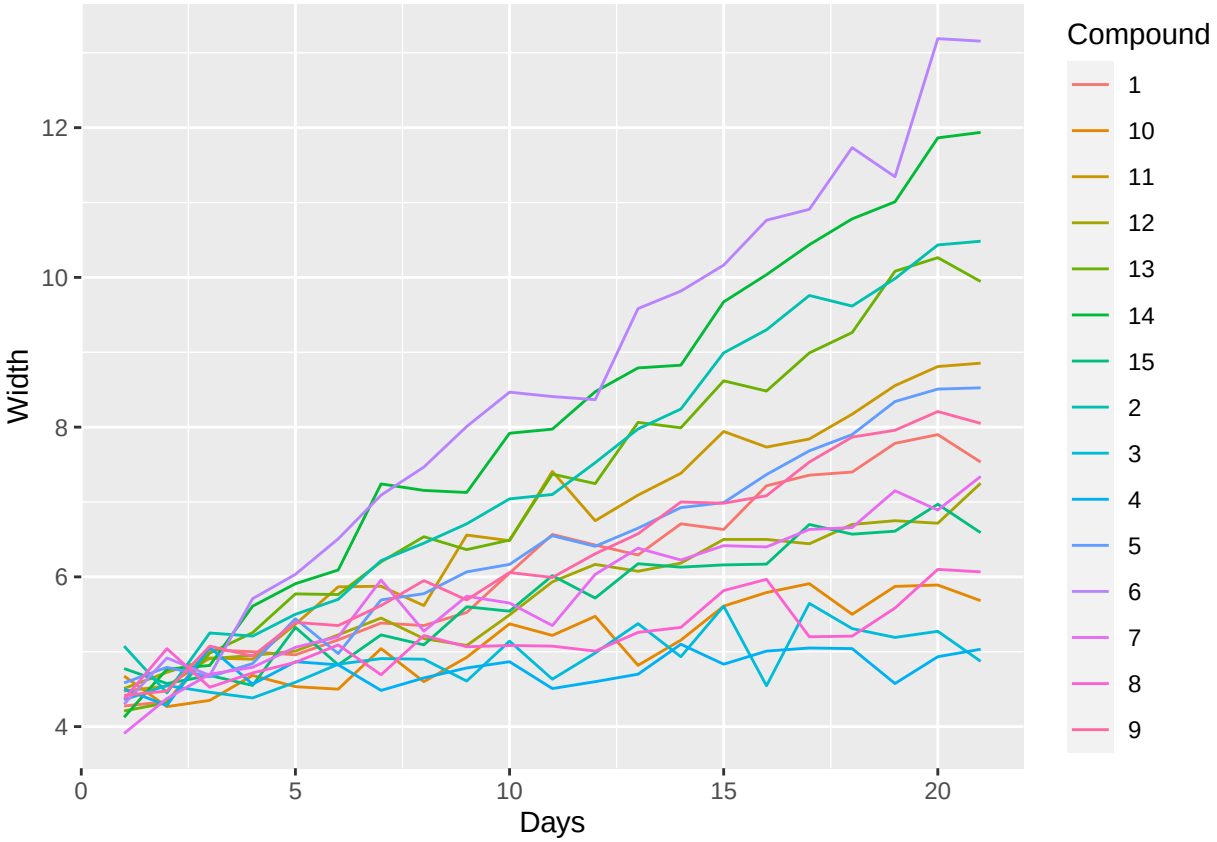
Discovering Associations

The data is loaded and manipulated

```
#read in data
path <- 'count_data_G6.csv'
d <- fread(path)
g <- fread('gaussian_data_G6.csv')
```

Below I plotted the mean width of the flower by day by compound on a given day.

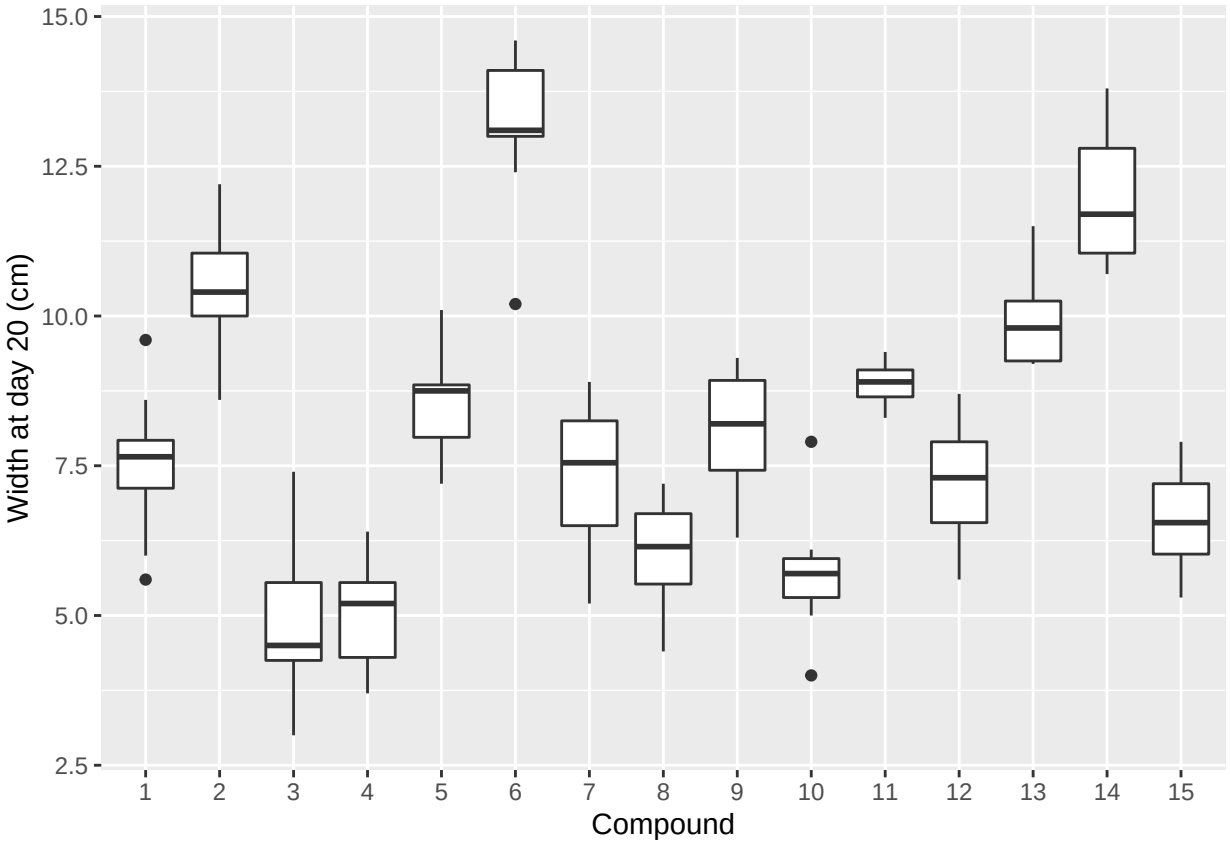
```
data_cc <- aggregate(Width ~ Compound + Days, data = dataG_combo, FUN = mean)

plot <- ggplot(data = data_cc)+
    geom_line(aes(x = Days, y = Width, color = Compound))

plot
```



A similar plot is performed, but in this case as a boxplot of the width at the day 20. This allows a bit more clearly to see the number of the compound.

```
gOld$Compound=as.factor(gOld$Compound)
ggplot(data=gOld, aes(x=Compound,y=T_20))+
  geom_boxplot()+ylab("Width at day 20 (cm)")
```

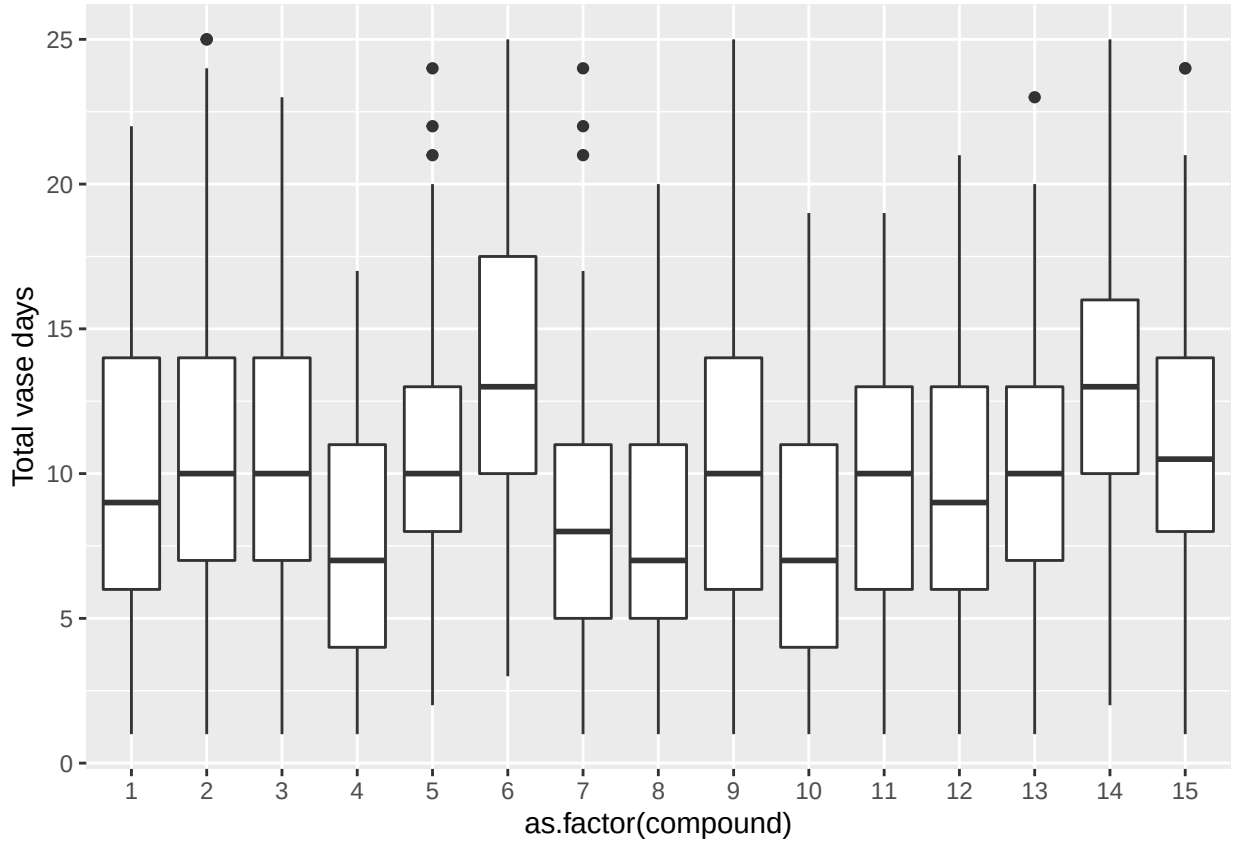
```
## Warning: Removed 10 rows containing non-finite values (stat_boxplot).
```



To compare the gaussian data with the Poisson data, we plot a similar boxplot for the total vase days.

```
ggplot(data=d, aes(x=as.factor(compound),y=tot.vase.days))+
  geom_boxplot()+ylab("Total vase days")
```

```
## Warning: Removed 60 rows containing non-finite values (stat_boxplot).
```



A combined table is generated with the mean values of the two previous boxplots.

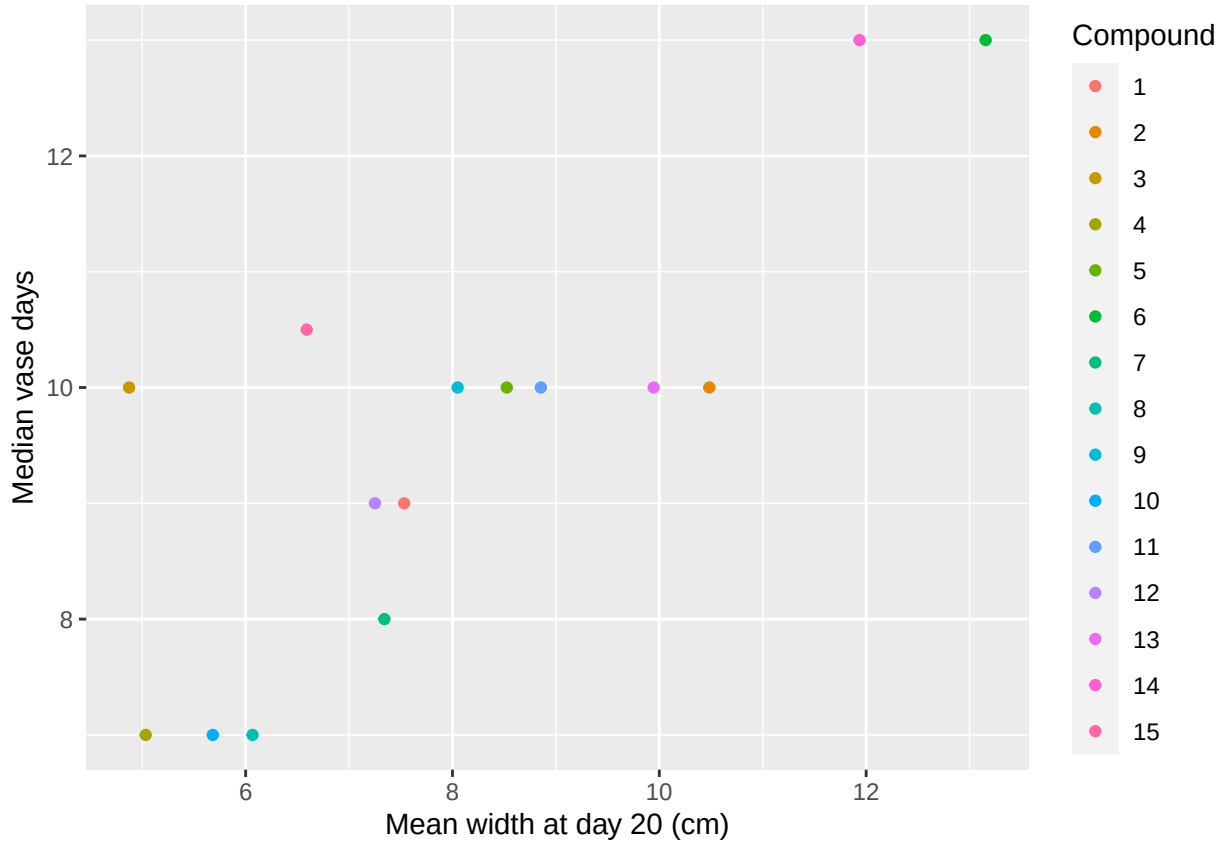
```
g.data.table=as.data.table(gOld)
g.data.table$Compound=as.factor(g.data.table$Compound)
g.data.table=na.omit(g.data.table, cols="T_20")
summary <- g.data.table[,.(MeanWidth=mean(T_20)),.(Compound)]
d.data.table=as.data.table(d)
d.data.table=na.omit(d.data.table, cols="tot.vase.days")
d.data.table$Compound=as.factor(d.data.table$compound)
summary2 <- d.data.table[,.(MedianVaseDays=median(tot.vase.days)),.(Compound)]
summary <- summary[summary2, on = .(Compound)]
summary
```

```
##     Compound MeanWidth MedianVaseDays
##  1:        1  7.533333            9.0
##  2:        2 10.483333           10.0
##  3:        3  4.872727           10.0
##  4:        4  5.033333            7.0
##  5:        5  8.525000           10.0
##  6:        6 13.155556           13.0
##  7:        8  6.066667            7.0
##  8:        9  8.050000           10.0
##  9:       10  5.681818            7.0
## 10:       11  8.854545           10.0
## 11:       12  7.250000            9.0
## 12:       13  9.945455           10.0
```

```
## 13:       14 11.936364           13.0
## 14:       15  6.590000           10.5
## 15:        7  7.341667            8.0
```

And then plotted

```
ggplot(data = summary,aes(x = MeanWidth, y = MedianVaseDays,color = Compound))+
    geom_point()+xlab("Mean width at day 20 (cm)")+ylab("Median vase days")
```



A mixed model is fitted

```
d$bushID<-as.factor(d$bushID)
d$compound<-as.factor(d$compound)
d$garden<-as.factor(d$garden)
d$species<-as.factor(d$species)
d$rater<-as.factor(d$rater)
m1 <-glmer(tot.vase.days~compound+species+garden+(1|rater)+(1|bushID)+(1|subplotID), family=poisson(lin
summary(m1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula: tot.vase.days ~ compound + species + garden + (1 | rater) + (1 |
##     bushID) + (1 | subplotID)
##    Data: d
```

```
##
##      AIC      BIC   logLik deviance df.resid
##   7222.6   7327.2  -3591.3   7182.6     1360
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.7555 -0.7878 -0.0459  0.6249  3.7431
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  bushID    (Intercept) 0.01129  0.1062
##  subplotID (Intercept) 0.05483  0.2342
##  rater     (Intercept) 0.12488  0.3534
## Number of obs: 1380, groups:  bushID, 96; subplotID, 16; rater, 6
##
## Fixed effects:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.135101   0.171288  12.465  < 2e-16 ***
## compound2    0.070574   0.046365   1.522 0.127975
## compound3    0.063235   0.046126   1.371 0.170398
## compound4   -0.272911   0.050428  -5.412 6.24e-08 ***
## compound5    0.062465   0.046940   1.331 0.183279
## compound6    0.346820   0.043616   7.952 1.84e-15 ***
## compound7   -0.167143   0.048947  -3.415 0.000638 ***
## compound8   -0.248028   0.050442  -4.917 8.78e-07 ***
## compound9    0.042653   0.046612   0.915 0.360164
## compound10  -0.223922   0.049918  -4.486 7.26e-06 ***
## compound11   0.003953   0.047305   0.084 0.933410
## compound12  -0.055446   0.047967  -1.156 0.247713
## compound13   0.046951   0.046540   1.009 0.313063
## compound14   0.292411   0.044307   6.600 4.12e-11 ***
## compound15   0.082048   0.045998   1.784 0.074466 .
## species2    -0.033620   0.028069  -1.198 0.231009
## garden2      0.186699   0.120430   1.550 0.121077
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Correlation matrix not shown by default, as p = 17 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it
```

And we plot the compund effect based on the model and compare it with the width data.

```
gOld2=as.data.table(gOld)
gOld2=na.omit(gOld2, cols="T_20")
gOld2=gOld2[,]
gSummary <- gOld2[,.(Mean=mean(T_20)),.(Compound)]
coefficients=summary(m1)$coefficients
coefficients=coefficients[c(1:15)]
coefficients=coefficients+coefficients[1]
coefficients[1]=coefficients[1]/2
gSummary$betas=coefficients
```

```
gSummary$Compound=as.factor(gSummary$Compound)
ggplot(data = gSummary,aes(x = Mean, y = betas,color = Compound))+
    geom_point()+xlab("Mean width at day 20 (cm)")+ylab("Compound effect based on the mixed model")
```