Considering Categorical Data

Colby Community College

Example 1

The following table summarizes two categorical variables from the `loans` data set.

		homeownership			
		rent	mortgage	own	Total
app_type	individual	3496	3839	1170	8505
	joint	362	950	183	1495
	Total	3858	4789	1353	10000

Example 1

The following table summarizes two categorical variables from the `loans` data set.

		`homeownership`			
		rent	mortgage	own	Total
`app_type`	individual	3496	3839	1170	8505
	joint	362	950	183	1495
	Total	3858	4789	1353	10000

Definition

A table that summarizes data for two categorical variables in this way is called a **contingency table**.

Example 1

The following table summarizes two categorical variables from the `loans` data set.

		homeownership			
		rent	mortgage	own	Total
app_type	individual	3496	3839	1170	8505
	joint	362	950	183	1495
	Total	3858	4789	1353	10000

Definition

A table that summarizes data for two categorical variables in this way is called a **contingency table**.

Definition

The **row totals** provide the total counts across each row.

The **column totals** provide the total counts down each column.

Note

You can also create a table that considers only a single variable.

Note

You can also create a table that considers only a single variable.

Example 2

homeownership	Count
rent	3858
mortgage	4789
own	1353
Total	10000

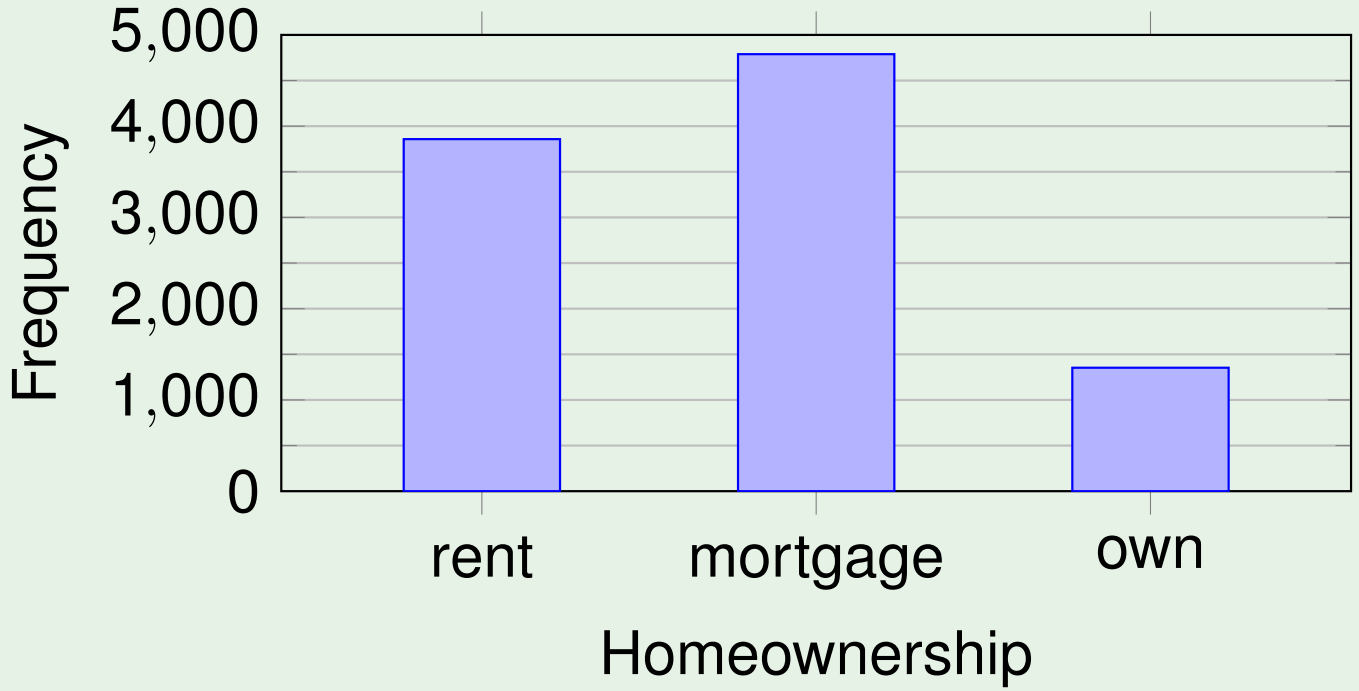
Definition

A **bar plot** plots a bar for each variable outcome, the height is the frequency of the outcome.

Definition

A **bar plot** plots a bar for each variable outcome, the height is the frequency of the outcome.

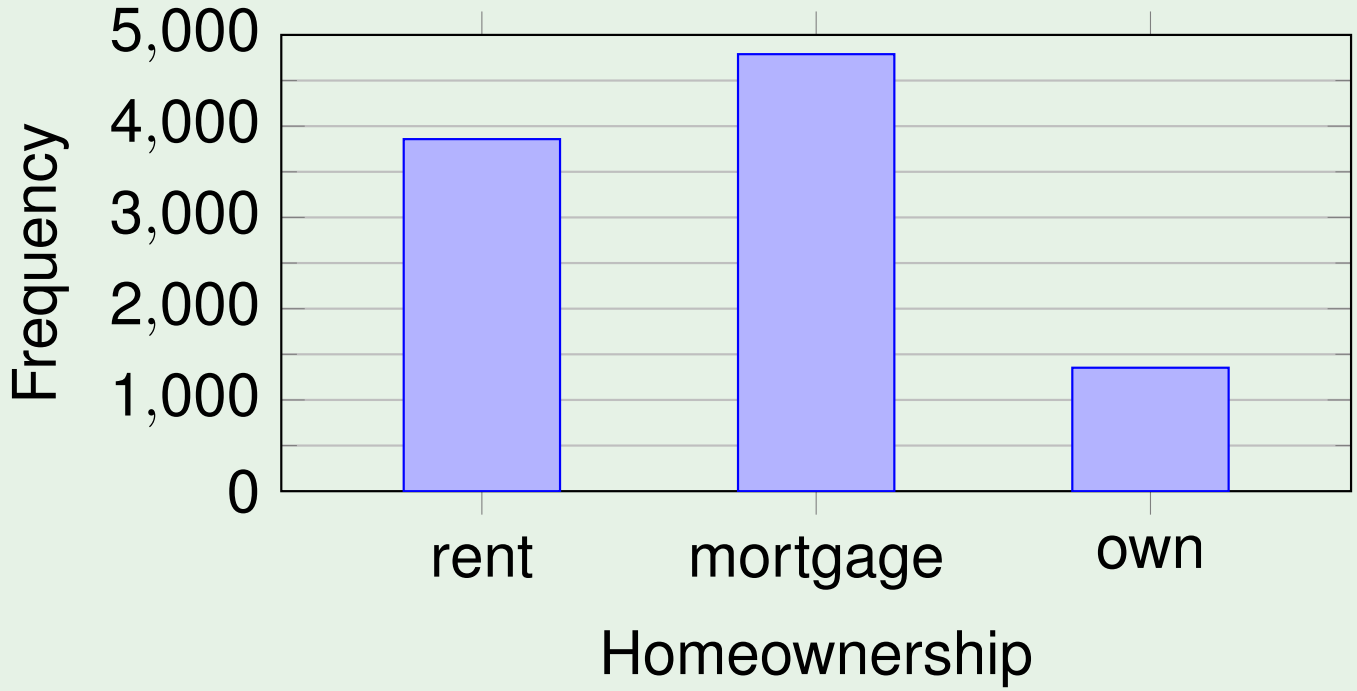
Example 3



Definition

A **bar plot** plots a bar for each variable outcome, the height is the frequency of the outcome.

Example 3



Note

A histogram has no gaps between the bars, where as bar plot does.

Note

Instead of using frequencies, we could instead use proportions.

Note

Instead of using frequencies, we could instead use proportions.

Example 4

To find the proportion, divide each frequency by the total count.

`homeownership`	Frequency	Proportion
rent	3858	
mortgage	4789	
own	1353	
Total	10000	

Note

Instead of using frequencies, we could instead use proportions.

Example 4

To find the proportion, divide each frequency by the total count.

`homeownership`	Frequency	Proportion
rent	3858	$\frac{3858}{10000} = 0.3858$
mortgage	4789	
own	1353	
Total	10000	

Note

Instead of using frequencies, we could instead use proportions.

Example 4

To find the proportion, divide each frequency by the total count.

`homeownership`	Frequency	Proportion
rent	3858	$\frac{3858}{10000} = 0.3858$
mortgage	4789	$\frac{4789}{10000} = 0.4789$
own	1353	
Total	10000	

Note

Instead of using frequencies, we could instead use proportions.

Example 4

To find the proportion, divide each frequency by the total count.

`homeownership`	Frequency	Proportion
rent	3858	$\frac{3858}{10000} = 0.3858$
mortgage	4789	$\frac{4789}{10000} = 0.4789$
own	1353	$\frac{1353}{10000} = 0.1353$
Total	10000	

Note

Instead of using frequencies, we could instead use proportions.

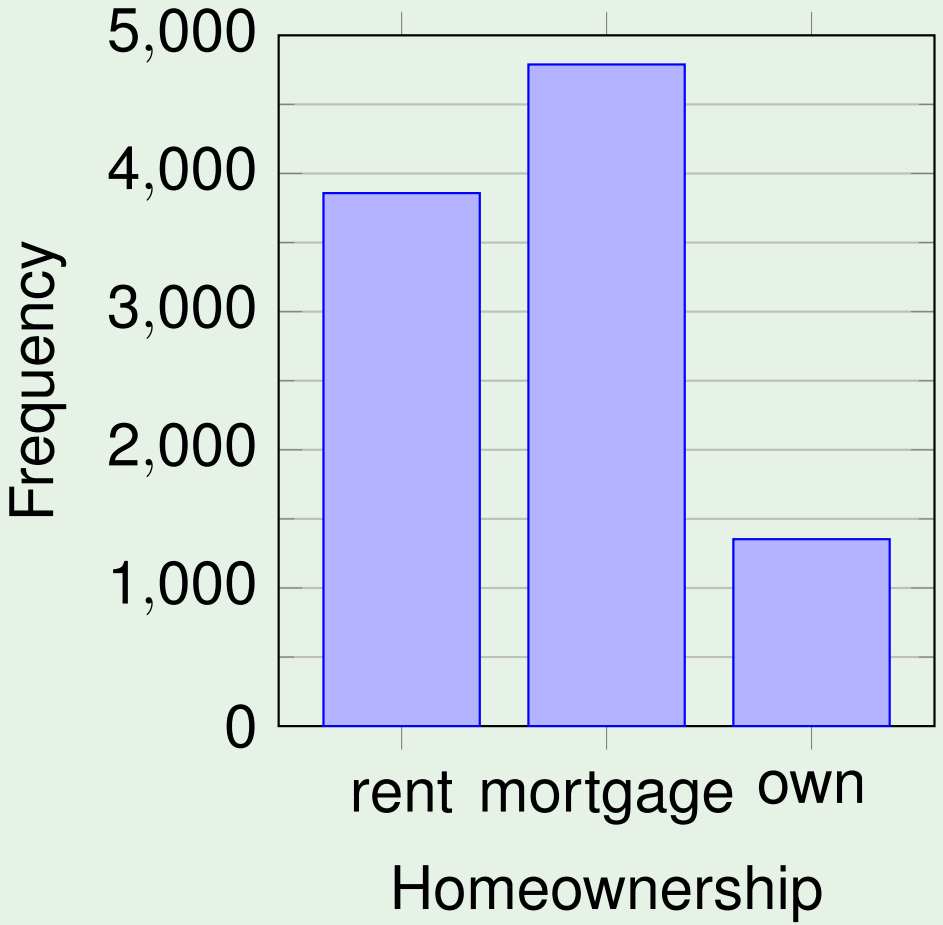
Example 4

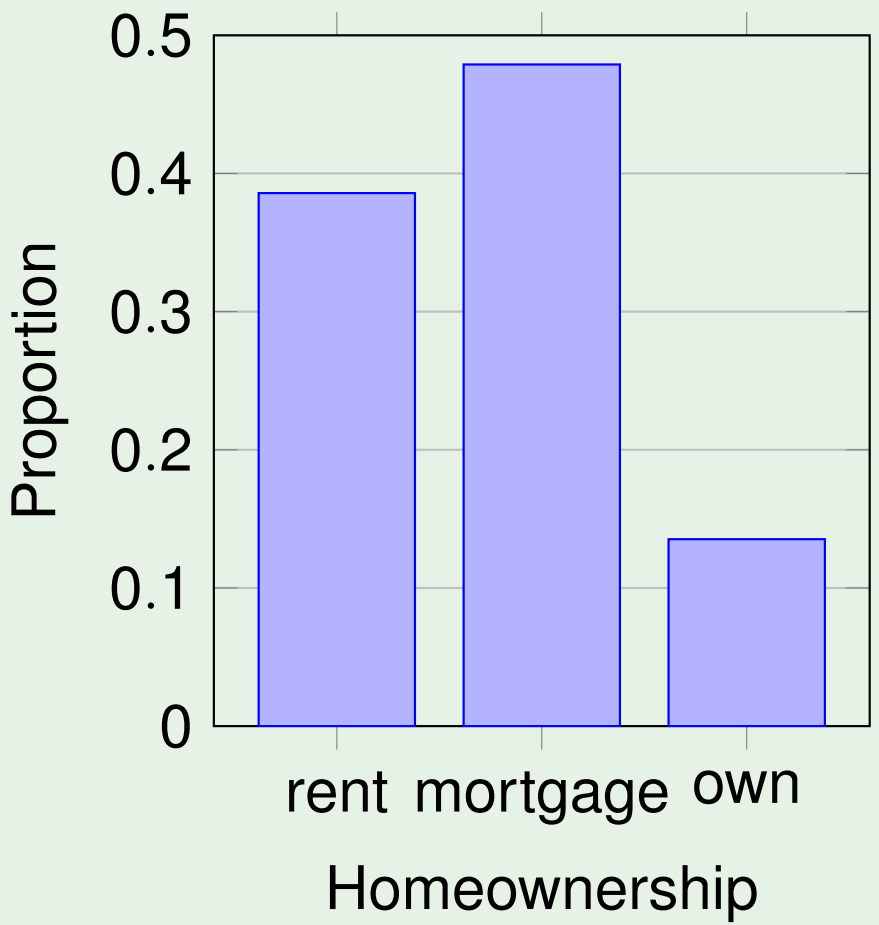
To find the proportion, divide each frequency by the total count.

`homeownership`	Frequency	Proportion
rent	3858	$\frac{3858}{10000} = 0.3858$
mortgage	4789	$\frac{4789}{10000} = 0.4789$
own	1353	$\frac{1353}{10000} = 0.1353$
Total	10000	1.0000

Example 5

Here are both the frequency and proportion for `homeownership`.





Example 6

Here, we use the **row proportions** for the contingency table from Example 1. Where we divide each number by the row total.

		homeownership			
		rent	mortgage	own	Total
app_type	individual	3496	3839	1170	8505
	joint	362	950	183	1495
	Total	3858	4789	1353	10000

↓ ↓ ↓

		homeownership			
		rent	mortgage	own	Total
app_type	individual	0.4111	0.4514	0.1376	1.0000
	joint	0.2421	0.6355	0.1224	1.0000
	Total	0.3858	0.4789	0.1353	1.0000

Example 6

Here, we use the **row proportions** for the contingency table from Example 1. Where we divide each number by the row total.

		homeownership			
		rent	mortgage	own	Total
app_type	individual	3496	3839	1170	8505
	joint	362	950	183	1495
	Total	3858	4789	1353	10000

↓ ↓ ↓

		homeownership			
		rent	mortgage	own	Total
app_type	individual	0.4111	0.4514	0.1376	1.0000
	joint	0.2421	0.6355	0.1224	1.0000
	Total	0.3858	0.4789	0.1353	1.0000

What does the number 0.4111 represent?

Example 6

Here, we use the **row proportions** for the contingency table from Example 1. Where we divide each number by the row total.

		homeownership			
		rent	mortgage	own	Total
app_type	individual	3496	3839	1170	8505
	joint	362	950	183	1495
	Total	3858	4789	1353	10000

↓ ↓ ↓

		homeownership			
		rent	mortgage	own	Total
app_type	individual	0.4111	0.4514	0.1376	1.0000
	joint	0.2421	0.6355	0.1224	1.0000
	Total	0.3858	0.4789	0.1353	1.0000

What does the number 0.4111 represent?

That 41.11% of those that applied as individuals are renters.

Example 7

Here, we use the **column proportions** for the contingency table from Example 1. Where we divide each number by the column total.

		homeownership			
		rent	mortgage	own	Total
app_type	individual	3496	3839	1170	8505
	joint	362	950	183	1495
	Total	3858	4789	1353	10000

↓ ↓ ↓

		homeownership			
		rent	mortgage	own	Total
app_type	individual	0.9062	0.8016	0.8647	0.8505
	joint	0.0946	0.1984	0.1353	0.1495
	Total	1.0000	1.0000	1.0000	1.0000

Example 7

Here, we use the **column proportions** for the contingency table from Example 1. Where we divide each number by the column total.

		homeownership			
		rent	mortgage	own	Total
app_type	individual	3496	3839	1170	8505
	joint	362	950	183	1495
	Total	3858	4789	1353	10000

↓ ↓ ↓

		homeownership			
		rent	mortgage	own	Total
app_type	individual	0.9062	0.8016	0.8647	0.8505
	joint	0.0946	0.1984	0.1353	0.1495
	Total	1.0000	1.0000	1.0000	1.0000

What does the number 0.9062 represent?

Example 7

Here, we use the **column proportions** for the contingency table from Example 1. Where we divide each number by the column total.

		homeownership			
		rent	mortgage	own	Total
app_type	individual	3496	3839	1170	8505
	joint	362	950	183	1495
	Total	3858	4789	1353	10000

↓ ↓ ↓

		homeownership			
		rent	mortgage	own	Total
app_type	individual	0.9062	0.8016	0.8647	0.8505
	joint	0.0946	0.1984	0.1353	0.1495
	Total	1.0000	1.0000	1.0000	1.0000

What does the number 0.9062 represent?

That 90.62% of renters applied as individuals.

Do Associations next.

Note

We ca