Keshanti Tidwell

10/16/17

Bioinformatics

Professor Kate Hertweck

R Project Questions

1. Figure 1 is quantitative data. I chose to use a boxplot because it is easy to see the minimum, median, and maximum wing length from both species, Bombus balteatus and Bombus Sylvicola. Based on the visualization, it is apparent that species Bombus balteatus has longer wings than Bombus Sylvicola. This could mean that they are better pollinators.

Figure 2 is quantitative data. I chose to use a scatterplot to depict the relationship between the caste of bees and their buzz frequency. From this figure it seems that worker bees have a higher buzz frequency than queens, which makes sense because they are the ones that have the primary job of pollinating plants, while the queen’s primary job is to make more bees.

Figure 3 is quantitative data as well and I used a boxplot again, because of how easy it is to see the min, median, and max values of buzz frequency amongst each of the species. From this figure you can see that Bombus sylvicola has a higher buzz frequency than Bombus balteatus, and that helps put the previous graph into perspective. Bombus sylvicola’s wings are shorter than Bombus balteatus which allows them to move their wings faster, hence the high frequency. So in actuality, Bombus sylvicola would have the advantage in pollination.

I ran an ANOVA statistical test to find out if the relationship that was shown between species and buzz frequency in figure 3, was significant. The p-value that was generated (9.05e-06) shows that it is incredibly significant.

1. Citations:

* R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
* RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>.
* Hadley Wickham, Romain Francois, Lionel Henry and Kirill Müller (2017). dplyr: A Grammar of DataManipulation. R package version 0.7.4.https://CRAN.R-project.org/package=dplyr
* R Special Interest Group on Databases (R-SIG-DB), Hadley Wickham and Kirill Müller (2017). DBI: R Database Interface. R package version 0.7. <https://CRAN.R-project.org/package=DBI>
* Hadley Wickham (2016). lazyeval: Lazy (Non-Standard) Evaluation. R package version 0.2.0. <https://CRAN.R-project.org/package=lazyeval>
* <ggplot citation goes here>
* H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

1. To improve interpretability and reuse of the data I analyzed I could have shared my data early in my “workflow” so that I could have shared my results and interpretation with Professor Hertweck before I wrote my analyses. That probably would have made them better.
2. I failed to implement Rule 2, which is “Developing a Prototype First.” Before writing any code, I should have clarified what I was trying to implement and everything that I needed to do. Instead I just coded as I went, which proved to be unproductive because I had so many questions as I went along because I did not have a full understanding of what I was doing. One rule that I think I did follow well is Rule 6, which instructs to “Use Pictures” to validate results. I used boxplots to turn the data I used into a visual component for better understanding