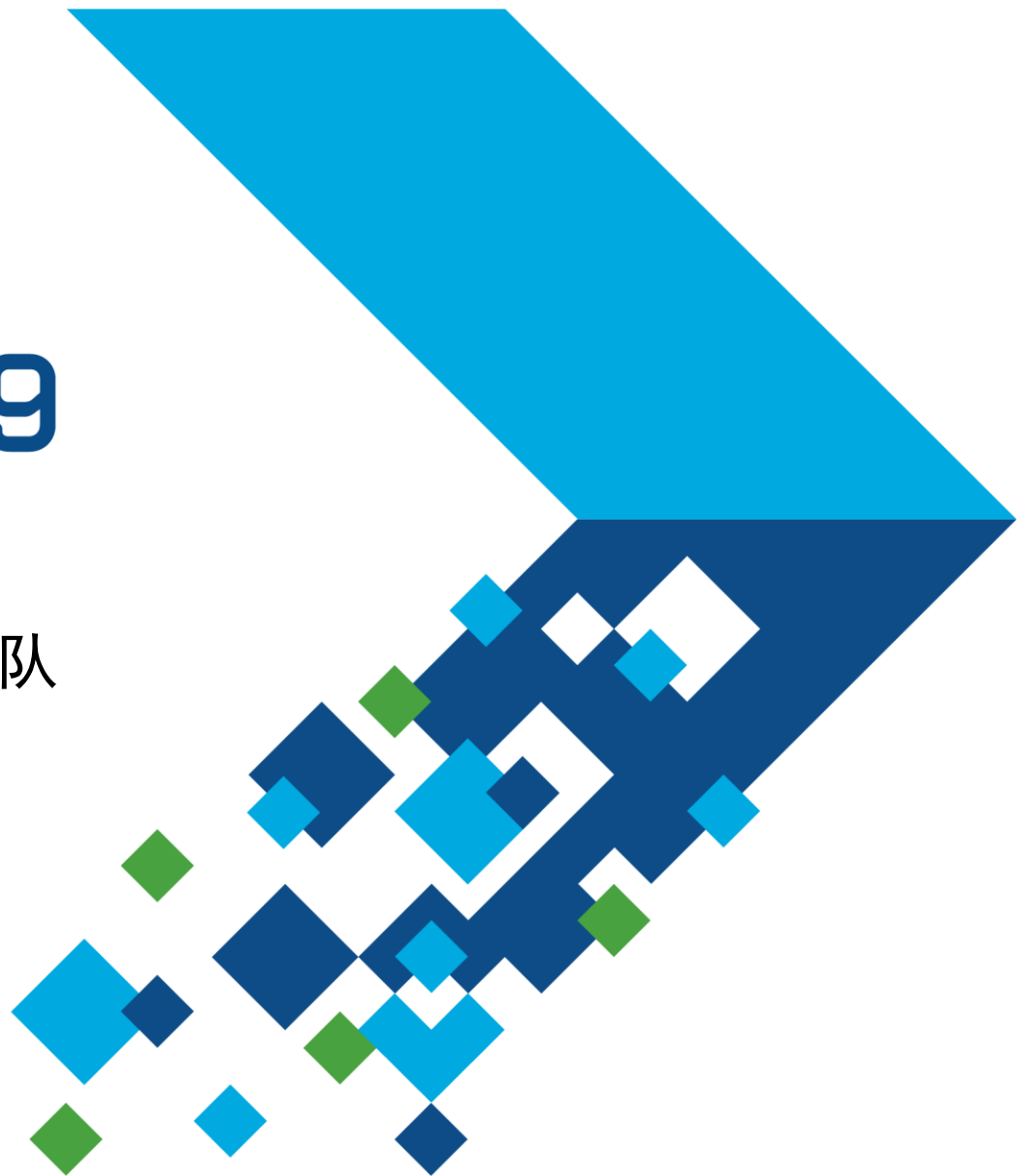


MATLAB EXPO 2019

驾驭数据洪流—建设以数据为中心的工程团队

赵晨星



团队如何更好地使用和利用数据之途径

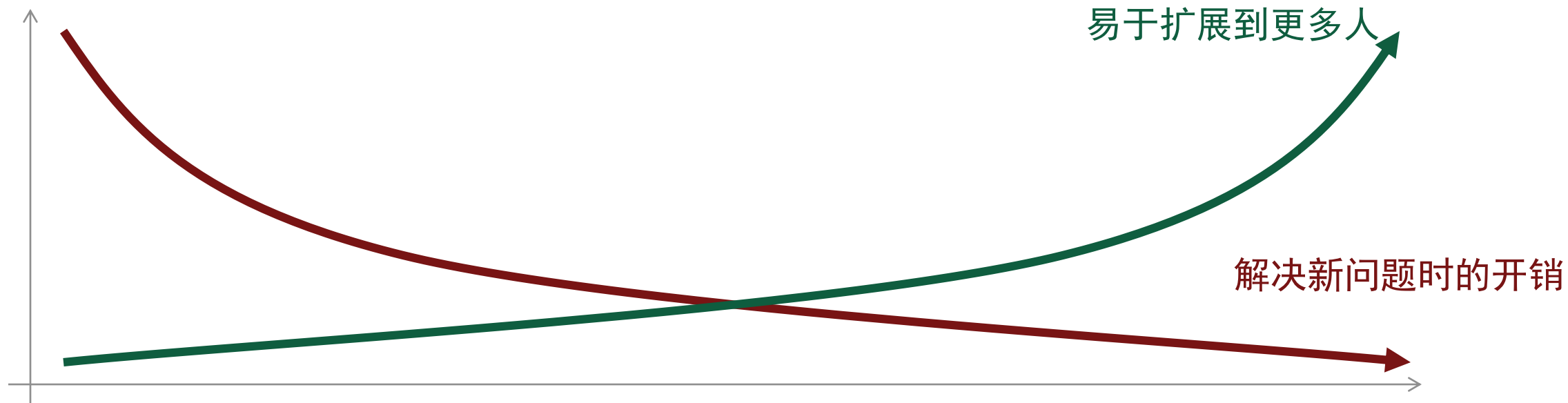
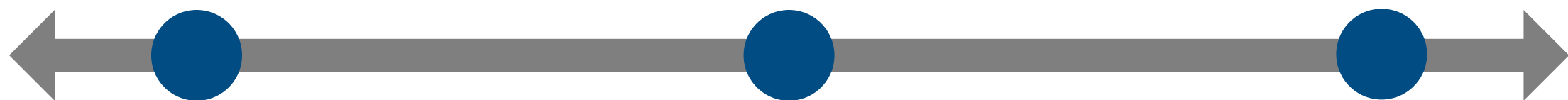


数据科学成熟度

临时的个人数据
分析

通用的分析工具

基础平台
经测试并提供文档

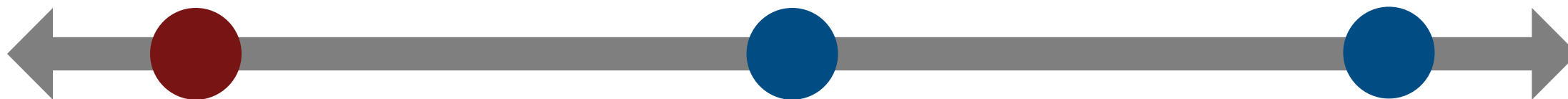


数据科学成熟度

临时的个人数据
分析

通用的分析工具

基础平台
经测试并提供文档



- **目标是快：减少洞察数据所需时间**

开始: 探索新的数据集

The image shows the MATLAB R2019a interface. The top menu bar includes 'MATLAB', 'Window', and 'Help'. The main toolbar is divided into sections: HOME, PLOTS, and APPS. Below the toolbar, there are several toolboxes: FILE (New Script, New Live Script, New, Open, Find Files, Compare, Import Data), VARIABLE (Save Workspace, Clear Workspace, New Variable, Open Variable, Clear Workspace), CODE (Analyze Code, Run and Time, Clear Commands), ENVIRONMENT (Layout, Set Path, Preferences, Add-Ons), and RESOURCES (Help, Community, Request Support, Learn MATLAB). The current folder is 'GettingStartedWithFlightData' located at 'Work > MATLAB'. The workspace is empty. The command window shows 'fx >>'. The current folder view shows two CSV files: 'flightData_1Hz.csv' and 'flightData_4Hz.csv'.

Workspace

Name	Value
------	-------

Current Folder

- flightData_1Hz.csv
- flightData_4Hz.csv

Command Window

```
fx >>
```

开始: 探索新的数据集

The image shows the MATLAB R2019a interface. The main window displays a script titled "Exploring Flight Data" with the following content:

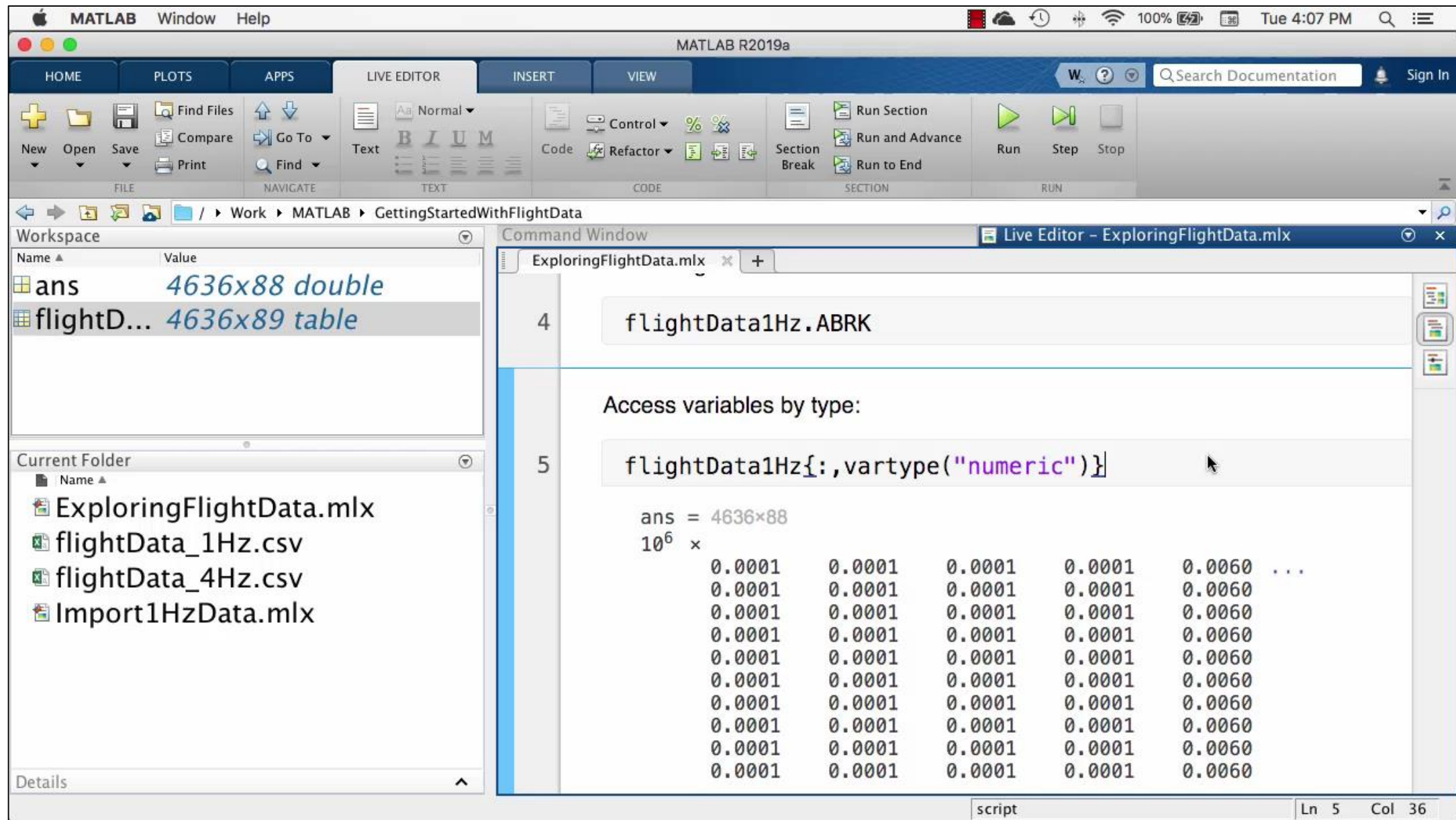
```
1 Import1HzData
```

flightData1Hz = 4636x89 table

	Time	ABRK	ACMT	AIL_1
1	10-May-20...	119.9836	59	91.8789
2	10-May-20...	119.9836	59	91.8994
3	10-May-20...	119.9836	60	91.8994
4	10-May-20...	119.9836	59	91.8994
5	10-May-20...	119.9836	60	91.8789
6	10-May-20...	119.9836	60	91.9607

The interface also shows the Workspace window with a variable named "flightD..." of size "4636x89 table". The Command Window shows the script execution output.

开始: 探索新的数据集



MATLAB R2019a

HOME PLOTS APPS LIVE EDITOR INSERT VIEW

Find Files Find Compare Print Go To Find

Normal Text B I U M

Control Refactor Section Break Run Section Run and Advance Run to End Run Step Stop

Work > MATLAB > GettingStartedWithFlightData

Workspace

Name	Value
ans	4636x88 double
flightD...	4636x89 table

Current Folder

- ExploringFlightData.mlx
- flightData_1Hz.csv
- flightData_4Hz.csv
- Import1HzData.mlx

Command Window

ExploringFlightData.mlx

```
4 flightData1Hz.ABRK
```

Access variables by type:

```
5 flightData1Hz{: , vartype("numeric")}]
```

```
ans = 4636x88
106 x
    0.0001    0.0001    0.0001    0.0001    0.0060 ...
    0.0001    0.0001    0.0001    0.0001    0.0060
    0.0001    0.0001    0.0001    0.0001    0.0060
    0.0001    0.0001    0.0001    0.0001    0.0060
    0.0001    0.0001    0.0001    0.0001    0.0060
    0.0001    0.0001    0.0001    0.0001    0.0060
    0.0001    0.0001    0.0001    0.0001    0.0060
    0.0001    0.0001    0.0001    0.0001    0.0060
    0.0001    0.0001    0.0001    0.0001    0.0060
    0.0001    0.0001    0.0001    0.0001    0.0060
```

script Ln 5 Col 36

开始: 探索新的数据集

The screenshot shows the MATLAB R2019a environment. The workspace contains variables: `ans` (1x1 StackedLineChart), `endtime` (1x1 datetime), `flightD...` (4636x89 table), and `starttime` (1x1 datetime). The Command Window shows a plot of four stacked signals labeled `off_1`, `off_2`, `off_3`, and `off_4`. Each signal shows a noisy baseline with sharp vertical spikes. The x-axis is labeled 'Time' and ranges from 17:15 to 18:00. The y-axis for each signal ranges from -300 to 0.

Missing Data
`ismissing`
`rmissing`
`fillmissing`

Outliers
`isoutlier`
`rmoutliers`
`filloutliers`

Change Points
`ischange`

Noisy Data
`smoothdata`

and more...

<https://www.mathworks.com/help/matlab/preprocessing-data.html>

开始: 探索新的数据集

MATLAB R2019a

HOME PLOTS APPS LIVE EDITOR INSERT VIEW

W Search Documentation Sign In

FILE NAVIGATE TEXT CODE SECTION RUN

Work > MATLAB > GettingStartedWithFlightData

Live Editor - ExploringFlightData.mlx

Workspace

Name	Value
ans	1x1 StackedLineChart
endtime	1x1 datetime
flightD...	4636x89 table
starttime	1x1 datetime

Current Folder

- ExploringFlightData.mlx
- flightData_1Hz.csv
- flightData_4Hz.csv
- Import1HzData.mlx
- Import4HzData.mlx

Command Window

ExploringFlightData.mlx

80
60
Time
17:15 17:30 17:45 18:00 18:15
May 10, 2001

Load 4Hz Data

Load data from same flight that was sampled at 4Hz.

```
15 Import4HzData
16 t4hz = table2timetable(flightData4Hz)
```

Synchronize 1Hz and 4Hz Data

Join the 1Hz and 4Hz data using the synchronize command. The synchronize command gives us flexibility in how the synchronize occurs. Here, we use the default synchronize method which synchronizes the

script

开始: 探索新的数据集

The screenshot shows the MATLAB R2019a interface. The workspace contains variables: 'ans' (1x1 StackedLineChart), 'endtime' (1x1 datetime), and two 'flightD...' tables (4636x89 and 18544x50). The current folder contains files like 'ExploringFlightData.mlx', 'flightData_1Hz.csv', 'flightData_4Hz.csv', 'Import1HzData.mlx', and 'Import4HzData.mlx'. The command window shows a table of data with columns for time and longitude. The live editor shows a script with the following code:

```

19 figure;
20 inFlight = t.WOW==1;
21 geoplot(t.LATP(inFlight),t.LONP(inFlight),'Linewidth',3);

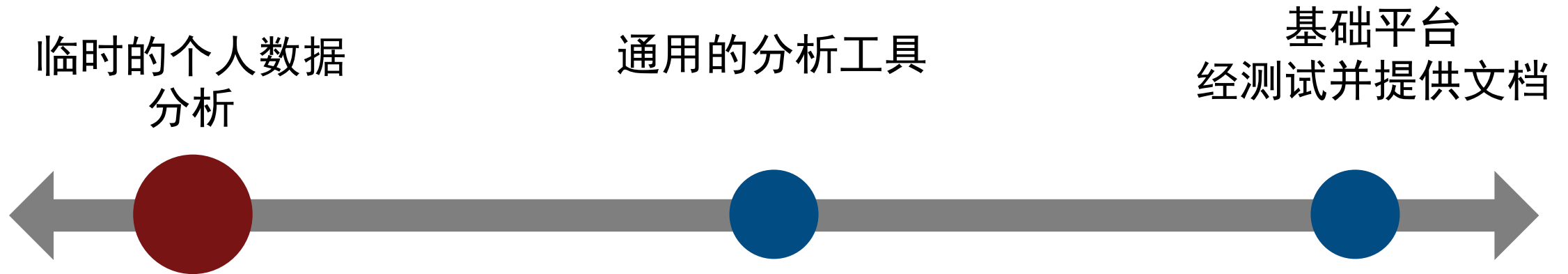
```

This section displays four different geographic visualization functions:

- geoplot**: A line plot showing a flight path on a map of North America, with latitude on the y-axis (39 to 42) and longitude on the x-axis (-86 to -82).
- geoscatter**: A scatter plot showing data points across a world map, with latitude on the y-axis (-60 to 60) and longitude on the x-axis (-100 to 100).
- geobubble**: A bubble plot showing data points on a world map, with latitude on the y-axis (30°S to 50°N) and longitude on the x-axis (210°W to 90°W). A scale bar indicates 5000 km and 2000 mi.
- geodensityplot**: A density plot showing data points on a world map, with latitude on the y-axis (10 to 60) and longitude on the x-axis (100 to 200).

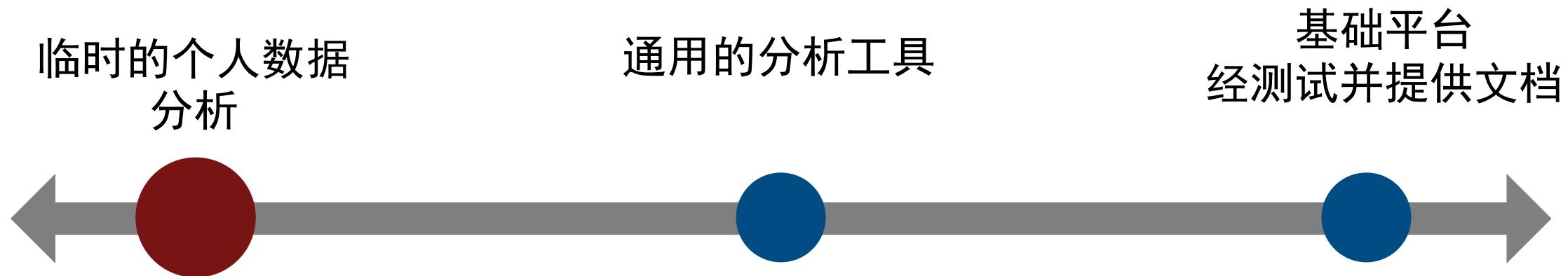
<https://www.mathworks.com/help/matlab/geographic-plots.html>

数据科学成熟度



- 探索并解析数据
- 记录分析
- 在以后的步骤中重用

数据科学成熟度



- 适用于不同的数据集
- 函数/代码
- MATLAB Apps
- 趋势: 处理**大数据**

飞行数据一览

- **35 架飞机**
- **180,000 次飞行**
- **300 GB 的数据**
- 数据源:
 - NASA Dash Link: Sample Flight Data
 - <https://c3.nasa.gov/dashlink/projects/85/>



大数据创造新机会



搜索数据以查找罕见事件，然后深入了解事件的背景



使用真实世界的数据来构建和验证符合现实条件的测试场景



执行整个机群的分析，以了解更高层次的趋势

大数据需要新工具

Create a datastore from all CSV files

```
ds = datastore('*.*csv')
```

Read a single file of data

```
data = read(ds);
```

Reset the datastore back to the first file

```
reset(ds);
```

Find the maximum value of "Y" in each file

```
X = [];
while hasdata(ds)
    data = read(ds);
    X(end+1) = max(data.Y);
end
```

Built-In Datastores	
General	datastore
	spreadsheetDatastore
	tabularTextDatastore
	fileDatastore
Database	databaseDatastore
Image	imageDatastore
	denoisingImageDatastore
	randomPatchExtractionDatastore
	pixelLabelDatastore
	augmentedImageDatastore
Audio	audioDatastore
Predictive Maintenance	fileEnsembleDatastore
	simulationEnsembleDatastore
Simulink	SimulationDatastore
Automotive	mdfDatastore

大数据需要新工具

自定义数据存储

- 为你的数据库定义数据存储区
- 使您能够控制如何加载和格式化数据
- MATLAB子类: “填空”
- 构建基础, 然后在您的分析中重用

```
function [data,info] = read(ds)
    ...
end
```

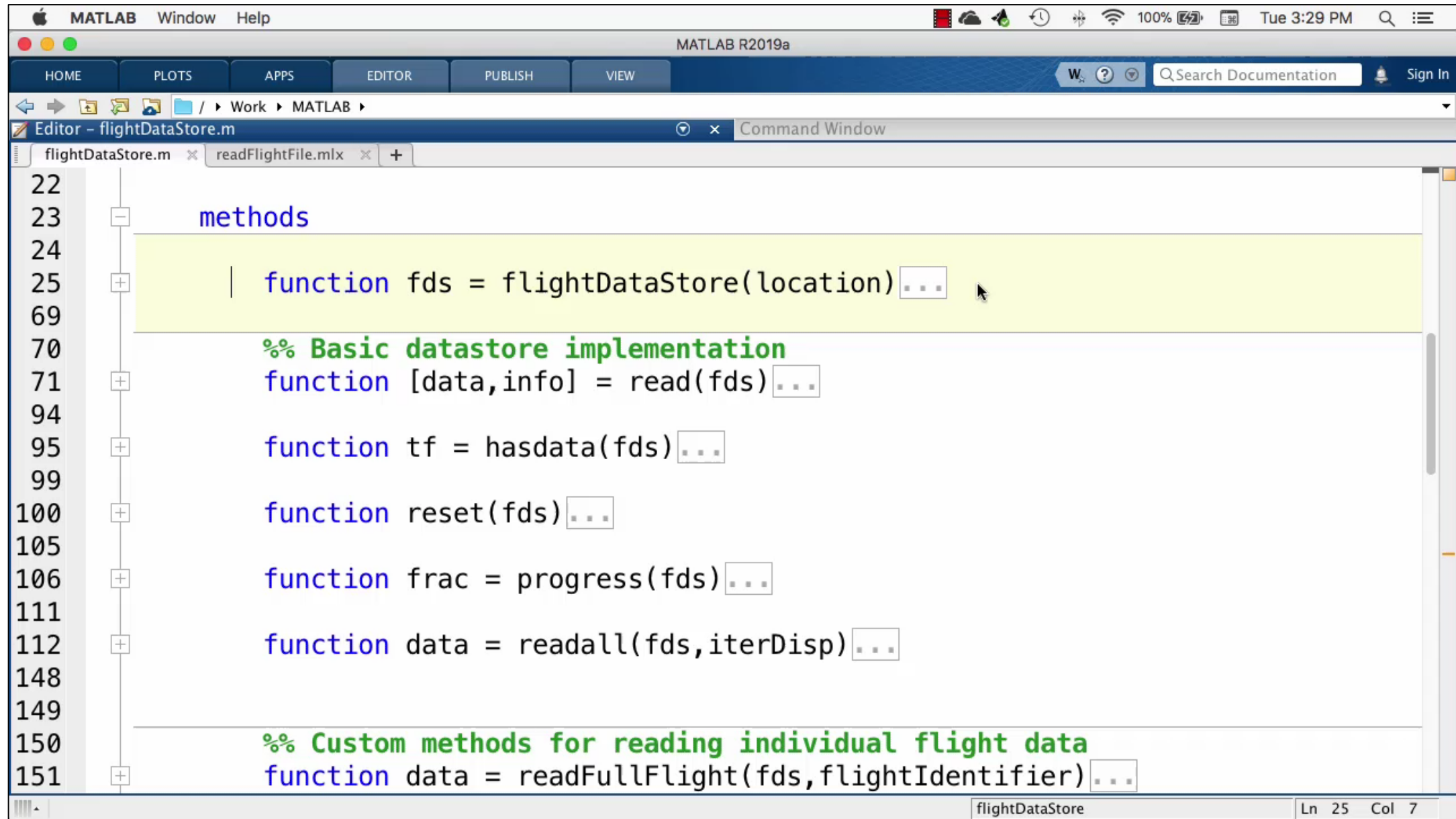
```
function tf = hasdata(ds)
    ...
end
```

```
function reset(ds)
    ...
end
```

```
function p = progress(ds)
    ...
end
```

```
function data = readall(ds)
    ...
end
```


给飞行数据自定义数据存储



```
MATLAB R2019a
HOME PLOTS APPS EDITOR PUBLISH VIEW
W Search Documentation Sign In
Work > MATLAB >
Editor - flightDataStore.m Command Window
flightDataStore.m readFlightFile.mlx +
22
23 methods
24
25 | function fds = flightDataStore(location) ...
69
70 %% Basic datastore implementation
71 function [data,info] = read(fds) ...
94
95 function tf = hasdata(fds) ...
99
100 function reset(fds) ...
105
106 function frac = progress(fds) ...
111
112 function data = readall(fds,iterDisp) ...
148
149
150 %% Custom methods for reading individual flight data
151 function data = readFullFlight(fds,flightIdentifier) ...
flightDataStore Ln 25 Col 7
```



发现异常事件, 再深入

The image shows the MATLAB R2019a interface. The top menu bar includes 'MATLAB', 'Window', and 'Help'. The title bar indicates 'MATLAB R2019a'. The main window is titled 'Live Editor - IncidentAnalysisExample.mlx' and contains the following content:

Incident Analysis

The data for tail 666 captures an engine fire that happened shortly after takeoff. This example analyzes data from that tail number to see if there is any interesting information that can be learned about how/when the fire occurred.

Search for Flights where the Engine 2 Fire Sensor Went True

Loop through all flights for this aircraft, and display some information if there is a fire.

```
1 ds = flightDataStore('/Volumes/Ext/Data/FlightData/666/');|
2 ds.SelectedVariableNames = {'FIRE_2'};
3 ds.AddFlightIdentifier = true;

4 results = table('Size',[0 2],'VariableTypes',{'categorical','double'},...
5     'VariableNames',{'Flight','Fire_Seconds'});
6 while hasdata(ds)
7     t = read(ds);
8     if any(t.FIRE_2)
9         results = [results; {t.Flight(1), sum(t.FIRE_2)}];
10    end
11 end
12 results
```

The status bar at the bottom shows 'script' and 'Ln 1 Col 59'.

执行“机群”数据计算

The image shows the MATLAB R2019a interface. The main window displays a script titled "FleetWideFuelEconomy.mlx" with the following content:

Fleet-Wide Calculations

Calculate fuel efficiency across all of the flights in the dataset.

Tall Arrays

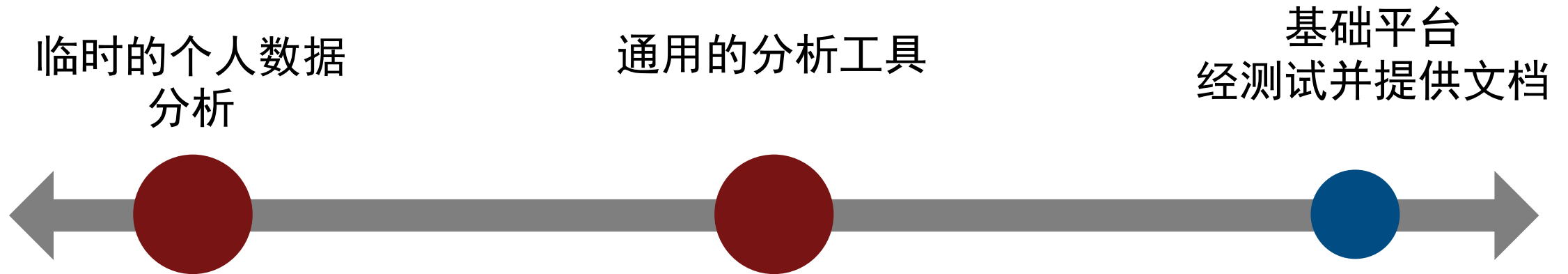
Creating a tall array from the datastore gives us access to many built-in functions.

```
1 ds = flightDataStore('/Volumes/Ext/Data/FlightData/');
```

TAS: True Airspeed (knots)
FF_N: Fuel Flow N (lbs/hr)
WOW: Weight On Wheels (logical)

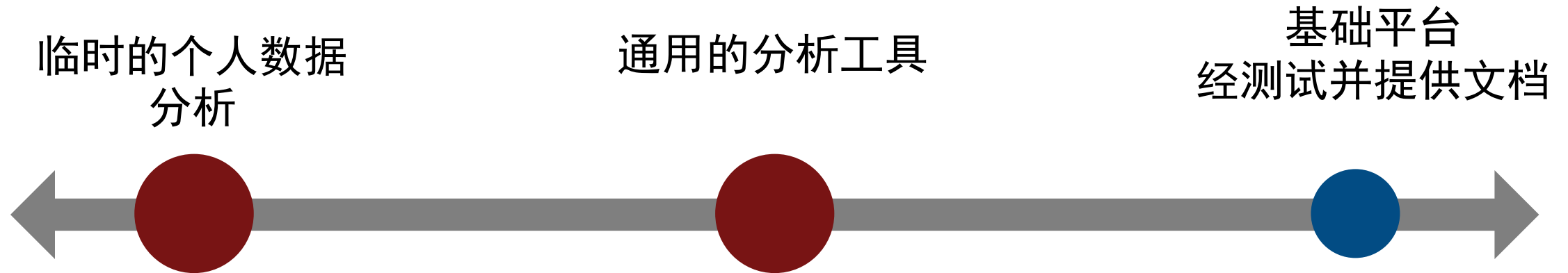
At the bottom, a partial line of code is visible: `ds.SelectedVariableNames = {'TAS', 'FF_N', 'FF_N', 'FF_N', 'FF_N', 'WOW'};`

数据科学成熟度



- 使数据导航更容易
- 每次分析数据集时都要重用

数据科学成熟度



- **协作:在公共代码基础上与他人协作**
- **验证:编写经过良好测试的软件**
- **共享:为他人构建工具**

MATLAB 工程

The screenshot displays the MATLAB R2019a desktop environment. The top menu bar includes 'MATLAB', 'Window', and 'Help'. The system status bar shows the time as 'Mon 1:51 PM' and battery level at '100%'. The main interface features a ribbon with tabs for 'HOME', 'PLOTS', and 'APPS'. The 'HOME' tab is active, showing various toolboxes and functions categorized into 'FILE', 'VARIABLE', 'CODE', 'ENVIRONMENT', and 'RESOURCES'. The 'Command Window' is open, showing the prompt 'fx >>' and a mouse cursor. The current project is 'Project - FlightDataToolbox'.

测试

The screenshot shows the MATLAB R2019a Project - FlightDataToolbox interface. The top toolbar includes sections for FILE, TOOLS, ENVIRONMENT, and SOURCE CONTROL. The main workspace displays a file explorer view of the project directory.

Name	Status	Git	Size	Date Modified
+tests	✓	·		2/28/19, 1:...
doc	✓	·		3/1/19, 11:...
flightDataStore.m	✓	●	13 KB	3/4/19, 11:...
FlightDataToolbox.mltbx	✓	●	426 KB	3/1/19, 11:...
info.xml	✓	●	2 KB	3/1/19, 9:5...
PackageToolbox.prj	✓	●	6 KB	3/1/19, 11:...
readFlightFile.mlx	✓	●	5 KB	2/26/19, 10...

The Command Window at the bottom shows the MATLAB prompt `fx >>`.

建立工具箱

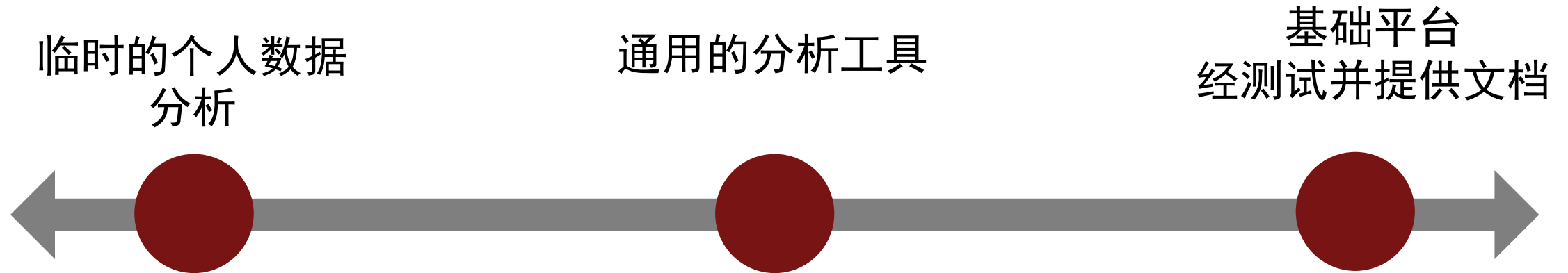
The screenshot shows the MATLAB R2019a interface. The top menu bar includes 'HOME', 'PLOTS', 'APPS', 'PROJECT', and 'PROJECT SHORTCUTS'. The 'PROJECT' tab is active, showing a toolbar with icons for file operations (New, Open, Add Files, Unsaved Changes, Share), search, custom tasks, and run checks. Below the toolbar, the 'ENVIRONMENT' and 'SOURCE CONTROL' sections are visible. The 'SOURCE CONTROL' section includes icons for Fetch, Push, Pull, Remote, Branches, and Stashes.

The Command Window displays the 'Project - FlightDataToolbox' view. The left sidebar shows 'Views' with 'Files' selected. The main area shows a table of files and folders in the project:

Name	Status	Git	Size	Date Modified
+tests	✓	.		2/28/19, 1:...
doc	✓	.		3/1/19, 11:...
flightDataStore.m	✓	●	13 KB	3/4/19, 11:...
FlightDataToolbox.mltbx	✓	●	426 KB	3/4/19, 2:0...
info.xml	✓	●	2 KB	3/1/19, 9:5...
PackageToolbox.prj	✓	●	6 KB	3/1/19, 11:...
readFlightFile.mlx	✓	●	5 KB	2/26/19, 10...

At the bottom of the Command Window, the 'Git' section shows the current branch as 'master', the branch status as 'Normal', and that it is 'Ahead of /origin/master'.

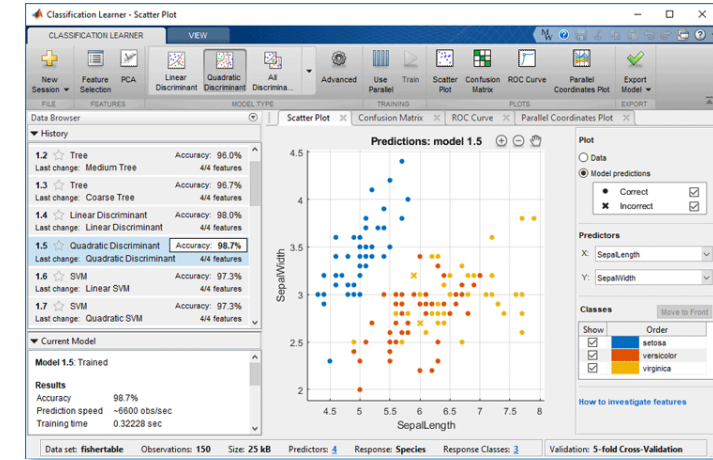
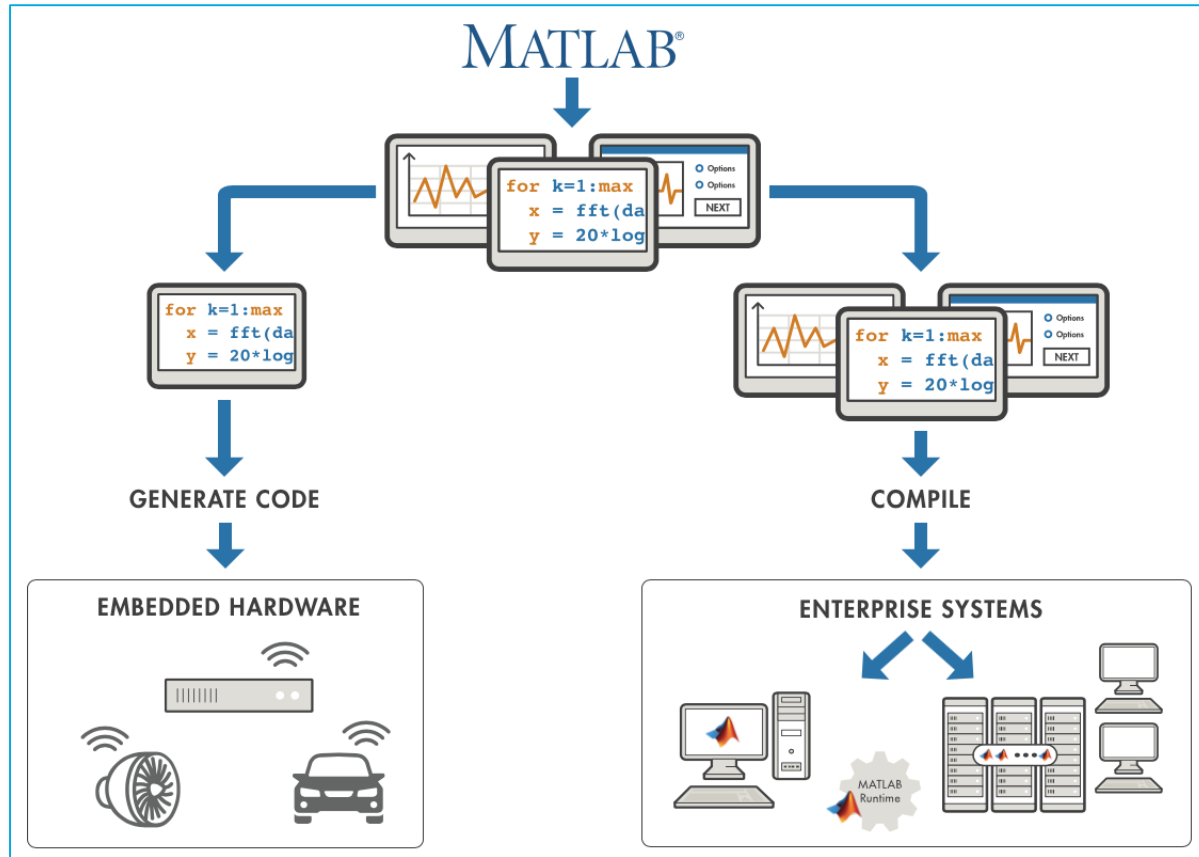
数据科学成熟度



- 向更大的用户群扩展
- 易于维护和共享

下一步?

高级分析和机器学习



为嵌入式系统构建和测试算法

将应用程序和分析部署到企业IT系统

要点

- MATLAB有许多新工具可以帮助您更好地处理和利用数据
- 为您/您的团队/您的组织创建工具来探索和分析数据
- 提高数据科学的成熟度是一个过程，MATLAB能起到很大的帮助