# New Orleans Tire Survey

## Mark Myer

Project setup & data import

Thursday, November 21, 2019

8:26 AM

# Objective:

Import data into R and set up initial project folder/file structure.

# Action Taken:

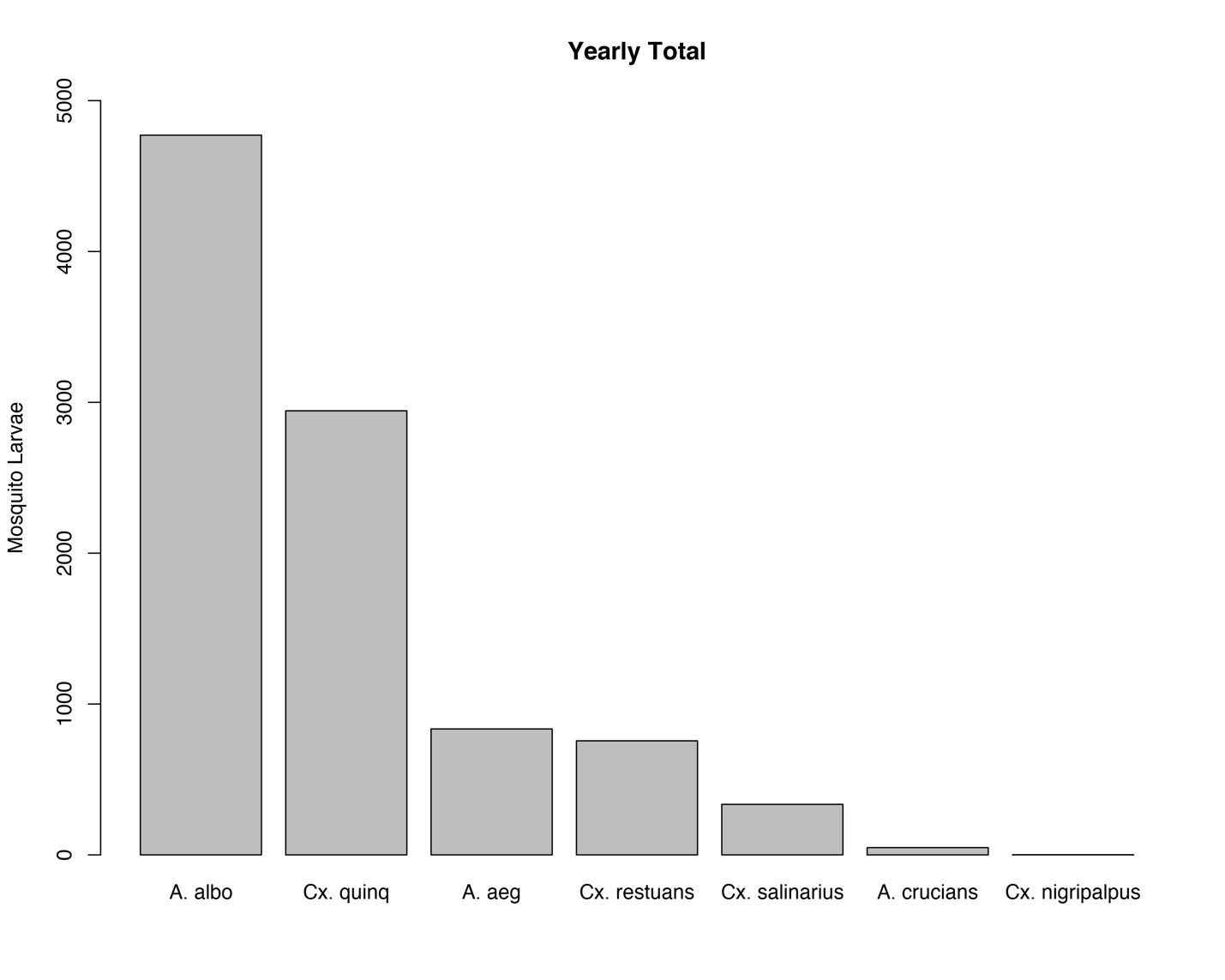
An Rproj project was created in the folder on the Q drive to hold code and data from this project. An initial piece of code was written to import and clean the data.

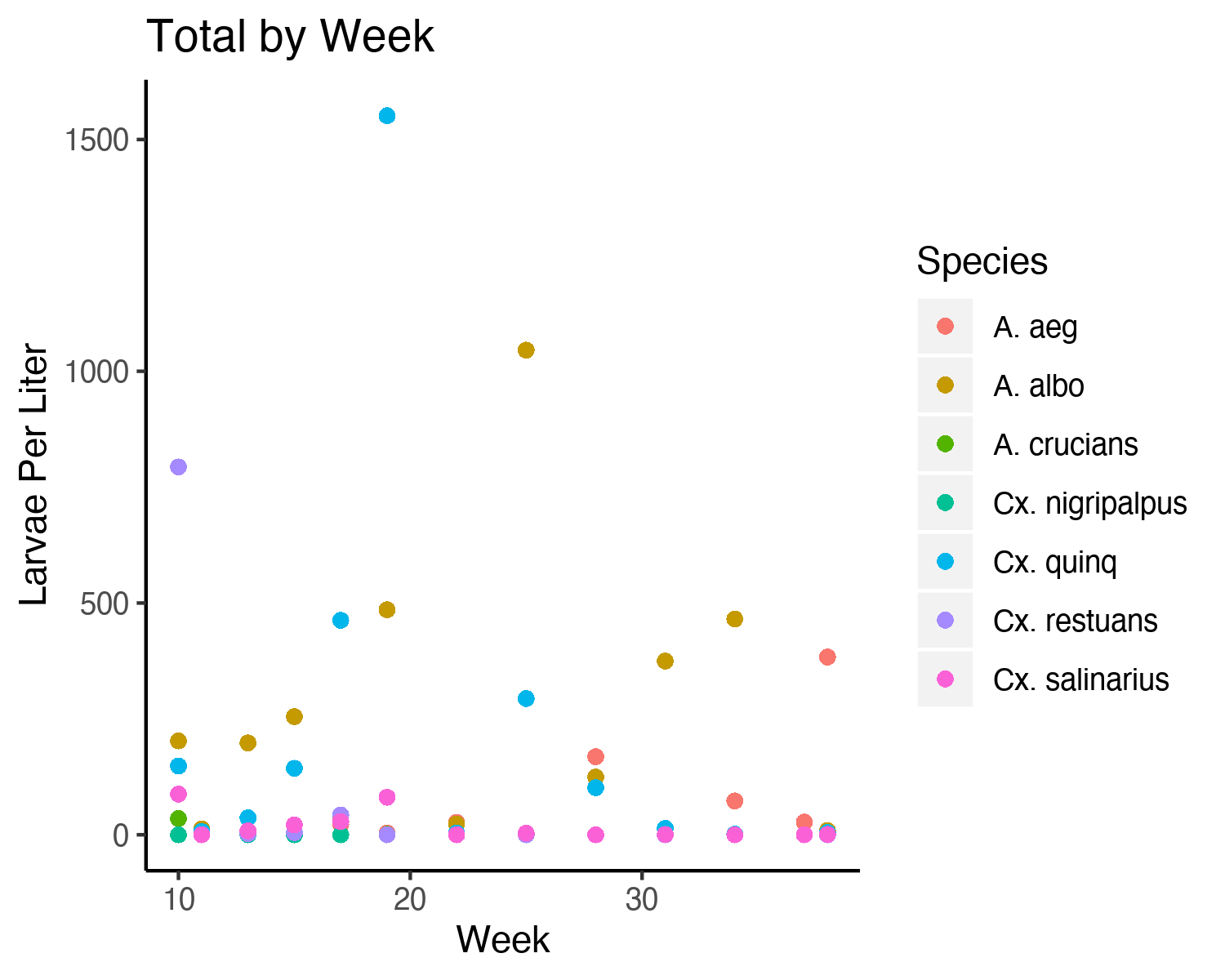
# Result/Product/Conclusion

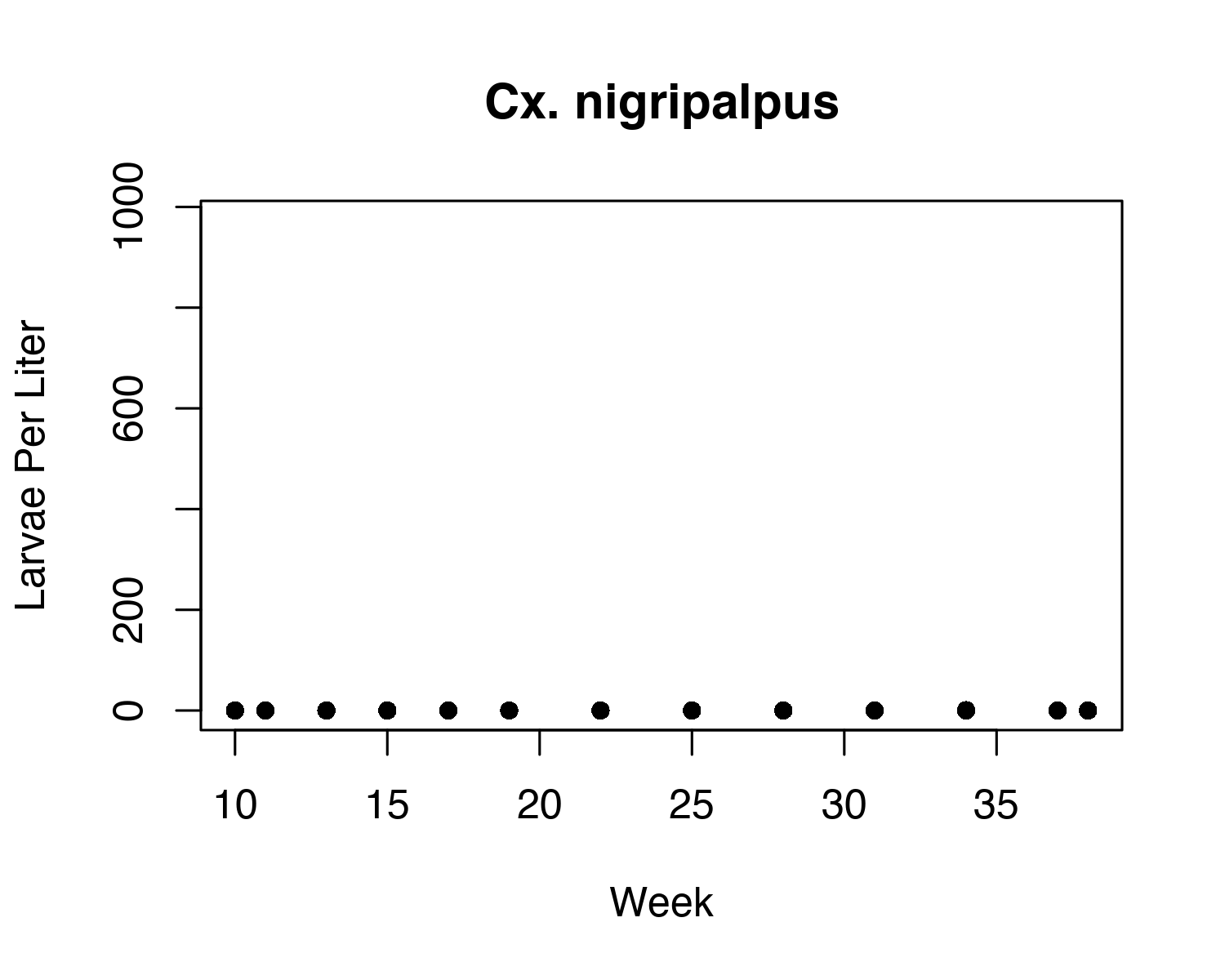
Errors and typos were identified using R and corrected in Excel.

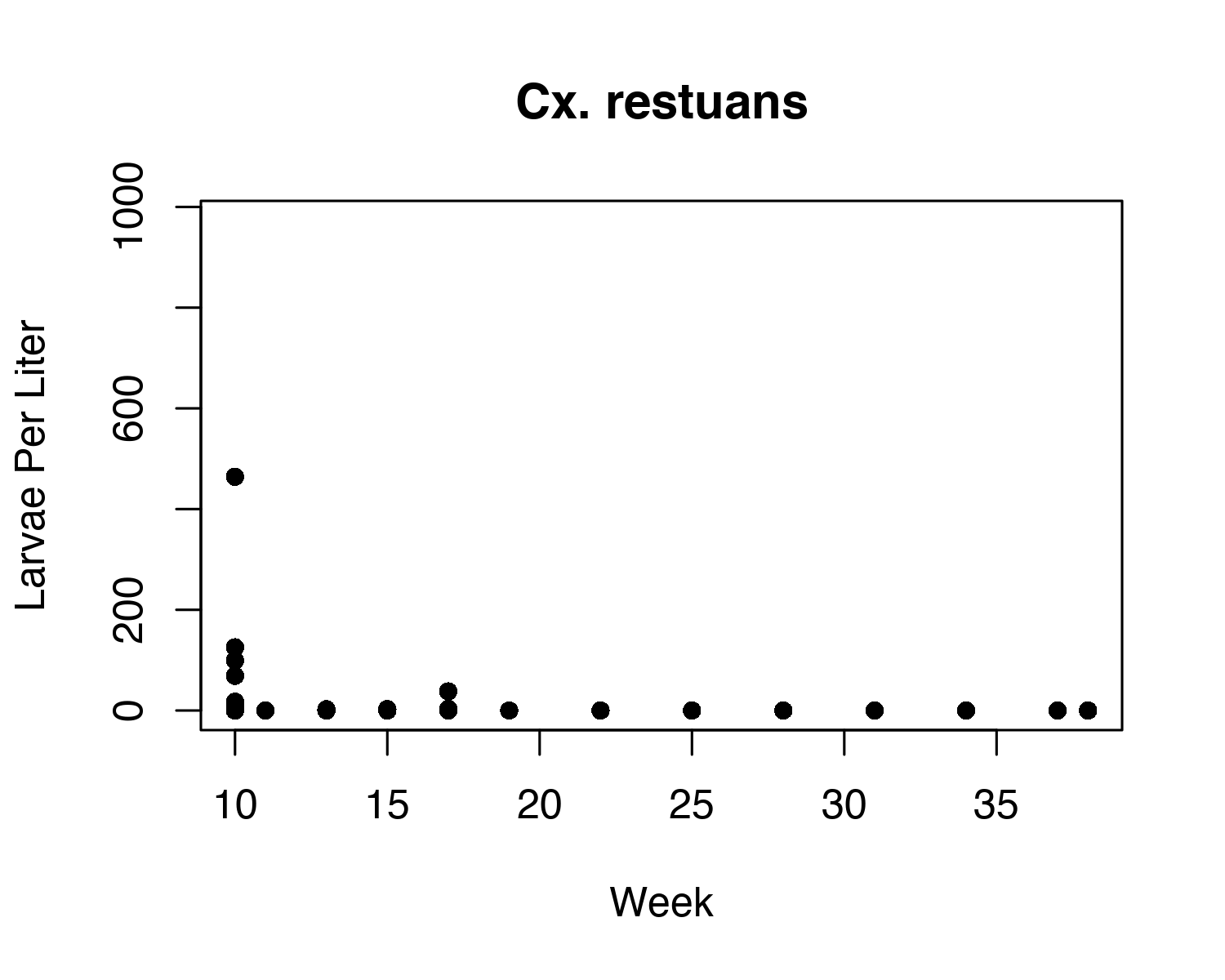
Data visualizations were created to check the distribution of mosquito larvae captures among species.

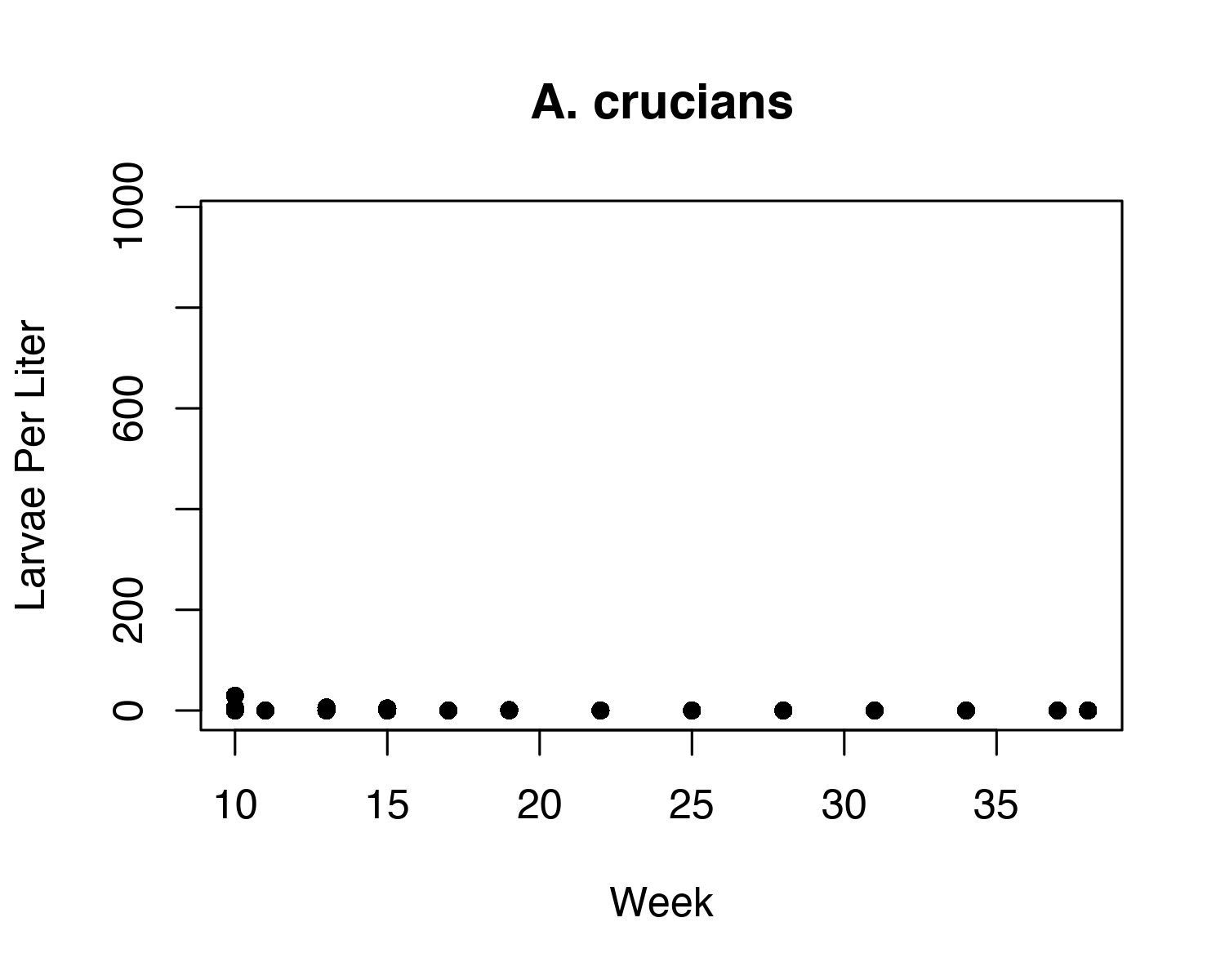
They revealed that the greatest number of larvae were Ae. albopictus, followed by Cx. quinquefasciatus. As a general trend, more quinqs were found before week 25, while more albos were found after.

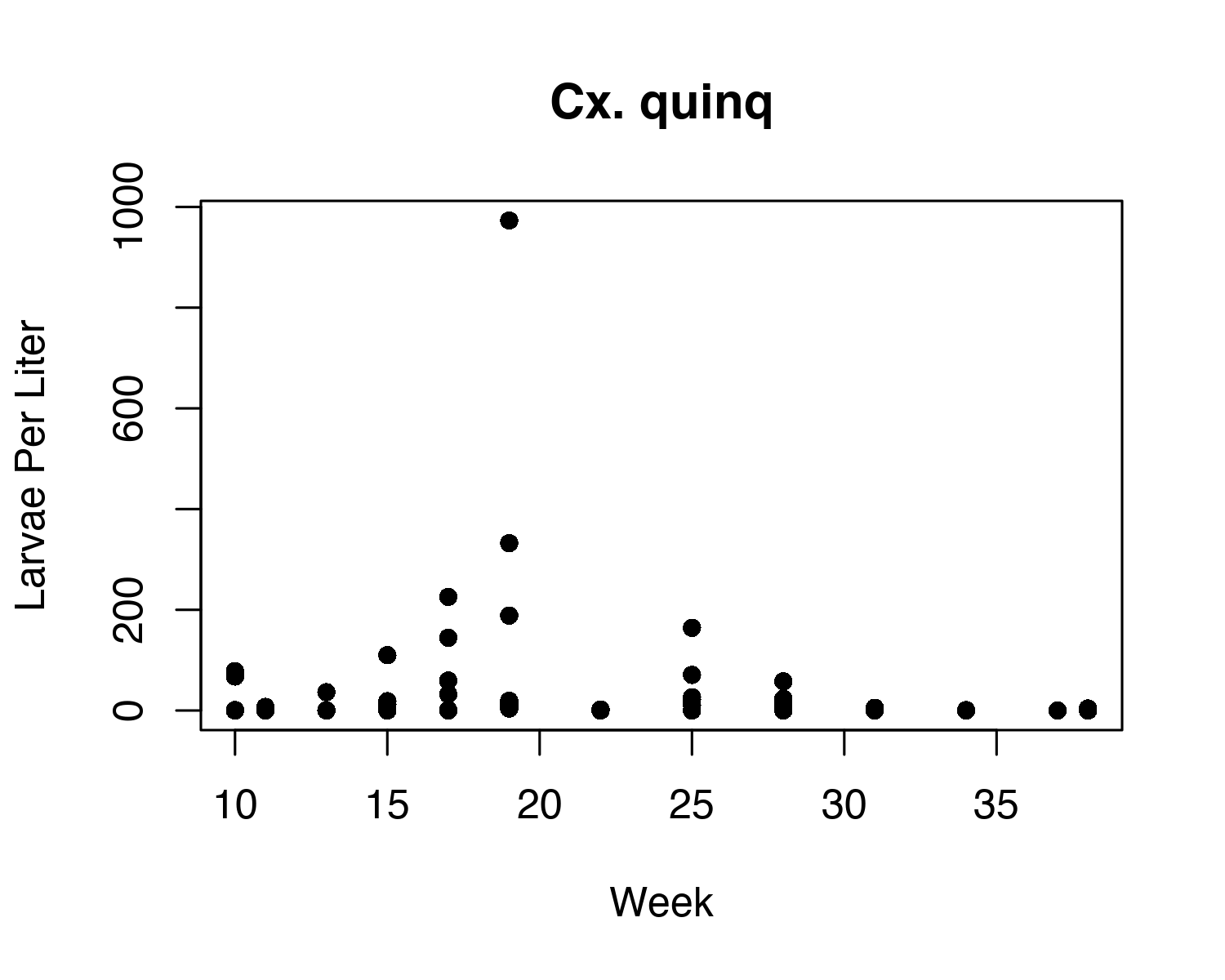


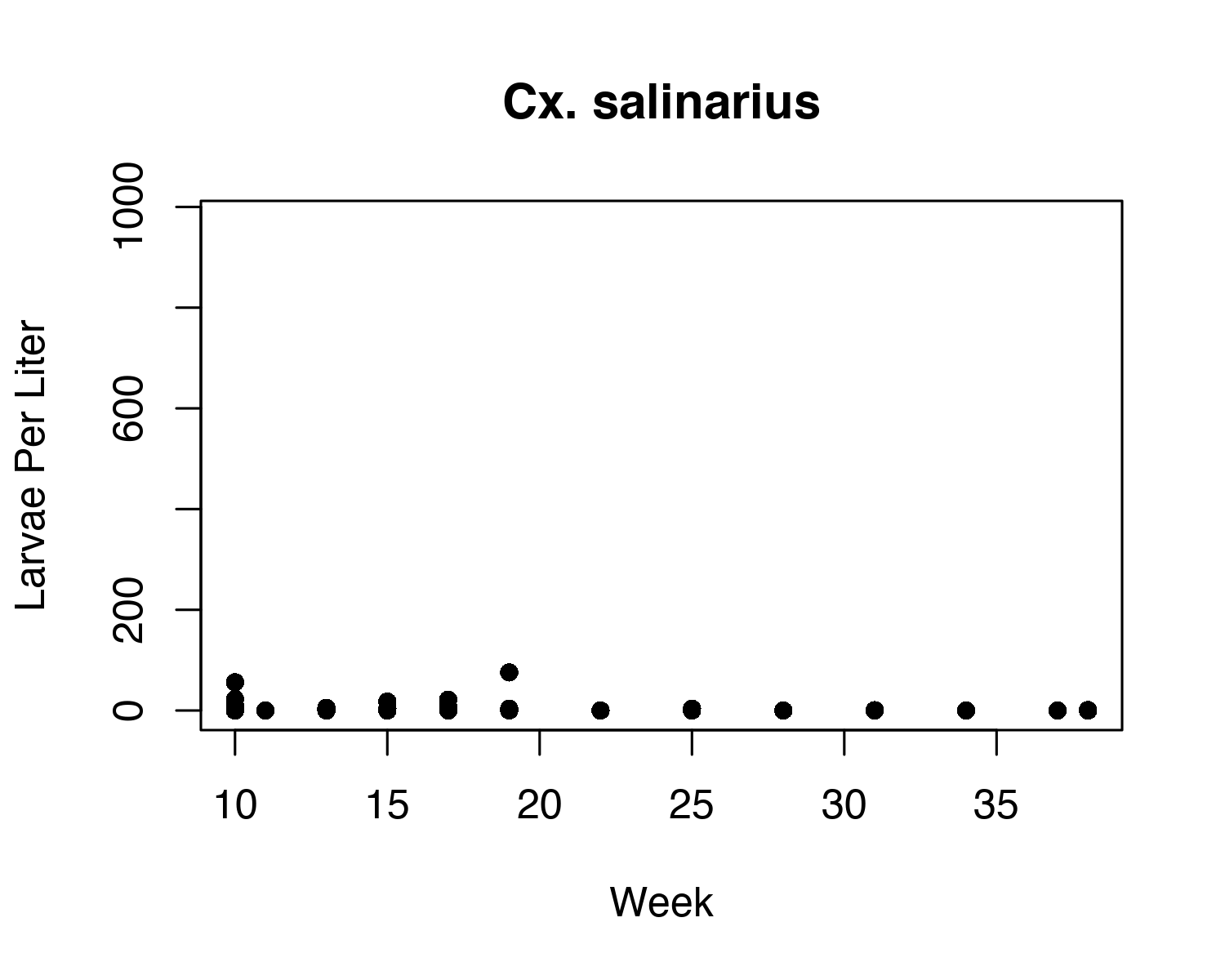


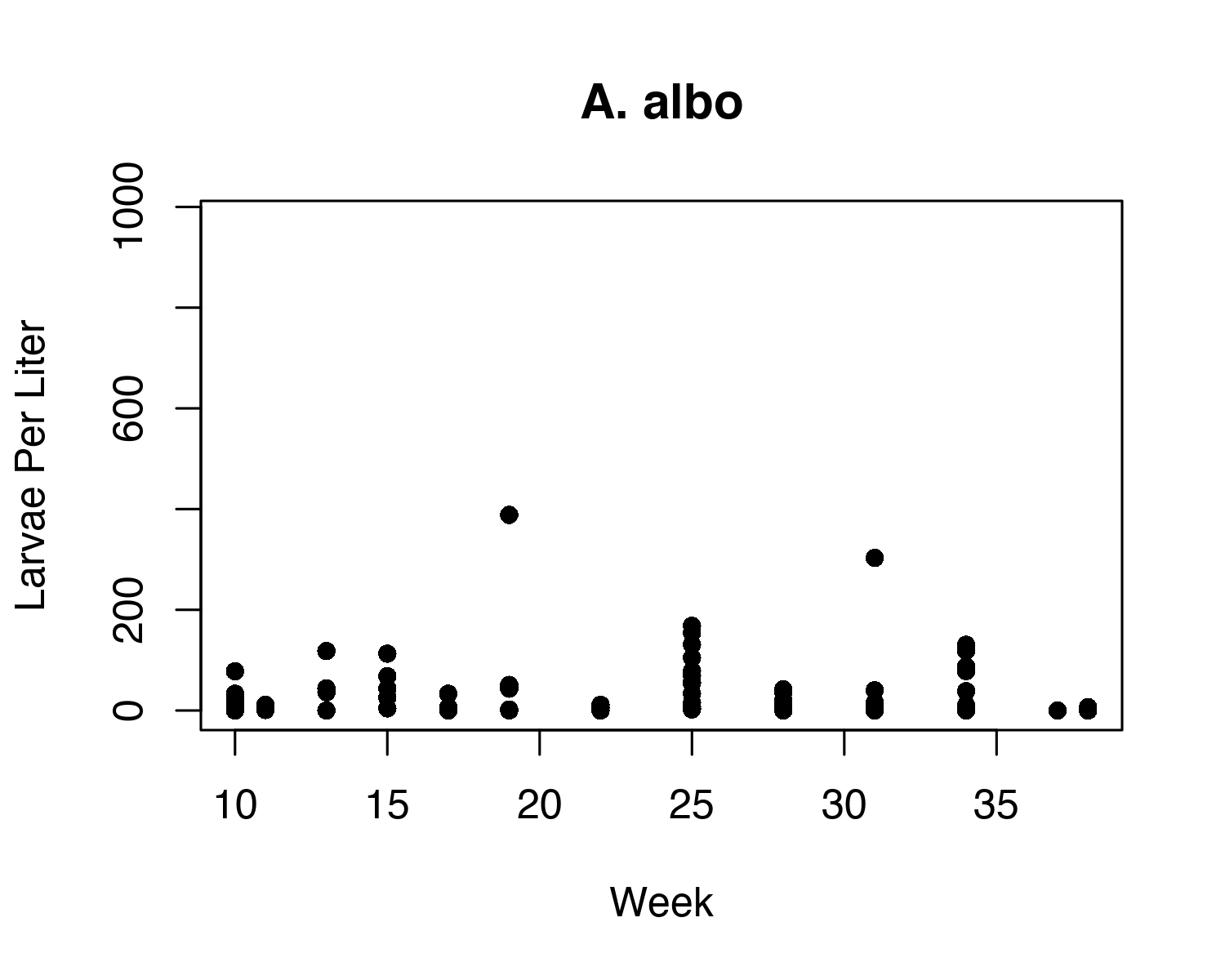


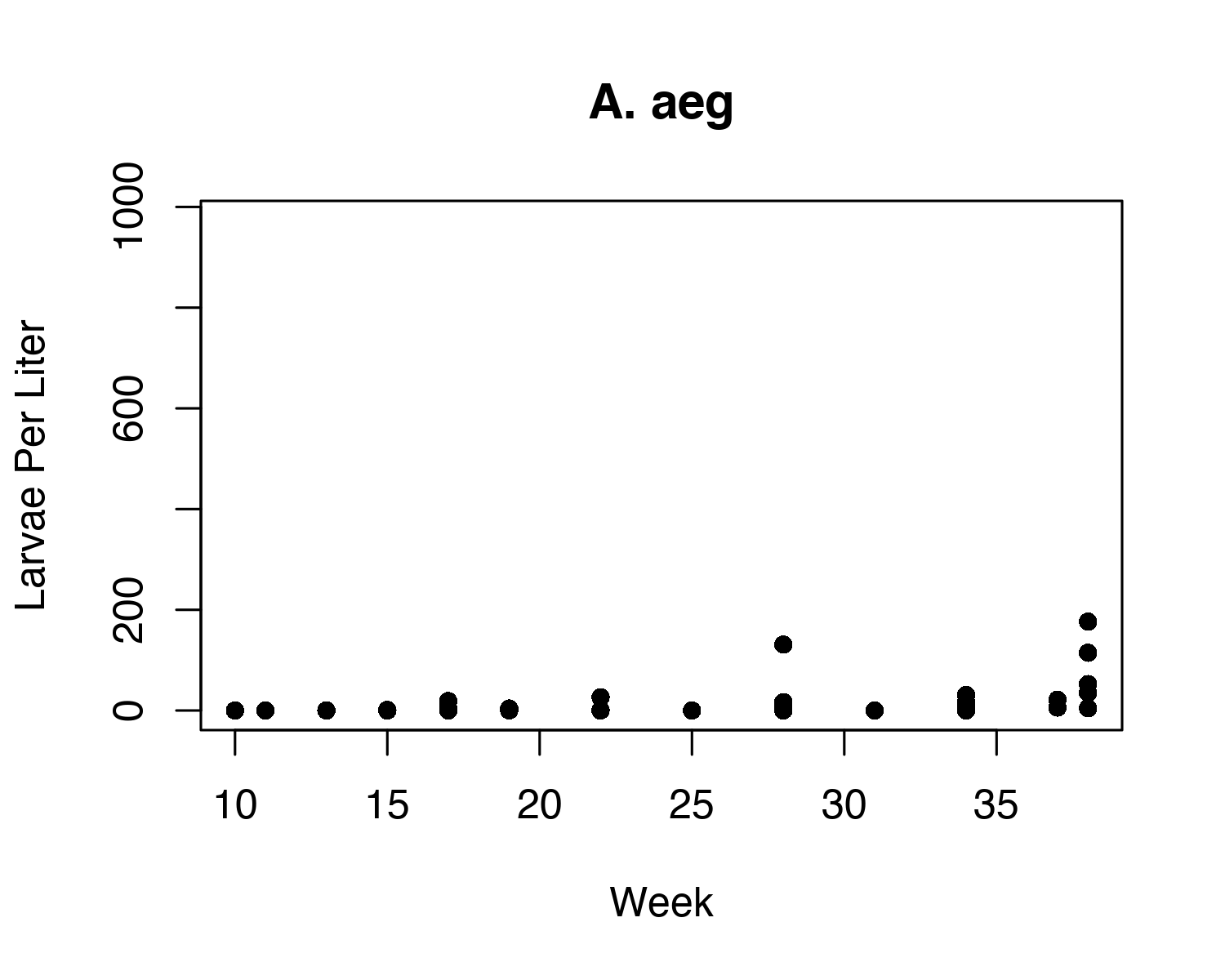












Correlation analysis

Friday, December 13, 2019

1:07 PM

**Objective:**

Check correlations between IVs and DVs for each mosquito species of interest to begin narrowing down predictor variables.

**Action Taken:**

Correlation matrices were created for each mosquito species of interest. There were not enough individuals of Cx. nigripalpus (1 larva) or A. crucians (48 larvae) found to create a model, and those species were not analyzed.

The function findCorrelation() from the *Caret* package was used to identify pairs of correlated variables (>0.75 correlation coefficient) among the potential predictors, excluding location and time, and to remove one from each pair in order to simplify variable selection.

Bayesian variable selection using GibbsBvs() from the *BayesVarSel* package was used to sample from all possible models and then rank variables based on their probability of inclusion in the best model. For each mosquito species, the top 5 variables were selected and kept for analysis.

**Result/Product/Conclusion**

The initial datasets included 75 potential predictor variables each (including location and time).

The findCorrelation() function reduced each set of DVs to approximately 40 variables, around a 40% reduction.

The results of Bayesian variable selection are as follows:

**A. aeg**

1. Medium intensity development
2. Copepods
3. Barren land
4. Developed open space
5. Low intensity development

**A. albo**

1. Mites
2. Seeds
3. Cultivated crops
4. Triseriatus
5. Professional degree holders

**Cx. salinarius**

1. Developed open space
2. Low intensity development
3. Medium intensity development
4. Vegetation in tires
5. Chironomids

**Cx. quinq**

1. Developed open space
2. Chironomids
3. Open water
4. Amount of water in tire
5. Area in km2

**Cx. restuans**

1. Professional degree holders
2. Amount of water in tire
3. Rented unoccupied housing
4. Deciduous forest
5. Fine particulate matter