*ggTukey: A Visual Analogue for Tukey's Range Test in R*

Carson Keeter
Department of Statistics
University of Wyoming
May 2019

Imagine that you have a series of data regarding mice and the amount of time spent on a running wheel (much like the one that powers my brain). The purpose of your study is to see when mice start running longer distances, say, after being brought from sea-level to 7,000ft in elevation. After collecting 15 days' worth of data, you would most likely utilize an *Analysis of Variance* (*ANOVA*, for short) to determine if there is a "main effect" of the variables contribution (*time*, in this case) to the outcome (*running distance*). The ANOVA results in a p-value less than your chosen alpha (for convention's sake we'll say $\alpha = 0.05$). But you're not done.

The *ANOVA* gives you an answer to your question if your question was in a 'yes or no' format. In your case, "yes" the *ANOVA* says, "as time goes on, there is *some kind of* change in running distance". "But when does that change happen?" you ask. "Not my job" says the *ANOVA*. Thankfully, Tukey has the solution.

*Tukey's Range Test* (commonly known as *Tukey's Honest Significant Difference Test*, *Tukey's HSD* for short) was developed by John Tukey to mediate this issue. *Tukey's HSD* makes a comparison between every combination of variables (days, in our case) using the *Studentized Range Distribution.* Essentially, it's a *T-test* with a family-wise error correction. The arcana can be seen here:

$$q_s = \frac{Y_A + Y_B}{SE}$$

*&*

$$q = \frac{\overline{Y}_{\max} - \overline{Y}_{min}}{S\sqrt{2/n}}$$

where,

$$q_s > q_\alpha \; connotes \; statistical \; significance$$

Thankfully, we don't have to do any of that by hand. In *R,* the function *TukeyHSD* does the pairwise comparison and reports back p-values for all the combinations of variables. But what if you have a fairly large number of comparisons you want to make? The output of *TukeyHSD* can get fairly onerous to sort through with large comparison, and there's no easy visualization that can describe the significant differences. UNTIL NOW!

The package *ggTukey* takes your *TukeyHSD* object and projects the data onto a simple x-y dot plot. Dots filled in with black signify statistical significance while white dots signify the opposite. Two functions within the package deal with numeric factors (*ggTukey_num*) and character factors (*ggTukey_char*). Below is an example using our previously given example:

Using the function *ggTukey_num*():

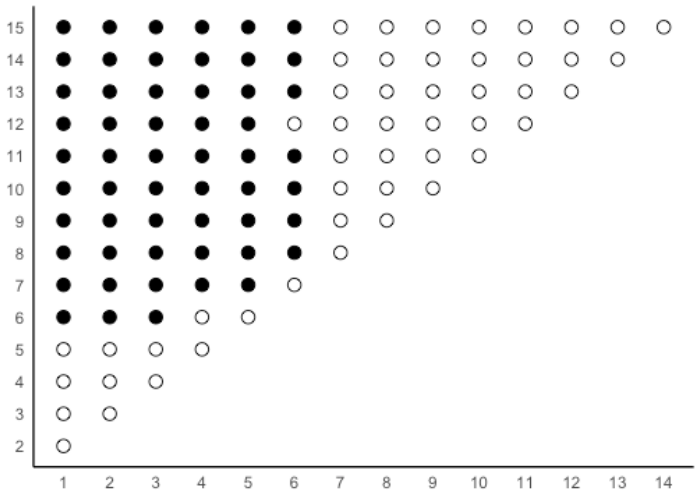


*Figure 1: Day-to-day comparison of running distance in mice during altitude acclimatization (black: $p < 0.05$; white: $p > 0.05$)*

In the current version (0.1.0), you are able to choose your own alpha (convention, again, being $\alpha = 0.05$ but you could choose a smaller $\alpha$ if your boss is a perfectionist). Also, this current version is able to handle day-to-day comparison as well as group comparison (if you wanted to compare differences between mice, for instance).

Using *ggTukey_char()*:

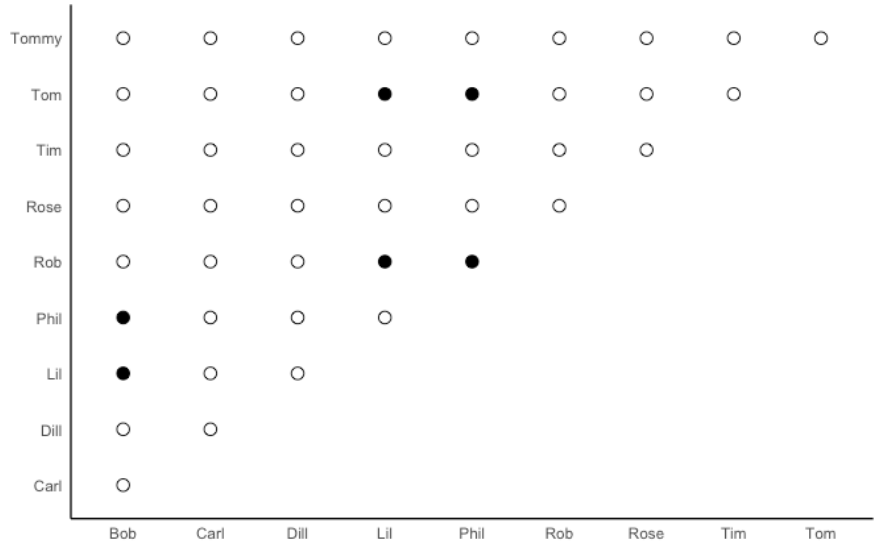


*Figure 2: Person-to-person comparison of something, not sure what (black: $p < 0.05$; white: $p > 0.05$)*

Example code can be found in the documentation, however the basic idea is creating an ANOVA object, creating a TukeyHSD object, then adding that TukeyHSD as an argument in either *ggTukey_num(object, a.pri = 0.05)* or *ggTukey_char(object, a.pri = 0.05)*.

Given that this is the first version of this package, inevitably there will be a few situations where this package doesn't work properly. Please contact Carson Keeter (keeterc1@gmail.com) when those situations arise. If possible, please send your code along with the purpose of your project and the issue will be recorded and fixed in future versions.

Carson Keeter