



Global Unemployment and Economic Indicators Analysis

Lujia Deng, Jon Kang, Xueyan Liu, Yunfei Zhang

Introduce Our Data



World Development Indicators

- Country based
- Year: 2014-2018
- 1000+ Attributes on
 - **Macroeconomics**
 - **Income and Geographic Group**
 - **Economic Development**
 - Agriculture & Rural Development
 - Education
 - Environment
 - Financial Sector
 - Health
 - Infrastructure
 - Science & Technology
 - Social Development
 - Urban Development

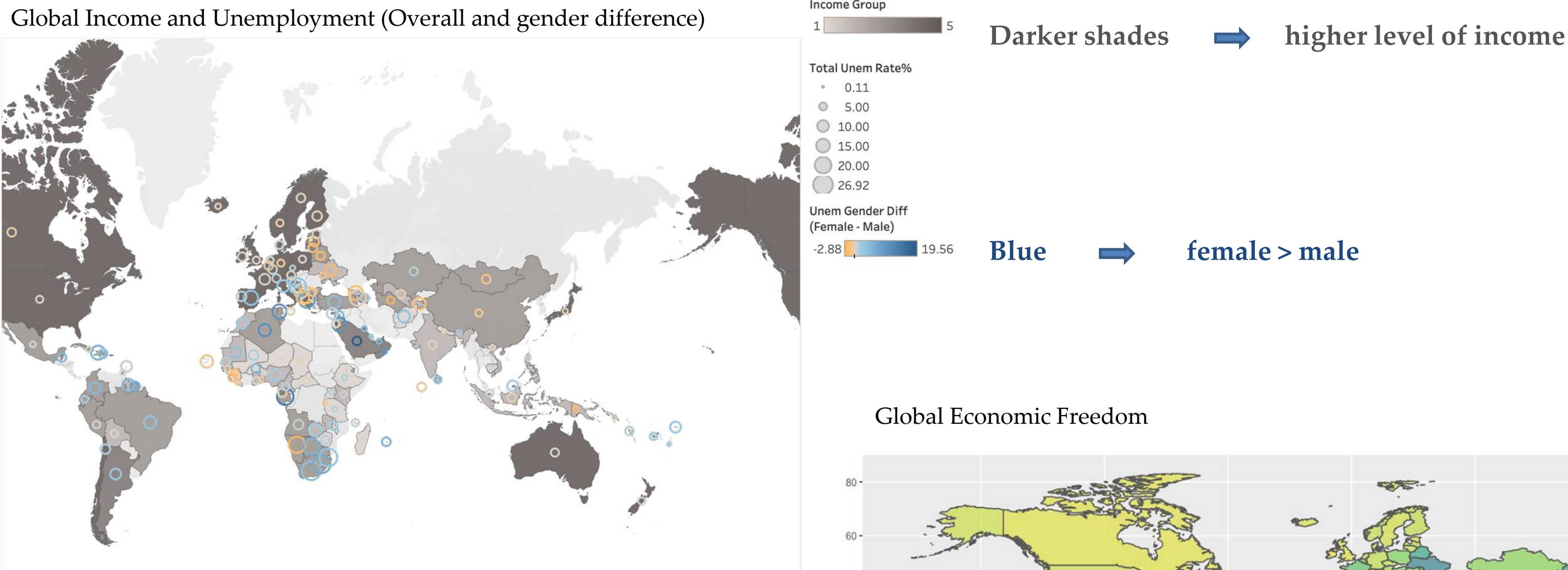


Economic Freedom Index

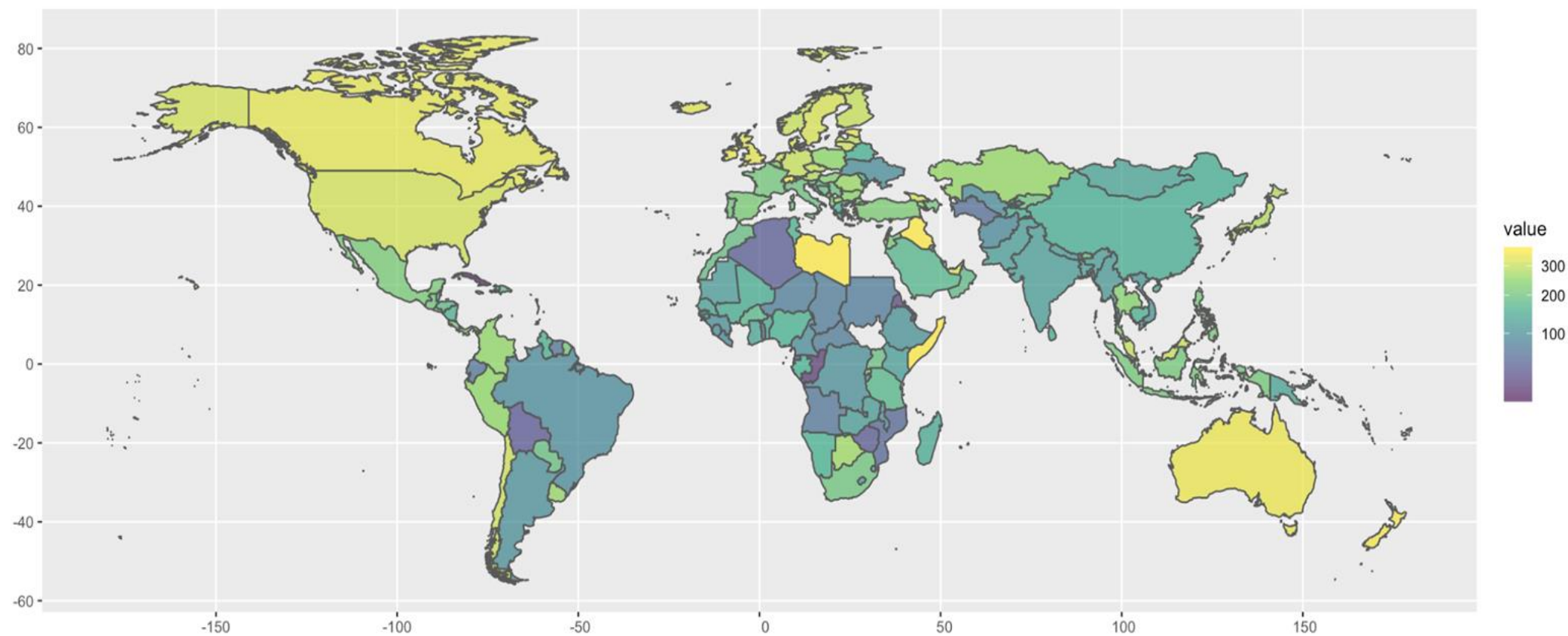
- Country based
- Year: 2014-2018
- 12 indicators on
 - **Rule of law** (property rights, judicial effectiveness, and government integrity)
 - **Government size** (tax burden, government spending, and fiscal health)
 - **Regulatory efficiency** (business freedom, Labor freedom, and monetary freedom)
 - **Market openness** (trade freedom, investment freedom, and financial freedom)
- All these features are graded on a scale of 0 to 100.

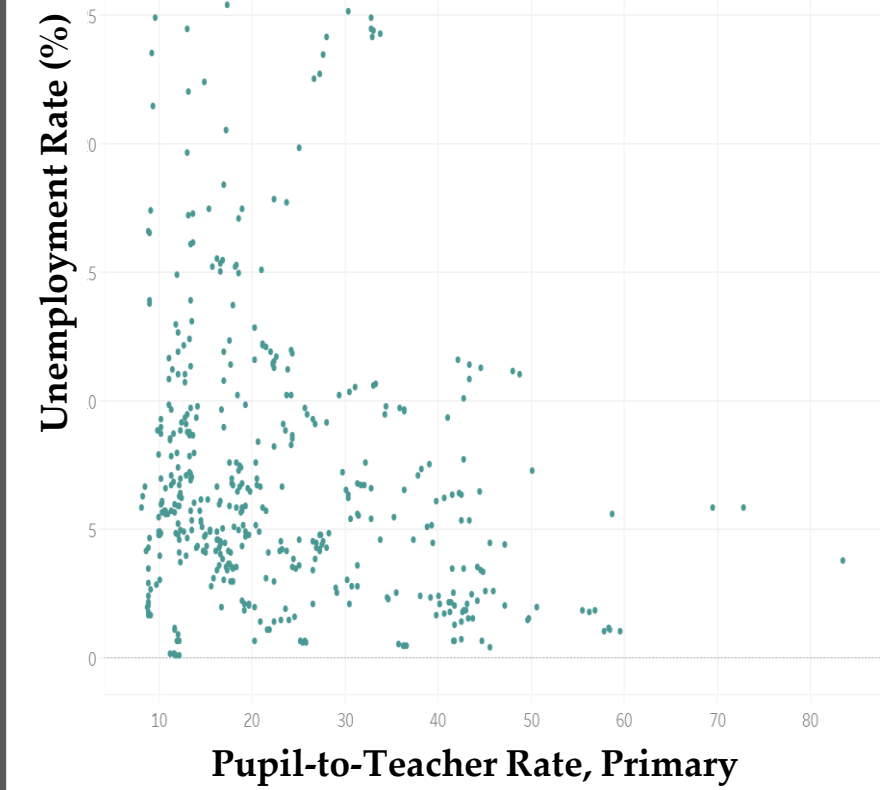
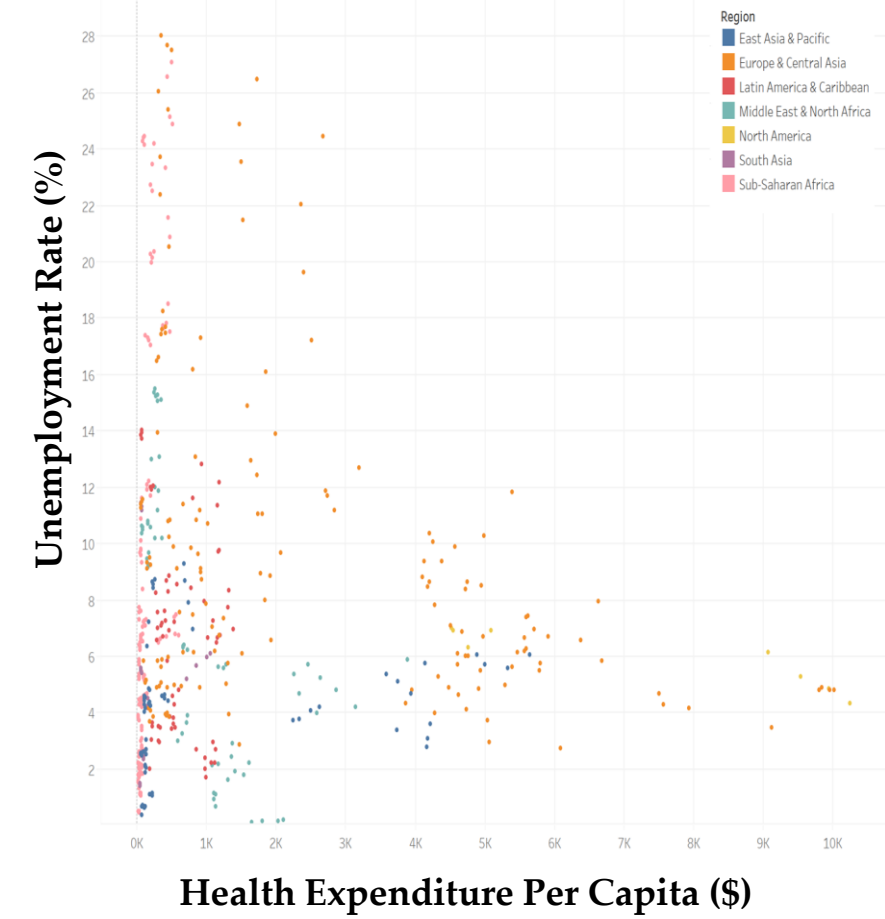
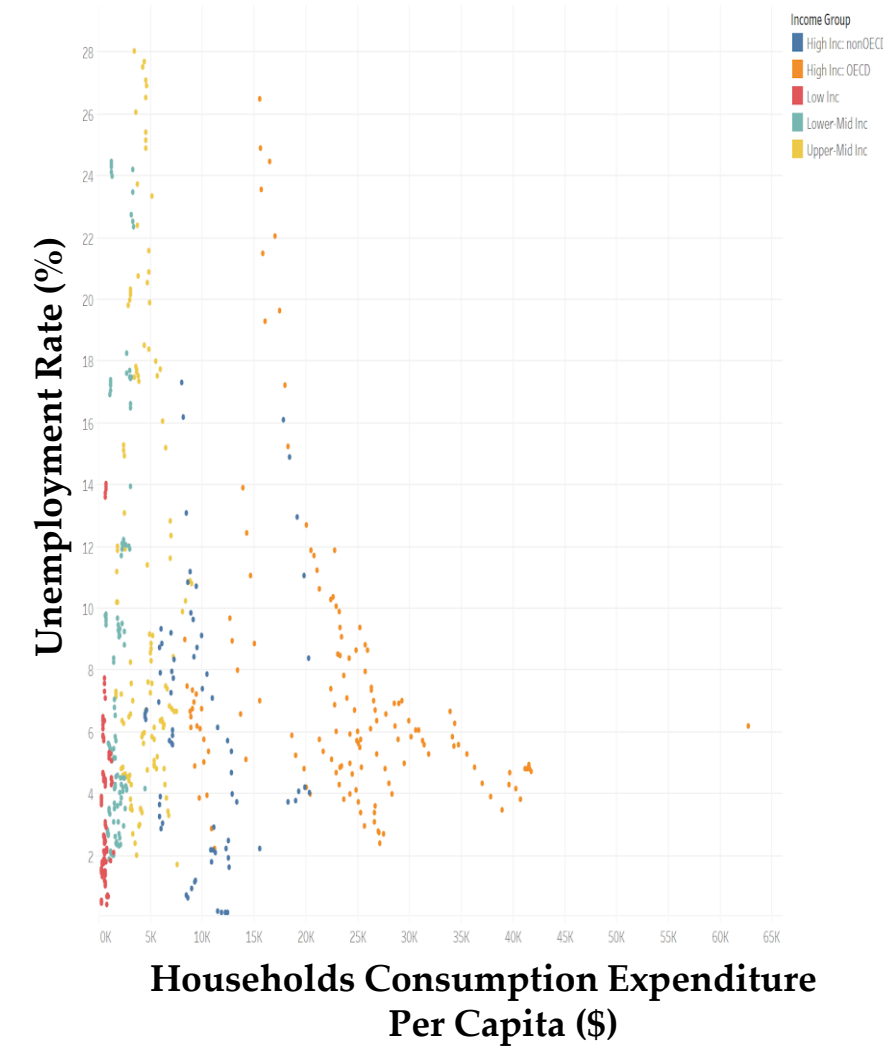
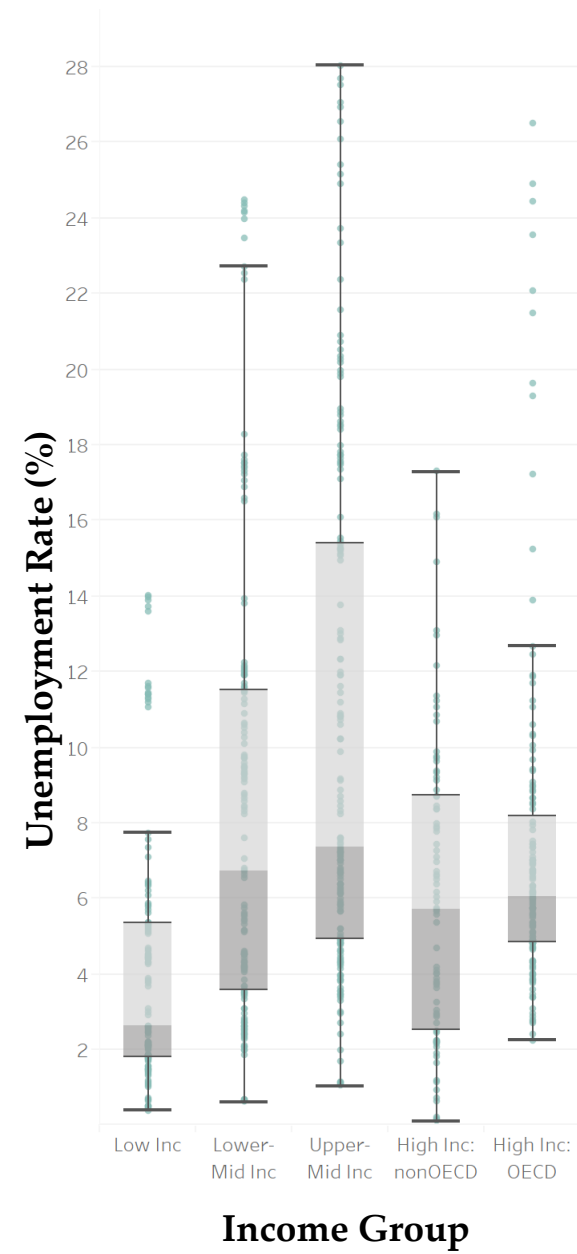
Exploration: Global Income and Economic Freedom

Global Income and Unemployment (Overall and gender difference)



Global Economic Freedom





Exploration: Unemployment and Income, Consumption, Health & Education

Methods that we used:



PCA

Aimed to reduce the dimensionality; applied regression for the pc that can explain more than 95% variability.



Random Forest

Applied random forest algorithm with cross validation, resulted in satisfactory results and understanding of feature importance



LASSO+Gradient Boosting Tree

PCA with Regression

Challenge: Missing Value

Solution:

- Removing missing values
 - Removing the columns with more than 50% miss value
 - Removing the incomplete rows
 - (32 cols, 534 rows)
- Replacing missing with:
 - Mean Value or
 - Random select from distribution or
 - Iterative PCA method ✓
 - (42 cols, 705 rows)

11%

Missing Value

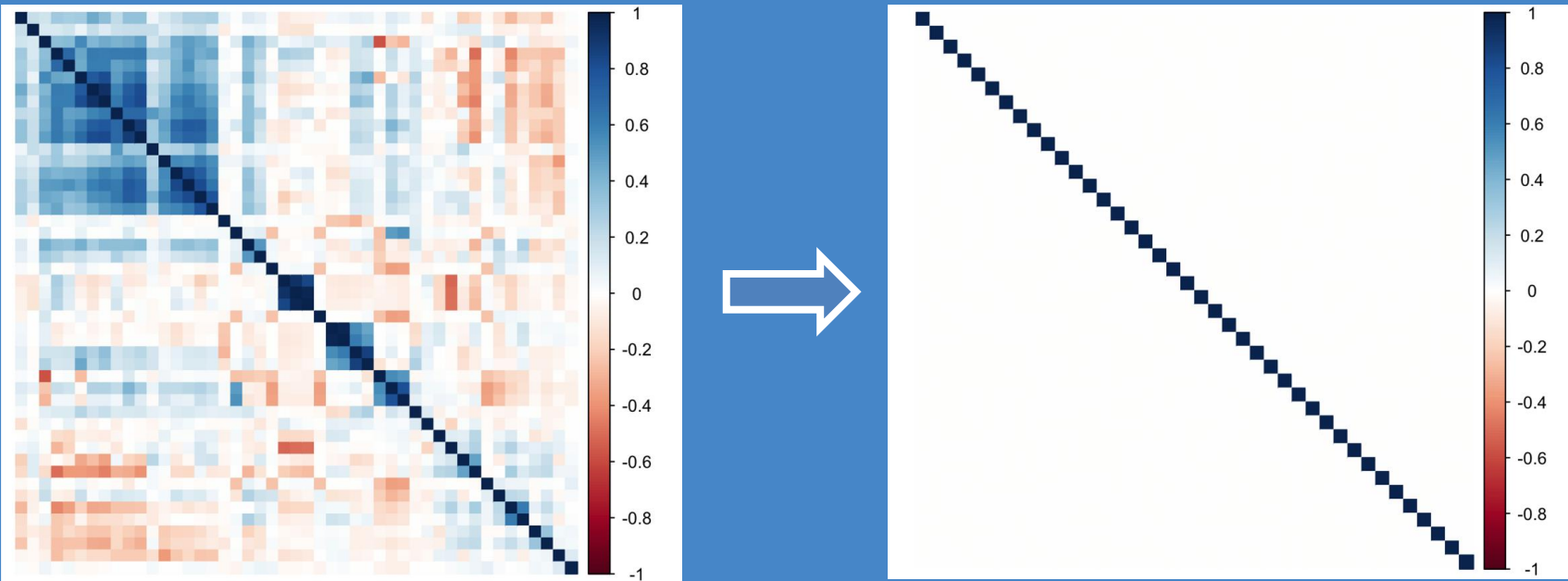


0%

Missing Value

PCA with Regression: Understanding the Results

- PCA removes correlation between variables.
- 95% of the variance explained by the principal components will be good enough.
- Reduced the dimensionality to around 22, (14 for removing the missing value)



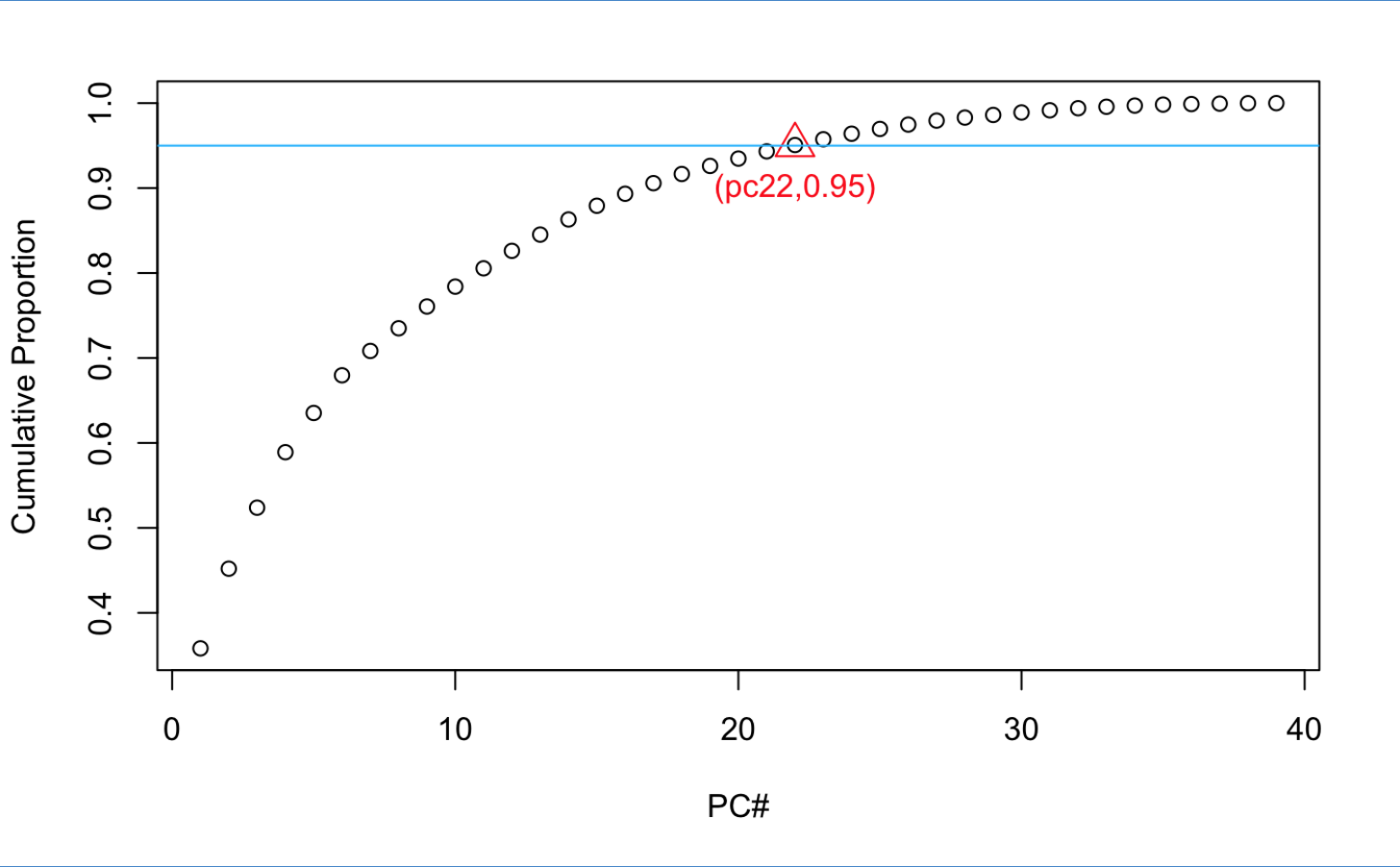
Importance of components:

| | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 | PC8 | PC9 | PC10 |
|------------------------|--------|--------|---------|---------|---------|---------|---------|---------|---------|---------|
| Standard deviation | 3.5566 | 1.8201 | 1.59297 | 1.51852 | 1.27730 | 1.25167 | 1.00620 | 0.97060 | 0.95193 | 0.90926 |
| Proportion of Variance | 0.3581 | 0.0938 | 0.07184 | 0.06529 | 0.04619 | 0.04436 | 0.02866 | 0.02667 | 0.02566 | 0.02341 |
| Cumulative Proportion | 0.3581 | 0.4519 | 0.52379 | 0.58907 | 0.63526 | 0.67962 | 0.70829 | 0.73496 | 0.76061 | 0.78402 |

| | PC11 | PC12 | PC13 | PC14 | PC15 | PC16 | PC17 | PC18 | PC19 | PC20 |
|------------------------|---------|--------|---------|---------|---------|---------|---------|--------|---------|---------|
| Standard deviation | 0.87178 | 0.8529 | 0.82186 | 0.79309 | 0.75527 | 0.70616 | 0.66331 | 0.6176 | 0.57910 | 0.55270 |
| Proportion of Variance | 0.02152 | 0.0206 | 0.01912 | 0.01781 | 0.01615 | 0.01412 | 0.01246 | 0.0108 | 0.00949 | 0.00865 |
| Cumulative Proportion | 0.80554 | 0.8261 | 0.84526 | 0.86307 | 0.87922 | 0.89334 | 0.90580 | 0.9166 | 0.92609 | 0.93474 |

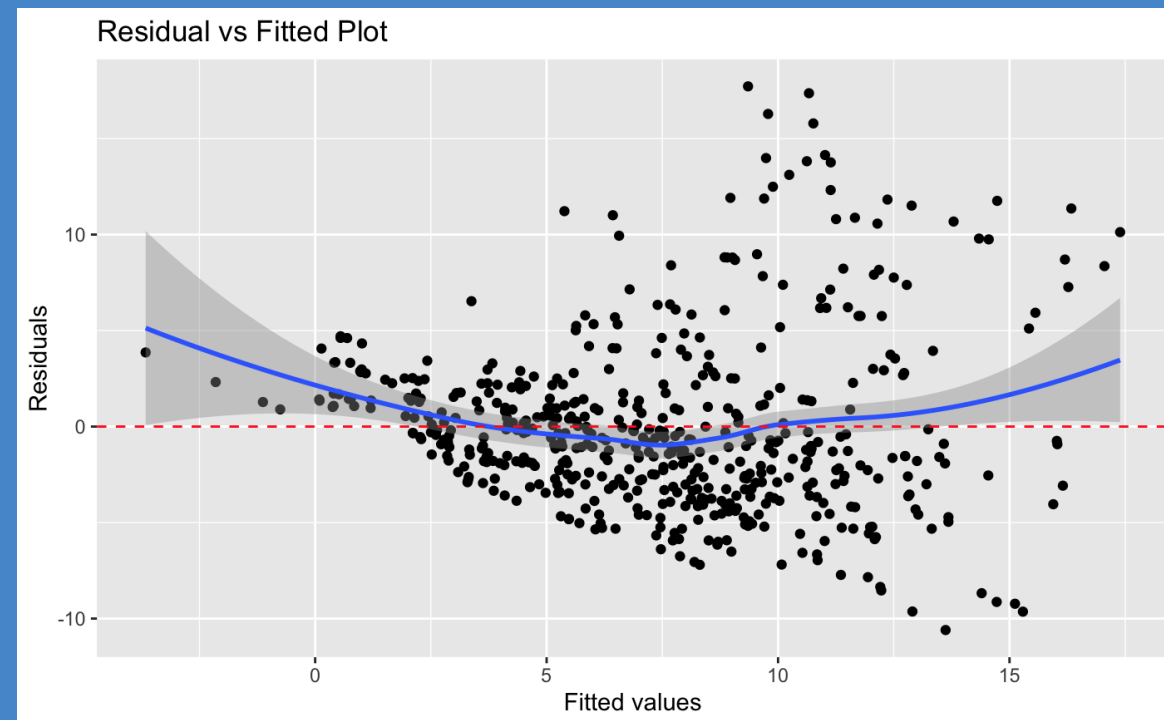
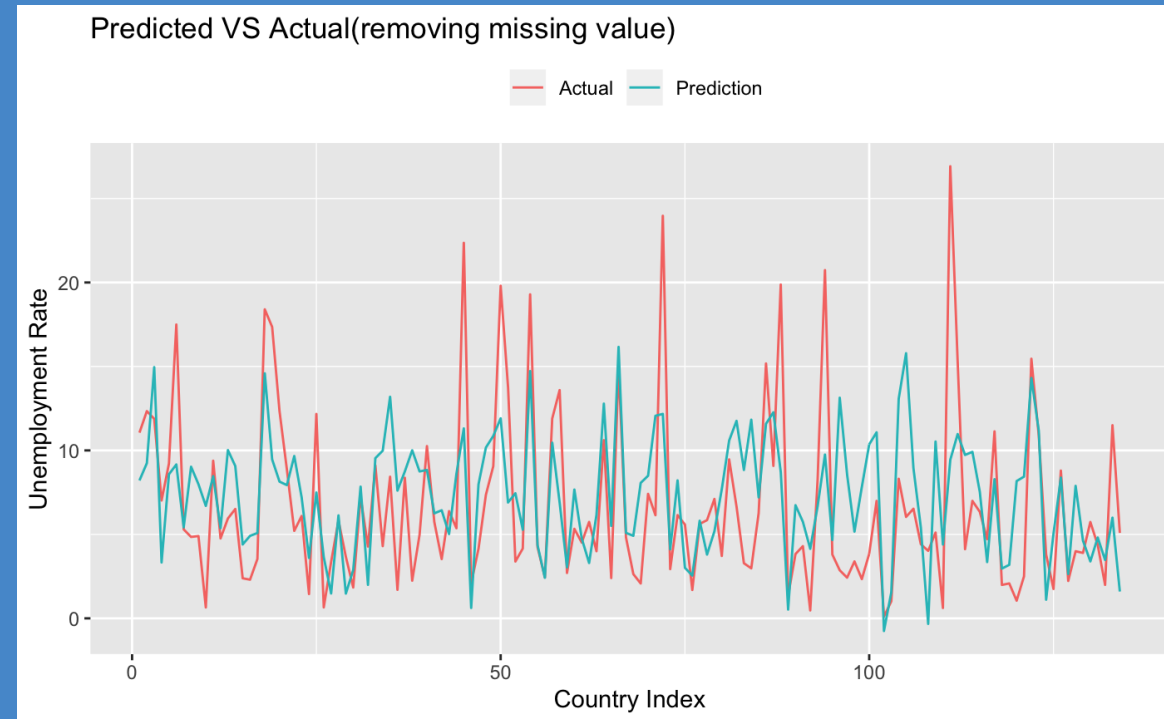
| | PC21 | PC22 | PC23 | PC24 | PC25 | PC26 | PC27 | PC28 | PC29 | PC30 |
|------------------------|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Standard deviation | 0.5480 | 0.51218 | 0.48942 | 0.48201 | 0.44526 | 0.42587 | 0.40649 | 0.35444 | 0.33012 | 0.32164 |
| Proportion of Variance | 0.0085 | 0.00743 | 0.00678 | 0.00658 | 0.00561 | 0.00514 | 0.00468 | 0.00356 | 0.00309 | 0.00293 |
| Cumulative Proportion | 0.9432 | 0.95067 | 0.95745 | 0.96403 | 0.96964 | 0.97478 | 0.97945 | 0.98301 | 0.98610 | 0.98903 |

| | PC31 | PC32 | PC33 | PC34 | PC35 | PC36 | PC37 | PC38 | PC39 |
|------------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| Standard deviation | 0.30050 | 0.29344 | 0.24768 | 0.21979 | 0.20264 | 0.14814 | 0.13884 | 0.12729 | 0.05538 |
| Proportion of Variance | 0.00256 | 0.00244 | 0.00174 | 0.00137 | 0.00116 | 0.00062 | 0.00055 | 0.00046 | 0.00009 |
| Cumulative Proportion | 0.99158 | 0.99402 | 0.99576 | 0.99712 | 0.99829 | 0.99891 | 0.99945 | 0.99991 | 1.00000 |



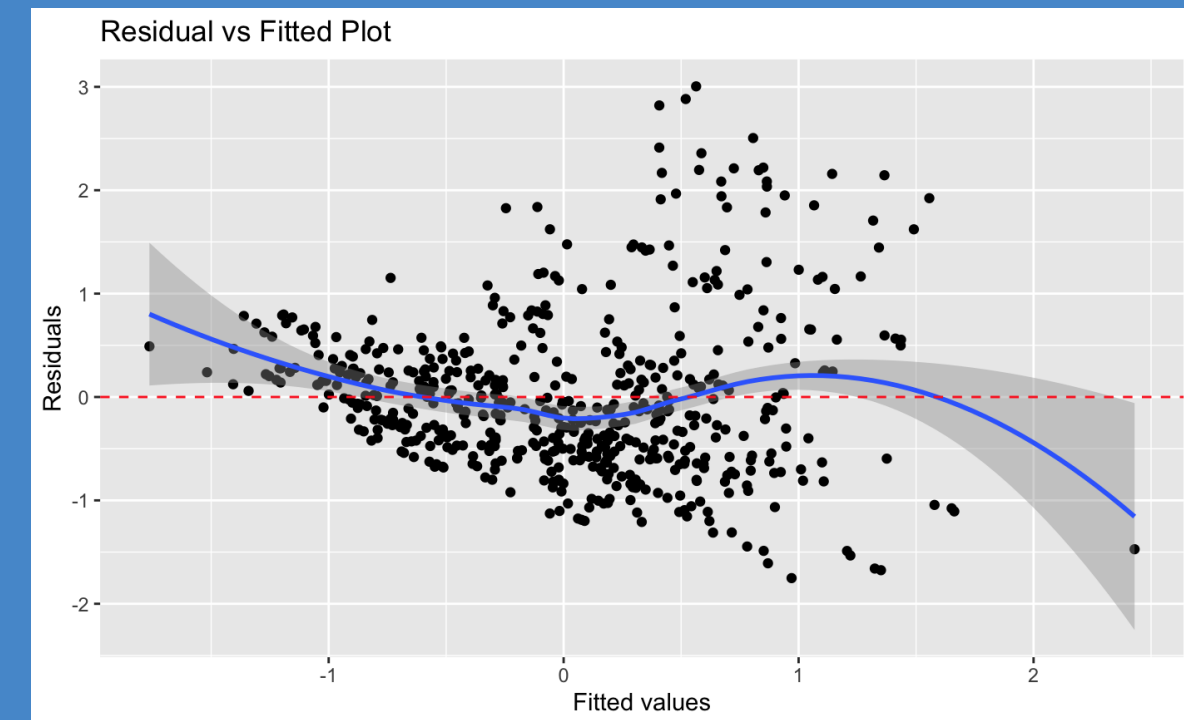
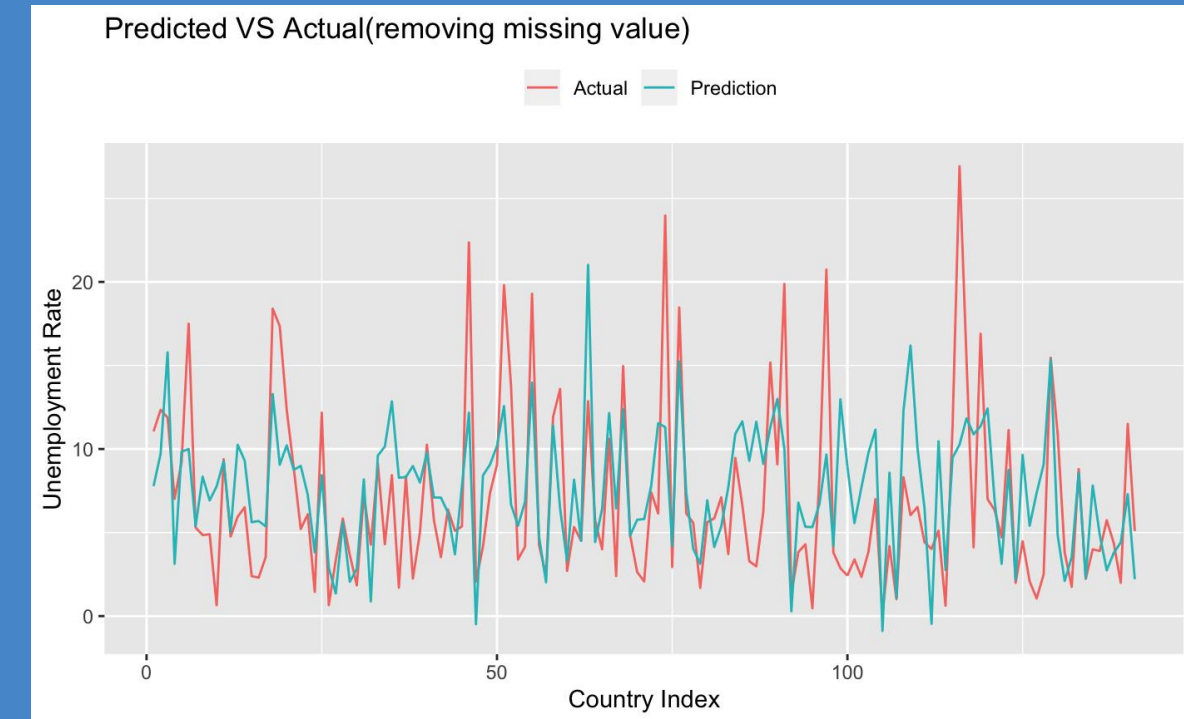
PCA with Regression: Understanding the Results

Removing the missing value



Train RMSE :
0.59
Test RMSE :
0.62

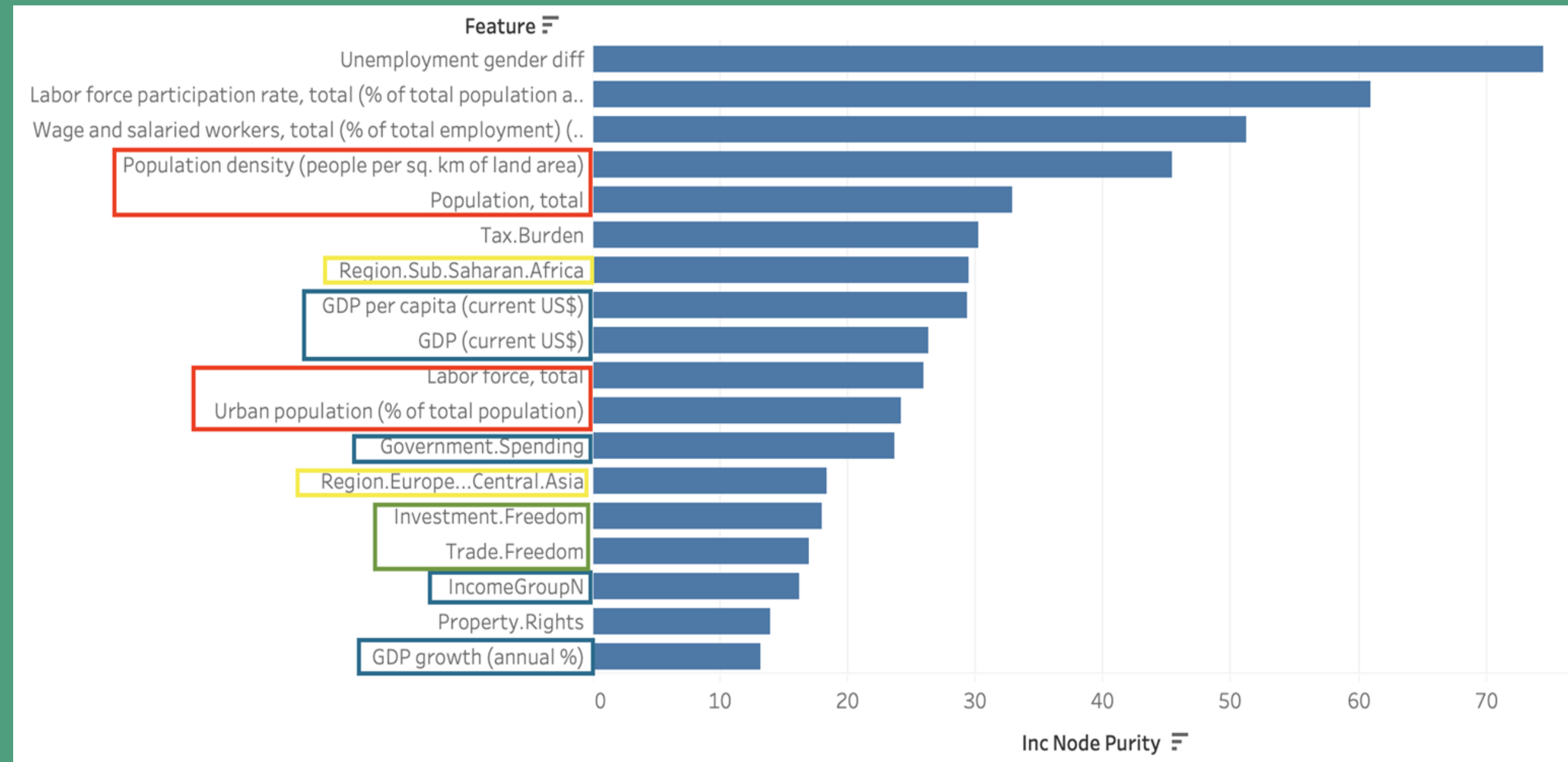
Replacing the missing value



Train RMSE :
0.58
Test RMSE :
0.60

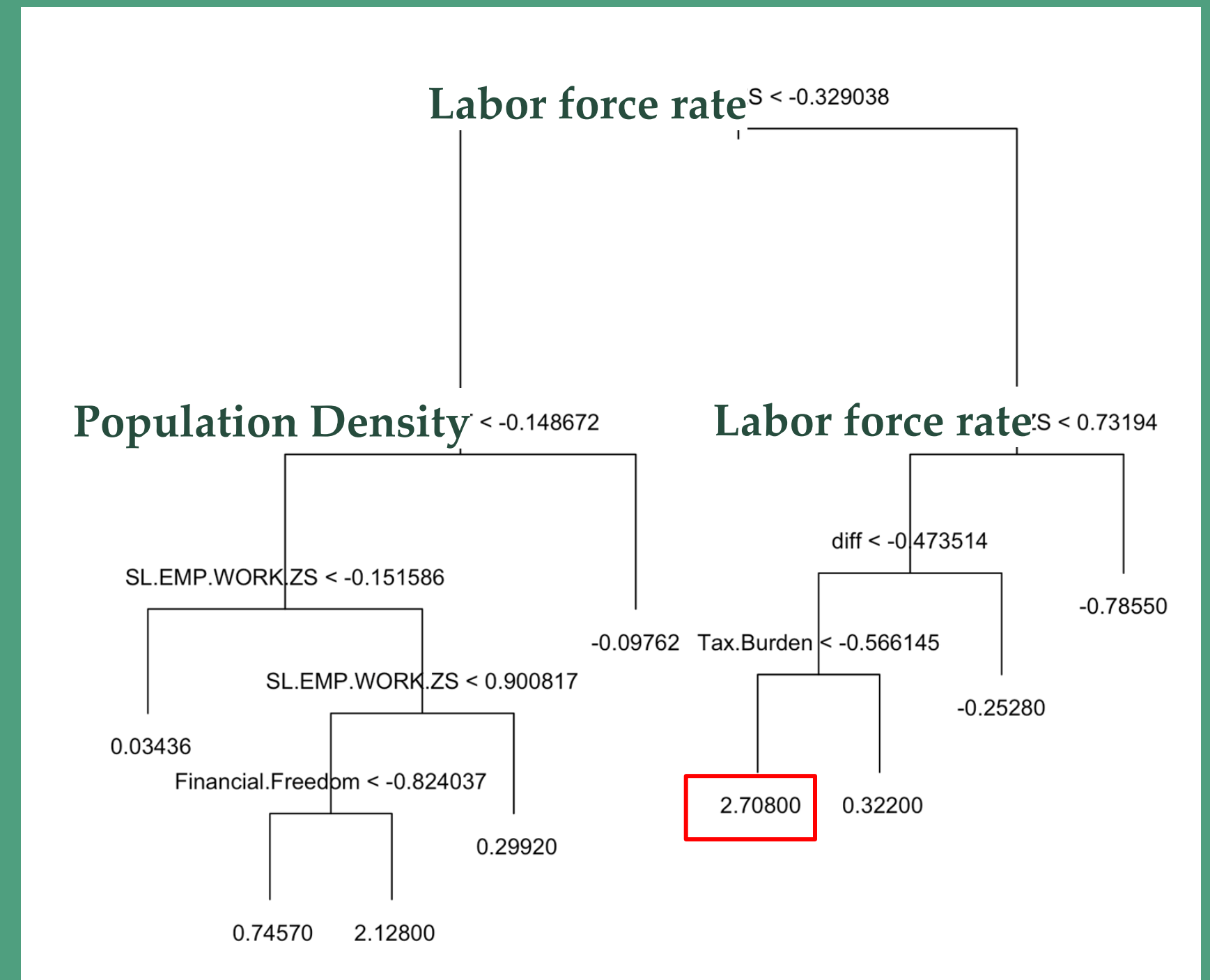
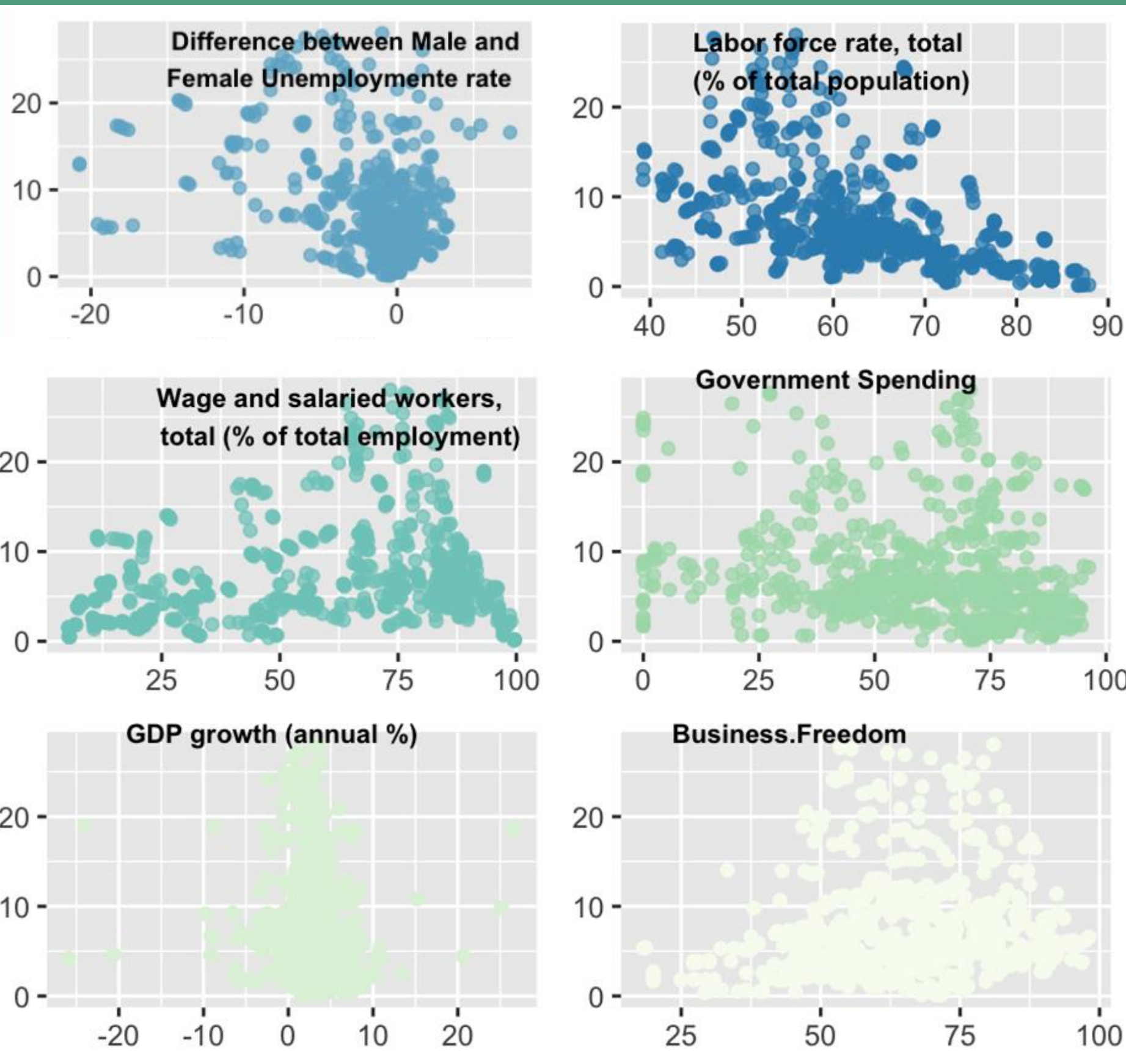
Random Forest Regression

- **Data Manipulation:**
 - Scaled
 - One-hot encoding
 - Feature Engineer
- **Response:** Unemployment rate
- **Predictors:**
 - all
 - fill missing value using rImpute
- **CV:** mtry, ntree, 10-fold
- **Test RMSE :** 0.74



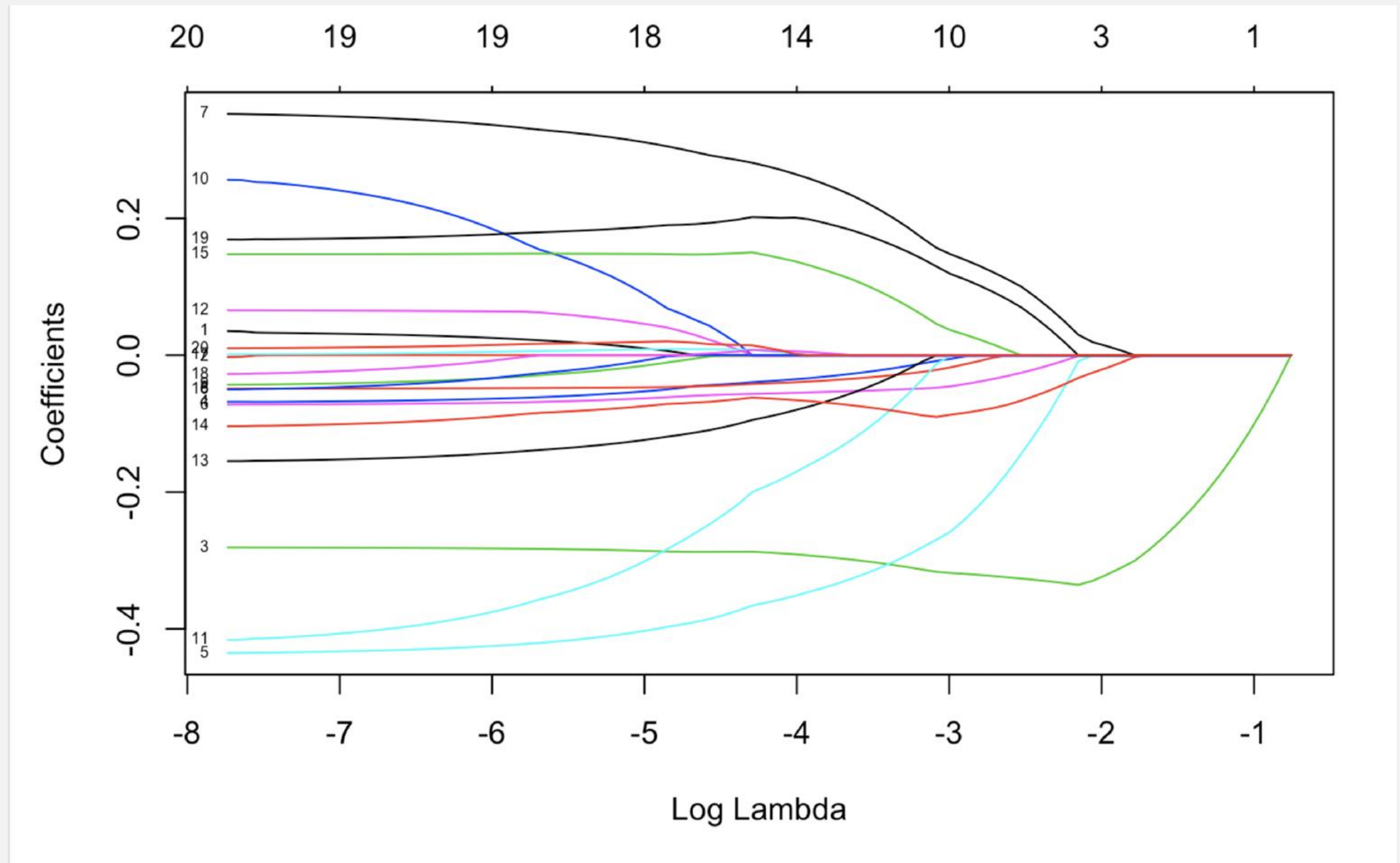
Random Forest Regression: Understanding the Results

- E.g. Higher the labor force rate, lower the unemployment rate



LASSO Feature Selection

1. Clean Data
2. Scale
3. Run LASSO



LASSO Feature Selection

From World Bank’s World Development Indicators data

- SL.TLF.CACT.ZS : Labor force participation rate, total
- NY.GDP.MKTP.CD : GDP (current US\$)
- NY.GDP.PCAP.CD : GDP per capita (current US\$)
- NY.GDP.MKTP.KD.ZG: GDP growth (annual %)
- SL.EMP.WORK.ZS: Wage and salaried workers, total (% of total employment)
- EN.POP.DNST: Population density (people per sq. km of land area)

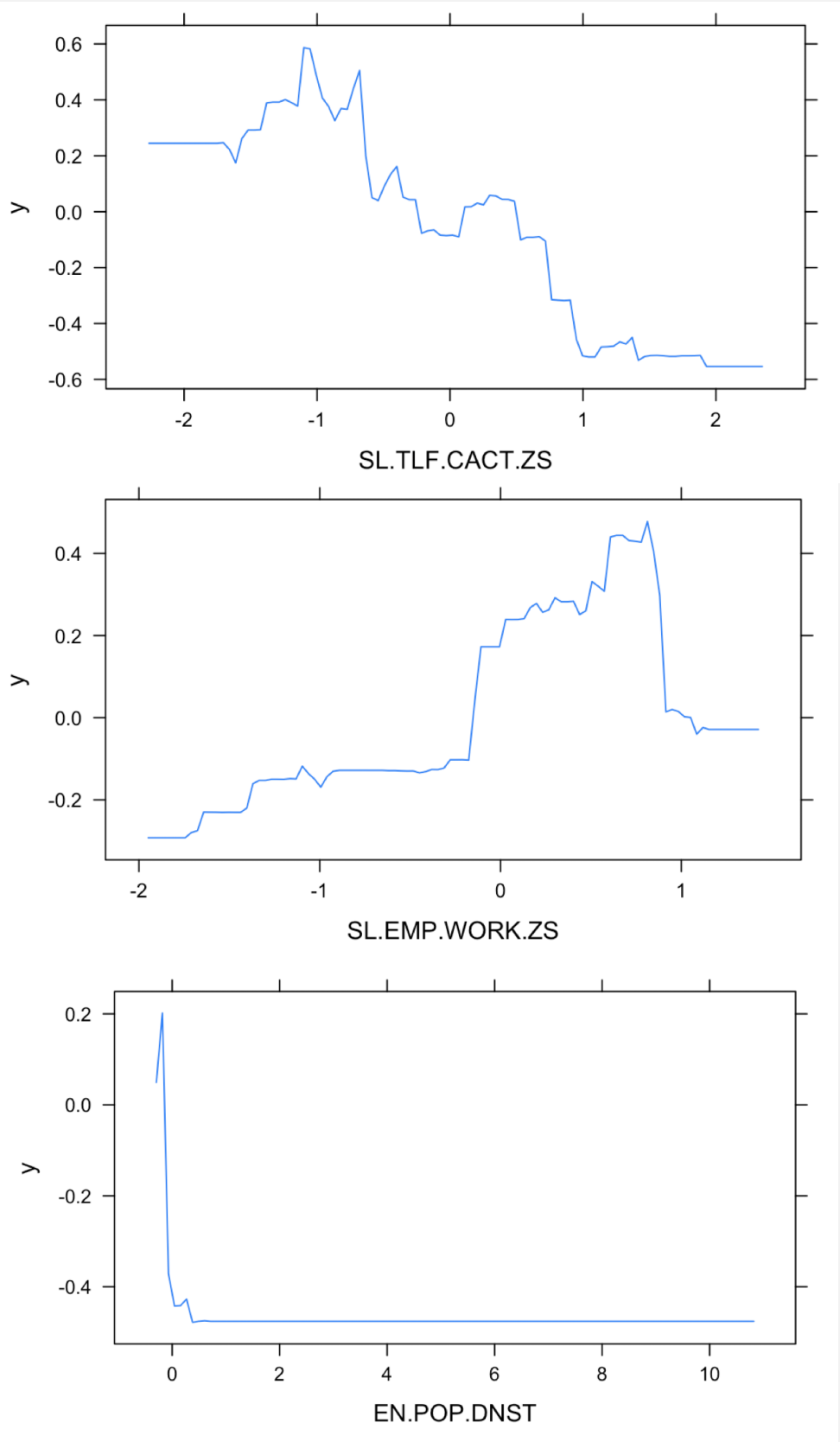
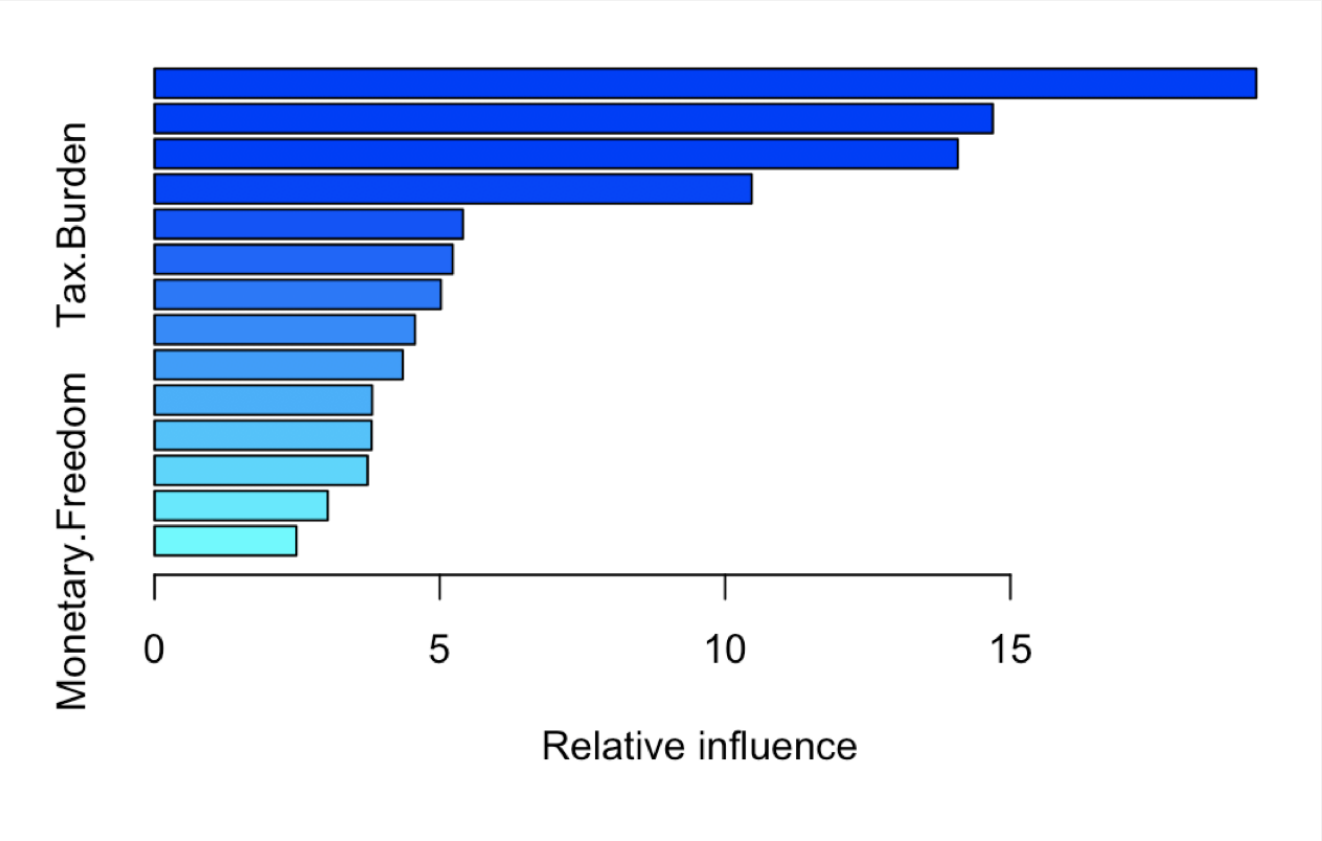
From Index of Economic Freedom data

- Overall.Score
- Property.Rights
- Government.Integrity
- Tax.Burden
- Government.Spending
- Business.Freedom
- Monetary.Freedom
- Investment.Freedom

| | |
|---|--------------|
| 21 x 1 sparse Matrix of class "dgCMatrix" | |
| | s0 |
| (Intercept) | 0.006472942 |
| SP.POP.TOTL | . |
| SL.TLF.TOTL.IN | . |
| SL.TLF.CACT.ZS | -0.305572352 |
| NY.GDP.MKTP.CD | -0.046905909 |
| NY.GDP.PCAP.CD | -0.356850021 |
| NY.GDP.MKTP.KD.ZG | -0.059486188 |
| SL.EMP.WORK.ZS | 0.237748249 |
| EN.POP.DNST | -0.014257341 |
| SP.URB.TOTL.IN.ZS | . |
| Overall.Score | 0.054346619 |
| Property.Rights | -0.240270465 |
| Government.Integrity | 0.046913336 |
| Tax.Burden | -0.105676320 |
| Government.Spending | -0.100982654 |
| Business.Freedom | 0.113155418 |
| Labor.Freedom | . |
| Monetary.Freedom | 0.003249899 |
| Trade.Freedom | . |
| Investment.Freedom | 0.212567685 |
| Financial.Freedom | . |

Gradient Boosting Tree

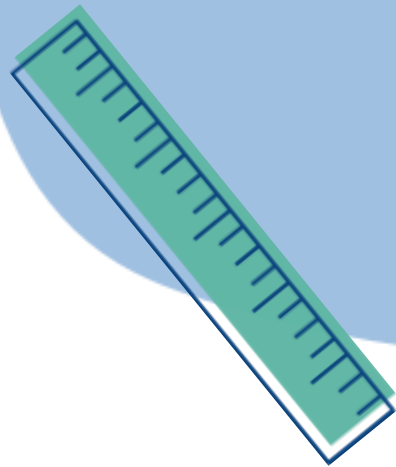
| | var<fctr> | rel.inf<dbl> |
|----------------------|----------------------|--------------|
| SL.TLF.CACT.ZS | SL.TLF.CACT.ZS | 17.872038 |
| SL.EMP.WORK.ZS | SL.EMP.WORK.ZS | 12.496255 |
| EN.POP.DNST | EN.POP.DNST | 11.459481 |
| NY.GDP.MKTP.CD | NY.GDP.MKTP.CD | 9.890386 |
| NY.GDP.MKTP.KD.ZG | NY.GDP.MKTP.KD.ZG | 7.443716 |
| Government.Spending | Government.Spending | 6.239493 |
| NY.GDP.PCAP.CD | NY.GDP.PCAP.CD | 6.145873 |
| Tax.Burden | Tax.Burden | 6.068504 |
| Investment.Freedom | Investment.Freedom | 5.458991 |
| Overall.Score | Overall.Score | 5.364466 |
| Government.Integrity | Government.Integrity | 4.320011 |
| Business.Freedom | Business.Freedom | 3.346121 |
| Monetary.Freedom | Monetary.Freedom | 2.110850 |
| Property.Rights | Property.Rights | 1.783816 |



Our Conclusions

Model Comparison

Gradient Boosting!



Most Related to Unemployment



1. National Income
2. Economic Freedom
3. Gender Difference in Unemployment

Next Steps

1. Find more potential related variables to improve models' performance
2. Identify causal relationships to inform policy making

