STA610 Lab 6 Team 4 Report

Cole Juracek Lauren Palazzo Lingyu Zhou Fan Zhu

2021-03-31

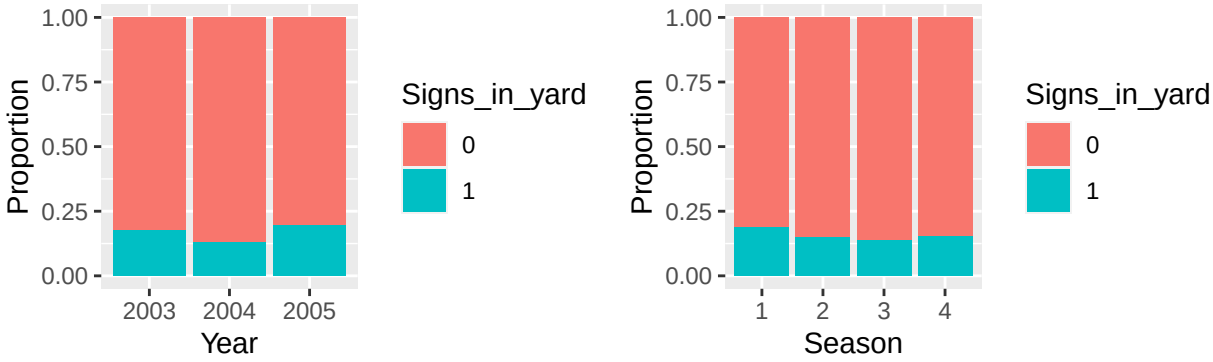
EDA

Response Variable: Signs_in_yard

Table 1: Frequency Table for Signs in Yard

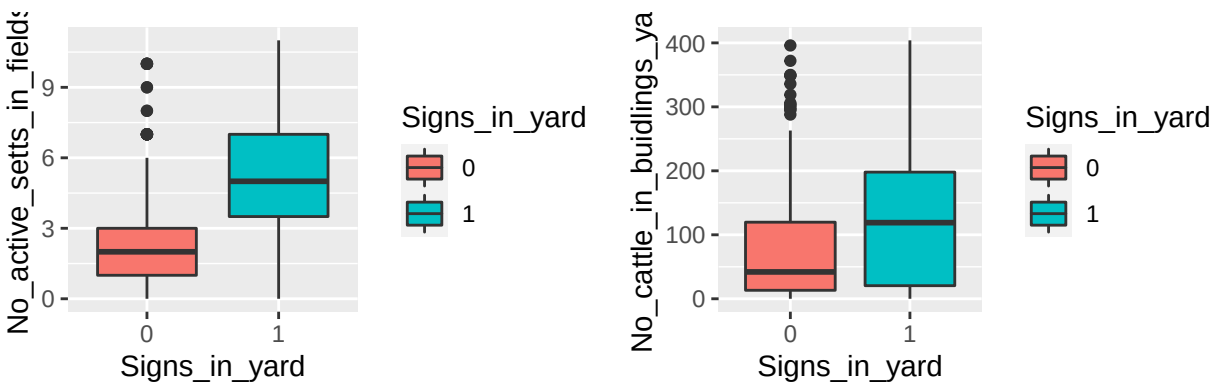
Signs in Yard	Freq
0	230
1	43

Year & Season



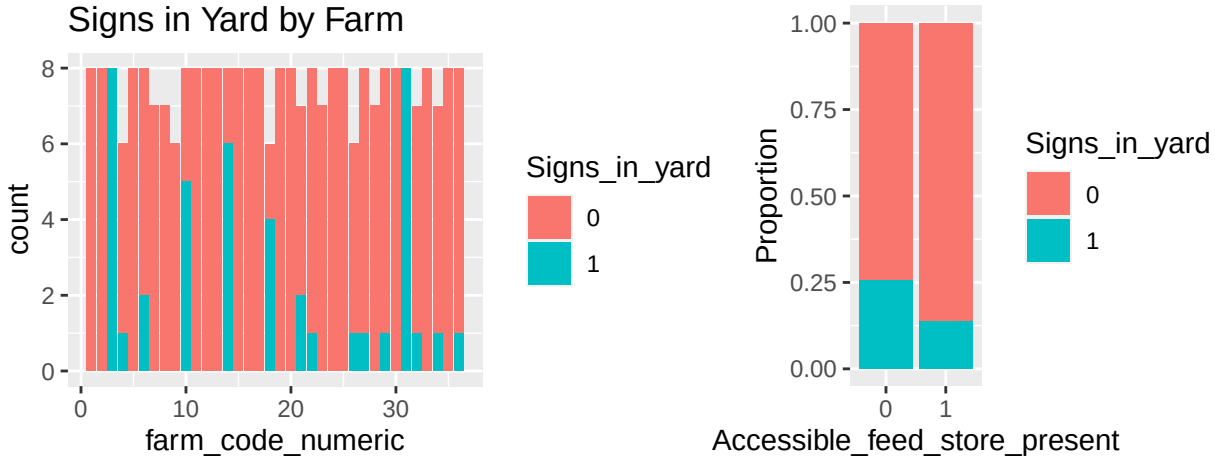
From the bar plot of Signs_in_yard by Year we can observe that the presence of badger activity in the farmyard in 2004 is lower than those in 2003 and 2005. From the bar plot of Signs_in_yard by Season we see that presence of badger activity is the highest in season 1 and the lowest in season 3. Even though badgers do not hibernate, they reduce activities during cold weathers so we will use a main effect of season in our model.

No_active_setts_in_fields & No_cattle_in_buidlings_yard



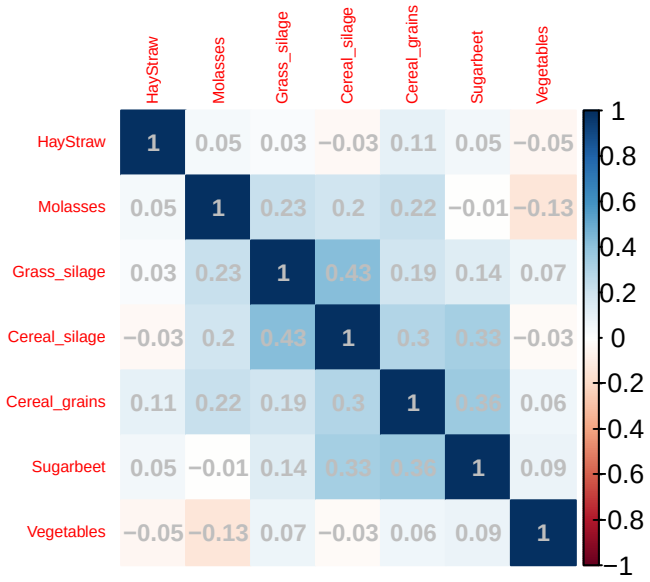
The median of No_active_setts_in_fields is much higher when there is badger activity in the farmyard than when there is no badger activity. The median of No_cattle_in_buidlings_yard is also much higher when there is badger activity in the farmyard than when there is no badger activity.

Farm



The patterns in the stacked bar chart of Signs_in_yard by farm indicate the incorporation of the random intercept by farm into our model. From the segmented bar chart for Accessible_feed_store_present we know that there are more presences of badger activities in the farmyard when there is Accessible_feed_store_present than when there is not.

Binary Variables



Model

Model Selection From the correlation plot we can observe that there are some significant correlations between the binary variables. Therefore, we need to exclude some of them to prevent the collinearity issue. For instance, Cereal_silage and HayStraw cannot exist in the same model because of their significant correlation coefficient. We did not use random intercept for season because there are only four seasons. Interaction terms are removed because they would result in failures to converge and singularity issues.

Based on these principles, we built multiple viable models by adding combinations of binary variables on a basic model $Signsinyard \sim Noactivesettsinfields + Nocattleinbuidlingsyard + (1|farmcodenumeric) + Season$.

Table 2: Model Comparisons with BIC

Model	Binary.Variables.Added.to.the.Basic.Model	BIC
Model4	HayStraw	193.1834
Model5	Sugarbeet	193.5937
Model6	Molasses	194.5083
Model7	Grass silage	194.8197
Model8	Cereal silage	194.8185
Model9	Cereal grains	194.8278
Model10	Cereal silage+Cereal grains	200.4277

From the model comparisons above we chose model 4 as our final model following the principle of parsimony.

Model Specification

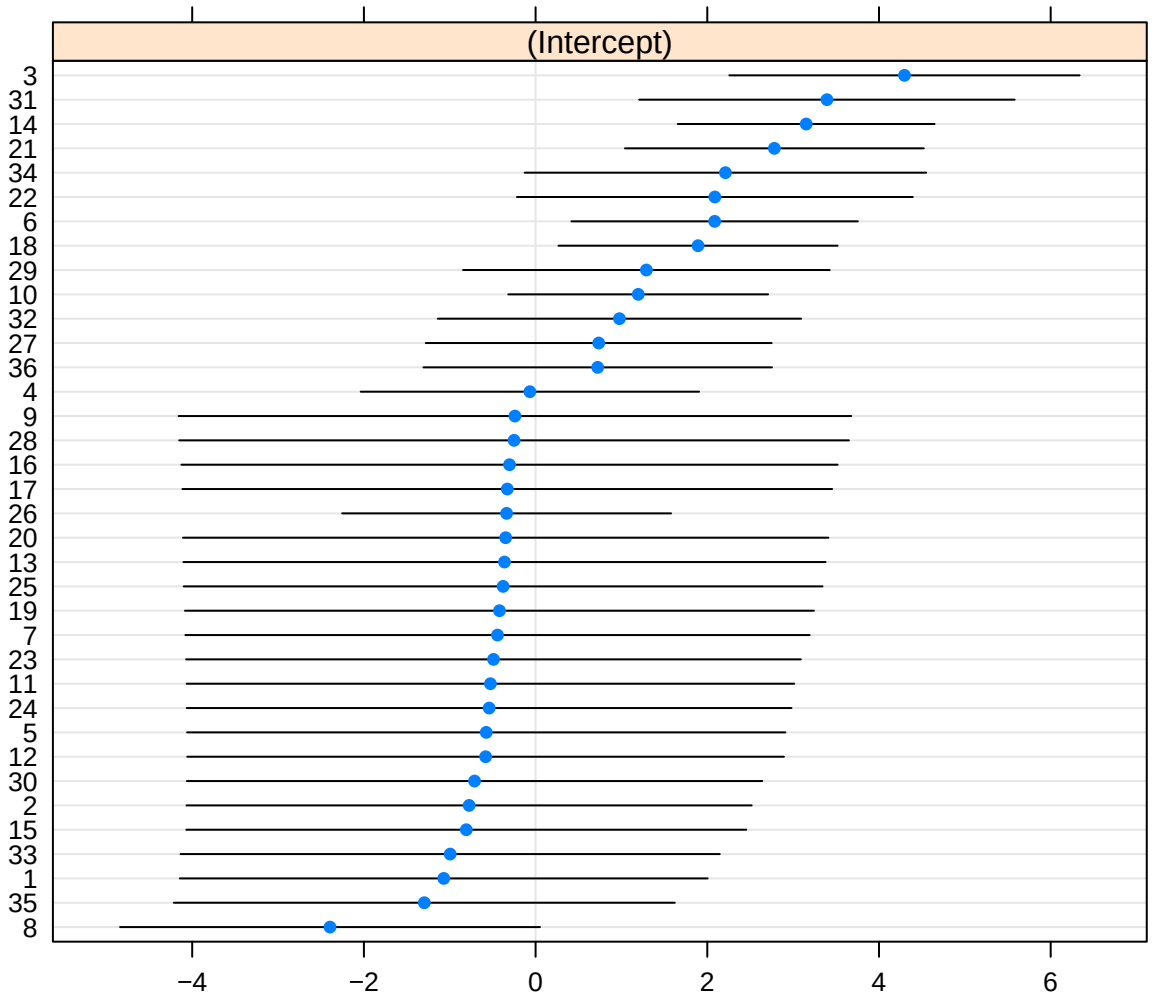
Result

Fixed Effects

	Estimate	Std.Error	t-value	P-value
(Intercept)	-4.6049271	1.1512197	-4.0000420	0.0000633
No_active_setts_in_fields	0.4878348	0.1490160	3.2737069	0.0010615
No_cattle_in_buidlings_yard	0.0051664	0.0036291	1.4236158	0.1545577
Season2	-0.2294778	0.7433120	-0.3087234	0.7575319
Season3	0.0185161	0.8013403	0.0231065	0.9815654
Season4	-0.3679229	0.6917097	-0.5319037	0.5947927
HayStraw	-0.7805303	0.6012218	-1.2982401	0.1942048

Random Effects

Group	Variance	Std.Dev.
Farm (Intercept)	4.94648457248147	2.22406937222773


farm_code_numeric
(Intercept)
3
31
14
21
34
22
6
18
29
10
32
27
36
4
9
28
16
17
26
20
13
25
19
7
23
11
24
5
12
30
2
15
33
1
35
8
−4
−2
0
2
4
6