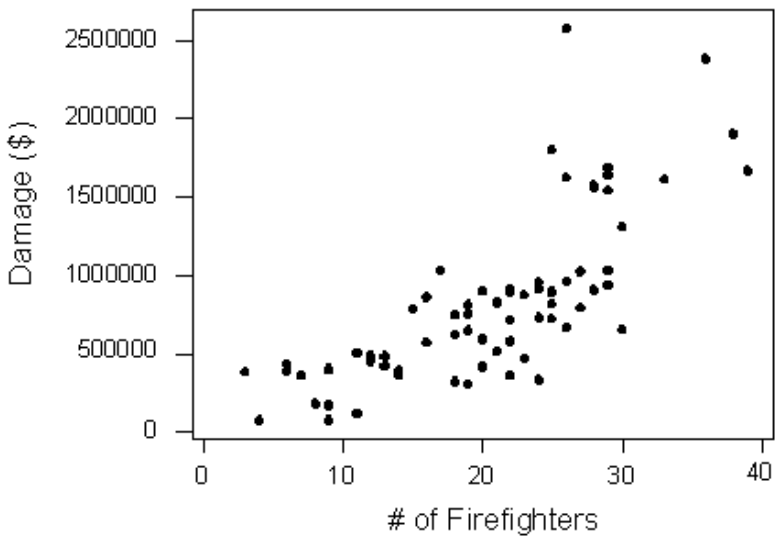
Causation and Lurking Variables: Impact on X and Y

Learning Objective: Recognize the distinction between association and causation, and identify potential lurking variables for explaining an observed relationship.

Let's start by looking at the following example:

Example: Fire Damage

The scatterplot below illustrates how the number of firefighters sent to fires (X) is related to the amount of damage caused by fires (Y) in a certain city.

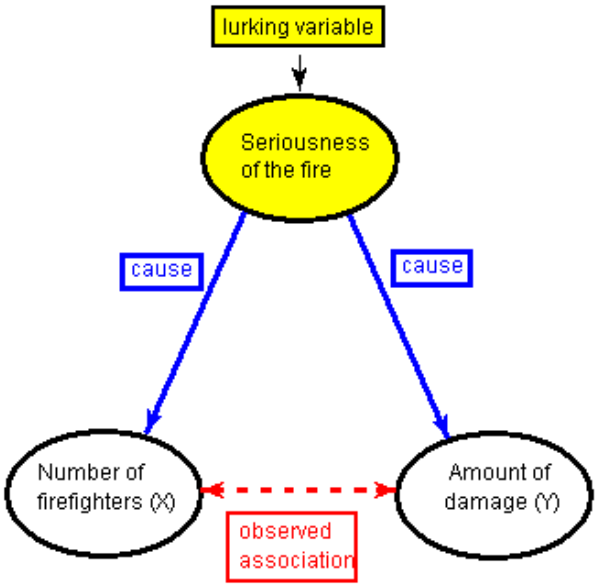


The scatterplot clearly displays a fairly strong (slightly curved) **positive** relationship between the two variables. Would it, then, be reasonable to conclude that sending more firefighters to a fire causes more damage, or that the city should send fewer firefighters to a fire, in order to decrease the amount

of damage done by the fire? Of course not! So what is going on here?

There is a **third variable in the background**—the seriousness of the fire—that is responsible for the observed relationship. More serious fires require more firefighters, and also cause more damage.

The following figure will help you visualize this situation:



Here, the seriousness of the fire is a **lurking variable.** A **lurking variable** is a variable that is not among the explanatory or response variables in a study, but could substantially affect your interpretation of the relationship among those variables.

In particular, as in our example, the lurking variable might have an effect on *both* the explanatory and the response variables. This common effect creates the observed association between the explanatory and response variables, even though there is no causal link between them. This possibility, that there might be a lurking variable (which we might not be thinking about) that is responsible for the observed relationship leads to our principle:

Principle

Association **does not** imply causation!
