Muhammad Raihan Muhaimin
Student number 301349274

Part 2: Data analysis with Matplotlib:

a) Produce at least **two figures** that illustrate the **max. temperature distribution over the entire globe** and enable a comparison of **different non-overlapping time periods**. Only show temperatures where you have data available.

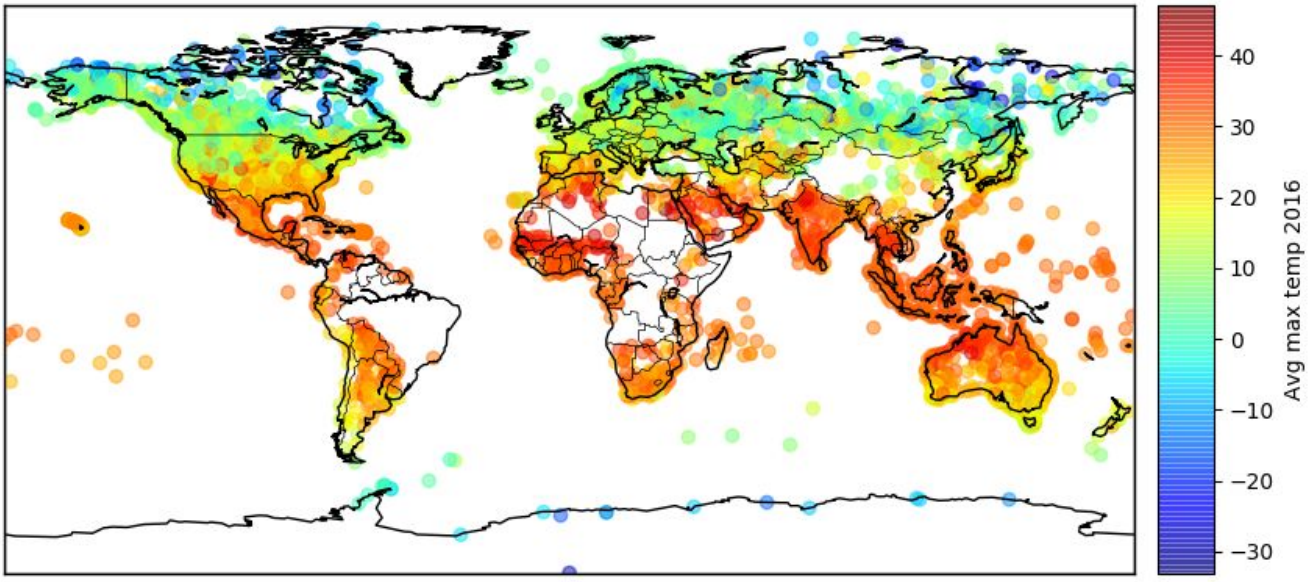

Fig: Avg of max temp reported by weather station throughout year 2016

The above figure shows average of all the max temperature reported by various weather station in year 2016. From our figure we get approximation about the hotter area and coolear area in the earth in year 2016

Muhammad Raihan Muhaimin
Student number 301349274

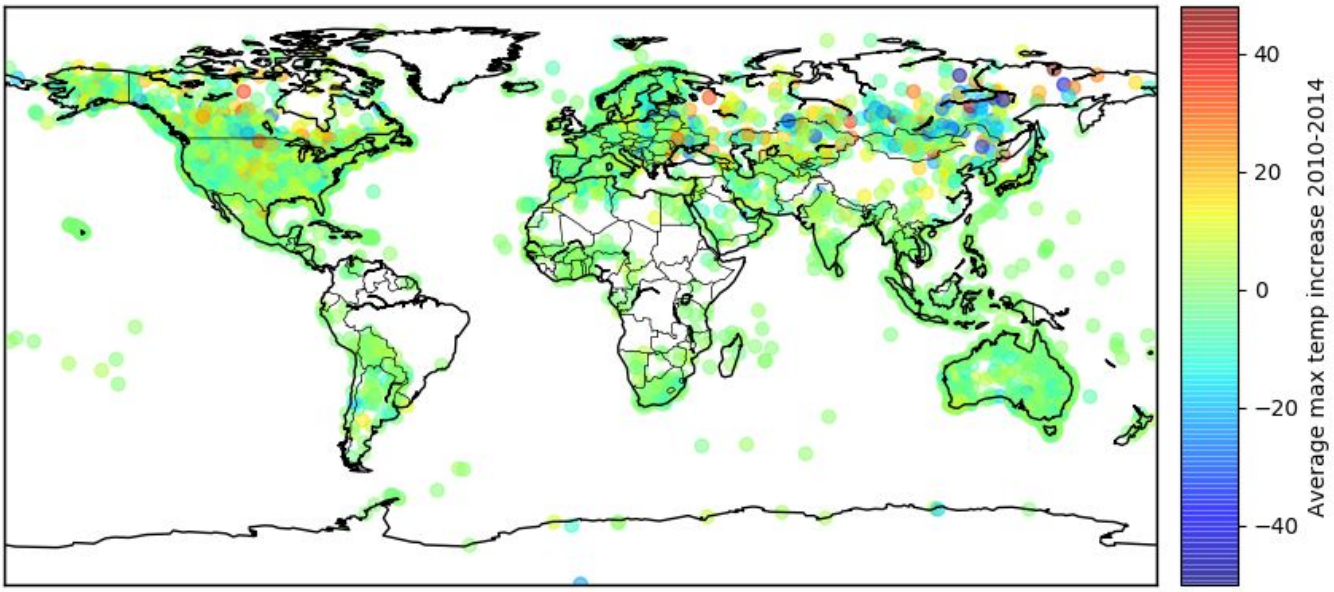

Fig: Increase of avg max temperature for each weather station from year 2010 - 2014

In the above figure we are observing the difference of avg max temperature of year 2014 and avg max temperature of 2010. From this figure we can see in the polar region of the earth, some of the weather station reported temperature increase over the time.

Muhammad Raihan Muhaimin
Student number 301349274

b1)

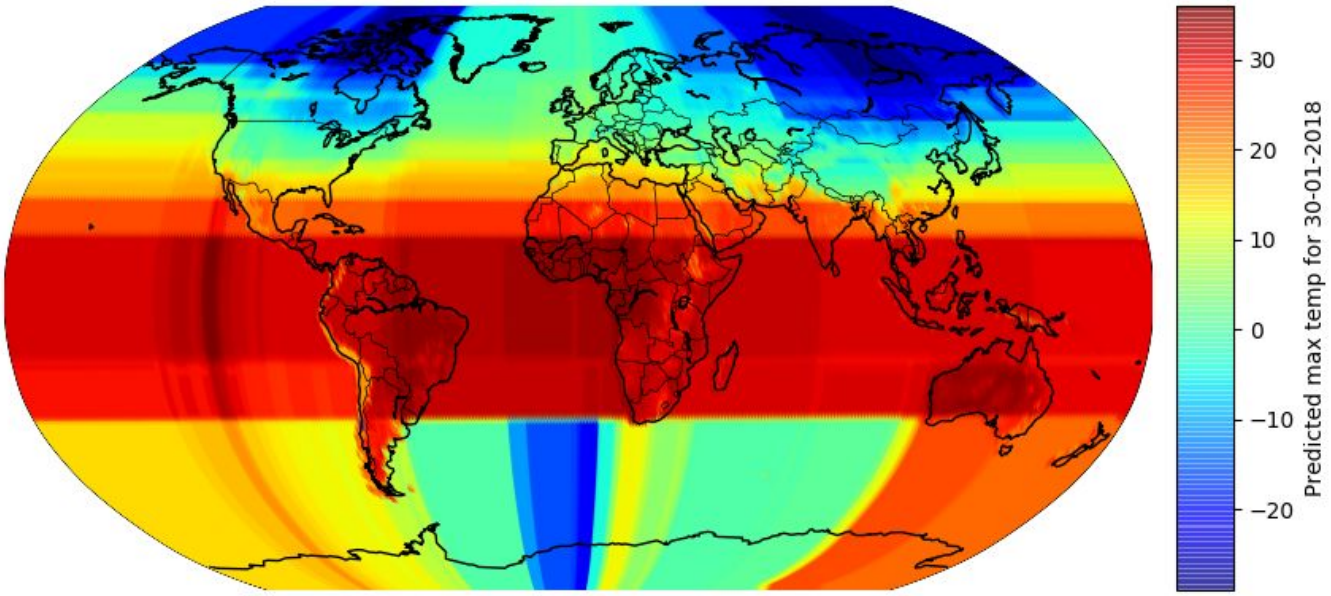

Fig: Use our weather prediction to create a dense plot of max temp for the present day

The above figure shows a dense plot of max temperature throughout the globe which was predicted by the weather prediction model we developed in Fall 2017 as a part of SFU Big Data masters course work. To train the model we took 75% datas from the tmax-2 dataset. This program takes latitude, longitude, elevation and the date of the current day to predict max temperature. Interestingly in the antarctica region especially in the south of Australia it predicted very hot, though the temperature there is much low. One possible reason for this can be we don't have enough weather station in that area. So lack of data leads to wrong prediction.

Muhammad Raihan Muhaimin
Student number 301349274

b2)

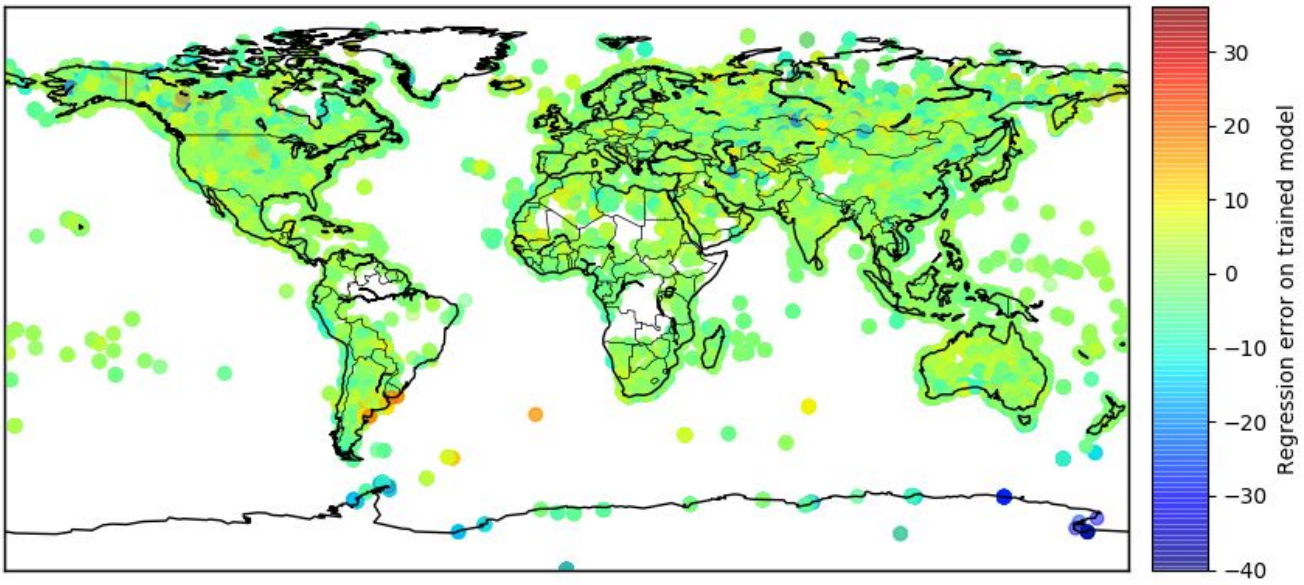

Fig: Evaluation of our prediction model using known data

Here we are using 25% of data of tmax-2 dataset which we keep separated from our training model to evaluate our prediction model. We can see majority of our prediction is within the range of 10 degrees of actual data. It also indicates our prediction error in antarctica region where it is showing our error is between -10 to -40 which means actual temperature is 10 to 40 degree lower than our predicted temperature.