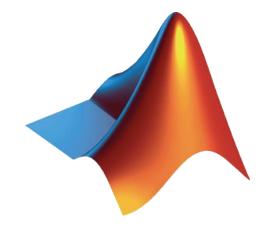


Beyond Excel: Enhancing Your Data Analysis with MATLAB

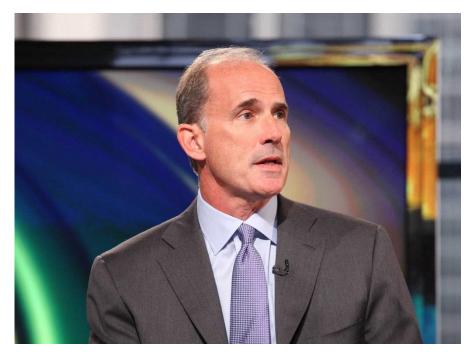
David Willingham Senior Application Engineer – Data Analytics



© 2016 The MathWorks, Inc.



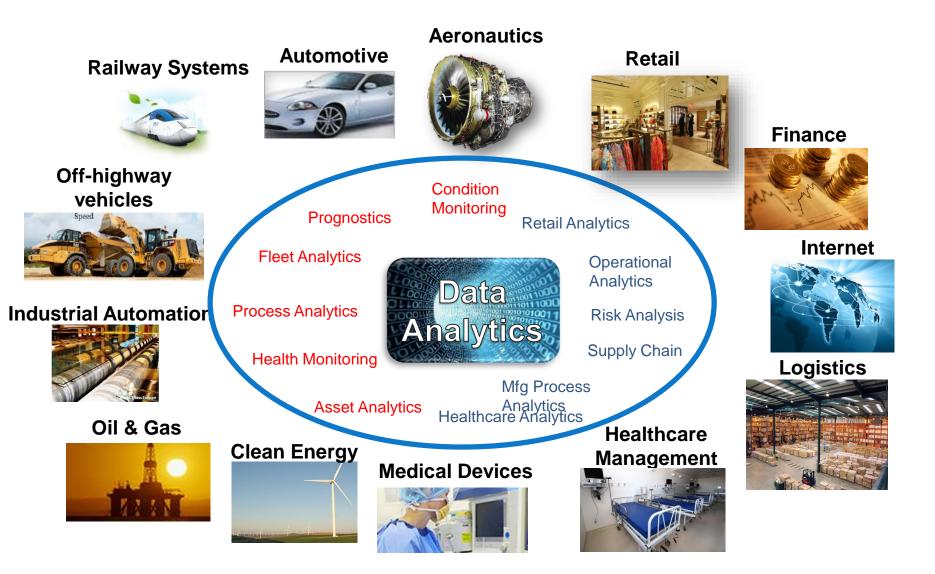
"Data is the sword of the 21st century, those who wield it the samurai."



Google's Former SVP - Jonathan Rosenberg

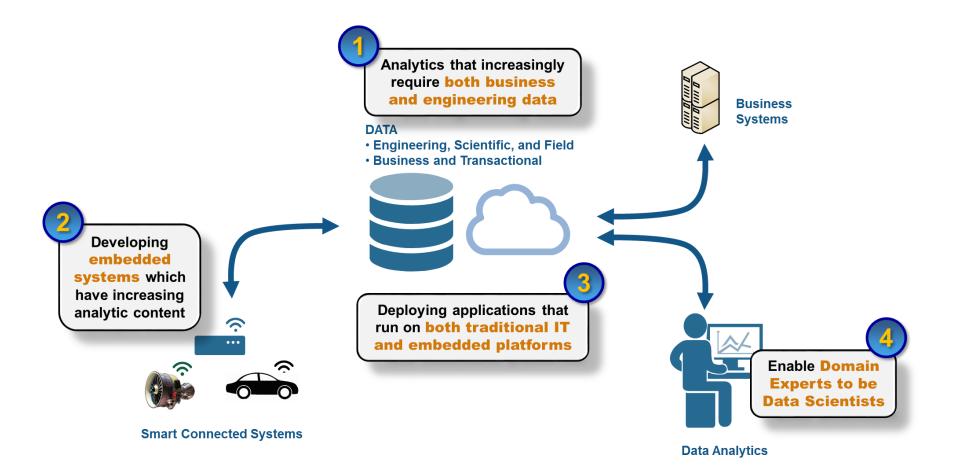
- Big data how to create it, manipulate it, and put it to good use.
- "If you want to work at Google, make sure you can use MATLAB."







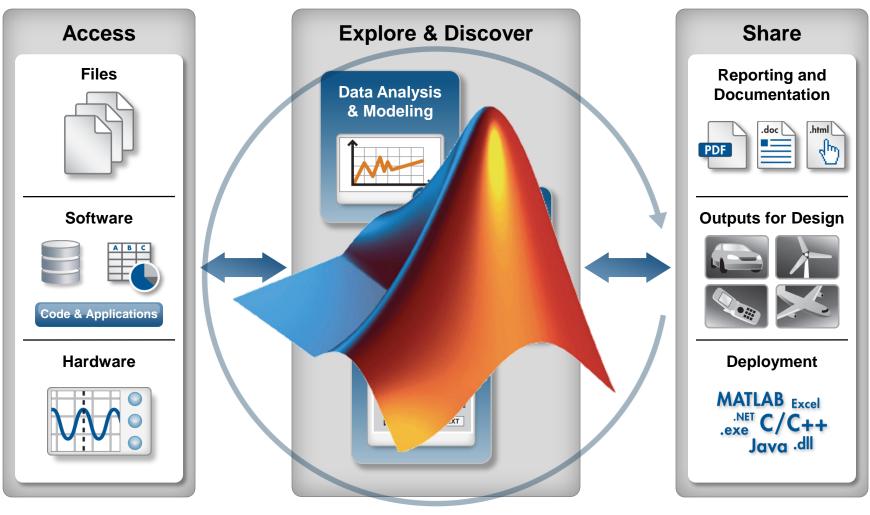
What Analytics Challenges are you facing?



4



Data Analytics Workflow



Automate



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



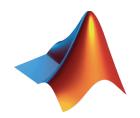
"MATLAB and MATLAB Compiler SDK enabled us to rapidly deliver a sophisticated portfolio analytics web application with confidence that it will return accurate results extremely quickly, ensuring a highly usable and stable platform for our clients."

> Lee Eriera Frontier Advisors



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

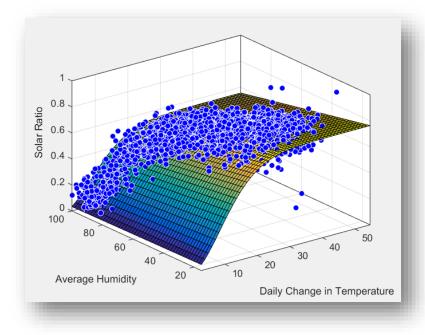
- Complex calculations
- Messy Data
- Speed of Execution
- Automation
- Batch Processing
- Report Generation
- Deployment



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

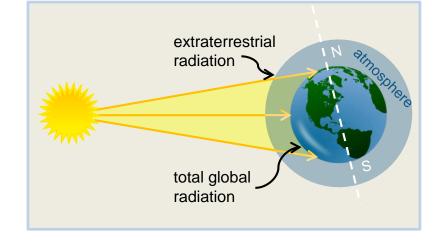




Modeling Global Solar Radiation

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

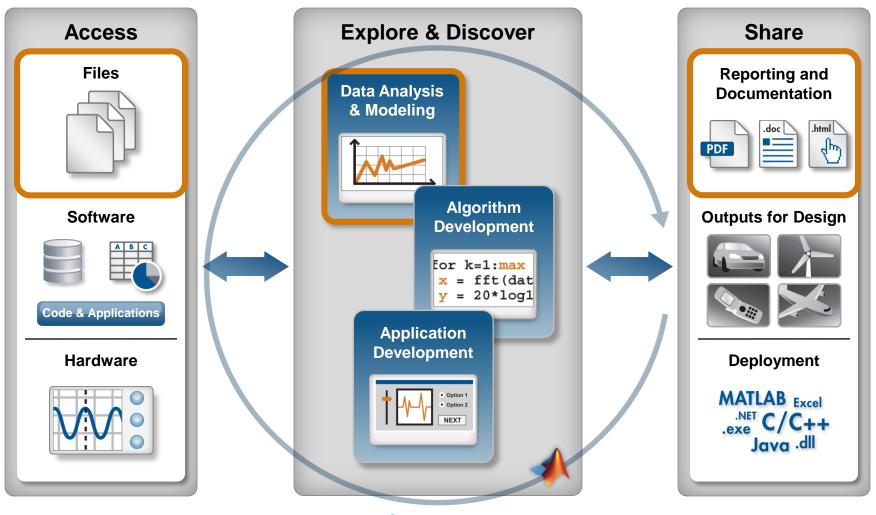
Solar Ratio (\mathbf{R}_{s}) = $\frac{\text{Global solar radiation}}{\text{Extraterrestrial solar radiation}}$ Daily Temperature Difference ($\Delta \mathbf{T}$) = $T_{\text{DailyMax}} - T_{\text{DailyMin}}$ \mathbf{H} is Relative Humidity $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{n}$ are the model coefficients





Demo Summary Solar Radiation Estimation

- **Products Used**
- MATLAB
- Curve Fitting Toolbox



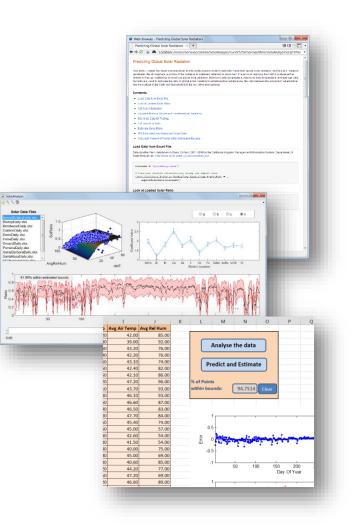


Sharing Results from MATLAB

Automatically generate reports

Create and package applications

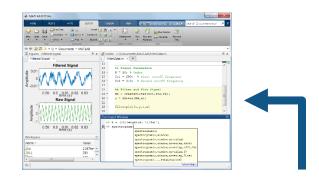
Deploy to other environments





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

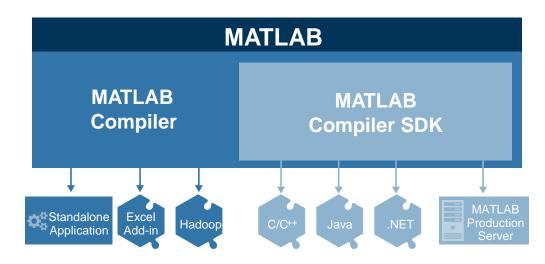






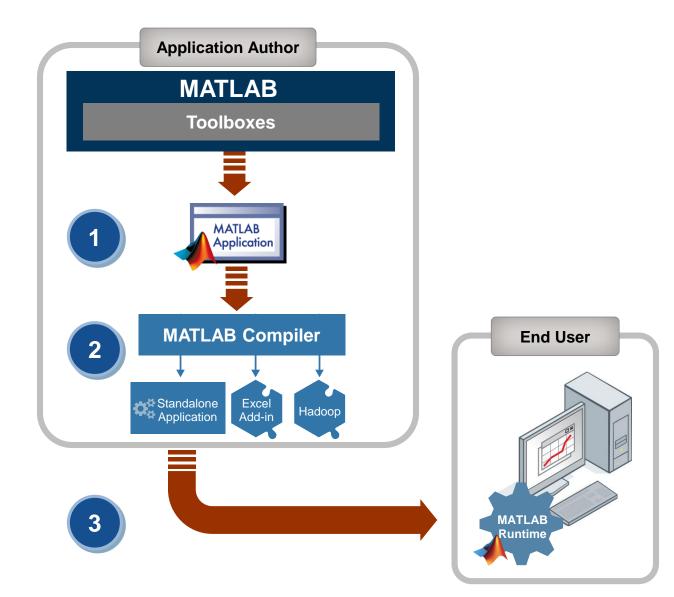
MATLAB Application Deployment

- 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

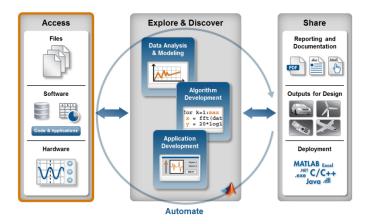




Common Data Analysis Challenges

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

	7719x5 <u>table</u>					
	1	2	3	4	5	6
	Date	Origin	FlightCount	AvgDelay	FractionLate	
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	
	101 01 2011 05 001	ICCO!	i i	r .		



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)

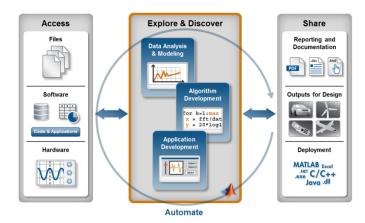
IOME PLOTS APPS	DATABA	ASE EXPLORE	R VI	EW			1 40	10 1	2
Connect	icttable.stockNumbe	• •	▼ sale	svolume.	StockN	umber	•	1±.:	
SOURCES		SQL CRI	TERIA						-
ase Explorer - dbtoolboxdemo									- 1
oolboxdemo ×									
base Browser	Data Preview	i							-88
toy_store	▼ sales_data	(15 X 9) >	4						
inventorytable	productDe	January	February	March	() 	April		May	
E mytable	Victorian Dol	1 1400	1100		981		882		7
	USB0::1689::874::CL Connection Connection status	s to usb manufac	RONIX,TDS 2024B,C turer id 1689 (mode un-2011 16:44:05: Ti	l code 874)	Connect	ed 🗌	Cancel CF:91.1CT	Dis	connect
	US80:1689:874:CU Connection Connection status Last identification Communicate <i>Qu</i> Sending data	s to usb manufac n request on 23-J onfigure Session	turer id 1689 (mode un-2011 16:44:05: Ti 1 Log	l code 874) KTRONIX,T Receivi	Connect DS 20248	ed 8,CU010114,		Dis	connect D1
	USB0:1689:874-CL Connection Connection status Last identification Communicate Co Sending data Data type: ASM	s to usb manufac n request on 23-J onfigure Session CII	turer id 1689 (mode un-2011 16:44:05: Ti n Log	I code 874) KTRONIX,T Receivi Data ty	Connect DS 20248 ng data pe:	eed		Dis	connect D1
	USB0-1689-874-CL Connection Connection status Last identification Communicate (cc Sending data Data type: ASt Data format: SSA	s to usb manufac n request on 23-J onfigure Session CII	turer id 1689 (mode un-2011 16:44:05: Ti 1 Log	I code 874) KTRONIX,T Receivi Data ty Data fo	Connect DS 2024E ing data ipe:	ed 8,CU010114,		Dis	connect D1
	USB0-1689-874-CU Connection Connection status Last identification Communicate (co Sending data Data type: ASS	s to usb manufac n request on 23-J onfigure Session CII	turer id 1689 (mode un-2011 16:44:05: Ti n Log	Receivi Data to Size (o	Connect DS 20248 ng data pe: prmat: ptional):	eed		Dis	connect D1
	USB0-1689-874-CC Connection Connection status Last identification Communicate CC Sending data Data type: ASI Data format: %Gy Data to write: "IDN?	s to usb manufac n request on 23-J onfigure Session CII n	turer id 1689 (mode un-2011 16:44:05: Ti 1 Log	Receivi Data ty Size (o) Respon	connect DS 20248 ng data pe: prional: ise:	eed	.CF:91.1CT	Dis	connect D1
	USB0-1689-874-CL Connection Connection status Last identification Communicate Cc Sending data Data type: ASS Data format %A) Data to write: "IDN?	s to usb manufac n request on 23-J onfigure Session CII n	turer id 1689 (mode un-2011 16:44:05: Ti 1 Log	Receivi Data ty Size (o) Respon	Connect DS 20248 Ing data pe:	sed 3,CU010114, ASCII %c	.CF:91.1CT	Dis	connect D1
	USB0-1689-874-CL Connection Connection status Last identification Communicate Cc Sending data Data type: ASS Data format %A) Data to write: "IDN?	s to usb manufac n request on 23-J onfigure Session CII n	turer id 1689 (mode un-2011 16:44:05: Ti x Log write	Code 874) CKTRONDK,T Data ty Data ty Data fc Size (o Respon data1	Connect DS 20248 Ing data pe:	xed ASCII ASCII %c VD,TDS 202 port	.CF:91.1CT	Dis	connect D1
	USB01689-874-CC Connection Connection statu Last identification Communicate of Sending data Data type: & Sa Data townte: %00 Data to write: "UNP Evaluate in wo Action Connecting to	to usb manufac a request on 23-J onfigure Session CII on onkspace before Ubata VISA-USB-C	turer id 1689 (mode un-2011 16:44:05: Ti x Log write	I code 874) KTRONIX,T Data ty Data fc Size (o Respondata1 Rear	connect DS 20248 ng data pe: prional): ise: (TEKTROI d) Ex Size	ASCII %c ND, TDS 202 port Fo	CF:91.1CT ?48,) ormat	Dis	connect D1
	USB0-1689-874-CL Connection Connection fattur Last identificatur Communicate (Sending data Data format: %a) Data to write: 1DN? Evaluate in wc Action Connecting to Write	s to usb manufac request on 23-J onfigure Session CII In Data VISA-USB-C *RST	turer id 1689 (mode un-2011 1644-05: Ti 1 Log write Query Write	I code 874) KTRONIX,T Data ty Data fc Size (o Respondata1 Rear	connect DS 20248 ng data pe: ptional): ise: (TEKTRO) d Ex Size 1x4	eed a,CU010114, ASCII %c vD,TDS 202 port Fo %s	(CF:91.1CT 248,) prmat	Dis	connect D1
	USB01689-874-CC Connection Connection statu Last identification Communicate of Sending data Data type: & Sa Data townte: %00 Data to write: "UNP Evaluate in wo Action Connecting to	i to usb manufac request on 23-J onfigure Session CII n Dritspace before [USA-USB-C *RST +IDN7	turer id 1689 (mode un-2011 1644-05: Ti 1 Log write Query Write	Action of the second se	connect DS 20248 ng data pe: prional): ise: (TEKTROI d) Ex Size	ASCII %c ND, TDS 202 port Fo	(CF:91.1CT 24B,) ormat \n	Dis	connect D1



Common Data Analysis Challenges

Handling complex and messy data

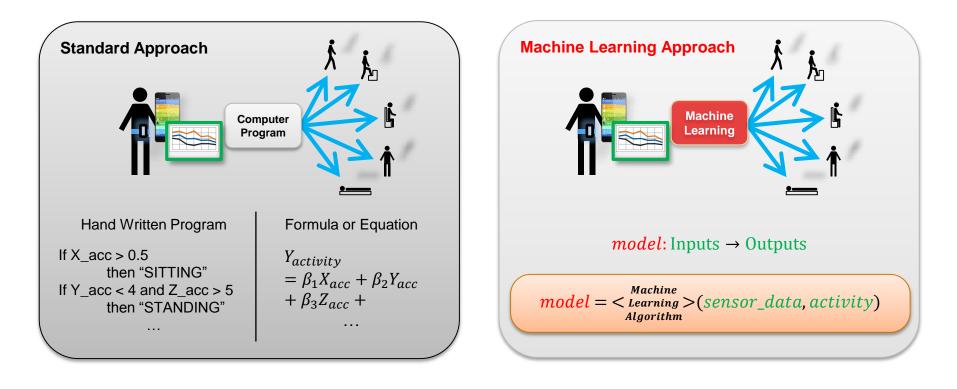
Modeling with many predictors





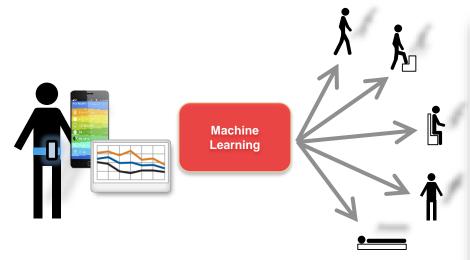
Machine Learning

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





Demo: Machine Learning Using Mobile Phone Data



Data:

3-axial Accelerometer data
3-axial Gyroscope data

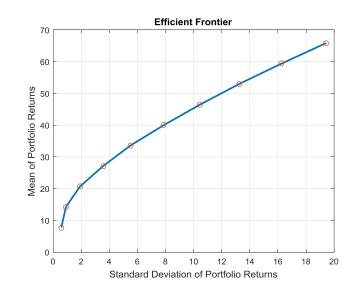
CLASSIFICATION LEAF	RNER	VIEW					20 🖪
Import Feature Data Selection	A Boosted Trees	A Bagged Trees	Subspace Subspace Discriminant KNN	D Train	Advanced	Confusion Matrix	Export Model
			SIFIER	1	RAINING	PLOTS	EXPORT
Data Browser			Scatter Plot X				
✓ History			Variable on X axis:			Scatter Plot	of humanA
SVM Linear SVM		86.4%	avg_body_gyro_x_test	•	0.35	×	
SVM BoxConstraint = 3		87.1%	Variable on Y axis:				• •
KNN		87.1%	stdv_total_acc_y_test	•	0.3		
Fine KNN		94.9%					
KNN NumNeighbors = 2		90.7%	Legend Correctly classified		0.25 test		
KNN NumNeighbors = 1		≡ 94.1%	 Walking ClimbingStairs Sitting 		acc / test	×	×
KNN NumNeighbors = 2		91.7%		_			× x
Ensemble NumLearners = 100		95.9% -	Misclassified - true class is X Walking	:	0.15	×	
 Current model 			ClimbingStairs		0.1	•	•
Type: Ensemble Preset: < Custom >				_	0.05		
Data Transformation: Nor Status: Trained	ne		Show Classifier Results				



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

File Exchange	Answers	Newsgroup	Link Exchange	Blogs	Trend	y Cody	Contes	
Trendy	,	File Exche				ATLAB Ansv		
Engaging Expansive Addictive		Sourang.	Modulation Technique su Banerji htzee Zachary Danzi	ger		how to custo Use of timer: that its period	mize a colo what's the of time is	
Cody Join the f	fun	Create Video of Rotating 3D Plot Alan Jennings Four parameters logistic regression - There and back again Gluesppe Cardillo Direct strength method calculator for columns at ambient temperature Ashkan				 CFA Green Channel Bill Median and Gradient be algorithms suada <u>Matab/simulink - Is ther</u> only two states <i>Robin</i>. How to transfer the siz 		
			on of effective squas nannel steel Ashkan	h load for	۵	image to the e MATLAB 7.5 operation Ma	doesn't cal	
MathWo	orks	Cody			Tr	endy		
Careers		Recent P	roblems		Р	opular Plot	5	
Tell us what Looking for u provide feed	isers to back on	Claudio Claudi	of lattice points within Gelmi ed fractions Claudio e name from the lege lle Michael C.	Jelmi	, mark		Briti Nev Hug	



Expand Your Analysis Capabilities

Machine learning (Statistics and Machine Learning Toolbox,

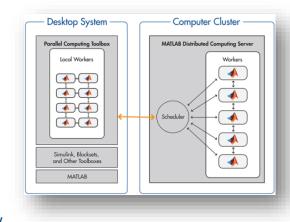
Neural Networks Toolbox)

 "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





