Hyperiondev

TASK

Exploratory Data Analysis on the '*Student Performance in Exams Report*' Dataset

Visit our website

Introduction

Summary of the data set

The '*Student Performance in Exams Report'* dataset is trying to investigate what factors influenced a group of American students' math, reading and writing scores.

This dataset can be located on Kaggle at [Students Performance in Exams | Kaggle]

Dataset approved by mentor Rudolph.

DATA CLEANING

SUMMARY OF THE METHODS AND VISUALISATIONS DONE DURING DATA CLEANING

- Columns:
    - *gender*: students' gender
    - *race/ethnicity*: student's race/ethnicity
    - *parental level of education*: student's parents' education level
    - *lunch*: students' lunch status. (Free/reduced lunch status implies that a student comes from a lower income household)
    - *test preparation course*: students' who completed a test preparation course (refers to a student's access to resources)
    - *math score*: students' math score
    - *reading score*: students' reading score
    - *writing score*: students' writing score

- It is not necessary to remove any of the columns from the database as they are all relevant to the data story.
- Dropped any duplicate rows.
- Checked data types.
- Visualised all the unique values in each column with the data type 'object' to ensure there were no variations of the same value.

MISSING DATA

ANY MISSING DATA? HOW DID YOU HANDLE IT

- Visualised the number of missing values.
- There is 0 missing data.

DATA STORIES AND VISUALISATIONS

THIS IS THE BULK OF THIS PROJECT. EXTRACT STORIES AND ASSUMPTIONS BASED ON VISUALISATIONS OF THE DATA

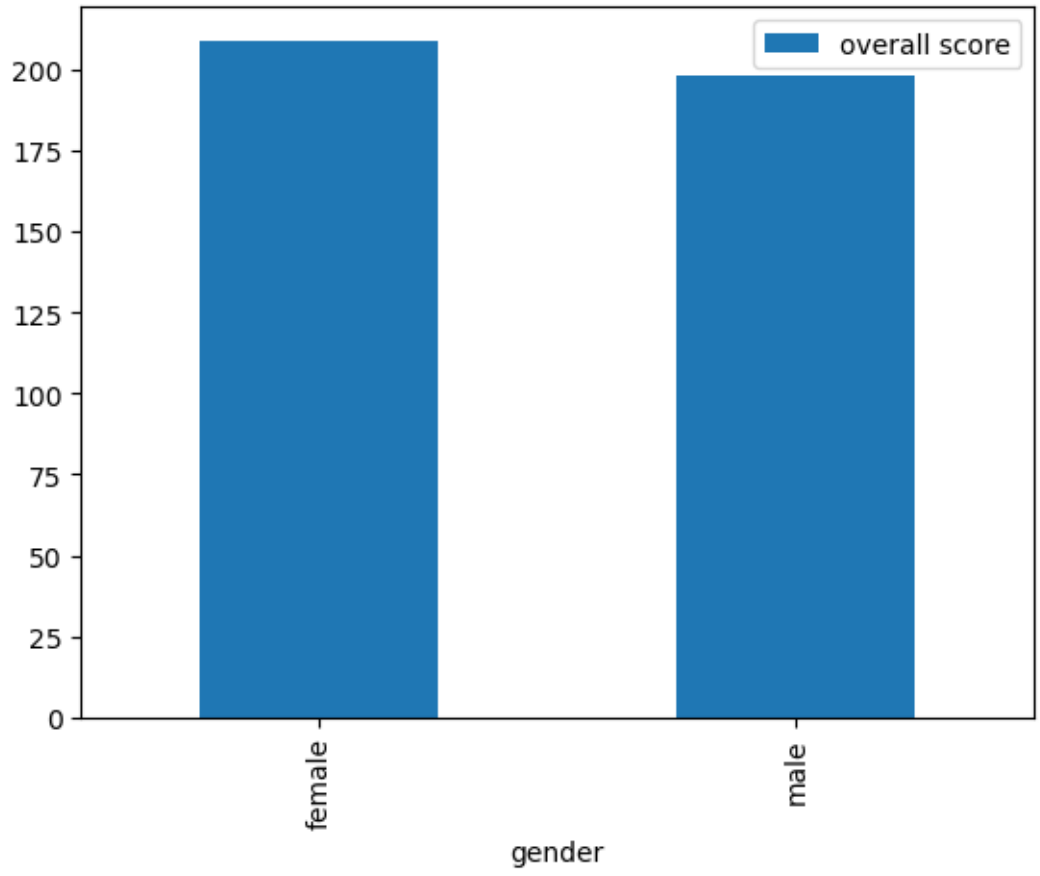
1. New Column

Create a new column called *overall score* which adds together every student's math, reading, and writing score.

2. Gender

Visualised the average math, reading, writing and overall score of male and female students. By comparing male and female students it can be shown that female students on average performed better than male students in reading and writing but male students scored higher in math.

A bar chart visualising the average overall score of male and female students shows that female students have a higher average score than male students.



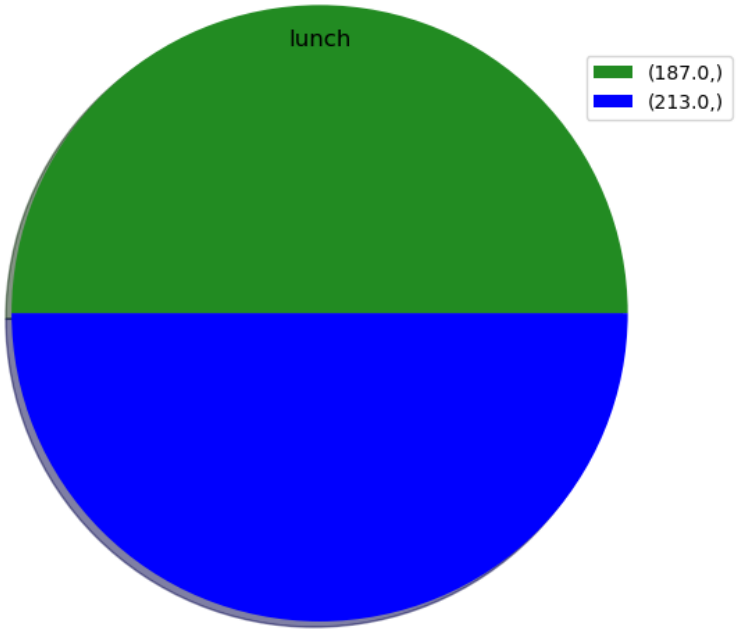
Gender will not be considered as an influencing factor on student's score as the difference of the average overall score is so small.

3. Lunch

Compared the average overall score of students who received a free/reduced lunch to students who received a standard lunch.

The average overall score of students on a free/reduced lunch is 183. The average overall score of students on a standard lunch is 213.

A pie chart indicates that lunch may be defining factor which influences the score of students.

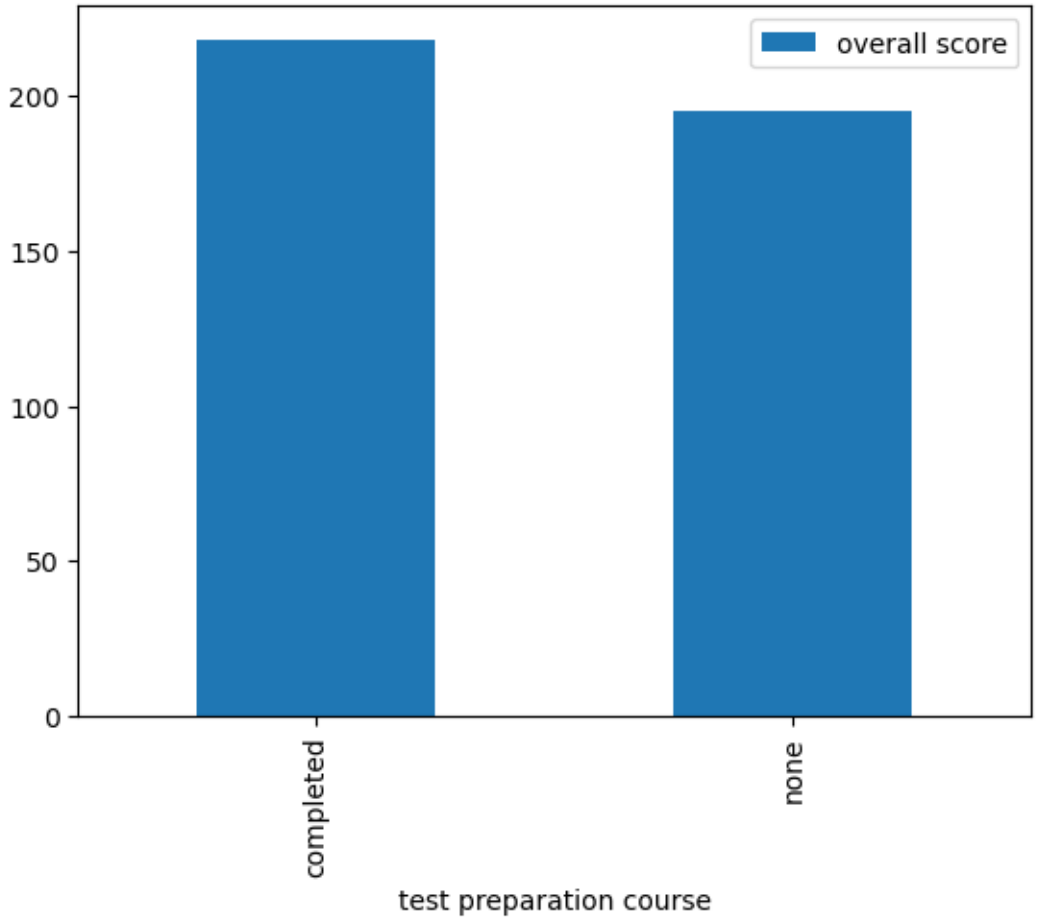


4. Test Preparation Course

Compared the average overall score of students who completed the test preparation course to students who did not complete the course.

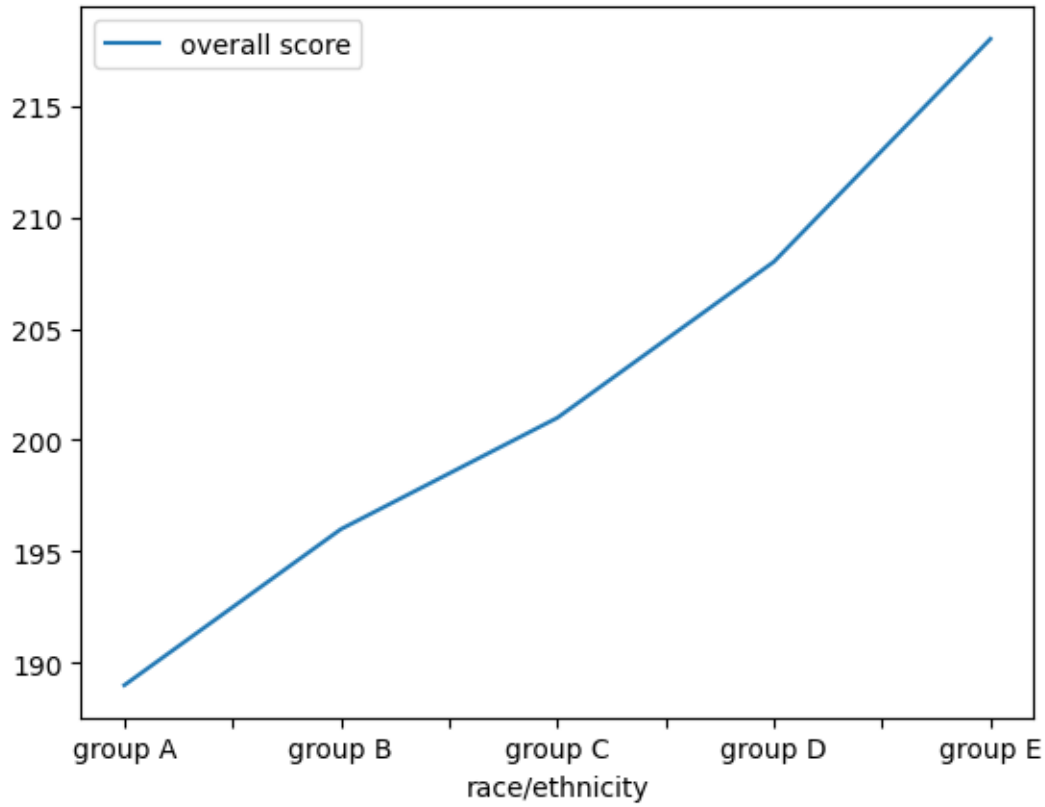
Students who completed the course scored higher on average than students who did not complete the course.

A bar chart shows that by completing the test preparation course a student could achieve a higher score.


overall score
200
150
100
50
0
completed
none
test preparation course

5. Race/Ethnicity

The average overall score of each ethnicity group is presented in a line graph.



The line graph shows that Group A had the lowest average overall score, and group E had the highest average overall score.

Going forward, this report will compare group A and group E to decipher which factor influenced a student's score the most.

6. Group A

Group A has the lowest average overall score of 189.

There are 89 students in group A.

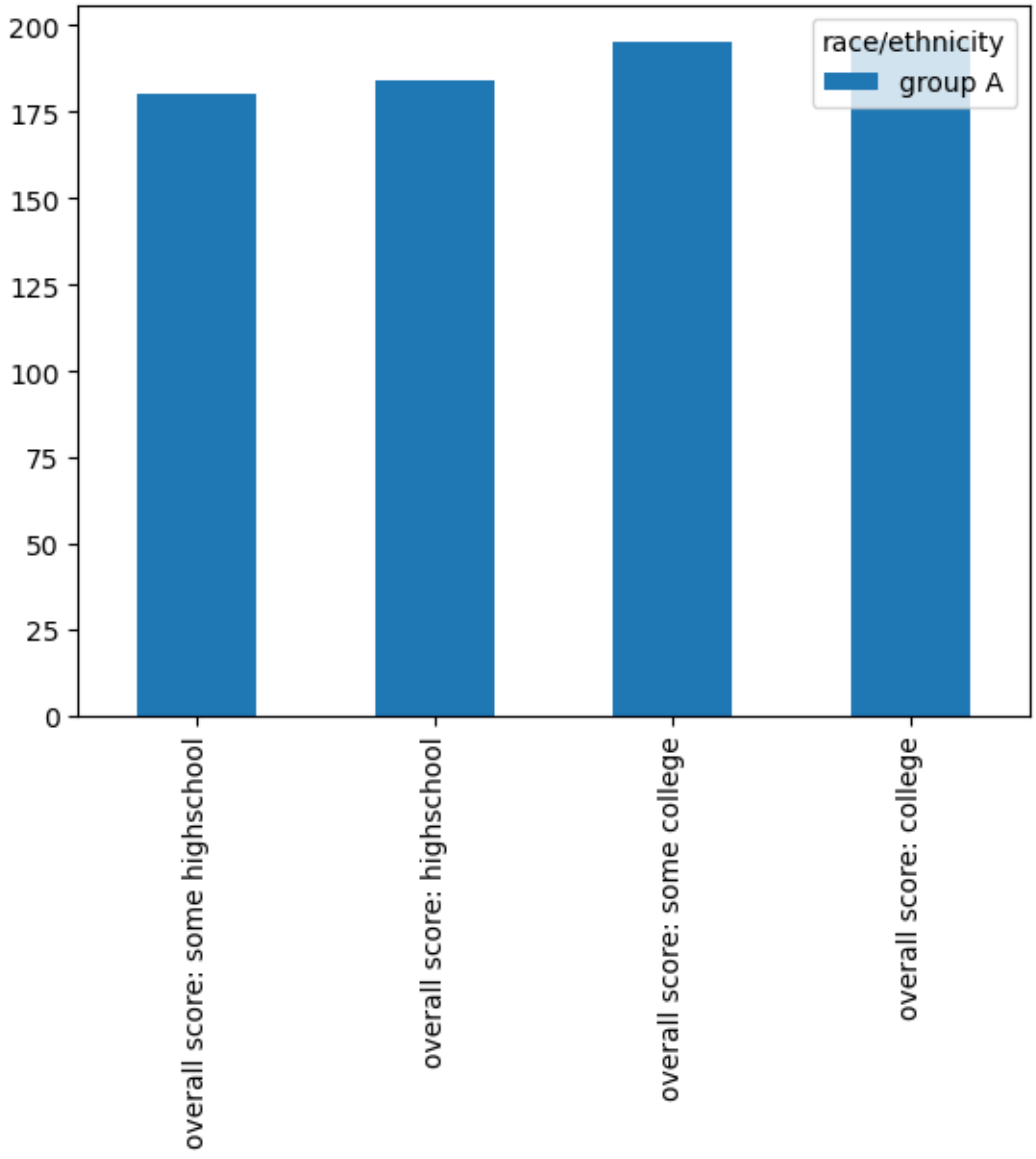
parental level of education

Within group A:

- 18 student's parental level of education is high school.
    - Their average score is 184.
- 24 student's parental level of education is some high school.
    - Their average score is 180.

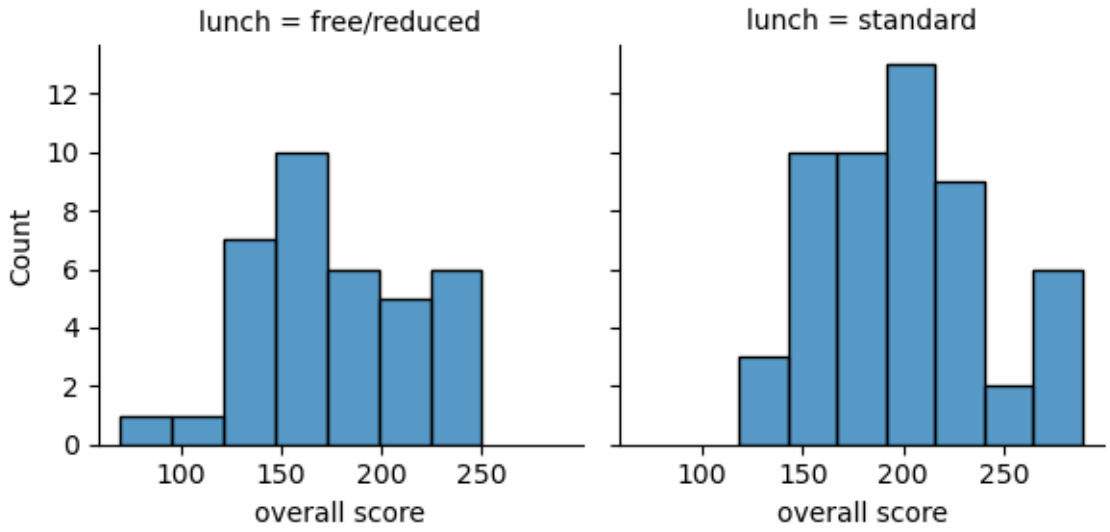
- 18 student's parental level of education is some college.
    - Their average score is 195.
- 29 student's parental level of education is college (have either an associate's, bachelor's, or master's degree).
    - Their average score is 196.

Bar chart shows that within group A, the higher the parents' education the higher the student's overall score.



lunch

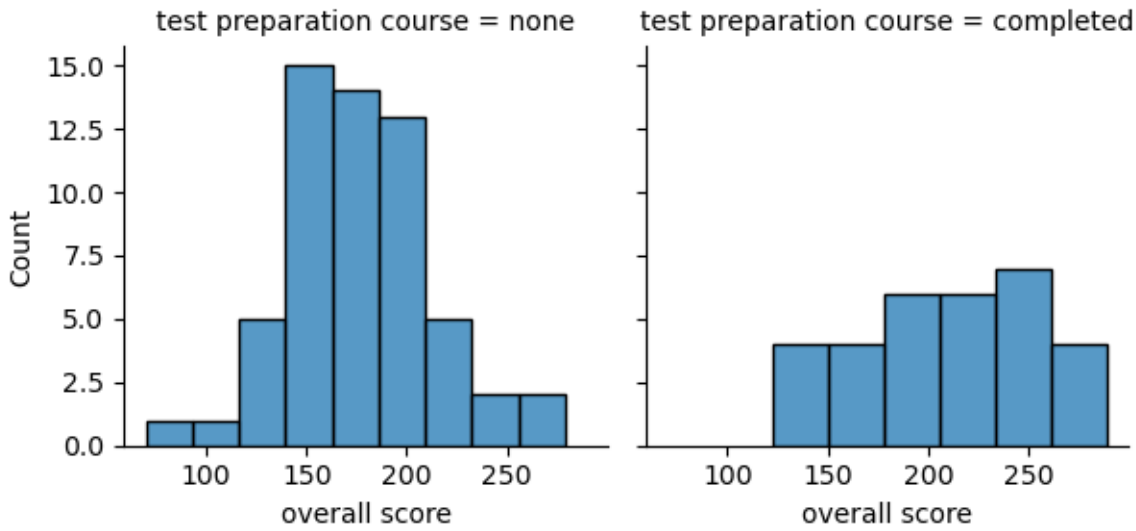
A multi-plot grid comparing students lunch status shows that on average students with a standard lunch scored higher overall than a student with a free/reduced lunch.



test preparation course

A lot more students in group A did not complete the course than students who did.

A multi-plot grid shows that students were more likely to score over 250 if they completed the test preparation course and that students were more likely to score under 150 if they did not complete the course. However, for the scores between 150 and 250 , completing the test preparation course does not seem to have made a significant difference.



Observations

Comparing the overall score of category A (a student from group A whose parental

educational level is some high school, has a reduced/free lunch and did not complete the test preparation course) to category B (a student from group A whose parental educational level is college, has a standard lunch and did complete the test preparation course) shows:

- That 4 out of the 10 students scored below 150 in category A. A further 2 scored below 170.
- Only 3 students scored above 180 in category A.
- All 6 students in category B scored above 180.

7. Group E

Group E has the highest average overall score of 218.

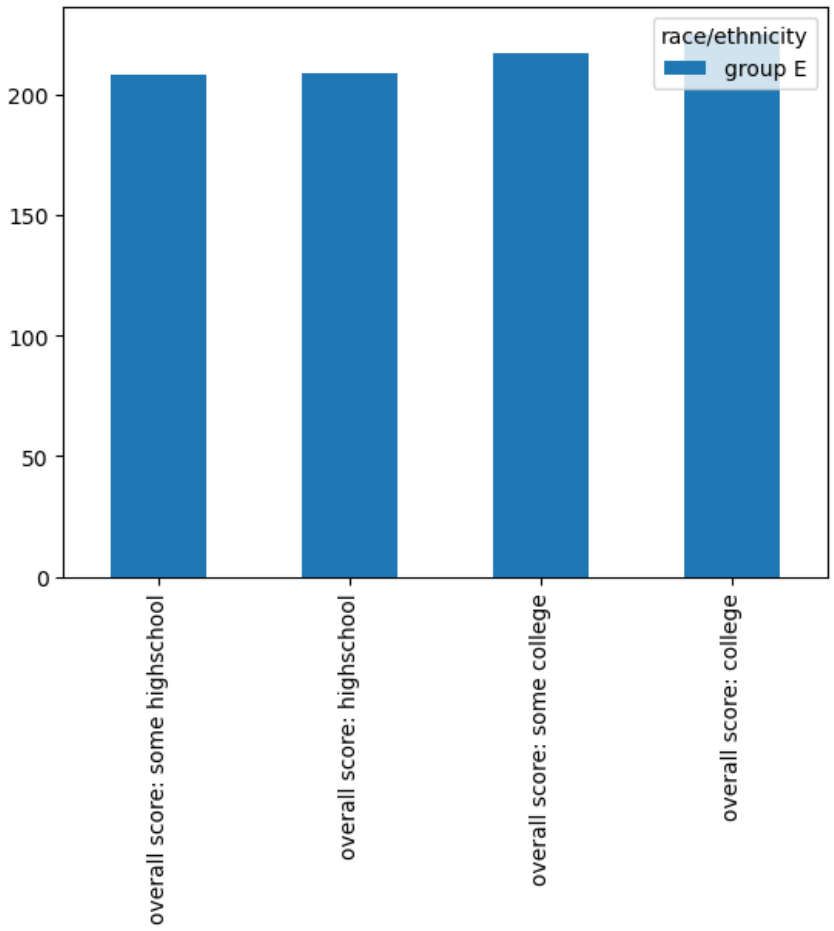
There are 140 students in group E.

parental level of education

Within group E:

- 22 student's parental level of education is high school.
    - Their average score is 209.
- 18 student's parental level of education is some high school.
    - Their average score is 208.
- 35 student's parental level of education is some college.
    - Their average score is 217.
- 65 student's parental level of education is college (have either an associate's, bachelor's, or master's degree).
    - Their average score is 225.

Bar chart shows that within group E, the higher the parents' education the higher the student's overall score.

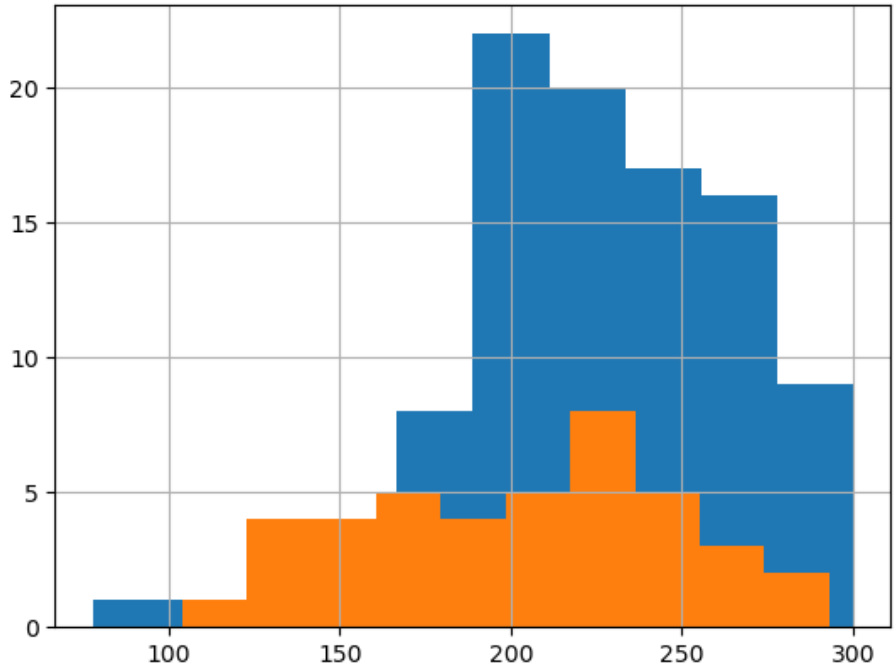


lunch

A histogram shows that most students in group E received a standard lunch.

By comparing students' lunch status, it can be shown that students with a standard lunch scored higher than students with a free/reduced lunch.

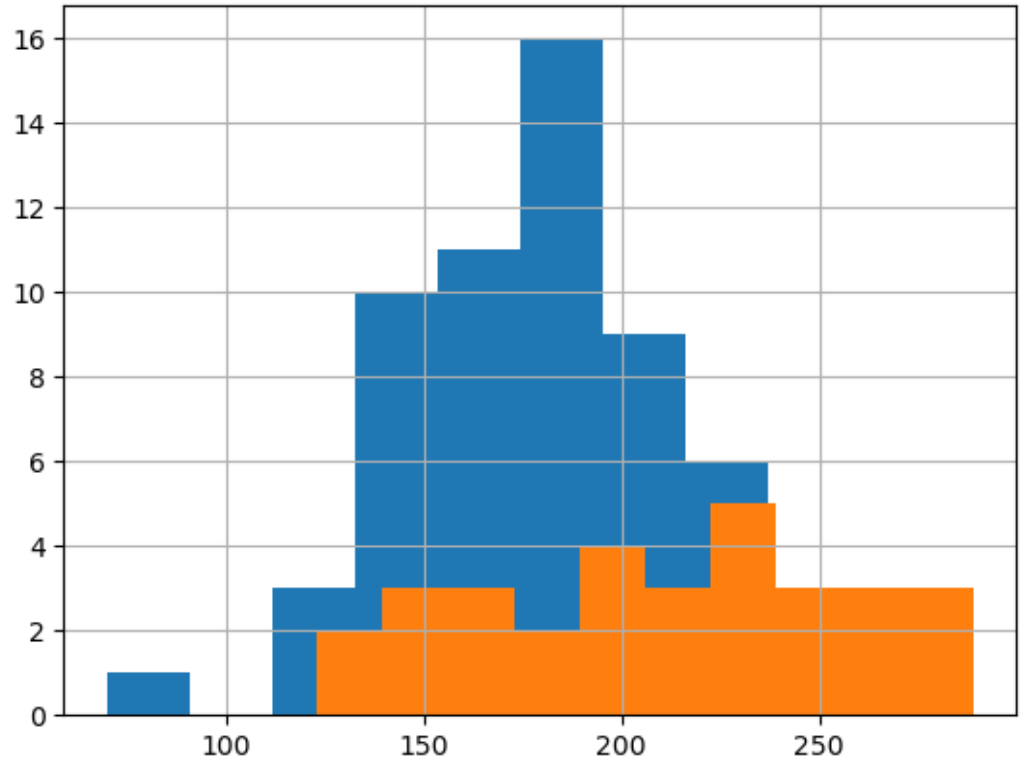
Unlike group A, there is only a slight disparity.



Test preparation course

A lot more students in group E did not complete the course than students who did.

A histogram shows that if you didn't complete the test preparation course you were more likely to score below 150 overall and if you did complete the course, you were more likely to score above 250.



Again, the middle scores do not seem to be affected by whether or not you completed the test preparation course.

Observations

Comparing the overall score of category C (a student from group E whose parental educational level is some high school, has a reduced/free lunch and did not complete the test preparation course) to category D (a student from group E whose parental educational level is college, has a standard lunch and did complete the test preparation course) shows:

- There are only four students who fall into category C.
- 2 of those students scored above 200 and the other 2 scored below 150.
- There are 21 students who fall into category D.
- Only 3 score below 200. The other students scored above 200.
- 1 student in category D scored 300. (100%)

8. Assumptions

The most influencing factor of a student's score in order:

1. Race/Ethnicity
2. Parental education level
3. Lunch
4. Test preparation course

THIS REPORT WAS WRITTEN BY: RACHEL BIRRELL