

Coding challenge 2

Mamata K C

2025-02-20

Advanced Data Visualization– 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: Mamata K C

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. 5pts. Using ggplot, create a boxplot of DON by Treatment so that the plot looks like the image below.
 - a. Jitter points over the boxplot and fill the points and boxplots Cultivar with two colors from the cbbPallete introduced last week.
 - b. Change the transparency of the jittered points to 0.6.
 - c. The y-axis should be labeled “DON (ppm)”, and the x-axis should be left blank.
 - d. The plot should use a classic theme
 - e. The plot should also be faceted by Cultivar

```
datum <- read.csv ("MycotoxinData.csv" , na.strings = "na") #loading data in R
str(datum) #viewing the structure of the data
```

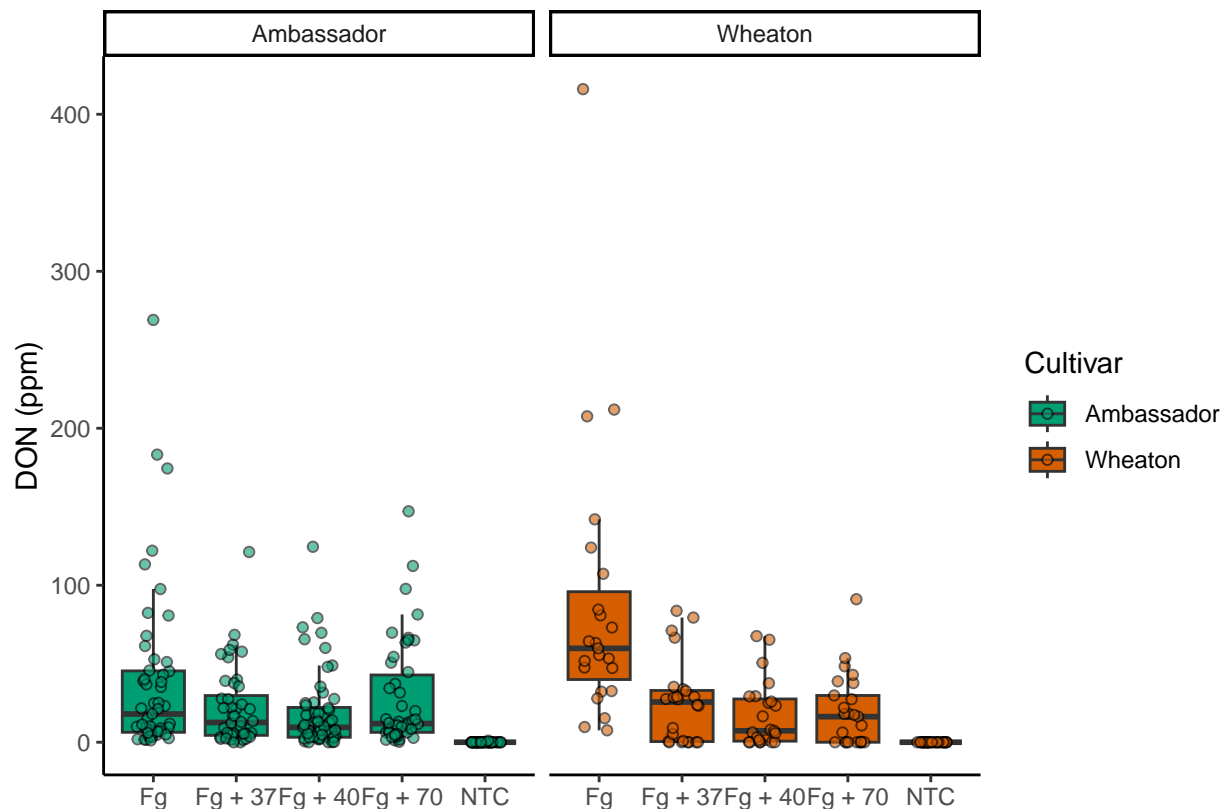
```
## 'data.frame': 375 obs. of 6 variables:
## $ Treatment : chr "Fg" "Fg" "Fg" "Fg" ...
## $ Cultivar : chr "Wheaton" "Wheaton" "Wheaton" "Wheaton" ...
## $ BioRep : int 2 2 2 2 2 2 2 2 2 3 ...
## $ MassperSeed_mg: num 10.29 12.8 2.85 6.5 10.18 ...
## $ DON : num 107.3 32.6 416 211.9 124 ...
## $ X15ADON : num 3 0.85 3.5 3.1 4.8 3.3 6.9 2.9 2.1 0.71 ...
```

```
library(ggplot2) #loading library ggplot2
cbbPalette <- c("#000000", "#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00", "#CC79A7")

boxplot <- ggplot(data = datum, aes(x = Treatment, y = DON, fill=Cultivar)) + #defining aesthetics x as
  geom_boxplot(position = position_dodge(), outlier.shape = NA) + #creating box plot without overlap a
  geom_point(position = position_jitterdodge(dodge.width = 0.8), aes(fill=Cultivar),alpha = 0.6, shape=)
  scale_fill_manual(values = c(cbbPalette[[4]], cbbPalette[[7]])) + #setting fills of boxplot and data
  xlab("") + #labelling x axis
  ylab("DON (ppm)") + #labelling y axis
  theme_classic() + #setting theme classic to make plain white background
  facet_wrap(~Cultivar) #separating the plots according to cultivars
boxplot
```

```
## Warning: Removed 8 rows containing non-finite outside the scale range
## ('stat_boxplot()').
```

```
## Warning: Removed 8 rows containing missing values or values outside the scale range
## ('geom_point()').
```

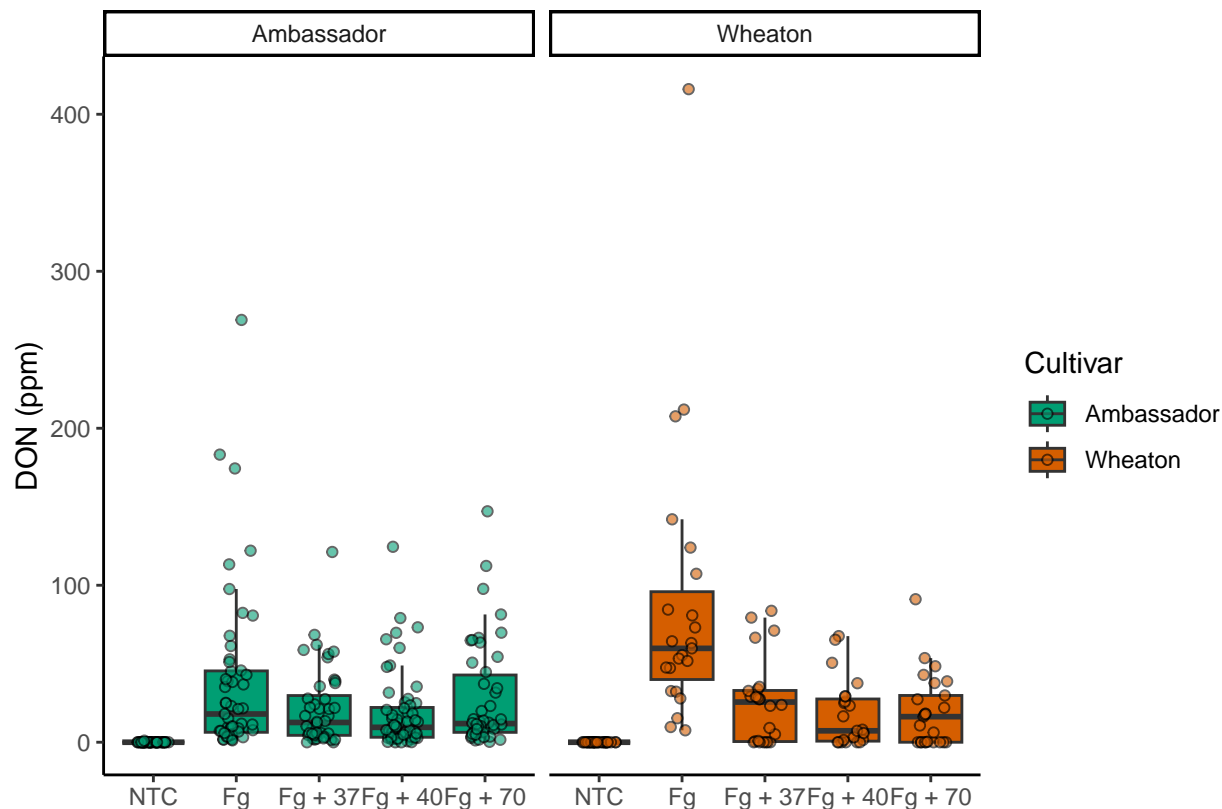


2. 4pts. Change the factor order level so that the treatment “NTC” is first, followed by “Fg”, “Fg + 37”, “Fg + 40”, and “Fg + 70”.

```
datum$Treatment <- factor(datum$Treatment, levels = c("NTC", "Fg", "Fg + 37", "Fg + 40", "Fg + 70")) #c
boxplot1 <- ggplot(data = datum, aes(x = Treatment, y = DON, fill = Cultivar)) + #defining aesthetics
  geom_boxplot(position = position_dodge(), outlier.shape = NA) + #creating box plot without overlap an
  geom_point(position = position_jitterdodge(dodge.width = 0.8), aes(fill=Cultivar),alpha = 0.6, shape=
  scale_fill_manual(values = c(cbbPalette[[4]], cbbPalette[[7]])) + #setting fills of boxplot and data
  xlab("") + #labelling x axis
  ylab("DON (ppm)") + #labelling y axis
  theme_classic() + #setting theme classic to make plain white background
  facet_wrap(~Cultivar) #separating the plots according to cultivars
boxplot1
```

```
## Warning: Removed 8 rows containing non-finite outside the scale range
## ('stat_boxplot()').
```

```
## Warning: Removed 8 rows containing missing values or values outside the scale range
## ('geom_point()').
```

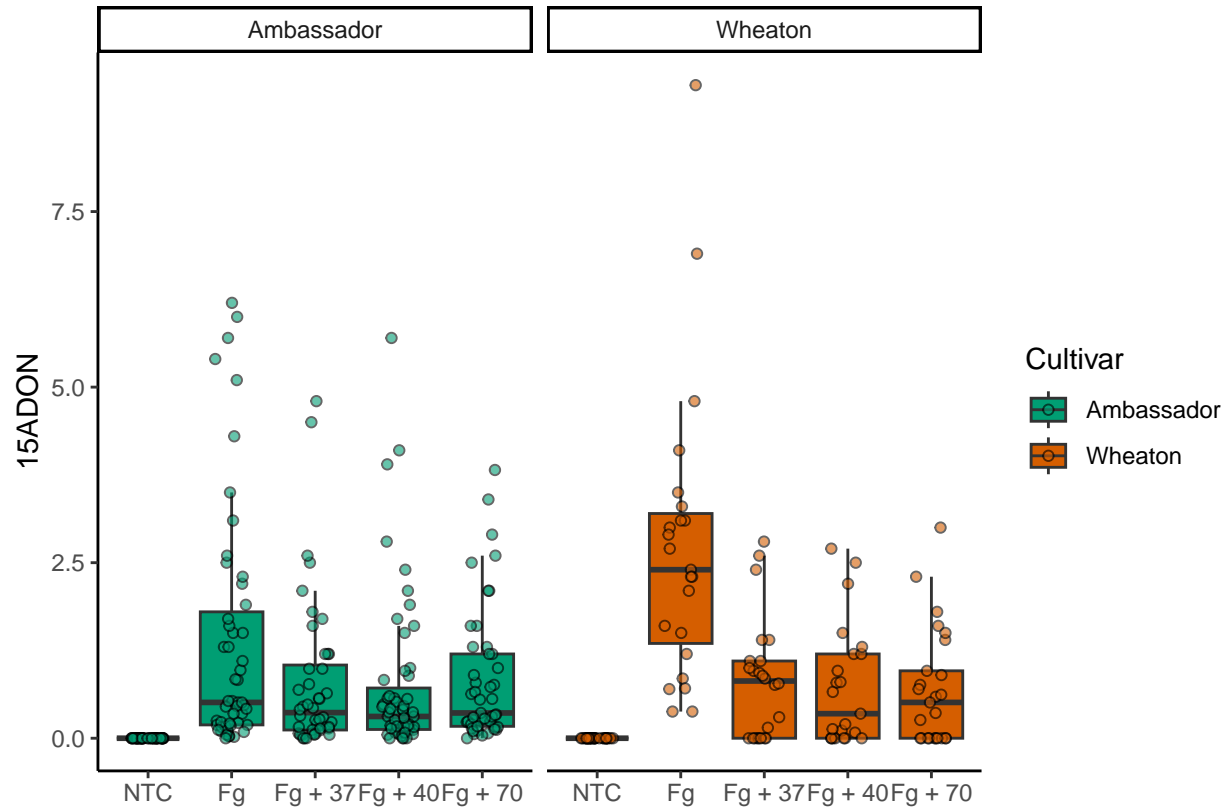


3. 5pts. Change the y-variable to plot X15ADON and MassperSeed_mg. The y-axis label should now be "15ADON" and "Seed Mass (mg)". Save plots made in questions 1 and 3 into three separate R objects.

```
boxplot_15ADON <- ggplot(data = datum, aes(x = Treatment, y = X15ADON, fill = Cultivar)) + #defining aes
  geom_boxplot(position = position_dodge(), outlier.shape = NA) + #creating box plot without overlap and
  geom_point(position = position_jitterdodge(dodge.width = 0.8), aes(fill=Cultivar),alpha = 0.6, shape=)
  scale_fill_manual(values = c(cbbPalette[[4]], cbbPalette[[7]])) + #setting fills of boxplot and data points
  xlab("") + #labelling x axis
  ylab("15ADON") + #labelling y axis
  theme_classic() + #setting theme classic to make plain white background
  facet_wrap(~Cultivar) #separating the plots according to cultivars
boxplot_15ADON
```

```
## Warning: Removed 10 rows containing non-finite outside the scale range
## ('stat_boxplot()').
```

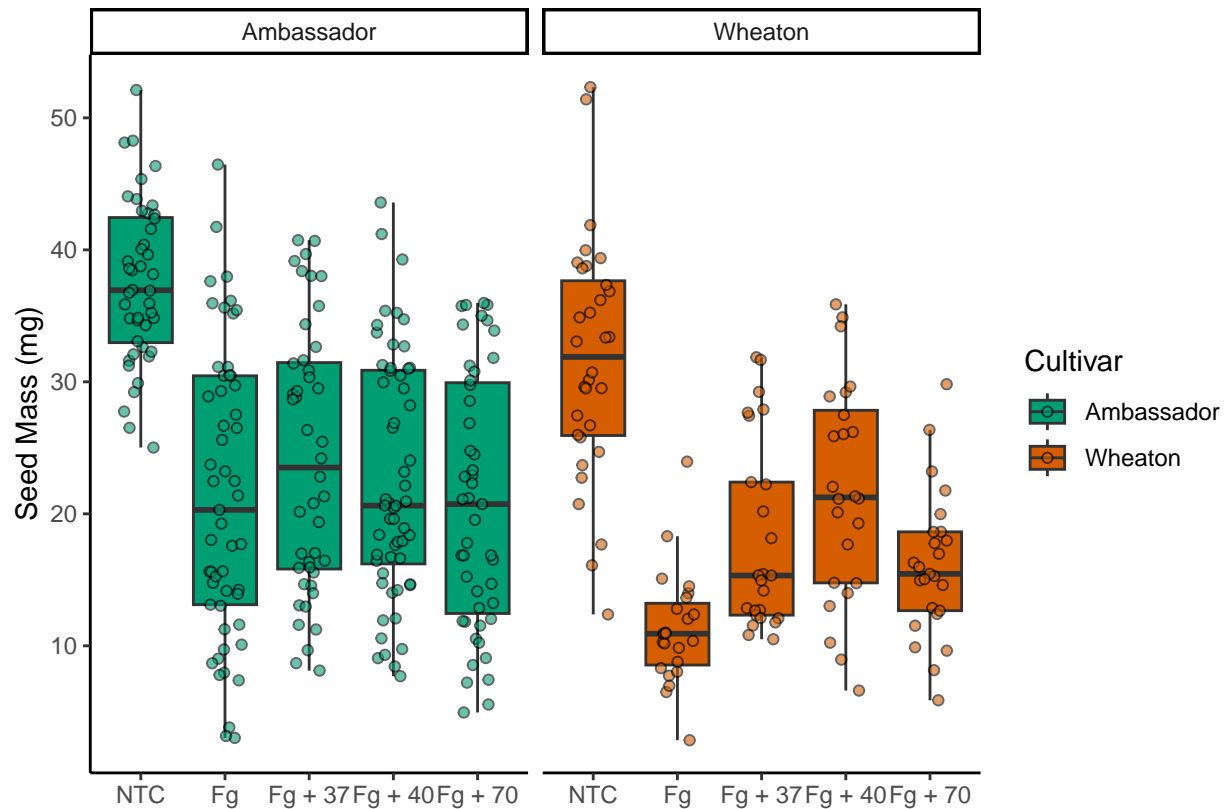
```
## Warning: Removed 10 rows containing missing values or values outside the scale range
## ('geom_point()').
```



```
boxplot_massperseed <- ggplot(data = datum, aes(x = Treatment, y = MassperSeed_mg, fill = Cultivar)) +
  geom_boxplot(position = position_dodge(), outlier.shape = NA) + #creating box plot without overlap and
  geom_point(position = position_jitterdodge(dodge.width = 0.8), aes(fill=Cultivar),alpha = 0.6, shape=1) +
  scale_fill_manual(values = c(cbbPalette[[4]], cbbPalette[[7]])) + #setting fills of boxplot and data points
  xlab("") + #labelling x axis
  ylab("Seed Mass (mg)") + #labelling y axis
  theme_classic() + #setting theme classic to make plain white background
  facet_wrap(~Cultivar) #separating the plots according to cultivars
boxplot_massperseed
```

```
## Warning: Removed 2 rows containing non-finite outside the scale range
## ('stat_boxplot()').
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range
## ('geom_point()').
```



4. 5pts. Use `ggarrange` function to combine all three figures into one with three columns and one row. Set the labels for the subplots as A, B and C. Set the option `common.legend = T` within `ggarrange` function. What did the `common.legend` option do?

a. HINT: I didn't specifically cover this in the tutorial, but you can go to the help page for the `ggarrange` function to figure out what the `common.legend` option does and how to control it.

```
library(ggpubr)
figure <- ggarrange(
  boxplot1, #plotting boxplot1
  boxplot_15ADON, #plotting boxplot_15ADON
  boxplot_massperseed, #plotting boxplot_massperseed
  labels = "AUTO", #setting labels as A, B and C
  nrow = 1, #combining figures with single row
  ncol = 3, #combining figures with 3 columns
  common.legend = TRUE, #combining legends for all three figures as one
  legend = TRUE #displaying legend
)
```

```
## Warning: Removed 8 rows containing non-finite outside the scale range
## ('stat_boxplot()').
```

```
## Warning: Removed 8 rows containing missing values or values outside the scale range
## ('geom_point()').
```

```
## Warning: Removed 8 rows containing non-finite outside the scale range
## ('stat_boxplot()').

## Warning: Removed 8 rows containing missing values or values outside the scale range
## ('geom_point()').

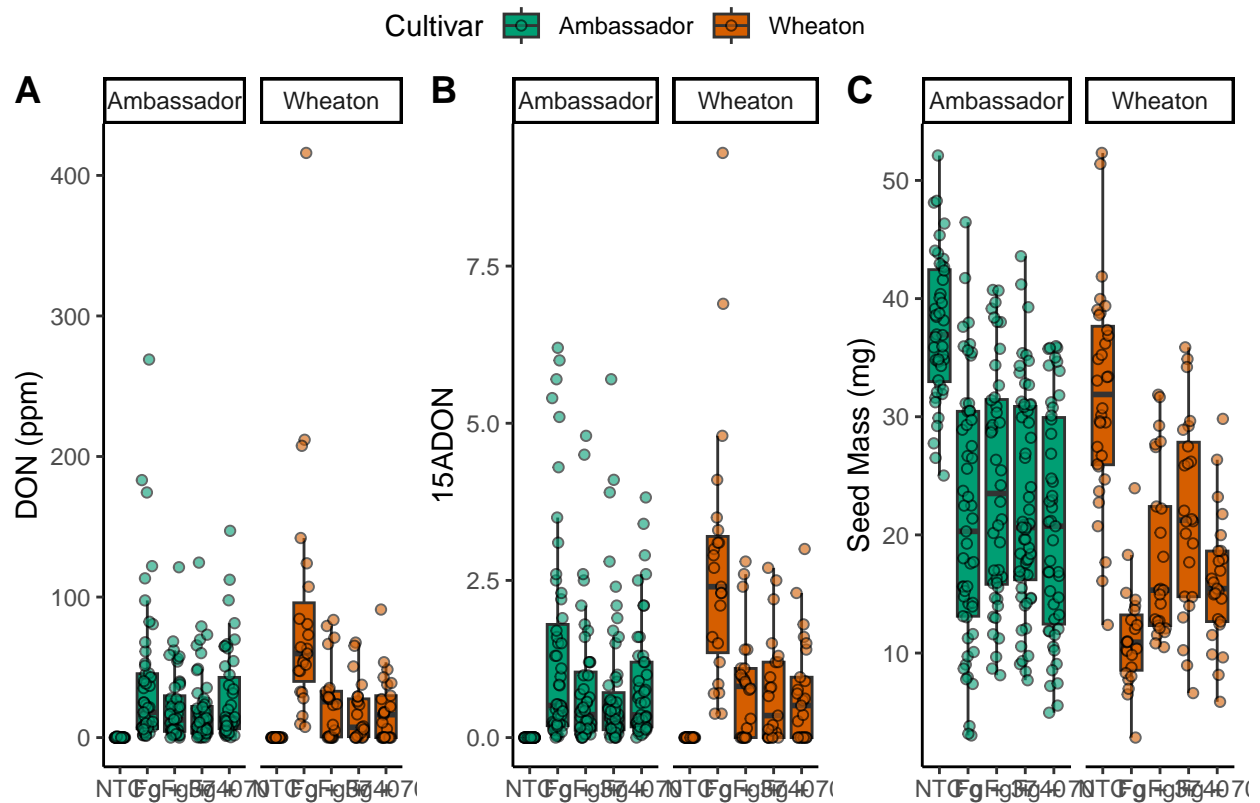
## Warning: Removed 10 rows containing non-finite outside the scale range
## ('stat_boxplot()').

## Warning: Removed 10 rows containing missing values or values outside the scale range
## ('geom_point()').

## Warning: Removed 2 rows containing non-finite outside the scale range
## ('stat_boxplot()').

## Warning: Removed 2 rows containing missing values or values outside the scale range
## ('geom_point()').
```

figure



The common.legend option within ggarrange function let us combine the legends and display only one legend for multiple plots that have been combined using ggarrange function. The option can be controlled by setting it to either TRUE or FALSE. If we set the option as TRUE R will give one legend across multiple plots and if we set it as FALSE, R will give separate legends for separate plots.

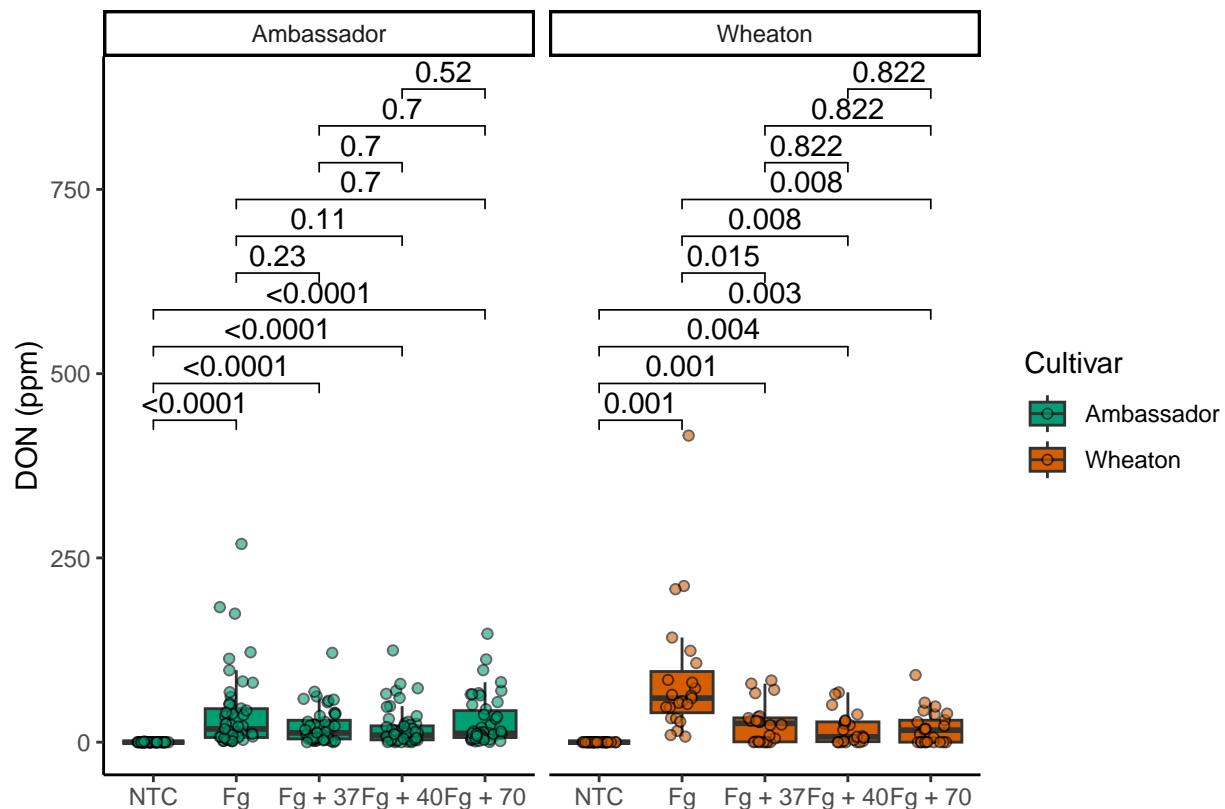
5. 5pts. Use `geom_pwc()` to add t.test pairwise comparisons to the three plots made above. Save each plot as a new R object, and combine them again with `ggarange` as you did in question 4.

```
boxplot1_pwc <- boxplot1 +  
  geom_pwc(aes(group = Treatment), method = "t.test", label = "p.adj.format") #pair-wise comparisons be  
boxplot1_pwc
```

```
## Warning: Removed 8 rows containing non-finite outside the scale range  
## ('stat_boxplot()').
```

```
## Warning: Removed 8 rows containing non-finite outside the scale range  
## ('stat_pwc()').
```

```
## Warning: Removed 8 rows containing missing values or values outside the scale range  
## ('geom_point()').
```



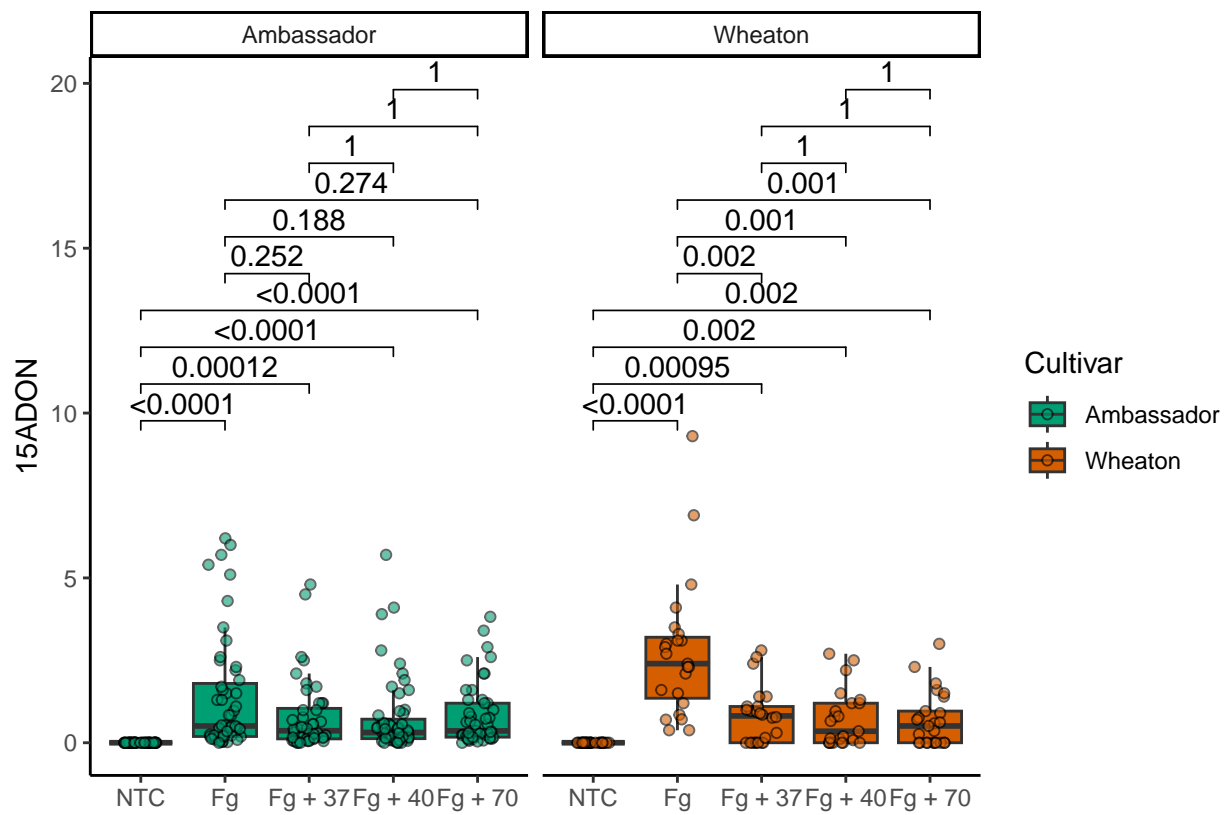
```
boxplot_15ADON_pwc <- boxplot_15ADON +  
  geom_pwc(aes(group = Treatment), method = "t.test", label = "p.adj.format") #pair-wise comparisons b  
boxplot_15ADON_pwc
```

```
## Warning: Removed 10 rows containing non-finite outside the scale range  
## ('stat_boxplot()').
```

```
## Warning: Removed 10 rows containing non-finite outside the scale range  
## ('stat_pwc()').
```



```
## Warning: Removed 10 rows containing missing values or values outside the scale range
## ('geom_point()').
```

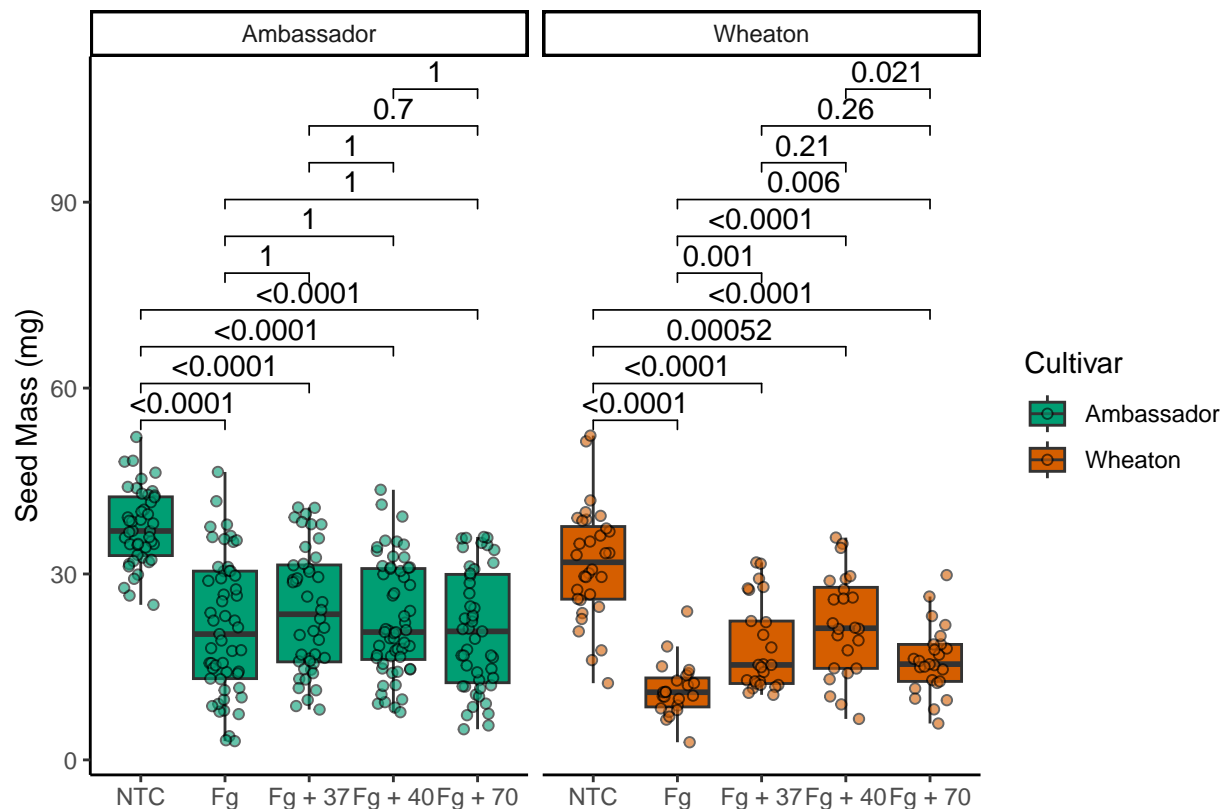


```
boxplot_massperseed_pwc <- boxplot_massperseed +
  geom_pwc(aes(group = Treatment), method = "t.test", label = "p.adj.format") #pair-wise comparisons be
boxplot_massperseed_pwc
```

```
## Warning: Removed 2 rows containing non-finite outside the scale range
## ('stat_boxplot()').
```

```
## Warning: Removed 2 rows containing non-finite outside the scale range
## ('stat_pwc()').
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range
## ('geom_point()').
```



```
figure_pwc <- ggarrange(
  boxplot1_pwc, #plotting boxplot1_pwc
  boxplot_15ADON_pwc, #plotting boxplot_15ADON_pwc
  boxplot_massperseed_pwc, #plotting boxplot_massperseed_pwc
  labels = "AUTO", #setting labels as A, B and C
  nrow = 1, #combining figures with single row
  ncol = 3, #combining figures with 3 columns
  common.legend = TRUE, #combining legends for all three figures as one
  legend = TRUE #displaying legend
)
```

```
## Warning: Removed 8 rows containing non-finite outside the scale range
## ('stat_boxplot()').
```

```
## Warning: Removed 8 rows containing non-finite outside the scale range
## ('stat_pwc()').
```

```
## Warning: Removed 8 rows containing missing values or values outside the scale range
## ('geom_point()').
```

```
## Warning: Removed 8 rows containing non-finite outside the scale range
## ('stat_boxplot()').
```

```
## Warning: Removed 8 rows containing non-finite outside the scale range
## ('stat_pwc()').
```

```
## Warning: Removed 8 rows containing missing values or values outside the scale range
## ('geom_point()').

## Warning: Removed 10 rows containing non-finite outside the scale range
## ('stat_boxplot()').

## Warning: Removed 10 rows containing non-finite outside the scale range
## ('stat_pwc()').

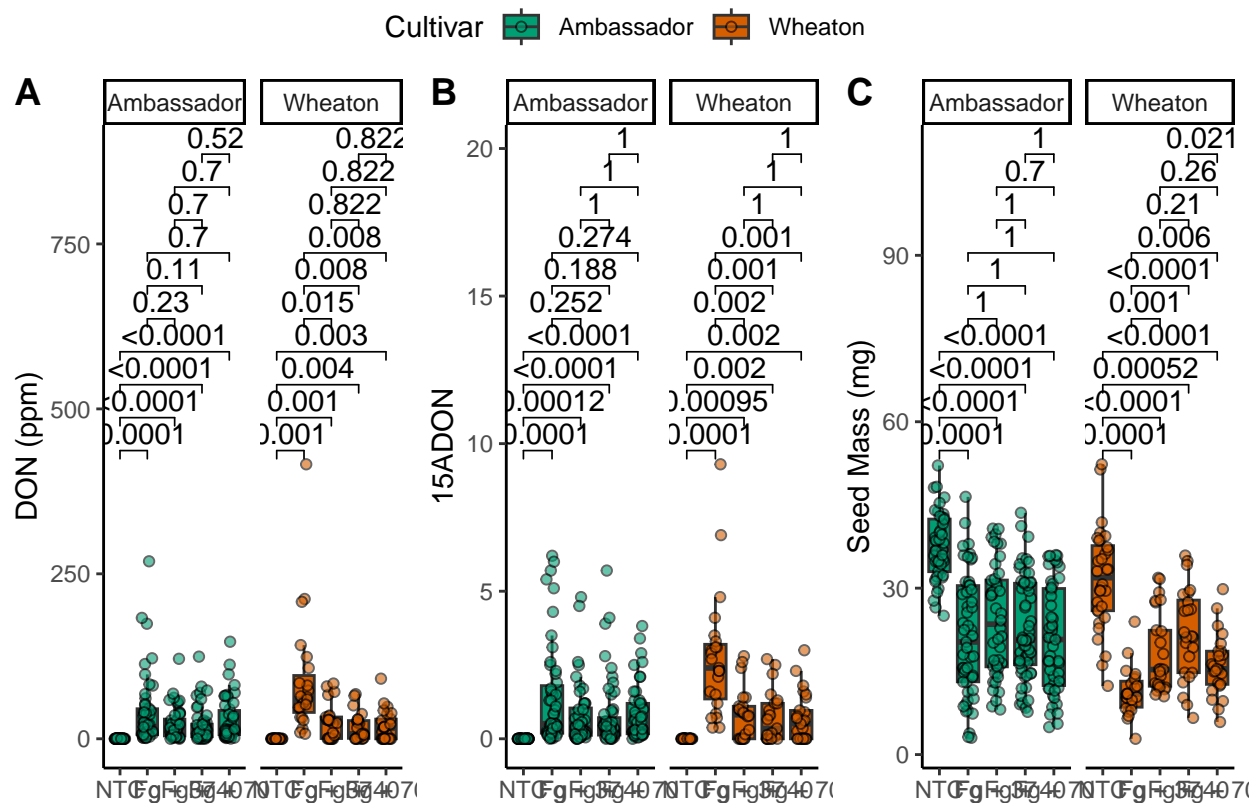
## Warning: Removed 10 rows containing missing values or values outside the scale range
## ('geom_point()').

## Warning: Removed 2 rows containing non-finite outside the scale range
## ('stat_boxplot()').

## Warning: Removed 2 rows containing non-finite outside the scale range
## ('stat_pwc()').

## Warning: Removed 2 rows containing missing values or values outside the scale range
## ('geom_point()').
```

figure_pwc



6. 1 pt. Annotate your code, push it to GitHub. Provide me a link to your GitHub

<https://github.com/mamata2423/ReproducibilityClass/blob/main/Advanced%20Visualization.Rmd>