

# For Botanist

*Stephen Brownsey*

*03/11/2019*

## Introduction

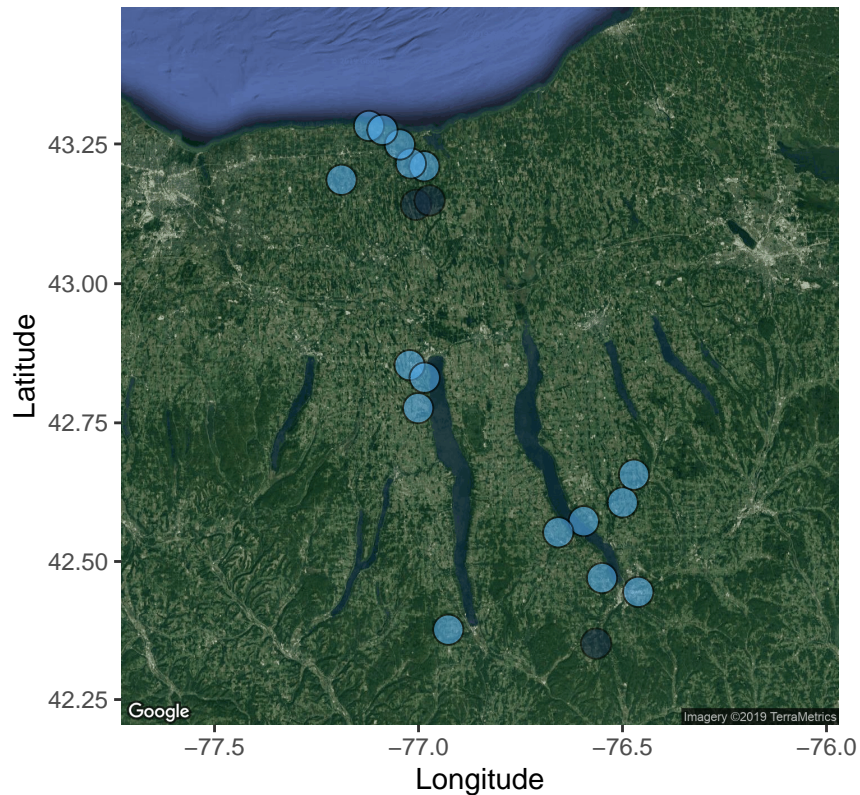
Questions on Bees in Orchards study. But in particular, I'd like to know your opinion on the subsetting of the data for the purpose of the decision models, in particular if there were any "obvious" ways to model it to start with. The data itself, doesn't seem to be great in my opinion for model building 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 wonder 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=640x640>

## Orchard Locations



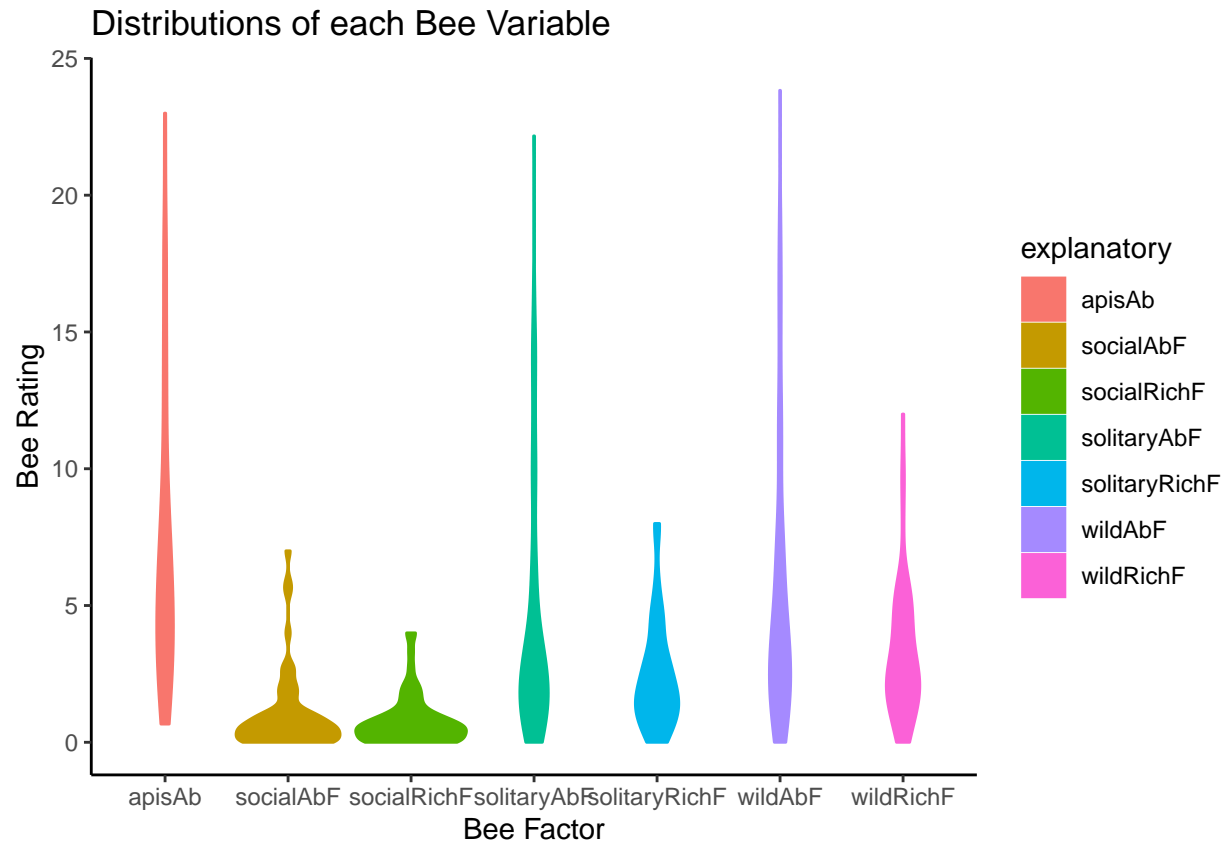
The variables in general are very correlated, as you would expect, but this means that for the decision models I would want to be looking at a subset of these but from a botanist's point of view which variables would she consider best? Not that clear in a .pdf but lots are upwards of 0.8/0.9 correlated.

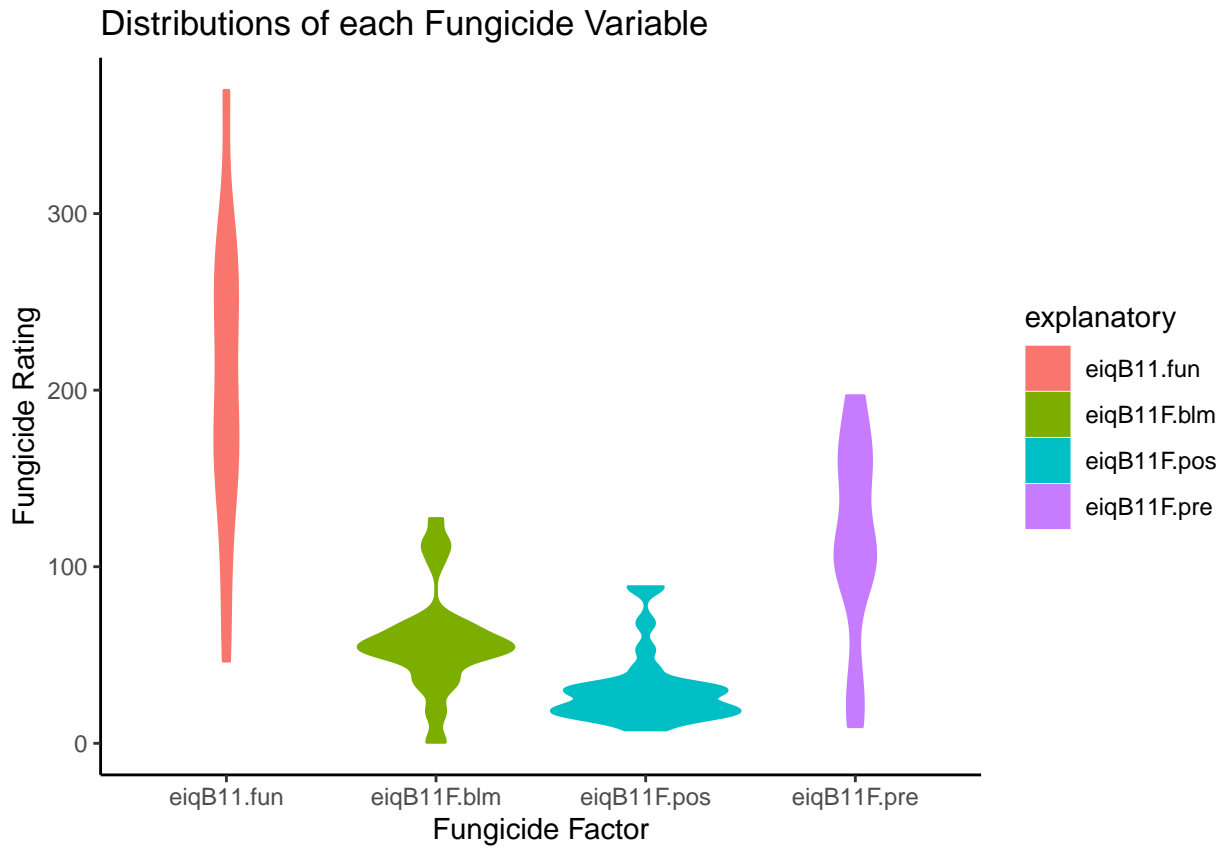
## GGTally

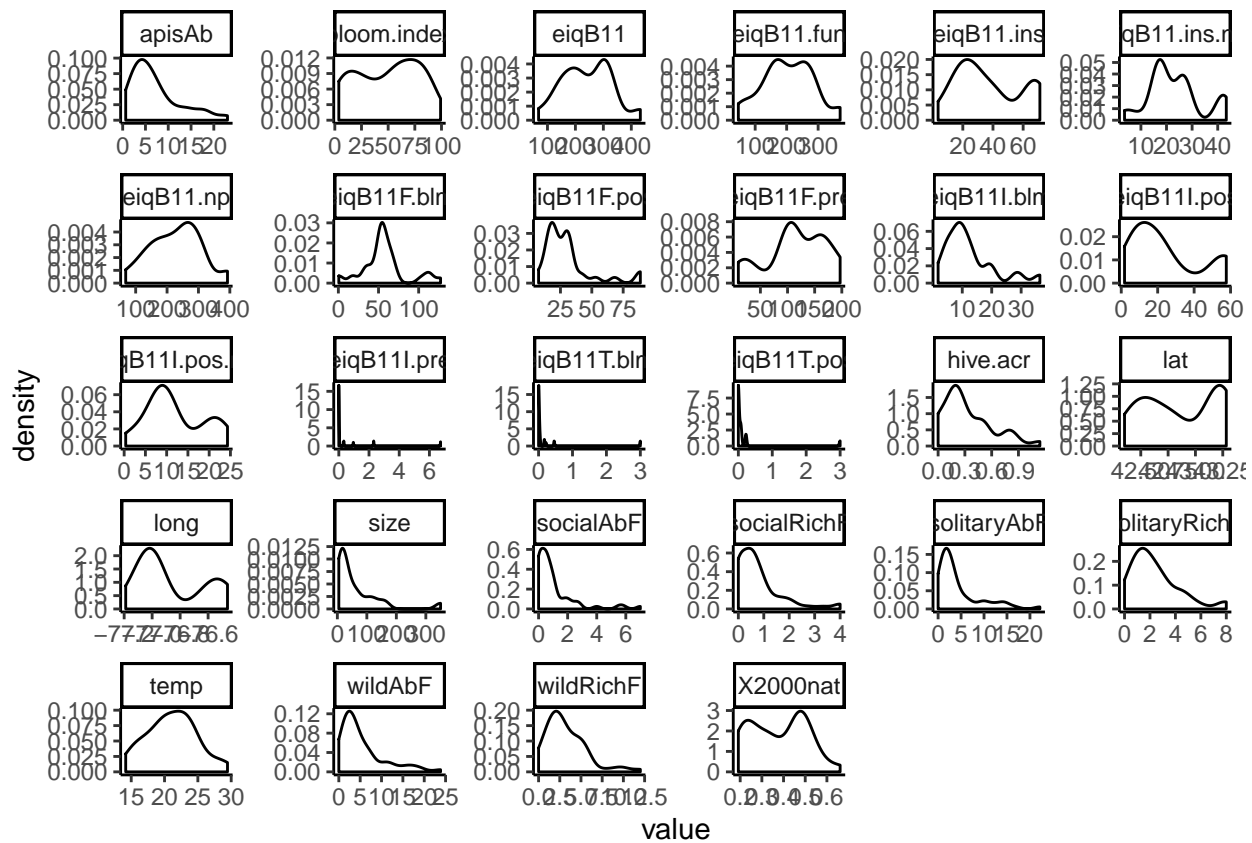
### Looking at individual variables

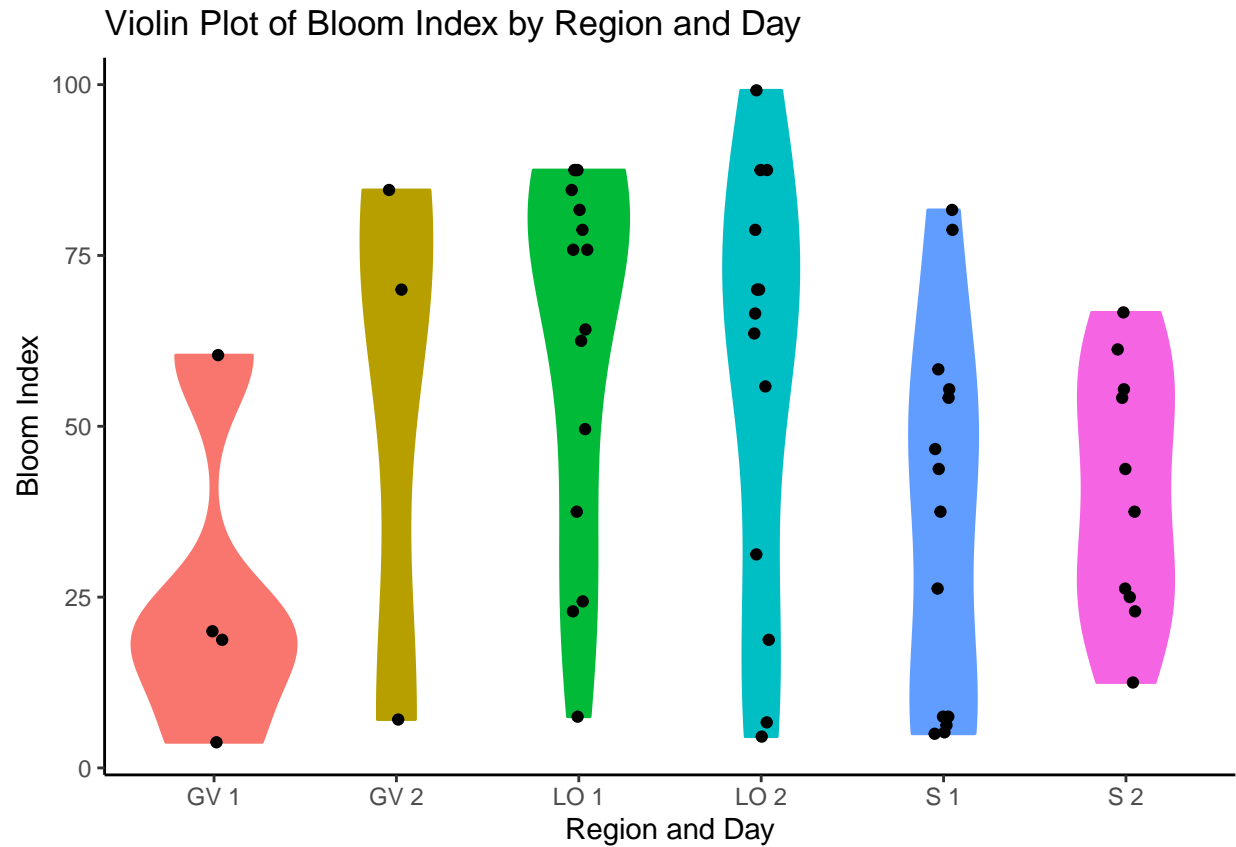
### Code post 25/10

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.





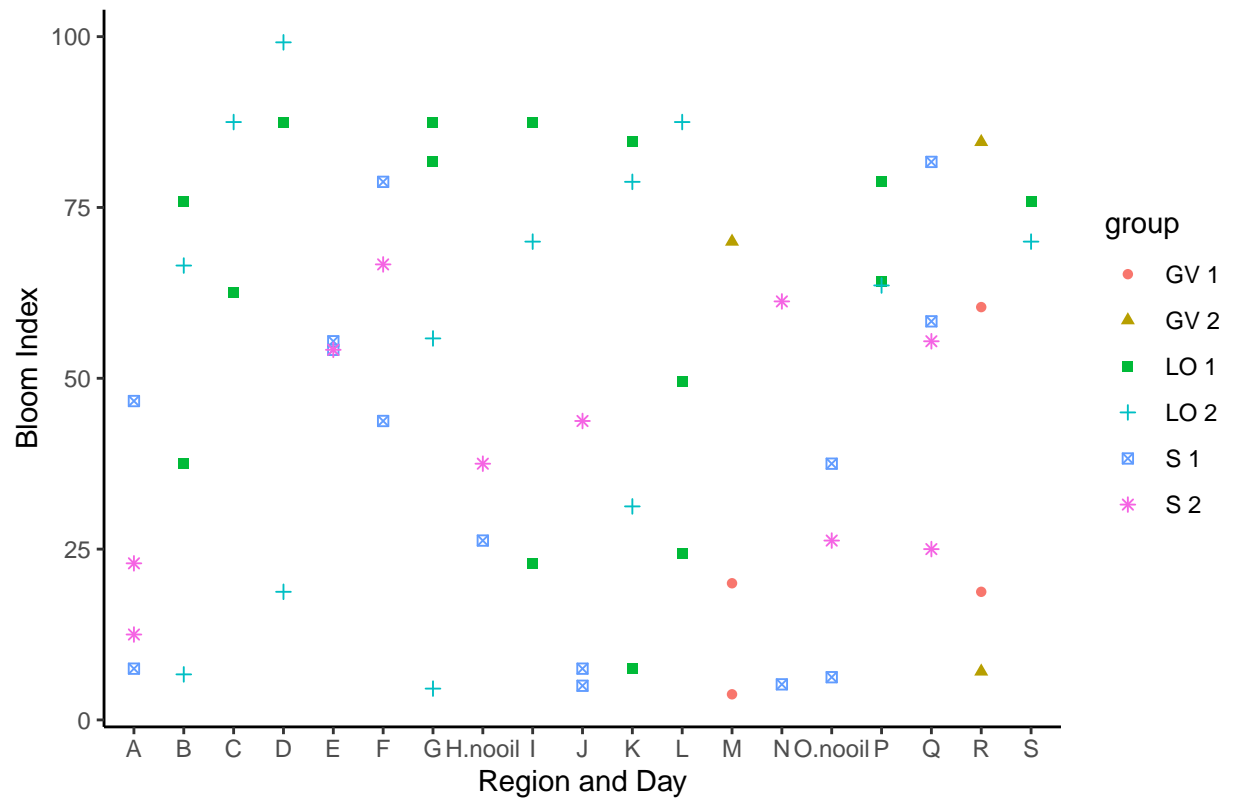




### Graphics from botanists chat

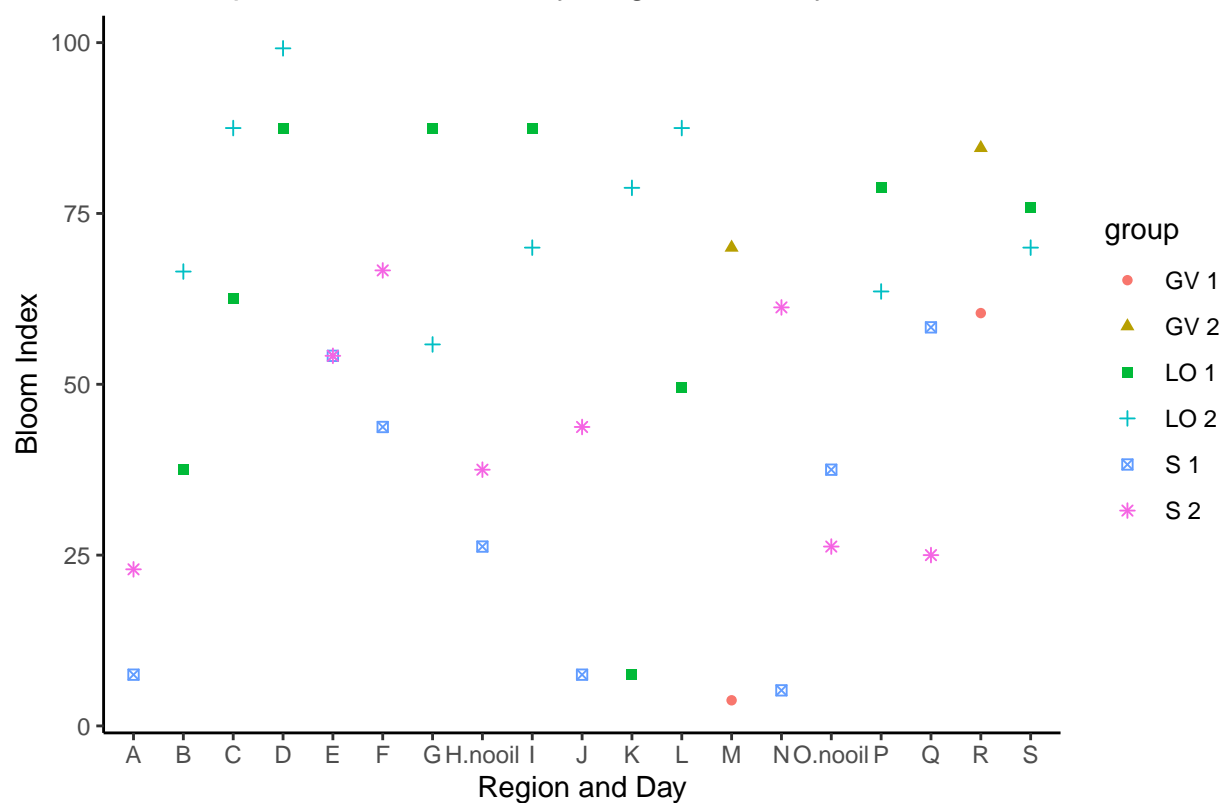
Another thing: Instead of the last figure with violin plots about bloom for region/day, could you do a scatter plot visualising the relationship between day 1 and day 2 bloom index?  $y$ =day 1 bloom index versus  $x$ =day 2 index. You can put all the all the three regions into the same scatter plot,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

Scatter\_plot of Bloom Index by Region and Day



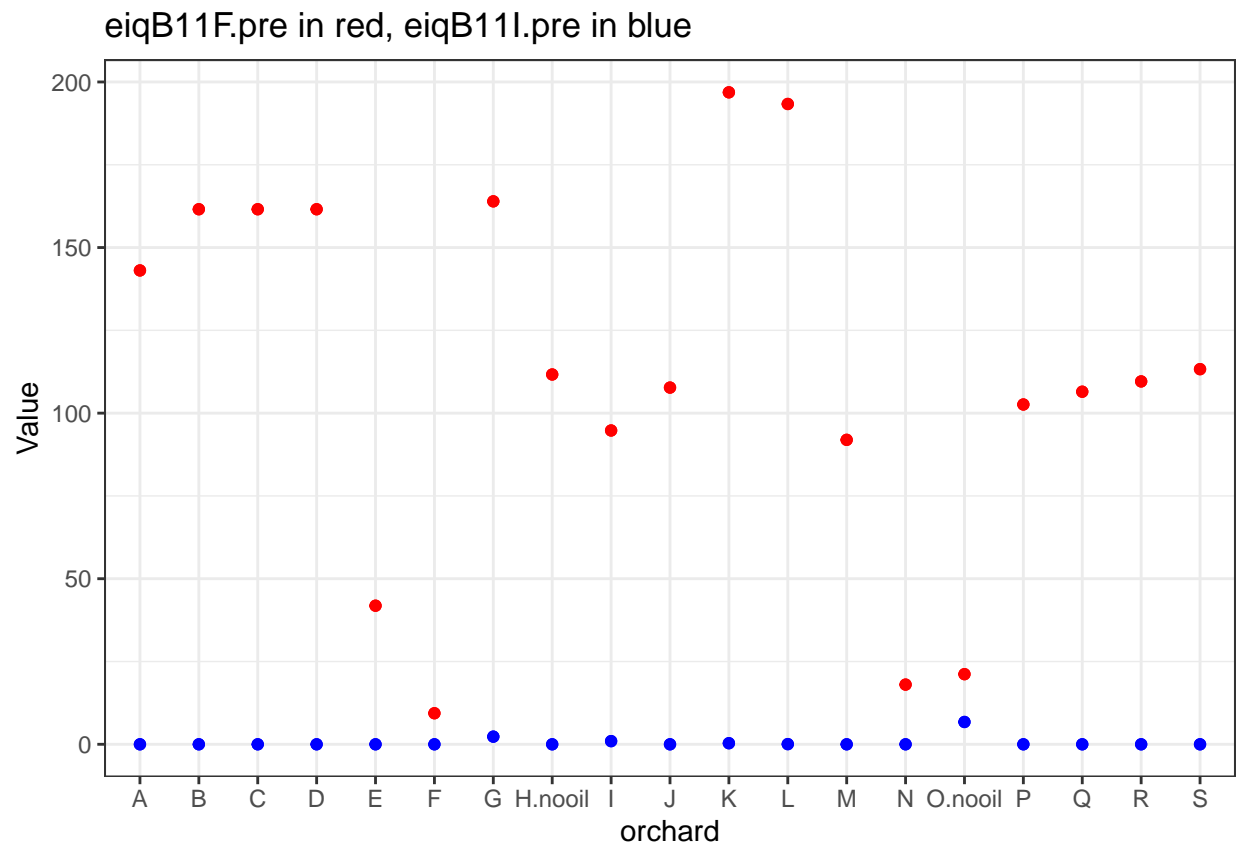
for each region.

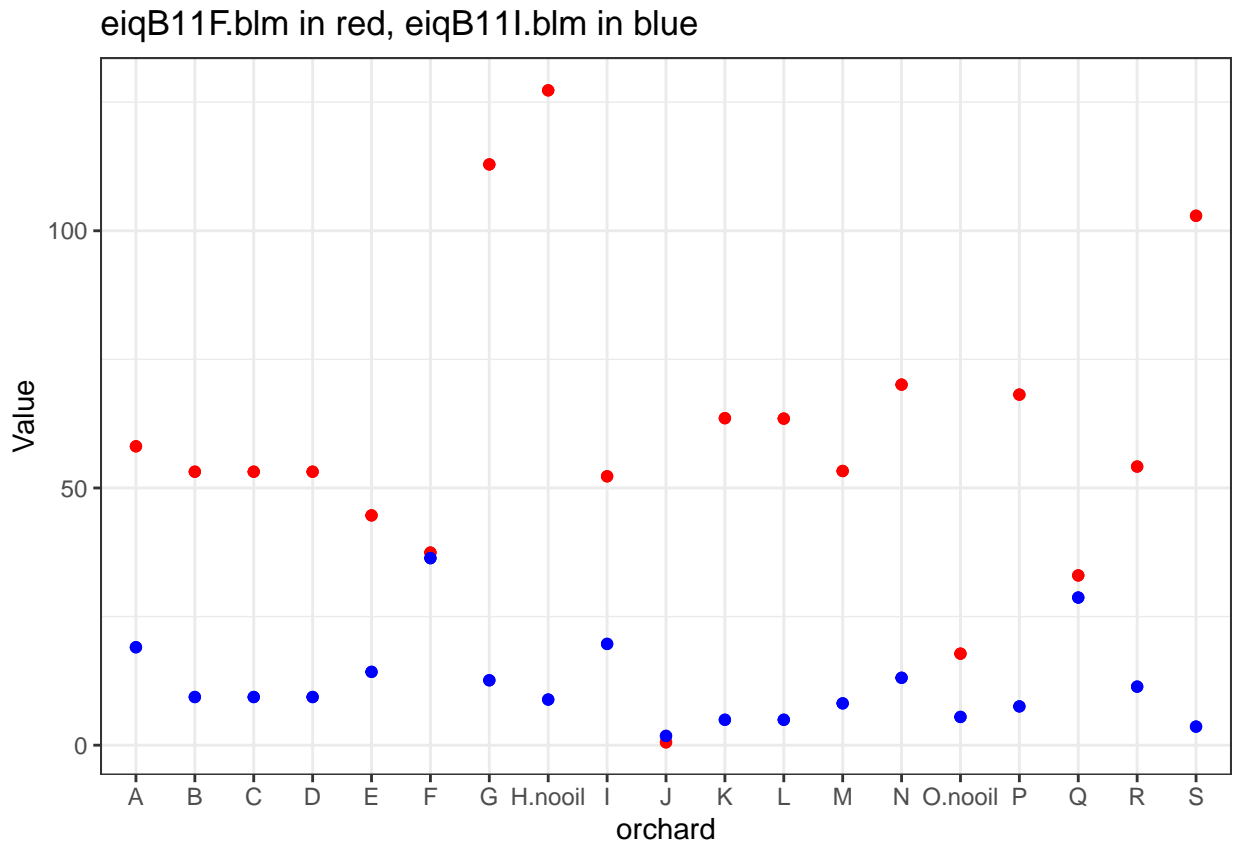
Scatter\_plot of Bloom Index by Region and Day for 2012 Data



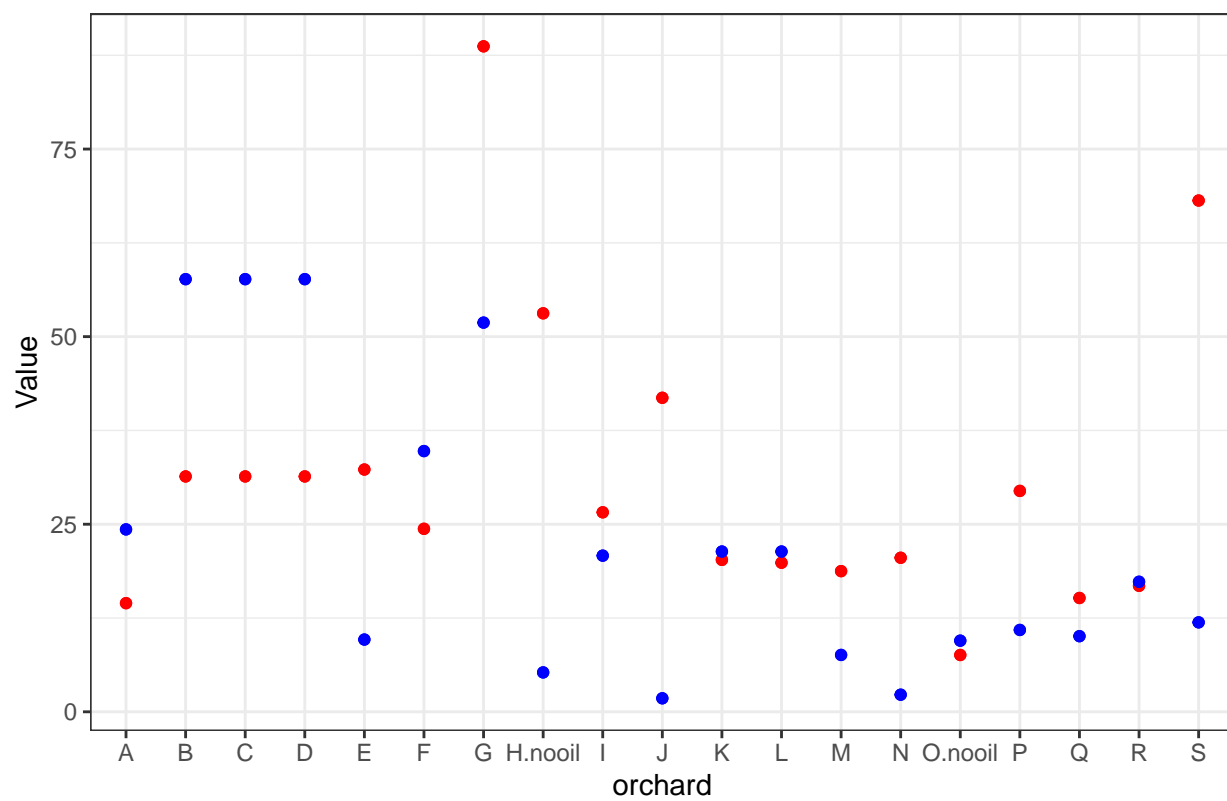
The first step for us is to understand the chemicals applications better, including the combinations. The variables in the data are actually scores developed by the authors that capture the impact they have, so that is quite ready to use and comparable. (See Park et al for details.)

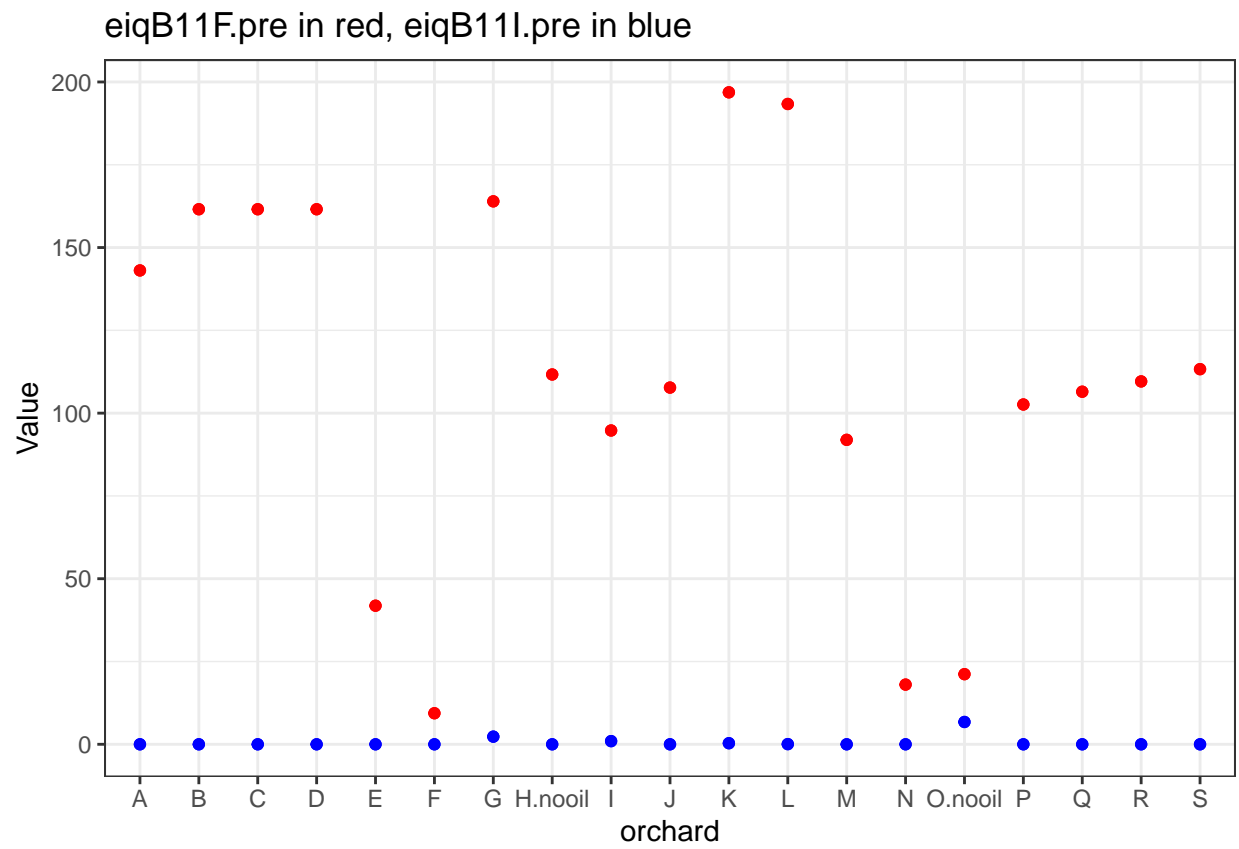




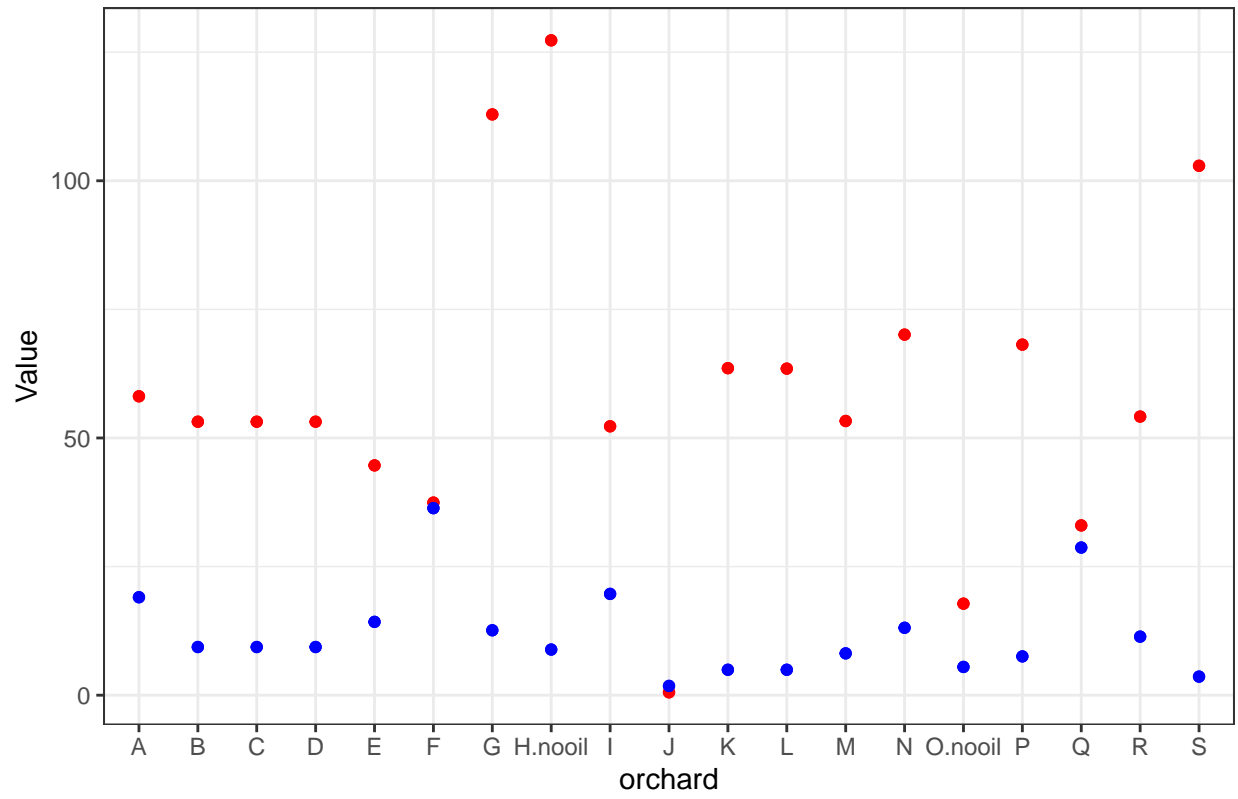


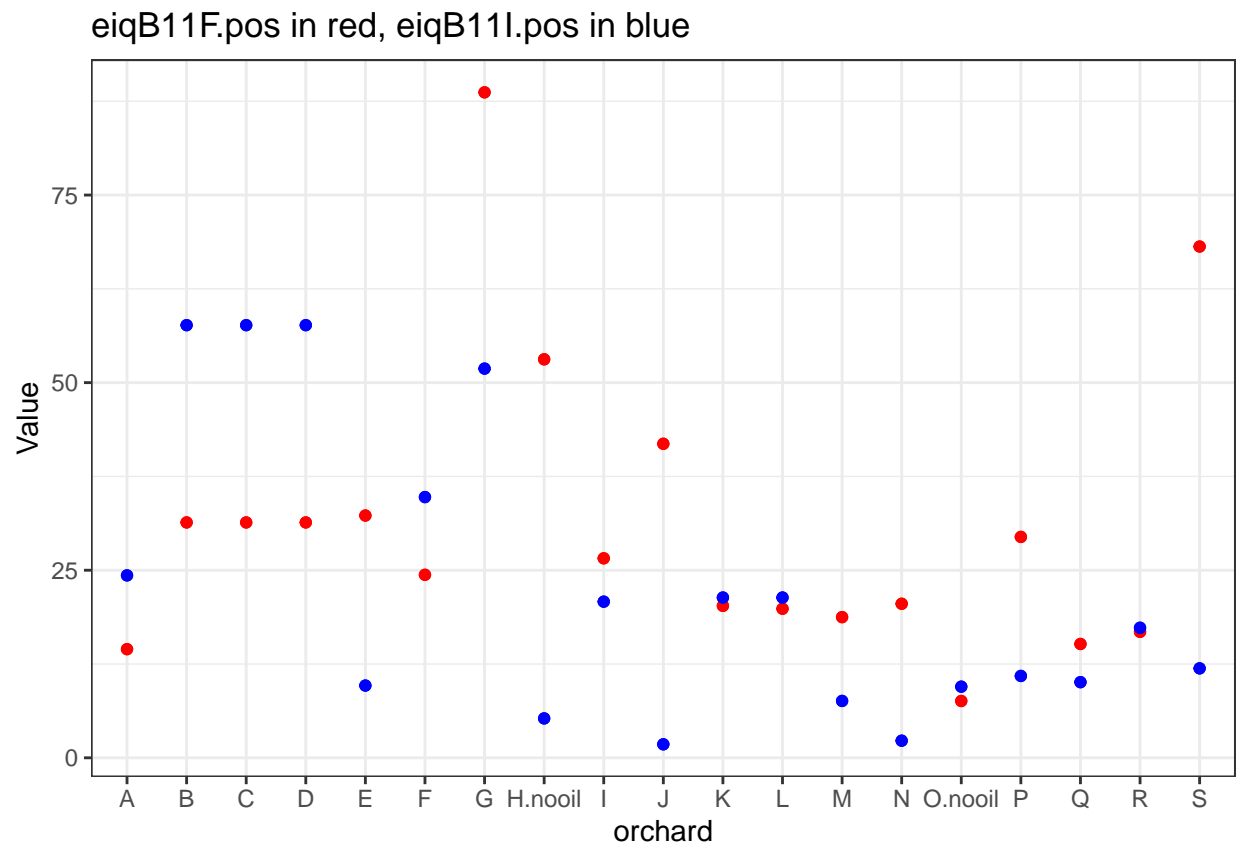
eiqB11F.pos in red, eiqB11I.pos in blue



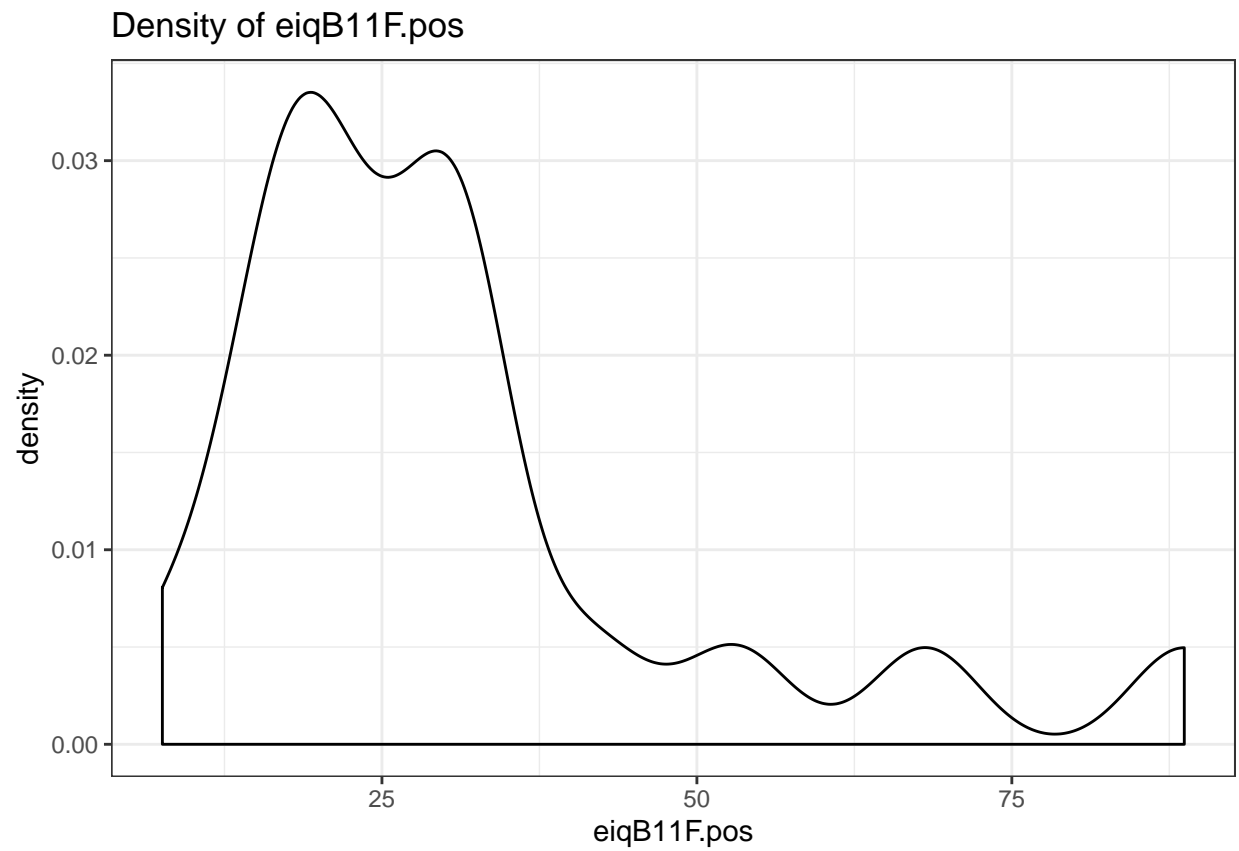


eiqB11F.blm in red, eiqB11I.blm in blue

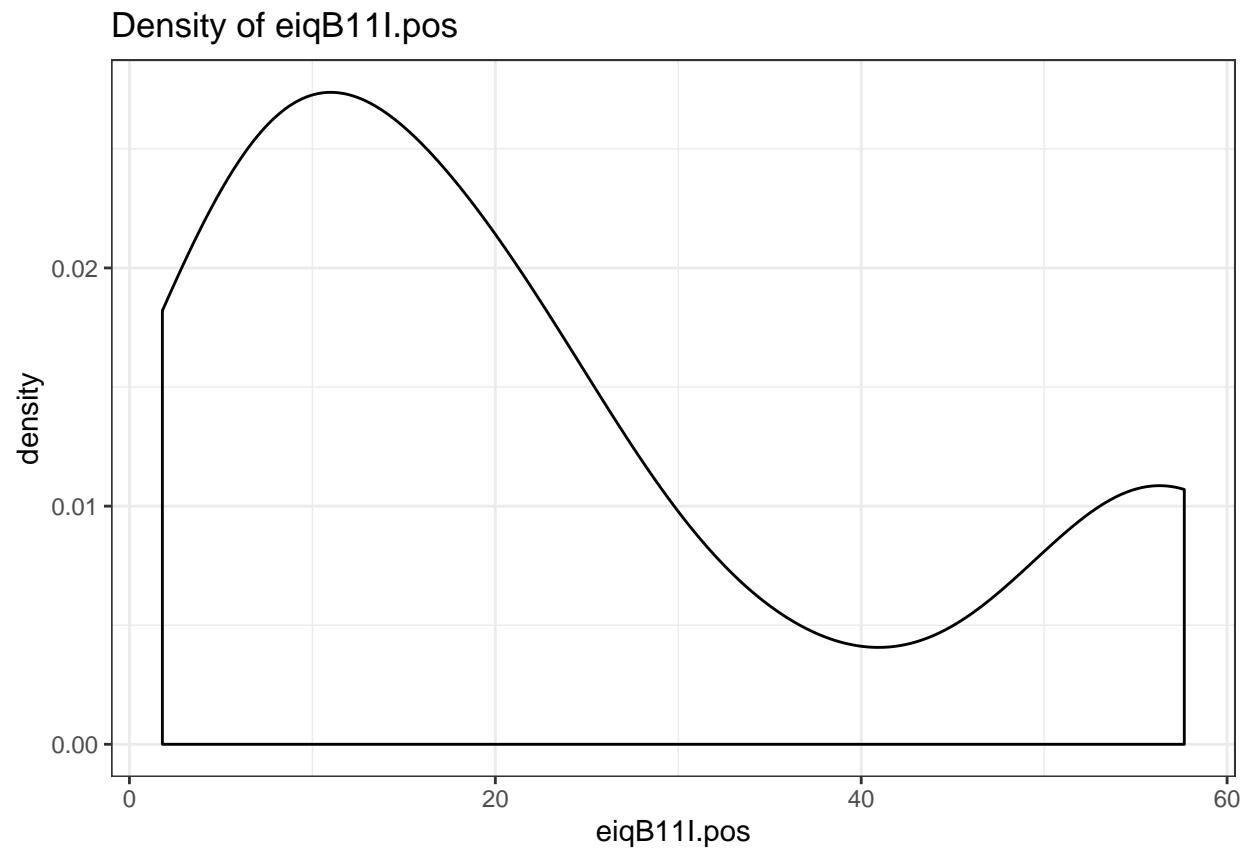




## [[1]]

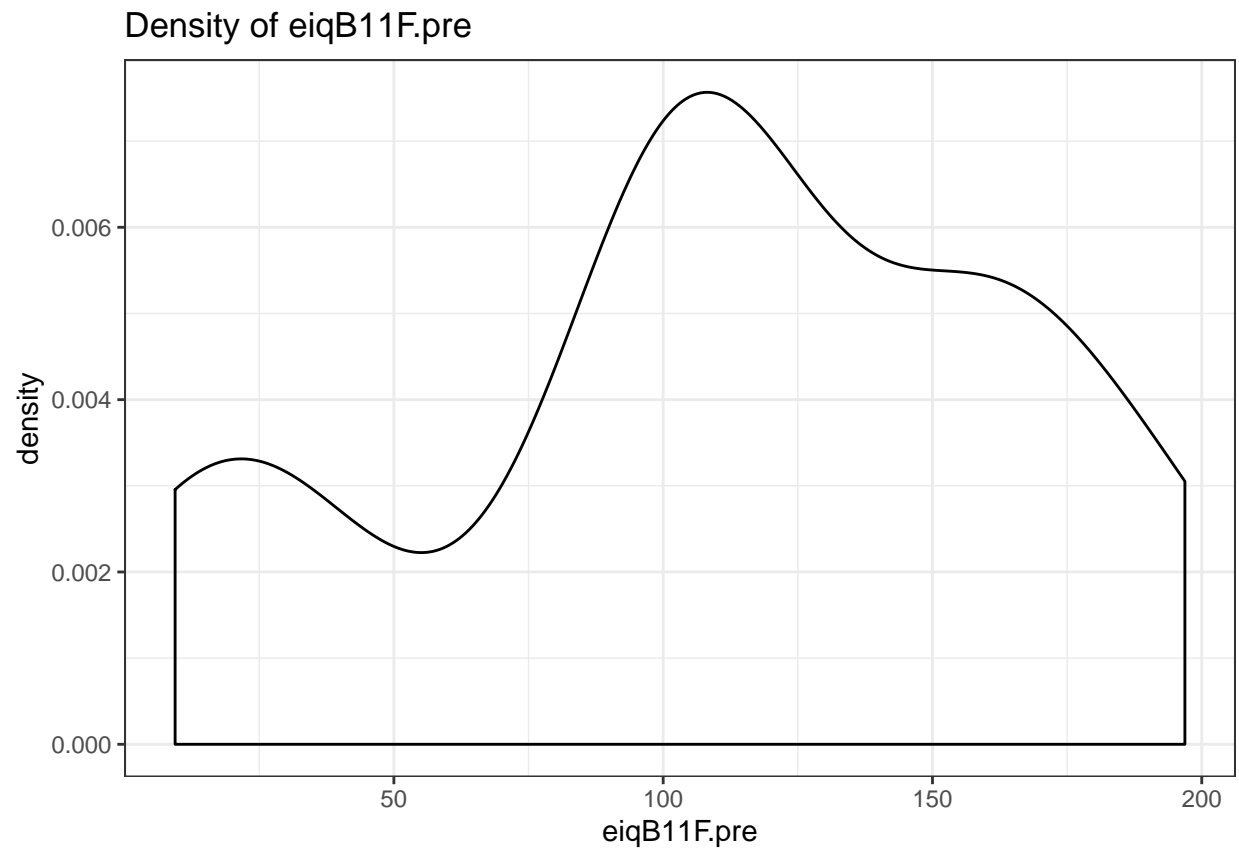


```
##  
## [[2]]
```

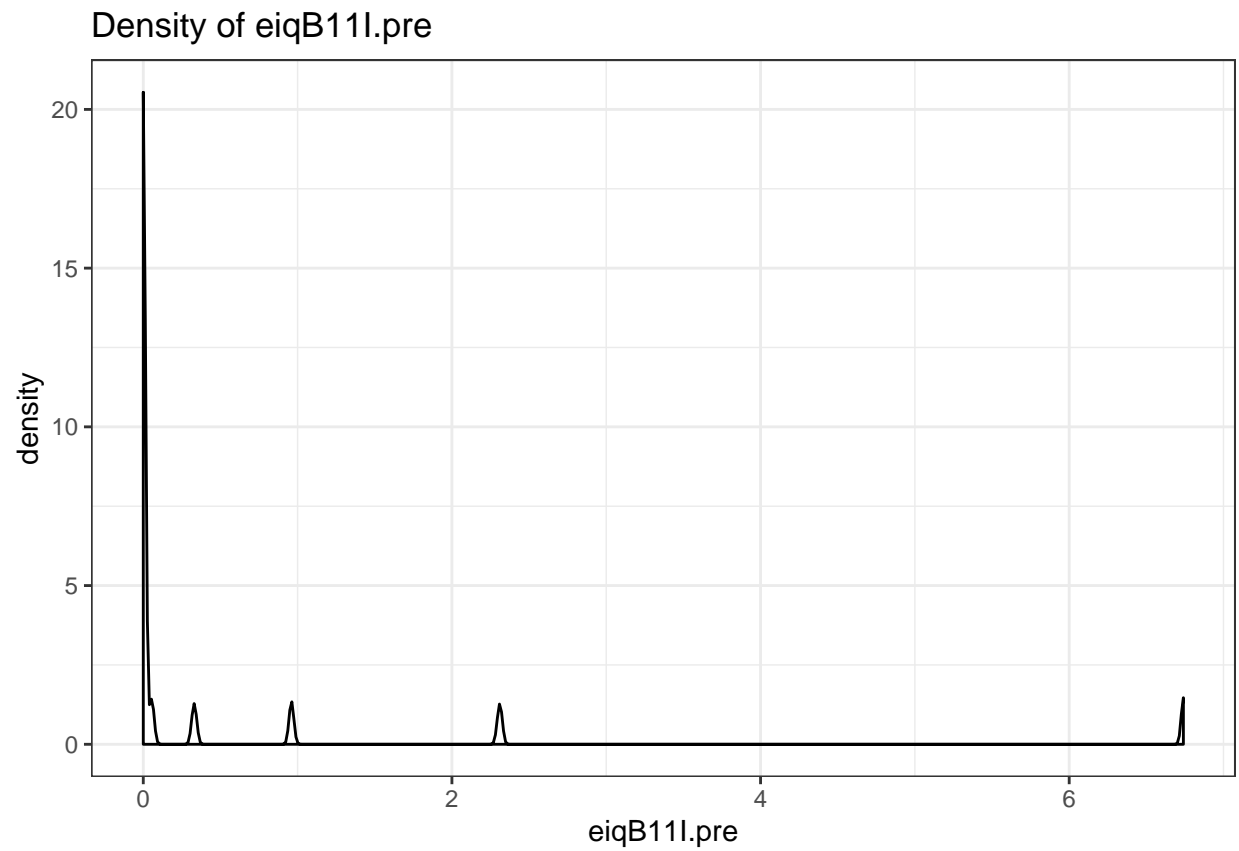


```
##  
## [[3]]
```

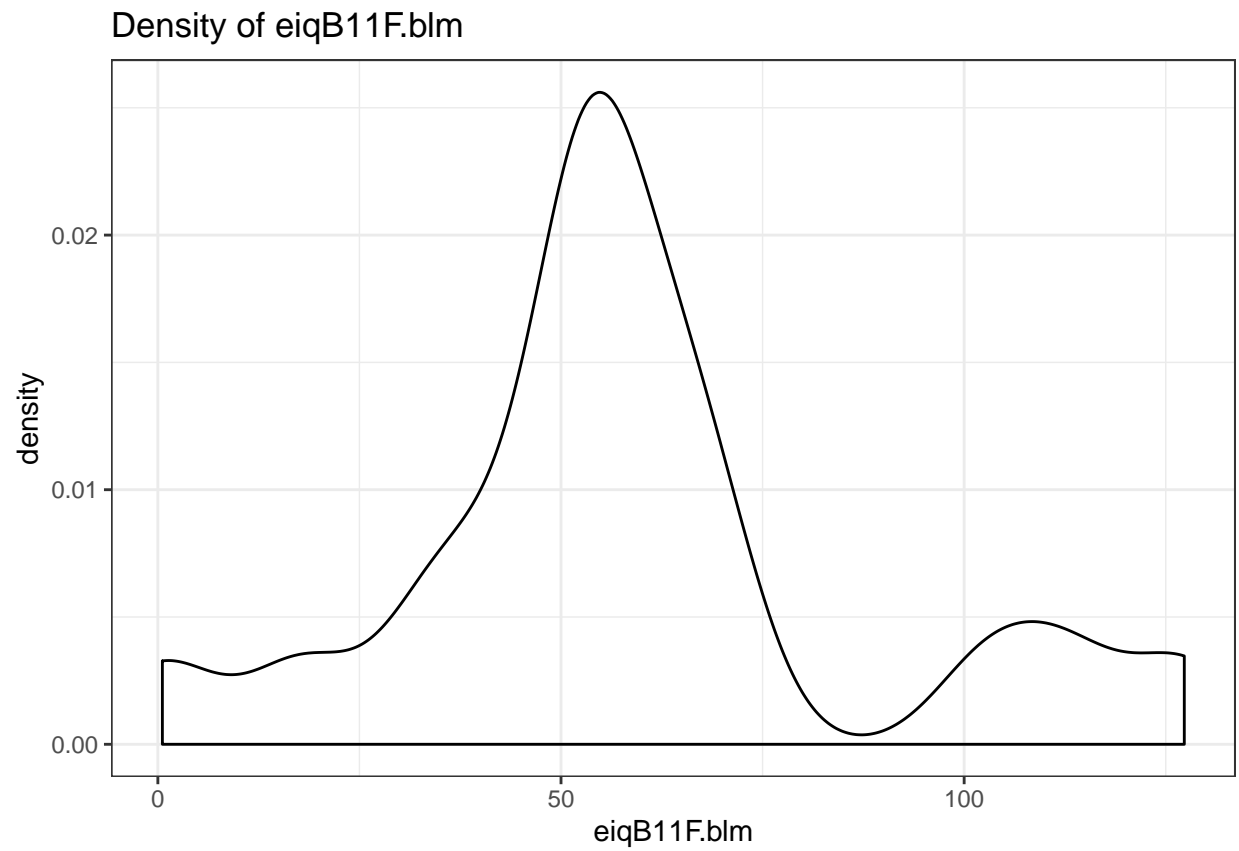




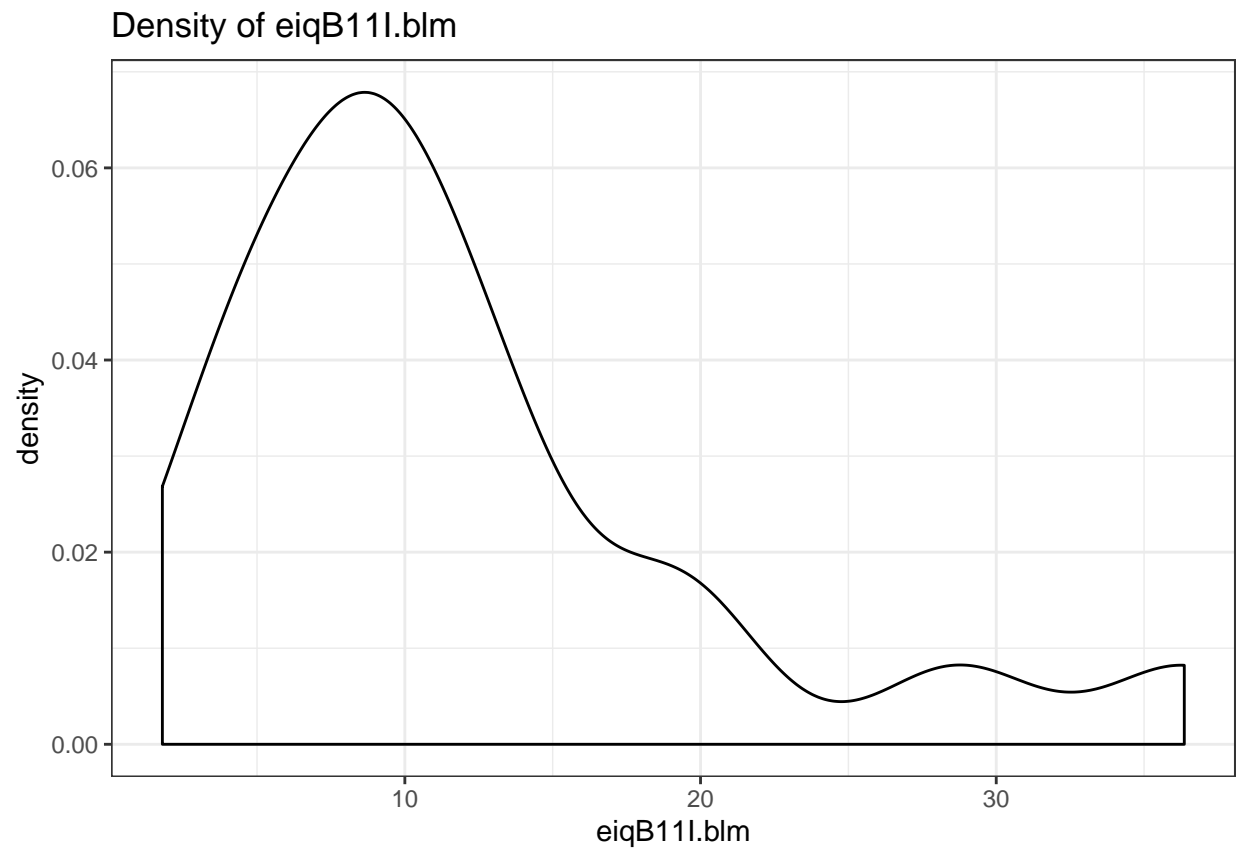
```
##  
## [[4]]
```



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



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



““