What makes a student perform well?

Binal Modi (bjm470) and Apurv Srivastav (as11419)

Project Description

Project Goal:
Using data collected from [Center for Machine Learning and Intelligent Systems], we aim to explore the dataset with different attributes of high performing students and poor performing students to understand the roles of these associated attributes in a student's performance.

In our project, we will be showing how various outbound factors play a role in students performance.To measure performance, we will consider average grades combining maths and Portuguese (native language) subjects' grades drawn from two Portuguese schools.

Data Set

Attributes:

Attribute	Description	Type	Range and Values
sex	Gender of Student 'F' - female or 'M' - male	Categorical	F, M
age	Student's age	Quantitative	15 - 22
address	student's home address type 'U' - urban or 'R' - rural	Categorical	U, R
Medu	mother's education 0 - without higher education, 1 - with higher education Although the original dataset has a range of values, we will consider them under educated category only if they have higher education for our purpose, so the new categories would be 0(without higher education) or 1(with higher education)	Categorical	0 - 1
Fedu	father's education 0 - without higher education, 1 - with higher education Although the original dataset has a range of values, we will consider them under educated category only if they have higher education for our purpose, so the new categories would be 0(without higher education) or 1(with higher education)	Categorical	0 - 1
Pedu	Parents education 0 - both non educated, 1 - both educated	Categorical	0 - 1

Mjob	mother's job 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other'	Categorical	Multiple
Fjob	father's job 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other'	Categorical	Multiple
studytime	weekly study time 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)	Quantitative	2 - 10
schoolsup	extra educational support yes or no	Categorical	Yes No
famsup	family educational support yes or no	Categorical	Yes No
activities	extra-curricular activities yes or no	Categorical	Yes No
romantic	with a romantic relationship yes or no	Categorical	Yes No
famrel	quality of family relationships 1 - very bad to 5 - excellent	Ordinal	1 - 5
freetime	free time after school 1 - very low to 5 - very high	Ordinal	1 - 5
goout	going out with friends 1 - very low to 5 - very high	Ordinal	1 - 5
Dalc	workday alcohol consumption 1 - very low to 5 - very high	Ordinal	1 - 5
Grade	G3 - final grade numeric: from 0 to 20, output target After manipulation A,B,C,D	Quantitative	Multiple

Analytics Question

Here are the questions we first figured out to address. These questions have their proxy task and proxy values.

How does family background influence a student's performance?

- Does parents education (having a higher secondary education or above) contribute to an individual's performance?
    - medu (Mother's education)
    - fedu (Father's education)
    - Grade
- If the parents' education does contribute to the performance, does the quality of family relationship also matters to improve their performance?
    - pstatus (parent's cohabitation status)
    - famrel (quality of family relationships)
    - Grade

How lifestyle contributes to their performance in school?

- Does having free time after school improve their performance?
    - Freetime (free time after school)
- Having free time but being in a romantic relationship or outing or alcohol consumption takes a toll on the performance?
    - romantic (with a romantic relationship)
    - goout (going out with friends)
    - dalc (workday alcohol consumption)
    - walc (weekend alcohol consumption)
    - Grade
- Even though frequent alcohol consumption, but if they have sufficient study time is the performance affected?
    - studytime (weekly study time)
    - dalc (workday alcohol consumption)
    - walc (weekend alcohol consumption)
    - Grade
- Studios students(with >10 hours study time) but no extracurricular activities have better performance than students with moderate (4-10 hours) study time and extracurricular activities?
    - activities (extra-curricular activities)
    - studytime (weekly study time)
    - Grade

How school and facilities contribute to students performance?

- Does the reason for choosing the school influence their performance?
    - reason (reason to choose this school)
    - Grade

- Does extra support from school improve their performance?
    - schoolsup (extra educational support from school)
    - Grade
- Although extra support from school but no family education support affect their performance?
    - schoolsup (extra educational support from school)
    - famsup (family educational support)
    - Grade

How student's gender and personal experiences influence their performance?

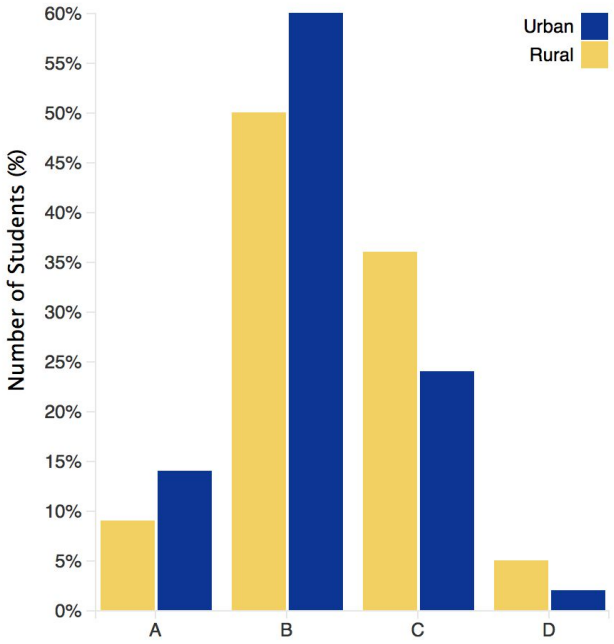
- How performance varies according to the sex and age of a student?
    - sex (gender of student)
    - age (age of student)
    - Grade
- Does the performance vary for students of the same age if they come from a rural or urban area?
    - address (student's home address type)
    - sex (gender of student)
    - age (age of student)
    - Grade
- If previous setbacks and failure affect their performance?
    - failures(number of past class failures)
    - Grade

Data Analysis:

Here are the list of questions we analysed.

How student's gender and personal experiences influence their performance?

- Does the performance vary for students if they come from a rural or urban area?
    - address (student's home address type)
    - Grade



Analysis:

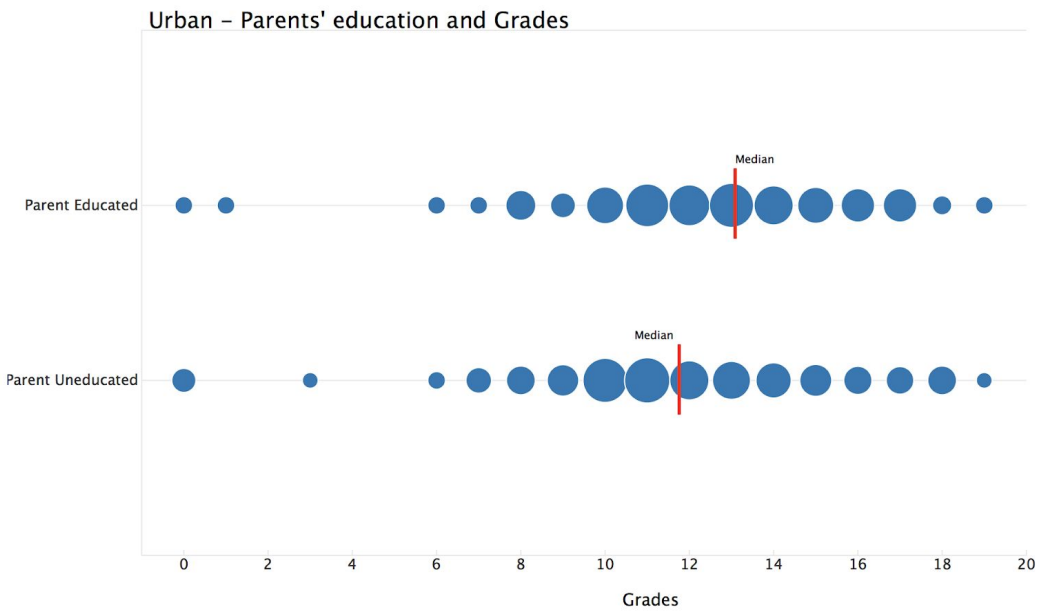
We can clearly see that the percentage of students scoring higher grades A and B us more in case of urban students and comparatively less among rural students.

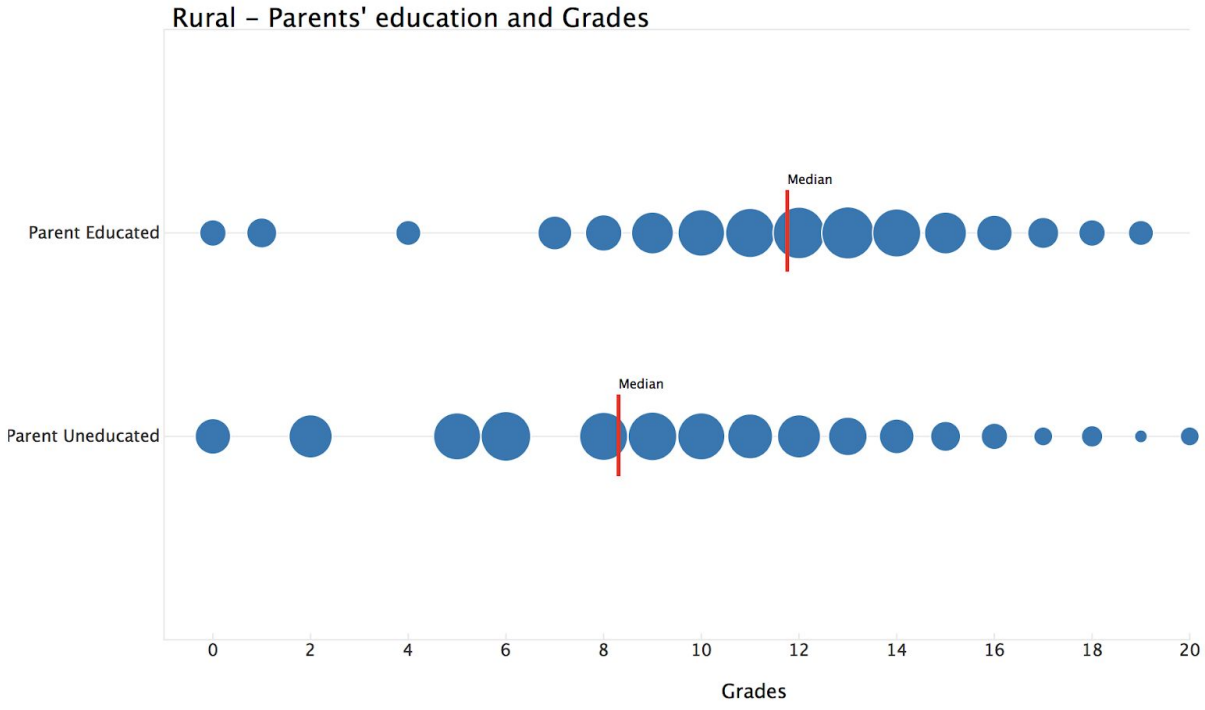
How does family background influence a student's performance?

- Does parents education (having a higher secondary education or above) contribute to an individual's performance?
    - parentEdu (taking in account both Mother's education and Father's education)
    - Grade

To answer this question we did some feature engineering. We categorised education into 2 categories - Educated and Uneducated. Were educated means the person who has complete his/her higher education atleast. We then created a new category called Parent Education which has value 1 if and only if both the parents are educated else we set the value as 0.

Graph and its analysis is below:



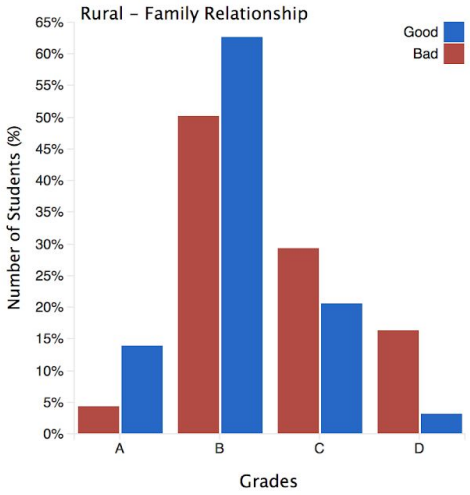


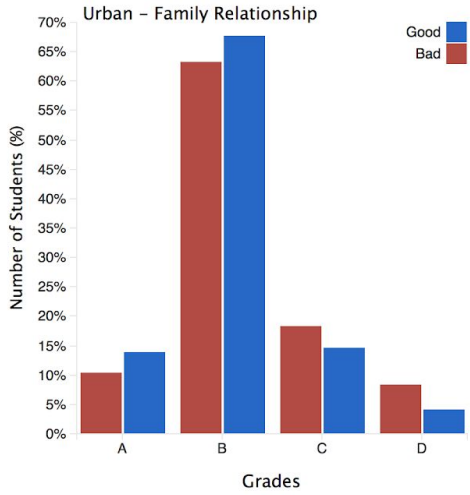
Analysis of the graph:
From the above graph we can clearly say that the education of parent do influence the performance of student of rural areas. But the same is not true for students who belong to urban area.

- If the parents' education does contribute to the performance, does the quality of family relationship also matters to improve their performance?
    - famrel (quality of family relationships - good or bad)
    - Grade (categorical - values - A [16-20], B [11-15], C [6-10], D [0-5])

We then extended our analysis to the family relation attribute. Here we only considered the family relation of student with his/her parents. We redefined the range of “famrel” attribute. We made it a binary value attribute where value 1 represents good relation with parents and 0 means bad relation with parents. Here we also converted the Grade attribute of (which was earlier numerical) into categorical type.

Here what we got the graph for the family relation.





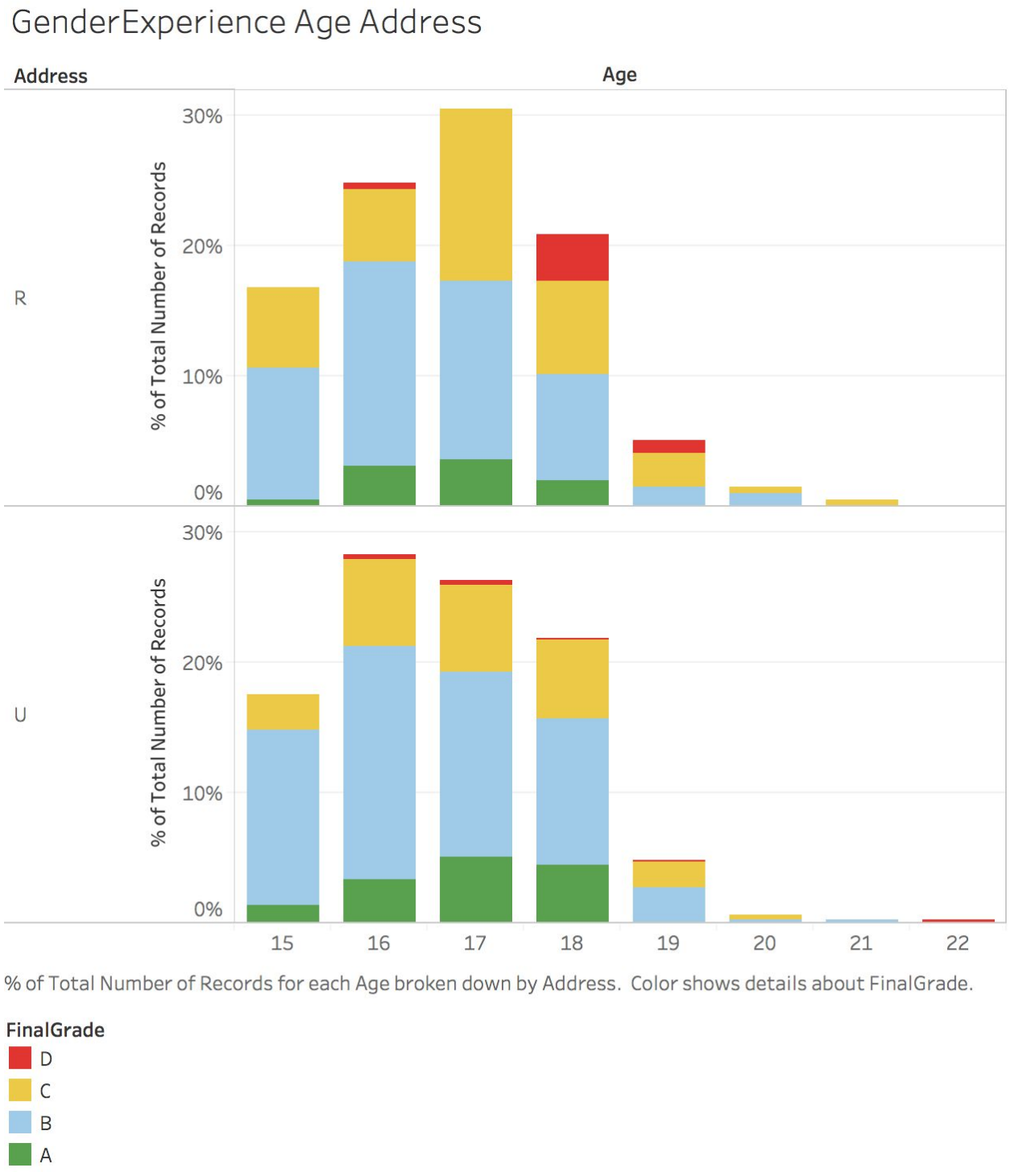
Analysis:
Seeing the above graph we can say that good family relation drives the score for the rural students. But in case of urban, students performance doesn’t have much influence of family relation status.

Some of the analysis which couldn’t make into the story:

How student’s gender and personal experiences influence their performance?

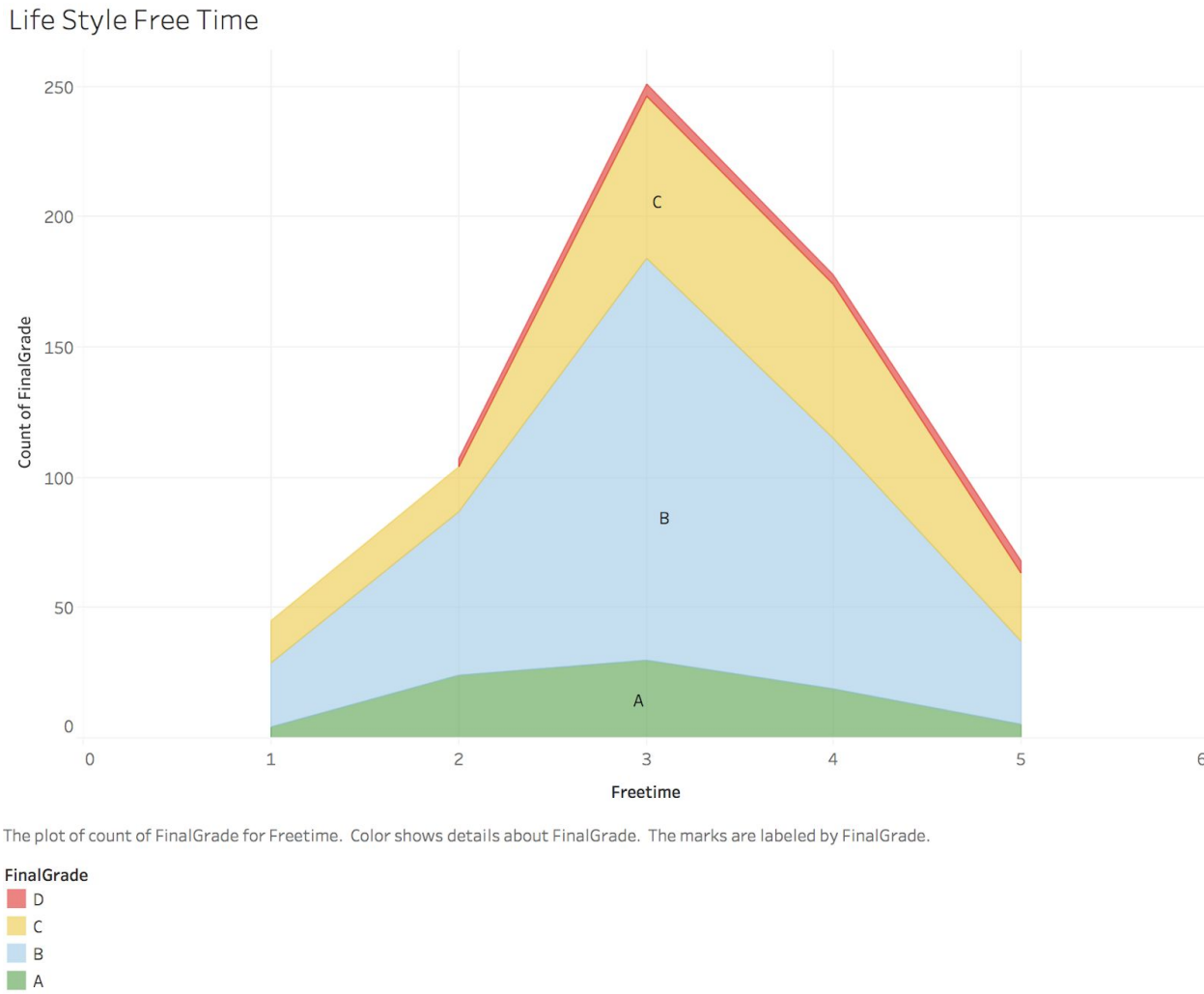
- Does the performance vary for students of the same age if they come from a rural or urban area?
    - address (student's home address type)

- age (age of student)
- Grade



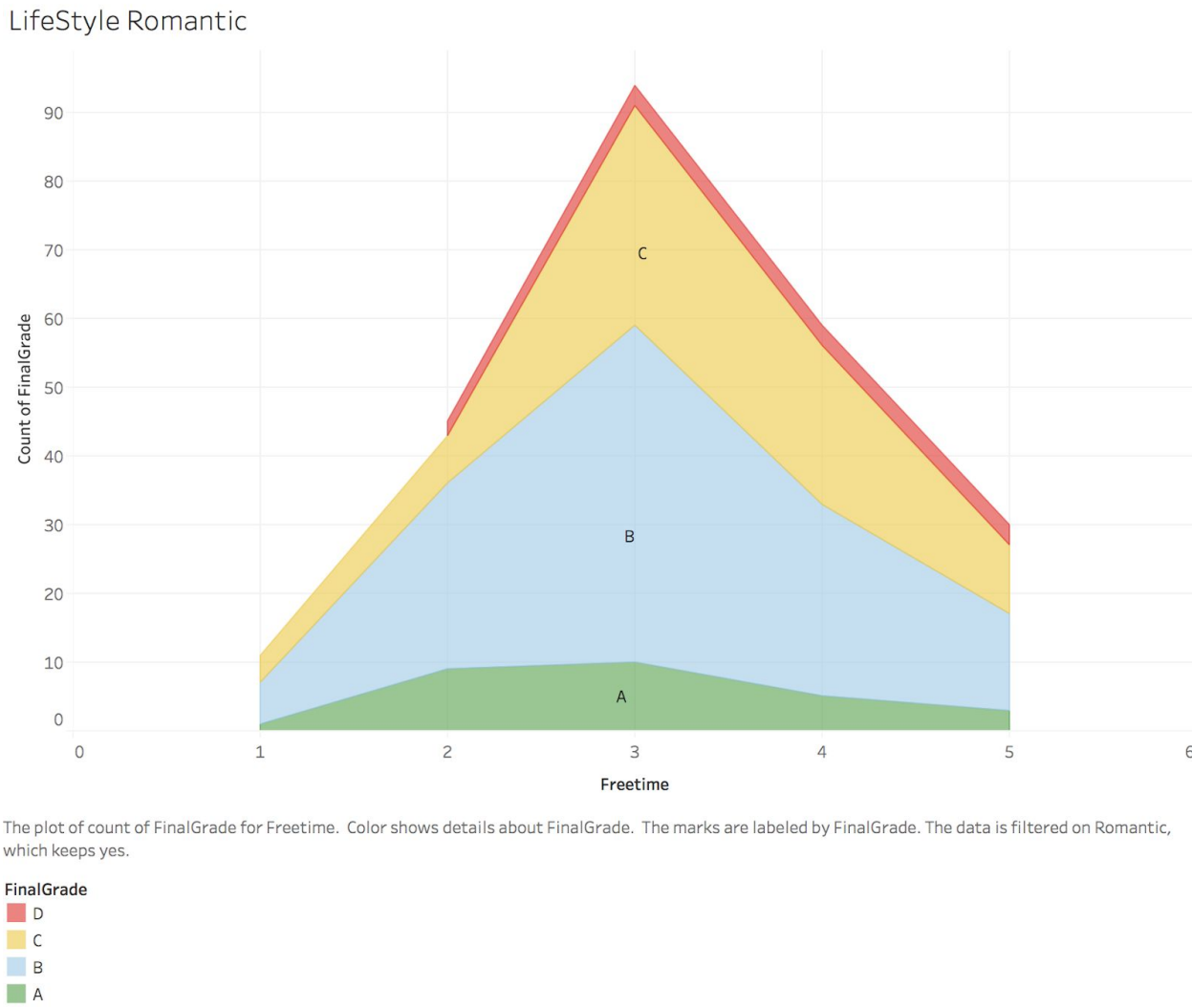
How lifestyle contributes to their performance in school?

- Does having free time after school improve their performance?
    - Freetime (free time after school)



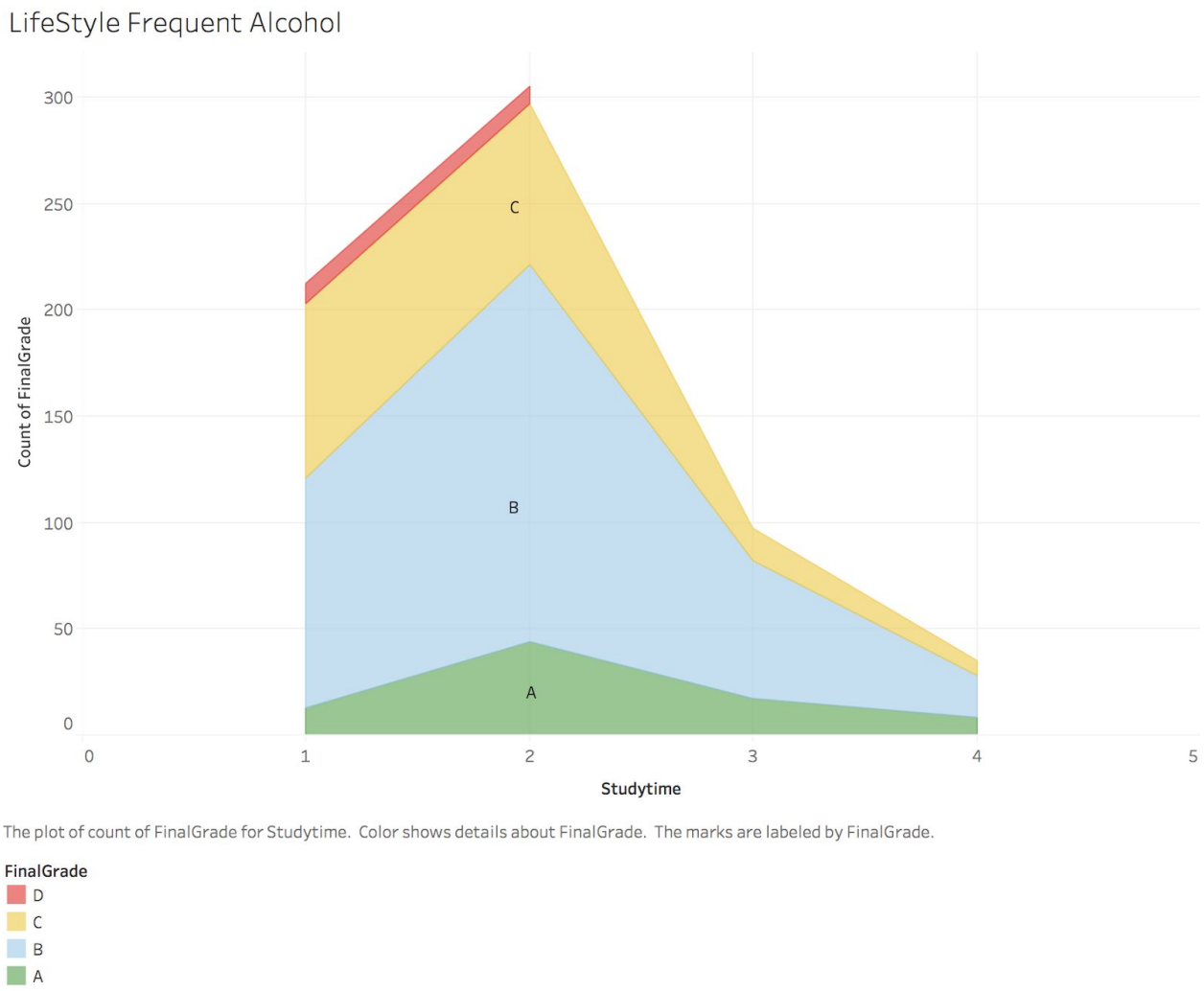
This shows that, even if they have free time after a certain amount, it doesnt contribute positively to their performance.

- Having free time but being in a romantic relationship or outing or alcohol consumption takes a toll on the performance?
    - romantic (with a romantic relationship)
    - goout (going out with friends)
    - dalc (workday alcohol consumption)
    - walc (weekend alcohol consumption)
    - Grade



This shows that having a romantic relationship does not. contribute much when filtered with freetime and shows same kind of trend.

- Even though frequent alcohol consumption, but if they have sufficient study time is the performance affected?
    - studytime (weekly study time)
    - dalc (workday alcohol consumption)
    - walc (weekend alcohol consumption)
    - Grade

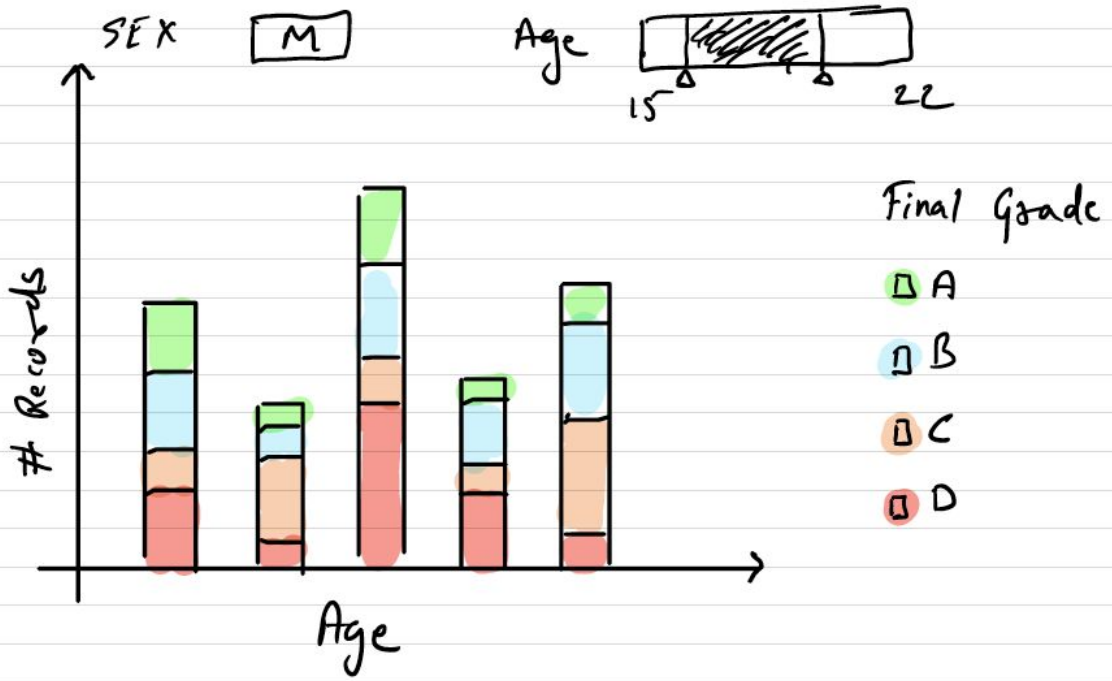


Alcoholic students with 3-4 study time failed to improve their grade though they managed to avoid D grade.

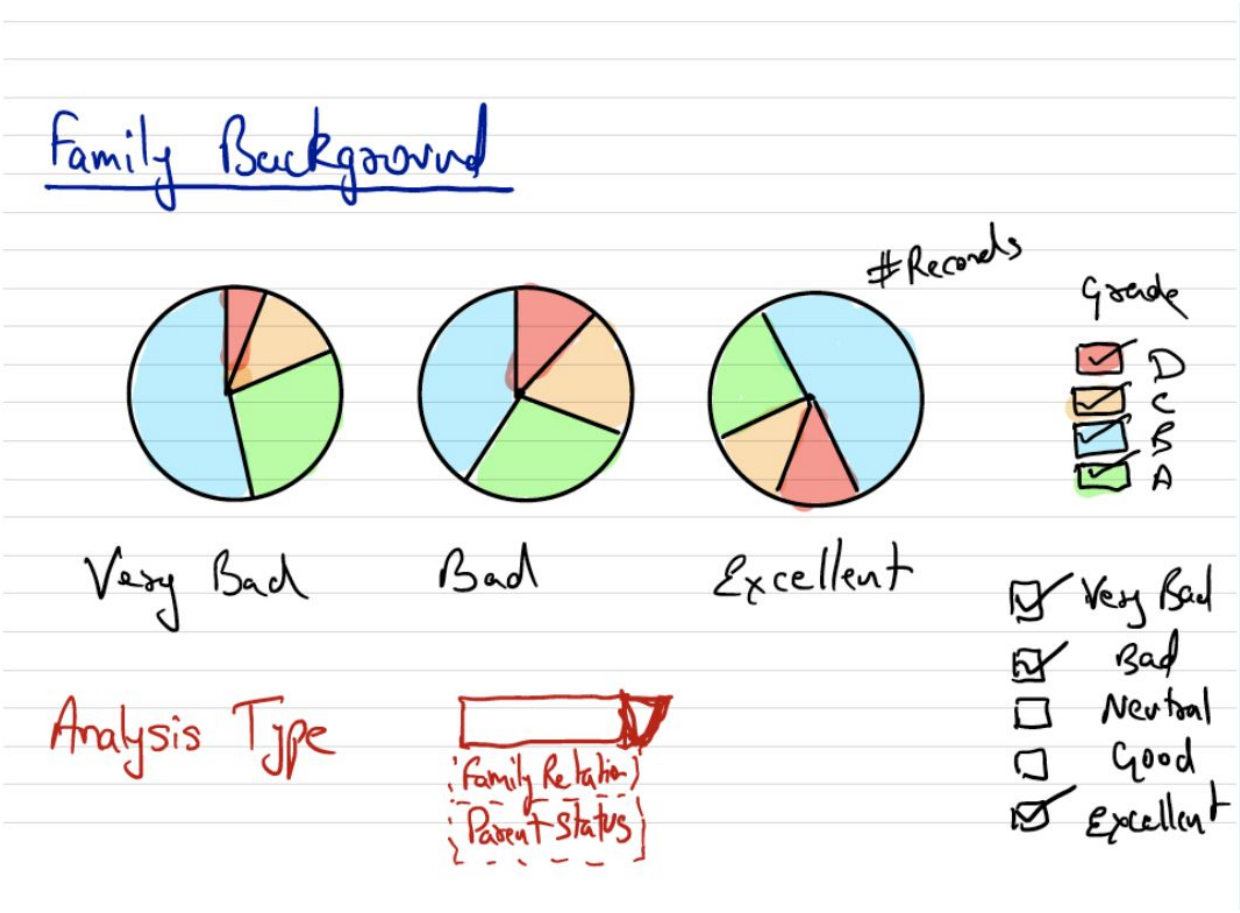
Sketching:

Here are some of the graphs we sketched out while working on our story designing.

Demographics



Student Background



NOTE:

After Data Analysis and sketching and discussing with the professor, we finally decided on the some changes to be made in the storyboard and as suggested just kept into consideration the family aspect background and location factor as it would be more interesting if we dug deeper in one aspect to share a story.

Also, since our initial graphs could not show the distribution properly we changed them in to bar graphs. And used scatter plot to show the parents education factor.

Implementation:

Github access link:
https://github.com/NYU-VIS-FALL2018/storytelling-group-17-students-performance
Demo link:
https://nyu-vis-fall2018.github.io/storytelling-group-17-students-performance/