

Anova test

Joshua Gichana

2022-12-03

1. Required packages

```
library(ggplot2)
library(ggpubr)
library(broom)
library(AICcmodavg)
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v tibble 3.1.8      v dplyr 1.0.10
## v tidyr 1.2.1      v stringr 1.4.1
## v readr 2.1.3      v forcats 0.5.2
## v purrr 0.3.5
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()
```

2. loading the data

```
crop.data <- read.csv('/cloud/project/crop.data.csv', header = TRUE , colClasses = c("factor", "factor"))
```

3. summary of the crop data

```
summary(crop.data)
```

```
## density block fertilizer yield
## 1:48 1:24 1:32 Min. :175.4
## 2:48 2:24 2:32 1st Qu.:176.5
## 3:24 3:32 Median :177.1
## 4:24 Mean :177.0
## 3rd Qu.:177.4
## Max. :179.1
```

4. One way test

```
one.way <- aov(yield ~ fertilizer , data = crop.data)
```

5. summary of the one way test

```
summary(one.way)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## fertilizer  2   6.07  3.0340   7.863 7e-04 ***
## Residuals  93  35.89  0.3859
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

6. Two anova test

```
two.way <- aov( yield ~ fertilizer + density , data = crop.data)
```

7. summary of the two test

```
summary(two.way)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## fertilizer    2  6.068    3.034    9.073 0.000253 ***
## density       1  5.122    5.122   15.316 0.000174 ***
## Residuals    92 30.765    0.334
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

8. checking interaction btwn independent variables

```
interaction <- aov( yield ~ fertilizer * density , data = crop.data)
```

9. summary of the interaction

```
summary(interaction)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## fertilizer      2  6.068    3.034    9.001 0.000273 ***
## density         1  5.122    5.122   15.195 0.000186 ***
## fertilizer:density 2  0.428    0.214    0.635 0.532500
## Residuals      90 30.337    0.337
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

10. adding a blocking variable

```
blocking <- aov( yield ~ fertilizer + density + block , data = crop.data)
```

11. summarizing the block model

```
summary(blocking)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## fertilizer      2  6.068    3.034    9.018 0.000269 ***
## density         1  5.122    5.122   15.224 0.000184 ***
## block           2  0.486    0.243    0.723 0.488329
## Residuals      90 30.278    0.336
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

12. The best fit model using the akaike information criterion

```
library(AICcmodavg)
```

```
model.set <- list(one.way, two.way , interaction , blocking)
```

```
model.names <- c("one.way", "two.way", "interaction", "blocking")
```

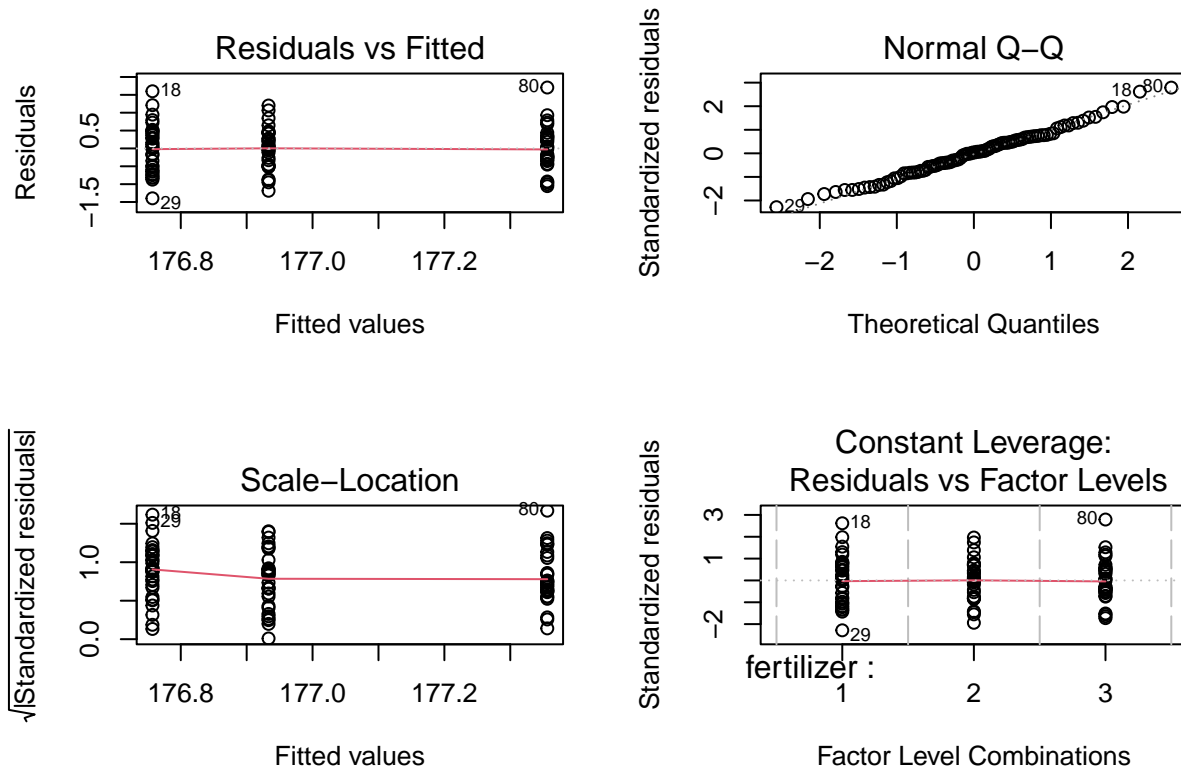
```
aictab(model.set, modnames=model.names)
```

```
##
## Model selection based on AICc:
##
##              K    AICc Delta_AICc AICcWt Cum.Wt    LL
## two.way      5 173.86         0.00   0.71   0.71 -81.59
```

```
## blocking      7 176.93      3.08   0.15   0.86 -80.83
## interaction    7 177.12      3.26   0.14   1.00 -80.92
## one.way       4 186.41     12.56   0.00   1.00 -88.99
```

14. Checking for homoscedasticity (common variance assumption) using diagnostic plots

```
par(mfrow = c(2,2)) # recreates a 2*2 plotting matrix
plot(one.way)
```



```
par(mfrow = c(1,1))
```

15. Doing a post-hoc test we use the turkey's Honestly Significant Difference (Turkey's HSD) post hoc test for pairwise comparison

```
tukey.two.way <- TukeyHSD(two.way)
```

```
tukey.two.way
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = yield ~ fertilizer + density, data = crop.data)
##
## $fertilizer
##      diff      lwr      upr    p adj
## 2-1 0.1761687 -0.16822506 0.5205625 0.4452958
## 3-1 0.5991256  0.25473179 0.9435194 0.0002219
## 3-2 0.4229569  0.07856306 0.7673506 0.0119381
##
## $density
##      diff      lwr      upr    p adj
```

```
## 2-1 0.461956 0.2275204 0.6963916 0.0001741
```

16 find out which group means are statistically different from one another

```
tukey.plot.aov <- aov(yield ~ fertilizer:density , data = crop.data)
```

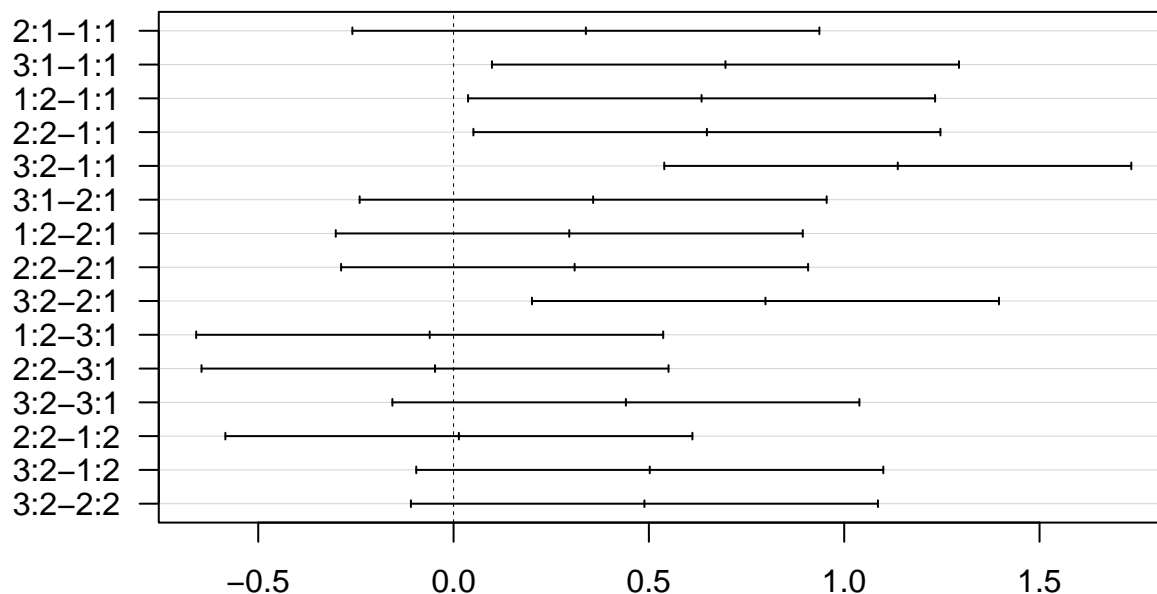
18. plotting the results in a graph

-When plotting the results of a model, it is important to display:

- the raw data
- summary information, usually the mean and standard error of each group being compared
- letters or symbols above each group being compared to indicate the groupwise differences.

```
tukey.plot.test <- TukeyHSD(tukey.plot.aov)
plot(tukey.plot.test , las = 1)
```

95% family-wise confidence level



Differences in mean levels of fertilizer:density

19. making a dataframe with group labels

```
mean.yield.data <- crop.data %>%
  group_by(fertilizer , density) %>%
  summarize(yield = mean(yield))
```

```
## `summarise()` has grouped output by 'fertilizer'. You can override using the
## `.groups` argument.
```

```
# adding labels
```

```
mean.yield.data$group <- c("a", "b", "b", "b", "b", "c")
```

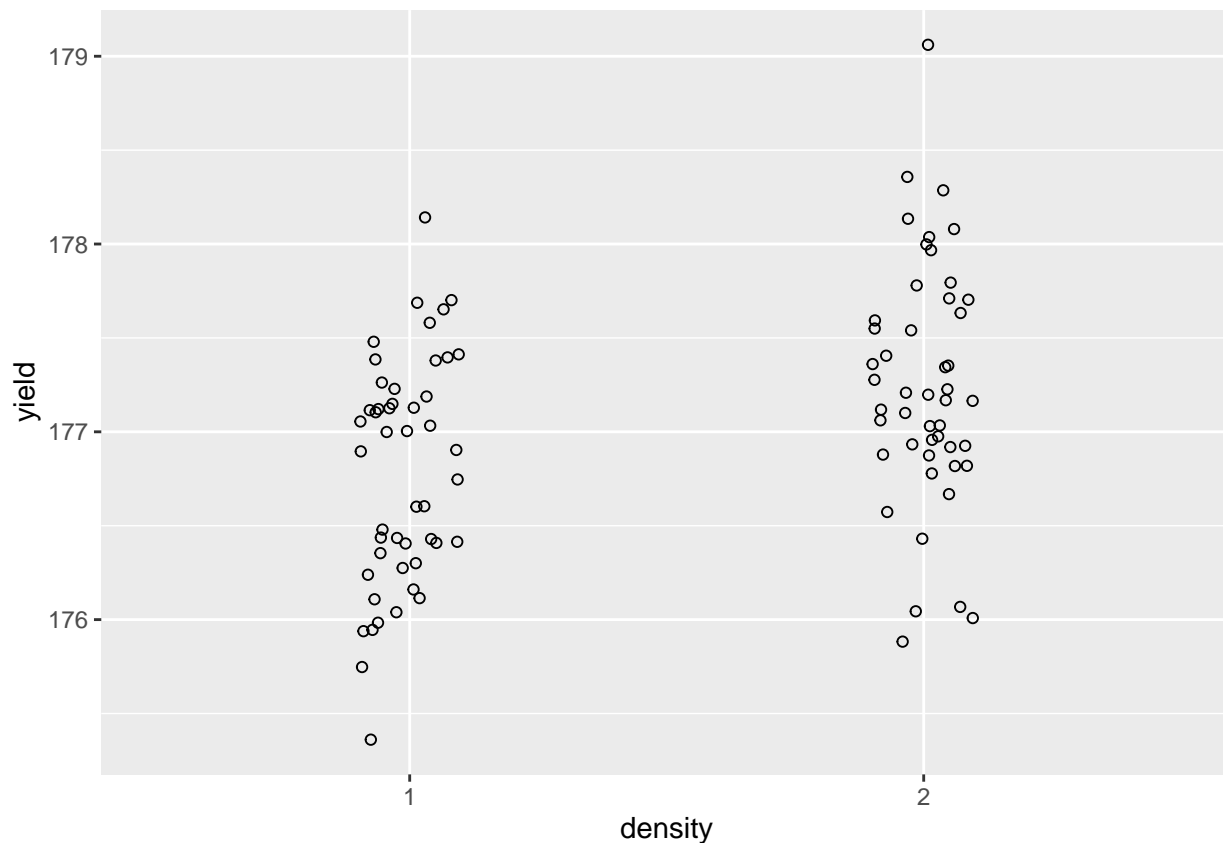
```
mean.yield.data
```

```
## # A tibble: 6 x 4
## # Groups:   fertilizer [3]
##   fertilizer density yield group
##   <fct>      <fct>  <dbl> <chr>
## 1 1          1      176. a
## 2 1          2      177. b
## 3 2          1      177. b
## 4 2          2      177. b
## 5 3          1      177. b
## 6 3          2      178. c
```

a) plot raw data

```
two.way.plot <- ggplot(data = crop.data ,mapping = aes(x = density , y = yield , group = fertilizer))+
  geom_point(cex = 1.5 , pch = 1.0 , position = position_jitter(w = 0.1 , h = 0))
```

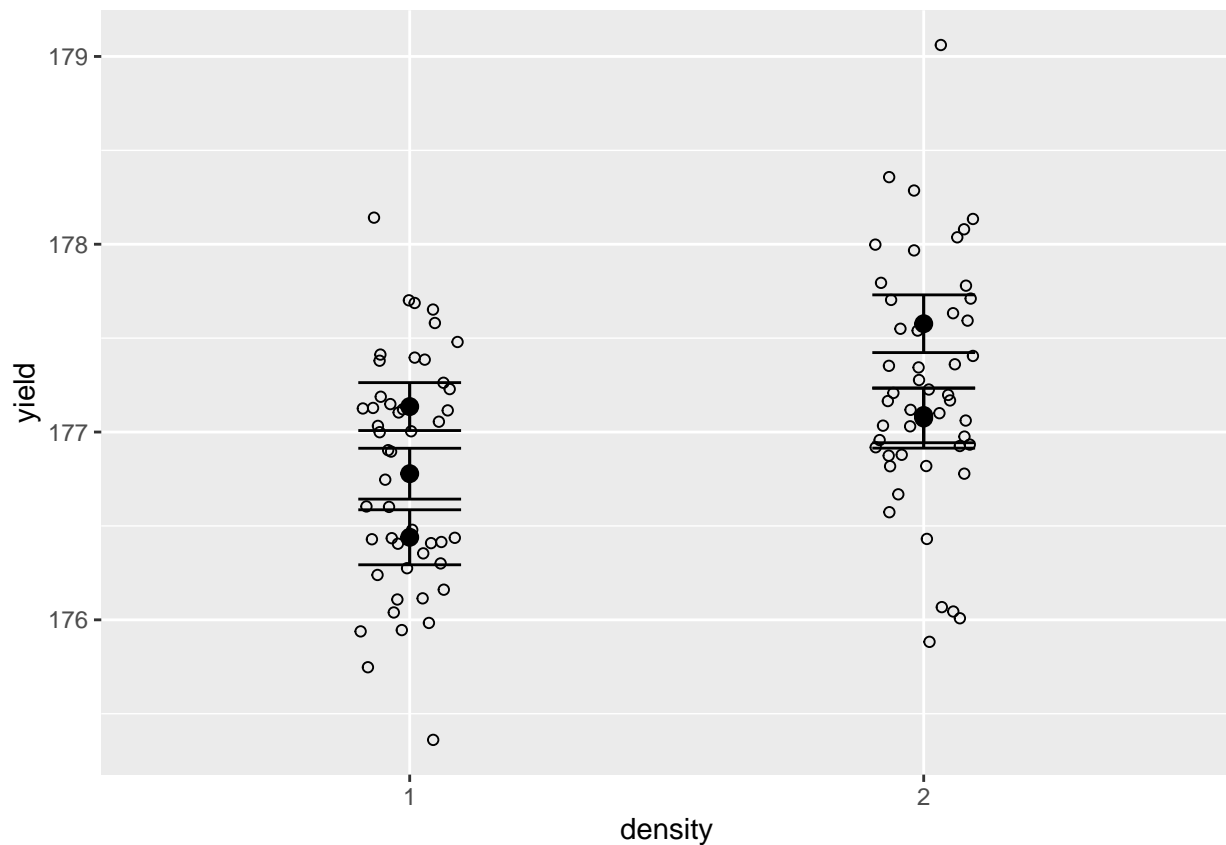
two.way.plot



b) Add means and standard errors to the graphs

```
two.way.plot <- two.way.plot +
  stat_summary(fun.data = 'mean_se' , geom = 'errorbar' , width = 0.2)+
  stat_summary(fun.data = 'mean_se' , geom = 'pointrange')+
  geom_point(data = mean.yield.data , mapping = aes(x = density , y = yield))
```

two.way.plot

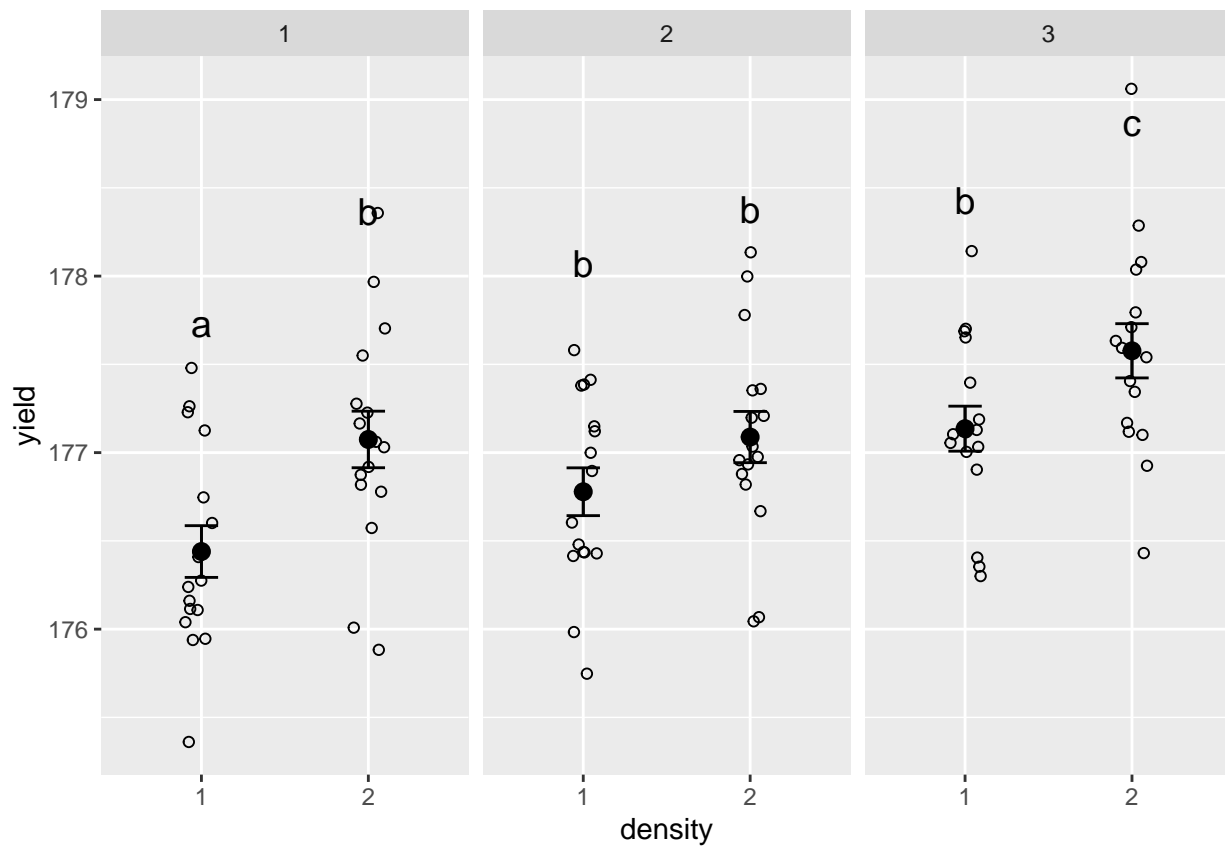


This is very hard to read, since all of the different groupings for fertilizer type are stacked on top of one another. We need to split the data.

split up the data

```
two.way.plot <- two.way.plot +
  geom_text(data = mean.yield.data ,label = mean.yield.data$group , vjust = -8 , size = 5)+
  facet_wrap(~ fertilizer)

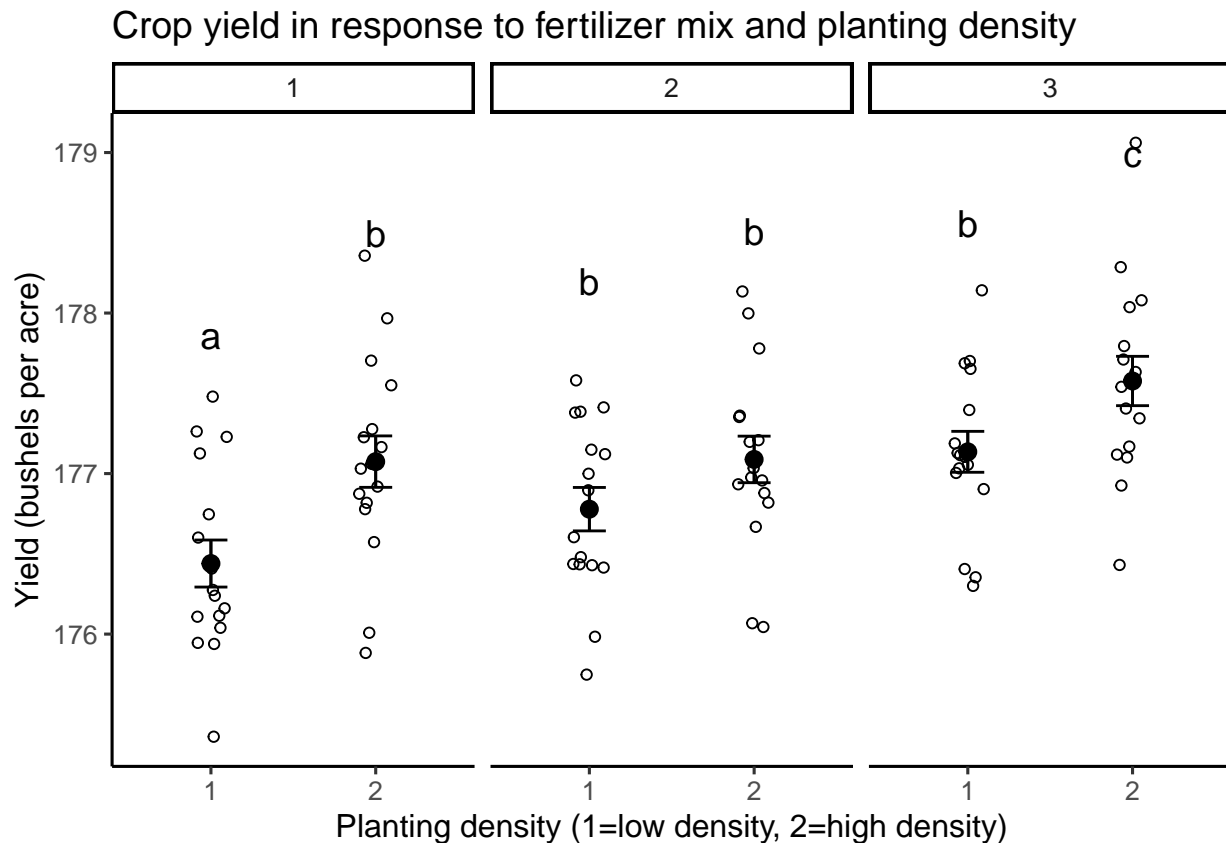
two.way.plot
```



20. making the graph ready for publication

```
two.way.plot <- two.way.plot +
  theme_classic2() +
  labs(title = "Crop yield in response to fertilizer mix and planting density",
       x = "Planting density (1=low density, 2=high density)",
       y = "Yield (bushels per acre)")

two.way.plot
```



21. Reporting the results

In addition to a graph, it's important to state the results of the ANOVA test. Include:

- A brief description of the variables you tested
- The F value, degrees of freedom, and p values for each independent variable
- What the results mean.
 - We found a statistically-significant difference in average crop yield by both fertilizer type ($F(2)=9.018$, $p < 0.001$) and by planting density ($F(1)=15.316$, $p < 0.001$).
 - A Tukey post-hoc test revealed that fertilizer mix 3 resulted in a higher yield on average than fertilizer mix 1 (0.59 bushels/acre), and a higher yield on average than fertilizer mix 2 (0.42 bushels/acre). Planting density was also significant, with planting density 2 resulting in an higher yield on average of 0.46 bushels/acre over planting density 1.
 - A subsequent groupwise comparison showed the strongest yield gains at planting density 2, fertilizer mix 3, suggesting that this mix of treatments was most advantageous for crop growth under our experimental conditions.