Bivariate Data

PIERIAN DATA

Bivariate Data

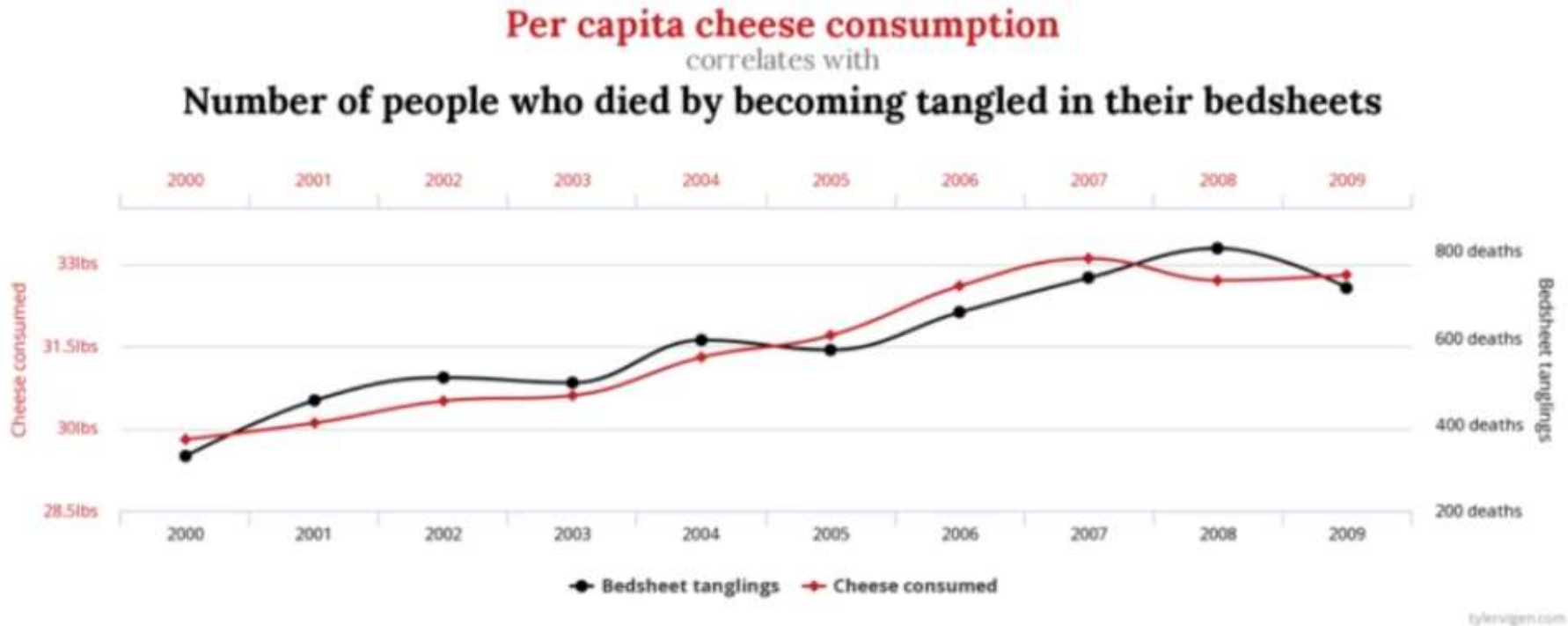
- Compares two variables
- By convention, the x-axis is set to the **independent variable**
- The y-axis is set to the **dependent variable**, or that which is being measured relative to x.

Bivariate Data

- Scatter plots may uncover a **correlation** between two variables
- They *can't* show **causality**!

Bivariate Data

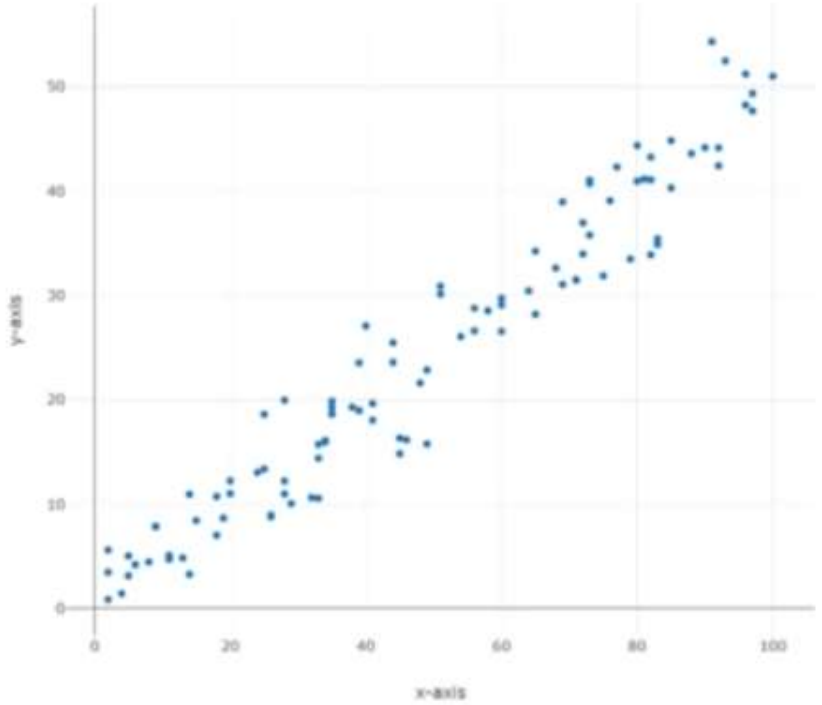
- **Correlation** between two variables
- Doesn't prove **causality**!



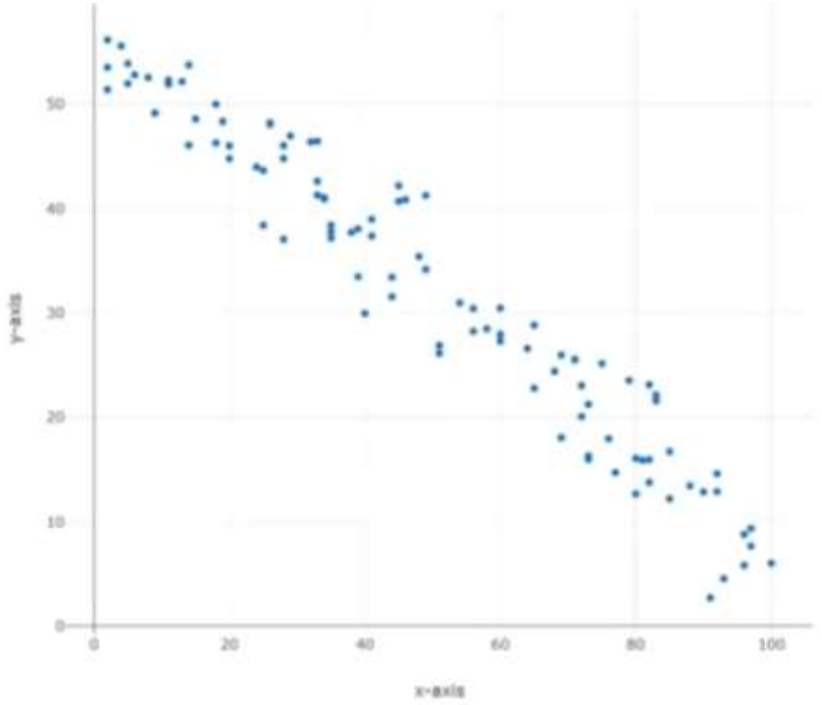
Bivariate Data

- More statistical analysis is needed to determine **causality**!
- For example: "Does increasing number of police officers decrease crime?"
- We would look at correlation, and do further analysis to understand causality.

Bivariate Data



Positive correlation



Negative or Inverse correlation

PIERIAN DATA

Covariance

- A common way to compare two variables is to compare their variances – how far from each item's mean do typical values fall?
- The first challenge is to match scale. Comparing height in inches to weight in pounds isn't meaningful unless we develop a **standard score** to **normalize** the data.

Covariance

- For simplicity, we'll consider the *population covariance:*

$$cov(X,Y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$$

Covariance Exercise

- Consider the following two tables:

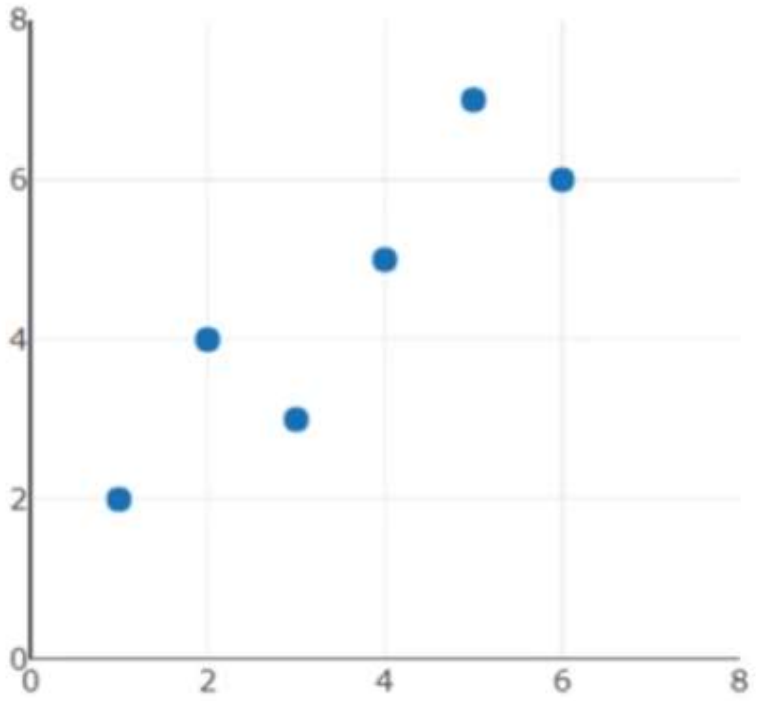
x	y
1	4
2	6
3	5
4	7
5	9
6	8

x	y
1	5
2	9
3	7
4	4
5	8
6	6

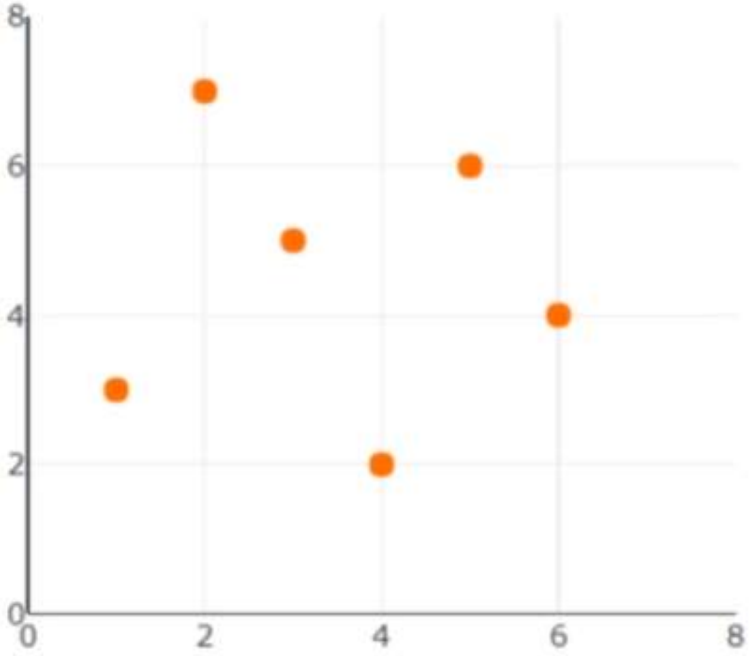
Covariance Exercise

- Plot them:

x	y
1	4
2	6
3	5
4	7
5	9
6	8



x	y
1	5
2	9
3	7
4	4
5	8
6	6



Covariance Exercise

- Calculate mean values:

x	y
1	4
2	6
3	5
4	7
5	9
6	8

$$\bar{x} = \frac{1+2+3+4+5+6}{6} = 3.5$$

$$\bar{y} = \frac{4+6+5+7+9+8}{6} = 6.5$$

x	y
1	5
2	9
3	7
4	4
5	8
6	6

$$\bar{x} = \frac{1+2+3+4+5+6}{6} = 3.5$$

$$\bar{y} = \frac{5+9+7+4+8+6}{6} = 6.5$$

Covariance Exercise

$\bar{x} = 3.5, \bar{y} = 6.5$

- Calculate $(x - \bar{x})$ and $(y - \bar{y})$:

x	y	(x - $\bar{x}$)	(y - $\bar{y}$)
1	4	-2.5	-2.5
2	6	-1.5	-0.5
3	5	-0.5	-1.5
4	7	0.5	0.5
5	9	1.5	2.5
6	8	2.5	1.5

x	y	(x - $\bar{x}$)	(y - $\bar{y}$)
1	5	-2.5	-1.5
2	9	-1.5	2.5
3	7	-0.5	0.5
4	4	0.5	-2.5
5	8	1.5	1.5
6	6	2.5	-0.5

Covariance Exercise

$\bar{x} = 3.5, \bar{y} = 6.5$

- Calculate $(x - \bar{x})(y - \bar{y})$:

x	y	$(x - \bar{x})$	$(y - \bar{y})$	$(x - \bar{x})(y - \bar{y})$
1	4	-2.5	-2.5	6.25
2	6	-1.5	-0.5	0.75
3	5	-0.5	-1.5	0.75
4	7	0.5	0.5	0.25
5	9	1.5	2.5	3.75
6	8	2.5	1.5	3.75

x	y	$(x - \bar{x})$	$(y - \bar{y})$	$(x - \bar{x})(y - \bar{y})$
1	5	-2.5	-1.5	3.75
2	9	-1.5	2.5	-3.75
3	7	-0.5	0.5	-0.25
4	4	0.5	-2.5	-1.25
5	8	1.5	1.5	2.25
6	6	2.5	-0.5	-1.25

Covariance Exercise

$\bar{x} = 3.5, \bar{y} = 6.5$

- Calculate sums:

x	y	$(x - \bar{x})$	$(y - \bar{y})$	$(x - \bar{x})(y - \bar{y})$
1	4	-2.5	-2.5	6.25
2	6	-1.5	-0.5	0.75
3	5	-0.5	-1.5	0.75
4	7	0.5	0.5	0.25
5	9	1.5	2.5	3.75
6	8	2.5	1.5	3.75
			Σ	**15.5**

x	y	$(x - \bar{x})$	$(y - \bar{y})$	$(x - \bar{x})(y - \bar{y})$
1	5	-2.5	-1.5	3.75
2	9	-1.5	2.5	-3.75
3	7	-0.5	0.5	-0.25
4	4	0.5	-2.5	-1.25
5	8	1.5	1.5	2.25
6	6	2.5	-0.5	-1.25
			Σ	**-0.5**

Covariance Exercise

$\bar{x} = 3.5, \bar{y} = 6.5$

- Calculate covariance:

x	y
1	4
2	6
3	5
4	7
5	9
6	8

$$cov(X,Y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$$

$$= \frac{15.5}{6} = \mathbf{2.583}$$

x	y
1	5
2	9
3	7
4	4
5	8
6	6

$$cov(X,Y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$$

$$= \frac{-0.5}{6} = \mathbf{-0.083}$$

Σ	15.5

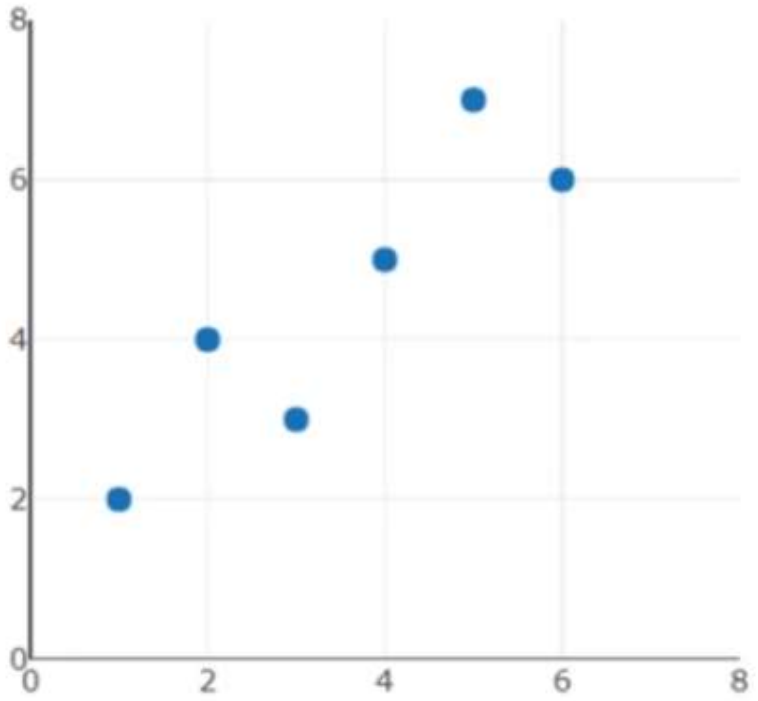
Σ	-0.5

PIERIAN DATA

Covariance Exercise

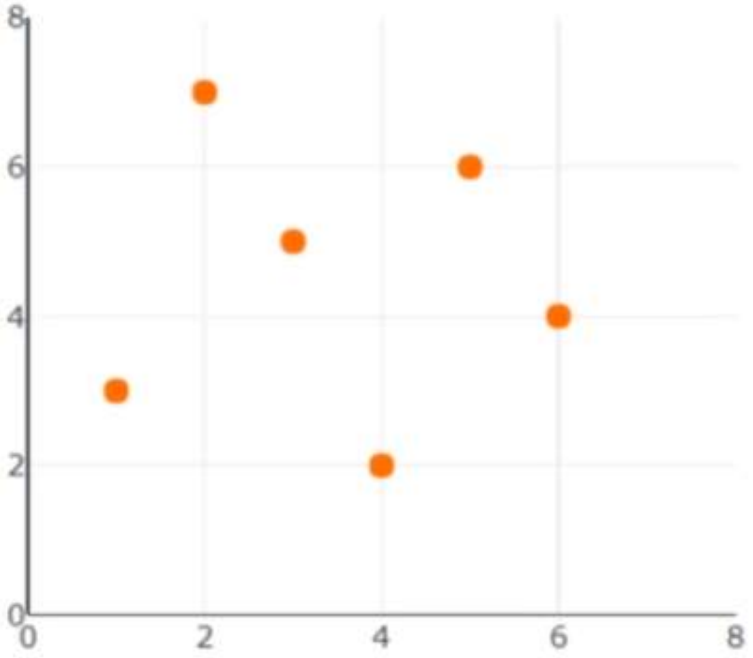
- Compare covariances:

x	y
1	4
2	6
3	5
4	7
5	9
6	8



cov(x,y) = 2.583

x	y
1	5
2	9
3	7
4	4
5	8
6	6



cov(x,y) = -0.083