National Education Longitudinal Study of 1988

INFX 598E: ADV. INFORMATION VISUALIZATION, WINTER 2015

ASSIGNMENT 2: EXPLORATORY DATA ANALYSIS

EDWARD NGUYEN

Initial Question

I chose the National Educational Longitudinal Study of 1988 as my domain and posed the question of whether there is a relationship between student academic performance in the subject of Math and their parents' educational background. The dataset provided by the study is sufficient for answering this question. I began by reviewing the dataset itself, having to refer to the document that explains the survey encodings.

After initially assessing the data, the following survey questions are most applicable to begin exploring my question and the data:

|	
F1S39A	Math Grade
BYS34A	Father's Highest Edu
BYS34B	Mother's Highest Edu

Transforming the data was necessary beforehand as there was missing data or data that didn't help answer my question. I directly imported the dataset in .csv format into Tableau and began plotting right away. Data excludes were handled in Tableau. A summary of the data I excluded is on the next page.

Mother Education Plot

Plotting survey data was tricky at first. Below is the initial plotted chart, and a summary of data I excluded from the chart, done directly in Tableau. I plotted math grade as an ordinal variable on the x-axis, the number of records/students on the y-axis, and color encoded separate lines to denote the mother's highest education level.
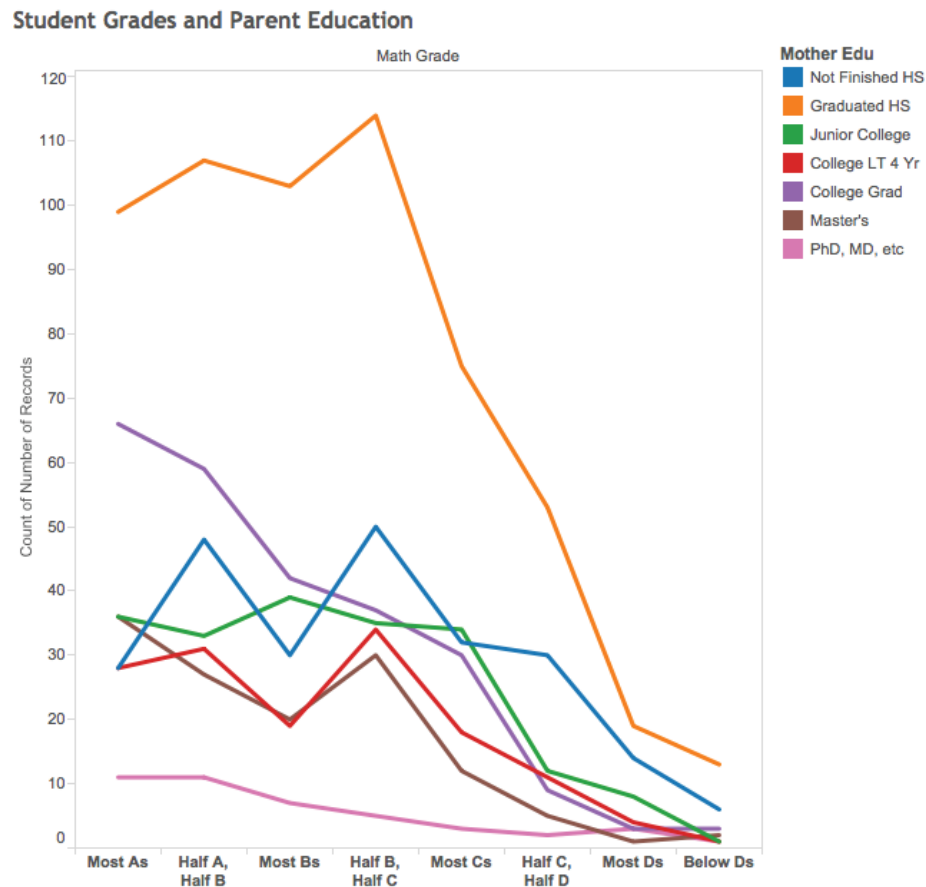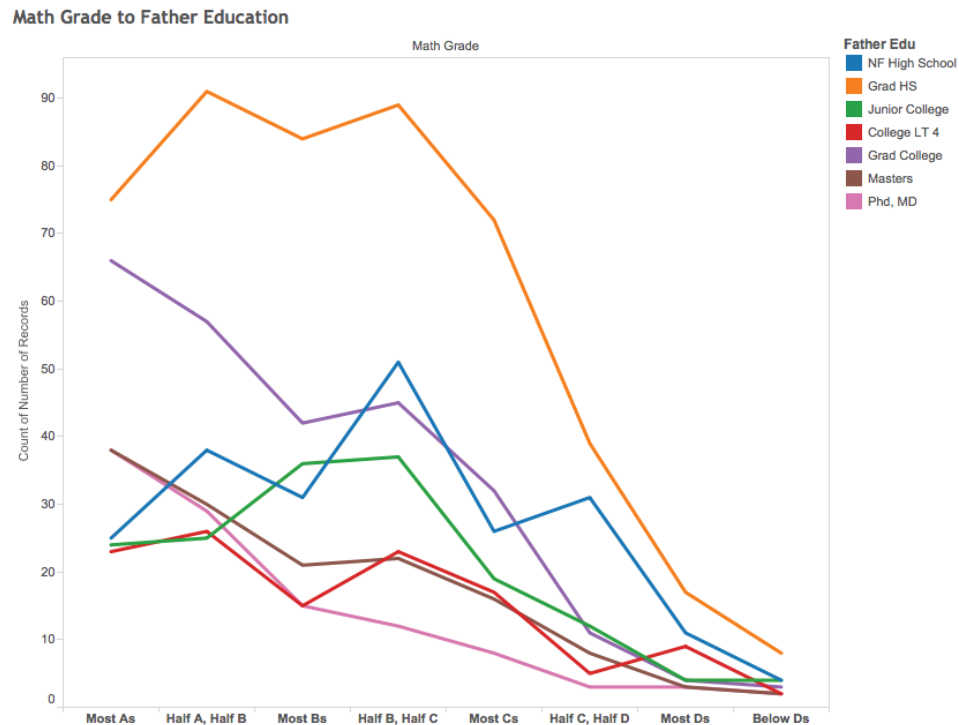

EXCLUDED VARIABLES:

F1S39A

01	Not Taking Subject
10	Subject Not Graded
96	Multiple Response
97	Refusal
98	Missing

BYS34A-B

08	Don't Know
97	Refusal
98	Missing

Father Education Plot

Below is the same plot, but for the father's Highest Education level.

Math Grade to Father Education

PLOTS RAISED MORE QUESTIONS

It seemed odd that most students that did better in Math pretty much came from parents that only had High School as their highest education level. I had a skeptical feeling that it might have been that there was a disproportionate about of parents that had only high school education, and plotted a simple histograms to see the amounts of both fathers and mothers and their educational levels (view on next page).

Parent Education Charts

Below are two histograms mapping mother and father's highest educational level, shown as a percentage of the total. This survey data was reported by the parents themselves
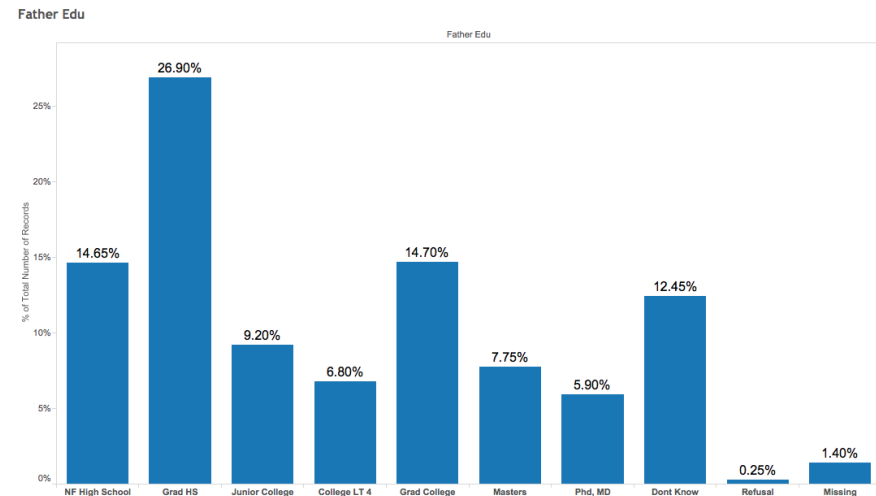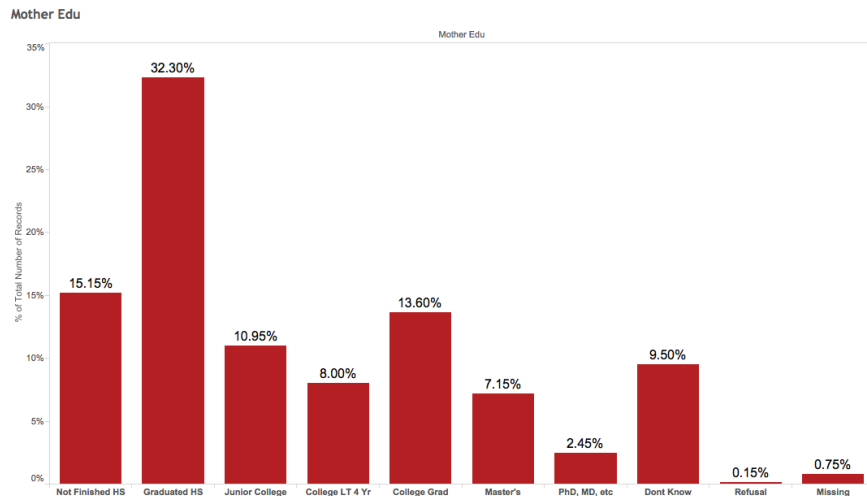

The histograms above confirmed my suspicion, showing 32% of Mother's had only a high school diploma, while father's had just about 27%. With about 1 out of every 3 students having a mother with only a high school education, students with A's had a significant chance of having a mother with this education off probability alone. This is the same expectation with fathers.

My follow-up question was to determine the proportion within each parent's educational level, the type of student they are more likely to have. This involved not comparing the educational levels to each other necessarily, but to view percentages within each educational level to see the breakdown.

Math Grade to Mother Education

The problem with my initial chart was again: the skew in the amount of parents' educational levels (mostly skewed to high school graduates), and showing the sheer total of students achieving each grade didn't help to determine the percentages, although if we isolate each line we can focus more easily on the proportions that way. My chart below was to make visualizing this information much easier.
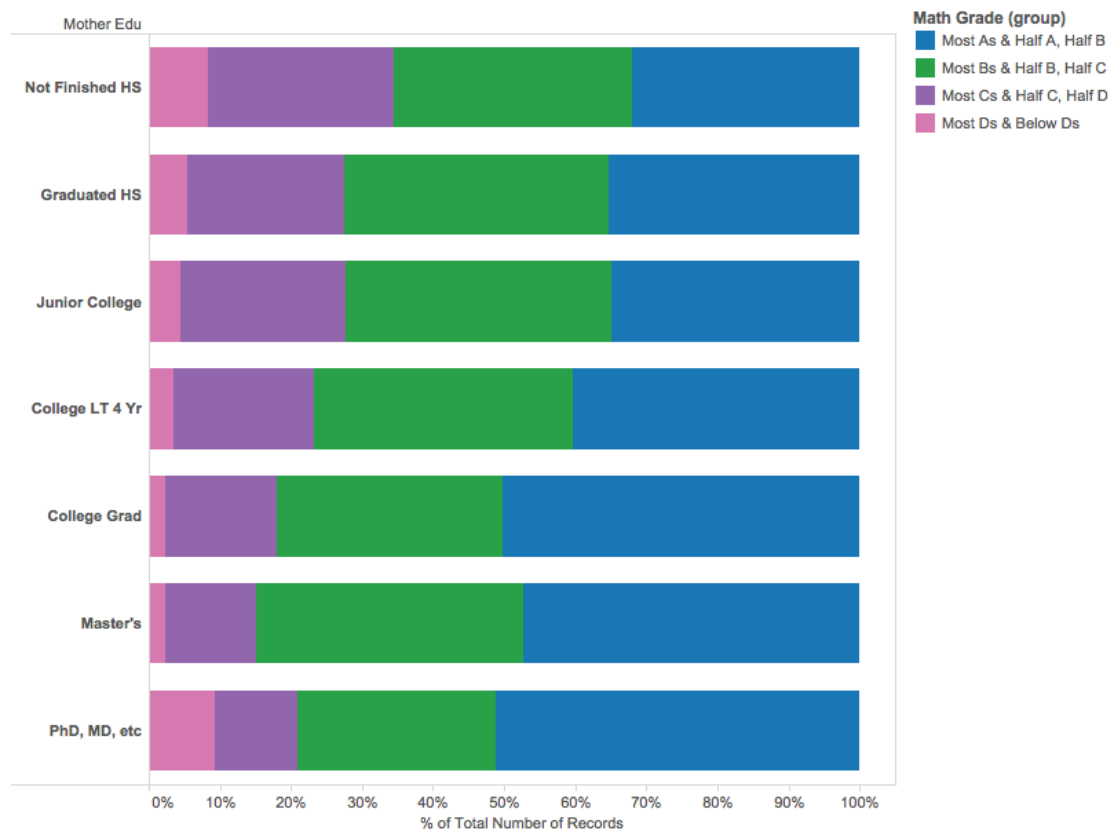
CONSOLIDATING GRADES

A big thing to note is the consolidation of grades - I grouped similar grade types, as eight was a bit much.

ANALYSIS

The answer to my question is much more clear in this visualization - is there a relationship between parent educaitonal level and success in Math? From here, we clearly see that Mother's with post-doctoral, Master's, and College Grads, have students that are mostly A or half A and half B students. Master's and College grad mothers also have almost zero students that have D's or lower.

Math Grade to Father Education

The problem with my initial chart was again: the skew in the amount of parents' educational levels (mostly skewed to high school graduates), and showing the sheer total of students achieving each grade didn't help to determine the percentages, although if we isolate each line we can focus more easily on the proportions that way. My chart below was to make visualizing this information much easier.
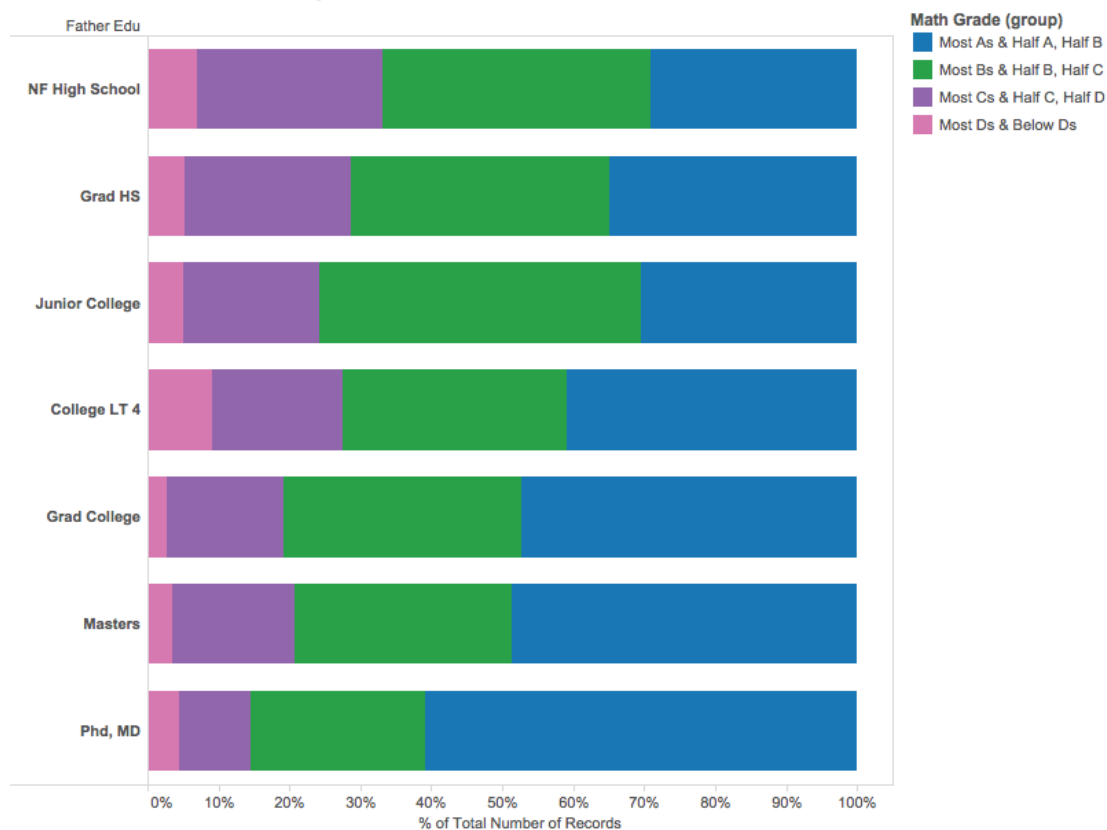
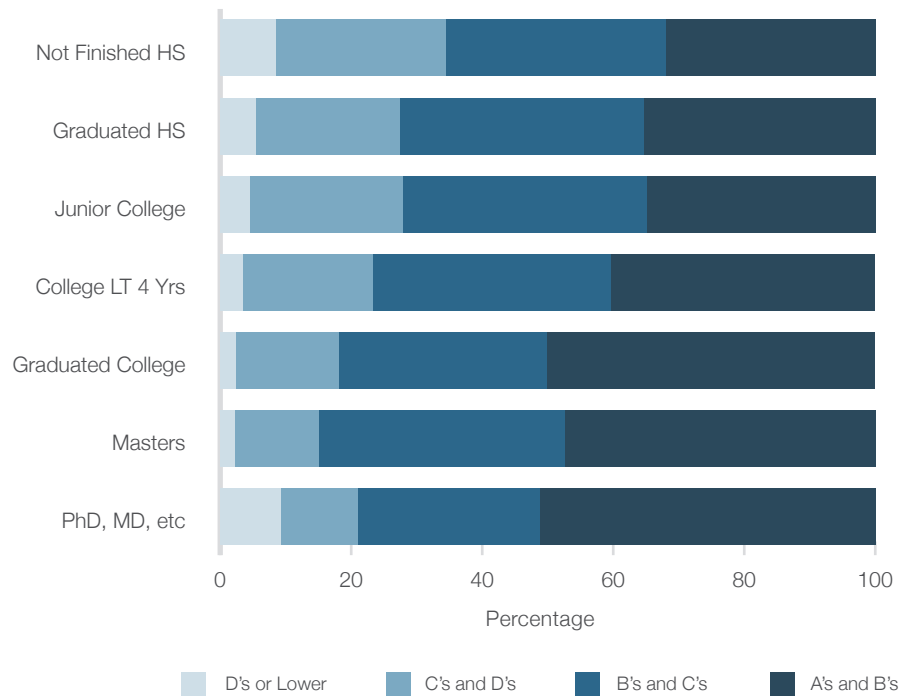## RATIONALE FOR THIS VISUALIZATION

We can see rows clearly, showing the Father's educational level. I used a stacked bar chart to visualize the percentages that each educational level produced in terms of student grade output. This makes it easy to focus on one educational level if interested, and also compare to other data on the chart.

ANALYSIS

Contrary to the initial plot, fathers who only graduated high school have only 30% students that are high performing, even though the number is high. They seem to be mostly B or C students. We see a trend that the higher the father's eduactional level, the higher the percentage the student has A's and B's, or higher.

Final Visualizations + Visual Design

MOTHER HIGHEST EDUCATION LEVEL to STUDENT MATH GRADES

FATHER HIGHEST EDUCATION LEVEL to STUDENT MATH GRADES