Project 5

Kavya Nair

May 2022

1 Introduction

Climate change is currently a very prevalent issue. Climate change is the long-term shift in climate and weather patterns largely resulting from increased industrial activity, which emits carbon dioxide into the air. Climate change also has very harsh and obvious effects in many places. We will be looking at California due to its frequent droughts. Climate data has often been modeled and found to be cyclical, or repeating. We will be examining California's levels of precipitation for 2015, 2016, 2017, 2018, and 2019 by using the recorded total monthly precipitation. We will be asking the question of: *Is the precipitation rate of California cyclical? How does the precipitation from 2015 to 2019 compare with normal precipitation rates for each month?* In order to answer these questions, we will choose specific cities, San Francisco from the north and Los Angeles from the south, and model their precipitation rates using a sinusoidal function. We will also examine the graphs of the normal precipitation rates to compare them with the data from 2015-2019.

2 Data Collection

To begin, we can use www.usclimatedata.com to get recorded total monthly precipitation for San Francisco and Los Angeles. After collecting the data for each month in 2015, 2016, 2017, 2018, and 2019, we have the graphs below.

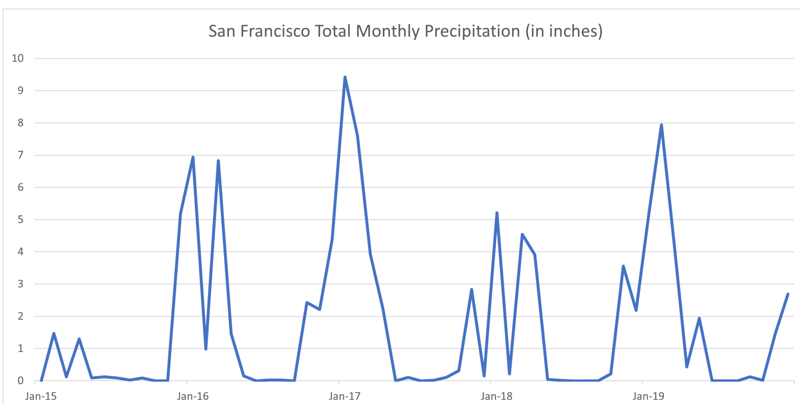

Figure 1

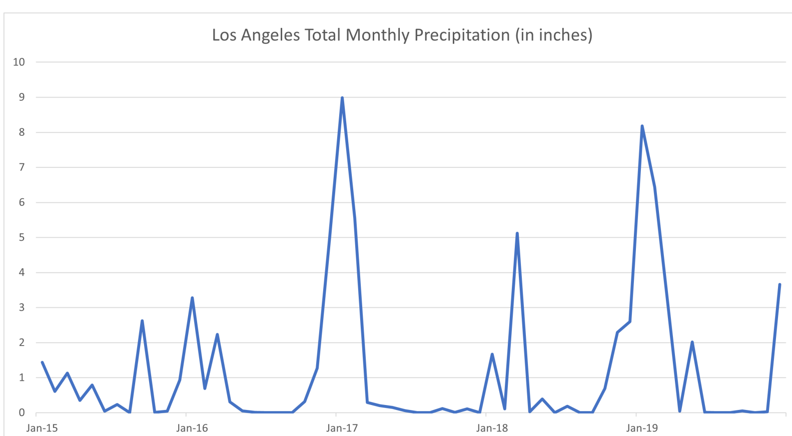

Figure 2

We can observe from these graphs that our data looks to be repeating itself, so we know that we can use a cyclical model.

We can now graph the normal precipitation rates for each month. The normal precipitation rates refers to the average historical total rainfall each month. Normal precipitation is the same value for all Januaries, Februaries, Marches, etc. but our data for precipitation rates looks at a span of five years. In order to easily compare the precipitation rates to the normal precipitation rates, we can graph the normal values over and over five times so that it appears as five total years.

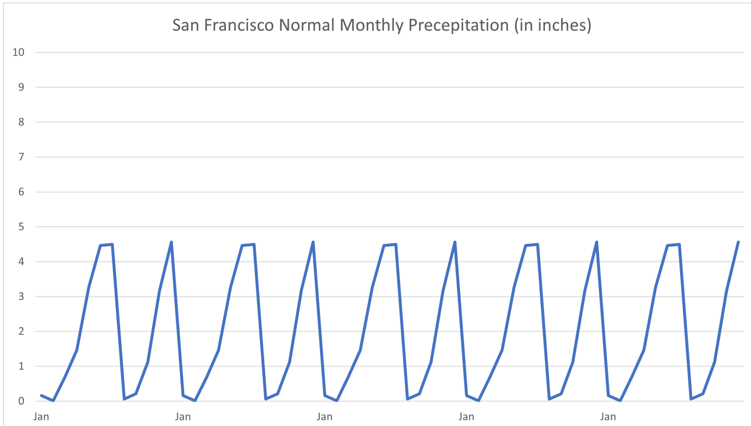

Figure 3

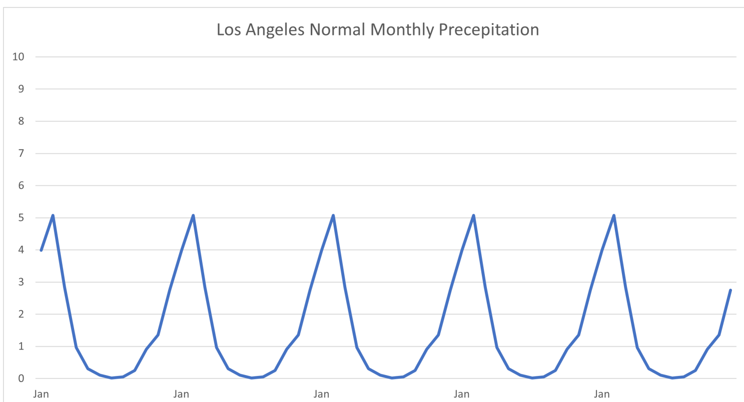

Figure 4

Now that we have constructed graphs for both the recorded precipitation rates as well as the normal precipitation rates, we can move on to constructing sinusoidal models for each of these graphs.

3 Constructing Sinusoidal Models

For both our San Francisco Monthly Precipitation Rates graph and our Los Angeles Monthly Precipitation Rates graph, we can see that our graphs roughly keep peaking at January, our period is a year, or 12 months.

3.1 San Francisco Recorded Monthly Precipitation Rates Model

Let's start by using the San Francisco Monthly Precipitation Rates graph and constructing a sinusoidal to model the data. The method we can use to construct this graph is beginning with the parent function $\sin x$ and then manipulating the different aspects to put it in the form $A \sin\left(\frac{2\pi}{B}(x - C)\right) + D$. We can approximate the amplitude as an average of the peaks with some tweaking. We can use the highest value produced by the amplitude and zero to determine the mean value. We can use phase shift to shift the graph continuously until it is placed in what looks like the most accurate position to represent our data. After performing these manipulations, we have the graph presented below.

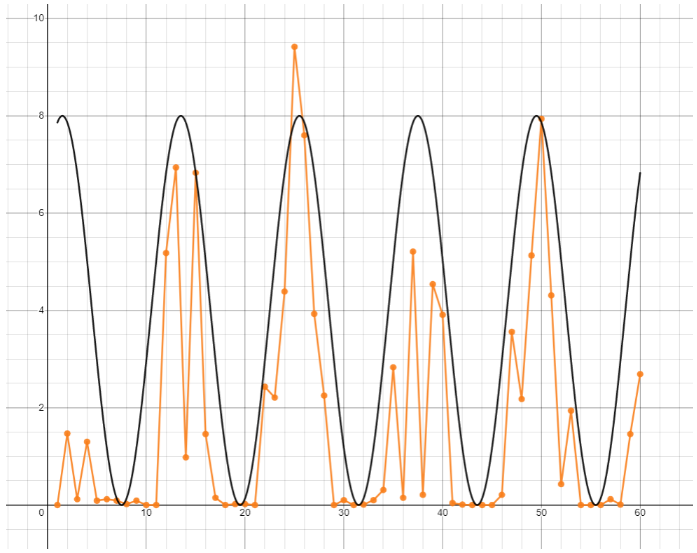

Figure 5

The sinusoidal in this graph is $4\left(\sin\frac{2\pi}{12}\left(x+1.5\right)\right)+4$ restricted to $\{1 \le x \le 60\}$.

3.2 Los Angeles Recorded Monthly Precipitation Rates Model

We can use this same process to model our Los Angeles Monthly Precipitation Rates. This gives us the sinusoidal model shown below.

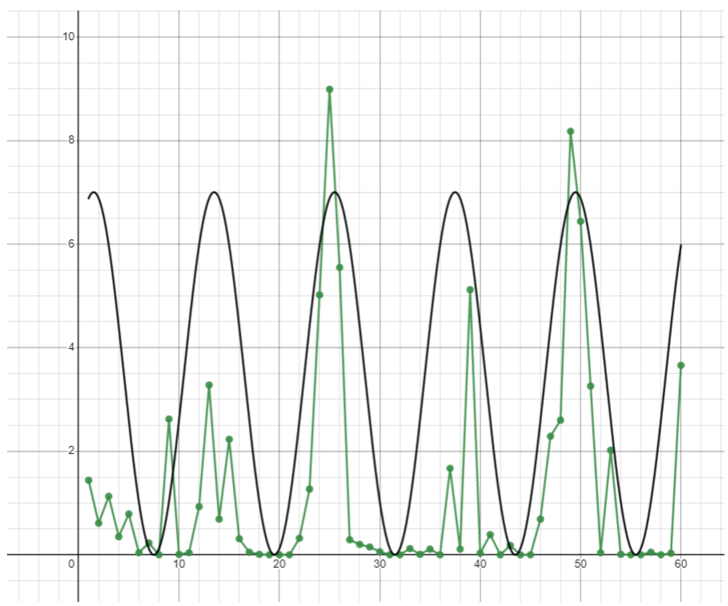

Figure 6

The sinusoidal in this graph is $3.502\left(\sin\left(\left(\frac{2\pi}{12}\right)\left(x+1.5\right)\right)\right)+3.502$ restricted to $\{1 \le x \le 60\}$.

3.3 San Francisco Normal Monthly Precipitation Rates Model

We must now model our San Francisco Normal Monthly Precipitation Rates. Because for this set, we have repeatedly graphed the same data, we can just take the peak as the maximum and the lowest as the minimum and go from there. But there is one change that needs to be made here. Observing the data, it looks like there are two peaks; one peak at roughly just before January and one seemingly in the middle of the year. Because it looks like there are around two peaks in a year, we can change the period for this function to be 6 instead of 12. Below is the graph obtained after making our manipulations to the function.

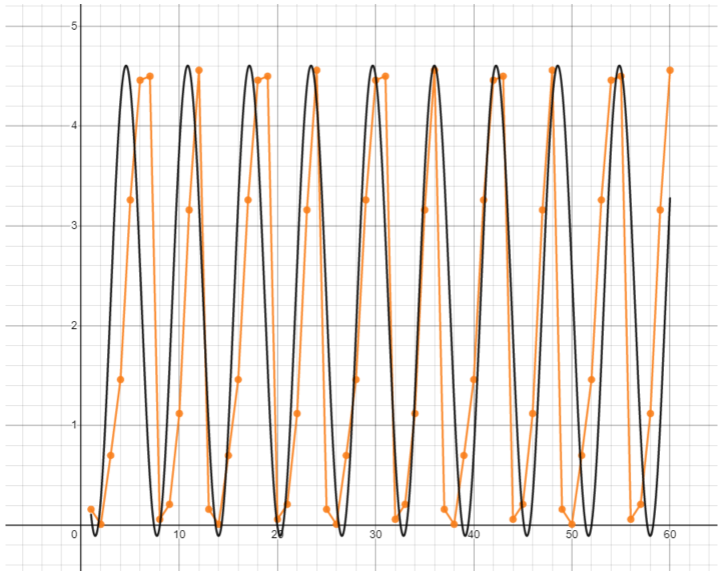

Figure 7

The sinusoidal in this graph is $2.25\left(\frac{2\pi}{6}\sin(x-3)\right)+2.25\,\{1\leq x\leq 60\}$ restricted to $\{1\leq x\leq 60\}$.

3.4 Los Angeles Normal Monthly Precipitation Rates Model

To construct the sinusoidal of our Los Angeles Normal Monthly Precipitation, we can use the method that was outlined for San Francisco's Normal Monthly Precipitation but change the period to 12 months again. Below is the graph we have made.

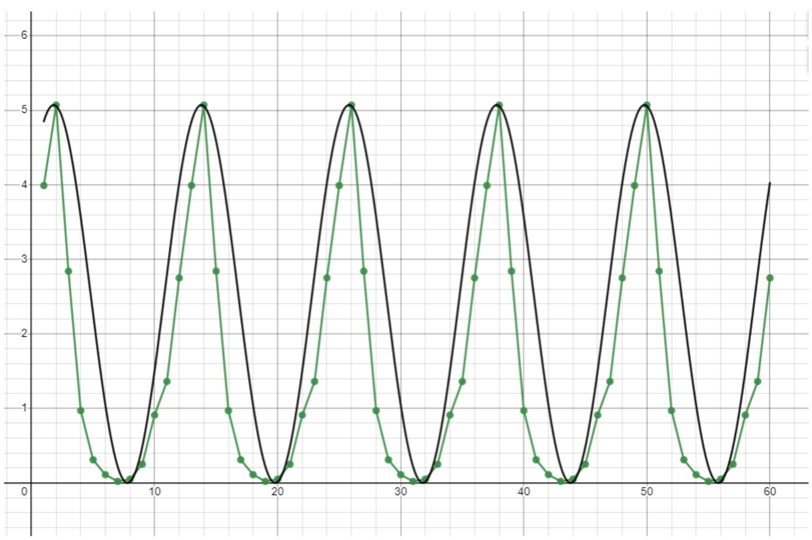

Figure 8

The sinusoidal in this graph is $\frac{5.07}{2}\sin\left(\frac{2\pi}{12}(x+1.2)\right)+\frac{5.07}{2}$ restricted to $\{1\leq x\leq 60\}$.

4 Calculating RMSE of the Models

Now that we have found the models, we want to test how good they are. To do this, we can find the RMSE of the data and the models we have to represent them. The computation shown below is of the RMSE of San Francisco's Total Monthly Precipitation Rates and the sinusoidal used to model it. We can use the same outlined process to calculate all RMSEs.

```
import math

# A, B, C, D in Asin(2pi/B(x-C))+D
a = 4
b = 12
```

```
c = -1.5
d = 4

# Actual Data
data1 = [0, 1.47, ... ,2.69]
data2 = []

# Calculates Actual Sinusoidal
for i in range(1,61):
    data2.append(a*math.sin(2 * math.pi / b * (i - c)) + d)

# RMSE calculation
def RMSE(da, db):
    summation = 0
    for i in range(len(da)):
        summation += (da[i] - db[i])**2 / len(da)
    return math.sqrt(summation)

# Return RMSE Value
print(RMSE(data1,data2))
```

Computation of RMSE (Code by Bohan Yao)

The sinusoidal function for San Francisco's recorded total monthly precipitation has an RMSE of 3.09124.
The sinusoidal function representing Los Angeles' recorded total monthly precipitation has an RMSE of 3.10974.
The sinusoidal function representing San Francisco's normal monthly precipitation rate has an RMSE of 2.1117.
The sinusoidal model for Los Angeles' normal monthly precipitation rate has an RMSE of 1.2712.

We have the RMSE's for all of the models we created for the specific data sets. But because part of our inquiry is the relationship to climate change, it would be helpful to see how the data from 2015-2019 aligns or strays from what is "normal." We can do this by computing the RMSE of San Francisco's normal precipitation sinusoidal and the recorded data from 2015-2019 as well as computing the RMSE of Los Angeles' normal precipitation sinusoidal and the recorded data from 2015-2019.

The sinusoidal for the normal precipitation rates of San Francisco, when used to model the recorded data from 2015-2019, has an RMSE of 3.1644. The data along with this sinusoidal is shown below.

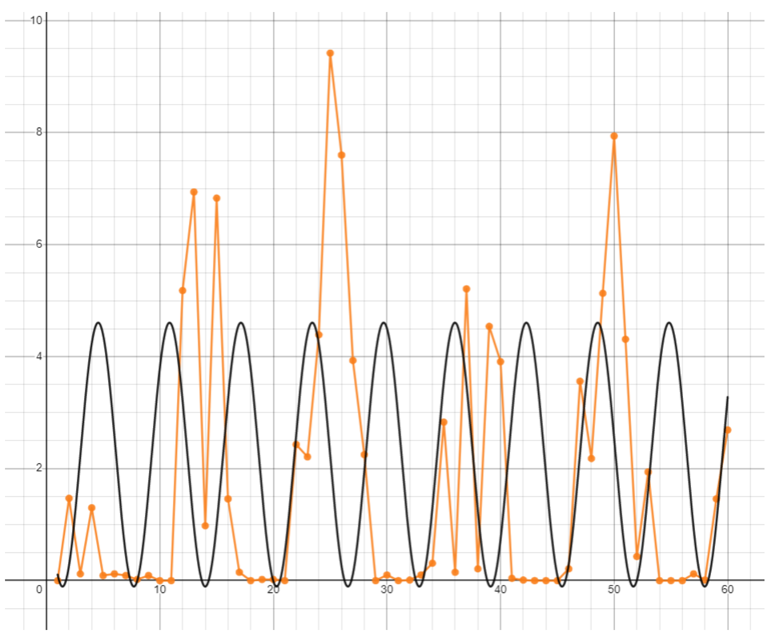

Figure 10

The sinusoidal for the normal precipitation rates of Los Angeles, when used to model the recorded data from 2015-2019, has an RMSE of 2.243. The data along with this sinusoidal is shown below.

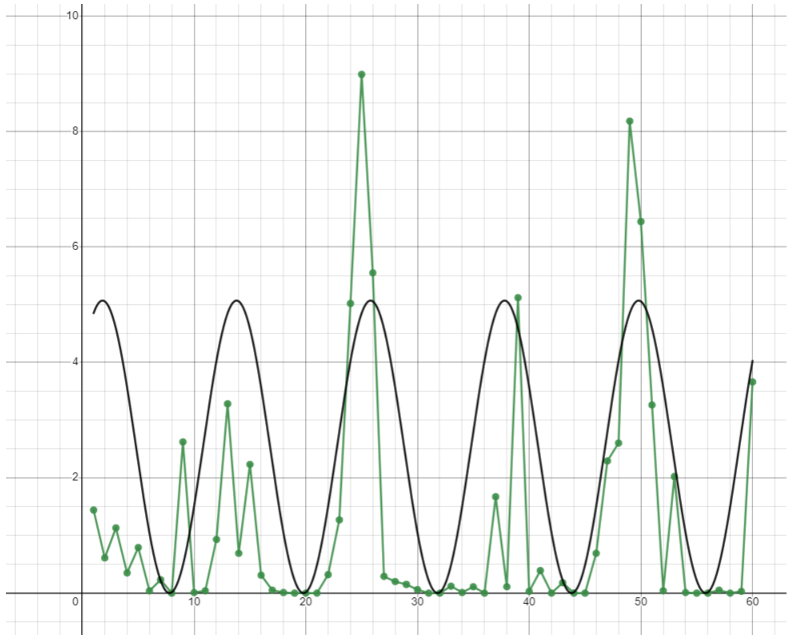

Figure 11

5 Data Analysis

We have constructed sinusoidal to model our four data sets, and we have also computed the RMSEs.

5.1 Using RMSE to Understand the Effectiveness of the Models

We were able to find that the model constructed for San Francisco's Total Monthly Precipitation Rates had an RMSE of 3.09124, which is not very good, but sufficiently models the data for our purposes. The model constructed for Los Angeles' Total Monthly Precipitation Rates was slightly worse than San Francisco's model, but still good enough to see the shapes of the data. The models we constructed to represent the normal rates for San Francisco and Los Angeles were relatively good models, with RMSEs of 2.1117 and 1.2712 respectively.

5.2 Understanding Why RMSE Does Not Accurately Convey the Effectiveness of Our Models

In order to determine the effectiveness of our models, we calculated the RMSE of the models and their original data sets. For both San Francisco and Los Angeles' Total Monthly Precipitation models, we received RMSEs of roughly 3. Considering the fact that the total rain recorded does not ever even reach 10, an RMSE of 3 is not good, and indicates that our models are not very accurate. However, through both the method we used to construct the model as well as the visual aspect, we can tell that the models we created are good at representing the data. We know this because the amplitude we chose is the average of the peaks, which is a good way to estimate the amplitude. The period is 12 months (as the expected cycle for something related to weather is yearly), except for one function, which looks like it is repeating twice a year. Our phase shift was chosen to match the graph as well as possible, and our mean value was the average between the highest and lowest value. We can still trust that our model is accurate because visually, it accurately models our rough maximums and minimums, as well as the shape of the graph. There are some dips in the middle of the general increases and decreases, which is likely a reason that the RMSE is high. These dips are not particularly significant here because the point of focus is mainly the the peaks. Because our graph so tightly covers the shape of the precipitation rates, we can use it as a general model to see the trend. We can consider what generally seem like random dips to be an aberration. Because these shapes align quite closely, this supports the idea that our model is correctly displaying the trends and general figure of our data.

5.3 Using the Normal Rates Models to Represent the Recorded Rates

We used the normal models to compare the recorded data to what is actually expected to occur.

When comparing San Francisco's recorded rates to the sinusoidal created for the normal rates, we got an RMSE of 3.1644. This RMSE score is almost the same as the RMSE score we got for the model constructed specifically to match the data. But, in the previous section, we established that RMSE alone is not enough to establish the effectiveness of our model in representing given data. While the RMSE scores are roughly the same, we can see visually that the peaks and valleys of this graph do not match with our data. This shows that the normal trends of precipitation rates are not present from 2015-2019, indicating that San Francisco's normal weather patterns strayed during this time.

When comparing our recorded data of Los Angeles' recorded rates to expected rates our RMSE was 2.243. The RMSE technically says that this model is better than the one we constructed specifically to match our data. Again referring to our analysis in section 5.2, we cannot trust this RMSE alone. But, looking at graph and how the peaks and valleys align, we can conclude that this actually is a better representation of the data. This tells us that Los Angeles' weather patterns did not stray from what was expected from 2015-2019.

5.4 Implications of Resulting Models and their Relationship to Climate Change

Looking at San Francisco's recorded rates and the comparison to the normal rates of precipitation, we can see that the models do not align. This indicates that the data from 2015-2019 on precipitation rates stray from what is considered normal for San Francisco. In the introduction, we defined climate change as long term shifts in climate and weather patterns. Because these five years show this difference in precipitation, this supports the fact that climate change is in fact occurring and its effects are clearly visible.

As for Los Angeles' comparison of recorded rates to normal rates of precipitation, these results match quite closely. While this may be the case, it does not mean that climate change does not exist because we already know that it is present. What it does tell us is that sometimes the effects are not quick to show up, or they show up in different ways that are not clearly visible. This refers to things like drought or forest fires, results of a water-deprived or extremely dry environment. These are different examples of climate change effects that are not clearly seen in our data.

5.5 Contextualizing Climate Change

Like we stated previously, industrial activity is a large contributor to climate change. In English, we discussed how high competition is what incites these rates of industrial activity. This then results in an excess of carbon emissions, which cause climate change. In our project, we examine an effect of this activity on the planet in a way that is clear to understand. The shift in rain rates in these cities in California show an obvious impact of climate change because they stray from what is expected. The discussion from English brings up points about what causes these changes in climate and weather patterns and further strengthens our understanding of why there is increased industrialized activity, and this connection can be considered more in the Further Inquiry section.

6 Conclusion

In conclusion, we can clearly see indications of climate change in San Francisco's data and how its weather patterns stray from what they are expected to be. We also explained why although Los Angeles did not stray significantly in precipitation rates, there are definitely still aspects of climate change present that are not seen within precipitation rates. We were also able to understand why RMSE does not always work in calculating how good a model is. When using a sinusoidal, we should also pay attention to the alignment and shape of the model with respect to the data in order to better understand if it is a proper fit.

7 Further Inquiry

Referencing our definition in the introduction, we mentioned that climate change is largely due to industrialization. As a path of further inquiry, we can look at the correlation between weather pattern shifts and the amount of industrialized activity across a long period of time, because this would let us know how related the two things are to each other. This is a new angle to look at the issue of climate change because so far, we have only mentioned the effects of climate change and not examine the potential causes.