**Scatterplot –**

**Including a categorical variable to designate different groups within the data**

For the voting data, make a scatterplot with different symbols for the southern states and the other states.

In this case, I choose the option to graph with Groups.

Notice that the dialog box not only gives places of the Y variable and X variable, but also for the Grouping variable. When you click on one of the empty cells to fill in, the only variables shown in the left-hand box are those which could be used. So “State” doesn’t show for the Y-variable or the X variable, because it is not a quantitative variable.
Notice that the southern states are now indicated with small squares, in red, while the non-southern states are indicated with small dots, in blue.

Later we will be interested in whether the outlier in the data – the point in the upper right, which is D.C. But, leaving aside that question for now, let’s look at the pattern shown by identifying the southern states in this relationship. Political analysts are interested in the pattern of the voting of southern states here.

We see that the data show that the southern states had somewhat lower percentages of voting Democratic in 1984 than one would expect if they were pretty much like all the other states. That is shown by the fact that their observations are all on the “lower edge” of that cloud of points.

In the early part of the 20th century, Democrats dominated over Republicans in politics in the southern states. But in the last quarter of the 20th century that support for Democrats weakened considerably.

No doubt it is relevant here that the Democratic candidate in 1980 was Jimmy Carter, who was (and still is) a southerner. But the Democratic candidate in 1984 was Walter Mondale - not a southerner. So it is not particularly surprising that the southern support for Carter was stronger than the southern support for Mondale.