

Topic 1: Introduction and Basics of Data

Course Overview: What do we learn in STOR 155?

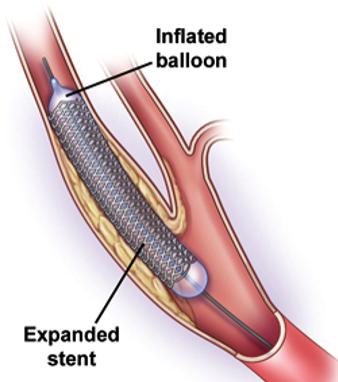
1. Identify a question or problem of interest

2. Collect relevant data

3. Analyze the data

4. Form a conclusion

Case Study: Do stents reduce the risk of stroke in at-risk patients?



Stents have been suggested as a tool for reducing the risk of stroke in certain at-risk patients. 451 patients agreed to participate in a study and were divided into two groups: **Control** (standard medical management) and **Treatment** (standard medical management + stent implanted). One year later, they were evaluated according to whether or not they had experienced a stroke.

Data for this study can be found in the Excel file **Stent365.xlsx** on Canvas (Module 1, Topic 1).

Summary of stent study outcomes for 451 patients:

	Stroke	No Event
Control	28	199
Treatment	45	179

Thoughts?

The Big Question: Do we have enough information to conclude that stents change the risk of stroke in this type of patient?

Coin flip example:

Stent example:

We use statistical tools to determine if the difference is so large that we should reject the notion that it was due to chance.

Data Sets: Vocabulary

Consider the spreadsheet `loan50.xlsx`

Each row of the sheet corresponds to...

Each column of the sheet corresponds to...

Two types of variables:

- Numerical

- Categorical

Classifying variables in the `loan50` dataset:

Relationships between Variables

The following data is from the spreadsheet `county.xlsx`

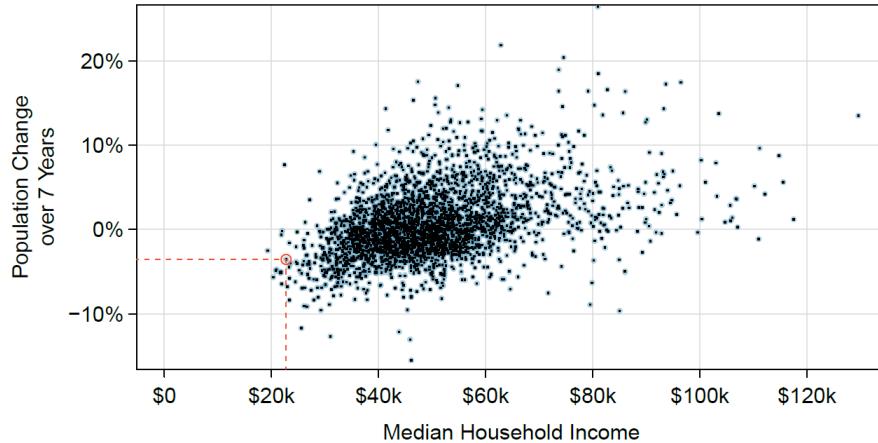
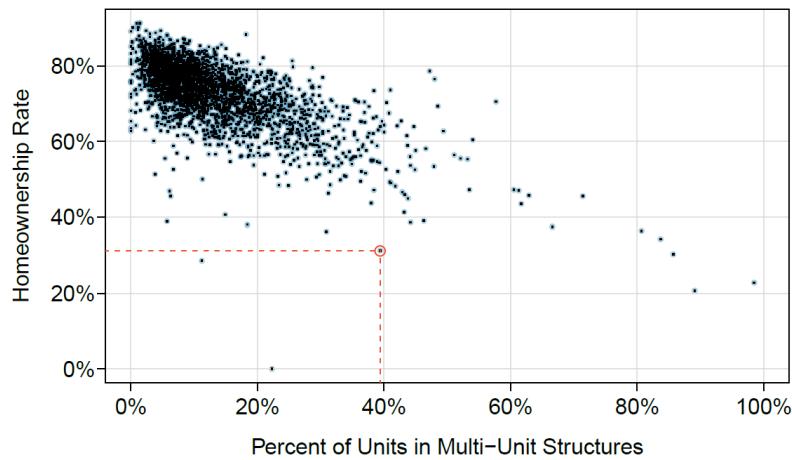


Figure 1.9: A scatterplot showing `pop_change` against `median_hh_income`. Owsley County of Kentucky, is highlighted, which lost 3.63% of its population from 2010 to 2017 and had median household income of \$22,736.

Comments on relationship between median household income and population change?



Comments on relationship between these two variables?

When examining the relationship between two variables, it is important to establish...

More examples:

- X: number of beers consumed
Y: blood alcohol content (BAC)
- X: type of treatment given (drug or placebo)
Y: outcome (cured or not)
- X: height of an individual
Y: weight of an individual
- X: SAT verbal score
Y: SAT math score

When two variables exhibit some form of relationship, we say they are...

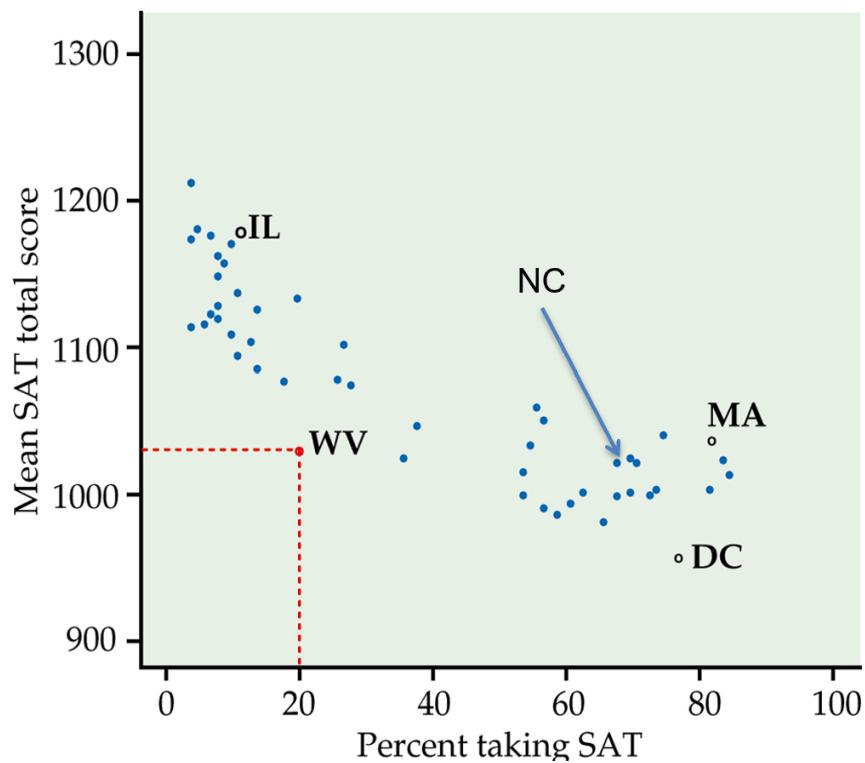
When two variables do not appear to be related, we say they are...

What we consider when describing the relationship between two variables:

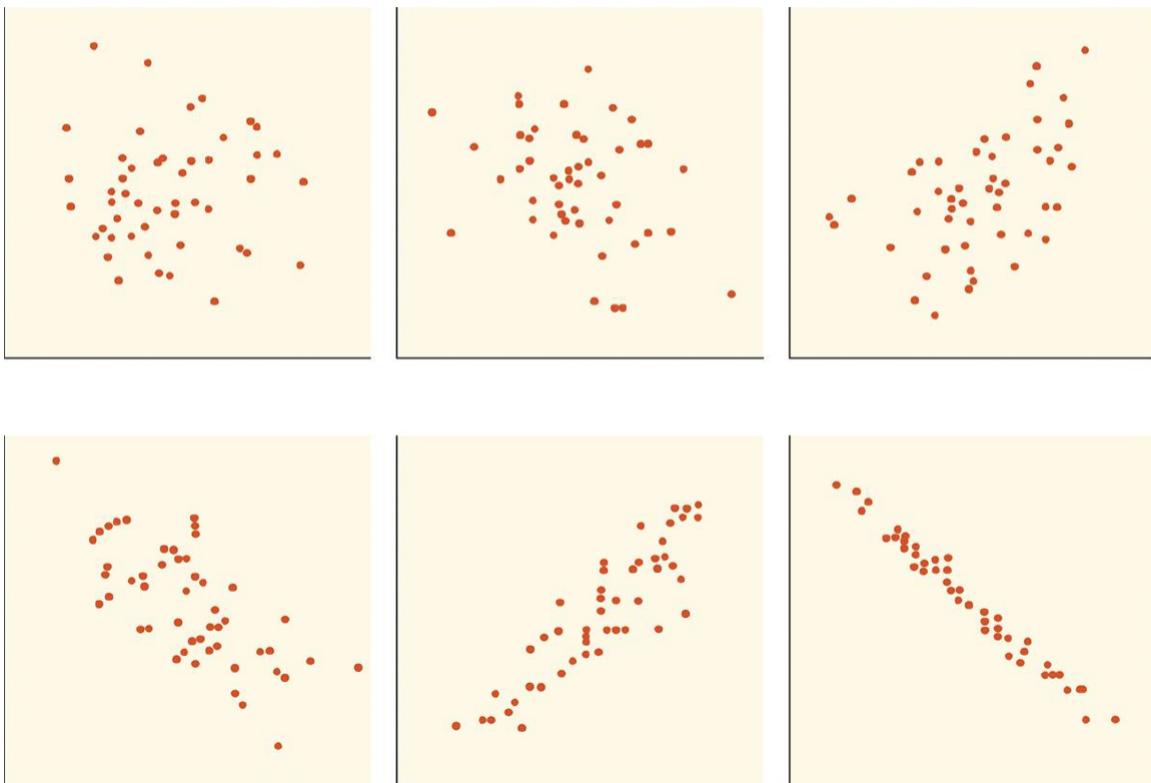
1.

2.

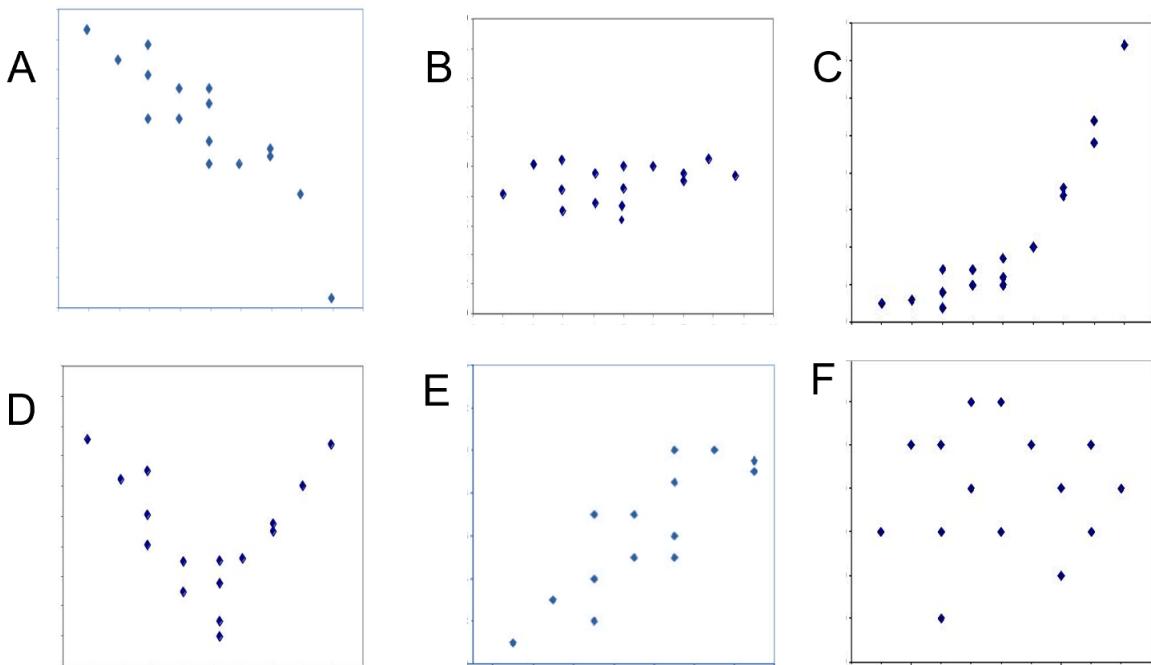
3.



Practice with describing direction and strength:



Practice with describing form:



Caution:

