**Introduction to visualization in R – 50 pts**

This assignment is intended to introduce you to visualizing your data in R, as well as help you practice some R coding. You may work with a partner to aid in your learning. However, your answers must be unique. For all questions show supporting code and plots as necessary.

**Data description**

Here we are going to analyze mycotoxin data that has been 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) mycotoxins Deoxynivalenol (DON) and the variant 15ADON from the wheat grain and this is your main response variable. In other words, we are interested to know if the fungal endophytes were able to reduce the concentration of DON on wheat heads. This was done across two wheat cultivars, a winter wheat variety Ambassador and 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>

**Data variables**

* Treatment
* Cultivar
* BioRep
* MassperSeed\_mg
* DON
* 15ADON

1. 10pts. Explain the following concepts about ggplot
   1. What is a geom?
   2. What is a facet?
   3. Explain the concept of layering.
   4. Where do you add x and y variables and map different shapes, colors, and other attributes to the data?
   5. Explain what ‘jittering’ does and why it’s important.
2. 5 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 read in the data and make the plot. Change the y label to “DON (ppm)” and make the x label blank.
3. 5 pts. Now convert this same data into a bar chart with standard-error error bars.
4. 5 pts. Add points to the foreground of the boxplot and bar chart you made in questions 2 and 3 that show the distribution of points over the boxplots or bars. Set the shape = 21 and outline color black (hint: use jitter\_dodge).
5. 3 pts. Change the color and fill of the points and boxplots to match some of the colors in the following colorblind pallet.

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

1. 5 pts. Add a facet to the plots based on cultivar.
2. 2 pts. Explore the themes and change to a theme of interest. <https://ggplot2.tidyverse.org/reference/ggtheme.html>
3. 5 pts. Add transparency to the points so you can still see the boxplot or bar in the background.
4. 5 pts. Explain what the boxplot and barcharts are showing. How are they similar? How are they different? Which representation of the data are better and why?
5. 5 pts. Explore two other ways 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?