Choose Healthy Chocolate

Mason Chen
Stanford OHS, Palo Alto, CA, USA, mason05@ohs.stanford.edu

Abstract

Many people like eating chocolate, but may have some concerns on health risk, especially to people with Cardiovascular or Neurovascular diseases. The objectives of this paper are to use Multivariate Statistics to define a health biometric on choosing a healthy chocolate to patients with heart disease. Chocolate, made from cocoa beans, contains flavonoids which contain antioxidants. Flavonoids are the most abundant polyphenols in human diet. Polyphenols have antioxidant properties which can prevent aging and is also beneficial to heart disease and diabetes patients. People with heart diseases should eat less of saturated fat, trans fat, sodium, and cholesterol. They should eat more dietary fiber. Cocoa flavanols promote healthy blood flow circulation from head to toe. The heart, brain, and muscle depend on a healthy circulatory system. Data has been collected on 20+ chocolate ingredient contents from 60+ different types of chocolate. Multivariate correlation study has found that (1) strong negative correlation between Cocoa and Sugar, and (2) strong positive correlation between Diet Fiber and Iron. Most dark chocolate contains more cocoa, and less sugar. Dietary fiber and iron are high in correlation because of the high cocoa percent. The above two correlations can be further explained by conducting the Hierarchical Clustering Analysis on separating the Dark Chocolate, Milk Chocolate and White Chocolate. The Cocoa and Calcium are the deciding factors to separate these three Chocolates. Based on Chocolate Science. These healthiest chocolate can actually help prevent heart disease. The same approach can be applied to help people with other diseases (cancer, diabetes...). In Big Data World, the Multivariate Statistics can help connect different data and explain the Science in a predictive or/and empirical modeling.

Keywords: STEM, Flavonoids, Chocolate, Statistics, Antioxidant

1. Introduction and Research

Many people like chocolate, but most have some concerns that chocolate is unhealthy. Are they sure whether eating chocolate is unhealthy? The objectives of this paper are to find out if eating chocolate is unhealthy, especially what diseases can be prevented by eating chocolate, and why can chocolate prevent those diseases, what nutritions let chocolate prevent those diseases, and how we can select the best chocolate for preventing those diseases.

1.1 Adopt STEM Approach

Author has decided to deploy STEM (Science, Technology, Engineering, Mathematics) methodology. All 4 STEM elements are critical to making this project successful. The following is the breakdown of four STEM elements. The science studied was cocoa bean nutrition, flavonoids, flavanols, and antioxidants. Understanding this chemistry science and foods science is useful when defining the project objectives and scope. Technology is the chocolate manufacturing process to produce the chocolate from coca beans. Systematic engineering problem solving techniques such as 5 whys were deployed to understand the root cause analysis. JMP Statistics, Graphical Analytics, Data Mining tools are conducted to analyze the chocolate science and product complexity in order to establish a Chocolate Health Index.

1.2 Literature Research

Is chocolate really unhealthy? In Fig.1, the chocolate consumption and the median life expectancy was plotted side by side [1]. The countries on the left are the world's biggest chocolate consumers in 2015. The number on the right is the median life expectancy for those countries. If chocolate is unhealthy, then the median life expectancy will go down from top to bottom. However, there is no significant correlation, and therefore according to this chart, chocolate has not been proven unhealthy.

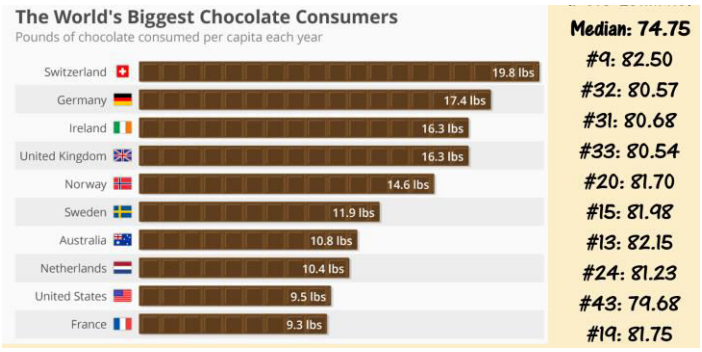

Figure 1. World's Biggest Chocolate Consumers vs. Life Expectancy

1.3 Study Chocolate Ingredients

Most chocolates are made from milk powder, sugar, cocoa butter (butter extraction from liquor), and cocoa liquor. Dark chocolate has more contents of cocoa liquor, cocoa butter and less of sugar. White chocolate has the opposite quantities. Milk chocolate composition is in the between Dark Chocolate and White Chocolate. Will these composition patterns indicate which chocolate type is healthier? Authors will like to study Chocolate Technology and Science in the following section.

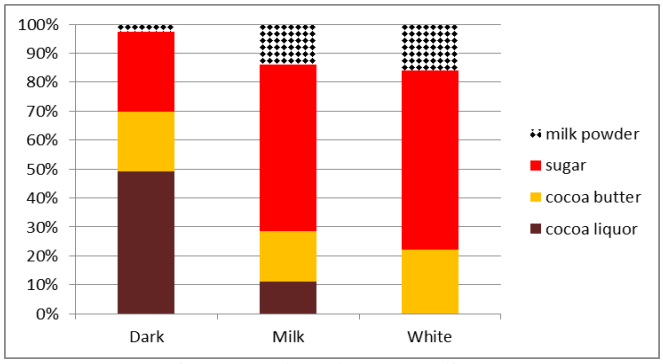

Figure 2. Chocolate Ingredients

2. Chocolate Technology and Science

This section will cover three subjects: (1) understand Chocolate Manufacturing Process and Technology, 2) Study Chocolate Foods Science, (3) systematic engineering problem solving, and (4) Chocolate Type Literature Research.

2.1 Chocolate Manufacturing Process and Technology

What chocolate manufacturing process is critical to determine which chocolate is healthy. In Fig. 3 chocolate process flow chart, Cacao pods are first collected from cacao trees. Then, cacao beans are eliminated of its pods and dried under the hot sun. Different type of cacao beans will decide what chocolate ingredient in rich. The dried beans are shipped to chocolate makers and washed thoroughly. Cacao beans are cooked in a certain high temperature and roasted to control certain chocolate attributes. Nibs, or the "fruit" of the cacao bean, are grinded into cocoa liquor. Cocoa butter is extracted from cocoa liquor. The remaining cocoa cake is grinded into cocoa powder. The cocoa liquor is added with cocoa butter, sugar, and other ingredients depending on the manufacturer and the chocolate type. The mixture is soothed under heating (conching) before put in molds, cooled, and shipped. Understand this chocolate manufacturing process will help us categorize chocolate types and their major ingredients as well as process control.

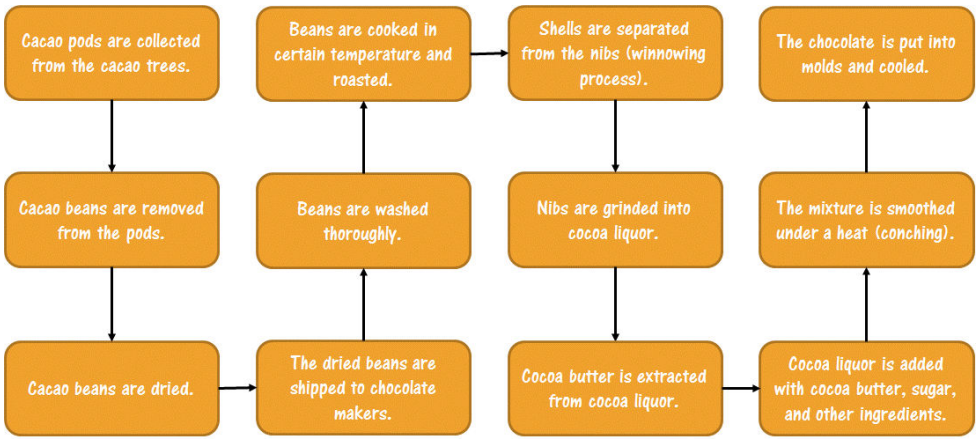

Figure 3. Chocolate Manufacturing Process [2]

2.2 Study Chocolate Chemistry and Science

Chocolate is actually a powerful source of antioxidant. Antioxidant prevents aging which is healthy for human body. Antioxidants are also essential to preventing heart disease since it increases blood flow. Apple and Blueberry are well known fruits with excellent antioxidant functions. In Fig.4, if chocolate's serving size is equal to apple, it has the most antioxidant amount among all the foods listed. The next objected interested is why Chocolate can have such rich antioxidant effects?

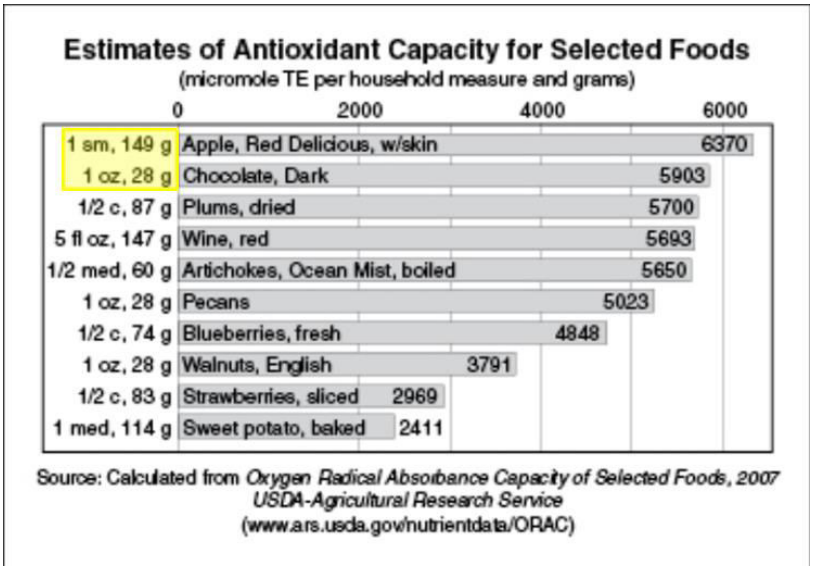

Figure 4. Estimates of Antioxidant Capacity

Chocolate, made from cocoa beans, contains flavonoids which contain rich antioxidants. Antioxidants prevent human aging and is also beneficial to heart disease and diabetes patients particularly. People with heart diseases should eat less of saturated fat, trans fat, sodium, cholesterol and eat more dietary fiber. It's critical to study Chocolate flavonoids chemistry and science to further understand the antioxidant benefit. Flavonoids are the most abundant polyphenols in human diet, representing 2/3 of those digested. Polyphenols are compounds found abundantly in natural food sources that have antioxidant properties. Flavonoids have the general structure of a 15-carbon skeleton: (1) consists of two phenyl rings (A and B) and heterocyclic ring (C), and (2) this carbon structure is abbreviated C6-C3-C6. There are seven different types of flavonoids (classified based on its chemical structure): Flavones, flavanol, flavanones, isoflavones, anthocyanidins, chalcones, catechins. Chocolate flavonoids are flavanols [3] (shown in Figure 5).

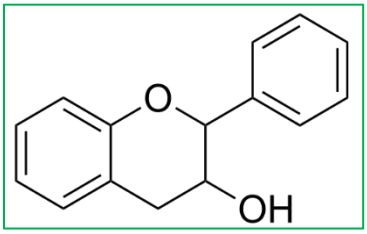

Figure 5. Flavanol Structure

Cocoa flavanols promote healthy blood flow from head to toe. The heart, brain, and muscle depend on a healthy circulatory system seen in Fig.6. Supporting healthy blood flow is essential to helping maintain exceptional health throughout life [4]. Flavanol benefits including living longer, weight control, and cardiovascular disease, cancer, diabetes, and neurodegenerative disease prevention [5,6,7].

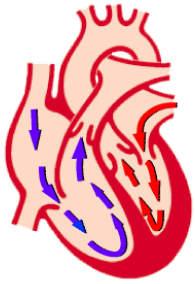
Figure 6. Healthy Blood Flow

2.3 Systematic Engineering Problem Solving

After gone through the literature research, chocolate manufacturing process and chocolate chemistry and foods science, authors have further conducted two engineering problem solving techniques: (1) 5-Whys analysis, and (2) SIPOC analysis in order to define the project scope for this paper.

5-Whys Analysis

1st Question: is chocolate unhealthy for everybody?
1st Answer: not really, chocolate may prevent some diseases.
2nd Question: which diseases can be prevented from chocolate?
2nd Answer: heart disease and diabetes.
3rd Question: why can chocolate prevent these diseases?
3rd Answer: because it has rich antioxidant.
4th Question: why does chocolate have antioxidant?
4th Answer: because cocoa beans have a lot of flavonoids which contains antioxidant.
5th Question: how can we select chocolate for patients with heart disease?
5th Answer: create a health index for choosing healthy chocolate products.

SIPOC analysis includes five elements: S(Supplier), I(Input), P(Process), O(Output), and C(Customer). SIPOC is a customer-driven team-building analysis which can help define the project scope and provide guidelines through project deployment. The main customers are people with heart diseases. People with heart diseases want chocolates with nutrition that will help stop their diseases. The key output deliverable is building a healthy index based on the input Chocolate Type, Technology, Chemistry and Science. The process is to use JMP Statistics software to build a transfer function which can predict the chocolate health index. Based on the above systematic 5-whys and SIPOC problem solving techniques, the main objective and scope of this paper will be building a reliable healthy index for

choosing healthy chocolates for patients with heart disease particularly.

2.4 Literature Research of Chocolate Product Types

In order to build a reliable health index, authors have further conducted literature research on Chocolate Product Types. There are main three commercial chocolate product types available in most stores: (1) Dark Chocolate, (2) Milk Chocolate, and (3) White Chocolate. Can the healthy index be based on the Chocolate Product Type? The characteristics of dark chocolate are plenty amounts of soluble fiber, rich minerals (iron, magnesium, copper, manganese, potassium, phosphorus, zinc, selenium), powerful source of antioxidant, improve blood flow and lower blood pressure, increases HDL (good cholesterol) and decreases LDL (bad cholesterol), lower risk of cardiovascular disease, and improve brain function [8]. The side risks are that dark chocolate may cause migraines, kidney stones, and caffeine. The characteristics of milk chocolate are some of dark chocolate plus calcium, heart healthy, boosts brain functions, slows signs of aging, fights colds, stops tooth decay, lowers blood pressure, and reduces stress [9]. The main concerns are lots of sugar, and some side effects like caffeine [10,11]. The characteristics of white chocolate are rich calcium, prevents hypertension and heart failure, increases blood flow, maintains cholesterol level, and reduces breast cancer [12,13]. The problems are the enormous amounts of sugar, obesity, and diabetes [14].

3 Graphical Analytics and Multivariate Statistical Analysis

This section 3 will build the statistical modeling in order to define the Health Index in section 4.

3.1 Data Collection Plan and Transform Data

It's critical to collect the right data regarding the Chocolate health information. Target Store was chosen as the main chocolate retailer since it had plenty chocolate products (enough sample size) and it was extremely convenient for collecting data. Figure 7 shows one sample of chocolate nutrition fact data that was collected from the Target Store. 60+ different types of chocolates were collected, and each had 20 variables. Ensuring good data quality is critical to screen out noise data. After certain data filtering process (not all information is relevant to define the health index), 8 variables were selected shown in Figure 7.

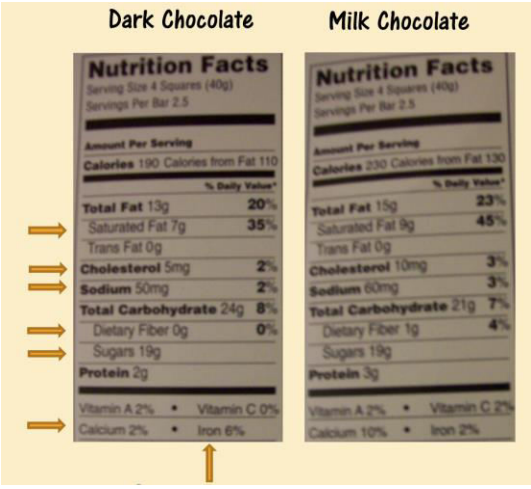

Figure 7. Variables that were chosen to be analyzed

The raw data collected have different distributions (sample mean and sample standard deviation). In order to eliminate the bias of the central tendency and spread, the raw data was transformed to become Z- Standardized Data as shown in Figure 8. After Z-transformation, all variables have new sampled distributions of mean at 0 and standard deviation of 1. The objective of this transformation is to eliminate any larger variation bias in building the statistical modeling of deriving the chocolate health index.

Chocolate%_1	Chocolate%_1	Calorie/Size_1	CalFat/Size_1	Tfat/Size_1	Sfat/Size_1	Cholest/Size_1	Sodium/Size_1	Carbs/Size_1
1.46313048	1.46313048	0.88587249	1.741211934	1.965213277	1.427055953	-1.112379727	-0.72659527	-2.484040893
1.208673005	1.208673005	0.518015041	1.176780474	0.909641981	0.750357618	-1.112379727	-0.573353644	-1.703445541
1.2595645	1.2595645	0.231903692	1.897998451	2.023856131	1.539839005	-1.112379727	-1.033078522	-2.137109629
0.547083571	0.547083571	0.88587249	1.14542317	1.085570534	0.33681974	-1.112379727	-1.033078522	-1.44324709
0.547083571	0.547083571	-0.422065106	0.39284789	0.381856333	0.33681974	-1.112379727	-1.033078522	-0.286809533
0.445300581	0.445300581	1.253729938	1.176780474	1.173534805	1.088706786	-1.112379727	-0.72659527	-1.18304864
0.547083571	0.547083571	-0.198338938	0.835150906	0.881863788	0.590086963	-1.112379727	-1.033078522	-0.9502395
0.445300581	0.445300581	0.518015041	1.176780474	1.173534805	1.088706786	-1.112379727	1.418787496	-0.922850189
-0.063614369	-0.063614369	1.458095188	1.709854631	1.73064188	1.915782527	-0.109165248	-0.692541575	-1.269781453
		-0.972775673	-0.947264232	-0.784827734	-0.834541114	0.075637418	1.547833076	0.967012239
		-1.076033903	-0.610585817	-0.673714964	-0.715822105	-1.112379727	0.192854487	0.638340516
		-1.076033903	-0.610585817	-0.673714964	-0.715822105	-1.112379727	0.192854487	0.638340516
		-1.076033903	-0.986873458	-1.025572059	-0.715822105	-1.112379727	-1.033078522	1.332203055
		1.866825687	1.709854631	1.73064188	1.915782527	0.141638375	0.499337739	-1.269781453
		-1.076033903	-0.986873458	-1.025572059	-1.166954338	-1.112379727	-1.033078522	1.332203055
-0.063614369	-0.063614369	4.09626477	-0.029050373	-0.001987771	0.104418294	-1.112379727	-1.033078522	0.0706348
		1.458095188	1.709854631	1.73064188	1.915782527	-0.109165248	-0.692541575	-1.269781453
0.547083571	0.547083571	0.612118109	0.579533231	0.578241688	0.175950897	-1.112379727	2.530680224	-0.765520894
-0.063614369	-0.063614369	-0.072267843	-0.208045551	-0.403685101	-0.138792515	-1.112379727	1.105176727	0.20265939
0.954215531	0.954215531	0.88587249	1.176780474	1.173534805	1.088706786	-1.112379727	-1.033078522	-1.703445541
		1.458095188	1.396281597	1.73064188	1.915782527	-1.112379727	-0.692541575	-1.269781453
		1.458095188	1.709854631	1.73064188	1.915782527	0.141638375	-0.692541575	-1.269781453
0.445300581	0.445300581	0.518015041	0.894564744	1.173534805	0.750357618	-0.209486695	1.725270748	-0.922850189
		-1.076033903	-1.363161098	-1.025572059	-0.715822105	-1.112379727	-1.033078522	1.332203055
		1.049364689	1.082708564	1.144213378	1.539839005	0.141638375	1.520948579	-0.980672071

Figure 8. Subset of Z- Standardized Data

3.2 Graphical Analytics

After raw data been Z-transformed, JMP interactive graphical analysis was conducted in order to uncover the comprehensive chocolate nutrition distributions. The objective is to find any interesting patterns of how the chocolate producers made of today's chocolate product types in order to provide us a better insight information regarding the chocolate health index. The first Dark Chocolate sampled distribution (Figure 9) was conducted of the eight variables and chocolate type to compare the different chocolate types' nutrition facts.

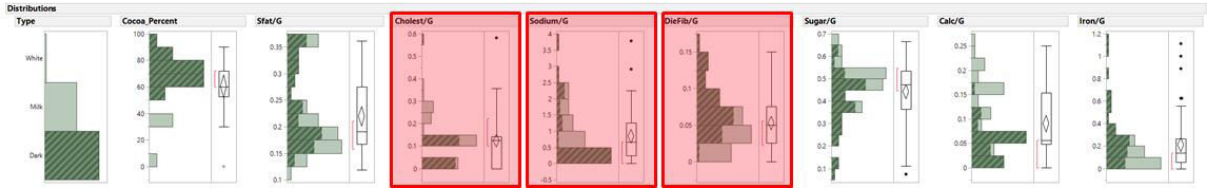

Figure 9. Dark Chocolate Distribution

After looking at the interactive graphical mode of nutrition distribution (only Dark Chocolate was selected), some interesting correlations were found. Dark chocolates mostly in common have low cholesterol, low sodium, and high dietary fiber. This helps prove that the hypothesis (dark chocolate is healthier than milk chocolate) may be correct. Milk chocolate (in Figure 10), on the other hand, does not show any significant correlation patterns among the variables analyzed. Most sampled distributions are near random (white noise). This observation may indicate there is no health requirement on formulating the chocolate nutrition ingredients for the milk chocolate. The distribution contrast between Dark Chocolate and Milk Chocolate has provided the first-hand information on how to derive the Chocolate Health Index.

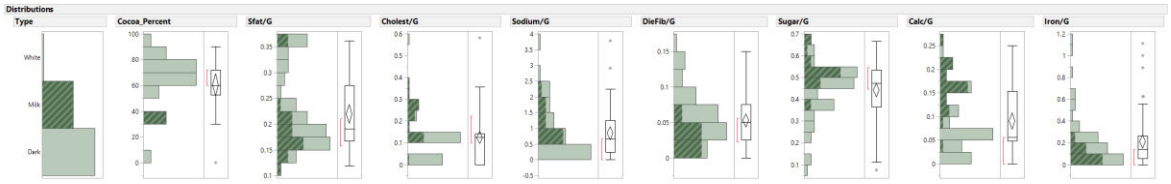

Figure 10. Milk Chocolate Distribution

The objective of this paper is to study how the Chocolate manufacturers chose healthier nutrition facts for

particular healthier chocolate types. A JMP multivariate correlation study [15] shown in Figure 11 and Figure 12 was further done to see if any chocolate type has strong correlation(s) between healthier nutritions and/or negative correlations between unhealthy nutritions. Correlations between <-0.75 and >0.75 threshold were set to identify any nutrition correlation pattern.

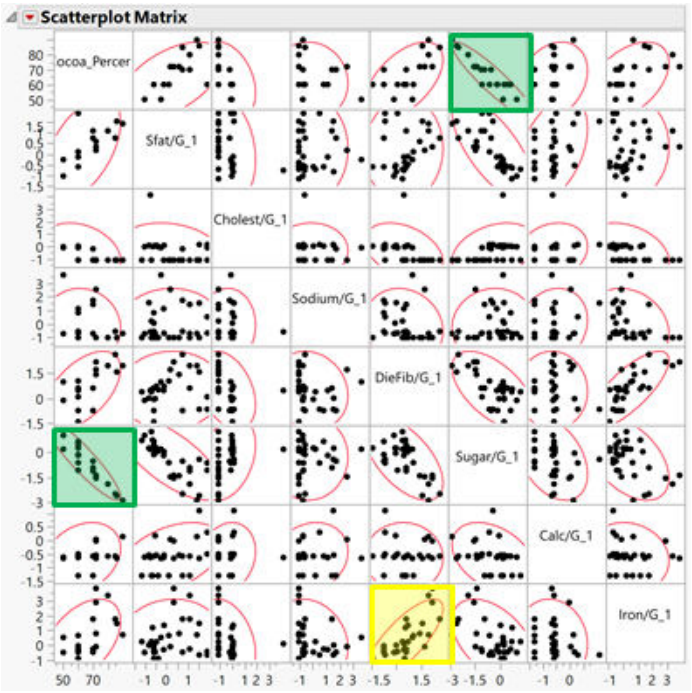

Figure 11. Scatterplot Matrix

Correlations

	Cocoa_Percent	Sfat/G_1	Cholest/G_1	Sodium/G_1	DieFib/G_1	Sugar/G_1	Calc/G_1	Iron/G_1
Cocoa_Percent	1.0000	0.5291	-0.3114	-0.0583	0.5482	-0.9162	0.2625	0.4597
Sfat/G_1	0.5291	1.0000	-0.1980	0.0184	0.0341	-0.7068	0.4161	0.0687
Cholest/G_1	-0.3114	-0.1980	1.0000	0.0302	-0.3666	0.3333	0.1732	-0.3304
Sodium/G_1	-0.0583	0.0184	0.0302	1.0000	-0.1344	0.0462	0.1667	-0.1862
DieFib/G_1	0.5482	0.0341	-0.3666	-0.1344	1.0000	-0.5804	-0.0207	0.7722
Sugar/G_1	-0.9162	-0.7068	0.3333	0.0462	-0.5804	1.0000	-0.3696	-0.4669
Calc/G_1	0.2625	0.4161	0.1732	0.1667	-0.0207	-0.3696	1.0000	-0.1037
Iron/G_1	0.4597	0.0687	-0.3304	-0.1862	0.7722	-0.4669	-0.1037	1.0000

Figure 12. Multivariate Correlations Table

The scatterplot matrix shows a visual diagram of the nutrition correlations among commercial chocolate products. The straighter and more diagonal the line, the stronger the correlation is between any two nutrition variables. Sugar and cocoa pair has a strong negative correlation of -0.9162. This shows that the higher the chocolate percent is, the lower the sugar percent. Since dark chocolate has a high chocolate percent, it also has low sugar. This indirectly indicates that dark chocolate is healthier with higher cocoa percent and lower sugar percent. The other identified strong positive correlation is between dietary fiber and iron. One science research has shown in Figure 11 that most dark chocolate products with 70%-85% are rich in Fiber and Iron Recommended Daily Allowance (RDA). Dietary fiber and iron are high in positive correlation because of the dark chocolate's high cocoa percent. Both graphical analyses have further provided why dark chocolate is healthier due to certain skewed nutrition preference.

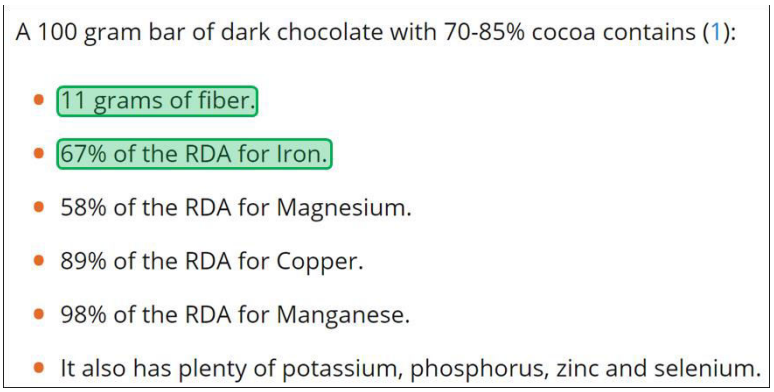

Figure 13. Iron and Dietary Fiber Correlation Research [11]

4. Results and Conclusions

This section will analyze the chocolate health index distribution and understand on how to choose better chocolate products, especially for patient with heart disease. The top four healthiest chocolates for heart disease were found in this paper. JMP software tools such as cluster analysis, correlation analysis, and distribution analysis were all applied to this project. Cocoa science, such as cocoa production, flavonoids, antioxidants, flavanol benefits, and the different types of chocolate, was learned throughout this paper. STEM approach is applied effectively to define the project scope by taking systematic scientific literature and engineering problem solving techniques. Further research may consider different health indexes for other diseases (cancer, diabetes, etc). The STEM approach could be applied more in daily life. The other opportunity is that antioxidants and flavonoid science will be further researched and proven through STEM framework.

References

1. Louis, P F. “Four Research-Proven Reasons to Eat Dark Chocolate for Your Health. ”Natural News, Science Daily, 27 Aug. 2013
2. www.health.harvard.edu/blog/sweet-dreams-eating-chocolate-prevents-heart-disease-201506168087.
3. www.cocoavia.com/why-it-works/cocoa-flavanol-benefits.
4. www.consumerlab.com/answers/is-dark-chocolate-or-cocoa-a-good-source-of-iron/dark_chocolate_cocoa_iron/.
5. www.prevention.com/food/healthy-eating-tips/dark-chocolate-and-antioxidants-0.
6. www.killerchocolate.com/articles/chocolate_antioxidants_health.php.
7. www.tuscany-diet.net/2014/01/22/flavonoids-definition-structure-classification/.
8. www.healthline.com/nutrition/7-health-benefits-dark-chocolate.
9. www.livestrong.com/article/354232-health-risks-of-dark-chocolate/.
10. www.onegreenplanet.org/natural-health/problems-with-chocolate-milk-campaigns/.
11. perfecthealthdiet.com/2014/03/white-chocolate/.
12. drhealthbenefits.com/food-bevarages/snack/health-benefits-white-chocolate.
13. www.benefitandharm.com/products/397-the-benefits-and-harms-of-white-chocolate.html.
14. infographic.statista.com/normal/chartoftheday_3668_the_worlds_biggest_chocolate_consumers_n.jpg.
15. http://www.jmp.com/support/help/Correlations_and_Multivariate_Techniques.shtml
16. Michael R. Anderberg. Cluster analysis for applications. Academic Press, New York, 1973. ISBN 0120576503.