Homework 1: Data & Data Management

1) (15 pts) Write a brief data management plan for your research. This can be for research you have done, or are planning to do. Use https://dmp.cdlib.org/ to help you write the plan, answering the questions listed. For some examples, see http://www.dataone.org/data-management-planning and http://www.dataone.org/best-practices.

2) (10 pts) You are co-ordinating a team of biologists sampling ponds around the state to look at the impact of food (as measured by Chlorophyll A) and turbidity (measured as Secchi Depth) on zooplankton abundance (measured as individuals per milliliter). You have the data. It's time to analyze it. But first, it's time to Q.C. it. So load it in (oh, and did I mention it's space delimited – see read.table!) and get fixing.

        a) Read the data in, and note which values are missing, if any.

        b) One of your team reported that their instrument may have been failing and produced anomalous Chlorophyll readings. Fix this.

        c) One of your team misread your meta-data, and entered zooplankton values in the wrong concentrations – numbers per liter instead of milliliter. So, they're off by a factor of 1000. Fix this.

        d) Extra credit (5 pts more): One of your team members thought their site was called "EE" instead of "E". Sure, you could find and replace this in the original data, but who knows what you might accidentally mess up.   Change the names back to "E" for those anomalous sites. Note that this is going to lead you down the path of factors versus numeric vectors, which we'll touch on in the coming weeks. Look at the help pages for the following functions: as.factor, as.numeric, levels, as.character. Good luck!