How to Download and Update Air Quality Data

Authors: Monika Roznere and Andy Bean

ToC:

**USING SAQGETR TO OBTAIN CURRENT DATA**

The major procedure description is given in the code which extracts the initial data set:

####Procedure####

# First, identify a city location and search for a meteorological site there.

# If there is one, extract its data and refine aq sites using it's lat/long.

# Find the closest aq site possible and export that data also.

# Then pivot the aq data and left join it to the met data.

# Finally, eliminate unneccesary rows and export the data as a csv.

# If there is a met site in a city, it is acceptable to take multiple aq sites

# in that same city to make multiple data sets. Especially if those aq sites

# are looking at different parts of the city (background vs. traffic, etc.).

We will break this down piece by piece with code examples. But we will first need the setup:

# Call Libraries

library(openair); #vis pkg

library(tidyverse); #helps vis pkg

library(worldmet); #met pkg

library(dplyr); #data manip fxns

library(saqgetr); #euro pkg

# Set Working Directory

setwd("/Users/…etc.");

# data for all sites and sort down to Germany

sites\_dat <- get\_saq\_sites();

de\_sites <- filter(sites\_dat, country == 'germany');

We need openair for pivoting/appending, tidyverse and worldmet for met data, and dplyr and saqgetr for aq data.

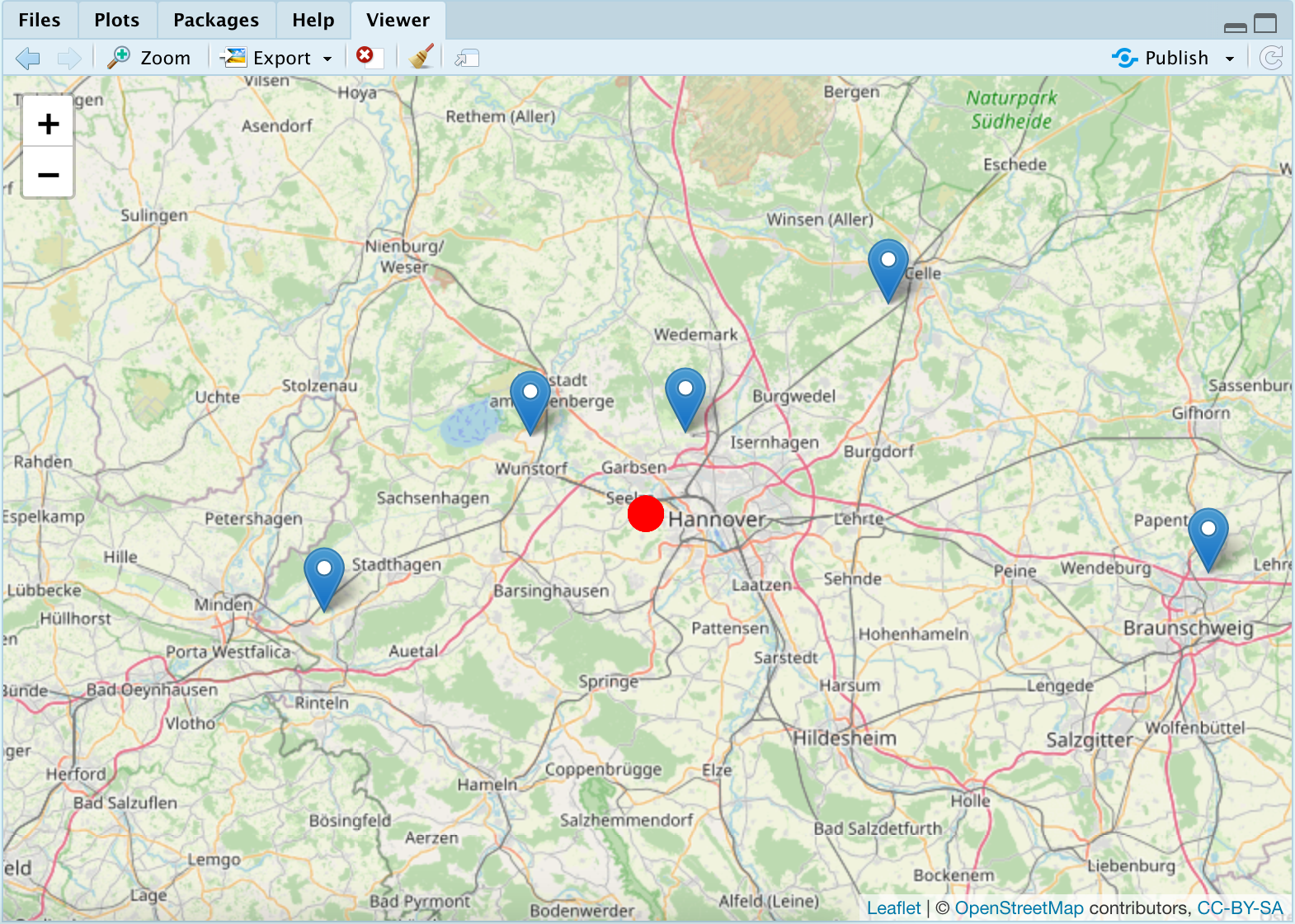
1. Identify a City and Search for Met Site

First, find the lat/long of a city you want to investigate (I did this through google earth). For this example, let’s take Hannover, Germany. Round the values and then run the getMeta command as follows:

# Look for met site in Hannover 52.3797932,9.6213879

getMeta(lat = 52.38, lon = 9.62, returnMap = TRUE);

This command will return an interactive map in the viewer window centered on the coordinates input (red dot) with available met sites highlighted with blue location icons. Once you find a site you like, click on the corresponding icon and copy it’s “code.” You’ll need this to download data. In the icon, you can also see info about how long the data has run for. These met sites are from a database stored by NOAA



1. Extract Data and Refine AQ Sites with Lat/Long

When you find the site you want met data from and have its code, now run the command importNOAA():

hannover\_met <- importNOAA(code = "103380-99999", year = 2014:2021);

cleaned\_h\_met <- select(hannover\_met, station:air\_temp);

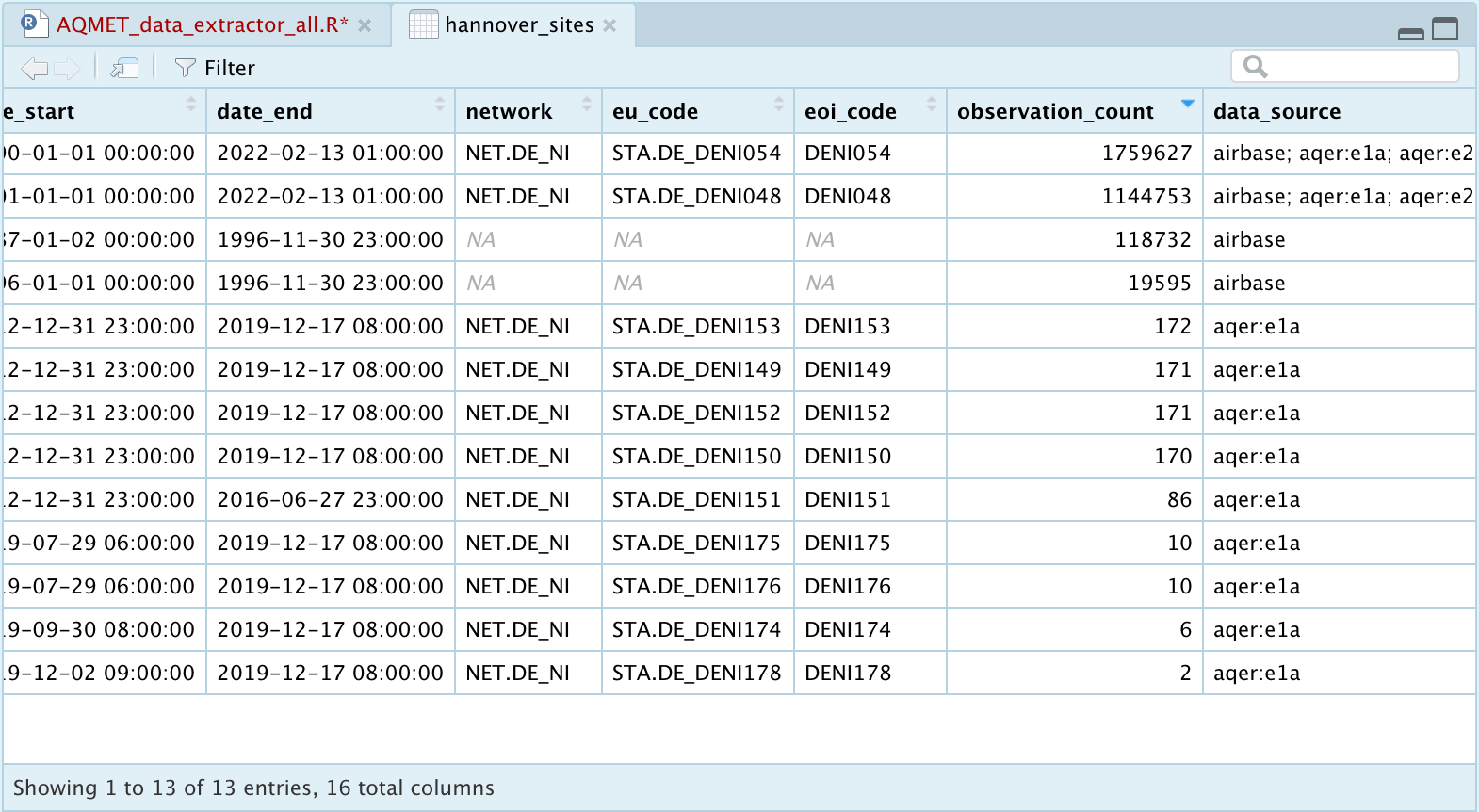
It is also recommended at this point that you clean down this met data by only taking the columns station through air\_temp, all others are not really helpful.

Next, once that data is saved. We need to refine our aq sites down to be closer to the met site (there are already ~2,000 aq sites in Germany which saqgetr has data for:

# filter all aq sites to this location

hannover\_sites <- filter(de\_sites, latitude > 52.3 & latitude < 52.7 & longitude > 9.5 & longitude < 9.9);

I tended to use a range of about .5 degrees around the lat and long of the met station (not the city, though this should be very similar and fine). Once this is done, open the new sites list you have created by clicking in the environment. Sort this list by number of data entries (highest on top) so you can see which sites actually have enough data for you to use.



1. Identify AQ Sites and extract their data

Once you have found a site you like, copy its code (farthest left column named “site”) for later use. Remember, a good site is close to the met site and has sufficient data for analysis. Furthermore, you might consider taking two sites very close to each other as long as both are still near the met site relatively and those two sites show something different (urban background vs. traffic for instance)

Now, run the command to extract the aq data:

# 1 take main background Hannover

hannover\_dat <- get\_saq\_observations(

site = "deni054",

start = 2014,

end = 2021,

verbose = TRUE

);

1. Pivot Data to Append and Export

The final step is putting the data in a format that openair can read.

First, we need to pivot the file into a format that can be joined. Right now, the aq data is logged by measurement type for each row, not measurement type by column over time. This can be done with some tidyverse formatting and the saq\_clean\_observations function. Then use left\_join to tack this onto the met data by date so that all data lines up by day and time. And finally, export the file with your desired name and format to your working directory

pivoted\_hannover <- hannover\_dat %>%

saq\_clean\_observations(summary = "hour", valid\_only = TRUE, spread = TRUE);

aq\_hannover\_main <- left\_join(cleaned\_h\_met, pivoted\_hannover, by = "date");

write.csv(aq\_hannover\_verk,"aq\_hannover\_verk.csv", row.names = TRUE);

**FUTURE DATA ACQUISITION**

How might we update this data or obtain something completely new if we chose?

1. Update by just changing years in the function calls

You will notice that most function calls simply ask for the years to download the data for. To update the data to be from 2015 - 2022, for instance, all one would have to do is change that in the years of the met and aq data.

However, note that changing dates and rerunning the entire file does carry some risk. Rerunning the whole file takes time, and more importantly, if a year’s data is missing for some reason (has happened for met sites before, from data that we \*were\* able to download two weeks prior), then the new file you produce will overwrite the old one, and you may no longer be able to see the data for a newly missing year. So please save backups in an extra location, or change the name as you export to not overwrite an old file.

1. Update by just extracting new and appending to bottom of existing

If perhaps an old year’s data is newly missing, but a new year’s is now available, you might consider just appending the new to existing data. If you have the sites of the old data set logged, simply run again to export the new year only (e.g. 2022).

Then, after that operation, you can download in your existing data, and append the new year to all of the previous data, and re-export the previous data.

1. Get a totally new site through the process above if in Germany

If you want a new site in Germany, then simply follow the process above with *saqgetr*.

* 1. If in UK?

If you want to add data from sites in you UK, you could actually use the saqgetr route again (because the saqgetr data is for European regions and UK is there). However, openair was designed by and for people in the UK, and so it has functions that will directly download several possible aq sites in a format that is already complete with the weather!

Try commands like “importAURN” and others available in the openair manual. A list of site codes should also be available there (or at least instructions on how to produce those codes).

* 1. If in USA?

If in USA, worldmet will still work to obtain met data, because NOAA data for meteorology already covers the US of course, in addition to its hundreds of locations outside of the States. However, new aq data will be needed. There are some sites that can be accessed through EPA portals such as: <https://www.epa.gov/outdoor-air-quality-data>

But, this may require a new API to download the data or another way to download and import it at reasonable speed in a format that can be combined with the met data still. This is the part that would require an additional investment of time. Alternatively, saqgetr or openair documentation (github for both) may have recommendations about how to obtain aq data for US sites (or other libraries that will obtain those)