**READ ME** for files created by Neptune & Company, Inc., September, 2016

(1) You will have to set your working directory at the beginning of most of the files.

Eg) setwd("~/repos/Phytoplankton-DataAnalysis/gitAnalysisFiles/analysis")

(2) If the big algae file is updated, you should start by re-running the code chunks in **Phytoplankton\_DataAnalysis.Rmd**. This code will create .Rdata files that are used in other .Rmd files. The Rdata files will then be loaded at the beginning of other .Rmd files.

(3) Because there were field duplicates in the first few data sets we were working with, there are still response variable names to indicate this and they are denoted with and \_Ravg at the end, or an \_Rmax at the end indicating whether it was created by averaging the field duplicates or taking the maximum. Now that there are no duplicates (and assuming that continues to be the case) it does not matter which one you use because they will be the same.

(3) Here is a list of the other files and a very brief description of each, though the file names are meant to be informative and these are not in any particular order. (Yellow highlighting means I still need to look through them one more time to possibly take out some code that is not needed. Orange means I haven’t looked through them recently at all.)

* **ExploratoryPlotsForOneLake.Rmd** - plots to make for a single lake where you tell the file what reservoir you are interested in at the beginning
* **SummerMaxInvestigation.Rmd** – Creates plots to explore the use of the summer maximum for total cell density and these also include the horizontal lines for the EPA health
* **SummerMaxInvestigation\_biovolume.Rmd –**Same as SummerMaxInvesigation.Rmd, but for biovolume instead of cell density. The two could easily be combined because they load the same data sets and use essentially the same plotting code (just response variable name differs).
* **CompareMeasures.Rmd –** Compares total cell density, total biovolume, and proportions for all taxa, cyanobacteria, and toxin producing. These plots provide a good check of the cyanobacteria data because toxin producing should be a subset of cyanobacteria.
* **SeasonalityInvestigation.Rmd** – Looks at seasonality within years. Still needs to be updated to use the most recent BGtotals\_maxsummer.Rdata file, but does work fine as is (just does work that we now do in when the maxsummer data file is created).
* **TemperatureInvestigation.Rmd –** Imports the temperature data file because it is not merged with the larger algae file. The focus is on the exploratory plots and looking if temperature can help account for different months sampled in different years. This file can also be the start to looking at other relationships with temperature and calculating other summary measures for temperature that might be meaningful covariates
* **LinearVsLoess.Rmd –** Creates plots comparing linear and loess (nonparametric smoother) for all reservoirs. It compares for total cell density, total biovolume, and the relative measures and focuses just on summer data (though not the summer max). Should add use of just the summer max to this file, or could just do this comparison for the summer max within another file. NOTE: The degree of smoothing is controlled by the span=1.5 in the geom\_smooth() function within the ggplot code. The shape of the line will depend on this parameter. I chose one that looked reasonable over many lakes.
* **AnalysisForOneLake.Rmd –** Goes through the plots, linear regression analysis , and checking of assumptions for a single lake using the summer maximum as the response variable and year as the
* **AnalysisForMultipleLakes.Rmd –** Goes through different methods of analysis for combining lakes to get an estimate of an overall slope (linear trend over years). Looks at multiple linear regression, as well as different ways of fitting varying intercepts and varying slopes models. I suggest getting Gelman and Hill (2007) as a great reference for multilevel models (includes R code). Zurr et al. is also useful, though I think they put too much emphasis on making analysis decisions based on AIC. I can help look through this again when you start playing with it.
* **ChangePointAnalysis.Rmd –** Just to provide a starting place for R packages and help files. Does not go through a whole analysis.
* **EFRanalysis.Rmd- (remove?)**
* **IntroAlgaePlots.Rmd (remove?)**
* **IntroAlgaePlots2.Rmd (remove?)**
* **PlotsSelectReservoirs.Rmd (remove?)**
* **NotesAboutAnalysisIdeas.Rmd –** [NEED TO READ THROUGH AND ADD SOME THINGS] This does not include code, but bullets about thoughts I have had about the analysis and future analyses. This is not meant to be formal, but just to share my thoughts. Feel free to ask any questions about it.