

Definition 1 (Histogram). A visual way of showing the frequencies (counts) of the number of datapoints within a certain range.

To make one, first decide on the bin width (size of the interval). This may change for every dataset.

Remark (1): The general the interval size (width) is given in question.

Make a table on the *relative frequency distribution* (how common are these values compared to the total). This table usually displays the class interval (bins), class midpoint, frequency (count) of points falling within the interval, and the relative frequency (frequency divided by the sample size). We can draw the histogram in two ways.

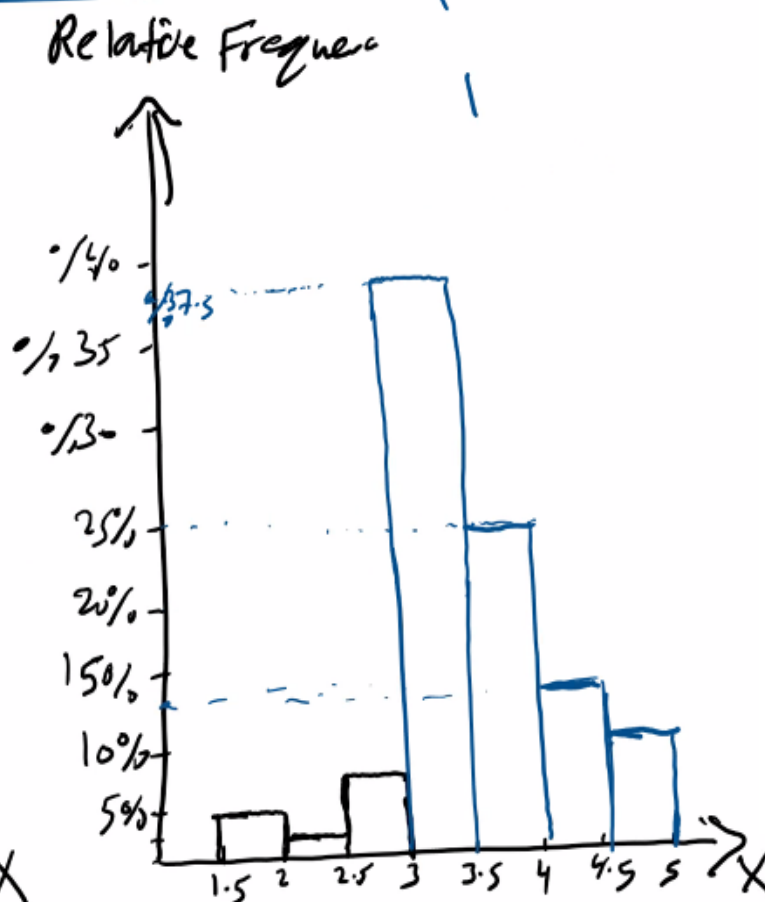
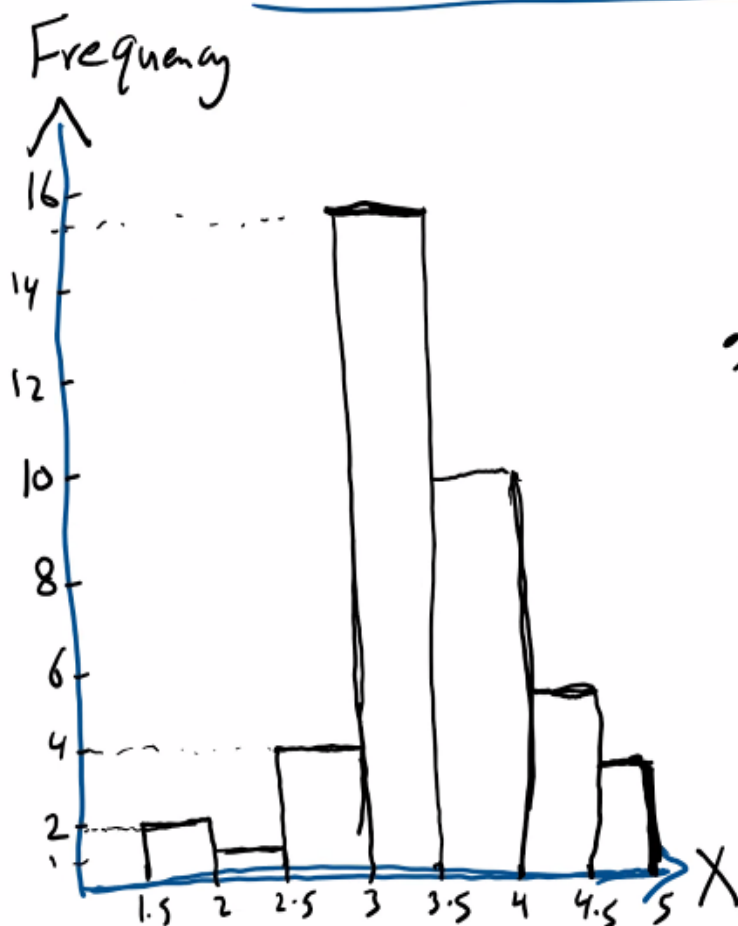
1. use midpoints on the x -axis, and use frequency on the y -axis
2. use midpoints on the x -axis, and use relative frequency on the y -axis

Example 1 (Car Batteries). The following data specifies the “life” of 40 similar car batteries recorded to the nearest tenth of a year. The batteries are guaranteed to last 3 years. (Data from pg 21 of our textbook.)

1.6	2.6	3.1	3.2	3.4	3.7	3.9	4.3
1.9	2.9	3.1	3.3	3.4	3.7	3.9	4.4
2.2	3.0	3.1	3.3	3.5	3.7	4.1	4.5
2.5	3.0	3.2	3.3	3.5	3.8	4.1	4.7
2.6	3.1	3.2	3.4	3.6	3.8	4.2	4.7

Draw a histogram for this dataset starting at 1.5 and using widths of 0.5.

Class interval	Frequency	Relative Frequency
[1.5, 2)	2	$\frac{2}{40} = 0.05 = 5\%$
[2, 2.5)	1	$\frac{1}{40} = 0.025 = 2.5\%$
[2.5, 3)	4	$\frac{4}{40} = 0.10 = 10\%$
[3, 3.5)	5	$\frac{5}{40} = 0.125 = 12.5\%$
[3.5, 4)	10	$\frac{10}{40} = 25\%$
[4, 4.5)	5	$\frac{5}{40} = 12.5\%$
[4.5, 5)	3	$\frac{3}{40} = 7.5\%$
	40	100%



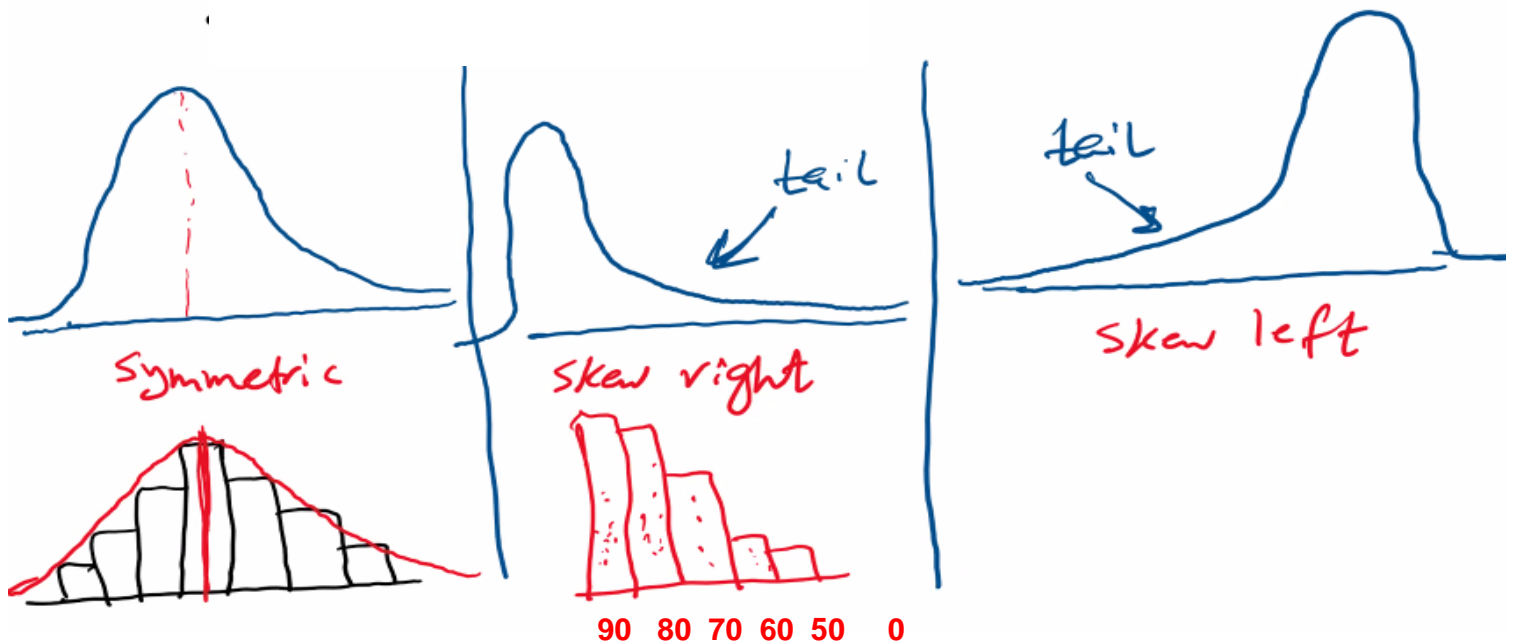
Q 1. How can we describe our data?

Look at:

- 1) Shape (Today)
- 2) Center
- 3) Spread (Chapter 3)

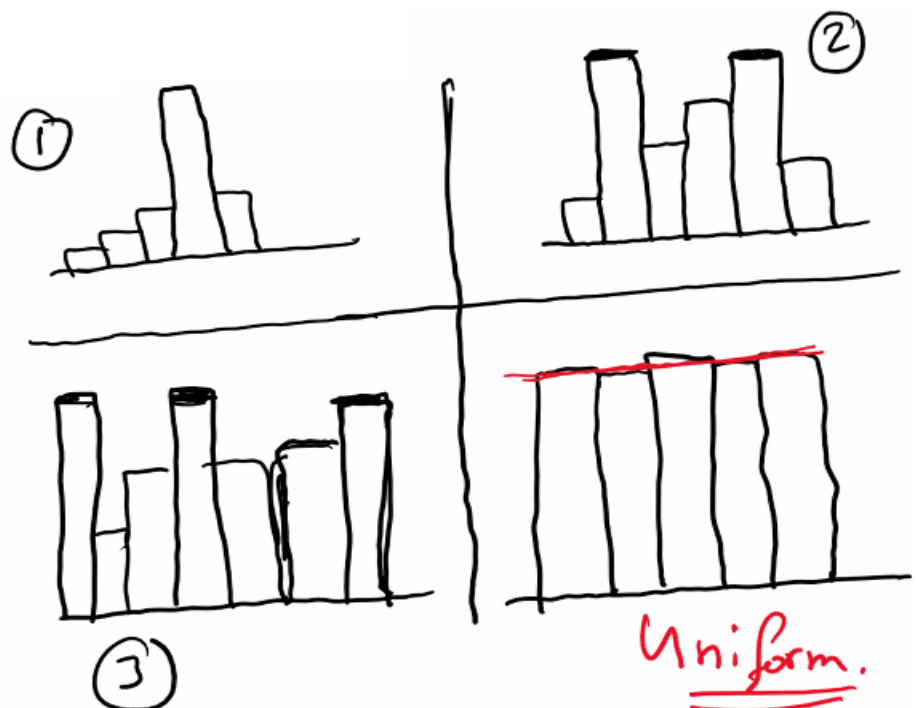
Shape

- Does the picture look symmetric or skewed?



- Does the picture have humps / modes?

- 1) 1 Peak: **Unimodal**
- 2) 2 Peaks: **Bimodal**
- 3) More than 2 peaks: **Multimodal**
- 4) Uniform: All approx. same height



- Are there any unusual features?

Definition 2 (Outliers). Points that are far away from the rest of the points.



Example 2. item The following scores represent the final examination grades for an elementary statistics course:

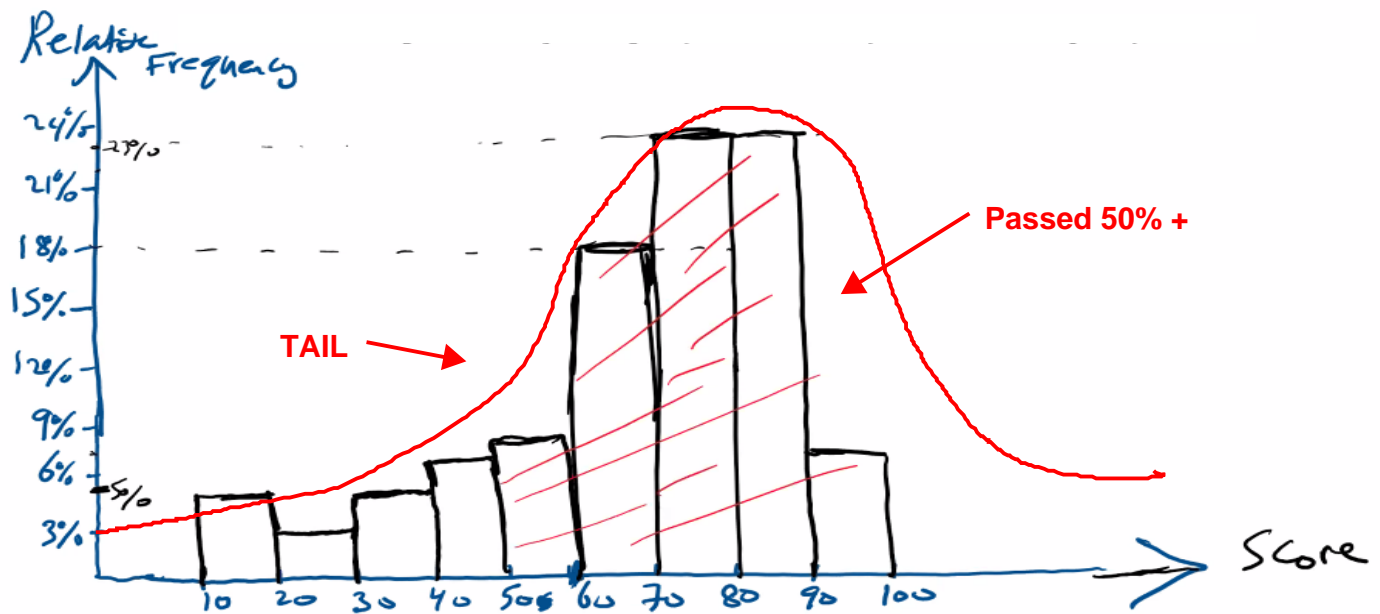
	23	60	79	32	57	74	52	70	82
	36	80	77	81	95	41	65	92	85
	55	76	52	10	64	75	78	25	80
Max Value →	98	81	67	41	71	83	54	64	72
	88	62	74	43	60	78	89	76	84
	48	84	90	15	79	34	67	17	82
	69	74	63	80	85	61			

← n = 60

1. Construct a relative frequency table. Start your intervals at 10, and use bins of size 10.

Class interval	Frequency	Relative Frequency
$[10, 20) \equiv 16-19$	3	$3/60 = 0.05 = 5\%$
$[20, 30)$	2	$2/60 = 0.03 = 3\%$
$[30, 40)$	3	5%
$[40, 50)$	4	7%
$[50, 60)$	5	8%
$[60, 70)$	11	18%
$[70, 80)$	14	23%
$[80, 90)$	14	23%
$[90, 100)$	4	7%
	60	

2. Draw the histogram corresponding to your results in your relative frequency table.



3. Is the histogram skewed? If so, is it skewed to the left or right?

Left skewed

4. If you needed to make a report of the center and spread of the data, what statistics should you use? (Examples: mean, median, mode, range, etc.)

Skip
(Related to chapter 3)

The number in the middle

Remarks:

1) If the data is skewed → Report the **Median** and IQR

2) If the data is not skewed → Report the Mean and Variance

2 3 **7** 9 100 (**Median = 7**, Mean = 22)