

# 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 it affects the food supply.

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 "cattle\_owid.csv", "soybean\_owid.csv", and "forest\_area.csv" from Our World in Data and "CW\_HistoricalEmissions\_ClimateWatch.csv" from Climate Watch Data.

The data sources I have curated are as follows:

```
## Rows: 13819 Columns: 40
## — Column specification —————
## Delimiter: ","
## chr (1): Country
## dbl (8): Year, Population, Production (t), production__tonnes__per_capita, ...
## lgl (31): Product, Yield (t/ha), Land Use (ha), area_harvested__ha__per_capi...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
## # A tibble: 13,819 × 40
##   Product Country      Year Population `Production (t)` production__tonnes__p...1
##   <lgl>    <chr>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 NA      Afghanistan 1961      8790140      43000      0.00489
## 2 NA      Afghanistan 1962      8969055      45800      0.00511
## 3 NA      Afghanistan 1963      9157463      47250      0.00516
## 4 NA      Afghanistan 1964      9355510      48000      0.00513
## 5 NA      Afghanistan 1965      9565154      48700      0.00509
## 6 NA      Afghanistan 1966      9783153      68000      0.00695
## 7 NA      Afghanistan 1967     10010037      65000      0.00649
## 8 NA      Afghanistan 1968     10247782      71000      0.00693
## 9 NA      Afghanistan 1969     10494491      75000      0.00715
## 10 NA     Afghanistan 1970     10752973      62000      0.00577
## # i 13,809 more rows
## # i abbreviated name: 1production__tonnes__per_capita
## # i 34 more variables: `Production per capita (kg)` <dbl>,
## #   `Yield (t/ha)` <lgl>, `Yield (kg/animal)` <dbl>, `Land Use (ha)` <lgl>,
## #   area_harvested__ha__per_capita <lgl>, `Land Use per capita (m²)` <lgl>,
## #   `Producing or slaughtered animals` <dbl>,
## #   `Producing or slaughtered animals per capita` <dbl>, `Imports (t)` <lgl>, ...
```

```
## Rows: 10 Columns: 10
## — Column specification —————
## Delimiter: ","
## dbl (10): Year, Population, Production (t), Yield (t/ha), Land Use (ha), Lan...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
## # A tibble: 10 × 10
##   Year Population `Production (t)` `Yield (t/ha)` `Land Use (ha)`
##   <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
## 1  2011  198185300      74815450      3.12      23968664
## 2  2012  199977710      65848856      2.64      24975258
## 3  2013  201721780      81724480      2.93      27906676
## 4  2014  203459650      86760520      2.87      30273764
## 5  2015  205188210      97464936      3.03      32181244
## 6  2016  206859580      96394820      2.90      33183120
## 7  2017  208504960     114732104      3.38      33959880
## 8  2018  210166600     117912450      3.39      34777936
## 9  2019  211782880     114316830      3.18      35895208
## 10 2020  213196300     121797710      3.28      37188168
## # i 5 more variables: `Land Use per capita (m²)` <dbl>,
## #   `Domestic supply (t)` <dbl>, `Domestic supply per capita (kg)` <dbl>,
## #   `Animal feed (t)` <dbl>, `Animal feed per capita (kg)` <dbl>
```

```
## Rows: 7846 Columns: 4
## — Column specification —————
## Delimiter: ","
## chr (2): entity, code
## dbl (2): year, forest_area
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
## # A tibble: 7,846 × 4
##   entity      code  year forest_area
##   <chr>      <chr> <dbl>      <dbl>
## 1 Afghanistan AFG    1990      0.0285
## 2 Afghanistan AFG    1991      0.0286
## 3 Afghanistan AFG    1992      0.0286
## 4 Afghanistan AFG    1993      0.0287
## 5 Afghanistan AFG    1994      0.0287
## 6 Afghanistan AFG    1995      0.0288
## 7 Afghanistan AFG    1996      0.0288
## 8 Afghanistan AFG    1997      0.0289
## 9 Afghanistan AFG    1998      0.0290
## 10 Afghanistan AFG    1999      0.0290
## # i 7,836 more rows
```

```
## Rows: 11115 Columns: 35
## — Column specification —————
## Delimiter: ","
## chr (4): Country, Source, Sector, Gas
## dbl (31): 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, ...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
## # A tibble: 11,115 × 35
##   Country Source Sector Gas `1990` `1991` `1992` `1993` `1994` `1995`
##   <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 AFG Climate... Total... All ... 11.6 11.9 11.5 11.7 11.7 12.2
## 2 AFG Climate... Total... All ... 9.24 9.51 9.16 9.29 9.34 9.79
## 3 AFG Climate... Energy All ... 2.28 2.14 1.68 1.64 1.57 1.50
## 4 AFG Climate... Indus... All ... 0.0527 0.0554 0.0581 0.0619 0.0646 0.0673
## 5 AFG Climate... Agric... All ... 8.07 8.39 8.40 8.48 8.52 8.95
## 6 AFG Climate... Waste All ... 1.23 1.32 1.40 1.49 1.58 1.67
## 7 AFG Climate... Land-... All ... -2.39 -2.39 -2.39 -2.39 -2.39 -2.39
## 8 AFG Climate... Bunke... All ... 0.0194 0.0195 0.0195 0.0195 0.0162 0.0162
## 9 AFG Climate... Total... CO2 2.05 1.94 1.53 1.53 1.49 1.46
## 10 AFG Climate... Total... CH4 6.74 7.02 7.12 7.21 7.46 7.83
## # i 11,105 more rows
## # i 25 more variables: `1996` <dbl>, `1997` <dbl>, `1998` <dbl>, `1999` <dbl>,
## # `2000` <dbl>, `2001` <dbl>, `2002` <dbl>, `2003` <dbl>, `2004` <dbl>,
## # `2005` <dbl>, `2006` <dbl>, `2007` <dbl>, `2008` <dbl>, `2009` <dbl>,
## # `2010` <dbl>, `2011` <dbl>, `2012` <dbl>, `2013` <dbl>, `2014` <dbl>,
## # `2015` <dbl>, `2016` <dbl>, `2017` <dbl>, `2018` <dbl>, `2019` <dbl>,
## # `2020` <dbl>
```

## 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)

Should we prioritise food production or environmental conservation 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).

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 “cattle\_owid.csv”, “soybean\_owid.csv”, and “forest\_area.csv” from Our World in Data and “historical\_emissions\_cw.csv” from Climate Watch Data.

For the “forest\_area.csv” dataset, I will filter the column “entity” for the rows consisting of “Brazil”. I will also filter the column “year” such that it only shows rows with “2011”, “2012”, ..., “2019”, and “2020”. I will use the columns “year” and “forest\_area”.

For the “soybean\_owid.csv” and “cattle\_owid.csv” datasets, I will filter the column “Country” for all the rows containing “Brazil” and I will filter the column “Year” such that it only shows rows with “2011”, “2012”, ..., “2019”, and “2020”. For “soybean\_owid.csv”, I will use the columns “Year”, “Production (t)”, “Land Use (ha)”, “Domestic supply (t)” and “Animal feed (t)”. For “cattle\_owid.csv”, I will use the columns “Year”, “Production (t)” and “Producing or slaughtered animals”

For the “historical\_emissions\_cw.csv” dataset, I will filter the column “Gas” for all the rows containing “All GHG”. I will filter the column “Country” for all the rows containing “BRA” which stands for my country of focus, Brazil. I will only look at the columns “2011”, “2012”, ..., “2019”, and “2020”. For the column “Sector”, I will filter out the rows such that it shows only “Land-Use Change and Forestry” and “Agriculture”.

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 cattle or soybean production with respect to time. Another would be between the emissions and forest area against time.

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 ggplot 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

Concept	Week
Make a simple function	Week-5
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	

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.