**Inspecting a DataFrame**

When you get a new DataFrame to work with, the first thing you need to do is explore it and see what it contains. There are several useful methods and attributes for this.

* .head() returns the first few rows (the “head” of the DataFrame).
* .info() shows information on each of the columns, such as the data type and number of missing values.
* .shape returns the number of rows and columns of the DataFrame.
* .describe() calculates a few summary statistics for each column.

homelessness is a DataFrame containing estimates of homelessness in each U.S. state in 2018. The individual column is the number of homeless individuals not part of a family with children. The family\_members column is the number of homeless individuals part of a family with children. The state\_pop column is the state's total population.

pandas is imported for you.

To better understand DataFrame objects, it's useful to know that they consist of three components, stored as attributes:

* .values: A two-dimensional NumPy array of values.
* .columns: An index of columns: the column names.
* .index: An index for the rows: either row numbers or row names.

You can usually think of indexes as a list of strings or numbers, though the pandas Index data type allows for more sophisticated options. (These will be covered later in the course.)

homelessness is available.

# Sorting rows

Finding interesting bits of data in a DataFrame is often easier if you change the order of the rows. You can sort the rows by passing a column name to .sort\_values().

In cases where rows have the same value (this is common if you sort on a categorical variable), you may wish to break the ties by sorting on another column. You can sort on multiple columns in this way by passing a list of column names.

| **Sort on …** | **Syntax** |
| --- | --- |
| one column | df.sort\_values("breed") |
| multiple columns | df.sort\_values(["breed", "weight\_kg"]) |

By combining .sort\_values() with .head(), you can answer questions in the form, "What are the top cases where…?".

homelessness is available and pandas is loaded as pd.