Effective Pest Treament That Protects Pollinators

https://github.com/shivanikuckreja/CitrolaKuckrejaSaltman_ENV872_EDA_FinalProject/tree/main/Project

Sam Saltman, Shivani Kuckreja, Jessica Citrola

Contents

1 Rationale and Research Questions **5**

2 Dataset Information **6**

3 Exploratory Analysis **7**

4 Analysis **14**
4.1 Question 1: Is there an exposure type that is more likely to cause mortality for bees vs. non-bee insects? . . . 14
4.2 Question 2: Are there chemicals that are more likely to cause mortality for bees vs. non-bee insects? . . . 19

5 Summary and Conclusions **22**

6 Appendix **23**

List of Tables

3 Species List 7

List of Figures

1 Bee Mortality by Exposure Type 10
2 Non-bee Mortality by Exposure Type 11
3 Bee Mortality by Chemical 12
4 Non-bee Mortality by Chemical 13
5 Residual Plot for Bee Mortality by Exposure Type 23
6 Residual Plot for Non-Bee Mortality by Exposure Type 24
7 Residual Plot for Bee Mortality by Chemical Type 24
8 Residual Plot for Non-Bee Mortality by Chemical Type 25

1 Rationale and Research Questions

Pollination is a critical component of agriculture. Bees are important pollinators, however, a decline in pollinators has been linked to extensive use of insecticides. Measuring hazardous and lethal toxicity as well as potential side effects for various pollinators could be utilized for research or management recommendations. Our analysis evaluates if there are exposure methods and chemicals that do not cause significant harm to bees while eliminating pests. The goal of our analysis is to determine potential treatment methods that reduce pests while having a non-lethal impact on bees.

Questions:

1. *Is there an exposure type that is more likely to cause mortality for bees vs. non-bee insects?*
2. *Are there chemicals that are more likely to cause mortality for bees vs. non-bee insects?*

2 Dataset Information

Data Source: The dataset was pulled from a repository created for Environmental Data Analytics at Duke University in 2020. The data collected is from several EPA studies on neonicotinoids and their effects on insects. The data we will be analyzing is the type of chemical administered, how it was administered, and how both of these variables impact insects.

In the wrangling process, we selected the relevant information to our topic. This includes the chemical type, chemical number, insect species, lifestage and age of the species, exposure type, the effect of the exposure and the measurement of the exposure. An example of an exposure type is giving food to a bee. An example of an effect is mortality, and the measurement is a more detailed analysis of the effect.

We converted all these selected categorical variables to factors to prep for analysis. In the next step, we processed two data frames – all bee species and all non-bee species. The split resulted in 2529 non-bee observations and 1407 bee observations. Lastly we added a mortality column using an ifelse statement. We did this to run a binomial glm in our analysis. We coded mortality as 1 and everything else as 0.

Detail	Description
Data Source	EPA ECOTOX Knowledgebase
——-	———
Retrieved From	https://cfpub.epa.gov/ecotox/help.cfm
——-	———
Date Range	1982-2019

```
kable(sapply(Processed1_Filter, class))
```

	x
CAS.Number	factor
Chemical.Name	factor
Species.Common.Name	factor
Organism.Lifestage	factor
Organism.Age	factor
Exposure.Type	factor
Effect	factor
Effect.Measurement	factor

3 Exploratory Analysis

Summary of all species in study

Table 3: Species List

	x
Honey Bee	667
Parasitic Wasp	285
Buff Tailed Bumblebee	183
Carniolan Honey Bee	152
Bumble Bee	140
Italian Honeybee	113
Japanese Beetle	94
Asian Lady Beetle	76
Euonymus Scale	75
Wireworm	69
European Dark Bee	66
Minute Pirate Bug	62
Asian Citrus Psyllid	60
Parastic Wasp	58
Colorado Potato Beetle	57
Parasitoid Wasp	51
Erythrina Gall Wasp	49
Beetle Order	47
Snout Beetle Family, Weevil	47
Sevenspotted Lady Beetle	46
True Bug Order	45
Buff-tailed Bumblebee	39
Aphid Family	38
Cabbage Looper	38
Sweetpotato Whitefly	37
Braconid Wasp	33
Cotton Aphid	33
Predatory Mite	33
Ladybird Beetle Family	30
Parasitoid	30
Scarab Beetle	29
Spring Tiphia	29
Thrip Order	29
Ground Beetle Family	27
Rove Beetle Family	27
Tobacco Aphid	27
Chalcid Wasp	25
Convergent Lady Beetle	25

	x
Stingless Bee	25
Spider/Mite Class	24
Tobacco Flea Beetle	24
Citrus Leafminer	23
Ladybird Beetle	23
Mason Bee	22
Mosquito	22
Argentine Ant	21
Beetle	21
Flatheaded Appletree Borer	20
Horned Oak Gall Wasp	20
Leaf Beetle Family	20
Potato Leafhopper	20
Tooth-necked Fungus Beetle	20
Codling Moth	19
Black-spotted Lady Beetle	18
Calico Scale	18
Fairyfly Parasitoid	18
Lady Beetle	18
Minute Parasitic Wasps	18
Mirid Bug	18
Mulberry Pyralid	18
Silkworm	18
Vedalia Beetle	18
Araneoid Spider Order	17
Bee Order	17
Egg Parasitoid	17
Insect Class	17
Moth And Butterfly Order	17
Oystershell Scale Parasitoid	17
Hemlock Woolly Adelgid Lady Beetle	16
Hemlock Wooly Adelgid	16
Mite	16
Onion Thrip	16
Western Flower Thrips	15
Corn Earworm	14
Green Peach Aphid	14
House Fly	14
Ox Beetle	14
Red Scale Parasite	14
Spined Soldier Bug	14
Armoured Scale Family	13
Diamondback Moth	13

	x
Eulophid Wasp	13
Monarch Butterfly	13
Predatory Bug	13
Yellow Fever Mosquito	13
Braconid Parasitoid	12
Common Thrip	12
Eastern Subterranean Termite	12
Jassid	12
Mite Order	12
Pea Aphid	12
Pond Wolf Spider	12
Spotless Ladybird Beetle	11
Glasshouse Potato Wasp	10
Lacewing	10
Southern House Mosquito	10
Two Spotted Lady Beetle	10
Ant Family	9
Apple Maggot	9
(Other)	670

Bar chart comparing exposure type to mortality count of bees. It initially looked like bees are more likely to die when chemical exposure comes from consuming food.The other exposure types do not look particularly significant

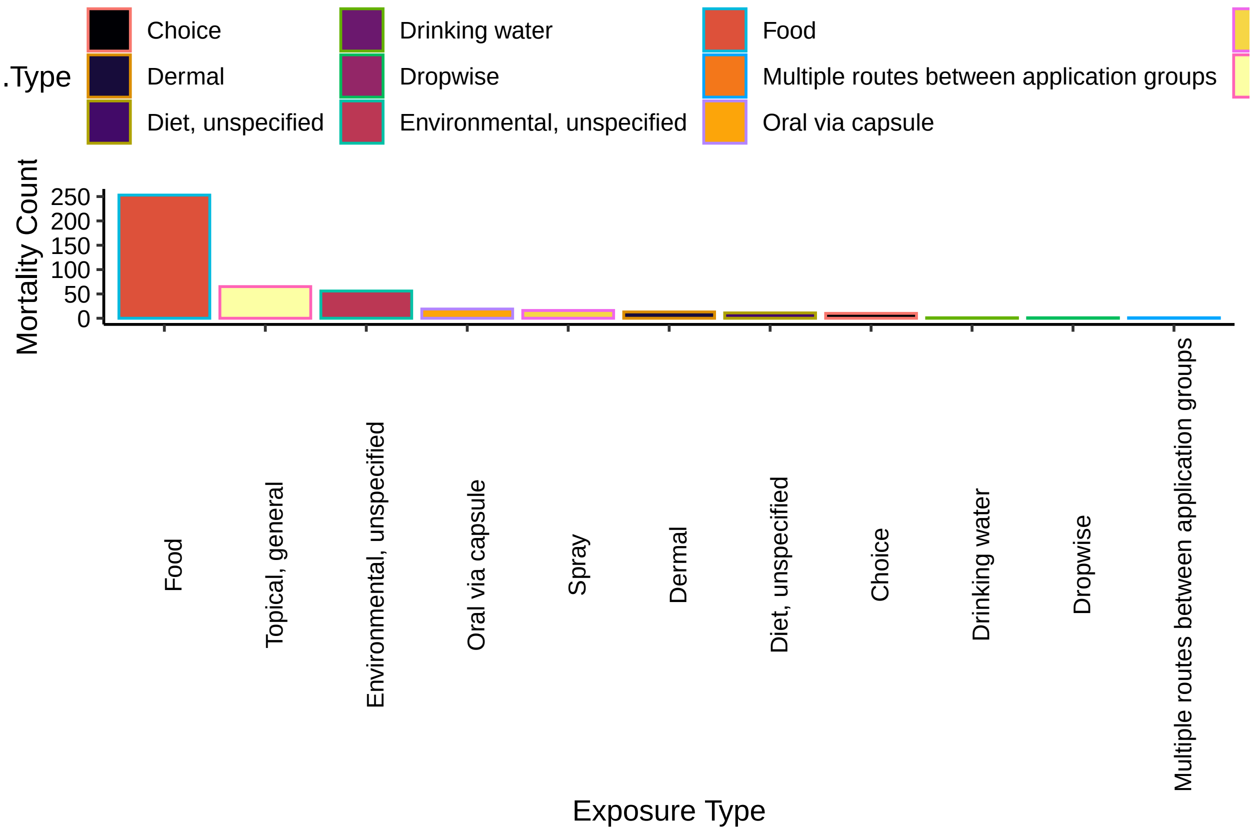

Figure 1: Bee Mortality by Exposure Type

Bar chart comparing exposure type to mortality count of non-bees. Exposure from an environmental source appears more lethal to non-bees species.

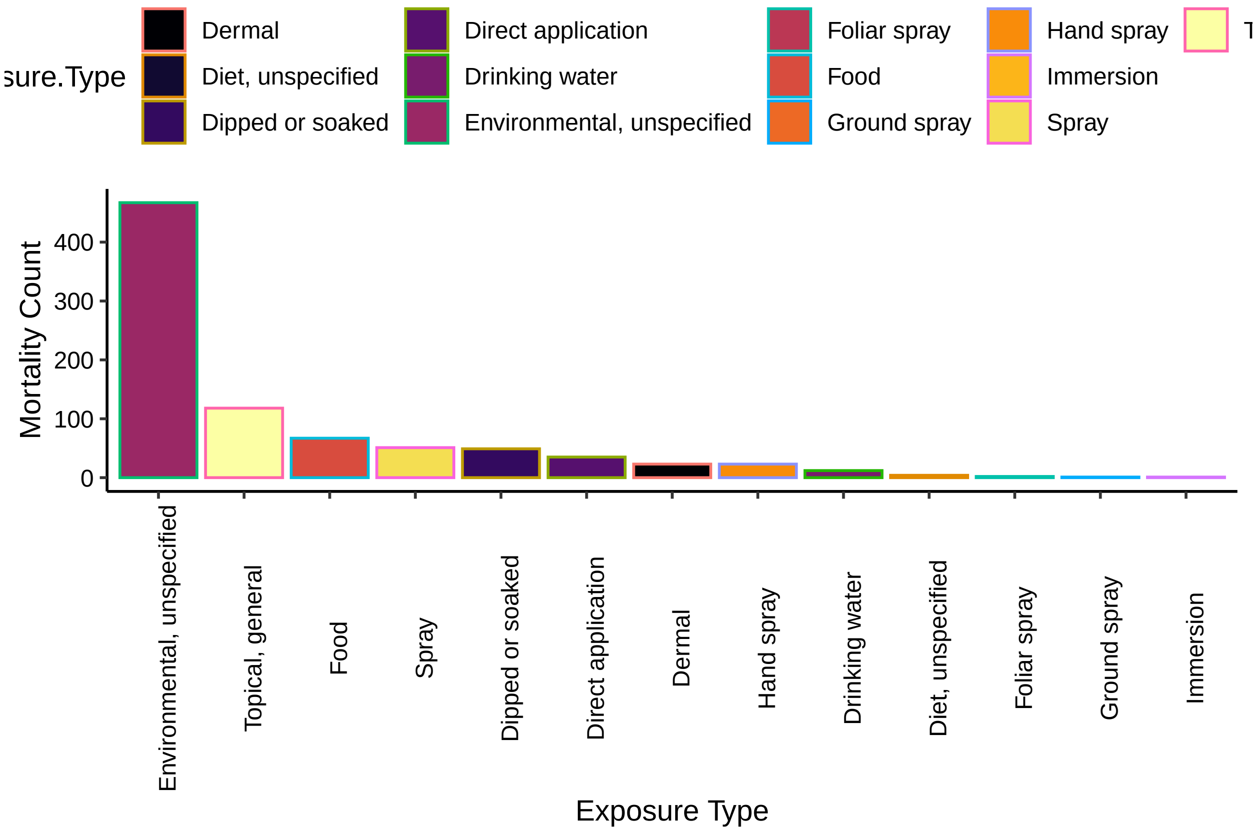

Figure 2: Non-bee Mortality by Exposure Type

The next area of exploration was looking into how many bee and non-bee samples died from exposure to certain chemical compounds. We used the chemical number as the chemical names are complex.

This bar chart looking at bees suggests that three types of chemical compounds in this dataset are toxic to bees.

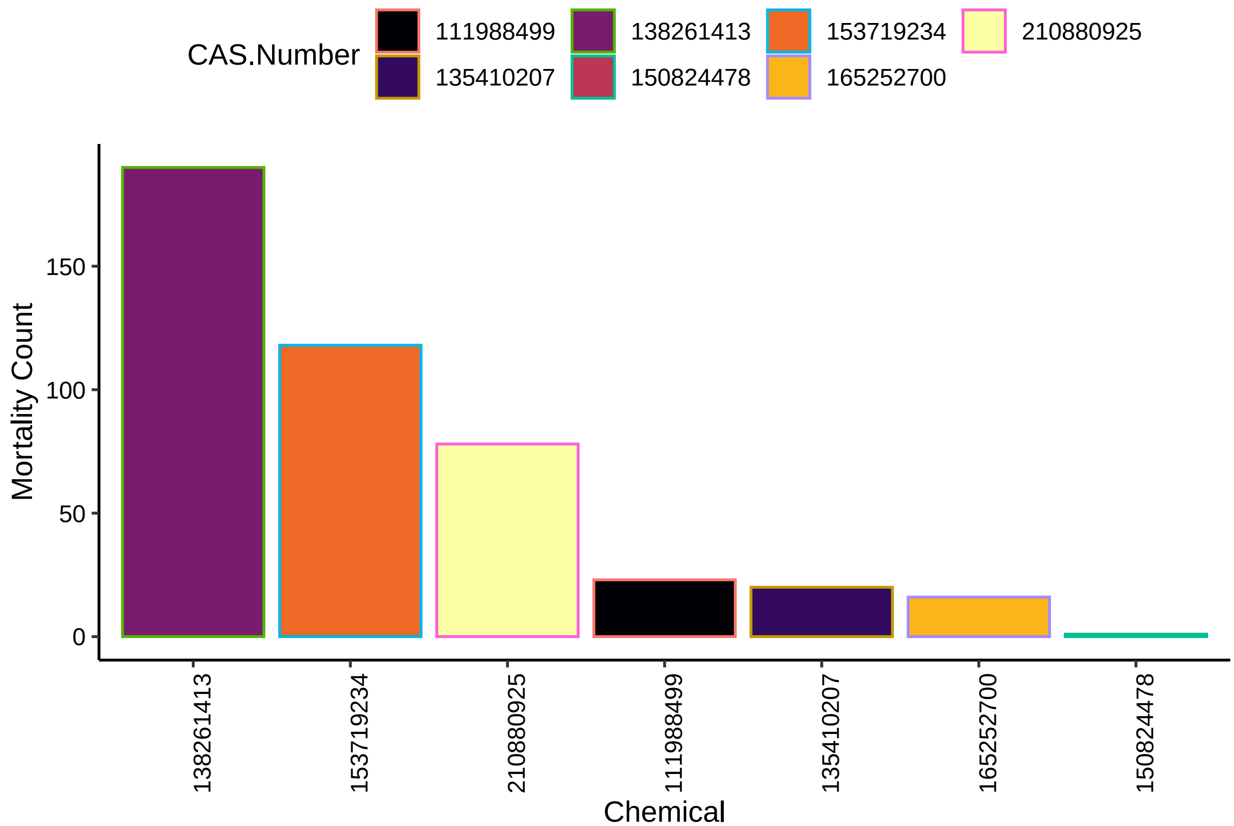

Figure 3: Bee Mortality by Chemical

This bar chart looking at bees suggests that three types of chemical compounds in this dataset are toxic to non-bees.

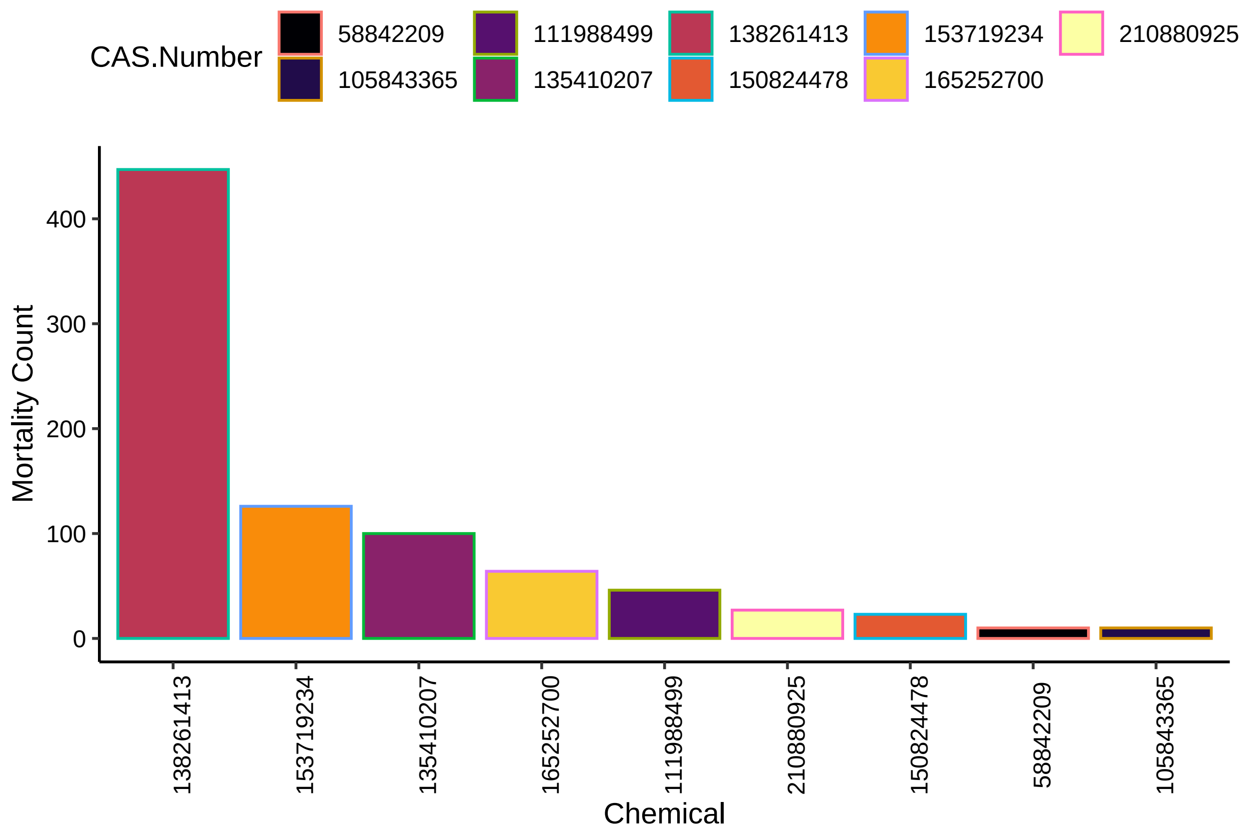

Figure 4: Non-bee Mortality by Chemical

Next we analyze to see if there is any statistical significance supporting these observations.

4 Analysis

We ran GLMs to analyze our categorical data. Using the if/else statement to create the mortality column, we were able to compare mortality against all other effects. We ran two GLMs for the bee category and two for the non-bee categories. The two types of GLMs we ran analyzed mortality against exposure type and mortality against chemical type.To understand if this regression was a fit, we ran a pseudo regression using the pR2 function on each GLM as well

The 'logit' model evaluated the effect of exposure type on mortality on bee species. The results of this model showed that topical exposure had a significant effect on bee mortality (p=.009). For every one unit change in topical, the odds of mortality increase by 1.1787. In comparison to non-bee species, 'logit2' had various exposure types that had a significant effect on mortality. The significant exposure types included dipped or soaked, direct application, drinking water, environmental (unspecified), foliar spray, ground spray, hand spray, spray, and topical.

The 'logit3' and 'logit4' models evaluated which chemical types have a lethal effect on bee and non bee species. The results of 'logit3' showed that nearly all chemical types with data had a significant effect on bees. However, the only chemical that was not significant (150824478) only had one entry. The results of 'logit4' showed that no chemical had a significant effect on non-bee species. The results of the four models are summarized in Table 1.

We evaluated the goodness-of-fit by finding the McFadden pseudo R-squared for each model. The model for exposure type and non-bee mortality ('logit2') had a good fit with a McFadden pseudo R-squared value of .214. However, the rest of the models had very low pseudo R-squared values. As a result, these models may not provide the most predictive information.

4.1 Question 1: Is there an exposure type that is more likely to cause mortality for bees vs. non-bee insects?

```
#Which exposure types have a significant effect (mortality) on bee species?
BeeMortalityExposureType$Mortality <- ifelse(BeeMortalityExposureType$Effect=="Mortality

#Making new column categorical
BeeMortalityExposureType$Mortality <- as.factor(BeeMortalityExposureType$Mortality)

logit <- glm(Mortality ~ Exposure.Type, data = BeeMortalityExposureType, family = "binom
summary(logit)
```

```
##
## Call:
## glm(formula = Mortality ~ Exposure.Type, family = "binomial",
##     data = BeeMortalityExposureType)
##
## Deviance Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -1.4620  -0.7742  -0.7742   1.1774   1.9728
##
## Coefficients:
##                                                   Estimate Std. Error
## (Intercept)                                        -0.5306     0.3985
## Exposure.TypeDermal                                -0.8747     0.5046
## Exposure.TypeDiet, unspecified                      0.5306     0.5836
## Exposure.TypeDirect application                   -16.0354  1696.7344
## Exposure.TypeDrinking water                        17.0967  2399.5448
## Exposure.TypeDropwise                              17.0967  2399.5448
## Exposure.TypeEnvironmental, unspecified             0.1252     0.4343
## Exposure.TypeFood                                  -0.5208     0.4052
## Exposure.TypeGround granular                      -16.0354  1073.1091
## Exposure.TypeHand spray                           -16.0354  1199.7724
## Exposure.TypeMultiple routes between application groups  -1.2611     1.1513
## Exposure.TypeOral via capsule                      17.0967   550.4935
## Exposure.TypeSpray                                  0.2587     0.5186
## Exposure.TypeTopical, general                       1.1787     0.4512
##                                                  z value Pr(>|z|)
## (Intercept)                                       -1.331    0.183
## Exposure.TypeDermal                               -1.734    0.083 .
## Exposure.TypeDiet, unspecified                     0.909    0.363
## Exposure.TypeDirect application                   -0.009    0.992
## Exposure.TypeDrinking water                        0.007    0.994
## Exposure.TypeDropwise                              0.007    0.994
## Exposure.TypeEnvironmental, unspecified            0.288    0.773
## Exposure.TypeFood                                 -1.285    0.199
## Exposure.TypeGround granular                      -0.015    0.988
## Exposure.TypeHand spray                           -0.013    0.989
## Exposure.TypeMultiple routes between application groups  -1.095    0.273
## Exposure.TypeOral via capsule                      0.031    0.975
## Exposure.TypeSpray                                 0.499    0.618
## Exposure.TypeTopical, general                      2.612    0.009 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1757.6  on 1406  degrees of freedom
## Residual deviance: 1621.4  on 1393  degrees of freedom
## AIC: 1649.4
##
## Number of Fisher Scoring iterations: 15
```

```r
#For every one unit change in Topical, the odds of mortality increase by 1.1787. The P
#Significance: Total topical samples 99 - mortality - 65
#Additional data collection is needed on: Exposure through soil contact, drinking

#Pseudo R2
 pR2(logit)
```

```
## fitting null model for pseudo-r2

##           llh       llhNull            G2      McFadden          r2ML
## -810.68616225 -878.77992574  136.18752697    0.07748671    0.09225597
##          r2CU
##    0.12934540
```

Exposure Effect on Bees	Pr(>
Topical, general	0.009 **

*Pseudo McFadden Score = 0.07748671

```r
#Which exposure types have a significant effect (mortality) on non bee species?

NonBeeMortalityExposure$Mortality <- ifelse(NonBeeMortalityExposure$Effect=="Mortality",

#Making new column categorical
NonBeeMortalityExposure$Mortality <- as.factor(NonBeeMortalityExposure$Mortality)

logit2 <- glm(Mortality ~ Exposure.Type, data = NonBeeMortalityExposure, family = "binom
summary(logit2)
```

```
##
## Call:
## glm(formula = Mortality ~ Exposure.Type, family = "binomial",
##     data = NonBeeMortalityExposure)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2581  -1.0150  -0.1775   1.2435   2.9497
##
## Coefficients:
##                                                                  Estimate
## (Intercept)                                                        1.1896
## Exposure.TypeDiet, unspecified                                    -0.4964
## Exposure.TypeDipped or soaked                                     -1.3921
## Exposure.TypeDirect application                                   -1.9780
## Exposure.TypeDrinking water                                       -1.3437
```

```
## Exposure.TypeEnvironmental, unspecified                       -1.5843
## Exposure.TypeFilmcoating                                      -18.7557
## Exposure.TypeFoliar spray                                      -5.5269
## Exposure.TypeFood                                              -0.4814
## Exposure.TypeGround granular                                  -18.7557
## Exposure.TypeGround spray                                      -5.3327
## Exposure.TypeHand spray                                        -2.8824
## Exposure.TypeImmersion                                         16.3765
## Exposure.TypeMisted                                           -18.7557
## Exposure.TypeMultiple routes within environmental exposures  -18.7557
## Exposure.TypePresent in soil                                  -18.7557
## Exposure.TypeSpray                                             -2.5710
## Exposure.TypeTopical, general                                   1.2785
## Exposure.TypeWatered                                          -18.7557
##                                                              Std. Error z value
## (Intercept)                                                      0.4317   2.756
## Exposure.TypeDiet, unspecified                                   0.9676  -0.513
## Exposure.TypeDipped or soaked                                    0.4727  -2.945
## Exposure.TypeDirect application                                  0.4774  -4.144
## Exposure.TypeDrinking water                                      0.5840  -2.301
## Exposure.TypeEnvironmental, unspecified                          0.4358  -3.635
## Exposure.TypeFilmcoating                                      3956.1804  -0.005
## Exposure.TypeFoliar spray                                        0.8324  -6.640
## Exposure.TypeFood                                                0.4812  -1.000
## Exposure.TypeGround granular                                   273.0027  -0.069
## Exposure.TypeGround spray                                        1.0965  -4.864
## Exposure.TypeHand spray                                          0.4877  -5.911
## Exposure.TypeImmersion                                        3956.1804   0.004
## Exposure.TypeMisted                                           1398.7210  -0.013
## Exposure.TypeMultiple routes within environmental exposures  2797.4420  -0.007
## Exposure.TypePresent in soil                                  2284.1018  -0.008
## Exposure.TypeSpray                                               0.4592  -5.599
## Exposure.TypeTopical, general                                    0.5430   2.355
## Exposure.TypeWatered                                          1142.0510  -0.016
##                                                              Pr(>|z|)
## (Intercept)                                                  0.005855 **
## Exposure.TypeDiet, unspecified                               0.607926
## Exposure.TypeDipped or soaked                                0.003227 **
## Exposure.TypeDirect application                              3.42e-05 ***
## Exposure.TypeDrinking water                                  0.021404 *
## Exposure.TypeEnvironmental, unspecified                      0.000278 ***
## Exposure.TypeFilmcoating                                     0.996217
## Exposure.TypeFoliar spray                                    3.14e-11 ***
## Exposure.TypeFood                                            0.317121
## Exposure.TypeGround granular                                 0.945227
```

```
## Exposure.TypeGround spray                                        1.15e-06 ***
## Exposure.TypeHand spray                                          3.41e-09 ***
## Exposure.TypeImmersion                                           0.996697
## Exposure.TypeMisted                                              0.989301
## Exposure.TypeMultiple routes within environmental exposures 0.994651
## Exposure.TypePresent in soil                                     0.993448
## Exposure.TypeSpray                                               2.16e-08 ***
## Exposure.TypeTopical, general                                    0.018538 *
## Exposure.TypeWatered                                             0.986897
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3233.2  on 2528  degrees of freedom
## Residual deviance: 2540.4  on 2510  degrees of freedom
## AIC: 2578.4
##
## Number of Fisher Scoring iterations: 16
```

```
#Dipped or soaked, Direct application, Environmental- unspecified, Foliar spray, Groun
#Plenty of samples for non-bees - correlations
#Conclusion: Spray could be an effective anthropogenic exposure technique to eliminate

#Pseudo R2
 pR2(logit2)
```

```
## fitting null model for pseudo-r2
```

```
##           llh       llhNull            G2      McFadden          r2ML
## -1270.1751454 -1616.5870098   692.8237287     0.2142859     0.2396312
##          r2CU
##     0.3321160
```

Non-Bee Exposure Type	Pr(>
Dipped or soaked	0.003227
Direct application	3.42e-05
Drinking water	0.021404
Environmental, unspecified	0.000278
Foliar spray	3.14e-11
Ground spray	1.15e-06
Hand spray	3.41e-09
Spray	2.16e-08
Topical, general	0.018538

*Pseudo McFadden Score = 0.2142859

4.2 Question 2: Are there chemicals that are more likely to cause mortality for bees vs. non-bee insects?

```
#Which chemical types have a significant effect (mortality) on bee species?

logit3 <- glm(Mortality ~ CAS.Number, data = BeeMortalityExposureType, family = "binomia
summary(logit3)
```

```
##
## Call:
## glm(formula = Mortality ~ CAS.Number, family = "binomial", data = BeeMortalityExposur
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5453  -0.7585  -0.7585   1.2286   1.7080
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)            0.8329     0.3788   2.199 0.027885 *
## CAS.Number135410207  -2.0268     0.4568  -4.437 9.10e-06 ***
## CAS.Number138261413  -1.9315     0.3879  -4.979 6.39e-07 ***
## CAS.Number150824478  11.7332   324.7439   0.036 0.971178
## CAS.Number153719234  -0.9526     0.3993  -2.385 0.017062 *
## CAS.Number165252700  -1.4276     0.4904  -2.911 0.003599 **
## CAS.Number210880925  -1.5066     0.4035  -3.734 0.000189 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1757.6  on 1406  degrees of freedom
## Residual deviance: 1689.6  on 1400  degrees of freedom
## AIC: 1703.6
##
## Number of Fisher Scoring iterations: 11
```

```
 #Need more data on 58842209 105843365 and 58842209, 150824478
 # Every other chemical is statistically significant - 111988499 135410207 138261413

#Pseudo R2
 pR2(logit3)
```

```
## fitting null model for pseudo-r2
```

```
##          llh       llhNull           G2     McFadden         r2ML
## -844.79573614 -878.77992574   67.96837921   0.03867201   0.04715907
##         r2CU
##   0.06611832
```

Chemical Effect on Bees	Pr(>
135410207	9.10e-06 ***
138261413	6.39e-07 ***
153719234	0.017062 *
165252700	0.003599 **
210880925	0.000189 ***

*Pseudo McFadden Score = 0.03867201

```
#Which exposure types have a significant effect (mortality) on non bee species?

logit4 <- glm(Mortality ~ CAS.Number, data = NonBeeMortalityExposure, family = "binomial
summary(logit4)
```

```
##
## Call:
## glm(formula = Mortality ~ CAS.Number, family = "binomial", data = NonBeeMortalityExpo
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5928  -0.8379  -0.8379   1.2907   1.9027
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)          1.557e+01  4.602e+02   0.034    0.973
## CAS.Number105843365  1.448e-09  6.509e+02   0.000    1.000
## CAS.Number111988499 -1.507e+01  4.602e+02  -0.033    0.974
## CAS.Number135410207 -1.583e+01  4.602e+02  -0.034    0.973
## CAS.Number138261413 -1.643e+01  4.602e+02  -0.036    0.972
## CAS.Number150824478 -1.463e+01  4.602e+02  -0.032    0.975
## CAS.Number153719234 -1.615e+01  4.602e+02  -0.035    0.972
## CAS.Number165252700 -1.580e+01  4.602e+02  -0.034    0.973
## CAS.Number210880925 -1.720e+01  4.602e+02  -0.037    0.970
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3233.2  on 2528  degrees of freedom
## Residual deviance: 3091.8  on 2520  degrees of freedom
## AIC: 3109.8
```

```
##
## Number of Fisher Scoring iterations: 14
```

```r
#There is no chemical that is effective at eliminating all non-bee species.
#Conclusion - exposure mechanism of chemical is a more reliable factor at understandin

#Pseudo R2
 pR2(logit4)
```

```
## fitting null model for pseudo-r2
```

```
##           llh       llhNull            G2      McFadden          r2ML
## -1.545888e+03 -1.616587e+03  1.413983e+02  4.373360e-02  5.437649e-02
##          r2CU
##  7.536291e-02
```

Chemical Effect on Non-Bees	Pr(>
Chemical	None

*Pseudo McFadden Score = 4.373360e-02

5 Summary and Conclusions

We found that the exposure technique influenced the mortality predictiveness of the neonicotinoids on the species in our dataset. Our model showed that topical exposure had a significant effect on mortality for bee species, but the pseudo R-squared suggests that the model might not be a goof fit to the data. Our model for exposure techniques on non bee species shows that several exposure techniques significantly affected mortality. The pseudo R-squared value of .214 suggests that this model is a good fit. Regarding the data on chemical types, our models showed that no chemical had a significant effect on non-bee species mortality. For bee species, we did find several chemical types that had a significant effect on mortality, but this model also had a low pseudo R-squared and may not be the most predictive for our data. Given our data on exposure types and their effects on bee and non-bee mortality, we suggest that spray exposure techniques could maximize the effectiveness on reducing unwanted pests while preventing mortality on bees.

6 Appendix

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced

## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```

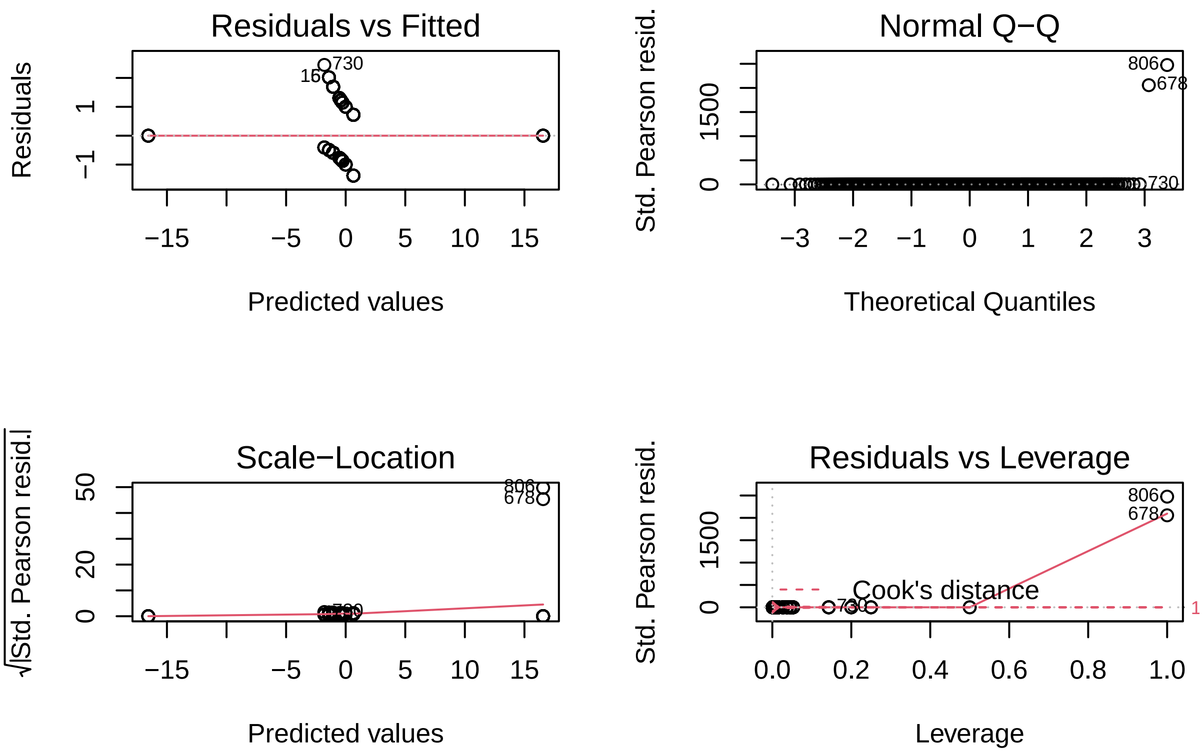

Figure 5: Residual Plot for Bee Mortality by Exposure Type

```
## Warning: not plotting observations with leverage one:
##   1437

## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced

## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced

## Warning: not plotting observations with leverage one:
##   880
```

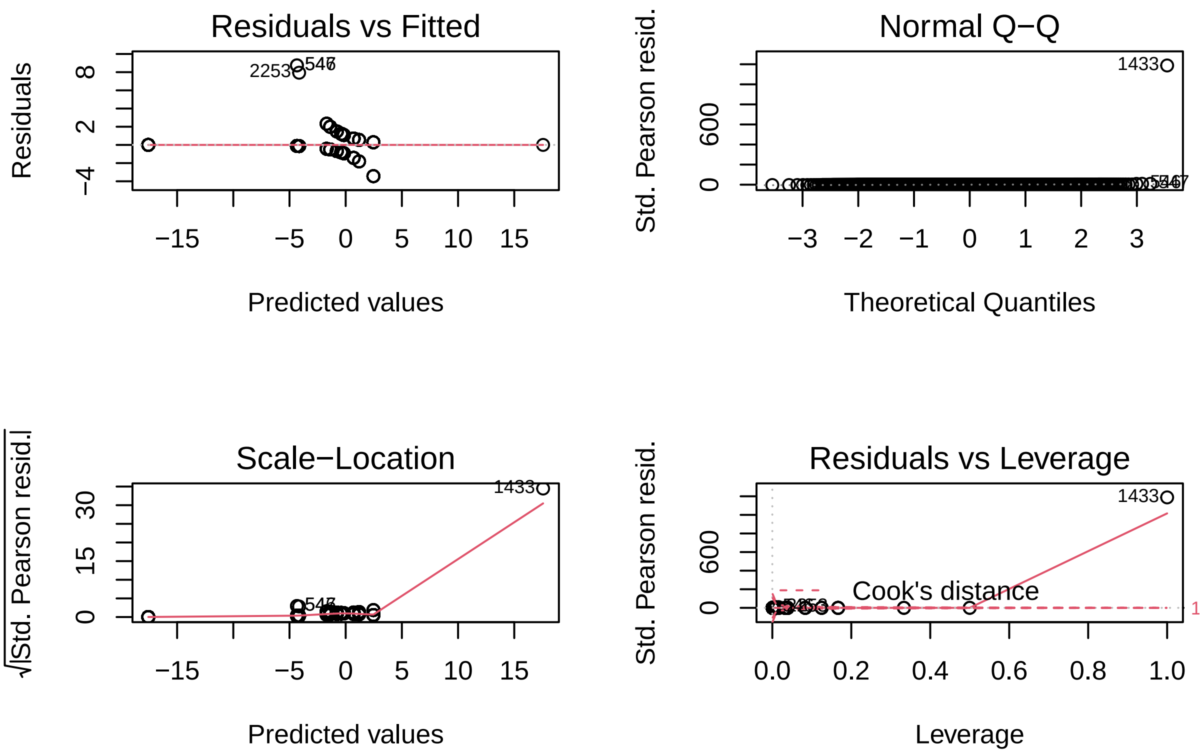

Figure 6: Residual Plot for Non-Bee Mortality by Exposure Type

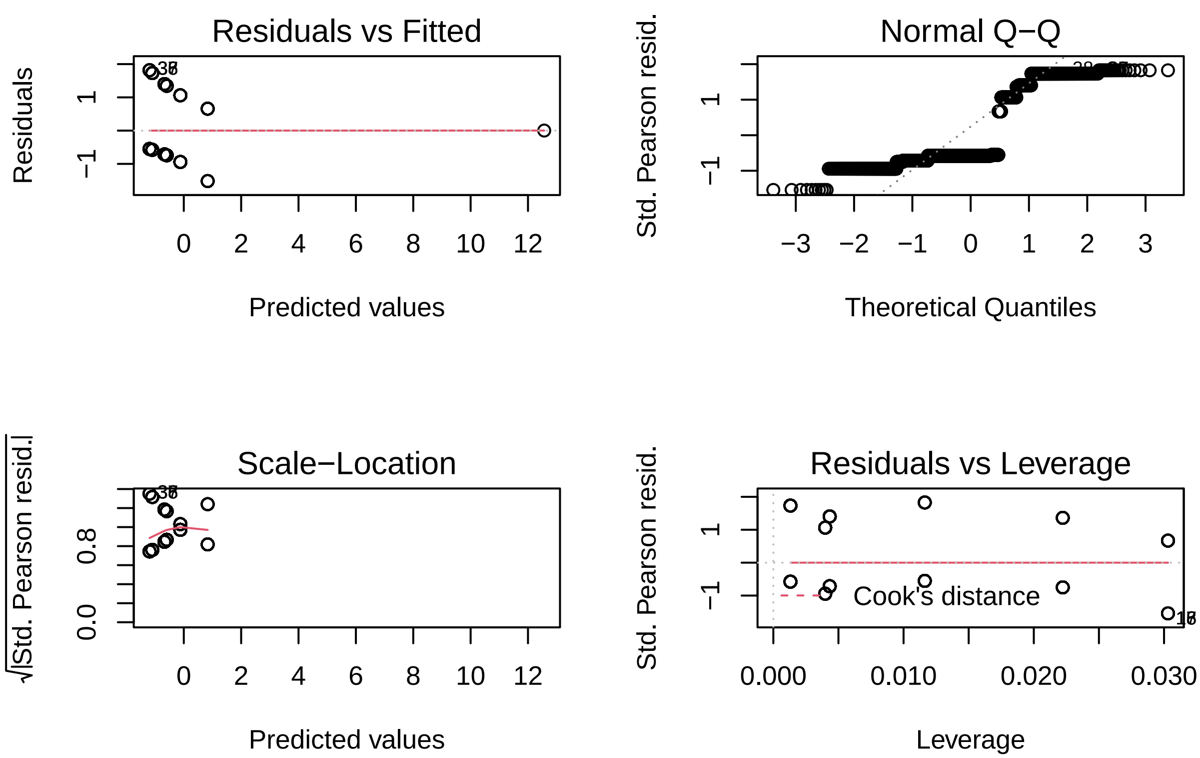

Figure 7: Residual Plot for Bee Mortality by Chemical Type

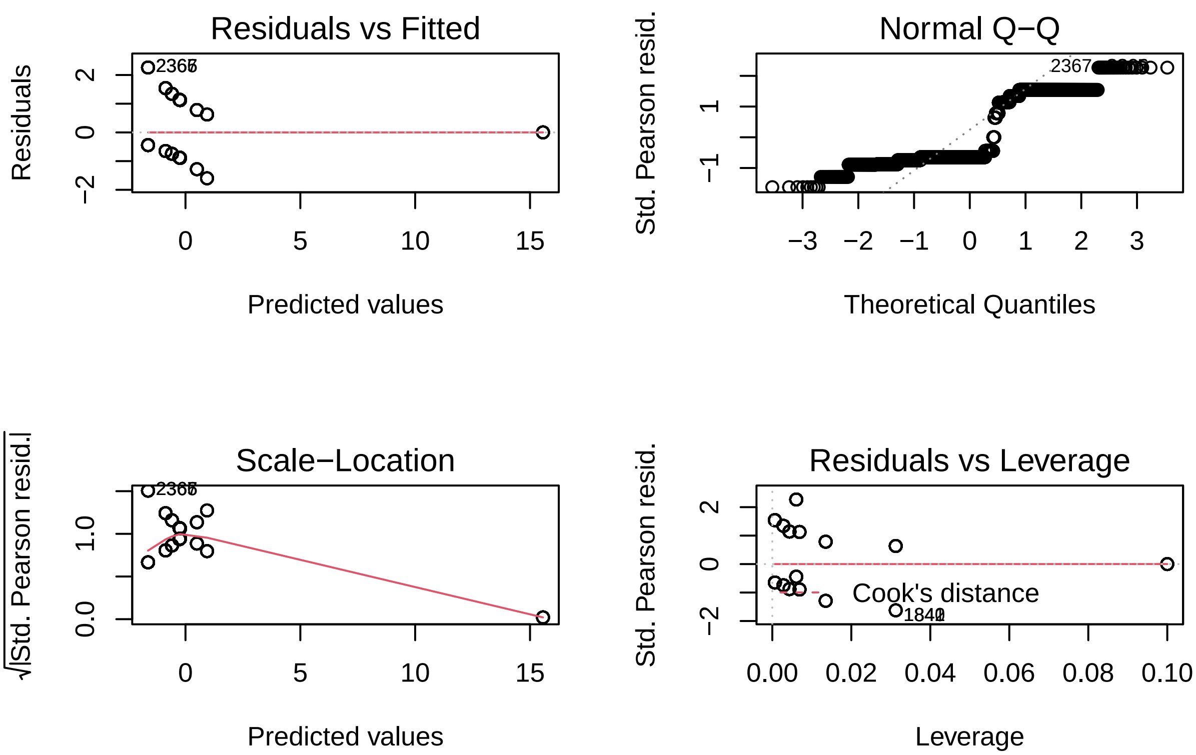

Figure 8: Residual Plot for Non-Bee Mortality by Chemical Type