Module 8 Capstone Project Final Report Option 1 United States Organization

David P. Jurist

MIS581-1 – Capstone Business Intelligence and Data Analytics

Colorado State University – Global Campus

Professor Jamia Mills, PhD

July 7, 2019

**Module 8 Capstone Project Final Report Option 1 United States Organization**

Following is the Module 8 Capstone Project final report for MIS581. Option #1 (United States organization) is selected. This Module 8 document completes a final writeup of data analytics and project conclusions. It includes discussion of the organization, analysis of its data for strategy, findings, and results of hypothesis testing. It is complemented by a separate PowerPoint presentation and delivery of an oral presentation in a YouTube video.

**Abstract**

The Capstone Portfolio Project documented in this writeup will contribute to the author's work in the future. The global food chain and the international agriculture industries are so complex and so vast that they cannot be comprehensively analyzed in this brief report of the overall project. Future work will consist of developing interrelated models of all the input and output measures from a macro- and micro-perspective with descriptive, prescriptive and predictive data analytics. The models are intended to eventually portray food crop commodity production volume and pricing, the production and consumption of finished foods, and the use of natural and manufactured resources in the various production processes. This initial work will be useful for understanding how to support the growing global population under stressed ecological systems with the sustenance needed for survival and health. These efforts will inform new solutions, predict future estimations, and prescribe sustainable choices.

**Introduction**

It has been acknowledged in the business data analytics community that the agriculture industry had long been underserved by advanced data analytics and business analytics. That has been changing, however, since large agribusiness firms such as ADM and Cargill have entered the business of providing data analytics as a professional service to landholders and growers. “Information and Communication Technologies tools on the ag domain is rapidly and drastically changing the relationship between farmers and the land, by impacting every aspect of products and processes related to it. Expressions like smart and predictive agriculture, precision farming, ‘aginformatics’ and farming 4.0 constantly contribute to depict the farming sector worldwide as an even more digital landscape where structures and systems are inextricably tangled with digital devices and services, bound to make the ag chain and industry increasingly accountable and sustainable” (Leone, 2017, p. 507). The digital age in agriculture is real, and presents opportunities for advanced analytics that other industries already apply innovatively. If this trend continues, it will benefit those who need it the most in this business and industry.

The natural resources for food production and its products have been inequitably distributed by territorial, political systems and international agribusinesses. Finite natural resources on our planet and its ecology must be conserved before they become depleted or extinct. How people select and consume food products should for their own sake be realigned for health and longevity, not only for glamourous epicureanism, convenience, cheap price, and flavor alone. Ranchers have been committing suicide in increasing numbers, because of their financial dilemmas, and small landowners and growers of subsistence crops have been impoverished in spite of their diligent manual work. Potable water is unequally available around the world. The population of the world may grow to nearly nine billion by 2035 because of population growth in Africa and Asia. The growing population will stress both natural resources and edible food, and threaten the life-support systems of the entire global ecology.

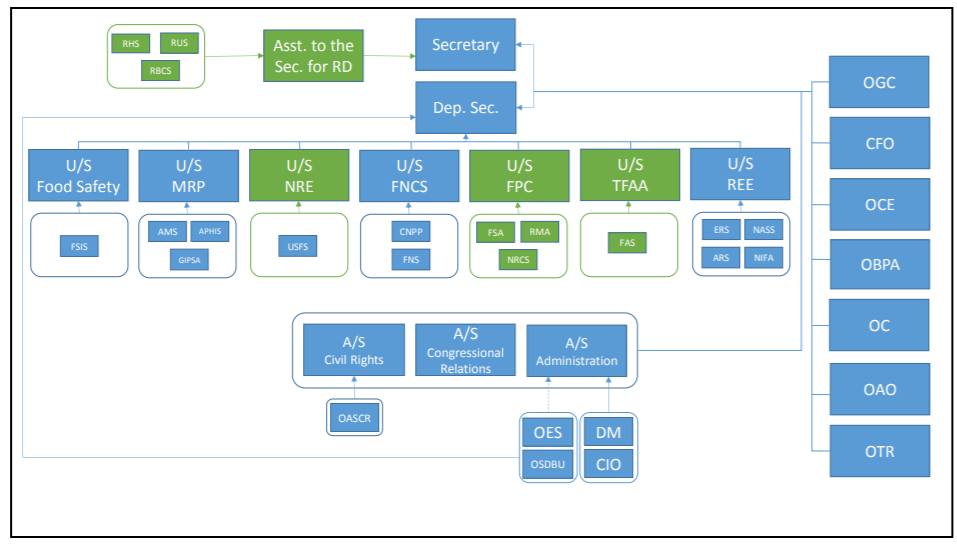
**The United States Organization**

The United States organization chosen for this project is the United States Department of Agriculture (USDA), because the United States has long been regarded as the breadbasket of the world. The mission statement and vision statement of the department state:

“We provide leadership on food, agriculture, natural resources, rural development, nutrition, and related issues based on public policy, the best available science, and effective management. We have a vision to provide economic opportunity through innovation, helping rural America to thrive; to promote agriculture production that better nourishes Americans while also helping feed others throughout the world; and to preserve our Nation's natural resources through conservation, restored forests, improved watersheds, and healthy private working lands. Our strategic goals serve as a roadmap for the Department to help ensure we achieve our mission and implement our vision” (USDA, 2018a, p. 1).

**Describe the Organization**

The USDA has been recognized as one of the top ten best places in the country to work. The USDA reports to the Executive Office of the United States Government. It comprises a main body in Washington, DC and in field offices, its independent agencies, and its staff. The Secretary of the USDA prepared the organization chart below in Figure 1 to depict the manner in which the USDA has organized itself with approval of Congress.



*Figure 1*. Organization of USDA Trade Functions, Under Secretary for Trade and Foreign Agricultural Affairs (TFAA), and Elevation of Rural Development. Retrieved from https://www.usda.gov/sites/default/files/documents/report-proposed-2017-reorg-usda-under-secretary-trade-foreign-ag-affairs.pdf

**History**

President Abraham Lincoln signed legislation to establish the United States Department of Agriculture on May 15, 1862. In his final message to Congress, President Lincoln called it “The People's Department.” “Through our work on food, agriculture, economic development, science, natural resource conservation and other issues, USDA has impacted the lives of generations of Americans” (USDA, 2018a, p. 4).

**Services and Products**

The USDA is a department of the executive branch of the United States Government. As a government body, it neither manufactures nor sells any products. Its role in the economy is to provide government services that fulfill its main mission to serve the American public, who pay the federal taxes supporting its people and operations. The department generates printed reports and Big Data (USDA, 2018b), which it collects from its field enumerators and statisticians.

**Location**

The USDA is headquartered at 1400 Independence Avenue, Washington, DC. The USDA and its agencies operate at 4,500 locations around the United States and in other countries.

**Number of Employees**

The USDA comprises twenty-nine agencies. It employs nearly 100,000 people who directly serve the American public generally and agricultural producers specifically.

**Revenue**

The USDA receives no revenue from sales of services and products. It is funded by federal taxes paid by the American public and approved by Congress to support the Executive Branch of the United States Government. The department’s 2020 budget (USDA, 2019a, p. 1) has allocated a total of $150 billion for its expenditures, of which $22 billion is for discretionary spending and $128 billion for mandatory programs.

**Why I Chose the Organization**

I chose the USDA, because they are an admirable organization guided toward a noble mission. They are likely the single most reliable source in the country for information about the American agriculture industry, its participants, and its components. I endeavor to learn more about how American agriculture participates in the entire global food chain, and how I may begin to model and simulate it with our analytical tools, as O’Leary (2017) summarizes the purposes of research, for expanding knowledge, for improving professional practice of the data sciences, to impact policy, and to benefit individual growers, processors, and consumers. Kleppel (2014) wrote, “A new kind of farming is emerging in America” (p. XIII), and I would like to participate in this important evolution by helping to improve our “prospects for revolutionary change in our food system” (Kleppel, 2014, p. XIII) with advanced data analytics that benefit all of its stakeholders.

**Objectives of Capstone Project**

1. Search, survey, analyze, and model data from diverse sources about global food-chain.
2. Select USDA data that can be modeled with data analytics for actionable business value (Hughes-Cromwick, & Coronado, 2019).
3. Demonstrate how the USDA can apply these tools to analyze data in a project.
4. Transmute business intelligence and data about food crops into actionable information.
5. Address ethics, privacy, and security considerations.
6. Lend a perspective for food-insecure nations and their food producers.
7. Lend a perspective for informed decisions by American producers and processors to meet their multiple agribusiness purposes.

**Overview of Study**

This study is an academic pursuit that has received no direct participation in the case studies and data analytics from such major participants in the industry as the U.S. Department of Agriculture (2018a), the Food and Agriculture Organization of the United Nations (2018), or the largest agribusinesses in the world (Lafarge, Cargill, Continental, United Nations, UNESCO, ADM, or Monsanto (2018)) in both the public and private sectors. It is intended to begin a lifelong body of work bringing current business analytics methods and data analytics tools to bear upon the vast sources of food crop volume reports and macro-economic influences on the supply of and demand for products. The study consists of a complex literature review and early analytics of selected data sets.

**Literature Review**

The complex literature review comprises a search, survey and critique of print media in the author’s personal library, online web pages, scholarly literature from the CSU Global Online library databases and catalogs, and searches of white papers online at distinguished business data analytics sites such as J.P. Morgan Chase Institute and McKinsey Agricultural. Publications on the USDA web pages were surveyed, such as the monthly *Outlook for U.S. Agricultural Trade* and monthly crop reports with volumes and prices. Other influential writings include Kleppel (2014), *Successful Farming* (Bedord, 2018, p. 49), *Eating Well* (Estabrook, 2019), and the predictive capabilities of the National Earth System (Carman, et al., 2017). Reviews such as the Longo (2008) study about fertilizers, pesticides, and environmental impacts will be entertained for at least preliminary input to the total modeling effort. “Estimating allocation of land to different uses is becoming increasing important in the light of population growth, urbanization, expansion of agriculture and global climate change” (Ngugi, 2013, p. 637). Conversely, the Roth (2011) journal article discusses the influences of other variables on agricultural revenue as the dependent dimension. Their primary survey “modeled the spatial patterns of urban growth under seven policy scenarios and calculated potential loss of annual agricultural revenue from each” (p. 1). A traceability model would be revealing for usage and consumption correlations to production, such as documented by Buskirk, Schweihofer, Rowntree, Clarke, Grooms, and Foster (2013). No personal discussions could be obtained.

**Research Design**

**Methodology**

Methodology for the project followed a predominantly quantitative approach, because a qualitative approach was not suitable for any data discovered. The data for this project is time series data for food crop volumes and prices. The IKANOW (2014) editorial lends a high-level, eight-step approach that guided the work; namely,

1. Determine the problem to solve.
2. Understand how these problems impact business and then develop use cases.
3. Measure the success of the project.
4. Ascertain if this problem is solved.
5. Decide if the solution should live in the cloud, on premise, or hybrid solution.
6. Evaluate data requirements.
7. Identify gaps.
8. Choose an agile or iterative approach.

**Methods**

No primary data was collected. All data was secondary data already published by the USDA and other sources. Data was gathered, examined, consolidated, prepared, reformatted, and saved from Excel spreadsheets into csv files, which were imported into tools such as Weka and Tableau. Exploratory data analysis (EDA) followed inductive modality. As many of the algorithms in the data analytics tools as are reasonable will be applied during future work for clear understanding of time series and other data, such as regression and neural networks.

**Limitations**

Paradoxically, the limitation on this project is its vast scope. A comprehensive model would encompass a global approach to the integrated model and its sub-models, which will be impossible to accomplish during so brief a project. It was a challenge to apply all the plausible algorithms on the selected data for exploratory data analysis.

**Security, Privacy, and Ethical Considerations**

Ethics are professional standards of conduct governing our treatment of Big Data “to consider serious ethical issues including whether certain uses of big data violate fundamental civil, social, political and legal rights” (Davis, 2012, p. viii). The ethical considerations surrounding identity, privacy, ownership and reputation remained a consistent guide for work throughout this project. “Philosophy and business don’t always get along well” (Davis, 2012, p. vii). Nevertheless, no security, ethical or privacy issues imposed any impediments during this project. Davis’ (2014) ethical framework of inquiry, analysis, articulation and action will continue to support the research and analysis on an individual level interfacing with other organizations that publish the data. The USDA and other data sets are openly provided on public web sites with the express intent of making the data available to us for business and academic research analytics. “ERS economists and social scientists develop and disseminate a broad range of science-based economic and statistical information to the public” (United States Department of Agriculture, 2018a, p. 2). USDA publicly documented its policies and standards for how it operates to protect privacy and security.

**Research Hypothesis and Testing the Null Hypothesis**.

“Hypothesis testing is a framework for testing whether a pattern in the data is ‘statistically significant,’ by looking at how likely it is to occur by random chance in the absence of an underlying real-world phenomena” (Cady, 2017, p. 296). Research expectations are that visualizations of early models of global food can be represented from EDA. Other more focused hypotheses were decided. “Formally stating a claim and the subsequent hypothesis test is done with a *null* and an *alternative* hypothesis. The null hypothesis is interpreted as the *baseline* or *no change* hypothesis and is the claim that is assumed to be true. The alternative hypothesis is the conjecture that you’re testing for, against the null hypothesis” (Davies, 2016, p. 386).

The initial research hypothesis with the data for this project is that global food production varies inversely to energy prices. It is anticipated that, since agriculture is so dependent on crude oil distillates for energy and the nitrogen fertilizers, food production would decline as energy prices increase, and would increase as energy prices decrease.

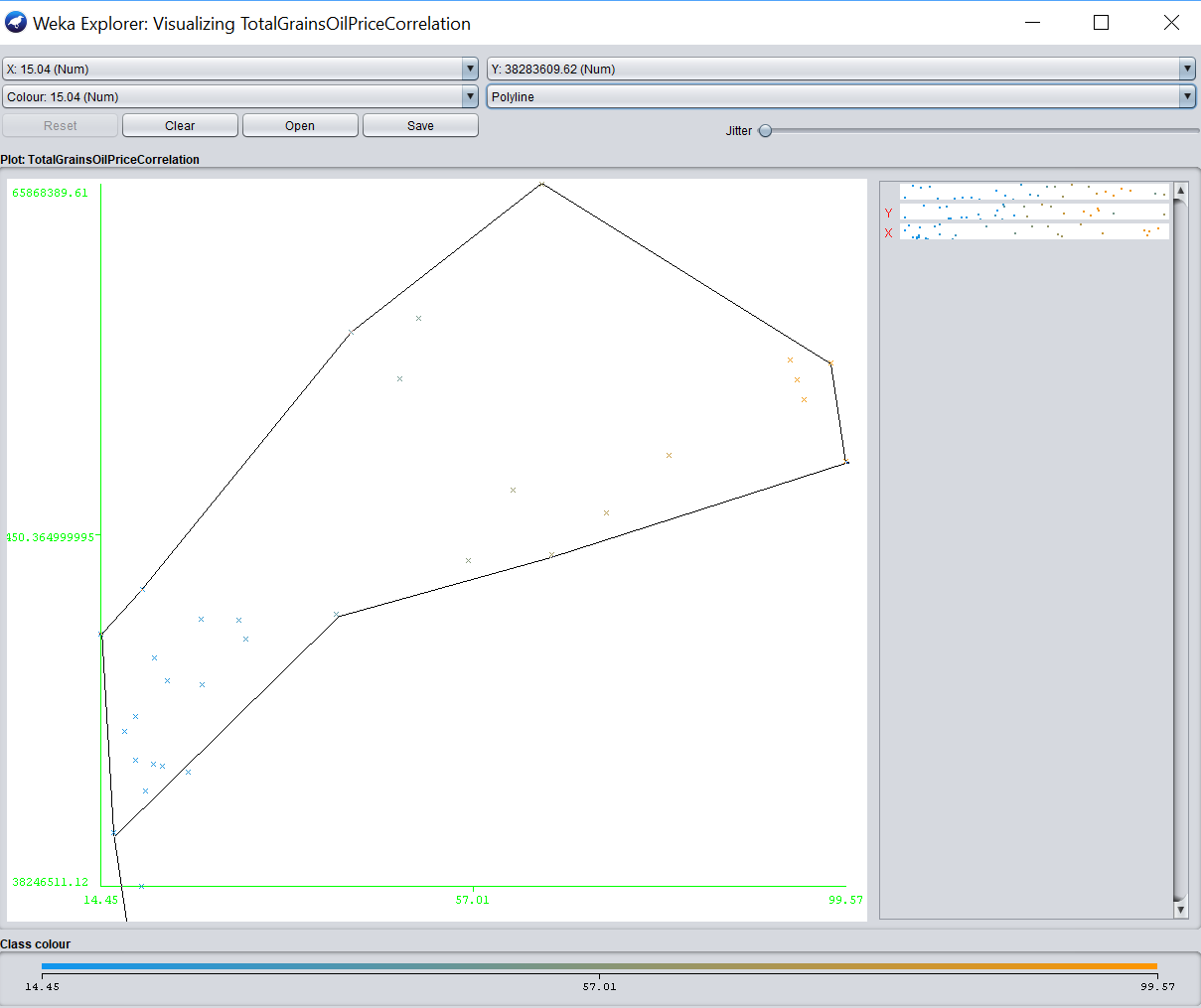
Hypothesis: The volume of global food-grain production is inversely proportional to West Texas Intermediate crude oil prices.

Let r be the correlation coefficient, and n = 32 (annual samples)

Null hypothesis (H0): r = - 1.00

Alternate hypothesis (HA): r > - 1.00

The alternate hypothesis fails the test. The sample of 32 years of data with Weka yielded r = - 0.4734, but does not show a close inverse correlation, as shown by the screenshot below, which depicts a scatterplot (x = WTI, y = Volume) bounded by a polyline.



*Figure 2*. Screenshot of correlation between WTI crude prices and global food-grain production from 1986 to 2018. Adapted from author’s desktop Weka application.

I anticipate that both of these hypotheses will be refined and tested with more detailed data for the various model components in future work with arbitrarily selected significance levels and *p*-values while testing means, proportions, and categorical variables, as documented by Davies (pp. 388-420). Other hypotheses will be tested in future work; namely, that the world population grows at a constant rate as a result of global food production growing, since families feel food-secure to increase births. Also, production of individual food-grain crops varies by country, with the production of and demand for other crops, and with the availability of water.

**Datasets**

The prevailing research question during the project was, “If we can feasibly develop a visualized model and simulation of the global food chain with the Big Data already available, what data elements will the model and simulation contain, and what engaging story may we tell about data found? Also, what is an effective, methodological approach to so ambitious an undertaking “for making it doable” (O’Leary, 2017, p. 123)?

My intention in this project was to select macro-data that represents essential elements of food agriculture to depict essential elements of a global model and simulation (Jurist, 2018, p. 2). Since the total scope of such in-depth research is too broad for this course, the hypotheses for this Capstone Project were necessarily narrowed as data analysis advanced and research progressed. As Machi and McEvoy (2016) explain, a focused perspective is essential. “Choice of perspective depends on the subject chosen for study and the unit of analysis from which the researcher has chosen to study it” (p. 24). In this case the perspective was with business analytics applied to quantitative data for volumes and prices.

The search for data in this course project was limited to few food-crop commodities and their supporting models (energy, natural resources, weather and insolation, fertilizers and pesticides), and excluded livestock and non-foodstuffs. The food-crop commodities selected are wheat, corn, barley, rice, oats and soy, consolidated in one total. Public datasets are generated by the U.S. Department of Agriculture (2018b), and are available online. Other data was considered but not selected from the Chicago Mercantile Exchange (CME), the Food and Agriculture Organization (FAO), United Nations (UN), World Health Organization (WHO), the Energy Information Agency (EIA), and U.S. Department of Health Services (USDHS). Data files were deposited on the author’s GitHub site (Jurist, 2019, p. 1). Links to project data files:

https://inflationdata.com/articles/inflation-adjusted-prices/historical-crude-oil-prices-table/ (McMahon, 2019)

https://www.usda.gov/oce/commodity/wasde/ (United States Department of Agriculture, 2019)

https://www.nass.usda.gov/Data\_and\_Statistics/index.php (United States Department of Agriculture, 2018b)

https://nassgeodata.gmu.edu/CropScape/ (United States Department of Agriculture, 2018c)

https://www.nass.usda.gov/Charts\_and\_Maps/Agricultural\_Prices/index.php (United States Department of Agriculture, 2018d)

https://data.worldbank.org/indicator/sp.pop.totl?end=2018&start=1960 (World Bank, 2019).

These datasets were chosen in order to begin characterizing four essential components that contribute to the American agriculture industry’s share of the global food chain; namely, energy consumption and cogeneration, use of natural resources (water supplies and consumption [such as described by Bisson and Lehr (2004)], and land use), insolation and weather, and fertilizer and pesticide applications.

**How the Organization Can Benefit from the Dataset**

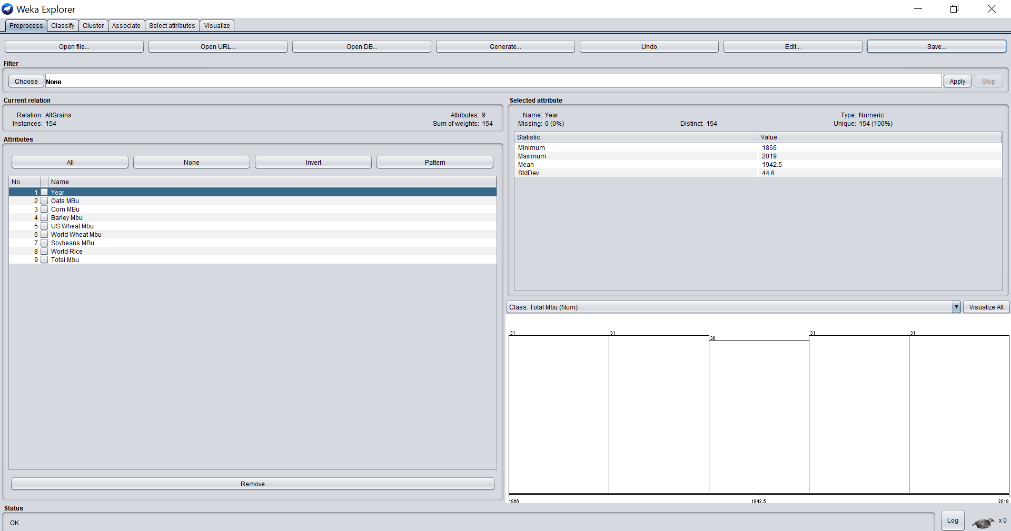
“The value of government data is difficult to measure, but it is clearly a substantial strategic asset for the US business sector. Such data are used . . . for purposes that include production and investment decisions, marketing and inventory management, and long-range strategic planning” (Hughes-Cromwick, & Coronado, 2019, p. 145). Visualized analytics of the datasets could partially depict the American components of the global food chain in order to solve problems and answer questions. Grave matters are at stake that analytics can elucidate for the sustainability and profitability of growers and processors, and for the health, nutrition, and longevity of consumers. “My audience is people, not only machines” (Cady, 2017, p. 122). This work will serve both institutions and individual growers, producers, processors, and marketers willing to accept its value proposition.

**Tools and Techniques**

Data was cleansed and prepared in Wordpad and Excel. Analytics were performed with Weka 3.9.3 because it is simple to use for flat file data, and some descriptive analytics with R CLI v. 3.4.4. **Tableau Desktop Public Edition (2018.2.3) is the preferred tool f**or visualizations, which appear on the Tableau Public repository (Jurist, 2019b)**.** No three-dimensional modeling was completed.

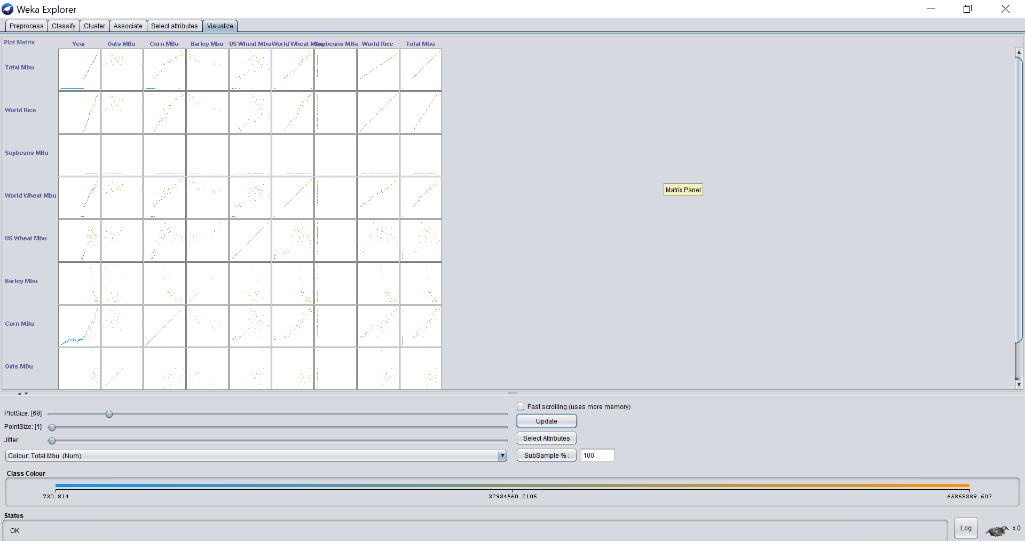
**Programming Code**

Algorithms in the standard library functions in Weka were executed. Following in Figure 3 is a screenshot of the Weka Explorer entrance to the analytics with a simple examination of the raw csv data file for the consolidation of all food crops introduced in this project.



*Figure 3.* Weka Explorer view of raw data file of all consolidated food crops. Adapted from the author’s desktop in Weka.

The next view in Figure 4 depicts a comprehensive visualization of all the crops included in the study all on one page in the Weka tool.



*Figure 4*. Weka Explorer view of time series for all consolidated food crops. Adapted from the author’s desktop in Weka.

The output shown above are examples of project analytics output. Screenshots and output have been uploaded to the author’s GitHub (Jurist, 2019) repository.

Script and output from R CLI 3.4.4 follows:

oilfoodcor.csv <- read.csv("C:\\R\\TotalGrainsOilPriceCorrelation.csv", header=T)

oilfoodcor.csv <- c(oil,food)

> data.frame(read.csv("C:\\R\\TotalGrainsOilPriceCorrelation.csv", header = F))

V1 V2 V3

1 1986 38283610 15.04

2 1987 38246511 19.17

3 1988 40342659 15.98

4 1989 41961892 19.64

5 1990 42706641 24.47

6 1991 42966041 21.50

7 1992 43051014 20.56

8 1993 43185791 18.45

9 1994 44327921 17.19

10 1995 44908172 18.44

11 1996 46299318 22.11

12 1997 47200256 20.61

13 1998 48153009 14.45

14 1999 49922748 19.26

15 2000 48699303 30.30

16 2001 48724380 25.95

17 2002 46146282 26.12

18 2003 47950857 31.12

19 2004 48934350 41.44

20 2005 51033841 56.49

21 2006 51276245 66.02

22 2007 52919097 72.32

23 2008 54967633 99.57

24 2009 53833779 61.65

25 2010 55199080 79.40

26 2011 57382300 94.87

27 2012 58140311 94.11

28 2013 58801863 97.91

29 2014 58927827 93.26

30 2015 58194929 48.69

31 2016 60021826 43.14

32 2017 60559256 50.88

33 2018 65868390 64.94

#Convert the data.frame into a simple variable name for easy use.

D <- data.frame(read.csv("C:\\R\\TotalGrainsOilPriceCorrelation.csv", header=T))

> nrow(D)

[1] 33

head(D)

YEAR FPROD WTI

1 1986 38283610 15.04

2 1987 38246511 19.17

3 1988 40342659 15.98

4 1989 41961892 19.64

5 1990 42706641 24.47

6 1991 42966041 21.50

# Calculate average global food production over 32 years (MBu).

> mean(D$FPROD)

[1] 49973852

# Calculate average price of West Texas Intermediate crude oil.

> mean(D$WTI)

[1] 43.78939

# Calculate correlation coefficient between WTI and food prod.

> cor(D$FPROD,D$WTI)

[1] 0.778973

**Findings**

The hypothesis test disproved the null hypothesis. More precise data about energy resources consumed solely by the agriculture industry and for specific crops will reveal better test results in the future. Much data is available for the United States, but little is available for other countries. Little has been analyzed and published by the USDA except with economic reports, but much data is available. Since food production is specifically a geographical activity, territorial and national mappings of the data will make for more revealing visualizations. Security and privacy are not prima facie a concern. USDA data and resources are driven largely toward the generation of data but not toward the analytics of it. The United States is the largest food producer in the world. Eventually as we discover the value of the data, we will find more ways to use it. International organizations are beginning to form standards for the growth and processing of crops and food products. Blockchain technology (Bedord, 2019) with IBM will enable new applications of technology for the decisions that will help feed the food insecure.

The initial null hypothesis was disproven by the data not demonstrating close inverse proportion. Also, following are detailed observations about the data and analytics.

1. Field food-crop commodities are grown as food for people, feed for livestock, and raw material for non-food manufacturing such as ethanol and packaging.
2. Corn is the most widely produced feed grain in the United States (95%).
3. Rice is the most abundantly produced food crop in the world, most of it being grown outside the United States and in Asia (China, Vietnam, India).
4. It is still necessary to acquire usage and consumption statistical data.
5. It is still necessary to acquire data for energy, natural resource consumption, fertilizer and pesticides.
6. It is still necessary to correlate United States and global population growth to the food production and consumption statistics to generate inferential statistical reasoning.
7. Data is organized and formatted differently in each raw data file, and must be reformatted for consistent comparisons and consolidations. This is a tedious, time-consuming task.
8. GeoDa can help with geographic dimensions, but Tableau will work better for this.
9. Must convert the data with metric tons of rice production into bushels with a conversion ratio of 49.992 bushels per metric ton.
10. Most time has been spent in the tedious tasks of data gathering, examination, cleansing, reformatting and preparation for the more fascinating work of data analytics.
11. Other dimensions besides time series data will suggest hypothetical causal correlations.

**Conclusion**

An interactive, predictive-prescriptive model of global food is feasible, but will require long-term construction of the fully integrated model with its contributory sub-models of energy, weather, natural resources, fertilizer and pesticide. “The value of government data is difficult to measure, but it is clearly a substantial strategic asset for the US business sector” (Hughes-Cromwick, & Coronado, 2019, p. 145). Ultimately the intention of these efforts will be to arrive at insights from descriptive, predictive and prescriptive analytics that extract that intrinsic value of United States government data for actionable business, financial, and agricultural decisions.

**Recommendations and Future Research**

It is recommended that this project be continued and expanded after the academic research project has concluded, and that with a dedicated, long-term, technical, wholistic, data-driven, visualized, prescriptive initiative it be undertaken toward modeling American and other international food interdependencies. More detailed models will be prepared with attention to causal relationships and the interdependencies of import-export and geographic data subsets. Such modeling with highly legible visualizations will serve those nations who are most food-insecure and the growers who lack highly sophisticated business data analytics to inform their decisions for investments, plantings, cultivations, harvests, distribution, and marketing.

References

Bedord, L. (February 6, 2019). Ibm helps agriculture understand, adopt blockchain. Retrieved from https://www.agriculture.com/technology/data/q-a-nigel-gopie-ibm

Bisson, R., & Lehr, J. (July 2004). *Modern groundwater exploration: Discovering new water resources in consolidated rocks using innovative hydrogeologic concepts, exploration, drilling, aquifer testing and management methods*. John Wiley & Sons, Inc. ISBN: 978-0-471-06460-2. 321 pages.

Buskirk, D. D., Schweihofer, J. P., Rowntree, J. E., Clarke, R. H., Grooms, D. L., & Foster, T. P. (2013). A traceability model for beef product origin within a local institutional value chain. *Journal of Agriculture, Food Systems, and Community Development, 3*(2), 33-43. Retrieved from https://csuglobal.idm.oclc.org/login?url=https://search-proquest-com.csuglobal.idm.oclc.org/docview/1326745845?accountid=38569

Cady, F. (2017). *The data science handbook*. Hoboken, New Jersey: Wiley & Sons, Inc.

Carman, J. C., Eleuterio, D. P., Gallaudet, T. C., Geernaert, G. L., Harr, P. A., Kaye, J. A., Uccellini, L. W. (2017). The national earth system prediction capability: Coordinating the giant. *Bulletin of the American Meteorological Society, 98*(2), 239-252. doi:http://dx.doi.org.csuglobal.idm.oclc.org/10.1175/BAMS-D-16-0002.1

Codex Alimentarius. (2018). International food standards. Available online at http://www.fao.org/fao-who-codexalimentarius/en/

Colorado State University. (2018). CSU water center. International water research. Available online at https://watercenter.colostate.edu/international/

Data.gov. (2019). U.S. Department of Agriculture. 1291 datasets found. Retrieved from https://catalog.data.gov/organization/usda-gov

Davies, T. (2016). *The book of R: a first course in programming and statistics*. San Francisco, California: No Starch Press, Inc. ISBN-10: 1-59327-651-6.

Davis, K. (2012). *Ethics of big data*. Sebastopol, CA: O'Reilly Media, Inc. ISBN: 9781449311797

Department of Health and Human Services. (2018). U.S. Food & Drug Administration. Food. Retrieved from https://www.fda.gov/Food/default.htm

Estabrook, B. (March 2019). From the fields to your table. Eating well. Des Moines, Iowa. Meredith Corporation. ISSN 1046\_1639.

Federal Reserve of Kansas City. (2018). Research and data. Agriculture and the economy. Retrieved from <https://www.kansascityfed.org/research/agriculture>

Food and Agriculture Organization of the United Nations. (2018). Retrieved from http://www.fao.org/home/en/

Fry, B. (2007). The seven stages of visualizing data. Visualizing data by Ben Fry. Chapter 1. Retrieved from https://www.oreilly.com/library/view/visualizing-data/9780596514556/ch01.html

Hughes-Cromwick, E., & Coronado, J. (2019). The value of US government data to US business decisions. *The Journal of Economic Perspectives, 33*(1), 131-146. doi:http://dx.doi.org.csuglobal.idm.oclc.org/10.1257/jep.33.1.131

IKANOW. (2014, February 10). 8 proven steps to starting a big data analytics project. Retrieved from http://www.ikanow.com/8-proven-steps-to-starting-a-big-data-analytics-project/

Jurist, D. (May 13, 2018). MIS500 module 8 portfolio project. Available online at CSU Global Online Course Archives.

Jurist, D. (2019). *GitHub. BusinessIntelligenceDataAnalyticsCapstone* [Data file repository]. Available online at https://github.com/HigherPurp0se?tab=repositories

Kane, G. C., Palmer, D., Nguyen-Phillips, A., Kiron, D., & Buckley, N. (2017). Achieving digital maturity. *MIT Sloan Management Review, 59*(1), n/a-0. Retrieved from https://csuglobal.idm.oclc.org/login?url=https://search-proquest-com.csuglobal.idm.oclc.org/docview/1950392650?accountid=38569

Kitchen, J., & Monaco, R. (2003). Real-time forecasting in practice. *Business Economics, 38*(4), 10-19. Retrieved from https://csuglobal.idm.oclc.org/login?url=https://search-proquest-com.csuglobal.idm.oclc.org/docview/199852502?accountid=38569

Kleppel, G. (2014). *The emergent agriculture: farming, sustainability and the return of the local economy*. British Columbia, Canada. New Society Publishers.

Leone, L. (2017). Addressing big data in EU and US agriculture: A legal focus. *European food and feed law review: EFFL, 12*(6), 507-518. Retrieved from https://csuglobal.idm.oclc.org/login?url=https://search-proquest-com.csuglobal.idm.oclc.org/docview/1979764247?accountid=38569

Longo, S., & York, R. (2008). Agricultural exports and the environment: A cross-national study of fertilizer and pesticide consumption. *Rural Sociology, 73*(1), 82-104. doi: http://dx.doi.org.csuglobal.idm.oclc.org/10.1526/003601108783575853

Machi, L., & MacEvoy, B. (2016). *The literature review – Six steps to success (3rd ed.)*. Thousand Oak, CA: Corwin – Sage Publishing. ISBN: 9781506336244

McMahon, T. (February 2019). Historical crude oil prices. Retrieved from https://inflationdata.com/articles/inflation-adjusted-prices/historical-crude-oil-prices-table/

Monsanto. (2018). Conceiving new growing tools. Data science. Retrieved from https://monsanto.com/innovations/data-science/

Ngugi, D., Annan, K., Mullen, J., & Bergstrom, J. (2013). Forecasting land use: An application of four models. *International Journal of Arts & Sciences, 6*(1), 637-649. Retrieved from https://csuglobal.idm.oclc.org/login?url=https://search-proquest-com.csuglobal.idm.oclc.org/docview/1419027011?accountid=38569

O’Leary, Z. (2017). *The essential guide to doing your research project (3rd ed.)*. Thousand Oaks, CA: Sage Publishing. ISBN: 9781473952089

Roth, N. E., Thorne, J. H., Johnston, R. A., Quinn, J. F., & McCoy, M. C. (2012). Modeling impacts to agricultural revenue and government service costs from urban growth. *Journal of Agriculture, Food Systems, and Community Development, 2*(4), 1-20. Retrieved from https://csuglobal.idm.oclc.org/login?url=https://search-proquest-com.csuglobal.idm.oclc.org/docview/1080968746?accountid=38569

Schimmelpfennig, D. E., & Norton, G. W. (2005). Assessing the value of economics research: the case of the bias in the consumer price index. *Contemporary Economic Policy, 23*(4), 625-635. doi:http://dx.doi.org.csuglobal.idm.oclc.org/10.1093/cep/byi048

U.S. Commodity Futures Trading Commission. (2018). Financial data. Retrieved from https://www.cftc.gov/MarketReports/financialfcmdata/index.htm

U.S. Department of Agriculture. (June 16, 2017). Report to congress. Retrieved from https://www.usda.gov/sites/default/files/documents/report-proposed-2017-reorg-usda-under-secretary-trade-foreign-ag-affairs.pdf

U.S. Department of Agriculture. (2018a). About the U.S. Department of Agriculture. Retrieved from https://www.usda.gov/our-agency/about-usda

U.S. Department of Agriculture. (2018b). Data. Retrieved from https://www.usda.gov/topics/data

U.S. Department of Agriculture. (2018c). Digital strategy at USDA. 1.2.0 Make open data, content, and web APIs the new default. Retrieved from https://www.usda.gov/digital-strategy

U.S. Department of Agriculture. (2019a). Budget. Retrieved from https://www.usda.gov/our-agency/about-usda/budget

United Nations. (2018). Department of economic and social Affairs. Statistics Division. Branches. U.N. trade statistics. Analytical tables. Retrieved from <https://unstats.un.org/unsd/trade/data/tables.asp>

United States Department of Agriculture. (2018a). Economic research service. Data products. Retrieved from https://www.ers.usda.gov/data-products

United States Department of Agriculture. (2018b). National agricultural statistics service. Retrieved from https://www.nass.usda.gov/Data\_and\_Statistics/index.php

United States Department of Agriculture. (2018c). National agricultural statistics service. CropScape - cropland data layer. Retrieved from https://nassgeodata.gmu.edu/CropScape/

United States Department of Agriculture. (2019b). National agricultural statistics service. Agricultural prices. Retrieved from https://www.nass.usda.gov/Charts\_and\_Maps/Agricultural\_Prices/index.php

United States Department of Agriculture. (May 10, 2019c). World agricultural supply and demand estimates. Office of the chief economist. Retrieved from https://www.usda.gov/oce/commodity/wasde/

U.S. Commodity Futures Trading Commission. (2018). Financial data. Retrieved from https://www.cftc.gov/MarketReports/financialfcmdata/index.htm

U.S. Department of Agriculture. (June 16, 2017). Report to congress. Retrieved from https://www.usda.gov/sites/default/files/documents/report-proposed-2017-reorg-usda-under-secretary-trade-foreign-ag-affairs.pdf

U.S. Department of Agriculture. (2018a). About the U.S. Department of Agriculture. Retrieved from https://www.usda.gov/our-agency/about-usda

U.S. Department of Agriculture. (2018b). Data. Retrieved from https://www.usda.gov/topics/data

U.S. Department of Agriculture. (2018c). Digital strategy at USDA. 1.2.0 Make open data, content, and web APIs the new default. Retrieved from https://www.usda.gov/digital-strategy

U.S. Department of Agriculture. (2019a). Budget. Retrieved from https://www.usda.gov/our-agency/about-usda/budget

United Nations. (2018). Department of economic and social Affairs. Statistics Division. Branches. U.N. trade statistics. Analytical tables. Retrieved from <https://unstats.un.org/unsd/trade/data/tables.asp>

United States Department of Agriculture. (2018a). Economic research service. Data products. Retrieved from https://www.ers.usda.gov/data-products

United States Department of Agriculture. (2018b). National agricultural statistics service. Retrieved from https://www.nass.usda.gov/Data\_and\_Statistics/index.php

United States Department of Agriculture. (2018c). National agricultural statistics service. CropScape - cropland data layer. Retrieved from https://nassgeodata.gmu.edu/CropScape/

United States Department of Agriculture. (2019b). National agricultural statistics service. Agricultural prices. Retrieved from https://www.nass.usda.gov/Charts\_and\_Maps/Agricultural\_Prices/index.php

United States Department of Agriculture. (May 10, 2019c). World agricultural supply and demand estimates. Office of the chief economist. Retrieved from https://www.usda.gov/oce/commodity/wasde/

Water & Wastewater International (WWI). (2011). *Revolutionary groundwater method reveals megawatersheds in Trinidad & Tobago*. Retrieved from http://www.waterworld.com/articles/wwi/print/volume-19/issue-7/regional-focus/revolutionary-groundwater-method-reveals-megawatersheds-in-trinidad-tobago.html

World Bank. (2019). Population, total. Retrieved from https://data.worldbank.org/indicator/sp.pop.totl?end=2018&start=1960

**Portfolio Project Outline**

* information about the organization (history, services and products, location, employee size, and revenue) and why you chose it;
* an explanation about what dataset you selected and why you chose this data;
* identifying your research null and alternative hypotheses;
* identifying the tools and techniques you used to analyze the data;
* addressing security, privacy, and ethical concerns when handling and analyzing the data;
* explaining your data analysis outcomes and whether you could support your null or alternative hypothesis; (if you did not find support for your alternative hypothesis, what conclusions can you draw about your data?);
* programming code, and screenshots of data analysis outcomes;
* uploading your project (code and documents, for example) to your GitHub account and include the link in your project write-up; and
* the write-up components submitted in Module 5.

Requirements:

#### Option #1:

#### ****Capstone Project****

#### ****Final Report and Slide Presentation:****

#### ****U.S. Organization****

For your Capstone Project, you will select a public dataset and identify two or three tools to analyze the data for an organization in the United States. The purpose is to show the organization how these tools can be used to analyze data in a project. You submitted your project write-up in Module 5 and a rough draft in Module 6. You will incorporate all feedback from your instructor and submit your Final Report, including a presentation with speaker notes.

Your Final Report should include:

* information about the organization (history, services and products, location, employee size, and revenue) and why you chose it;
* an explanation about what dataset you selected and why you chose this data;
* identifying your research null and alternative hypotheses;
* identifying the tools and techniques you used to analyze the data;
* addressing security, privacy, and ethical concerns when handling and analyzing the data;
* explaining your data analysis outcomes and whether you could support your null or alternative hypothesis; (if you did not find support for your alternative hypothesis, what conclusions can you draw about your data?);
* programming code, and screenshots of data analysis outcomes;
* uploading your project (code and documents, for example) to your GitHub account and include the link in your project write-up; and
* the write-up components submitted in Module 5.

Your Capstone Project Final Report should:

* include a reference page with at least five scholarly or peer-reviewed sources from the [CSU-Global Library (Links to an external site.)Links to an external site.](http://csuglobal.libguides.com/libraryhome);
* be 7-10 pages in length, not including the title and references pages; and
* be well-written and in conformity with the [CSU-Global Guide to Writing and APA (Links to an external site.)Links to an external site.](http://csuglobal.libguides.com/apacitations)

Your slide presentation (use PowerPoint or another software program) should:

* be 10-15 slides in length;
* include information about your final report and contain speaker notes as if you were presenting directly to upper management members within the organization; and
* include a reference slide with at least 5-10 scholarly or peer-reviewed sources from the [CSU-Global Library (Links to an external site.)Links to an external site.](http://csuglobal.libguides.com/libraryhome).

For your oral presentation do the following:

* Practice your speech several times in VoiceVibes to improve your delivery and presentation. Once you have improved your speech, you will record a video using your presentation slides and other materials in [(Links to an external site.)Links to an external site.](https://support.google.com/meet/answer/7557124?co=GENIE.Platform%3DDesktop&hl=en) or another video recording tool. You will then upload your speech to your YouTube channel.
* Make your presentation at least 10 minutes long.
* Complete and attach the self-evaluation form detailing your experience and thoughts for improving your presentation for the next speech, and submit it. Include in the form the link to your YouTube video.

The VoiceVibes and YouTube channel instruction sheets are in the Module 8 folder with the assignment. The self-evaluation form and public datasets file are also in the Module 8 folder with the assignment.