UBC Okanagan (Biology): Guidelines for instructors and students on how to present tables and figures

These guidelines are based on current "best practices" in Biology (and science in general)[1]. They aim to achieve consistency among faculty, instructors, and students in how data are summarized and presented within lab reports and research papers.

1. Tables

Table 1. Summary of trait measurements made on individuals of *Solidago* spp. collected within shaded and open habitats in the vicinity of Portland, Oregon.

	Habitat: shaded (n = 20)		Habitat: open (n† = 18)	
Trait	Mean (sd)	95% Confidence limits	Mean (sd)	95% Confidence limits
Leaf area (cm^2)	4.59 (0.974)	4.14, 5.05	4.54 (0.972)	4.24, 5.15
Leaf mass (mg)	2.52 (0.765)	2.15, 2.89	2.62 (0.705)	2.25, 2.99
Root mass (mg)	9.97 (2.754)	8.67, 11.26	9.90 (2.454)	8.37, 11.16

† data for two individuals misplaced

- Heading is above the table (this varies among disciplines)
- Table should be interpretable on its own thanks to an informative heading and judicious use of footnotes
- Sample sizes and units always included
- Horizontal lines only; typically only occurring above and below headings, and at bottom of table

2. Descriptive / summary statistics

- Round numbers to one more digit for measures of centre (e.g. mean), and 2 more digits for measures of spread (e.g. sd) than was used in measuring the data

For detailed guidelines about significant digits, consult the following webpage:
http://www.astm.org/SNEWS/SO_2008/datapoints_so08.html

- Units are preceded by a space within text passages: "Average height was 34.2 cm (± 3.43 SEM)."

Describing numerical variables

- Report mean with standard deviation, and additionally median with inter-quartile range for variables that exhibit a non-normal frequency distribution (e.g. is skewed) or that includes outliers
- Parameter estimates should be accompanied by measures of uncertainty, i.e. the *standard error of the mean* (SEM) or confidence *interval* (notation: lower limit – upper limit); confidence *limits*: (lower limit, upper limit)
- Confidence intervals are strongly encouraged because they inform about *effect size*
- Measures of uncertainty for an estimate such as SEM can be preceded by a ± sign; do not make the common mistake of reporting a ± sign with a standard deviation, as it is not a measure of uncertainty in an estimate

Describing categorical variables

- Report a frequency table (raw data) or a summary table with proportions for categories (the main descriptive statistic of interest), along with the confidence interval for the proportion if appropriate

3. Results of statistical tests

- Test statistics (e.g. Student's t or an "F" from ANOVA) should be rounded to 2 decimal places, and associated P-values should report 3 decimal places, or if smaller than 0.001, then <0.001. P-values do not indicate effect size, so reporting $P = 10^{-6}$ is not more impressive than $P < 0.001$
- Concluding statements should, in the absence of a table, include the test, test statistic value, degrees of freedom (df) or sample sizes, P-value, and confidence interval (if appropriate) in parentheses: "On average, hair loss was significantly greater among fathers compared to childless men (Welch's 2-sample t-test; $t = 4.23$; $n_F = 18$, $n_C = 20$; $P = 0.018$; 95% CI for difference: 9.34 – 18.22%)". Your *Methods* section should clearly state the significance level (α), and this should be decided prior to the study
- Regression and ANOVA results should be shown in a standard ANOVA table format[2]

[1] http://www.nature.com/collections/qghhqm/pointsofsignificance and
http://blogs.nature.com/methagora/2013/07/data-visualization-points-of-view.html

[2] http://www.biostathandbook.com/onewayanova.html (this site is an excellent resource for many topics)

4. Figures

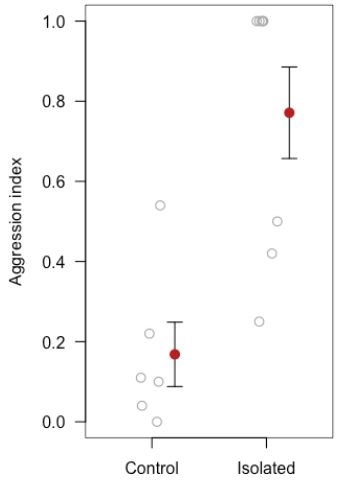

Figure 1. Aggression was significantly higher among isolated ants (n = 8) compared to the control group (n = 6) (see text for details). Shown are individual observations (grey circles), group means (solid circles) with +/- 1 SEM.

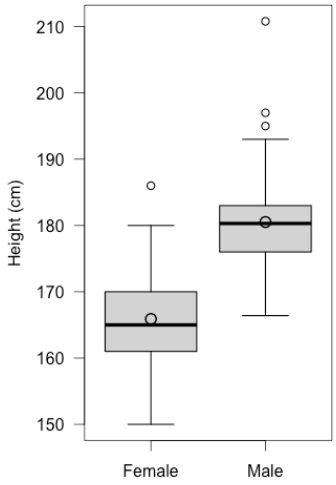

Figure 2. Height of male (n = 64) and female (n = 90) students within BIOL202. Thick horizontal lines represent group medians, large circles represent group means, boxes delimit 1st to 3rd quartiles, whiskers extend to 1.5 x IQR, and small circles represent extreme observations.

- Figure heading appears below the graph, and should enable the figure to be interpreted on its own
- Heading can include statistical statements (Fig. 1), or simply describe what is being shown (Fig. 2)
- Sample size(s) reported, and the first time a particular type of graph is shown (e.g. boxplot), details of graph features must be provided. Subsequent figures of the same type can refer to the first for details.
- Use hollow symbols so that overlapping points can be seen (Fig. 3)
- Orient all text horizontal (except y-axis label), including all tick labels
- Place axis tick marks outside of figure border to avoid overlapping with observations

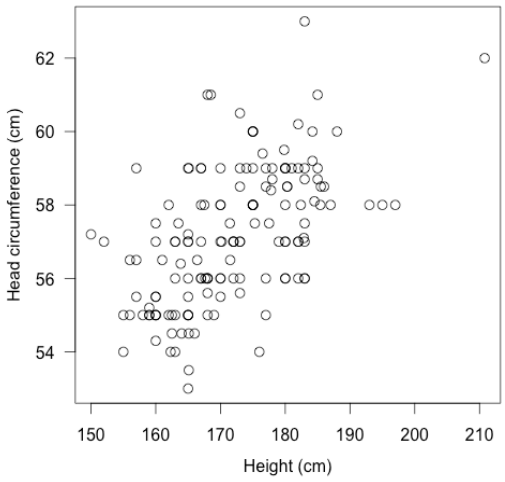

Figure 3. Head circumference versus height for n = 150 BIOL202 students. The positive association is highly significant (Pearson r = 0.82; P < 0.001).

- Data points should not touch axes
- Fitted lines (e.g. least-squares regression) included in figures should be fully explained in the heading, e.g. "Line represents a least-squares linear regression line, $y = 0.3 + 4.5x$ (F = 5.65, df = 36; P = 0.021)".
- For more complex statistics (e.g. lines associated with mixed effects models) refer the reader to the text for details

- Bar plots should **only** be used to visualize categorical data (e.g. proportion of students with brown or blue eyes), or counts (number of flies on scat)[3]

- When comparing numerical data among categories or groups (Figs. 1,2) use "stripcharts" (Fig. 1) when sample sizes are small (i.e. <20) and boxplots otherwise (Fig. 2)[1]

[3] http://www.nature.com/nmeth/journal/v11/n2/full/nmeth.2807.html