Web Visualizer for Functional Map Of The World

Chinedum Ajabor
Computer Engineering and Sciences
Florida Institute of Technology
Melbourne, United States
cajabor2018@my.fit.edu

Ethan Banks
Computer Engineering and Sciences
Florida Institute of Technology
Boca Raton, United States
ebanks2019@my.fit.edu

Grant Butler
Computer Engineering and Sciences
Florida Institute of Technology
Melbourne, United States
gbutler2020@my.fit.edu

Michael Gourdine
Computer Engineering and Sciences
Florida Institute of Technology
Melbourne, United States
mgourdine2019@my.fit.edu

Michael Hon
Computer Engineering and Sciences
Florida Institute of Technology
Melbourne, United States
mhon2014@my.fit.edu

Riad Terrence Mohammed
Computer Engineering and Sciences
Florida Institute of Technology
Melbourne, United States
rmohammed2019@my.fit.edu

Matthew Dominic Rampersad
Computer Engineering and Sciences
Florida Institute of Technology
Melbourne, United States
mrampersad2019@my.fit.edu

Thomas van Haastrecht
Computer Engineering and Sciences
Florida Institute of Technology
Melbourne, United States
tvanhaastrec2015@my.fit.edu

Tyler Zars
Computer Engineering and Sciences
Florida Institute of Technology
Melbourne, United States
tzars2019@my.fit.edu

***Abstract*—The Functional Map of the World (FMOW) dataset[1] is a large dataset of satellite images augmented with additional information on where and when the images were taken, as well as the circumstances, such as cloud cover or the sun's position at the time of the image, and contents, such as buildings. An interactive web visualizer was constructed which shows on a map where images are located and allows a user to see the satellite images at those locations. To gain a better understanding of the data, exploratory data analysis (EDA) was performed on the dataset. While the images themselves did not contain much useful information, their labels provided some extra information on the locations of important landmarks, such as power plants and sports fields. This allowed the processing of calculating correlations between these landmarks and other social, economic, or climate-related phenomena, from other datasets.**

***Index Terms*—Big Data, Visualization, FMOW, EDA**

I. Introduction

Working with datasets that span the world comes with many different challenges. Oftentimes the data is not uniformly distributed, which makes it difficult to get unbiased results. More developed countries had significantly more information than less developed countries. Different locations are affected by more than just their location. Different countries have access to varying amounts of resources and have different cultures, which impact every facet of life. The goal of this research into the functional map of the world (FMOW) dataset was to find out how useful this dataset is as a representation of different things on Earth. A web visualizer similar to road pavement temperature datamaps[2] was constructed to view and explore the images of the dataset including the location and additional metadata. To that end, few research questions was initially produced ,to observe if a satisfactory conclusions could be made from the dataset, augmented with other datasets.

We looked into the following:

- Is obesity linked to the number of recreational facilities in a country?
- Does the proximity of nuclear power plants affect an area's air and water quality?
- Are nuclear power plants responsible for a rise in cancer cases?
- Do countries produce food and power at similar rates?
- Do countries with a high Gross Domestic Product (GDP) value higher education more?

II. Literature Review

A. Nuclear Power and Environmental Impacts

While the public perception of nuclear power is less than stellar[3], the actual harm caused by modern power plants is quite minimal. A 2011 study comparing the environmental impacts caused by both fossil fuel and nuclear power plants showed that nuclear power is much cleaner than fossil fuels[4]. In fact, for the foreseeable future it seems to be the only feasible green and economic energy solution. While there are differences between nuclear fusion and fission plants, the results seem to suggest that both options would be beneficial in comparison to fossil fuel solutions. Other green energy solutions simply lack scale, as they are not capable of meeting the energy requirements of today's society, at least not on their own. As such, nuclear energy should be a safe and more environmentally friendly alternative to fossil fuel plants.

B. Nuclear Power Plants and Cancer

As nations across the globe continue to work on implementing renewable and clean energy, nuclear power has become

a main player to provide this sustainable power. Nuclear energy has plenty of benefits but some health effects from radioactive material are present; this paper breaks down the health of children around the area of nuclear power plants analyzing outliers and trends in the health of the youth[5]. While the study mainly focuses on persons in their childhood, the conclusions drawn by the authors hold true with members of any generation showing that there are no extra cancer or health issues when individuals live within the bounds of a nuclear power plant. The reference group of individuals the authors studied was in a 15 to 50 kilometer radius of power plants, while the focus of the study looking for negative effects was less than 15 kilometers away from a nuclear power plant. This study concluded that even those individuals less than five kilometers away from a nuclear power plant were at no risk to their health that wasn't also present at 50 kilometers away. This dataset was collected over a period of 30 years from 1974 to 2004 with no changes in normal health of anyone in the region near the nuclear power plants.

C. Obesity and Recreation Centers

According to the World Health Organization (WHO), overweight and obesity are defined as abnormal or excessive fat accumulation that presents a risk to health. Of central concern, a body mass index is used to measure these conditions, whereby a BMI over 25 is considered overweight and over 30 is seen as obese. The article reference studied and monitored children in the age range 5-8 years old with approximately 30 recreation centers, specific to the San Diego County, California. The results showed no significant fluctuation in subjects' BMI, except for girls, who experienced a slight decrease in BMI[6]. This paper expanded the San Diego research on a global scale by collecting obesity and recreation center data from a vast list of countries. It was observed that a slightly negative correlation to the abundance of recreational facilities in those countries were evident. The correlation coefficient was -0.0309. The slightly negative correlations means that possibly, the more recreational facilities in a country, the lower their BMI and relative obesity levels.

D. Food and Energy Production Comparison

During times of inconsistent weather patterns as a result of climate change, and ongoing wars, countries are being made to turn their attention towards ensuring not only food security, but also energy security for their population. In an effort to address food security, the World Bank Group has allocated approximately $30 billion to assist countries such as Jordan, Bolivia and Ghana in projects which will see their agricultural infrastructure receive much needed strengthening [7]. Moreover, when it comes to energy security, measures taken to ensure this would vary based on the source of energy, whether it be oil, natural gas, nuclear or electricity. For example, to achieve natural gas security, apart from boosting production, countries must also focus in obtaining investments in distribution infrastructure and improving LNG shipping capacity[8]. Our chosen research topic aims to highlight whether countries have prioritized one of these concerns over the other, or whether they tend to go hand in hand.

E. GDP and HEI

While every country has a differing number of Higher Education Institutions (HEI)[9], this paper focuses on countries in the European Union. This paper takes into effect the idea of HEI's not existing at all in the European Union and the effects this would have on the various countries Gross Domestic Product (GDP) and the countries ability to preserve through times of turmoil. This paper concluded that the countries emulated with no HEI's would have a roughly 11% lower GDP than those same countries if they had HEI's. This data correlates with the idea of how countries with more institutions for higher education will have a larger GDP than those countries with less HEI's. While the paper does conclude that the simulated environments may be counteracted by HEI's in other countries, the positive correlation in the end between the country having more HEI's and higher GDP's is telling of the importance of HEI's.

III. Proposed Approach or Methods

Fig. 1. Process Workflow

For each of the research questions, additional datasets were needed in order to augment the FMOW dataset, so that the information on the different topics was being covered. Depending on the data used, restructuring or filtering of the data was needed. By following a data analysis workflow model, steps were accomplished allowing us to ensure that each group member was working in tandem towards the same end goal. While data was gathered from different sources, it was all cleaned and group members worked together to help share cleaned data and information they were gathering from their exploratory data analysis. The FMOW web visualization was built off of FMOW data and after testing was deployed at the end of the project.

Data Gathering was done by collecting all the JSON object containing the metadata of all of the images. The metadata contained image file names, image box dimension, category, ground sampling distance, raw location ,UTM zones, country code, cloud coverage, pand resolution, multi resolution, target azimuth, sun azimuth and nadir angle. Most of these metadata was cleaned off for the web application and data exploratory analysis. For web visualization, the dataset needed to be cleaned off and reduced in size thus only certain attributes remained such as file name, category and location as the purpose of information in the other metadata such as resolution and angle could not be determined. Data enrichment and aggregation was done with other dataset for the purpose of providing some results for our research questions for our hypothesis.

A. Nuclear Power Plant Pollution

For the question relating to the pollutive effects of nuclear power plants, air quality and water pollution dataset was utilized which covered the United States of America. As most countries do not have good enough data on these metrics, or access to nuclear power plants at all, the decision was made to only focus on the US. The air quality dataset had a collection of points from many different cities in the US, however it did not have air quality data directly surrounding the location of the power plants. As such, calculations on the distance from the nearest power plant to each of the air quality data points was done. Calculating the correlation between the distance to the nearest power plant and our two metrics for the pollution of the environment, would give us a good statistical measurement for how pollutive power plants really are.

B. Nuclear Power Plants and Cancer

The horrors of what can happen when nuclear power plants malfunction is well-known, but how much do these facilities affect our health when they are functioning properly? The goal of this question was to see if nuclear power plants are a possible cause for a rise in cancer cases. For this question, cancer data was used from the World Cancer Research Fund (WCRF), which listed the number of cancer cases per country. As data could not get more local than the country level, decision was made to aggregate the FMOW data. For each country, the count of the nuclear power plants was taken, so that correlates could be observersed with the rate of cancer cases for each country. The WCRF also had data on the deaths caused by cancer. Similar analysis was performed on this data.

C. Obesity and Recreation

We wanted to see if the abundance of recreational facilities where people could engage in physical activity had a relationship to the obesity rates in countries with average Body Mass Index greater than 30. A Body Mass Index (BMI) is an indicator for if a person is underweight, has regular weight, overweight, or obese. We compared these countries' average BMI to the abundance of recreational facilities, specifically, basketball courts, golf courses, track and field centers, soccer fields, tennis courts, baseball fields, American football fields, and swimming pools, to see if there is a relationship between them.

D. Food and Power

After conducting the research necessary to examine the effects of power plants, decision to extend the investigation to determine how high a priority Energy Production was for various countries, as compared to Food Production. The dataset captured energy plants such as Nuclear Power plants, Wind Farms and Solar Farms within the various countries, as well as several Crop Fields. In order to compare the two categories, relevant columns and rows were filtered from the initial dataset, resulting in a dataset which consisted of only Energy Plants and Food Production. Each category was then tallied per country and the correlation was calculated.

E. How does GDP effect Higher Education Institutions?

Countries with a higher GDP are generally more developed than those with lower GDP. As access to higher education institutions is often costly, with many people being unable to afford to attend them due to financial limitations. However, HEI also serve as an investment in a country's economy as skilled laborers are able to be more productive and produce more value towards a nation's GDP. We compared the GDP of countries with the number of HEI located in them.

IV. Results and Discussion

A. Nuclear Power Plant Pollution.

The correlation between the distance to power plants, and the two pollutive metrics was negligible. As shown in the table below, the correlations were 0.015 and -0.017. Values were

TABLE I
CORRELATION

	Air Quality	Water Pollution	Distance
Air Quality	1.000000	-0.379472	0.014774
Water Pollution	-0.379472	1.000000	-0.016518
Distance	0.014774	-0.016518	1.000000

plotted on map, to observe whether the locations of the power plants seemed to affect the surrounding pollution levels. While the darkest area is on the East half of the country, which is also where most of the power plants are, the power plants are not surrounded by darker areas than the rest, the dark patches seem to be relatively uniformly distributed.

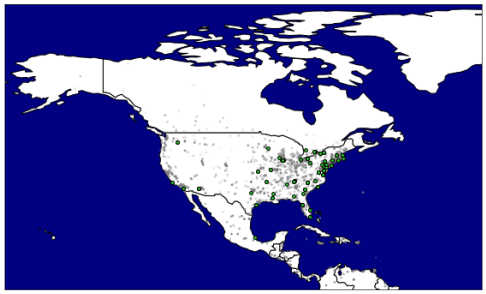

Fig. 2. Water pollution data and power plants in the US

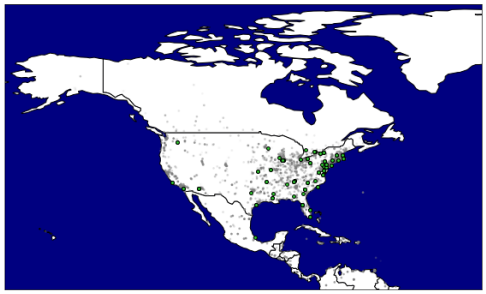

Fig. 3. Air quality data and power plants in the US

These results align with the current consensus on the pollutive effects of power plants. Power plants do not produce direct

carbon dioxide emissions [10], but instead have radioactive waste as a byproduct. While this waste is certainly not good for the environment, it is usually handled with care and stored in special storage casks, which prevents the waste from harming the environment. As such, it was expected for power plants to not have a large effect on the air and water quality of its surroundings.

B. The Relation between Power Plants and Cancer Cases

The rate of cancer incidents and the rate of cancer deaths were related to the number of power plants in a country, in order to find a correlation between these metrics. The correlation for cancer incidents was **0.244** which is a slightly positive correlation. The correlation for cancer deaths was much lower, at **0.023**. While there is a slightly positive correlation between the number of power plants and the rate of cancer cases, the correlation for cancer deaths is very close to 0.

These values do somewhat align with the results from Spix, Claudia, et al. (2008) [5]. Their study covers a more accurate dataset in a smaller area, but in both cases, proximity to a power plant leads to an increased risk of cancer. While one might expect a similar situation for cancer related deaths, this difference may be explained by the abundance of power plants in more developed countries. Countries with more power plants will be more likely to have a well developed health care system.

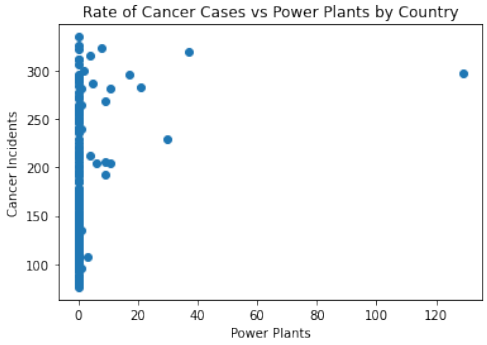

Fig. 4. Cancer Incidents

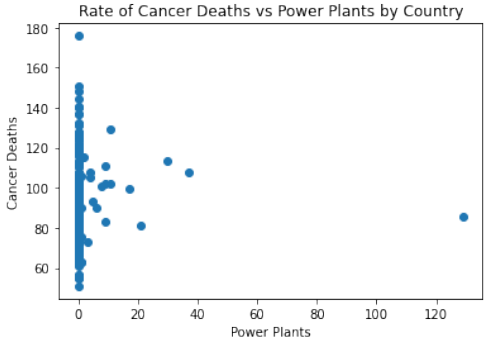

Fig. 5. Cancer Deaths

C. Obesity and Recreation.

After considering countries with average BMI greater than or equal to 30, slightly negative correlation was discovered to the abundance of recreational facilities in those countries. The correlation coefficient was **-0.0309**. The slightly negative correlations means that possibly, the more recreational facilities in a country, the lower their BMI and relative obesity levels. The correlation is not strong enough to conclude more than this. Possibly, with more abundant recreational facilities in a country, people tend to use them more to try and keep their obesity rates lower.

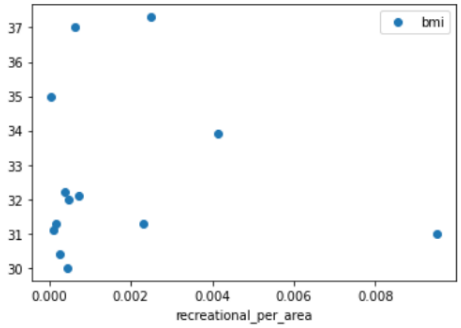

Fig. 6. BMI vs. Recreational Facilities

D. Food and Power

It was found that from the dataset, there was a correlation of **0.4** between the frequency of Energy Plants and Food Production which tells us that countries tend to focus on both Energy and Food production together. Moreover, the data pointed to China and the United States for having the most Energy Plants, and France and Italy for having the most Food Production. The following graph was plotted to visualize the data:

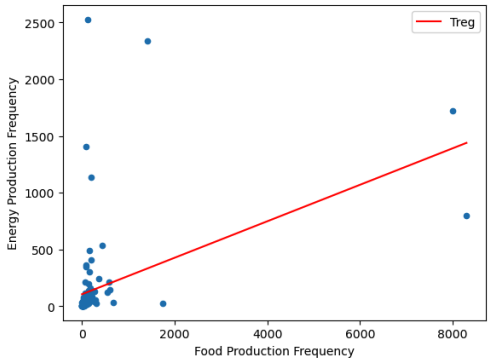

Fig. 7. Foods Vs. Energy

Hence, it is observed that most countries do not prioritize energy and food production over one another, but rather in parallel. However, as show in the graph, there are some isolated cases where some countries have demonstrated a clear preference for one over the other.

E. How does GDP influence higher education?

When a country has a greater number of higher education institutes, it leads to an increase in GDP and GDP Per Capita. While the correlation is not perfect, it is a **0.938** correlation that the more higher education institutions a country has, the higher its GDP will be. On the other side of GDP Per Capita, the correlation between it and higher education institutions is slightly positive correlation at **0.173**. Countries like the USA, China, United Kingdom, and Canada all have a high GDP and due to their large landmass they typically have a large number of higher education institutions to support their widespread population. On the flip side, the countries with high GDP Per Capita are smaller and with not such a huge population that there doesn't need to be a massive number of higher education institutions to educate the population of the country. The countries with high GDP Per Capita like Qatar and Luxembourg have a tiny number of higher education institutes but still have a massive GDP Per Capita explaining the lower correlation between GDP Per Capita and higher education institutions.

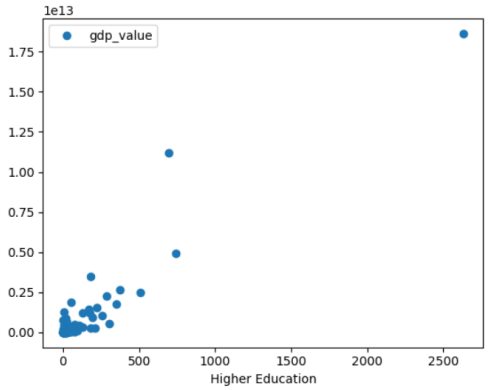

Fig. 8. Education vs. GDP Value

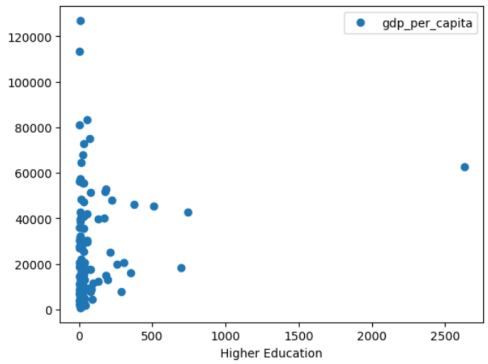

Fig. 9. Education vs. GDP per Capita

F. Web visualization

At the end of the analysis, a web application was built to provide a visualization of the FMoW dataset with the location of each image. The web application currently allows for filtering of country code, image category and sensor platform. The purpose of this is give the user an idea of what kind of categories the data consisted of and where the data is being gathering from.

V. Conclusion

To conclude, the FMOW dataset was very suited to answering many different questions, pertaining to different locations and landmarks. While the dataset is not perfectly distributed and thus may not be capable of producing unbiased results in every situation, it is particularly suited towards aggregating its data.

This research covered a somewhat narrow range of topics, thus it cannot be said with confidence that say that the dataset is useful in all contexts, but it certainly seems to be suitable as a representation of many different categories in the world. Having a singular dataset that covers many different types of landmarks could be very interesting for many different types of research, as it already has all of its data in one format.

Future research may be able to use this dataset and other processed images of the world to build initial hypothesises of research. We found the data in provided in this FMOW was accurate to real life values done by other research while not being particularly tailored to a single item, rather being broad enough that it could answer multiple different questions in a wide variety of categories. With the plan to use Machine Learning on the dataset to continue building the number of items marked, this dataset proves to be a strong starting point for keeping tally of items worldwide without needing specialized information gathering.

REFERENCES

[1] G. Christie et al. “Functional Map of the World”. CVPR. 2018.

[2] G. Sreejith. “Pavement temperature data maps for worldwide locations”. International Journal of Pavement Research and Technology 15.4 (2022), pp. 948–956.

[3] Slovic, Paul, et al. ”Nuclear power and the public.” Cross-cultural risk perception. Springer, Boston, MA, 2000. 55-102.

[4] Javidkia, F., M. Hashemi-Tilehnoee, and V. Zabihi. ”A comparison between fossil and nuclear power plants pollutions and their environmental effects.” Journal of Energy and Power Engineering 5.9 (2011).

[5] Heinävaara, S., Toikkanen, S., Pasanen, K. et al. Cancer incidence in the vicinity of Finnish nuclear power plants: an emphasis on childhood leukemia. Cancer Causes Control 21, 587–595 (2010). https://doi.org/10.1007/s10552-009-9488-7

[6] Elder JP, Crespo NC, Corder K, Ayala GX, Slymen DJ, Lopez NV, Moody JS, McKenzie TL. Childhood obesity prevention and control in city recreation centres and family homes: the MOVE/me Muevo Project. Pediatr Obes. 2014 Jun;9(3):218-31.

[7] World Bank Group, “Food security: Rising food insecurity in 2022”, World Bank, 17-Nov-2022. https://www.worldbank.org/en/topic/agriculture/brief/food-security-update

[8] Ayoo, Collins. ”Towards Energy Security for the Twenty-First Century”. Energy Policy, edited by Tolga Taner, IntechOpen, 2020. 10.5772/intechopen.90872.

[9] José M. Pastor, Carlos Peraita, Lorenzo Serrano & Ángel Soler (2018) Higher education institutions, economic growth and GDP per capita in European Union countries, European Planning Studies, 26:8, 1616-1637, DOI: 10.1080/09654313.2018.1480707

[10] US Energy Information Administration. Nuclear power and the environment - U.S. Energy Information Administration (EIA). https://www.eia.gov/energyexplained/nuclear/nuclear-power-and-the-environment.php