

POLI 30 D: Political Inquiry
Professor Umberto Mignozzetti
(Based on DSS Materials)

**Lecture 07 | Measuring Population
Characteristics II**

Before we start

Announcements:

- ▶ Quizzes and Participation: On Canvas.
- ▶ Github page:
<https://github.com/umbertomig/POLI30Dpublic>
- ▶ Piazza forum: <https://piazza.com/ucsd/winter2023/17221>

Before we start

Recap: We learned:

- ▶ The definitions of theory, scientific theory, and hypotheses.
- ▶ Data, datasets, variables, and how to compute means.
- ▶ Causal effect, treatments, outcomes, and randomization.
- ▶ Sampling, descriptive statistics, and descriptive plots for one variable.

Great job!

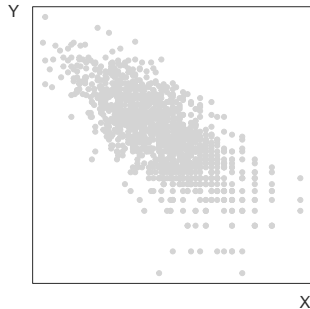
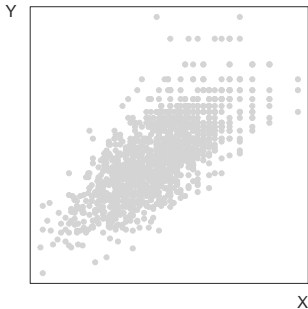
- ▶ Do you have any questions about these contents?

Plan for Today

- Exploring the Relationship Between Two Variables
 - Scatter plots
 - Correlations

Scatter Plots

- A **scatter plot** enables us to visualize the relationship between two variables by plotting one against the other



Scatter Plots

Imagine we have two variables:

X	Y
4	2
8	5
10	3

We can create the scatter plot by plotting one point at a time.

Scatter Plots

Imagine we have two variables:

X	Y
4	2
8	5
10	3

>> First, let's plot this point: $(x_1, y_1) = (4, 2)$

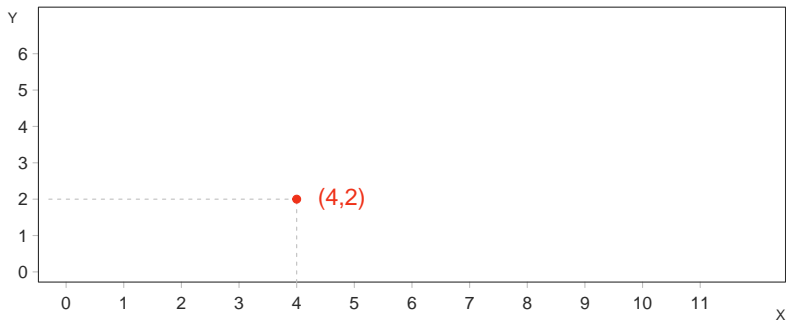
Scatter Plots

Imagine we have two variables:

X	Y
4	2
8	5
10	3

>> First, let's plot this point: $(x_1, y_1) = (4, 2)$

Scatter Plots

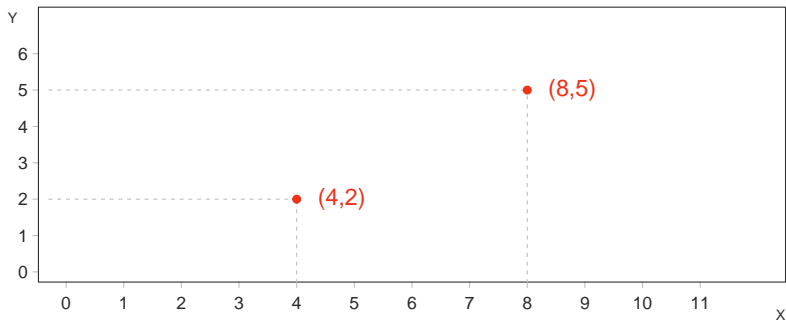


Scatter Plots

Imagine we have two variables:

X	Y	
4	2	>> First, let's plot this point: $(x_1, y_1) = (4, 2)$
8	5	>> Now, let's plot this point: $(x_2, y_2) = (8, 5)$
10	3	

Scatter Plots

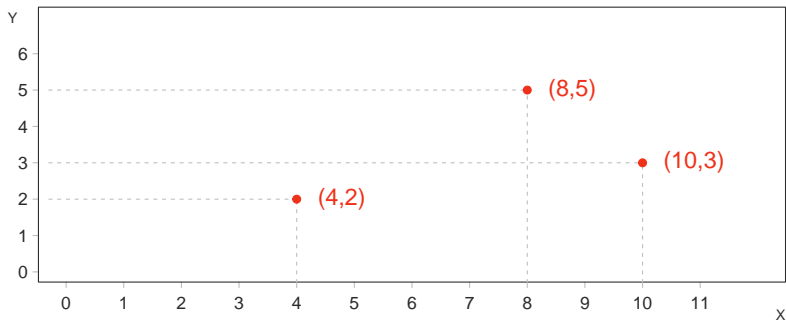


Scatter Plots

Imagine we have two variables:

X	Y	
4	2	>> First, let's plot this point: $(x_1, y_1) = (4, 2)$
8	5	>> Now, let's plot this point: $(x_2, y_2) = (8, 5)$
10	3	>> Finally, let's plot this point: $(x_3, y_3) = (10, 3)$

Scatter Plots



Scatter Plots

- ▶ R functions: `plot()` or `ggplot() + geom_point()`
- ▶ How many arguments are required?
 - ▶ two; the two variables

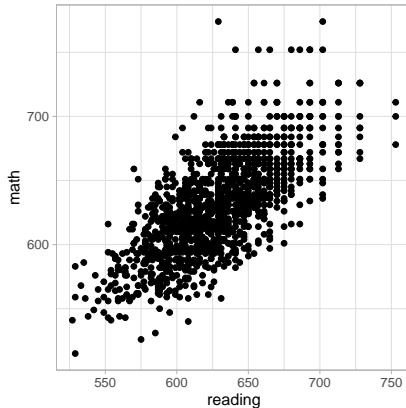
Scatter Plots

- Let us use the data from Project STAR:

```
head(star, 3) # shows first observations
##   classtype reading math graduated
## 1    small    578   610         1
## 2  regular    612   612         1
## 3  regular    583   606         1
```

- Unit of observation?
 - students; each observation represents a student
- Unit of measurement of *reading* and *math*?
 - points

Scatter Plots



- What can we learn from this scatter plot?

Correlation Coefficient

- ▶ The **correlation coefficient** is a statistic that summarizes the relationship between two variables with a number
 - ▶ denoted as $\text{cor}(X,Y)$
- ▶ $\text{cor}(X,Y)$ summarizes the **direction** and the **strength** of the **linear association** between X and Y
- ▶ $\text{cor}(X,Y)$ ranges from -1 to 1

Correlation Coefficient

The sign reflects the **direction** of the linear association:

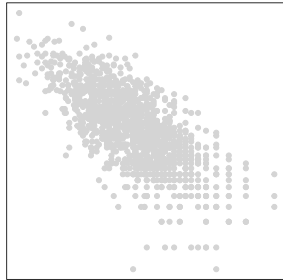
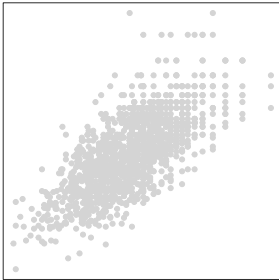
- ▶ $\text{cor}(X,Y) > 0$ (tends to see one **increasing** when the other increases)
- ▶ $\text{cor}(X,Y) < 0$ (tends to see one **decreasing** when the other increases)

The absolute value reflects the **strength** of the linear association:

- ▶ $|\text{cor}(X,Y)| = 0$ if there is no linear association
- ▶ $|\text{cor}(X,Y)| = 1$ if there is a perfect linear association
- ▶ $|\text{cor}(X,Y)|$ increases as the linear association becomes stronger

Correlation Coefficient

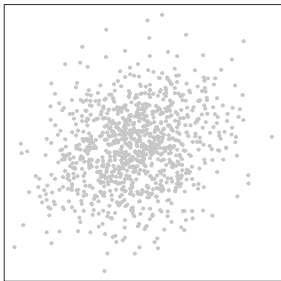
Direction of the linear association between two variables:



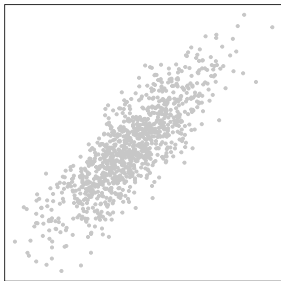
positive linear association	vs.	negative linear association
positive correlation	vs.	negative correlation

Correlation Coefficient

Strength of the linear association between two variables:

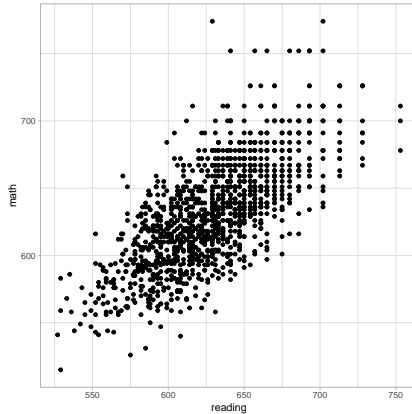


weak linear association
absolute value closer to 0



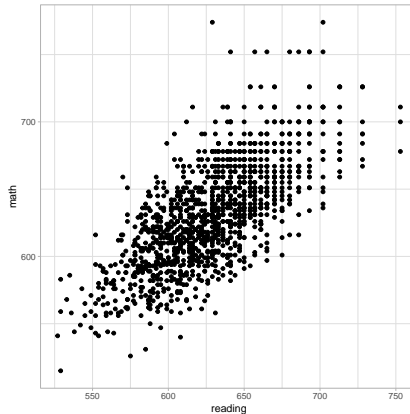
vs. **strong** linear association
absolute value closer to 1

Correlation Coefficient



- Do you expect the correlation between *reading* and *math* grades to be positive or negative?

Correlation Coefficient



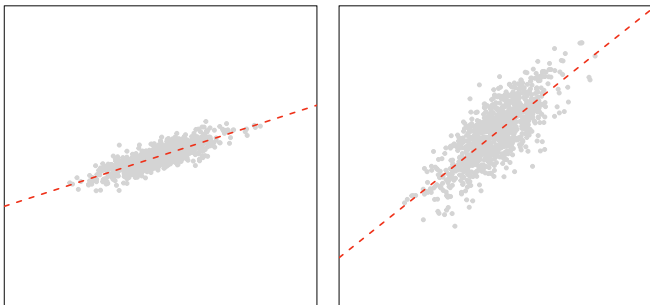
- Do you expect the absolute value of the correlation between *reading* and *math* to be closer to 1 or to 0?

Correlation Coefficient

- ▶ R function: `cor()`
- ▶ How many required arguments?
 - ▶ two; the two variables
- ▶ Does the order of the arguments matter?
 - ▶ no; $\text{cor}(X,Y) = \text{cor}(Y,X)$
- ▶ What is the code to compute the correlation between *reading* and *math*?
 - ▶ Answer:

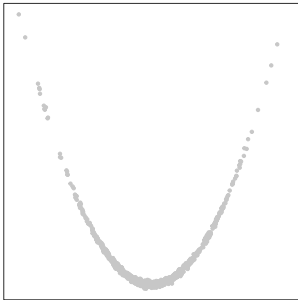
```
cor(star$reading, star$math)
## [1] 0.7161218
```
 - ▶ Is the correlation what we expected?

Correlation Coefficient



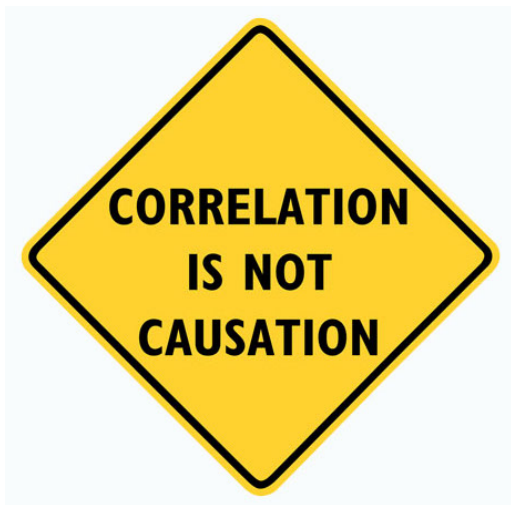
- Line of best fit is steeper in the first or second scatterplot?
- Is correlation higher in the first or second scatterplot?

Correlation Coefficient



- ▶ $\text{cor}(X,Y) \approx 0$
- ▶ Does this mean that there is no relationship between the two variables? No. Check the dino!

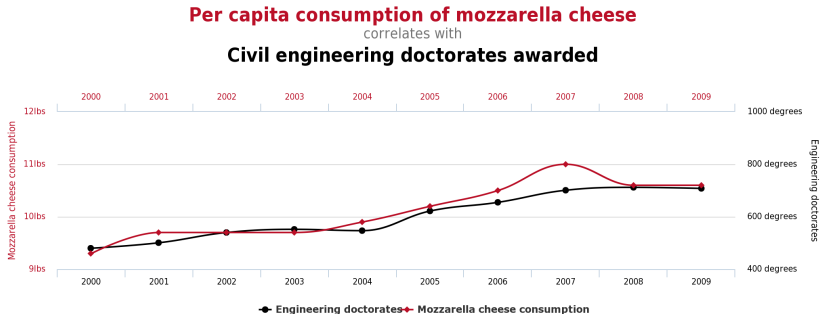
Correlation does not necessarily imply causation



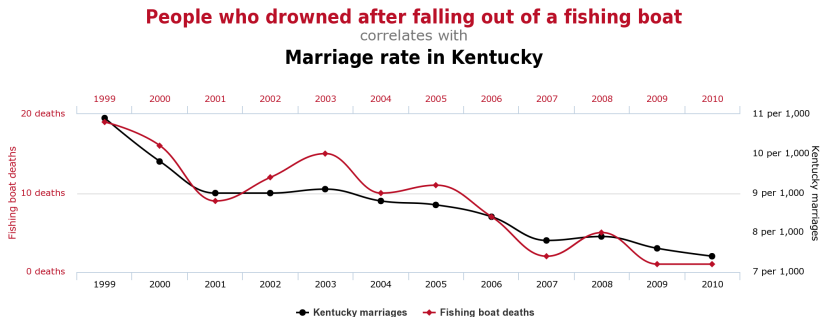
Correlation does not necessarily imply causation

- ▶ Just because two variables have a strong correlation does not mean that changes in one variable cause changes in the other
- ▶ Example: *reading* and *math* are highly correlated.
 - ▶ Does that mean that if you study *math* you learn *reading*?!

Correlation does not necessarily imply causation



Correlation does not necessarily imply causation



► More on this later in the semester!

Summary

- ▶ **Today's Class:**
 - ▶ Exploring the Relationship Between Two Variables
 - ▶ Scatterplots
 - ▶ Correlations
- ▶ Next class:
 - ▶ Prediction and Linear Regression

Questions?

See you in the next class!