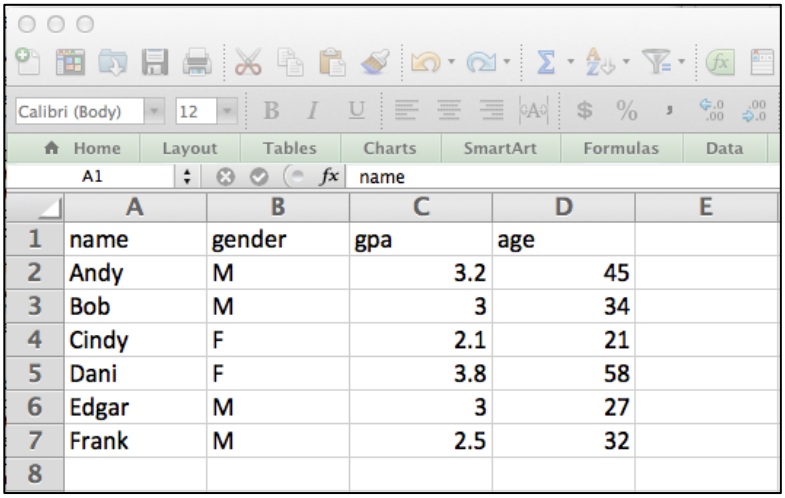
Module 2: Data SNACN

Goals for Module 2: Data SNACN

~Participants will format a spreadsheet to be compatible with SNACN software. Participants will load the spreadsheet into the SNACN software.

CSV Files

CSV is a file extension that stands for "comma separated values." A CSV file is essentially a text file in which information is separated by commas. A CSV is a non-program specific file format that allows spreadsheet software to read text and organize it into rows and columns. Spreadsheet software, like Microsoft Excel, automatically organizes text into different cells of a table based on the location of the commas. CSV files are very versatile files that can be read easily by virtually any software program than can read text or spreadsheets on any computer. Below is a screenshot of a sample CSV file. It is a spreadsheet that contains seven rows and four columns.



	A	B	C	D	E
1	name	gender	gpa	age	
2	Andy	M	3.2	45	
3	Bob	M	3	34	
4	Cindy	F	2.1	21	
5	Dani	F	3.8	58	
6	Edgar	M	3	27	
7	Frank	M	2.5	32	
8					

When you open a CSV file in a text-editing program such as TextEdit on Macs or Notepad on Windows you get the same information, but you can see the text separated by commas that underlie the CSV file. In the text-editing program are the same seven rows seen in the spreadsheet software, but instead of four columns each row contains three commas separating the data that would otherwise belong to each separate cell across four columns.

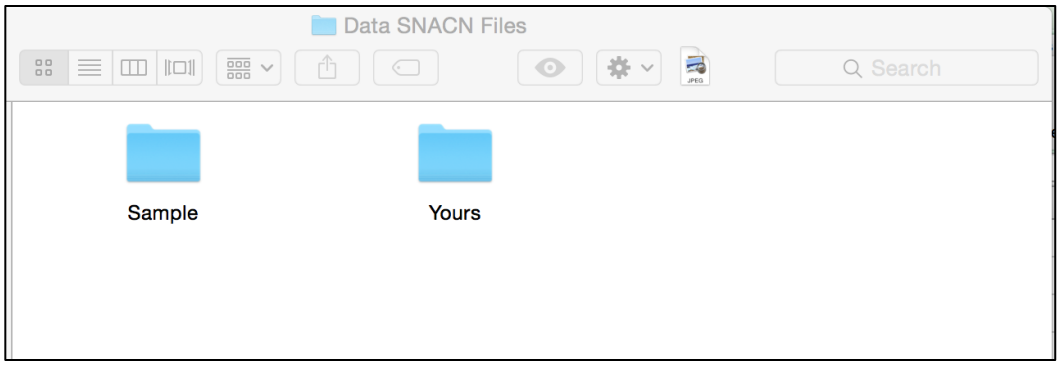
The CSV file format is compatible with SNACN and easily imports data into the software.

Arrest.csv

In order to successfully import arrest networks into SNACN, we must first master the spreadsheet requirements of SNACN. The good news is that most arrest records are similar enough that reformatting them to meet SNACN requirements should just take a little bit of work in your spreadsheet software, like Microsoft Excel. The bad new is that small mistakes in the spreadsheet could cause it not to load, and small mistakes are difficult to diagnose. For this reason, we will start by constructing our very own fake arrest spreadsheet in Microsoft Excel of a small network before working with larger examples. Follow the instructions below to craft an **`Arrest.csv`** of your own.

1. Open a new spreadsheet in Excel.

Located in your **`Data SNACN Files`** folder for Module 2 are two subfolders: **`Sample`** and **`Yours`**. See the screenshot below.



You need to save your new spreadsheet in the **`Yours`** subfolder. The reason for this is that all SNACN files require the exact same file name for the software to know which file to load. A single folder on a computer cannot hold files with the same names, so these separate subfolders keep different **`Arrest.csv`** files separate.

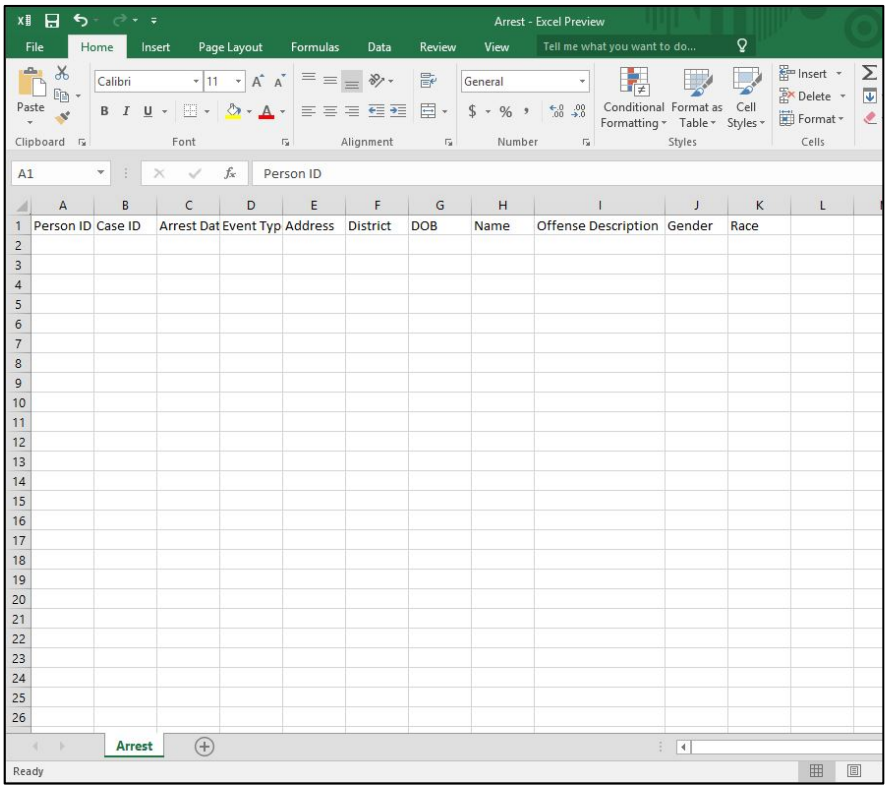
2. Save your newly opened and currently empty spreadsheet in the **`Yours`** subfolder with the exact file name: **`Arrests.csv`**
   The default format to save files in Microsoft Excel is an Excel workbook with the extension ".xlsx". We need to override this default by selecting **`Comma Separated Values (.csv)`** next to **`Format`** to change the file extension. You shouldn't need to type the **`.csv`** in the file name of **`Arrests.csv`** because changing the format automatically changes the file extension.

You should be able to see the complete file name at the top of your Excel window. If not, return the **`Yours`** subfolder to inspect the file name and file format. Once you are satisfied that the file is saved correctly, you can begin filling in the spreadsheet. First we will add the header row.

3. In row 1 of the **`Arrest.csv`** type the following headings exactly as they appear below. Replicate the order. Be sure to include only a single space between words and no spaces after the last word. Replicating the capitalization is not necessary, but it is recommended. There are 11 headings, so you will use columns A through K in the spreadsheet, typing one heading per cell in the following order across the first row.

```
Person ID
Case ID
Arrest Date
Event Type
Address
District
DOB
Name
Offense Description
Gender
Race
```

Hopefully some of these headings look familiar to you and you can imagine the equivalent headings that your agency uses.

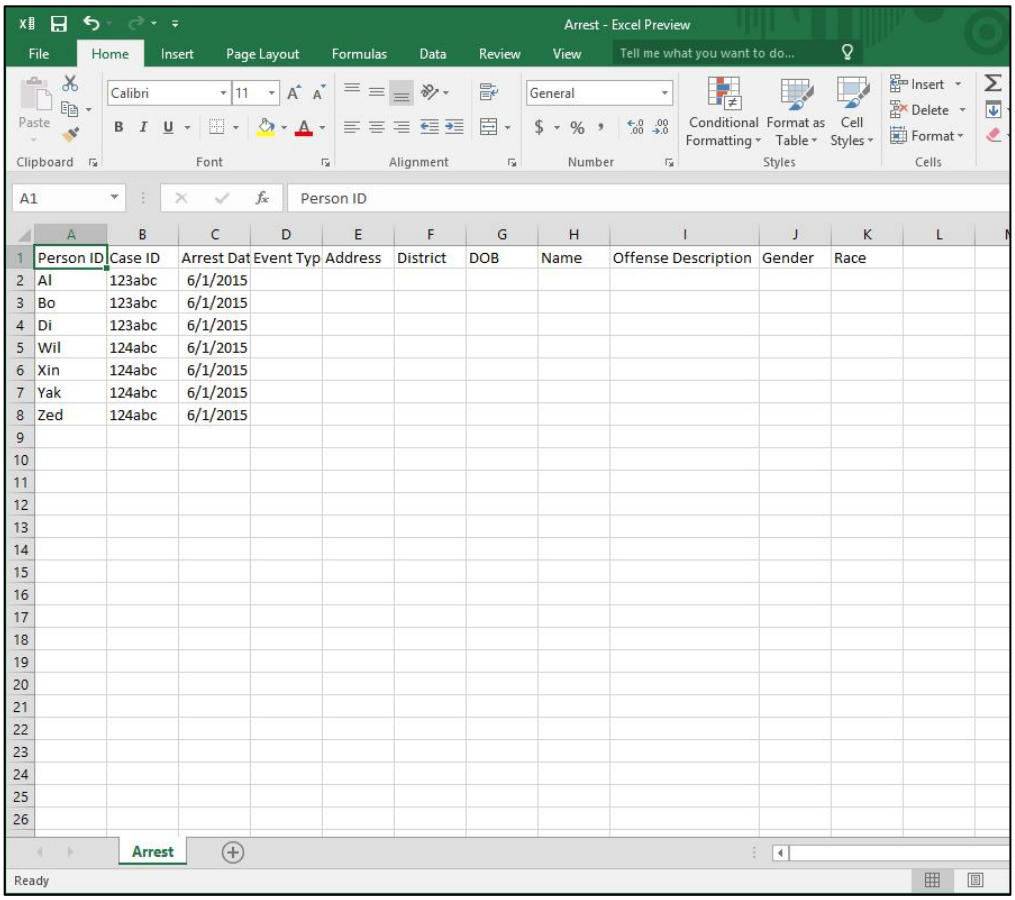


Now we can add some fake examples. Let's create two group arrest events. The first arrest will contain 3 people and the second arrest will contain 4 different people.

4. Create 3 unique Person IDs to represent the 3 people in the first group arrest. Person IDs can be names, numbers, or a combination of letters and numbers. They just need to be unique to each person. Create a unique Case ID number for this group arrest. Case IDs can also include letters and/or numbers. The Case ID will be the same for the 3 people in this group arrest. Create an Arrest Date for this event. We recommend choosing a date that has passed in the last year. (This is only because we will use the calendar function as a filter to check that our data loaded correctly, and more recent dates will help you avoid scrolling through months of calendars to find your dates.) Avoid using future dates. Optionally, you can fill in some of the other columns of this spreadsheet, but we will not be using the other columns in this lab.

5. Repeat the previous step for a second group arrest that includes 4 people. You will need 4 new Person IDs and 1 new Case ID. The date can be the same or different.

Here is an example screenshot of this exercise that will upload correctly in the SNACN software.

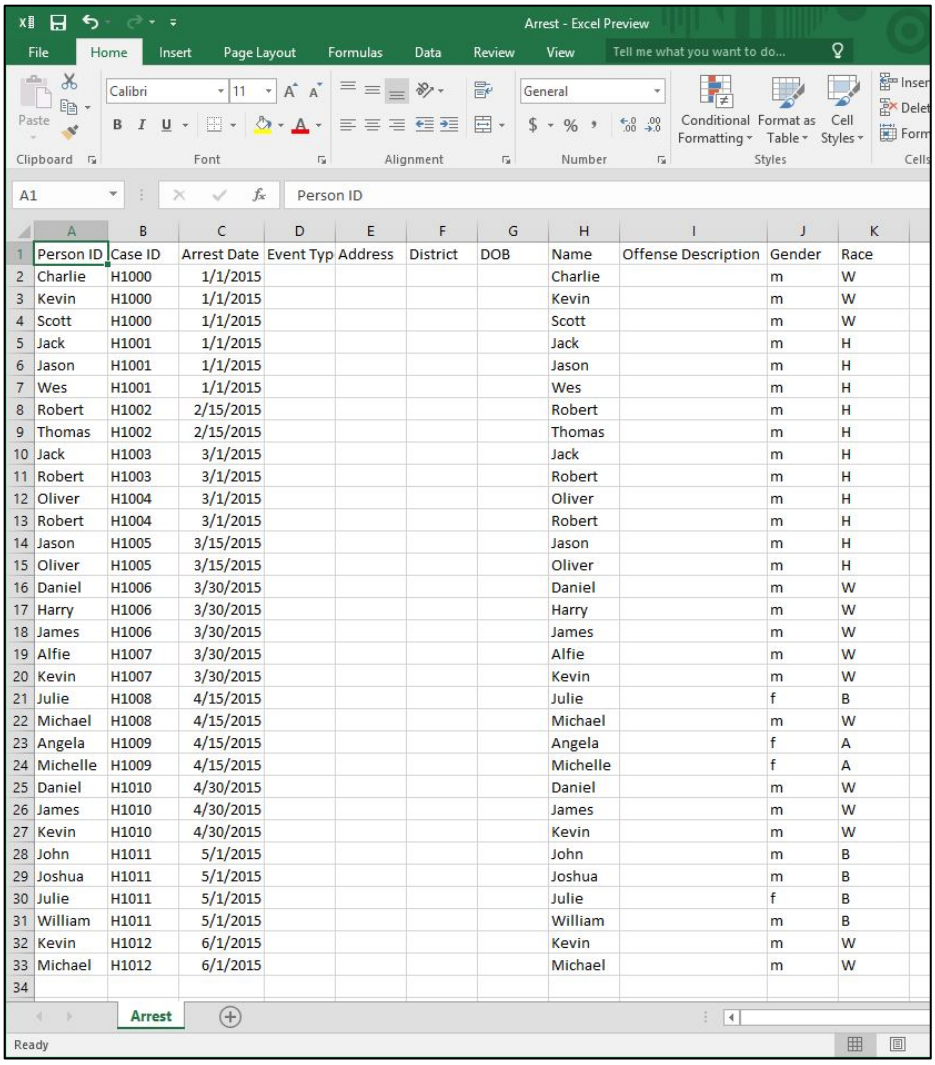


	A	B	C	D	E	F	G	H	I	J	K
1	Person ID	Case ID	Arrest Dat	Event Typ	Address	District	DOB	Name	Offense Description	Gender	Race
2	Al	123abc	6/1/2015								
3	Bo	123abc	6/1/2015								
4	Di	123abc	6/1/2015								
5	Wil	124abc	6/1/2015								
6	Xin	124abc	6/1/2015								
7	Yak	124abc	6/1/2015								
8	Zed	124abc	6/1/2015								

6. Save your `Arrest.csv` and close Microsoft Excel.

Data SNACN Files

This lab includes another premade **`Arrest.csv`** file that you can practice loading into SNACN software. This file is stored in the **`Sample`** subfolder located in the **`Data SNACN Files`** folder in order to avoid conflict between files with the same name. Open this premade **`Arrest.csv`** now to examine its contents. Be sure that the **`Arrest.csv`** from the previous task is closed. What differences do you notice between it and the one you made in the previous task?



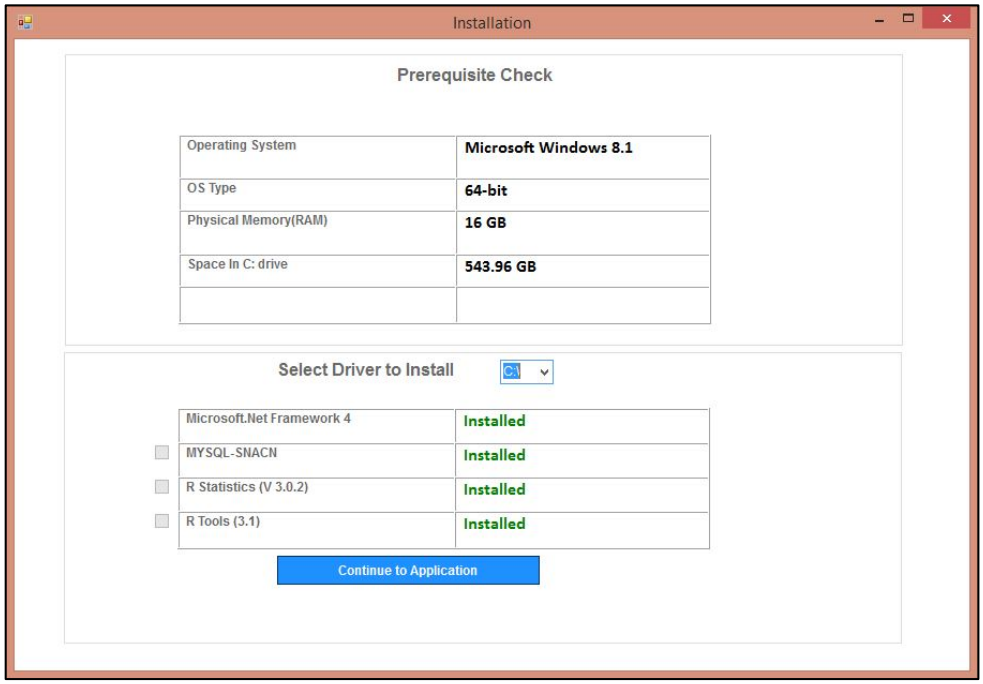
	A	B	C	D	E	F	G	H	I	J	K
1	Person ID	Case ID	Arrest Date	Event Typ	Address	District	DOB	Name	Offense Description	Gender	Race
2	Charlie	H1000	1/1/2015					Charlie		m	W
3	Kevin	H1000	1/1/2015					Kevin		m	W
4	Scott	H1000	1/1/2015					Scott		m	W
5	Jack	H1001	1/1/2015					Jack		m	H
6	Jason	H1001	1/1/2015					Jason		m	H
7	Wes	H1001	1/1/2015					Wes		m	H
8	Robert	H1002	2/15/2015					Robert		m	H
9	Thomas	H1002	2/15/2015					Thomas		m	H
10	Jack	H1003	3/1/2015					Jack		m	H
11	Robert	H1003	3/1/2015					Robert		m	H
12	Oliver	H1004	3/1/2015					Oliver		m	H
13	Robert	H1004	3/1/2015					Robert		m	H
14	Jason	H1005	3/15/2015					Jason		m	H
15	Oliver	H1005	3/15/2015					Oliver		m	H
16	Daniel	H1006	3/30/2015					Daniel		m	W
17	Harry	H1006	3/30/2015					Harry		m	W
18	James	H1006	3/30/2015					James		m	W
19	Alfie	H1007	3/30/2015					Alfie		m	W
20	Kevin	H1007	3/30/2015					Kevin		m	W
21	Julie	H1008	4/15/2015					Julie		f	B
22	Michael	H1008	4/15/2015					Michael		m	W
23	Angela	H1009	4/15/2015					Angela		f	A
24	Michelle	H1009	4/15/2015					Michelle		f	A
25	Daniel	H1010	4/30/2015					Daniel		m	W
26	James	H1010	4/30/2015					James		m	W
27	Kevin	H1010	4/30/2015					Kevin		m	W
28	John	H1011	5/1/2015					John		m	B
29	Joshua	H1011	5/1/2015					Joshua		m	B
30	Julie	H1011	5/1/2015					Julie		f	B
31	William	H1011	5/1/2015					William		m	B
32	Kevin	H1012	6/1/2015					Kevin		m	W
33	Michael	H1012	6/1/2015					Michael		m	W
34											

Import Data into SNACN

Likely there will be an icon on your desktop from the SNACN installation process completed in Module 1. If you don't have a desktop icon that looks like the graphic above, then locate SNACN among other programs on your computer.

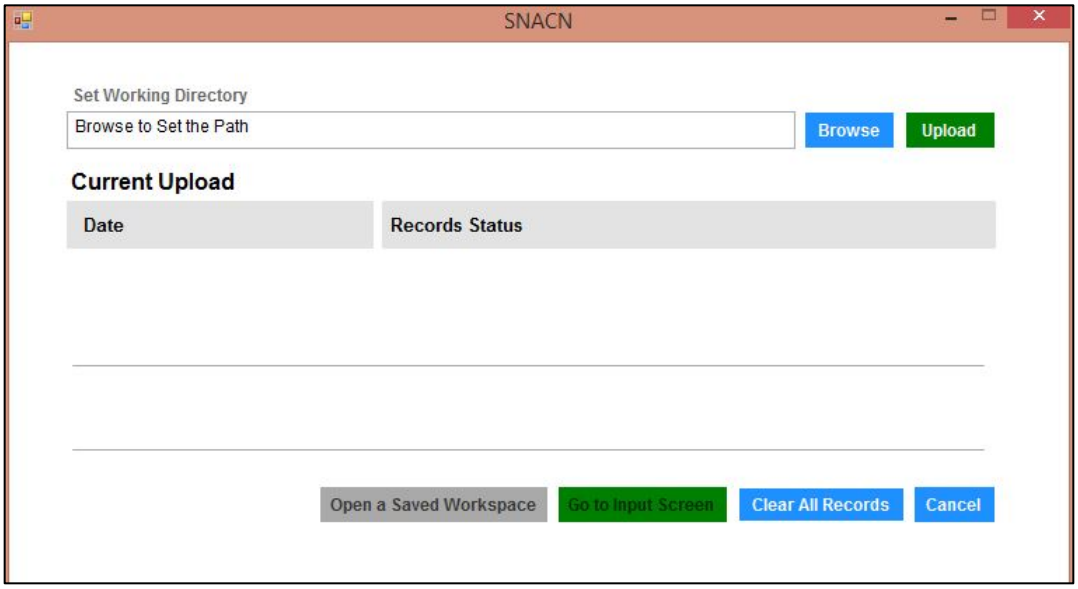
1. Open SNACN on your computer.

This will bring you to a pre-requisite check widow. Hopefully everything installed correctly during Module 1 as indicated by the list of green `Installed` confirmations.



2. Click on the blue `Continue to Application` button at the bottom of the window.

This brings us to the window where we upload our data. If you uploaded any data in the previous session then you might see the counts stored here.

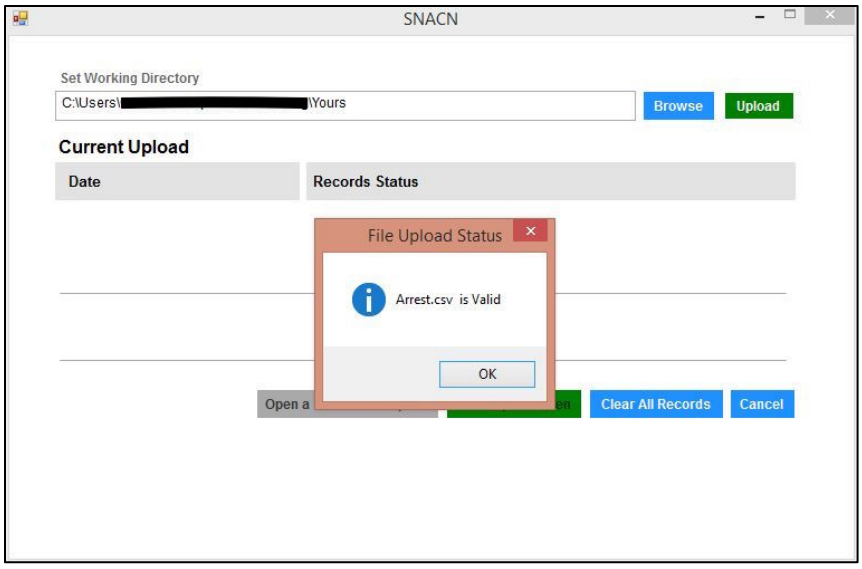


First let's upload the **`Arrest.csv`** that you made of the two group arrests. We stored this file in the **`Yours`** subfolder within the **`Data SNACN Files`** folder for the module.

3. Click on the blue **`Browse`** button toward the top of the window. Locate the **`Yours`** subfolder. Click **`OK`**.

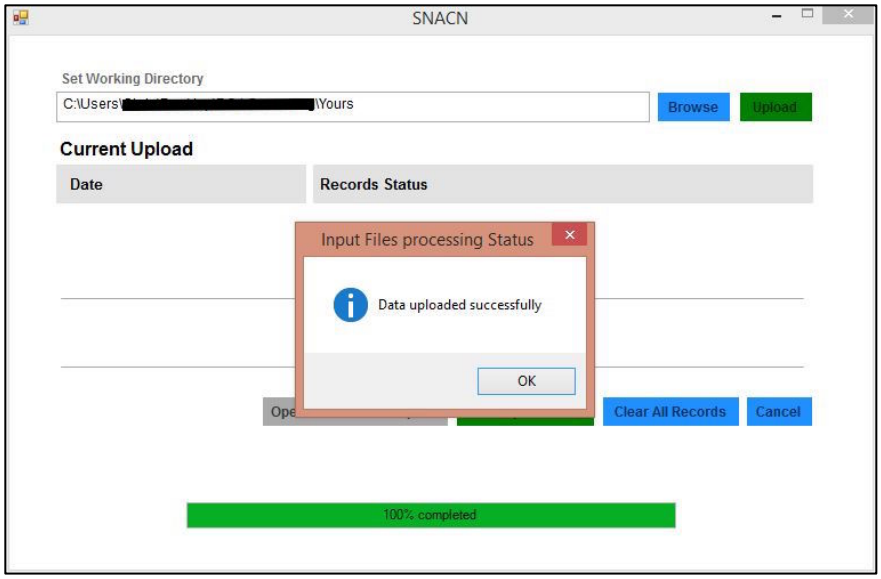
You will see your computer's path to the **`Yours`** subfolder now located in the **`Set Working Directory`** bar.

4. Click on the green **`Upload`** button next to **`Browse`** to see if the software can find the **`Arrest.csv`** file that it needs.

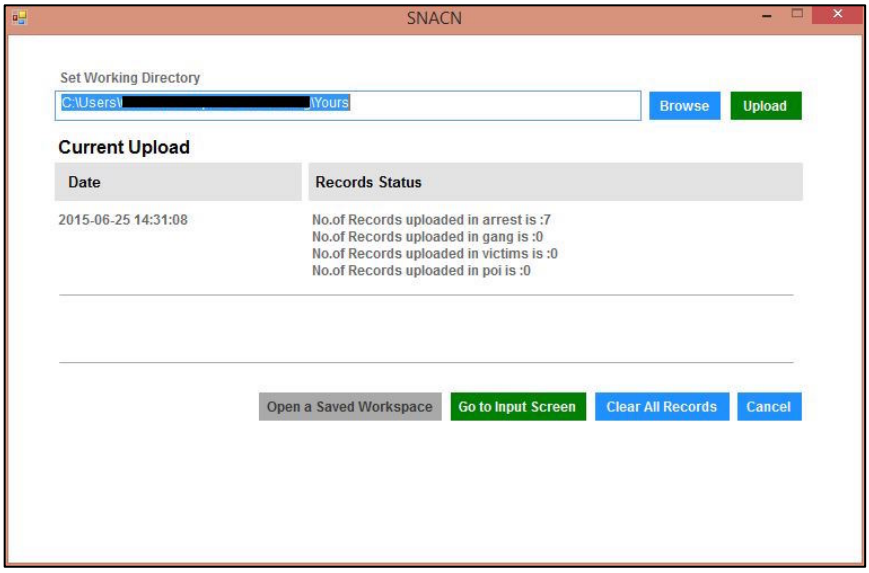


Feel validated if you see the **File Upload Status** pop-up window like the one above. The SNACN software recognizes your **Arrest.csv** file.

5. Click the **OK** button in the pop-up validation window to upload the **Arrest.csv** and watch your data load.



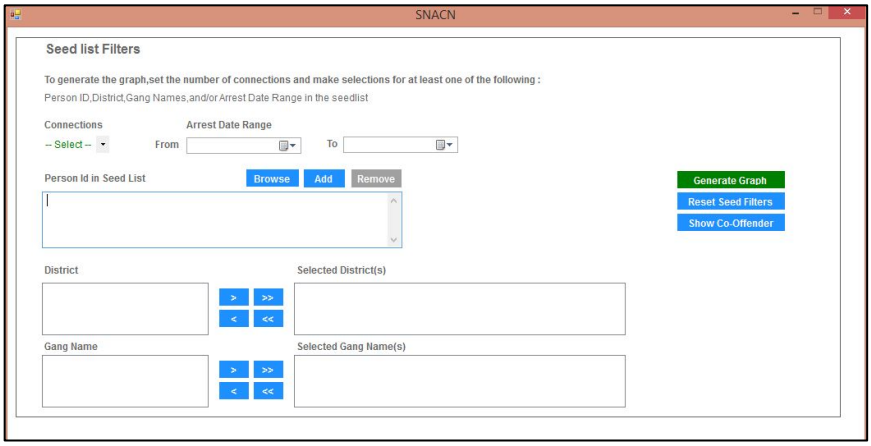
5. Click the **OK** button in the pop-up **Input Files Status** window to see how many records imported.



As you can see in the top row of **Records Status**, we have 7 cases of individuals being arrested. You can also see the time that you uploaded these data, which might be useful if you return to the SNACN software at a later time and it continues to show this timestamp.

Module 3: Visualization focuses more on how to generate and customize network images in SNACN, but let's make a basic network image now to check that we structured the **`Arrest.csv`** file correctly. Recall that we wanted to create a file that would display 1 group arrest of 3 people and 1 group arrest of 4 people. Using the most basic graphing commands, we can see how we did.

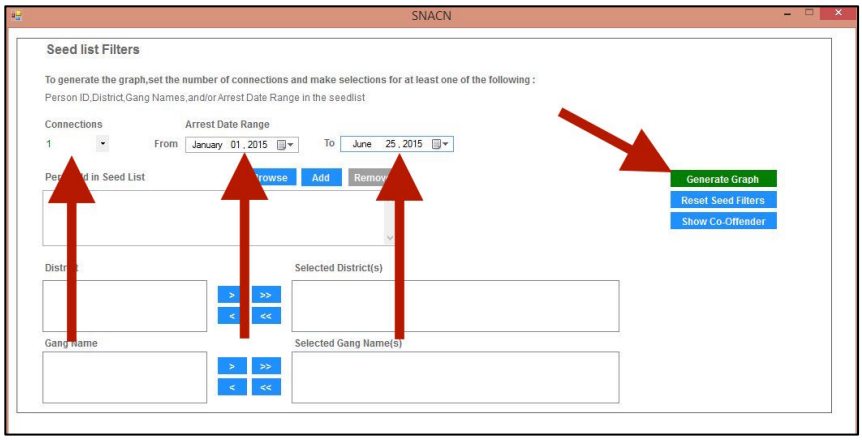
6. From the Upload window, click on the green **`Go to Input Screen`** button located near the bottom of the window, which will take you to a new **`Seed List Filters`** window shown in the screenshot below.



This screen is designed to extract small sections of an arrest network of interest based on a particular set of nodes. But for our purposes, our network is so small that we just want to make a network image of the entire dataset. In order to do this, we use the **`Connections`** and the **`Arrest Date Range`** filters.

7. Scroll through the calendars in the **`Arrest Date Range`** to select the earliest date you entered in your **`Arrest.csv`** and the latest date you entered in your **`Arrest.csv`**. Earlier we recommended using dates in 2015 that had passed, so hopefully you don't have to click through too many calendar months to get to the beginning of your time period. The end of your date range can be today's date since that range should cover what we put in our **`Arrest.csv`**.

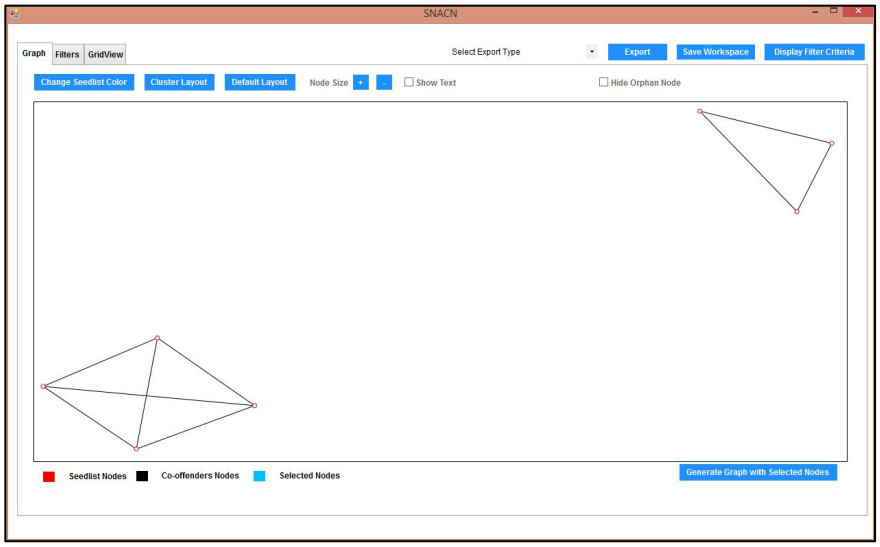
8. Click on the dropdown arrow underneath **`Connections`** and select the number **`1`**. We will go into more detail about what values higher than 1 mean in Module 4: Analytics, but for now we are telling the SNACN to include the direct connections to every person arrested within our date range or every person within a single tie of those arrested within our date range. This will include everyone in the network because of the date range.



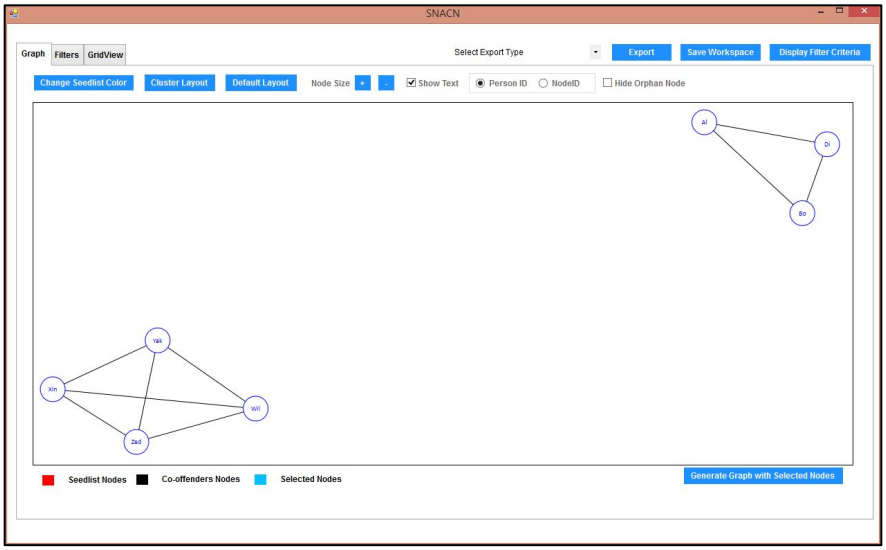
Once the **`Connections`** and the **`Arrest Date Range`** filters are set, we can generate the network.

9. Click on the green **`Generate Graph`** button located on the right side of the window.

Viola! A new window will open containing the arrest network using the default plotting settings of SNACN. If everything went accordingly, we should have a triangle for our group of 3 and a square with an x through it for our group of 4.



9. Checking the **`Person ID`** in the **`Arrest.csv`** is easy. Click the **`Show Text`** box toward the top middle of this window. Use the plus sign next to **`Node Size`** to increase the circle to be big enough to fit the names. And just for fun, click on **`Change Seedlist Color`** to change the color of the circles around the nodes.



10. Return to the beginning of this section of the lab on importing data and try uploading and making a basic network image of the **`Arrest.csv`** file in the **`Sample`** folder.

Review of Module 2: Data SNACN

CSV is a file extension that stands for "comma separated values." A CSV file is essentially a text file in which information is separated by commas. CSV files are very versatile files that can be read easily by virtually any software program than can read text or spreadsheets on any computer.

`Arrest.csv` files are a set format that can be imported into SNACN software. They are two-mode data and require a particular set of headings. Because the file names are the same across different `Arrest.csv` files use subfolders to differentiate them, so saving a new `Arrest.csv` doesn't delete over and replace other files with the same name.

Loading two-mode data into SNACN software requires setting the working directory to the subfolder where the `Arrest.csv is stored` and uploading the data. You can quickly inspect the entire contents of the `Arrest.csv` file by including all of the dates in a date range filter and setting the connections to 1.