Case Study #1

Jose Zavala

2023-01-30

Abstract

Researchers Blackman and Catalina aimed to understand the relationship between stages of full moon and any possible effects on mental health hospital admissions. There has been suggestions that a full moon has an effect on individuals behavior and therefore could potentially affect their mental health. Admissions to a mental hospital were recorded for a period of a full calendar year to assess for mental health effects. To measure the effect of a full moon in mental health, three measurements of moon progression were measured: average number of admission recordings to a mental health hospital before, during and after a full moon. Considering there is a full moon every month we have three admissions measurements per month. A two-way ANOVA was performed to compare the effect of different stages of full moon and month effects on mental health hospital admissions. The two-way ANOVA found that there was a statistically significant difference in the mean number of admissions to the hospital for both moon phase and month.

Analysis

Data

Month	Before	During	After
Jan	10.4	9	12.9
Feb	11.5	13	13.5
Mar	13.8	16	13.1
Apr	15.4	25	15.8
May	15.7	14	13.3
Jun	11.7	14	12.8
Jul	15.8	20	14.5
Aug	6.4	5	5.8
Sep	7.1	13	9.2
Oct	6.5	14	7.9
Nov	8.6	12	7.7
Dec	8.1	6	11.0

The purpose of this study was determine what effect, if any, does the full moon have on the mental health of individuals. For this study, Blackman and Catalina set out to collect data of individuals who presented to the psychiatric emergency room at the North Richmond Community Mental Health Center in Staten Island, New York. They looked at data from a 12 month period ranging from August, 1971 through July 1972 and obtained the average number of patients seen 4-13 days before the full moon, during the full moon, and 4-13 days after the full moon.

In this study the population of interest could be represented by all individuals who are dealing with mental health related issues and subsequent sample would be the people who presented themselves to the psychiatric

emergency room. This analysis focuses on 3 variables: mean number of admissions (continuous), the month (categorical/nominal), and phase of moon (categorical). The month variable has 12 levels, one for each month, and the phase of moon variable has 3 levels. The corresponding moon levels are: before, during, and after the full moon.

We compute the overall mean number of admissions for the entire sample within the 12 months and find that the mean number of admissions is 11.93. The remainder of the summary statistics can be found on Table 1. Next also computed the summary statistics for each month as well as for each moon phase summarized in Tables 2 and 3 respectively. For 6 months we can observe that the average number of admissions is higher than the overall mean for the entire 12 months. When grouped by phase of moon we observe a higher average during and after the full moon when compared to the overall average.

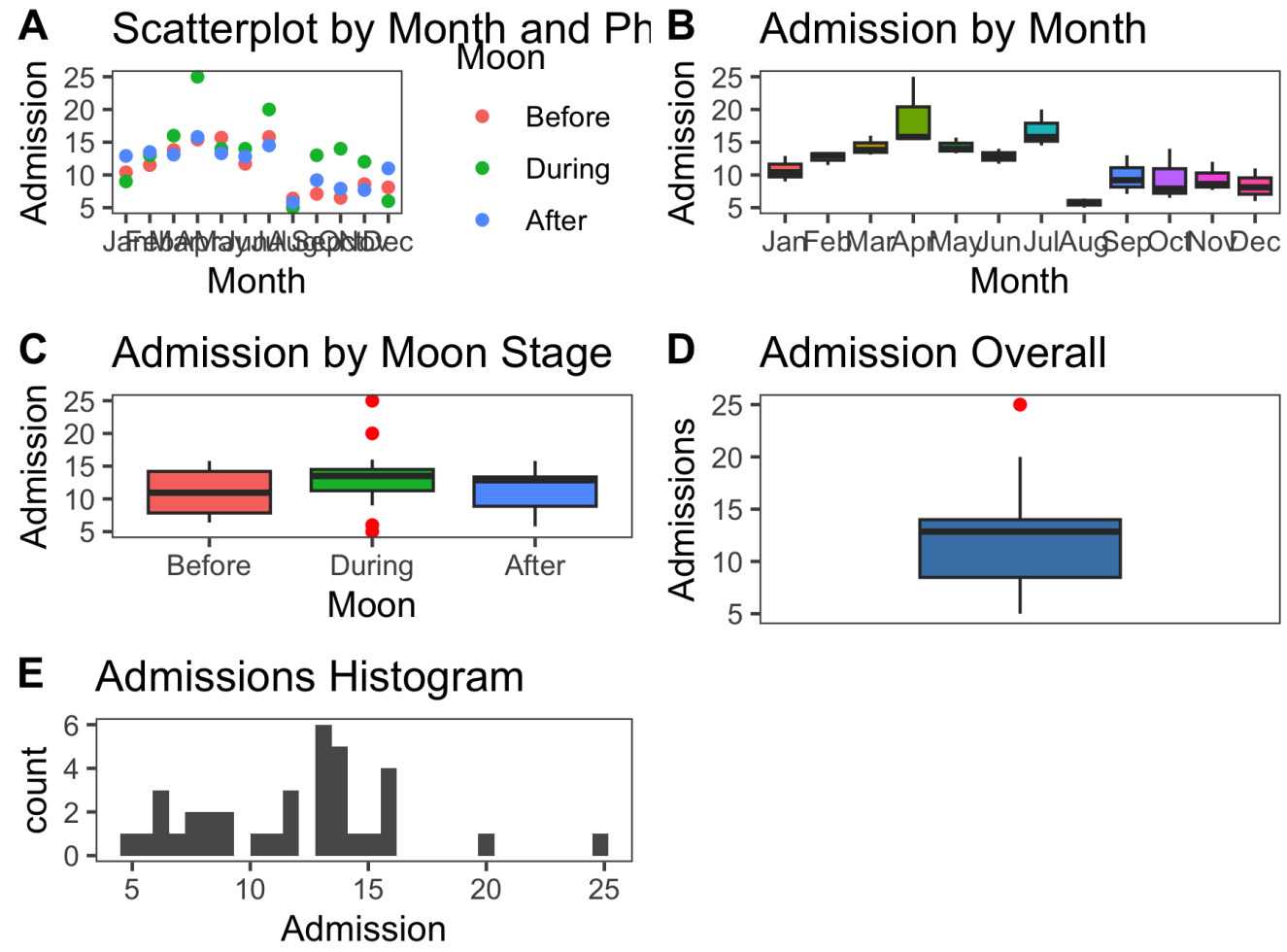
```
## # A tibble: 3 x 6
##   Moon    Mean StandDev Median   Min   Max
##   <fct>  <dbl>    <dbl>  <dbl> <dbl> <dbl>
## 1 Before  10.9     3.62   11.0   6.4  15.8
## 2 During  13.4     5.5    13.5   5    25
## 3 After   11.5     3.11   12.8   5.8  15.8
```

```
## # A tibble: 12 x 6
##    Month  Mean StandDev Median   Min   Max
##    <fct> <dbl>    <dbl>  <dbl> <dbl> <dbl>
##  1 Jan   10.8      1.98   10.4   9    12.9
##  2 Feb   12.7      1.04   13    11.5  13.5
##  3 Mar   14.3      1.51   13.8  13.1  16
##  4 Apr   18.7      5.43   15.8  15.4  25
##  5 May   14.3      1.23   14    13.3  15.7
##  6 Jun   12.8      1.15   12.8  11.7  14
##  7 Jul   16.8      2.87   15.8  14.5  20
##  8 Aug    5.73     0.7     5.8   5     6.4
##  9 Sep    9.77     2.99    9.2   7.1  13
## 10 Oct    9.47     3.99    7.9   6.5  14
## 11 Nov    9.43     2.27    8.6   7.7  12
## 12 Dec    8.37     2.51    8.1   6    11
```

```
## # A tibble: 1 x 5
##    Mean StandDev Median   Min   Max
##   <dbl>    <dbl>  <dbl> <dbl> <dbl>
## 1  11.9     4.23   12.8     5    25
```

Figure A shows a scatter plot of the mean number of admissions by month and for each of the 3 moon phases. We can observe that the during moon phase appears to have higher mean admissions for more months. This is corroborated by the boxplots shown in Figure C which show that the higher mean as well as the outliers detected for the during moon phase.

Figure B shows boxplots by month and we observe that from August through December there are lower average number of admissions while the peak occurs in April. Figures D and E visualize the distribution of the average number of admissions. The histogram shows a vague bell shape curve although it could be interpreted as a slight right skew.



Model

We conducted a two way ANOVA model to test the impact that both categorical variables have on the average admissions with the subsequent null hypotheses:

$H01$: Mean for Admission is equal for all levels of Moon phase

$H02$: Mean for Y is equal for all levels of Month

$H03$: Mean for Admission is not interacted by Moon phase and Month

We will not be testing null hypothesis 3 which is testing for an interaction since such a model would be a saturated model and is of no use statistically.

We conduct a two way ANOVA using Admission as the response variable with Moon phase and Month as the predictor variables at a confidence level of 0.05. Both predictors are significant with Moon having a p-value of 0.0453 and 0.0000508 for Month. We can reject both null hypotheses and find that there is a difference in means between both phases of the moon and varying by month.

We can also see how each category for both predictors influences admissions by looking at the estimated coefficients. We observe "During" and "After" categories add to average admissions when compared to before. For the "month" predictor we see that the reference month is January and February, March, April, May, June, July all add to average admissions while August, September, October, November, and December all subtract from it. The months that seem to have the strongest impact are April and August.

```
two.way = aov(Admission ~ Moon + Month, data = df)
summary(two.way)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Moon         2   41.5   20.76   3.573   0.0453 *
## Month       11  455.6   41.42   7.129 5.08e-05 ***
## Residuals   22  127.8    5.81
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

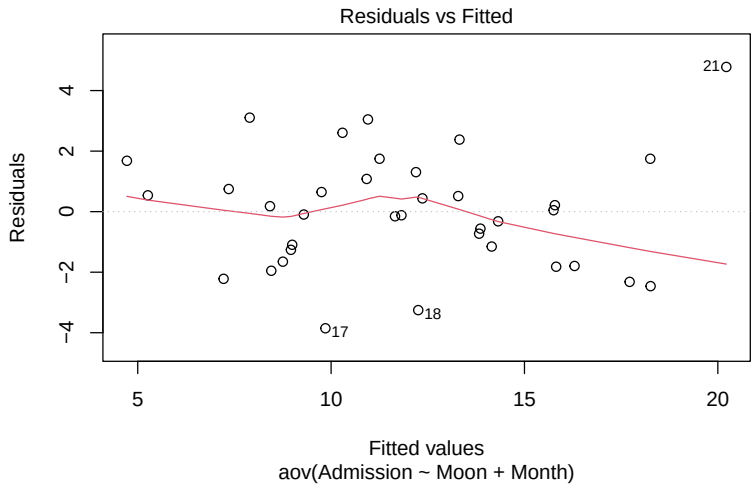
```
two.way$coefficients
```

```
## (Intercept)  MoonDuring   MoonAfter    MonthFeb    MonthMar    MonthApr
##   9.7527778   2.5000000   0.5416667   1.9000000   3.5333333   7.9666667
##    MonthMay    MonthJun    MonthJul    MonthAug    MonthSep    MonthOct
##   3.5666667   2.0666667   6.0000000  -5.0333333  -1.0000000  -1.3000000
##    MonthNov    MonthDec
##  -1.3333333  -2.4000000
```

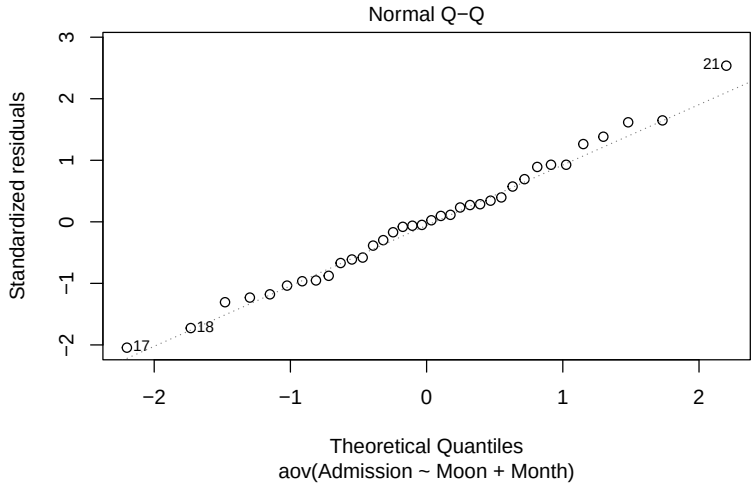
There are 3 model assumptions we need to verify for a two way ANOVA model. We did not see an extreme skew in the histogram plot of the average number of admissions so we will conclude normality being met. We will assume independent samples, although this cannot be ascertained due to the manner in which the data was collected and presents an opportunity for improvement. The homogeneity of variance assumption will be tested for both Moon and Month variables using Levene's Test at an alpha level of 0.05 and in fact the residuals vs fitted plot of the model shows possible signs of concern. Both of the Levene tests for homogeneity of variance conclude that we can assume equal variances since neither p-value is smaller than our alpha level, thus no evidence to conclude heterogeneity.

1. The populations from which the samples were obtained must be normally or approximately normally distributed.
2. The samples must be independent.
3. The variances of the populations must be equal.

```
#check model assumptions via plots. QQplot looks approx normal. Residuals vs fitted (homogeneity of var
plot(two.way,1)
```




```
plot(two.way,2)
```




```
#levene test for equal variances. Both do not reject null, can assume equal variances
leveneTest(Admission ~ Month, data = df)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group 11  0.4326 0.9256
##       24
```

```
leveneTest(Admission ~ Moon, data = df)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  2  0.7122  0.498
##       33
```

Next we will conduct a Tukey's test to see where the significance differences lie.

```
HSD.test(two.way,"Moon",console=TRUE, group = T)
```

```
##
## Study: two.way ~ "Moon"
##
## HSD Test for Admission
##
## Mean Square Error:  5.809975
##
## Moon,  means
##
##        Admission      std  r Min  Max
## After   11.45833 3.109358 12 5.8 15.8
## Before  10.91667 3.619853 12 6.4 15.8
## During  13.41667 5.501377 12 5.0 25.0
##
## Alpha: 0.05 ; DF Error: 22
## Critical Value of Studentized Range: 3.552594
##
## Minimun Significant Difference: 2.471963
##
## Treatments with the same letter are not significantly different.
##
##        Admission groups
## During  13.41667      a
## After   11.45833     ab
## Before  10.91667      b
```

There are significant differences in the average admissions during the full moon and before, but not after the full moon.

```
monthresults = HSD.test(two.way,"Month",console=TRUE, group = T)
```

```
##
## Study: two.way ~ "Month"
##
## HSD Test for Admission
##
## Mean Square Error:  5.809975
##
## Month,  means
##
##     Admission       std r  Min  Max
## Apr 18.733333 5.4307765 3 15.4 25.0
## Aug  5.733333 0.7023769 3  5.0  6.4
## Dec  8.366667 2.5106440 3  6.0 11.0
## Feb 12.666667 1.0408330 3 11.5 13.5
## Jan 10.766667 1.9756855 3  9.0 12.9
## Jul 16.766667 2.8746014 3 14.5 20.0
## Jun 12.833333 1.1503623 3 11.7 14.0
## Mar 14.300000 1.5132746 3 13.1 16.0
```

```
## May 14.333333 1.2342339 3 13.3 15.7
## Nov  9.433333 2.2678918 3  7.7 12.0
## Oct  9.466667 3.9878984 3  6.5 14.0
## Sep  9.766667 2.9905406 3  7.1 13.0
##
## Alpha: 0.05 ; DF Error: 22
## Critical Value of Studentized Range: 5.144324
##
## Minimun Significant Difference: 7.15904
##
## Treatments with the same letter are not significantly different.
##
##     Admission groups
## Apr 18.733333      a
## Jul 16.766667     ab
## May 14.333333    abc
## Mar 14.300000    abc
## Jun 12.833333   abcd
## Feb 12.666667   abcd
## Jan 10.766667    bcd
## Sep  9.766667    bcd
## Oct  9.466667     cd
## Nov  9.433333     cd
## Dec  8.366667     cd
## Aug  5.733333      d
```

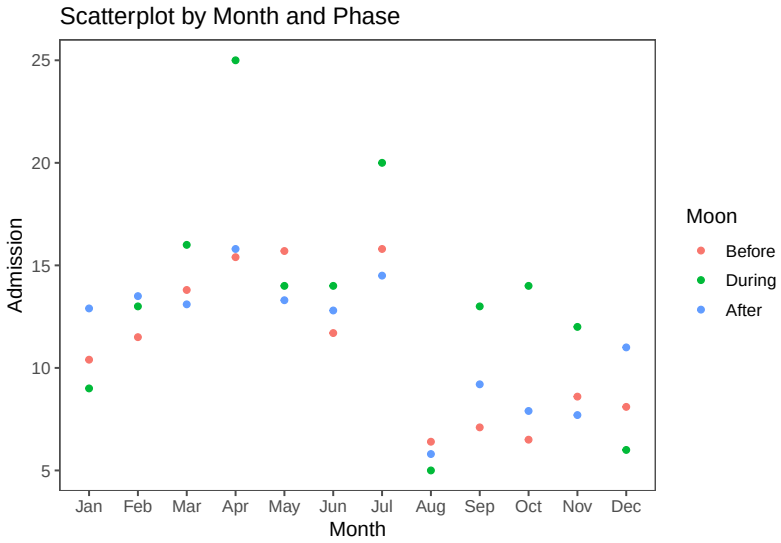
We can see that when grouped by month there are several averages which are not significantly different from one another, such as the average admissions of October, November, and December. However, there are stark differences in average admissions between certain months, such as April and August which corroborates the results from our ANOVA model showing that both moon phase and month have different average admissions across those groups.

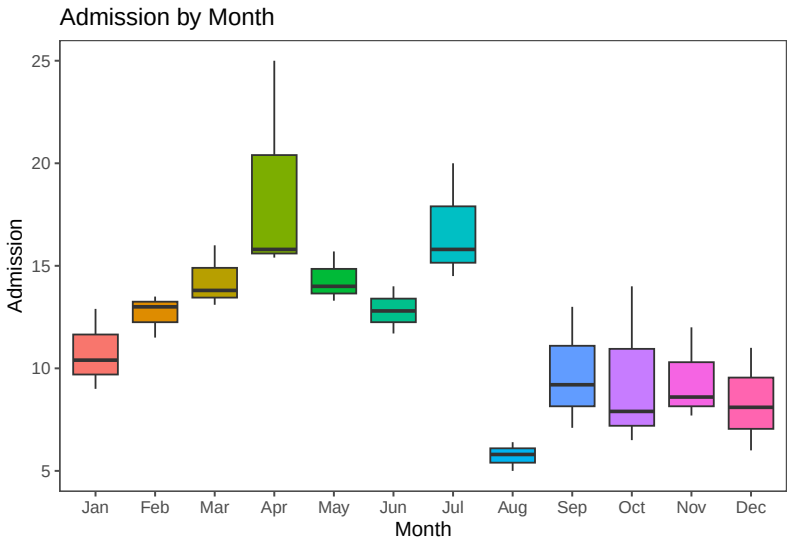
Conclusion and Discussion

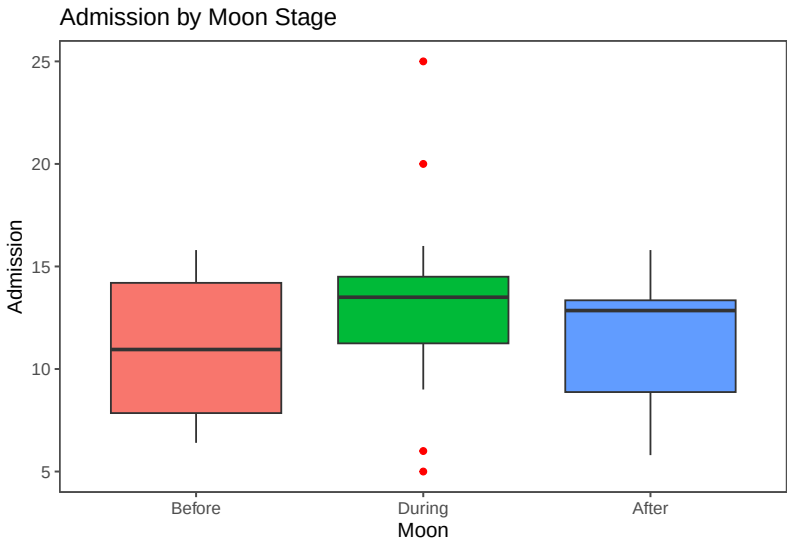
Overall, the analysis did find that there is a relationship between both phases of the moon and the month with regards to their impacts on average admissions.

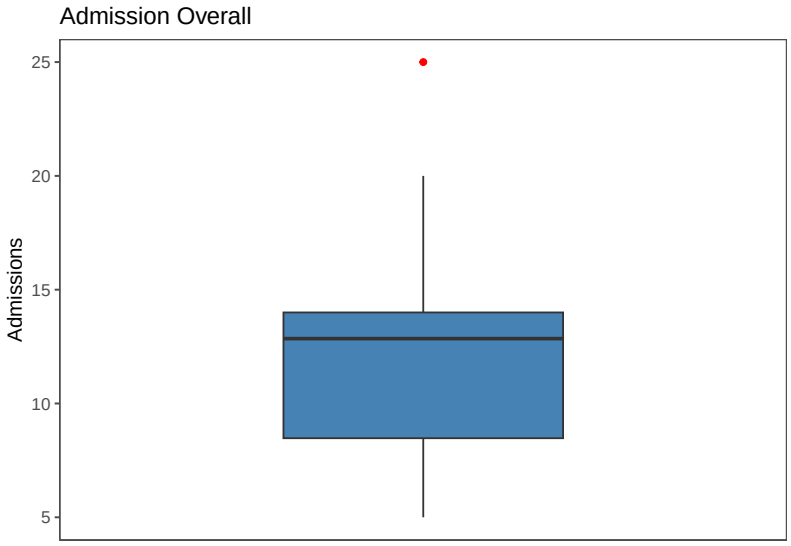
While there is a statistical significance of that the phases of the moon seem to have on admissions, it is important to be wary of this result and consider potential lurking variables which could be affecting this relationship. The study did not aim to gather data on what types of admissions into the emergency room, nor any length of stay. Additionally, the study only looked at a period of 1 year, with 3 measurements per month for the average number of admissions during the specified phase of the moon.

Appendix

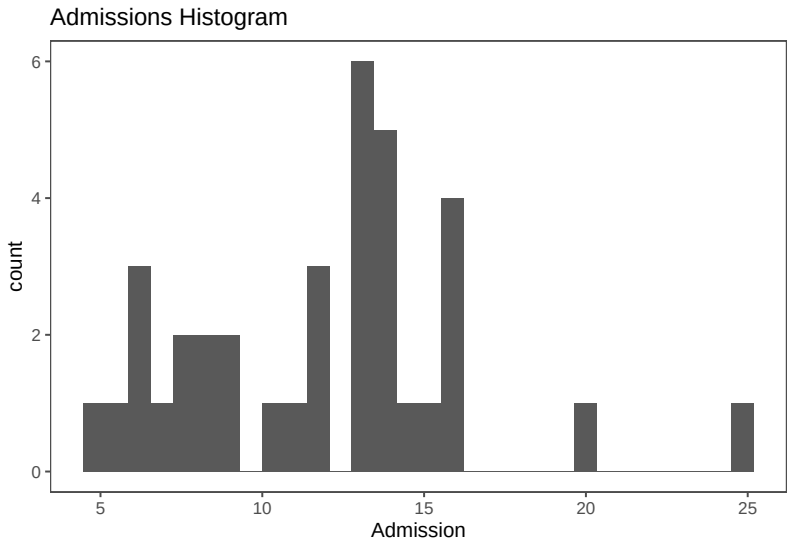

Scatterplot by Month and Phase
Admission
25
20
15
10
5
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month
Moon
Before
During
After


Admission by Month
Admission
25
20
15
10
5
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month






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




```
library(tidyverse)
library(ggthemes)
library(car)
library(agricolae)
library(ggpubr)
options(readr.show_col_types = FALSE)
data = read_csv("/Users/jz/Local Desktop/Depaul/winter 2023/MAT 442/projects/MentalHealth.csv", col_nam
df = data

df3 = df
df3$Moon = as_factor(df$Moon)
df3$Month = factor(df$Month)

df$Moon = as_factor(df$Moon)
df$Month = factor(df$Month,  levels=month.abb)
df2 = df %>% dplyr::group_by(Month, Moon) %>%
 tidyr::spread(Moon, Admission) %>%
 as.data.frame()



#summary stats for overall admission rates

df2 %>% ggtexttable(rows = NULL) %>%
 tab_add_title(text = "Data", face = "bold", size = 10)
df %>% group_by(Moon) %>% summarise(Mean = round(mean(Admission), 2), StandDev = round(sd(Admission),2)
                                    Median = median(Admission), Min = min(Admission),
                                    Max = max(Admission))

df %>% group_by(Month) %>% summarise(Mean = round(mean(Admission), 2), StandDev = round(sd(Admission), 
                                     Median = median(Admission), Min = min(Admission),
                                     Max = max(Admission))

df %>% summarise(Mean = round(mean(Admission), 2), StandDev = round(sd(Admission), 2),
                                     Median = median(Admission), Min = min(Admission),
```

```
                                    Max = max(Admission)) #%>% ggtexttable(rows = NULL)
knitr::include_graphics("/Users/jz/Local Desktop/Depaul/winter 2023/MAT 442/projects/moon.png")

two.way = aov(Admission ~ Moon + Month, data = df)
summary(two.way)
two.way$coefficients
#check model assumptions via plots. QQplot looks approx normal. Residuals vs fitted (homogeneity of var
plot(two.way,1)
plot(two.way,2)
#levene test for equal variances. Both do not reject null, can assume equal variances
leveneTest(Admission ~ Month, data = df)
leveneTest(Admission ~ Moon, data = df)
HSD.test(two.way,"Moon",console=TRUE, group = T)
monthresults = HSD.test(two.way,"Month",console=TRUE, group = T)

#interaction = aov(Admission ~ Moon + Month + Moon:Month, data = df)

#summary(interaction)
#does admission depend on moon stage and the month? Two way ANOVA
two.way = aov(Admission ~ Moon + Month, data = df)
summary(two.way)
two.way$coefficients
#check model assumptions via plots. QQplot looks approx normal. Residuals vs fitted (homogeneity of var
plot(two.way,1)
plot(two.way,2)
#levene test for equal variances. Both do not reject null, can assume equal variances
leveneTest(Admission ~ Month, data = df)
leveneTest(Admission ~ Moon, data = df)

#both are significant although Month appears to be more significant than moon stage
HSD.test(two.way,"Month",console=TRUE, group = T)
HSD.test(two.way,"Moon",console=TRUE, group = T)

TukeyHSD(two.way, which = "Moon")
TukeyHSD(two.way, which = "Month")
p1 = ggplot(df) + geom_point(aes(Month, Admission, color = Moon)) + theme_few()  + ggtitle("Scatterplot

p2 = ggplot(df, aes(x=Month, y=Admission, fill=Month)) +
  geom_boxplot(outlier.color = "red", show.legend = F) +
  ggtitle('Admission by Month')  + theme_few()

p3 = ggplot(df, aes(x=Moon, y=Admission, fill=Moon)) +
  geom_boxplot(outlier.color = "red", show.legend = F) +
  ggtitle('Admission by Moon Stage')  +theme_few()

p4 = ggplot(df) + geom_boxplot(aes(y=Admission), outlier.color = "red", fill = "steelblue") +
  scale_x_discrete() + labs(y = "Admissions") + ggtitle("Admission Overall") + theme_few()


p5 = ggplot(df) + geom_histogram(aes(x = Admission)) + theme_few() + ggtitle("Admissions Histogram")

#ggplot(df, aes(x = Admission, fill = Moon)) +
 # geom_histogram() + theme_few()
```

```r
moonplot = ggarrange(p1, p2, p3, p4, p5, labels = c("A", "B", "C", "D", "E"), nrow = 4, ncol = 2)

ggsave("moon.png", width = 6, height = 6)
p1

p2

p3

p4

p5
```