

# Obtaining Climate Records

Marc Los Huertos

September 12, 2018

## 1 Terrestrial Meteorological Data

### 1.1 Selected History of Meteorology

The Korean's are first to develop the first way of recording climate data—they invited the first standardized range gauges, and even more impressively, a coordinated network of rain gauges and records to assist in rice planting with goals to improve food security.

In 1607, Galileo Galilei constructed a thermoscope, but it wasn't until 1654, that Ferdinando II de Medici established the first weather observing network, consisting of meteorological stations in Florence, Cutigliano, Vallombrosa, Bologna, Parma, Milan, Innsbruck, Osnabrck, Paris and Warsaw. The collected data were sent to Florence at regular time intervals. In 1714, Gabriel Fahrenheit created a reliable scale for measuring temperature with a mercury-type thermometer. In 1742, Anders Celsius, a Swedish astronomer, proposed the “centigrade” temperature scale, the predecessor of the current Celsius scale.

The arrival of the electrical telegraph in 1837 afforded, for the first time, a practical method for quickly gathering surface weather observations from a wide area. This data could be used to produce maps of the state of the atmosphere for a region near the Earth's surface and to study how these states evolved through time.

To make frequent weather forecasts based on these data required a reliable network of observations, but it was not until 1849 that the Smithsonian Institution began to establish an observation network across the United States under the leadership of Joseph Henry. Similar observation networks were established in Europe at this time.

### 1.2 Purpose

This document is a guide for getting started on the climate data analysis project. Here you will find:

- different sources for climate data
- a description of how to download/access the data

### 1.3 Climate Data Online (CDO) system

Climate Data Online (CDO) provides free access to the National Climate Data Center's archive of global historical weather and climate data in addition to station history information. These data include quality controlled daily, monthly, seasonal, and yearly measurements of temperature, precipitation, wind, and degree days as well as radar data and 30-year Climate Normals.

There are several ways to obtain online climate data and the CDO-web interface is probably the easiest, using the following URL: <https://www.ncdc.noaa.gov/cdo-web/>.

Climate data, such as temperature and precipitation are recorded at each station – so these are the observational units we will generally be interested in for our analyses. Even if you don't have a location of interest nailed down, this interface can help you find sites and evaluate the length of their records. For the purpose of our project we recommend picking a site that has data going back at least into the the 1940's as it might be difficult to come up with insightful observations about sites with less data.

Below is a description of the steps to get data from the site:

**Follow Link** Click on the following link to reach the CDO website:

<https://www.ncdc.noaa.gov/cdo-web/>

**Click Browse Datasets** after clicking the icon you should see a list of available data sets by type

**Expand Link** Daily Summaries

**Select the Search Tool** additionally, you can use the mapping tool to evaluate which areas of the globe have available datasets but if you have a particular area in mind it might be easiest to search for it using the search tool

**Specify Parameters**

- ensure that “Daily Summaries” is selected from the dropdown menu in **Select Weather Observation Type/Dataset**
- in **Select Data Range** use the calendar tool change the start date to earliest date available (It should be January 1, 1763) and the end date to the most recently available date. This will enable you to see how far back the data record for the station goes.
- in the drop down menu **Search For** select whether you want to search for climate data by country, state, county, zip code etc. Select the appropriate option for the search term that you will enter in the box below.

For example if you want to search for climate data in your hometown, you might consider searching by state or county and then typing in the name below

- in the **Enter a Search Term** box, type in the name of the county, state or other possible identifier of the place for which you would like to find a climate record.

**Start Search** Once you have initiated a search, the next window will give a list of records – click in these records to see what they might be composed of. Then decide which station you will want to evaluate.

**Add to Cart** Don't worry, there is no charge.

**View All items** Once you have added the climate record to your cart you should see an orange box pop up on the top right corner that says “Cart (Free Data) 1 item.” Mouse over the box after which you should have the option to click “view all items”.

**Select Cart Options** in **Select the Output Format** be sure to click the button for **Custom GHCN-Daily CSV** to ensure that the data file is delivered as a .csv file (comma separated values)

**Click ‘Continue’ Button**

**Station Detail & Data Flag Options** Change to metric

**Select Data Types** be sure to check the boxes for precipitation and air temperature to include these in the data sets. If fact, I suggest you click on the + button and select maximum and minimum temperatures, which usually have the longest records.

**Click ‘Continue’ Button** NOTE: Common Errors: Error – \*Text order size is 3,948 Station Years, which exceeds our capacity of 1,000 Station Years. Please select fewer stations/locations, or reduce the date range.

If you get this error, you need to reduce the number of stations selected or make the date range smaller. NOTE: it's not always obvious that you have selected more than one station. When you are looking at the cart, follow the "View Full Details" link and then "Show all Stations". I generally sort by earliest records and find a station that is still collecting data and has nearly 100% data coverage.

**Complete Order** Type in your email address and you will get a confirmation email relativley soon... the data may take up to a couple days to be delivered but are usually delivered within a couple hours.

### 1.3.1 Evaluating the Data

Once you get the data, we'll need to evaluate what we have and format it for our purposes. Often you'll find problems when you begin analyzing it, but it's good to understand the structure of the data before you run into problems.

The CHCND documentation helps us to understand the data format, [https://www1.ncdc.noaa.gov/pub/data/cdo/documentation/GHCND\\_documentation.pdf](https://www1.ncdc.noaa.gov/pub/data/cdo/documentation/GHCND_documentation.pdf).

## 1.4 Historical Climatology Network-Daily (GHCND) data

Alternatively, one can use the FTP site to access the data in a different format (e.g. dyl). The site's url is <ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/gsn/>.

In addition, the CHCND documentation, as noted above, is very helpful in understanding the data format, [https://www1.ncdc.noaa.gov/pub/data/cdo/documentation/GHCND\\_documentation.pdf](https://www1.ncdc.noaa.gov/pub/data/cdo/documentation/GHCND_documentation.pdf).

## 1.5 Advanced Users: rnoaa library tools

There is a relatively new tool to access NOAA data using an R package, called **rnoaa**. To use the tool, several dependent packages need to be installed and loaded, including **rNOAA**. Next users need to get an API code, which can be easily done by following the instructions described by this link: <https://www.ncdc.noaa.gov/cdo-web/token>.

Using the **rnoaa** package is relatively strait forward, although I have been getting timeout errors that I can't solve.

Here's a URL for a tutorial: [https://ropensci.org/tutorials/rnoaa\\_tutorial.html](https://ropensci.org/tutorials/rnoaa_tutorial.html).

```
library(plyr)
library(rnoaa)
```

### 1.5.1 List of Cities

rNOAA has a simple function to list the cities:

```
ncdc_locs(locationcategoryid='CITY', sortfield='name', sortorder='desc')

## $meta
## $meta$totalCount
## [1] 1987
##
## $meta$pageCount
## [1] 25
##
## $meta$offset
## [1] 1
##
##
## $data
##      mindate   maxdate      name datacoverage      id
## 1  1892-08-01 2018-07-31   Zwolle, NL         1.0000 CITY:NL000012
## 2  1901-01-01 2018-09-08   Zurich, SZ         1.0000 CITY:SZ000007
## 3  1957-07-01 2018-09-08 Zonguldak, TU         1.0000 CITY:TU000057
```

```
## 4 1906-01-01 2018-09-08 Zinder, NG 0.9025 CITY:NG000004
## 5 1973-01-01 2018-09-08 Ziguinchor, SG 1.0000 CITY:SG000004
## 6 1938-01-01 2018-09-08 Zhytomyra, UP 0.9723 CITY:UP000025
## 7 1948-03-01 2018-09-08 Zhezkazgan, KZ 0.9302 CITY:KZ000017
## 8 1951-01-01 2018-09-08 Zhengzhou, CH 1.0000 CITY:CH000045
## 9 1941-01-01 2018-06-30 Zaragoza, SP 1.0000 CITY:SP000021
## 10 1936-01-01 2009-06-17 Zaporijhzhya, UP 1.0000 CITY:UP000024
## 11 1957-01-01 2018-09-08 Zanzibar, TZ 0.8016 CITY:TZ000019
## 12 1973-01-01 2018-09-08 Zanzjan, IR 0.9105 CITY:IR000020
## 13 1893-01-01 2018-09-09 Zanesville, OH US 1.0000 CITY:US390029
## 14 1912-01-01 2017-06-19 Zahle, LE 0.9819 CITY:LE000004
## 15 1951-01-01 2018-09-08 Zahedan, IR 0.9975 CITY:IR000019
## 16 1860-12-01 2018-09-08 Zagreb, HR 1.0000 CITY:HR000002
## 17 1975-08-29 2018-09-08 Zacatecas, MX 0.9306 CITY:MX000036
## 18 1947-01-01 2018-09-08 Yuzhno-Sakhalinsk, RS 1.0000 CITY:RS000081
## 19 1893-01-01 2018-09-10 Yuma, AZ US 1.0000 CITY:US040015
## 20 1942-02-01 2018-09-10 Yucca Valley, CA US 1.0000 CITY:US060048
## 21 1885-01-01 2018-09-09 Yuba City, CA US 1.0000 CITY:US060047
## 22 1998-02-01 2018-09-08 Yozgat, TU 1.0000 CITY:TU000056
## 23 1893-01-01 2018-09-10 Youngstown, OH US 1.0000 CITY:US390028
## 24 1894-01-01 2018-09-10 York, PA US 1.0000 CITY:US420024
## 25 1869-01-01 2018-09-10 Yonkers, NY US 1.0000 CITY:US360031
##
## attr(,"class")
## [1] "ncdc_locs"
```

The function queries the NOAA website and retrieves city names and the dates of the climate records. Importantly, the records include the station ID, which can be used to college the data for that city.

NOTE2: It would be nice to make a map of how concentrated the stations spatially.

## 1.5.2 Getting Data

```
ncdc_stations(datasetid='GHCND', locationid='FIPS:12017', stationid='GHCND:USC00084289')

## $meta
## NULL
##
## $data
##   elevation   mindate   maxdate latitude      name
## 1      12.2 1899-02-01 2018-09-10  28.8029 INVERNESS 3 SE, FL US
##   datacoverage      id elevationUnit longitude
## 1           1  GHCND:USC00084289      METERS   -82.3126
##
```

```
## attr("class")  
## [1] "ncdc_stations"
```

## **1.6 Evaluating Records**

TBD

## **1.7 Export Options**

TBD

## **2 Sea Surface Temperature Data**

TBD

## **3 Satellite Data**

TBD

## **4 Ice-Core Data**

TBD