Observations and Results:

Note: Please zoom in to see the charts.

1. Study of Stations

* The code parses the stations file and skips the lines which does not represent the metadata.

Output from console skipping the lines

Unable to parse line 1:Integrated Surface Database Station History, April 2018

Unable to parse line 2:

Unable to parse line 3: USAF = Air Force station ID. May contain a letter in the first position.

Unable to parse line 4: WBAN = NCDC WBAN number

Unable to parse line 5: CTRY = FIPS country ID

Unable to parse line 6: ST = State for US stations

Unable to parse line 7: ICAO = ICAO ID

Unable to parse line 8: LAT = Latitude in thousandths of decimal degrees

Unable to parse line 9: LON = Longitude in thousandths of decimal degrees

Unable to parse line 10: ELEV = Elevation in meters

Unable to parse line 11:BEGIN = Beginning Period Of Record (YYYYMMDD). There may be reporting gaps within the P.O.R.

Unable to parse line 12: END = Ending Period Of Record (YYYYMMDD). There may be reporting gaps within the P.O.R.

Unable to parse line 13:

Unable to parse line 14:Notes:

Unable to parse line 15:- Missing station name, etc indicate the metadata are not currently available.

Unable to parse line 16:- The term "bogus" indicates that the station name, etc are not available.

Unable to parse line 17:- For a small % of the station entries in this list, climatic data are not

Unable to parse line 18: available. To determine data availability for each location, see the

Unable to parse line 19: 'isd-inventory.txt' or 'isd-inventory.csv' file.

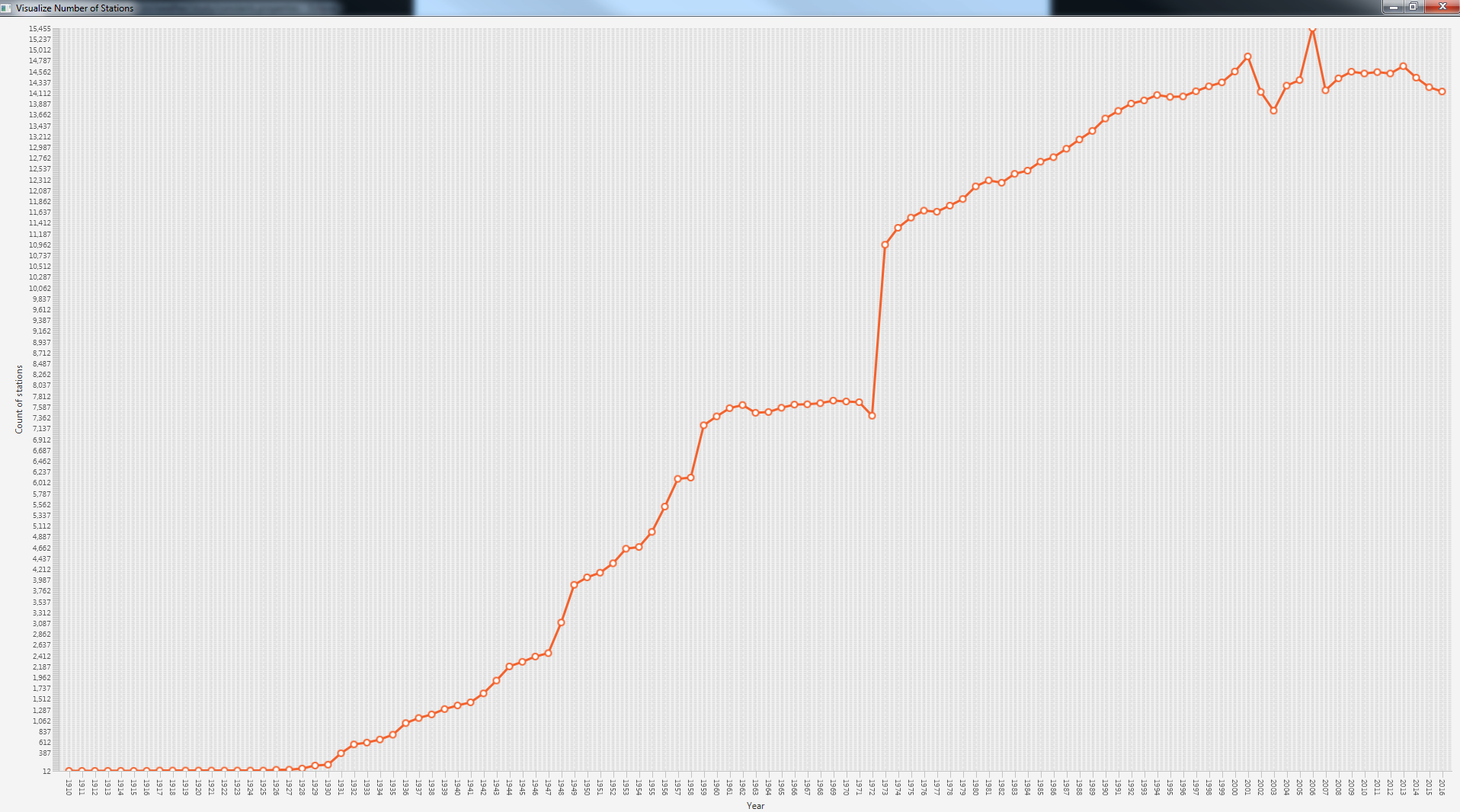
Unable to parse line 20:

Unable to parse line 21:USAF WBAN STATION NAME CTRY ST CALL LAT LON ELEV(M) BEGIN END

Unable to parse line 22:

* Chart showing number of stations each year. Data comes from the station metadata file isd-history.txt:
  + There is sudden growth in number of weather stations ~1972 .
  + Reached plateau in ~2000.

Number of total stations worldwide from 1910 to 2016



* Derivative chart showing change in number of stations for each year. Data comes from the station metadata file isd-history.txt:

Change in Number of Stations from Previous Year (Mathematical Derivative) from 1911 to 2016



Stations were growing in numbers, except a couple of times when there were dips in the above chart.

1. Study of Weather for different stations. I tried to generate an average monthly air temperature chart for Oslo and New Delhi (Safdarganj). I chose these 2 as one is non-polluted and the other is polluted. Looking at the charts, it does not look like pollution is impacting the structure of air temperature curve.

* File 014843-99999-final.txt: Data from Oseberg Station (the city is Oslo in Norway, which is considered to be one of the most non-polluted and green cities in the world). Data is from 1995-2017. 2018 data in this graph is predicted; the 2018 data is not read from the file.

Location of First Analysis



Prediction output (console) for Oslo in 2018 using an average of the previous years (1995-2017).

Prediction for month 1 of 2018: 5.288727109119397 degrees of Celsius in OSEBERG

Prediction for month 2 of 2018: 4.737114214791209 degrees of Celsius in OSEBERG

Prediction for month 3 of 2018: 5.111511219517693 degrees of Celsius in OSEBERG

Prediction for month 4 of 2018: 6.473096056488887 degrees of Celsius in OSEBERG

Prediction for month 5 of 2018: 8.475702306041965 degrees of Celsius in OSEBERG

Prediction for month 6 of 2018: 10.87952519437852 degrees of Celsius in OSEBERG

Prediction for month 7 of 2018: 13.32729040330503 degrees of Celsius in OSEBERG

Prediction for month 8 of 2018: 14.087891859835274 degrees of Celsius in OSEBERG

Prediction for month 9 of 2018: 12.615970157278367 degrees of Celsius in OSEBERG

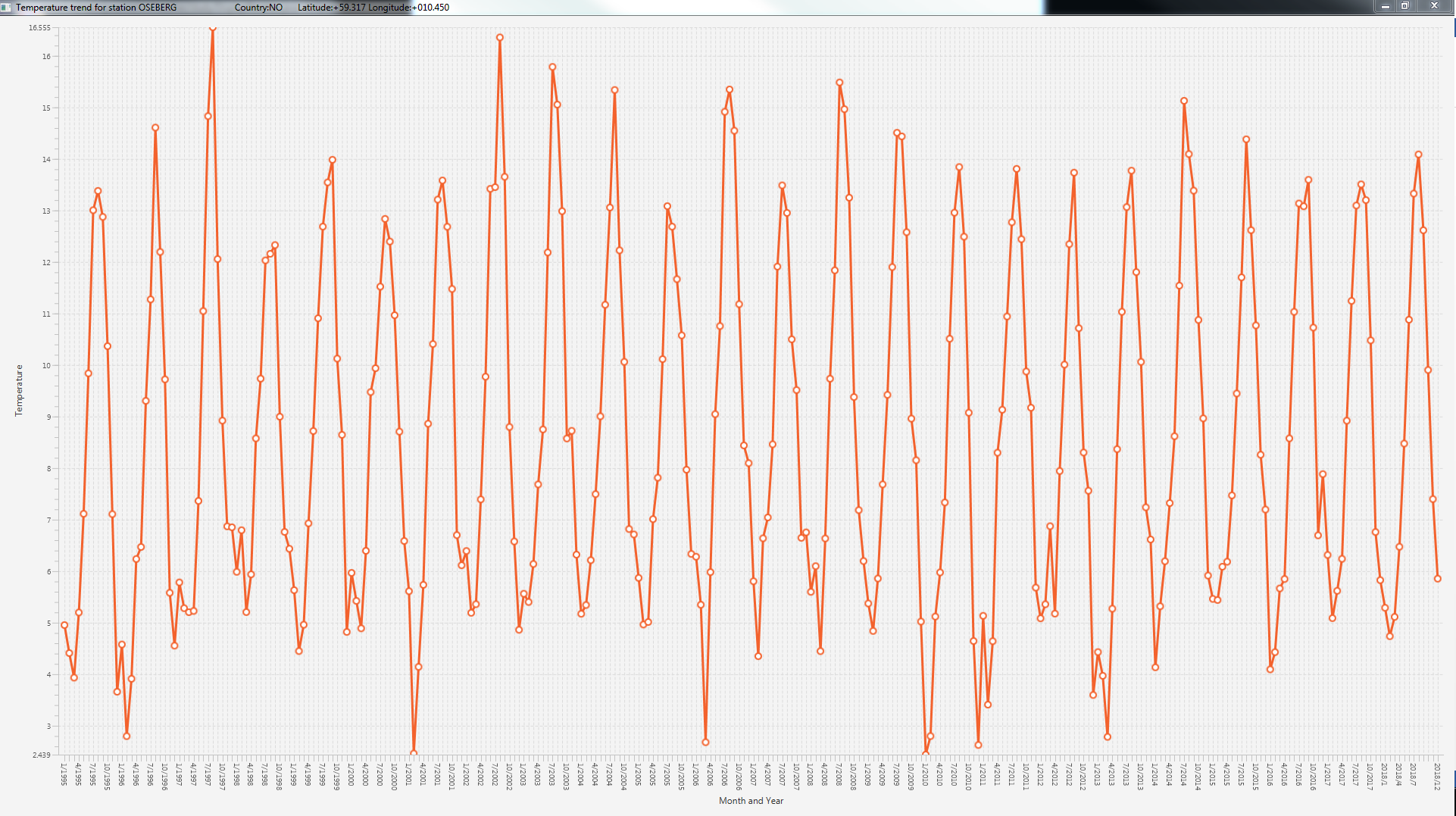
Prediction for month 10 of 2018: 9.903255955832027 degrees of Celsius in OSEBERG

Prediction for month 11 of 2018: 7.398440318614769 degrees of Celsius in OSEBERG

Prediction for month 12 of 2018: 5.853991728003633 degrees of Celsius in OSEBERG

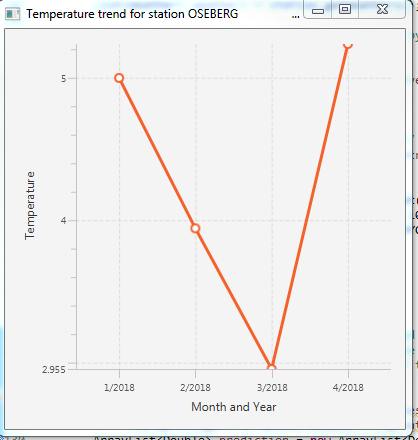
==============

Average Monthly Temperature (degrees Celsius) in Oslo from 1995 to 2017 (plus the predicted 2018 curve)



We also observe here that initially after every 4 years, the average temperature gets cooler.

* File 014843-99999-2018.txt – The chart below maps the real 2018 data so far from the Oseberg Station.

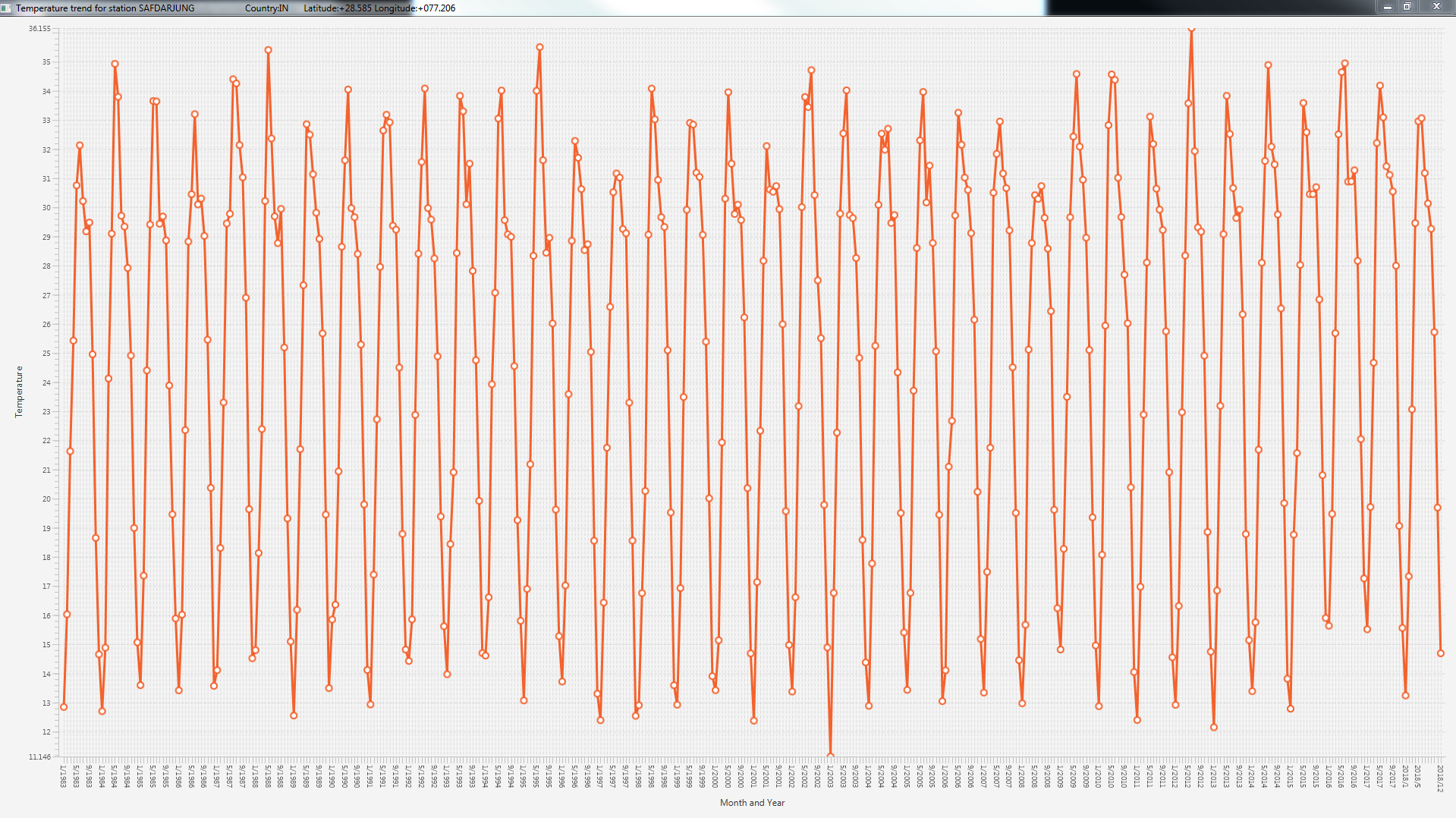


* File - 421820-99999-final.txt – Data for station SAFDARGUNJ (From Delhi, India which is considered one of the most polluted cities in the world) from 1983 to 2018

Location of Second Analysis

* 

Average Monthly Temperature (degrees Celsius) in Safdargunj from 1995 to 2017



Looking at the pattern, there does not seem to be any visible difference. Does not look like any global warming effect, at least from the data from last ~25 years in Delhi.

Code flow and enhancements:

* The file names to be parsed are in the constants.properties file.
* Fixed file formats field lengths and year range for chart are hard-coded but can be moved in properties file as well.
* One test case has been added for loading stations as that is first thing required for any charts to generate. Other test cases can be added as well.
* Code filters out the data with missing air temperatures from the calculations.
* The code currently loads one weather file at a time. It expects the file name as 014843-99999-final.txt. The USAF number 014843 and WBAN number 99999. It can be enhanced to read multiple files at a time too. Currently I have ftp the files and cat them in single file for data from one station to study it.
* The code can be easily enhanced to study other weather properties for a given station/country and find the correlation between different properties.
* I have used the average method to predict the temperatures but more accurate regression or neural network algorithms should be used.
* The code prints out in the console the line number and contents from files which it’s not able to parse. Those can be configured to write to a file.