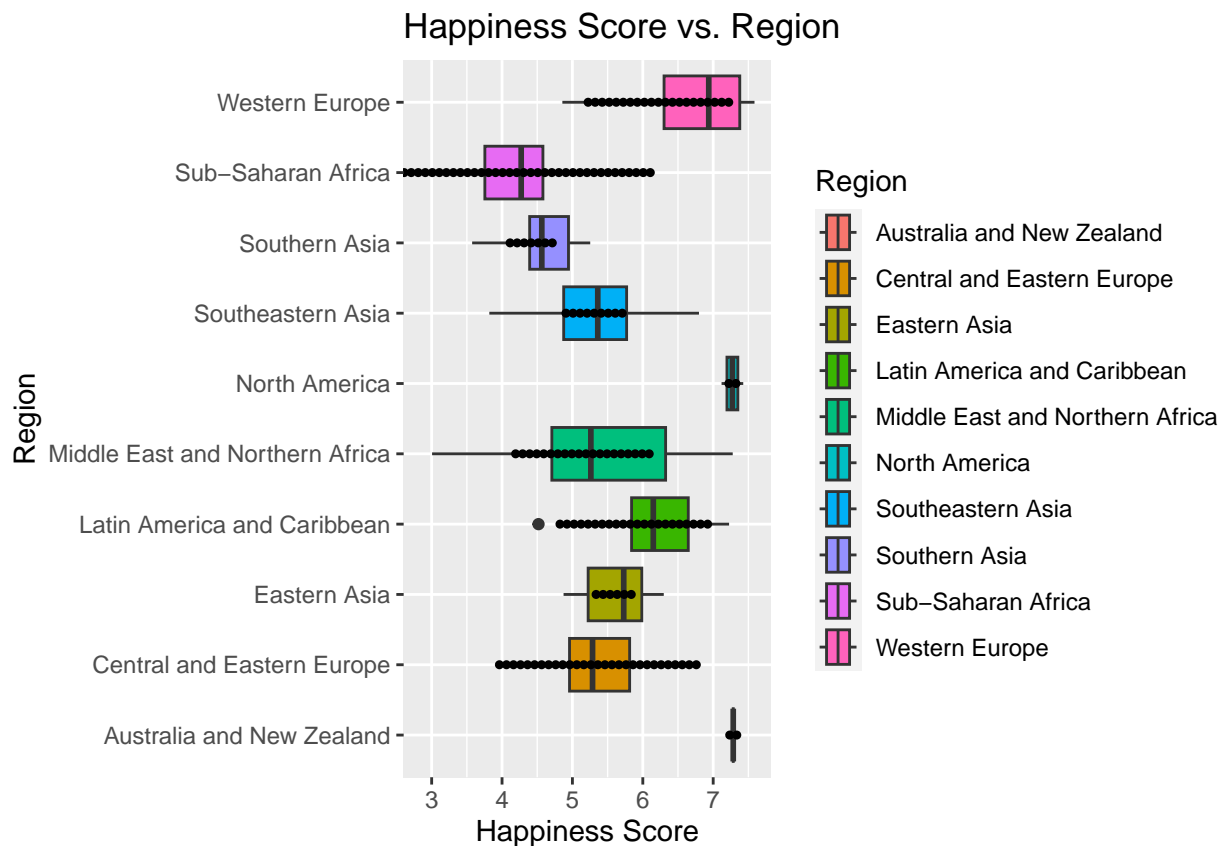
Above are the countries with the top 10 highest and lowest happiness scores.. The top 10 is mostly made up of Western European countries, while 8 of the lowest 10 rankings nations are from Sub-Saharan Africa. The USA is ranked 15/158 (15th out of 158).

Boxplot (box and whisker) of the Happiness Scores.

Individual scores per country per region are portrayed as black dots to show density/skewness.

```
ggplot(data=whr2015, aes(x=Happiness_Score, y=Region, fill=Region)) +
  geom_boxplot() +
  geom_dotplot(binaxis='y', stackdir='center', dotsize=0.5, fill='black', binwidth = 0.2) +
  labs(x='Happiness Score', title='Happiness Score vs. Region')
```

What is the average happiness score per region?

```
region_average2015 <- sqldf("SELECT Region, AVG(Happiness_Score) FROM whr2015
                            GROUP BY Region
                            ORDER BY AVG(Happiness_Score) desc")
region_average2015
```

```
##                               Region AVG(Happiness_Score)
## 1          Australia and New Zealand             7.285000
## 2                      North America             7.273000
## 3                     Western Europe             6.689619
## 4          Latin America and Caribbean            6.144682
## 5                       Eastern Asia             5.626167
## 6   Middle East and Northern Africa             5.406900
## 7          Central and Eastern Europe            5.332931
## 8                  Southeastern Asia             5.317444
## 9                      Southern Asia             4.580857
## 10                 Sub-Saharan Africa             4.202800
```

Correlation matrix

```
whr2015.cor <- correlate(whr2015[3:11])
```

```
## Correlation computed with
## * Method: 'pearson'
## * Missing treated using: 'pairwise.complete.obs'
```

```
whr2015.cor
```

```
## # A tibble: 9 x 10
##    term   Happiness_Score Standard_Error     GDP  Family Life_Exp Freedom    Trust
##    <chr>            <dbl>          <dbl>   <dbl>   <dbl>    <dbl>   <dbl>    <dbl>
## 1 Happi~              NA         -0.177   0.781   0.741    0.724   0.568    0.395
## 2 Stand~          -0.177             NA  -0.218  -0.121   -0.310  -0.130   -0.178
## 3 GDP              0.781         -0.218  NA        0.645    0.816   0.370    0.308
## 4 Family           0.741         -0.121   0.645  NA        0.531   0.442    0.206
## 5 Life_~           0.724         -0.310   0.816   0.531   NA       0.360    0.248
## 6 Freed~           0.568         -0.130   0.370   0.442    0.360  NA        0.494
## 7 Trust            0.395         -0.178   0.308   0.206    0.248   0.494  NA
## 8 Gener~           0.180        -0.0884 -0.0105   0.0875   0.108   0.374    0.276
## 9 Dysto~           0.530         0.0840  0.0401   0.148    0.0190  0.0628  -0.0331
## # i 2 more variables: Generosity <dbl>, Dystopia_Res <dbl>
```
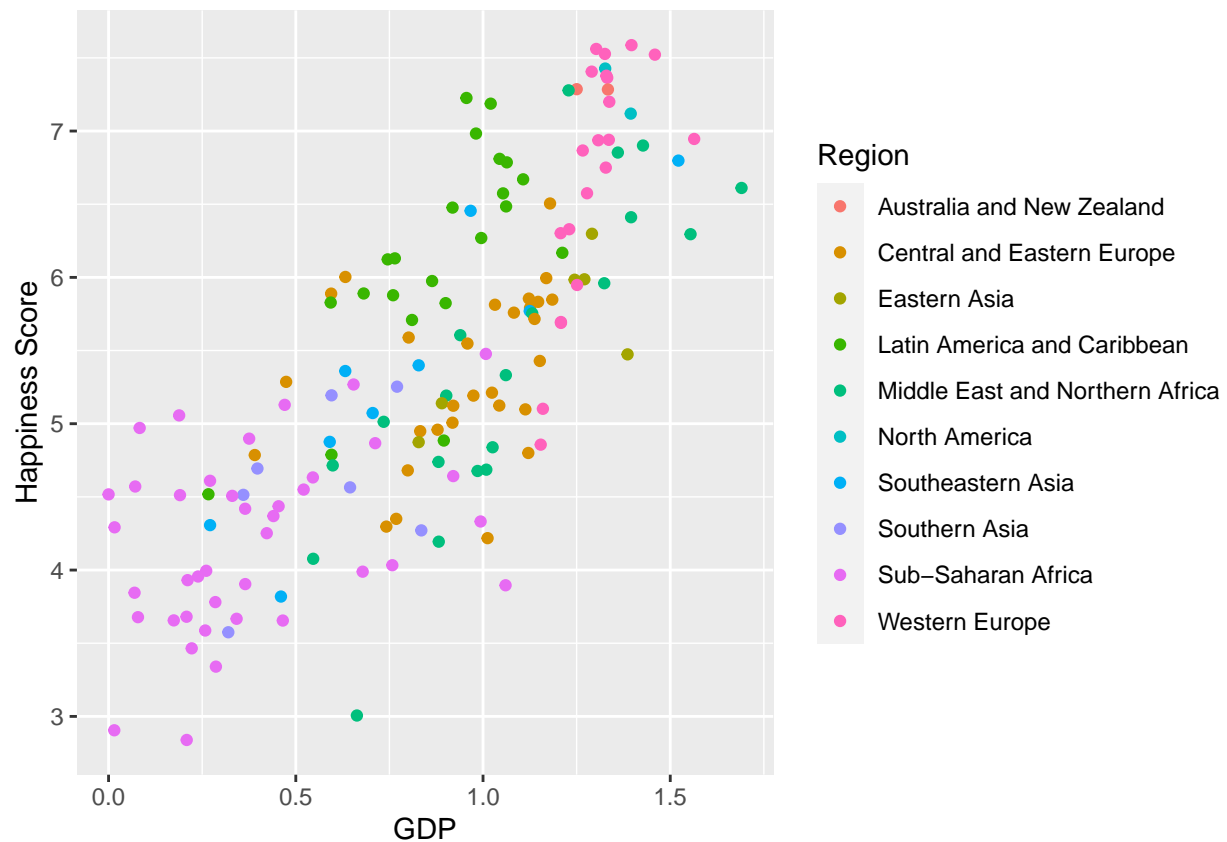
Based on the values of the correlation matrix, **GDP** has the strongest positive relationship with the variable **Happiness_Score**.

Visualize the relationship from above

Graph of the relationship between variables **GDP** and **Happiness Score**.

```
ggplot(whr2015, aes(x = GDP, y = Happiness_Score, group = Region)) +
  geom_point(aes(col=Region)) +
  labs(y = "Happiness Score")
```


It appears that countries with a higher GDP has a higher happiness score.

Create a model to prdict a nation's happiness score

Split data into train and test

```
set.seed(1)

#use 70% of dataset as training set and 30% as test set
sample <- sample(c(TRUE, FALSE), nrow(whr2015), replace=T, prob=c(0.7,0.3))
train <- whr2015[sample, ]
test <- whr2015[!sample, ]

#create the model
whr_lm <- lm(Happiness_Score ~ GDP+Family+Life_Exp+Freedom+Trust+Generosity+Dystopia_Res, data=train)
summary(whr_lm)
```

```
## 
## Call:
## lm(formula = Happiness_Score ~ GDP + Family + Life_Exp + Freedom +
##     Trust + Generosity + Dystopia_Res, data = train)
## 
## Residuals:
##         Min         1Q     Median         3Q        Max
## -6.131e-04 -2.288e-04  1.988e-05  2.298e-04  4.906e-04
## 
## Coefficients:
##                Estimate Std. Error   t value Pr(>|t|)
## (Intercept)  -2.532e-05  1.436e-04    -0.176     0.86
## GDP           1.000e+00  1.342e-04  7451.425   <2e-16 ***
## Family        1.000e+00  1.446e-04  6913.606   <2e-16 ***
## Life_Exp      1.000e+00  1.859e-04  5377.901   <2e-16 ***
## Freedom       9.996e-01  2.495e-04  4005.721   <2e-16 ***
## Trust         1.000e+00  2.655e-04  3765.958   <2e-16 ***
## Generosity    1.000e+00  2.356e-04  4246.088   <2e-16 ***
## Dystopia_Res  1.000e+00  4.835e-05 20682.631   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.0002874 on 106 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 2.714e+08 on 7 and 106 DF,  p-value: < 2.2e-16
```

Output the predicted Happiness Scores for the test data

```
predictions <- predict(whr_lm, newdata=test, type='response')
as.single(predictions)
```

```
##  [1] 7.522214 7.406059 7.378191 7.119565 6.946227 6.940305 6.901348 6.867302
##  [9] 6.575062 6.411171 6.302495 6.294758 6.167708 6.122795 5.987084 5.959677
## [17] 5.889036 5.770440 5.604646 5.548160 5.473856 5.332513 5.286405 5.252882
## [25] 5.211935 5.192300 5.128861 5.123174 4.971137 4.959376 4.948817 4.876034
## [33] 4.799603 4.693976 4.680790 4.677375 4.564673 4.514426 4.419337 4.194281
## [41] 3.988768 3.819222 3.677731 3.655915
## attr(,"Csingle")
## [1] TRUE
```

Compared to the actual test dataset's Happiness Scores

```
test$Happiness_Score
```

```
##  [1] 7.522 7.406 7.378 7.119 6.946 6.940 6.901 6.867 6.575 6.411 6.302 6.295
## [13] 6.168 6.123 5.987 5.960 5.889 5.770 5.605 5.548 5.474 5.332 5.286 5.253
## [25] 5.212 5.192 5.129 5.123 4.971 4.959 4.949 4.876 4.800 4.694 4.681 4.677
## [37] 4.565 4.514 4.419 4.194 3.989 3.819 3.678 3.656
```

The outputs are pretty close.