Introduction

This study aims to examine the academic performance of kindergarten students across different income groups over the academic year 1998-99. Specifically, we'll analyze the changes in reading and math scores, using general knowledge scores as a baseline, to understand how socio-economic factors might influence educational outcomes in early childhood.

Research Questions:

1. Does income group influence the changes in reading and math scores over time, controlling for baseline general knowledge?
2. How does the interaction between income group and initial general knowledge scores influence the changes in reading and math scores from fall to spring?

When conducting ANOCOVA tests, several assumptions are made regarding the data:
Independence: Each student's data is separate and doesn't affect anyone else's
Linearity: The relationship between baseline knowledge and score changes is straight-line
Homogeneity of Regression Slopes: The effect of baseline knowledge on score changes is consistent across income groups
Normality: The differences between the observed and expected score changes form a bell-shaped curve
Homogeneity of Variances: The spread of score changes is similar across all income groups
No Multicollinearity: The independent variables and their interaction don't overly overlap in what they explain
Additivity: The interaction effect straightforwardly combines with the main effects

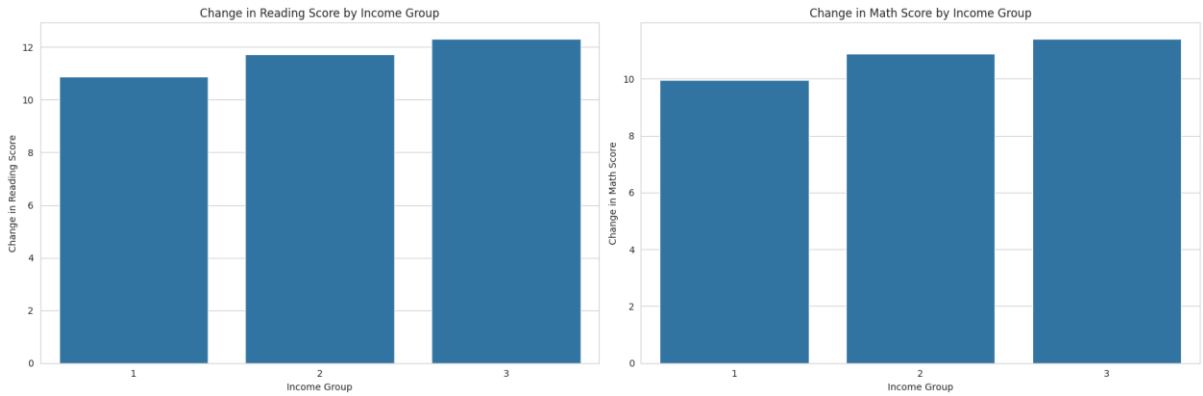
Data Cleaning

The raw dataset has a total of **9** columns with **11933** rows. Below we outlined our data and its' dictionary of what each column means.

fallreadingscore	Student's reading score of the fall term
fallmathscore	Student's math score of the fall term
fallgeneralknowledgescore	Student's general knowledge score of fall term
springreadingscore	Student's reading score of the spring term
springmathscore	Student's math score of the spring term
springgeneralknowledgescore	Student's general knowledge score of spring term
totalhouseholdincome	The total household income
incomeinthousands	The total household income (thousands)
incomegroup	Income group category derived from household income

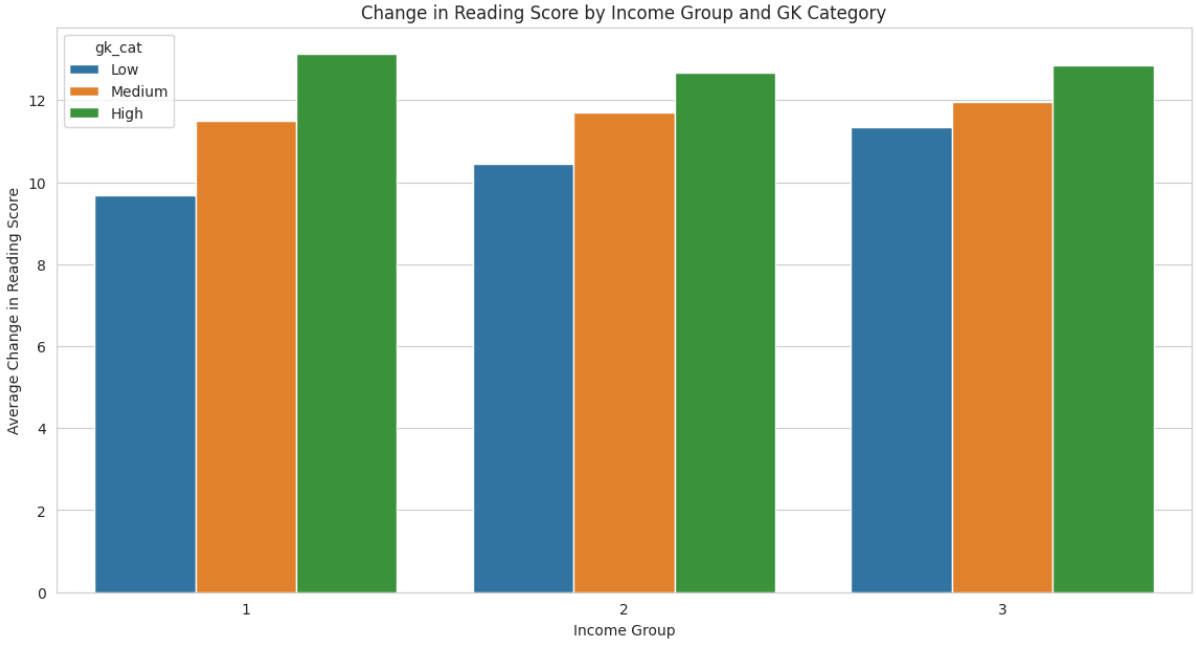
Exploratory Data Analysis

We now create plots to visually assess the interaction between income group and the changes in scores over time.



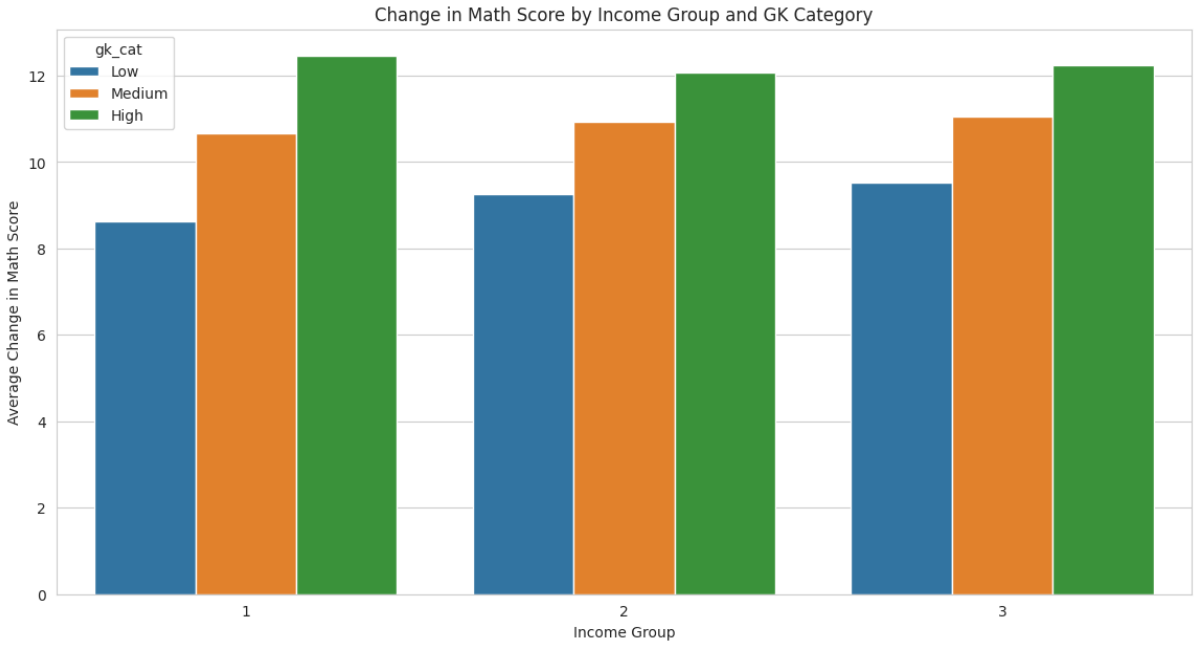
The bar plots indicate the average changes in reading and math scores across three income groups. From the visual data, we can observe that all income groups show an increase in both reading and math scores. However, the plots do not show a stark difference in score changes between income groups, suggesting that income may not be a strong differentiator in how reading and math scores changed over this period.

Therefore, we may interested in if there exists any interaction between income group, initial general knowledge (GK) levels, and the changes in reading and math scores. We categorize initial general knowledge scores into 3 categories, and then calculate average changes in reading and math scores by income group and general knowledge.



This plot shows that students from all income groups demonstrate an increase in reading scores, and this increase appears to be influenced by their initial general knowledge category. For example, those in the 'High' GK category show a larger increase in reading scores across all income groups compared

to those in 'Low' and 'Medium' categories.



Similar to the reading scores, this plot also indicates that students with higher initial general knowledge tend to have a larger increase in math scores, regardless of income group.

By comparing the changes in scores across both income and general knowledge categories, we can infer that initial general knowledge might play a role in how much a student's reading and math scores improve over time. These plots suggest that general knowledge could be a significant factor in academic progress, potentially even more so than income group. However, in order to quantify the research question and to determine if the observed patterns are statistically significant, we need to conduct the ANOCOVA analysis.

ANOCOVA

Let's restate our research questions here: Does income group influence the changes in reading and math scores over time, controlling for baseline general knowledge?

We use the income group as the categorical variable, and use the fall general knowledge as the baseline. After applying the ANOCOVA analysis, we get the following results for the change in reading score:

	sum_sq	df	F	PR(>F)
C(incomegroup)	287.485906	2.0	2.251247	1.053126e-01
fallgeneralknowledgescore	14054.124684	1.0	220.110317	2.354473e-49
Residual	761671.036393	11929.0		

As we can see from the table, the income group factor has a p-value of 0.105, indicating that there is not a statistically significant difference in the change in reading scores among the different income groups when controlling for the baseline general knowledge score. Also, the

baseline general knowledge score (fall 1998) significantly influences the change in reading scores, with a p-value close to 0, suggesting a strong relationship.

Similarly, we get the following results for the change in math score:

	sum_sq	df	F	PR(>F)
C(incomegroup)	55.879616	2.0	0.624286	5.356614e-01
fallgeneralknowledgescore	22425.932956	1.0	501.083959	9.42525e-109
Residual	533880.499781	11929.0		

As we can see from the table, the income group factor has a p-value of 0.536 for math scores, suggesting that income group does not significantly influence the change in math scores when controlling for the baseline general knowledge. Also, the baseline general knowledge score significantly affects the change in math scores as well, with a p-value very close to 0, indicating a substantial effect.

For the research question "How does the interaction between income group and initial general knowledge scores influence the changes in reading and math scores from fall to spring?", we want to conduct ANCOVA tests that include an interaction term between income group and initial general knowledge scores. This analysis will help us understand if the impact of income group on the changes in reading and math scores is moderated by the students' initial general knowledge levels.

ANCOVA for Reading Scores with Interaction Term:

	sum_sq	df	F	PR(>F)
C(incomegroup)	287.485906	2	2.255280	1.05e-01
fallgeneralknowledgescore	14054.124684	1	220.50466	1.94e-49
C(incomegroup):fallgeneralknowledgescore	1489.617873	2	11.685811	8.509e-06
Residual	760181.418519	11927		

The interaction between income group and fall general knowledge is statistically significant (p = 8.50986e-06), with an F-statistic of approximately 11.69. This suggests that the effect of income group on the change in reading scores is different at different levels of initial general knowledge.

ANCOVA for Math Scores with Interaction Term:

	sum_sq	df	F	PR(>F)
C(incomegroup)	55.879616	2	0.625281	5.351e-01
fallgeneralknowledgescore	22425.932956	1	501.88277	6.42e-109
C(incomegroup):fallgeneralknowledgescore	939.110149	2	10.508441	2.756e-05
Residual	532941.389632	11927		

Similar to reading scores, there is a significant interaction effect (p = 2.75586e-05) with an F-statistic of about 10.50. This indicates that the initial general knowledge level modifies the impact of income group on the change in math scores from fall to spring.

For both reading and math scores, the initial general knowledge of the students has a moderating effect on how income group influences their score changes. In practical terms, this means that the relationship between socio-economic status and educational outcomes is more complex than simply comparing across income groups; it's also affected by where students start in terms of their general knowledge.

Conclusion

Now, given the significant interaction, educational interventions may need to be tailored not only to students' socio-economic backgrounds but also to their starting levels of general knowledge to be most effective.

These findings should be used to guide educational policy and teaching strategies, acknowledging that the 'one-size-fits-all' approach may not be suitable for educational improvements across diverse socio-economic and knowledge-based student groups.