

# Factors That Affect Student Performance In Exams

AUTHOR  
Mutkallam Warraich

PUBLISHED  
May 8, 2025

## Introduction to the Dataset

The dataset we will be analyzing today focuses on various factors that affect student performance in exams. The “Student Performance Factors” dataset was found in Kaggle<sup>1</sup> and was meant to be used for data analysis for research on projects. There are 20 attributes that were recorded for 6,607 students with detailed information on each. The attributes cover a wide range of factors, including study habits, attendance, parental involvement, and more. The factors are listed below with the last column being their performance in the exams. Along with this, you can also see the first few rows and columns of the data to understand its layout.

	Column.Names			
1	Hours_Studied			
2	Attendance			
3	Parental_Involvement			
4	Access_to_Resources			
5	Extracurricular_Activities			
6	Sleep_Hours			
7	Previous_Scores			
8	Motivation_Level			
9	Internet_Access			
10	Tutoring_Sessions			
11	Family_Income			
12	Teacher_Quality			
13	School_Type			
14	Peer_Influence			
15	Physical_Activity			
16	Learning_Disabilities			
17	Parental_Education_Level			
18	Distance_from_Home			
19	Gender			
20	Exam_Score			

	Hours_Studied	Attendance	Parental_Involvement	Access_to_Resources
1	23	84	Low	High
2	19	64	Low	Medium
3	24	98	Medium	Medium
4	29	89	Low	Medium
5	19	92	Medium	Medium
6	19	88	Medium	Medium

## Fair Principles

The dataset follows FAIR principles in multiple ways. For findability(F), the dataset was published on the website, Kaggle. It is also full of metadata and has unique identifiers. For accessibility(A), the dataset is available in multiple formats and is also on a public sourced website. For interoperability(I), the data uses standard formats and also commonly used vocabulary for compatibility with different IDES and languages. For reusability(R), the dataset has detailed information with no missing values for better analysis. It is also stored in a reliable way to download and use.

## Research Question

The “Student Performance Factors” dataset is highly valuable for understanding how performance is impacted by different attributes. 19 factors were recorded to analyze their relationship to exam scores. For our research, we want to identify which factors have the most significant impact on exam scores.

# Exploring the Data

By using a correlation matrix, we can take a look at the correlation coefficient of every variable with respect to the response variable, Exam Score. However, this method can only be used for quantitative variables. As a result, the variables with the highest correlation seem to be Hours Studied, Attendance, Previous Scores, and Tutoring Sessions. All of these correlation coefficients are positive, indicating that as the number for these variables increase, exam score also increases.

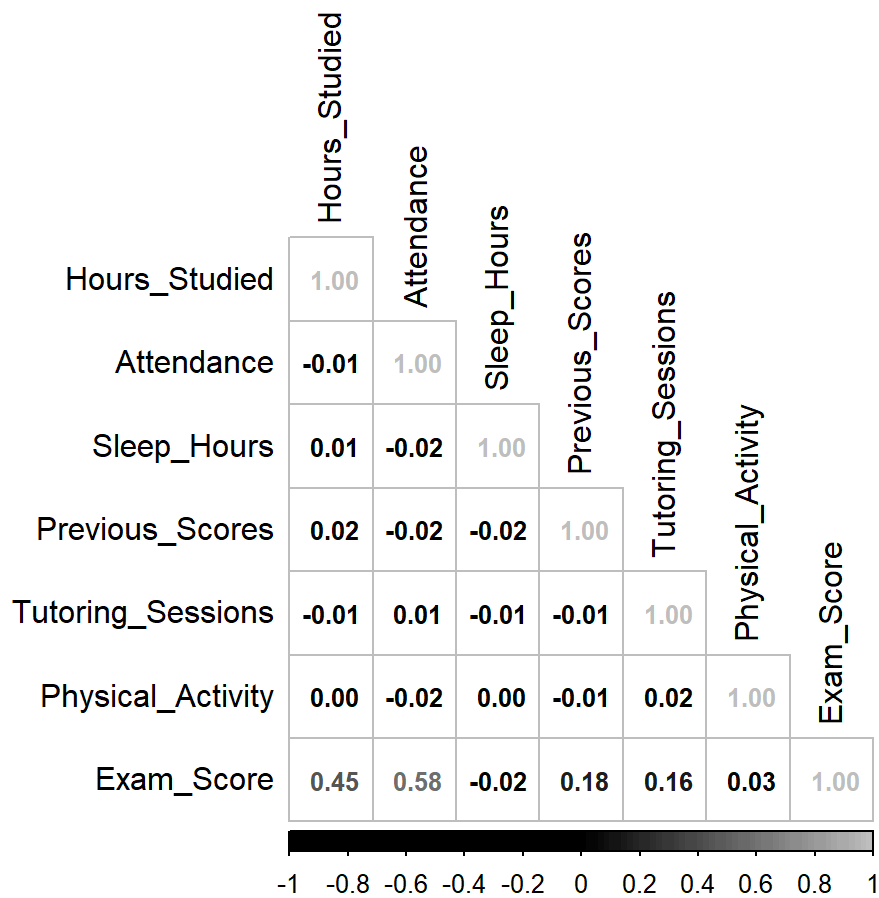
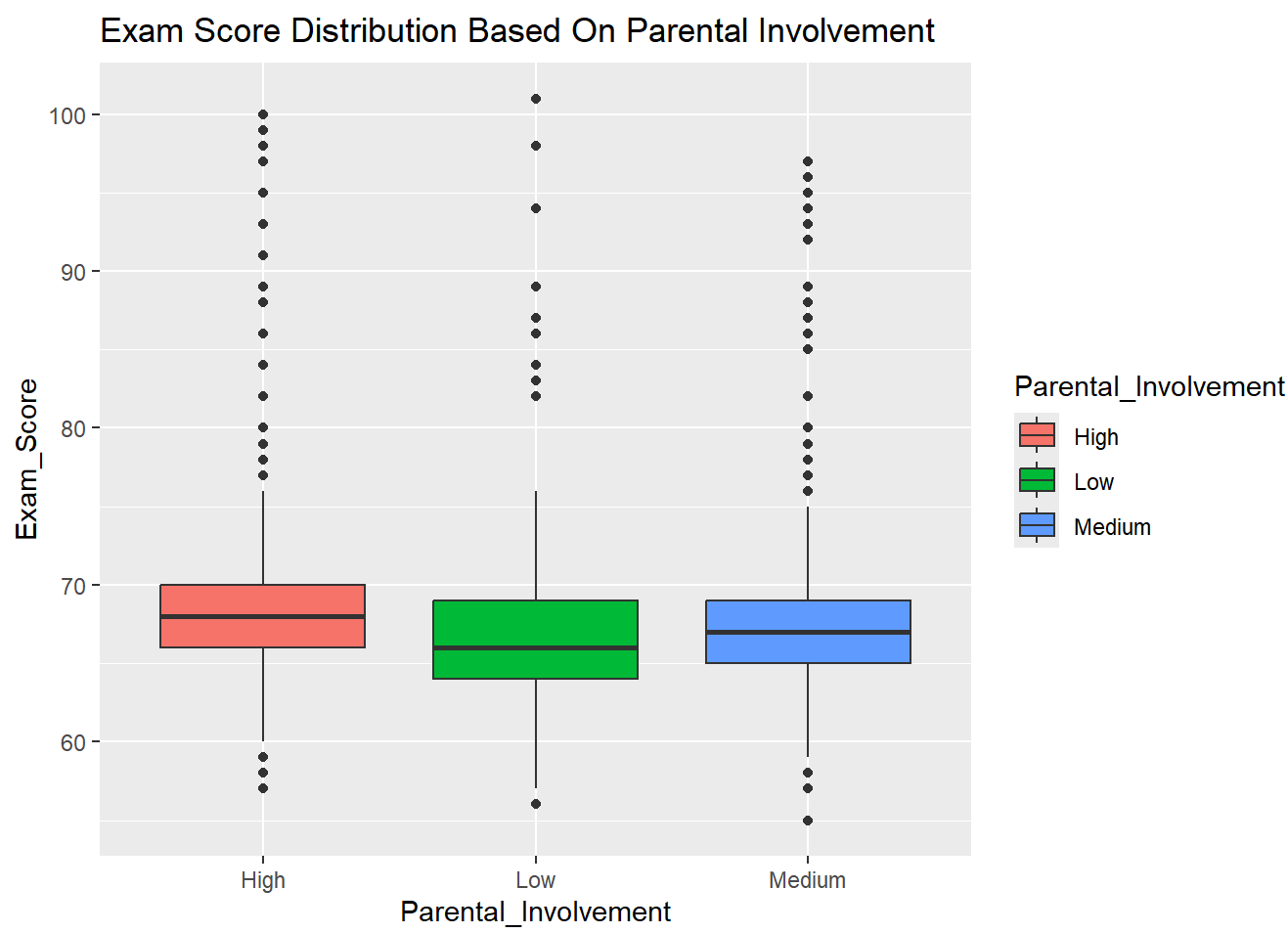
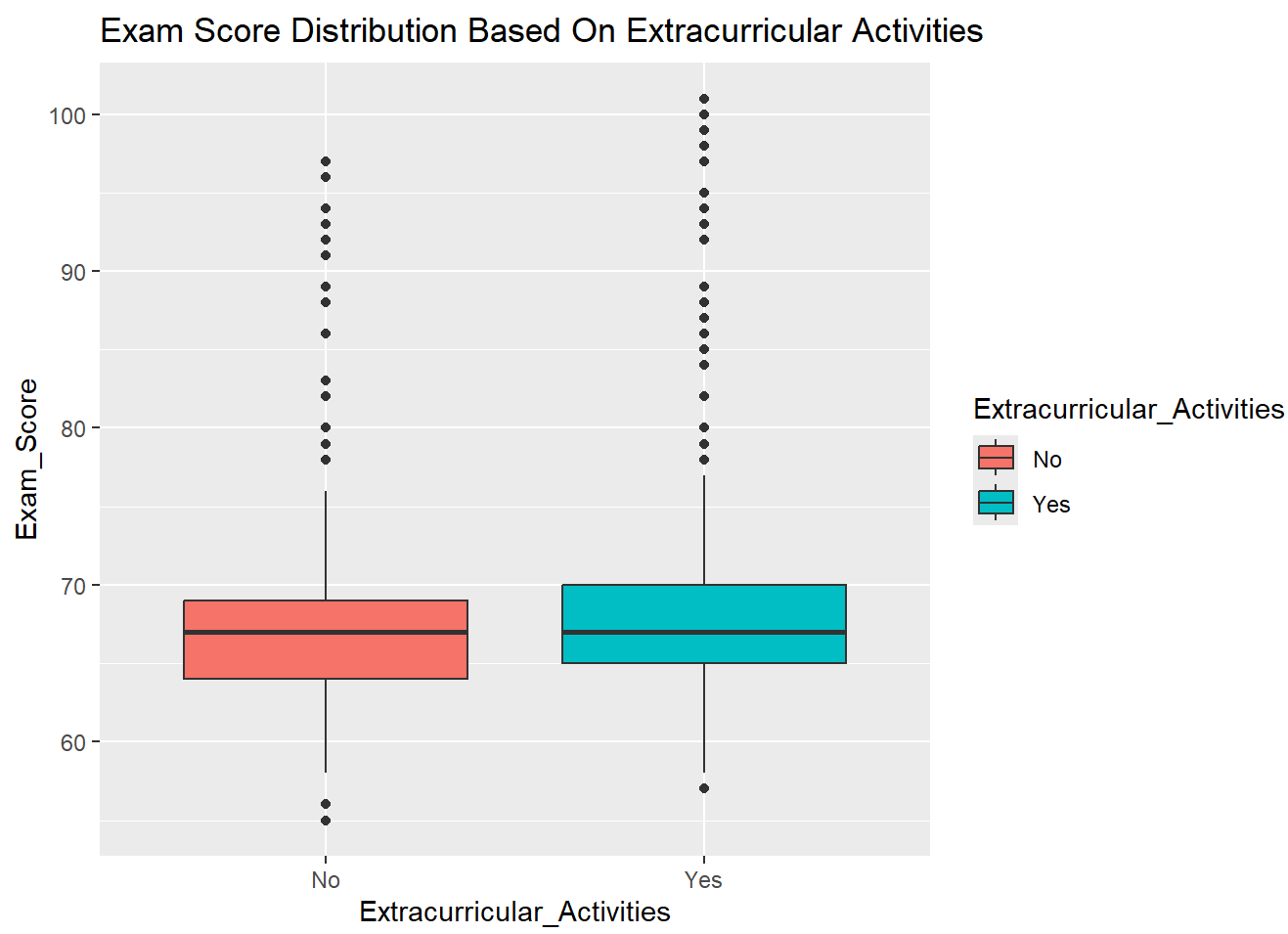
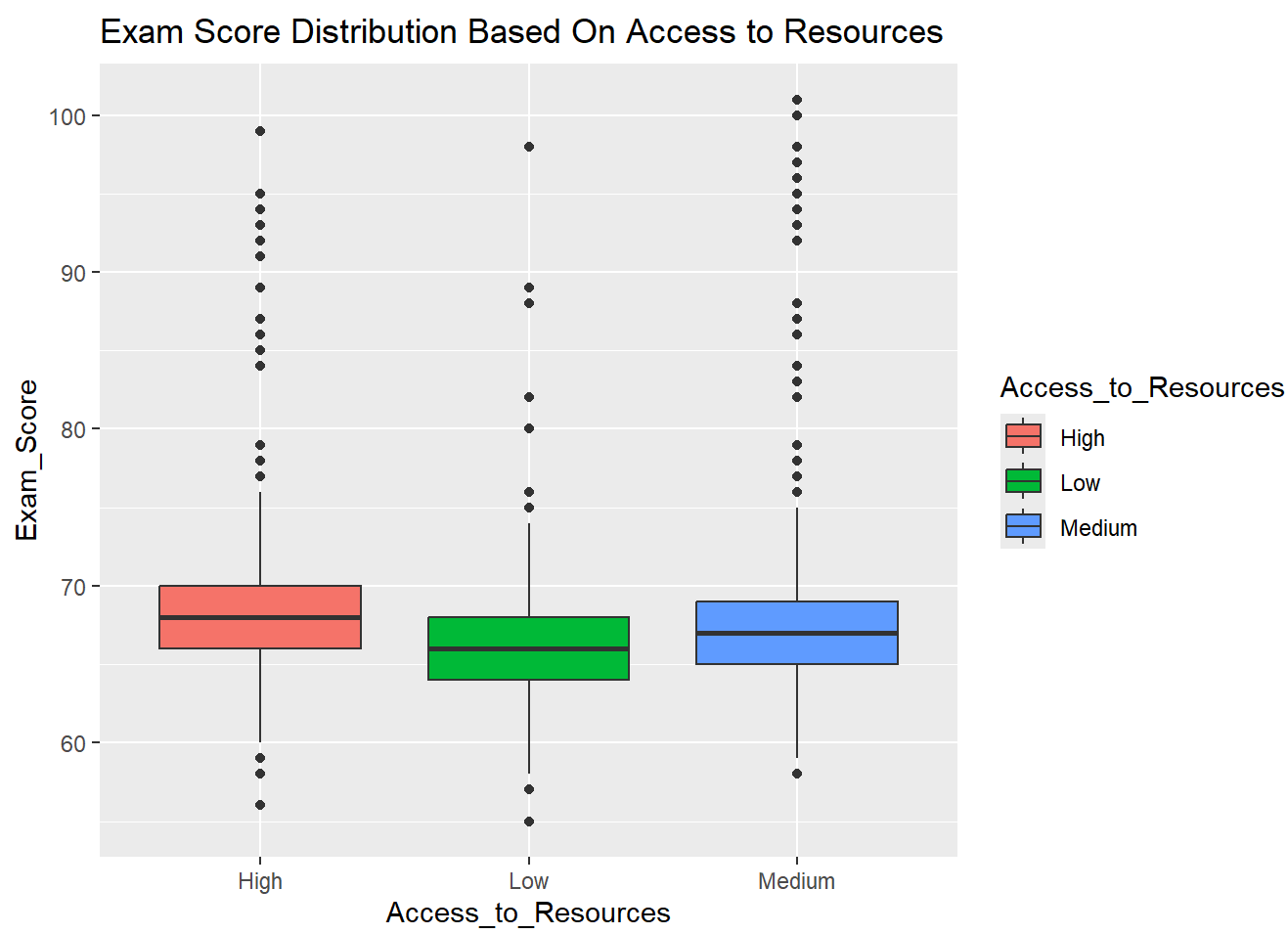


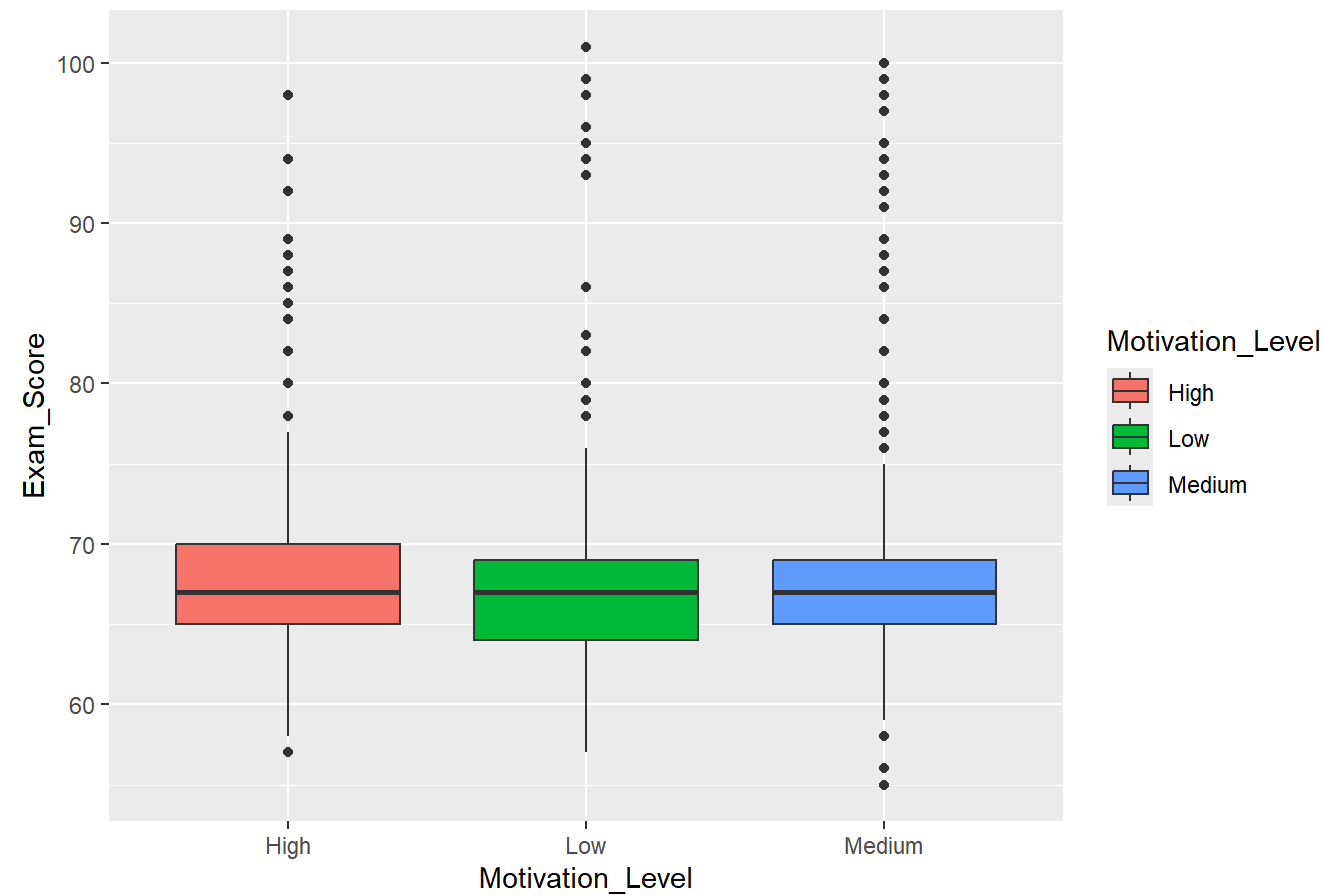
Figure 1: Correlation Matrix

Judging from the correlation matrix, predictors are not highly correlated, so multicollinearity is not a problem. As for the categorical variables, we can use side-by-side boxplots to view the relationship between them and students’ exam scores.



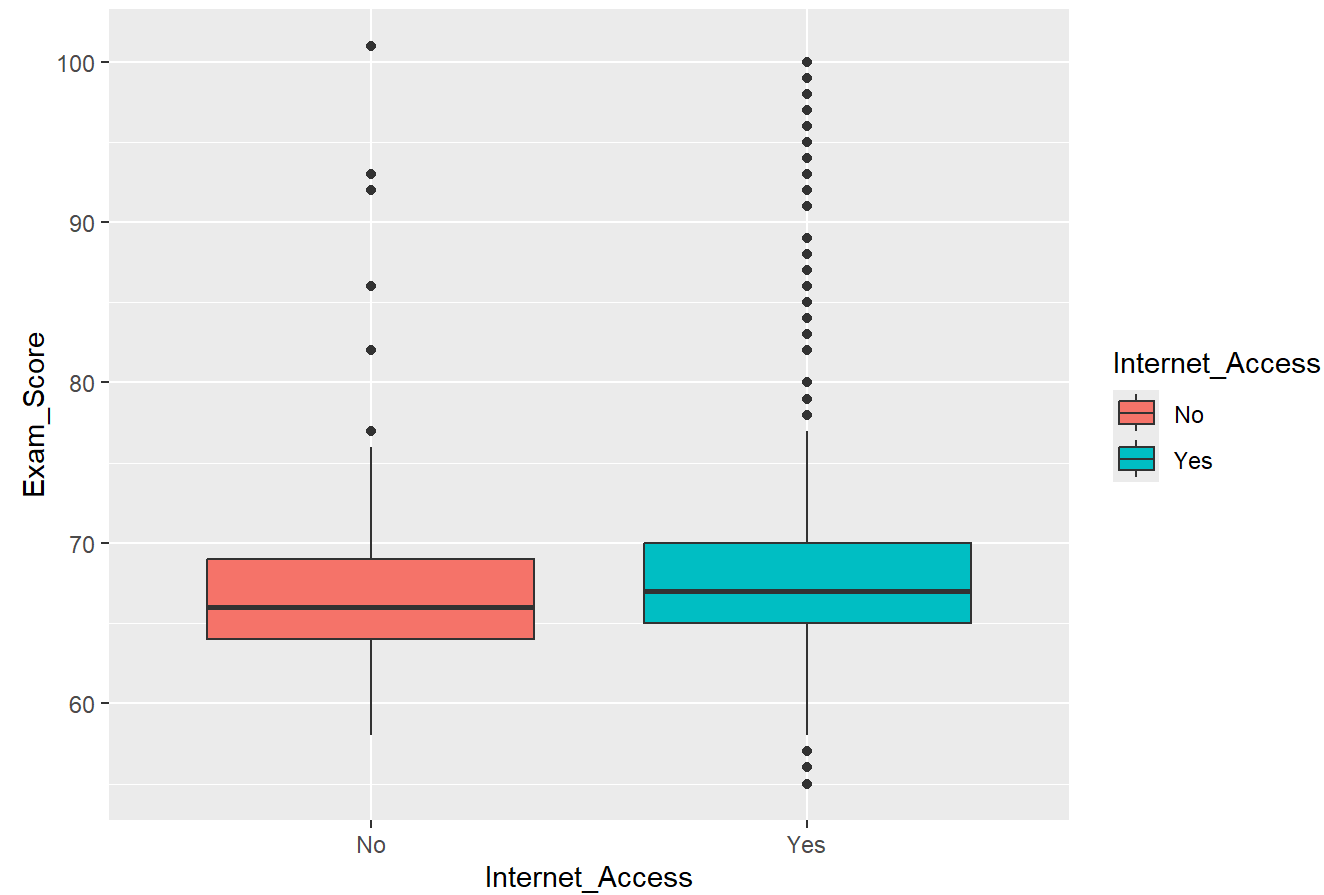


Exam Score Distribution Based On Motivation Level



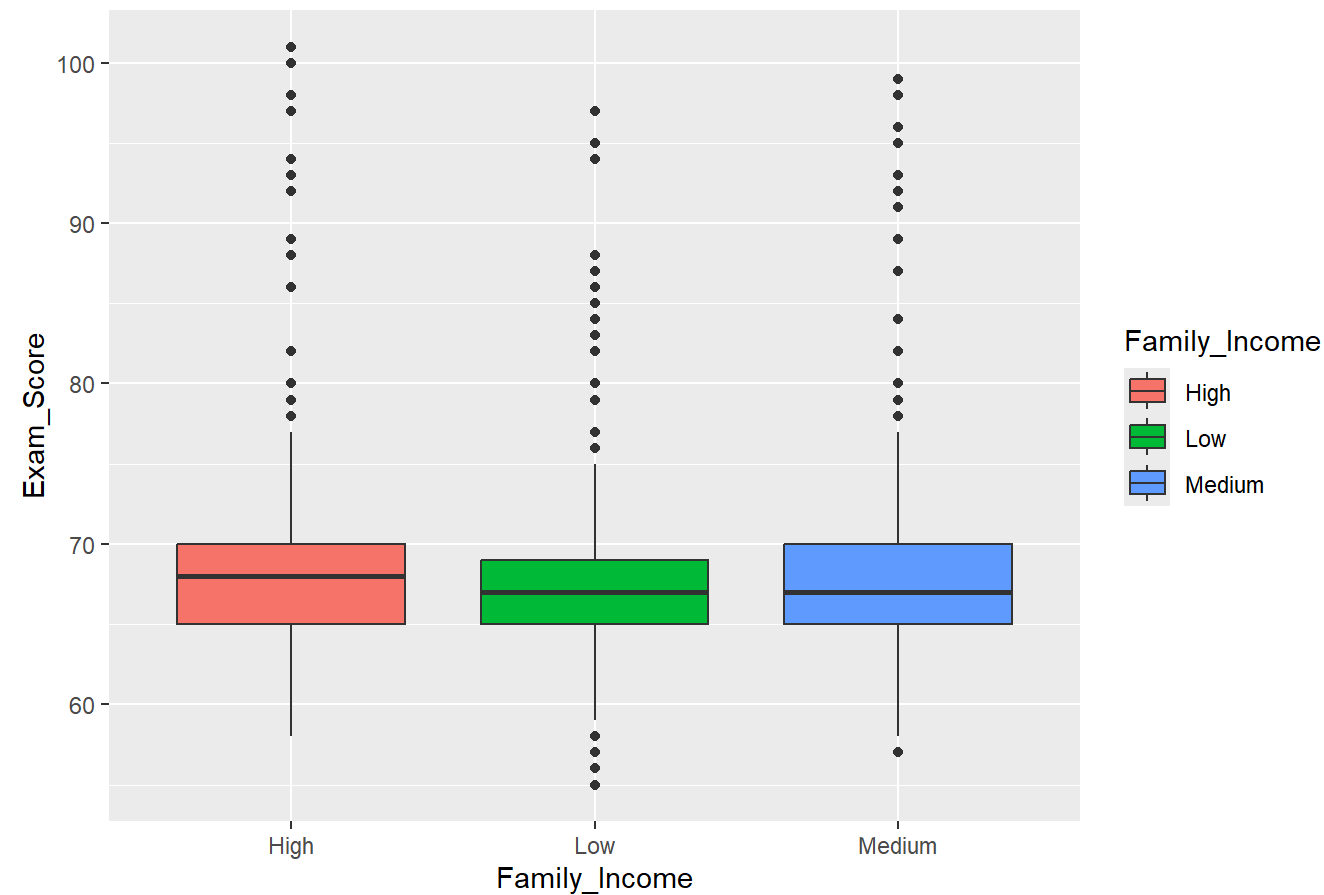
Motivation Level

Exam Score Distribution Based On Internet Access



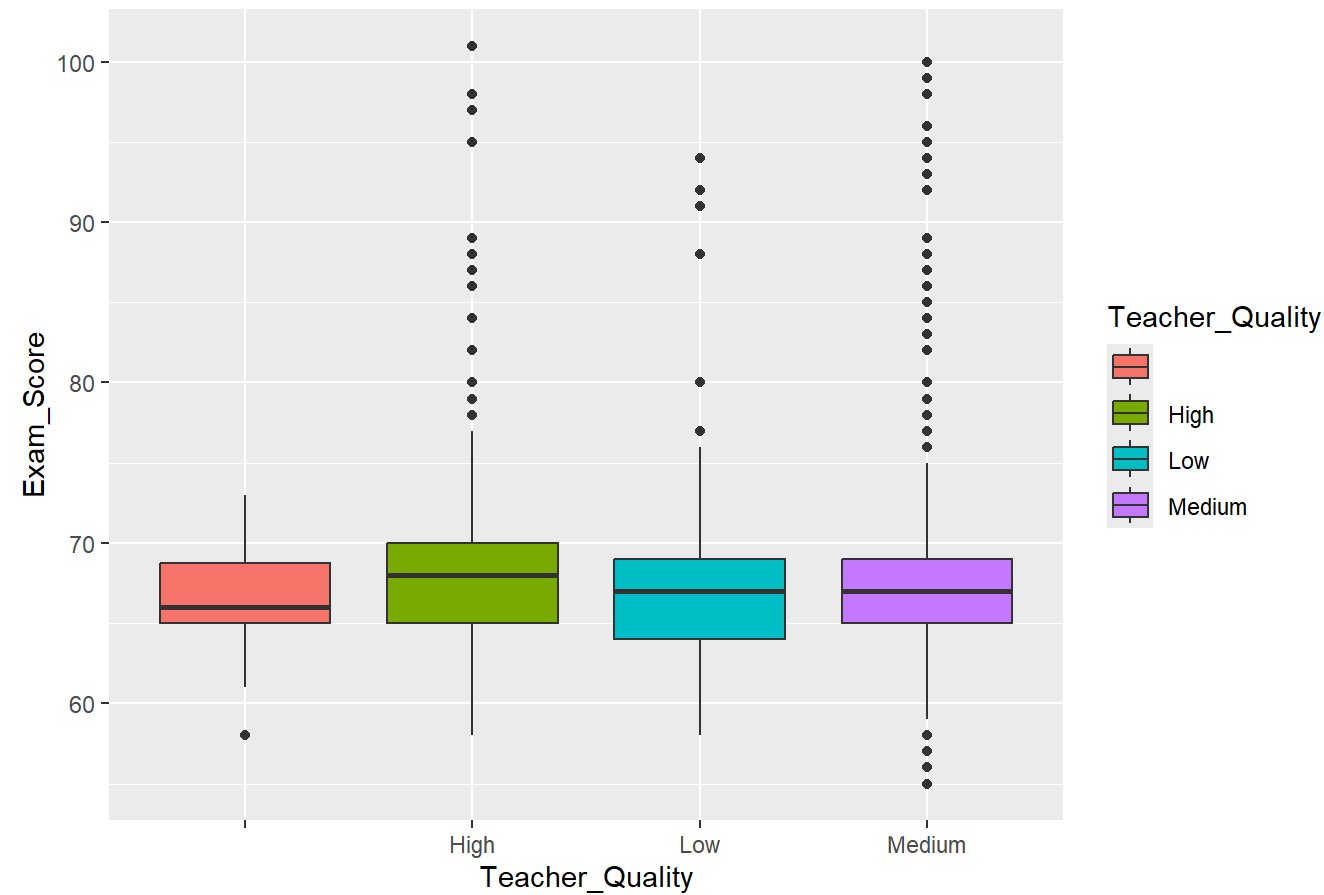
Internet Access

Exam Score Distribution Based On Family Income



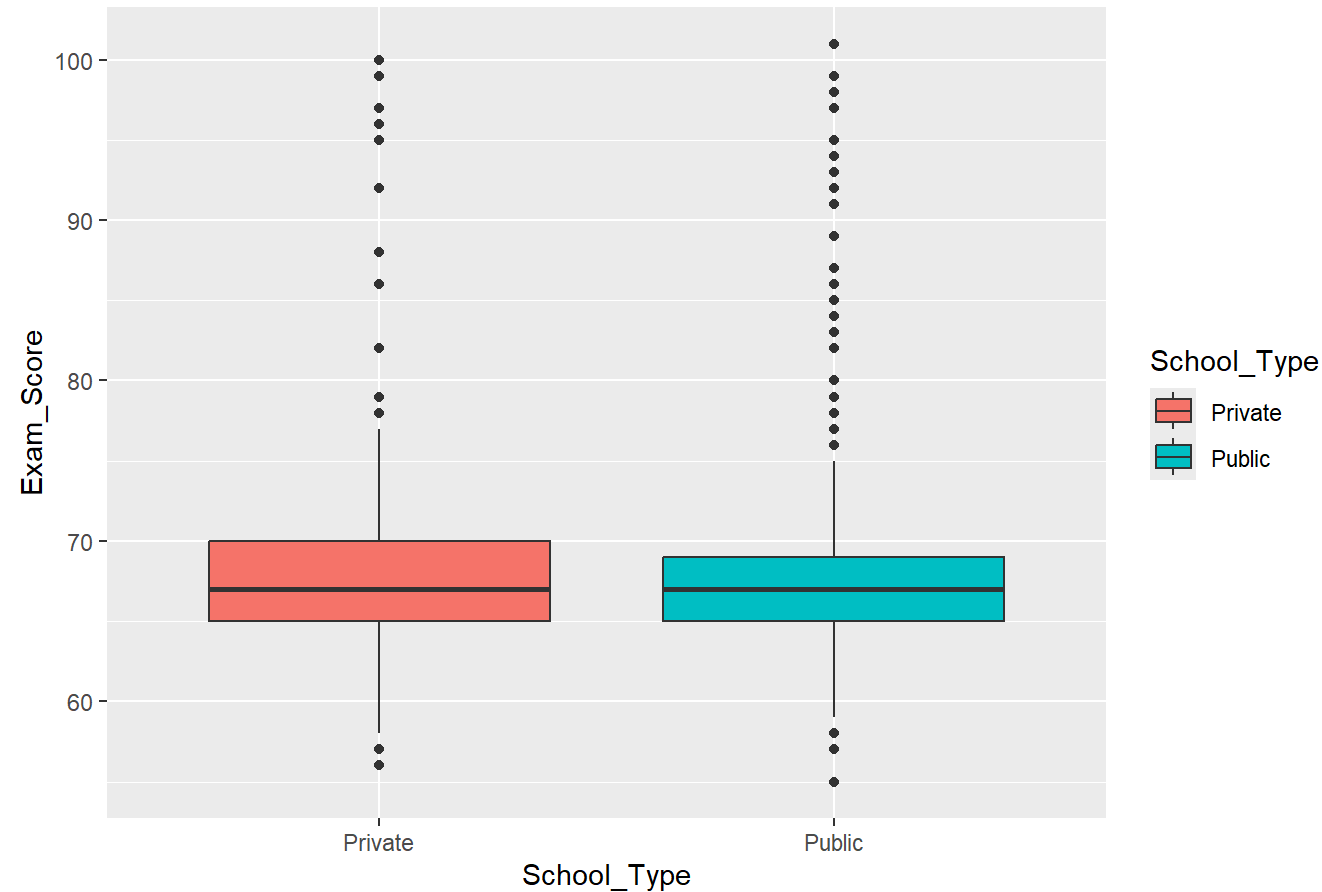
Family Income

Exam Score Distribution Based On Teacher Quality



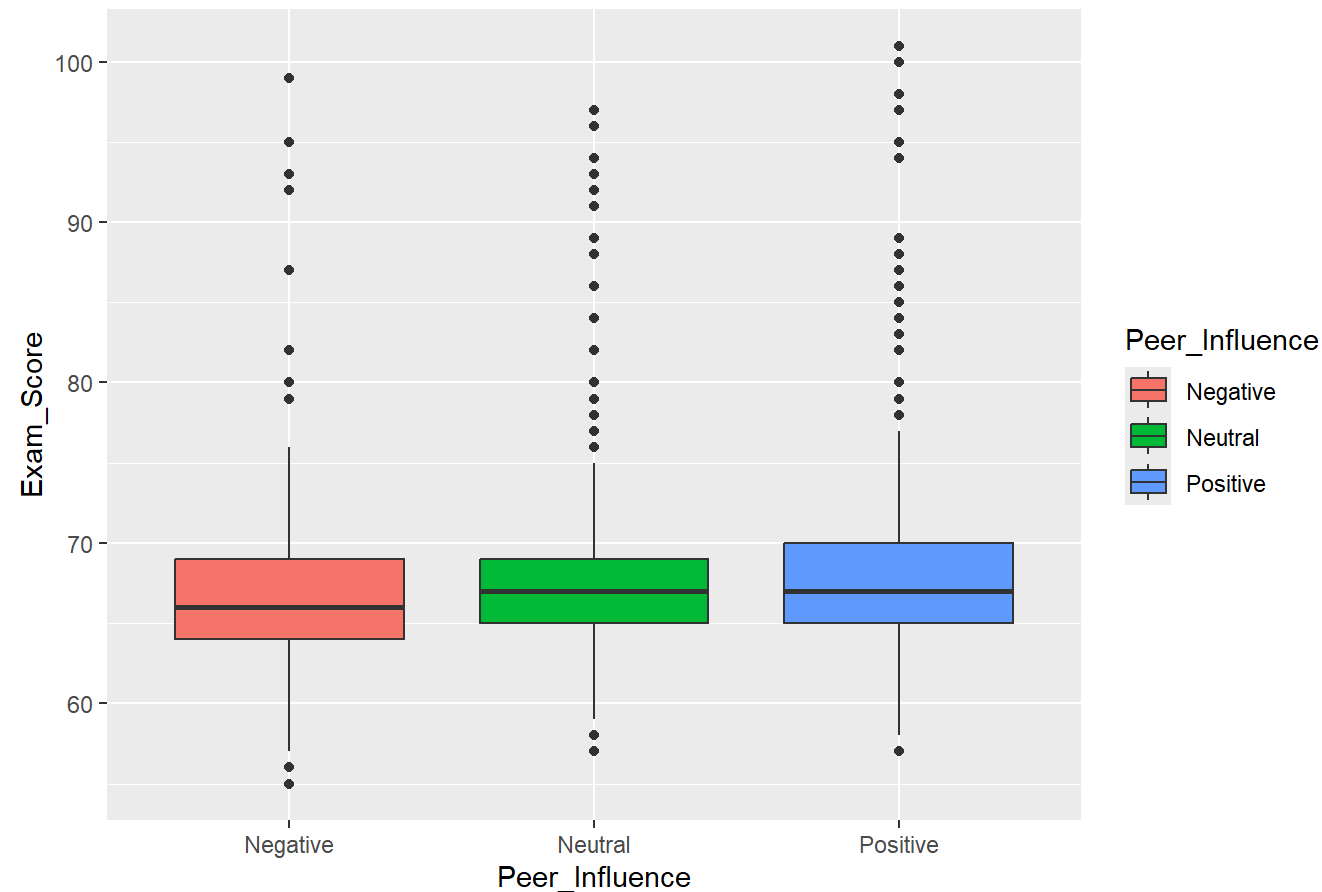
Teacher Quality

Exam Score Distribution Based On School Type



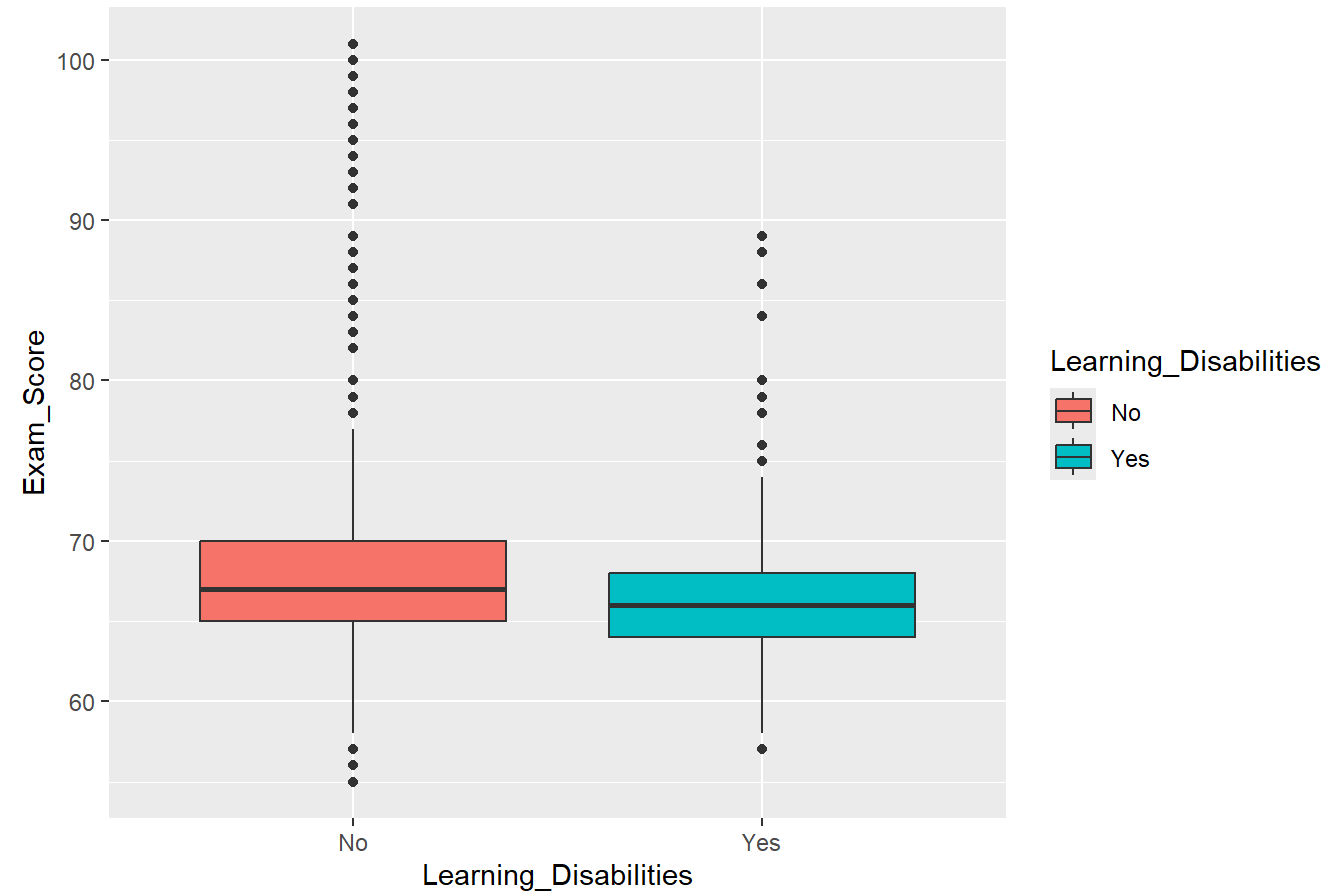
School Type

Exam Score Distribution Based On Peer Influence



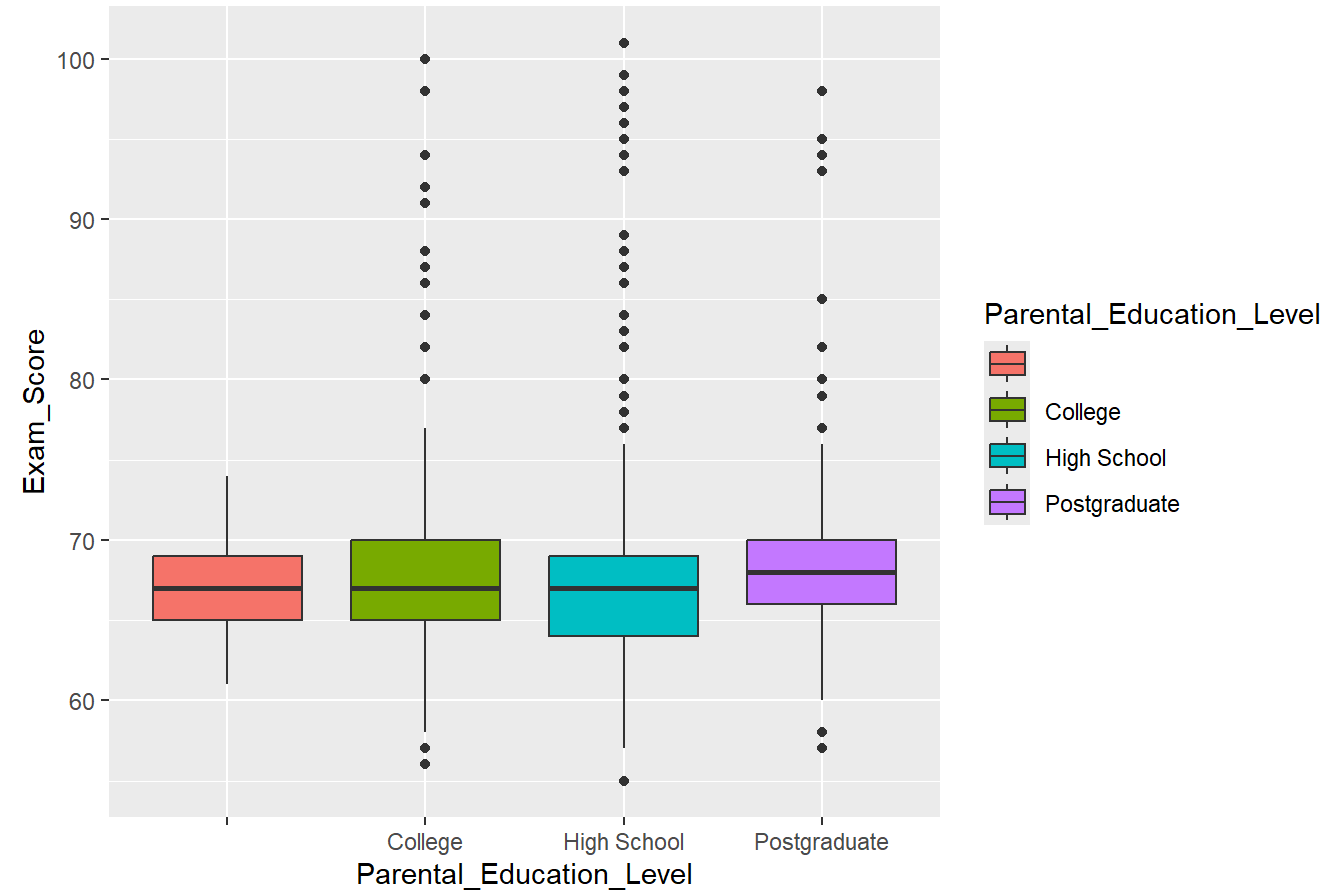
Peer Influence

Exam Score Distribution Based On Learning Disabilities

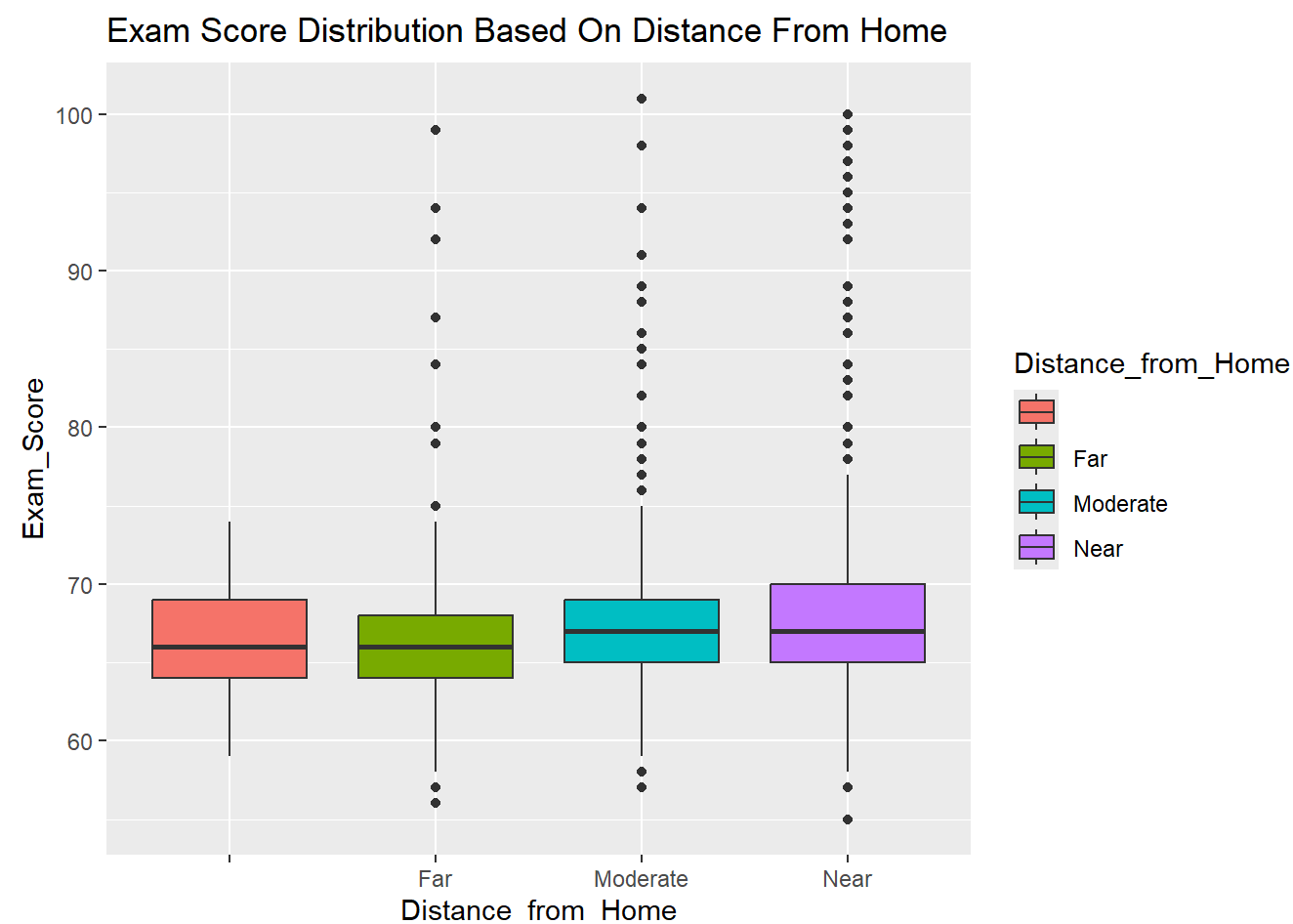


Learning Disabilities

Exam Score Distribution Based On Parental Education Level



Parental Education Level



Distance From Home



Gender

Figure 2: Boxplots for Categorical Variables

Now that we have done some exploratory data analysis, we can take a closer look at some attributes to see how they make up the dataset before starting to answer our research question.

For the following graph, we can see that the average score per tutoring sessions generally increases. This happens till 6 where it starts to level out.



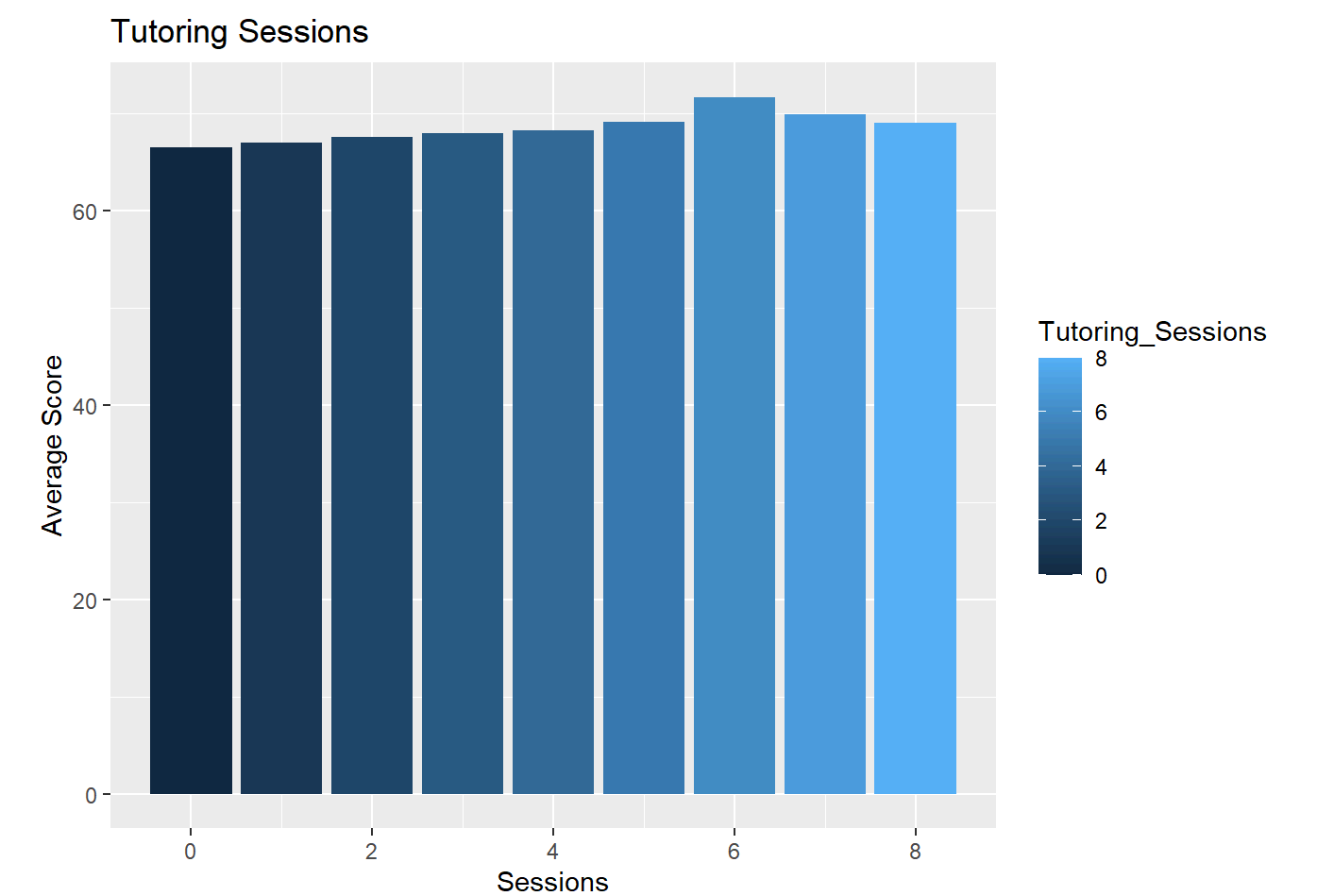


Figure 3: Tutoring Sessions and Average Score

For the following pie chart, we can see the access to resources from low to high and how they are distributed among the population.

### Count of Access to Resources

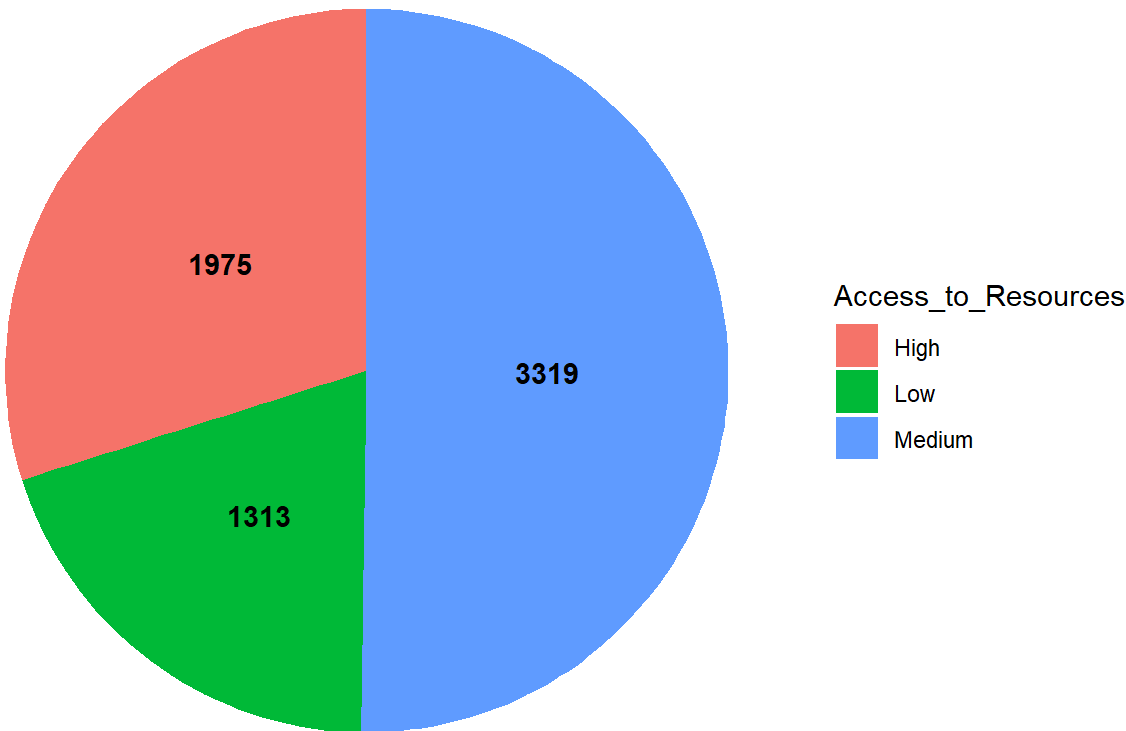


Figure 4: Pie Chart

## Methodology

Now that we have done exploratory data analysis, we can start answering our research question. To determine the best model to fit the data, Akaike’s Information Criteria (AIC) can be used.

$$AIC = n\log\left(\frac{SSE}{n}\right) + 2k$$

Start: AIC=17952.57  
Exam\_Score ~ 1

	Df	Sum of Sq	RSS	AIC
+ Attendance	1	33760	66226	15233
+ Hours_Studied	1	19840	80146	16493
+ Previous_Scores	1	3065	96921	17749
+ Access_to_Resources	2	2882	97104	17763
+ Parental_Involvement	2	2494	97492	17790
+ Tutoring_Sessions	1	2450	97536	17791
+ Parental_Education_Level	3	1098	98888	17886
+ Peer_Influence	2	1021	98965	17889
+ Family_Income	2	894	99092	17897
+ Distance_from_Home	3	868	99119	17901
+ Motivation_Level	2	773	99213	17905
+ Learning_Disabilities	1	724	99263	17907
+ Teacher_Quality	3	622	99364	17917
+ Extracurricular_Activities	1	414	99572	17927
+ Internet_Access	1	265	99721	17937
+ Physical_Activity	1	77	99909	17949
<none>			99986	17953
+ Sleep_Hours	1	29	99957	17953
+ School_Type	1	8	99978	17954
+ Gender	1	0	99986	17955

Step: AIC=15232.79  
Exam\_Score ~ Attendance

	Df	Sum of Sq	RSS	AIC
+ Hours_Studied	1	20358	45868	12808
+ Previous_Scores	1	3491	62736	14877
+ Access_to_Resources	2	3114	63112	14919
+ Parental_Involvement	2	2651	63576	14967
+ Tutoring_Sessions	1	2197	64030	15012
+ Peer_Influence	2	1095	65132	15127
+ Family_Income	2	1043	65183	15132
+ Motivation_Level	2	811	65416	15155
+ Parental_Education_Level	3	799	65428	15159
+ Distance_from_Home	3	636	65590	15175
+ Teacher_Quality	3	603	65623	15178
+ Learning_Disabilities	1	521	65706	15183
+ Extracurricular_Activities	1	431	65795	15192
+ Internet_Access	1	386	65840	15196
+ Physical_Activity	1	167	66059	15218
<none>			66226	15233
+ Sleep_Hours	1	6	66220	15234
+ Gender	1	4	66222	15234
+ School_Type	1	0	66226	15235
- Attendance	1	33760	99986	17953

Step: AIC=12807.96  
Exam\_Score ~ Attendance + Hours\_Studied

	Df	Sum of Sq	RSS	AIC
+ Access_to_Resources	2	3209	42659	12333
+ Previous_Scores	1	3089	42779	12349
+ Parental_Involvement	2	2893	42975	12382
+ Tutoring_Sessions	1	2390	43478	12456
+ Family_Income	2	1034	44834	12661
+ Peer_Influence	2	1005	44863	12666
+ Motivation_Level	2	895	44973	12682
+ Parental_Education_Level	3	866	45002	12688
+ Distance_from_Home	3	717	45151	12710
+ Teacher_Quality	3	632	45236	12722
+ Extracurricular_Activities	1	468	45400	12742

+ Learning_Disabilities	1	430	45438	12748
+ Internet_Access	1	359	45509	12758
+ Physical_Activity	1	151	45717	12788
+ Sleep_Hours	1	16	45852	12808
<none>			45868	12808
+ Gender	1	2	45866	12810
+ School_Type	1	0	45868	12810
- Hours_Studied	1	20358	66226	15233
- Attendance	1	34278	80146	16493

Step: AIC=12332.79  
Exam\_Score ~ Attendance + Hours\_Studied + Access\_to\_Resources

	Df	Sum of Sq	RSS	AIC
+ Parental_Involvement	2	3067	39592	11844
+ Previous_Scores	1	2941	39719	11863
+ Tutoring_Sessions	1	2454	40205	11943
+ Family_Income	2	1050	41610	12172
+ Peer_Influence	2	1008	41651	12179
+ Motivation_Level	2	877	41782	12200
+ Parental_Education_Level	3	881	41778	12201
+ Distance_from_Home	3	714	41945	12227
+ Teacher_Quality	3	664	41995	12235
+ Extracurricular_Activities	1	490	42169	12258
+ Learning_Disabilities	1	415	42244	12270
+ Internet_Access	1	391	42268	12274
+ Physical_Activity	1	164	42495	12309
<none>			42659	12333
+ Sleep_Hours	1	10	42649	12333
+ School_Type	1	3	42656	12334
+ Gender	1	2	42657	12334
- Access_to_Resources	2	3209	45868	12808
- Hours_Studied	1	20453	63112	14919
- Attendance	1	34520	77179	16248

Step: AIC=11843.86  
Exam\_Score ~ Attendance + Hours\_Studied + Access\_to\_Resources +  
Parental\_Involvement

	Df	Sum of Sq	RSS	AIC
+ Previous_Scores	1	3061	36531	11314
+ Tutoring_Sessions	1	2465	37127	11421
+ Family_Income	2	1006	38586	11678
+ Motivation_Level	2	952	38640	11687
+ Peer_Influence	2	934	38658	11690
+ Parental_Education_Level	3	903	38689	11697
+ Distance_from_Home	3	697	38895	11732
+ Teacher_Quality	3	636	38956	11743
+ Extracurricular_Activities	1	549	39043	11754
+ Learning_Disabilities	1	429	39163	11774
+ Internet_Access	1	366	39226	11784
+ Physical_Activity	1	171	39421	11817
<none>			39592	11844
+ Sleep_Hours	1	7	39585	11845
+ School_Type	1	7	39585	11845
+ Gender	1	0	39592	11846
- Parental_Involvement	2	3067	42659	12333
- Access_to_Resources	2	3383	42975	12382
- Hours_Studied	1	20711	60304	14622
- Attendance	1	34707	74300	16001

Step: AIC=11314.2  
Exam\_Score ~ Attendance + Hours\_Studied + Access\_to\_Resources +  
Parental\_Involvement + Previous\_Scores

	Df	Sum of Sq	RSS	AIC
+ Tutoring_Sessions	1	2534	33997	10841
+ Family_Income	2	1054	35477	11125
+ Peer_Influence	2	997	35534	11135
+ Motivation_Level	2	928	35603	11148
+ Parental_Education_Level	3	935	35596	11149
+ Distance_from_Home	3	676	35855	11197
+ Teacher_Quality	3	635	35896	11204
+ Extracurricular_Activities	1	538	35994	11218
+ Learning_Disabilities	1	446	36085	11235
+ Internet_Access	1	357	36174	11251
+ Physical_Activity	1	189	36342	11282
<none>			36531	11314
+ School_Type	1	4	36527	11316
+ Sleep_Hours	1	2	36529	11316
+ Gender	1	0	36531	11316
- Previous_Scores	1	3061	39592	11844
- Parental_Involvement	2	3187	39719	11863
- Access_to_Resources	2	3229	39760	11870
- Hours_Studied	1	20309	56840	14233
- Attendance	1	35105	71636	15762

Step: AIC=10841.15

Exam\_Score ~ Attendance + Hours\_Studied + Access\_to\_Resources +  
 Parental\_Involvement + Previous\_Scores + Tutoring\_Sessions

	Df	Sum of Sq	RSS	AIC
+ Family_Income	2	1038	32959	10640
+ Peer_Influence	2	1013	32984	10645
+ Motivation_Level	2	923	33073	10663
+ Parental_Education_Level	3	923	33073	10665
+ Teacher_Quality	3	632	33365	10723
+ Distance_from_Home	3	630	33367	10724
+ Extracurricular_Activities	1	530	33467	10739
+ Learning_Disabilities	1	462	33535	10753
+ Internet_Access	1	375	33621	10770
+ Physical_Activity	1	165	33832	10811
<none>			33997	10841
+ School_Type	1	2	33994	10843
+ Sleep_Hours	1	1	33996	10843
+ Gender	1	0	33997	10843
- Tutoring_Sessions	1	2534	36531	11314
- Previous_Scores	1	3130	37127	11421
- Parental_Involvement	2	3201	37197	11432
- Access_to_Resources	2	3293	37290	11448
- Hours_Studied	1	20505	54502	13957
- Attendance	1	34840	68837	15500

Step: AIC=10640.37

Exam\_Score ~ Attendance + Hours\_Studied + Access\_to\_Resources +  
 Parental\_Involvement + Previous\_Scores + Tutoring\_Sessions +  
 Family\_Income

	Df	Sum of Sq	RSS	AIC
+ Peer_Influence	2	982	31978	10445
+ Parental_Education_Level	3	923	32036	10459
+ Motivation_Level	2	901	32058	10461
+ Teacher_Quality	3	642	32317	10516
+ Distance_from_Home	3	616	32343	10522
+ Extracurricular_Activities	1	543	32417	10533
+ Learning_Disabilities	1	485	32474	10544
+ Internet_Access	1	373	32586	10567
+ Physical_Activity	1	186	32774	10605
<none>			32959	10640
+ School_Type	1	1	32958	10642

+ Gender	1	0	32959	10642
+ Sleep_Hours	1	0	32959	10642
- Family_Income	2	1038	33997	10841
- Tutoring_Sessions	1	2518	35477	11125
- Parental_Involvement	2	3156	36115	11240
- Previous_Scores	1	3178	36137	11247
- Access_to_Resources	2	3303	36262	11267
- Hours_Studied	1	20487	53447	13832
- Attendance	1	34986	67945	15418

Step: AIC=10444.63

Exam\_Score ~ Attendance + Hours\_Studied + Access\_to\_Resources +  
 Parental\_Involvement + Previous\_Scores + Tutoring\_Sessions +  
 Family\_Income + Peer\_Influence

	Df	Sum of Sq	RSS	AIC
+ Parental_Education_Level	3	918	31059	10258
+ Motivation_Level	2	901	31077	10260
+ Teacher_Quality	3	655	31323	10314
+ Distance_from_Home	3	633	31344	10318
+ Extracurricular_Activities	1	528	31450	10337
+ Learning_Disabilities	1	466	31512	10350
+ Internet_Access	1	386	31592	10366
+ Physical_Activity	1	187	31790	10408
<none>			31978	10445
+ School_Type	1	1	31977	10446
+ Gender	1	0	31977	10447
+ Sleep_Hours	1	0	31978	10447
- Peer_Influence	2	982	32959	10640
- Family_Income	2	1006	32984	10645
- Tutoring_Sessions	1	2534	34512	10946
- Parental_Involvement	2	3081	35059	11048
- Previous_Scores	1	3241	35219	11080
- Access_to_Resources	2	3302	35279	11090
- Hours_Studied	1	20389	52366	13701
- Attendance	1	35049	67027	15332

Step: AIC=10258.1

Exam\_Score ~ Attendance + Hours\_Studied + Access\_to\_Resources +  
 Parental\_Involvement + Previous\_Scores + Tutoring\_Sessions +  
 Family\_Income + Peer\_Influence + Parental\_Education\_Level

	Df	Sum of Sq	RSS	AIC
+ Motivation_Level	2	911	30148	10065
+ Teacher_Quality	3	655	30404	10123
+ Distance_from_Home	3	646	30413	10125
+ Extracurricular_Activities	1	520	30539	10148
+ Learning_Disabilities	1	449	30611	10164
+ Internet_Access	1	392	30667	10176
+ Physical_Activity	1	214	30845	10214
<none>			31059	10258
+ School_Type	1	2	31057	10260
+ Gender	1	0	31059	10260
+ Sleep_Hours	1	0	31059	10260
- Parental_Education_Level	3	918	31978	10445
- Peer_Influence	2	977	32036	10459
- Family_Income	2	1006	32065	10465
- Tutoring_Sessions	1	2523	33582	10772
- Parental_Involvement	2	3104	34163	10884
- Previous_Scores	1	3276	34335	10919
- Access_to_Resources	2	3316	34375	10924
- Hours_Studied	1	20455	51515	13599
- Attendance	1	34732	65791	15215

Step: AIC=10065.32

Exam\_Score ~ Attendance + Hours\_Studied + Access\_to\_Resources +  
Parental\_Involvement + Previous\_Scores + Tutoring\_Sessions +  
Family\_Income + Peer\_Influence + Parental\_Education\_Level +  
Motivation\_Level

	Df	Sum of Sq	RSS	AIC
+ Teacher_Quality	3	671	29477	9922.7
+ Distance_from_Home	3	649	29499	9927.5
+ Extracurricular_Activities	1	499	29648	9957.0
+ Learning_Disabilities	1	448	29700	9968.4
+ Internet_Access	1	369	29779	9985.9
+ Physical_Activity	1	222	29926	10018.6
<none>			30148	10065.3
+ School_Type	1	1	30147	10067.1
+ Gender	1	1	30147	10067.1
+ Sleep_Hours	1	0	30148	10067.3
- Motivation_Level	2	911	31059	10258.1
- Parental_Education_Level	3	929	31077	10259.8
- Peer_Influence	2	976	31124	10271.8
- Family_Income	2	984	31132	10273.6
- Tutoring_Sessions	1	2519	32666	10593.4
- Parental_Involvement	2	3179	33327	10723.7
- Previous_Scores	1	3252	33400	10740.2
- Access_to_Resources	2	3294	33442	10746.5
- Hours_Studied	1	20541	50689	13496.2
- Attendance	1	34780	64928	15132.0

Step: AIC=9922.67

Exam\_Score ~ Attendance + Hours\_Studied + Access\_to\_Resources +  
Parental\_Involvement + Previous\_Scores + Tutoring\_Sessions +  
Family\_Income + Peer\_Influence + Parental\_Education\_Level +  
Motivation\_Level + Teacher\_Quality

	Df	Sum of Sq	RSS	AIC
+ Distance_from_Home	3	654	28823	9780.4
+ Extracurricular_Activities	1	479	28998	9816.3
+ Learning_Disabilities	1	447	29030	9823.7
+ Internet_Access	1	369	29108	9841.5
+ Physical_Activity	1	234	29243	9872.0
<none>			29477	9922.7
+ School_Type	1	1	29476	9924.3
+ Gender	1	1	29476	9924.5
+ Sleep_Hours	1	0	29477	9924.7
- Teacher_Quality	3	671	30148	10065.3
- Parental_Education_Level	3	929	30407	10121.8
- Motivation_Level	2	927	30404	10123.2
- Peer_Influence	2	989	30466	10136.8
- Family_Income	2	993	30470	10137.7
- Tutoring_Sessions	1	2516	31993	10461.8
- Parental_Involvement	2	3144	32621	10588.2
- Previous_Scores	1	3258	32735	10613.2
- Access_to_Resources	2	3323	32800	10624.5
- Hours_Studied	1	20564	50041	13417.3
- Attendance	1	34763	64240	15067.6

Step: AIC=9780.41

Exam\_Score ~ Attendance + Hours\_Studied + Access\_to\_Resources +  
Parental\_Involvement + Previous\_Scores + Tutoring\_Sessions +  
Family\_Income + Peer\_Influence + Parental\_Education\_Level +  
Motivation\_Level + Teacher\_Quality + Distance\_from\_Home

	Df	Sum of Sq	RSS	AIC
+ Extracurricular_Activities	1	496	28327	9667.7
+ Learning_Disabilities	1	447	28376	9679.1
+ Internet_Access	1	379	28444	9694.9

+ Physical_Activity	1	228	28595	9730.0
<none>			28823	9780.4
+ School_Type	1	2	28821	9782.0
+ Gender	1	1	28822	9782.3
+ Sleep_Hours	1	0	28823	9782.4
- Distance_from_Home	3	654	29477	9922.7
- Teacher_Quality	3	676	29499	9927.5
- Motivation_Level	2	931	29754	9986.3
- Parental_Education_Level	3	943	29766	9987.0
- Family_Income	2	979	29802	9997.0
- Peer_Influence	2	1007	29830	10003.3
- Tutoring_Sessions	1	2469	31292	10321.4
- Parental_Involvement	2	3125	31948	10456.6
- Previous_Scores	1	3237	32060	10481.5
- Access_to_Resources	2	3319	32142	10496.6
- Hours_Studied	1	20636	49459	13346.0
- Attendance	1	34544	63367	14983.2

Step: AIC=9667.71

Exam\_Score ~ Attendance + Hours\_Studied + Access\_to\_Resources +  
 Parental\_Involvement + Previous\_Scores + Tutoring\_Sessions +  
 Family\_Income + Peer\_Influence + Parental\_Education\_Level +  
 Motivation\_Level + Teacher\_Quality + Distance\_from\_Home +  
 Extracurricular\_Activities

	Df	Sum of Sq	RSS	AIC
+ Learning_Disabilities	1	440	27887	9566.2
+ Internet_Access	1	384	27943	9579.5
+ Physical_Activity	1	229	28098	9616.2
<none>			28327	9667.7
+ School_Type	1	1	28326	9669.4
+ Gender	1	1	28326	9669.5
+ Sleep_Hours	1	0	28327	9669.7
- Extracurricular_Activities	1	496	28823	9780.4
- Teacher_Quality	3	656	28982	9812.9
- Distance_from_Home	3	671	28998	9816.3
- Motivation_Level	2	908	29235	9872.2
- Parental_Education_Level	3	936	29263	9876.4
- Family_Income	2	991	29318	9891.0
- Peer_Influence	2	992	29319	9891.0
- Tutoring_Sessions	1	2461	30788	10216.2
- Parental_Involvement	2	3182	31509	10367.2
- Previous_Scores	1	3222	31549	10377.4
- Access_to_Resources	2	3343	31670	10400.8
- Hours_Studied	1	20679	49006	13287.2
- Attendance	1	34554	62881	14934.3

Step: AIC=9566.21

Exam\_Score ~ Attendance + Hours\_Studied + Access\_to\_Resources +  
 Parental\_Involvement + Previous\_Scores + Tutoring\_Sessions +  
 Family\_Income + Peer\_Influence + Parental\_Education\_Level +  
 Motivation\_Level + Teacher\_Quality + Distance\_from\_Home +  
 Extracurricular\_Activities + Learning\_Disabilities

	Df	Sum of Sq	RSS	AIC
+ Internet_Access	1	389	27498	9475.5
+ Physical_Activity	1	237	27650	9511.8
<none>			27887	9566.2
+ Gender	1	2	27885	9567.8
+ School_Type	1	1	27885	9567.9
+ Sleep_Hours	1	0	27887	9568.2
- Learning_Disabilities	1	440	28327	9667.7
- Extracurricular_Activities	1	489	28376	9679.1
- Teacher_Quality	3	654	28541	9713.4
- Distance_from_Home	3	671	28558	9717.3

- Motivation_Level	2	907	28794	9773.8
- Parental_Education_Level	3	919	28806	9774.4
- Peer_Influence	2	973	28860	9788.9
- Family_Income	2	1013	28899	9797.9
- Tutoring_Sessions	1	2477	30363	10126.4
- Parental_Involvement	2	3197	31084	10279.3
- Previous_Scores	1	3238	31124	10289.9
- Access_to_Resources	2	3328	31215	10307.2
- Hours_Studied	1	20587	48474	13217.0
- Attendance	1	34365	62251	14869.8

Step: AIC=9475.47

Exam\_Score ~ Attendance + Hours\_Studied + Access\_to\_Resources +  
 Parental\_Involvement + Previous\_Scores + Tutoring\_Sessions +  
 Family\_Income + Peer\_Influence + Parental\_Education\_Level +  
 Motivation\_Level + Teacher\_Quality + Distance\_from\_Home +  
 Extracurricular\_Activities + Learning\_Disabilities + Internet\_Access

	Df	Sum of Sq	RSS	AIC
+ Physical_Activity	1	245	27253	9418.3
<none>			27498	9475.5
+ Gender	1	2	27496	9476.9
+ School_Type	1	1	27497	9477.3
+ Sleep_Hours	1	0	27498	9477.5
- Internet_Access	1	389	27887	9566.2
- Learning_Disabilities	1	445	27943	9579.5
- Extracurricular_Activities	1	494	27992	9591.1
- Teacher_Quality	3	654	28152	9624.7
- Distance_from_Home	3	682	28180	9631.3
- Motivation_Level	2	884	28382	9680.5
- Parental_Education_Level	3	924	28422	9687.8
- Peer_Influence	2	987	28485	9704.5
- Family_Income	2	1011	28509	9710.0
- Tutoring_Sessions	1	2495	29993	10047.3
- Parental_Involvement	2	3170	30668	10192.3
- Previous_Scores	1	3229	30727	10207.0
- Access_to_Resources	2	3360	30858	10233.1
- Hours_Studied	1	20563	48061	13162.6
- Attendance	1	34484	61982	14843.1

Step: AIC=9418.29

Exam\_Score ~ Attendance + Hours\_Studied + Access\_to\_Resources +  
 Parental\_Involvement + Previous\_Scores + Tutoring\_Sessions +  
 Family\_Income + Peer\_Influence + Parental\_Education\_Level +  
 Motivation\_Level + Teacher\_Quality + Distance\_from\_Home +  
 Extracurricular\_Activities + Learning\_Disabilities + Internet\_Access +  
 Physical\_Activity

	Df	Sum of Sq	RSS	AIC
<none>			27253	9418.3
+ Gender	1	3	27250	9419.6
+ School_Type	1	1	27252	9420.0
+ Sleep_Hours	1	0	27253	9420.3
- Physical_Activity	1	245	27498	9475.5
- Internet_Access	1	397	27650	9511.8
- Learning_Disabilities	1	453	27706	9525.3
- Extracurricular_Activities	1	495	27747	9535.1
- Teacher_Quality	3	667	27920	9572.0
- Distance_from_Home	3	676	27929	9574.1
- Motivation_Level	2	892	28145	9627.1
- Parental_Education_Level	3	953	28205	9639.3
- Peer_Influence	2	990	28243	9650.0
- Family_Income	2	1036	28289	9660.9
- Tutoring_Sessions	1	2467	29720	9988.9
- Parental_Involvement	2	3179	30431	10143.1



- Previous_Scores	1	3252	30505	10161.2
- Access_to_Resources	2	3375	30627	10185.6
- Hours_Studied	1	20542	47794	13127.8
- Attendance	1	34596	61849	14831.0

```
Call:
lm(formula = Exam_Score ~ Attendance + Hours_Studied + Access_to_Resources +
  Parental_Involvement + Previous_Scores + Tutoring_Sessions +
  Family_Income + Peer_Influence + Parental_Education_Level +
  Motivation_Level + Teacher_Quality + Distance_from_Home +
  Extracurricular_Activities + Learning_Disabilities + Internet_Access +
  Physical_Activity, data = data)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.3409	-0.4275	-0.1688	0.0897	29.9844

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	41.374861	0.482024	85.836	< 2e-16 ***
Attendance	0.198722	0.002174	91.388	< 2e-16 ***
Hours_Studied	0.294845	0.004187	70.419	< 2e-16 ***
Access_to_ResourcesLow	-2.055236	0.072611	-28.305	< 2e-16 ***
Access_to_ResourcesMedium	-1.003401	0.057952	-17.314	< 2e-16 ***
Parental_InvolvementLow	-1.985092	0.072774	-27.278	< 2e-16 ***
Parental_InvolvementMedium	-1.059759	0.058428	-18.138	< 2e-16 ***
Previous_Scores	0.048867	0.001744	28.021	< 2e-16 ***
Tutoring_Sessions	0.497116	0.020370	24.404	< 2e-16 ***
Family_IncomeLow	-1.080863	0.069498	-15.552	< 2e-16 ***
Family_IncomeMedium	-0.586035	0.069577	-8.423	< 2e-16 ***
Peer_InfluenceNeutral	0.518434	0.068019	7.622	2.85e-14 ***
Peer_InfluencePositive	1.026948	0.067761	15.155	< 2e-16 ***
Parental_Education_LevelCollege	0.331686	0.219807	1.509	0.131351
Parental_Education_LevelHigh School	-0.155897	0.217977	-0.715	0.474511
Parental_Education_LevelPostgraduate	0.825286	0.222247	3.713	0.000206 ***
Motivation_LevelLow	-1.060525	0.072785	-14.571	< 2e-16 ***
Motivation_LevelMedium	-0.543124	0.066296	-8.192	3.05e-16 ***
Teacher_QualityHigh	0.541835	0.235503	2.301	0.021436 *
Teacher_QualityLow	-0.509740	0.244365	-2.086	0.037019 *
Teacher_QualityMedium	-0.005094	0.233189	-0.022	0.982571
Distance_from_HomeFar	-0.440626	0.261416	-1.686	0.091933 .
Distance_from_HomeModerate	-0.052766	0.253317	-0.208	0.835002
Distance_from_HomeNear	0.468349	0.251308	1.864	0.062416 .
Extracurricular_ActivitiesYes	0.558805	0.051141	10.927	< 2e-16 ***
Learning_DisabilitiesYes	-0.854987	0.081729	-10.461	< 2e-16 ***
Internet_AccessYes	0.929040	0.094901	9.790	< 2e-16 ***
Physical_Activity	0.187559	0.024378	7.694	1.64e-14 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.035 on 6579 degrees of freedom  
 Multiple R-squared: 0.7274, Adjusted R-squared: 0.7263  
 F-statistic: 650.3 on 27 and 6579 DF, p-value: < 2.2e-16

Judging from the results of the AIC procedure, it looks like some variables were not included in the final model. These variables include Sleep\_Hours, School\_Type, and Gender. Variables with a high p-value (> 0.05) may not be statistically significant. If this were the case, it is best to look at variables that are statistically significant. This includes all of the variables except for Parental\_Education\_Level, Teacher\_Quality and Distance\_from\_Home. With all of these variables, a larger slope indicates that the variable has a more dramatic effect on the response variable, exam score. The top five variables with the most impact on exam score based on the AIC procedure are Access\_to\_Resources, Parental\_Involvement, Family\_Income, Peer\_Influence, and Motivation\_level respectively.

```
data = read.csv("StudentPerformanceFactors.csv")
data.frame(Column.Names = names(data))
```

```

head(data[1:4])
library(corrplot)
numeric_data <- data[sapply(data, is.numeric)]
correlation_matrix <- cor(numeric_data, method = "pearson")
corrplot(correlation_matrix,
          method = "number",
          type = 'lower',
          tl.col = "black",
          col = colorRampPalette(c("black", "black", "grey"))(100),
          number.cex = .8
        )
library(ggplot2)
ggplot(data, aes(x = Parental_Involvement, y = Exam_Score, fill = Parental_Involvement)) +
  geom_boxplot() +
  ggtitle("Exam Score Distribution Based On Parental Involvement")
ggplot(data, aes(x = Access_to_Resources, y = Exam_Score, fill = Access_to_Resources)) +
  geom_boxplot() +
  ggtitle("Exam Score Distribution Based On Access to Resources")
ggplot(data, aes(x = Extracurricular_Activities, y = Exam_Score, fill = Extracurricular_Activities)) +
  geom_boxplot() +
  ggtitle("Exam Score Distribution Based On Extracurricular Activities")
ggplot(data, aes(x = Motivation_Level, y = Exam_Score, fill = Motivation_Level)) +
  geom_boxplot() +
  ggtitle("Exam Score Distribution Based On Motivation Level")
ggplot(data, aes(x = Internet_Access, y = Exam_Score, fill = Internet_Access)) +
  geom_boxplot() +
  ggtitle("Exam Score Distribution Based On Internet Access")
ggplot(data, aes(x = Family_Income, y = Exam_Score, fill = Family_Income)) +
  geom_boxplot() +
  ggtitle("Exam Score Distribution Based On Family Income")
ggplot(data, aes(x = Teacher_Quality, y = Exam_Score, fill = Teacher_Quality)) +
  geom_boxplot() +
  ggtitle("Exam Score Distribution Based On Teacher Quality")
ggplot(data, aes(x = School_Type, y = Exam_Score, fill = School_Type)) +
  geom_boxplot() +
  ggtitle("Exam Score Distribution Based On School Type")
ggplot(data, aes(x = Peer_Influence, y = Exam_Score, fill = Peer_Influence)) +
  geom_boxplot() +
  ggtitle("Exam Score Distribution Based On Peer Influence")
ggplot(data, aes(x = Learning_Disabilities, y = Exam_Score, fill = Learning_Disabilities)) +
  geom_boxplot() +
  ggtitle("Exam Score Distribution Based On Learning Disabilities")
ggplot(data, aes(x = Parental_Education_Level, y = Exam_Score, fill = Parental_Education_Level)) +
  geom_boxplot() +
  ggtitle("Exam Score Distribution Based On Parental Education Level")
ggplot(data, aes(x = Distance_from_Home, y = Exam_Score, fill = Distance_from_Home)) +
  geom_boxplot() +
  ggtitle("Exam Score Distribution Based On Distance From Home")
ggplot(data, aes(x = Gender, y = Exam_Score, fill = Gender)) +
  geom_boxplot() +
  ggtitle("Exam Score Distribution Based On Gender")
library(dplyr)
library(ggplot2)

tutor_data <- data %>%
  group_by(Tutoring_Sessions) %>%
  summarise(
    Mean_Exam_Score = mean(Exam_Score, na.rm = TRUE))

ggplot(tutor_data, aes(x = Tutoring_Sessions, y = Mean_Exam_Score, fill = Tutoring_Sessions)) +
  geom_bar(stat = "identity") +
  labs(
    title = "Tutoring Sessions",
    x = "Sessions",
    y = "Average Score")

```

```

Resources_data <- data %>%
  group_by(Access_to_Resources) %>%
  summarise(count = n())

ggplot(Resources_data, aes(x = "", y = count, fill = Access_to_Resources)) +
  geom_bar(stat = "identity") +
  coord_polar(theta = "y") +
  geom_text(aes(label = count), position = position_stack(vjust = 0.5), fontface = 'bold') +
  labs(
    title = "Count of Access to Resources") +
  theme_void() +
  theme(
    plot.title = element_text(hjust = 0.5, size = 16, face = "bold"))
nullModel = lm(Exam_Score ~ 1, data = data)
fullModel = lm(Exam_Score ~ Hours_Studied + Attendance + Parental_Involvement + Access_to_Resources, data = data)
stepBothAIC = step(nullModel, scope = list(lower = nullModel, upper = fullModel), direction = "both")
summary(stepBothAIC)
aic_value3 = AIC(stepBothAIC)

```

## Footnotes

1. Me, Practice Data Analysis With. "Student Performance Factors." Kaggle, 26 Nov. 2024, [www.kaggle.com/datasets/lainguyn123/student-performance-factors](https://www.kaggle.com/datasets/lainguyn123/student-performance-factors). ↩