

Investigating the Economic Health of the United Nations

Peter Ilhardt

November 5, 2019

There could be trouble ahead for Saudi Arabia's economy

By Varsha Koduvayur for CNN Business Perspectives

Updated 11:08 AM ET, Sun November 3, 2019

The Post's View • Opinion

The economy is in good shape, but two big clouds hang over it

BUSINESS NEWS OCTOBER 17, 2019 / 4:17 PM / 17 DAYS AGO

China's GDP growth to grind to 27-1/2 year low as tariffs hit demand

Economy | Market Outlook

U.S. GDP Growth - Surprising, A Little, On The Upside

Nov. 2, 2019 3:03 PM ET | 1 comment | 1 Like | Includes: BAPR, BAUG, BIBL, BJUL, BJUN, BOCT, CHGX, CRF, ...

How Asia transformed from the poorest continent in the world into a global economic powerhouse

The New York Times

Economy Grew at 1.9% Rate in Quarter, Hit by Trade Fight and Global Weakness

Economics

Taiwan Bucks Asia Slowdown With Fastest Growth in 5 Quarters

India struggles but Bangladesh's GDP rides high on manufacturing, export boom

Article 1 of the UN Charter: “To achieve international co-operation in solving international problems of an economic, social, cultural, or humanitarian character...”

What Factors Most Influence a Nation's Wealth?

- **Sources:** United Nations Statistics Division (UNdata), Wikipedia
- **Data:** Economic and social aggregate metrics for 182 nations as measured in 2005, 2010, and 2015
- **Aim:** Predict GDP per capita in 2015 and determine most predictive metrics



Modeling Workflow

Load and Clean Data

- Web scraping and querying
- Normalization, transformation
- Sparse features and imputation

Exploratory Data Analysis

- Univariate and bivariate relationships
- Principal components analysis

Build Models

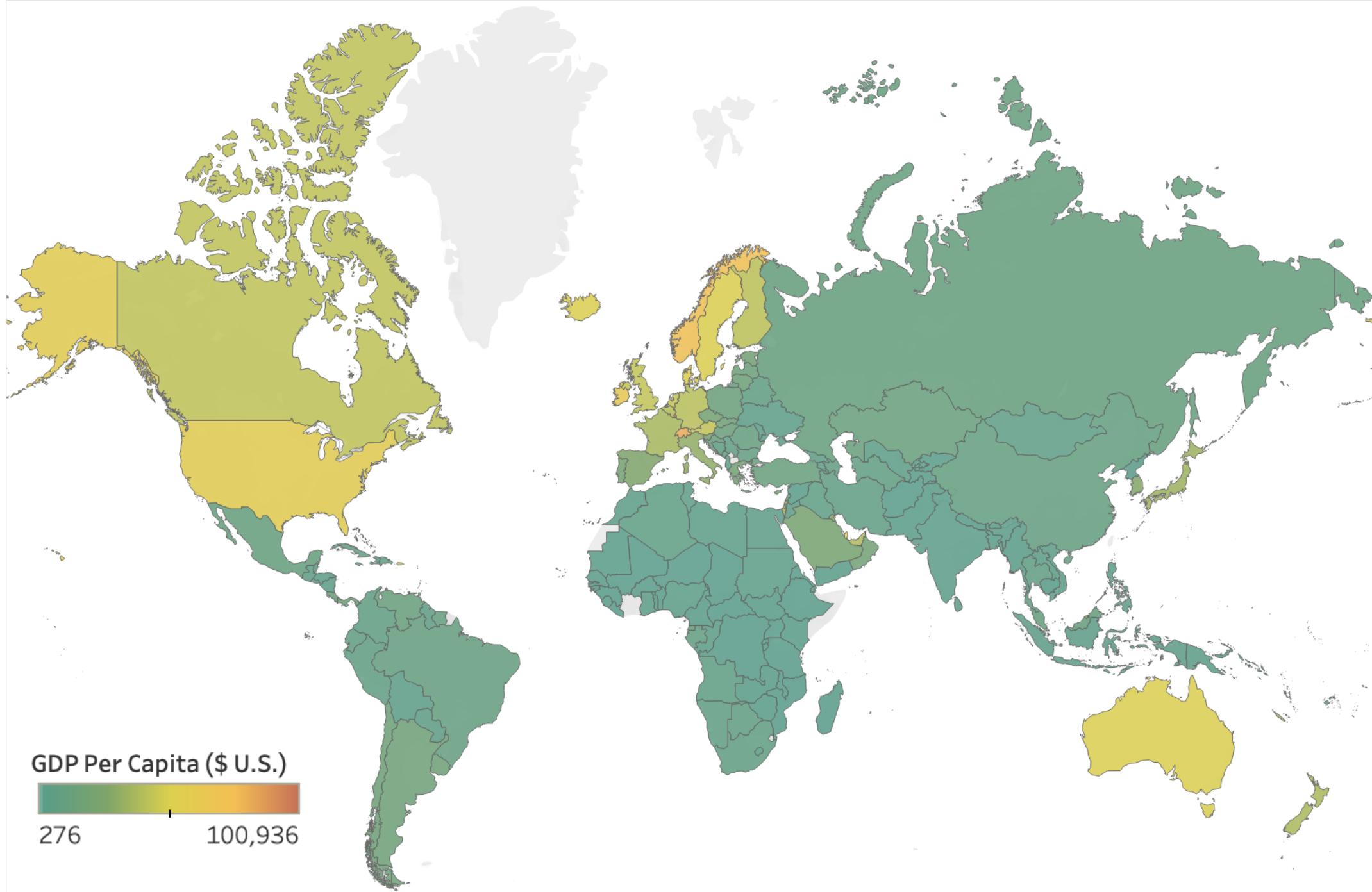
- Multiple linear regression
- Decision trees and random forests
- Cross-validation

Optimize Models

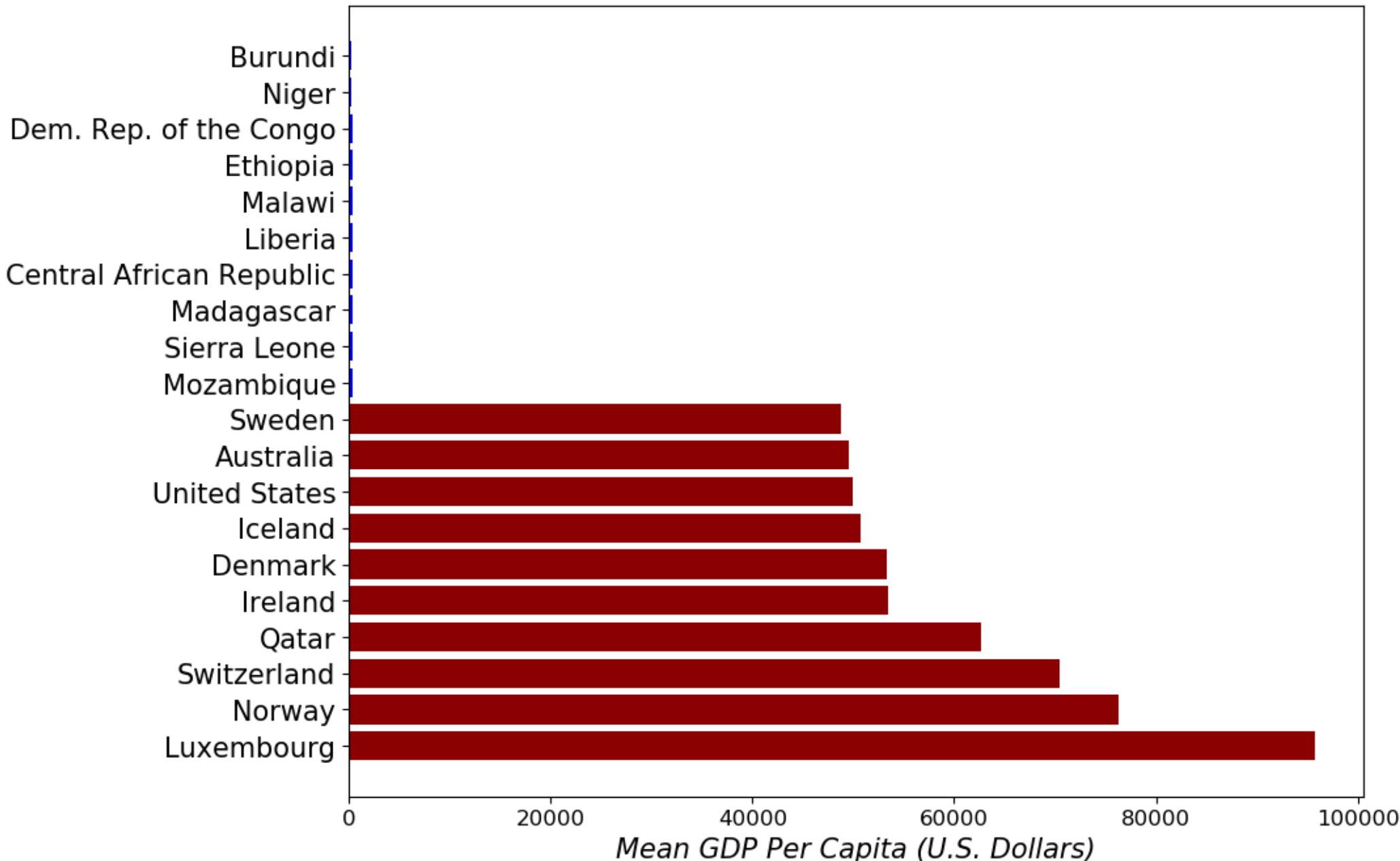
- Regularization
- Feature selection
- Parameter tuning

Evaluate and Interpret

- Evaluate model fit and predictions
- Interpret significant features

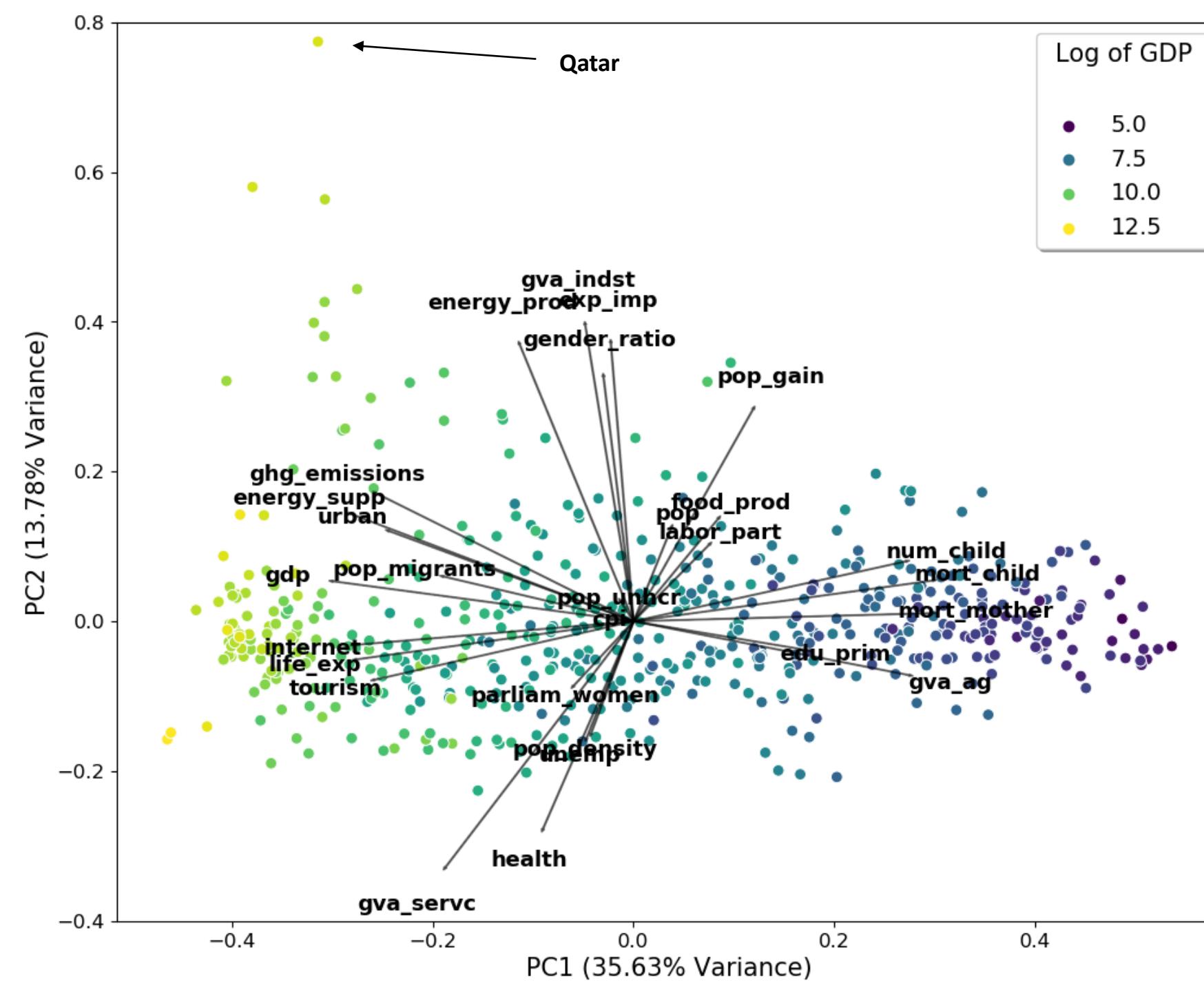


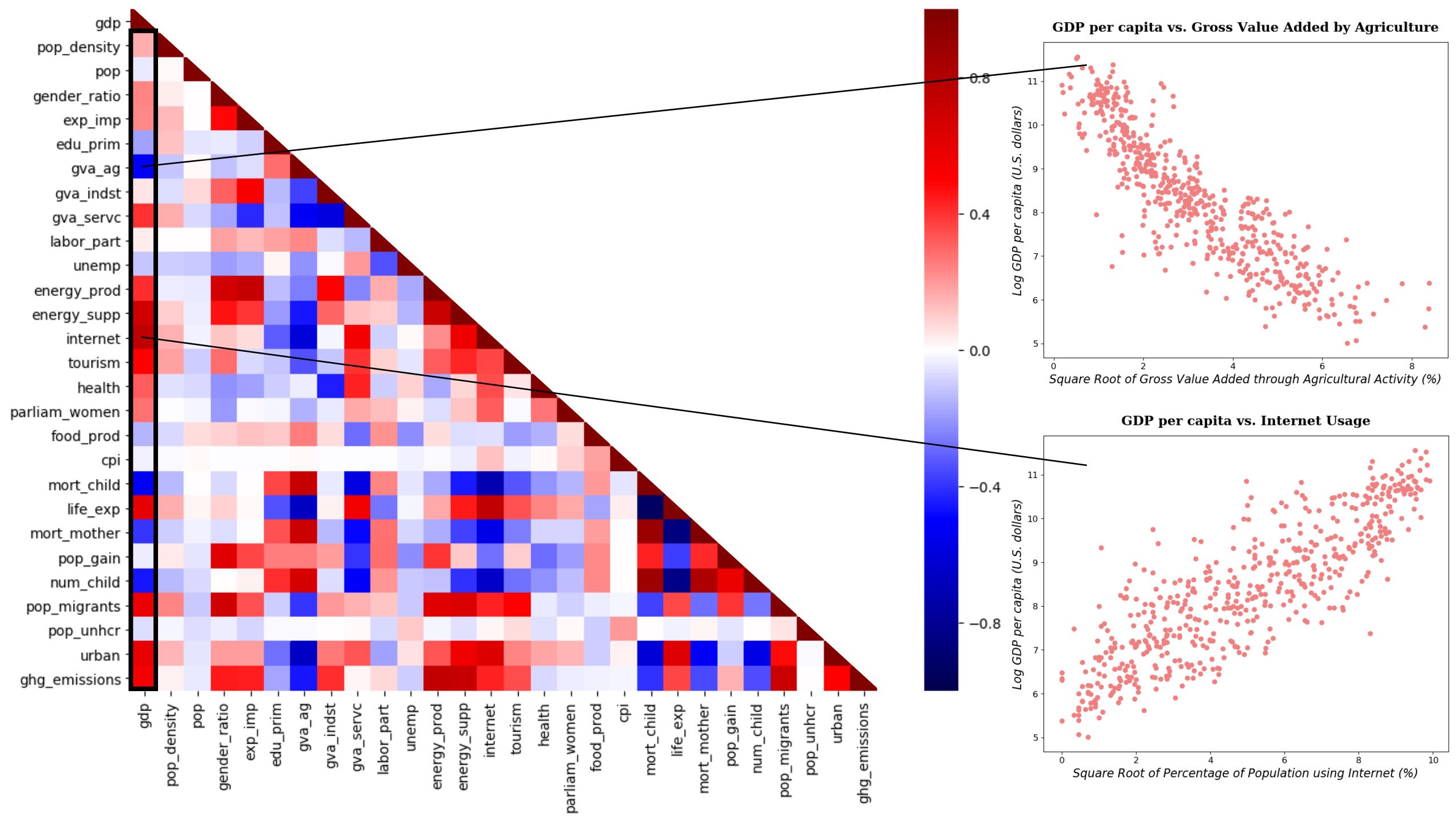
Countries with the highest and lowest mean GDP per capita



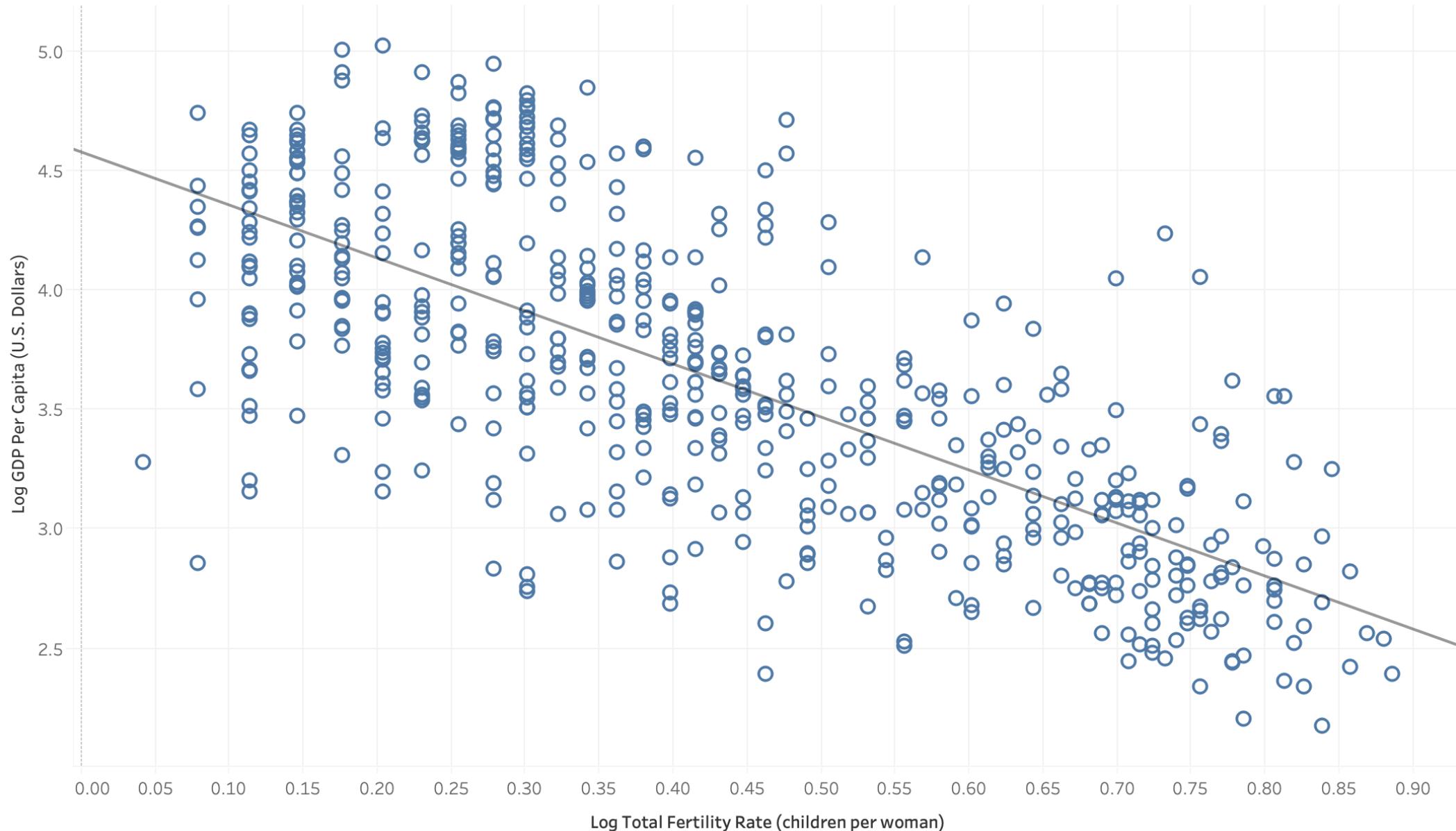
PCA

- Reduces dimensions by generating new features that capture variance
- Allows us to “project” the data to a lower-dimensional space to visualize sample-feature relationships
- GDP appears correlated with features such as life expectancy and urbanization





Confounding Variables



Confounding Variables



Multiple Linear Regression

$$GDP = b_0 + b_1(Urban) + b_2(Population)$$

Fit a model using 2005 and 2010 UN data for all countries:

$$GDP = -12,900 + 435.76(Urban) - 1.17(Population)$$

Predict the GDP of Australia in 2015 (*Urban* = 85.7, *Population* = 23.8):

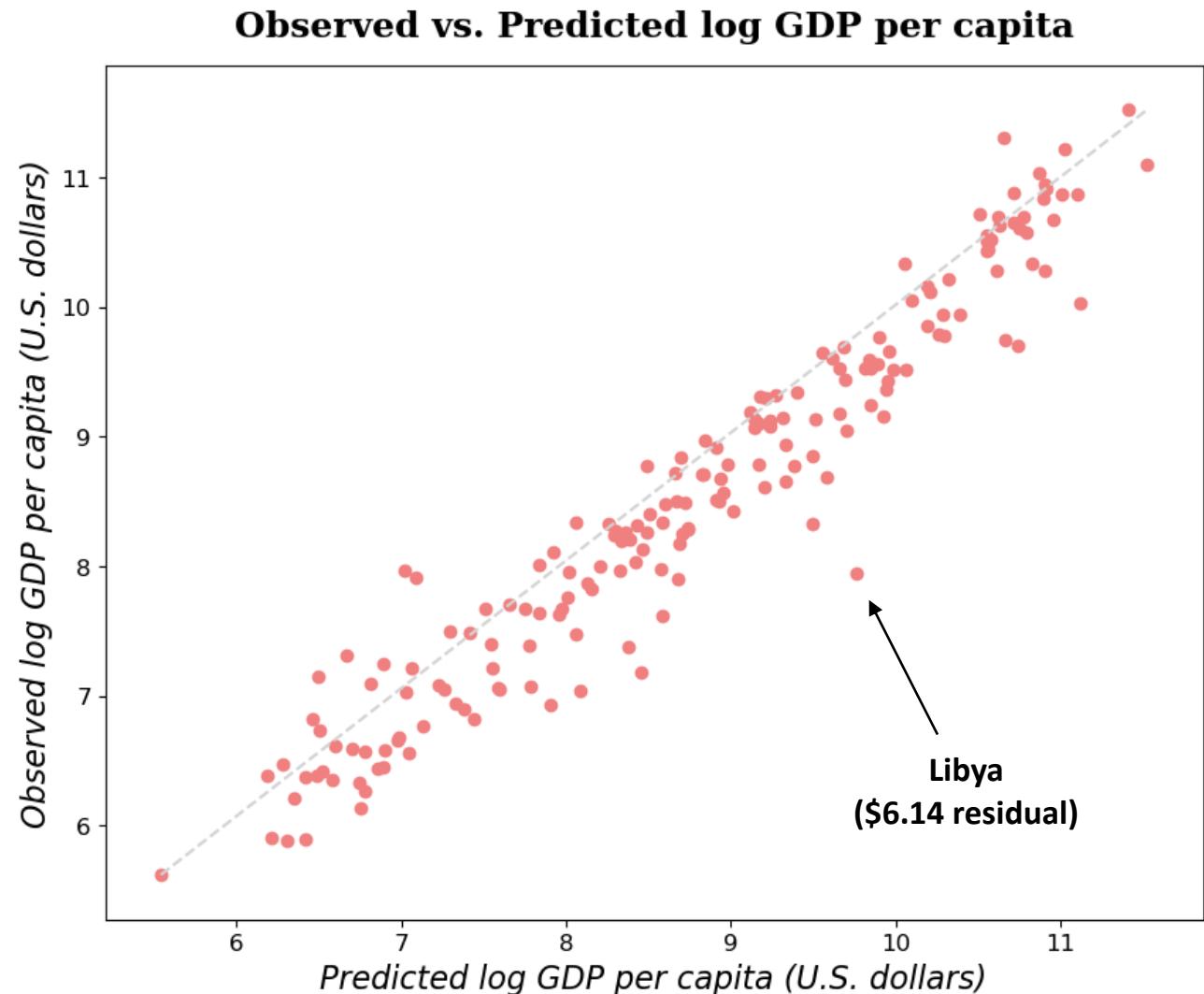
$$\begin{aligned} GDP &= -12,900 + 435.76(85.7) - 1.17(23.8) \\ &= \$24,417 \text{ per capita} \end{aligned}$$

Compare to actual GDP of Australia in 2015 (\$52,388):

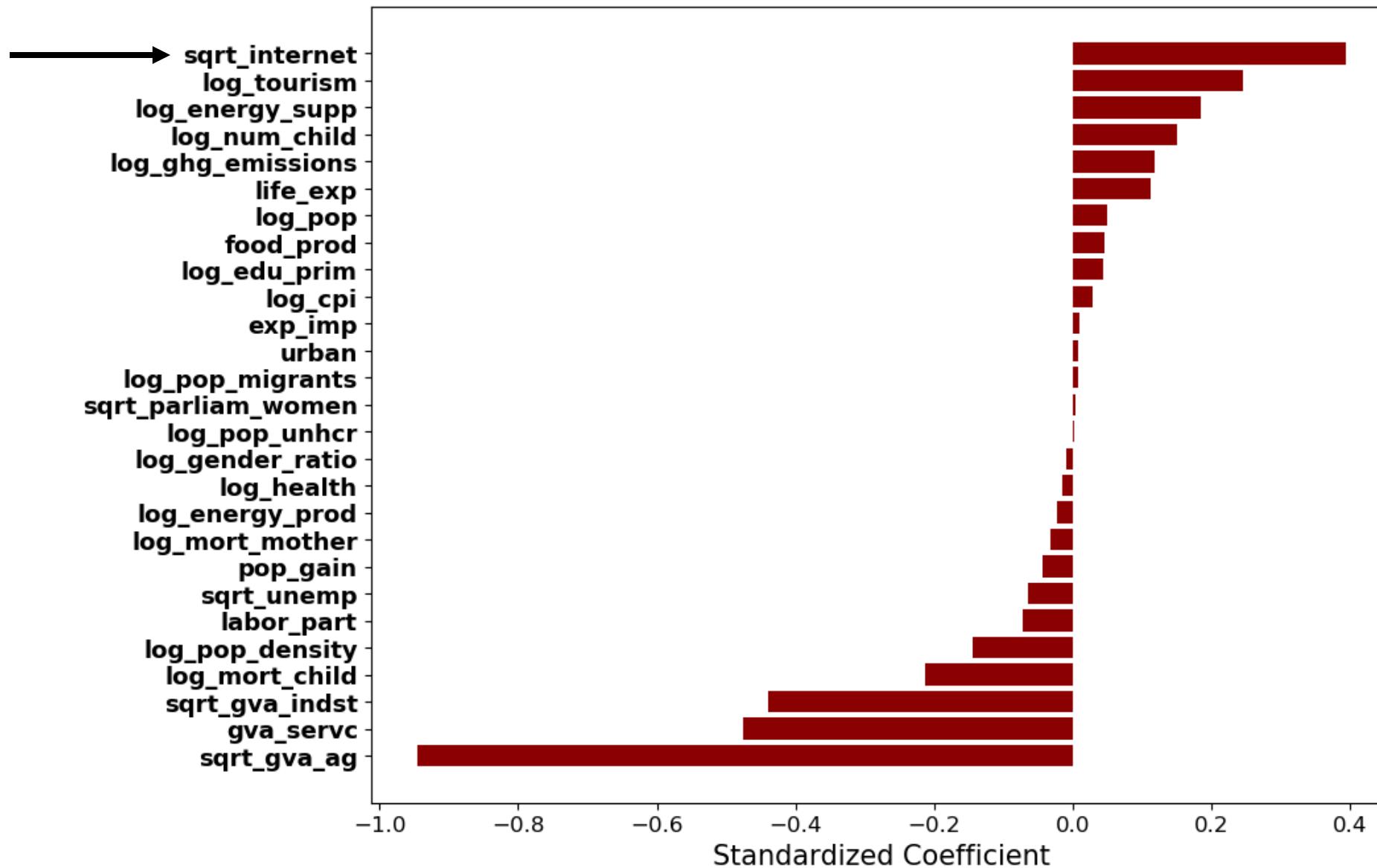
$$Error = \$52,388 - \$24,417 = \$27,971$$

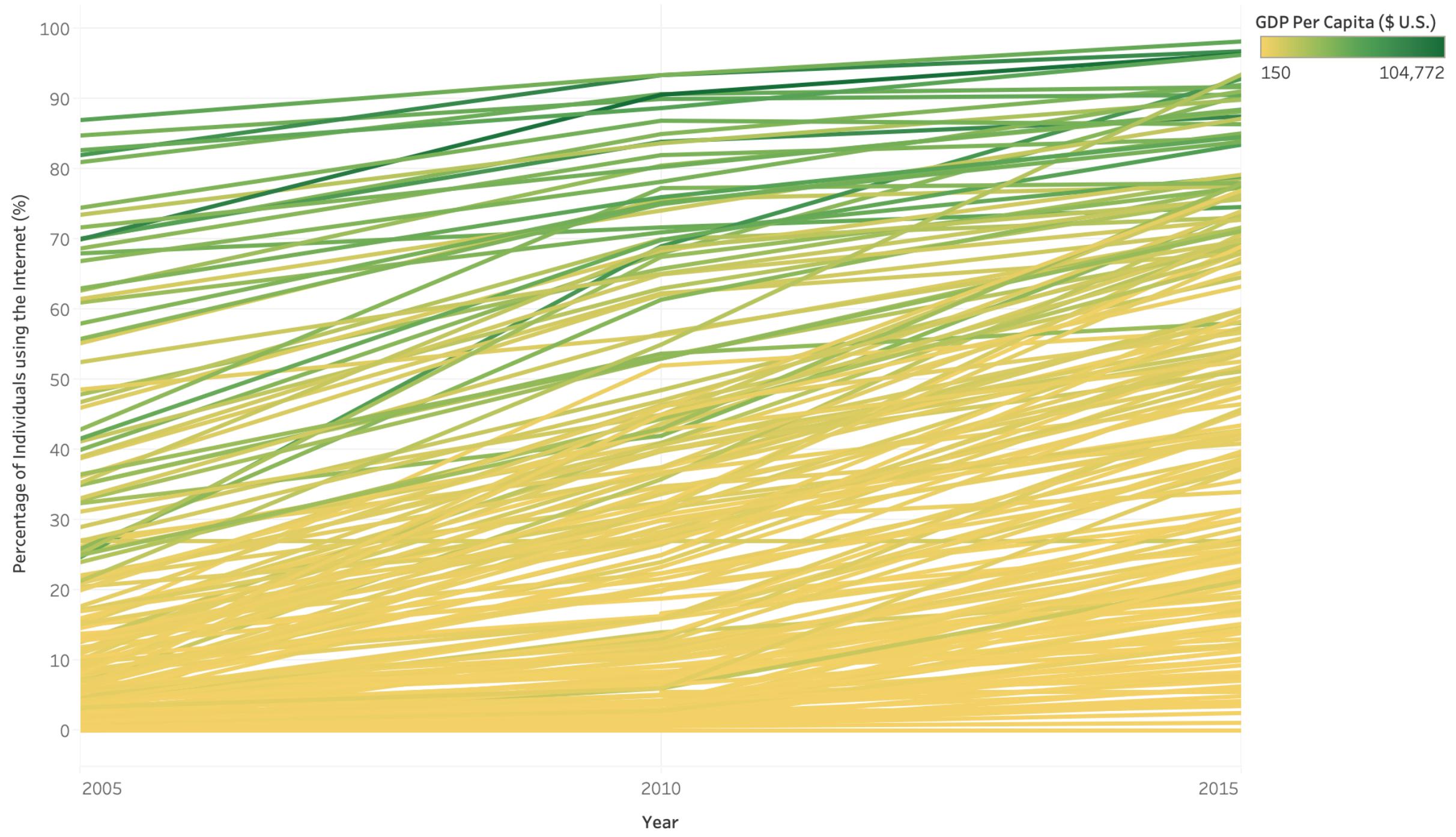
Predicted ~90% of Variance in 2015 Data

- Best Predictive Model had R^2 value of 0.905 on test data (0.953 on training data)
- Typical deviation of $\sim \$1.50$ U.S.
- Accurately reproduced GDP
- Best model kept 14 predictors and determined 13 were not significant



Strongest predictors include social metrics

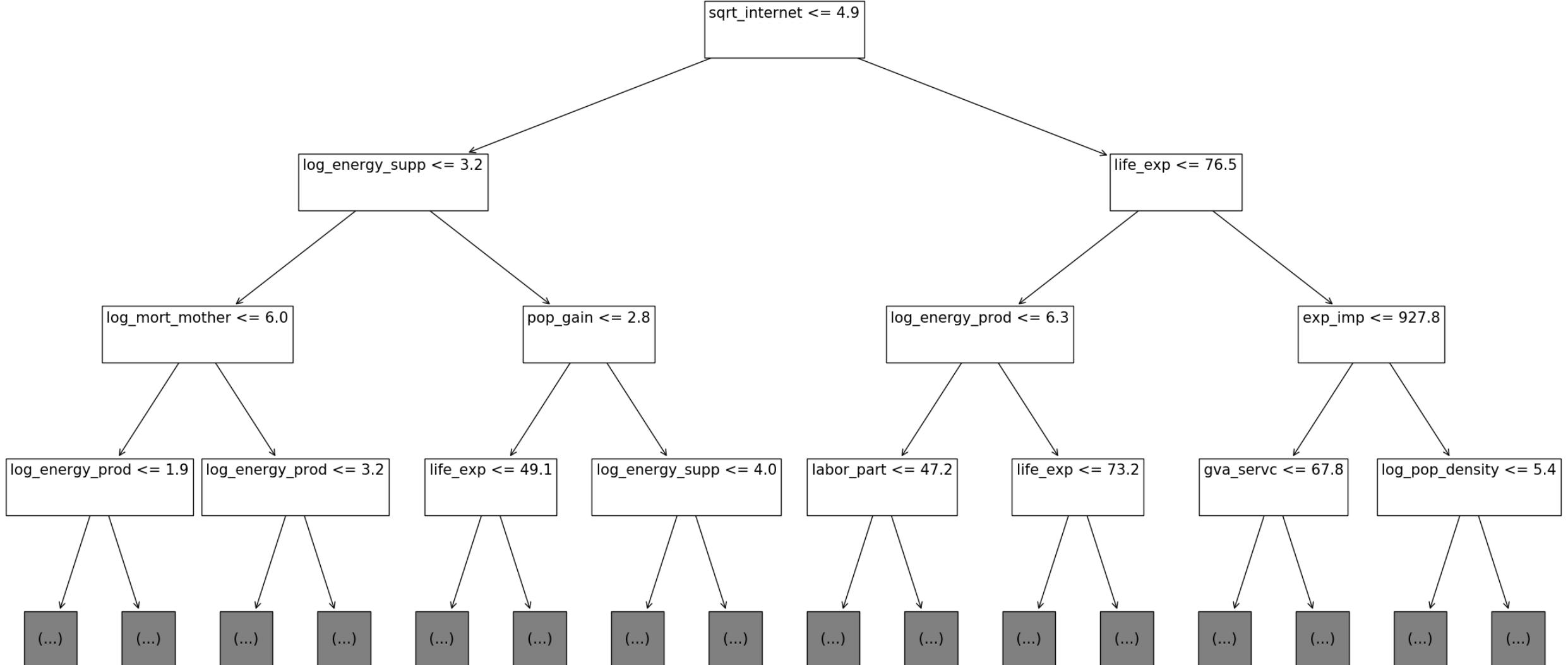




Which Features Were Significant?

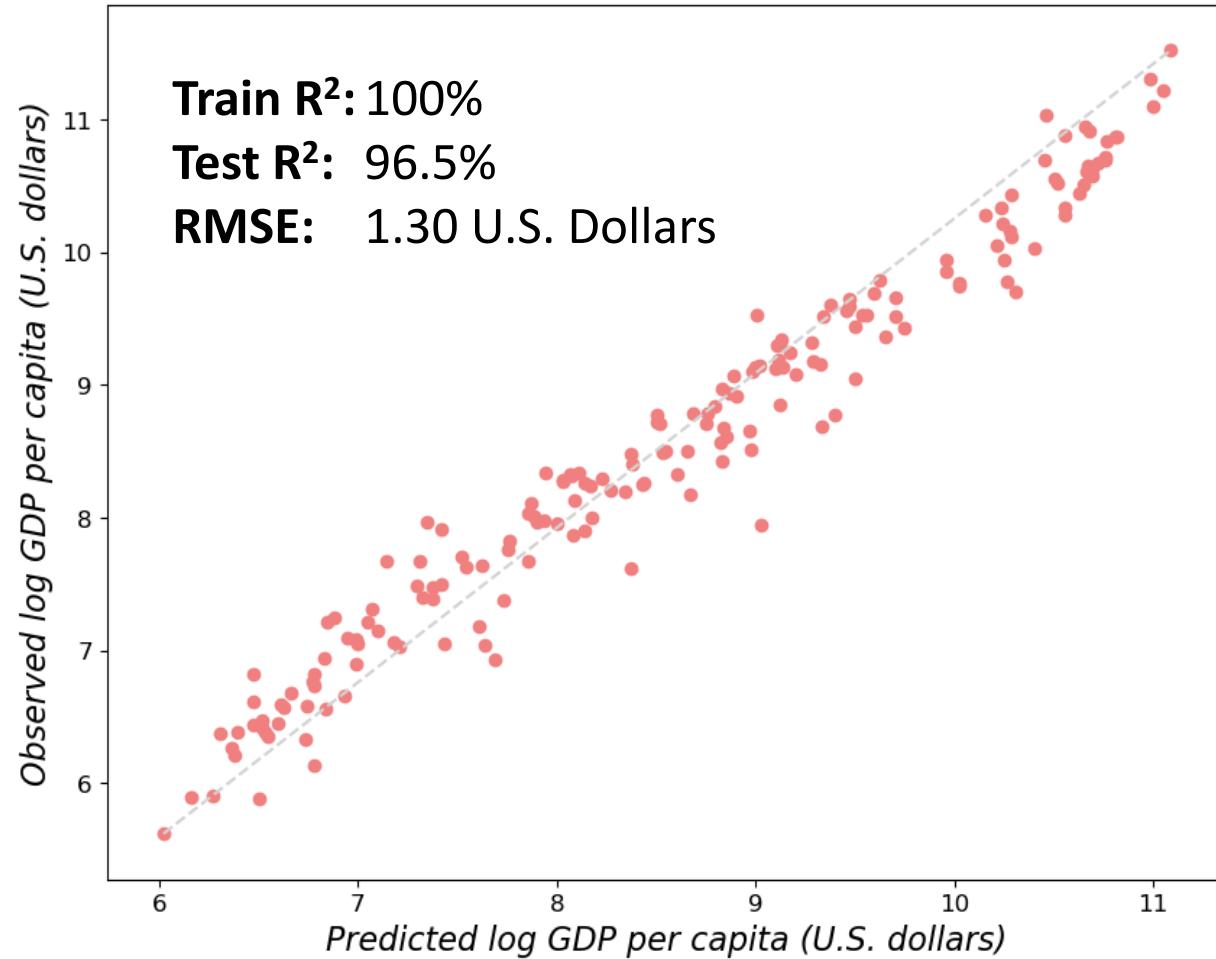
Not Predictive	Positively Related	Negatively Related
Exports - Imports Difference	Percentage of Population using Internet	Gross Value Added by Agriculture
Year-to-Year Population Gain	Tourism Expenditure	Infant Mortality Ratio
Total Population	Energy Consumed per Capita	Maternal Mortality Ratio
Gender Ratio	GHG Emissions per Capita	Population Density
Energy Produced per Capita	Life Expectancy	Labor Force Participation Rate
Percentage of GDP Spent on Health	Percentage of Population Being Migrants	Unemployment Rate

Random Forest for Improved Accuracy

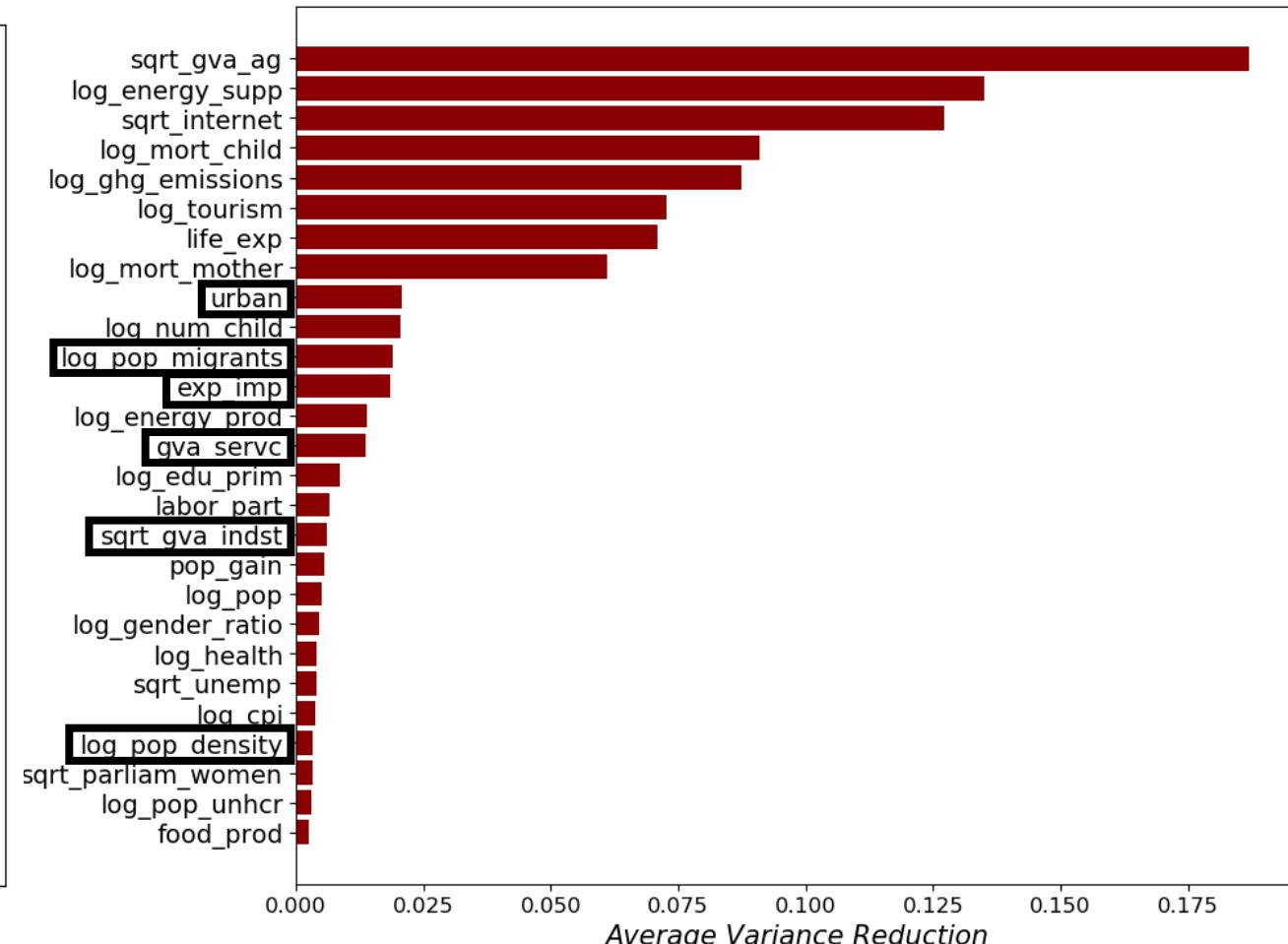


Random Forest Predictions and Features

Observed vs. Predicted log GDP per capita



Random Forest Feature Importance





Key Insights

- Countries with higher GDP per capita characterized by better access to healthcare and technology, larger energy consumption/GHG emissions, service-driven economies, and diverse urban populations
- Social metrics such as education, gender distribution, and population growth largely non-predictive, but family size surprisingly relevant

Thank you



Appendix

Multiple Linear Regression

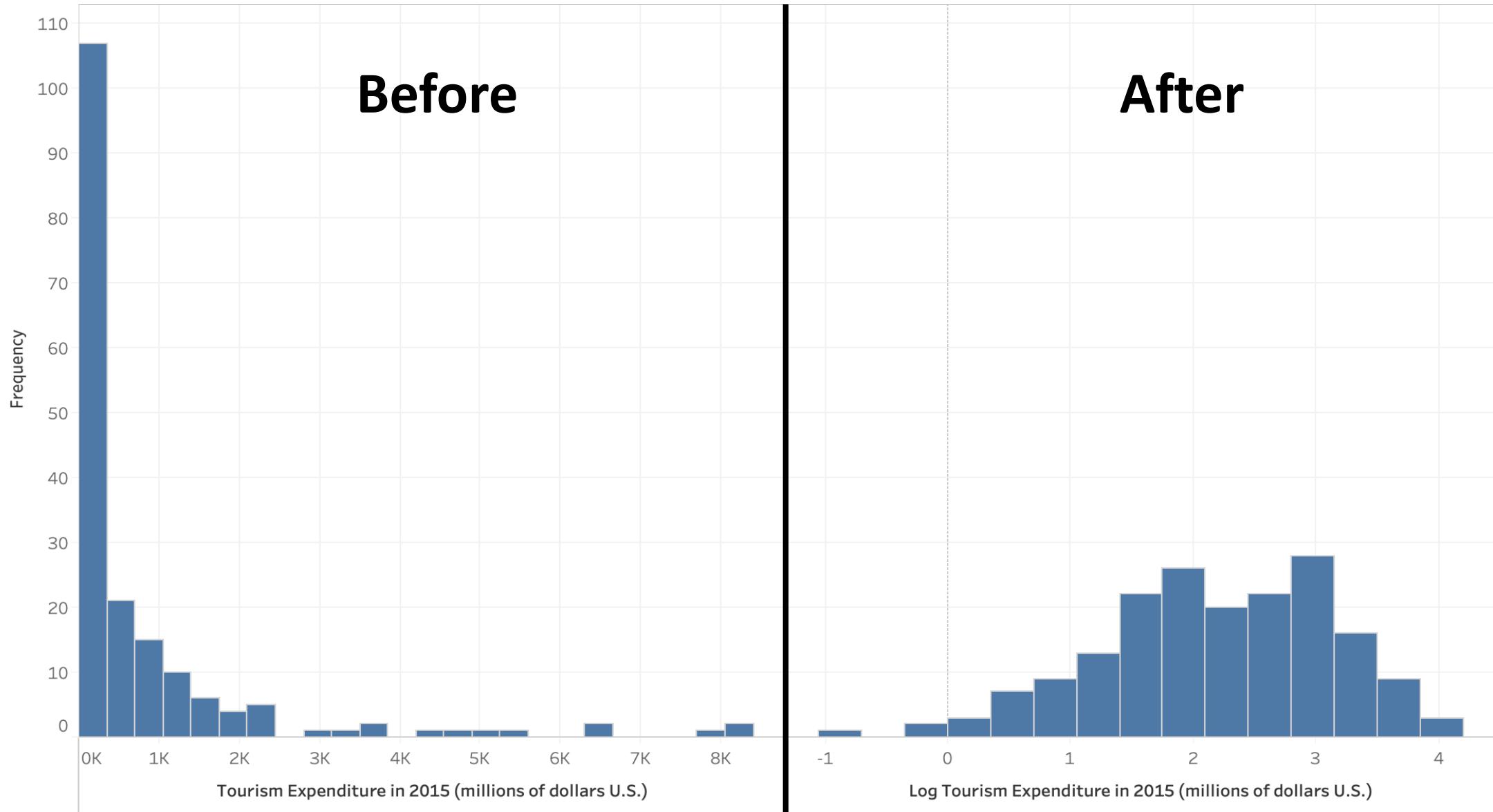
- Allows us to model a quantitative feature (e.g. GDP) using a linear combination of predictor variables (e.g. GHG emissions, population, etc.)
- Estimates a coefficient for each predictor that serves as a “weight” for how important that variable is
- Considers the predictor variables as a group in determining the impact of each one individually
- Highly flexible, interpretable, and fast

$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p + \varepsilon$$

Random Forest

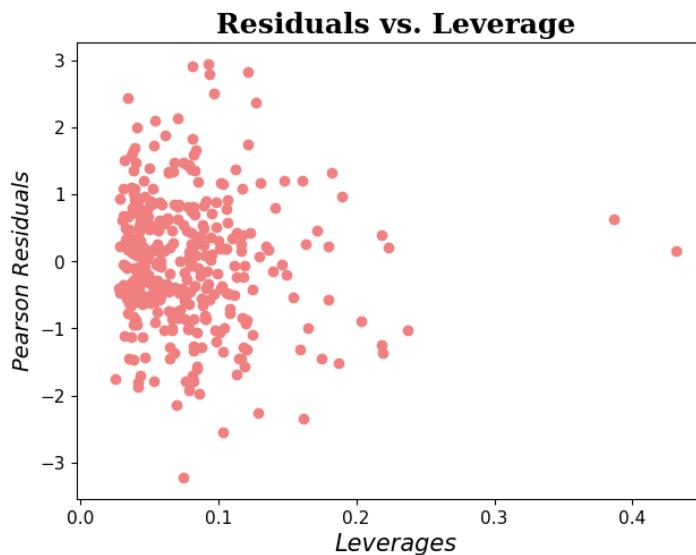
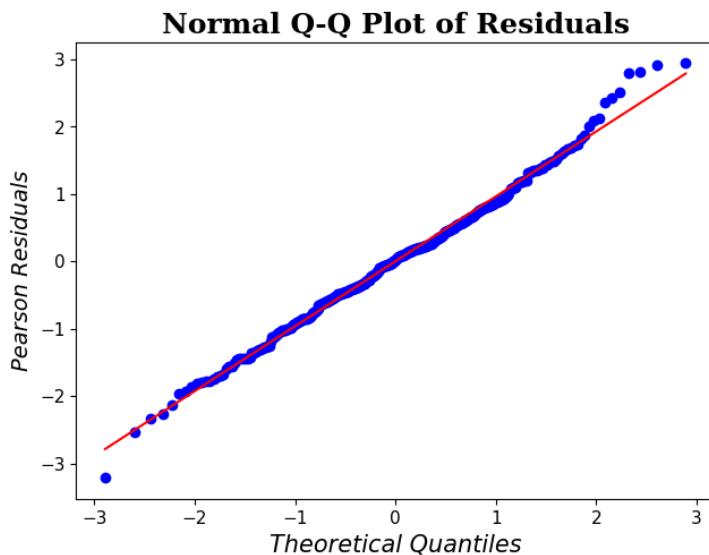
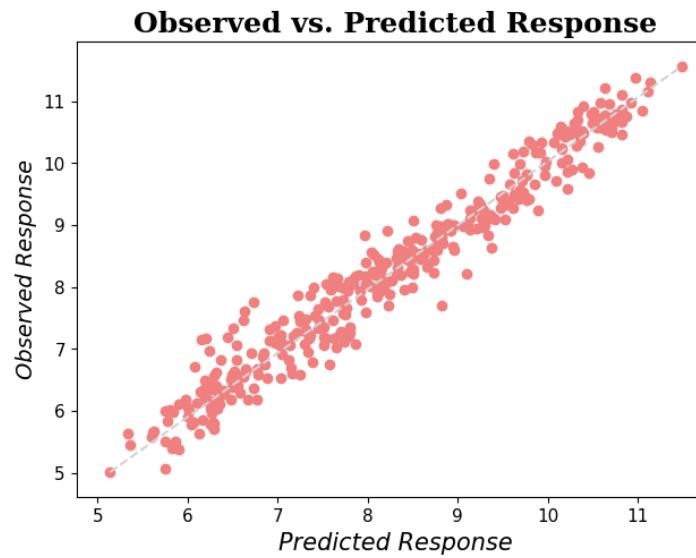
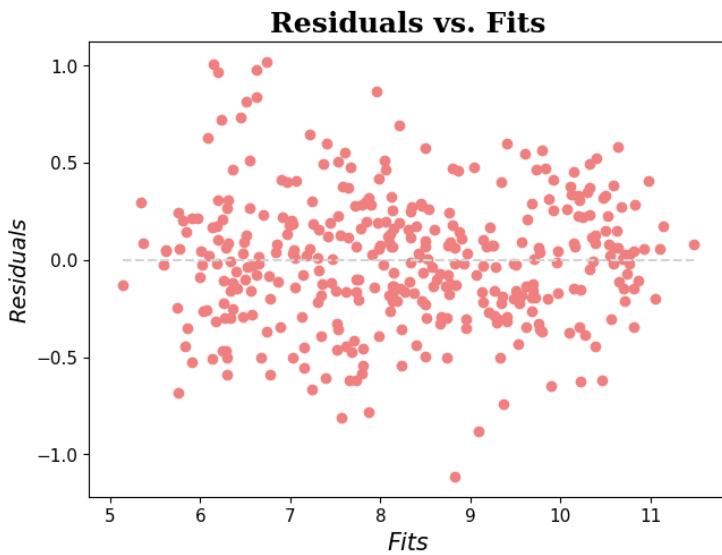
- Ensemble method that combines the outputs of many **decision trees** for regression
- Decision trees make a series of binary splits (on one feature at a time) to reduce the variance in the data as efficiently as possible
- The bottom nodes of each tree assign predictions to new samples by simply averaging the response values (GDP) of the training samples in that node
- Random forests randomize the samples and features used to build each tree in order to avoid overfitting the training data

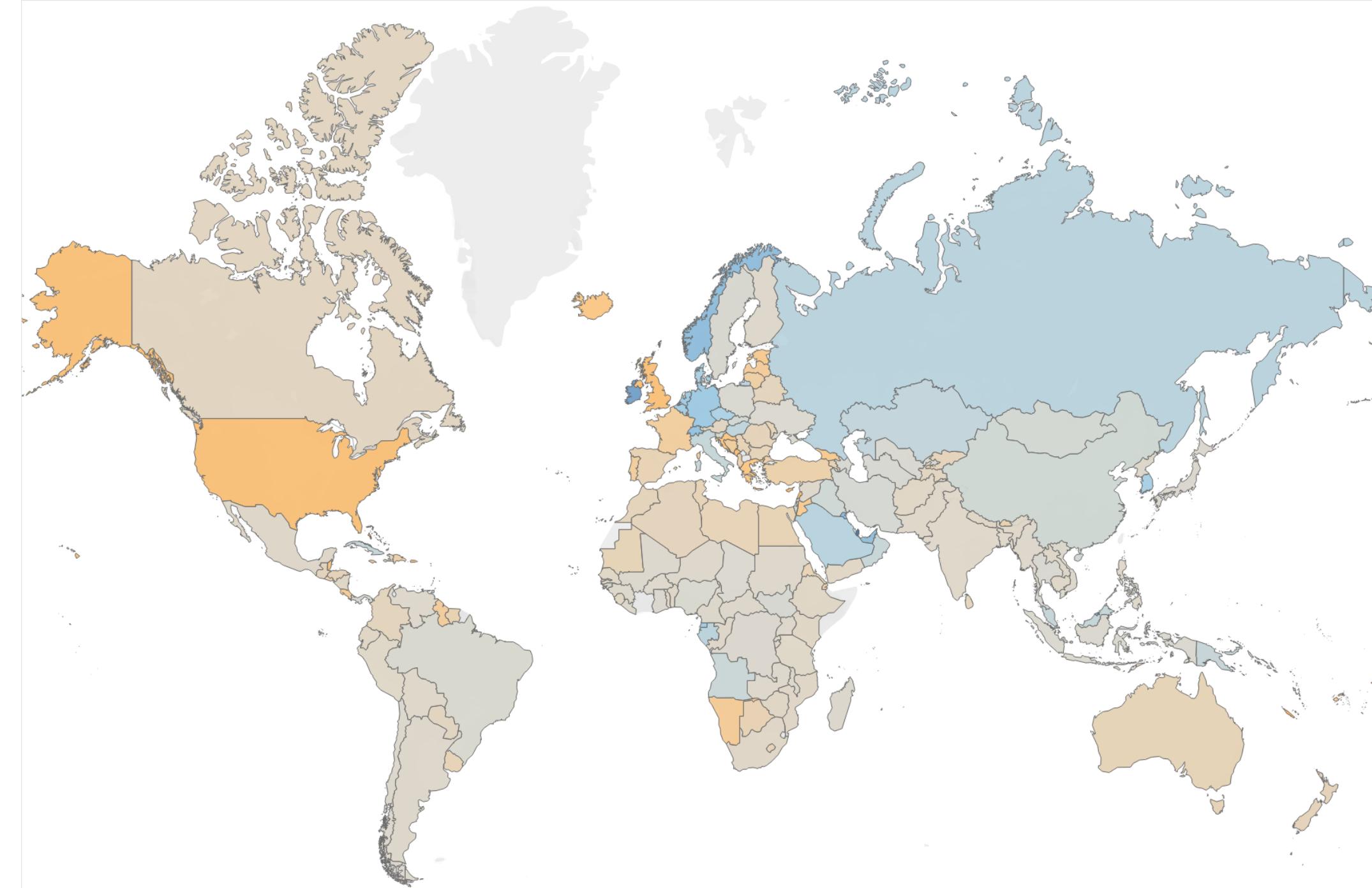
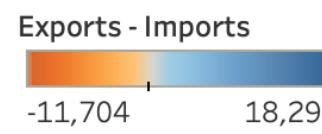
Feature Transformations



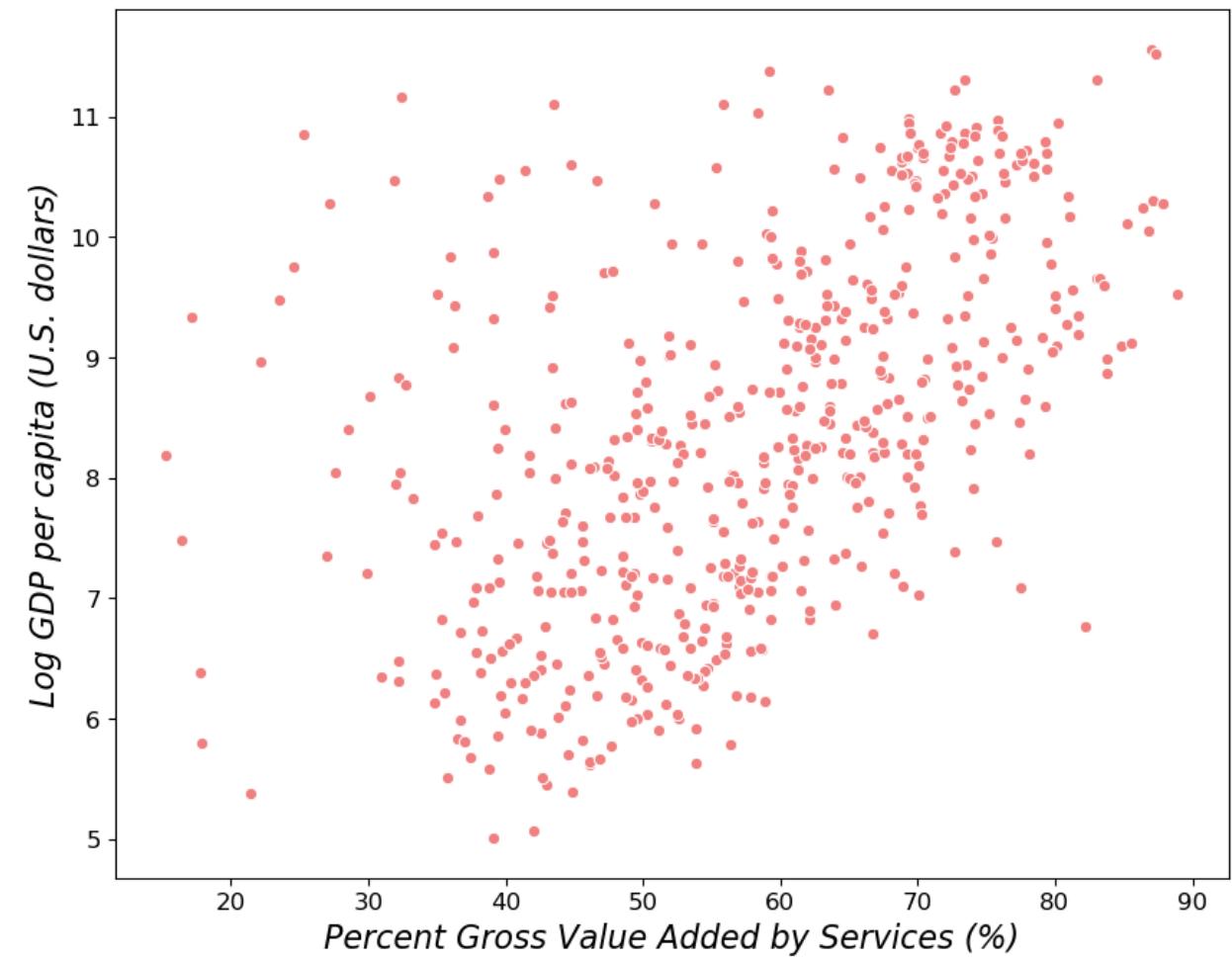
Regression Errors

Regression Diagnostic Plots





GDP per capita vs. Gross Value Added by Service



GDP per capita vs. Population Gain

