More on bar diagram and histogram

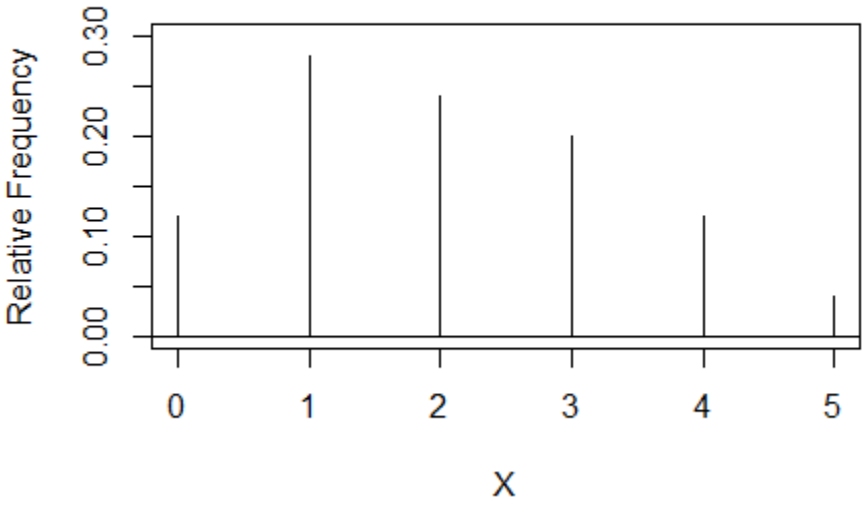
Example (discrete variable):

Let $X =$ number of machine failures in an area in a week. Let $n = 50$.
Data: $2, 4, 1, 0, \cdots, 1, 3$.

Frequency table:

x	Frequency	Relative Frequency
0	6	0.12
1	14	0.28
2	12	0.24
3	10	0.20
4	6	0.12
5	2	0.04
Total	50	1.00

Bar Diagram



- We can call it bar diagram though we used lines instead of bars. We should NOT call it a line diagram. Line diagrams are different.
- We used relative frequencies instead of frequencies (to compare with probabilities later).
- Height of a bar (line) is proportional to the relative frequency.

Example (continuous variable):

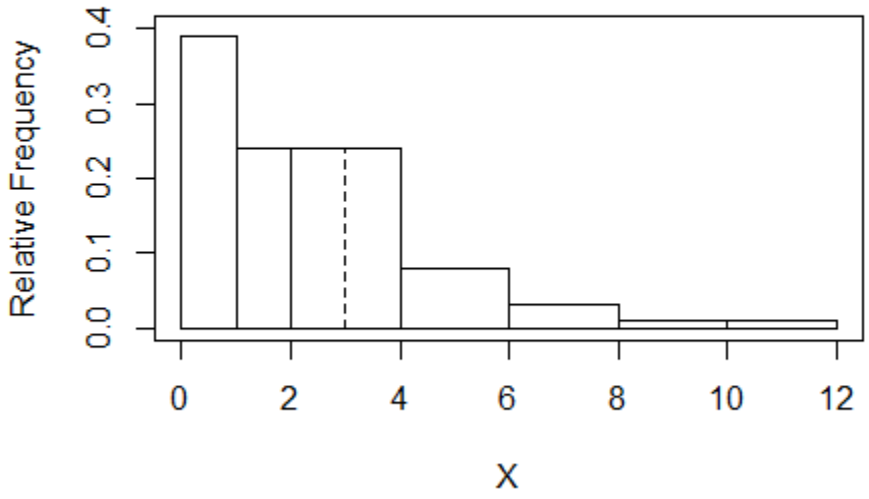
Let $X =$ lifetime (years) of an electric component. Let $n = 100$.

Data: $2.85, 0.54, 1.29, 11.23, \cdots, 3.55$.

Frequency table (class intervals with unequal widths):

Class	Frequency	Relative Frequency	Density = rel. fr. / class width
0 – 1	39	0.39	0.39
1 – 2	24	0.24	0.24
2 – 4	24	0.24	0.12
4 – 6	8	0.08	0.04
6 – 8	3	0.03	0.015
8 – 10	1	0.01	0.005
10 – 12	1	0.01	0.005
Total	100	1.00	Total is meaningless

Incorrect Histogram



In the above histogram, "height" of a bar is proportional to relative frequency. The class '$1-2$' has relative frequency 0.24. The class '$2-4$' also has relative frequency 0.24. These two classes received bars of equal heights (though the class '$2-4$' has more width). Why is it wrong? To understand this, a vertical dotted line is drawn at $X = 3$. Even if this dotted line is not drawn, people will imagine it subconsciously.

Then, people will think that there are equal number of values in '$1-2$', '$2-3$' and '$3-4$'. That is, they will thank that 24% of all values are in '$2-3$' and an additional 24% are in '$3-4$', which is clearly not the case.

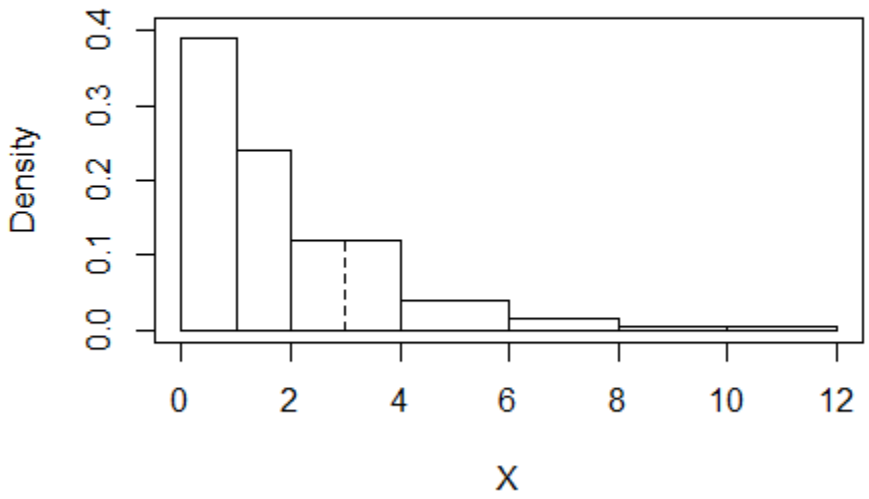
In the following histogram, "area" of a bar is proportional to relative frequency. This means, height does NOT represent relative frequency and we should NOT label the y-axis "relative frequency" or "frequency". Height of a bar represents 'density' which is defined as:

$$\text{density} = \frac{\text{relative frequency}}{\text{class width}}$$

That is,

$$\underset{\text{(area of bar)}}{\text{relative frequency}} = \underset{\text{(height of bar)}}{\text{density}} \times \underset{\text{(width of bar)}}{\text{class width}}$$

Correct Histogram



Here, people will think that 12% of all values are in '$2-3$' and an additional 12% in '$3-4$'. This is a simplified extension of the fact that 24% of values are in '$2-4$'.

We discussed an example with unequal class intervals. When all the class intervals are equal, density is proportional to relative frequency. However, we should use 'density' in both cases to maintain similarity and to make probability discussion easier.