Graphing Grouped Continuous Data Lesson Plan

Marney Pratt, Smith College

11/11/2019

Focus:

This lesson helps students know some of the options for how to graph grouped continuous data (such as those involved in doing a t-test or ANOVA) and how to choose the best option.

Overview:

Graphing grouped continuous data is very common in Biology. For example, comparing the growth rate of caterpillars fed two different diets or the height of trees grown in drought compared to well-watered conditions. However, common graph types used can make it impossible for a reader to critically evaluate and interpret the data and accompanying conclusions. This lesson demonstrates how to make several different types of graphs that are more effective at allowing the readers to critically evaluate the data. At the end of the lesson, you will choose what you think is the best graph to represent the data to turn in to your instructor.

Learning objectives:

After this lesson, you will

- Understand the definition of variable types, normal distribution, and some basic descriptive statistics (mean, standard deviation, median, interquartile range)
- Know how to use some clues from exploring your data to assess whether your data are normally distributed
- Learn about different graph options for situations with a continuous dependent variable and a categorical independent variable
- Be able to choose the best graph for situations with a continuous dependent variable and a categorical independent variable

Lesson sequence:

1. Introduce the data set and desire to visualize a numerical dependent variable (insect density) as a function of a categorical independent variable (forest location)
2. Define variable types
    - Ask questions about variable types to test understanding
3. Make a bar graph of mean and SE (code is given)
    - Ask questions about the graph (what is the sample size? How many points are above the top of the bar? is the sample distribution normal?)
    - Discuss problems with bar graphs showing mean & SE, point to Weissgerber et al 2019 for more information
4. Introduce how to use histogram to assess if normal (but mention this is useful only when n>30 or so)
5. Decide if data are normally distributed
6. Introduce boxplot (code is given)

- Compare boxplot to histogram
- add different options for boxplot (adding points, adding mean, outliers, etc)

7. Show violin plot (code is given) and compare to other graphing options
8. Introduce dot plots with mean or median and why it is best to show all the data with a measure of central tendency when sample size is small
9. Have student choose the best graph to present the data and turn this in outside of swirl (using lab notebook, learning management software, paper, Google Form, etc)

Pre-lesson activities:

Before completing this lesson, students should know how to do some basic things in RStudio including how to look at and understand the structure of a data frame, be able to filter and summarize data using dplyr, and they should understand the basic "grammar of graphics" used in ggplot. It is also helpful to be familiar with variable types, normal distributions, descriptive statistics (especially mean, standard deviation, median, and interquartile range), and what is represented in a histogram and boxplot.

Post-lesson activities:

At the end of the lesson, students will choose the best graph to represent the data, save the graph outside of R, and then turn the graph in by method of choice (paper printout, learning management system, lab notebook, etc)