Initial Data Analysis

AUTHOR
Zehaan Naik

Data Description

We've collected the data for a total of 17460 countries using two methods:

1. BDP: This query method gave us most amount of data for all the variables but resulted in similar values for all 5 years. This leads me to believe that this query method is not suited for obtaining time-series data. However, this gave us good insights on our values of interest for the year 2023 and would be useful for some time static analysis.
2. BDH: This query method gave us sparse data for only 6 of the 11 values of interest. However this data showed different values for each year which leads us to believe that this method is good to collect time series data. We will use this to track growth and change across time for several variables.

Variables of Interest:

Following are the variables available in BDH or time query data.

1. Country
2. CEO Compensation
3. Board average age
4. CEO Average tenure
5. Chairman Average tenure
6. Board Average tenure

Following are the extra variables available in the BDP data.

1. CEO Duality
2. Board Size
3. % of independent board members
4. % of female board members
5. Number of board meetings in a year

For the time query data, the number of non-NaN data points in the data set are 3221.

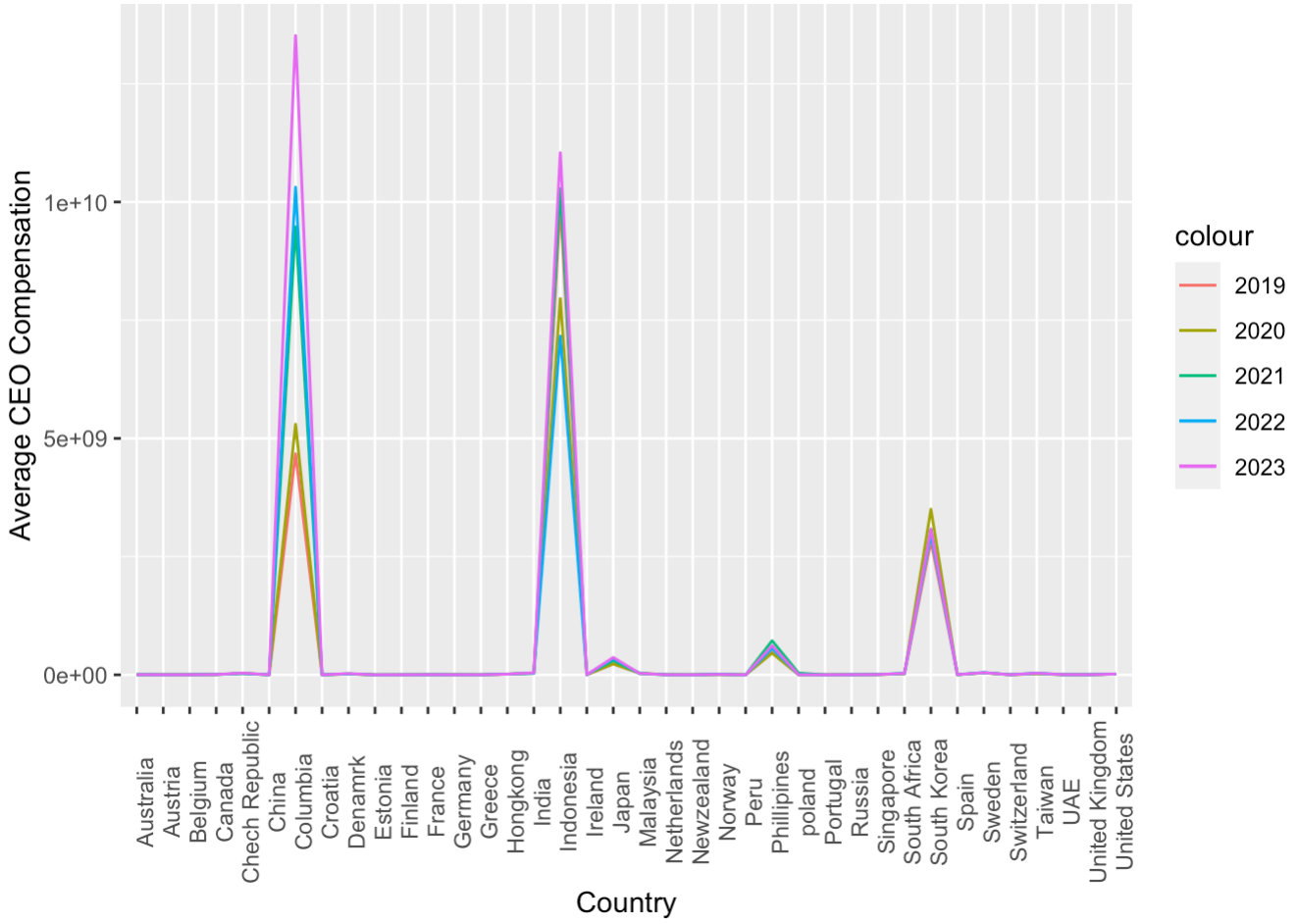
```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```
[1] 3221
```

Country Analysis

The first bit of analysis that we want to perform on this data is to group the data based on the country of origin and try to observe trends in the growth of CEO compensation as a metric for the growth of the company.

The following plot describes the average compensation of a CEO in various countries across 5 years:



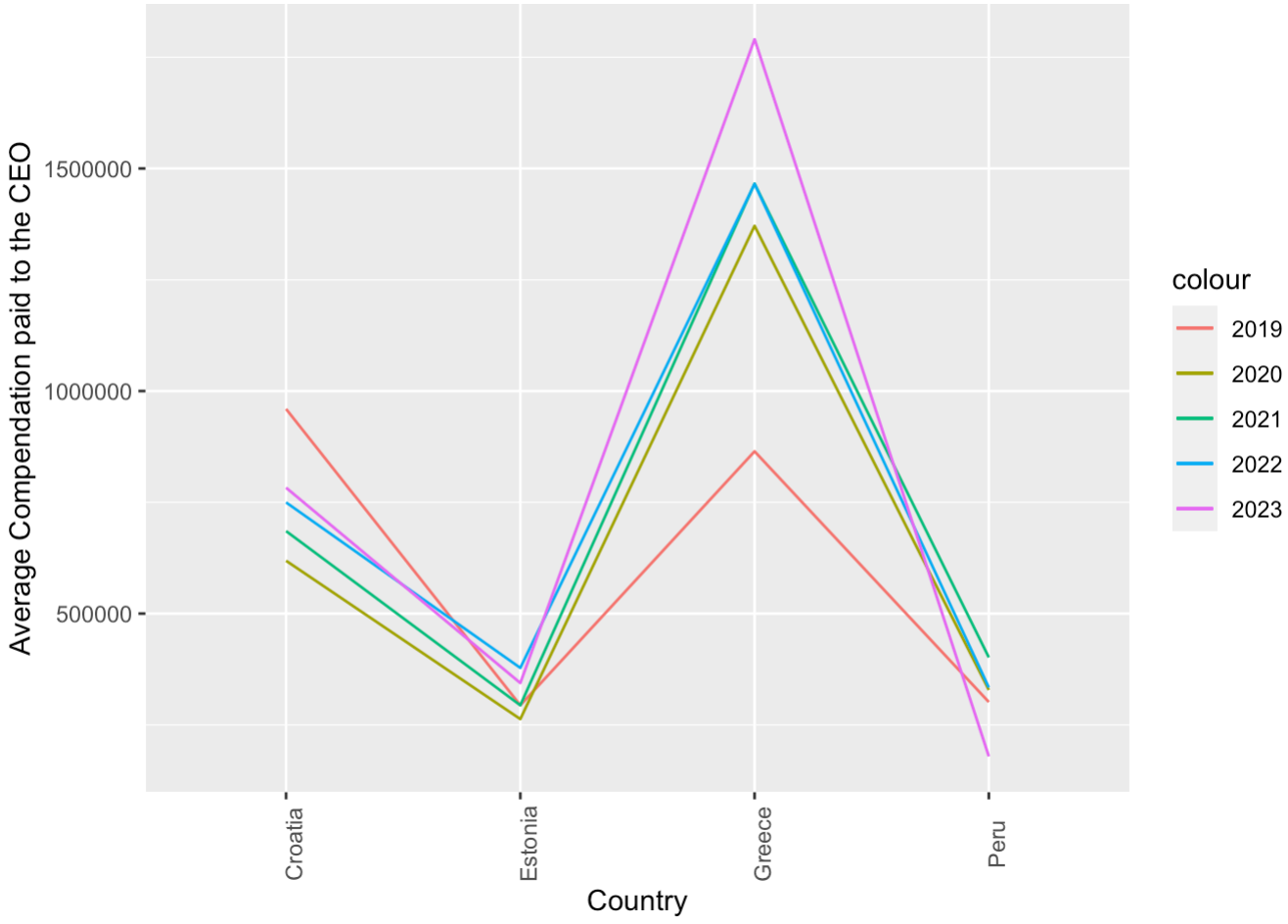
We observe that there are a few outliers in our data and this is more evident in the quantile analysis of the compensation values:

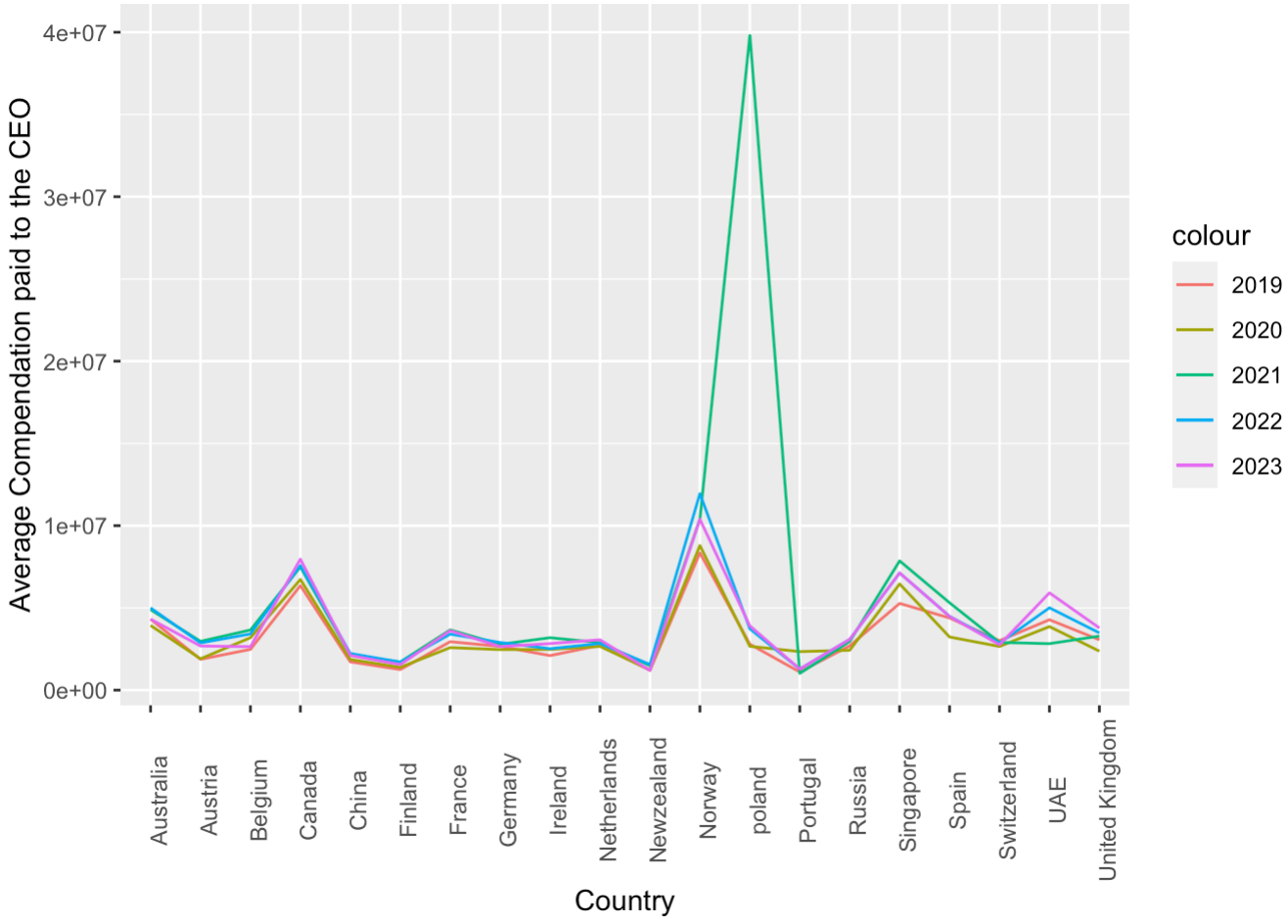
```
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max. 
2.940e+05 2.195e+06 4.299e+06 4.846e+08 2.368e+07 9.895e+09 
```

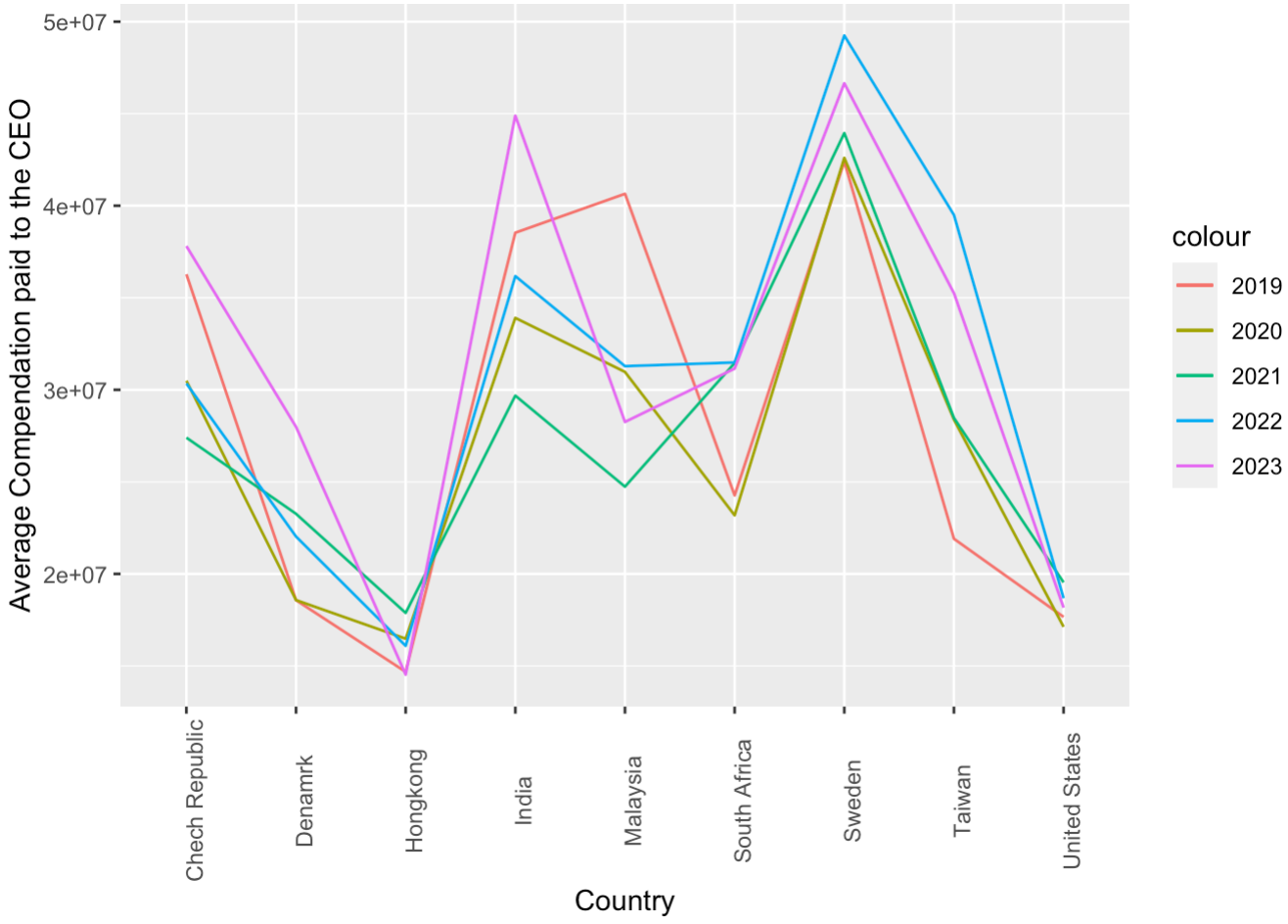
Hence, we divide the countries into 4 classes:

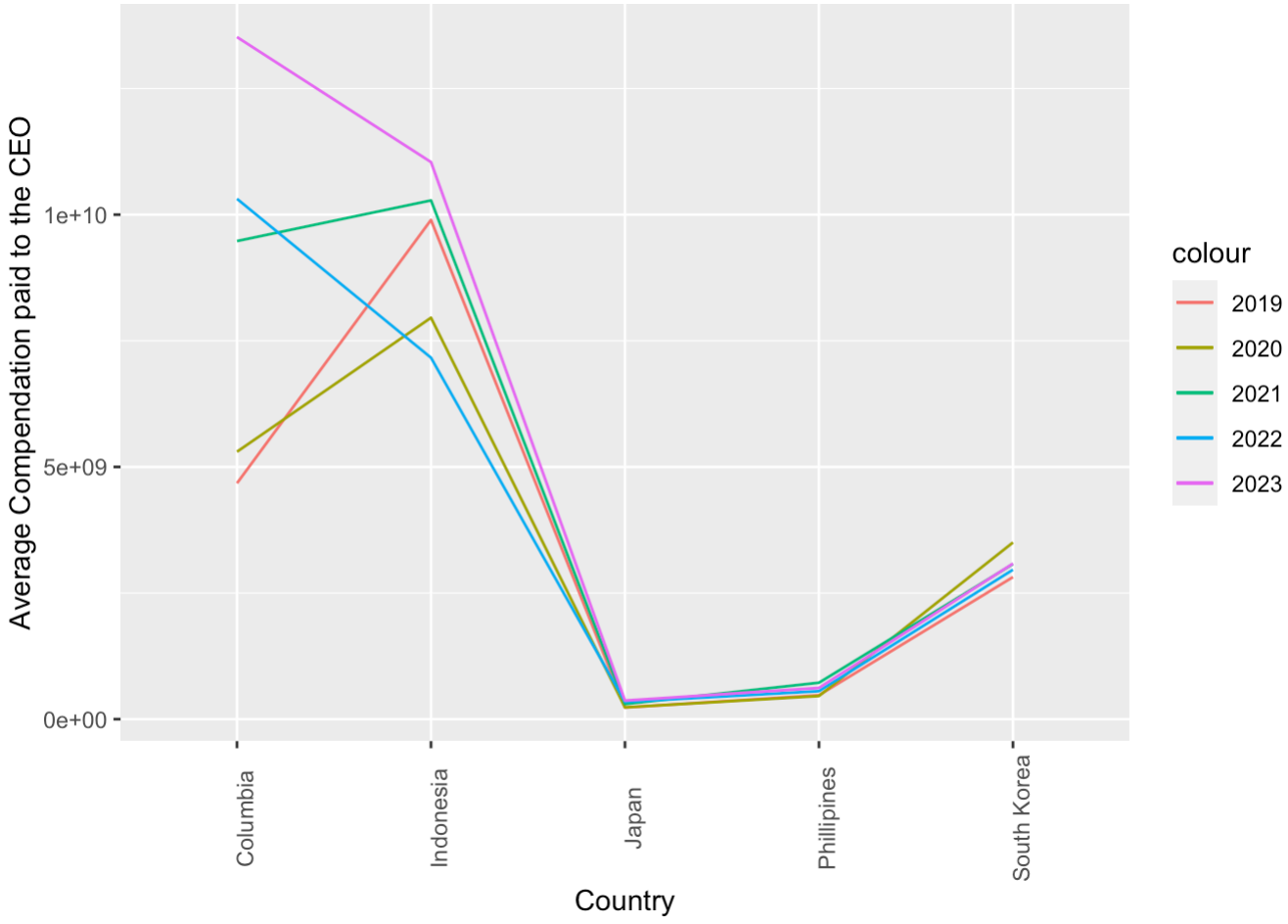
1. Average compensation in the range 1e5-1e6
2. Average compensation in the range 1e6-1e7
3. Average compensation in the range 1e7-1e8
4. Average compensation greater than 1e8

The following plots better illustrate the comparative between the CEO compensation over the years between various countries.


Average Compendation paid to the CEO
1500000
1000000
500000
Croatia
Estonia
Greece
Peru
Country
colour
2019
2020
2021
2022
2023

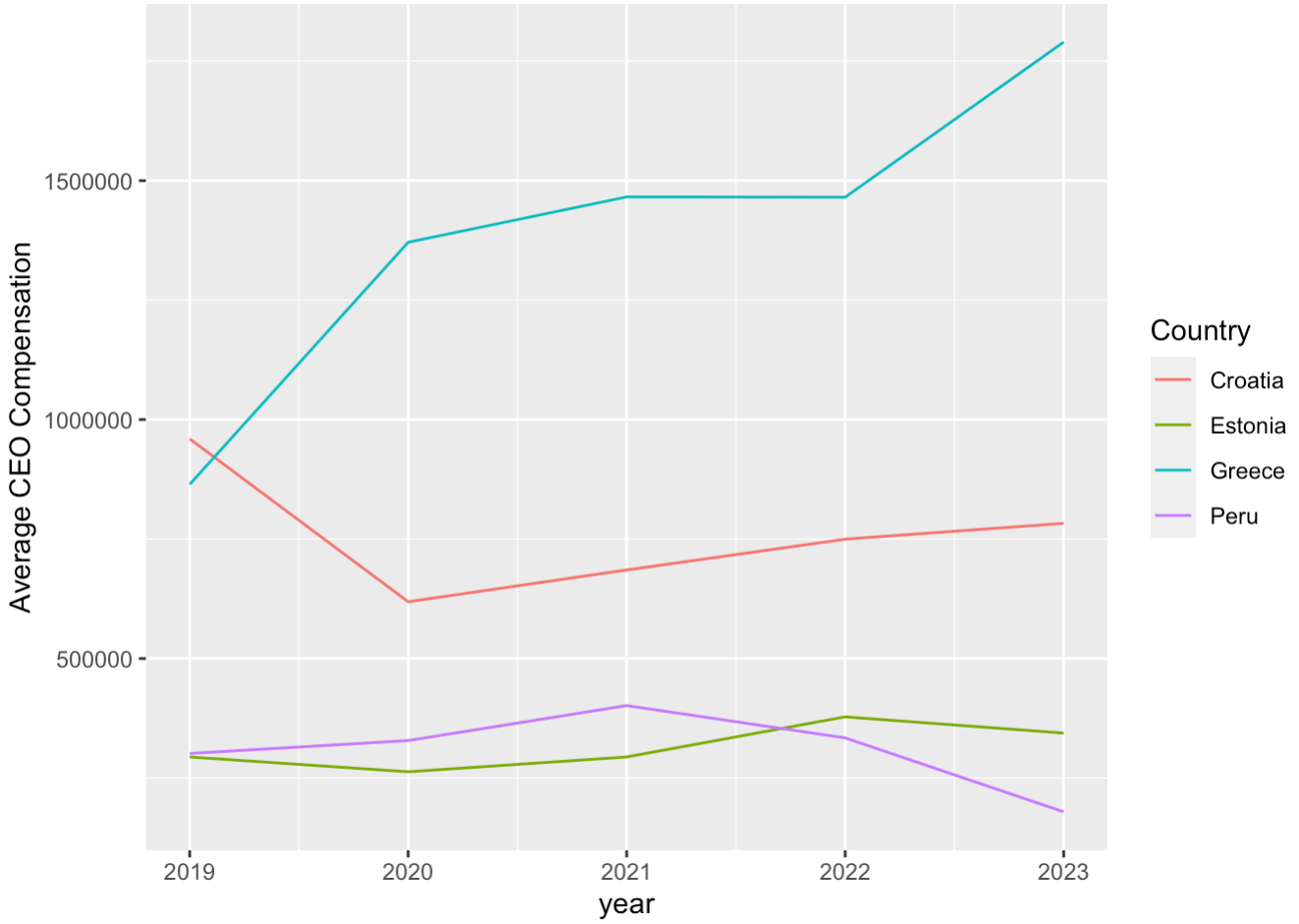

Average Compendation paid to the CEO
4e+07
3e+07
2e+07
1e+07
0e+00
Australia
Austria
Belgium
Canada
China
Finland
France
Germany
Ireland
Netherlands
Newzealand
Norway
poland
Portugal
Russia
Singapore
Spain
Switzerland
UAE
United Kingdom
Country
colour
2019
2020
2021
2022
2023

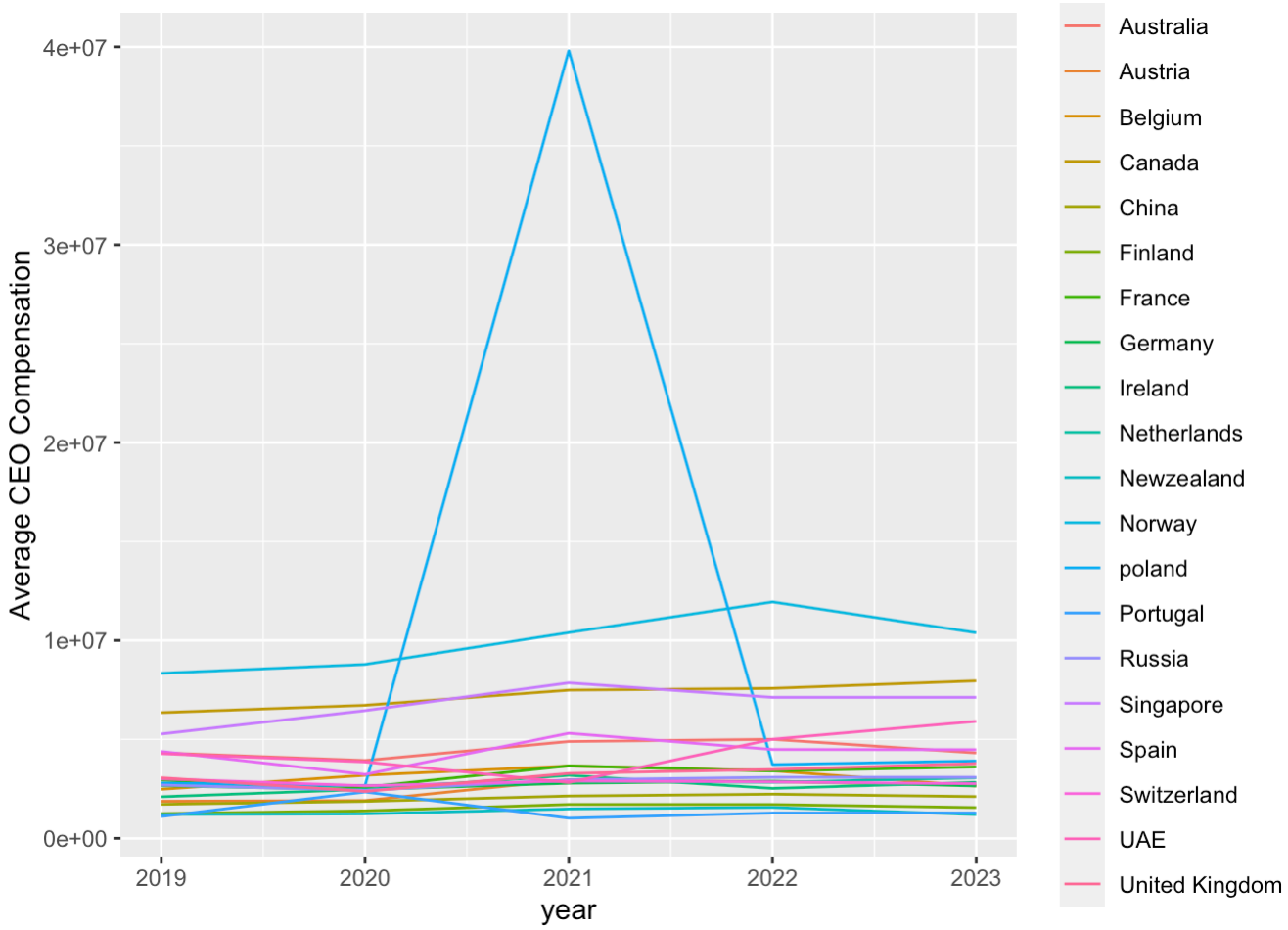


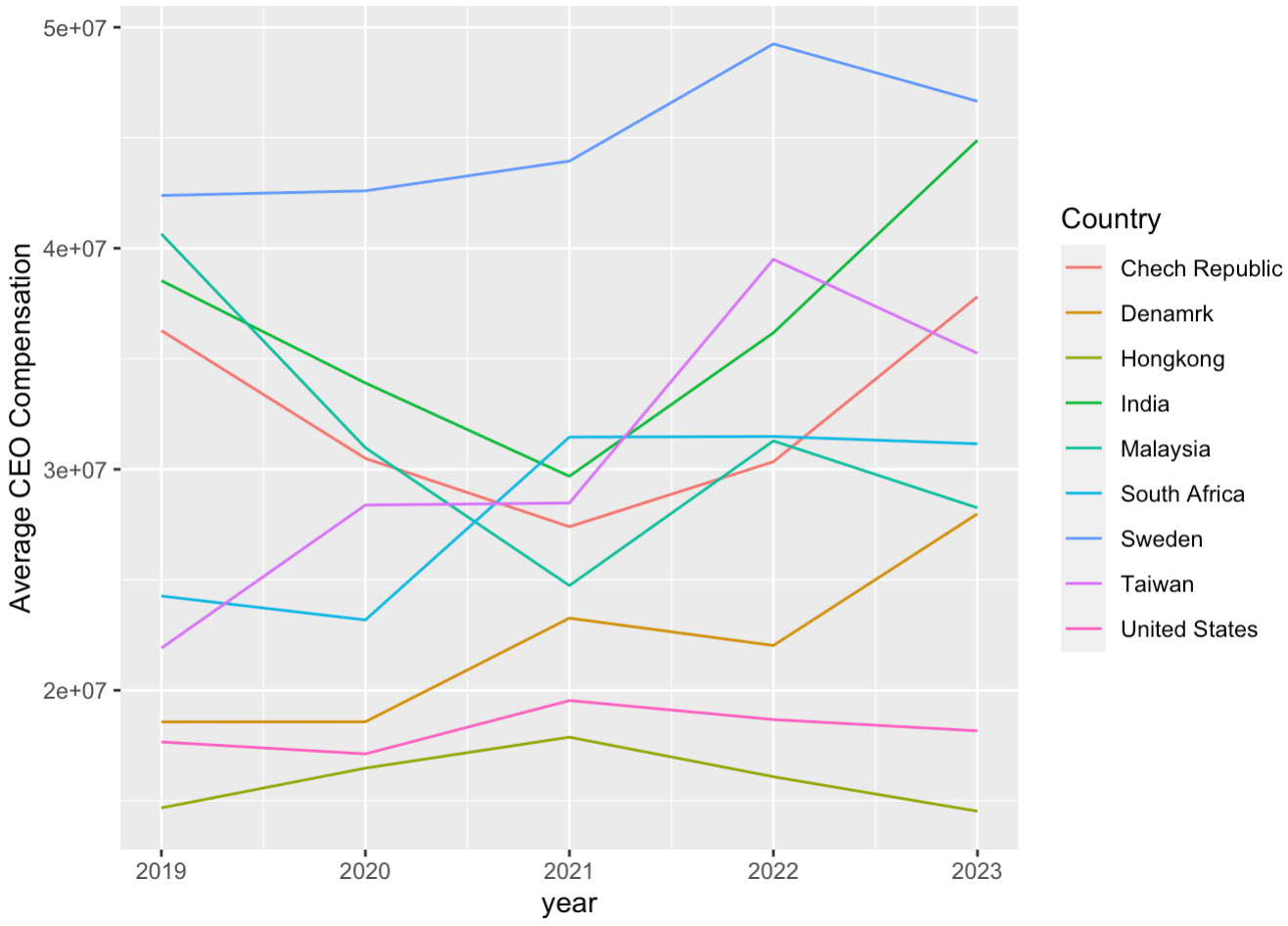


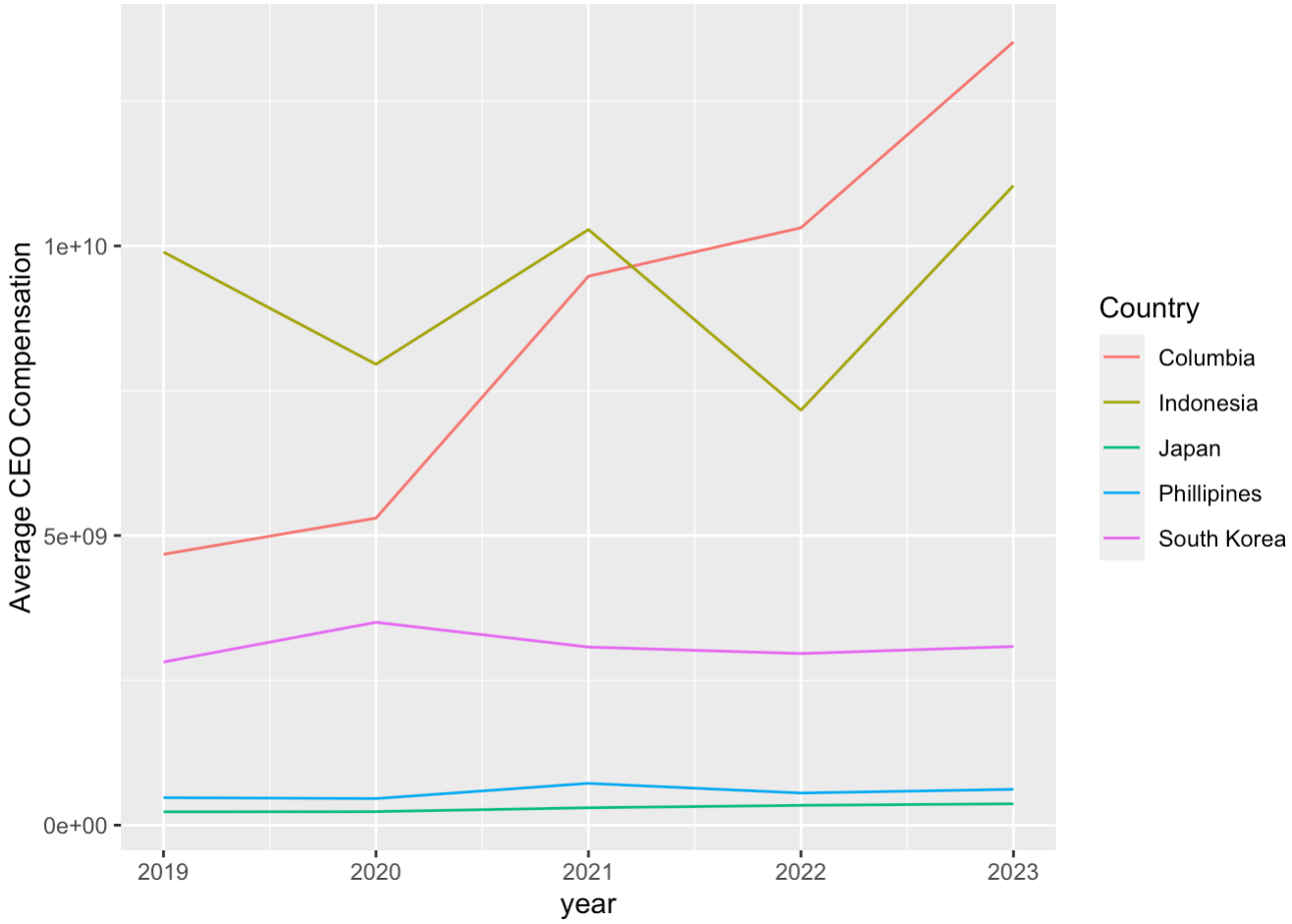
Another interesting analysis can be drawn from the following plots:

These plots show the rise or fall of CEO compensation in various countries over the course of 2019-2023. These trends motivate us to reason this fall or growth with the other parameters that we have at our arsenal.




Average CEO Compensation
4e+07
3e+07
2e+07
1e+07
0e+00
2019
2020
2021
2022
2023
year
Australia
Austria
Belgium
Canada
China
Finland
France
Germany
Ireland
Netherlands
Newzealand
Norway
poland
Portugal
Russia
Singapore
Spain
Switzerland
UAE
United Kingdom


Average CEO Compensation
5e+07
4e+07
3e+07
2e+07
2019
2020
2021
2022
2023
year
Country
Chech Republic
Denamrk
Hongkong
India
Malaysia
South Africa
Sweden
Taiwan
United States


Average CEO Compensation
1e+10
5e+09
0e+00
2019
2020
2021
2022
2023
year
Country
Columbia
Indonesia
Japan
Phillipines
South Korea