* **The executive summary**
* The executive summary should brieﬂy describe the source and background of your data and summarize the questions you addressed and your key results (e.g., explained in the context of your speciﬁc application). Also mention any caveats. Summarize your recommendations on how these results can or cannot be used.

**Purpose**

School Rank Analyzer provides ranking for elementary, middle and high schools in all of the 50 states.The current dataset is for elementary schools.Objective of this study is to provide a predictive model to establish rankings based on available parameters. There were different variables, some of which included number of students,student teacher ratio, number of full time teachers, percent of Americans, percent of African American population, percent of Hispanics, percent of Asians,low grades,High Grades,low grade,high grade,is title school,is charter or is private school.

**Interesting Findings along the way**

Counties with lower number of schools have better Ratings. Hispanic ethnicity is most the dominant group for Title schools. No strong correlation identified between Ranking and considered variables. Influencers able to explain 58% of the Ranking process .

**The Development process**

Initially, the model suffered from overpruning . Then I tried changing the value of bestcp as .006.

What I did was right but the conclusion was not satisfactory. The predictive power of the model is too coarse. The success rate of the predictive model was partial because it . Random forest tree or boosting might have produced better results .But if even random forest produces unsatisfactory results then this means that data set is bad for prediction. And that the variables are not a reasonable proxy for predicting rank.it turns out that the tree is not flexible enough, and that is the best we can do with this model. So we cannot expect much with this model for the current dataset. I used the training set to test, but the reality maybe even worse.So, with this model when I predict the best schools, I am right but if I predict the so so schools, my prediction may be off, one way or another way.If I am saying that its good ,most likely I am right.if ,you go towards the top rank, the more accurate it is likely to be.So if I say 1500 , the worse it can be is around a 100, but if it is say 700, it can be anywhere in between 300 to 1200.For the bottom 50% I would not be very accurate , but for the top 50 % , I think it will be fine.A big advantage of what I am doing is that I can give a summary of what variables matter for example ,the department of education can come up with an intervention strategy.With random forest I could have got a better predictive power but the level of interpretability would have gone down.i have squeezed out as much as I could from this model.Of course other models could have give better results.

Trained on the training and validate with the test. prune on modelcart1.

**Details of variables**

Details of variables

Type of variables

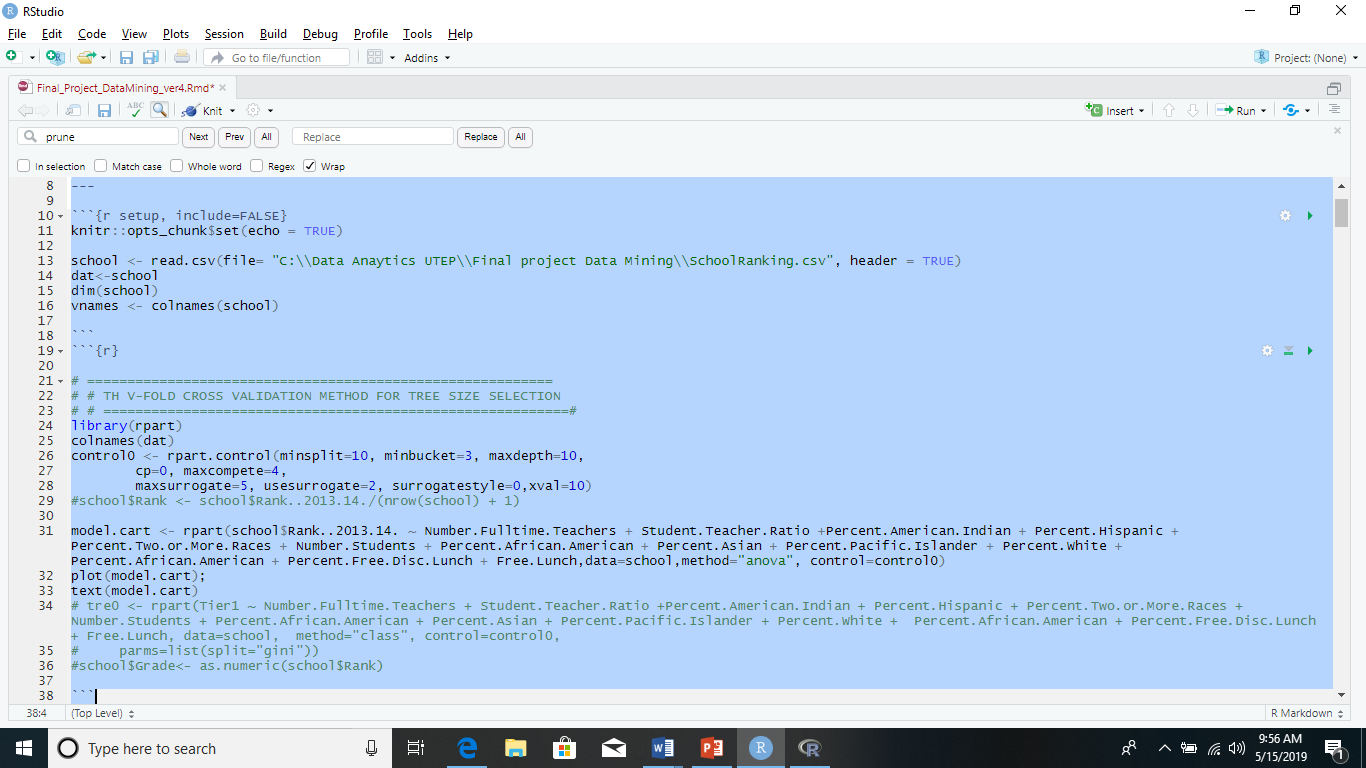
Nature of problem

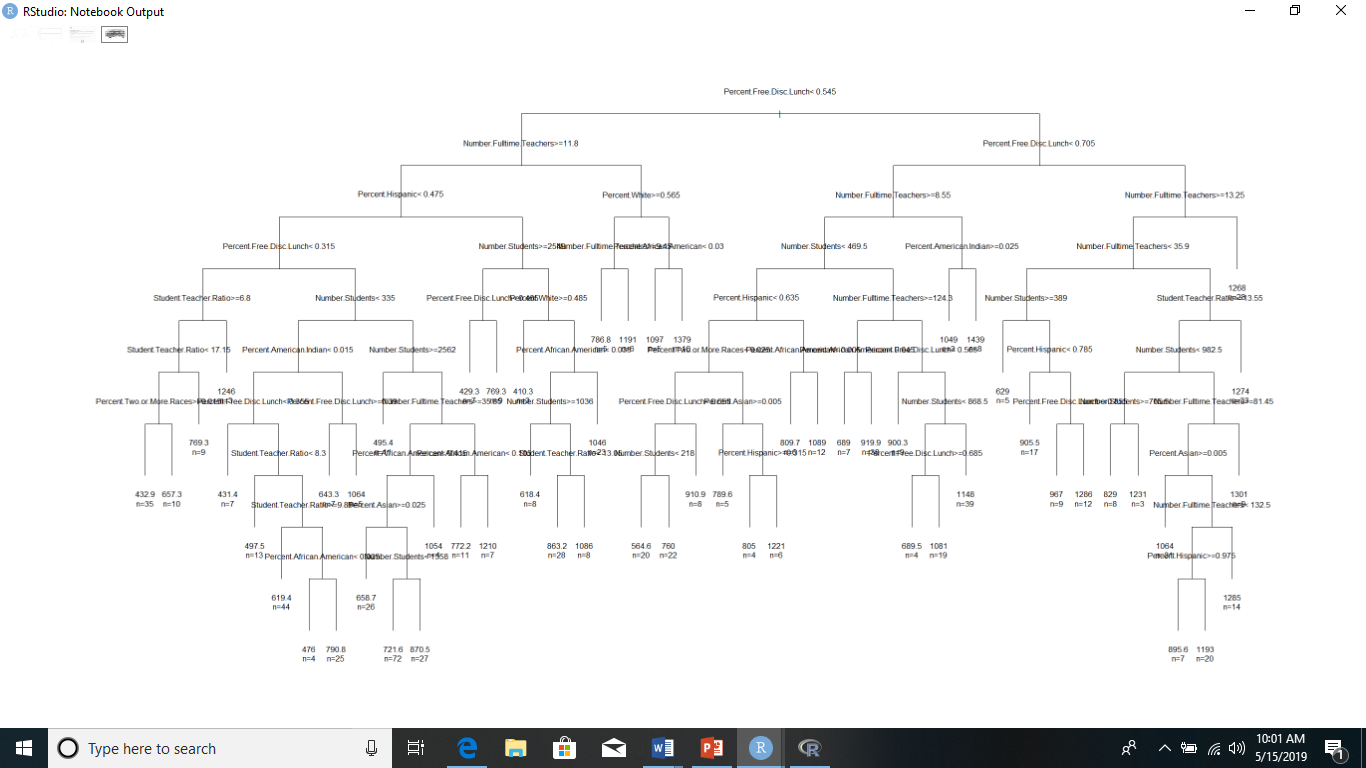
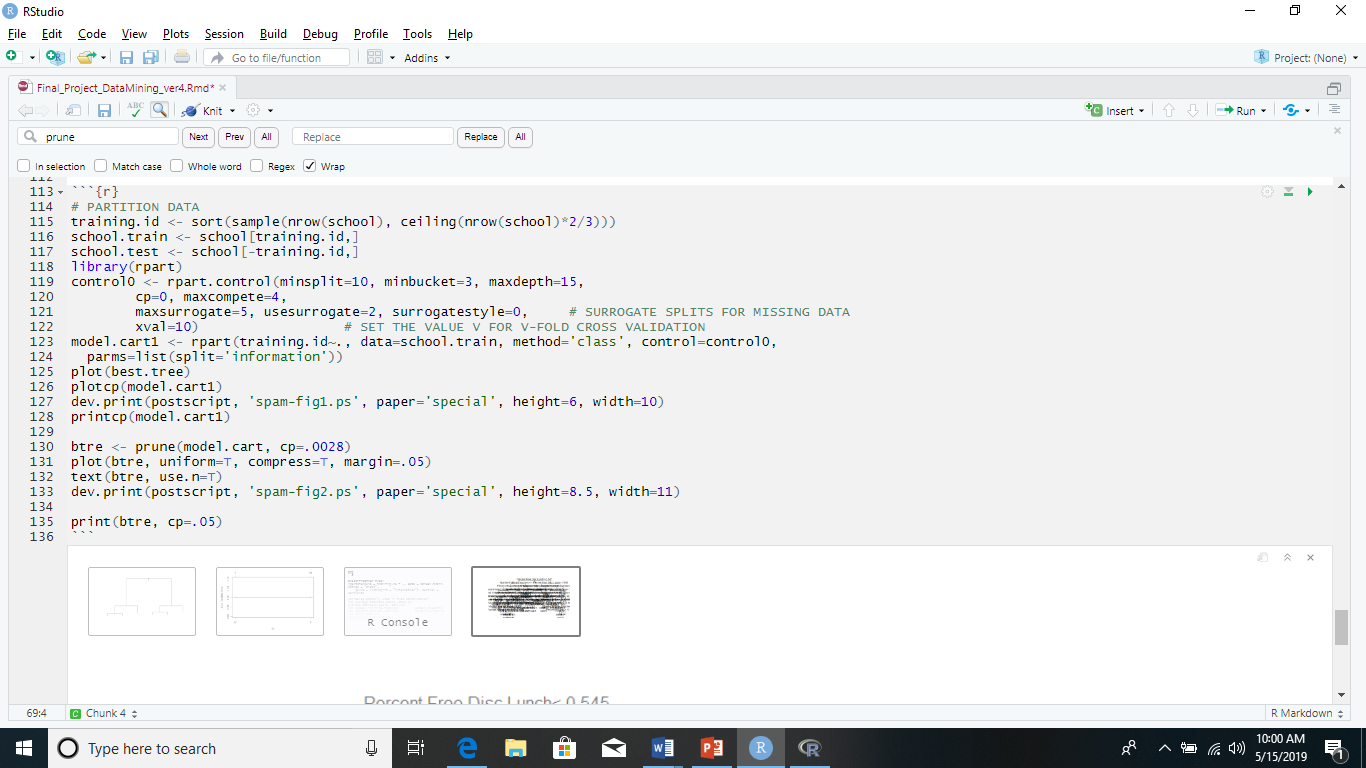
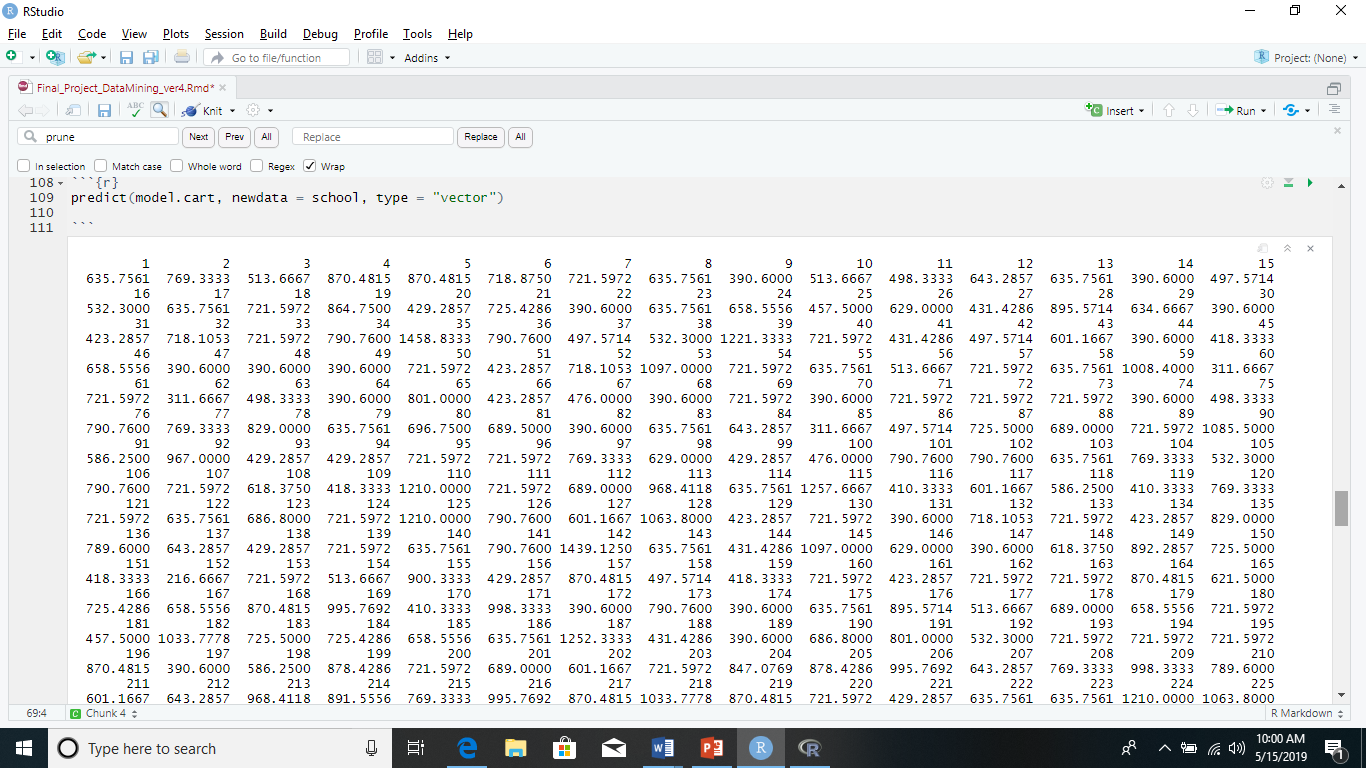
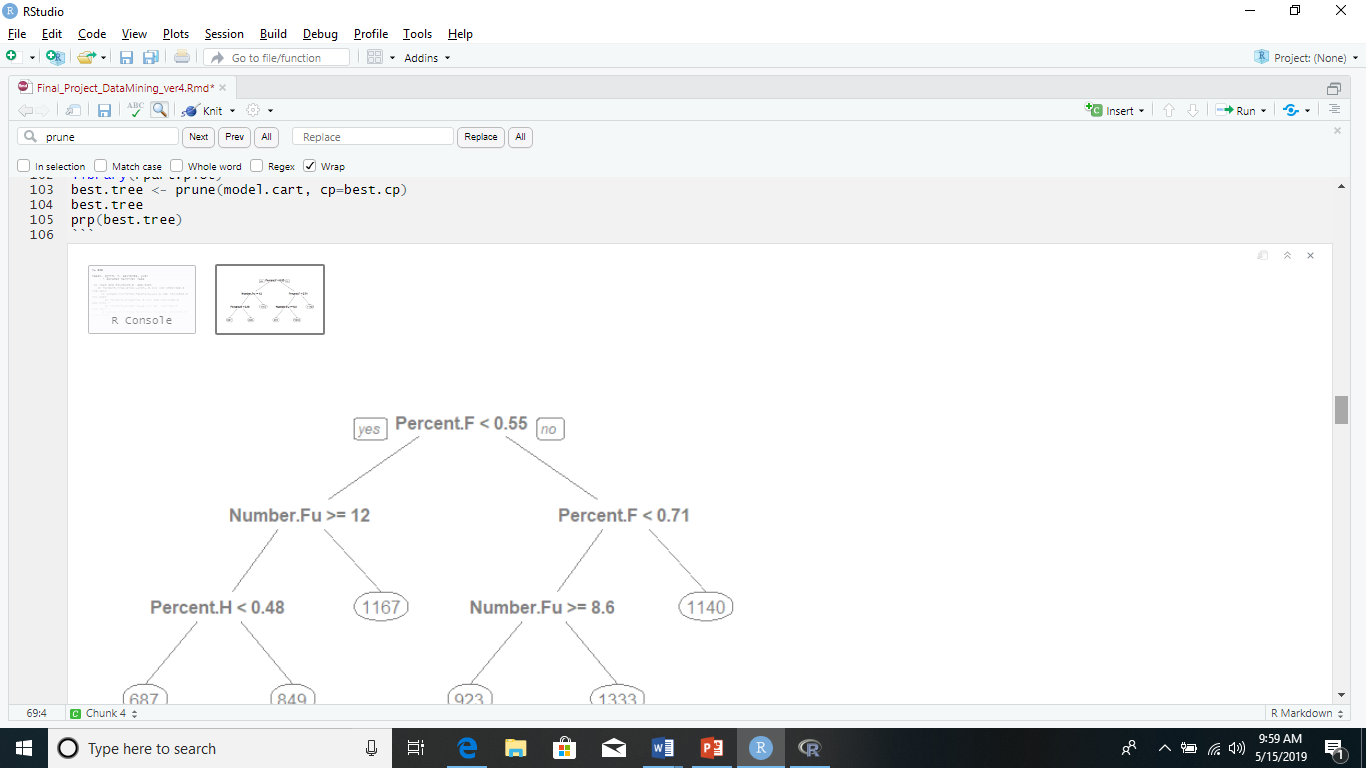
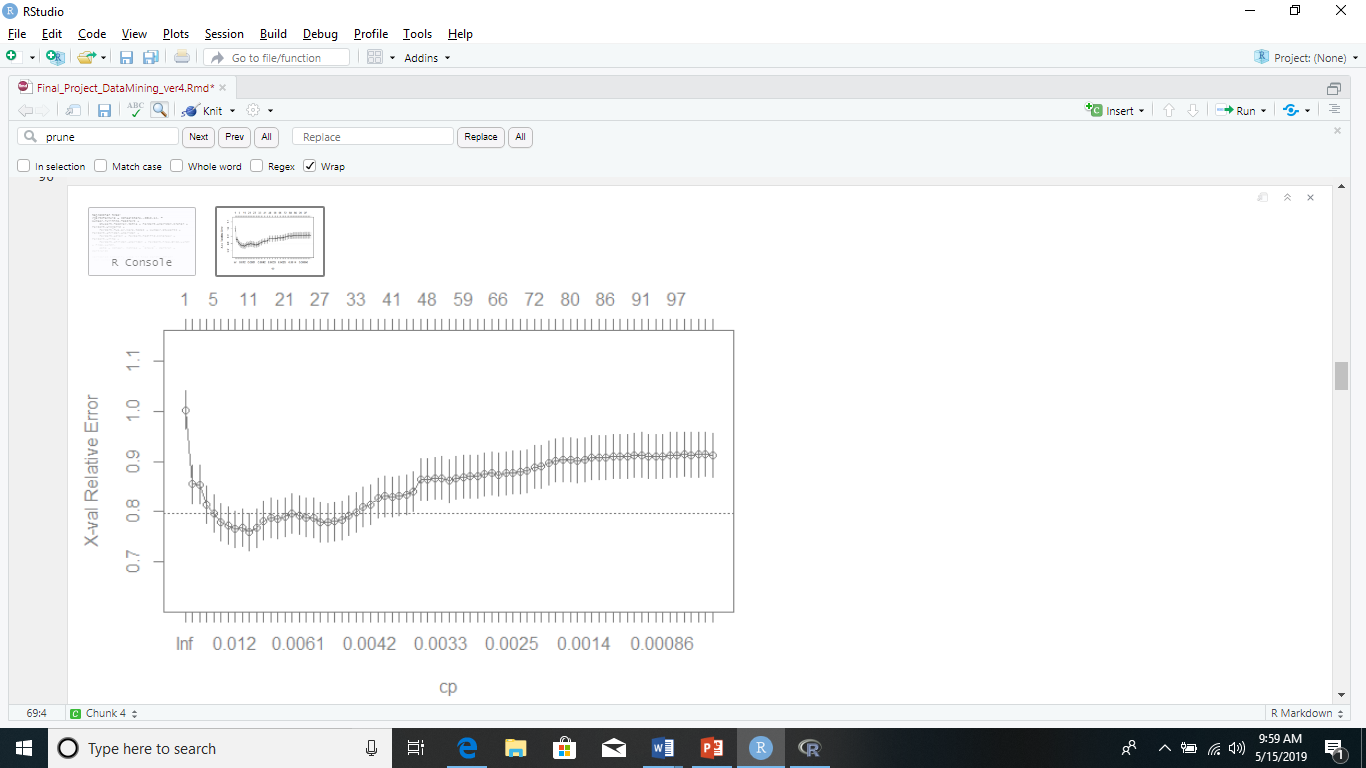
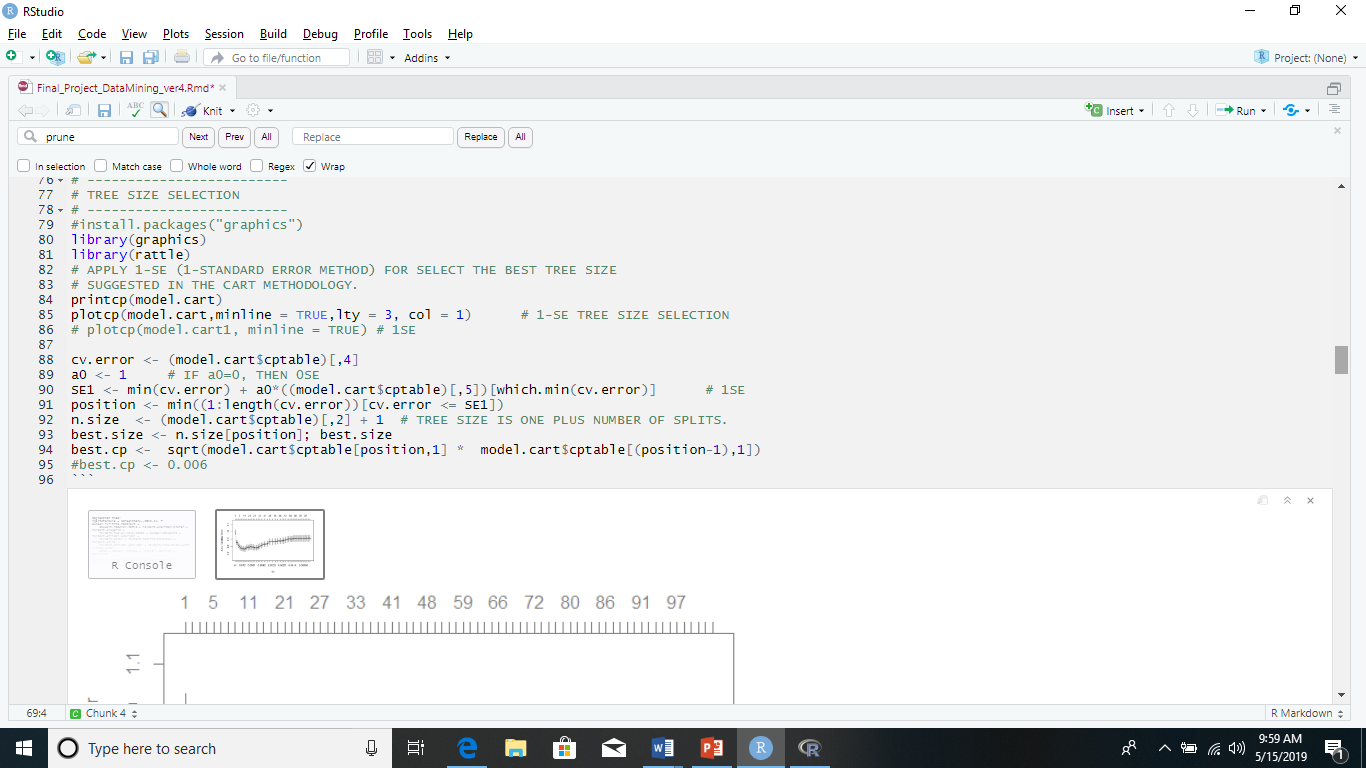
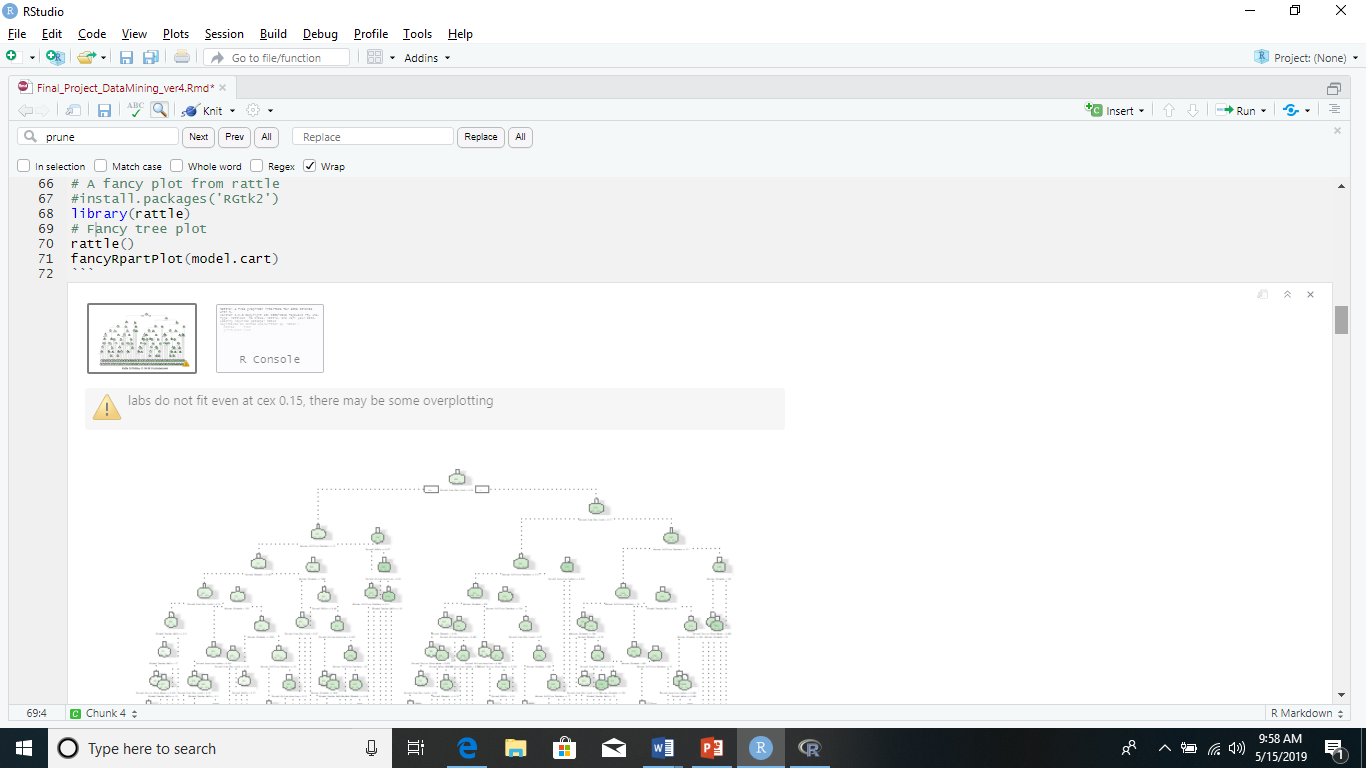
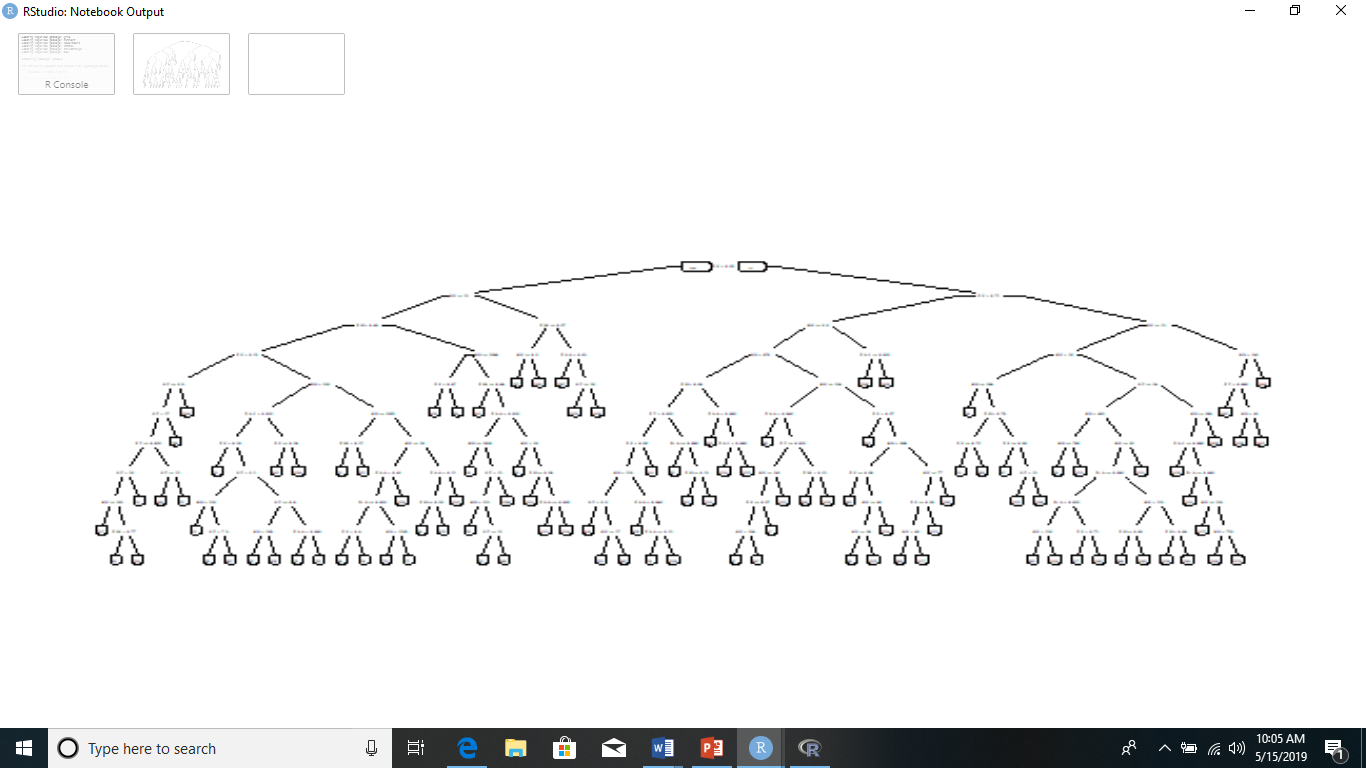
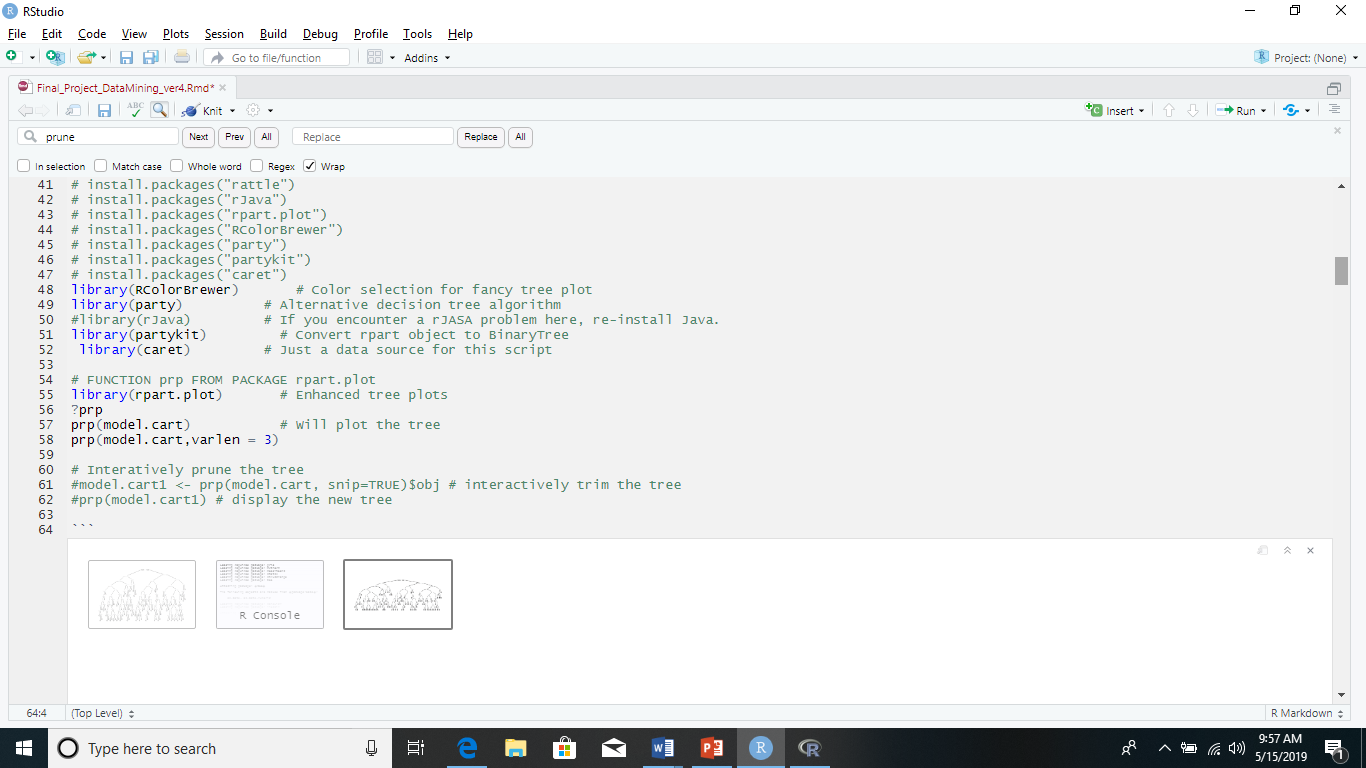
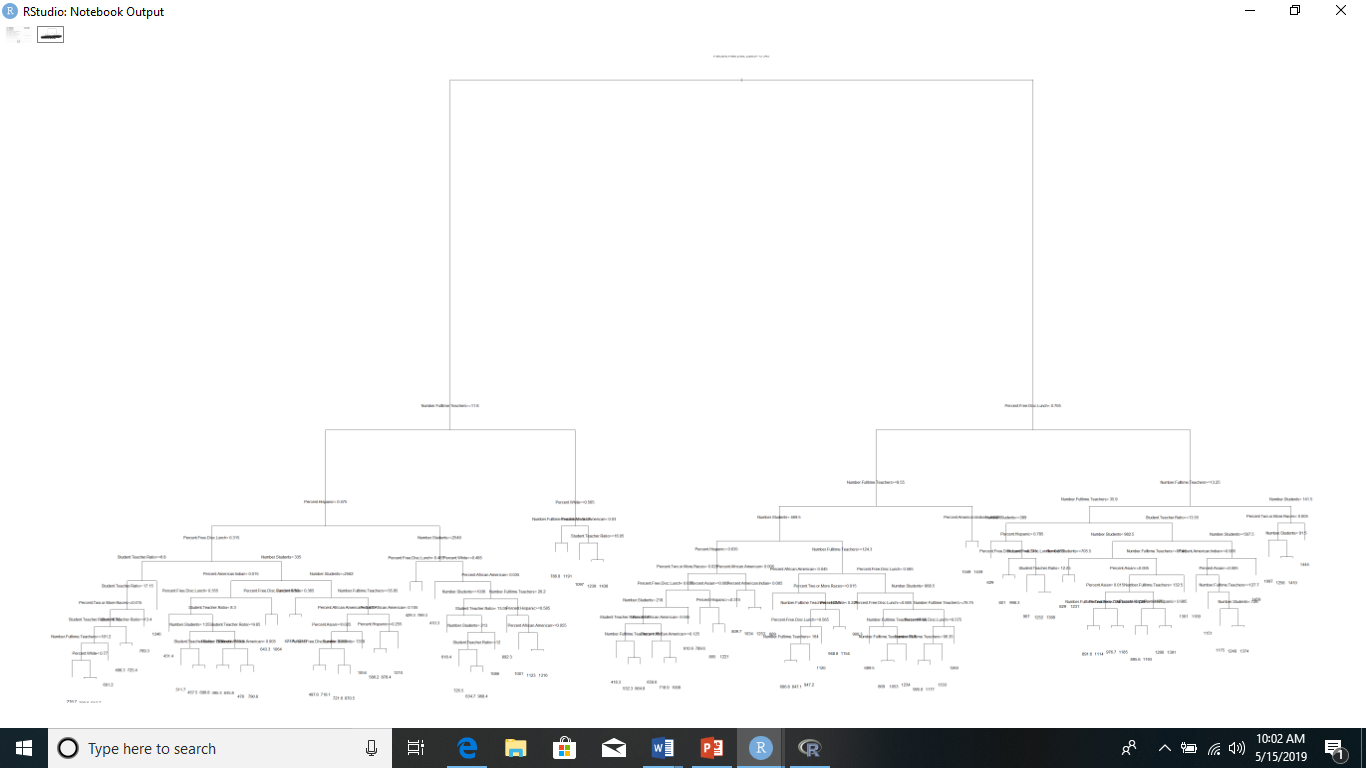
What kind of model was used

why Regression-type problems are generally those where we attempt to predict the values of a continuous/ categorical variable from one or more continuous and/or [categorical predictor variables](http://www.statsoft.com/textbook/statistics-glossary/c.aspx?button=c#Categorical Predictor).

**About the model**

Its not perfect, but not useless .





* Every year Schooldigger calculates school rankings based on STAAR Reading, STAAR Mathematics, STAAR Writing, STAAR Science test scores released by the Texas Education Agency. There are 18 schools sharing ranks

No duplicate values found in the dataset

Assigned PK – 100, KG – 101

Ignored 7 schools with Blank Titles, 3 schools with no student count info

83 schools are new with no 2013-14 data for standard scores – these are deleted

In total, 101 (2%) records deleted.

98% of original data cleaned and retained

These rankings were last updated on July 17, 2018, and are based on the most recent test scores available. We update rankings as we receive new test scores from the Texas Education Agency, usually on a yearly basis.

* Our sources of data include the National Center for Education Statistics, U.S. Department of Education and the Texas Education Agency