Exploring Global Crop Yield Trends: An Analysis of the Factors Impacting Crop Yields Over Time

https://github.com/satyaandrew/C7083-Assessment.git

22358100

Contents

1. Introduction:

Crop yields are a key component of global food production, providing the world's population with essential staple crops like wheat, rice, maize, and soybeans. However, crop yields are influenced by a complex array of factors, including land use, fertilizer application, and mechanization. In this portfolio, we'll explore trends in global crop yields and investigate how changes in land use, fertilizer use, and tractor density have impacted crop yields over time.

The datasets used in this portfolio are assigned from the GitHub "rfordatascience" that compromised the data from an article from Our World in Data. The article is - Hannah Ritchie, Max Roser and Pablo Rosado (2022) - "Crop Yields". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/crop-yields' [Online Resource] Five datasets are used to create visualisations for crop yields and countries from 1885 to 2019 but there is limited data up to 1960. The datasets are:

1.1 Key Crop Yields

A dataset with over 13000 observations that give the yield of different crops over the years 1961 to 2018. the following 14 variables are described:

variable	class	description
Entity	character	Country or Region Name
Code	character	Country Code (note is NA for regions/continents)
Year	double	Year
Wheat (tonnes per hectare)	double	Wheat yield
Rice (tonnes per hectare)	double	Rice Yield
Maize (tonnes per hectare)	double	Maize yield
Soybeans (tonnes per hectare)	double	Soybeans yield
Potatoes (tonnes per hectare)	double	Potato yield
Beans (tonnes per hectare)	double	Beans yield
Peas (tonnes per hectare)	double	Peas yield
Cassava (tonnes per hectare)	double	Cassava (yuca) yield
Barley (tonnes per hectare)	double	Barley
Cocoa beans (tonnes per hectare)	double	Cocoa
Bananas (tonnes per hectare)	double	Bananas

1.2 Arable land

A dataset with the area of land used for agriculture from years 1961 to 2014 with 11280 observations of 4 variables

variable	class	description
Entity	character	Country or Region Name
Code	character	Country Code (note is NA for regions/continents)
Year	double	Year
Arable land needed to produce a fixed quantity of crops ((1.0 = 1961))	double	Arable land normalized to 1961

1.3 Fertilizer

A dataset with the recordings of Nitrogen fertilizer use from 1961 to 2018 with 11965 observations.5 variables

variable	class	description
Entity	character	Country or Region Name
Code	character	Country Code (note is NA for regions/continents)
Year	double	Year
Cereal yield (tonnes per hectare)	double	Cereal yield in tonnes per hectare
Nitrogen fertilizer use (kilograms per hectare)	double	Nitrogen fertilizer use kg per hectare

1.4 Land use

A dataset with recordings of the change in land use for arable agriculture with a range of year from 10000 BCE to 2019 and over 49000 observations. This data contains lots of missing values due to population data having been added from Gapminder. 6 variables

variable	class	description
Entity	character	Country or Region Name
Code	character	Country Code (note is NA for regions/continents)
Year	double	Year
cereal yield index	double	Cereal yield index
change to land area used for cereal production since 1961	double	Change to land area use for cereal production relative since 1961
total population (gapminder)	double	Total population from gapminder data

1.5 Tractor

A dataset recording yield of crops and the number of tractors per 100 square kilometres of arable land with a range of years from 10000 BCE to 2019 AD and over 49000 observations. This data contains lots of missing values due to population data having been added from Gapminder. 6 variables

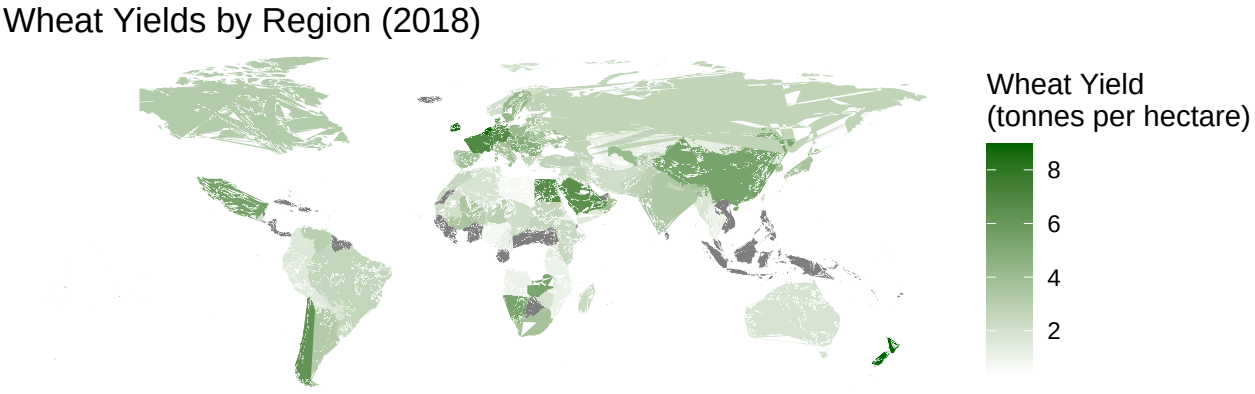
variable	class	description
Entity	character	Country or Region Name
Code	character	Country Code (note is NA for regions/continents)
Year	double	Year
Tractors per 100 sq km arable land	double	Number of tractors per 100 sq km of arable land
Cereal yield (kilograms per hectare) (kg per hectare)	double	Cereal yield in kg per hectare
Total population (Gapminder)	double	Total population from gapminder

```
## Warning: package 'ggplot2' was built under R version 4.2.2
```

```
## Warning: package 'reshape2' was built under R version 4.2.2
```

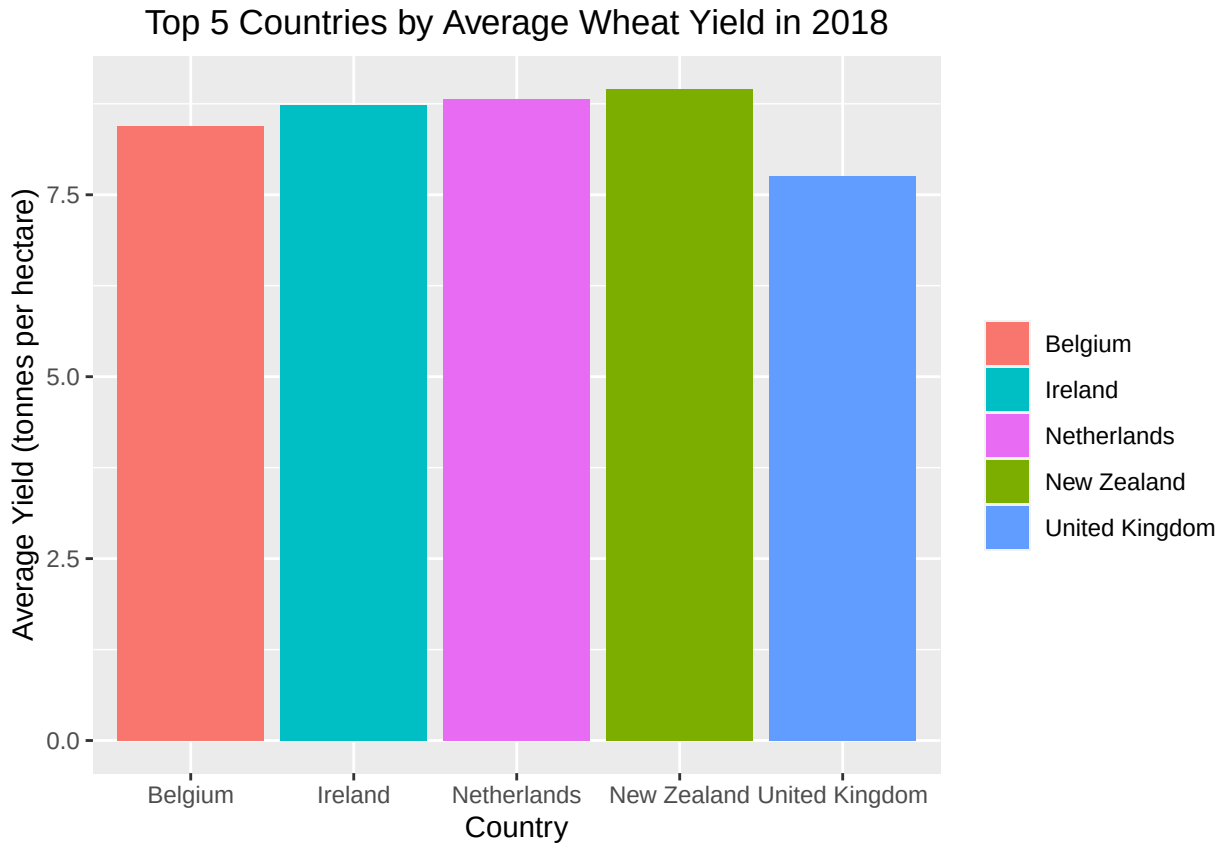
02. Chloropleth map of Wheat yield around the globe in the year 2018

This is a choropleth map of wheat yields by region in the year 2018. Each region is shaded with a color gradient that represents the wheat yield per hectare, with darker shades indicating higher yields. The legend on the right side of the map indicates the range of wheat yields represented by each color.



03. Bar chart of top 5 countries by average crop yield in 2018

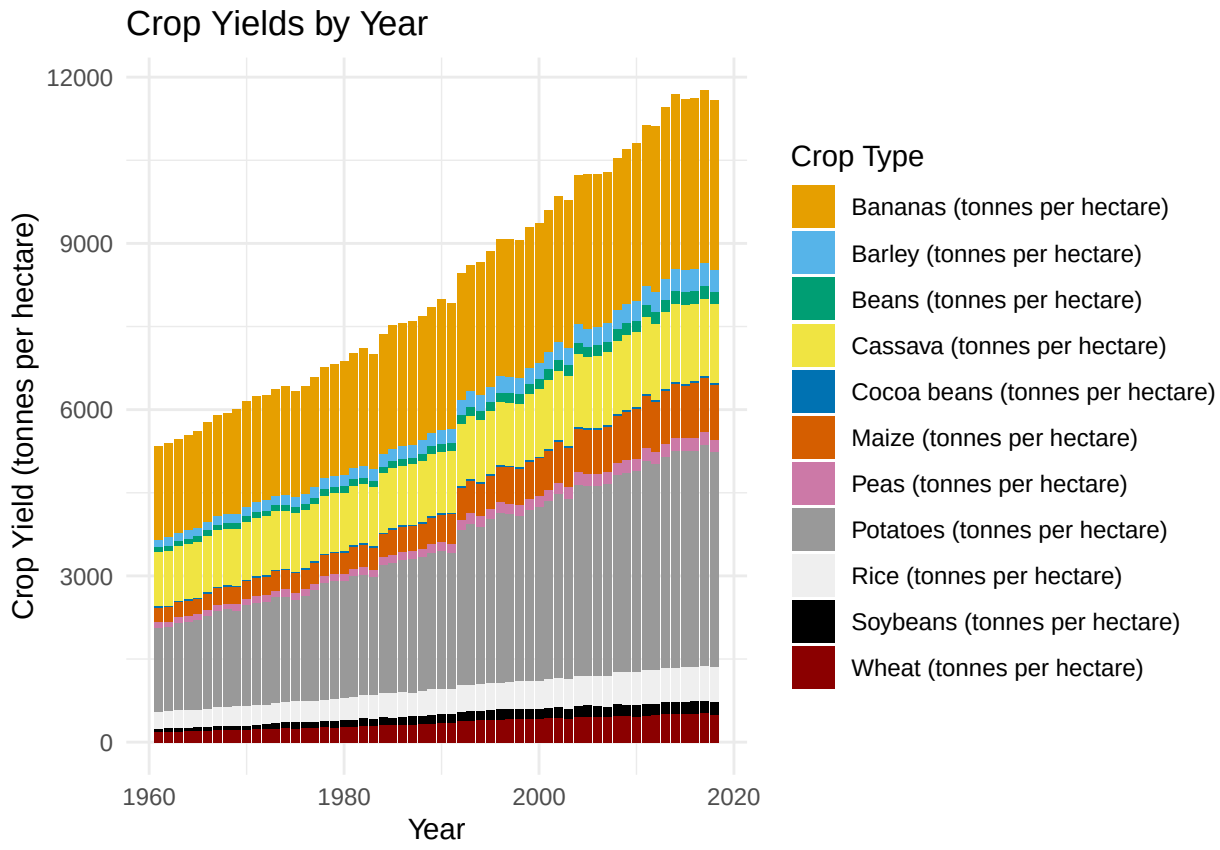
This bar chart shows the average wheat yield in 2018 for the top 5 wheat-producing countries. The countries are ranked from left to right based on their yield, with the country with the highest yield on the left and the lowest on the right.



04. Stacked Bar chart for Yields of different crops from 1961 to 2018.

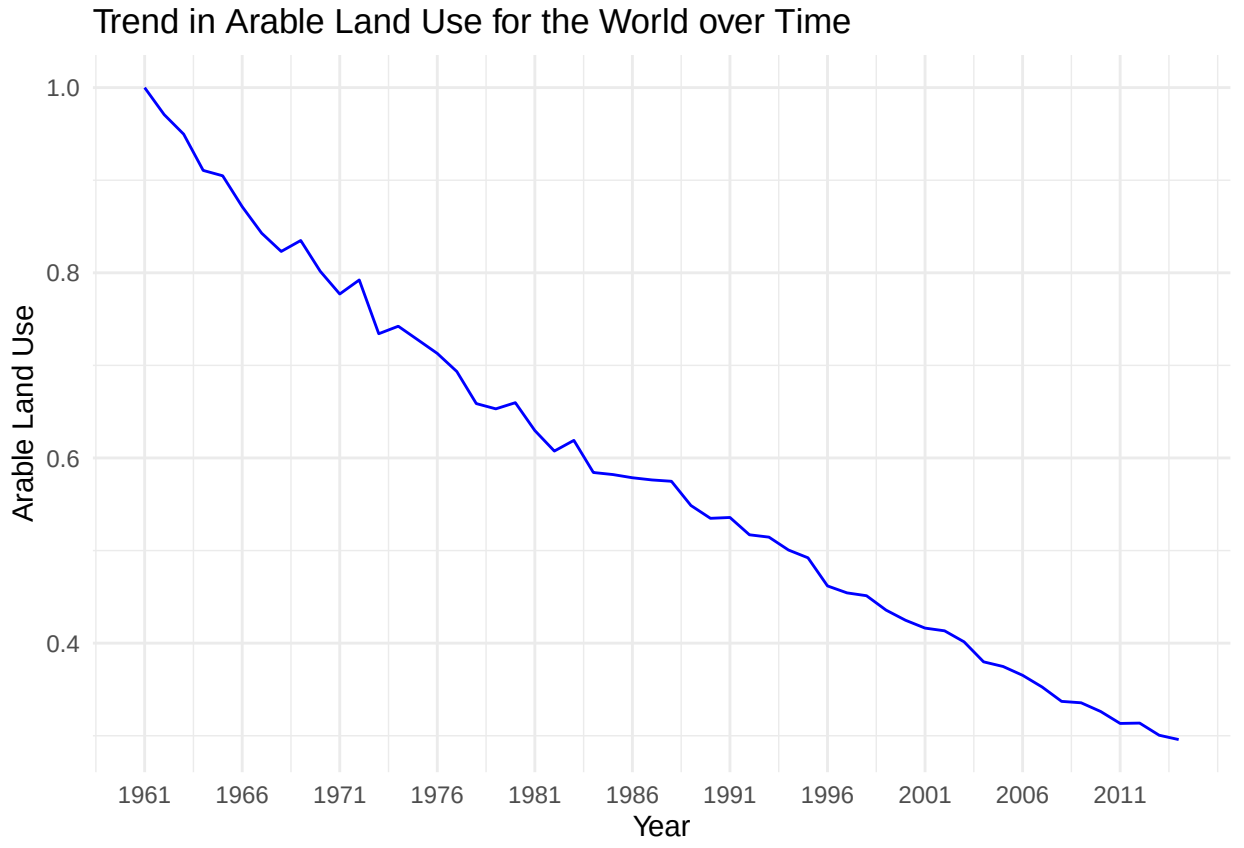
Each bar represents a year and the height of each segment within the bar represents the yield of a specific crop

```
## Warning: Removed 58819 rows containing missing values (‘position_stack()‘).
```



05. Arable land use line chart

A line chart showing the trend in arable land use for the world over time. The x-axis represents the year, while the y-axis represents the arable land needed to produce a fixed quantity of crops. The line plot shows the changes in the amount of arable land needed over time.

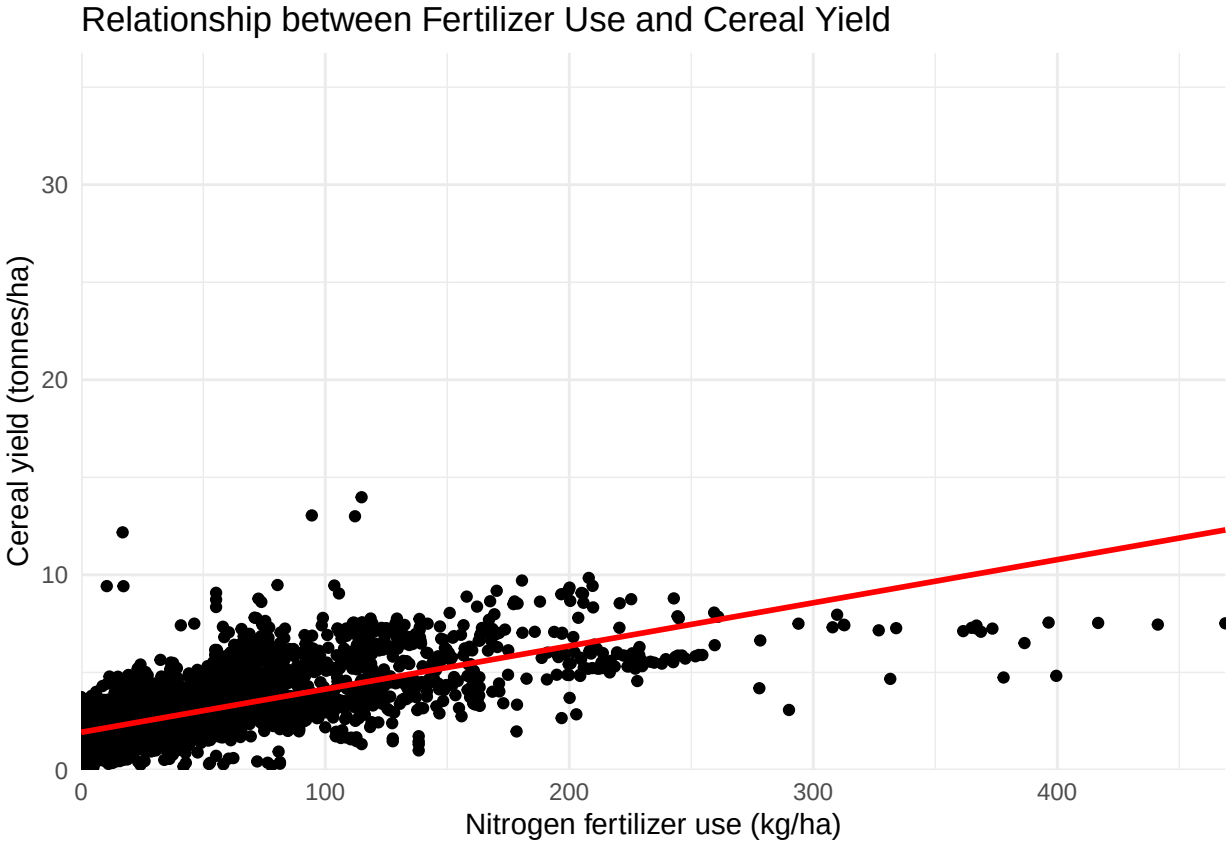


06.Relation between fertilizer use and cereal yield in Scatter plot

A scatter plot showing the relationship between nitrogen fertilizer use (in kilograms per hectare) and cereal yield (in tonnes per hectare). Each point on the plot represents a different country. The red line represents the linear regression line fitted to the data, showing the overall trend in the data. The plot suggests a positive relationship between fertilizer use and cereal yield, with countries using higher amounts of nitrogen fertilizer generally having higher cereal yields.

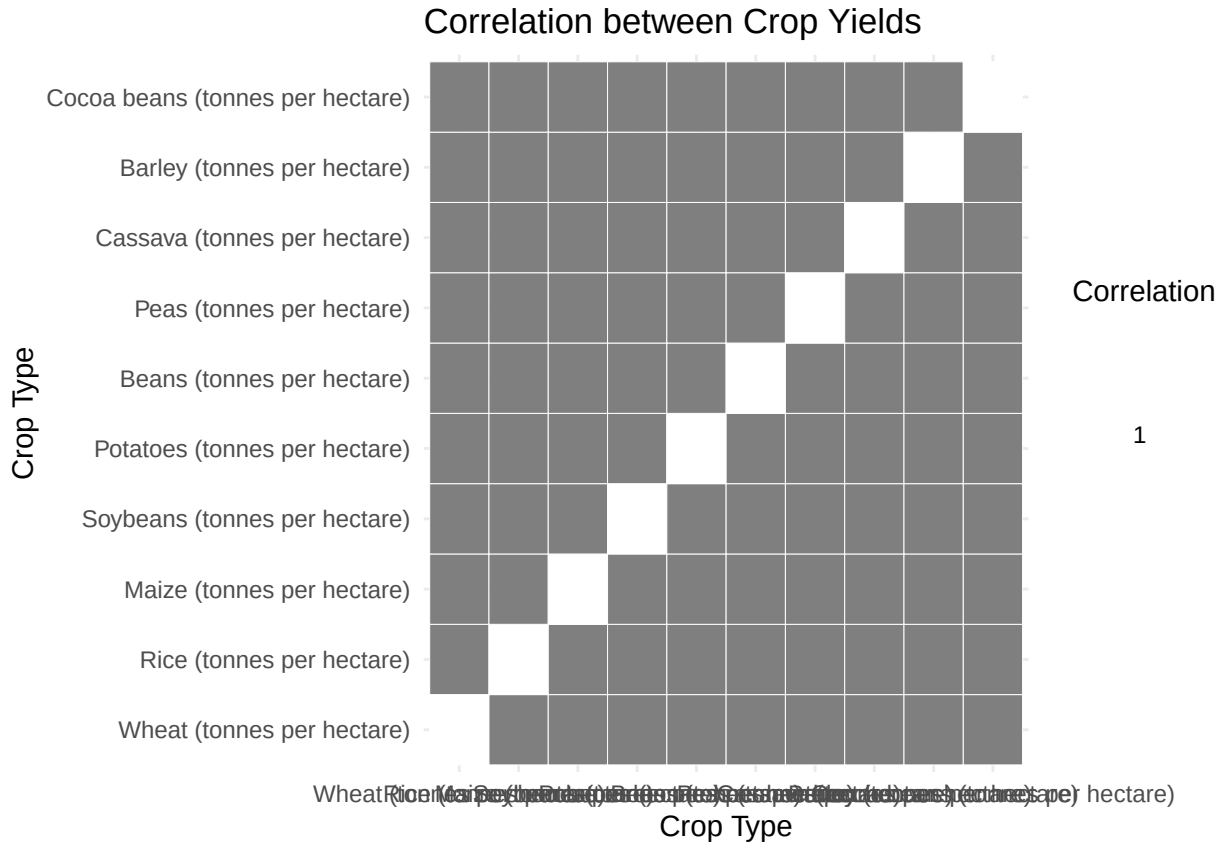
```
## Warning: Removed 9127 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 9127 rows containing missing values (`geom_point()`).
```



07.Heatmap showing the correlation between crop yields.

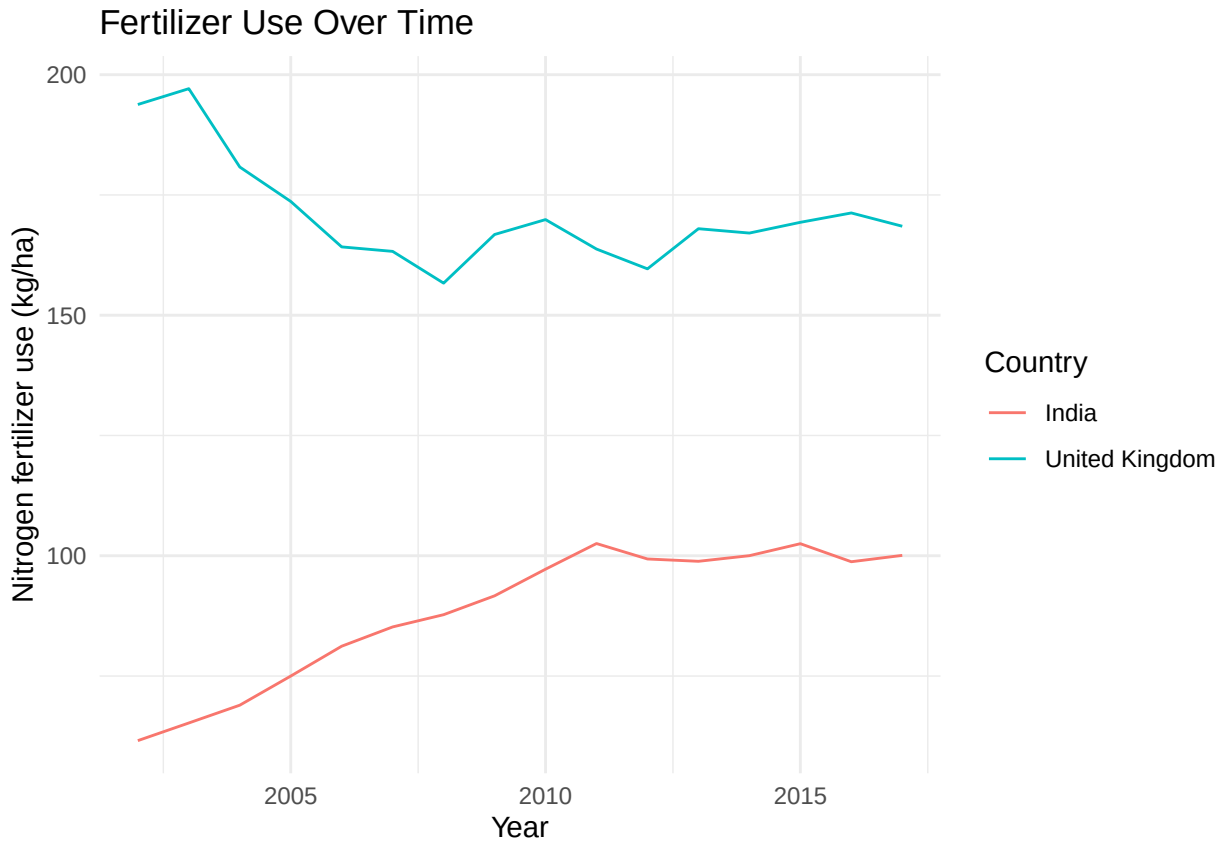
The heatmap visualizes the correlation between different crop yields. The correlation values are displayed as colors, with red indicating negative correlation, green indicating positive correlation, and white indicating no correlation.



08. Comparing two countries for Fertilizer use and yield

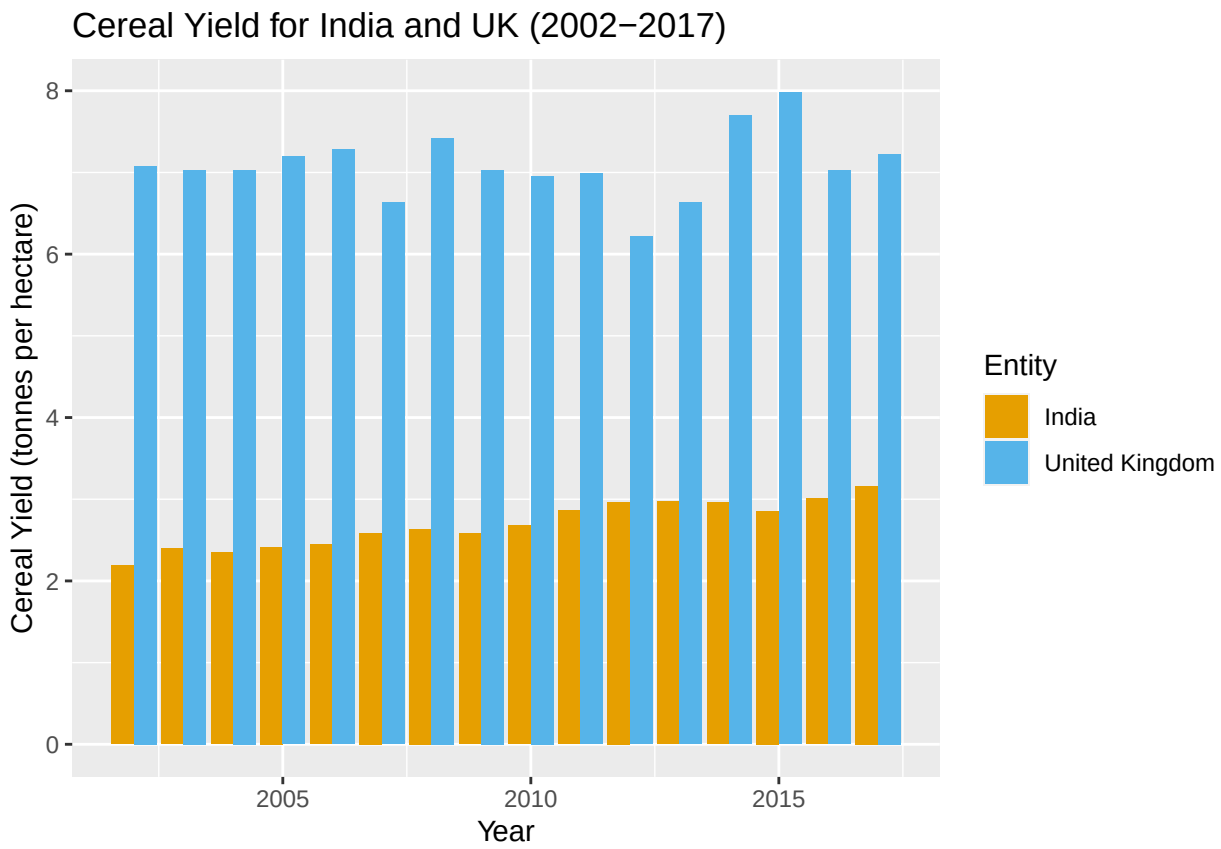
08.1 Line chart

A line chart of the data, with the x-axis representing years and the y-axis representing nitrogen fertilizer use. The lines are colored by country, with India shown in one color and the UK in another. The plot shows the trend in fertilizer use for the two countries over time.



08.2 Dodged bar chart

The plot suggests that the United Kingdom consistently had higher cereal yields than India over the 2002-2017 period. The yields for both countries seem to fluctuate over time, with some years having higher yields



than others.

09. Conclusion

In conclusion, global crop yields are a complex and multifaceted topic that requires a holistic approach to analysis. By examining trends in global crop yields and investigating the impact of various factors like land use, fertilizer use, and mechanization, we can gain a better understanding of the challenges and opportunities facing global food production. With careful planning and thoughtful policy decisions, we can work towards a more sustainable and resilient global food system that can meet the needs of a growing population in the years to come.

10. Critique

10.1 Good Visualization

The data visualization in appendix 1 shows the percentage of people living in poverty in India, divided into different states. The colours used in the map represent different levels of poverty, with the darker shade of red indicating a higher percentage of people living in poverty. The strengths of this visualization are its simplicity and effectiveness in communicating the overall pattern of poverty in India. The map provides a clear visual representation of the percentage of people living in poverty in each state, allowing the viewer to quickly compare poverty levels across different regions. First, the use of a map is generally an effective way to convey geographic data, and the choice of a colour gradient to represent the percentage of people below the poverty line is visually intuitive. Second, the legend is clear and concise, with a simple colour scale and clear category labels. This makes it easy for the viewer to understand the data being presented. Finally, the title and caption provide additional context and information about the data being presented, which is helpful for understanding the significance of the visualization. Overall, while there may be some room for improvement in terms of visual design and interactivity, the image appears to be a clear and effective representation of the percentage of people living below the poverty line in India.

10.2 Bad Visualization

Appendix 2 pie chart is a poor example of data visualization for several reasons. It attempts to show Cricketers' salaries. Firstly, it fails to accurately represent the data. The chart does not add up to 360 degrees, which is the standard measure for pie charts. This makes it difficult to compare the different slices accurately and undermines the chart's effectiveness. Secondly, the colour coding used in the chart is inconsistent and does not serve any purpose. The colours used for each slice of the chart do not convey any meaningful information, and the choice of colours appears to be random. This makes it difficult for the viewer to understand the data and reduces the chart's effectiveness. Furthermore, the labels used in the chart are unclear and do not provide adequate information. The chart lacks a title, and it is unclear what exactly the chart is representing. The labels used for the slices are also not clear, as they do not include any specific numbers or percentages. Finally, the chart is not visually appealing, and it lacks any design elements to make it more engaging. The use of a simple pie chart with no additional visuals or design elements makes the chart uninteresting and difficult to interpret. The inaccurate representation of the data, inconsistent use of colour, unclear labelling, and lack of visual appeal make it difficult for viewers to understand and engage with the data.

11. References

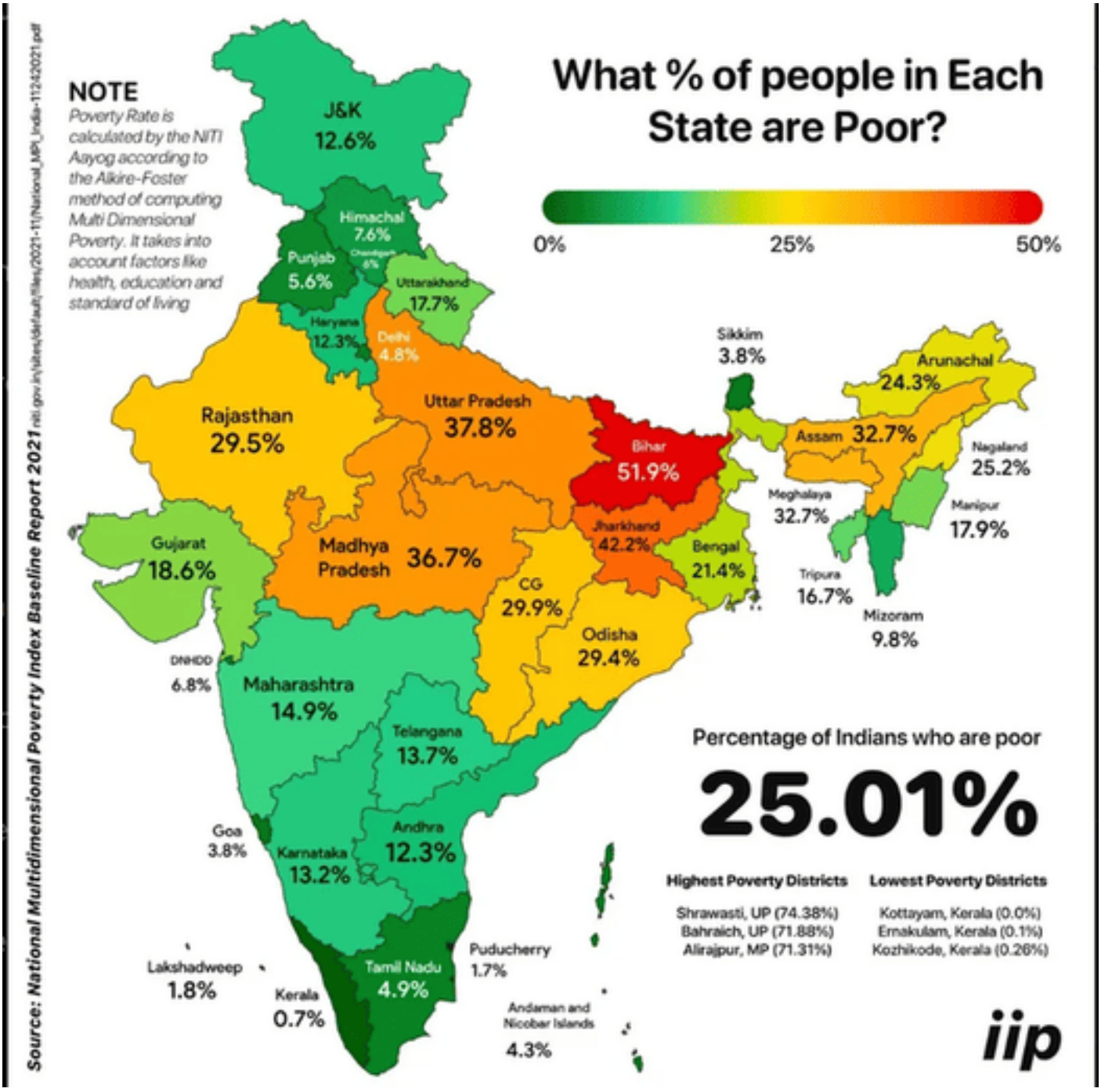
Midway, S.R., 2020. Principles of effective data visualization. Patterns, 1(9), p.100141.

Imgur 2023 from:https://imgur.com/a/uvjRNYu [Accessed on 02-03-2023]

Reddit 2023 u/cough___potato [online] Available from:https://www.reddit.com/r/IndiaSpeaks/comments/11ioey4/we_might_all_knew_this_but_organized_data_helps/ [Accessed on 05-03-2023]

12.Appendices

12.1 Appendix 1,Example for good visualization



12.2 Appendix 2,Example for bad visualization