

# Frequency Tables and Histograms

## Steps to Creating and Evaluating a Histogram

**Step 1.** Source relevant data cube. In this example we have downloaded the 'Homeless per state under the age of 18' data cube from the 2016 Australian Bureau of Statistics Census.

This data will be used to calculate the mean, median and range for the whole data set.

Table 1. Homeless per State under the age of 18.

|          | NSW   | VIC   | QLD   | SA  | WA    | TAS | NT    | ACT |
|----------|-------|-------|-------|-----|-------|-----|-------|-----|
| Under 12 | 3,963 | 3,372 | 2,979 | 804 | 1,208 | 212 | 3,132 | 183 |
| 12-18    | 2,677 | 2,010 | 1,710 | 638 | 741   | 162 | 1,899 | 109 |

**Step 2.** To calculate the **mean**, add up all of the numbers and then divide by the number of numbers (the count).

$$3,963 + 3,372 + 2,979 + 804 + 1,208 + 212 + 3,132 + 183 + 2,677 + 2,010 + 1,710 + 638 + 741 + 162 + 1,899 + 109 = 25,799$$

We've added up 16 numbers, so we divide by 16 to get the mean:

$$25,799 / 16 = 1,612 \text{ (rounded to the nearest whole number)}$$

The mean is 1,612.

**Step 3.** Calculate the **median**. This can be done by ordering the numbers by size, finding the middle numbers, and then taking the mean of those.

109; 162; 183; 212; 638; 741; 804; | **1,208; 1,710**; | 1,899; 2,010; 2,677; 2,979; 3,132; 3,372; 3,963

The middle numbers are 1,208 and 1,710.

Their mean is  $(1,208 + 1,710) / 2 = 1,459$

The median of the whole set is 1,459.

**Step 4.** Calculate the **range** by subtracting the lowest number from the highest number.

$$3,963 - 109 = 3,854$$

The range is 3,854.

**Step 5.** Create a **frequency table** for the data to determine a range for the set of numbers.

The frequency table for the "Homeless under the age of 18" data would look like Table 2.

Table 2. Homelessness Frequency per 400

| Range       | Frequency |
|-------------|-----------|
| 0-400       | 4         |
| 401-800     | 2         |
| 801-1,200   | 1         |
| 1,201-1,600 | 1         |
| 1,601-2,000 | 2         |
| 2,001-2,400 | 1         |
| 2,401-2,800 | 1         |
| 2,801-3,200 | 2         |
| 3,201-3,600 | 1         |
| 3,601-4,000 | 1         |

**Step 6.** Create your histogram using the the range as the x-axis, and the frequency as the y-axis.

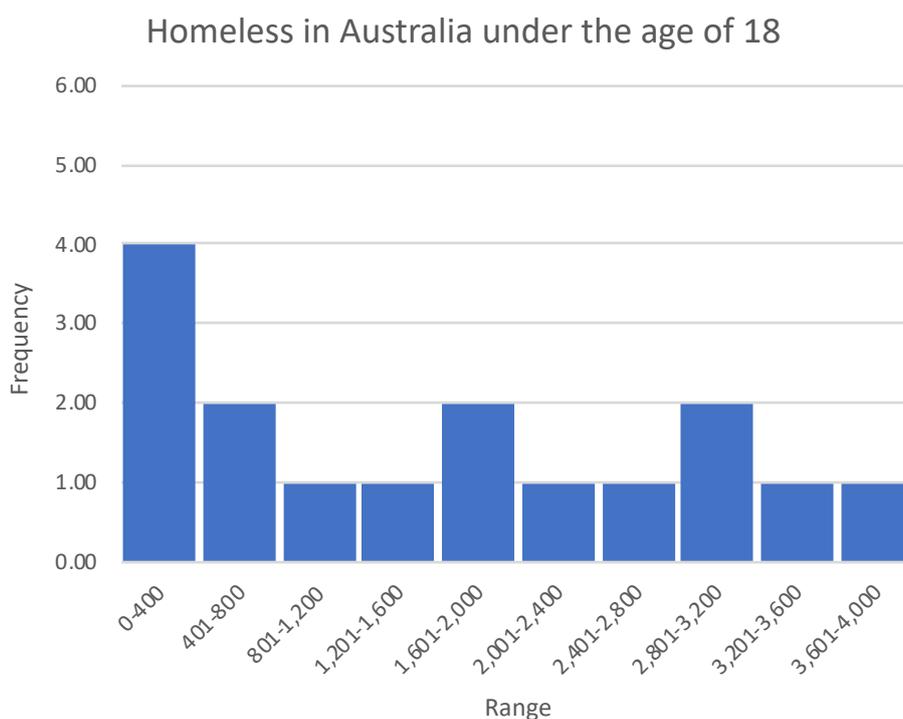
**Step 7.** Consider if the histogram is skewed, symetrical or bi-modal.

A histogram is skewed right if the bars towards the left are taller and there is a tail to the right.

A histogram is skewed left if the bars towards the right are taller and there is a tail to the left.

A histogram is symmetrical if it is similar (doesn't need to be exact) on both sides.

A histogram is bi-modal if it has two 'peaks' - i.e., two values occur more frequently than others.



The histogram above is skewed right, which means that the lowest values have the highest frequencies.

When a histogram is skewed right, the mean is greater than the median, which corresponds to our calculations from Steps 2 and 3.

Data Source: ABS, [Census of Population and Housing: Estimating homelessness](#), 2016