

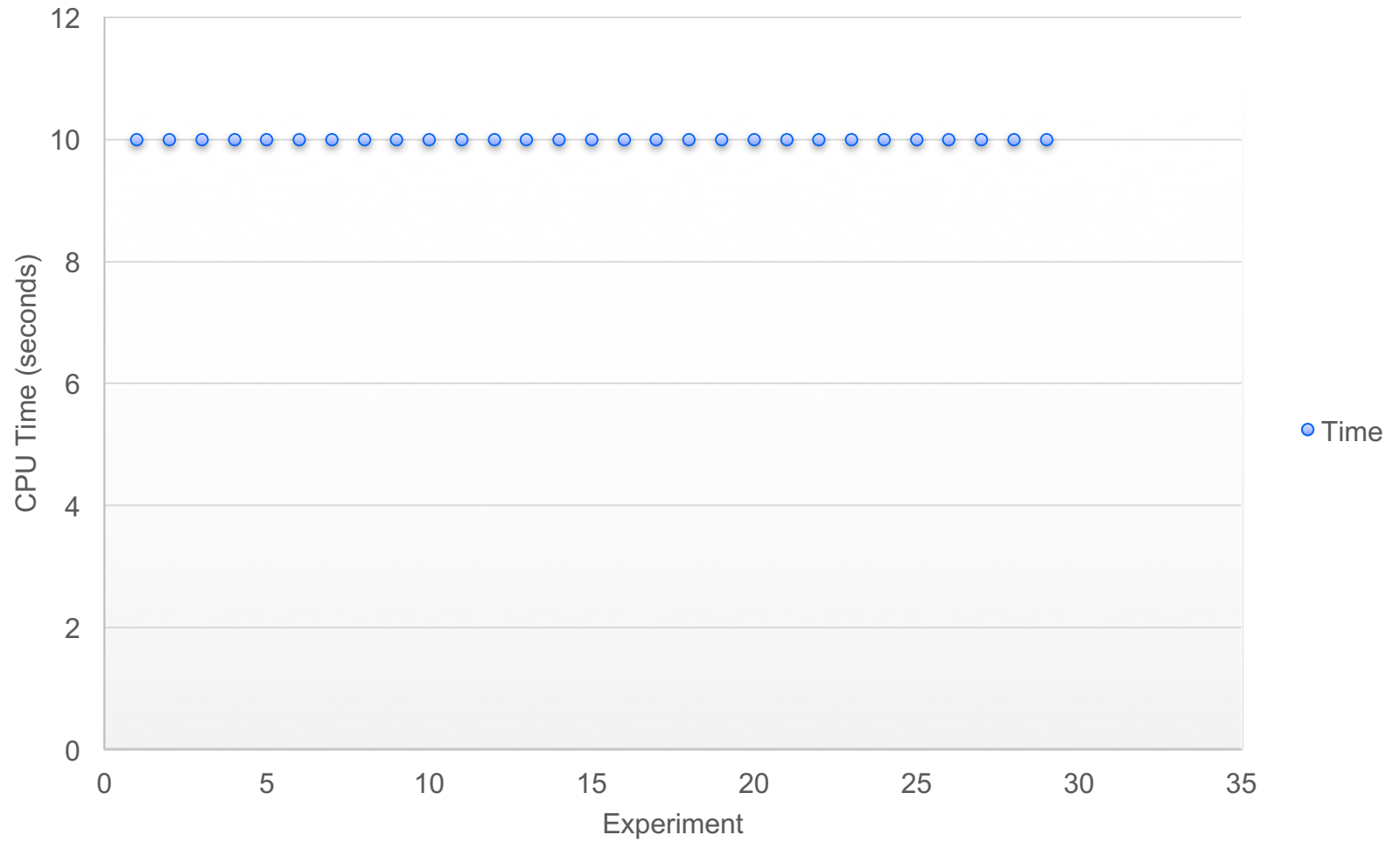
A Little Bit of Statistics

Methods & Tools for Software Engineering (MTSE)
Fall 2017

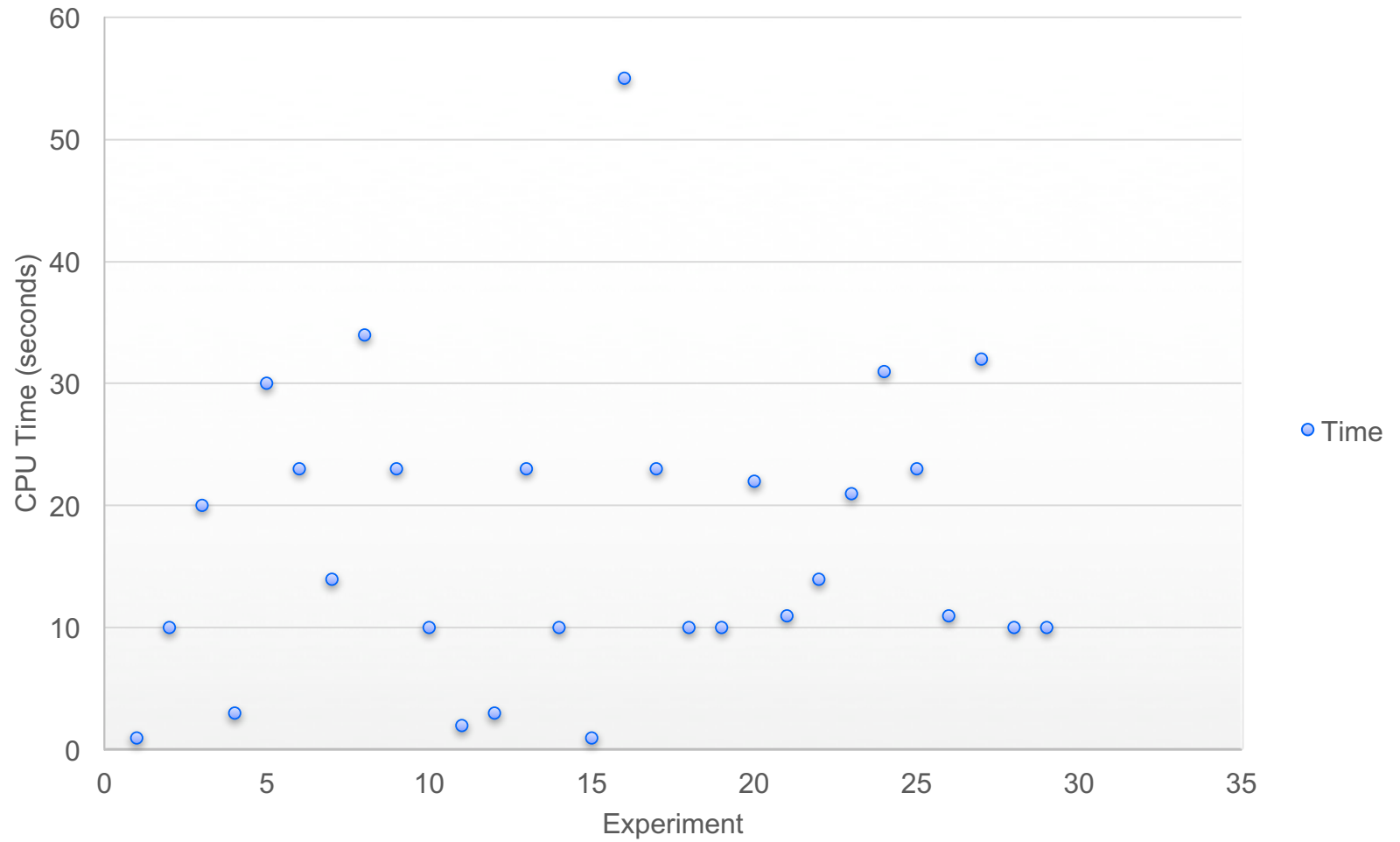
Prof. Arie Gurfinkel



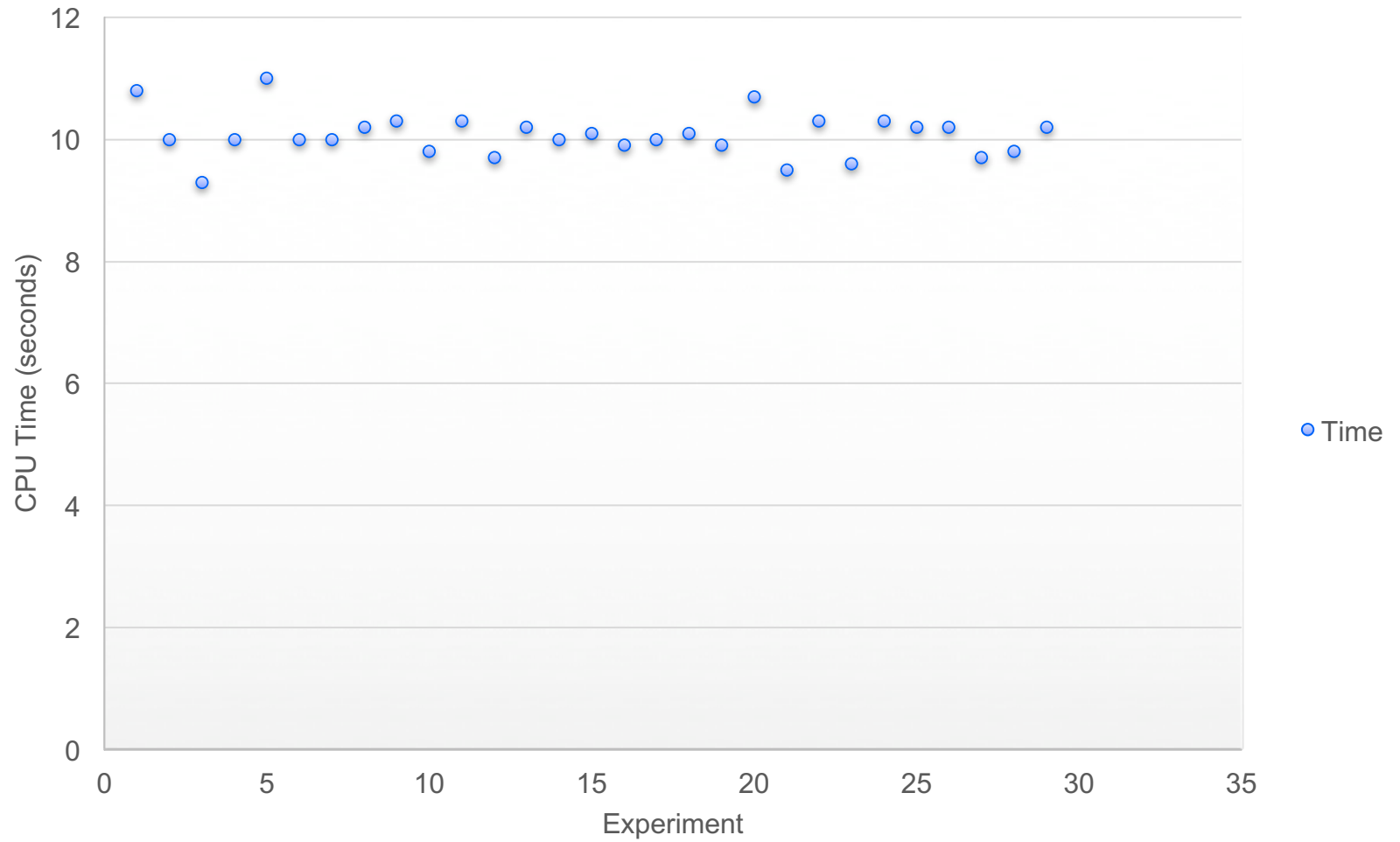
Running Time



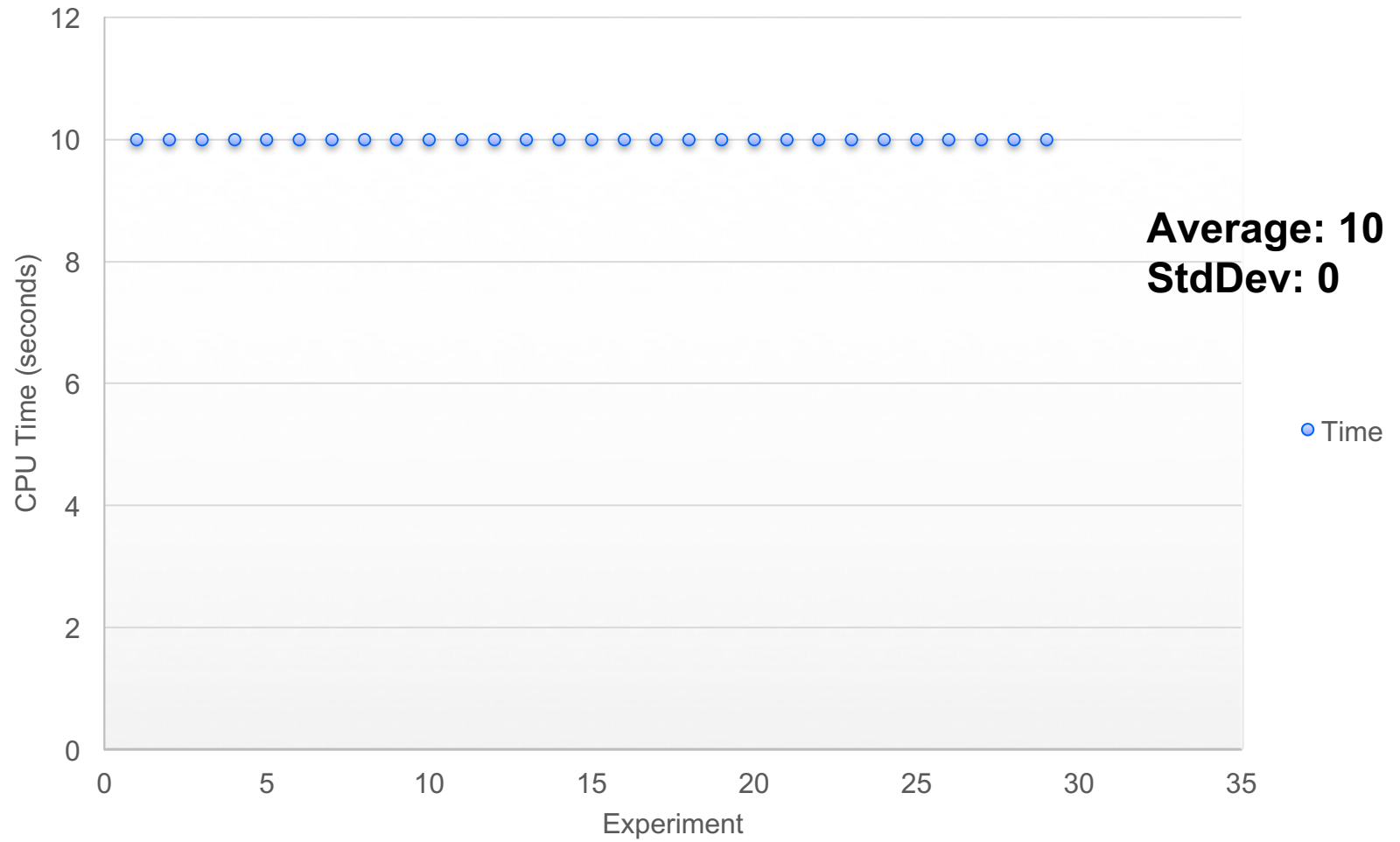
Running Time



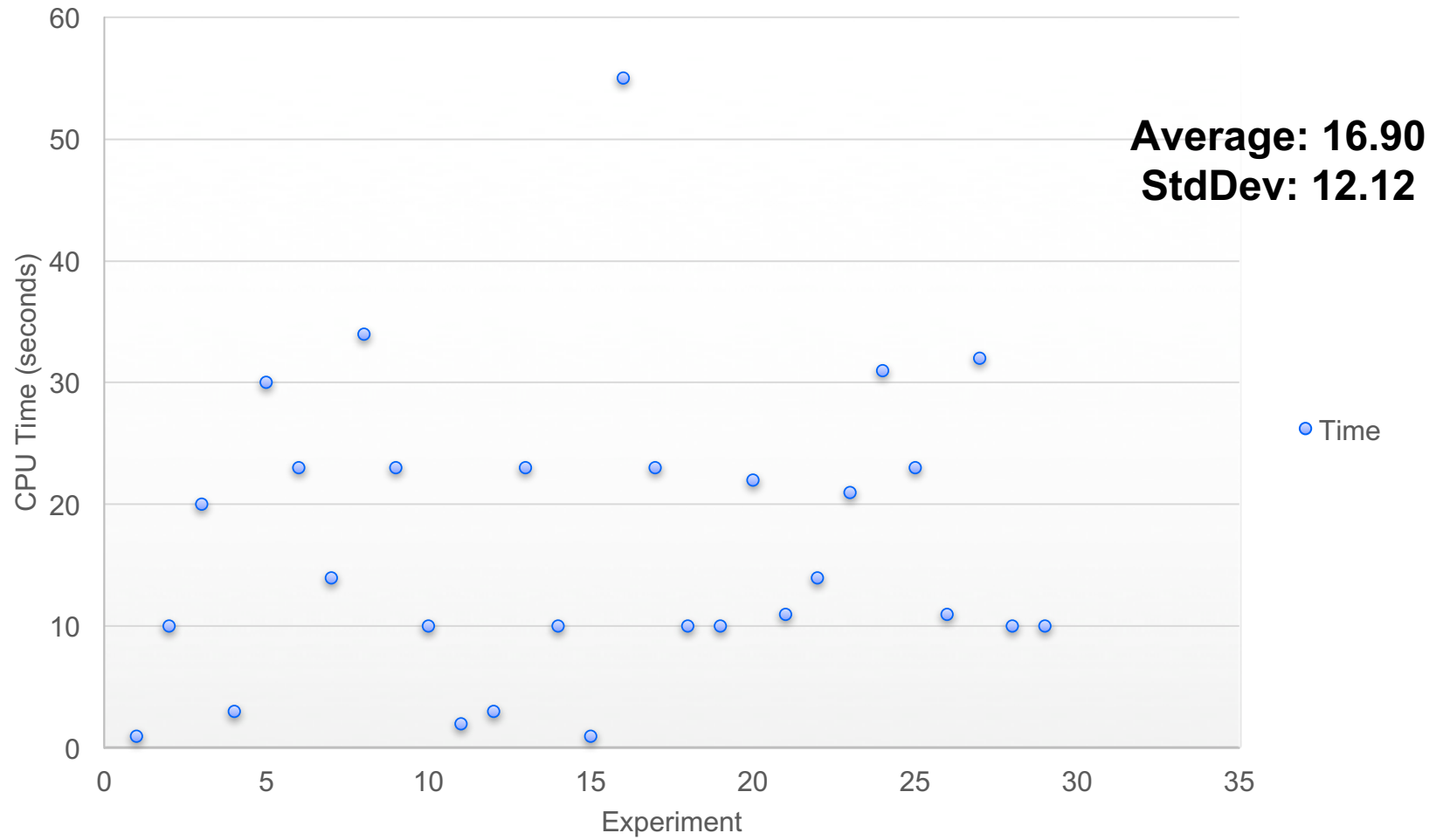
Running Time



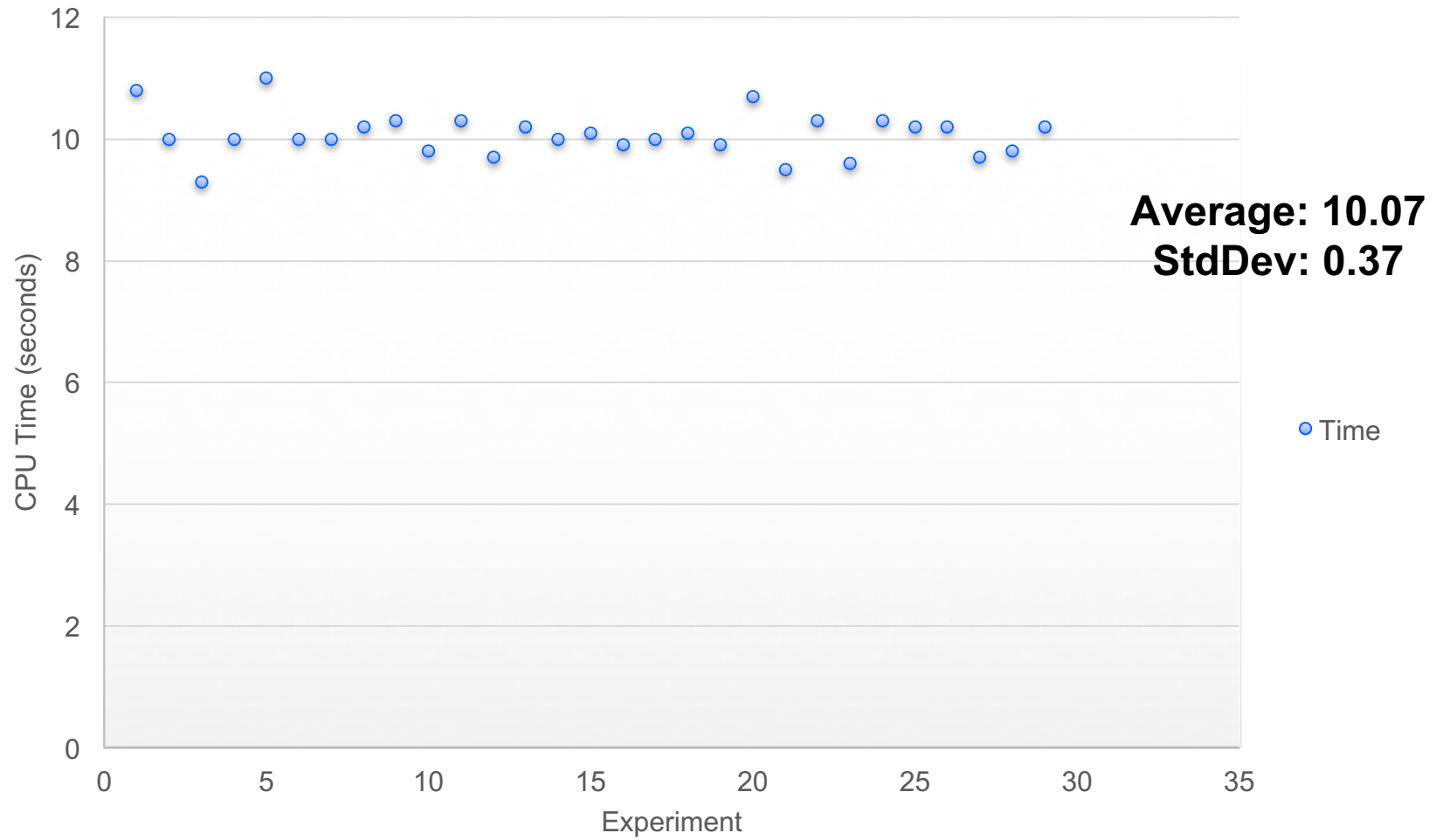
Running Time



Running Time



Running Time



Average, Variance, and Standard Deviation

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

$$\textit{Variance}(X) = \textit{Average} \left(\{(x_i - \textit{Average}(X))^2 \mid x_i \in X\} \right)$$

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

<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?

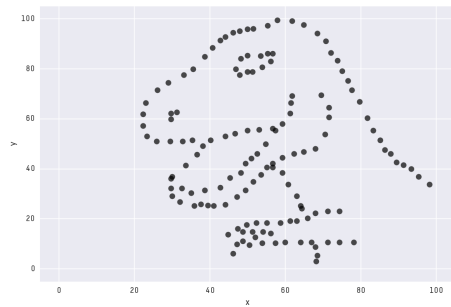
- are statistics enough?

by Jef Mallett

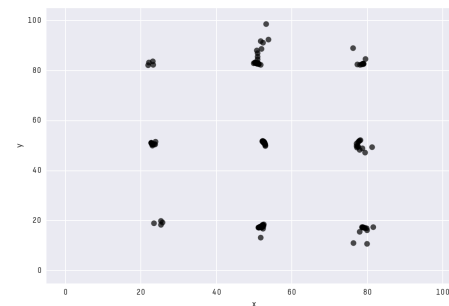
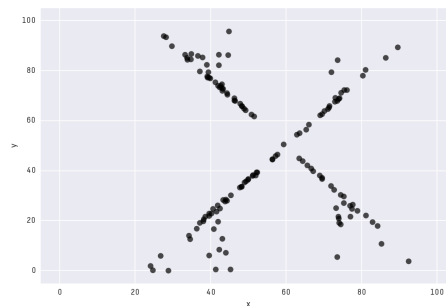
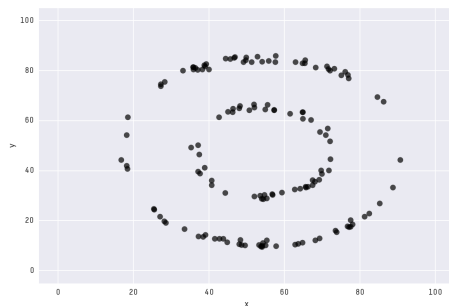
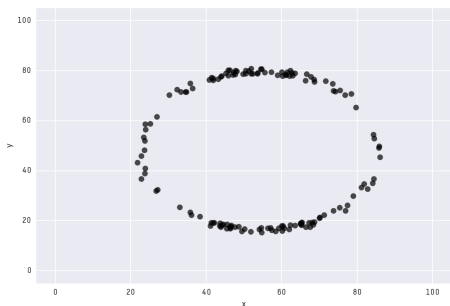
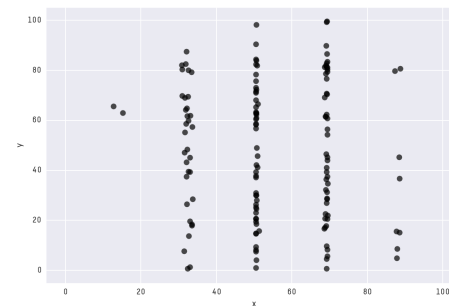
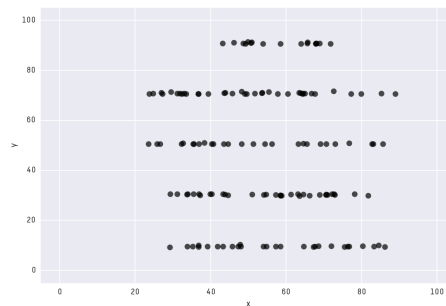
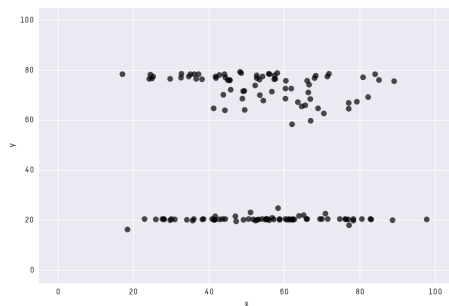
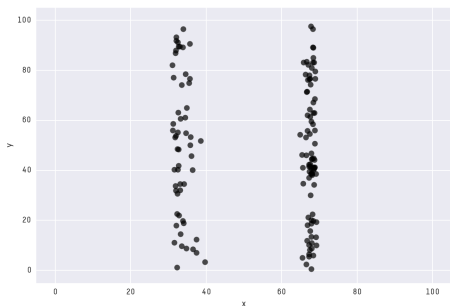
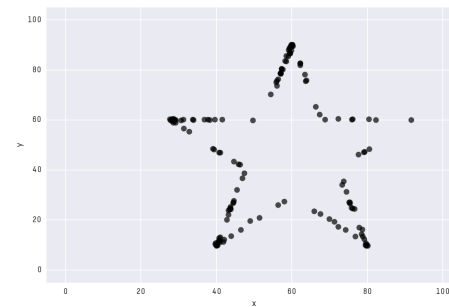
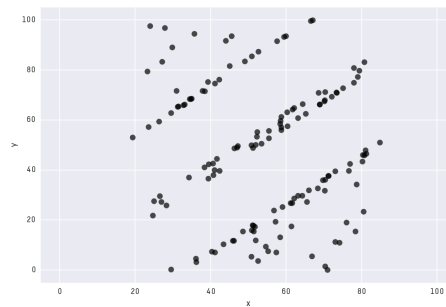
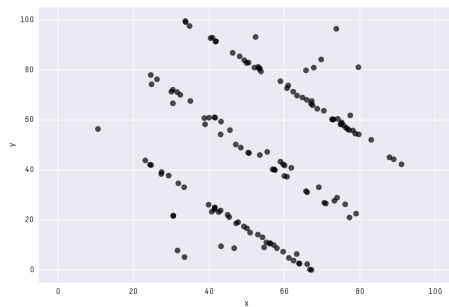
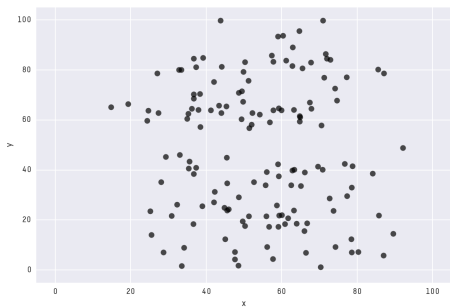
May 08, 2006

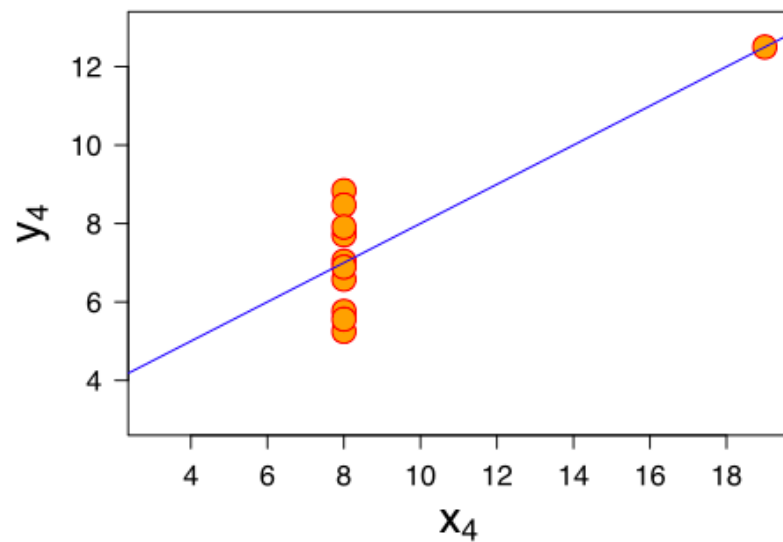
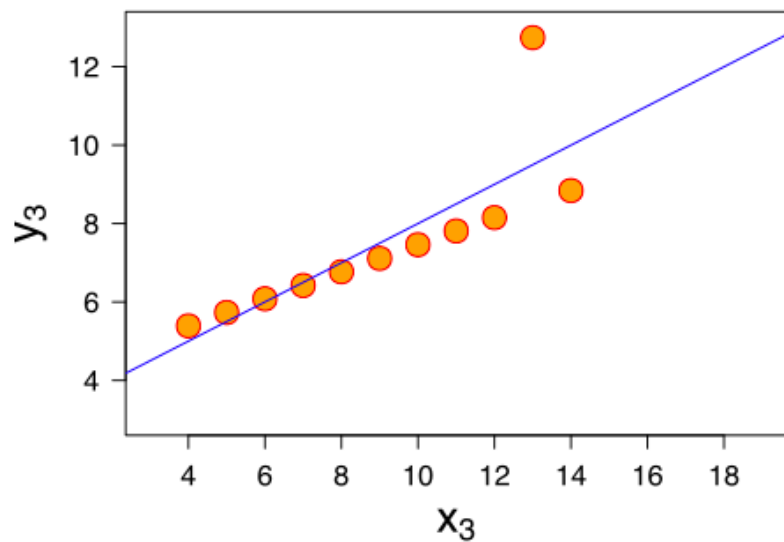
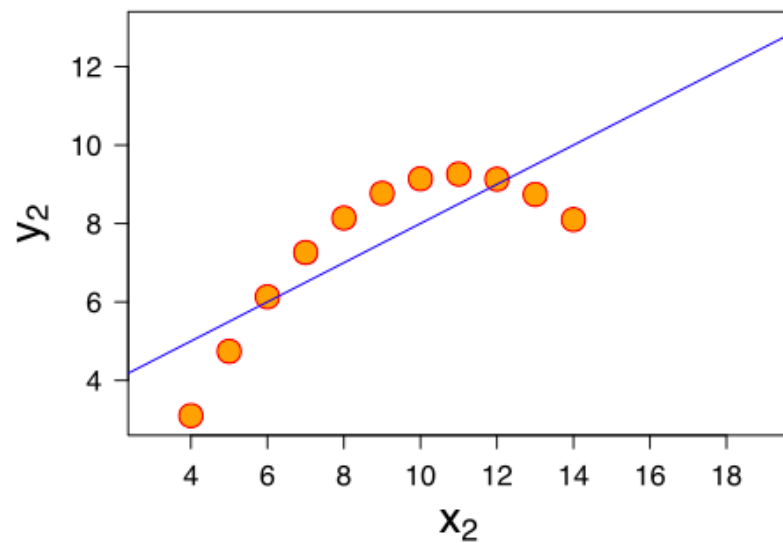
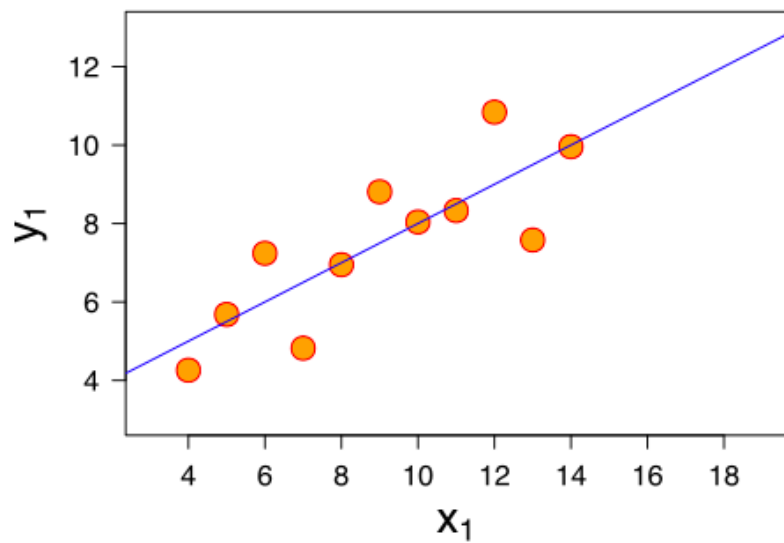


Let's look at some data



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





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