

# Chapter Six: Scatterplots, Association, and Correlation

## Review: Explanatory and Response Variables

So remember from previous discussions, when we are looking at a relationship between two variables we assign them the roles of \_\_\_\_\_ and \_\_\_\_\_ variable. In Chapter Four we looked at a relationship between a categorical variable and a quantitative variable. Now in chapter six, we will look at the relationship between two \_\_\_\_\_ variables.

For two quantitative variables we:

- 
- 

So now, let's take a moment to run through some examples. Label which variable is the explanatory variable and which variable is the response in the following scenarios.

1. ACT composite score and freshman GPA

- Explanatory Variable:
- Response Variable:

2. Physical Response times and sleep deprivation

- Explanatory Variable:
- Response Variable:

3. Height of Corn stalk and fertilizer type

- Explanatory Variable:
- Response Variable:

4. Number of cigarettes smoked per day and life expectancy

- Explanatory Variable:
- Response Variable:

5. Number of headaches and number of glasses of water consumed

- Explanatory Variable:
- Response Variable:

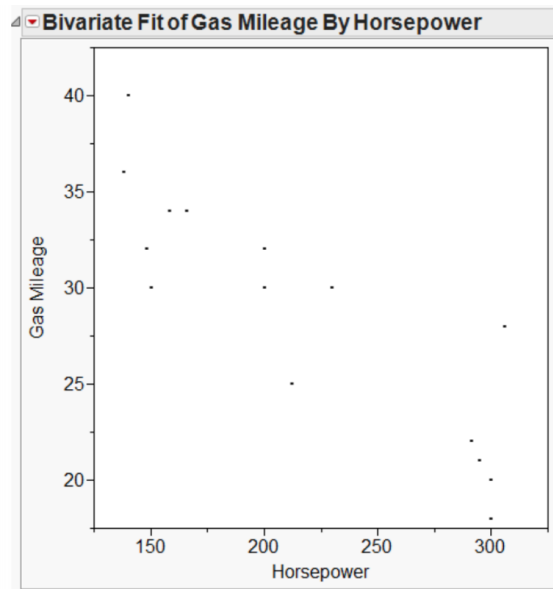
6. Hours of sleep at night and dollars spent on coffee

- Explanatory Variable:
- Response Variable:

## Scatterplots

A \_\_\_\_\_ is used to display the relationship between two quantitative variables. The explanatory variable is put on the \_\_\_\_\_ and the response variable is on the \_\_\_\_\_.

**Example** Suppose we have data on 15 different vehicle models from 2007. For this example, we are interested in studying horsepower and gas mileage.



To interpret this scatterplot, we consider the form, strength, direction/association, and outliers.

### Form

•

•

•

**Strength**

- 

- 

- 

**Direction/Association**

- 

- 

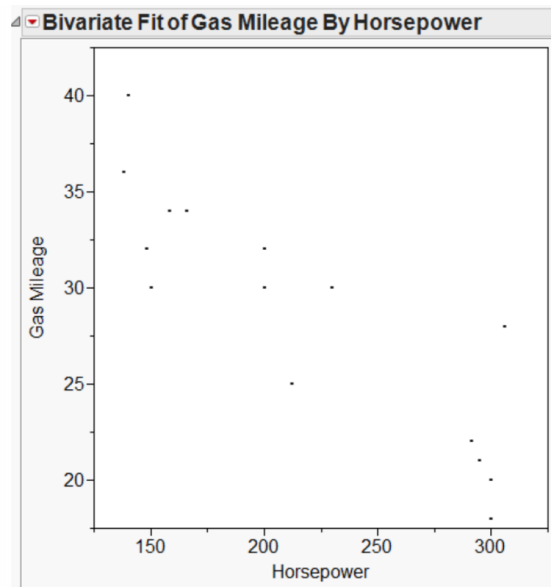
**Outliers**

- 

- 

-

So let's return to our example:



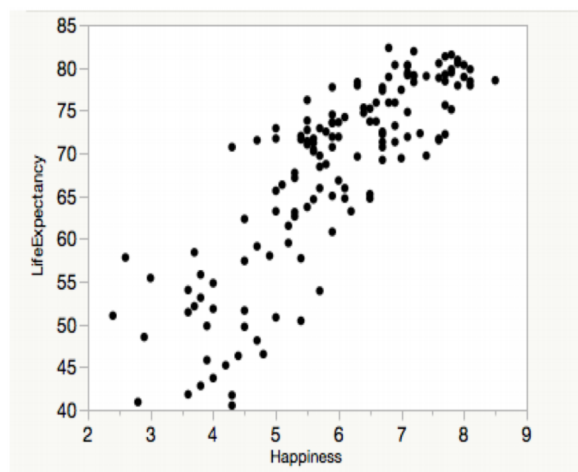
How would you describe this scatterplot using the categories we just discussed?

**Example** *The Happy Planet Index was a presentation given by Nic Marks in a 2010 TED talk. One of the key points of the talk was that we should be measuring the well-being of the people in countries, rather than measuring economic information on the countries. One variable that captures the well-being of people is called the Happiness variable. According to the Happy Planet Index: 2012 Report, 'The Happy Planet Index (HPI) is an efficiency measure which captures the degree to which long and happy lives are achieved per unit of environmental impact.' Large values of the Happiness variable indicate greater happiness, health, and well-being of the country's citizens. What are the 5 Ws of this dataset?*

What is the explanatory variable and what is the response variable?

What kind of relationship do you think exists between these two variables?

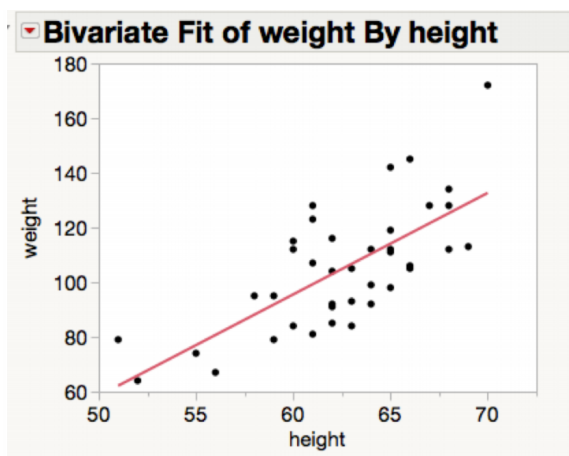
Describe the form, strength, and direction of the scatterplot provided. Are there any outliers? Any points that deviate from that form?



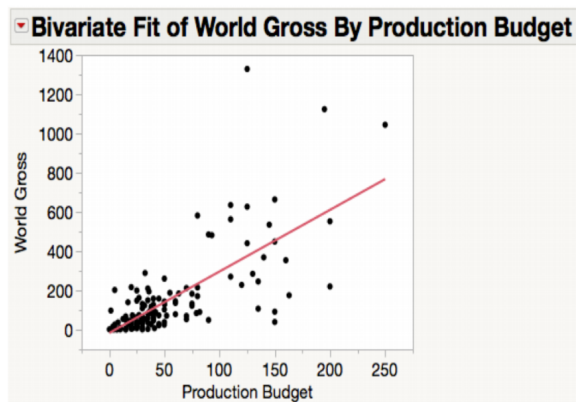
For the following scatterplots, answer these two questions:

1. What is the explanatory variable
2. What is the response variable
3. Describe the form, strength, and direction of the scatterplot provided. Are there any outliers?
4. Any points that deviate from that form?

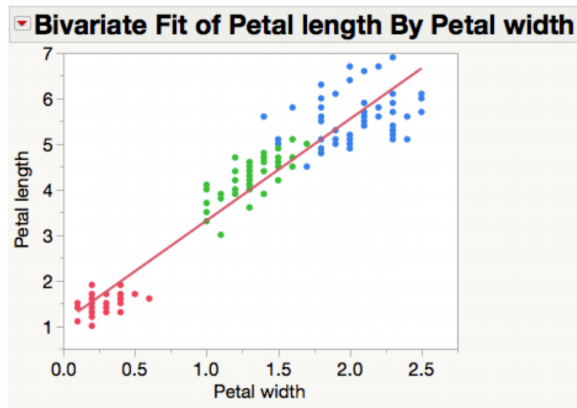
### Example



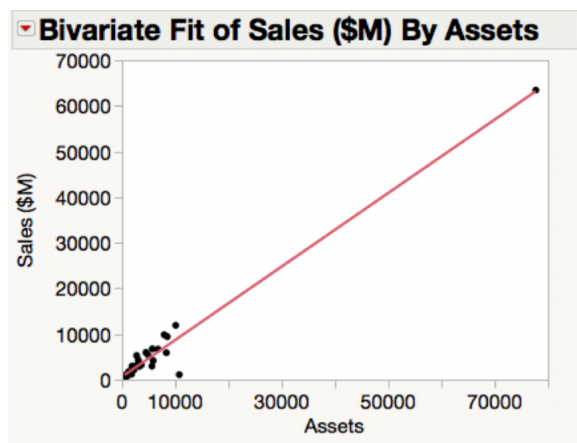
### Example



### Example



### Example



### Example

*Notice how this is the same data as the previous example, just the outlier is removed*

