

Week-12-document

Ho Zhi Yi

2023-11-07

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:** According 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 analyze 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

- List the visualizations 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
 - gas type against time
 - region against time
 - **Bar graphs** of
 - gas type against industry
 - gas type against region
 - 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.
- 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
- What concepts incorporated in your project were taught in the course and which ones were self-learned? (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-learned concepts)

No.	Concepts	Week
1	Reading csv file	Week
	Accessing elements of a vector	3
2	Manipulating data	Week
	<ul style="list-style-type: none"> • tidy data (filter(), summarise()) • choosing row/column • piping operators (%>%) • calculating mean and sum 	4
3	Functions	Week
	<ul style="list-style-type: none"> • writing functions 	5
4	Iterations	Week
	<ul style="list-style-type: none"> • for loops 	6

No.	Concepts	Week
5	Visualizing data using Shiny ggplot2	Week 7
6	<ul style="list-style-type: none"> • creating line and bar graphs Visualizing data using Shiny	Week 8
7	<ul style="list-style-type: none"> • creating Shiny app for line and bar graphs Exploratory Data Analyses	Week 9
8	Self-learnt <ul style="list-style-type: none"> • pivot_longer() function • 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

- (1) What is the theme of your data story?
- (2) Why is it important to address this question?
- (3) Why do you think the data sources that you have curated can help you answer the question?
- (4) What are the insights from the data and how are they depicted in plots?
- (5) How did you implement this entire project? Were there any new concepts that you learnt to implement some aspects of it?

Write-up

Write-up including the list of references (APA style, include websites you use to source data) should not exceed more than 1200 words. The minimum requirement is 500 words.