Nils Baker Case Study

Chaman Preet Kaur and Chris Atterbury

October 15, 2015

1 Executive Summary

This section is to be written last and contains a short (approx 250 words) summary of the case study as a whole.

2 Background

The financial success of a bank depends on its total holdings, which in turn depend on the size of its customer base. The case study revolves around Nils Baker, vice president of a regional retail bank in the US. Based on the feedback received from a promising customer, he develops a hypothesis that having more physical branches motivates people to open a checking account with the bank. In order to make a strategic business plan based on this hypothesis, it needs to be proven statistically. The data available for analysis includes: total households in a particular area, total number of checking accounts with the bank in that area and the presence of a physical branch or an ATM in that area. Setting up a physical branch involves a lot of investment and there are many maintenance fees, but if it can be proven to help gain more checking accounts in turn expanding the customer base and adding to the the total holdings, it can be a profitable business move. This is what we explore here.

3 Models

3.1 Initial Considerations

The data that we have received to pursue Nils Baker's hypothesis includes four columns. Initially, they are *ID*, *Total.Households.in.Area*, *Households.with.Account*, and *Inside.Outside.Footprint*. First, we drop the last two rows of the data set because they are empty and change the column names to *ID*, *Total.Households*, *Accounts*, and *Footprint*, respectively. Next, further preprocessing of the data is performed so that we have numeric data (see **Additional Codeblock 1** in the *Appendix* for the code used to do this).
The first column, *ID*, is identical to the row number and, thus, will not be considered in this case study. Each row itself is a separate Metropolitan Statistical Area (MSA). For our purposes, we need only understand that each row contains a diferent geographic region. The next two columns, *Total.Households* and *Accounts*, contain the number of total households and the number of those households that have a checking account with the bank. The last column, *Footprint*, was originally coded with either "Inside" or "Outside." A region was considered "Inside" if there was both a physical bank location and an ATM in the region. This was recoded as 1 for easier analysis of the data. A region was considered "Outside" if there was not a physical bank location and only an ATM in the region. This was recoded as 0.
Finally, we will add a column called *Accts.Hsehld* that takes the *Accounts* column and divides it by the *Total.Households* column. We may want to choose this as the response variable since this allows us to get closer to evaluating Nils Baker's hypothesis without as much interpretation of the model.
With the data in a usable form, we can now view the correlation matrix to get an initial feel for the relationships amongst the variables.

Figure 1

```
##                          ID Total.Households    Accounts  Footprint
## ID                1.0000000      -0.67957529 -0.66181262  0.3030236
## Total.Households -0.6795753       1.00000000  0.91121152 -0.3004534
## Accounts         -0.6618126       0.91121152  1.00000000 -0.2171381
## Footprint         0.3030236      -0.30045335 -0.21713807  1.0000000
## Accts.Hsehld      0.1732239      -0.08685308  0.07169036  0.1523331
##                  Accts.Hsehld
## ID                 0.17322388
## Total.Households  -0.08685308
## Accounts           0.07169036
## Footprint          0.15233308
## Accts.Hsehld       1.00000000
```

We can ignore the values involving `ID` since this is basically just a running counter of which number region we see. Its correlation with `Total.Households` is based on the fact that the data set's observations are organized by the number of households in the region with the largest first. We see that `Accounts` has a strong correlation with `Total.Households`, which is reasonable since we expect that regions with more households will have more accounts, just based on volume. Also, this explains the negative correlation between `Accounts` and `ID`. The numbers indicate what we would expect for `Footprint`; the correlations with the households and accounts do not show anything. That is, the locations of physical banks were not chosen based on the number of households or accounts in a region. This correlation matrix ends up showing us that trying to predict `Accts.Hsehld` may be difficult, but hopefully combining a few features will help. Please see **Additional Figure 1** in the *Appendix* for the pairs plot that shows these relationships graphically.

3.2 Procedure

The goal in this case study, as stated by Nils Baker, is to ascertain whether or not the presence of a physical bank in a region increases the likelihood of a given household possessing a checking account. We will start with Simple Linear Regression (SLR) models, before considering more complex Multiple Linear Regression (MLR) models.

3.3 SLR Models

3.3.1 Predicting Number of Accounts

We will start by evaluating SLR models for `Accounts`. Even though these models may be more difficult to interpret in terms of Nils Baker's hypothesis, if we find an especially good model, then we may make an exception. The R^2 values obtained for these three models were as follows:

Figure 2

Model	R^2	R^2_a	p-value
Accounts ~ Total.Households	0.8303	0.8289	$<$ 2.2e-16
Accounts ~ Footprint	0.04715	0.03907	0.01721
Accounts ~ Accts.Hsehld	0.00514	-0.003292	0.4365

(see **Additional Codeblock 2** in the *Appendix* for the code)
Considering the results of the correlation matrix above, none of these results are particularly surprising. The total number of households in a region proves to be a good predictor of the number of accounts in the region.

The coefficient of determination, denoted R^2, is high, so we know that a significant amount of the variation in `Accounts` is explained by `Total.Households`. Additionally, the p-value for the coefficient and the model are significant. However, this model only serves to highlight the drawbacks of choosing `Accounts` as the response variable since we can only determine that regions with more households have more accounts. We can say nothing about how likely a given household is to have an account.
The other two models, which use `Footprint` and `Accts.Hsehld`, are exceptionally poor. `Footprint` is the feature that most interests us since it tells us if a physical bank is in the region. We would have been surprised to see the presence of a physical bank be a good predictor of the gross number of accounts, so seeing that it is not fits in with what we would expect. Now we will look at SLR models predicting the number of accounts per household in a region.

3.3.2 Predicting Accounts per Household

Now we look at how each of the individual potential predictors does with `Accts.Hsehld` as the response variable. By using this response variable, we can easily understand which potential predictors are involved in determining the likelihood of a given household having an account with the bank. The R^2 values obtained for these three models were as follows:

Figure 3

Model	R^2	R^2_a	p-value
Accts.Hsehld ~ Total.Households	0.007543	-0.0008672	0.3456
Accts.Hsehld ~ Accounts	0.00514	-0.003292	0.4365
Accts.Hsehld ~ Footprint	0.02321	0.01493	0.09672

(see **Additional Codeblock 3** in the *Appendix* for the code) Once again, these results are aligned with those from the correlation matrix. All three SLR models are quite poor. Their respective R^2 values are very low, and their respective p-values are high. With a higher p-value we cannot reject the null hypothesis that these models are just as good as the average at predicting `Accts.Hsehld`. Interestingly, the R^2 and p-values are best, albeit still terrible, for the SLR that used `Footprint`. Depending on what we find when looking at MLR models, Nils Baker's hypothesis still has a chance at being correct.
Now we will turn our attention to MLR models to try to find one that does a decent job at predicting `Accts.Hsehld`.

3.4 MLR Models

We feel that the `Accts.Hsheld` is a more appropriate response variable than `Accounts` since it is a ratio of total accounts to the total number of households in that area. Thus, we explore MLR models with `Accts.Hsheld` initially.

3.4.1 Models with Accts.Hsheld as response variable

3.4.1.1 Initial Considerations MLR models can get quite complicated. We can look at how different terms interact with each other, quadratic terms, etc. It is imperative that we keep Nils Barker's hypothesis in mind as we look at MLR models. We want to see how well `Footprint` predicts `Accts.Hsehld`. Thus, we will focus on generating models that will help us answer that question.

3.4.1.2 Ramsey RESET Test The Ramsey RESET Test attempts to determine if quadratic terms add value to a model. It does not take into consideration the interaction between quadratic terms and other variables, so this will only help to give us an intuition. The results of the test are below.

Figure 4

```
##
##  RESET test
##
## data:  lm(Accts.Hsehld ~ Accounts + Total.Households, data = d)
## RESET = 1.9605e-05, df1 = 1, df2 = 116, p-value = 0.9965
```

```
##
##  RESET test
##
## data:  lm(Accts.Hsehld ~ Accounts + Total.Households + Footprint, data = d)
## RESET = 0.039219, df1 = 1, df2 = 115, p-value = 0.8434
```

We see that the p-value for the first RESET test is about as bad as is possible at 0.9965. The second is not much better. This test was run just for possible quadratic terms. We will not evaluate any higher order terms with such a poor value. We did not run the test with `Footprint` in the model initially because a qualitative variable cannot be squared. Thus, we expect that quadratic terms will be most valuable for their interactions, if at all. Now we turn our attention to actual MLR models.

3.4.1.3 Footprint with one other term In the correlation matrix (**Figure 1** above) we saw that `Accounts` and `Total.Households` have a very high correlation (0.911). Thus, we will explore models with only one or the other.
The code and summaries for each of the models explored can be found in **Additional Codeblock 4** in the *Appendix*. However, they will not be displayed here because none offered promising results. The R^2 value only goes up to 0.119. When R^2 is adjusted to account for the number of predictors, R_a^2 only gets up to a miniscule 0.08036 . With these dispiriting results we turn our attention to MLR models with both `Accounts` and `Total.Households`.

3.4.1.4 Using all potential predictors We know that error cannot be unexplained. Thus, we can look at the R^2 value with all possible quadratic and interaction terms to get a sense of just how high we can hope to get that value when building our model and then start to delete variables that are insignificant. Additionally, we can compare R_a^2 to make sure that our chosen model improves on this. See **Additional Codeblock 5 and 6** in the *Appendix* for the full summaries.
For the model `Accts.Hsehld ~ (. - ID + I(Accounts^2) + I(Total.Households^2))^2` we get $R^2 = 0.6141$ and $R_a^2 = 0.5585$. On comparing this model with a less bloated model, `Accts.Hsehld ~ . - ID + Footprint:Accounts + Total.Households:Footprint + I(Accounts^2) + I(Total.Households^2) + Footprint:I(Accounts^2) + Footprint:I(Total.Households^2)`, we find all coefficients are significant at the $\alpha = 0.05$ level. That is, we can be at least 95% confident in the significance of these coefficients. $R^2 = 0.5706$ and $R_a^2 = 0.5354$. These values are close to the obese one above. Also, the interaction terms with `Footprint` are all the most significant ones.
Finally, we turn to a peculiar model that highlights the problem with having a response variable built off of two of its potential predictors. When the model is `log(Accts.Hsehld) ~ log(Accounts) + log(Total.Households)`, or any variation that contains those three elements, all variation is explained and $R^2 = R_a^2 = 1$ (see **Codeblock 7** in the *Appendix* for the full summary). This becomes obvious when looking at the math below.

Proof

Let $A = Accounts$ and $H = Total.Households$.
Thus $Accts.Hsehld = \frac{A}{H}$.
Then, $\log \frac{A}{H} = b_0 + b_1 \log A + b_2 \log H \iff \log A - \log H = b_0 + b_1 \log A + b_2 \log H$.
If we let $b_0 = 0$, $b_1 = 1$, $b_2 = -1$, then both sides are equal.
Thus, we know we know both sides of the regression equation will always be equal. □

This highlights the shortcoming of this approach and must cause us to reconsider it when we only have these three potential predictors.
The low R^2 values for the other models force us to reconsider our response variable too. We have seen previously that `Accounts` regressed on `Total.Households` gives the best R^2 value so far of 0.8303. Thus, we further explored models with `Accounts` as the response variable.

3.4.2 Models with Accounts as response variable

3.4.2.1 Ramsey RESET Test We first check if there is a possibility of a higher order term with the Ramsey RESET Test.

Figure 5

```
##
##  RESET test
##
## data:  lm(Accounts ~ Total.Households + Footprint, data = d)
## RESET = 4.4374, df1 = 1, df2 = 116, p-value = 0.03732
```

```
##
##  RESET test
##
## data:  lm(Accounts ~ Total.Households + Footprint, data = d)
## RESET = 3.955, df1 = 1, df2 = 116, p-value = 0.04909
```

The p-value of 0.03732 obtained from the test indicates the potential value of second order term. The p-value of 0.04909 obtained for the potential third order term indicates that an even higher order term might be valuable too.

3.4.2.2 Exploring all models We explored many models (see **Additional Codeblock 8** in the *Appendix* for the summary) with degree 2 and degree 3 terms and obtained the following results for R^2 and R_a^2:

Figure 6

Model	R^2	R_a^2
Accounts ~ Total.Households + Footprint + I(Total.Households^2)	0.8415	0.8374
Accounts ~ Total.Households + Footprint + Total.Households:Footprint	0.8923	0.8895
Accounts ~ Total.Households + Footprint + I(Total.Households^2) + Total.Households:Footprint	0.8934	0.8897
Accounts ~ Total.Households + Footprint + I(Total.Households^2)		

Model	R^2	R_a^2
+ I(Total.Households^3)		
+ Total.Households:Footprint	0.8957	0.8911
Accounts ~ Total.Households + Footprint		
+ I(Total.Households^2)		
+ I(Total.Households^2):Footprint	0.8882	0.8843
Accounts ~ Total.Households + Footprint		
+ I(Total.Households^2)		
+ Total.Households:Footprint		
+ I(Total.Households^2):Footprint	0.8934	0.8887

Looking at the values for all these models we find `Accounts ~ Total.Households + Footprint + Total.Households:Footprint` as the most parsemonius model with high R^2 and R_a^2 values. Thus we choose it to be the best model that can explain the data. There are some common problems with all the models listed in **Figure 6** as the they are not normal and the residual plots do not look very convincing for any of these (see **Figures 6** and **Figure 7** below, where this is discussed in more detail). However, the given the number of observations and potential predictors, we give this model our attention.

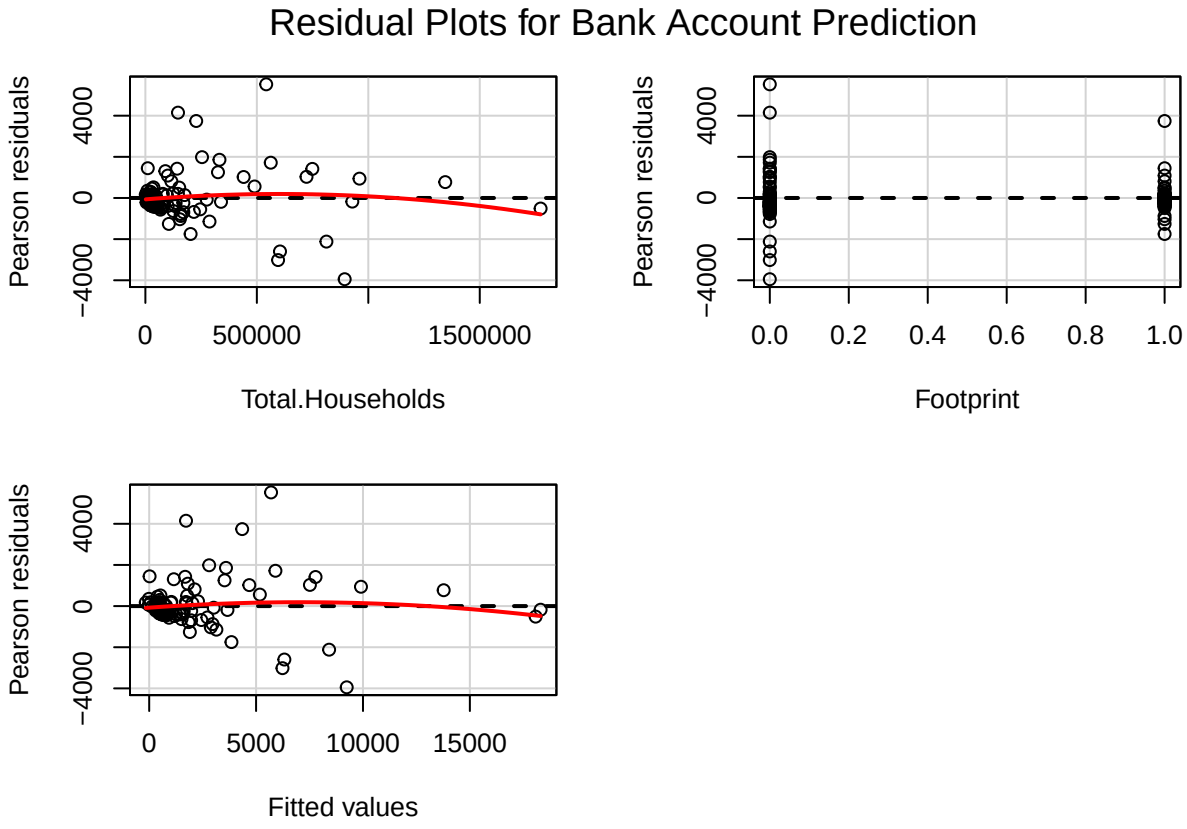
3.5 Results

The best model that we could generate from the given data was `Accounts ~ Total.Households + Footprint + Total.Households:Footprint`.

3.5.1 Homoskedasticity and Normality of error terms

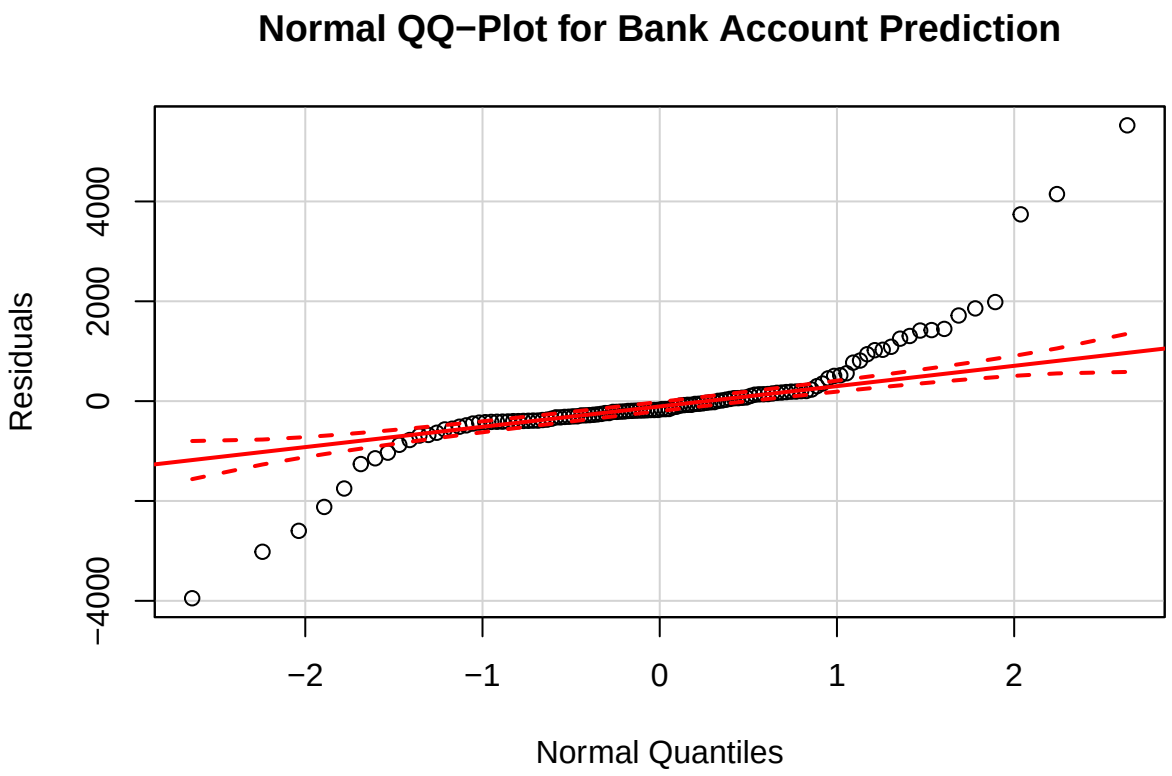
On plotting the residuals and looking for the normality of the residuals, we find that the data is not normal and the error terms do not quite show constant variance.

Figure 7




```
##                   Test stat Pr(>|t|)
## Total.Households     -1.097    0.275
## Footprint            -0.644    0.521
## Tukey test           -1.042    0.298
```

Figure 8



The non-constant variance of the error terms is further confirmed by using the Breusch-Pagan test. It shows that we must reject the null hypothesis that error terms are homoskedastic (i.e. have constant variance).

Figure 9

```
ncvTest(lm(Accounts ~ Total.Households + Footprint +
                  Total.Households:Footprint, data = d))
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 53.32331    Df = 1     p = 2.829296e-13
```

```
qchisq(.95, 1)
```

```
## [1] 3.841459
```

3.5.2 Significance of Footprint

The p-value of `Footprint` is relatively low at 0.063 (see `lm_9` in **Codeblock 8** of the *Appendix* for the full summary of the model). While not below the standard $\alpha = 0.05$, it is very close and the interaction term with `Footprint` is extremely significant with a p-value of 1.48e-12. The other models that we tried tended to show this pattern as well (see **Codeblock 8** in the *Appendix* for the full summary of each model). That is, `Footprint` is not insignificant, but its real value is in its interaction terms. This plays a key role in determinig if having a physical branch motivates people to open checking accounts with the bank.
The coefficient for `Footprint` is negative. In fact, it is negative enough that it drops the intercept below zero when there is a physical bank in a given region. The interaction term increases the slope. At the very least, this sends mixed messages about the significance of `Footprint`, which will need to be explored in the *Conclusions* section below.

4 Conclusions

As stated at the beginning of this case study, our goal is to ascertain if the presence of a physical bank influences the likelihood of persons to hold an account with the bank. Our data set gives us the number of households, the number of accounts, and whether or not a physical bank resides within a given region. We looked at models that tried to predict a compound variable, the number of acconts per household, and models that tried to predict the total number of accounts. While we would have liked to be able to use the compound variable for its ease of interpretation, this was not possible. Finally, we found a model that did a good job of predicting the number of accounts based on the number of households, whether or not a physical bank is located in the region, and the interaction term between the two.

In drawing conclusions from our models we must be frank, no model was wholly satisfying. The few models explained a very significant amount of the variance in the number of accounts. The ones that did lacked in other areas. Our chosen model lacked the desired normality of error terms at its tails and their variance was not as constant as desired. One may look to **Codeblock 8** in the *Appendix* to see that models that should have corrected these issues were deficient in other areas and often did not correct the problems satisfactorally anyway. Thus, we can draw no conclusions with the desired certainty for which we hope.

We are able to make some conjectures about Nils Baker's hypothesis. As noted in the *Results* section, interaction terms with `Footprint` tended to be rather significant, even when the other was a higher order term. `Footprint` itself tended to have some significance, usually when an interaction term was included. Since we are predicting the gross number of accounts, and we know that the number of households in the region is highly correlated with it, the interaction terms are the strongest signal that `Footprint` may have some impact on the number of accounts. However, we cannot escape the fact that coefficient for `Footprint` is negative. When predicting the number of accounts per household, `Footprint` was usually positive. However, when predicting `Accounts`, about half gave `Footprint` a negative coefficient, mostly the higher order ones. This makes ambiguity inescapable when trying to draw conclusions.

To make any confident conclusions about whether or not the presence of a physical bank impacts the number of accounts, we require more data. More observations may help, but there are likely more factors at work in predicting the number or accounts, or accounts per household. A more successful model might try to predict accounts per household using `Footprint`, but also using something like average income in the region. Also, Nils Baker's hypothesis itself does not take into account costs associated with opening up a physical branch. This might suggest that we want a very high correlation between accounts per household and presence of a physical bank in a region to justify opening more physical bank locations.

Thus, we use the significance of the interaction terms to conjecture that there may be some sort of tenuous relationship between the presence of a physical bank and the likelihood of households opening up accounts. However, the coefficients we found in models for `Footprint` suggest that we need further information to draw any true conclusions.

5 Appendix

5.1 Additional Codeblocks

Additional Codeblock 1

```
d <- read.csv("41330723.csv", header = TRUE, stringsAsFactors = FALSE)
d <- d[1:120, ] # last two rows contain no data
names(d) <- c("ID", "Total.Households", "Accounts", "Footprint")

for (i in 2:3) {
  d[[i]] <- gsub(",", "", d[[i]])
  d[[i]] <- as.numeric(d[[i]])
}
d[["Footprint"]][d[["Footprint"]] == "Outside"] <- 0
```

```
d[["Footprint"]][d[["Footprint"]] == "Inside"] <- 1
d[["Footprint"]] <- as.numeric(d[["Footprint"]])
d[["ID"]] <- as.numeric(d[["ID"]])
d[["Accts.Hsehld"]] <- d[["Accounts"]] / d[["Total.Households"]]
```

```
head(d)
```

```
##   ID Total.Households Accounts Footprint Accts.Hsehld
## 1  1          1772960    17563         0  0.009906033
## 2  2          1345209    14547         0  0.010813933
## 3  3           960434    10847         0  0.011293853
## 4  4           928274    18133         1  0.019534103
## 5  5           893995     5291         0  0.005918378
## 6  6           812137     6297         0  0.007753618
```

Additional Codeblock 2

```
lm_1 <- lm(Accounts ~ Total.Households, data = d)
summary(lm_1)
```

```
##
## Call:
## lm(formula = Accounts ~ Total.Households, data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4645.8  -358.6  -223.9    77.7  7823.8
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.260e+02  1.447e+02   1.562    0.121
## Total.Households 1.086e-02  4.521e-04  24.029   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1366 on 118 degrees of freedom
## Multiple R-squared:  0.8303, Adjusted R-squared:  0.8289
## F-statistic: 577.4 on 1 and 118 DF,  p-value: < 2.2e-16
```

```
lm_2 <- lm(Accounts ~ Footprint, data = d)
summary(lm_2)
```

```
##
## Call:
## lm(formula = Accounts ~ Footprint, data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2540.2 -1778.4  -819.1   -40.1 16943.4
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2627.2      395.4   6.645 9.86e-10 ***
## Footprint    -1437.6      594.9  -2.416   0.0172 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3236 on 118 degrees of freedom
## Multiple R-squared:  0.04715,    Adjusted R-squared:  0.03907
## F-statistic: 5.839 on 1 and 118 DF,  p-value: 0.01721
```

```
lm_3 <- lm(Accounts ~ Accts.Hsehld, data = d)
summary(lm_3)
```

```
##
## Call:
## lm(formula = Accounts ~ Accts.Hsehld, data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2671.2 -1594.0 -1433.4   143.6 16047.8
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1742.8      439.6   3.965 0.000126 ***
## Accts.Hsehld  17527.1    22448.6   0.781 0.436504
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3307 on 118 degrees of freedom
## Multiple R-squared:  0.00514,    Adjusted R-squared:  -0.003292
## F-statistic: 0.6096 on 1 and 118 DF,  p-value: 0.4365
```

Additional Codeblock 3

```
lm_4 <- lm(Accts.Hsehld ~ Total.Households, data = d)
summary(lm_4)
```

```
##
## Call:
## lm(formula = Accts.Hsehld ~ Total.Households, data = d)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.010135 -0.006645 -0.003025  0.001904  0.121737
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.492e-02  1.432e-03  10.423   <2e-16 ***
## Total.Households -4.235e-09  4.472e-09  -0.947    0.346
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.01351 on 118 degrees of freedom
## Multiple R-squared:  0.007543,   Adjusted R-squared:  -0.0008672
## F-statistic: 0.8969 on 1 and 118 DF,  p-value: 0.3456
```

```
lm_5 <- lm(Accts.Hsehld ~ Accounts, data = d)
summary(lm_5)
```

```
##
## Call:
## lm(formula = Accts.Hsehld ~ Accounts, data = d)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.009282 -0.006456 -0.003580  0.002018  0.122535
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.365e-02  1.444e-03   9.454 3.94e-16 ***
## Accounts    2.932e-07  3.756e-07   0.781    0.437
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01353 on 118 degrees of freedom
## Multiple R-squared:  0.00514,    Adjusted R-squared:  -0.003292
## F-statistic: 0.6096 on 1 and 118 DF,  p-value: 0.4365
```

```
lm_6 <- lm(Accts.Hsehld ~ Footprint, data = d)
summary(lm_6)
```

```
##
## Call:
## lm(formula = Accts.Hsehld ~ Footprint, data = d)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.011787 -0.006085 -0.003160  0.002384  0.120077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.012411   0.001637   7.580 8.62e-12 ***
## Footprint   0.004125   0.002464   1.674   0.0967 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0134 on 118 degrees of freedom
## Multiple R-squared:  0.02321,    Adjusted R-squared:  0.01493
## F-statistic: 2.803 on 1 and 118 DF,  p-value: 0.09672
```

Additional Codeblock 4

```
summary(lm(Accts.Hsehld ~ Footprint + Total.Households, data = d))
```

```
##
## Call:
## lm(formula = Accts.Hsehld ~ Footprint + Total.Households, data = d)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.011920 -0.006122 -0.003293  0.002291  0.119948
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       1.293e-02  1.979e-03   6.536 1.72e-09 ***
## Footprint         3.758e-03  2.592e-03   1.450    0.150
## Total.Households -2.202e-09  4.667e-09  -0.472    0.638
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01345 on 117 degrees of freedom
## Multiple R-squared:  0.02506,    Adjusted R-squared:  0.008395
## F-statistic: 1.504 on 2 and 117 DF,  p-value: 0.2266
```

```
summary(lm(Accts.Hsehld ~ Footprint + Total.Households +
             Footprint:Total.Households, data = d))
```

```
##
## Call:
## lm(formula = Accts.Hsehld ~ Footprint + Total.Households + Footprint:Total.Households,
##     data = d)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.011634 -0.006183 -0.003389  0.002127  0.120225
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 1.306e-02  2.022e-03   6.461 2.54e-09 ***
## Footprint                   3.299e-03  2.922e-03   1.129    0.261
## Total.Households           -2.754e-09  4.950e-09  -0.556    0.579
## Footprint:Total.Households  5.284e-09  1.532e-08   0.345    0.731
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0135 on 116 degrees of freedom
## Multiple R-squared:  0.02606,    Adjusted R-squared:  0.0008719
## F-statistic: 1.035 on 3 and 116 DF,  p-value: 0.38
```

```
summary(lm(Accts.Hsehld ~ Footprint + Total.Households +
             Footprint:Total.Households + I(Total.Households^2),
             data = d))
```

```
##
```

```
## Call:
## lm(formula = Accts.Hsehld ~ Footprint + Total.Households + Footprint:Total.Households +
##     I(Total.Households^2), data = d)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.011703 -0.006198 -0.003313  0.002467  0.120160
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  1.336e-02  2.363e-03   5.652 1.17e-07 ***
## Footprint                    3.091e-03  3.056e-03   1.011    0.314
## Total.Households            -5.533e-09  1.246e-08  -0.444    0.658
## I(Total.Households^2)        2.172e-15  8.934e-15   0.243    0.808
## Footprint:Total.Households   6.161e-09  1.580e-08   0.390    0.697
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01355 on 115 degrees of freedom
## Multiple R-squared:  0.02656,    Adjusted R-squared:  -0.007298
## F-statistic: 0.7844 on 4 and 115 DF,  p-value: 0.5375
```

```
summary(lm(Accts.Hsehld ~ Footprint + Total.Households +
             Footprint:Total.Households + I(Total.Households^2) +
             Footprint:I(Total.Households^2),
           data = d))
```

```
##
## Call:
## lm(formula = Accts.Hsehld ~ Footprint + Total.Households + Footprint:Total.Households +
##     I(Total.Households^2) + Footprint:I(Total.Households^2),
##     data = d)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.012253 -0.006095 -0.003269  0.002092  0.119637
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     1.329e-02  2.381e-03   5.583 1.62e-07 ***
## Footprint                       3.836e-03  3.870e-03   0.991    0.324
## Total.Households               -4.936e-09  1.265e-08  -0.390    0.697
## I(Total.Households^2)           1.706e-15  9.090e-15   0.188    0.851
## Footprint:Total.Households     -9.514e-09  5.207e-08  -0.183    0.855
## Footprint:I(Total.Households^2)  1.769e-14  5.597e-14   0.316    0.753
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01361 on 114 degrees of freedom
## Multiple R-squared:  0.02741,    Adjusted R-squared:  -0.01524
## F-statistic: 0.6426 on 5 and 114 DF,  p-value: 0.6676
```

```r
summary(lm(Accts.Hsehld ~ Footprint + Accounts, data = d))
```

```
##
## Call:
## lm(formula = Accts.Hsehld ~ Footprint + Accounts, data = d)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.011271 -0.005797 -0.003439  0.002640  0.119953
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.123e-02  1.916e-03   5.860 4.35e-08 ***
## Footprint   4.772e-03  2.520e-03   1.894   0.0607 .
## Accounts    4.497e-07  3.806e-07   1.182   0.2397
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01338 on 117 degrees of freedom
## Multiple R-squared:  0.03472,    Adjusted R-squared:  0.01822
## F-statistic: 2.104 on 2 and 117 DF,  p-value: 0.1265
```

```r
summary(lm(Accts.Hsehld ~ Footprint + Accounts +
             Footprint:Accounts, data = d))
```

```
##
## Call:
## lm(formula = Accts.Hsehld ~ Footprint + Accounts + Footprint:Accounts,
##     data = d)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.015178 -0.005548 -0.003203  0.002283  0.119780
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         1.193e-02  2.024e-03   5.896 3.74e-08 ***
## Footprint           3.328e-03  2.855e-03   1.166    0.246
## Accounts            1.821e-07  4.548e-07   0.400    0.690
## Footprint:Accounts 8.906e-07  8.296e-07   1.074    0.285
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01337 on 116 degrees of freedom
## Multiple R-squared:  0.04422,    Adjusted R-squared:  0.0195
## F-statistic: 1.789 on 3 and 116 DF,  p-value: 0.1531
```

```r
summary(lm(Accts.Hsehld ~ Footprint + Accounts +
             Footprint:Accounts + I(Accounts^2),
             data = d))
```

```
##
```

```
## Call:
## lm(formula = Accts.Hsehld ~ Footprint + Accounts + Footprint:Accounts +
##     I(Accounts^2), data = d)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.011156 -0.005046 -0.003119  0.002268  0.118062
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         9.681e-03  2.239e-03   4.324 3.27e-05 ***
## Footprint           3.825e-03  2.818e-03   1.357   0.1773
## Accounts            2.240e-06  1.037e-06   2.161   0.0328 *
## I(Accounts^2)      -1.593e-10  7.238e-11  -2.201   0.0298 *
## Footprint:Accounts  1.435e-06  8.529e-07   1.683   0.0951 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01316 on 115 degrees of freedom
## Multiple R-squared:  0.08284,    Adjusted R-squared:  0.05094
## F-statistic: 2.597 on 4 and 115 DF,  p-value: 0.03992
```

```
summary(lm(Accts.Hsehld ~ Footprint + Accounts +
             Footprint:Accounts + I(Accounts^2) +
             Footprint:I(Accounts^2),
           data = d))
```

```
##
## Call:
## lm(formula = Accts.Hsehld ~ Footprint + Accounts + Footprint:Accounts +
##     I(Accounts^2) + Footprint:I(Accounts^2), data = d)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.014667 -0.005026 -0.003006  0.001939  0.115490
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               1.109e-02  2.298e-03   4.826 4.36e-06 ***
## Footprint                -2.095e-04  3.343e-03  -0.063   0.9501
## Accounts                  9.525e-07  1.181e-06   0.806   0.4218
## I(Accounts^2)            -5.963e-11  8.485e-11  -0.703   0.4836
## Footprint:Accounts        6.617e-06  2.539e-06   2.607   0.0104 *
## Footprint:I(Accounts^2) -3.380e-10  1.563e-10  -2.163   0.0326 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01295 on 114 degrees of freedom
## Multiple R-squared:  0.119,  Adjusted R-squared:  0.08036
## F-statistic:  3.08 on 5 and 114 DF,  p-value: 0.01206
```

Additional Codeblock 5

```r
lm_all <- lm(Accts.Hsehld ~ (. - ID + I(Accounts^2) + I(Total.Households^2))^2,
             data = d)
summary(lm_all)
```

```
##
## Call:
## lm(formula = Accts.Hsehld ~ (. - ID + I(Accounts^2) + I(Total.Households^2))^2,
##     data = d)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.023668 -0.002990 -0.000442  0.002224  0.067033
##
## Coefficients:
##                                       Estimate Std. Error t value
## (Intercept)                           1.421e-02  2.309e-03   6.155
## Total.Households                     -2.303e-07  6.991e-08  -3.294
## Accounts                              1.649e-05  5.187e-06   3.179
## Footprint                             4.773e-03  3.204e-03   1.490
## I(Accounts^2)                        -1.645e-09  1.657e-09  -0.992
## I(Total.Households^2)                 4.947e-13  1.847e-13   2.678
## Total.Households:Accounts            -8.893e-12  2.363e-11  -0.376
## Total.Households:Footprint           -5.193e-07  1.193e-07  -4.353
## Total.Households:I(Accounts^2)       -8.068e-15  5.562e-15  -1.451
## Total.Households:I(Total.Households^2) -5.887e-19  2.365e-19  -2.490
## Accounts:Footprint                    3.012e-05  5.980e-06   5.037
## Accounts:I(Accounts^2)                2.793e-13  1.975e-13   1.415
## Accounts:I(Total.Households^2)        8.509e-17  5.784e-17   1.471
## Footprint:I(Accounts^2)              -3.212e-09  7.623e-10  -4.213
## Footprint:I(Total.Households^2)       1.009e-12  3.746e-13   2.693
## I(Accounts^2):I(Total.Households^2)   8.509e-22  6.355e-22   1.339
##                                      Pr(>|t|)
## (Intercept)                          1.43e-08 ***
## Total.Households                      0.00135 **
## Accounts                              0.00195 **
## Footprint                             0.13937
## I(Accounts^2)                         0.32332
## I(Total.Households^2)                 0.00860 **
## Total.Households:Accounts             0.70747
## Total.Households:Footprint           3.15e-05 ***
## Total.Households:I(Accounts^2)        0.14990
## Total.Households:I(Total.Households^2)  0.01438 *
## Accounts:Footprint                   2.00e-06 ***
## Accounts:I(Accounts^2)                0.16018
## Accounts:I(Total.Households^2)        0.14428
## Footprint:I(Accounts^2)              5.38e-05 ***
## Footprint:I(Total.Households^2)       0.00825 **
## I(Accounts^2):I(Total.Households^2)   0.18352
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.008973 on 104 degrees of freedom
## Multiple R-squared:  0.6141, Adjusted R-squared:  0.5585
```

```
## F-statistic: 11.03 on 15 and 104 DF,  p-value: 1.699e-15
```

Additional Codeblock 6

```
summary(lm(Accts.Hsehld ~ . - ID + Footprint:Accounts +
             Total.Households:Footprint + I(Accounts^2) +
             I(Total.Households^2) + Footprint:I(Accounts^2) +
             Footprint:I(Total.Households^2),
           data = d))
```

```
##
## Call:
## lm(formula = Accts.Hsehld ~ . - ID + Footprint:Accounts + Total.Households:Footprint +
##     I(Accounts^2) + I(Total.Households^2) + Footprint:I(Accounts^2) +
##     Footprint:I(Total.Households^2), data = d)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.024939 -0.003457 -0.001472  0.002135  0.068849
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     1.182e-02  1.653e-03   7.150 1.01e-10 ***
## Total.Households               -9.749e-08  2.366e-08  -4.121 7.32e-05 ***
## Accounts                        9.509e-06  2.362e-06   4.026 0.000104 ***
## Footprint                       7.859e-03  2.680e-03   2.932 0.004096 **
## I(Accounts^2)                  -5.234e-10  1.846e-10  -2.836 0.005439 **
## I(Total.Households^2)           5.402e-14  1.860e-14   2.904 0.004450 **
## Accounts:Footprint              3.403e-05  4.873e-06   6.984 2.30e-10 ***
## Total.Households:Footprint     -6.379e-07  7.951e-08  -8.023 1.22e-12 ***
## Footprint:I(Accounts^2)        -3.203e-09  5.120e-10  -6.255 7.78e-09 ***
## Footprint:I(Total.Households^2)  1.244e-12  1.672e-13   7.443 2.33e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.009204 on 110 degrees of freedom
## Multiple R-squared:  0.5706, Adjusted R-squared:  0.5354
## F-statistic: 16.24 on 9 and 110 DF,  p-value: < 2.2e-16
```

Additional Codeblock 7

```
summary(lm(log(Accts.Hsehld) ~ log(Accounts) + log(Total.Households),
           data = d))
```

```
##
## Call:
## lm(formula = log(Accts.Hsehld) ~ log(Accounts) + log(Total.Households),
##     data = d)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -5.944e-15 -1.305e-15 -2.050e-16  5.510e-16  5.545e-14
```

```
##
## Coefficients:
##                          Estimate Std. Error    t value Pr(>|t|)
## (Intercept)           -5.189e-15  5.577e-15 -9.300e-01    0.354
## log(Accounts)          1.000e+00  8.972e-16  1.115e+15   <2e-16 ***
## log(Total.Households) -1.000e+00  9.426e-16 -1.061e+15   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.344e-15 on 117 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 6.265e+29 on 2 and 117 DF,  p-value: < 2.2e-16
```

Additional Codeblock 8

```
lm_7 <- lm(Accounts ~ Total.Households + Footprint, data = d)
summary(lm_7)
```

```
##
## Call:
## lm(formula = Accounts ~ Total.Households + Footprint, data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4626.9  -452.8  -213.8   184.0  7423.0
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      7.718e+00  1.997e+02   0.039    0.969
## Total.Households 1.109e-02  4.710e-04  23.535   <2e-16 ***
## Footprint        4.122e+02  2.616e+02   1.576    0.118
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1357 on 117 degrees of freedom
## Multiple R-squared:  0.8338, Adjusted R-squared:  0.831
## F-statistic: 293.6 on 2 and 117 DF,  p-value: < 2.2e-16
```

```
lm_8 <- lm(Accounts ~ Total.Households + Footprint + I(Total.Households^2),
           data = d)
summary(lm_8)
```

```
##
## Call:
## lm(formula = Accounts ~ Total.Households + Footprint + I(Total.Households^2),
##     data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4994.2  -402.2  -129.2   255.3  6973.0
##
## Coefficients:
```

```
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)           -2.467e+02  2.233e+02  -1.105   0.2715
## Total.Households       1.359e-02  1.153e-03  11.786   <2e-16 ***
## Footprint              5.354e+02  2.618e+02   2.045   0.0431 *
## I(Total.Households^2) -2.028e-09  8.543e-10  -2.374   0.0193 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1331 on 116 degrees of freedom
## Multiple R-squared:  0.8415, Adjusted R-squared:  0.8374
## F-statistic: 205.3 on 3 and 116 DF,  p-value: < 2.2e-16
```

```
lm_9 <- lm(Accounts ~ Total.Households + Footprint +
           Total.Households:Footprint, data = d)
summary(lm_9)
```

```
##
## Call:
## lm(formula = Accounts ~ Total.Households + Footprint + Total.Households:Footprint,
##     data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3948.3  -381.3  -168.7   167.7  5524.4
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 2.515e+02  1.644e+02   1.530    0.129
## Total.Households            1.005e-02  4.025e-04  24.980  < 2e-16 ***
## Footprint                  -4.460e+02  2.376e+02  -1.877    0.063 .
## Total.Households:Footprint  9.880e-03  1.245e-03   7.933 1.48e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1097 on 116 degrees of freedom
## Multiple R-squared:  0.8923, Adjusted R-squared:  0.8895
## F-statistic: 320.3 on 3 and 116 DF,  p-value: < 2.2e-16
```

```
lm_10 <- lm(Accounts ~ Total.Households + Footprint + I(Total.Households^2) +
             Total.Households:Footprint, data = d)
summary(lm_10)
```

```
##
## Call:
## lm(formula = Accounts ~ Total.Households + Footprint + I(Total.Households^2) +
##     Total.Households:Footprint, data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4113.9  -332.0  -109.1   191.9  5315.0
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)                 1.442e+02  1.912e+02   0.754    0.452
## Total.Households            1.107e-02  1.008e-03  10.977  < 2e-16 ***
## Footprint                  -3.700e+02  2.473e+02  -1.496    0.137
## I(Total.Households^2)      -7.929e-10  7.228e-10  -1.097    0.275
## Total.Households:Footprint  9.560e-03  1.278e-03   7.480 1.61e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1097 on 115 degrees of freedom
## Multiple R-squared:  0.8934, Adjusted R-squared:  0.8897
## F-statistic: 240.9 on 4 and 115 DF,  p-value: < 2.2e-16
```

```
lm_11 <- lm(Accounts ~ Total.Households + Footprint + I(Total.Households^2) +
            I(Total.Households^3) + Total.Households:Footprint, data = d)
summary(lm_11)
```

```
##
## Call:
## lm(formula = Accounts ~ Total.Households + Footprint + I(Total.Households^2) +
##     I(Total.Households^3) + Total.Households:Footprint, data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3702.8  -309.0   -56.3   206.3  5224.9
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -2.202e+01  2.175e+02  -0.101   0.9195
## Total.Households            1.387e-02  2.044e-03   6.785 5.46e-10 ***
## Footprint                  -3.299e+02  2.470e+02  -1.335   0.1844
## I(Total.Households^2)      -6.398e-09  3.640e-09  -1.758   0.0814 .
## I(Total.Households^3)       2.426e-15  1.544e-15   1.571   0.1189
## Total.Households:Footprint  9.736e-03  1.275e-03   7.636 7.48e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1090 on 114 degrees of freedom
## Multiple R-squared:  0.8957, Adjusted R-squared:  0.8911
## F-statistic: 195.7 on 5 and 114 DF,  p-value: < 2.2e-16
```

```
lm_12 <- lm(Accounts ~ Total.Households + Footprint + I(Total.Households^2) +
            I(Total.Households^2):Footprint, data = d)
summary(lm_12)
```

```
##
## Call:
## lm(formula = Accounts ~ Total.Households + Footprint + I(Total.Households^2) +
##     I(Total.Households^2):Footprint, data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4252.4  -309.6  -175.1   128.7  5187.8
##
```

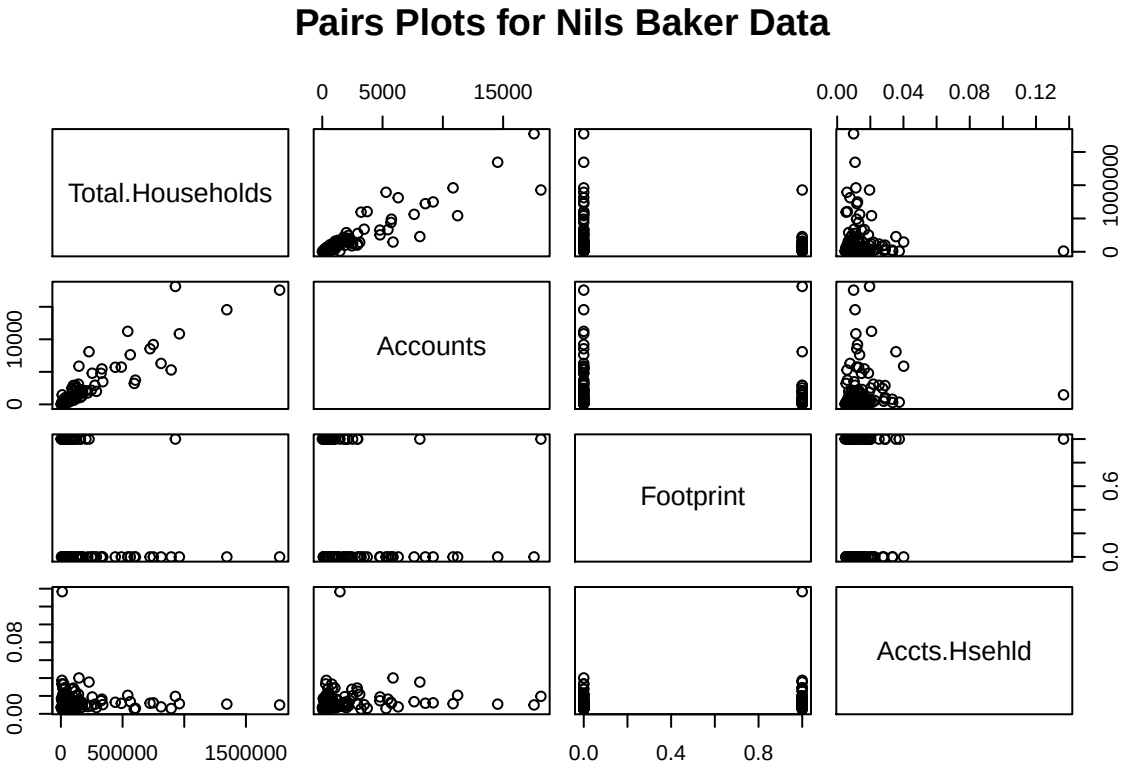
```
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     7.166e+01  1.939e+02   0.370    0.712
## Total.Households                1.164e-02  1.013e-03  11.492  < 2e-16 ***
## Footprint                       1.280e+02  2.285e+02   0.560    0.577
## I(Total.Households^2)          -1.169e-09  7.312e-10  -1.599    0.113
## Footprint:I(Total.Households^2)  9.750e-09  1.407e-09   6.930 2.59e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1123 on 115 degrees of freedom
## Multiple R-squared:  0.8882, Adjusted R-squared:  0.8843
## F-statistic: 228.4 on 4 and 115 DF,  p-value: < 2.2e-16
```

```
lm_13 <- lm(Accounts ~ Total.Households + Footprint + I(Total.Households^2) +
             Total.Households:Footprint + I(Total.Households^2):Footprint,
            data = d)
summary(lm_13)
```

```
##
## Call:
## lm(formula = Accounts ~ Total.Households + Footprint + I(Total.Households^2) +
##     Total.Households:Footprint + I(Total.Households^2):Footprint,
##     data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4111.6  -331.4  -106.7   182.9  5317.9
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      1.456e+02  1.927e+02   0.756   0.4514
## Total.Households                 1.105e-02  1.024e-03  10.793   <2e-16 ***
## Footprint                       -3.875e+02  3.132e+02  -1.237   0.2185
## I(Total.Households^2)           -7.819e-10  7.357e-10  -1.063   0.2902
## Total.Households:Footprint       9.928e-03  4.214e-03   2.356   0.0202 *
## Footprint:I(Total.Households^2) -4.160e-10  4.530e-09  -0.092   0.9270
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1101 on 114 degrees of freedom
## Multiple R-squared:  0.8934, Adjusted R-squared:  0.8887
## F-statistic: 191.1 on 5 and 114 DF,  p-value: < 2.2e-16
```

5.2 Additional Figures

Additional Figure 1


Pairs Plots for Nils Baker Data
Total.Households
Accounts
Footprint
Accts.Hsehld
0
5000
15000
0.00
0.04
0.08
0.12
0
1000000
0
10000
0.0
0.6
0.00
0.08
0
500000
1500000
0.0
0.4
0.8