

Data Format

The data we will use is from the National Climatic Data Center (NCDC, <http://www.ncdc.noaa.gov/>). The data is stored using a line-oriented ASCII format, in which each line is a record. The format supports a rich set of meteorological elements, many of which are optional or with variable data lengths. For simplicity, we focus on the basic elements, such as temperature, which are always present and are of fixed width. Fig. 1

```
0057
332130  # USAF weather station identifier
99999   # WBAN weather station identifier
19500101 # observation date
0300    # observation time
4
+51317  # latitude (degrees x 1000)
+028783 # longitude (degrees x 1000)
FM-12
+0171   # elevation (meters)
99999
V020
320     # wind direction (degrees)
1       # quality code
N
0072
1
00450   # sky ceiling height (meters)
1       # quality code
C
N
010000  # visibility distance (meters)
1       # quality code
N
9
-0128   # air temperature (degrees Celsius x 10)
1       # quality code
-0139   # dew point temperature (degrees Celsius x 10)
1       # quality code
10268   # atmospheric pressure (hectopascals x 10)
1       # quality code
```

Figure 1: Input data format

shows a sample line with some of the salient fields highlighted. The line has been split into multiple lines to show each field; in the real file, fields are packed into one line with no delimiters.

```
(0, 0067011990999991950051507004...9999999N9+00001+9999999999...)
(106, 0043011990999991950051512004...9999999N9+00221+9999999999...)
(212, 0043011990999991950051518004...9999999N9-00111+9999999999...)
(318, 0043012650999991949032412004...0500001N9+01111+9999999999...)
(424, 0043012650999991949032418004...0500001N9+00781+9999999999...)
```

Figure 2: Sample input data

To visualize the way the map works, consider the sample lines of input data as shown in Fig.2. (some unused columns have been dropped to fit the page, indicated by ellipses): These lines are presented to the map function as the key-value pairs: