

Internet's impact on the World

By: Group 7



Content

01

Overview

02

Analytic Goals

03

Process Taken and Data Clean Up

04

Data Analysis

05

Other Goals driven by our Analysis

06

Conclusions

Our Team



Andrew
Carson



Manuel Parra



Alghalia
Alsammak



Jie Zhu



Syed Shah

Analytic Goals

- 1. Does Internet usage has an effect on tax revenue Worldwide?**
- 2. Does Internet usage has something to do with Carbon emissions?**
- 3. Does Internet usage drives electricity consumption?**



Tools & Technologies

This section details the various software applications and statistical analysis tools utilized throughout the project. It covers the methodologies applied to process, analyze, and visualize the data, ensuring accurate and insightful interpretations of global internet usage trends.



Python



Mean, Standard Deviation, Linear Regression, Correlation Coefficient



Pandas, Numpy



Matplotlib



Github

Process Taken

DATA CLEANING

Step 1

DATA FRAMING

Step 2

IN-DEPTH DATA
ANALYSIS

Step 3

Data Clean Up and Data Frame Setup

World Bank DataSet Description

Data range: 1960-2022

Countries: 268

Features: 48

Areas:

- macro-economical
- social
- political
- environmental data

Data Source: <https://www.kaggle.com/datasets/nicolasgonzalezmunoz/world-bank-world-development-indicators>



Project Data Frame(s) Setup

```
#Extract specific columns for analysis
#df_project = df[['date','country','population', 'individuals_using_internet%']]
df_project = df[['date','country', 'individuals_using_internet%']]

#Check for missing values and drop them
df_clean = df_project.dropna()

#Final project dataframe for analysis
df_clean
```

	date	country	individuals_using_internet%
30	1990-01-01	Afghanistan	0.000000
31	1991-01-01	Afghanistan	0.000000
32	1992-01-01	Afghanistan	0.000000
33	1993-01-01	Afghanistan	0.000000
34	1994-01-01	Afghanistan	0.000000
...
17265	2017-01-01	Zimbabwe	24.400000
17266	2018-01-01	Zimbabwe	25.000000
17267	2019-01-01	Zimbabwe	26.588274
17268	2020-01-01	Zimbabwe	29.298565
17269	2021-01-01	Zimbabwe	34.813971

8044 rows × 3 columns

7 Key Dimensions Deep Dive

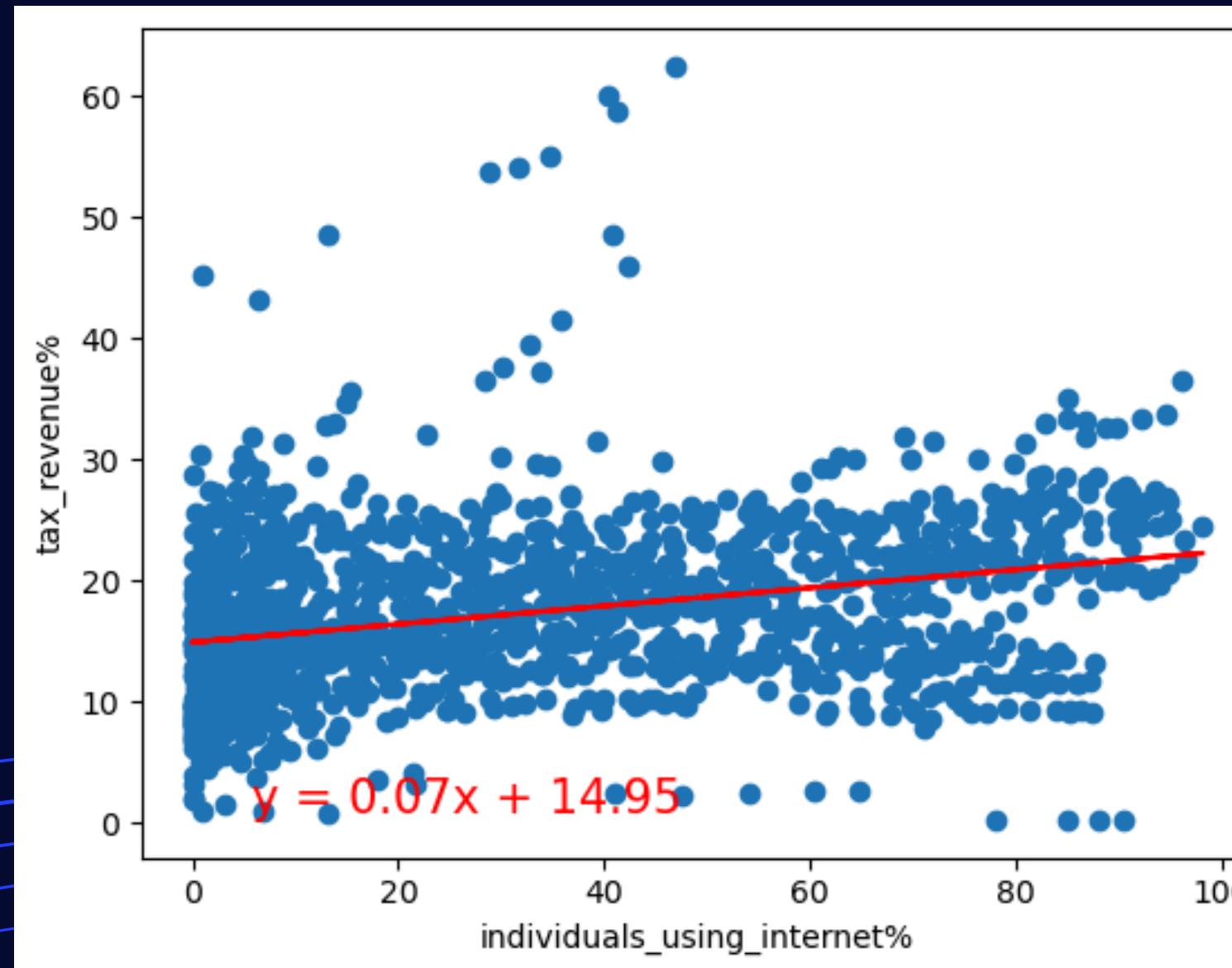
- Individuals_using_internet
- Electric_power_consumption
- Tax_revenue%
- Control_of_corruption_estimate
- Death_rate
- Intentional_homicides
- Life_expectancy_at_birth

```
<class 'pandas.core.frame.DataFrame'>
Index: 1106 entries, 228 to 17260
Data columns (total 12 columns):
 #   Column           Non-Null Count Dtype  
 --- 
 0   date             1106 non-null   datetime64[ns]
 1   country          1106 non-null   object  
 2   population       1106 non-null   float64 
 3   individuals_using_internet% 1106 non-null   float64 
 4   tax_revenue%     1106 non-null   float64 
 5   electric_power_consumption 1106 non-null   float64 
 6   control_of_corruption_estimate 1106 non-null   float64 
 7   CO2_emissions    1106 non-null   float64 
 8   access_to_electricity%    1106 non-null   float64 
 9   death_rate        1106 non-null   float64 
 10  life_expectancy_at_birth 1106 non-null   float64 
 11  intentional_homicides   1106 non-null   float64 
dtypes: datetime64[ns](1), float64(10), object(1)
memory usage: 112.3+ KB

df_clean.columns
Index(['date', 'country', 'population', 'individuals_using_internet%', 'tax_revenue%', 'electric_power_consumption', 'control_of_corruption_estimate', 'CO2_emissions', 'access_to_electricity%', 'death_rate', 'life_expectancy_at_birth', 'intentional_homicides'], dtype='object')

# Country Data Frame Setup
df_countries = df_clean.groupby('country').max().sort_values(by='individuals_using_internet%', ascending=False)
```

1. Does internet usage has an effect on tax revenue Worldwide?



Initial Thoughts

- Historically, tax collection has historically been a difficult process, as it would require outsourcing tax collectors to go and collect the taxes, and many taxes would fall through the cracks. But with the internet, more taxes could be collected more efficiently.
- However, this is offset with economic liberalization, as collecting high taxes is not actually the goal of a government. The government collects taxes to provide services to the population. But the government extracting money from the populace slows the economy down. And the slower the economy, the slower its economic growth

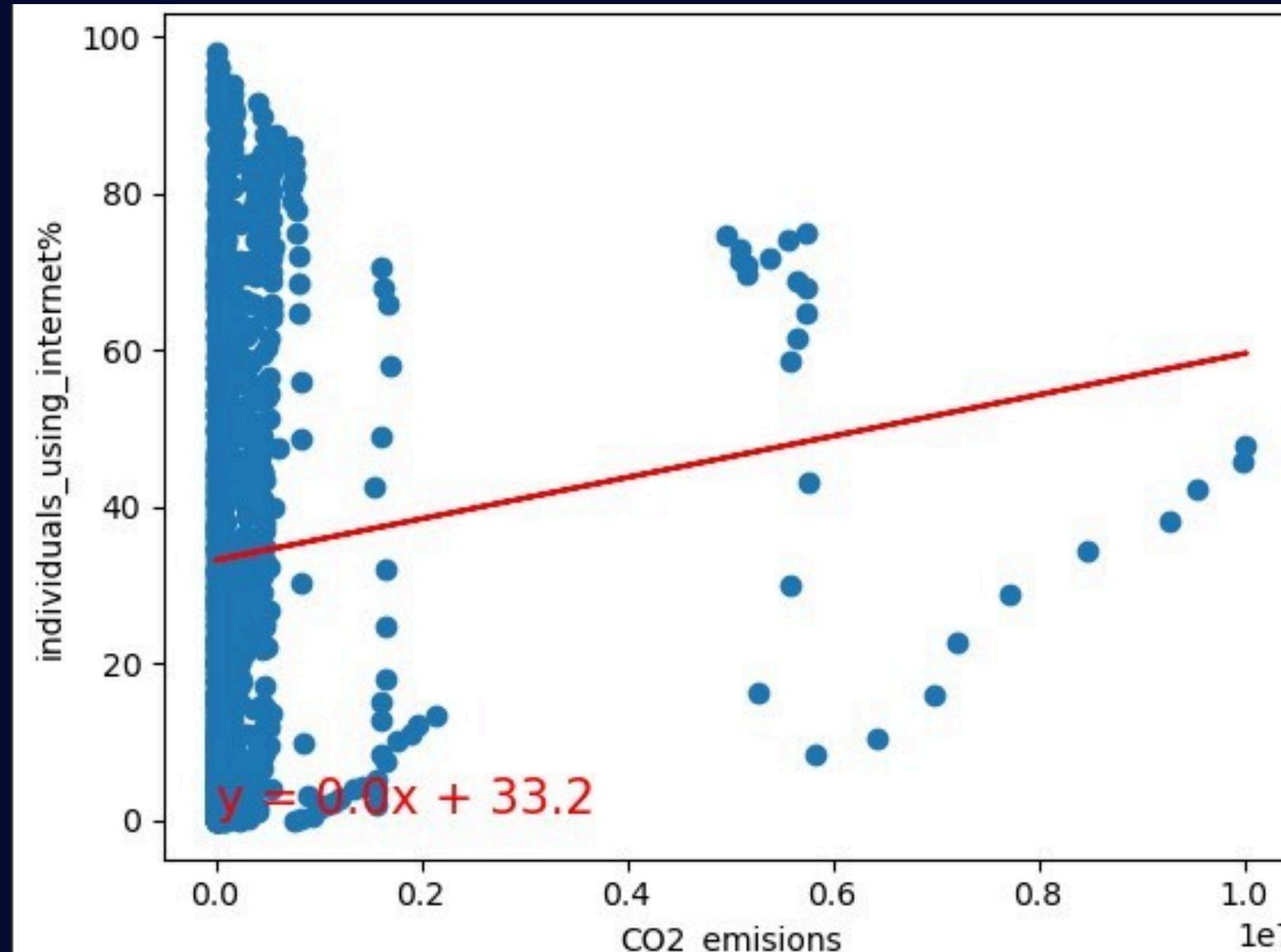
Quantitative Analysis

- The slope is 0.07 and y intercept is 14.95
- Correlation coefficient is low at .30

Qualitative Analysis

- There seems to be a very weak correlation
- The variables individuals using internet and tax revenue do not go hand in hand

2. Does Internet usage has something to do with Carbon emissions



Initial thoughts

- One would assume that the more intense usage, the more developed they are, and the more carbon emissions

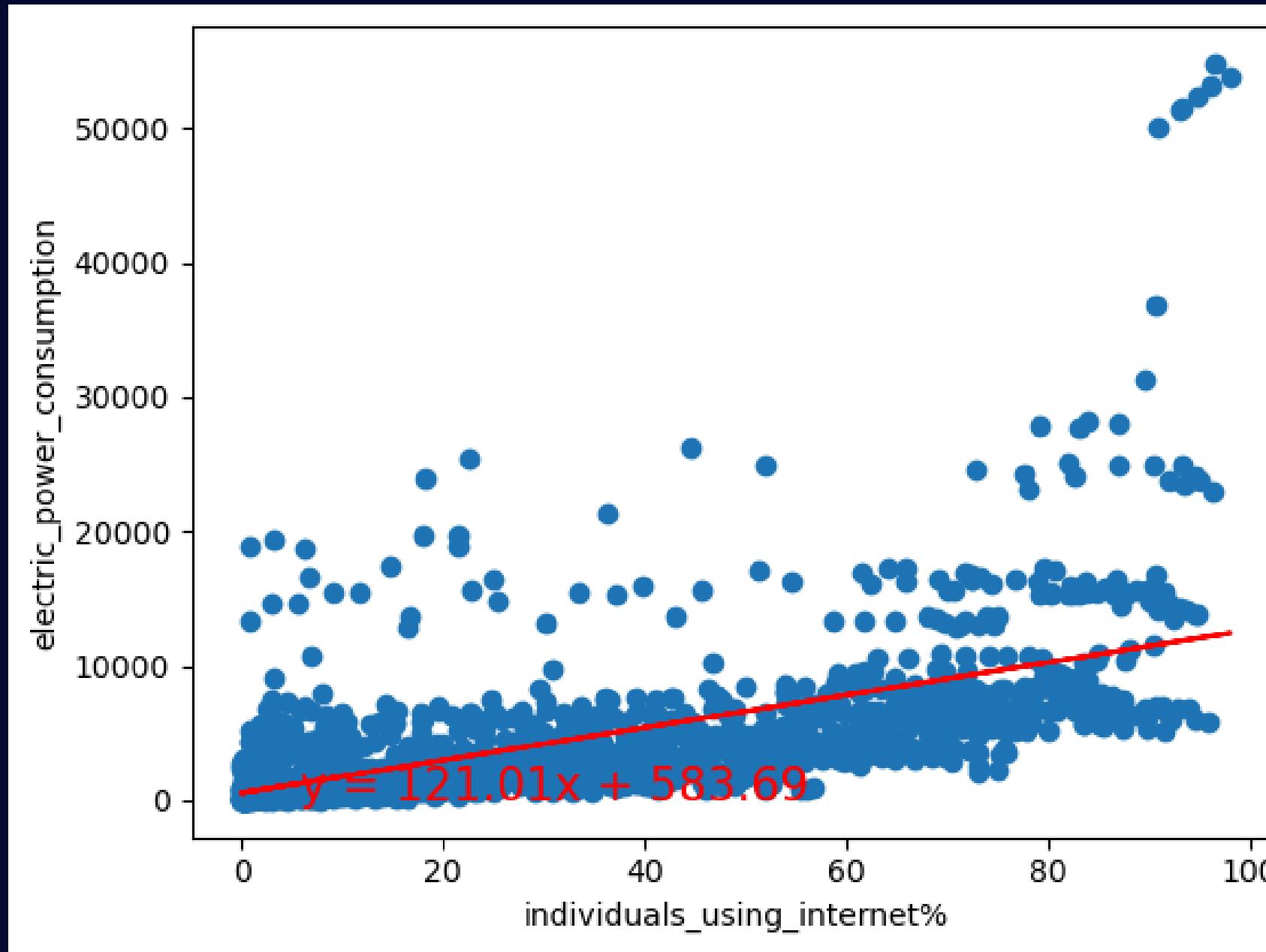
Quantitative analysis

- The slope is 0 and y intercept is 33.2
- Correlation coefficient is low at 0.08
- Most countries have very low emissions across the board. The countries that do have higher carbon emissions are the outliers, not the norm.

Qualitative analysis

- There is no real correlation
- The variables individuals using internet and CO2 emissions do not go hand in hand

3. Does Internet usage drives electricity consumption?



Initial Thoughts

- Internet usage might correlate to more electric consumption via computers and database usage
- Internet usage may also be linked to more sophisticated or advanced economies which naturally demand more electric need and therefore more advanced supply

Quantitative Analysis

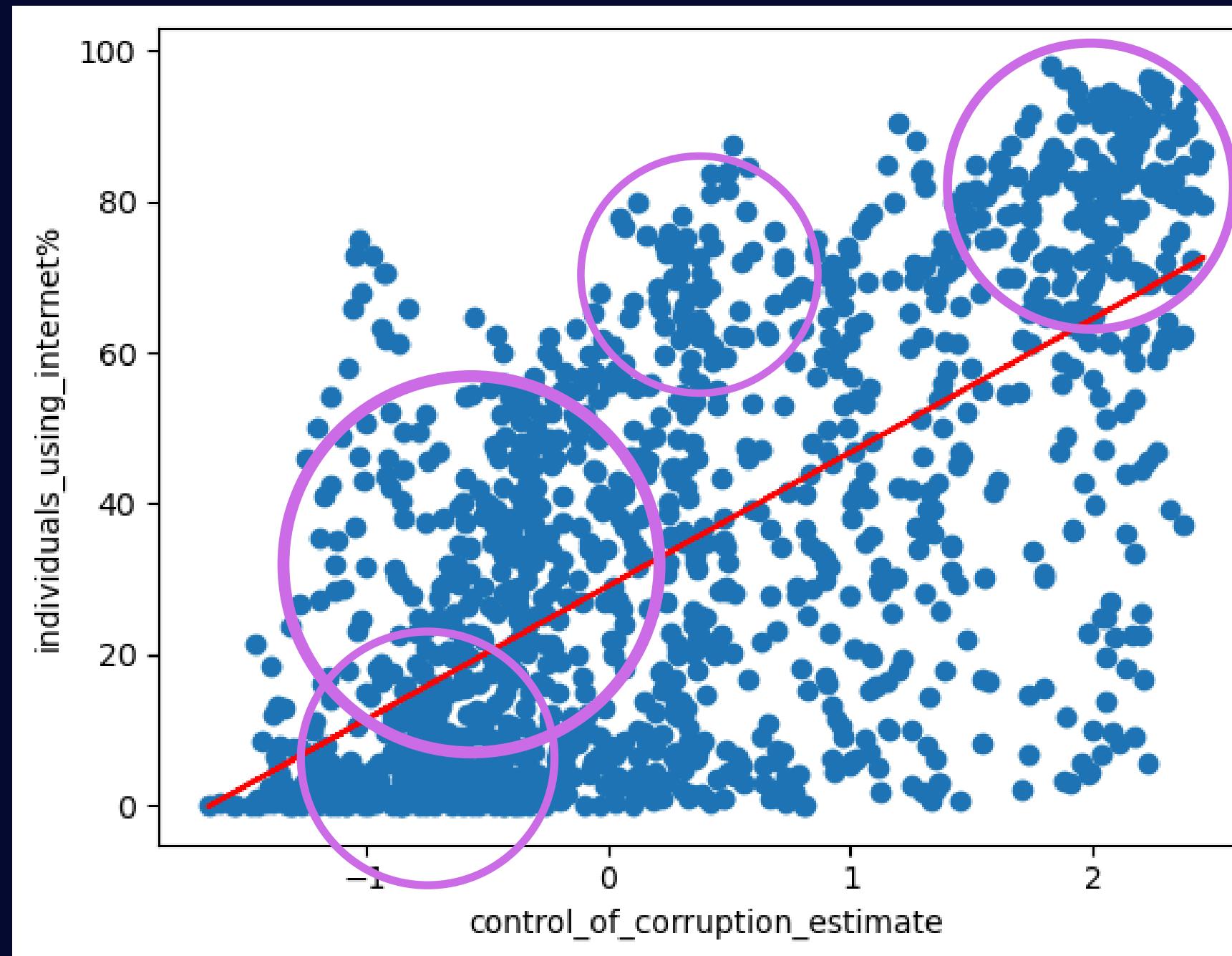
- The slope is 121.01 and y intercept is 583.69
- Correlation coefficient is moderate at 0.58
- In the top right hand corner there is a tight grouping of outliers that have extremely high consumption

Qualitative Analysis

- The correlation is not as strong as we would expect
- Internet usage does to some degree move with electric consumption however the effect is not pronounced
- The infrastructure that predicts power consumption is not always tied to the infrastructure that provides internet usage (they do not go hand in hand)

Other Goals driven by our Analysis

1- How Internet usage impacts corruption?



Initial Thoughts

- We hypothesized that more access there is to information the less corruption is able to flourish. Countries such as North Korea point to the extremes of this trend.

The Graph

- The slope is .8 and y intercept is -4.5
- Correlation coefficient is strong at .618
- Countries appear to very spread out, considerably evenly across the grid of the chart
- There are several 'groups' or 'clumps' of countries in purple that might suggest there are several ways to consider this relationship. Perhaps corruption relates to the internet in 3 or 4 standard predictable styles.

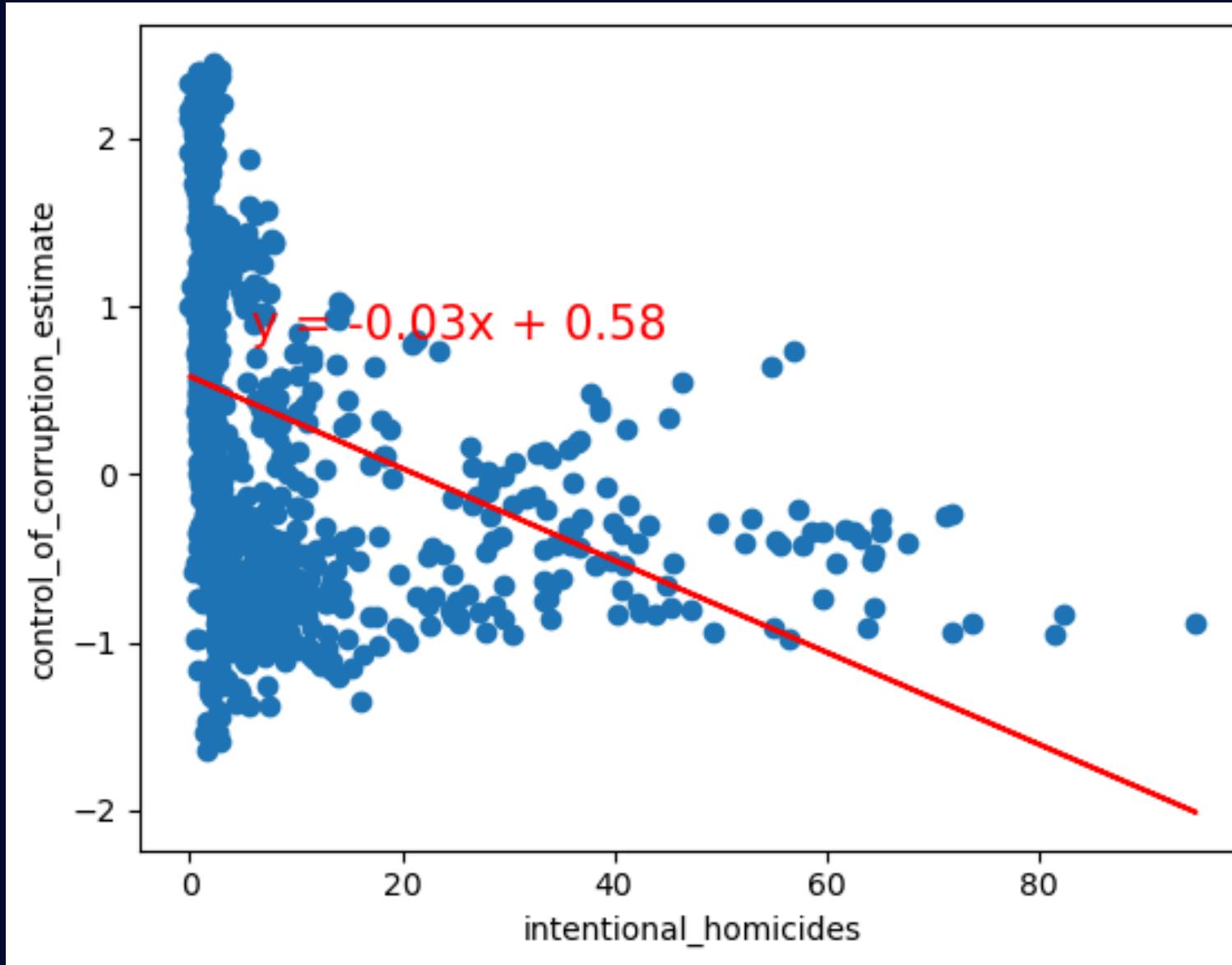
Real World Conclusions

- We see in the bottom left extremely low access to the internet (<20%) leads to higher predictable corruption
- However corruption can also exist just as reliably from 30% - 60% internet usage, perhaps in more hypothetically stratified societies where privileged classes enjoy unfettered access and other classes are locked out allow for certain forms of corruption.
- We also have to question the underlying data collection and how accurately it reflect the nuances of corruption worldwide

Other Goals driven by our Analysis

2- How corruption affects homicides?

Initial Thoughts



Quantitative Analysis

Independent Value (x) : Intentional homicides

Dependent Value(y): Control of Corruption estimate

- Slope: -0.03
- Intercept: 0.58
- Correlation Coefficient: -0.33

Hence there is **no positive correlation**, which means that when X increases, Y doesn't increase

Qualitative Analysis

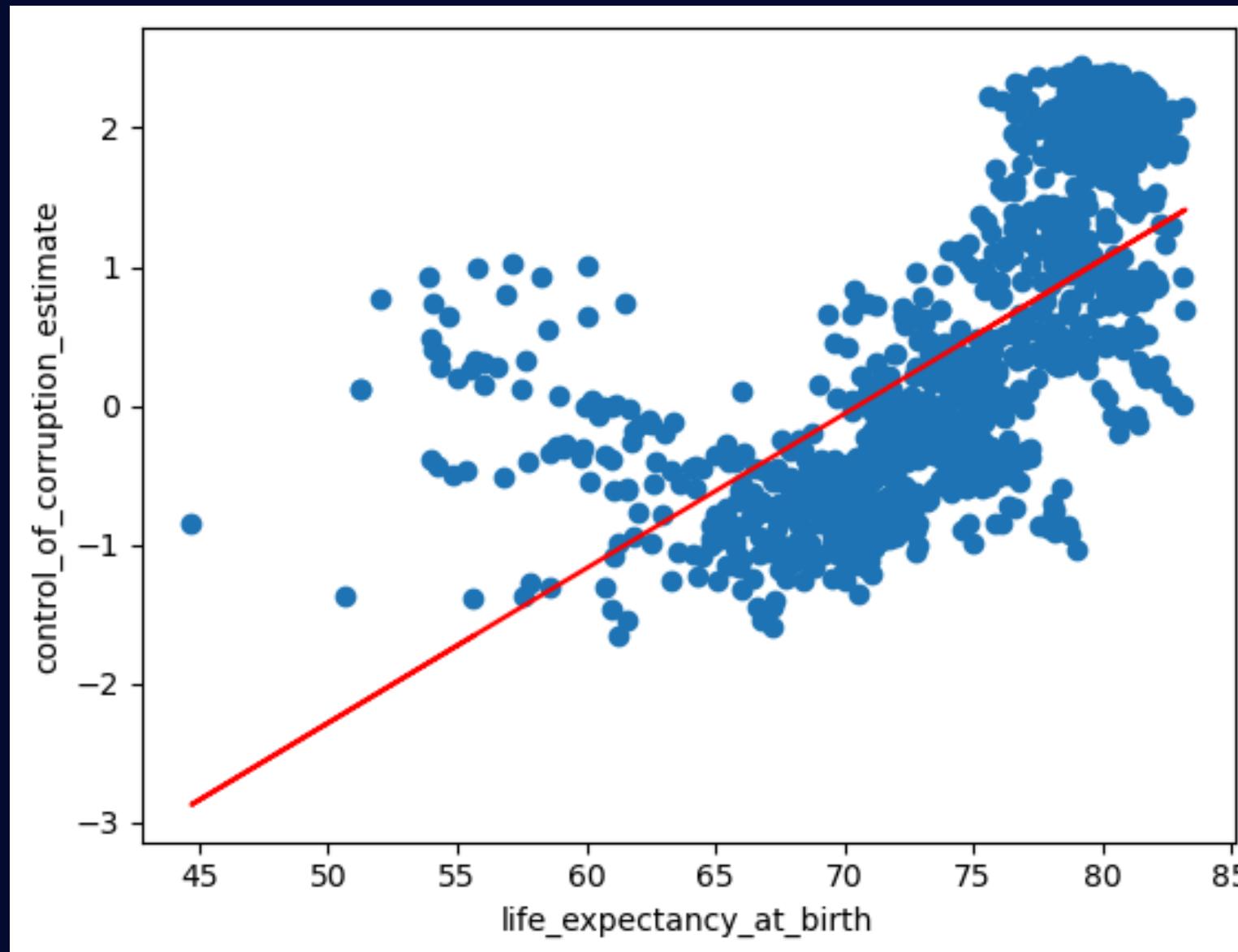
The Dynamics of corruption requires parties playing together, for a mutual benefit (against the law), which needs low homicides to function, it is our thought that this dynamic explains the lack of correlation

Other Goals driven by our Analysis

3- How Life expectancy is affected by corruption?

Initial Thoughts

Quantitative Analysis



Independent Value (x) : Life Expectancy

Dependent Value(y): Control of Corruption estimate

- Slope: 0.11
- Intercept: -7.84
- Correlation Coefficient: 0.65

There is **positive correlation**, which means that when X increases, Y doesn't increase

Qualitative Analysis

Countries with low corruption, can afford better medical services, and overall have more open societies which tend to guarantee a better life and a longer one.



Conclusions

1- Does internet usage has an effect on tax revenue Worldwide?
Low Correlation - Correlation Coefficient: 0.30

2. Does Internet usage has something to do with Carbon emissions?

Very Low Correlation - Correlation Coefficient: 0.08

3. Does Internet usage drives electricity consumption?

Moderate Correlation - Correlation Coefficient: 0.58

Honorable mention:

How Internet usage impacts corruption?

Strongish (High) Correlation - Correlation Coefficient 0.65

Thank You!

