**Kylie Blake**

All my answers are in purple

**Introduction to visualization in R – 25 pts**

PLEASE READ THIS BEFORE CONTINUING

This assignment is designed to help you visualize data in R and practice your R coding skills. You may collaborate with a partner to enhance your learning experience. Please ensure the following:

* Collaboration: If you worked with a partner, include their name.
* Submission: Only one person should submit the assignment.
* Answers: Provide written answers to all questions in the current document or a rendered markdown file.
* Code Submission: Submit your R code in an .R or .Rmd file by sharing a link to your GitHub repository.

Partner 1: NA

Partner 2: NA

**Data description**

Here, we will analyze mycotoxin data collected from wheat heads grown in the greenhouse inoculated with *Fusarium graminearum* (Fg) with or without fungal endophytes to protect against *Fg*. The experiment was performed by inoculating the heads with Fg (positive control), Fg + 1 of 3 endophytes (Fg + 40, Fg + 70, or Fg + 37), and the non-treated control (NTC). Therefore, the experiment was five treatments with 6-10 technical replicates and three experimental replicates. The authors quantified the parts per million (ppm) of mycotoxins Deoxynivalenol (DON) and the variant 15ADON from the wheat grain, and this is your primary response variable. In other words, we are interested in knowing if the fungal endophytes could reduce the concentration of DON on wheat heads. This was done across two wheat cultivars, a winter wheat variety, Ambassador, and a spring wheat variety, Wheaton. These data are published here:

* Noel, Z.A., Roze, L.V., Breunig, M., Trail, F. 2022. Endophytic fungi as promising biocontrol agent to protect wheat from *Fusarium graminearum* head blight. *Plant Disease.* <https://doi.org/10.1094/PDIS-06-21-1253-RE>
* The following github repository is associated with this paper: <https://github.com/noelzach/EndophyteBiocontrol>

1. **5 pts. Explain the following concepts about ggplot** 
   1. What three elements do you need to produce a ggplot?
      1. Data frame, aesthetics (x and y-variables, color,etc), and geom function to call in which type of figure you’re interested in making (scatterplot, line graph, boxplot,etc).
   2. What is a geom?
      1. Geom is function in ggplot2 that allows you to create a visual representation of the data frame you are working with in the figure of your choosing (box plot, line graph, etc.)
   3. What is a facet?
      1. Facet is a function in ggplot2 that allows you to plot your data into multiple panels in one figure.
   4. Explain the concept of layering.
      1. To make a figure in ggplot, geom functions are called in as layers. Therefore, once you establish the data frame and aesthetics of a figure, the way in which you call in the geoms, such as geom\_point() and geom\_bar(), will determine which function is executed first. This can impact your figure visually depending on the order of functions.
   5. Where do you add x and y variables and map different shapes, colors, and other attributes to the data?
      1. The x and y variables and other attributes are added in the aes() function when you first call in the data frame to ggplot.
2. 4 pts. Make a boxplot using ggplot with DON as the y variable, treatment as the x variable, and color mapped to the wheat cultivar. Show the code you use to load the libraries you need to read in the data and make the plot. Change the y label to “DON (ppm)” and make the x label blank.

See R code

1. 2 pts. Now convert this data into a bar chart with standard-error error bars using the stat\_summary() command.

See R code

1. 2 pts. Add points to the foreground of the boxplot and bar chart you made in question 3 that show the distribution of points over the boxplots. Set the shape = 21 and the outline color black (hint: use jitter\_dodge).

See R code

1. 2 pts. Change the fill color of the points and boxplots to match some colors in the following colorblind pallet.

cbbPalette <- c("#000000", "#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00", "#CC79A7")

See R Code

1. 2 pts. Add a facet to the plots based on cultivar.

See R Code

1. 2 pts. Add transparency to the points so you can still see the boxplot or bar in the background.

See R Code

1. 2 pts. Explore one other way to represent the same data <https://ggplot2.tidyverse.org/reference/> . Plot them and show the code here. Which one would you choose to represent your data and why?

I selected the violin plot because it retains the key features of a traditional box plot while also displaying the data distribution density. This figure helps the audience see how the concentrations of mycotoxins Deoxynivalenol (DON) differ in the winter and spring wheat varieties across treatments, as well as across samples. This way of visualizing the treatments also helps the intended audience understand where 1) DON was highest and 2) which endophyte treatments were more successful at reducing DON concentrations. I also changed the DON concentrations to a log10 scale since the data appeared too small on the given scale.

A screenshot of a computer program

AI-generated content may be incorrect.

1. 4 pts. Annotate your code and push it to github. Now, find a partner if you don’t already have one and have that person “fork” the repository containing the code you just pushed to github. Describe what just happened when you “forked” the repository. Please include the links (URLs) below to answer this question.

I forked Sam Donahoo’s class repository here: <https://github.com/kzb0180/PLPA6820_SP25>. When I forked a repository I made a copy of hers that will not impact or modify any of her original files in any way. I believe it also shows this as a branch of her repository now.