**Step 1) Data analysis**

**a) compiling data**

This is your starting point: have all you data in a format that you can use to plot. You may have many different data sources, e.g. solid phase, dissolved, gaseous. I prefer to have everything in one big spreadsheet with the meta data (data on data, such as collection time and date, comments etc) and IDs in several columns. If you use replicates, consider calculation the average and standard deviations in another tab. If you wonder how to organize this, consider how excel (or most other programs) work. Typically you type in a command one time and then have excel apply this command to the rest of the data automatically. You don’t want to get in the way of this, otherwise this means you are typing in commands for hours…….

Therefore put some thought in data organization, never work of the original (copy it and play around with the copy). All this may take some time to set up but will make your life easier.

Label everything with ID and unit, rule of thumb is that another person should be able to make sense of you spreadsheet without your help.

**b) exploring and looking for patterns**

This part of the data analysis is important, it gives us our starting point. Have you hypotheses in mind and begin to make plots that contain data that can test your hypotheses. The type of visualization depends on the type of questions you want to answer.

For example, if you hypothesize that Sample A should have a higher carbon content as sample B (for whatever reason) a column plots could be a good choice because it allows for direct comparison.

Another approach is to look for correlations. You can generate a correlation matrix using jmp or EXCEL (you need the data add on). To learn how to do this please google it and follow one of the tutorials. Its actually quite simple.

**c) patterns and meaning**

This step makes use of the information (from step b) where you looked for patterns. Always keeping your hypotheses in mind (or you objectives or questions) select plots that are important. For example one plot could contain a time series (Carbon decreasing over time), another could be a correlation (as C decreased in the gas phase it also decreases in the aqueous phase) or similar. When we discuss data I will ask you questions about this (e.g. does nitrogen follow the same trend?) and we may try new ways to plot the data. For this know your data (spend time plotting and thinking about what this means) and have the spreadsheet organized.

When you have a plot you like, check it meaningfulness. E.g. you have a plot that shows pH decreasing when CO2 increases. Is this meaningful for your study or does this simply confirm something that we (humans) already knew?

Or you see several correlations, is this meaningful or is this simply an autocorrelation (if A drives B and B correlates with C, could it be that it is simply that A drives C as well?).

Also, plot data from several sources on the sample plot (e.g. CO2, DOC and DOM).

**d) drafting the results section**

I like to have the best plots (that may make it into the paper) copy pasted on a ppt and annotated. E.g. you have a plot and underneath you write “correlation between x and y for montmorillonite suggests suppression of microbial activity, in agreement with hypothesis 1”. This makes discussing data much easier.

Importantly, the result section cannot become meaningful without knowledge of the literature. For example if you find an interesting correlation, hit the books (and peer reviewed publications) to check if such observations have been made before. Add a comment on your ppt (e.g. in agreement with findings of Eimers et al. 2013, found increasing DOC with decreasing Ca).