# Group contract

**Group number:** L02G06

**Name:** Zichun Han

**GitHub link:** https://github.sydney.edu.au/ypan2680/L02G06.git

I agree to:

\* Abide by the terms of this contract in relation to the group assessment for DATA2002/2902.

\* Store all my written and code contributions to the assessment in the GitHub repository.

\* Keep a record of my other contributions to the assessment (e.g., discussions, emails, meetings attended). A copy of this may be requested by the coordinator.

\* All members will collaborate effectively and contribute fairly to group tasks, projects, and other learning activities, as outlined in Section 8(f) of the Student Charter.

\* The main communication platforms will be \*\*WhatsApp\*\* for informal discussions and \*\*GitHub\*\* for task updates.

\* Members should check \*\*WhatsApp\*\* daily and \*\*GitHub\*\* at least every two days, or whenever notified that an update has been pushed.

\* Additional discussions will take place during lab sessions, with extra meetings via Zoom scheduled as needed throughout the week.

I understand that:

\* My agreement to these terms is indicated through the act of submitting this document in Canvas.

\* If I fail to meet my obligations as detailed in this group contract, I have failed to meet the assessment requirements for DATA2002/2902 and may be awarded a mark of zero for some or all of the project components.

# Exploratory data analysis

**Data set:** Student Performance

**Dependent variable:** G3

图片包含 图形用户界面

描述已自动生成**Looking at the pattern of missing data:**

There is no missing data.

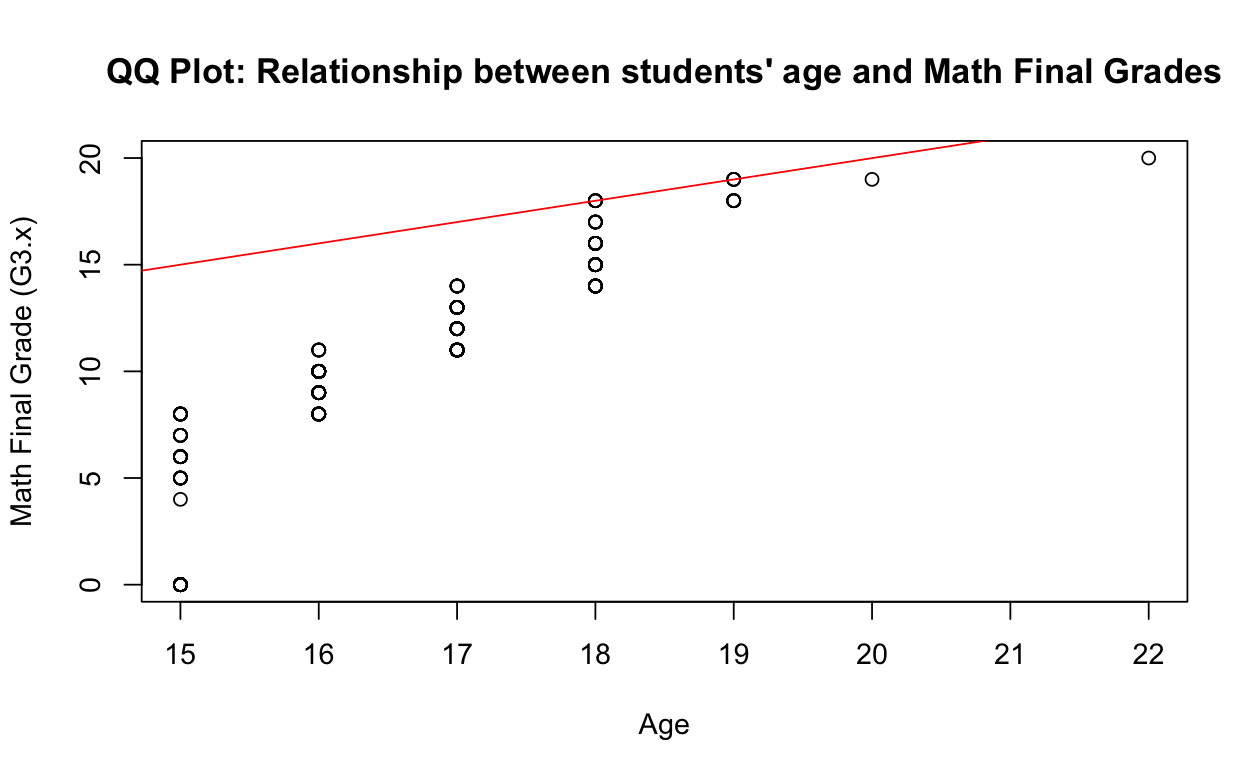
**Distribution of Final Grades (G3)**

图表, 直方图

描述已自动生成

**Description** This histogram shows the overall distribution of Math Final Grade, and we can accurately see that there were more zeros, which may indicate that some students did not get the G3 grade. And the scores are mostly concentrated in the medium level, most students’ scores are between 10 and 15 points. The number of students with higher grades (15 points or more) decreased significantly, with only a small number of students achieving a score close to 20 points. In general, the distribution of scores is not uniform, more students have low scores, and zero scores account for a large proportion, and their scores are concentrated around 10 points.

**To explore the relationship between final grades (G3) and age:**



**Description** This QQ plot shows the relationship between math final grade (G3.x) and age. Most of the data points in the graph are concentrated in younger age groups (15 to 17 years old), where math achievement is more spread out. As we get older, the data points become sparser. As can be seen from the red lines in the figure, there is no clear linear relationship between math performance and age. Although there was a slight upward trend in performance with age, the overall distribution did not strictly follow the trend line. Among older students, there were a small number of outliers who scored higher (near 20) or lower in math.

**To compare final grades (G3) based on the mother's job (Mjob):**

**图表, 箱线图

描述已自动生成**

**Description** From this box-plot, it can be seen that students whose mothers work in health-related jobs have relatively high median final grades, and the interquartile shows that the grade distribution is more concentrated and the overall performance is better. Among students with mothers at home, the median achievement was lower and the distribution was wider, showing large differences in achievement. Students whose mothers worked in the service sector had a more even distribution of grades, with a median slightly higher than that of stay-at-home mothers, but slightly lower than in the health field. Students whose mother is a teacher, the achievement distribution is closer to the health field, and the performance is better. For mothers in other occupations, the median is similar to those in the service industry. Health, teaching and other professions, student achievement has some outlier, the achievement of individual students and the majority of the gap is large, and are reflected in the achievement is too low.

**Conclusion**

From several charts, final math grades (G3.x) are influenced by a number of factors, including gender, age, and the mother's occupation. First, male and female students have similar median grades, but female grades are more widely distributed and there are more low grades. There is no clear linear relationship between age and achievement, and while achievement is relatively concentrated among younger students, the distribution becomes more dispersed as they age. When it comes to mothers' occupations, children of mothers who work in health and education perform better, with higher median and more concentrated grades, while students whose mothers stay at home have greater differences in their grades. Overall, the mother's occupation, the student's age and gender all have some influence on the performance of math final grades.