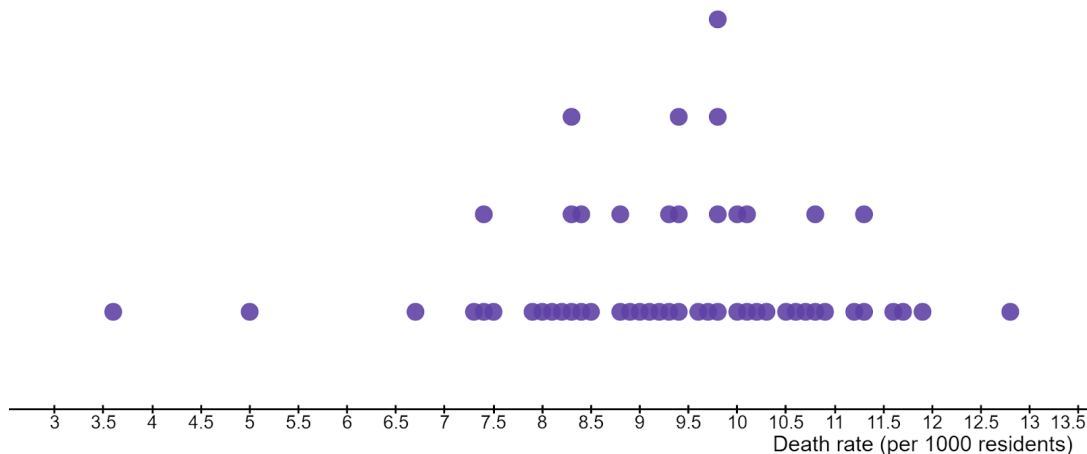


9.1: Scatterplots

Below is the death rate (per 1000 residents) for 53 randomly selected cities. Each dot represents a city. We can describe the shape, center, and spread for the distribution. Using the dotplot, we are able to answer questions about the distribution of death rates for the 53 cities. However, this dotplot does not reveal any possible explanations for the death rates. If we examine these explanations, we may be able to reduce high death rates.



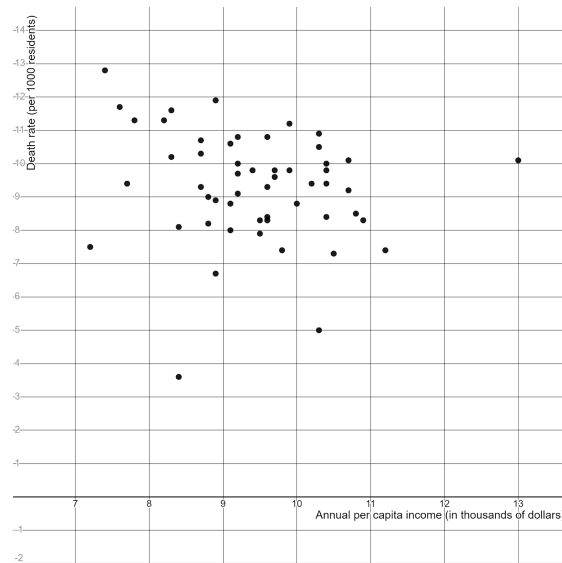
Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

We turn to a new type of graph called a scatterplot. When we have *bivariate data* (data with two quantitative variables), we can express the data in a **scatterplot**, as shown on the graph below. Scatterplots can help us determine relationships between two variables by noticing patterns.

Sometimes, if we observe a strong relationship between two variables, we use the model to make predictions. We try to predict the value of one variable using another variable. The *explanatory variable* is used to make predictions of the *response variable* values, and it is conventional to represent the explanatory variable on the horizontal axis or x-axis of the scatterplot. The response variable is then represented on the vertical axis or y-axis of the scatterplot.

Reading Scatterplots

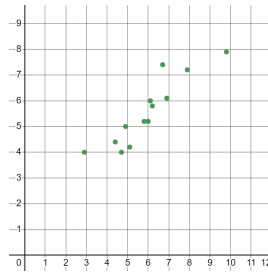
The scatterplot below describes graphically the relationship between annual per capita income (in thousands of dollars) and the death rate (per 1000 residents) of 53 randomly selected cities.



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

1. For the scatterplot above, what does a point on the scatterplot represent?
2. What are the explanatory and response variables?
3. Estimate the death rate in the city with the highest annual per capita income.
4. Estimate the annual per capita income of the city with the highest death rate.

5. What are two possible categories that could have data/results like are shown in the scatterplot below? Be creative!



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

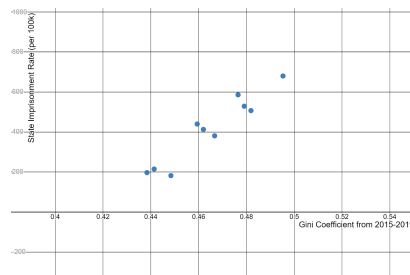
a. A point on the scatterplot could represent:

b. The x-variable could represent:

c. The y-variable could represent:

Recognizing Patterns in Data

The Gini coefficient is a ratio which quantifies the amount of inequality in a population. It is a number between 0 and 1 where 0 represents perfect equality and 1 represents perfect inequality. Below is a scatterplot that shows the Gini coefficient for 2015-2019 and the state imprisonment rate (per 100k) for 10 randomly selected states.



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

6. Does there appear to be a relationship between the two variables? Explain.

Direction

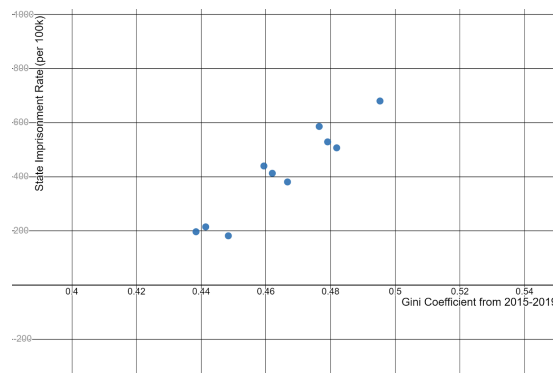
7. Do you expect states with higher Gini coefficients to have lower or higher state imprisonment rates (per 100k)? Explain.

The **direction** of a relationship is a pattern that can be recognized from a scatterplot. If the points trend upward, and as the values of x increase, so do the values of y , we say that the direction is **positive**. If the points trend downward, and as the values of x increase, the values of y decrease, we say that the direction is **negative**.

8. Think of an example of two variables whose scatterplot would have a negative direction.

Strength

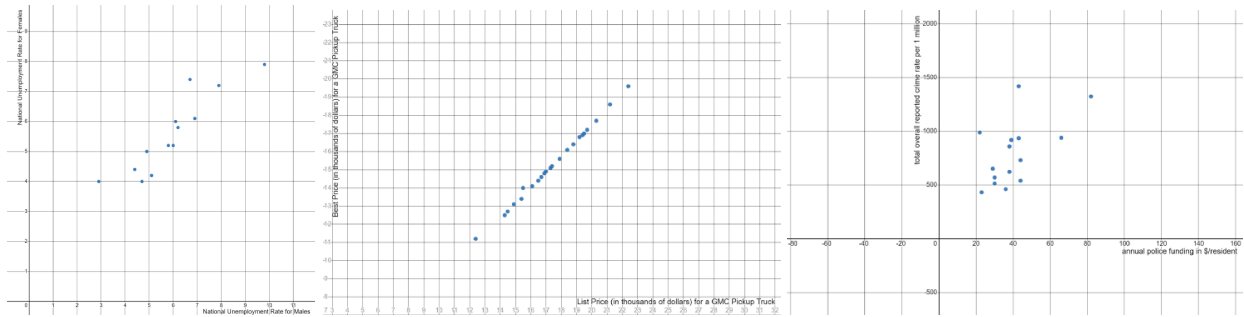
Below is a scatterplot that shows the Gini coefficient for 2015-2019 and the state imprisonment rate (per 100k) for 10 randomly selected states.



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

9. Use the pattern in the data to predict the state imprisonment rate (per 100k) for a state with a Gini coefficient of 0.52.

10. Of the following three scatterplots, which is the easiest to make predictions from? Explain.



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

The **strength** of a relationship is another pattern that can be recognized from a scatterplot. The association between two variables is considered **strong** when the points are close to some path or curve. When relationships are strong, it is easier to make predictions using the scatterplot and stronger relationships make for more accurate predictions. The association is **weak** if points are widely scattered from a path or curve.

This page titled [9.1: Scatterplots](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Hannah Seidler-Wright](#).

- [Current page](#) by Hannah Seidler-Wright is licensed [CC BY-NC-SA 4.0](#).
- [1.2: The Statistical Analysis Process](#) by Hannah Seidler-Wright is licensed [CC BY-NC-SA 4.0](#).