**Benthic REA - diver vs. diver comparisons**

*M. Ferguson 5/22/2018*

Required data files:

* ALL\_REA\_ADULTCORAL\_RAW.rdata
* ALL\_REA\_JUVCORAL\_RAW.rdata

Required function files:

* Benthic\_functions.R
* core\_functions.R

Script file:

- REA Benthic Diver Comparison.R

Currently, code is developed to conduct inter-diver comparisons for the following benthic summary metrics:

1) Coral colony density (adults and juvs)

2) Coral colony size (adults and juvs)

3) % old dead

4) % recent dead

5) Bleaching prevalence

6) Total disease prevalence

7) Acute disease prevalence

8) Chronic disease prevalence

9) COTS prevalence

Metrics are generated at the segment-level (base level) in order to make comparisons between divers within a site (it’s best to do diver comparisons within a site, not between sites). The default pooling field for metrics generated is at the Genus-level (however, this can be changed to the desired pooling field, e.g. species-level). Note that each function generates metrics for total scleractinians (SSSS). The ‘diver vs diver’ function and code is currently designed to use metrics generated for total scleractinians, however, additional size comparisons between divers can be made for a given species and/or morphology of interest (see code at the end of ‘REA Benthic Diver Comparison.R’ script). The ‘diver vs diver’ function generates/saves one .png file, displaying a plot for each benthic summary metric.

Note that for each plot, the x-axis limits are set by the data provided for each plot. Consequently, if there are very large outliers present, this may yield large x-axis limits, which may make it difficult to visualize differences between divers. For better visualization for a given plot, add ‘x\_range’ within the function argument list to manipulate the x-axis limits to the desired limit range (e.g. divervsdiver <- function(data, date1, date2, date3, x\_range).

*Summary of inter-diver comparisons*

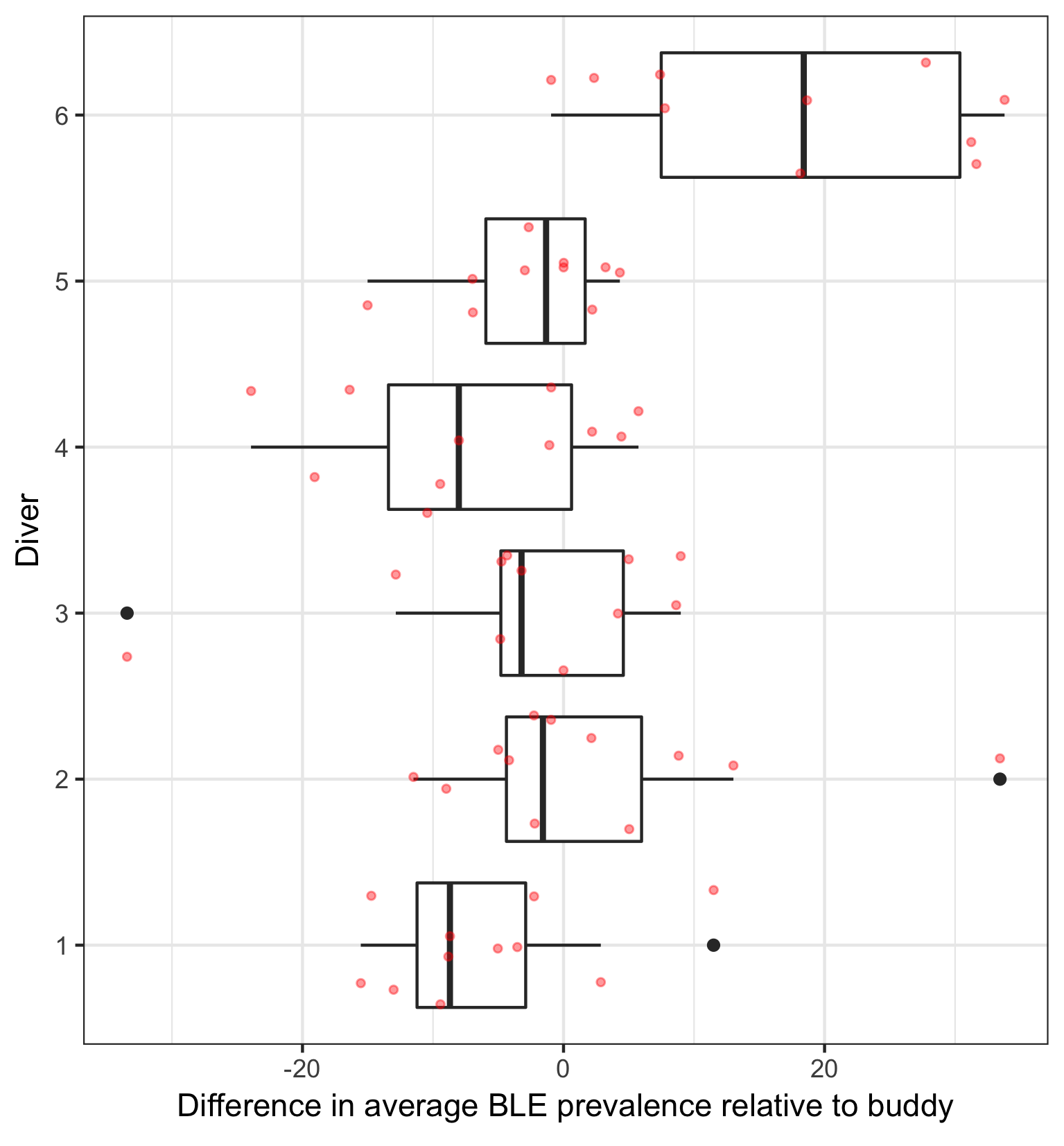
Benthic REA inter-diver comparisons are based off of fish REA inter-diver comparisons. Dive teams can dive together for several days but it is good practice to switch dive teams during each leg for variability. When generating comparisons, it’s better to have a higher sample size (~8-10 sites per team = ~ 2-3 day period) in order to generate more robust comparisons.

The following are assumptions made when generating comparisons:

* Diver estimates should be fairly similar at a site (i.e. - segments within a site vary some but don’t vary drastically)
* On average, a diver’s estimates should be higher than their buddy(ies) about half the time and lower about half the time

Inter-diver comparisons are made using boxplots, which are useful for visually comparing the “centers” and “spreads” of data (here, benthic summary metrics between divers). Quick summary of boxplots: The line in the middle of the box is the median. The box itself represents the middle 50% of the data. The box edges are the 25th and 75th percentiles. The vertical size of the boxes is the interquartile range, or IQR. They measure the spread of the data, similar to standard deviation. The whiskers represent the reasonable extremes of the data. These are the minimum and maximum values that do not exceed a certain distance from the middle 50% of the data (by default in R, the distance = 1.5×IQR). Outliers are the black data points outside of the whiskers (note that there will inherently be outliers for each diver in each comparative plot). Red ‘jittered’ dots on each plot are the average estimate for each diver at a site.

For each boxplot inter-diver comparison, we look at the median and spread of a diver’s counts relative to their buddy, with the aim that each diver’s estimates are centered on zero, and that a diver’s estimates are not too widely spread.



In the example plot above for bleaching prevalence (estimates generated after 3 days of diving), we see that most divers’ estimates are closely centered around zero, except for diver 6, whose estimates are much higher than the other divers. Diver 6’s estimates are also more widely spread than the other divers. Diver 1 and 4 are also slightly lower than the other divers but the estimates are not widely spread. From this plot, we would not conclude that estimates for Divers 1, 4, and 6 are inaccurate…instead we would keep track of these divers’ estimates and run comparisons as surveys continue over the course of the cruise leg. If diver 6’s estimates continue to be higher than the other divers (> 50% of the time) throughout the cruise, than this is cause for concern. Similarly, if diver 1 and 4’s estimates are continually lower than the other divers (> 50% of the time) throughout the cruise, than this is also cause for concern. This may prompt further investigation of bleaching data for Divers 1,4, and 6 in order to course-correct how these divers are collecting data.

The ultimate idea of these inter-diver comparative plots is to use them as a tool to visualize differences in estimates between divers. The key point is not that you expect divers to be identical, but ***they should be higher than their buddy(ies) half the time and lower half the time***. The comparisons aren’t meant to determine which individual or set of diver estimates are “right” or “wrong” but, rather, the comparisons can allow us to see trends or anomalies in a diver’s data. Based off anomalies you see in a diver’s data after running comparisons, its good practice to investigate that diver’s data to see what may be causing the anomaly in their data (as in the example given above for diver estimates of bleaching prevalence). As an additional example, inter-diver comparisons may show that Diver 1’s adult colony density estimates are consistently higher than their buddy(ies) (> 50% of the time)…this could indicate that Diver 1 is seeing more colonies than there actually are. This can happen when it’s difficult to tell where a colony begins and ends (i.e. - when large colonies are separated by bits of CCA, have large sections of old dead, etc.; encrusting Montipora species can be a problem child). These can be recorded as being several different colonies when they are actually the same colony. The idea is to catch these errors earlier rather than later in order to correct how divers’ collect data.