Creating a Grade

I am trying to come up with a scoring system for the ESG index we are creating in class.

This means I need a standard way to measure each metric on a scale. I suppose the way people understand scales is a grading system of 0 to 100%. In life higher grades general mean Good and low grades mean Bad. So in the case of CO2 emissions someone who contributes 0 to Global emissions would get a perfect score of 100% and someone who contributes a lot would get less.

I have already created the Percentage of Global Contribution for each country so all I need to do is subtract it from 100. In the case of the US they contribute 23% so their score would be low at 77%. However even that seems too gracious. It seems I should apply a curve as well.

A logical way to do this is with a bell curve with 5 sections from A-F.

However how to I scale the rankings into this grading system based on a distribution?

In this scenario virtually every country should get an A grade except for the real outliers who should get an F.

I can’t just sort and divide equally because that ignores magnitude.

It seems like I need to look at the differences for each value from the mean. In percentage terms the average percent contribution to Global C02 is something like .05%.

A lot of countries are a lot less than that, and some are very much more.

So say I have that distance from this ‘mean’. The distances have to be less than 100.

If the mean was 50 the max difference would be +50 and – 50, or 50 in absolute terms.

But again it feels like I come back to needing a measure for dispursion.

Suppose I had a set of Prices

And I wanted to find outliers

Maybe I would say show me prices under 5, over 95, and divide the rest into 30, 30, 30 buckets. Why?

What if the range of prices was 75 to 100. The range is 25. Divide the range into 5 equal parts?

What if instead of bucketing on percentages I bucket on absolutes?

What are ‘good’ measures of these values?

What if I come up with a rating scheme that leaves this dbags out? Then divide THAT by 4 and put the dbags in the last bucket? What’s so bad about that? It just depends on the data right? Call that my methodology and call it a day.

Lets say for CO2 %

10+ F

6<10 D

4<6 C

1<4 B

< 1 A

Got it! Both US and China show a consistent score of F but US shows a steady decline while China shows a steady increase.

I would add an adjustment factor that credits the Country a level in rank if they show a 0.5 decrease and penalize them if they show an increase > 0.1.

PRESENTATION NOTES

**Quantifying Good and Evil: Using Data to Create an ESG Scoring System**

We all have causes that are important to us:

[Natalie Holloway murdered in Aruba] [Organ Harvesting in China] [Health impact of Chernobyl]

We all want to make the world a better place, or at least not subsidize those who are making it worse.

Our project was about creating a framework to analyze

That is what ESG Scoring is about: Creating a standard framework for measuring performance on specific Issues, broken down by Environmental, Social, and Governance factors.

**What is ESG?**

Our group wanted to explore ways to use Data to do that. This is referred to as Environmental, Social, Governance investing or ESG.

The World Bank is an International Financial Institution tasked with supporting projects across the Developing world. As part of their operations they have amassed a massive data set

The World Bank is a global

* Eradicate Extreme Poverty and Hunger
* Achieve Universal Primary Education
* Promote Gender Equality
* Reduce Child Mortality
* Improve Maternal Health
* Combat HIV/AIDs, Malaria, and Other Diseases
* Ensure Environmental Sustainability

But how to do it? And can we use data and technology to create a framework to help us do that?

Our group wanted to explore ways to use data to make the world a better place.

**What Data is Out There, and How to Access It?**

* The World Bank Open Data Project: WBData API
* Functions for search\_indicaors, get\_data, get\_dataframe
* Pandas to do Data Scrubbing
* Matplotlib and Plotly to perform Visualizations

**Creating our own ESG Scoring System**

Methodology

1. Identify representative factors across Environmental, Social, and Governance
2. Assess the Quality and Scope of Data Coverage for the identified factors
3. Create a Normalized Scoring System for each measure: 5 Tiers A-F
4. Establish a calculation frequency
5. Score each Country based on Relative Performance Each Year
6. Weight and aggregate individual country measures into a Global ESG Score

Results

* Colorized Pie Chart of ESG
* Line Chart of Historical Environmental Performance for China vs US
* Global Aggregate Performance

**Post-Mortem and Next Steps**

* The amount of data available is staggering. The World Development Bank datasets alone have MILLIONS of series going back decades. However, the depth of coverage and quality varies dramatically. David and Matt were correct – we spent over 80% of our time reviewing and cleaning data
* Far more statistical analysis needs to be done on the available metrics to identify those sufficient coverage and unique descriptive power
* Many of the Distributions were not Standard. They were typically very skewed distributions and further analysis is required using Clustering to improve the rating system and ensure that Countries are not unfairly bucketed
* Adjust Scores to Reward/Penalize based on Trend

NOTES

Look how fortunate we all are in this room. Here we are in New York City, one of the great cities of the world, at one of the top universities on the planet, having the opportunity to learn new skills that we are interested in and that will help us be more successful.

It is easy to take all this for granted.

For example, I feel very passionate about human rights and about people being able to live free from fear.

We all watch the news and see people boycotting different Companies and Countries on the basis of their performance on various topics.

You may feel strongly about the environment and the effects of Climate Change and want to do what you can to promote these beliefs through your investment choices.

Creating a Personal ESG Investment Framework

Our goal was to create a framework that would allow us to measure the relative performance of every country around the world on a

Establishing an ESG Framework

The first step in quantifying you

**ESG Framework**

We analyzed data sets across the following categories and

* Environmental  
  + Natural Resources – Forest Acreage
  + Pollution – CO2 Emissions
  + Energy – Fossil Fuel %
  + Wildlife Protection – Threatened Species
* Social
  + Human Rights
  + Health
  + Education
  + Exclusion
* Governance
  + Political – Property Ownership
  + Legal – Prison Population
  + Corruption – Rule of Law

A screenshot of a map

Description automatically generated

Over 80% of people in the world have access to Electricity

A screenshot of a map

Description automatically generated

Global Emissions of Carbon have Exceeded 34B Metric Tons per Year

A screenshot of a cell phone

Description automatically generated

Worldwide, over 90% of Young Women Complete Primary Education

**Global CO2 Emissions**

![A screenshot of a cell phone

Description automatically generated]()

**Global CO2 Emissions**

A screenshot of a cell phone

Description automatically generated

* Elephants in the Room: China CO2 output is almost double the next highest producer
* With Long Tails: Emissions for high producers 46x greater than lesser producers

**CO2 Emissions for China vs US**

A close up of a map

Description automatically generated

* While US CO2 Emissions are still among the highest in the world levels are DECREASING
* China CO2 output outpaced the US in 2005 and are still INCREASING

A fair measure should reward positive behavior

**Sources**

* Data
  + MSCI ESG Government Ratings: <https://www.msci.com/documents/10199/e092c439-34e1-4055-8491-86fb0799c38f>
  + BP Statistical Review of World Energy: <https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html>
  + Doing Business Data Catalogue: <https://datacatalog.worldbank.org/dataset/doing-business>
  + Quandl World Bank Governance Indicators: <https://www.quandl.com/data/WWGI-World-Bank-Worldwide-Governance-Indicators/documentation>
  + Our World in Data Renewable Energy: <https://ourworldindata.org/renewable-energy>
  + Atlas of Sustainable World Bank Development Goals 2018 <https://books.google.com/books?id=89RdDwAAQBAJ&pg=PT107&lpg=PT107&dq=EG.ELC.ACCS.RU.ZS&source=bl&ots=2qTVZRuBd0&sig=ACfU3U1mM2ExRtOJXWZ4N5yhoMjdglRMdg&hl=en&sa=X&ved=2ahUKEwjr7NqUnb7kAhVBwlkKHUMpBAAQ6AEwB3oECAgQAQ#v=onepage&q=EG.ELC.ACCS.RU.ZS&f=false>
* Technology
  + World Bank API: <https://wbdata.readthedocs.io/en/latest/fetcher.html>
  + Plotly: <https://plot.ly/python/bubble-charts/>
  + Stack Overflow: <http://www.stackoverflow.com>

**Appendix: Using the World Bank Data API**

List the Hundreds of Available Databases

A screenshot of a cell phone

Description automatically generated

Perform Global Searches for Individual Series

A screenshot of a cell phone

Description automatically generated

Retrieve a Pandas DataFrame A screenshot of a cell phone

Description automatically generated

JSON Data Package

A screenshot of a cell phone

Description automatically generated