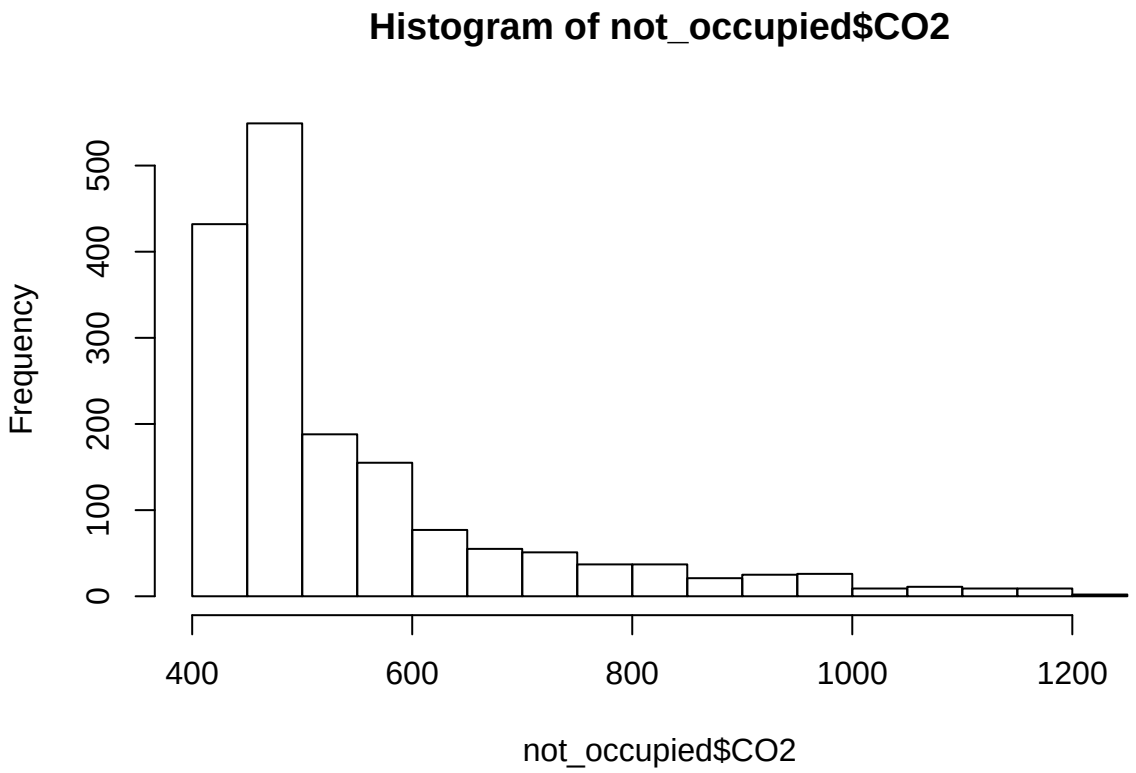
day_5_nov_3_2016_occupancy_boxcox

Going to try to do ANOVA tomorrow. Before then, I want to transform the variables to normal via box-cox and verify with QQplot.

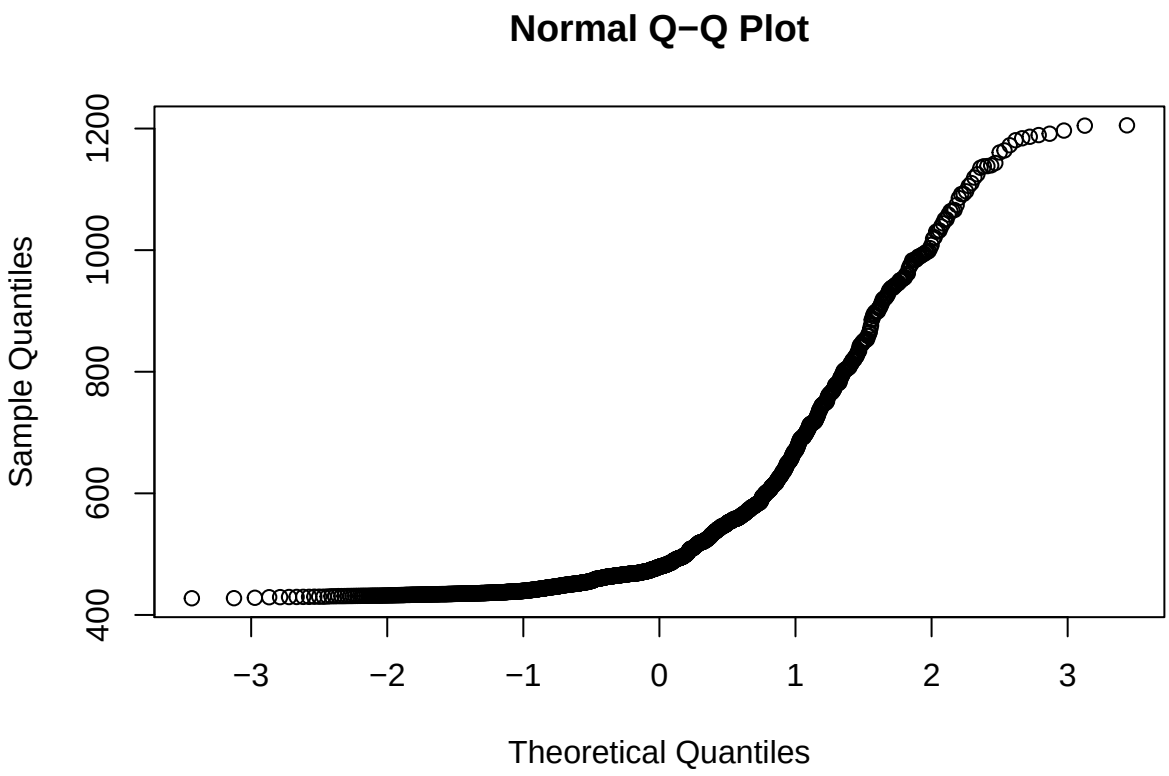
Here are some good slides on box-cox: https://www.ime.usp.br/~abe/lista/pdfm9cJKUmFZp.pdf

Let's visualize the CO2 for not occupied.

```
hist(not_occupied$CO2)
```




```
qqnorm(jitter(not_occupied$CO2))
```




```
summary(not_occupied$CO2)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   427.5   449.8   479.2   547.6   576.5  1205.0
```

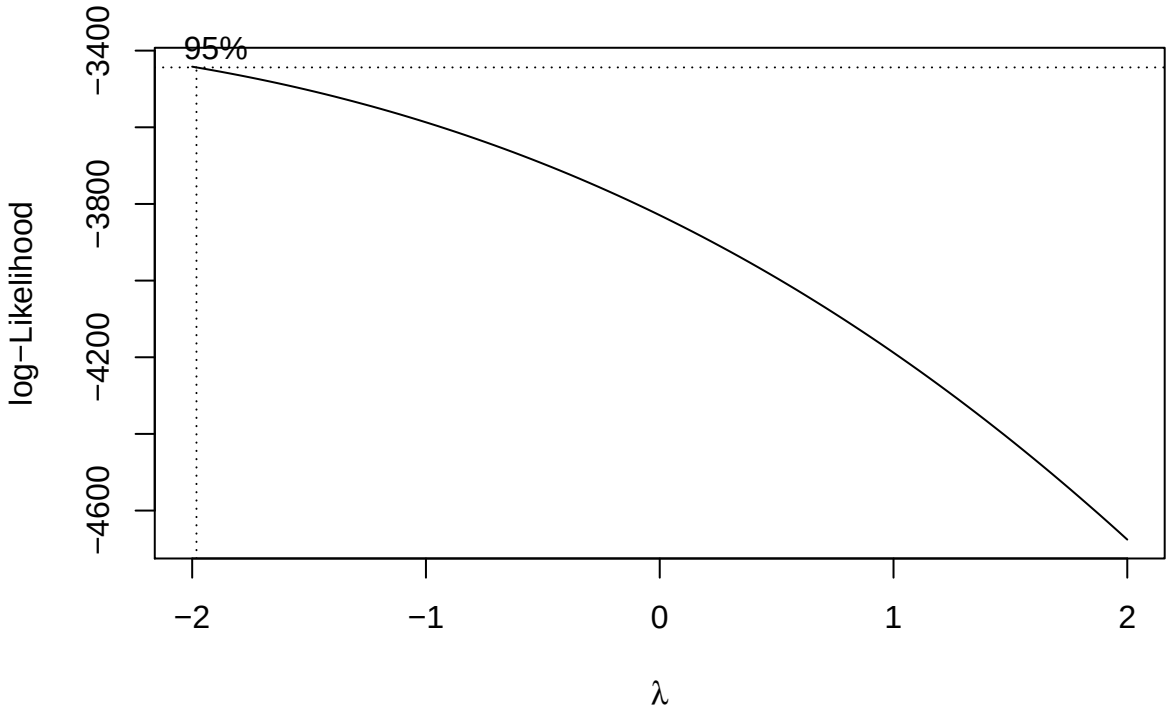
We see some really heavy positive skew. From the qqplot, we see that the line is not close to the 45 degrees line that we want to see, for a normal distribution.

Since the data is strictly positive, lambda_2 in box-cox should prob be negative or 0.

```
library(MASS)
(m <- lm(CO2 ~ Occupancy, data=not_occupied))
```

```
##
## Call:
## lm(formula = CO2 ~ Occupancy, data = not_occupied)
##
## Coefficients:
## (Intercept)    Occupancy
##       547.6           NA
```

```
bc <- boxcox(m)
```



Here Occupancy is 0 all the time, so the lm doesn't do anything...

Boxcox tries to get

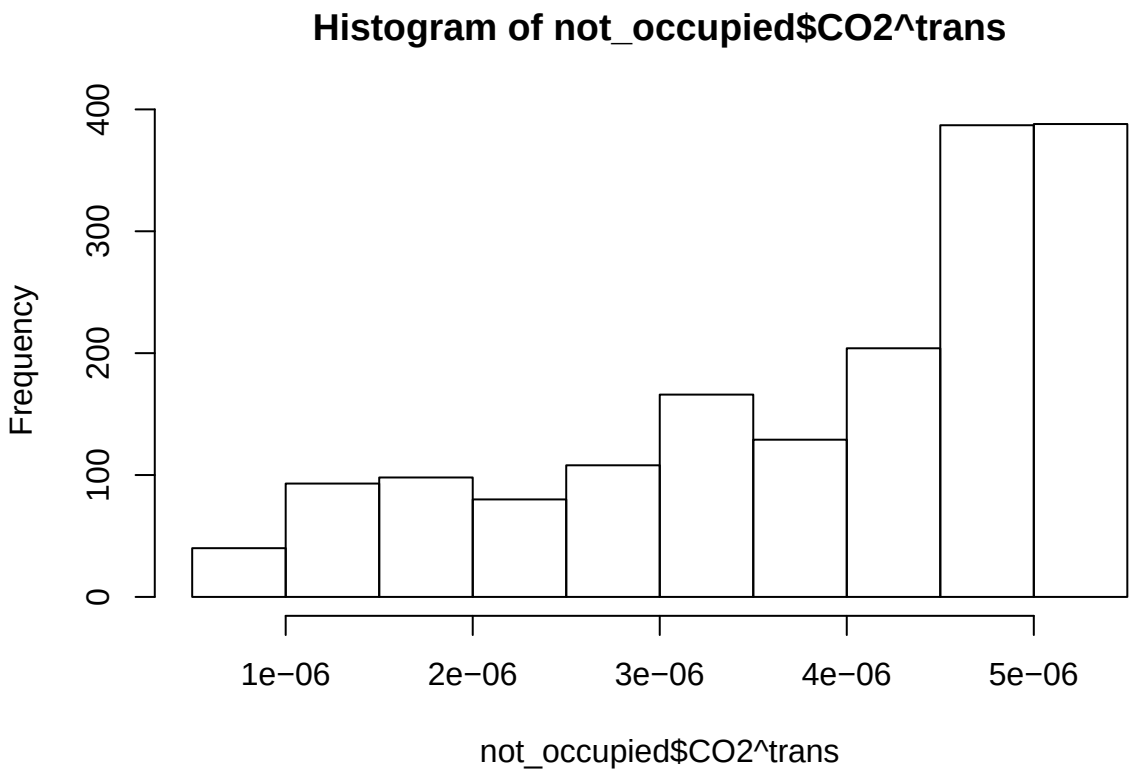
$$y - X\beta \; N(0, \sigma^2)$$

Since β is 0, it simply tries to transform the y.

```
(trans <- bc$x[which.max(bc$y)])
```

```
## [1] -2
```

```
hist(not_occupied$CO2^trans)
```




```
hist(not_occupied$CO2^-5)
```