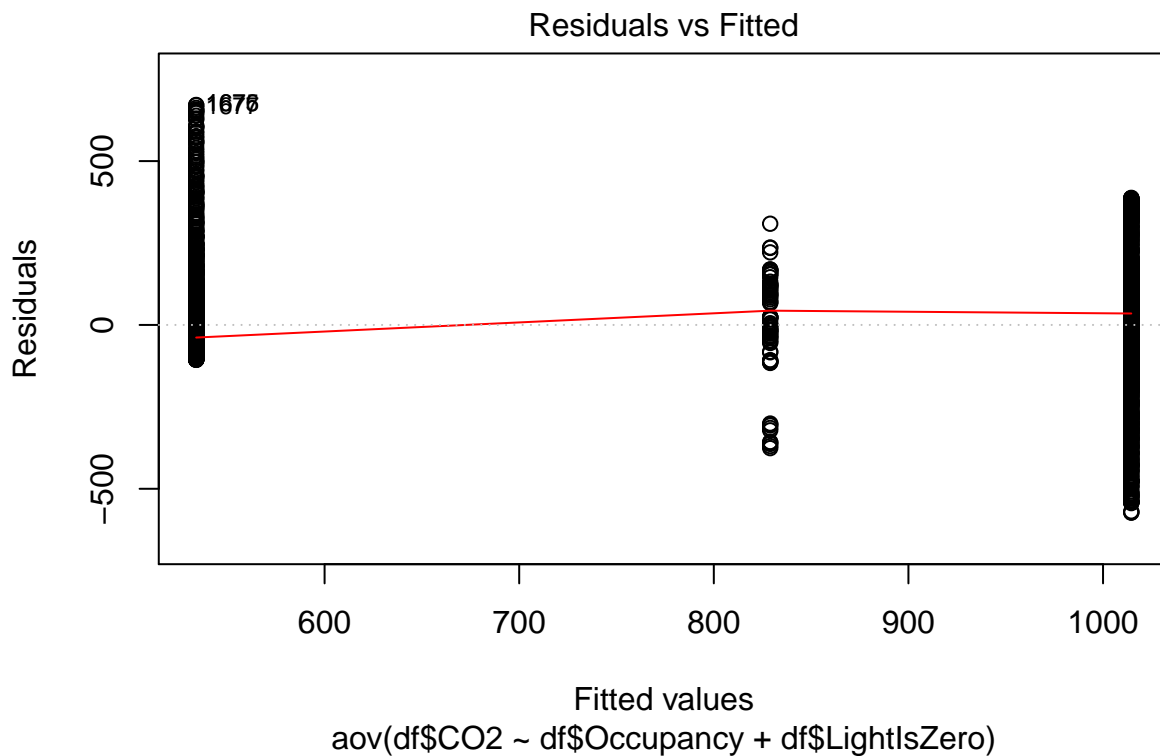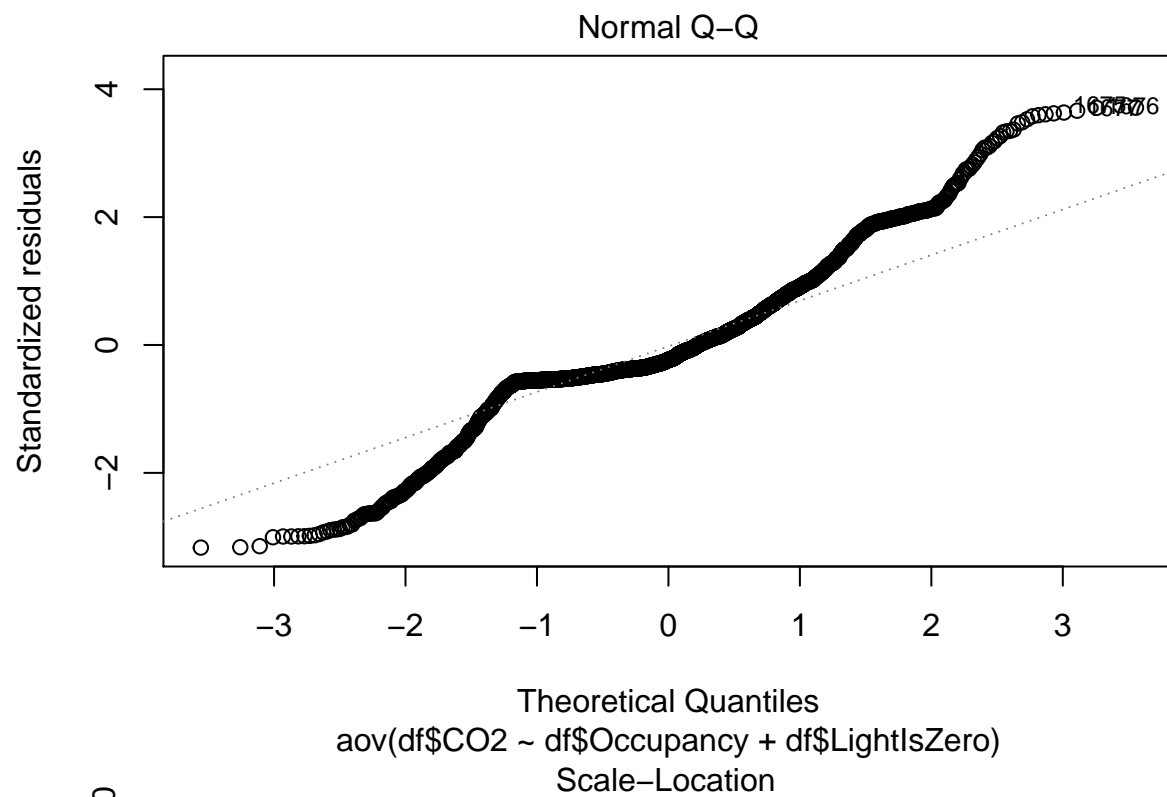# day_7_nov_7_2016_occupancy_tucky_anova

Going to try Anova with Tukey's method, which explains which groups are dissimilar.
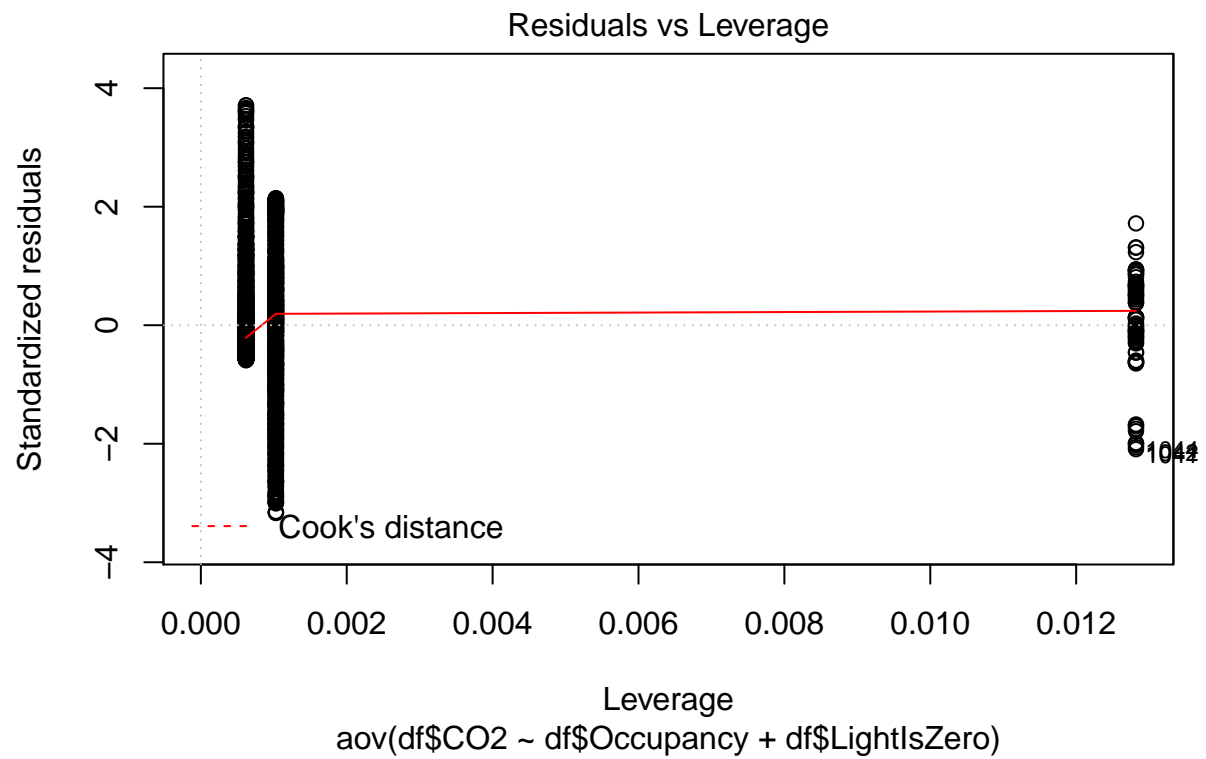
```
df$LightIsZero <- (df$Light == 0) * 1
df$Occupancy <- factor(df$Occupancy)
df$LightIsZero <- factor(df$LightIsZero)
myfit <- aov(df$CO2 ~ df$Occupancy + df$LightIsZero)
(tukey <- TukeyHSD(x=myfit, 'df$Occupancy', conf.level=0.95))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = df$CO2 ~ df$Occupancy + df$LightIsZero)
##
## $`df$Occupancy`
##         diff      lwr      upr p adj
## 1-0 466.9041 452.6286 481.1795     0
```

```
plot(myfit)
```


Residuals vs Fitted

aov(df$CO2 ~ df$Occupancy + df$LightIsZero)

Normal Q-Q

Standardized residuals

Theoretical Quantiles
aov(df$CO2 ~ df$Occupancy + df$LightIsZero)

Scale-Location

√|Standardized residuals|

Fitted values
aov(df$CO2 ~ df$Occupancy + df$LightIsZero)

2

Residuals vs Leverage

aov(df$CO2 ~ df$Occupancy + df$LightIsZero)

The anova doesn't work well haha, because there are only 3 different combo of factors. When Light is never 0, when occupied.

It's also cool to see Cook's D, that alongside the qq plot pretty much gives context on what the outliars are.