Homework: Coding Notes Data Visualization 2
PLPA-5820 (Spring 2025)

Gabriel A. A. Silva

2025-02-20

Disclaimer

Since I mixed up the assignments last week and submitted the "Homework: Coding Notes Data Visualization 2" as the "Homework: Coding Notes Data Visualization 1", I am ressubmitting the same assignment for the second week. I again apologize for the confusion, please let me know if you need me to redo the assignment.

Instructions

Watch the videos associated with the introduction to data visualization in R and turn in your notes (R code).

Notes

I was familiar with the packages before. The demonstrated way to add significance letters was interesting, I'm used to doing differently but I think I'll start using this way.

I am unsure if the package `tidyplots` is known by you, but I think it is a good package for data visualization by people not familiar with ggplot2's syntax. You can find more info at https://tidyplots.org/.

Adding significance letters to represent multiple pairwise comparisons of an ANOVA

```
# Load libraries
library(tidyverse)
```

```
-- Attaching core tidyverse packages ---------------------- tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.5.1     v tibble    3.2.1
v lubridate 1.9.4     v tidyr     1.3.1
v purrr     1.0.4
-- Conflicts ----------------------------------------- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
```

```r
library(lme4)
```

```
Loading required package: Matrix

Attaching package: 'Matrix'

The following objects are masked from 'package:tidyr':

    expand, pack, unpack
```

```r
library(emmeans)
```

```
Welcome to emmeans.
Caution: You lose important information if you filter this package's results.
See '? untidy'
```

```r
library(multcomp)
```

```
Loading required package: mvtnorm
Loading required package: survival
Loading required package: TH.data
Loading required package: MASS

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
```

```r
library(multcompView)

# Load data
STAND <- read.csv("Bull_richness.csv", na.strings = "na")

# Filter data
ave_stand <- STAND %>%
  group_by(Crop, Collection, Rep, Fungicide, Treatment) %>%
  summarize(
    ave.stand = mean(richness, na.rm=TRUE))
```

```
`summarise()` has grouped output by 'Crop', 'Collection', 'Rep', 'Fungicide'.
You can override using the `.groups` argument.
```

```r
# Run ANOVA
lm <- lmer(ave.stand ~ Treatment * Crop * Fungicide + (1|Rep), data = ave_stand)
```

```
boundary (singular) fit: see help('isSingular')
```

```r
# estimate lsmeans of variety within siteXyear
lsmeans <- emmeans(lm, ~ Treatment + Fungicide | Crop )

# contrast with Tukey ajustment
Results_lsmeansEC <- multcomp::cld(lsmeans, alpha = 0.05, reversed = TRUE,
                                   details = TRUE,  Letters = letters)
Results_lsmeansEC
```

```
$emmeans
Crop = Corn:
 Treatment Fungicide emmean   SE   df lower.CL upper.CL .group
 No-till   F          12.92 1.52 70.3     9.88     16.0  a
 No-till   C          12.19 1.52 70.3     9.16     15.2  a
 Conv.     C          10.47 1.52 70.3     7.44     13.5  a
```

```
 Conv.      F               7.36 1.52 70.3      4.32      10.4  a

Crop = Soy:
 Treatment Fungicide emmean   SE   df lower.CL upper.CL .group
 Conv.      C             18.81 1.52 70.3    15.77     21.8  a
 No-till    C             18.58 1.52 70.3    15.55     21.6  ab
 No-till    F             13.15 1.52 70.3    10.12     16.2   bc
 Conv.      F             12.14 1.52 70.3     9.10     15.2    c

Degrees-of-freedom method: kenward-roger
Confidence level used: 0.95
P value adjustment: tukey method for comparing a family of 4 estimates
significance level used: alpha = 0.05
NOTE: If two or more means share the same grouping symbol,
      then we cannot show them to be different.
      But we also did not show them to be the same.

$comparisons
Crop = Corn:
 contrast                  estimate   SE df t.ratio p.value
 Conv. C - Conv. F            3.111 2.15 85   1.445  0.4754
 (No-till C) - Conv. F        4.833 2.15 85   2.244  0.1197
 (No-till C) - Conv. C        1.722 2.15 85   0.800  0.8543
 (No-till F) - Conv. F        5.556 2.15 85   2.580  0.0553
 (No-till F) - Conv. C        2.444 2.15 85   1.135  0.6690
 (No-till F) - (No-till C)    0.722 2.15 85   0.335  0.9869

Crop = Soy:
 contrast                  estimate   SE df t.ratio p.value
 (No-till F) - Conv. F        1.014 2.15 85   0.471  0.9653
 (No-till C) - Conv. F        6.444 2.15 85   2.993  0.0186
 (No-till C) - (No-till F)    5.431 2.15 85   2.522  0.0637
 Conv. C - Conv. F            6.667 2.15 85   3.096  0.0139
 Conv. C - (No-till F)        5.653 2.15 85   2.625  0.0495
 Conv. C - (No-till C)        0.222 2.15 85   0.103  0.9996

Degrees-of-freedom method: kenward-roger
P value adjustment: tukey method for comparing a family of 4 estimates
```

```r
# Extracting the letters for the bars
sig.diff.letters <- data.frame(Results_lsmeansEC$emmeans$Crop,
                               Results_lsmeansEC$emmeans$Treatment,
```

```
                                Results_lsmeansEC$emmeans$Fungicide,
                                str_trim(Results_lsmeansEC$emmeans$.group))
colnames(sig.diff.letters) <- c("Crop",
                                "Treatment",
                                "Fungicide",
                                "Letters")

# for plotting with letters from significance test
ave_stand2 <- ave_stand %>%
  group_by(Crop, Treatment, Fungicide) %>%
  dplyr::summarize(
    ave.stand2 = mean(ave.stand, na.rm=TRUE),
    se = sd(ave.stand)/sqrt(4)) %>%
  left_join(sig.diff.letters)
```

`summarise()` has grouped output by 'Crop', 'Treatment'. You can override using
the `.groups` argument.
Joining with `by = join_by(Crop, Treatment, Fungicide)`

```
### Stand bars ####
ggplot(ave_stand, aes(x = Crop, y = ave.stand)) +
  stat_summary(fun=mean,geom="bar") +
  stat_summary(fun.data = mean_se, geom = "errorbar", width = 0.5) +
  ylab("Richness") +
  geom_jitter(width = 0.02, alpha = 0.5) +
  geom_text(data = ave_stand2,
            aes(label = Letters, y = ave.stand2+(3*se)),
            vjust = -0.5) +
  xlab("")+
  theme_classic() +
  theme(
    strip.background = element_rect(color="white", fill="white",
                                    size=1.5, linetype="solid"),
    strip.text.x = element_text(size = 12, color = "black"),
    axis.text.x = element_text(angle = 45, vjust = 1, hjust=1)) +
  facet_wrap(~Treatment + Fungicide, scales = "free_y")
```

Warning: The `size` argument of `element_rect()` is deprecated as of ggplot2 3.4.0.
i Please use the `linewidth` argument instead.