Statistics Part III — Graphical Display of Data

Joe Nahas

University of Notre Dame

UNIVERSITY OF
NOTRE DAME

Statistic Outline

1. Background:
   - A. Why Study Statistics and Statistical Experimental Design?
   - B. References
2. Basic Statistical Theory
   - A. Basic Statistical Definitions
     - i. Distributions
     - ii. Statistical Measures
     - iii. Independence/Dependence
       - a. Correlation Coefficient
       - b. Correlation Coefficient and Variance
       - c. Correlation Example
   - B. Basic Distributions
     - i. Discrete vs. Continuous Distributions
     - ii. Binomial Distribution
     - iii. Normal Distribution
     - iv. The Central Limit Theorem
       - a. Definition
       - b. Dice as an example

Statistic Outline (cont.)

3. Graphical Display of Data
   - A. Histogram
   - B. Box Plot
   - C. Normal Probability Plot
   - D. Scatter Plot
   - E. MatLab Plotting
4. Confidence Limits and Hypothesis Testing
   - A. Student's t Distribution
     - i. Who is "Student"
     - ii. Definitions
   - B. Confidence Limits for the Mean
   - C. Equivalence of two Means

Plots

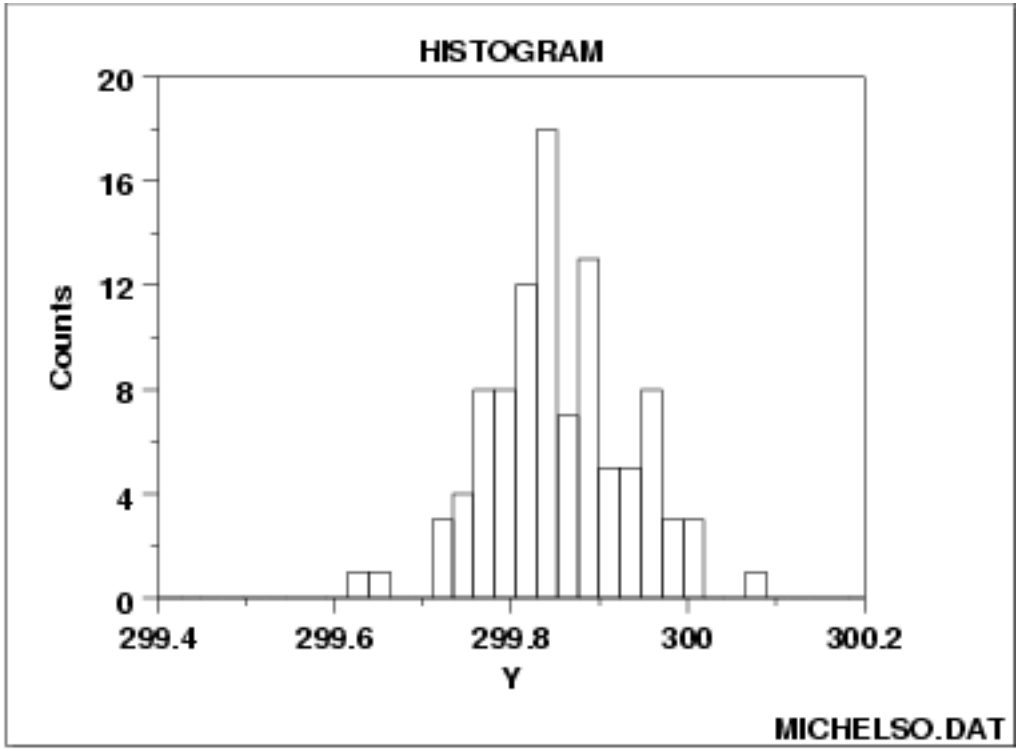
- Histogram
  - Look at distribution of data
- Box Plot
  - Compare multiple similar distributions of data
- Normal Probability Plot
  - Compare a distribution of data to the normal distribution
- Scatter Plot
  - Compare two distributions primarily for correlation

- These plots are available in the MatLab Statistical Package

Histograms

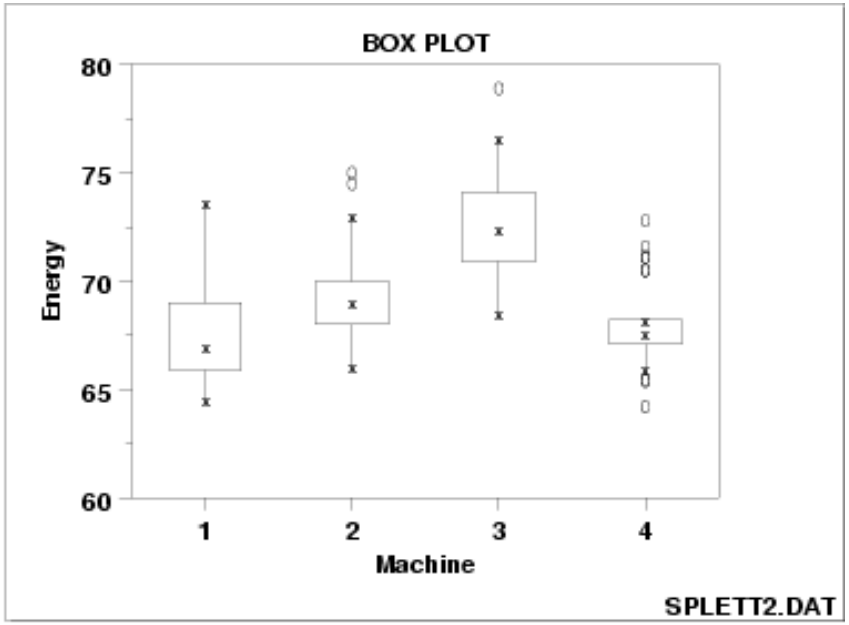
The histogram:

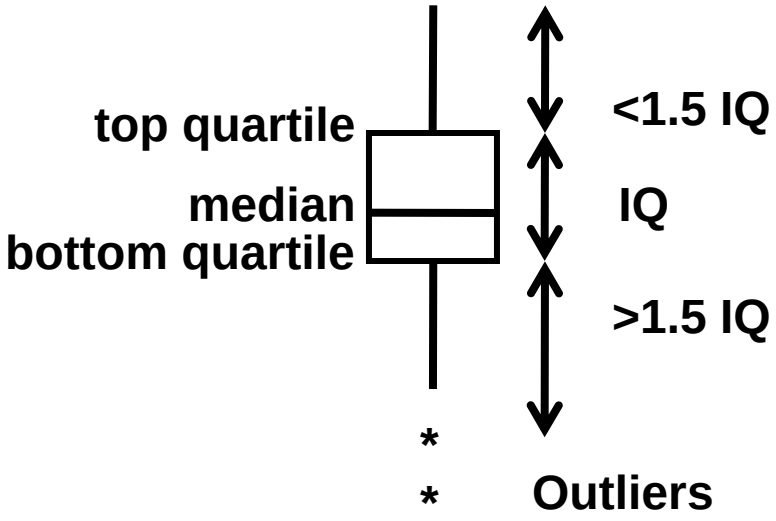
- **The area of the rectangle is proportional to the frequency of the observations within the specified interval.**
- **If the area of a specific rectangle is n_i/N, then the area of the whole histogram is 1.**



Box Plot

- The midpoint of the box is the median of the distribution.
- The top and bottom of the box are the 25% (lower quartile) and 75% (upper quartile) points of the distribution.
- The whiskers extend to the lowest and largest data points as long as they are within 1.5X of the lower quartile to upper quartile range (IQ) from the top or bottom quartile.
- Other points (outliers) are denoted by circles or asterisks.

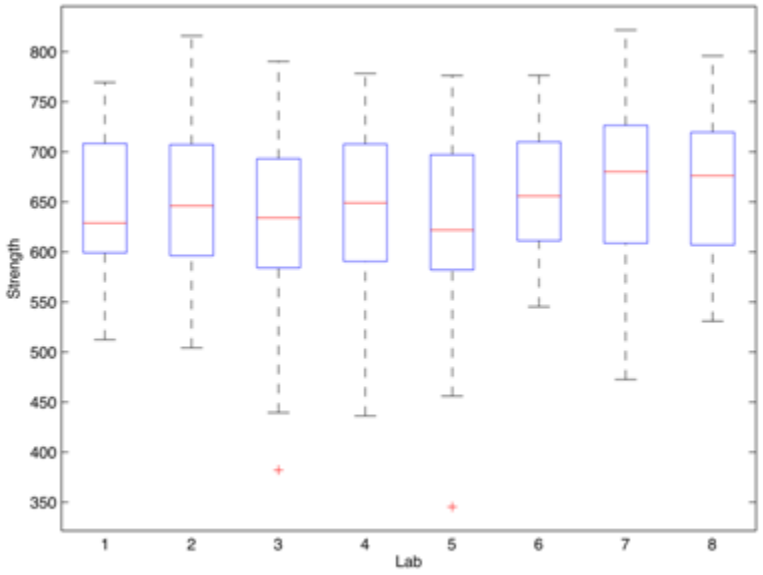




MatLab Box Plot

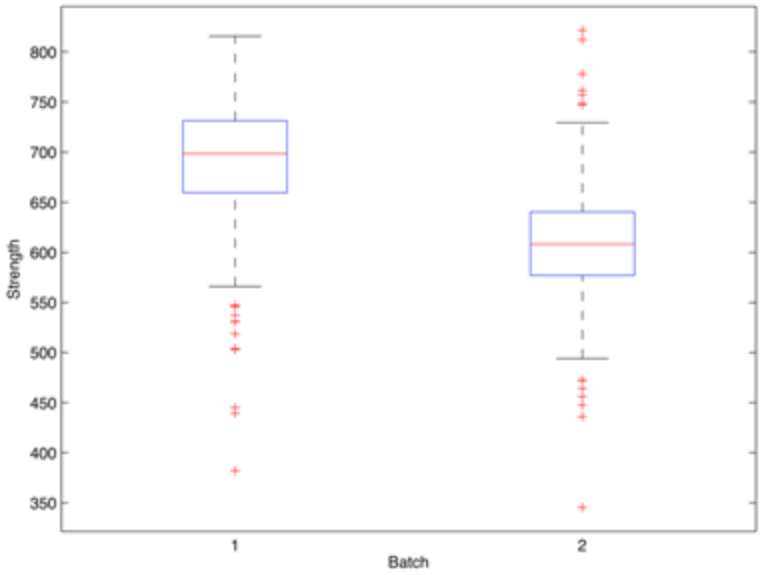
- **MatLab Box from NIST ESH Ceramics Data**

Ceramic Strength by Lab



Not much difference in distribution

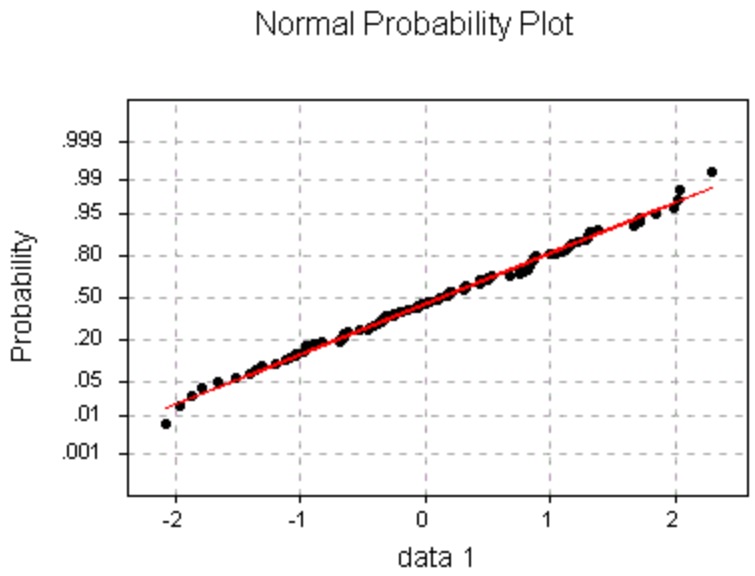
Ceramic Strength by Batch Order



Different medians
Similar Spread
Many outliers

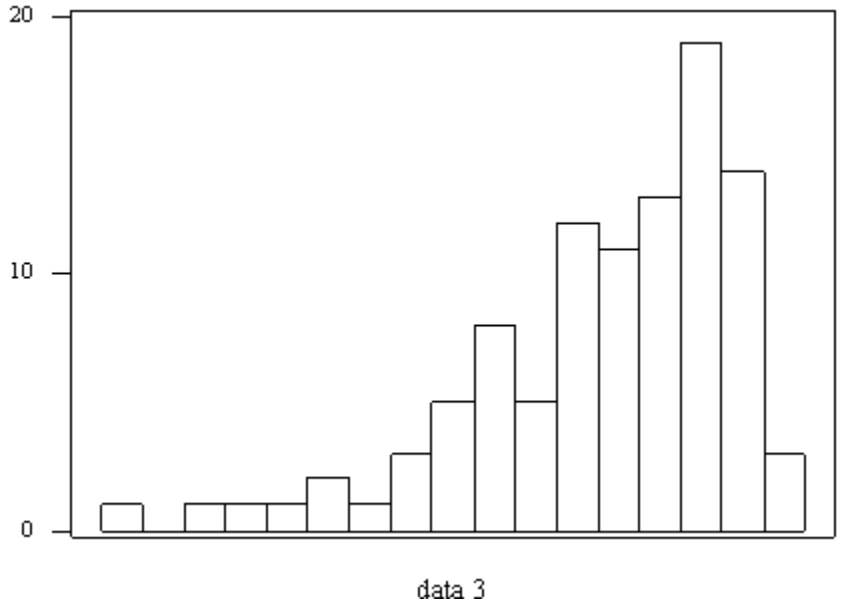
Normal Probability Plot

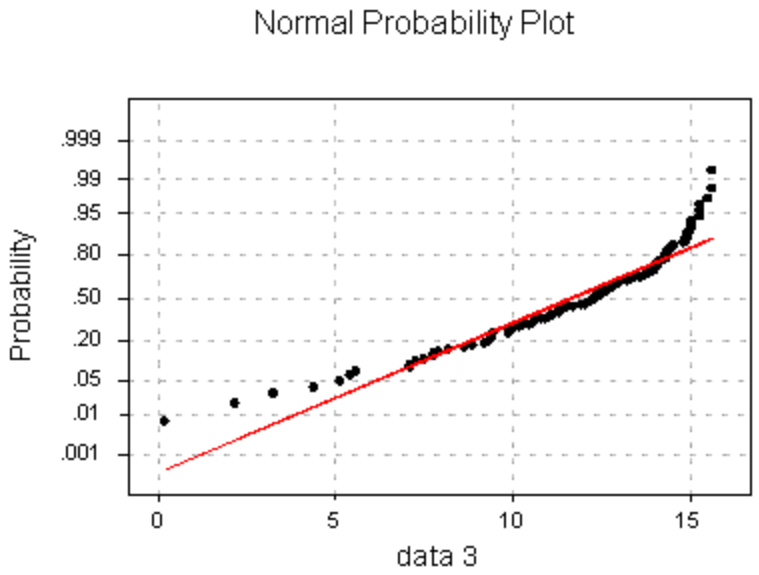
- The data values are plotted against a normal distribution.
- If the distribution is Normal, the data will show a straight line.
- If the distribution is not Normal, the data will deviate from the line.



Normal Probability Plot

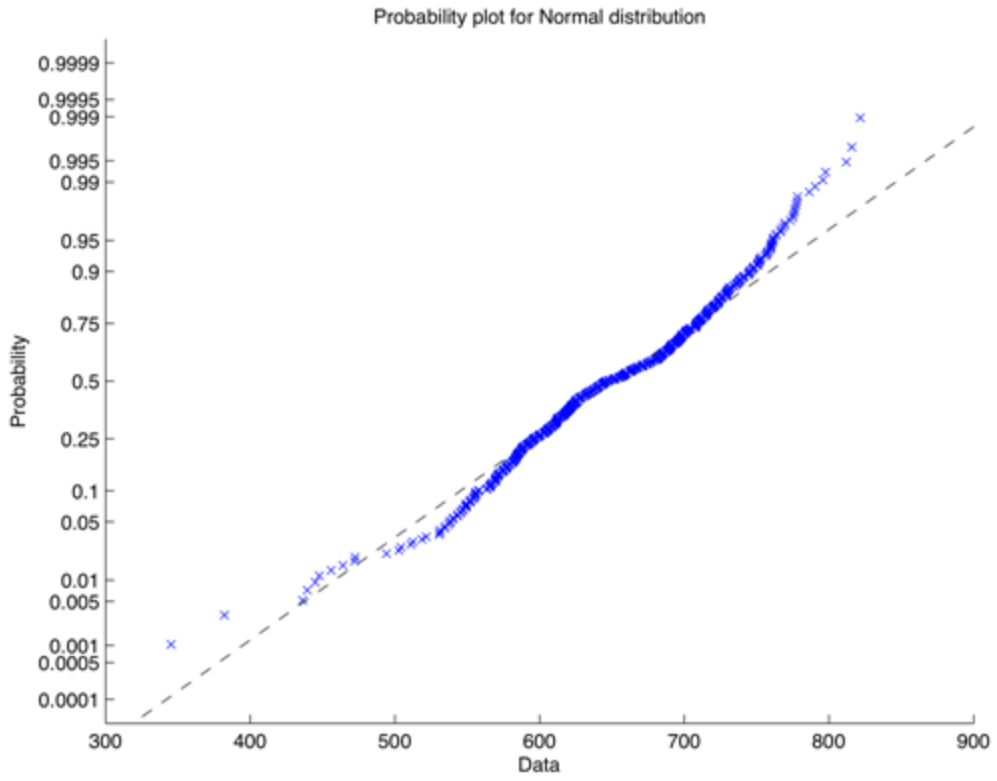
- **Normal Probability Plot with right skewed data**





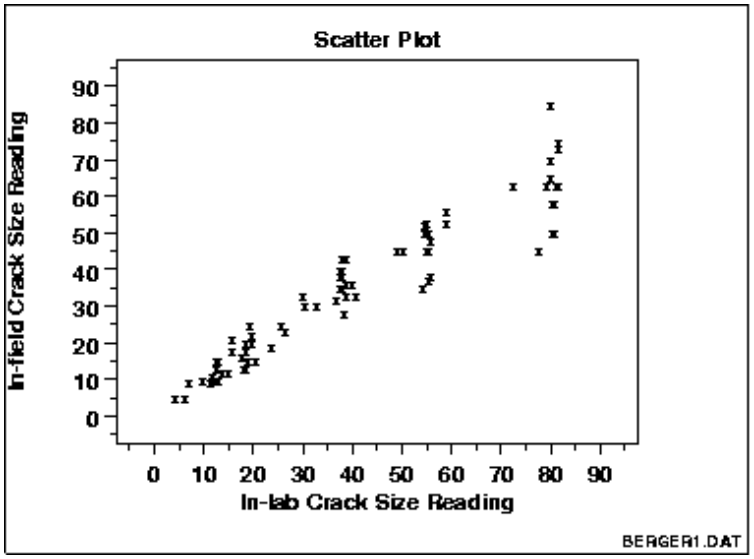
Normal Probability Plot

Normal Probability Plot of Ceramics Data from MatLab



Scatter Plot

- **Two factors from the same data set are plotted against each other to check for correlation**

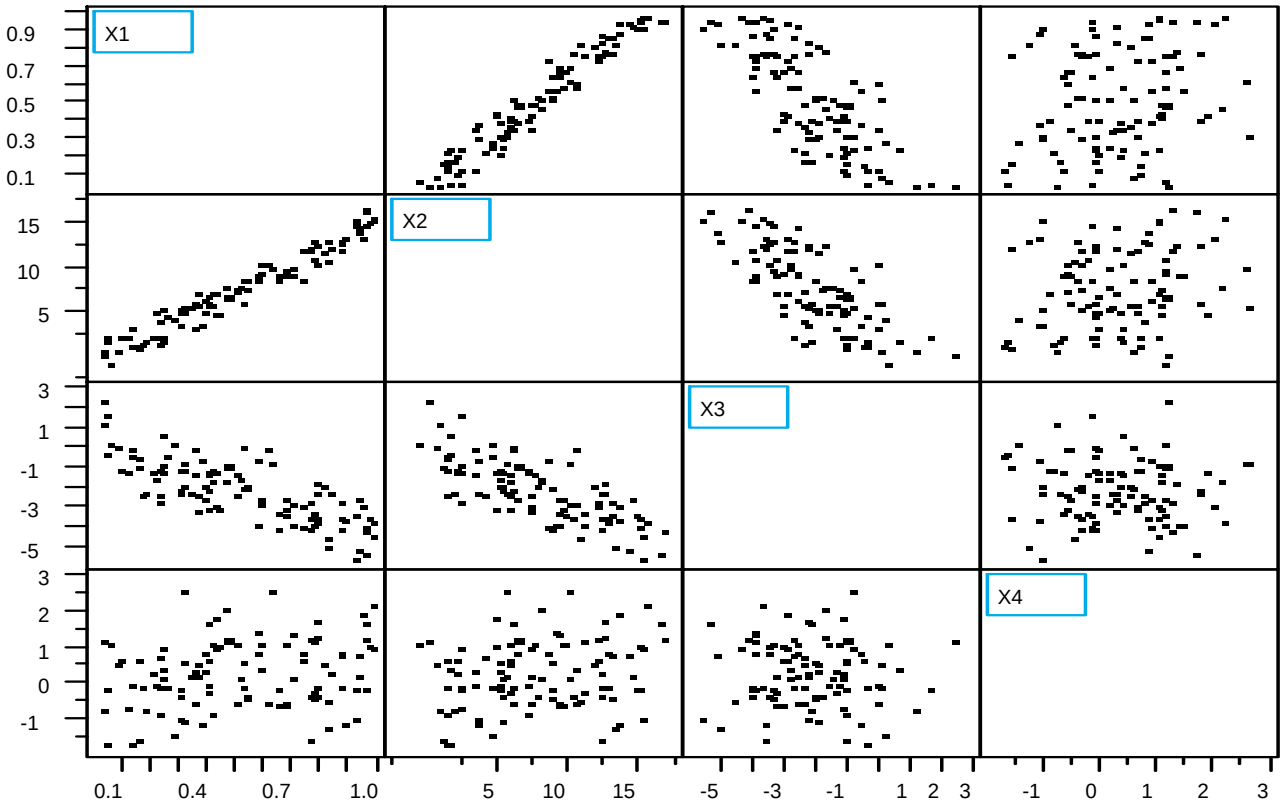


Correlation Example

Correlations

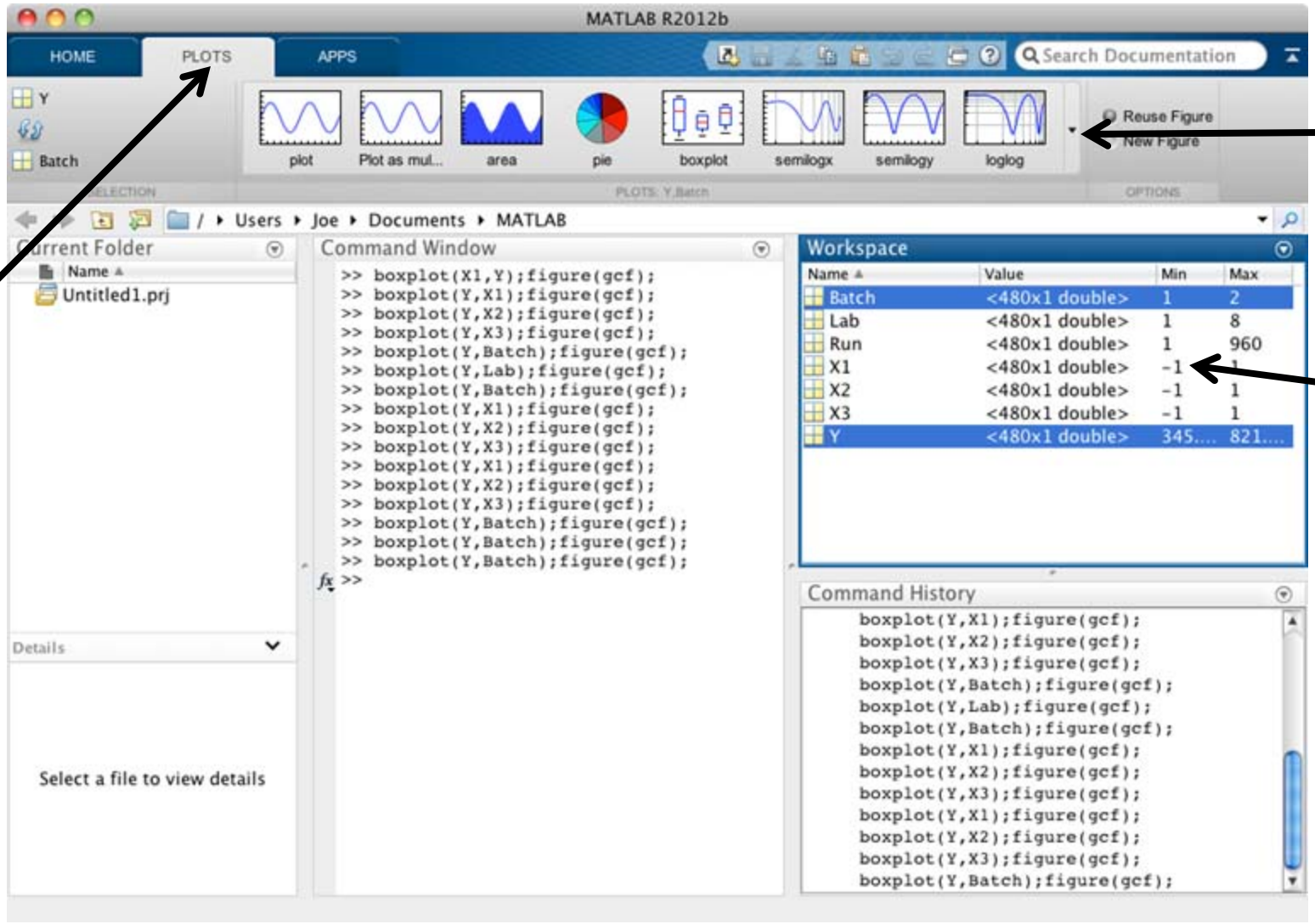
Variable	X1	X2	X3	X4
X1	1.0000	0.9719	-0.7728	0.2089
X2	0.9719	1.0000	-0.7518	0.2061
X3	-0.7728	-0.7518	1.0000	-0.0753
X4	0.2089	0.2061	-0.0753	1.0000

Scatter Plot Matrix



Poolla & Spanos

MatLab R2012b

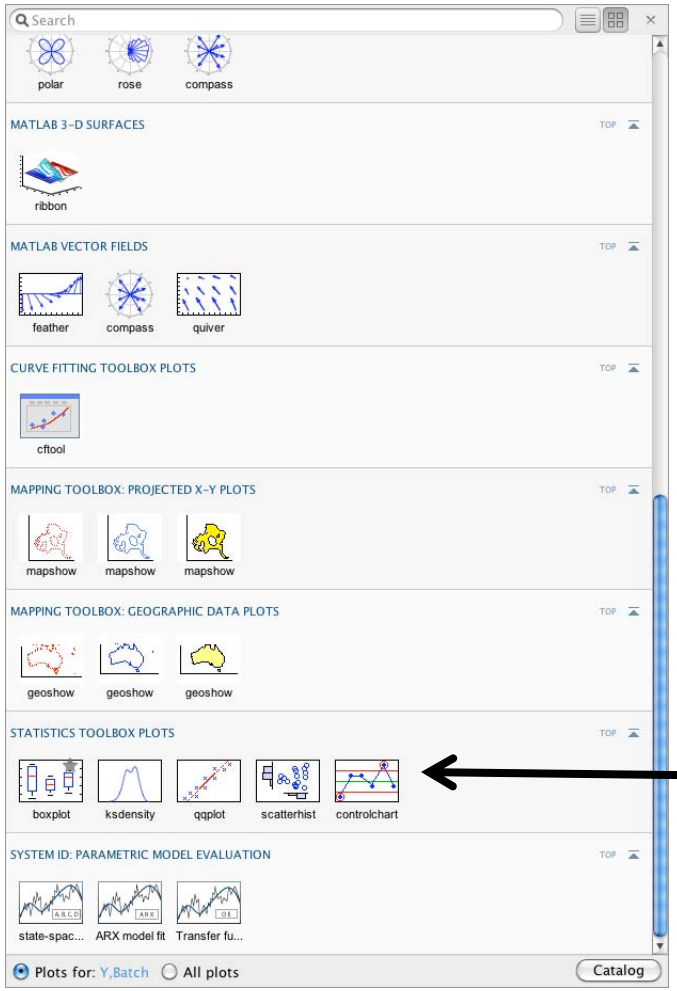


Plots Tab

More Plots Pull-Down

Data

MatLab Plots



Statistics Toolbox Plots
Available plots depends on data selected