"STEAMS" Methodology of Conducting Chocolate Science Research

Mason Chen[1]
[1]Stanford University OHS, Palo Alto, CA, USA, mason05@ohs.stanford.edu

Abstract
This paper adopts STEAMS (Science, Technology, Engineering, Artificial Intelligence, Math, Statistics) methodology. The objectives of this paper are to introduce the benefits of integrating all 6 "STEAMS" elements, especially living in the Big Data World. Chocolate Science Research case study was demonstrated to present this novel "STEAMS" concept as compared to current "STEM" or" STEAM" approach. There are three core visions of this "STEAMS" methodology: (1) replace "Art" with "Artificial Intelligence", (2) separate "Statistics" from "Math", and (3) integrate all six "STEAMS" elements. Adding the "Artificial Intelligence" element can trigger and enhance the effectiveness of "Scientific" Research and "Math" algorithms. Separating the "Statistics" element can conduct more effective risk management and draw practical conclusions. Due to the previous two benefits, integrating all 6 "STEAMS" elements is becoming a natural critical thinking way for most scientists and engineers striving in the modern Big Data era. It's critical and urgent for educators and teachers to migrate from their traditional STEM approach to the new "STEAMS" approach to educate our next generations in their early school learning and career development.

Key Words: Science, Technology, Engineering, Artificial Intelligence, Math, Statistics

1. Introduction

"STEM" (Science, Technology, Engineering, Math) or "STEAM" (Science, Technology, Engineering, Art, Math) are popular in School Education is a term used to group together these academic disciplines [1,2]. This term is typically used when addressing education policy and curriculum choices in schools to improve competitiveness in science and technology development. It has implications for workforce development, national security concerns and immigration policy. The acronym came into common use shortly after an interagency meeting on science education held at the US National Science Foundation. In the early 1990's, a summer program called STEM Institute is arranged for talented under-represented students in the Washington, DC area. Based on the program's recognized success and expertise in STEM education [3], that NSF was first introduced to the acronym STEM.

1.1 Criticism of STEM

The focus on increasing participation in STEM fields has attracted many criticisms:

(1) the efforts of the U.S. government to increase the number of STEM graduates, the science and engineering occupations have been flat or slow-growing, and unemployment as high or higher than in many comparably-skilled occupations [4].

(2) The STEM Crisis Is a Myth": there was a "mismatch between earning a STEM degree and having a STEM job in the United States, with only around ¼ of STEM graduates working in STEM fields, while less than half of workers in STEM fields have a STEM degree [5].

(3) Based on the data, science should not be grouped with the other three STEM categories, because, while the other three generally result in high-paying jobs, "many sciences, particularly the life sciences, pay below the overall median for recent college graduates [6].

(4) Efforts to remedy the perceived domination of STEM subjects has led to intense efforts to diversify the STEM workforce. Some critics feel that this practice in higher education, as opposed to a strict meritocracy, causes lower academic standards [7].

1.2 STEAM vs. STEM

STEAM fields are Science, Technology, Engineering, Art[8] , and Math [9]. STEAM is designed to integrate STEM subjects into various relevant education disciplines. These programs aim to teach students innovation, to think critically and use engineering or technology in imaginative designs or creative approaches to real-world problems while building on students' mathematics and science base [10]. STEAM programs add art to STEM curriculum by drawing on design principles and encouraging creative solutions.[11-14].

1.3 Artificial Intelligence and Digital Art

In the modern Big Data Society, Artificial intelligence (AI) is becoming a new dominant Data Science. AI sometimes called machine intelligence or machine learning, is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans. In computer science AI research is defined as the study of "intelligent agents": any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals [15]. Artificial intelligence was founded as an academic discipline in 1956, and in the years since has experienced several waves of optimism.

Digital art is an artistic work or practice that uses digital technology as an essential part of the creative or presentation process. Since the 1970s, various names have been used to describe the process, including computer art and multimedia art [16-17]. Recently, a lot of articles are about AI in art and design: a lot of the feature imagery, unsurprising as they're largely created by people from a science rather than arts background. And their research is often presented in the ultra-detailed format of scientific papers - heavy on words and, strangely to us. This digital art application of AI is being referred to as 'generative'. Generative AI will drive the next generation of apps for auto-programming, content development, visual arts, and other creative, design, and engineering activities. By 2019, most leading AI providers will offer tools and libraries for building AI-powered natural-language generation, image manipulation, and other generative use cases. Already, generative AI has proven itself - in both research and in commercial applications. In Generative graphics**:** AI can abstract visual patterns from artwork and then apply those patterns in the fanciful re-rendering of photographic images with the hallmark features of that artwork [18].

2. STEAMS

Instead of using classical STEM or STEAM, here, a new holistic “STEAMS” methodology is introduced. There are several novel concepts embedded in this new “STEAMS” methodology: (1) replace “Art” with “Artificial Intelligence”, (2) separate “Statistics” from “Math”, and (3) integrate all six “STEAMS” elements. “STEAMS” (Science, Technology, Engineering, Artificial Intelligence, Math, Statistics) methodology will be demonstrated through a Chocolate project. The chemistry “Science” studied was cocoa bean nutrition, flavonoids, flavanols, and antioxidants. “Technology” is the chocolate manufacturing process to produce the commercial chocolate products from coca beans. Systematic “Engineering” problem solving techniques such as 5 whys and SIPOCs were deployed to understand the root cause analysis. “Artificial Intelligence” clustering algorithms were utilized to recognize the clustering patterns hidden among Chocolate nutritions and products. Clustering Distance “Math” was utilized to explain the clustering algorithm. Distribution "Statistics" was conducted to visualize the effect of using "Artificial Intelligence" and "Math" algorithms. All 6 STEAMS elements are critical to making this Chocolate research project successful. The following sections will provide more details for each "STEAMS" element.

2.1 Understand Chocolate Anti-Oxidant "Science"

Chocolate contains flavonoids and antioxidants which can prevent aging and beneficial to heart disease and diabetes patients [19-21]. Chocolate is a powerful source of antioxidant which prevents human body aging/heart disease since it increases blood flow. Apple and blueberry are well known fruits with rich amounts of antioxidant. If chocolate's serving size is equal to apple, Chocolate contains even more antioxidant amount. Based on research [22-27], cardiovascular heart disease (CHD) risk is lower if taking more than 3 Chocolate servings per week (1 serving = 30 g). A typical cardiovascular disease is Atrial Fibrillation (AF). Antioxidants can prevent heart disease is because it reduces free radical formation. Free radicals are atoms with an odd number of electrons. When radicals form, they become highly reactive which causes cells to function poorly or die [28]. Excess free radicals initiate Cardiovascular disease (CVD) by damaging blood vessel. Bad cholesterol, Low-Density Lipoproteins (LDL), can also and only cause CVD after the oxidation of free radicals [29]. The oxidized components attract macrophages which absorb & deposit cholesterol [30]. Section 2.1 "Science" won't deviate much from "STEM" to "STEAMS".

2.2 Chocolate Product "Technology"

There are three main commercial chocolate product types available in most stores: dark, milk, and white. 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 High Density High-Density Lipoproteins (HDL, good cholesterol) and decreases Low-Density Lipoproteins (LDL, bad cholesterol), lower risk of cardiovascular disease, and improves brain function [31-32]. 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. The main concerns are lots of sugar, and some side effects like caffeine. The characteristics of white chocolate are rich calcium, prevents hypertension and heart failure, increases blood flow, maintains cholesterol level, and reduces breast cancer. The problems are the enormous amounts of sugar, obesity, and diabetes. Commercial chocolate production process flow chart is shown in Figure 1.

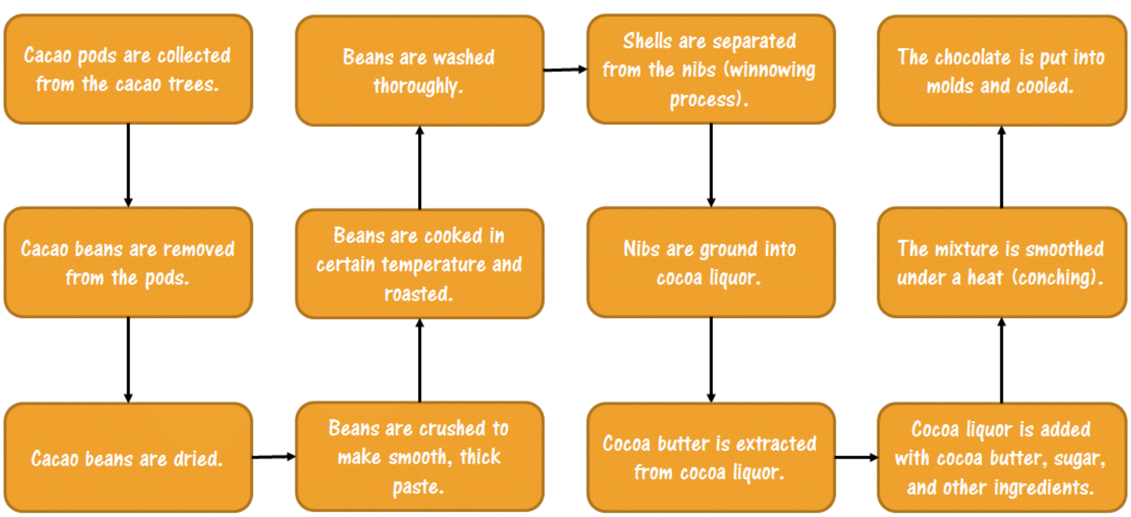

Figure 1: Chocolate Production Process Flow Chart

2.3 Systematic Engineering Problem Solving

Authors have previously done the multivariate statistics on understanding the Chocolate Science [33-35]. Understanding the Chocolate Product and Technology will help define the project scope and facilitate the remaining "Engineering, "Artificial Intelligence", "Math", and "Statistics" modules. 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. The Engineering techniques are not limited to the two choices here.

"5-Whys" Analysis

1^{st} Question: is chocolate unhealthy for everybody?
1^{st} Answer: not really, chocolate may prevent some diseases.

2^{nd} Question: which diseases can be prevented from chocolate?
2^{nd} Answer: heart disease and diabetes.
3^{rd} Question: why can chocolate prevent these diseases?
3^{rd} Answer: because it has rich antioxidant.

4^{th} Question: why does chocolate have antioxidant?
4^{th} Answer: because cocoa beans have a lot of flavonoids which contains antioxidant.

5^{th} Question: how can we select chocolate for patients with heart disease?
5^{th} Answer: create a health index for choosing healthy chocolate products.

"5-whys" analysis can clarify the project scope and explore first-level engineering root cause analysis. SIPOC analysis as shown in Figure 2 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.

Supplier	Input	Process	Output	Customer
JMP (SAS)	JMP Statistics Version 13 Software	Measure and Analyze Phase	Chocolate that has the most antioxidants	People with heart diseases
Target	Chocolate nutrition facts		Healthy choice	

Figure 2: SIPOC Engineering Analysis of Chocolate Science and Technology

2.4 Artificial Intelligence

In order to analyze Chocolate Science and Nutrition pattern, the "Variable Clustering" method [36] is used for grouping similar nutrition variables into representative clusters which are a linear combination of all variables in the same cluster. The cluster can be represented most by the variables identified to be the most representative members (higher R-Square with own cluster in Figure 3). The most representative variable in the cluster can be used to explain most of the variation in the data analyzed. Typically, dimension reduction using Cluster Variables is often more interpretable than dimension reduction using principal components. Based on JMP Clustering Variable analysis, four clusters are identified as shown in Figure 1:

Clustering Variables method can effectively explore the Chocolate nutrition clustering patterns which can explain the common foods science well. Adopting this dimension-reduction clustering algorithm can help simplify the predictive modeling by enhancing the signal-noise ratio, particularly in a very complicated/coupled design or system behavior. The "Artificial Intelligence" element can discover the patterns which may match well with traditional "Scientific" research and help "Engineering" critical thinking more effectively and efficiently. This "A" element is becoming a core and critical element in the new "STEAMS" approach.

Cluster Members

Cluster	Members	RSquare with Own Cluster	RSquare with Next Closest	1-RSquare Ratio
1	Calories (g)	0.789	0.314	0.308
1	Calories_from_Fat (g)	0.976	0.456	0.044
1	Total_Fat (g)	0.977	0.426	0.04
1	Saturated_Fat (g)	0.935	0.361	0.101
2	Cocoa_Percent	0.742	0.366	0.406
2	Cholesterol (mg)	0.811	0.387	0.309
2	Vitamin_A	0.505	0.126	0.566
2	Vitamin_C	0.412	0.016	0.598
2	Calcium	0.726	0.079	0.297
3	Sodium (mg)	0.345	0.013	0.664
3	Carbs (g)	0.876	0.185	0.152
3	Sugar (g)	0.874	0.416	0.216
4	Dietary_Fiber (g)	0.888	0.403	0.187
4	Protein (g)	0.73	0.358	0.421
4	Iron	0.803	0.269	0.269

Figure 3: Cluster Chocolate Nutrition Variables

2.5 Math

The Section 2.4 "Artificial Intelligence" clustering patterns were identified based on clustering distance math of calculating the dissimilarity of nutritions among chocolate products. There are several cluster math algorithms: (1) Average, (2) Centroid, (3) Ward, (4) Single, and (5) Complete (Citation). Will these 5 different clustering algorithms have the same results? If any difference, how to select which algorithm to explore the clustering patterns best?

In Figure 4, three existing clusters (Green, Yellow, Red) are going to join next. Which two clusters should bond first? The joining sequence is determined by the clustering distance algorithms. Centroid, Single, and Complete algorithms are compared. The Centroid algorithm connects Green cluster and Yellow cluster through the purple line connecting the two cluster means (purple triangles). The Single algorithm groups Green cluster and Red cluster by the closest points between these two clusters. The Complete algorithm groups Yellow cluster and Green cluster by the farthest points between these two clusters. Depending on which distance algorithm chosen, the clustering sequence and pattern may be different. We must dive into the mathematical calculations for each clustering distance algorithm and understand the benefits and limitations of each math algorithm in order to choose the best algorithm to draw reliable clustering patterns and results.

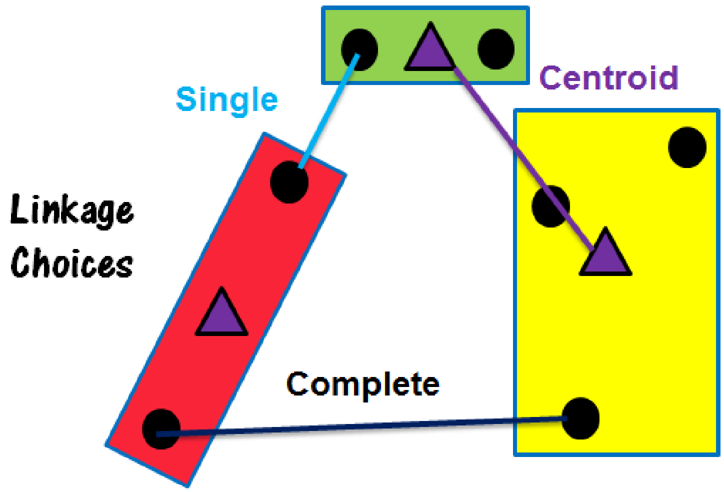

Figure 4: Diagram of the Centroid, Single, and Complete Clustering Methods

Here will compare five major clustering distance algorithms [37-43]. The calculations of the five different clustering algorithms are shown in Figure 5. The basic Algebra "Math" algorithms are utilized in the modern "Artificial Intelligence" clustering method. In the new "STEAMS" approach, "Math" is the foundation of evolving the modern "Artificial Intelligence". Adding the "Artificial Intelligence" element can trigger the exploring of the Math elements and Science research.

Average Linkage Distance for the average linkage cluster method is:

$$D_{KL} = \sum_{i \in C_K} \sum_{j \in C_L} \frac{d(x_i, x_j)}{N_K N_L}$$

Average

Centroid Method Distance for the centroid method of clustering is:

$$D_{KL} = \left\| \overline{x}_K - \overline{x}_L \right\|^2$$

Ward's Distance for Ward's method is:

$$D_{KL} = \frac{\left\| \overline{x}_K - \overline{x}_L \right\|^2}{\frac{1}{N_K} + \frac{1}{N_L}}$$

ANOVA

Single Linkage Distance for the single linkage cluster method is:

$$D_{KL} = \min_{i \in C_K} \min_{j \in C_L} d(x_i, x_j)$$

Minimum

Complete Linkage Distance for the Complete linkage cluster method is:

$$D_{KL} = \max_{i \in C_K} \max_{j \in C_L} d(x_i, x_j)$$

Figure 5: JMP Clustering Distance Algorithms [25]

2.6 Statistics

How will these algorithms impact the clustering patterns? The statistical distribution analysis was conducted for comparing two "Artificial Intelligence" clustering "Math": (1) Ward algorithm, and (2) Single algorithm. In Figure 6, only the first Ward cluster was selected which happens to be the healthier half of the dark chocolate (selected in dark green). The Single clusters are also listed aside for direct comparison. The second cluster for Ward consists of the unhealthy dark chocolate while the third consists of both milk chocolate and white chocolate. This direct comparison may indicate that Ward algorithm is better at dividing healthy and unhealthy clusters than the Single algorithm. The "Statistics" element can further enhance previous "Artificial Intelligence" and "Math" elements to characterize the Chocolate Product and Chocolate "Technology". This "Statistics" finding can help consumers pick better and healthy chocolate products. It's necessary here to pull the "Statistics" element from "Math" family in order to conduct risk management and draw more practical conclusions. "Math" element itself may be too ideal as compared to "Statistics" element.

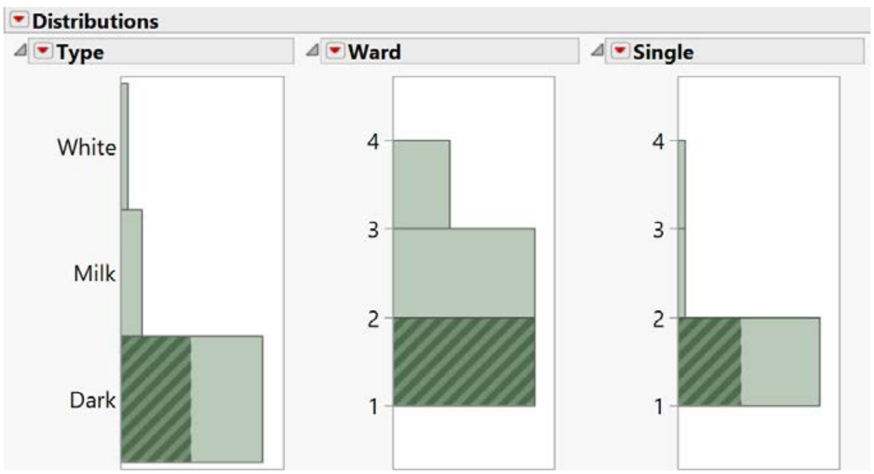

Figure 6: Distribution Analysis for the 1st Ward Cluster (selected) among Chocolate Types

3. Conclusions

"STEAMS" methodology is very successful on understanding Chocolate Science Research and Nutrition Food "Science". Understanding the Chocolate Product and "Technology" can be done effectively through a systematic "Engineering" problem solving framework. Modern "Artificial Intelligence" methods can explore the Chocolate Science Patterns which can further help consumers pick their healthy chocolate products based on their preferred nutritions needed. Clustering distance "Math" algorithm is very critical on deciding the clustering sequence and clustering patterns. "Statistics" can help conduct risk management and draw practical conclusions. In the modern Big Data era, most scientists and engineers shall adopt this "STEAMS" methodology and integrate all 6 elements seamlessly and collectively.

Acknowledgements

Authors would like to thank my Statistics Advisor Dr. Charles Chen and the Biology Advisor Mr. Patrick Giuliano for helping and supporting us throughout this "STEAMS" project.

References

1. Science, Technology, Engineering, and Mathematics (STEM) Education: A Primer". Fas.org. Retrieved 2017-08-21.

2. Rita Colwell , Compare: "Guest commentary: A "STEM" in Collier County to reach their future". Retrieved 2016-06-09.
3. President Bush Honors Excellence in Science, Mathematics and Engineering Mentoring | NSF - National Science Foundation". Retrieved 2018-10-03.
4. Teitelbaum, Michael S. "The Myth of the Science and Engineering Shortage". The Atlantic.
5. Charette, Robert N. (August 30, 2013). "The STEM Crisis Is a Myth". IEEE Spectrum.
6. Casselman, Ben (September 12, 2014). "The Economic Guide To Picking A College Major". FiveThirtyEight.
7. MacDonald, Heather (Spring 2018). "How Identity Politics Is Harming the Sciences". City Journal. Manhattan Institute. Retrieved 21 June 2018.
8. "STEAM Rising: Why we need to put the arts into STEM education". Slate. Retrieved 2016-11-10.
9. "Virginia Tech and Virginia STEAM Academy form strategic partnership to meet critical education needs". Virginia Tech News. 31 July 2012.
10. Jolly, Anne. "STEM vs. STEAM: Do the Arts Belong?". EdWeek.org. Education Week: Teacher. Retrieved 6 September 2016.
11. Pomeroy, Steven Ross. "From STEM to STEAM: Science and Art Go Hand-in-Hand". blogs.scientificamerican.com. Scientific American. Retrieved 17 November 2016.
12. Eger, John. "National Science Foundation Slowly Turning STEM to STEAM". www.huffingtonpost.com. Huffington Post. Retrieved 17 November 2016.
13. "STEAM learning matters. Science, Technology, Engineering, Arts and Mathematics". www.cbs8.com. CBS News. Retrieved 17 November 2016.
14. Schank, Roger C. (1991). "Where's the AI". AI magazine. Vol. 12 no. 4. p. 38.
15. Boom of the 1980s: rise of expert systems, Fifth Generation Project, Alvey, MCC, SCI.
16. Christiane Paul (2006). *Digital Art*, pp. 7–8. Thames & Hudson.
17. Lieser, Wolf. Digital Art. Langenscheidt: h.f. ullmann. 2009, pp. 13–15
18. James Kobielus InfoWorld | on December 22, 2017 "How generative AI is changing art and design", https://www.digitalartsonline.co.uk/features/illustration/how-generative-ai-is-changing-art-design/
19. Petyaev, Ivan M., and Yuriy K. Bashmakov. "Dark Chocolate: Opportunity for an Alliance between Medical Science and the Food Industry?" Current Neurology and Neuroscience Reports., U.S. National Library of Medicine, 2017,
20. Magrone, Thea, et al. "Cocoa and Dark Chocolate Polyphenols: From Biology to Clinical Applications." Current Neurology and Neuroscience Reports., U.S. National Library of Medicine, 2017,
21. Allen, R. R., Carson, L., Kwik-Uribe, C., Evans, E. M., & Erdman, J. W. (2008 April), "Daily consumption of a dark chocolate containing flavanols and added sterol esters affects cardiovascular risk factors in a normotensive population with elevated cholesterol", Journal of Nutrition. 138(4):725-31.
22. Wensem, van. "Overview of Scientific Evidence for Chocolate Health Benefits." Environmental Toxicology and Chemistry, Wiley-Blackwell, 26 Dec. 2014.
23. Panche, A. N., et al. "Flavonoids: An Overview." Current Neurology and Neuroscience Reports., U.S. National Library of Medicine, 2016,
24. Latif, R. (2013, March). Chocolate/cocoa and human health: a review. The Netherlands Journal of Medicine. 71(2):63-8.
25. Patel Wang, J., Varghese, M., Ono, K., Yamada, M., Levine, S., Tzavaras, N., Pasinetti, G. M. (2014). Cocoa extracts reduce Oligomerization of amyloid-β: implications for cognitive improvement in Alzheimer's disease. Journal of Alzheimer's disease, 41(2):643-50.
26. R. K., Brouner, J., & Spendiff, O. (2015, December). Dark chocolate supplementation reduces the oxygen cost of moderate intensity cycling. Journal of the International Society of Sports Nutrition 2015, 12:47.
27. Crichton, G. E., Elias, M. F., Alkerwi, A. (2016, May). Chocolate intake is associated with better cognitive function: The Maine-Syracuse Longitudinal Study. 100:126-32

28. "Antioxidants: In Depth." National Center for Complementary and Integrative Health, U.S. Department of Health and Human Services, 4 May 2016
29. Maxwell, Simon R. J., and Gregory Y. H. Lip. Advances in Pediatrics., U.S. National Library of Medicine, Oct. 1997
30. Bentson, Jacob. "Mechanics of Plaque Formation." Arteriosclerosis, Thrombosis, and Vascular Biology, 2014
31. LeWine, Howard, and M.D. "Sweet Dreams: Eating Chocolate Prevents Heart Disease." Harvard Health Blog, 17 June 2015,
32. Mason C., (2018 July) "Multivariate Statistics of Antioxidant Chocolate", IWSM Bristol Proceedings, Vol 2 37-40
33. Mason C., (2018 July), "Choose Healthy Chocolate", IEOM Europe Proceedings, 434-441
34. Wu, Anna Dong. "Starbucks and Cardiovascular Disease Prevention." IEOM, IEOM Society, 26 July 2018,
35. Harris, C.W. and Kaiser, H.F. (1964), "Oblique Factor Analytic Solutions by Orthogonal
36. Milligan, G.W. (1980), "An Examination of the Effect of Six Types of Error Perturbation on Fifteen Clustering Algorithms," Psychometrika, 45, 325–342.
37. Hartigan, J.A. (1981), "Consistence of Single Linkage for High–Density Clusters," Journal of the American Statistical Association, 76, 388–394.
38. Florek, K., Lukaszewicz, J., Perkal, J., and Zubrzycki, S. (1951a), "Sur La Liaison et la Division des Points d'un Ensemble Fini," Colloquium Mathematica, 2, 282–285.
39. Jardine, N. and Sibson, R. (1971), Mathematical Taxonomy, New York: John Wiley and Sons.
40. McQuitty, L.L. (1957), "Elementary Linkage Analysis for Isolating Orthogonal and Oblique Types and Typal Relevancies," Educational and Psychological Measurement, 17, 207–229.
41. Sokal, R.R. and Michener, C.D. (1958), "A Statistical Method for Evaluating Systematic Relationships," University of Kansas Science Bulletin, 38, 1409–1438.
42. Sneath, P.H.A. (1957) "The Application of Computers to Taxonomy," Journal of General Microbiology,17, 201–226.
43. "Correlations and Multivariate Techniques." Multiple Linear Regression