Document-Entries

Tan Ee Xuan

2023-10-18

Week 9

Answer the following questions in an R Markdown file,

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

The topic I have chosen to finalise on would be on deforestation in Brazil and how agriculture and livestock produce contributed to deforestation.

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

I have decided to use the datasets from Our World in Data and from Climate Watch Data. To be more specific, the datasets I intend to use would be “sugarcane.csv”, “soybean.csv”, “maize.csv”, “cattle.csv”, “pigmeat.csv”, and “poultry.csv” from Our World in Data and “historical_emissions.csv” from Climate Watch Data.

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)

How has Agriculture and Livestock Production Contributed to Deforestation in Brazil?

(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)

According to the United Nations (n.d.), the process of climate change is accelerating as a result of greenhouse gas emissions that humans produce and are responsible for. The effects of climate change will have a devastating impact on countries and communities all around the world, especially for the poor who may depend on the environment for their livelihoods (MercyCorps, 2021). Brazil is one of the largest emitters of greenhouse gasses due to “deforestation, agriculture and other land-use” (Gratten, 2022).

(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).

Our World in Data

I have decided to use the datasets from Our World in Data and from Climate Watch Data. To be more specific, the datasets I intend to use would be “sugarcane.csv”, “soybean.csv”, “maize.csv”, “cattle.csv”, “pigmeat.csv”, and “poultry.csv” from Our World in Data.

For the “sugarcane.csv”, “soybean.csv”, “maize.csv”, “cattle.csv”, “pigmeat.csv”, and “poultry.csv” datasets, I have filtered the column “Country” for all the rows containing “Brazil” and I filtered the column “Year” such that it only shows rows with “2011”, “2012”, …, “2019”, and “2020” from its respective raw datasets that I obtained from Our World in Data.

For each of the “sugarcane.csv”, “soybean.csv”, and “maize.csv” datasets, I will use the columns “year”, “population”, “production_kg”, “production_kgpercapita”, “landuse_ha”, “animalfeed_kg”, “animalfeed_kgpercapita”.

For each of the “cattle.csv”, “pigmeat.csv”, and “poultry.csv” datasets, I will use the columns “year”, “population”, “production_t”, “production_kgpercapita”, “producing_or_slaughtered_animals”, “producing_or_slaughtered_animals_percapita”.

Climate Watch Data

For the “historical_emissions.csv” dataset, I filtered the column “Gas” for all the rows containing “All GHG”. I also filtered the column “Country” for all the rows containing “BRA” which stands for my country of focus, Brazil. I will only look at the years from “2011”, “2012”, …, “2019”, and “2020”. I will use the columns “year”, “agriculture”, “land_use_change_and_forestry”, and “total.”

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

I struggled to find specific datasets that would provide me with the information that I need to answer my question. Some of the datasets available only helps answer part of the question. I also realised that I had to come up with specific indicators or variables that I want to focus on such that I would not be overwhelmed with the large amount of data available.

Week 11

(1) 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?)

There will be different graphs shown for each relationship I want to observe. I will be able to clearly visualise and make comments on how certain trends have influenced each other or suggest what might happen in the future.

Some of the relationships I hope to study would be between the soybean, maize, sugar cane, cattle, poultry, and pigmeat production with respect to time. The land use over time and percentage of crop used for animal feed over time for the soybean, maize, and sugar cane datasets.

(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)

I will be using shiny and plotly to make the story interactive. There will be a drop-down feature for the user to interact with and learn more about the trends of the variable that they are interested in.

(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)

Concept	Week
Include graphics	Week-1
ggplot	Week-2
Variables	Week-3
Manipulate data (filter, select, group_by, etc)	Week-4
Make a simple function	Week-5

Concept	Week
strsplit	
ifelse	
nchar	
Loops	Week-6
Sys.sleep	
substr	
ggplot graphs (geom_bar, geom_point, geom_violin, facet_wrap, facet_grid)	Week-7
Shiny App	Week-8
tribble	Week-9
pivot_longer, pivot_wider	Week-9
data.frame	
plotly, add_trace	
fluidrow	
wellPanel, tabsetPanel	

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

I had some difficulty in coming up with ideas on what I could do to make the story interactive. To overcome this I went to search up some information and draw some inspiration from others.

Week 12

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

Finding a way to organise all the information without it being too messy, the different features that can be used in the UI aspect was very useful in helping me arrange items. For the part on the server, I had to try to find ways to simplify my code such that it would not be too complicated. Understanding which parts of the code did what was quite useful. I had to Google a lot to try to understand what was wrong with my code. I managed to pick up some hacks from all the coding and researching (that helped speed up the process a little). For instance, I realised I could press an arrow at the left bar of RStudio that essentially shortens the function into one line. It helped me look at each code better and figure out the larger picture.

Week 13

(1) What is the theme of your data story?

The theme would be on sustainability and the environment. The story will study the trends in agriculture and livestock production in Brazil and how such produce can contribute to deforestation in Brazil.

The data story will first mention Brazil's role as a superpower in agriculture and livestock production. Brazil is one of the largest food producers in the world, with the fourth-largest Gross Domestic Product (GDP) value for agricultural produce and one of the largest populations of livestock in the world (Schneider et al., 2023). The produce can be exported to feed people from other countries, showing the significance of Brazil's contribution to the food supply worldwide.

The story would then raise the issue of how Brazil's agriculture and livestock production is also the leading cause of deforestation in Brazil. Farmers may seek to clear rainforests to start plantations and create pastures (Moran, 2018). Excessive clearing of rainforests drives climate change, indirectly affecting people worldwide. The title of agriculture superpower that Brazil holds is hence not without consequences.

Through visuals and descriptions, the data story will investigate the trends, the different types, and the amount of agriculture and livestock produced in Brazil. This allows better context as to what contributes to deforestation in Brazil, allowing for more targeted efforts to reduce deforestation in Brazil.

(2) Why is it important to address this question?

Brazil is one of the largest emitters of greenhouse gasses due to "deforestation, agriculture and other land-use" (Gratten, 2022). These greenhouse gases contribute to climate change that affects countries worldwide. Understanding how agriculture and livestock production contributed to deforestation in Brazil is crucial as it can help provide solutions that reduce deforestation and protect biodiversity.

Addressing the question can show how specific crops or livestock may emit more greenhouse gases, allowing for more targeted solutions. For instance, rearing livestock usually causes greenhouse gas emissions as it requires cleared land for pastures and crops grown for feed. Meat products, therefore, indirectly fuel the production and hence emissions of certain crops. Changing consumption habits, like cutting out meat, is an example of an approach to indirectly reduce deforestation and hence emissions. Addressing the question can therefore drive future solutions to mitigate deforestation.

Additionally, addressing the question can help preserve biodiversity. Understanding the agriculture and livestock produce that contributes to deforestation allows better measures to prevent such clearing of rainforests. The Amazon rainforest is the world's largest tropical carbon sink (Pannett, 2021). It holds over 3 million species in an interconnected and balanced ecosystem (Thomson, 2020). Land cleared for pastures and plantations threatens the balance, causing a decline in biodiversity. This potentially leads to the loss of the healthcare and food security benefits that biodiversity provides. Addressing the question helps reduce deforestation, therefore preserving biodiversity.

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

The historical emissions dataset provides information on the causes of greenhouse gas emissions in Brazil (Climate World Data, n.d.). Deforestation, from agriculture or livestock production, is often the cause of the emissions. This dataset allows a better understanding of the contributors of greenhouse gas emissions and deforestation in Brazil.

Other datasets include commonly produced crops in Brazil, including maize, soybean, and sugarcane (Ritchie et al., 2023). The variables in each datasheet include the amount produced, the land used, and the amount used for animal feed. Each dataset shows variables from 2011-2020. This hopes to study the trend of that particular crop's production and its trajectory. From the dataset, I can learn more about the crops that drive agriculture production in Brazil and the purpose of the produce, such as for animal feed. This can better contextualise the agriculture production in Brazil, potentially allowing for the adoption of practices that can reduce deforestation from crop production. Hence, the datasets on the different crops can help answer how agriculture production contributed to deforestation.

Lastly, studying datasets on meat production like cattle, pigmeat and poultry, coupled with the crop production datasets, can provide insights into deforestation in Brazil (Ritchie et al., 2023). For instance, the trends of livestock production, in addition to the portion of crops allocated for animal feed, can explain how

the resources are being allocated for livestock production. This can therefore contextualize deforestation as a result of livestock production.

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

An insight from the data would be the extent of greenhouse gas emissions contributed by agriculture and land use in Brazil. The bar graph generated from the historical emissions dataset clearly shows that more than half of the total emissions in Brazil came from agriculture land use and forestry from 2011 to 2020. Since deforestation is often the cause of emissions, agriculture or livestock production activities are therefore the main drivers of deforestation.

From the data, soybean appears to be one of the largest contributors to deforestation. Despite the production of sugarcane being significantly higher, the amount of land used for soybean production is the most significant and is still increasing as of 2020. This shows that much of the land area cleared is being used for soybean production. Soybean production is hence the most significant crop contributor to deforestation.

The data shows that the maize produced largely served as animal feed. About half of the maize produced is used as animal feed. This is significantly more than the portion of soybean or sugarcane produced used for animal feed. Maize is one of the most popular animal feeds for poultry (Jacob, n.d.). From the data, poultry production, in mass and quantity, is the highest among the meat types from the period of 2011 to 2020. Cattle production, in terms of mass, comes in second, followed by pigmeat. The data hence supports the idea of maize production for animal feed.

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

Before starting the project, I researched different potential topics and datasets. I also experimented with visualisation features in the Shiny App. Coming up with a placeholder helped me organise and visualise my thoughts before starting on the application. This planning would later guide me on the rest of the project.

I initially started with a write-up on deforestation and agriculture and livestock production in Brazil. This information would later be rephrased to be included in the website to provide context on the data that would later be analysed. I learnt to include videos, images, and website links into the page itself, using codes like “<img src=“”>” and other codes for style. In addition to experimentation, the website also involved a lot of research. This research would serve as a foundation for the data analysis in the next segment of the project.

When constructing the application, I referenced the placeholder for the data used and the way it is presented. I decided to mainly use graphs as it allows for the visualisation of trends. I settled on using “plot_ly()” instead of “ggplot()” as I preferred the aesthetics and workings of the former. To compare the different datasets, I used codes like “selectizeInput()” for the UI aspect. Codes like “tabsetPanel()” also allowed for exploring different topics by alternating between the tabs created. With the visualisations, I was able to analyse the information from the datasets better and make more conclusions about the topic.

References can be found in Final Submission Folder