Quantitative Framework for Coffee Rust, Production and Futures

Laura H. Kahn
School of Informatics and Computing
Indiana University
Bloomington, IN USA
lkahn@indiana.edu

***Abstract*—More than two billion cups of coffee are consumed worldwide each day [14]. The livelihood of 120 million people depends on the coffee supply chain [33]. Coffee rust leads to production losses of over $500 million worldwide and may affect futures prices [15]. Coffee rust is caused by the coffee berry borer, or *Hemileia vastatrix* fungus, at temperatures from 10-30 C, and is one of the main diseases that attacks the coffee arábica plant [19]. Coffee is the second largest traded commodity worldwide, with about $100 billion in volume traded annually [32, 4]. Understanding the relationship between coffee rust, production quantities and futures prices is important to anyone affected by the coffee supply chain. This research offers a more quantitative framework for describing and visualizing the relationship between coffee rust, amount of coffee produced and futures prices.**

Keywords—econometrics; hemileia vastatrix; pricing; commodity futures; coffee trading; statistical analysis

I. Introduction

The *arábica* variety of coffee accounts for 75% of total bean production worldwide [14]. In 2011, coffee had a $90 billion retail value in the U.S. alone in 2011 [20]. In 2014, coffee in Brazil accounted for 6.9% of exports and generated revenues of U.S. $6.66 billion. Brazil, Vietnam, Colombia, Indonesia and Ethiopia were the top five coffee producing countries in 2017 [14]. Brazil, Colombia and Papua New Guinea produce 48% of the world's *arábica* supply. Vietnam and Indonesia produce coffee of the *robusta* variety, which is outside the scope of this research.

Coffee rust is caused by the coffee berry borer, or *Hemileia vastatrix* fungus at temperatures from 10-30 C, and is one of the main diseases of coffee arábica worldwide [19]. Coffee rust varies in severity with levels from 1-6 [10]. The higher the percentage of rust that covers a leaf, the higher the percentage of rust. Understanding the relationship between coffee rust, production quantities and futures prices is important to anyone affected by the coffee supply chain and commodities market.

Coffee futures are standardized, exchange-traded contracts in which the contract buyer agrees to take delivery, from the seller, a specific quantity of coffee at a predetermined price on a future delivery date. Coffee futures are traded on average 252 days on the New York Stock Exchange (NYSE) from 0930-1330 daily [6].

The research will attempt to establish a quantitative framework for the relationship between coffee rust, production and futures prices. The research will also attempt to examine if past data of rust-infected coffee plants and futures can contribute to this quantitative framework.

There has been no known research to date that has shown a mathematical connection between historical coffee rust occurrence rates in multiple coffee-producing countries and worldwide coffee futures prices. There is no known research to date that explores the relationship between past coffee futures prices and coffee rust variables and quantifies the relationship.

Since coffee is the second most popular commodity traded by volume, having the ability to understand how these prices are related to rust and / or production will benefit those businesses that invest in coffee futures [32].This research could be used to further research to increase profits for anyone trading coffee futures.

The *hypotheses* are that*:*

a. More rain increases rust.
b. Higher temperatures increase rust.
c. More rust decreases production.
d. More production decreases futures.
e. More rust decreases futures.

A. Related Work

Yap explores the effect of precipitation, exchange rate and import volume on Brazilian coffee futures prices [36]. Yap uses simple line plots to compare the variables. Magrach visualizes how climate change impacts where coffee is grown using a choropleth map of *robusta* and *arábica* coffee varieties worldwide [25]. Avelino (2006) discusses how rust is affected by minimum, maximum temperature, average rainfall and whether the plant is shaded [2]. Avelino creates a visual decision tool to determine the level of rust. Avelino (2015) also suggests a possible link between coffee rust and prices in general but does not discuss futures prices [3]. Avelino uses a heat map of Central America to visualize precipitation anomalies and a bar plot to compare effect of weather properties on rust.

Lamouroux suggests soil pH, soil structure and temperature range between coffee plots influence the amount of coffee rust [23]. Lamouroux visualizes the qualitative relationship between variables with three-dimensional bar plots. Montague uses artificial intelligence machine learning models to predict 27 futures prices based on daily historical prices data but does not include any visualizations [26]. Kim uses artificial neural networks to predict stock prices for several commodities [22]. Since no electronic copy of Kim's research is available, any visualizations cannot be discussed. Luaces uses regression and classification to predict coffee fruit load. Luaces visualizes the model effectiveness with line plots [24]. Cintra uses decision trees for coffee rust warning in Brasil from 1998-2016 [5]. Cintra visualizes the nodes and decision pathways for the selected decision tree.

As previously mentioned, none of the existing research visualizes the relationship between coffee rust, production and futures.

II. DATA

A. Acquisition

The final data set includes 337 observation of five variables: average monthly *Rain* (mm), average monthly *Temperature* (◦C), Rust (percent disease that covers the plant), *Production* (measured in 1000-60 kg bags of coffee beans) and *Futures* (in US dollars). Data from Brasil was from 2005-2006 and 2008-2009. Data from Colombia was from 1995 and 2011-2013. Data from Papua New Guinea was from 1989-1991.

As with most data science projects, the analysis was limited by data that is open (publically available at no cost). There is no freely available rust data from some of the other top coffee producing nations such as Vietnam or Indonesia. Since Vietnam and Indonesia do not produce the arábica variety of coffee, the lack of data from these countries is irrelevant to the scope of this research. Since Brazil and Colombia are in the top five coffee producers worldwide and they along with Papua New Guinea produce 48% of the arábica coffee, the data is considered a reasonable sample of all coffee-producing regions.

1. Challenges

Challenges to getting the data included lacking domain expertise, an initial lack of historical futures data, a language barrier and combining data from multiple sources.

Gathering enough accurate data with all the necessary variables was challenging at the beginning of the project and took a significant amount of time. Existing literature from experts with coffee production and rust data was relied upon since the author did not have this domain knowledge. A learning curve in knowing how to search and find the necessary data was part of the data acquisition process.

For the coffee futures price data, at the initial phase of the project proposal printed historical charts scanned at low resolution were the only available data [17]. Since no more precise data was found, initially manual approximations were made from the scanned images for prices on the first day of each month rounded to the nearest decimal point.

Fortunately after the initial project proposal deadline, more accurate futures price data from the US Commodities Futures Trading Commission was located [9]. This resource contained much more accurate data digitally available to three decimal points from 1986-2012. These amounts were reported more frequently at four to five times each month (except in 1989-1991 where there were only 2 reports per month). These amounts were used as the authoritative data source for futures.

All documents regarding coffee rust occurrences in Brazil were in Portuguese [11, 12, 18, 20, 27, 36] so Google Translate was used as the best approximation to translate the relevant figures and data. Supplemental rust data from one of the more prolific coffee rust researchers [10], was requested and received in Spanish. Having additional knowledge in Spanish helped provide the necessary translation for this data.

On average, each country had five different data sources that had to be manipulated and cleaned before the analysis and visualization. Units of measurement were verified for each variable to ensure they were consistent for each country.

2. Data Sources

For Brasil, the rust data from June 2008-December 2009 is from Table 1 in [20]. The rust data from September 2005-August 2006 is obtained from Figure 4 in [18].

For Papua New Guinea from 1989-1991, the *Rust* data was obtained by calculating the average of Onaningka, Kayokite and Bena coffee growing areas by month shown in Figure 1 [23].

For Colombia, the *Rust* data from January – December 1995 was calculated by subtracting the percent alive leaves from 100. The percent alive leaves was calculated by taking an average of alive leaves in Supia, La Catalina and Narnjal 60 days after flowering in Figure 1 [28]. The same percent rust value was used for all months in 1995. The *Rust* data from August 2011 – March 2013 was calculated using an average of 16 samples [10].

For Brasil, Papua New Guinea and Colombia, the average monthly *Rainfall* and *Temperature* were obtained from the World Bank [33]. *Production* data was obtained from the 'Paises Productores' tab of [15]. *Futures* data at the NYSE "C" Contract market was obtained from [9].

3. Data Assumptions

In order to understand the scope of the research, it is important to understand the assumptions and limitations of the data. Latitude and longitude for Supia, Colombia is 5.4559N, -75.6504W [10] and Cauca, Colombia is 3.3599N, -76.6386W. A negative value represents South and West.

Rainfall, *Temperature*, *Rust* and *Production* values were used for each date within the same month. *Rainfall*, *Temperature* and *Rust* are reported monthly. *Production* is only reported annually so monthly amounts were calculated by dividing the yearly amount by 12. The only data that is available bi-weekly is the *Futures* data.

Coffee futures are reported worldwide not by country on the NYSE.

Altitudes are similar for Brasil (1010 m), Columbia (1310 m and 1761 m) and Papua New Guinea (1410-1770 m) as are temperatures so using data from these three coffee-growing regions is acceptable since growing conditions are similar. The variables formed a correlative model where environmental variables such as *Temperature* and *Rainfall* were initially assumed to influence coffee *Production* and *Rust* [24].

The impact of climate change on coffee growing regions has been studied by [21] and [1]. However, any possible implications of climate change on the *Rust*, *Production* or *Futures* was not included in the analysis.

All *Rust*, *Production* and *Futures* data available publically was assumed to be accurate. Since the data is open on the web and not under a license or proprietary in any way, the data is not considered private. The data also does not contain any personal identifiable information so the reuse of the data for the research does not violate any ethical principles.

4. Variable Selection

Weather and physical crop properties that affect rust were included to make the simplest and easiest to interpret model possible. The literature suggests that two main physical crop variables determine coffee rust - air temperature and rainfall amounts [17, 35, 4]. The literature also suggests that farmers apply rust control measures such as pesticides when crop leaves have 5-20% brown spots [12]. Certain crop management variables such as fertilization and pesticide use were not included to avoid introducing confounding variables into the analysis. Furthermore, it was assumed that adding these crop management variables would not add a significant amount of value to quantifying the relationships between variables and the hypotheses. Crop management practices are known methods to control coffee rust after it reaches a 5% threshold [11].

Other attributes such as soil humidity, pH and day length may be significant variables that affect rust but were not included in the analysis since there was no consistent and reliable source data for Brasil, Colombia and Papua New Guinea [29, 23].

5, Pre-processing

Data cleansing is an important pre-processing step of the data acquisition process. Minimal processing was sufficient. The 'Date' variable was standardized as MM/DD/YY format. The remaining numerical variables were standardized to two decimal places. The *Temperature* variable was renamed to 'Temp'.

Below is a screenshot of the first few rows of pre-processed data in CSV format.

	Date	Country	Rain	Temp	Rust	Production	Futures
0	9/6/2005	Brasil	71.45	25.28	8.50	2783.33	84.49
1	9/13/2005	Brasil	71.45	25.28	8.50	2783.33	84.54
2	9/20/2005	Brasil	71.45	25.28	8.50	2783.33	85.02
3	9/27/2005	Brasil	71.45	25.28	8.50	2783.33	86.51
4	10/4/2005	Brasil	117.67	26.82	0.50	2783.33	87.37
5	10/11/2005	Brasil	117.67	26.82	0.50	2783.33	87.86
6	10/18/2005	Brasil	117.67	26.82	0.50	2783.33	85.87
7	10/25/2005	Brasil	117.67	26.82	0.50	2783.33	84.81

Fig. 1. Pre-processed Data

B. Exploratory Analysis

Exploratory data analysis begins by looking at summary statistics of the data. Average monthly *Rain* ranges from 0.2 - 407.7 mm per month, average monthly *Temperature* ranges from 23.36-27.16° C, average monthly *Rust* ranges from 0.33-50%, monthly calculated *Production* ranges from 80.33 - 3832.67 (1000-60kg bags of beans) and bi-weekly *Futures* range from $21.98-175.18 US dollars.

	Rain	Temp	Rust	Production	Futures
count	337.000000	337.000000	337.000000	337.000000	337.000000
mean	183.548276	25.119281	16.415549	1731.397834	93.192611
std	96.050734	0.829998	12.286022	1286.144220	45.692033
min	0.200000	23.360900	0.330000	80.330000	21.980000
25%	115.400000	24.515500	6.570000	1010.330000	40.238000
50%	185.609000	24.990200	15.900000	1073.170000	106.243000
75%	256.832000	25.747800	22.670000	2783.330000	132.765000
max	407.700000	27.164000	50.000000	3832.670000	175.183000

Fig. 2. Summary Statistics

The next step was creating a plot of all the variables (*Rain*, *Temp*, *Rust*, *Production* and *Futures*) with a logarithmic scale. A log scale was chosen since the regular scale line plot wasn't very helpful to gain any insights due to the large range of values. The y-axis was changed to a logarithmic scale to see if any patterns could be detected in the data. A few possible conclusions can be drawn from this plot: *Rain* might be correlated to *Rust* and *Futures;* and *Temperature* does not appear to change a lot so it might not be a relevant variable. The plot is still very difficult to read since it uses a light yellow color for *Futures*, uses red and green that color-blind people cannot see, and it has too many variables on one chart. This is not the best visualization to use to draw definite conclusions about possible relationships between variables. Since this visualization could be difficult to interpret and since it did not add to telling the story of coffee rust and futures, it was not included in the final visualization presentation.

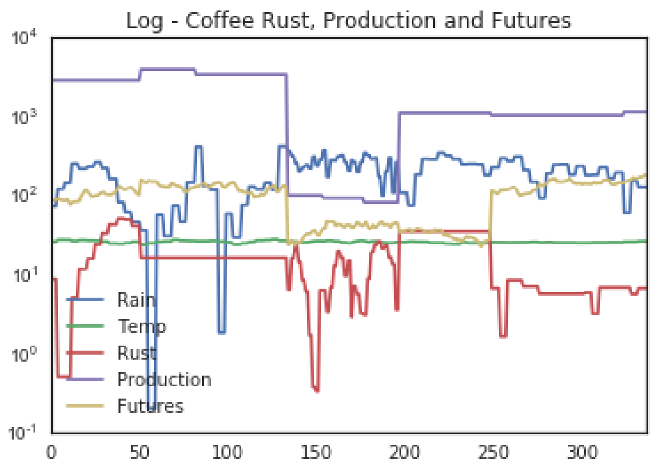

Fig. 3. Line Plot of Rust, Production and Futures

Next, a time-series plot of Coffee *Futures* was created. The color brown was used for encoding the futures variable over time in keeping with the web page theme. The futures vary quite a bit from about \$82-\$178 and there is a strange drop in prices between 150-250 days. Going back to the original data shows that futures minimum at \$21.98 in 1995. This graph confirms anecdotal comments from financial experts that *Futures* prices are volatile [31].

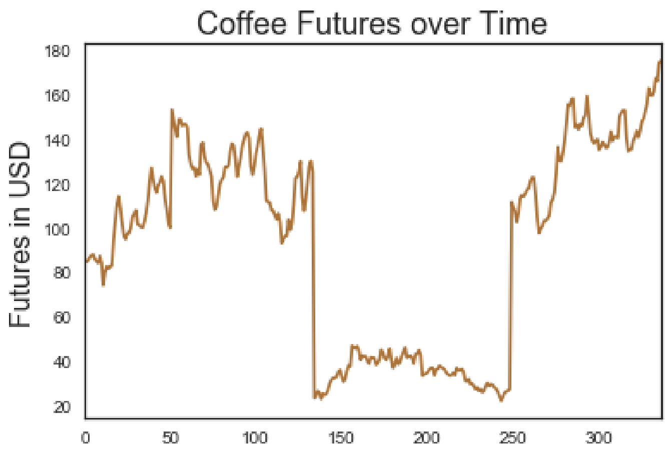

Fig. 4. Line Plot of Coffee Futures over Time

A line plot of *Rust* versus *Production* was created to quantify and visualize the relationship between the variables. The *Production* variable was plotted in black while the *Rust* variable was plotted with a brown color (#663300) to help the user distinguish between the two plots. The scales of the variables are so different from one another (production from 0-3800 and futures from 0-178) that any meaningful relationships are difficult to see in this visualization. The *morning coffee* color palette was used for all plots as described by [7].

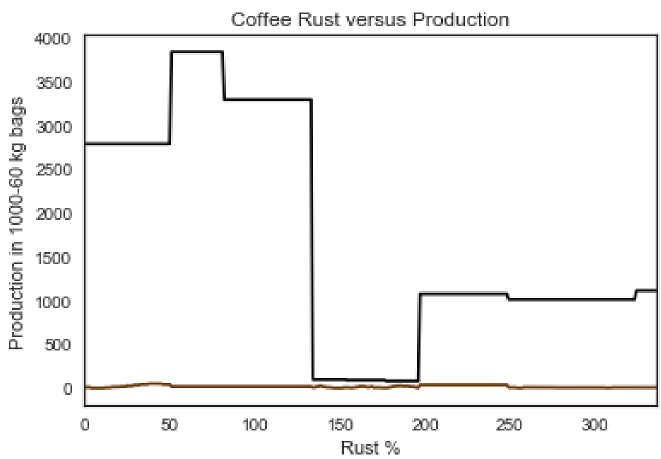

Fig. 5. Line Plot of Rust versus Production

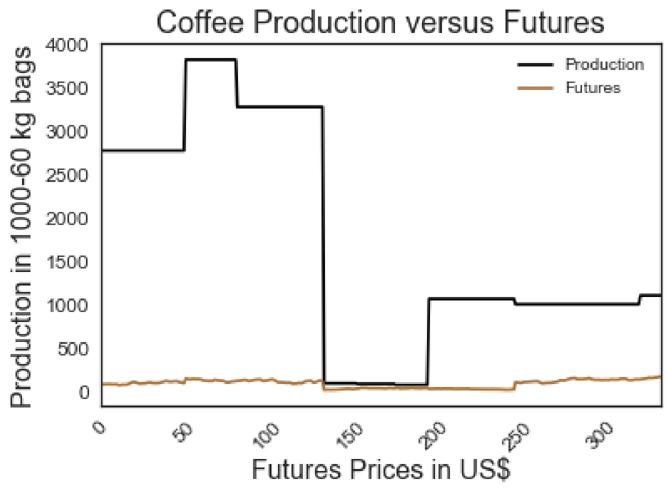

Fig. 6. Line Plot of Production versus Futures

A line plot of *Production* versus *Futures* was created to quantify and visualize the relationship between the variables. The *Production* variable was plotted in black while the *Futures* variable was plotted with a brown color (#663300) to help the user distinguish between the two plots. The scales of the variables are so different from one another (*Production* from 0-3800 and *Rust* from 0.33-50%) that any meaningful relationships are difficult to see in this visualization.

A line plot of *Rust* versus *Futures* was created to quantify and visualize the relationship between the variables. The scales of the variables are so different from one another (*Rust* from 0.33-50% and *Futures* from $21.98-175.18) that any meaningful relationships are difficult to see in this visualization. There may be a positive correlation between the variables but it's difficult to make any definite conclusions. The *morning coffee* color palette was also used as described above.

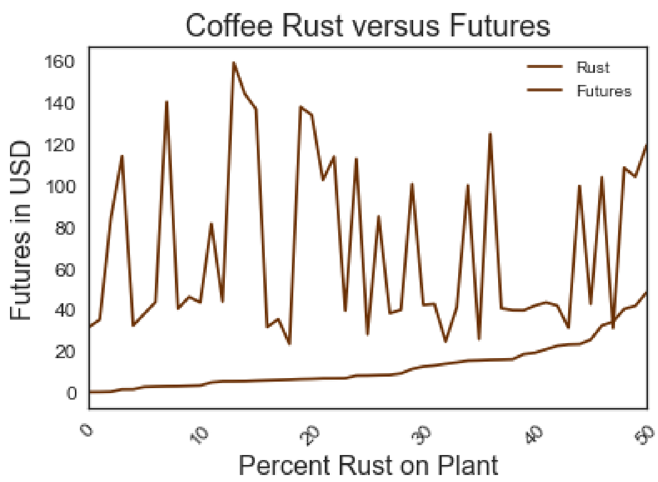

Fig. 7. Line Plot of Rust versus Futures

III. METHODS

As was briefly discussed with the line plot visualizations, there appears to be some correlation between *Rust* and *Futures*. There may be a correlation between *Rust* and *Production*. In order to verify and quantify any correlation between the five variables, another type of visualization needs to be created - *a correlation matrix*. In a correlation matrix, variables that are positively correlated (> 0) change in the same direction. Variables that are negatively correlated (< 0) change in the opposite direction.

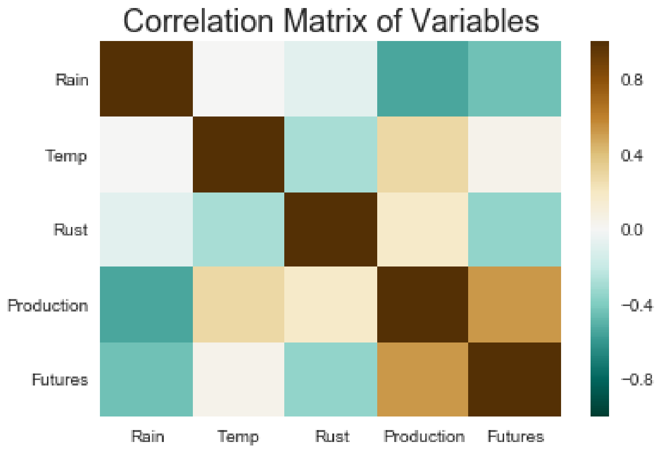

Fig. 8. Correlation Matrix

The correlation matrix shows that *Rain* and *Temperature* are not correlated to each other since they have a value of 0 on the color scale on the right of the visual. *Rain* is negatively correlated with production (-0.6), meaning that if the rain decreases, production increases. Therefore the initial hypothesis cannot not accepted.

Rain and *Rust* are negatively correlated to each other in a very small amount of -0.1. An increase in rain will decrease the amount of rust so this *hypothesis* that an increase in *Rain* will increase the amount of *Rust* cannot be not accepted.

Temperature is negatively correlated to *Rust* (-0.2) so if temperature increases, rust decreases. Based on this visualization, this *hypothesis* that increasing *Temperatures* increases *Rust* cannot be accepted.

Temperature is positively correlated to *Production* (0.3), meaning that if temperature increases, production decreases. This is the expected finding since increasing temperatures usually mean the coffee *Rust* can grow more easily and reduce production. Recall from the summary statistics above that the range of *Temperatures* in our data is 23.36 - 27.16° C. Since coffee rust thrives in temperatures from 10 - 30° C and our data falls within this range, it is logical that temperature would affect the amount of coffee rust. The *hypothesis* that increased *Temperature* increases coffee *Rust* can be accepted.

Production is highly correlated (0.8) with *Futures*. If production increases, futures prices increase. Since one of the initial hypotheses was that futures would decrease if production increases, the *hypothesis* cannot be accepted.

To understand how the coffee data is distributed, a histogram was created for *Rust*, *Production* and *Futures*. *Temperature* and *Rain* were not included since this visualization does not add any more insight into the story of quantifying *Rust*, *Production* and *Futures*.

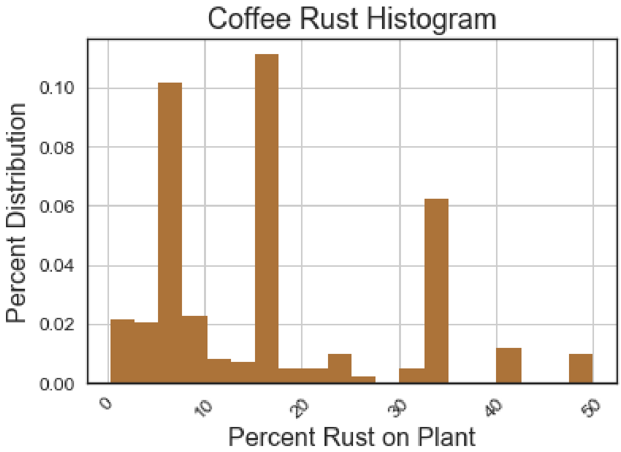

Fig. 9. Coffee Rust Histogram

From the coffee Rust histogram with a bin size of 20, we see most of the rust is about 5%, 18% and 33%. A histogram does not reveal why these peaks occur at these values. From the Coffee Production histogram with a bin chosen of 10, coffee production peaks at 1000 and then 3000 - 4000 (1000 - 60kg bags). There is a gap in the data between 1000 and 3000 since Papua New Guinea's production is much lower (80.33-97.92 in 1989-1991) than Colombia's production of 1010.33 in 2011-2012.

Fig. 10. Coffee Rust Histogram

The production histogram was not included in the final visualization presentation to avoid confusing the reader and because this visualization does not contribute significantly to accepting or failing to accept the hypotheses.

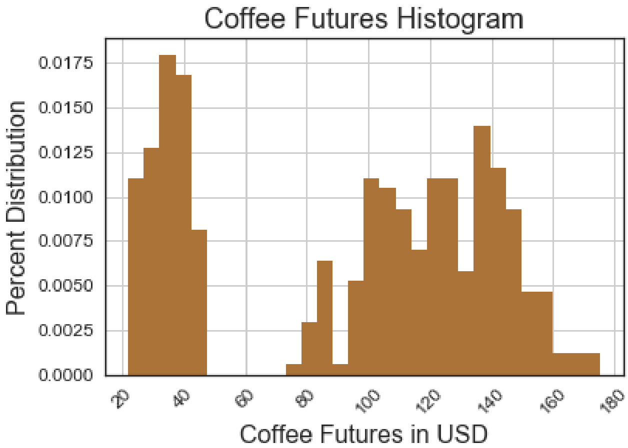

Fig. 11. Coffee Futures Histogram

The last histogram shows the distribution of coffee futures data. The default bin size of 10 was changed to 30 since this bin size best represented the variation in the data. There is not a high frequency of any futures data. The highest data distribution value is about 0.014 (when futures = $140). The gap in the histogram follows the gap in the data, with the futures ranging from $20-$40 in 1989-1991 and 1995 and futures ranging from $78-$175.18 from 2005-2013. This gap is probably caused by natural increases in futures prices between 1995 and 2005. Most farmers begin coffee rust pest management practices when rust exceeds 3%. The relationship between time and futures needs to be understood better to see if this is the reason for the bimodal histogram. Recall that the data set has futures prices in US$ from 1989-2013.

Kernel Density Estimates (KDE) were not plotted on any of the histograms since the KDE spillover below zero values. This biased plot with negative values happens since each of the *Rust*, *Production*, and *Futures* variables plotted are always positive. Rather than apply a logarithmic or other transformation to plot the KDE, these estimates were not included. Adding KDE to the histograms would not any additional value to telling the story of coffee rust, production and futures.

IV. Results

Results that present a quantifiable relationship between coffee *Rust*, *Production*, and *Futures* will now be discussed. A linear regression plot was created to quantify the relationship between *Rust* and *Production* using Python. The plot shows the relationship between *Rust* and *Production*. The colors were changed so that the regression line would be a brown that matched the morning joe color palette in the other visualizations and the markers were black circles that would be easy to read. The regression shows that as rust increases, production increases. The slope of *Production* is 0.222 and the normalized root mean square error (NRMSE) is 0.012.

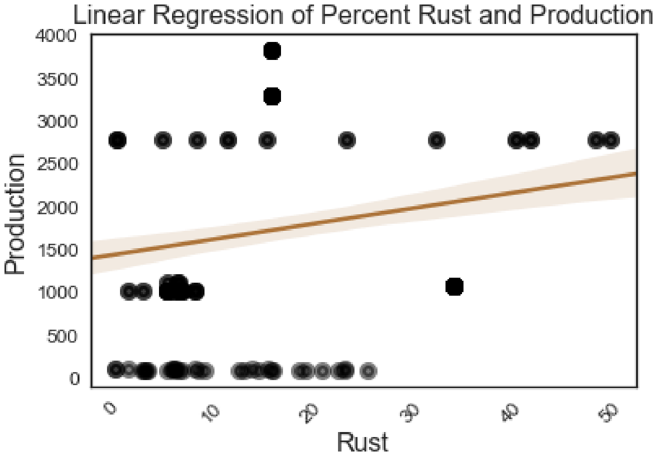

Fig. 12. Linear Regression of Rust versus Production

A linear regression plot was created for *Production* versus *Futures*. From this plot, there a positive correlation between the variables meaning if *Production* increases, *Futures* increase. The slope for *Futures* is 0.049 and the normalized root mean square error (NRMSE) is 0.301.

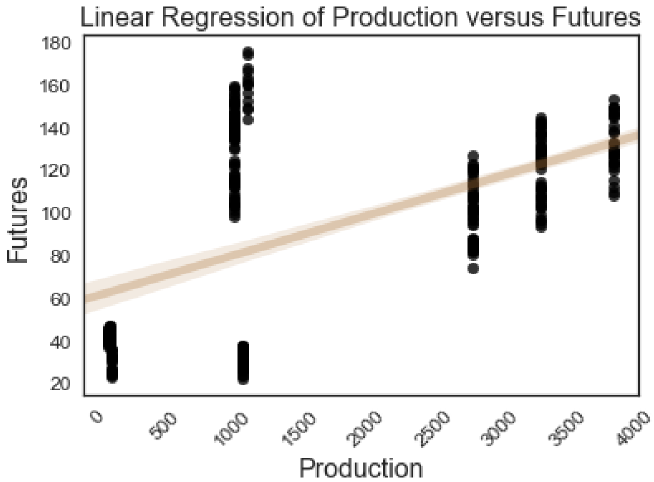

Fig. 13. Linear Regression of Futures versus Production

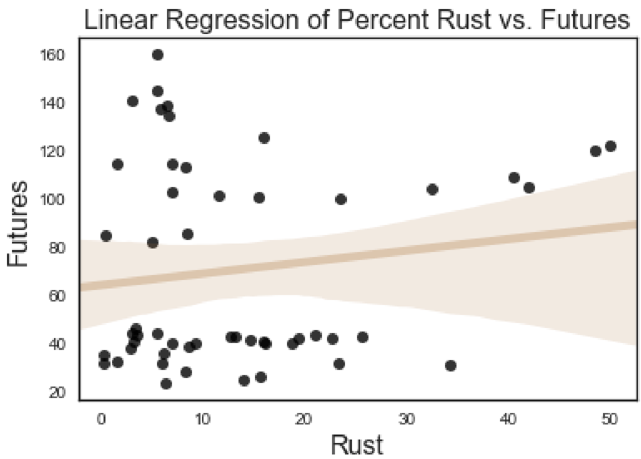

Fig. 14. Polynomial Regression of Futures versus Production

The final regression plot shown in Figure 14 is not a linear regression since a linear regression did not best fit the data. A polynomial regression shows the relationship between rust and futures being positive. The slope for *Rust* is 0.724 and the normalized root mean square error is 0.112

Results for the *Production*, *Futures* and *Rust* variables are presented in Table 1.

Variable	**Slope**	**NRSME**
Production	0.222	0.012
Futures	0.049	0.301
Rust	0.724	0.112

TABLE I. SLOPE AND NORMALIZED ROOT MEAN SQUARE ERROR RESULTS

A quick look at the x-variable (*Rust*) and y-variable (*Futures*) shows all but two of the first 12 observations have a positive correlation. This validates the positive correlation from the polynomial regression.

	Rust	Futures
0	0.33	31.620000
1	0.37	35.365000
2	0.50	84.378750
3	1.63	114.272500
4	1.67	32.495000
5	2.83	38.200000
6	3.00	43.800000
7	3.11	140.525000
8	3.17	40.620000
9	3.33	46.355000
10	3.50	43.650000

Fig. 15. Several Rust versus Futures Observations

V. Discussion

From the first linear regression of *Rust* versus *Production*, the hypothesis that as *Rust* increases, *Production* decreases *cannot be accepted* since the variables have a positive correlation. This plot comes close to but does not touch any of the points so a linear regression may not be the best fit for the data.

From the second linear regression of *Production* versus *Futures*, the hypothesis that as production increases, futures decreases *cannot be accepted*. This plot touches lots of the points which means it is probably the best fit for the data.

From the third polynomial regression of *Rust* versus *Futures*, the hypothesis that more Rust decreases Futures *cannot be accepted*. This plot touches only a few data points but comes much closer to fitting the data than the original linear regression used.

From the Slope results in Table 1, *Futures* changes the least with a value of 0.049 while *Rust* changes the most with a value of 0.724.

A deeper discussion on how and why the results were presented in a visual format is discussed in Appendix I.

VI. Conclusion

This research offers a quantitative framework for understanding the relationship between coffee rust, amount of coffee produced and futures prices. The final results show which of the hypotheses can be accepted.

The visualizations show the following:

- More rain does not affect coffee rust
- Higher temperatures = more coffee rust
- More coffee rust = more production
- More coffee rust = higher futures prices
- More coffee production = higher future prices

Recall, the original hypotheses

a. More rain increases rust. → *fail to accept*
b. Higher temperatures increase rust. → *accept*
c. More rust decreases production. → *fail to accept*
d. More production decreases futures. → *fail to accept*
e. More rust decreases futures. → *fail to accept*

The research shows there is a quantifiable link between the amount of rust coffee plants have and the amount of coffee produced. There is a quantifiable link between the amount of rust coffee plants have and coffee futures prices. Past data of rust-infected coffee plants from Brasil, Colombia and Papua New Guinea and futures prices worldwide can be used to show these connections. The visualizations helped to accept or not accept the original hypotheses.

Appendix A – Visual Presentation

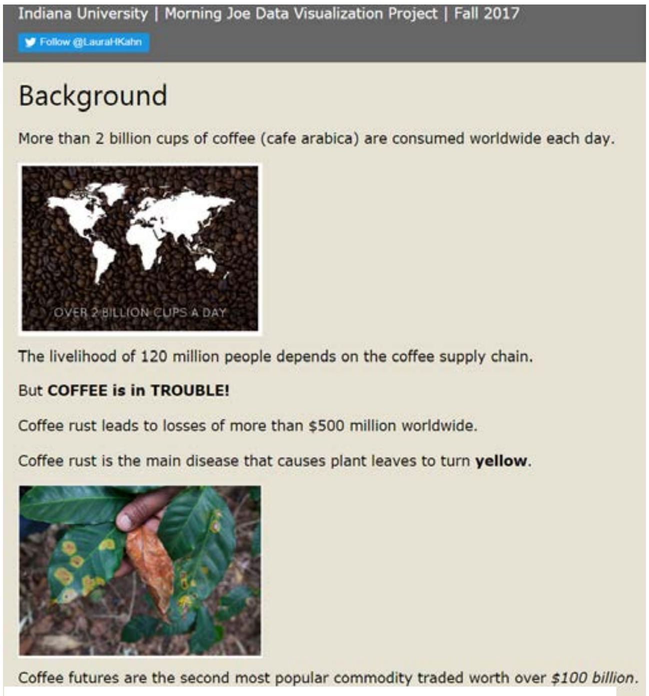

Fig. 1. HTML Web format

Fig. 2. JavaScript WordPress Template

Selecting a Web Page Format
To tell the story of coffee rust, production and futures and present the results for a general audience, a web page article was created using JavaScript WordPress templates. Here are the main reasons this format was chosen over the manually created HTML page.

1. Improved ability to insert image slideshows and images with a circle frame more easily.

2. It follows current web standards for presenting information visually and is Section 508 accessible.

3. It automatically detects the user device (desktop, tablet, phone) and scales the content appropriately.

4. It was created from an existing website (https://thedatalass.com). This site is publically available for the foreseeable future, contributing to the ideals of open science and open data.

The webpage contains the following content sections: *Introduction*, *Scope*, *Analysis and Results*, *Appendix* and *References*.

Introduction
Rather than duplicate work from scratch, existing unlicensed static images with clear resolution were used to draw the reader's eye attention to the research significance to the project.

Fig. 3. Image 1 – Coffee beans in the background of world map

It was important to emphasize how many cups of coffee were consumed each day to visualize the importance coffee plays in our daily lives and why this project is important. The last static image added to the slideshow at the top of the page was a chart of coffee futures prices [9]. This image was chosen since it shows the reader what is meant by coffee futures.

Click for Chart	Current Session								Prior Day		Opt's
	Open	High	Low	Last	Time	Set	Chg	Vol	Set	Op Int	
Dec'15	134.70	137.50	134.60	135.75	11:49 Oct 14	-	1.40	21527	134.35	86297	Call Put
Mar'16	137.75	140.90	137.75	139.15	11:49 Oct 14	-	1.40	7093	137.75	43090	Call Put
May'16	140.90	142.85	140.70	141.60	11:49 Oct 14	-	1.80	952	139.80	21607	Call Put
Jul'16	141.40	144.15	141.40	142.95	11:49 Oct 14	-	1.55	696	141.40	9806	Call Put
Sep'16	144.75	145.95	144.00	144.65	11:49 Oct 14	-	1.60	509	143.05	7068	Call Put
Dec'16	146.00	148.00	146.00	146.75	11:49 Oct 14	-	1.60	218	145.15	9141	Call Put
Mar'17	150.00	150.00	150.00	150.00	11:49 Oct 14	-	2.80	90	147.20	1837	Call Put
May'17	-	-	-	148.85 *	11:49 Oct 14	-	-	-	148.85	700	Call Put
Jul'17	152.35	153.10	151.95	151.95	11:49 Oct 14	-	1.35	39	150.60	431	Call Put
Sep'17	153.35	153.35	153.00	153.00	11:49 Oct 14	-	1.35	13	151.65	439	Call Put
Dec'17	155.75	155.85	154.00	154.35	11:49 Oct 14	-	1.15	3	153.20	530	Call Put
Mar'18	157.05	157.05	155.70	155.70	11:49 Oct 14	-	1.15	1	154.55	98	Call Put

Commodity Futures Price Quotes For Coffee (ICE Futures) (Price quotes for ICE Futures Coffee delayed at least 10 minutes as per exchange requirements) REFRESH DATA

Figure 4. Image 2 – Coffee Futures Price Quotes

These images were added as a slideshow in the *Introduction* section.

Following this slideshow of static images, another two static images showing what coffee rust looks like on a coffee plant and the coffee berries. The images were added with circle frames.

Figure 5. Image 3 – Coffee Rust

Scope
Following an introduction to the topic, country-specific static maps of coffee growing regions were added to the web page *Scope* section. The web page features a map of Brasil's coffee growing regions that has been slightly modified from the black and white image below.

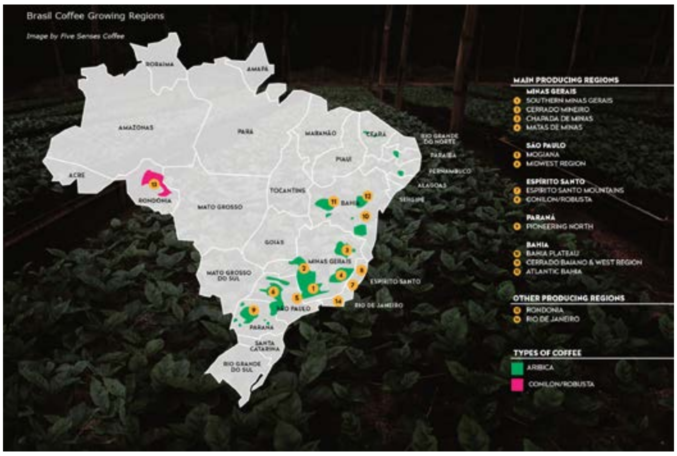

Fig. 6. Before image of Brasil Coffee Growing Regions

The Before image in Figure 6 shows administrative boundaries of Brasil (spelling in country's native language) and has a legend with a number that shows where each type of coffee is grown on the map. The Before Image has extra circle numbers, state names, and green sections within states that distracts the reader from the research focus so these encodings were removed. The goal of the modified map was to communicate that certain administrative areas of Brasil produced coffee. A blank administrative map of Brasil was modified from Wikipedia. The colors of the administrative boundaries were that produced coffee were changed to dark brown, a title and legend were added and the image was resized.

Fig. 7. After image of Brasil Coffee Growing Regions

The same process was followed to modify maps of Colombia and Papua New Guinea to show coffee growing regions. The goal of these maps was to give the user a geospatial reference to which countries were part of the project.

The map of coffee growing regions in Colombia was derived from information in Columbian Connection [8]. This map gives the reader geospatial context of each of the regions within the country and also where Colombia is located in South America with an inset at the bottom right corner. The color palette in the original image from the Columbian Connection is not distinct enough to show the different coffee-growing regions so it was modified. All the slightly different shades of blue and green were replaced with more contrasting colors following data visualization design principles. A blank SVG map was obtained from Wikimedia and modified.

Fig. 8. Before image of Colombia Coffee Growing Regions

Fig. 9. After image of Colombia Coffee Growing Regions

The coffee growing regions of Papua New Guinea include the Eastern Highland Province, Western Highland Province and Simbu [23]. A map color encoding these regions was created using the same colors as the other country maps. The original blank country map with administrative regions from Wikimedia Commons was used to make the final map including a descriptive title.

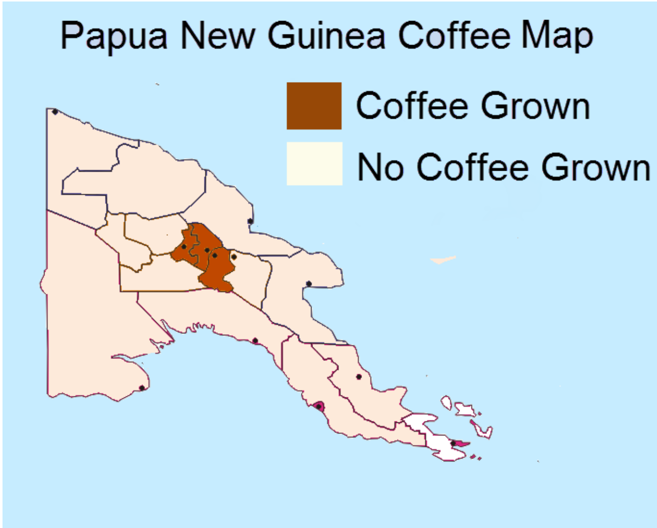

Fig. 10. Papua New Guinea Coffee Growing Regions

Analysis and Results

Content for the *Analysis and Results* section of the web page is described in the Methods, Results and Discussion above.

Appendix

Since the WordPress template does not allow for adding any code, a separate HTML file using D3 JavaScript library was used to create an interactive box plot. This visualization shows coffee Rust intensity level versus Futures. The x-axis represents the futures prices in USD while the y-axis represents the rust severity levels. A user can hover over any of the different level box plots to see the minimum, median and maximum futures prices for that particular level. Since the visualization shows *Rust* level rather than *Rust* percent, it is a different measurement than the original scope and is included in the Appendix.

Failed Experiments

In addition to the visualizations discussed in the Methods and Results section, there were three failed experiments of visualizations that were not included in the final web page.

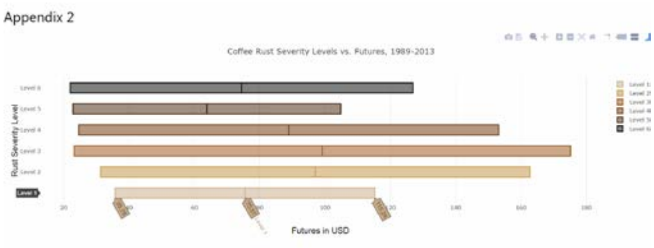

Fig. 11. Interactive Box Plot of Rust versus Futures

Fig. 12. Failed Experiment – Choropleth Map of Brasil Coffee Regions

One failed experiment was creating a choropleth map visualization that would show amount of coffee production for each country (similar to Figure 12). Administrative shapefiles that show the country's boundaries are freely available for each country. However, multiple shapefiles existed for each country and there was a technical barrier knowing how to combine these files into one shapefile per country before adding the coffee data. Furthermore, even if the technical barrier was overcome, this file would have had to been converted into a GeoJSON file before creating a custom choropleth map with D3 JavaScript library. All examples of existing geographic visualizations were made with one single shapefile. Even if the technical issues would have been resolved, it was decided that this type of choropleth map would only emphasize the *Production* variable rather than the relationship between *Rust*, *Production* and *Futures*.

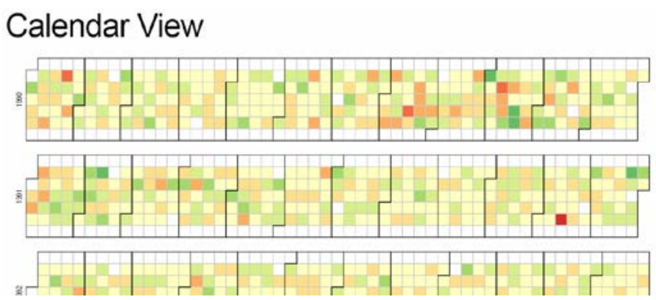

Fig. 13. Failed Experiment – Calendar View of Daily Futures

At the beginning of the project, there was a plan to include an interactive Calendar View of Futures. This would have allowed the user to hover over a block in the visualization to see that day's future price. A Calendar View visualization would only emphasize the futures variable. Therefore, it was decided not to include this type of visualization since it too would draw attention away from the relationship between *Rust*, *Production* and *Futures*.

The final failed experiment was an interactive line chart of the amount of cumulative coffee rust over time in all three countries. The chart appears to be blank at first. However, when a user clicks on the Show Data button, the line plot appears in real-time and the user sees the futures prices changing from 1989-2013. The Reset button clears the line graph off the chart. Since there are multiple gaps in the data, this visualization would imply that it's acceptable to interpolate the missing data points. This would be an inaccurate way to interpret the data so the visualization was not included.

Appendix 1

Fig. 14. Failed Experiment – Interactive Line Plot of Futures over Time

References

[1] Alves, Marcelo de Carvalho. Cafeicultura De Precisão Na Proteção de Plantas: Monitoramento de Pragas E Doenças. Expocafé 2012 - slide 28. https://pt.slideshare.net/cafeicultura/expocafe-marcelo. Accessed 17 October 2017.

[2] Avelino, J., H. Zelaya, A. Merlo, A. Pineda, M. Ordoñez and S. Savary. The intensity of a coffee rust epidemic is dependent on production situations. *Ecological Modeling*, 197 (2006): 431-447

[3] Avelino, J., Cristancho, M., Georgiou, S. et al. The coffee rust crisis in Colombia and Central America (2008-2013): impacts, plausible causes and proposed solutions. Food Sec. (2015) 7: 303. https://doi.org/10.1007/s12571-015-0446-9.

[4] Bock, K.R., 1962b. Dispersal of uredospores of *Hemieia vastatrix* under field conditions. Trans. Br. Mycol. Soc. 45, 63-74.

[5] Cintra, M., C.A.A. Meira, M.C. Monard, H.A. Camargo. And L.H.A. Rodrigues. The Use of Fuzzy Decision Trees for Coffee Rust Warning in Brazilian Crops. 2011 11th International Conference on Intelligent Systems Design and Applications. DOI:10.1109/ISDA.2011.6121847. Accessed 17 October 2017.

[6] Coffee C Futures. Intercontinental Exchange. https://www.theice.com/products/15/Coffee-C-Futures. Accessed 27 November 2017.

[7] Color-Hex Color Palettes. http://www.color-hex.com/color-palette/50402. Accessed 21 November 2017.

[8] Colombian Connection Coffee. http://colombianconnection.com.au/coffee/. Accessed 5 December 2017.

[9] Commodity Futures Price Quotes for Coffee. http://data.tradingcharts.com/futures/quotes/kc.html. Accessed 11 September 2017.

[10] Corrales, D., Ledezma, A., Peña, A., Hoyos, J., Figueroa, A. & Corrales, J. (2014). A new dataset for coffee rust detection in Colombian crops base on classifiers. Revista S&T, 12(29), 9-23.https://www.icesi.edu.co/revistas/index.php/sistemas_telematica/article/viewFile/1802/2333. Accessed 4 November 2017.

[11] Custudio, Adriano Augusto de Paiva et al . Comparison and validation of diagrammatic scales for brown eye spots in coffee tree leaves. **Ciênc. agrotec.**, Lavras , v. 35, n. 6, p. 1067-1076, Dec. 2011. http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-70542011000600005&lng=en&nrm=iso. Accessed 3 October 2017.

[12] Cunha, R.L; Mendes, A.N.G.; Chalfoun, S.M. Controle químico da ferrugem do cafeeiro (Coffea arabica L.) e seus efeitos na produção e preservação de enfolhamento. Ciência e Agrotecnologia, v. 28, n.5, p.990-996, 2004.

[13] De Carvalho Alves, M., da Silva, F.M., Sanches, L. et al. Geospatial analysis of ecological vulnerability of coffee agroecosystems in Brazil. Appl Geomat (2013) 5: 87. https://doi.org/10.1007/s12518-013-0101-0

[14] Economics of coffee. Wikipedia: https://en.wikipedia.org/wiki/Economics_of_coffee. Accessed 3 October 2017.

[15] Escobar, Kimberly. "Coffee Rust is killing Latin American Plants." http://www.dbknews.com/2016/10/28/umd-research-grant-coffee-rust-fungus/. Accessed 18 October 2017.

[16] Estadísticas y Precios. Instituto del Café de Costa Rica. http://www.icafe.cr/sector-cafetalero/informacion-de-mercado/estadisticas-y-precios/. Accessed 4 December 2017.

[17] Futures Historical Trading Charts. http://futures.tradingcharts.com/historical/CF/1989/3/linewchart.html. Access 27 September 2017.

[18] Galli, F.; Carvalho, P.C.T. Doenças do cafeeiro. In: GALLI, F. (Coord.). Manual de fitopatologia. São Paulo: Editora Ceres, 1980. 587p. V.2. p.128-140.

[19] *Hemileia Vastatrix*. Wikipedia.https://en.wikipedia.org/wiki/Hemileia_vastatrix. Accessed 22 October 2017.. www.farmdoc.illinois.edu/nccc134. Accessed 2 December 2017.

[20] Japiassú, L. B, A. W.R. Garcia, A.E. Miguel, M.S. A. Mendonça, C.H.S. Carvalho, R.A. Ferreira. Influencia da Carga Pendente, Do Espaçamento e de Fatores Climáticos no Desenvolvimento da Ferrugem do Cafeeiro. Estação de Avisos Fitossanitários Do Mapa/Fundação Procafé. http://www.sapc.embrapa.br/arquivos/consorcio/spcb_anais/simposio4/p292.pdf. Accessed 10 October 2017.

[21] Jaramillo J., Muchugu E., Vega F.E., Davis A., Borgemeister C., Chabi-Olaye A. (2011). Some Like It Hot: The Influence and Implications of Climate Change on Coffee Berry Borer (*Hypothenemus hampei)* and Coffee Production in East Africa. PLoS ONE 6(9): e24528. https://doi.org/10.1371/journal.pone.0024528

[22] Kim, K. and W. Lee. Stock market prediction using artificial neural networks with optimal feature transformation. Neural Computing and Applications (2004), 13: 225-260. doi: 10.1007/s00521-004-0428-x.

[23] Lamouroux, N., F. Pellegrin, D. Nandris and F. Kohler. The Coffea arabica Fungal Pathosystem in New Caledonia: Interactions at Two Different Spatial Scales. J. Phytopathology 143, 403-413 (1995). http://onlinelibrary.wiley.com/doi/10.1111/j.1439-0434.1995.tb00284.x/abstract.

[24] Luaces, O., L. H.A. Rodrigues, C. A.A. Meira, A. Bahamonde. Using nondeterministic learners to alert on coffee rust disease. Expert Systems with Applications 38(2011): 14276-14283.

[25] Magrach A, Ghazoul J (2015). Climate and Pest-Driven Geographic Shifts in Global Coffee Production: Implications for Forest Cover, Biodiversity and Carbon Storage. PLoS ONE 10(7): e0133071. doi:10.1371/journal.pone.0133071.

[26] Montage, D. Algorithmic Trading of Futures via Machine Learning. Stanford University CS229 Course material. http://cs229.stanford.edu/proj2014/David%20Montague,%20Algorithmic%20Trading%20of%20Futures%20via%20Machine%20Learning.pdf. Accessed 11 September 2017.

[27] Paulo, E.M., Montes, S.M.N.M., & Fischer, I.H.. (2013). Progresso temporal da ferrugem alaranjada em cultivares de cafeeiro no Oeste de São Paulo. *Arquivos do Instituto Biológico*, *80*(1), 59-64. https://dx.doi.org/10.1590/S1808-16572013000100009.

[28] Perez-Ariza, CB, Nicholson, AE & Flores, MJ 2012, Prediction of coffee rust disease using Bayesian networks. in A Cano, M Gomez-Olmedo & TD Nielsen (eds), *Proceedings of the Sixth European Workshop on Probabilistic Graphical Models.* University of Granada, Granada Spain, pp. 259 - 266, European Workshop on Probabilistic Graphical Models (PGM), Granada Spain, 1 January.

[29] Quiñones, P., Javier, A. Effects of Daylength and Soil Humidity on the Flowering of Coffee Coffea Arabica L. in Colombia. Rev.Fac.Nal.Agr.Medellín, Medellín , v. 64, n. 1, p. 5745-5754, June 2011. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0304-28472011000100001&lng=en&nrm=iso. Accessed 3 Oct. 2017.

[30] Ruiz-Cárdenas, R. and P. Baker. Life table of Hypothenemus hampei in relation to coffee berry phenology under Columbian field conditions. Sci. Agric. (Piracicaba, Braz.), v.67, n.6, p.658-668, November/December 2010. http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-90162010000600007. Accessed 1 October 2017.

[31] Seven Things You Must Know about Coffee Futures. https://tradingsim.com/blog/7-things-you-must-know-about-coffee-futures/. Accessed 12 September 2017.

[32] Slawson, Nicola. Coffee and climate change: what you need to know. The Guardian. https://www.theguardian.com/sustainable-business/2016/sep/27/coffee-climate-change-supply-chain-risk-smallholders-fairtrade-tech. Accessed 22 November 2017.

[33] World Bank Climate Change Knowledge Portal. http://sdwebx.worldbank.org/climateportal/index.cfm?page=country_historical_climate&ThisCCode=PNG. Accessed 1 November 2017.

[34] Yap, Jamie. Stat 531 Midterm Project. https://ionides.github.io/531w16/midterm_project/project10/midtermproject.html. Accessed 27 November 2017.

[35] Zambolim, L.; Vale, F.X.R.; Pereira, A.A.; Chaves, G.M. Café (Coffea arábica L.): Controle de doenças causadas por fungos, bactérias e vírus. In: VALE, F.X.R.; ZAMBOLIM, L. (Ed.). Controle de doen-ças de plantas. Viçosa: Suprema Gráfica e Editora, 1997. pp.83-180.