Class Activity 4

Courtney Hodge

2024-07-19

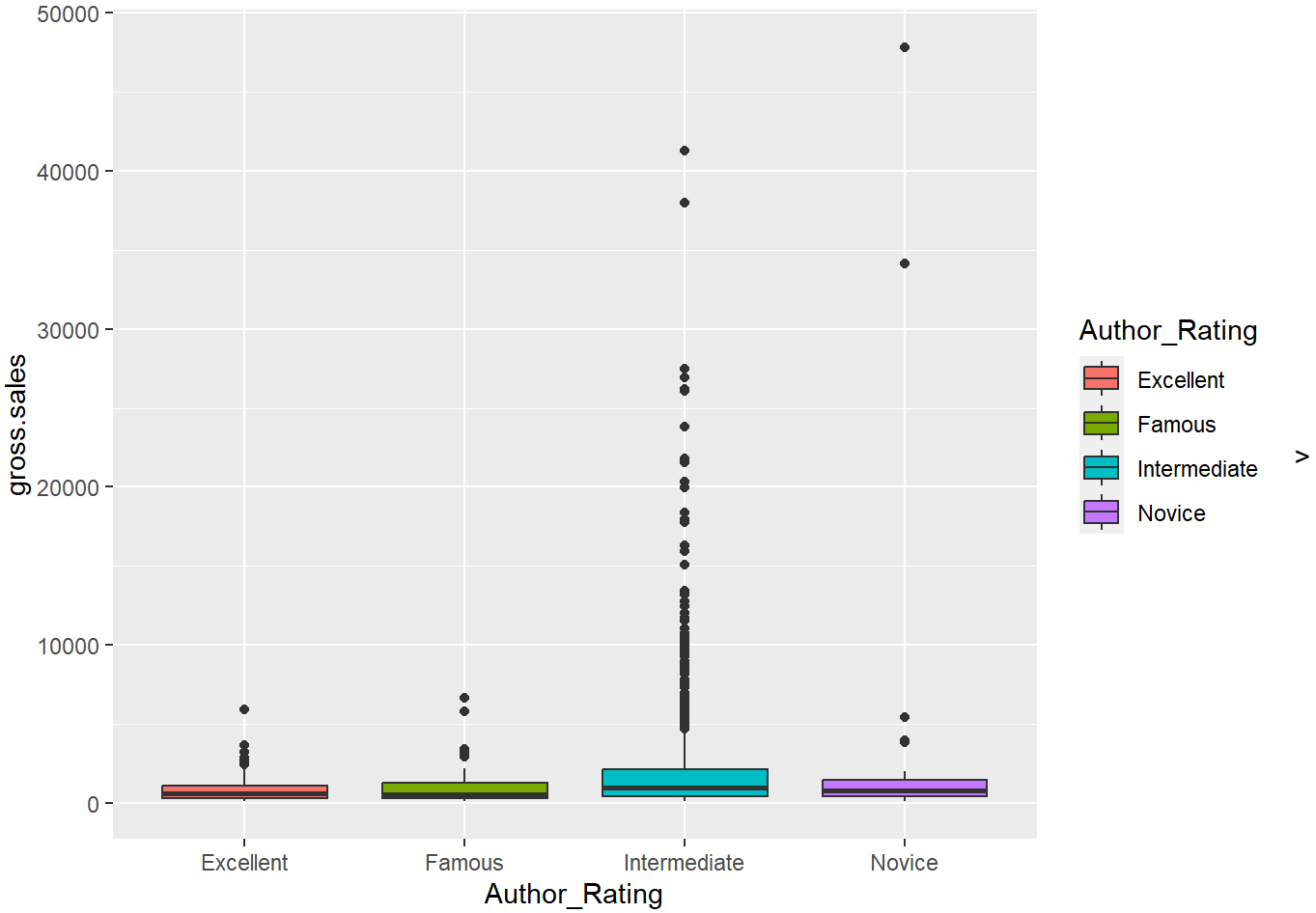
```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.3     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.4     ✓ tibble    3.2.1
## ✓ lubridate 1.9.3     ✓ tidyr     1.3.0
## ✓ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
e errors
```

Startups Data

```
BookData <- read.csv("C:\\Users\\hodge\\Downloads\\Books_Data_Clean.csv")
```

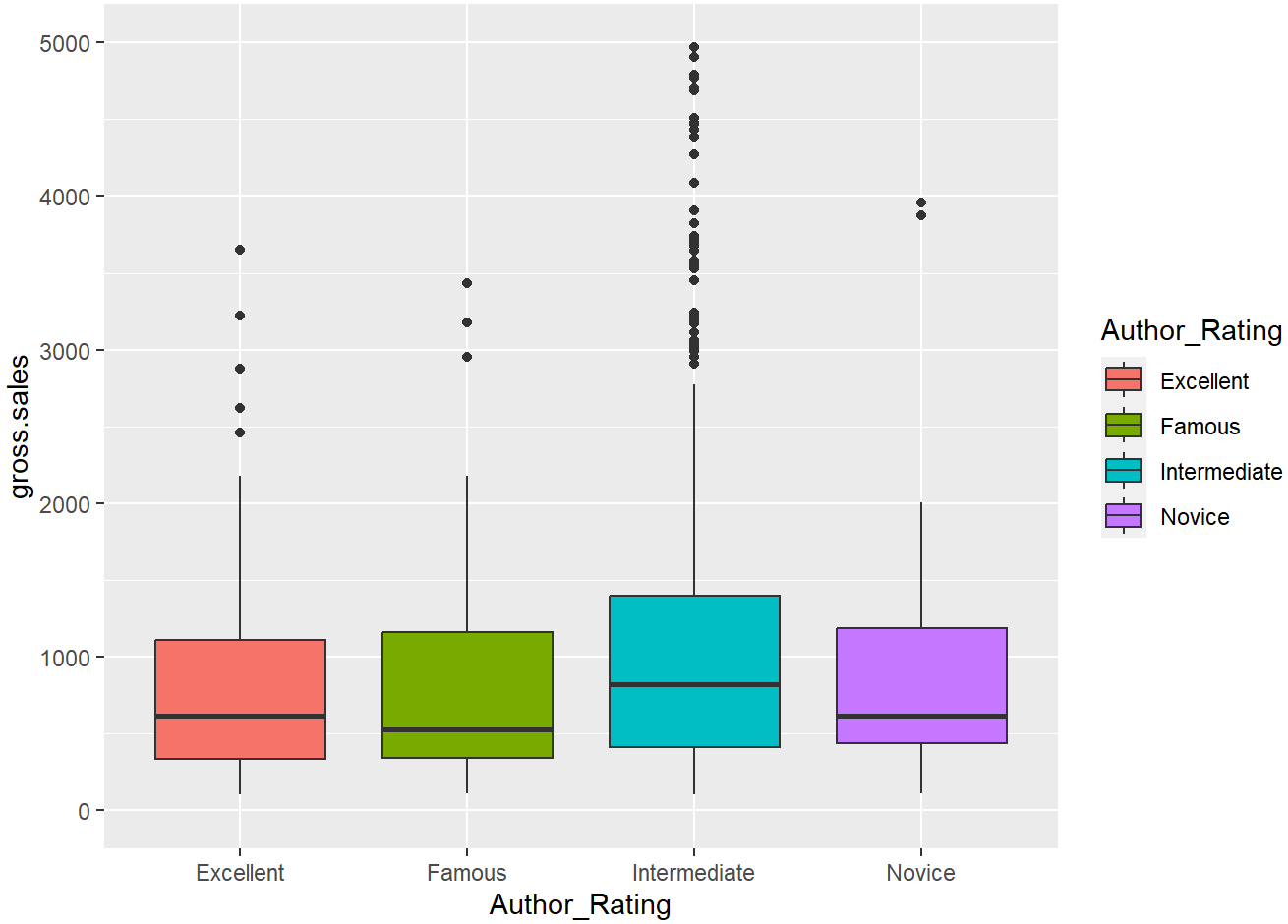
```
ggplot(BookData, aes(x = Author_Rating, y = gross.sales, fill = Author_Rating )) +
  geom_boxplot()
```



zooming into graph

```
ggplot(BookData, aes(x = Author_Rating, y = gross.sales, fill = Author_Rating )) +
  geom_boxplot() + ylim(0, 5000)
```

```
## Warning: Removed 91 rows containing non-finite values (`stat_boxplot()`).
```



Hypotheses

H0: Mu A = Mu B = Mu C = Mu D HA: At least one of them are different

```
anova <- aov(gross.sales~Author_Rating, data = BookData)

summary(anova)
```

```
##                Df    Sum Sq   Mean Sq F value   Pr(>F)
## Author_Rating    3 7.881e+08 262704446   17.75 3.03e-11 ***
## Residuals     1066 1.578e+10  14803673
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since our p-value of 3.03e-11 is incredibly small and less than 5%, then p-val < 0.05 means that we reject the null-hypothesis.

multiple hypothes testing

MuFamous-MuExcellent

MuIntermediate-MuExcellent

MuNovice-MuExcellent

MuIntermediate-MuFamous

MuNovice-MuFamous

MuNovice-MuIntermediate

```
TukeyHSD(anova, conf.level = 0.95)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = gross.sales ~ Author_Rating, data = BookData)
##
## $Author_Rating
##                              diff        lwr      upr     p adj
## Famous-Excellent         373.6263 -1082.4068 1829.659 0.9119276
## Intermediate-Excellent  1679.8549  1025.9664 2333.743 0.0000000
## Novice-Excellent        3034.1830  1153.2751 4915.091 0.0002092
## Intermediate-Famous     1306.2286  -110.1401 2722.597 0.0829421
## Novice-Famous           2660.5567   398.6229 4922.491 0.0134907
## Novice-Intermediate     1354.3281  -496.0455 3204.702 0.2358077
```

MuFamous-MuExcellent –> - > + (inconclusive)

MuIntermediate-MuExcellent –> + > +

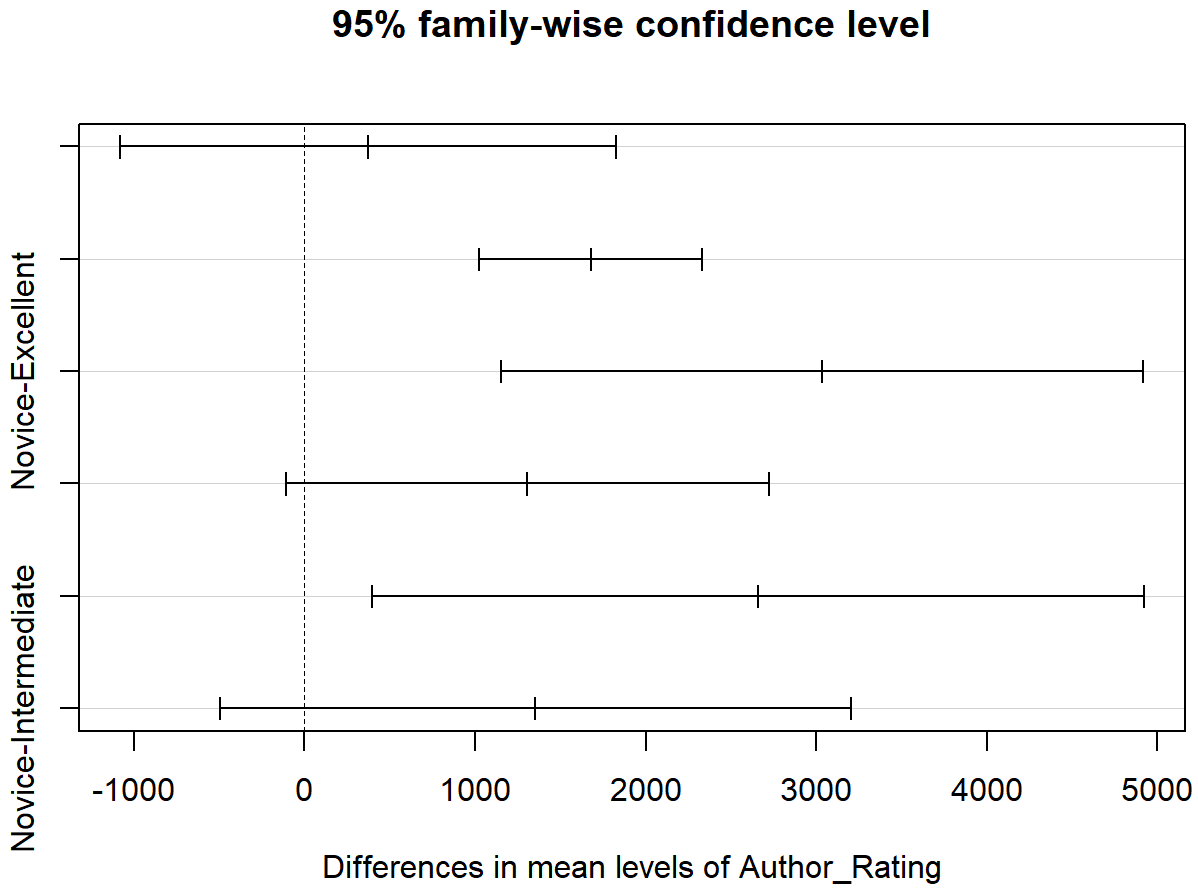
MuNovice-MuExcellent –> + > +

MuIntermediate-MuFamous –> - > + (Incolcusive)

MuNovice-MuFamous –> + > +

MuNovice-MuIntermediate –> - > + (incolcusive)

```
plot(TukeyHSD(anova, conf.level = 0.95))
```



Based on the analysis above, it looks like drug A is the most effective drug in this experiment. we found that by looking at the lower and upper bounds and noticing that A is positive, but C and B are negative.

We have enough statistical evidence to conclude that the mean pain rating for Drug A is much lower than Drugs B and C. Becuase this is an experiment, Drug A is more effective in treating migraine headaches compared to Drugs B and C