Week-13-document

Ho Zhi Yi

2023-11-15

Week 9

1. What is the topic that you have finalized? (Answer in 1 or 2 sentences)

   - **Topic:** Greenhouse Gas Emissions Trends in the World
   - Analyzing Greenhouse Gas Emissions Trends based on:
   - industry

```
distinct(data,Industry)
```

```
##                                                                  Industry
## 1                                  Agriculture, Forestry and Fishing
## 2                                                        Construction
## 3                 Electricity, Gas, Steam and Air Conditioning Supply
## 4                                                       Manufacturing
## 5                                                              Mining
## 6                                          Other Services Industries
## 7                                                   Total Households
## 8                                     Total Industry and Households
## 9                                        Transportation and Storage
## 10 Water supply; sewerage, waste management and remediation activities
```

   - region

```
distinct(data,Country)
```

```
##                            Country
## 1                Advanced Economies
## 2                            Africa
## 3                          Americas
## 4                              Asia
## 5          Australia and New Zealand
## 6                      Central Asia
## 7                      Eastern Asia
## 8                    Eastern Europe
## 9  Emerging and Developing Economies
## 10                           Europe
## 11                              G20
## 12                               G7
## 13   Latin America and the Caribbean
```

```
## 14                   Northern Africa
## 15                  Northern America
## 16                   Northern Europe
## 17                           Oceania
## 18                South-eastern Asia
## 19                     Southern Asia
## 20                   Southern Europe
## 21                Sub-Saharan Africa
## 22                       Western Asia
## 23                    Western Europe
## 24                             World
```

- gas types

```
distinct(data,Gas_Type)
```

```
##            Gas_Type
## 1    Carbon dioxide
## 2 Fluorinated gases
## 3    Greenhouse gas
## 4           Methane
## 5     Nitrous oxide
```

2. What are the data sources that you have curated so far? (Answer 1 or 2 sentences)

    - Annual greenhouse gas emissions by activity and by region (2010 to 2021)

Week 10

1. What is the question that you are going to answer? (Answer: One sentence that ends with a question mark that could act like the title of your data story)

    - Who contributed the most to greenhouse gas emissions?

2. Why is this an important question? (Answer: 3 sentences, each of which has some evidence, e.g., *"According to the United Nations..."* to justify why the question you have chosen is important),

    - **Environmental Impact:** According to the United Nations, understanding the factors driving greenhouse gas emissions is crucial to address climate change and its environmental impact. Reducing emissions is essential to mitigate the consequences of global warming, such as more frequent extreme weather events and rising sea levels.
    - **Economic Implications:** According to Earth.org, reducing carbon emissions would decrease the number of deaths related to air pollution and help to ease pressure on healthcare systems.
    - **Policy and Mitigation:** Accoriding to the International Monetary Fund (IMF), to accelerate cuts to emissions, policymakers need detailed statistics to assist them in devising effective mitigation measures that can deliver the fastest and least disruptive pathway toward net zero emissions.

3. Which rows and columns of the dataset will be used to answer this question? (Answer: Actual names of the variables in the dataset that you plan to use).

    - **Columns:**

  - **Year:** To track changes over time.

F2010	F2011	F2012	F2013	F2014	F2015	F2016	F2017	F2018	F2019	F2020	F2021

  - **Industry:** To examine emissions trends by sector.
  - **Country:** To analyse regional variations in emissions.
  - **Gas_Type:** To understand the contributions of different greenhouse gases.
- **Rows:**
  - for Industry:
    * Manufacturing
    * Electricity, Gas, Steam and Air Conditioning Supply
    * Transportation and Storage
    * Agriculture, Forestry and Fishing
    * Construction
  - for Country:
    * Africa
    * Americas
    * Asia
    * Europe
    * Oceania
  - for Gas_Type:
    * Carbon dioxide
    * Fluorinated gases
    * Methane
    * Nitrous oxide

4. *Include the challenges and errors that you faced and how you overcame them.*

- **Challenge 1**
  - the regions available in the **Country column** contains some regions that overlap, hence I had to select suitable regions to ensure that there are no overlapping countries in the regions used for analysis
  - the industries available in the **Industry column** contains some industries that overlap, hence I had to select suitable industries to ensure that there are no overlapping industries used for analysis
  - ***solution***: select relevant variables in the Country and Industry columns through filtering the data followed by creating a new data frame by selecting the relevant columns as shown below

```
new_data <- data %>%
  filter(Country %in% c("Africa", "Americas", "Asia", "Europe", "Oceania")) %>%
  filter(Industry %in% c("Manufacturing", "Electricity, Gas,
                         Steam and Air Conditioning Supply", "Transportation and Storage",
                         "Agriculture, Forestry and Fishing", "Construction")) %>%
  select(Country, Industry, Gas_Type, F2010, F2011, F2012, F2013, F2014, F2015,
         F2016, F2017, F2018, F2019, F2020, F2021)
```

- **Challenge 2**
  - column names were not ideal and straight forward to use
  - the individual years were not named in numerical form

  - the Country column was a bit confusing because the values under this column are regions rather than specific countries
  - ***solution***: rename the columns accordingly as shown below

```
new_names <- c("Region", "Industry", "Gas_Type", 2010, 2011, 2012, 2013, 2014, 2015,
               2016, 2017, 2018, 2019, 2020, 2021)

new_data %>% set_names(new_names)
```

Week 11

1. List the visualisations that you are going to use in your project (Answer: What are the variables that you are going to plot? How will it answer your larger question?)

   - **Line graphs** of
     - industry against time
     - each industry against time using shiny
     - region against time
     - each region against time using shiny
     - gas type against time
     - each gas type against time using shiny
   - **Bar graphs** of
     - each gas type against industry using shiny
     - each gas type against region using shiny
   - These graphs will assist in pinpointing the region, industry, and specific type of gas that exerts the most significant influence on the greenhouse gas effect, thereby providing valuable information for crafting policies targeted at reducing the primary source of emissions.

2. How do you plan to make it interactive? (Answer: features of ggplot2/shiny/markdown do you plan to use to make the story interactive)

   - Create shiny apps to allow users to select various variables to view the plots for the chosen variable
   - Use ggplot2 with plotly to make the plots interactive

3. What concepts incorporated in your project were taught in the course and which ones were self-learnt? (Answer: Create a table with topics in one column and Weeks in the other to indicate which concept taught in which week is being used. Leave the entry of the Week column empty for self-learnt concepts)

No.	Concepts	Week
1	Reading csv file Accessing elements of a vector	Week 3
2	Manipulating data • tidy data (filter( ), summarise( )) • choosing row/column • piping operators (%>%) • calculating mean and sum	Week 4

No.	Concepts	Week
3	Functions • writing functions	Week 5
4	Iterations • for loops	Week 6
5	Visualizing data using Shiny ggplot2 • creating line and bar graphs	Week 7
6	Visualizing data using Shiny • creating Shiny app for line and bar graphs	Week 8
7	Exploratory Data Analyses • pivot_longer( ) function	Week 9
8	**Self-learnt** • using kable( ) to display the sample data in a table form • sapply( ) function • creating styles.scss file to edit the theme of the website as it allows me to define variables, making it easier to edit the colours of the website • using panel-tabset to create tabs to show the plot and code in different tabs • using plotly to make the plots interactive (including plots in shiny app)	-

4. *Include the challenges and errors that you faced and how you overcame them.*

- **Challenge 1**
  - values of variables were not in the correct data type, hence returning errors when generating the plots
  - ***solution***: convert the values to specific data types using
    - as.character(), as.numeric(), as.integer()
- **Challenge 2**
  - the data is too wide, making it difficult to manipulate for calculations
  - ***solution***: use pivot_longer() to reshape the data to a longer format

Week 12

1. *Include the challenges and errors that you faced and how you overcame them (if any)*

- **Challenge 1**
  - the height of the bar graphs plotted may be misleading as the y axis is auto adjusted to fit the values for each graph
  - ***solution***: incorporate plotly into all the plots to show the actual emissions value and make it interactive at the same time
- **Challenge 2**
  - in the 4th shiny app of plotting the graphs for each gas type against industry, the industry names were too long, so they are overlapping one another. also, the legend is too long hence when incorporating plotly into the graphs, the graphs are shrunk to the point that they can't be seen

- ***solution***: use the first letter of each industry as the labels to make it shorter using substring()

Week 13 (Final Submission)

Diary Entry

1. What is the theme of your data story?

   - The theme of the data story revolves around understanding and addressing greenhouse gas emissions trends from 2010 to 2021. The analysis focuses on industries, regions, and gas types to identify key contributors to greenhouse gas emissions. The question that we are trying to answer is: Who is the highest contributor?

2. Why is it important to address this question?

   - Greenhouse gas emissions is a primary driver of climate change. Identifying the most significant source of emissions allows policymakers, scientists, and the public to make informed decisions and implement targeted strategies for mitigation. The urgency of addressing this question stems from the immediate and long-term impacts of climate change on ecosystems, economies, and human well-being.

3. Why do you think the data sources that you have curated can help you answer the question?

   - **Credibility of the International Monetary Fund (IMF):** The emissions data was obtained from the IMF website. The IMF is a reputable international organization that collects and disseminates economic and financial data. It is known for its rigorous data collection processes and commitment to providing accurate and reliable information. This enhances the credibility of the greenhouse gas emissions data obtained.
   - **Comprehensive Coverage:** The dataset covers a range of industries, regions, and gas types, providing a comprehensive view of greenhouse gas emissions. This allows holistic analysis of the data.
   - **Standardised Unit of Measurement:** The use of a standardised unit of measurement (million metric tons of CO2 equivalent) ensures consistency and facilitates meaningful comparisons across different categories. This is crucial for accurate analysis and interpretation of the emissions data.

4. What are the insights from the data and how are they depicted in plots?

   - **Line Graphs**: Highlight emissions trends (2010-2021) for industries, regions, and gas types
     - Identified that the Electricity, Gas, Steam, and Air Conditioning Supply industry, Asia, and Carbon Dioxide are the highest contributors to greenhouse gas emissions based on industry, region and gas type respectively.
   - **Bar Graphs**: Visualize the total emissions of each gas type across industries and regions
     - Proved that the Electricity, Gas, Steam, and Air Conditioning Supply industry in Asia is the highest contributor to carbon dioxide gas emissions.

5. How did you implement this entire project? Were there any new concepts that you learnt to implement some aspects of it?

   - **Data Cleaning:** Cleaned the dataset to ensure that the dataset is suitable for analysis
   - **Data Visualisation:** Utilised line graphs and bar graphs to visualise the emissions data
   - **New concepts:** Integrated plotly into the ggplot2 plots and shiny app to make the plots interactive

Write-up

Background

Greenhouse gas emissions, primarily from human activities, drive climate change by trapping heat in the atmosphere, impacting human health and ecosystems (Basics of Climate Change | US EPA, 2023). The global focus on greenhouse gas emissions has increased, with 54% of the world's population considering climate change a very serious problem (Budiman, 2020). Recognising the environmental consequences, there is a pressing need to understand and address the intricacies of these emissions.

Reducing carbon emissions not only helps mitigate climate change but also has direct benefits for public health. It decreases deaths related to air pollution and eases pressure on healthcare systems (Sarao, 2021). To accelerate emission cuts, policymakers require detailed statistics to devise effective mitigation measures, aiming for the fastest and least disruptive path toward achieving net-zero emissions (International Monetary Fund, 2022). This underscores the importance of accurate data analysis in shaping strategies to tackle the immediate and long-term challenges posed by greenhouse gas emissions.

This growing acknowledgment underscores the necessity to study and analyse greenhouse gas emissions data. As societies grapple with their carbon footprint, a deeper understanding of the factors driving emissions is crucial for informed decision-making and effective strategies to mitigate the impacts of climate change.

In this writeup, we will delve into an analysis of greenhouse gas emission trends covering the period from 2010 to 2021. The focus will be to examine the contributions of different industries, regions, and gas types to greenhouse gas emissions. The dataset used for this analysis was sourced from the International Monetary Fund, ensuring a credible and reliable foundation for the investigation. The primary objective is to uncover the highest contributor to greenhouse gas emissions by meticulously discerning patterns and trends within the provided data. This analysis seeks to provide valuable insights that can guide strategies and decisions aimed at tackling the challenges presented by the most significant greenhouse gas emissions.

Approach

To comprehensively analyse greenhouse gas emission trends from 2010 to 2021, the approach involves a systematic examination of data focusing on industries, regions, and gas types. The following steps outlines the methodology:

1. Data Collection:

The dataset, encompassing annual greenhouse gas emissions data from 2010 to 2021, was procured from the International Monetary Fund, which is a renowned source known for its precision and reliability. The information is categorised based on industry, region, and gas type.

Link to IMF dataset: https://climatedata.imf.org/datasets/c8579761f19740dfbe4418b205654ddf

2. Cleaning and Formatting:

To ensure precision in the analysis, relevant variables related to industry, region, gas type, and yearly emissions (2010 to 2021) were filtered from the dataset, ensuring that only data essential to the investigation was retained for further analysis. Also, a critical aspect of data clarity lies in proper naming and formatting. Therefore, the column names were renamed to be clear, reflecting the nature of the data they encapsulate. This enhances the data's readability and at the same time make it easier to manipulate. Lastly, due to the fact that there were overlapping industries and regions, selection of suitable industries and regions was needed.

3. Exploratory Data Analysis:

Various statistical and graphical techniques were employed to extract insights into the overarching patterns and distributions of greenhouse gas emissions data from the dataset. Key visualisations including line graphs and bar graphs were generated.

4. Interactive Visualisations:

Through the utilisation of plotly, ggplot2 and shiny app, the visualisations offered interactivity for exploring greenhouse gas emissions data. Users can zoom, hover, and focus on specific data points, ensuring a dynamic and engaging experience. This approach encourages active user involvement in the exploration process, enhancing the understanding of greenhouse gas emissions trends.

5. Identification of Highest Contributor of Greenhouse Gas Emissions:

Through an in-depth analysis of data trends from the visualisations, the goal was to pinpoint the industry, region, and gas type that emerged as the highest contributor of greenhouse gas emissions spanning from 2010 to 2021.

Subsequently, bar graphs were used to prove if the identified industry and region were indeed the highest contributors to the specific gas type associated with the highest emissions.

6. Insights and Recommendations:

Extracted insights from the line and bar graphs provided a comprehensive understanding of the dynamics of greenhouse gas emissions. Through recognising potential shortcomings in the analysis, there is room for improvements to enhance accuracy of the insights. These insights forms the groundwork for potential recommendations and strategies, addressing the challenges posed by the identified highest contributor.

Final Outcome

The visualisations led to a conclusive finding that the primary contributor to greenhouse gas emissions is the Electricity, Gas, Steam, and Air Conditioning Supply industry in Asia, particularly associated with carbon dioxide emissions. The line graphs distinctly illustrated that this identified industry, Asia region, and carbon dioxide gas type exhibited the highest contributions to greenhouse gas emissions spanning from 2010 to 2021. Furthermore, the supporting evidence from the bar graphs substantiates the initial identification, providing additional confirmation that the Electricity, Gas, Steam, and Air Conditioning Supply industry in Asia is indeed the highest contributor to carbon dioxide gas emissions.

Therefore, with this conclusion, suggested policies and strategies can be directed towards mitigating the impact of carbon dioxide emissions from the Electricity, Gas, Steam, and Air Conditioning Supply industry in Asia, contributing to a more effective approach in addressing the challenges posed by greenhouse gas emissions.

References

Basics of climate Change | *US EPA*. (2023, November 1). US EPA. https://www.epa.gov/climatechange-science/basics-climate-change#

Budiman, A. (2020, July 27). *Global Concern about Climate Change, Broad Support for Limiting Emissions | Pew Research Center*. Pew Research Center's Global Attitudes Project. https://www.pewresearch.org/global/2015/11/05/global-concern-about-climate-change-broad-support-for-limiting-emissions/#

Climate Change Indicators Dashboard. (2021, March 19). https://climatedata.imf.org/datasets/c8579761f19740dfbe4418b205654ddf/explore

International Monetary Fund. (2022, November 28). *Bridging data gaps can help tackle the climate crisis.* IMF. https://www.imf.org/en/Blogs/Articles/2022/11/28/bridging-data-gaps-can-help-tackle-the-climate-crisis#

Sarao, J. (2021, September 14). *Reducing carbon emissions will benefit the global economy. Here's how.* Earth.Org. https://earth.org/carbon-emissions-economy/#