Organizing Qualitative Data

Recall: qualitative variables take non-numeric values.

Def: The number of times a particular value occurs is called its frequency or count.

Frequency Distributions

Def: A frequency distribution of qualitative data is a listing of the distinct values and their frequencies.

To construct a freqency distribution,

1. List the distinct values of the observations in the data in the first column of a table.
2. Count the number of observations for each distinct value and place them in the second column.

Suppose I asked 21 people whether they prefer chocolate or vanilla.

- Chocolate
- Vanilla
- Vanilla
- Chocolate
- Chocolate
- Chocolate
- Vanilla

- Chocolate
- Vanilla
- Chocolate
- Chocolate
- Vanilla
- Chocolate
- Chocolate

- Vanilla
- Chocolate
- Chocolate
- Chocolate
- Chocolate
- Vanilla
- Chocolate

Create a frequency distribution for these data:

Step 1: List the distinct values of the observations in the first column on a table.

Flavor Preference	**Frequency**
Chocolate	
Vanilla	

Step 2: count the number of observations for each distinct value.

- Chocolate
- Vanilla
- Vanilla
- Chocolate
- Chocolate
- Chocolate
- Vanilla

- Chocolate
- Vanilla
- Chocolate
- Chocolate
- Vanilla
- Chocolate
- Chocolate

- Vanilla
- Chocolate
- Chocolate
- Chocolate
- Chocolate
- Vanilla
- Chocolate

Placing these counts in the second column,

Flavor Preference	**Frequency**
Chocolate	14
Vanilla	7

Relative Frequency Distributions

Def: The relative frequency is the ratio of the frequency (count) to the total number of observations.

$$\text{relative frequency} = \frac{\text{frequency}}{\text{number of observations}}$$

Def: A relative frequency distribution of qualitative data is a listing of the distinct values and their relative frequencies.

Flavor Preference	**Frequency**
Chocolate	14
Vanilla	7

- Relative frequency of chocolate $= \frac{14}{21} = \frac{2}{3}$.
- Relative frequency of vanilla $= \frac{7}{21} = \frac{1}{3}$.

So the relative frequency distribution is

Flavor Preference	**Relative Frequency**
Chocolate	$2/3 \approx 0.667$
Vanilla	$1/3 \approx 0.333$

Pie Charts

Def: A pie chart is a disk divided into wedge-shaped pieces proportional to the relative frequency to the qualitative data.

To construct a pie chart:

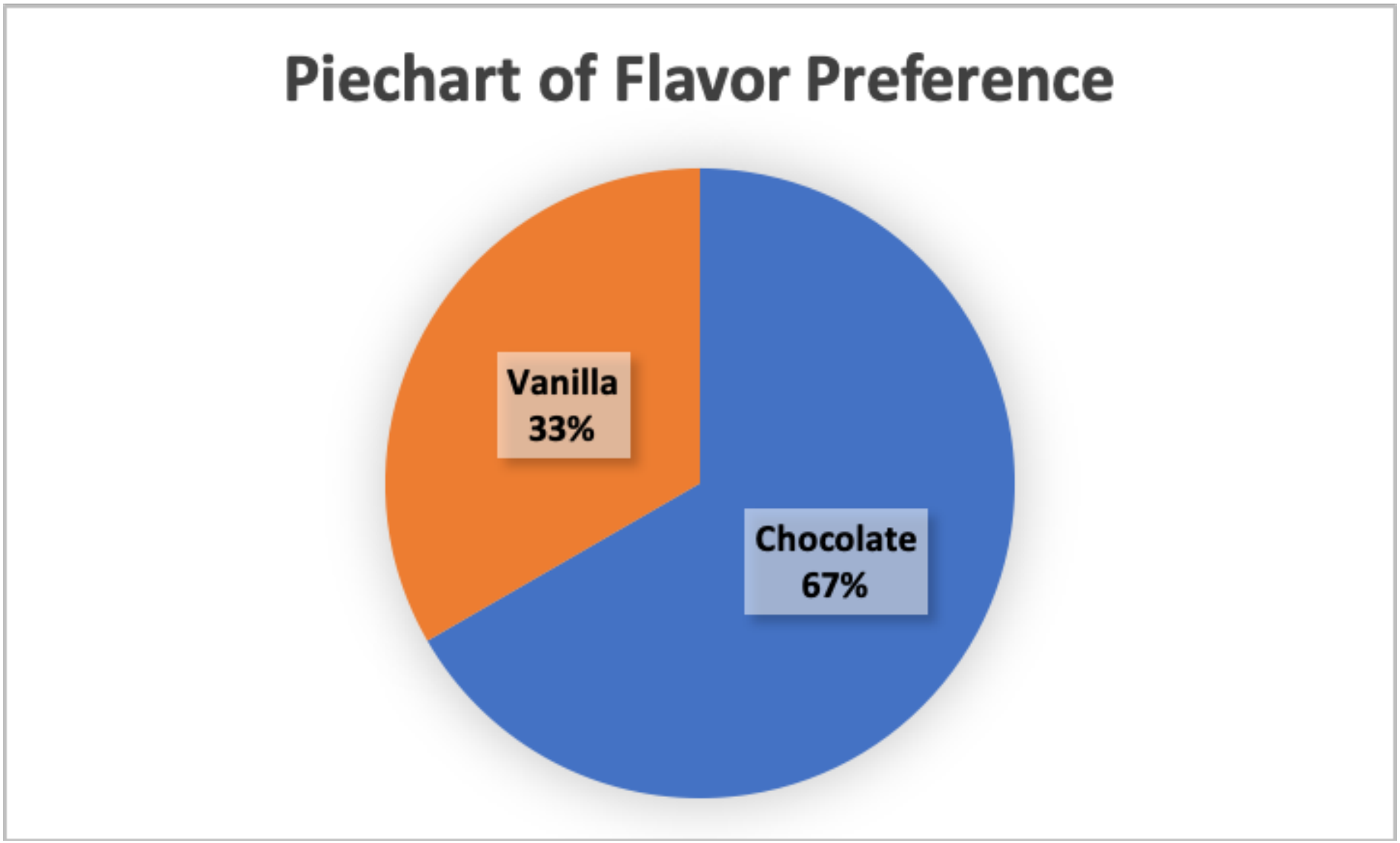
1. Obtain a relative frequency distribution of the data.
2. Divide the disk into wedge-shaped pieces proportional to the relative frequencies.
   - Since a circle has 360°, we multiply each relative frequency by 360 to get the angle for each wedge.
3. Label each wedge with the distinct values and their relative frequencies.

Earlier, we came up with the following frequency distribution:

Flavor Preference	**Relative Frequency**
Chocolate	$2/3 \approx 0.667$
Vanilla	$1/3 \approx 0.333$

To divide into wedges,

- Chocolate has angle $2/3 \times 360° = 240°$.
- Vanilla has angle $1/3 \times 360° = 120°$.

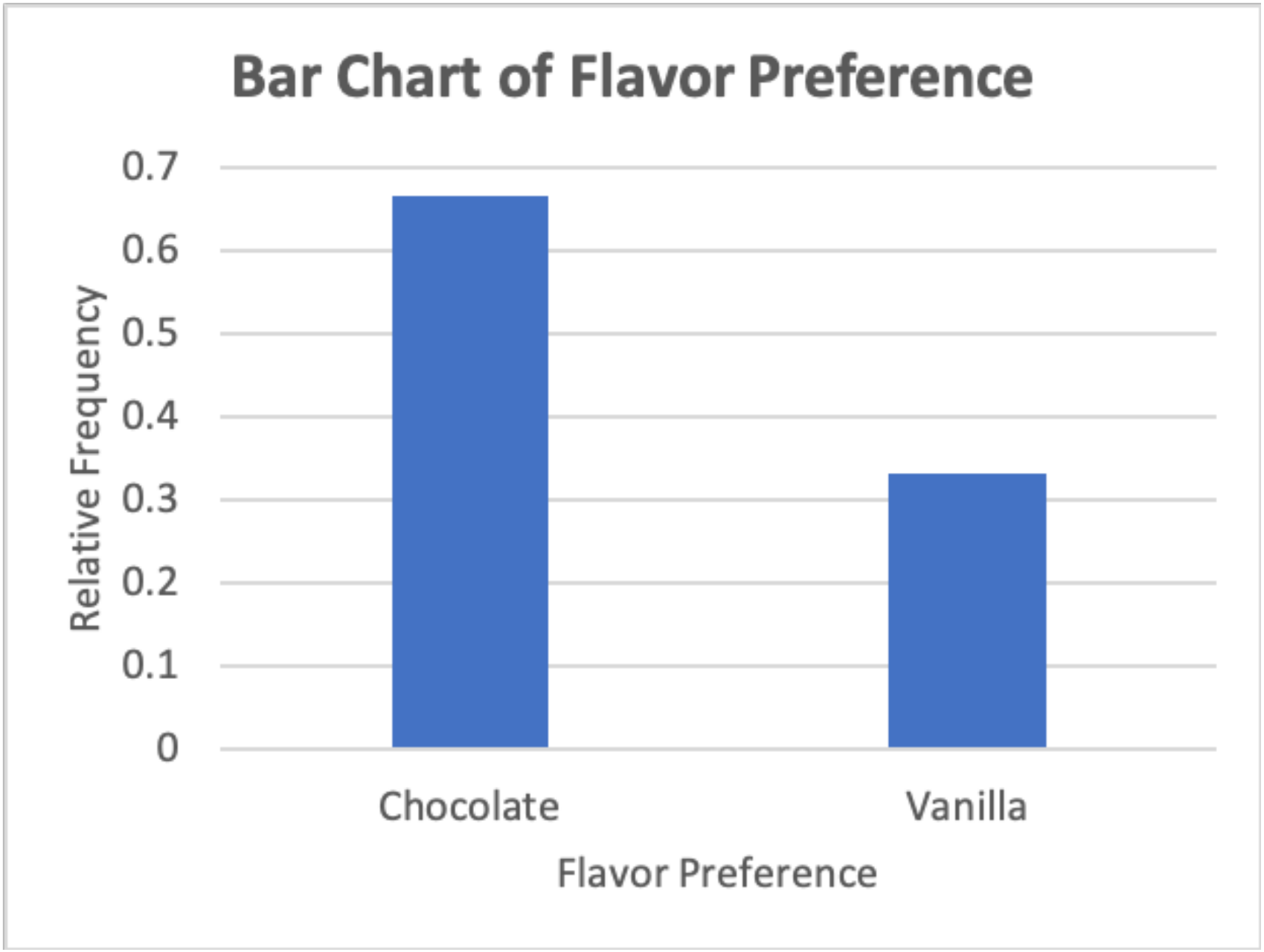

Piechart of Flavor Preference
Vanilla
33%
Chocolate
67%

Bar Charts

Def: A bar chart displays the distinct values of the qualitative data on a horizontal axis and relative frequencies (or frequencies or percents) of those values on a vertical axis.

The relative frequency of each distinct value is represented by a vertical bar whose height is equal to the relative frequency of that value.

The bars should be positioned so they do not touch each other.


Bar Chart of Flavor Preference
Relative Frequency
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Chocolate
Vanilla
Flavor Preference

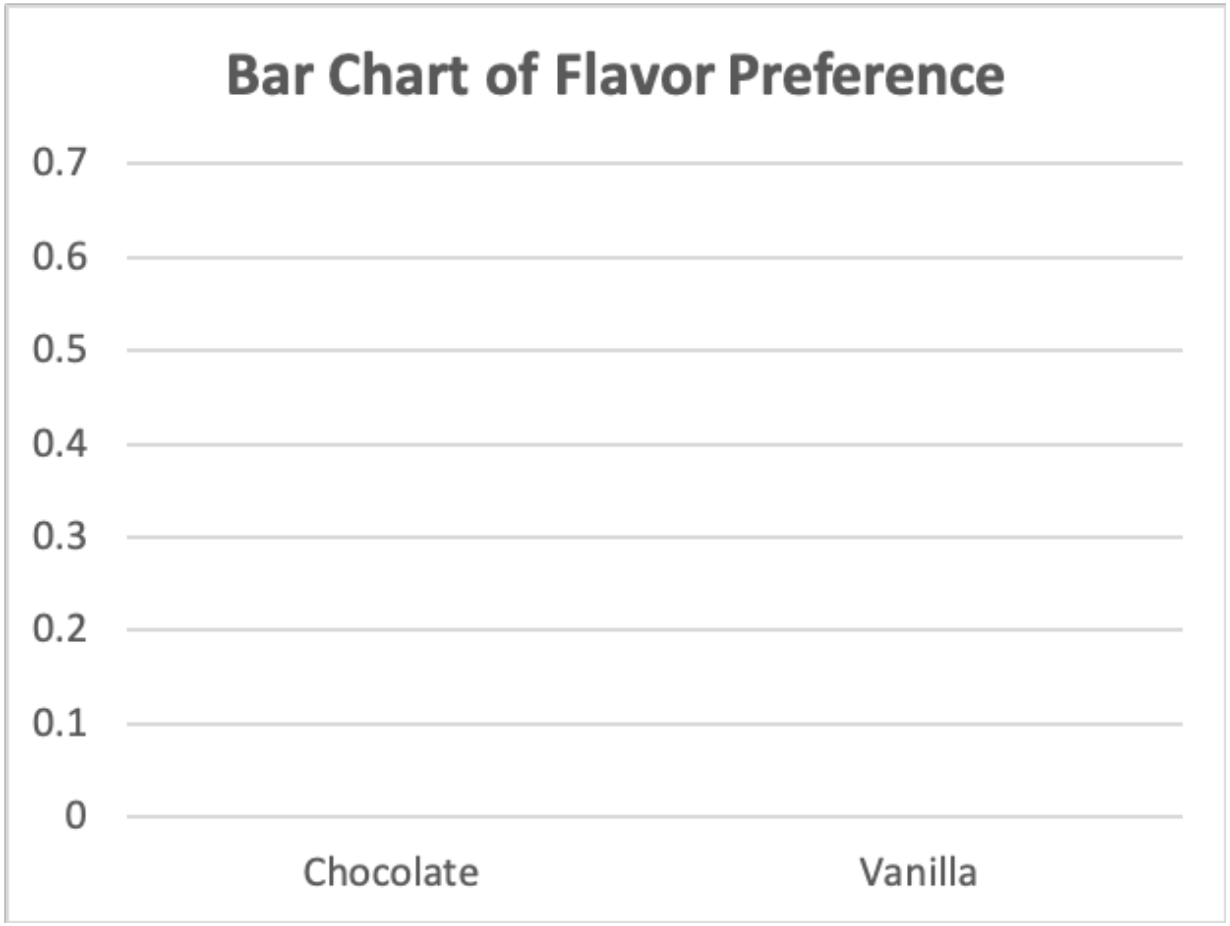
To construct a bar chart,

1. Obtain a relative frequency distribution of the data.
2. Draw a horizontal axis on which to place the bars and a vertical axis on which to display the relative frequencies.
3. For each distinct value, construct a vertical bar whose height equals the relative frequency of that value.
4. Label the bars with the distinct values. Label your axes.

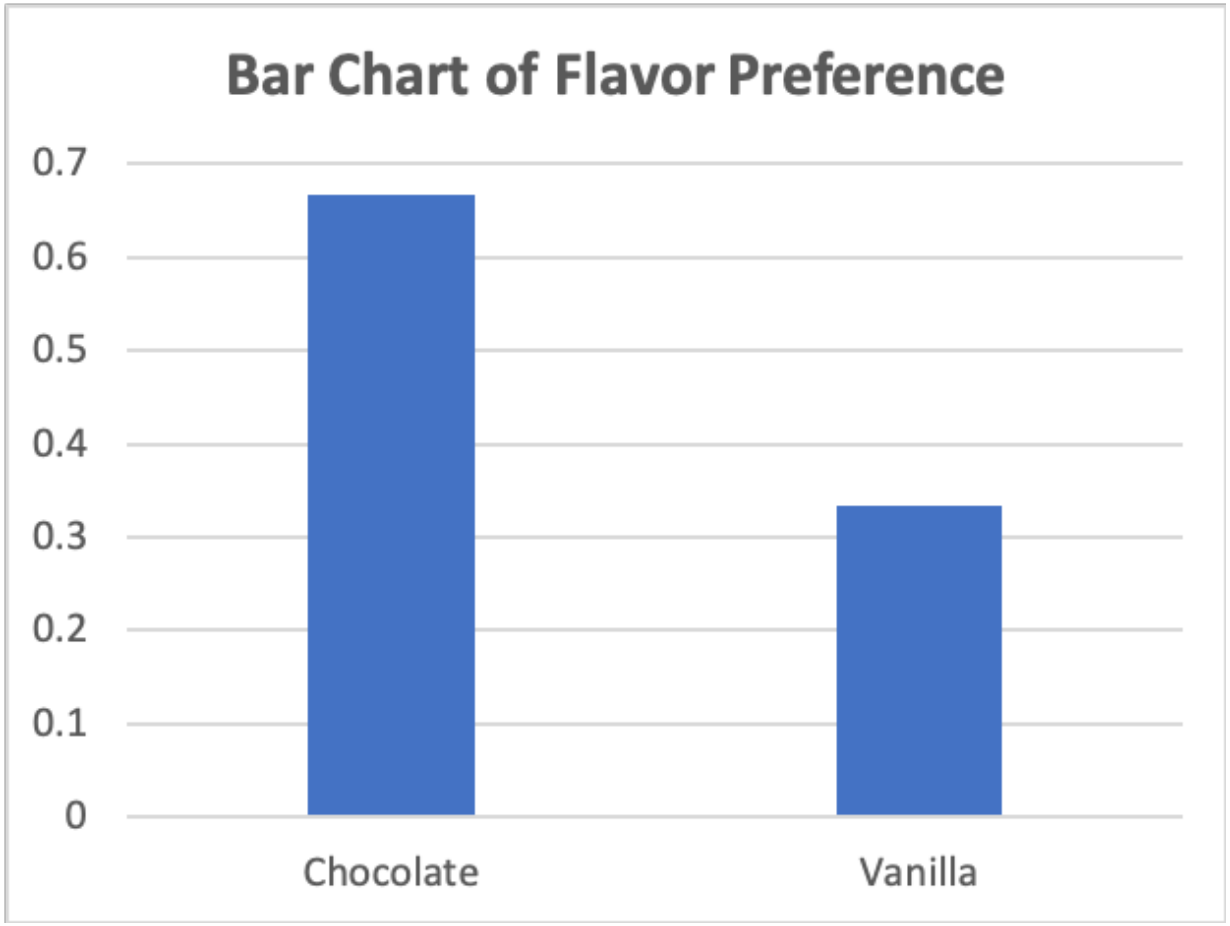
Step 1: Obtain a relative frequency distribution of the data.

Flavor Preference	**Relative Frequency**
Chocolate	$2/3 \approx 0.667$
Vanilla	$1/3 \approx 0.333$

Step 2: Draw a horizontal axis on which to place the bars and a vertical axis on which to display the relative frequencies.



Step 3: For each distinct value, construct a vertical bar whose height equals the relative frequency of that value.



Step 4: Label everything.