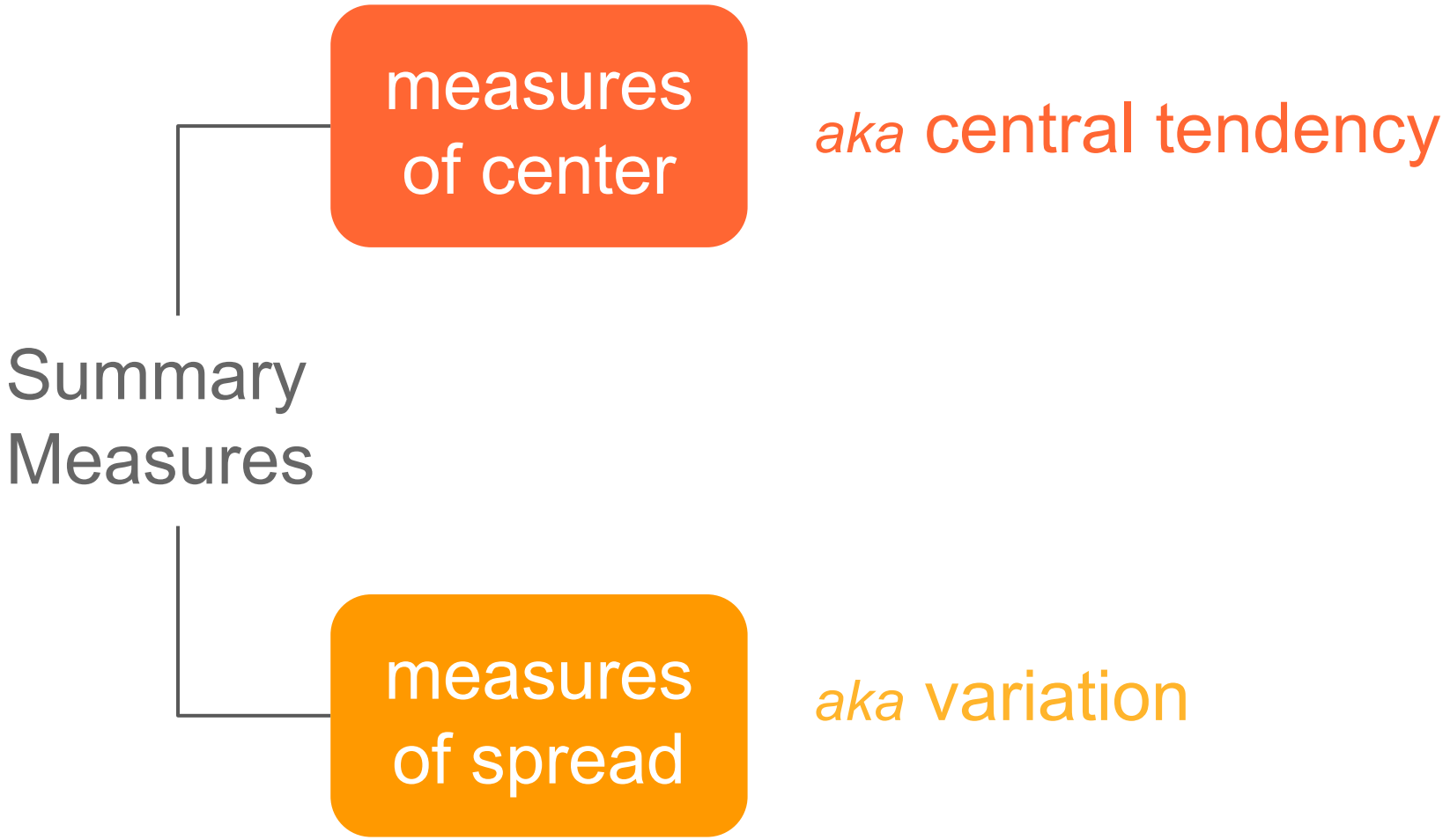
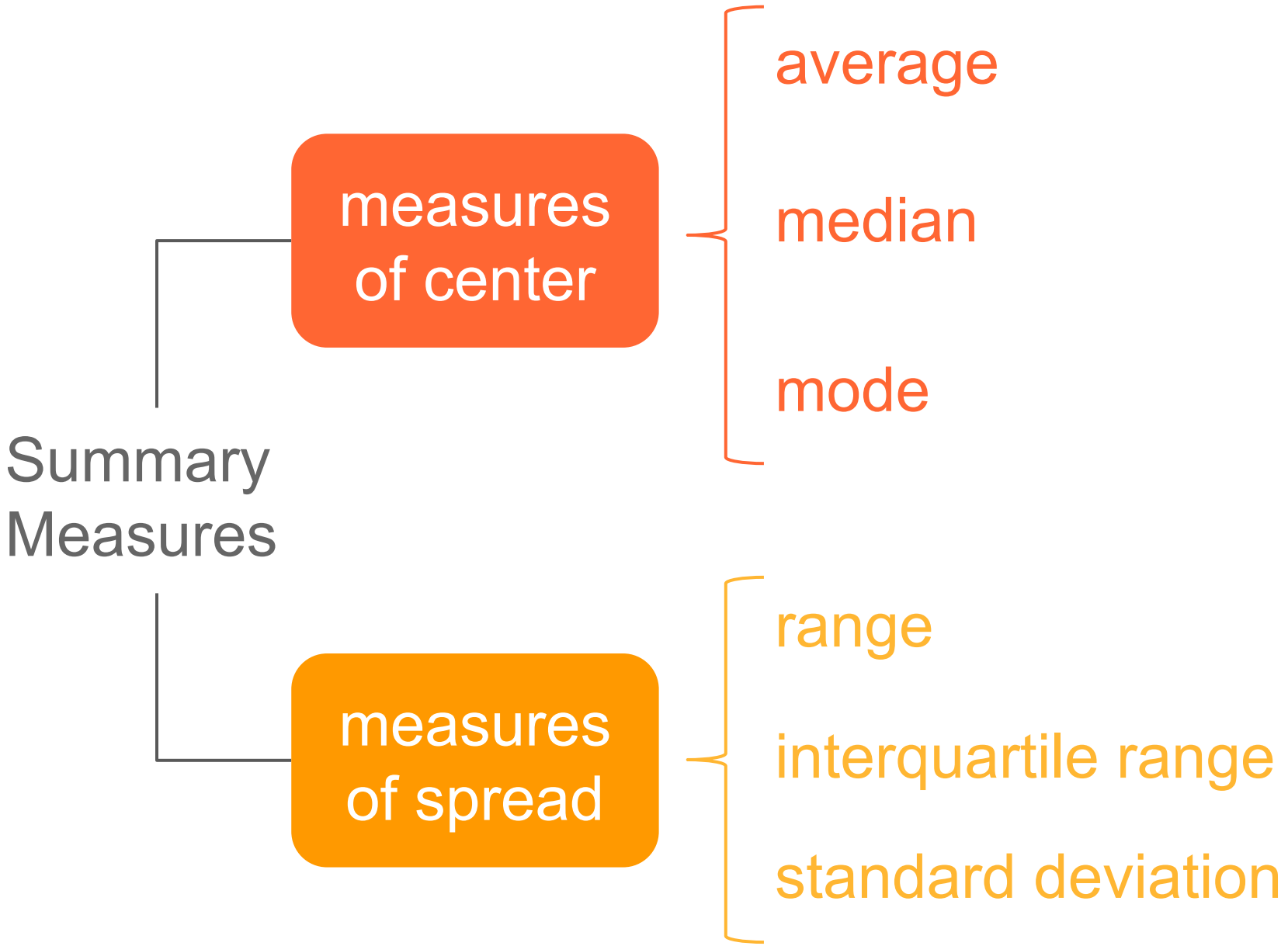
Boxplots

Gaston Sanchez


measures of center
aka central tendency
Summary Measures
measures of spread
aka variation


Summary Measures
measures of center
average
median
mode
measures of spread
range
interquartile range
standard deviation

Preliminary Concepts

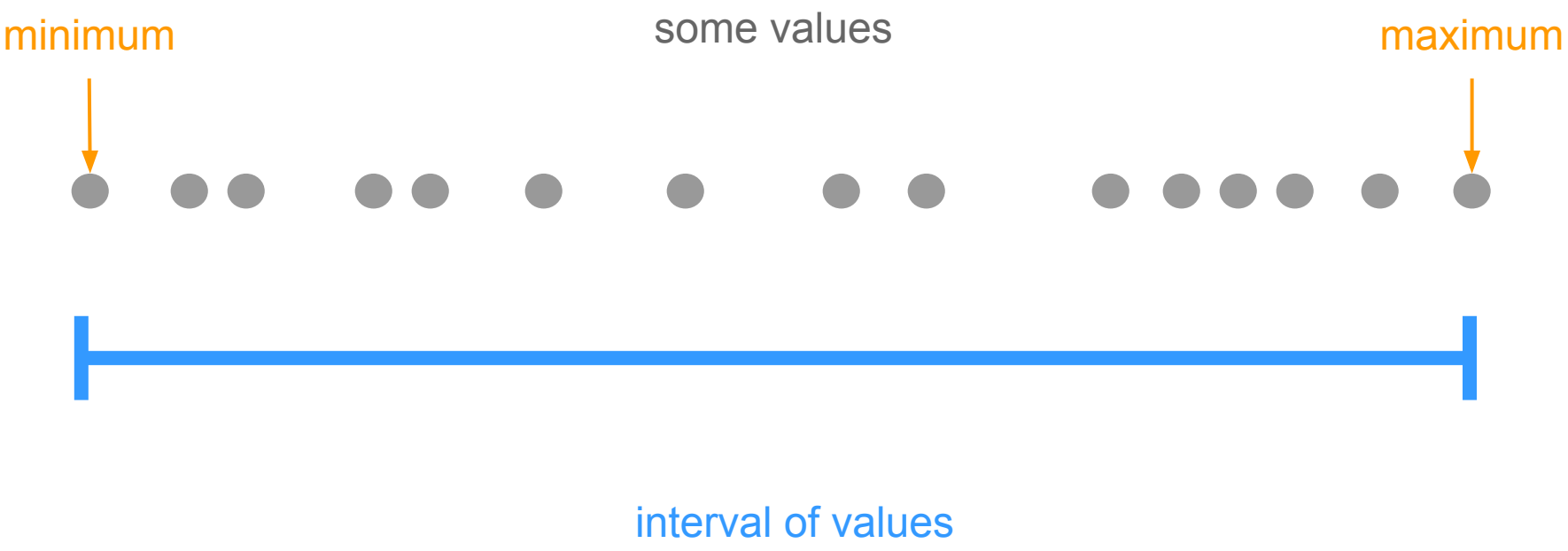
Preliminary Concepts

Minimum

Maximum

Quantiles

Quantiles



Special Quantiles

Number of quantiles	Name
2	Median
3	Terciles
4	Quartiles
5	Quintiles
6	Sextiles
10	Deciles
100	Percentiles

Special Quantiles

Number of quantiles	Name
2	Median
3	Terciles
4	**Quartiles**
5	Quintiles
6	Sextiles
10	Deciles
100	Percentiles

important for measures of spread

Do not confuse

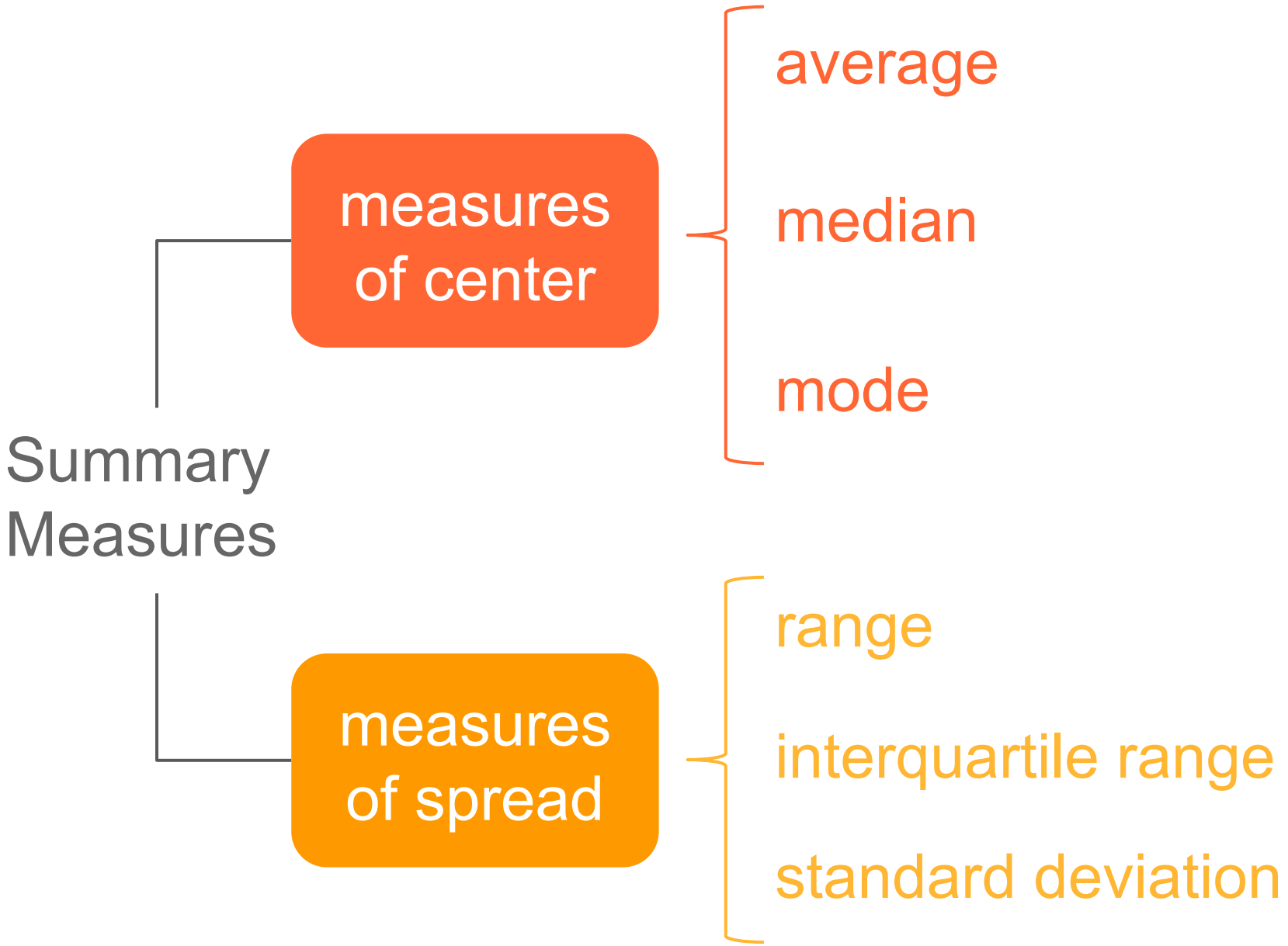
Quantile *"N"* $\neq$ Quartile *"R"*

Range

Looking for a value that reflects the amount of spread

Spread Value

Is there a **"representative"** value that tells us how much variation a variable has?

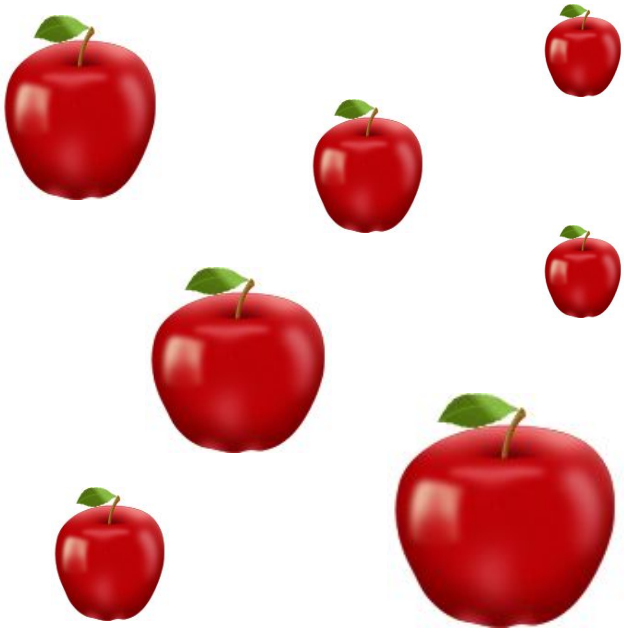

Summary Measures
measures of center
average
median
mode
measures of spread
range
interquartile range
standard deviation

Range

Difference between the largest and smallest values of a data distribution

Range = maximum - minimum

Apple dataset



num	Weight oz	Carbs	Acidity	Shape
1	5	20.0	medium	round
2	6	24.3	high	oval
3	7	25.0	medium	round
4	7	25.5	low	square
5	6	24.7	medium	round
6	8	26.1	low	round
7	6	25.2	high	square
8	9	23.7	high	oval
9	10	21.0	low	round
10	8	27.4	medium	oval

Finding the Range

Apple weight values

5, 6, 7, 7, 6, 8, 6, 9, 10, 8

maximum = 10

minimum = 5

Range = 10 - 5 = 5

About the Range

The range is one measure of variability.

The range is the distance spanned by the data.

It is an overall measure of spread.

It does not take into account any measure of center as a reference value.

Inter-Quartile Range (IQR)

Spread with Percentiles

We can describe the spread of a distribution by giving several percentiles.

The most common is to use the 25th and the 75th percentiles

Spread with Percentiles

1st quartile Q_1 = *25th percentile*

2nd quartile Q_2 = *50th percentile*

3rd quartile Q_3 = *75th percentile*

4th quartile Q_4 = *100th percentile*

Spread with Percentiles

1st quartile $\mathbf{Q_1}$ = *25th percentile*

2nd quartile Q_2 = *50th percentile*

3rd quartile $\mathbf{Q_3}$ = *75th percentile*

4th quartile Q_4 = *100th percentile*

Finding the SD

Apple weight values

5, 6, 7, 7, 6, 8, 6, 9, 10, 8

Ordered values

5, 6, 6, 6, 7, | 7, 8, 8, 9, 10

median

Q_1 Q_3

IQR = 8 - 6 = 2

Boxplots

Box-plots

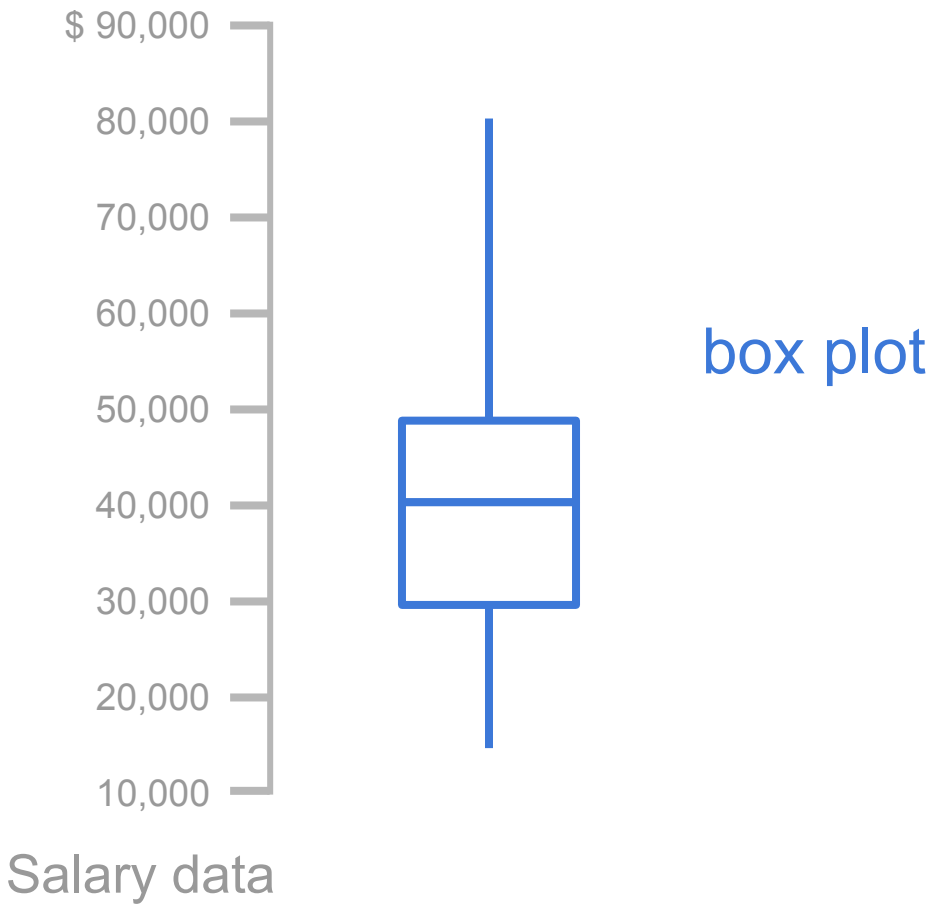
Box-and-whisker plots, most commonly known as box plots, are a simple yet very effective way to display the distribution of values by using five summary indicators.

Box-plots based on 5-number summary

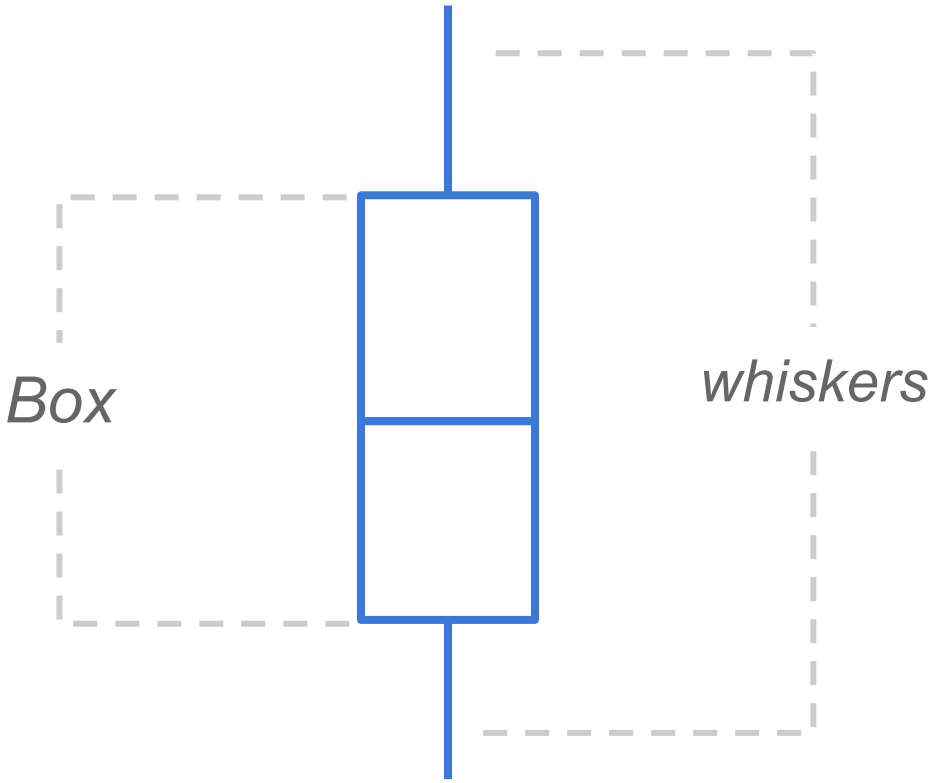
5 summary indicators:

1. minimum
2. 25th percentile (Q1)
3. 50th percentile (Q2) or median
4. 75th percentile (Q3)
5. maximum

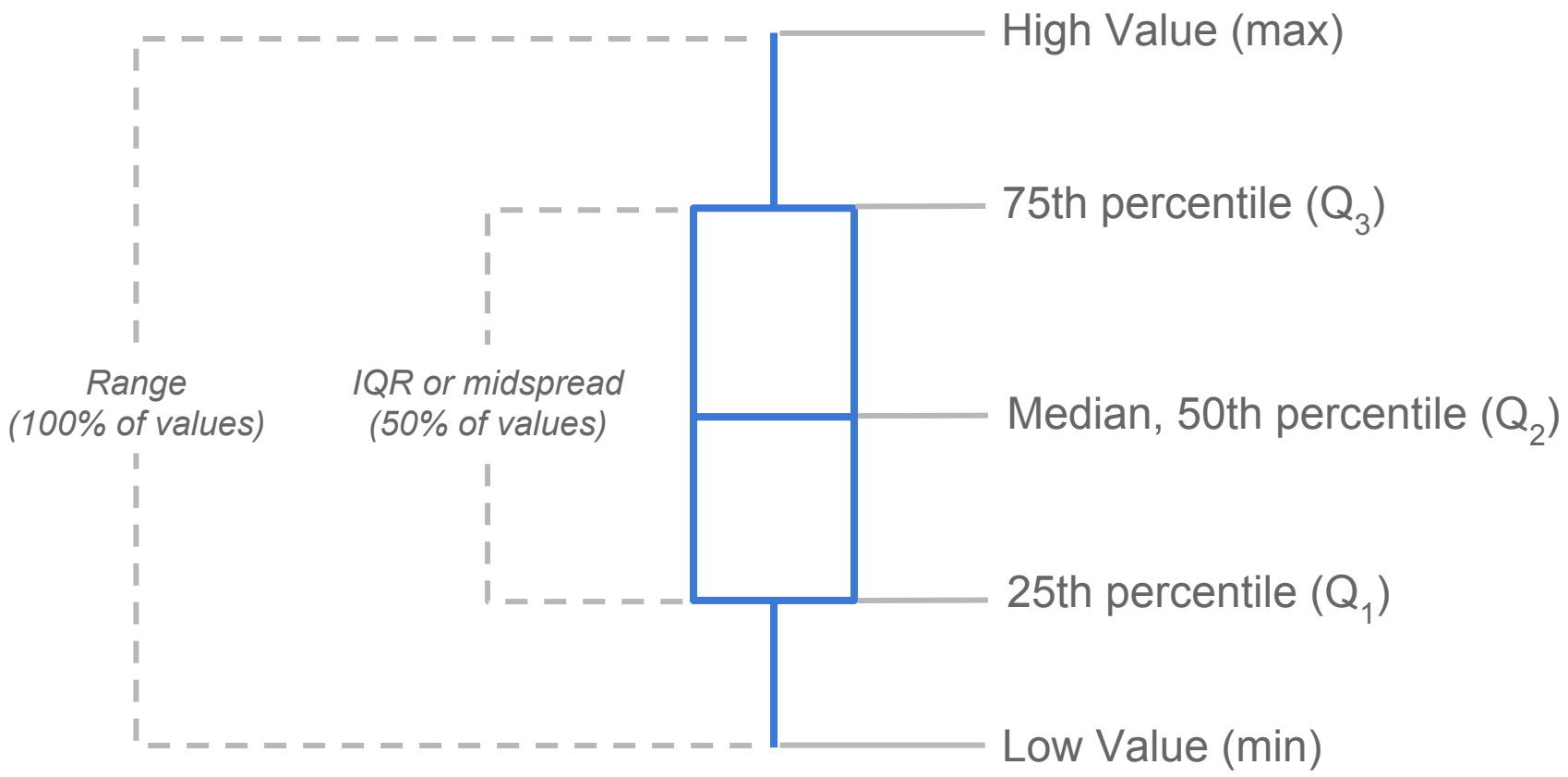
Very simple boxplot example



Anatomy of a boxplot



Anatomy of a boxplot



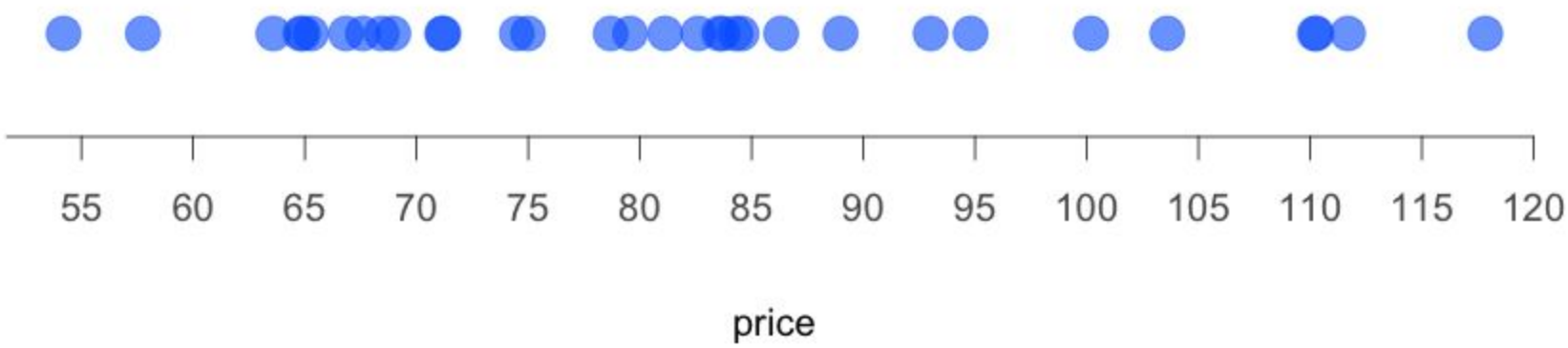

NFC PLAYOFF
DATE AND TIME TBA.
CANDLESTICK PARK
LEVEL
LOWER RESERVE
SEC.
LR4
ROW
10
SEAT
14
WC TICKET: $89.00
DIV TICKET: $104.00
PRICE INCLUDES ALL TAXES. NO REFUND OR EXCHANGE.
VISA
NFL
PLAYOFFS
LOWER RESERVE
SEC.
LR4
ROW
10
SEAT
14

NFL Average Ticket Prices

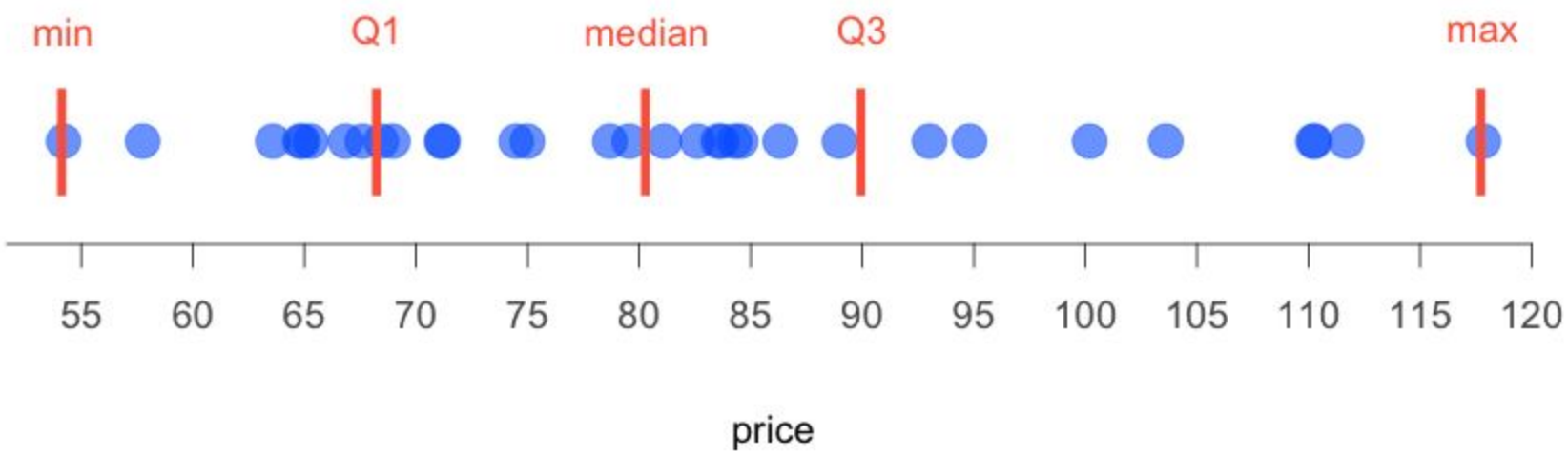
Source: Team Marketing 2013

Num	Team	Price	Num	Team	Price
1	browns	54.20	17	steelers	81.13
2	bills	57.75	18	packers	82.61
3	buccaneers	63.59	19	49ers	83.54
4	raiders	64.80	20	falcons	83.71
5	chiefs	64.92	21	broncos	84.27
6	titans	65.28	22	chargers	84.55
7	panthers	66.84	23	colts	86.32
8	lions	67.60	24	texans	88.98
9	jaguars	68.44	25	eagles	93.01
10	bengals	68.96	26	redskins	94.80
11	dolphins	71.14	27	ravens	100.19
12	seahawks	71.21	28	bears	103.60
13	rams	74.49	29	cowboys	110.20
14	saints	74.99	30	jets	110.28
15	vikings	78.69	31	giants	111.69
16	cardinals	79.56	32	patriots	117.84

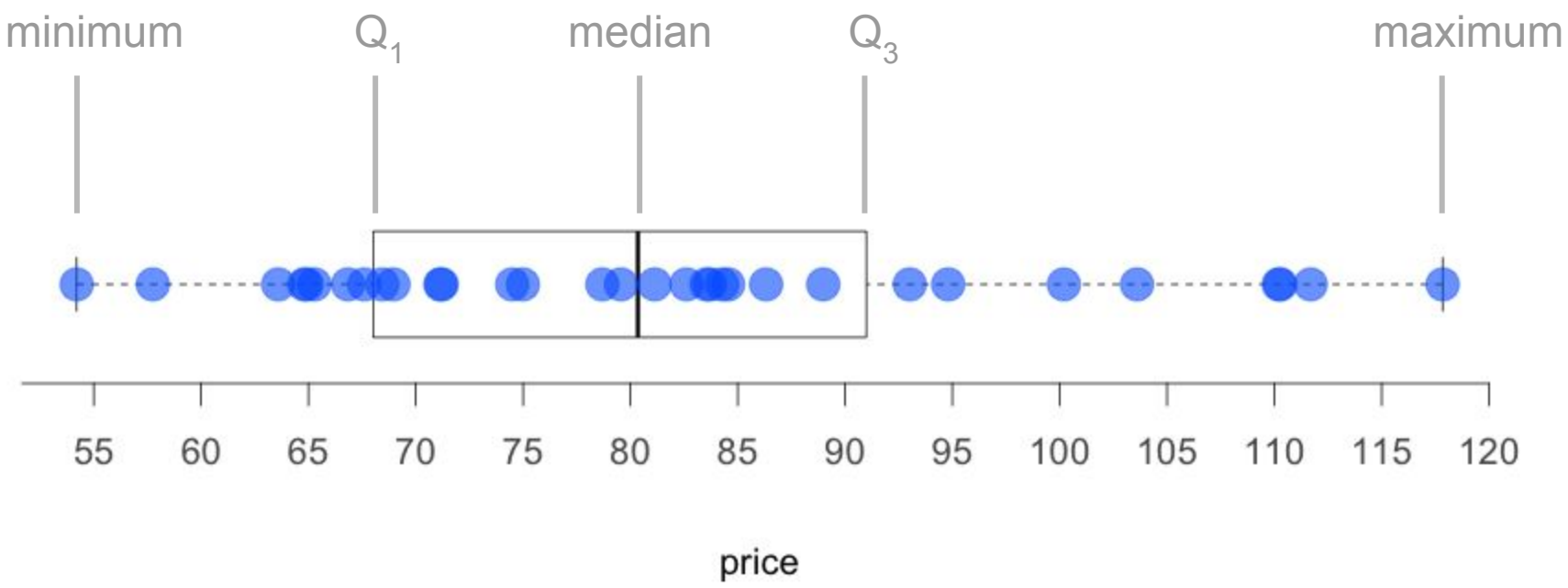
NFL Ticket prices



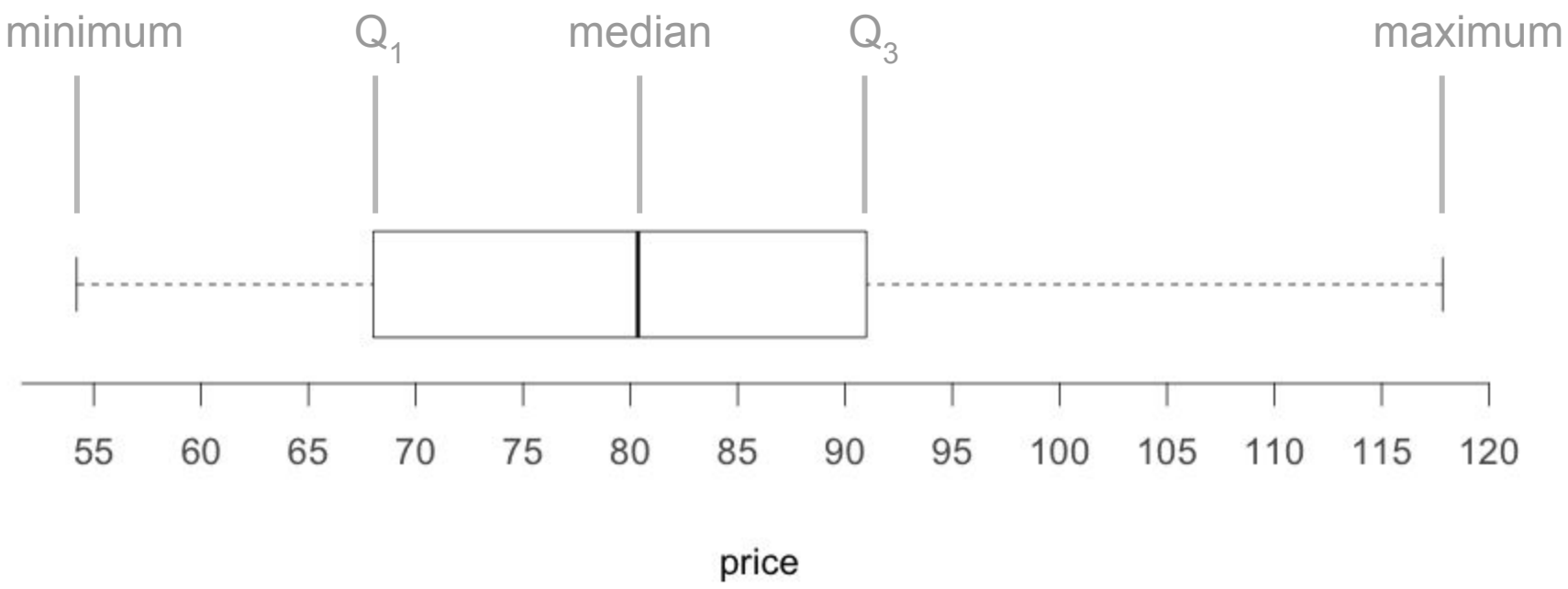
NFL Ticket prices



NFL Ticket prices



NFL Ticket prices



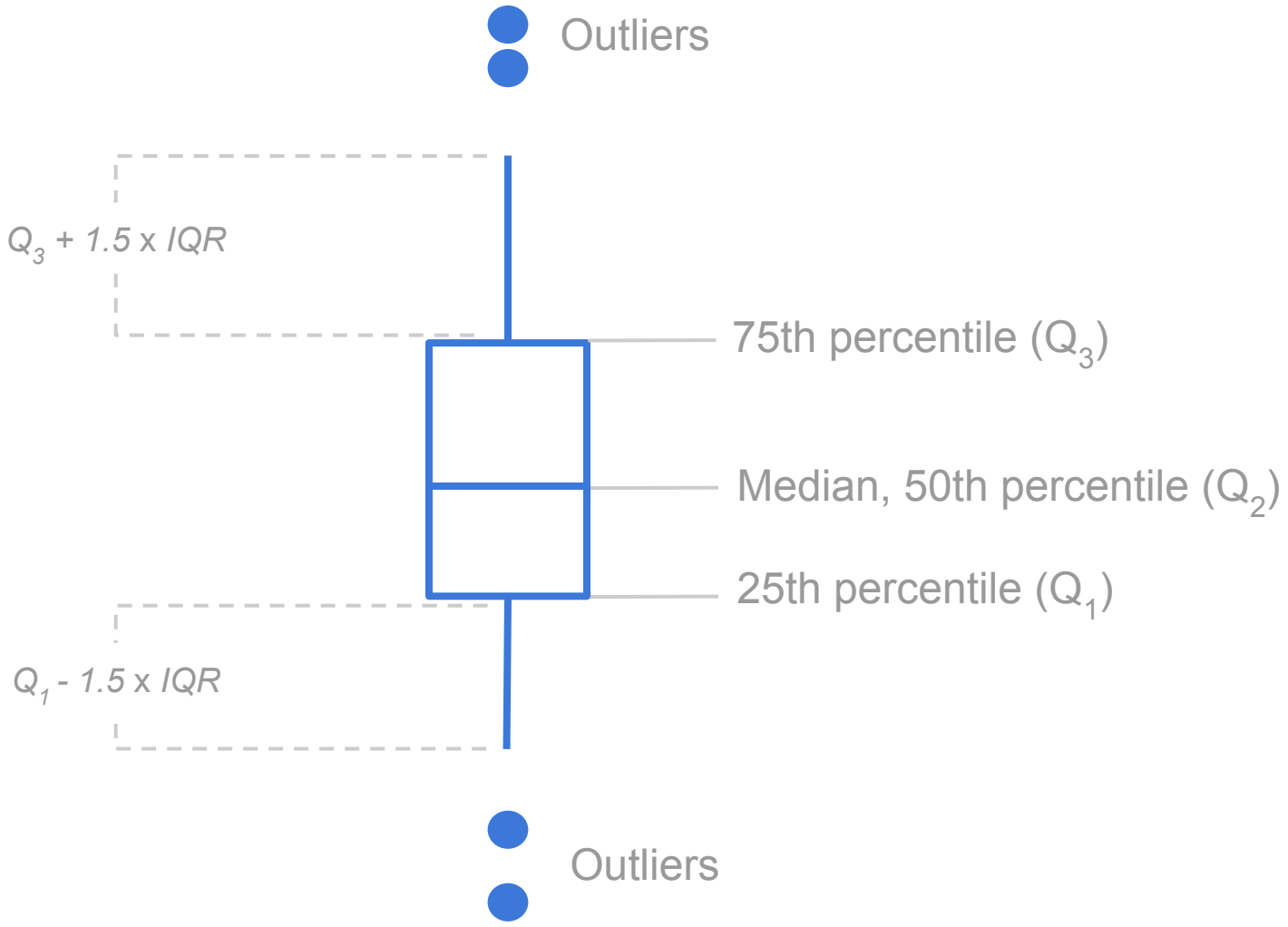
Normally, no dots are plotted

Boxplots and outliers

The 1.5 x IQR rule for outliers

Call an observation a suspected outlier if it falls more than **1.5 x IQR** above the third quartile or below the first quartile

Anatomy of a boxplot



Your Turn

5-number summary

Summary	*Value*
minimum	110
Q_1 (25th percentile)	182
Q_2 (50th percentile)	225
Q_3 (75th percentile)	320
maximum	440

Graph a boxplot