

Section 2.1

Chapter 2 - Relationships b/t Variables

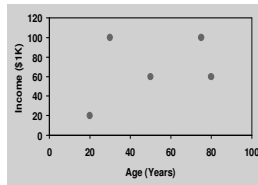
	1 Variable (Ch. 1)	2 Variables (Ch.2)
Graphical Summaries	histograms, etc.	scatterplots
Numerical Summaries	center, spread	correlation
Models	density curve	regression

Exploring the Relationship b/t Two Variables

- We label the two variables as "x" and "y."
- **Question:** Is there an **association** between the two variables? That is, do they tend to vary together?
- **Graphical Solution: Scatterplot**  
A scatterplot is the most common way of graphically exploring the relationship between two **quantitative** variables.

Example

Name	Age	Income (\$ 1K)
Anna	20	20
Bert	50	60
Chad	80	60
Dave	30	100
Erin	75	100



Purpose

1. Explore the nature of the relationship between two variables
2. Show that one variable can explain the variation in the other variable

**Response (Dependent) Variable** - measures an outcome of a study  
**Explanatory (Independent) Variable** - explains or causes changes in the response variable.

In a response/explanatory setting, **always** plot the explanatory (x) variable on the horizontal axis.

Response/Explanatory Variables

In the following situations, indicate if a response-explanatory relationship exists. If so, which variable is the response variable and which variable is the explanatory variable?

- The amount of time spent studying for an exam and the grade on the exam.  
Response: Grade on the Exam  
Explanatory: Time spent studying

- The height of husband and the height of wife.  
No response/explanatory relationship
- The crop yield and the amount of yearly rainfall.  
Response: Crop Yield  
Explanatory: Rainfall

Note: Calling one variable explanatory and the other response doesn't necessarily mean that changes in one *causes* changes in the other.

## Examining a Scatterplot

---

When examining a scatterplot we want to look for an overall pattern and observations that noticeably deviate from this pattern. Overall patterns are described by:

1. Form
2. Direction
3. Strength

An **outlier** is an individual value that falls outside the overall pattern of the relationship.

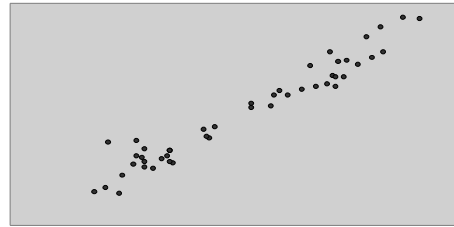
Statistics 528 - Lecture 6  
Prof. Kate Calder

7

## Form

---

### Linear



Statistics 528 - Lecture 6  
Prof. Kate Calder

8

## Form

---

### Curved



Statistics 528 - Lecture 6  
Prof. Kate Calder

9

## Form

---

### Clustered



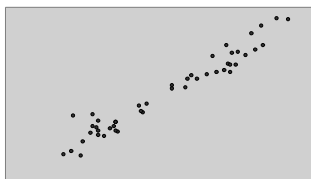
Statistics 528 - Lecture 6  
Prof. Kate Calder

10

## Direction

---

**Positive Association** – As values of the x variable increase, values of the y variable increase.



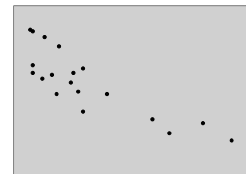
Statistics 528 - Lecture 6  
Prof. Kate Calder

11

## Direction

---

**Negative Association** – As values of the x variable increase, values of the y variable decrease.

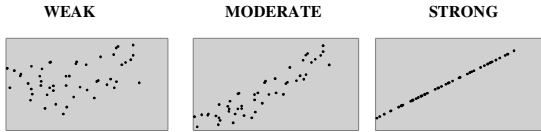


Statistics 528 - Lecture 6  
Prof. Kate Calder

12

## Strength

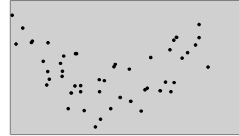
**Strength** refers to how scattered the points are about a line or curve.



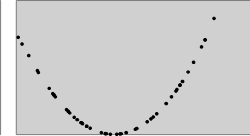
Statistics 528 - Lecture 6  
Prof. Kate Calder 13

## Strength

WEAK CURVE



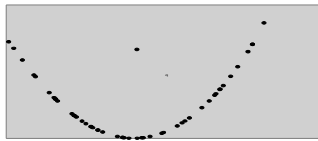
STRONG CURVE



Statistics 528 - Lecture 6  
Prof. Kate Calder 14

## Outliers

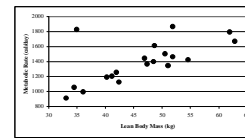
An individual is an **outlier** if it falls outside the overall pattern of the relationship.



Statistics 528 - Lecture 6  
Prof. Kate Calder 15

## Example

The plot shown below is a scatterplot of the lean body mass of some individuals and their metabolic rate.



Lean body mass – body weight excluding fat  
Metabolic rate – rate body consumes energy

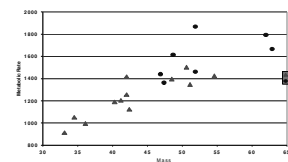
Statistics 528 - Lecture 6  
Prof. Kate Calder 16

- What is the form of this relationship?
  - The form is linear; the points follow a straight line.
- Is there a positive or negative association? Why?
  - Positive Association: As lean body mass increases, metabolic rate increases.
- How strong is the relationship?
  - The relationship is fairly weak.
- Do there appear to be any outliers? If so, which point? Does it strengthen or weaken the linear relationship?
  - There is one outlier - someone with a low body mass and high metabolism. It weakens the linear relationship.

Statistics 528 - Lecture 6  
Prof. Kate Calder 17

## Adding a Categorical Variable

Use different colors or symbols to plot points when you want to add a categorical variable to a scatterplot. Below is the plot of the lean body mass data with women marked as triangles and men as circles.



Statistics 528 - Lecture 6  
Prof. Kate Calder 18