Group 2 - Data Visualization Report

University of Southern Denmark

Faculty of Engineering

Maersk Mc-Kinney Moller Institute

5th semester - Data Visualization

Course Code - SM-DV

Project period: 01. September 2022 - 18. December

Group 2

Nikolas James Duus Kristiansen nkris19@student.sdu.dk

Mette Marie Storgaard Hansen meha419@student.sdu.dk

Teacher

Parisa Niloofar parni@mmmi.sdu.dk

Abstract

The purpose of this paper is to explore the relation between different types of produced energy within the United States of America, and how it has evolved between the years 1973 and 2022. Several questions have been designed for the data, such as "Has climate laws and global warming had an impact in the US produces their energy?", "Which historical events impacted the data and can it be seen?" and lastly "What has been the distribution of produced energy throughout 1973-2022 and where is it heading?". These questions were answered with selected variables, which were then visualized.

In the visualizations it can be seen there are major differences between the three main categories of energy, these being renewable energy, nuclear energy and fossil fuels. With fossil fuels being the highest produced energy every year, ranging from 1973 to 2022, the other categories are no where near the amount like fossil fuels. However, the visualizations did show that Renewable and Nuclear are increasing, albeit slowly. This evolution can be attributed to the fact that no climate laws were taken into effect to limit the production of fossil fuels and therefore limiting pollution, but instead only incentives about renewable energy were created. Resulting in a growing production of fossil fuels alongside renewable and nuclear energy.

The visualizations also showed fluctuations in data, for instance, in coal production significant dips in data can be seen. These can be explained by nationwide worker strikes and/or less demand of product due to cheaper energy alternatives such as natural gas. Every energy type has increases and decreases in their data, all which could be backed up with sources regarding historical events. Overall the visualizations gave great insight to how the energy production has evolved throughout time, and which energy is the highest ranking overall. It also gave the possibility of seeing where the data may be heading, such as future increase in coal, wind and natural gas production.

II Table of contents

III List of Figures

Figures

1. Motivation and Background

The motivation for this project is to demonstrate an understanding of how visual representations can help in analyzing and understanding complex data. The tasks for this project consist of finding a data set that the group would like to explore in depth, and from that data set generate questions that will need to be answered through visualizations. *The group worked equally in creating the visualizations and answering the questions.*

The data set that the group decided to explore was "Energy Production within the United States of America", between the years 1973 to 2022. This data set is explained further in Section 3. The data set was chosen with the aim to shed some light into the way the United States of America (US) produces their energy. Throughout the last couple of decades, the world has become increasingly aware of global warming and the consequences it has in the environment.

Therefore, there was an interest to try and see if through the energy production, whether or not the US had made any changes or advancements, in favor of creating a more sustainable way of producing energy. Furthermore, there was also an interest in seeing how the data evolved throughout time, and how it was impacted by historical events.

2. Project Objectives

Based on the motivation and background for this project, seen in Section 1, some questions were elicited. The questions are as follows.

1. *Did climate laws have an impact on energy production in the US?*
2. *Since the world has become more aware of global warming and the problems surrounding it, can the effects be seen in the way the US produces its energy?*
3. *What historical events impacted the data, and where can it been seen?*
4. *What has been the distribution of produced energy sources throughout the years?*
5. *Where is the production of energy heading?*

The focal point for these questions was seeing and understanding the history in the data. Figuring out if there are any possible explanations as to why the data evolved as it did. Also to see where the data might be heading in the future. The answers to these questions will be discussed and presented in Section 5.

3. Data

The data was acquired from the U.S. Energy Information Administration's (EIA) website, eia.gov. The data was accompanied with a statistical report describing the data set in full detail, produced by the EIA.

The data set consists of 50 observations of 14 variables. The variables are Year, CoalProduction, NaturalGasProduction, CrudeOilProduction, NaturalGasPlantLiquidsProduction, NuclearElectricPowerProduction, HydroelectricPowerProduction, GeothermalEnergyProduction, SolarEnergyProduction, WindEnergyProduction, BiomassEnergyProduction, TotalRenewableEnergyProduction, TotalPrimaryEnergyProduction and TotalFossilFuelsProduction. Their observations ranges from the years, 1973 to 2022.

Year is a qualitative variable, specifically categorical nominal. The energy productions are all quantitative variables, specifically numerical discrete. The data set's last three variables are a summation of values from the different categories of energy productions, i.e total production categories. The categories are TotalRenewableEnergyProduction (Renewable energy), which contained hydroelectric power, geothermal, solar, wind and biomass energy production. TotalPrimaryEnergyProduction, which is a total of *all* energy productions. TotalFossilFuelsProduction (Fossil fuels) which contained coal, natural gas, crude oil and natural gas plant liquids production. Leaving NuclearElectricPowerProduction (Nuclear energy) only to be categorized in TotalPrimaryEnergyProduction.

This resulted in some issues when creating visualizations dealing with the total production categories. None of the total production categories contained nuclear energy, except for TotalPrimaryEnergyProduction, but that category also contained renewable energy and fossil fuels. Resulting in NuclearElectricPowerProduction being used as a total production category, for itself, in order to get the full overview of every total production category.

When processing the data the group ran into problems, with regards to the data formatting within the csv file. The variables, as seen in the beginning of this section, were originally spaced between each word. This became a problem since RStudio replaced the space with a dot character, which in turn made RStudio incapable at parsing this character during the creation of the visualizations. This led to some data cleanup, where every variable's space in the csv file was deleted manually. Lastly, when creating the visualizations some of the observations of the variables were "Not available" or "N/A". This was easier to clean up, with the help of one line of code which replaced every N/A with 0, resulting in easier plotting of the data.

4. Visualization

This chapter will present what thoughts the group had in mind before and while designing the visualizations. Furthermore, this chapter will show all the graphs that the group choose to create, in order to answer the questions from section 2. Lastly, the chapter will contain a description of what features the group would like to add, to increase the quality of the visualizations.

4.1. Design

When designing the visualizations certain aspects were kept in mind. Aspects such as ensuring that the graphs have proper labels, titles, color maps and a suitable visual channel.

The biggest design decision was made when creating the graphs. Ensuring every graph was properly presented, in relation to the respective types of the variables. The graphs chosen for this visualization are Scatter plot, Time series and Stacked Bar charts.

Since the group worked with a simple data-set, containing one type of categorical variable and several variables of the type of numerical, it was limited in the aspect of visually encoding the differences in the data-set.

The easiest way to differentiate between the different numerical variables was through color. Color was vital when making these visualizations, and it was kept in mind by making the color map as color-blind friendly, as possible. Additionally, to create clear distinctions between the data and making use of a broad palette of colors, instead of picking one color palette such as different shades of blue.

With color differentiating the different types of energy, it was also kept in mind to properly label and title the different graphs, ensuring every data point was explained, and every graph was properly explained. The next subsection will show the visualizations.

The first type of graph that the group decided to depict were two scatter plot diagrams with a trend line. These graphs were chosen to see the overall trend of the production of Nuclear, Fossil fuels and Renewable energy, and how they have moved throughout the years. On Figure 4.1 it can be seen that fossil fuels is much higher than renewable and nuclear. Meaning that that the production of fossil fuels is much higher, than the other types.

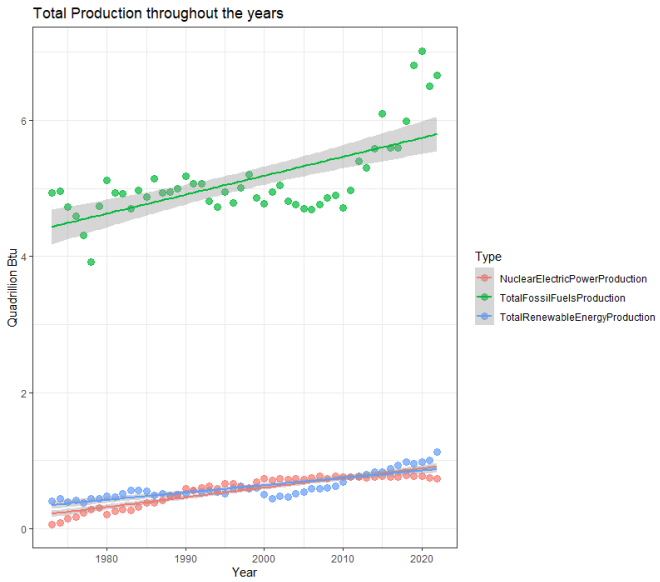

Figure 4.1 Wind, Coal and Nuclear production throughout the years

It can also be seen that renewable and nuclear are hard to differentiate between. Figure 4.2 was then produced to give a more up close view, on how Nuclear and Renewable energy have moved throughout the years.

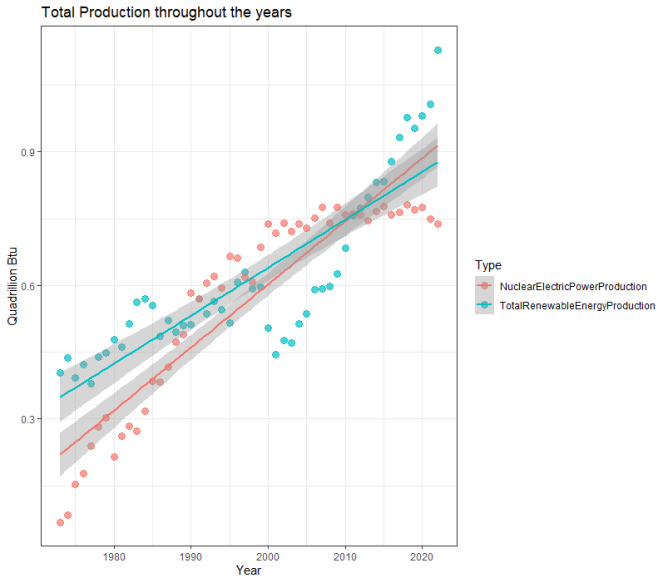

Figure 4.2 Renewable and Nuclear energy production throughout the years

It can be seen that nuclear energy is increasing, compared to renewable energy, around 2010.

The next graph that the group chose was a stacked bar chart, which can be seen on Figure

4.3. This graph gives an indication of which percentage in total, all the production types are for 1973 and 2022. It was chosen to only show the data available from the first recorded year and the last recorded year, in order to see how each of the production types has moved from start til end.

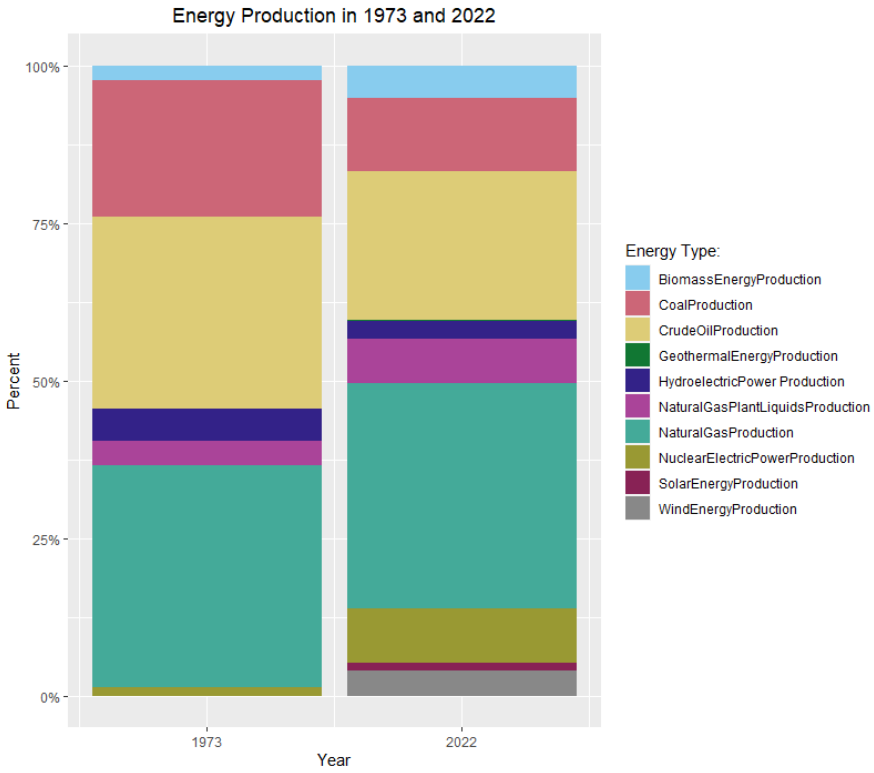

Figure 4.3 Stacked bar chart showing percentage for 1973 and 2022

Here it can be seen that Natural gas, crude oil and coal production have the highest percentages during 1973. In 2022 these three types are still the highest percentages, but coal and crude oil to a lesser extent. Natural gas still dominates percentage wise, throughout both years.

It can also be seen that biomass energy and nuclear electric production is increasing in comparing the two years. Solar and wind energy was first recorded after 1980, which is why it is not present in the first bar but present in the second one.

The next graphs are time series charts, which can be seen on Figure 4.4, 4.5 and 4.6.

The time series charts represent one of each production type from each category. Coal was chosen to be visualized as it is one of the worst polluting fossil fuels, wind was chosen for the renewable energy category, and lastly Nuclear energy production.

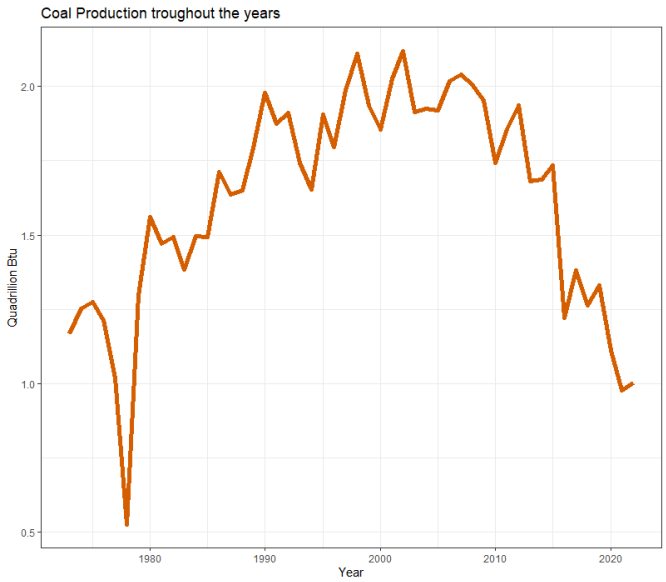

Figure 4.4 Time series chart showing coal production throughout the years

In this graph it can be seen that the production of coal fluctuates throughout time. A sudden decrease can be seen around 1980 on the x axis. This will be further explained in Section 5.

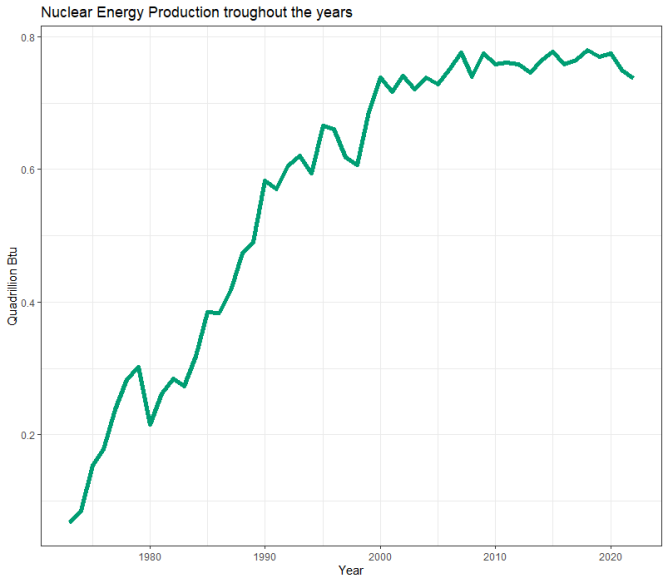

Figure 4.5 Time series chart showing nuclear production throughout the years

For nuclear energy a rapid and then steady increase in production can be seen. Although the production is nowhere near the production of coal.

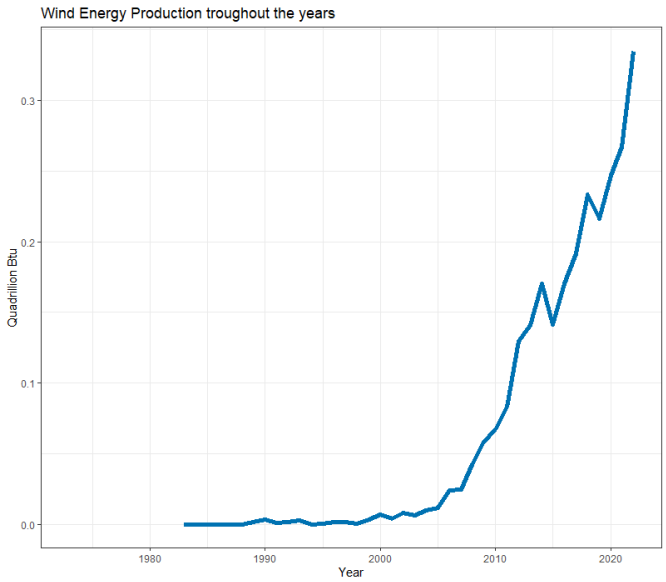

Figure 4.6 Time series chart showing wind production throughout the years

For wind energy it can be seen that it was first reported after 1980. It saw a slow increase after that until late 2000's where a rapid increase can be seen. The production of this type of energy is not as high as nuclear nor coal.

The next graph is a stacked bar chart, which can be seen on Figure 4.7. This was developed to give a different perspective on how renewable, nuclear and fossil fuels have moved throughout the years, in comparison to the previous time series charts. It was also chosen to give an overview over which energy takes up the majority of total produced for each recorded year.

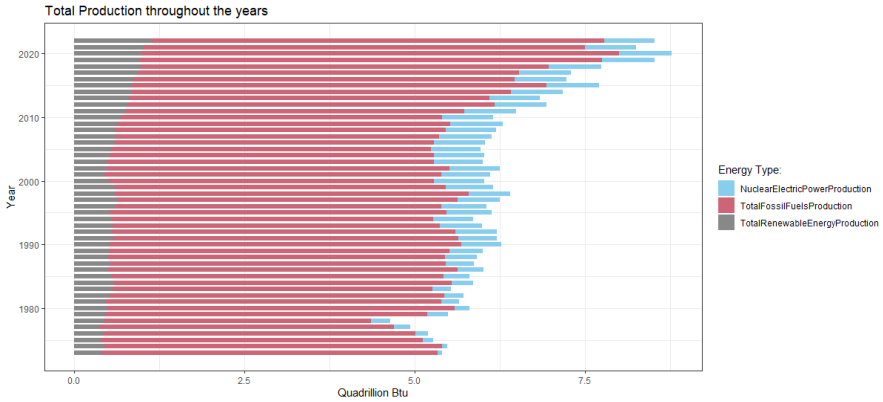

Figure 4.7 Horizontal stacked bar chart over the production of Wind, Nuclear and Wind

It can be seen that the production of renewable and nuclear energy are increasing, but so is coal. It is visible that Coal is the majority throughout 1973 to 2022.

The next Figure 4.8 shows a snippet taken from the animated graph. The snippet was taken at year 2022 and it is possible to see a trail of dots for each production type's value, for each year. The purpose of the graph is to show the evolution of the different energy production types.

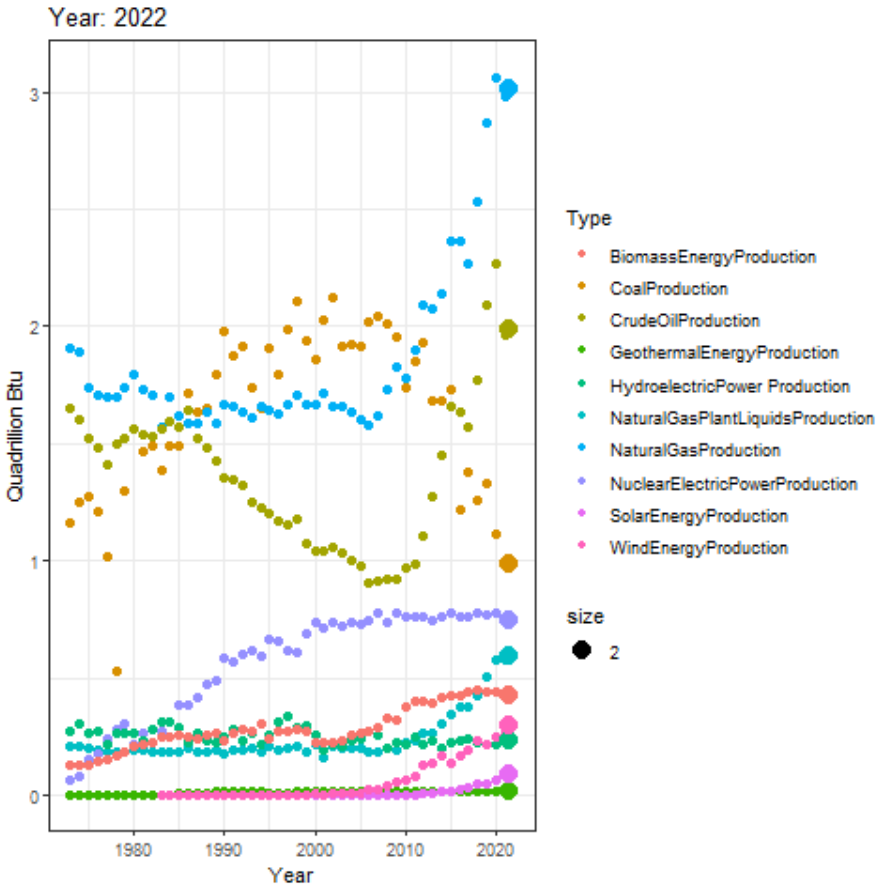

Figure 4.8 A snippet of the animated graph

The full animation can be accessed and seen insert link here, something dropbox idk.

The final Figure 4.9 shows a snippet taken from the interactive graph. In this graph it is possible to interact with the time span, selecting an interval of years for further data inspection. It is also possible to select which energy production type(s) one wants to see evolve throughout the selected time span.

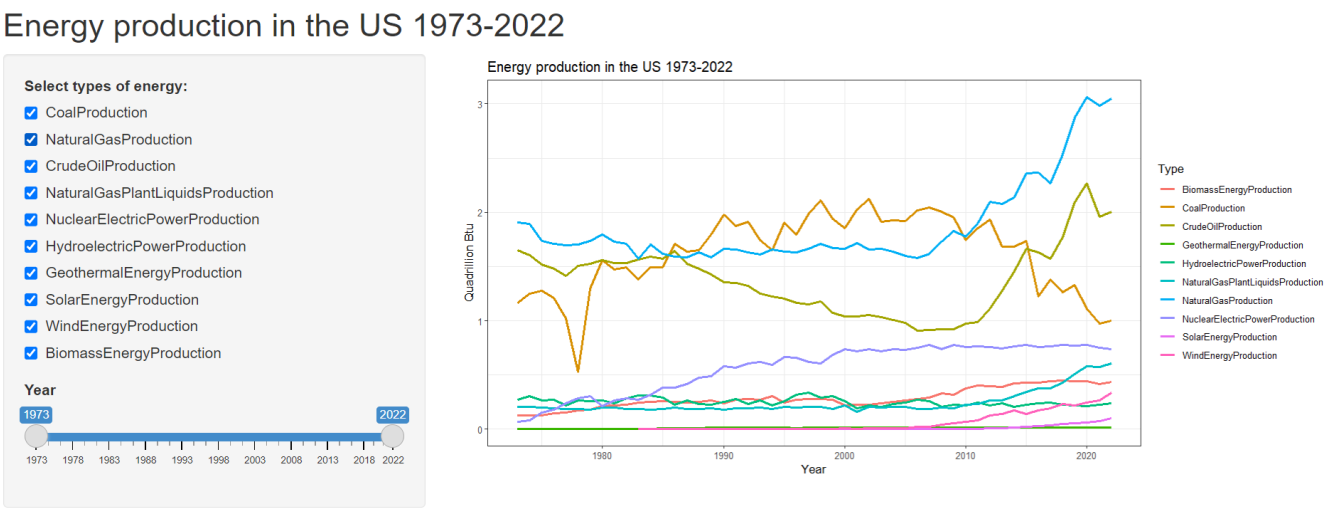

Figure 4.9 A snippet of the interactive graph

The full interactive graph can be accessed and seen here.

4.2. Optional features

- **Optional Features**: List the features which you consider to be nice to have, but not critical.

Some features that would have been nice to have could have been for graph 4.1. Instead of creating a separate graph in order to see the distinction between renewable and nuclear, the graph could have been interactive. Making it possible to zoom in and see the data more clearly.

For the interactive graph it would have been nice to have the option to move the cursor, to highlight specific data points on the graph.

5. Results

By inspecting the produced graphs it can be seen that the overall trend for all production types are on the rise, during the last couple of decades. It can be seen that fossil fuels have always been prevalent in how the US produces its energy, and renewable and nuclear are a relatively newer forms of energy production. The main difference between the production types are therefore the scale of which they are produced, where fossil fuel production greatly outweigh renewable and nuclear. Therefore, to answer question 1, the group is unable to answer this question by looking at the graphs, and by researching what climate laws that the US implemented during the years, nothing of significance appeared.

From most of the graphs it can be seen that some of the renewable energy production types only start being produced at around the 1990s and 2000s. These energy types are solar and wind energy. This could be due to the effect of the U.S. federal government establishing incentives to use renewable energy for the production of energy[1]. Therefore, to answer question 2 it is possible to see the effects of renewable energy production with some starting to be produced in the 1990s and 2000s, with the production increasing every year. However, if it is compared to fossil fuels production, which also keeps increasing, the effects of this incentive are minimal.

To best answer question 3, the group decided that the three time series graphs would represent best. The idea was to choose one energy type from each category and watch their respective progression throughout the time frame.

Going chronologically in order of the graphs, the first graph is the time series chart over coal production, Figure 4.4. As mentioned in Section 4, a large decrease in coal production can be seen around late 1970s. This is due to the fact that there was a national coal miner's strike in US during this time period[2], which halted most of the US' coal production, thus the significant change in data. When the strike was over, 1978, production is once again increasing, which is also visible on the figure. Up until 2015 where production dropped again, this time due to cheap natural gas prices[3]. Right before 2020 a significant decrease can be seen once more, according to EIA[2] this is due to less demand, but EIA also states that 2021 increases due to rising natural gas prices.

The second time series chart is Nuclear Energy, Figure 4.5. Two significant dips in the graph can be seen, the first is around 1980. This could be attributed to the fact that the production of the nuclear plants themselves were halted, due to the high cost[4]. The second dip is around 1996-1997, and it may be the result of a state ban on nuclear power plants for the second-largest coal producer in the US, West Virginia[5]. Overall, from

the looks of the data it seems there is a slow and steady increase in nuclear production. The influence of high cost power plants is the main contributor to the slow and steady data, but power plant meltdowns such as Chernobyl, 1986, put a moratorium on nuclear power[6].

The third time series chart is Wind energy, Figure 4.6. As mentioned in the answer to question 2, incentives were established to produce more renewable energy. Wind energy being a direct product from that, data started being reported after 1980. A rapid increase can be seen in 2008 to late 2000s, this is due to new wind projects being completed around this time frame, and accounting for 42% of the new power producing capacity[7].

The graph chosen to help answer how the distribution looks, is Figure 4.7, which gives a clear insight into the distribution of produced energies, by energy category. In the figure it can be seen nuclear production is the lowest produced energy throughout 1973 to 2022, renewable energies started out small but surpassed nuclear throughout most years within the time frame and is slowly on the rise to more and more increase in production. The highest producing energy type is fossil fuels, throughout the entire time frame. Only having minor dips in production, fossil fuel production is increasing more and more compared to both renewable and nuclear. So where is the energy production heading? By the looks of the visualizations nuclear is very slowly increasing its production, renewable is increasing more rapidly compared to nuclear, but it will be a while before it surpasses fossil fuels. Fossil fuels are still increasing the fastest and it does not look like it is planning to stop anytime soon. With no impending climate laws regulating the production of fossil fuels, fossil fuels will stand to be the main manufacturer of energy for the US.

6. Conclusion

In conclusion, the group has found five questions that there was interest in answering and in order to answer these questions, nine different graphs have been created. The graphs were produced while having color blind people in mind and additionally being aware of what graphs are most suitable to answer the questions presented in Section 2. The questions that the group elicited were also answered in Section 5, where the main points that can be drawn out is that even though the US has started producing more and more renewable energy they have no decreased the amount of energy they get from fossil fuels, but rather increased that as well.

7. Discussion

The group had a few problems during this course, however, the main problem would be finding the most suitable visual channel. This means that in order for a proper analysis to take place the most fitting types of graphs had to be chosen, in order to present the data with the least amount of bias. Towards the end of the course it was possible to present what the group had accomplished so far and to get the opportunity to get feedback on what graphs and thought the group had. The group found this experience particularly help full since one of the biggest problems, being finding a suitable visual channel, became less of a problem as a result of the feedback. Furthermore, the presentation made sure that the group was on the right track, and also gave insight into what had to be chanced in order to improve the visualizations.

In addition, the groups thought on the course as a whole has been a good learning experience. The classes were well structured with a lot of background knowledge was shared on the basic understanding behind visualization. There was plenty of opportunities to make sure the group was on the right track with the tasks that were required for the course. The group can therefore say that no improvement can be found within the course as everything that was expected from the course was delivered.

Bibliography

[1] EIA, *Renewable energy*, Accessed on 15/12-22. [Online]. Available: https://www.eia.gov/energyexplained/renewable-sources/incentives.php.

[2] R. B. EIA, *2019 u.s. coal production falls to its lowest level since 1978*, Accessed on 15/12-22. [Online]. Available: https://www.eia.gov/todayinenergy/detail.php?id=44536.

[3] B. Magill, *U.s. coal production dropped to 30-year low in 2015*, Accessed on 15/12-22. [Online]. Available: https://www.climatecentral.org/news/us-coal-production-hits-30-year-low-19897.

[4] K. NIEMEYER, *Chain reaction: The (slow) revival of us nuclear power*, Accessed on 15/12-22. [Online]. Available: https://arstechnica.com/science/2012/03/chain-reaction-the-slow-revival-of-us-nuclear-power/.

[5] NPR, *Coal-dependent west virginia eliminates ban on nuclear power*, Accessed on 15/12-22. [Online]. Available: https://www.npr.org/2022/02/08/1079339405/west-virginia-ban-nuclear-power-coal.

[6] D. V. Boom, *Nuclear power is clean and safe. why aren't we using more of it?* Accessed on 15/12-22. [Online]. Available: https://www.cnet.com/science/how-nuclear-power-plants-could-help-solve-climate-crisis/.

[7] J. Shepard, *Wind energy grows by record 8,300 mw in 2008*, Accessed on 15/12-22. [Online]. Available: https://eepower.com/news/wind-energy-grows-by-record-8300-mw-in-2008/#.