FIT5145 Workshop Week 8

Objectives

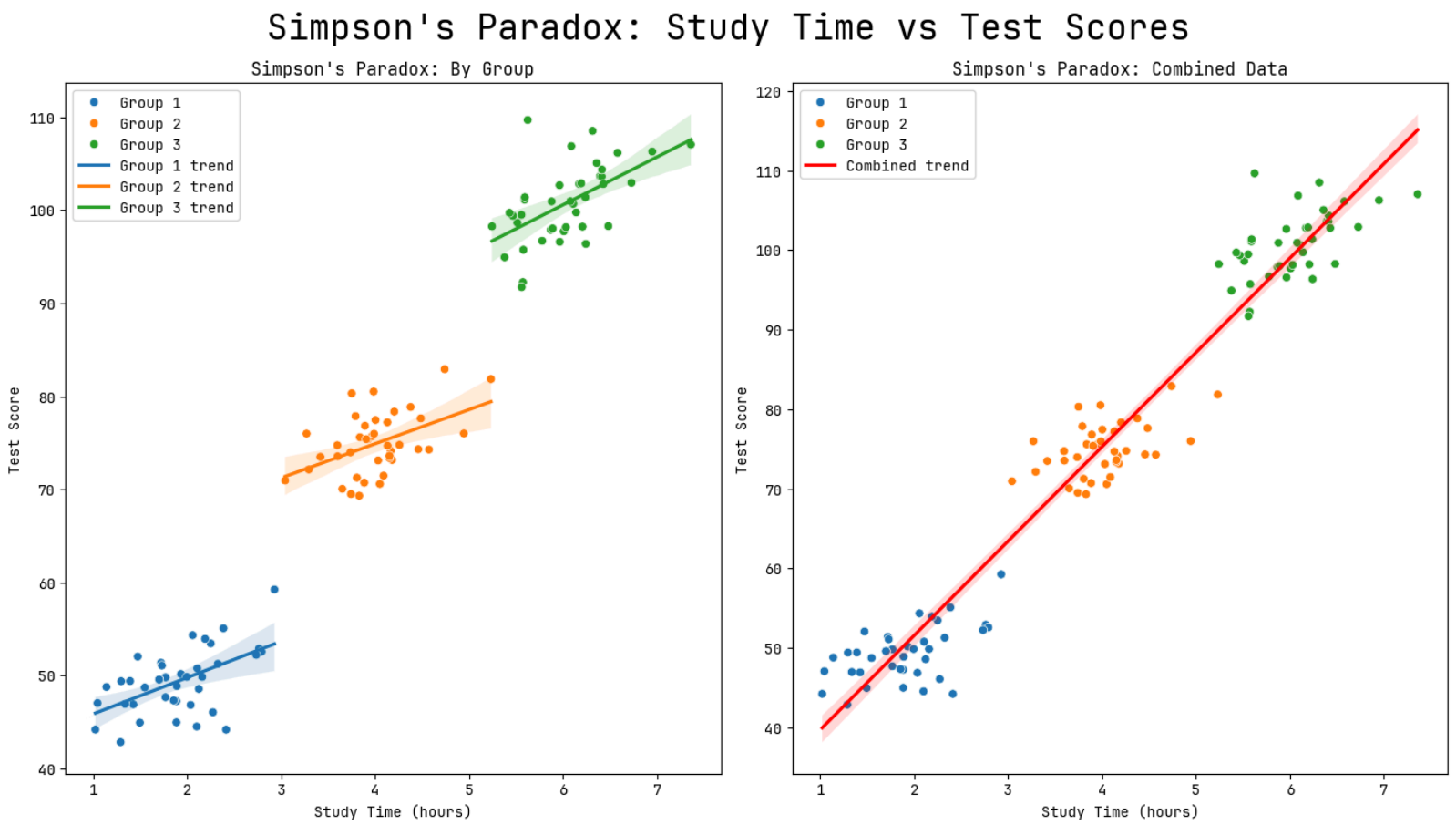
- Run a regression modelling on a dataset
- Understand the difference between correlation and causation
- Implement classification/regression trees

Core Concept: Considerations in Correlation vs Causation

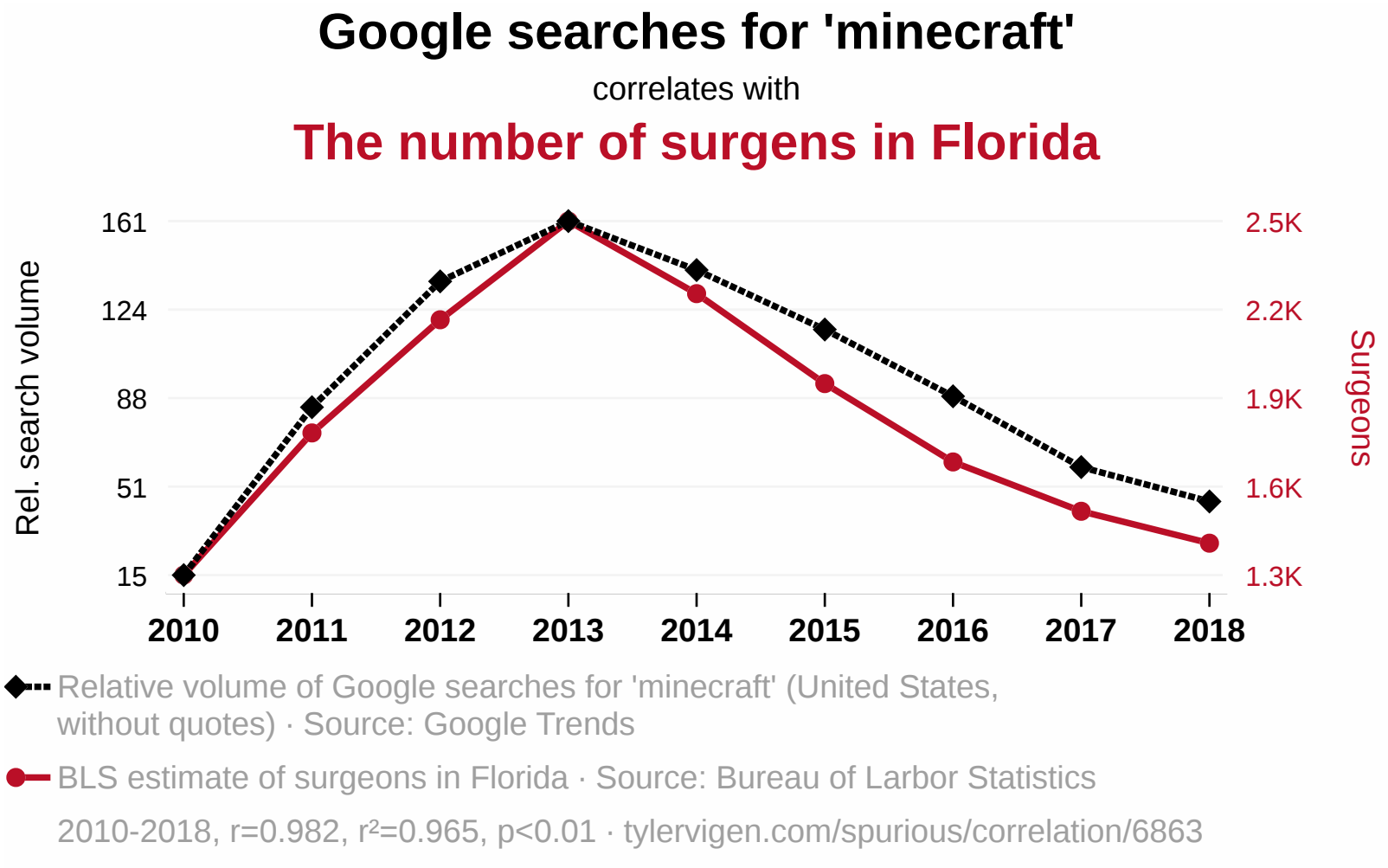
What correlation means

- strength and direction of linear relationship between two variables
- establishes association, not mechanism
- computationally simple, intuitive to interpret
- But...

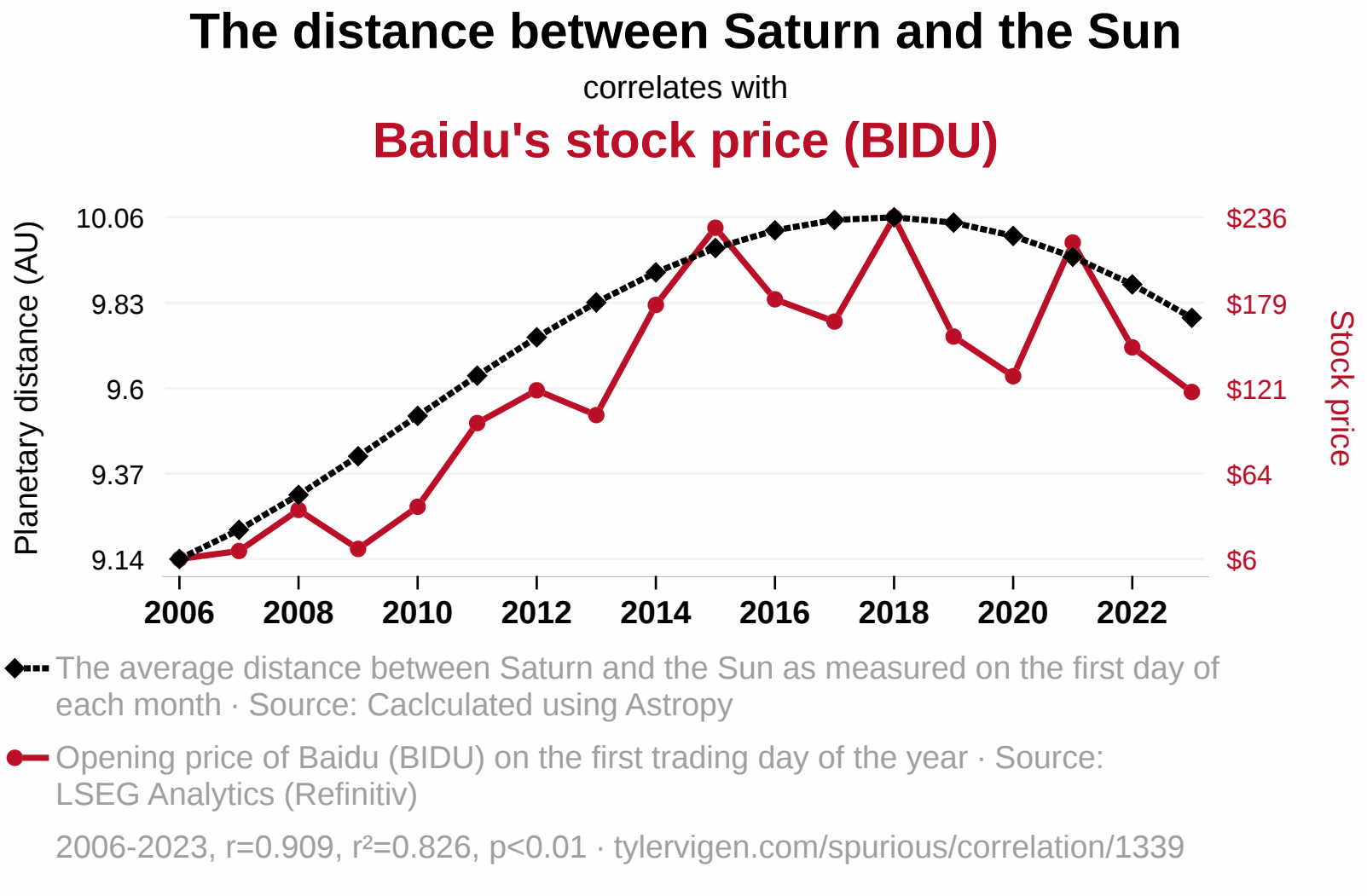
Correlation can be misleading



Correlation can be spurious



Correlation can be spurious



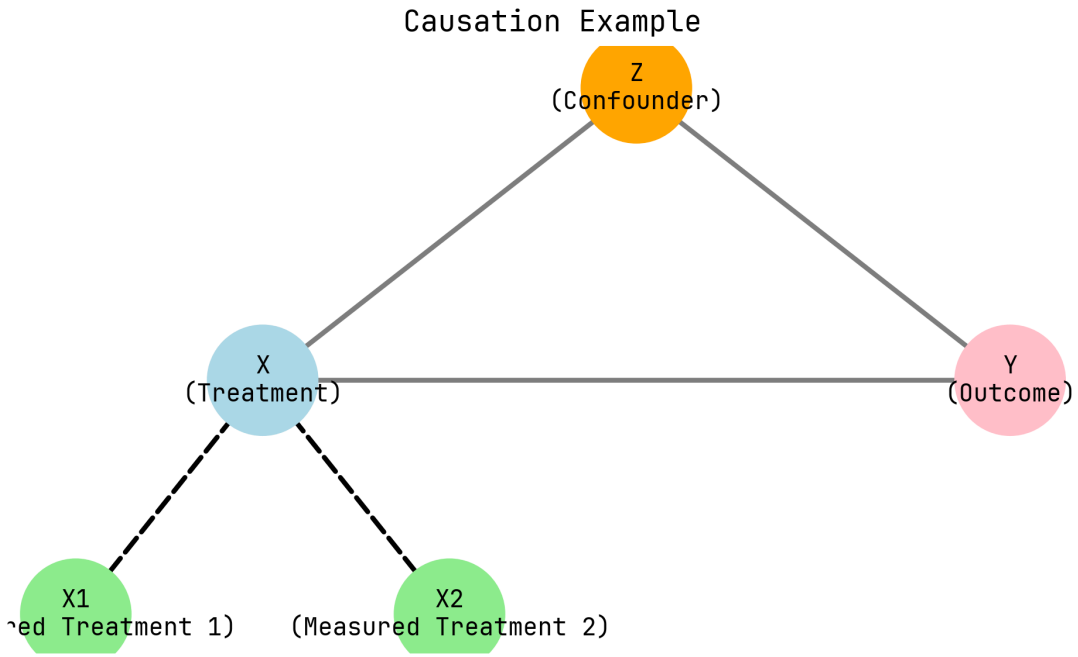
Causation is another beast

Correlation

- Statistical association
- Does not imply causation
- Can be established through observation

Causation

- Statistical association + mechanism
- Needs theoretical justification



So what?

Today's Agenda

Coding Tasks

- Activity 8.1: Regression modelling [1 ~ 1.5 hours]
- Activity 8.4: Classification trees [<1 hour]

Self-guided

- Activity 8.2: Fundamentals of regression modelling
- Activity 8.3: Regression modelling + Modelling **correlation** vs. **causation**