# Relationships Between Variables

We’ve already seen:

* Different ways to summarizes variables in a dataset
* Different ways to visualize a variable in a dataset

What we’re going to learn:

* How to find the relationship between the variables in a dataset?

[Pivot Table](https://en.wikipedia.org/wiki/Pivot_table)

A pivot table is a table of grouped values that aggregates the individual items of a more extensive table within one or more discrete categories.

[Common pivot table uses](https://searchsqlserver.techtarget.com/definition/pivot-table)

* To identify the maximum and minimum values of a dataset
* To generate a list of unique values; e.g. which states have ordered a specific product

[Python | Pandas.pivot\_table()](https://www.geeksforgeeks.org/python-pandas-pivot_table/)

[Contingency Table or Cross Tabulation](https://humansofdata.atlan.com/2016/01/cross-tabulation-how-why/)

Cross tabulation, also known as Contingency table, is a method to quantitatively analyze the relationship between multiple variables; it is usually performed on *categorical data*.



Benefits of Cross Tabulation:

* Eliminates confusion while interpreting the data
* Helps in deriving innumerable insights

[Python | Pandas.crosstab()](https://www.geeksforgeeks.org/pandas-crosstab-function-in-python/)

\*\* When creating a cross tabulation using pandas.crosstab, you won’t get the correct answer if you use the dot syntax to specify the column (e.g. df.day). Instead use the brackets syntax (e.g. df[‘day’]).

Lab - 3

Exercise -1

1. Open “tips.csv”
2. Create a cross tabulation for the “day” of the week and “party size”.
3. Create a stacked bar graph and visualize the cross tabulation.

Exercise -2

1. Calculate the joint distribution of the table and sum over the columns;
2. Take a look at the “Marginal Distribution” in Lab 3 instructions.

Exercise -3

1. Sum up the day/size data frame over the column to get the total size for each day
2. Divide the data frame by the total counts over each row.
3. Use bar plot to show the conditional distribution.

Exercise -4

1. Compute which proportion of the parties on Saturday is of size two.
2. Compute which proportion of the two-person parties are dined on Saturday.