**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: Anet Filipova

Partner 2: Brynleigh Payne

**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?

**Dataframe, aesthetic mappings and layers.**

* 1. What is a geom?

**A layer**

* 1. What is a facet?

**Separate plots for each set of data**

* 1. Explain the concept of layering.

**We put layers with the function geom() and it is basically adding elements one by one to the plot (e.g. a background, data points, etc.)**

* 1. Where do you add x and y variables and map different shapes, colors, and other attributes to the data?

**We add the x and y variables in the main aesthetic mapping using the function aes() directly after you specify your dataframe. You can map different shapes, colors and other attributes either there or to each additional layer afterwards.**

1. 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 th e plot. Change the y label to “DON (ppm)” and make the x label blank.
2. 2 pts. Now convert this data into a bar chart with standard-error error bars using the stat\_summary() command.
3. 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).
4. 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")

1. 2 pts. Add a facet to the plots based on cultivar.
2. 2 pts. Add transparency to the points so you can still see the boxplot or bar in the background.
3. 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?

We chose a violin plot because it better represents the distribution of the each data point and also the density.

**ggplot(myco, aes(x = Treatment, y = DON, fill = Cultivar)) +**

**stat\_summary(fun = mean, position = "dodge") +**

**stat\_summary(fun.data = mean\_se, position = "dodge") +**

**xlab("") +**

**ylab("DON (ppm)") +**

**geom\_violin(trim=FALSE, position = position\_dodge(width = 0.9), scale = "width", adjust = 0.6) +**

**geom\_point(color = "black", position = position\_jitterdodge(dodge.width = 0.9), shape = 21, alpha = 0.5) +**

**scale\_fill\_manual(values = cbbPalette) +**

**facet\_wrap(~Cultivar, scale = "free")**

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.