GENERAL ASSEMBLY

DATA SCIENCE
CLASS 3: DATA VISUALIZATION

AGENDA

I. THE IMPORTANCE OF VISUALIZATION
II. VISUALIZATION AS A MEDIUM

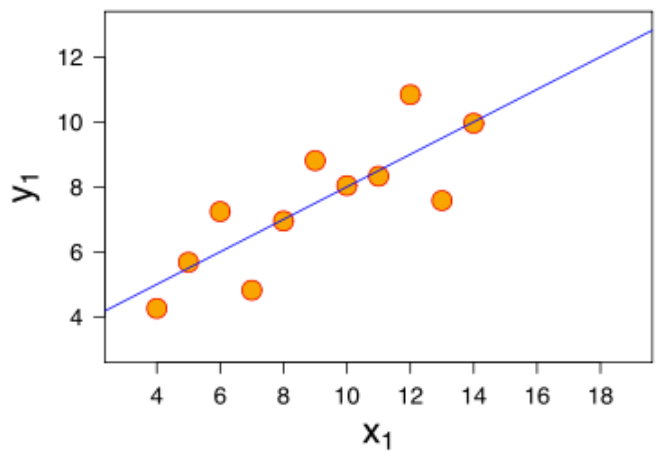
LAB:
III. VISUALIZATION IN R

I. THE IMPORTANCE OF VISUALIZATION

VISUALIZATION VS. SUMMARY STATISTICS

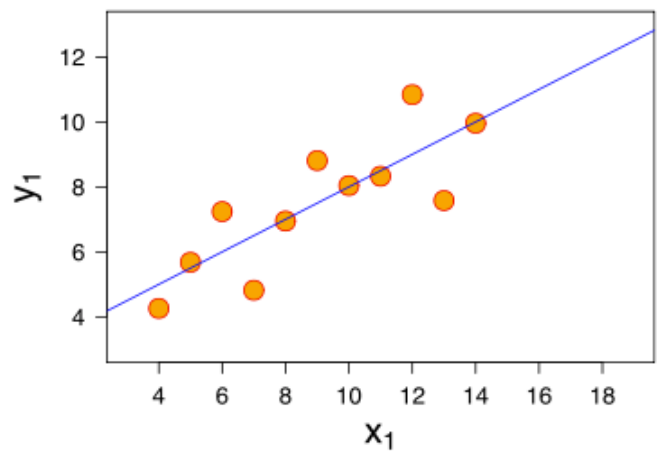
Consider the following dataset:

- eleven (x, y) points



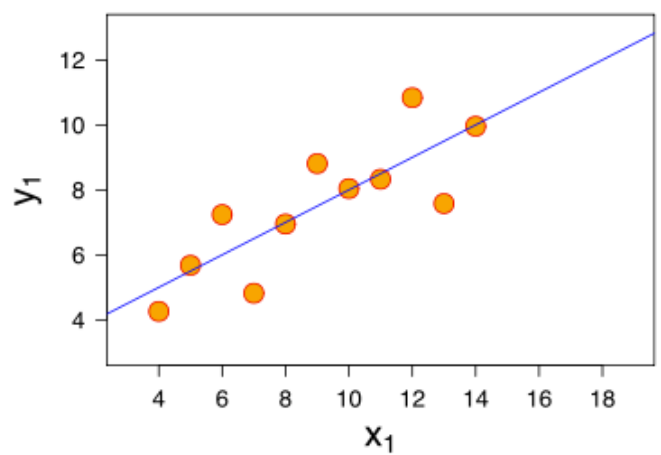
Consider the following dataset:

- eleven (x, y) points
- mean of x = 9, mean of y = 7.5



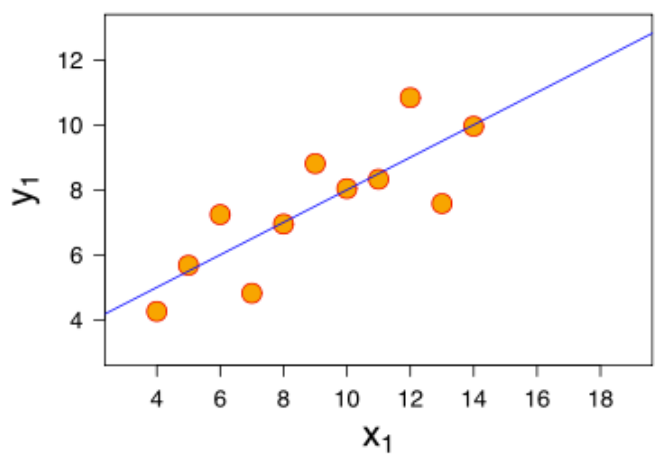
Consider the following dataset:

- eleven (x, y) points
- mean of x = 9, mean of y = 7.5
- variance of x = 11, variance of y = 4.1



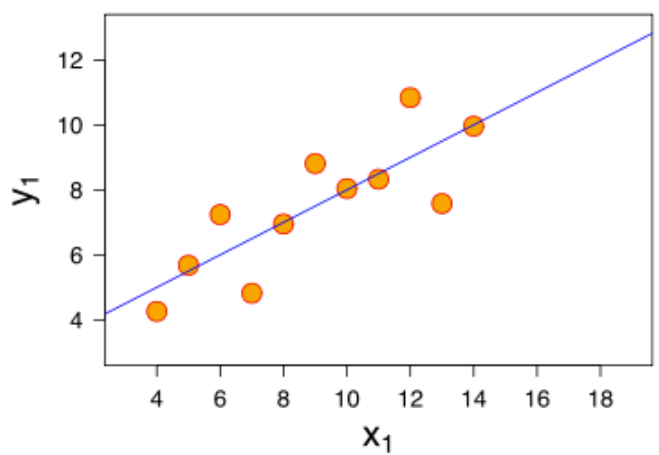
Consider the following dataset:

- eleven (x, y) points
- mean of x = 9, mean of y = 7.5
- variance of x = 11, variance of y = 4.1
- correlation of x and y = 0.8



Consider the following dataset:

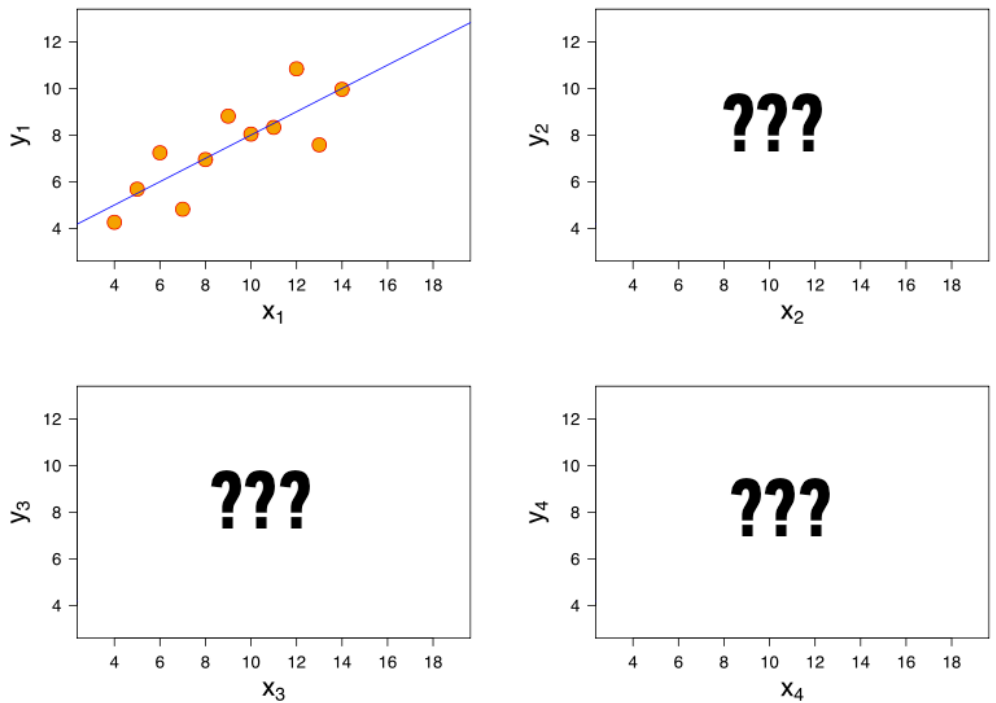
- eleven (x, y) points
- mean of x = 9, mean of y = 7.5
- variance of x = 11, variance of y = 4.1
- correlation of x, y = 0.8
- line of best fit: y = 3.00 + 0.500x



EXERCISE – WHY VISUALIZE DATA?

Now, suppose I give you three more datasets with exactly the same characteristics...

Q: how similar are these datasets?



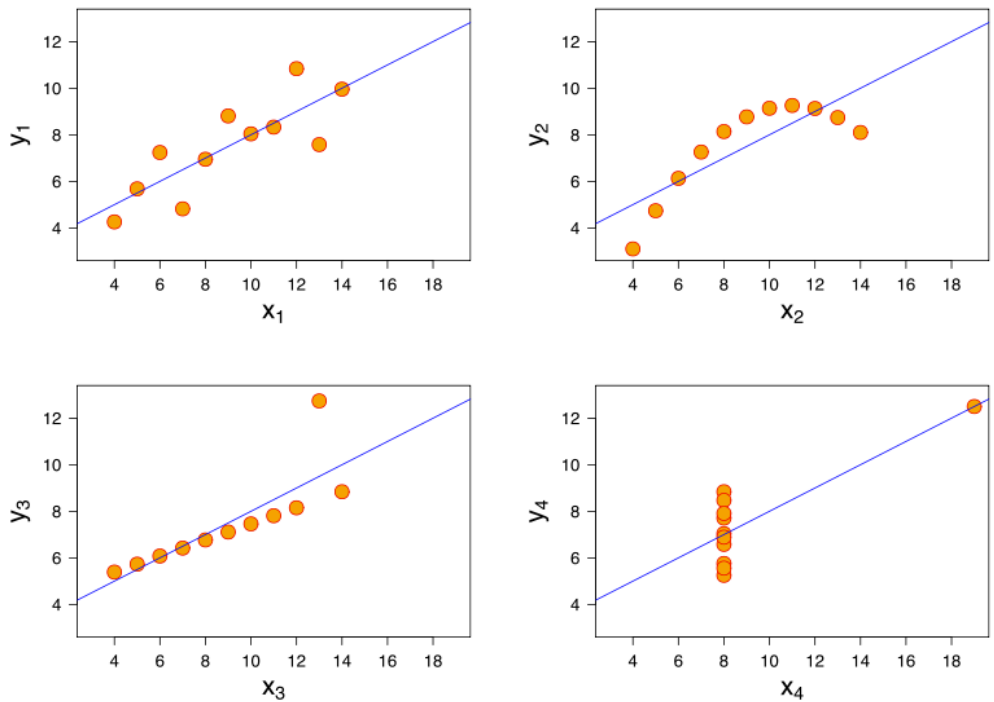
EXERCISE – WHY VISUALIZE DATA?

Now, suppose I give you three more datasets with exactly the same characteristics.

Q: how similar are these datasets?

A: not very!

http://en.wikipedia.org/wiki/Anscombe's_quartet



II. VISUALIZATION AS A MEDIUM

III. VISUALIZATION IN R