

Program: program_09.py
Program creator: Hannah Walcek

This program uses DataQualityChecking.txt as input and goes through 4 data quality checks: removing no data values, checking for gross errors, swapping max temp and min temp when max temp is less than min temp, and checking for daily temperature range exceedance. It also produces plots of precipitation, max temp, min temp and wind speed before and after cleaning.

Check 1: No Data

This check replaces all values of -999 with NaN. This drastically changes the scale of the graph by removing such a large number. Table 1 shows there are 2 replacements per variable except for wind speed. These extreme values can be shown in figures 1, 2, and 3.

Check 2: Gross Error

This check uses certain parameters for each variable and ensures values stay within these parameters. Any values outside become NaN. Table 1 shows quite a few gross errors for precipitation and maximum temperature and 2 for both minimum temperature and wind speed. Some of these gross errors can be easily seen in figures 1 and 2 where there are values around 200 and in figure 4 where there is a value of 40.

Check 3: Swapped

This check looks for any max temperature value that is less than the min temperature value and swaps them. This ensures consistency for temperature. As shown in Table 1, this only effects the temperature columns and makes 4 swaps. These are harder to spot in the provided figures.

Check 4: Range Fail

This check removes any temperature values where the max and min range exceeds 25 degrees Celsius. This temperature difference is nearly impossible so it is important to remove and replace with NaN. Table 1 shows the range fail only effects the temperature columns and there are 5 dates where this check fails. These are harder to spot in the provided figures.

Table 1: Table of checks for DataQualityChecking.txt

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

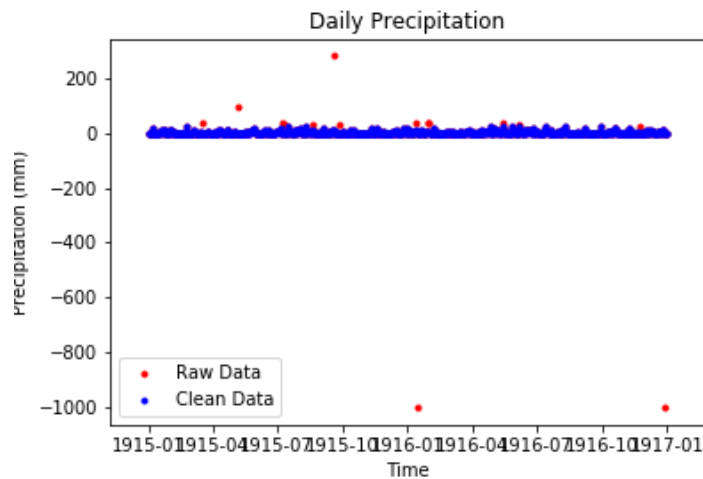


Figure 1: Daily precipitation with raw data and clean data shown. The scale is completely skewed due to outliers from the raw data, especially the null -999 values. There is also an extreme value over 200 mm.

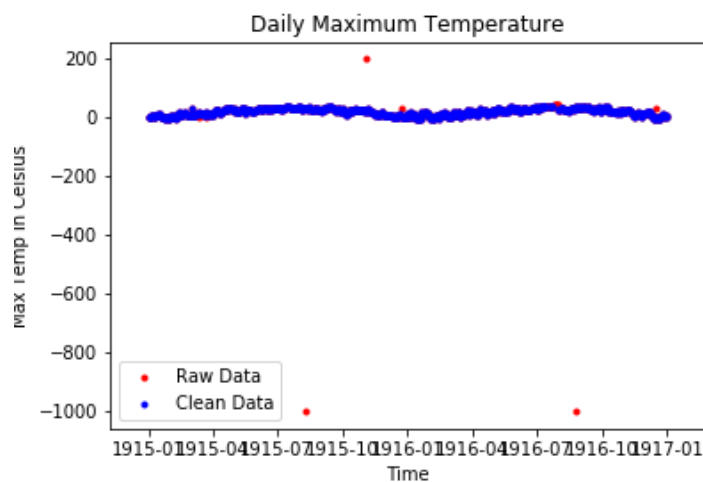


Figure 2: Daily maximum temperature with raw data and clean data shown. The scale is completely skewed due to outliers from the raw data, especially the null -999 values. There is also an extreme value over 200 degrees celsius.

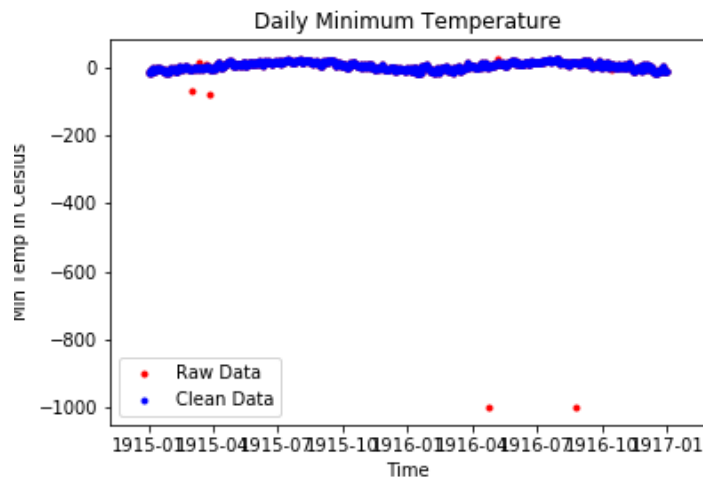


Figure 3: Daily minimum temperature with raw data and clean data shown. The scale is completely skewed due to outliers from the raw data, especially the null -999 values. There are also some low values out of typical range.

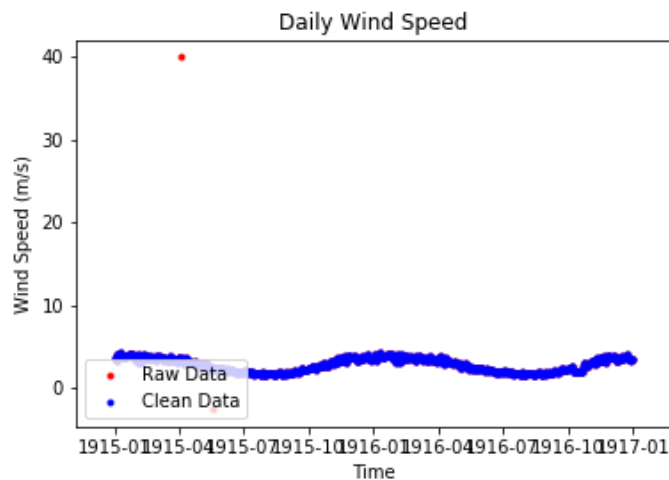


Figure 4: Daily wind speed with raw data and clean data shown. The scale is completely skewed due to a large outlier at 40 m/s.