Agricultural Commodities x Natural Disasters
Anna Leoni
22 August 2024

What do ice cream, laundry detergent, packaged bread, lipstick and biodiesel all have in common? Although it would be accurate to say that these are all items one can get from the local grocery store and gas station, there is another connection: palm oil [8]. Palm oil has experienced a meteoric rise in the last few decades, with world production of palm fruit quadrupling from 1995 to 2015; not to mention, it is now found in close to 50% of products at said grocery store [7]. In a changing climate and economic landscape, it may be worth considering where all of this palm oil comes from that we find in so many of our daily products. A quick search on Google returns a somewhat scary statistic, that in the 2023/2024 market, just one country produced 59% percent of the world's palm oil [4]! Facts like these inspired this research project.

In the face of a changing climate that exhibits "increases in the frequency and intensity of hot extremes, marine heatwaves, heavy precipitation, and, in some regions, agricultural and ecological droughts"[3], it is crucial to look into the relationship between weather events and agricultural production to understand our current situation. To have such market instability that a single country is responsible for the majority of production of a commodity could wreak havoc on the global market if a natural disaster or weather event destroyed production outlook. This research project aims to begin the exploration of the relationship between natural disasters and production levels for the top agricultural commodities over a thirty-year period. Data was collected from FAOSTAT (Food and Agriculture Organization of the United Nations) for commodities data, EM-DAT (The International Disaster Database) for natural disaster data, as well as the WorldBank's World Integrated Trade Solution for GDP data.

To outline the approach to this problem and what follows in this paper, the process went as such: first determine the top commodities produced globally, then uncover which countries produce the most of these top items. Finally, examine disasters in these countries and how they affect the country's production of the commodity.

To accomplish the first task of determining the top commodities globally, global data from FAOSTAT was loaded, cleaned, and filtered to the years 1990-2020. The dataframe was then sorted and the top commodities along with their production (in tons) was returned. The top 5 commodities out of the 211 in the data set were 'Sugar cane,' 'Maize (corn),' 'Rice,' 'Wheat,' and

commodity_names	
Sugar cane	4.621549e+10
Maize (corn)	2.392682e+10
Rice	1.992251e+10
Wheat	1.983415e+10
Raw milk of cattle	1.742084e+10
Potatoes	1.001947e+10
Sugar beet	8.093485e+09
Other vegetables, fresh n.e.c.	7.074710e+09
Cassava, fresh	6.841776e+09
Soya beans	6.729444e+09
Oil palm fruit	6.403669e+09
Barley	4.560225e+09
Tomatoes	4.035462e+09
Sweet potatoes	3.524895e+09
Meat of pig with the bone, fresh or chilled	3.008732e+09
Bananas	2.658771e+09
Raw milk of buffalo	2.607484e+09
Watermelons	2.434383e+09
Meat of chickens, fresh or chilled	2.309549e+09
Grapes	2.038514e+09

Figure 1.
Top 20 global commodities cumulative production (in tons) for years 1990-2020.

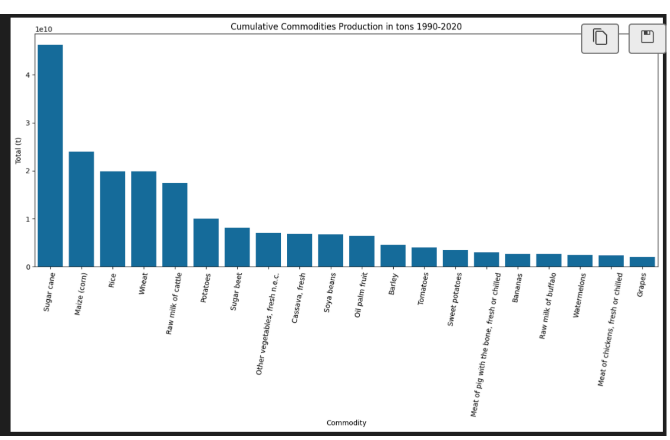


Figure 2.
Top 20 global commodities cumulative production (in tons) for years 1990-2020

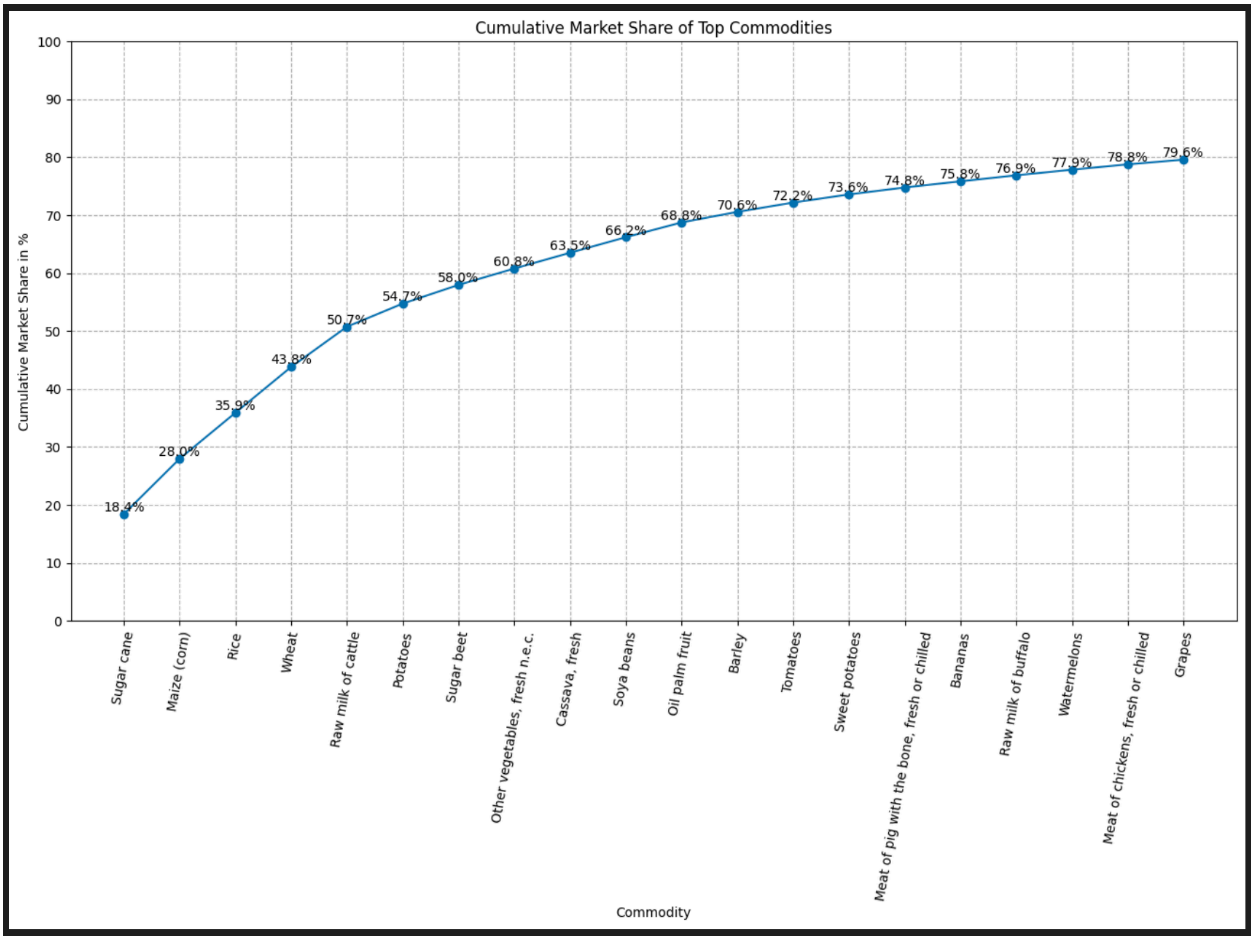


Figure 3.
Cumulative Market Share of Top Commodities 1990-2020

'Raw milk of cattle;' and accounted for 50.7% of all of the tons of commodities produced in the market between 1990-2020, with sugar cane claiming a whopping 18.4%. Given the press that palm oil has received lately, it seemed appropriate to include it in the study. 'Oil palm fruit' was ranked 11th in production of top global commodities, so this project focused on the top 11 commodities to include it. To visualize this information, a summary of the top commodities and their cumulative production in tons for the years 1990-2020 in Figure 1, as well as a bar chart of the tons of production and line plot of 'Cumulative Production Share of Top Commodities' were produced and can be seen in Figures 2 and 3.

The next step was finding which countries contributed the most to the production of these items. To do this, a more detailed dataset that included data for individual countries and regions from FAOSTAT was loaded and cleaned. Graphs for each commodity mapped the production for top-producing countries over the years 1990-2020. There were some noticeable differences in the various commodities graphs. The function to produce the graphs limited the number of countries to be included to the top 5, however, given different production trends, some commodities had a consistent 5 to 6 countries across all 30 years and some had up to 11 different countries across that time period that were counted in the top 5 in various years. This seemed interesting, but not concerning, unlike another trend in the graphs. There seemed to be several commodities, namely sugar cane, palm fruit, vegetables and corn that had one country that accounted for around 40%

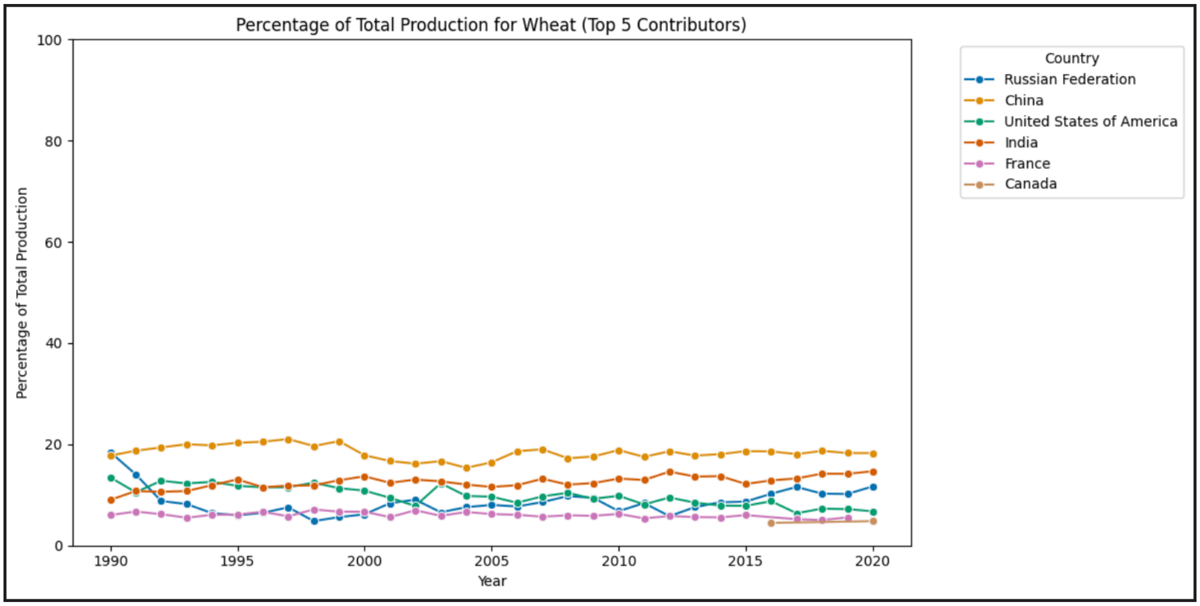


Figure 4.
Percentage of Total Production for Wheat (Top 5 Contributors). Exhibits a balanced share in production.

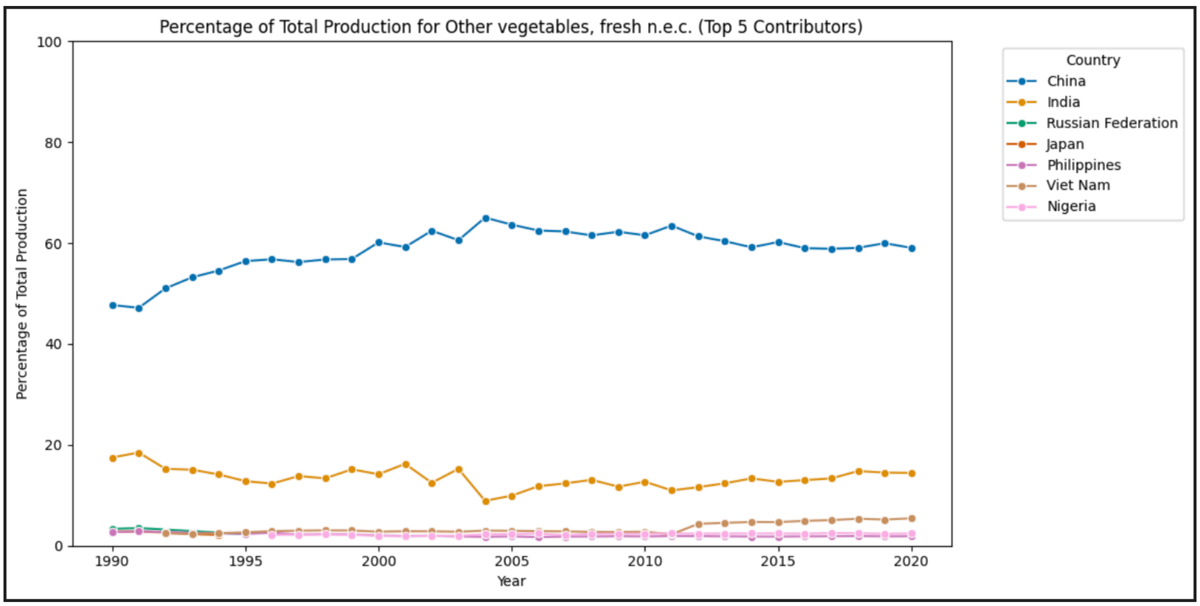


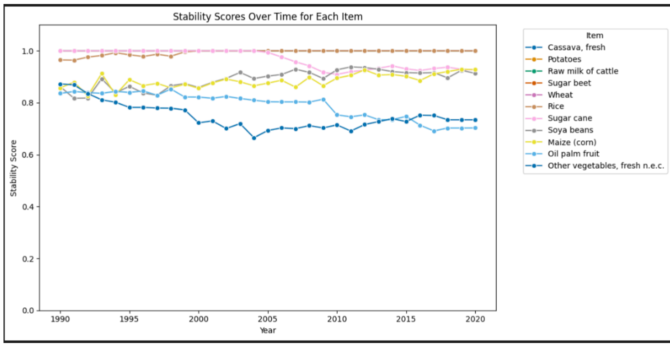
Figure 5.
Percentage of Total Production for Other Vegetables (Top 5 Contributors). Exhibits an imbalance in production percent across different countries with China producing around 60% after 2000.

of production. To explore this phenomenon more, feature engineering for contribution categories and stability analysis was conducted.

The feature engineering for contribution categories started out with calculating production quartiles across all countries, but all of the top countries plotted ended up being in the 'XL contributor' category this way. The next approach was calculating quartiles for only the top contributors included in the graphs, but China ended up being an 'XL contributor' for both Wheat and Other vegetables, which, as seen above in Figures 4 and 5 may be true, but the goal is to differentiate between these percentages rather than report the top contributor. The final and implemented approach was to create percentage buckets with 0%—10%:S, 11%—20%:M, 21%

—39%:L, and >40%: XL. Although perhaps somewhat arbitrary, it did capture and classify the desired information as intended.

To explore the stability of each commodity, a customized version of the Herfindahl-Hirschman Index calculation was applied to each year of production for the commodity and analyzed. This index provides information on market share to identify monopolies and quantify market competition (5). For the sake of this analysis, the HHI was normalized to be between 0 and 1 and then inverted so that 1 meant stable and 0 meant really unstable. To get an idea of the trend over the years, this resulting stability score was averaged for the years 1990-1995 and then for the years 2015-2020 and the difference was calculated to determine if stability had increased,



Item	Overall Average Stability (1990-2020)	Trend
Cassava, fresh	1.000000	Stable
Maize (corn)	0.885106	Increasing
Oil palm fruit	0.789766	Decreasing
Other vegetables, fresh n.e.c.	0.746859	Decreasing
Potatoes	1.000000	Stable
Raw milk of cattle	1.000000	Stable
Rice	0.993666	Increasing
Soya beans	0.890899	Increasing
Sugar beet	1.000000	Stable
Sugar cane	0.967533	Decreasing
Wheat	1.000000	Stable

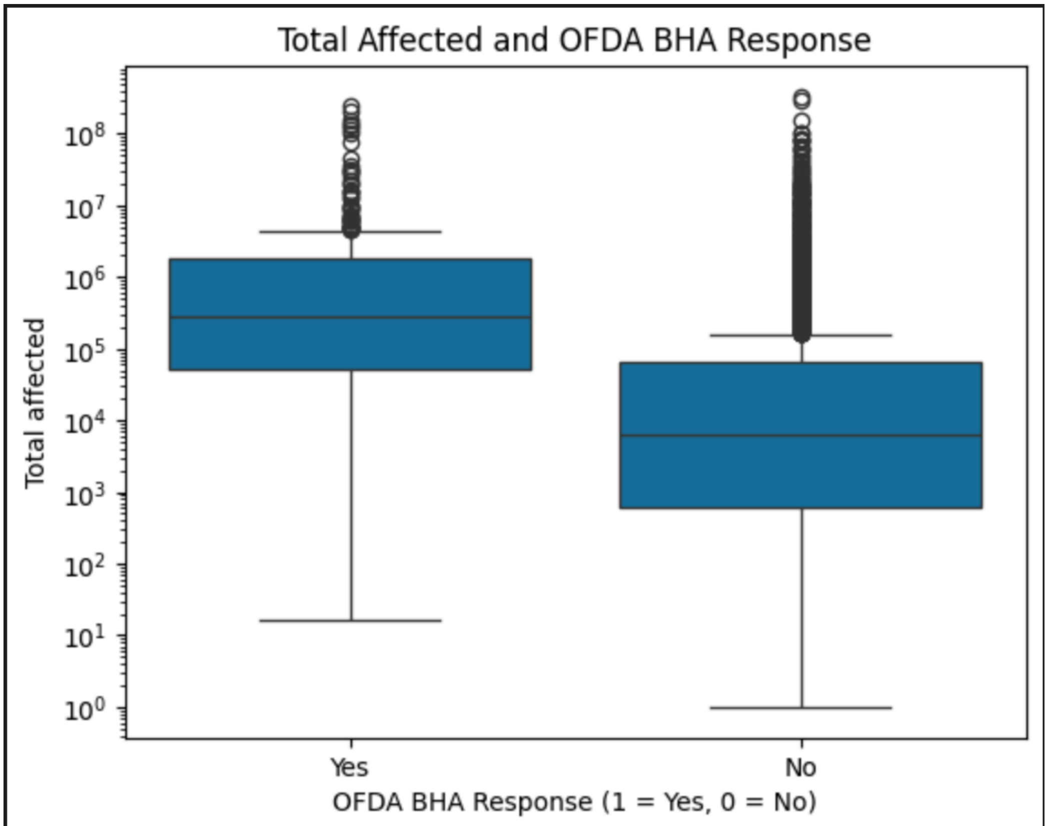
Figure 6.
Commodity stability over time 1990-2020.

Figure 7.
Stability summary table.

decreased, or remained stable over the course of the 30 year period. The stability score for each commodity market and each year was plotted and a table was created to summarize the information. The data in the ‘Contributor Category’ and the ‘Stability’ features were valuable information to include in the dataframe, but they were difficult to add due to different structures. The Commodities data frame is grouped by commodity and country and year. The stability index only applies to a single commodity per year. Instead of making a complicated multi-index that would eliminate duplicate information in the Stability column, it made more sense to be aware that stability applied to a year and a crop, not a specific country. All countries that produce an item in a year share the same entry in the Stability index. The contributor category applied to country, item and year.

Trade data was the next area of interest in gathering data. Originally, the plan was to look at import and export data and GDP. GDP data was collected from World Bank’s WITS data and import and export data was again collected from FAOSTAT. Due to the commodities in the export data being different than the production data, exports had to be excluded from the analysis. As an example of this, in the export data, Raw sugar from sugar cane and sugar beets was grouped together and palm oil was the export whereas in production there were individual measures of cane sugar and beet sugar and palm fruit for oil before being processed. Additionally, import data for all commodities was included in the imports metrics. Further research to determine the top imported commodities during a natural disaster needs to conducted to filter this data for a more accurate imports and therefore more accurate description of events.

The final group of data to be included in this analysis was natural disaster data. This was collected from EM-DAT. This data was quite comprehensive as far as variables, but contained a lot of NaNs making it not as accessible as hoped. Another problem with the NaNs was that they eliminated a lot of features for measuring disaster intensity: ‘Magnitude’, ‘Total Affected’, ‘Total Deaths’, ‘Reconstruction Costs’, ‘Total Damage’, etc. To solve this problem, a bootstrap analysis was conducted on mean (and median) difference between disasters that required an OFDA/BHA




```
Group Yes Size: 295
Group No Size: 4009
Bootstrapped Mean for Group Yes: 6149099.486471185
95% Confidence Interval for Group Yes: [3609057.00957627 9448644.07415254]
Bootstrapped Mean for Group No: 892328.1287031679
95% Confidence Interval for Group No: [ 666862.68899351 1177514.99564729]
Mean Difference: 5256771.357768018
```

Figure 8.
OFDA/BHA Response and Total Affected, proxy calculation results

Response (OFDA - Office of U.S. Foreign Disaster Assistance, BHA - Bureau for Humanitarian Assistance) and those that did not with Total Affected, as it was the most complete out of these potential measures. This resulted in a statistically significant difference between means (and medians too), with the disasters requiring aid also affecting more people (Figure 8.). These results encourage the idea that the OFDA/BHA Response binary could serve as a proxy for disaster intensity, with the idea that a disaster that required a response was likely to be more severe than a disaster that did not.

A bar-plot of the natural disasters revealed that floods and storms were by far the most common type of disaster. A stacked bar-plot of disaster types by country revealed that China,

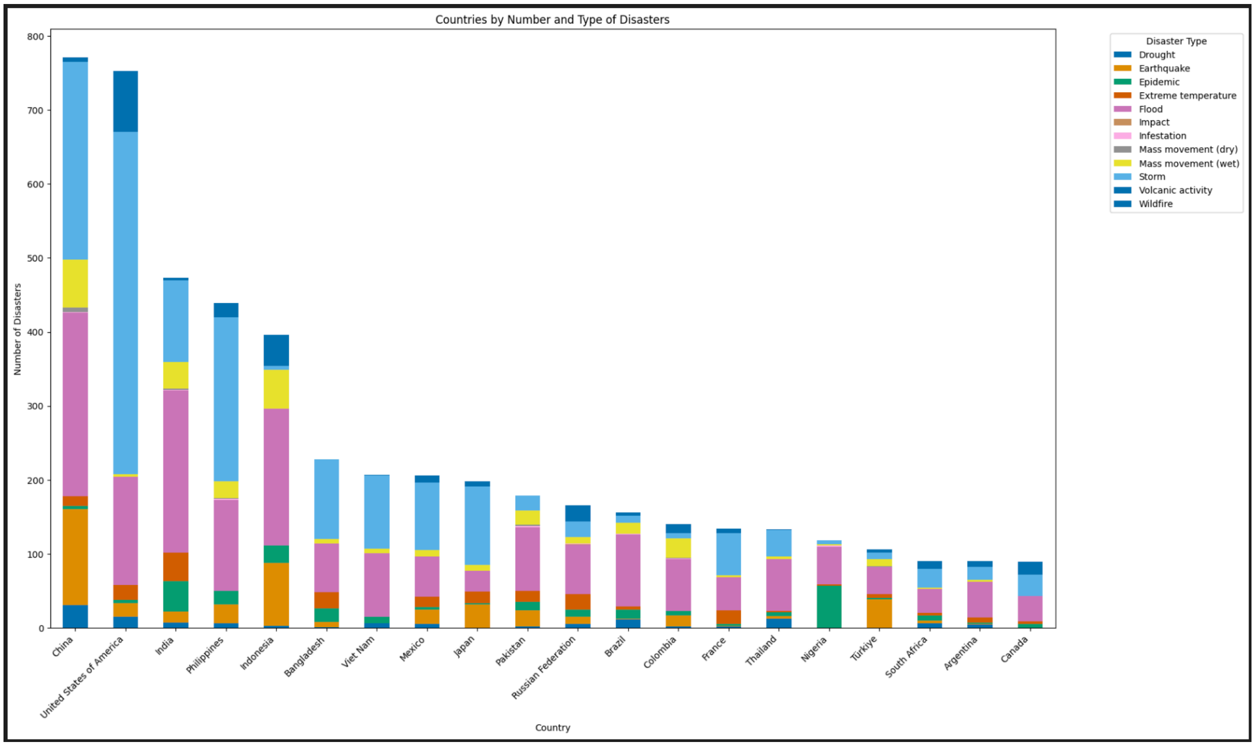


Figure 9.
Stacked bar-plot of countries by number and type of disasters.

USA, India, Philippines and Indonesia experienced the most disasters in this 30 year period, as seen in Figure 9. Two Plotly bubble charts were also produced in the notebook, but serve much better as an interactive tool than a two dimensional visual.

The final part in the approach to solving this problem was combining the data into a single dataframe and analyzing it. This was the most challenging part of the process given the different structures of the dataframes. GDP and Import data were straightforward to combine because each had a single year, country and value. Much like the stability index, this data could be combined with the commodities dataframe acknowledging that the data applied to a country and year rather than the commodity produced. The challenge began when trying to incorporate natural disaster data.

The natural disasters dataframe had a much different structure and was not organized by country or year, but rather Disaster Group, Disaster Subgroup, Disaster Type, and Disaster Subtype. It had a Start Year and an End Year, but in order to get a more accurate idea of the disaster timeframe on commodity production, it was necessary to adjust this. One preliminary, and soon discarded, idea was to make dummy variables to handle different disaster types and corresponding dummy variables for the OFDA/BHA Response. However, if there were multiple floods, for example, this method could easily spiral out of control with columns. Instead, a Year Range column (a list of years) was determined by start and end years and then exploded so that each disaster, regardless of duration, was noted as occurring in the year that it spanned. This added more complexity to the structure of the dataframe because ‘Country’ and ‘Year’ columns could have duplicates if multiple disasters happened in a single year in that country. A complicated multi-index approach was considered and then dropped in favor of combining this form of the natural disasters dataframe with the merged dataframe. Although this approach did add more rows to the merged dataframe, it ultimately met the needs for this analysis.

To get an idea of the combined data, Expected Production was calculated on a rolling 3 year average and plotted with Production for each Item, as well as Disaster Count for each Country as seen in Figure 10. Next, the percent change in Production by Disaster Type was plotted on a heat map. It appears that commodity production tended to to have a positive relationship with floods. Production, Disaster Count, Disaster Duration and OFDA/BHA

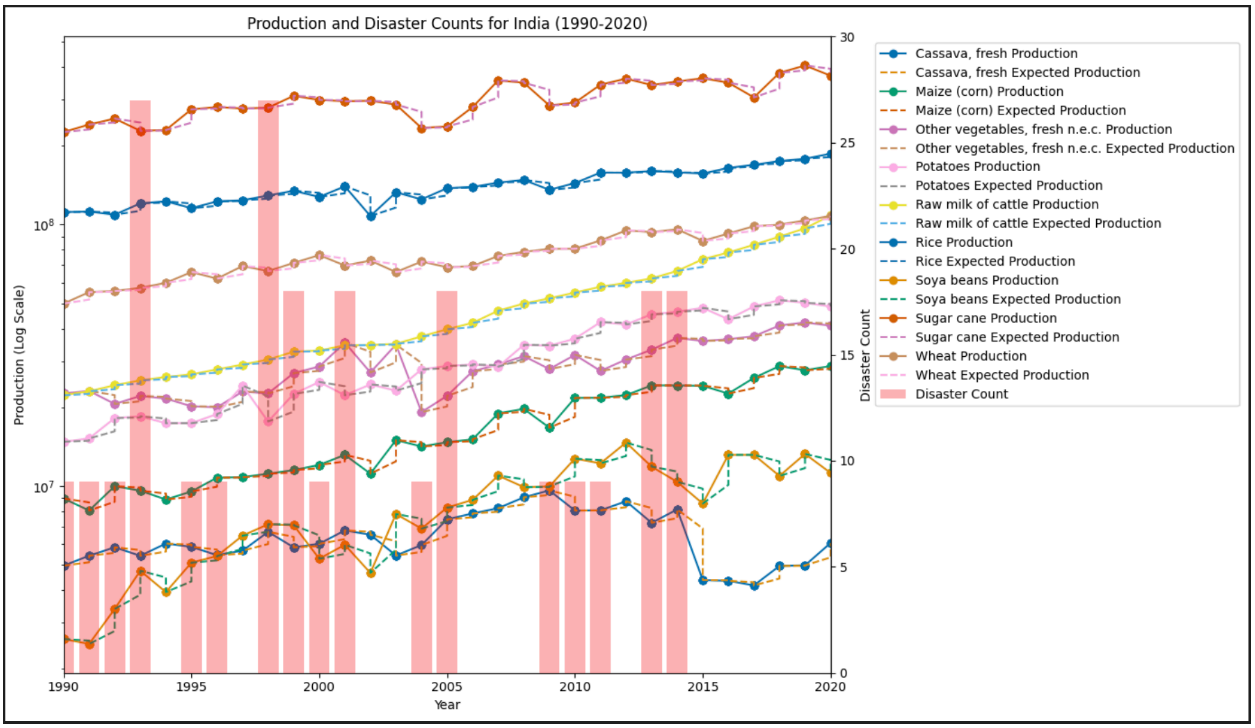


Figure 10.
Production and Disaster Counts for India.

Response were plotted on a correlation heat map. This showed all very weak correlations. To explore the suspicion that more severe disasters might increase imports, a box-plot of the severity proxy, OFDA/BHA Response, was plotted against imports. There didn’t seem to be much of a difference in mean, with the less severe disasters (OFDA/BHA Response == 0) having more variance and slightly higher imports. Finally, after determining the non-normally distributed nature of the data and that the OFDA/BHA Response Count was indeed a count variable, a poisson generalized linear model was performed using the statsmodel library on

‘Year’ and the more severe disasters, filtered by disasters that required a response in ‘OFDA/BHA Response Count,’ and their effect on ‘Production.’

The results from this analysis indicate statistically significant relationships for all country-commodity pairs. These results should be approached with caution. The magnitude of these

```
Significant results for Brazil, Cassava, fresh:
==============================================================================================
                              coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------------------
const                      21.5722      0.005   4295.568      0.000      21.562      21.582
Year                       -0.0023    2.5e-06   -922.282      0.000      -0.002      -0.002
OFDA_BHA_Disaster_Count    -0.0139   7.02e-05   -198.197      0.000      -0.014      -0.014
==============================================================================================
The p-value for OFDA_BHA_Disaster_Count is 0.0000
The impact is negative (disasters associated with decreased production).
Pseudo R-squared (deviance explained): 0.0273

==================================================

Significant results for Brazil, Soya beans:
==============================================================================================
                              coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------------------
const                    -113.9277      0.003    -3.3e+04      0.000    -113.934    -113.921
Year                        0.0656   1.72e-06    3.82e+04      0.000       0.066       0.066
OFDA_BHA_Disaster_Count     0.0225   4.36e-05    516.862      0.000       0.022       0.023
==============================================================================================
The p-value for OFDA_BHA_Disaster_Count is 0.0000
The impact is positive (disasters associated with increased production).
Pseudo R-squared (deviance explained): 0.9808

==================================================
```

Figure 11.
GLM results for Brazil, Cassava and Soya Beans

effects varied greatly across different commodities and countries. The range of pseudo R-squared values is also of note. In the case of Brazil, cassava had a pseudo R-squared of 0.0273 and for soybeans had a pseudo R-squared of 0.9808 (Figure 11). This shows that the model had varying levels of explanatory power across the different commodities and countries. Countries and commodities also had different responses to disasters, with some commodity production levels responding positively and some negatively. More exploration into disaster type could prove useful in understanding this outcome, since this differentiation was not taken into account when looking at disasters and production for the countries and commodities (Figure 12). As an example, cassava does not survive well in flood conditions [5], whereas rice thrives in flood conditions but struggles in droughts [1]. Running a more detailed analysis based on disaster type would be advised. Additionally, this model doesn’t account for time lags in disasters effects, which could further complicate the findings.

The results of this exploratory data analysis are not conclusive, though it seems that natural disasters had varying effects on the commodities and countries included in the analysis. This could be due to a few things. Due to a lack of geographic information regarding where commodities are produced and geographic information about disasters, a disaster occurring far away from a crop-producing field in the same country will have little to no effect on the production of the crop, yet still be counted as an event in the analysis. Another point previously brought to attention in the exploratory data analysis phase was that there is no effective measure

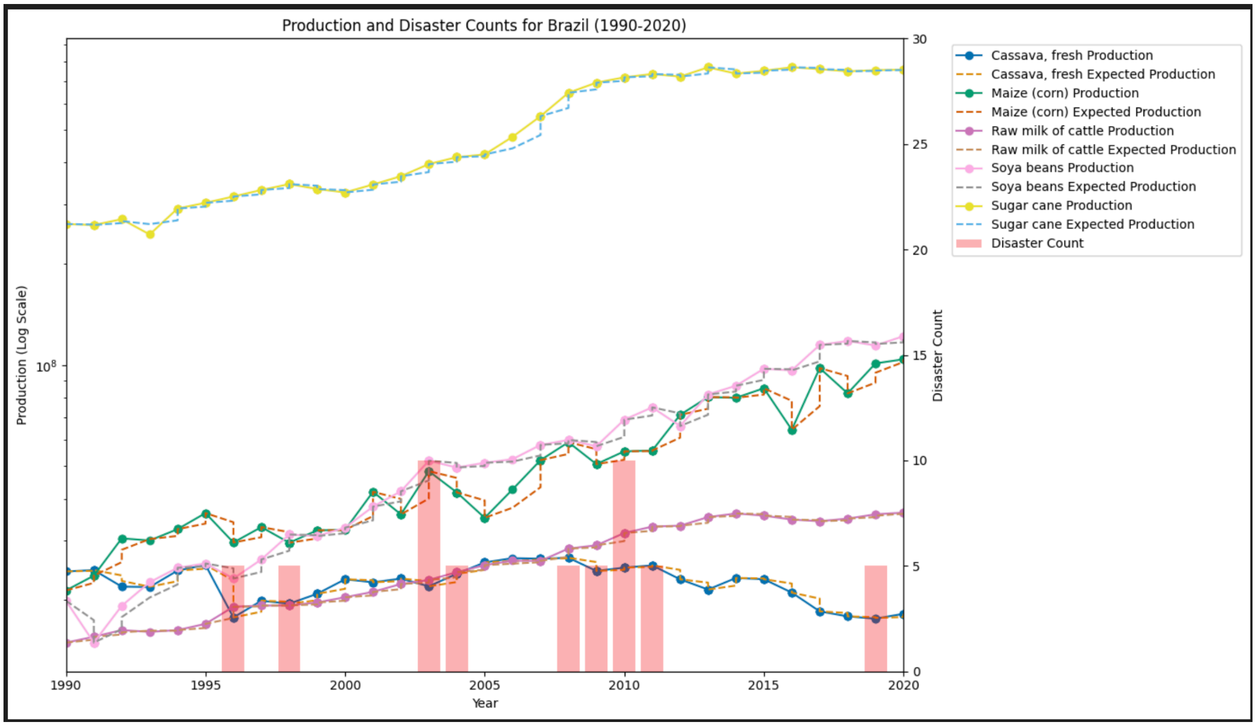


Figure 12.
Production and Disaster Counts for Brazil 1990-2020

of severity of the disaster provided in the data. Although OFDA/BHA Response served as a binary proxy (severe or not severe), there are problems with accepting this as a good proxy. The primary issue with this is that disasters that occur in metropolitan areas, or densely populated areas, are likely to affect more people than disasters that happen in rural or sparsely populated regions. With this in mind, if a disaster occurs in a city, affects many people, and gets an OFDA/BHA Response, this may be an adequate measure of intensity, but not as it relates to the farmland or commodity producing regions. This analysis only looked at the top 11 commodities, and although they account for 68.8% of all the agricultural commodities production in the 30-yr period, including all commodities might give a more holistic picture of the relationship. More research into natural disaster trends over time as well as a time series to determine seasonality and the upward trends in production to help balance findings would be beneficial to this analysis. A final point is that it has been suggested that the effect of extreme weather on crop production may not have had a noticeable or measurable impact until 2019 [6]. Including more years would also be beneficial to this analysis.

Although these results are not conclusive, it is important to consider the inflexibility of our agricultural practices in a changing climate. With individual countries contributing over 40% (and some close 60%) to global production of a single product, several decreasing commodity market stability scores, and the increasing frequency and severity of weather events, it may be beneficial to commodities markets to diversify production locations. With a growing population and limited resources, the future of the food supply chain may be at stake.

Resources:

1. Arora, N. K., Fatima, T., Mishra, I., Verma, M., Mishra, J., & Mishra, V. (2022). Rice yields and quality under climate change: A review. Scientific Reports, 12(1), 1-18. Retrieved from https://www.nature.com/articles/s41598-022-06123-9
2. Hayes, A. (2023). Herfindahl-Hirschman Index (HHI). Investopedia. Retrieved from https://www.investopedia.com/terms/h/hhi.asp
3. IPCC. (2021). Summary for policymakers. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Retrieved from https://www.ipcc.ch/report/ar6/wg1/chapter/summary-for-policymakers/
4. United States Department of Agriculture Foreign Agricultural Service. (2023). Production, supply and distribution. Retrieved from https://fas.usda.gov/data/production/commodity/4243000
5. USDA NRCS Plant Materials Program. (2011). Cassava Manihot esculenta Crantz. Retrieved from https://plants.usda.gov/DocumentLibrary/plantguide/pdf/pg_maes.pdf
6. Vaughan, A. (2023). The climate impact of feeding ourselves is getting worse and worse. New Scientist. Retrieved from https://www.newscientist.com/article/2443472-the-climate-impact-of-feeding-ourselves-is-getting-worse-and-worse/
7. Verité. (n.d.). Palm oil. Retrieved from https://verite.org/project/palm-oil-3/
8. World Wildlife Fund. (n.d.). Which everyday products contain palm oil? Retrieved from https://www.worldwildlife.org/pages/which-everyday-products-contain-palm-oil