Environment, Social, Governance Data

By Lindsay, Elizabeth, Karene and Michaela

Background on data

Since 2019, The Word Bank publishes a Sovereign ESG Data Framework divided into 3 pillars (Environmental, Social and Governance). It includes data in connection with their 17 Sustainable Development Goals. The purpose is to help potential investors and shed light on countries' sustainability performance.

Environment	Social	Governance
Emissions & pollution	Education & skills	Human rights
Natural capital endowment & management	Employment	Government effectiveness
Energy use & security	Demography	Stability & rule of law
Environment/climate risk & resilience	Poverty & inequality	Economic environment
Food security	Health & nutrition	Gender
	Access to services	Innovation

Background on data

CO2 Emissions	Fertility rate
Prevalence of overweight	School enrollment
Population density	Fossil Fuels
Government effectiveness	Internet Use
Hospital Beds	Voice and Accountability
GDP	Women in Parliament
Population density	Political Stability
Rule of Law	Women in Labor Force

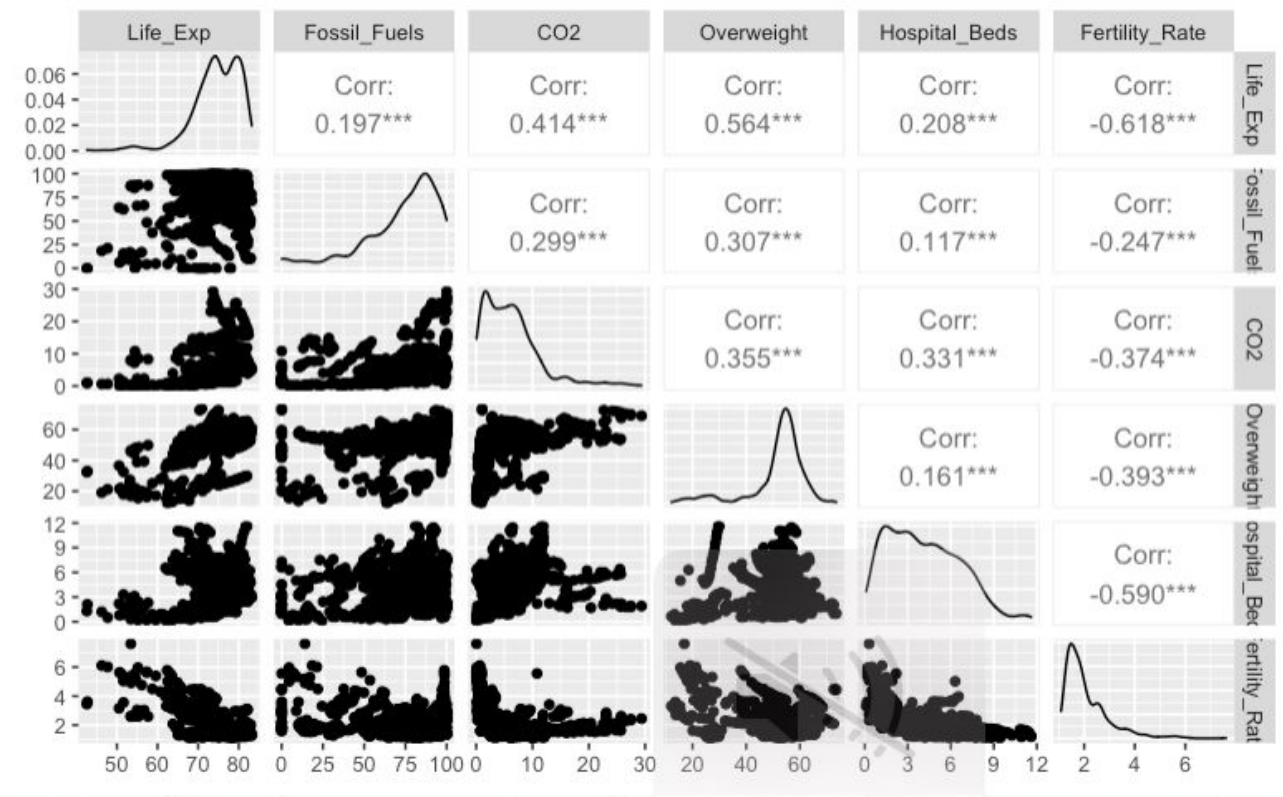
How are data collected?

ESG Data comes from different sources. Their main source is the World Bank.

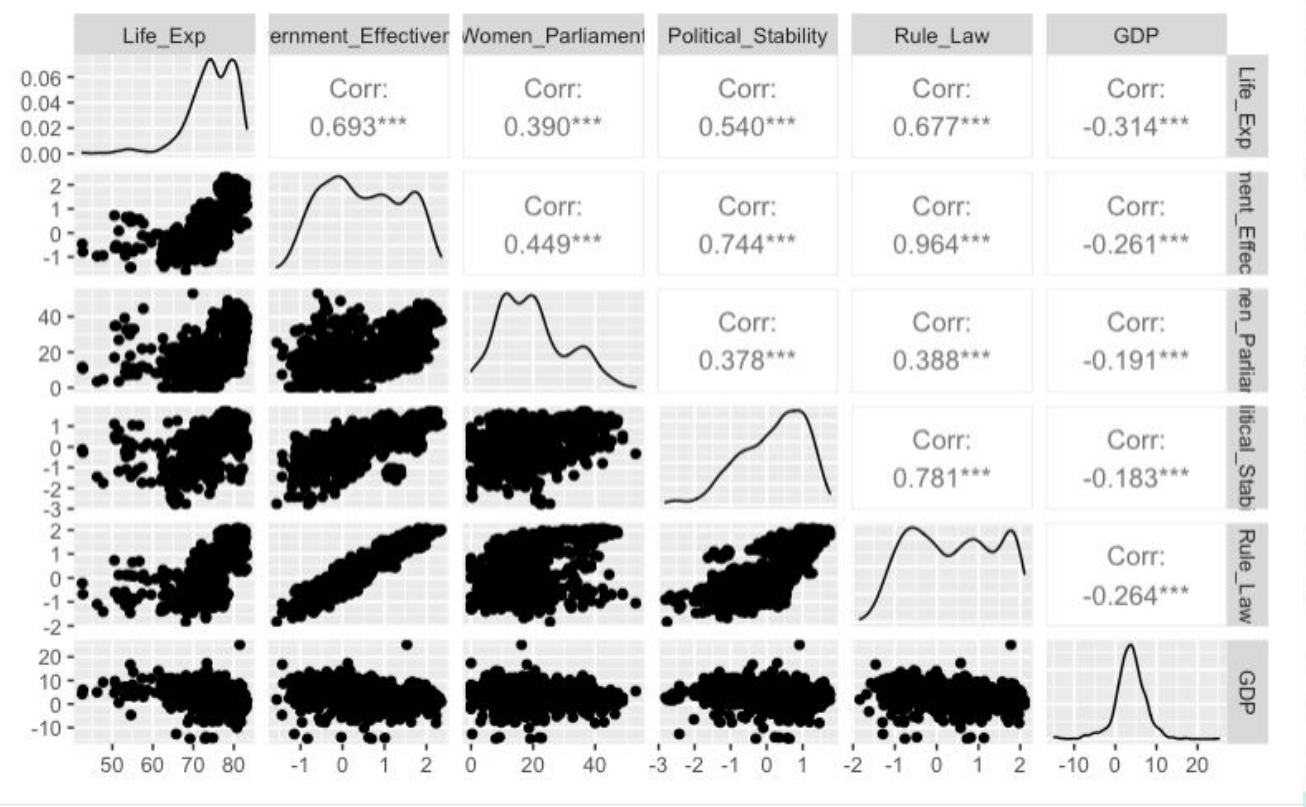
Our Interest

How the variables impact life expectancy across groups of people, such as gender, development status, and income, regions of the world including Latin America, European Union, Central Asia, and East Asia and the Pacific, and over a time span of 20 years.

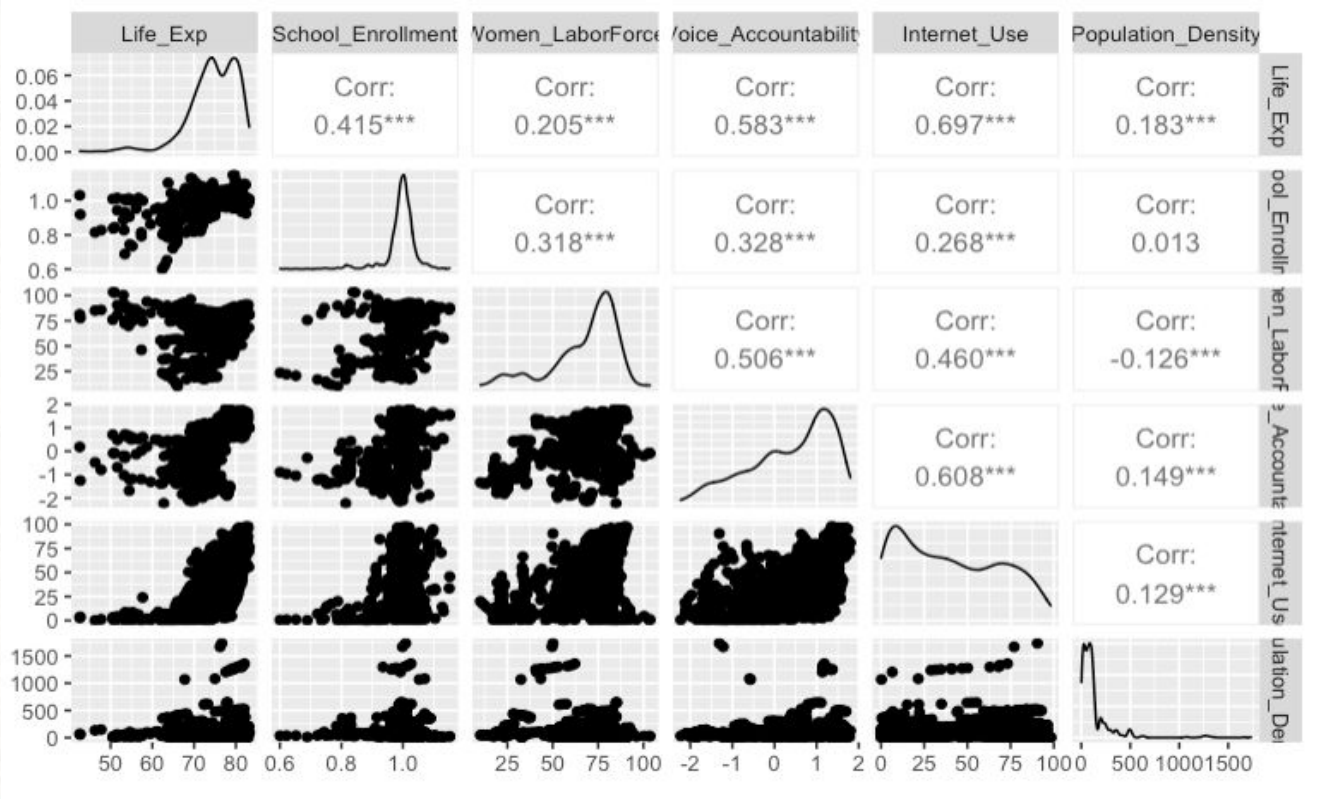
Exploratory data analysis



Exploratory data analysis

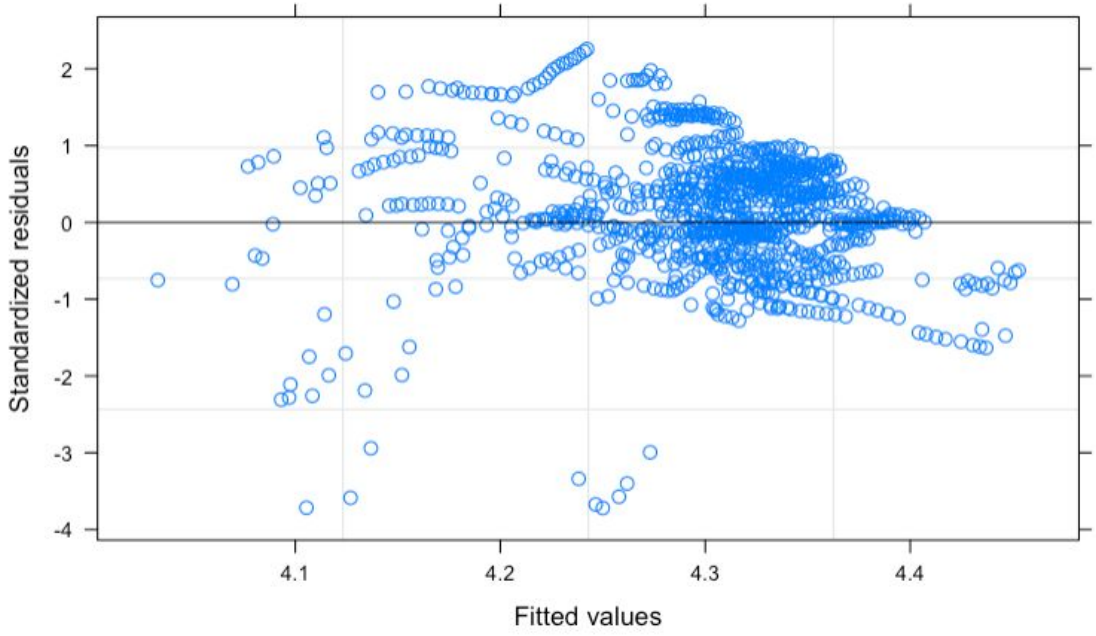


Exploratory data analysis



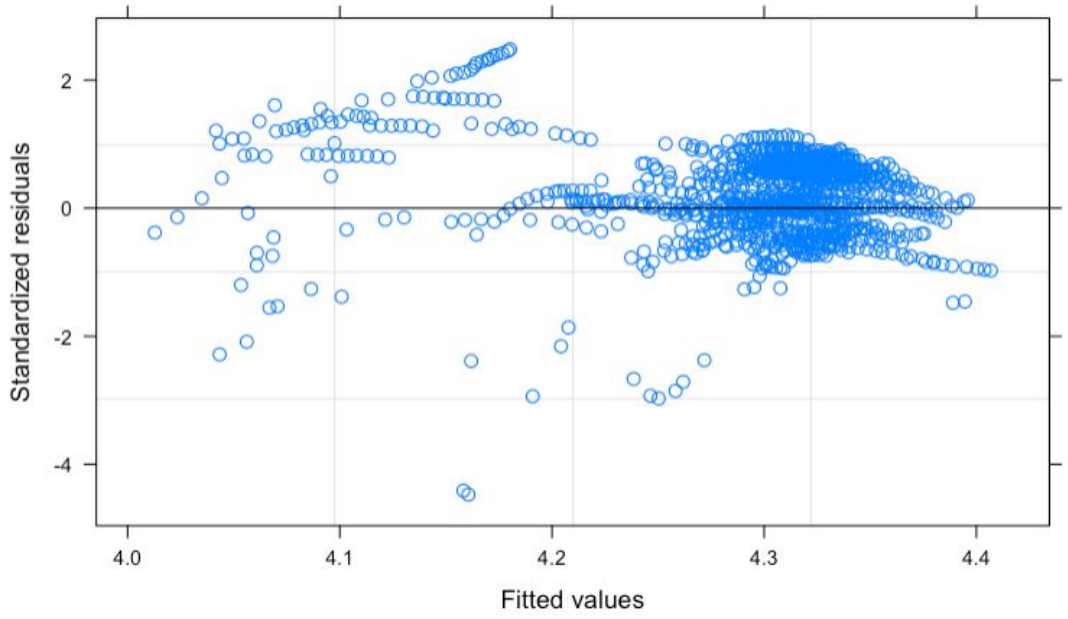
Exploratory data analysis

- Residual Analysis before transformations



Exploratory data analysis

- Residual Analysis after transformations



Exploratory data analysis

```{r}
vif(lmall)
```

Fossil_Fuels	log_co2	Overweight	log_fertility
1.497934	1.728013	1.825351	1.096047
GDP	log_pop	School_Enrollment	log_wlabor
1.104213	1.093718	1.082381	1.036934
log_internet	log_beds	Government_Effectiveness	log_wparliament
1.781132	1.028817	1.118236	1.106067
Political_Stability	Rule_Law	Voice_Accountability	
1.058078	1.146834	1.045391	

- Variance Inflation Factor shows very little/no multicollinearity

Statistical methods

Overall F-Test

- P-value for overall F-test: 8.303119e-130
- Doesn't address which variables are most important, but does say at least one is super important!

Transformations:

- Performed log transformations about about half of the variables
- Some variables had negative values so we couldn't transform them

Simple Linear Regression

- Used simple linear regression to identify relevant variables and determine the need for transformations before continuing on
- Can't see how predictor variables interact with each other by just looking at one at a time

Final regression model

log(Predicted life expectancy) = 3.89 - .00009*Fossil Fuels - .00007*log(CO2) + .005*Overweight - .0059*log(Fertility) - .000038*GDP + .024*log(Population Density) + .028*School Enrollment - .00075*log(Women's Labor Participation Rate) + .0035*log(Internet Usage) + .0019*log(Hospital Beds) + .0013*Government Effectiveness + .000053*log(Women's Participation in Parliament) + .00075*Political Stability + .0027*Rule of Law - .00038*Voice and Accountability

Statistical inferences and results

Countries that have a 10% higher rate of female participation in the labor force, but have the same values of every other predictor, tend to have 1.00007 years higher life expectancy.

Countries that have a 10% higher rates of females in legislative bodies, but have the same values of every other predictor, tend to have a life expectancy of 1.000000 additional year.

A country that consumes 1% more of their energy from fossil fuels, but have the same values of every other predictor, tends to have a life expectancy of 1.00009 years less.

A country that has a score of 1 higher on a scale of -2.5 to 2.5 of government effectiveness, but have the same values of every other predictor, tends to have a life expectancy of 1.001 additional years.

A country with 1% higher school enrollment, but have the same values of every other predictor, tends to have a life expectancy of 1.028 additional years.

Limitations

- Use of broom.mixed package resulted in slight variations of functions and their output
  - glance() for overall F-test did not provide a p-value, had to calculate by hand
- The data was incomplete.
  - Lots of NAs
- The dataset had negative numbers which made it difficult to transform.
  - GDP, Political Stability, Rule of Law, etc.

CONCLUSIONS

- Lots of moderately relevant variables, but none particularly stood out
- Average result of 1 year change in life expectancy
- There is opportunity for further research to be done in the field of predicting life expectancy.
  - The ESG Data set has more variables we didn't analyze
  - Could do more specific research by region or smaller timeline

Thanks!