**Martone Lab – UBC – Hakai Rocky Shore Seaweed Surveys**

**Protocol for preparing tidy data for manipulation and analysis**

How to prepare raw data (as entered into Excel from field sheets) for import into R for combining data for all years and sites (i.e. every tab of every Excel file).

For more information about tidy data see <https://en.wikipedia.org/wiki/Tidy_data> and <https://cran.r-project.org/web/packages/tidyr/vignettes/tidy-data.html>

**Basic Steps:**

1. Make tab names consistent: Same site names (e.g. West Beach) and zone names (e.g. HIGH) for every tab. The only thing that should differ is the year (e.g. “West Beach HIGH 2017”).
2. Control the dimensions of the data
   1. Delete any columns after column K. This is usually a data summary appended as new columns. Delete several columns after the data summaries to control for errant data entry.
   2. Delete any blank rows. Often seaweed and invertebrate data are separated by a blank row (usually highlighted in grey). These will mess up the data input.
   3. Resolve herbivore data into rows. This is difficult in many cases because the data are somewhat ambiguous as entered. However, this must be done to keep the data in a consistent format.
   4. Try not to change species names in the raw files. They will be cross referenced using the full species key spreadsheet.