MATH 19B
Winter Quarter 2020

Selected Research Papers

Sanya Srivastava

Group Number - 80

Pedro Morales

MATH-19B

11 March 2020

DEFINING A PROBABILITY GRAPH FOR PREDICTING THE FURTHER PRICES OF CRYPTOCURRENCY (BITCOINS)

ABSTRACT

Bitcoin is a type of cryptocurrency. “A cryptocurrency is a digital or virtual currency that is secured by cryptography, which makes it nearly impossible to counterfeit or double-spend virtual currency”. (Frankenfield, Jake. “Cryptocurrency.” Investopedia, https://www.investopedia.com/terms/c/cryptocurrency.asp. Accessed on 22 January 2020.) This project would deal with defining a probability graph for predicting the future prices of bitcoins. The main idea is to form a graph with observable patterns and to predict the further prices based on those patterns. There are some minute patterns such as - most of the time the graph spikes sometime in the month of January or late December. These patters when further studied in detail could reveal bigger and more important details which would contribute towards getting a substantial conclusion.

INTRODUCTION

There is no company or individual who actually controls the prices of bitcoins but the prices are self controlled and regulated. The prices of bitcoins are highly irregular and are controlled by the following factors -

1) The supply of bitcoin and market demand for it
2) Regulations governing its sale

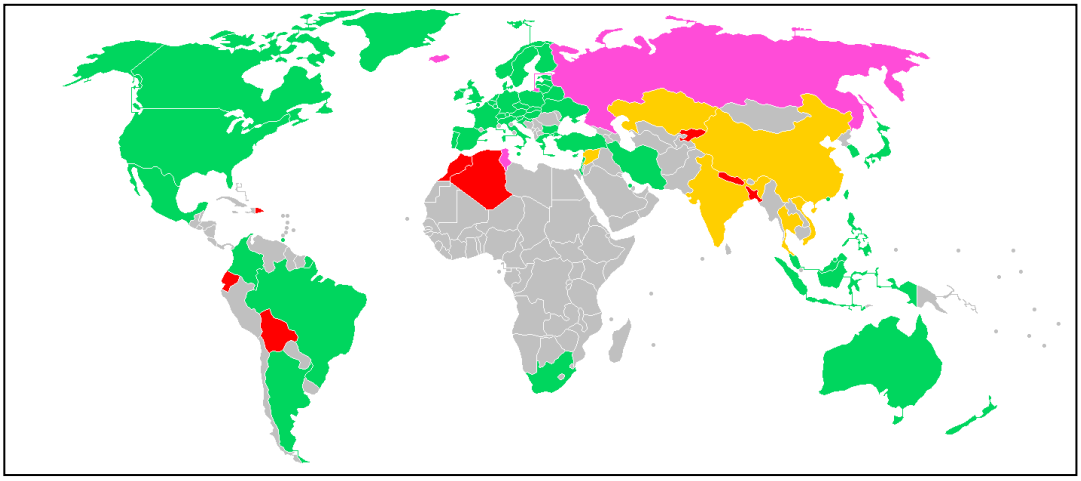
3) The number of competing cryptocurrencies

4) Regulations and Taxes

If the demand is increased, the price of the bitcoin goes up and if the demand is decreased, the price of the bitcoin goes down. The demand plays an essential role in controlling the prices of bitcoins. So the market controls the price of the bitcoins and the people who buy it are unknowingly responsible for its price fluctuation.

Another factor is the regulations and taxes. Each country has its own cryptocurrency rules and a sudden change in that policy might drastically increase or decrease the price of cryptocurrencies. For example, there was a drastic change in the bitcoin prices in China in 2017. In 2017, the Chinese government had put a ban on the ICO (Initial Coin Offering) that had resulted in the dropping of the bitcoin prices by 2000 dollars in 2 weeks of time. The bitcoin price is also affected by the rules and regulations made by the government regarding its sale. For example, in some countries like Algeria, Egypt, and Morocco, it is illegal to buy or sell bitcoins.

In the map below, the countries shaded in green refer to the countries where it is legal to



"Legal Status of Bitcoin", Legality Of Bitcoin By Country Or Territory, 6 March 2020, https://en.wikipedia.org/wiki/Legality_of_bitcoin_by_country_or_territory

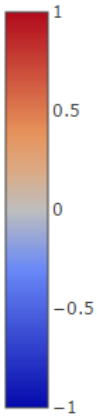
use and trade bitcoins, the countries shaded in pink refer to the countries where bitcoins are not prohibited directly by the law but there is some restriction due to the old laws, the

countries shaded in yellow refer to the countries where there are some legal restrictions on the usage of bitcoins, and lastly, the countries shaded in red are the countries where bitcoins are illegal.

The prices of other cryptocurrencies also affect bitcoin prices.

Cryptocurrency Correlation Matrix, 90-Days

	BTC	ETH	XRP	BCH	XLM	LTC	XEM	DASH	XMR	ETC	LSK	ZEC	SC	REP	FCT	NXT	S&P 500	VIX	GLD
BTC	1	0.5	0.25	0.19	0.52	0.57	0.46	0.45	0.6	0.43	0.51	0.52	0.43	0.36	0.49	0.47	0.01	-0.3	-0.1
ETH	0.5	1	0.47	0.48	0.46	0.74	0.63	0.65	0.69	0.73	0.58	0.79	0.49	0.63	0.7	0.31	-0.16	-0.08	0.13
XRP	0.25	0.47	1	0.14	0.56	0.37	0.53	0.29	0.32	0.32	0.42	0.48	0.36	0.26	0.44	0.1	-0.05	-0.08	0.16
BCH	0.19	0.48	0.14	1	0.13	0.34	0.32	0.64	0.54	0.57	0.19	0.52	0.37	0.44	0.4	0.23	-0.25	-0.05	0.02
XLM	0.52	0.46	0.56	0.13	1	0.42	0.52	0.39	0.57	0.47	0.44	0.52	0.45	0.3	0.59	0.3	-0.05	-0.1	0.08
LTC	0.57	0.74	0.37	0.34	0.42	1	0.61	0.6	0.62	0.57	0.49	0.65	0.47	0.47	0.55	0.39	-0.02	-0.12	0.04
XEM	0.46	0.63	0.53	0.32	0.52	0.61	1	0.53	0.57	0.51	0.5	0.59	0.51	0.46	0.57	0.39	0.05	-0.17	0.03
DASH	0.45	0.65	0.29	0.64	0.39	0.6	0.53	1	0.76	0.53	0.52	0.74	0.58	0.56	0.65	0.38	-0.09	-0.06	0.04
XMR	0.6	0.69	0.32	0.54	0.57	0.62	0.57	0.76	1	0.65	0.59	0.81	0.57	0.57	0.74	0.43	-0.13	-0.14	0.05
ETC	0.43	0.73	0.32	0.57	0.47	0.57	0.51	0.53	0.65	1	0.44	0.66	0.39	0.51	0.66	0.31	-0.1	-0.12	0.09
LSK	0.51	0.58	0.42	0.19	0.44	0.49	0.5	0.52	0.59	0.44	1	0.73	0.58	0.48	0.67	0.38	-0.06	-0.07	0.07
ZEC	0.52	0.79	0.48	0.52	0.52	0.65	0.59	0.74	0.81	0.66	0.73	1	0.6	0.6	0.79	0.39	-0.16	-0.07	0.11
SC	0.43	0.49	0.36	0.37	0.45	0.47	0.51	0.58	0.57	0.39	0.58	0.6	1	0.45	0.58	0.49	-0.08	-0.01	0.12
REP	0.36	0.63	0.26	0.44	0.3	0.47	0.46	0.56	0.57	0.51	0.48	0.6	0.45	1	0.57	0.32	-0.13	0.01	0.14
FCT	0.49	0.7	0.44	0.4	0.59	0.55	0.57	0.65	0.74	0.66	0.67	0.79	0.58	0.57	1	0.35	-0.17	-0.06	0.01
NXT	0.47	0.31	0.1	0.23	0.3	0.39	0.39	0.38	0.43	0.31	0.38	0.39	0.49	0.32	0.35	1	0.03	-0.17	-0.06
S&P 500	0.01	-0.16	-0.05	-0.25	-0.05	-0.02	0.05	-0.09	-0.13	-0.1	-0.06	-0.16	-0.08	-0.13	-0.17	0.03	1	-0.55	0.01
VIX	-0.3	-0.08	-0.08	-0.05	-0.1	-0.12	-0.17	-0.06	-0.14	-0.12	-0.07	-0.07	-0.01	0.01	-0.06	-0.17	-0.55	1	-0.03
GLD	-0.1	0.13	0.16	0.02	0.08	0.04	0.03	0.04	0.05	0.09	0.07	0.11	0.12	0.14	0.01	-0.06	0.01	-0.03	1



Rousey, Mariia. "*Cryptocurrency Correlation Matrix, 90 Days". What Affects The Price Of One Bitcoin, 1 November, 2019,* https://changelly.com/blog/what-affects-the-price-of-1-bitcoin/

The above Matrix is a representation of how the price of a cryptocurrency is influenced by the other cryptocurrency's prices. 1 means that the assets and prices are affected by external factors, 0 means the price is not affected, -1 means there is a negative correlation, movement of the prices and assets occurs in the opposite direction. Bitcoin is the most famous and successful cryptocurrency and controls approximately 70% of the cryptocurrency market so it is not much affected by other cryptocurrencies.

In this paper, a probability curve is designed using machine learning that is capable of predicting the future prices of bitcoins with the minimum possible error. Three machine learning models are designed and then are tested using the test data to check their accuracy. The data is predicted on the basis of the price history of the bitcoins.

A policy change of any country that is capable of affecting the bitcoin prices, that might occur in the future would make with the prices predicted, by the machine learning models developed for this research, inaccurate. This might happen because such future events do not have a pattern and such events have not been taken into account while developing the machine learning models and probability curves.

METHODOLOGY

The method of machine learning is used to define a probability curve for the prices of the bitcoins. At first, the data of the prices of bitcoins of each day from 2014 to 2020 February is collected. The data is collected from Yahoo Finance. The machine learning model is trained using this data. Three methods are used - linear, polynomial, and a Radial Basis Function (RBF) to predict which model give the best value.

From the data file, four kinds of values are accessible - Open (Price of the bitcoin during the first transaction of the day), High (Highest value reached by the bitcoin in the day), Low (Lowest Value reached by the bitcoin in the day), Close (Price of the bitcoin when the last transaction of the day was made). For constructing the machine learning models, the Closing value of the bitcoins of each day has been taken into consideration.

FACTOR(S) NOT TAKEN INTO CONSIDERATION -

- The volume of bitcoins has not been taken into consideration for this research.
- Events where policies were changed and the prices were affected.
- Prices of other cryptocurrencies.
- High, Low, and Open prices of each day.

FACTOR(S) TAKEN INTO CONSIDERATION -

- Closing Price History of bitcoins of the past years (from February 2014).

BITCOIN PRICE DATASET, TEST DATASET, AND TRAIN DATA SET VALUES-

Bitcoin Price Data Set - Value of Bitcoins from 17th September 2014 to May 24, 2019.

Train Data - Value of Bitcoins from 17 September 2014 to 22 February 2019.

Test Data1 - Value of Bitcoins from 22 February 2019 to 26 February 2020.

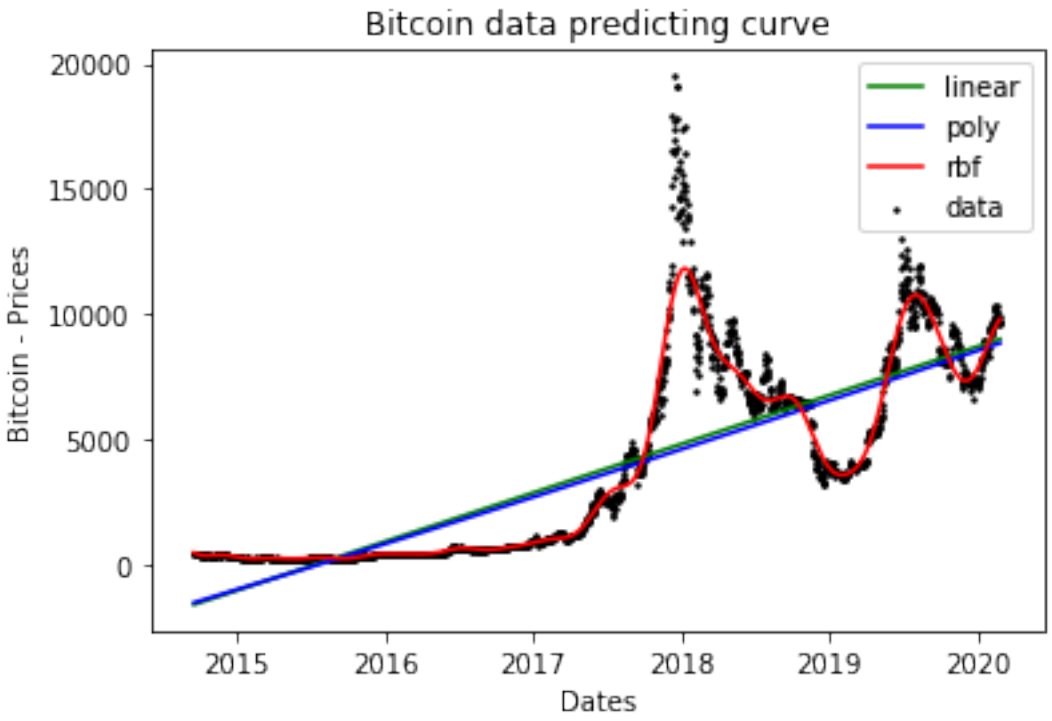
Test Data 2 - Value of Bitcoins from 22 February 2019 to 24 February 2019.

Test Data 3 - Value of Bitcoins from 14 May 2019 to 24 May 2019.

PYTHON LIBRARIES USED FOR CONSTRUCTING THE MACHINE LEARNING MODEL AND GRAPHS -

- Matplotlib
- Numpy
- Csv
- Sklearn

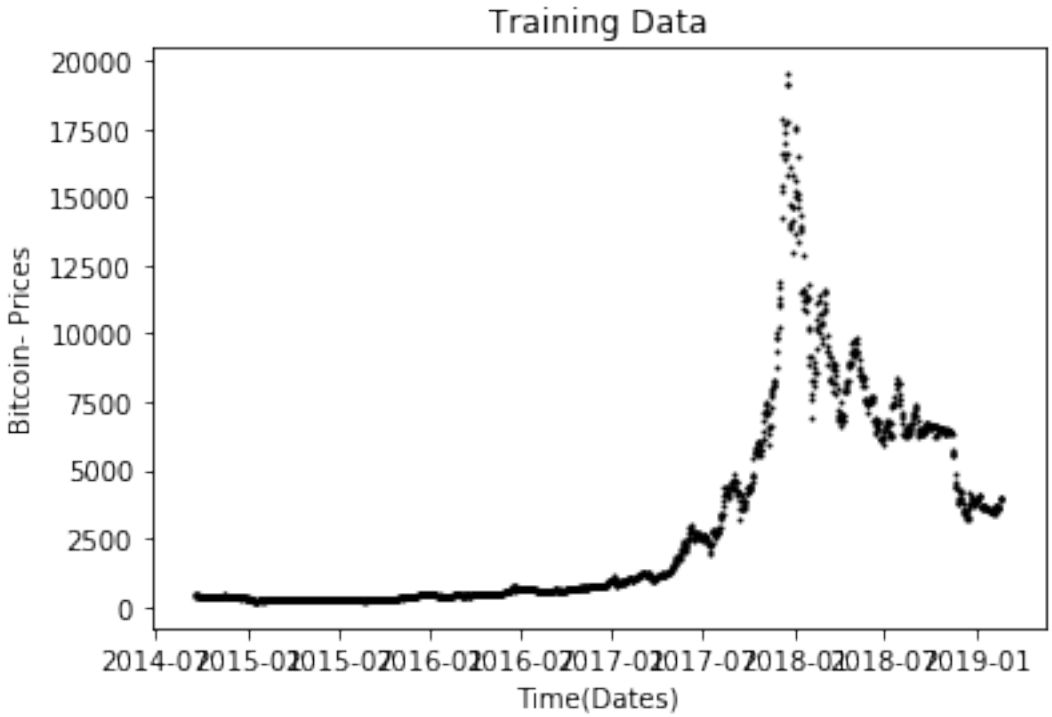
The graph containing the price of the bitcoins of each day from 17th September 2014 to 22nd February 2020 is shown below -



This graph is constructed using the entire Bitcoin Price Data Set available. The black dots refer to the actual data (the bitcoin prices) that is plotted on the graph. The green line

that is greatly overlapped by the blue line of the polynomial function (of degree 2) is of the linear function. The red curve is of the Radial basis function. From this model, it becomes obvious that the red curve of the RBF is capable of giving a better prediction. To test the accuracy of this model, we divide the entire data into two sets - Training data and Testing data. The testing data is further divided into three data sets (Test Data 1, Test Data 2, and Test Data 3). All the data is not used to train the machine but a part of it is used. The rest of the data is used to test which model gives the most accurate data and to which extent.

The graph obtained by the Training data from 17th September 2014 to 22nd February 2019 is -



For this paper, the concept of SVR has been used that is Support Vector Regression, a method by which optimum machine learning models for predicting future data can be obtained. The Support Vector Regression (SVR) uses the same principles as the Support Vector Machine (SVM) for classification, with a few differences. SMR uses a function

called kernel to map a data into higher dimensions. For example, a 1 dimensional data is converted into higher dimensions for better accuracy and for obtaining a better hyperplane that can predict data more accurately. Hyperplane is defined as a line that is used to predict the future values. There are two lines on either side of the hyper plane to create a boundary called boundary lines .

Assuming that we have a linear model and the boundary lines are at a distance of x unites from the hyper plane, and the equation of the hyper plane is -

$$Ax + B = 0$$

The equations of the boundary lines would be -

$$Ax + b = x \text{ and } Ax + B = - x$$

The same concept is used for the polynomial and the RBF.

In the linear function, the data is mapped into just a single dimension, in the polynomial function, the data is mapped into multiple dimensions and in the RBF, the data is mapped into infinite dimensions.

The accuracy of these graphs are calculated with three kinds of test data - A small test data consisting of predictions for a few days , and a huge test data set consisting of predictions for 2 to 6 months, and a data set having vales between the small and the large data set. The accuracy of the models is tested by calculating the root mean squared error they give when tested with each test data set.

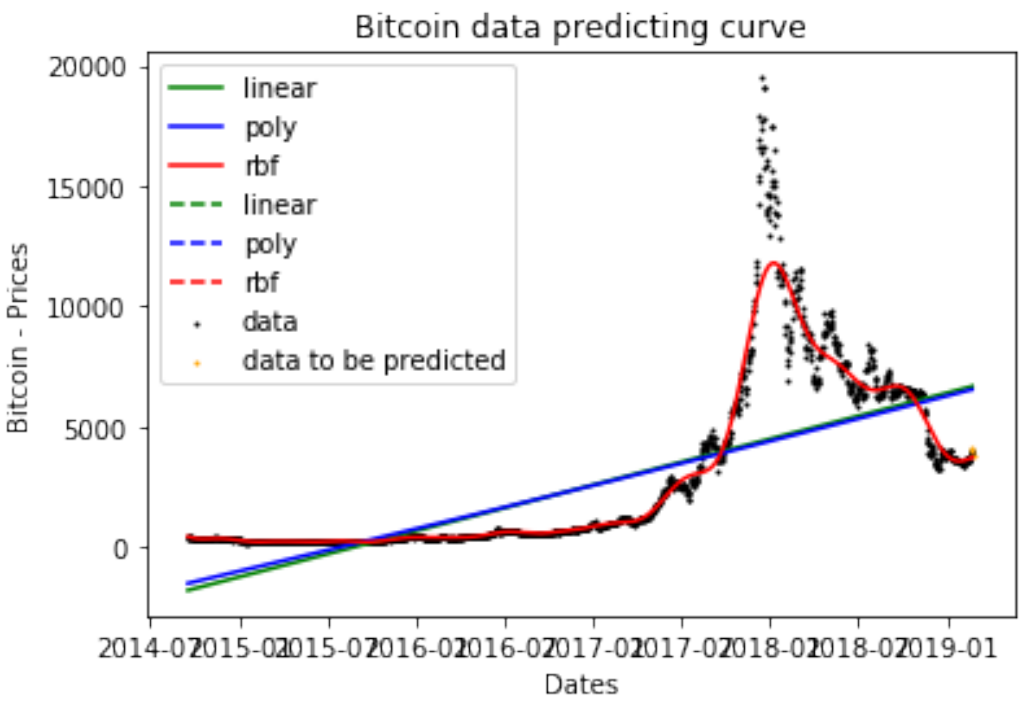
$$RMSErrors = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}}$$

Holmes, Susan. RMS Error, 28 November, 2000,
\ http://statweb.stanford.edu/~susan/courses/s60/split/node60.html

yi hat represents the predicted data and yi represents the original data. n is the number of the days for which this data was predicted, for example if a data is predicted for 3 days, n would be equal to 3.

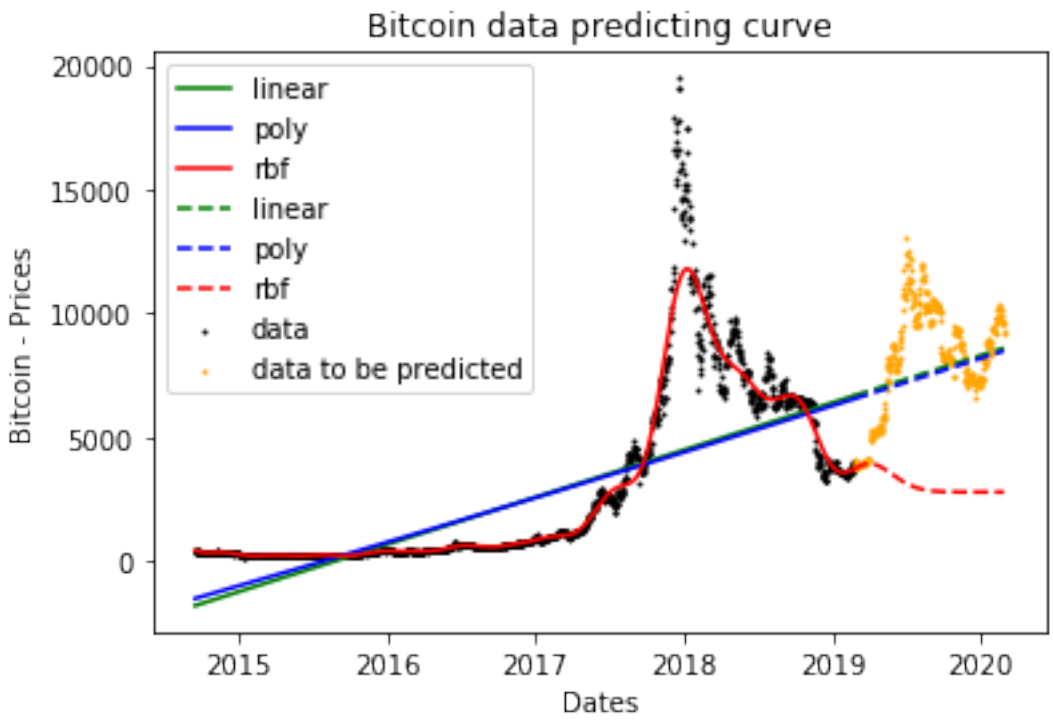
RESULTS

If the future prices of a few days have to be predicted (such as 3 days or 6 days) the RBF function predicts the values with the least amount of error out of the 3 models. The root mean square error with the RBF when a prediction is made of the values of the next 5 to 6 days is approximately 280.60 which is way lesser as compared to the linear and poly function which give the root mean square error of 2708.0 and 2579.04 respectively.



In the above graph, the orange dots represent the data that needs to be predicted and it overlaps with the red dashed line. This red dashed line is the line that shows the data predicted by the RBF graph. The green dashed line and blue dashed line which are overlapped by their respective non dashed lines represent the data predicted by the linear function and the polynomial function.

If the future price of bitcoins of each day has to be predicted for a big period of time such as of each day for the future 6 months (when the test data is quite big), the curve of the linear function would give the most accurate result while the curve of the RBF would give the least accurate predictions. The accuracy of the polynomial function lies between the linear function and the RBF in this case. The root mean square error given by the RBF is 5620.584. The root mean square error by the linear function is 2055.822 and by the poly function is 2091.515.



This graph shows the values predicted by the respective three functions for the data set that was given to this program (depicted with an orange dashed line). The red dashed

line depicts the data predicted by the RBF function. The blue dashed line depicts the data predicted by the polynomial function and the green dashed line depicts the data predicted by the linear function. It is clear by this graph that the linear and poly functions give the most accurate values for the future values (values taken from the test data).

The final findings meet the objective of the central research question. The central research question was to construct a probability curve, by observing the patterns, that is capable of predicting the future prices of bitcoins. The aim could have been achieved more accurately with a better machine learning model that could have been developed by using more number of factors that are responsible for the prices of the bitcoins.

LIMITATIONS OF THE MODEL/RESEARCH

- The values predicted by the curves might be inaccurate because prices of bitcoins fluctuate frequently and by large values.
- Different curves have to be used to predict the future prices of bitcoins depending on the time period for which the values have to be obtained. This model can be improved if the future prices can be predicted using one curve or function.
- This research does not specify a formula that could be used for predicting the future price of bitcoins. All the future prices are supposed to be predicted by the machine learning code and by the curve developed using the code.
- Events that may influence the price of bitcoins like the banning of the ICO in China in 2017 have not been taken into consideration for this project. So the models developed in this research may give out an inaccurate result if such policy changes happen.
- The prices of bitcoins depend on many factors and all of them have not been taken into consideration. This model is developed only using the closing prices of each day the past few years.

POSSIBLE IMPORVEMENTS

- To construct a formula that is based on the probability curves and can be used to predict the values of the function.
- To involve the volume of bitcoins as another parameter in predicting the future prices. The total volume of bitcoins of each day could be used as another parameter to train the machine learning model developed and used for this research. Including more parameters might increase the accuracy of the model.
- Other machine learning models can be tested to see if they give better predictions or two models can somehow be integrated together to form a better model for predicting data.
- The following factors could be taken into consideration for improving the model -
    - Prices of Volume of the Bitcoin
    - Open, Low, and High prices
    - Prices of Other Cryptocurrencies

RESOURCES USED

- nolo.com (https://www.nolo.com/legal-encyclopedia/decoding-cryptocurrency.html)
- CoinMarketCap.com (https://coinmarketcap.com/currencies/bitcoin/historical-data/?start=20191016&end=20191230)
- pluralsight.com (https://www.pluralsight.com/guides/the-cryptography-of-bitcoin)
- cryptocompare.com (https://www.cryptocompare.com/wallets/guides/how-does-bitcoin-cryptography-work/)
- Yahoo Finance (https://finance.yahoo.com/quote/BTC-USD/history?period1=1551052800&period2=1582416000&interval=1mo&filter=history&frequency=1mo)

- Wikipedia (https://en.wikipedia.org/wiki/Radial_basis_function)
- http://pages.cs.wisc.edu/~matthewb/pages/notes/pdf/svms/RBFKernel.pdf
- medium.com (https://medium.com/coinmonks/support-vector-regression-or-svr-8eb3acf6d0ff)
- stanford.edu (http://statweb.stanford.edu/~susan/courses/s60/split/node60.html)
- investopedia.com (https://www.investopedia.com/terms/b/blockchain.asp)
- changelly.com (https://changelly.com/blog/what-affects-the-price-of-1-bitcoin/)
- *wikipedia.org (https://en.wikipedia.org/wiki/Legality_of_bitcoin_by_country_or_territory)*

The Fibonacci Sequence, Golden Ratio and its Applications

Nora Jones and Rana Rahimi
rarahimi@ucsc.edu
nogjones@ucsc.edu

March 11, 2020

Abstract

Discovering what the Fibonacci Sequence represents and how it applies to the world began by exploring the background of Leonardo Bigollo Pisano—Fibonacci himself. To understand the sequence and the relationship to the golden ratio, research was done in regards to its history, its mathematical origins, and its mathematical applications in the real world. The relationship between the numbers and the sequence has been defined by scholars as an irrational number known as the Golden Ratio. The Golden Ratio was often employed in the arts and architecture, due to it supposedly being the most aesthetically pleasing physical ratio. The sequence itself and the Golden Ratio both are extremely present in science as well as art. It has been used as an explanation for many natural phenomena such as plant appendage arrangement and seed distribution. Its relevance in the universe and its irrationality make it an extremely fascinating sequence and number to understand and analyze whether from a purely mathematical perspective or from an applied one.

1 Introduction

The main purpose of researching Fibonocci's sequence is to discover its importance from the time it was discovered to its applications in the present day. The sequence was discovered in the thirteenth century, and was an extremely significant discovery for its time, as the Middle Ages were a dead period for the development of mathematics. When publishing *Liber Abaci,* the book proving and introducing the Fibonacci sequence, Fibonacci made it possible in his works for Western Europeans to become acquainted with Hindu Arabic Numerals, as he used them when proposing his sequence modeled by rabbit breeding. Before then, Ancient Greecians developed Phi and applied it to their art and architecture, and later, scholars were able to relate his sequence with the golden ratio ϕ discovered far before his time. The importance of understanding and using Fibonacci's sequence now is its relevance in science, music, artitecture and art. It can serve as an explanation as to why plant petals and seeds are arranged in particular ways; these numbers tend to allow plants to expend the least

amount of energy and be the most efficient. Many Ancient Greecian and Renaissance sculptors and artists, utilized the golden ratio and Fibonacci sequence in their work as it was seen as the ideal representation of aesthetic beauty. In general, it's important to understand the various applications in the real world, as many see mathematics as an abstract subject with no real world application which is why many deter away from engaging in mathematics itself.

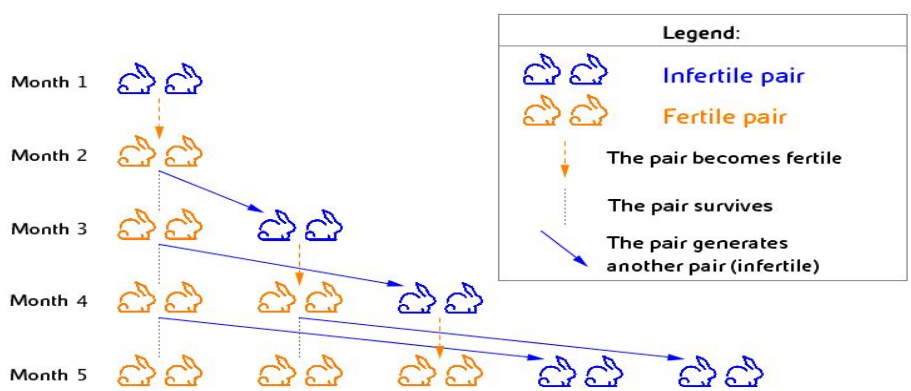
2 Methodology

At first, very broad research was conducted about understanding the Fibonacci sequence and how it is represented in mathematics using supplementary videos that math professors and teachers have posted on YouTube. The research led to the discovery of the applications of the Fibonacci Sequence and subsequently, the golden ratio, which is the consistent ratio that is found between two consecutive numbers in the sequence. From there, formally deriving this number, also known as ϕ was studied by reading an article from the University of Georgia. After deriving the golden ratio ϕ itself, its application in the mathematical world was studied by reading *Euclid's Elements* and scholarly interpretations of his research—exploring the number's relevance in geometry. Research also led to finding out how the Fibonacci sequence was utilized outside of the mathematical world in forms of art and architecture. Performing the research allowed for an in depth understanding of the sequence and its applications that a general education in school doesn't address.

3 Research

3.1 Research of Origin

Fibonacci originated the sequence by solving a famous problem called the "Rabbit Problem". The rabbit problem asks, "How many pairs of rabbits will be born in a year, starting from a single pair. If each month each pair gives birth to a new pair which becomes reproductive from the second month?" Fibonacci answered this by creating the following sequence: 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89... The diagram below also represents the solution to the problem.



3.2 Mathematical Research

The Fibonacci Sequence is defined as $x_n = x_{n-1} + x_{n-2}$. The series begins with 0 and continues following this pattern: 0, 1, 1, 2, 3, 5, 8, 13, 21... and so on. These numbers, when the larger number is divided by the previous number in the series (i.e. $x_n \div x_{n-1}$), always equal a consistent ratio known as the golden ratio. This exact, golden number has been derived utilizing a multitude of methods.

As stated above, it is known that the fibonacci sequence follows the pattern where $x_n = x_{n-1} + x_{n-2}$. As the numbers in the sequence approach infinity, then the ratio between them become closer and closer to a constant ratio, therefore,

$$\lim_{n\to\infty} \frac{x_n}{x_{n-1}} = \lim_{n\to\infty} \frac{x_{n-1}}{x_{n-2}} .$$

If this ratio is called ϕ, then,

$$\lim_{n\to\infty} \frac{x_n}{x_{n-1}} = \lim_{n\to\infty} \frac{x_{n-1}}{x_{n-2}} = \phi .$$

It is known that $x_n = x_{n-1} + x_{n-2}$, therefore,

$$\lim_{n\to\infty} \frac{x_{n-1} + x_{n-2}}{x_{n-1}} = \lim_{n\to\infty} \frac{x_{n-1}}{x_{n-2}} ,$$

$$\lim_{n\to\infty} 1 + \frac{x_{n-2}}{x_{n-1}} = \lim_{n\to\infty} \frac{x_{n-1}}{x_{n-2}} .$$

If $\lim_{n\to\infty} \frac{x_{n-1}}{x_{n-2}} = \phi$, then,

$$1 + \frac{1}{\phi} = \phi ,$$

$$\phi + 1 = \phi^2 ,$$

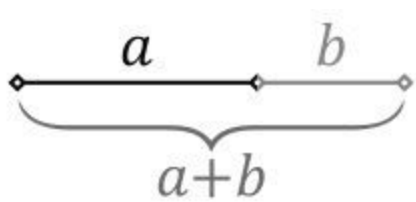
$$0 = \phi^2 - \phi - 1 .$$

If the quadratic formula is used, then the golden ratio, or ϕ, can be defined as

$$\phi = \frac{1 \pm \sqrt{5}}{2} \text{ or } \approx 1.61803...$$

After the Fibonacci Sequence was established, mathematicians such as Euclid began to notice and study its appearance in geometry. The geometric relationship of the golden ratio can be illustrated as,

where $a + b$ is to a as a is to b, or

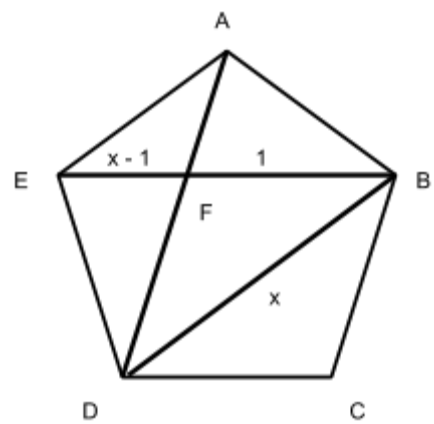


$$\frac{a+b}{a} = \frac{a}{b} = \phi .$$

One of the instances in which the golden ratio is present in geometry is within a regular (equiangular and equilateral) pentagon. Consider a regular pentagon—the inscribing of such within a circle is illustrated in Euclid's Elements Book IV, Proposition XI. If one side of the pentagon is defined as 1, then another diagonal line is drawn from one vertex to a non-adjacent vertex defined as *x*. This, subsequently, forms two similar triangles:

$$\Delta AEF \sim \Delta DBF .$$

Because these triangles are similar, then



$$\frac{AE}{BD} = \frac{AF}{DF} .$$

Substituting the values of *x* and 1,

$$\frac{1}{x} = \frac{x-1}{1} ,$$

$$1 = x^2 - x ,$$

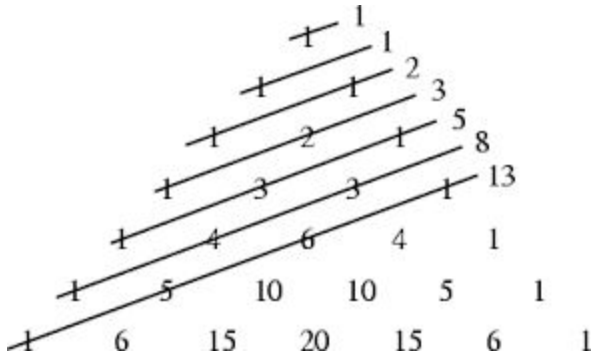
$$0 = x^2 - x - 1 .$$

If the quadratic formula is used, then

$$x = \frac{1 \pm \sqrt{5}}{2} \text{ or } \approx 1.61803...$$

Which is ϕ or the golden ratio. Thus, the ratio between the side of a regular polygon and a diagonal is always the golden number.

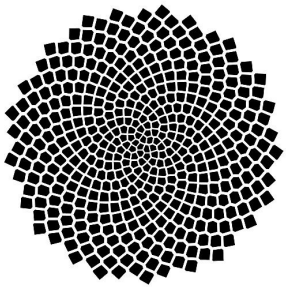
Interestingly, the Fibonacci sequence also appears in Pascal's triangle—a triangle formed by the sums of the numbers diagonally above it beginning with 1. If the diagonals of the numbers are added, they equal—sequentially—Fibonacci numbers.



3.3 Applied Research

3.31 Nature

The golden ratio can be applied in the understanding of phyllotaxis, which is the arrangement of leaves, branches, flowers or seeds in plants on an axis. For example, in many flowers like daisies and sunflowers, the seeds are arranged in the center such that the space they occupy is optimized. If seeds are arranged around an axis, for example, in a ratio of 1, then there would be a line of seeds protruding outwards. If they were arranged in a ratio of ½, then there would be two spokes protruding outward; a ratio of ⅓ would cause three spokes and so on. If one were to arrange the seeds using the ratio of an irrational number (such as π), seeds would begin to arrange themselves in a sporadic formation, but then eventually begin to resort themselves back into a three-spoke pattern. However, when the seeds are arranged in the ratio of ϕ, then the seeds arrange themselves in the most efficient arrangement they can be in, and do not revert back to any spoked pattern as shown below.



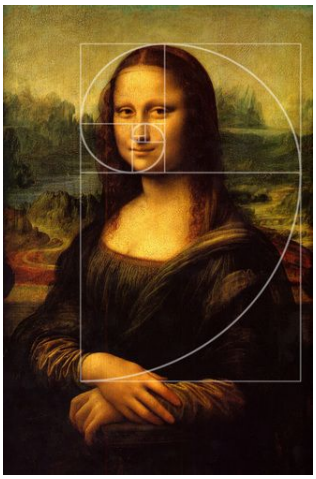
3.32 Music

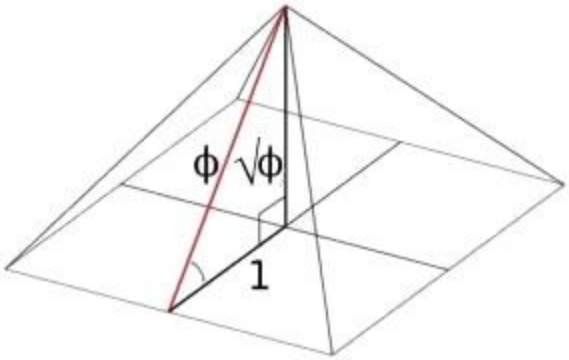
Fibonacci's sequence is also present in music. For example, the intervals between keys on a piano of the same scales are Fibonacci numbers. The importance of this is the fact that Fibonacci numbers harmonize well with each other. In fact, the golden ratio that relates to the numbers is what

Stradivarius, a famous instrumental crafter used in order to create his string instruments such as violins, cellos, and guitars.

3.33 Art and Architecture

The application of the Fibonacci sequence and the golden ratio, can also be seen in art and architecture. Leonardo da Vinci uses this golden ratio in creating the Mona Lisa. The reason behind this was because the golden ratio is seen to be the most aesthetically pleasing especially in forms of art. Moreover, the golden ratio is used in architecture due to its aesthetically pleasing appearance. For example, the perimeter of the Great Pyramid in Egypt, divided by twice its vertical height is the value of the golden ratio.





4 Conclusion/Discussion

Through the research, and synthesis of information collected, the origins of the Fibonacci sequence was identified and the number phi was derived. A geometrical application, using a regular pentagon was further proved using phi. After the mathematical research conducted, the application of the Fibonacci sequence and the golden ratio beyond mathematics was studied. The Fibonacci sequence and golden ratio are used in phyllotaxis, creating musical instruments, designing famous structures, and even in creating art pieces. Due to its pre-existence in nature and famous architectural pieces it was easy to identify the ways that the sequence is applied in the tangible universe. However, its prevalence in modern day applications were more limited. Nevertheless, in the future, applications of the sequence and the golden ratio could be used in coding and in predicting the stock market.

5 References

Benavoli, A., et al. "Fibonacci Sequence, Golden Section, Kalman Filter and Optimal Control." *Signal Processing*, Elsevier, 20 Feb. 2009, www.sciencedirect.com/science/article/pii/S0165168409000449.

Euclid, et al. *Euclids Elements: Books I-VI, XI and XII*. Dent, 1979.

Fiorenza, et al. "From Fibonacci Sequence to the Golden Ratio." *Journal of Mathematics*, Hindawi, 18 Mar. 2013, www.hindawi.com/journals/jmath/2013/204674/.

Geometrical Substantiation of *Phi*, the Golden Ratio and the Baroque of Nature, Architecture, Design and Engineering- Md. Akhtaruzzaman , Amir A. Shafie

Sudipta Sinha . "The Fibonacci Numbers and Its Amazing Applications." International Journal of Engineering Science Invention (IJESI), vol. 6, no. 9, 2017, pp. 07–14

TED, June 2013, www.ted.com/talks/arthur_benjamin_the_magic_of_fibonacci_numbers.

The Fibonacci sequence and the golden ratio in music - Robert van Gend

Vorobiev, N. N., and Mircea Martin. *Fibonacci Numbers*. Birkhäuser, 2002.

Getting Your Goat: The History of One of the Most Controversial Questions in Mathematics

Abstract:

This paper will be taking an analytical look at the Monty Hall problem, a controversial question within statistics, and whether or not the initial answer posited by Steve Selvin in 1975 is correct.

Introduction:

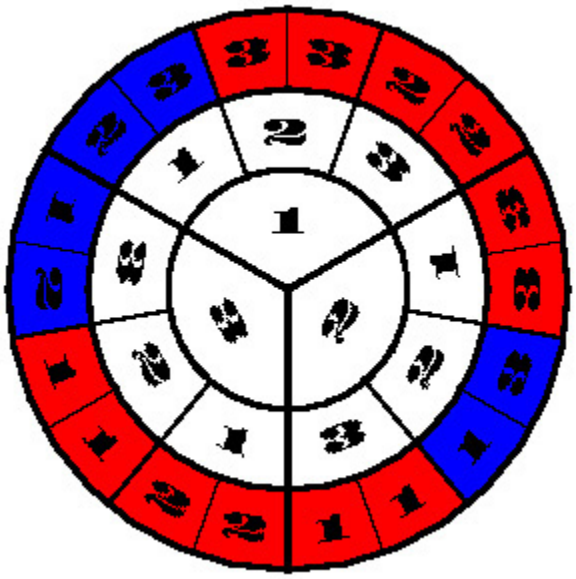
The Monty Hall problem is a famous statistics problem created in 1975. Steve Selvin, a professor at UC Berkeley, posited the problem by creating a scenario where the person being asked the question plays the role of a contestant of the tv show Let's Make a Deal, hosted by Monty Hall. The contestant is trying to open a door in a set of three doors that holds a car behind it, while the other two have goats. When the contestant chooses a door, Monty Hall opens one of the two doors they didn't pick, and reveals a goat. The question is whether or not it is statistically advantageous for the contestant to switch their choice to the other unopened door. Selvin controversially wrote that it would, in fact, be advantageous for the contestant to switch, having a ⅔ chance of hiding the car.

In 1990, Marilyn vos Savant, a writer for Parade, would introduce the greater world to this problem through her Ask Marilyn column, which focused on logic puzzles and brainteasers. After giving the same conclusion as Selvin, it quickly became infamous, with thousands decrying the answer as wrong. This answer would be debated to this day. By reading books on statistics and online publications made by various universities, the paper will compare the answers found by both the creator of the

problem, Steve Selvin, and those that agree with him, and what methods were used to find them.

Methods:

Selvin and his believers use the following logic to support their claim: The chances of picking the door with the car behind it is ⅓, meaning the chance the car is behind another door is ⅔. The three scenarios are you picked the correct door, meaning Monty Hall can open either of the two doors, you picked the wrong door, and Monty Hall has to open the one other door with the goat, or you picked the wrong door, and Monty Hall will open the one other door with the goat. This repetition is purposeful, as this means that when picking an initial door, there is a ⅔ chance of initially picking a door with a goat and having the other door with a goat be revealed. In other words, this gives picking the remaining closed door a ⅓ chance of leading to a goat, or ⅔ chance of leading to a car.



This image serves as a visual representation of the possible outcome, with the innermost circle being the door chosen, the middle circle being the correct door, and the outermost circle being the door opened.

A more mathematical approach deals with Bayes' theorem, in the case of this problem

$$P(C_1|D_3) = \frac{P(D_3|C_1)P(C_1)}{P(D_3|C_1)P(C_1) + P(D_3|C_2)P(C_2) + P(D_3|C_3)P(C_3)}.$$

Where Ci is the event the car is behind the door, Di is the event Monty Hall opens a door with a goat, and i being a number of a door 1, 2, or 3. Solving this equation, you can see the chance of the car being behind the door you initially chose is ⅓.

Despite the vast majority of people (around 79-87% of people according to a 1999 cross cultural study) initially believing that the opening would lead to 50/50 chance of having picked the correct door, thus switching giving no mathematical advantage, in fact the chance remains ⅓. We can see this when applied to the real world as well. Leonard Mleodinow, in his book The Drunkard's Walk, claimed that in the game show the problem is based on, contestants that chose to switch were twice as likely to win as those that chose their initial door. Humorously, Monty Hall himself in a 1991 interview with the New York Times claimed that if switching is the best option, if the host is required to open a door.

<u>Results:</u>

A variety of sources, from books published by noted mathematicians to posts made by various universities to the titular host himself, that cover the topic agrees with the initial answer given by Selvin, using a variety of methods including mathematical

equations and logistics. No sources were found that offered a counter argument to the problem that weren't simply being pedantic about the admittedly contrived nature of the original problem.

Conclusion:

It can reasonably be concluded that, given the specificities of the question, the answer to the original problem is that it is more advantageous to switch doors when given the chance. Furthermore, study into the topic shows the sheer depths of misunderstanding of the problem among the general populace, regardless of cultural background. Much of this can be attributed to the wording of the original problem, though a lack of knowledge when it comes to mathematics also contributes to this issue. The limited scope of this study meant that questioning a larger pool of people and being certain they couldn't just look up the answer online was unlikely. A way to extend this study is by researching how different people of different demographics approach this problem. While a study cited in this paper researched the seemingly continuous problem to answer the question across cultures, further exploration could be had with people of various ages and education within a culture. Research could also be expanded to measure responses to problems similar in nature to the Monty Hall problem, such as the lesser known Sleeping Beauty paradox.

Resources:

Cetina, D. (2019, May 10). Don't Switch! Why Mathematicians' Answer to the Monty Hall Problem is Wrong. Retrieved from https://ima.org.uk/4552/dont-switch-mathematicians-answer-monty-hall-problem-wrong/

"Let's Make A Deal." *The Monty Hall Page*, math.ucsd.edu/~crypto/Monty/montybg.html.

Rosenthal, J. S. (2008). Monty Hall, Monty Fall, Monty Crawl. *Math Horizons*, *16*(1), 5–7. doi: 10.1080/10724117.2008.11974778

Saenen, Lore, et al. “Why Humans Fail in Solving the Monty Hall Dilemma: A Systematic Review.” *Psychologica Belgica*, Ubiquity Press, 1 June 2018, www.ncbi.nlm.nih.gov/pmc/articles/PMC6194549/.

Wheelan, C. (2014). *Naked statistics stripping the dread from the data*. New York, NY: W.W. Norton & Company.

A Survey of Statistical Mechanics

By Alexei Martsinkovskiy

Abstract: Statistical mechanics is simply put, the study of systems and how they interact in nature. More precisely, it is the study of how matter behaves based on the behavior of the constituent molecules. This can be derived largely using classical mechanics, quantum mechanics, and mathematical methods. In this paper, we will explore the simpler concepts in statistical mechanics, focusing especially on the background of the topic in relation to entropy and thermodynamics.

Background: We must first define some important concepts. An ideal gas is a gas that is assumed to not interact with other gas particles. This means not only do electrostatic interactions take place, but also the gas is assumed to have no mass. Entropy, in classical mechanics is the availability of a system to do work. According to the 2nd law of thermodynamics, all-natural systems approach a state of maximum entropy. For example, we can think of a glass of boiling water cooling. As it cools towards room temperature, it tends towards a state of being unavailable to do work (with the work here being heating up the surroundings). In Boltzmann's study of thermodynamics, he proposed that a system is made up of a number of microstates. Each microstate is defined as the position and momentum of each individual particle. When you add up all the properties of the microstates, you should come to the state of the system. This is also referred to as the macrostate. In statistical mechanics, with the assumption that each possible microstate is equally probable, and all other microstates are impossible, the formula for entropy is $S = k_b \ln(\Omega)$ With S being the total entropy of the system, and k_b is a constant, and Ω is the number of possible microstates for that system. This is where the argument that energy = disorder comes from. This is not technically true. A more correct way is that as entropy increases, more ways of distributing energy within the system are possible.

Derivation of Entropy

Let us say that a model of an ideal class consists of a number of particles. We can represent these in a space, with an axis for every coordinate of each particle. Thus, this is an abstract space, with 3N dimensions (N= number of particles). This is only the positional state of the system. We must also add an axis for each component of each particles velocity, therefore we get an abstract 6N dimensional space. It is important to note that the potential energy of the system is 0, since the particles do not interact in any way. This is untrue, but we do assume so anyway. If we also make the fairly logical assumption that the position of each particle is independent of its momentum, we get $P(q,p) = P_q(q)P_p(p)$. In other words, we can represent the probability of a particle having position q and momentum p as the product of the probability of its position being q and its velocity being p. We are assuming the gas is ideal, so the only two things that

contribute to entropy are the momentum and the velocity of the particles. Therefore, the entropy of an ideal gas can be expressed as the sum of the entropy given by its position (configurational entropy) and entropy contributed by its momentum (energy dependent entropy).

The configurational entropy is only dependent on 2 factors. The number of particles (N) and the volume of the container or system (V). Therefore, $S_q = S_{q(V,N)}$. Let us consider a system containing two boxes containing N boxes of distinguishable ideal gas particles. We can call the boxes A and B, with volumes V_A and V_B. The number of particles in box A is N_A and the number of particles in N_B is given by $N_B = N - N_A$. The number of particles in each box can be either contained (a sealed box) or by making a hole that leads to the other box. These boxes are a closed system however, and the number of particles will still be kept constant. Assuming the particles are kept constant, the probability density of the particles when the wall is removed can be written as the following: (note: this math is beyond me and is taken from an Introduction to Statistical Mechanics and Thermodynamics).

$$P_q(q) = P_N(\{\vec{r_j}\}) = \prod_{j=1}^{N} P_1(\overrightarrow{r_j})$$

If we further assume that it is equally likely the particle will be anywhere in the system, the probability of it being in box A is $\frac{V_A}{V}$. If N particles are going back and forth between boxes A and B, the distribution of the particles in box A, N_A is a binomial distribution. It can be given by the formula:

$$P(N_A|N) = \frac{N!}{N_A!\,(N - N_A!)}\left(\frac{V_A}{V}\right)^{N_A}\left(1 - \left(\frac{V_A}{V}\right)\right)^{N-N_A}$$

Additionally, $N_A + N_B$ must equal N. We must also consider the width of the probability distribution for N_A, especially since we are dealing with such large numbers. This is given by

$$\delta N_A = \left[N\left(\frac{V_A}{V}\right)\left(1 - \frac{V_A}{V}\right)\right]^{\frac{1}{2}} = \left[N\left(\frac{V_A}{V}\right)\left(\frac{V_B}{V}\right)\right]^{\frac{1}{2}} = \left[\langle N_A\rangle\left(\frac{V_B}{V}\right)\right]^{\frac{1}{2}}$$

It is important to recognize difference between N_A and $\langle N_A\rangle$. The number of particles in box A, N_A is a distinct property of our system. It will change with time and is never constant. It is an integer (you can't have half a particle; it is either the whole particle or no particle). Meanwhile, $\langle N_A\rangle$ is not an integer and is used to describe the system. It does not depend on time and will not change. It is a statistical variable used to describe box A, but is not a property of box A. The magnitude of fluctuation from the actual number of particles is the standard deviation of N_A. The order of this is equal to the following equation: $\sqrt{\langle N_A\rangle}$. This means that if we have around 10^{20} particles in box A, the actual number of particles we have in said box will vary by a factor of 10^{10} around $\langle N_A\rangle$. This may make $\langle N_A\rangle$ seem useless, but we also have to remember real life experiments. A experiment with a margin of error of 1% is pretty good, and this variation in

particles isn't all that large compared to that. If the width of the probability distribution is small, $\langle N_A \rangle$ is a useful feature in our system. The equation for the width of our probability distribution is:

$$\frac{\delta N_A}{\langle N_A \rangle} = \frac{1}{\langle N_A \rangle}\left[\langle N_A \rangle \left(\frac{V_B}{V}\right)\right]^{\frac{1}{2}} = \sqrt{\frac{1}{\langle N_A \rangle}}\sqrt{\frac{V_B}{V}}$$

For example, for 10^{20} particles, our uncertainty is 10^{-10}, which is much smaller than we can measure experimentally. This was even less of an issue as thermodynamics was being developed, as the experiments were less refined. Additionally, during that time the atomic theory wasn't accepted by everyone. This meant that thermodynamics was based upon the masses of a sample, rather than upon the number of atoms.

For us to continue deriving entropy, we must remember that the system is a composite of 2 subsystems. This can be shown by this function:

$$\Omega_q(N, V) = \frac{V^N}{N!}$$

We can then rewrite our probability distribution into a much more legible form:

$$P(N_A, N_B) = \frac{\Omega_q(N_A, V_A)\Omega_q(N_B, V_B)}{\Omega_q(N, V)}$$

Since the ln function is monotonic, then the maximum of the probability density function and the maximum of the natural log of the probability density function will occur at the same values of N_A and N_B. It is then easier for us to use the natural log of the probability density function than the function itself. This can be seen as:

$$\ln[P(N_A, N_B)] = \ln\left[\frac{V_A^{N_A}}{N_A!}\right] + \ln\left[\frac{V_B^{N_B}}{N_B!}\right] - \ln\left[\frac{V^N}{N!}\right]$$

$$= \ln \Omega_q(N_A, V_A) + \ln \Omega_q\,(N_B, V_B) - \ln \Omega_q(N, V)$$

Here, we can see that the first term in our system is only dependent on Box A, the second term on box B, and the third term is dependent on the whole system.

We can define a function:

$$S_q(N, V) = k \ln\left(\frac{V^N}{N!}\right) + kXN = k \ln \Omega_q(N, V) + kXN$$

K and X are both arbitrary constants. The maximum of this function is the following:

$$S_{q,tot}\,(N_A, V_A, N_B, V_B) = k \ln[P(N_A, N_B)] + S_q(N, V) = \; S_q(N_A, V_A) + S_q(N_B, V_B)$$

As always, $N_B + N_A$ must equal N. In this case, the width of the probability distribution is equal to $\sqrt{\langle N_A \rangle}$, and since it is normalized, the peak must be proportional to $1/\sqrt{\langle N_A \rangle}$. At equilibrium values of N_A and N_B this gives the following approximation:

$$\ln P(N_A, N_B)\,|_{eqil} \approx -\frac{1}{2}\ln\left(2\pi\langle N_A\rangle\langle\frac{V_B}{V}\rangle\right)$$

Since $N > \langle N_A \rangle \gg \ln(\langle N_A \rangle)$, the term $k\ln[P(N_A, N_B)]$ in the $S_{q,\,tot}$ equation is completely negligible during equilibrium, and we can completely ignore it. Therefore, the total equilibrium is the following:

$$S_{q_{tot}}(N_A, V_A, N_B, V_B) = S_q(N_A, V_A) + S_q(N_B, V_B)$$

Since we have found the maximum of the functions for all these variables at equilibrium, we can then derive the maximum entropy from the configuration of the particles. Note that these calculations do not include any calculations for entropy due to momentum. It is also extremely significant that we can add these. The entropy derived by these equations follows Boltzmann's definition of entropy being a log of probability. The fact that the entropy is at its maximum value during equilibrium also happens to prove the second law of thermodynamics. So therefore, the math shown here is also a derivation of the second law of thermodynamics.

Conclusion: Using probability theory, physics, and math, we can see that we can derive the 2nd law of thermodynamics from a relatively simply system. Additionally, this also shows how to very precisely calculate the entropy, given you have that much knowledge about the system. Therefore, shows that many of Boltzmann's theories can be proved – and have been proved.

Citations:

University of Texas. "Number of Microstates." *Numbers of Microstates*, University of Texas, 2014, ch301.cm.utexas.edu/section2.php?target=thermo%2Fsecond-law%2Fmicrostates-boltzmann.html

Eastman, Peter. "Introduction to Statistical Mechanics." *Introduction to Statistical Mechanics - Introduction to Statistical Mechanics*, Stanford, 2014, web.stanford.edu/~peastman/statmech/.

Frig, Roman. "What Is Statistical Mechanics? ." *Roman Frig*, romanfrigg.org.

Swendsen, Robert H. *An Introduction to Statistical Mechanics and Thermodynamics*. Oxford University Press, 2020.

Planck, Max. *Treatise on Thermodynamics: Max Planck*. Longmans, Green, 1927.

Gibbs, J. Willard. *Elementary Principles of Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics*. Yale University Press, 1902.

Mathematical Implications to Modern-day Computer Programming

Maxwell Muscat

University of California, Santa Cruz

Abstract:

Mathematics is used ubiquitously in the world of computer programming. From abstract concepts derived from lectures to specific subjects and equations, mathematics is crucial for those trying to become a programming wizz. The abstract concepts of analytical reasoning and algorithms learned in the math classroom can also be applied when solving problems related to computer programming. More specifically, basic math concepts like algebra and arithmetic are ubiquitous in writing programs, while higher-level concepts have more specific applications. Overall, this essay will reveal how math is used in computer programming, why taking any math class can help one become a better programmer.

Many people would agree if they were told mathematics is needed in computer programming; it has been ingrained in our culture to believe this. However, the follow-up question "why" may make some second guess themselves. Yes, it has indeed been a well known fact that math knowledge is fundamental for computer programming, however many people do not know why. Math is critical for computer programming because it teaches abstract topics that are closely related to computer programming. For example it is used to write complex algorithms and can be applied in problem solving strategies. Mathematics teaches various abstract concepts that will help any and all future computer programmers. Examples of said concepts include analytical skills and algorithms. More specifically, mathematics is needed for writing basic functions (arithmetic and algebra), game design (calculus), and signal processing (differentials and fourier transforms).

From an impartial perspective, it is obvious that math and computer science are not the same subjects. However, when looking at the two abstractly, there are many similarities. First, both mathematics and computer science require a sense of analytical reasoning. According to the article "What is Analytical Thinking and How to Develop", analytical reasoning is the "ability to apply logical thinking to break down complex problems into smaller components in order to solve a problem" (Cole, 2019). Math is a perfect way to improve analytical reasoning because it is dependent on logic. All math problems require our brains to think critically about a problem with limited information to solve, using logic and reasoning to do so. However, how does this relate to computer programming? When looking at the two abstractly, is a math problem and computer program not the same thing? Both math and computer science problems require people to think critically with limitations –in a computer program this closely refers to syntax but can

also include other constraints– and use logic to solve the problem. In mathematics there is a way that is most logical in solving a problem, and this also applies to computer programming; there is always a way to make programs have use less lines of code, thus being more efficient. Good programmers know how to apply what they have learned to solve problems most efficiently, just like mathematicians knowing the best equations and strategies to use to most efficiently solve a math problem.

Another abstract math concept that can be applied when learning computer programming are algorithms. An algorithm is a process or set of rules to be followed in calculations or other problem-solving operations. Algorithms are used all the time in mathematics. From simple algorithms like FOIL, an algorithm to multiply two binomials, to something as complicated as differentials, an algorithm with many applications, almost every "strategy" taught in math is to some extent an algorithm. Learning the basics of these algorithms is crucial to understanding program algorithms. With the help of the analytical thinking gained by taking math classes all their lives, computer programmers can further understand what an algorithm is, and know the method of learning and applying strategies to solve problems both in class and in the real world, or in this case the virtual one.

In addition to math's abstract applications to programming, there are also many straightforward ways in which mathematics can be used in computer programming strategies. Lower level maths such as arithmetic and algebra are ubiquitous in computer programming. From simple tasks like finding whether an index is odd or even to more complicated tasks such as generating prime numbers, arithmetic and algebra are bound to be involved. In fact, Keith Perry, author of the article "What Types of Math Do Computer Programmers Use?" states that

"addition, subtraction, multiplication and division is used in almost every program written" (2019). However, this does not really answer the question at hand. Algebra and arithmetic are math subjects that can be applied in many cases, so it makes sense that all students must take it; so what about higher levels of math? At the University of California Santa Cruz, for example, computer science students are required to take math classes beyond calculus two in order to declare their major, but why is that? Calculus is a high level math that, for most CS majors, is the first step to even just declaring their major. Calculus is built off of few ideas, two of which being derivation and integration. One real world application of this is physics, more specifically position, velocity, and acceleration. Taking the derivative of the position formula gets one velocity, which can then get acceleration when taken again. The reverse is true for integration. In the article "What Math Skills do I need for a Game Design Degree" the author states that "Because time and motion are involved in a game loop, programmers … use continuous math concepts from calculus" (2020). When the author mentions "motion," they are talking about physics; which is a big deal in game design. Almost all games require some deal of physics, and calculus is used to create it through the three equations listed previously.

Another example of how math is used in computer programming is fourier transform and signal processing. The fourier transform is a "tool that breaks a waveform (a function or signal) into an alternate representation, characterized by sine and cosines" (Bevelacqua, 2010). In his book "Applied Underwater Acoustics," author D.A. Abraham states that signal processing "condenses measurements to extract information about some distant state of nature" (2017, p. 743). Fourier transform is critical in signal processing because it transforms whatever needs to be analyzed into something that can be analyzed using signal processing. Some signals can also

only be processed in its digital or analogue form, and fourier's transform helps alternate between the two.

In conclusion, maths will be used in almost any program written. The abstract concepts of analytical reasoning and algorithms are critical when wanting to learn how to program because those same abstract skills are needed. When looking at specific examples, it was found that math from all levels are needed in computer programming, from algebra and arithmetic all the way up to calculus and the fourier transform. However, the higher the level of math learned, the less application it will have. Once in college, students will quickly learn that they will have to take specific math classes based upon what they want to do specifically with their computer science degree.

References:

Abraham, D.A. *Applied Underwater Acoustics*. Elsevier, 2017.

Bevelacqua, Pete. "Fourier Transforms." *Fourier Transform*, 2010, www.thefouriertransform.com/.

Cole, Lauran. "What Is Analytical Thinking Ability, How to Develop?: MentalUP." *MentalUP.co*, MentalUP, 16 Oct. 2017, www.mentalup.co/blog/what-is-analytical-thinking-ability-and-how-to-develop.

Harkiran. "What Is the Importance of Mathematics in Computer Science?" *GeeksforGeeks*, 2 Aug. 2019, www.geeksforgeeks.org/what-is-the-importance-of-mathematics-in-computer-science/.

Perry, Keith. "What Types of Math Do Computer Programmers Use?" *Techwalla*, 2019, www.techwalla.com/articles/what-types-of-math-do-computer-programmers-use.

"What Math Skills Do I Need for a Game Design Degree?" *Online Game Design Schools*, www.onlinegamedesignschools.org/faq/math-skills-game-design-degree/.

Sphere Tracing

Adrian Guerra

March 12, 2020

Abstract

Sphere tracing is a type of ray tracing which represents the object to be rendered as a signed distance function. Sphere tracing is well suited to rendering pathological surfaces, since it can render surfaces with discontinuous or undefined derivatives.
Sphere tracing is more accurate than rasterization and traditional ray tracing, which both require objects to be approximated as polygon meshes.
Sphere tracing can be considerably faster than traditional approaches which require a polygon mesh for more dynamic objects whose geometry is continuously changing in non-trivial ways.

Introduction

Sphere tracing is a type of ray tracing which represents objects using signed distance functions. Sphere tracing involves marching a ray towards the object being rendered by repeatedly sampling its distance function. Distance functions can be modified and combined using various operations, which can be used to build up complicated shapes from the distance functions of simple primitives.

1 Signed Distance Functions

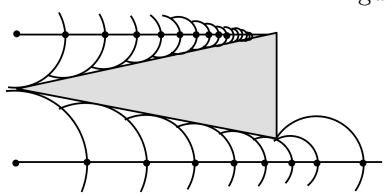
1.1 Definition

For an object A, the corresponding signed distance function f_A is a function such that

- $|f_A(x)|$ = the distance between x and the closest point to x on the surface of A
- $f_A(x) > 0$ for all points x outside of A
- $f_A(x) \leq 0$ for all points x inside A
- $f(x) = 0$ for all points x on the surface of A

In other words, $A = \{x : f(x) \leq 0\}$.

Figure 1: Hit vs. Miss



(Hart 1995)

2 Sphere Tracing

2.1 Ray-Signed Distance Function intersection

Given a ray with starting position $p_0 \in \mathbb{R}^n$ and (normalized) direction vector $d \in \mathbb{R}^n$, and an object A represented by a signed distance function $f_A : \mathbb{R}^n \to \mathbb{R}$, we can determine whether the ray intersects A based on the limit of the sequence

$$p_{i+1} = p_i + d \times f(p_i)$$

If the sequence diverges, then the ray did not intersect.

If the sequence converges, then the ray did intersect, and the location of the intersection is the value that the sequence converged to.

3 Operations on Signed Distance Functions

Signed distance functions for more complicated shapes can be built up from the distance functions of basic primitives.

This document includes some examples of the operations that can be performed on signed distance functions. More comprehensive lists can be found in Quilez 2008 and Hart 1995.

3.1 Set Operations

Given two objects A and B, represented by two distance functions f_A and f_B respectively, set operations on A and B can be represented as the following operations on f_A and f_B.

$$f_{A \cup B}(x) = \min(f_A(x), f_B(x))$$

(Quilez 2008)

$$f_{A \cap B}(x) = \max(f_A(x), f_B(x))$$

(Quilez 2008)

$$f_{A-B}(x) = \max(-f_A(x), f_B(x))$$

(Quilez 2008)

3.2 Rotation & Translation

Given a function $t : \mathbb{R}^n \to \mathbb{R}^n$ that performs rotation or translation, and a signed distance function $f_A : \mathbb{R}^n \to \mathbb{R}$ representing an object A, we can construct a signed distance function $f_B : \mathbb{R}^n \to \mathbb{R}$ representing an object $B = \{t(x) : x \in \mathrm{A}\}$

$$f_B(p) = f_A\left(t^{-1}(p)\right)$$

(Quilez 2008)

3.3 Scale

Given a signed disance function $f_A : \mathbb{R}^n \to \mathbb{R}$ representing an object A, and a scale factor s, we can construct a signed distance function $f_B : \mathbb{R}^n \to \mathbb{R}$ representing an object $B = \{x \times r : x \in \mathrm{A}\}$

$$f_B(p) = f_A\left(\frac{p}{s}\right) \times s$$

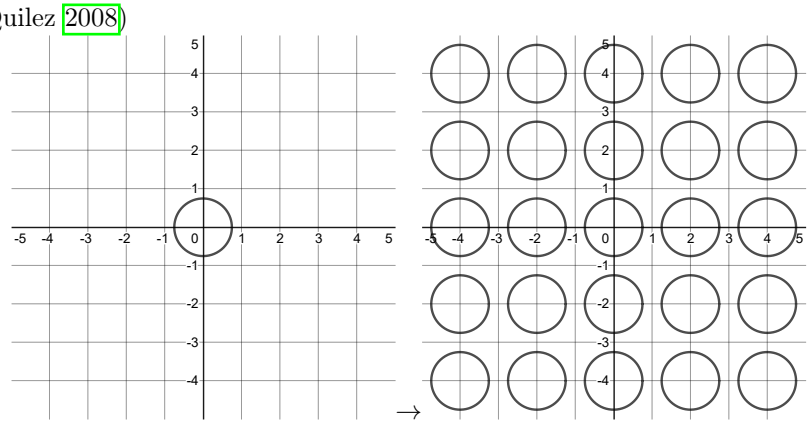
(Quilez 2008)

3.4 Infinite Repetition

Given a signed disance function $f_A : \mathbb{R}^n \to \mathbb{R}$ representing an object A, and a tuple $c \in \mathbb{R}^n$ representing the bounds of the domain to be repeated, we can construct a signed distance function $f_B : \mathbb{R}^n \to \mathbb{R}$ representing an object B defined as the infinite tiling of A bounded in c.

$$f_B(p) = f_A\left(\mathrm{mod}\left(p + \frac{c}{2}, c\right) - \frac{c}{2}\right)$$

(Quilez 2008)



References

Hart, John (June 1995). "Sphere Tracing: A Geometric Method for the Antialiased Ray Tracing of Implicit Surfaces". In: *The Visual Computer* 12. DOI: 10.1007/s003710050084.

Quilez, Inigo (2008). *distance functions*. http://iquilezles.org/www/articles/distfunctions/distfunctions.htm.

Gender Equality in Mathematics Higher Education

Jack Campbell & Josephine Ward
(Group 32)
MATH 19B-Winter 2020
Morales
11 March 2020

ABSTRACT

The purpose of this study was to examine the gender differences in the mathematics faculty of universities within the United States. Gender inequality in mathematics has been prevalent since the dawn of mathematics and continues to persist into our modern era. Previously the lack of representation stemmed from the oppression and lack of opportunity for women, but now is largely reliant on the embedded culture and norms of STEM, mathematics in particular, being a “man's domain”. This study examines thirty five universities from 11 states. Samples included the primary public university system for California, Maine, South Carolina, Texas, and Nebraska as well as the Ivy League schools. The sample was taken using publicly available information on the Universities’ websites. Analysis was conducted based on the percentage of the mathematics department faculty (professors and lecturers) who identified as female. The cumulative average percentage female for all 35 schools was 23.7% with a maximum of 66.6% and a minimum of 0%. The data shows a weak positive correlation between the year founded and percentage of the mathematics department being female. It also showed a weak correlation of lower ranked schools having higher percentages of women. Finally, the departments with the highest percentages (30-60%) of women were consistently the smallest departments, averaging less than 15 faculty members while the departments with 30-200 faculty members percentage of female faculty was significantly lower (5-15%). Future research should include a broader sample and more research into the effects of the lack of representation as well as mitigation efforts that could be undertaken to even out the field.

INTRODUCTION

Hebert Turnballs' book, *Great Mathematicians*, lists the dozen most important mathematicians from history and describes their impacts on the world. Not one of them is a woman.[1] If we were to consider common historical knowledge, teaching, and even textbooks, it would appear as if women were absent in the realm of mathematical study until our very recent modern era. This is decidedly untrue. Looking deeper, we are able to identify dozens of genius women from Hypatia in 300 AD to Amalia Noether in the mid-to-late 1900s who redefined the fields of calculus, astronomy, physics, and so much more. However, the names of Maria Agnesi, Emilie du Châtelet, Caroline Herschel, Sophie Germain, or Mary Fairfax Somerville are not household names as Plato and Gauss are, despite often working in tandem with these men, or working tirelessly on unique discoveries equally as important. Unfortunately, even when these women are finally given recognition in more modern books and research, they are often first described in terms of their beauty, next by their relation to some famous man, third by their notable scandals and lovers, and then finally by their mathematical prowess and achievements. Claudia Henrion describes this irritating situation perfectly: "How often do you see a memorial of a male mathematician that says 'although he did good mathematics, his physical appearance left something to be desired'? Yet in her case it was considered relevant and important"[2].

All of these historical women had to overcome great, unimaginable by today's standards, barriers to receive their education. Due to their gender alone, their social class, marital status, and parental status confined their opportunities. Those who did manage to enter the field of mathematics were often ostracized, seen as unnatural, or forced to be independent and self-reliant, issues that men of these times did not face while attempting a mathematics degree or research. Yet despite this struggle and relentless perseverance, their work and names are often lost to history. However, regardless of how many influential women in mathematics one can name in any period, it is important to remember that, "By glamorizing the exceptional case, we often manage to preserve the myth of equality in education"[3]. These women were the exception, not the rule. For centuries, the average, or even exceptional, woman was unable to obtain an education beyond basic reading and writing, earn degrees, or hold any semblance of a teaching or research position at a university.

These archaic rules and restrictions seem like relics of the past, however the culture that dictated these restrictions remain even today. Women are still

---

[1] Turnbull, Herbert W. *The Great Mathematicians*. 1929.
[2] Henrion, Claudia. *Women in Mathematics: The Addition of Difference*. 1997. P. 68
[3] Olsen, L. M. *Women in Mathematics*. MIT. 1974. P. 164

incredibly underrepresented in higher mathematics. Even with official bans on women attending universities gone, women still face social, economic, and most importantly cultural barriers to this field of study. Studies dating back to the 1970's show that the lack of women in mathematics starts with high school. Male students are far more likely to receive encouragement from peers, teachers, parents, and counselors to pursue mathematics compared to female students. [4] This perpetuates a cultural stigma that diminishes the confidence of female students and creates a "sex-role identity"[5] perception of math being a male subject. This cultural perception then in turn leads to social barriers and a "conflict between academic achievement and popularity"[6]. All of this culminates into fewer women pursuing a higher education in mathematics, to this day.

This gap has been steadily closing in recent decades, but as our and many others' research will show, it is far from gone. Claudia Henrion interviewed dozens of women who still teach and research mathematics today and found that they all faced issues of being viewed as a "women mathematician, not just a mathematician"[7] and many of the women expressed that "When a woman mathematician enters a room, attends a meeting, goes to a conference, or applies for a job, the first thing that is noticed is that she is a woman"[8] , once again not something men have to cope with. Woman is not the "default assumption".[9] Furthermore, more often than not, women are forced to put far more time and effort into how they present themselves in order to be respected. Men are allowed to be aloof, eccentric, ambitious, loud, or forceful without being disregarded while women are not.[10] Furthermore, *all* of the women interviewed expressed the absolute necessity of keeping their personal lives separate from their professional, all the while still coping with the expectation and burden of being wives and mothers. To this day, the expectations of female involvement in childrearing outweighs that of men, regardless of employment status. We often see women being lauded as "supermoms'' in the media for maintaining full time employment and a disproportionate role in their children's lives. Lack of access to child care continues to be a significant barrier to women succeeding in all fields, not just mathematics. As Henrion puts it, "Women should not have to be heroic in order to be mathematicians AND have children"[11].

In our research we took a small sample of American universities, examined their mathematics faculty, and calculated what percentage of said faculty were

---

[4] Chipman, Susan F, et al., editors. *Women and Mathematics: Balancing the Equation*. Lawrence Erlbaum, Inc, 1985. P. 96
[5] Chipman, p. 103
[6] Chipman, p. 62
[7] Henrion, p. 43
[8] Henrion, p. 70
[9] Henrion. p.70
[10] Henrion, p.72
[11] Henrion, p.85

female. Due to the American population being roughly 50% and in 2020 now OVER 50% of university students are female, the expected values were anticipated to be roughly 50% female in each department. As our research shows, this hypothesis was inaccurate.

METHODOLOGY

When faced with a systemic issue as massive and widespread as gender inequality in mathematics, it is necessary by confines of time, resources, and breadth to limit the scope of research. Our goal was to take a sample that was precise and representative of the nation as a whole. In this particular case, our total population was all American collegiate mathematics departments faculty.

We blocked the country by state and into strata of regions (i.e. West coast, Midwest, South, etc). A representative state was chosen at random from each strata, with the exception of the West Coast region, which California was selected on the basis of this being our home institution, the largest state in the country, and widely regarded as containing one of the most prestigious, liberal, and massive public university systems. The states selected for our representative sample were California, Maine, South Caroline, Texas, and Nebraska. For each state, the primary public university system (comparable to our own UC system) was selected as the sub-population. Additionally, the Ivy league schools were sampled as a point of reference comparing the "elite" private schools to public universities. 35 schools in total were sampled.

The main variable considered in this study was the number of female faculty members in the Universities mathematics department. For the purpose of this research we define faculty to mean full time professors, part time professors, associate professors, visiting professors, and lecturers. Essentially, any person with the word "professor" in their job title (excluding emeritus and in memoriam professors) or lecturer was included in this sample. Full time university researchers who do not teach were not included.

Our method of data collection was to use each universities' mathematics department's website. Using the faculty directory, the number of total faculty members (defined above) were counted by hand. Next, the number of female faculty members were counted. A potential source of error, separate from the general human error factor of counting by hand, is the theoretical misgendering of faculty members if their gender was not explicitly stated in bios, websites, or essays. Faculty members not listed on websites were beyond our ability to count in this study. Multiple samples taken did not include lecturers in their directories as indicated with a dash in the data sample below.

Data was consolidated into tables[12] and the percentage of female faculty members was found by dividing the total number of identified women in the department by the number of total faculty members in the department. Further

[12] See fig 1-6 below

data was gathered on each university's overall US ranking, math department ranking, and year founded to identify trends. Several schools' rankings were unavailable or inaccessible with our resources. An overall average percentage was also taken for each state as well as a cumulative average for all 35 departments selected.

Analysis is conducted by comparing the percentage of female mathematics faculty to the other universities within the state as well as universities outside of the state. Graphs[13] comparing rankings and year founded to percentage were also created to identify trendlines (excluding the unranked schools).

[13] See figs. 7-10

RESULTS

35 schools → 23.73% average

Fig 1. California

CALIFORNIA

School	Overall US Ranking	Math Dept. Ranking	Year Established	# of female professors	# of female lecturers	Total Faculty	% Female
UC Los Angeles	20	7	1919	23	-	182	12.6
UC Santa Cruz	84	71	1965	4	1	34	14.7
UC Santa Barbara	34	39	1891	7	0	50	14
UC San Diego	37	19	1960	14	-	104	13.4
UC Merced	104	144	2005	8	3	30	36.6
UC Riverside	91	71	1907	9	0	56	16.07
UC Irvine	36	39	1965	7	2	84	10.7
UC Berkeley	22	2	1868	18	2	147	7.35
UC Davis	39	34	1905	9	2	60	18.33
CA Average:							**15.8%**

Fig 2. Texas

TEXAS

School	Overall US Ranking	Math Dept. Ranking	Year Established	# of female professors	# of female lecturers	Total Faculty	% Female
UT Arlington	293-381	136	1895	7	14	58	36.2
UT Austin	48	n/a	1883	12	9	92	13
UT Dallas	147	117	1969	9	8	69	24.6
UT El Paso	293-381	n/a	1914	7	9	59	27.11
UT Rio Grande Valley	293-381	n/a	2013	8	15	83	27.7
UT San Antonio	293-381	n/a	1969	3	10	52	25
Texas Average:							**25.6%**

Fig 3. Maine

MAINE

School	Overall US Ranking	Math Dept. Ranking	Year Established	# of female professors	# of female lecturers	Total Faculty	% Female
UM Maine	202	n/a	1820	3	3	30	20
UMA Augusta	Not ranked	n/a	1965	2	3	8	62.5
UMA Farmington	Not ranked	n/a	1864	4	-	13	30.7
UMA Presque Isle	Not ranked	n/a	1903	0	0	1	0
UMA Southern Maine	Not ranked	n/a	1878	2	10	26	45.15
Maine Average:							**31.66%**

Fig 4. South Carolina

SOUTH CAROLINA

School	Overall US Ranking	Math Dept. Ranking	Year Established	# of female professors	# of female lecturers	Total Faculty	% Female
USC Columbia	104	86	1801	3	-	38	7.89
USC Aiken	Not ranked	n/a	1961	2	2	19	21.05
USC Beaufort	Not ranked	n/a	1959	2	2	10	40
USC Sumter	Not ranked	n/a	1966	1	1	3	66.6
USC Average:							**33.3885%**

Fig 5. Nebraska

NEBRASKA

School	Overall Ranking	Math Dept. Ranking	Year Established	# of female professors	# of female lecturers	Total Faculty	% Female
UN Kearney	Not ranked	n/a	1905	2	4	12	50
UN Lincoln	139	66	1869	11	-	38	28.9
UN Omaha	293-381	n/a	1908	5	3	23	34.7
Nebraska Average:							**37.8%**

Fig 6. The Ivy League

THE IVY LEAGUE

School	Location	Overall US Ranking	Math Dept. Ranking	Year Established	# of female professors	# of female lecturers	Total Faculty	% Female
Brown University	Providence, Rhode Island	14	14	1764	4	1	28	17.8
Columbia University	NYC, New York	3	7	1754	7	-	49	14.28
Cornell University	Ithaca, New York	17	13	1865	3	3	41	14.63
Dartmouth College	Hanover, New Hampshire	12	53	1769	6	-	23	26
Harvard University	Cambridge, Massachusetts	2	2	1636	2	4	26	23
Princeton University	Princeton, New Jersey	1	1	1746	7	7	170	8.2
Yale University	New Haven, Connecticut	3	9	1701	3	2	57	8.7
University of Pennsylvania	Philadelphia, Pennsylvania	6	16	1740	4	4	60	13.3
Ivy Average:								**15.7%**

Fig 7. Graph - year founded vs percentage female

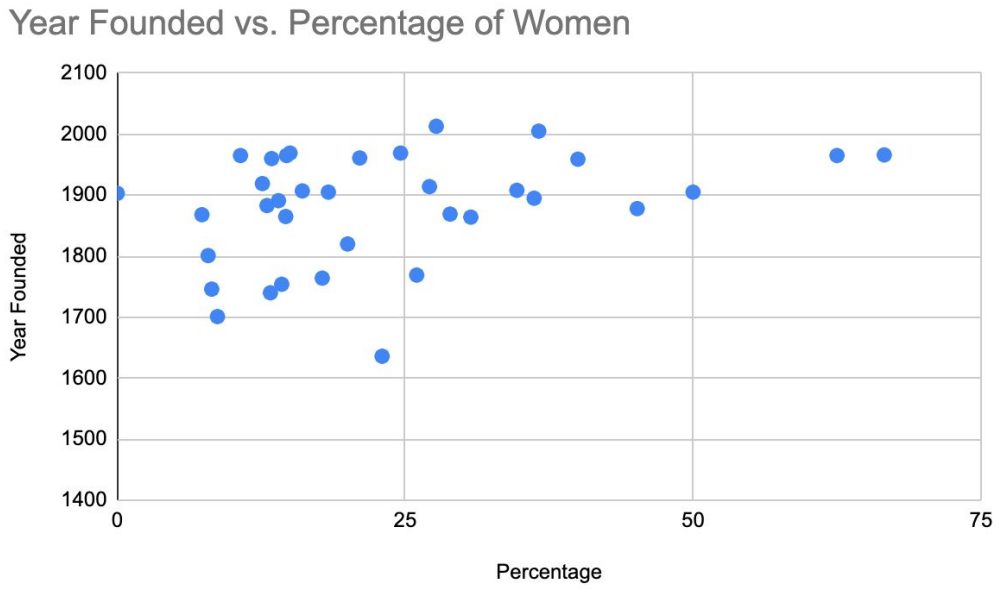
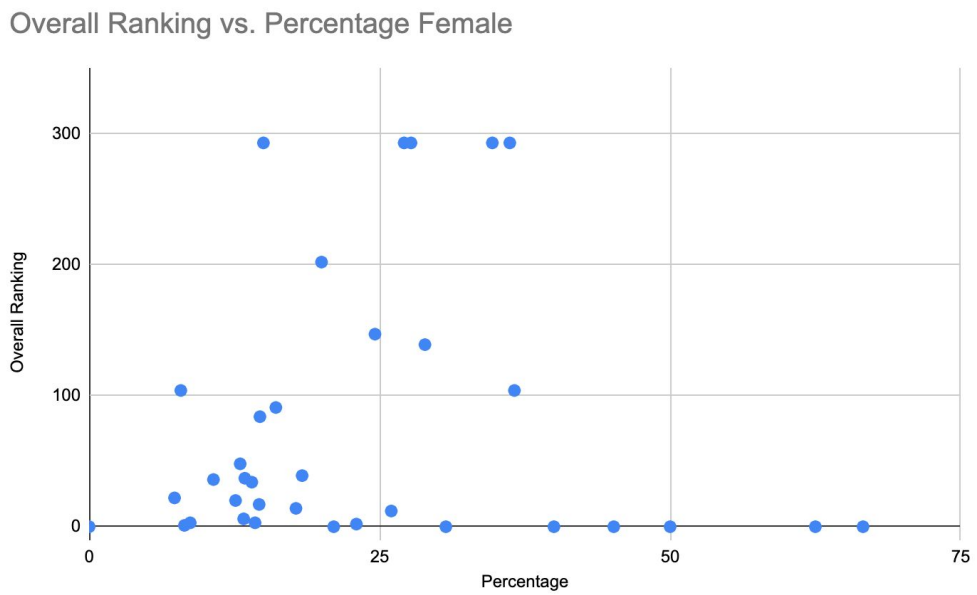
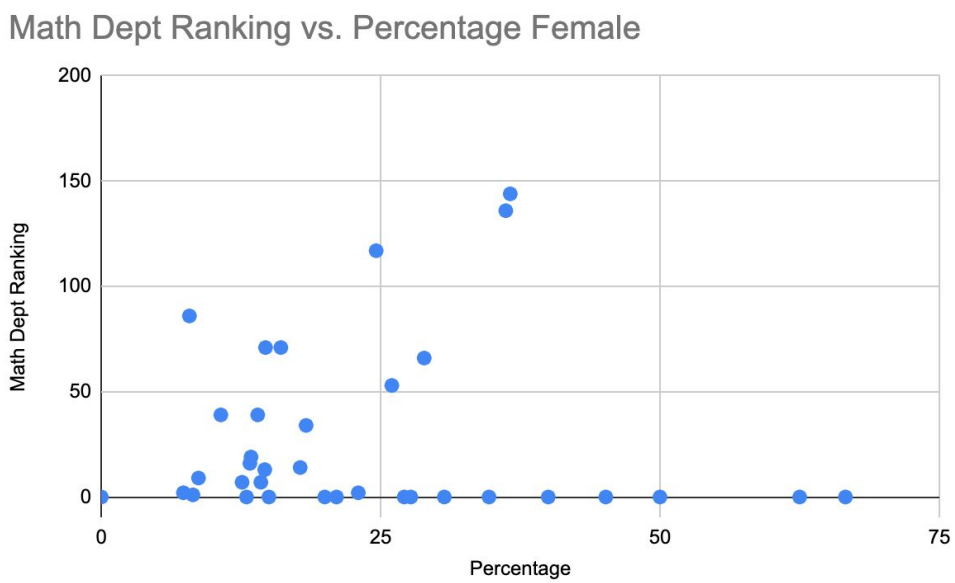


Fig 8. Graph - Overall US ranking vs percentage female



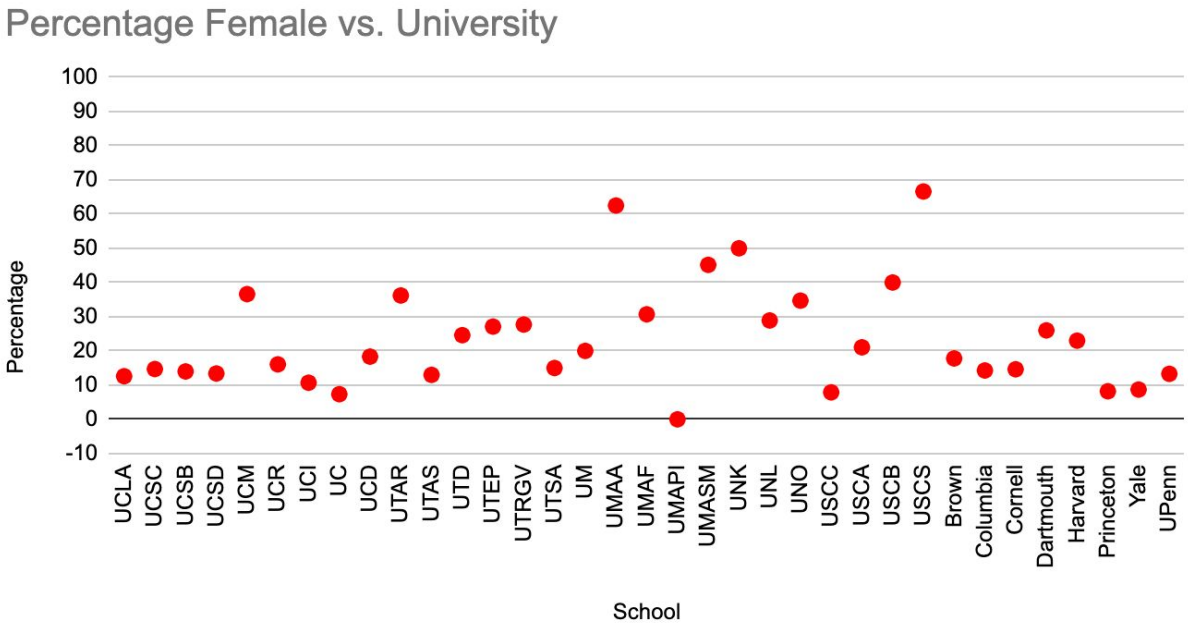
**Note: 0 ranking (lies on x-axis) = unranked by our resources*

Fig 9. Graph - math department ranking vs percentage female



**Note: 0 ranking (lies on x-axis) = unranked by our resources*

Fig 10. Graph - Percentage female vs University



DISCUSSION

The results of this study were ultimately disheartening. Based on the gender distribution of the United States as well as the average gender distribution of universities, we expected the percentages to be roughly 50%. However, for the mathematics departments represented, this is not so. We found in our research that the average math department in the United States is less than 24% female.[14] As faculty are in theory a representation of student populations, this traces back to a lack of female student participation in higher education mathematics. While the percentage of female students from each university was not taken, an expected trend for co-ed schools is for the general student body to be at or above 50% female, but for the math departments to be significantly lower. Faculty are exclusively those holding degrees in mathematics, causing the amount of female mathematics students to be the largest indicator of the number of female faculty.

After compiling all of our sample data, we found the state of Nebraska to have the highest average percentage of women of female faculty in the mathematics department with 37.8%. Following this, we found South Carolina to be second with 33.39%, Maine with 31.66%, Texas with 25.6%, California with 15.8%, and lastly the Ivy League with 15.7%. The highest university women/math faculty percentage that was found was, despite (or perhaps because of) its small size, University of Maine Augusta with 62.5%. The lowest university women/math faculty percentage that was found was University of Maine Presque Isle with 0%. We are going to set aside this statistic due to the universities math program only consisting of one faculty member. The next lowest percentage of women in math is the University of California Berkeley with 7.25%. Finally, the departments with the highest percentages (30-60%) of women were consistently the smallest departments, averaging less than 15 faculty members while the departments with 30-200 faculty members percentage of female faculty was significantly lower (5-15%).

Other factors we also took into account were the year the university was founded and its rankings. We expected the newer schools to show a higher percentage of women because of recent increased social pressure to stimulate equality. Our data proved this to be reasonably accurate, if not to the strength anticipated. As for rankings, it was expected to see the higher ranked schools, such as the Ivy League, to have a more even distribution of gender due to increased public scrutiny and media attention. To the extent we were able to collect, this mostly proved false. The highest ranked schools generally had the lower percentages of women in their mathematics departments. However, trend lines are

[14] See results

difficult to ascertain as many of the school's rankings were beyond our capabilities to find, causing them to be excluded from the data set.

There were also potential sources of error when gathering and computing the given data. Human error is one factor with potential miscalculations being made and accidental inaccuracies in the transfer of data from university websites to spreadsheets. Additionally, many of the websites information was pulled from did not include photos or full lists of their faculty. We realized this could lead to many types of error including misgenderment, undercoverage, as well as a decrease in the overall accuracy of our data. Additionally, given our data was sampled on February 18th, 2020, it is beyond our ability to account for potential changes in faculty positions between quarters or semesters. Lastly, due to our small sample size, anomalies on both ends have the potential to skew the data set.

Due to the limitations of research already discussed, the scope of this project was not as extensive as it could have been. For further research, it would be ideal to take a much larger sample in order to more accurately reflect the given population. Additionally, this research focused primarily on the end result of diminished female participation, not the root causes. Further research should investigate not only the historical trends that have barred the majority of women from excelling in mathematics, but also the current barriers still in place, looking into why women still are not proportionally represented in today's lecture halls. Furthermore there were significant limitations and intentional oversights in how the faculty was grouped based on professorial status. For the purposes of this research, all types of professors and lecturers (excluding emeritus and in memoriam) were considered equal in status and representation. In reality, there are significant differences in tenure status, full time, part time, and seniority. Future research should delve into specific gender ratios for the individual components and identify trends in department advancement.

This research is incredibly relevant and important as academic equality is something that should be continually strived for until it is achieved. Our data shows that despite what many think, equality has not been satisfied. Historical biases and barriers must be accounted for and overcome. Previous studies[15] indicate no difference in mathematical ability between boys and girls and that the only differences in higher education for men and women is the participation rate. This is research that should be continued in the future and solutions discovered to make up for the tragic disparity.

[15] Chipman, et al

REFERENCES

Turnbull, Herbert W. *The Great Mathematicians*. 1929.

Olsen, L. M. *Women in Mathematics*. MIT. 1974.

Chipman, Susan F, et al., editors. *Women and Mathematics: Balancing the Equation*. Lawrence Erlbaum, Inc, 1985.

Henrion, Claudia. *Women in Mathematics: The Addition of Difference*. 1997.

Gallagher, Ann M, and James C Kaufman, editors. *Gender Differences in Mathematics*. The University of Cambridge, 2005.

Daniel Bojin, Fan Xia, and Sean Osborne

Professor Morales

Math 19B

11 March 2020

Carbon Dioxide Emission Rates in the US From 1940-2019

Abstract:

The purpose of our project is to track the different sources of carbon dioxide emissions over time to determine any trends in the data found. Since there are many contributing factors to carbon emissions, the main sectors tracked were transportation, agriculture, stationary, electricity, and residential/commercial sources. To find trends, the data for carbon dioxide emissions in metric tons were researched and recorded for five years: 1940, 1960, 1980, 2000, and 2019. A graph was then compiled to demonstrate how carbon emissions have changed over time and from what sources.

Introduction:

Throughout the years, global warming has been an increasing cause for concern that can be directly linked to carbon dioxide emissions from human activity. As technology improved, it assisted in the dramatic growth of the population which correlates to the increasing carbon emission rates. Between 1940 and 2019, there were several historical and important events that affected the rates of these emissions, including World War II, the population boom, and also a large push for the environmental movement as well as a fight against climate change. Since there is not much data online compiling statistics for the amount of carbon dioxide produced from different sources and showing trends throughout the decades, this became the objective for our project.

Methodology:

Overall:

To find the amount of carbon dioxide emissions during the five different years, we separated the sources into five different sectors: transportation, agriculture, stationary, electricity, and residential/commercial sources. For each individual sector, we researched and estimated the

total amount of carbon dioxide emissions in metric tons per year. This was done either by finding the data directly from published research journals or estimating the amount of CO_2 emissions using conversion factors or graphs.

*Since methodology for each source is so different, individual methodologies of calculating carbon emissions are attached below and explained in greater detail

Methodology for Agriculture:

To determine the amount of CO2 emissions from the agriculture sector, I first researched what contributed to carbon dioxide emissions. I found that although there were several factors in agriculture that produced greenhouse gases, only crushed limestone/dolomite and urea fertilizer emitted CO2. To find the metric tons of carbon dioxide emissions each year, I first found a conversion factor between tons of crushed limestone/dolomite to metric tons of carbon dioxide. Likewise, I did the same with urea fertilizer. I found data from a variety of scientific research sources or government websites that estimated the tons of crushed limestone or urea fertilizer produced in 1940, 1960, 1980, 2000, and 2019. Some data did not date back far enough or close enough to 2019 so to combat this, I created graphs with data points from known amounts of limestone/dolomite and urea fertilizer and created a line of best fit.

[1] https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_053252.pdf
[2] https://www.usda.gov/oce/climate_change/Quantifying_GHG/Chapter3S.pdf
[3] https://www.energy.gov/sites/prod/files/2013/11/f4/stone.pdf
[4] https://www.usgs.gov/centers/nmic/crushed-stone-statistics-and-information
[5] http://www.rocksandminerals.com/aggregate.htm
[6] https://www.dnr.wa.gov/publications/ger_b52_limestone_res_western_wa_1.pdf
[7]https://books.google.com/books?id=jaTJzL5y0LQC&pg=PA779&lpg=PA779&dq=tons+of+crushed+limestone+US+1980&source=bl&ots=9rX_k8Sl2M&sig=ACfU3U0HgaoN9thHTkRPm6sqc17U3KCkBg&hl=en&sa=X&ved=2ahUKEwjXj_-to-fnAhWaGDQIHVI-CWYQ6AEwEnoECAoQAQ#v=onepage&q=tons%20of%20crushed%20limestone%20US%201980&f=false
[8] https://www.usda.gov/oce/climate_change/Quantifying_GHG/Chapter3S.pdf
[9] https://www.ers.usda.gov/data-products/fertilizer-use-and-price.aspx
[10] https://naldc.nal.usda.gov/download/CAT78694121/PDF
[11] https://www.earth-syst-sci-data.net/10/969/2018/essd-10-969-2018.pdf
[12] https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions
[13] https://www.epa.gov/sites/production/files/2017-12/documents/ag_module_users_guide.pdf

Methodology for Electricity:

To calculate the amount of carbon dioxide emitted from electricity use, I first found a conversion factor from kWh to pounds and then converted that value to metric tons. From the US Energy Information Administration, I was able to find a data table of total electricity used in the US for the past few decades. Since the data did not date back far enough to the 1940s, I plotted the known data on a graph for the years close to 1940 and created a line of best fit to estimate a number.

[1] https://www.eia.gov/tools/faqs/faq.php?id=74&t=11
[2] https://www.eia.gov/energyexplained/electricity/electricity-in-the-us.php

Methodology for Residential/Commercial :

For residential and commercial carbon dioxide emissions, I was able to find a graph of the total emissions for the past few decades. To find the total carbon emissions, I added the CO2 emissions from the residential and commercial sectors. Since the data did not date back far enough to 1940, I created a graph with known data points near 1940 and created a line of best fit to estimate. Additionally, no information for the year 2019 was published and because the recent years do not have a clear trend in the direction it was heading, I decided to choose a data point near 2019, 2018, to estimate.

[1]https://www.statista.com/statistics/204879/us-carbon-dioxide-emissions-by-sector-since-1950/

Methodology for Transportation:

In order to calculate the carbon emission for each vehicle type, I first used a converging factor of carbon dioxide emission per gallon of gasoline. This factor is the same for each vehicle type since they all use the same type of gasoline (this is typically not true but let's assume that it is for less complication). Then I needed to find miles per gallon for each vehicle type since some sources online only provided the average amount driven (km or miles) per singular vehicle. This also differs every year (1940, 1960, 1980, 2000, and 2019) since models for cars, airplanes and heavy duty trucks always change. Not always for the better, hence the slow decrease in mpg (miles per gallon) for heavy duty trucks as time goes on. Finally, I needed to find the amount of cars, for example, that were used for each year, in order to find the total metric tons of carbon dioxide every year. This information was provided in one of the previous sources I found to determine the quantity for each vehicle every targeted year. Since heavy duty trucks and general

aviation/airplanes weren't particularly used by the general public during the 1940, I only included the amount used by the military during WW2.

[1] https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P100WUHR.pdf
[2]https://www.epa.gov/regulations-emissions-vehicles-and-engines/regulations-greenhouse-gas-emissions-commercial-trucks
[3] https://www.eia.gov/totalenergy/data/annual/showtext.php?t=pTB0208
https://theicct.org/sites/default/files/publications/ICCT_Aircraft-FE-Trends_20150902.pdf
[4]https://business.edf.org/insights/green-freight-math-how-to-calculate-emissions-for-a-truck-move/
[5]https://e360.yale.edu/digest/transportation-replaces-power-in-u-s-as-top-source-of-co2-emissions
[6] https://www.bbc.com/news/science-environment-49349566
[7] https://www.bts.gov/content/number-us-aircraft-vehicles-vessels-and-other-conveyances

Methodology for Stationary Sources (by Sean):

Finding the stats for the stationary sources was fairly difficult. To start, I had to narrow down and generalize which sources I would find stats for, as being specific and finding stats for every single source would've been nearly impossible given our limited resources. In order to find my stats, I would often have to find the overall amount of greenhouse gases released in the U.S for each of the given years (1940, 1960, 1980, 2000, and 2019). I'd then used official statements by the Environmental Protection Agency (or the EPA) that stated what percentage of those greenhouse gases were specifically carbon dioxide, as well as statements and graphs stating what percentages of those greenhouse gases came from my specific sources, like industry and power plants. Using those, I could then estimate how many metric tons of carbon dioxide were released during each year. If I found links that gave those specific numbers, I would use those instead, but that wasn't too common since finding those numbers were often difficult, especially for the earlier years. In some instances, I would have to lowball my calculations using trends that stayed constant, while making sure they still aligned with our other calculations.

[1] https://www.eia.gov/tools/faqs/faq.php?id=77&t=11
[2] https://www.epa.gov/ghgemissions/overview-greenhouse-gases
[3] https://cfpub.epa.gov/ghgdata/inventoryexplorer/#industrialprocesses/allgas/source/all
[4] https://www.eia.gov/environment/emissions/carbon/
[5] https://www.macrotrends.net/countries/USA/united-states/carbon-co2-emissions
[6https://www.reference.com/science/main-sources-carbon-dioxide-atmosphere-2a0cdf8c2919dd14
[7] https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions

Results:

- **Agriculture:**

Background info on liming:
https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_053252.pdf

https://www.usda.gov/oce/climate_change/Quantifying_GHG/Chapter3S.pdf : (directly from the usda so more accurate or reliable I guess?)
This results in a carbon coefficient or emission factor of 40/1000= -0.04kg C per kg CaCO3 ⇐ this is for liming

Equation 3-23: Change in Soil Carbon Stocks from Lime Application

$$\Delta C_{Lime} = M \times EF \times CO_2MW$$

Where:

ΔC_{Lime}	= Annual change in soil carbon stocks from lime application (metric tons CO_2-eq)
M	= Annual application of lime as crushed limestone or dolomite (metric tons of crushed limestone or dolomite year^{-1})
EF	= Metric ton CO_2 emissions per metric ton of lime -0.04 (metric ton carbon (metric ton lime)$^{-1}$)
CO_2MW	= Ratio of molecular weight of CO_2 to carbon (44/12) (metric tons CO_2 (metric tons C)$^{-1}$)

Liming of soils and urea fermentation create CO2 emissions but greenhouse gases are far beyond just this, especially in agriculture sector
https://www.energy.gov/sites/prod/files/2013/11/f4/stone.pdf
75% of crushed stone is limestone/dolomite
1.7 billion tons of crushed stone in 2000
https://www.usgs.gov/centers/nmic/crushed-stone-statistics-and-information
1.14 billion tons for the first 9 months of 2019
(1.14/9)(12)= 1.52 billion tons in 2019
1.52(0.7)= 1.064 billion tons of limestone/dolomite
1.2 billion tons for the first 9 months of 2000
(1.2/9)(12)=1.6 billion tons in 2000
1.6(0.7)= 1.12 billion tons of limestone/dolomite
http://www.rocksandminerals.com/aggregate.htm
350 million tons of aggregates in 1940
71% limestone/dolomite
350(.71) = 248.5 million tons
https://www.dnr.wa.gov/publications/ger_b52_limestone_res_western_wa_1.pdf

451,253,000 tons of limestone/dolomite 1960
https://books.google.com/books?id=jaTJzL5y0LQC&pg=PA779&lpg=PA779&dq=tons+of+crushed+limestone+US+1980&source=bl&ots=9rX_k8Sl2M&sig=ACfU3U0HgaoN9thHTkRPm6sqc17U3KCkBg&hl=en&sa=X&ved=2ahUKEwjXj_-to-fnAhWaGDQIHVI-CWYQ6AEwEnoECAoQAQ#v=onepage&q=tons%20of%20crushed%20limestone%20US%201980&f=false
646 million tons of crushed limestone/dolomite 1981
In 1981, consumption of crushed stone decreased 11% from last year to 873 million
(646/873)=0.74
0.89x=873
x=980.9 tons in 1980
980.0(0.74)=725.84 million tons of crushed limestone/dolomite in 1980

Liming:

Year	Crushed Limestone/dolomite (metric tons)	ΔC_{Lime}
1940	248,500,000	36,446,666.67
1960	451,253,000	66,183,773.33
1980	725,840,000	106,456,533.3
2000	1,120,000,000	164,266,666.7
2019	1,064,000,000	156,053,333.3

Equation 3-28: CO_2 Emissions from Urea Fertilization

$$C_{Urea} = M \times EF \times CO_2MW$$

Where:

C_{Urea} = Annual release of carbon from urea added to soil (metric tons CO_2-eq year^{-1})

M = Annual amount of urea fertilization (metric tons urea year^{-1})

EF = Emission factor or proportion of carbon in urea, 0.20 (metric ton C (metric ton urea)$^{-1}$)

CO_2MW = Ratio of molecular weight of CO_2 to carbon (44/12) (metric tons CO_2 (metric tons C)$^{-1}$)

https://www.usda.gov/oce/climate_change/Quantifying_GHG/Chapter3S.pdf
https://www.ers.usda.gov/data-products/fertilizer-use-and-price.aspx
Tons of urea fertilization in 1960 (year ending in June 30)= 142,198

Tons of urea fertilization in 1970 = 533,535
Tons of urea fertilization in 1980 = 2,144,648
Tons of urea fertilization in 1990 = 3,738,827
Tons of urea fertilization in 2000 = 4,697,642
Using a line of best fit to calculate 2019 because I could not find data for that year yet, I used points from 2010-2015 and got 8,223,600

https://naldc.nal.usda.gov/download/CAT78694121/PDF
1940 total amount of anhydrous ammonia was 475,000

https://www.earth-syst-sci-data.net/10/969/2018/essd-10-969-2018.pdf
< 0.3 Tg N yr−1 from 1850-1940 of fertilizer consumption
9.8 Tg N yr−1 in 1980

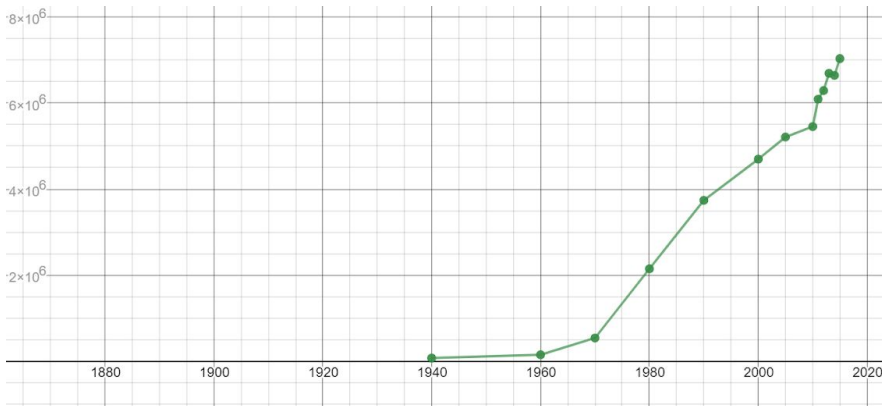
0.3/9.8=0.0306
0.0306(2,144,648)=65652.5
>65652.5 tons of urea in 1940

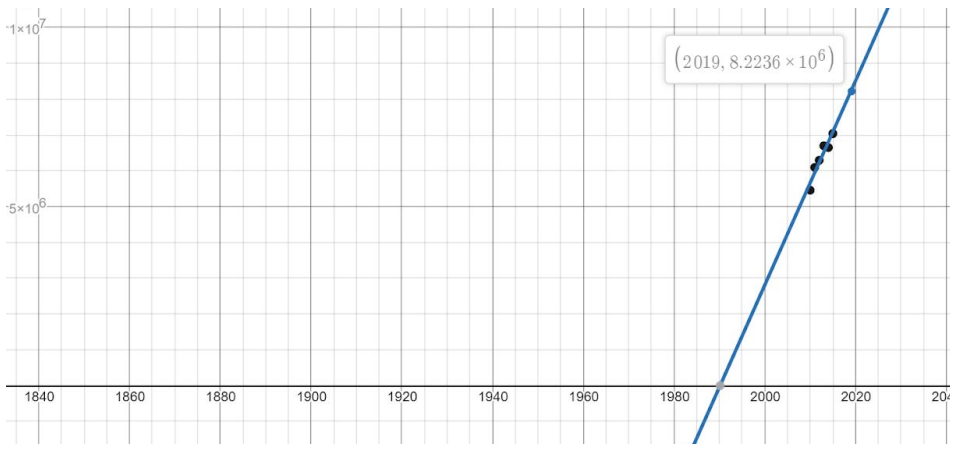
Urea as Fertilizer:

Year	Amount of urea fertilization (metric tons)	C_{Urea}
1940	65652.5	48,145.167
1960	142,198	104,278.53
1980	2,144,648	1,572,741.867
2000	4,697,642	3,444,937.467
2019	8,223,600	6,030,640

https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions

https://www.epa.gov/sites/production/files/2017-12/documents/ag_module_users_guide.pdf





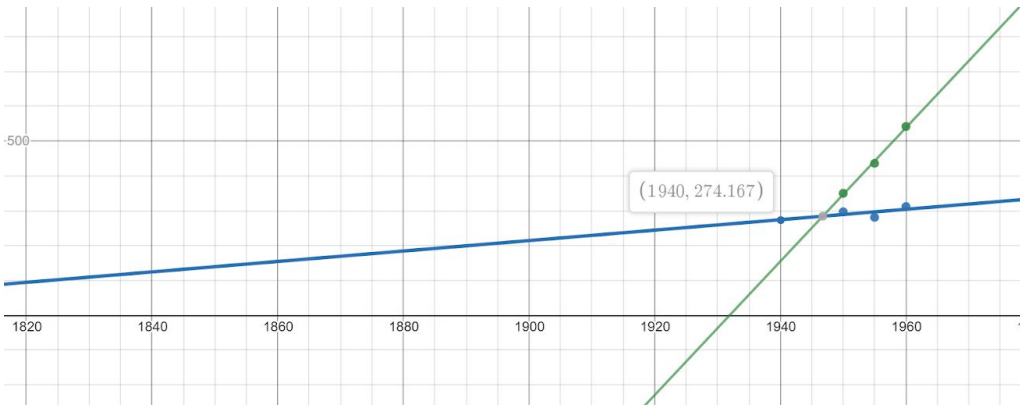
- **Indirect/Direct Residential and Commercial:**

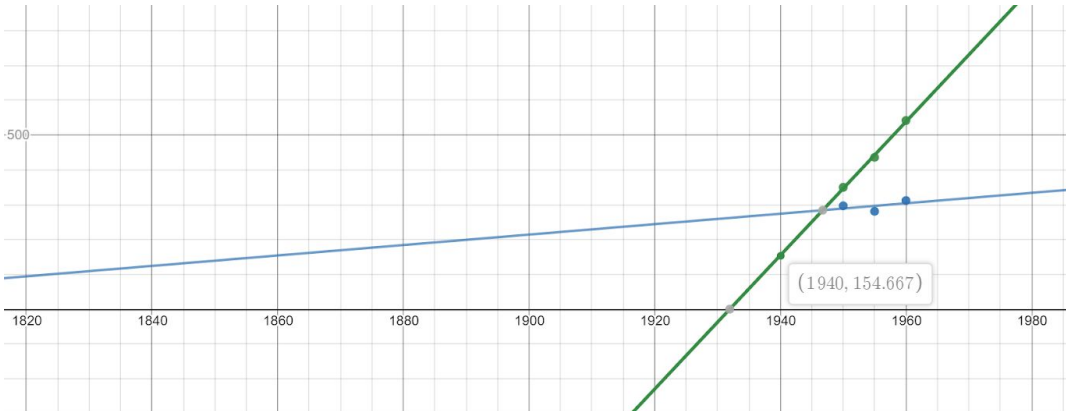
https://www.statista.com/statistics/204879/us-carbon-dioxide-emissions-by-sector-since-1950/
https://www.c2es.org/document/decarbonizing-u-s-buildings/
According to https://www.c2es.org/document/decarbonizing-u-s-buildings/ , residential and commercial sectors are just as large as transportation and electricity is due to indirect sources of carbon dioxide such as "electricity generated off-site".

Year	Carbon Breakdown (Million metric tons)	Total Carbon Emission (metric tons)
1940	Residential Sector: 154.667 Commercial Sector: 274.167	428,834,000
1960	Residential Sector: 542	854,000,000

	Commercial Sector: 312	
1980	Residential Sector: 911 Commercial Sector: 662	1,573,000,000
2000	Residential Sector: 1185 Commercial Sector: 1022	2,207,000,000
2019 (2018 could not find for 2019)	Residential Sector: 1019 Commercial Sector: 893	1,912,000,000

Blue = Commercial Sector
Green = Residential Sector





- **Electricity:**

https://www.eia.gov/tools/faqs/faq.php?id=74&t=11
In 2018, 4.17 trillion kilowatthours (kWh) which resulted in 1.87 billion metric tons of carbon dioxide (2.06 short tons) = 0.99 lbs of CO2 per kWh
1lbs = 0.0005 tons
1 metric ton = 2204.62 lbs

1 short ton = 2000 lbs

https://www.eia.gov/energyexplained/electricity/electricity-in-the-us.php

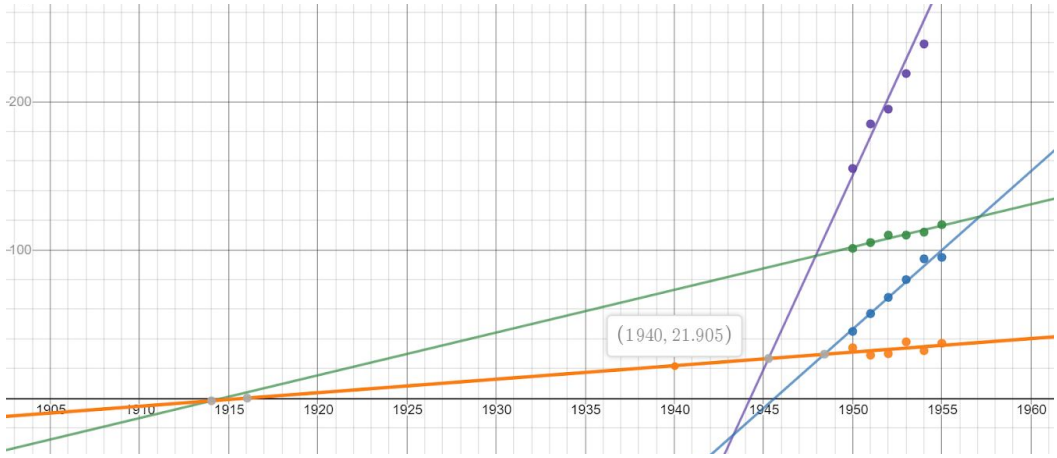
Year	kWh	CO2 (metric tons)
1940	Renewables: 73.095 billion kWh Petroleum and others: 21.905 billion kWh Total: 95 billion kWh	42,660,413.13
1960	Coal: 403 billion kWh Natural Gas: 158 billion kWh Nuclear: 1 billion kWh Renewables: 150 billion kWh Petroleum and other: 48 billion kWh Total: 760 billion kWh	341,283,305.1
1980	Coal: 1162 billion kWh Natural Gas: 346 billion kWh Nuclear: 251 billion kWh Renewables: 285 billion kWh Petroleum and other: 246 billion kWh Total: 2290 billion kWh	1,028,340,485
2000	Coal: 1,966 billion kWh Natural Gas: 601 billion kWh Nuclear: 754 billion kWh Renewables: 356 billion kWh Petroleum and other: 124 billion kWh Total: 3801 billion kWh	1,706,865,582
2019 (technically 2018 but similar)	Coal: 1,146 billion kWh Natural Gas: 1,468 billion kWh Nuclear: 807 billion kWh Renewables: 713 billion kWh Petroleum and other: 44 billion kWh Total: 4178 billion kWh	4.178e12(0.99)/2204.62= 1,876,160,064

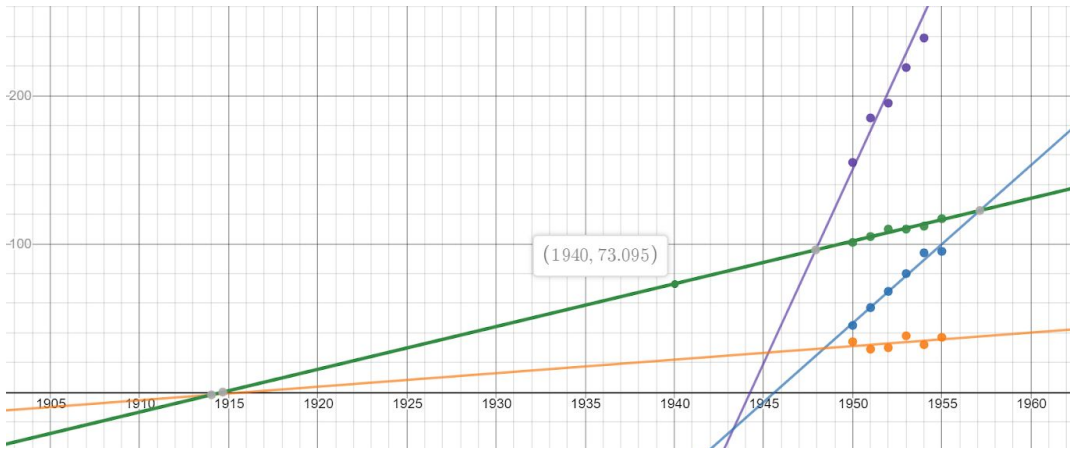
Purple = coal

Blue = natural gas

Green = renewables

Orange = petroleum and others





- **Transportation**

Cars: A typical passenger vehicle emits about 4.6 metric tons of carbon dioxide per year. This assumes the average gasoline vehicle on the road today has a fuel economy of about 22.0 miles per gallon and drives around 11,500 miles per year.
Every gallon of gasoline burned creates about 8,887 grams of CO2.

General Aviation (Local Planes): CO2 emissions from aviation fuel are 3.15 grams per gram of fuel, which gives CO2 emissions from an aircraft of 115 g per passenger km. At a cruising speed of 780 km per hour, this is equivalent to 90 kg CO2 per hour. So for the aircraft, the emissions are around 90 kg CO2 per hour.

Heavy duty trucks: the truck fleet in the U.S. consumed about 94,890,000 million gallons of fuel each day in a year, and emitted a total of 436.5 million metric tons of carbon dioxide. The average freight truck in the U.S. emits 161.8 grams of CO2 per ton-mile, with about only 6.0 mpg.

- 1940
    - Cars:
        - Average of 4,680,000 cars and about 10 to 15 miles per gallon.
        - About 43,056,000 metric tons of carbon dioxide per year
    - General Aviation:
        - Average of 587 aircrafts.
        - About 36,453 metric tons of carbon dioxide per year (Not fully accurate).
    - Heavy duty vehicles:
        - Average of 560,000 trucks and about 10 miles per gallon.
        - About 5,383,125 metric tons of carbon dioxide per year.
    - Overall:
        - 48,475,578 metric tons of carbon dioxide per year.
- 1960
    - Cars:
        - Average of 61,671,390 cars and about 12.4 miles per gallon.
        - About 503,318,118 metric tons of carbon dioxide per year.
    - General Aviation:
        - Average of 76,500 airplanes.
        - 4,750,298 metric tons of carbon dioxide per year.
    - Heavy duty vehicles:
        - Average of 686,500 heavy duty trucks and about 8 miles per gallon.
        - About 32,687,900 metric tons of carbon dioxide per year.
    - Overall:
        - 540,756,316 metric tons of carbon dioxide per year.
- 1980
    - Cars:
        - Average of 121,600,843 cars and about 13.5 to 27.5 miles per gallon.
        - About 600,292,942 metric tons of carbon dioxide per year.
    - General Aviation:
        - Average of 211,000 airplanes.
        - About 13,104,900 metric tons of carbon dioxide per year.
    - Heavy duty vehicles:
        - Average of 1,416,869 heavy duty trucks and about 5.4 miles per gallon.
        - About 260,055,000 metric tons of carbon dioxide per year.
    - Overall:
        - 873,452,842 metric tons of carbon dioxide per year.

- 2000
    - Cars:
        - Average of 225,821,241 cars and about 21 miles per gallon.
        - About 1,088,243,000 metric tons of carbon dioxide per year.
    - General Aviation:
        - 172,602,783 metric tons of carbon dioxide per year.
    - Heavy duty vehicles:
        - 348,300,000 metric tons of carbon dioxide per year.
    - Overall:
        - 1,609,145,783 metric tons of carbon dioxide per year.
- 2019
    - Cars:
        - Average of 272,480,899 cars and about 22 miles per gallon.
        - About 1,253,412,000 metric tons of carbon dioxide per year.
    - General Aviation:
        - 129,200,000 metric tons of carbon dioxide per year.
    - Heavy duty vehicles:
        - Average of 15,500,000 heavy duty trucks and about 6 miles per gallon.
        - 436,500,000 metric tons of carbon dioxide per year.
    - Overall:
        - 1,819,112,000 metric tons of carbon dioxide per year.

- **Stationary Sources**

https://www.eia.gov/environment/emissions/carbon/
There was a total of 5,269 million metric tons of carbon released in 2019, and a total of 5,867 million metric tons released in 2000. According to the EPA's statement of 14% of all carbon coming from industry, then roughly 737.66 million metric tons of carbon came from industry in 2019, and 821.38 million metric tons were released in 2000. This is roughly 737,660,000 metric tons and 821,380,000 metric tons from 2019 and 2000, respectively.

https://www.macrotrends.net/countries/USA/united-states/carbon-co2-emissions
1980 released 4,723,209 kilotons of carbon, and 1960 released 2,890,696 kilotons of carbon. Using the 14% calculation from earlier, that means that 661,249.26 kilotons of carbon were released by industry in 1980, and 404,697.44 kilotons were released by industry in 1960. This comes out to about 599,875,238 metric tons released by industry in 1980, and 367,135,341 metric tons in 1960.

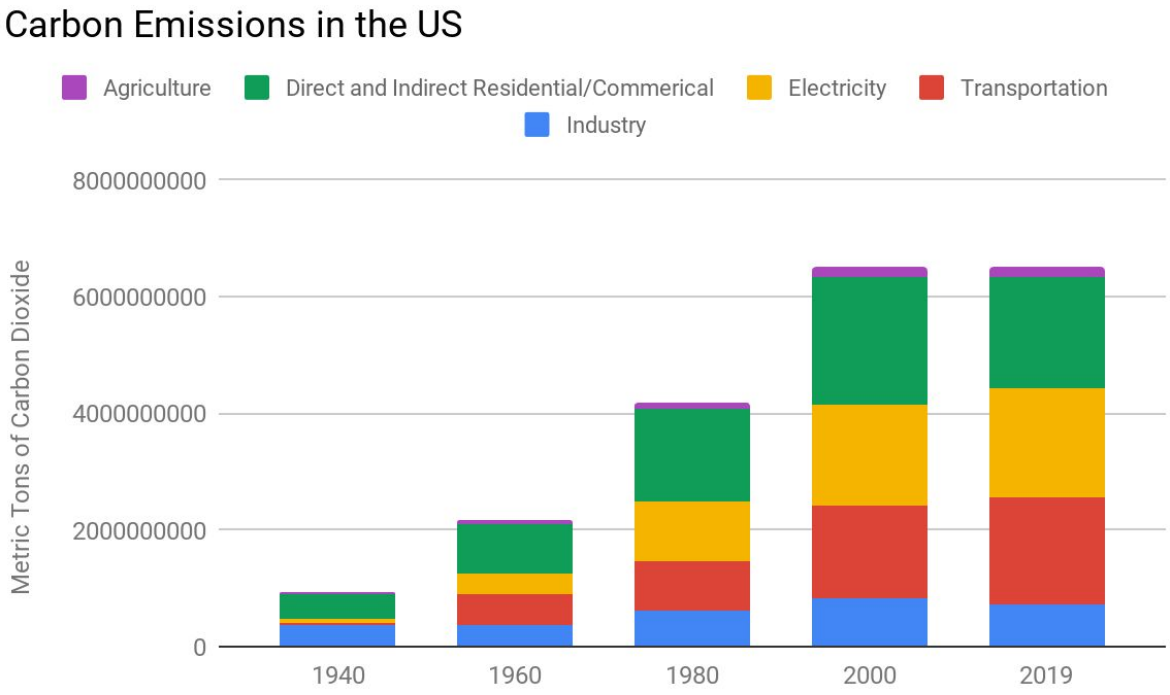
https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions
In 1940, 1.87 billion metric tons of carbon were released. Given that carbon makes up 82% of all greenhouse (according to the EPA), then there were 2,280,487,804 metric tons of greenhouse gases released overall. Based on the EPA's graphs of greenhouse gases by economic sector, on average, 19.75% of greenhouse gases come from industry. This means that 450,396,341.29 metric tons of greenhouse gases are released by industry. If 82% of those gases are carbon, then that means that 369,324,999.86 metric tons of carbon are released by industry.

Industry's Carbon Footprint

Year	Metric Tons of Carbon Released
1940	369,324,999.86
1960	367,135,341
1980	599,875,238
2000	821,380,000
2019	737,660,000

- **Data Tables and Charts**

	A	B	C	D	E	F
1		Industry	Transportation	Electricity	Direct and Indir	Agriculture
2	1940	369,325,000	48,475,578	42,660,413.13	428,834,000	36494811.84
3	1960	367,135,341	540,756,316	341,283,305.10	854,000,000	66288051.86
4	1980	599,875,238	873,452,842	1,028,340,485	1,573,000,000	108030275.2
5	2000	821,380,000	1,609,145,783	1,706,865,582	2,207,000,000	167711604.2
6	2019	737,660,000	1,819,112,000	1,876,160,064	1,912,000,000	162083973.3



We entered the numbers into a spreadsheet to create a stacked column chart. This was the best chart to visualize trends in the data. From this chart, we are able to see that the total amount of carbon dioxide emissions over the years have drastically increased but have stagnated in recent years. This could be due to the popularization of sustainable products and technology.

Discussion:

Our main finding throughout this project is the metric tons of carbon dioxide for every respected year charted. By looking at our findings, we can see an increasing trend of CO2 throughout the decades, which strongly correlates to the increase in global warming. The largest sectors contributing to carbon emissions are transportation, electricity, and indirect/direct

residential and commercial sources. Since it was unfeasible for us to personally record and track the amount of carbon emissions, our main source of obtaining statistics was through other professional studies and charts. We were limited in our ability to find accurate data from less recent years. Therefore, we had to use pre-existing data to estimate numbers to the best of our abilities. Additionally, a few of the categories overlapped in the amount of carbon emissions calculated. Electricity was one factor that may have been accounted for more than once since residential and commercial sources include electricity. This may result in the residential and commercial sector appearing to contribute more CO2 emissions than it actually does. Another obstacle that we faced was finding the carbon emission for trains in the transportation category. Since trains change in size during every transport, we have to take into account the weight that correlates to the amount of gas used in order to move this massive vehicle. To resolve this problem, we decided to change this type of transportation into heavy duty vehicles, which was easier to find exact numbers in order to find the metric tons of carbon dioxide per year. For future work, it would be best if we could have access to exact statistics rather than estimating or personally collecting data on carbon dioxide emissions.