

IST 707: Data Analytics

World Happiness Report Analytics

Samantha Brennen-Lisko, Holden Herrell, Jogesh Pugazhendhi, & Matthew Sutherland

Topics

- About the Data
- Exploratory Analysis
- Models
- Results
- Conclusions

About the Data

World Happiness Report

Source: Kaggle

(<https://www.kaggle.com/unsdsn/world-happiness>)

Description:

Data set ranking the happiness score of 155 countries around the world. Released annually resulting from Gallup polls.

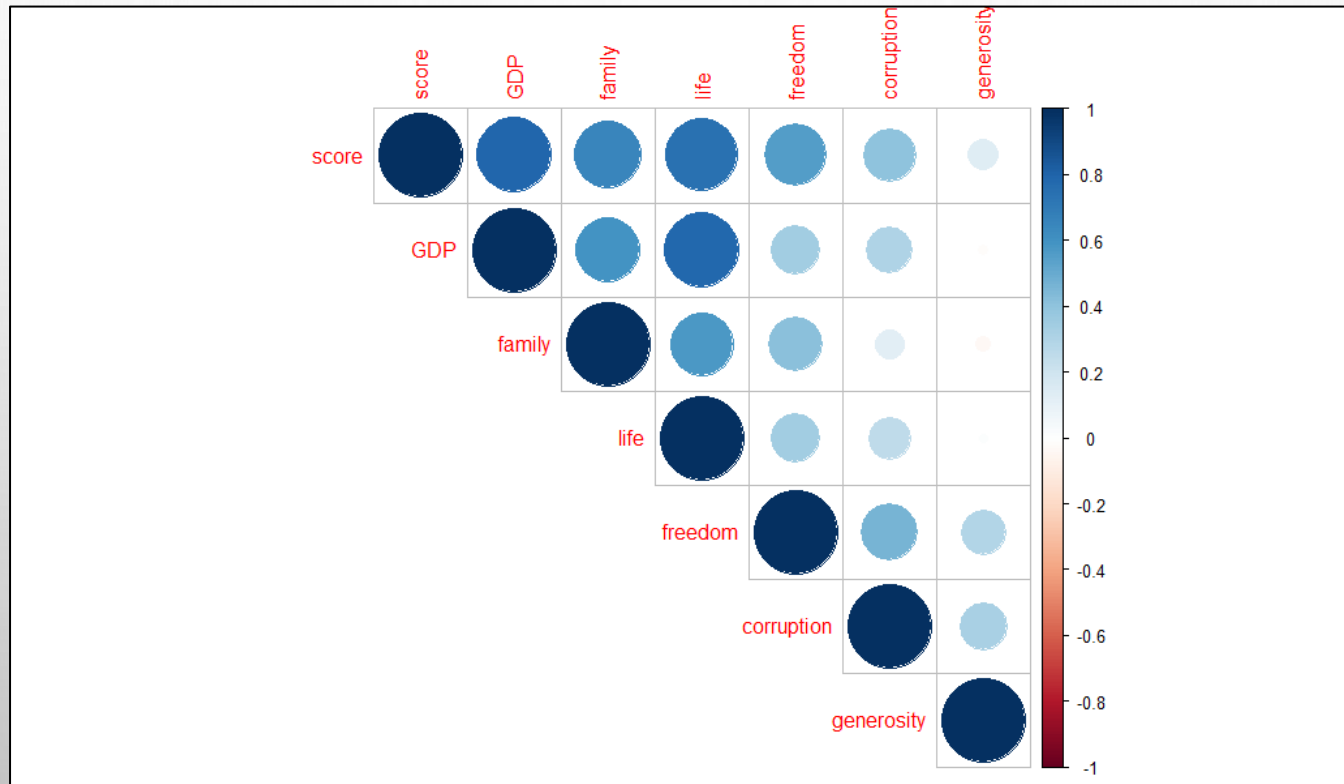
Key Fields: (Indexes) GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perceptions of corruption

About the Data

- Data consisted of 2015-2019 World Happiness Report survey results
- Not all years had same data fields
- Some partially existing fields were added (i.e. continent) to years where they were missing
- New fields were generated to support modeling (Discretized Happiness score i.e. 3 to 4)

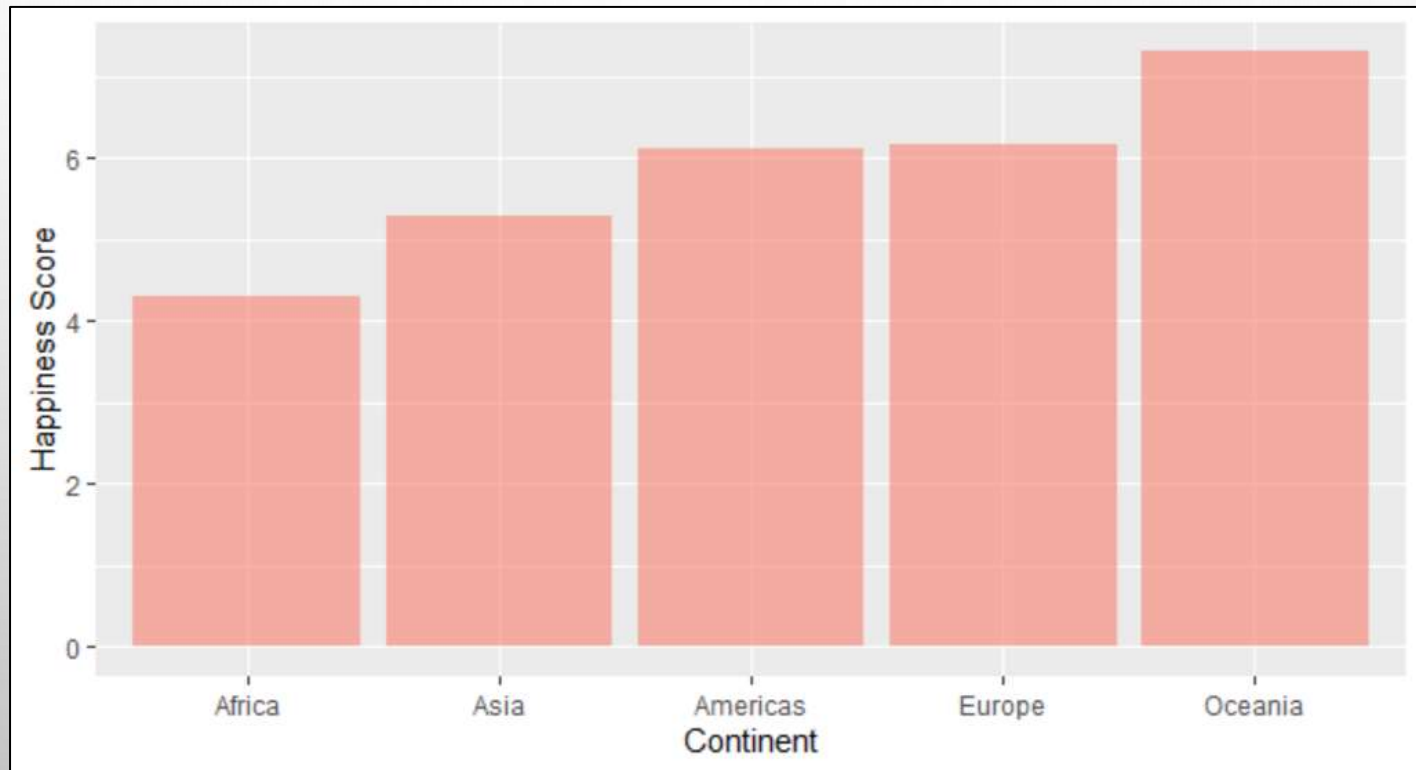
Exploratory Analysis

- Correlation of attributes



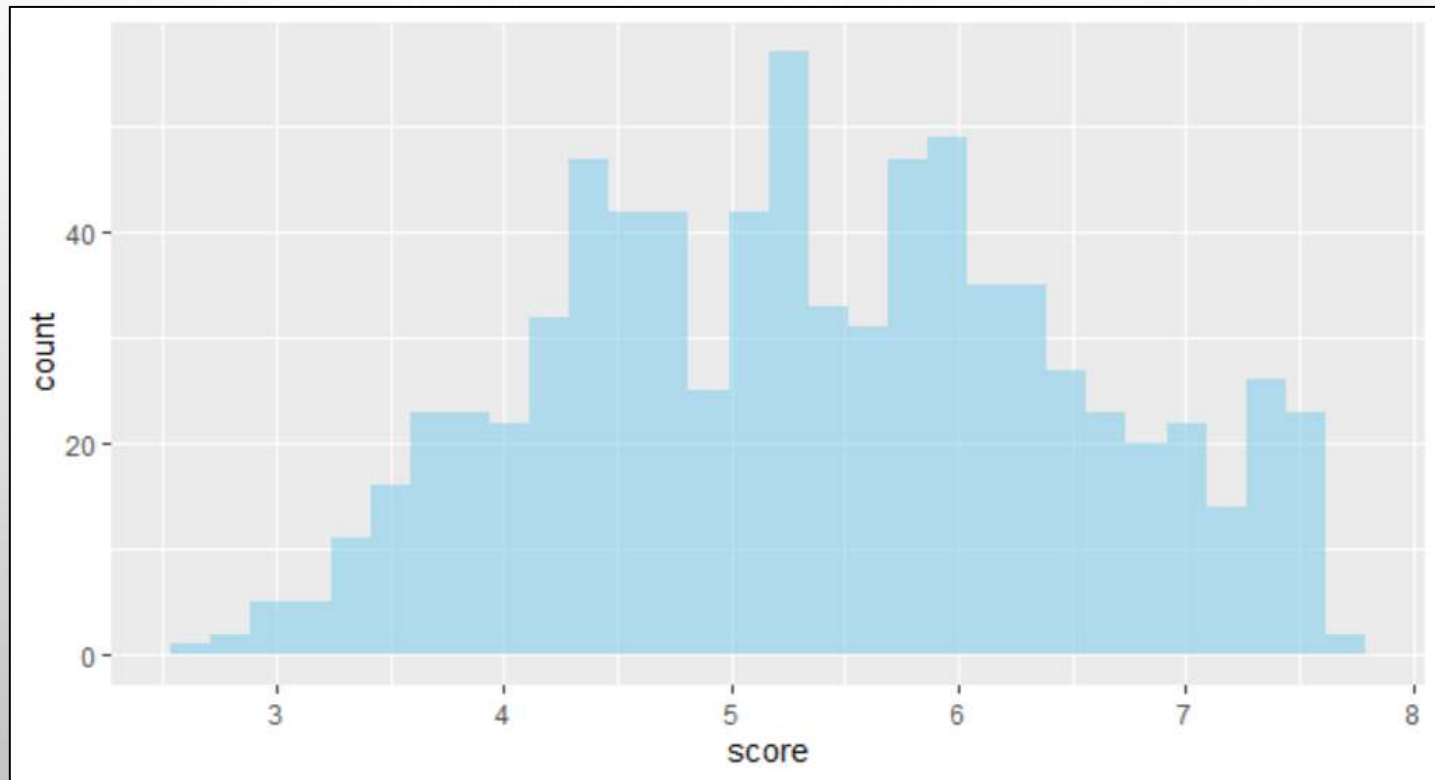
Exploratory Analysis

- Happiness Score by continent



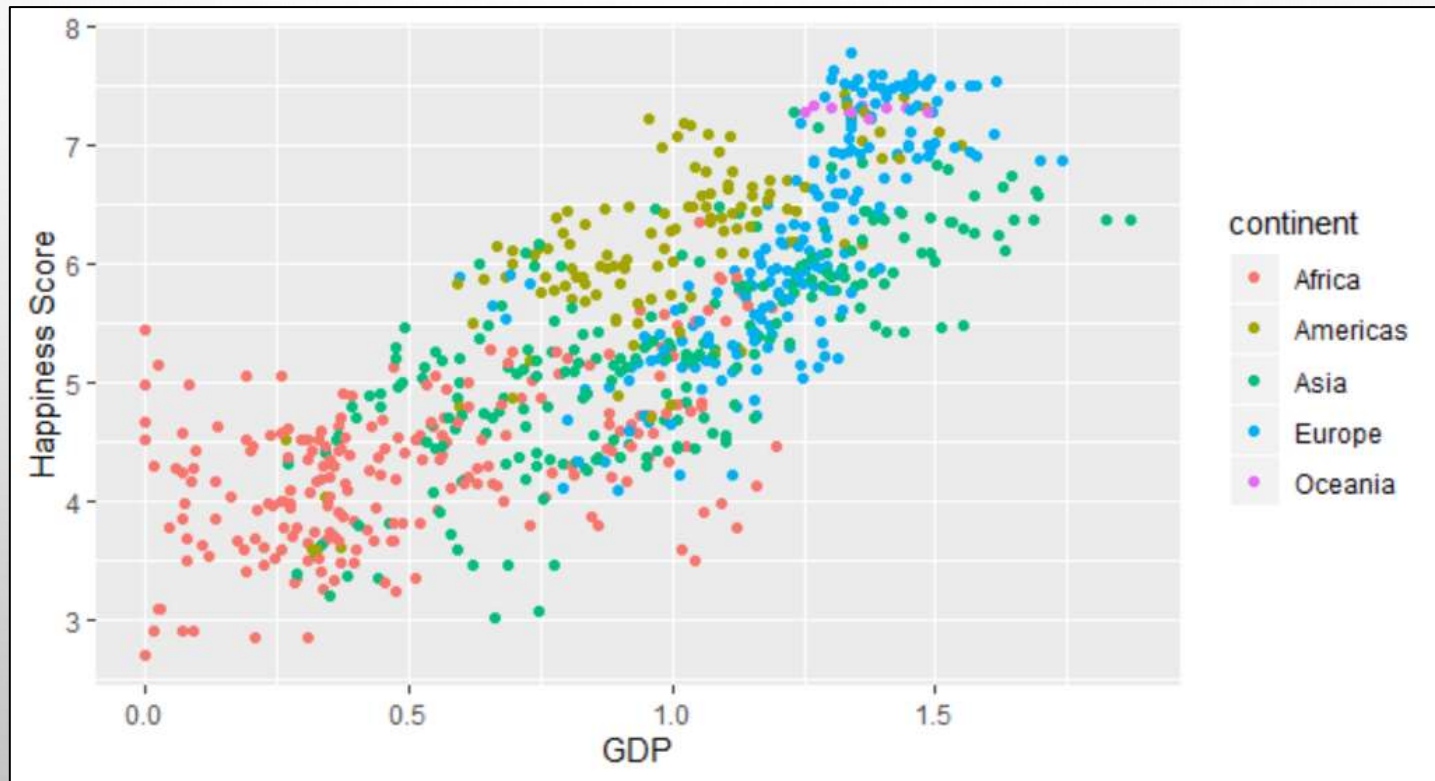
Exploratory Analysis

- Happiness Score distribution



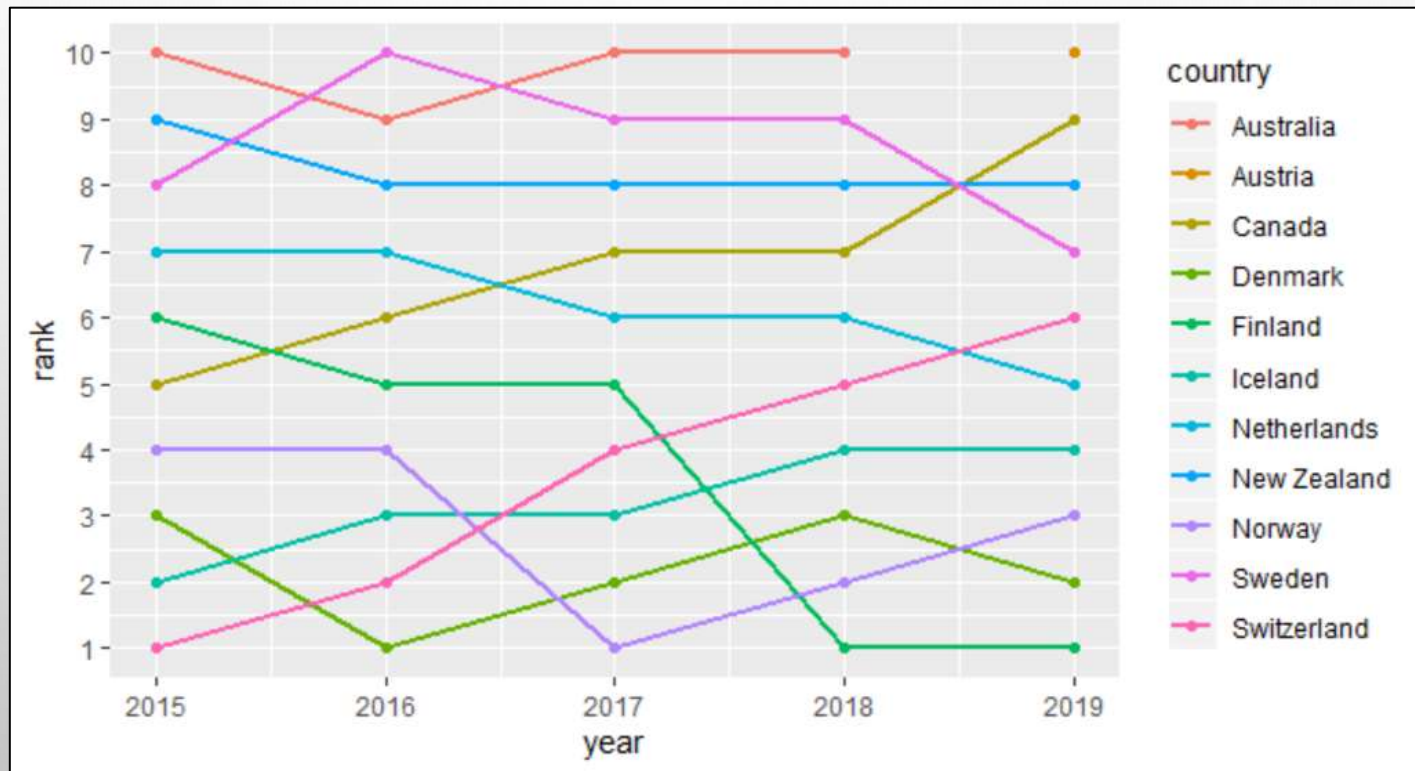
Exploratory Analysis

- Happiness Score vs GDP



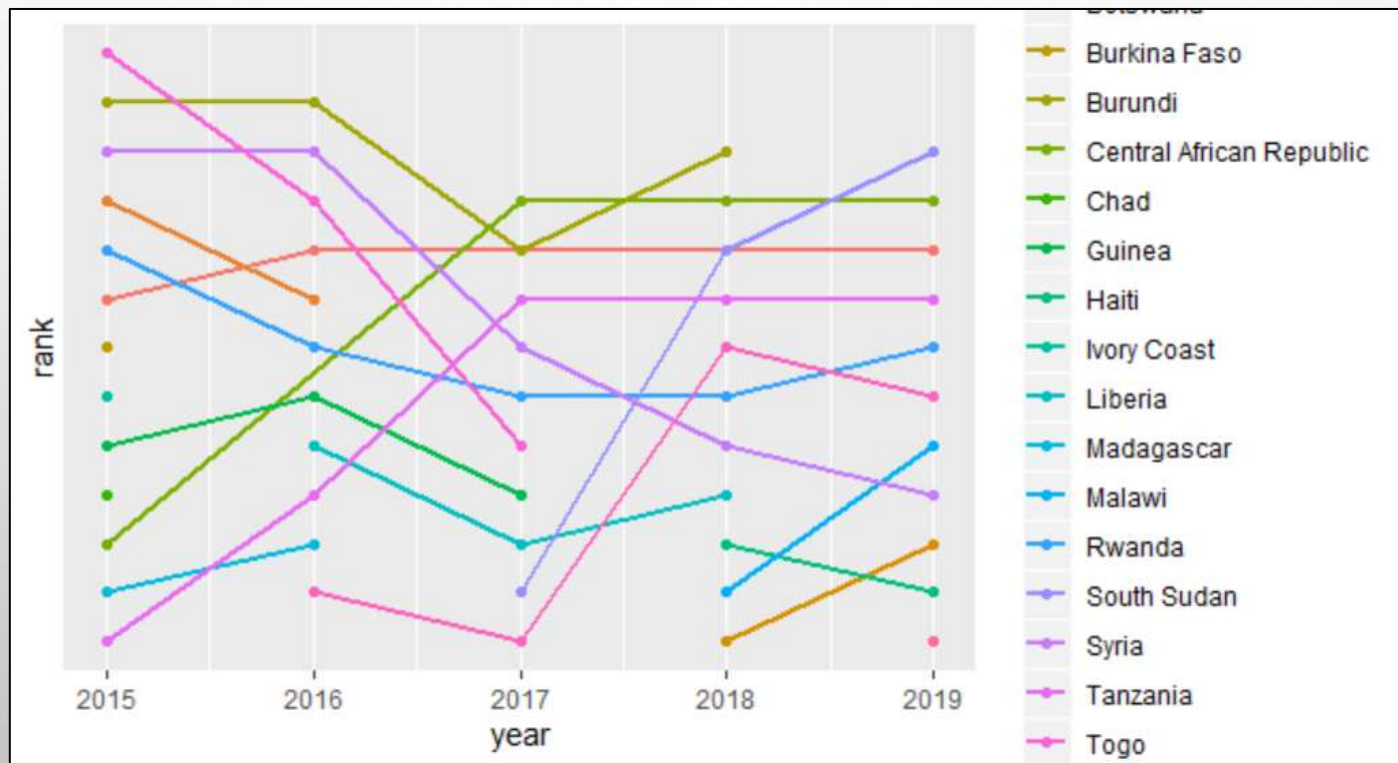
Exploratory Analysis

- Top ten countries ranked by year



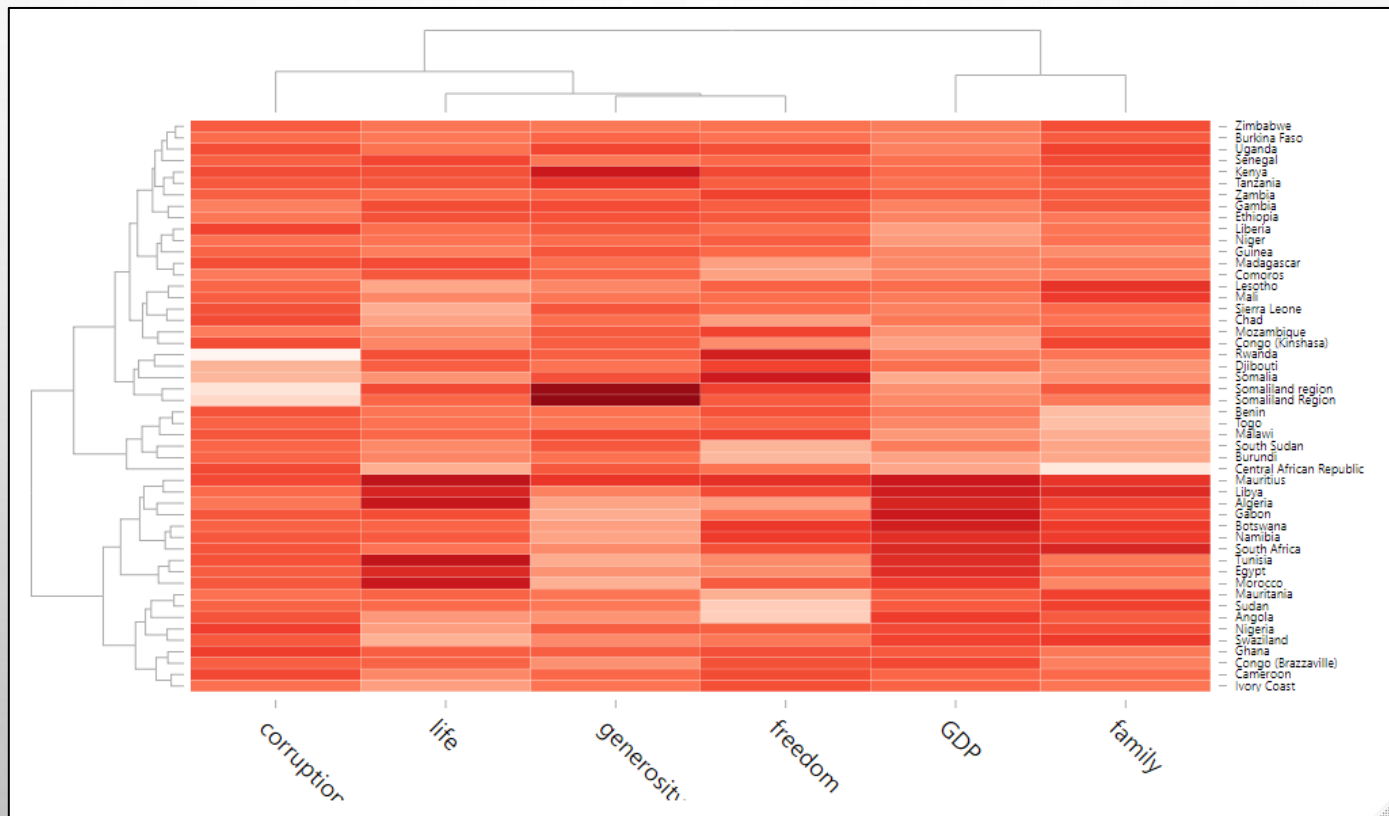
Exploratory Analysis

- Bottom ten countries ranked by year



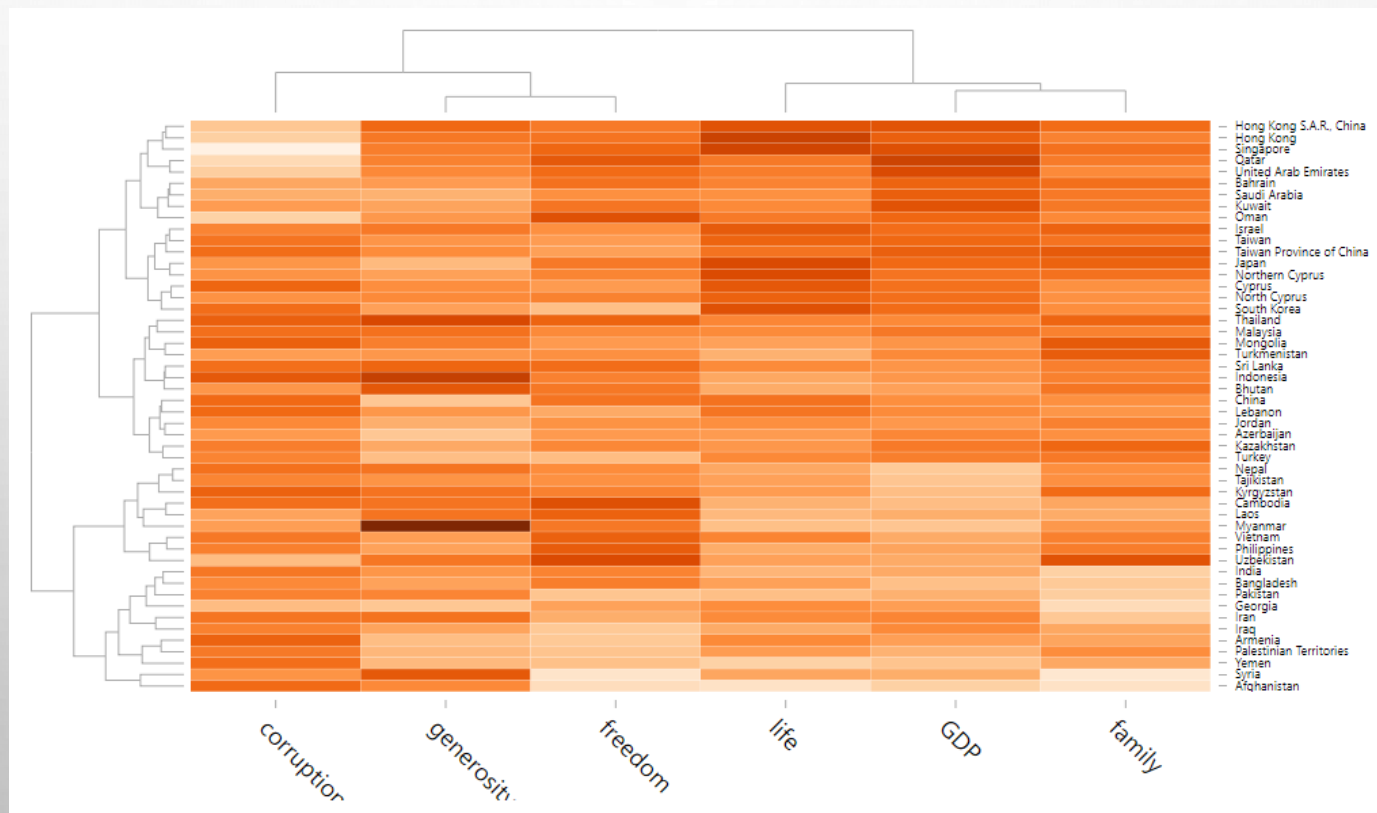
Exploratory Analysis

- Heat Map: Africa



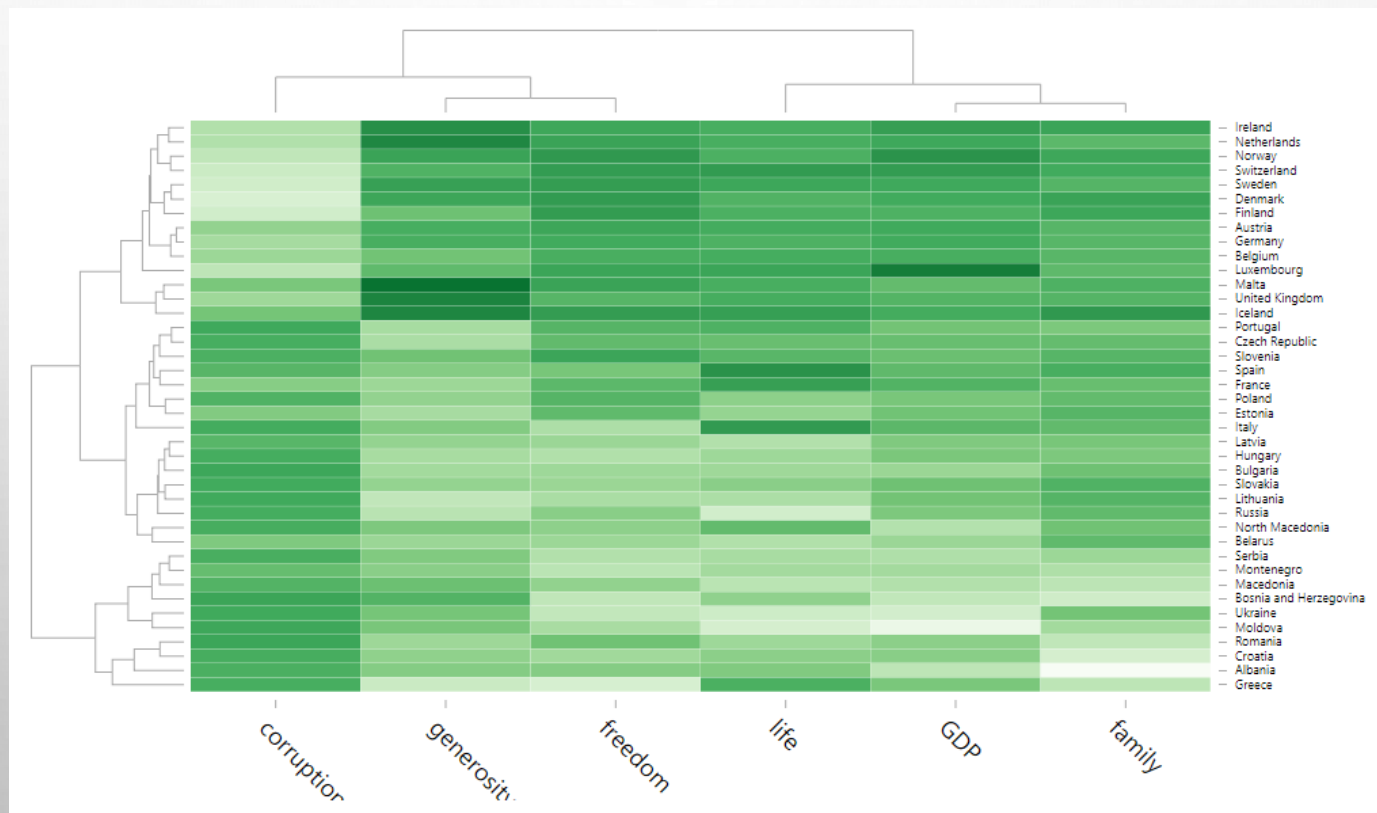
Exploratory Analysis

- Heat Map: Asia



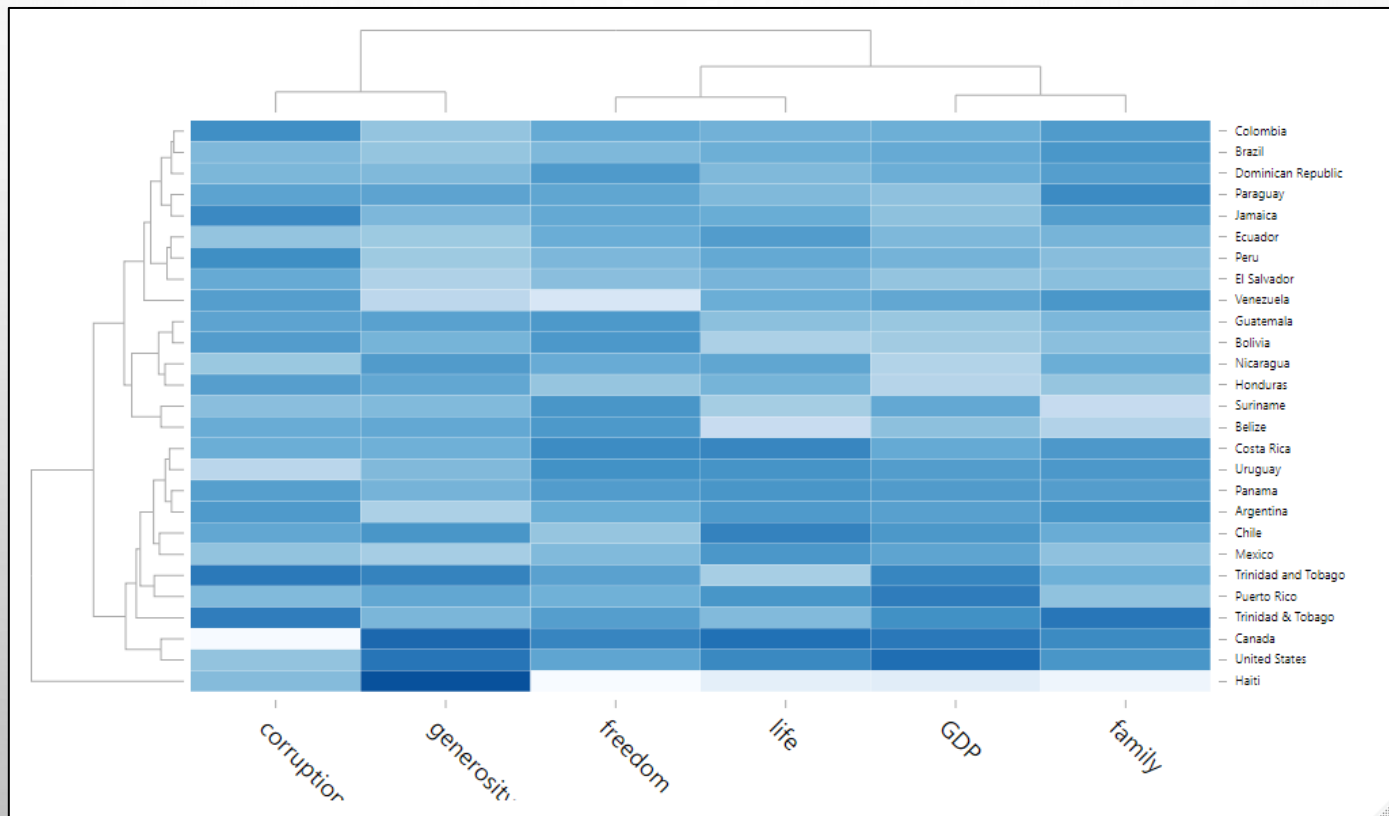
Exploratory Analysis

- Heat Map: Europe



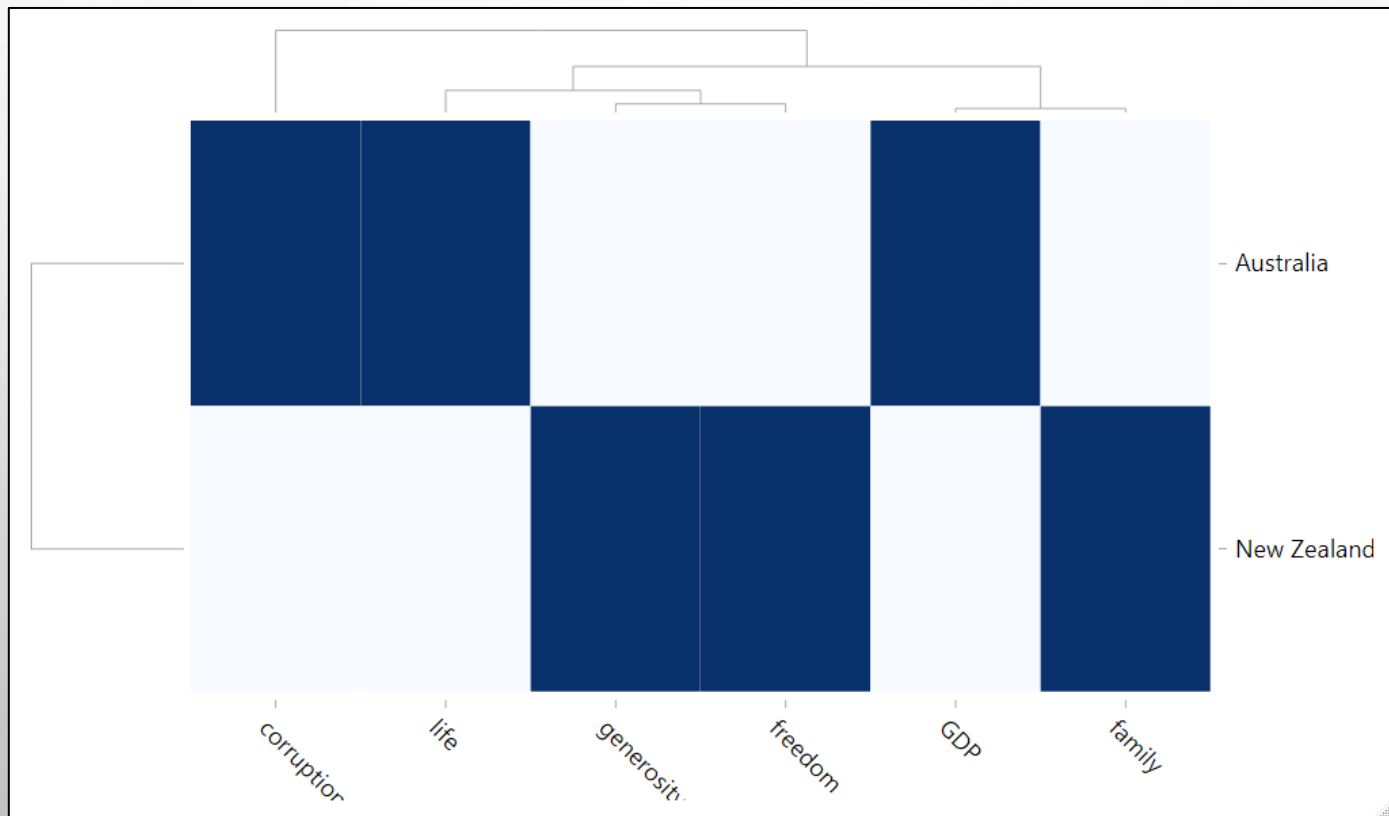
Exploratory Analysis

- Heat Map: Americas

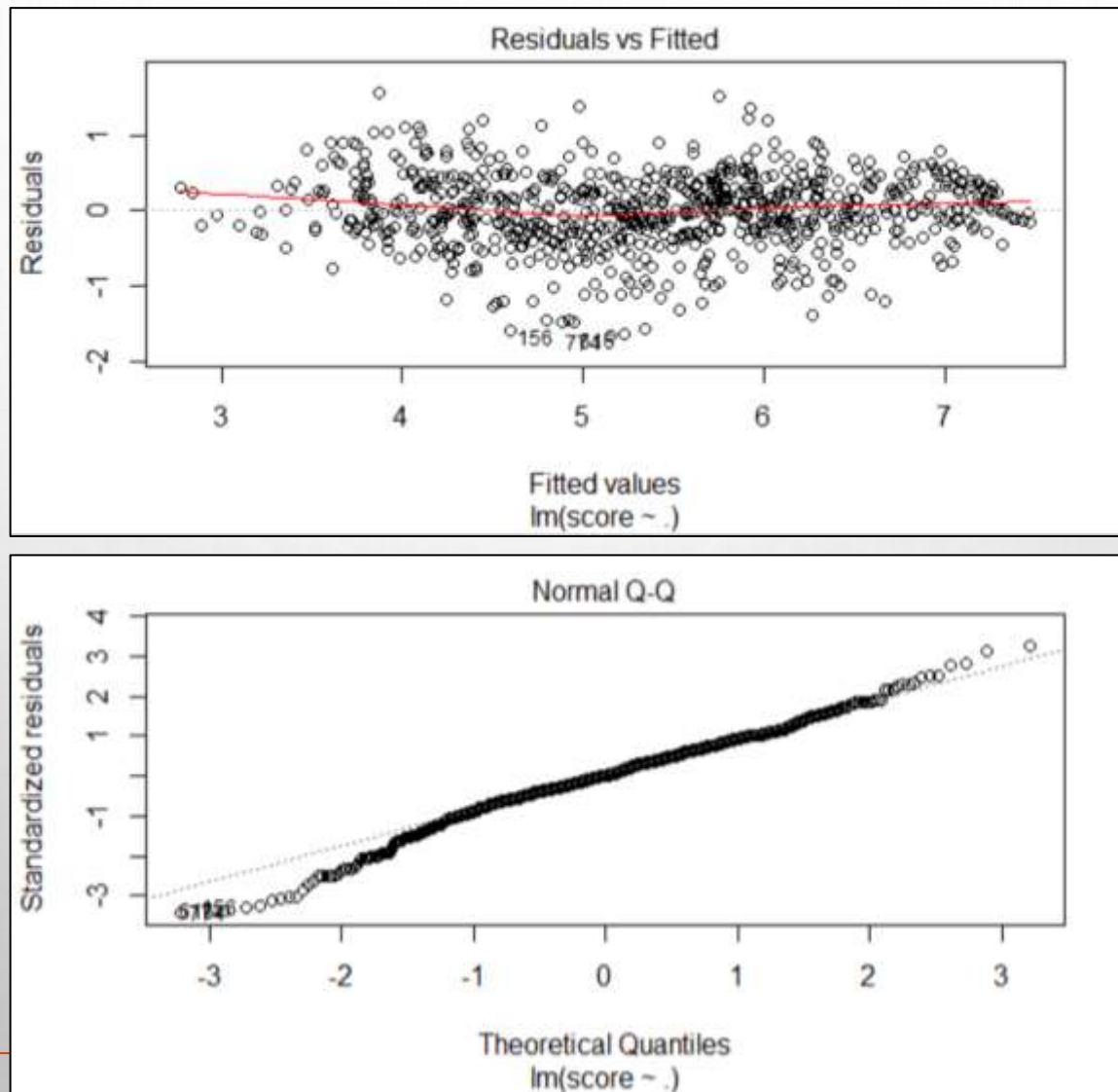


Exploratory Analysis

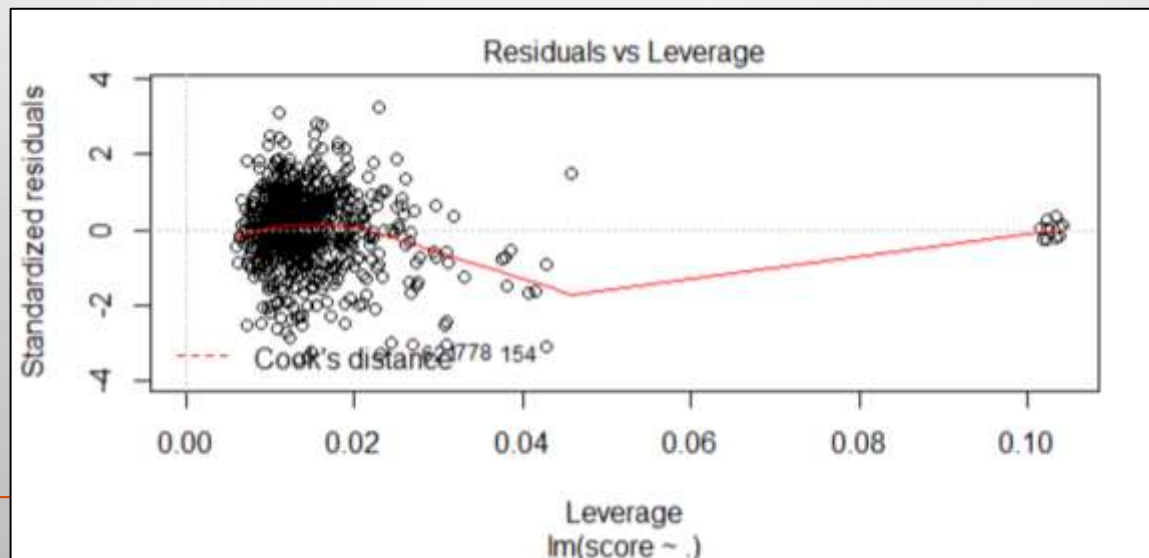
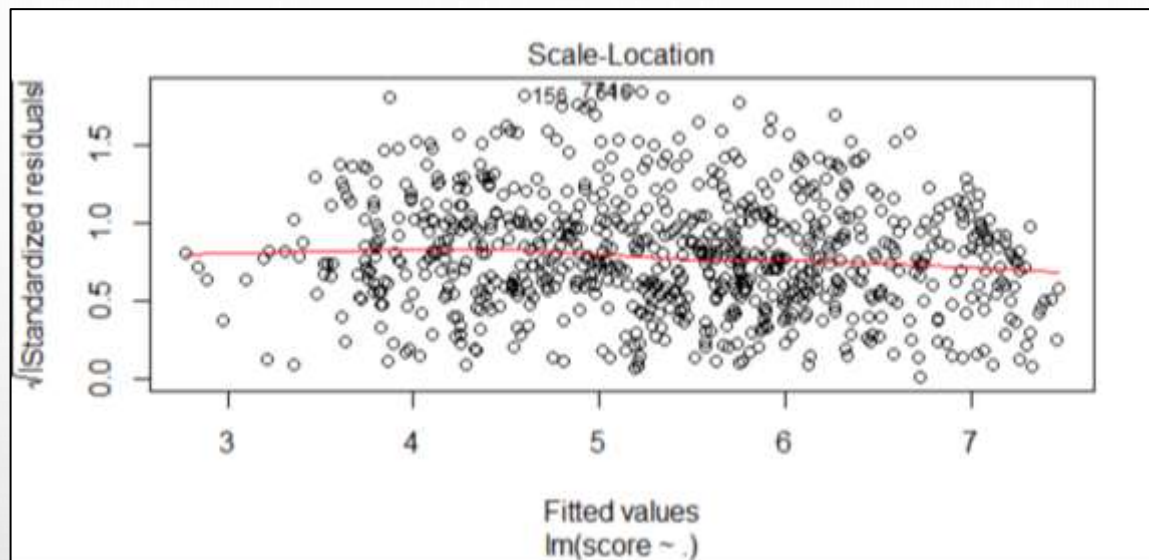
- Heat Map: Oceania



Models—Linear Regression



Models—Linear Regression



Models—Linear Regression

- Output: $2.1712 + \underline{1.1439}GDP + 0.7010Family + \underline{0.9613}Life + \underline{1.2937}Freedom + 0.7731Corruption + 0.9049Generosity$
- Freedom, GDP, and Life Expectancy are most impactful in determining Happiness Score
- $R^2 = 0.7702$

Models—Random Forest

- RMSE = 0.4519984

Models—SVM

Models—Naïve Bayes

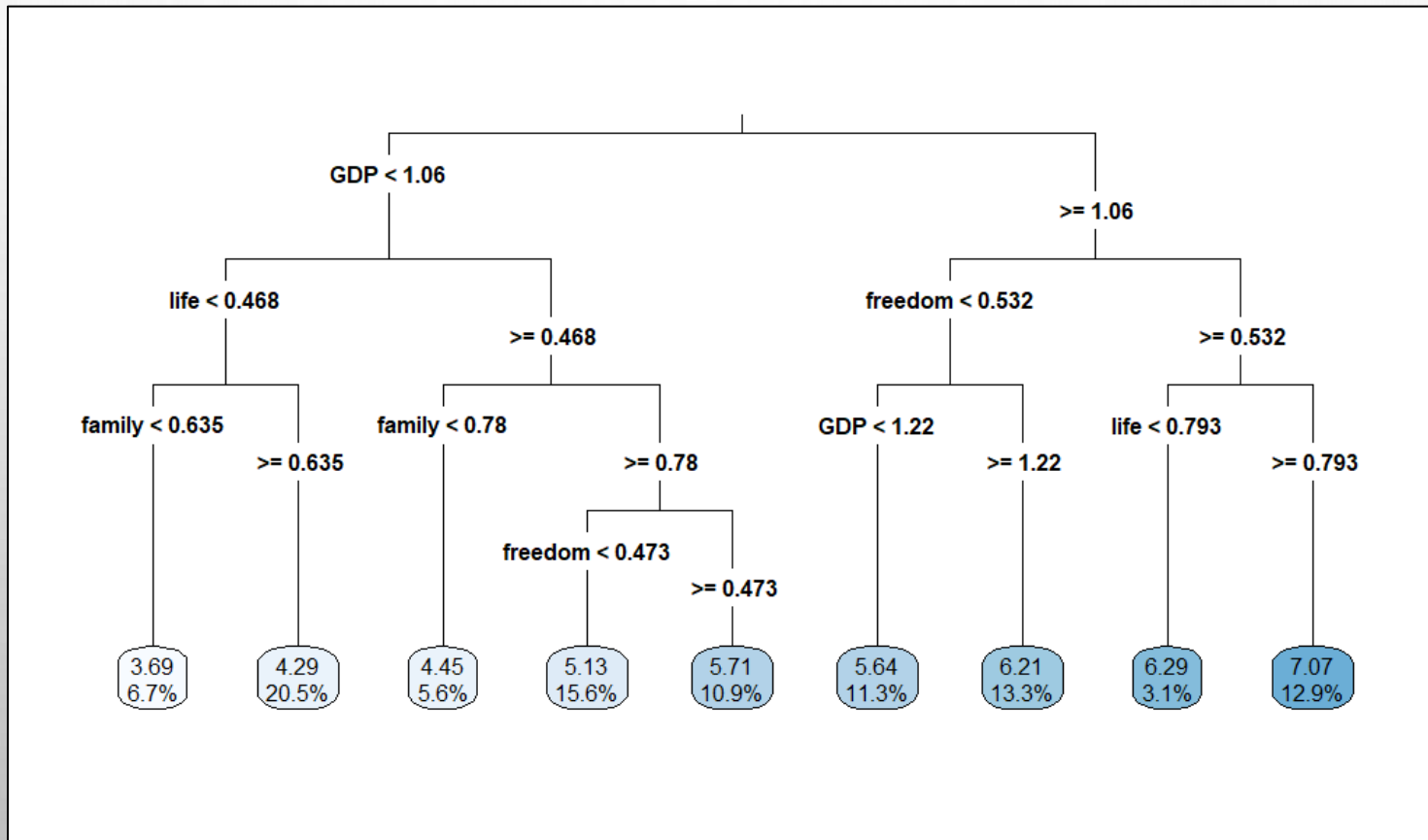
- Used to predict Happiness Score discretized group (3 to 4, 4 to 5, etc.)
- Confusion Matrix

	Predicted					
	Class	3 to 4	4 to 5	5 to 6	6 to 7	7 to 8
Actual	3 to 4	2	3	0	0	0
	4 to 5	0	9	3	0	0
	5 to 6	0	1	10	3	0
	6 to 7	0	0	1	6	2
	7 to 8	0	0	0	1	2

- Accuracy: 67.442%

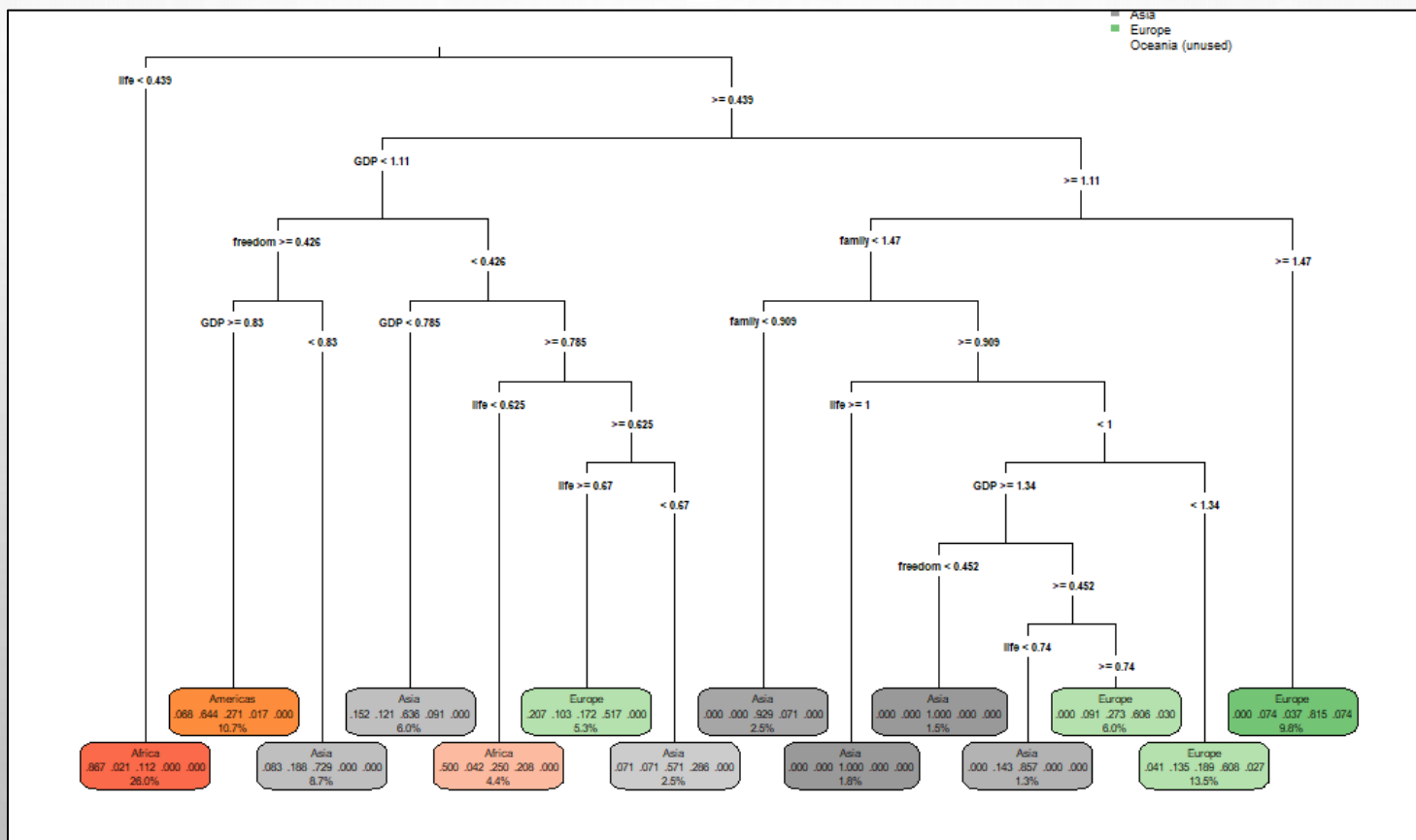
Models—Decision Tree

- Predicting Happiness Score



Models—Decision Tree

- Predicting Continent



Results

Predicting Happiness Score

- Linear Model: Freedom is most impactful on Happiness Score, 77% Accurate
- Random Forest:
- SVM:
- Decision Trees: GDP is most impactful on Happiness Score

Predicting Continent

- Life Expectancy, GDP biggest identifiers

Predicting Happiness Score Group

- Naïve Bayes: 67% Accurate

Conclusions

- GDP appears to be the most influential in determining a country's Happiness Score
- Life expectancy and Freedom are also substantial in determining a country's Happiness Score