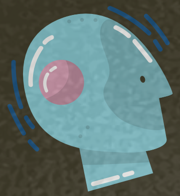
AI Analysis of Grades of Portuguese High School Students

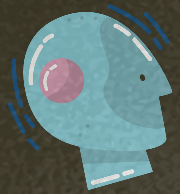
By: John Cervone and William Pitera

Executive Summary Pt 1



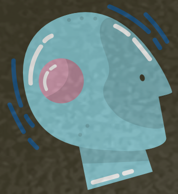
- Predicting student performance using machine learning to identify factors influencing academic success.
- Developing a predictive model to analyze student demographics, social conditions, and academic records to forecast final grades.

Executive Summary Pt 2



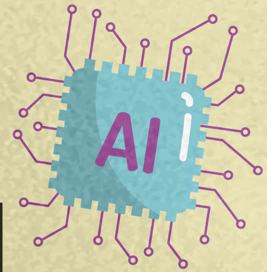
- Utilizes the Student Performance Dataset and follows a comprehensive pipeline.
- Preprocesses data to ensure clean and structured input for the model.
- Builds and optimizes machine learning models, including Random Forest and Gradient Boosting, for high accuracy.
- Evaluates the model's predictive power, targeting 75% classification accuracy or an R-squared of 0.80.

Executive Summary Pt 3



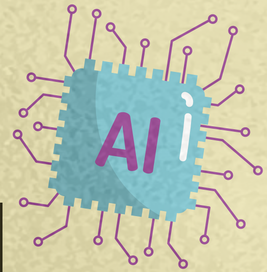
- Explores key factors impacting performance, including prior academic scores, study time, and absences, offering actionable recommendations for educators and policymakers.
- Demonstrates how machine learning enhances educational outcomes and identifies at-risk students.
- Paves the way for targeted interventions and future research in education.

Overview of Data Collection, Cleanup, & Exploration Processes



- Uses the Student Performance Dataset from the UCI Machine Learning Repository.
- Includes data on students' demographics, social factors, and academic performance in Portuguese and mathematics classes.

Overview of Data Collection, Cleanup, & Exploration Processes

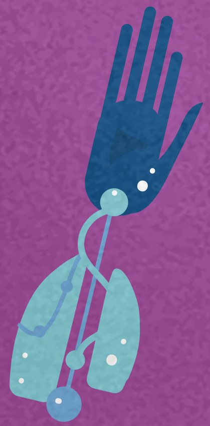


- Contains 33 attributes, including prior grades (G1, G2), study time, parental education, and absences.
- Uses the final grade (G3) as the target variable for prediction.

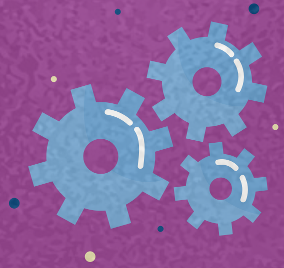
Data Cleanup

- Minimal missing data in the dataset.
- Basic checks confirmed completeness, eliminating the need for imputation.

Data Cleanup

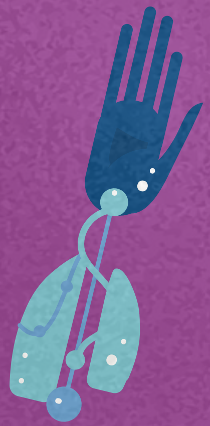


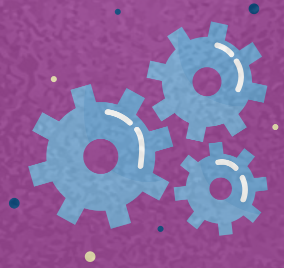
- Encoded categorical features like school, sex, address, and parental status using one-hot encoding.
- Ensured compatibility with machine learning models.



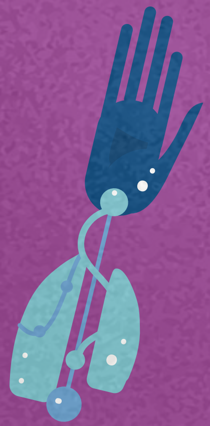
Data Cleanup

- Normalized numerical variables such as G1, G2, and absences.
- Ensured consistent scales to enhance model performance.

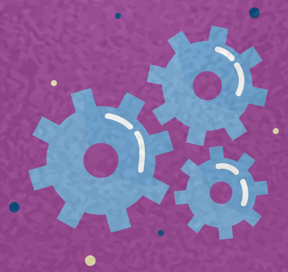




Data Cleanup



- Conducted correlation analysis to identify impactful features influencing final grades.
- Key predictors included prior grades (G1, G2), study time, and family education levels.



Data Exploration

- Performed descriptive analysis to understand distributions, averages, and variability across attributes.
- Identified strong positive correlations between G1, G2, and G3 scores.

Data Exploration

- Used histograms and box plots to reveal patterns and outliers in numerical features like absences and grades.

Data Exploration

- Class Imbalance: The target variable (G3) displayed slight imbalances, with more students clustering in mid-performance ranges.
- Informed the choice of evaluation metrics and model selection.

The Approach

- Followed a systematic and iterative approach to predict student performance and identify key factors influencing academic outcomes.
- The process was structured as follows:

Approach to Data Preprocessing

- Exploration: Explored the dataset to understand its structure, key variables, and potential challenges (e.g., missing data or outliers).
- Cleaning: Cleaned the data by encoding categorical variables (e.g., school, address) using one-hot encoding and scaling numerical variables (e.g., grades, absences) for standardization.
- Feature Selection: Performed correlation analysis to prioritize features with the most predictive power, focusing on prior grades, study time, and family education levels.

Approach to Model Development

- Baseline Model: Implemented a simple Logistic Regression model to establish a baseline accuracy for classification.
- Advanced Models: Explored an advanced model, Random Forest, to improve predictive performance.
- Hyperparameter Tuning: Fine-tuned parameters like the number of trees and depth for the Random Forest model using grid search and cross-validation to achieve optimal results.

Approach to Model Evaluation

- Metrics: Evaluated model performance using accuracy, precision, recall, and F1-score for classification tasks, and R-squared for regression tasks.
- Visualization: Used confusion matrices and feature importance plots to gain insights into model behavior and identify influential variables.

Approach to Documentation & Presentation

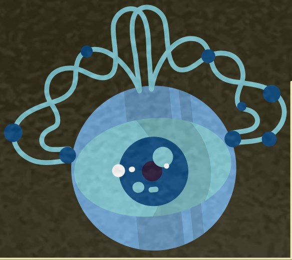
- GitHub Repository: Documented the entire workflow, from data preprocessing to model evaluation, in a clean, organized repository. Included a polished README file with an overview of the project.
- Result Analysis: Summarized the final results, including the best-performing model and key takeaways, for the presentation.

Approach to Iterative Refinement Pt 1

- Iterative Approach: Regularly revisited earlier steps, refining data preprocessing, feature engineering, and model selection to address issues and enhance performance.
- Outcome: The iterative approach ensured a robust final solution that exceeded project benchmarks.

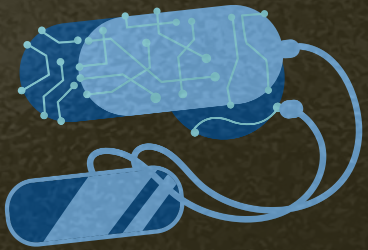
Approach to Iterative Refinement Pt 2

- Approach: Combined data exploration, advanced modeling techniques, and rigorous evaluation.
- Outcome: Successfully achieved project goals and gained valuable insights into factors driving student performance.



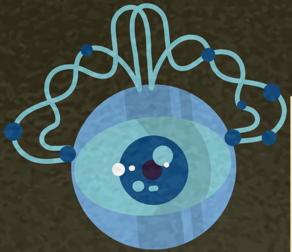
Future Research Directions Pt 1

- Analyzing Additional Factors:
  a. Incorporate external data sources (e.g., school funding, teacher-student ratios, neighborhood socioeconomic status) to understand their influence on academic performance.
  b. Study the impact of psychological and emotional factors (e.g., stress levels, peer pressure, motivation) that were not included in the dataset.



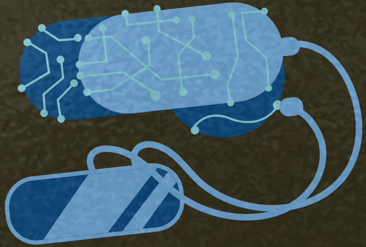
Future Research Directions Pt 2

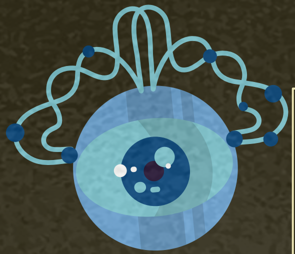
- Investigating Longitudinal Trends:
  - Conduct a longitudinal analysis to track how student performance evolves over multiple academic years.
  - Identify early warning signs for at-risk students.



Future Research Directions Pt 3

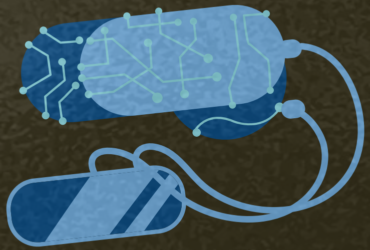
- Exploring Intervention Strategies:
  - Research how targeted interventions (e.g., extra tutoring, counseling, or parental engagement) impact academic performance.
  - Simulate and compare potential outcomes of different educational strategies using predictive modeling.

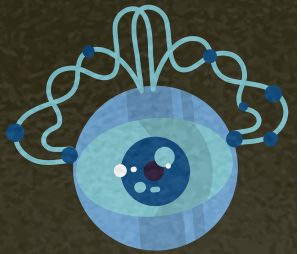


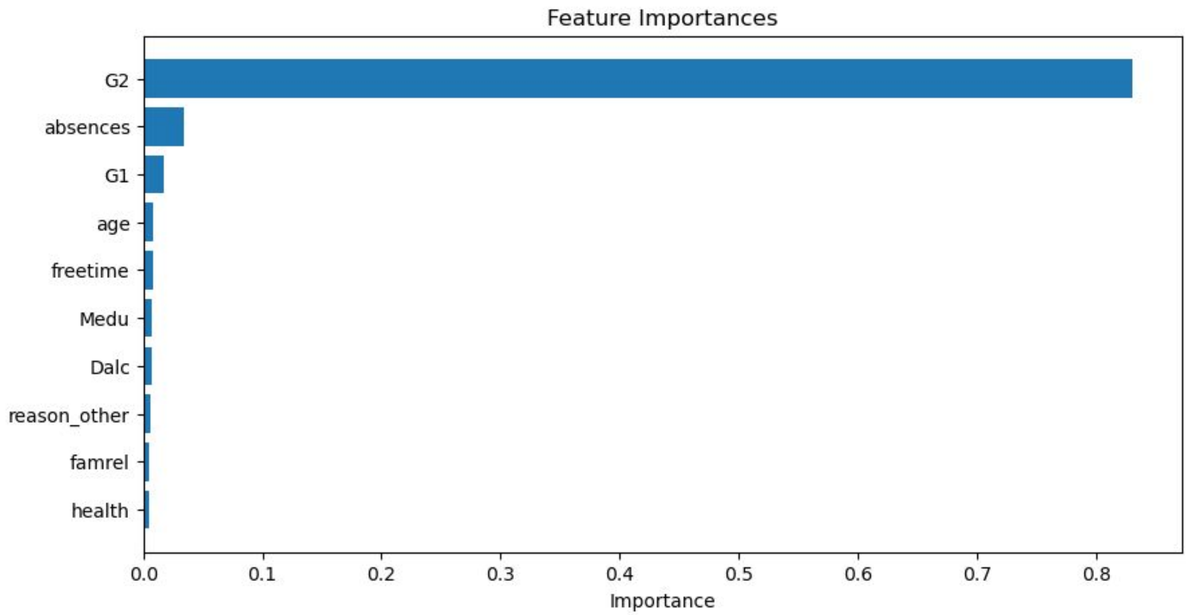


Future Research Directions Pt 4

- Advancing Modeling Techniques:
  - Experiment with deep learning models, such as neural networks, to capture more complex relationships in the data.
  - Incorporate Explainable AI (XAI) techniques to better interpret and communicate the model's predictions to educators and policymakers.

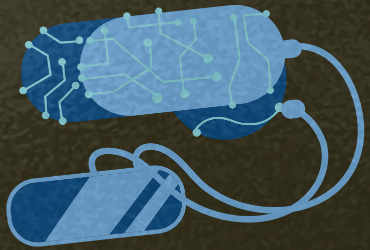






Top 10 Feature Importance

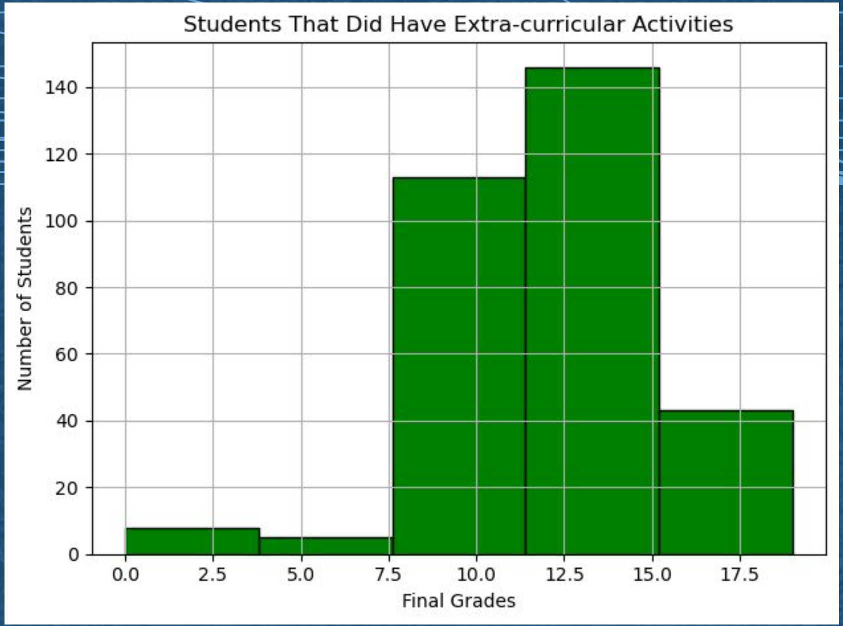
G2 = 83% importance
Absence = 3% Importance
G1 = 1% Importance
The rest are > 1% importance

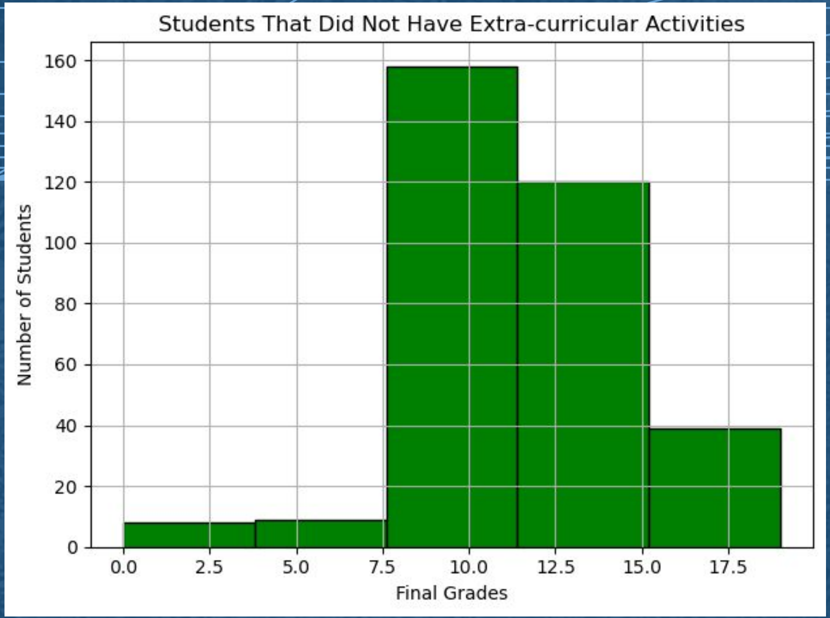


Findings and Results To Questions

How does a student pursuing extracurricular activities affect the students final grades?

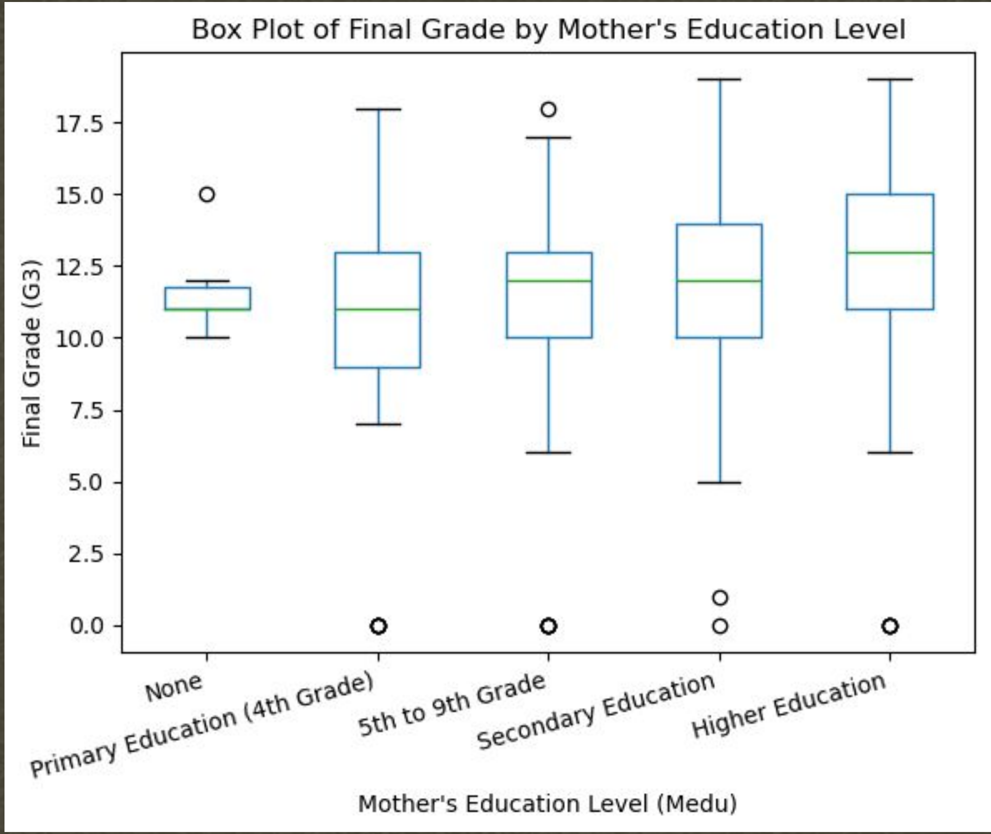
By: John


Students That Did Have Extra-curricular Activities
Number of Students
Final Grades
0
20
40
60
80
100
120
140
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5


Students That Did Not Have Extra-curricular Activities
Number of Students
Final Grades
0
20
40
60
80
100
120
140
160
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

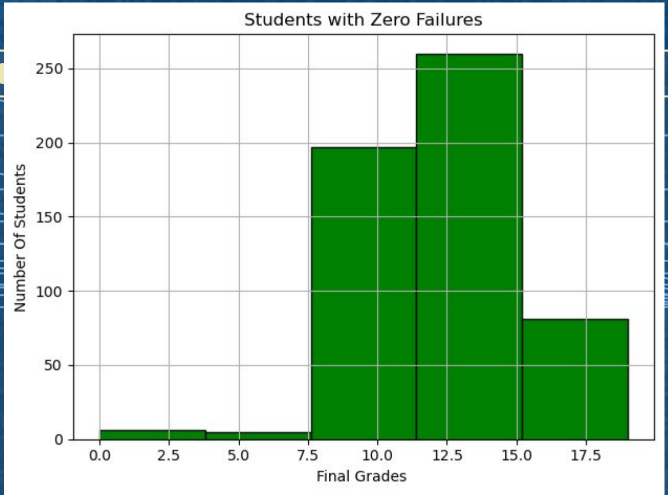
Does the mothers education level affect the students grades?

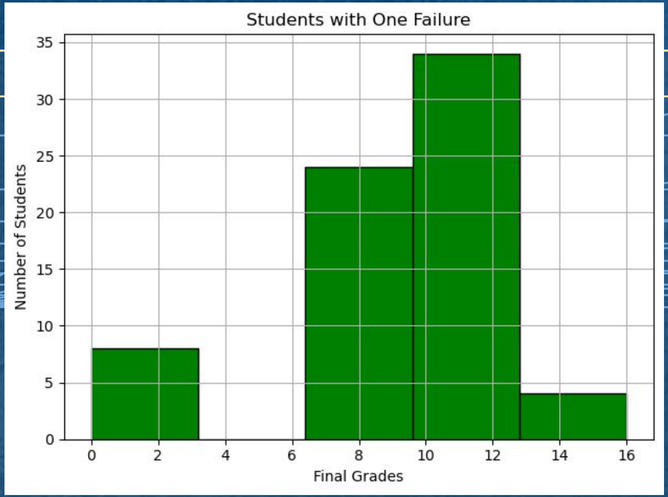
By: Will

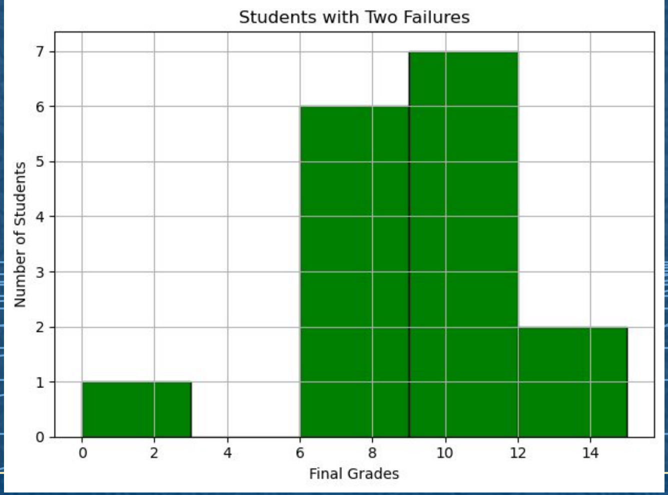

Box Plot of Final Grade by Mother's Education Level
Final Grade (G3)
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
None
Primary Education (4th Grade)
5th to 9th Grade
Secondary Education
Higher Education
Mother's Education Level (Medu)

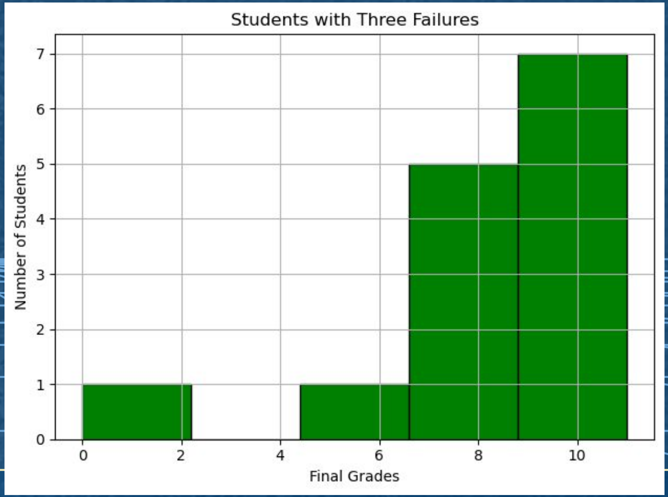
What is the relationship between number of failures and grades?

By: John


Students with Zero Failures
Number Of Students
Final Grades
0
50
100
150
200
250
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

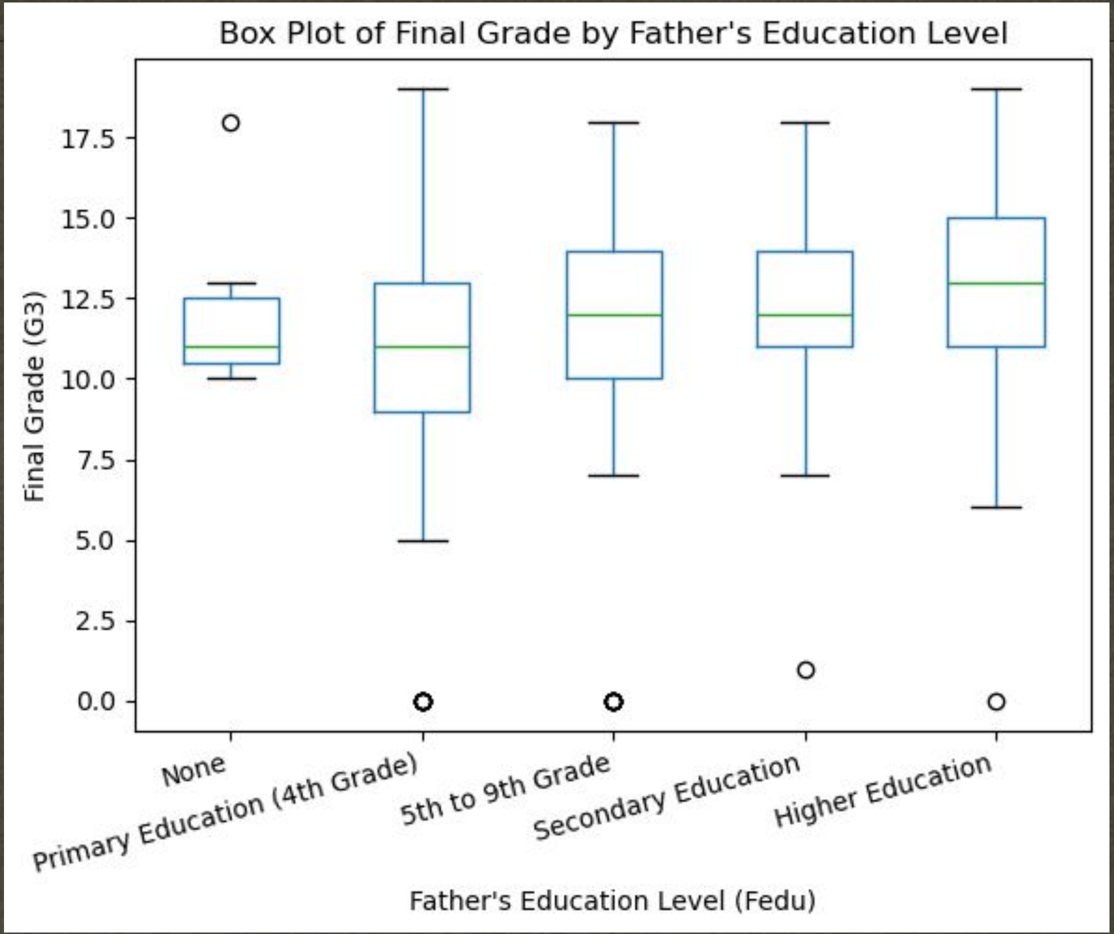

Students with One Failure
Number of Students
Final Grades
0
5
10
15
20
25
30
35
0
2
4
6
8
10
12
14
16


Students with Two Failures
Number of Students
Final Grades
0
1
2
3
4
5
6
7
0
2
4
6
8
10
12
14


Students with Three Failures
Number of Students
Final Grades
0
1
2
3
4
5
6
7
0
2
4
6
8
10

Does the fathers education level affect the students grades?

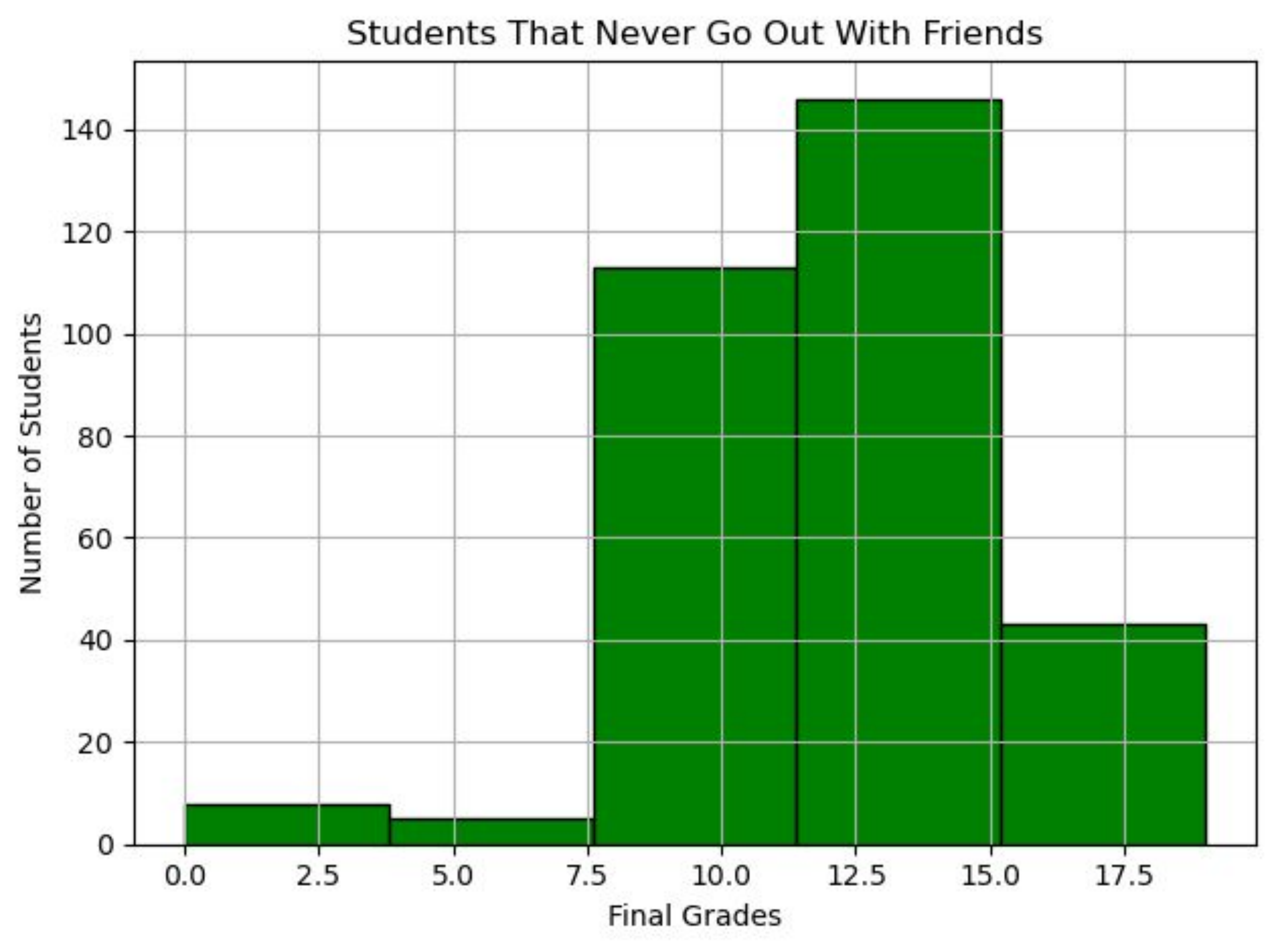
By: Will

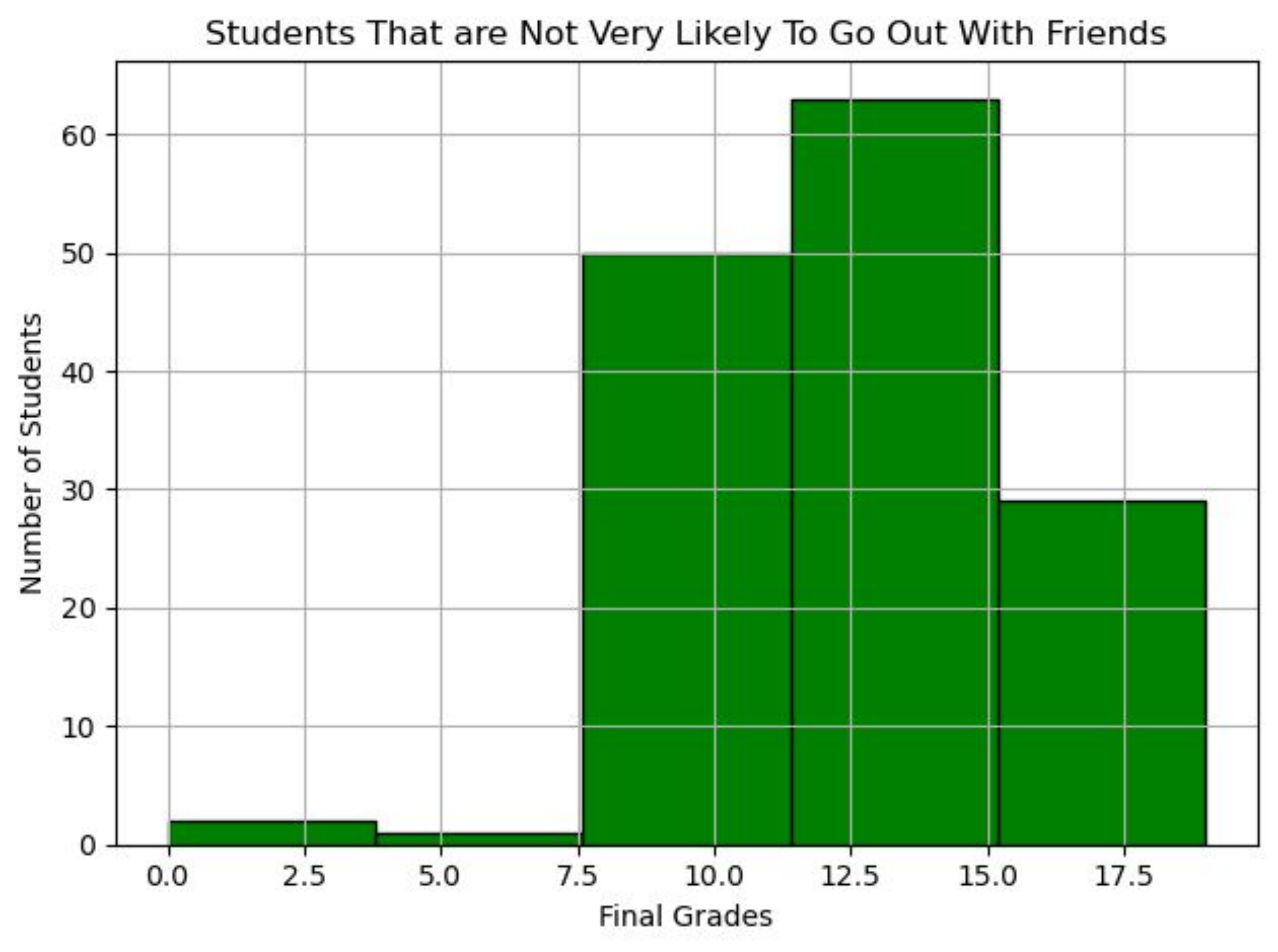


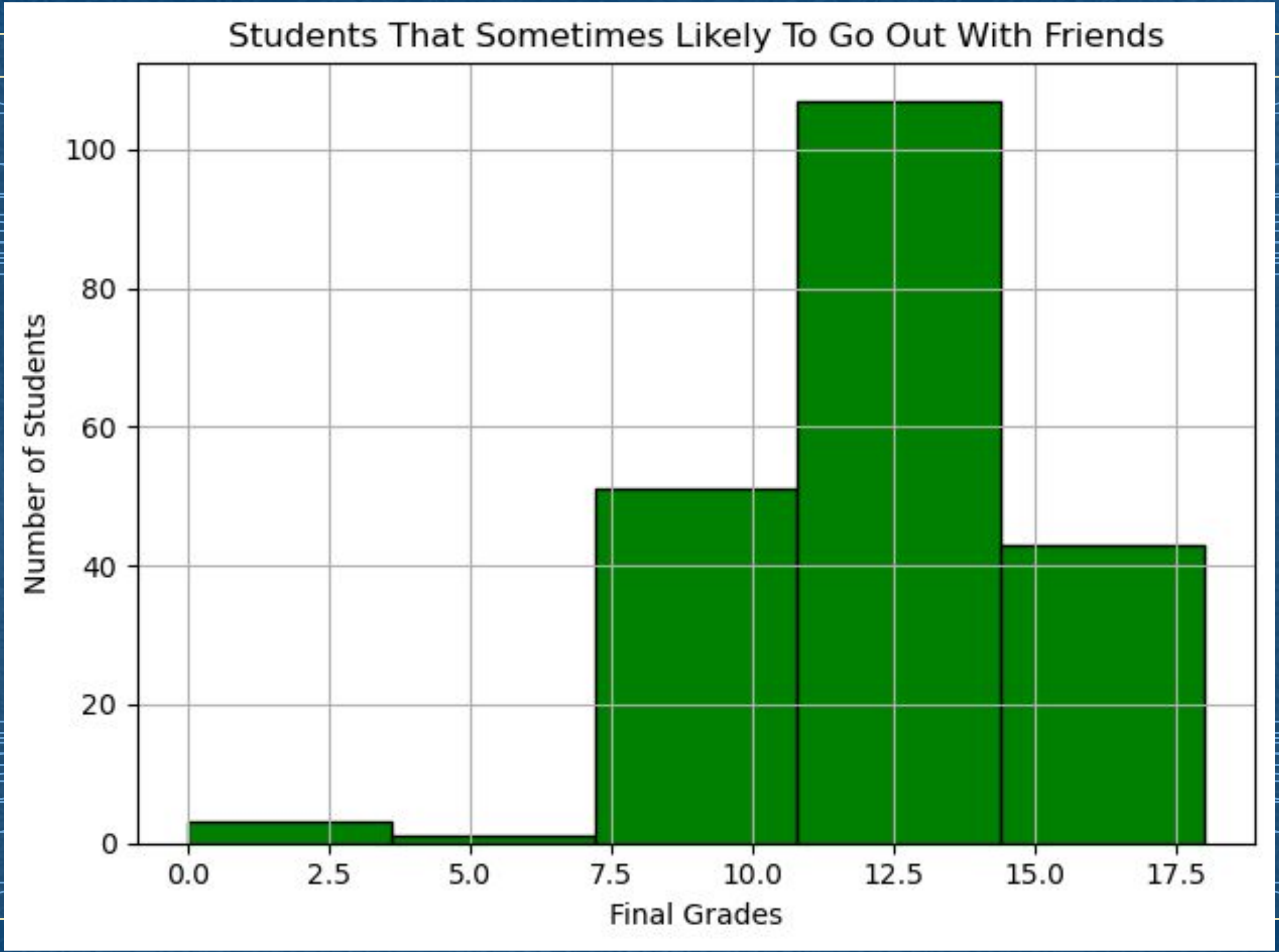
Box Plot of Final Grade by Father's Education Level
Final Grade (G3)
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
None
Primary Education (4th Grade)
5th to 9th Grade
Secondary Education
Higher Education
Father's Education Level (Fedu)

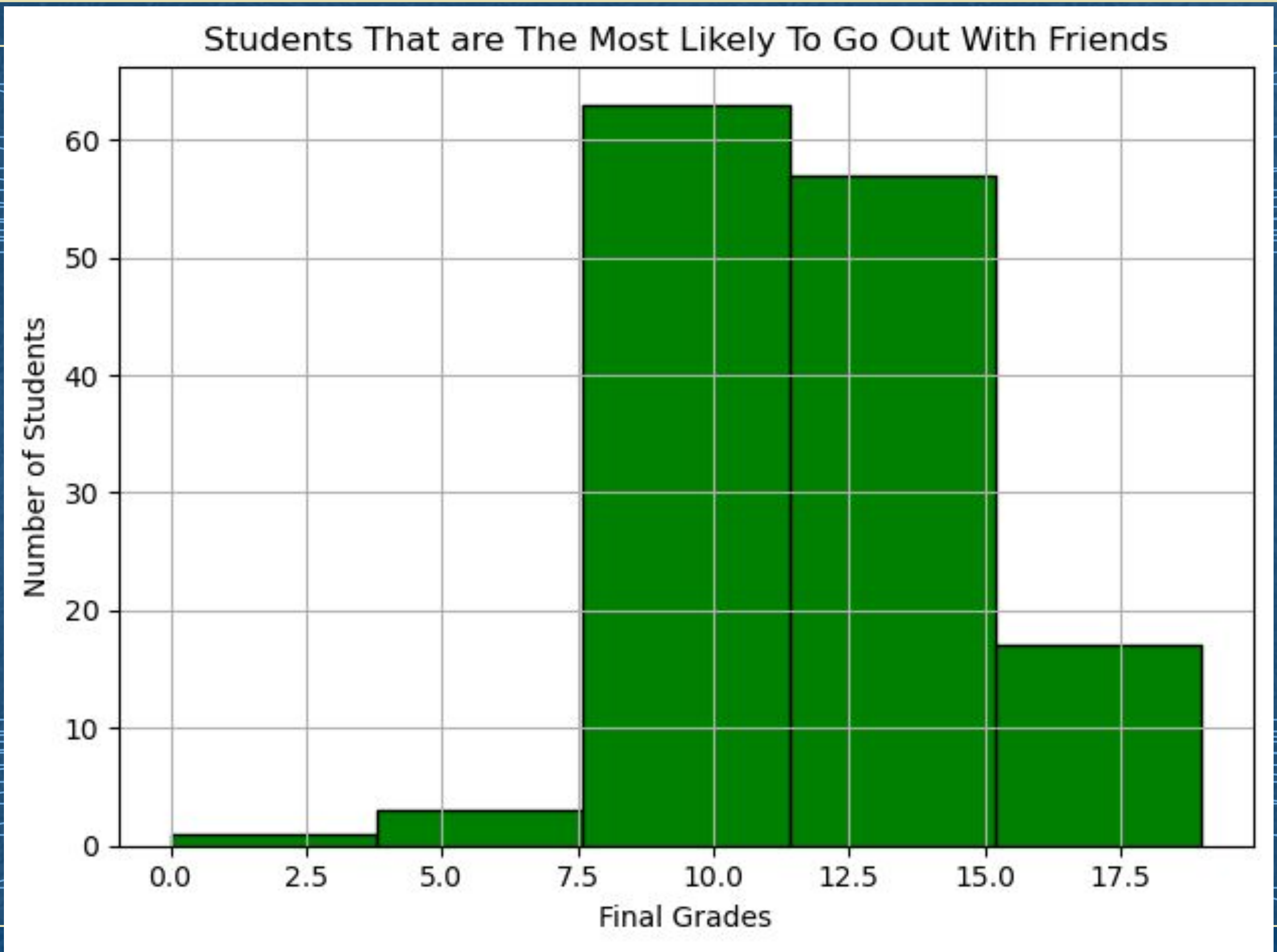
Does going out with friends correlate to good grades?

By: John

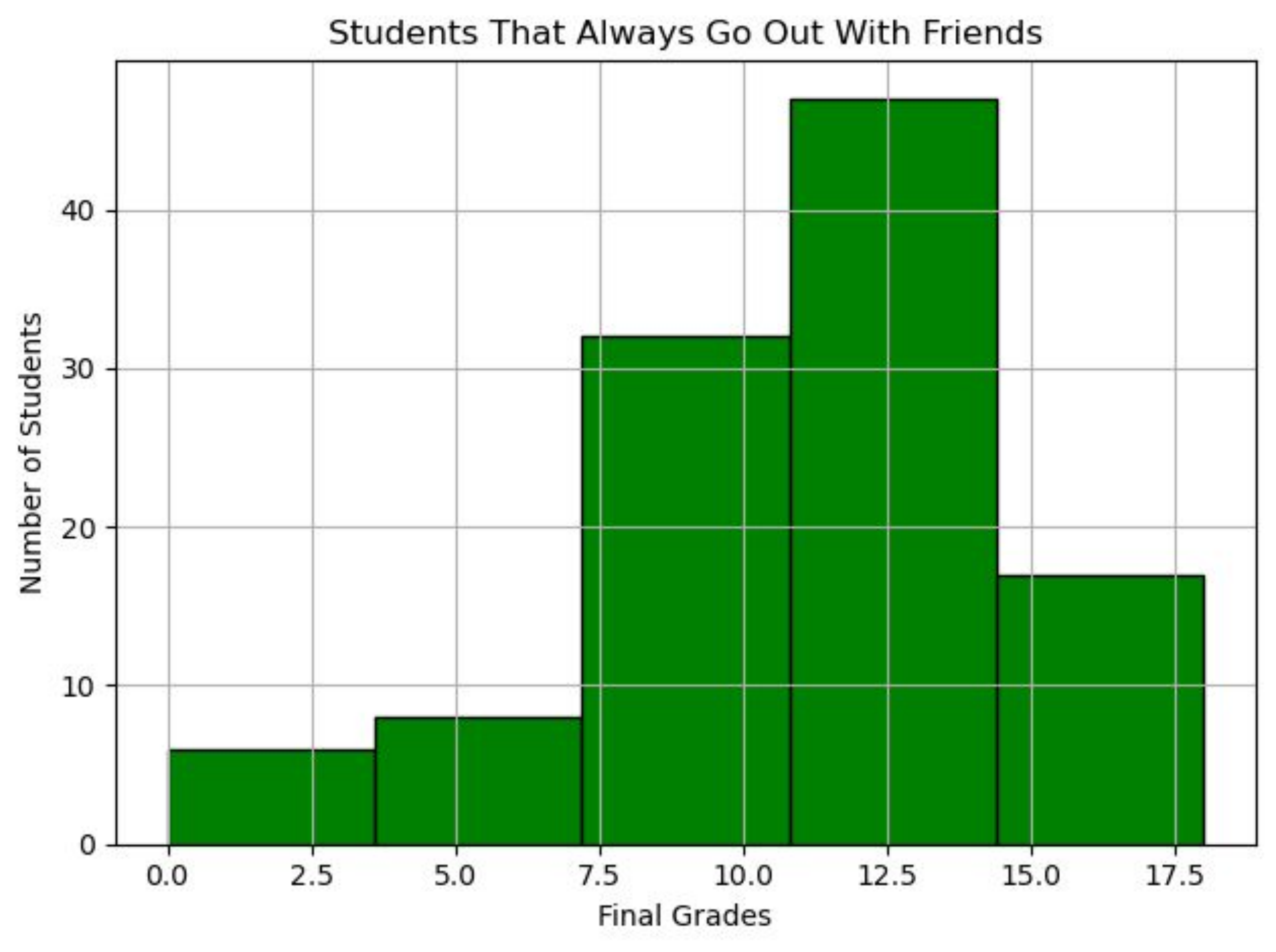

Students That Never Go Out With Friends
Number of Students
Final Grades
0
20
40
60
80
100
120
140
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

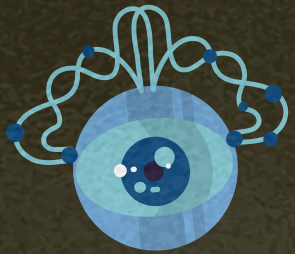

Students That are Not Very Likely To Go Out With Friends
Number of Students
0
10
20
30
40
50
60
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Final Grades


Students That Sometimes Likely To Go Out With Friends
Number of Students
0
20
40
60
80
100
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Final Grades



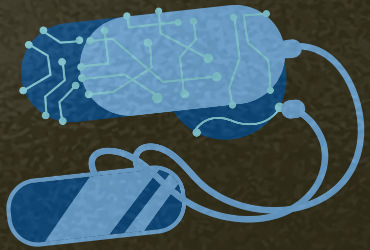
Students That are The Most Likely To Go Out With Friends
Number of Students
0
10
20
30
40
50
60
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Final Grades

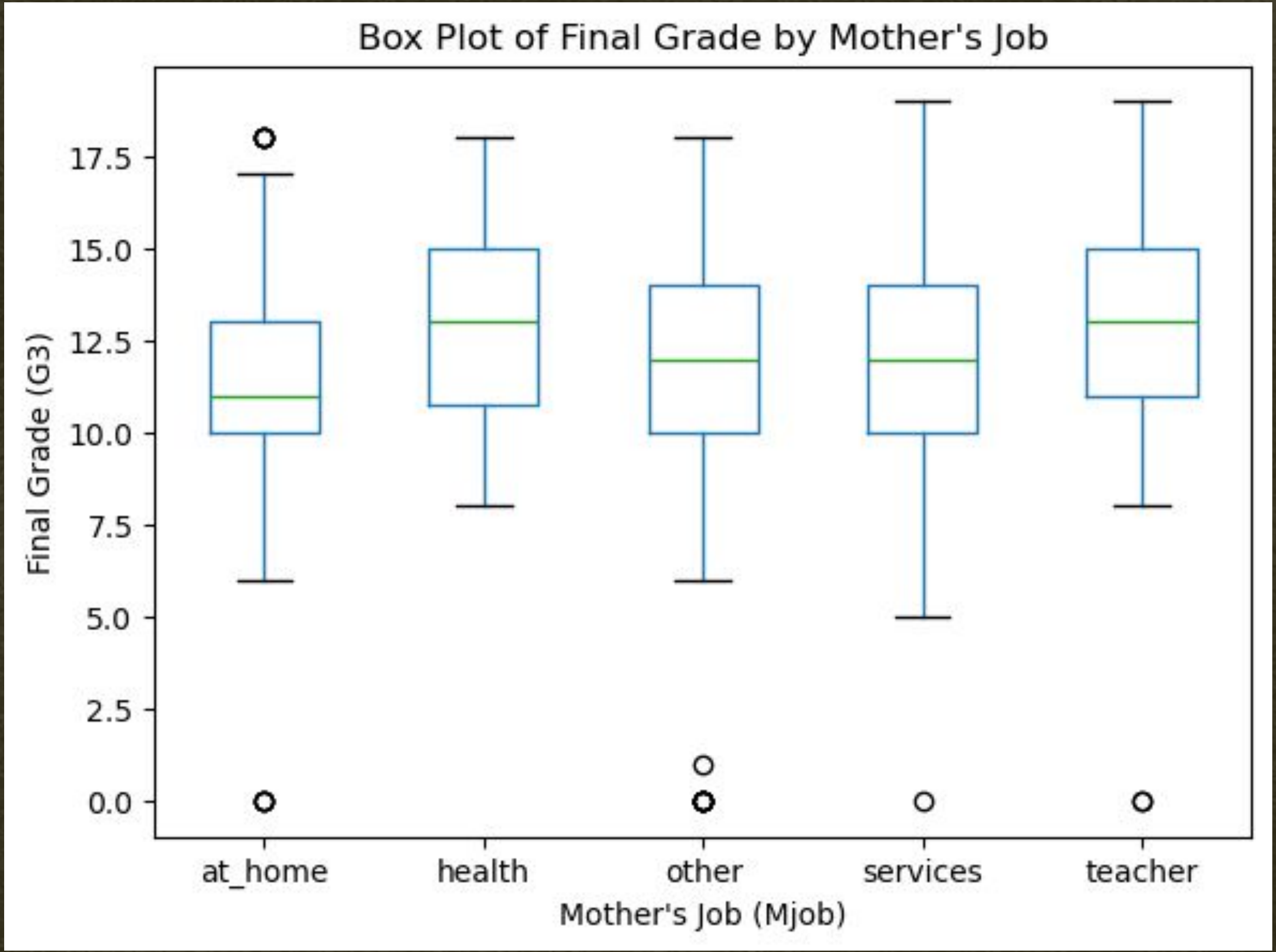

Students That Always Go Out With Friends
Number of Students
0
10
20
30
40
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Final Grades



Does the mothers jobs affect the students grades?

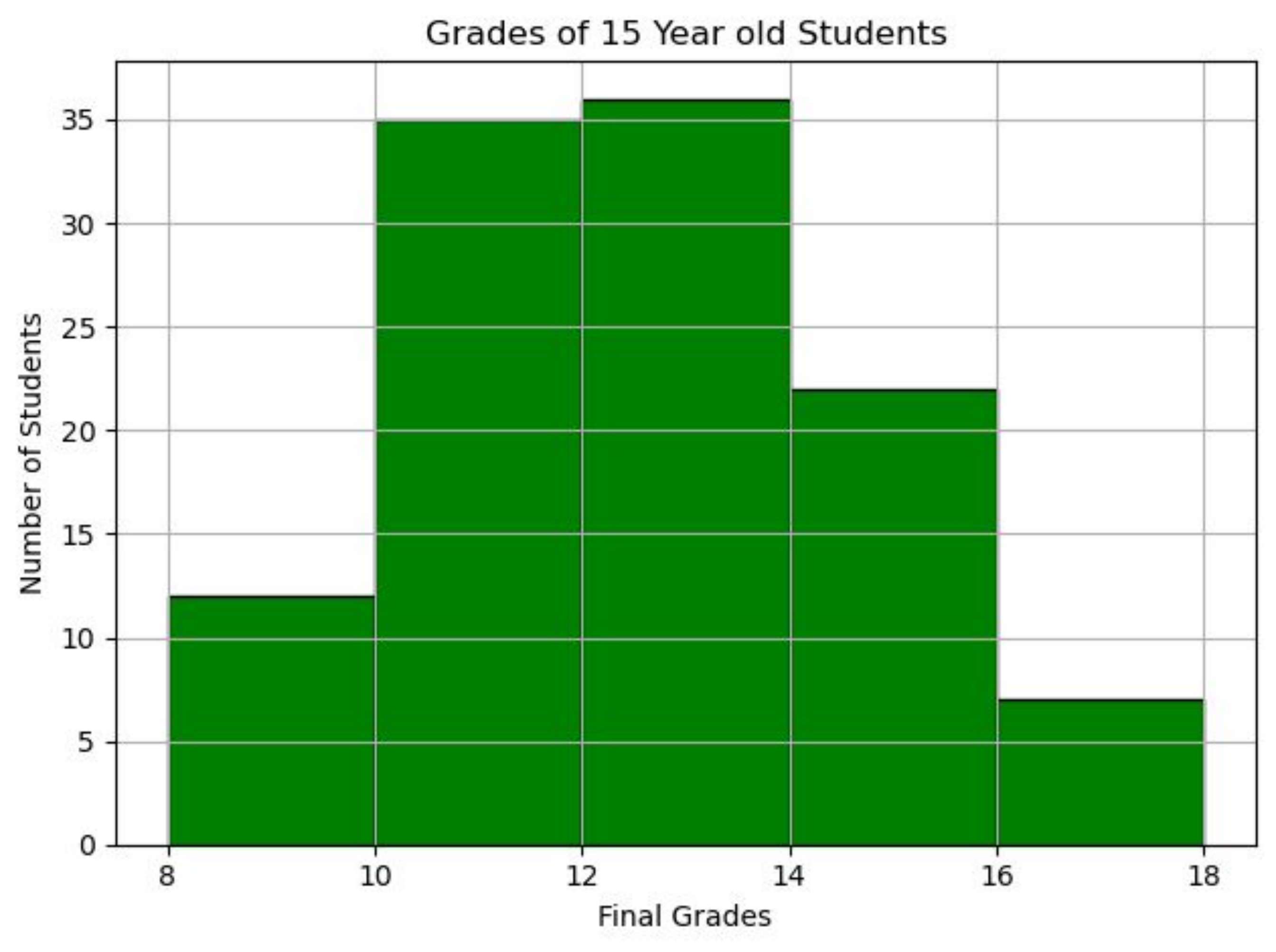
By: Will

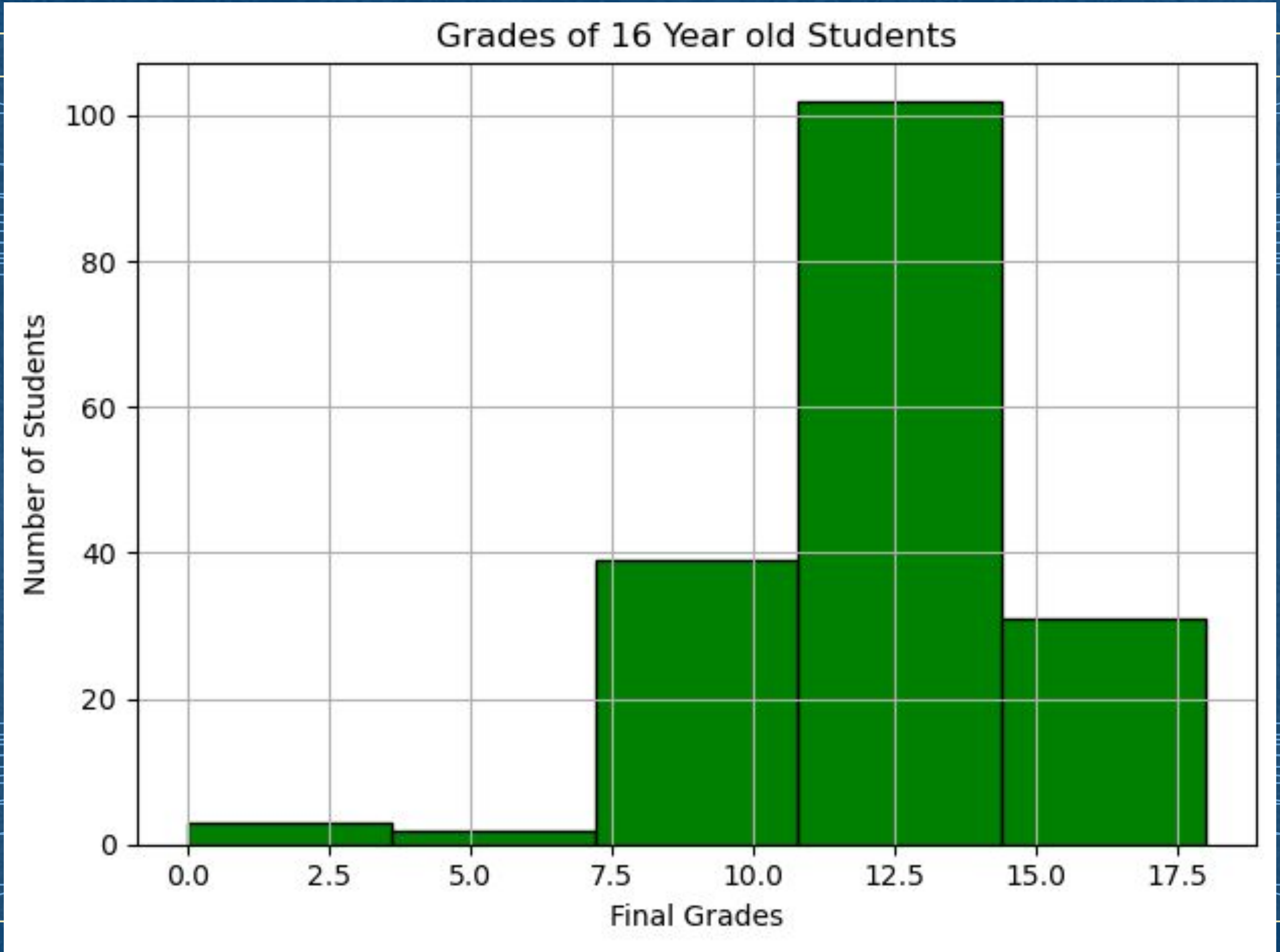


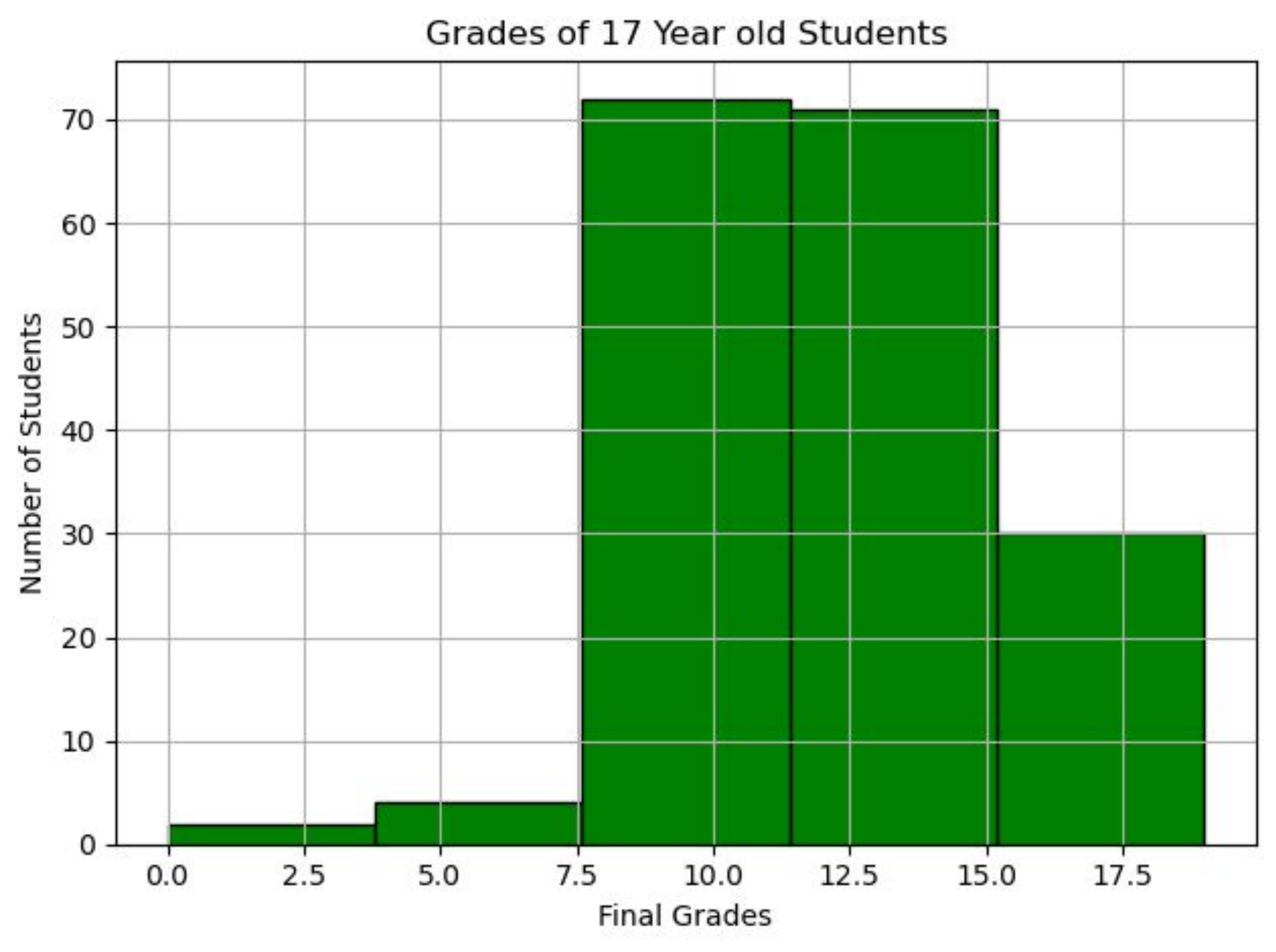

Box Plot of Final Grade by Mother's Job
Final Grade (G3)
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
at_home
health
other
services
teacher
Mother's Job (Mjob)

Does age correlate to grades?

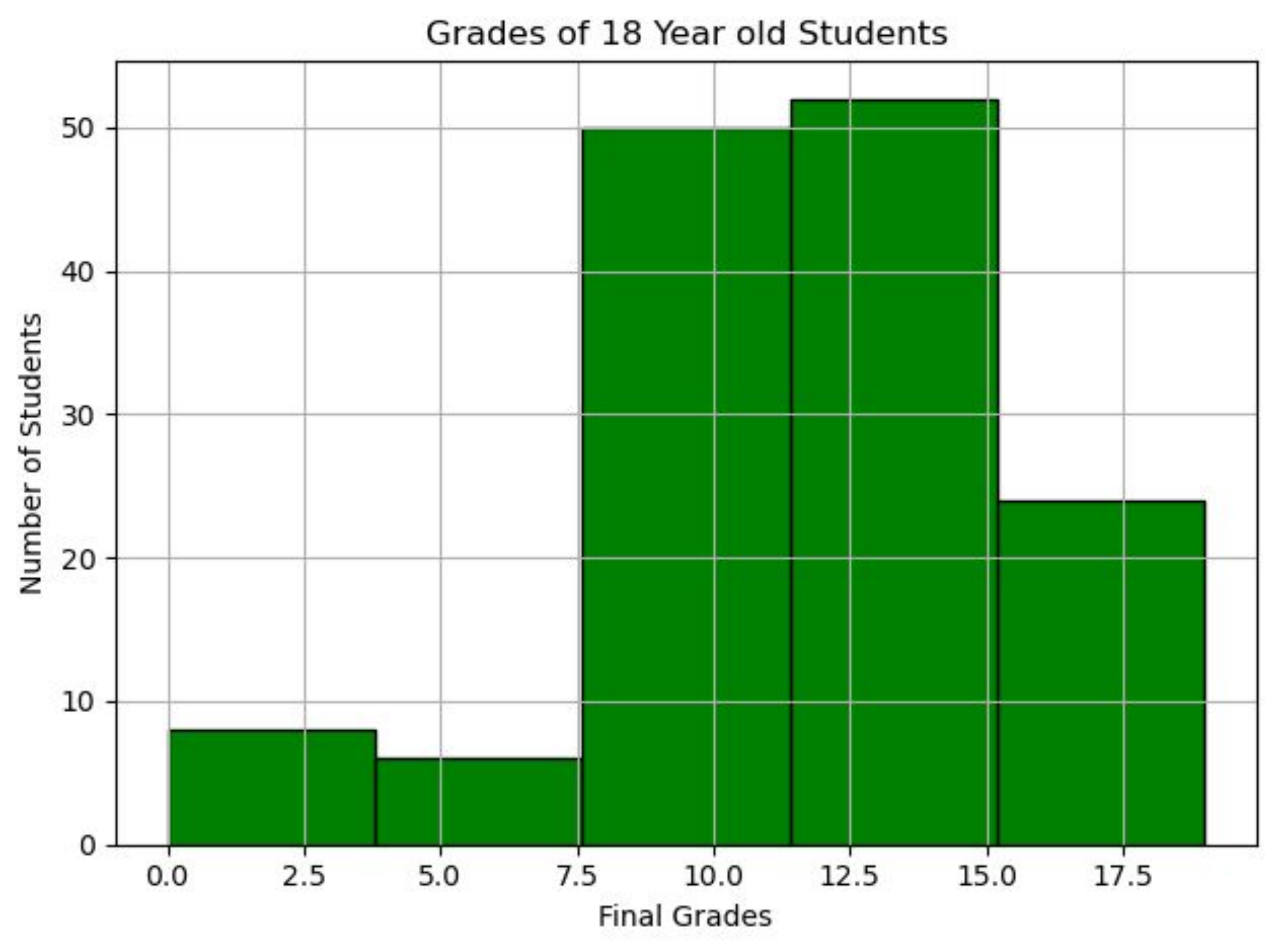
By: John

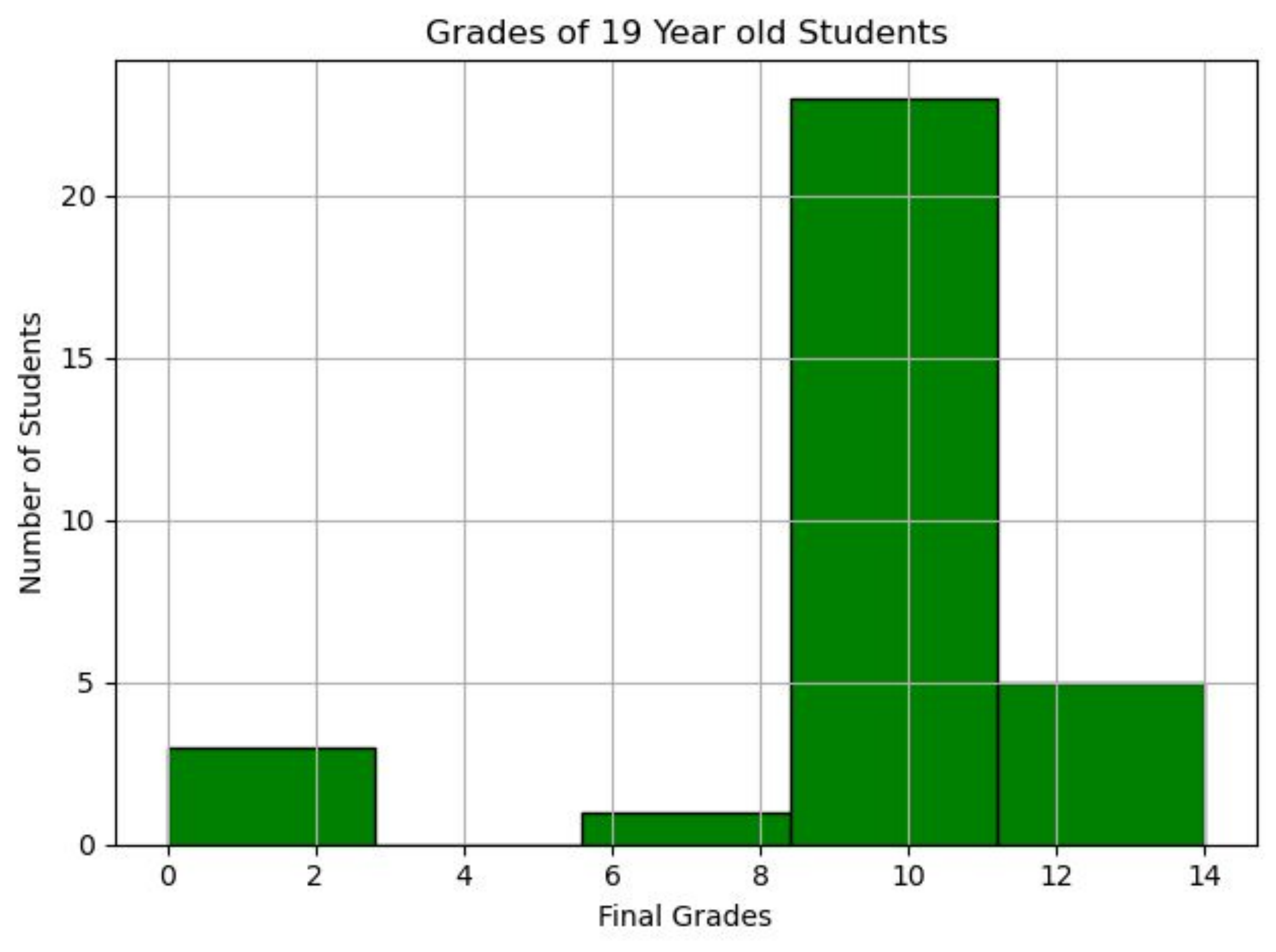

Grades of 15 Year old Students
Number of Students
0
5
10
15
20
25
30
35
8
10
12
14
16
18
Final Grades

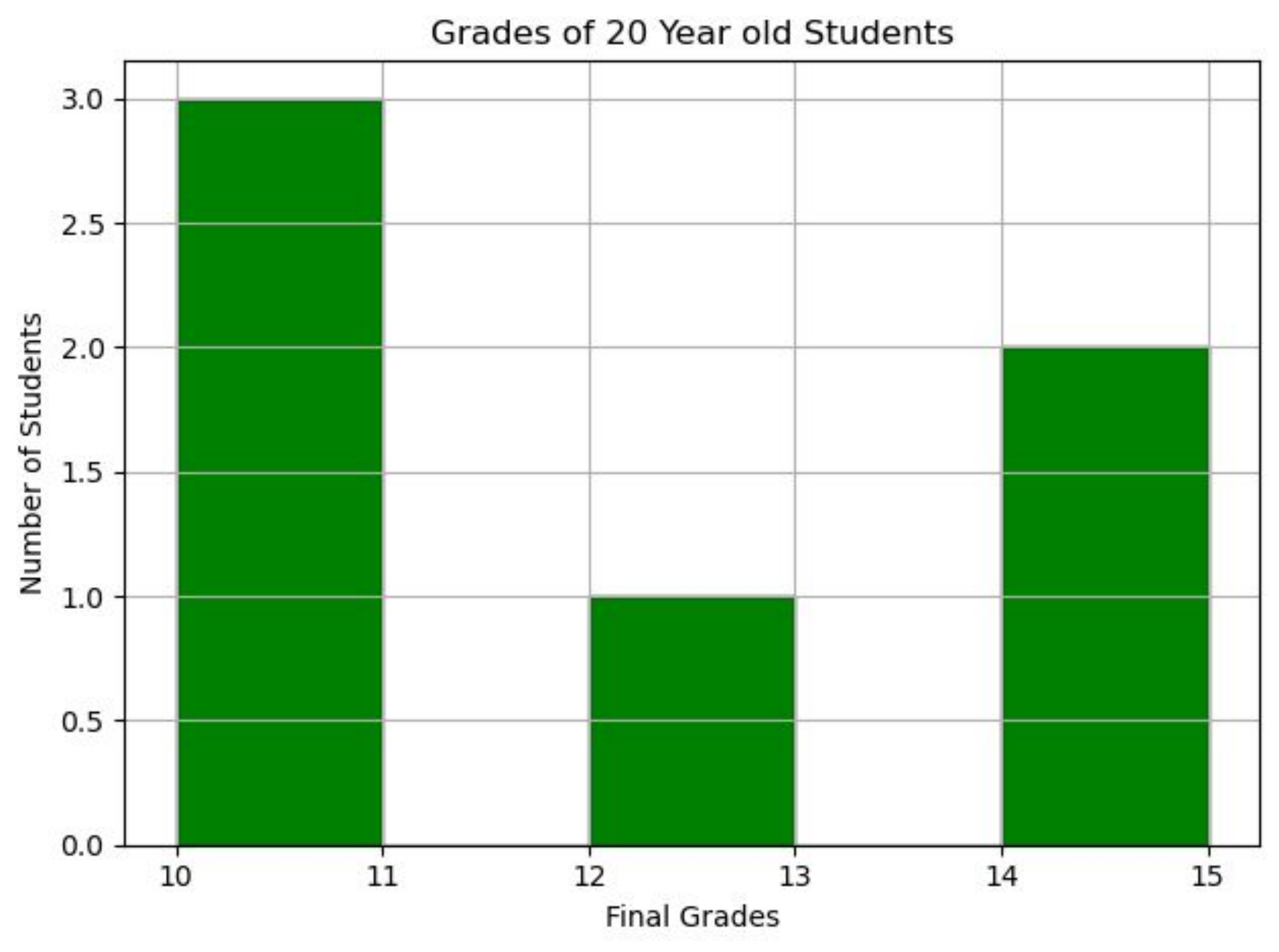

Grades of 16 Year old Students
Number of Students
Final Grades
0
20
40
60
80
100
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

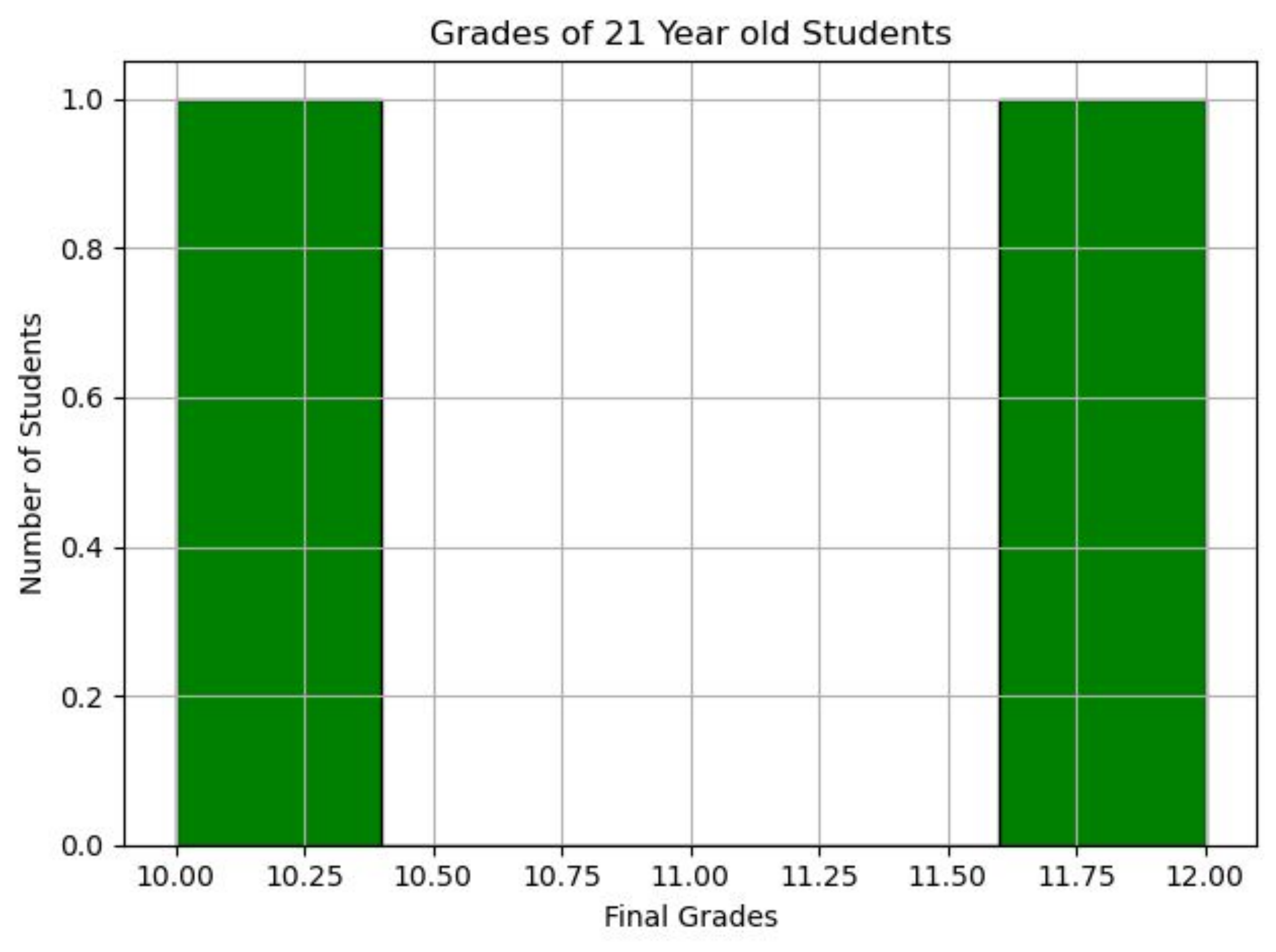


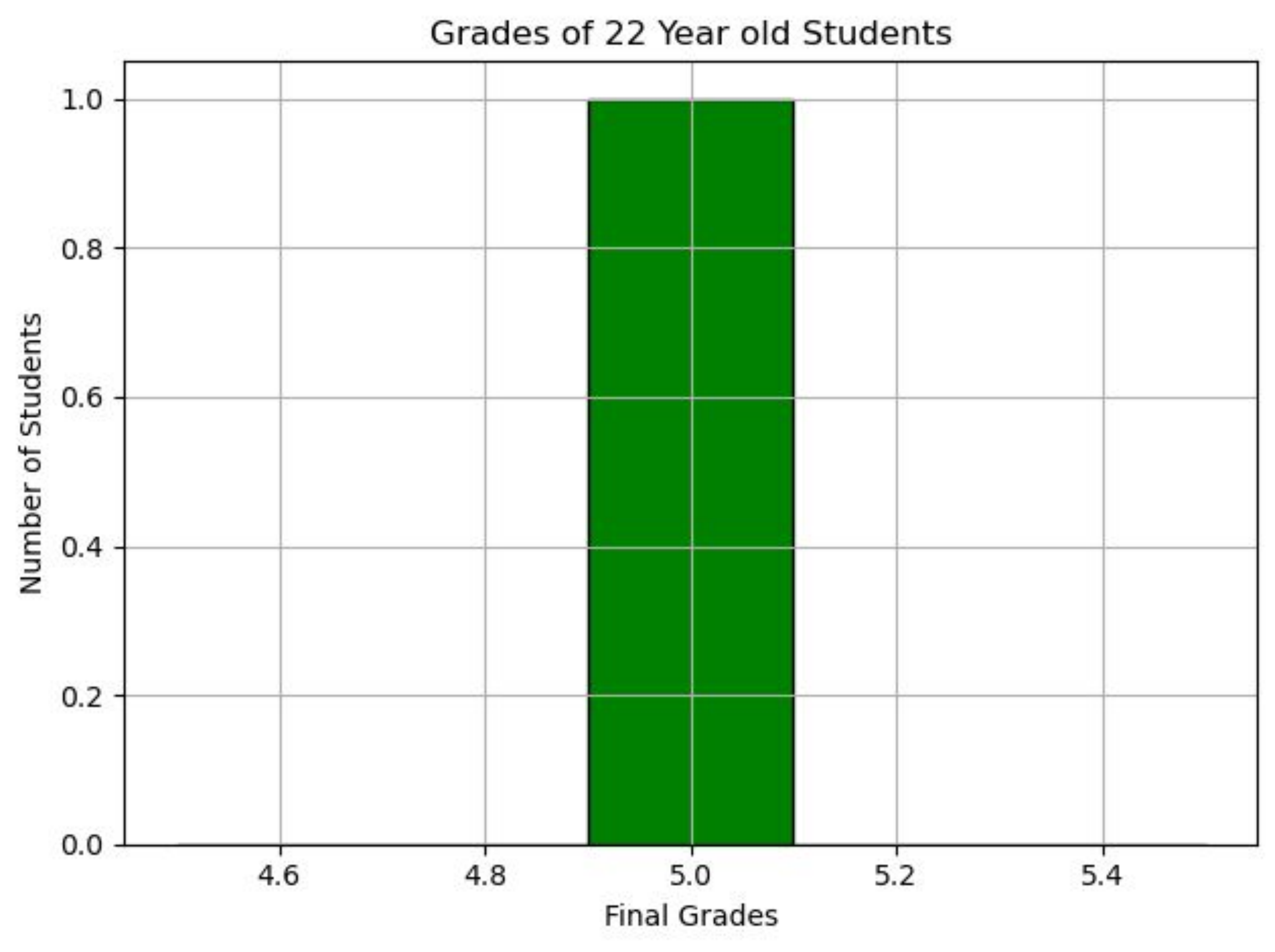
Grades of 17 Year old Students
Number of Students
0
10
20
30
40
50
60
70
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Final Grades

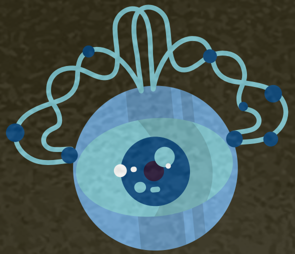

Grades of 18 Year old Students
Number of Students
0
10
20
30
40
50
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Final Grades


Grades of 19 Year old Students
Number of Students
Final Grades
0
5
10
15
20
0
2
4
6
8
10
12
14


Grades of 20 Year old Students
Number of Students
3.0
2.5
2.0
1.5
1.0
0.5
0.0
10
11
12
13
14
15
Final Grades

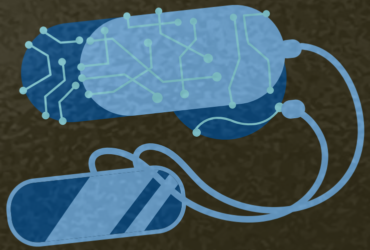

Grades of 21 Year old Students
Number of Students
1.0
0.8
0.6
0.4
0.2
0.0
10.00
10.25
10.50
10.75
11.00
11.25
11.50
11.75
12.00
Final Grades

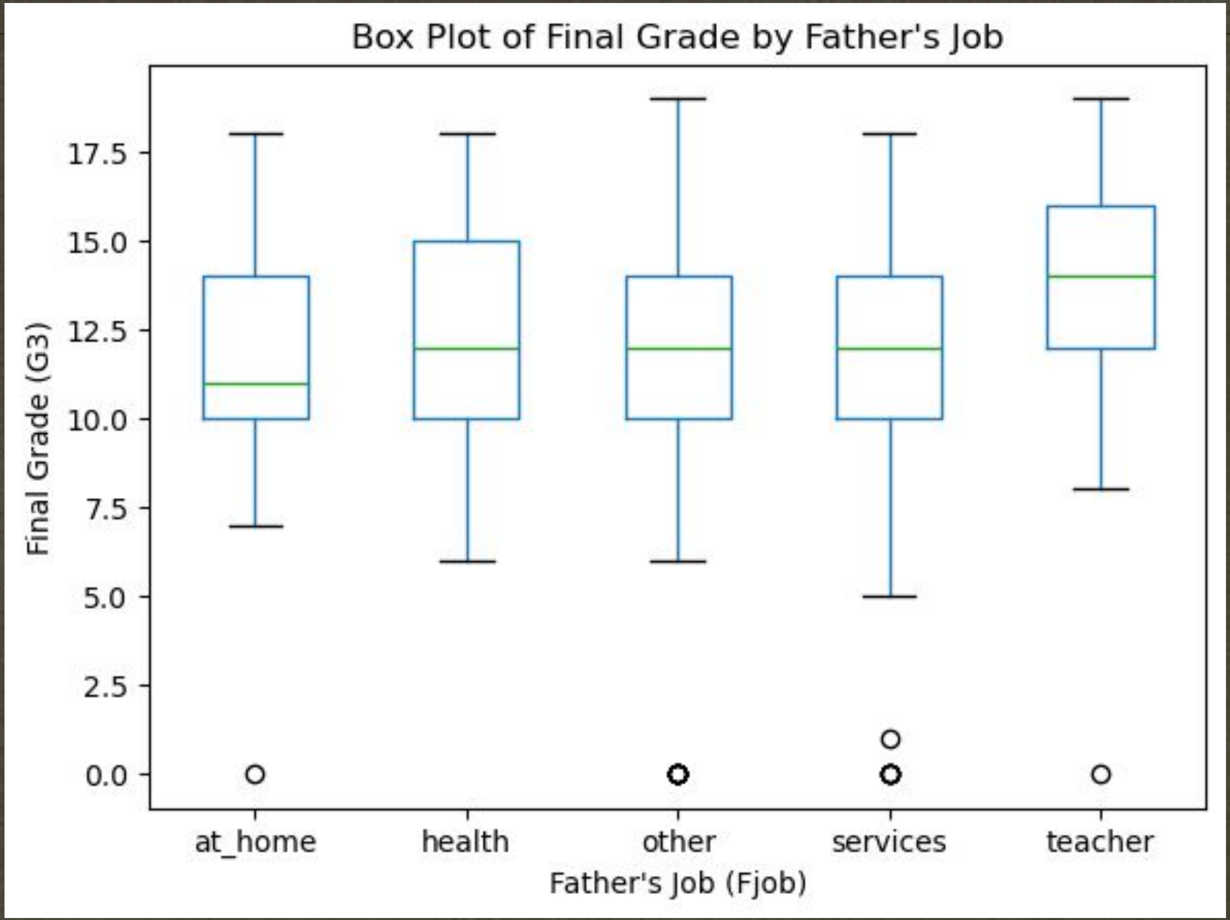

Grades of 22 Year old Students
Number of Students
1.0
0.8
0.6
0.4
0.2
0.0
4.6
4.8
5.0
5.2
5.4
Final Grades

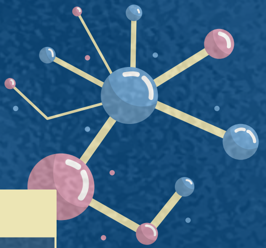


Does the fathers jobs affect the students grades?

By: Will




Box Plot of Final Grade by Father's Job
Final Grade (G3)
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
at_home
health
other
services
teacher
Father's Job (Fjob)



Thank You For Listening!

Any Questions?