

Module 2c – Five number summary (Min, Q1, Median, Q3, Max) and Box Plots

Median (Min, Q1, Median, Q3, Max) aka five number summary

- [Against all Odds - Unit 6](#)
- We do this when we are describing a single (univariate), *skewed continuous/quantitative* variable's *center* and *spread/variability* numerically
- **Center:** Median (the middle value)
 - Step 1 – Order the array of numbers from least to greatest
 - Step 2 – Find the middle number
 - $\text{Position} = n + 1 / 2$ (where n = total number of values)

Min, Quartile 1, Quartile 3, Max (process)

- **Spread/Variability:** Min, Q1, Q3, Max
- The Min is the smallest value
- Q1 is the median of the bottom half of values
- Q3 is the median of the top half of values
- The Max is the largest value
- IQR (interquartile range)=Q3 minus Q1

Median Examples

Data A

9.26

9.14

9.13

8.77

8.74

8.14

8.1

7.26

6.13

4.74

3.1

$n=11$ (11 values in the array)

$\text{Position}=(11+1)/2=6$ (median position)

Median=8.14

Q1: (exclude 8.14 and find median for bottom half)

$n=5$, $\text{position}=(5+1)/2=3$

Q1=6.13

Q3: Repeat for top half of values

Q3=9.13

Min=3.1

Max=9.26

Median Examples

Data B

12.5

8.84

8.47

7.91

7.71

7.04

6.89

6.58

5.76

5.56

5.25

5.2

$n=12$ (12 values in the array)

$\text{Position}=(12+1)/2=6.5$

$\text{Median}=(7.04+6.89)/2=6.97$

Q1: (include 6.89 and find median for bottom half)

$n=6$, $\text{position}=(6+1)/2=3.5$

$Q1=(5.76+5.56)/2=5.66$

Q3: Repeat for top half of values

$Q3=8.19$

$\text{Min}=5.2$

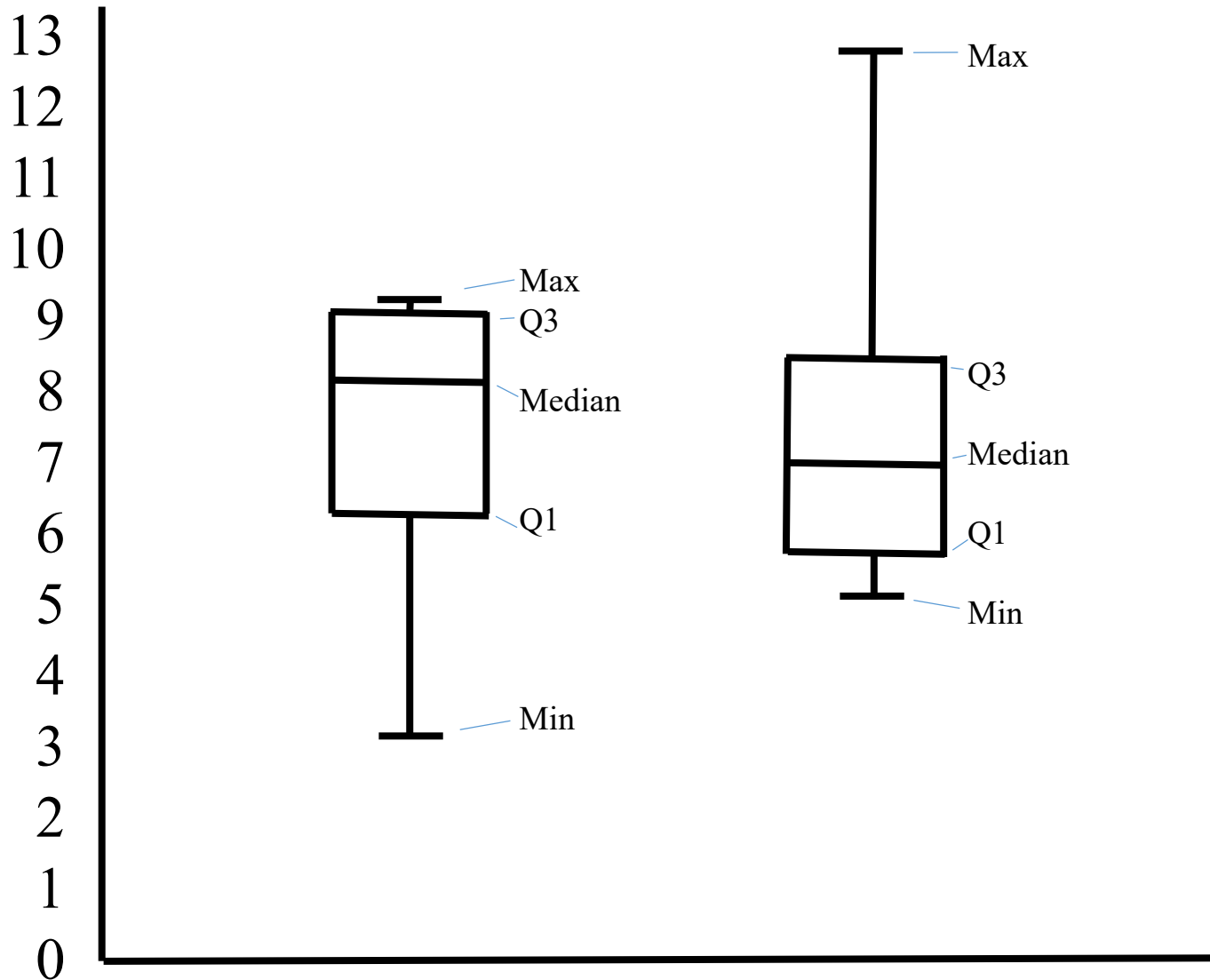
$\text{Max}=12.5$

Making a Boxplot (with five number summary)

Data A		Data B	
	Max:	9.26	Max: 12.5
		9.14	8.84
	Q3:	9.13	Q3: 8.47
		8.77	7.91
		8.74	7.71
	Median:	8.14	Median: 7.04
		8.1	6.89
		7.26	6.58
	Q1:	6.13	Q1: 5.76
		4.74	5.56
	Min:	3.1	Min: 5.25
Average:		7.5	Average: 7.5
Standard Deviation:		2.0	Standard Deviation: 2.0

Making a Boxplot

(with five number summary)



Activity

- Explain how the fact that two arrays of numbers can have the same average and standard deviation but different shapes/distributions reinforces why we use median/standard deviation in some cases and the five number summary in others.