

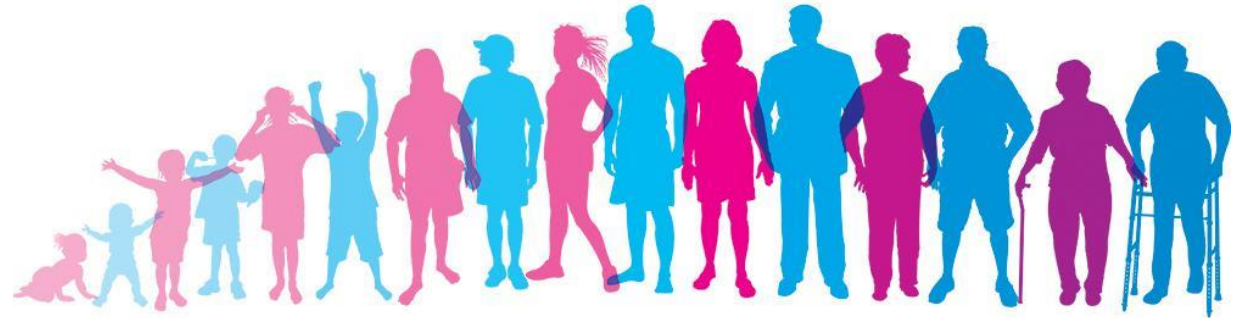


Life Expectancy Prediction

A PROJECT BY SOUMYADIP KUNDU
SPRINGBOARD DATA SCIENCES CAREER TRACK

Background

- Key Statistic for predicting population health
- Regularly monitored by WHO and health departments of all countries
- Various global characteristics affect its trend.
- Changed significantly in the past 20 years due to increased growth in health sector



Problem Statement

Assessing the relationship between the various predictive factors that affect an individual's life expectancy in 193 countries.



- Generated by The World Health Organization's Global Health Repository
- Downloaded through Kaggle
- 2938 rows and 22 columns

Dataset

Exploratory Data Analysis

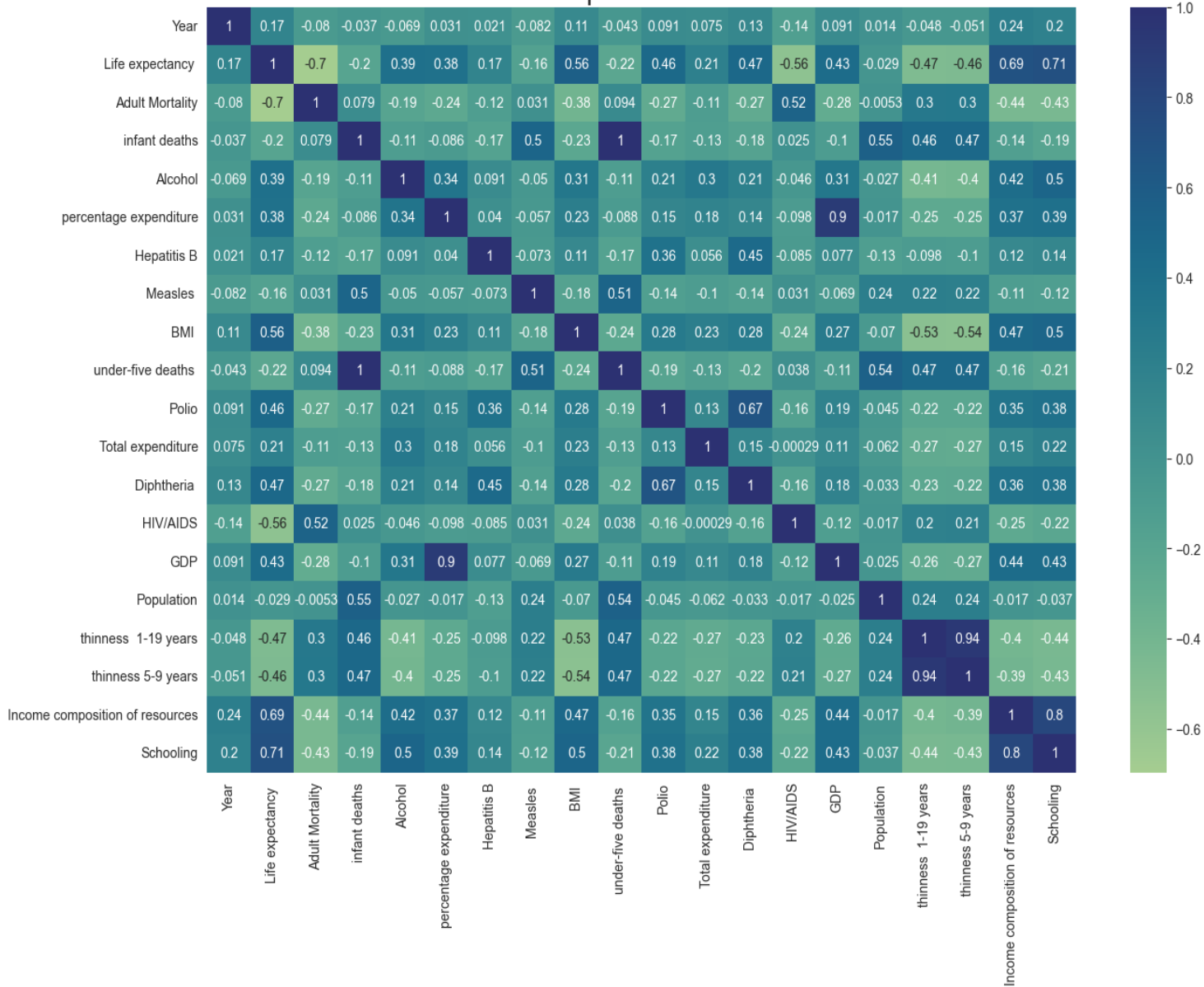


Stock Market Report

330.40	341.94	354.71	367.00	380.00
395.22	411.50	428.00	445.00	460.00
475.73	492.70	510.00	528.00	545.00
561.00	579.00	597.00	615.00	633.00
650.00	669.00	688.00	707.00	726.00
745.00	764.00	783.00	802.00	821.00
840.00	859.00	878.00	897.00	916.00
935.00	954.00	973.00	992.00	1011.00
1030.00	1049.00	1068.00	1087.00	1106.00
1125.00	1144.00	1163.00	1182.00	1201.00
1220.00	1239.00	1258.00	1277.00	1296.00
1311.00	1330.00	1349.00	1368.00	1387.00
1408.00	1427.00	1446.00	1465.00	1484.00
1503.00	1522.00	1541.00	1560.00	1579.00
1598.00	1617.00	1636.00	1655.00	1674.00
1693.00	1712.00	1731.00	1750.00	1769.00
1784.00	1803.00	1822.00	1841.00	1860.00
1879.00	1898.00	1917.00	1936.00	1955.00
1974.00	1993.00	2012.00	2031.00	2050.00



Correlation heatmap between different variables



Correlation Heatmap

Top Positive correlations:

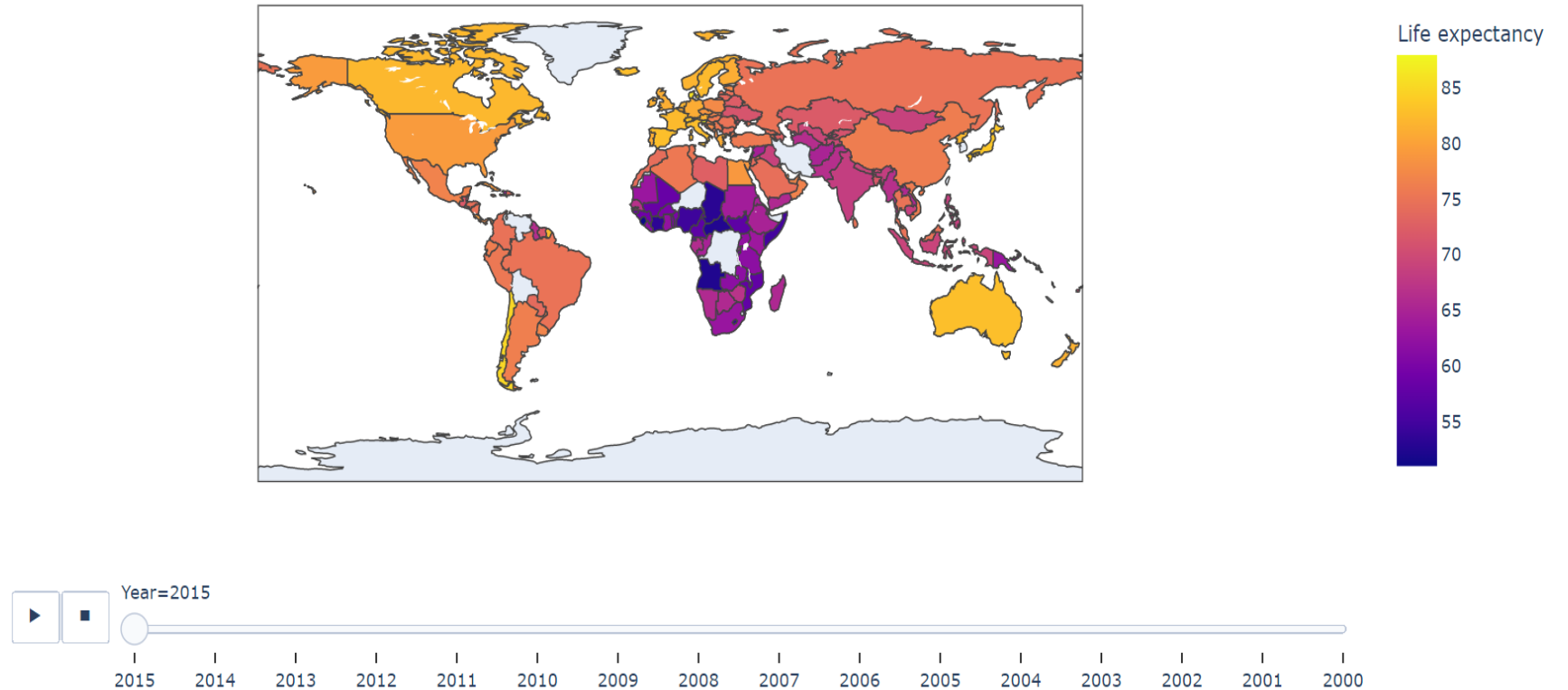
- Schooling
- Income

Top Negative correlations:

- Adult Mortality
- HIV/AIDS

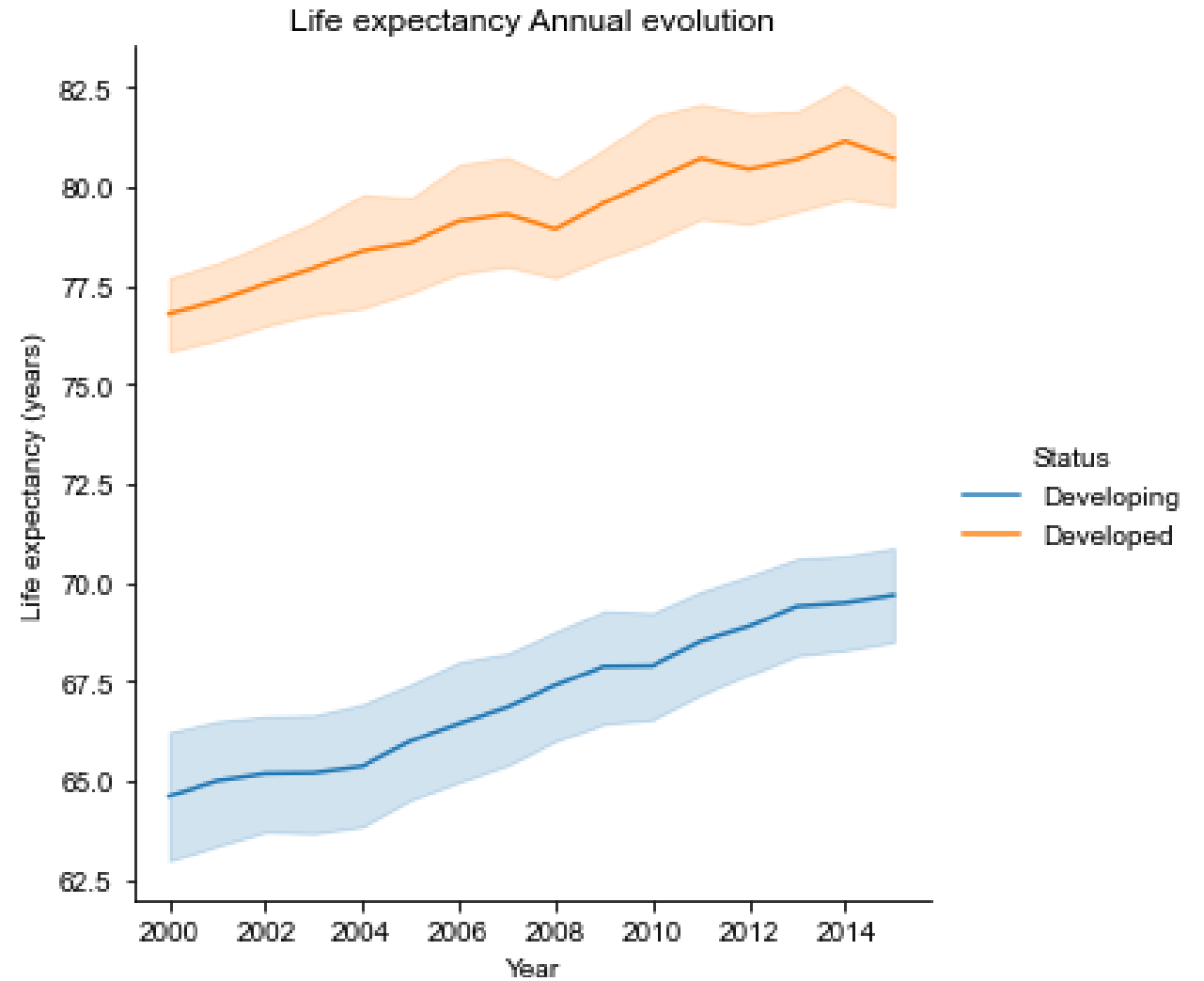
Chloropleth map of Life Expectancy

Life Expectancy over the years

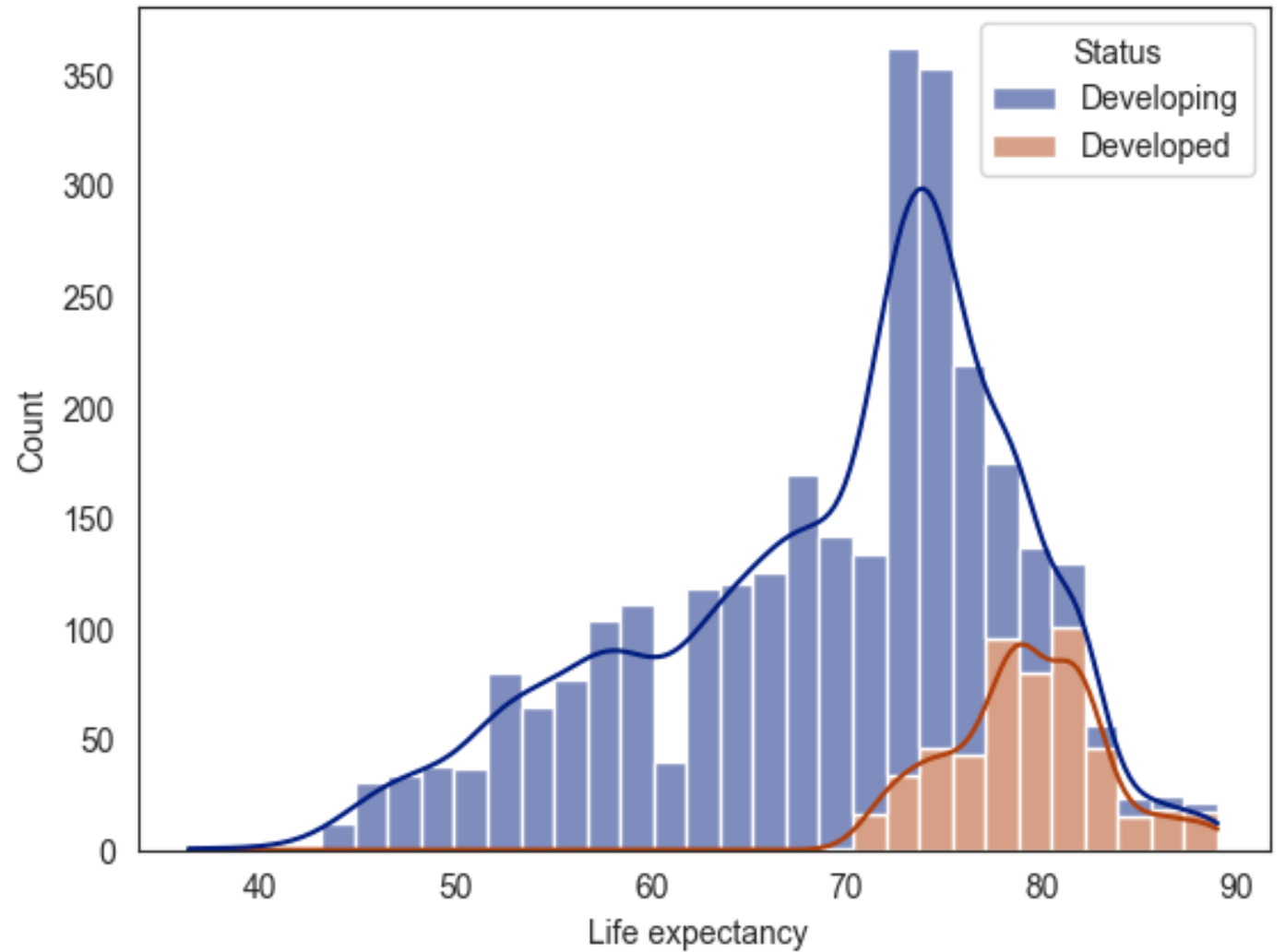


Animated version: <https://skundu01.github.io/>

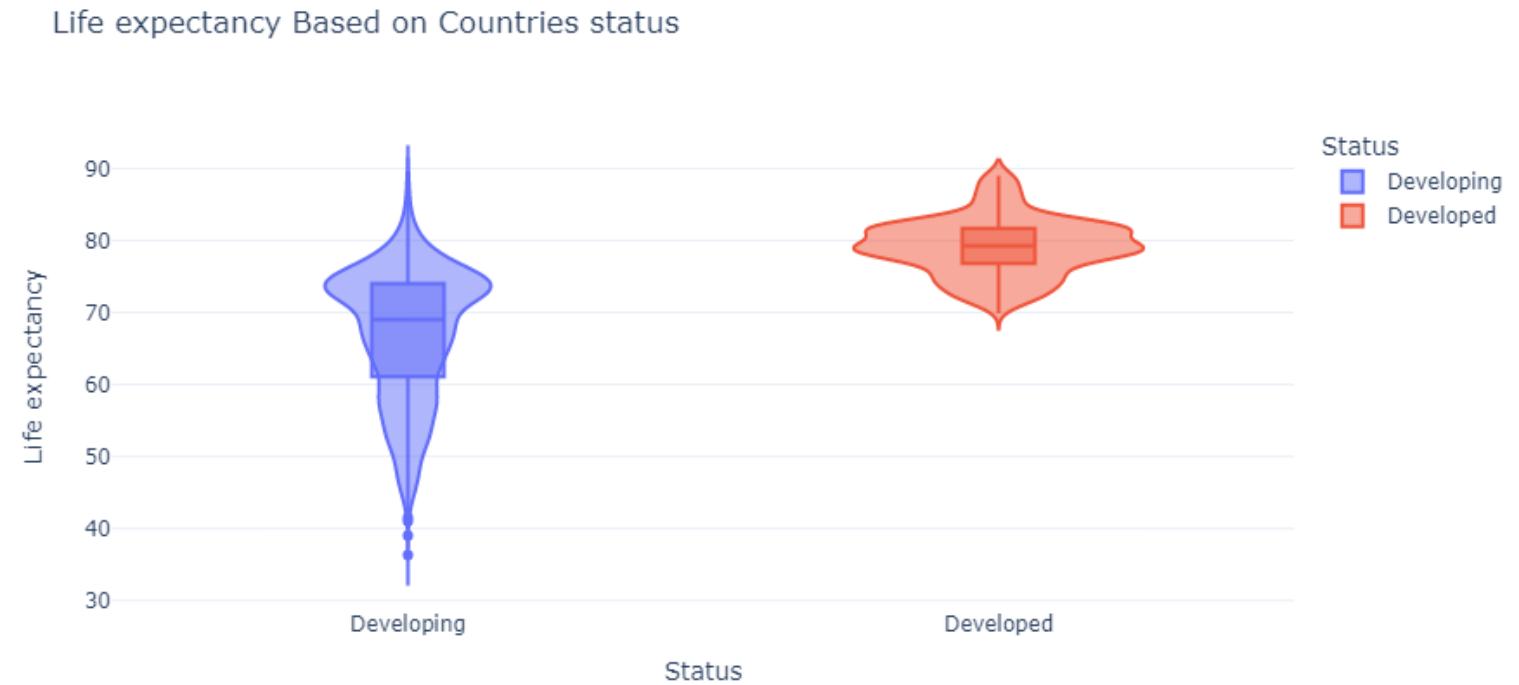
Annual Evolution of Life Expectancy



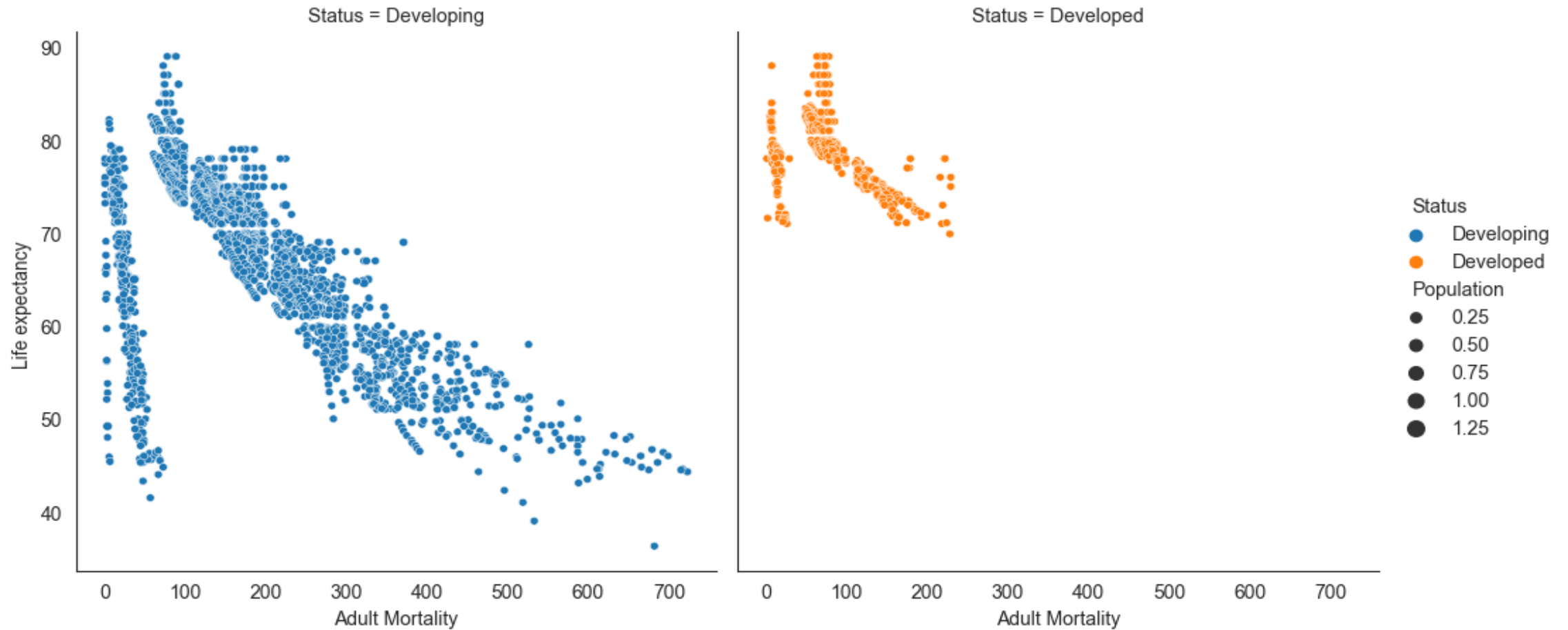
Distribution of Life Expectancy in Developing vs Developed Countries



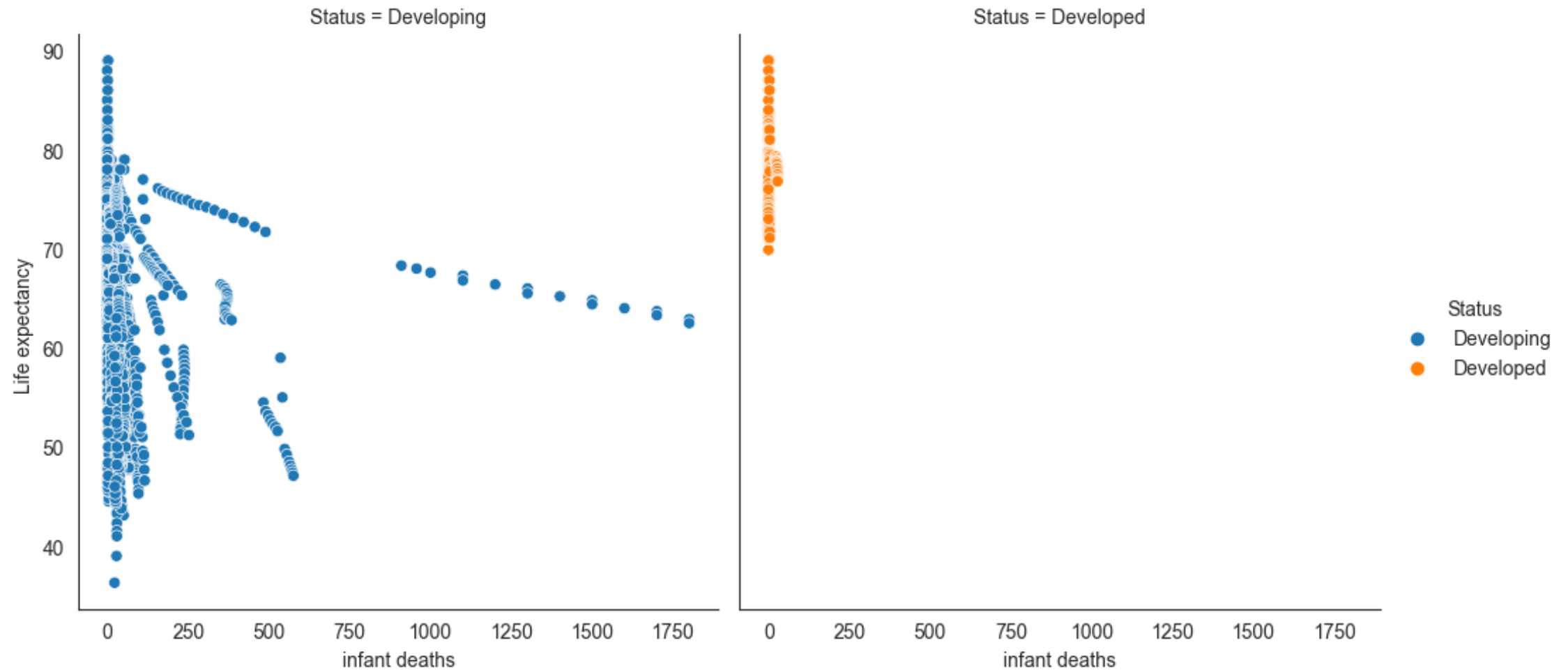
Violin plot of Life expectancy based on status



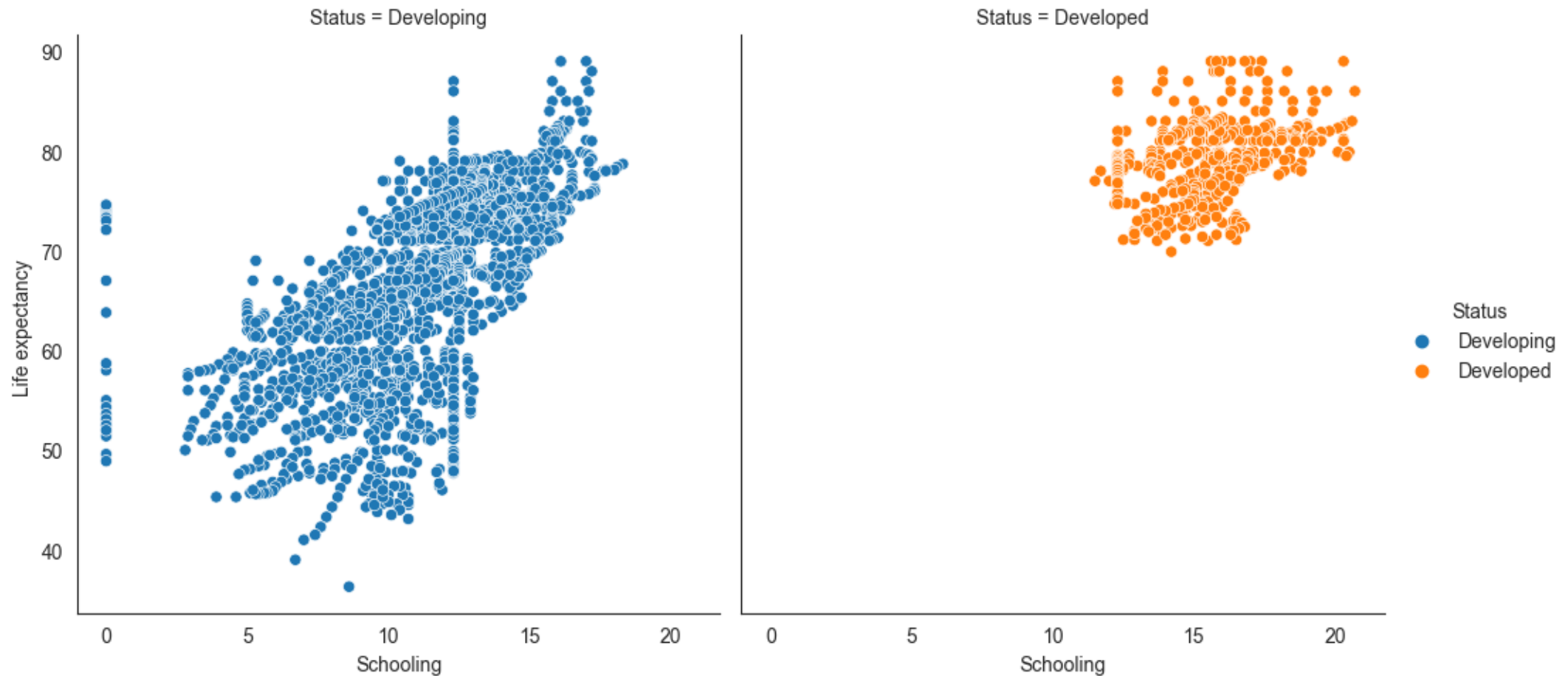
Life Expectancy vs Adult Mortality



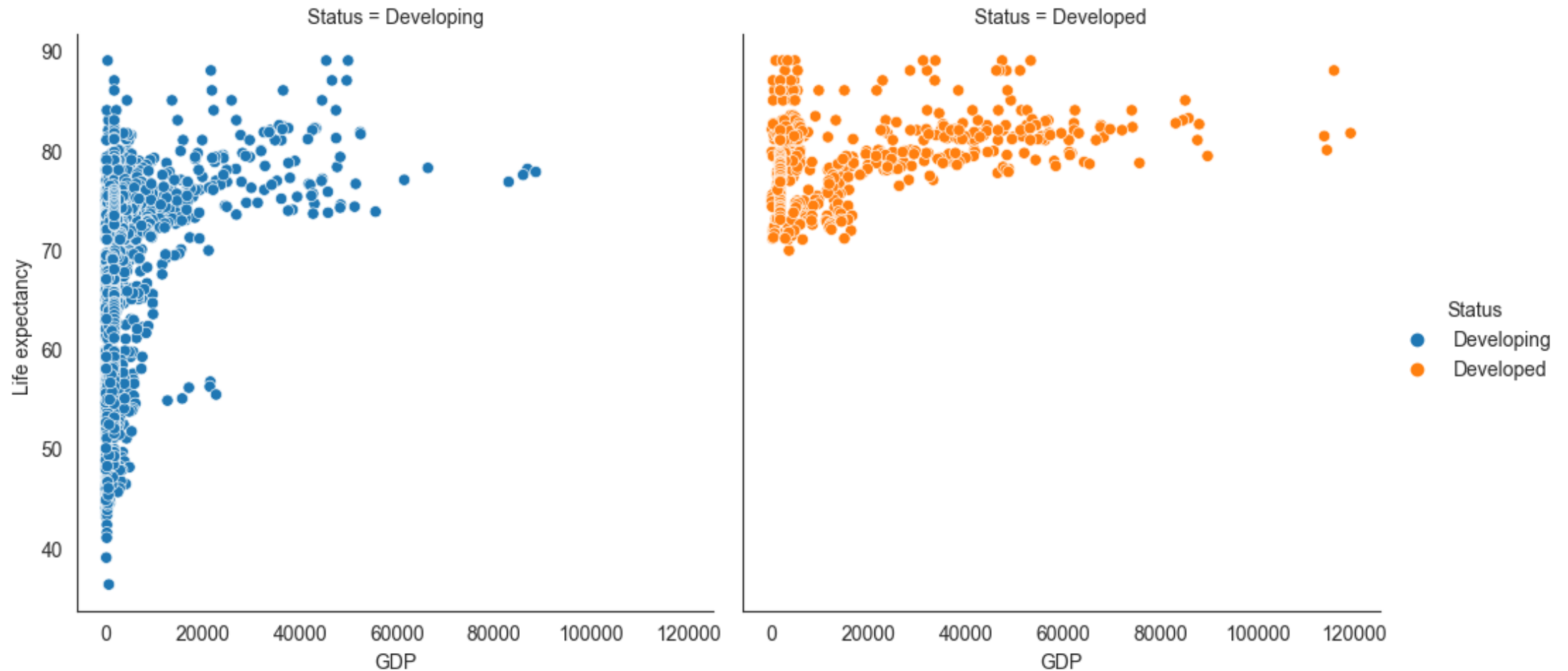
Life Expectancy vs Infant Deaths



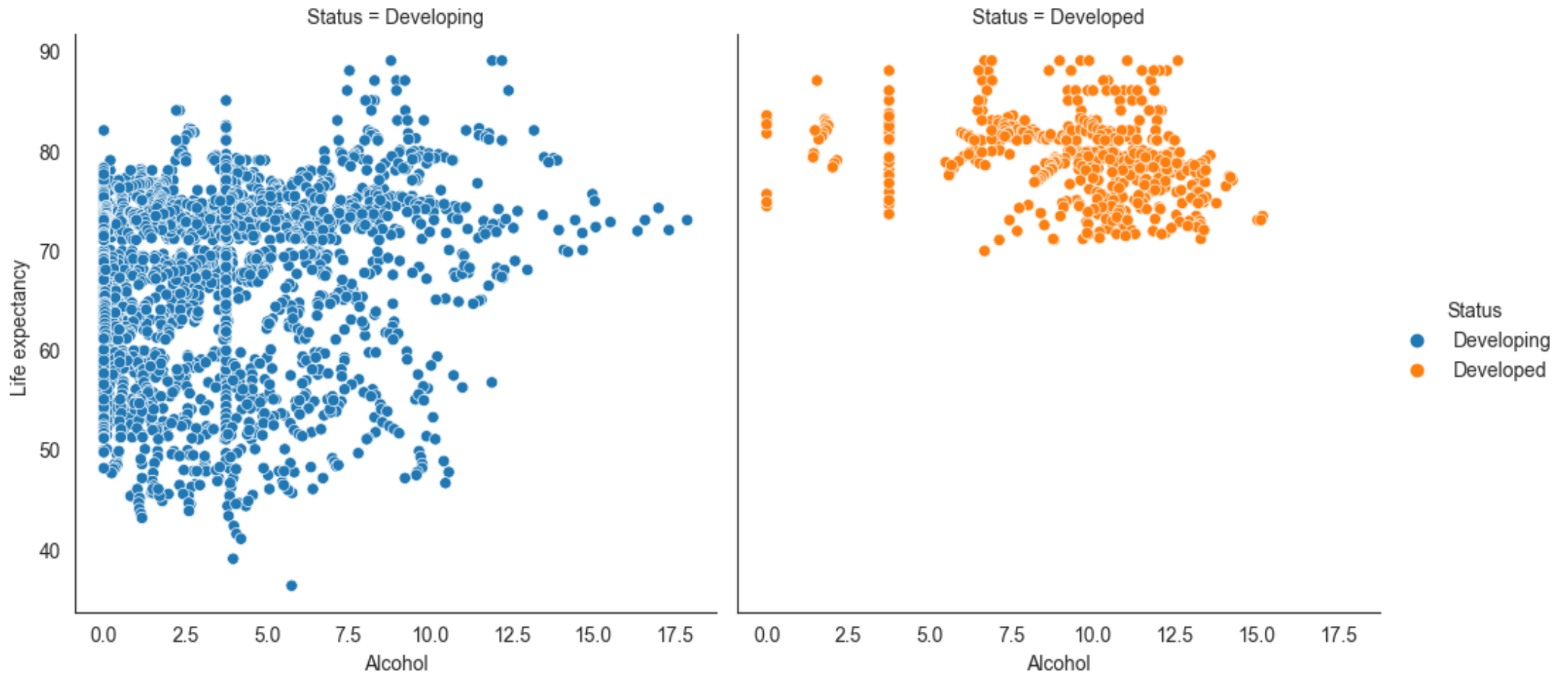
Life Expectancy vs Schooling



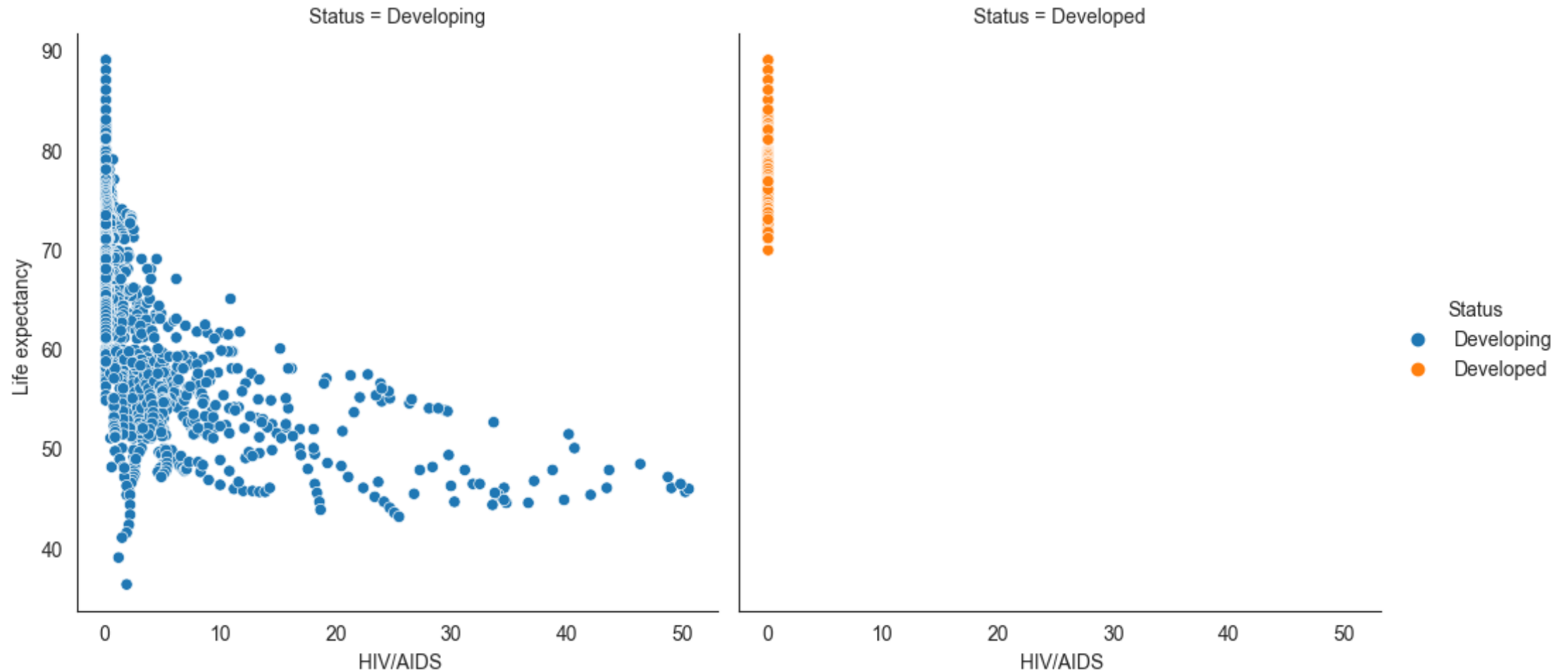
Life Expectancy vs GDP



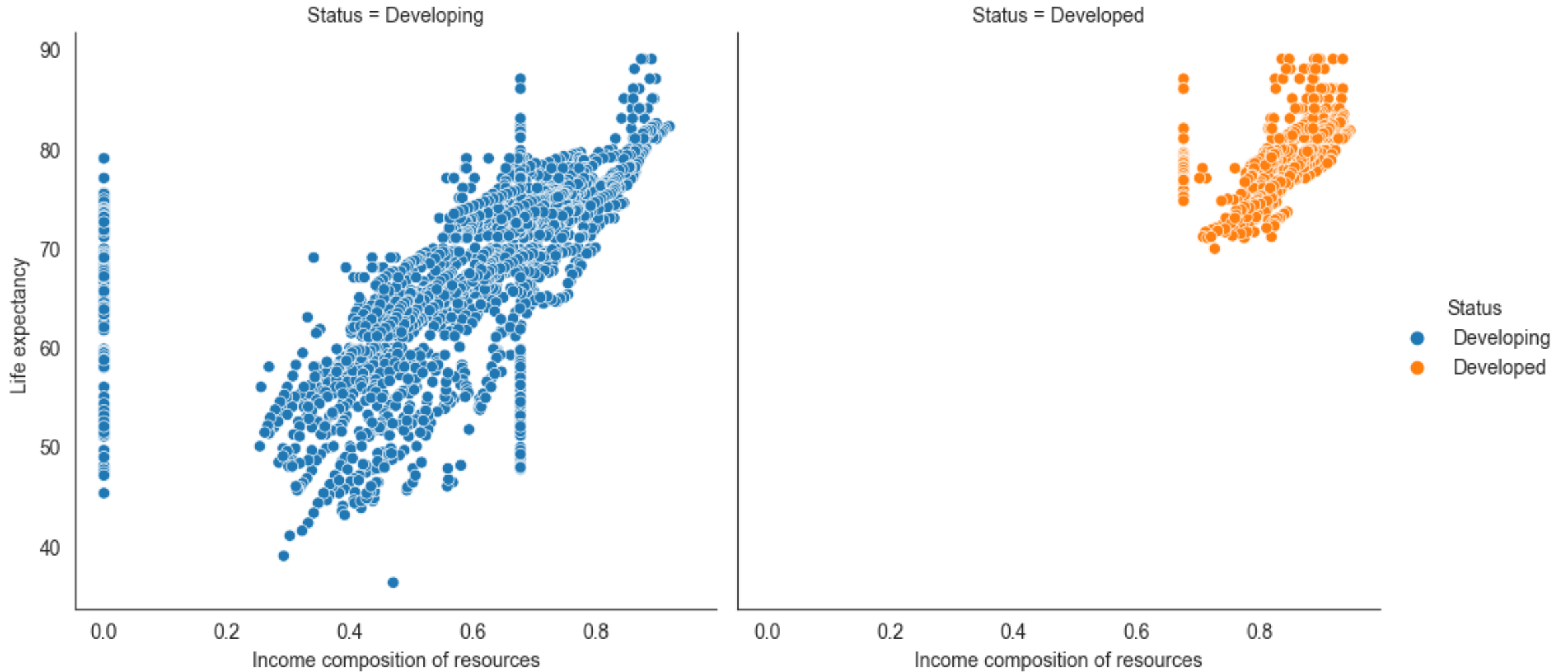
Life expectancy vs Alcohol Consumption



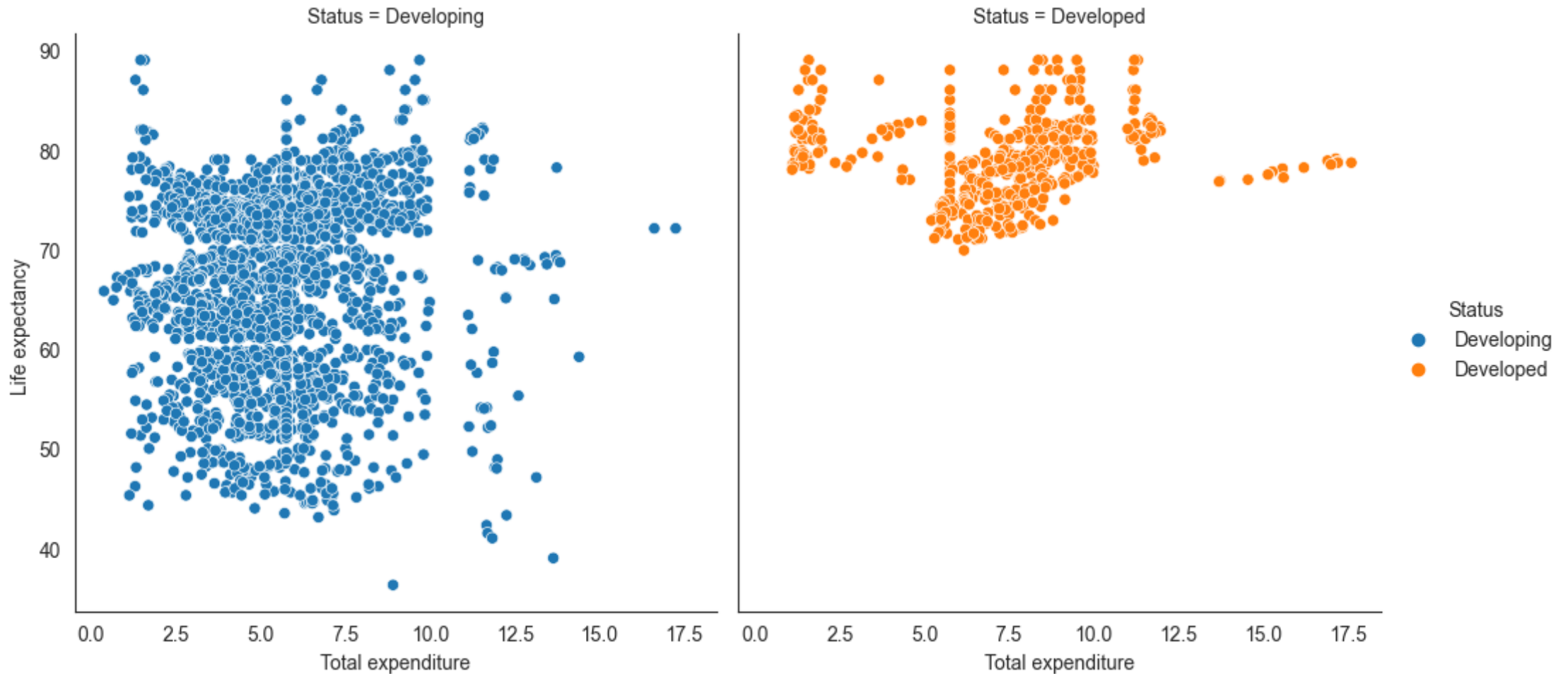
Life Expectancy vs HIV/AIDS



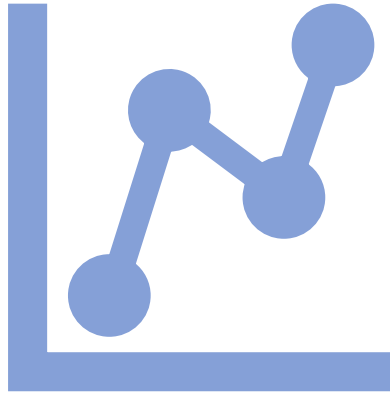
Life Expectancy vs Total Income



Life Expectancy vs Total Expenditure



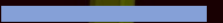
Data Cleaning and Preprocessing



- 3 columns contained about 25% missing data
- Simple Imputer used to impute values with the median values
- Numerical values normalized by MinMaxScaler
- Categorical values encoded by Label Encoder
- 80-20 train-test split created with shuffle



Data Modeling

The background of the slide is a dark, textured surface with a complex pattern of overlapping, semi-transparent yellow and green lines and circles. The lines form a grid-like structure, while the circles are scattered throughout, some appearing as nodes in a network. The overall effect is a modern, digital aesthetic.



Models used

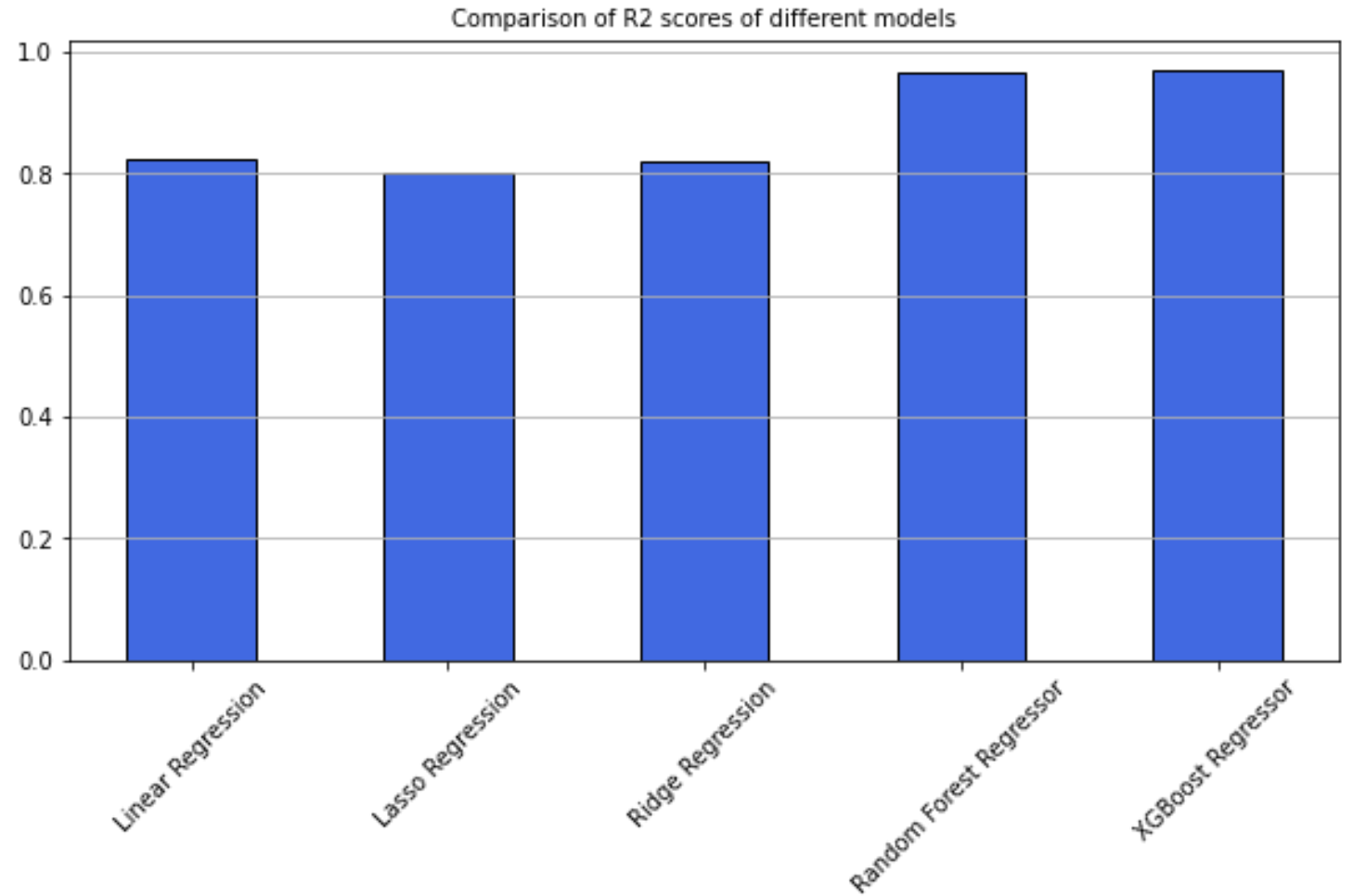
- Linear Regression
- Lasso Regression
- Ridge Regression
- Random Forest Regressor
- XG Boost Regressor

Model Performance Comparison

Model Name	RMSE	R2
Linear Regression	0.074506	0.822111
Lasso Regression	0.078636	0.801845
Ridge Regression	0.074777	0.820814
Random Forest Regression	0.031716	0.967766
XG Boost Regression	0.030095	0.970976

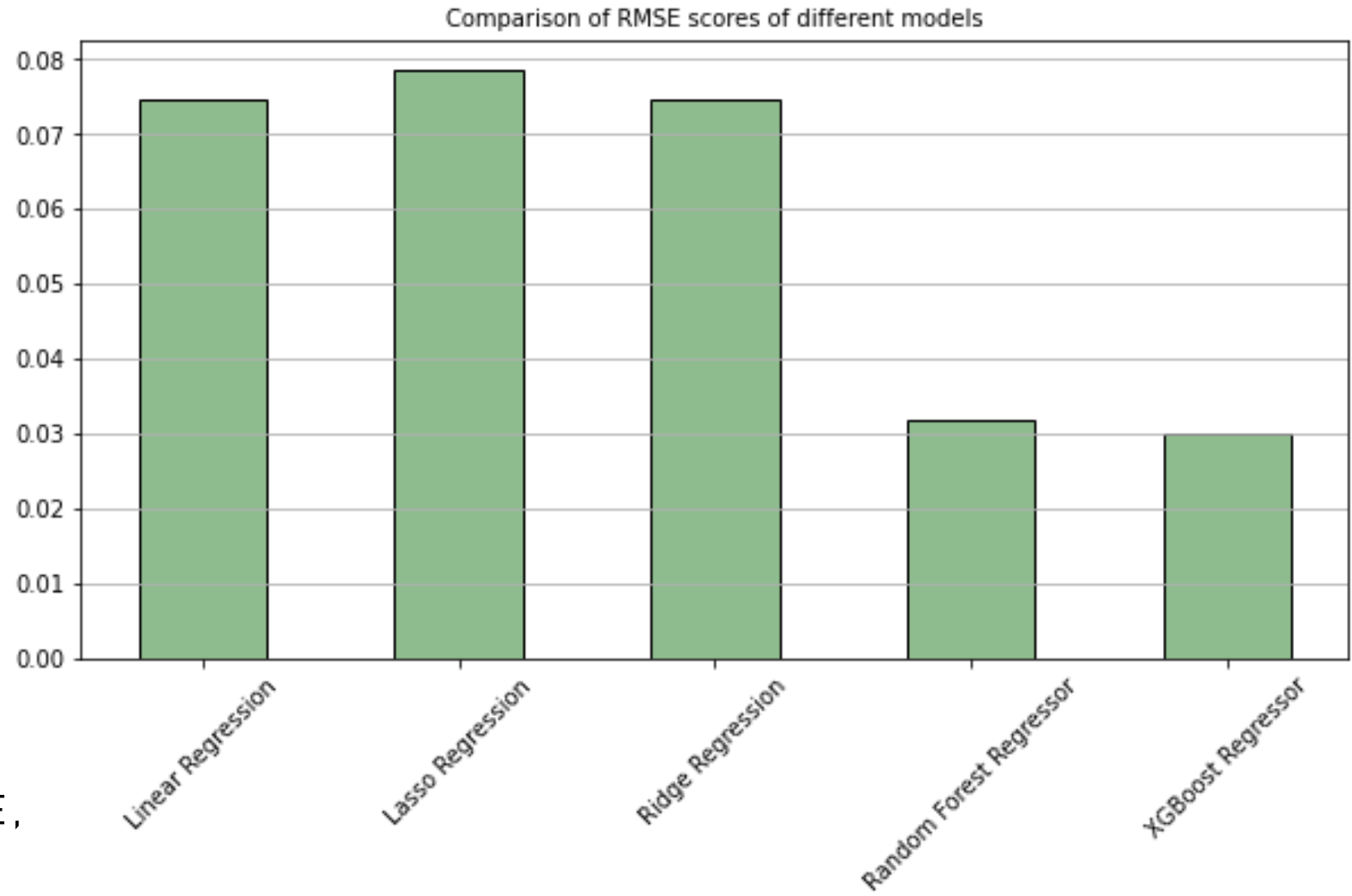
Model Performance Metrics

HIGHER THE R2 SCORE,
BETTER THE MODEL



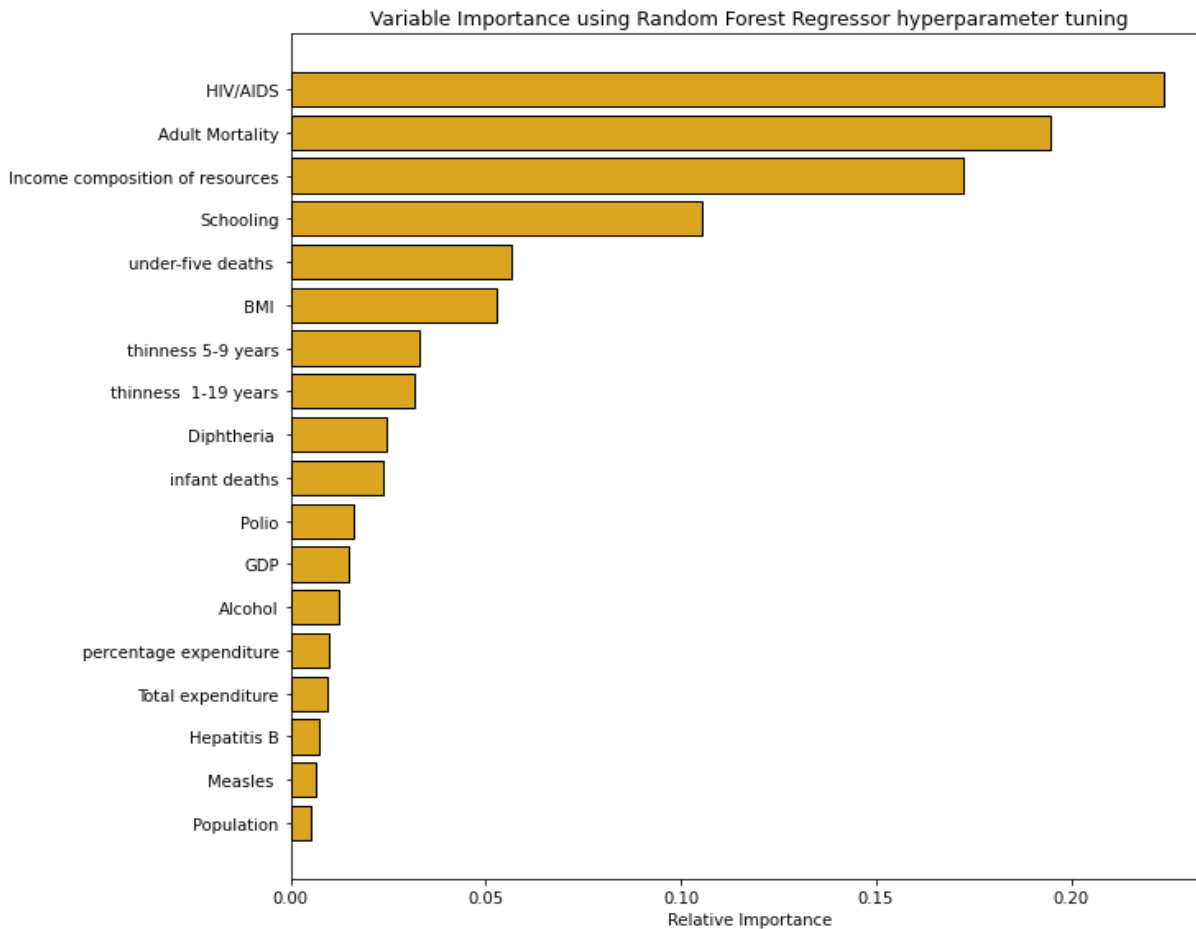
Model Performance Metrics contd...

LOWER THE RMSE SCORE,
BETTER THE MODEL

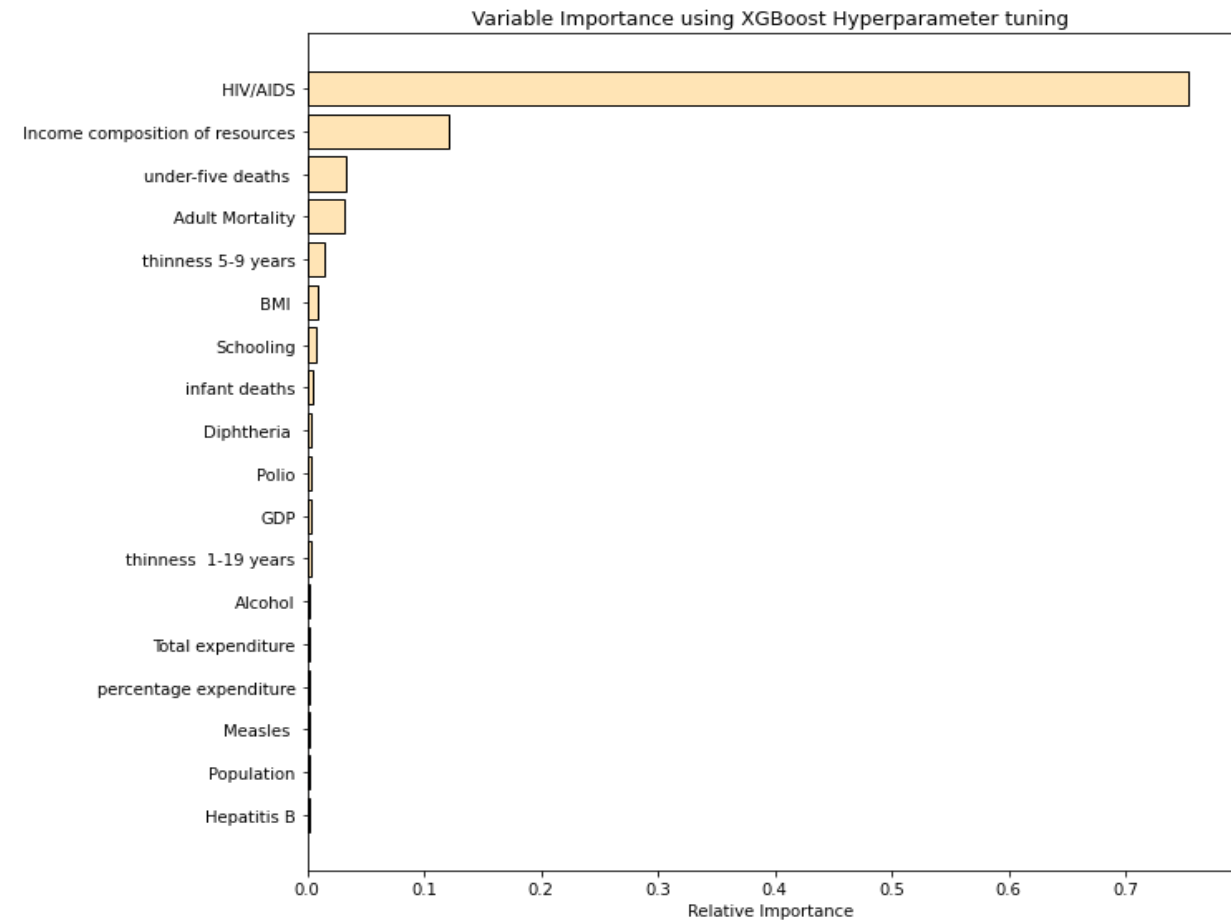


Hyperparameter tuning and Feature Selection

RANDOM FOREST



XG BOOST



Conclusion

Most important predictors according to Random Forest

- HIV/AIDS
- Adult Mortality
- Income
- Schooling

Most important predictors according to XG Boost

- HIV/AIDS
- Income
- Infant deaths
- Adult Mortality

Acknowledgements

- Kenneth Gil-Pasquel (Mentor)
- Springboard team
- World Health Organization
- Kaggle
- Cover images – Google images