

Section 5.1 Scatter Plots

Bivariate data is data for two different variables (usually related in some way). The two types of data can be classified as either a response variable or an explanatory variable.

A **response variable** measures the outcome of a study. An **explanatory variable** attempts to explain the observed outcomes.

The value of a dependent variable relies upon that of an independent variable. Thus a **response variable** corresponds to a **dependent variable** while an **explanatory variable** corresponds to an **independent variable**.

Example 1: Identify the response and explanatory variables.

a. Amount of time worked and amount of paycheck.

Explanatory:

Amount of time worked

Response:

Amount of paycheck

b. Fish activity and water temperature.

Explanatory:

Water temp

Response:

Fish activity

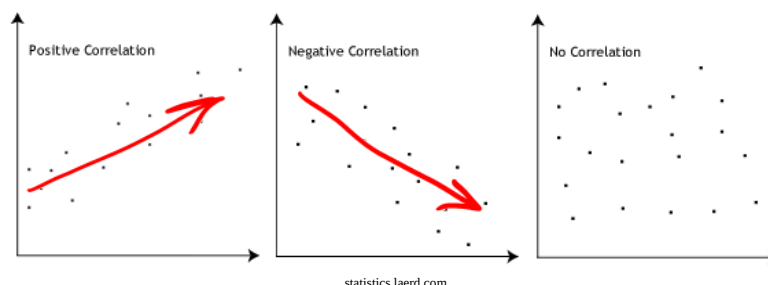
The most common graphical display used to study the association between two variables is called a **scatter plot**. The R command is: `plot()` The purpose of this kind of graph is to demonstrate relationships and trends within the paired data.

To interpret a scatter plot we will look at the direction, form and strength.

Direction: Look to see if the variables have a positive or negative association.

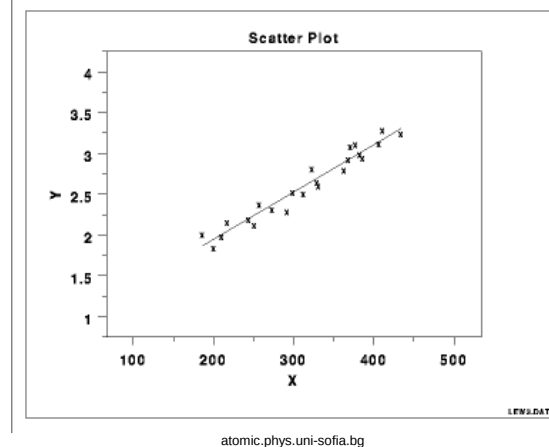
Two variables are said to be **positively** related if larger values of one variable tend to be associated with larger values of the other.

Two variables are said to be **negatively** related if larger values of one variable tend to be associated with smaller values of the other.



- **Form:** Refers to the shape. Most of what we will see in this class are linear forms.
- **Strength:** Want to know how closely the points “follow” the shape. The closer the points fall in to a straight line, the stronger we say the relationship is.

Strong Positive Relationship



Example 2: The bivariate data given below relate the high temperature reached on a given day and the number of bottled water sold from a particular vending machine.

Temperature (in degrees)	Bottled Water (16 oz)
90	30
91	32
88	29
93	33
92	31
89	29
90	30
91	31
92	32
94	34

a. Construct a scatter plot.

In R: Create two lists of values:

Command:

```
temp=c(90,91,88,93,92,89,90,91,92,94)
```

```
water=c(30,32,29,33,31,29,30,31,32,34)
```

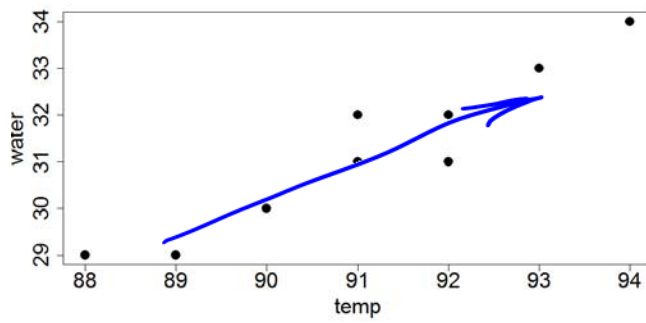
Construct the scatter plot:

Command:

```
plot(temp,water,pch=16,cex=2,cex.lab=2,cex.axis=2)
```

Handwritten annotations:

- bigger dots* (pointing to `pch=16`)
- labels* (pointing to `cex.lab=2`)
- enlarge font* (pointing to `cex.axis=2`)
- exp* (pointing to `temp`)
- res* (pointing to `water`)
- filled dots* (pointing to `pch=16`)



b. Does there appear to be a linear relationship between the two variables?

YES

c. Based on the scatter plot, would you characterize the relationship as positive or negative?

Positive