

Presentation of Data

- Histogram
- Box & Whisker Plot
- Cumulative Frequency Graph

You can present data textually, tabularly, and graphically

written in IP
form with texts
and figures

tables

Bar Graph
Pie chart
Line Graph
Pictograph
All above graphs

Histogram

Histo- + -gram
Pole, Mast Graph

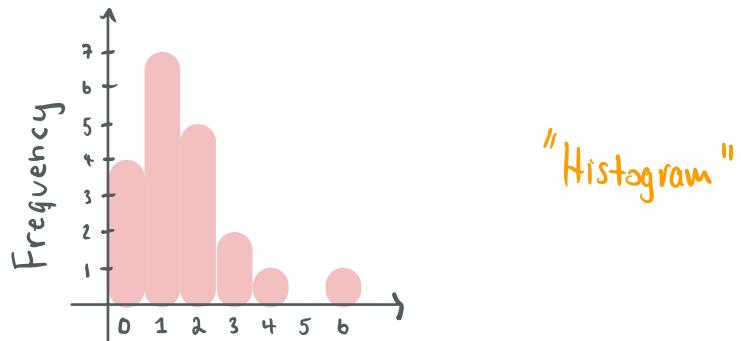
- * Most commonly used to present statistical data
- * Histogram is a bar graph that plots Frequency of values

Ex Consider the data set: 1, 4, 2, 1, 0, 2, 1, 0, 1, 2, 1, 0, 0, 2, 2, 3, 1, 1, 3, 6

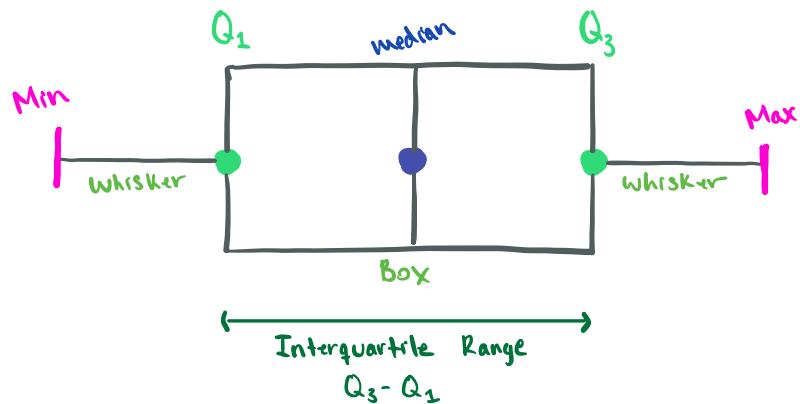
(a) Convert the data into a Frequency Table

Number	Frequency
0	4
1	7
2	5
3	2
4	1
5	0
6	1

b) Plot data like so



Box & Whisker Plot



ex) The following data is the number of points Katy Perry scored over the span of 6 basketball games

5 10 16 17 18 20 32

a) Find the values :

Upper Extreme : 32

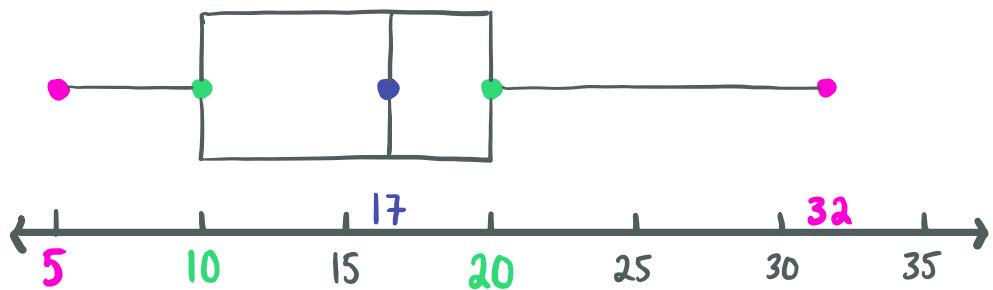
Lower Extreme: 5

Median: 17

Upper Quartile (Q₃): 20

Lower Quartile (Q₁): 10

(b) Construct Box & Whisker Plot



Cumulative Frequency Graph

Cumulative Frequency (CF) - the sum of all the frequencies up to a particular value

Ex] The number of visitors (n) to Goofy Goober Fun Park was recorded on 200 separate days of the year

(a) How many days did GGF have less than 100 visitors?

$$16 + 38 = 54 \text{ days}$$

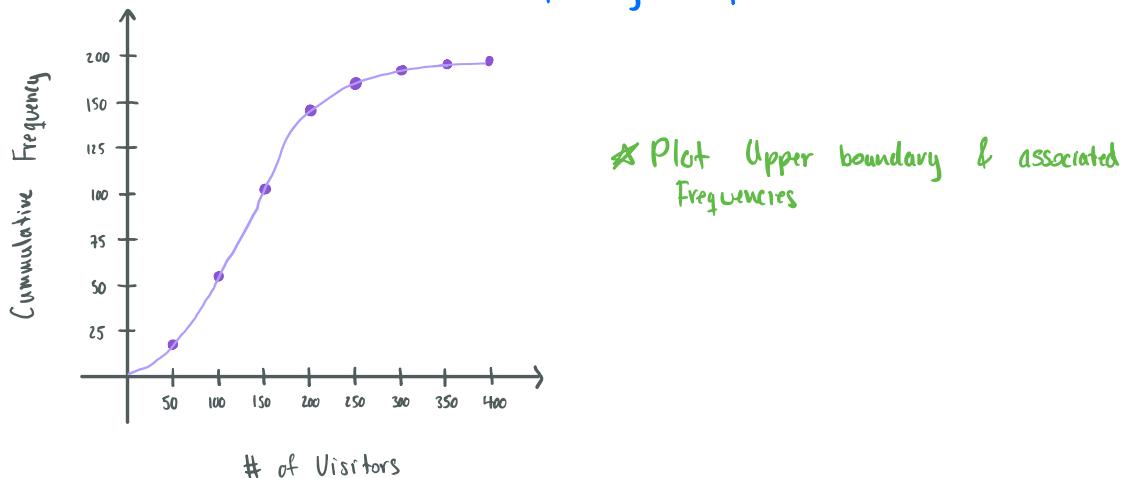
n	Frequency
$0 \leq n < 50$	16
$50 \leq n < 100$	38
$100 \leq n < 150$	50
$150 \leq n < 200$	36
$200 \leq n < 250$	32
$250 \leq n < 300$	19
$300 \leq n < 350$	6
$350 \leq n < 400$	3

$$\sum F = 200$$

(b) Complete this table with the upper boundaries and Cumulative Frequencies

n	Cumulative Frequency
$n < 50$	16
$n < 100$	54
$n < 150$	104
$n < 200$	140
$n < 250$	172
$n < 300$	191
$n < 350$	197
$n < 400$	200

⑤ Draw a Cumulative Frequency Graph for this data



⑥ Use your Cumulative Frequency Graph to estimate the following values:

$$\text{Median} = 146 \text{ visitors}$$

The median occurs at the 100th day $\Rightarrow y=100$
Thus, when $y=100 \Rightarrow x=146$

$$\text{Lower Quartile } (Q_1) = 95 \text{ visitors}$$

Q_1 occurs at the $\frac{1}{4}(200) = 50^{\text{th}}$ day
Thus, when $y=50 \Rightarrow x=95$

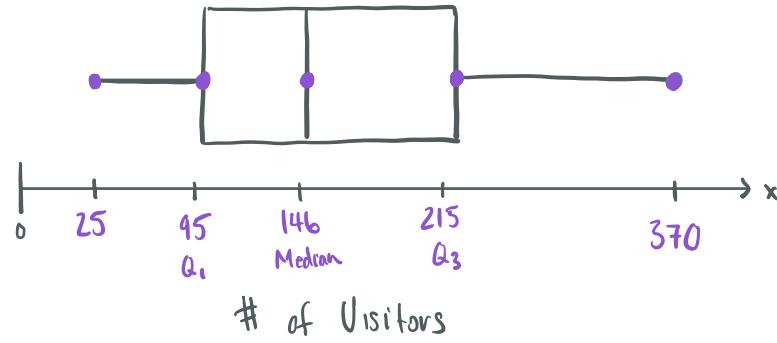
$$\text{Upper Quartile } (Q_3) = 215 \text{ visitors}$$

Q_3 occurs at the $\frac{3}{4}(200) = 150^{\text{th}}$ day
Thus, when $y=150 \Rightarrow x=215$

⑦ Find an estimate for the 85th percentile

$$y-\text{int}: 0.85(200) = 170^{\text{th}} \text{ day} \Rightarrow x = 247 \text{ visitors}$$

⑧ Provided that lowest number of visitors was 25 and the highest number was 370, draw a box-and-whisker plot to represent this data.



⑨ Determine any Outliers

$$IQR = Q_3 - Q_1 = 215 - 95 = 120$$

$$\underline{z} > Q_3 + 1.5(IQR) = 215 + 1.5(120) = \underline{395} \Rightarrow \boxed{\text{None}}$$

$$\underline{z} < Q_1 - 1.5(IQR) = 95 - 1.5(120) = \underline{-85} \Rightarrow \boxed{\text{None}}$$

Data Value	Frequency	Cumulative Frequency
3	1	1 ← 1 number is 3
4	1	2 ← 2 numbers are 4 or less
5	3	5 ← 5 numbers are 5 or less
6	7	12 ← 12 numbers are 6 or less
7	15	27 ← 27 numbers are 7 or less
8	8	35 ← 35 numbers are 8 or less
9	5	40 ← all numbers are 9 or less
Total	40	

Ex

Example 10

Data on the shoe sizes of a group of students is shown in the table.

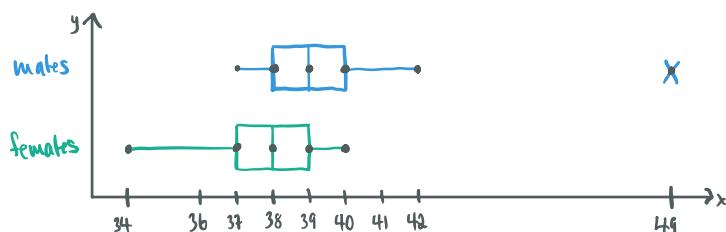
Shoe size	34	36	37	38	39	40	41	42	49
Females	2	2	10	8	7	3	0	0	0
Males	0	0	3	7	12	9	3	1	1

Draw a box-and-whisker plot for the females and for the males and compare the two plots.

State whether the box plots are symmetrical.

Comment on whether there are any outliers.

Draw the box plots again showing any outliers clearly. Outliers are represented by crosses.

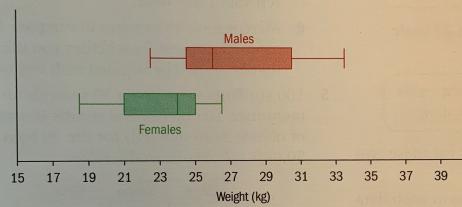


$$\begin{aligned}
 & \text{Males: } Q_1 \rightarrow 8^{\text{th}} = 36 \\
 & \quad Q_3 \rightarrow 8^{\text{th}} = 40 \\
 & \quad \text{Median} \rightarrow 39 \\
 & \quad Q_5 \rightarrow 24^{\text{th}} = 40 \\
 & \quad Z < 35 \\
 & \quad Z > 43 \Rightarrow \text{Outlier: } 49 \\
 & \text{Females: } Q_1 \rightarrow 8^{\text{th}} = 36 \\
 & \quad Q_3 \rightarrow 8^{\text{th}} = 40 \\
 & \quad \text{Median} \rightarrow 38 \\
 & \quad Q_5 \rightarrow 24^{\text{th}} = 39 \\
 & \quad Z < 34 \Rightarrow \text{None} \\
 & \quad Z > 42
 \end{aligned}$$

Ex

Investigation 10

The box-and-whisker plots show the weights, in kilograms, of 24 male poodles (upper) and 24 female poodles (lower).



- 1 Write down the median for both groups.
- 2 Write down the IQR for both groups.
- 3 Write down the percentage of female poodles that weigh less than 24 kg.
- 4 Write down the percentage of male poodles that weigh between 26 and 30.5 kg.
- 5 Compare the two box plots and discuss the differences.
- 6 **Factual** Which measures of central tendency and dispersion can you read from a box-and-whisker plot and from a histogram?

① Males: 26 kg
Females: 24 kg

② $IQR_M: 6 \text{ kg}$
 $IQR_F: 4 \text{ kg}$

③ 50%

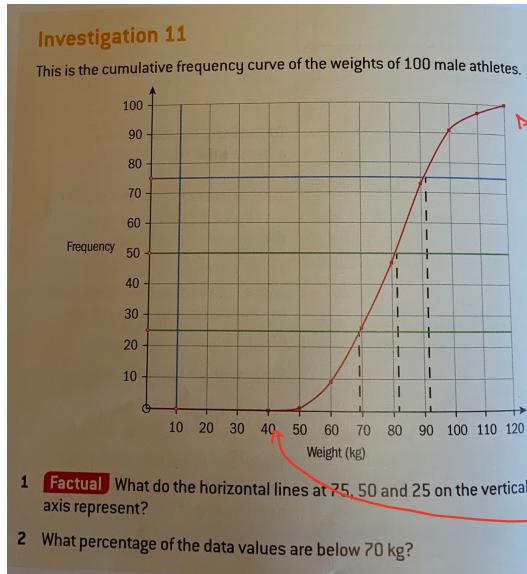
④ 25%

⑤ Males more dispersed

⑥ Central T: Median

Dispersions: IQR, Range

ex



- (1) $y = 75 \rightarrow 25\%$ $x = Q_1$
 $y = 50 \rightarrow 50\%$ $x = \text{Median}$
 $y = 25 \rightarrow 75\%$ $x = Q_3$
- (2) 25% Since $y = 25\%$ of total CF
- (3) $10\% \text{ of } 100 = 10 \text{ kg}$
- (4) $Z < 70 - 1.5(20) = 40 \text{ None}$
 $Z > 90 + 30 = 120 \text{ None}$

3. If 90% of the athletes weigh more than x kg, how could you find the value of x ?

4. Are there outliers?

Practice Problems

Pg 126-127 Q1-9

Pg 131-132 Q1-6