For Botanist

Stephen Brownsey

03/11/2019

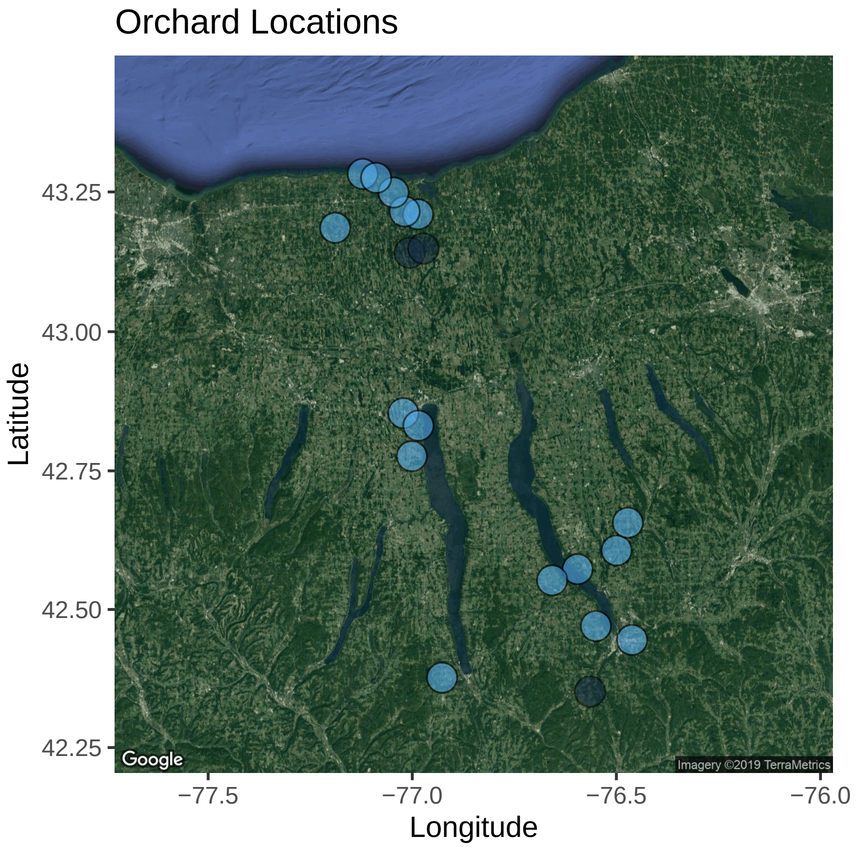
Introduction

Questions on Bees in Orchards study. But in particular, I'd like to know your opinion on the subsetting of the data for the purpose of the decision models, in particularly if there were any "obvious" ways to model it to start with. The data itself, doesn't seem to be great in my opinion for model building as it's quite hard to see how we can extract relationships that are that meaningful out of for the following reasons: The values of all the pesticides e.t.c are the same for each orchard at each visit occasion (All the values are the same each day for each year, which I find hard to believe) and obviously the bee abundance and richness levels vary alot depending on the visit but all the values are the same for these parameters (I ran best subsets regression on the the dataset and these values were the most significant which I also thought was odd as I was expecting bloom and temperature to have a significant affect as these are the only ones which change for each visit). Am I also right in thinking: richness is diversity of bee species and abundance is number of bees - in which case would you look at them combined or separately or just 1 of them when building the models (I can model all scenarios but wondered what atleast would be your recommendation of starting point.) I have decided to only take the data from 2012 but in terms of further subsetting your opinion would be appreciated. In the data we have the pesticide levels, but these are never 0. So in the specification that Julia sent me done by the botanist it mentioned something along the lines of *Three isolated time periods which I would be interested in: Before Bloom, During Bloom and After Bloom. Before bloom decision points: apply fungicide, apply insecticide, apply both and apply neither. During bloom decision points: apply fungicide, apply insecticide, apply thinner, apply two of the three, apply all, apply nothing. After bloom decision points: apply insecticide, apply thinner, apply both, apple nothing. These can be visualised by the diagram below.* But from the data we have I don't quite see how I could follow this and extract this data to use for the purpose of modelling and wander the next best way to approach it. Perhaps a very simplified approach with the actual data and then a theoretically application if I had data on it? Not really sure how to approach the modellign side of things if I'm perfectly honest.

Map Analysis

This is an image of where each of the orchards is located, more for interest but would you think that they would go around taking readings of different orchards based on similar days when they are close to each other? If so, would it make sense to apply some models to just one area?

```
## Source : https://maps.googleapis.com/maps/api/staticmap?center=42.85269,-76.850661&zoom=9&size=640x6
```



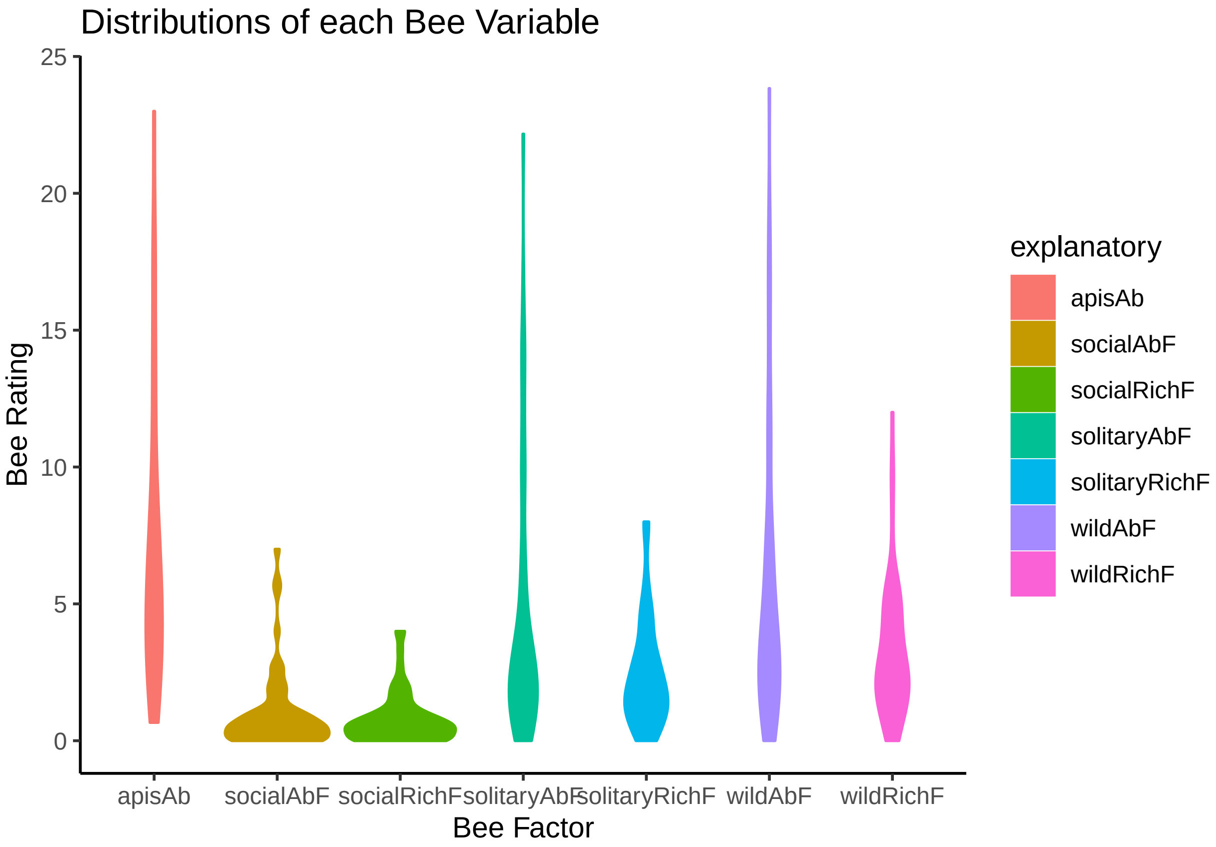
The variables in general are very correlated, as you would expect, but this means that for the decision models I would want to be looking at a subset of these but from a botanist's point of view which variables would she consider best? Not that clear in a .pdf but lots are upwards of 0.8/0.9 correlated.

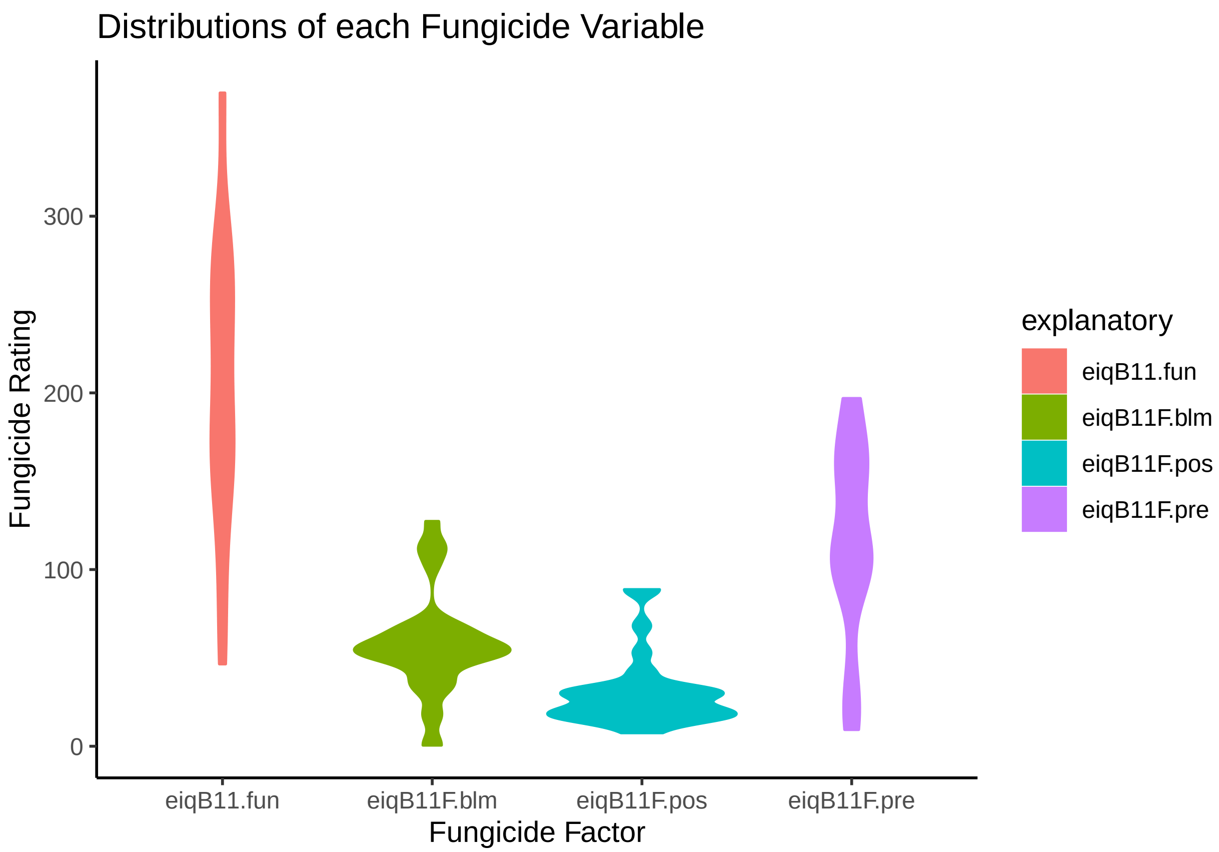
GGTally

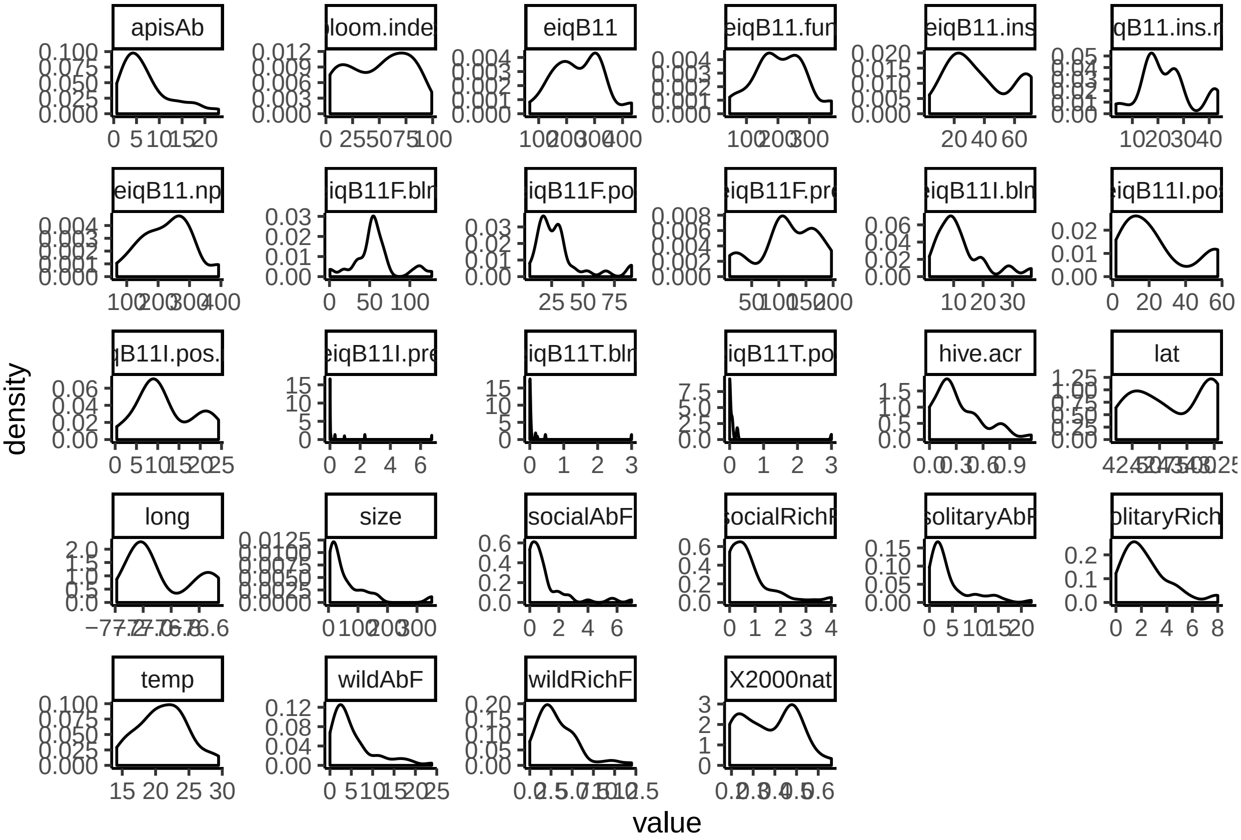
Looking at individual variables

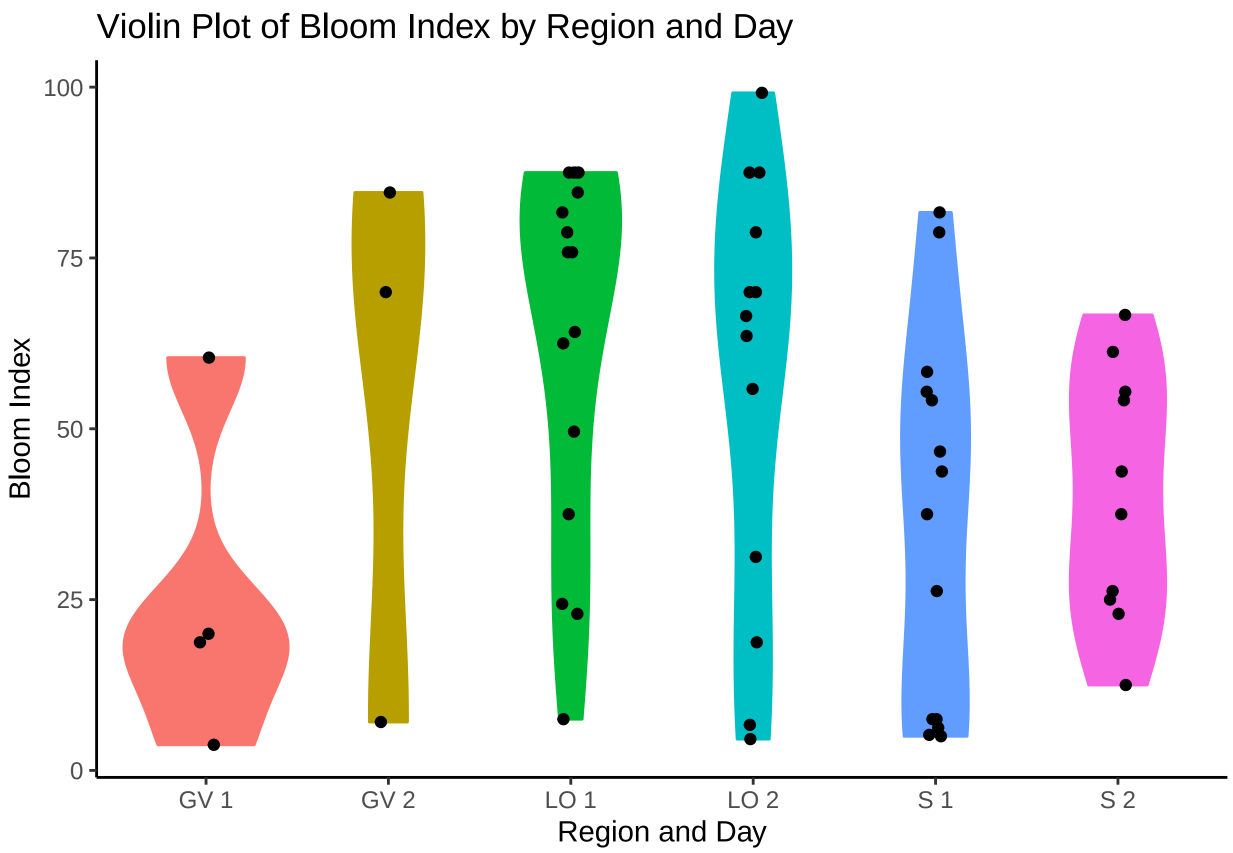
Code post 25/10

Violin plots of the data - I am a fan of what these show, but in particular note the one for bloom.index (at the bottom), I had thought that day 1 would be < 50% bloom and day 2 post 50% bloom but this clearly shows this is not the case - thoughts? Have also included a distributions plot of the variables in the dataset for interest as it could be helpful when thinking about these things.


Distributions of each Bee Variable
Bee Rating
25
20
15
10
5
0
apisAb
socialAbF
socialRichF
solitaryAbF
solitaryRichF
wildAbF
wildRichF
Bee Factor
explanatory
apisAb
socialAbF
socialRichF
solitaryAbF
solitaryRichF
wildAbF
wildRichF

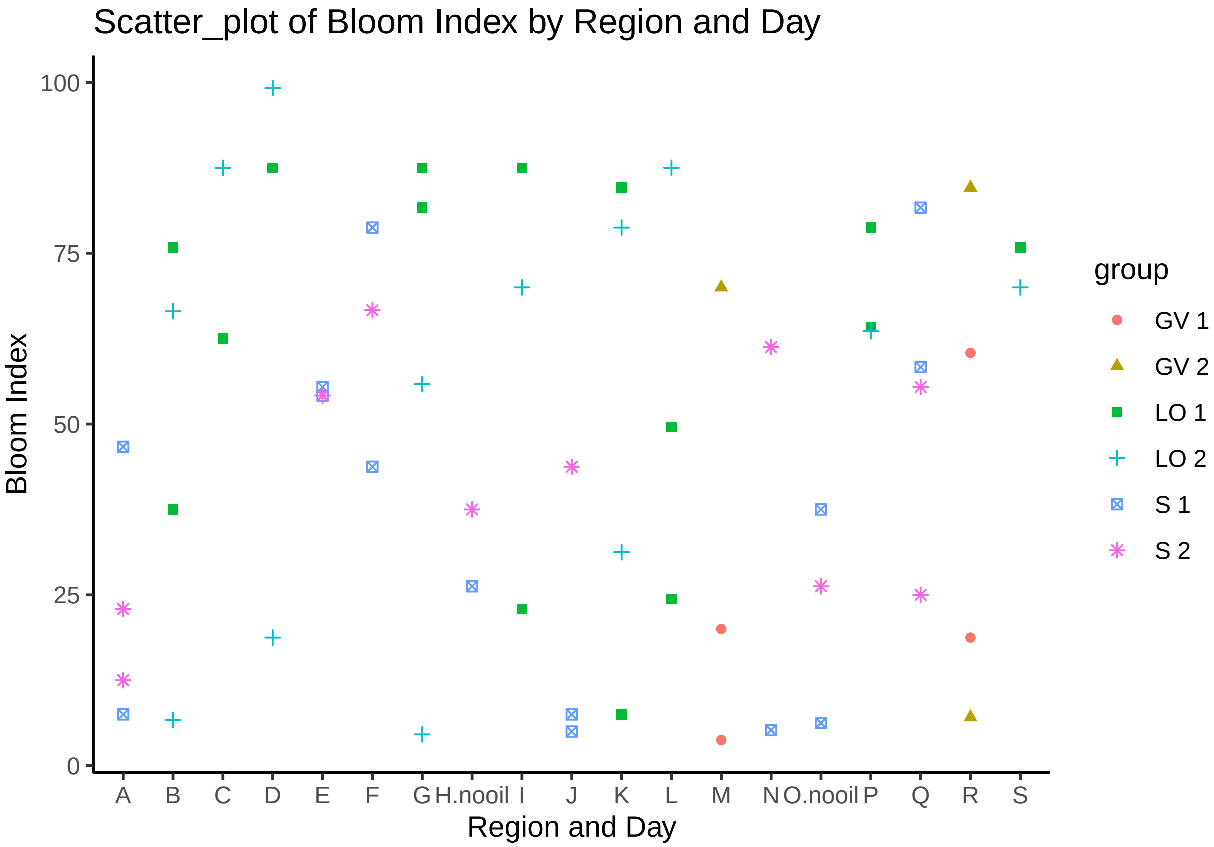

Distributions of each Fungicide Variable
Fungicide Rating
300
200
100
0
eiqB11.fun
eiqB11F.blm
eiqB11F.pos
eiqB11F.pre
Fungicide Factor
explanatory
eiqB11.fun
eiqB11F.blm
eiqB11F.pos
eiqB11F.pre


apisAb
loom.inde
eiqB11
eiqB11.fur
eiqB11.ins
qB11.ins.r
eiqB11.np
iqB11F.blr
iqB11F.po
eiqB11F.pr
eiqB11I.blr
eiqB11I.po
qB11I.pos.
eiqB11I.pr
iqB11T.blr
iqB11T.po
hive.acr
lat
long
size
socialAbF
ocialRichF
solitaryAbF
olitaryRich
temp
wildAbF
wildRichF
X2000nat
density
value

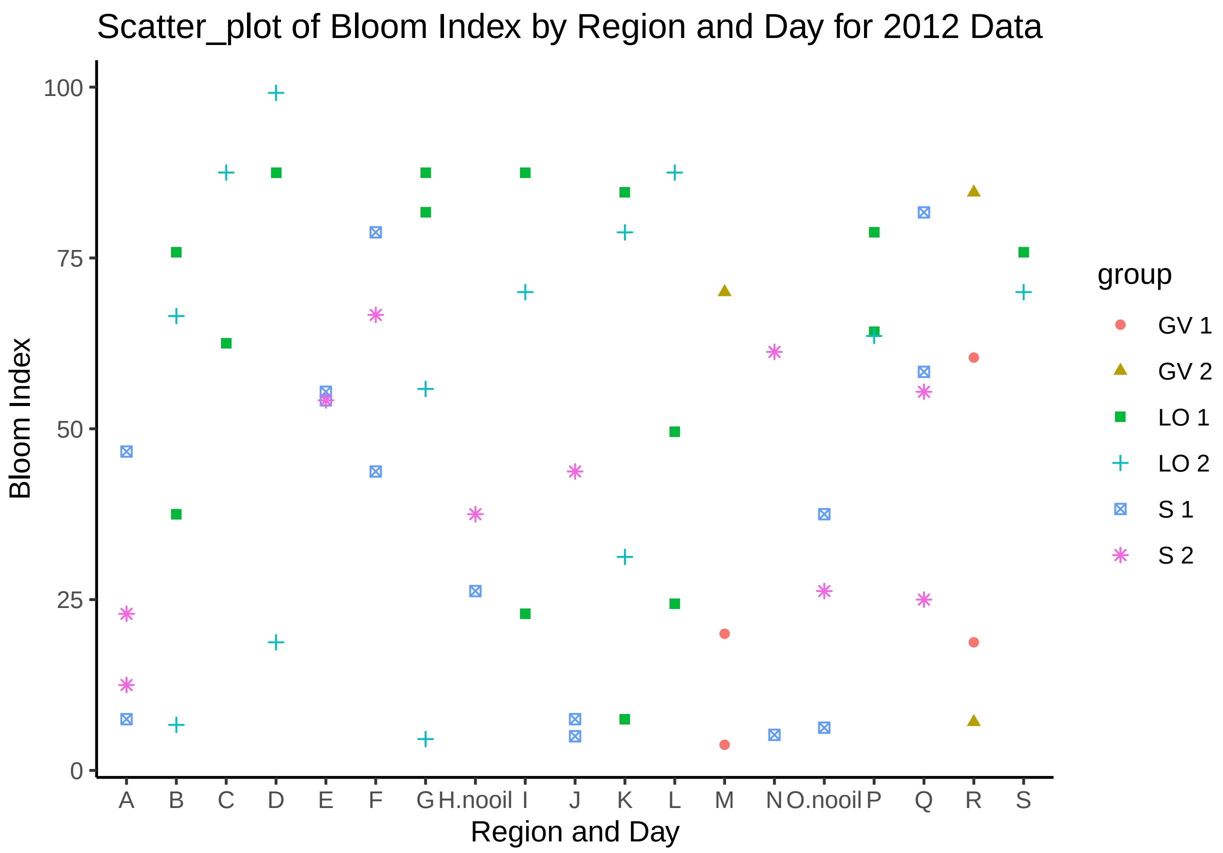


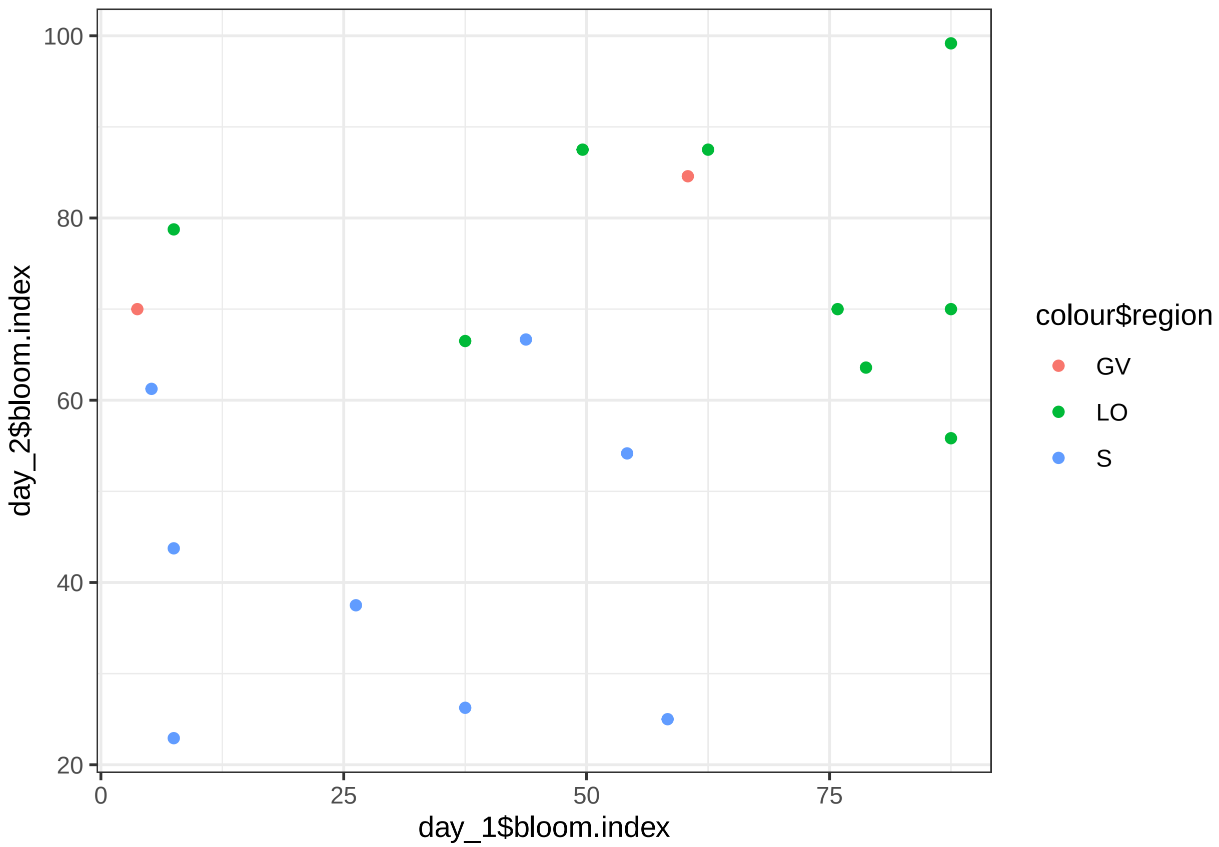
Graphics from botanists chat

Another thing: Instead of the last figure with violin plots about bloom for region/day, could you do a scatter plot visualising the relationship between day 1 and day 2 bloom index? y=day 1 bloom index versus x=day 2 index You can put all the all the three regions into the same scatter plot, using a different colour/plotting characters

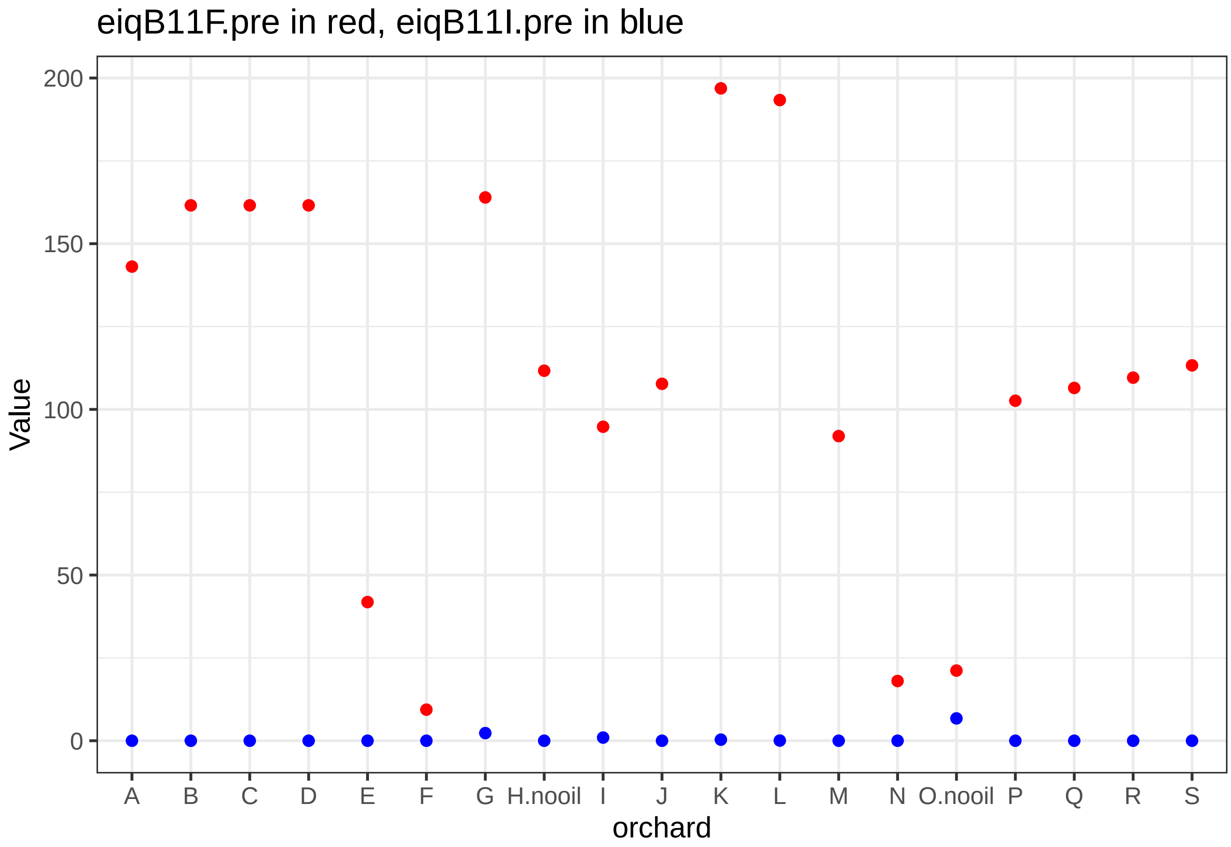


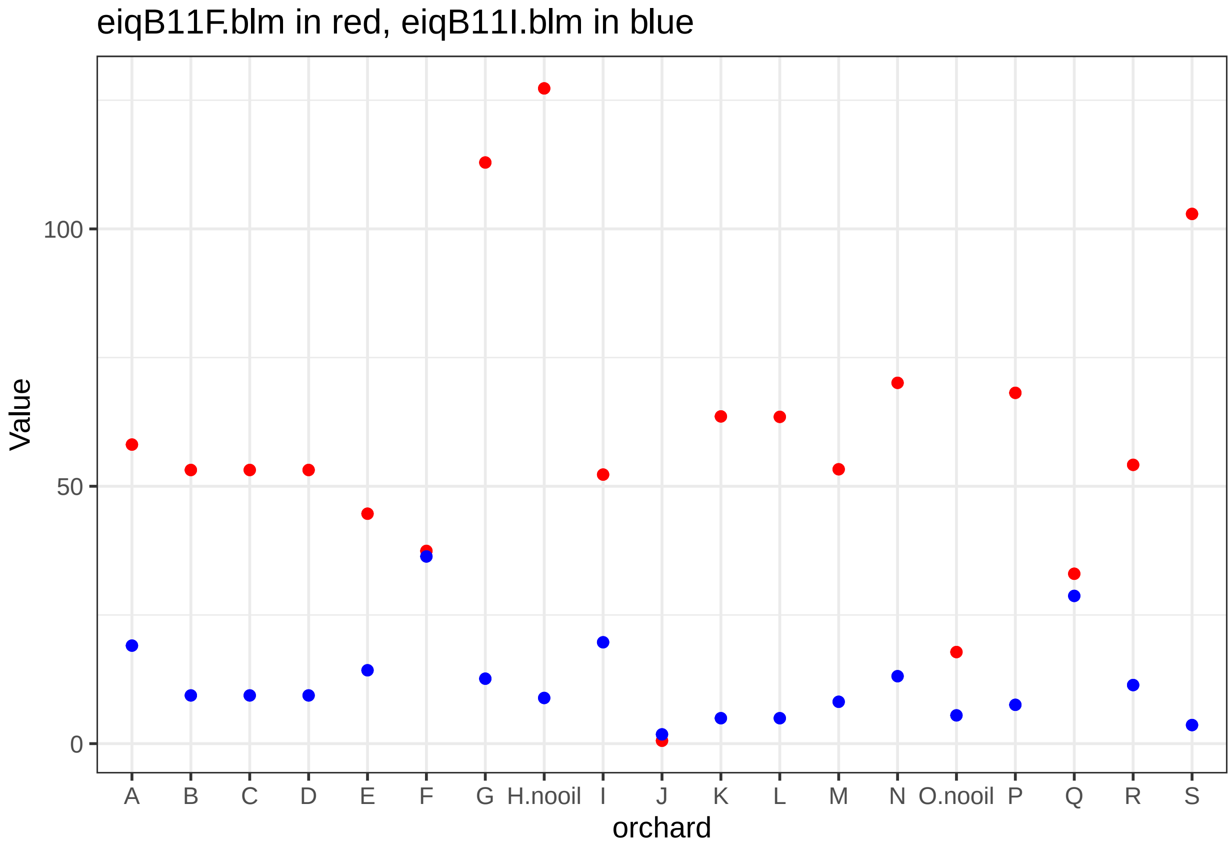
for each region.

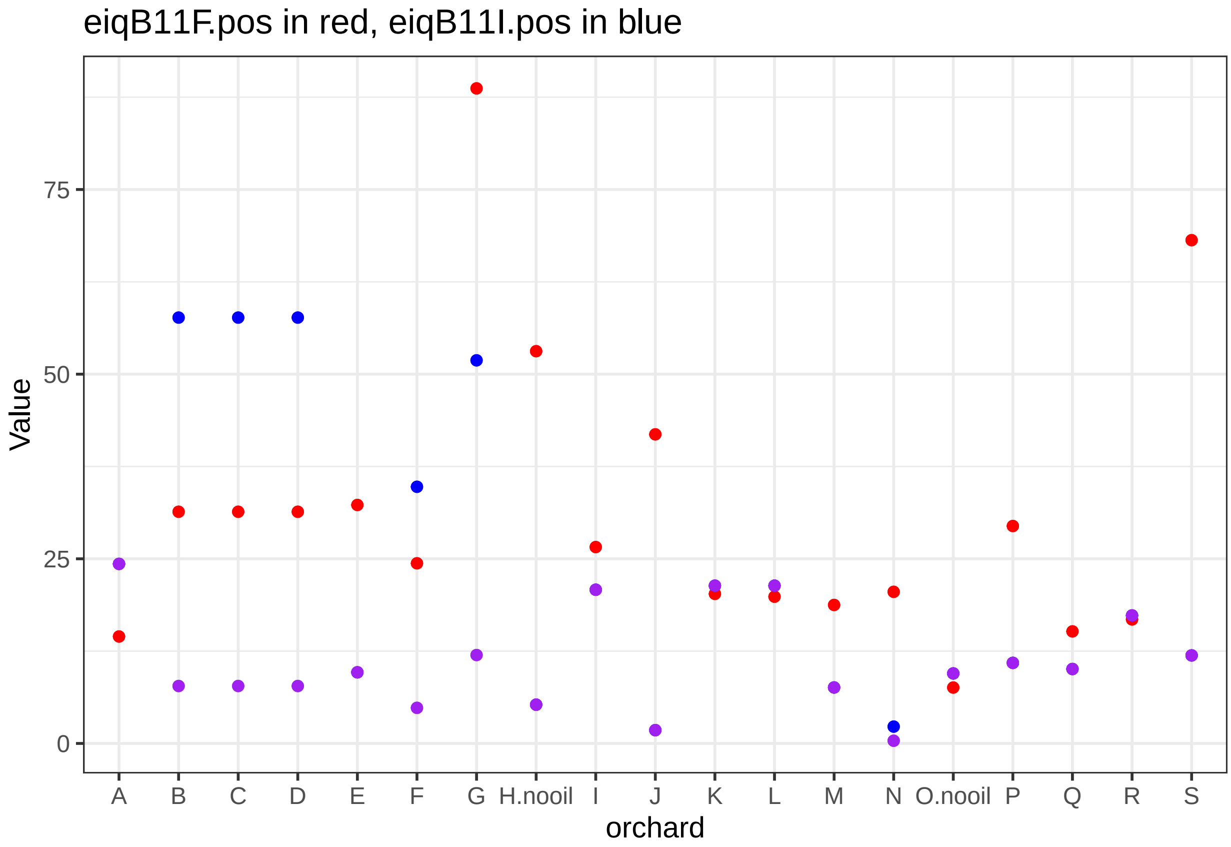

Scatter_plot of Bloom Index by Region and Day for 2012 Data
Bloom Index
100
75
50
25
0
A
B
C
D
E
F
G
H.nooil
I
J
K
L
M
N
O.nooil
P
Q
R
S
Region and Day
group
GV 1
GV 2
LO 1
LO 2
S 1
S 2



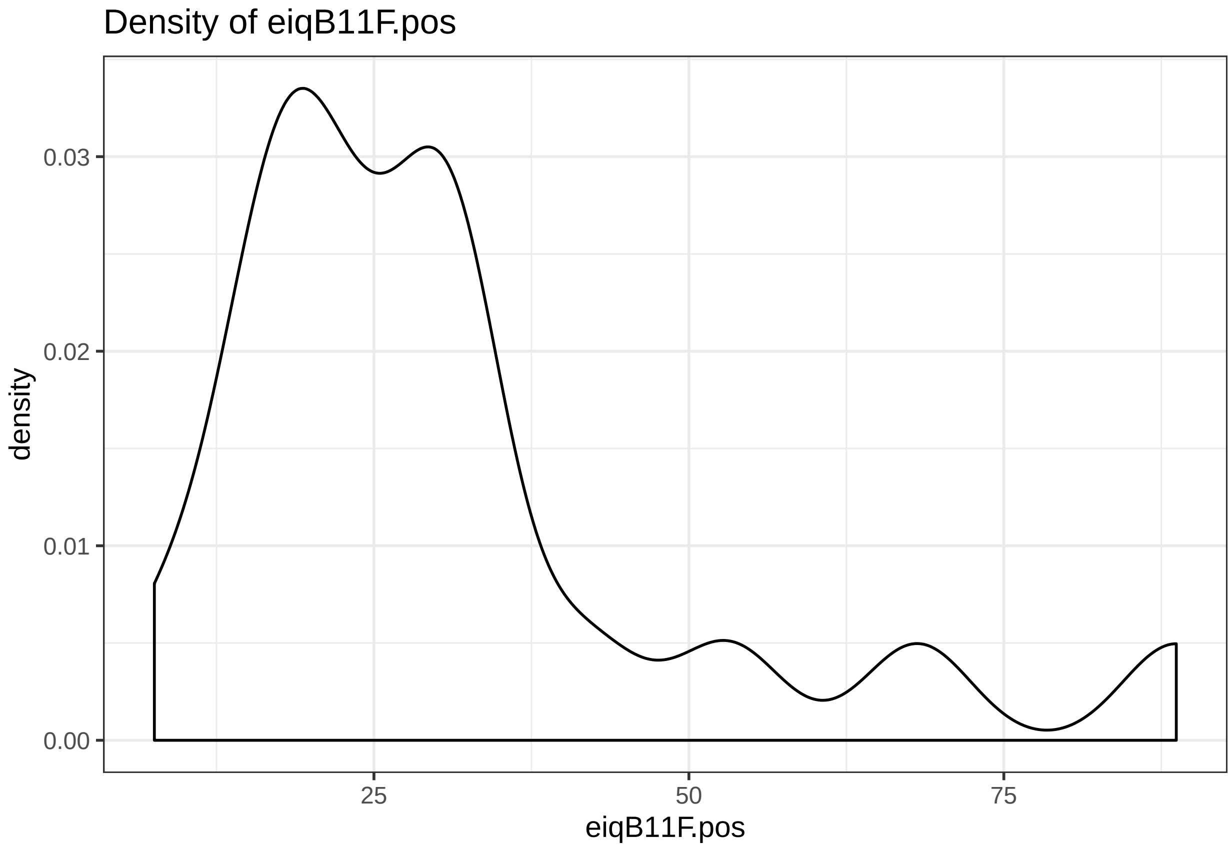
The first step for us is to understand the chemicals applications better, including the combinations. The variables in the data are actually scores developed by the authors that capture the impact the have, so that is quite ready to use and comparable. (See Park et al for details.)


eiqB11F.pre in red, eiqB11I.pre in blue
Value
200
150
100
50
0
A
B
C
D
E
F
G
H.nooil
I
J
K
L
M
N
O.nooil
P
Q
R
S
orchard

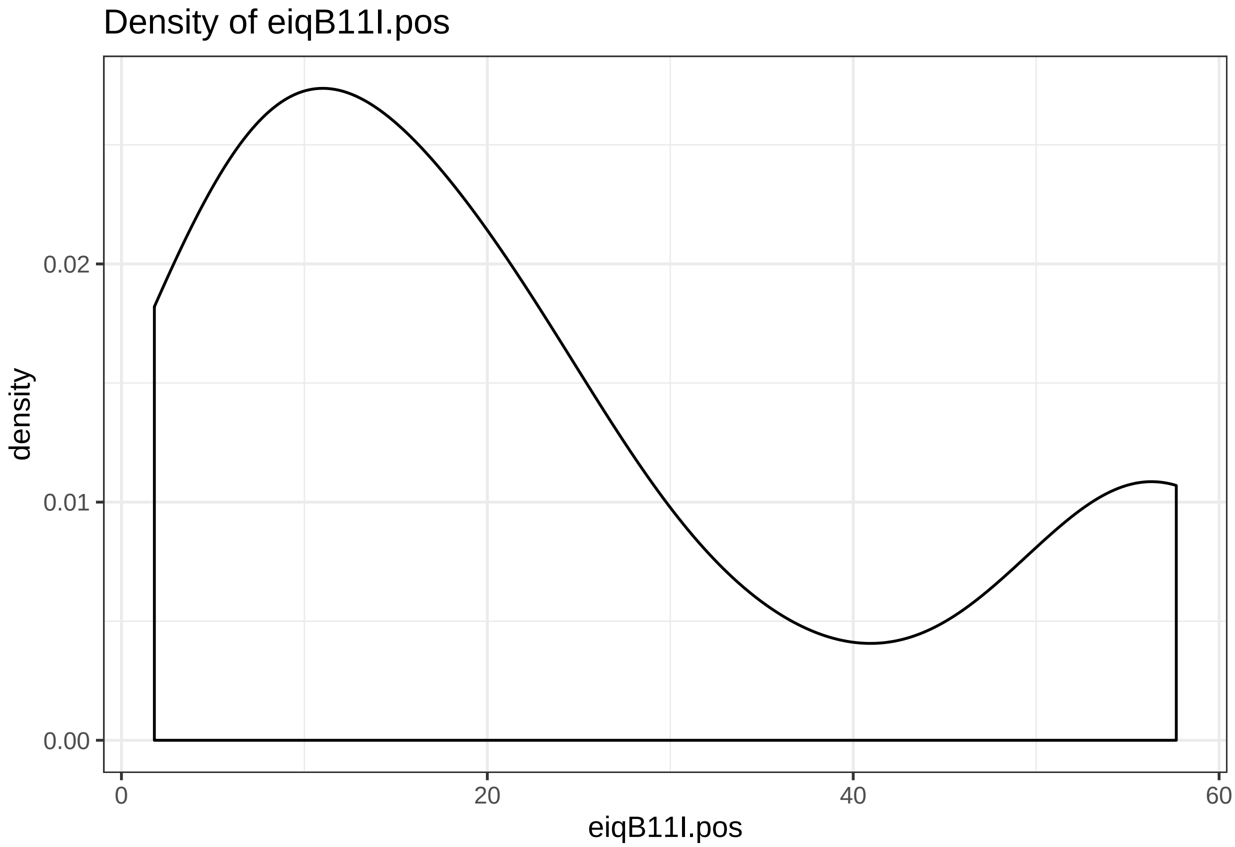

eiqB11F.blm in red, eiqB11I.blm in blue
Value
100
50
0
A
B
C
D
E
F
G
H.nooil
I
J
K
L
M
N
O.nooil
P
Q
R
S
orchard


eiqB11F.pos in red, eiqB11I.pos in blue
Value
0
25
50
75
orchard
A
B
C
D
E
F
G
H.nooil
I
J
K
L
M
N
O.nooil
P
Q
R
S

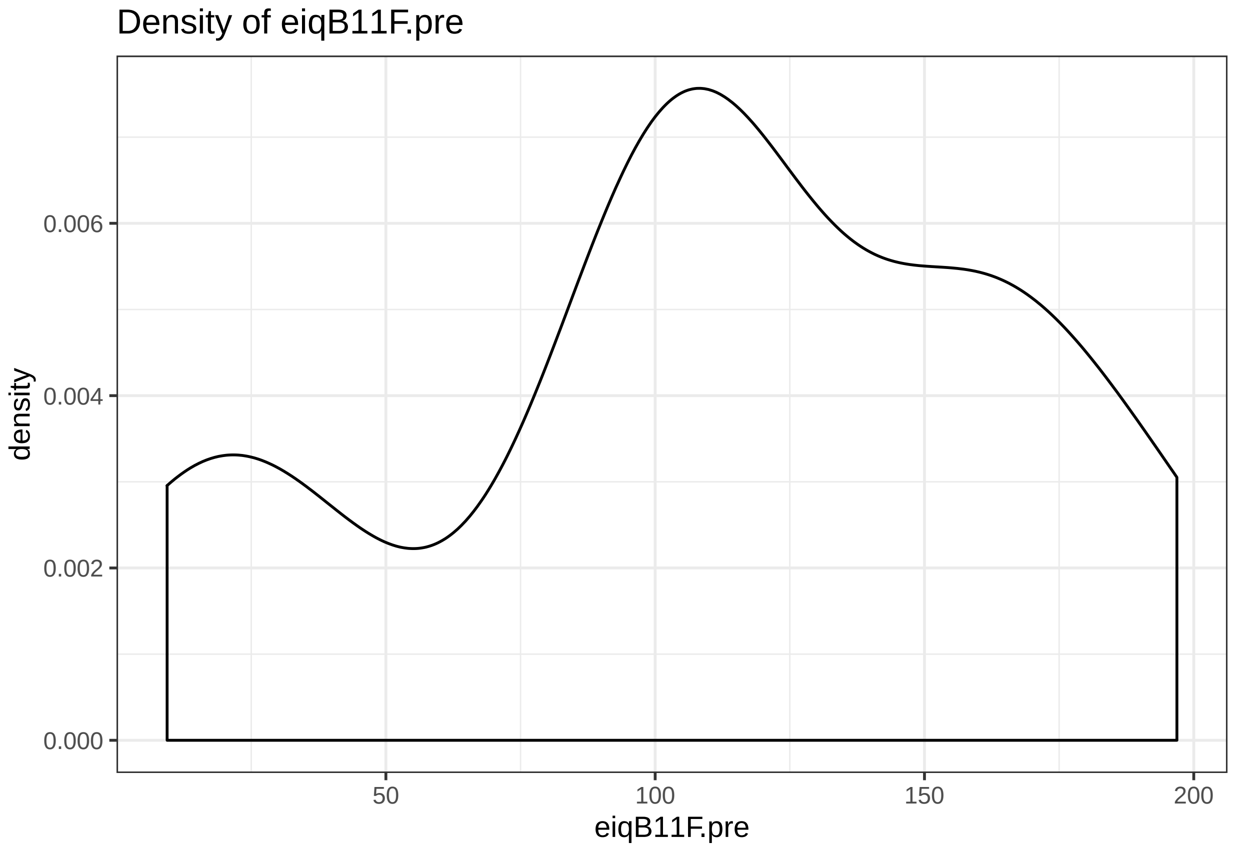
```
## [[1]]
```


Density of eiqB11F.pos
density
0.03
0.02
0.01
0.00
25
50
75
eiqB11F.pos

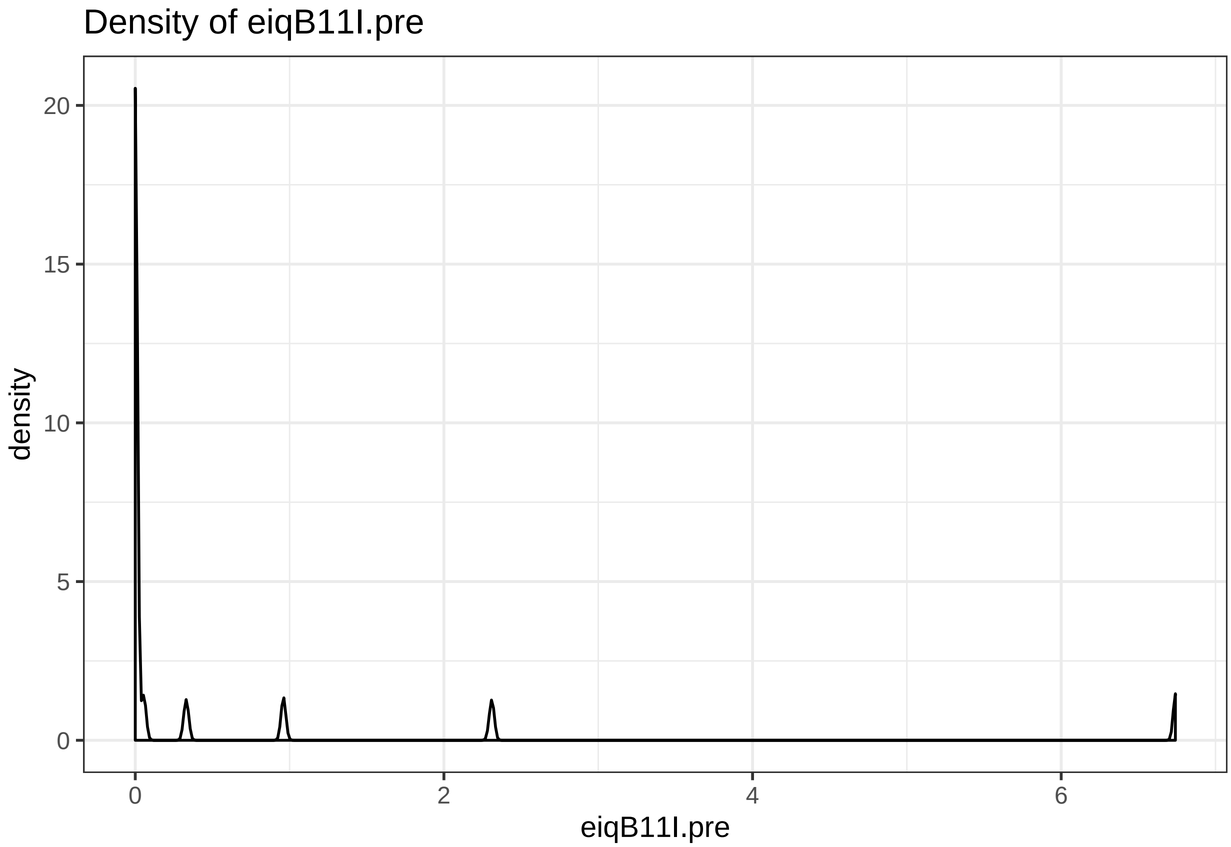
```
##
## [[2]]
```


Density of eiqB11I.pos
density
0.00
0.01
0.02
0
20
40
60
eiqB11I.pos

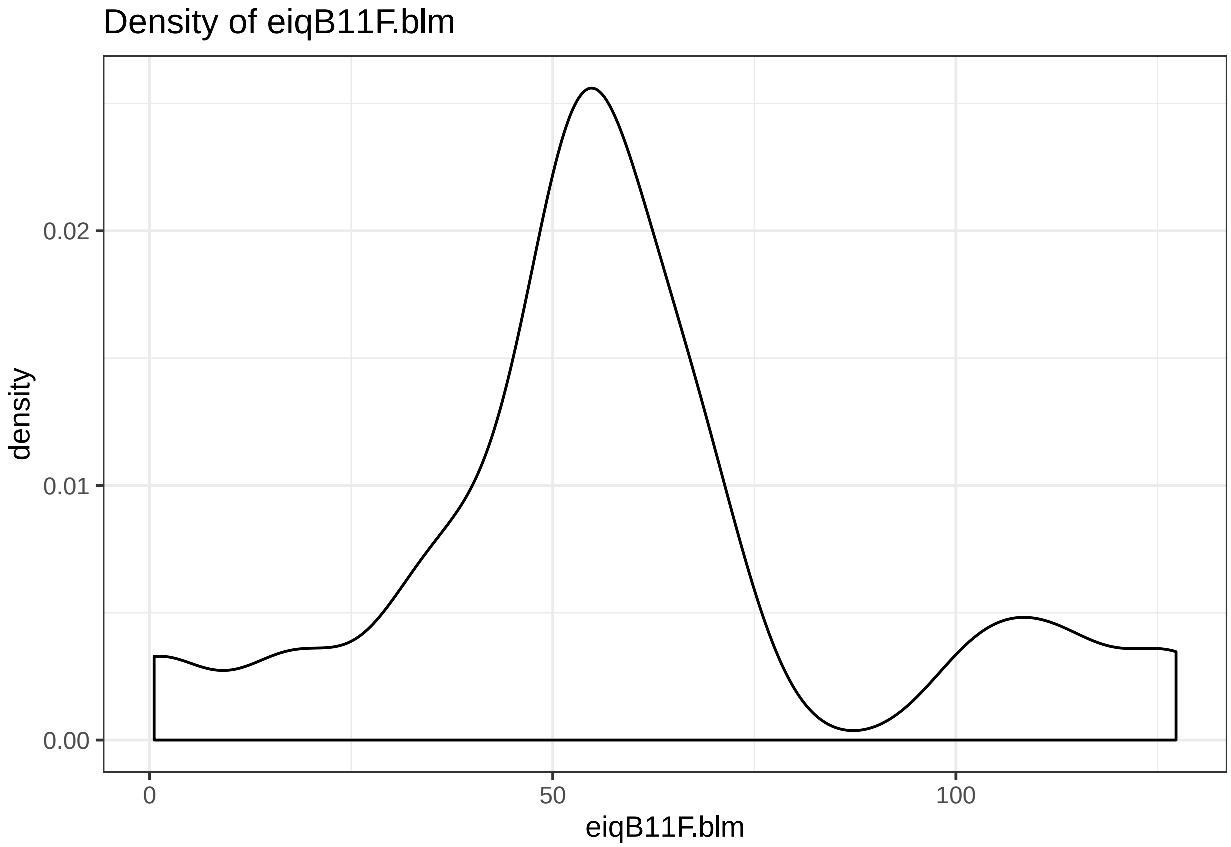
```
##
## [[3]]
```


Density of eiqB11F.pre
density
0.006
0.004
0.002
0.000
50
100
150
200
eiqB11F.pre

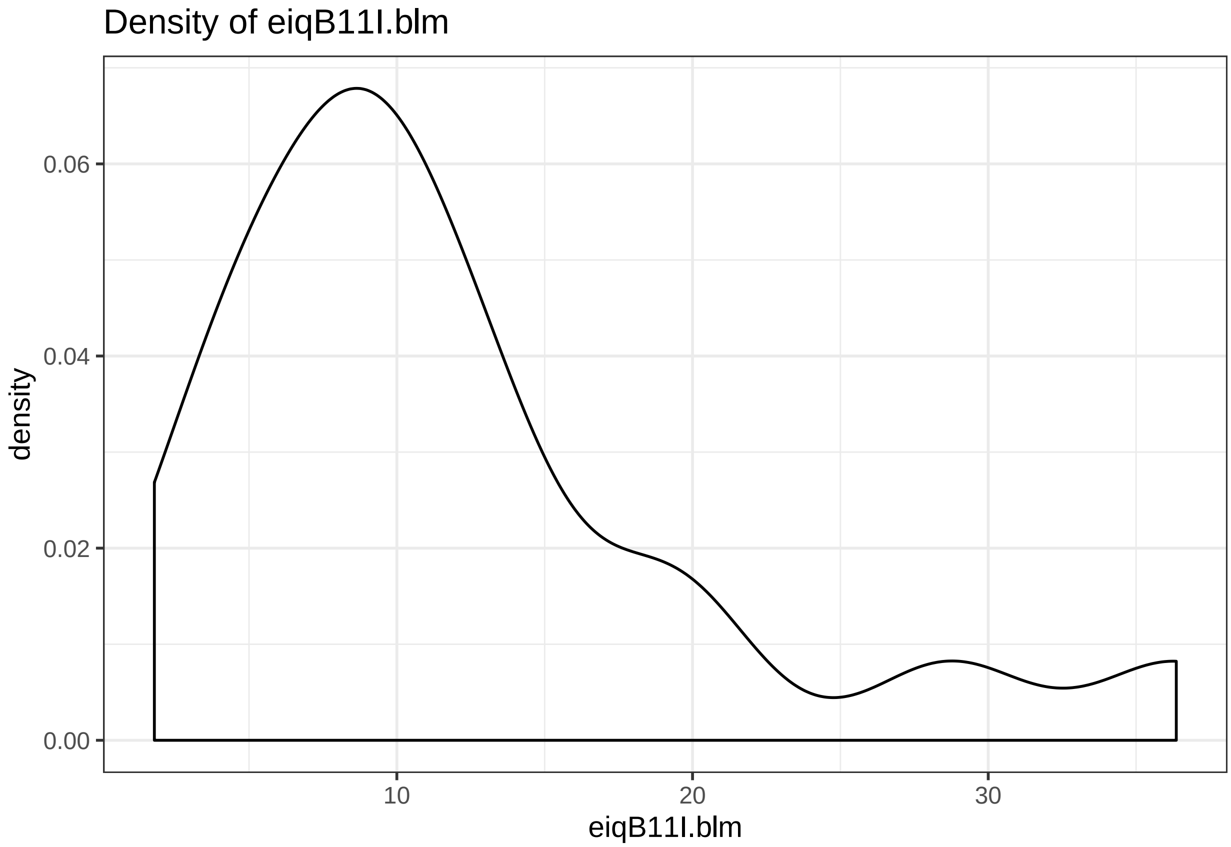
```
##
## [[4]]
```



```
##
## [[5]]
```

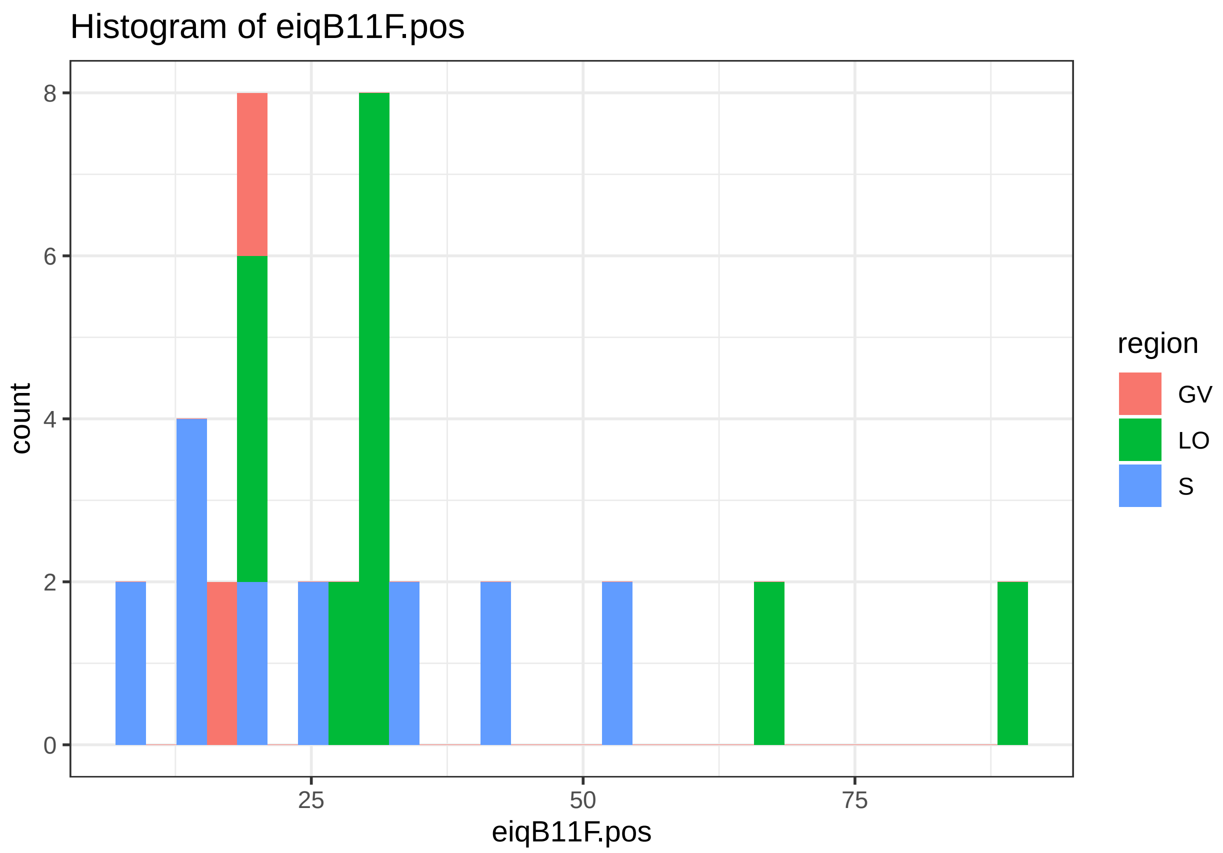

Density of eiqB11F.blm
density
0.02
0.01
0.00
0
50
100
eiqB11F.blm

```
##
## [[6]]
```



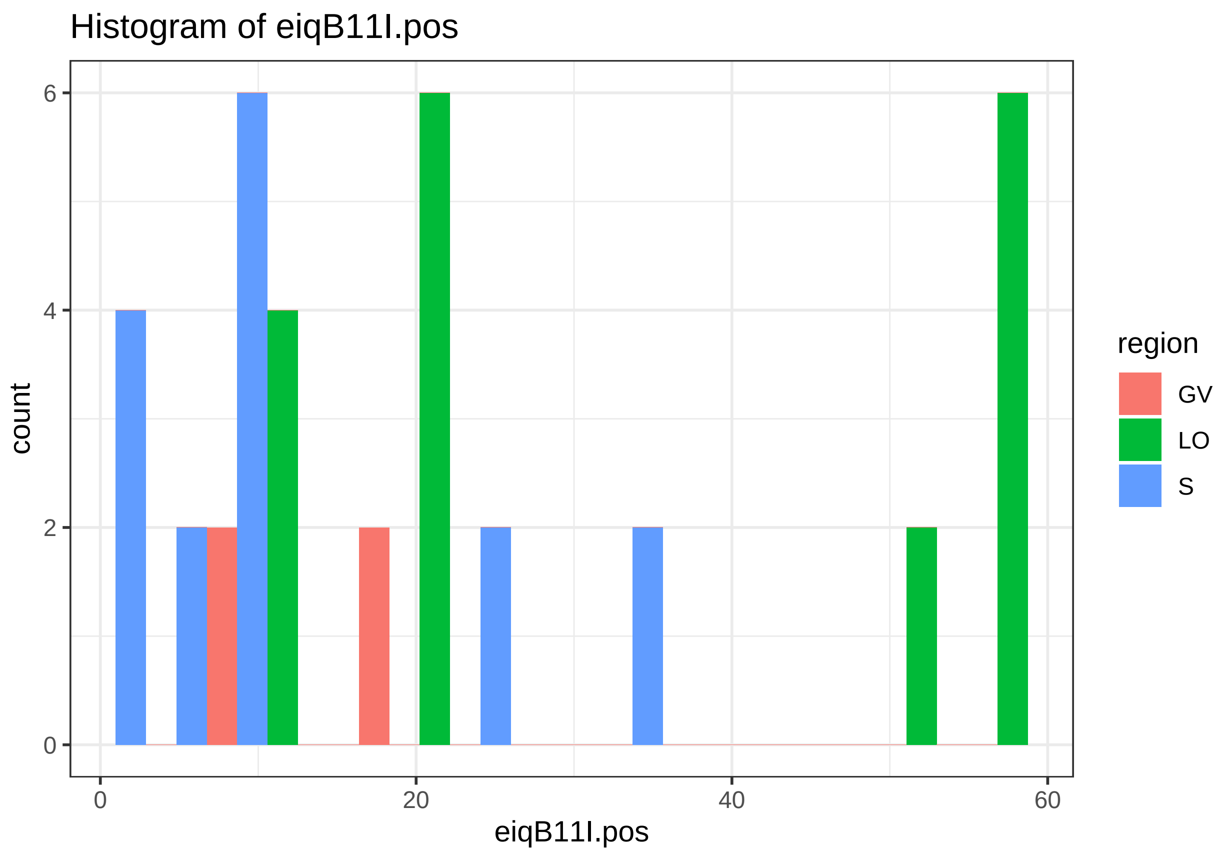
```
## [[1]]

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



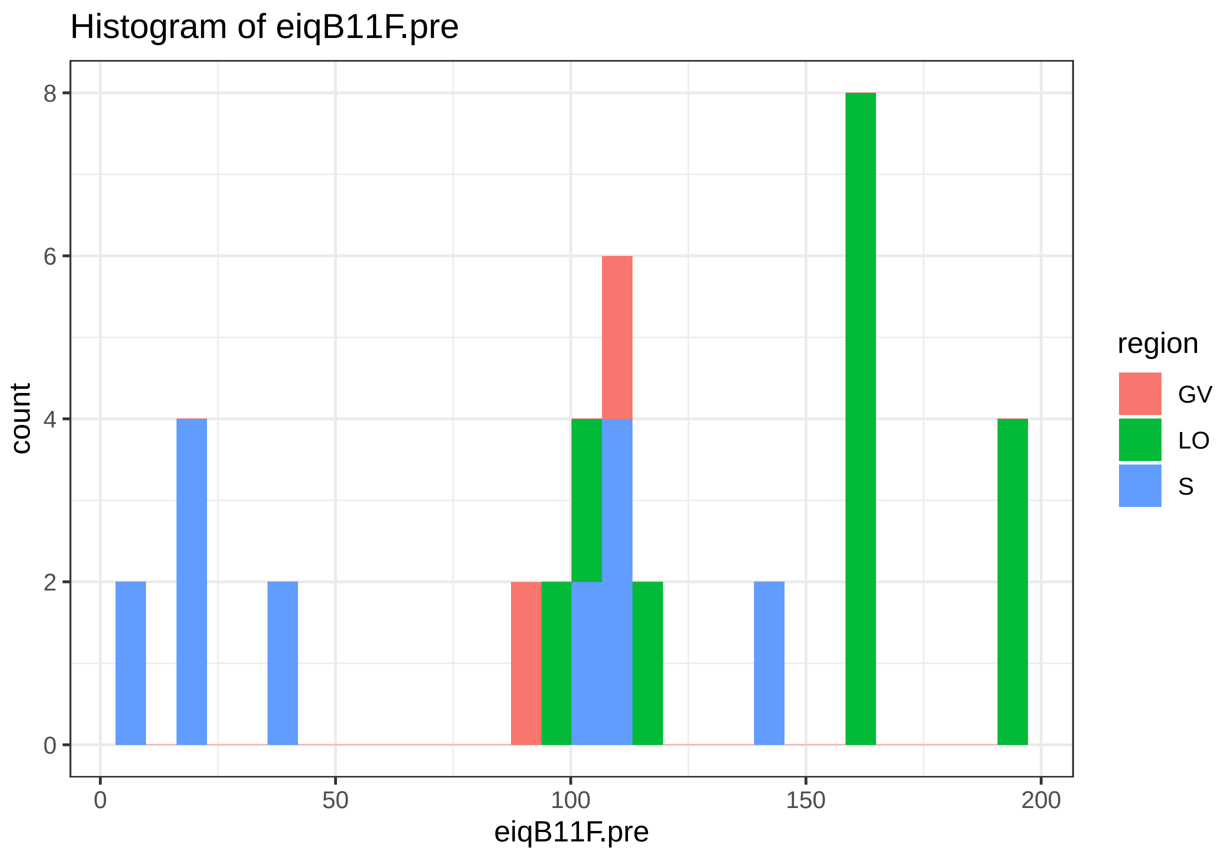
```
##
## [[2]]
```

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



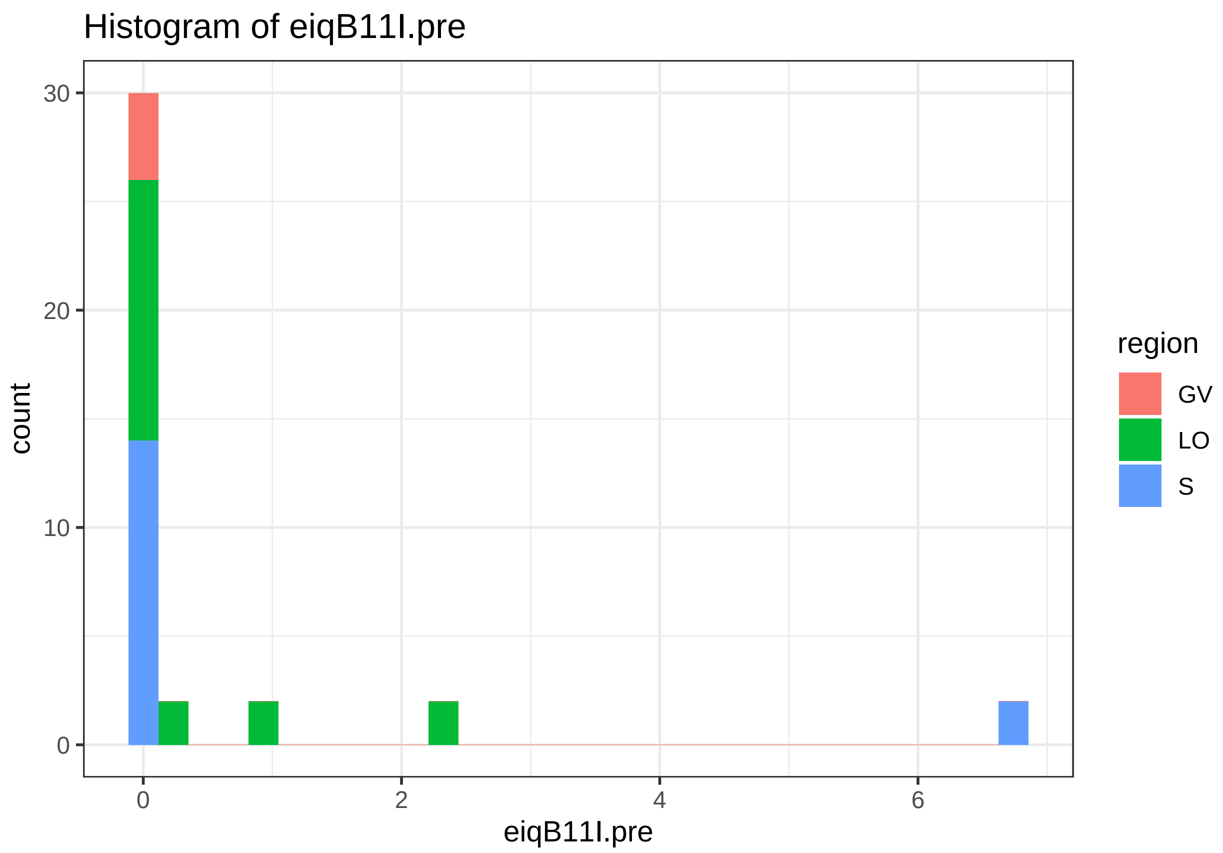
```
##
## [[3]]
```

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



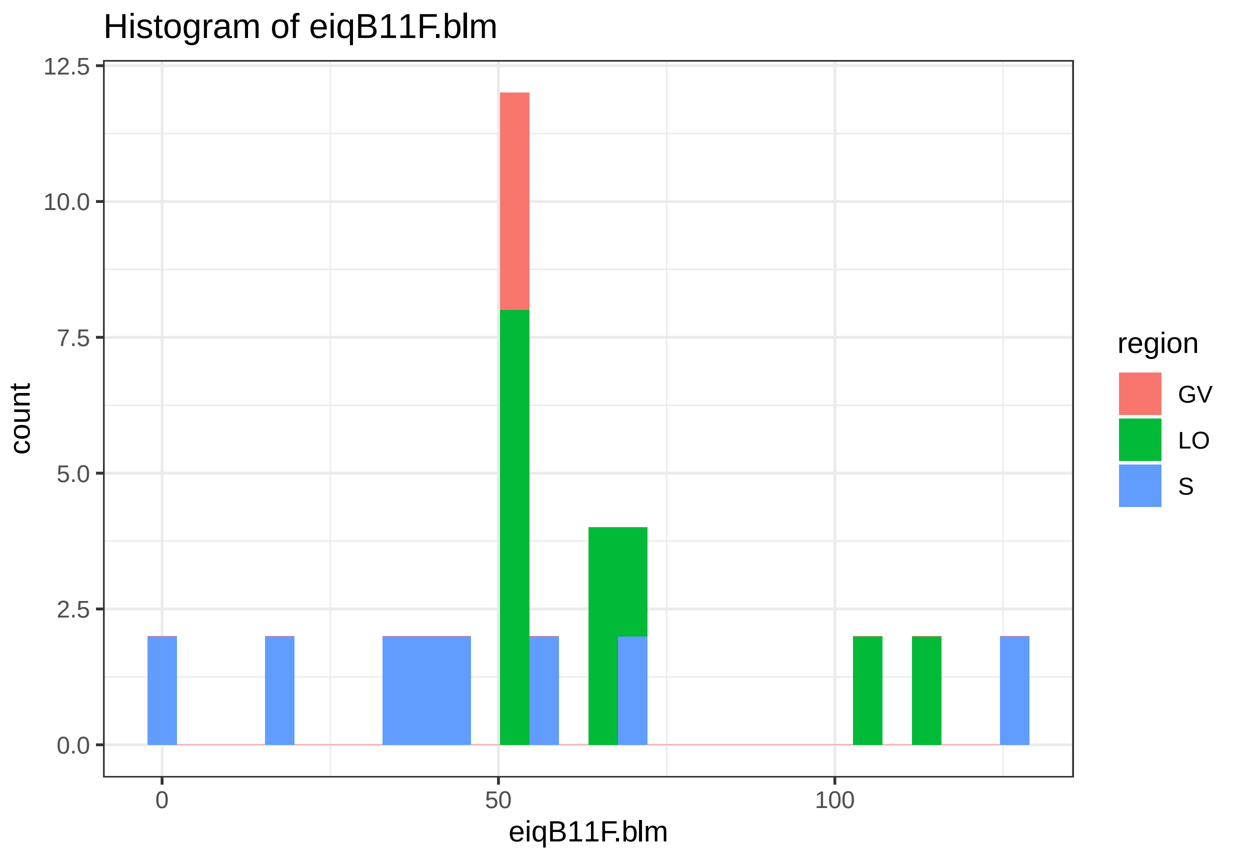
```
##
## [[4]]
```

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



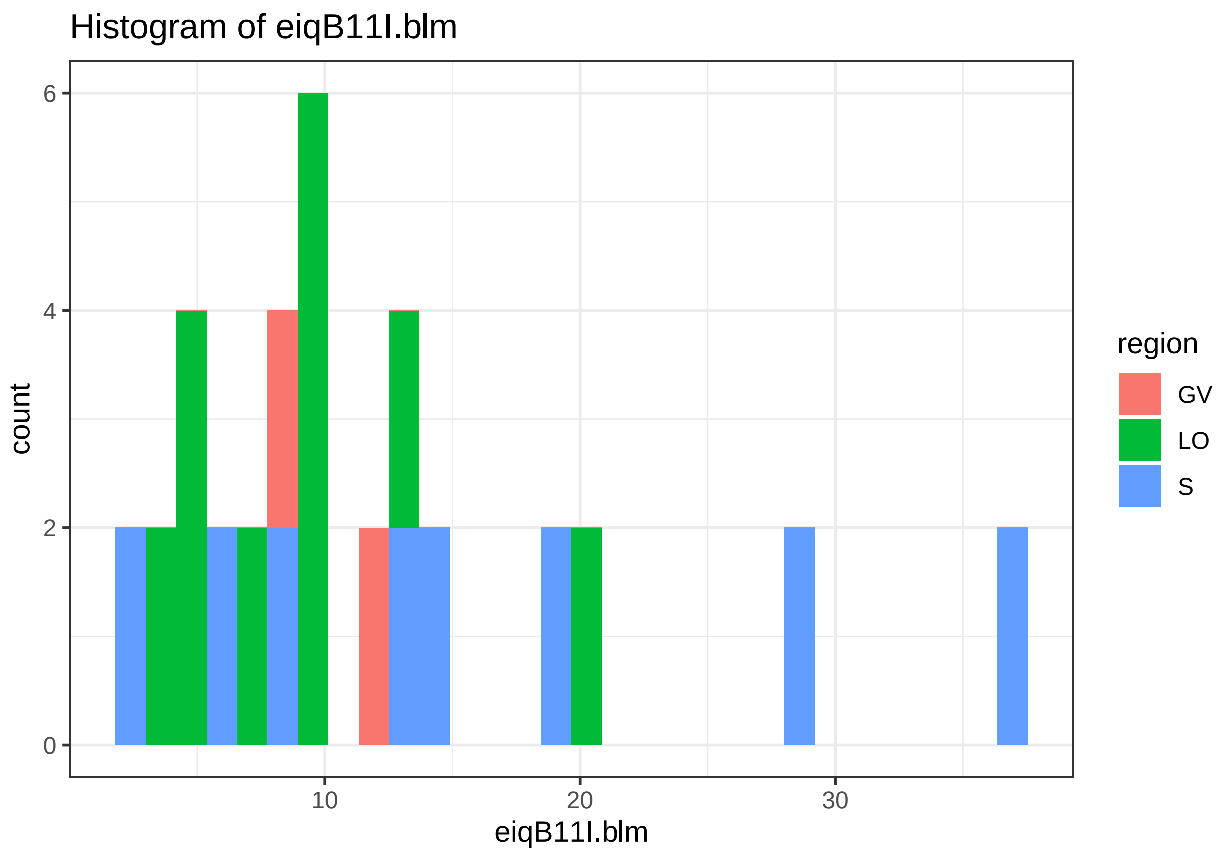
```
##
## [[5]]
```

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

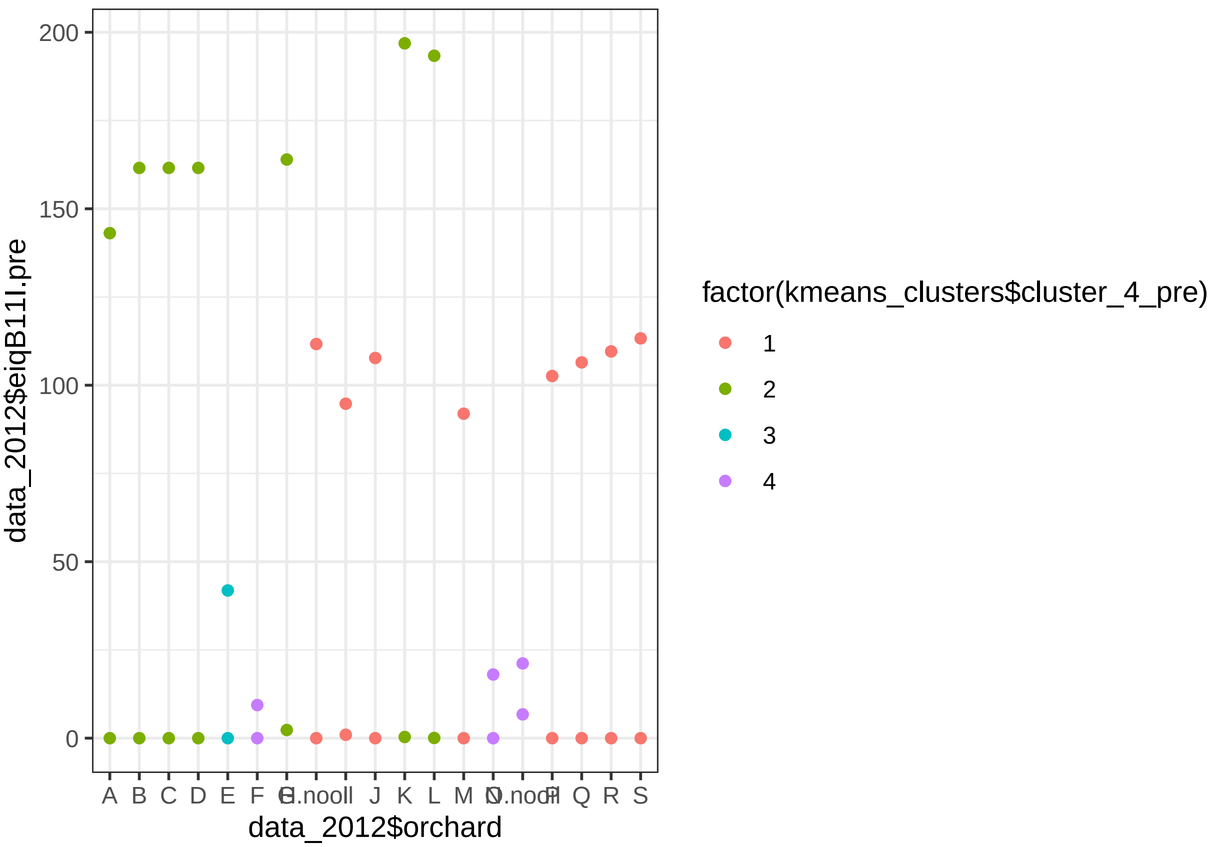


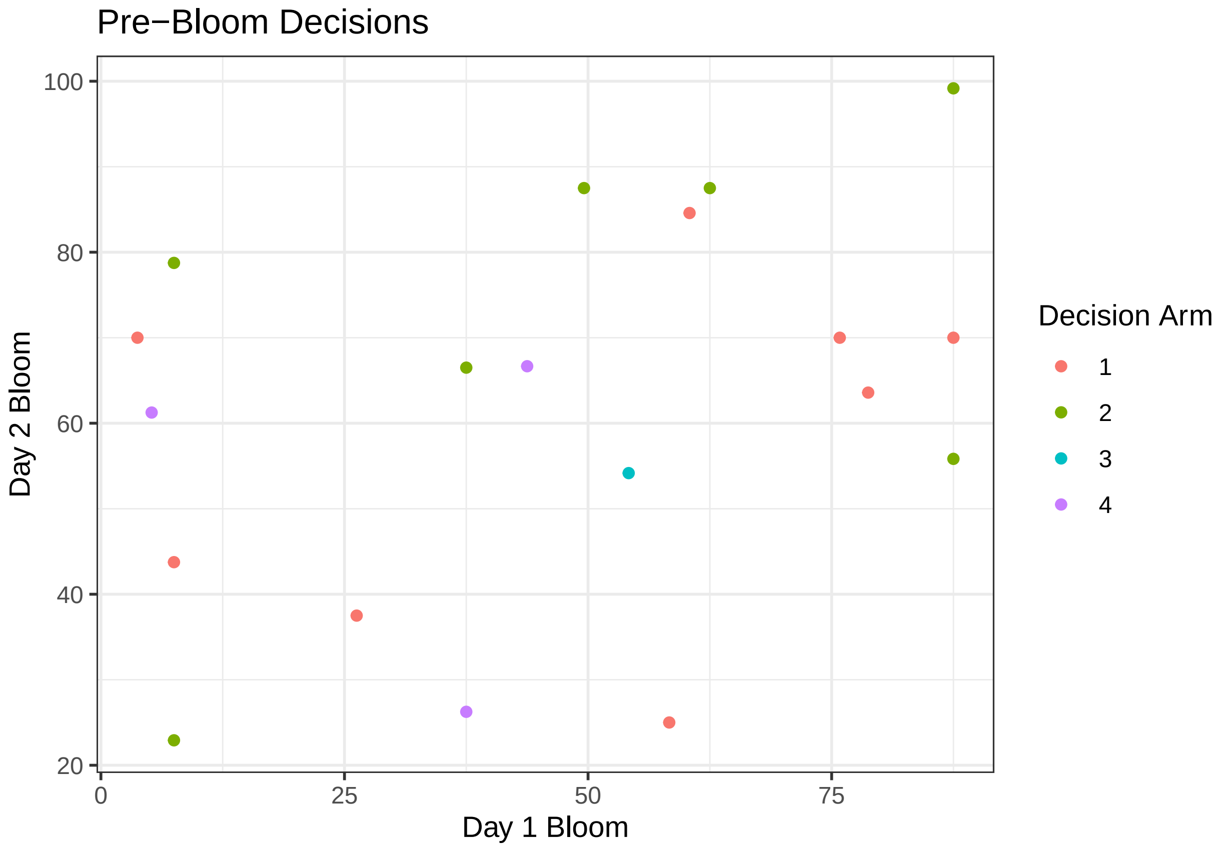
```
##
## [[6]]
```

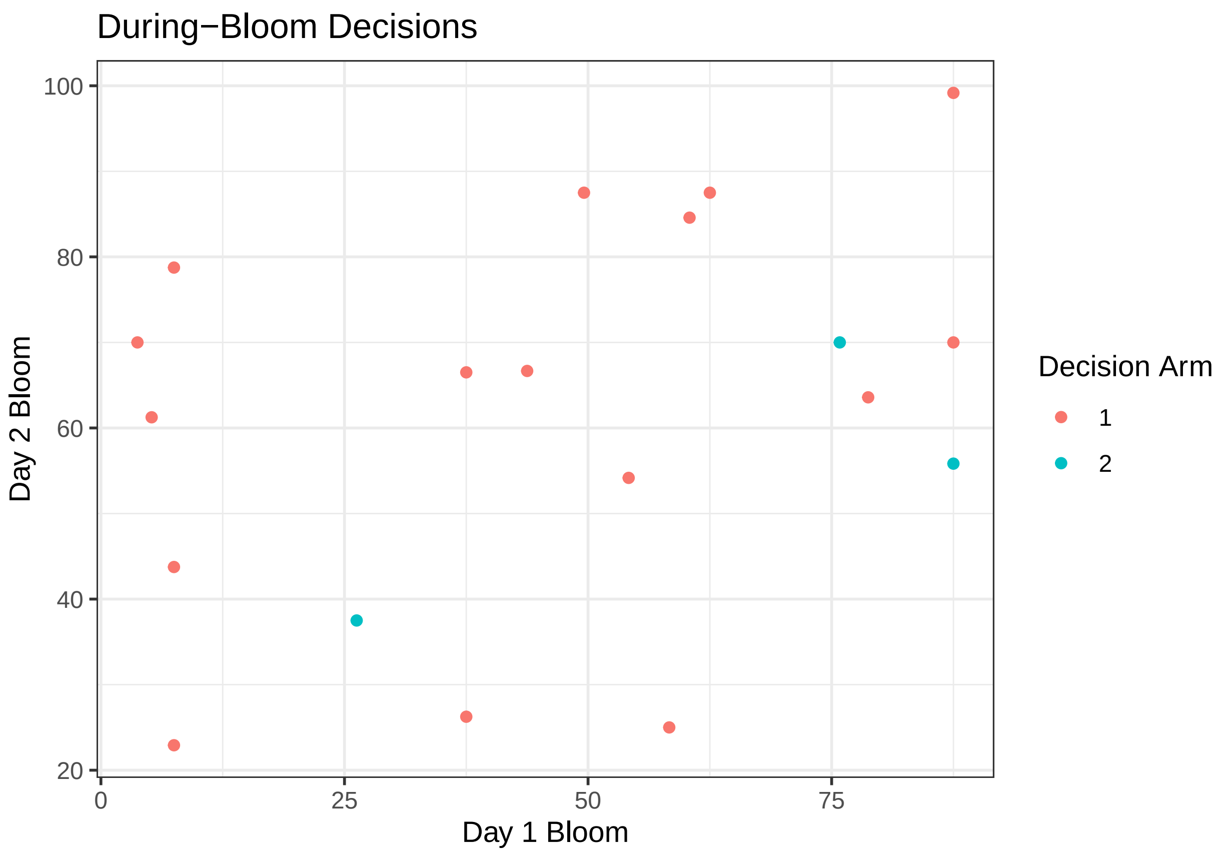
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

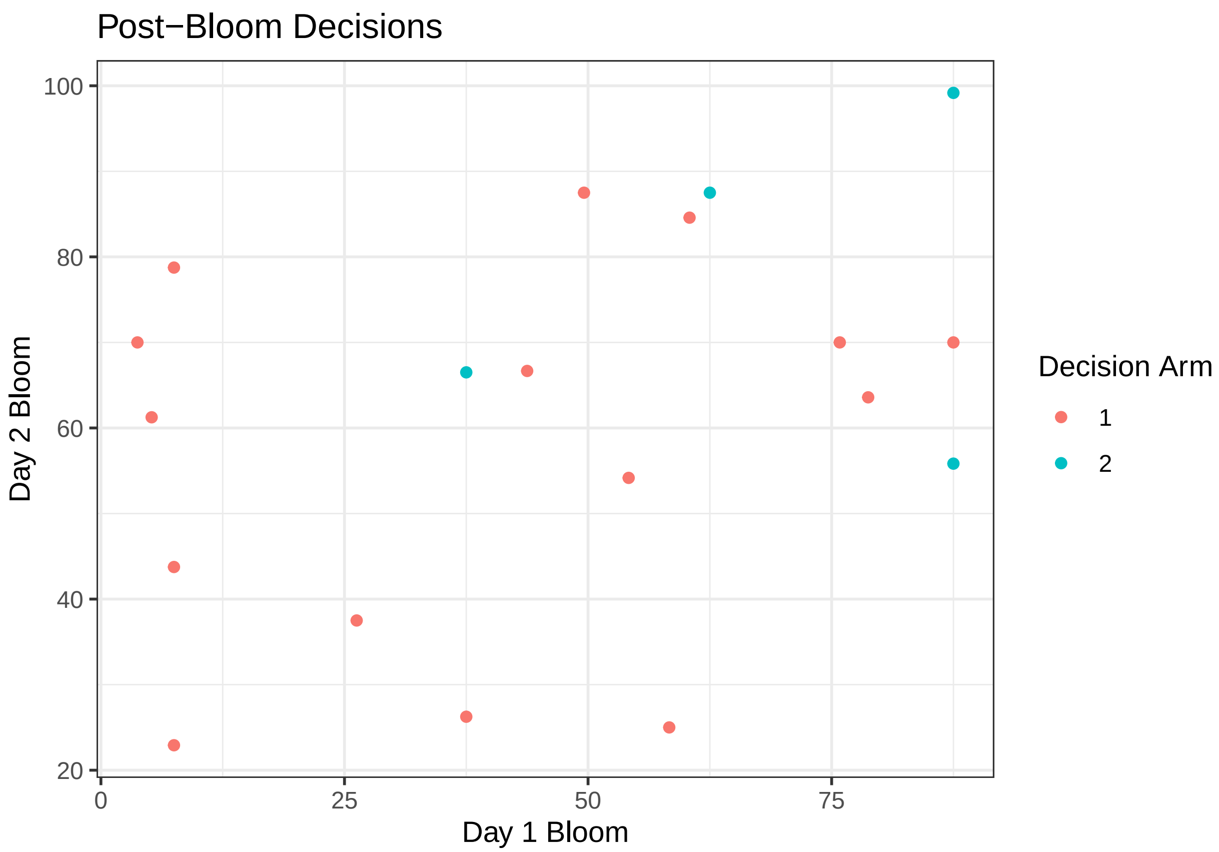


Looking at kmeans for clustering of orchards to decide on decision model.


data_2012$eiqB11I.pre
200
150
100
50
0
A B C D E F GH.noolI J K L M NO.noPI Q R S
data_2012$orchard
factor(kmeans_clusters$cluster_4_pre)
● 1
● 2
● 3
● 4


Pre−Bloom Decisions
Day 2 Bloom
100
80
60
40
20
0
25
50
75
Day 1 Bloom
Decision Arm
● 1
● 2
● 3
● 4


During−Bloom Decisions
Day 2 Bloom
100
80
60
40
20
0
25
50
75
Day 1 Bloom
Decision Arm
● 1
● 2


Post−Bloom Decisions
Day 2 Bloom
100
80
60
40
20
0
25
50
75
Day 1 Bloom
Decision Arm
● 1
● 2