Modeling Grade Distribution

A PROJECT REPORT SUBMITTED IN PARTIAL FULFILLMENT
FOR THE REQUIREMENTS OF THE DEGREE OF

Bachelors of Science in Mathematics

BY

Jasmine Gray

UNDER THE GUIDANCE OF

Jebessa Mijena, Phd.

0pt*0*0

0pt*0*0

0pt*0*0

School of Arts and Sciences

Georgia College and State University

DECEMBER 2017

0.1 Abstract

The purpose of this project is to see how multiple variables are associated to a students overall performance in courses from various departments. We looked at the length, level, semester, professor rank, course title and start time, in twenty-four- hour format, of the course as well as specific characteristics about the course such as number of students and professor to see how these are correlated to the distribution of grades. In the first phase, we started out collecting data from the Georgia College grade distribution for all college courses from all departments we studied from the past 7 semesters. We did not use online courses due to the fact that they had no set length to the class which would make one of our variables nonexistent in that instance. During the second phase, which was analysis, we used diverse models such as regression models, tree based models and random forest models to analyze the association between our variables. With this information, we used the statistical software R and a variety of its functions to evaluate the data we collected. Our goal is to show how these eight variables can affect a students performance in the classroom. Here is a sample of a few of the results that we have found most interesting. We found that the highest percent of Ws (withdraws) was in Summer 2017 in the class CHEM 1151K. Also, out of all 890 courses we studied 22 total had zero percent of As. There was also roughly 36% of the courses we studied with less than 16 students.

Chapter 1

Introduction

When it comes to a students performance in a collegiate level course there are many variables that could play a role that are outside of the students control as well as some that can be changed. Outside of looking at the students study habits and breakdown of how the student did in each part of the course we looked solemnly at the end grade percentages. The class time and professor are some things that the student has control over choosing but class size is one thing you can not control. When students are choosing classes for upcoming semesters they mainly look at the prerequisites, the time and professor, but what if they looked deeper. In this study we covered a total of 10 variables. The response variables in our study were the percentage of each letter grade (A,B,C,D,F,W). The predictor variables we covers were semester, level, length, size, time, college and professor rank. We covered a total of 7 different semesters, from Summer 2015 to Spring 2017. Each semester had a mix of courses from one hour to longer. In each of the semesters covered there were courses from each level from 1000 to 4000 which is freshman to senior level courses.

In this study we covered a total of 16 various departments within Georgia College and from them we broke them down into five different colleges. The first was the college of Business which covered course from Economics, Finance, Marketing, Management, Business Communications and Accounting. These courses took up a total of 21% of the data we covered. The next college we partitioned into was the college of Science which covered the departments of Math, Math Education, Physics and Chemistry. These course took up a total of 31% of the data. Next, was the college of Health Sciences in which we only covered Nursing, these courses took up only 7% of the data. Then, we had the

college of information technology which covered the departments of Computer Science and Management and Information Systems, these departments were a total of 10% of the data. The last college was the college of social science which covered the departments of English, Mass Communications and Psychology which took up 31% of my data. Although, some colleges took up less of the data than others we also looked at the length of the courses. The two choices for length of course were hour classes and none hour classes. We also covered class time.In this study we took the times of the beginning of the course and categorized them into four different categories. The first was morning classes. These classes were course that started between 8:00 am and 10:59 am. Next was midday classes which started between 11:00 am and 1:59 pm. Then, we had afternoon classes which started between 2:00 pm and 4:59 pm. Finally, we had evening classes which were any classes that started at or after 5:00 pm.

In this study we also covered rank of professor. We broke down our faculty into 8 various ranks. Out of those 8 ranks 4 of them were full-time and the other 4 were temporary positions that people would hold. The 4 full-time were: full professor, associate professor, assistant professor and lecturer. When beginning as a professor you start at the rank of associate professor then move up to assistant professor and then full professor. Lecturer isn't categorized in this process. The 4 part-time positions were: part-time, teaching fellow, graduate assistant and instructor. All of our part- time positions came from our English and Computer Science departments. They stated that instructor meant that they were not on a tenure track line and had no PhD. For graduate assistant and teaching fellow all were graduate students but the teaching fellows came from the Masters in Fine Arts program and taught 2-courses while the graduate assistants came from the Masters of English and Rhetoric and taught 1-course. The part-time positions were about 10% of our data while the full-time was about 90%. The final variable was the size of class.

There have been many studies on how the size of a course plays a roll in a students performance on a secondary level, but most have come back with the same result. Most can not find a significant correlation between the size of a course and the students performance primarily because of the fact of the study analyzing to many variables causing class size to be hard to study. In our study we observed class sizes from as small as 5 students to as large as 73 students. We broke our class sizes down into 4 different categories. The first was small classes which were classes with less than 16 students. These courses were a total of 36% of all of the courses we studied. Then, we had average classes which were classes between the age of 16 and 34 students and about 43% of the data in this study. The next 18 of the courses we covered were there large classes with between 35 and 54

students. The final group was packed classes with between 55 and 74 students which was 3%.

Chapter 2

Methodology

2.1 Type of Research

The type of research used in this study is quantitative research. Quantitative research is research that gathers information using numerical data.

2.2 Data Collection and Clean-Up

In the process of collecting data we used the Georgia College paws website. Within this website we used the grade distribution as well as the course schedule. When collecting we only collected information for courses that were taught in class. This is because online courses would make one of our variables nonexistent. We also only did courses in which the professor only taught one section, because if a professor taught multiple sections in the grade distribution it combined all students into one. This would cause multiple variables to be incorrectly counted. In total we had 890 data points. After collecting all of the data, then came the clean up. In the process of clean up we went through and got rid of courses that had less than 5 courses total being that they would not give a good representation of that department.

2.3 Analysis

In this study we used three different models to analyze the data. The three models were the multiple regression model, regression trees and random forest models. Multiple regression models use many predictive variables to determine the correlation between them and the response variable. Being able to find outliers as well as being able to see how each variable is correlated to the response variable. Regression trees are a way of producing prediction models from data. There are a few advantages of this type of model and one of them is that nonlinear relationships between select parameters do not effect the trees performance. The assumption of linearity is not required for regression trees making it valid to run on nonlinear parameters unlike other models. Regression trees are also easy to read and interpret, which makes it easier to explain and for everyone to visualize everything. The last model is the random forest. Random forest is a mega estimator that fits multiple regression trees. The greatest advantage of random forest models are that they are most accurate of all models. When it came to comparing these 3 models we are comparing errors of the squared residuals of testing data. By looking at those we can see the best models. When it came to sampling our data we broke it up into 2 parts: training and testing. Training data was 70% of the data with exactly 623 data points while testing data was 30 % with 267 data points.

Chapter 3

Models

There was no model for the regression tree or random forest model but there was one for the multiple regression:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \epsilon,$$

where x_1 =college, x_2 =semester, x_3 =level, x_4 =time of class, x_5 =size of class, x_6 =length of class, x_7 =rank of professor $\epsilon \sim N(0, \sigma^2)$.

Chapter 4

Results and Discussion

We used a variety of graphs and models to analyze our data. These models were able to give us a visual representation of our data.

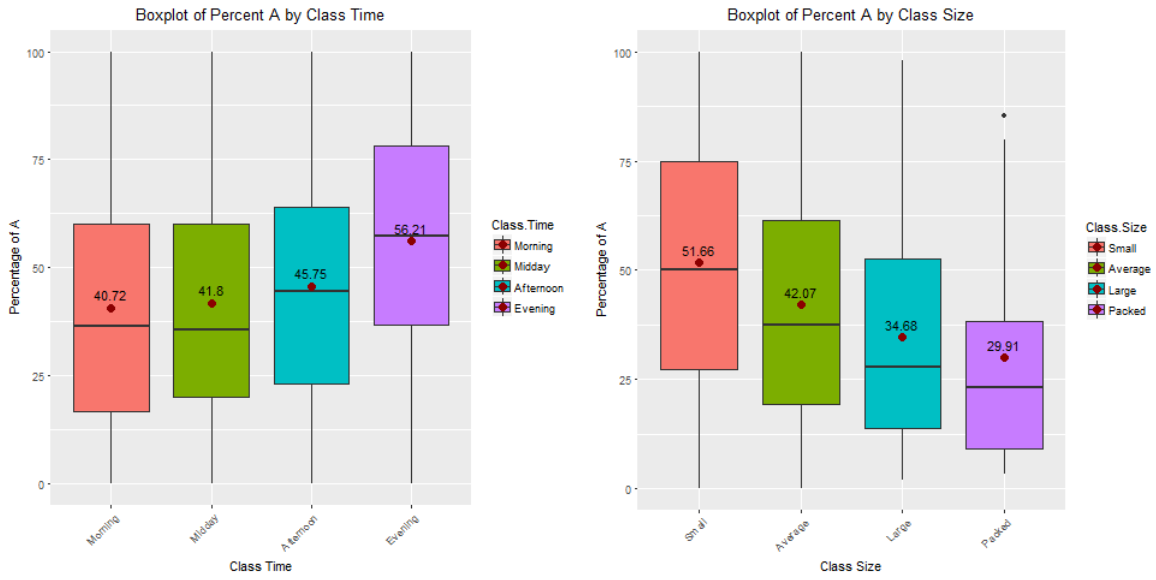


Figure 4.1: Percent A by time and size

The overall highest mean in percent A was in the evening classes with a mean of 56.21%. There was an overall 15.49% difference from the highest mean and lowest which came from the morning classes. The greatest difference was between evening classes and afternoon classes with roughly a 10% difference. When it came to class size it was interesting to see how as the class size went up the percentage of A's went down. From the small classes to the packed classes there was a 21.7% decrease in A's.

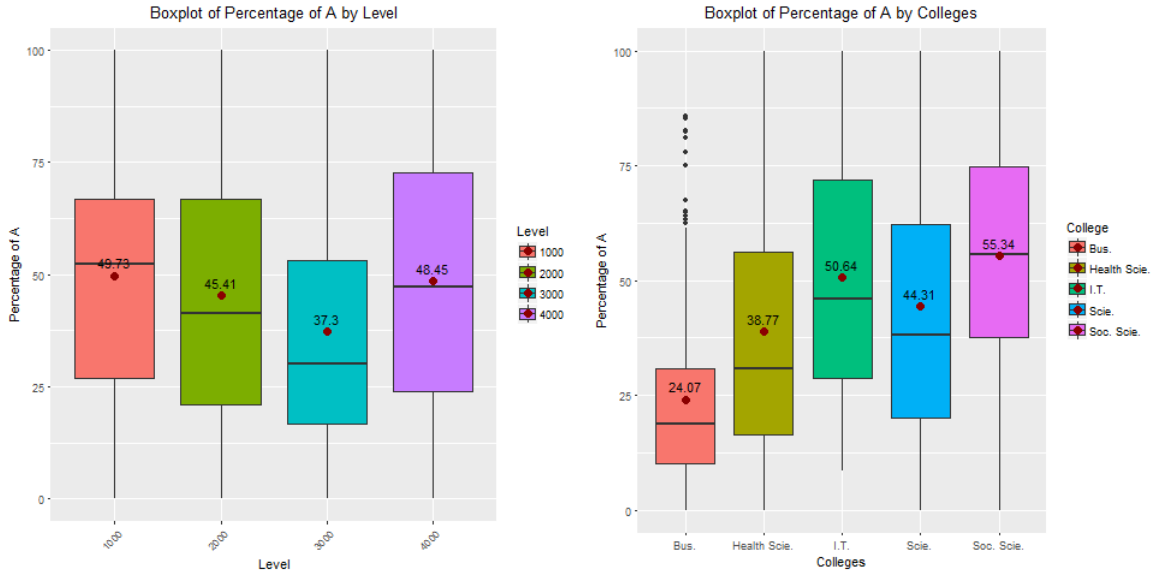


Figure 4.2: Percent A by level and college

The level with the highest percentage of A's was the 3000 level courses which are junior level courses had the overall lowest mean of A's. The college with the highest mean of A's was the college of Social Science while the college of business had the lowest with 30% less A's than that of social science.

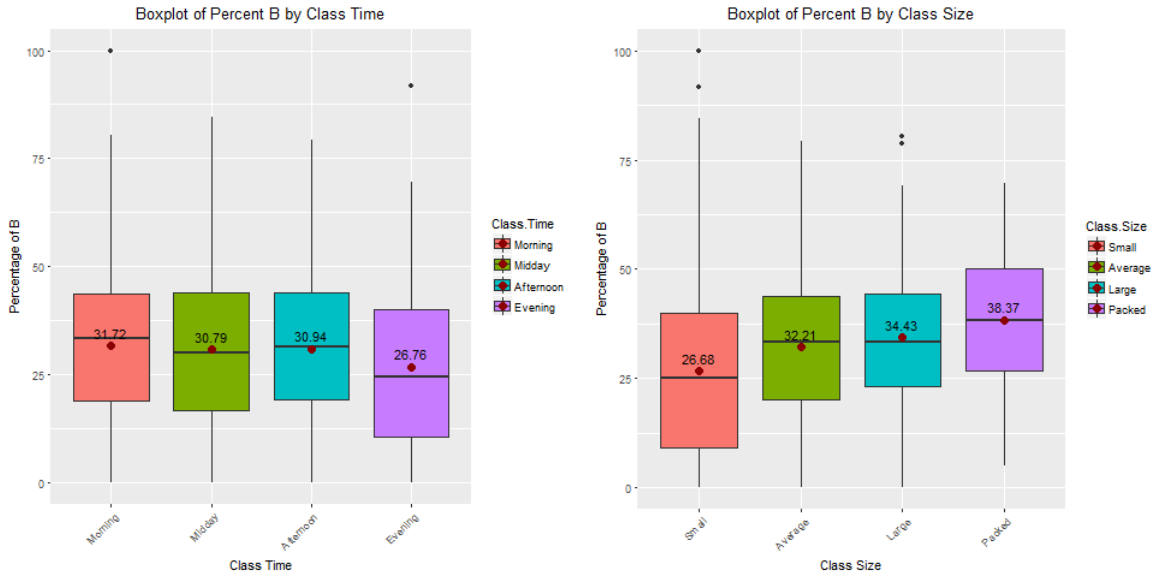


Figure 4.3: Percent B boxplots

For percentage of B, we were able to see that the time if day did not play a significant role in the percentage of B's given except for evening classes which had the lowest percentage of B's. When it came to class size, the packed classes had the highest mean of B's although everyone had close to the same widespread across the board.

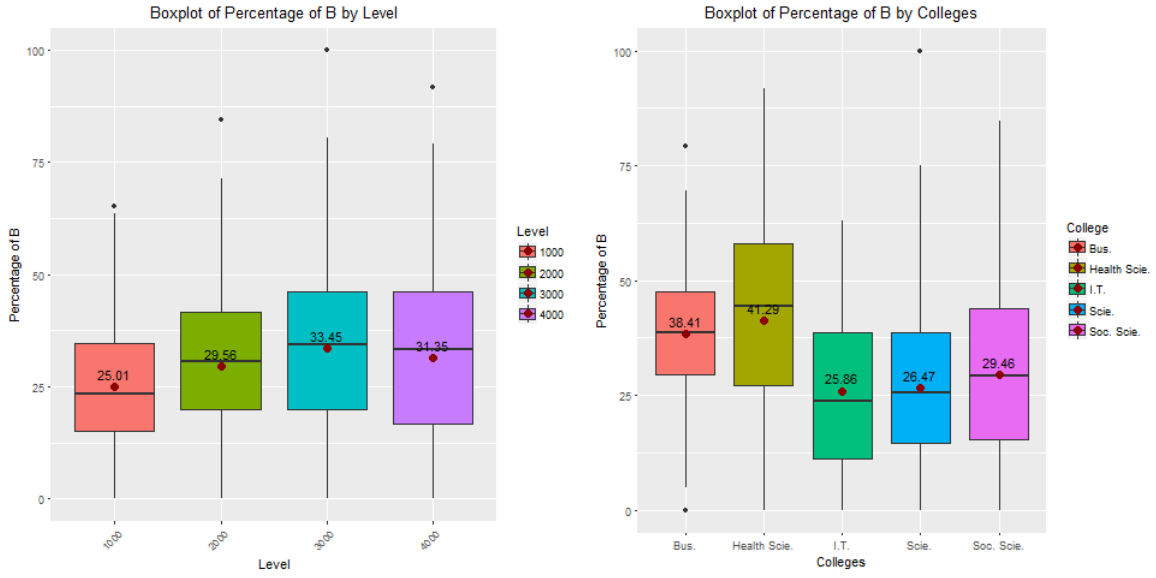


Figure 4.4: Percent B boxplots

The highest mean of B's came from the 3000 and 4000 level courses with a 3000 level courses 2.1% higher. The college with the highest mean of B's was the college of health science in which we covered just the nursing classes.

Figure 4.5: Percent C boxplots

or percentage of C's we can see that for the morning and midday classes there mean was off by less than 1% and both had a great widespread. When class size was considered we were able to see that large and packed classes had a difference of less that .3%.

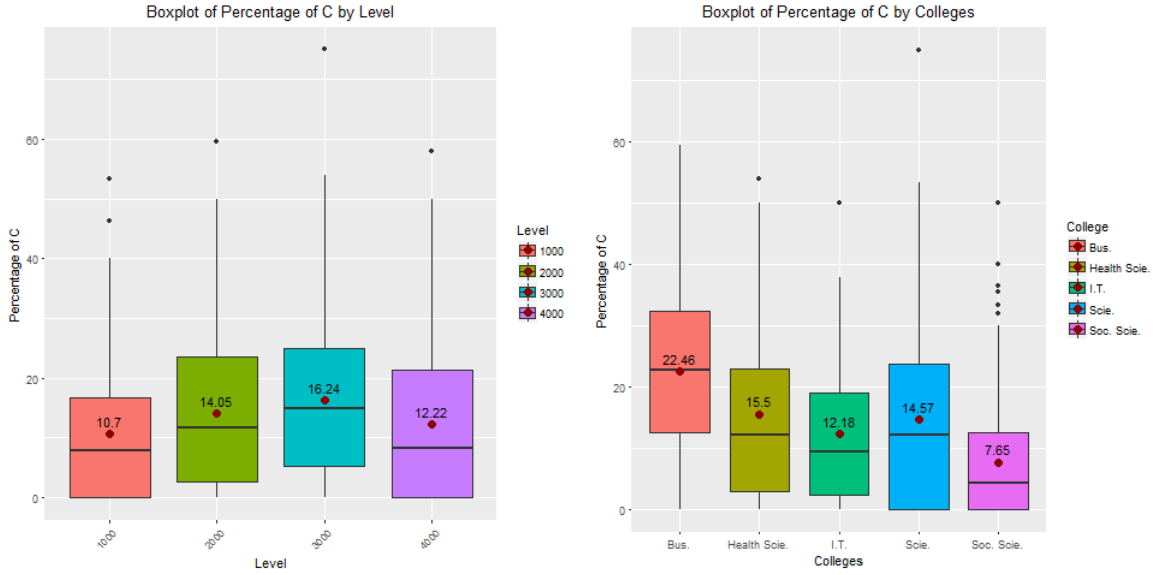


Figure 4.6: Percent C boxplots

The greatest mean of C's came from the 3000 level courses. The college with the highest mean of C's was the college of Business which I covered accounting, business communication, marketing, management, economics and finance.

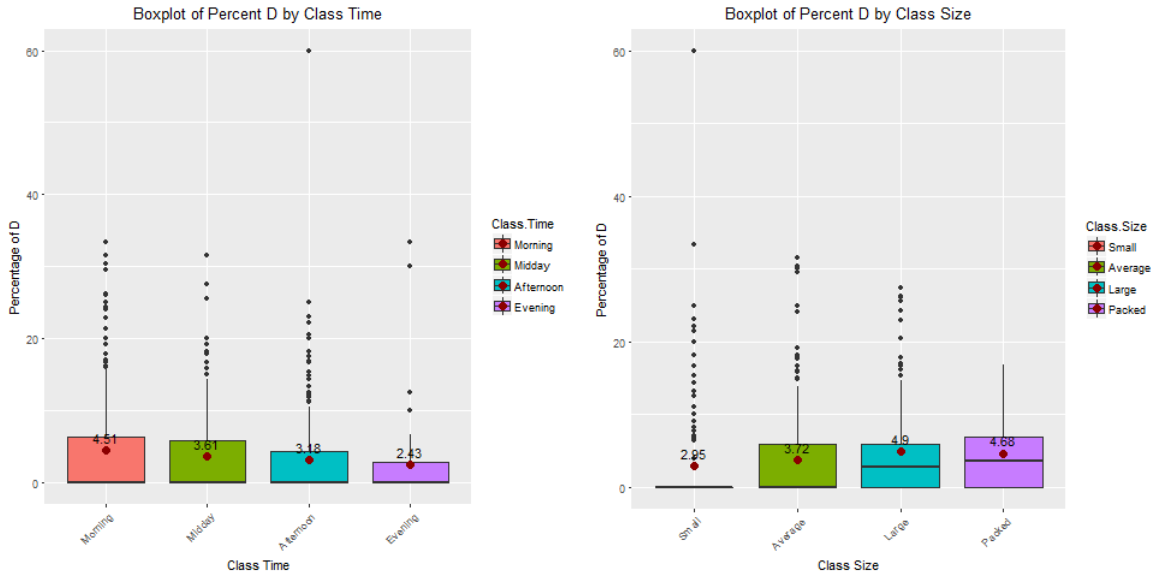
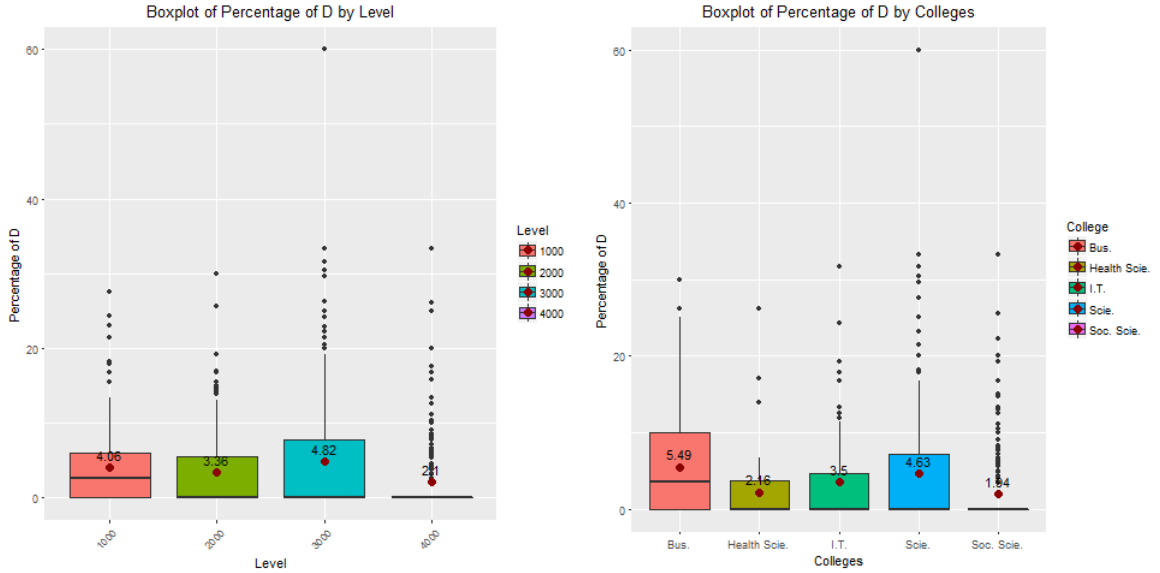


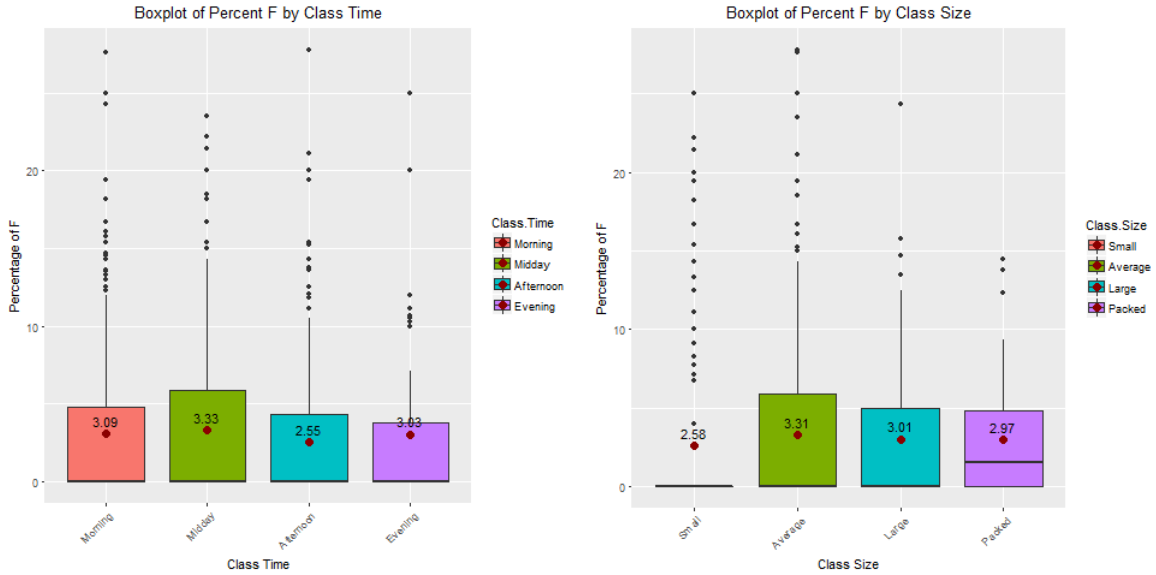
Figure 4.7: Percent D boxplots

When it came to class time the percentage of D's went down as the classes got later, with highest mean of D's being in morning classes. The mean of percent D for all class sizes was between 3 and 5 percent. An interesting fact was how the mean of percent D for small classes was an outlier on the boxplot.



When it comes to the percentages of D by level the means were less that a 2 percent difference between them as well as the mean percentage of D for 4000 level courses is an outlier in the data. The college of Business had the highest mean of D's across the board with a mean of 5.49%.

Figure 4.8: Percent D boxplots



The percentage of F's had the highest mean in midday classes and also average sized classes.

Figure 4.9: Percent F boxplots

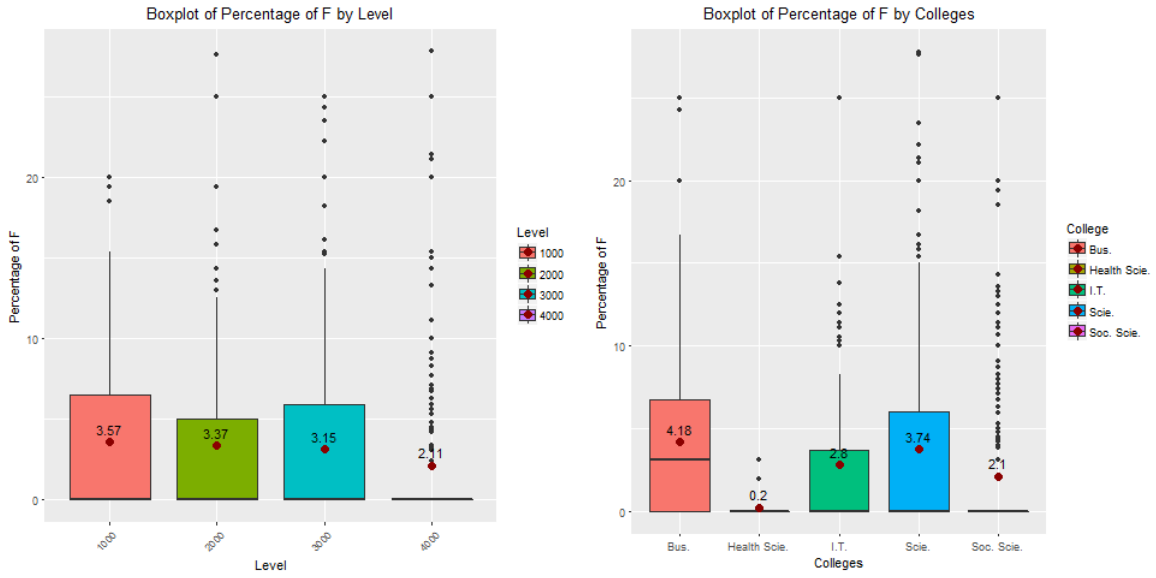


Figure 4.10: Percent F boxplots

When it comes to level the widespread of F's are close to the same for all levels except for the 4000 level which is lower than all. Once again, the highest percentage is from the college of business.

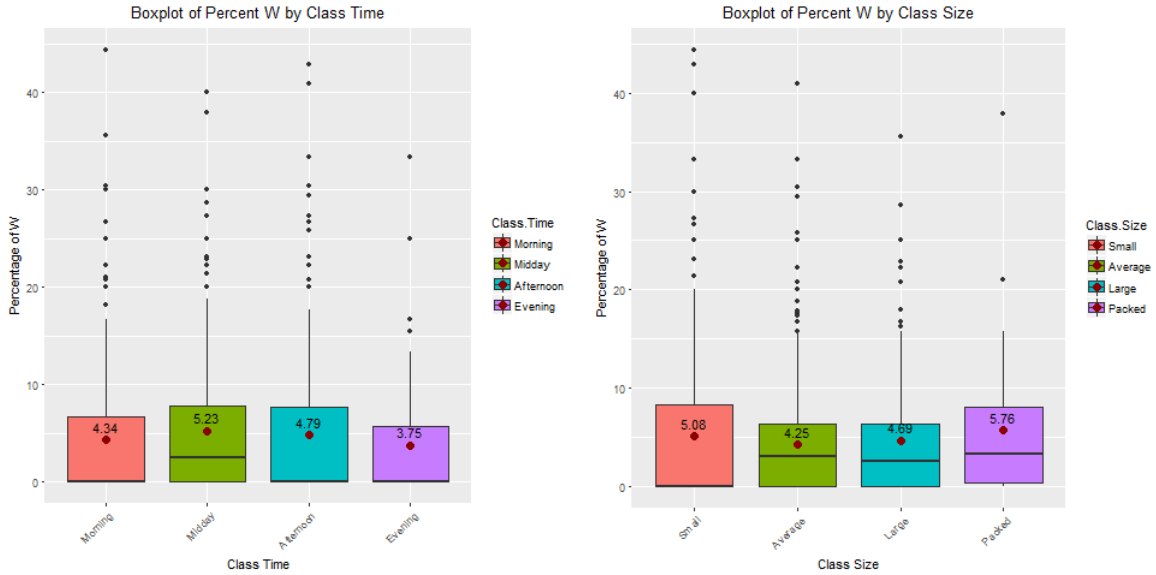


Figure 4.11: Percent W boxplots

With percentage of W's the mean was close to constant for time as well as size with slight variability.

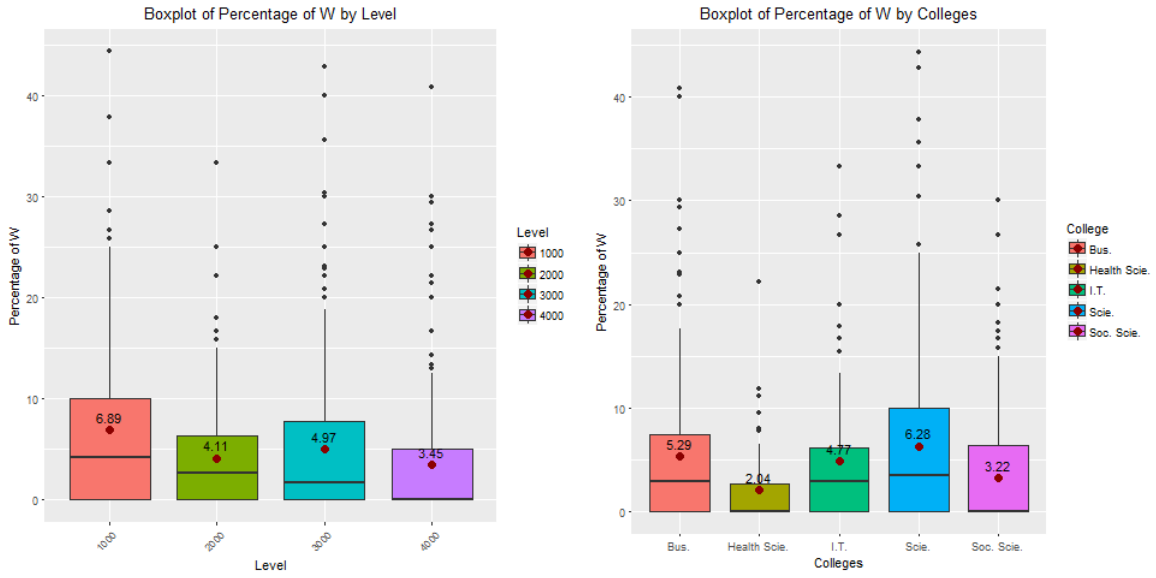


Figure 4.12: Percent W boxplots

The level with the highest mean of W's was the 1000 level. This was not a shocker being that in the 1000 level is freshman classes where students are still trying to find what they want to study. The college with the highest mean of W's is science.

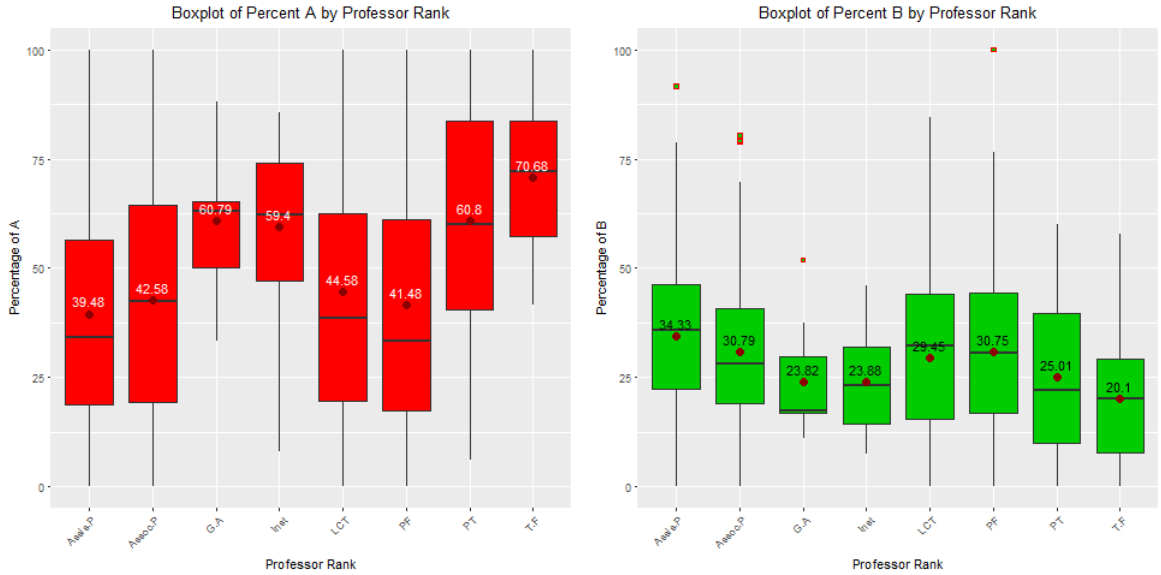
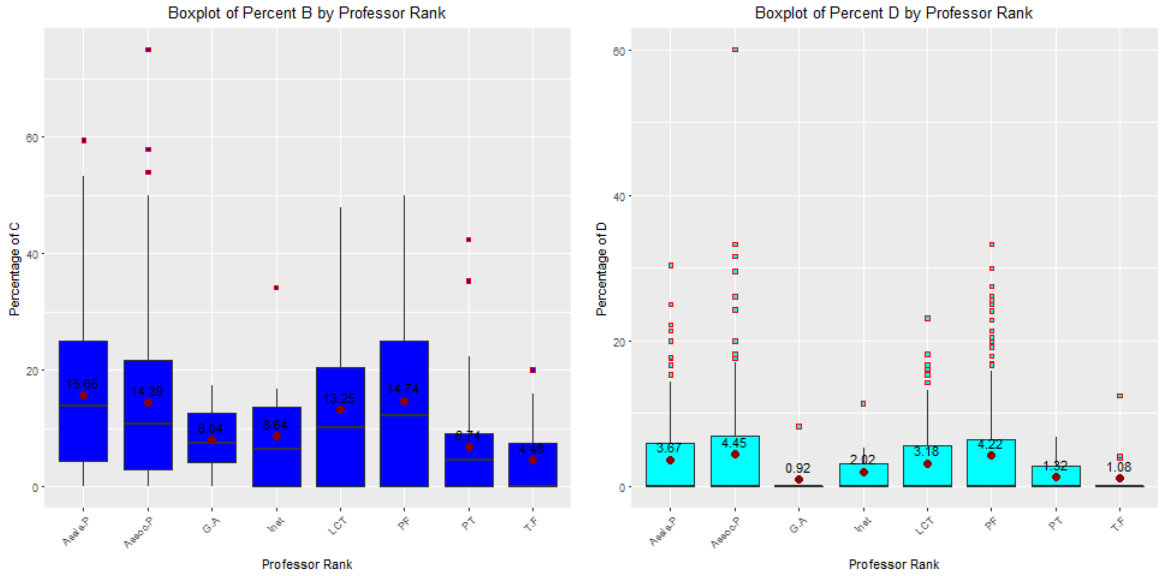


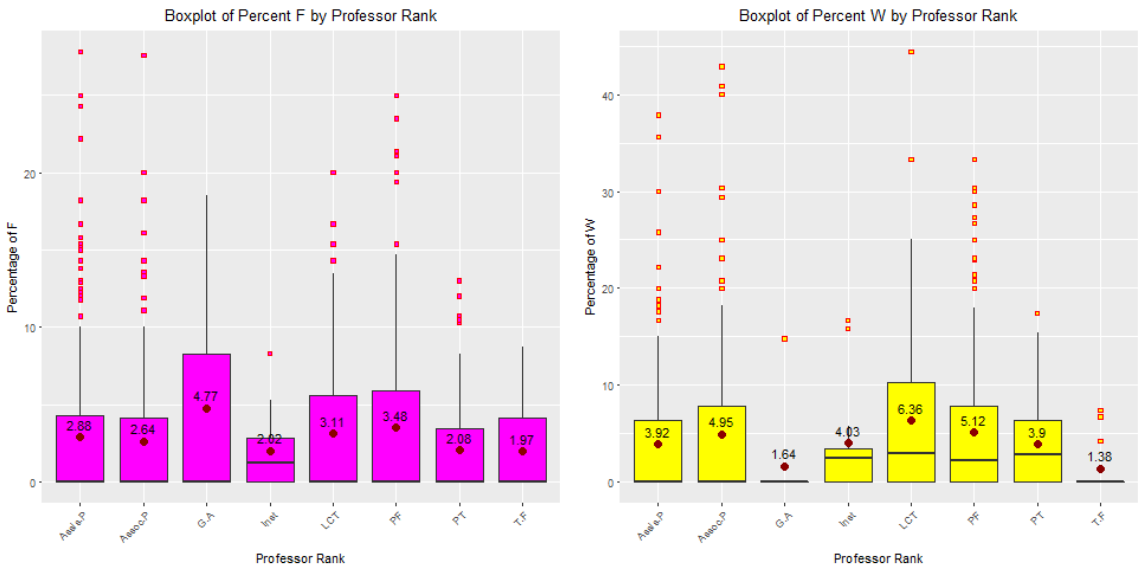
Figure 4.13: Boxplot of professor

When it came to rank of professor it was very interesting. With percent A we saw that all of the full-time positions had lower means of A's than the part-time positions. There was roughly a 15% decrease in percent A means by the full time positions.



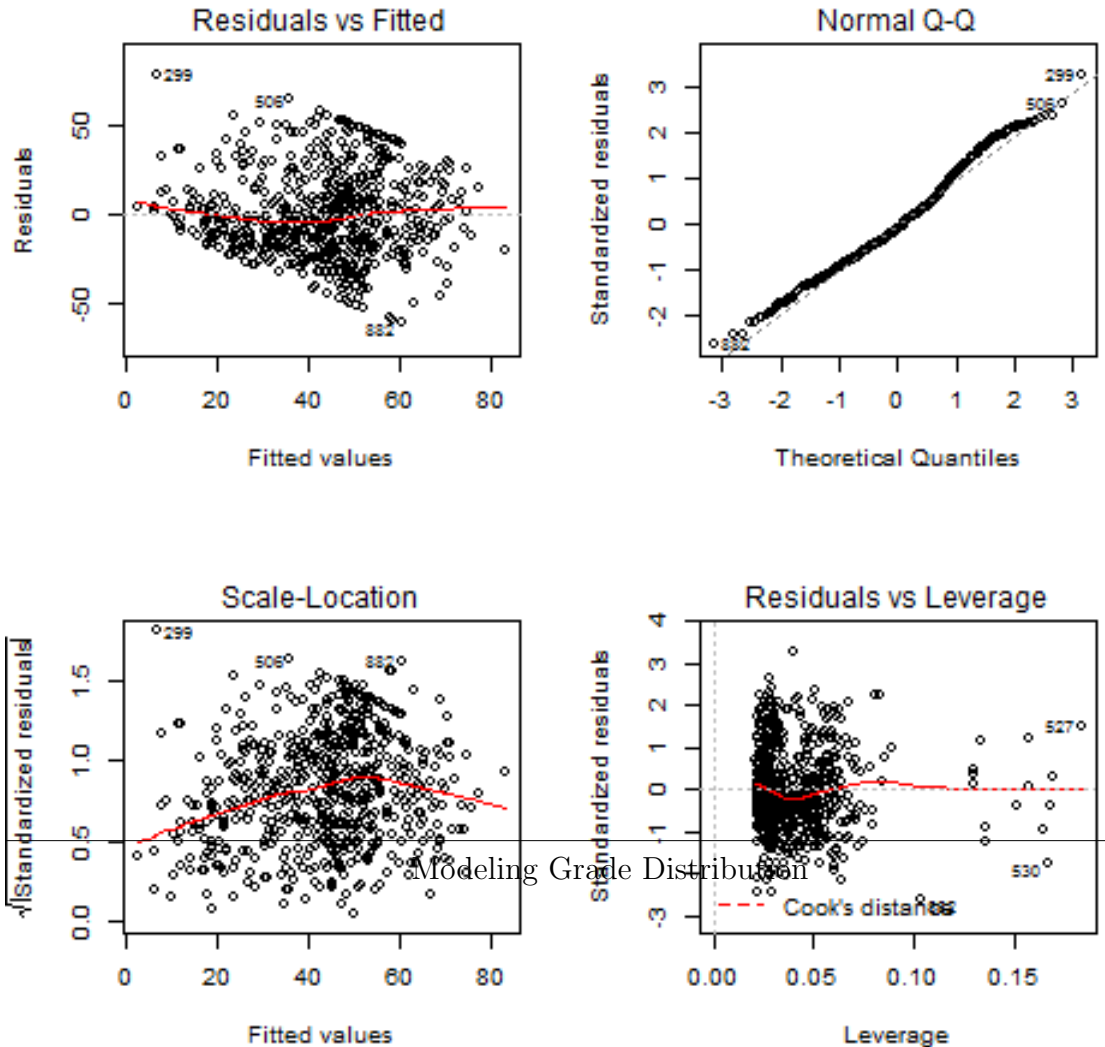
Percent B, C and D were a flip from percent A, where all full-time positions had higher means than those part-time positions across the board.

Figure 4.14: Boxplot of professor



When it came to percentage of F's and W's the means fluctuated between full time and part time positions . But the highest mean of F's were given by the graduate assistants while the highest mean of W's was given by the lecturers.

Figure 4.15: Boxplot by professor



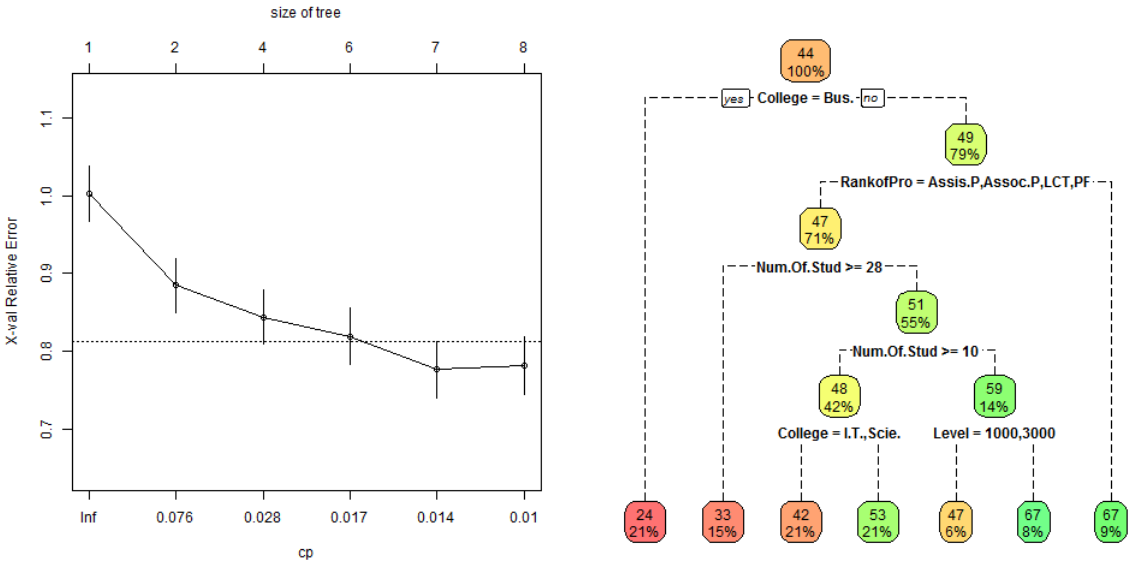


Figure 4.16: Percent A tree and cp

We can see from the cp that the tree with the best prediction will have 7 nodes. The college of business which took up 21% of our data has 24% of A. While courses other than business taught by instructors, graduate assistants, part-time and teaching fellows take up 9& of our data and have 67% A's.

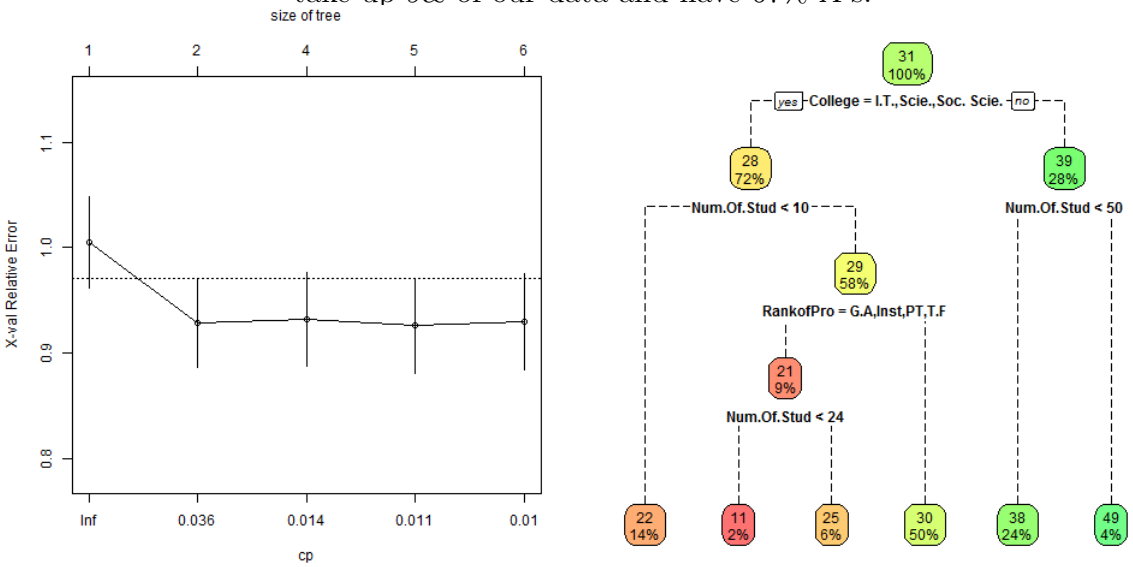


Figure 4.17: Percent B tree and cp

From the cp for percent B we can see that the tree with 6 nodes would be best. A class from the college of I.T, science and social science with less that 10 students took up 14% of the data and had 22% B's. A class taught in the college of Business or health science with more than 50 students took up 4% of the data and had 49% B's.

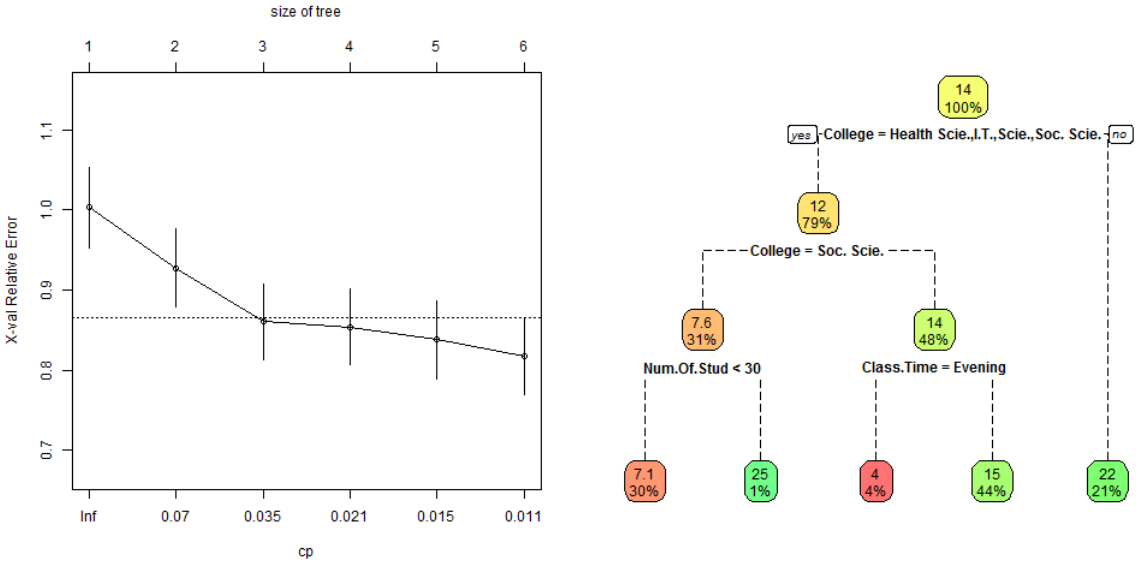


Figure 4.18: Percent C tree and cp

From the cp we can see that the best tree for prediction of percentage of C will have 6 nodes. Although the best tree would have 6 nodes ours has 5. We can see that courses in the college of Business took up 21% of the data and had 22% C's, while the college of social science with greater than 30 students had 25% C's.

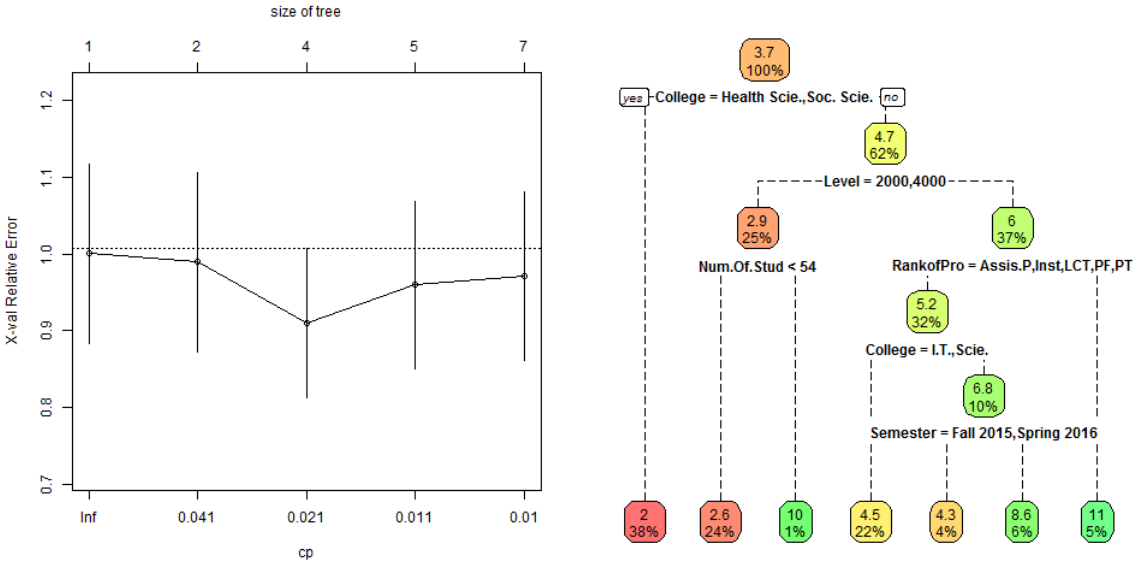


Figure 4.19: Percent D tree and cp

The tree with the best prediction of percent D will have 4 nodes but our tree has 7. We can see from our regression tree that courses in the college of health science and social science took up 38% of our data but only had 2% D's. The college of science, I.T, and business from level 200 and 400 with 54 or more students took up 1% of our data and had 10% D's.

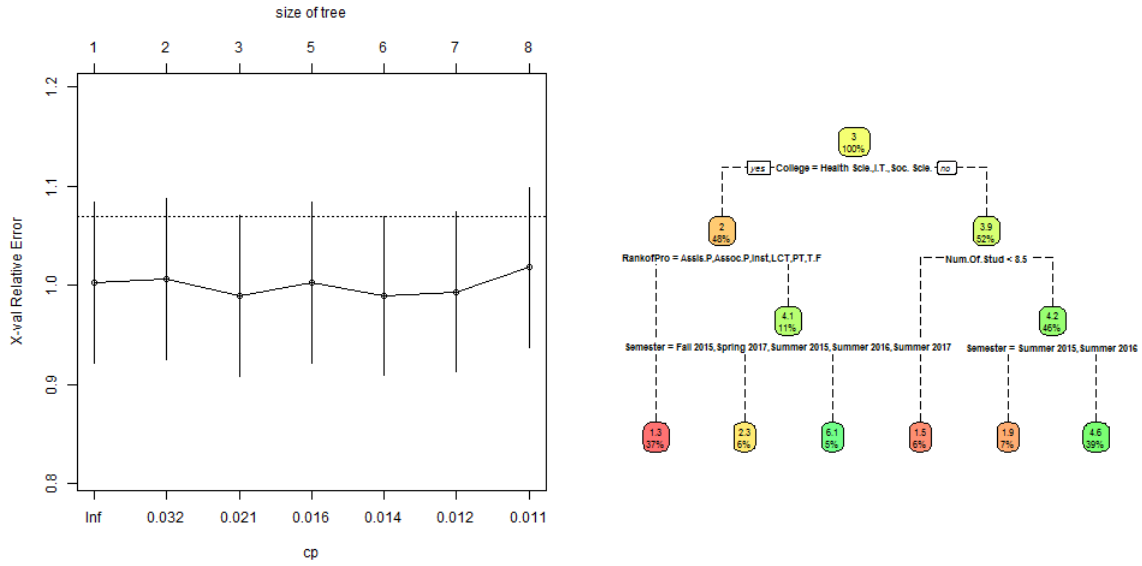


Figure 4.20: Percent F tree and cp

The best tree for percent F has 6 nodes. We can see from the college of science and business with more than 8.5 students taught in summer 2015 or summer 2016 had 1.9% F's.

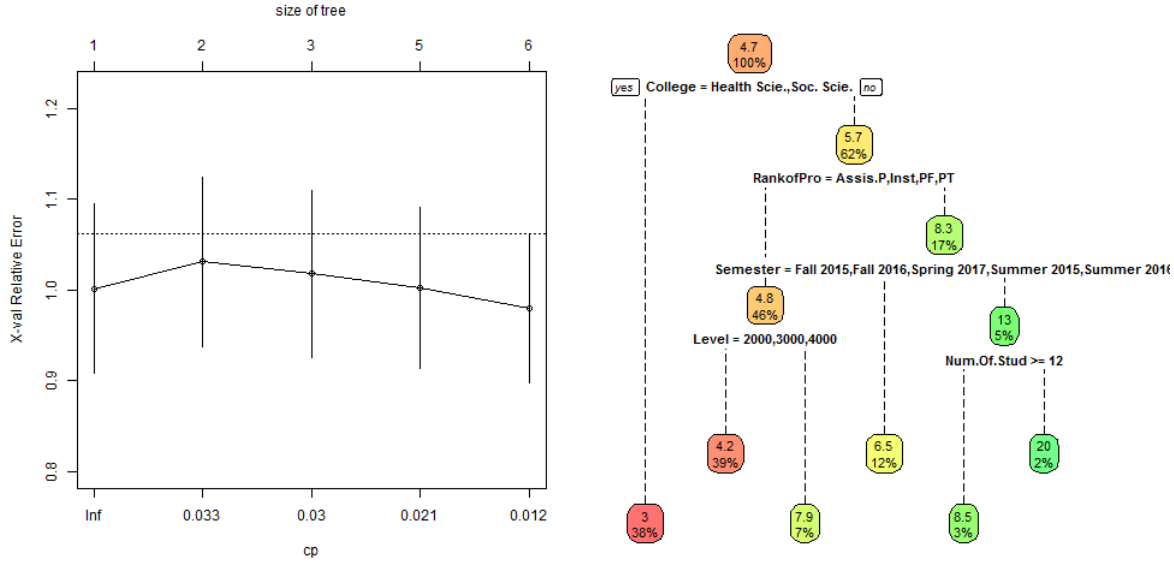


Figure 4.21: Percent W tree and cp

The best tree for percent W will have 6 nodes. From the colleges of Health science and social science they have a total of 3% W's. While the other colleges with assistant, instructor, professor and part-time in the semesters of spring 2016 and summer 2017 in 1000 level courses had 7.9% W's.

From this diagnostics plot we can see all of our assumptions are satisfied. By looking at the regression for percentage of A we can see that all of the colleges are significantly different from the base college which is the college of business with college of Health Science having 25% more A's than the college of business. We can also see that some semesters like Summer 2016 are also significantly different from the base case of Fall 2015. Summer 2016 has 11% more A's than Fall 2015. When it came to percent A we expected for time of class to be significantly different from the boxplots looked at earlier but the only one that is are the evening and afternoon classes although the evening classes are close to not being significant.

When it comes to percentage of B's we can see that the rank of associate professor has a significant difference from the base case of assistant professor in which associate professor have 5.28% less B's than assistant professors. We can also see that the college of science has a significant difference from the college of business when it comes to the percentage of B's. We can also see that none of the semesters were significantly different from the base semester in our study.

For percent C we can see that all colleges had a significant difference from the base college, with the college of social science having 12% less C's than the college of business. We can also see that afternoon and evening classes where significantly different from the base of morning classes when it came to percentage of C's.

Percent D was significantly different for some of its levels as well as some of its colleges. We can see that part-time gave roughly 3% less D's than the base case of assistant professors. When it came to percentage of D's the number of students did not play a significant role in the percentage of D's received.

When talking about percent of F's the college of social science and health science are significantly different from the base case with both having less F's than the college of business. Summer 2015 and 2017 also were significantly different from the base case. When it came to rank of professor there was only one that had a significant difference. This was shocking because when looking at the boxplot we see that the means fluctuated. All colleges as well as all levels were significantly different from the base case of business and 1000 level courses and they all had a less W's than their base case. With the 4000 level courses having 5.5% less W's than the 1000 level courses.Number of students played a significant role in the number of students who do not complete the course. This was not surprising being that a bigger class makes it harder to have that one-on-one connection with the professor.

The root mean square errors (RMSE) of test data are given for all three models considered:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum(\hat{y} - y)^2}$$

	Mult. Regression	Regression Trees	Random Forest
Percent A	24.73	25.48	24.13
Percent B	28.20	16.62	16.16
Percent C	37.21	11.93	11.58
Percent D	43.41	5.69	5.64
Percent F	43.45	4.67	4.70
Percent W	42.48	6.79	5.86

This chart gives us the error from all three models ran on the training data. From here we can see that the random forest model is the best model with the smallest error across the board.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	21.4862	6.3165	3.40	0.0007
2000	-0.3732	3.7230	-0.10	0.9202
3000	-1.3007	3.6072	-0.36	0.7185
4000	6.0242	3.9450	1.53	0.1273
Fall 2016	4.8241	2.9745	1.62	0.1054
Spring 2016	6.5942	3.0063	2.19	0.0287
Spring 2017	7.8077	2.9895	2.61	0.0092
Summer 2015	10.1848	5.2845	1.93	0.0544
Summer 2016	11.0590	5.1008	2.17	0.0305
Summer 2017	3.3198	5.5839	0.59	0.5524
Assoc.P	-1.1674	2.9577	-0.39	0.6932
G.A	13.5081	9.4922	1.42	0.1552
Inst	34.2966	9.6649	3.55	0.0004
LCT	3.2717	3.4037	0.96	0.3368
PF	3.3495	2.7038	1.24	0.2159
PT	24.4771	5.5089	4.44	0.0000
T.F	21.9650	6.8921	3.19	0.0015
one.Hour.ClassYes	5.0482	2.8664	1.76	0.0787
Midday	0.9162	2.6599	0.34	0.7306
Afternoon	6.6392	2.6989	2.46	0.0142
Evening	9.3787	4.6812	2.00	0.0456
Num.Of.Stud	-0.3796	0.0959	-3.96	0.0001
Health Scie.	25.1384	4.5822	5.49	0.0000
I.T.	19.1162	4.2216	4.53	0.0000
Scie.	18.8085	3.1276	6.01	0.0000
Soc. Scie.	23.7014	3.2504	7.29	0.0000

Table 4.1: Percent A regression

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	34.7140	4.6652	7.44	0.0000
2000	0.7468	2.7497	0.27	0.7860
3000	2.5893	2.6642	0.97	0.3315
4000	1.9079	2.9136	0.65	0.5128
Fall 2016	-1.8149	2.1969	-0.83	0.4091
Spring 2016	-4.0391	2.2204	-1.82	0.0694
Spring 2017	-1.4339	2.2079	-0.65	0.5163
Summer 2015	3.2979	3.9030	0.84	0.3985
Summer 2016	-0.1136	3.7673	-0.03	0.9760
Summer 2017	2.2983	4.1241	0.56	0.5776
Assoc.P	-5.2866	2.1845	-2.42	0.0158
G.A	-8.2765	7.0107	-1.18	0.2383
Inst	-18.1365	7.1382	-2.54	0.0113
LCT	-4.9341	2.5139	-1.96	0.0501
PF	-3.5180	1.9970	-1.76	0.0786
PT	-12.4305	4.0688	-3.06	0.0024
T.F	-14.1084	5.0903	-2.77	0.0058
one.Hour.ClassYes	-5.4308	2.1171	-2.57	0.0106
Midday	1.3424	1.9645	0.68	0.4947
Afternoon	-2.4460	1.9934	-1.23	0.2203
Evening	-2.3048	3.4574	-0.67	0.5053
Num.Of.Stud	0.2791	0.0709	3.94	0.0001
Health Scie.	-3.9166	3.3843	-1.16	0.2476
I.T.	-8.6178	3.1179	-2.76	0.0059
Scie.	-9.4224	2.3100	-4.08	0.0001
Soc. SScie.	-2.2919	2.4007	-0.95	0.3401

Table 4.2: Percent B regression

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	18.4012	3.1368	5.87	0.0000
2000	3.3605	1.8489	1.82	0.0696
3000	3.4721	1.7914	1.94	0.0531
4000	1.6708	1.9591	0.85	0.3941
Fall 2016	-2.9706	1.4772	-2.01	0.0448
Spring 2016	-3.0366	1.4929	-2.03	0.0424
Spring 2017	-3.5144	1.4846	-2.37	0.0182
Summer 2015	-1.5293	2.6243	-0.58	0.5603
Summer 2016	-3.8849	2.5331	-1.53	0.1256
Summer 2017	-0.8860	2.7730	-0.32	0.7495
Assoc.P	1.8553	1.4688	1.26	0.2071
G.A	-0.6300	4.7139	-0.13	0.8937
Inst	-13.5023	4.7996	-2.81	0.0051
LCT	-0.1903	1.6903	-0.11	0.9104
PF	-1.1311	1.3427	-0.84	0.3999
PT	-5.6086	2.7358	-2.05	0.0408
T.F	-1.9885	3.4227	-0.58	0.5615
one.Hour.ClassYes	0.3196	1.4235	0.22	0.8224
Midday	-0.7465	1.3209	-0.57	0.5722
Afternoon	-2.5439	1.3403	-1.90	0.0582
Evening	-6.1618	2.3247	-2.65	0.0082
Num.Of.Stud	0.1884	0.0476	3.95	0.0001
Health Scie.	-10.8157	2.2755	-4.75	0.0000
I.T.	-6.4508	2.0965	-3.08	0.0022
Scie.	-6.8830	1.5532	-4.43	0.0000
Soc. SScie.	-12.0344	1.6142	-7.46	0.0000

Table 4.3: Percent C regression

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	7.8048	1.6192	4.82	0.0000
2000	-0.6598	0.9544	-0.69	0.4896
3000	-0.0774	0.9247	-0.08	0.9333
4000	-2.6547	1.0113	-2.63	0.0089
Fall 2016	-0.0248	0.7625	-0.03	0.9740
Spring 2016	-0.7579	0.7706	-0.98	0.3258
Spring 2017	-0.2261	0.7663	-0.30	0.7681
Summer 2015	-4.1635	1.3546	-3.07	0.0022
Summer 2016	-1.3923	1.3075	-1.06	0.2874
Summer 2017	-2.4368	1.4314	-1.70	0.0892
Assoc.P	1.8912	0.7582	2.49	0.0129
G.A	-2.4093	2.4333	-0.99	0.3225
Inst	-2.7944	2.4775	-1.13	0.2598
LCT	-0.3357	0.8725	-0.38	0.7005
PF	0.3637	0.6931	0.52	0.6000
PT	-3.0667	1.4122	-2.17	0.0303
T.F	-1.0660	1.7667	-0.60	0.5465
one.Hour.ClassYes	0.0837	0.7348	0.11	0.9093
Midday	-1.8186	0.6818	-2.67	0.0079
Afternoon	-1.3842	0.6919	-2.00	0.0459
Evening	-0.5890	1.2000	-0.49	0.6237
Num.Of.Stud	0.0009	0.0246	0.04	0.9721
Health Scie.	-3.7447	1.1746	-3.19	0.0015
I.T.	-1.5616	1.0822	-1.44	0.1495
Scie.	-1.0523	0.8017	-1.31	0.1898
Soc. SScie.	-3.6105	0.8332	-4.33	0.0000

Table 4.4: Percent D regression

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.5556	1.2710	5.16	0.0000
2000	-0.0119	0.7491	-0.02	0.9873
3000	-1.0849	0.7259	-1.49	0.1355
4000	-1.6084	0.7938	-2.03	0.0432
Fall 2016	0.4646	0.5985	0.78	0.4379
Spring 2016	0.3639	0.6049	0.60	0.5476
Spring 2017	-0.9077	0.6015	-1.51	0.1319
Summer 2015	-4.2971	1.0634	-4.04	0.0001
Summer 2016	-1.4875	1.0264	-1.45	0.1478
Summer 2017	-2.3939	1.1236	-2.13	0.0335
Assoc.P	-0.0058	0.5952	-0.01	0.9923
G.A	1.8454	1.9100	0.97	0.3344
Inst	-1.8075	1.9448	-0.93	0.3531
LCT	0.2740	0.6849	0.40	0.6893
PF	-0.1169	0.5441	-0.21	0.8299
PT	-2.5966	1.1085	-2.34	0.0195
T.F	-0.9992	1.3868	-0.72	0.4715
one.Hour.ClassYes	-0.2920	0.5768	-0.51	0.6129
Midday	-0.3261	0.5352	-0.61	0.5425
Afternoon	-1.0116	0.5431	-1.86	0.0630
Evening	0.1786	0.9419	0.19	0.8497
Num.Of.Stud	-0.0110	0.0193	-0.57	0.5702
Health Scie.	-4.1608	0.9220	-4.51	0.0000
I.T.	-1.4315	0.8495	-1.69	0.0925
Scie.	-0.8300	0.6293	-1.32	0.1877
Soc. SScie.	-2.4705	0.6541	-3.78	0.0002

Table 4.5: Percent F regression

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	11.1508	1.8356	6.07	0.0000
2000	-3.1173	1.0819	-2.88	0.0041
3000	-3.6391	1.0483	-3.47	0.0006
4000	-5.5263	1.1464	-4.82	0.0000
Fall 2016	-0.5475	0.8644	-0.63	0.5267
Spring 2016	0.8336	0.8737	0.95	0.3404
Spring 2017	-1.7392	0.8688	-2.00	0.0458
Summer 2015	-3.4607	1.5357	-2.25	0.0246
Summer 2016	-4.1275	1.4823	-2.78	0.0055
Summer 2017	-0.1353	1.6227	-0.08	0.9336
Assoc.P	2.6256	0.8595	3.05	0.0024
G.A	-4.0376	2.7585	-1.46	0.1438
Inst	1.9505	2.8087	0.69	0.4877
LCT	1.8221	0.9891	1.84	0.0660
PF	0.9963	0.7858	1.27	0.2053
PT	-1.0067	1.6010	-0.63	0.5297
T.F	-3.6686	2.0029	-1.83	0.0675
one.Hour.ClassYes	0.3234	0.8330	0.39	0.6980
Midday	0.6262	0.7730	0.81	0.4182
Afternoon	0.6704	0.7843	0.85	0.3931
Evening	-0.2870	1.3604	-0.21	0.8330
Num.Of.Stud	-0.0770	0.0279	-2.76	0.0059
Health Scie.	-2.5525	1.3316	-1.92	0.0557
I.T.	-1.2452	1.2268	-1.01	0.3105
Scie.	-0.6045	0.9089	-0.67	0.5063
Soc. Scie.	-3.4549	0.9446	-3.66	0.0003

Table 4.6: Percent W regression

Chapter 5

Conclusion

When it comes to a students performance in a collegiate level course there are a lot of things that a student can control as well as a lot of things that are set in stone by the university. In this study we were able to look at a few of the things students get to chose as well as a few they can consider when registering for future semester. There are many things that can affect a student getting an A and a F. Although we did not cover the students prior knowledge of a course we can still say that these variables are important. In some select courses the size does not effect your grade overall in the course while in others adding one student could increase or decrease your chances of getting the grade you were hoping for. Pay close attention to the size of a course and the official ranking of your course because that could be a role in whether you get an A of and F.

5.1 References

- James, G., Witten, D., Hastie, T., Tibshirani, R. An Introduction to Statistical Learning with Applications in R.
- Wickham, H. ggplot 2: Elegant Graphics for Data Analysis
- Class size and student performance: 198494 by Graham Gibbs , Lisa Lucas Vanessa Simonite