

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:

```
library(tidyuesdayR)
library(tidyverse)
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✓ dplyr      1.1.3      ✓ readr      2.1.4
## ✓ forcats    1.0.0      ✓ stringr    1.5.0
## ✓ ggplot2    3.4.4      ✓ tibble     3.2.1
## ✓ lubridate  1.9.3      ✓ tidyr      1.3.0
## ✓ purrr      1.0.2
## — Conflicts — tidyverse_conflicts() —
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
read_csv("cattle_owid.csv")
```

```
## 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: `production__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>, ...
```

```
read_csv("soybean_owid.csv")
```

```
## Rows: 12405 Columns: 40
## — Column specification —————
## Delimiter: ","
## chr  (1): Country
## dbl (35): Year, Population, Production (t), production__tonnes__per_capita, ...
## lgl  (4): Product, Yield (kg/animal), Producing or slaughtered animals, Prod...
##
## 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: 12,405 × 40
##   Product Country      Year Population `Production (t)` production__tonnes__p...1
##   <lgl>    <chr>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 NA      Afghanistan 2010      28189672      NA      NA
## 2 NA      Afghanistan 2011      29249156      NA      NA
## 3 NA      Afghanistan 2012      30466484      NA      NA
## 4 NA      Afghanistan 2013      31541216      NA      NA
## 5 NA      Afghanistan 2014      32716214      NA      NA
## 6 NA      Afghanistan 2015      33753500      NA      NA
## 7 NA      Afghanistan 2016      34636212      NA      NA
## 8 NA      Afghanistan 2017      35643420      NA      NA
## 9 NA      Afghanistan 2018      36686788      NA      NA
## 10 NA     Afghanistan 2019      37769496      NA      NA
## # i 12,395 more rows
## # i abbreviated name: `production__tonnes__per_capita`
## # i 34 more variables: `Production per capita (kg)` <dbl>,
## #   `Yield (t/ha)` <dbl>, `Yield (kg/animal)` <lgl>, `Land Use (ha)` <dbl>,
## #   area_harvested__ha__per_capita <dbl>, `Land Use per capita (m²)` <dbl>,
## #   `Producing or slaughtered animals` <lgl>,
## #   `Producing or slaughtered animals per capita` <lgl>, `Imports (t)` <dbl>, ...
```

```
read_csv("forest_area.csv")
```

```
## 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
```

```
read_csv("historical_emissions_cw.csv")
```

```
## 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.