Scatterplots

Scatterplots
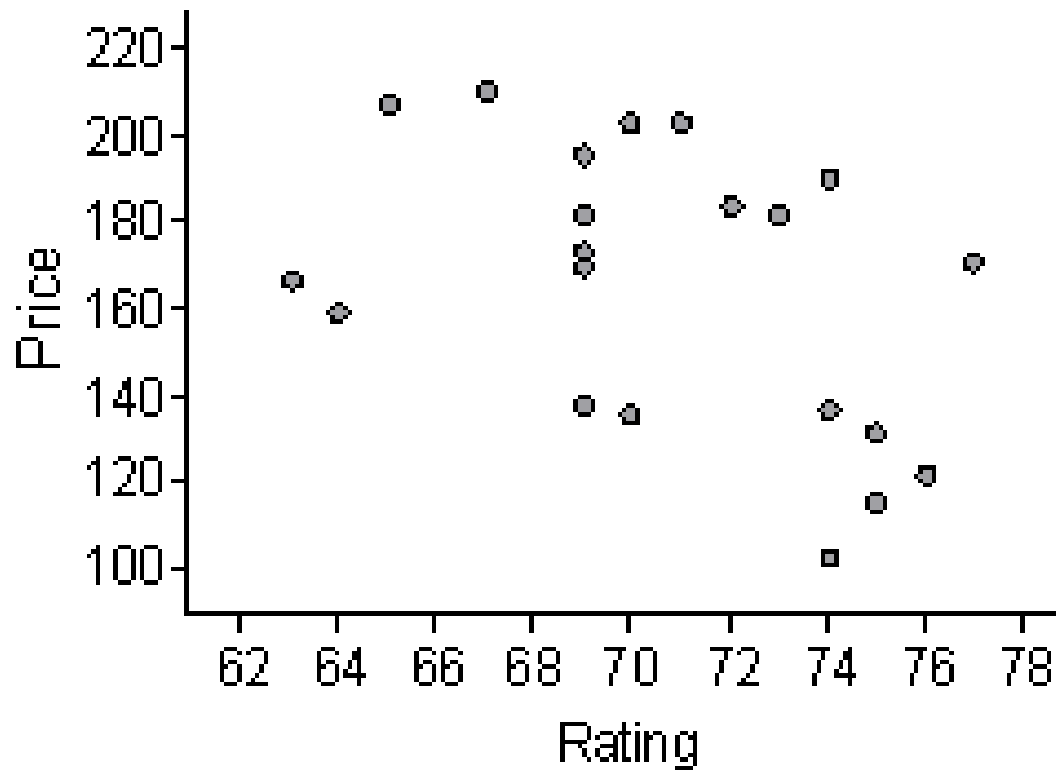
a.k.a. *Scatter Plots*
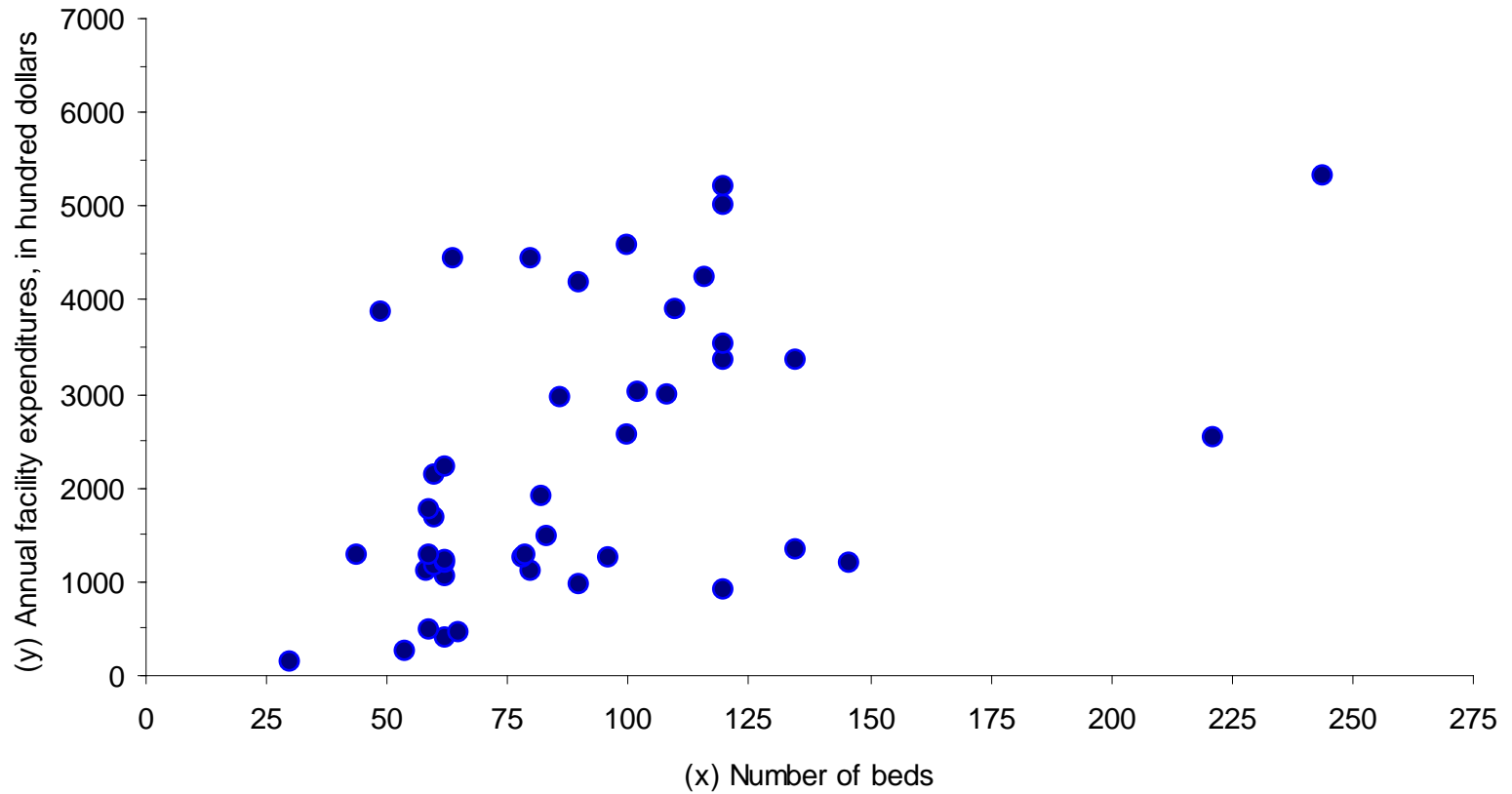
Scatter Diagrams
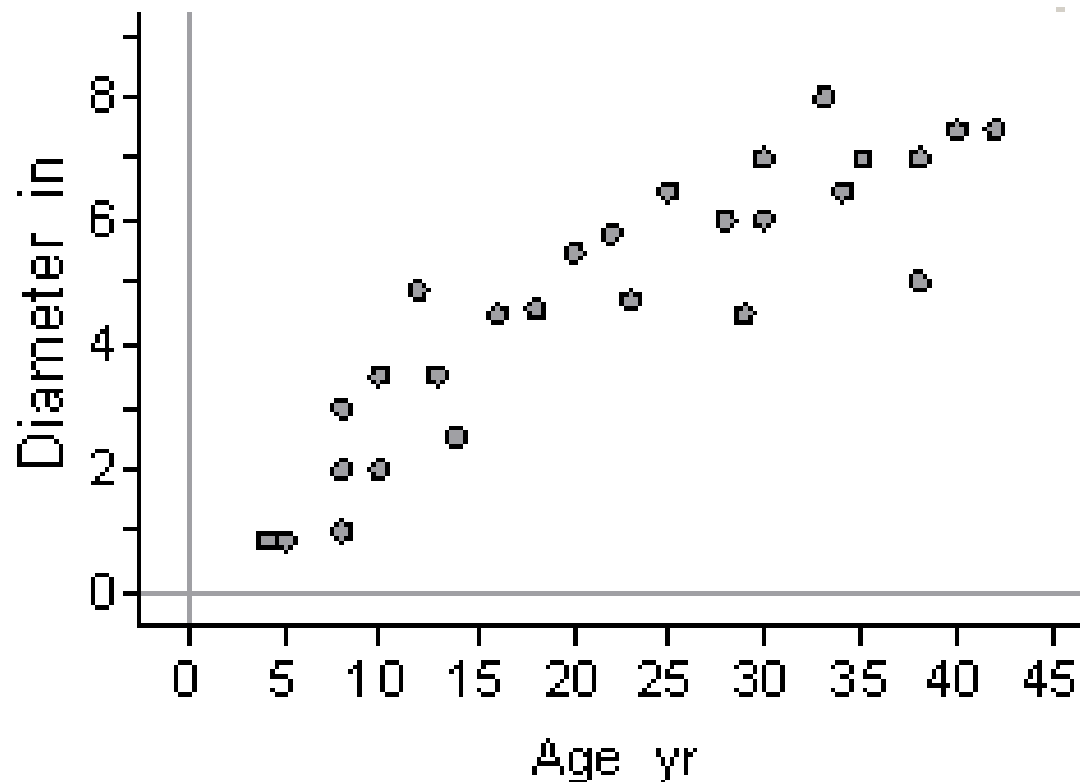
Scatterplot

Rating and Price of Eyeglasses

Scatter Plot

Scatter Diagram

Scatterplot
a.k.a. *Scatter Plot, Scatter Diagram*

- Shows the relationship between two quantitative variables

Scatterplot
a.k.a. *Scatter Plot, Scatter Diagram*

- **Shows the relationship between two quantitative variables**

  - **Explanatory variable – "cause"**
    - **Influences second variable**
    - **Independent variable**

Scatterplot
a.k.a. *Scatter Plot, Scatter Diagram*

- **Shows the relationship between two quantitative variables**

  - **Response variable – "effect"**
    - **Responds to changes in explanatory variable**
    - **Dependent variable**

Explanatory vs. Response Variable

- **For grades - amount of time spent studying for an exam vs. grade on the exam**

- **For a person - weight vs. height**

- **For crops (ex. Corn or Wheat) - amount of rainfall during the growing season vs. the yield for the crop**

- **For SAT scores – SAT Mathematics score vs. SAT Verbal score**

Explanatory vs. Response Variable

- For consumer product evaluation (ex. Eyeglasses) - rating for eyeglasses vs. price for eyeglasses
- For a nursing home - the annual facility expenditures vs. the number of beds
- For an oak tree - the age of the tree vs. the diameter of the tree
- For colleges/universities - the retention rate vs. the graduation rate

Creating a Scatterplot

- **Identify**
  - **Explanatory variable**
  - **Response Variable**

Creating a Scatterplot

- Set up
  - Axes – explanatory variable on horizontal axis; response variable on the vertical axis

Creating a Scatterplot

- **Set up**
  - **Scale (numbers with tick marks)**
    - **use range for *explanatory* variable to determine scale on *horizontal* axis**
    - **use range for *response* variable to determine scale on *vertical* axis**

Creating a Scatterplot

- Graph points
  - Ordered pairs are of the form (x, y) for which
    - x is the value of the explanatory variable which corresponds to the value y for the response variable

Creating a Scatterplot

- **Graph points**
  - **Ordered pairs are of the form (x, y) for which**
    - **x is the value of the explanatory variable which corresponds to the value y for the response variable**

CAUTION: Correspondence of values matters!! You cannot reorder individual columns in the data table!!!

Describing Pattern in Scatter Plot

- **Identify**
    - **Variables and Cases**
        - **units of measure**
        - **range of each variable**
- **Describe the overall shape**
    - **Linearity**
    - **Clusters**
    - **Outliers**

Describing Pattern in Scatter Plot

- **Identify**
  - **Variables and Cases**
    - **units of measure**
    - **range of each variable**
- **Describe the overall shape**
  - **Linearity – points scattered (close to; far from) about a possible line**

Describing Pattern in Scatter Plot

- **Identify**
  - **Variables and Cases**
    - **units of measure**
    - **range of each variable**
- **Describe the overall shape**
  - **Clusters – groups of points**

Describing Pattern in Scatter Plot

- **Identify**
  - **Variables and Cases**
    - **units of measure**
    - **range of each variable**
- Describe the overall shape
  - Outliers – points separated (away from) other points in the graph

Describing Pattern in Scatter Plot

- Positive trend
  - as the value of the explanatory variable increases, the value of the response variable increases
- Negative trend
  - as the value of explanatory variable increases, the value of the response variable decreases

Describing Pattern in Scatter Plot

- For explanatory variable x and response variable y,
  - Positive trend
    - as x increases, y increases
  - Negative trend
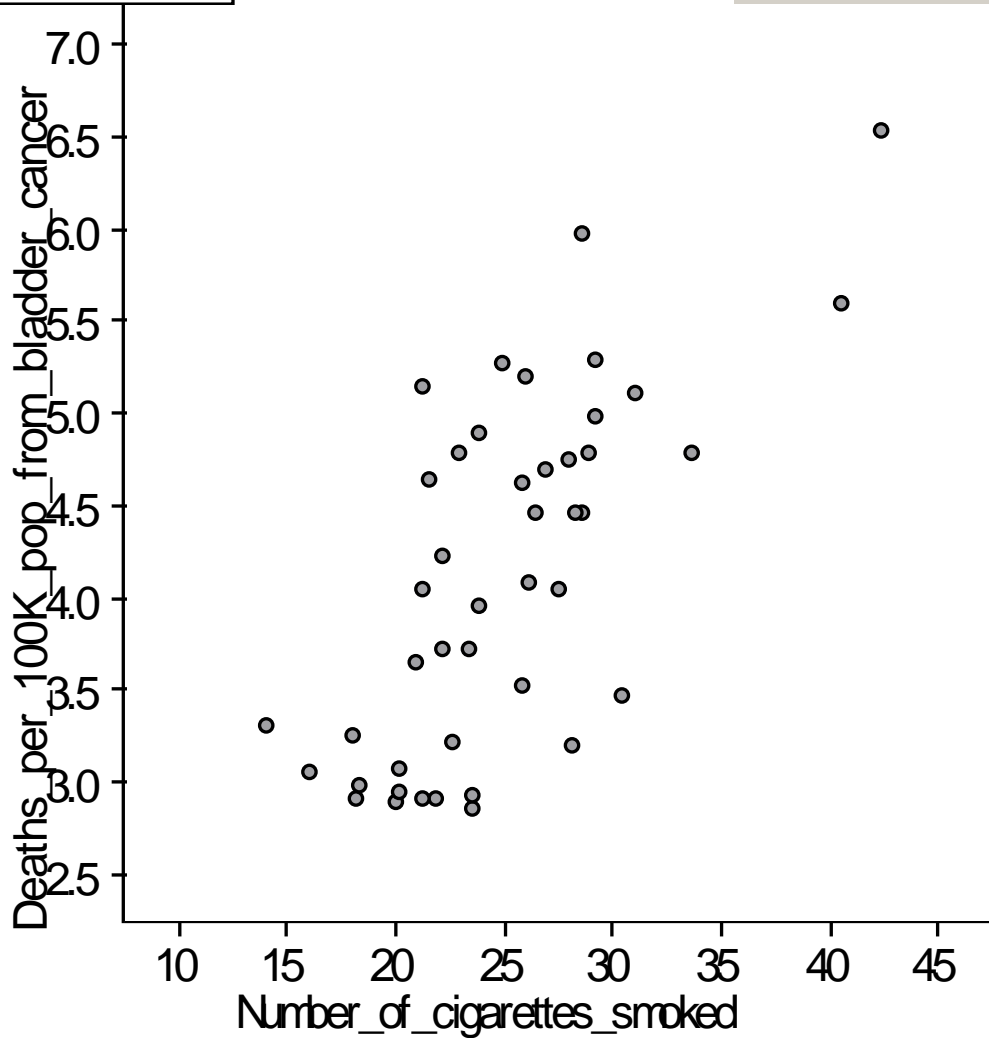   - as x increases, y decreases

Describing Pattern in Scatter Plot

- **Describe the strength of linear trend**
  - **Strong – points closely grouped/gathered around a possible line**
  - **Weak – points scattered/grouped farther from a possible line**
  - **Varies in strength**
    - **Heteroscedasticity – points fan out at one end**
  - **Constant strength – points grouped relatively same manner around possible line**

Describing Pattern in Scatter Plot

- Plausible explanations for pattern?
  - Consider variables
  - Consider context of data
- Reasonable to conclude one variable is dependent on another?
  - One affects another?
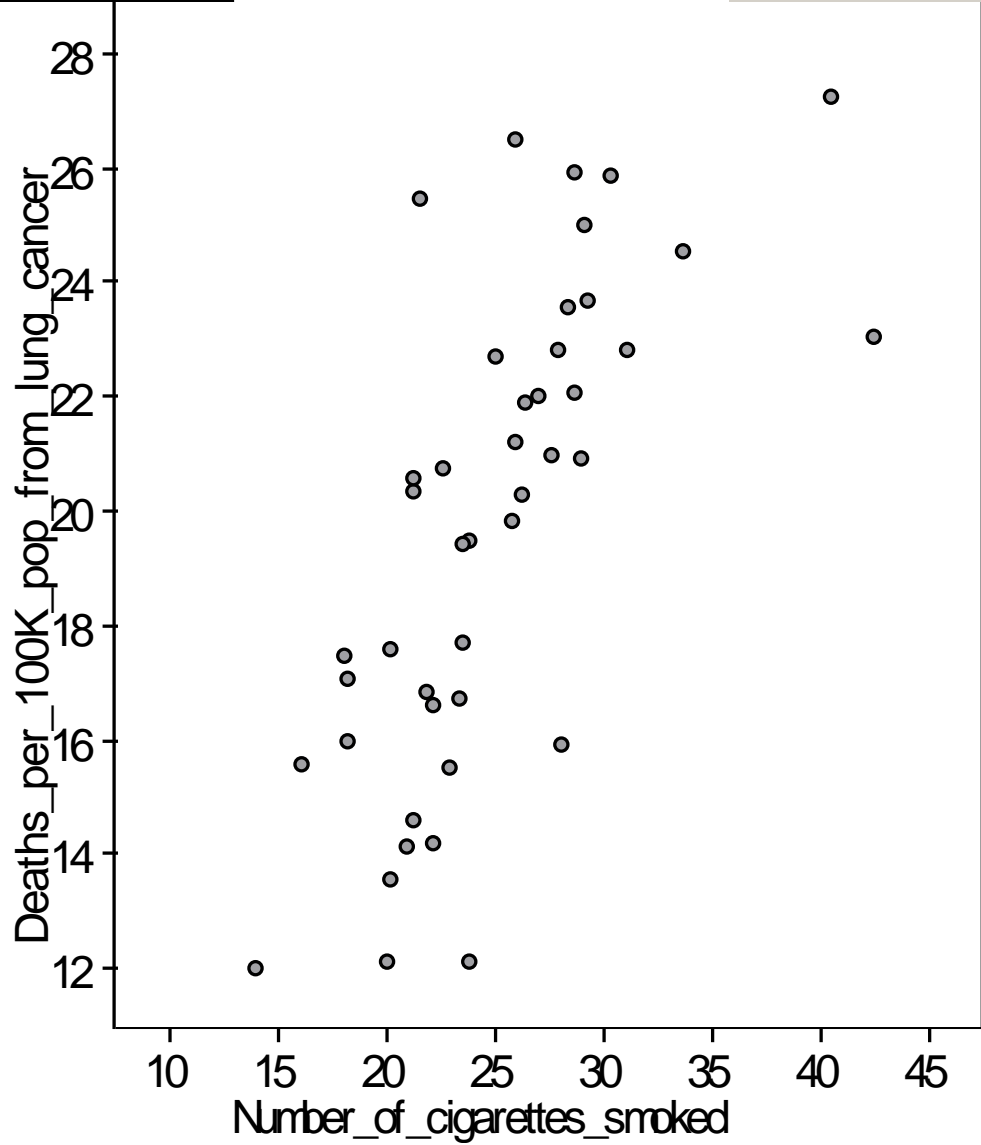- Possible for third variable to affect both?
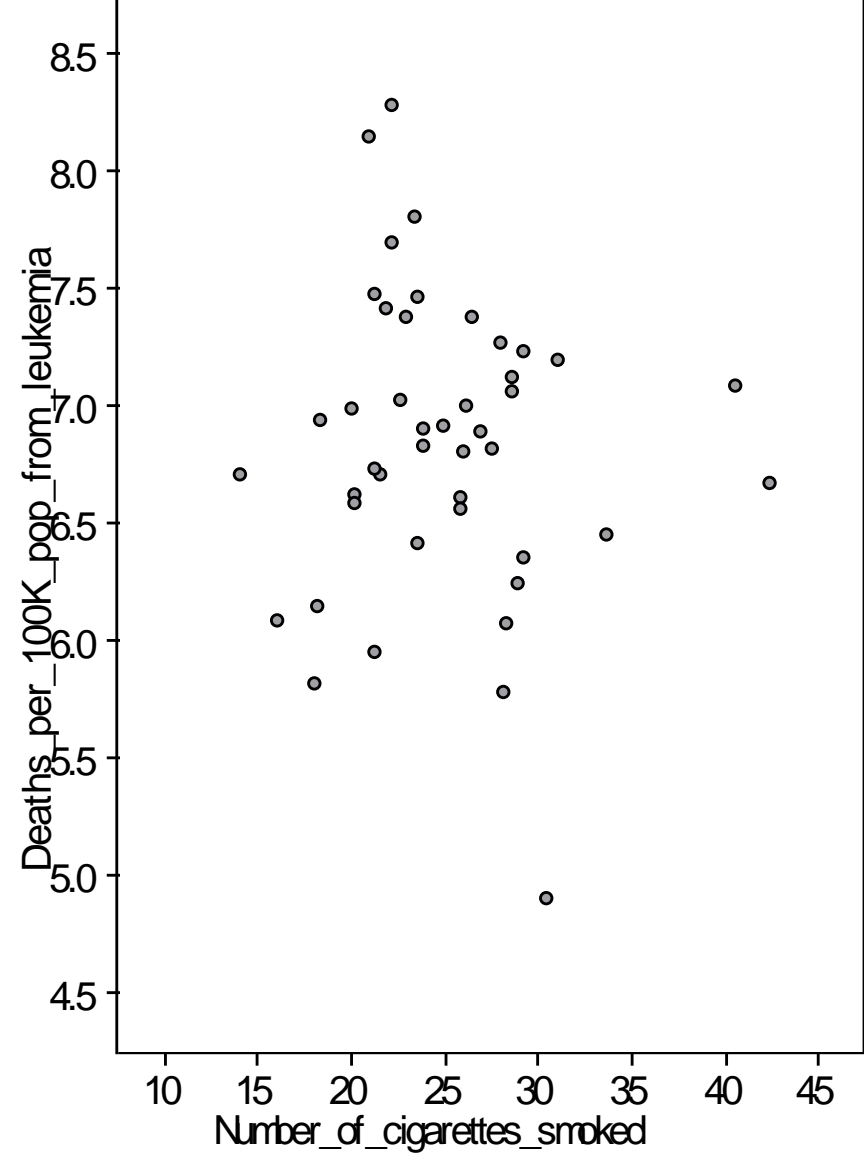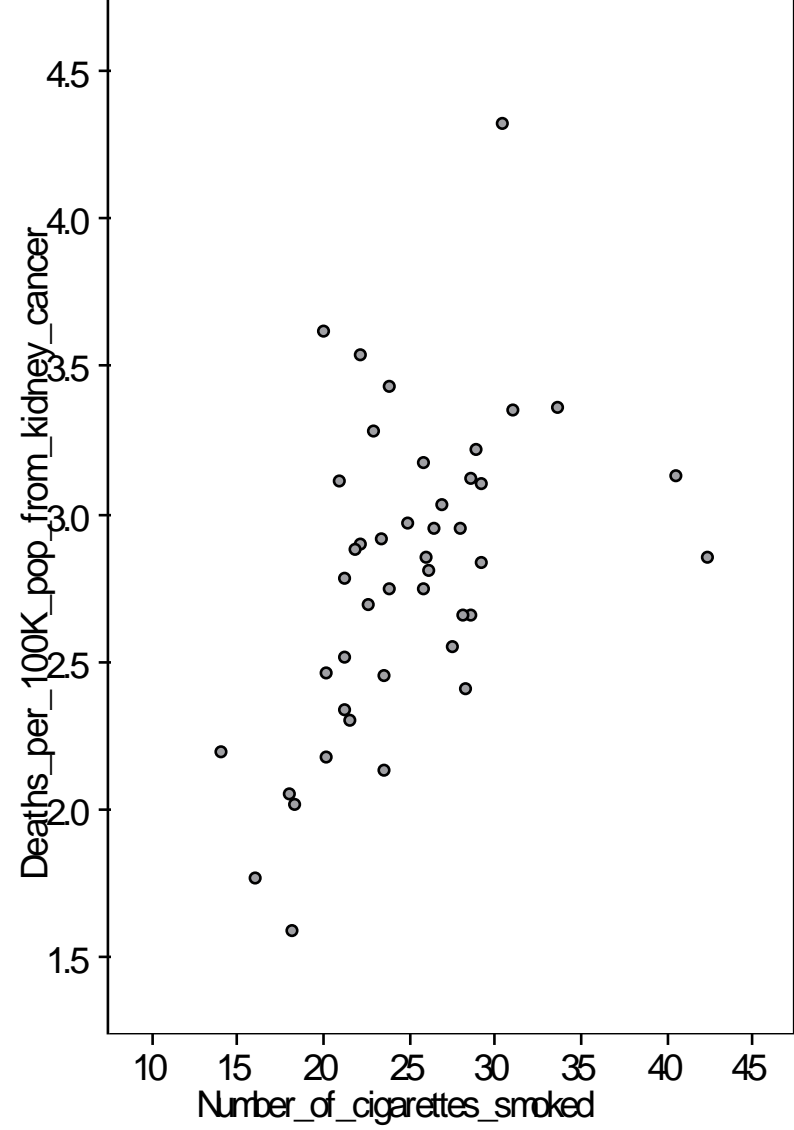 - Lurking variable

cigcancerdat

Scatter Plot ▼

Deaths_per_100K_pop_from_bladder_cancer

Number_of_cigarettes_smoked

Compare the scatter plots and determine which displays the data with the strongest linear relation.