Definition: If the distribution of one variable is the same for all categories of a second variable, then the two variables are said to be **independent**.

- Do the variables anger level and CHD presence appear to be independent?
- Would you expect gender and political view to be independent?

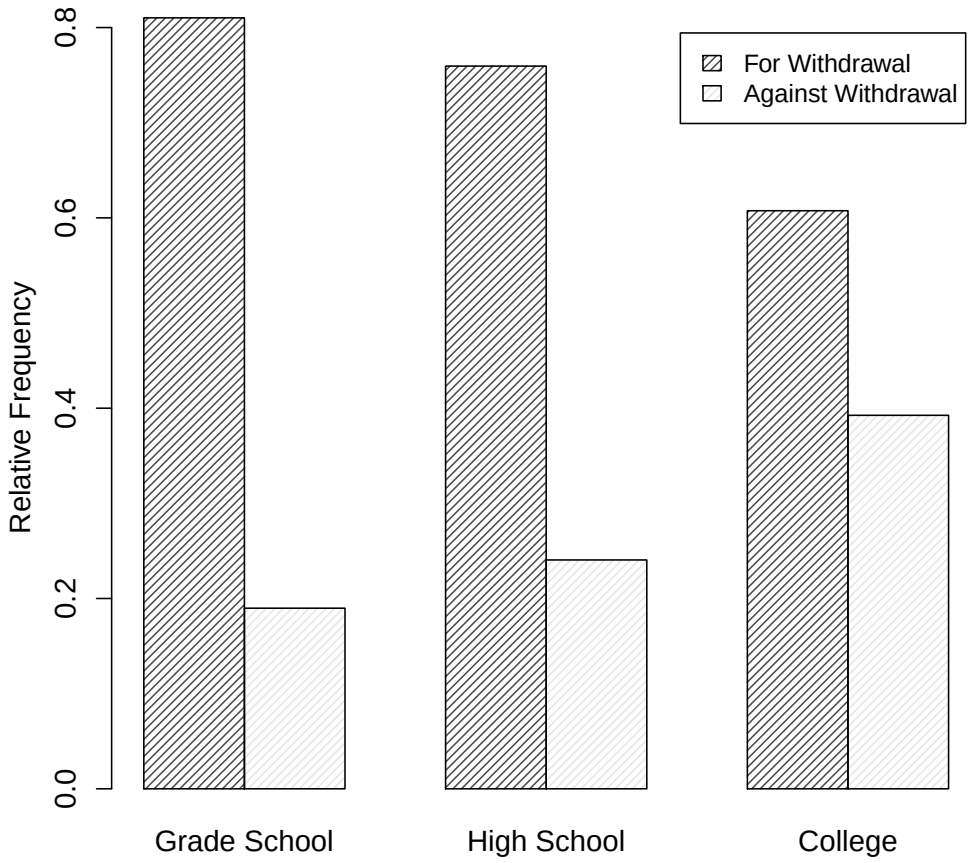
Example: A January 1971 Gallup Poll asked " A proposal has been made in Congress to require the U.S. Government to bring home all U.S. troops before the end of this year. Would you like to have your congressman vote for or against this proposal?" The two-way table below displays the results from this poll.

		For or Against	
		For Withdrawal	Against Withdrawal
Education Level	Grade School	286	67
	High School	281	89
	College	195	126

Following the **Think, Show, Tell** philosophy:

1. **THINK**: Identify the variables and the W's. Be certain that the data are counts and that the categories do not overlap so that no individual is counted twice.

2. **SHOW**: Make an appropriate display to see whether there is a difference in the relative proportions.



3. **TELL**: Interpret your findings.

Simpson's Paradox: Consider the following problem taken largely from a book exercise: Most patients who undergo surgery make routine recoveries and are discharged as planned. Others suffer excessive bleeding, infection, or other post-surgical complications and have their discharges from the hospital delayed. Suppose one of the hospitals in Missoula (let's say hospital A) takes on quite a few more surgeries than the other hospital (hospital B). Suppose we collected data to see how many surgical patients had their discharges delayed by post-surgical complications and found the results shown below.

	Discharge Delayed		
Procedure	Yes	No	Total
Hospital A	130	870	1000
Hospital B	30	270	300
Total	160	1140	1300

Are there a higher percentage of patients from Hospital A or Hospital B delayed by post-surgical complications?

The Paradox: Suppose we consider a third variable, whether or not the surgery was considered to be major or minor. A table incorporating this additional information is given on the next page:

	Major Surgery: Discharge Delayed	
	Yes	No
Hospital A	120	680
Hospital B	10	40

	Minor Surgery: Discharge Delayed	
	Yes	No
Hospital A	10	190
Hospital B	20	230

Now consider the following pair of questions:

1. Among the cases where the patient had major surgery, was there a higher percentage of discharge delays in Hospital A or Hospital B?

2. Among the cases where the patient had minor surgery, was there a higher percentage of discharge delays in Hospital A or Hospital B?

Answers:

1. 120 out of 800 or 15.00% of patients having major surgery at Hospital A had a delayed discharge, and 10 out of 50 or 20.00% of the patients having major surgery at Hospital B had a delayed discharge.

2. 10 out of 200 or 5.00% of patients having minor surgery at Hospital A had a delayed discharge, and 20 out of 250 or 8.00% of the patients having minor surgery at Hospital B experienced a delayed discharge.

What happened here?!? Although a higher overall percentage of patients had a delayed discharge at Hospital A, Hospital B had a higher percentage of delayed discharges for both Major and Minor surgeries. Can you explain this paradox?

Bottom Line: Additional variables, such as surgery severity here, can play an important role in the analysis of data, and can change our perceptions and conclusions. Variables of this type are known as **lurking** or **confounding variables**, and will be studied further in Chapters 7 & 13.