Graphical Descriptive Statistics
Part II: Depicting the relationship between two variables

Eco 2470: Economic Statistics

Fall, 2019. (Chapters 2-3)

(last updated Sep 11, 2019)

Describing the Relationship between Two Nominal Variables

To describe the relationship between two nominal variables, we must remember that we are permitted only to determine the frequency of the values. As a first step we need to produce a cross-classification table, which lists the frequency of each combination of the values of the two variables

Example 2.4 Newspaper Readership Survey

A survey was conducted to analyze the relationship between newspapers read and occupation. A sample of newspaper readers was asked to report which newspaper they read: *Globe and Mail* (1) *Post* (2), *Star* (3), *Sun* (4), and to indicate whether they were blue-collar worker (1), white-collar worker (2), or professional (3).

Example 2.4

Reader	Occupation	Newspaper
1	2	2
2	1	4
3	2	1
.	.	.
.	.	.
352	3	2
353	1	3
354	2	3

Determine whether the two nominal variables are related.

Cross-Classification Table of Frequencies

	Newspaper				
Occupation	*G&M*	*Post*	*Star*	*Sun*	Total
Blue collar	27	18	38	37	120
White collar	29	43	21	15	108
Professional	33	51	22	20	126
Total	89	112	81	72	354

Row Relative Frequencies

	Newspaper				
Occupation	*G&M*	*Post*	*Star*	*Sun*	Total
Blue collar	.23	.15	.32	.31	1.00
White collar	.27	.40	.19	.14	1.00
Professional	.26	.40	.17	.16	1.00
Total	.25	.32	.23	.20	1.00

Graphing the Relationship Between Two **Interval** Variables…

We are frequently interested in how two interval variables are related.

To explore this relationship, we employ a ***scatter diagram***, which plots two variables against one another.

The ***independent*** variable is labeled X and is usually placed on the horizontal axis, while the other, ***dependent*** variable, Y, is mapped to the vertical axis.

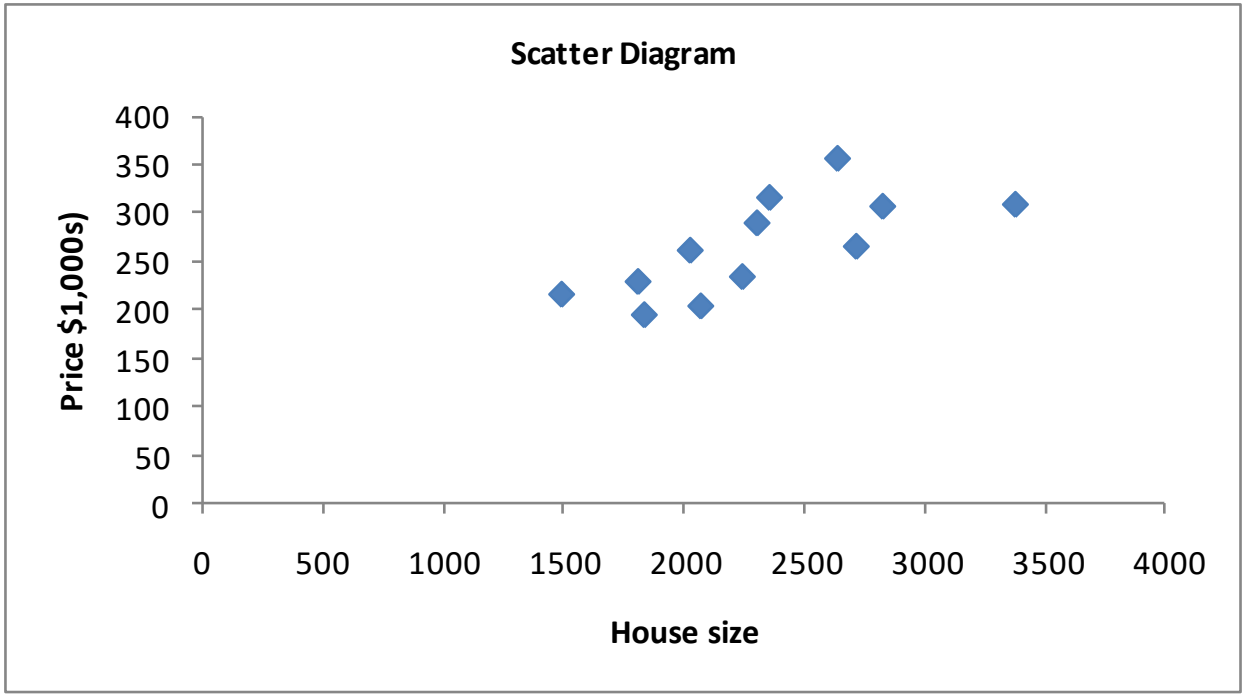
Example 3.7

A real estate agent wanted to know to what extent the selling price of a home is related to its size. To acquire this information he took a sample of 12 homes that had recently sold, recording the price in thousands of dollars and the size in hundreds of square feet as follows:

Size	2354	1807	2637	2024	2241	1489	3377	2825	2302	2068	2715	1833
Price	315	229	355	261	234	216	308	306	289	204	265	195

Example 3.7

It appears that in fact there is a relationship, that is, the greater the house size the greater the selling price…

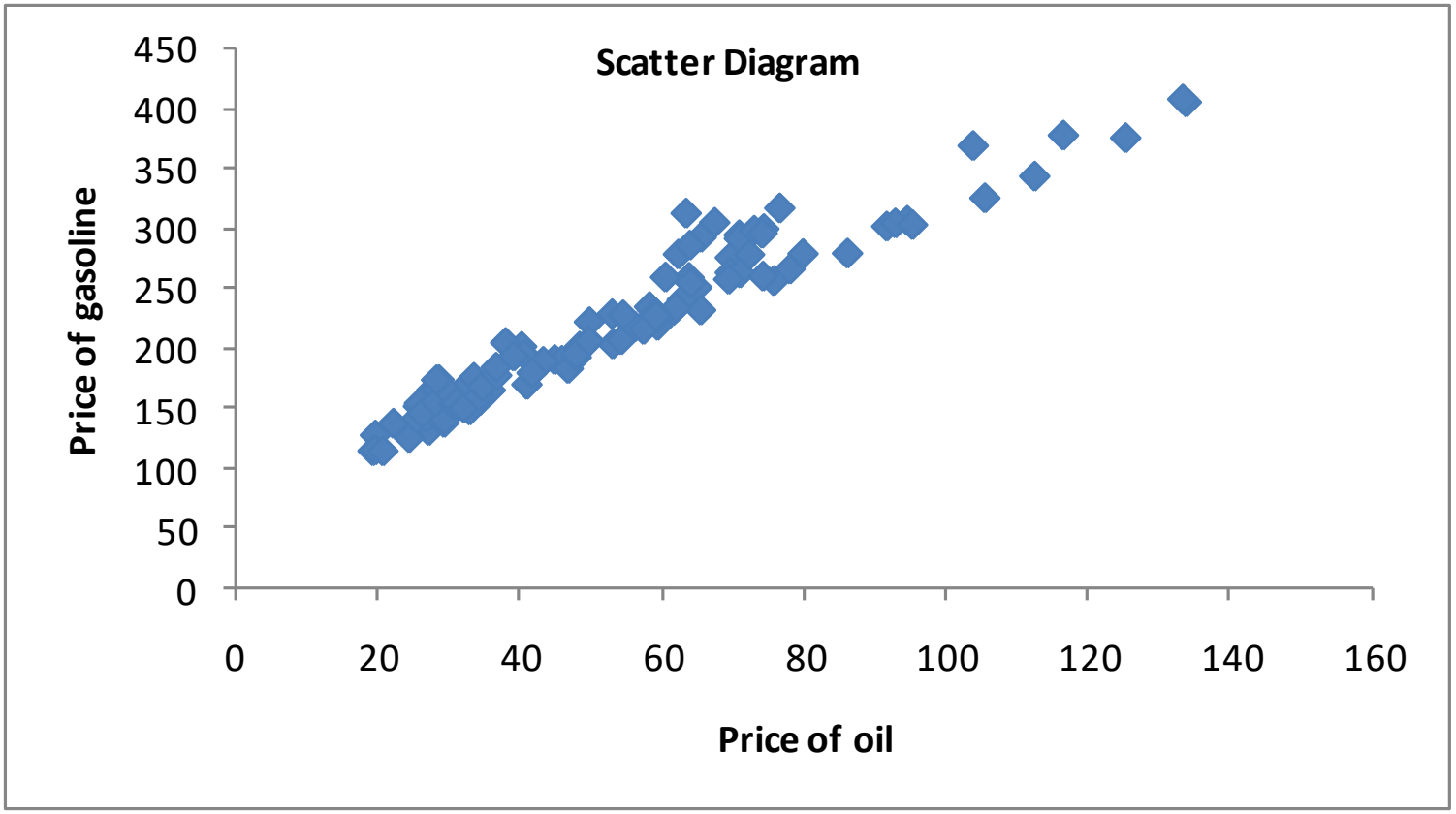


Chapter 3-Opening Example

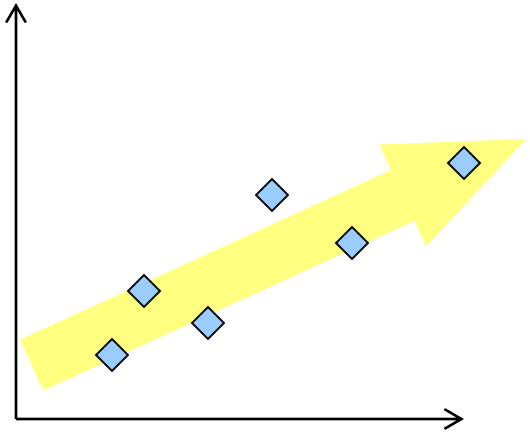
Were Oil Companies Gouging Customers 2000-2009:

Many drivers complained that the oil companies were guilty of price gouging. That is, they believed that when the price of oil increased the price of gas also increased, but when the price of oil decreased, the decrease in the price of gasoline seemed to lag behind. Let's look at the relationship between oil and gas prices.

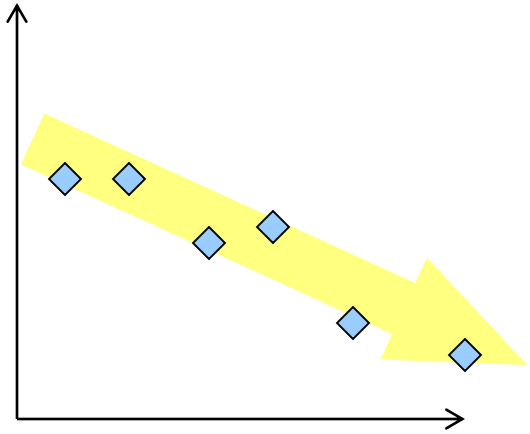
Chapter 3-Opening Example



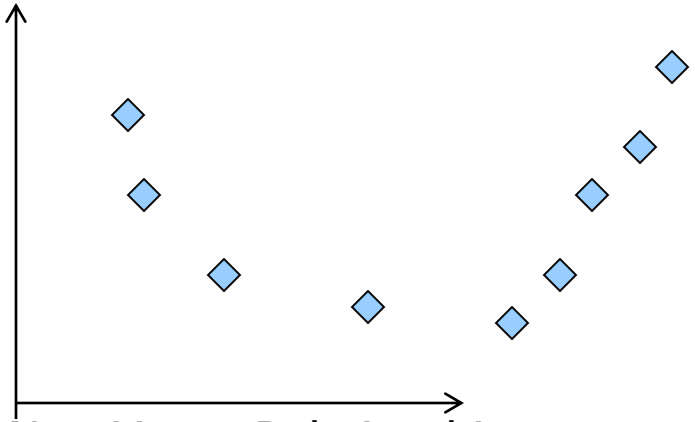
Patterns of Scatter Diagrams...



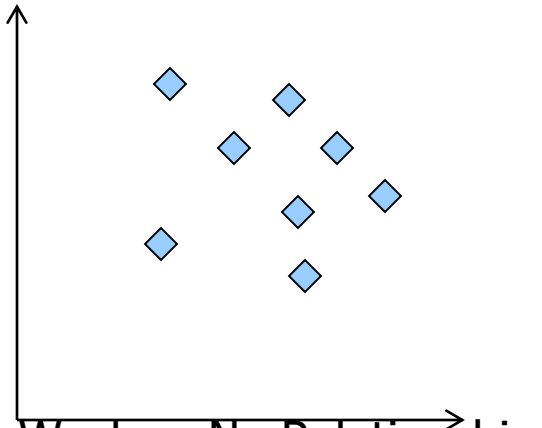
Positive Linear Relationship



Negative Linear Relationship



Non-Linear Relationship



Weak or No Relationship

Interpretation of Scatter Diagrams: Relationship versus causation

- Oil prices and gas prices are strongly positively related
- So increases in oil prices are often (not always!) associated with increases?
- Do oil price increases cause gas price increases?
- Do gas prices increases cause oil price increases?
- Is there something else driving both price increases?
- Or is it a combination of the above factors?

Direct Causation

Direct Causation: When changes in the independent variable (X-variable) cause changes in the dependent variable (Y-variable).

This can create a relationship between the two variables.

X-Variable: ➔ Y-Variable

Example: Education ➔ Earnings

Indirect Causation

Indirect Causation: When changes in the X and Y-Variables are both caused by a third variable (say Z)

- You will observe a relationship in the scatter diagram even if X does not impact Y and Y does not impact X.

- Be careful interpreting such a relationship.

X-Variable: Y-Variable

Z-Variable

Earnings and Education Revisited

- You were not selected into university at random.
- You chose to go UofG.
- And, UofG chose you too!
- Presumably this had something to do with your ABILITY

  ABILITY  EDUCATION
- If you perform well at your job, you may get a raise
- This also has something to do with your ABILITY.

- ABILITY  EARNINGS

Earnings and Education (continued)

Relationship between earnings & education may be due to: a

Direct Effect: EDUCATION → EARNINGS

Indirect Effect EDUCATION → ABILITY → EARNINGS

Or **BOTH** EDUCATION → EARNINGS, EDUCATION → ABILITY → EARNINGS

Reverse Causality

- **Reverse Causality:** When changes in the dependent variable (Y-Variable) cause changes in the independent variable (X-Variable)
- Put another way, the causation goes in the opposite direction as expected.
- You see a relationship in the scatter diagram, but the interpretation is opposite to what you would think.

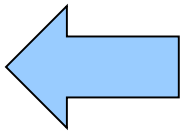
Example: Police and Crime (Steve Levitt, author of Freakonomics)

- X-Variable: POLICE (number of police)
- Y-Vabiable: CRIME (number of reported crimes)
- Direct Causation: Hiring more police should reduce crime
- Suggests: Negative relationship between POLICE & CRIME

Police and Crime continued

- Puzzling Finding: The data show a positive relationship between police and crime (in the US): U.S. cities with larger police forces tend to have higher crime rates

- Levitt's Explanation: Cities with higher crime rates hire more police

Reverse Causation: POLICE  CRIME

Bottom Line: Due to indirect and reverse causation care is needed when interpreting relations between variables

Summary II…

	Interval Data	Nominal Data
Single Set of Data	Histogram	Frequency and Relative Frequency Tables, Bar and Pie Charts
Relationship Between Two Variables	Scatter Diagram	Cross-classification Table, Bar Charts