World Happiness Report

Final Project

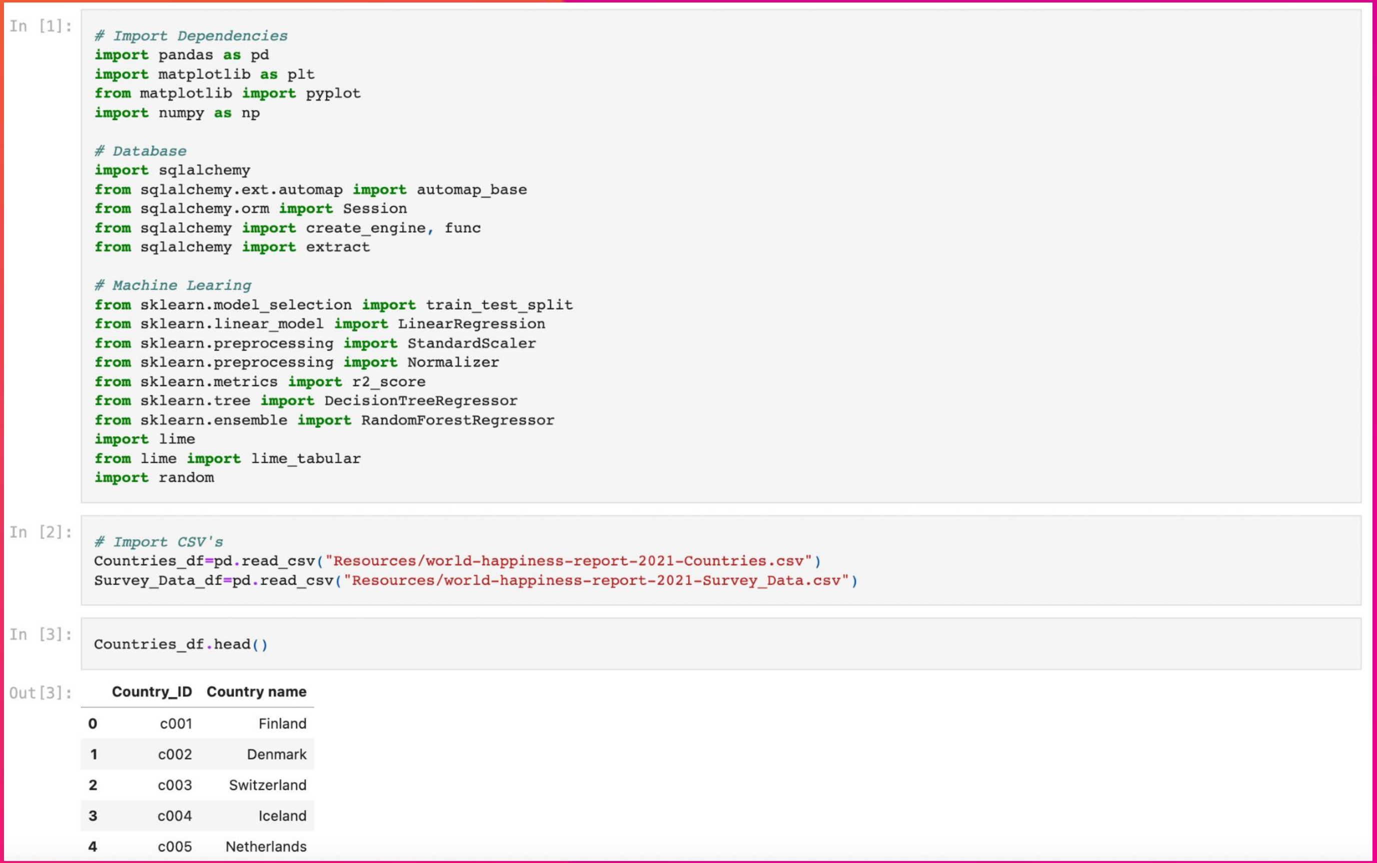
Mareike Van Wie
Erik Guerra
Marianne Valdespino

Languages

Python

Packages used:

- Pandas
- Numpy
- Sklearn
- Matplotlib
- Lime



In [1]:

```
# Import Dependencies
import pandas as pd
import matplotlib as plt
from matplotlib import pyplot
import numpy as np

# Database
import sqlalchemy
from sqlalchemy.ext.automap import automap_base
from sqlalchemy.orm import Session
from sqlalchemy import create_engine, func
from sqlalchemy import extract

# Machine Learing
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import Normalizer
from sklearn.metrics import r2_score
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor
import lime
from lime import lime_tabular
import random
```

In [2]:

```
# Import CSV's
Countries_df=pd.read_csv("Resources/world-happiness-report-2021-Countries.csv")
Survey_Data_df=pd.read_csv("Resources/world-happiness-report-2021-Survey_Data.csv")
```

In [3]:

```
Countries_df.head()
```

Out[3]:

	Country_ID	Country name
0	c001	Finland
1	c002	Denmark
2	c003	Switzerland
3	c004	Iceland
4	c005	Netherlands

Preprocessing and cleaning data

- Dropped columns
- Changed the index to name of countries
- Returned matrix and series of the data using `X = df2.iloc[:,1:]`

  `y = df2.iloc[:, 0]`

- Scaled the data using `scaler = Normalizer().fit(X_train)`

Splitting the data

- The data was split using a test size of point three and randomness of 101
- We then scaled the data and input it into our first model, a Linear Regression model

Models

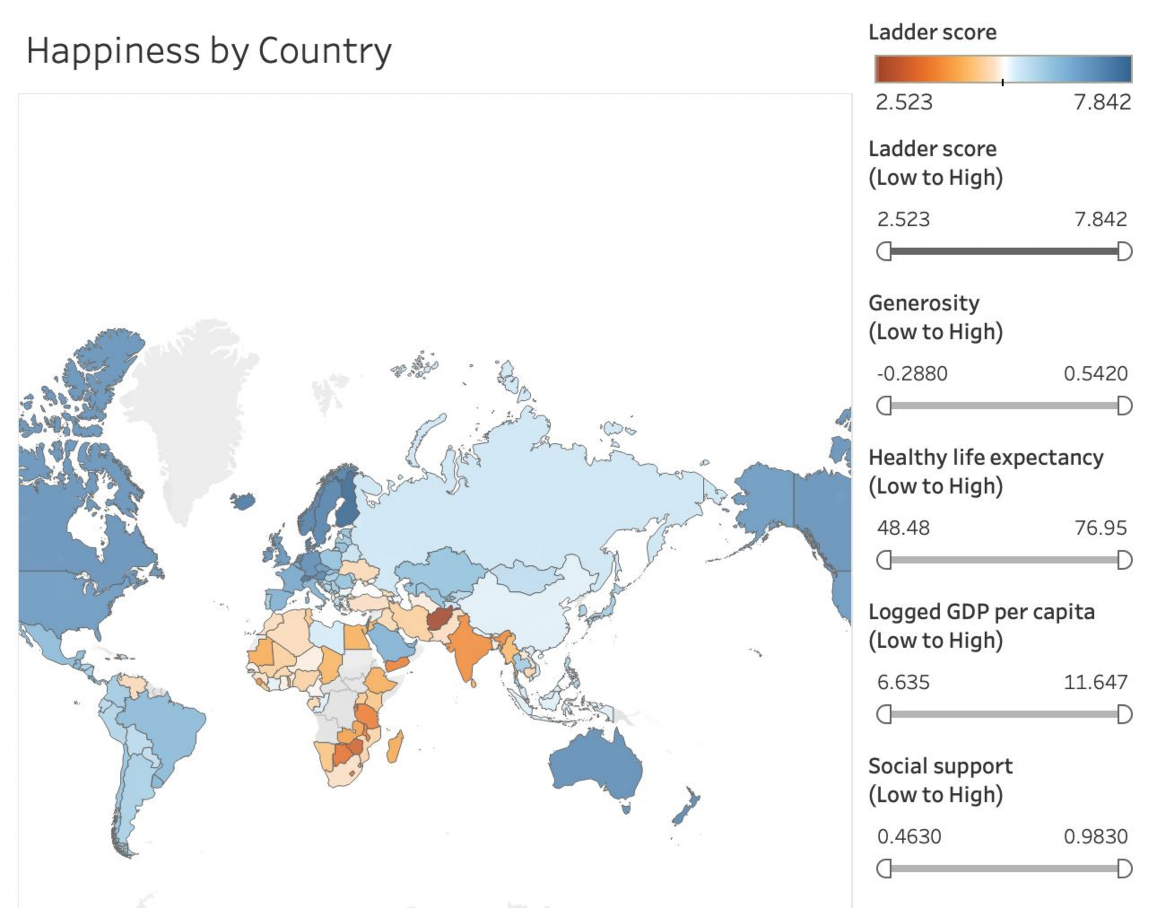
- We decided to compare **three models** to see which one is best with our data
- We found the **<u>linear regression model</u>** was best with a .993 accuracy, decision tree with .992 and random forest with .997
- We chose these three because:
   a. Linear is a good place to start, its advantage is estimation procedure simple and easy to understand
   b. Decision trees allow all aspects to be challenged, however, we also understood it could lead to overfitting of the data
   c. Random forest is quick, allows for high dimensionality and has a low bias

Methods for Visualizations - Tableau

Interactive World Map:

Sliders for each variable

Click here to view on Tableau



Happiness by Country

Ladder score

2.523 7.842

Ladder score (Low to High)

2.523 7.842

Generosity (Low to High)

-0.2880 0.5420

Healthy life expectancy (Low to High)

48.48 76.95

Logged GDP per capita (Low to High)

6.635 11.647

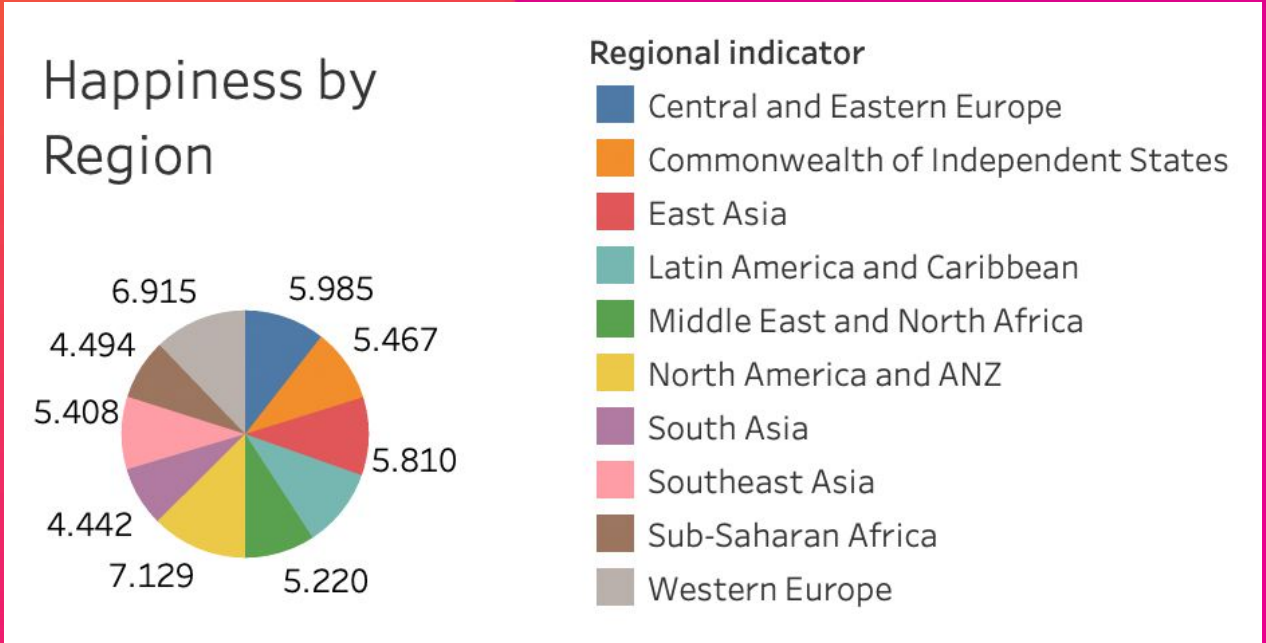
Social support (Low to High)

0.4630 0.9830

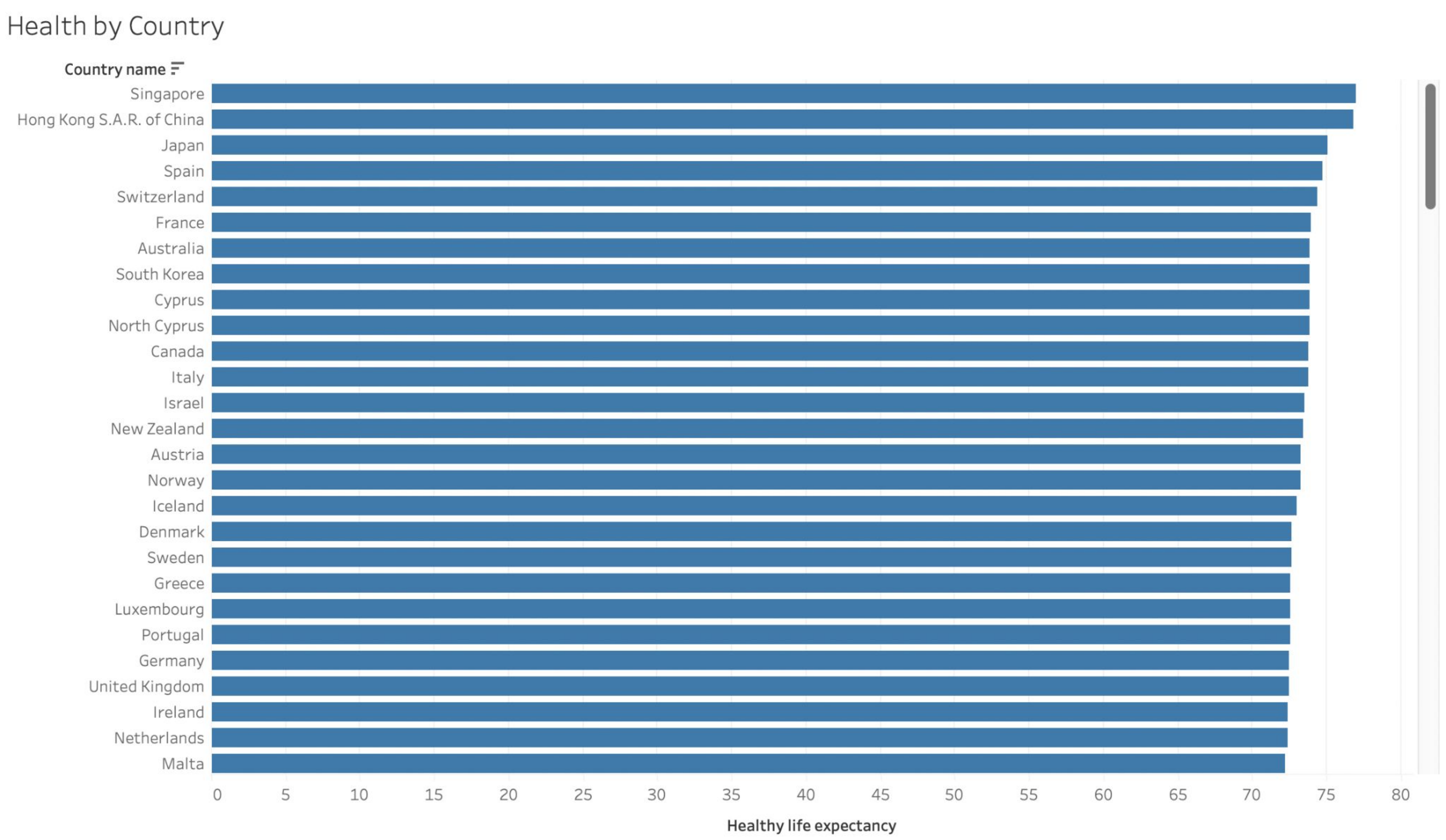
Methods for Visualizations - Tableau

Pie Chart:

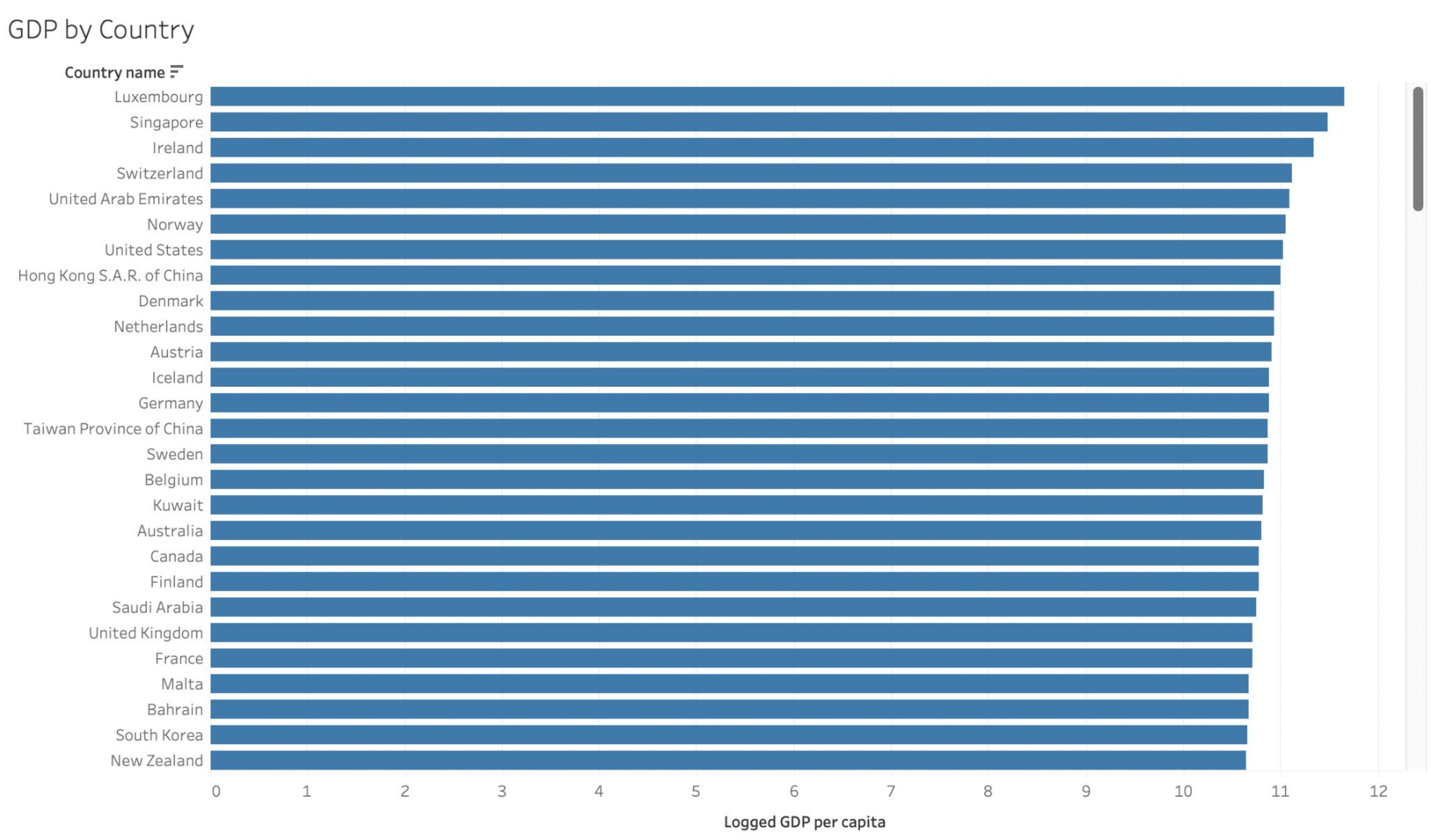
Happiness by Region



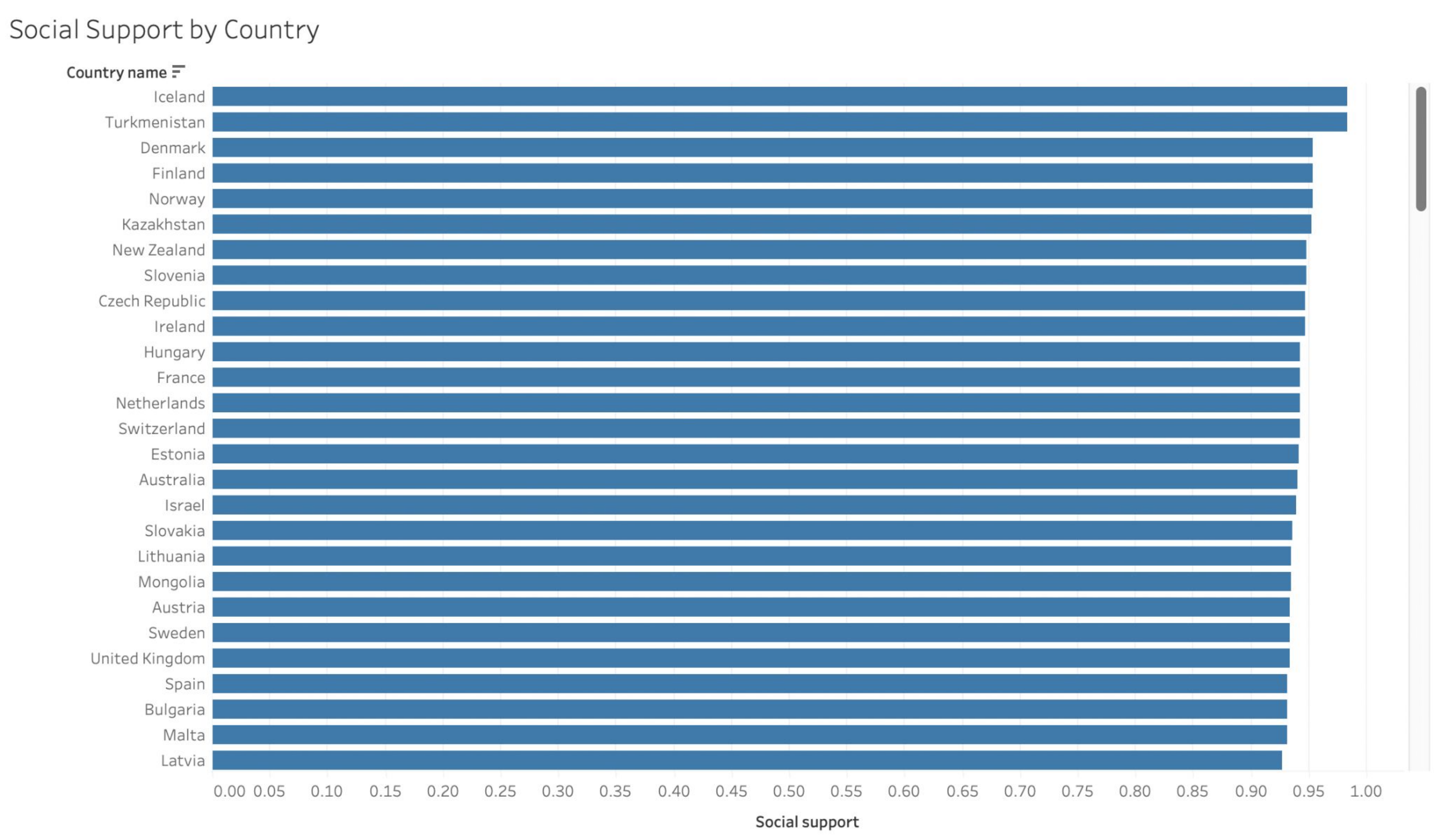
Methods for Visualizations - Tableau



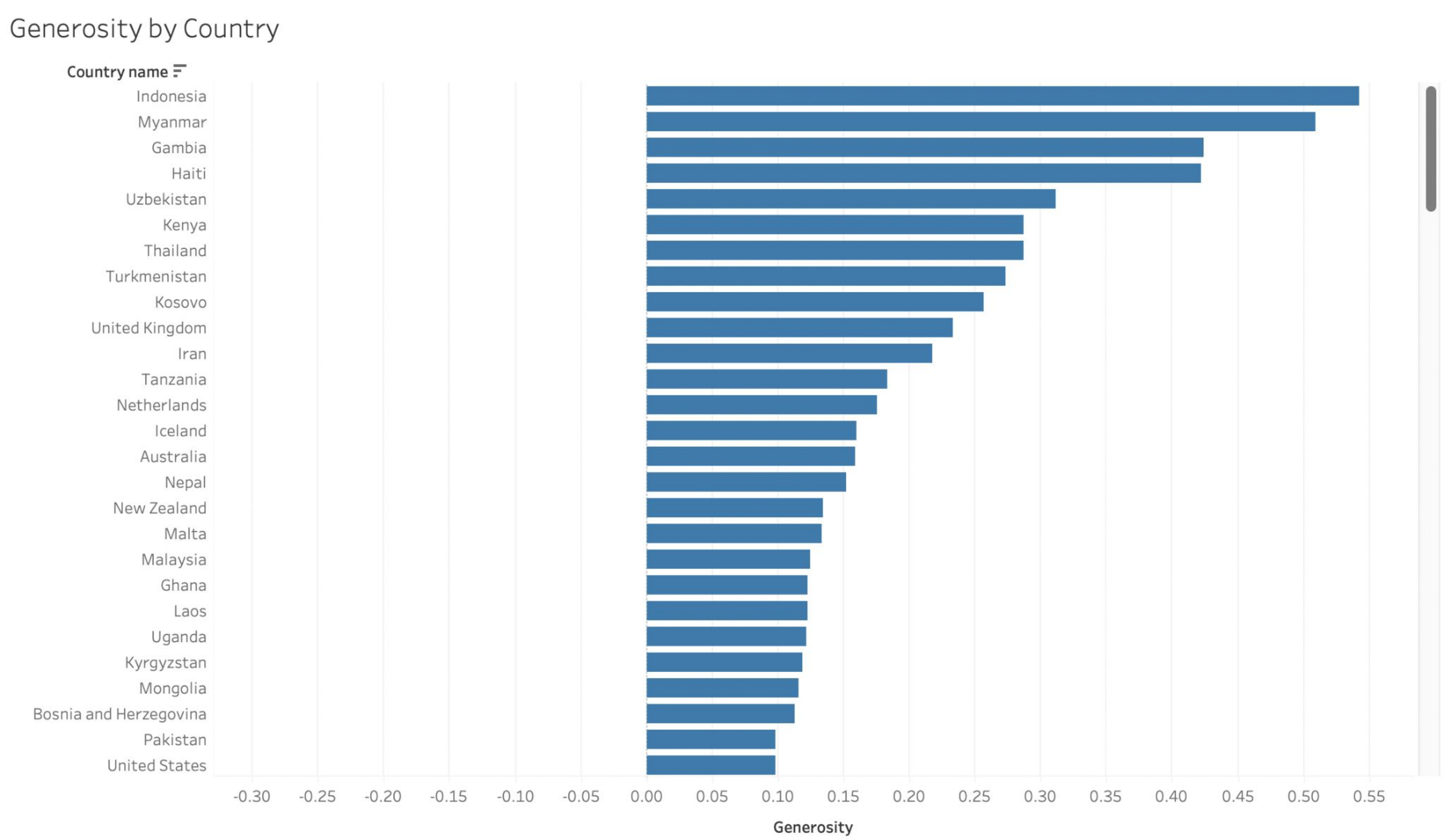
Methods for Visualizations - Tableau



Methods for Visualizations - Tableau


Social Support by Country

Methods for Visualizations - Tableau


Generosity by Country
Country name
Indonesia
Myanmar
Gambia
Haiti
Uzbekistan
Kenya
Thailand
Turkmenistan
Kosovo
United Kingdom
Iran
Tanzania
Netherlands
Iceland
Australia
Nepal
New Zealand
Malta
Malaysia
Ghana
Laos
Uganda
Kyrgyzstan
Mongolia
Bosnia and Herzegovina
Pakistan
United States
Generosity