case_3.2

April 22, 2020

1 What are spatial and temporal trends in crop production?

[2]:
```
### Load relevant packages
import pandas                   as pd
import numpy                    as np
import matplotlib.pyplot        as plt
import seaborn                  as sns
import statsmodels.formula.api as sm
import os

%matplotlib inline
plt.style.use('ggplot')
```

1.1 Introduction (5 mts)

Business Context. In 2019, the United Nations issued a report on the current state of food security, which indicated that 2 billion people are current experiencing moderate or severe food insecurity. The report also suggests and advocates for changes "to live in a world without hunger, food insecurity and malnutrition in any of its forms." [1] The report suggests investing wisely to reduce variability in food production and increase the capacity to withstand food shortages and distribution disruptions in times of economic or political turmoil.

An important aspect of food security is, of course, the food itself. Who is producing which foods, and how has that changed over time? Understanding the production, yield, and land usage of the world's most vulnerable populations can hopefully yield insights into what policy changes may help them the most.

[1] FAO, IFAD, UNICEF, WFP and WHO. 2019. *The State of Food Security and Nutrition in the World 2019. Safeguarding against economic slowdowns and downturns.* Rome, FAO.

Business Problem. You are a researcher for a think tank that is interested in proposing scientific investigations and policy solutions based on your insights to the following question: **"What are spatial and temporal trends in crop production?"**

Analytical Context. In this case, we will use crop data collected by the Food and Agriculture Organization of the United Nations, a dataset known as FAOSTAT, to identify spatial and temporal trends in food production.

In the last case, you became familiar with basic concepts in information gathering and information sufficiency. In this case, you will further explore techniques in information gathering and assessing information sufficiency. You will continue to use skills in interpreting charts, figures, and data summaries. Note that unlike the previous case, which was looking at a relationship between preidentified factors and clear outcomes of interest, here we will be doing more exploratory analysis, where the desired insights are far more ambiguous.

The case is structured as follows: (1) you will understand and focus on a particular outcome of interest (coffee); (2) you will investigate noteworthy geographical changes in the data over time; (3) you will compare production and exports across particular regions of interest; and finally (4) you will evaluate information sufficiency and provide recommendations for further investigation.

1.2 Understand and narrow an outcome of interest (20 mts)

The question of interest:

> What are spatial and temporal trends in crop production?

is ambiguous and broad. This is because the goal is clear (improve worldwide food security), but the variables to investigate, outcomes to measure, and questions to ask are not. In the previous case, we started with a clear question and we needed to gather and assess the necessary data. Here, we will start with the FAOSTAT dataset and explore potential insights we could glean from it.

An important first step when presented with a problem like this is to narrow the scope of the problem and focus the investigation on one area of the problem before moving on to the next. It is often useful when gathering information to **explore what is already known about the topic at hand in order to find sub-problems or hypotheses which are worth further analysis**. Specifically, here we want to look at recent findings surrounding spatial and temporal trends in crop production, then spot any unexplained phenomena or gaps in current understanding to pinpoint a potentially interesting hypothesis.

Read each statement and then answer the corresponding question below.

1.2.1 Exercise 1:

1.1 One major question in reducing food insecurity: is the issue mostly in production, or in distribution? In other words, do we produce enough food to feed all the mouths, or do we simply not get it into the right hands fast enough? In July 2012, a research paper was published with the provocative title:

> "We Already Grow Enough Food for 10 Billion People... and Still Can't End Hunger." [2]

Which of the following data points are *necessary and sufficient* to make the statement above? Multiple selections may be required.

(a) The amount of food grown in the world;

(b) The amount of food thrown away in the world;

(c) The amount of food a person needs to be food secure;

(d) Additional information is required.

Answer. The answer is (a), (c), and (d). The validity of the claim that "We Already Grow Enough Food for 10 Billion People... and Still Can't End Hunger" rests on three key numbers: 1. the amount of food we produce; 2. the amount of food each person needs to be secure; 3. the number of people in the world (in particular, the existence of hungry people).

The first two numbers let you conclude that we already grow *enough* food for ten billion people, and the third number lets you make the second part of the statement: "still can't end hunger."

Notice that if the claim in the first question is true, then optimizing food distribution vs. increasing food production may be what is required to achieve international food security. Next, we investigate what is known about food distribution.

1.2 In November 2017, a research article investigated the FAOSTAT dataset to understand the international food trade network. In particular, they investigated what is called the "community structure," which formalizes which groups of countries or regions trade more frequently with each other.

> "Our estimates indicate that the probability [that two countries belong] to the same food trade community depends more on geopolitical and economic factors – such as geographical proximity and trade agreements co-membership – than on country economic size and/or income."

Which of the following pieces of evidence, if true, would make you *skeptical* of the statement above? Multiple selections may be required.

(a) A country's economic size is highly related to its "geopolitical and economic factors."

(b) A country's economic size is highly variable over time.

(c) A country's economic size is highly related to the food trade community it is in.

(d) None of the above.

Answer. (a) only, because if a country's economic size is highly related to the geopolitical and economic factors that were measured, then it is not clear that you can conclude that food trade community membership depends *more* on "geopolitical and economic factors" than "country economic size and/or income."

(b) is incorrect because no part of the statement makes a claim related to change over time.

(c) is incorrect because if you know that economic size is highly related to food trade community membership, then

---

The previous question highlights the importance and complexity of food trade network communities. In the next question, we zoom in on a potential trade opportunity that could significantly reduce food insecurity.

1.3 Killian Stokes is an adjunct lecturer on Business and Global Development at the Quinn School of Business in UCD and the co-founder of Moyee Coffee Ireland, the world's first FairChain coffee. He noticed that Ethiopia is one of the countries in the world with the highest number of food insecure people, despite a thriving agricultural sector and high rates of crop production. In particular, it exports some of the world's finest coffee, which has led him to conclude:

> If Ethiopia started to develop its coffee industry, it could trade its way out of poverty.

Which of the following pieces of evidence, if true, would be *sufficiently convincing* to conclude the statement above?

(a) Ethiopia has not developed its coffee industry and is still very impoverished.

(b) Another country very geopolitically and economically similar to Ethiopia developed its coffee export industry in the last few decades and is now far more food secure.

(c) When Ethiopia's coffee export business was more developed in the past, it was less impoverished.

(d) Every country that has developed its coffee industry has been able to trade its way out of poverty.

Answer. This is a tricky question, but (b) is the most convincing. (a) is not sufficiently convincing because it doesn't make any claims about the state of Ethiopia in a world where its coffee industry is developed, which is what we would need. (d) is not sufficiently convincing because it is possible that all of the other countries that developed their coffee industries could be geopolitically and economically extremely different from Ethiopia, so much to the point that this outcome would not necessarily be expected for Ethiopia. (c) is a close second, as it makes a claim about the state of Ethiopia in a world where its coffee industry is developed, but the country may have changed in other ways in the time that has passed. This leaves (b), which is most convincing because it references an economic, geopolitical, and temporal analogue to Ethiopia that has demonstrated its success.

You may not realize it, but in the above exercises you began to see examples of **confounding variables**, **correlation vs. causation**, and **probability**. You will learn more about these concepts more formally in future cases.

[2] Holt-Giménez, Eric & Shattuck, Annie & Altieri, Miguel & Herren, Hans & Gliessman, Steve. (2012). *We Already Grow Enough Food for 10 Billion People … and Still Can't End Hunger.* Journal of Sustainable Agriculture - J SUSTAINABLE AGR. 36. 595-598.

[3] Torreggiani, Sofia Torreggiani & Mangioni, Giuseppe & Puma, Michael J. Puma & Fagiolo, Giorgio. (2017). *Identifying the community structure of the international food-trade multi network.* arXiv:1711.05784.

1.3 Investigate geographical changes in coffee production over time (30 mts)

Recall the third claim in the previous section: *If Ethiopia started to develop its coffee industry, it could trade its way out of poverty.* This seems like a hypothesis where knowing certain pieces of info would likely validate or invalidate it, as opposed to our original question of interest, which was

purely exploratory and could not be deemed to be right or wrong in any way. In the remainder of this case, we will investigate the potential validity of this hypothesis by looking at spatial and temporal trends in coffee production.

Consider the following plots from the FAOSTAT dataset. The first set captures coffee production trends in the decade 1961-1971, and the second set captures coffee production trends in the recent decade 2007-2017. The 1960s is when much of the earliest data in FAOSTAT becomes reliably available. Additionally, several trade policies and international negotiations took place in that decade to establish trade agreements that are still in place today. Make note of the relevant time frames, axes, titles, and legends to understand what each graph is depicting.

1.3.1 Coffee production 1961-1971:

1.3.2 Coffee production 2007-2017:

1.3.3 Question:

Make sure you can answer the following warm-up questions about the graphs above:

1. Why are some of the countries in the world maps white?
2. Why is the crop called "Coffee, green"?
3. What is the difference between "Ethiopia PDR" and "Ethiopia"?

Here are the answers to the warm-up questions above:

1. These are countries where data is not collected or reported to FAOSTAT.
2. "Coffee, green" refers to raw coffee, as opposed to coffee that has been roasted. Most exported coffee is not roasted.
3. This was a socialist government in place in Ethiopia for some time prior.

Now let's proceed:

1.3.4 Exercise 2:

Notice the shift in share of coffee production from the 1960s to 2010s for Asia: 6.7% to 30.2%. Based on the plots above, what is likely the MOST significant driver of this?

(a) Brazil stagnating on its coffee production, which let Asian countries take a larger share of the pie.

(b) Vietnam developing its coffee production industry in a major way.

(c) Angola leaving the coffee production stage, which let Asian countries take a larger share of the pie.

(d) Cannot determine from the plots provided.

Answer. (b), because Vietnam's coffee industry exploded in those decades. The answer is not (a), since the plots do not suggest Brazil stagnated in its coffee production, since its production is significantly higher in the 2010s than in the 1960s. The answer is not (c), since while Angola did

leave the top 10 coffee production stage, the plots do not suggest it was a huge player. The answer is not (d), since the plots do strongly suggest that Vietnam had a significant enough effect to cause the shift to Asia.

1.3.5 Exercise 3:

Based on the results of Exercise 2 and the hypothesis we are investigating, come up with a sensible way to proceed.

Answer. From Exercise 2, we see that both Brazil and Vietnam significantly increased increased their coffee production within the last few decades. Furthermore, from Exercise 1.3, we know that finding an analogue for Ethiopia - that is, a country that was geopolitically and economically very similar to Ethiopia within the past few decades - and looking at how the coffee industry impacted their development is a good strategy. Hence, it makes sense to determine whether Brazil and/or Vietnam may be suitable analogues for Ethiopia if it develops its coffee industry.

1.4 Compare coffee production in Ethiopia vs. Brazil (20 mts)

We can see from the previous section that many of the main players in the coffee production stage have shifted significantly since the 1960s, when many trade agreements began to be established. In this task, we will look at temporal trends in coffee harvest, production, and yield for Ethiopia (our country of interest) and Brazil, the largest coffee producer. Understanding the characteristics of Brazil's coffee production may yield insights into how Ethiopia can grow its coffee industry (and potentially "trade its way out of poverty").

1.4.1 World:

1.4.2 Ethiopia vs. Brazil:

1.4.3 Exercise 4:

Based on these plots alone, is Brazil's coffee production industry in the first two decades (1960-1980) comparable to Ethiopia's coffee production industry in the last two decades (1990-2010)? Why or why not?

Answer. No, Brazil's coffee production industry is not a good analogue for Ethiopia's in the last decade. From the charts, we can see that coffee yield has been steadily increasing since 1990, while Ethiopia's yield has remained roughly constant. Furthermore, Brazil's area harvested was decreasing from 1960-1980, while Ethiopia's was increasing; This means that Brazil has concentrated on reducing the land area dedicated to coffee production while investing in other methods to increase yield. However, Ethiopia has a much smaller and less diverse geographical area than Brazil (which is massive) and it is unclear if their soil exhibits the same characteristics as in Brazil which allowed for that type of yield improvement.

1.5 Compare coffee production in Ethiopia vs. Vietnam (20 mts)

Now let's look at the graphs below comparing the coffee area usage, production amount, and yield for Ethiopia and Vietnam:

1.5.1 Exercise 5:

Based on these plots alone, is Vietnam's coffee production industry in the first two decades (1960-1980) comparable to Ethiopia's coffee production industry in the last two decades (1990-2010)? Why or why not?

Answer. Yes. Vietnam harvests similar amounts of land for coffee as Ethiopia, and Ethiopia's coffefe yield in 1990-2010 is comparable to Vietnam's coffee yield in 1960-1980. Furthermore, Vietnam massively expanded its dedicated land for coffee production and was able to increase both its yield as well as its total production through this process. Given that Ethiopia has slightly expanded its land in recent years, this suggests that Ethiopia could benefit from a more effort from increasing its land usage dedicated to coffee, much like Vietnam.

We have seen that while Brazil's and Ethiopia's coffee industries may not be comparable, it appears that Ethiopia's and Vietnam's are similar in their initial starting conditions. However, they are different in their subsequent actions, which gives insight into how Ethiopia can change its actions to expand its coffee industry. Specifically, it can look to increase its yield (more investigation is required to determine how Vietnam did this), and also devote more land to crops (in particular, to coffee).

Given that Ethiopia and Vietnam had similar starting conditions, and Vietnam was able to successfully trade its way out of poverty by developing its coffee industry, this suggests that Ethiopia could do the same.

1.6 Compare coffee exports in Ethiopia vs. Vietnam (15 mts)

Now that we've determined that Vietnam could be a suitable analogue for Ethiopia, it makes sense to look at where Vietnam's produced coffee was consumed to gather insight into how Ethiopia can distribute future coffee production. Consider the following two plots depicting the coffee export value and quantity over time for Ethiopia and Vietnam, then answer the questions below.

1.6.1 Exercise 6:

6.1 Based on these plots alone, what are some differences between coffee exports in Vietnam vs. Ethiopia? Select all that apply.

(a) Vietnam's coffee export value has been increasing since 1990, while Ethiopia's coffee export value has remained roughly constant.

(b) Vietnam's coffee export quantity has been increasing since 1990, while Ethiopia's coffee export quantity has remained roughly constant.

(c) Both Vietnam's and Ethiopia's coffee export quantity have been increasing since 1990, but Vietnam's coffee export quantity has been increasing at a much higher rate.

(d) There are no differences, Vietnam and Ethiopia have entirely comparable coffee exports.

Answer. (c)

6.2 How does this information change (or not) your conclusions from question 4.4?

(a) This strengthens the conclusion that there is room for growth in Ethiopia's coffee industry by producing and exporting more coffee.

(b) This does not change the conclusion that Ethiopia could economically benefit from devoting more land to coffee production.

(c) This weakens the conclusion that Ethiopia could economically benefit from devoting more land to coffee production.

(d) The information is not related to the conclusions made in 4.4.

Answer. (a)

1.7 Summarize conclusions and assess information sufficiency (20 mts)

Now, let's pull together everything we've gathered so far to see what we can conclude and what requires more information and analysis to make claims about. In addition to the analysis we did above, let's consider the following piece of information:

> *While 100% of coffee is grown in the coffee belt, 99.9% of all coffee we drink is roasted in Europe or America. Coffee is exported out of the coffee belt as raw green bean and so... most of the jobs, income and profits from coffee are exported out of the coffee belt."*

More specifically, here is the list of the world's 20 biggest coffee drinkers (kilogram per capita per year): [4]

Country	annual kg/capita
Finland	12
Norway	9.9
Iceland	9
Denmark	8.7
Netherlands	8.4
Sweden	8.2
Switzerland	7.9
Belgium	6.8
Luxembourg	6.5
Canada	6.2
Bosnia and Herzegovina	6.1
Austria	5.9
Italy	5.8
Slovenia	5.8

Country	annual kg/capita
Brazil	5.5
Germany	5.5
Greece	5.4
France	5.1
Croatia	4.9
Cyprus	4.8

1.7.1 Exercise 7:

7.1 For each of the following statements, state whether you have:

(a) `Sufficient information to support`

(b) `Initial but not sufficient information to support`

(c) `No information to support`

(d) `Initial but not sufficient information to refute`

(e) `Sufficient information to refute`

based on all of the plots and tables you have investigated in this case.

1. Ethiopia has substantial room for growth in its coffee industry.
2. Vietnam has maxed out the size of its coffee industry.
3. Brazil's coffee production will remain a world leader for another century.
4. Ethiopia should fix its trade relations with Germany, a major coffee importer interested in Ethiopian beans.

Answer.

1. (b)
2. (c)
3. (d)
4. (c)

1.7.2 Question:

For each statement in the previous exercise where you determined you did not have sufficient evidence, can you think of additional studies and datasets you would want to analyze to be confident in your conclusions?

7.2 Consider the following policy recommendation for Ethiopia:

> Our analysis has determined that Ethiopia has substantial room for growth in its coffee industry and could increase its export quantity for coffee by devoting more of its land to crop production, specifically coffee.

Which of the following datasets or studies would be MOST important for you to study next in order to gain more confidence in this recommendation?

(a) A simulation of Ethiopia's coffee production quantity over the next two decades based on models for increases in land usage and coffee yield.

(b) A table of coffee imports over time for each country, in order to identify which countries would be the best targets for Ethiopia to export to.

(c) A simulation of other top coffee producers' coffee quantities produced over the next two decades, in order to assess how crowded the coffee production market will get.

(d) All of the above would be useful and important datasets to have.

Answer. (d)

Think about *why* you think this information would be so important and which issues and uncertainties in the conclusion they would address.

[4] Oliver Smith. October 1, 2017. Countries that drink the most coffee. *The Telegraph*. https://www.telegraph.co.uk/travel/maps-and-graphics/countries-that-drink-the-most-coffee/

1.8 Conclusions (5 mts)

In this case, we sought to understand crop production and yield patterns over seasons, years, and different geographies in order to gain insight into what can improve resilience to food shortages and ultimately international food security. In particular, we focused on how to improve poverty in Ethiopia through trade.

In particular, we saw that Brazil's and Ethiopia's coffee industries are likely not very comparable. However, the similarities and differences between Ethiopia's and Vietnam's are promising. The plots tentatively suggest that Ethiopia could develop its coffee industry by increasing the yield on its land, though more investigation is required to determine how Vietnam achieved such high yields (e.g. importing high-quality pesticides?), as well as by devoting more land to crops.

1.9 Takeaways (5 mts)

Here, you have further practiced your skills in reading charts and plots, gathering information, and assessing information sufficiency. You also saw that problems and questions can often be posed to you in vague and unanswerable ways, and that it is your job to narrow that down to a much more tractable hypothesis. You saw that a very effective way to do this was to read up on recent developments in the domain, in order to find something very specific with a more clear "right/wrong" answer. This again highlights a key point, which will come up again and again in this course: domain knowledge and/or expertise is essential to every part of the data science & analytics process.

The way that you resolved the hypothesis at hand here was generally qualitative - through interpreting charts & graphs, and using logical thinking skills. In future cases, you will learn how to validate/invalidate hypotheses that you generate via more quantitative and technical methods.