

STAT230 Work Team Project

Work Team 7: Evan, Becca, Laith

Introduction & Motivation

We are interested in how we can attempt to predict the profitability of farms by different farming practices. Specifically we are curious about sustainable agriculture decisions and decisions about crop types. From a conservationist perspective, it is important that sustainable choices are economically viable for farmers.
<https://www.nass.usda.gov/AgCensus/>

Our data is sourced from the United States Census of Agriculture, between the years 2009 and 2017, along with county demographic information from Wikipedia, sourced from the U.S Census. The data is collected by county, often summing the a variable across all operations in a given county. This limits our specificity, but we are still able to estimate what is typical for farms across the county by looking at averages for each county.

Our response variable is reported farm income, and we are planning to predict it with a number of variables (population density, farm acres, bee colonies, grazing rotation, fertilizer use).

Grazing rotation refers to the number of farming operations in a county that reported using either crop cycling or alternating patterns of grazing to minimize impact. Farm acres and bee colonies are counts that are summed across all farms within a state. Fertilizer use is the total dollar amount spent on fertilizers in that year by all farming operations in a given county. Population density was estimated based on county sizes and populations based on the U.S census via Wikipedia. We feel that this combination of predictors is appropriate because it examines many dimensions of farming practices. It includes information about context of the farming (population density, farm size) as well as the types of methods of sustainable or unsustainable practices carried out in each of these counties. It is valuable to understand how economically viable sustainable farming is in the short term in order to access if further supportive measures should be taken in order to ensure both sufficient food production, as well as producing that food in ways that will limit harm to the Earth.

Dataset & Wrangling

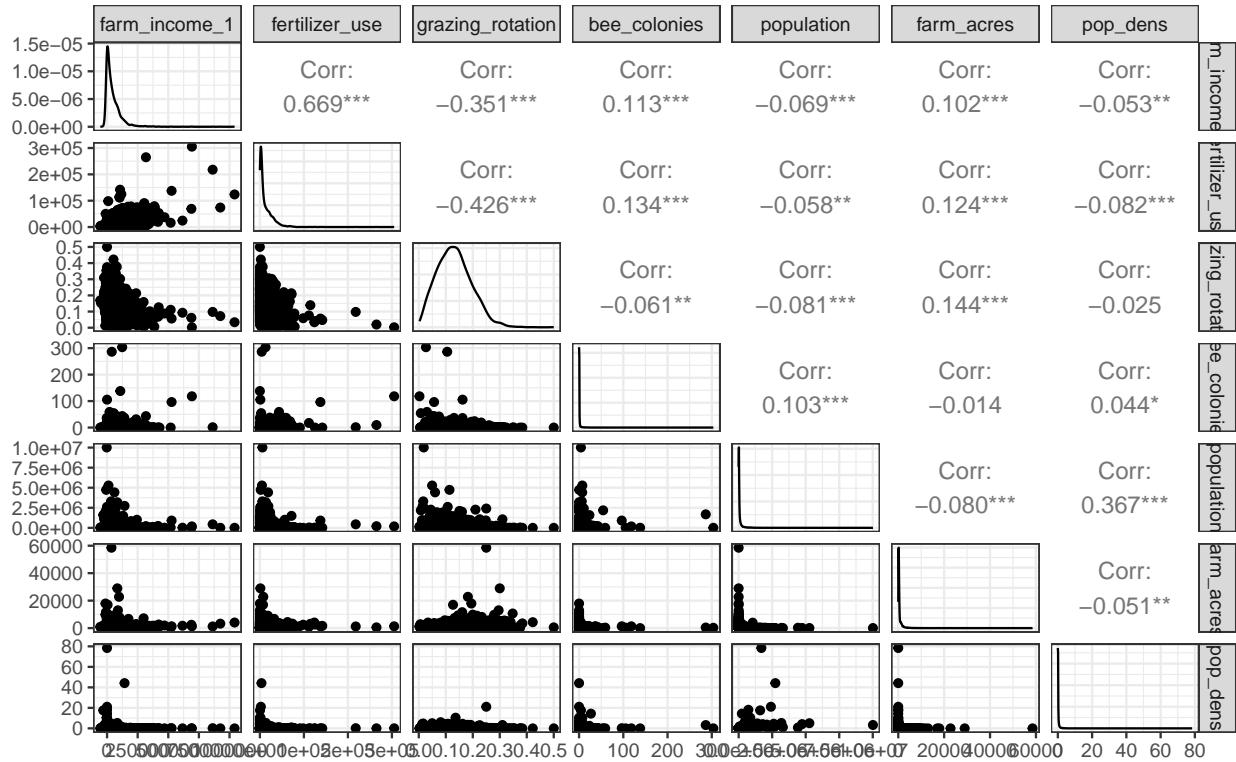
The initial combining, and wrangling of this large data set can be found at <https://github.com/evanmacarthur/Stat230Proj1>. The data wrangling was then perfected as shown in appendix 1

Exploratory Data Analysis (EDA)

Explore distributions and associations graphically and numerically.

Table 1: Summary of Farm Income

	min	Q1	median	Q3	max	mean	sd	n	missing
	-55150	5494.5	27177	66913	1041275	46857	67968	3076	0



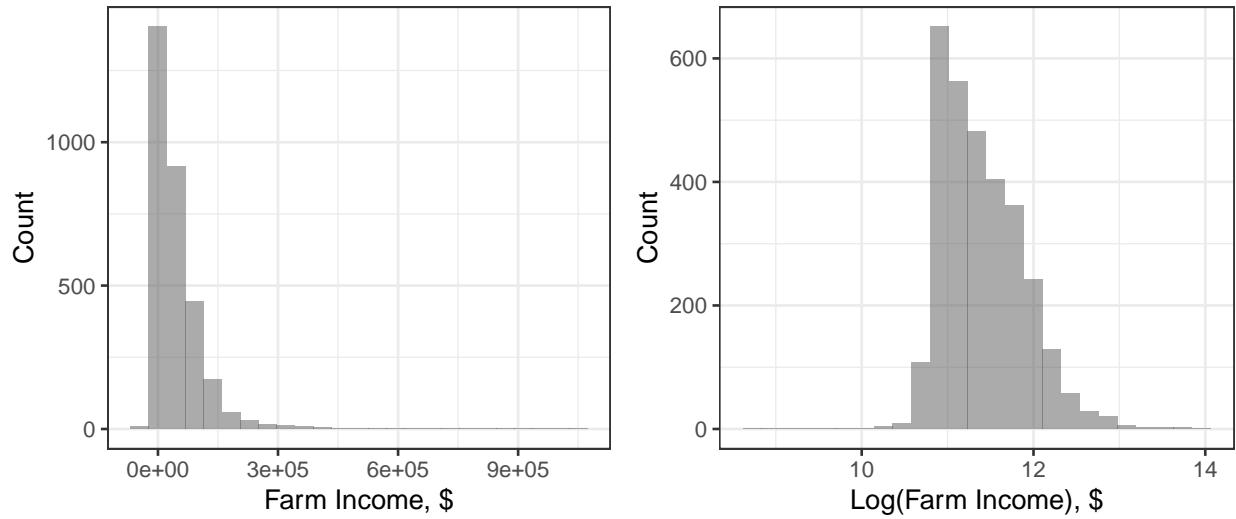
There are significant correlations with all variables, but lots of these associations do not look linear. Transforming the data may help improve these relationships.

The mean is 46857, the median is 27177, we can also tell it looks right skewed from the histogram, and the IQR is 61419

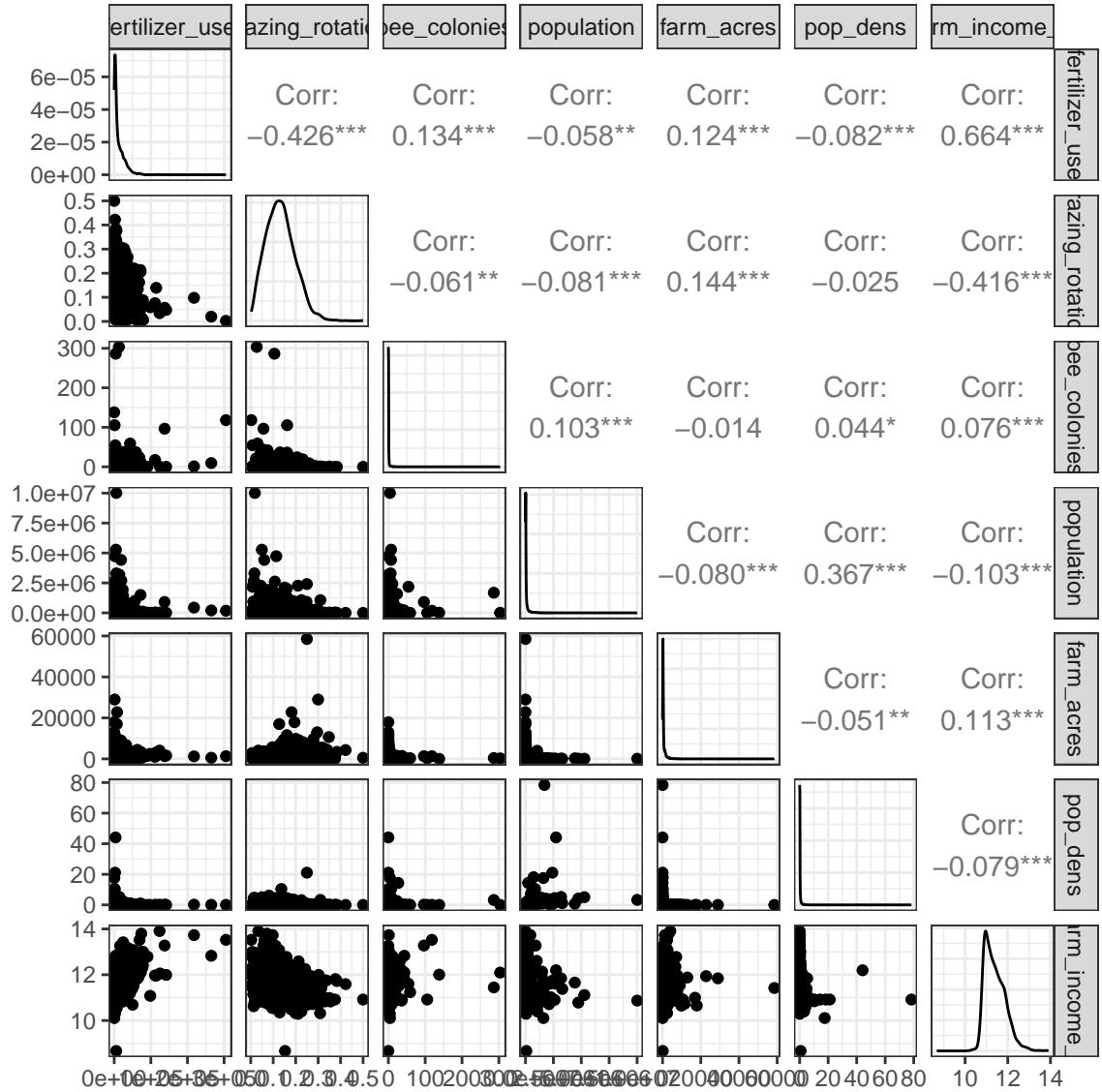
We log-transposed+transformed farm income such that it better matched a normal distribution.

Table 2: Summary of Transformed Farm Income

min	Q1	median	Q3	max	mean	sd	n	missing
8.6733	11.013	11.319	11.712	13.908	11.4	0.48011	3075	0



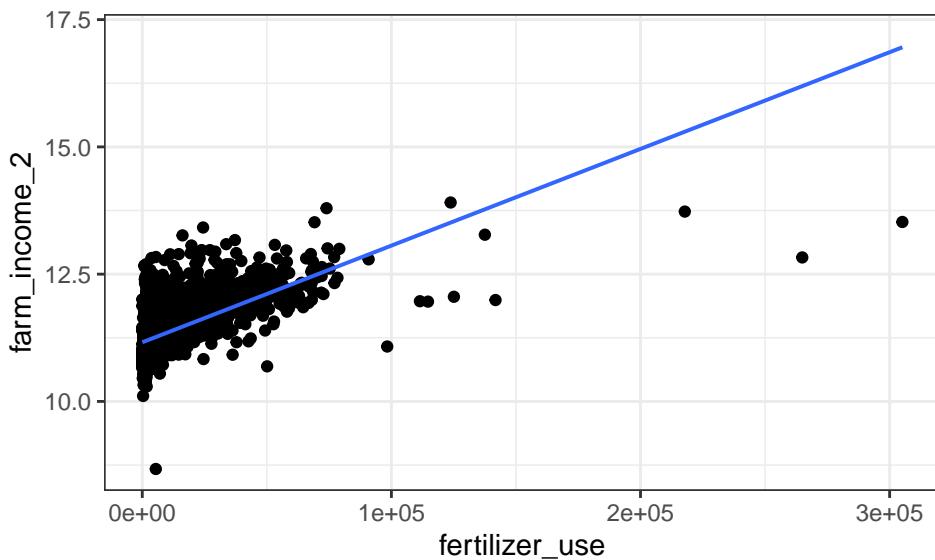
Looks much improved, the mean is 11.4, and median is 11.319, and the histogram still is not ideal, but better. The IQR = 0.699, likely due to the sampling methods, which only include farms that carry out above \$1200 of business annually. It seems like this may introduce a this sort of sharp drop-off in log(farm income)



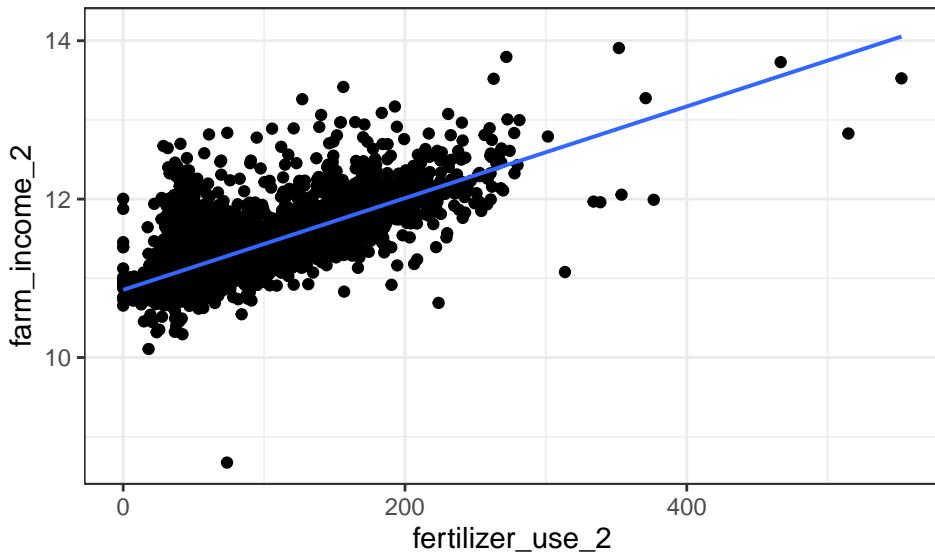
This improved linearity slightly, but still questionable, lets look at predictors.

	min	Q1	median	Q3	max	mean	sd	n	missing
	0	2105.3	6182	18511	305114	12600	16789	3069	6

	min	Q1	median	Q3	max	mean	sd	n	missing
	0	45.883	78.626	136.05	552.37	94.397	60.751	3069	6



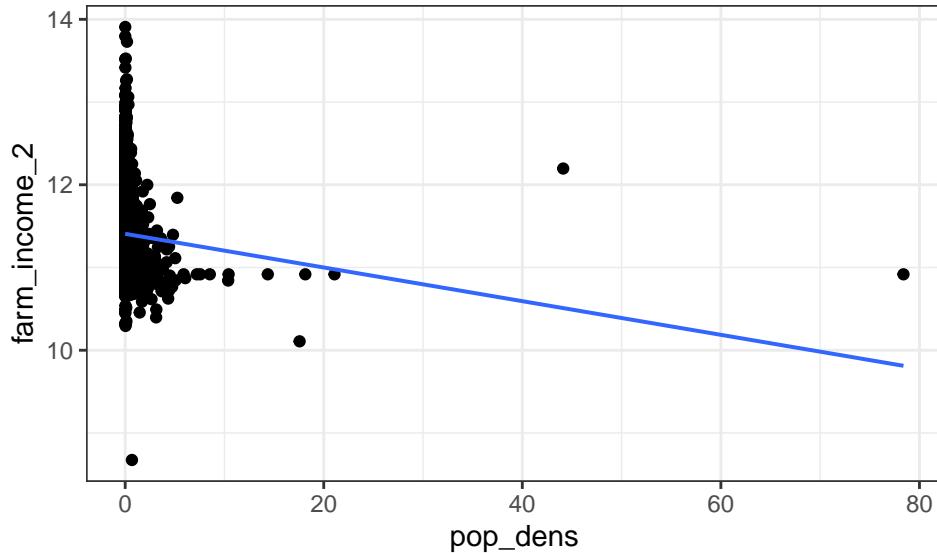
This does not look great, heavily skewed distribution and it is difficult to assess the scatterplot, the mean is 12600, the median is 6182, and the IQR is 16406. Will try transforming.



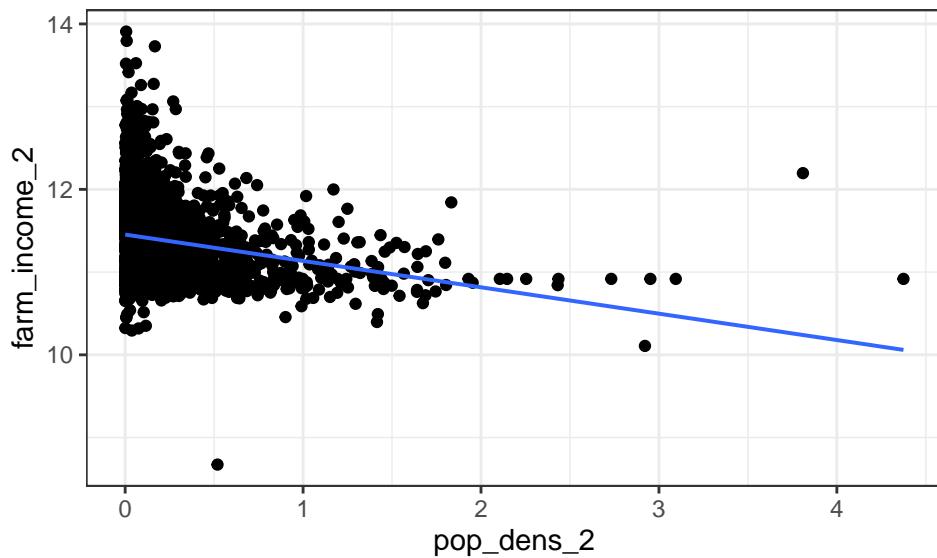
Once again, this improves how the data looks. The distribution is still somewhat right skewed, with a mean of 94.397 and a median of 78.628, and an IQR of 90.167, but this is an improvement, but linearity looks better based on the scatterplot.

	min	Q1	median	Q3	max	mean	sd	n	missing
	0.00015	0.02472	0.06443	0.16653	78.391	0.29922	1.8532	3061	14

	min	Q1	median	Q3	max	mean	sd	n	missing
	0.00015	0.02442	0.06244	0.15404	4.3744	0.16477	0.30615	3061	14



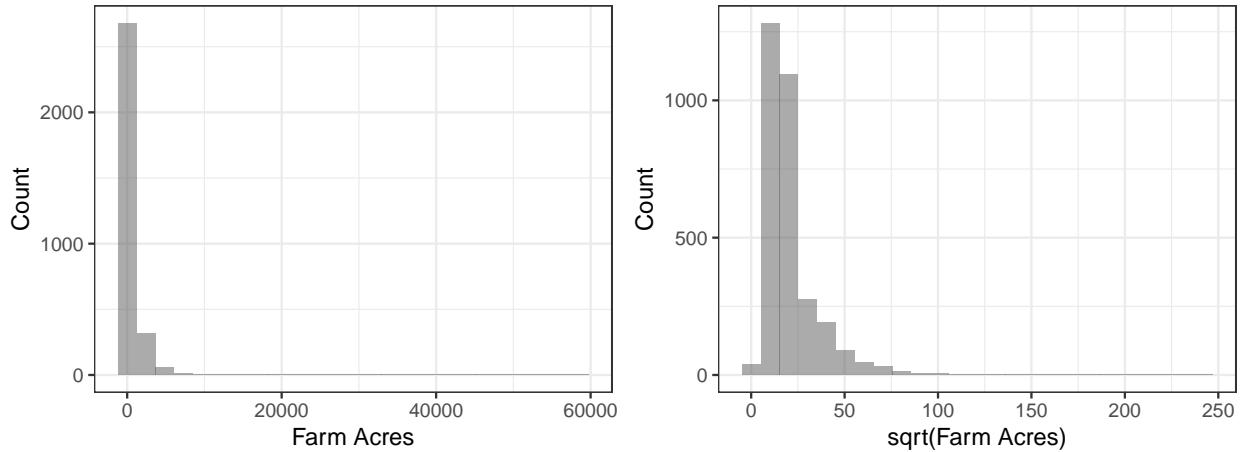
This one is also not great, scatterplot not very useful, the histogram is right skewed, the mean is 0.29922 and median is 0.064434, and the IQR is 0.14181.



This one still does not look great after transforming, but it probably helped. The graph is still right skewed, but not as strongly, with a mean of 0.16477 and a median of 0.062443, and IQR of 0.12962, and the relationship based on the scatter plot is not very useful.

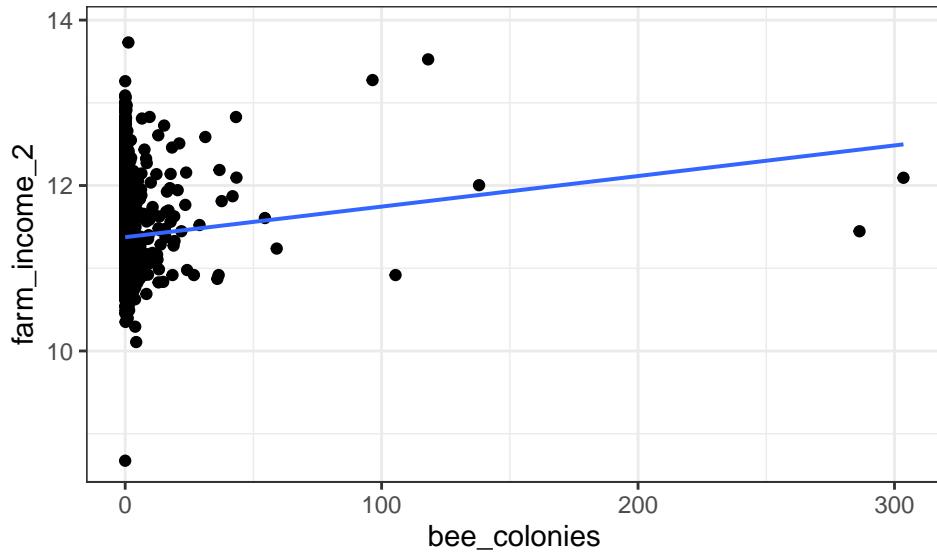
	min	Q1	median	Q3	max	mean	sd	n	missing
	0	12.525	16.353	22.971	241.9	20.833	14.833	3075	0

	min	Q1	median	Q3	max	mean	sd	n	missing
	0	156.87	267.43	527.67	58518	653.95	1663.4	3075	0



Once again, unhelpful scatterplot, right skewed distribution, mean of 291893, median of 179306 and IQR of 282196

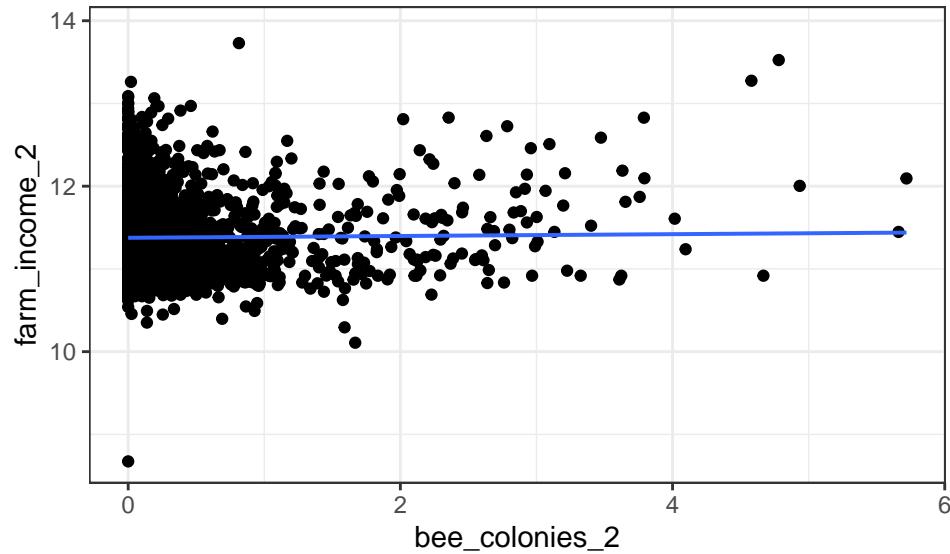
This transformation improved it somewhat, the histogram is less right skewed, but the relationship does not look very linear based on the scatter plot, the mean is 20.833 and the median is 16.353, and the IQR is 10.446



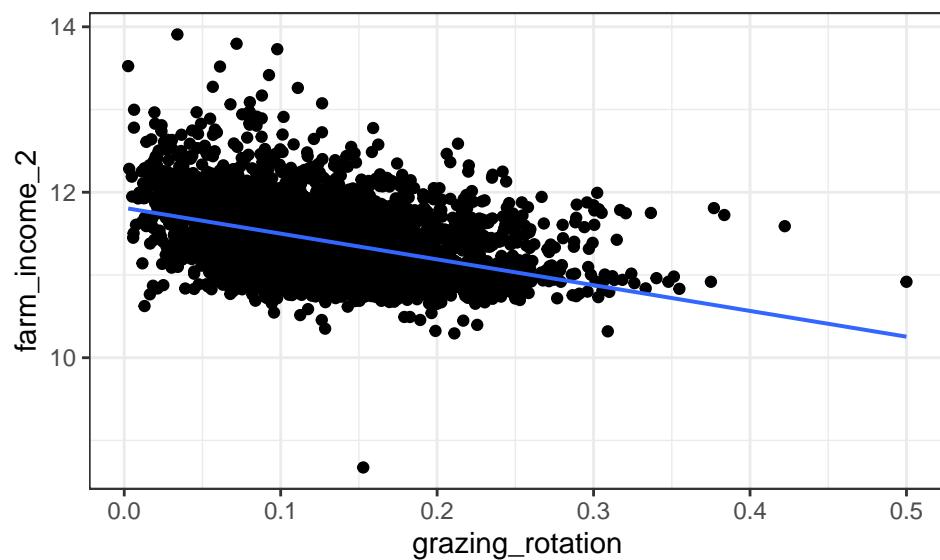
Very strongly right skewed histogram, scatterplot is not useful for assessing linearity, the mean is 1.1837, the median is 0.16083, and the IQR is 0.35789

	min	Q1	median	Q3	max	mean	sd	n	missing
	0	0.04734	0.16083	0.40523	303.5	1.1837	9.5011	2849	226

	min	Q1	median	Q3	max	mean	sd	n	missing
	0	0.04625	0.14913	0.3402	5.7187	0.32371	0.56401	2849	226



After transforming the data, the histogram looks somewhat better, the main part of the distribution is very normal, but we have introduced a large number of 0s into the dataset, which represent real values, so probably shouldn't be removed. the mean is 4.1501 the median is 4.5433 and the IQR is 2.4136. The scatter plot similarly seems to show a fairly linear relationship aside from a large number of 0 values.

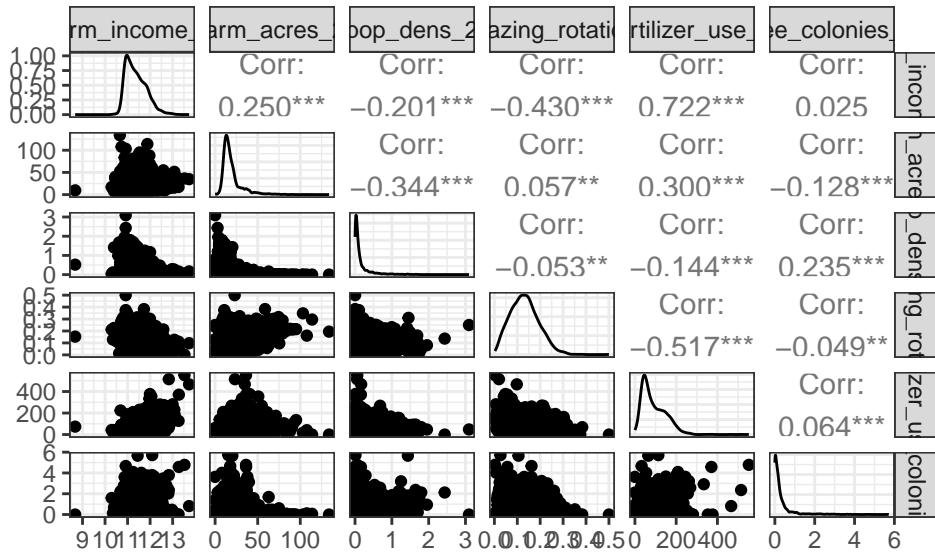


This distribution looks pretty good. Relatively normal histogram, and the scatter plot shows a

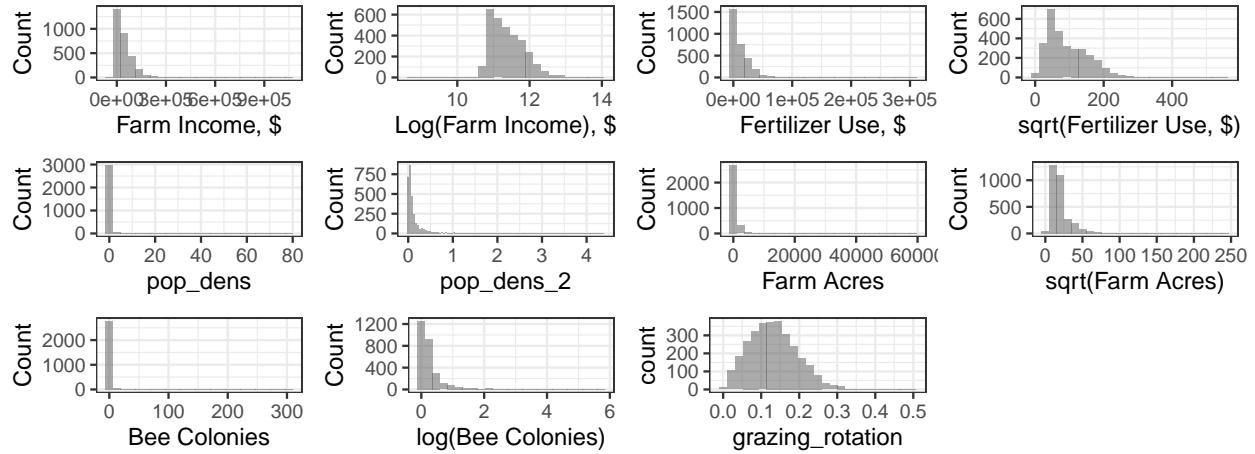
	min	Q1	median	Q3	max	mean	sd	n	missing
	0.00253	0.08572	0.12806	0.17316	0.5	0.13218	0.06355	3034	41

relatively linear relationship. mean is 0.13218, median is 0.12806 and IQR is 0.08744

After we have tried to improve all of the response variables, lets take a look at the ggpairs again



These generally look much better, we still have a lot of 0 values in the population density, but they are real, so the data is not actually normally distributed. The scatter plots generally look much more linear than before transforming the variables, and the correlations are similar to before transforming. Bee colonies is no longer correlated though I removed missing values to make my following tests and operations easier.



Model selection

```
## [1] "which"   "rsq"     "rss"      "adjr2"    "cp"       "bic"      "outmat"   "obj"
```

	adjr2	CP	bic	rss	farm_acres_2	pop_dens_2	grazing_rotation	fertilizer_use_2	bee_colonies_2
1 (1)	0.521	100.154	-2053.1	283.48			*		
2 (1)	0.530	44.510	-2101.5	277.86		*		*	
3 (1)	0.536	5.694	-2134.1	273.89		*	*	*	
4 (1)	0.537	4.283	-2129.6	273.56	*	*	*	*	
5 (1)	0.537	6.000	-2121.9	273.53	*	*	*	*	*

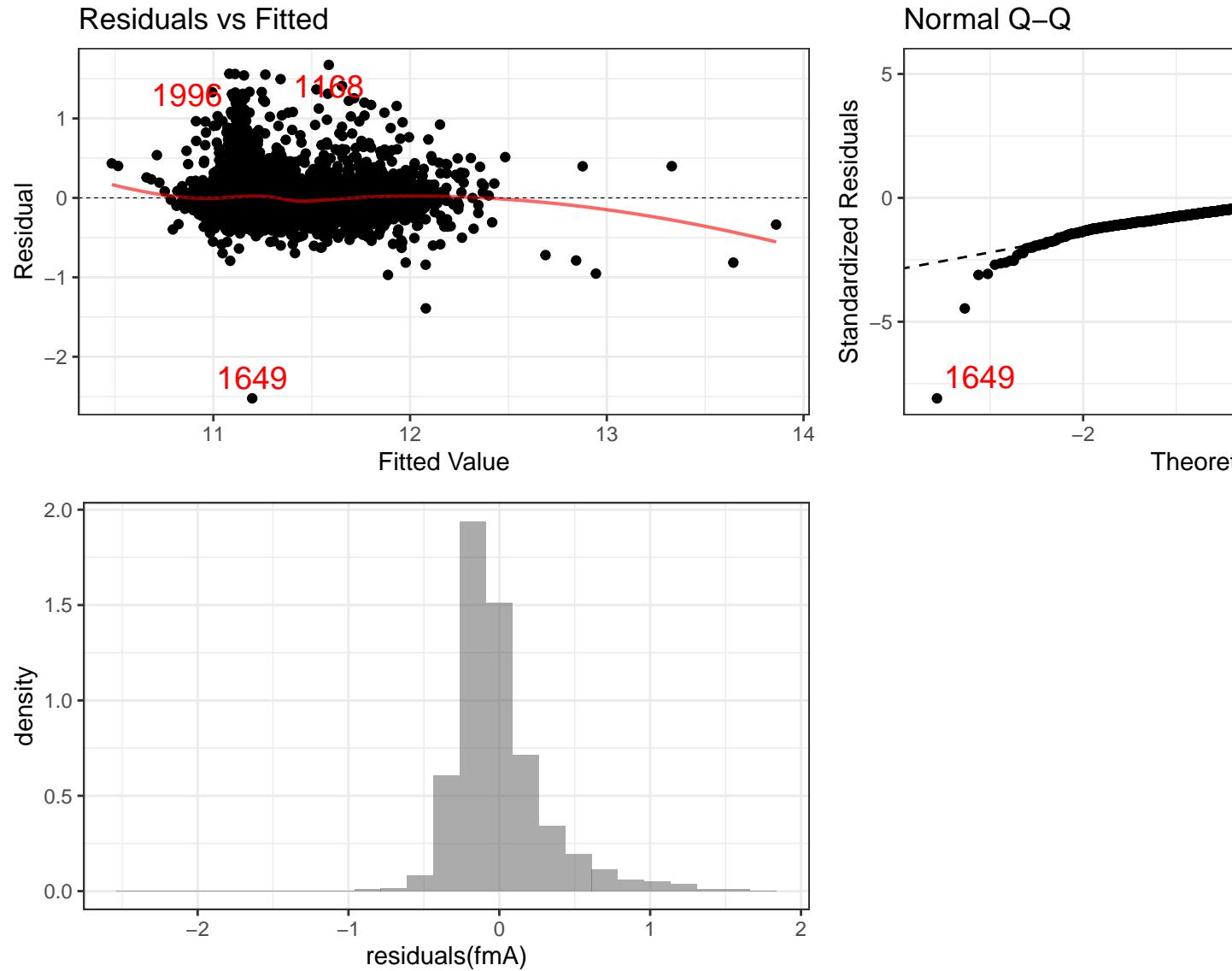
```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## farm_income_2 ~ 1
##
## Final Model:
## farm_income_2 ~ fertilizer_use_2 + pop_dens_2 + grazing_rotation +
##                 farm_acres_2
##
##          Step Df  Deviance Resid. Df Resid. Dev      AIC
## 1                      2812      591.50 -4384.4
## 2 + fertilizer_use_2  1  308.01801  2811      283.48 -6451.4
## 3      + pop_dens_2   1   5.61712  2810      277.86 -6505.7
## 4 + grazing_rotation 1   3.97730  2809      273.89 -6544.3
## 5      + farm_acres_2 1   0.33237  2808      273.55 -6545.7

## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## farm_income_2 ~ farm_acres_2 + pop_dens_2 + grazing_rotation +
##                 fertilizer_use_2 + bee_colonies_2
##
## Final Model:
## farm_income_2 ~ farm_acres_2 + pop_dens_2 + grazing_rotation +
##                 fertilizer_use_2
##
##          Step Df  Deviance Resid. Df Resid. Dev      AIC
## 1                      2807      273.53 -6544.0
## 2 - bee_colonies_2   1  0.027553   2808      273.55 -6545.7
```

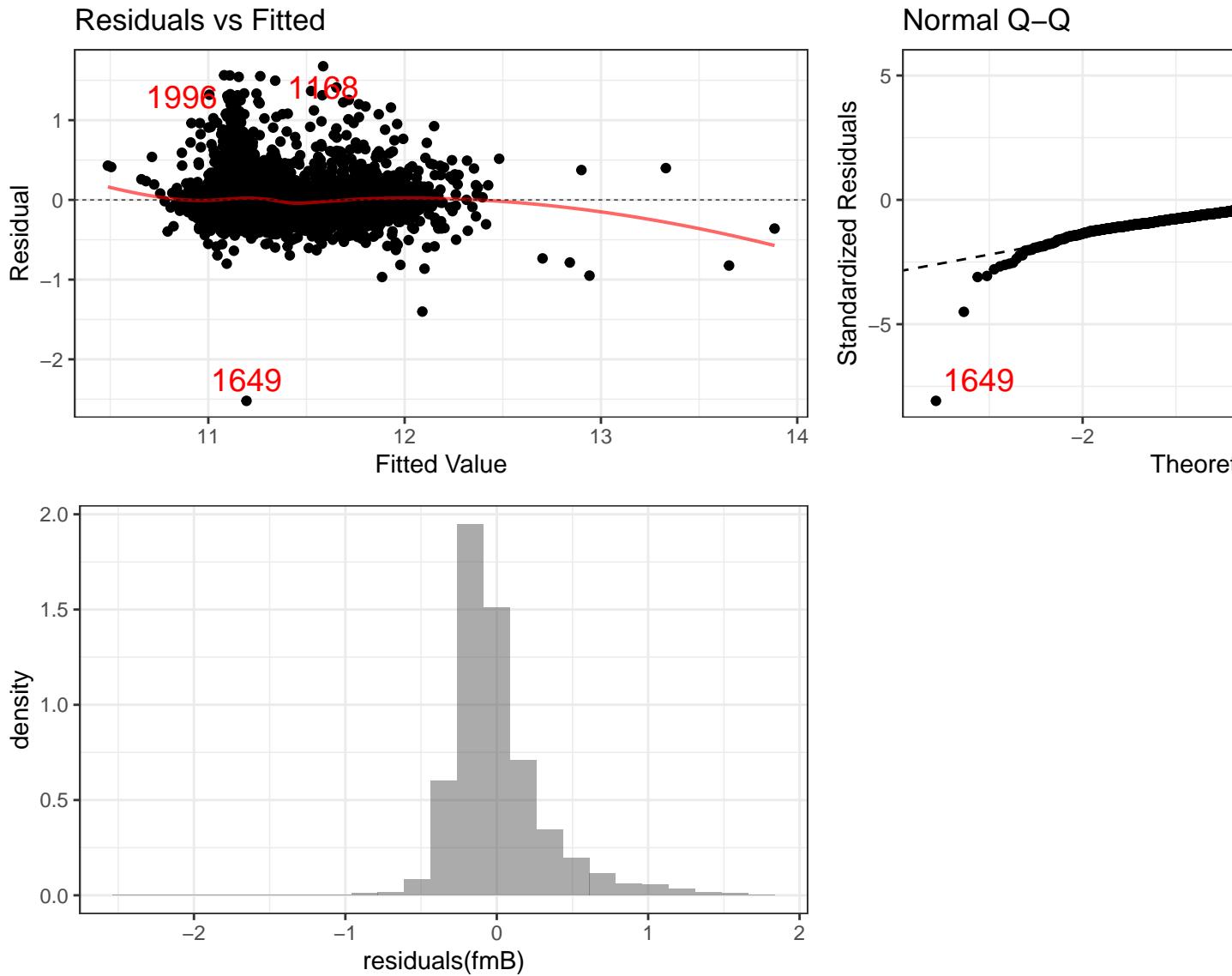
The final model suggested by best subsets regression is the 4-predictor model with pop_dens_2, grazing_rotation, fertilizer_use_2, and farm_acres_2 without bee colonies 2. This model had

the lowest C_p score of 4.2828. This model was also recommended by stepwise regression and backwards elimination, with the lowest AIC score of -6545.7. However, since the 5-predictor model including bee_colonies_2 had similar AIC, and C_p scores, we will manually compare the model strength and significance.

```
## `geom_smooth()` using formula = 'y ~ x'
```



```
## `geom_smooth()` using formula = 'y ~ x'
```



The residuals vs fitted plot, QQ plot, and histogram of the residuals of both models are practically identical and will not serve as a point of difference. The linearity condition is mostly met as there is mostly symmetry over the line residual=0 on the residuals vs fitted values plot, and there does not seem to be major bending or curving of points. The equal variance condition is questionable as there seems to be funneling in of the spread of points on the residual vs fitted value plot as you move to higher fitted values. The normality condition is also questionable as, though the QQ-plot is mostly a straight line with mostly no bending, there does seem to be a very large and bent tail at the high end. Furthermore, the histogram of the residuals, while it is unimodal and centered close to 0, there does seem to be a slight right skew. We will assume independence and proceed. Finally, there appear to be several unusual points in both models. We compare these conditions with those of the model of untransformed data (Appendix 4).

Compare conditions and fit to un-transformed model

Conditions are definitely worse in the untransformed model. The data is even more skewed, with many extreme residual values that disrupt normality and equal variance. Therefore, it looks

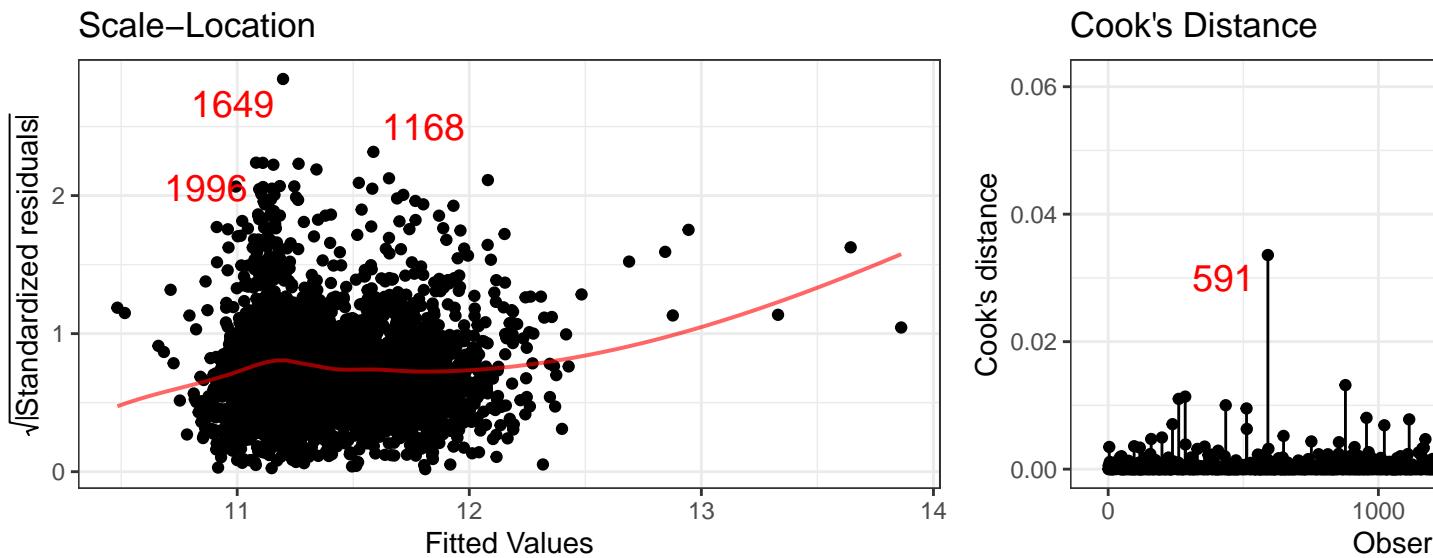
like our transformations were helpful, even if they didn't completely fix the data to fit modeling conditions.

```
##               Estimate Std. Error t value Pr(>|t|) 
## (Intercept) 11.023637  0.024966 441.55 < 2e-16 ***
## pop_dens_2   -0.179728  0.023776 -7.56  5.5e-14 ***
## grazing_rotation -0.772752  0.116352 -6.64  3.7e-11 ***
## fertilizer_use_2  0.005092  0.000129 39.57 < 2e-16 ***
## farm_acres_2    0.001030  0.000558   1.85   0.065 . 
## 
## Residual standard error: 0.312 on 2808 degrees of freedom
## Multiple R-squared:  0.538, Adjusted R-squared:  0.537 
## F-statistic:  816 on 4 and 2808 DF, p-value: <2e-16

##               Estimate Std. Error t value Pr(>|t|) 
## (Intercept) 3.16e+04  2.50e+03 12.64 < 2e-16 ***
## pop_dens     -5.19e+03  1.23e+03 -4.22 2.6e-05 ***
## grazing_rotation -1.12e+05  1.51e+04 -7.45 1.2e-13 ***
## fertilizer_use  2.28e+00  5.82e-02 39.23 < 2e-16 ***
## farm_acres    1.83e+00  9.06e-01   2.02   0.044 * 
## 
## Residual standard error: 43100 on 2808 degrees of freedom
## Multiple R-squared:  0.47, Adjusted R-squared:  0.469 
## F-statistic:  621 on 4 and 2808 DF, p-value: <2e-16
```

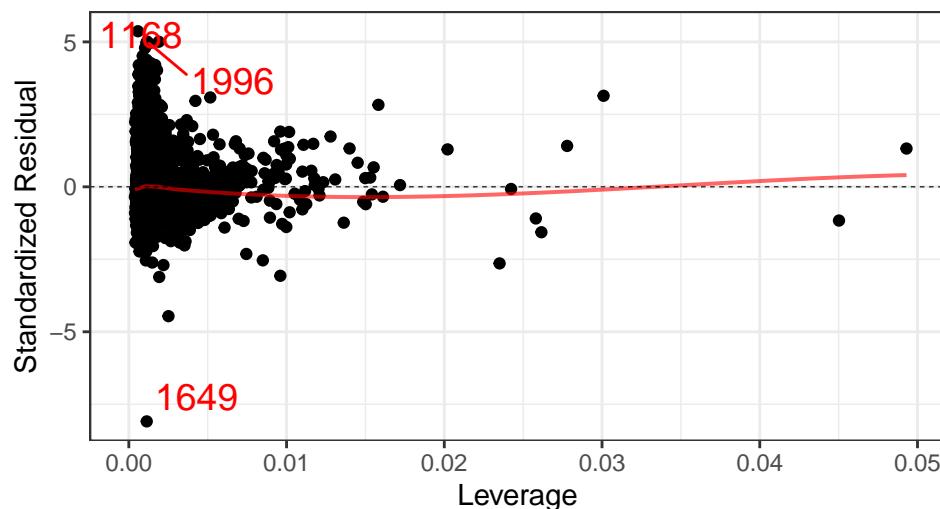
Transforming our variables resulting in reducing our model significance, as `farm_acres` is significant in the untransformed model, with `farm_acres_2` is not significant in the second model. However, the improvement in meeting conditions seems sufficient to justify continuing with one of the transformed models. The transformed model also has a higher R^2_{adj} by 6.8%, which means that the transformations also improved model strength.

```
## 'geom_smooth()' using formula = 'y ~ x'
```



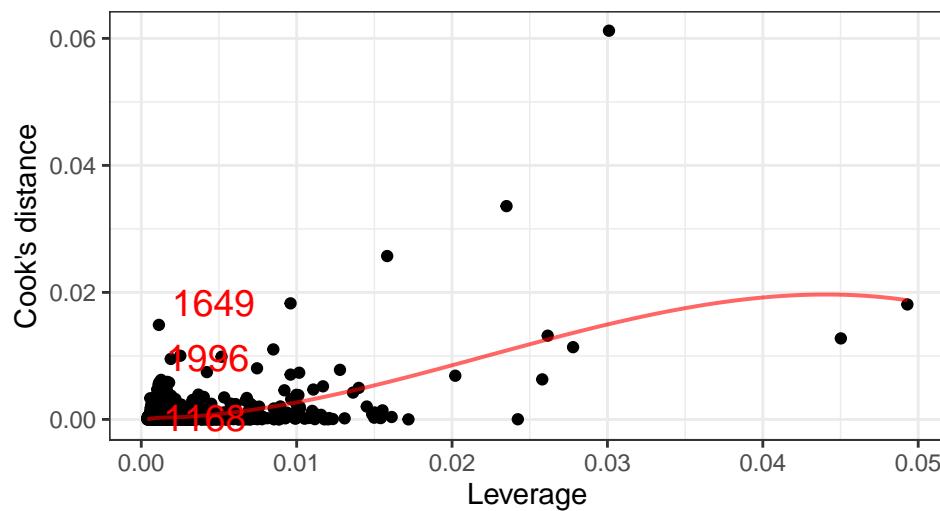
```
## 'geom_smooth()' using formula = 'y ~ x'
```

Residuals vs Leverage



```
## `geom_smooth()` using formula = 'y ~ x'
```

Cook's dist vs Leverage



```
## [1] 236
```

```
## [1] 136
```

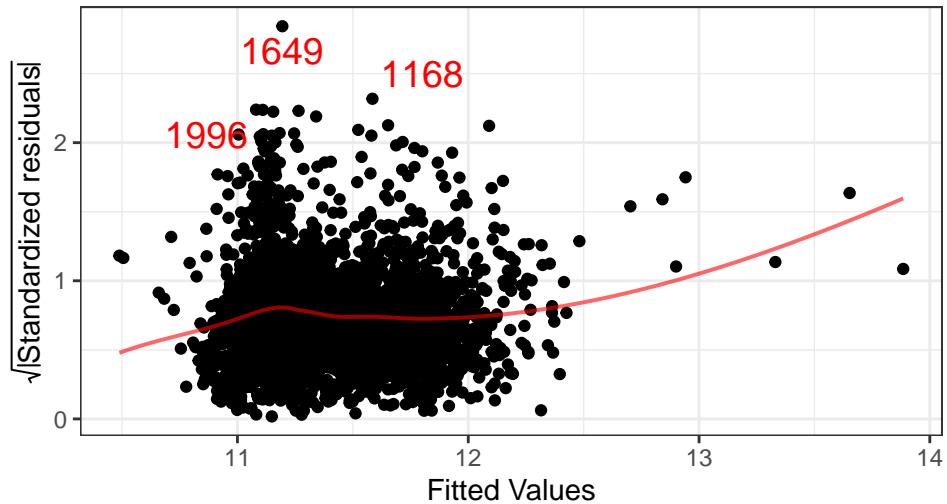
```
## [1] 153
```

```
## [1] 62
```

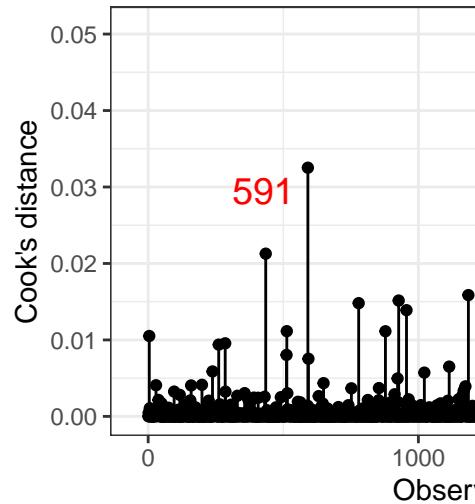
```
## [1] 0
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

Scale–Location

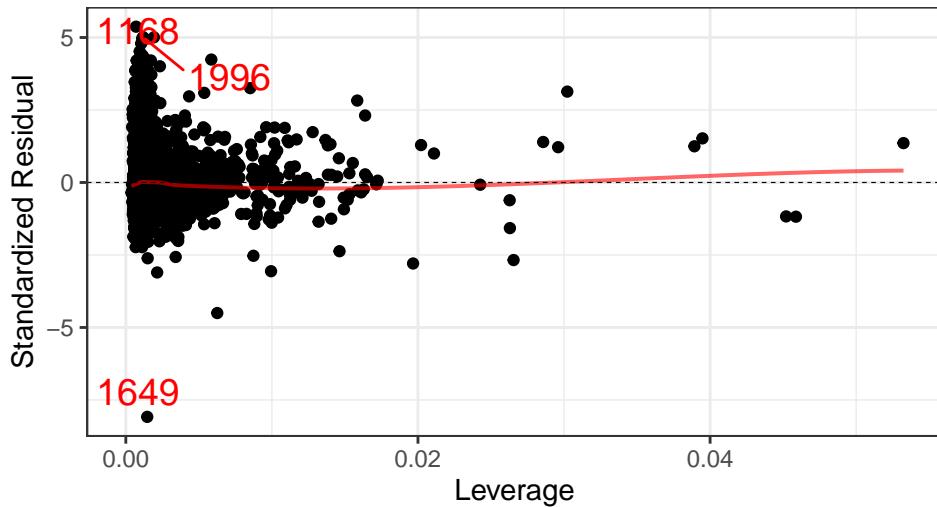


Cook's Distance

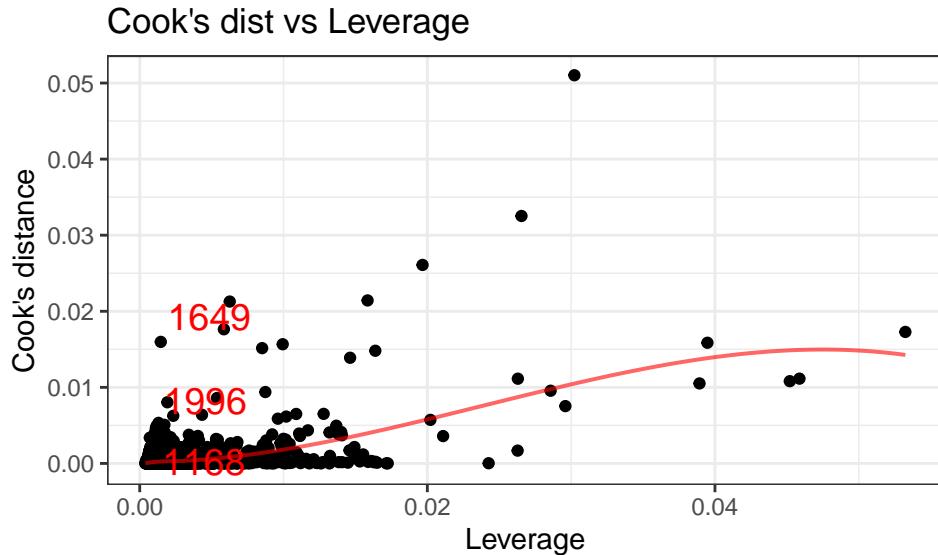


```
## `geom_smooth()` using formula = 'y ~ x'
```

Residuals vs Leverage



```
## `geom_smooth()` using formula = 'y ~ x'
```



```

## [1] 337
## [1] 207
## [1] 153
## [1] 62
## named numeric(0)
## [1] 2813      6
## [1] 0.0035549
## [1] 0.0053324
## [1] 71

```

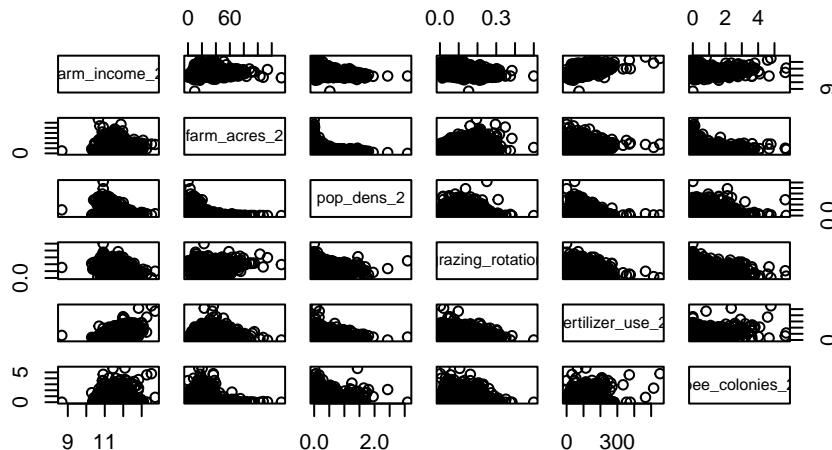
fmA: High leverage would be $2(k+1)/n = 2(4+1)/2813 = 0.0035549$ and problematic leverage would be $3(k+1)/n = 0.0053324$. There are 236 points with high leverages, 136 of which have problematic leverages. There are 91 points with the absolute value of their standardized residuals above 2 but not 3, making them outliers. There are 62 points with an absolute value of their standardized residuals above 3, making them big outliers. There do not appear to be many influential points, as there are no points with Cook's distances above or near 0.5.

fmB: High leverage would be $2(k+1)/n = 2(5+1)/2813 = 0.0035549$ and problematic leverage would be $3(k+1)/n = 0.0053324$. There are 337 points with high leverages, 207 of which have problematic leverages. There are 91 points with the absolute value of their standardized residuals above 2 but not 3, making them outliers. There are 62 points with an absolute value of their standardized residuals above 3, making them big outliers. There do not appear to be many influential points, as there are no points with Cook's distances above or near 0.5. So, fmA seems to have 71 less points with problematic leverages. They have the same number of big outliers, and neither have influential points.

```

##          farm_income_2 farm_acres_2 pop_dens_2 grazing_rotation
## farm_income_2      1.000000   0.250045 -0.200636     -0.430220
## farm_acres_2       0.250045   1.000000 -0.343740      0.056543
## pop_dens_2        -0.200636  -0.343740   1.000000     -0.052888
## grazing_rotation   -0.430220   0.056543 -0.052888      1.000000
## fertilizer_use_2    0.721624   0.299724 -0.144407     -0.517474
## bee_colonies_2      0.024889  -0.128397  0.235198     -0.049328
##          fertilizer_use_2 bee_colonies_2
## farm_income_2       0.721624    0.024889
## farm_acres_2        0.299724   -0.128397
## pop_dens_2         -0.144407    0.235198
## grazing_rotation    -0.517474   -0.049328
## fertilizer_use_2     1.000000    0.063654
## bee_colonies_2      0.063654    1.000000

```

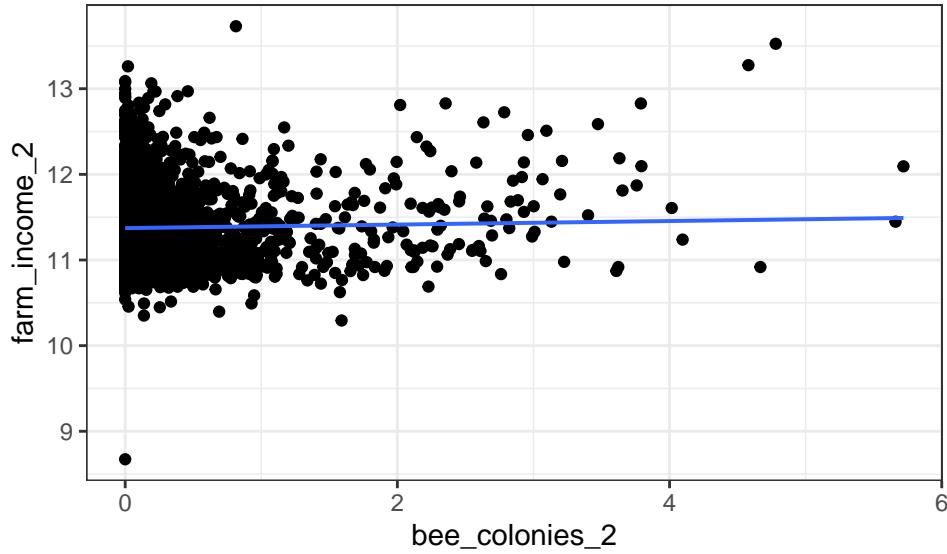


```

##          pop_dens_2 grazing_rotation fertilizer_use_2      farm_acres_2
##              1.1428            1.4723            1.6138            1.2880
##          farm_acres_2      pop_dens_2 grazing_rotation fertilizer_use_2
##              1.2984            1.1975            1.4745            1.6378
##          bee_colonies_2
##              1.0786

```

Thankfully, for both models, none of the VIF values are above 5 which would strongly indicate multicollinearity. The inclusion of bee_colonies_2 in fmB has little impact on VIF. The correlation matrix for the bee_colonies_2 pairs isn't too concerning with correlation coefficients of 0.23 or lower.

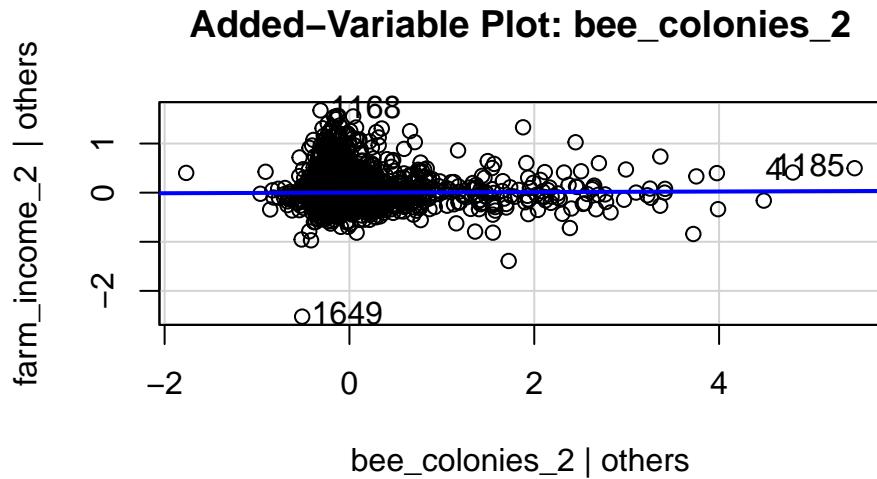


```

## [1] 0.024889

##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.37228   0.00998 1139.36 <2e-16 ***
## bee_colonies_2 0.02070   0.01568    1.32    0.19
##
## Residual standard error: 0.459 on 2811 degrees of freedom
## Multiple R-squared:  0.000619, Adjusted R-squared:  0.000264
## F-statistic: 1.74 on 1 and 2811 DF, p-value: 0.187

```



> Looking at a simple linear regression model of the relationship between farm_income_2 and bee_colonies_2, I see a weak, positive, unclearly linear relationship between farm_income_2 and farm_acres_2 with several outliers and a correlation coefficient of 0.024889. The AV plot of the additional predictor of fmB, bee_colonies_2, shows a horizontal band that is curvilinear. This indicates that bee_colonies_2 contains very little additional information about farm_income_2, beyond that from the other predictors. Both of these indicate that bee_colonies_2 may not be a significant or strong predictor for farm_income_2.

```

## Analysis of Variance Table
##
## Model 1: farm_income_2 ~ pop_dens_2 + grazing_rotation + fertilizer_use_2 +
##           farm_acres_2
## Model 2: farm_income_2 ~ farm_acres_2 + pop_dens_2 + grazing_rotation +
##           fertilizer_use_2 + bee_colonies_2
## Res.Df RSS Df Sum of Sq    F Pr(>F)
## 1    2808 274
## 2    2807 274   1     0.0276 0.28    0.59

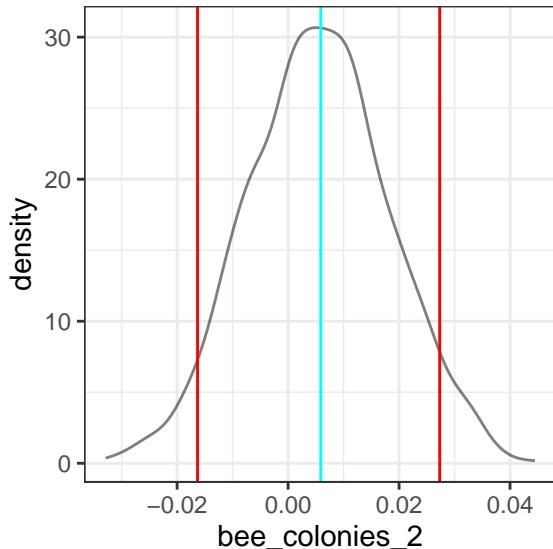
```

$$H_0 : \beta_{farm,acres,2} = 0 \quad H_A : \beta_{farm,acres,2} \neq 0$$

Since the p-value of 0.59 is large, we fail to reject the null hypothesis and do not have enough evidence to conclude the alternative hypothesis that the predictor `bee_colonies_2` is significant.

Bootstrapping

Since the normality condition was not met for either model, we will test the significance of `bee_colonies_2` using bootstrapping. We assumed that our bootstrapped distribution was non-normal and skewed. The resulting 95% confidence interval for the coefficient of `bee_colonies_2` is (-0.016319, 0.027324). Since this interval includes 0 (see figure below), we are not 95% confident that the `bee_colonies_2` is not 0, and conclude that this predictor is not significant.



Final evaluation of 4 and 5 term models

We conclude that fmA, the four term model is preferable to fmB, the five term model that includes bee colonies as a predictor. As described above, fmA and fmB have very similar deviations from model conditions. However, fmB has about 100 more points with high leverage than fmA, indicating that fmB has more potential to make inaccurate predictions than fmA. After controlling for the effects of the other four predictors, bee colonies has an insignificant effect on farm income, as indicated by a nested f test and individual coefficient t stat with p values of 0.59. We double checked this result with a bootstrap test since our model did not fit conditions, but still found

that the effects of bee colonies on farm income were insignificant. Furthermore, as we saw in our initial model selection process, fmA is a stronger model than fmB. Both models have an adjusted R^2 of 0.537, and fmA has lower AIC and Cp scores than fmB. Overall, we found no reasons why fmB is preferable and many reasons why fmA is preferable. We will therefore proceed with analyzing fmA.

Outlier removal

One outlier, at index 1649, stands out in our equal variance plot and has the largest standard residual value in fmA. We tried removing it and reassessing conditions and model fit. See the appendix for our specific model results for this new model, fmA3. Removing this outlier decreased residual variance at the outlier's fitted value and taking out such an extreme value improved data normality. However, both equal variance and normality are still very questionable. Removing this outlier did slightly decrease the significance of pop_dens_2 ($p = 1.1e-13$ compared to $5.5e-14$ with the outlier) and farm_acres_2 ($p = 0.072$ compared to 0.065 with the outlier). However, removing the outlier increased R^2 by 0.5%, from 0.537 to 0.542. Given that taking out this outlier improved model conditions and increased model strength slightly, and since the impacts on model significance were small and less important than improvement to model conditions, we decided that it made sense to leave the outlier out of our model.

Interactions

Finally, we decided to test some interactions in our model. We specifically tested all interactions involving population density, since we thought that the population density of a county could modulate the effects of farm size and farming practices on farm income. For example, more populous counties might have more access to resources that would help them profitably employ fertilizer and grazing rotation practices. See the appendix for specific model results for this model, fmC. We first checked the conditions of our model with these three interaction terms. Equal variance, normality, and linearity looked very similar to our model without interactions. Looking at model significance, we found that all three interactions were significant or close to significant: $p=0.0561$ for pop_dens_2:grazing_rotation; $p=0.0274$ for pop_dens_2:fertilizer_use_2; $p=0.0199$ for pop_dens_2:farm_acres_2. However, adding these interaction terms barely increased the overall strength of our model: R^2 adjusted changed from 0.542 to 0.543. Although these interactions are interesting and fairly significant, they are not strong predictors of farm income and do not improve model conditions. We decided not to include them in our final model in order to simplify our model. This allowed us to interpret the effects of individual predictors on farm income.

Final Conclusion:

Overall, our final model was fmA3, which has outlier 1649 removed and has no interaction terms. This model indicates that 54.2% of the variability in farm_income_2 can be explained by the multiple linear relationship between farm_income_2 and pop_dens_2, grazing_rotation, fertilizer_use_2, and farm_acres_2. Additionally, this model indicates that average farm income in a county increases with increasing population density, proportion of farms that use grazing rotations, average fertilizer use, and average farm size. Specifically, average farm income increases by \$0.31 for every 1 person/acre density increase; farm income increases by \$0.05 for every 10% increase in the percent of farms that use grazing rotations; farm income increases by \$1.00 for every \$1 increase in average farm fertilizer expenditure; and farm income increases by \$1.00 for every 1 acre increase in average farm size. That is, counties with bigger farms that use more fertilizer, counties with more farms that use grazing rotations, and counties with large population densities are likely to contain more profitable farms.

Fitted model equation:

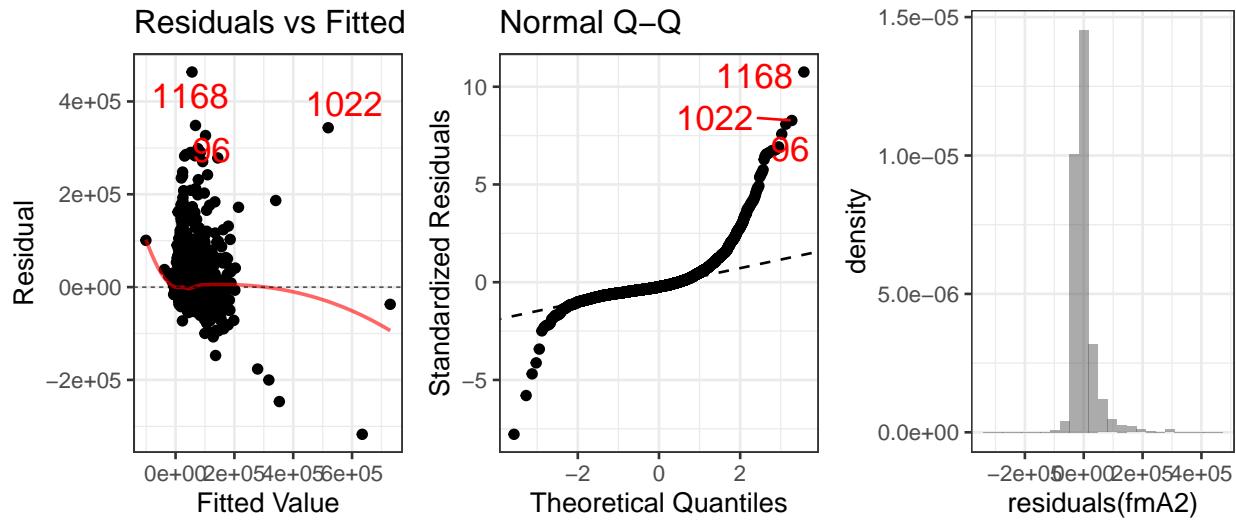
$$\widehat{\text{farm income}} = 61262 + 0.309(\text{popdensity}) + 0.466(\text{grazingrotation}) + 1.005(\text{fertilizeruse}) + 1.001(\text{farmacres})$$

Appendices

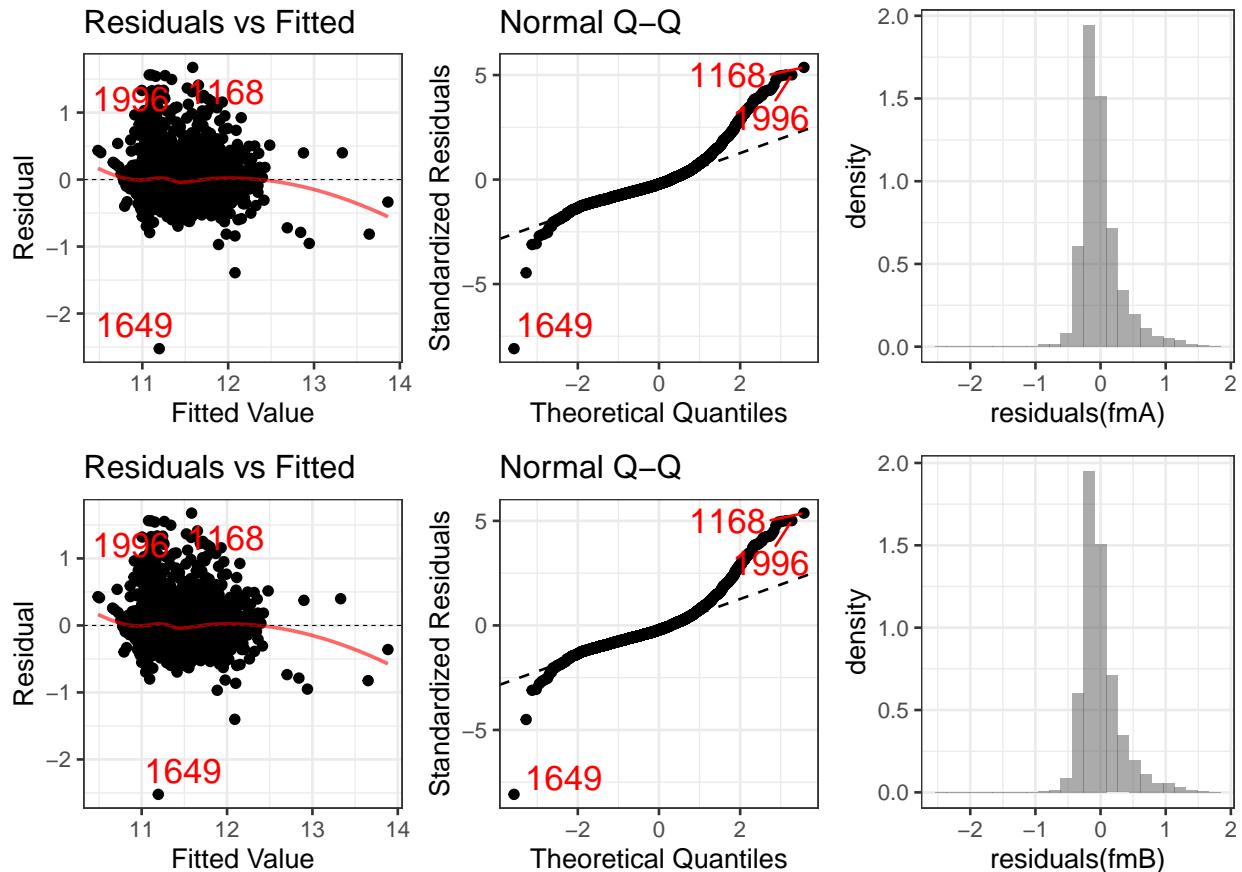
Appendix 1: Data Wrangling

Appendix 2. Plots of observed vs theoretical values, and Q-Q plots for the initial three models

```
## `geom_smooth()` using formula = 'y ~ x'
```



```
## `geom_smooth()` using formula = 'y ~ x'  
## `geom_smooth()` using formula = 'y ~ x'
```



Becca Appendix

Final evaluation of 4 and 5 term models

fmA summary:

```
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 11.023637  0.024966 441.55 < 2e-16 ***
## pop_dens_2   -0.179728  0.023776 -7.56 5.5e-14 ***
## grazing_rotation -0.772752  0.116352 -6.64 3.7e-11 ***
## fertilizer_use_2  0.005092  0.000129 39.57 < 2e-16 ***
## farm_acres_2     0.001030  0.000558    1.85    0.065 .  
## 
## Residual standard error: 0.312 on 2808 degrees of freedom
## Multiple R-squared:  0.538, Adjusted R-squared:  0.537 
## F-statistic: 816 on 4 and 2808 DF, p-value: <2e-16
```

fmB summary:

```
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 11.02278   0.02502 440.55 < 2e-16 ***
## farm_acres_2  0.00106   0.00056   1.89    0.059 .  
## 
```

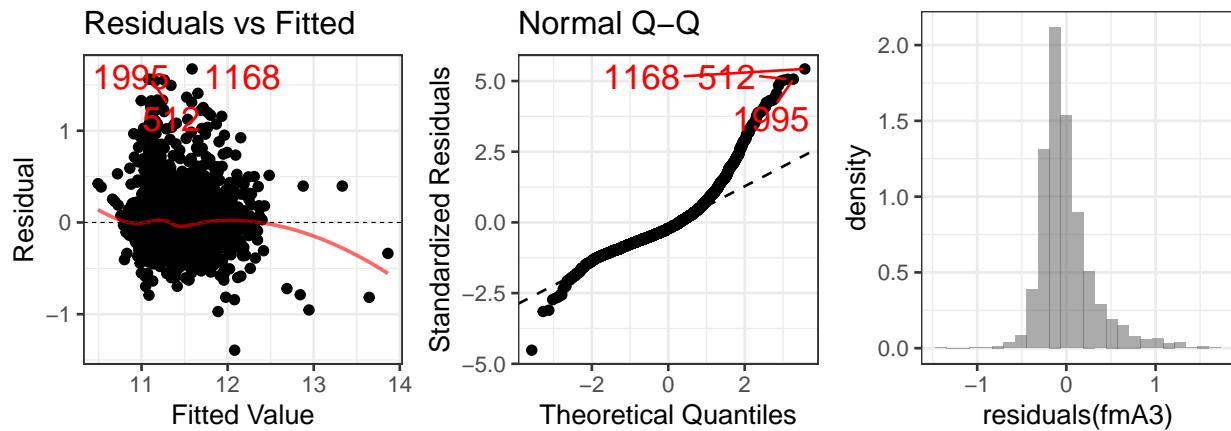
```

## pop_dens_2      -0.18250   0.02434   -7.50  8.7e-14 ***
## grazing_rotation -0.77515   0.11646   -6.66  3.4e-11 ***
## fertilizer_use_2  0.00508   0.00013   39.21 < 2e-16 ***
## bee_colonies_2    0.00590   0.01109    0.53   0.595
##
## Residual standard error: 0.312 on 2807 degrees of freedom
## Multiple R-squared:  0.538, Adjusted R-squared:  0.537
## F-statistic:  653 on 5 and 2807 DF, p-value: <2e-16

```

Evaluation of model without outlier, fmA3

```
## `geom_smooth()` using formula = 'y ~ x'
```



fmA3 summary:

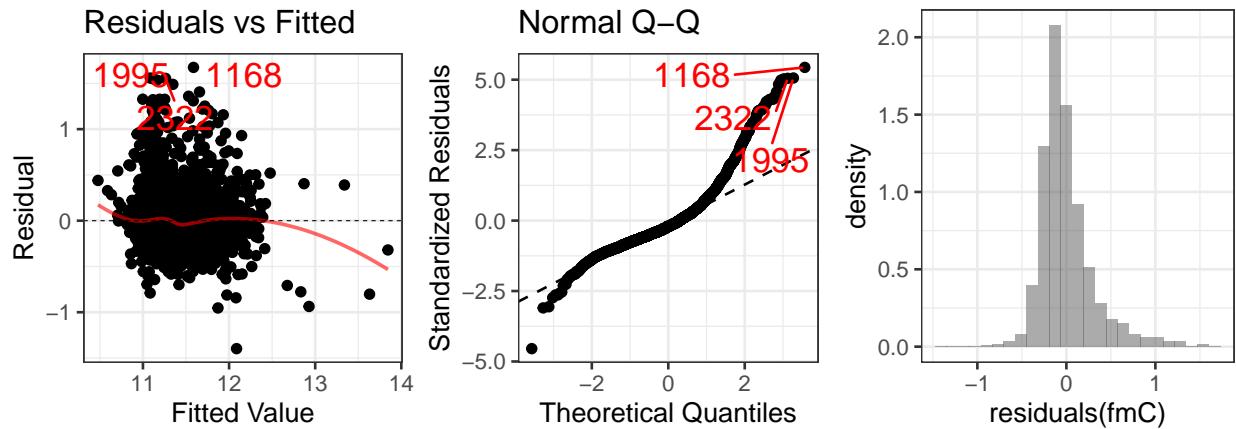
```

##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)  11.022920  0.024677 446.68 < 2e-16 ***
## pop_dens_2   -0.175508  0.023507  -7.47 1.1e-13 ***
## grazing_rotation -0.763733  0.115014  -6.64 3.7e-11 ***
## fertilizer_use_2  0.005098  0.000127  40.07 < 2e-16 ***
## farm_acres_2    0.000991  0.000551    1.80   0.072 .
##
## Residual standard error: 0.309 on 2807 degrees of freedom
## Multiple R-squared:  0.543, Adjusted R-squared:  0.542
## F-statistic:  833 on 4 and 2807 DF, p-value: <2e-16

```

Evaluation of model with interactions, fmC

```
## `geom_smooth()` using formula = 'y ~ x'
```



fmC summary:

```
##                                     Estimate Std. Error t value Pr(>|t|) 
## (Intercept)                 11.049113   0.026524 416.56 < 2e-16 ***
## pop_dens_2                  -0.249320   0.077285  -3.23  0.0013 ** 
## grazing_rotation             -0.888741   0.129364  -6.87  7.9e-12 ***
## fertilizer_use_2              0.004965   0.000144 34.48 < 2e-16 ***
## farm_acres_2                 0.001302   0.000561   2.32  0.0203 *  
## pop_dens_2:grazing_rotation  0.729898   0.381994   1.91  0.0561 .  
## pop_dens_2:fertilizer_use_2   0.001493   0.000677   2.21  0.0274 *  
## pop_dens_2:farm_acres_2      -0.012375   0.005313  -2.33  0.0199 * 
## 
## Residual standard error: 0.308 on 2804 degrees of freedom
## Multiple R-squared:  0.544, Adjusted R-squared:  0.543 
## F-statistic: 478 on 7 and 2804 DF, p-value: <2e-16
```

Reverse transforming variables

```
coef.intercept<-fmA3$coefficients["(Intercept)"]
exp(coef.intercept)
```

```
## (Intercept)
##       61262

#population density
coef.pop<-fmA3$coefficients["pop_dens_2"]
exp(coef.pop)/exp(1)
```

```
## pop_dens_2
## 0.30866
```

```
#grazing
coef.graze<-fmA3$coefficients["grazing_rotation"]
exp(coef.graze)/1

## grazing_rotation
##          0.46592

#fertilizer
coef.fert<-fmA3$coefficients["fertilizer_use_2"]
exp(coef.fert)/(1^2)

## fertilizer_use_2
##          1.0051

#acreage
coef.acre<-fmA3$coefficients["farm_acres_2"]
exp(coef.acre)/(1^2)

## farm_acres_2
##          1.001
```