SUMMARIZING FROG AND TOAD COUNT DATA

This set of protocols will take you through all the steps necessary for summarizing the frog and toad data for each NAAMP route that has been assigned to you.

BEFORE YOU START, HERE ARE SOME IMPORTANT DATA REMINDERS, BOTH FOR THIS PROJECT AND FOR YOUR LIFE IN GENERAL.

- Don't leave data entries blank – if they're blank someone (including you) might just think you forgot about them. If a variable comes out as zero, then enter 0. If you can't calculate something for whatever reason, enter NA for "Not Available."

- Be careful and consistent with formatting. Data analysis software can do amazing things, but it can't read your mind (at least not yet). For example, a software package won't know that "N" means the same thing as "no" or even the same as "n". So, please follow the templates exactly and use consistent notation with things like capitalization.

- Never, ever guess! I know this goes counter to all the student skills you've accumulated over your life. But, please think of yourself as a scientist, not a student. If you have some extra NAs, no problem, we can drop these and still have plenty of data to find out what we want to find out. But incorrect entries can really mess things up as far as the data analysis goes.

- If you have questions, please ask me (marsh@nceas.ucsb.edu) or ask your instructor. This project is challenging and we don't expect anyone to understand everything at the beginning. Asking questions is part of the process.

NOTES ABOUT THE FROG CALL DATA

The raw NAAMP data are collected by volunteers and entered into a web-form for submission to USGS. The NAAMP coordinator for each state then looks these over for problems before they go into the database. In general, we've found very little obvious weirdness with the NAAMP data. However, inaccuracies are always a possibility, so you should always check for anything that looks surprising.

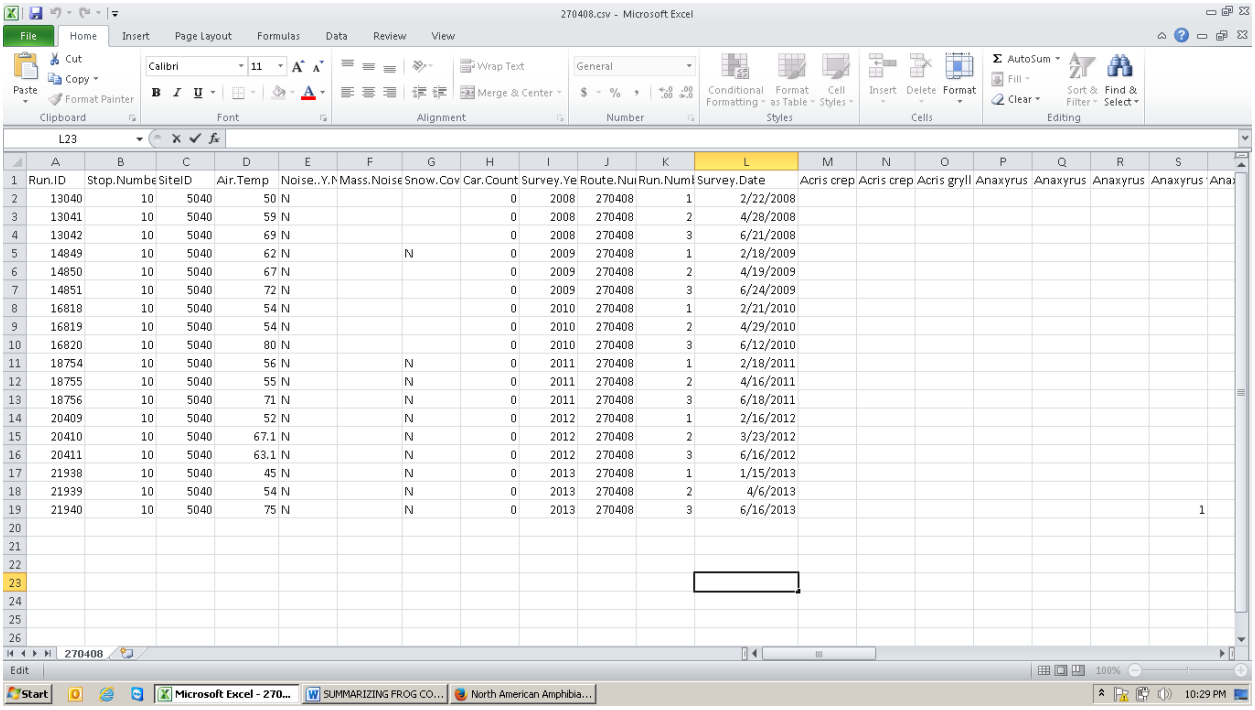
The NAAMP data are stored by USGS as a set of records associated with each run (e.g. the number of the survey for that year, the date of the survey, the number of days since rainfall), a set of records associated with each stop (e.g. air temperature, moon visibility, the number of cars that went by during the survey) and a set of records for frog and toad counts (which species heard and their calling intensity). Storing the data this way ends up saving a lot of space, which is important since the total dataset probably encompasses around 40 or 50,000 surveys. Unfortunately, with the raw data, you can't really see what's going on, so we've repackaged the raw data for this project. The repackaging combines variables from each of the three datasets, eliminates many variables that we won't need (e.g. moon phase), and it also allows you to see which species were *not heard* at a given stop. Still, the

repackaged data need some work to get into a good format to be analyzed, which is where this protocol comes in.

PROTOCOL

Start by opening each data file in Excel or other spreadsheet software. Go to the data site (http://mirrors.nceas.ucsb.edu/trn/) and type in the user name and password given to you by your instructor. Then, click on the folder for 2014. Download the folder for your route. Note the contents in each folder. First, there a csv/excel file that contains the frog and toad data for that stop. We'll be using this. Then there's a folder numbered that contains all the map data (see protocol for getting starting in qGIS).

A sample data sheet will look something like this (this one's 270408):



I. FIRST, WE'LL TAKE YOU THROUGH A SUMMARY OF EACH OF THESE VARIABLES (I.E. THE COLUMNS):

A) **Run.ID**. A run is a survey of the 10 stops within a route. *Each time an individual does a NAAMP survey, this survey is assigned a unique Run ID.* Typically, there are 3 runs for each route in each year, corresponding to three different time periods in which amphibians might be calling. Sometimes there are fewer than 3 runs when an individual can't do all three surveys. On occasion, there will be multiple runs in the same year with the same run number (i.e. 1, 2, or 3). These represent cases where an individual surveyed their route more than once within a designated survey period.

B) **Stop.Number**. There are 10 stops along each route at which an individual listens for frogs. However, we have included one randomly-chosen stop from each route. If we did more stops than this, the landscapes around the stops (at least at 5 and 10 km) would largely overlap and we wouldn't be getting independent data. All the runs within each file should all have the same Stop Number – if for some reason this is not the case, please let me know.

C) **SiteID**. Each stop has a unique SiteID. We will use these IDs to relate the frog data at each site to the landscape data at each site. *In theory, it is possible that SiteIDs would vary for a given stop if that stop had been moved at some point. If you find one of these, please let your instructor know.*

D) **Air.Temp**. This is air temperature in Fahrenheit at the time the stop was surveyed. You may not end up using this for anything, but it's there just in case.

E) **Noise.Y.N**. Sometimes during a survey there will be noise that could interfere with one's ability to hear frogs (e.g. traffic, people talking, etc.). Surveyors note when such noise is present (Y), and we'll likely test analyze our data both including and excluding sites that tend to be noisy.

F) **Mass.Noise.Index.** This is something called the Massachusetts Noise Index, an index of 1-5 for the type/strength of any noise heard. This variable has not been collected for the majority of states, so don't be surprised in yours are all blank.

G) **Snow.Cover.** This was included only accidentally, but I didn't want to have to reprocesses all the files just to remove it.

H) **Car.Count.** In many (but not all) of the surveys, surveyors recorded the number of cars that went by during the survey of each stop. This may end up being important data, as it provides an assessment of nocturnal traffic at the stop that could have negative consequences for frogs and toads (i.e. they could get squished). Please note: '0's FOR CAR COUNT ARE NOT THE SAME AS EMPTY CELLS. A zero means that no cars went by, whereas an empty cell means that these data were not collected. You need to pay attention to this when figuring out the average number of cars for the stop.

I) **Survey.Year**. This is just the year in which the survey took place.

J) **Route.Number**. Each route has a route number. The first two digits of this number are unique to whichever state contains the route. Make sure the route number in your spreadsheet is the same as the route number in the name of the folder.

K) **Run.Number**. Again, this is just the number of the survey (1,2, or 3) within each year. These numbers correspond to the time period for the survey, so if a survey is missing it could be 1,2 or 3 (i.e. you could have a run 2 and 3 in a given year, but no run 1).

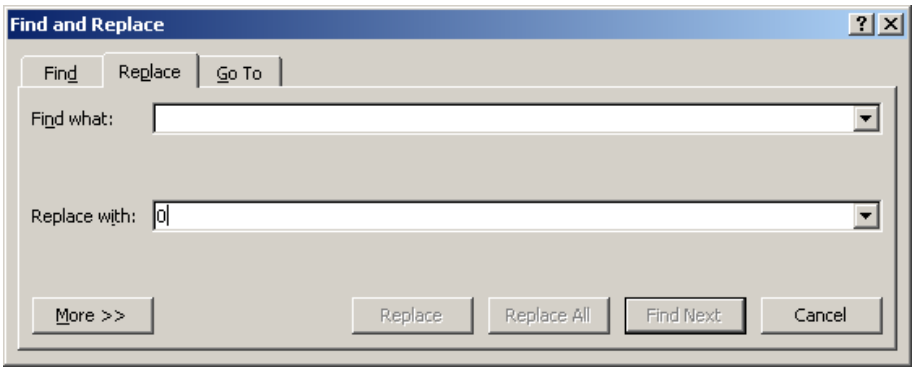
L) **Survey.Date.** This is the date of the survey. We won't formally use these data but it's nice to have this information in case there's anything confusing going on with the run numbers and years.

M, N, O) Call intensity data for each species. From here on out, the datasheet lists each species and then a blank (not detected) or a 1,2, or 3 for the intensity of calling. Here's what the numbers mean:

1: Individuals can be counted; there is space between calls
2: Calls of individuals can be distinguished but there is some overlapping of calls
3: Full chorus, calls are constant, continuous and overlapping

You can think of these as corresponding to "a few", "some", and "many" if that makes it easier to remember. In some cases, we may consider the call intensity, but in others we'll just look at whether or not a given species is detected.

Blank spaces in the call intensity data correspond to zeros (NAAMP doesn't record anything when species are *not heard*.) To do your calculations, you might want to have zeros instead of these empty spaces. To do this in Excel, just block the whole section you want to add zeros to. Then click on "Replace" at the far end of the main (Home) Excel screen. If you just leave the Find what: space blank and type in 0 in the Replace with: space, Excel will replace all the blanks in this section with zeros. If you do this, just look over the sheet carefully and make sure this did what you wanted.



Caution – elsewhere in these data sheets (i.e. outside of call intensities) blank spaces do not necessarily indicate zeros, so do not do this find/replace unless you're told that it's OK.

II. NOW WE'LL TAKE YOU THROUGH WHAT YOU NEED TO CALCULATE FROM THESE DATA.

You will need to summarize the frog and toad data for each stop. Following are all the variables you need to summarize for each stop. The data templates are given state by state through the Data tab on the project website: LINK TO STATE TEMPLATES

The state-specific templates are set up to make your life easier in a few different ways. First, we've eliminated species that don't occur at all in that state. This is why there are many fewer species listed in the state templates (i.e. many fewer columns) than in the raw data. Second, the templates include an NA ("not available") anywhere a particular species has not been in observed in that county. If you're working on a small state (e.g. DE) or a state that's not very ecologically variable (e.g. MA), you won't

necessarily have many of these. But for large states (MO, MN) or very variable states (e.g. NC, VA), you will find many NAs where species occur in only part of the state.

One note about this: you may occasionally have frogs recorded in the NAAMP data that are listed as NA in the data templates. These could be mistakes (i.e. the frog wasn't actually there) or they could represent new records of species. Since we have no way of knowing, we're going to play at safe and exclude any frogs heard outside of their known range. So, if it's "NA" in the template, it should stay "NA", even if the NAAMP data show something there.

OK, now for the template. The first three columns are the identifying information for each stop. These columns are: A) SiteID, B) Stop number ("Stop.Number"), C) Route Number ("Route.Number"). These can just get copies straight from the raw data for that stop/route. The fourth column is D) Entered.By This is your complete initials, so that we know who to ask if we have questions. Then come the columns with actual tabulations/calculations.

E) TOTAL NUMBER OF SURVEYS ("Num.Surveys"). This is just the number of surveys (i.e. rows) for each stop/ SiteID.

F) TOTAL NUMBER OF YEARS ("Num.Years"). This is the number of years in which the stop was surveyed at least once. You can just look at the dates when the surveys were done and count how many years there were.

G) MEAN CAR COUNT ("Car.Count"). This is the mean (average) of the car count data for each stop. ***Make sure to exclude empty cells when averaging – these are runs when cars were not counted, they are not the same as zero car (which are indicated with an actual 0).***

H) SURVEYS WITH NOISE ("Noise.Num"). This is the number of surveys for each stop in which Noise was noted (Noise = "Y"). You'll include surveys with noise in all your calculations, but when we're all done, we'll also do an analysis that excludes stops with frequent noise to make sure we get the same results.

AFTER YOUR LIST OF SPECIES: SPECIES RICHNESS ("Richness"). This column follows the set of species for the state, but it's important so I wanted to point it out here. This is species richness for each stop (i.e. number of species detected at each stop) with a species considered present if it was recorded in at least one survey. It's probably easiest to tally up which species were present first (see below) and then calculate the species richness at the end, which is why this column follows the individual species columns.

I, J, K, etc., up to "Richness"

For each species, we'll record two summary statistics, so you'll have two columns for each species. The first set of columns for each species comes before "Richness" and the second set comes after "Richness." This first set of columns registers whether a species is present at a stop based on whether it was ever heard at that stop. Presence will be recorded as a "1" and absence as a "0". For *Acris crepitans* (for example), this column would be called ACCR; for *Lithobates clamitans* this column would

be called LICL. Most of the abbreviations for species will just be the first two letters of the genus name following by the first two letters of the species name. However, there are a couple of cases where this doesn't work, so please consult the SPECIES ABBREVIATIONS (LINK HERE) file under the Data tab if you're at all confused about how the columns match up with the raw data.

In two cases, you may be asked to make adjustments in terms of species records to improve the quality of the data and to make it easier to combine results across states. First, *Hyla versicolor* & *H. chyroscelis* (Gray treefrog and Copes' Gray Treefrog) have subtly different calls and some (though not all) states combine them so that surveyors don't have to distinguish them. We will lump them as well using the column called "sum.HYCV" (*Hyla chrysoscelis* + *versicolor*). For this combined column, if you have any record of HYCH or HYVE or HYCV or HYCH/HYVE then "sum.HYCV" = 1. If none of these (HYCH, HYVE, HYCV, HYCH/HYCE) are present, then sum.HYCV=0.

Second, the chorus frogs in the *Pseudacris triserata* complex are a mess taxonomically. Some biologists separate *P. triserata* into four distinct species, some split them into three species, and some treat them as a single species with 4 subspecies. Further complicating this, some of the species and/or subspecies hybridize where they meet. Given all this, we're just going to combine the complex into as a single column called "sum.PSTR". The following species/complexes are all included in sum.PSTR: PSFE,PSFK, PSFT, PSKA, PSMA, PSMT, PSMF, PSNI, PSNF, PSTR. *So, if any one (or more) of these species is present, then sum.PSTR=1*. Otherwise, sum.PSTR = 0.

Whether or not either of these species complexes is a problem will depend on your state. If your instructor does not tell you how to deal with these for your state, be sure to ask them if you have any questions about it.

Once you've recorded which species are present, you can add these down the row to get the value for species richness. However, please remember to use sum.HYCV and sum.PSTR in place of the individual columns that make up these. Gray treefrogs and the *P. triserata* complex can each add only 1 unit to species richness for each stop.

In addition to presence/absence for each species, we'll also record the mean calling index for each species. This will allow us to check whether the same landscape features that affect presence/absence also affect abundance (or at least abundance of calling males).

The mean calling index for each species is just the mean of all the recorded values (including zeros) for that species. For example, if *Acris crepitans* was recorded in 8 surveys as: 1, 3, 0, 0, 0, 2, 3, 1, then the mean calling index (ACCR.MEAN) is equal to 10 (the sum of the calling index values) divided by 8 (the number of surveys of the stop), which is 1.25.

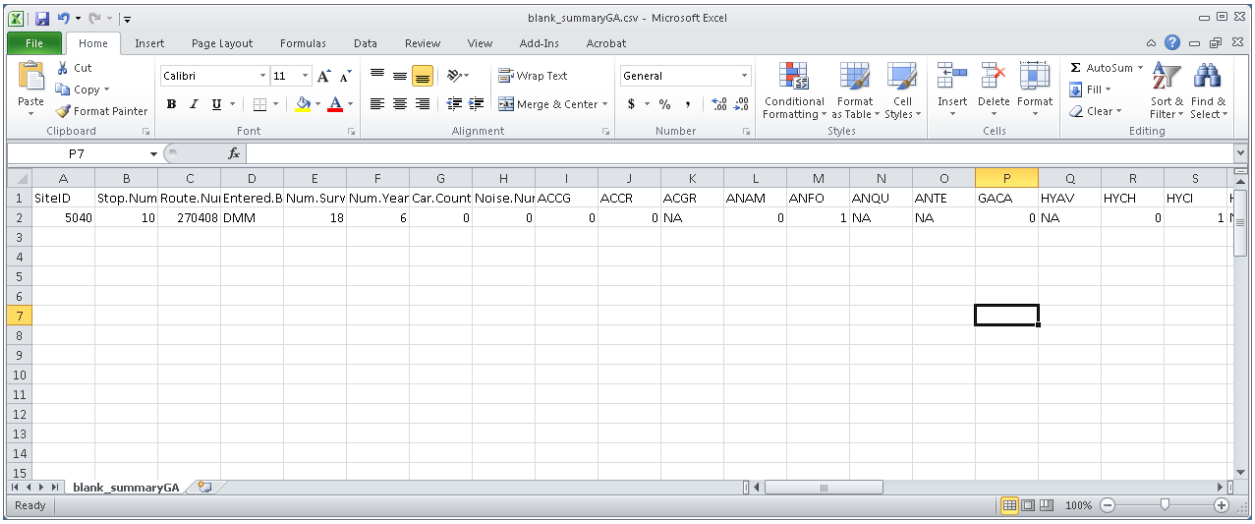
ONE CAUTION – IF YOU WANT TO USE THE BUILT-IN AVERAGE FUNCTION IN EXCEL, YOU'll NEED TO ADD ZEROS BEFORE AVERAGING. OTHERWISE, EXCEL WILL JUST AVERAGE THE NON-ZERO VALUES AND YOU'LL GET THE WRONG ANSWER. IF YOU DON'T WANT TO ADD ZEROS, YOU CAN SUM DOWN

THE ROW INSTEAD OF AVERAGING AND THEN HAVE EXCEL DIVIDE THE SUM BY THE TOTAL NUMBER OF SURVEYS OF THE ROUTE/STOP.

To give another example, for the data column below, the two columns would be called “HYGR” and “HYGR.mean”. The values entered into each of these would be “1” (since the species was heard in at least one survey), and “0.25” (since the mean calling index is 3/12 = 0.25)

Hyla grati
0
0
0
0
0
0
0
1
0
0
0
2

Your final data sheet should look something like this (this is for 270408):



Congratulations – you’ve now summarized the frog and toad calling data. Now move on to the protocols for GIS and compiling the land use data.