A Data Research on Climate Change

Ning Shangguan

5/3/2021

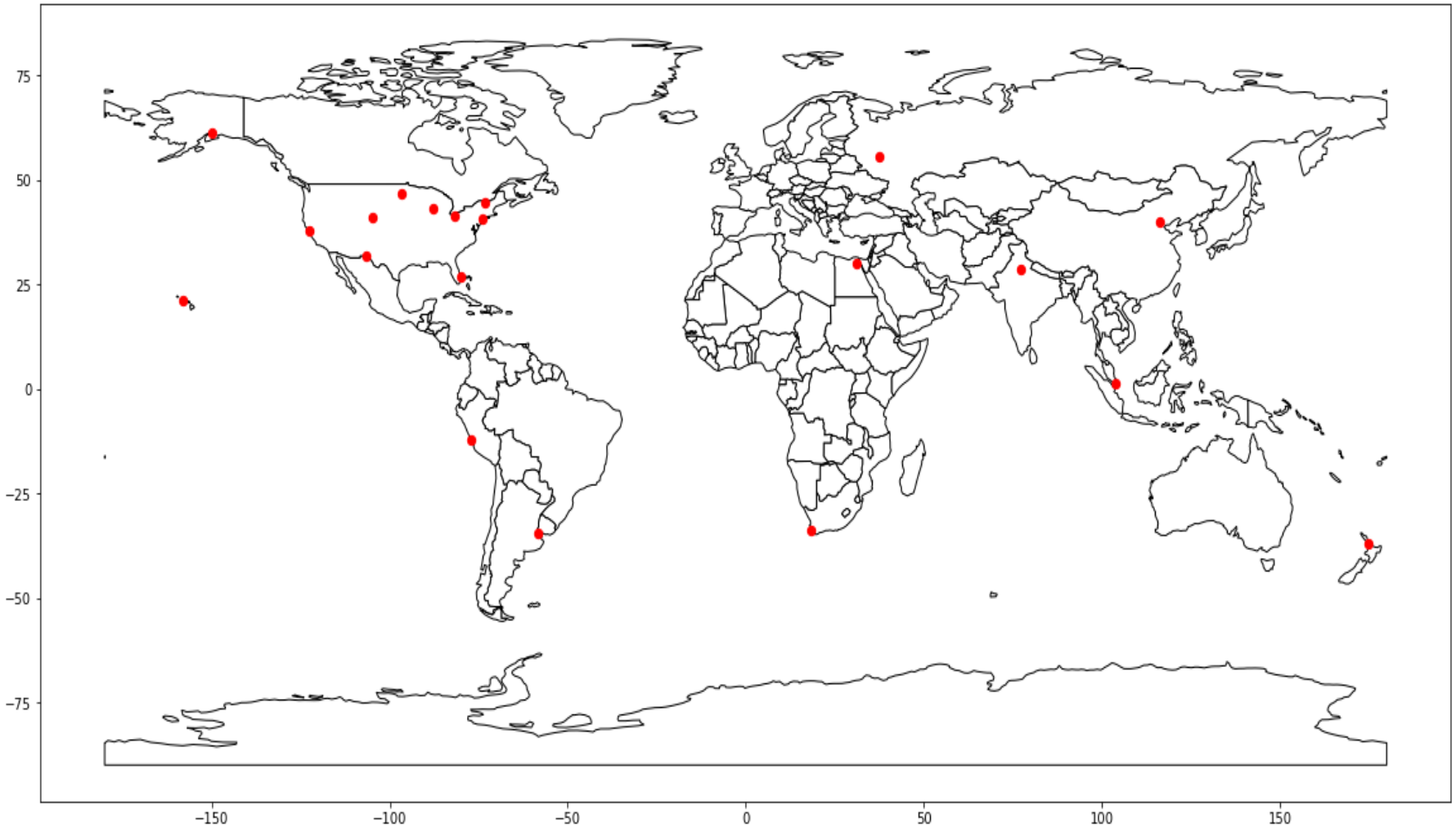
Questions about the Climate Change

1. Is the climate change real?

2. How does the climate change look like?

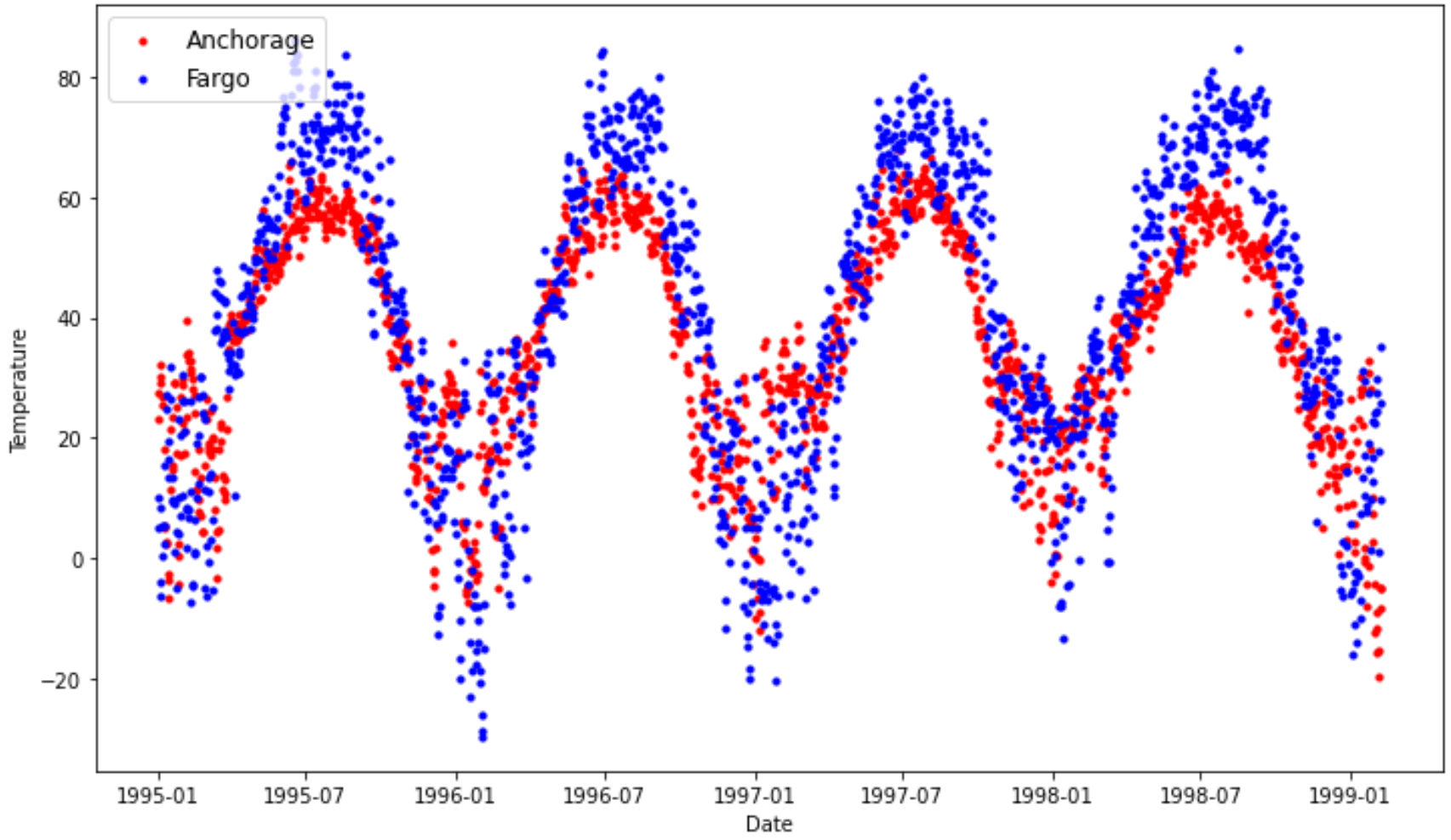
3. Is the CO_2 emission from fossil fuel causing the climate change?

Totally 20 cities were selected to see the mean temperature change over the 20 years (1995-2015)



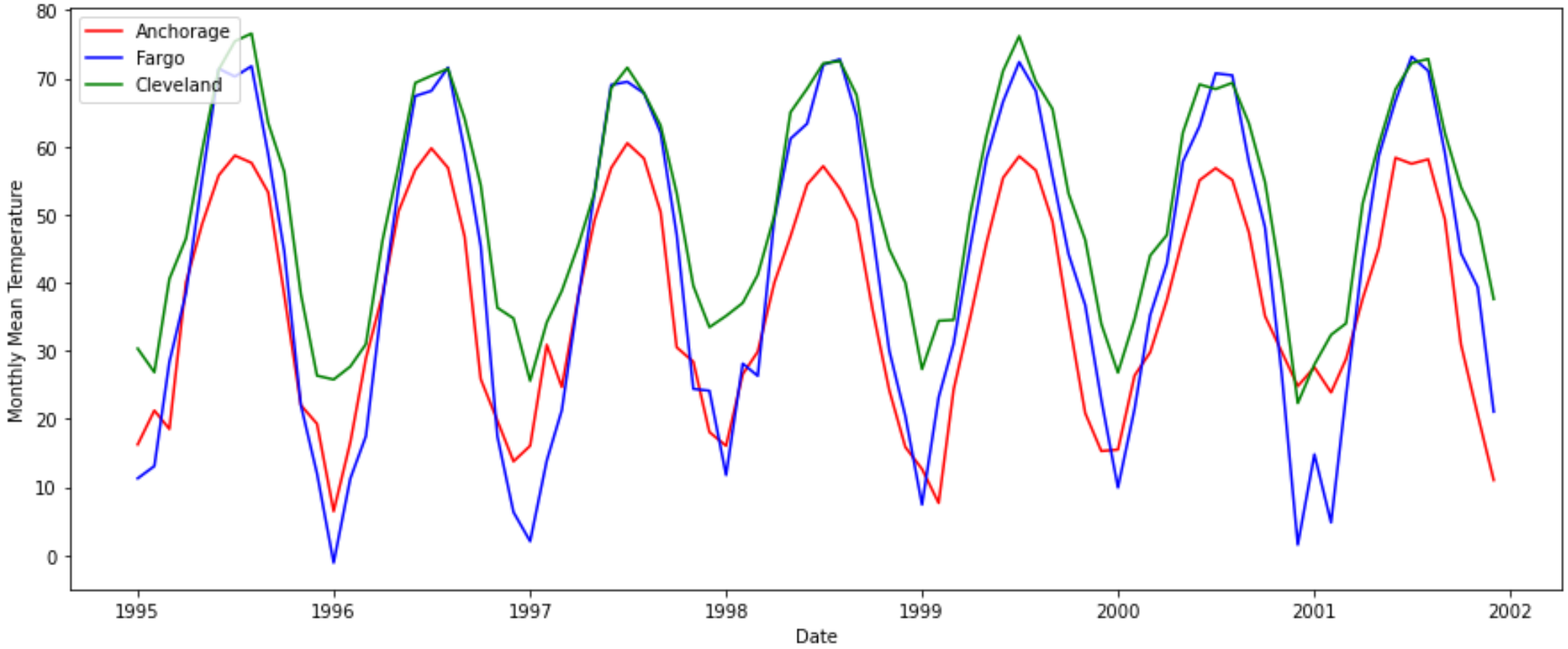
Data Source:
https://academic.udayton.edu/kissock/http/weather/

Daily Mean Temperature of Cities

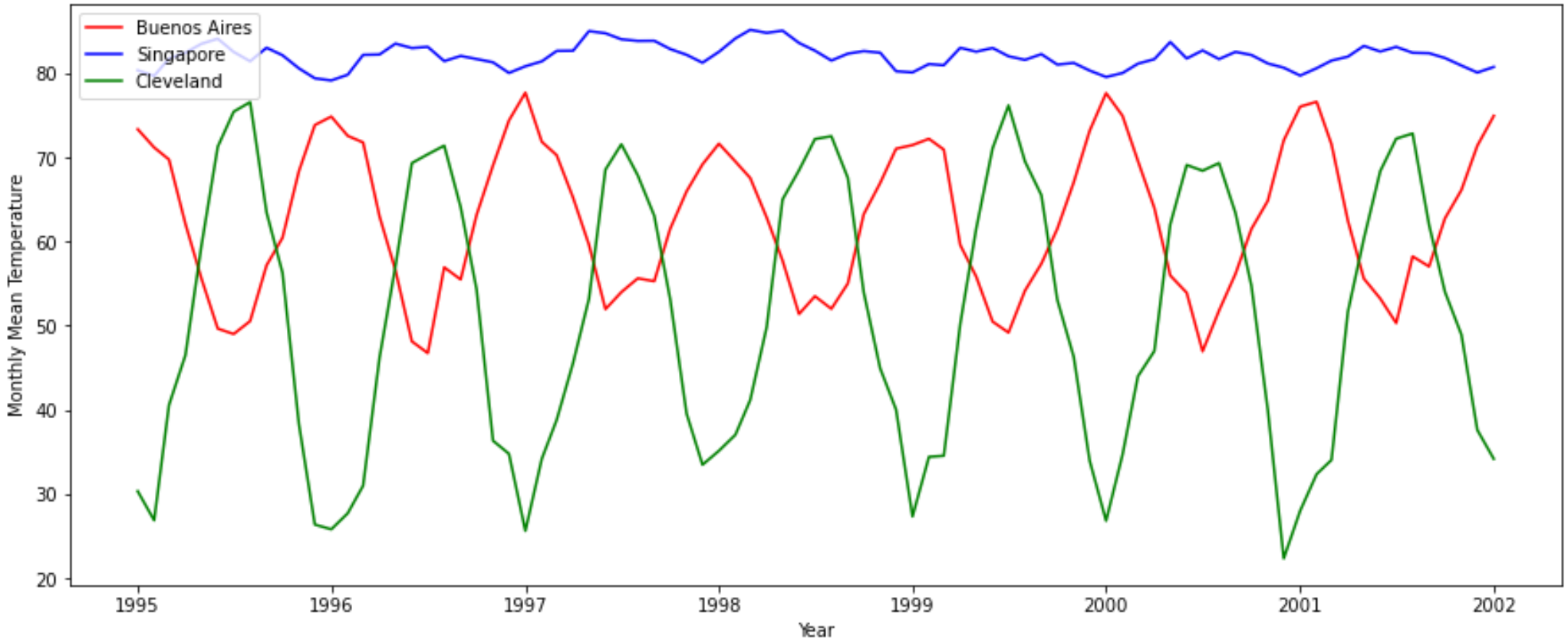


Too many data points,
Difficult to see clearly.

Monthly Mean Temperature of US Cities



Monthly Mean Temperature of International Cities



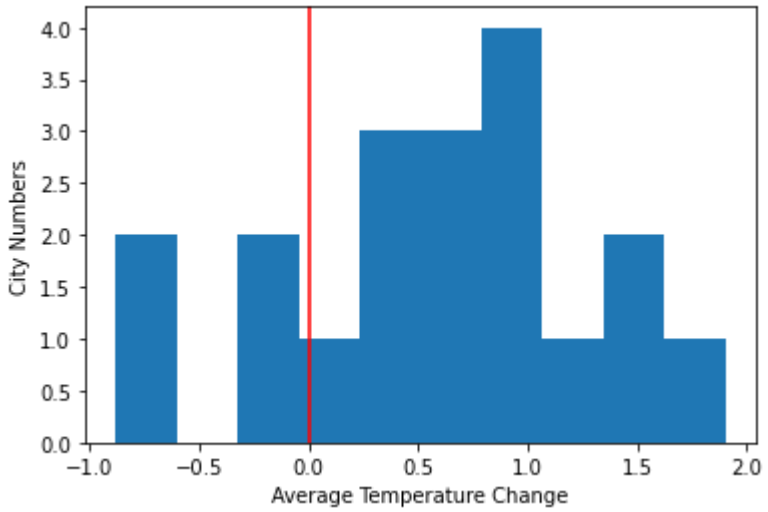
Yearly Mean Temperature Comparisons

Climate Change is still not obvious to see

Mean Temperature Changes

Take mean temperature of 1995-1999 (T1) and mean temperature of 2010-2014 (T2)
ΔT=T2-T1

AK_Anchorage_Temp	1.131106
ND_Fargo_Temp	1.083954
FL_West_Palm_Beach_Temp	0.657804
Vermont_Burlington_Temp	1.061802
TX_El_Paso_Temp	1.430969
Wyoming_Cheyenne_Temp	0.768072
San_Francisco_Temp	0.513417
Cleveland_Temp	0.820345
Milwaukee_Temp	0.385734
New_York_City	0.533899
Honolulu_Temp	-0.102903
SA_Cape_Town	1.023740
Egypt_Cairo	1.613335
India_Delhi	1.005422
China_Beijing	-0.821714
Singapore_Temp	-0.119140
New_Zealand Auckland	0.080367
Russia_Moscow	1.955066
Argentina Buenos_Aires	1.046030
Peru_Lima	-0.839321



The mean change of 20 cities is 0.66 °F

Is it statistically significant?

Hypothesis Test:

Assumption: there was no temperature change between (1995-1999 and 2010-2014). The yearly mean temperatures of each city followed the standard distribution.

Permutation:

1. Use the average of 20 years temperature and the standard deviation of the yearly mean temperature of each city to generate 10000 data (as the sample pool of the yearly mean temperatures) for each city.

2. Randomly pick 5 data (5a) from the pool as the yearly mean temperature of 1995-1999, and other 5 data (5b) as the those of 2010-2014.

ΔT = Mean(5b)-Mean(5a)

Hypothesis Test:

C = Average of 20 cities of ΔT

I repeated 1000 times and obtained a list of c.

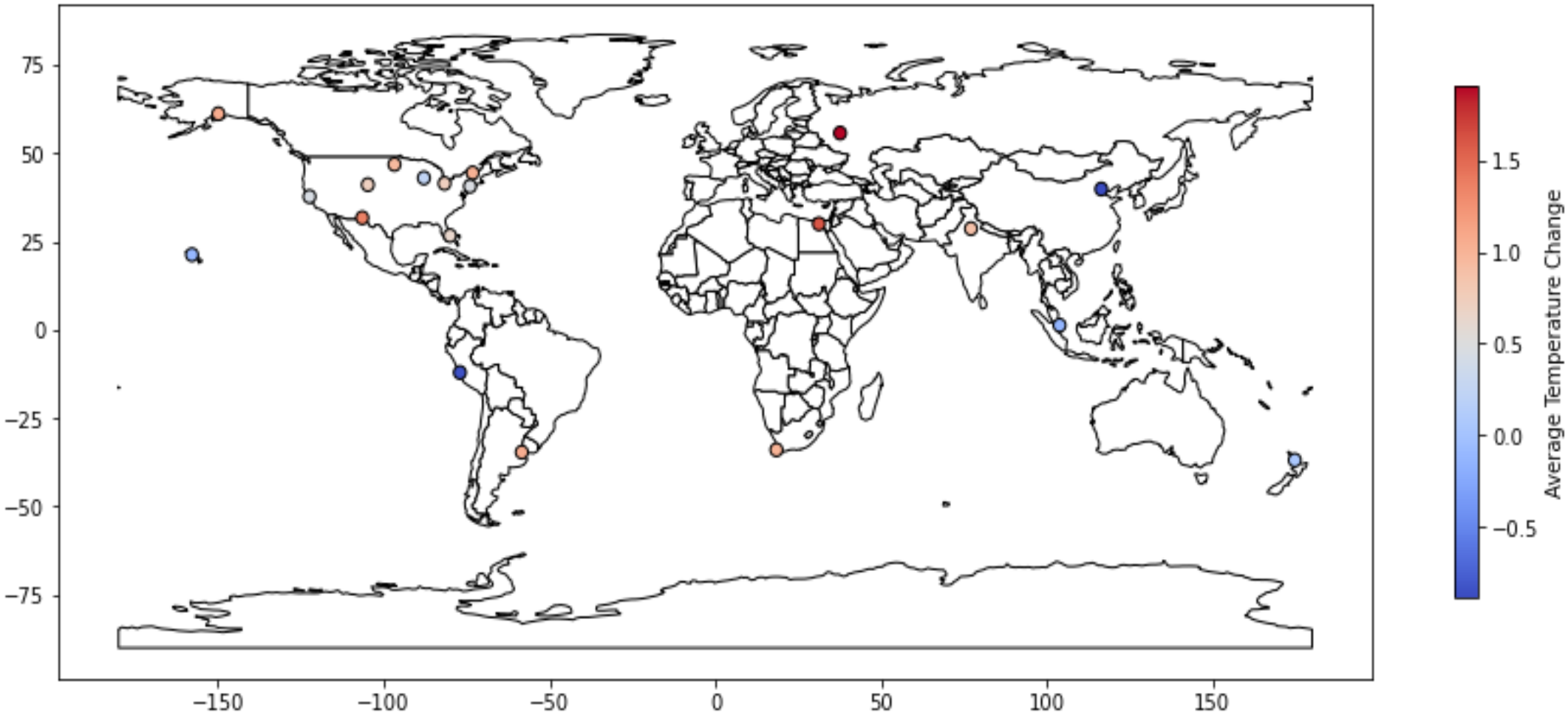
5% -95% of c is between -0.27 and 0.27 (°F)

1% -99% of c is between -0.41 and 0.41 (°F)

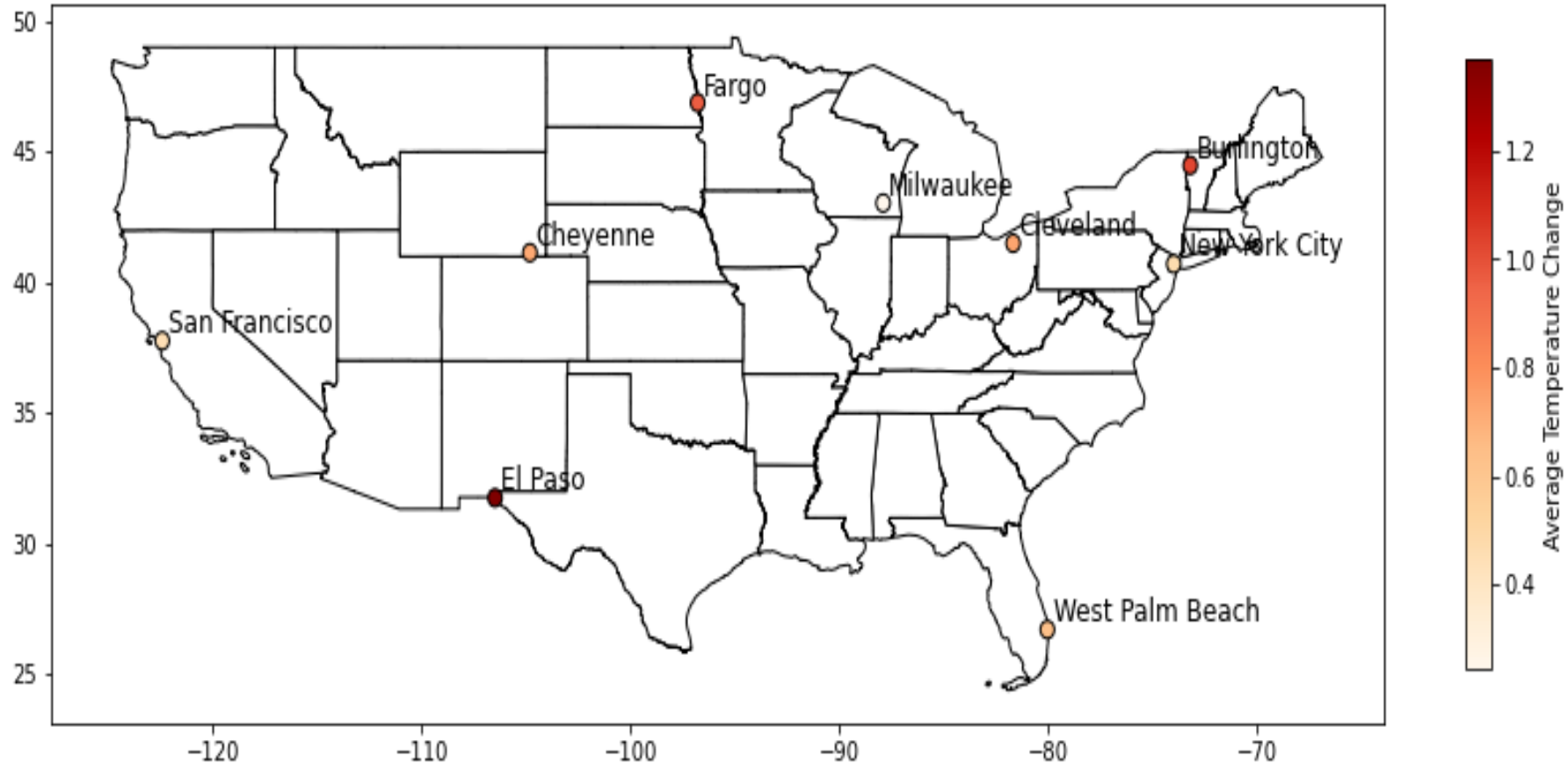
The real data of 20 cities during the 20 years gave a C of 0.66°F . It is well above 99% of the possible data in the permutation test.

The original hypothesis that there was no temperature change is wrong. The overall temperature of 20 cities increased between the years of 1995-1999 and 2010-2014.

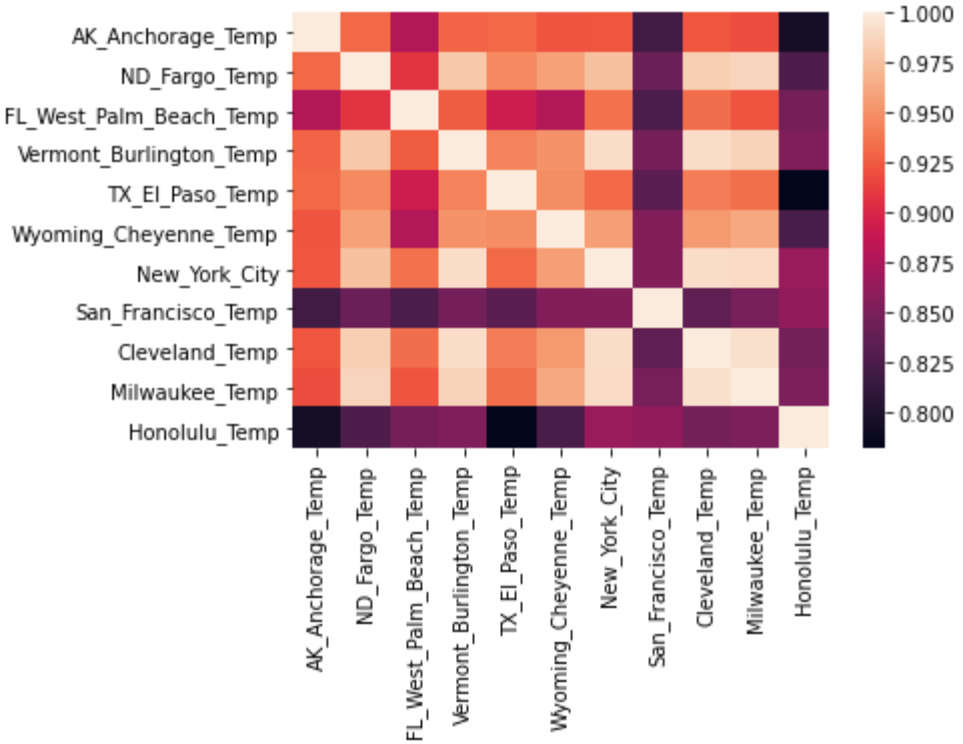
Cities and Their Five Years Mean Temperature Changes

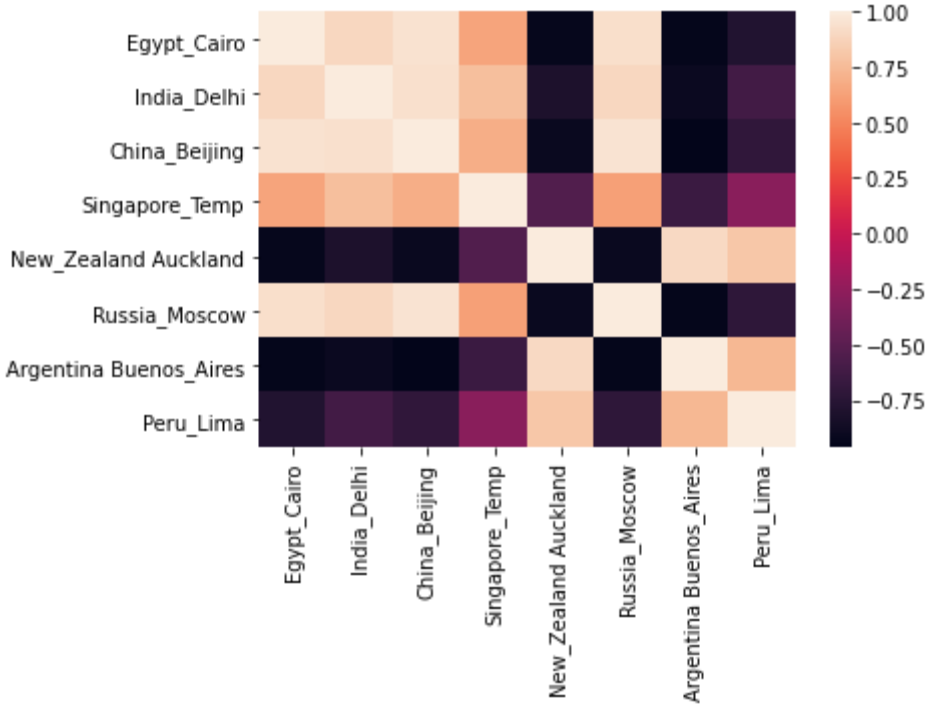


US Cities and Their Five Years Mean Temperature Changes

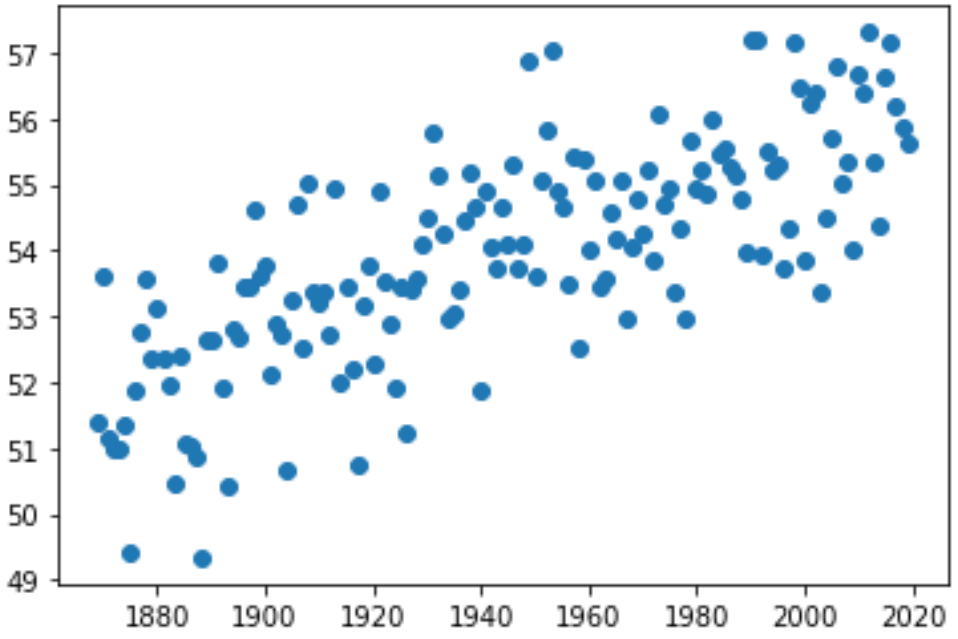


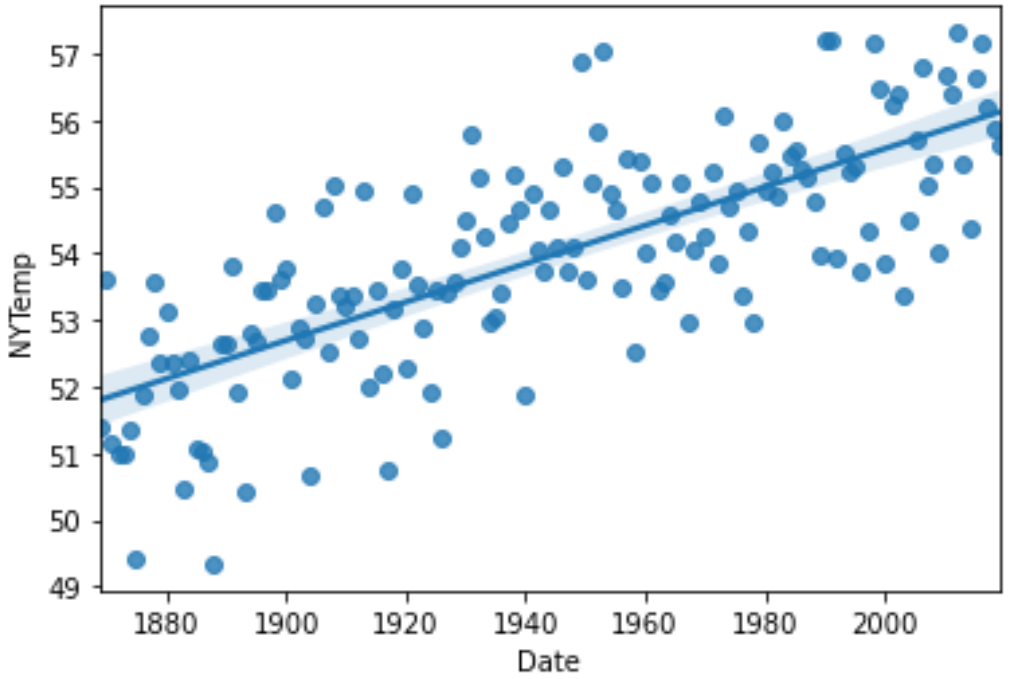
Correlation of Monthly Mean Temperature of Cities





Yearly Mean Temperature of New York City (1869-2019)





www.weather.gov/media/okx/Climate/CentralPark/monthlyannualtemp.pdf

Temperature vs Year Correlation: 0.746

T=51.88+0.0278(Y-1869) (°F)

Every 15 years, the mean temperature of NYC increases 0.53 °F, it is consistent with the previous result of NYC ΔT=0.47 °F (15 years of time span, mean temperature of 5 years).

Prediction I:

If I know some other cities (Burlington and Cleveland) of monthly mean temperature, can I predict the monthly mean temperature of New York city?

KNN model: K =2, Train Score= 0.993, Test Score=0.981

Linear Regression: Train Score= 0.994, Test Score= 0.990, MAE=1.28°F

Gradient Boosting: Train Score= 0.998, Test Score= 0.988, MAE= 1.36 °F

Conclusion: It is easy to predict a city's temperature if you know the temperature of other US cities at the same time.

Prediction II:

Based on the 150 years of historical data of the monthly mean temperature of New York city, can we predict the future temperatures of NYC?

KNN model: K=1, Test Score =0.892; K=2, Test Score =-0.272..

Linear Regression: Train Score = 0.965, Test Score= 0.959, MAE= 2.45°F

Gradient Boosting: Train Score = 0.969, Test Score= 0.931, MAE= 2.89°F

Random Forest: Test Score= 0.929, MAE= 2.95°F

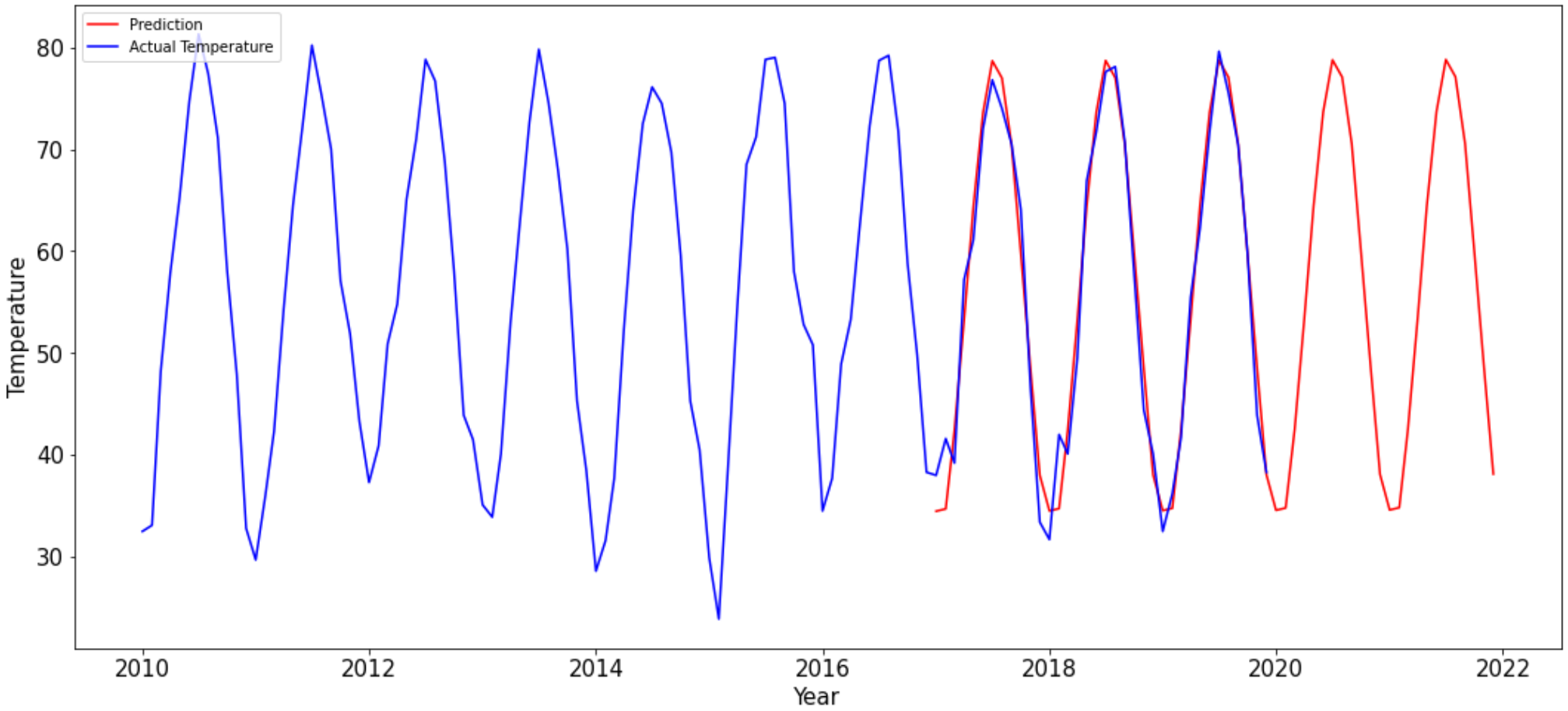
Prediction of Year 2100 (Jan-Dec, (yearly mean)) :

Linear Regression: 36.7, 37.3, 45.0, 55.7, 66.8, 75.9, 81.1,79.4, 72.8, 62.1, 50.9, 40.5. (58.7)

Gradient Boosting: 33.8, 34.7, 42.3, 52.8, 62.8, 71.3, 76.3, 74.7, 67.9, 57.2, 47.9, 29.7. (54.3)

Year 2018: 31.7, 42, 40.1, 49.5, 66.9, 71.7, 77.6, 78.1, 70.7, 57.7, 44.4, 40.1. (55.9)

NYC Actual Monthly Mean Temperature and the Prediction by Linear Regression



Prediction III

If I know the monthly mean temperature of NYC, Milwaukee, Cleveland and Burlington, can I predict the monthly mean temperature of NYC next year?

KNN model: K =4, Train Score= 0.951, Test Score=0.916

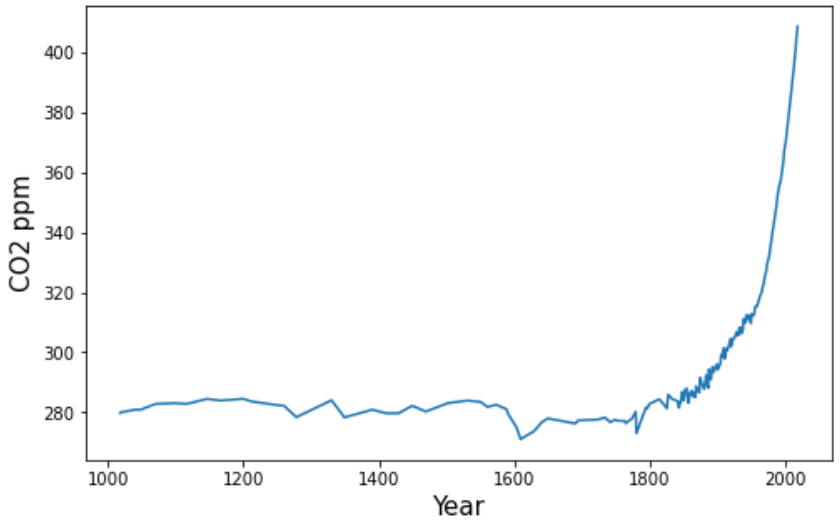
Linear Regression: Train Score= 0.969, Test Score= 0.969， MAE=2.19

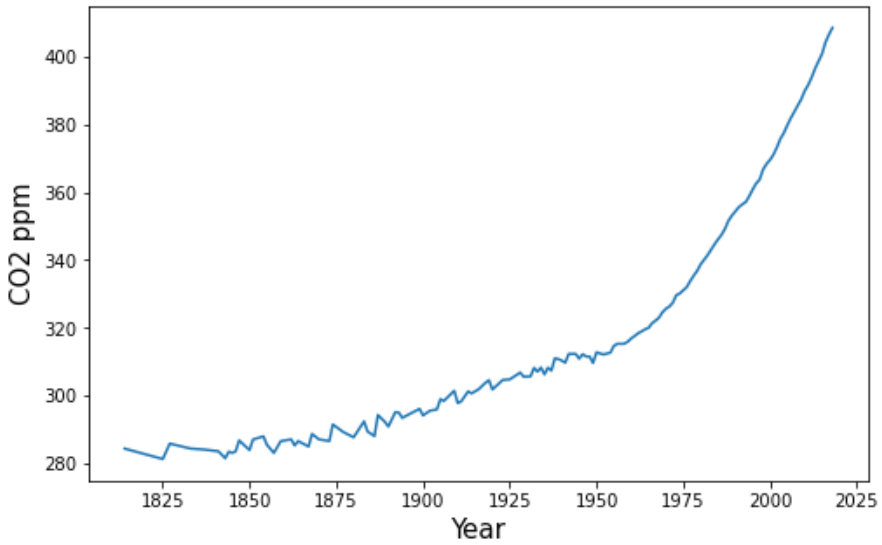
Gradient Boosting: Train Score= 0.992, Test Score= 0.914

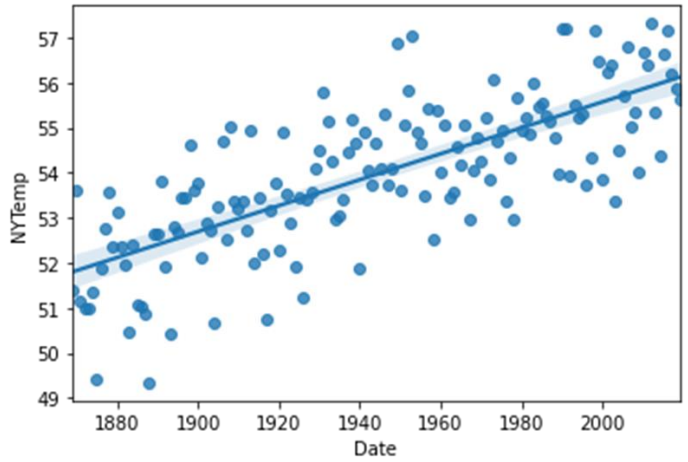
Conclusion: Adding data of other US cities help predicting a US city's temperature next year slightly better.

The Remaining Question: The Relationship Between CO_2 Concentration and the Earth Mean Temperature

Data from https://ourworldindata.org/grapher/global-co-concentration-ppm







With the recent years of rapid increase of CO_2 concentration in atmosphere, could we still expect the linear increase of the mean temperature?

Summary:

1. The climate change of earth is a complicated phenomenon, 20 years of climate data is not big enough to give us a clear understanding. During the 20 years, we see many variations and fluctuations of mean temperature of each year.
2. Overall, most of the selected cities have increased mean temperature over 5 years compared with 15 years ago, it shows the trend of the earth temperature increase and the variations of the climate change.
3. The 150 years of weather data of New York City shows a clear picture of the yearly mean temperature increase trend.
4. The relationship between the CO_2 concentration in atmosphere and the mean temperature of earth still needs more research and waits for a better answer.