

# alli\_data

*Ciera Martinez*

*July 11, 2017*

**Purpose:** The purpose is to to correlate compounds to oviposition index. Mike wants to use multivariable regression or decision/regression to predict behavior. He also said to use linear or nonparametric.

```
## Library Prereqs
library(lme4)
```

```
## Warning: package 'lme4' was built under R version 3.3.2
```

```
## Loading required package: Matrix
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```
library(languageR)
library(xlsx)
```

```
## Loading required package: rJava
```

```
## Loading required package: xlsxjars
```

```
library(reshape2)
```

```
## Functions
```

```
## modelcheck <- function(model, h = 8, w = 10.5) { ## because plot(lmer.obj) doesn't work rs <- residu
```

## Read in Data

```
aq_DF <- read.csv("../data/output/combinedData.csv", row.names = 1)

dim(aq_DF)
```

```
## [1] 12 40
```

```
summary(aq_DF)
```

```
##      yeast      Fly.Line  X.eggs.on.yeast.side X.eggs.on.CTL.side
## Yeast.1:6  RW1005:12   Min.   : 66.0           Min.   : 3.00
## Yeast.2:6                1st Qu.:134.5         1st Qu.: 7.25
##                        Median :192.0           Median :10.00
```

```

##                               Mean    :173.7           Mean    :19.00
##                               3rd Qu.:212.0           3rd Qu.:14.00
##                               Max.    :231.0           Max.    :96.00
##      Total.eggs  Oviposition.index X1.hexanol..2.ethyl_1765
##  Min.    : 71.0  Min.    :0.4056   Min.    :0
## 1st Qu.:143.0  1st Qu.:0.8513   1st Qu.:0
## Median :205.0  Median :0.8754   Median :0
## Mean    :192.7  Mean    :0.8348   Mean    :0
## 3rd Qu.:239.2  3rd Qu.:0.9267   3rd Qu.:0
## Max.    :323.0  Max.    :0.9516   Max.    :0
## X1..3.dioxolane..2.4.5.trimethyl_347
##  Min.    :0.3499
## 1st Qu.:0.3499
## Median :0.3833
## Mean    :0.3833
## 3rd Qu.:0.4167
## Max.    :0.4167
## X2.methoxy.4.vinylphenol..vanilla..coffee..clove._3379
##  Min.    :0.0
## 1st Qu.:0.0
## Median :0.5
## Mean    :0.5
## 3rd Qu.:1.0
## Max.    :1.0
## X3.2H..Thiophenone..dihydro.2.methyl..blackberry._1514
##  Min.    :0.0
## 1st Qu.:0.0
## Median :0.5
## Mean    :0.5
## 3rd Qu.:1.0
## Max.    :1.0
## X7.octenoic.acid..ethyl.ester_2692 acetic.acid..very.tiny._199
##  Min.    :0.0                               Min.    :0.1867
## 1st Qu.:0.0                               1st Qu.:0.1867
## Median :0.5                               Median :0.2157
## Mean    :0.5                               Mean    :0.2157
## 3rd Qu.:1.0                               3rd Qu.:0.2448
## Max.    :1.0                               Max.    :0.2448
## acetic.acid..2.phenylethyl.ester._3100
##  Min.    :0
## 1st Qu.:0
## Median :0
## Mean    :0
## 3rd Qu.:0
## Max.    :0
## acetic.acid..2.phenylethyl.ester_3043
##  Min.    :0
## 1st Qu.:0
## Median :0
## Mean    :0
## 3rd Qu.:0
## Max.    :0
## acetic.acid..2.phenylethyl.ester_3071 butanoic.acid..ethyl.ester_589
##  Min.    :0.003214                          Min.    :0.0000

```

```

## 1st Qu.:0.003214          1st Qu.:0.0000
## Median :0.003804          Median :0.4886
## Mean   :0.003804          Mean   :0.4886
## 3rd Qu.:0.004395          3rd Qu.:0.9772
## Max.   :0.004395          Max.   :0.9772
## decanoic.acid..ethyl.ester_3810 ethane..1...diethoxy..very.tiny._357
## Min.   :0                  Min.   :0.3499
## 1st Qu.:0                  1st Qu.:0.3499
## Median :0                  Median :0.3833
## Mean   :0                  Mean   :0.3833
## 3rd Qu.:0                  3rd Qu.:0.4167
## Max.   :0                  Max.   :0.4167
## ethanol_36 ethanol._75 ethyl.acetate_135
## Min.   :0.8914 Min.   :0.6249 Min.   :0.1165
## 1st Qu.:0.8914 1st Qu.:0.6249 1st Qu.:0.1165
## Median :0.9457 Median :0.7307 Median :0.5582
## Mean   :0.9457 Mean   :0.7307 Mean   :0.5582
## 3rd Qu.:1.0000 3rd Qu.:0.8365 3rd Qu.:1.0000
## Max.   :1.0000 Max.   :0.8365 Max.   :1.0000
## Hex.5.enoic.acid..ethyl.ester_1526 Hexanoic.acid..ethyl.ester_1595
## Min.   :0.0              Min.   :0.0
## 1st Qu.:0.0              1st Qu.:0.0
## Median :0.5              Median :0.5
## Mean   :0.5              Mean   :0.5
## 3rd Qu.:1.0              3rd Qu.:1.0
## Max.   :1.0              Max.   :1.0
## isoamyl.acetate..very.tiny._220 isoamyl.acetate..very.tiny._422
## Min.   :0.09595          Min.   :0
## 1st Qu.:0.09595          1st Qu.:0
## Median :0.11091          Median :0
## Mean   :0.11091          Mean   :0
## 3rd Qu.:0.12586          3rd Qu.:0
## Max.   :0.12586          Max.   :0
## isoamyl.acetate_924 isoamyl.acetate_937 isoamyl.acetate_945
## Min.   :0.4239          Min.   :0.002092 Min.   :0
## 1st Qu.:0.4239          1st Qu.:0.002092 1st Qu.:0
## Median :0.5785          Median :0.002956 Median :0
## Mean   :0.5785          Mean   :0.002956 Mean   :0
## 3rd Qu.:0.7331          3rd Qu.:0.003820 3rd Qu.:0
## Max.   :0.7331          Max.   :0.003820 Max.   :0
## isoamyl.alcohol_373 isobutanol_159 isobutyl.acetate_490
## Min.   :0.3936          Min.   :0.2645 Min.   :0
## 1st Qu.:0.3936          1st Qu.:0.2645 1st Qu.:0
## Median :0.5815          Median :0.2999 Median :0
## Mean   :0.5815          Mean   :0.2999 Mean   :0
## 3rd Qu.:0.7695          3rd Qu.:0.3352 3rd Qu.:0
## Max.   :0.7695          Max.   :0.3352 Max.   :0
## methyl.anthranilate..concord.grapes._3522 Nonanal_2202
## Min.   :0.0              Min.   :0.8803
## 1st Qu.:0.0              1st Qu.:0.8803
## Median :0.5              Median :0.8987
## Mean   :0.5              Mean   :0.8987
## 3rd Qu.:1.0              3rd Qu.:0.9170
## Max.   :1.0              Max.   :0.9170

```

```
## octanoic.acid.ethyl.ester_2741 oxime..methoxyphenyl..small._1092
## Min. :0.3183 Min. :0.0000
## 1st Qu.:0.3183 1st Qu.:0.0000
## Median :0.6592 Median :0.3351
## Mean :0.6592 Mean :0.3351
## 3rd Qu.:1.0000 3rd Qu.:0.6701
## Max. :1.0000 Max. :0.6701
## phenylethyl.alcohol_2157 phenylethyl.alcohol_2249 phenylethyl.alcohol_2291
## Min. :0 Min. :0.3383 Min. :0.0000
## 1st Qu.:0 1st Qu.:0.3383 1st Qu.:0.0000
## Median :0 Median :0.4627 Median :0.3268
## Mean :0 Mean :0.4627 Mean :0.3268
## 3rd Qu.:0 3rd Qu.:0.5872 3rd Qu.:0.6536
## Max. :0 Max. :0.5872 Max. :0.6536
## propanoic.acid..ethyl.ester_308 triacetin_3589
## Min. :0.1400 Min. :0
## 1st Qu.:0.1400 1st Qu.:0
## Median :0.2606 Median :0
## Mean :0.2606 Mean :0
## 3rd Qu.:0.3812 3rd Qu.:0
## Max. :0.3812 Max. :0
```

```
str(aq_DF)
```

```
## 'data.frame': 12 obs. of 40 variables:
## $ yeast : Factor w/ 2 levels "Yeast.1","Yeast.2": 1 1 1 1 1 1 1 1 1 1 1 1
## $ Fly.Line : Factor w/ 1 level "RW1005": 1 1 1 1 1 1 1 1 1 1 1 1
## $ X.eggs.on.yeast.side : int 184 204 231 207 227 200 150 230 137 ...
## $ X.eggs.on.CTL.side : int 14 8 8 44 96 12 14 10 10 3 ...
## $ Total.eggs : int 198 212 239 251 323 212 164 240 147 ...
## $ Oviposition.index : num 0.859 0.925 0.933 0.649 0.406 ...
## $ X1.hexanol..2.ethyl_1765 : int 0 0 0 0 0 0 0 0 0 0 ...
## $ X1..3.dioxolane..2.4.5.trimethyl_347 : num 0.35 0.35 0.35 0.35 0.35 ...
## $ X2.methoxy.4.vinylphenol..vanilla..coffee..clove._3379: int 1 1 1 1 1 1 0 0 0 0 ...
## $ X3.2H..Thiophenone..dihydro.2.methyl..blackberry._1514: int 1 1 1 1 1 1 0 0 0 0 ...
## $ X7.octenoic.acid..ethyl.ester_2692 : int 1 1 1 1 1 1 0 0 0 0 ...
## $ acetic.acid..very.tiny._199 : num 0.187 0.187 0.187 0.187 0.187 ...
## $ acetic.acid..2.phenylethyl.ester._3100 : int 0 0 0 0 0 0 0 0 0 0 ...
## $ acetic.acid..2.phenylethyl.ester_3043 : int 0 0 0 0 0 0 0 0 0 0 ...
## $ acetic.acid..2.phenylethyl.ester_3071 : num 0.00321 0.00321 0.00321 0.00321 0.00321 ...
## $ butanoic.acid..ethyl.ester_589 : num 0.977 0.977 0.977 0.977 0.977 ...
## $ decanoic.acid..ethyl.ester_3810 : int 0 0 0 0 0 0 0 0 0 0 ...
## $ ethane..1...diethoxy..very.tiny._357 : num 0.35 0.35 0.35 0.35 0.35 ...
## $ ethanol_36 : num 1 1 1 1 1 ...
## $ ethanol._75 : num 0.836 0.836 0.836 0.836 0.836 ...
## $ ethyl.acetate_135 : num 0.116 0.116 0.116 0.116 0.116 ...
## $ Hex.5.enoic.acid..ethyl.ester_1526 : int 1 1 1 1 1 1 0 0 0 0 ...
## $ Hexanoic.acid..ethyl.ester_1595 : int 1 1 1 1 1 1 0 0 0 0 ...
## $ isoamyl.acetate..very.tiny._220 : num 0.096 0.096 0.096 0.096 0.096 ...
## $ isoamyl.acetate..very.tiny._422 : int 0 0 0 0 0 0 0 0 0 0 ...
## $ isoamyl.acetate_924 : num 0.733 0.733 0.733 0.733 0.733 ...
## $ isoamyl.acetate_937 : num 0.00382 0.00382 0.00382 0.00382 0.00382 ...
## $ isoamyl.acetate_945 : int 0 0 0 0 0 0 0 0 0 0 ...
## $ isoamyl.alcohol_373 : num 0.769 0.769 0.769 0.769 0.769 ...
```

```
## $ isobutanol_159 : num 0.335 0.335 0.335 0.335 0.335 ...
## $ isobutyl.acetate_490 : int 0 0 0 0 0 0 0 0 0 0 ...
## $ methyl.anthranilate..concord.grapes._3522 : int 1 1 1 1 1 1 0 0 0 0 ...
## $ Nonanal_2202 : num 0.88 0.88 0.88 0.88 0.88 ...
## $ octanoic.acid.ethyl.ester_2741 : num 1 1 1 1 1 ...
## $ oxime..methoxyphenyl..small._1092 : num 0 0 0 0 0 ...
## $ phenyethyl.alcohol_2157 : int 0 0 0 0 0 0 0 0 0 0 ...
## $ phenyethyl.alcohol_2249 : num 0.587 0.587 0.587 0.587 0.587 ...
## $ phenyethyl.alcohol_2291 : num 0 0 0 0 0 ...
## $ propanoic.acid..ethyl.ester_308 : num 0.14 0.14 0.14 0.14 0.14 ...
## $ triacetin_3589 : int 0 0 0 0 0 0 0 0 0 0 ...
```

```
# Im'm going to force all into num
colnames(aq_DF)
```

```
## [1] "yeast"
## [2] "Fly.Line"
## [3] "X.eggs.on.yeast.side"
## [4] "X.eggs.on.CTL.side"
## [5] "Total.eggs"
## [6] "Oviposition.index"
## [7] "X1.hexanol..2.ethyl_1765"
## [8] "X1..3.dioxolane..2.4.5.trimethyl_347"
## [9] "X2.methoxy.4.vinylphenol..vanilla..coffee..clove._3379"
## [10] "X3.2H..Thiophenone..dihydro.2.methyl..blackberry._1514"
## [11] "X7.octenoic.acid..ethyl.ester_2692"
## [12] "acetic.acid..very.tiny._199"
## [13] "acetic.acid..2.phenylethyl.ester._3100"
## [14] "acetic.acid..2.phenylethyl.ester_3043"
## [15] "acetic.acid..2.phenylethyl.ester_3071"
## [16] "butanoic.acid..ethyl.ester_589"
## [17] "decanoic.acid..ethyl.ester_3810"
## [18] "ethane..1...diethoxy..very.tiny._357"
## [19] "ethanol_36"
## [20] "ethanol._75"
## [21] "ethyl.acetate_135"
## [22] "Hex.5.enoic.acid..ethyl.ester_1526"
## [23] "Hexanoic.acid..ethyl.ester_1595"
## [24] "isoamyl.acetate..very.tiny._220"
## [25] "isoamyl.acetate..very.tiny._422"
## [26] "isoamyl.acetate_924"
## [27] "isoamyl.acetate_937"
## [28] "isoamyl.acetate_945"
## [29] "isoamyl.alcohol_373"
## [30] "isobutanol_159"
## [31] "isobutyl.acetate_490"
## [32] "methyl.anthranilate..concord.grapes._3522"
## [33] "Nonanal_2202"
## [34] "octanoic.acid.ethyl.ester_2741"
## [35] "oxime..methoxyphenyl..small._1092"
## [36] "phenyethyl.alcohol_2157"
## [37] "phenyethyl.alcohol_2249"
## [38] "phenyethyl.alcohol_2291"
## [39] "propanoic.acid..ethyl.ester_308"
```

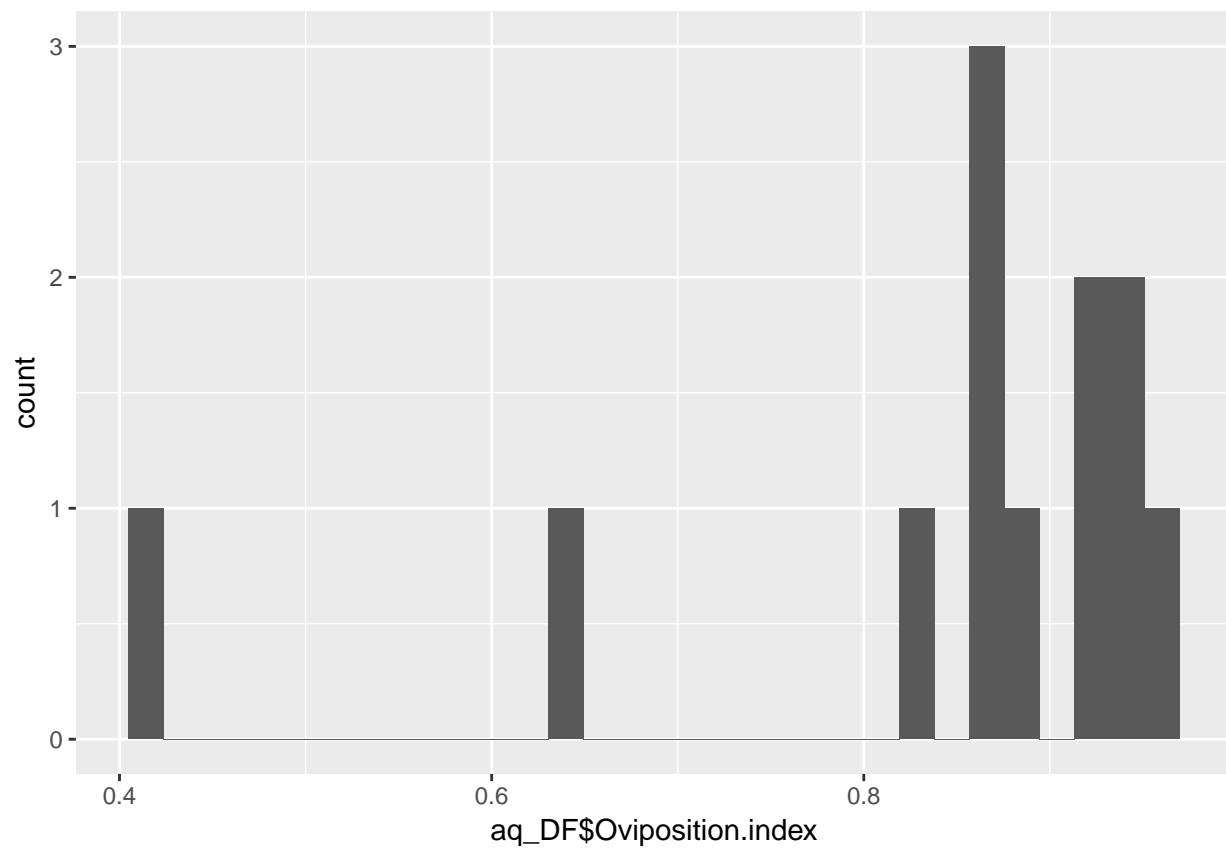
```
## [40] "triacetin_3589"
```

```
aq_DF[,7:40] <- apply(aq_DF[,7:40], 2, function(x) as.numeric(as.character(x)))
```

## Visualize Distributions

```
ggplot(aq_DF, aes(aq_DF$Oviposition.index)) +  
  geom_histogram()
```

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



```
## I think that we need more measurements.
```