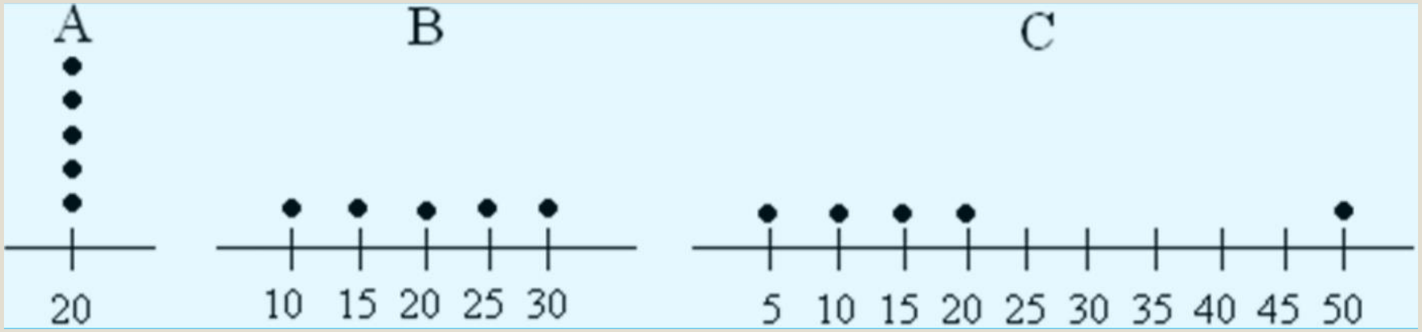
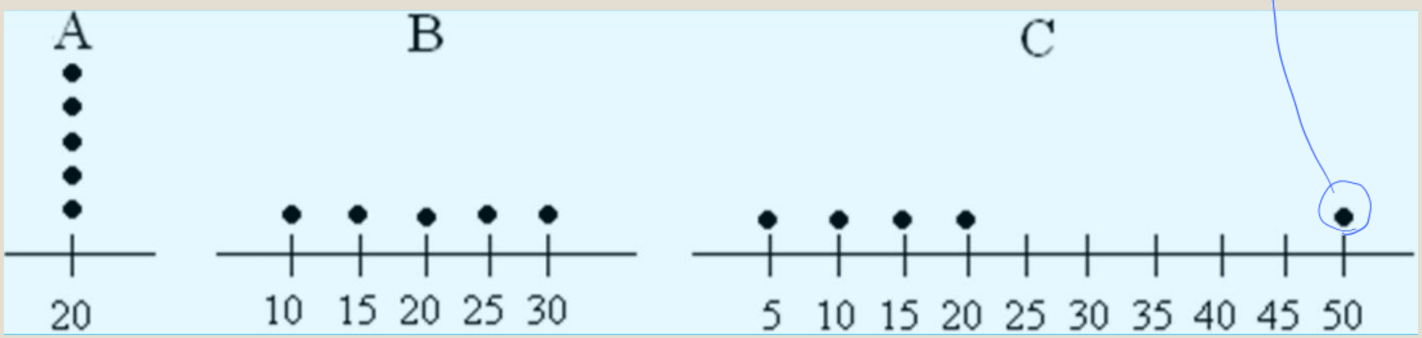
What is the same?



- All have 5 data points (n = 5)
- All have a mean/average of 20
- All have at least one point equal to 20

What is different?

large outlier!



range = 0 range = 20 range = 45

- shape, range (variability/precision)

- What is one measure of center that we could use to compare to all data points to describe how variable the data points are?

- Mean

x_i = i^{th} index data point from data set X (ex X_3 = 3^{rd} index data point)

$\bar{x}$ = mean of values of all datapoints in X

Calculate the deviations from the mean

	A x_i	Deviation from mean $x_i - \bar{x}$	B y_i	Deviation from mean $y_i - \bar{y}$	C z_i	Deviation from mean $z_i - \bar{z}$
x_1	20	0	10	-10	5	-15
x_2	20	0	15	-5	10	-10
x_3	20	0	20	0	15	-5
x_4	20	0	25	5	20	0
(x_n) x_5	20	0	30	10	50	30

`

- How can we convert this to one number to describe the variability?

  - Average? $\Longrightarrow$ problem is all equal 0

  - Get rid of the negatives
    - ↳ square the deviations from the mean

Calculate the squared deviations

A x_i	$x_i - \bar{x}$	Squared Deviation $(x_i - \bar{x})^2$	B y_i	$y_i - \bar{y}$	Squared Deviation $(y_i - \bar{y})^2$	C z_i	$z_i - \bar{z}$	Squared Deviation $(z_i - \bar{z})^2$
20	0	0	10	-10	100	5	-15	225
20	0	0	15	-5	25	10	-10	100
20	0	0	20	0	0	15	-5	25
20	0	0	25	5	25	20	0	0
20	0	0	30	10	100	50	30	900

outlier effect

- Now, how can we convert this to one number to describe the variability?

Average the squared deviations

Sum of the squared deviations

A x_i	$x_i - \bar{x}$	Squared Deviation $(x_i - \bar{x})^2$	B y_i	$y_i - \bar{y}$	Squared Deviation $(y_i - \bar{y})^2$	C z_i	$z_i - \bar{z}$	Squared Deviation $(z_i - \bar{z})^2$
20	0	0	10	-10	100	5	-15	225
20	0	0	15	-5	25	10	-10	100
20	0	0	20	0	0	15	-5	25
20	0	0	25	5	25	20	0	0
20	0	0	30	10	100	50	30	900
	Sum =	0		Sum =	250		Sum =	1250

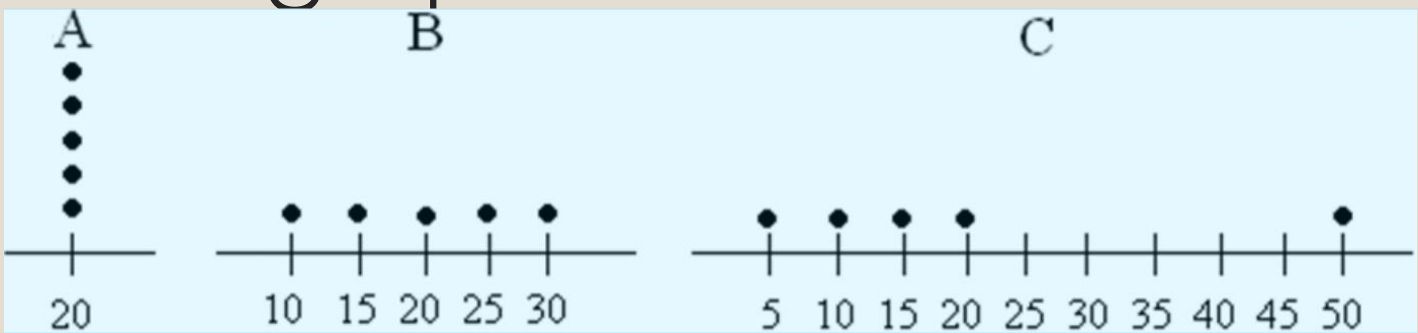
"Average" $\frac{0}{n-1} = \frac{0}{4} = 0$ $\qquad \frac{250}{n-1} = \frac{250}{4} = 62.5$ $\qquad \frac{1250}{n-1} = \frac{1250}{4} = 312.5$

$\boxed{\text{Divide by } n-1\text{, not } n}$

s^2 = variance

s = standard deviation

Compare s^2 (variance) to the graphs



Dataset	s^2
A	0
B	62.5
C	312.5

As variance (s^2) increases, the spread of the data looks larger

In the process we squared the deviations

- How do you "un-square" a value?

square-root!

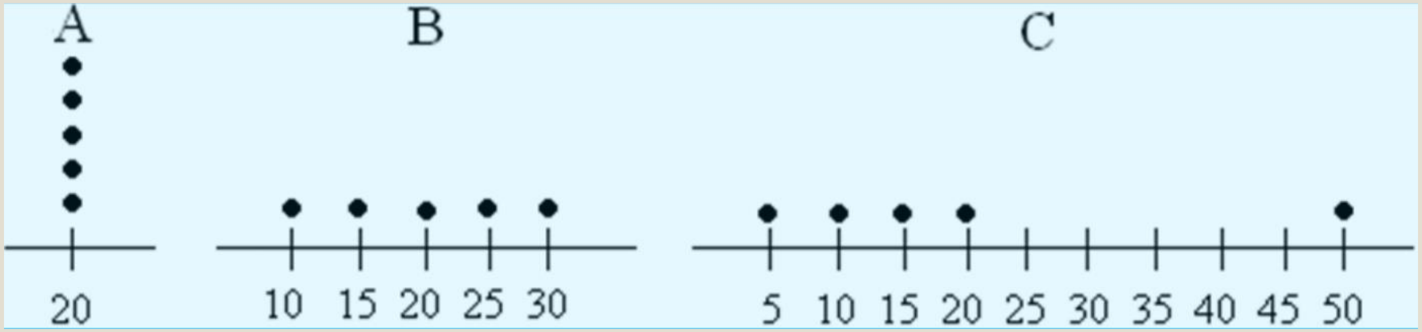
Standard Deviation

Dataset	s^2	s
A	0	0
B	62.5	7.91
C	312.5	17.68

- What do the standard deviations represent?

How far, on average, each value is from the mean

Compare s (standard deviation) to the graphs



Dataset	s^2	s
A	0	0
B	62.5	7.91
C	312.5	17.68

As standard deviation (s) increases, the spread looks larger in the plot

Standard Deviation

$\sum$ = summation from index $i=1$ to index $i=n$

$$s=\sqrt{\frac{(y_1-\bar{y})^2+(y_2-\bar{y})^2+\ldots+(y_n-\bar{y})^2}{n-1}}=\sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}}$$

Can s be negative? No.
Does s have units? Yes.
Can s = 0? Yes