

Paper Diagrams

Paul J Palmer

11/01/2019

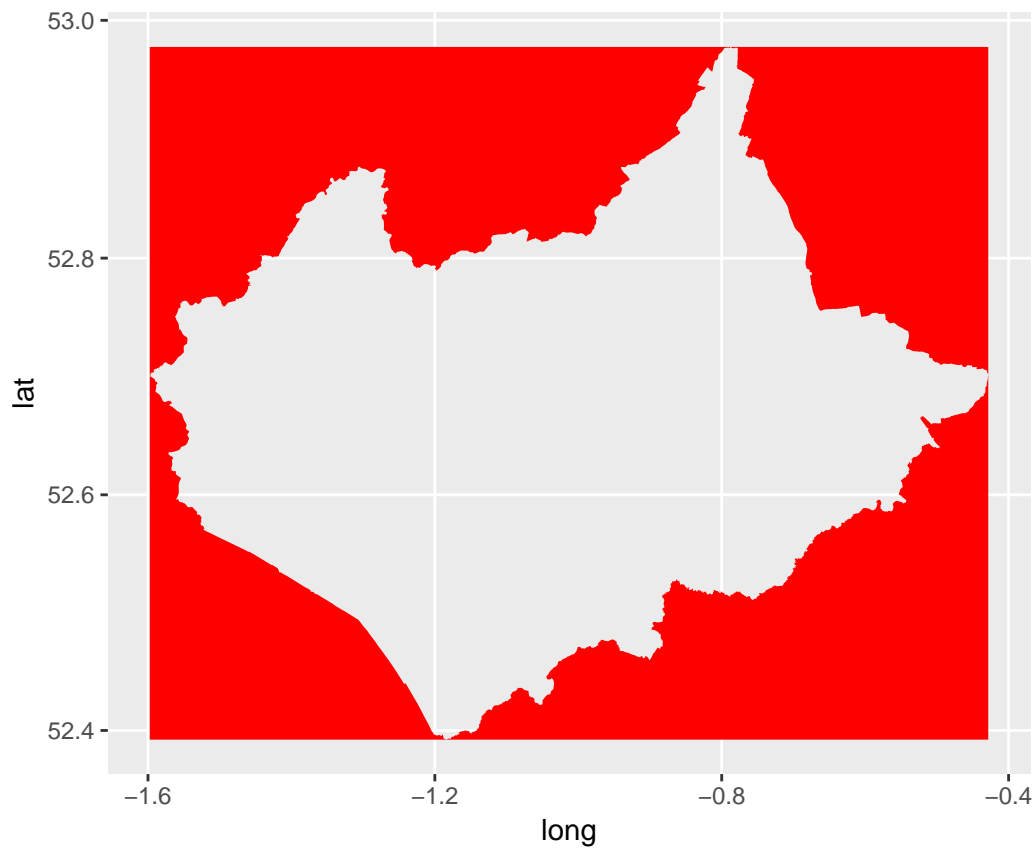
Diagrams for Academic Paper

These diagrams been created dynamically. The version with the full background code will be lodged with the data for reproducibility. The released base data will be anonymised using the guidelines published by Butterfly Conservation and contains no personal details that can be associated with individual records.

The appendices were created prior to the summary diagrams. This ‘back to front’ workflow is needed to simplify the narrative, but adds complexity to the production of this document as it is necessary to ensure that the execution order of the code chunks do not affect the results. Note that the position of the setup option is at the head of the document.

Code setup

Setup map



Data Preparation

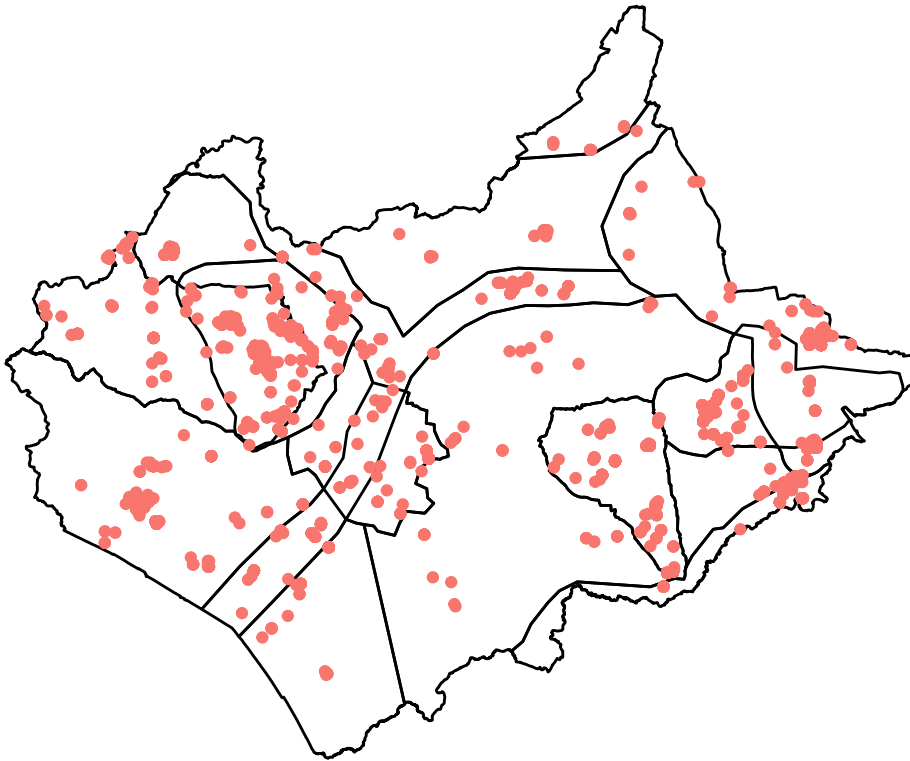
The data has been cleaned using the PrepareDataForETL package written by author PJP. Of particular note, were the need to convert dates to the ISO yyyy-mm-dd format due to the use of multiple formats. For ease of analysis, additional spatial information have been added to the data relating to the underlying geology and major land use features. Thus Greater Leicester and Rutland Water may be used in analytically consideration of records. The inclusion of these features early within the analysis helps to ensure independence of code chunk execution.

Records for Green Carpet Moth

The data comprises of 2207 observations. Several calculated fields have been added:

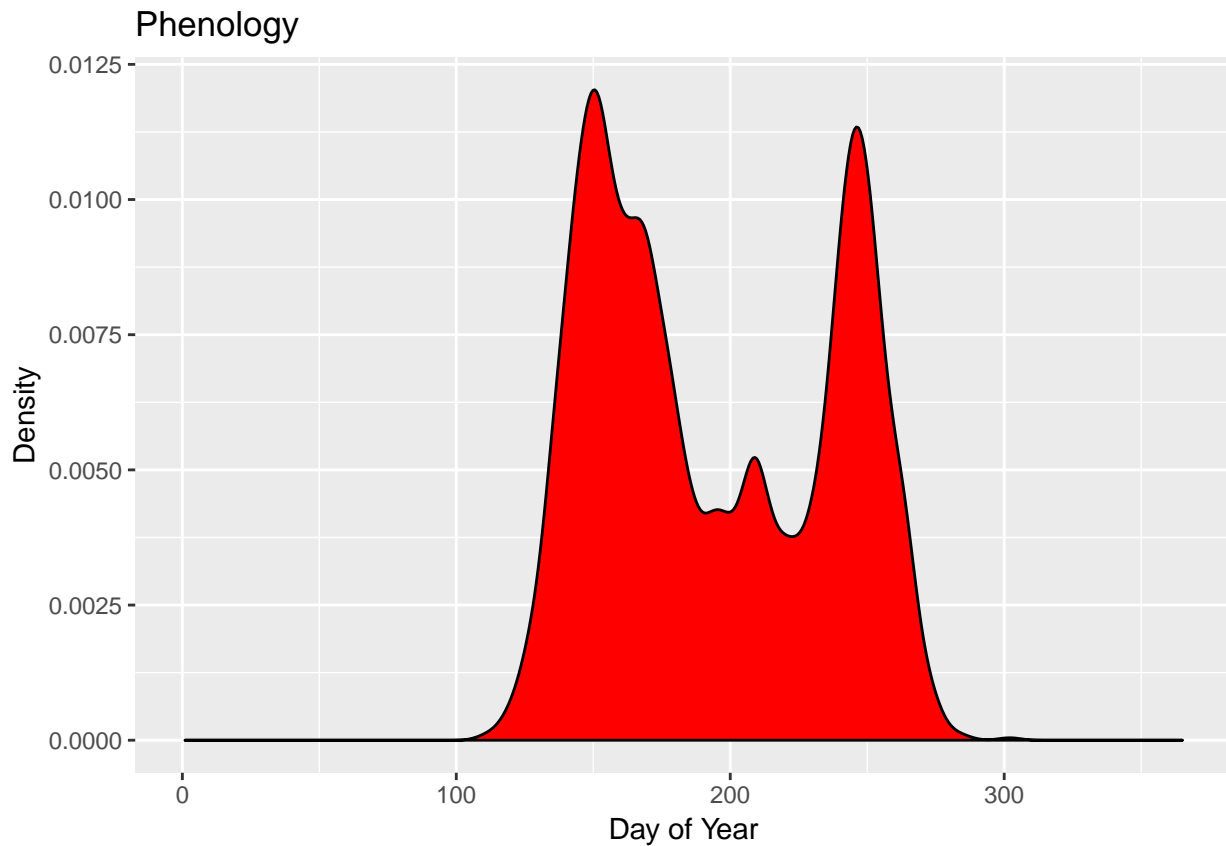
- *Latitude and Longitude* Calculated from the observation National grid references.
- *Area* Named regions associated with underlying map polygons. All records have “All records” as part of the named area since not all strictly fall within the county boundary due to changes over time. Note also that a point may fall within several polygons, so it is necessary to filter plots to ensure that each record is only counted once;
- *Broods* Observations have been marked as single or double on the basis of flight time. This has then been used various plots. All records prior to 1985 have been considered single brooded.

Distribution all records



This map shows all 2207 observations marked on the map of Leicestershire with underlying geographical areas marked. As indicated above, filtering is possible by area. Although the taxon appears from this map to be widely distributed and common, when the flight time is taken into consideration, the picture becomes far more complex. See Appendix 1 for plots by region that help to illustrate this point.

Phenology plot of all records using ggplot



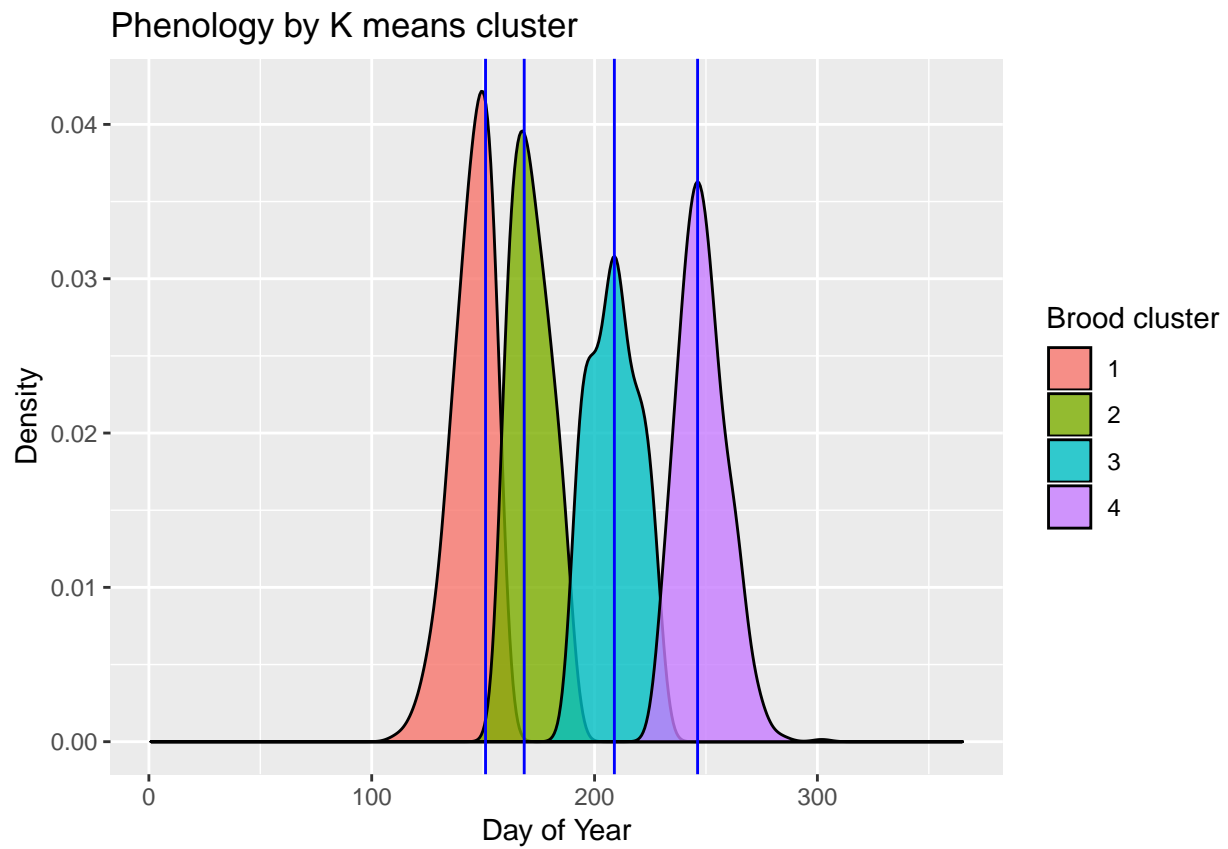
The chart above shows the phenology for all records. There are clearly mixed distributions in the plot so the following section uses a k-means clustering to separate them.

Clustering by K means algorithm

A more sophisticated method to separate the data is to use the well founded k-means method to cluster the data. Latitude, Longitude and Day of Year (DOY) were used as variables in the algorithm. After normalisation of all variables to the range 0 to 1 a multiplication weighting of 8 was applied to the DOY. This value was selected empirically, but in practice any higher value did not affect the outcome. Unexpectedly, 4 clusters gave the most compelling results.

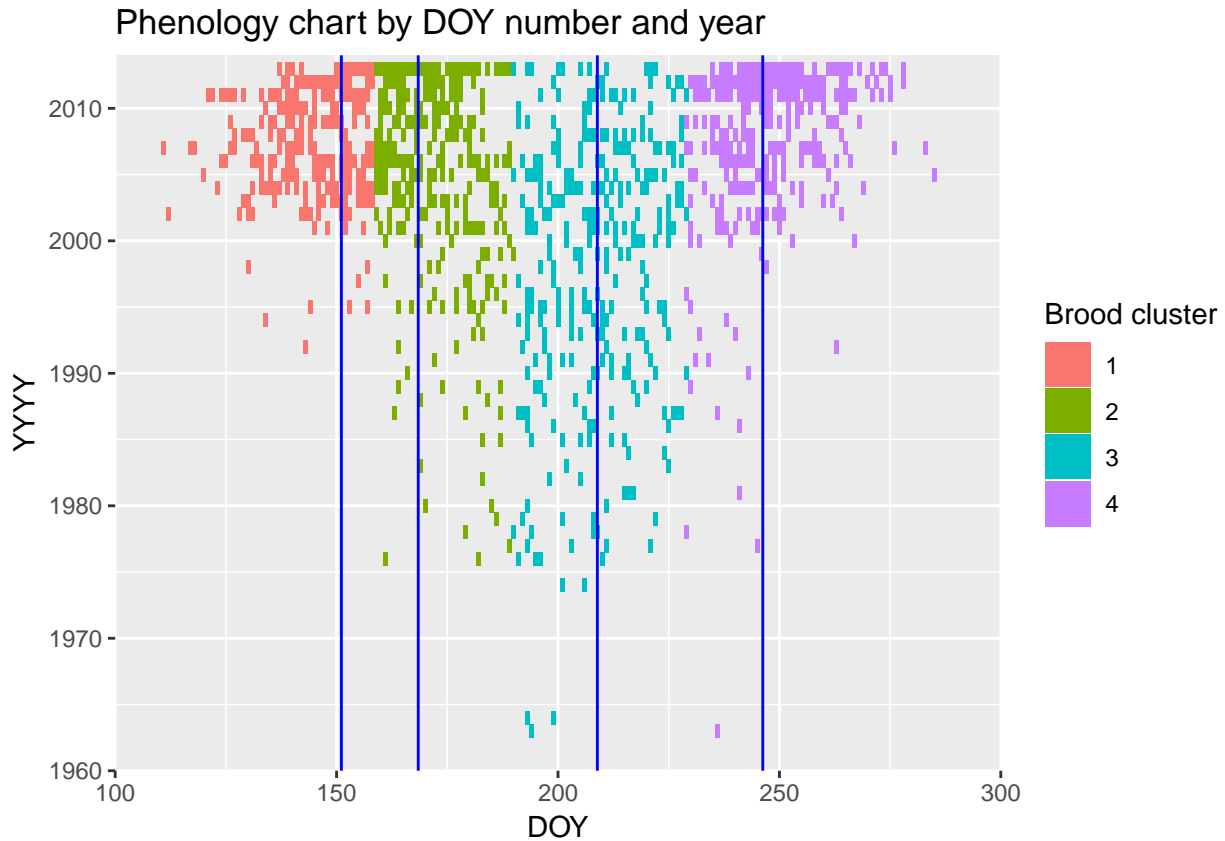
Considering the 3rd cluster as single brooded and the other three clusters as double brooded, gave almost identical results to the simple selection by day of year window.

This is an unexpected result and currently leaves open the question on what charts are the most useful to generate next.



The vertical lines mark the brood peak dates of 31-May, 17-June, 27-July, 03-September.

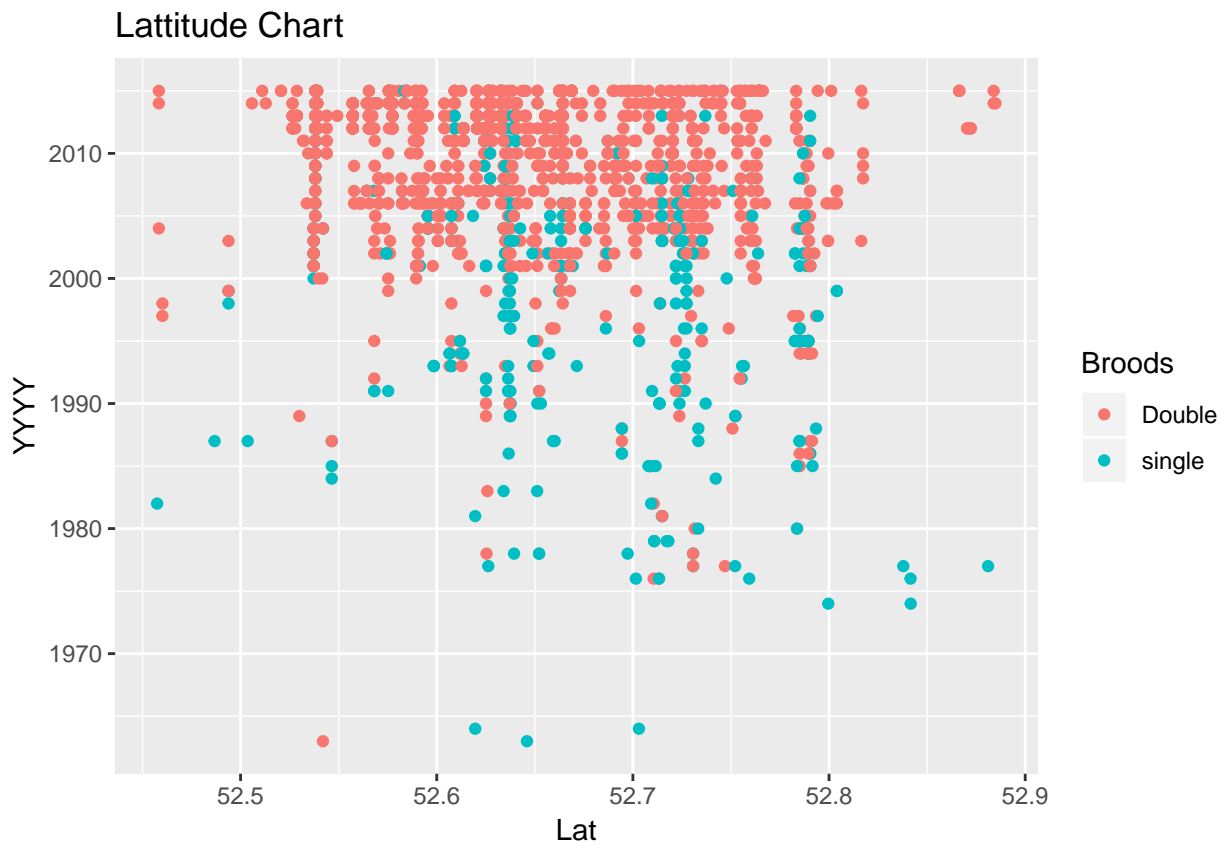
Phenology by Year including k-means groups



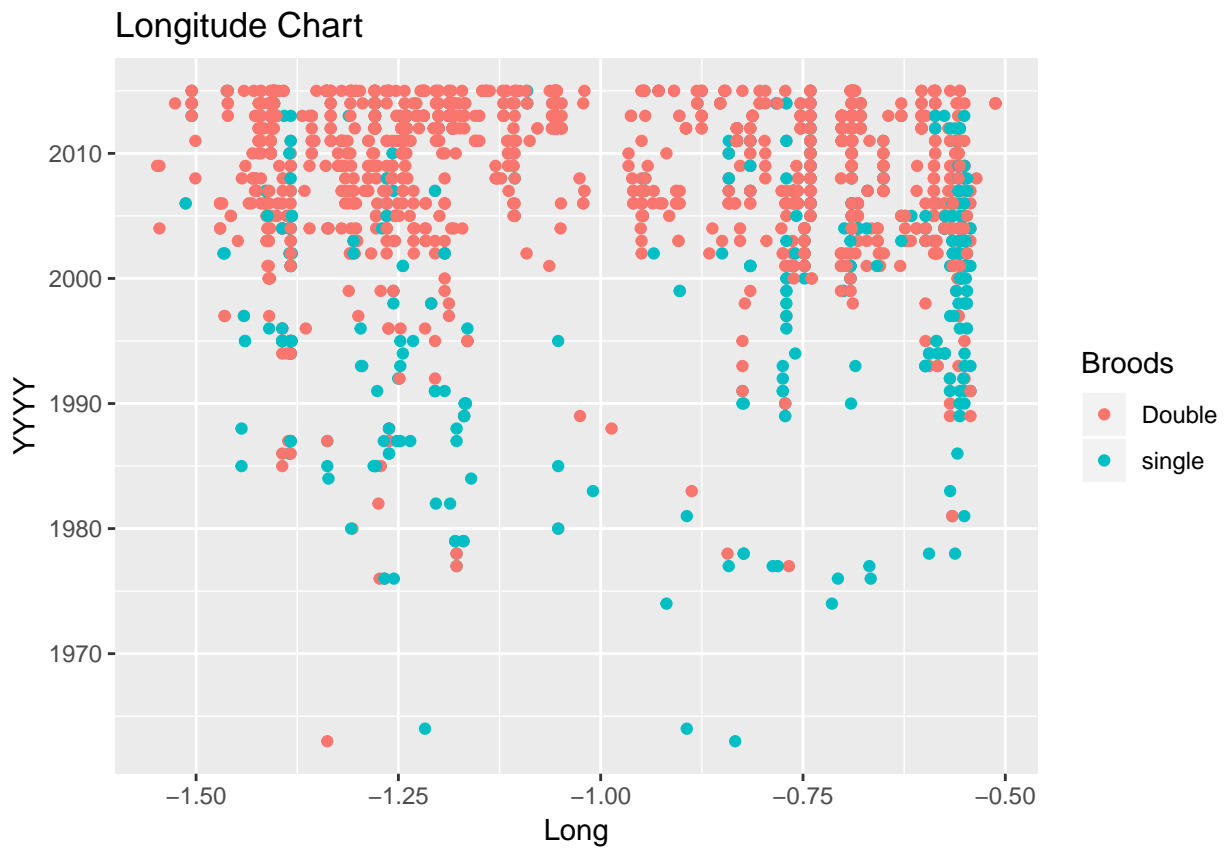
Marking the broods on the phenology chart reveals a complex picture that changes through time. The single brood 3 is constant through time. Broods 2 and 4 occur occasionally prior to 2000. Brood 1 starts post 1990 and post 2000 becomes an annual feature. Note that prior to 1990 the population appears to be single brooded. After 1990 flight times start to change and by 2000 it appears to be double brooded.

Plotting the data by group, shows that the single brooded group is separate to the other three groups which overlap in distribution.

Latitude plot

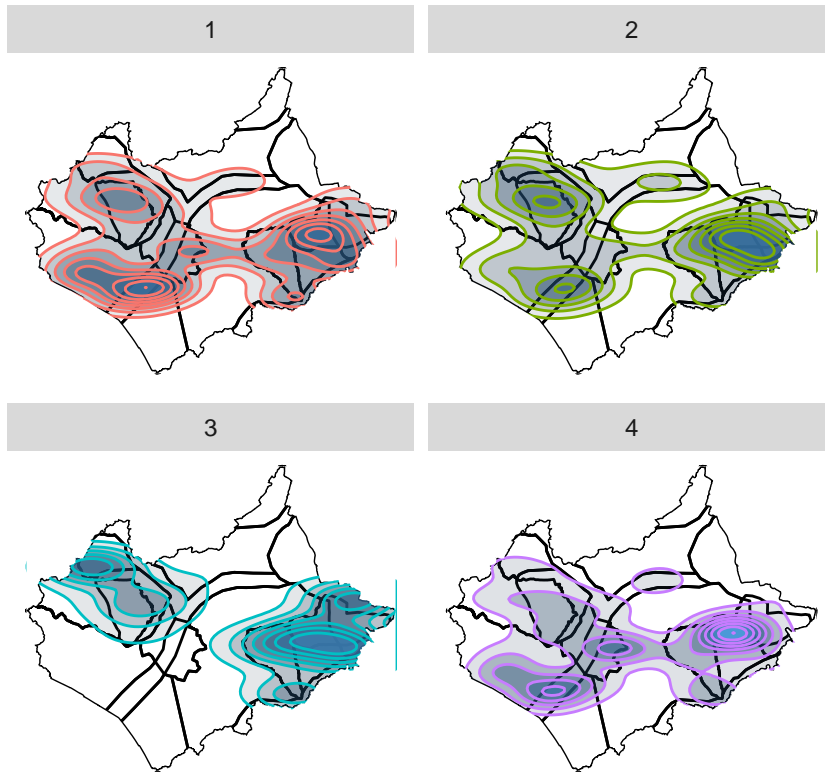


Plot by Longitude



Plot by brood

All data by brood cluster

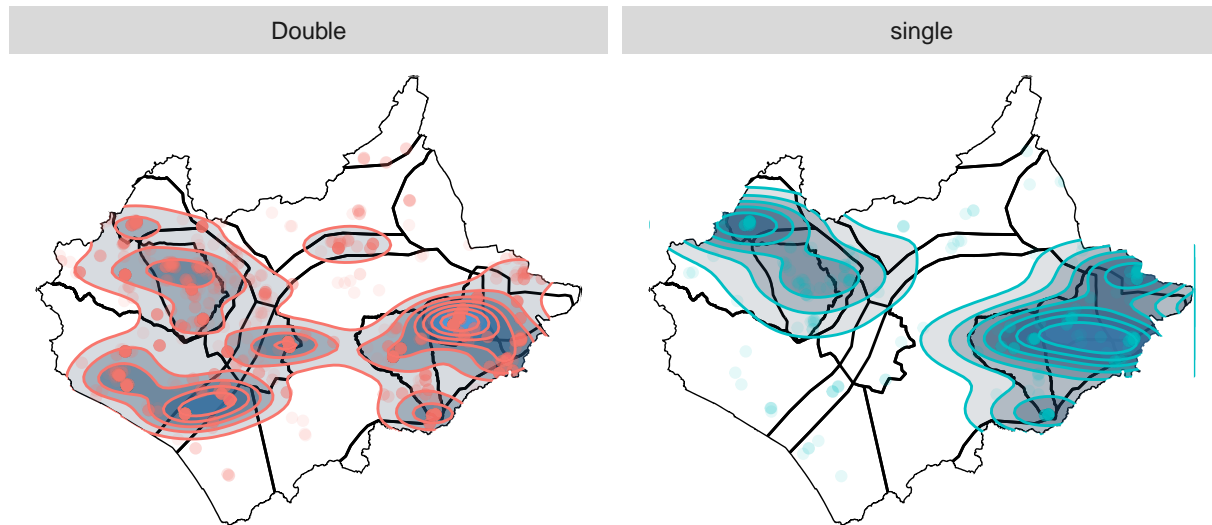


The plots above suggest that brood 1, 2 and 4 represent an single population as the observed distributions are visually identical. Brood 3, the single brooded population, is distinguished by a separate distribution as well as flight time.

Single and double broods

From the data above, broods 1,2 and 4 have been marked as double brooded and brood 3 marked as single brooded.

All data



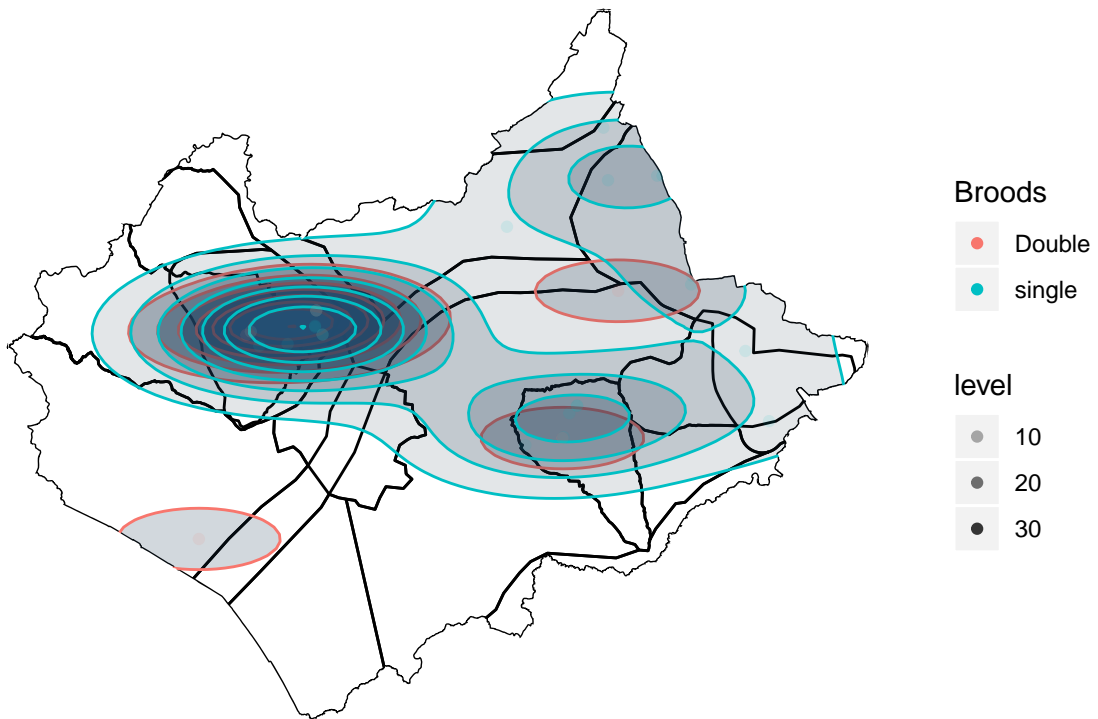
This improved plot give a better sense of the distribution of records. The contour lines mark estimates of equal observation density. The visual overlay of the two density plots is assisted by the use of transparency and coloured contour lines. Note that in the East, the county of Rutland supports both populations. There is some overlap in the extreme West. Centrally the population is almost exclusively double brooded. The picture is far more complex than expected.

Pre 1980 records

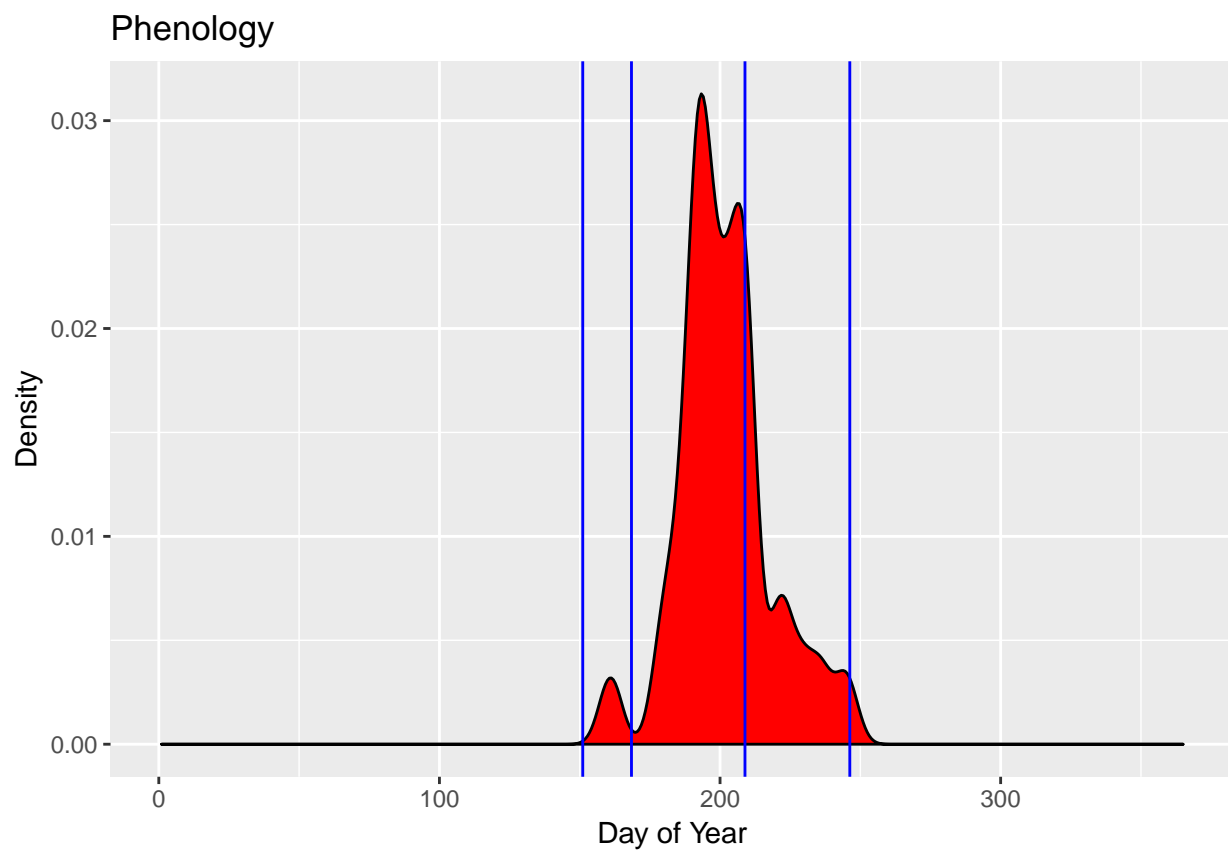
Pre 1980 the taxon was noted to be single brooded and tended to be confined to the Northern part of the county.

Pre 1980 distribution

Pre 1980 data



Pre 1980 Phenology

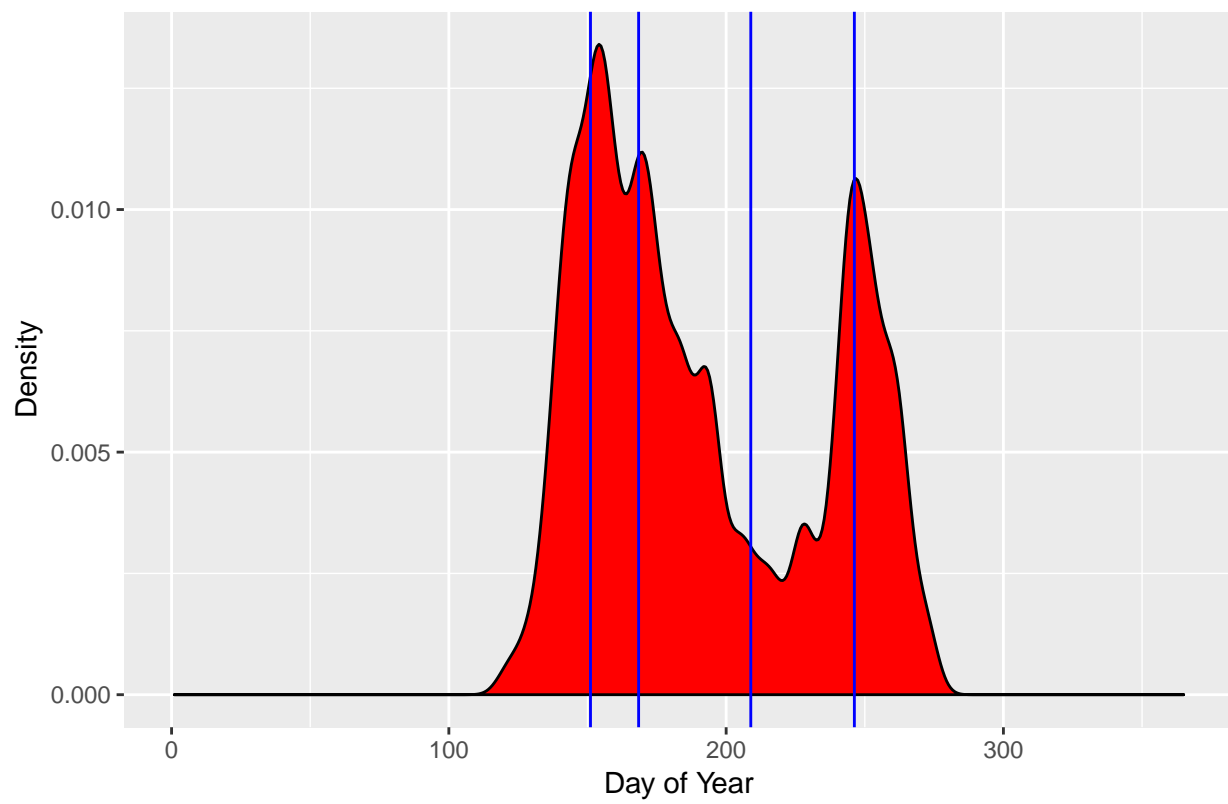


Appendix 1 Post 1980 data by area

Post 1980 Charnwood Forest



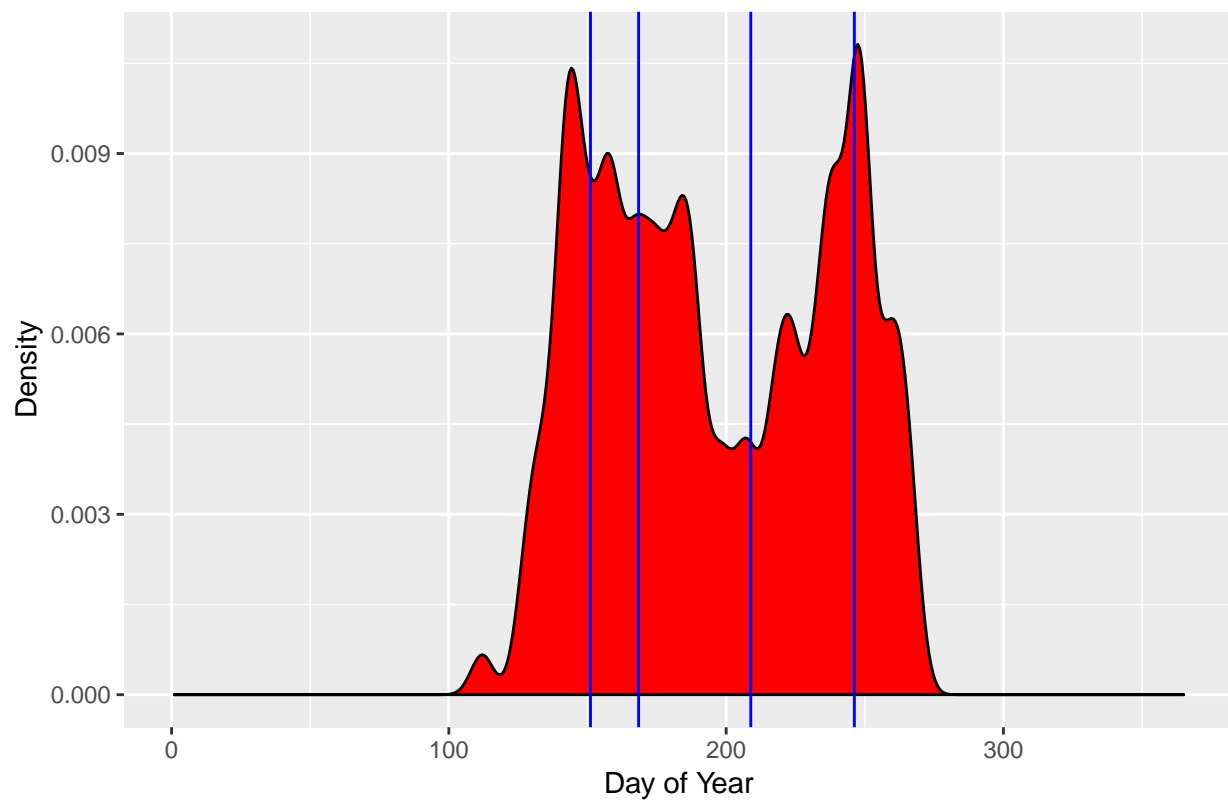
Phenology



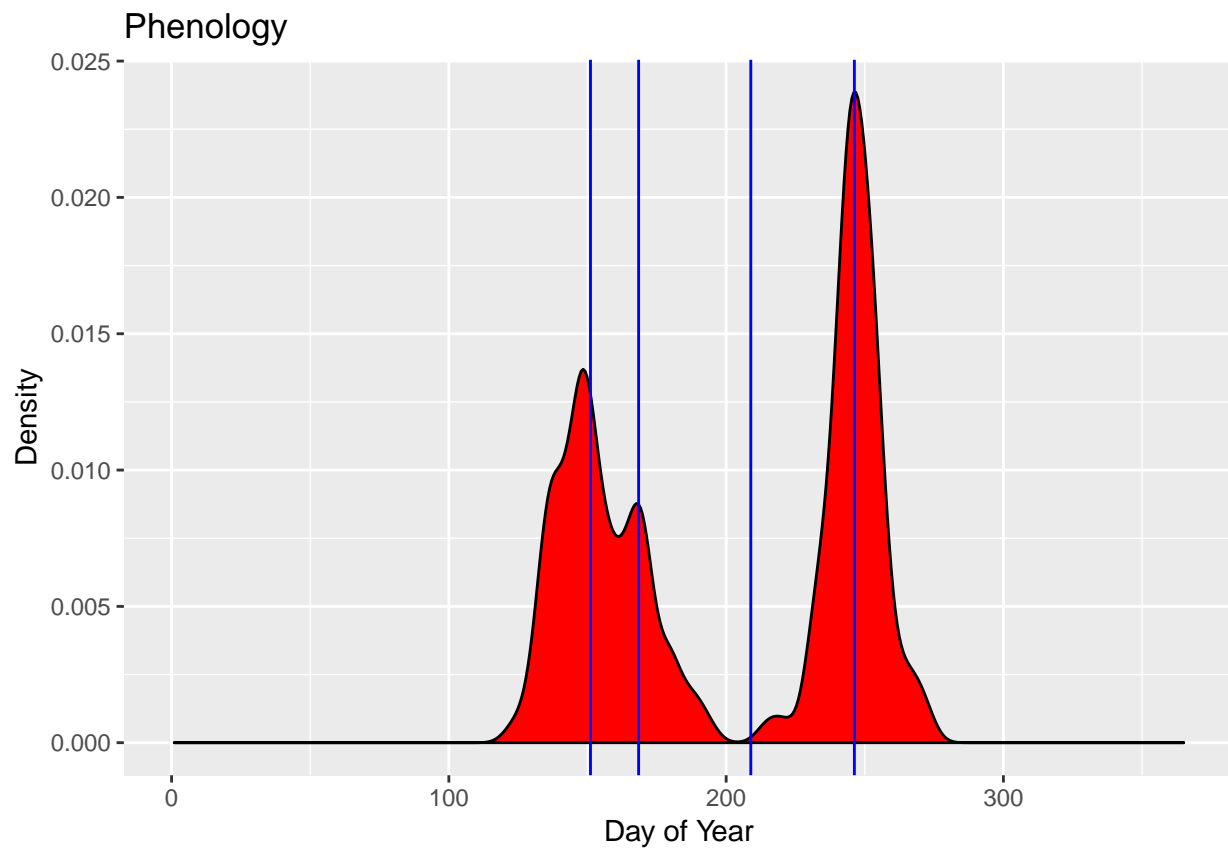
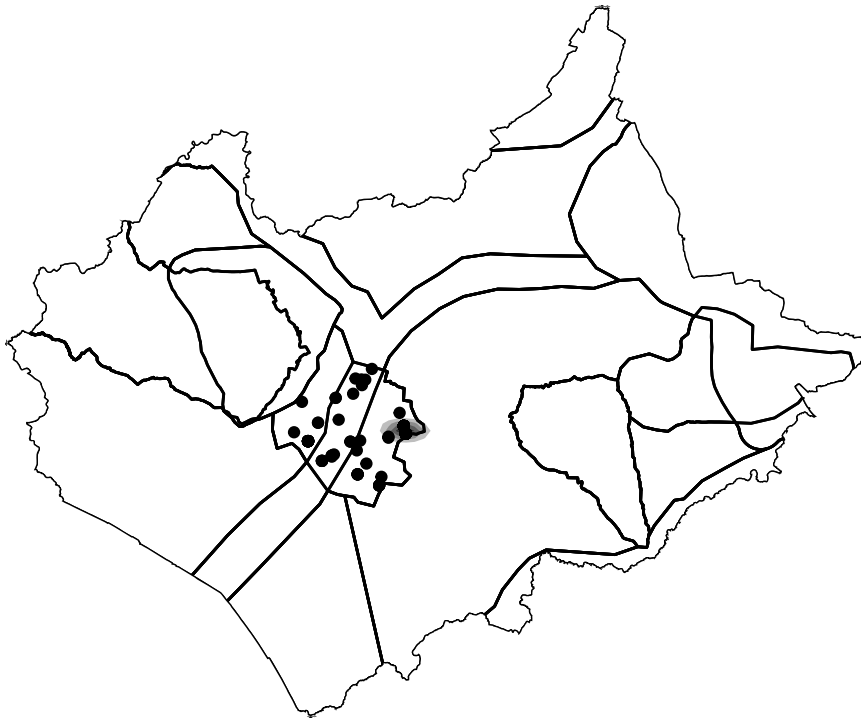
Post 1980 East Leics & Rutland Clays



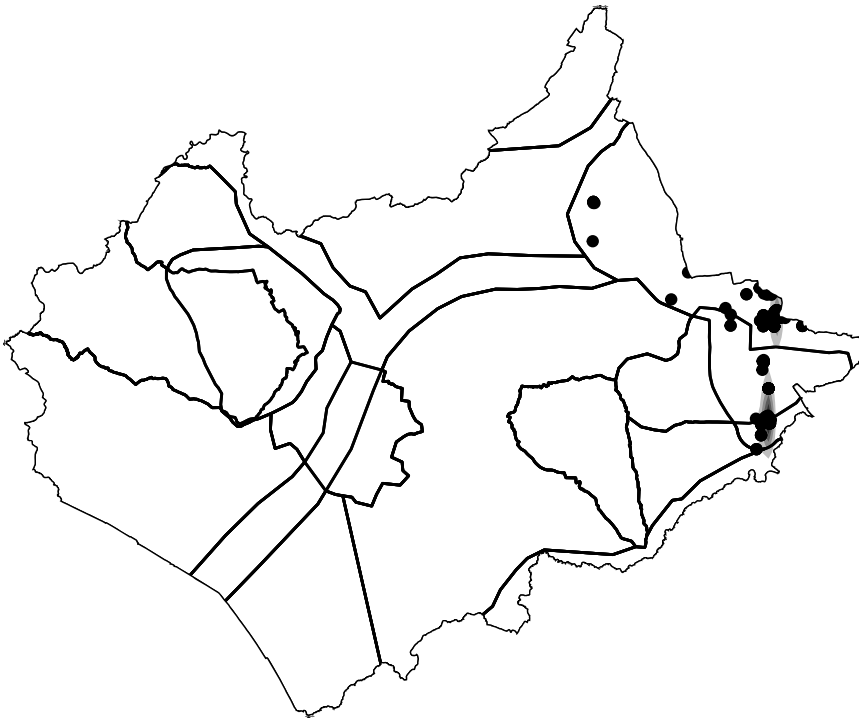
Phenology



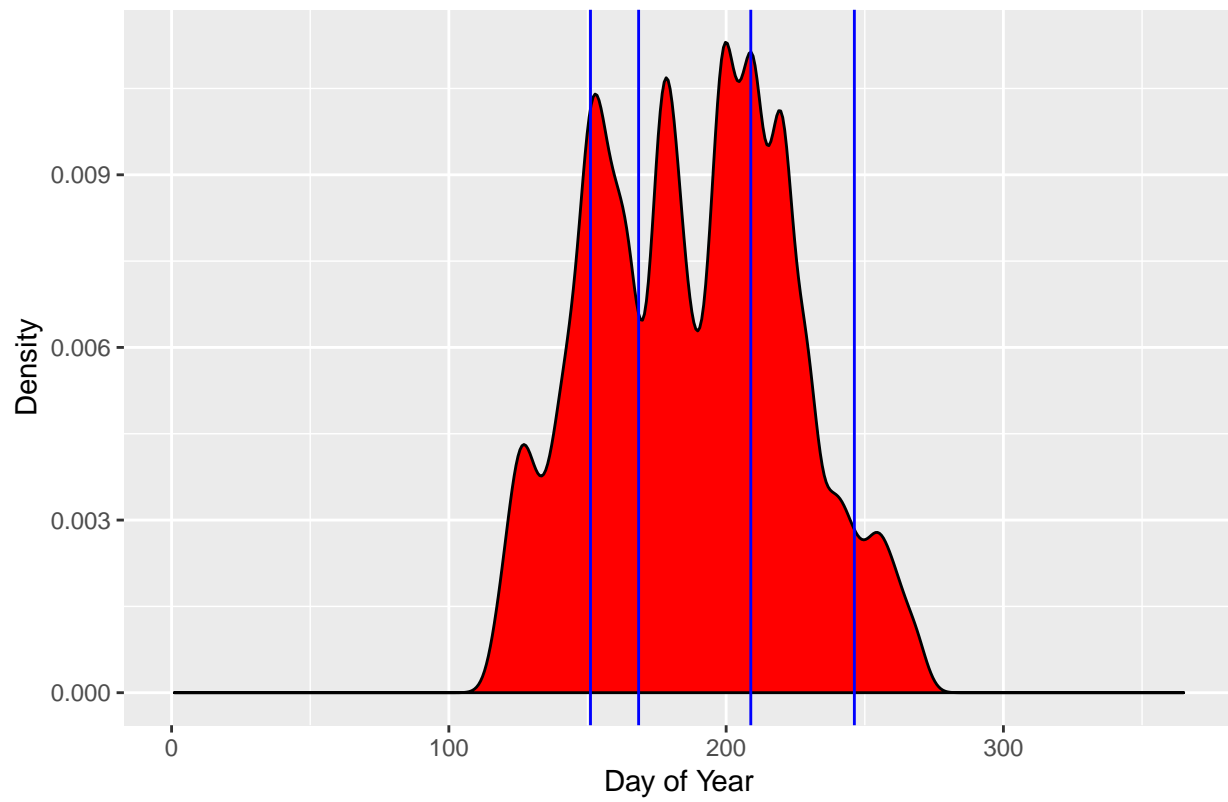
Post 1980 Greater Leicester



Post 1980 Jurassic Limestone



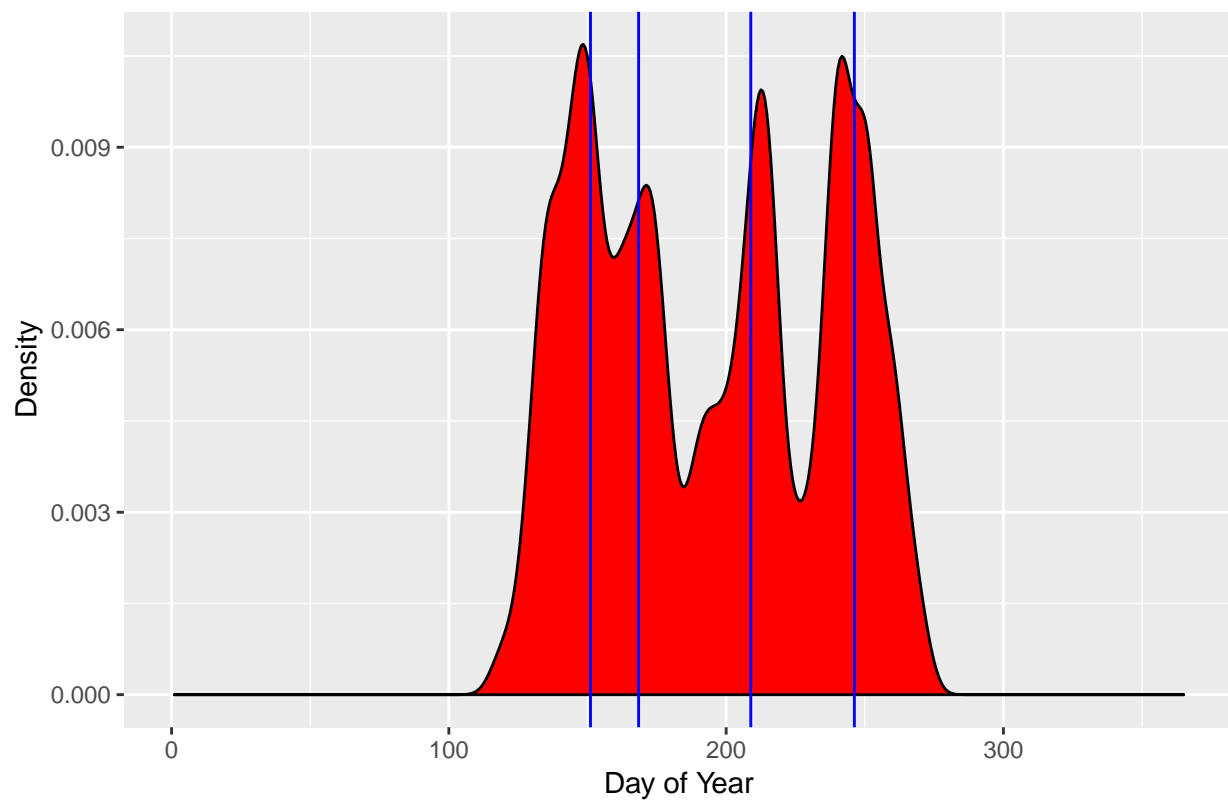
Phenology



Post 1980 Leighfield Forest



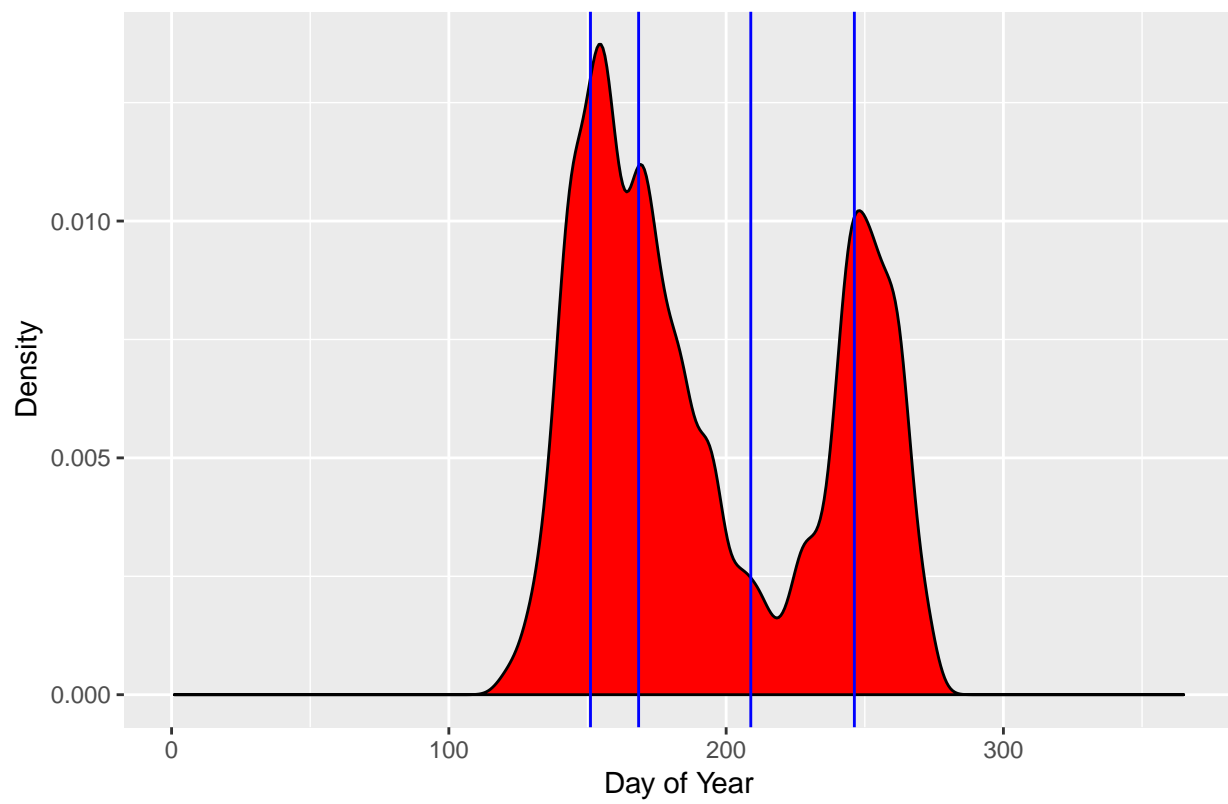
Phenology



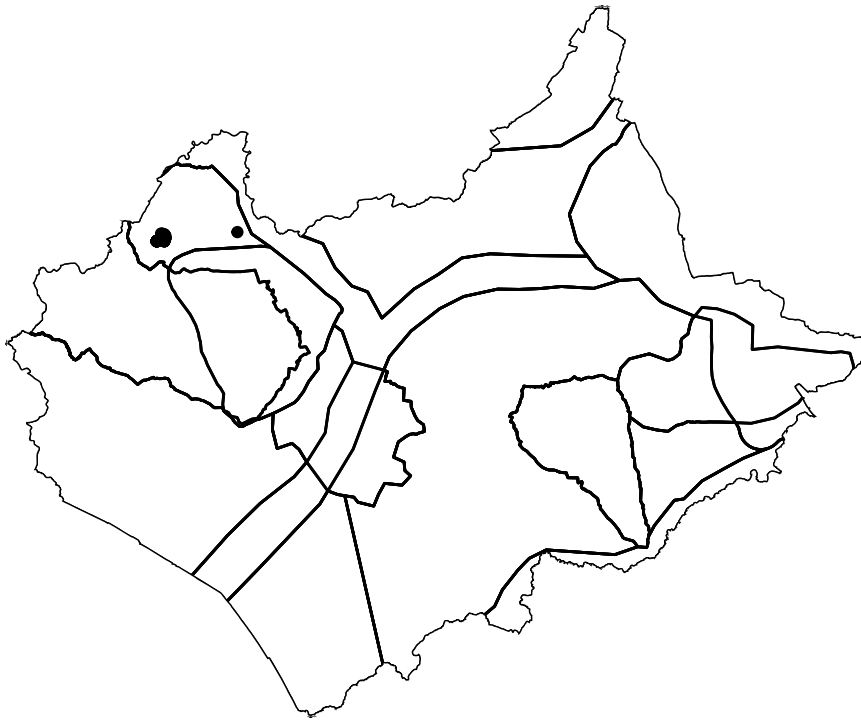
Post 1980 National Forest



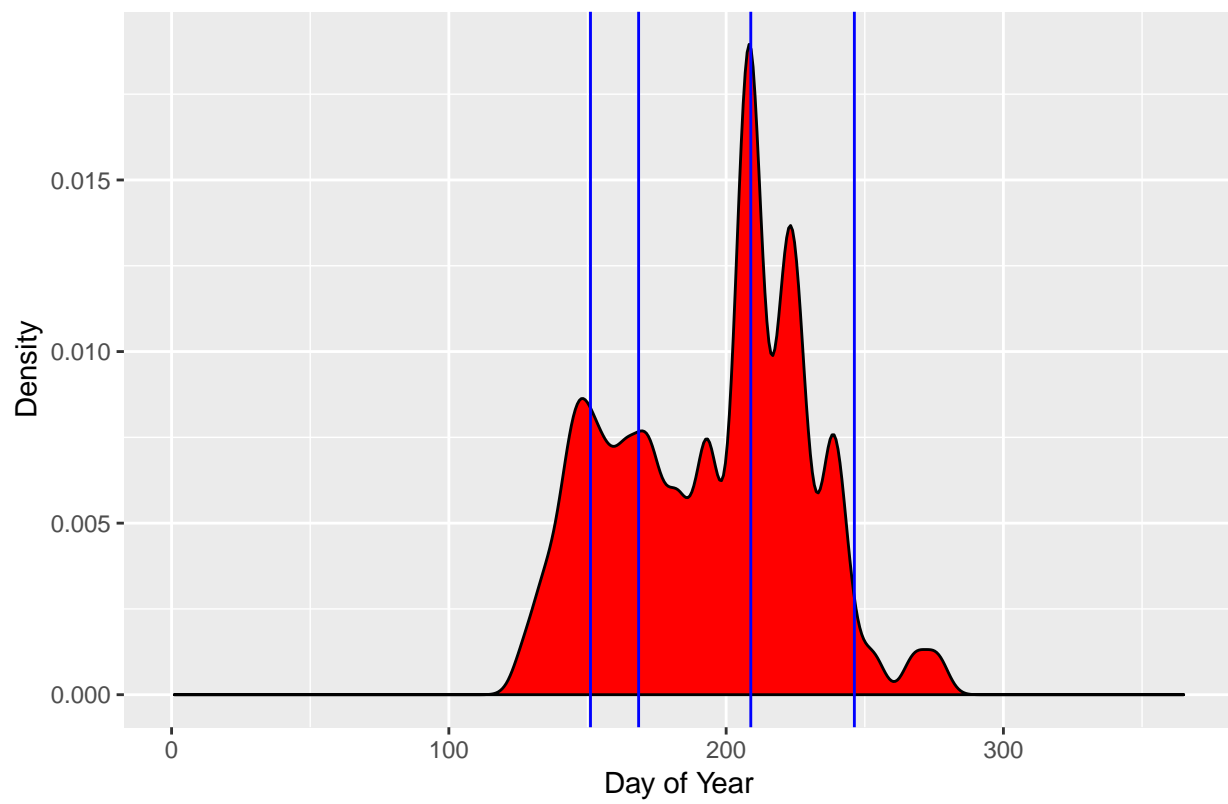
Phenology



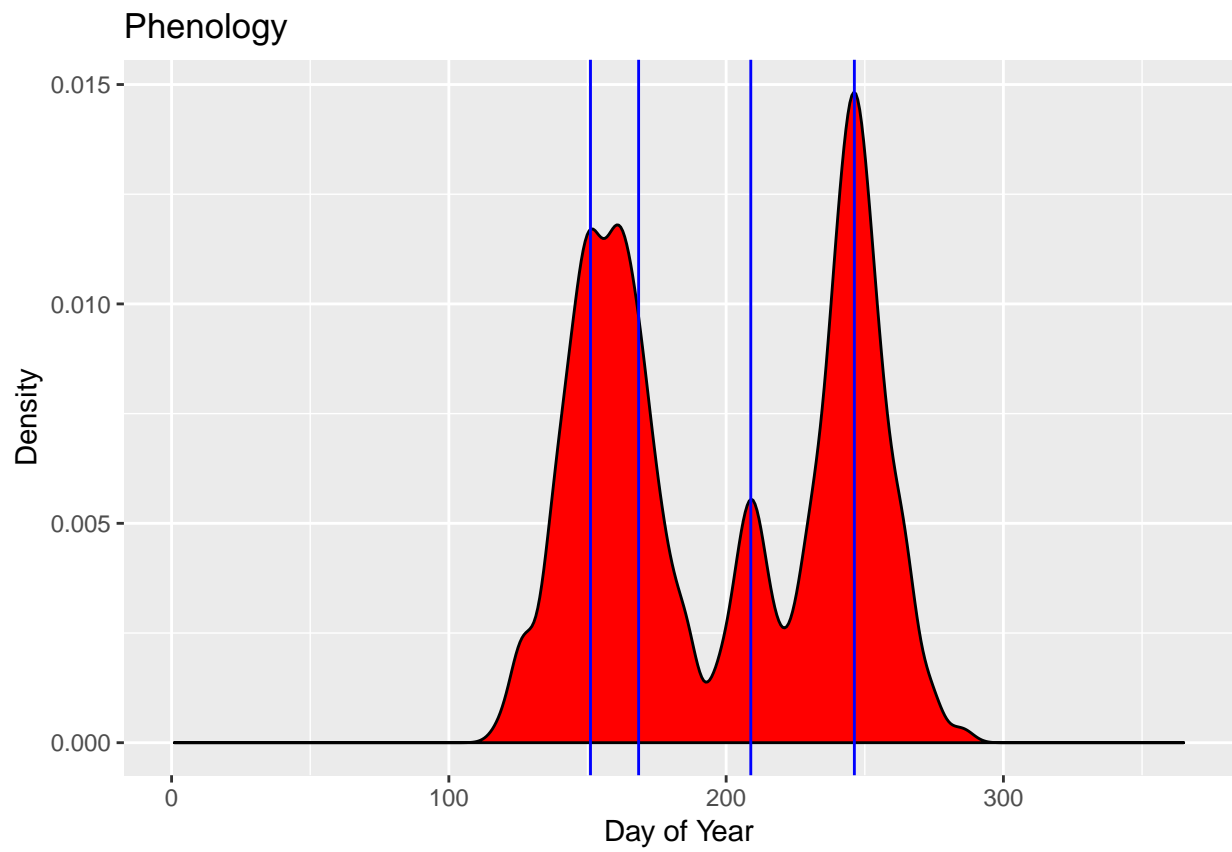
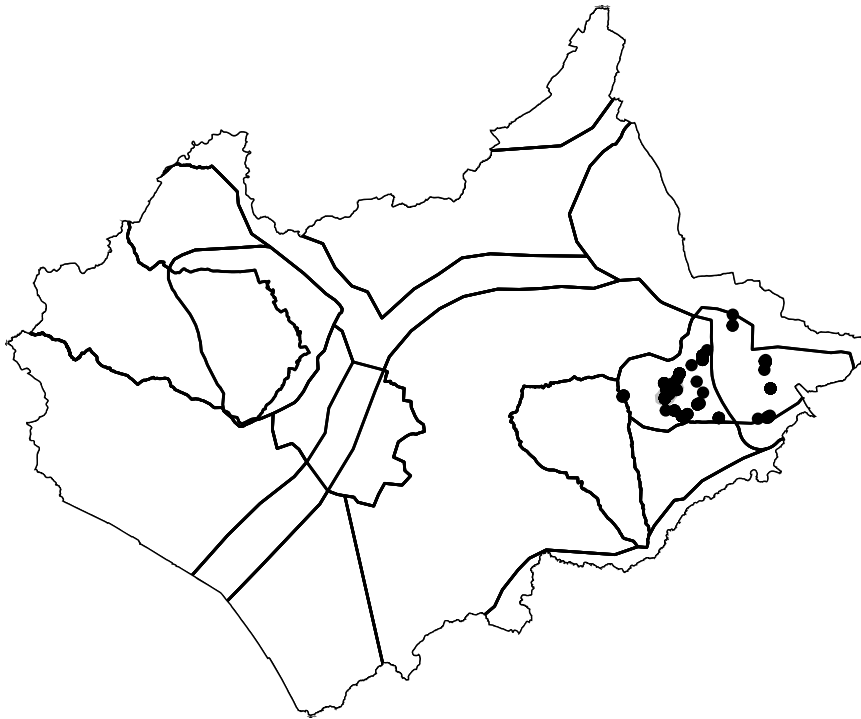
Post 1980 North West Leics Clays & Limestone



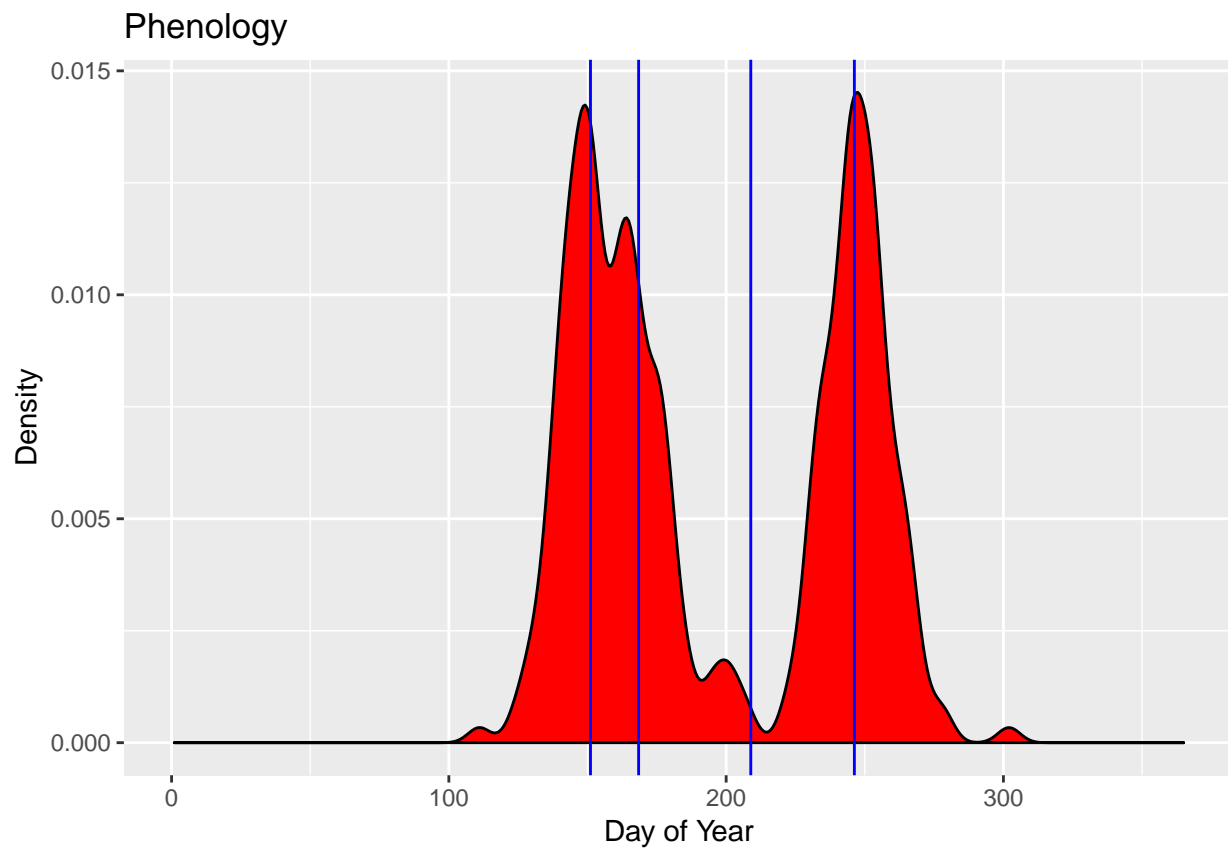
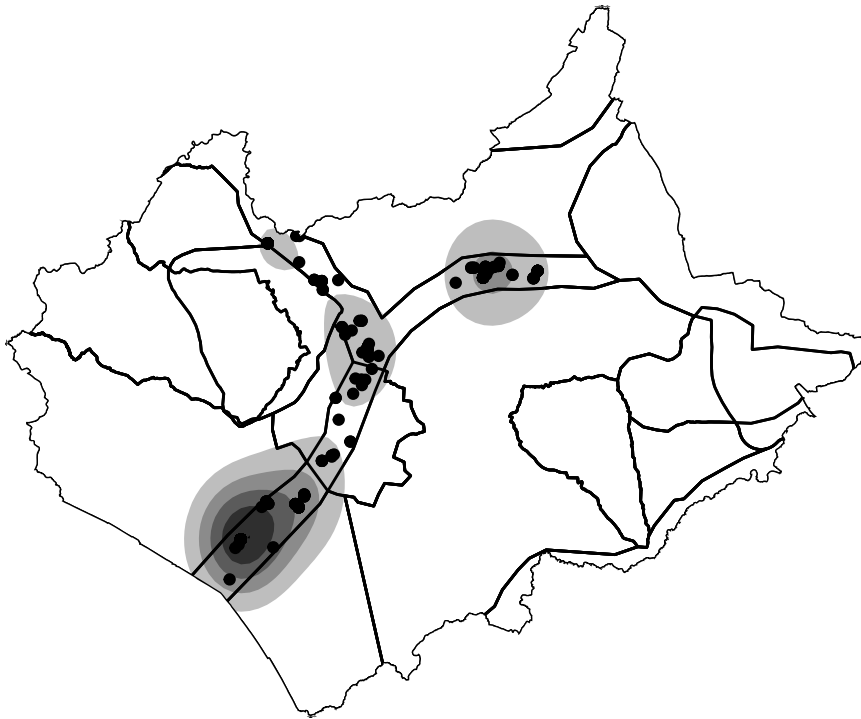
Phenology



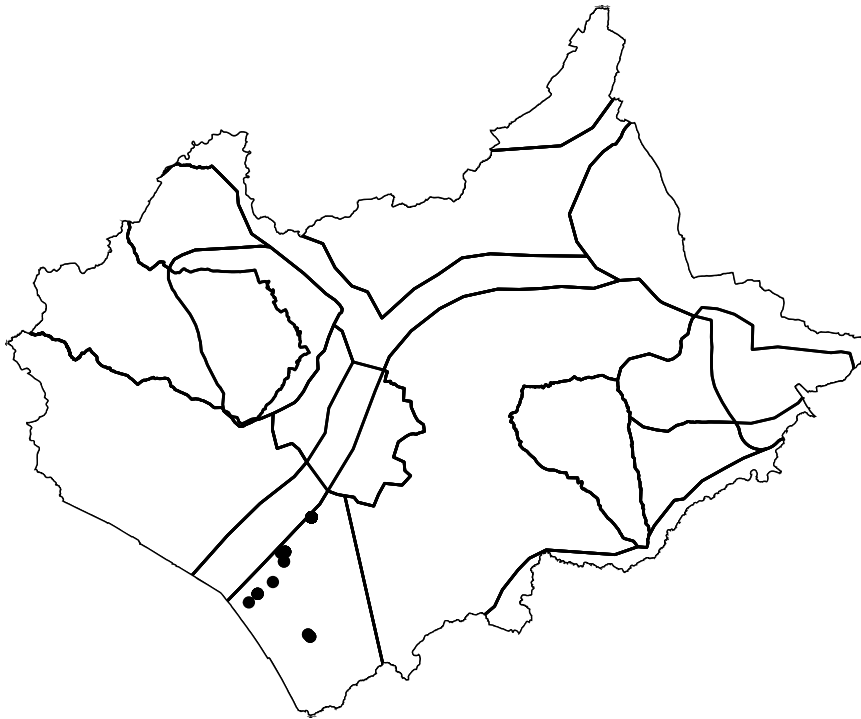
Post 1980 Rutland Water



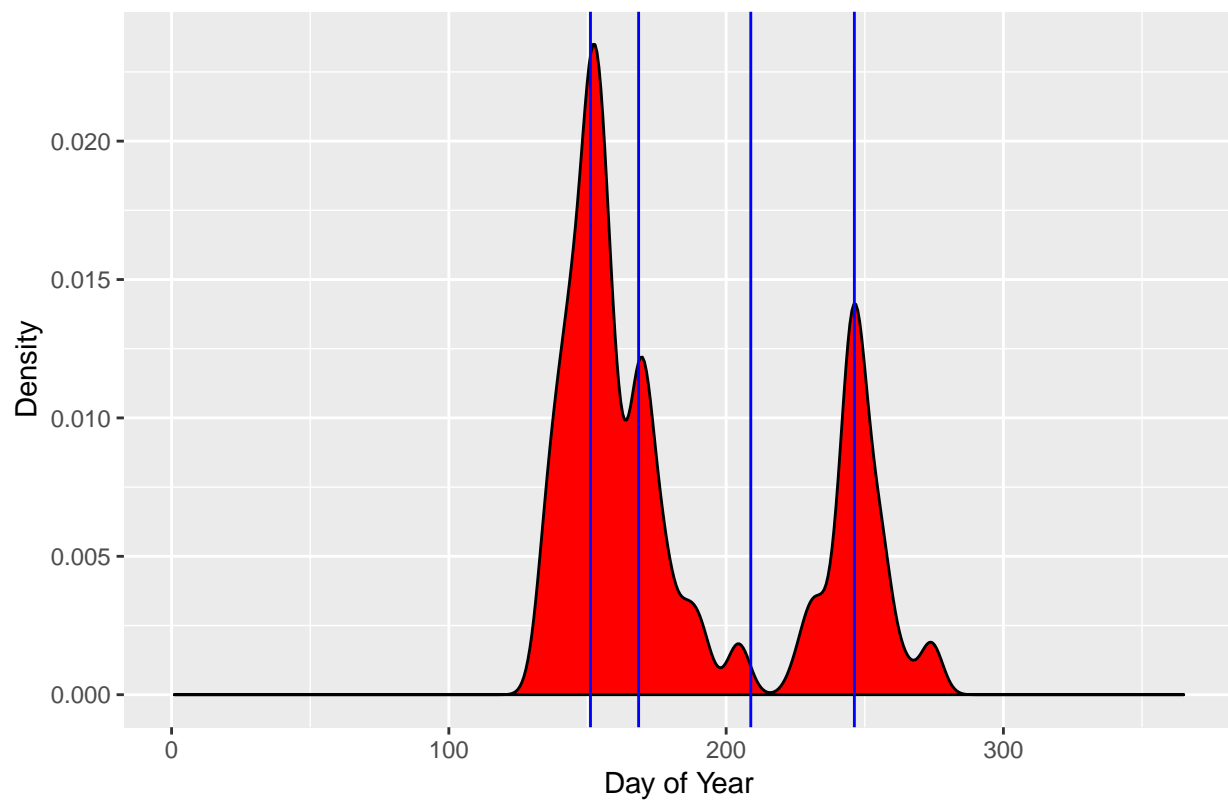
Post 1980 Soar & Wreake Valley



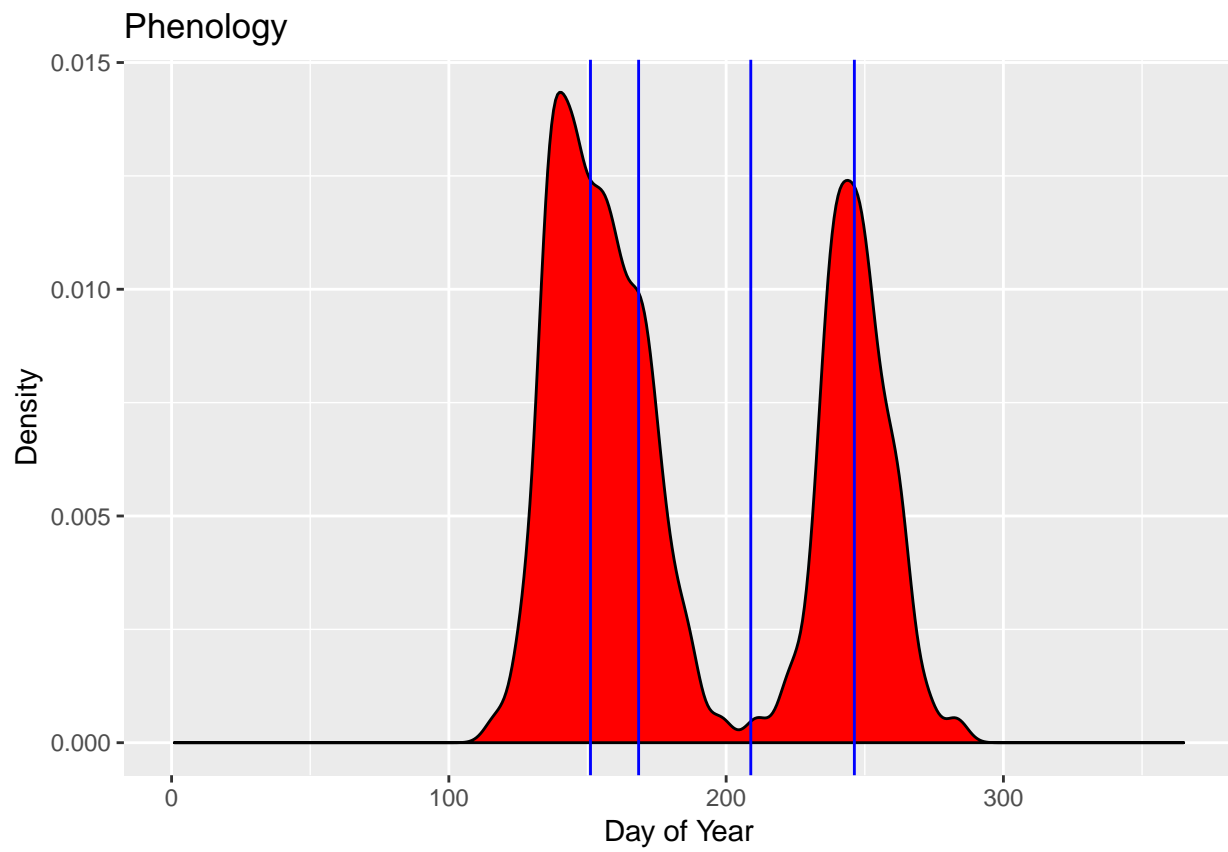
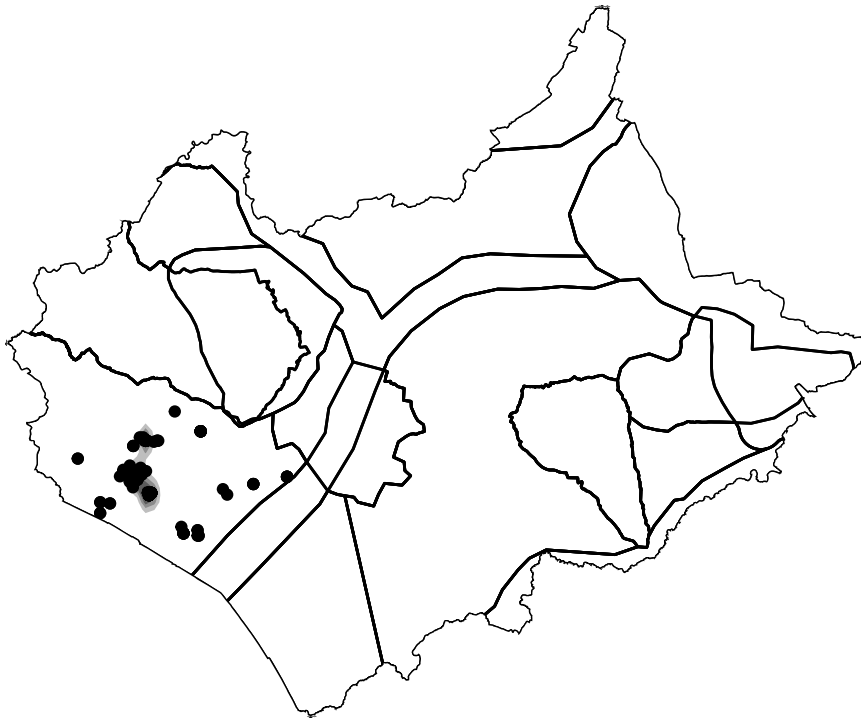
Post 1980 South Leicestershire



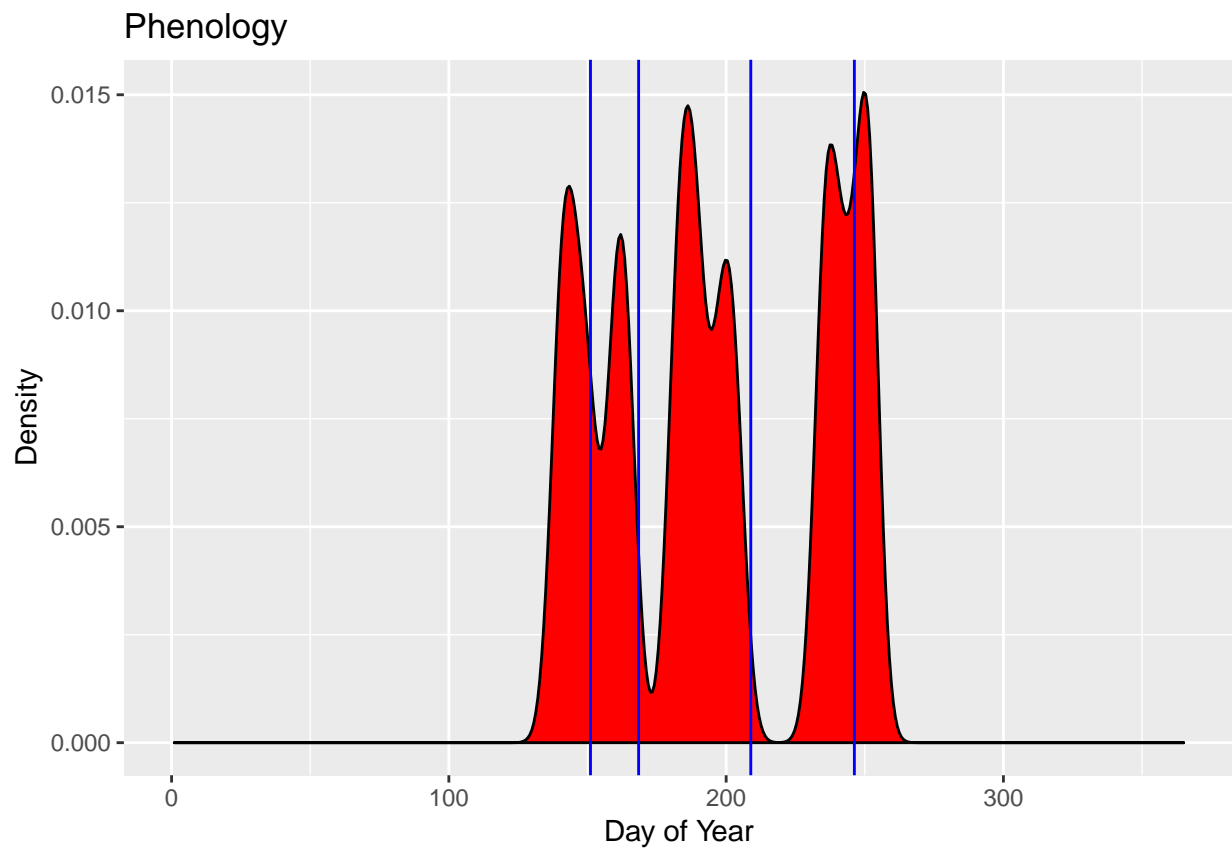
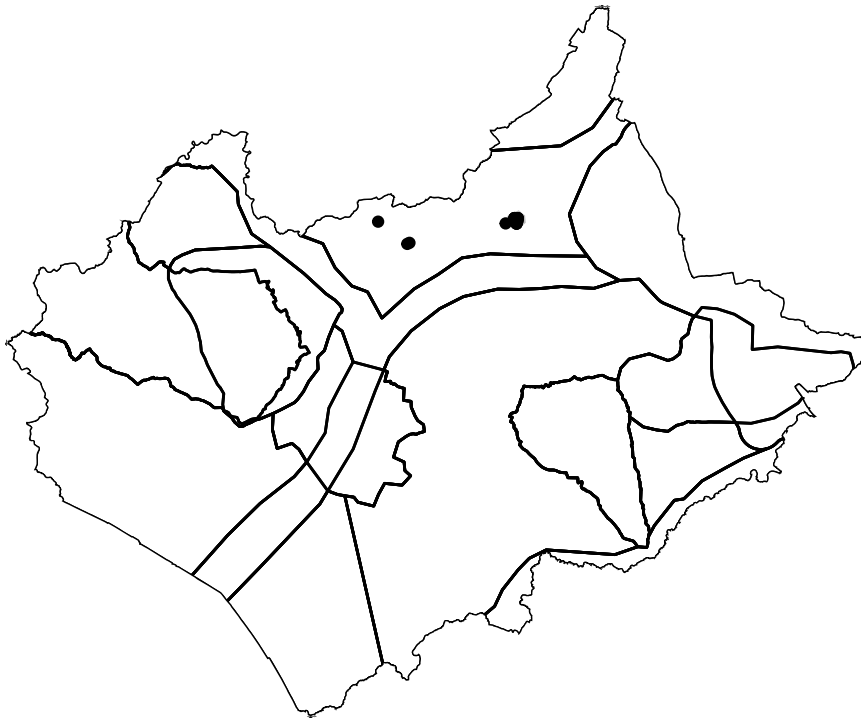
Phenology



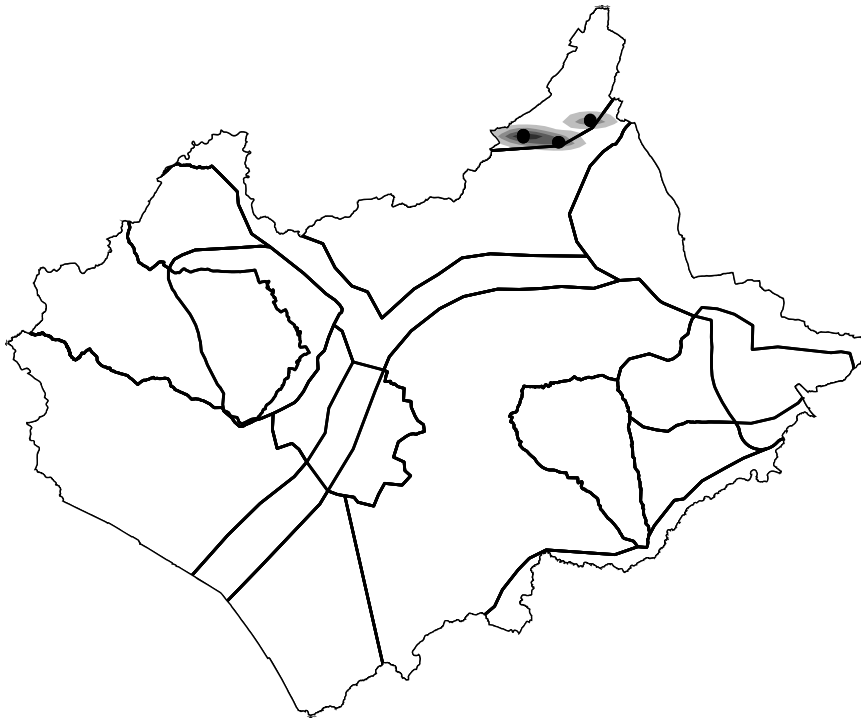
Post 1980 South West Leicestershire



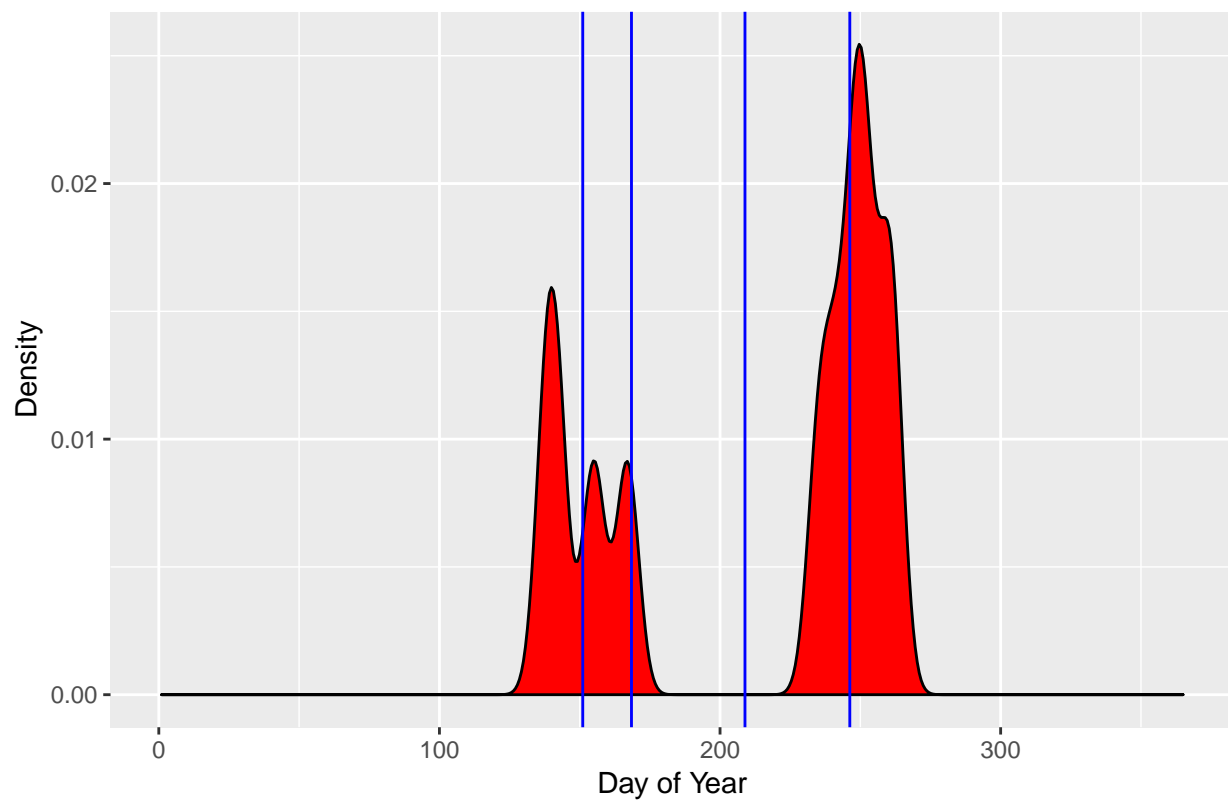
Post 1980 The Wolds



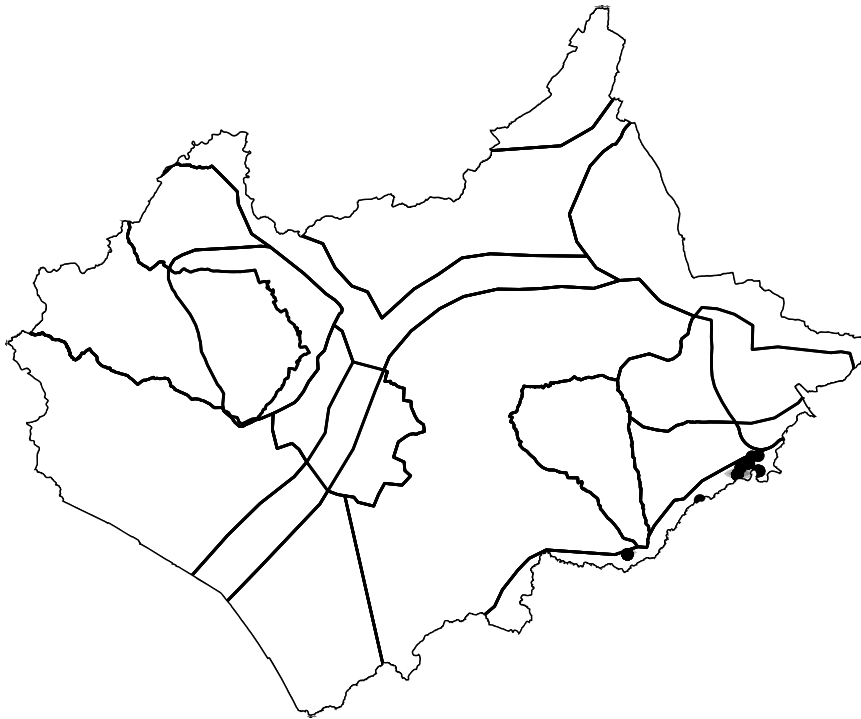
Post 1980 Vale of Belvoir



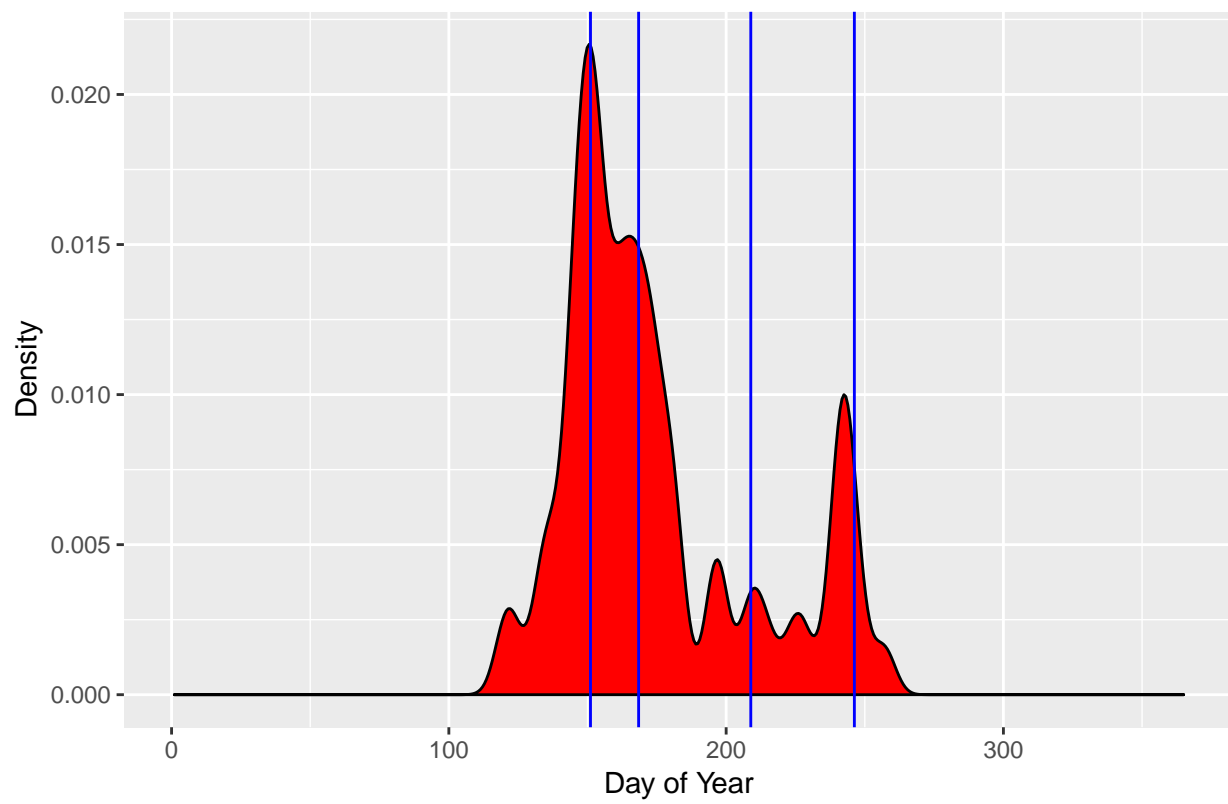
Phenology



Post 1980 Welland Valley



Phenology



Plot by Groups

Post 1985 data, all groups

