By the end of this activity, you will be able to:

1. Identify the major components in semi-structured data from a weather station
2. Create plots of weather station data

**Step 1.** **Open a terminal shell.** Open a terminal shell by clicking on the square black box on the top left of the screen.



Change into the sensor directory:

1

cd Downloads/big-data-2/sensor

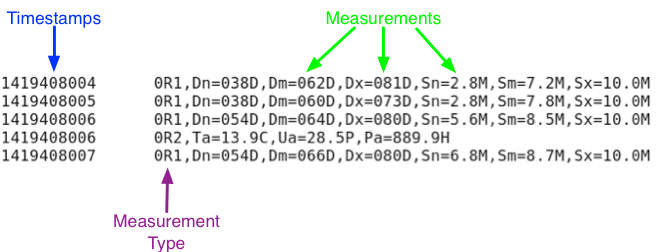
Run ls to see the sensor data files and scripts:

https://d3c33hcgiwev3.cloudfront.net/imageAssetProxy.v1/C787xQaTEeahzgq54E2myw_f98e28ea2f30948dcecce4981268b8ad_ls.png?expiry=1574640000000&hmac=Pa1J-FJtgAI0NP1b16BdPovNJqDKfl9XCy9zKn9A4rE

**Step 2. View weather station data.** Twenty-four hours of weather station data is in the file *wx-data.txt*. View this data by running *more wx-data.txt*:

1

more wx-data.txt



Each line in the file contains a timestamp in the first column and a set of measurements in the second column. These column are separated by a tab. The measurements are also a set of columns separated by commas.

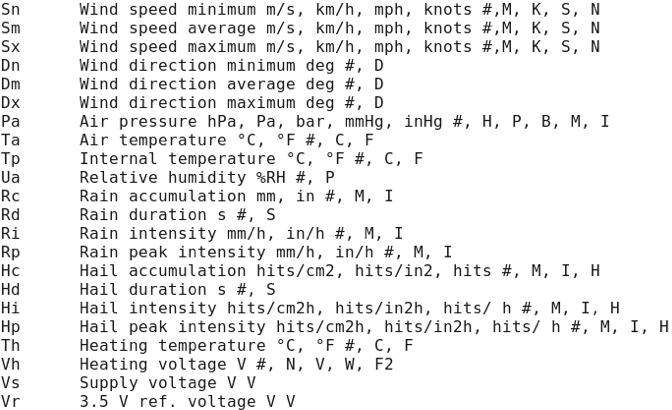
The timestamp is the number of seconds since January 1, 1970. By looking at the successive timestamps in the file, we can see that the measurements arrive about every second.

The measurements are prefixed with *R0, R1, R2*, etc. Each R-value holds a different set of measurements. For example, *R1* has *Dn, Dm, Dx*, etc. Additionally, the different types are measured at different frequencies. R1 is measured every second, but R0 and R2 are much less frequent.

**Step 3. Look at key for measurements.** We can learn what the measurement fields mean by looking at the measurements key file. Open a new terminal and *cd* to *Downloads/big-data-2/sensor*. Run more *wxt-520-format.txt*:

1

more wxt-520-format.txt

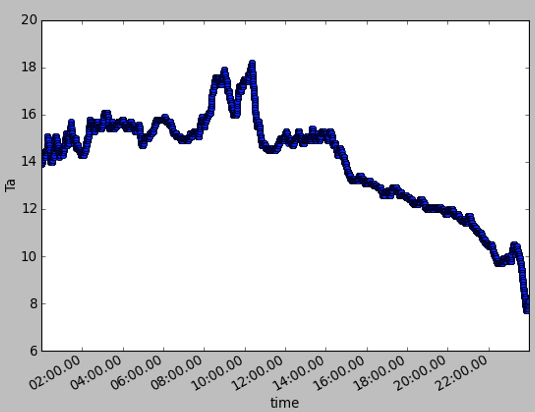


This file provides a list of what each field means. For example, *Ta* is the Air temperature, and *Dn* is the Wind direction minimum.

**Step 4. Plot data.** Run *plot-data.py wx-data.txt Ta* to plot the air temperature:

1

./plot-data.py wx-data.txt Ta



**Note:**If you see the following error**:**ImportError: No module named matplotlib.pyplot, your python environment is being incorrectly pointed to. To find the Math Plot Library, edit the first line to

1

#!/usr/bin/env python

You can plot other measurements in the file by using the field name as the last argument. For example, use Dn to plot the maximum wind direction:

1

./plot-data.py wx-data.txt Dn