

ABE65100 Lab 9 METADATA File

I. Description

This metadata file is designed to accompany the python program file named "program_09.py", created by Joshua Tellier on 3/18/2020 using the Purdue University scholar cluster. The program is designed to take an input file named "DataQualityChecking.txt", import the dataset, and apply four data checking corrections to it. The program then makes four plots comparing each variable before and after data quality checking. Finally, it exports two files, one which contains the corrected data and on ehwhc provides a summary of the corrections made.

II. Check Summary

The four data quality checks are as follows:

Check 1: Remove No Data Values

This check replaces the placeholder "no data" number (i.e. -999) with a NaN value that is understood by python. This check removed 2 values in each of the first three variables. The wind speed variable had no missing data and so was unaffected (see section III).

Check 2: Gross Errors

This check removes any values for each variable that are clearly out of the realm of possibility. For example, a negative precipitation value would be removed by this check. This check removed at least two values from each variable, and many more values in both the Precipitation and Max Temp variables (see section III).

Check 3: Tmax-Tmin Swap

For this check, any data observation in which the max temperature was less than the minimum temperature are swapped. It is therefore assumed that these values were transcribed incorrectly and should have been switched. In the dataset, 4 values were swapped in this way (see section III).

Check 4: Tmax-Tmin Range

For this check, any dates in which the difference between Max and Min temperature was greater than 25 degrees celsius were expunged. 5 days worth of temperature values were removed from the dataset due to this check (see section III).

III. Correction Reference Table

	Precipitation	Max Temp	Min Temp	Wind Speed
1. No Data	2	2	2	0
2. Gross Error	15	14	2	2
3. Swapped	0	4	4	0
4. Range	0	5	5	0

IV. Variable Correction Plots

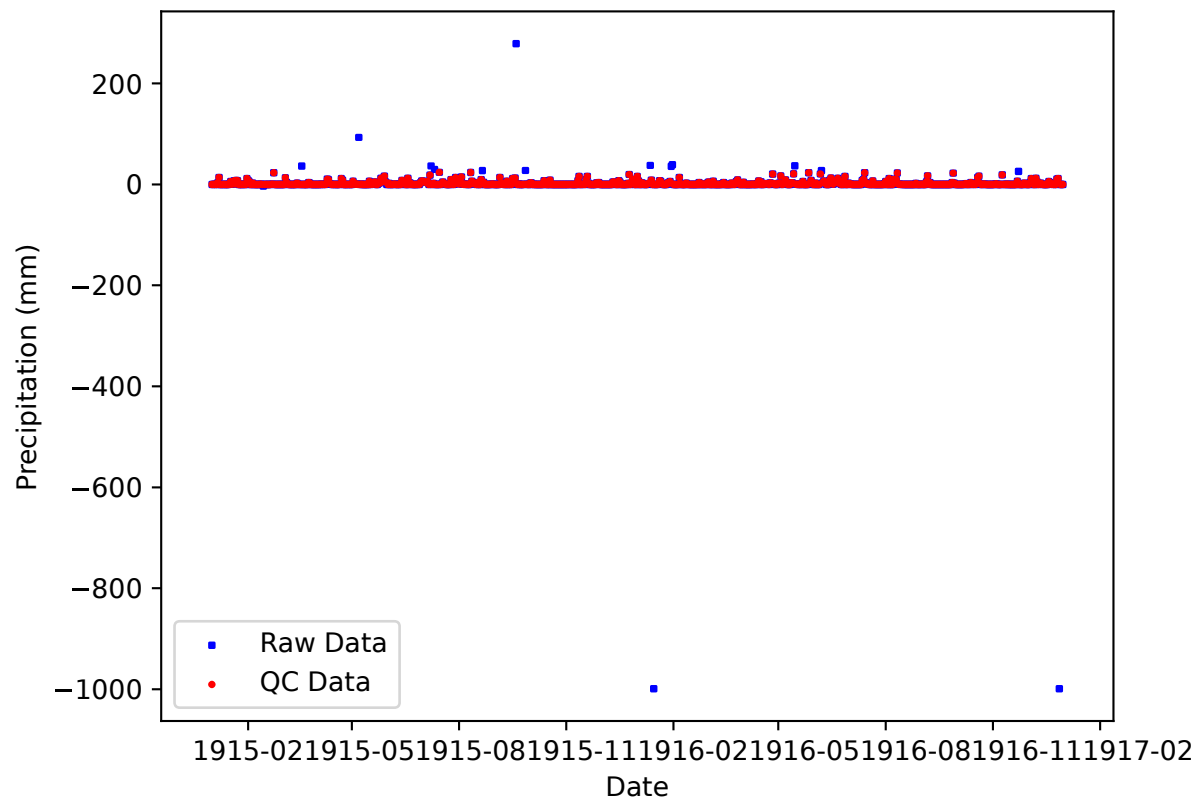


Figure 1: Precipitation scatter plot showing differences in the dataset before and after quality checking. Two -999 values and several gross error values were removed (visible in blue).

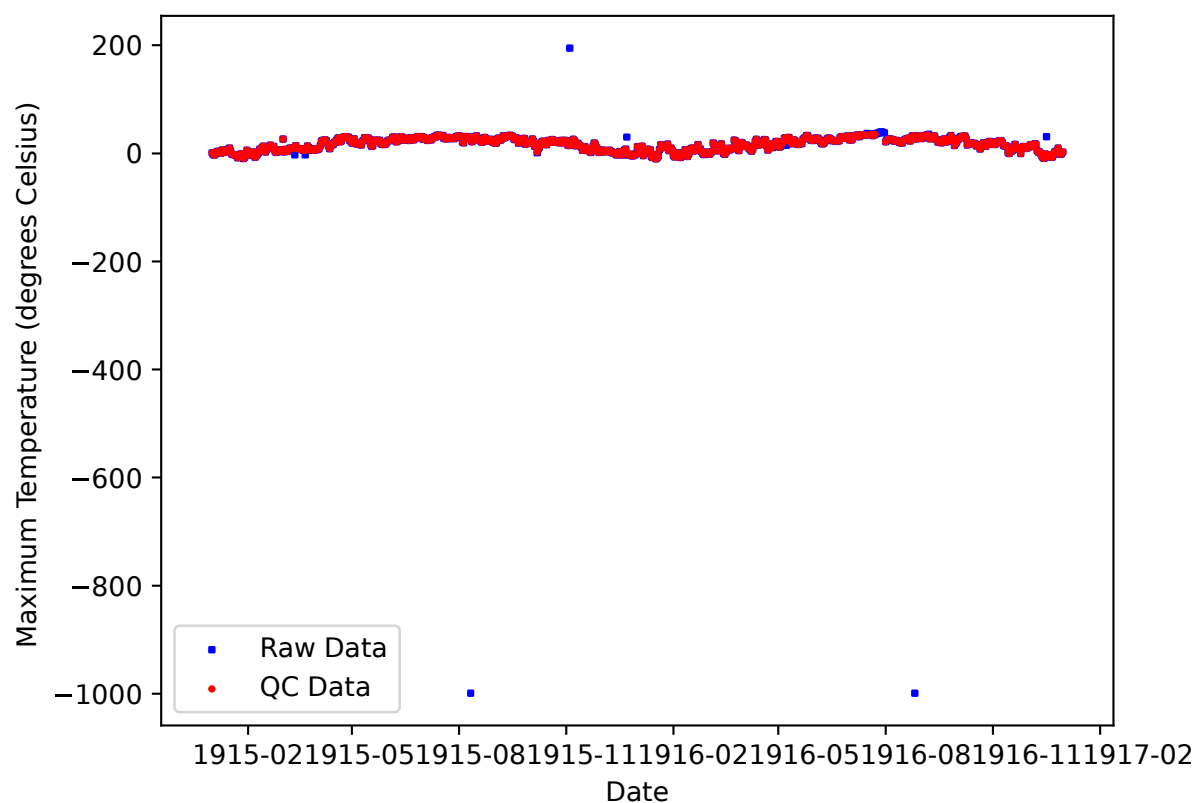


Figure 2: Maximum Temperature scatter plot showing differences in the dataset before and after quality checking. Two -999 values and several gross error values were removed (visible in blue). Visible blue dots near the red trend are values that were swapped with minimum temperature and/or failed the range check.

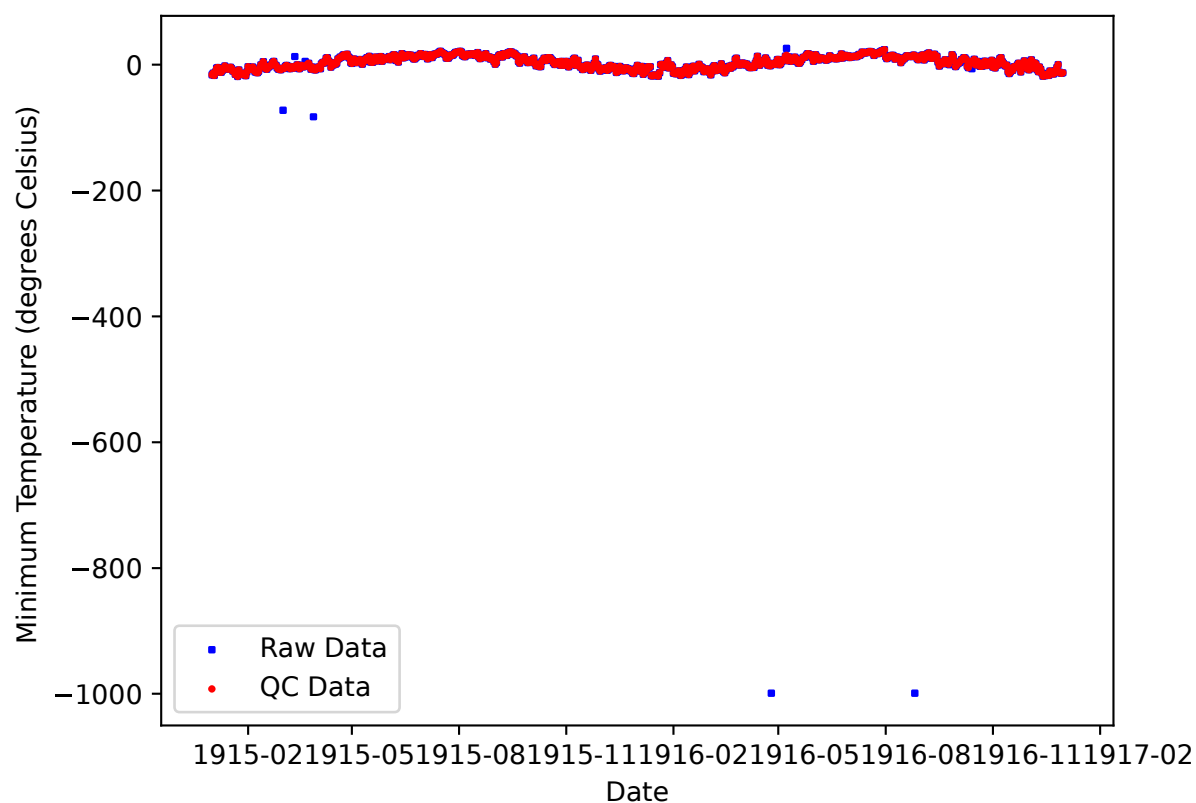


Figure 3: Minimum Temperature scatter plot showing differences in the dataset before and after quality checking. Two -999 values and several gross error values were removed (visible in blue). Visible blue dots near the red trend are values that were swapped with minimum temperature and/or failed the range check.

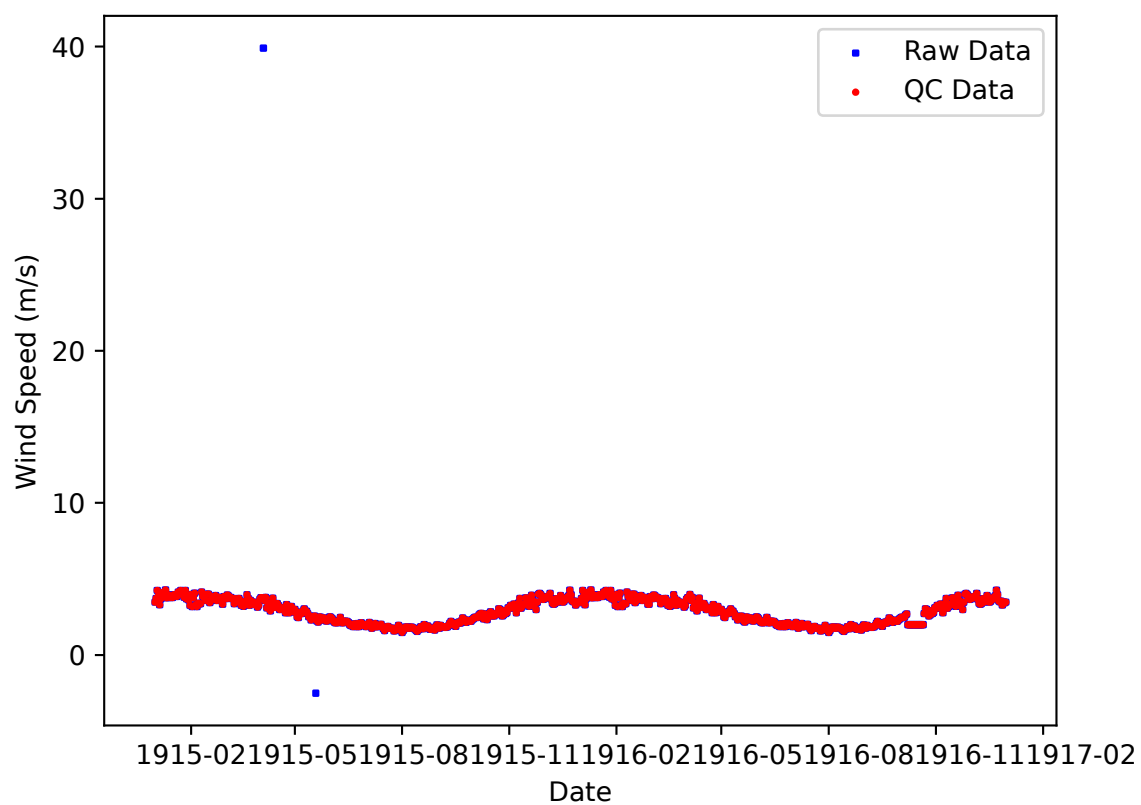


Figure 4: Wind speed scatter plot showing differences in the dataset before and after quality checking. There were no -999 values for this variable. A few gross error values were removed (visible in blue).