Metadata

Modifier: Linji Wang
Original Author: Keith Cherkauer
Date of Modification: 04/03/2020
Filename: porgram_09.py
File Type: Python 3 script
Purpose: This python script check the quality of the dataset "DataQualityChecking.txt". Specifically removes no data values, identifies gross errors, inconsistencies in variables, and range errors. Also, creates output for the data quality check including corrected data, and plots.

Input Data: "DataQualityChecking.txt"
Format: .txt space separated values
Descriptions:
This file contains precipitation, max temperature, minimum temperature, and wind speed data from 01/01/1915 to 12/31/1916 with a daily time step. No data values are denoted as -999.

Scrip Structure Note:
This python scrip performs the following data quality checks:

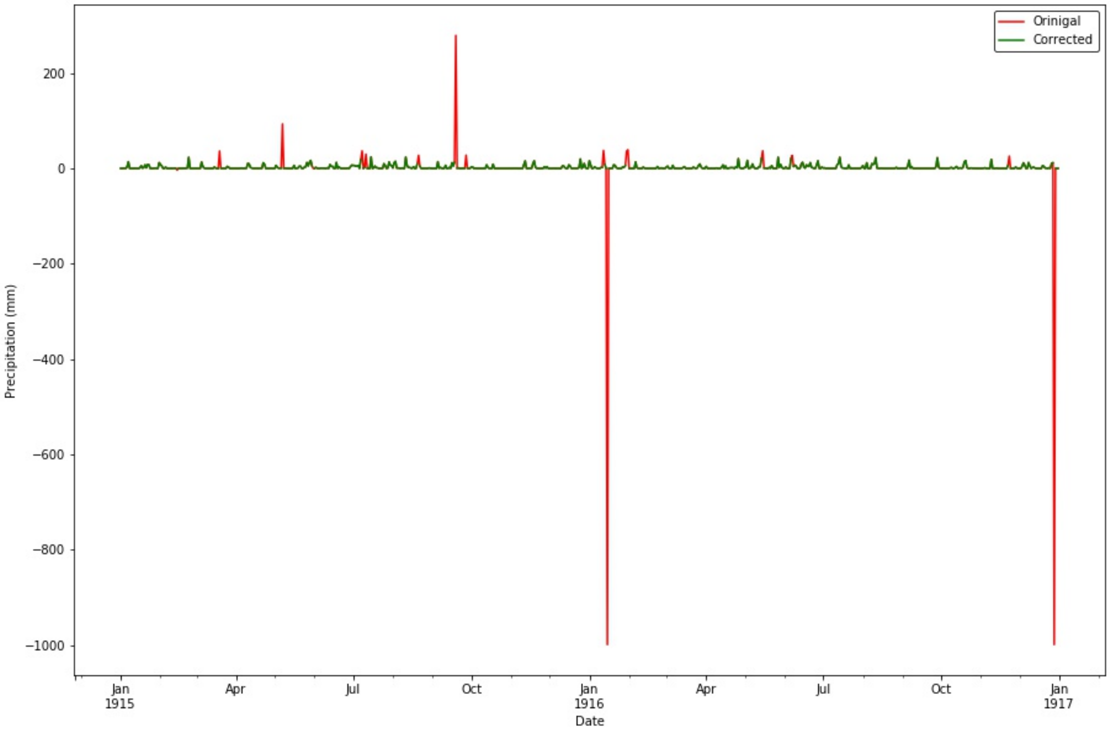
- Check 1: Removes No Data values.
  - Replace all values of -999 in this file with the NumPy NaN values.
  - Record the number of values replaced for each data type in the dataframe ReplacedValuesDF with the index "1. No Data"
- Check 2: Check for gross errors
  - Identify all values outside the thresholds 0 <= P <= 25; -25 <= T <= 35, 0 <= WS <= 10 and replace with NaN
  - Record the number of values replaced for each data type in the dataframe ReplacedValuesDF with the index "2. Gross Error"
- Check 3: Swap Max Temp and Min Temp when Max Temp is less than Min Temp.
  - Check that all values of Max Temp are greater then for Min Temp for the current day's observations.
  - Where they are not, swap the values.
  - Record the number of values replaced for each data type in the dataframe ReplacedValuesDF with the index "3. Swapped"
- Check 4: Check for daily temperature range exceedence.
  - Identify days with temperature range (Max Temp minus Min Temp) greater than 25°C.
  - When range is exceeded replace both Tmax and Tmin with NaN.
  - Record the number of values replaced for each data type in the dataframe ReplacedValuesDF with the index "4. Range Fail"

After the data quality checks, this script create outputs for the corrected datafram, a summary of data quality check result and plots for comparing data before and after quality checks.

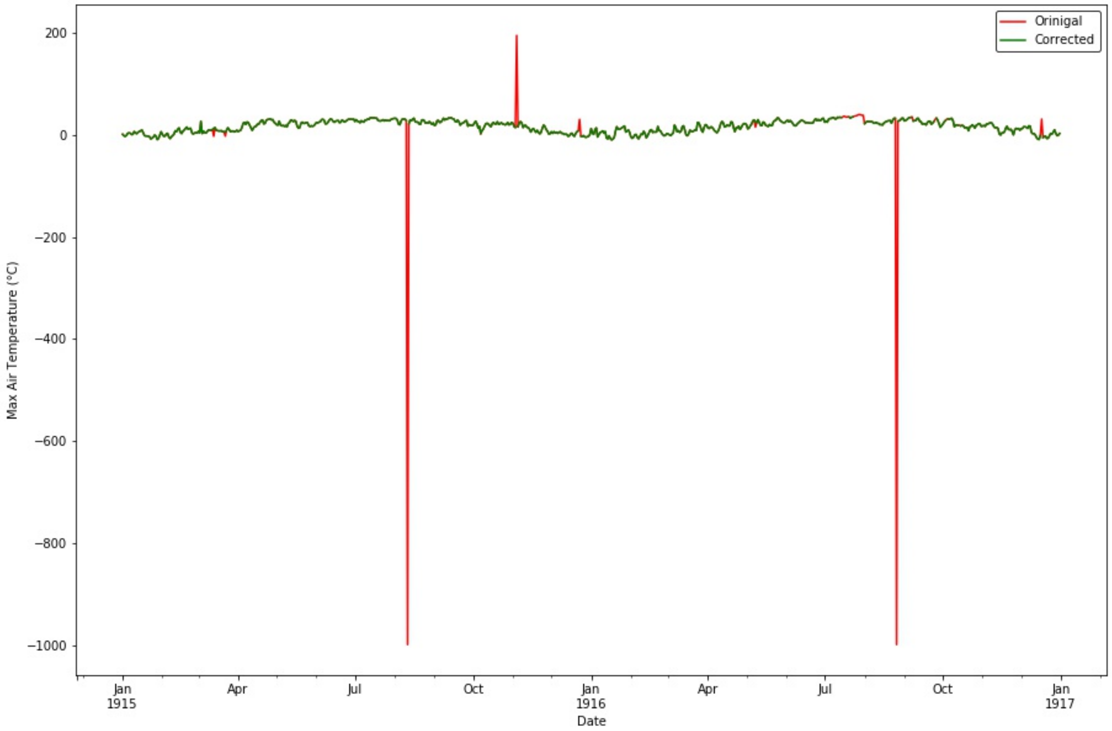
Summary of Data Quality Check:

	Precip	Max Temp	Min Temp	Wind Speed
1. No Data	0.0	0.0	0.0	0.0
2. Gross Error	15.0	14.0	2.0	2.0
3. Swapped	0.0	4.0	4.0	0.0
4. Range Fail	0.0	5.0	5.0	0.0

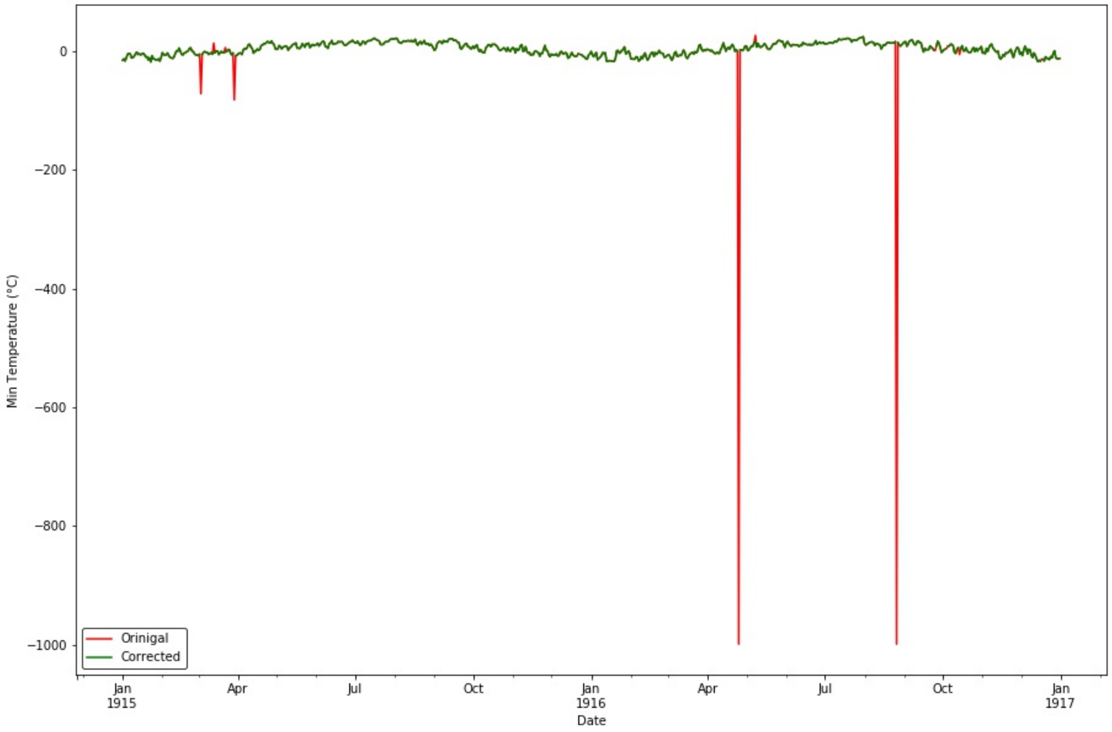
Precipitation Data Quality Check Result:



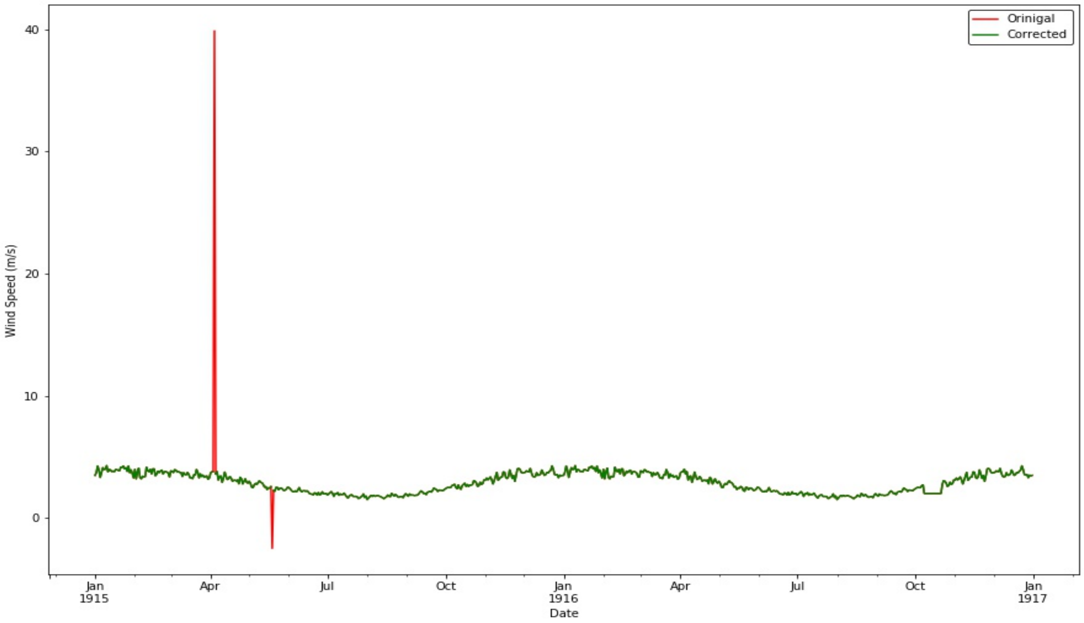
Max Air Temperature Data Quality Check Result Plot:



Min Air Temperature Data Quality Check Result Plot:



Wind Speed Data Quality Check Result Plot:


Orinigal
Corrected
Wind Speed (m/s)
Jan
Apr
Jul
Oct
Jan
Apr
Jul
Oct
Jan
Date