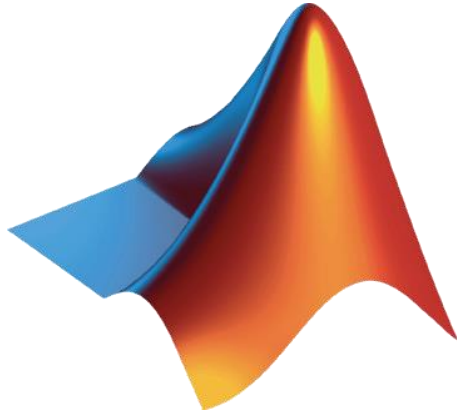
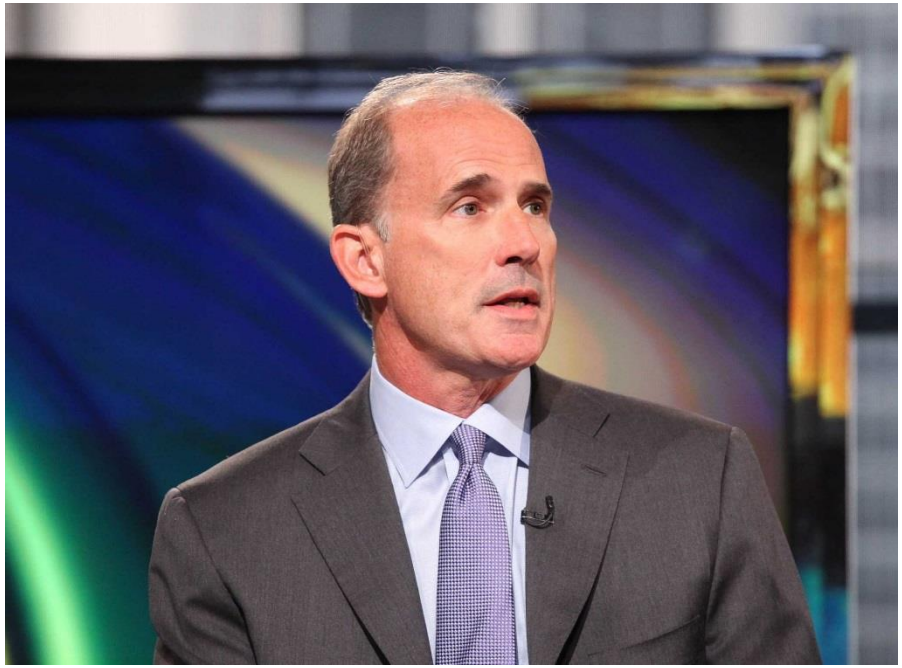
Beyond Excel: Enhancing Your Data Analysis with MATLAB

David Willingham
Senior Application Engineer – Data Analytics



© 2016 The MathWorks, Inc.

“Data is the sword of the 21st century, those who wield it the samurai.”



Google’s Former SVP - Jonathan Rosenberg

- Big data — how to create it, manipulate it, and put it to good use.
- “If you want to work at Google, make sure you can use MATLAB.”

MathWorks

Railway Systems

Automotive

Aeronautics

Retail

Finance

Off-highway vehicles

Condition Monitoring

Prognostics

Retail Analytics

Fleet Analytics

Operational Analytics

Data Analytics

Process Analytics

Risk Analysis

Health Monitoring

Supply Chain

Mfg Process Analytics

Asset Analytics

Healthcare Analytics

Internet

Industrial Automation

Logistics

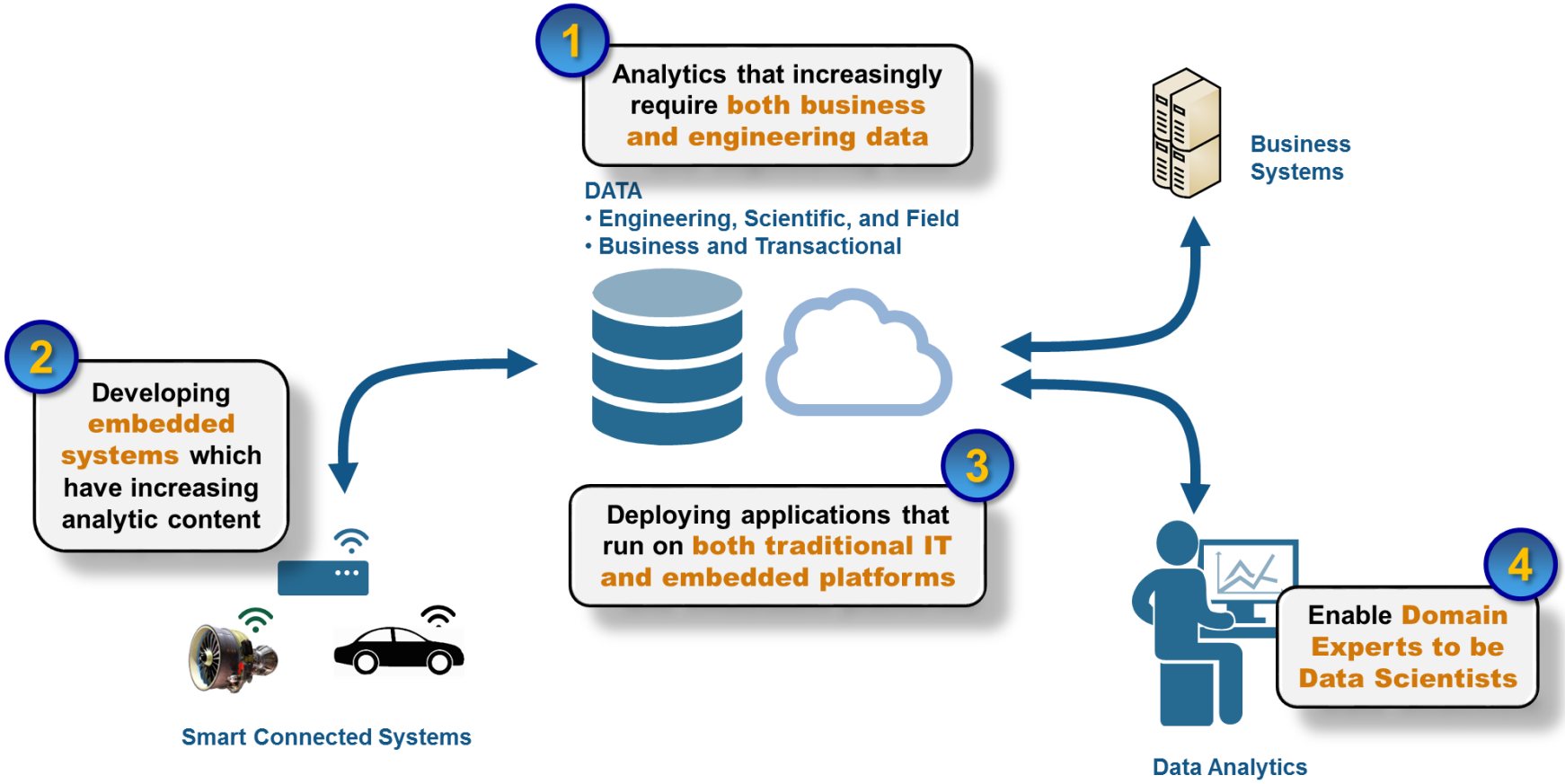
Oil & Gas

Clean Energy

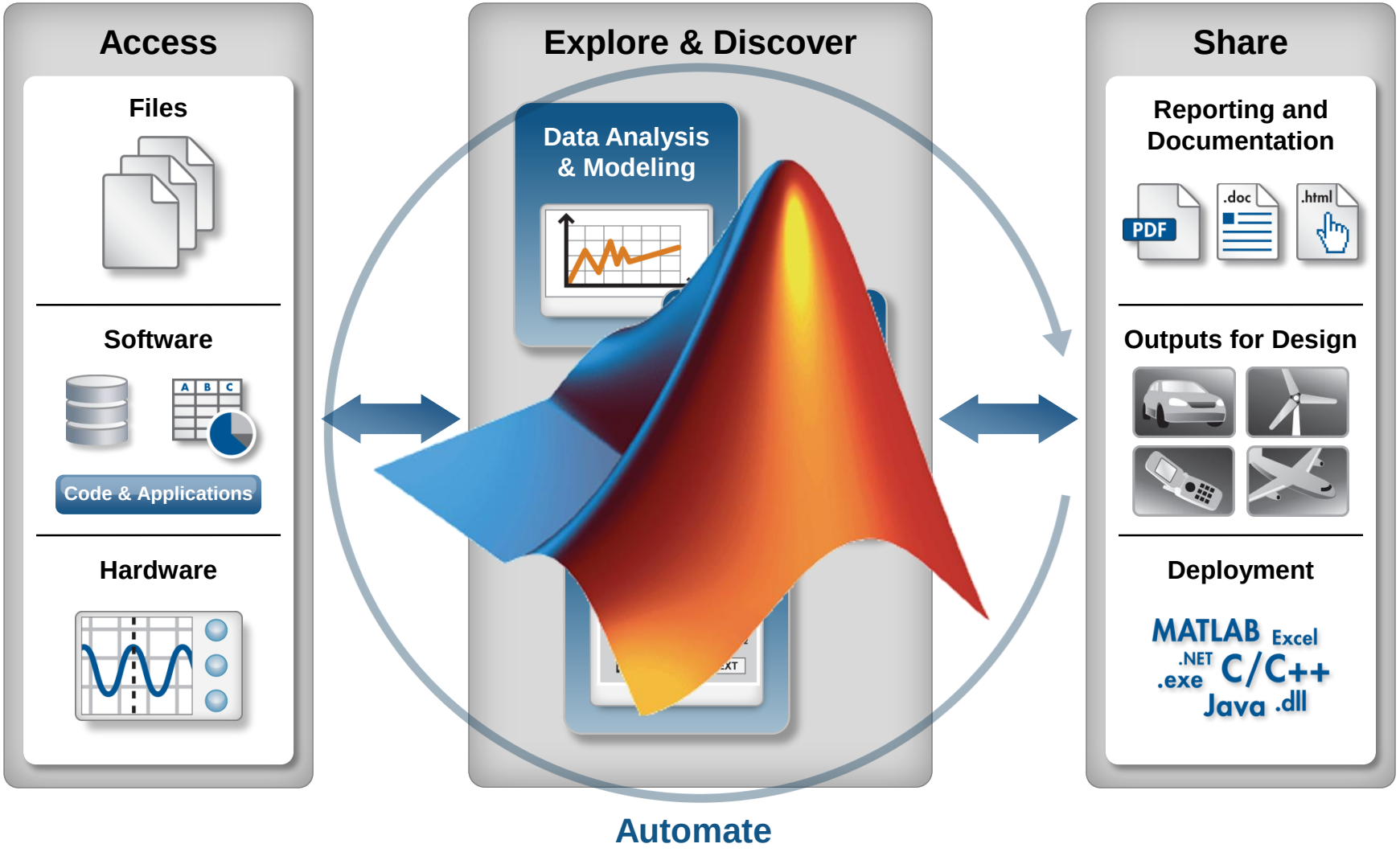
Medical Devices

Healthcare Management

What Analytics Challenges are you facing?



Data Analytics Workflow



Frontier Advisors Develops Web-Based Platform for Portfolio Analytics

Challenge

Provide clients with an industry-first web platform for portfolio modeling and analytics

Solution

Use MATLAB to develop and test analytics modules, and use MATLAB Compiler SDK to deploy them into a production .NET environment

Results

- Quantitative development decoupled from interface development
- Stable, responsive system deployed
- Rapid delivery of new features enabled

"MATLAB and MATLAB Compiler SDK enabled us to rapidly deliver a sophisticated portfolio analytics web application with confidence that it will return accurate results extremely quickly, ensuring a highly usable and stable platform for our clients."

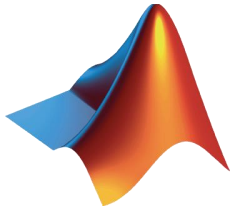
Lee Eriera

Frontier Advisors

Link to user story

Today's Objectives

- Introduce you to data analysis with MATLAB
- Show how you can overcome common data analysis challenges with MATLAB
- Demonstrate multiple ways of sharing your analysis and results with others



Common Data Analysis Challenges *using Excel*

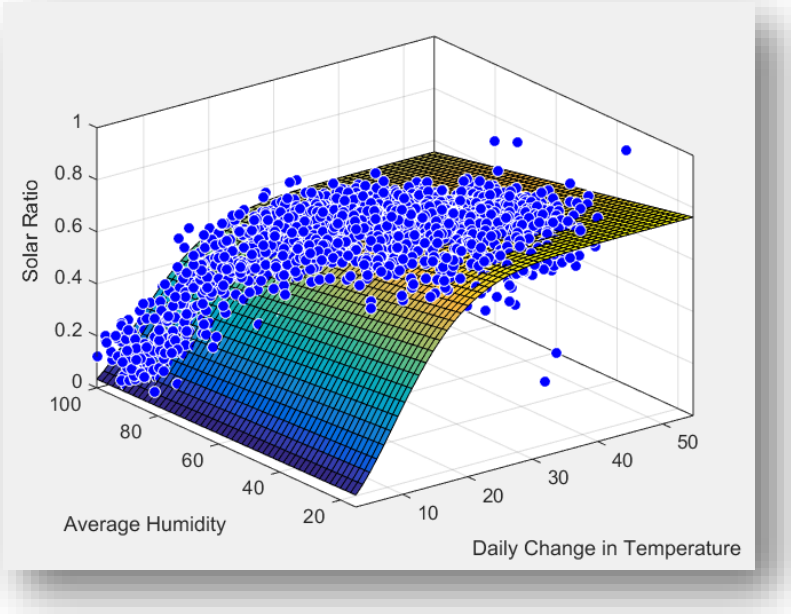
- Complex calculations
- Messy Data
- Speed of Execution
- Automation
- Batch Processing
- Report Generation
- Deployment

Demo: Solar Radiation Estimation

Introduction to Data Analysis with MATLAB

- Goal:
  - Estimate daily mean global solar radiation given low cost and easily obtained measurements

- Approach:
  - Process historical measurements
  - Develop predictive model
  - Document analysis in a report
  - Apply analysis on multiple files



Modeling Global Solar Radiation

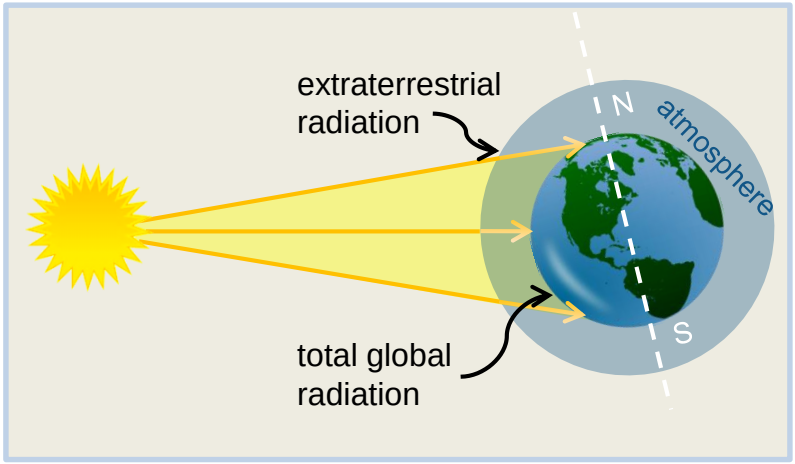
$$R_s = a\,(1 + bH)(1 - e^{-c\,\Delta T^n})$$

Solar Ratio (R_s) = $\dfrac{\text{Global solar radiation}}{\text{Extraterrestrial solar radiation}}$

Daily Temperature Difference (ΔT) = $T_{DailyMax} - T_{DailyMin}$

H is Relative Humidity

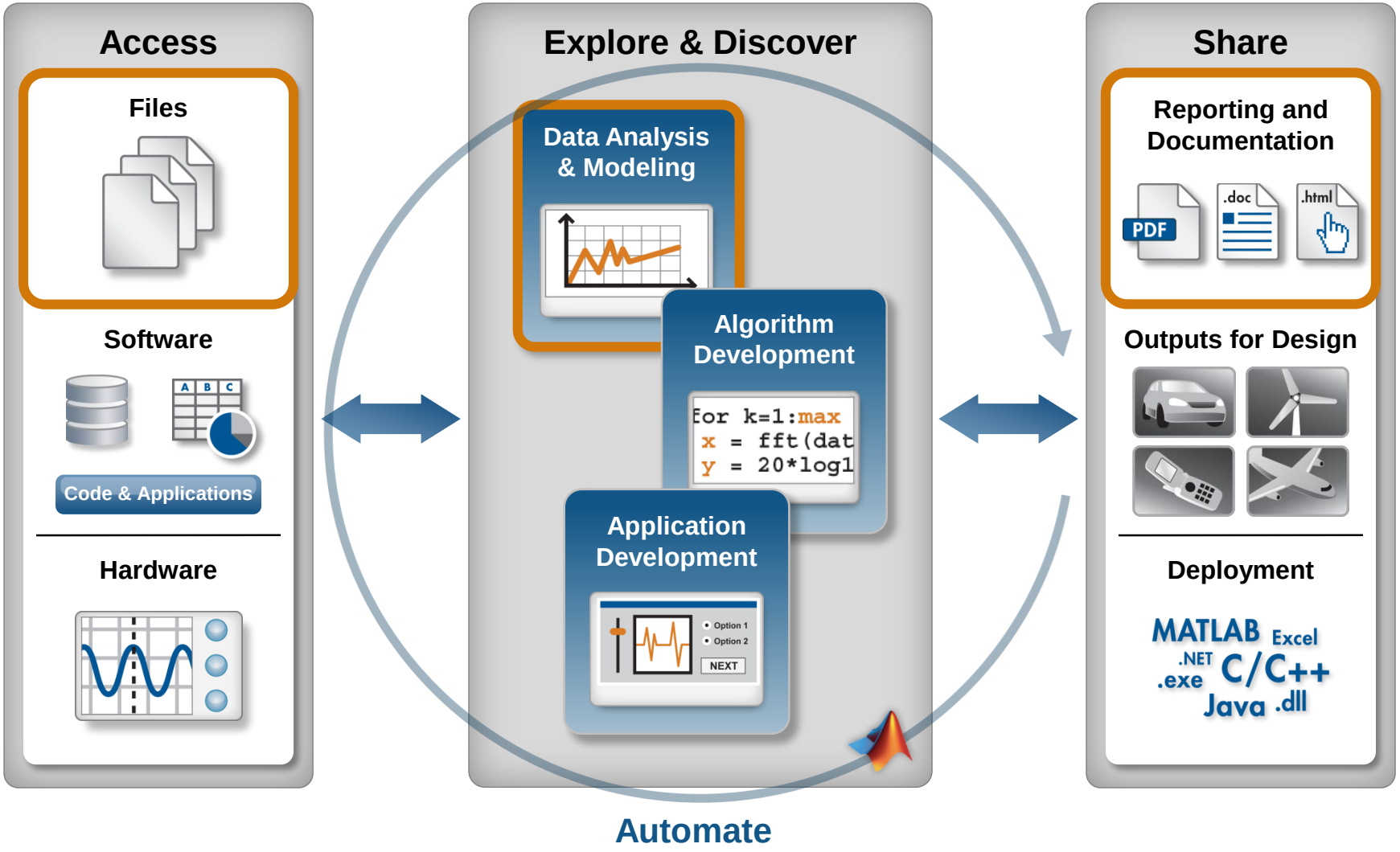
a,b,c,n are the model coefficients



Demo Summary

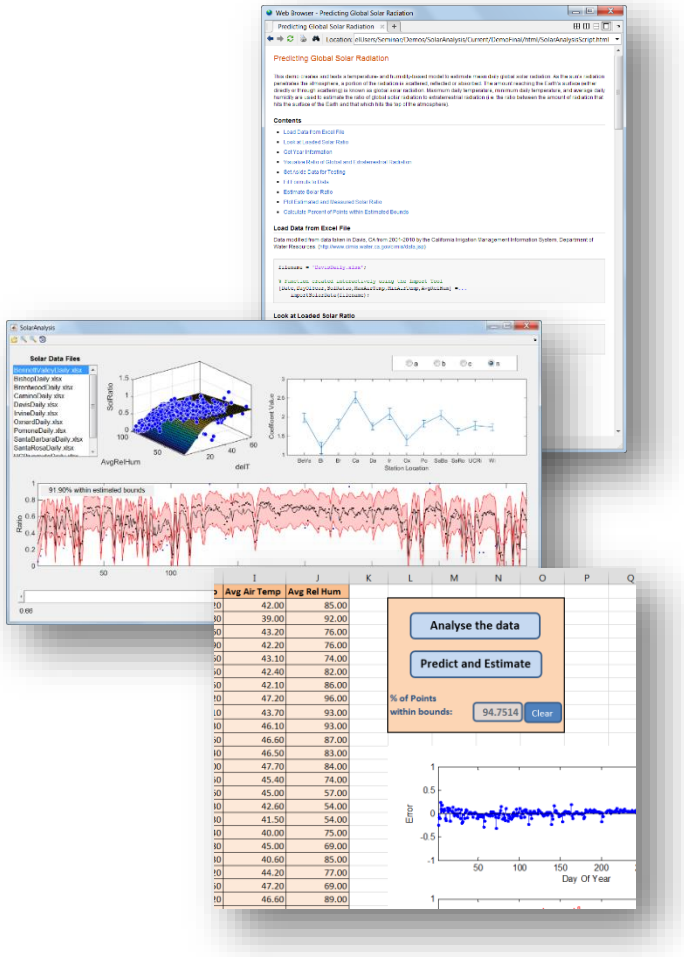
Solar Radiation Estimation

Products Used
- MATLAB
- Curve Fitting Toolbox



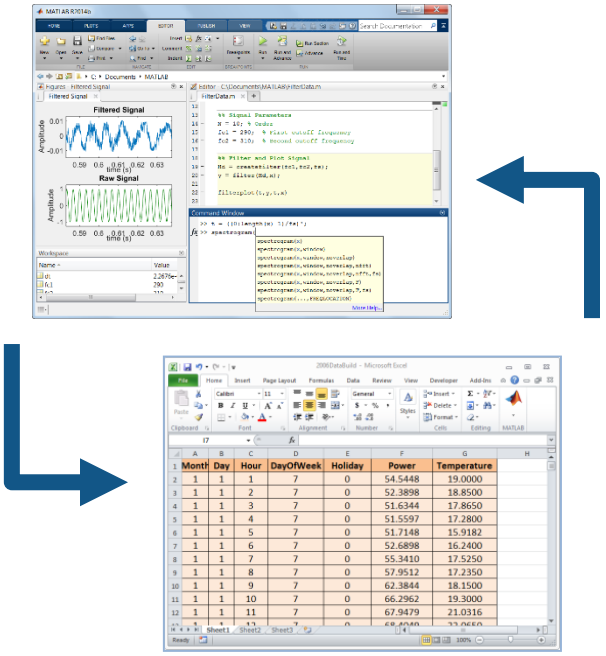
Sharing Results from MATLAB

- Automatically generate reports
- Create and package applications
- Deploy to other environments



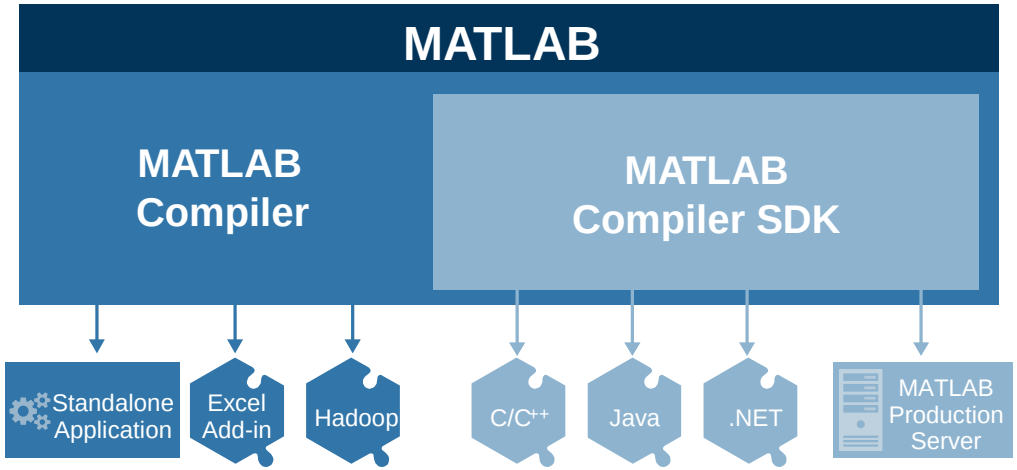
Using MATLAB with Excel

- Passing data between MATLAB and Excel
  - MATLAB
- Accessing MATLAB from an Excel spreadsheet
  - MATLAB
  - Spreadsheet Link EX
- Deploying MATLAB as an Excel add-in
  - MATLAB
  - MATLAB Compiler

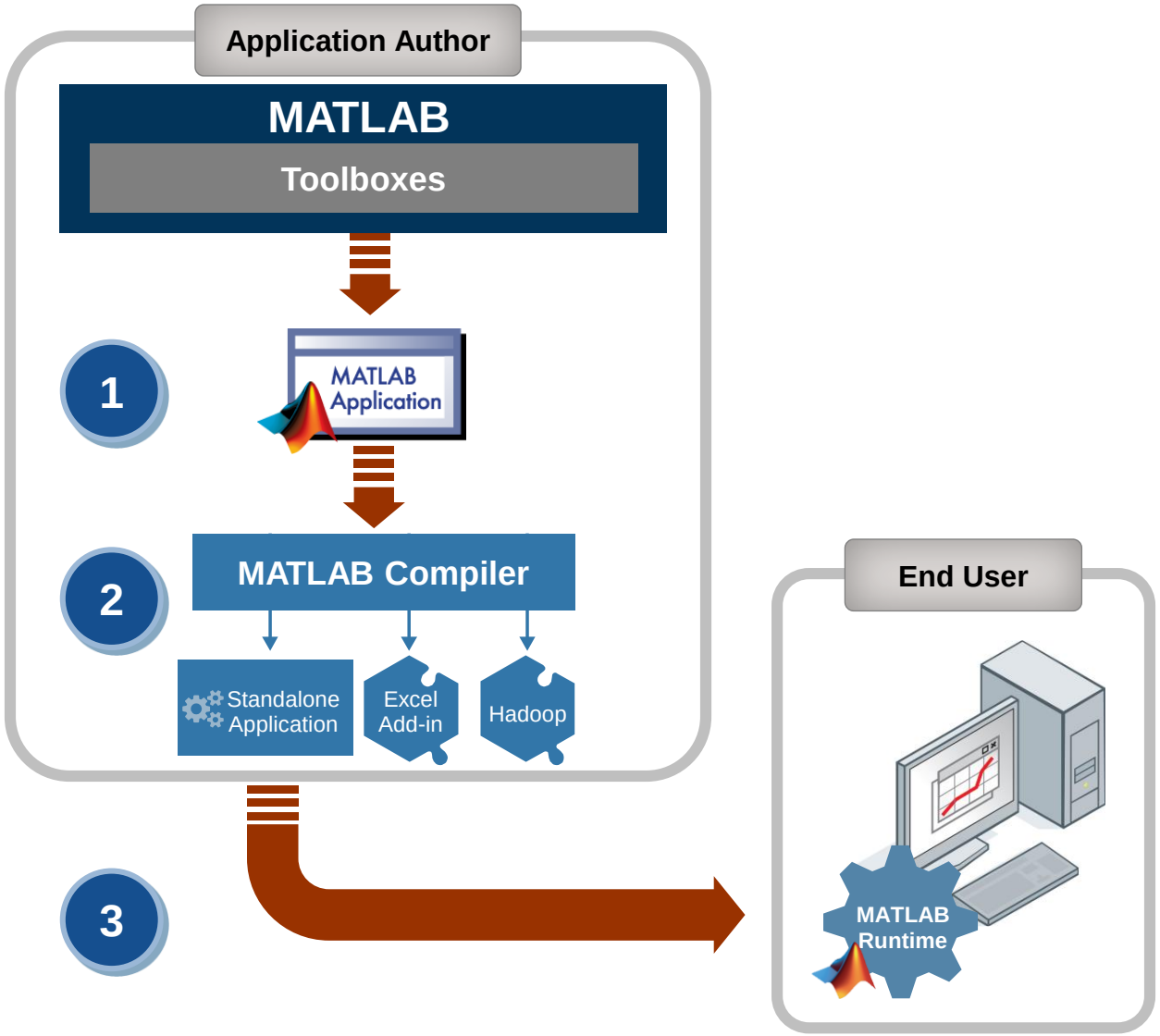


MATLAB Application Deployment

- Share MATLAB programs with people who do not have MATLAB
  - Royalty-free distribution
  - Encryption to protect your intellectual property
- Create both standalone applications and components for integration
- Deploy to desktop, web, and enterprise applications



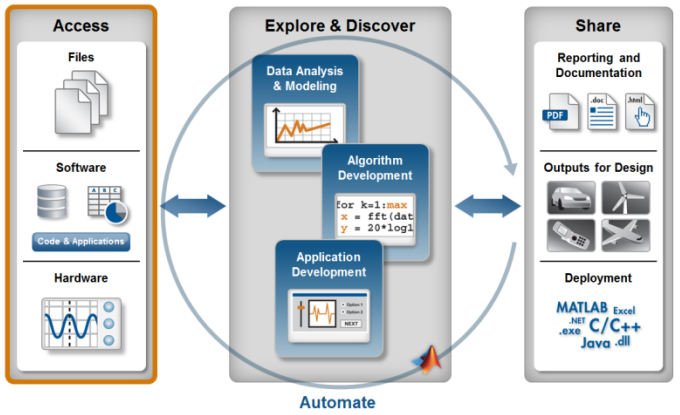
Sharing Standalone Applications



Common Data Analysis Challenges

- Handling complex and messy data

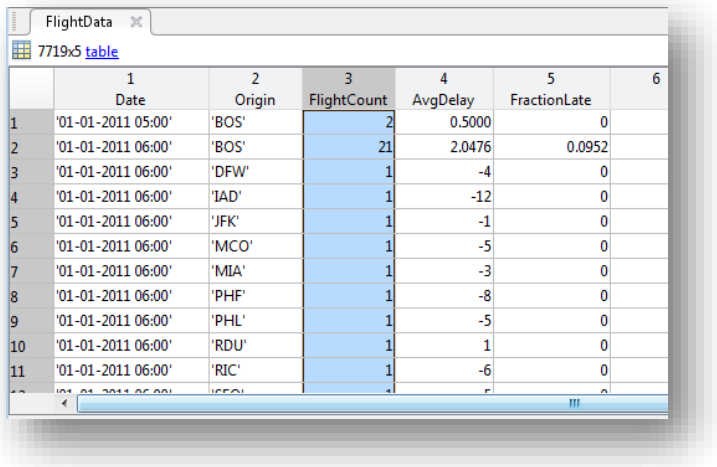
- Modeling with many predictors



Demo: Preparing Late Plane Data

Handling Complex and Messy Data

- Goal:
  - Prepare late plane data for further analysis
- Approach:
  - Load mixed data from files
  - Filter data and replace missing data
  - Merge observations from different time intervals into a single data set


FlightData

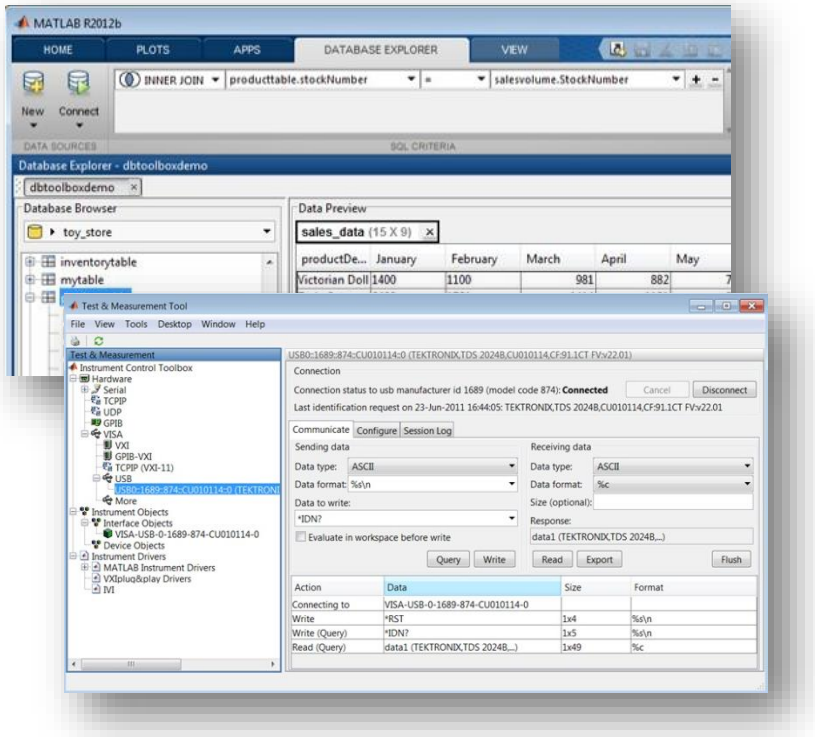
7719x5 table

	1 Date	2 Origin	3 FlightCount	4 AvgDelay	5 FractionLate	6
1	'01-01-2011 05:00'	'BOS'	2	0.5000	0	
2	'01-01-2011 06:00'	'BOS'	21	2.0476	0.0952	
3	'01-01-2011 06:00'	'DFW'	1	-4	0	
4	'01-01-2011 06:00'	'IAD'	1	-12	0	
5	'01-01-2011 06:00'	'JFK'	1	-1	0	
6	'01-01-2011 06:00'	'MCO'	1	-5	0	
7	'01-01-2011 06:00'	'MIA'	1	-3	0	
8	'01-01-2011 06:00'	'PHF'	1	-8	0	
9	'01-01-2011 06:00'	'PHL'	1	-5	0	
10	'01-01-2011 06:00'	'RDU'	1	1	0	
11	'01-01-2011 06:00'	'RIC'	1	-6	0	

Accessing Data from MATLAB

Access | Explore & Discover | Share

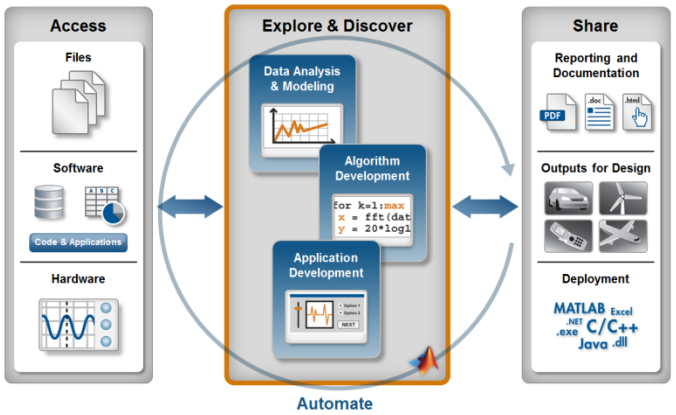
- Files
  - Excel, text, or binary
  - Audio and video, image
  - Scientific formats and XML
- Web Services
  - JSON, CSV, and image data
- Applications and languages
  - C/C++, Java, FORTRAN
  - COM, .NET, shared libraries
  - Databases (*Database Toolbox*)
- Measurement hardware
  - Data acquisition hardware *(Data Acquisition Toolbox)*
  - Stand-alone instruments and devices *(Instrument Control Toolbox)*



Common Data Analysis Challenges

- Handling complex and messy data

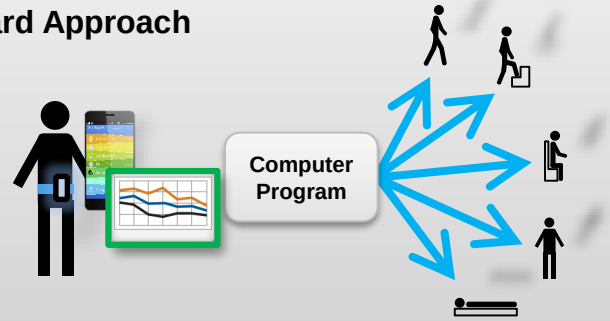
- Modeling with many predictors



Machine Learning

Machine learning uses **data** and produces a **program** to perform a **task**

Standard Approach



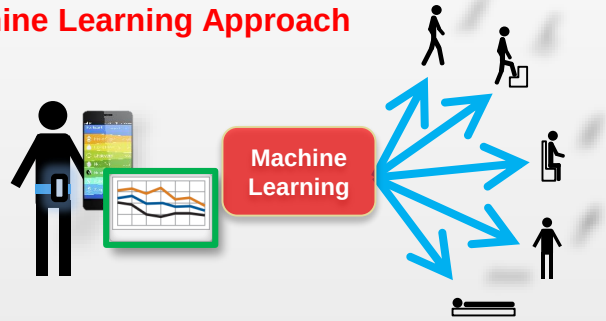
Hand Written Program

If X_acc > 0.5
then "SITTING"
If Y_acc < 4 and Z_acc > 5
then "STANDING"
…

Formula or Equation

$$Y_{activity} = \beta_1 X_{acc} + \beta_2 Y_{acc} + \beta_3 Z_{acc} + \ldots$$

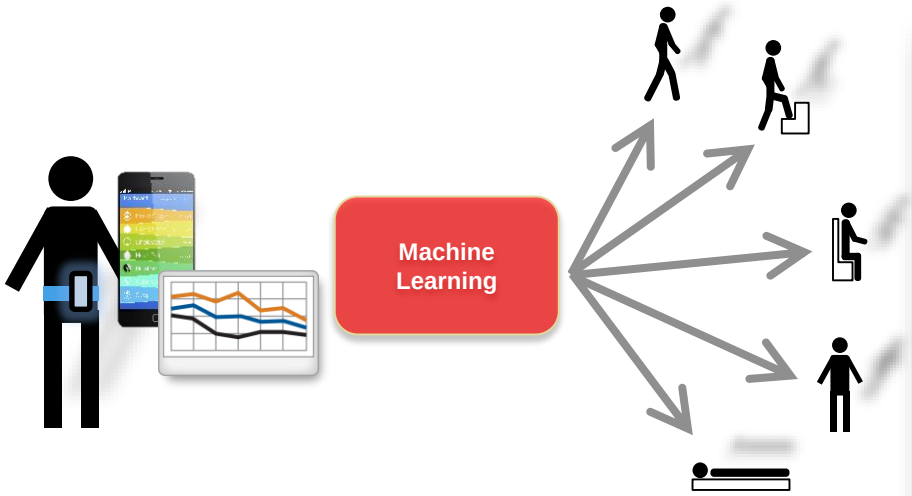
Machine Learning Approach

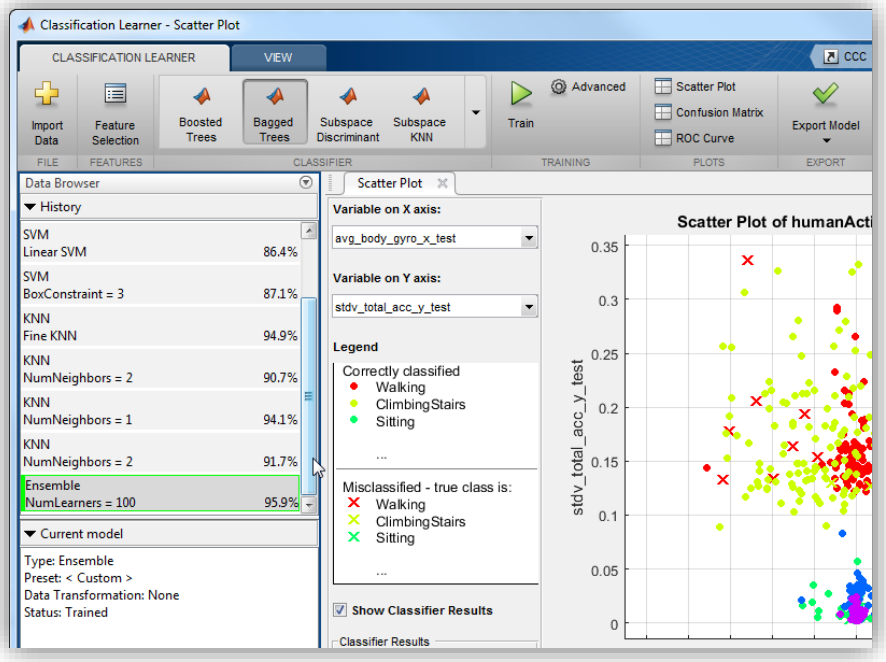


$$model: \text{Inputs} \rightarrow \text{Outputs}$$

$$model = <^{\textbf{Machine Learning Algorithm}}>(sensor_data, activity)$$

Demo: Machine Learning Using Mobile Phone Data





Data:

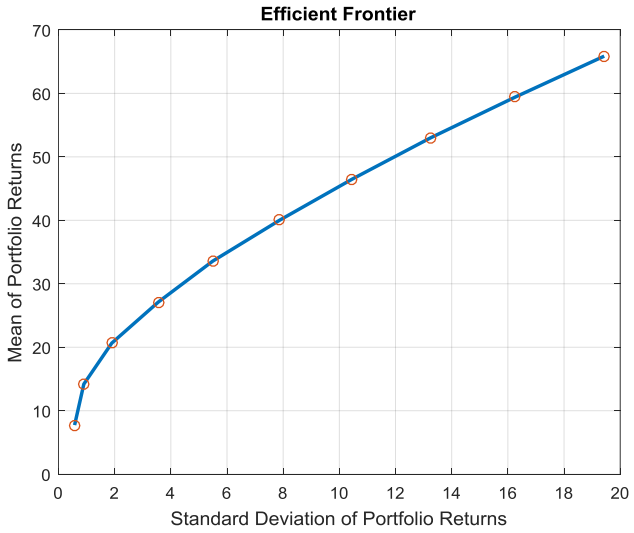
- 3-axial Accelerometer data
- 3-axial Gyroscope data

Demo: Portfolio Optimisation

Computing the Efficient Frontier

- Goal: Compute an Efficient Frontier in:
  - Excel Solver
  - MATLAB

- Compare the 2 approaches in:
  - Performance
  - Automation of Workflow



Workflow Portfolio Optimization

- Convert prices to returns.
- Expected Returns.
- Covariance matrices
- Calculate Efficient Frontier
  - Optimize to Maximise the return
  - Optimize to Minimise the risk
  - Optimze multiple times between Min Risk & Max Return

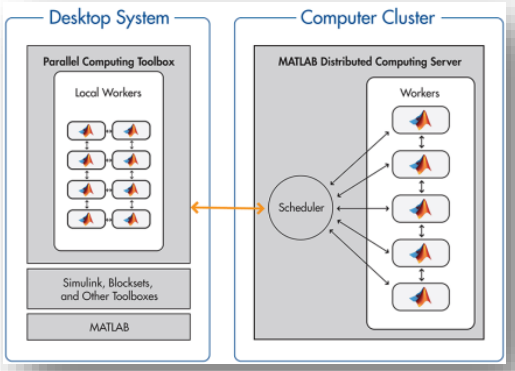
MATLAB Central

- Community for MATLAB and Simulink users
  - Over 70k daily visits
- File Exchange
  - Access more than 10k free files including functions, apps, examples, and models
- MATLAB Answers
  - Ask programming questions or search
  - 18k+ community-answered Questions
- Blogs
  - Read commentary from engineers who design, build, and support MATLAB and Simulink

Expand Your Analysis Capabilities

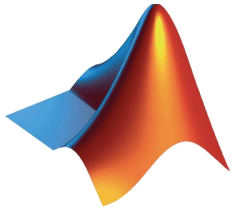
- Machine learning
  (Statistics and Machine Learning Toolbox, Neural Networks Toolbox)
  - "Learn" from your data without assuming an equation as a model
  - www.mathworks.com/machine-learning

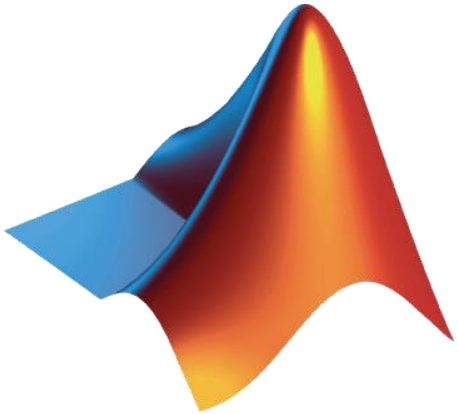
- Parallel programming
  (Parallel Computing Toolbox)
  - Speed up your analysis using multicore computers, GPUs, and computer clusters
  - http://www.mathworks.com/products/parallel-computing/



Today's Objectives

- Introduce you to data analysis with MATLAB
- Show how you can overcome common data analysis challenges with MATLAB
- Demonstrate multiple ways of sharing your analysis and results with others





© 2016 The MathWorks, Inc.