Analyzing *C. elegans* single animal tracking data for Dennis 2017

Emily Jane Dennis

9/22/2017

Purpose:

To plot all single animal tracking data for the manuscript.

Background:

This is an R Markdown document. http://rmarkdown.rstudio.com.

The code required to get these graphs is in the gray boxes, along with information on how to run the statistical tests reported in the paper and any plots that were used to determine if these statistical tests were generally appropriate.

This was run in RStudio, using R version 3.3.1 (2016-10-31), Sincere Pumpkin Patch on a x86_64-apple-darwin13.4.0

Before these data can be analyzed, I used the following block to:

1. Load all libraries required for analysis and plotting (ensure they're all installed ahead of time)
2. Load these data
3. Subset data to use later to plot

I found this resource especially helpful http://www.bmj.com/content/312/7038/1079

```
library(plyr)
library(dplyr)
library(ggplot2)
library(ggthemes)
library(reshape2)
library(MASS)
library(car)
library(multcompView)
```

Pre-processing of raw MATLAB output files

Each processed file went through these steps: 1. 60 minute videos were taken at 3fps using LED illumination to evenly illuminate the plate. less than 15 animals were added to each plate for each experiment.

2. The videos were analyzed in MATLAB version 2014b using previously published custom scripts (Cho *et al* eLife 2016). Emily Dennis manually linked all trajectories until tracks were complete. Incomplete trajectories were discarded, as were any animals who did not move more than 3 body lengths away from the start position. If an animal exited the field of view, the track was truncated but considered 'complete'. Also, a paralytic (NaN3) was added to each odor source, and trajectories of animals that reached the odor source were therefore truncated.
3. Each pause in each trajectory of each animal was extracted, along with the last frame of the pause as well as the length of the pause. Pauses less than one second in length (3 frames) were discarded.

4. The data were then unblinded, organized, and exported as a .csv file. For each pause, the genotype and plate treatment were collapsed into one variable (Ne = N2 animals on ethanol plate, Sd = str-217 mutants on a plate containing 0.15% DEET). The headings/column names for data are as follows:

Column headings in spread sheet

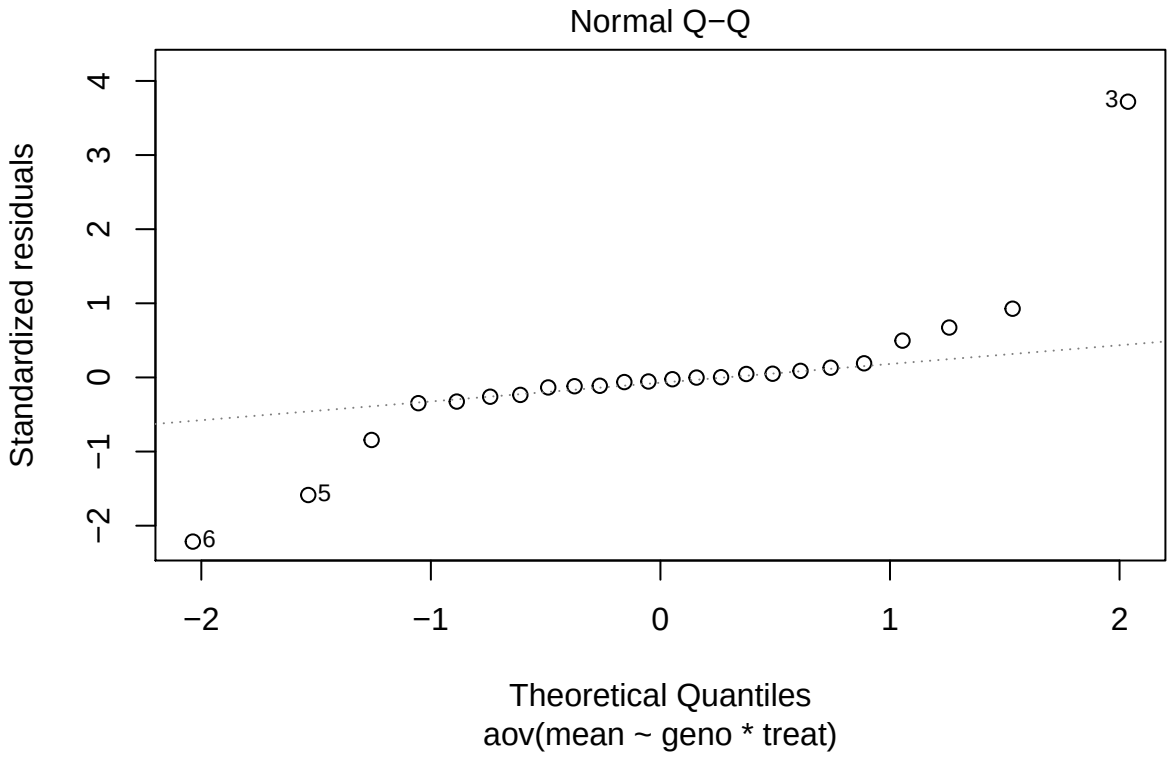
genotype *includes plate information and genotype of animal* datelist *date acquired, camera used (of 4 recorded simultaneously), and video number (if multiple rounds of experiments were conducted on the same day) e.g. 20160825c1v1 was the first video taken with computer/camera 1 on August 25th, 2016* animal *computer-generated ID for each animal (1-N)* fram *frame of the last frame of the pause* lens *length (in frames) of the pause*

First, I processed all of these data to make summary data for statistics.

Fig 4d

When chemotaxing to isoamyl alochol, wild-type animals exhibit many very long pauses that raise the average pause length

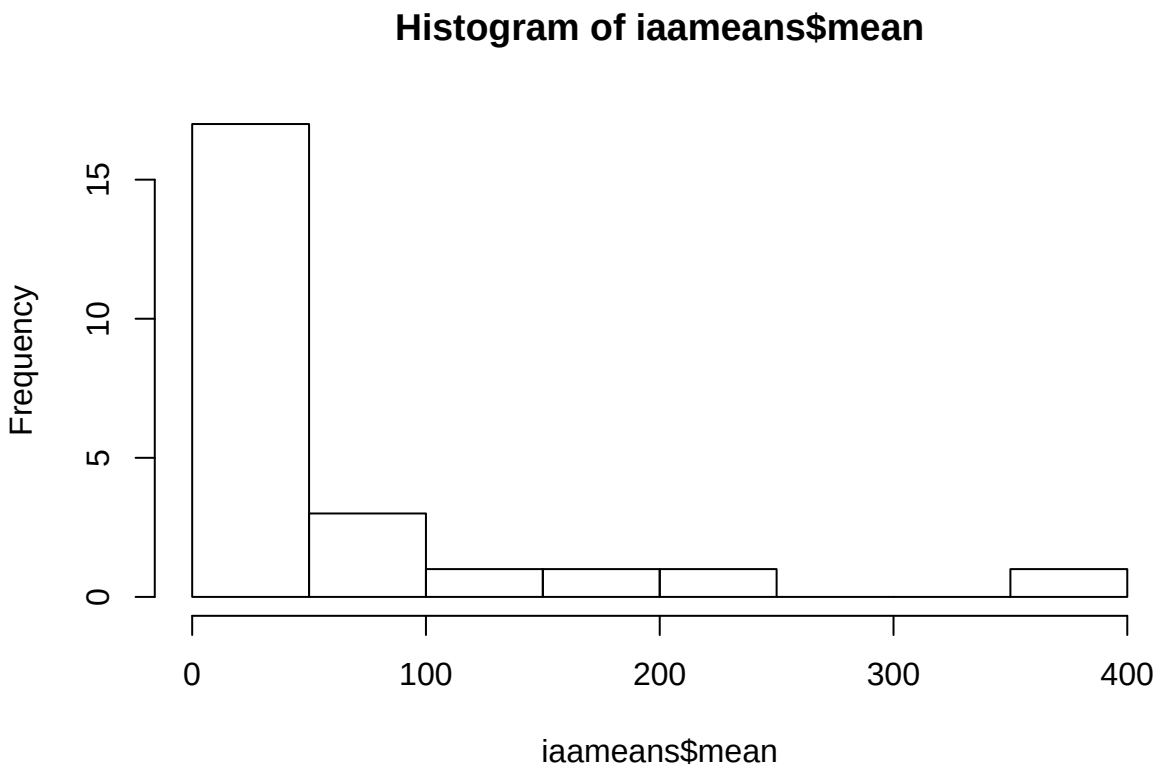
```
#load data and trim any longer tracks (on a few days, the computer failed to end recording movie at end
iaa <- read.csv('tracking_iaa.csv')
iaa <- subset(iaa,fram<10800)

#take up the average pause length for each animal on each plate
iaaanimal <- ddply(iaa,.(genotype,datelist,animal),summarize,lengths=mean(lens,na.rm=TRUE))

#makes means of means, each plate result is an average of all individual animal mean pause lengths on t
iaameans <- ddply(iaaanimal,.(genotype,datelist),summarize,mean=mean(lengths,na.rm=TRUE))

iaameanssummary <- ddply(iaameans,.(genotype),summarize,means=mean(mean),sd=sd(mean),N=length(mean),se=s


iaameans$geno <- substr(iaameans$genotype,1,1)
iaameans$treat <- substr(iaameans$genotype,2,2)

iaastat.aov = aov(mean~geno*treat,data=iaameans)
plot(iaastat.aov,which=2)
```




```
#This looks like a really bad fit.
```

The Q-Q plot for these data looks terrible (non-normal residuals) so I looked at the histogram and it's obviously very skewed

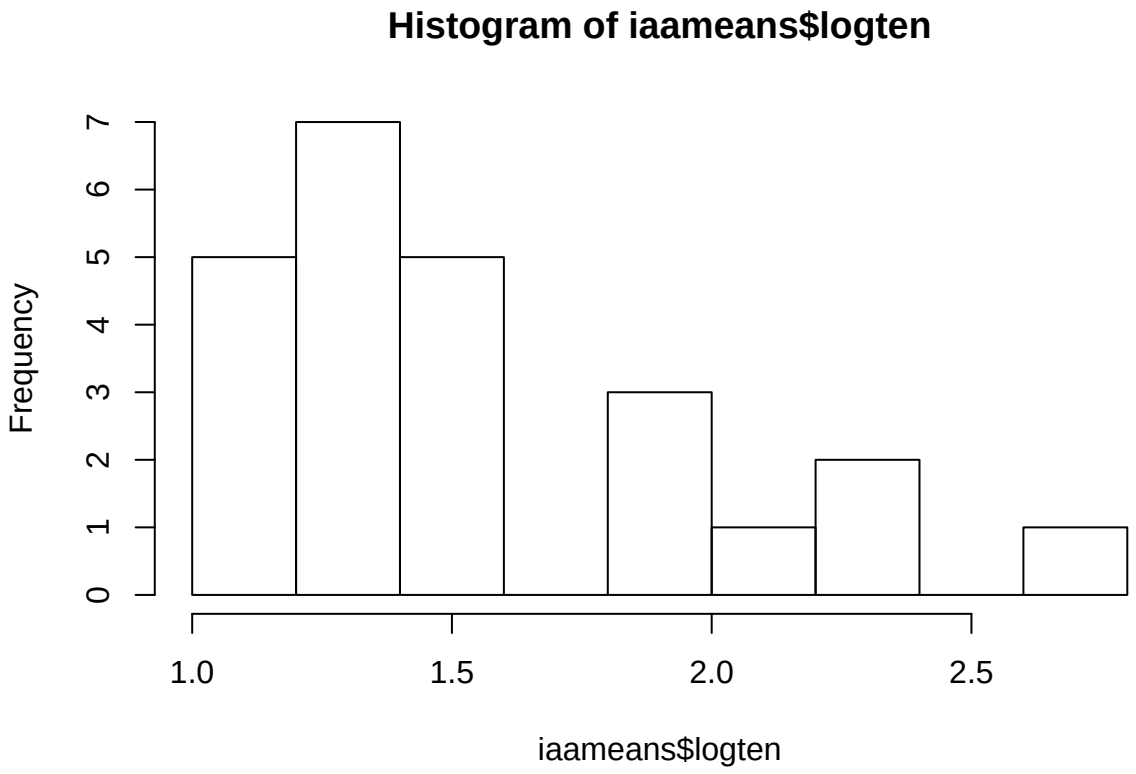
```
hist(iaameans$mean,breaks=10)
```



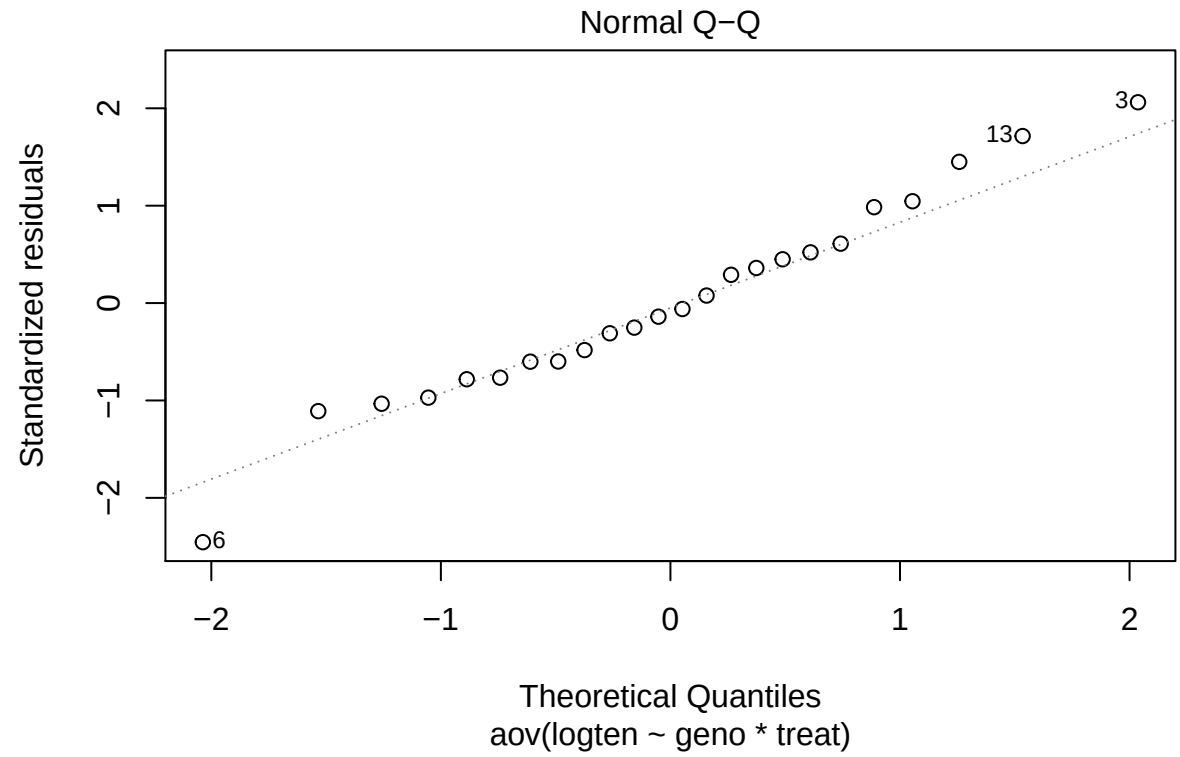

```
#It is obvious to me that the actual effect (change in mean) also changes the variance (adding long pau
```

I then did a log10 transform and checked the QQ and histograms which look improved

```
iaameans$logten <- log10(iaameans$mean)

hist(iaameans$logten)
```




```
iaastatbox.aov = aov(logten~geno*treat,data=iaameans)
plot(iaastatbox.aov,which=2)
```



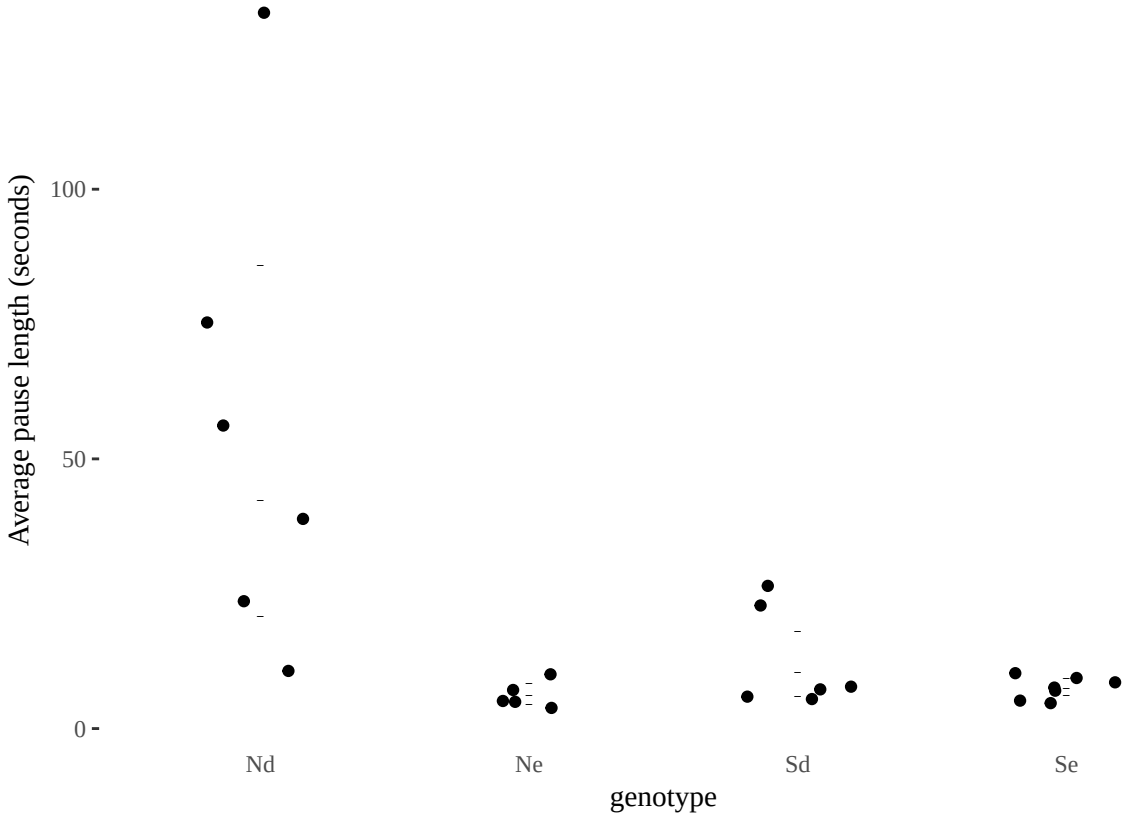

```
#This looks more reasonable!
```

I then ran an ANOVA on these transformed data and backtransformed to get the mean and confidence intervals, which are plotted along with these data

```
iaatukey <- TukeyHSD(aov(formula=logten~geno*treat,data=iaameans))

iaatukey.levels <- iaatukey$`geno:treat`[,4]

 multcompLetters(iaatukey.levels)['Letters']
```

```
## $Letters
## S:d N:e S:e N:d
## "a" "a" "a" "b"
```

```
iaameanssummaryadj <- ddply(iaameans,.(genotype),summarize,means=mean(logten),sd=sd(logten),N=length(log

iaameanssummaryadjback <- iaameanssummaryadj
iaameanssummaryadjback$means <- 10^(iaameanssummaryadj$means)
iaameanssummaryadjback$cim <- 10^(iaameanssummaryadj$cim)
iaameanssummaryadjback$cip <- 10^(iaameanssummaryadj$cip)

ggplot(iaameans,aes(x=genotype,(y=mean/3))) +
  geom_jitter(height=0,width=0.2) +
  theme_tufte() +
  ylab("Average pause length (seconds)") +
  theme(axis.ticks.x=element_blank()) +
  geom_point(data=iaameanssummaryadjback,aes(x=genotype,y=(means/3)),shape='-') +
  geom_point(data=iaameanssummaryadjback,aes(x=genotype,y=(cip/3)),shape='-') +
  geom_point(data=iaameanssummaryadjback,aes(x=genotype,y=(cim/3)),shape='-')
```



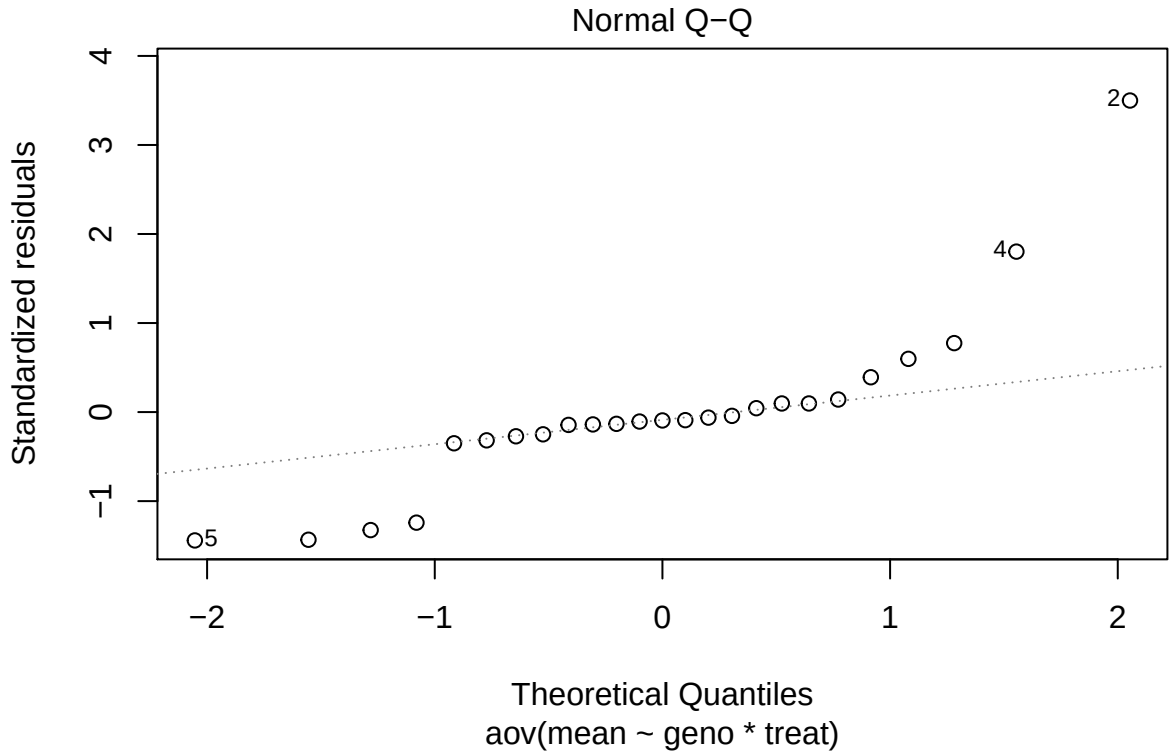
I used the same transformation on all data collected in a similar fashion, checking that the Q-Q plots were improved and the histograms look less skewed.

Fig 4e

On DEET-agar, wild-type animals show an increase in pausing behavior, and *str-217* animals are resistant to this effect of DEET

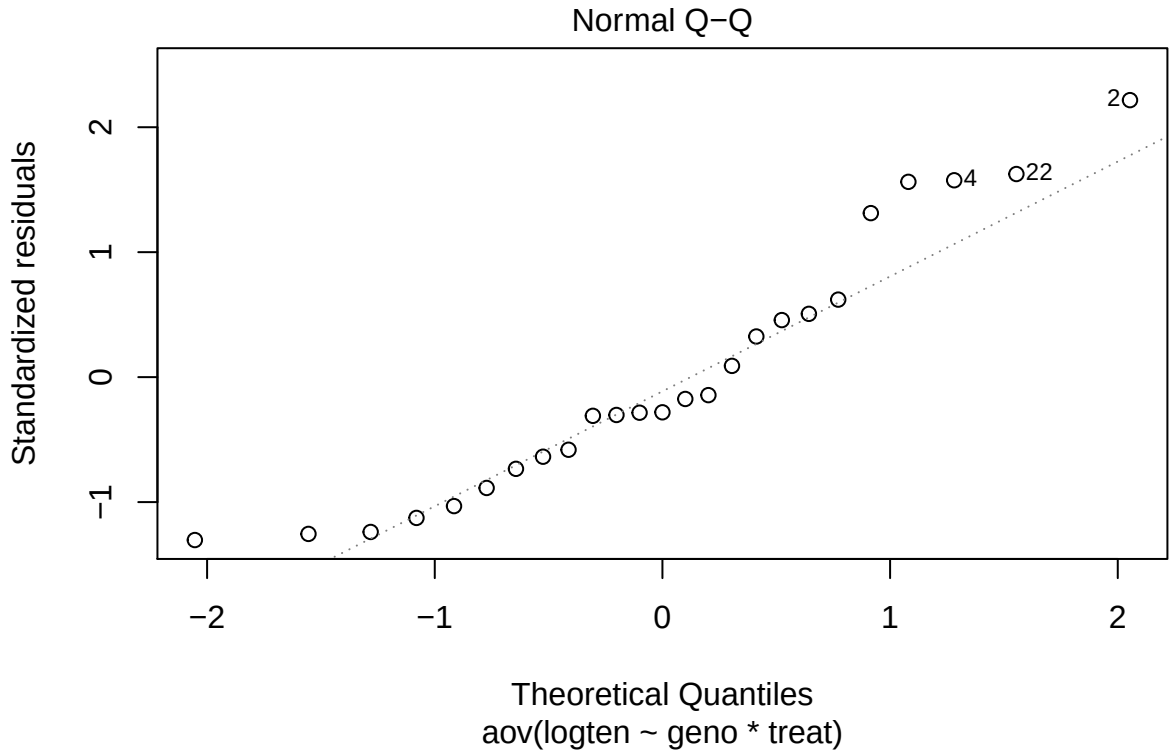
While the mean of means isn't an ideal measure, it allowed us to easily compare across groups and differences in pause frequency.

```
#load data
noodor <- read.csv("tracking_noodorant2.csv")

#Extract the mean pause length for each animal
noodoranimal <- ddply(noodor,.(genotype,datelist,animal),summarize,meanlen=mean(lens,na.rm=TRUE))

#Take means of animal means, such that each plate is the average of all animals on that plate
noodormeans <- ddply(noodoranimal,.(genotype,datelist),summarize,mean=mean(meanlen,na.rm=TRUE))

noodormeanssummary <- ddply(noodormeans,.(genotype),summarize,means=mean(mean),sd=sd(mean),N=length(mea

noodormeans$geno <- substr(noodormeans$genotype,1,1)
noodormeans$treat <- substr(noodormeans$genotype,2,2)

noodorstat.aov = aov(mean~geno*treat,data=noodormeans)
plot(noodorstat.aov,which=2)
```

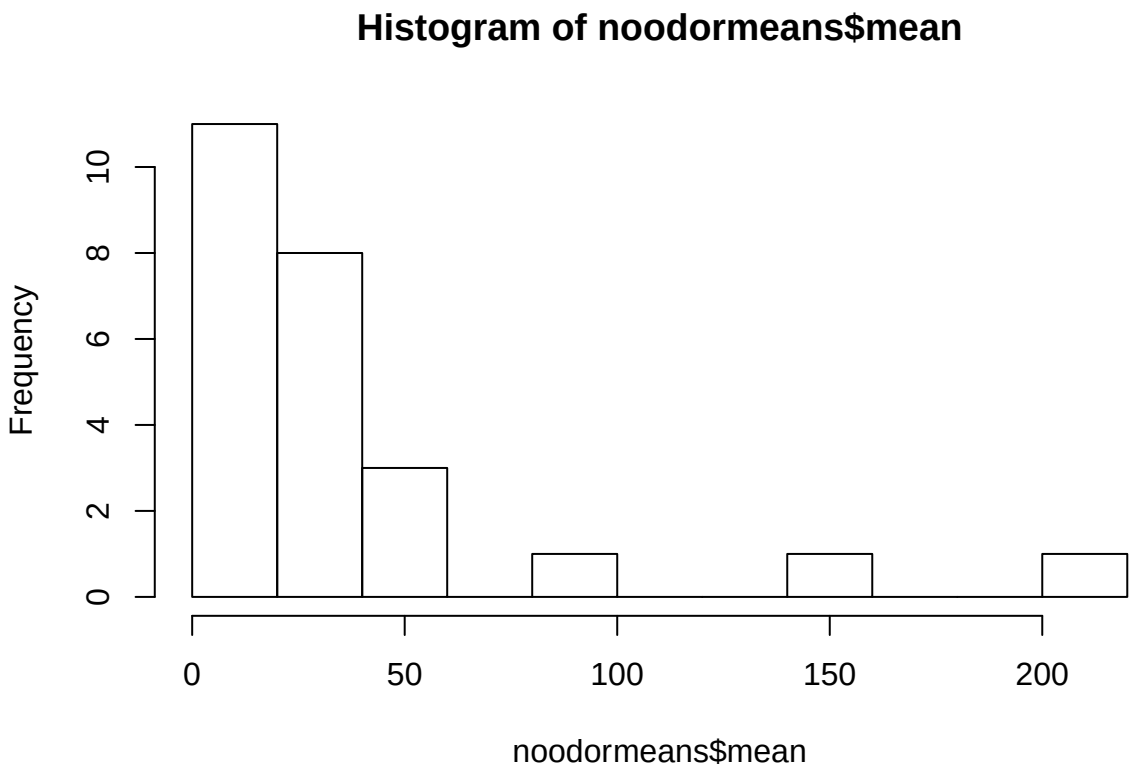



```
#This looks like a really bad fit
```

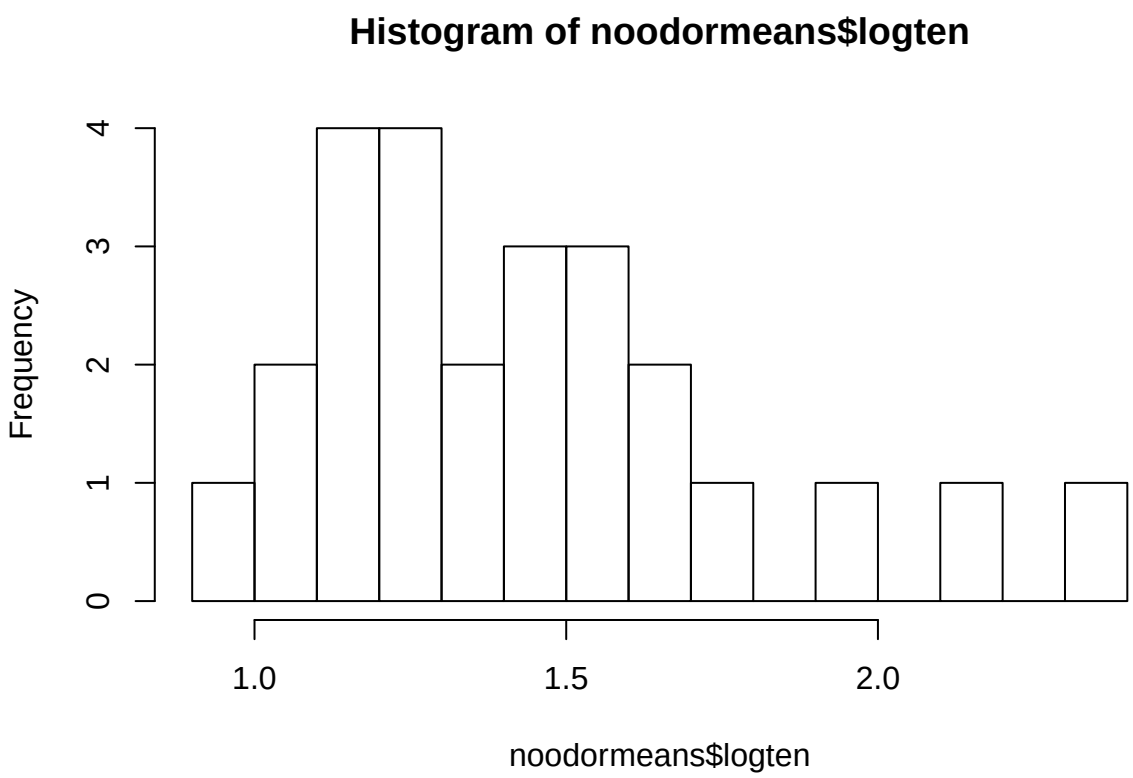
```r
noodormeans$logten <- log10(noodormeans$mean)

noodorstatbox.aov = aov(logten~geno*treat,data=noodormeans)
plot(noodorstatbox.aov,which=2)
```




```r
#This looks more reasonable!

hist(noodormeans$mean,breaks=10)
```




```
hist(noodormeans$logten,breaks=10)
```




```
#Histograms are also improved
```

```
noodortukey <- TukeyHSD(aov(formula=logten~geno*treat,data=noodormeans))

noodortukey.levels <- noodortukey$`geno:treat`[,4]

 multcompLetters(noodortukey.levels)['Letters']
```

```
## $Letters
## S:d N:e S:e N:d
## "a" "a" "a" "b"
```

```
noodormeanssummaryadj <- ddply(noodormeans,.(genotype),summarize,means=mean(logten),sd=sd(logten),N=leng

noodormeanssummaryadjback <- noodormeanssummaryadj
noodormeanssummaryadjback$means <- 10^(noodormeanssummaryadj$means)
noodormeanssummaryadjback$cim <- 10^(noodormeanssummaryadj$cim)
noodormeanssummaryadjback$cip <- 10^(noodormeanssummaryadj$cip)

ggplot(noodormeans,aes(x=genotype,(y=mean/3))) +
  geom_jitter(height=0,width=0.2) +
  theme_tufte() +
  ylab("Average pause length (seconds)") +
  theme(axis.ticks.x=element_blank()) +
  geom_point(data=noodormeanssummaryadjback,aes(x=genotype,y=(means/3)),shape='-') +
  geom_point(data=noodormeanssummaryadjback,aes(x=genotype,y=(cip/3)),shape='-') +
  geom_point(data=noodormeanssummaryadjback,aes(x=genotype,y=(cim/3)),shape='-')
```

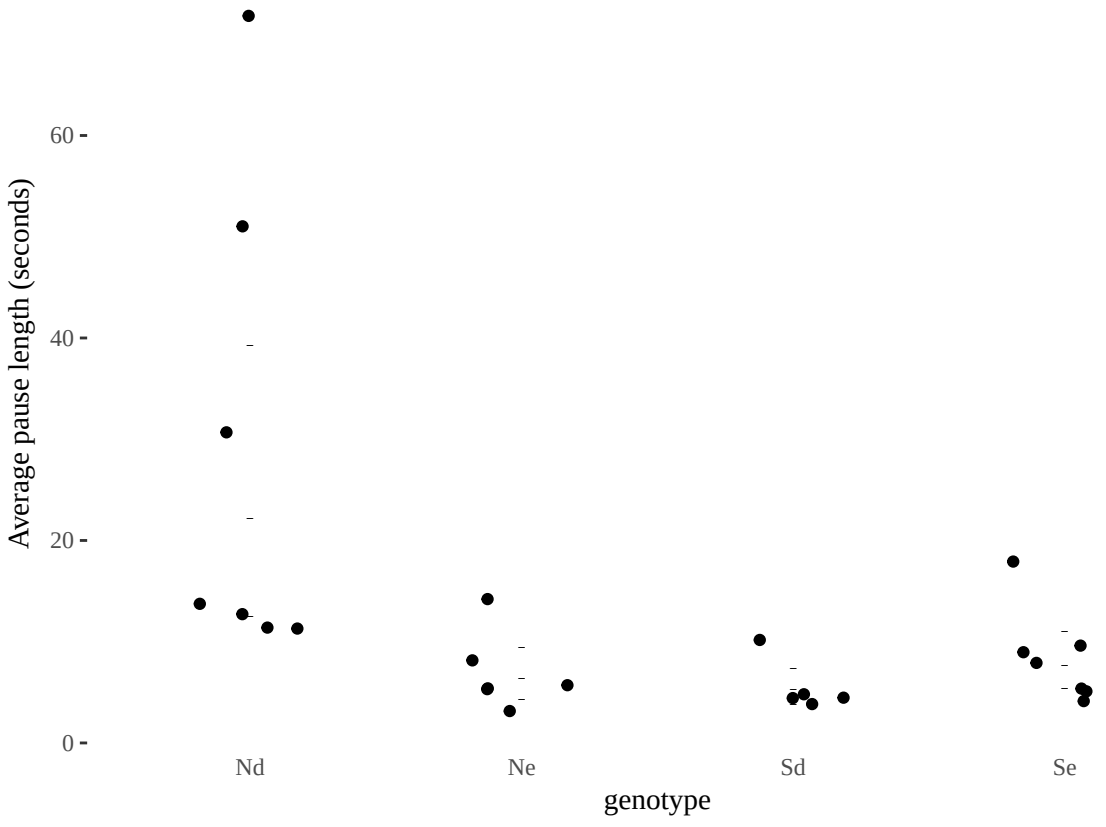
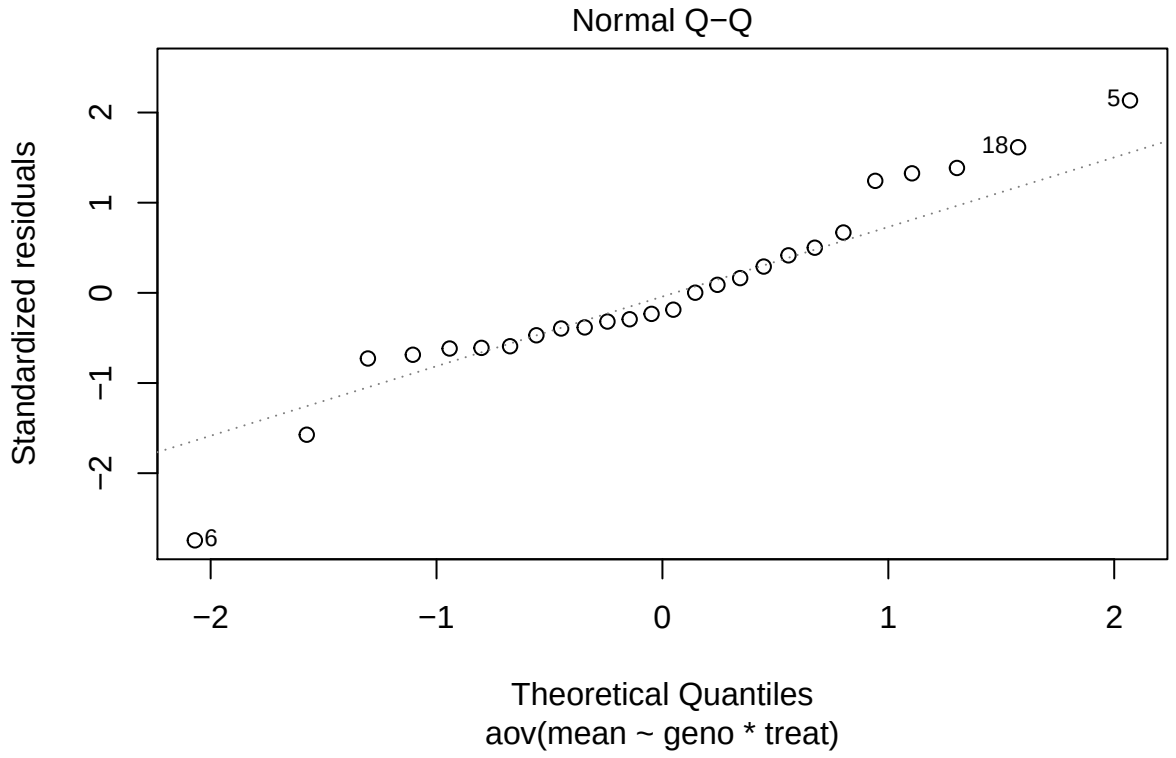


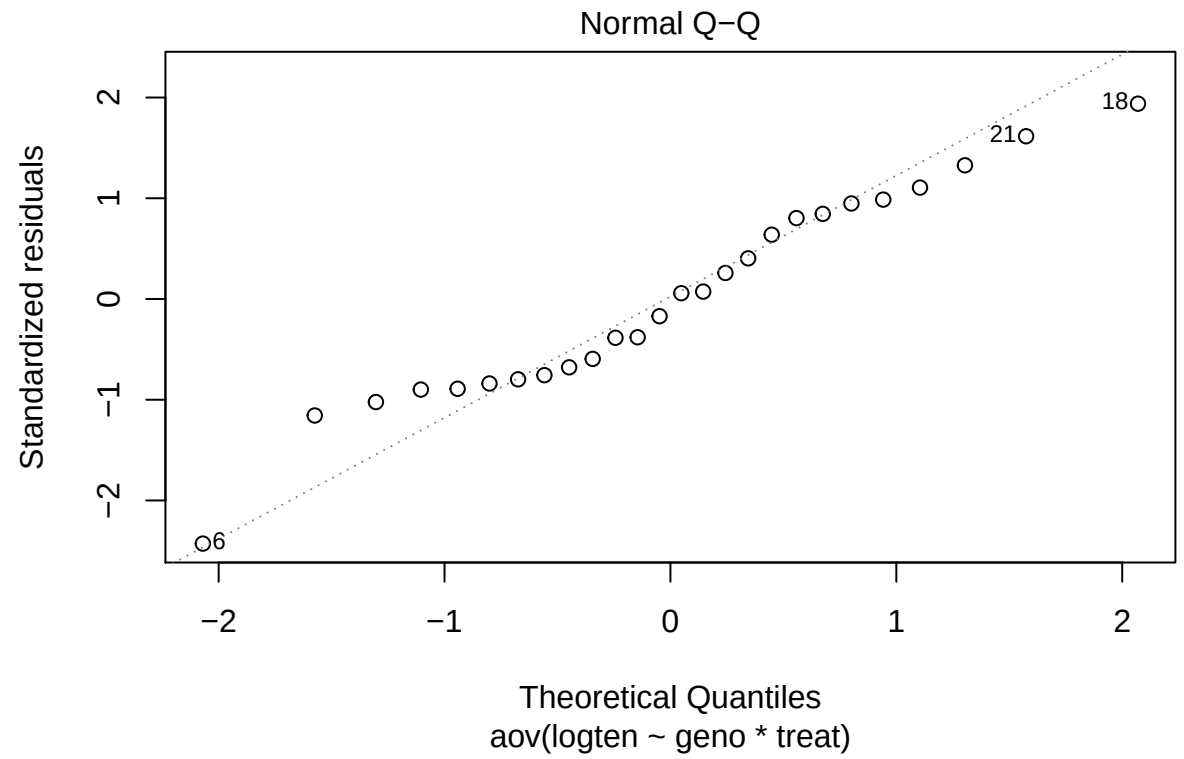
Fig 4f

On DEET-agar, wild-type animals show an increase in pausing behavior, and *ADL::Tetanus toxin* animals are resistant to this effect of DEET

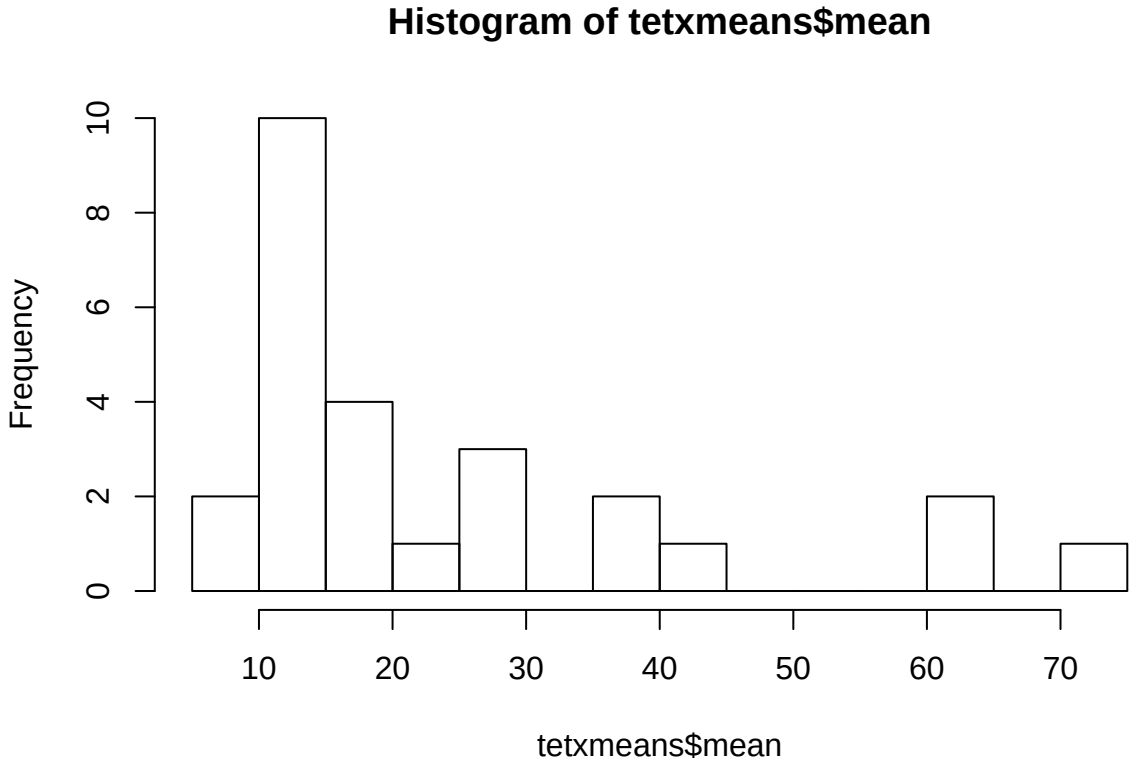
```
#load data
tetx <- read.csv('tracking_tetx.csv')

#Get mean pause length for each animal on each plate
tetxanimal <- ddply(tetx,.(genotype,datelist,animal),summarize,lengths=mean(lens,na.rm=TRUE))

#Take means of animal means, such that each plate is the average of all animals on that plate
tetxmeans <- ddply(tetxanimal,.(genotype,datelist),summarize,mean=mean(lengths,na.rm=TRUE))

tetxmeanssummary <- ddply(tetxmeans,.(genotype),summarize,means=mean(mean),sd=sd(mean),N=length(mean),s


tetxmeans$geno <- substr(tetxmeans$genotype,1,1)
tetxmeans$treat <- substr(tetxmeans$genotype,2,2)

tetxstat.aov = aov(mean~geno*treat,data=tetxmeans)
plot(tetxstat.aov,which=2)
```



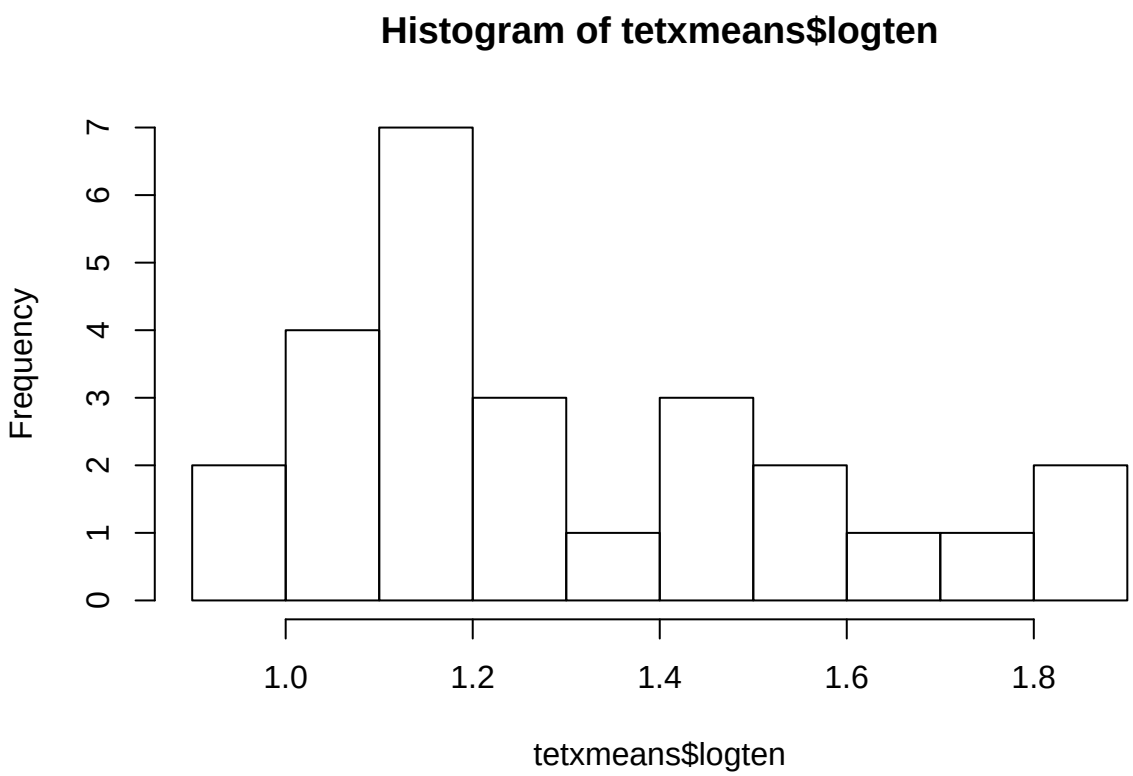

```
#This looks like a  bad fit

tetxmeans$logten <- log10(tetxmeans$mean)

tetxstatadj.aov = aov(logten~geno*treat,data=tetxmeans)
plot(tetxstatadj.aov,which=2)
```




```
#Looks better


hist(tetxmeans$mean,breaks=10)
```



```
hist(tetxmeans$logten,breaks=10)
```




```
#Histograms are also improved


tetxtukey <- TukeyHSD(aov(formula=logten~geno*treat,data=tetxmeans))

tetxtukey.levels <- tetxtukey$`geno:treat`[,4]

multcompLetters(tetxtukey.levels)['Letters']
```

```
## $Letters
## X:d N:e X:e N:d
## "a" "a" "a" "b"
```

```
tetxmeanssummaryadj <- ddply(tetxmeans,.(genotype),summarize,means=mean(logten),sd=sd(logten),N=length(

tetxmeanssummaryadjback <- tetxmeanssummaryadj
tetxmeanssummaryadjback$means <- 10^(tetxmeanssummaryadj$means)
tetxmeanssummaryadjback$cim <- 10^(tetxmeanssummaryadj$cim)
tetxmeanssummaryadjback$cip <- 10^(tetxmeanssummaryadj$cip)

ggplot(tetxmeans,aes(x=genotype,(y=mean/3))) +
  geom_jitter(height=0,width=0.2) +
  theme_tufte() +
  ylab("Average pause length (seconds)") +
  theme(axis.ticks.x=element_blank()) +
  geom_point(data=tetxmeanssummaryadjback,aes(x=genotype,y=(means/3)),shape='-') +
  geom_point(data=tetxmeanssummaryadjback,aes(x=genotype,y=(cip/3)),shape='-') +
  geom_point(data=tetxmeanssummaryadjback,aes(x=genotype,y=(cim/3)),shape='-')
```

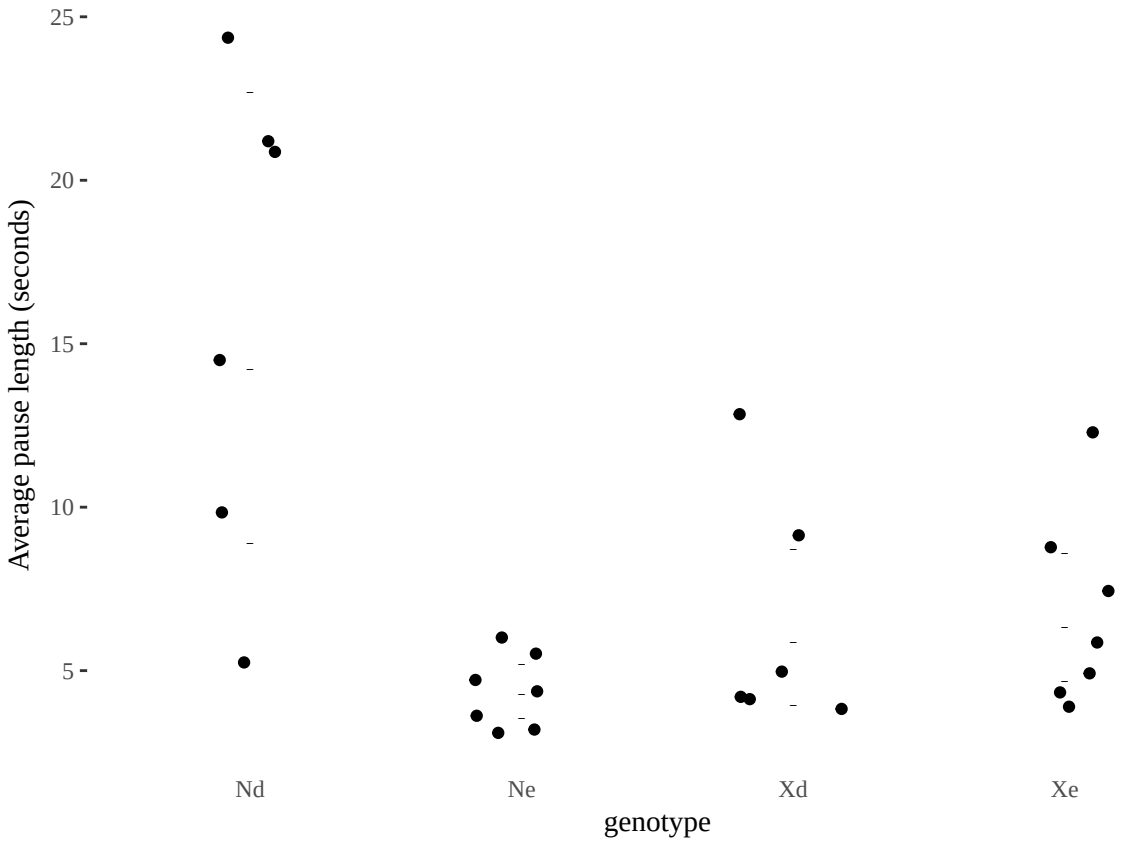


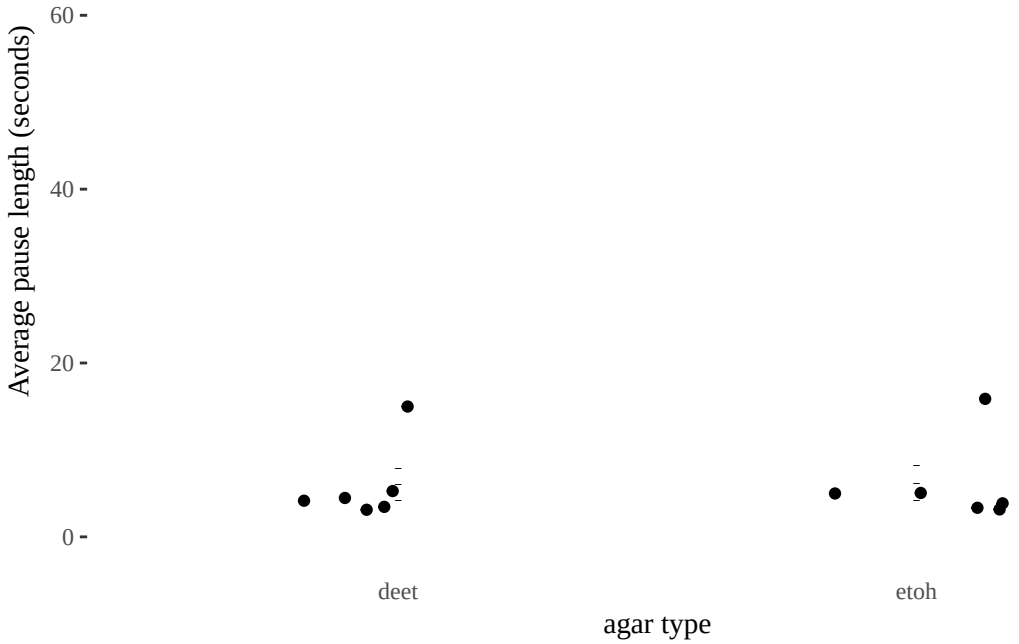
Fig 4g

Pyrazine chemotaxis is able to overcome DEET-induced increased pause lengths.

```
#Load data
pyr <- read.csv('tracking_pyr.csv')

#Get mean pause duration for each animal
pyranimal <- ddply(pyr,.(X,datelist,animal),summarize,lengths=mean(lens,na.rm=TRUE))

#Makes means of means, each plate is a mean of each animal
pyrmeans <- ddply(pyranimal,.(X,datelist),summarize,mean=mean(lengths,na.rm=TRUE))

pyrmeanssummary <- ddply(pyrmeans,.(X),summarize,means=mean(mean),sd=sd(mean),N=length(mean),se=sd/sqrt

ggplot(pyrmeans,aes(x=X,(y=mean/3))) +
  geom_jitter(height=0,width=0.2) +
  theme_tufte() +
  ylim(0,75) +
  ylab("Average pause length (seconds)") +
  xlab('agar type') +
  ggtitle('Average pause lengths of wild-type animals chemotaxing to pyrazine on DEET-agar and solvent-a
  theme(axis.ticks.x=element_blank()) +
  geom_point(data=pyrmeanssummary,aes(x=X,y=(means/3)),shape='-') +
  geom_point(data=pyrmeanssummary,aes(x=X,y=(mep/3)),shape='-') +
```

```
  geom_point(data=pyrmeanssummary,aes(x=X,y=(mem/3)),shape='-')
```

Average pause lengths of wild−type animals chemotaxing to pyrazine on DEET−ag




```
pyrdeet <- subset(pyrmeans,X=='deet')
pyretoh <- subset(pyrmeans,X=='etoh')

t.test(pyrdeet$mean,pyretoh$mean)
```

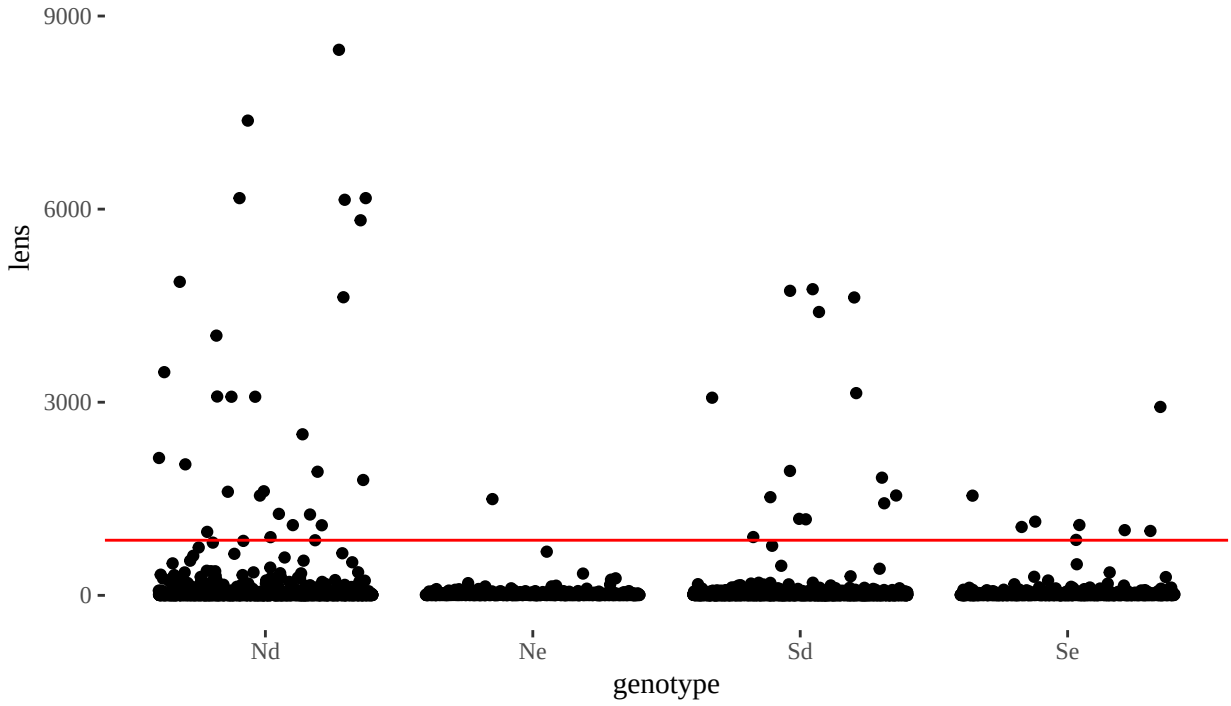
```
##
##  Welch Two Sample t-test
##
## data:  pyrdeet$mean and pyretoh$mean
## t = -0.049965, df = 9.9392, p-value = 0.9611
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.56624  17.75249
## sample estimates:
## mean of x mean of y
##  17.73097  18.13784
```

These data were not included in the figures but I found them useful in thinking about and interpreting these data

IAA chemotaxis

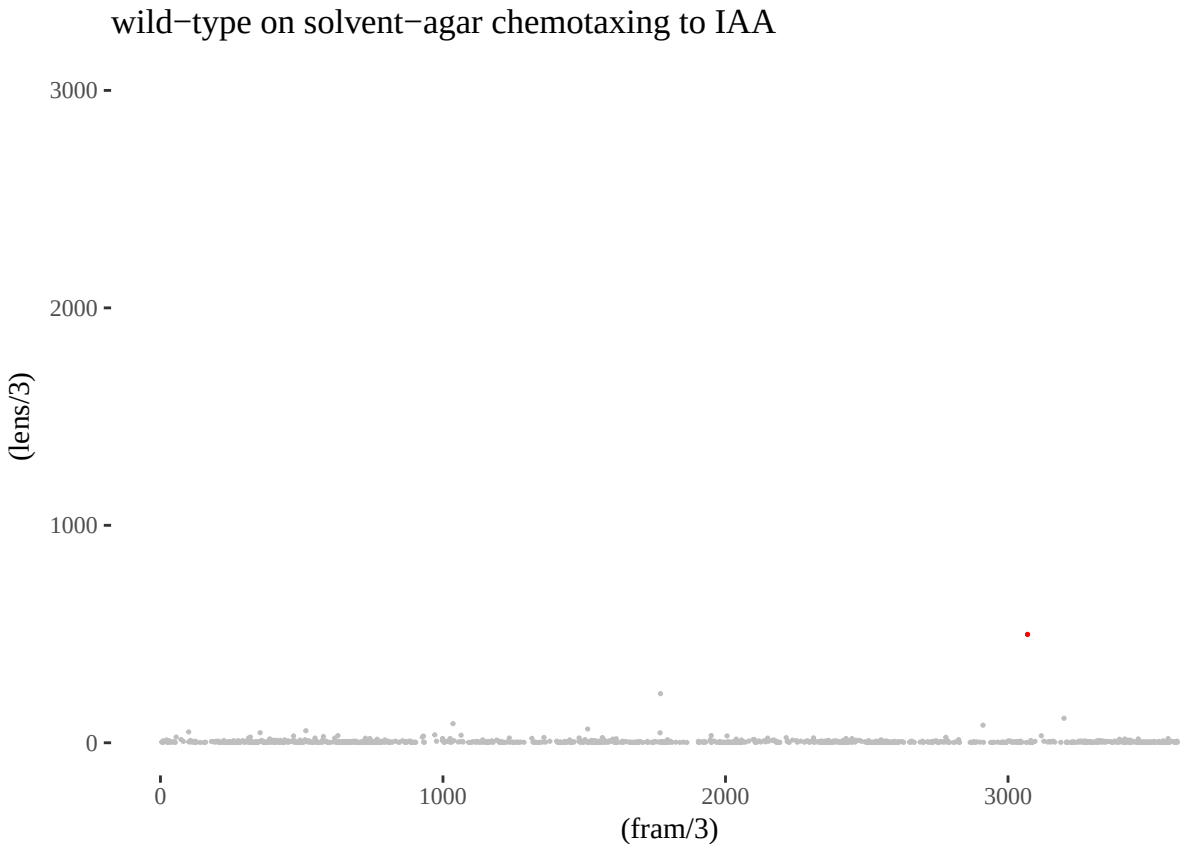
- I plotted all pauses, by genotype and drew a red line at 3 standard deviations above the mean of all pause lengths
- I also plotted all pauses for each genotype/treatment combination, plotted by the last frame of the pause (x) and the length of the pause (y). Pauses were colored red if they were 3 standard deviations above the GROUP mean (above the red line in the first graph)

```
ind <- subset(iaa,genotype=='Nd')
ine <- subset(iaa,genotype=='Ne')
isd <- subset(iaa,genotype=='Sd')
ise <- subset(iaa,genotype=='Se')

m <- mean(iaa$lens)
sdm <- sd(iaa$lens)
minlength <- m+(3*sdm)

above3sdofmean <- subset(iaa,lens>minlength)

ggplot(iaa,aes(x=genotype,y=lens)) + geom_jitter() + theme_tufte() + ylim(0,10800) + geom_hline(yinterc
```

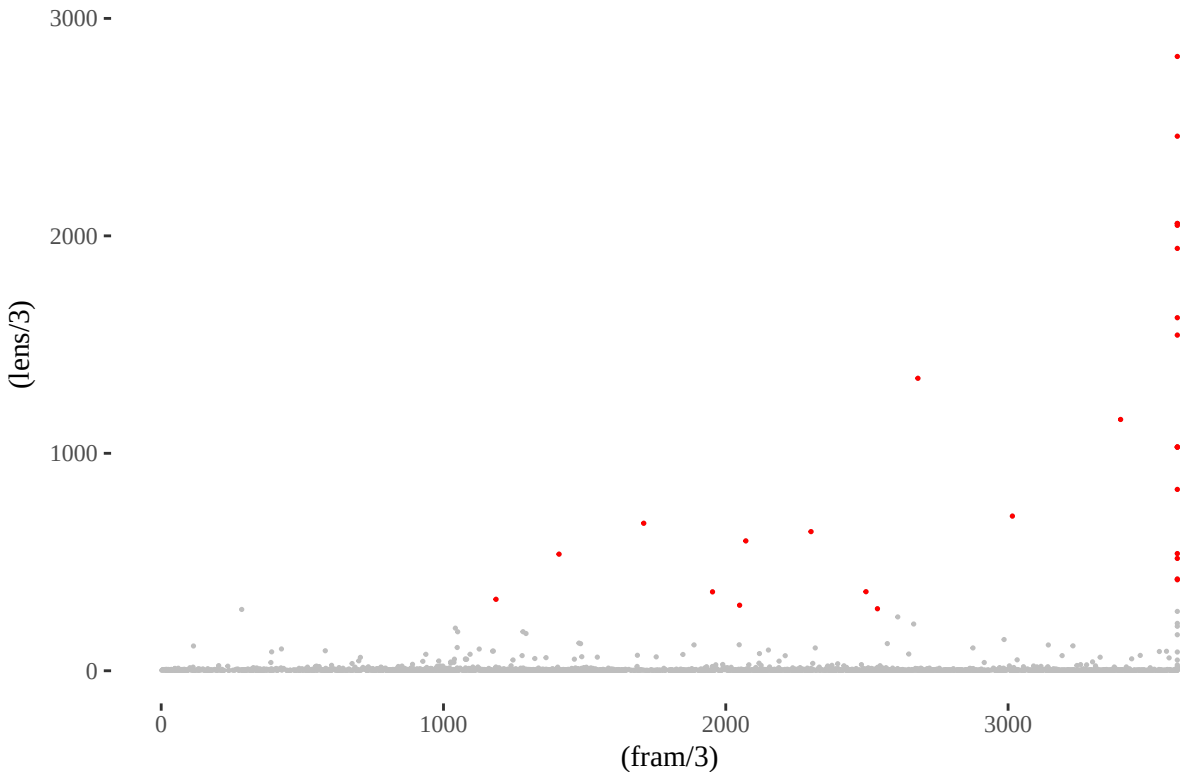



```
ndabove <- subset(above3sdofmean,genotype=='Nd')
neabove <- subset(above3sdofmean,genotype=='Ne')
sdabove <- subset(above3sdofmean,genotype=='Sd')
seabove <- subset(above3sdofmean,genotype=='Se')

ggplot(ine,aes(x=(fram/3),y=(lens/3))) +
  geom_point(size=0.25,color='gray') +
```

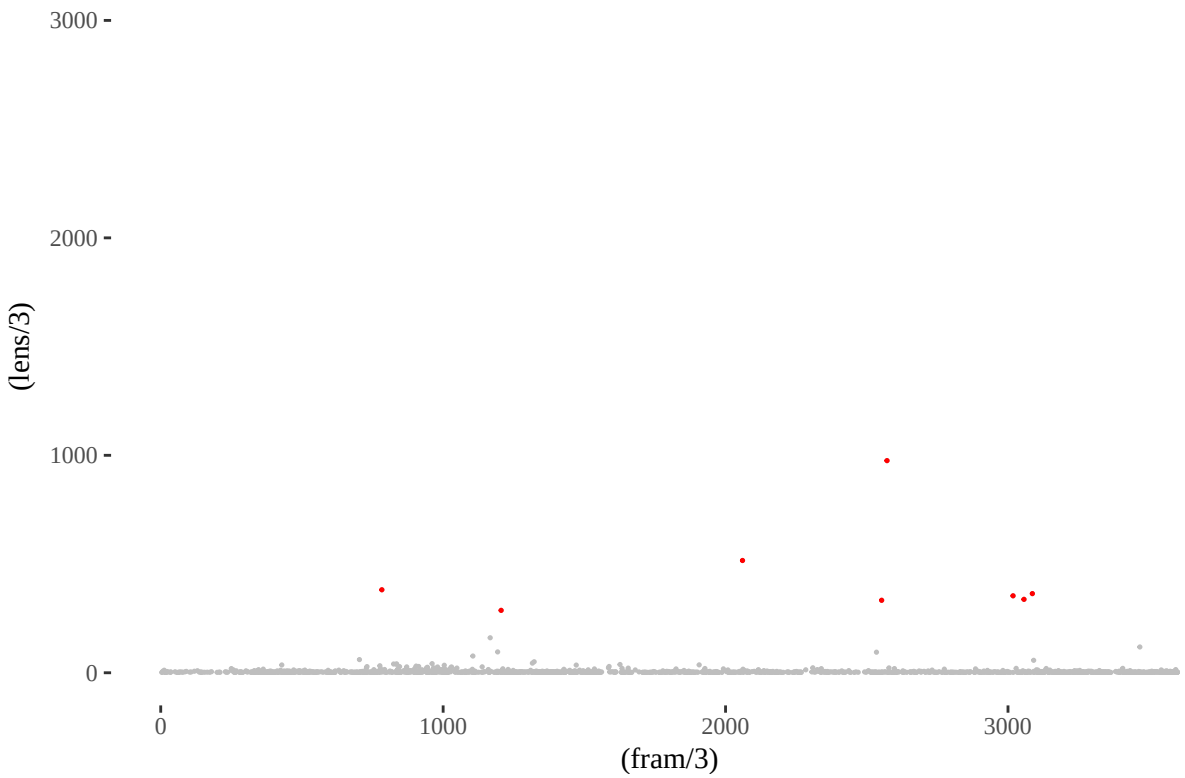
```
  theme_tufte() +
  ggtitle('wild-type on solvent-agar chemotaxing to IAA') +
  geom_point(data=neabove,aes(x=(fram/3),y=(lens/3)),color='red', size=0.25) +
  ylim(0,3000)
```



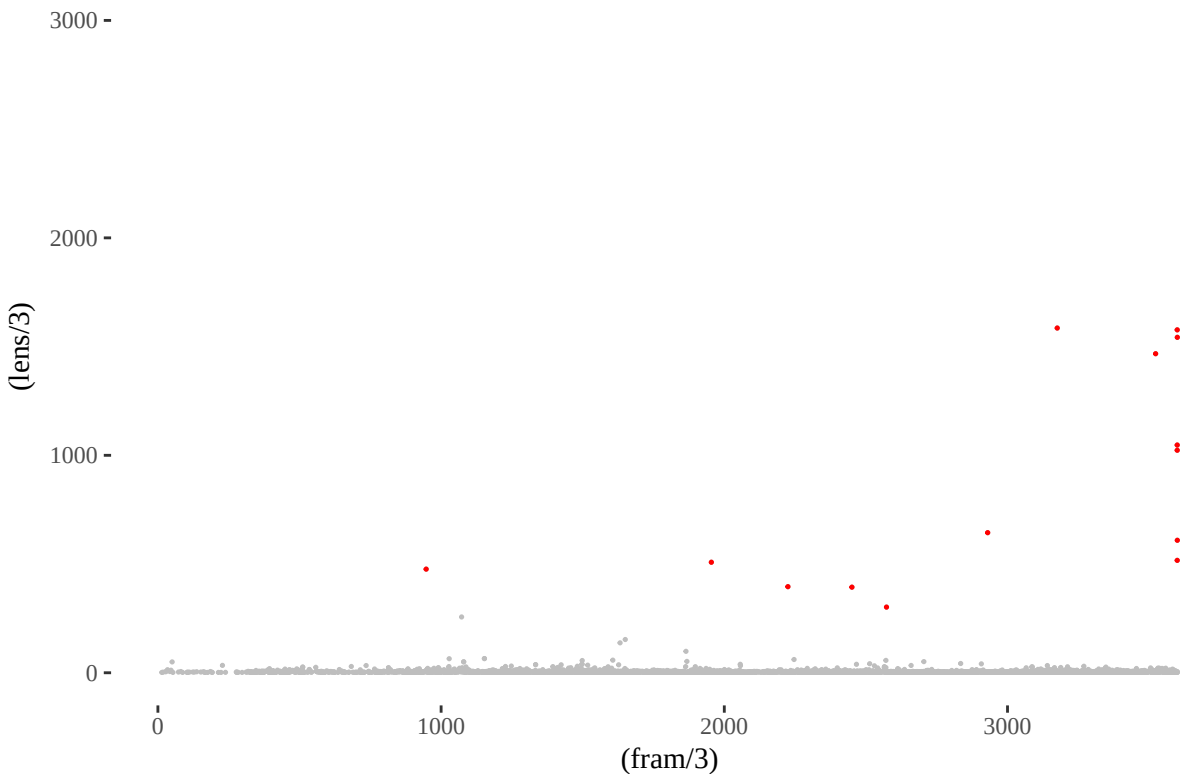

```
ggplot(ind,aes(x=(fram/3),y=(lens/3))) +
  geom_point(size=0.25,color='gray') +
  theme_tufte() +
  ggtitle('wild-type on DEET-agar chemotaxing to IAA') +
  geom_point(data=ndabove,aes(x=(fram/3),y=(lens/3)),color='red', size=0.25) +
  ylim(0,3000)
```




```
ggplot(ise,aes(x=(fram/3),y=(lens/3))) +
  geom_point(size=0.25,color='gray') +
  ggtitle('str-217 mutants on solvent-agar chemotaxing to IAA') +
  theme_tufte() +
  geom_point(data=seabove,aes(x=(fram/3),y=(lens/3)),color='red', size=0.25) +
  ylim(0,3000)
```

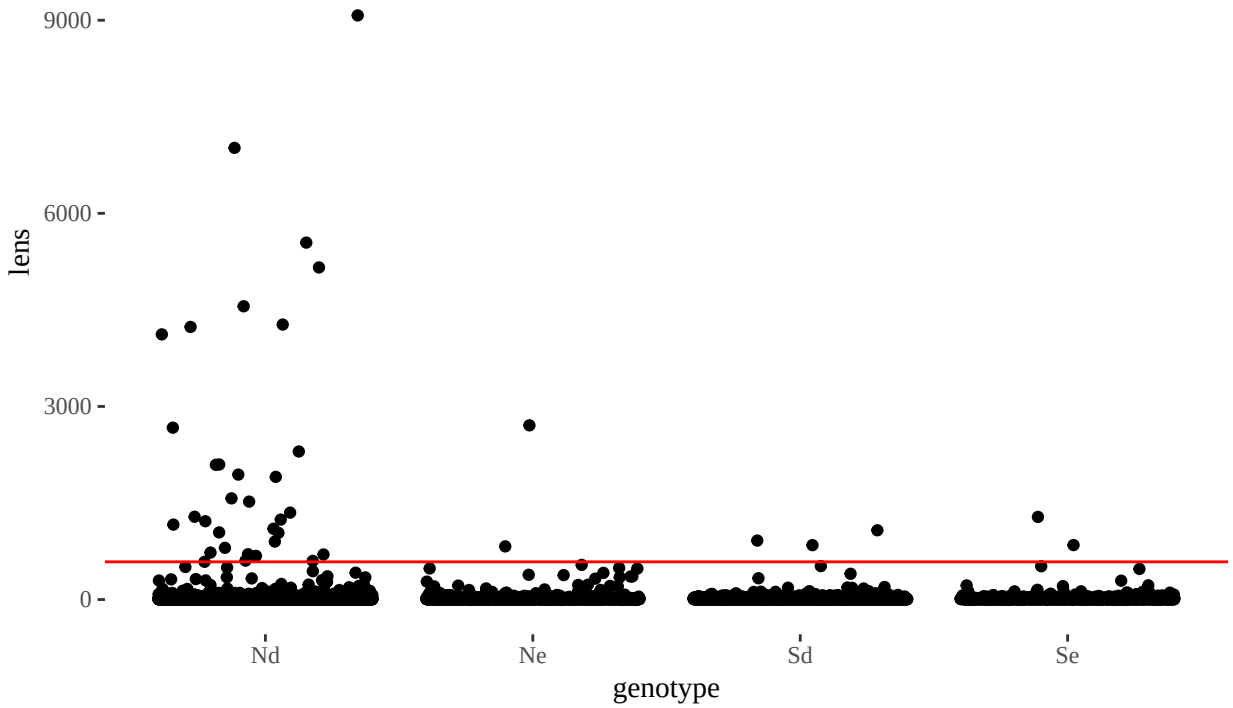



```
ggplot(isd,aes(x=(fram/3),y=(lens/3))) +
  geom_point(size=0.25,color='gray') +
  ggtitle('str-217 mutants on DEET-agar chemotaxing to IAA') +
  theme_tufte() +
  geom_point(data=sdabove,aes(x=(fram/3),y=(lens/3)),color='red', size=0.25) +
  ylim(0,3000)
```



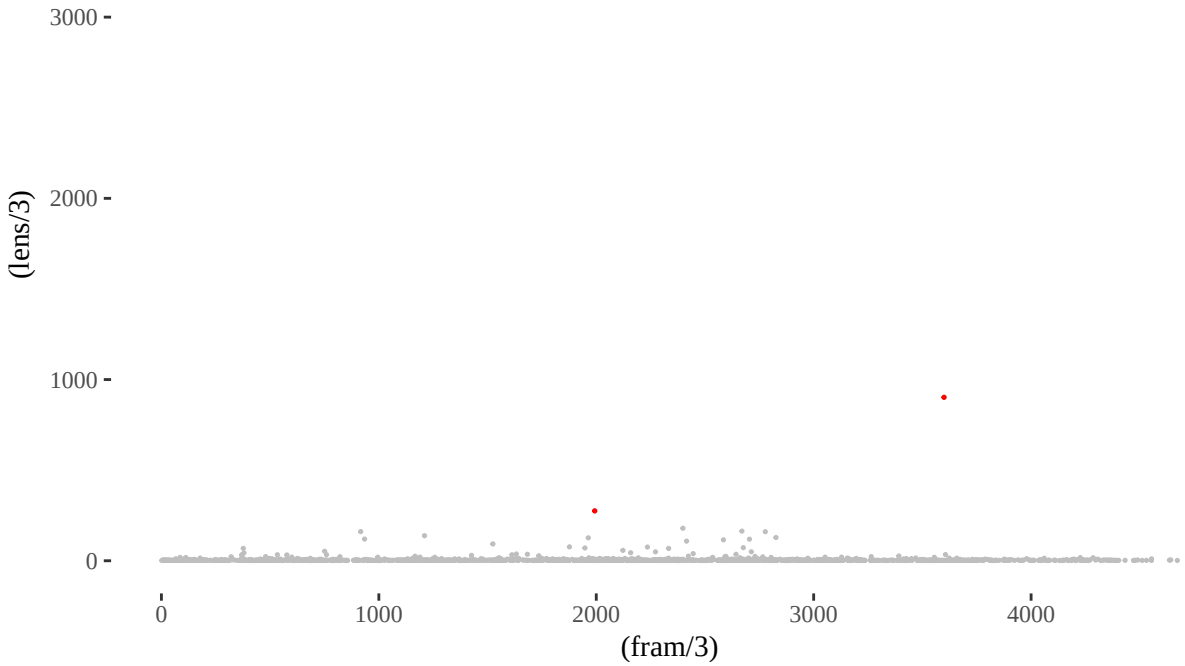
No odorant

```
nond <- subset(noodor,genotype=='Nd')
none <- subset(noodor,genotype=='Ne')
nosd <- subset(noodor,genotype=='Sd')
nose <- subset(noodor,genotype=='Se')

nom <- mean(noodor$lens)
nosdm <- sd(noodor$lens)
nominlength <- nom+(3*nosdm)

noabove3sdofmean <- subset(noodor,lens>nominlength)

ggplot(noodor,aes(x=genotype,y=lens)) + geom_jitter() + theme_tufte() + ylim(0,10800) + geom_hline(yint
```



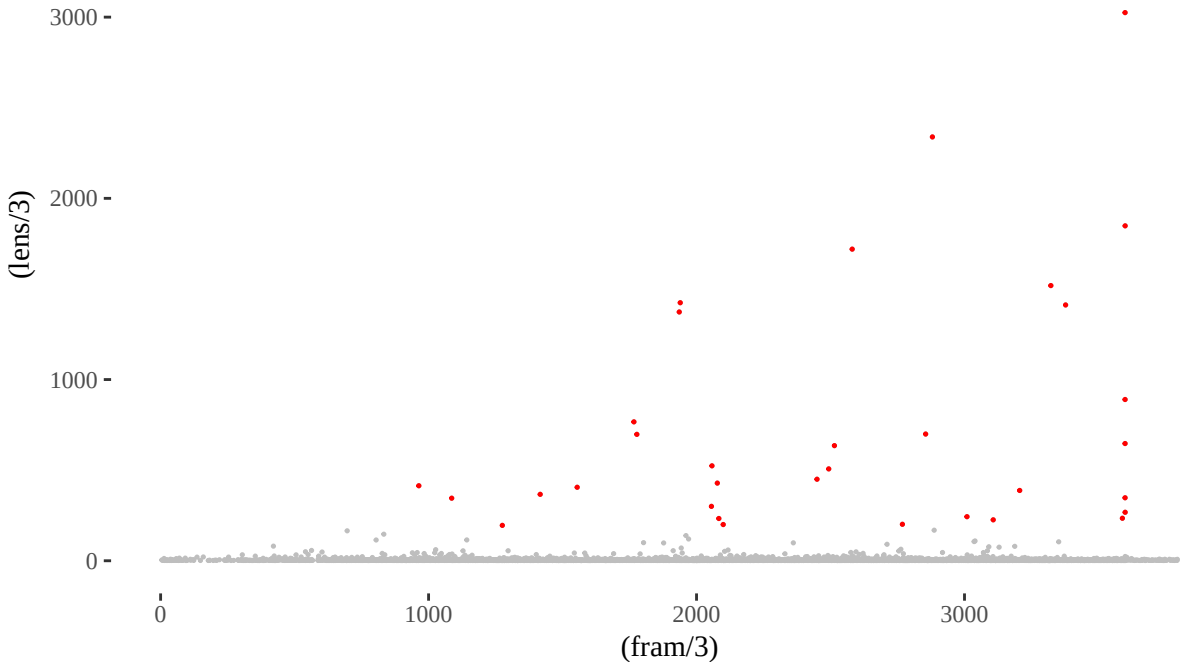

```
nondabove <- subset(noabove3sdofmean,genotype=='Nd')
noneabove <- subset(noabove3sdofmean,genotype=='Ne')
nosdabove <- subset(noabove3sdofmean,genotype=='Sd')
noseabove <- subset(noabove3sdofmean,genotype=='Se')

ggplot(none,aes(x=(fram/3),y=(lens/3))) +
  geom_point(size=0.25,color='gray') +
  theme_tufte() +
  ggtitle('wild-type on solvent-agar') +
  geom_point(data=noneabove,aes(x=(fram/3),y=(lens/3)),color='red', size=0.25) +
  ylim(0,3600)
```

wild−type on solvent−agar

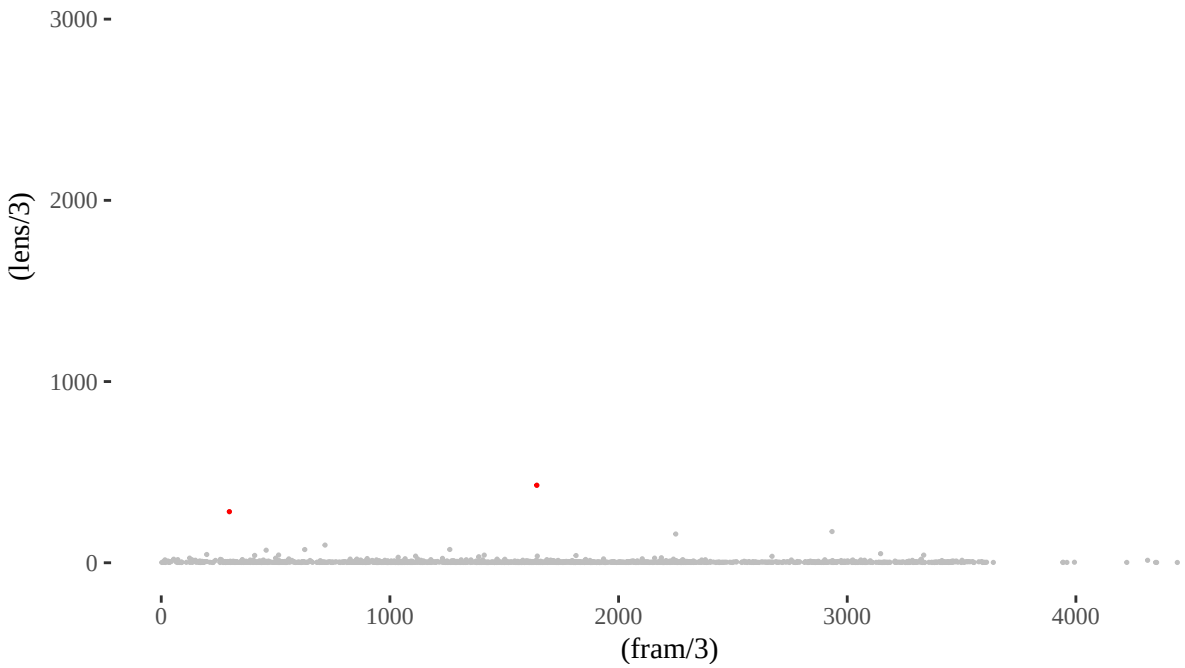



```
ggplot(nond,aes(x=(fram/3),y=(lens/3))) +
  geom_point(size=0.25,color='gray') +
  theme_tufte() +
  ggtitle('wild-type on DEET-agar') +
  geom_point(data=nondabove,aes(x=(fram/3),y=(lens/3)),color='red', size=0.25) +
  ylim(0,3600)
```

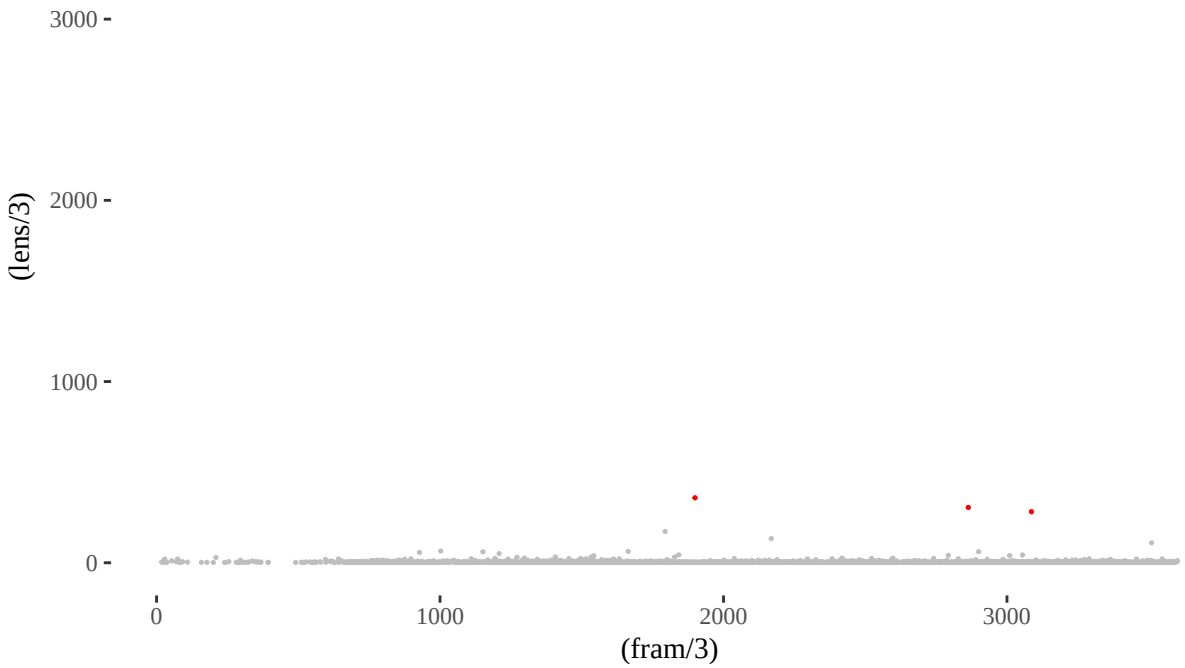
wild−type on DEET−agar




```
ggplot(nose,aes(x=(fram/3),y=(lens/3))) +
  geom_point(size=0.25,color='gray') +
  ggtitle('str-217 mutants on solvent-agar') +
  theme_tufte() +
  geom_point(data=noseabove,aes(x=(fram/3),y=(lens/3)),color='red', size=0.25) +
  ylim(0,3600)
```

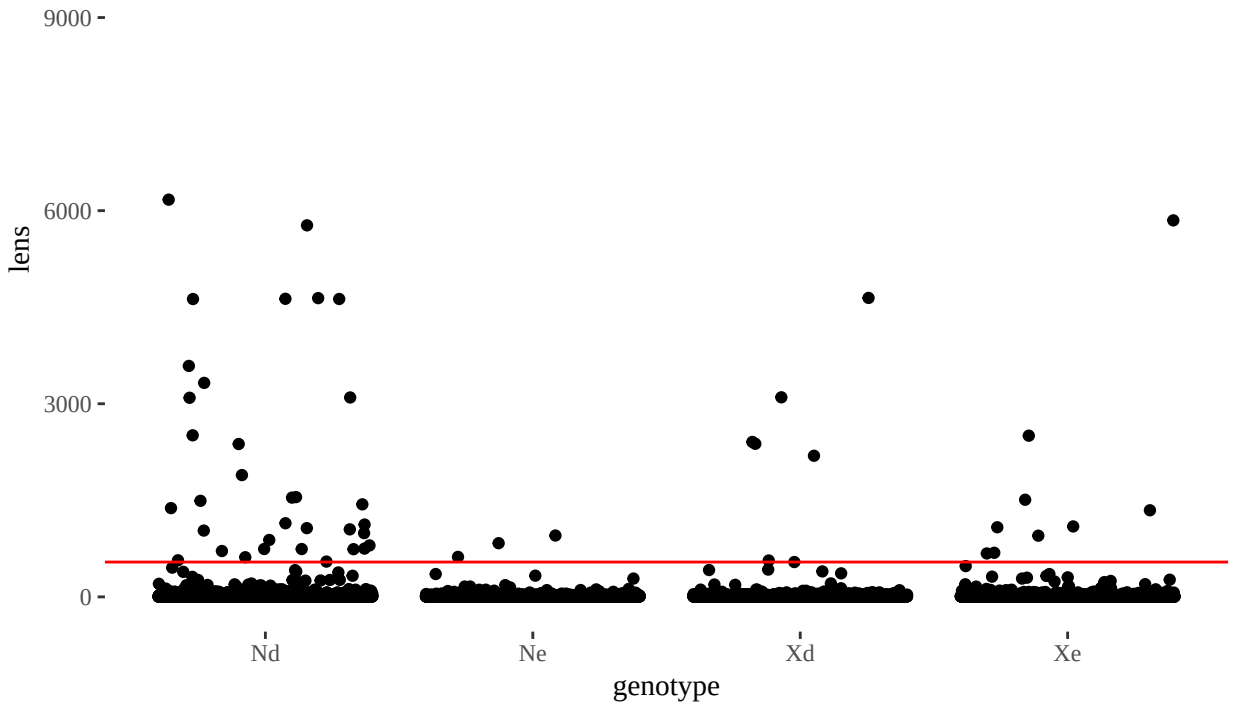



```
ggplot(nosd,aes(x=(fram/3),y=(lens/3))) +
  geom_point(size=0.25,color='gray') +
  ggtitle('str-217 mutants on DEET-agar') +
  theme_tufte() +
  geom_point(data=nosdabove,aes(x=(fram/3),y=(lens/3)),color='red', size=0.25) +
  ylim(0,3600)
```



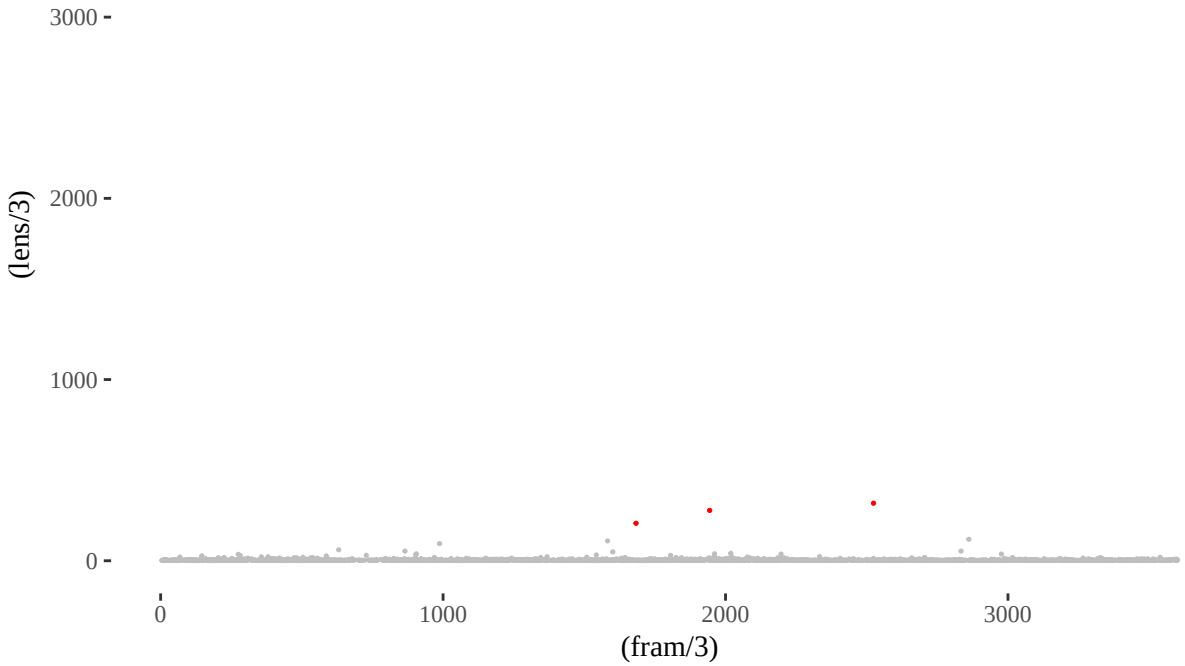
for no odorant testing of ADL::TeTX

```
noxnd <- subset(tetx,genotype=='Nd')
noxne <- subset(tetx,genotype=='Ne')
noxd <- subset(tetx,genotype=='Xd')
noxe <- subset(tetx,genotype=='Xe')

noxm <- mean(tetx$lens)
noxsdm <- sd(tetx$lens)
noxminlength <- nom+(3*noxsdm)

noxabove3sdofmean <- subset(tetx,lens>noxminlength)

ggplot(tetx,aes(x=genotype,y=lens)) + geom_jitter() + theme_tufte() + ylim(0,10800) + geom_hline(yinter
```



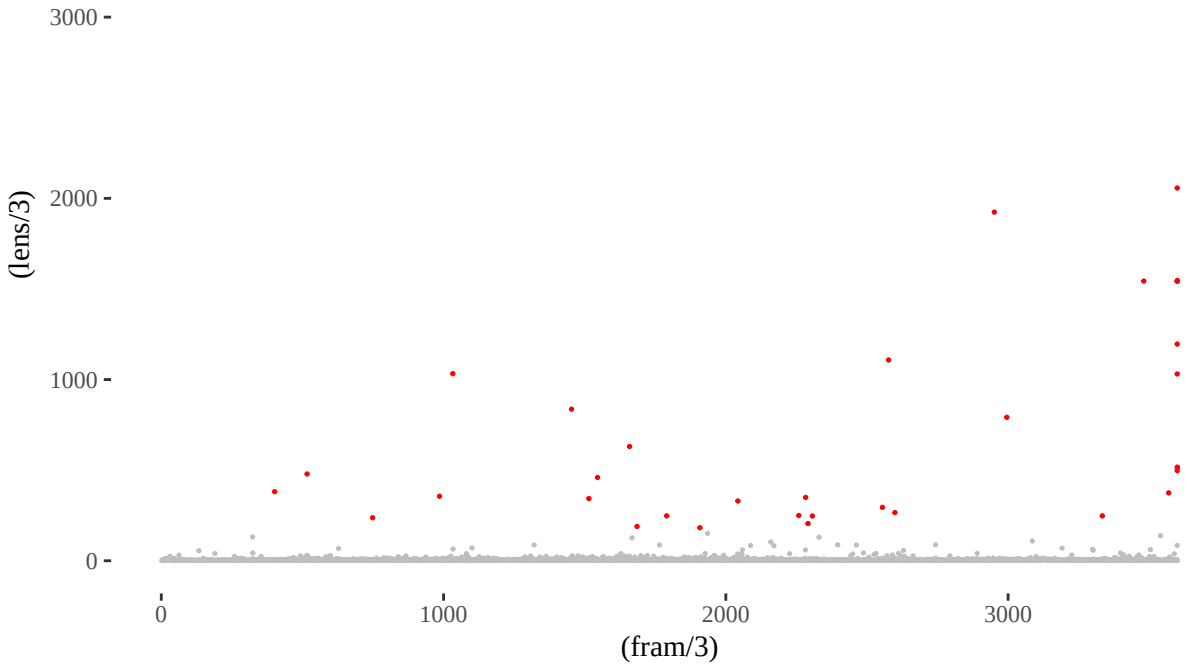

```
noxndabove <- subset(noxabove3sdofmean,genotype=='Nd')
noxneabove <- subset(noxabove3sdofmean,genotype=='Ne')
noxdabove <- subset(noxabove3sdofmean,genotype=='Xd')
noxeabove <- subset(noxabove3sdofmean,genotype=='Xe')

ggplot(noxne,aes(x=(fram/3),y=(lens/3))) +
  geom_point(size=0.25,color='gray') +
  theme_tufte() +
  ggtitle('wild-type on solvent-agar') +
  geom_point(data=noxneabove,aes(x=(fram/3),y=(lens/3)),color='red', size=0.25) +
  ylim(0,3600)
```

wild−type on solvent−agar



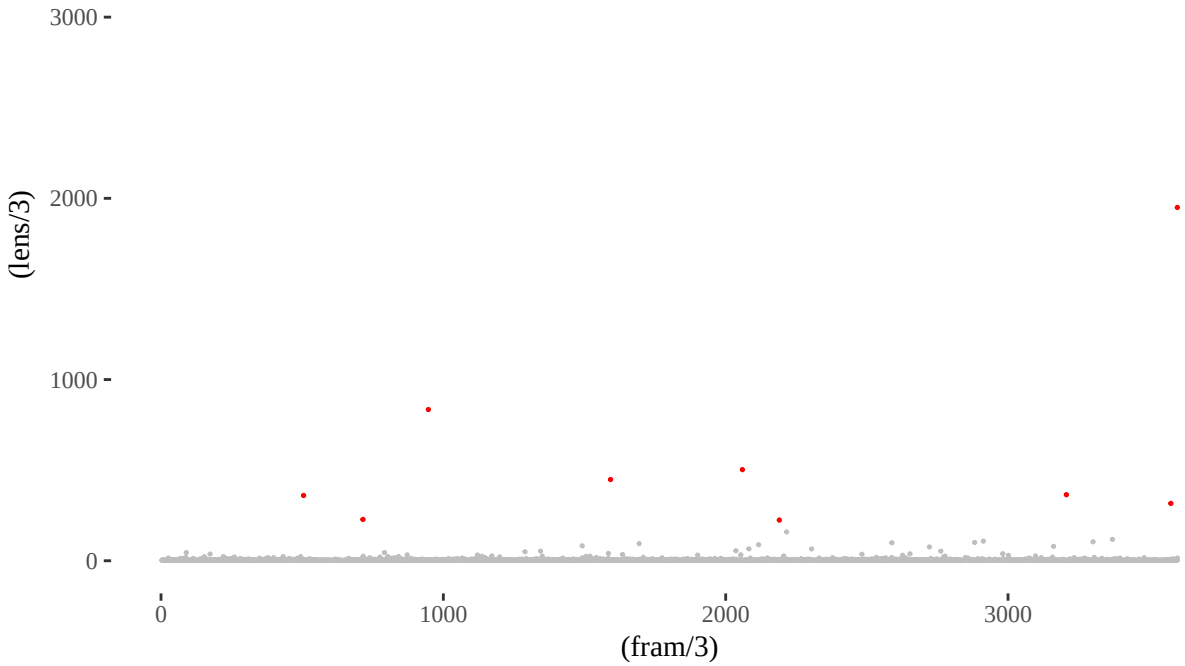

```
ggplot(noxnd,aes(x=(fram/3),y=(lens/3))) +
  geom_point(size=0.25,color='gray') +
  theme_tufte() +
  ggtitle('wild-type on DEET-agar') +
  geom_point(data=noxndabove,aes(x=(fram/3),y=(lens/3)),color='red', size=0.25) +
  ylim(0,3600)
```

wild−type on DEET−agar

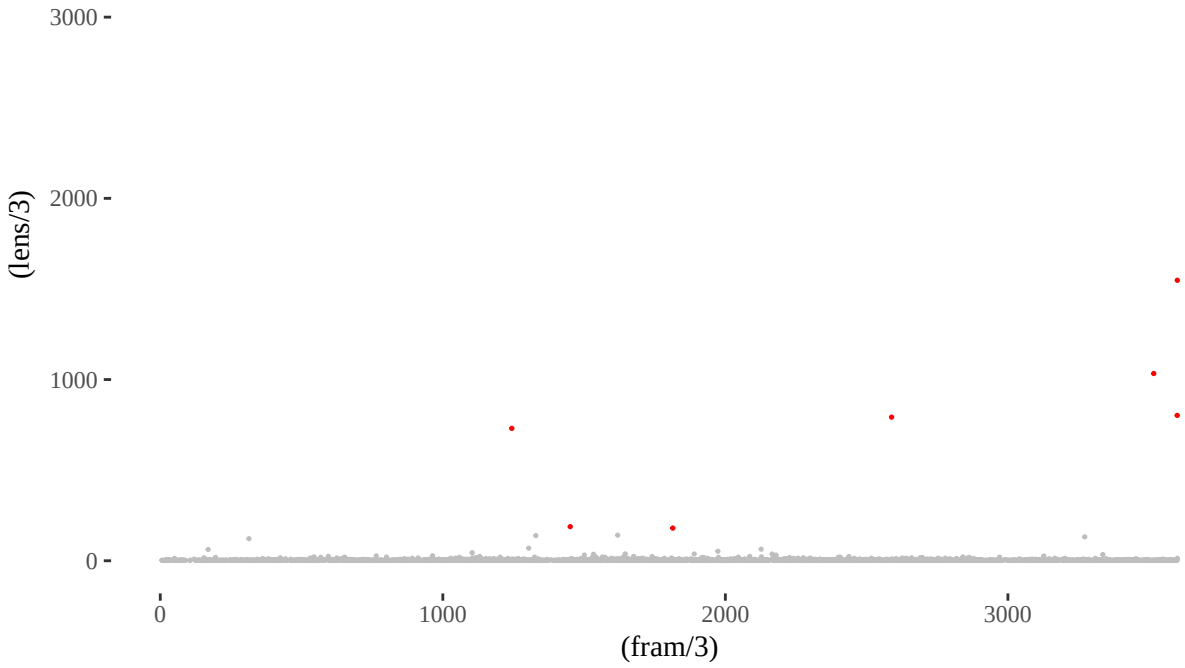



```
ggplot(noxe,aes(x=(fram/3),y=(lens/3))) +
  geom_point(size=0.25,color='gray') +
  ggtitle('ADL::TeTX animals on solvent-agar') +
  theme_tufte() +
  geom_point(data=noxeabove,aes(x=(fram/3),y=(lens/3)),color='red', size=0.25) +
  ylim(0,3600)
```

ADL::TeTX animals on solvent−agar

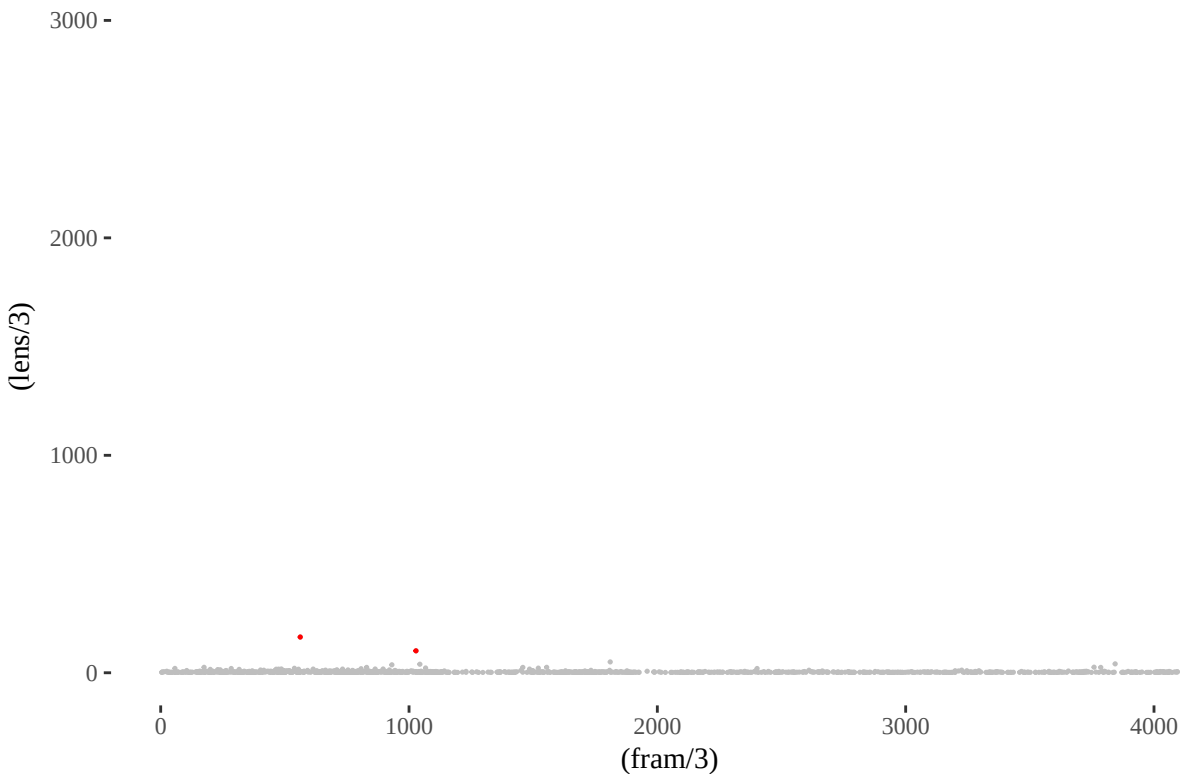



```
ggplot(noxd,aes(x=(fram/3),y=(lens/3))) +
  geom_point(size=0.25,color='gray') +
  ggtitle('ADL::TeTX animals on DEET-agar') +
  theme_tufte() +
  geom_point(data=noxdabove,aes(x=(fram/3),y=(lens/3)),color='red', size=0.25) +
  ylim(0,3600)
```

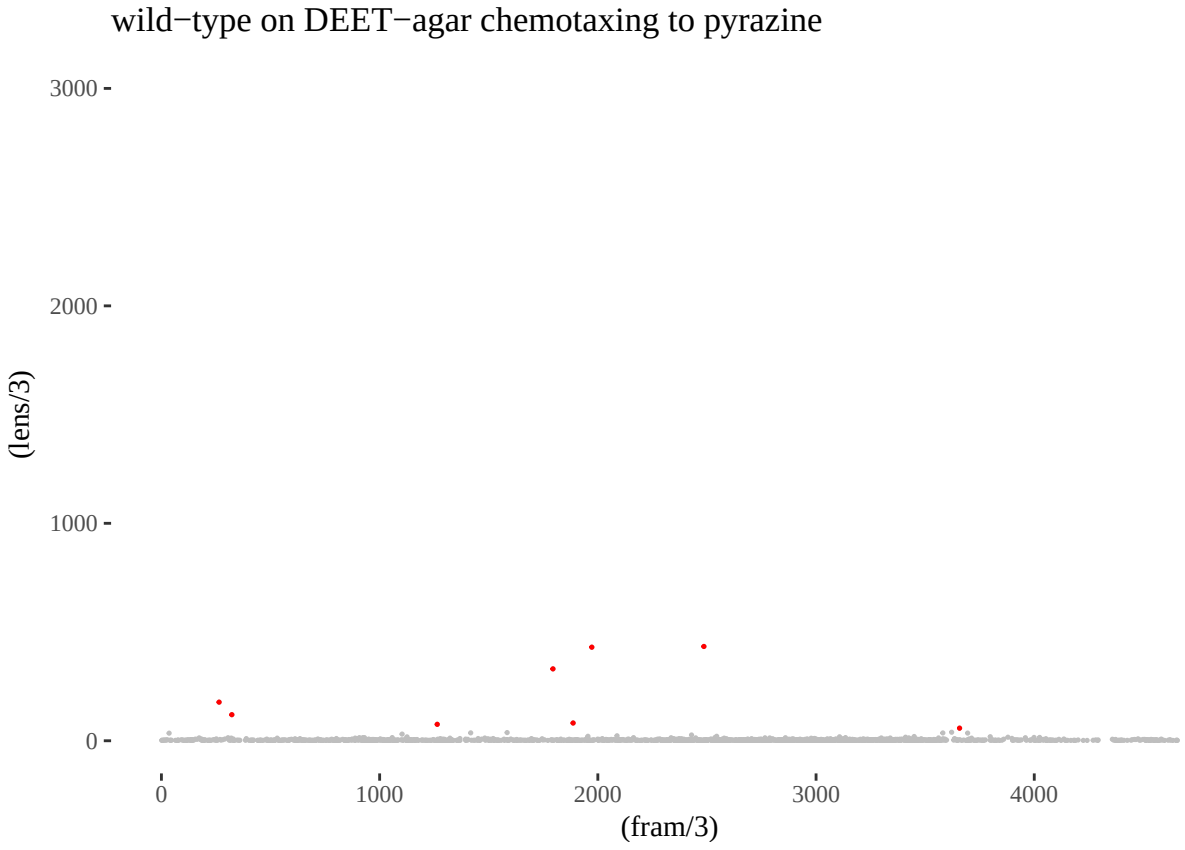


Pyrazine chemotaxis

```
pyrd <- subset(pyr,X=='deet')
pyre <- subset(pyr,X=='etoh')

mpyr <- mean(pyr$lens)
sdmpyr <- sd(pyr$lens)
minlengthpyr <- mpyr+(3*sdmpyr)

above3sdofmeanpyr <- subset(pyr,lens>minlengthpyr)
pyrdabove <- subset(above3sdofmeanpyr,X=='deet')
pyreabove <- subset(above3sdofmeanpyr,X=='etoh')

ggplot(pyre,aes(x=(fram/3),y=(lens/3))) +
  geom_point(size=0.25,color='gray') +
  theme_tufte() +
  ggtitle('wild-type on solvent-agar chemotaxing to pyrazine') +
  geom_point(data=pyreabove,aes(x=(fram/3),y=(lens/3)),color='red', size=0.25) +
  ylim(0,3000)
```




```
ggplot(pyrd,aes(x=(fram/3),y=(lens/3))) +
  geom_point(size=0.25,color='gray') +
  theme_tufte() +
  ggtitle('wild-type on DEET-agar chemotaxing to pyrazine') +
  geom_point(data=pyrdabove,aes(x=(fram/3),y=(lens/3)),color='red', size=0.25)  +
  ylim(0,3000)
```


wild−type on DEET−agar chemotaxing to pyrazine
3000
2000
1000
0
(lens/3)
0
1000
2000
3000
4000
(fram/3)