Devon Morgan

Due: Tuesday, May 1st

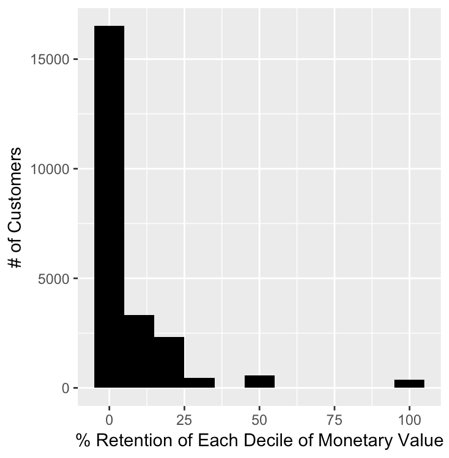
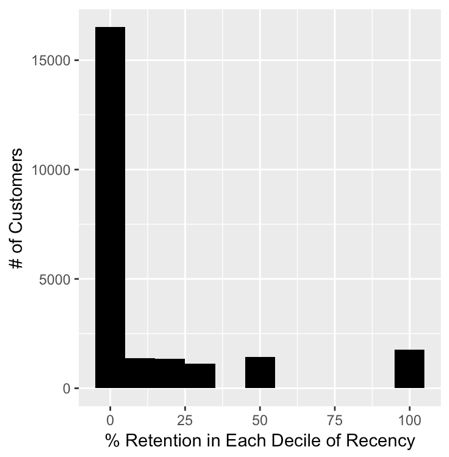
CDNow Case Assignment

**Part 1: Traditional RFM – Decile Analysis**

1. For both recency and monetary, create new columns that place each customer into a decile based on ow they rank. Calculate and plot the percent retention in each decile of recency and percent retention in each decile of monetary. Do these variables seem to make any difference to retention?

Based on the graphs below, recency and monetary value do not seem to make any

difference to retention.



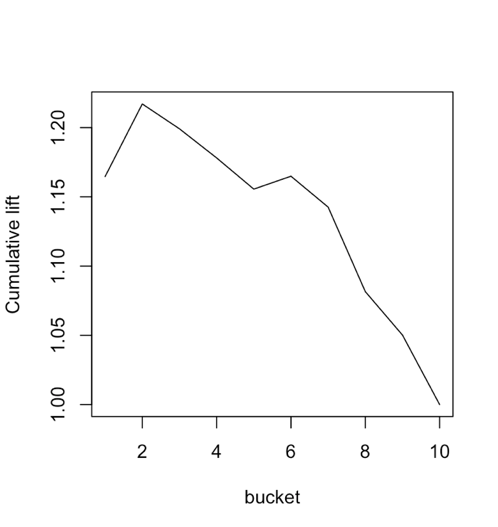
1. Now try to do the same analysis for frequency. What problems do you run into?

After running my code, I received an error message that stated: “Error: Aesthetics must be either length 1 or the same as the data (23570): x.” This occurred because the frequency variable does not have enough levels, meaning that most people have purchased one time or a few more times but there is not enough variation in that variable to make deciles.

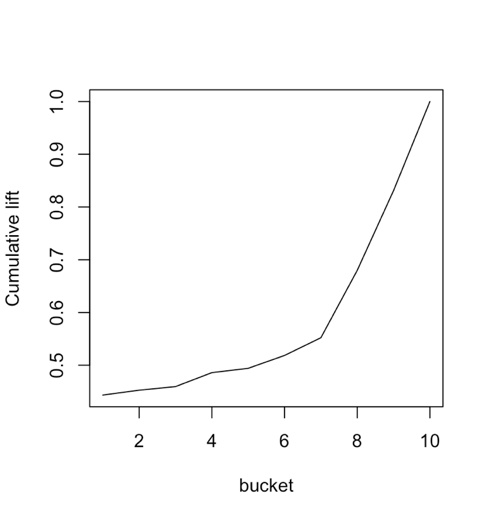
**Part 2: Visualization with Cumulative Lift Charts**

1. Create cumulative lift chars for recency and monetary (the idea behind this kind of chart is described in the handout “Lift Charts” on D2L). Based on these charts, does recency or monetary seem to be more important to predicting retention? Why do you conclude that?

After looking at the lift charts for recency and monetary, monetary seems to be more important when predicting retention because it shows a larger pump in response rate by targeting fewer people.

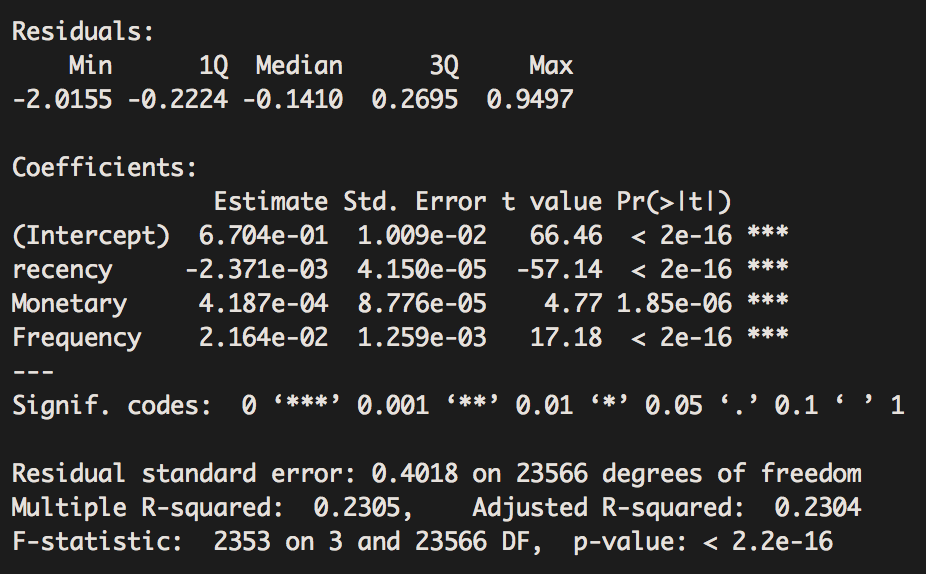


Monetary



Recency

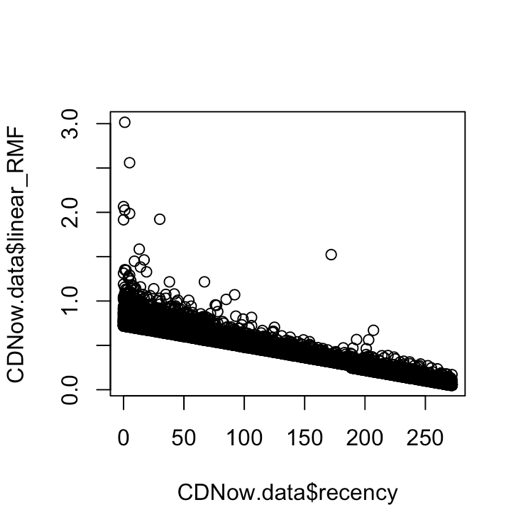
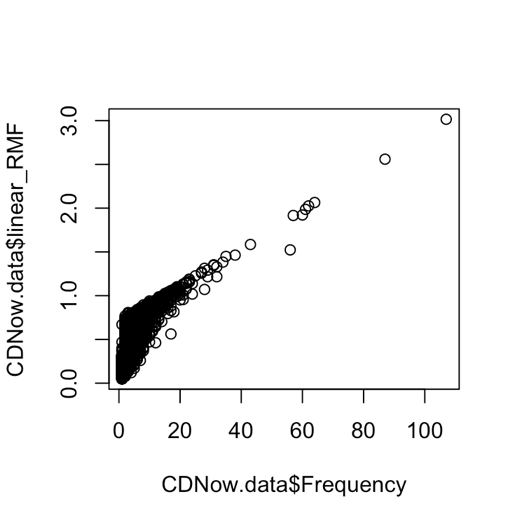
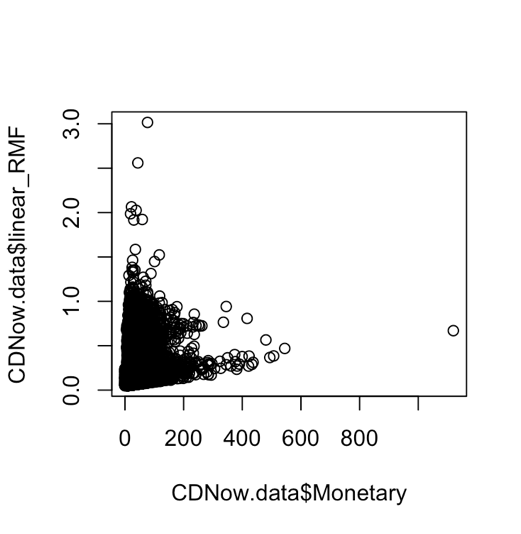
**Part 3: Predicting Retention with Regression**

1. Linear Regression: Use linear regression to predict retention using recency, monetary, and frequency as predictors. Interpret the model output including parameter estimates and significant levels. What are your conclusions?

* Intercept: The predicted value when all variables equal 0. This is the prediction for customers who did not purchase in Time2 and who last purchased on September 30th, 1997.
* Recency: The effect of recency controlling for monetary and frequency. An increase in 1 unit of recency is predicted to decrease retention by -2.371, meaning that adding one more day since last purchase is predicted to decrease retention by -2.371.
* Monetary: The effect of monetary controlling for recency and monetary. An increase in 1 unit of monetary is predicted to increase retention by 4.187, meaning that retention is predicted to increase by 4.187 if the average order size in dollars increases by 1 unit.
* Frequency: The effect of frequency controlling for recency and monetary. An increase in 1 unit of frequency is predicted to increase retention by 2.164, meaning that a 1 unit increase in the number of orders is predicted to increase retention by 2.164.
* Significance: All variables are significant.

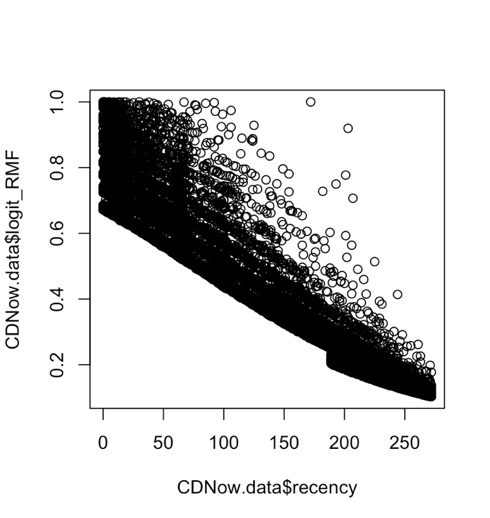
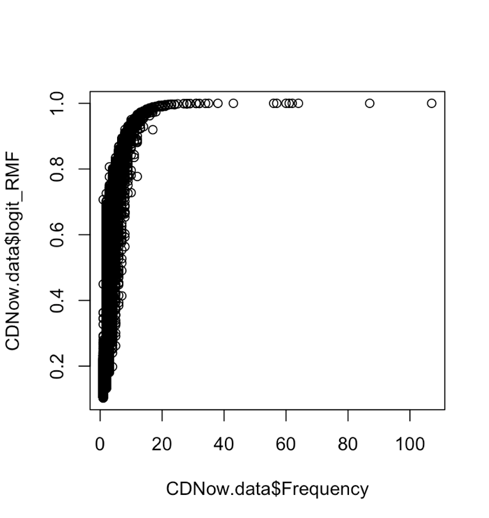
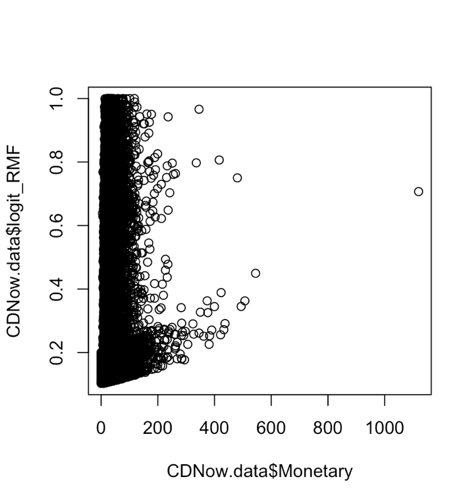
Conclusion: Recency, monetary, and frequency are good predictors for retention.

1. Generate a prediction for each customer. Plot the predictions as a function of recency, frequency, and monetary value (thus, create three plots, one for each predictor). What are the strengths and weaknesses of this regression approach?

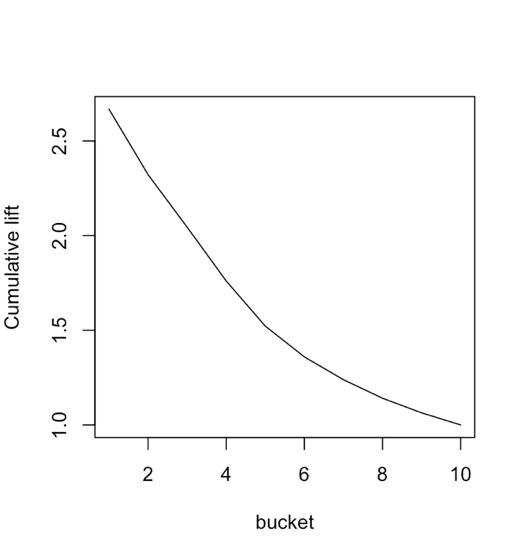
A weakness of this approach is that the linear line is not bounded. The linear regression shows how there are x values that exist for predictions outside the range of greater than 1. On the other hand, this approach can predict whether a new observation will fall on or near the function.

1. Logistic Regression: Now do all the same steps using logistic regression. Which model is preferable, linear or logistic? Why? Do the models lead to the same or different conclusions?

The logistic model is preferred because it applies a link function to our prediction, which minimizes the sum of squares with a fit line. These models lead to different conclusions because the logistic regression treats the two groups as separate. For instance, for the values that above 1 are turned slightly below 1 but with a linear regression the values are not altered if they lie outside the range.



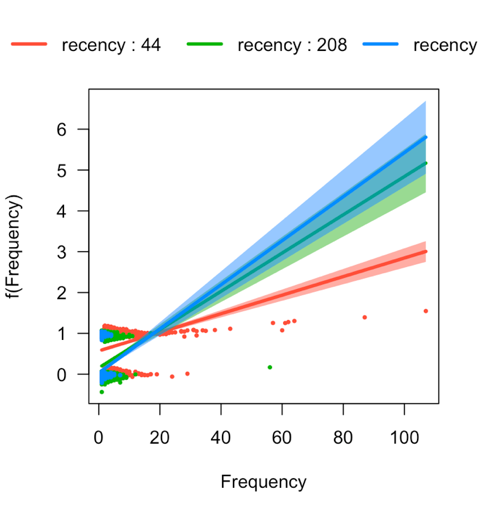
**Part 4: Visualizing Logistic Regression with a Lift Chart**

1. Create a cumulative lift chart based on the predictions of your logistic regression analysis. First bucket customers into deciles based on the predictions of the logistic model. Use these deciles for your cumulative chart. Compare this chart to the ones you created in part 2. What do you conclude?

Logistic Predictions

The cumulative lift chart based on the predictions from the logistic regression analysis seems to be more accurate compared to the cumulative lift chart in part 2. When looking at the monetary cumulative lift chart, it wouldn’t make sense to target group 2 more than group 1 since they are in a lower income bracket. With a steep curve, we can conclude that the logic regression chart is more reliable in analyzing the pump in response rate by targeting fewer people.

**Part 5: Interaction Model and Cross-Validation**

1. Run a new logistic regression that includes recency, frequency, monetary value, and the interaction between recency and frequency. Is there a significant interaction between recency and frequency? What does this interaction mean?

Yes, there is a significant interaction between recency and frequency. This interaction implies that the slopes are different, and the effect of frequency differs for the number of days since the last purchase at the end of Time1.

1. Use the CVTools package to cross-validate this model and the logistic model from part 3. Use K=5 and R=5 for your parameters. What does the cross-validation show? Which model seems to be better based on the cross-validation? What does it mean to be “better” in this case?

The cross validation gave a 5-fold CV result of 0.04162691. This shows that the model is strong in predictive performance. The term “better” in this case means a model that maximizes the number of its, distinguishes good customers and bad customers, targets good customers and ignores bad customers, and increases predictive ability.

**Part 6: Conclusions and Managerial Implications**

What are some things CDNow might do in light of your analyses to sustain and grow their business? Remember, CDNow essentially went out of business before the dotcom bust, after dramatic growth early on. Looking back what might they have done differently by using customer analytics?