

## WHAT ARE "CONNECTIONS"?

The **Connections** section of DATA2GO.NYC enables users to do two things: explore the correlation between two variables across New York City's fifty-nine community districts and examine how individual community districts stack up when compared with other city districts on the given variables. **The Connections page asks and answers the question: Is there a correlation between variable x and variable y, and if so, how strong is the relationship?**

On the **Connections** section are about forty x variables, and for each x variable, a unique set of y variables. By clicking on the dots, you can obtain the (x,y) coordinates for every NYC community district. You can also compare two sets of coordinates by clicking on one dot and then passing your cursor over a second dot.

This section includes a subset of the full DATA2GO.NYC dataset; these indicators were selected based on the social science literature on these topics. Researchers interested in exploring the full range of correlations can access the complete DATA2GO.NYC datasets by clicking the "Download" tab.

One caveat is that the data come from a variety of years. While all the data were the most recent available in the spring of 2018, the original data sources conduct their studies with different frequency. This should not, however, undermine the value of the analysis for pinpointing need, identifying areas for further research, developing fact-based solutions, and more.

## WHAT IS A CORRELATION?

A correlation refers to an array of statistical measurements that describe the relationship between two variables. Correlations can be very useful to understand whether, and to what extent, one variable relates to another. They are usually a first step leading to further research on these topics.

The Connections section of DATA2GO.NYC uses Pearson's correlation, which measures the linear relationship between two variables. Pearson's correlation yields a number ranging from -1 to +1 and is represented with the symbol  $r$ . If  $r$  is positive, the linear relationship between the two variables is positive, meaning that as one variable increases across the fifty-nine NYC districts, the other also tends to increase. If  $r$  is negative, then as one variable increases, the other tends to decrease. **The further  $r$  is from zero (in either a positive or negative direction), the stronger the relationship between the two variables.**

For example, the correlation between the indicators Median Household Income (\$) and Completed at Least a Bachelor's Degree (% of adults 25+) has an  $r$  value of 0.90, meaning that, across NYC's fifty-nine community districts, these two variables have a strong, positive correlation. In plain English, we can say that as median household income increases, the share of adults with at least a bachelor's degree tends to increase. There is a strong relationship between these two variables.

## CORRELATION DOES NOT IMPLY CAUSATION

Correlations are statements of statistical relationships. The Pearson's correlation describes the linear relationship between two variables. **But it does not prove that a change in one variable causes a change in the other variable.** For example, ice cream sales and murder rates tend to rise in tandem in places with hot summers. Does that mean that ice cream consumption causes murder, or that murder makes people crave ice cream? Of course not. Research shows that they both rise in the summer months because warm weather makes ice cream a particularly appealing treat and summer is a time when people are more likely to get together and to be outside, where they come into greater contact with one another. Determining causal relationships requires extensive research and subject matter expertise.

We hope these connections spark discussion, spur understanding, and suggest new solutions for our city's greatest challenges.

**Happy exploring!**