Scatterplots and Correlation

Objectives

1. Create and analyze scatterplots
2. Determine the type of correlation of a scatterplot

Scatterplots

Scatterplot

A **scatterplot** is a visual display which can be used to examine an association between two variables, usually x and y.

Scatterplots

Scatterplot

A **scatterplot** is a visual display which can be used to examine an association between two variables, usually x and y.

The independent variable, x, is called the **explanatory variable** and the dependent variable, y, is called the **response variable**.

Scatterplots

Scatterplot

A **scatterplot** is a visual display which can be used to examine an association between two variables, usually x and y.

The independent variable, x, is called the **explanatory variable** and the dependent variable, y, is called the **response variable**.

Scatterplots allow us to see if there is a relationship between the two variables.

Example 1

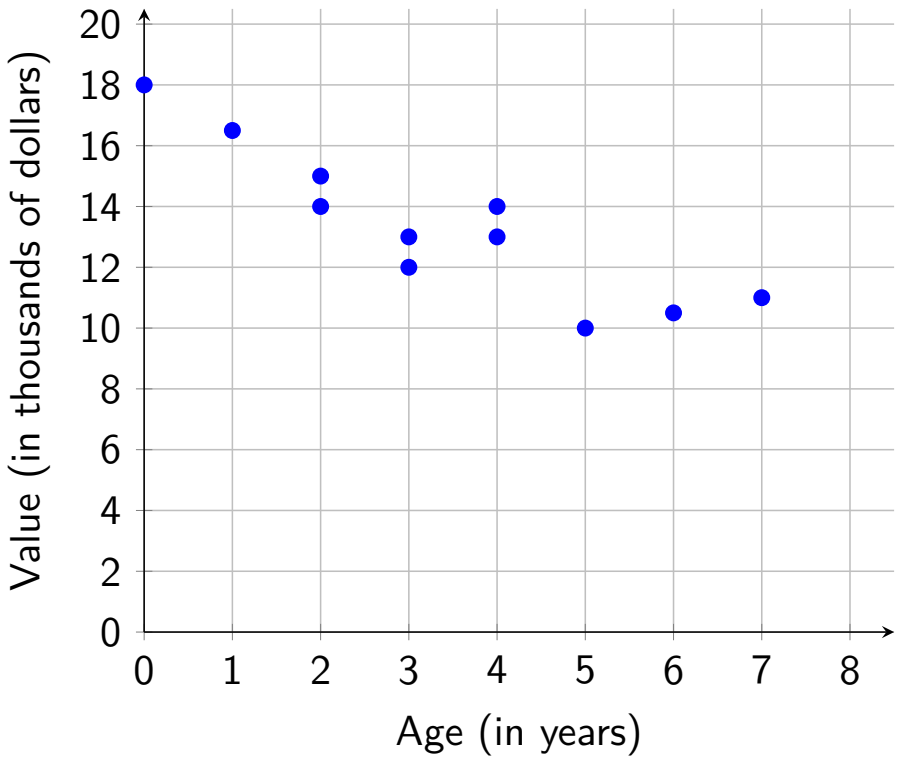
The table below shows the age of a certain model of car (in years) with the cars current value (in thousands of dollars). Create a scatterplot for the data.

Age	**Value**
2	15
3	12
3	13
2	14
4	13
5	10
6	10.5
1	16.5
0	18
4	14
7	11

Example 1

The table below shows the age of a certain model of car (in years) with the cars current value (in thousands of dollars). Create a scatterplot for the data.

Age	Value
2	15
3	12
3	13
2	14
4	13
5	10
6	10.5
1	16.5
0	18
4	14
7	11



Objectives

1. Create and analyze scatterplots

2. Determine the type of correlation of a scatterplot

Direction of Points

Often times, the data in a scatterplot has some pattern to it.

Direction of Points

Often times, the data in a scatterplot has some pattern to it.

Correlation

A **correlation** between two variables examines how the response variable's (y) values change as the explanatory variable's (x) values change.

Direction of Points

Often times, the data in a scatterplot has some pattern to it.

Correlation

A **correlation** between two variables examines how the response variable's (y) values change as the explanatory variable's (x) values change.

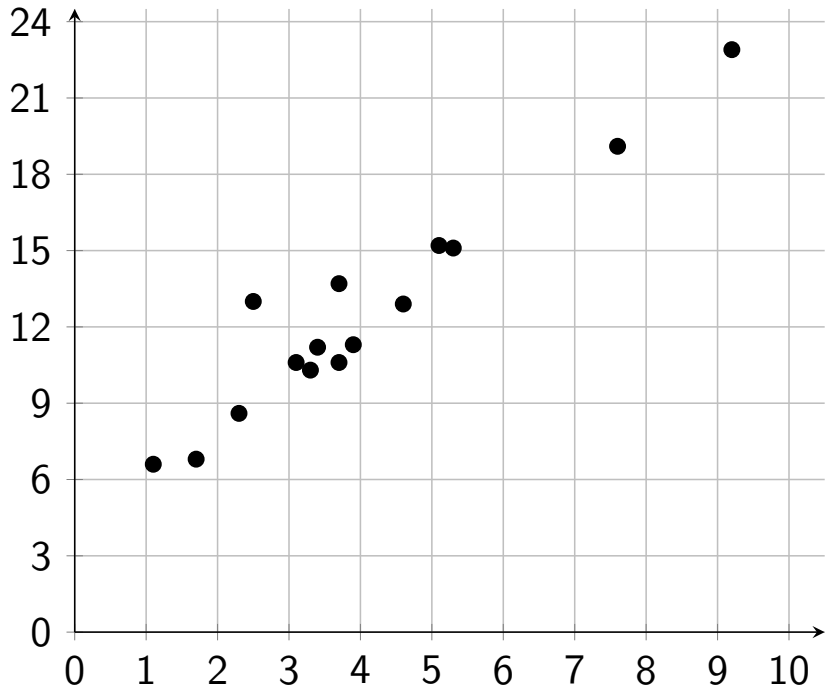
We will examine three correlation types: positive, negative, and none (a.k.a. no correlation)

Positive Correlation

As x increases, so does y.

Positive Correlation

As x increases, so does y.

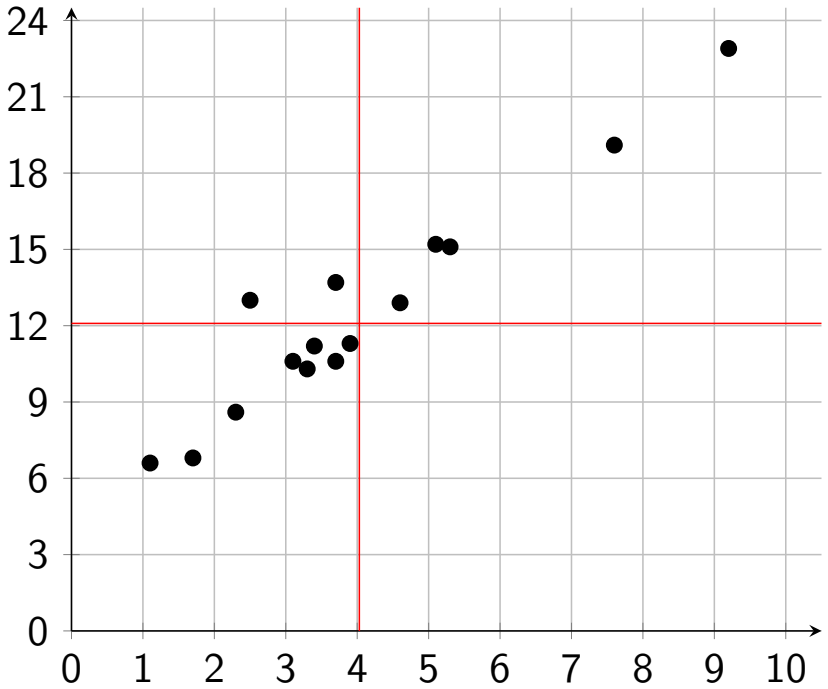


Quadrants from Means

We can also get a *general idea* of the type of correlation by looking at the counts of observations in the quadrants formed by the means of x and y.

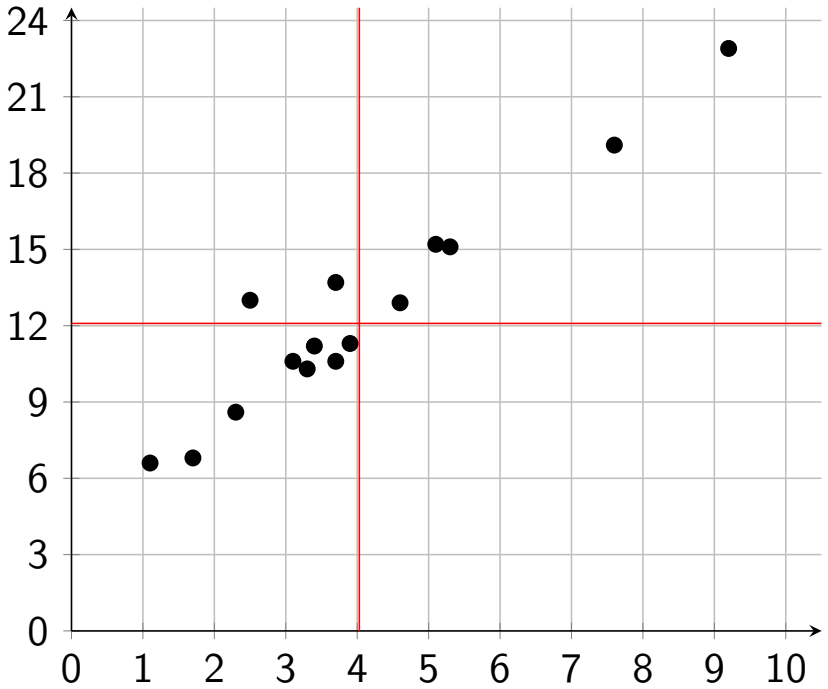
Quadrants from Means

We can also get a *general idea* of the type of correlation by looking at the counts of observations in the quadrants formed by the means of x and y.



Quadrants from Means

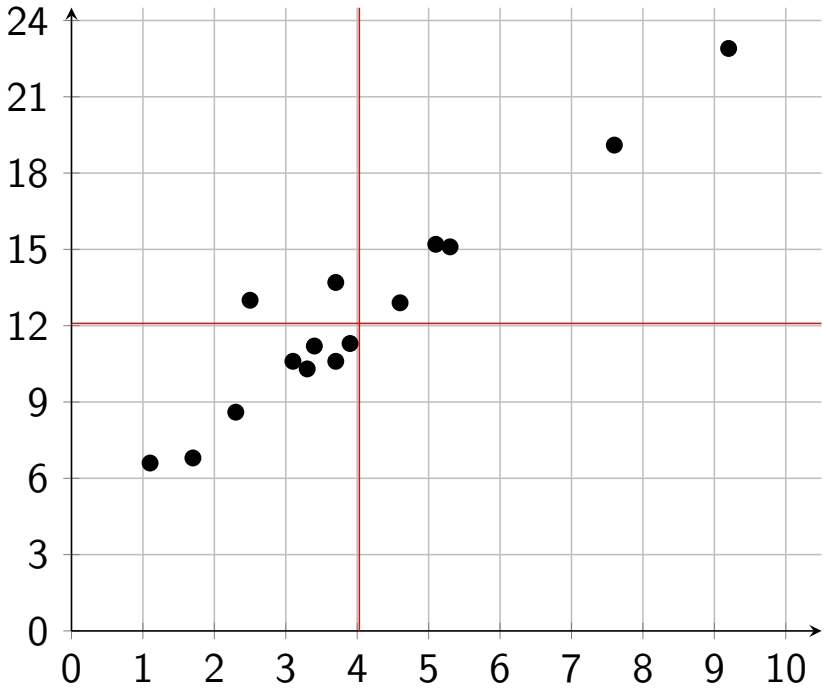
We can also get a *general idea* of the type of correlation by looking at the counts of observations in the quadrants formed by the means of x and y.



Q1: 5 values
Q3: 8 values
Total $= 13$

Quadrants from Means

We can also get a *general idea* of the type of correlation by looking at the counts of observations in the quadrants formed by the means of x and y.

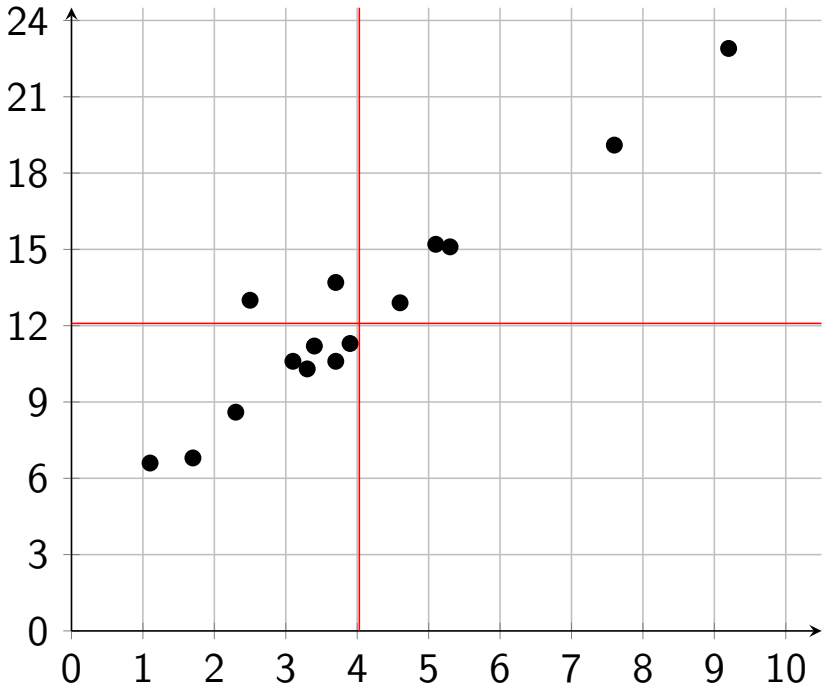


Q1: 5 values
Q3: 8 values
Total $= 13$

Q2: 2 values
Q4: 0 values
Total: 2

Quadrants from Means

We can also get a *general idea* of the type of correlation by looking at the counts of observations in the quadrants formed by the means of x and y.



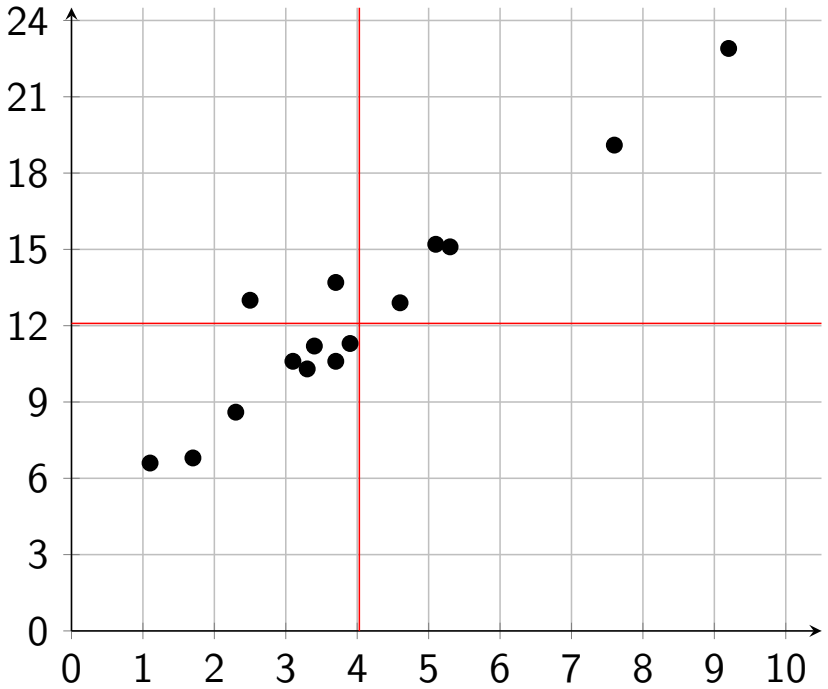
Q1: 5 values
Q3: 8 values
Total $= 13$

Q2: 2 values
Q4: 0 values
Total: 2

11 more points in Q1 and Q3

Quadrants from Means

We can also get a *general idea* of the type of correlation by looking at the counts of observations in the quadrants formed by the means of x and y.



Q1: 5 values
Q3: 8 values
Total $= 13$

Q2: 2 values
Q4: 0 values
Total: 2

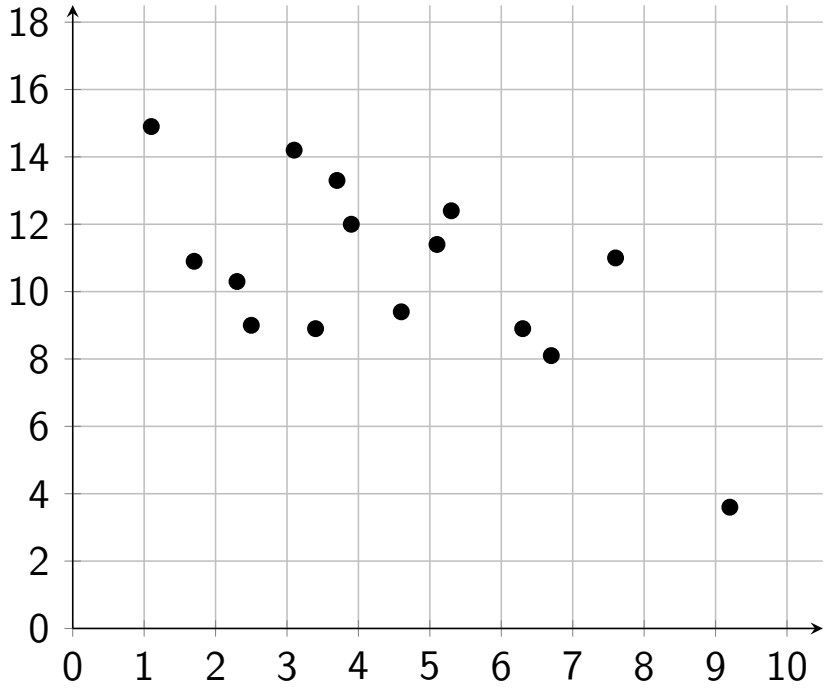
11 more points in Q1 and Q3 suggests positive correlation

Negative Correlation

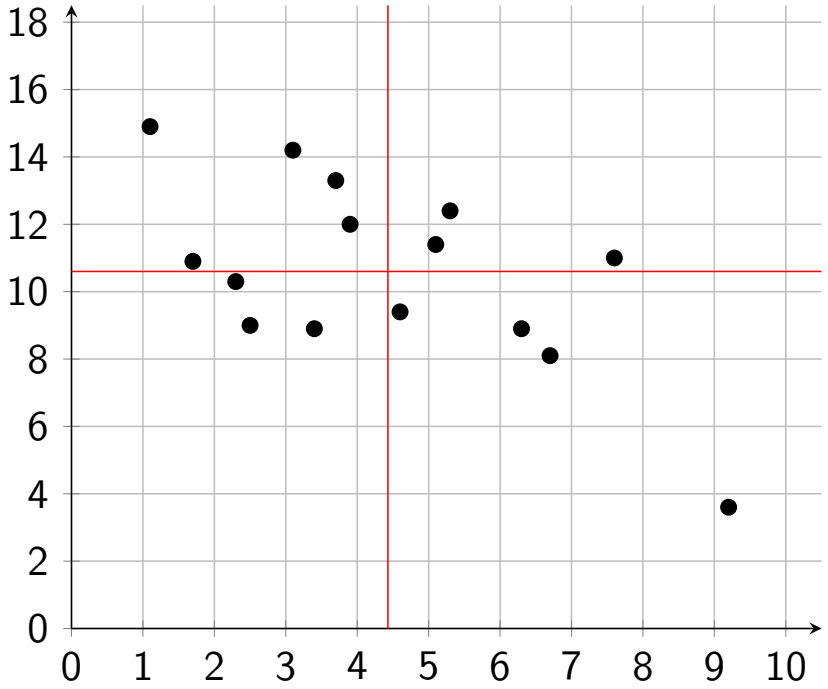
As x increases, y decreases.

Negative Correlation

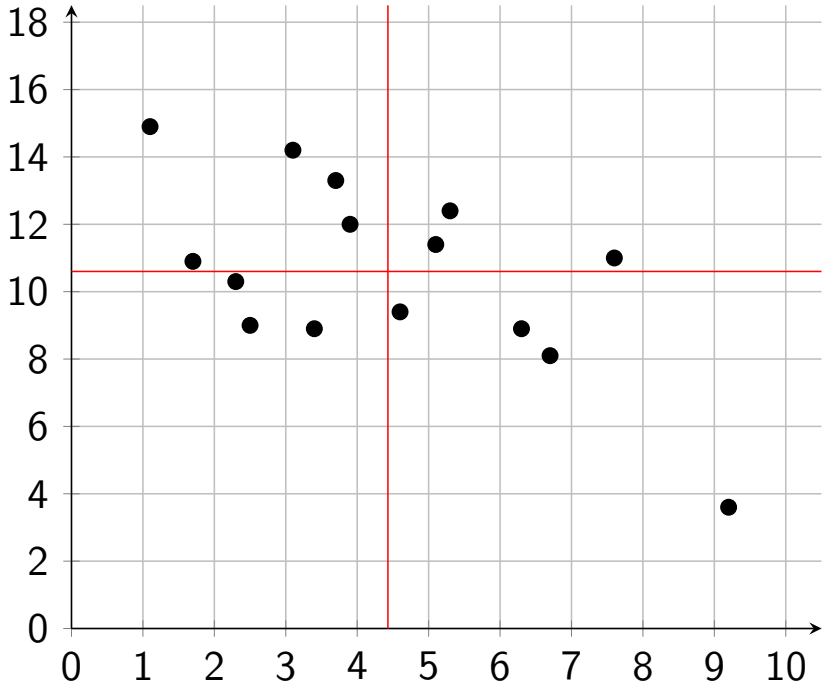
As x increases, y decreases.



Quadrants from Means



Quadrants from Means

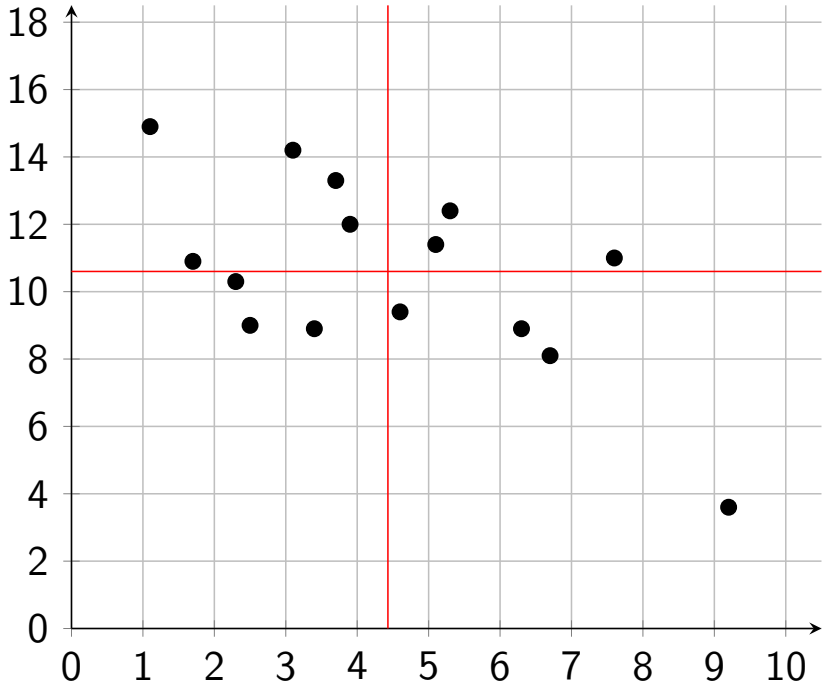


Q1: 3 values

Q3: 3 values

Total = 6

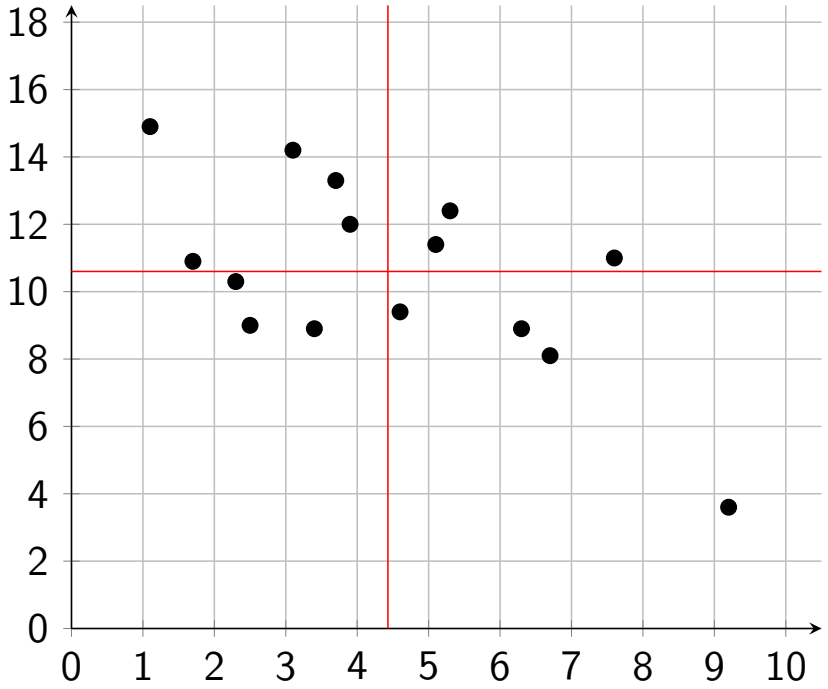
Quadrants from Means



Q1: 3 values
Q3: 3 values
Total = 6

Q2: 5 values
Q4: 4 values
Total: 9

Quadrants from Means

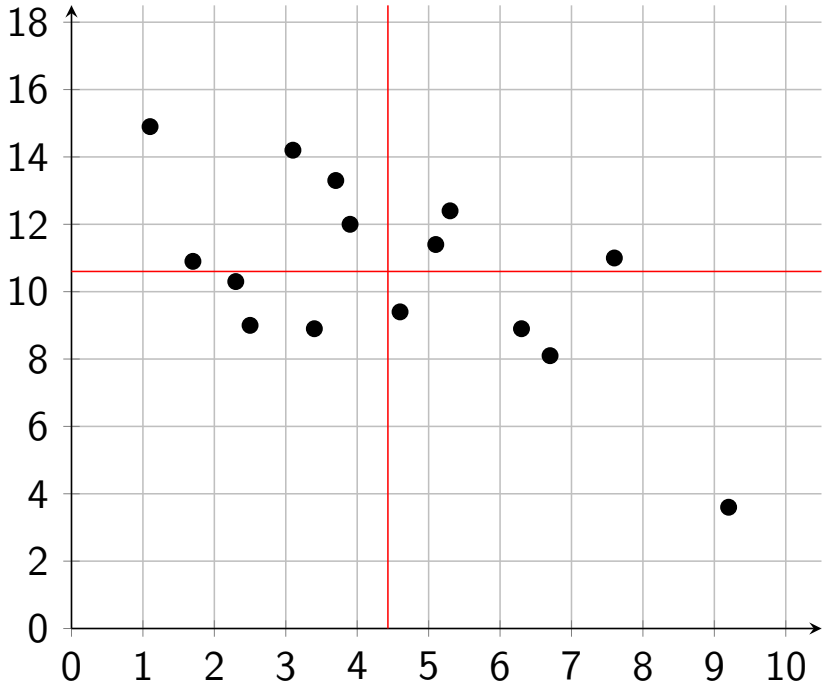


Q1: 3 values
Q3: 3 values
Total = 6

Q2: 5 values
Q4: 4 values
Total: 9

3 more points in Q2 and Q4

Quadrants from Means



Q1: 3 values
Q3: 3 values
Total = 6

Q2: 5 values
Q4: 4 values
Total: 9

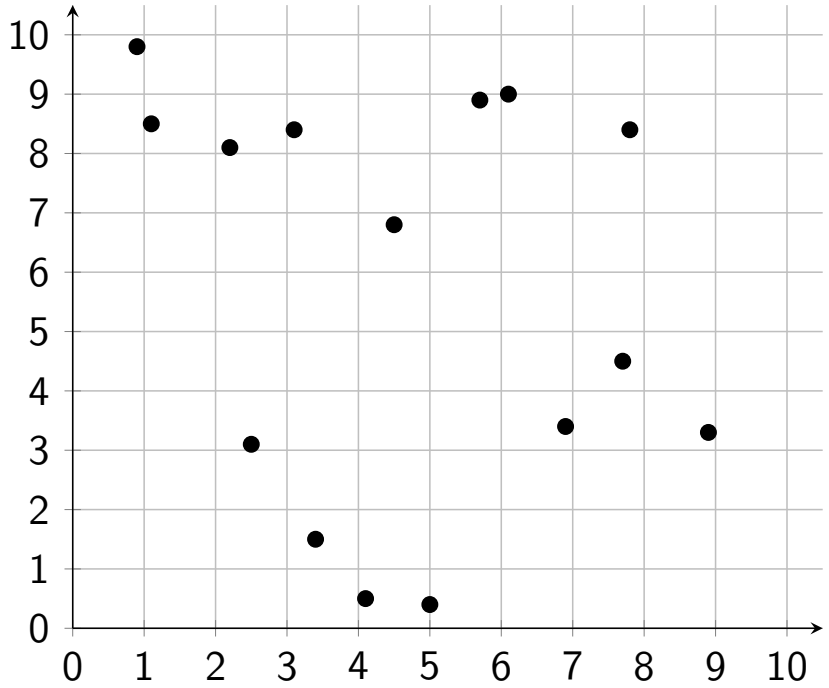
3 more points in Q2 and Q4 suggests a very weak negative correlation

No Correlation

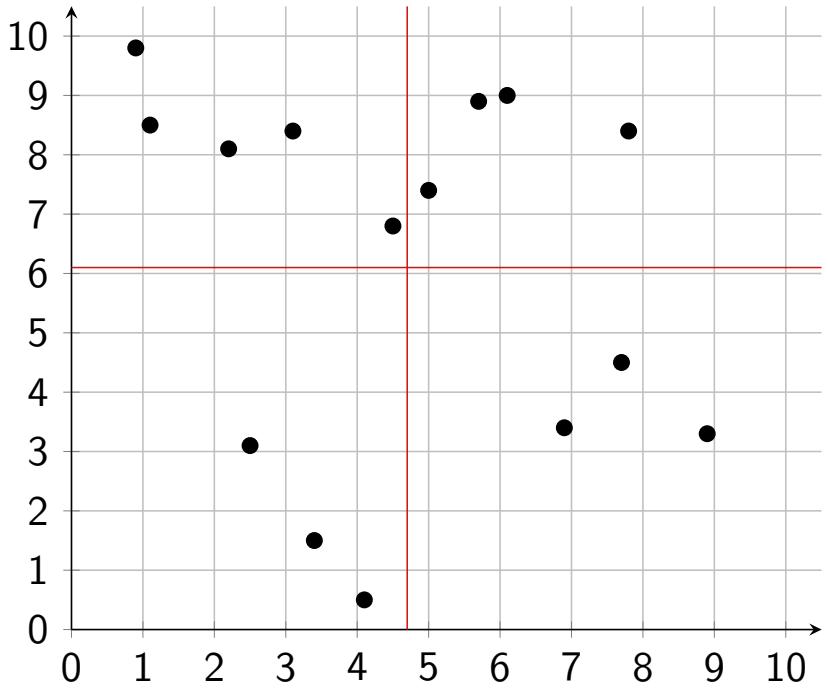
There is no visible pattern between x and y.

No Correlation

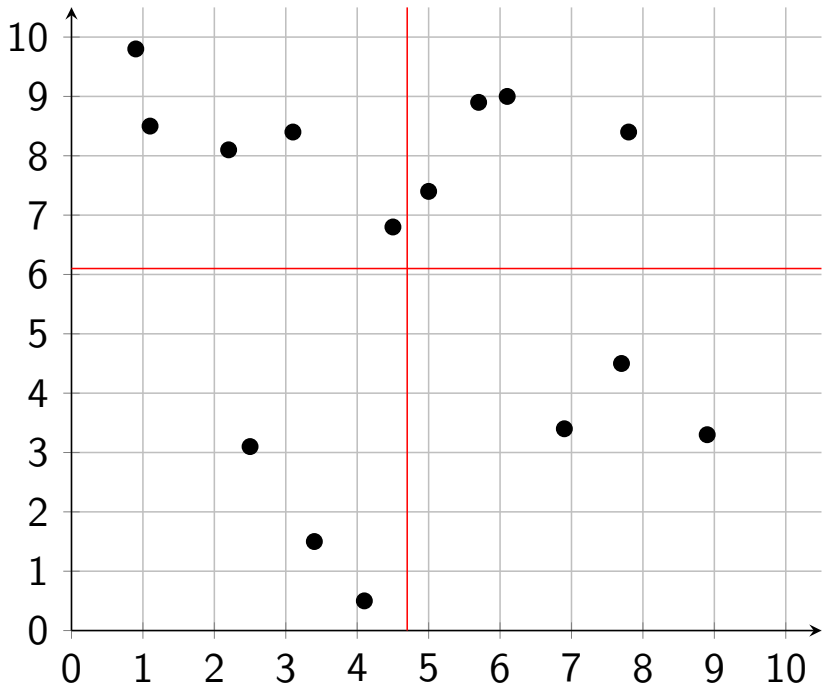
There is no visible pattern between x and y.



Quadrants of Means



Quadrants of Means

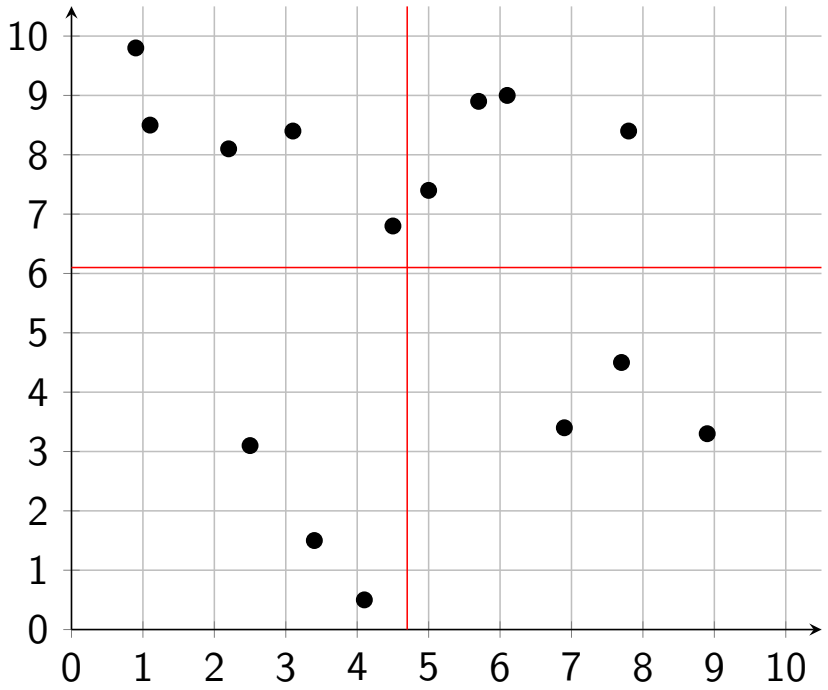


Q1: 4 values

Q3: 3 values

Total = 7

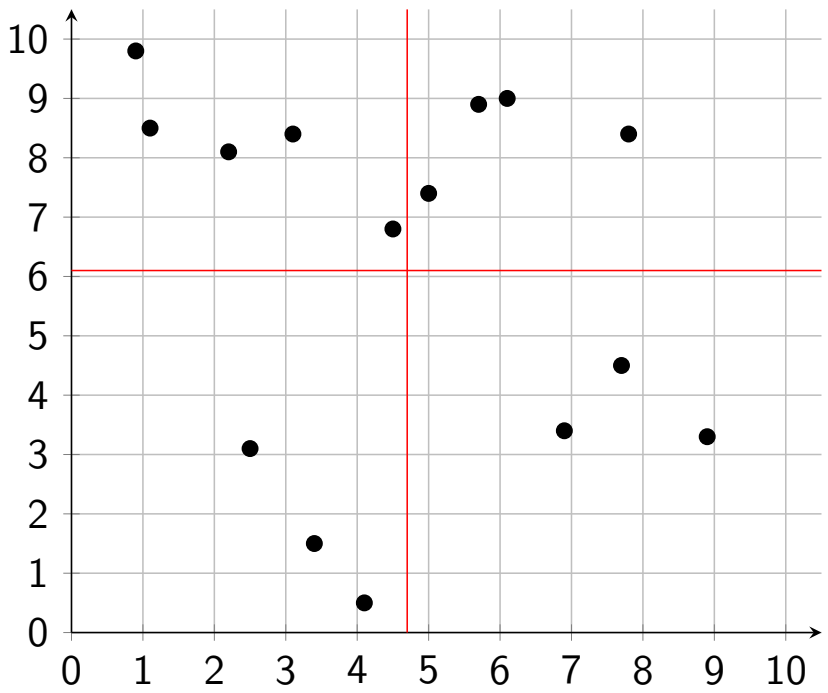
Quadrants of Means



Q1: 4 values
Q3: 3 values
Total = 7

Q2: 5 values
Q4: 3 values
Total: 8

Quadrants of Means

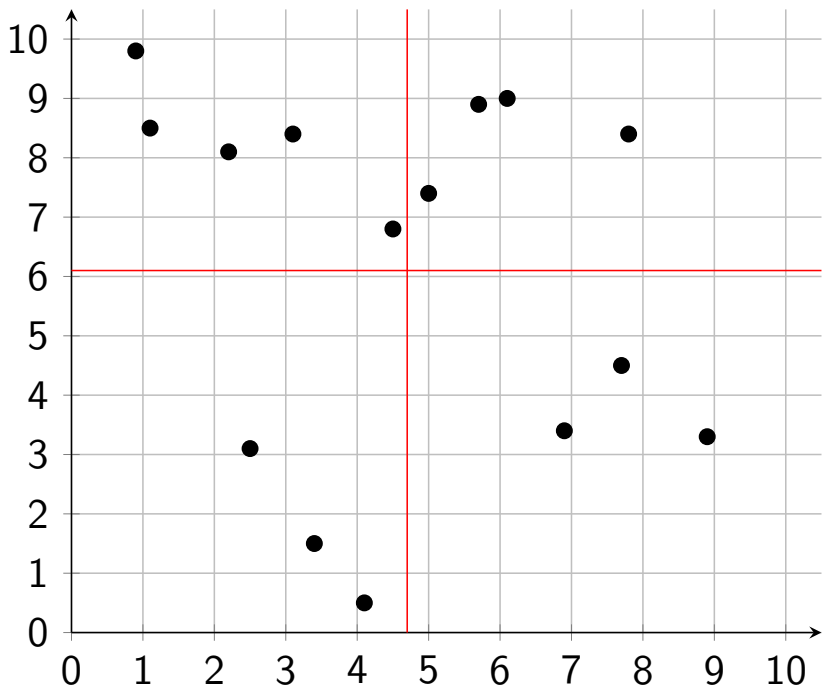


Q1: 4 values
Q3: 3 values
Total = 7

Q2: 5 values
Q4: 3 values
Total: 8

1 more point in Q2 and Q4

Quadrants of Means



Q1: 4 values
Q3: 3 values
Total = 7

Q2: 5 values
Q4: 3 values
Total: 8

1 more point in Q2 and Q4 suggests almost no correlation

Correlation vs. Causation

It is important to note that just because there may be a strong correlation (an **association**) between two variables **it does not mean that one causes the other to happen.**

Correlation vs. Causation

It is important to note that just because there may be a strong correlation (an **association**) between two variables **it does not mean that one causes the other to happen.**

For instance, we can look at the association of a dog's paw size with the dog's weight.

Correlation vs. Causation

It is important to note that just because there may be a strong correlation (an **association**) between two variables **it does not mean that one causes the other to happen.**

For instance, we can look at the association of a dog's paw size with the dog's weight.

We might notice that larger paws tend to have larger weights, but we can not conclude that large paws cause a large weight.

Correlation vs. Causation

It is important to note that just because there may be a strong correlation (an **association**) between two variables **it does not mean that one causes the other to happen.**

For instance, we can look at the association of a dog's paw size with the dog's weight.

We might notice that larger paws tend to have larger weights, but we can not conclude that large paws cause a large weight.

If there is a strong correlation there may be lurking variables or confounding at play.

Lurking Variables and Confounding

Lurking Varaible

A **lurking variable** is an explanatory variable that is not considered in a study or experiment that has an influence on the response variable.

Lurking Variables and Confounding

Lurking Varaible

A **lurking variable** is an explanatory variable that is not considered in a study or experiment that has an influence on the response variable.

Confounding

Confounding occurs when we can not distinguish the effect(s) one (or many) explanatory variables has (have) on a response variable.