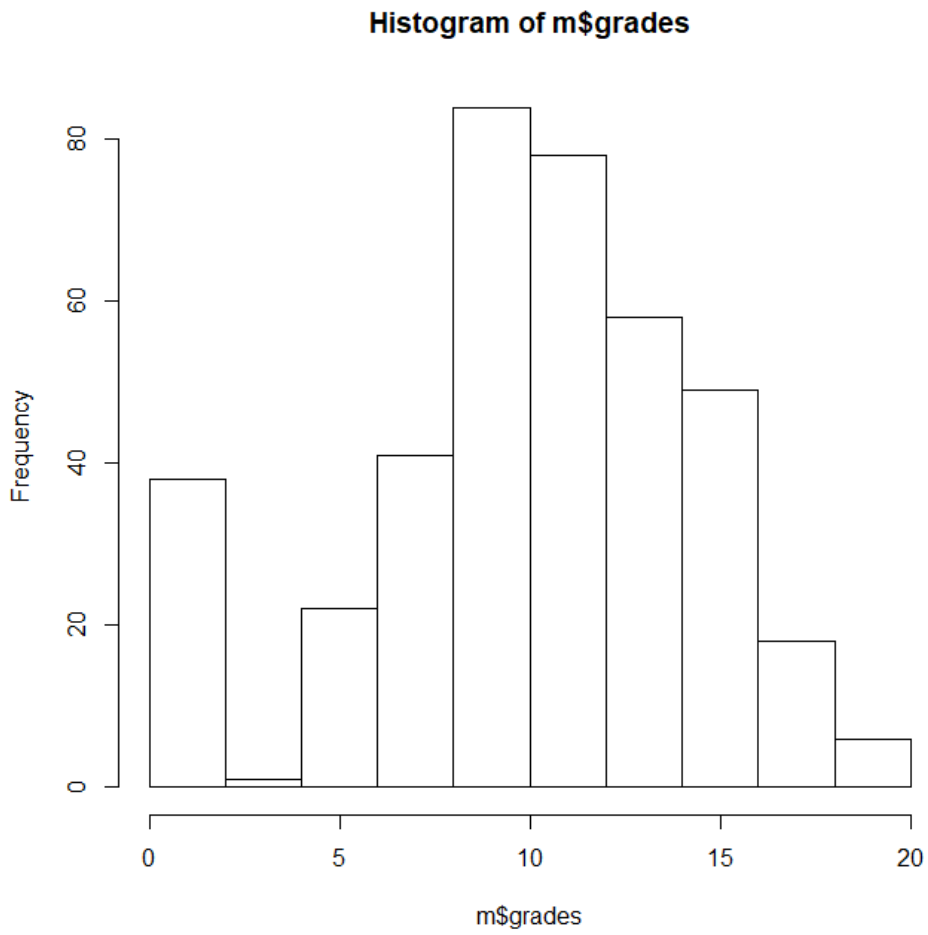
Feedbacks for Case Study 2

Ling-Chieh Kung

Always start from a histogram

- To see whether we need to separate observations into groups.



Correlation coefficients

- Don’t do that for:
  - Medu and Fedu
  - gender and school
  - Etc.
- It’s good to have a complete correlation coefficient table “internally” to check multicollinearity among quantitative variables.
  - However, do not put the whole table into the slides or report.
  - Just put the quantitative part.

Qualitative vs. quantitative

- Gender → definitely qualitative
- Mother education → almost always qualitative
- Study time (0-2, 2-5, 5-10, 10 or above) → either
  - Set to 1, 3.5, 7.5, and 12.5 if treated as quantitative
- Health (or any kind of rating) → often quantitative
- Absence → definitely quantitative

Absence^2

- If we have a model with absence and absence^2, the interpretation is:
  - If the coefficient for absence^2 is positive: Some absence hurts the grades, but a lot of absence helps the grades (or there are some students getting high grades with a lot of absence).
  - If the coefficient for absence^2 is negative: The other way

Outliers

- Let’s assume that there is no wrong value
- If some observations have “weird” values for the dependent variable
  - Avoid removing them
  - We are trying to explain why that happens
- If some observations have “weird” values for the independent variable
  - It is more acceptable to remove them
  - They are just “special”

Counter-intuitive findings

- We have some intuitions at the beginning
  - Descriptive statistics strengthens some intuitions
- It is good (in many cases) to show that the intuitions are not true
  - By using, e.g., regression