Guide1 : Obtaining State Weather Station Data

Marc Los Huertos

February 2, 2024 (ver. 0.8)

1 Introduction

1.1 Goals

Using a list of active weather stations in the United States, you will download the data from five stations and read these into the R environment.

1.2 Selected History of Climate Science

Geologists have known the climate has been changing over the Earth's history. But what causes these changes has been a major research area for over 100 years. There are numerous drivers that contribute to changing climates – including the arrangement of the continents on the planet, the distance to the sun, energy generated by the sun, volanic activity, and the composition of the Earth's atmosphere.

It's the last one that we'll spend time because the Earth's temperature are changing pretty dramatically over the last 100 years and the cause is no mystery – the human activity that has released carbon dioxide (CO_2) into the atmosphere. The two main sources of CO_2 is from land use change, e.g. deforestration, and the burning of fossil fuels, e.g. coal, oil, and natural gas.

The first person to propose the role of CO_2 on the Earth's atmosphere was a Swedish scientist Svante Arrhenius, who figured out that CO_2 absorbs infarred light. Moreover, he deduced that the Earth's temperature was actually warmer than it might otherwise be if CO_2 was not part of the Earth's atmosphere.

1.3 Why Look at Individual Stations?

I don't think there is a single, perfect way to analyze and communicate climate change. But the beauty of the network of stations in the USA and around the world is that these stations record weather as expecienced by local people. And while indiviudual stations may not represent the overall regional and global patterns well, this give us a mechanism to connect local experiences to regional or global processes.

Of course, some may fixate on the local pattern and remain unconvinced of the larger context and for those folks, there may be better ways to communicated climate data.

However, I would be remiss in failing to mention that some may fixate on local patterns and use these patterns to ignore or to dimiss the patterns in other regions.

Finally, the impacts of climate change are highly specific to the region in question. Thus, once someone understands the impacts on climate change in their region, they my not be able to appreciate how differnet the climate impacts might affect other peoples, who maybe more vulneratble, around the globe.

Thus, with these weaknesses in mind, I will pursue this project with an eye to address these other issues at later stages.

1.4 Approach

1.4.1 NOAA Data Records

The US National Oceanic and Atmospheric Administration (NOAA) maintains several sources of digital weather data from the USA and beyond. These data have been collected from stations around the country to support a wide range of human activities that include farming, aviation, shipping, and even armed conflict.

At various times, these records have been used to evaluate long-term climate change with varying success. Without a doubt, these data are not perfect, but they remain that foundation of an effective adn professionally maintained environmental monitoring program that engenders integrity, even when facing budget cuts.

I will use these data to select for a station with a long record for each state in the USA. Future projects might evaluate the record for stations around the world, but we will see about that.

1.4.2 R Programming Language

R is an open source programming environment that has become one of the most popular tools for statiticians and data scientists. Capitalizing on the open source framework, a wide range of libraries or packages have been developed to faciliate data processing, analysis, and graphical displays.

For the project we'll use a few packages that need to be installed and loaded. Using the lower right panel, you can install the packages by clicking on the "Packages" tab and then "Install" and then type in the package name and click "Install".

1.5 Revised Documentat Structure

Based on some feedback in the last week, I have re-orgnaized the document with bullets of the steps and then the code. In the following sections, I'll descriptions

about "what the code does and why". But several folks see the text and get tuned out.

There are two weaknesses with this approach that I can see. First, if the code does not work, you will have a really hard time figuring out why, since you have no idea what the code is supposed to do. Second, if the code works, you will not really be learning any R coding. Just how to copy and paste.

I have created a bunch of functions, which can then be used to collect Let's try this and see how it works! Please download and run the code called Guide1.R, which has the functions for this handout.

2 R Functions for Getting Weather Station Data

The following functions are used to select the weather stations, download data, and read into R.

Subset Inventory Data with my.state This function requires two parameters to be set: filename and my.state.

Here's the function:

```
## function (filename, my.state)
## {
##     inventory.active.oldest <- read.csv(filename)
##     my.inventory = subset(inventory.active.oldest, STATE == my.state)
##     return(my.inventory)
## }
```

I suggest you use file.choose() and assign filename.csv to the path and name of the csv file.

Example of how to use the function:

```
my.state <- "CA"
filename.csv <- "/home/mwl04747/RTricks/04_Regional_Climate_Trends/stations.active.olde
my.inventory <- readInventory.fun(filename.csv, my.state)
```

Download Weather Station Data and Read into R This functions requires two parameters: data path and dataframe of my.inventory. The data path is a folder that you can always get your station data from if you need it. For example, if you need to update the data, you can just download the data from the NOAA website and then read it into R.

Your data path will be different, I created a directory to separate my data by class year, you don't need that!

```
## function (datafolder, my.inventory = my.inventory)
## {
##     for (i in 1:nrow(my.inventory)) {
##         url = paste0("https://www.ncei.noaa.gov/pub/data/ghcn/daily/by_station/",
##             my.inventory$ID[i], ".csv.gz")
##         download.file(url, paste0(datafolder, my.inventory$ID[i],
##             ".csv.gz"), quiet = FALSE, mode = "w", cacheOK = TRUE)
##         assign(paste0("station", i), read.csv(gzfile(paste0(datafolder,
##             my.inventory$ID[i], ".csv.gz")), header = FALSE))
##         print(paste("Index (Loop) ", i, " Completed."))
##         print("Think about something you are grateful for today!")
##     }
##     names(station1) <- c("ID", "DATE", "ELEMENT", "VALUE", "M-FLAG",
##         "Q-FLAG", "S-FLAG", "OBS-TIME")
##     names(station3) <- names(station2) <- names(station1)
##     names(station5) <- names(station4) <- names(station1)
## }
```

An example of how to use the function:

```
datapath = "/home/mwl04747/RTricks/04_Regional_Climate_Trends/Data/SP24"
downloadStations.fun(datapath, my.inventory)
```

3 Explaining the Code

3.1 Selecting Weather Records by State

There are numerous ways to analyze temperature records, where stations can be analyzed individually or records could be sampled and analyzed in spatially in grids. Each of these are valid approaches depending on the question to be addressed.

Here are the questions we will address:

- What stations have the longest meterological records in the USA?
- Can we determine the reliability of these stations?
- Finally, is there a temperature trend?

3.2 Defining Path & Read Data

First, we install some packages and read in the data. I suggest you create a folder for the project (I created one called "04_Regional_Climate_Trends") and then used the function here() to get the working directory and read the csv into R. This might be easier than the file.choose() option, but you can use that if you prefer.

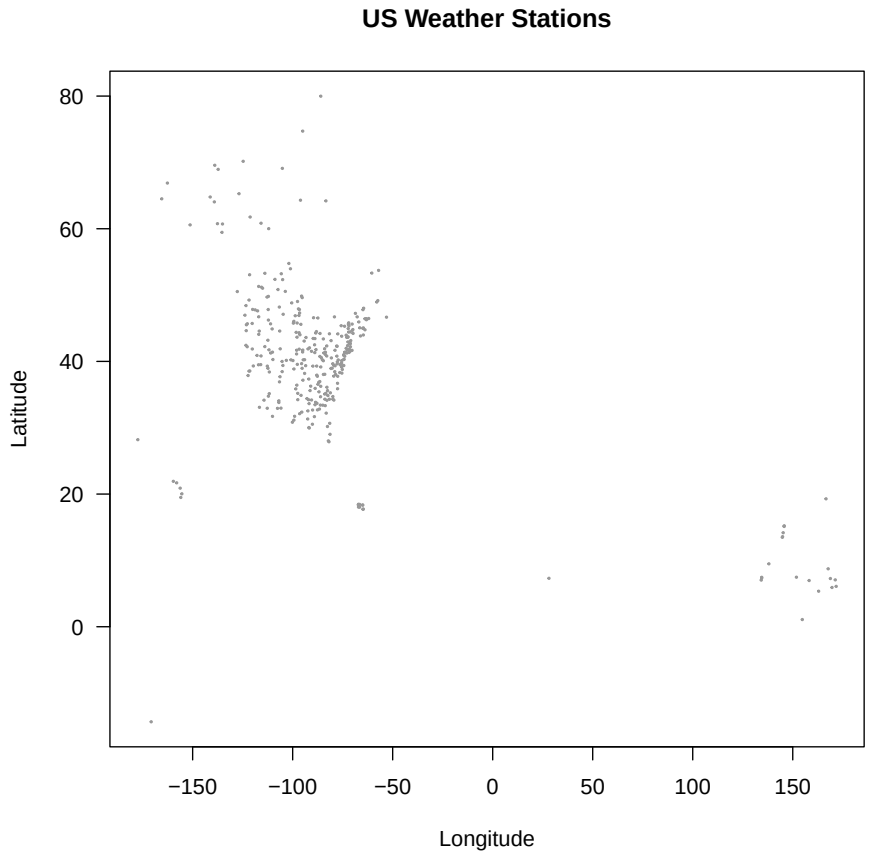
```
library(here)
library(xtable)

stations.active.oldest = read.csv(
  here("04_Regional_Climate_Trends", "stations.active.oldest.csv"))

# OR
# use file.choose() to select the file
# filename = "MY.PATH/04_Regional_Climate_Trends/stations.active.oldest.csv"
# stations.active.oldest = read.csv(filename)
```

3.3 Map US Weather Stations

Here's a map of the weather stations in the dataset. Pretty lame map! We'll make a better one later.[1]

```
plot(stations.active.oldest$LONGITUDE,
     stations.active.oldest$LATITUDE,
     xlab = "Longitude",ylab = "Latitude",
     pch=20, cex=0.3, col='gray60', las=1,
     main = "US Weather Stations")
```

[1] Evelyn/Brody: This is a good change to see how to use R for map making! First we need to transform that data.



3.4 Select and Evaluate State Data

```
stations.unique =
  unique(stations.active.oldest[,c("STATE", "STATE_NAME")])

xtab = xtable(stations.unique)
```

The each of you will select a state – see the Google Sheet sign up so we have a diverse set of states.

```
my.state = "CA" # change the "CA" to your state
```

4 Download Data from NOAA

4.1 Subset Station Data by State

This uses the stations.active.oldest file to download the data from the NOAA website based on the state you have choose.

```
my.stations = subset(stations.active.oldest, STATE == my.state)

# Download Updated Station Data
i=1
here::here("04_Regional_Climate_Trends", my.stations$ID[i])

## [1] "/home/mwl04747/RTricks/04_Regional_Climate_Trends/USC00043157"
```