

**Learning Goal:** Distinguish between association and causation. Identify lurking variables that may explain an observed relationship.

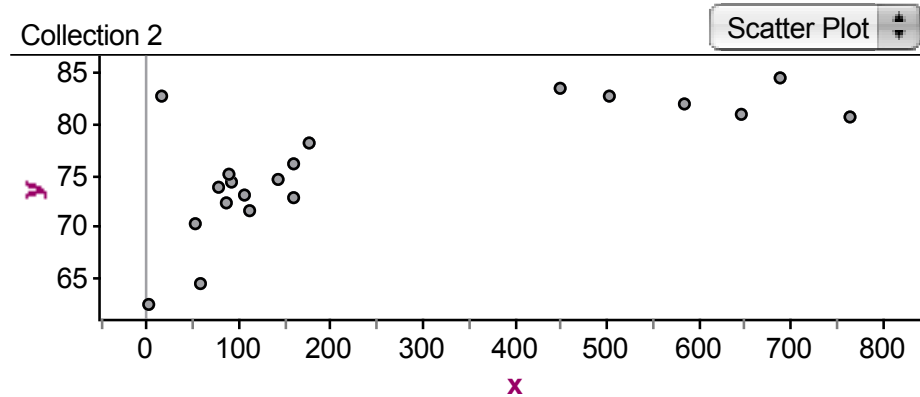
A *lurking variable* is a variable that is not measured in the study. It is a third variable that is neither the explanatory nor the response variable, but it affects your interpretation of the relationship between the explanatory and response variable.

- 1) To understand the above ideas, read this excerpt from *A Mathematician Reads the Newspaper* by John Allen Paulos.

"A more elementary widespread confusion is that between correlation and causation. Studies have shown repeatedly, for example, that children with longer arms reason better than those with shorter arms, but there is no causal connection here. Children with longer arms reason better because they're older! Consider a headline that invites us to infer a causal connection: BOTTLED WATER LINKED TO HEALTHIER BABIES. Without further evidence, this invitation should be refused, since affluent parents are more likely both to drink bottled water and to have healthy children; they have the stability and wherewithal to offer good food, clothing, shelter, and amenities. Families that own cappuccino makers are more likely to have healthy babies for the same reason. Making a practice of questioning correlations when reading about "links" between this practice and that condition is good statistical hygiene." (p. 137)

- a) In this example, "children with longer arms reason better than those with shorter arms," what is the explanatory variable? The response variable? The lurking variable?
- b) Explain what it means to say "there is no causal connection" between these two variables.
- c) What is "good statistical hygiene" to Paulos?

- 2) For the 20 countries with the largest population for 2009 the scatterplot shows  
x = internet users per 1000 people  
y = life expectancy (years)  
(World Almanac Book of Facts, 2009)



The correlation coefficient is 0.72, which is strong. Larger numbers of internet users per 1,000 correlate with longer life expectancy. Someone who confuses correlation with causation might suggest that an easy way to improve a country's life expectancy is to get more people onto the internet, which is a ridiculous cause-and-effect statement. Identify a lurking variable that might be explaining the strong association between life expectancy and the number of internet users per 1,000.