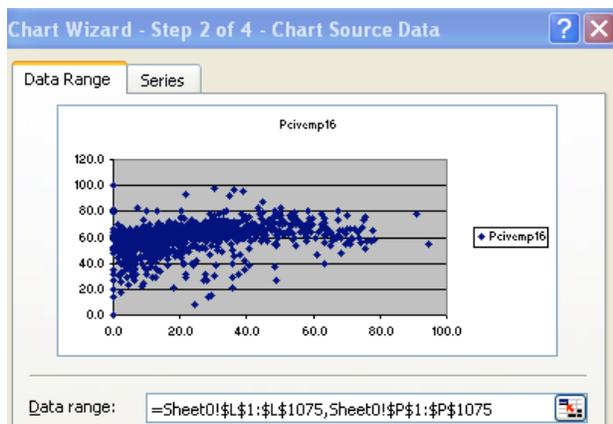


Exercise 8. Association Between Variables

Purpose: One important use of graphs is to look at the association of two or three variables. In most cases a scattergram of just two variables is created so that one can look to see if a change in one variable for an observation results in a systematic change in the second variable. Sometimes many variables are systematically paired in an array of scattergrams. One can then scan over the scattergrams to note any systematic patterns between various pairings of the variables.

Associations

1. Open *CalifCities* data file in *Excel*.
2. Select the columns labeled *Place*, *P25BAdeg*, and *Pcivemp16* so that all cells in each column are highlighted.
3. Select *Insert > Chart* and scroll down to *XY (Scatter)* in the first *Chart Wizard* window. Then click *Next*.
4. The second window will give you a quick look at your scattergram.



5. Click *Next* to move to *Step 3* of the *Chart Wizard*.
6. Under the *Titles* tab enter the title *Civilian Employed vs Education*. For the *X* axis enter *Percent Persons Age 25 and Older with a BA Degree or Higher*. For the *Y* axis

enter *Percent Persons Age 16 and Older Employed in Civilian Occupations*.

7. Under the *Legend* tab deselect the *Show legend* button. Then select *Next*.
8. At *Step 4* select the *As new sheet:* button and click *Finish*.

These two variables exhibit a fairly strong relationship since the points tend to form a positively oriented linear cluster. Thus, in cities with a higher percent of persons with at least a bachelor's degree there also tends to be higher percent of persons with civilian employment. However, for a given change in education there is not a lot of change in civilian employment.

Since there is a distinct trend in this data one might decide to apply correlation and regression methods to the variables.

While you might normally do correlation and regression in a program such as *SPSS* you can calculate the regression line in *Excel*.

9. Select *Chart > Add Trendline*.
10. In the *Add Trendline* palette under the *Type* tab click the *Linear* box if not already checked.
11. Select the *Options* tab and click the *Display equation on chart* button and then the *Display R-squared value on chart* button.

For these two variables the R^2 is 0.11 which is not very powerful for estimating employment from education. The trend line is $Y = 0.2055 + 55.041$

Exercises

1. Look through the *CalifCities* table to see if there are other pairs of variables that could be graphed in a scattergram. For example, education and income are strongly related.