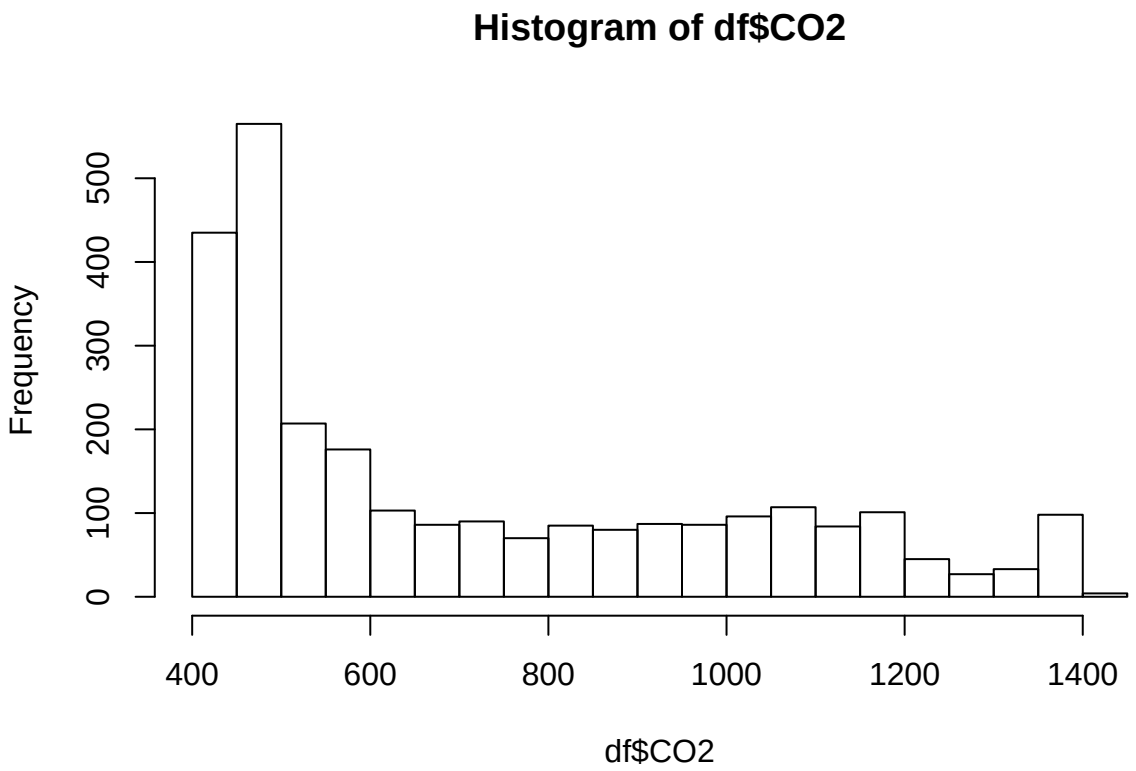
day_4_nov_2_2016_occupancy_anderson_darling

Going to look at the normality of some variables.

```
hist(df$CO2, breaks=20)
```



We see that the CO2 data has a positive skew. Eye balling it says that it's not normal. Let's see what Anderson-Darling says.

```
library(ADGofTest)
# sanity check
ad.test(rnorm(100, 100, 10), pnorm, 100, 10)
```

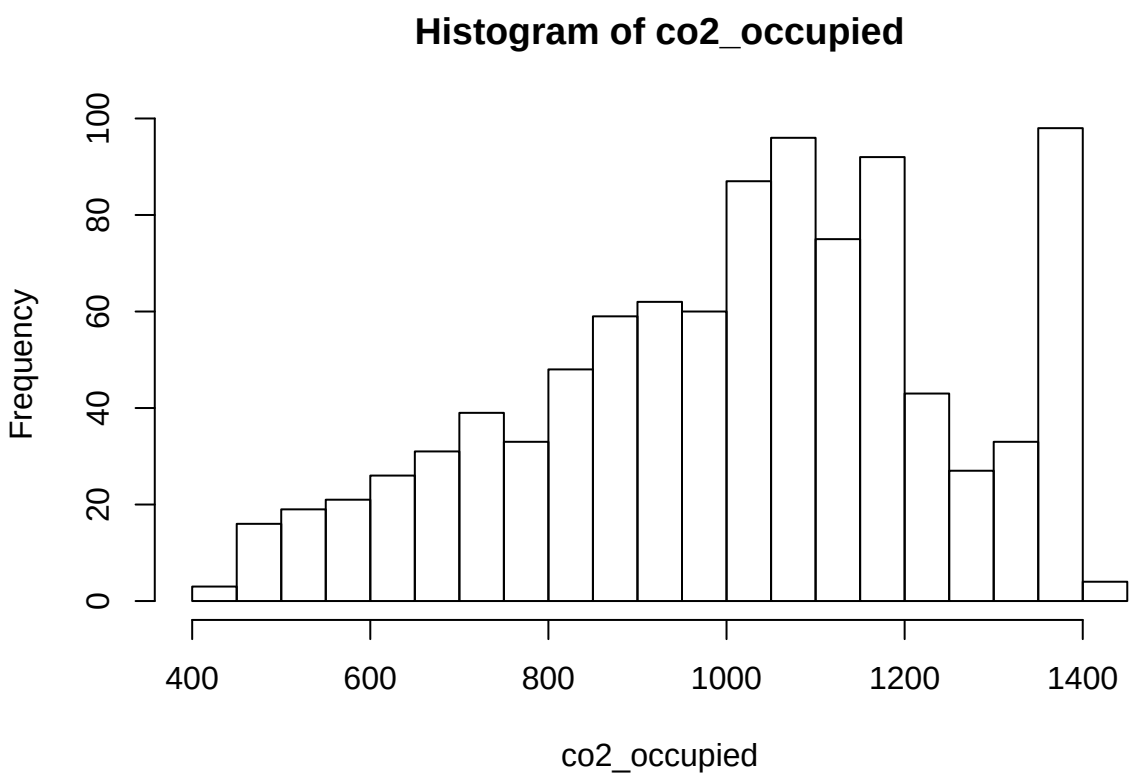
```
##
##  Anderson-Darling GoF Test
##
## data:  rnorm(100, 100, 10)  and  pnorm
## AD = 1.6279, p-value = 0.1488
## alternative hypothesis: NA
```

```
ad.test(df$CO2, pnorm, mean(df$CO2), sd(df$CO2))
```

```
##
##  Anderson-Darling GoF Test
##
## data:  df$CO2  and  pnorm
## AD = 150.35, p-value = 2.251e-07
## alternative hypothesis: NA
```

So our hypothesis was correct. The null that CO2 is normally distributed is false.

```
co2_occupied <- df[df$Occupancy == 1,]$CO2
hist(co2_occupied, breaks=25)
```




```
ad.test(co2_occupied, pnorm, mean(co2_occupied), sd(co2_occupied))
```

```
##
##  Anderson-Darling GoF Test
##
## data:  co2_occupied  and  pnorm
## AD = 5.4158, p-value = 0.001819
## alternative hypothesis: NA
```

We see that if we take just the occupied co2 values, the data is more normally distributed.