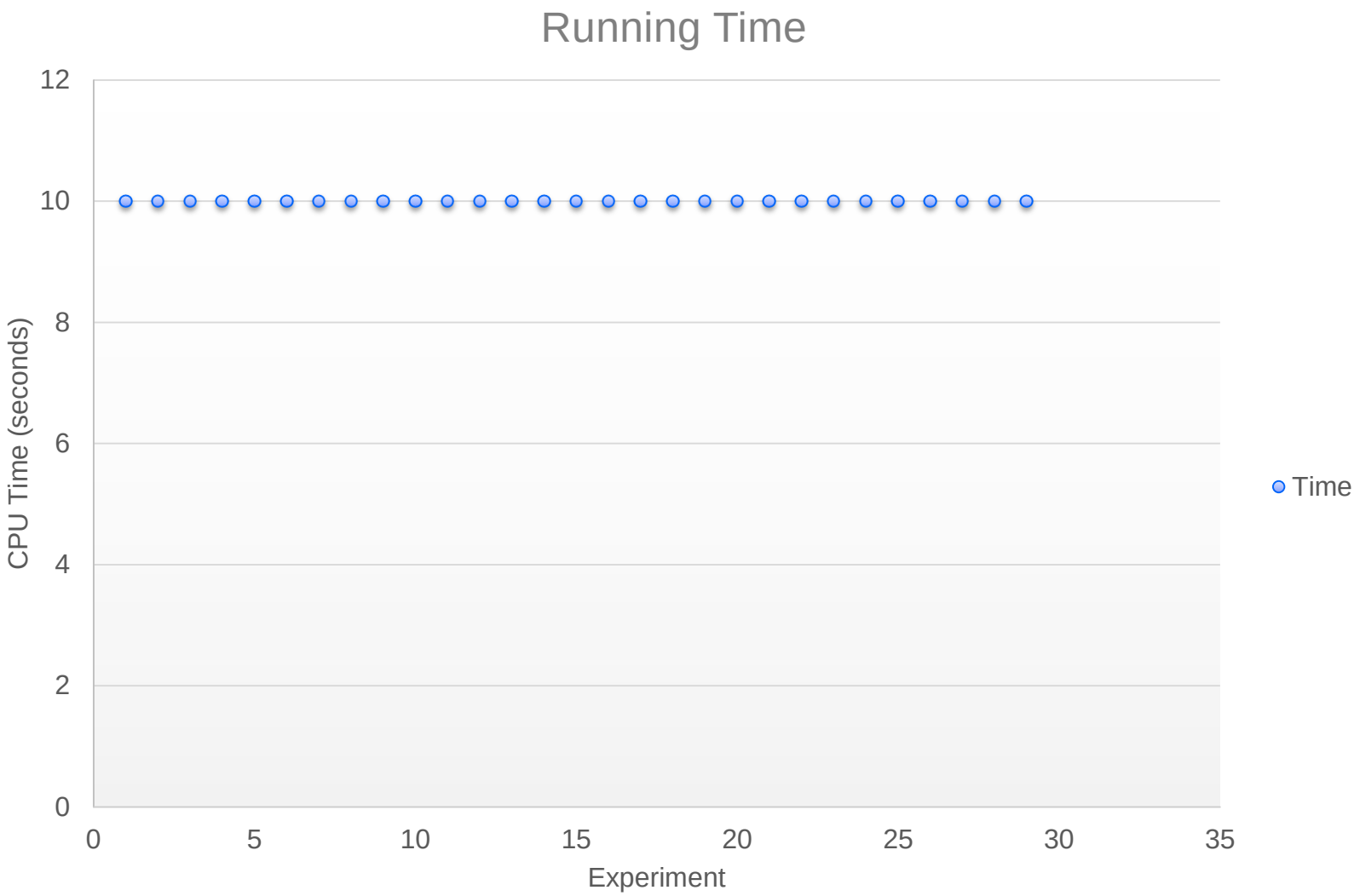
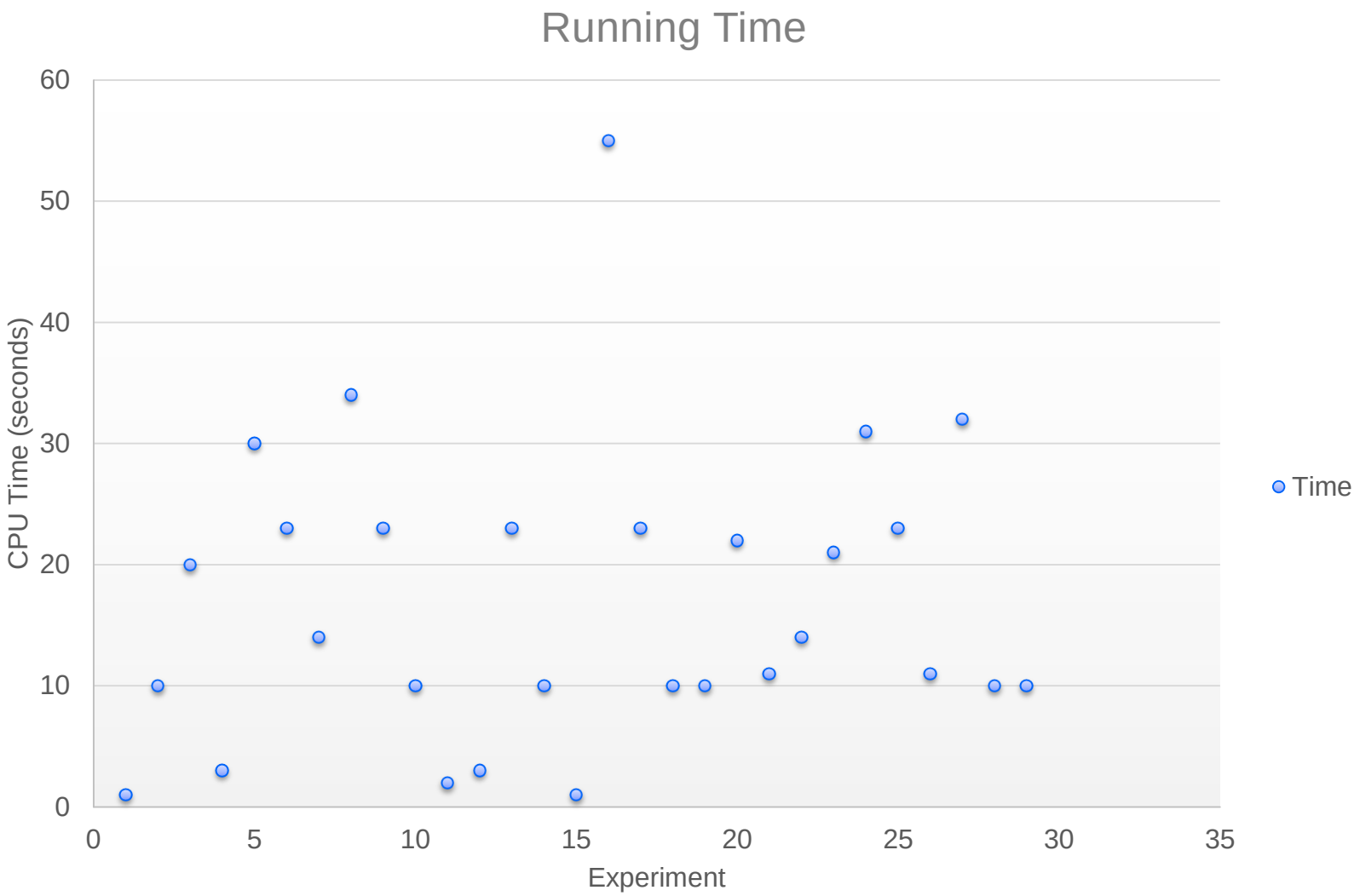
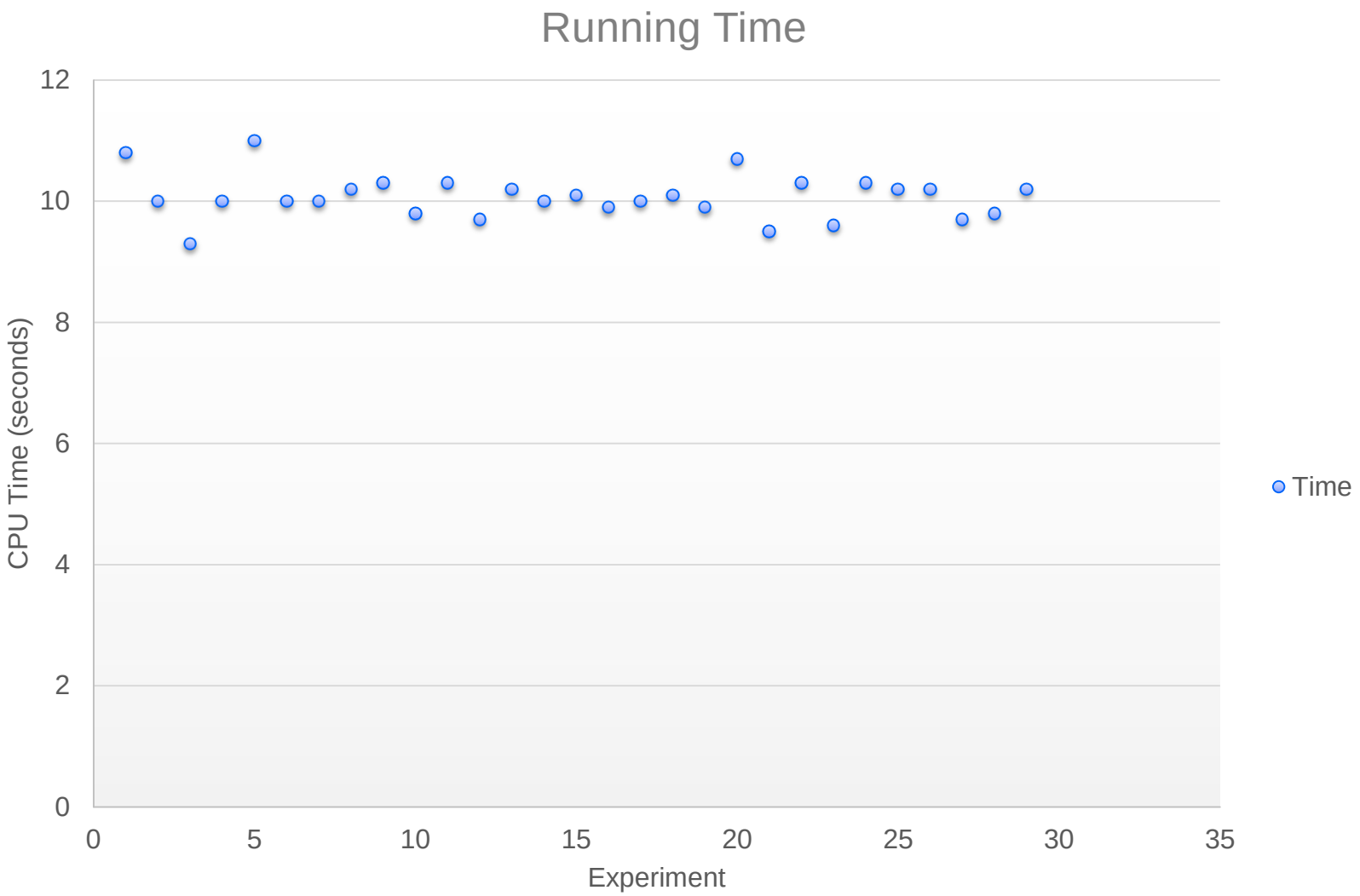
A Little Bit of Statistics

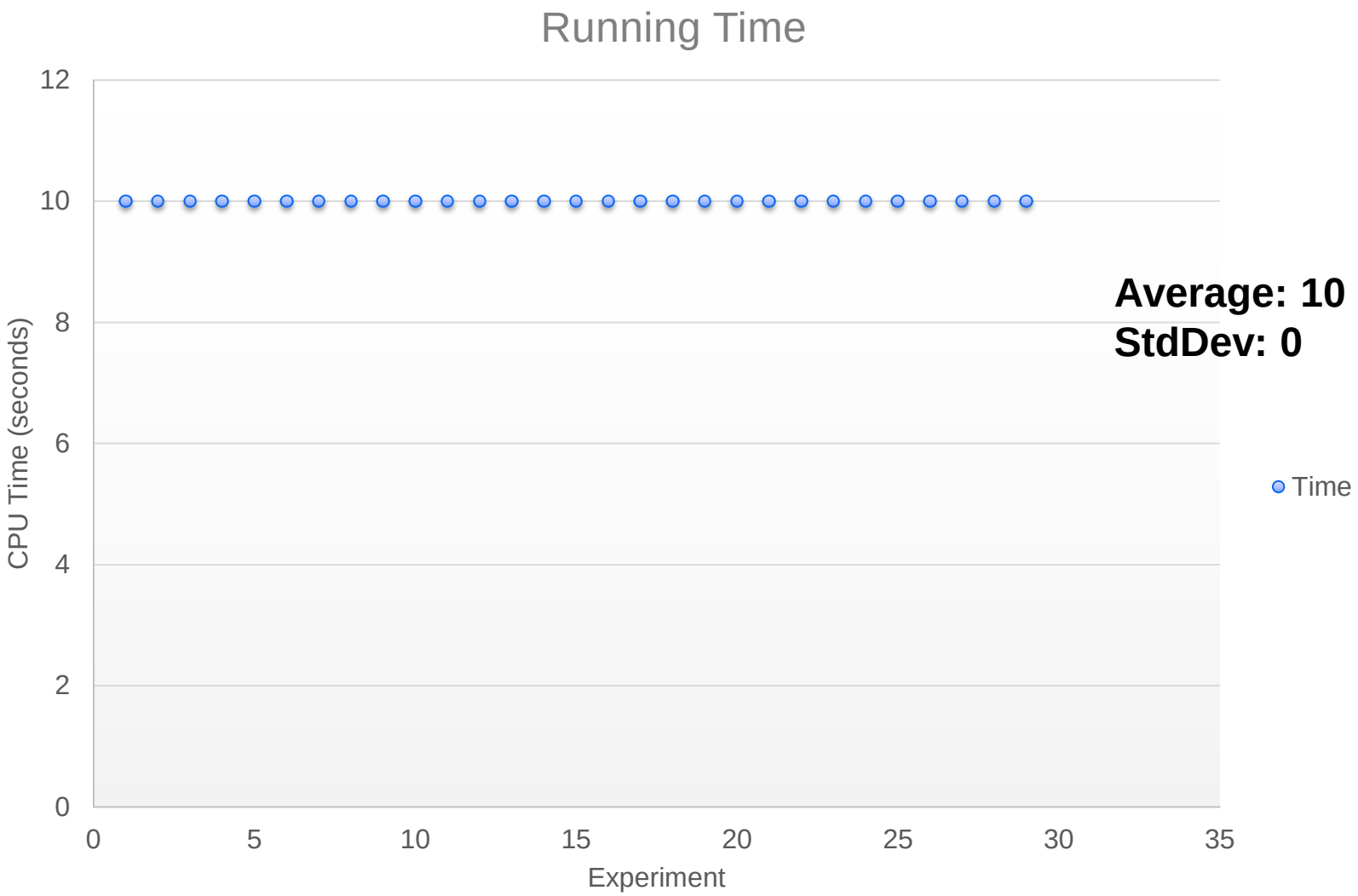
Methods & Tools for Software Engineering (MTSE)
Fall 2017

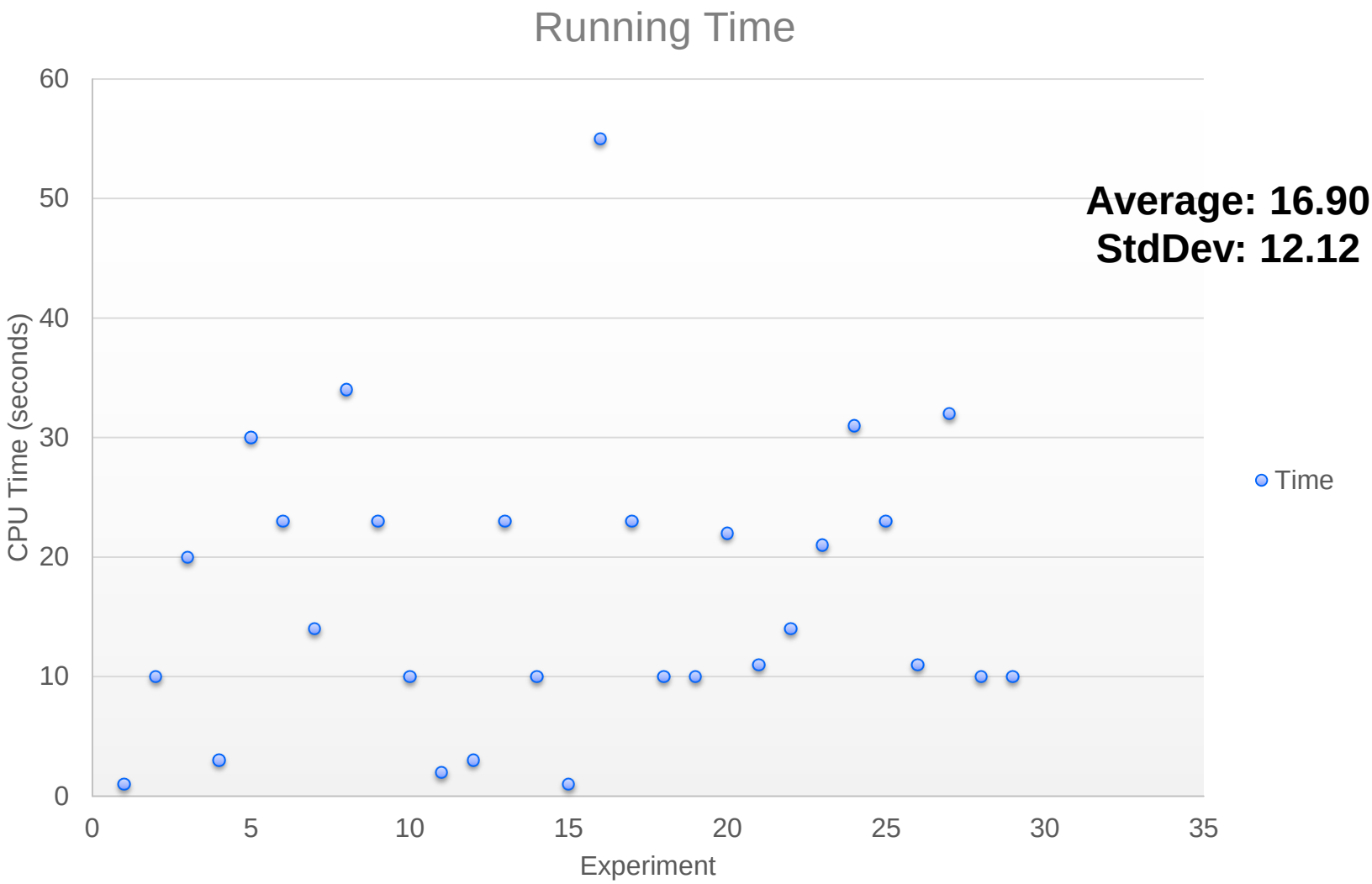
Prof. Arie Gurfinkel


Running Time
CPU Time (seconds)
12
10
8
6
4
2
0
0
5
10
15
20
25
30
35
Experiment
Time


Running Time
CPU Time (seconds)
60
50
40
30
20
10
0
0
5
10
15
20
25
30
35
Experiment
Time

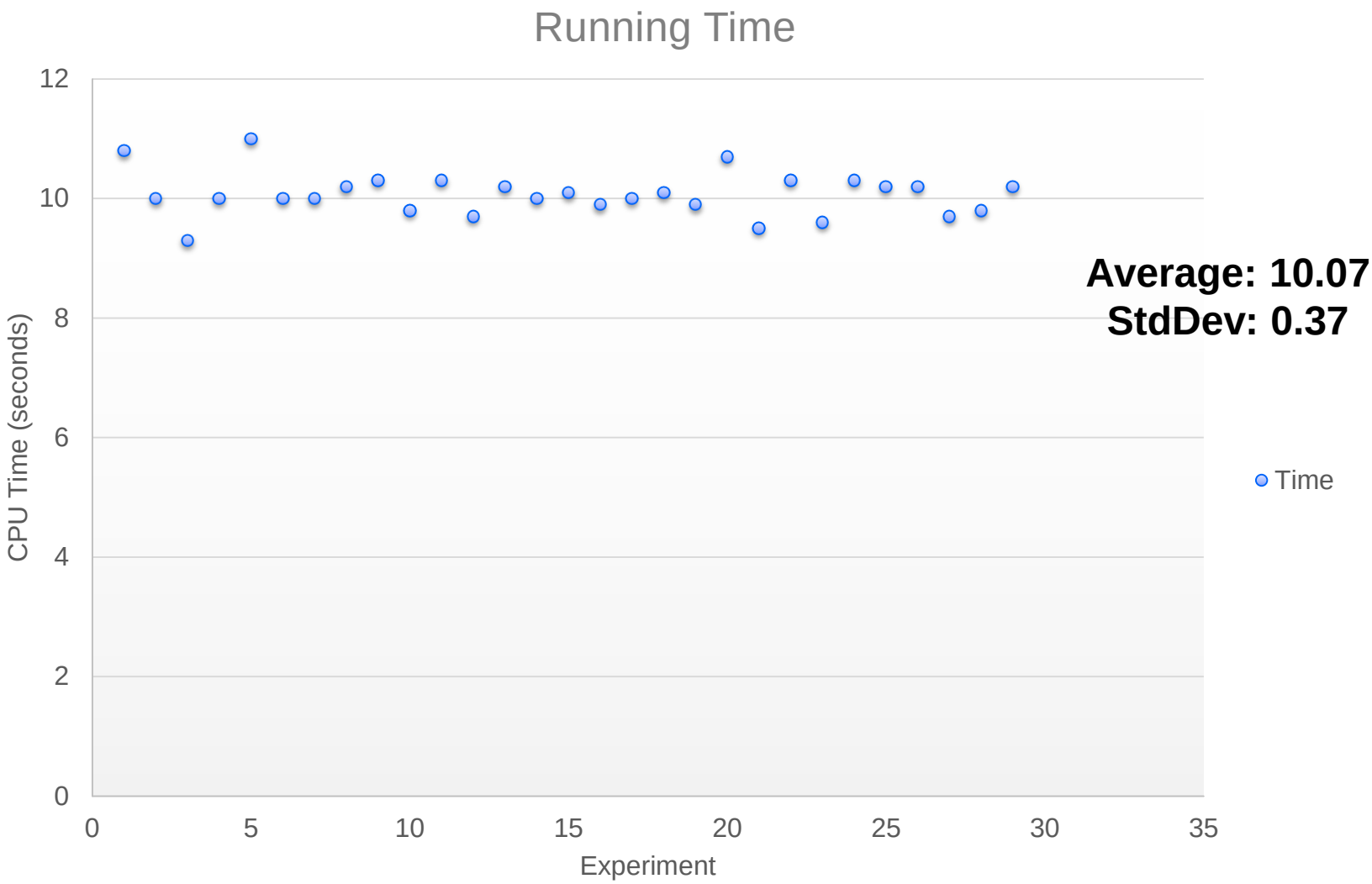

Running Time
CPU Time (seconds)
12
10
8
6
4
2
0
0
5
10
15
20
25
30
35
Experiment
Time


Running Time
CPU Time (seconds)
12
10
8
6
4
2
0
0
5
10
15
20
25
30
35
Experiment
Time
Average: 10
StdDev: 0



Average: 16.90
StdDev: 12.12

UNIVERSITY OF WATERLOO


Running Time
CPU Time (seconds)
0
2
4
6
8
10
12
0
5
10
15
20
25
30
35
Experiment
Time
Average: 10.07
StdDev: 0.37

Average, Variance, and Standard Deviation

$$Average(X) = \frac{x_1 + \ \ldots + x_n}{n}$$

$$Variance(X) = Average\left(\{\left(x_i - Average(X)\right)^2 \;\middle|\; x_i \in X\}\right)$$

$$StdDev(X) = \sqrt{Variance(X)}$$

http://www.mathsisfun.com/data/standard-deviation.html

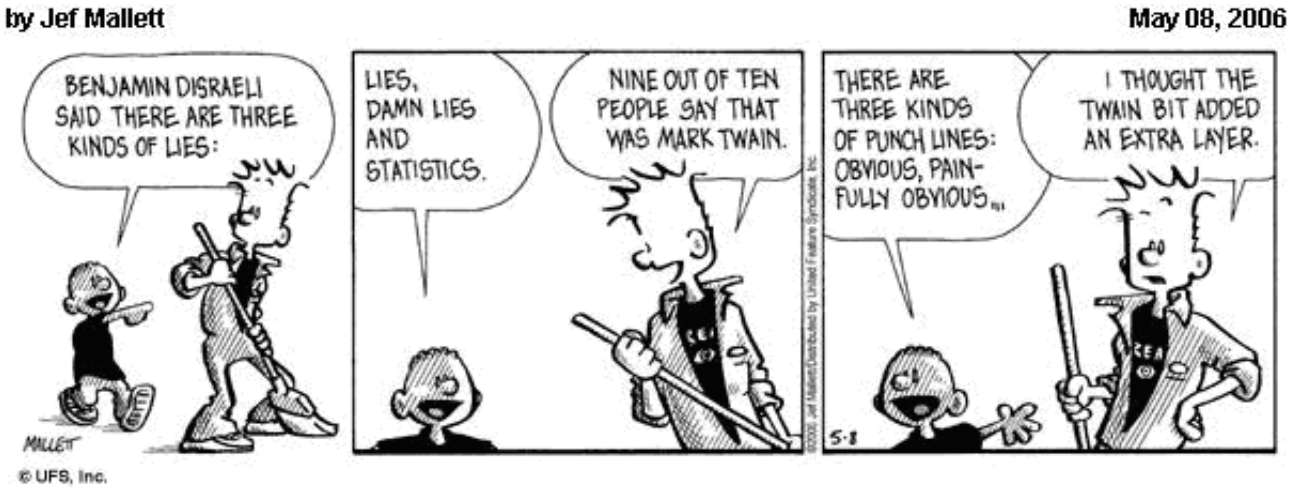
Statistical Summary

Statistics summarizes data

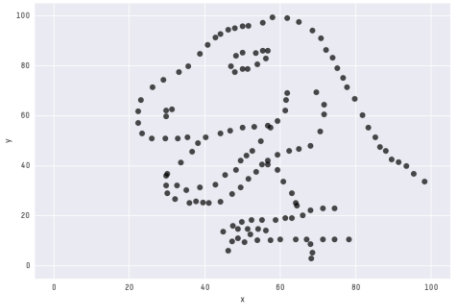
- hides (unimportant) detail
- highlights important detail
- simplifies interpretation

Given statistics about the data, do we still need to look at the data?

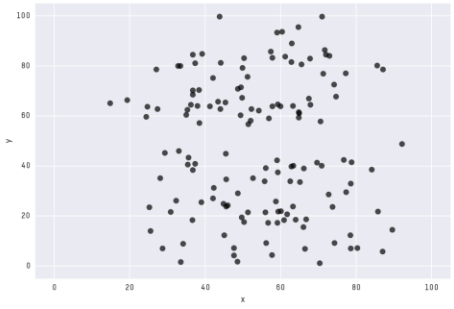
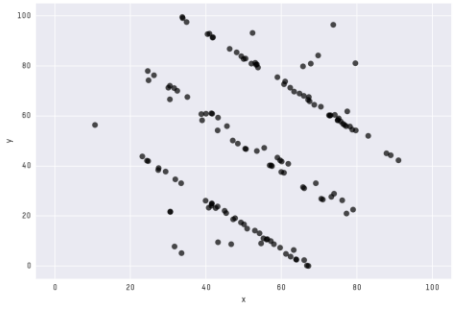
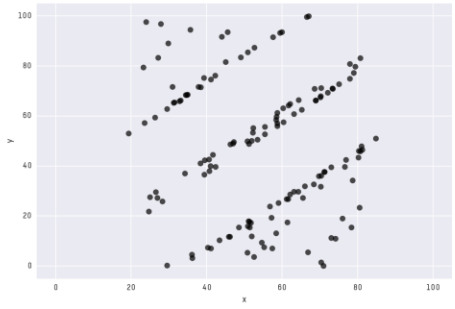
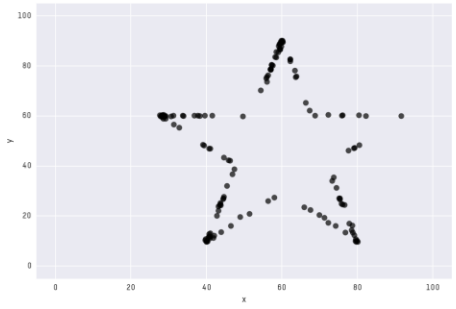
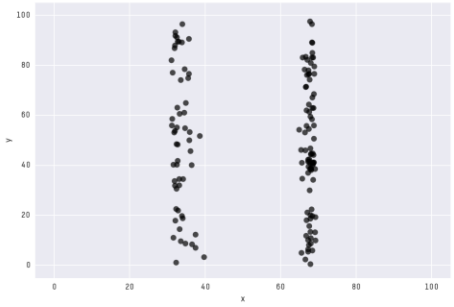
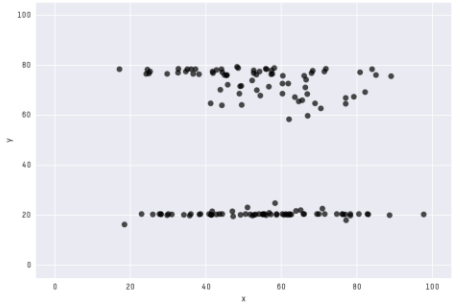
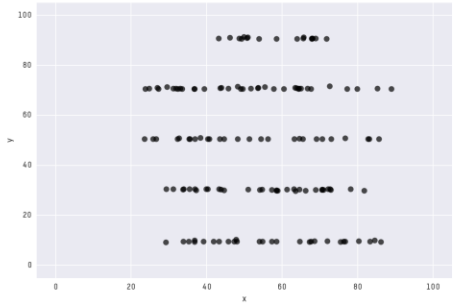
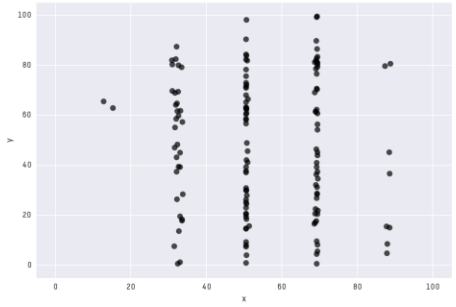
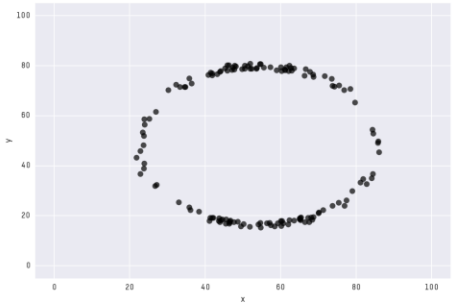
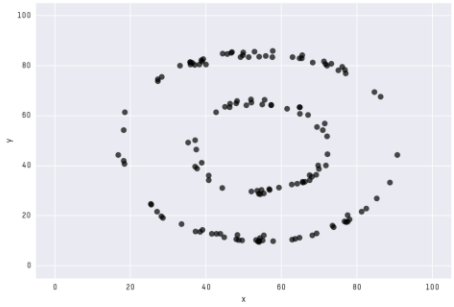
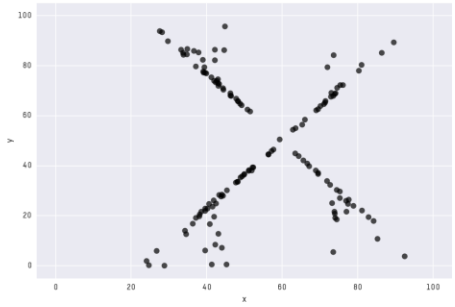
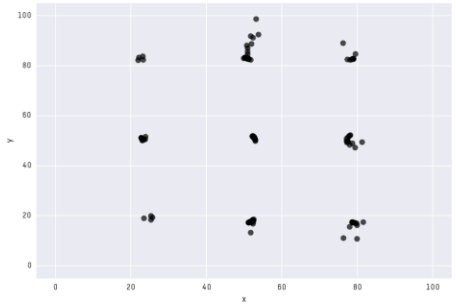
- are statistics enough?

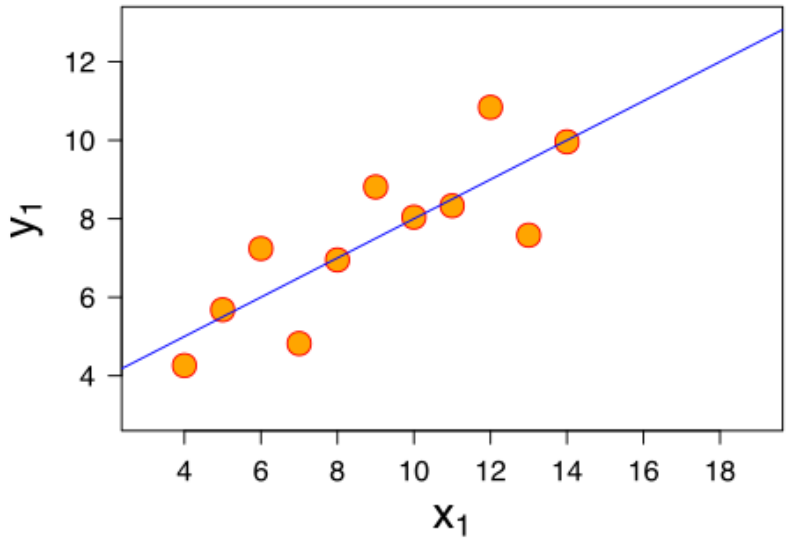


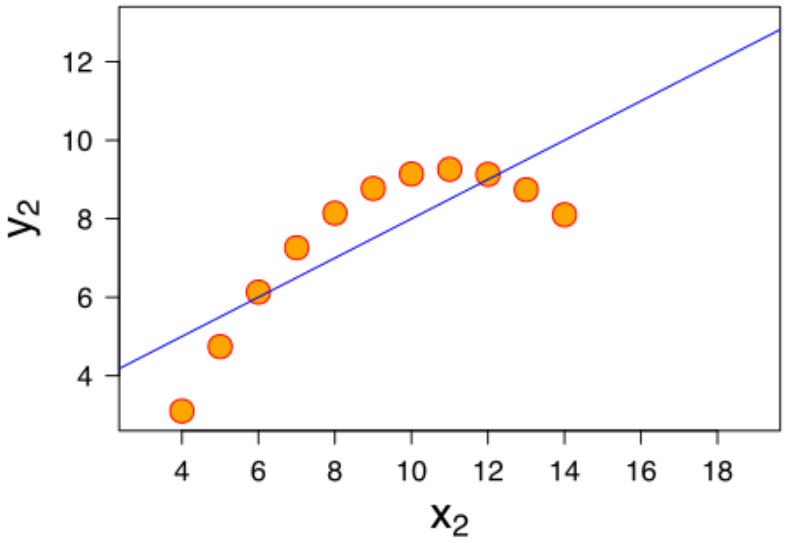
Let’s look at some data

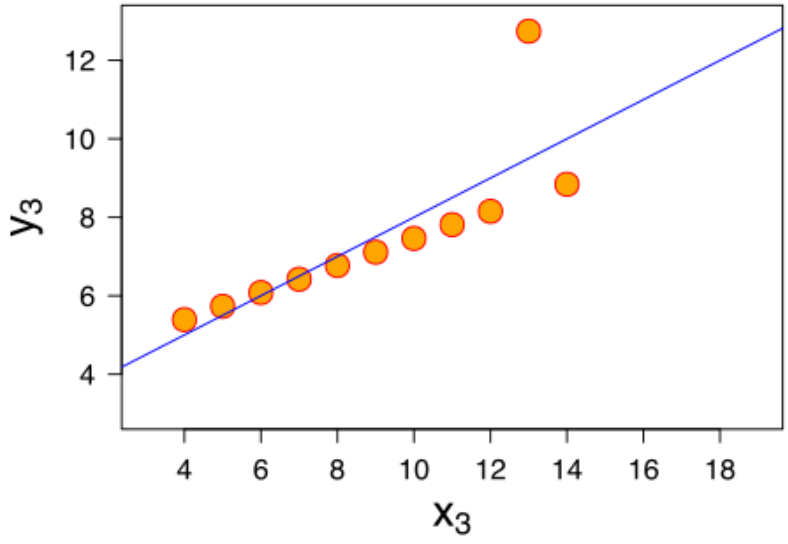


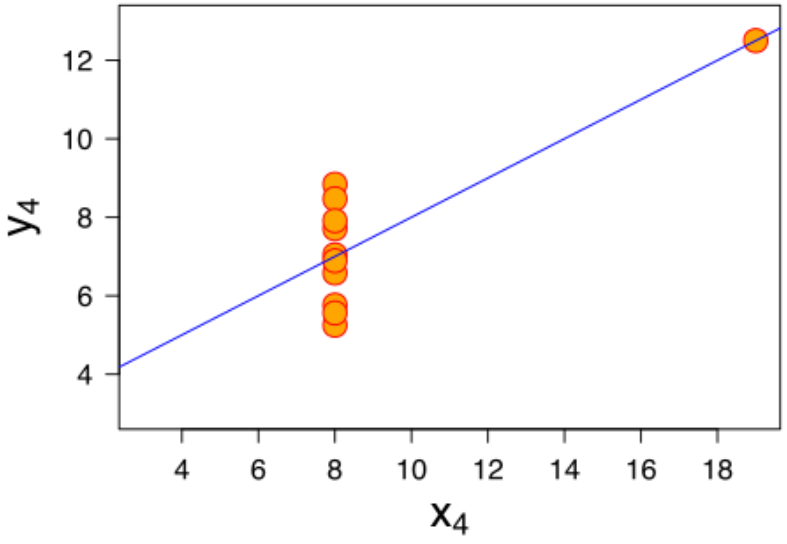
X Mean: 54.26
Y Mean: 47.83
X SD : 16.76
Y SD : 26.93
Corr. : -0.06


y1
x1
4
6
8
10
12
14
16
18
4
6
8
10
12


y2
x2
4
6
8
10
12
14
16
18
4
6
8
10
12


y3
x3
4
6
8
10
12
14
16
18
4
6
8
10
12


y4
x4
4
6
8
10
12
14
16
18
4
6
8
10
12

Summary

Data requires careful analysis

Statistics help summarize important aspects of the data

- hide unimportant detail
- highlight important detail

However, what is hidden might be IMPORTANT!

References

http://www.mathsisfun.com/data/standard-deviation.html

https://www.autodeskresearch.com/publications/samestats

Anscombe Quartet
https://en.wikipedia.org/wiki/Anscombe%27s_quartet