

Cover Page

Team 7 - KWM² - due: 6/11/2021



MAE 10 FINAL PROJECT

3. Final Project Report

3.1 Abstract

This project aims to serve as a counter to the emerging demographic of Climate Change Deniers. Accompanying them are various misinformative talking points that counter scientific claims through logical fallacies. For example, the first talking point we will address is that of the climate not changing rapidly, which will be addressed by showing environmental data plots over time. The second point we will rebuke is that there are too many outliers in these kinds of statistics, which we will filter through the datasets for and remove using Matlab.

3.2 Nomenclature

Define all variables and symbols that will appear in this document.

Variables and symbols
IQR.m
Vec: stores data from text file as a 1-dimensional vector firstquartile: finds and stores 25th percentile of vector thirdquartile : finds and stores 75th percentile of vector tolerance: calculates the maximum distance a data point can be before it is an outlier middle: stores the midpoint of the vector outliers: vector to be removed from vec to produce corrected results vector upper: calculates upper bounds based on thirdquartile lower: calculates lower bounds based on firstquartile corrected: vector without outliers that is then returned
emission_subplot_figure.m
years: time vector, including years from 1993 to 2018 CAemissions: California carbon emissions vector FLemissions: Florida carbon emissions vector NYemissions: New York carbon emissions vector figure(2): subplot of the three emission graphs
RealMedianIncomeData.m
yearsCA: time vector, including years from 1984 to 2019 yearsNY: time vector, including years from 1984 to 2019 yearsFI: time vector, including years from 1984 to 2019 RealMedianIncomeCA: vector of Median Income in California RealMedianIncomeNY: vector of Median Income in New York RealMedianIncomeFI: vector of Median Income in Florida figure(1): subplot of the three median income vs time graphs
compare_figure.m

years: time vector for emissions Yearss: time vector for income CAemissions: California carbon emissions vector FLemissions: Florida carbon emissions vector NYemissions: New York carbon emissions vector CAmedianincome: median CA income vector FLmedianincome: median FL income vector NYmedianincome: median NY income vector figure(1): subplot comparing emissions and income over time
income_post_IQR_figure.m
years: time vector 1 years2: time vector 2 CAmedianincome: median CA income vector FLmedianincome: median FL income vector NYmedianincome: median NY income vector CAmedianincome2: median CA income vector post IQR test FLmedianincome2: median FL income vector post IQR test NYmedianincome2: median NY income vector post IQR test figure(1): subplot comparing income post IQR test and time
emissions_subplot_figure_post_IQR.m
years: time vector for emissions CAemissions: California carbon emissions vector FLemissions: Florida carbon emissions vector NYemissions: New York carbon emissions vector CAemissions2: California carbon emissions vector post IQR test FLemissions2: Florida carbon emissions vector post IQR test NYemissions2: New York carbon emissions vector post IQR test figure(1): subplot comparing emissions post IQR test and time
statistical_figure.m
Calculate-forms a standard deviation of our data set Means-forms a mean of our data set Medians-forms a median of our dataset Modes-forms a mode from our data set Deviation-forms an absolute standard deviation of our dataset Variance-forms the variance of our data set

3.2 Introduction

Summarize the purpose of this document so that someone unfamiliar with the project can understand what they are about to read. This is where you can talk about what you intended and what you accomplished (related to this project/team).

The purpose of this document is to learn more about climate change and observe factors that can influence the climate change of certain states in the United States. We aimed to see whether or not there is statistically clear correlation between how negatively climate change has impacted the following states: California, Florida, and New York. Trends in Median Salary of those states across the past few decades were observed to see if there is any indication that it can relate to the carbon emissions of that area. We hope to work together to interpret the data we researched and to see if we can come to a conclusion on how carbon emissions can be correlated with how affluent the area is.

3.3 Background of the project

Why should anyone care about the problem you are trying to solve? Why does it need a “solution”? (This can be the same as in the Proposal Document, but if there are any updates, include those here).

In our project we will be talking about Climate Change, which is a long term change in weather patterns that affects earth’s local and global climates. People should care about climate change because it will have multiple different long term effects on the way we live our day to day lives.

With increasing temperatures there will come more heat waves and droughts. There will be rising sea levels as we will see stronger and more intense natural disasters. Overall, climate change will bring ugly things onto our world. This issue isn’t one that is far away from us, study shows that “2020 was the second warmest year in the 141-year record for the combined land and ocean surface”. (*climate change: global temperature*).

In our project we decided to study the relationship between socioeconomic status and climate change (by using median income datasets and carbon emission datasets). For, when it comes to climate change, the most socioeconomically disadvantaged group will feel the effects of the heating globe first. Additionally, this same group of people will have the least resources available to them in order to handle problems associated with a change in climate. By examining our datasets we hope to prove this relationship between socioeconomics and carbon emissions.

Understanding climate change is the first step towards tackling climate change. By looking at and examining the data that's out there and finding trends that exist it breaks down this

3.4 Brief Overview of the Solution Using MATLAB

What did you create using MATLAB to address the problem? How did it address the problem you have identified? What adjustments did you make to your project as you were working on the solution? Who will find your solution useful (i.e. who is the intended user?)

Our solution is a simple set of figures, intended to prove the relationships discussed . Our solution refutes arguments that climate change is fraudulent by showing clear trends in data such as an increasing mean, median, and standard deviation in carbon emissions over time through code and

graphs, while also reducing random errors by automatically removing outliers using a basic Matlab program. The first function essentially returns significant statistical figures from a dataset while the second removes any outliers from a dataset that are more than 1.5 times the interquartile range away from the 25th or 75th percentiles. From there, those functions are applied to various datasets about environmental statistics and modeled in plots. The intended user is anyone who might need access to such figures to clearly show that meteorological data converges on the occurrence of severe Climate Change over the past 2-3 decades, as well as anyone who might need a program that removes IQR outliers or provides a 5-point summary.

3.5 Data Collection/Acquisition

What data did your team use? Did you collect or create the data set or did you use an existing data set (if so, make sure to include it in Appendix B)?

In our project we used an existing data set. We relied on data sets collected from the US Energy Information Administration and the Economic Research Federal Reserve Bank of St. Louis. Focusing on 1990 to 2018 what was the restriction for our data set for carbon emissions. And came along with other challenges such as limiting some of the findings we can come up with, since the data on CO2 emissions was primarily sourced from energy-related sources this data represents about 80% of all carbon emissions making it less representative of the dangers of increased CO2 emissions rates.

3.6 Data Analysis

Describe how you analyzed your data. Describe how you determined relevant figures to create and why they are important/useful.

We analyzed the data by looking at trends and highlighting relevant relationships between the sets with respect to time or location. From there, specifically, we identified the relationships between median income and time, emissions and time, emissions and location, and median income and location. After plotting these relationships and identifying those trends, we exported the plots from Matlab. Then, we filtered the data to exclude any outliers for carbon emissions and median income and proceeded to plot the filtered data on a separate graph to show that there were few, if any, outliers.

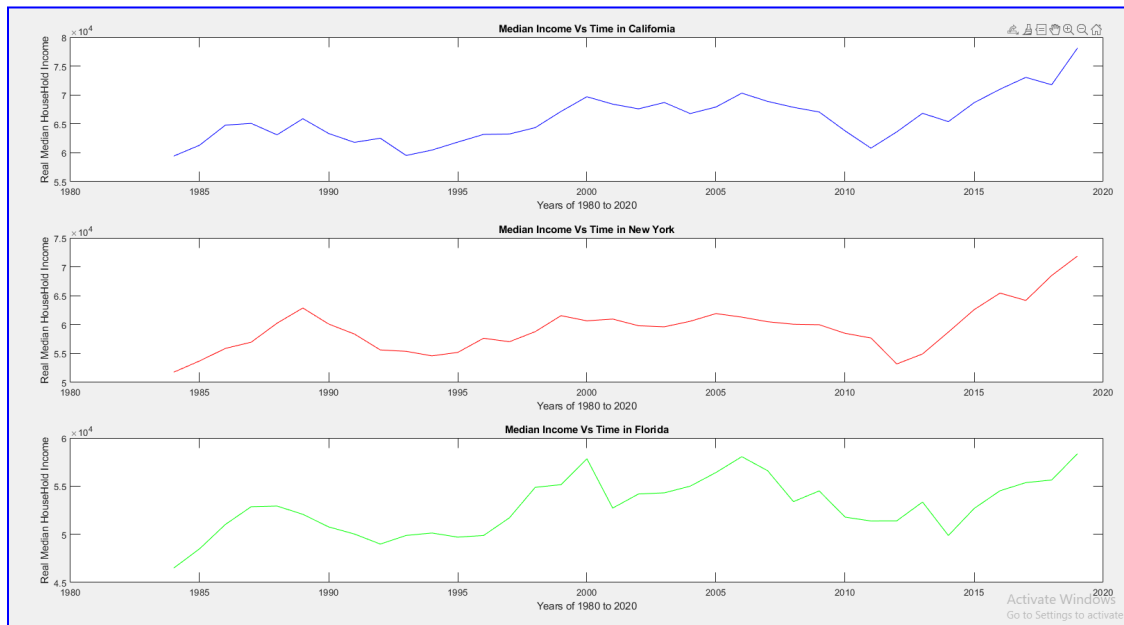
3.7 Solution and Demonstration

In this section, describe your solution including the five figures you created. Make sure each figure has an appropriate figure caption and is described in the narrative.

Legend:

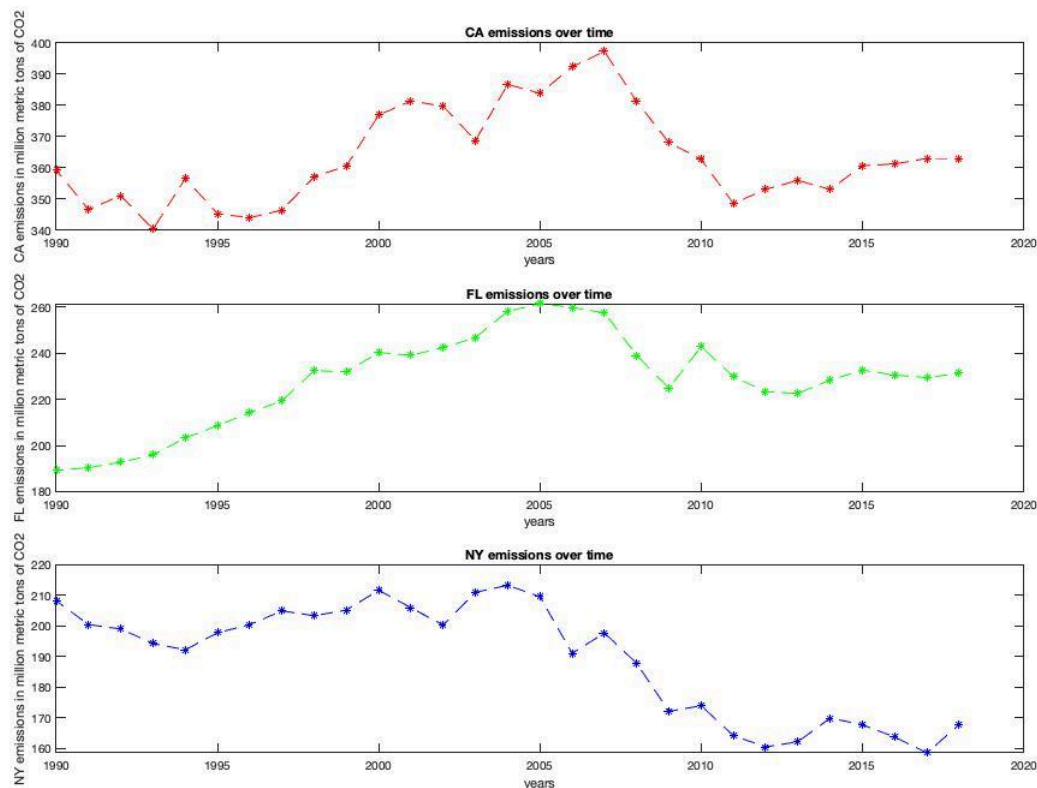
Color	State
	California
	Florida
	New York

1- Median income graphed against time for the three states we picked, Florida, California, and New York:



When looking at this graph, we can compare the median incomes of the three states CA, NY, and FL. By looking at the graph closer we can see that in 2019 the CA median income was around \$80,000, the FL median income was around \$60,000, and the NY median income was around \$70,000. As we move forward and examine the data we can keep in mind that in terms of median income: **CA > NY > FL**, keeping this in mind will help us prove the relationship between socio-economics and climate change.

2- Emissions in million metric tons of CO₂ graphed against time for the three states we picked, Florida, California, and New York:

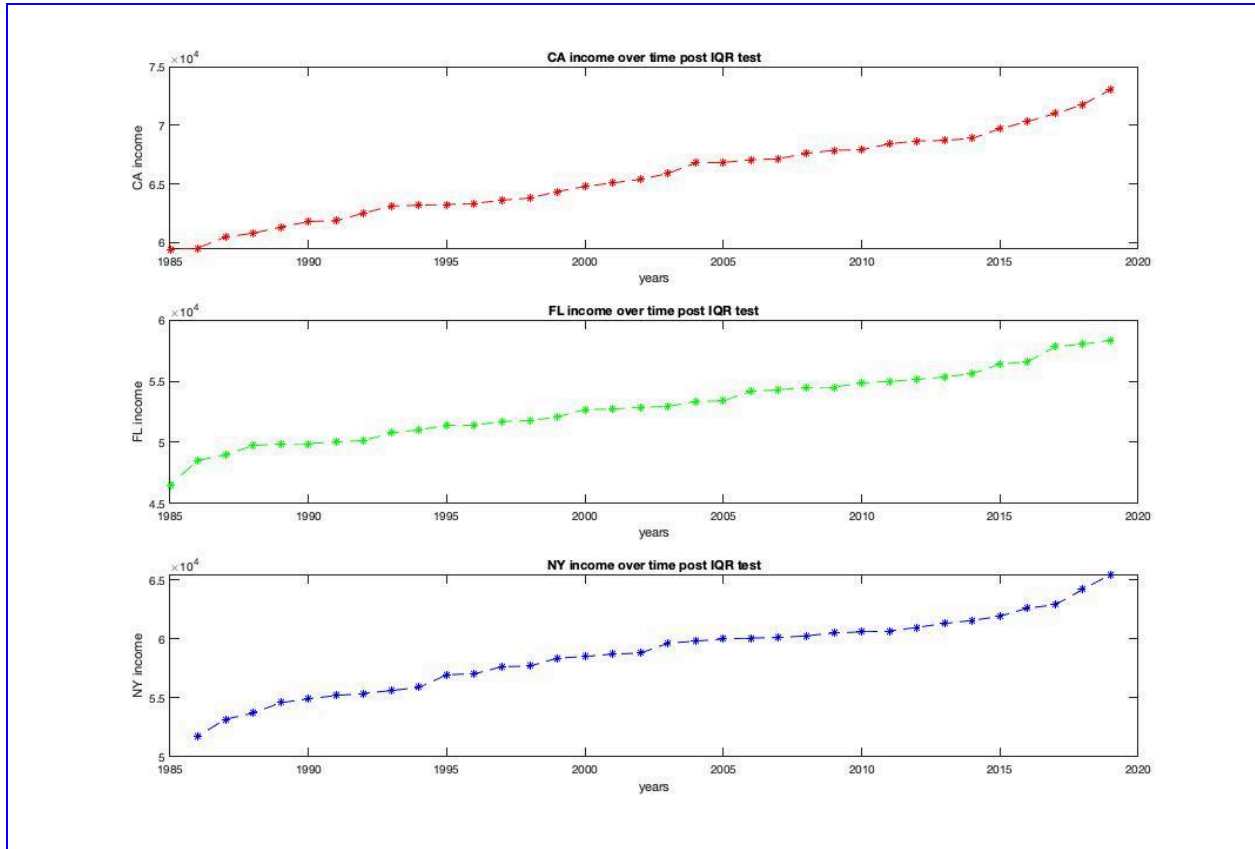


When looking at this graph, we can compare the emissions of the three states CA, NY, and FL. By looking at the graph closer we can see that in 2018 emissions in CA were around 400 million metric tons of CO₂, the emissions in FL were around 200 million metric tons of CO₂, and the emissions in NY were around 200 as well. As we move forward and examine the data we can keep in mind that in terms of emissions: **CA > NY ~ FL**, keeping this in mind will help us prove the relationship between socio-economics and climate change.

What is the IQR test? Why is it important?

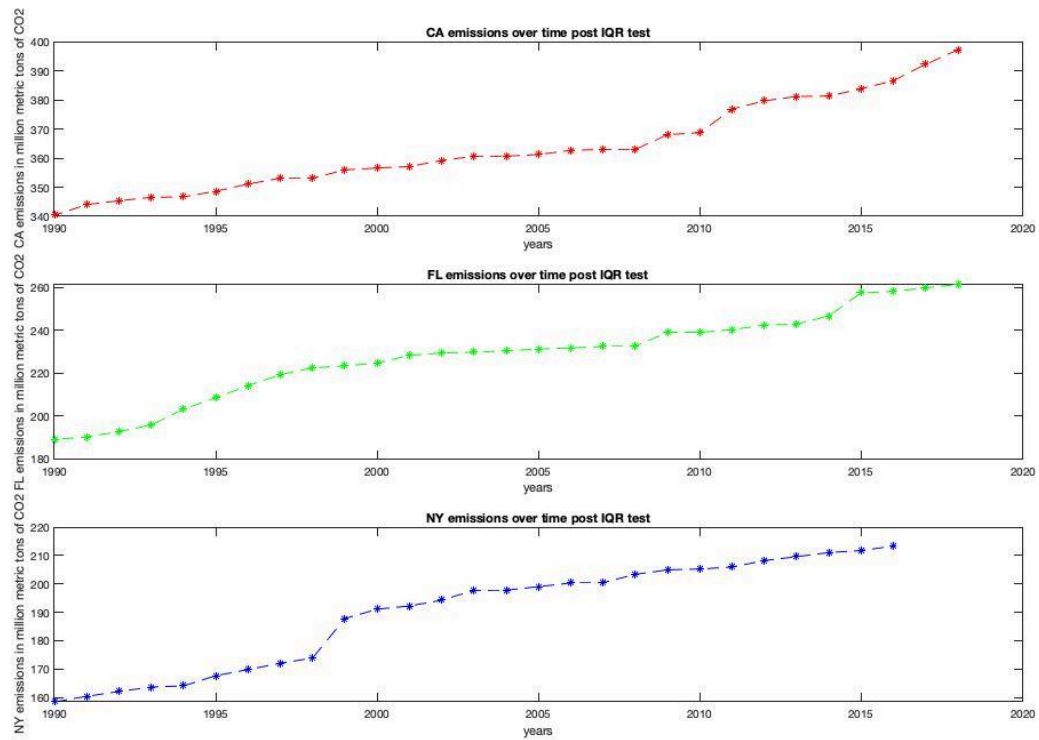
The InterQuartile Range, or IQR, is the difference between the third and first quartiles (or 75th and 25th percentiles) in a dataset of discrete values such as these. Any values in the set beyond 1.5 times the IQR are considered outliers and should be removed. This is significant as it confirms something generally about datasets over time, as well as datasets by year. It either confirms that within a year, there are few outliers and the average rate of carbon emissions or other environmental factors is actually increasing by year at an alarming rate, or that they have comparatively raised so much over time, the upper or lower ends of the spectrum might be considered an outlier, which is bad as there is almost no gradual adjustment, just a rapid increase.

3- Median income **post IQR test** graphed against time for the three states we picked, Florida, California, and New York:



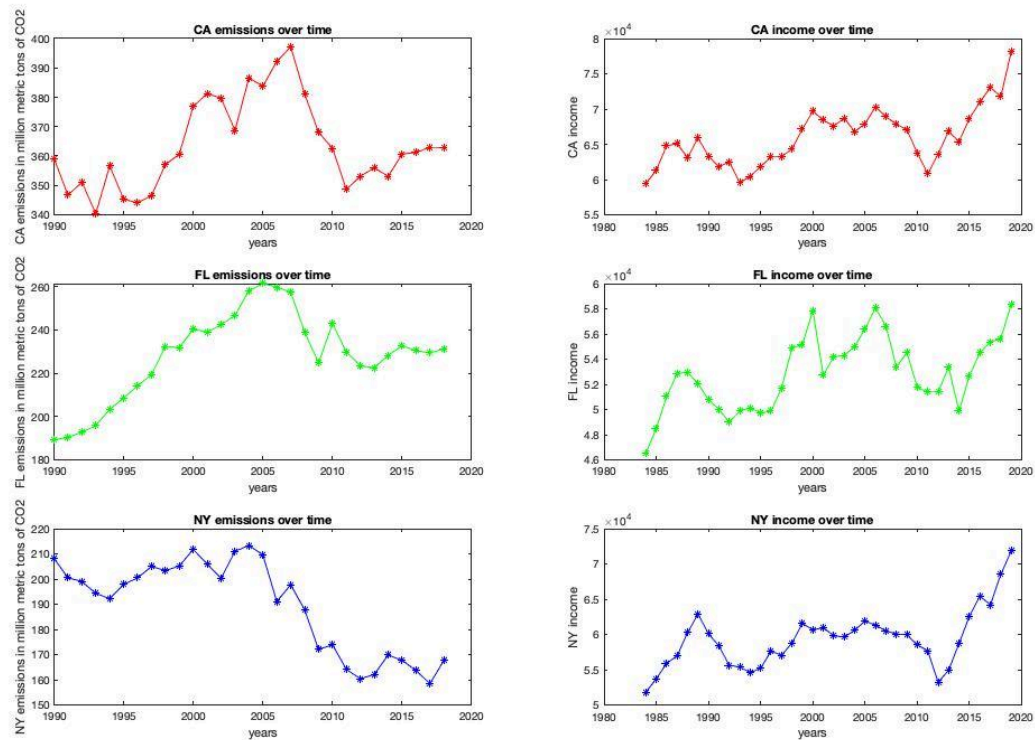
Median income graphed after running the IQR test and removing outliers. We see that the relationship remains in terms of median income: **CA > NY > FL**, keeping this in mind will help us prove the relationship between socio-economics and climate change.

4 - Emissions in million metric tons of CO₂ **post IQR test** graphed against time for the three states we picked, Florida, California, and New York:



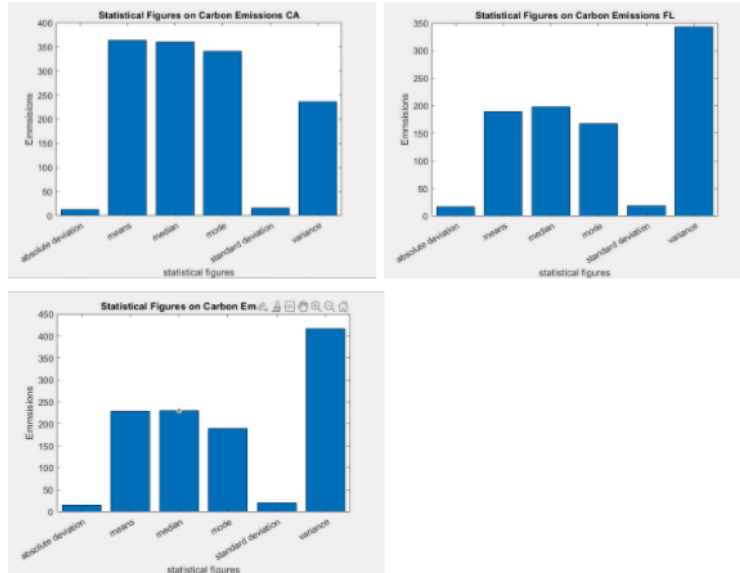
Carbon emissions graphed after running the IQR test and removing outliers. We see that the relationship remains the same in terms of carbon emissions: $CA > NY \sim FL$, keeping this in mind will help us prove the relationship between socio-economics and climate change.

5 - Emissions and income over time



When we place the two sets of graphs side by side we can conclude the relationship between climate change and socio-economics. We see that as the income of states increase, the carbon emissions also increase.

6- Statistical figures:



When looking at the statistical figure regarding the data set that we are looking at we can further affirm that the data points we used are comparable enough to draw conclusions from.

3.8 Technical Lessons Learned

What technical skills did you learn as a result of this project? Describe the functions and MATLAB tools that you learned as a result of this project.

Technical skills that accompanied the completion of the project were fairly general matlab functions and statistical skills such as finding reliable data sources, plotting with descriptive labels and for the figures, and usage of preprogrammed matlab functions, as well as navigating a vector. These preprogrammed functions include statistically significant markers such as mean, median, mode, standard deviation, etc. and return a single value. Additionally, a for loop was used with concatenation to produce the Outliers vector that was removed from the original vector. The Outliers were removed from the original vectors by the setdiff function, which accepts one dataset as a parameter, then removes common values from the second dataset parameter from the first set, and then returns that sorted vector.

Graphically, we also learned how to import spreadsheet data files into Matlab for plotting, as well as the practical skills associated with producing, graphing, and labelling such plots in desired configurations.

3.9 Professional Lessons Learned

How did your team function? What was effective? What will you do differently next time you work on a team project?

Our team collectively has a heavy interest in climate change. This allowed for us to come together and create a project based on how climate change is occurring in our communities and what has relationships to climate change. The team functioned with a passion to not only do well in the coding itself, but also to create a meaningful project that allows us to learn more about carbon emissions. To manage our time well, especially with finals coming up, we create a table of specific deadlines and meetings. This gave us space for error if we ever happened to not finish by our own deadlines. This was effective because it kept us well organized. We also learned that next time we should be more clear about the specific tasks we distribute to team members. The few bumps we had throughout the project was that we missed one small part of our project when assigning tasks. We now know that the best way to avoid that accident is to create a table that breaks down all the tasks and their descriptions, so that it clarifies things for team members.

What other professional skills did your team members learn through this project?

We were able to further develop our communication, teamwork, problem solving, and organization skills. In order for our team to function properly, we needed to make sure our communication was top notch so that there were no misunderstandings. Our teamwork was what brought our project alive. Without it, we would have one sided effort. This allowed for every member to contribute fairly to the project and every one of us is proud of it. As for problem solving, at the birth of our project, we hit a rut with our data sets. Our goal to compare several local communities and temperature change was too ambitious, so we had to alter our proposal to something more simplistic. We then came up with what we have now while still keeping our climate change theme in mind.

3.10 Summary and Conclusion

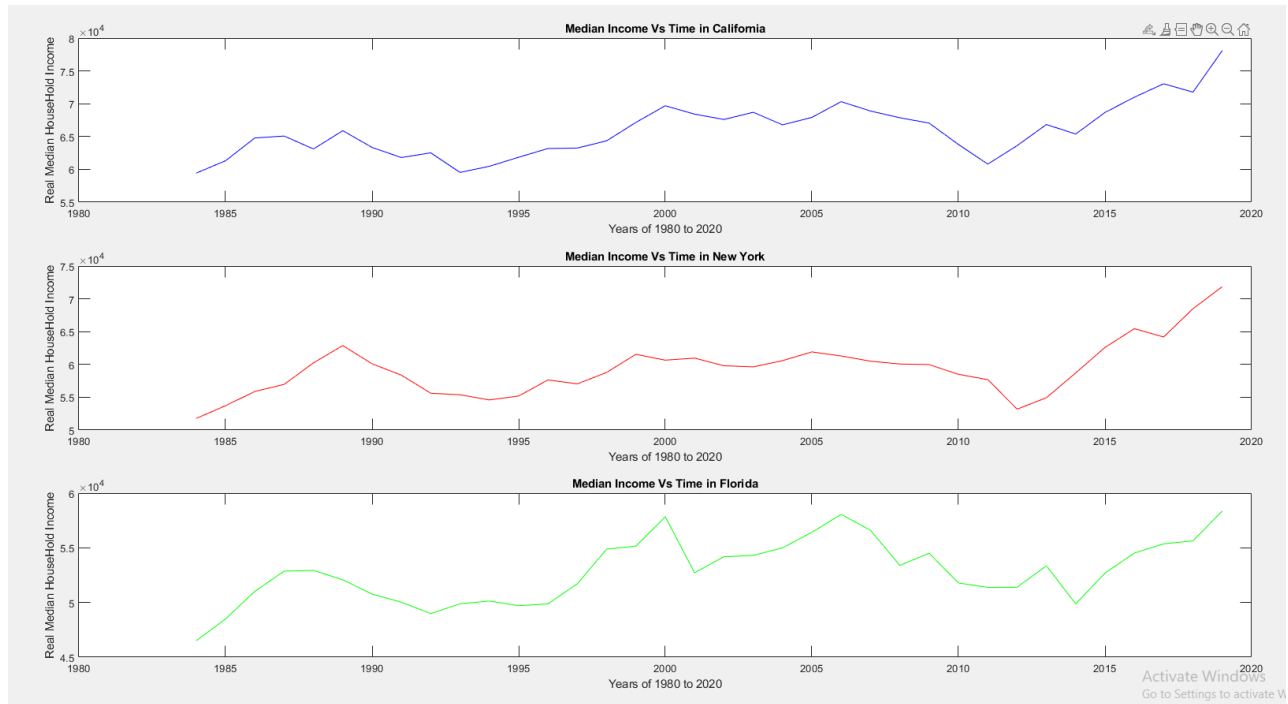
Summarize your project and lessons learned.

Our project's goal was to simplify and cement the argument that Climate Change is unnatural, and most definitely the result of human processes, shown in as recent as a mere two decades. As a result, this would refute the argument that the earth naturally heats up over time, which while true, is meant to be far more granular than the monstrous rate at which it is occurring currently. This was accomplished by us learning to program a pair of user-defined functions that would provide the statistical measures of center for various datasets to show a trend over time, as well as a secondary function that would remove outliers to show that this is occurring at a very fast and man-made rate, and not the byproduct of random natural error or phenomenon. If the rate were to worsen over the years, it would even detect the oldest or newest values as outliers if Climate Change worsens enough. This was then shown when we plotted these datasets with respect to time, to simply show the conclusion that markers such as carbon emissions are increasing at a concerning speed.

Appendix A

MATLAB Code.. Ensure that they are appropriately commented and organized so someone outside of your team can

RealMedianIncome.m:



```
%%RealMedianIncomeData.m
% California Real Median Income 1984-2019 vs Time

yearsCA = readvars('California_real_mesian.xlsx','Range','B12:B47'); % California Years 1984 to 2019
RealMedianIncomeCA = readvars('California_real_mesian.xlsx','Range','C12:C47'); % California Real Median Income

% New York Real Median Income 1984-2019 vs Time

yearsNY = readvars('NewYork_real_median.xlsx','Range','B12:B47'); % New York Years 1984 to 2019
RealMedianIncomeNY = readvars('NewYork_real_median.xlsx','Range','C12:C47'); % New York Real Median Income

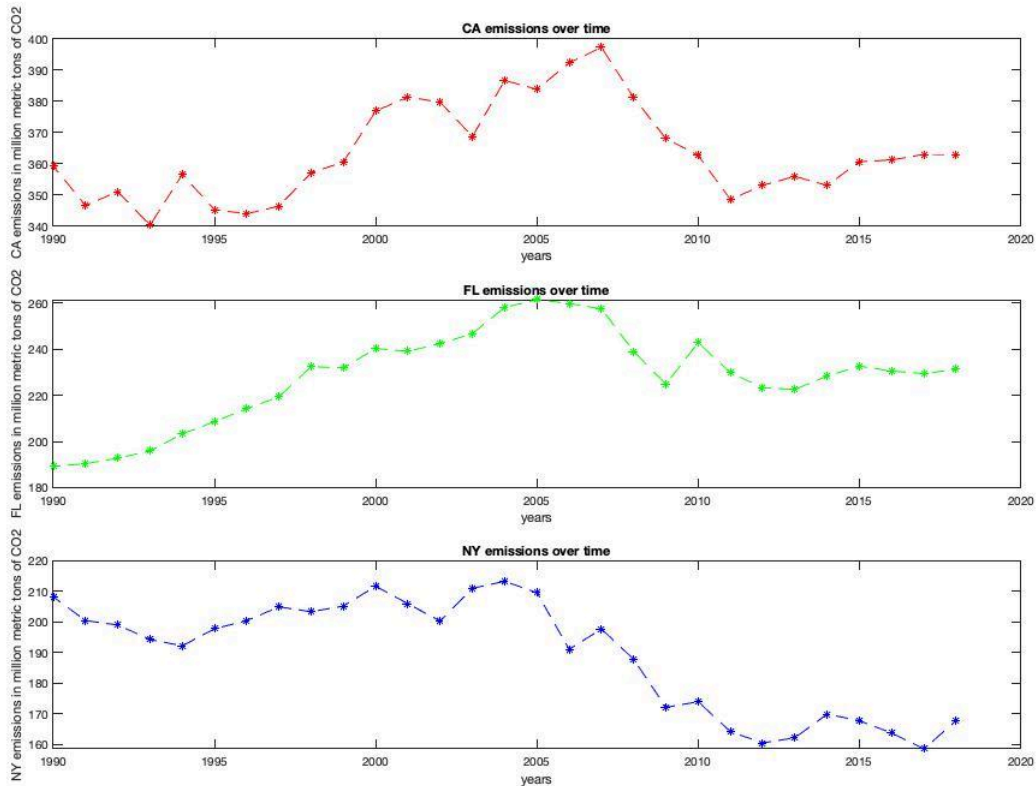
% Florida Real Median Income 1984-2019 vs Time
|
yearsFL = readvars('Florida_real_median.xlsx','Range','B12:B47'); % Florida Years 1984 to 2019
RealMedianIncomeFL = readvars('Florida_real_median.xlsx','Range','C12:C47'); % Florida Real Median Income

figure(1) % plots all 3 relationships between Real Median Income and Time (CA, NY, FL)
subplot(3,1,1)
plot(yearsCA,RealMedianIncomeCA,'b')
xlabel('Years of 1980 to 2020')
ylabel('Real Median HouseHold Income')
title('Median Income Vs Time in California')

subplot(3,1,2)
plot(yearsNY,RealMedianIncomeNY,'r')
xlabel('Years of 1980 to 2020')
ylabel('Real Median HouseHold Income')
title('Median Income Vs Time in New York')

subplot(3,1,3)
plot(yearsFL,RealMedianIncomeFL,'g')
xlabel('Years of 1980 to 2020')
ylabel('Real Median HouseHold Income')
title('Median Income Vs Time in Florida')
```

emissions_subplot_figure.m:



```
%% Making the figures for emissions vs time

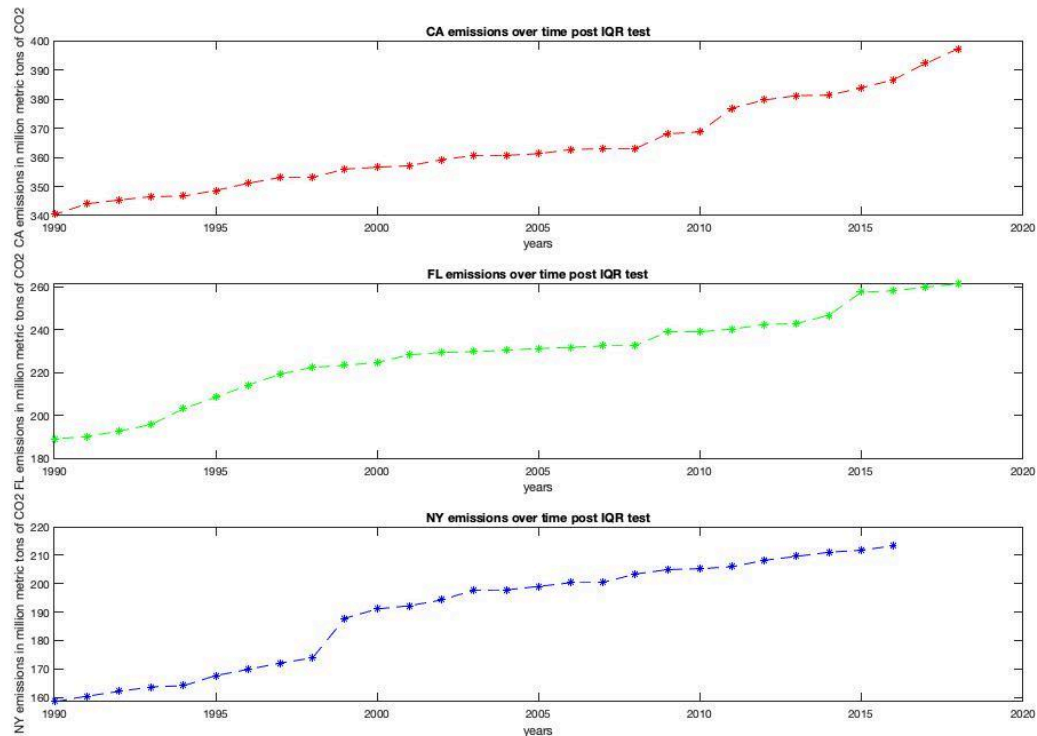
% loading in all of the data
years = xlsread('table1' , 'B5:AD5');           % set up the time vector

CAemissions = xlsread('table1' , 'B10:AD10');    % set up the CAemissions vector
FLemissions = xlsread('table1' , 'B15:AD15');    % set up the FLemissions vector
NYemissions = xlsread('table1' , 'B38:AD38');    % set up the NYemissions vector

% Making the subplot figure

subplot(3,1,1)
plot(years,CAemissions,'r*--')
xlabel('years')
ylabel('CA emissions in million metric tons of CO2')
title('CA emissions over time')
subplot(3,1,2)
plot(years,FLemissions,'g*--')
xlabel('years')
ylabel('FL emissions in million metric tons of CO2')
title('FL emissions over time')
subplot(3,1,3)
plot(years,NYemissions,'b*--')
xlabel('years')
ylabel('NY emissions in million metric tons of CO2')
title('NY emissions over time')
```

emissions_subplot_figure_post_IQR.m:



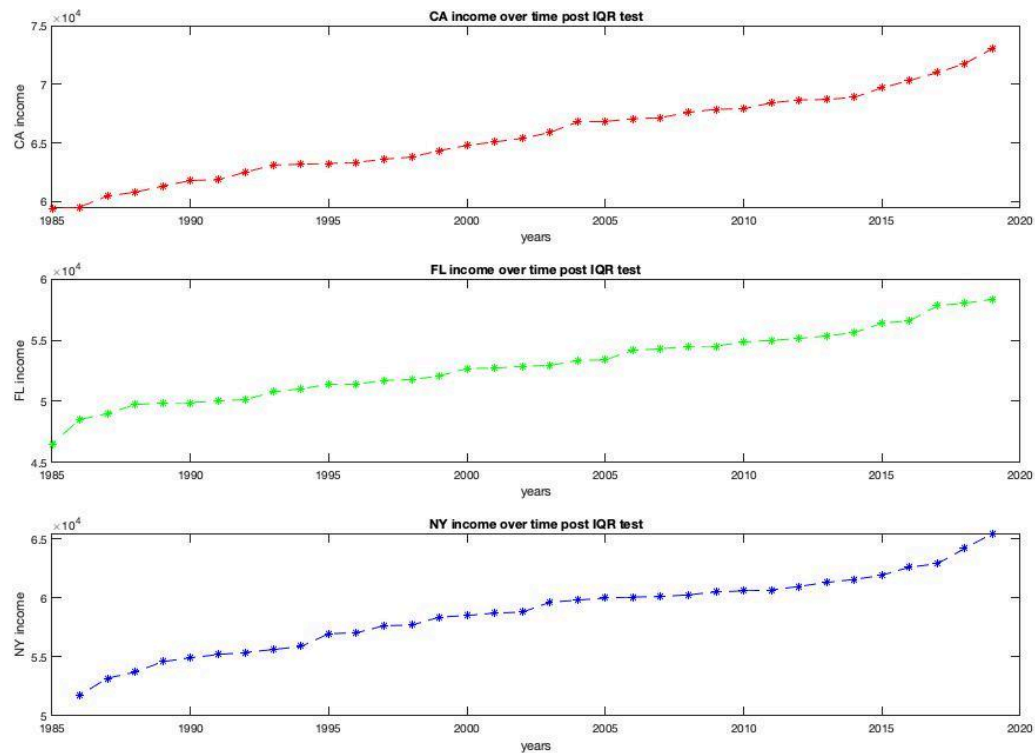
```
%%
% loading in all of the data
years = xlsread('table1' , 'B5:AD5');
years2 = xlsread('table1' , 'B5:AB5');

CAemissions = xlsread('table1' , 'B10:AD10');
FLemissions = xlsread('table1' , 'B15:AD15');
NYemissions = xlsread('table1' , 'B38:AD38');

CAemissions2 = IQR_2(CAemissions);
NYemissions2 = IQR_2(NYemissions);
FLemissions2 = IQR_2(FLemissions);

subplot
subplot(3,1,1)
plot(years,CAemissions2,'r*--')
xlabel('years')
ylabel('CA emissions in million metric tons of C02')
title('CA emissions over time post IQR test')
subplot(3,1,2)
plot(years,FLemissions2,'g*--')
xlabel('years')
ylabel('FL emissions in million metric tons of C02')
title('FL emissions over time post IQR test')
subplot(3,1,3)
plot(years2,NYemissions2,'b*--')
xlabel('years')
ylabel('NY emissions in million metric tons of C02')
title('NY emissions over time post IQR test')
```


income_post_IQR_figure.m:



```
%%
```

```
% setting up time vectors
```

```
years = readvars('NewYork_real_median.xls','Range','A13:A47');
```

```
years2 = readvars('NewYork_real_median.xls','Range','A14:A47');
```

```
%setting up income vectors
```

```
CAMedianincome = readvars('California_real_mesian.xls','Range','B12:B47');
```

```
FLMedianincome = readvars('Florida_real_median.xls','Range','B12:B47');
```

```
NYMedianincome = readvars('NewYork_real_median.xls','Range','B12:B47');
```

```
%running IQR test on income vectors
```

```
FLMedianincome2 = IQR_2(FLMedianincome);
```

```
CAMedianincome2 = IQR_2(CAMedianincome);
```

```
NYMedianincome2 = IQR_2(NYMedianincome);
```

```
% plotting the subplot |
```

```
subplot
```

```
subplot(3,1,1)
```

```
plot(years,CAMedianincome2,'r*--')
```

```
xlabel('years')
```

```
ylabel('CA income')
```

```
title('CA income over time post IQR test')
```

```
subplot(3,1,2)
```

```
plot(years,FLMedianincome2,'g*--')
```

```
xlabel('years')
```

```
ylabel('FL income')
```

```
title('FL income over time post IQR test')
```

```
subplot(3,1,3)
```

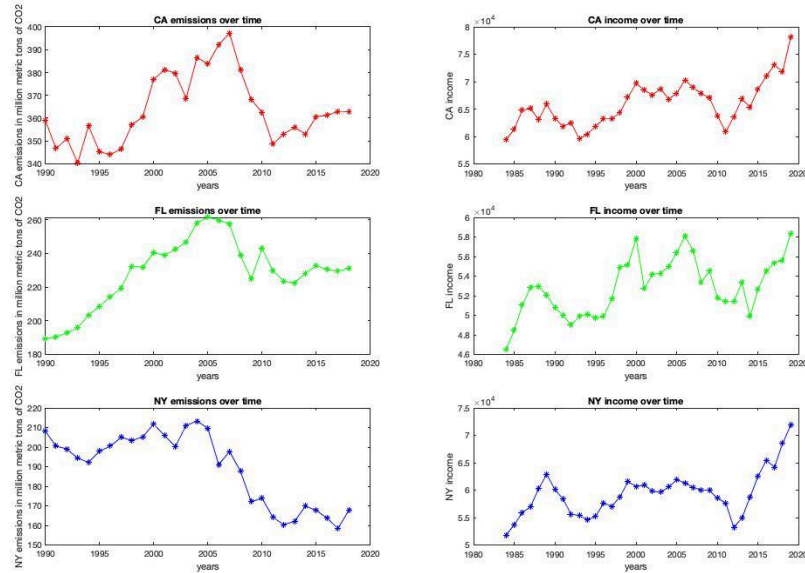
```
plot(years2,NYMedianincome2,'b*--')
```

```
xlabel('years')
```

```
ylabel('NY income')
```

```
title('NY income over time post IQR test')
```

compare_figure.m:



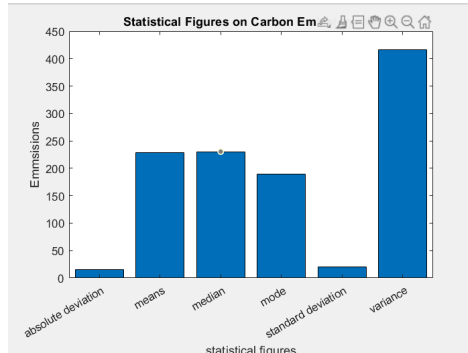
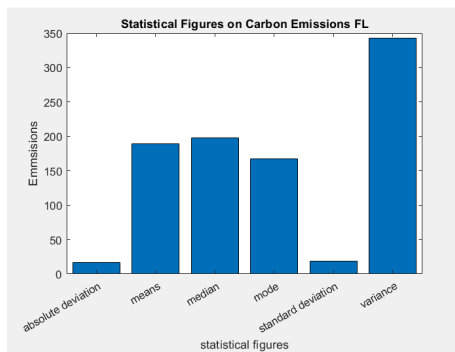
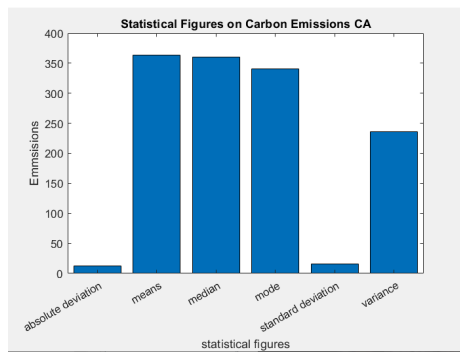
```
%%
% loading in all of the data
years = xlsread('table1', 'B5:AD5');
yearss = readvars('NewYork_real_median.xls', 'Range', 'A12:A47');
CAemissions = xlsread('table1', 'B10:AD10'); % set up the CAemissions vector
FLemissions = xlsread('table1', 'B15:AD15'); % set up the FLemissions vector
NYemissions = xlsread('table1', 'B38:AD38'); % set up the NYemissions vector
CAmedianincome = readvars('California_real_mesian.xls', 'Range', 'B12:B47');
FLmedianincome = readvars('Florida_real_median.xls', 'Range', 'B12:B47');
NYmedianincome = readvars('NewYork_real_median.xls', 'Range', 'B12:B47');

%creating subplot |
subplot(3,2,1)
plot(years,CAemissions,'r*-')
xlabel('years')
ylabel('CA emissions in million metric tons of CO2')
title('CA emissions over time')
subplot(3,2,2)
plot(yearss,CAMedianincome,'r*-')
xlabel('years')
ylabel('CA income')
title('CA income over time')
subplot(3,2,3)
plot(years,FLemissions,'g*-')
xlabel('years')
ylabel('FL emissions in million metric tons of CO2')
title('FL emissions over time')
subplot(3,2,4)
plot(yearss,FLmedianincome,'g*-')
xlabel('years')
ylabel('FL income')
title('FL income over time')
subplot(3,2,5)
plot(years,NYemissions,'b*-')
xlabel('years')
ylabel('NY emissions in million metric tons of CO2')
title('NY emissions over time')
subplot(3,2,6)
plot(yearss,NYmedianincome,'b*-')
xlabel('years')
ylabel('NY income')
title('NY income over time')
```

IQR.m :

```
%%  
function corrected = IQR(input)  
% loads data file. Input should be of the format 'Data.txt'  
vec = input;  
% finds 25th and 75th percentile (or 1st and 3rd quartile)  
firstquartile = prctile(vec, 25);  
thirdquartile = prctile(vec, 75);  
% calculates boundaries for outlier  
tolerance = iqr(vec) * 1.5;  
% assigns middle to determine whether to use upper or lower bound  
middle = median(vec);  
% initializes array of outlier values to remove  
outliers = [];  
    for i = 1:size(vec)  
        % stores distance from upper and lower bounds  
        upper = abs(vec(i) - thirdquartile);  
        lower = abs(vec(i) - firstquartile);  
        % checks whether it's above or lower than average to see if upper or lower is used  
        if ((vec(i) >= middle) && (upper > tolerance)) || ((vec(i) < middle) && (lower > tolerance))  
            % adds value to vector of identified outliers  
            outliers = [outliers, vec(i)];  
        end  
    end  
    %returns array with values of vec that are NOT in outliers  
    corrected = setdiff(vec, outliers);  
end
```

statistical_figure.m:



```
function [calculate,means, medians,modes,deviation,variance] = statistical_figures(input)
% calculate the standard deviation of the data set
calculate=std(input)
% calculate the mean of the data set
means=mean(input)
%calculate the median of a data set
medians=median(input)
%calculate the mode of a data set
modes=mode(input)
%%calculate the absolute deviation of a data set
deviation=mean( abs(input - mean(input)) )
%%calculate the variance of a data set
variance=var(input)
%presents data
Y=[calculate means medians modes deviation variance]
x=categorical({'standard deviation' 'means' 'median' 'mode' 'absolute deviation' 'variance'})
bar(x,Y)
title('Statistical Figures on Carbon Emissions')
xlabel('statistical figures')
ylabel('Emmsisions')
end
```

Appendix B

If applicable, include your data source here so that someone else could easily access the data you used for your project.

US emissions source

<https://www.eia.gov/environment/emissions/state/>

Florida income source

<https://fred.stlouisfed.org/series/MEHOINUSFLA672N>

California income source

[Real Median Household Income in California \(MEHOINUSCAA672N\) | FRED | St. Louis Fed \(stlouisfed.org\)](#)

New York income source

<https://fred.stlouisfed.org/series/MEHOINUSNYA672N>

Bibliography

1. The effects of Climate Change

<https://climate.nasa.gov/effects/>

2. Climate change: Global temperatures

<https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature>