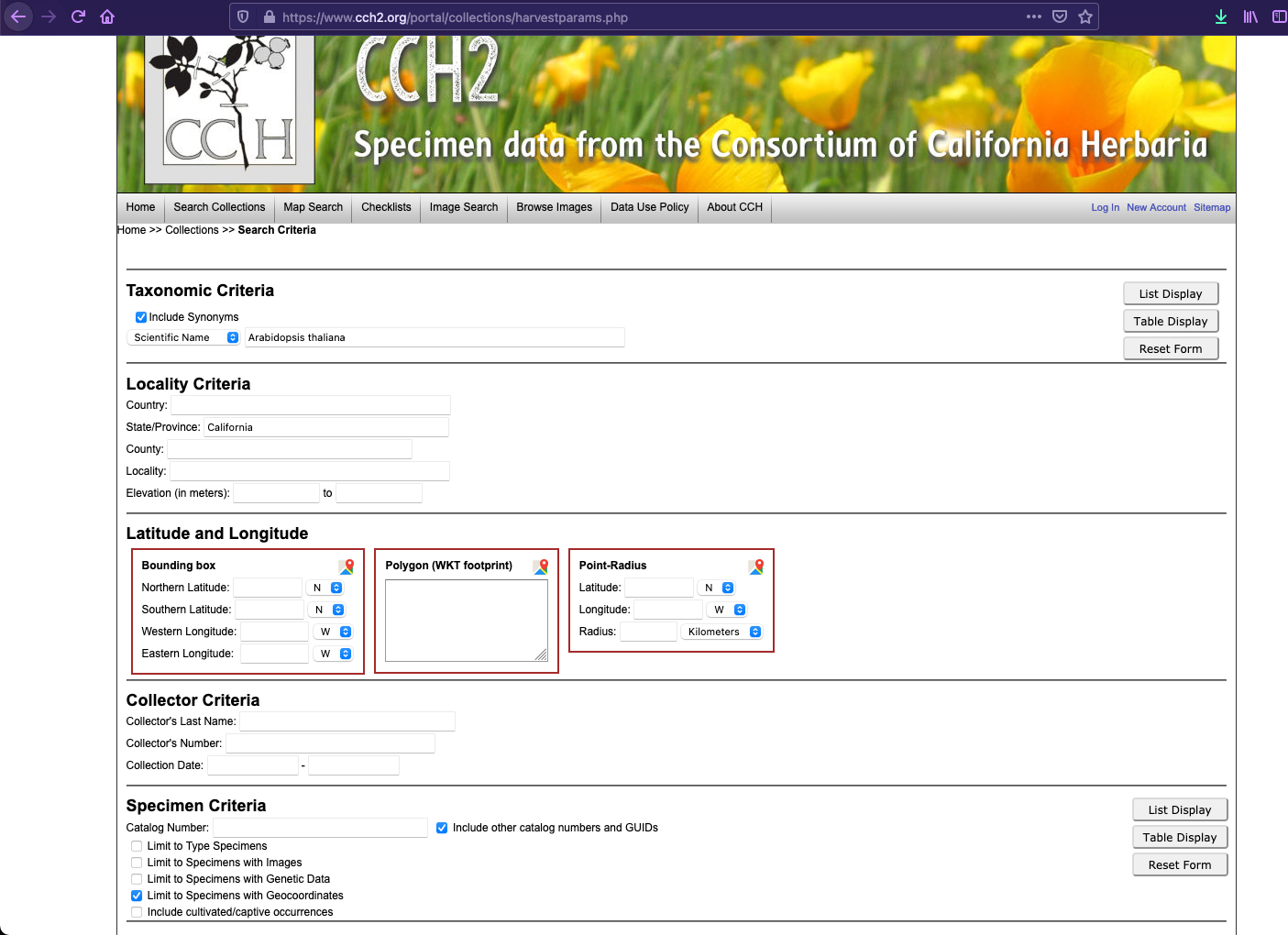
**Creating maps from herbarium records**

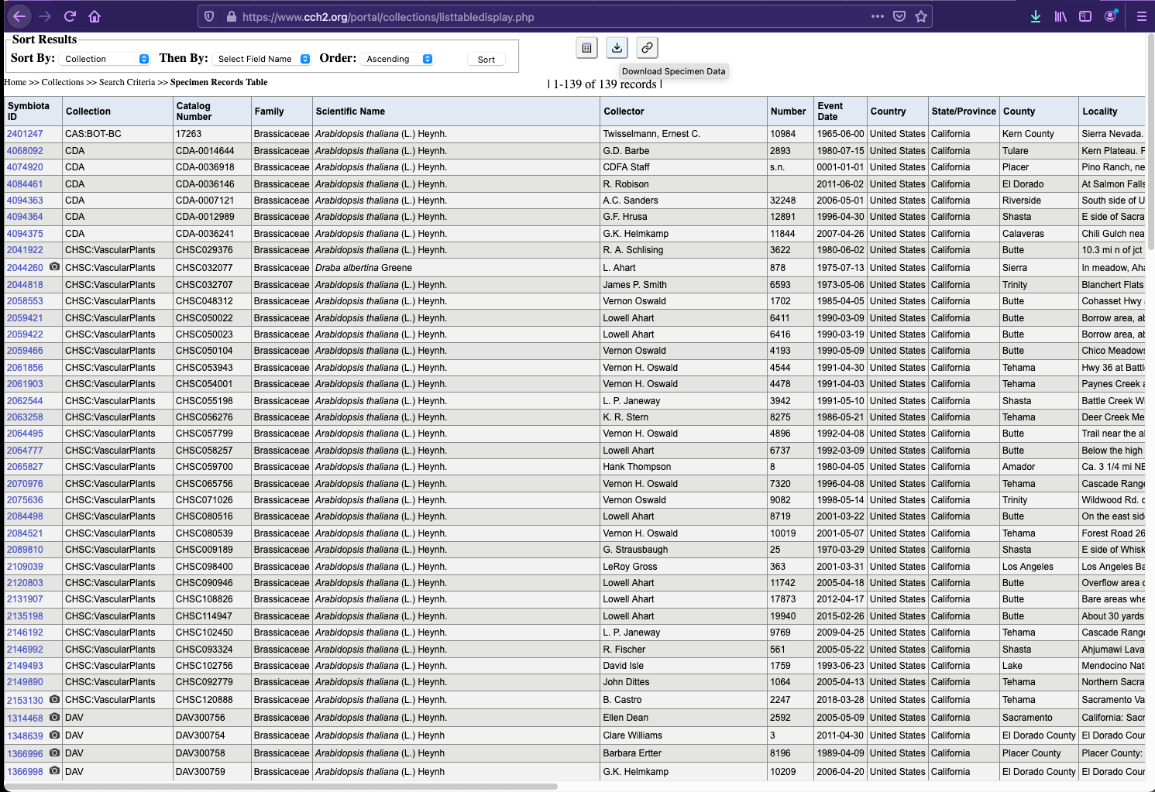
The purpose of this exercise is to create a map showing the location of records compiled in the database. In addition, the points in the map will be colored based on time intervals to identify range expansions or range contractions. This exercise can be used in class to help students understand the colonization process of invasive species (lag phase, exponential growth, points of introduction, etc.). It can also be used to help students understand range contractions of endangered species over time.

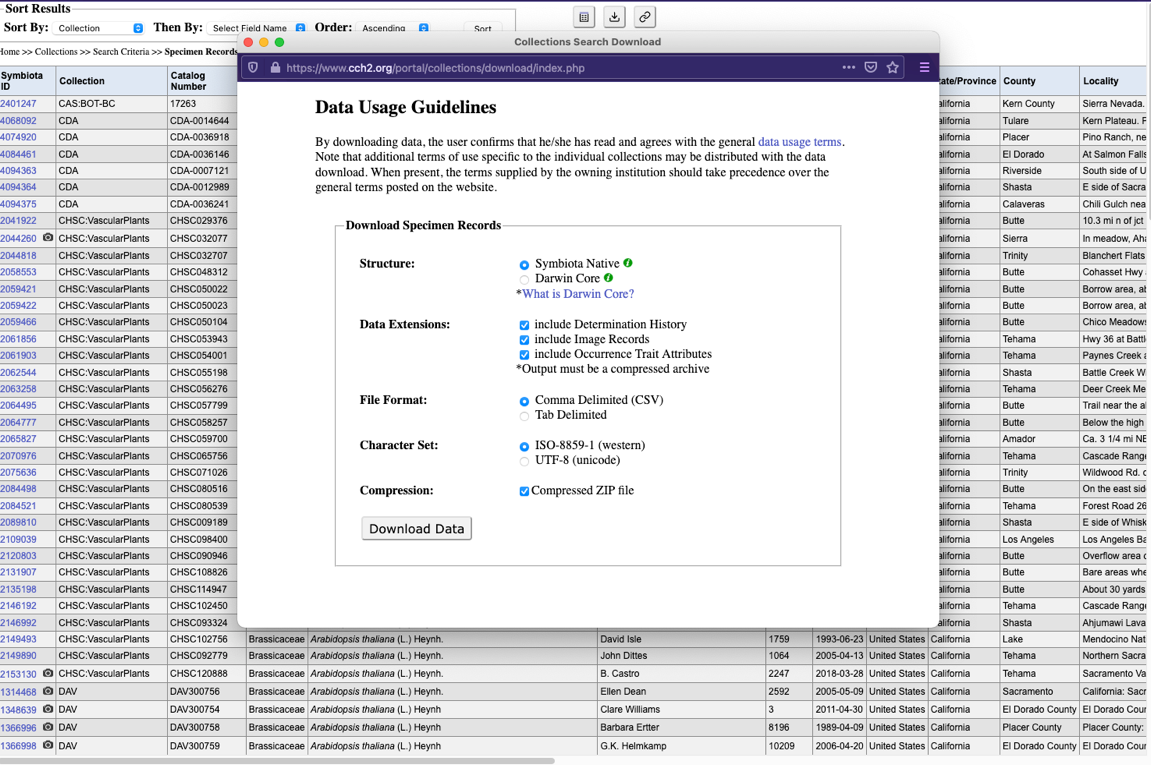
Methods:

1. Go to the CCH2 website <https://www.cch2.org> and to the tab “Search Collections”. In the criteria input your species name, geographic region of interest. For this example, we are using *Arabidopsis thaliana* in the state of California (see image below).

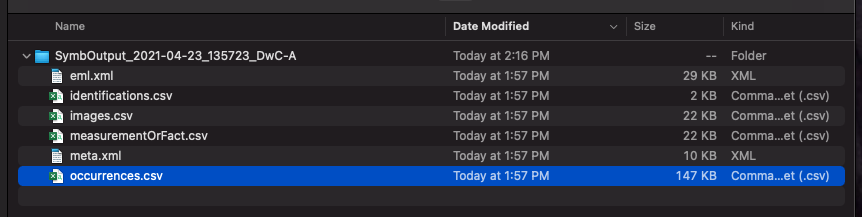


2. Select table display and download the data. When you click the Download Specimen data make sure you have in “File Format: Comma Delimited (CSV) (see images below). This last step is very important, otherwise R will have trouble reading the data set.

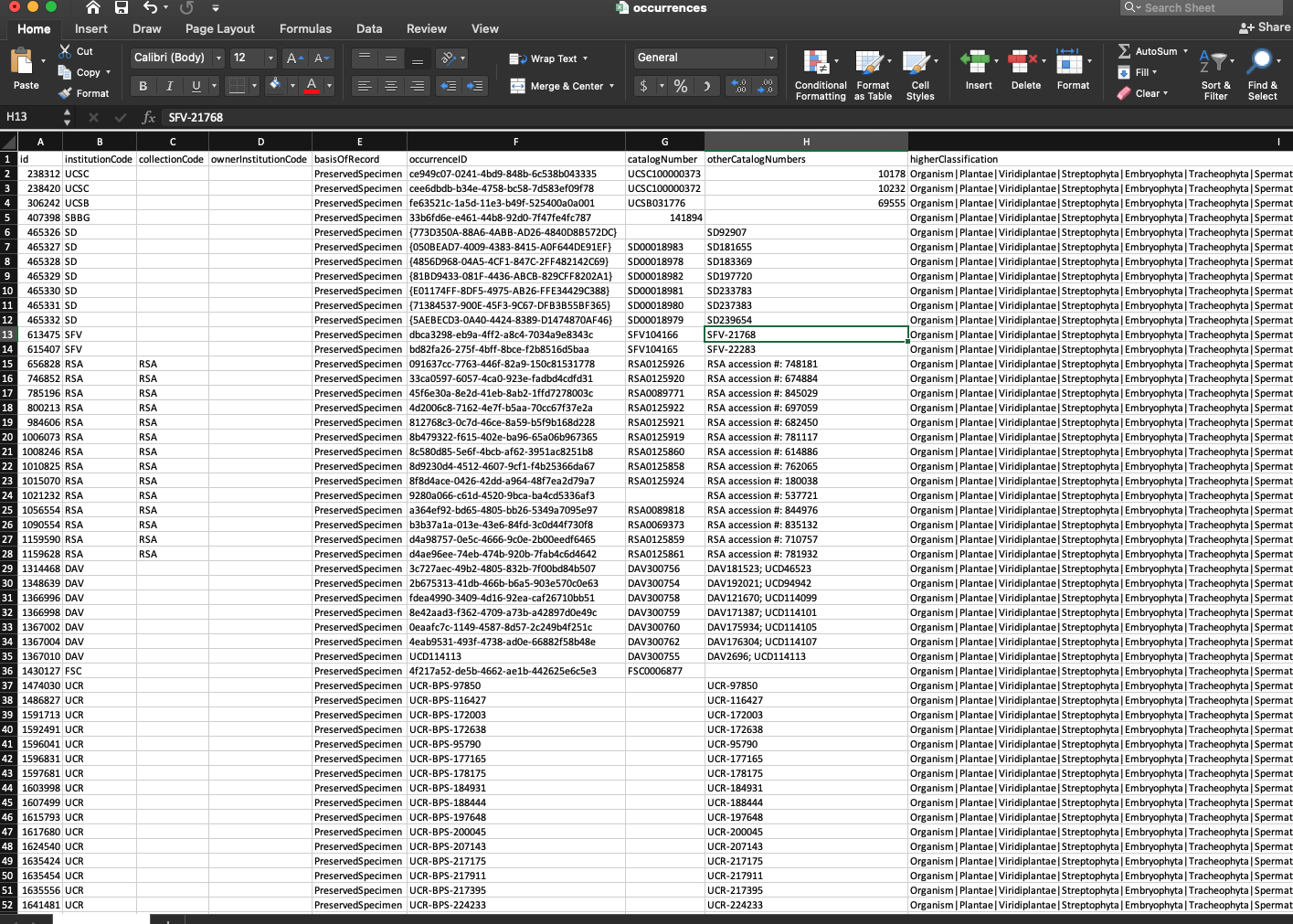




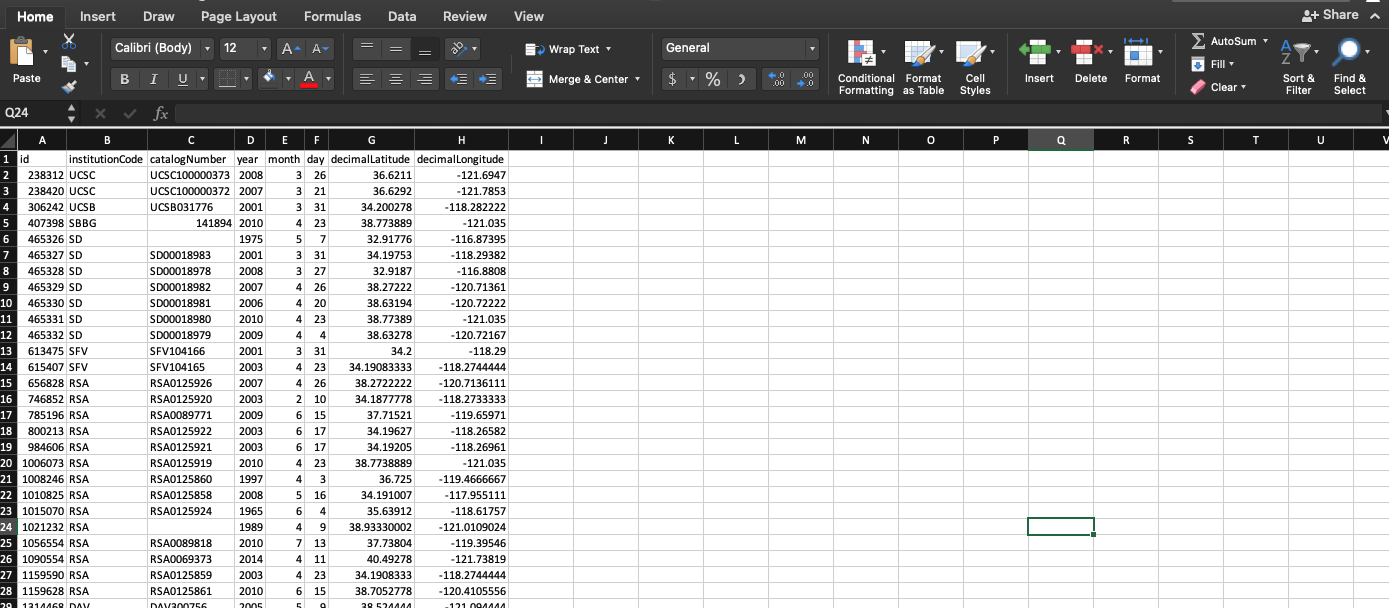
3. CSV files can be open in Excel in case you need to do any modification. After downloading the data, you have to unzip it (if you didn’t mark Compressed ZIP file you can skip the unzipping, note that downloading compressed files will save you download time). Once the folder is unzip you will see a set of files, the one you want is called occurrences.csv (see image)



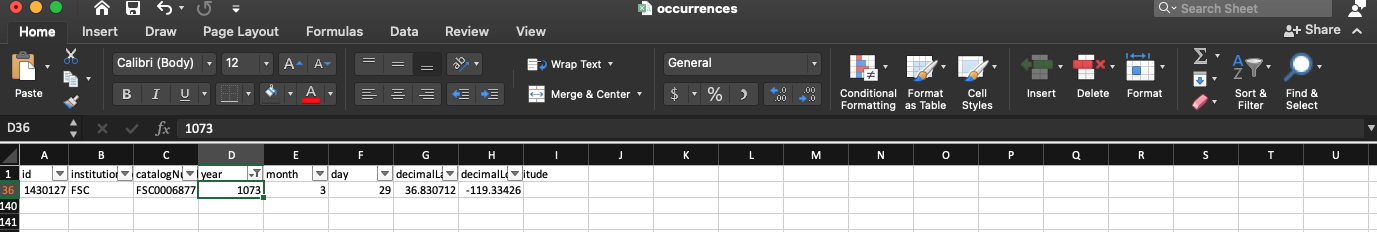
If you open your file in Excel, it will look like this:



Because we cannot choose exactly which field to download and R will complain about symbols, we have to clean the data. Remove all columns except the id, institutionCode, catalogNumber, year, month, day, decimalLatitude and decimalLongitude. Then sort based on latitude or longitude and remove all entries with no data for these two parameters. Now the data set will look like this:



One last check, filter the data set and check the range of dates for “year”. In some cases when there is no collection date, curators input in the date field year 1. If you find any like this, remove the whole entry. There could also be other year entries that were typed incorrectly, remove those too. For example, for this data set there was one 1 and also this (obviously 1073 is not a correct collection date):



4. Let’s move now to R. Open R studio and set your working directory (remember that when the cursor is between quotes you can use Tab to navigate folders). Then, load your data and use a name that is easy to remember so you won’t have issues invoking it later.

setwd("Downloads/SymbOutput\_2021-04-23\_135723\_DwC-A/")

mapDB <- read.csv("occurrences.csv")

5. Now that the data has been loaded remember to always double check that the file has properly loaded. We can do that using the command str which will show us the dataframe size and the type of data in each column.

str(mapDB)

And it prints:

'data.frame': 139 obs. of 8 variables:

$ id : int 238312 238420 306242 407398 465326 465327 465328 465329 465330 465331 ...

$ institutionCode : chr "UCSC" "UCSC" "UCSB" "SBBG" ...

$ catalogNumber : chr "UCSC100000373" "UCSC100000372" "UCSB031776" "141894" ...

$ year : int 2008 2007 2001 2010 1975 2001 2008 2007 2006 2010 ...

$ month : int 3 3 3 4 5 3 3 4 4 4 ...

$ day : int 26 21 31 23 7 31 27 26 20 23 ...

$ decimalLatitude : num 36.6 36.6 34.2 38.8 32.9 ...

$ decimalLongitude: num -122 -122 -118 -121 -117 ...

5. To make crating the plot easier we are going to load latitude and longitude and also year. Be careful to spell the header of the column where the data we want to load is located, after the $ you could use tab to autocomplete.

Latitude <- mapDB$decimalLatitude

Longitude <- mapDB$decimalLongitude

Year <- as.numeric(mapDB$year) #### here we use the as.numeric because when we did str(mapDB) the data in year was int instead of numeric.

6. Time to create the plot. First load the libraries

library(maps)

library (RColorBrewer)

First, we create our color palette with one color for each time interval. For this part you need to know how many time intervals you want to create for your map. To figure out the range of dates you have:

range(Year)

[1] 1914 2018

In this case it says we have from year 1914 to 2018, that is 104 years. We are going to make 4 intervals of 26 years for each interval. The intervals will be (1914-1939), (1940-1965), (1966-1991) and (1992-2018). Select the same number of colors as intervals and find colors that are different. The colors name can be found here <https://www.nceas.ucsb.edu/sites/default/files/2020-04/colorPaletteCheatsheet.pdf>

mypal <- colorRampPalette(c("orange", "royalblue", "palegreen4", "maroon2")) (4) ##this line creates your color palette and assigns 4 colors to 4 intervals.

colbreak <- seq(min(Year), max(Year), length.out=5) ## this line specifies the range of values for the intervals and the number of break for such range of values (note that you need one more break than number of intervals).

mapDB$colorcolumn <- cut(Year, breaks = colbreak, labels = mypal) ## putting together the range of values and the palette

mapDB$colorcolumn <- as.character(mapDB$colorcolumn) ## variable must be a character, turning it into a character

plot(Longitude, Latitude, col = mapDB$colorcolumn, pch=20) ## creating the latitude and longitude plot with the information for the color.

map(add=TRUE) # adding the map layer

legend("bottomleft", legend=c("1914-1939", "1940-1965", "1966-1991", "1992-2018"), col=c("orange", "royalblue", "palegreen4", "maroon2"), cex=1, pch=20) ## adding a legend to the map

7. And this is the result:

In yellow we can see the first samples of *Arabidopsis thaliana* collected in the state of California. This species was allegedly introduced in the US. In the map we can see that it is mostly found inland and along the mountain ranges. The fact that we have several distanced yellow dots could point out to more than one introduction event or to a lack of sampling in the region between them. Said that, Northern CA looks like an introduction point that then further spread towards the South (few yellow and blue dots that stay put for a long period -lag phase- in the north followed by green and then a much larger and farter spread in pink -exponential growth and spread- ).

