Mediating a work conflict (source: radicalmath.org)

We have this information concerning wages at a fictional company:

Number of people in each position	Position	Yearly individual salary	Total salary per position
1	President	$200,000	$200,000
3	Vice Presidents	$100,000	$300,000
5	Managers	$50,000	$250,000
10	Supervisors	$30,000	$300,000
11	Workers	$28,000	$308,000
20	Workers	$20,000	$400,000
22	Workers	$18,000	$396,000
6	Workers	$16,000	$96,000

The union leader, who represents the 59 workers of the company, claims the average yearly salary is $18,000 and suggests all workers get a raise of $7,000 a year. How did the union leader obtain such an "average"?

The company owners claim the average yearly salary in the company is $28,846. They propose each worker receive a raise of $1,000 a year. How did the company owners obtain this "average"?

Mediating a work conflict (source: radicalmath.org)

We have this information concerning wages at a fictional company:

Number of people in each position	Position	Yearly individual salary	Total salary per position
1	President	$200,000	$200,000
3	Vice Presidents	$100,000	$300,000
5	Managers	$50,000	$250,000
10	Supervisors	$30,000	$300,000
11	Workers	$28,000	$308,000
20	Workers	$20,000	$400,000
22	Workers	$18,000	$396,000
6	Workers	$16,000	$96,000

The union leader, who represents the 59 workers of the company, claims the average yearly salary is $18,000 and suggests all workers get a raise of $7,000 a year. How did the union leader obtain such an "average"? $18,000 is actually the mode: the most frequent salary.

The company owners claim the average yearly salary in the company is $28,846. They propose each worker receive a raise of $1,000 a year. How did the company owners obtain this "average"? This is the mean of *everyone's* salary.

On "Averages"

Given a set of n data points $x_1, x_2, \ldots, x_n$, we have:

- The mean $\bar{x}$ ("x bar") is the sum of all the data points, divided by the number of points:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

- The mode is the most frequent value appearing in the data points.
- The median is the number "in the middle" once the list has been ordered from smallest to largest.

We also have a measure of the spread of the data, given by the standard deviation σ ("sigma"), which is related to how far are the data points from the mean:

$$\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}$$

Can we describe all our data using just a few numbers?

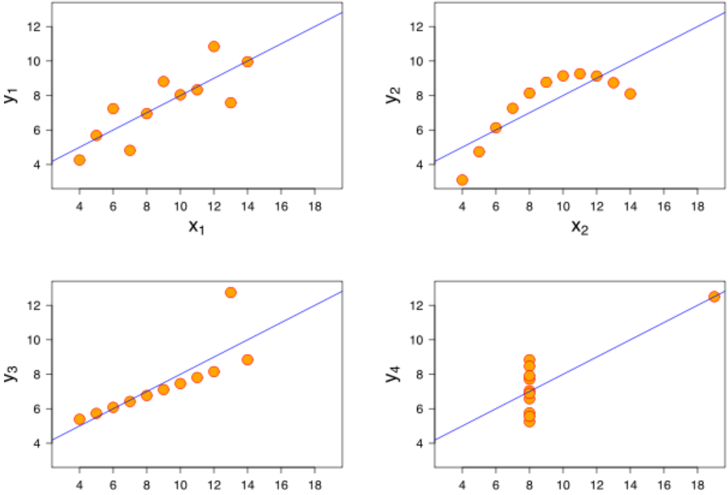
Here a few questions to ponder for today:

- Can you find two sets of three numbers that have the same mean but look very different?
- Can you find two sets of three numbers that have the same mean *and* the same standard deviation but look very different?
- Can you find two sets of 10 numbers that have the same mean but look very different?
- Can you find two sets of 10 numbers that have the same mean *and* the same standard deviation but look very different?

Challenge: can you find a set of 10 different (x, y) points such that the mean of the x's is 9, the standard deviation of the x's is 11, while the mean of the y's is 7.5?

Super Challenge! Can you find two different answers to the above challenge, that look very different?!

Anscombe's Quartet (1973. Source: Wikipedia.)



This is Anscombe's Quartet, demonstrating the importance of graphing data and the effects of outliers on statistical properties.