Lecture 2: Data Basics

Data Sets: Vocabulary

Open 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
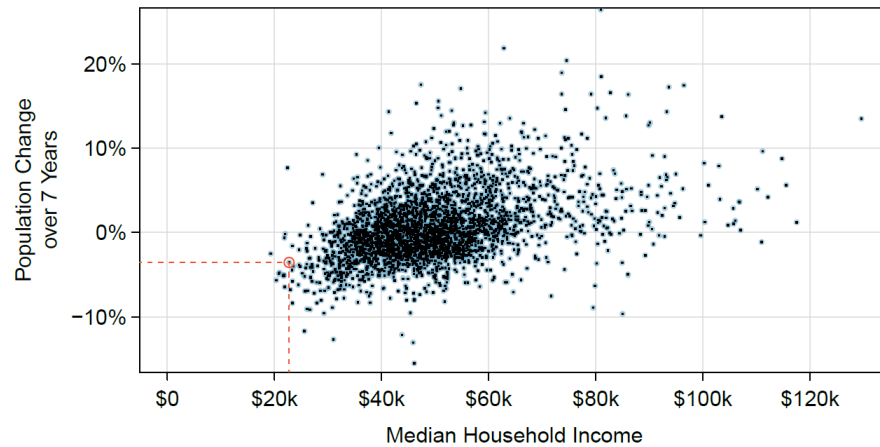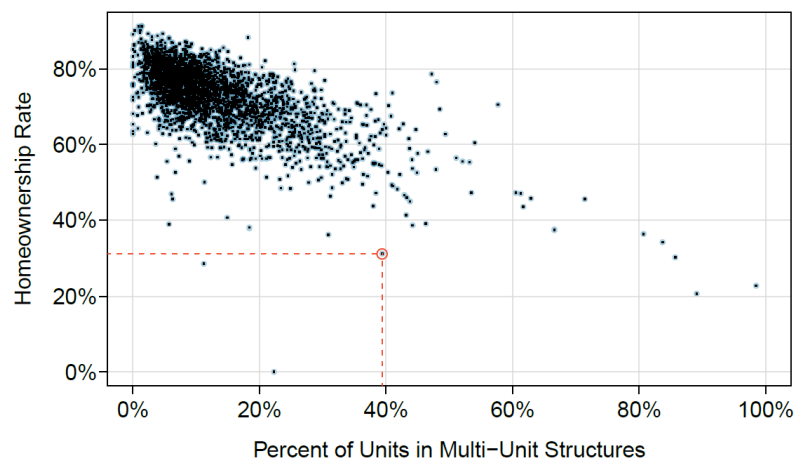
Open the spreadsheet `county.xlsx`

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

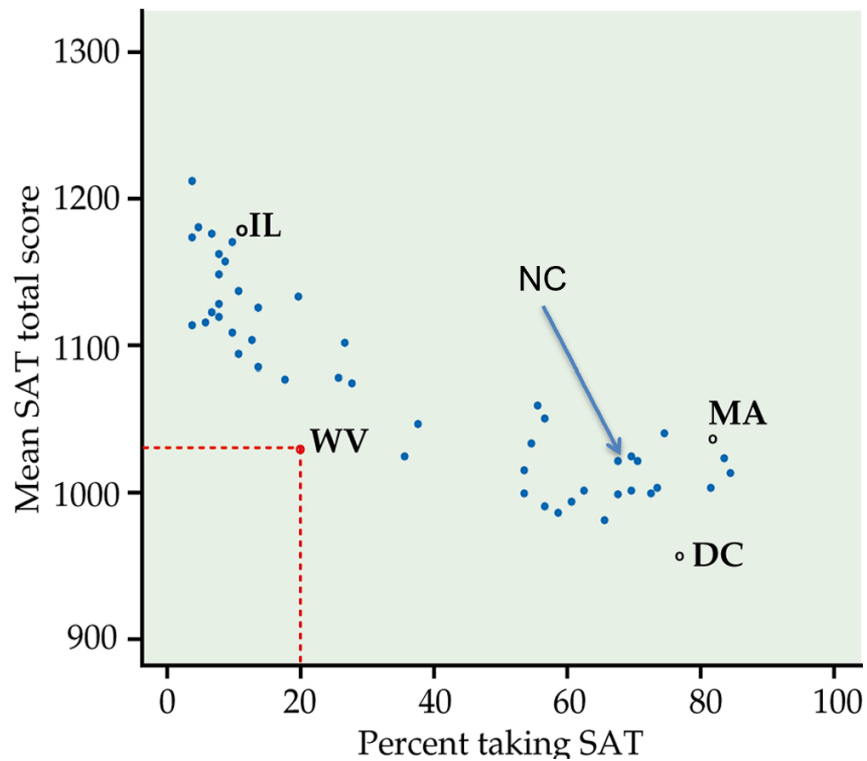When two variables exhibit some form of relationship, we say they are...

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

What we consider when describing the relationship between two variables:
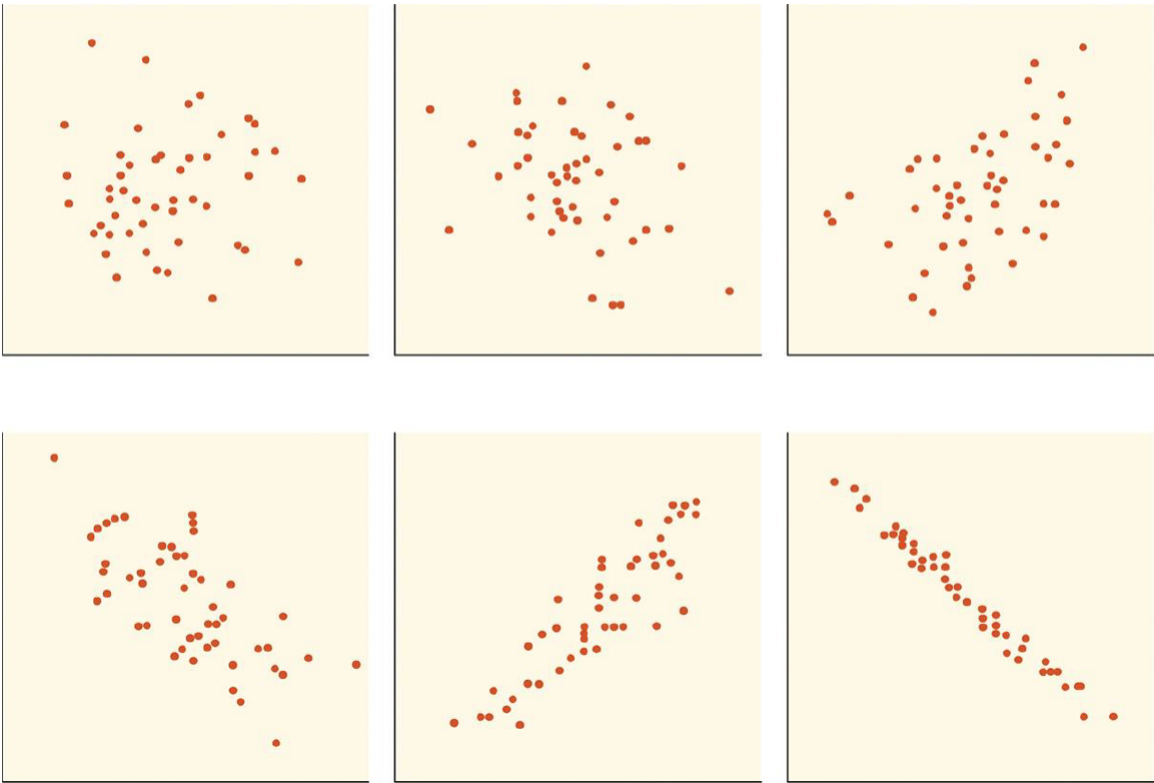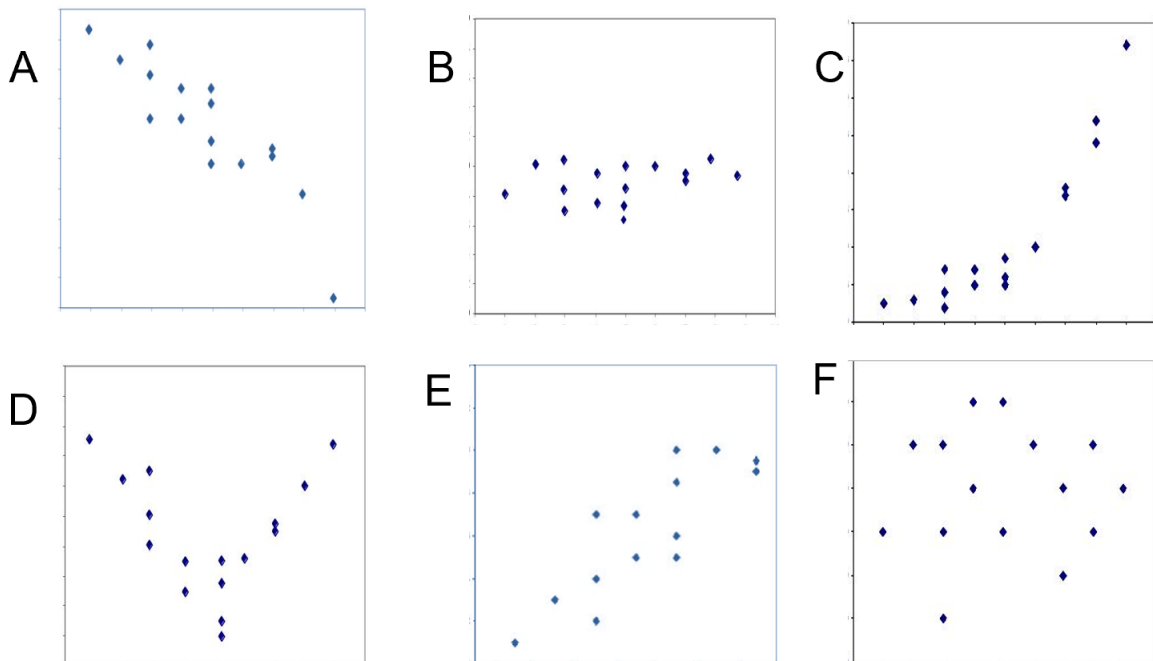
    1.

    2.

    3.

Practice with describing direction and strength:

Practice with describing form:

Caution: