MATLAB을 활용한 금융 데이터 분석 및 모델링

May. 19th, 2022

장규환
Application Engineer @ MathWorks
pauljang@mathworks.com

MathWorks Finance CME 2022

MathWorks®

Agenda

- 다양한 금융 데이터 활용 사례
  - 금융 회사 사례
  - 금융 데이터 연구사례
- 금융 데이터 과학
  - 다양한 금융 데이터 소개
  - MATLAB을 활용한 금융 데이터 과학 기법
  - 금융 데이터 전처리 및 시각화 기능 소개
- 금융 데이터를 활용한 모델링 기법 소개
  - 금융 데이터를 활용한 AI 모델링 기법 소개
  - AI 모델 공유 소개

Calibrating the Rating System

- Overseeing a portfolio of bonds
- Improve rating engine using machine learning

	B	C	D	E	F	G
1	WC_TA	RE_TA	EBIT_TA	MVE_BVTD	S_TA	Industry
2	0.049	0.220	0.041	2.400	0.489	6
3	0.145	0.240	0.051	1.514	0.280	4
4	0.170	0.177	0.056	1.685	0.455	7
5	0.164	0.289	0.064	1.029	0.222	3
6	0.078	0.066	0.044	0.417	0.147	9
7	0.160	0.371	0.060	2.653	0.445	9
8	0.224	0.384	0.065	1.789	0.283	6
9	0.273	0.213	0.062	0.426	0.159	2
10	0.015	-0.138	0.036	0.108	0.084	6
11	0.006	0.175	0.031	3.279	0.686	8
12	0.565	0.755	0.115	5.264	0.545	7
13	0.030	-0.215	0.042	0.146	0.090	2

H
Ratings
AA
BBB
A
BBB
BB
AA
A
BB
BB
AA
AAA
CCC

MATLAB R2021b

C: ▸ Users ▸ pauljang ▸ OneDrive - MathWorks ▸ Desktop ▸ Finance CME_1207 ▸ demo16 ▸ simple-credit-classification-master

라이브 편집기 - C:₩Users₩pauljang₩OneDrive - MathWorks₩Desktop₩Finance CME_1207₩demo16₩simple-credit-classification-master₩Script1_CreditRating_...

Script1_CreditRating_Classification.mlx | Script2_ProbabilityOfDefault_LogisticRegression.mlx | Script3_DefaultRatesForecasts.mlx

Credit Classification

1) Classify based on the treebagger classification function 2) Validate model using the confusionmat validation function 3) Use the predict function to estimate

Load historical data

```
load CreditData
RatingCategories = {'AAA' 'AA' 'A' 'BBB' 'BB' 'B' 'CCC'};
```

Setup Data

```
FinRatios = [Data.RE_TA, ...
             Data.MVE_BVTD, ...
             Data.Industry];

Ratings    = categorical(Data.Rating,RatingCategories,'Ordinal',true);
RatingsNum = double(Ratings);
```

명령 창

```
>> doc fitlm
>>
```

작업 공간

이름

- Data
- FinRatios
- RatingCategories
- Ratings
- RatingsNum

Zoom: 100% | UTF-8 | LF | 스크립트 | 라인 32 | 열 21


80%
20%

"Looks like you've got all the data – what's the holdup?"

Agenda

- 다양한 금융 데이터 활용 사례
  - 금융사의 데이터 활용 사례
  - 금융 데이터 연구사례
- 금융 데이터 과학
  - 다양한 금융 데이터 소개
  - MATLAB을 활용한 금융 데이터 과학 기법
  - 금융 데이터 전처리 및 시각화 기능 소개
- 금융 데이터를 활용한 모델링 기법 소개
  - 금융 데이터를 활용한 AI 모델링 기법 소개
  - AI 모델 공유 소개

HSBC Adopts MATLAB For Model Risk Management

Challenge

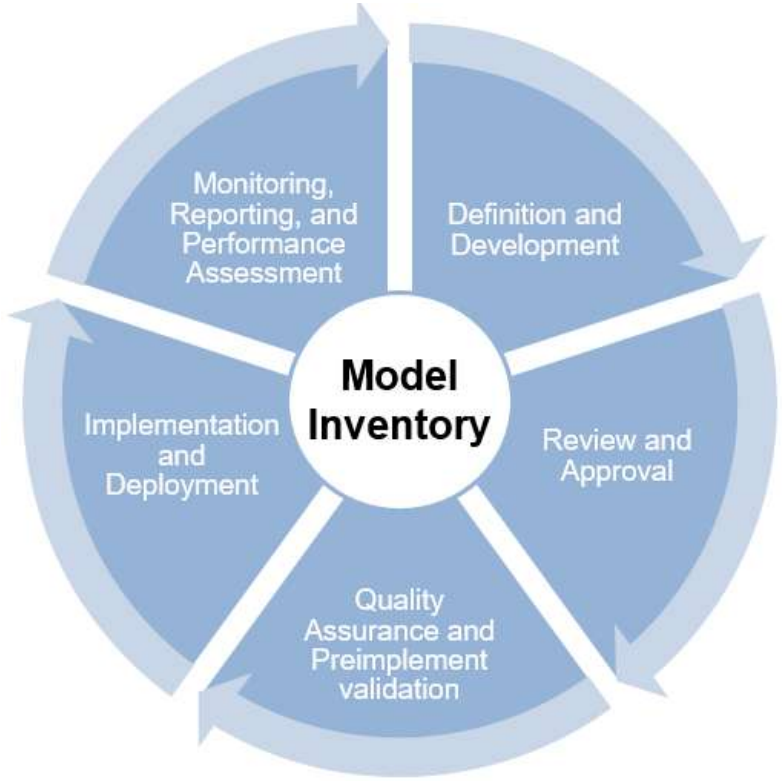
Improve the pace, transparency, and reproducibility of the model development and review processes

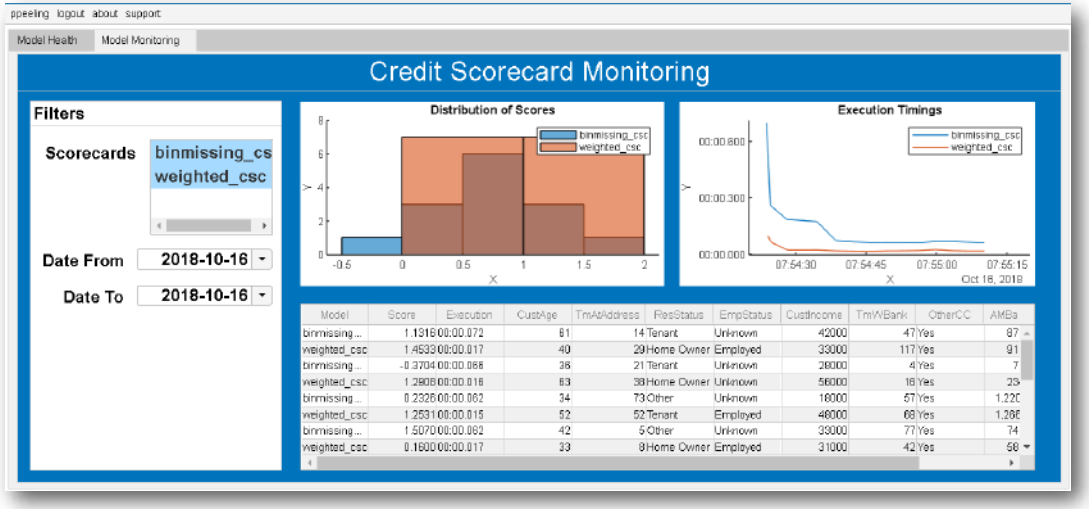
Solution

Use of MATLAB toolboxes for risk modeling and automatic template-driven reports. Production Server used to execute models and dashboards to monitor them.

Results

- Complete model lineage achieved
- All stages of the model workflow are supported
- Models automatically document as-you-build
- Template driven report generation
- Web-deployed model review environment
- MATLAB dashboards easily created and deployed to the web to monitor model performance





Scaling Financial Research: JP Morgan

Challenge

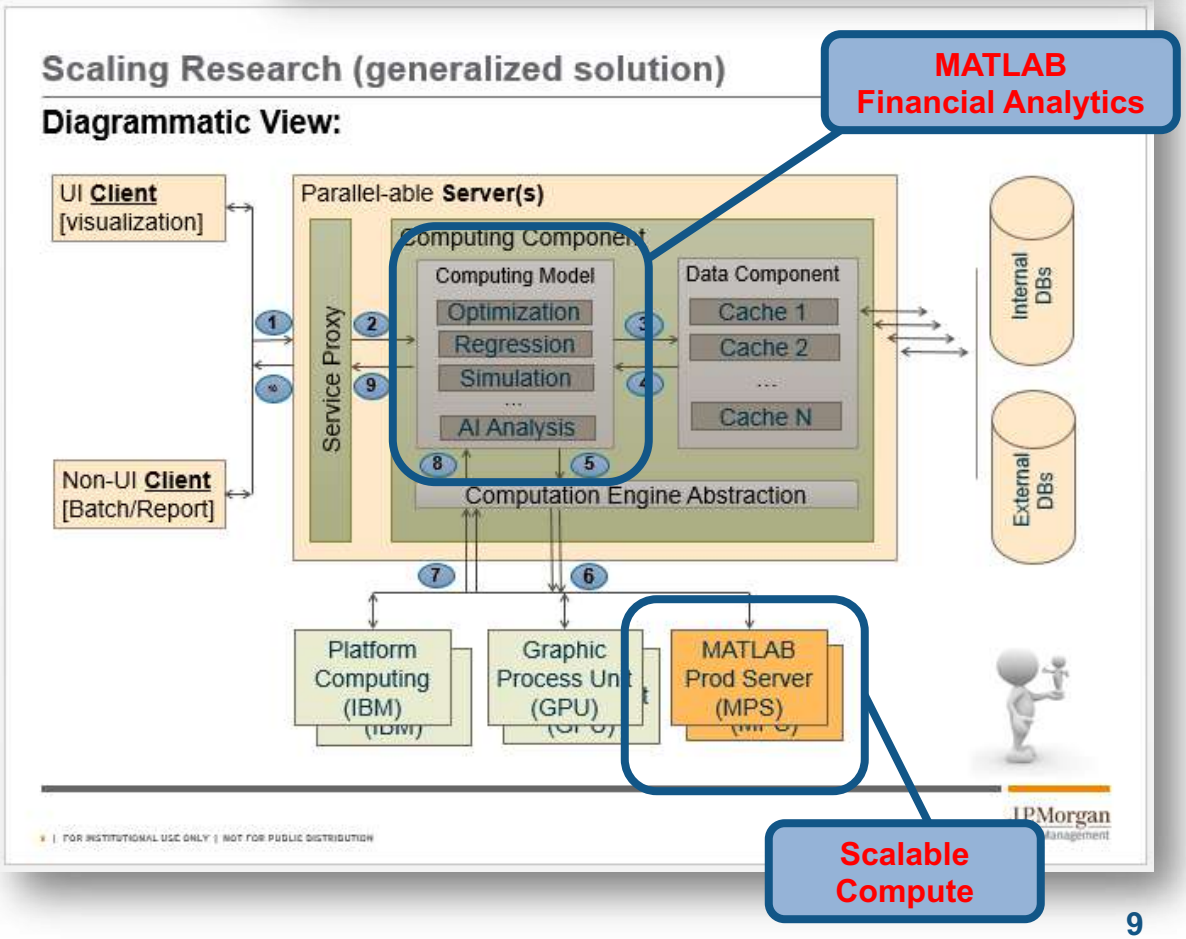
- Quantitative research models can take a day to run
- Buy-side research team need real-time modeling to transition from “Research Activity” to scalable “Production activity”

Solution

- Transition from Quantitative research (Desktop) to Production (Server)
- Advanced visualization, scalable compute, integrated with enterprise data platform
- Financial models supported: Alpha, risk, transaction cost, portfolio construction, scenario analysis, stress-tests

Benefit

- Researchers can run models in minutes, not hours
- Professionally supported scalable solution for IT



NIBC

- Started discussions on the benefit of Model Risk Management.
- NIBC is using Deep Learning Toolbox from MATLAB

New technologies

Simplification

NIBC's annual report 2019

Net interest income

Net interest income remained relatively stable with a 1% increase to EUR 209 million in H1 2019 compared to EUR 207 million in H1 2018. Net interest income continues to be affected by the adoption of IFRS 9 in combination with the revision of NIBC's hedging framework. In 2017 and 2018 the old hedging portfolio (with relatively high interest rates) was unwound and replaced with new swaps at current (low)interest rates. As a result, the gain until maturity is recognised in net interest income, due to the lower coupon that NIBC pays on the hedging derivatives. The effect on net interest income in H1 2019 is estimated at EUR 19 million (H1 2018: EUR 28 million).

The effects of our rebalancing strategy for the corporate loan portfolio are reflected in the 2019 year-to-date origination spread, which decreased from 2.99% in 2018 to 2.54% in H1 2019. This is partially explained by the fact that lower and more granular risk, and therefore lower yield, loans are originated, resulting in a transition to a lower risk profile of the portfolio, and consequently lower

NIBC worked with MATLAB consultant in 2019 for IFRS9 compliance.

Development of GUI's / Applications to enable non-MATLAB users to access MATLAB based codes and use it in parallel in their daily work.

Process Optimization

Optimization of processes started with MATLAB Production Server in 04/2019. Separation of Development and Production. Here the initial request from NIBC to support with the set-up.

Automation

Risk Management runs a MATLAB oriented platform (DTAP) for algorithm / model and application development that facilitates the deployment of models. MATLAB codes are linked to data dictionaries (necessary if data is changing).

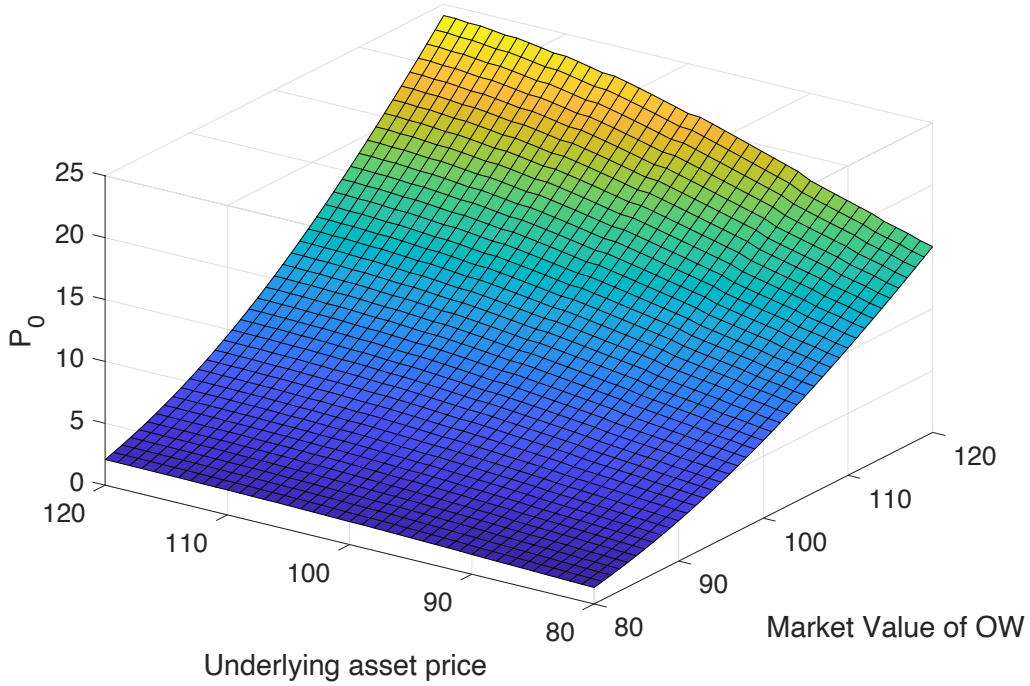
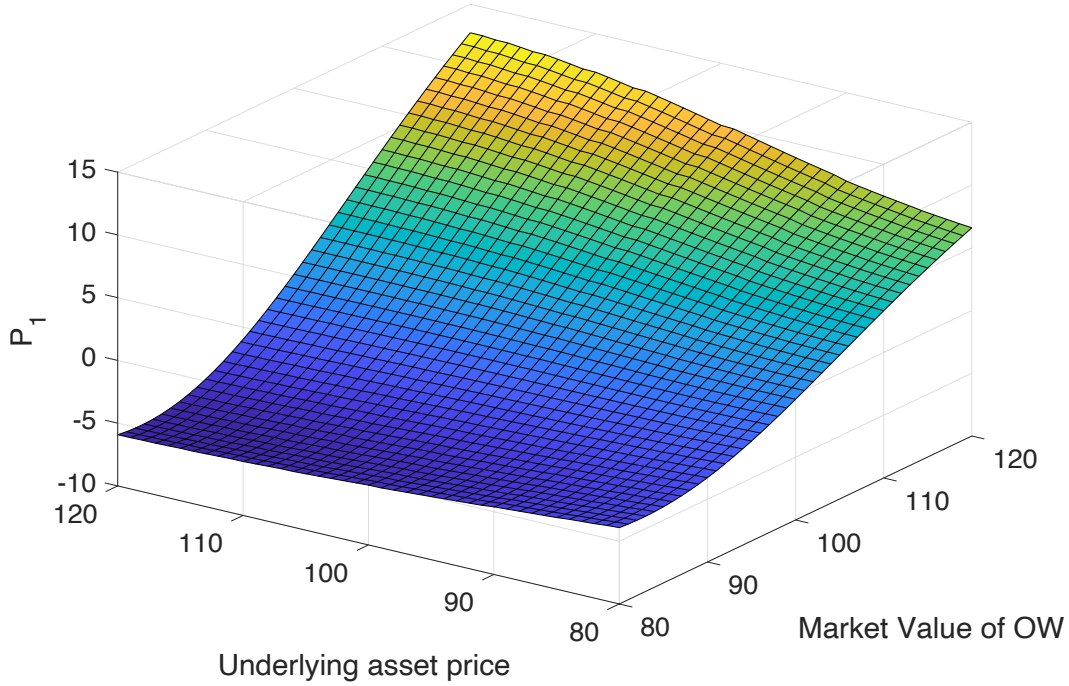
Agenda

- 다양한 금융 데이터 활용 사례
  - 금융사의 데이터 활용 사례
  - 금융 데이터 연구사례
- 금융 데이터 과학
  - 다양한 금융 데이터 소개
  - MATLAB을 활용한 금융 데이터 과학 기법
  - 금융 데이터 전처리 및 시각화 기능 소개
- 금융 데이터를 활용한 모델링 기법 소개
  - 금융 데이터를 활용한 AI 모델링 기법 소개
  - AI 모델 공유 소개

Pusan Univ Research: Option Pricing Numerical Experiments

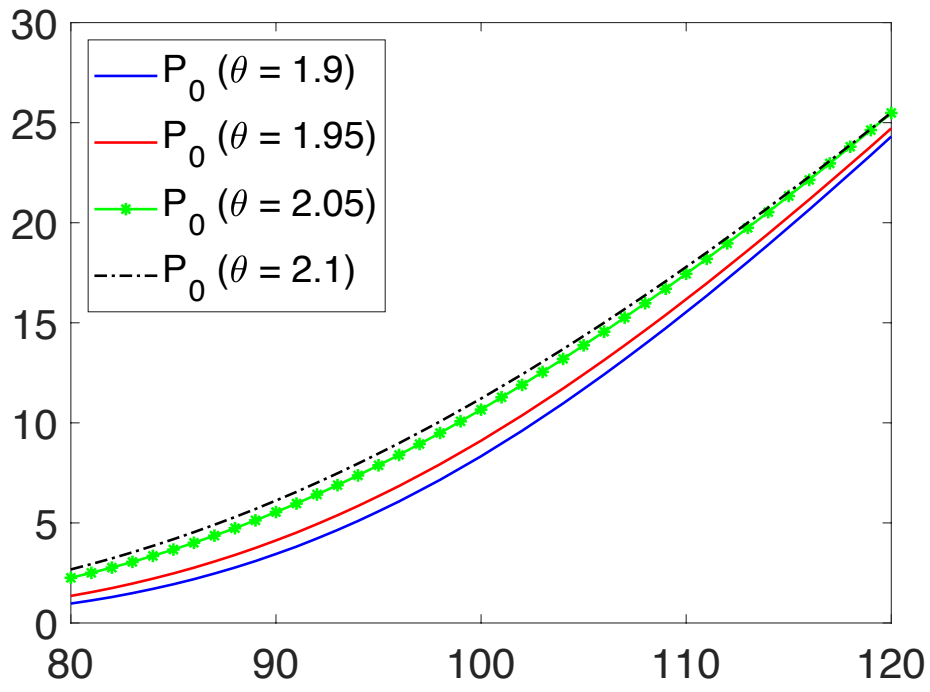
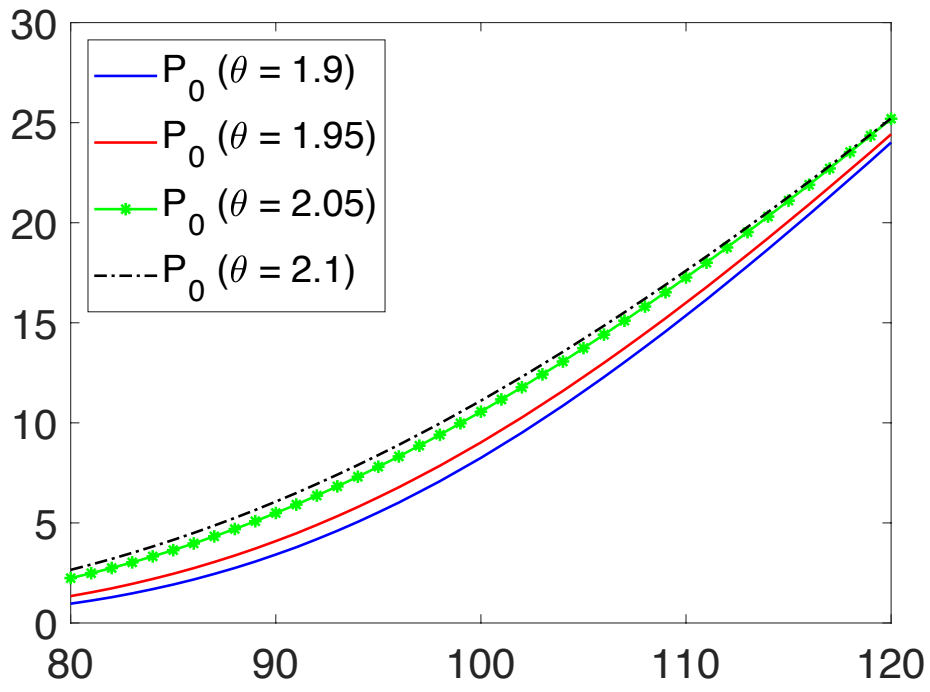
- Price surface ($\theta = 2.05$)

Link

$P_0(t,x,y)$	$P_1(t,x,y)$
	

Option Pricing Numerical Experiments(continue)

- Value of the correction price $(y_0 = 100)$

$D = 60$	$D = 80$
	

Gachon Univ Research: Fit the LDA Model – bag-of-words

- The LDA model is fit using a bag-of-words model.

- A bag-of-words model:
  - The bag-of-words model is a simplifying representation used in natural language processing.
  - In this model, a text is represented as the bag (multiset) of its words, disregarding grammar and even word order but keeping multiplicity. (from Wikipedia)

- Fit an LDA model with the number of topics.
  - ```
    numTopics = 4;
    mdl = fitlda(bag,numTopics)
    ```

```
bag = bagOfWords(documents)
```

```
tbl = topkwords(bag,10)
```

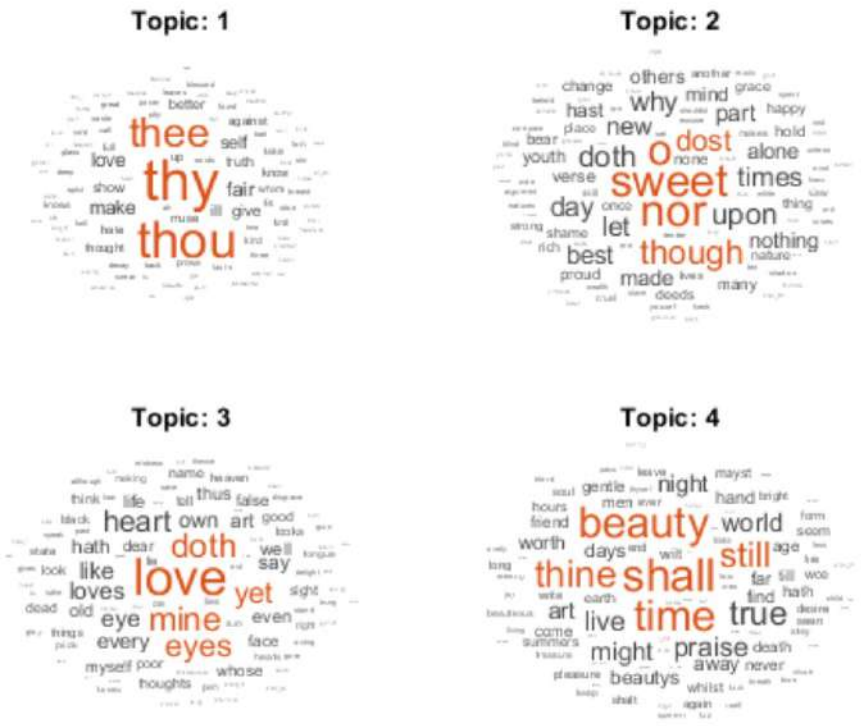
```
tbl=10×2 table
     Word       Count
    _______     _____

    "thy"        281
    "thou"       234
    "love"       162
    "thee"       161
    "doth"        88
    "mine"        63
    "shall"       59
    "eyes"        56
    "sweet"       55
    "time"        53
```

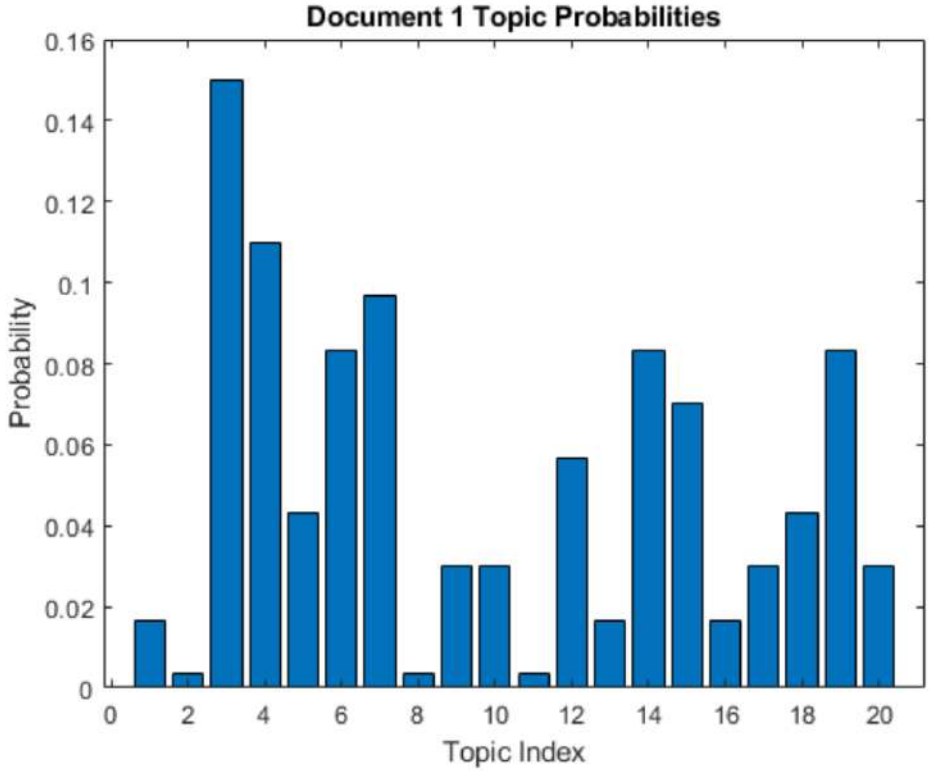
Link
```

Visualization

Visualize the topics using <u>word clouds</u>



View the topic probabilities of the first document in the training data.

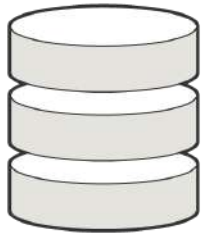


Agenda

- 다양한 금융 데이터 활용 사례
  - 금융사의 데이터 활용 사례
  - 금융 데이터 연구사례
- **금융 데이터 과학**
  - 다양한 금융 데이터 소개
  - MATLAB을 활용한 금융 데이터 과학 기법
  - 금융 데이터 전처리 및 시각화 기능 소개
- 금융 데이터를 활용한 모델링 기법 소개
  - 금융 데이터를 활용한 AI 모델링 기법 소개
  - AI 모델 공유 소개

So many Finance Data Sources

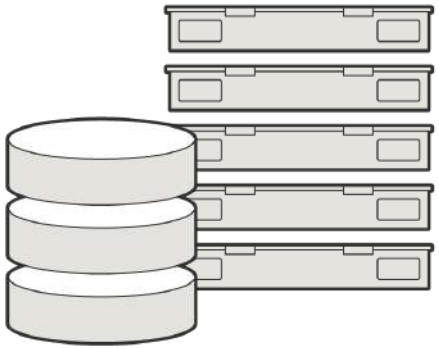
Local disk
Shared folders
Databases



Datafeeds

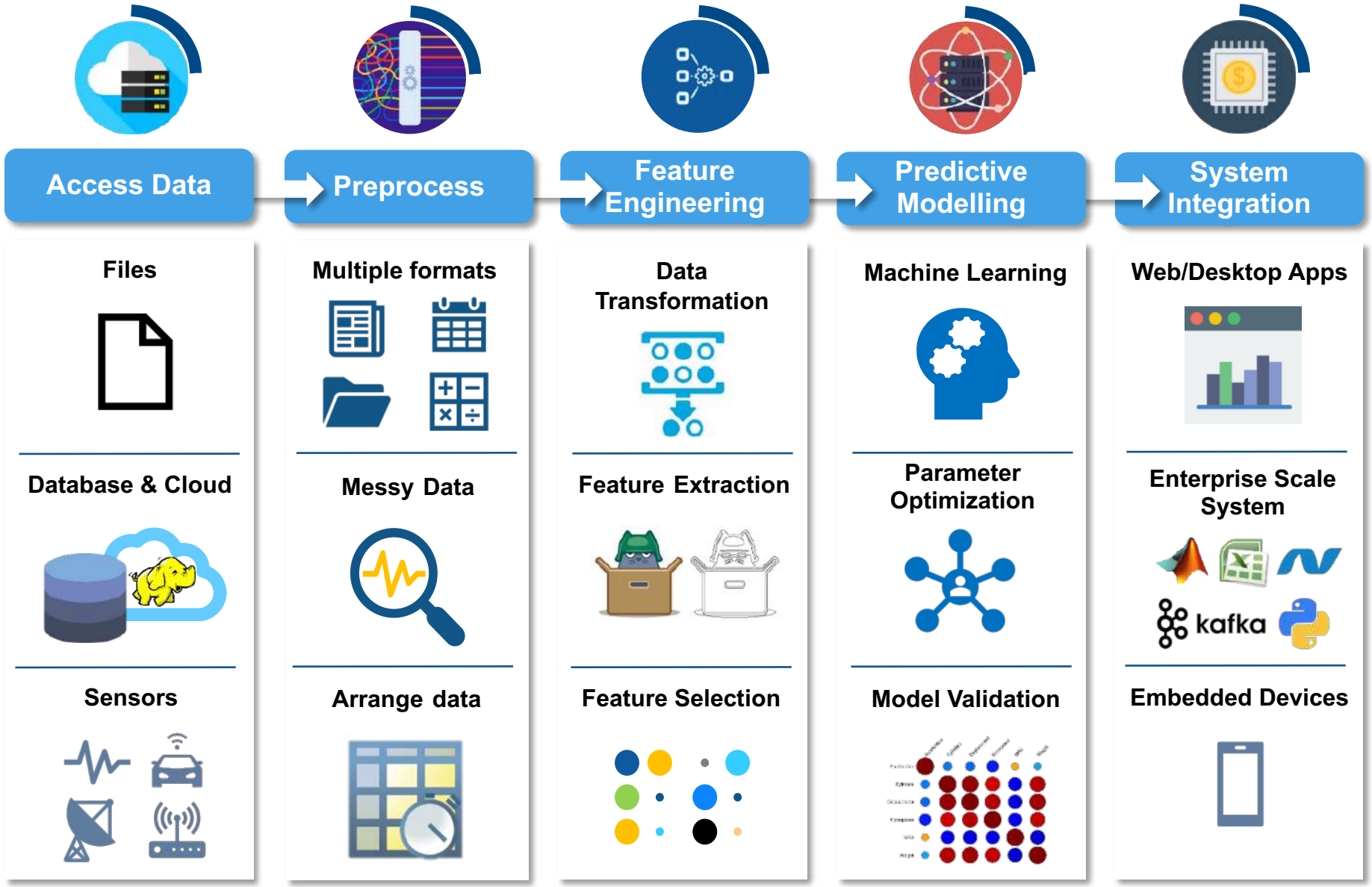
Webpages

Flat files/Excel



Spark+Hadoop

Finance Data Science Workflow



Typical Challenges in Data Cleaning, Management in Finance

- Drowning in Data
  - Data Volume and Variety
  - Different sources, types, sizes
  - Garbage-in garbage-out
- Poor Data Quality
  - Poorly formatted files
  - Irregularly sampled data
  - Redundant, Missing data, Outliers
- Need for more customized analytics
  - No one size fits all

“How can I spend more time adding value and less time juggling data?”

다양한 금융 데이터 플랫폼

- Bloomberg
  - 주식, 채권, 환율, 상품(Commodity) 등
  - 리서치, 실시간 데이터, 뉴스

Retrieve Current and Historical Data Using Bloomberg

This example shows how to connect to Bloomberg® and retrieve current and historical Bloomberg® market data. For details about Bloomberg® connection requirements, see Data Server Connection Requirements. To ensure a successful Bloomberg connection, perform the required steps before executing a connection function. For details, see Installing Bloomberg and Configuring Connections.

Connect to Bloomberg®

Create a Bloomberg® Desktop connection.

```
c = blp;
```

Alternatively, you can connect to the Bloomberg® Server using blpsrv or Bloomberg® B-PIPE® using bpipe.

Retrieve Current Data

Format MATLAB® data display for currency.

```
format bank
```

Retrieve closing and open prices for Microsoft®.

```
sec = 'MSFT US Equity';
fields = {'LAST_PRICE';'OPEN'}; % closing and open prices

[d,sec] = getdata(c,sec,fields)
```

```
d =

  struct with fields:

    LAST_PRICE: 62.32
          OPEN: 62.48
```

다양한 금융 데이터 플랫폼

- Technologies X_TRADER
  - 금, 은, 플루토늄, 팔라듐, 알루미늄 등 귀금속과 기본 금속류, 원유, 가솔린, 등유 같은 에너지, 고무, 금 옵션에 대한 선물 금융 데이터

Create Order Using X_TRADER

R2022a

This example shows how to connect to Trading Technologies® X_TRADER® and create a market order.

Connect to Trading Technologies X_TRADER

```
c = xtrdr;
```

Create Instrument for Contract

Create an instrument for a contract of CAISO NP15 EZ Gen Hub 5 MW Peak Calendar-Day Real-Time LMP Futures with an expiration date of August 2014 on the Chicago Mercantile Exchange.

```
createInstrument(c,'Exchange','CME','Product','ZF',...
                 'ProdType','Future','Contract','Aug14',...
                 'Alias','SubmitOrderInstrument3')
```

Register Event Handler for Order Server

Register an event handler to check the order server status.

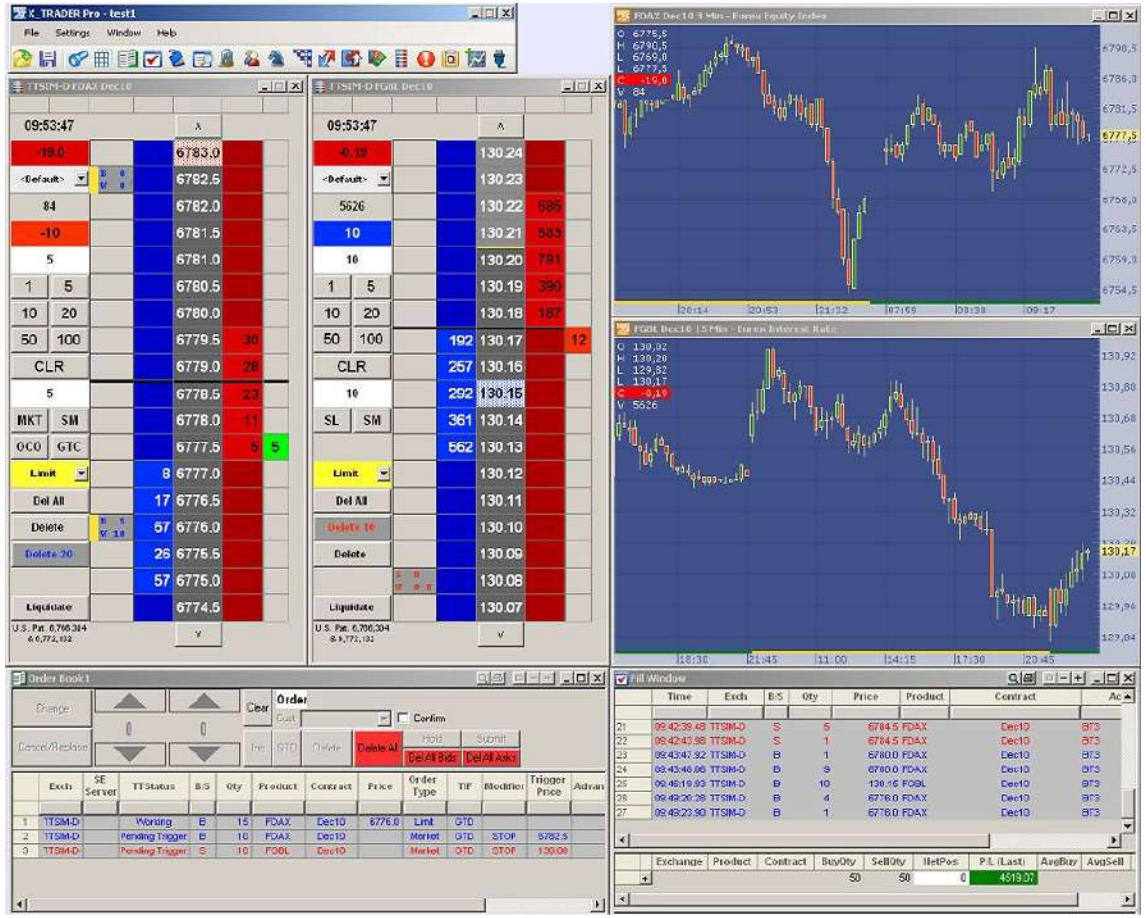
```
sExchange = c.Instrument.Exchange;
c.Gate.registerevent({'OnExchangeStateUpdate', ...
    @(varargin)ttordersserverstatus(varargin{:},sExchange)})
```

Create Order Set and Set Order Properties

Create an empty order set. Then, set order set properties. Setting the first property to true (1) enables the X_TRADER API to send order rejection notifications. Setting the second property to true (1) enables the X_TRADER API to add order pairs for all order updates to the order tracker list in this order set. Setting the third property to ORD_NOTIFY_NORMAL sets the X_TRADER API notification mode for order status events to normal.

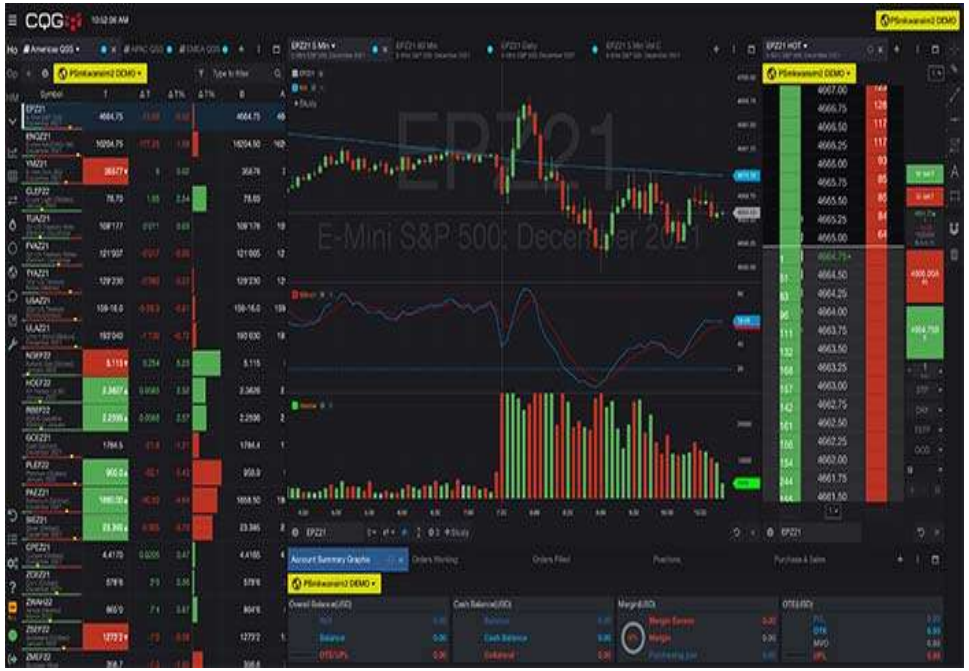
```
createOrderSet(c)

c.OrderSet(1).EnableOrderRejectData = 1;
c.OrderSet(1).EnableOrderUpdateData = 1;
c.OrderSet(1).OrderStatusNotifyMode = 'ORD_NOTIFY_NORMAL';
```



다양한 금융 데이터 플랫폼

- CQG
  - 증권사 HTS와 유사한 형태
  - 금융 데이터 제공 및 데이터 차트 및 분석자료 제공



Datafeed Toolbox Support

Supported Data Providers

- Bloomberg Desktop
- Bloomberg B-PIPE
- Bloomberg Server
- Datastream Web Services R2018b
- Refinitiv Machine Readable News
- Refinitiv Tick History R2018a
- FactSet Data Server
- Federal Reserve Economic Data (FRED)
- Haver Analytics R2018b
- IHS Markit
- Money.Net
- Quandl
- SIX Financial Information
- Twitter R2017b

Supported Trading Systems

- Bloomberg EMSX
- Trading Technologies X_TRADER
- CQG
- Wind Data Feed Services R2018a

Support Packages in File Exchange

- FRED® REST R2019a
- RavenPack® Analytics R2018b
- Barchart® OnDemand R2019a

금융 데이터 종류

- Fundamental Data
  - 기업의 순자산, 부채, 매출액, 순이익
  - 주식, 채권, 파생상품, 외환, 원자재 등
- Market Data
  - 가격
  - 거래량
  - 호가
  - 배당
  - 듀레이션
- Analysis Data
  - 애널리스트 기업 분석
  - 예상 실적 전망치
  - 기업 신용 평가 등
- Alternative Data
  - SNS: Twitter, Facebook
  - CCTV 영상
  - 구글 검색량(네이버 검색량)
  - 이미지 데이터(위성 사진 등)

금융 데이터 종류

- Fundamental Data
  - 공시나 감독 기관에 제출하는 자료
  - 거시경제 데이터
  - 분기별 재무제표 데이터
  - 자산, 부채, 매출, 비용 및 이익 등의 데이터
  - 주로 투자하고자 하는 대상에 대한 정보가 있음
  - 데이터의 양이 매우 적은 편
- Alternative data
  - 주로 비 구조화된 형태의 데이터
  - SNS, 위성사진, CCTV 영상, 상품 리뷰 등
- Market Data
  - 가격, 수익률, 변동성, 거래량, 배당/쿠폰, 시중 금리, 호가 데이터 등 거래소와 같은 시장에서 생성되는 데이터
  - 생성되는 데이터가 많음
  - 시계열 데이터 많음
- Analytics data
  - 다른 데이터를 기반으로 분석을 통해 만들어진 새로운 데이터
  - 애널리스트의 추천 정보나 신용 평가 정보, 다음 실적 예측치
  - 데이터 자체가 이미 원본 데이터에서 어느 정도의 가공을 통해 정보를 추출해낸 형태임

금융 데이터 형태

- 금융 데이터 형태
  - 시계열 데이터
    - 주가, 파생상품의 가격 등
  - 문자형 데이터
    - SNS, 상품 리뷰 등의 데이터 형태
  - 숫자형 데이터
    - 기본적인 금융 데이터 형태
  - 빅 데이터
    - 방대한 양의 비정형 데이터 포함

```
ans = 508×1 string array
    "Walmart: "you wanna destroy Amazon?" Google: "bet" $WMT $GOOG
    "$WMT wants next level customer service w/highly personalized
    "Ironic prelude to $DIS buying $TWTR soon IMO $AAPL $GOOG $SPY
    "$AMZN the $WMT threat grows each and every day  https://t.co/
    "MU Investments Co. Ltd. Sells 30 Shares of Alphabet Inc. $GOO
    "Ad $ are going to $GOOG and $FB away from wppgy  #Advertising
    "Big bullish unusual option activity detected: $SPX, $GOOG, $C
    "REPORT: Apple to build data center in Iowa: https://t.co/jwH6
    "RT @theflynews: REPORT: Apple to build data center in Iowa: h
```

person	year	income	age	sex
1	2001	1300	27	1
1	2002	1600	28	1
1	2003	2000	29	1
2	2001	2000	38	2
2	2002	2300	39	2
2	2003			

Date	Close	High	Low	Open	Volume
2017-Aug-17 09:30:00	925.87	925.87	925.78	925.78	13585
2017-Aug-17 09:31:00	923.58	925.45	923.26	925.45	5667
2017-Aug-17 09:32:00	925.2	925.48	923.985	924.42	9254
2017-Aug-17 09:33:00	925.5407	925.5407	925.27	925.27	1500
2017-Aug-17 09:34:00	924.63	925.72	924.63	925.64	1505
2017-Aug-17 09:35:00	925	925.39	924.98	925.06	1112
2017-Aug-17 09:36:00	924.7	925.22	924.11	925.195	3092
2017-Aug-17 09:37:00	924.85	924.9799	924.21	924.3	1285
2017-Aug-17 09:38:00	925.04	925.86	924.46	924.87	6171
2017-Aug-17 09:39:00	926.15	926.15	925.13	925.13	1400
2017-Aug-17 09:40:00	926.14	926.17	925.38	925.76	1000
2017-Aug-17 09:41:00	925.9364	926.295	925.93	926.16	525
2017-Aug-17 09:42:00	926.55	926.62	926.11	926.5	700
2017-Aug-17 09:43:00	926.25	926.25	925.83	925.88	700
2017-Aug-17 09:44:00	925.94	925.94	925.03	925.51	1400

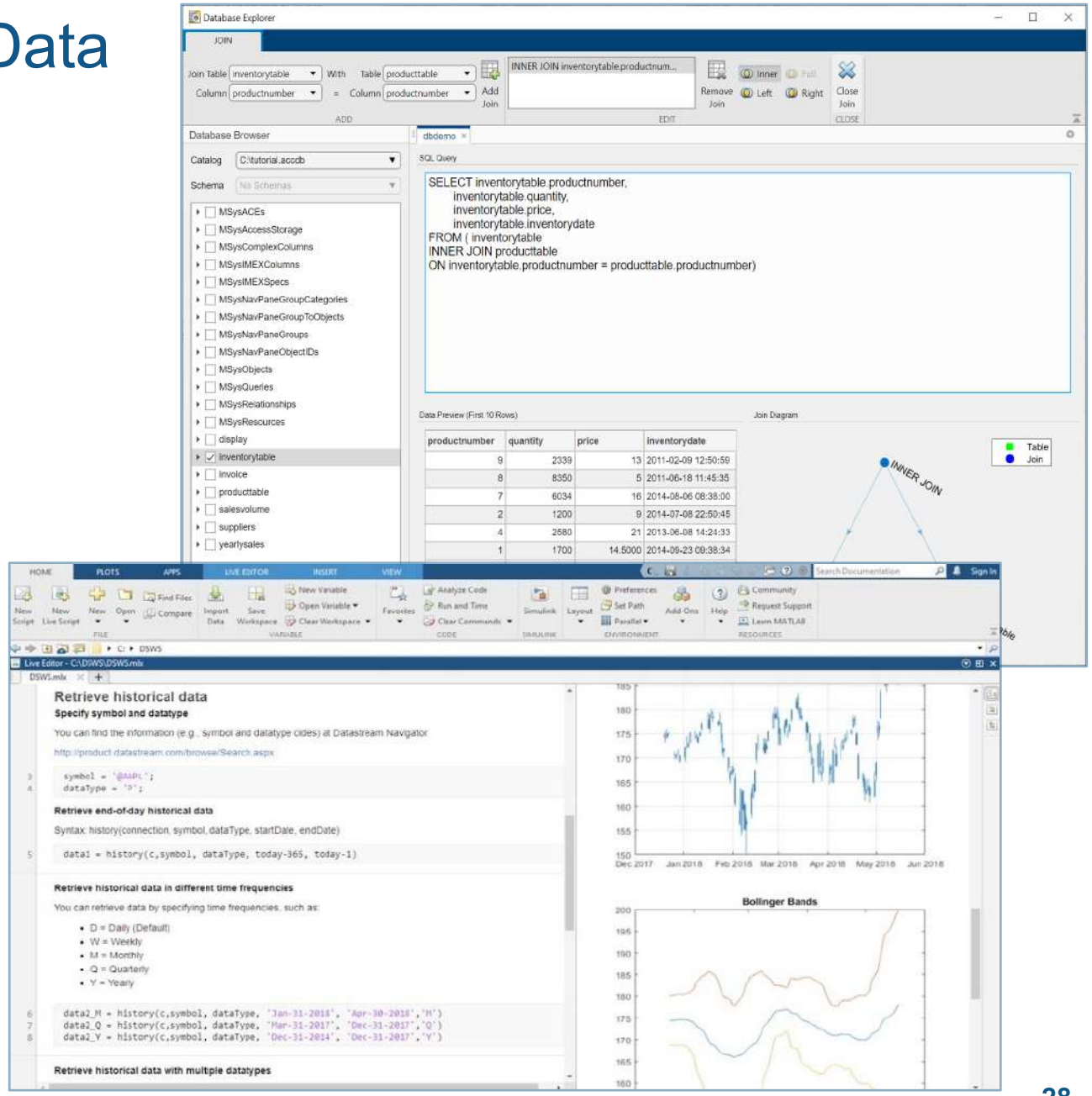
Access Financial and Alternative Data

Databases: Exchange data with relational and nonrelational databases

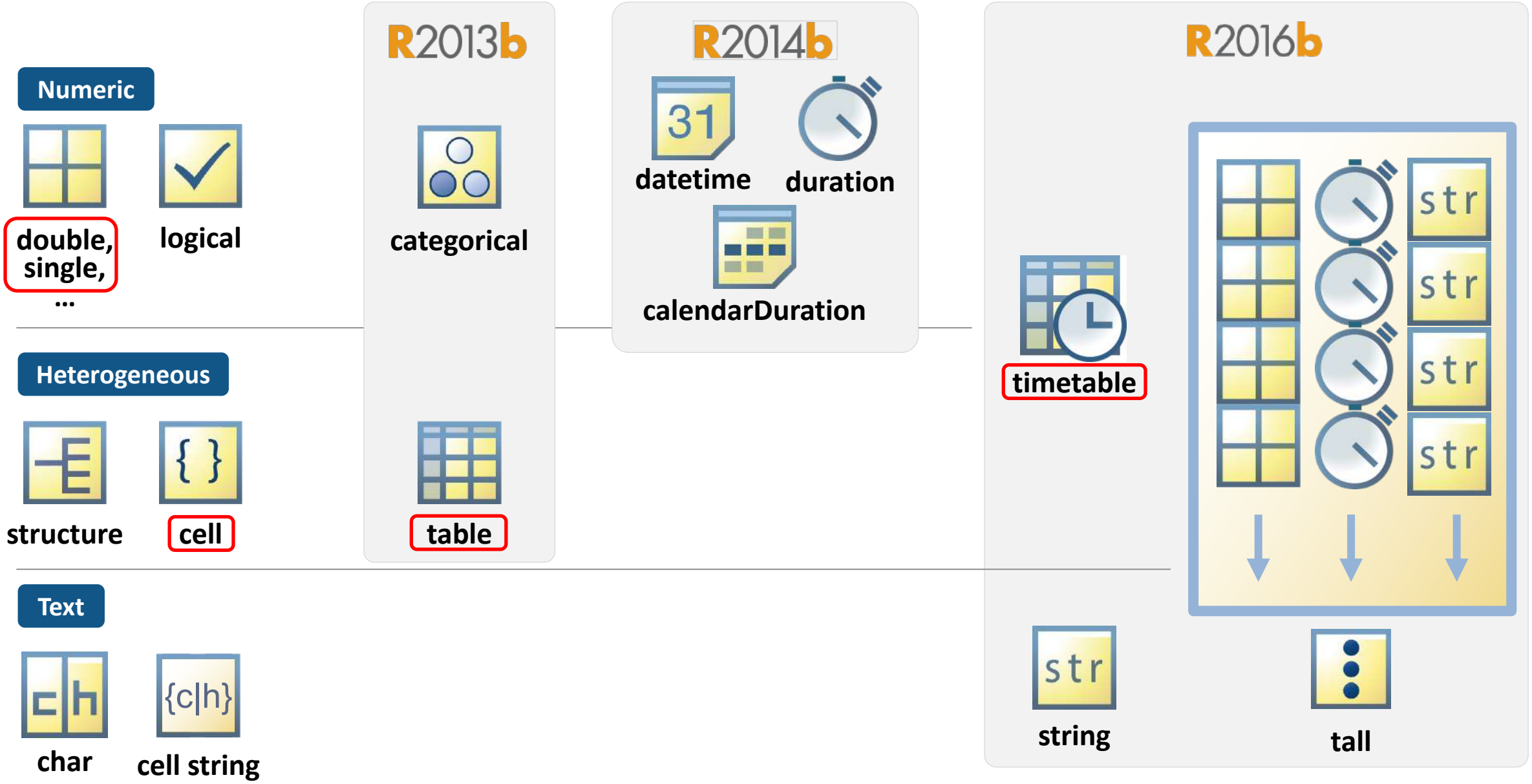
- Additional native interface support for popular SQL and NoSQL databases
- Customize import options and generate dynamic SQL queries with **databaseImportOptions**
- Performance improvements to JDBC, ODBC, and native database access workflows

Data feeds: Access financial data from data service providers

- Trading Toolbox functionality merged into Datafeed Toolbox in R2021a



MATLAB Data Types express more types of finance data naturally

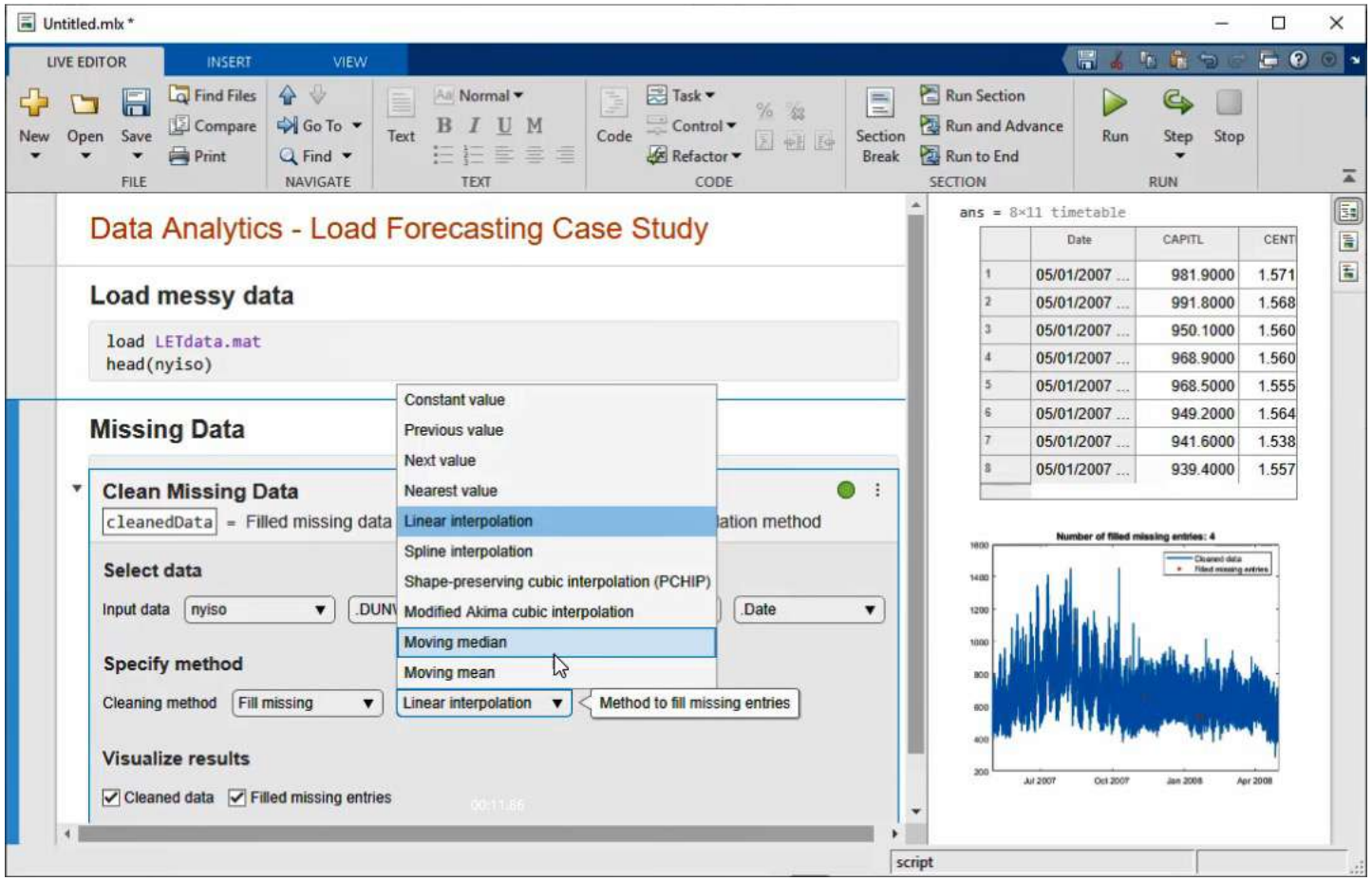


Agenda

- 다양한 금융 데이터 활용 사례
  - 금융 회사 사례
  - 금융 데이터 연구사례
- **금융 데이터 과학**
  - 다양한 금융 데이터 소개
  - MATLAB을 활용한 금융 데이터 과학 기법
  - 금융 데이터 전처리 및 시각화 기능 소개
- 금융 데이터를 활용한 모델링 기법 소개
  - 금융 데이터를 활용한 AI 모델링 기법 소개
  - AI 모델 공유 소개

Introduce MATLAB Live Editor

- Live scripts are living whitepapers
- Use contextual hints when calling functions
- Automatically generate code when interacting with plots and tables in the output
- Add Live Tasks to interactively explore parameters and options
  - Tasks are lightweight apps that can be included in scripts



Create Live scripts that combine code, output, and formatted text in an executable notebook

R2016a

Live Editor - C:\MATLAB\SunriseSunset.mlx

LIVE EDITOR | INSERT | VIEW

New | Open | Save | Find Files | Compare | Print | Go To | Find | Text | Normal | B I U M | Code | Section Break | Run Section | Run and Advance | Run to End | Run | Step | Stop

FILE | NAVIGATE | TEXT | CODE | SECTION | RUN

Estimating Sunrise and Sunset

We can calculate sunrise and sunset times from the following equations.

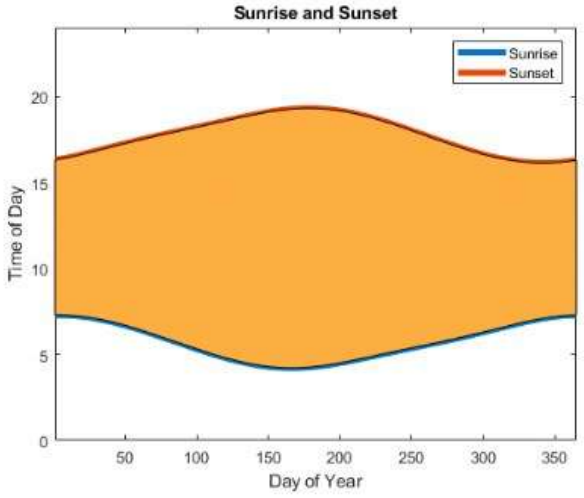
$$\text{sunrise} = 12 - \frac{\cos^{-1}(-\tan\phi\tan\delta)}{15^\circ} - \frac{SC}{60} \qquad \text{sunset} = 12 + \frac{\cos^{-1}(-\tan\phi\tan\delta)}{15^\circ} - \frac{SC}{60}$$

```
long =   -71 ;
lat  =    42 ;
timeZone = 'Eastern (UTC-5)';

sc = solarT                    imeZone);
delta = asi                    (360*(days - 81)/365));
sunrise = 1                    )*tand(delta))/15 - sc/60;
sunset  = 12 + acosd(-tand(lat)*tand(delta))/15 - sc/60;

plot(days, sunrise, days, sunset, 'LineWidth', 4)
title('Sunrise and Sunset')
xlabel('Day of Year')
```

'Eastern (UTC-5)'
'Central (UTC-6)'
'Mountain (UTC-7)'



Live Editor

https://www.mathworks.com/products/matlab/live-editor.html

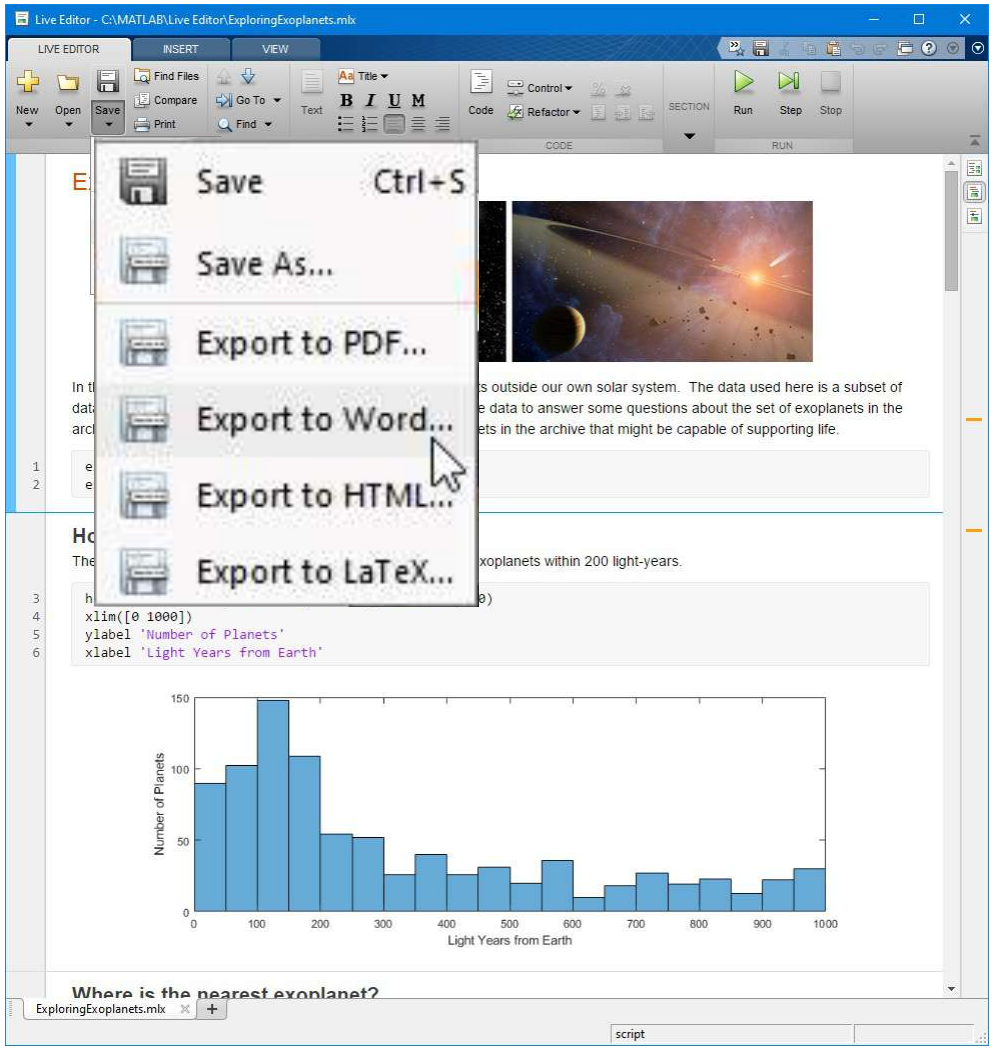
Document as you go – *your script is your report*

Code + Output + Formatted Text

↓

Executable Notebook

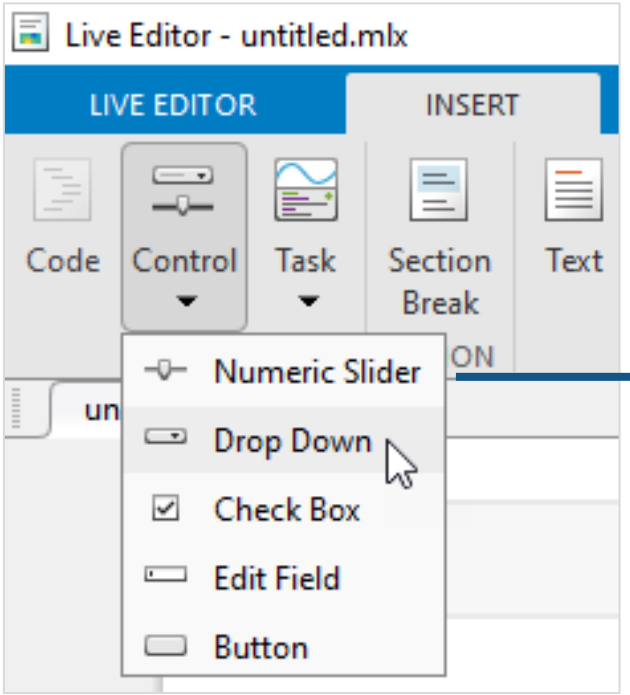
- Divide code into sections
- Embed outputs next to the code
- Add rich text formatting, equations, images, and hyperlinks
- Include animations with embedded controls R2021a, and export R2021b
- Programmatically control fonts
- Save directly to PDF, HTML, Word, and LaTeX

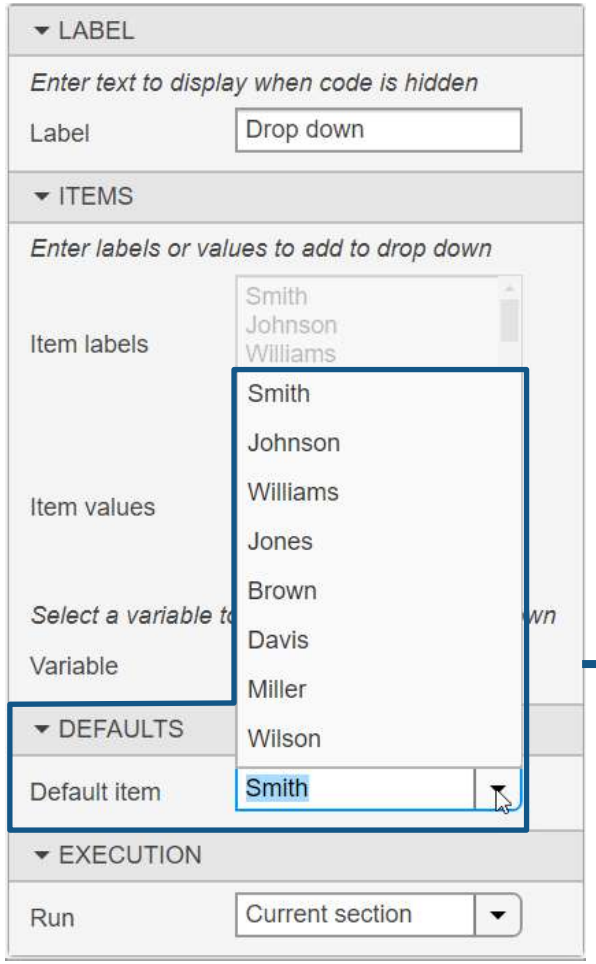


Turn your script into an app – *in seconds*

```
lastnames = ["Smith", "Johnson", "Williams",
"Jones", "Brown", "Davis", "Miller", "Wilson"];
```

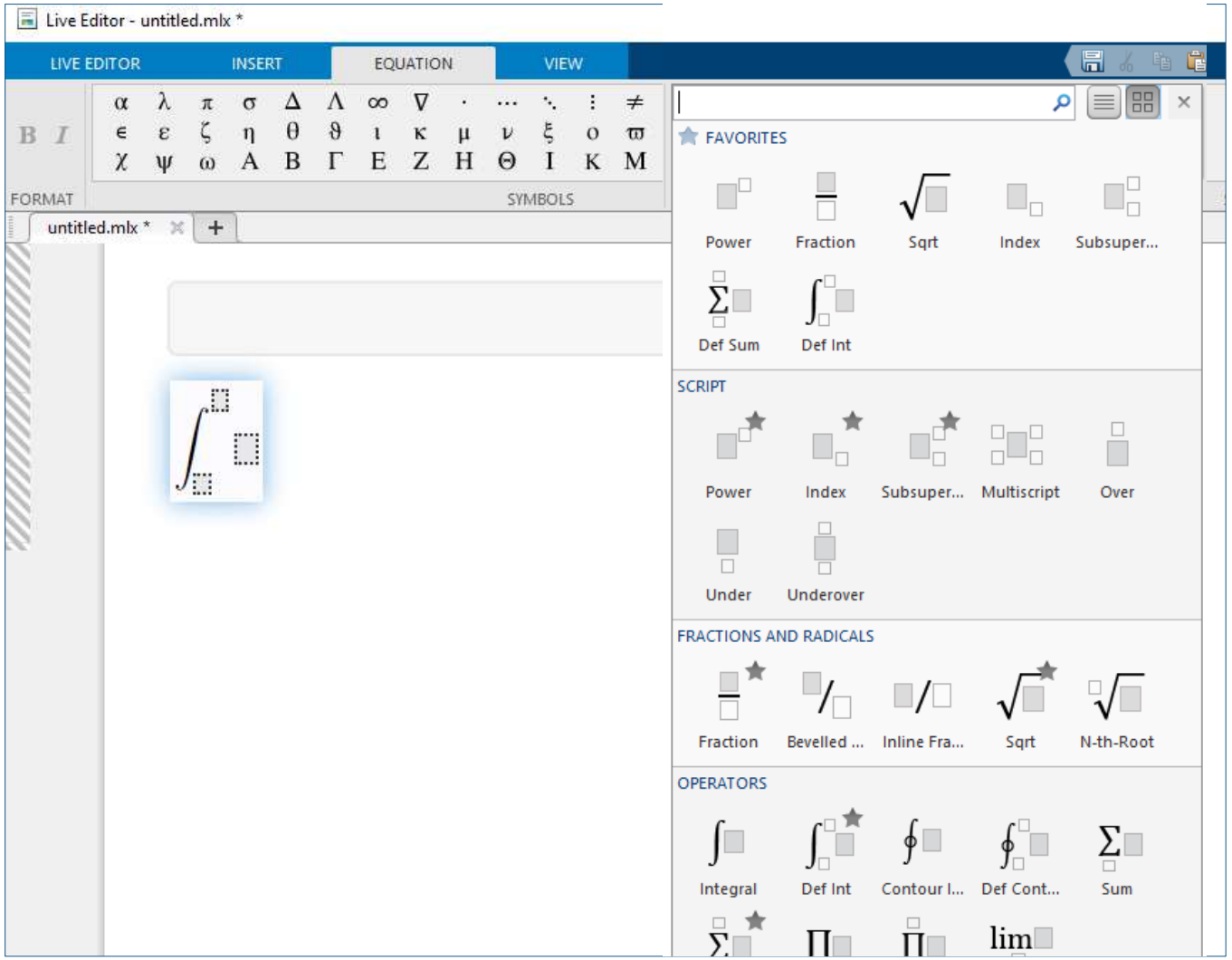
- Add interactive controls
  - Sliders
  - Drop downs
  - Check boxes
  - Edit fields
  - Buttons
- Hide code to create a small app
- Link variables to drop down items and slider values
- Set default values for sliders, drop-down lists, check boxes, and edit fields R2021b



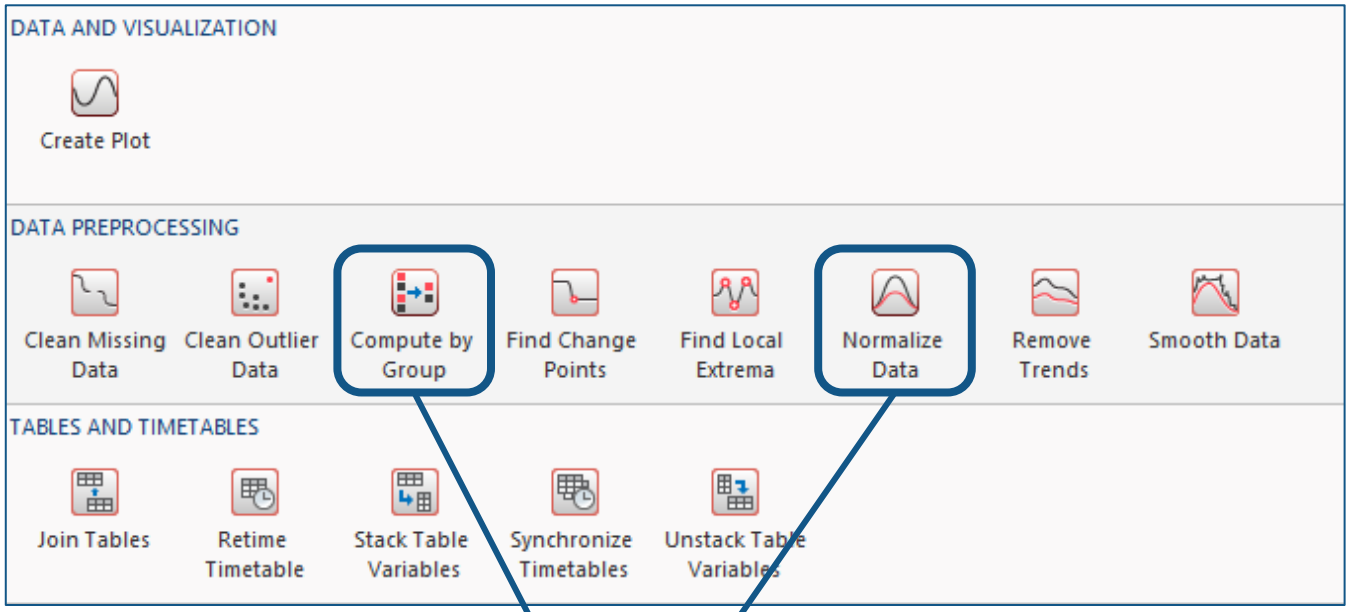


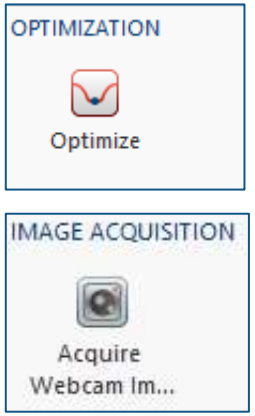
Explore, code, and debug more efficiently

- Interactively explore embedded figures and tables
- Use contextual hints for function arguments and file names
- Troubleshoot with a fully integrated debugger
- Use bookmarks to navigate
- Select and edit rectangular areas of code
- Interactively edit equations



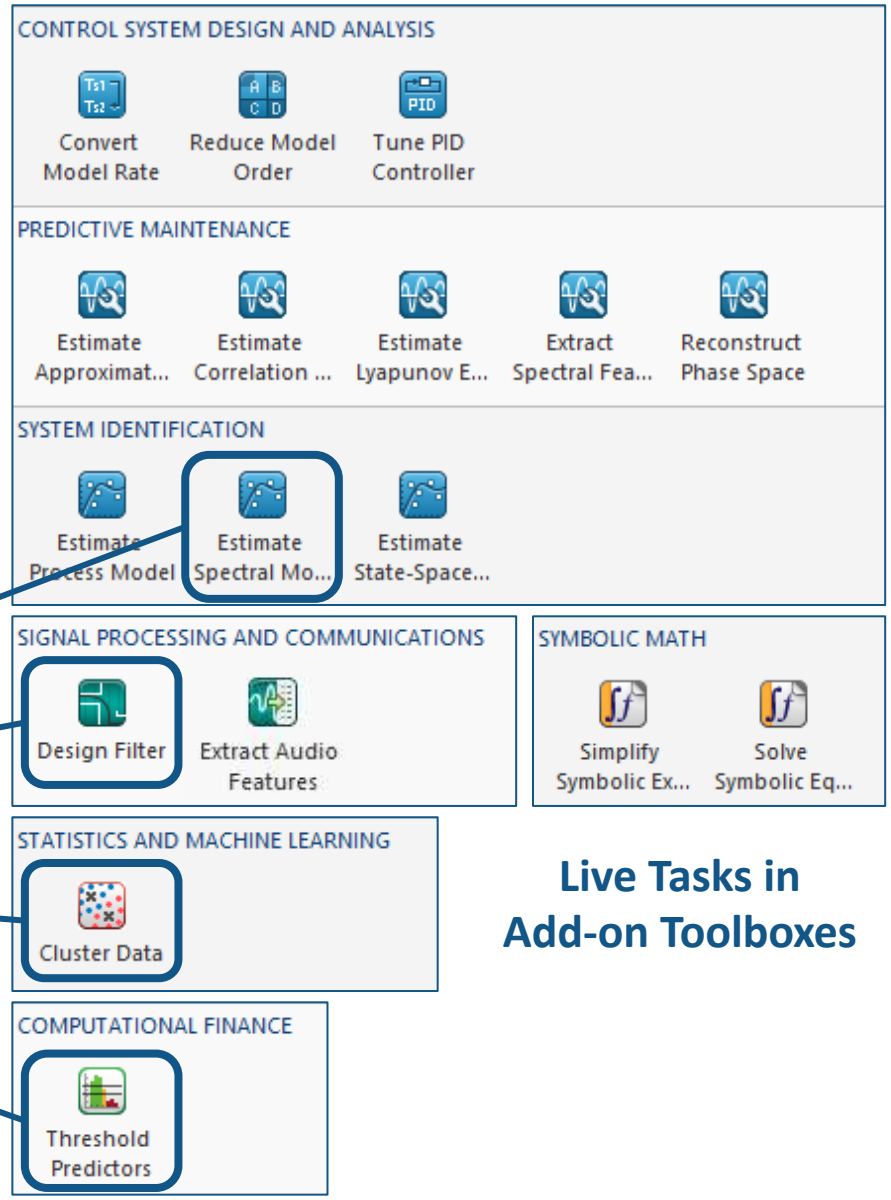
There are many Live Editor tasks – *with more available each release*





Live Tasks in MATLAB

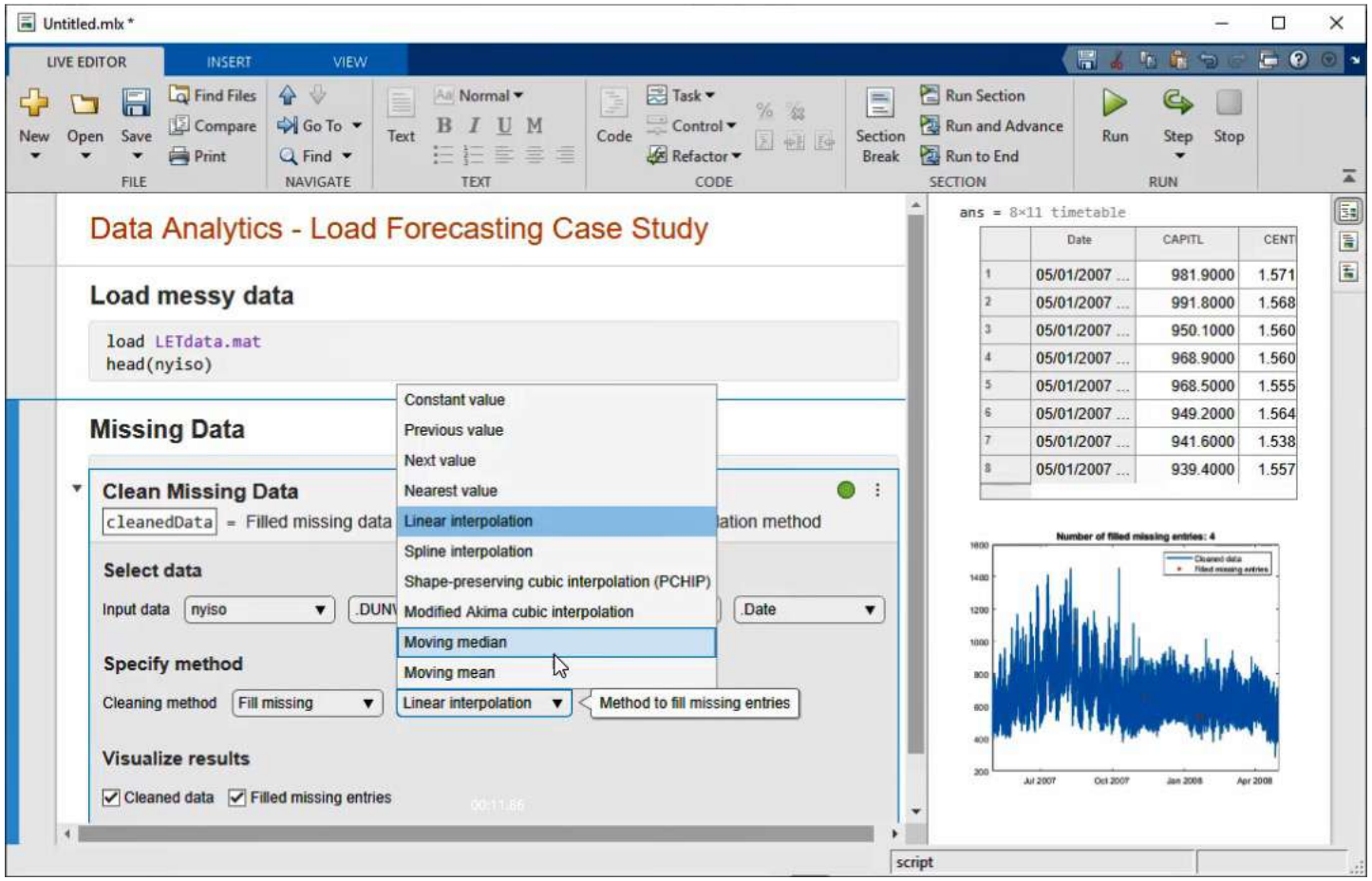
R2021b



Live Tasks in Add-on Toolboxes

Program – *without coding*

- Automatically generate code when interacting with plots and tables in the output
- Add Live Editor tasks to interactively explore parameters and options



https://www.mathworks.com/help/matlab/matlab_prog/add-live-editor-tasks-to-a-live-script.html

Use MATLAB and Simulink with no downloads or installations.

Collaborate using projects

R2021b

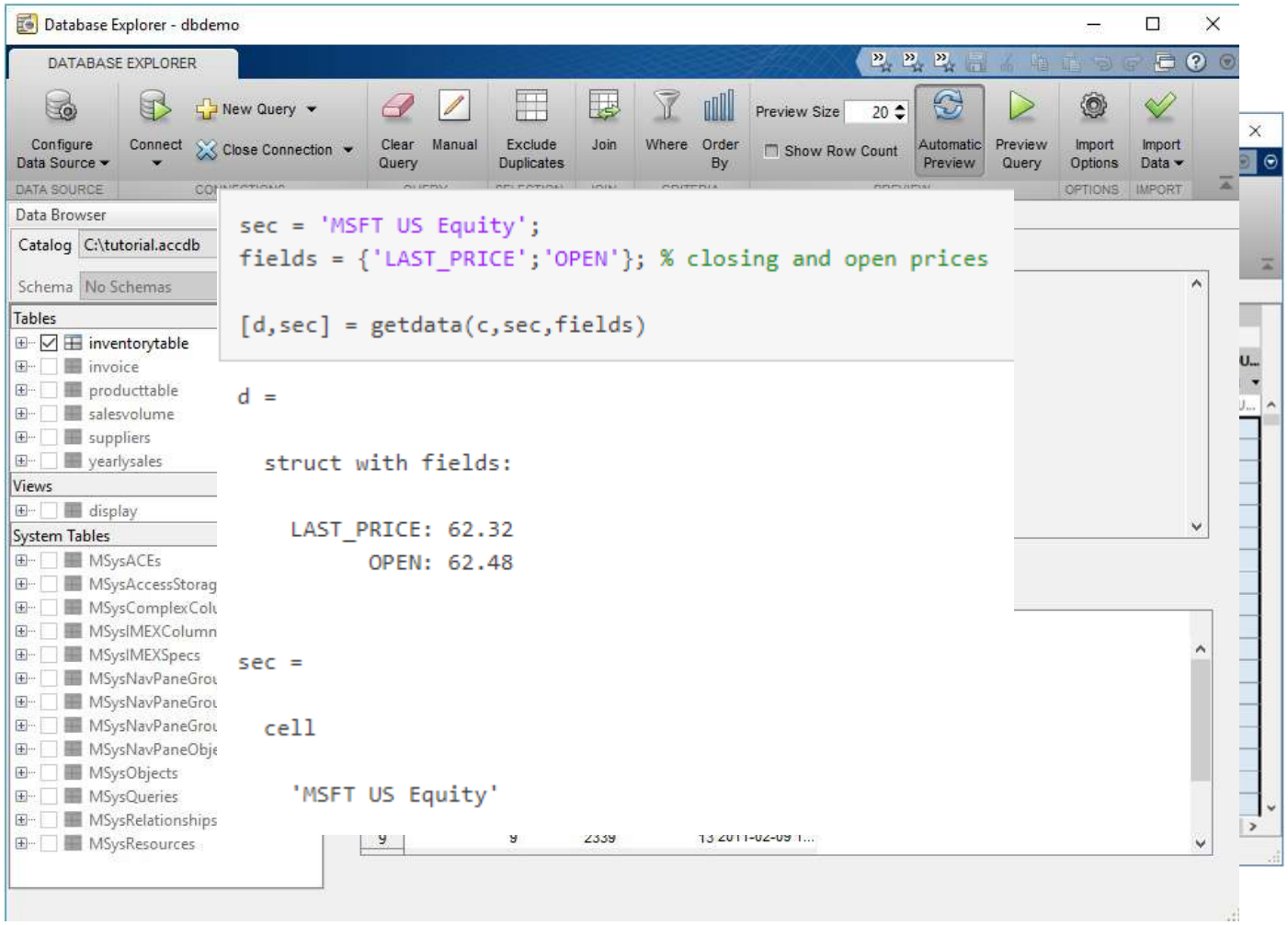
Collaborate with others through online sharing and publishing

Store, manage, and access your files anywhere.

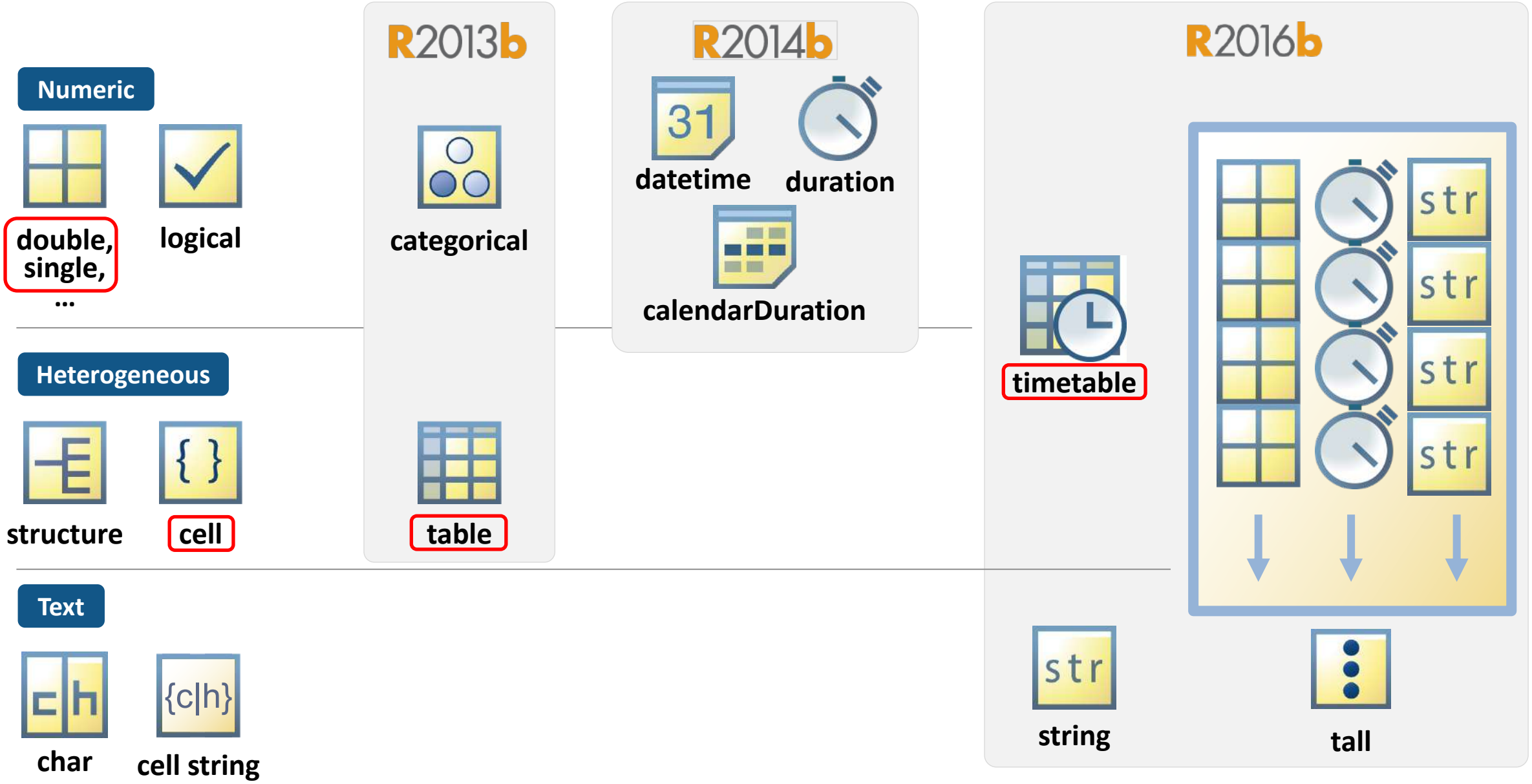
All **named** academic, commercial and home user accounts are eligible

Data Explorer

- 파일로 데이터를 가진 경우
  - Use the import tool to import data
  - Generate code for similar files
- DB로 데이터를 가진 경우
  - Use Database Explorer App for SQL to view and query*
  - Or connect and query via code*
- 실시간 데이터를 받는 경우
  - MATLAB can connect to a host of live datafeeds**

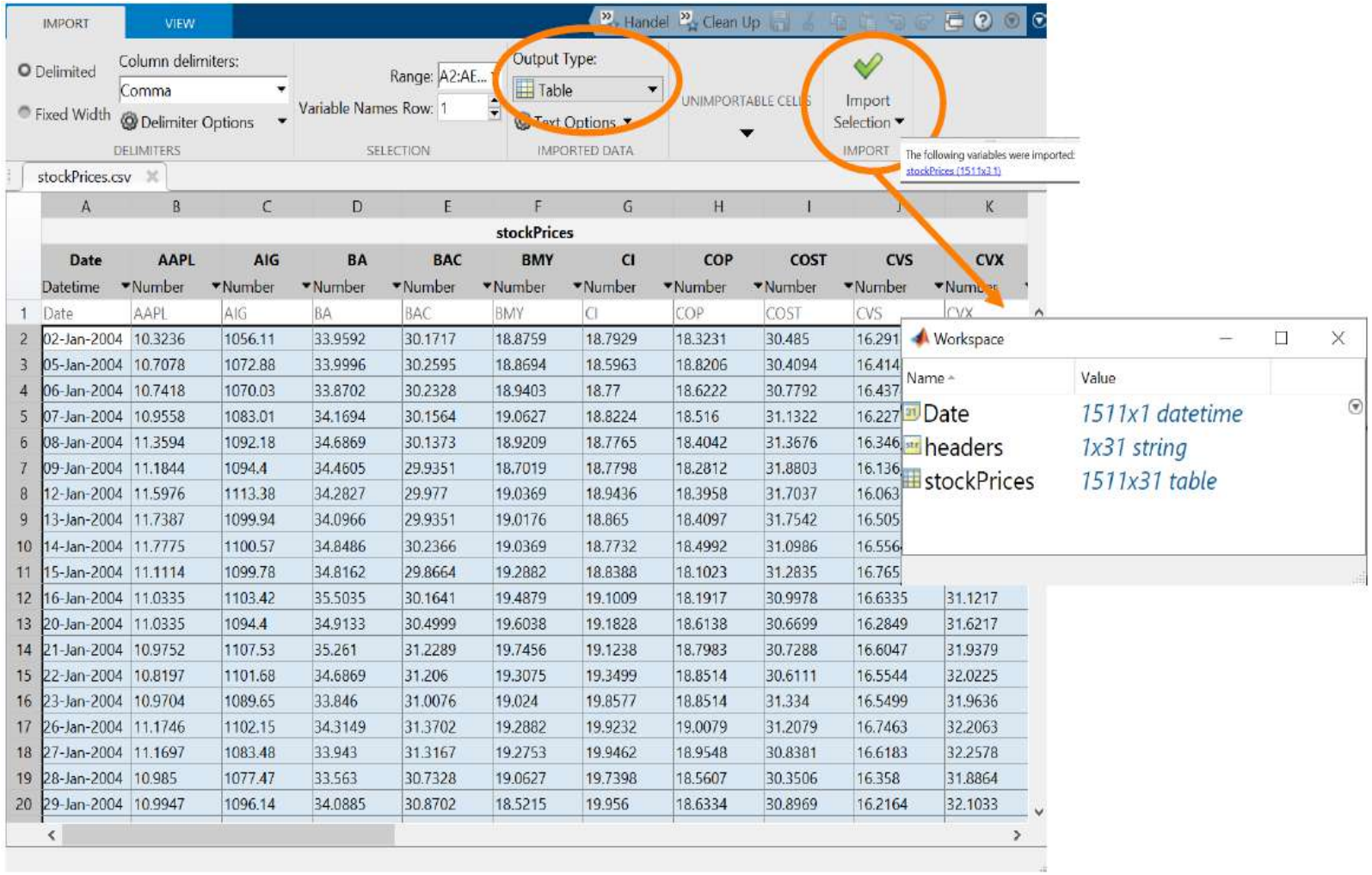


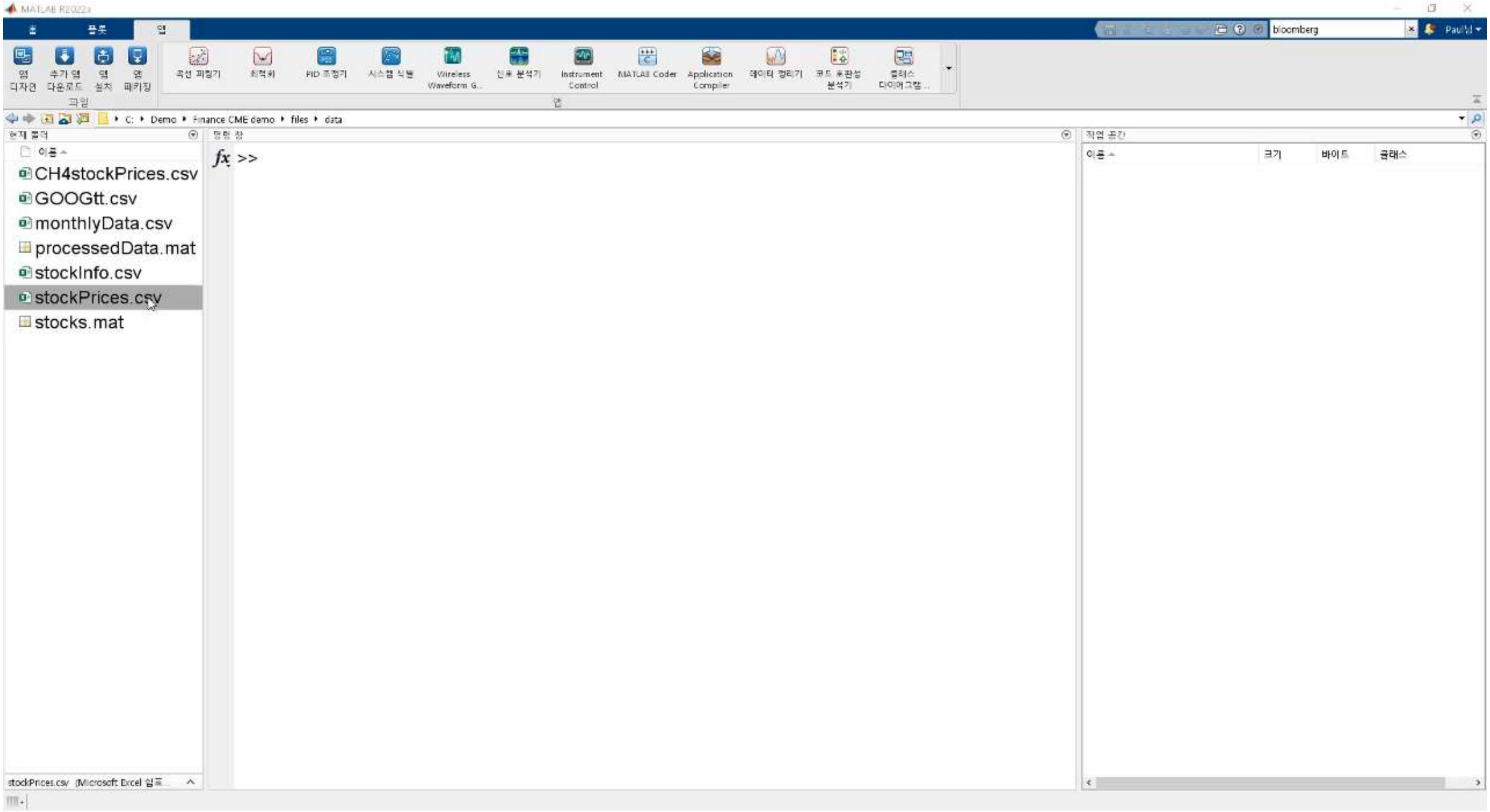
MATLAB Data Types express more types of data naturally



Access data interactively using the Import Tool

- Select data types
- Choose what to do with missing data
- Generate MATLAB code




MATLAB R2022a
홈
플롯
앱
bloomberg
앱 디자인
추가 앱 다운로드
앱 설치
앱 패키징
곡선 피팅기
최적화
PID 조정기
시스템 식별
Wireless Waveform G...
신호 분석기
Instrument Control
MATLAB Coder
Application Compiler
데이터 정리기
코드 호환성 분석기
클래스 다이어그램 ...
C: ▸ Demo ▸ Finance CME demo ▸ files ▸ data
현재 폴더
이름
CH4stockPrices.csv
GOOGtt.csv
monthlyData.csv
processedData.mat
stockInfo.csv
stockPrices.csv
stocks.mat
명령 창
fx >>
작업 공간
이름
크기
바이트
클래스
stockPrices.csv (Microsoft Excel 쉼표...

Tables

R2013b

- **For:**
  - Mixed-type tabular data
  - Include metadata
- **Provides:**
  - Flexible indexing
  - Data organization
    - joins, stack/unstack, etc.

```
data(1:10,["begin time","state","event type","event narrative","damage total ($)"])
```

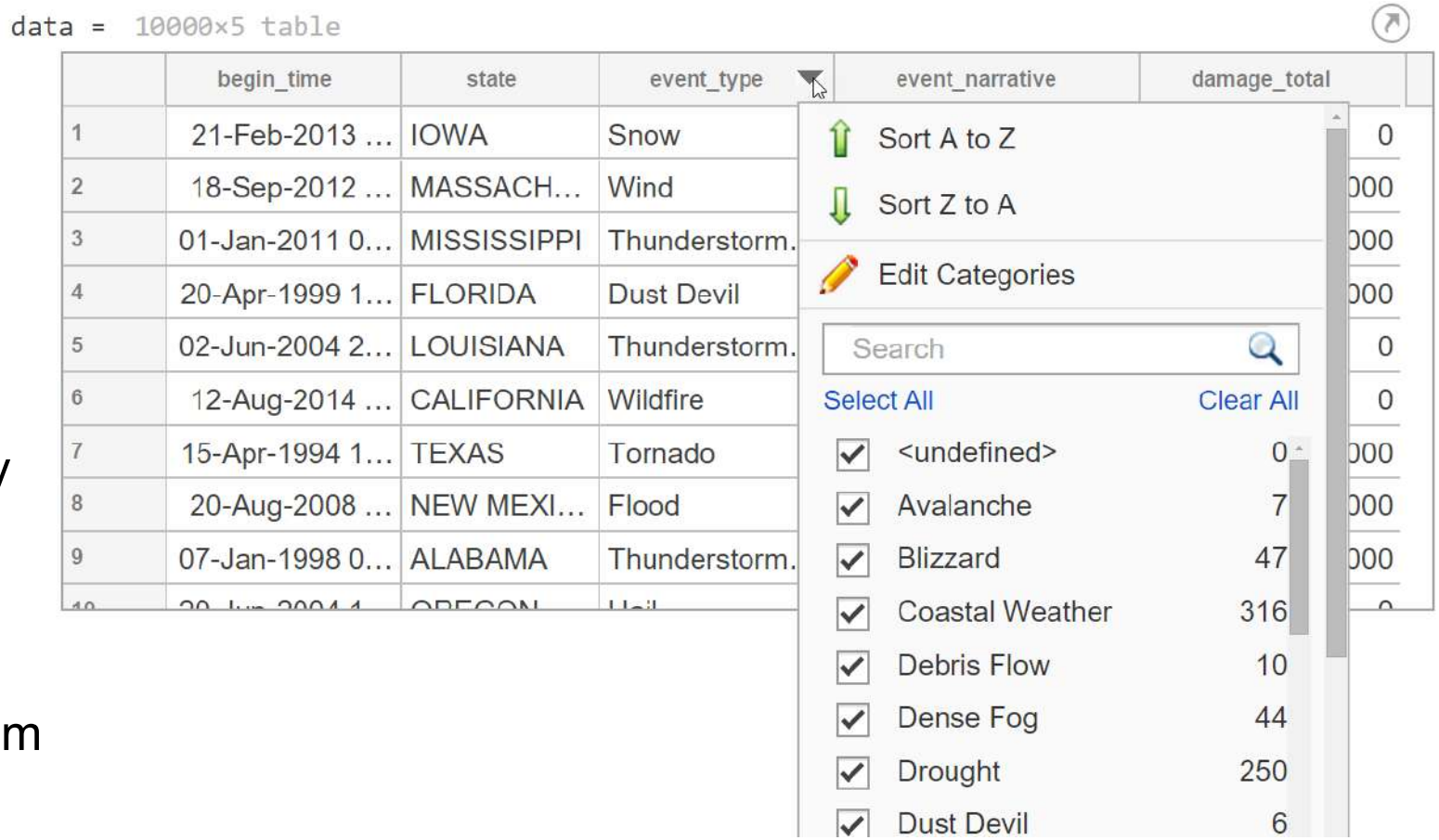
ans = 10×5 table

	begin time	state	event type	event narrative	damage...
1	07-Apr-09 01:00:00	KENTUCKY	Winter Weather	"white out conditions and 1....	0
2	10-Jul-05 16:00:00	ALABAMA	Tropical Storm	"several trees and power lin...	47
3	19-Feb-13 13:11:00	NEVADA	High Wind	"this gust occurred at mid-sp...	0
4	01-Sep-13 00:00:00	ARKANSAS	Drought	"severe drought was occurri...	0
5	12-Sep-13 15:25:00	ARKANSAS	Thunderstorm...	"trees were blown down."	5
6	27-Aug-17 04:57:00	TEXAS	Flash Flood	"numerous roads closed du...	150
7	10-Dec-15 00:17:00	MONTANA	High Wind	"at the saco us-2 dot site, su...	0
8	27-May-07 03:00:00	TEXAS	Flash Flood	"scattered showers and thun...	0
9	03-Sep-11 14:35:00	ILLINOIS	Hail	"golf ball sized hail fell in mo...	0

Categorical Arrays

R2013b

- **For:**
  - Discrete non-numeric data drawn from a finite set of possible values
- **Provides:**
  - Improved memory efficiency
  - Ordered categories *("good" > "poor")*
  - "Protected" status which prevents new categories from being inadvertently added



Date and Time Arrays

R2014b

- **For:**
  - Representing a point in time – `datetime`
  - Representing elapsed time – `duration`, `calendarDuration`

- **Provides:**
  - Same data type for computation and display
    - Add, subtract, sort, compare
    - Plotting and axes labels
    - Customize display formats
    - Nanosecond precision
  - Accounts for time zones, daylight savings time, leap years and leap seconds

```
data.storm_duration = abs(data.end_timestamp - data.begin_timestamp)
```

data = 10004×4 table

	begin_timestamp	end_timestamp	event_type	storm_duration
1	21-Feb-2013 14:00:00	22-Feb-2013 03:0...	Snow	13:00:00
2	18-Sep-2012 17:53:00	18-Sep-2012 23:...	Wind	05:17:00
3	01-Jan-2011 01:20:00	01-Jan-2011 01:2...	Thunderstorm...	00:00:00
4	20-Apr-1999 11:20:00	20-Apr-1999 10:2...	Dust Devil	01:00:00
5	02-Jun-2004 22:00:00	02-Jun-2004 21:0...	Thunderstorm...	01:00:00
6	12-Aug-2014 02:00:00	31-Aug-2014 23:...	Wildfire	477:59:00
7	15-Apr-1994 12:37:00	15-Apr-1994 11:3...	Tornado	01:00:00
8	20-Aug-2008 16:49:00	20-Aug-2008 16:...	Flood	00:19:00
9	07-Jan-1998 06:00:00	07-Jan-1998 06:0...	Thunderstorm...	00:02:00
10	29-Jun-2004 15:40:00	29-Jun-2004 14:4...	Hail	00:55:00

Timetables

R2016b

- **For:**
  - Time-stamped tabular data
- **Provides:**
  - Indexing by time, time range, or within a tolerance around a time
  - Retiming to create a constant sample rate
  - Synchronizing multiple timetables to align based on time stamps
  - All supported table functionality

```
data(timerange("01-Jan-2017","17-Mar-2017"),:)
```

ans = 161×4 timetable

	begin_timestamp	state	event_type	event_narrative	damage_total
1	21-Jan-2017 13:02:00	GEORGIA	Thunderstorm…	"a tree was blown d…	0
2	21-Jan-2017 05:14:00	ALABAMA	Tornado	"the tornado first tou…	750
3	05-Jan-2017 04:00:00	OHIO	Winter Weather	"the county garage …	0
4	05-Mar-2017 18:00:00	OREGON	Snow	"there were reports …	0
5	04-Feb-2017 12:15:00	WYOMING	Wind	"the wydot sensor a…	0
6	08-Feb-2017 08:00:00	INDIANA	Winter Weather	"the observers locat…	0
7	18-Jan-2017 18:00:00	CALIFORNIA	Winter Weather	"a spotter in moonri…	0
8	07-Feb-2017 07:00:00	CALIFORNIA	Flood	"major flooding from…	0
9	13-Jan-2017 15:00:00	KANSAS	Ice Storm	"ice accretion was 3…	0
10	23-Jan-2017 00:00:00	NEW YORK	Wind	"a mesonet station …	50

Strings

R2016b

- **For:**
  - Text data
- **Provides:**
  - Improved execution speed and memory usage vs `cellstr`
  - Manipulate, compare, and store text data efficiently
  - Simplified text manipulation functions
  - `pattern` object for text searching in strings

```
>> "image" + (1:3) + ".png"
   1×3 string array
   "image1.png"   "image2.png" "image3.png"
```

Previously:

```
if ~isempty(strfind(textdata,"Dog"))
```

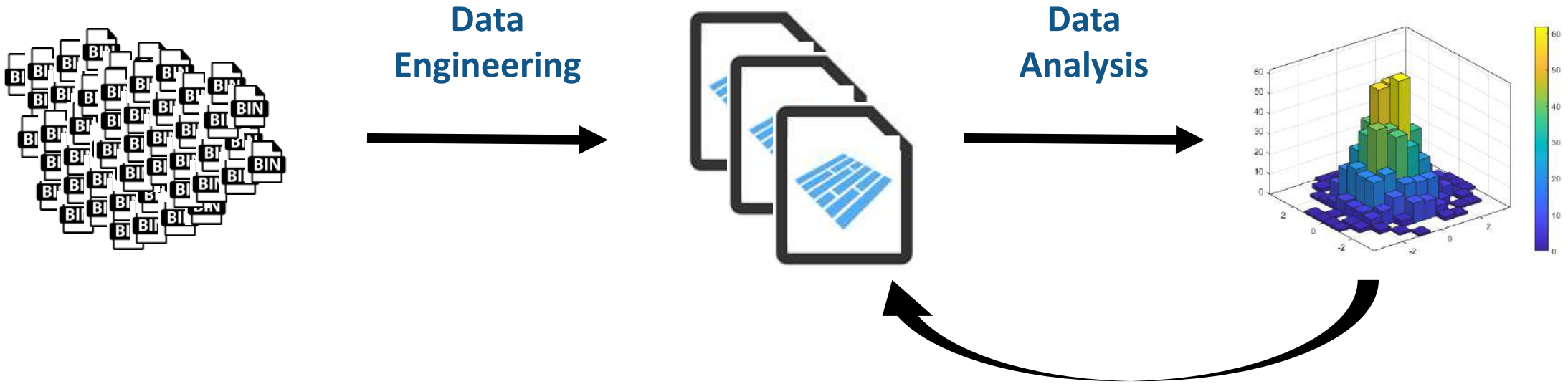
Now:

```
if contains(textdata,"Dog")
```

```
pat = "R" + digitsPattern(4) + ("a"|"b")
str = ["String was introduced in R2016b."
       "Pattern was added in R2020b."];
extract(str,pat)
```

```
ans =
   2x1 string array
     "R2016b"
     "R2020b"
```

Finance Big Data importing and Analysis

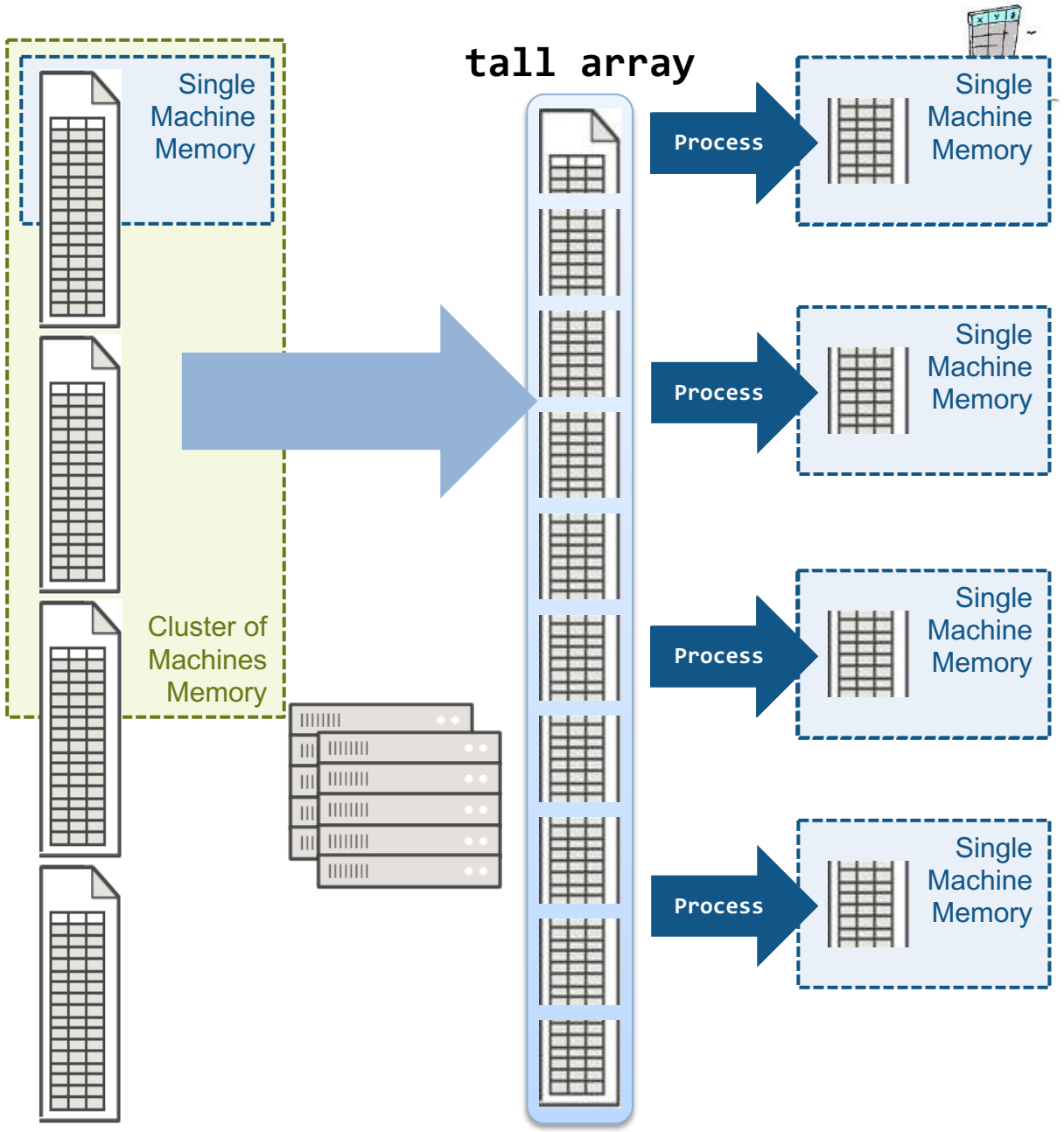



```
datastores()
+
writeall()
```

```
datastores()
+
tall()
```

tall Arrays R2016b

- New data type designed for data that doesn't fit into memory
- Lots of observations (hence "tall")
- Looks like a normal MATLAB array
  - Supports numeric types, tables, datetimes, strings, etc…
  - Supports several hundred functions for basic math, stats, indexing, etc.
- With Parallel Computing Toolbox, process several "chunks" at once
- Can scale up to clusters with MATLAB Parallel Server



Big Data Analysis Without Big Changes

One file

Access Data

```
measured = readtable('PumpData.csv');
measured = table2timetable(measured);
```

Preprocess Data

Select data of interest

```
measured = measured(timerange(seconds(1),seconds(2)),'Speed')
```

Work with missing data

```
measured = fillmissing(measured,'linear');
```

Calculate statistics

```
m = mean(measured.Speed);
s = std(measured.Speed);
```

One hundred files

Access Data

```
measured = datastore('PumpData*.csv');
measured = tall(measured);
measured = table2timetable(measured);
```

Preprocess Data

Select data of interest

```
measured = measured(timerange(seconds(1),seconds(2)),'Speed')
```

Work with missing data

```
measured = fillmissing(measured,'linear');
```

Calculate statistics

```
m = mean(measured.Speed);
s = std(measured.Speed);
```

```
[m,s] = gather(m,s);
```

Agenda

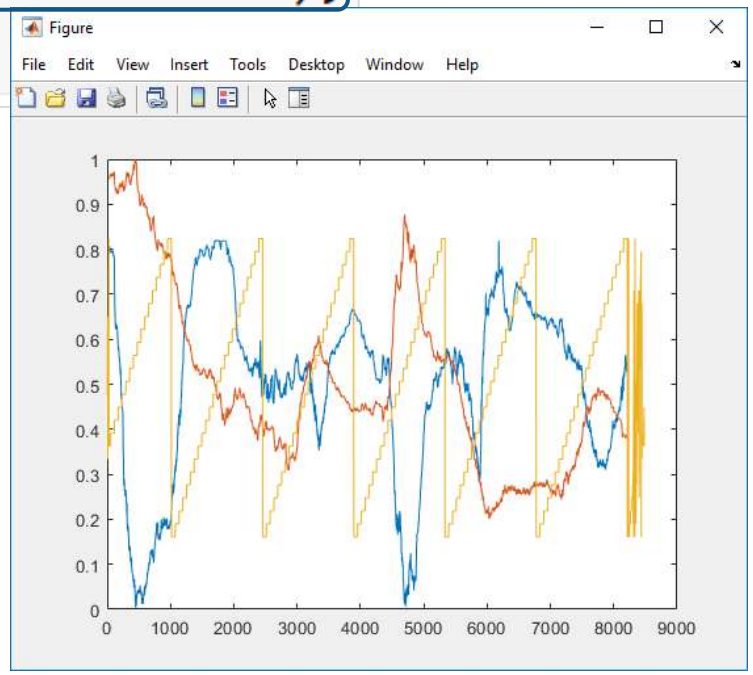
- 다양한 금융 데이터 활용 사례
  - 금융 회사 사례
  - 금융 데이터 연구사례
- 금융 데이터 과학
  - 다양한 금융 데이터 소개
  - MATLAB을 활용한 금융 데이터 과학 기법
  - 금융 데이터 전처리 및 시각화 기능 소개
- 금융 데이터를 활용한 모델링 기법 소개
  - 금융 데이터를 활용한 AI 모델링 기법 소개
  - AI 모델 공유 소개

Preprocessing task for Finance Data

- Synchronize by time
- Find, fill, and remove missing
- Work with outliers
- Data Smoothing
- Normalize, rescale data

Preprocess

```
t = synchronize(t1,t2,t3);
t = fillmissing(t,"linear");
t = rmoutliers(t);
t = smoothdata(t,"movmedian");
t = normalize(t);
```



Preprocessing and analyzing data is easier than ever

Import

```
t1 = readtimetable("s3://bucket_name/file.txt");
```

Preprocess

```
t = synchronize(t1,t2,t3);
t = fillmissing(t,"linear");
t = rmoutliers(t);
t = smoothdata(t,"movmedian");
t = normalize(t);
```

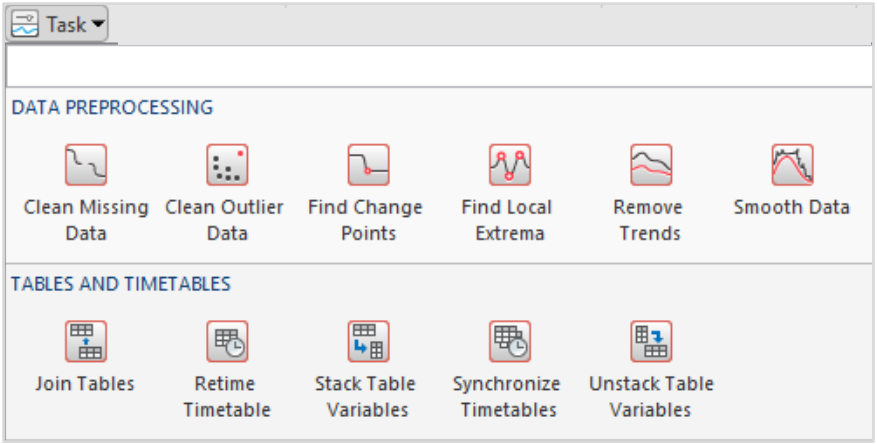
Explore

```
top5 = topkrows(t,5,"RH");
byTime = groupsummary(t,"Time","year","mean");
scaled = grouptransform(t,"State","rescale");
chgpts = ischange(t,"variance","Threshold",20);
```

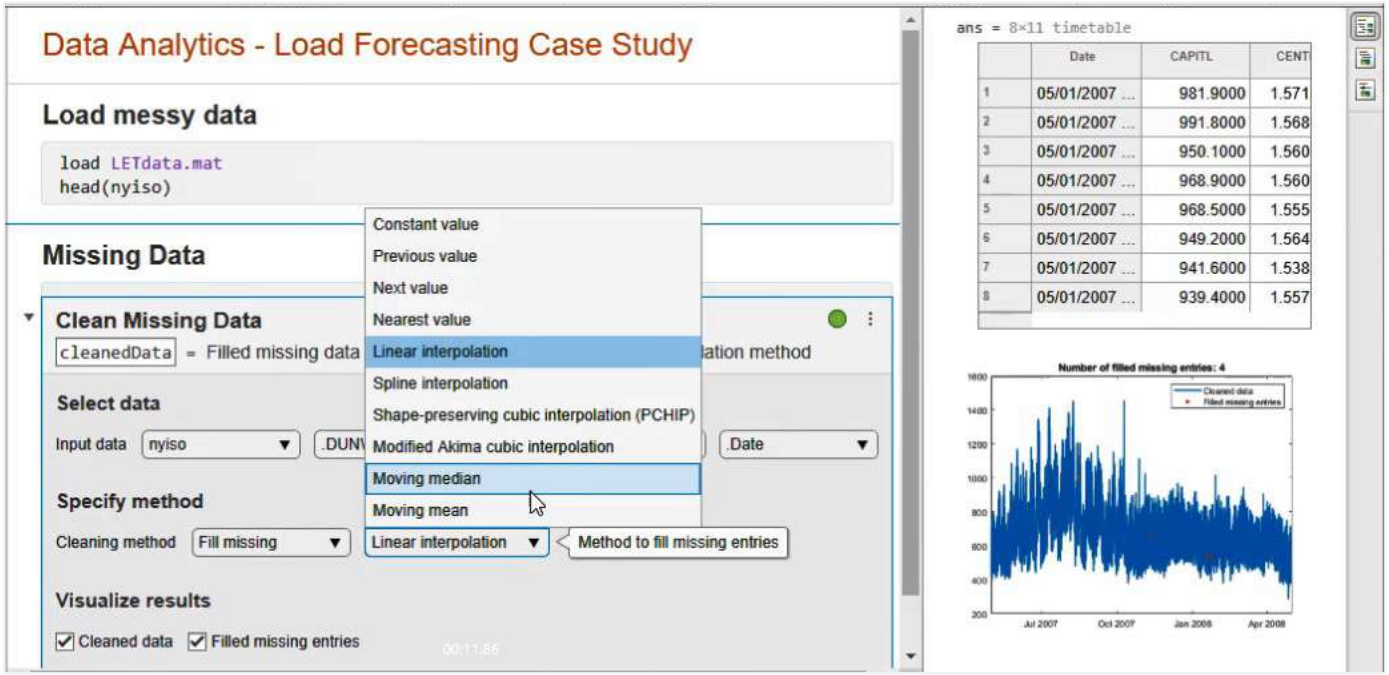
Visualize

```
stackedplot(t);
geoplot(t.Lat,t.Lon,t.RH);
heatmap(t,"State","AQILabel");
scatterhistogram(t.RH,t.DP);
```

Live Editor tasks for preprocessing data and manipulating tables and timetables



- Tasks are apps that can be included in scripts
- Preprocessing tasks allow you to:
  - Interactively explore parameters and options
  - Preview results based on those parameters and options
- Additional tasks for interactively manipulating tables and timetables
- Automatically generate the corresponding MATLAB code
- Save the task as part of the script for subsequent use by others



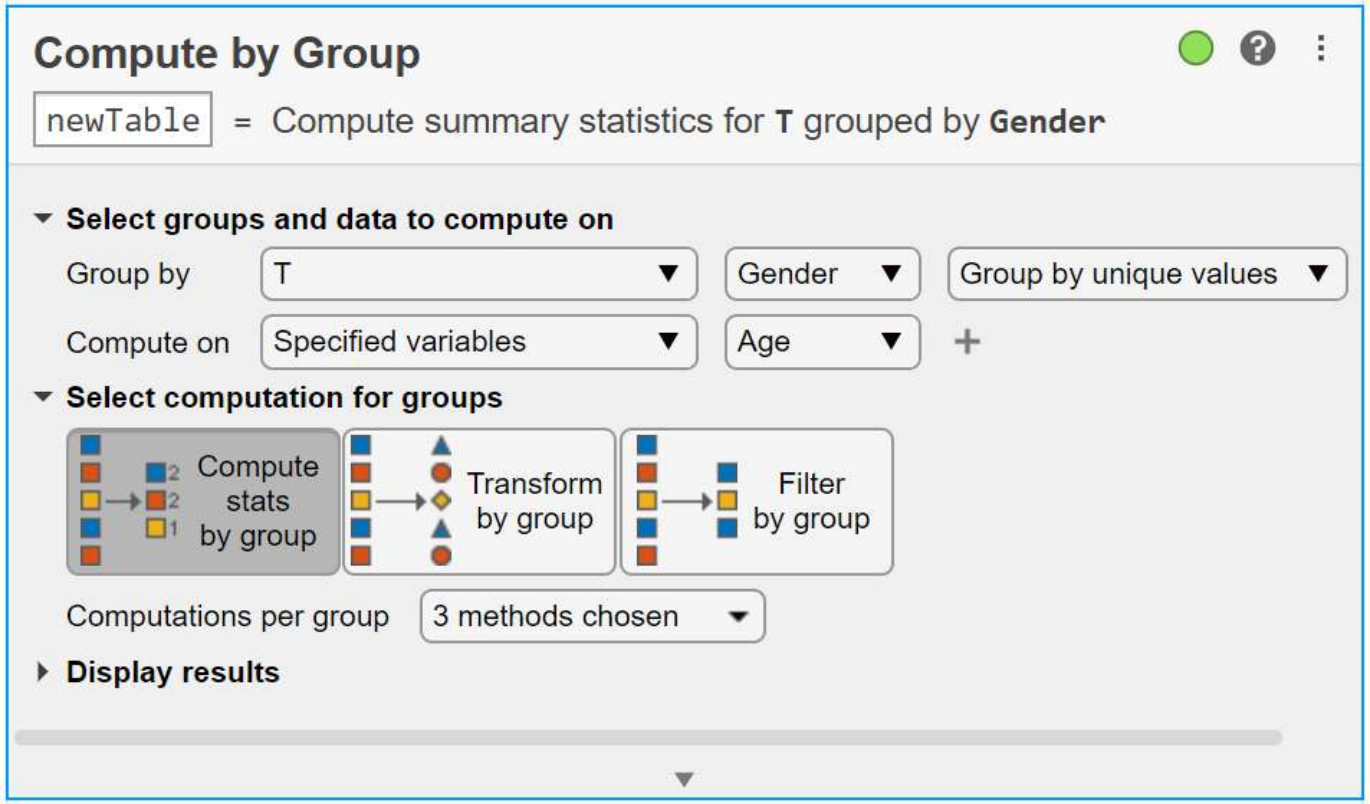
Compute By Group Live Editor Task

Interactively summarize, filter, or transform groups of data

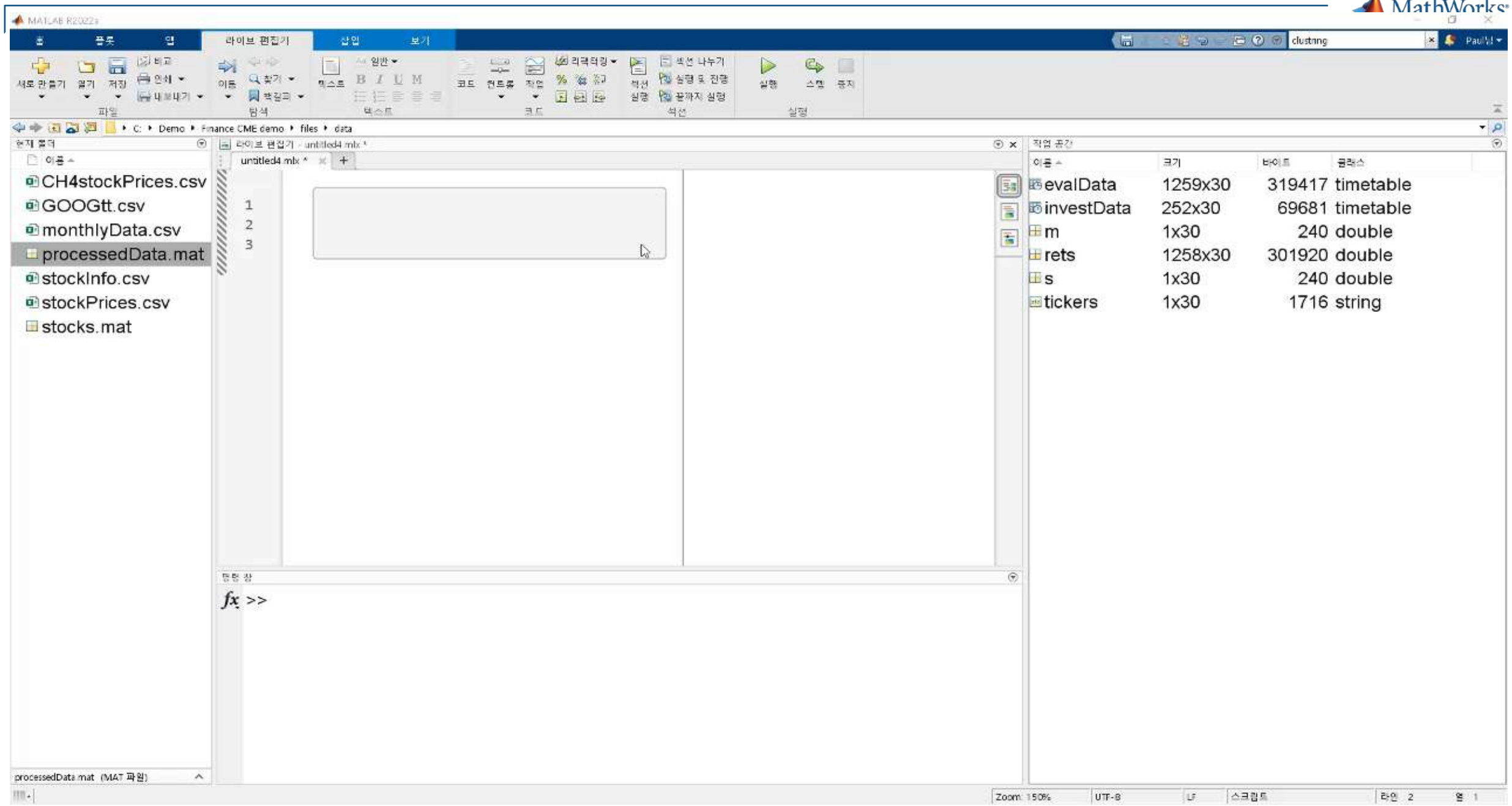
groupsummary
Group summary computations

groupfilter
Filter by group

grouptransform
Transform by group



Compute by Group Live Task


MATLAB R2022a
라이브 편집기
삽입
보기
clustring
C: ▸ Demo ▸ Finance CME demo ▸ files ▸ data
CH4stockPrices.csv
GOOGtt.csv
monthlyData.csv
processedData.mat
stockInfo.csv
stockPrices.csv
stocks.mat
processedData.mat (MAT 파일)
untitled4.mlx
evalData 1259x30 319417 timetable
investData 252x30 69681 timetable
m 1x30 240 double
rets 1258x30 301920 double
s 1x30 240 double
tickers 1x30 1716 string
>>

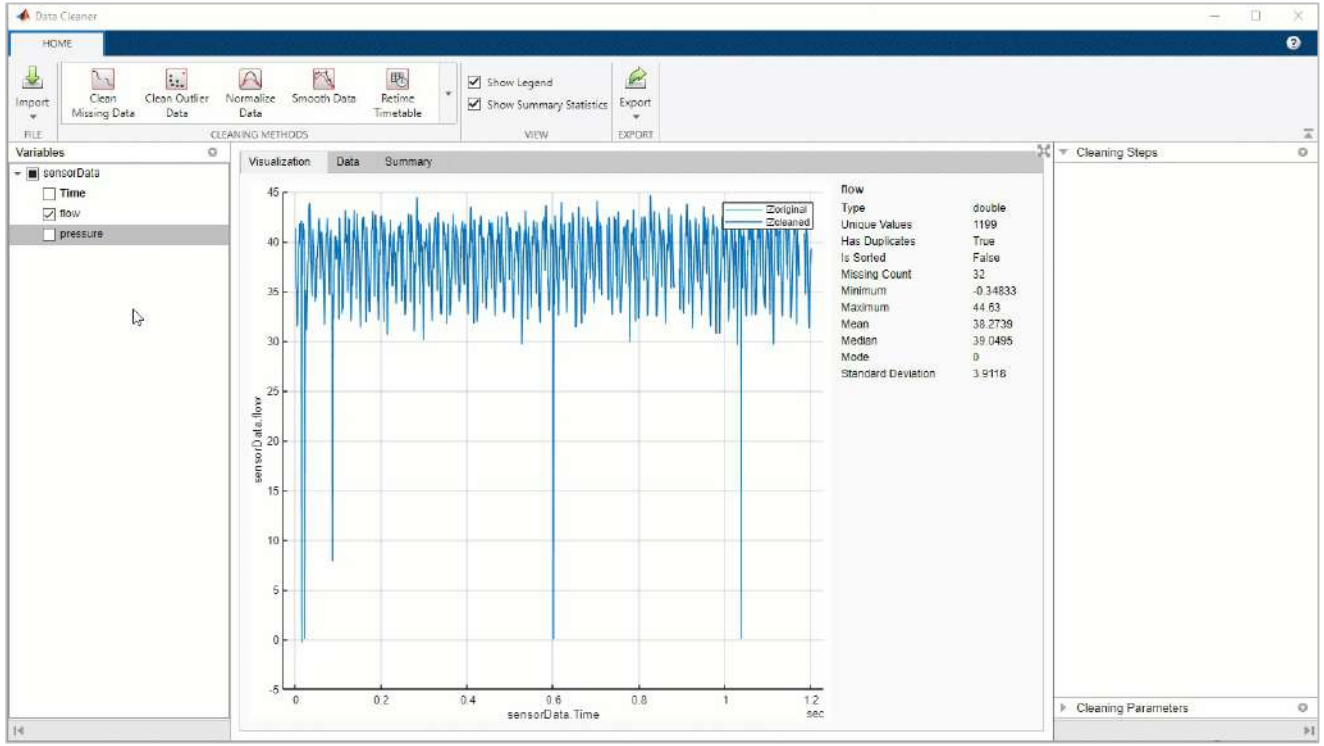
Interactively preprocess and organize column-oriented data

R2022a

- Get insights on data impurities
- Interactively clean, verify and adjust
- Export to workspace or generate MATLAB code

- Works on time-based signals in timetables

- **Supports:**
  - Missing data
  - Outlier data
  - Normalize and smoothing
  - Retiming
  - Stacking and unstacking

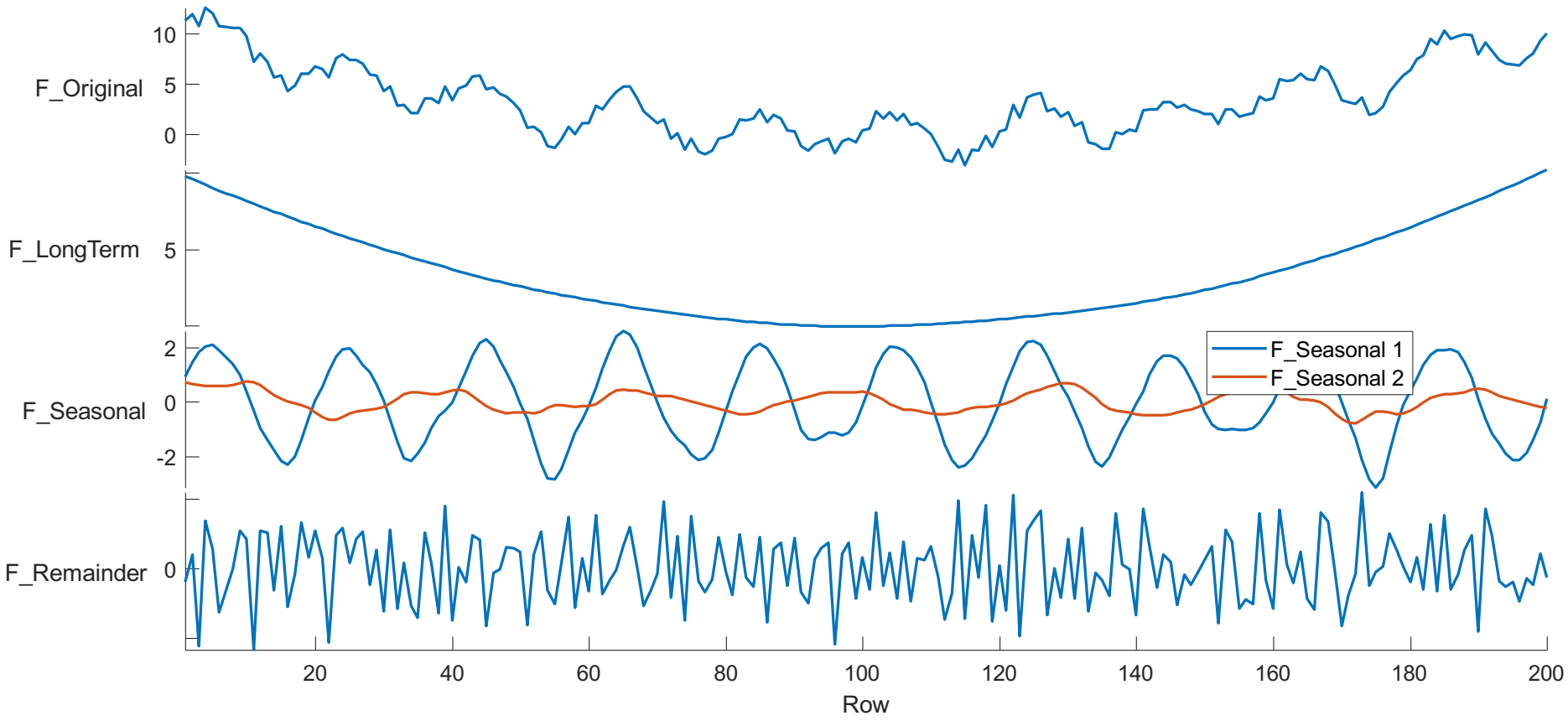


Data Cleaner

https://www.mathworks.com/help/matlab/ref/datacleaner-app.html

trenddecomp

Decompose data into long-term and seasonal trends



Signal

Long term trend

Seasonal trends

Remainder

MATLAB R2022a

C: ▸ Users ▸ pauljang ▸ OneDrive - MathWorks ▸ Documents ▸ MATLAB ▸ Examples ▸ R2022a ▸ matlab ▸ DecomposeVectorOfDataExample

untitled4.mlx * | DecomposeVectorOfDataExample.mlx

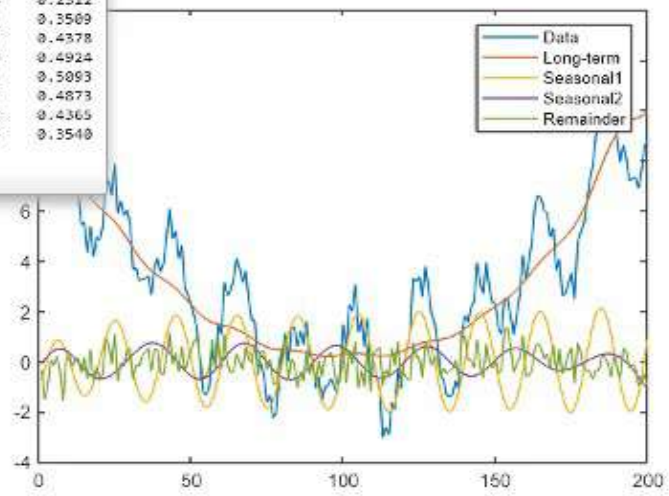
Create a vector of data A that contains a long-term trend, two seasonal trends with different periods, and a noise component.

```
t = (1:200)';
trend = 0.001*(t-100).^2;
period1 = 20;
period2 = 30;
seasonal1 = 2*sin(2*pi*t/period1);
seasonal2 = 0.75*sin(2*pi*t/period2);
noise = 2*(rand(200,1) - 0.5);
A = trend + seasonal1 + seasonal2 + noise;
```

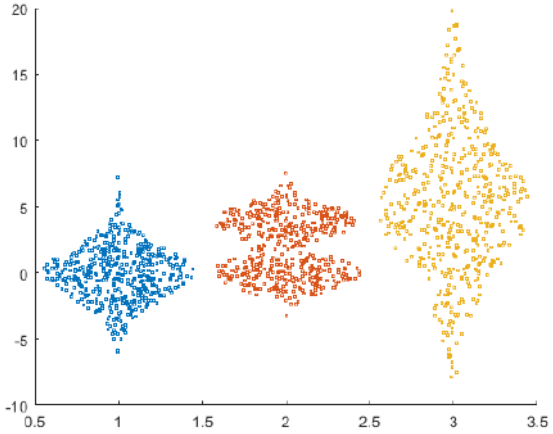
Decompose the data using the SSA algorithm. Plot the data, the trends, and the remainder.

```
[LT,ST,R] = trenddecomp(A);
plot([A LT ST R]);
legend("Data","Long-term","Seasonal1","Seasonal2","Remainder")
```

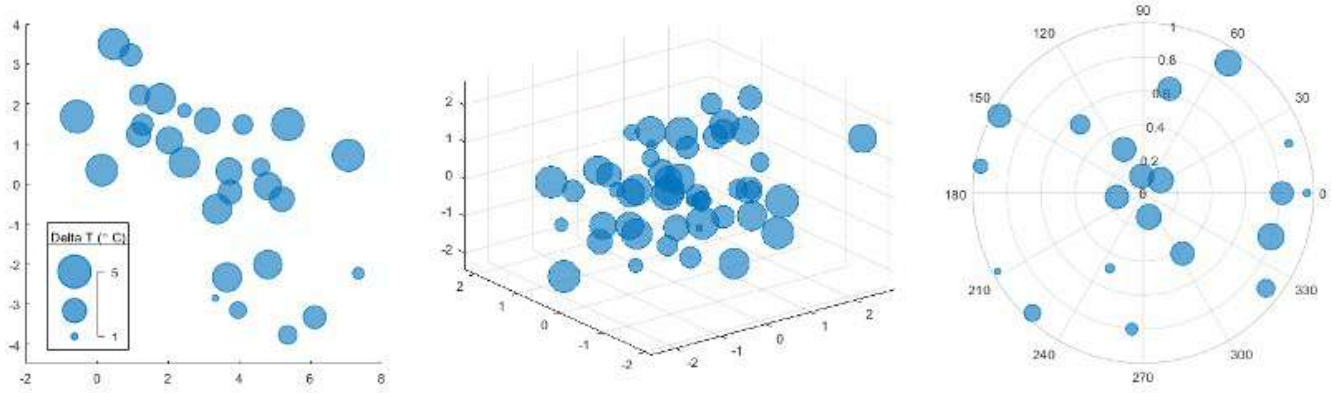
```
ST = 200×2
   -0.7430   -0.0723
   -0.2965    0.0833
    0.1434    0.2312
    0.5131    0.3509
    0.7659    0.4378
    0.8770    0.4924
    0.8401    0.5093
    0.6691    0.4873
    0.3915    0.4365
    0.0346    0.3540
       ⋮
```



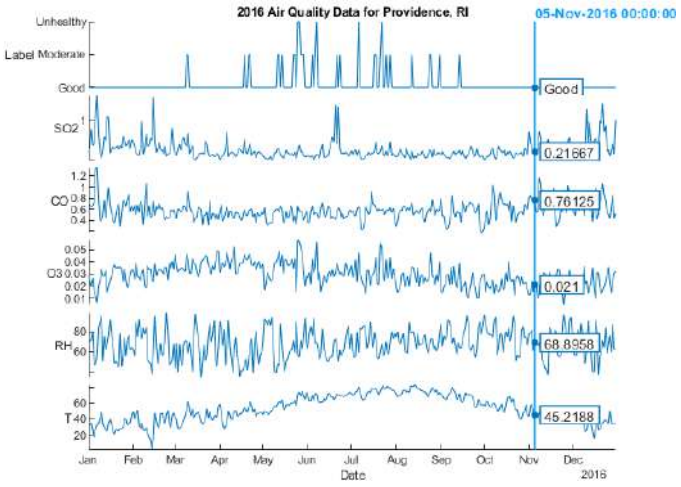
Charts and visualizations for Finance Data



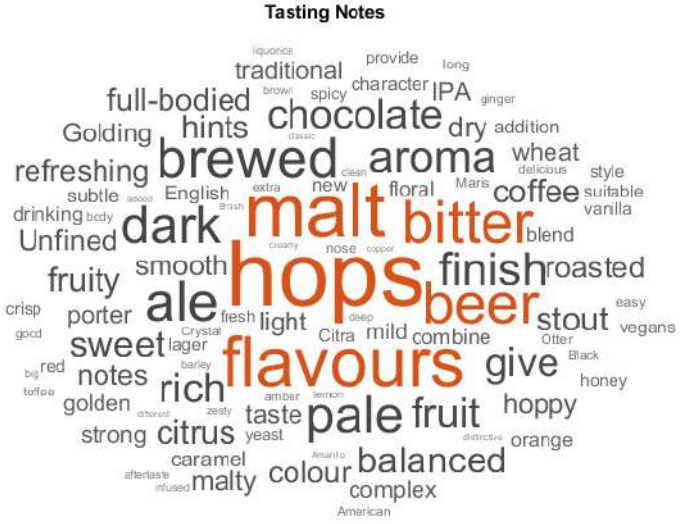
swarmchart



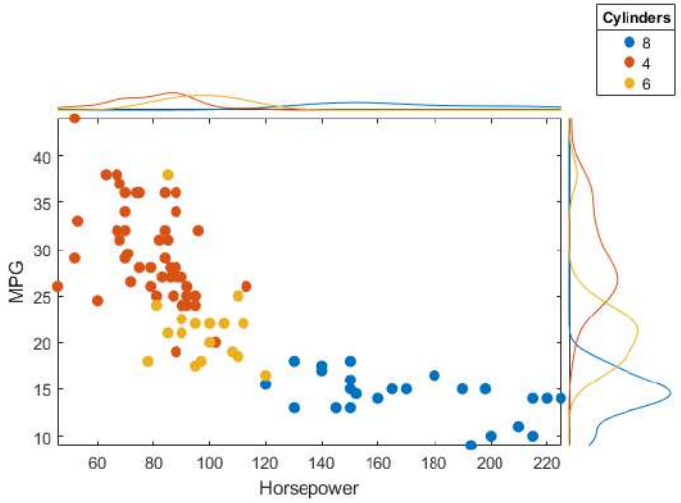
bubblechart, bubblechart3, polarbubblechart



stackedplot

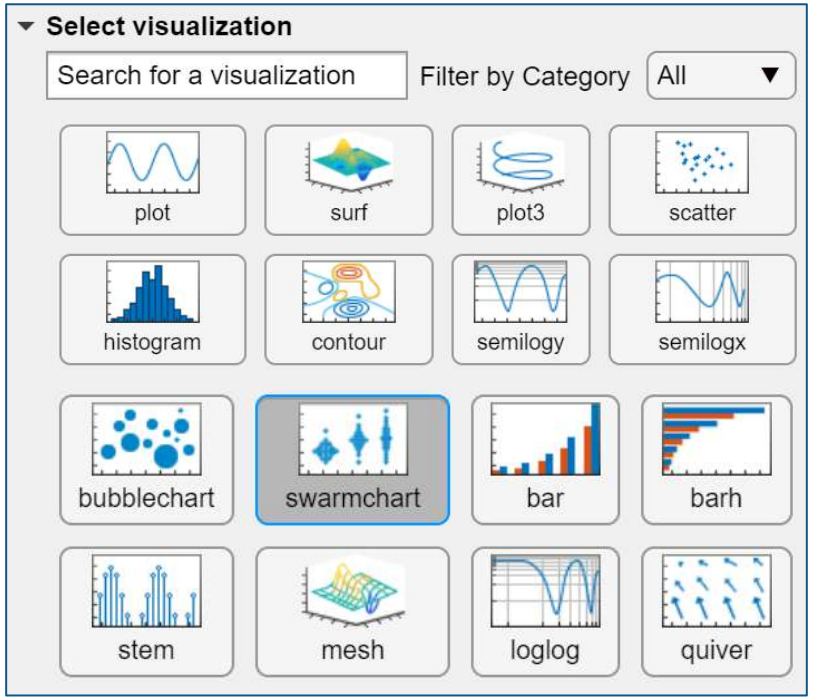


wordcloud

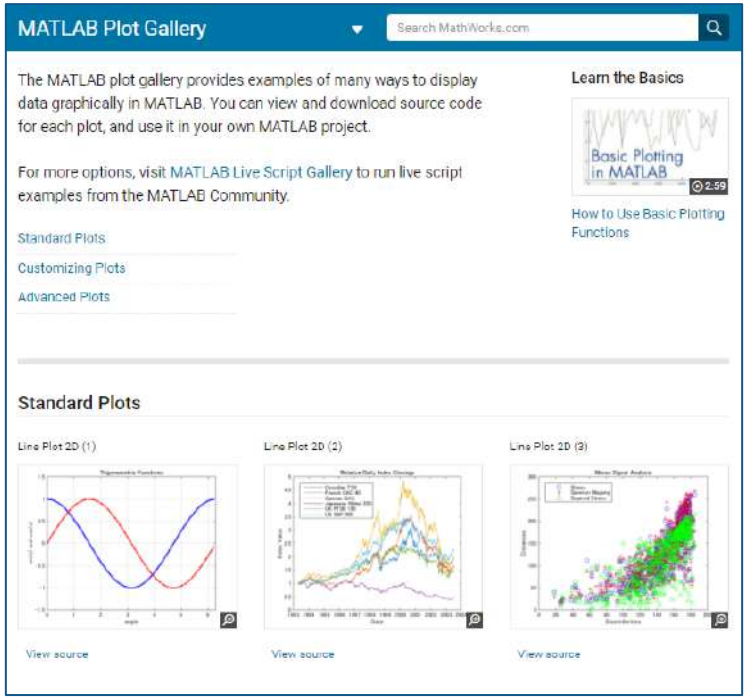


scatterhistogram

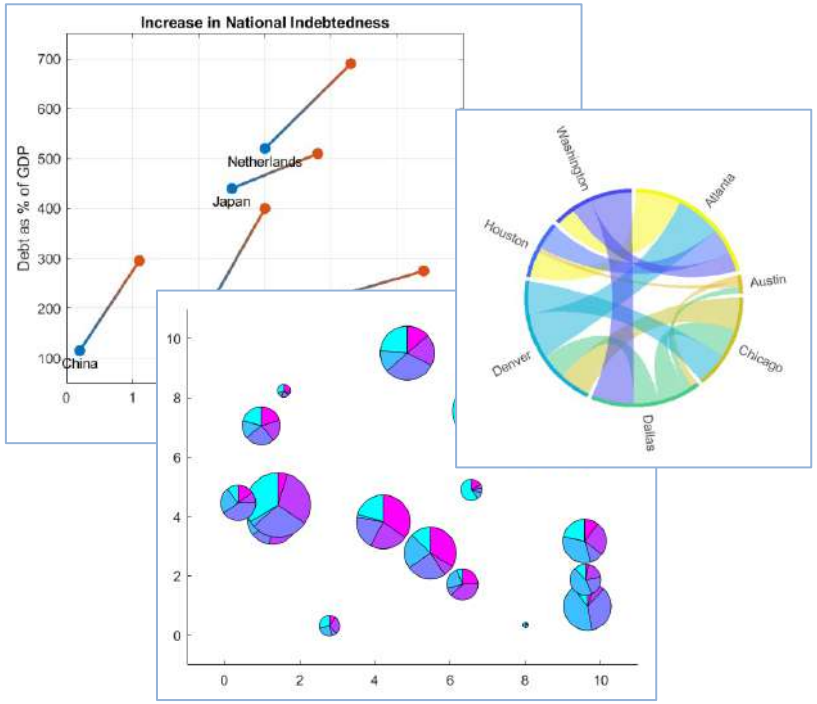
Finance Data Visualization



Create Plot Live Task



MATLAB Plot Gallery



Community Contributions

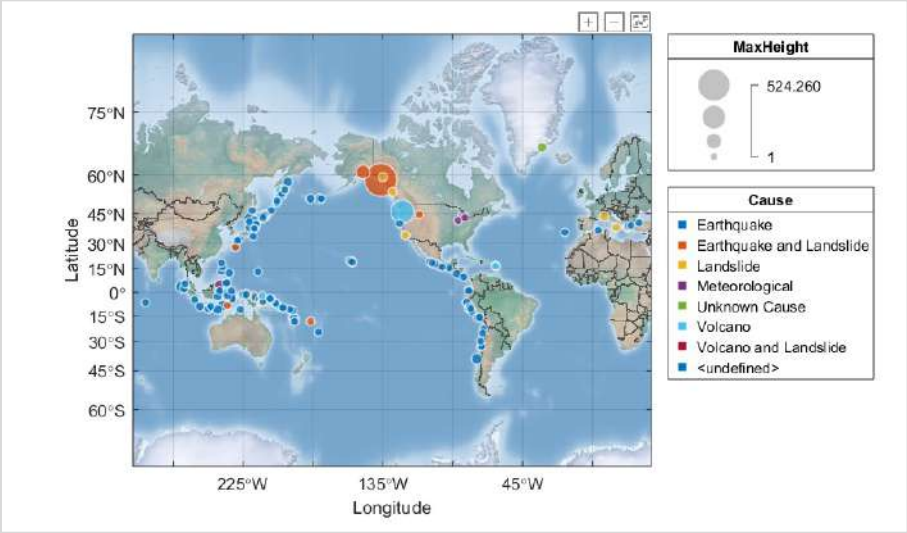
Custom Visualizations using Chart Authoring Framework

https://github.com/MATLAB-Graphics-and-App-Building
https://www.mathworks.com/matlabcentral/fileexchange?q=Tag%3Achartcontainer

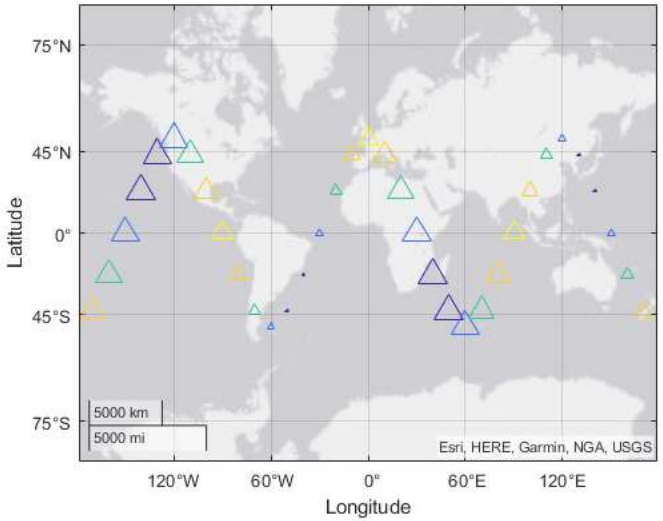
Geographic Plots

Visualize data on maps

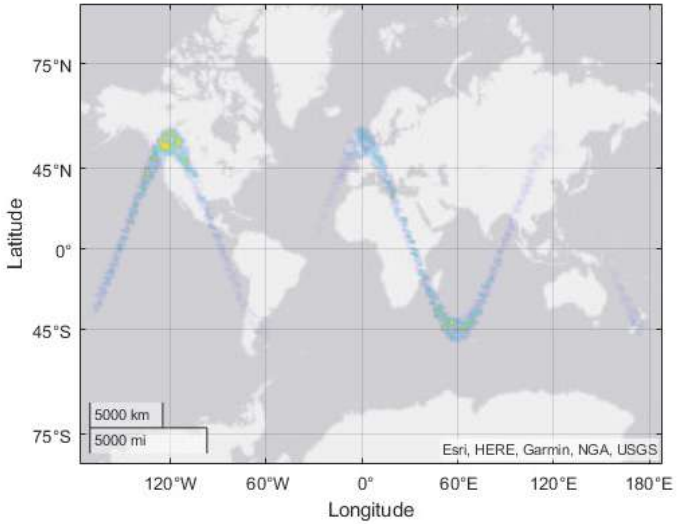
- Visualize latitude and longitude data over interactive maps
- Create line, scatter, point density, and bubble plots over maps
- Customize map imagery using basemaps



geobubble



geoscatter

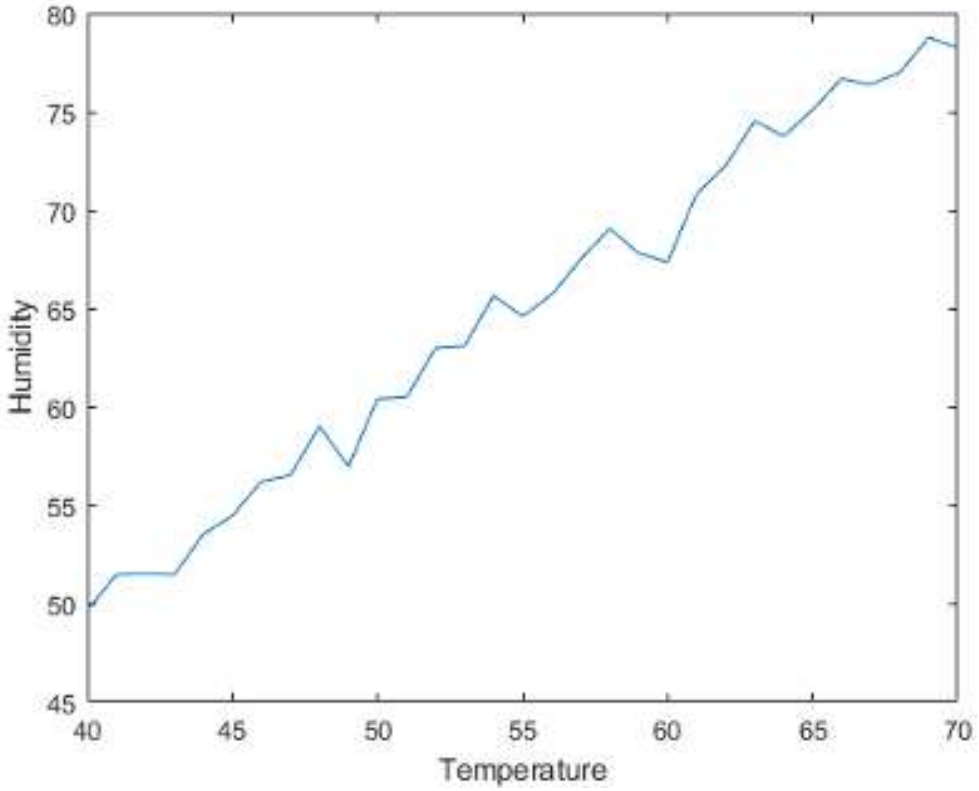


geodensity

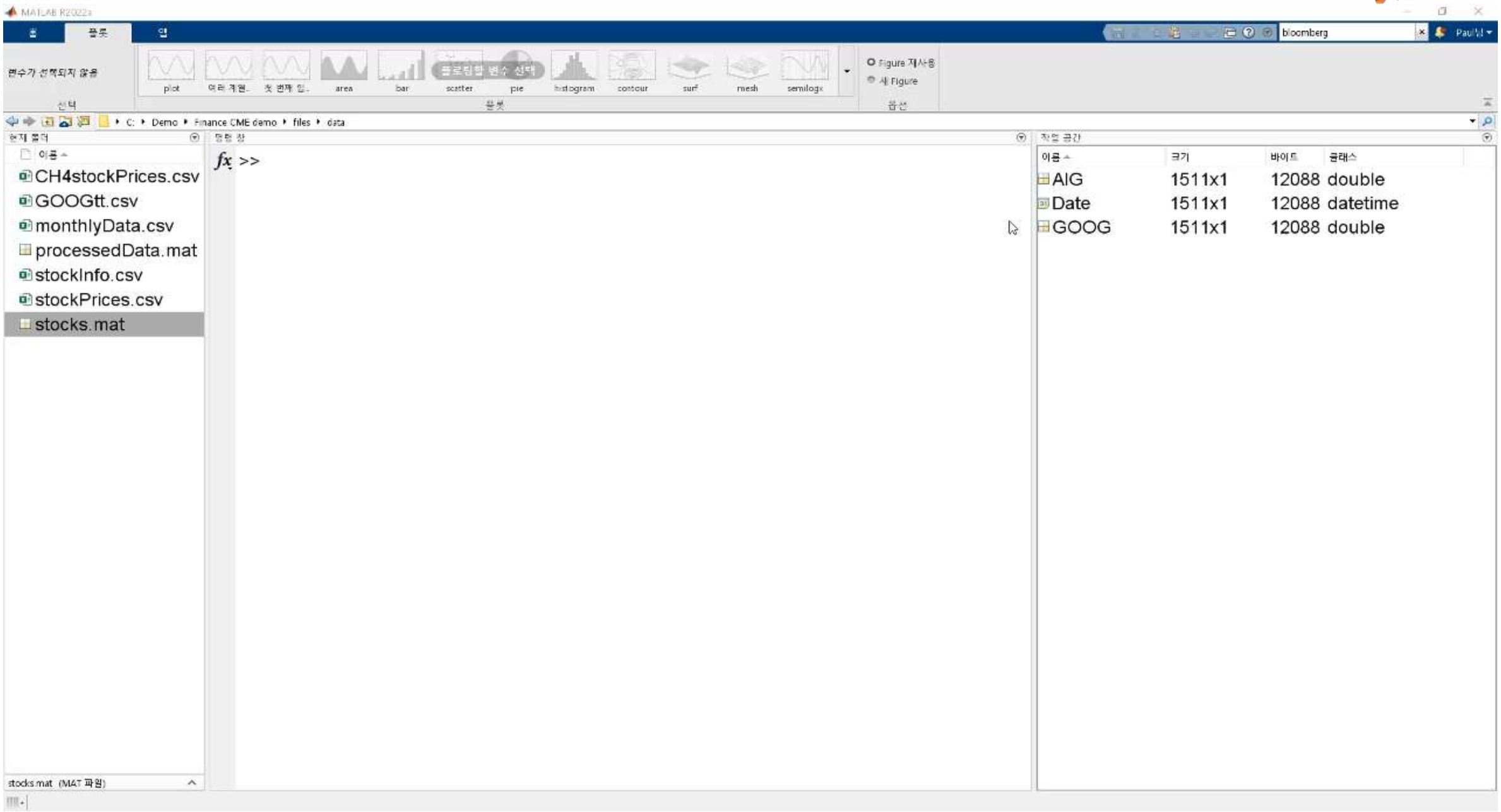
https://www.mathworks.com/help/matlab/geographic-plots.html

Pass tables directly to plotting functions

- Functions supported include:
  - **`plot`**, **`plot3`**, **`polarplot`**, and other line plotting functions (R2022a)
  - Scatter plots, bubble charts, and swarm charts (R2021b)
- Axis labels and legend (if present) are automatically added
- Default row labels of data tips are the names of the table variables associated with the data point




```
T = table(Temperature,Humidity);
plot(T,"Humidity","Temperature")
```

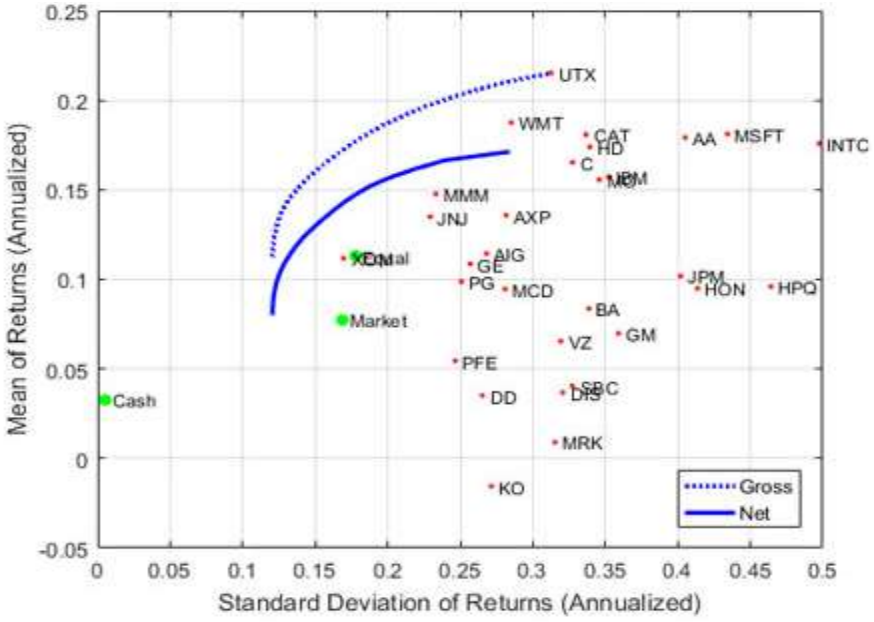

MATLAB R2022a
플롯
앱
bloomberg
변수가 선택되지 않음
plot
area
bar
scatter
pie
histogram
contour
surf
mesh
semilogx
플롯
Figure 재사용
새 Figure
옵션
C: ▸ Demo ▸ Finance CME demo ▸ files ▸ data
현재 폴더
이름
CH4stockPrices.csv
GOOGtt.csv
monthlyData.csv
processedData.mat
stockInfo.csv
stockPrices.csv
stocks.mat
stocks.mat (MAT 파일)
명령 창
>>
작업 공간
이름
크기
바이트
클래스
AIG 1511x1 12088 double
Date 1511x1 12088 datetime
GOOG 1511x1 12088 double

Agenda

- 다양한 금융 데이터 활용 사례
  - 금융 회사 사례
  - 금융 데이터 연구사례
- 금융 데이터 과학
  - 다양한 금융 데이터 소개
  - MATLAB을 활용한 금융 데이터 과학 기법
  - 금융 데이터 전처리 및 시각화 기능 소개
- 금융 데이터를 활용한 모델링 기법 소개
  - 금융 데이터를 활용한 AI 모델링 기법 소개
  - AI 모델 공유 소개

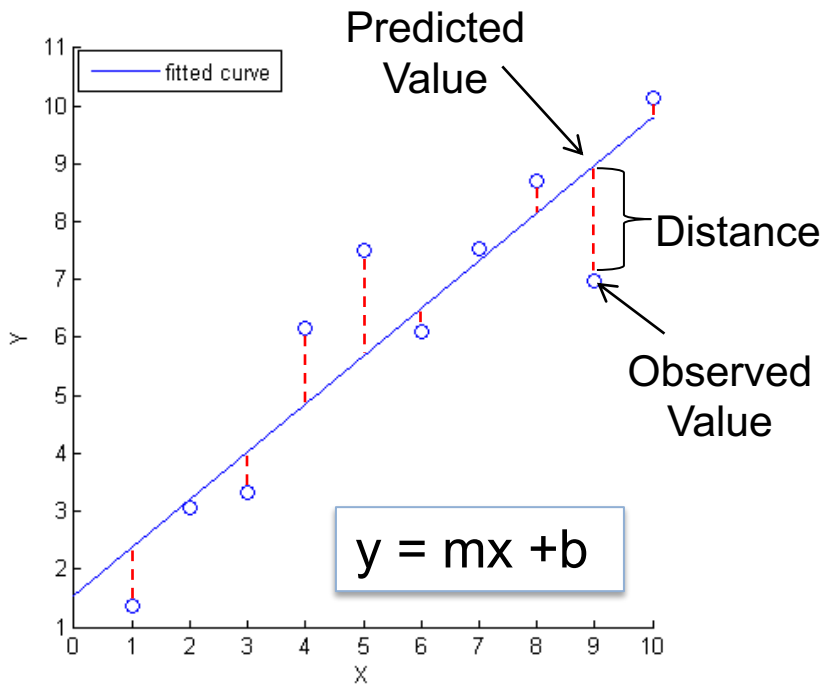
AI Applications in Finance

- Algorithmic trading
- Portfolio optimization
- Trade execution
- Market-making



What is Regression?

- Type of predictive modeling
- Specify a model that describes Y as a function of X
- Estimate a set of coefficients that minimizes the difference between predicted and actual
- Typically, minimize the sum of the squared errors (the sum of the squared residuals)



Regression Techniques

- Regression techniques require that the user specify a model.
- Model specification describes the dynamics of the system.

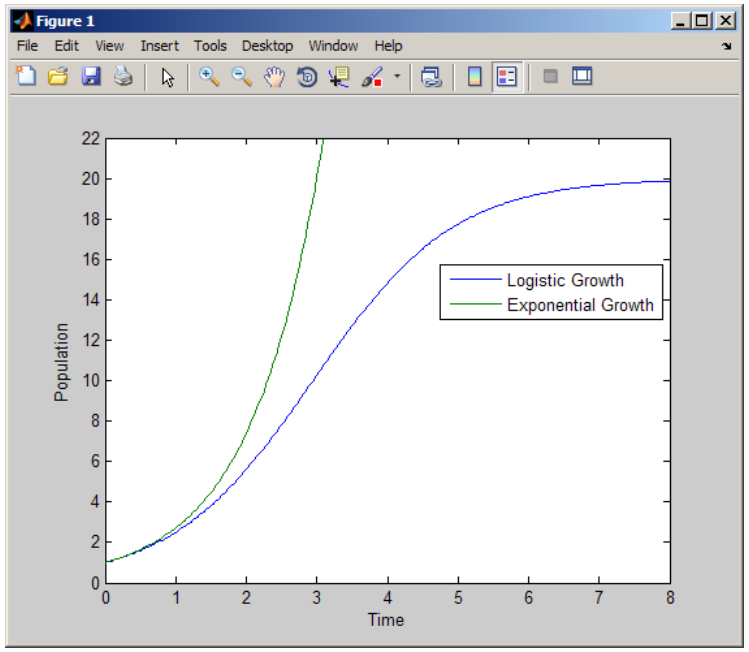
Example - population models

Logistic Growth

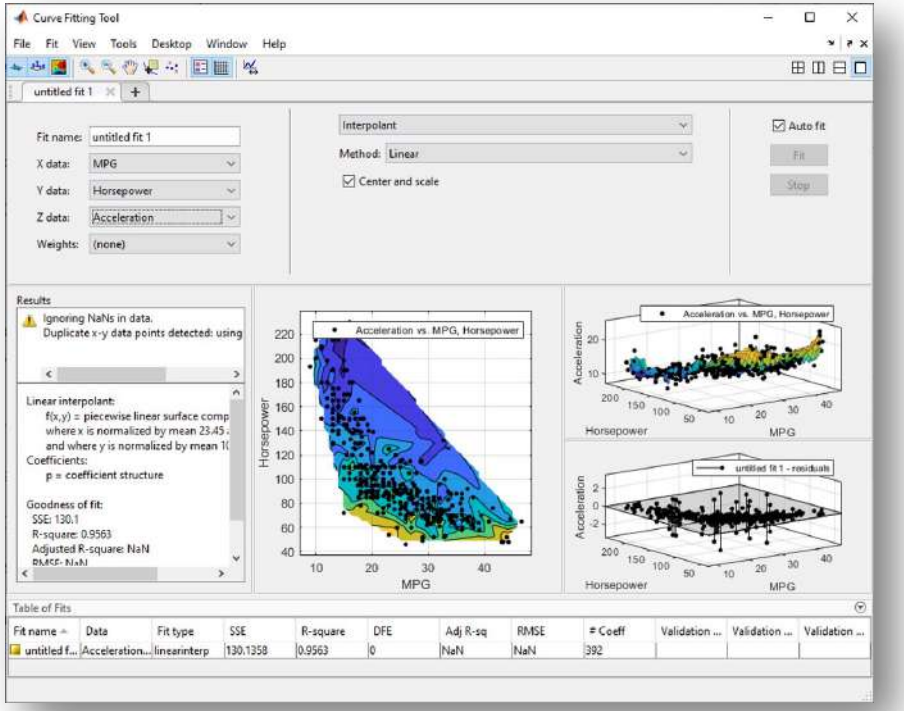
$$N_t = \frac{N_0 \times K}{N_0 + (K - N_0) \times \exp(-r_0 \times t)}$$

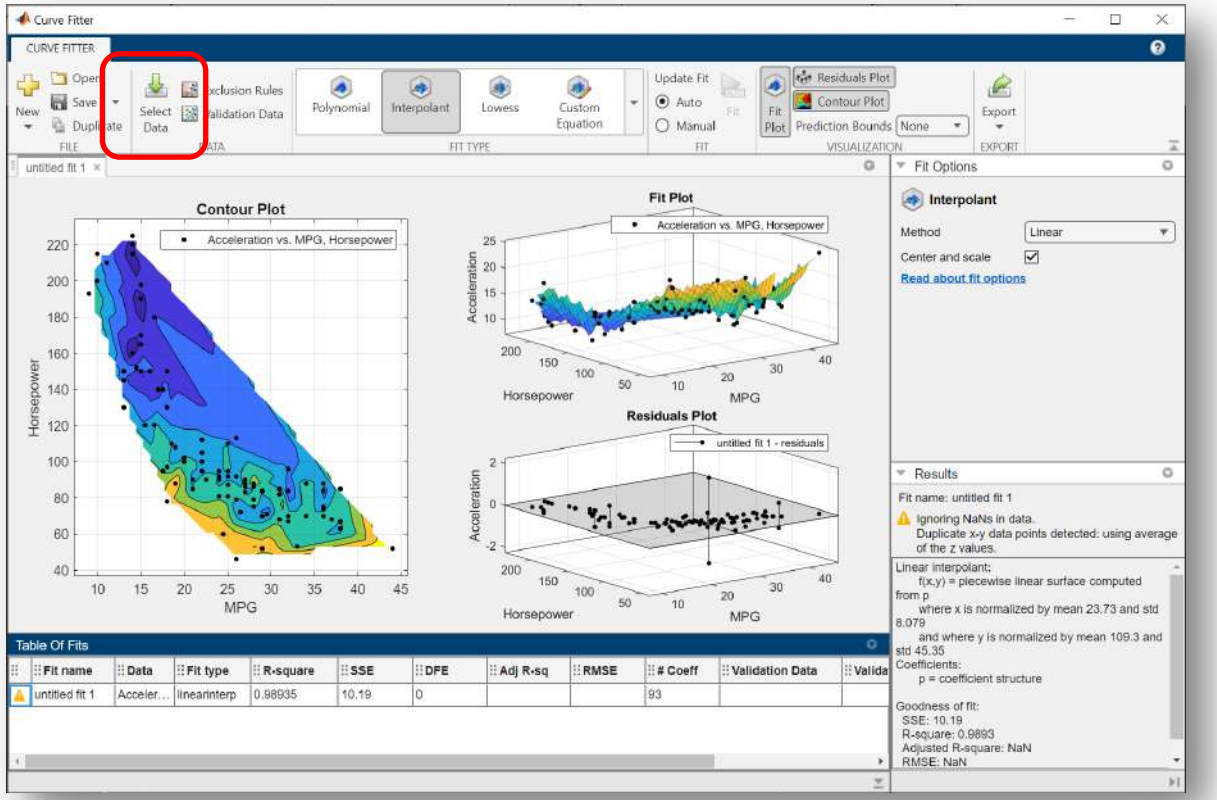
Exponential Growth

$$N_t = N_0 \times e^{(r \times t)}$$



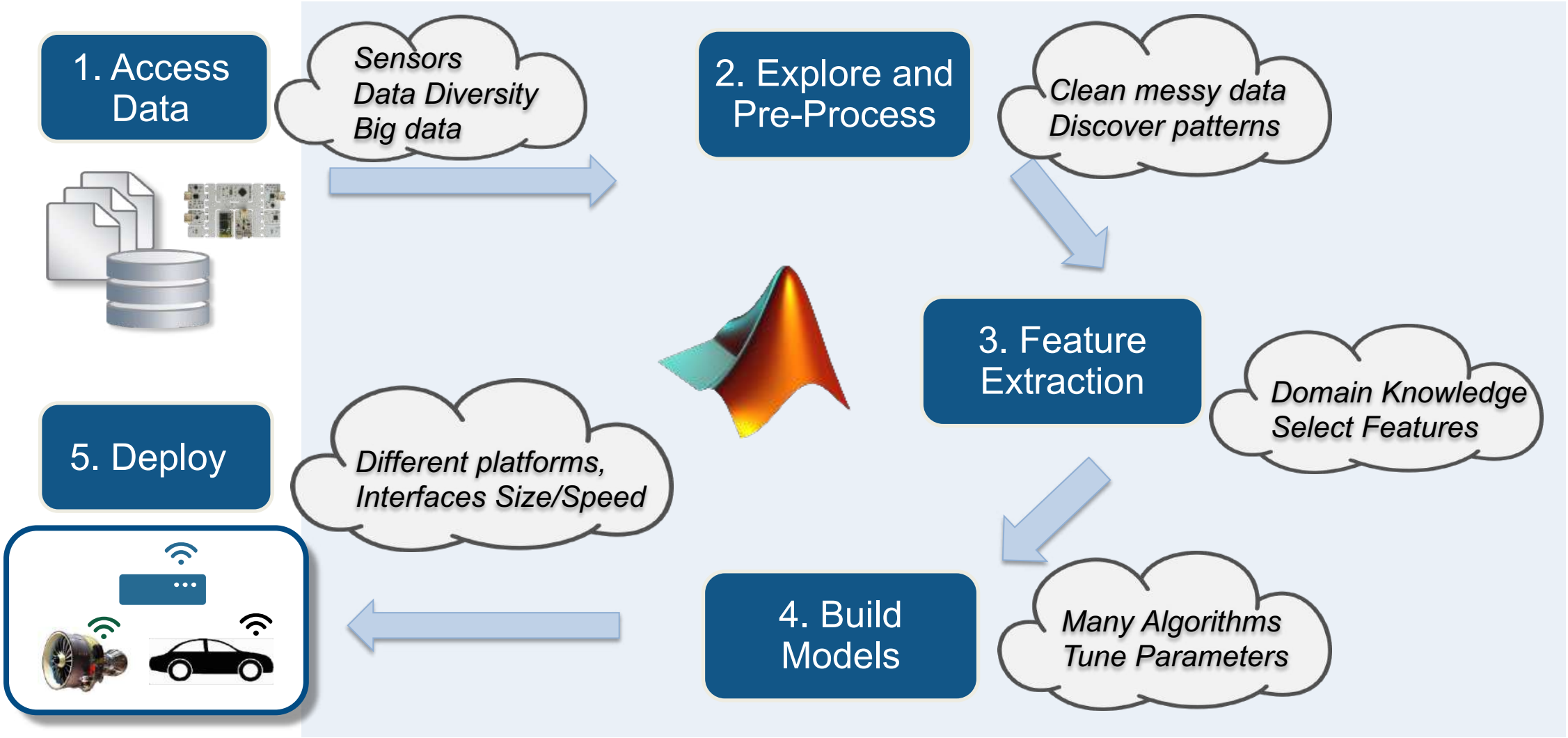
Curve Fitter now implemented as AppContainer-based app





With `table` support!

Financial Modeling Workflow with Machine Learning



MATLAB as a Platform for Finance Modeling

- MATLAB can support traditional modeling as well as your advanced analytics work

- Consolidated platform for:
  - Optimization
  - Regression
  - Time Series Modeling
  - Machine Learning
  - Deep Learning
  - Natural Language Processing
  - Reinforcement Learning

Math, Statistics, and Optimization

Curve Fitting Toolbox
Deep Learning Toolbox
Global Optimization Toolbox
Optimization Toolbox
Partial Differential Equation Toolbox
Statistics and Machine Learning Toolbox
Symbolic Math Toolbox
Text Analytics Toolbox

Data Science and Deep Learning

Deep Learning Toolbox
Predictive Maintenance Toolbox
Reinforcement Learning Toolbox
Statistics and Machine Learning Toolbox
Text Analytics Toolbox

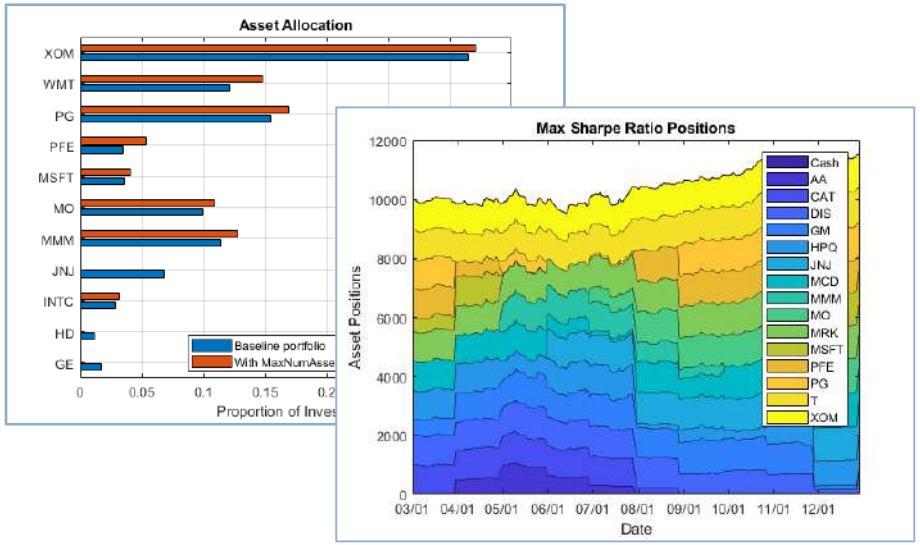
Computational Finance

Database Toolbox
Datafeed Toolbox
Econometrics Toolbox
Financial Instruments Toolbox
Financial Toolbox
Risk Management Toolbox
Spreadsheet Link
Trading Toolbox

Asset Management and Instrument Pricing

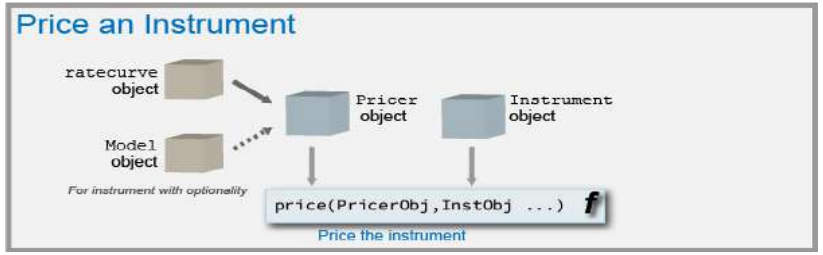
Analyze financial data and develop financial models

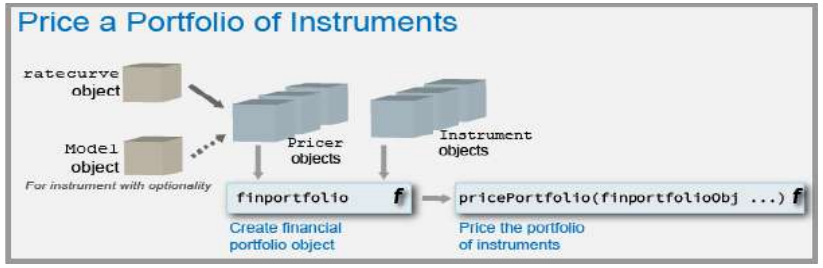
- New investment strategy backtest framework
- Portfolio Optimization
  - Manage a risk parity portfolio
  - Integrate Environmental, Social, Governance (ESG) criteria



Design, price, and hedge complex financial instruments

- Object-based framework for pricing financial instruments
- Interest-rate instruments for alternative reference rate (ARR) securities
- Support for pricing inflation derivatives like inflation swaps
- Price complex options using Deep Learning





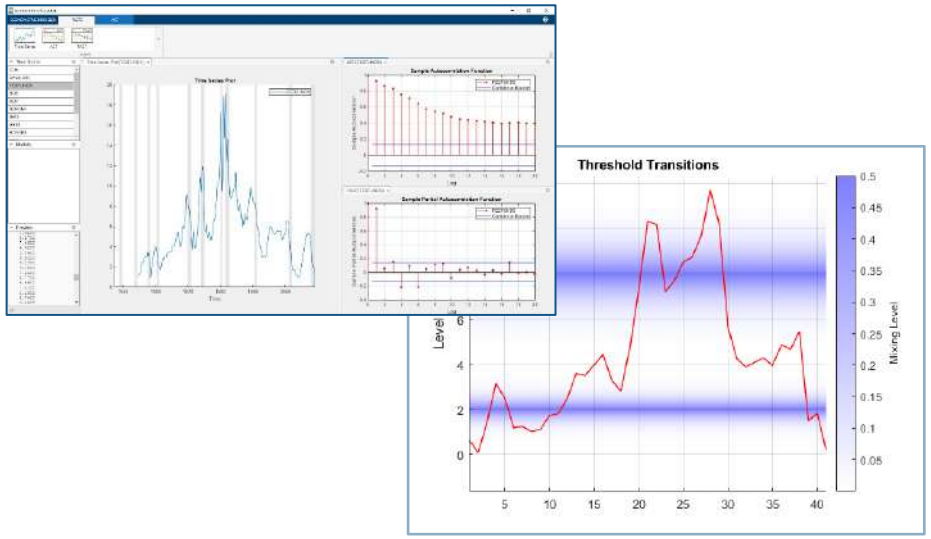
Risk Management and Time Series Modeling

Develop Credit Risk and Market Risk models

- Lifetime expected credit loss (ECL) calculator at individual and portfolio level
- Validation metrics and plots for consumer and corporate credit risk models
- Responsible AI: Evaluate fairness of Credit Scoring models
- Insurance: Unpaid claims estimation techniques

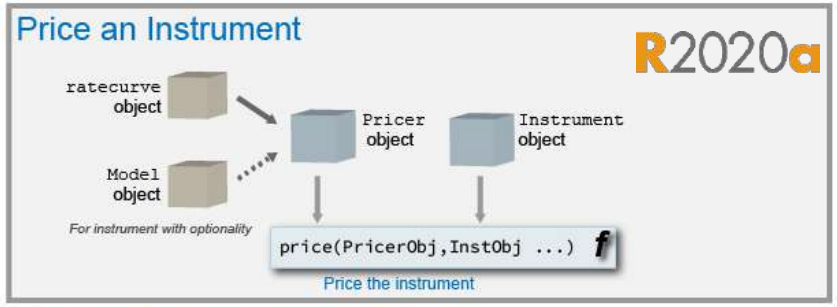
Analyze and model time series data

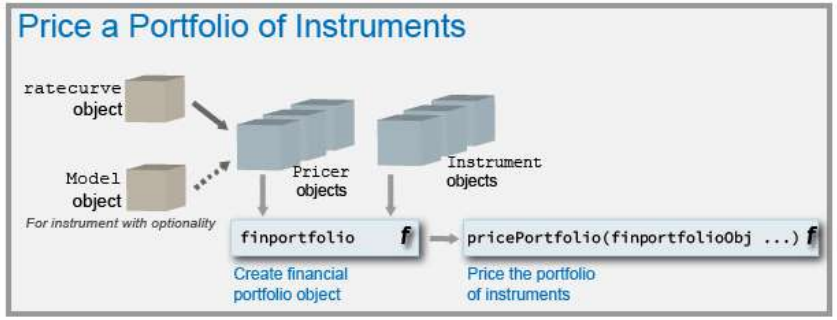
- Multivariate time series modeling support in the Econometric Modeler app
- Model time series in the presence of regime-switches



Design, Price, and Hedge Complex Financial Instruments

- **Object-based framework for pricing financial instruments**
- Monte Carlo Simulation pricing supported for pricing equity, FX, and Commodity instruments
- Construct inflation curves and price inflation bonds and swaps
- Vanna-Volga Method: Vanilla, barrier, double barrier, One-touch and double one-touch options
- Price Variance swaps
- Double Barrier options with Closed Form and Finite Differences



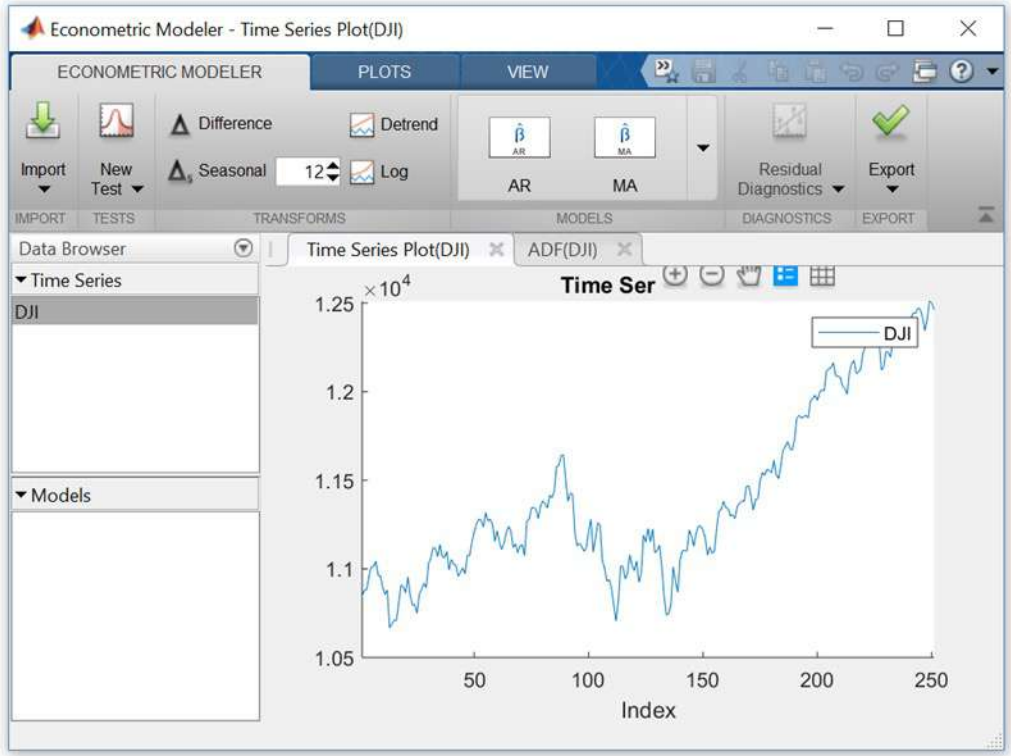


Pricing Interest-rate options with negative

Normal Volatility Model (Bachelier Model)	Shifted Black Model	Shifted SABR Model
European swaption (swaptionbynormal)	European swaption (swaptionbyblk)	Implied Black volatility (blackvolbysabr)
Cap (capbynormal)	Cap (capbyblk)	Option sensitivity (optsensbysabr)
Floor (floorbynormal)	Floor (floorbyblk)	
Caplet volatilities (capvolstrip)		
Floorlet volatilities (floorvolstrip)		

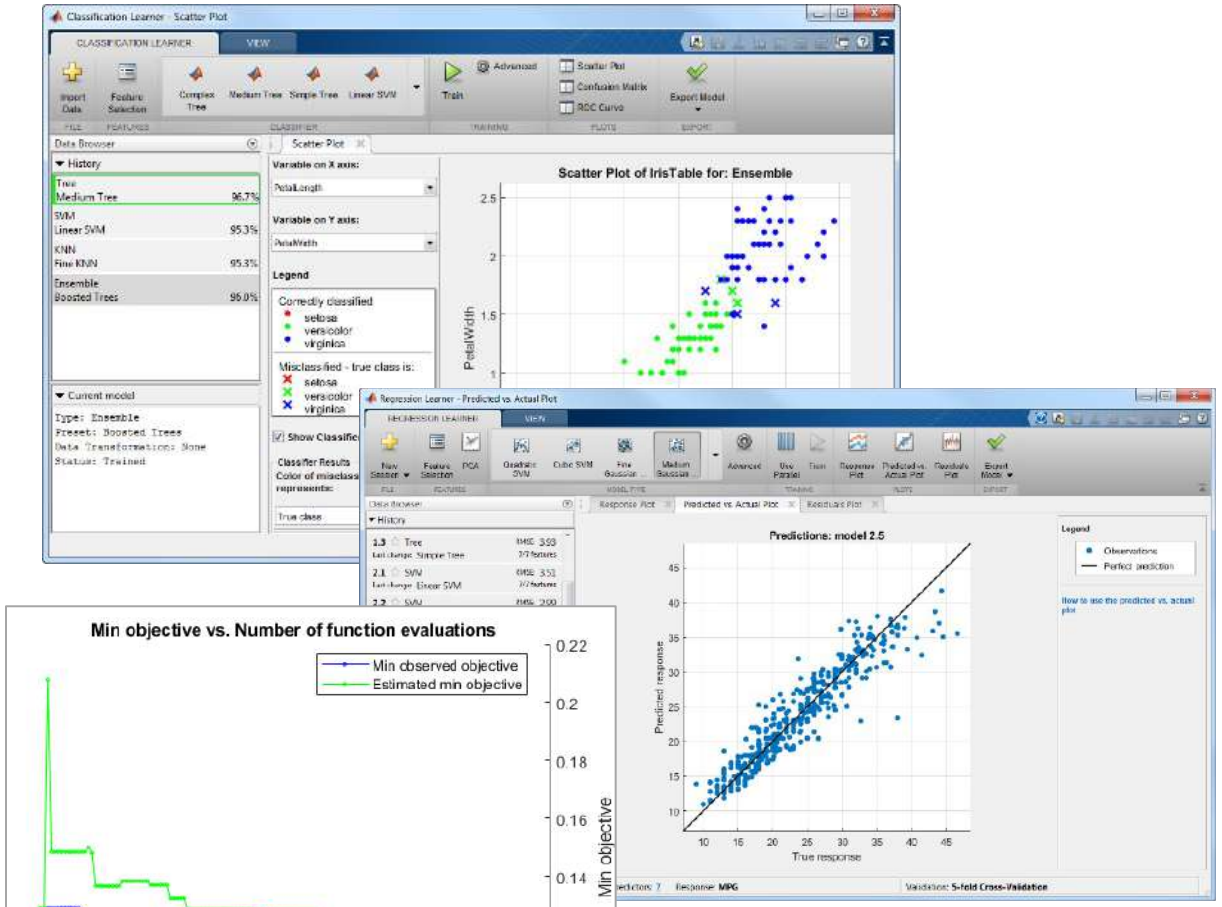
Interactive Econometric Modeling and Reporting

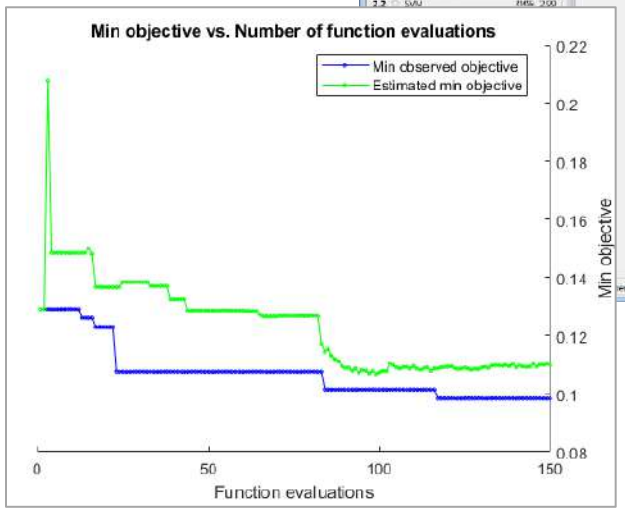
- Econometric Modeler App
- Popular econometric models, tests, and visualizations available as interactive workflow
- Document your entire model development workflow with one click



Train, Compare, and Optimize Machine Learning Models

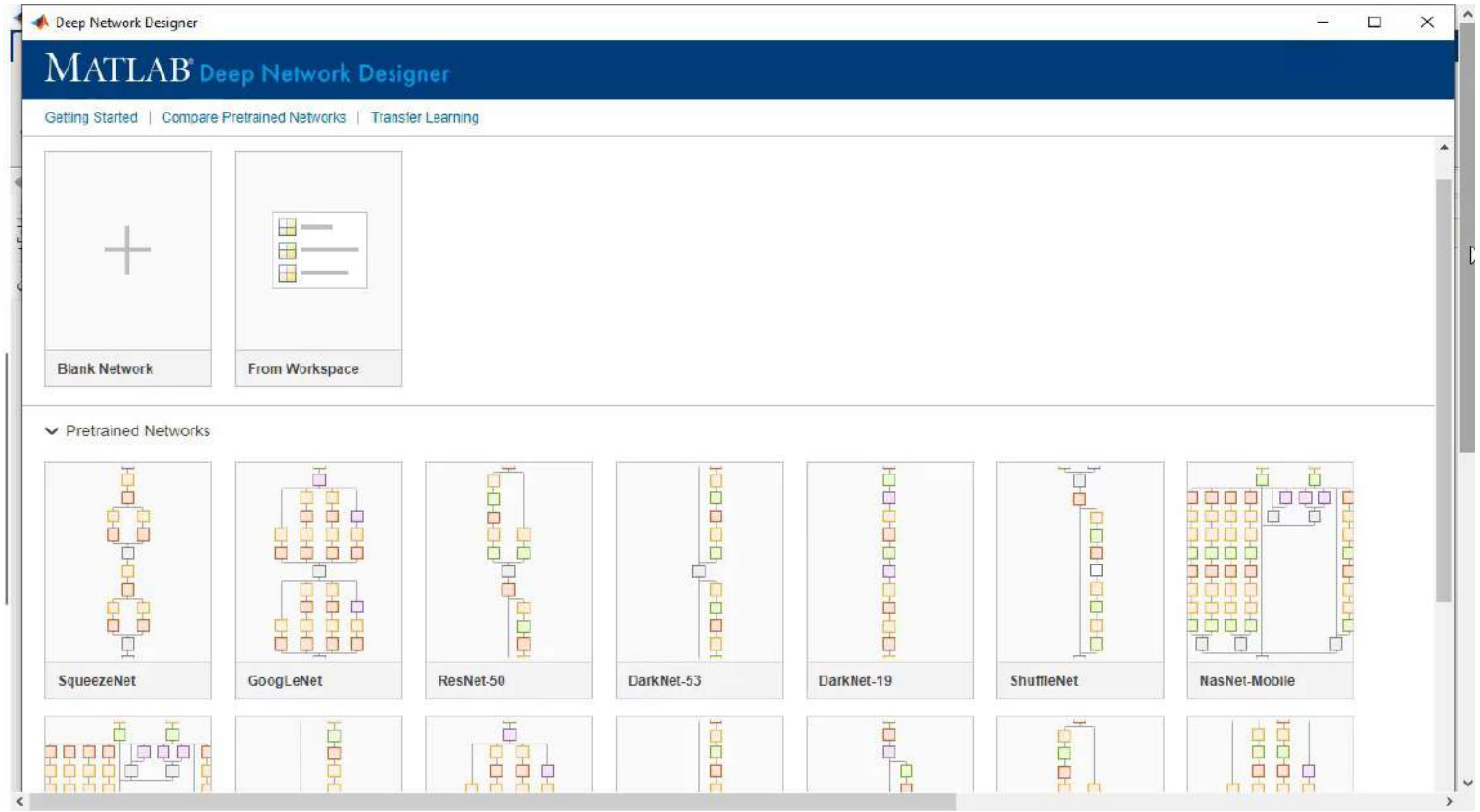
- Use apps to interactively train and compare regression and classification models
  - Use Bayesian optimization in the apps to tune hyperparameters
- Automatically pick the best model via optimization using AutoML functionality
- Support for Incremental/Online learning and semi-supervised learning
- Shallow Neural Networks now supported in Classification/Regression Learner Apps





Deep Learning: Efficiently Design Deep Networks

- Deep Network Designer
  - Easily visualize and design a network
  - Start from scratch or with a pretrained model for transfer learning



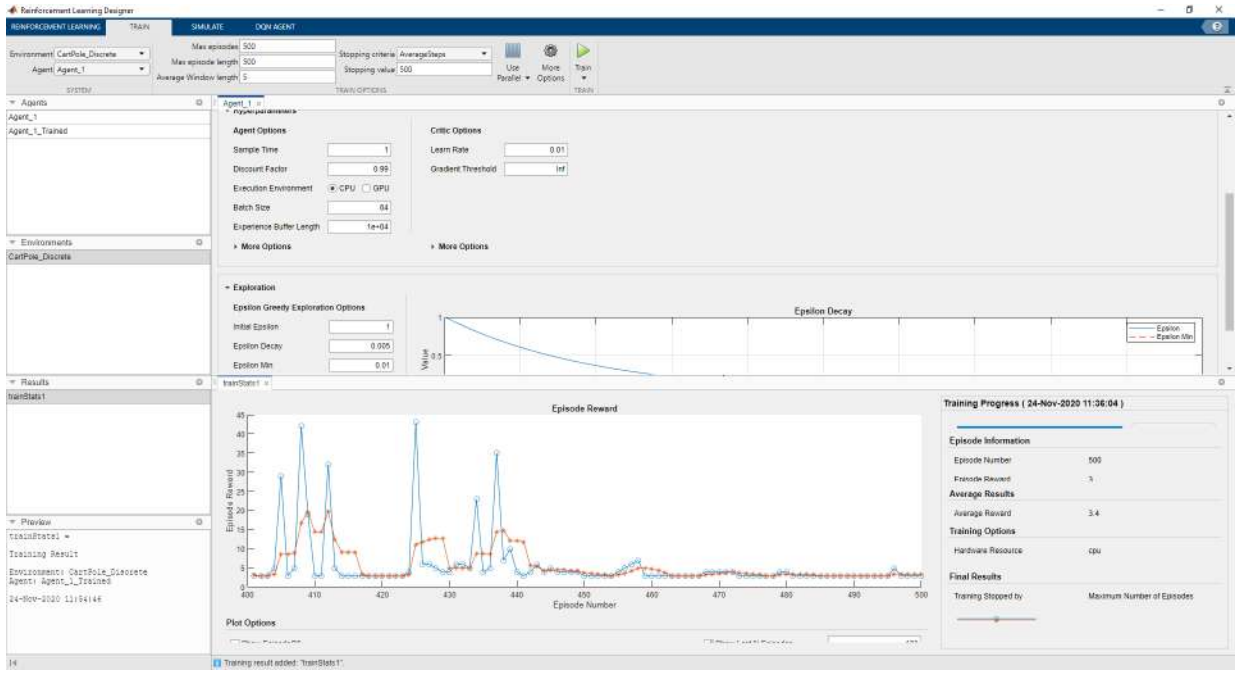
Reinforcement Learning

- **Design, train, and simulate agents interactively with the Reinforcement Learning Designer App**

- Design Environments and Train Reinforcement Learning Agents
  - Q-Learning, DDPG, Actor-Critic agents, TD3, and more

- Warm-start reinforcement learning available via imitation learning

- Create RNN/LSTM-based policies for many kinds of agents

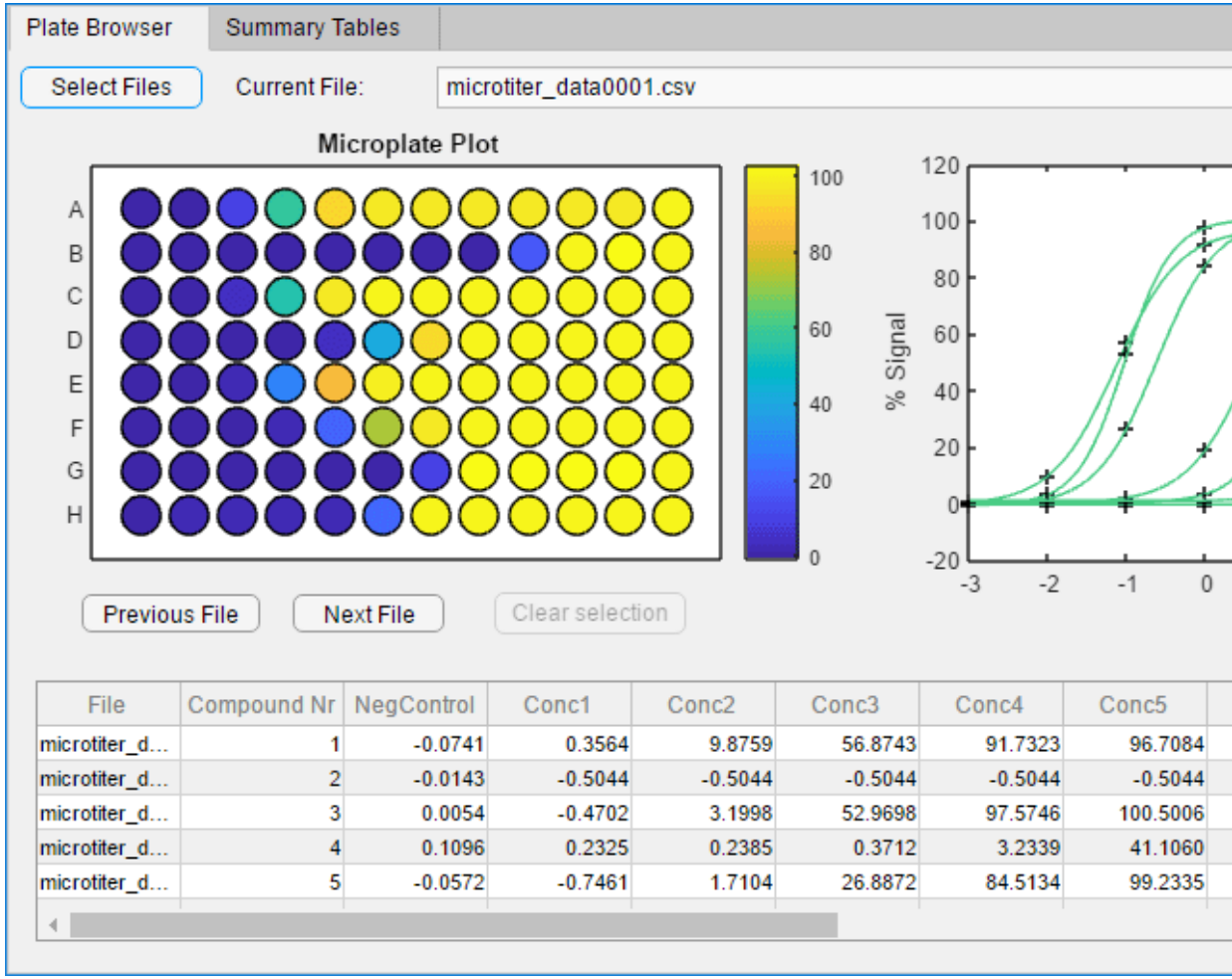


Agenda

- 다양한 금융 데이터 활용 사례
  - 금융 회사 사례
  - 금융 데이터 연구사례
- 금융 데이터 과학
  - 다양한 금융 데이터 소개
  - MATLAB을 활용한 금융 데이터 과학 기법
  - 금융 데이터 전처리 및 시각화 기능 소개
- 금융 데이터를 활용한 모델링 기법 소개
  - 금융 데이터를 활용한 AI 모델링 기법 소개
  - AI 모델 공유 소개

Create and Share Apps as a Vehicle for our Models

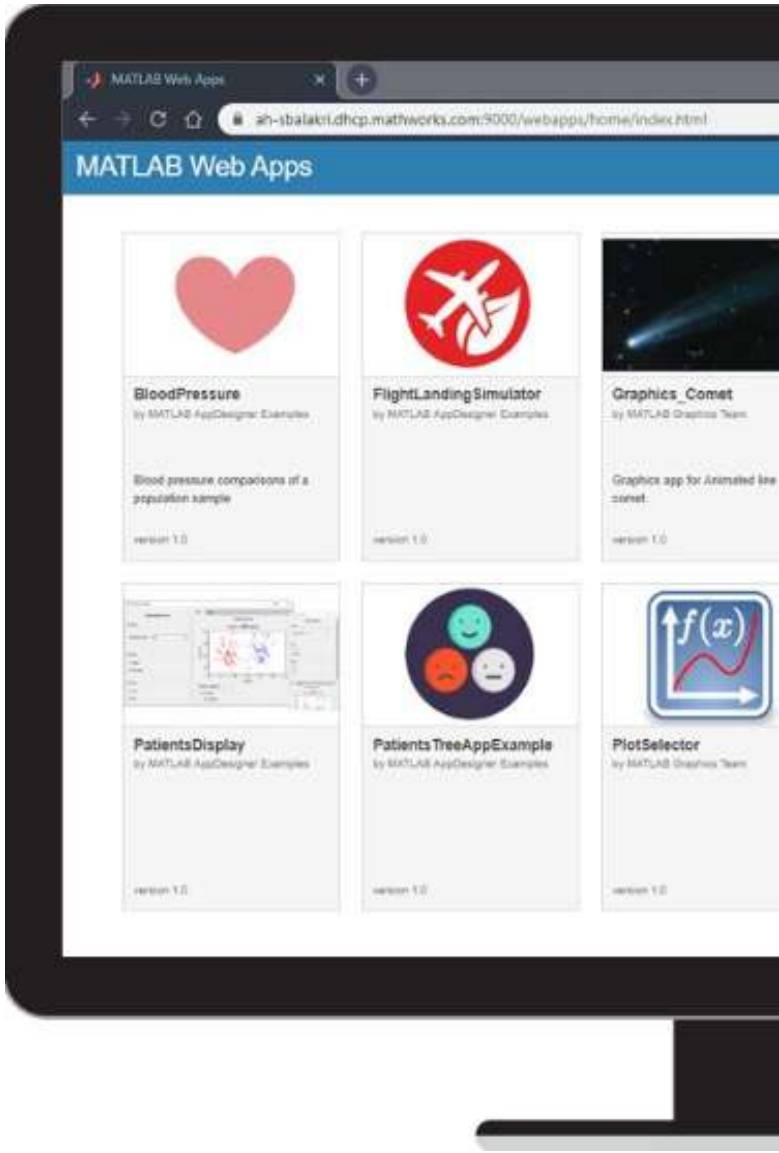
- App designer can be used to design GUIs for your models
- No need to learn another language to make your models interactive



File	Compound Nr	NegControl	Conc1	Conc2	Conc3	Conc4	Conc5
microtiter_d...	1	-0.0741	0.3564	9.8759	56.8743	91.7323	96.7084
microtiter_d...	2	-0.0143	-0.5044	-0.5044	-0.5044	-0.5044	-0.5044
microtiter_d...	3	0.0054	-0.4702	3.1998	52.9698	97.5746	100.5006
microtiter_d...	4	0.1096	0.2325	0.2385	0.3712	3.2339	41.1060
microtiter_d...	5	-0.0572	-0.7461	1.7104	26.8872	84.5134	99.2335

Share MATLAB Apps as Browser-Based Web Apps

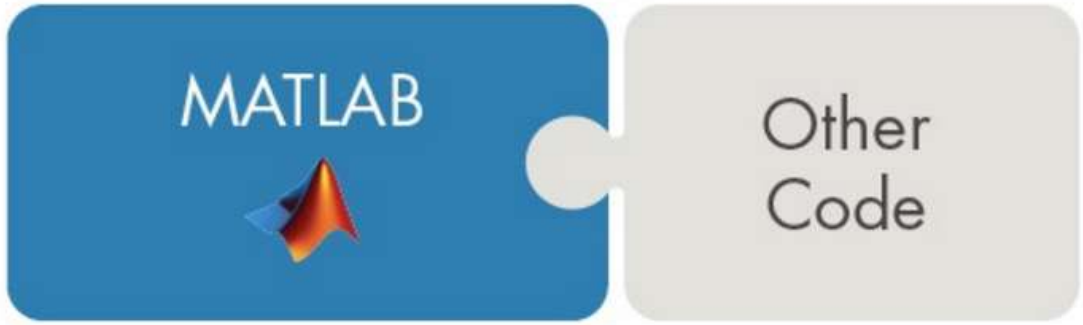
- MATLAB Web App Server
- Provides:
  - Authentication using OpenID Connect and LDAP
  - Support for apps developed in different releases of MATLAB or Simulink
  - Role-based access and editing privileges
  - **Policy-based access control: decide group permissions at an app-by-app level**



https://www.mathworks.com/products/matlab-web-app-server.html

Using MATLAB with Other Languages

Calling Libraries Written in Another Language From MATLAB



- Java
- Python
- C
- C++
- Fortran
- COM components and ActiveX® controls
- RESTful, HTTP, and WSDL web services

Execute Python functions out of process R2019b

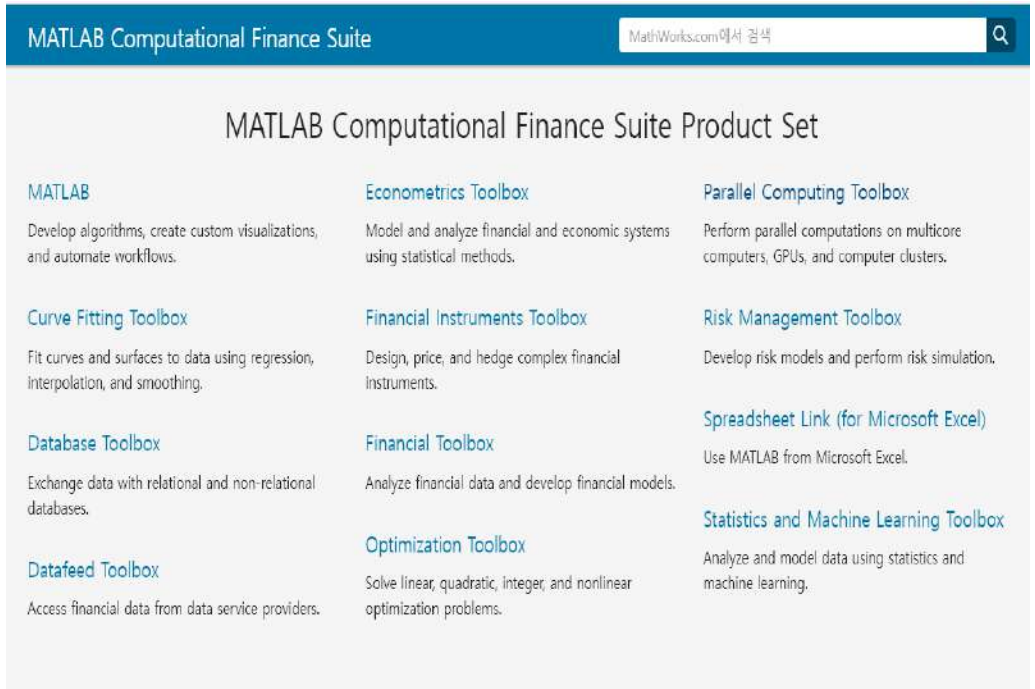
Call C++ libraries directly from MATLAB R2019a

Calling MATLAB from Another Language

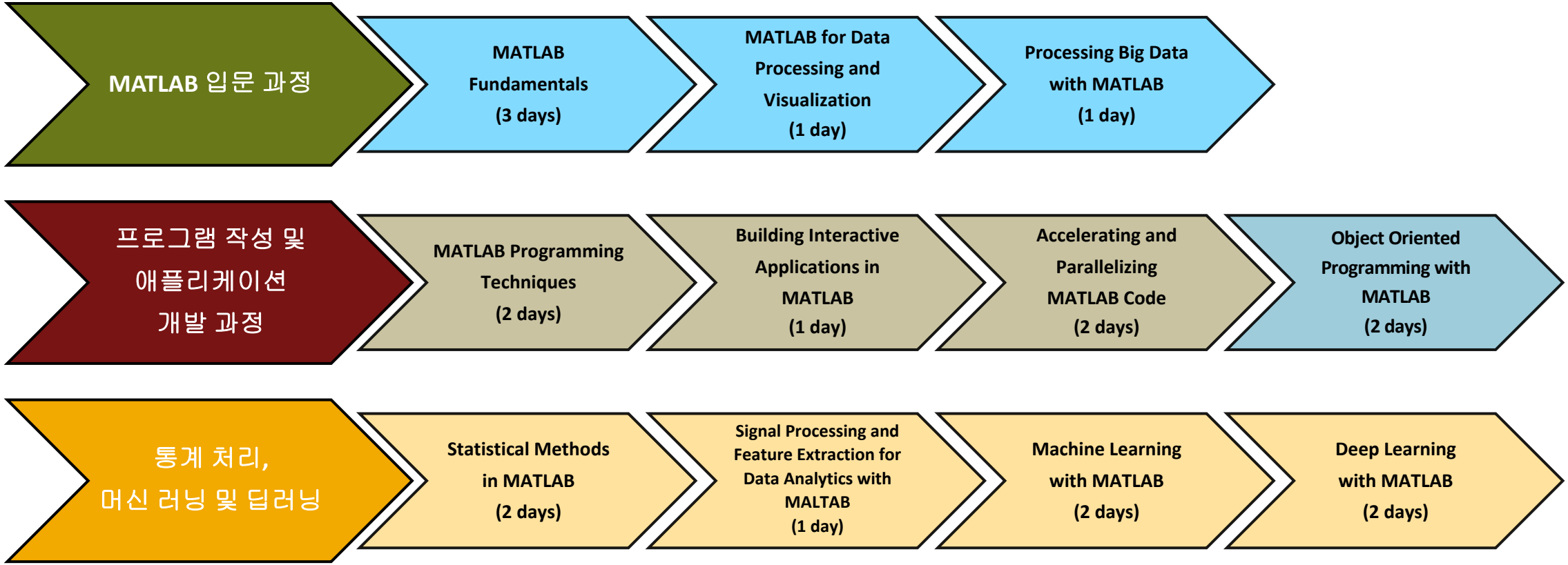
- Java
- Python
- C/C++
- Fortran
- COM Automation server

MATLAB Computational Finance Suite

- MATLAB 금융제품군 사용할 수 있음(Link)
  - 다양한 데이터베이스 및 데이터 피드에 액세스
  - 기계 학습 또는 계량 경제학 모델을 사용하여 재무 데이터 예측
  - 주식 파생 상품, 금리 파생 상품 또는 하이브리드 파생 상품과 같은 가격 금융 상품
  - 대화형 응용 프로그램을 사용하여 신용 성과 기록표 모델링 수행



금융 데이터 사이언스 분야 교육과정



- MathWorks는 데이터 사이언스에 MATLAB을 쉽게 적용하기 위한 단계별 교육 커리큘럼을 제공하며, 대면 또는 비대면 형식으로 교육을 진행하고 있습니다.

금융 분야 집중과정

Prerequisites
For analysts who are new to MathWorks tools.

Risk Management
For risk analysts, risk managers, portfolio managers, and other financial professionals who analyze, assess, and manage risk.

Portfolio Management
For portfolio managers implementing asset allocation strategies.

Time-Series Modeling
For professionals performing quantitative data analytics.

Application Development
For professionals developing and maintaining complex, standalone applications.

Self-paced Online Course
* All online courses can be offered in a classroom

Soft Prerequisite

** Onsite trainings can be customized