Data Bootcamp Final Project

Data Analysis on Deregulation of Coffee Market

Joanna Patsalis and Victoria McClemens

Data Bootcamp Final Project
Data Analysis on Deregulation of Coffee Market
Joanna Patsalis and Victoria McClemens

Abstract

In this project, we analyzed the coffee industry in an effort to understand the impact that the deregulation of the International Coffee Agreement (ICA) in 1989 had on the production and prices for growers of coffee. We learned that production was dominated by one major player (Brazil) and that prices and production have risen over the past 20 years with periods of intense volatility. We also learned that there is weak correlation between the price and production of coffee. In future work, we plan to build on our analysis to understand consumption trends, trends by different types of coffee, and the correlations between price, production, and variables including economic indices and climate change proxies.

Business Understanding

As coffee drinkers, we were intrigued to learn a bit more about some of the economic forces in the coffee industry that allow us to enjoy our favorite beverage. We've noticed many trends in our daily lives including: more coffee shops on every block in NYC, more options within coffee shops, more emphasis on organic and fair-trade coffee, etc. Given that coffee is a major export commodity that is enjoyed globally, we planned to analyze the beginning of the supply chain and understand what trends were existing among producers.

Until 1989, trade in coffee was regulated by the International Coffee Agreement (ICA) - similar to an OPEC for coffee - with the aim of keeping prices stable between $1.20-$1.40 per pound.[1] However, in 1989, the International Coffee Organization (ICO) who oversees the ICA, failed to come to an agreement on export quotas. "This breakdown was disastrous for many people along the supply chain, as the ICO's composite indicator price for coffees dropped by nearly 75% in the subsequent 5 years, from $1.34 USD per pound in 1989 to an average of $0.77 per pound through 1995."[2] Deregulation also created increased supply opportunities in the free market that we see today. Furthermore, the popularity of coffee has increased significantly over this time frame. Global consumption is at all-time high (evidenced by the increases in the number of coffee shops and M&A activity of coffee chains).[3]

---

[1] https://www.theguardian.com/business/2003/may/29/fairtrade.famine
[2] https://dailycoffeenews.com/2014/07/17/a-brief-history-of-global-coffee-production-as-we-know-it-1963-2013/
[3] https://dailycoffeenews.com/2018/03/21/current-coffee-consumer-trends-inside-the-ncas-2018-report/

Given the increased consumption of coffee combined with the recent deregulation, we expect to see either increased production with consistent pricing or stable production with increased pricing. We also expect the industry to become more concentrated as the larger producers (historically, Brazil has been the largest producer) are able to benefit from economies of scale and produce more coffee at cheaper prices than smaller-producing countries. Our plan was to create a series of graphs to understand the following:

- Which are the major countries producing Robusta coffee and how has their relative strength changed since deregulation began?
- How has the revenue for these countries changed over time
- Have overall production levels increased since deregulation (in keeping up with the global increases in consumption) and by how much?
- How has the price to growers changed over the deregulation period?
- What is the correlation between price and production

Data Understanding

We used ICO's Crop Data as the data set for our analysis. We found an interesting data set on Kaggle studying coffee production and consumption that inspired our project: https://www.kaggle.com/sbajew/icos-crop-data. This data set included the following variables: country (of production), year, month, total production (measured by 60kg bags of green coffee), domestic consumption, exportable consumption, and gross opening stock. The entire data set had 1568 rows and 7 columns.

The second data set we brought into our analysis was ICO's growers price data: http://www.ico.org/new_historical.asp. Pricing data was separated by the 2 types of coffee. This dataset included the 2 main types of coffee: Arabica and Robusta. For a quick note on their differences: Arabica beans are higher quality than Robusta and can charge a premium in the market. "Arabica beans are said to have a softer, sweeter and less harsh taste. Specialty roasts and high-end coffee is a luxury that many Westerners treat themselves to. This is the foundation on which Starbucks has built its empire - luxury for the common man. Robusta coffee beans are largely used for instant coffees, and as specialty roasts become more desired in Western consumers' minds, it's easy to see why Robusta hovers quite low."[4] To simplify our analysis and merge our datasets, we chose to study Robusta because we had more price detail on Robusta.

---

[4] https://datahero.com/blog/2014/03/20/coffee-production-visualization-through-the-years/

Data Bootcamp Final Project
Data Analysis on Deregulation of Coffee Market
Joanna Patsalis and Victoria McClemens

Data Preparing

We went through many steps to clean the data before beginning our analysis. First, we imported a series of packages such as pandas and numpy to our initial ICO data set. We had to change many column names, create several variables, and drop others. Below are some of the steps we took.

1. We changed the 'Year' column. This column was initially a string and showed consecutive years (i.e. 1990/1991). We changed this column to just show the first year of the data point and we changed the string to an integer.

2. We dropped the following columns: month (because different countries had different harvest months, and this did not add any value to our analysis) and gross opening stock (since we didn't plan to study this).

3. We renamed columns to make them easier to read.

4. We dropped all incomplete rows since there were many NaNs in our initial data set.

5. We added a column for 'internal consumption' to show what percentage of a country's total production which was consumed domestically. We also added a column with a newly created variable to show revenue brought into country which was growers price multiplied by a country's total production. We needed to reset this revenue column to be a float instead of a string.

6. We imported an additional data set from the ICO website giving us the price for growers for production in each country. We had to choose a specific type of coffee, Robusta coffee, to get one price per country per year. To combine the two datasets, we merged on the existing dataset on country and year variable.

7. We had data for a total of 27 countries; however, some of these had data for only 3 years while others had data for 28 years. We dropped countries that had less than 20 years of data which left us with 13 countries.

8. We created variables for each country to study trends over our time horizon.

Data Analysis: We uncovered many interesting observations and insights which are detailed below.
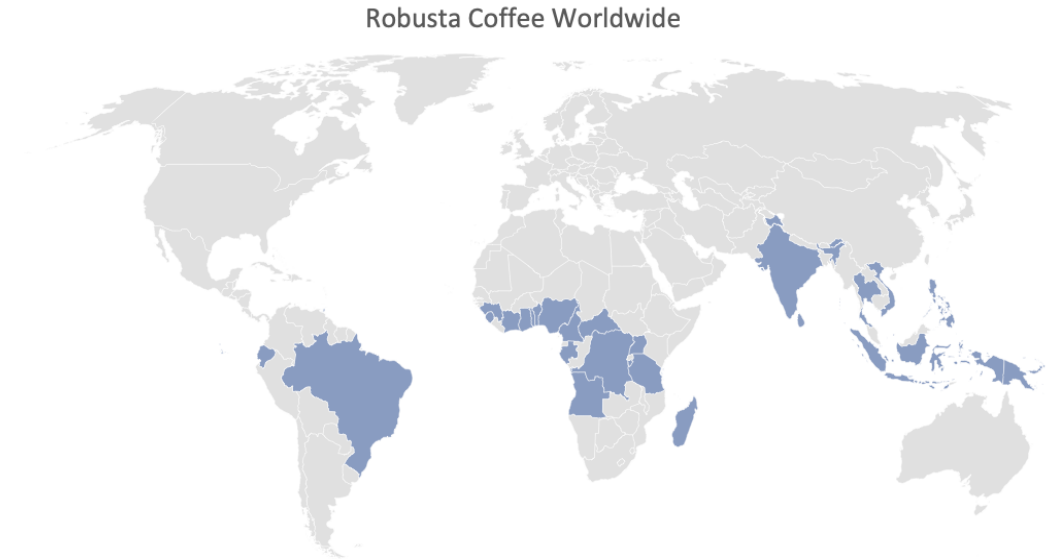
Data Bootcamp Final Project
Data Analysis on Deregulation of Coffee Market
Joanna Patsalis and Victoria McClemens

Countries

Once we cleaned the data, we wanted to take a closer look to which countries are producing Robusta coffee the most. There are only a few countries producing coffee and they all have similar climates. Because "coffee crops favor rich soil, steady temperatures, and moderate sunshine and rain, it grows only between the Tropic of Cancer and the Tropic of Capricorn."

Here, we created a map and highlighted all the coffee-producing countries:

Robusta Coffee Worldwide

Because this list is so low and we know global consumption is high, we expect this short list of countries to have very high production capabilities. From those countries we have selected only a few that we at least 20 years of growers pricing to continue our research. Those countries are: Angola, Brazil, Central African Republic, Côte d'Ivoire, Ecuador, Gabon, India, Papua New Guinea, Philippines, Thailand, Togo, Uganda, Vietnam.
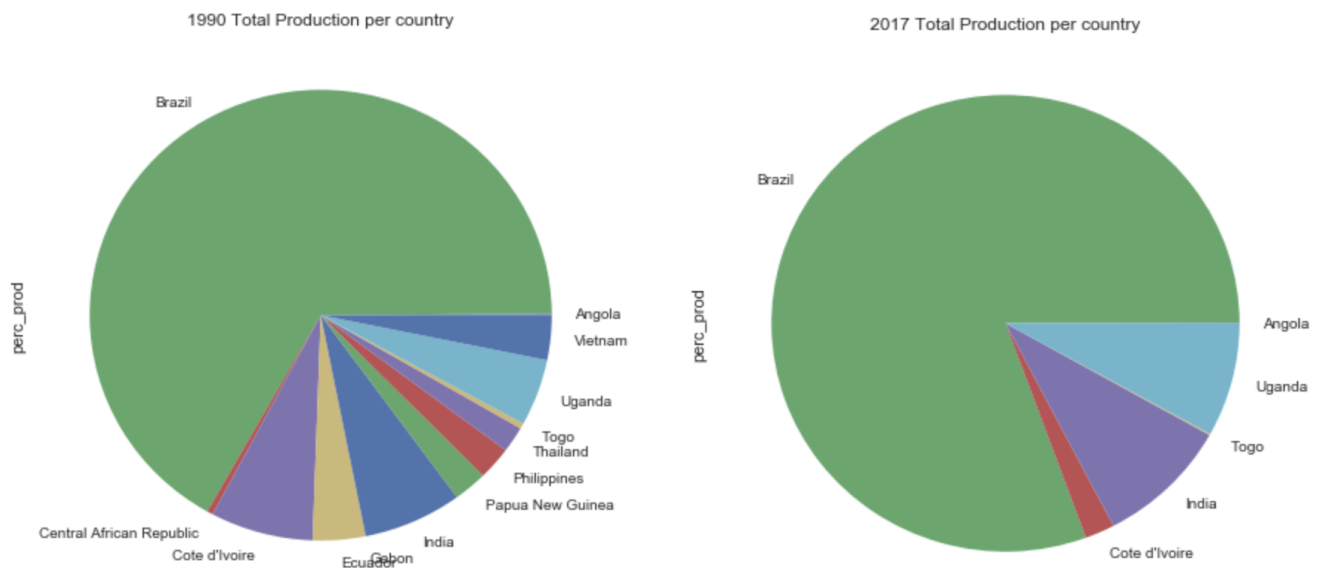
Brazil dominated overall production. From our data, it was obvious that Brazil was the largest producer of coffee, producing most of global production. Below, we created 2 pie charts - one from 1990 (the first year in our data set) and one from 2017 (the last year in our data set) - to show the global breakdown of production for Robusta coffee. In 1990, we see that Brazil is responsible for almost ⅔ of global Robusta production while remaining production is split pretty evenly among some of the other countries in the data set. In 2017, we see that Brazil makes up over ¾ of global production of Robusta coffee. In 2017, we also see far less countries in our data set producing Robusta coffee. We hypothesize that this was caused by two reasons. First, a lot of countries who were capable of growing both Robusta and Arabica coffee may have shifted production to Arabica with negligible production in Robusta coffee so they would have dropped out of our dataset. Second, many of the smaller producing countries may not have been able to compete with larger producers after deregulation and may have dropped out of the market.
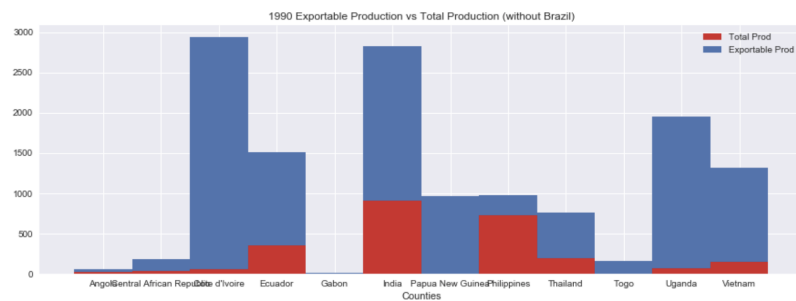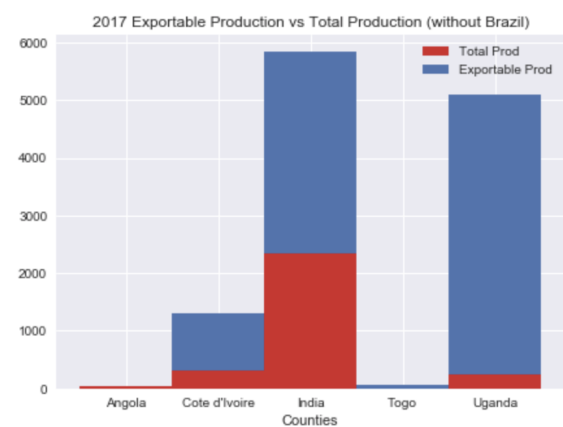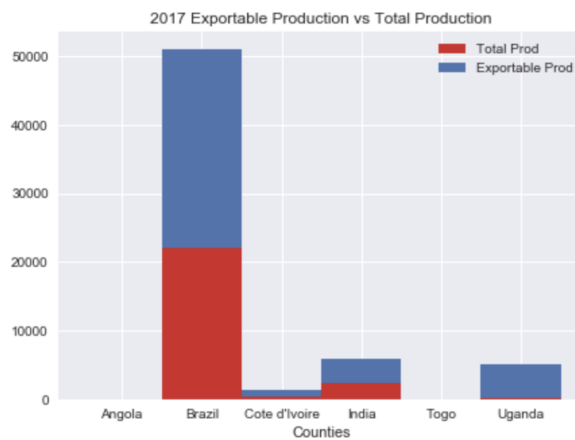
Because so few countries produce all of the coffee for global consumption, our hypothesis was that the majority of their total production would be exported. This was confirmed by our analysis. Below, we created a stacked graph showing the exportable production as part of total production.

From this, we had a few key insights. First, in looking at the charts from 1990, we notice that there are many countries producing and exporting coffee which we expected to see from the data. However, Brazil is such a dominant part of the data set, its difficult to read the data of other countries. Therefore, we removed Brazil to understand the ratio of exportable production to total production at the beginning of the deregulation period and noticed that India, Cote d'Ivoire, and Uganda, amongst others, were big exporters into the global market.

Next, we did the same analysis for 2017, the most recent year in the data set. We noticed that there were fewer countries which reinforced our initial hypothesis that the free market for coffee likely squeezed out many of the smaller producers and allowed Brazil, India, and other large producers to become even more dominant. Because there are now fewer producers satisfying the global market of consumption, the ratio of exportable coffee to total production for these countries is even higher.
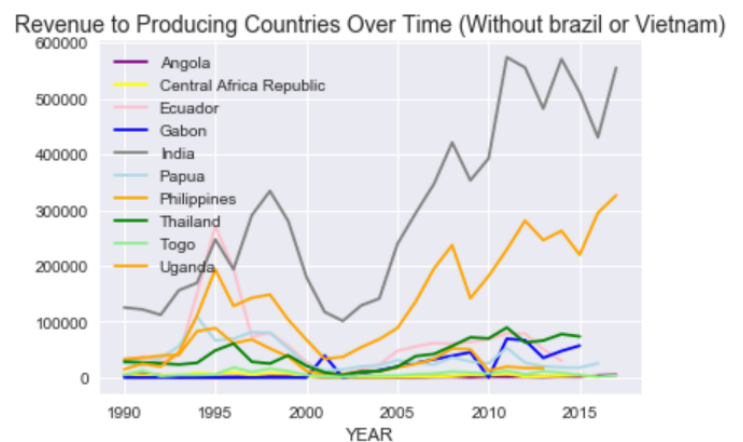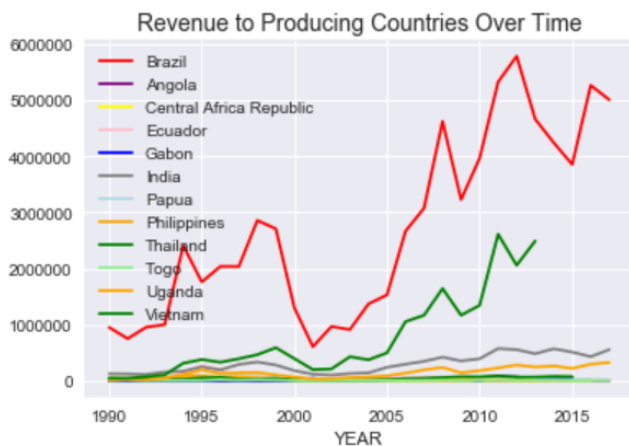
Data Bootcamp Final Project
Data Analysis on Deregulation of Coffee Market
Joanna Patsalis and Victoria McClemens

Revenue to countries

      Our hypothesis regarding revenue was that the free market would enable coffee producers to bring more revenue into their economies. We were particularly interested in revenue because this is a dynamic function of both price and production which both seem quite volatile based on our data set. Our graphs confirm that, since deregulation in the industry, countries have made more revenue from coffee. In particular, we see that Vietnam and Brazil have benefitted the most from deregulation with significantly more revenue.


      The graphs above show the revenue flow (growers' price multiplied by total production of the year) for each country throughout the years. To get a better understanding of any fluctuations in revenue among other countries, we removed Vietnam and Brazil to create a separate chart which also shows an increase in revenue where India and Philippines stand out. To understand what is driving the increase in revenue we want to further analyze any trends in growers' prices and total production separately.
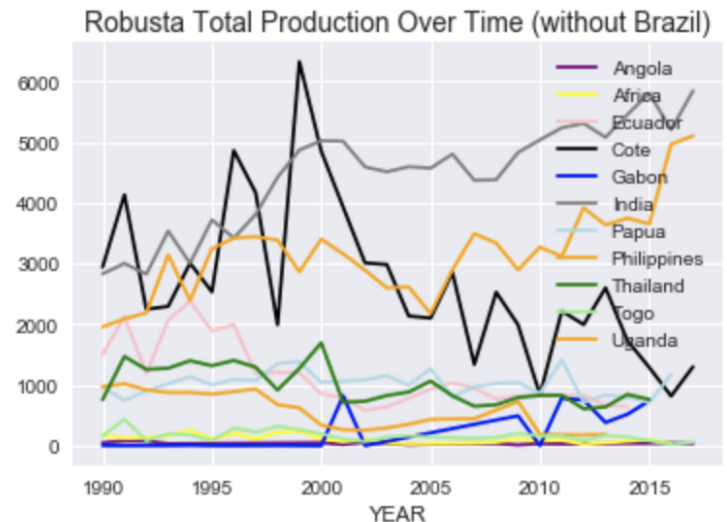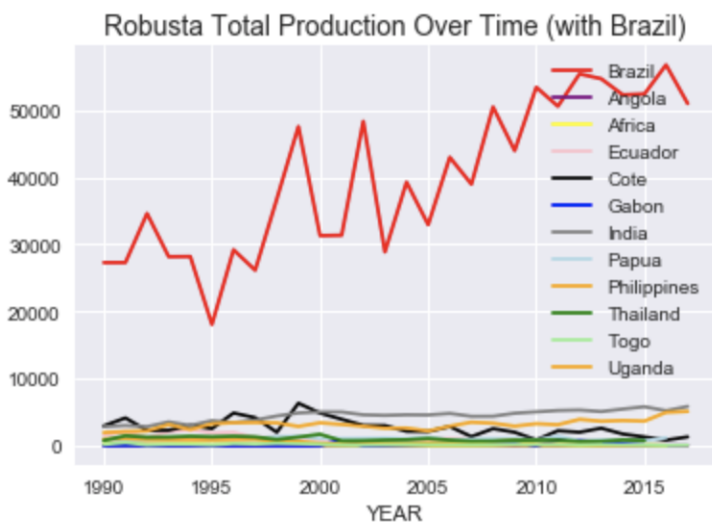
Histogram of Price Frequencies
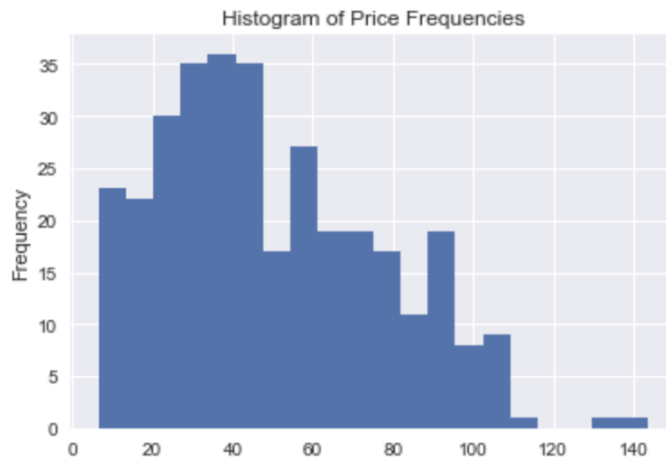
For different countries we saw that there was a range on growers' prices for different countries and wanted to see if there were any insights with regards to distribution. Below we created a histogram that shows us the wide range in prices of coffee across our data with a low of $6.44 to $143.62 with $27.07 as standard deviation and $49.75 as a mean. We believe this is because some countries are poorer and have less regulations than others. Furthermore, many market forces - like overproduction in the late 1990s (discussed in subsequent sections) and changes in countries trade agreements - we see that there are a few outliers that likely are not indicative of an equilibrium price.

Production trends

As hypothesized, the following charts confirm that production has gone up over the past 20 years with some ebbs and flows worth noting. Brazil is again standing out from the rest of the countries. We took a deeper dive by removing Brazil's production from the second chart to understand the fluctuations in production for other countries.

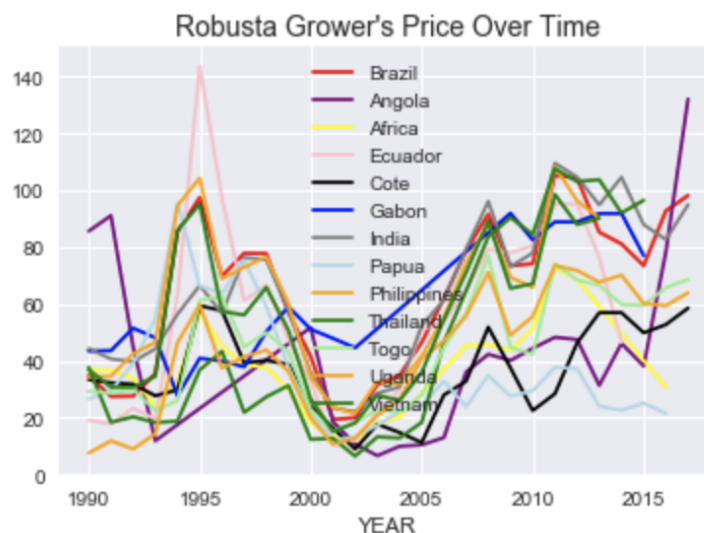Robusta Total Production Over Time (with Brazil)

Robusta Total Production Over Time (without Brazil)

We noticed a spike in the late 1990s, followed by a crash around 2000-2004, and mixed results by country after 2004. We believe this is because in the 1990s, producers were taking advantage of the free market. Our charts confirmed our earlier hypothesis that deregulation may have favored larger players who could produce as much as possible by trying to exploit economies of scale and by becoming more efficient with production, logistics, and exporting. This is evidenced by Brazil's rapid increase in production.

We did some further research and learned that producers got greedy with the introduction to the free market and overproduced coffee in the late-90s and early-2000s (evidenced by the spike in graph).[5] This led to the crash in both production and prices around 2004 as exporting countries were likely backfilled with the amount of coffee they had produced and couldn't sell into the saturated consumption market.[6] Another possible reason for the crash in 2004 may have been Hurricane Katrina which destroyed the Port of New Orleans. The Port of New Orleans is one of the largest importers of coffee in the world and take hundreds of thousands of tons from Brazil every year.[7] It is the port to one of the most dominant coffee markets in the world and was destroyed in 2005 which can further explain the fluctuations in production.

Growers' Price Trends

To understand the drivers of revenue per country we created a graph to see any changes in growers' prices over time. Our hypothesis was that the price of Robusta would have remained

Robusta Grower's Price Over Time

---

[5] https://www.theguardian.com/business/2004/apr/04/theobserver.observerbusiness11
[6] https://www.theguardian.com/business/2004/apr/04/theobserver.observerbusiness11
[7] https://www.crt.state.la.us/louisiana-state-museum/online-exhibits/coffee-trade-and-port-of-new-orleans/new-orleans-and-coffee/

the same in the late 1990s (since increased demand was being met by increased supply), but would've dropped off with the overproduction in the late 1990s. Interesting, it seems that there was a lot of variation in pricing in the 1990s which is likely because we did not factor in other variables such as changing preferences between robusta and arabica. However, as predicted, overall pricing dropped significantly with overproduction but has steadily increased since.

Production and Price Correlation

Our hypothesis was that there would be a negative correlation between production and price given the supply and demand economics. However, once we plotted overall price and production, we were unable to find a strong correlation. Moreover, the correlation coefficient is 0.23499 which is very low.
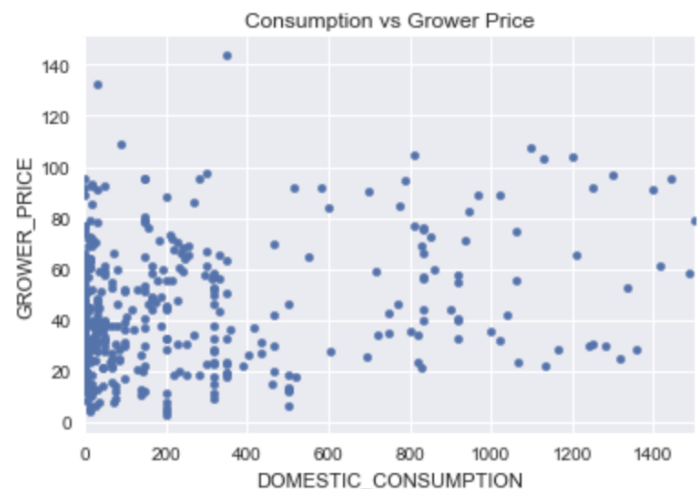
We believe the lack of correlation is due to a few factors. First, because coffee is a commoditized good with price inelasticity, the elasticity quotient should be close to zero with a steep slope on a graph showing that demand changes very little with price across all countries. Next, the price and production data points exhibited in these scatter plots covers multiple countries over many years in which we have already demonstrated that there had been volatility in price and production. Additionally, in our outside research, we learned that the price of coffee was set by future markets (i.e. the expected production in future years) rather than the current quantity in the market so there was not a strong correlation. Lastly, according to a recent WSJ article, coffee prices have been below the cost of production since the global financial crisis, leading some producers to abandon crops and some to migrate for new jobs. The shift is being

driven by currency fluctuations that are encouraging sales and production in Brazil, spurring a record crop that is driving down prices for other coffee-growing nations which may explain irregularities in the relationship between price and production.[8]

In a future analysis, we would likely try to isolate one country over a time series or to isolate one year in which we evaluate many countries to try to identify the correlation between price and production. We could also explore the machine learning techniques discussed, to determine what predictive value price and production have on one another. For example, what the percentage change in price would be with a corresponding change in production. Moreover, it will be interesting to look at what drives the growers' prices and do a regression analysis where different variables - such as exchange rates, climate factors, inflation, economy - act as the dependent variable and price is the independent variable.
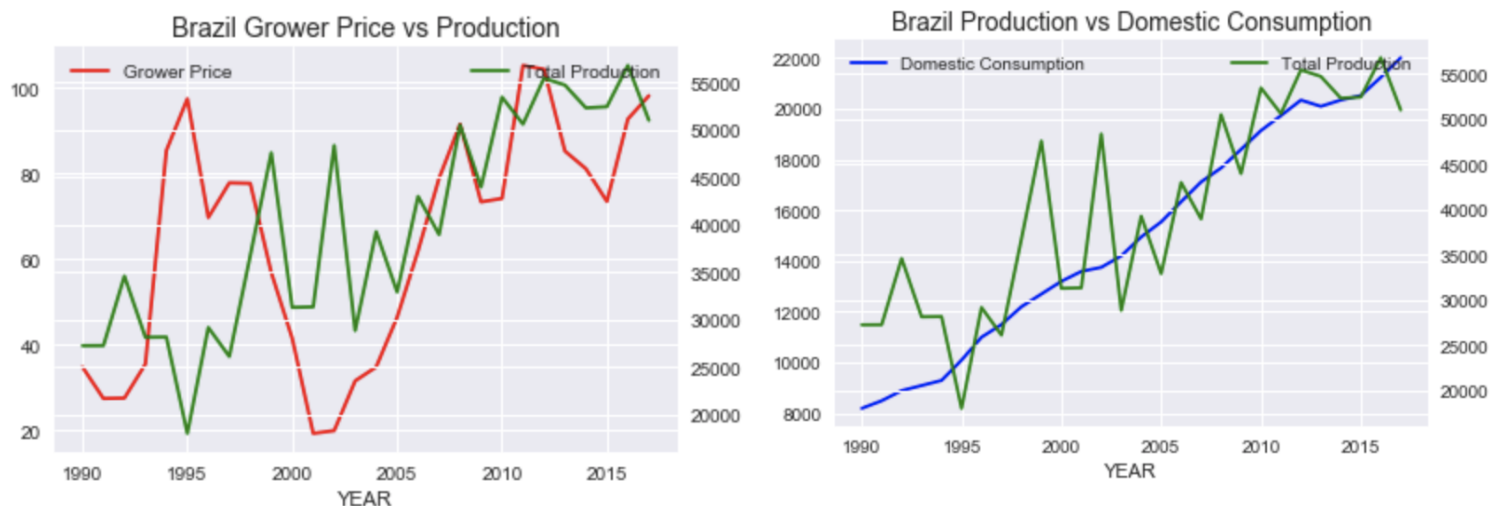
Taking a deeper dive into Brazil, which is the bigger outlier in our dataset, we saw that prices sharply increased in the 1990s and dropped to the lowest point in the early 2000s, production continued to fluctuate. It is interesting that in 1995 prices have reached a maximum and production was at the lowest level. Upon further research, we learned that this was caused by expectations of frost damage in Brazil which led traders to believe that expected demand would outstrip supply for the 1994-1995 season.[9] Those fears turned out to be unmerited and the ensuing overproduction of coffee in Brazil caused prices to drop as the market became saturated in 2000. After 2005, we see that growers prices, domestic consumption and production are moving together.

---

[8] https://www.wsj.com/articles/low-coffee-bean-prices-brew-trouble-for-farmers-11545228000?mod=hp_lead_pos9
[9] http://www.web.net/~bthomson/fairtrade/fair666.html

Conclusion and Future Work

Overall, we learned a lot about coffee production through our analysis of ICO data. Some of our biggest takeaways were:

- There is a weak correlation between growers price and production
- Brazil dominated global production for Robusta coffee
- After deregulation, the free market caused a greater disparity between larger producers and small producers
- There is a spike in production in the late 1990s and overall pricing dropped significantly with overproduction in the early 2000s, leading to a crash in 2000-2004 but has steadily increased since

With additional time, we would love to further analyze this data set. First, in a further analysis, we would like to use random forest to predict the production and prices of future years. Given that coffee prices are primarily set through the future markets of New York and London[10], we think the ability to predict prices or production would help us forecast the industry trends. Another way to predict growers price and production in the future would be to run a regression with all the variables we consider important for those countries, such as income, state of the economy, cost of production, regulation and exchange rates. We believe that climate change would also have effect on coffee production. Given the narrow range of latitudes and climates in which coffee can grow, we wanted to understand how vulnerable these areas were to climate change. Our hypothesize would be that rising temperatures and sea levels would decrease the supply of coffee in the long-run thus increasing prices. We could test this by identifying the average temperatures of the different coffee-producing regions over time and see how production correlates with variations in weather, natural disasters, etc.

We chose to study just Robusta; however, we realize that this gave us an incomplete picture of the coffee market. With more time and further analysis, we would be interested to see what the trends look like for Arabica and how they compare to the Robusta trends. Our hypothesis would be that fewer countries would be able to produce high volumes of Arabica (because it requires higher elevations to grow with a specific climate), however, developed countries - especially in the US and Europe - would have increased consumption significantly in recent history driving up price and volume of Arabica.

---

[10] https://dailycoffeenews.com/2018/10/23/coffee-futures-investor-behavior-overwhelms-market-fundamentals/

In a further analysis, we would also like to study consumption of coffee. Some of the world's largest coffee consumers - like the US and Europe - are not included in our data set because they are not producers. Given that the rise in production and prices correlates to increased global demand, we would like analyze the correlation between consumption and production.

Finally, we were interested to study the socio-economic and political impacts of coffee prices on the producing countries. In the past few years, we've read much about the inflation, unemployment, and political corruption in countries like Brazil and Venezuela. We would be interested to see how these problems have been exasperated with the deregulation of coffee in which prices have been below production costs at times.[11]

Code can be found in notebook:

https://github.com/joP77/Data_Bootcamp_Final_Project
https://github.com/vlm287/Data_Bootcamp_Final_Project

---

[11] https://www.wsj.com/articles/low-coffee-bean-prices-brew-trouble-for-farmers-11545228000?mod=hp_lead_pos9