Week 6

J. Lacasa

10/14/2021

Packages

```
library(tidyverse)
```

```
## Registered S3 methods overwritten by 'tibble':
##   method     from
##   format.tbl pillar
##   print.tbl  pillar

## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.7
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## Warning: package 'ggplot2' was built under R version 4.0.5

## Warning: package 'dplyr' was built under R version 4.0.5

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(car)
```

```
## Warning: package 'car' was built under R version 4.0.5

## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

```
library(multcomp)
```

```
## Warning: package 'multcomp' was built under R version 4.0.5
```

```
## Loading required package: mvtnorm
```

```
## Warning: package 'mvtnorm' was built under R version 4.0.5
```

```
## Loading required package: survival
```

```
## Loading required package: TH.data
```

```
## Loading required package: MASS
```

```
## Warning: package 'MASS' was built under R version 4.0.5
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
##
## Attaching package: 'TH.data'
```

```
## The following object is masked from 'package:MASS':
##
##     geyser
```

```
library(multcompView)
```

```
## Warning: package 'multcompView' was built under R version 4.0.5
```

```
library(emmeans)
```

```
## Warning: package 'emmeans' was built under R version 4.0.5
```

Data

```
usda <- read.csv("../data/USDAsurvey_00-18.csv")
legume <- read.csv("../data/legume.csv")
```

```r
colnames(legume)
```

```
##   [1] "IDCrop"
##   [2] "Crop_Sequence_Treatment_Name"
##   [3] "Crop_Site_Growing_Season_ID"
##   [4] "Crop_Growing_Season_Year_First"
##   [5] "Crop_Growing_Season_Year_Last"
##   [6] "Crop_Growing_Season_Number"
##   [7] "Crop_Species_Scientific_Name"
##   [8] "Crop_Species_Common_Name"
##   [9] "Crop_Species_Legume"
##  [10] "Crop_Date_Seeding"
##  [11] "Crop_Date_Harvest"
##  [12] "Crop_Date_From_Seeding_To_Harvest_Day_Number"
##  [13] "Crop_Following_Number"
##  [14] "Crop_Multiple_Following_For_Same_Preceding"
##  [15] "Crop_Across_Treatment_Averaged_Value"
##  [16] "Crop_Across_Treatment_Averaged_Value_Type"
##  [17] "Crop_Across_Species_Same_Treatment_Value"
##  [18] "Crop_Across_Species_Same_Treatment_Value_Type"
##  [19] "Crop_Replicate_Number"
##  [20] "Crop_Yield_Grain"
##  [21] "Crop_Yield_Grain_Unit"
##  [22] "Crop_Yield_Grain_Error"
##  [23] "Crop_Yield_Grain_Error_Type"
##  [24] "Crop_Yield_Grain_DM_Percentage"
##  [25] "Crop_Biomass_Aerial"
##  [26] "Crop_Biomass_Aerial_Unit"
##  [27] "Crop_Biomass_Aerial_Error"
##  [28] "Crop_Biomass_Aerial_Error_Type"
##  [29] "Crop_Biomass_Aerial_DM_Percentage"
##  [30] "Crop_Biomass_Aerial_Definition"
##  [31] "Crop_Biomass_Aerial_Stage_Detailed"
##  [32] "Crop_Biomass_Aerial_Stage_Simplified"
##  [33] "Crop_Harvest_Index"
##  [34] "Crop_Harvest_Index_Error"
##  [35] "Crop_Harvest_Index_Error_Type"
##  [36] "Crop_N_Quantity_Grain"
##  [37] "Crop_N_Quantity_Grain_Unit"
##  [38] "Crop_N_Quantity_Grain_Error"
##  [39] "Crop_N_Quantity_Grain_Error_Type"
##  [40] "Crop_N_Quantity_Aerial"
##  [41] "Crop_N_Quantity_Aerial_Unit"
##  [42] "Crop_N_Quantity_Aerial_Error"
##  [43] "Crop_N_Quantity_Aerial_Error_Type"
##  [44] "Crop_N_Quantity_Aerial_Definition"
##  [45] "Crop_N_Percentage_Grain"
##  [46] "Crop_N_Percentage_Grain_Error"
##  [47] "Crop_N_Percentage_Grain_Error_Type"
##  [48] "Crop_N_Percentage_Aerial"
##  [49] "Crop_N_Percentage_Aerial_Error"
##  [50] "Crop_N_Percentage_Aerial_Error_Type"
##  [51] "Crop_N_Percentage_Aerial_Definition"
```

```
##  [52] "Crop_N_Harvest_Index"
##  [53] "Crop_N_Harvest_Index_Error"
##  [54] "Crop_N_Harvest_Index_Error_Type"
##  [55] "Crop_N_Fixed_Quantity_Aerial"
##  [56] "Crop_N_Fixed_Quantity_Aerial_Unit"
##  [57] "Crop_N_Fixed_Quantity_Aerial_Error"
##  [58] "Crop_N_Fixed_Quantity_Aerial_Error_Type"
##  [59] "Crop_N_Fixed_Quantity_Aerial_Definition"
##  [60] "Crop_N_Fixed_Percentage_Aerial"
##  [61] "Crop_N_Fixed_Percentage_Aerial_Error"
##  [62] "Crop_N_Fixed_Percentage_Aerial_Error_Type"
##  [63] "Crop_N_Fixed_Percentage_Aerial_Method"
##  [64] "Crop_N_Fixed_Percentage_Aerial_Reference_Species"
##  [65] "Crop_N_Fixed_Percentage_Aerial_Stage_Detailed"
##  [66] "Crop_N_Fixed_Percentage_Aerial_Stage_Simplified"
##  [67] "Crop_Protein_Quantity_Percentage_Grain"
##  [68] "Crop_Protein_Quantity_Percentage_Grain_Unit"
##  [69] "Crop_Protein_Quantity_Percentage_Grain_Error"
##  [70] "Crop_Protein_Quantity_Percentage_Grain_Error_Type"
##  [71] "Crop_N_Balance_Simplified"
##  [72] "Crop_N_Balance_Simplified_Unit"
##  [73] "Crop_N_Balance_Simplified_Error"
##  [74] "Crop_N_Balance_Simplified_Error_Type"
##  [75] "Crop_N_Balance_Simplified_Equation"
##  [76] "Crop_N_Soil_Quantity_Percentage_Seeding"
##  [77] "Crop_N_Soil_Quantity_Percentage_Seeding_Unit"
##  [78] "Crop_N_Soil_Quantity_Percentage_Seeding_Type"
##  [79] "Crop_N_Soil_Quantity_Percentage_Seeding_Error"
##  [80] "Crop_N_Soil_Quantity_Percentage_Seeding_Error_Type"
##  [81] "Crop_N_Soil_Quantity_Percentage_Seeding_Depth"
##  [82] "Crop_N_Soil_Quantity_Percentage_Seeding_Depth_Unit"
##  [83] "Crop_N_Soil_Quantity_Percentage_Seeding_Date"
##  [84] "Crop_N_Soil_Quantity_Percentage_Harvest"
##  [85] "Crop_N_Soil_Quantity_Percentage_Harvest_Unit"
##  [86] "Crop_N_Soil_Quantity_Percentage_Harvest_Type"
##  [87] "Crop_N_Soil_Quantity_Percentage_Harvest_Error"
##  [88] "Crop_N_Soil_Quantity_Percentage_Harvest_Error_Type"
##  [89] "Crop_N_Soil_Quantity_Percentage_Harvest_Depth"
##  [90] "Crop_N_Soil_Quantity_Percentage_Harvest_Depth_Unit"
##  [91] "Crop_N_Soil_Quantity_Percentage_Harvest_Date"
##  [92] "Crop_Water_Use_Balance"
##  [93] "Crop_Water_Use_Balance_Unit"
##  [94] "Crop_Water_Use_Balance_Error"
##  [95] "Crop_Water_Use_Balance_Error_Type"
##  [96] "Crop_Water_Use_Balance_Equation"
##  [97] "Crop_Water_Use_Balance_Efficiency_Grain"
##  [98] "Crop_Water_Use_Balance_Efficiency_Grain_Unit"
##  [99] "Crop_Water_Use_Balance_Efficiency_Grain_Error"
## [100] "Crop_Water_Use_Balance_Efficiency_Grain_Error_Type"
## [101] "Crop_Water_Use_Balance_Efficiency_Aerial"
## [102] "Crop_Water_Use_Balance_Efficiency_Aerial_Unit"
## [103] "Crop_Water_Use_Balance_Efficiency_Aerial_Error"
## [104] "Crop_Water_Use_Balance_Efficiency_Aerial_Error_Type"
## [105] "Crop_Water_Use_Balance_Efficiency_Aerial_Definition"
```

```
## [106] "IDRotation_CropSystem"
```

```
nrow(legume)
```

```
## [1] 8581
```

```
unique(legume$Crop_Species_Scientific_Name)
```

```
##  [1] "Cicer arietinum"           "Vicia faba"
##  [3] "Lens culinaris"            "Pisum sativum"
##  [5] "Vigna radiata"             "Arachis hypogaea"
##  [7] "Glycine max"               "Triticum aestivum"
##  [9] "Lupinus albus"             "Lupinus angustifolius"
## [11] "Lupinus atlanticus"        "Lupinus pilosus"
## [13] "Vicia narbonensis"         "Lathyrus cicera"
## [15] "Lathyrus ochrus"           "Lathyrus sativus"
## [17] "Vicia benghalensis"        "Vicia sativa"
## [19] "Lupinus luteus"            "Avena sativa"
## [21] "Brassica napus"            "Brassica rapa"
## [23] "Brassica juncea"           "Triticum turgidum"
## [25] "Crambe abyssinica"         "Carthamus tinctorius"
## [27] "Helianthus annuus"         "Phaseolus vulgaris"
## [29] "Hordeum vulgare"           "Triticum sativum"
## [31] "Vigna subterranea"         "Vigna unguiculata"
## [33] "Cajanus cajan"             "Sorghum bicolor"
## [35] "Linum usitatissimum"       "Fallow"
## [37] "Panicum miliaceum"         "Triticum durum"
## [39] "Ã-Triticosecale"           "Vigna mungo"
## [41] "Oryza sativa"              "Zea mays"
## [43] "Vigna aconitifolia"        "Cyamopsis tetragonoloba"
## [45] "Lablab purpureus"          "Phaseolus lunatus"
## [47] "Solanum lycopersicum"      "Ipomoea batatas"
## [49] "Fagopyrum esculentum"      "Brassica chinensis"
## [51] "Vicia ervilia"             "Brassica campestris"
## [53] "Pennisetum glaucum"        "Vigna angularis"
## [55] "Vicia pannonica"           "Secale cereale"
## [57] "Lathyrus aphaca"           "Lathyrus clymenum"
## [59] "Trigonella foenum-graecum" "Vicia articulata"
## [61] "Vicia villosa"             "Vicia hybrida"
## [63] "Ricinus communis"          "Trifolium repens"
## [65] "Sinapis alba"              "Lupinus mutabilis"
## [67] "Gossypium hirsutum"        "Macrotyloma uniflorum"
## [69] "Sesamum indicum"           "Citrullus lanatus"
```

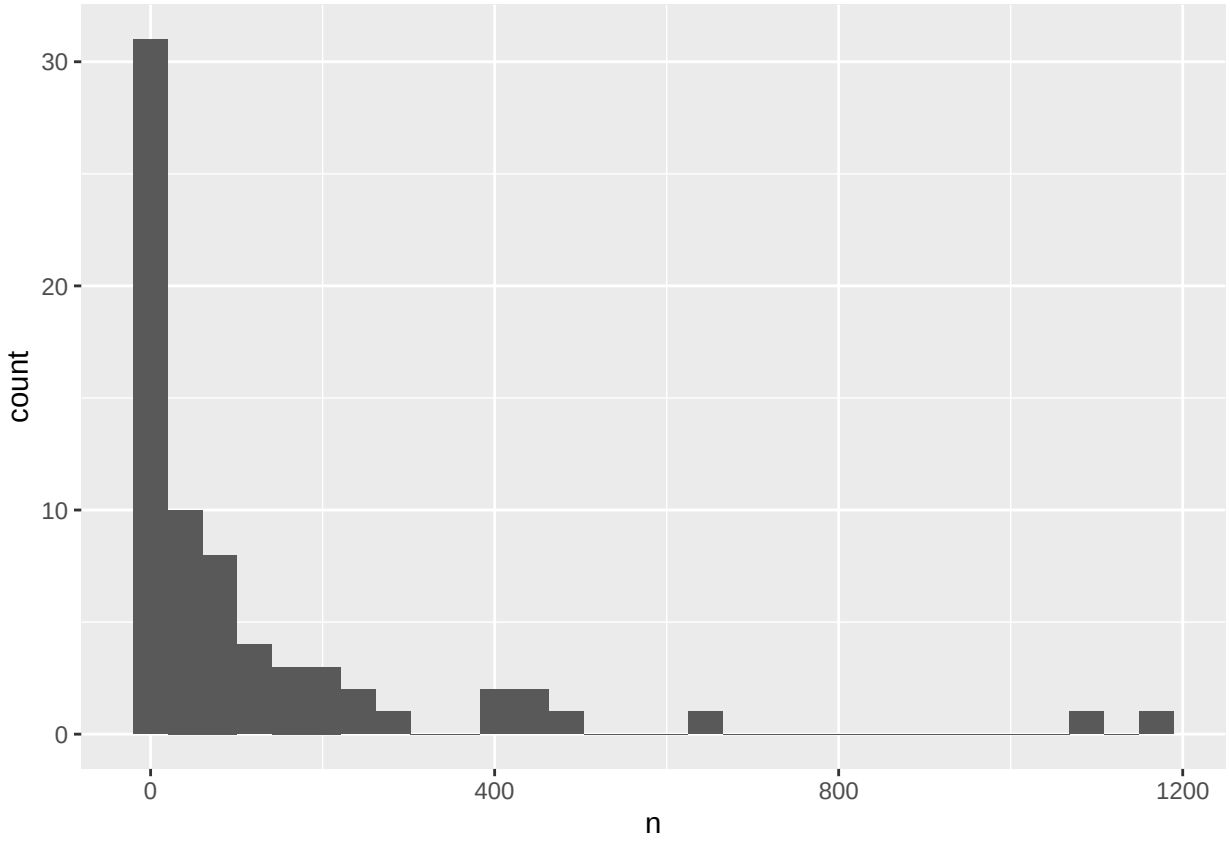
```
legume$Crop_Species_Scientific_Name %>% unique()
```

```
##  [1] "Cicer arietinum"           "Vicia faba"
##  [3] "Lens culinaris"            "Pisum sativum"
##  [5] "Vigna radiata"             "Arachis hypogaea"
##  [7] "Glycine max"               "Triticum aestivum"
##  [9] "Lupinus albus"             "Lupinus angustifolius"
## [11] "Lupinus atlanticus"        "Lupinus pilosus"
```

```
## [13] "Vicia narbonensis"         "Lathyrus cicera"
## [15] "Lathyrus ochrus"           "Lathyrus sativus"
## [17] "Vicia benghalensis"        "Vicia sativa"
## [19] "Lupinus luteus"            "Avena sativa"
## [21] "Brassica napus"            "Brassica rapa"
## [23] "Brassica juncea"           "Triticum turgidum"
## [25] "Crambe abyssinica"         "Carthamus tinctorius"
## [27] "Helianthus annuus"         "Phaseolus vulgaris"
## [29] "Hordeum vulgare"           "Triticum sativum"
## [31] "Vigna subterranea"         "Vigna unguiculata"
## [33] "Cajanus cajan"             "Sorghum bicolor"
## [35] "Linum usitatissimum"       "Fallow"
## [37] "Panicum miliaceum"         "Triticum durum"
## [39] "Ã-Triticosecale"           "Vigna mungo"
## [41] "Oryza sativa"              "Zea mays"
## [43] "Vigna aconitifolia"        "Cyamopsis tetragonoloba"
## [45] "Lablab purpureus"          "Phaseolus lunatus"
## [47] "Solanum lycopersicum"      "Ipomoea batatas"
## [49] "Fagopyrum esculentum"      "Brassica chinensis"
## [51] "Vicia ervilia"             "Brassica campestris"
## [53] "Pennisetum glaucum"        "Vigna angularis"
## [55] "Vicia pannonica"           "Secale cereale"
## [57] "Lathyrus aphaca"           "Lathyrus clymenum"
## [59] "Trigonella foenum-graecum" "Vicia articulata"
## [61] "Vicia villosa"             "Vicia hybrida"
## [63] "Ricinus communis"          "Trifolium repens"
## [65] "Sinapis alba"              "Lupinus mutabilis"
## [67] "Gossypium hirsutum"        "Macrotyloma uniflorum"
## [69] "Sesamum indicum"           "Citrullus lanatus"
```

```
legume %>%
  group_by(Crop_Species_Scientific_Name) %>%
  summarise(n = length(Crop_Yield_Grain)) %>%
  ggplot(aes(n))+
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

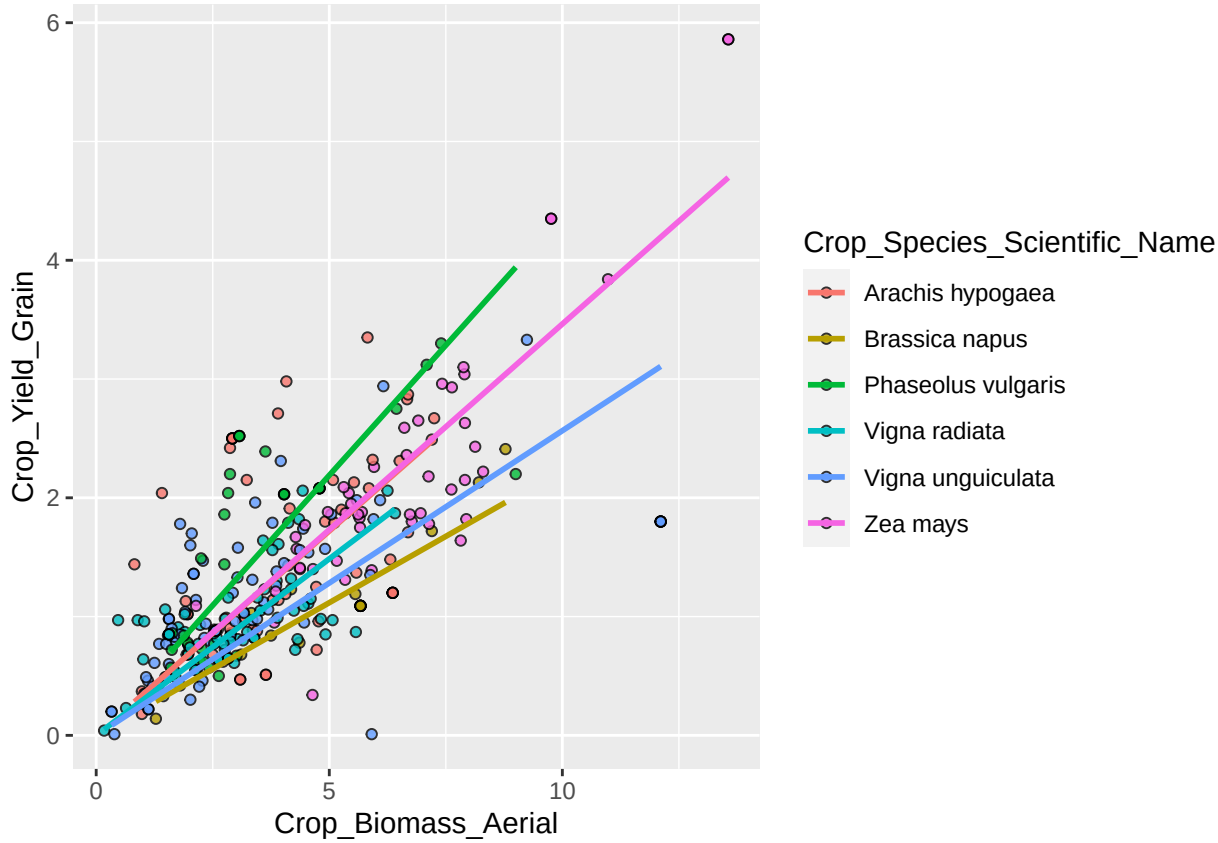



```
legume1 <- legume %>%
  group_by(Crop_Species_Scientific_Name) %>%
  mutate(n = length(Crop_Yield_Grain)) %>%
  ungroup() %>%
  filter(n %>% between(200, 300)) # n >= 200 & n <=300
```

```
legume1 %>%
  drop_na(Crop_Biomass_Aerial) %>%
  ggplot(aes(Crop_Biomass_Aerial, Crop_Yield_Grain))+
  geom_point(aes(fill = Crop_Species_Scientific_Name),
             shape = 21,
             alpha = .8)+
  geom_smooth(aes(color = Crop_Species_Scientific_Name),
              method = 'lm', se = FALSE, formula = y~x-1)
```

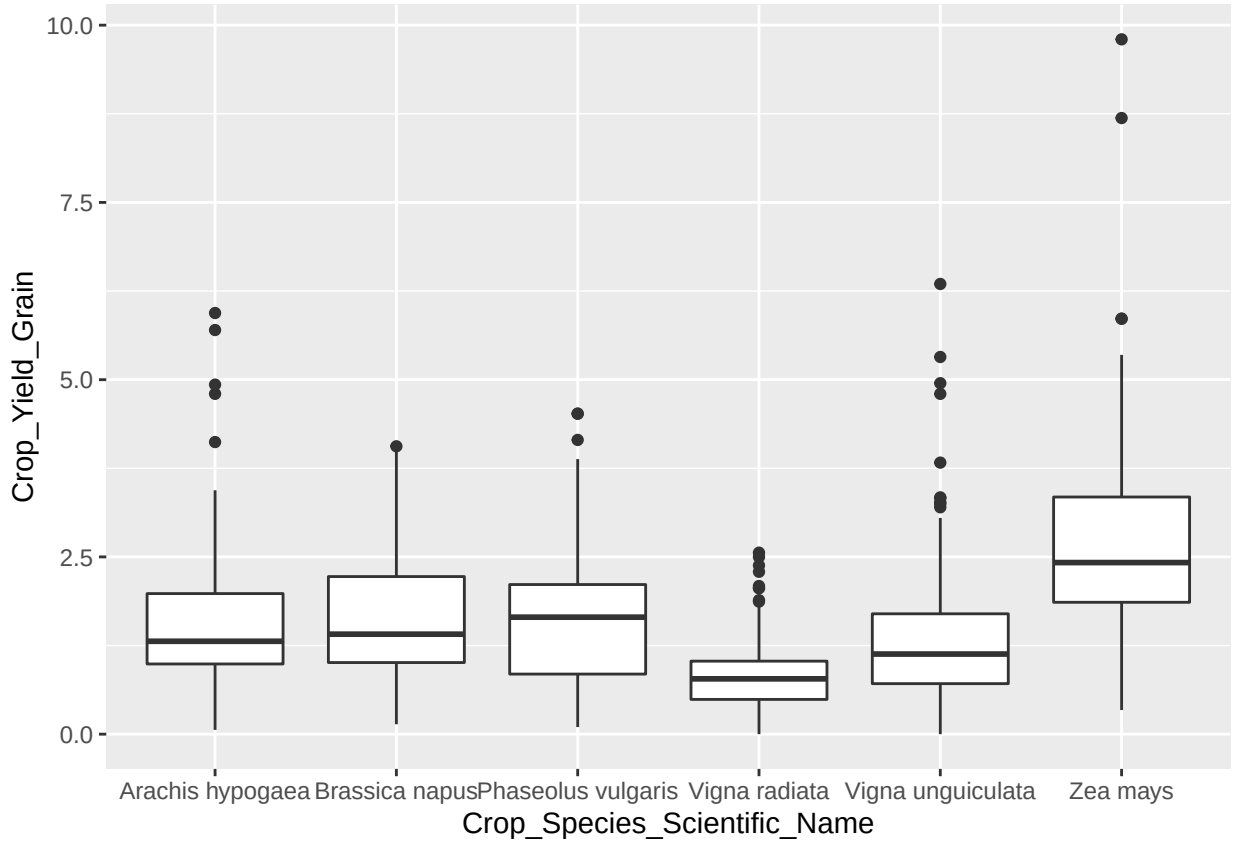
```
## Warning: Removed 4 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 4 rows containing missing values (geom_point).
```




```
legume1 %>%
  ggplot(aes(x = Crop_Species_Scientific_Name, y = Crop_Yield_Grain))+
  geom_boxplot()
```

```
## Warning: Removed 45 rows containing non-finite values (stat_boxplot).
```




```
 # y ~ 1 + x
m1 <- lm(Crop_Yield_Grain ~ Crop_Species_Scientific_Name -1,
         data = legume1)
summary(m1)
```

```
##
## Call:
## lm(formula = Crop_Yield_Grain ~ Crop_Species_Scientific_Name -
##     1, data = legume1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3883 -0.5783 -0.1741  0.4218  7.0717
##
## Coefficients:
##                                            Estimate Std. Error t value
## Crop_Species_Scientific_NameArachis hypogaea  1.52819    0.05613   27.23
## Crop_Species_Scientific_NameBrassica napus    1.70263    0.06230   27.33
## Crop_Species_Scientific_NamePhaseolus vulgaris  1.61562    0.06019   26.84
## Crop_Species_Scientific_NameVigna radiata     0.83562    0.06301   13.26
## Crop_Species_Scientific_NameVigna unguiculata  1.31787    0.06561   20.09
## Crop_Species_Scientific_NameZea mays          2.72831    0.06893   39.58
##                                            Pr(>|t|)
## Crop_Species_Scientific_NameArachis hypogaea   <2e-16 ***
## Crop_Species_Scientific_NameBrassica napus     <2e-16 ***
## Crop_Species_Scientific_NamePhaseolus vulgaris   <2e-16 ***
```

```
## Crop_Species_Scientific_NameVigna radiata        <2e-16 ***
## Crop_Species_Scientific_NameVigna unguiculata    <2e-16 ***
## Crop_Species_Scientific_NameZea mays             <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9325 on 1338 degrees of freedom
##   (45 observations deleted due to missingness)
## Multiple R-squared:  0.765,  Adjusted R-squared:  0.7639
## F-statistic: 725.8 on 6 and 1338 DF,  p-value: < 2.2e-16
```

```
# anova()

# Prefer using type III tests
car::Anova(m1, type = 3)
```

```
## Anova Table (Type III tests)
##
## Response: Crop_Yield_Grain
##                              Sum Sq   Df F value    Pr(>F)
## Crop_Species_Scientific_Name 3786.3    6  725.78 < 2.2e-16 ***
## Residuals                    1163.4 1338
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

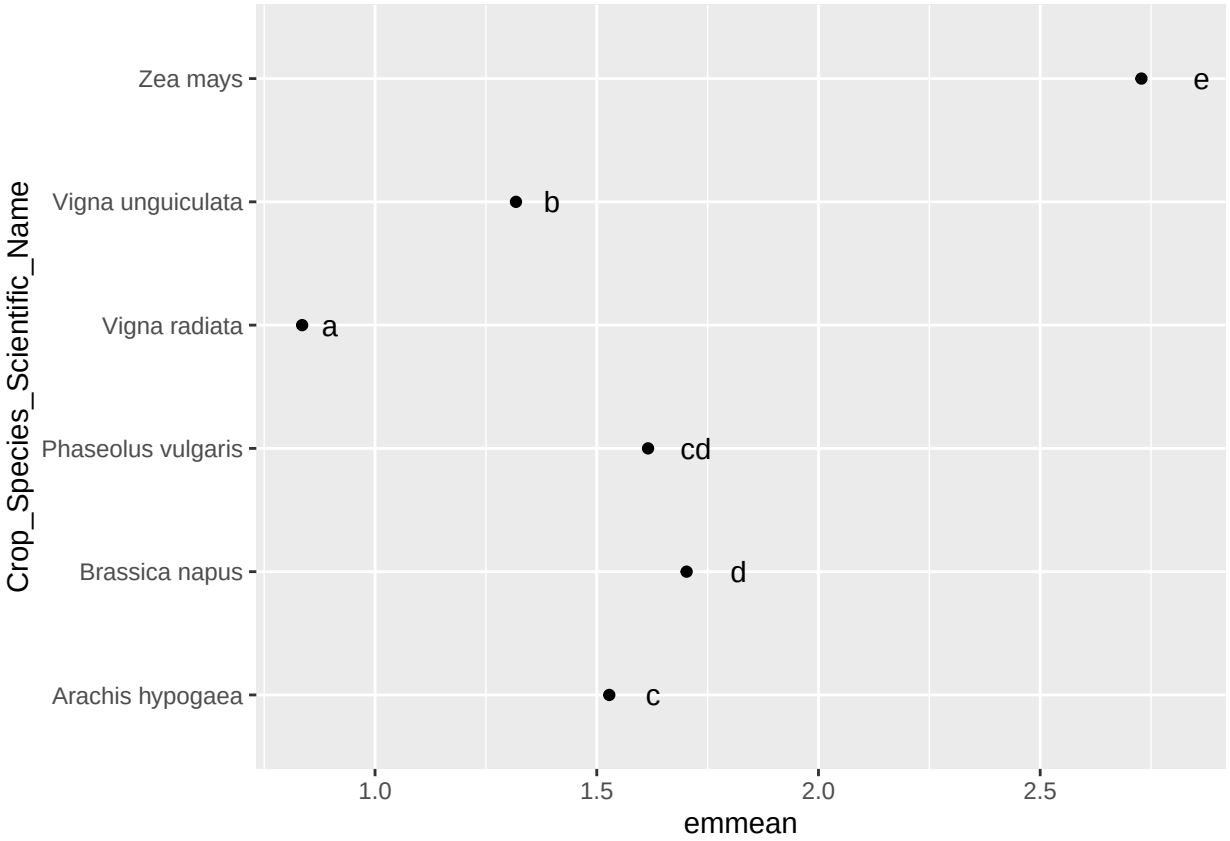
```
# reject H0: at least one mean is different
```

```
legume_means <- emmeans(m1, specs = ~ Crop_Species_Scientific_Name)

legume_means_comp <- cld(legume_means,
                         level = .05,
                         decreasing = TRUE,
                         adjust = "none",
                         Letters = "abcdefgh")
class(legume_means_comp)
```

```
## [1] "summary_emm" "data.frame"
```

```
legume_means_comp %>%
  ggplot(aes(emmean, Crop_Species_Scientific_Name))+
  geom_errorbarh(aes(xmin = lower.CL, xmax = upper.CL), height = 0)+
  geom_point(aes())+
  geom_text(aes(label = .group), nudge_x = 0.09)
```


Zea mays
Vigna unguiculata
Vigna radiata
Phaseolus vulgaris
Brassica napus
Arachis hypogaea
Crop_Species_Scientific_Name
e
b
a
cd
d
c
1.0
1.5
2.0
2.5
emmean