

# Analysis of Happiness Score Dataset

Prasham Bhuta [github.com/Prashambhuta](https://github.com/Prashambhuta)

16/03/2020

## Dataset

Dataset ranks 155 countries, from the period: 2015 to 2017, by their happiness through 6 indicators.

- economic production
- family
- health
- freedom
- trust (government corruption)
- generosity

The final indicator is dystopia residual. Dystopia is a made up country which ranks lowest in all of the above 6 criteria. The score generated in such a way is used to compare dystopia residual with happiness score.

- low dystopia residual = low happiness score
- high dystopia residual = high happiness score

The dataset is cleaned and provided courtesy Kian

## My Exercise

- To generate plots of indicators vs happiness score.
- To calculate Paerson's correlation coefficient and analyse the dependency of happiness score across all indicators.

## Let's Get Started

### Importing necessary library

```
library(readr)
library(dplyr)
```

### Generating data from csv

```
data <- read.csv("data_sets/World_Happiness_2015_2017.csv", stringsAsFactors = FALSE, header = T)
```

### Analysing the data

```
str(data)
```

```
## 'data.frame': 470 obs. of 11 variables:
## $ Country : chr "Switzerland" "Iceland" "Denmark" "Norway" ...
## $ Happiness.Rank : int 1 2 3 4 5 6 7 8 9 10 ...
## $ Happiness.Score : num 7.59 7.56 7.53 7.52 7.43 ...
## $ Economy..GDP.per.Capita. : num 1.4 1.3 1.33 1.46 1.33 ...
## $ Family : num 1.35 1.4 1.36 1.33 1.32 ...
## $ Health..Life.Expectancy. : num 0.941 0.948 0.875 0.885 0.906 ...
## $ Freedom : num 0.666 0.629 0.649 0.67 0.633 ...
## $ Trust..Government.Corruption.: num 0.42 0.141 0.484 0.365 0.33 ...
## $ Generosity : num 0.297 0.436 0.341 0.347 0.458 ...
## $ Dystopia.Residual : num 2.52 2.7 2.49 2.47 2.45 ...
## $ Year : int 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 ...
```

### Cleaning column headers

```
colnames(data) <-c("Country", "Happiness.Rank", "Happiness.Score", "Economy", "Family", "Health", "Freedom", "Trust", "Generosity", "Dystopia.Residual", "Year")
str(data)
```

```
## 'data.frame': 470 obs. of 11 variables:
## $ Country : chr "Switzerland" "Iceland" "Denmark" "Norway" ...
## $ Happiness.Rank : int 1 2 3 4 5 6 7 8 9 10 ...
## $ Happiness.Score : num 7.59 7.56 7.53 7.52 7.43 ...
## $ Economy : num 1.4 1.3 1.33 1.46 1.33 ...
## $ Family : num 1.35 1.4 1.36 1.33 1.32 ...
## $ Health : num 0.941 0.948 0.875 0.885 0.906 ...
## $ Freedom : num 0.666 0.629 0.649 0.67 0.633 ...
## $ Trust : num 0.42 0.141 0.484 0.365 0.33 ...
## $ Generosity : num 0.297 0.436 0.341 0.347 0.458 ...
## $ Dystopia.Residual: num 2.52 2.7 2.49 2.47 2.45 ...
## $ Year : int 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 ...
```

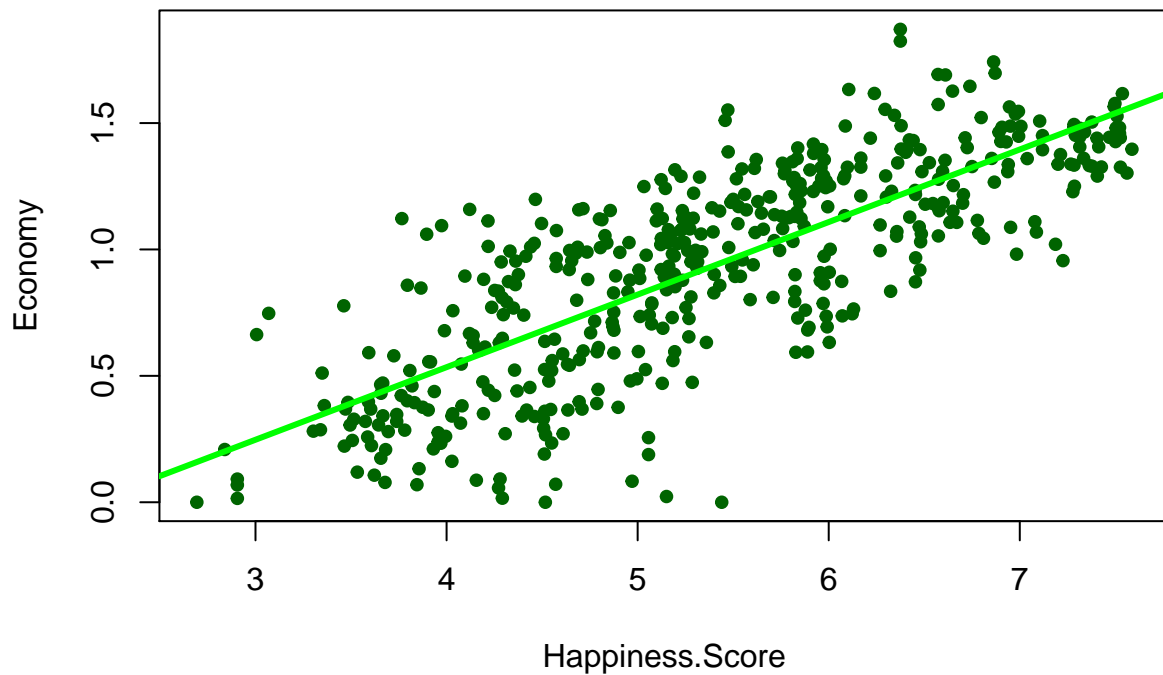
### Correlation between Economy and Happiness Score

```
corr_econ <- cor(data %>% select(Economy, Happiness.Score), method= "pearson")
cat("corr_econ (R):", corr_econ[2], "\nR^2:", corr_econ[2]^2)
```

```
## corr_econ (R): 0.7854496
## R^2: 0.6169311
```

```
data_econ <- data %>% select(Happiness.Score, Economy)
plot(data_econ, main="Economy vs Happiness Score", col="darkgreen", type="p", pch=20, cex=1.2)
abline(lm(data$Economy ~ data$Happiness.Score), col="green", lwd=3)
```

## Economy vs Happiness Score



The correlation coefficient (R) is 0.785 and the square (R<sup>2</sup>) is 0.616.

Happiness score is strongly related to the economy of the country. High economy means

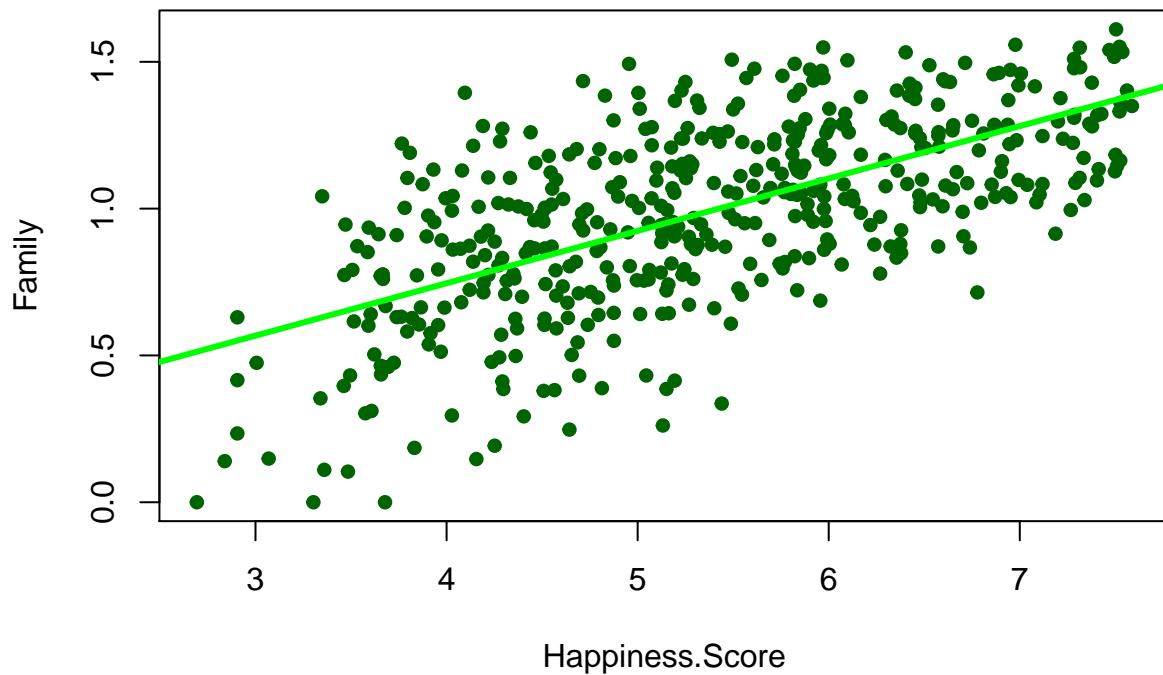
- higher living standards,
- less economical disparity,
- monetary satisfaction across all line of work.

## Family vs Happiness Score

```
data_fam <- data %>% select(Happiness.Score, Family)
corr_fam = cor(data_fam, method="pearson")[2]
cat("R:", corr_fam, "\nR^2:", corr_fam^2)
```

```
## R: 0.636532
## R^2: 0.4051729
```

```
plot(data_fam, col="darkgreen", pch=20, cex=1.3)
abline(lm(data_fam$Family ~ data_fam$Happiness.Score), col="green", lwd=3)
```



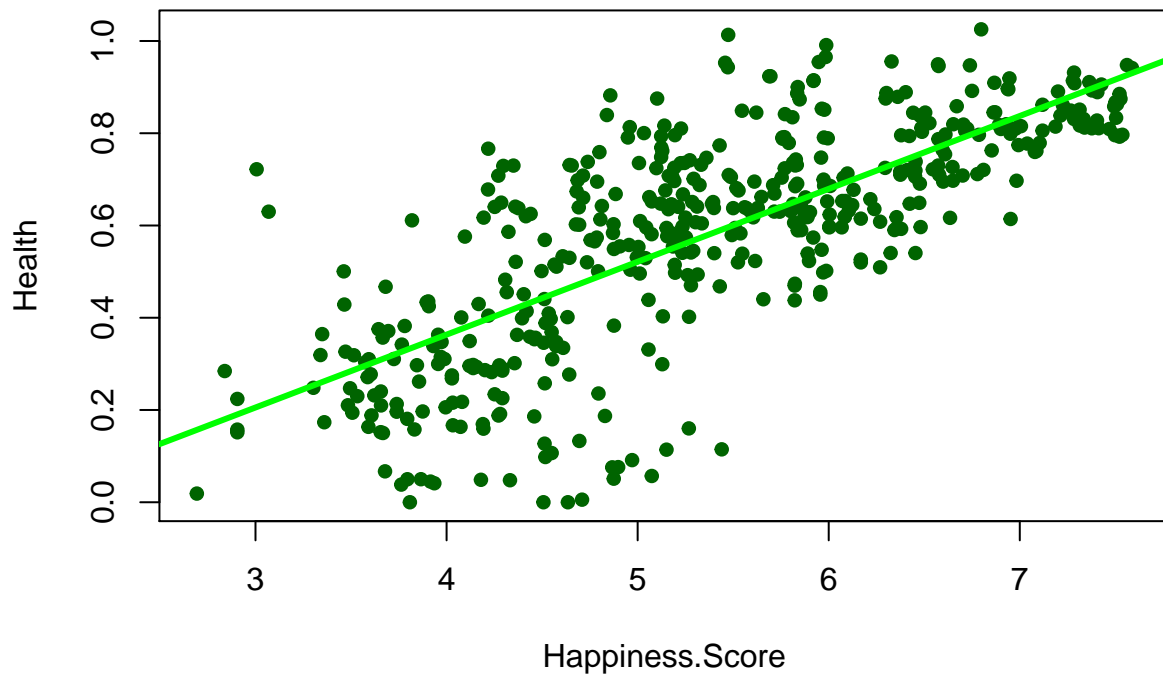
R is 0.63 while R<sup>2</sup> is 0.40, this shows that Happiness Score is strongly dependent on Family. Happy members of family contribute to increase in Family indicator and Happiness Score.

## Health vs Happiness Score

```
data_health <- data %>% select(Happiness.Score, Health)
corr_health <- cor(data_health, method = "pearson")[2]
cat("R:", corr_health, "\nR^2:", corr_health^2)
```

```
## R: 0.7480404
## R^2: 0.5595644
```

```
plot(data_health, col="darkgreen", pch=20, cex=1.3)
abline(lm(data_health$Health ~ data_health$Happiness.Score), col="green", lwd=3)
```



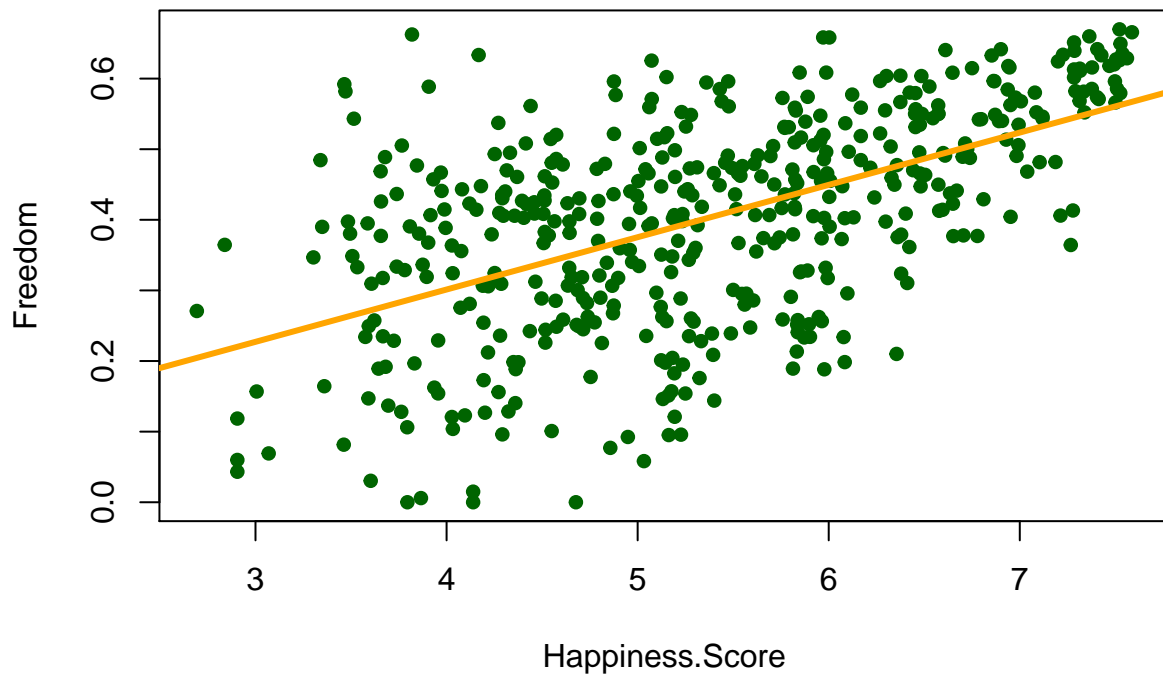
The R2 coefficient is 0.55, this indicates that health has positive relation to Happiness. People with higher life expectancy are happier compared to countries with lower life expectancy.

## Freedom vs Happiness Score

```
data_freedom <- data %>% select(Happiness.Score, Freedom)
corr_freedom <- cor(data_freedom, method = "pearson")[2]
cat("R:", corr_freedom, "\nR^2:", corr_freedom^2)
```

```
## R: 0.5603534
## R^2: 0.3139959
```

```
plot(data_freedom, col="darkgreen", pch=20, cex=1.3)
abline(lm(data_freedom$Freedom ~ data$Happiness.Score), col="orange", lwd=3)
```



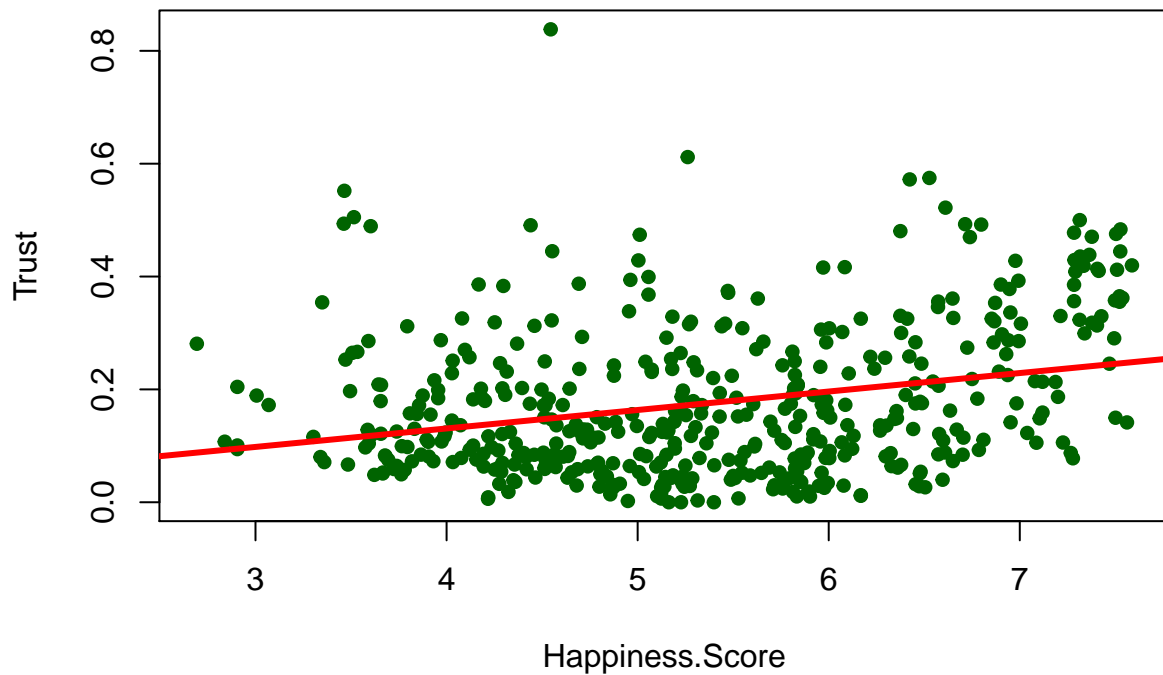
The R2 coefficient for Freedom vs Happiness score is 0.31. Compared to Economy, Family and Health; the coefficient is lower. However as per the data, happiness has strong dependence on Freedom. Freedom includes freedom to act upon a wish, such as changing careers, choosing partners, starting new business etc. People which such liberty are believed to be happier.

## Trust vs Happiness Score

```
data_trust <- data %>% select(Happiness.Score, Trust)
corr_trust <- cor(data_trust, method = "pearson")[2]
cat("R:", corr_trust, "\nR^2:", corr_trust^2)
```

```
## R: 0.2821296
## R^2: 0.07959714
```

```
plot(data_trust, col="darkgreen", pch=20, cex=1.3)
abline(lm(data_trust$Trust ~ data$Happiness.Score), col="red", lwd=3)
```



The R2 coefficient for Trust on government is 0.07. This indicates that the role of a government in the happiness of an human being is low. As long as government boosts the economy, and provides support & health to the families, people are happy.

## Generosity vs Happiness Score

```
data_generosity <- data %>% select(Happiness.Score, Generosity)
corr_generosity <- cor(data_generosity, method = "pearson")[2]
cat("R:", corr_generosity, "\nR^2:", corr_generosity^2)
```

```
## R: 0.2216836
## R^2: 0.04914363
```

```
plot(data_generosity, col="darkgreen", pch=20, cex=1.3)
abline(lm(data_generosity$Generosity ~ data$Happiness.Score), col="red", lwd=3)
```



The R2 coefficient for Generosity (0.04) is the lowest among all indicators. Generosity or helpfulness of other people has no to minimum say in the overall happiness of the society.

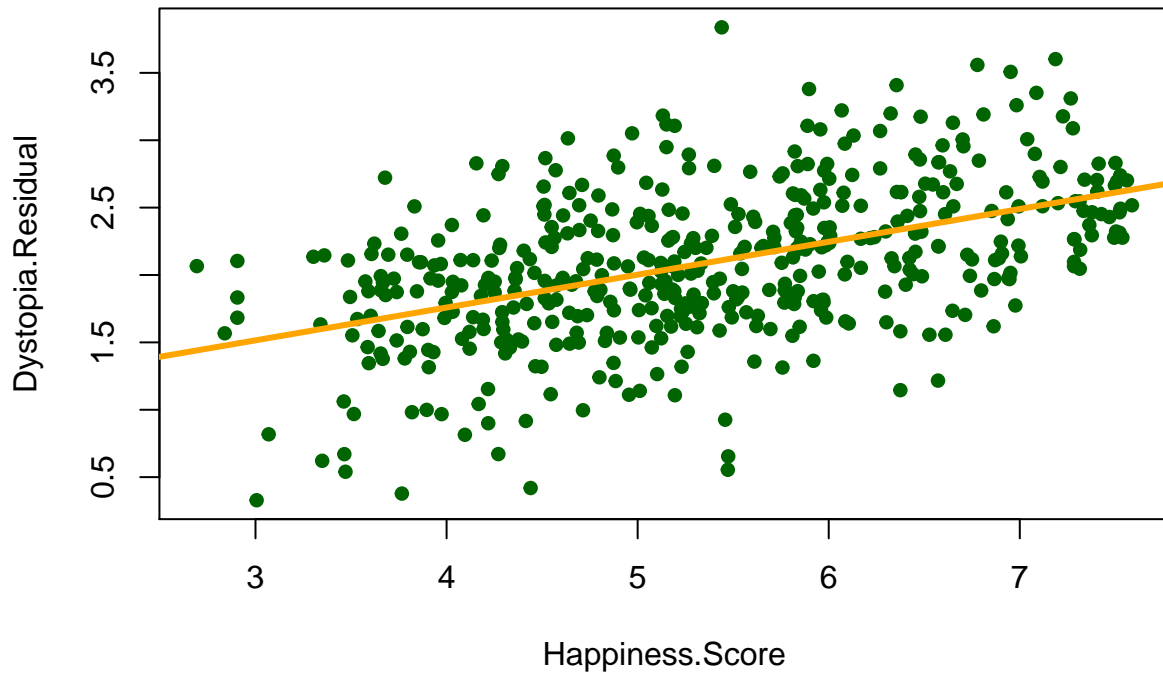
### Dystopia Residual vs Happiness Score

```
data_dys_res <- data %>% select(Happiness.Score, Dystopia.Residual)
corr_dys_res <- cor(data_dys_res, method = "pearson")[2]
cat("R:", corr_dys_res, "\nR^2:", corr_dys_res^2)
```

```
## R: 0.4897472
## R^2: 0.2398523
```

```
plot(data_dys_res, col="darkgreen", pch=20, cex=1.3)
abline(lm(data_dys_res$Dystopia.Residual ~ data$Happiness.Score), col="orange", lwd=3)
```





The R2 coefficient for Dystopia Residual is 0.24. Dystopia Residual means how much the said country is doing better than the worst country i.e. **Dystopia**. It is strong indicator and shows that countries with averages higher than the median are more happier than country whose indicator score lie close to that of Dystopia.

## Conclusion

From the data we can conclude which indicators are major influence on happiness, and which indicators are negligible. The following table depicts the dependency:

Level of Dependency	Indicator 1	Indicator 2	Indicator 3
Strongly Dependent	Economy (0.62)	Health (0.55)	Family (0.40)
Mildly Dependent	Freedom (0.31)	Dystopia Residual (0.24)	-
Weakly Dependent	Trust [Government] (0.08)	Generosity (0.05)	-

Check out the Jupyter Notebook for the same