

MATLAB × データサイエンス

～最新機能紹介と業務一元化の手引き～

MathWorks Japan

アプリケーションエンジニアリング部

田口 美紗

データサイエンス？

AI？



ディープラーニング？



機械学習？



データサイエンスで大切なこと

✓ 可視化・異常値チェックなど前処理が不可欠

✗ しかし、前処理には非常に時間がかかる

Web セミナーの目的

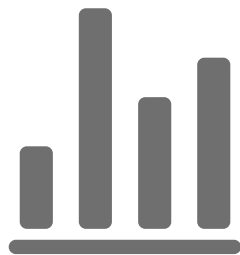
**MATLAB® を用いて効率よく
データサイエンスのプロセスを
進めて頂きたい！**



アジェンダ

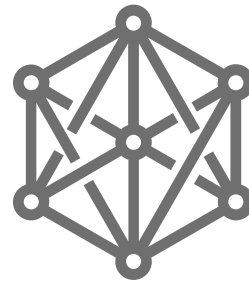
① アドホックな解析

- データ読み込み
- データアクセス
- 前処理
- 可視化



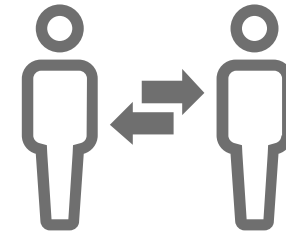
② ビッグデータ解析

- ビッグデータの読み込み
- ビッグデータの解析・深掘り



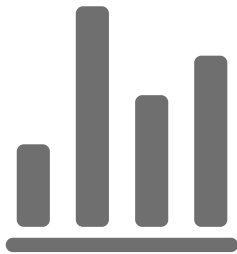
③ プロジェクト化

- プロジェクトの作成・共有
- 大人数で解析



①アドホックな解析

- データ読み込み
- データアクセス
- 前処理
- 可視化



フライトデータの解析
を例にご紹介

振動センサ

燃料センサ



時間

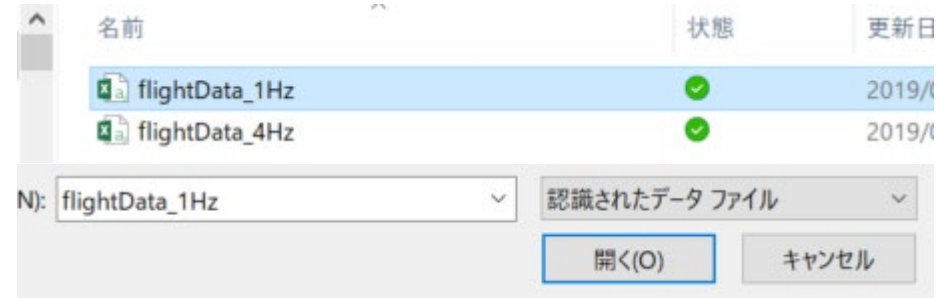
緯度・経度

火災センサ



データ読み込み

① GUI からのデータ読み込み

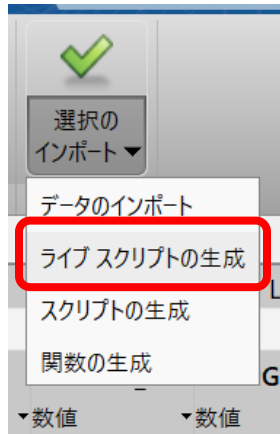


インポート		表示		MATLAB training											
<input type="radio"/> 区切り	列区切り文字: コンマ	範囲: A3:CK4...	変数名行: 1	出カタイプ: table	<input type="checkbox"/> 置き換え	次を含むインポートできないセル	NaN	<input checked="" type="checkbox"/>	選択のインポート						インポート
<input type="radio"/> 固定幅	区切り文字...			テキストオプション											
flightData_1Hz.csv															
flightData1Hz1															
	Time	ABRK	ACMT	AIL_1	AIL_2	ALTS	APFD	ATEN	A_T	BLV	BPGR_1	BPGR_2	BPYR_1	BPYR_2	CALT
テキスト	▼数値	▼数値	▼数値	▼数値	▼数値	▼数値	▼数値	▼数値	▼数値	▼数値	▼数値	▼数値	▼数値	▼数値	▼数値
1	Time	ABRK	ACMT	AIL_1	AIL_2	ALTS	APFD	ATEN	A_T	BLV	BPGR_1	BPGR_2	BPYR_1	BPYR_2	CALT
2		DEG		DEG	DEG	FEET					PSI	PSI	PSI	PSI	
3	10-May-20...	119.983558...	59	91.8788909...	91.5924835...	6000	2	0	1	0	0	43.9454231...	2373.05297...	0	0
4	10-May-20...	119.983558...	59	91.8993530...	91.5924835...	6000	2	0	1	0	0	43.9454231...	2373.05297...	0	0
5	10-May-20...	119.983558...	60	91.8993530...	91.5924835...	6000	2	0	1	0	0	43.9454231...	2368.17016...	0	0
6	10-May-20...	119.983558...	59	91.8993530...	91.5720214...	6000	2	0	1	0	0	43.9454231...	2368.17016...	0	0
7	10-May-20...	119.983558...	60	91.8788909...	91.5720214...	6000	2	0	1	0	0	43.9454231...	2368.17016...	0	0
8	10-May-20...	119.983558...	60	91.9607238...	91.5720214...	6000	2	0	1	0	0	43.9454231...	2368.17016...	0	0
9	10-May-20...	119.983558...	59	91.9198150...	91.5515670...	6000	2	0	1	0	0	43.9454231...	2368.17016...	0	0
10	10-May-20...	119.983558...	60	91.8788909...	91.5924835...	6000	2	0	1	0	0	43.9454231...	2368.17016...	0	0
11	10-May-20...	119.983558...	59	91.8175201...	91.5515670...	6000	2	0	1	0	0	48.8282470...	2368.17016...	0	0
12	10-May-20...	119.983558...	60	91.8175201...	91.5311126...	6000	2	0	1	0	0	43.9454231...	2368.17016...	0	0
13	10-May-20...	119.983558...	60	91.8175201...	91.5311126...	6000	2	0	1	0	0	43.9454231...	2368.17016...	0	0
14	10-May-20...	119.983558...	59	91.8175201...	91.5106506...	6000	2	0	1	0	0	43.9454231...	2363.28735...	0	0
15	10-May-20...	119.983558...	60	91.8175201...	91.5311126...	6000	2	0	1	0	0	48.8282470...	2368.17016...	0	0
16	10-May-20...	119.983558...	59	91.8379821...	91.5106506...	6000	2	0	1	0	0	43.9454231...	2363.28735...	0	0
17	10-May-20...	119.983558...	60	91.8379821...	91.5311126...	6000	2	0	1	0	0	43.9454231...	2363.28735...	0	0
18	10-May-20...	119.983558...	60	91.8379821...	91.4901885...	6000	2	0	1	0	0	43.9454231...	2910.16357...	0	0
19	10-May-20...	119.983558...	59	91.8584365...	91.5311126...	6000	2	0	1	0	0	43.9454231...	2910.16357...	0	0



データ読み込み

② GUI 操作と等価のライブスクリプト 自動生成



テキストファイルからのデータのインポート

次のテキストファイルからデータをインポートするスクリプト:

ファイル名: C:\Users\mtaguchi\OneDrive - MathWorks\Training\datacentricengineeringteams-master\exploringflightdata\flightDa

MATLAB からの自動生成日: 2019/04/11 17:51:50

インポート オプションの設定

```

1  opts = delimitedTextImportOptions("NumVariables", 89);
2
3  % 範囲と区切り記号の指定
4  opts.DataLines = [3, Inf];
5  opts.Delimiter = ",";
6
7  % 列名と型の指定
8  opts.VariableNames = ["Time", "ABRK", "ACMT", "AIL_1", "AIL_2", "ALTS", "APFD", "ATEN", "A_T", "BLV", "BPGR_1",
9  opts.VariableTypes = ["string", "double", "double", "double", "double", "double", "double", "double", "double", "double",
10 opts = setvaropts(opts, 1, "WhitespaceRule", "preserve");
11 opts = setvaropts(opts, 1, "EmptyFieldRule", "auto");
12 opts.ExtraColumnsRule = "ignore";
13 opts.EmptyLineRule = "read";
14
15 % データのインポート
16 flightData1Hz1 = readtable("C:\Users\mtaguchi\OneDrive - MathWorks\Training\datacentricengineeringteams-master\

```




データ読み込み

③ ライブスクリプトの関数化 → 以降の読み込みを自動化

ファイル名(N):

ファイルの種類(T):

保存(S) キャンセル

コマンドウィンドウ

MATLAB のご利用がはじめての場合は、[入門](#)をご覧ください。

```
>> Import1HzData
flightData1Hz =
4636×89 table
    Time    ABRK    ACMT    AIL_1    AIL_2    ALTS    APFD    ATEN    A_T    BLV    BPGR_1    BPGR_2    BPYR_1    BPYR_2    CALT    CASS    CRSS    DFGS    DWPT
    _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____
    NaT    119.98    59    91.879    91.592    6000    2    0    1    0    0    43.945    2373.1    0    0    128    33.926    1    42188
    NaT    119.98    59    91.899    91.592    6000    2    0    1    0    0    43.945    2373.1    0    0    128    33.926    1    42188
    NaT    119.98    60    91.899    91.592    6000    2    0    1    0    0    43.945    2368.2    0    0    128    33.926    1    42188
    NaT    119.98    59    91.899    91.572    6000    2    0    1    0    0    43.945    2368.2    0    0    128    33.926    1    42188
    NaT    119.98    60    91.879    91.572    6000    2    0    1    0    0    43.945    2368.2    0    0    128    33.926    1    42188
    NaT    119.98    60    91.961    91.572    6000    2    0    1    0    0    43.945    2368.2    0    0    128    33.926    1    42188
    NaT    119.98    59    91.92    91.552    6000    2    0    1    0    0    43.945    2368.2    0    0    128    33.926    1    42188
    NaT    119.98    60    91.879    91.592    6000    2    0    1    0    0    43.945    2368.2    0    0    128    33.926    1    42188
    NaT    119.98    59    91.818    91.552    6000    2    0    1    0    0    48.828    2368.2    0    0    128    33.926    1    42188
    NaT    119.98    60    91.818    91.531    6000    2    0    1    0    0    43.945    2368.2    0    0    128    33.926    1    42188
    NaT    119.98    60    91.818    91.531    6000    2    0    1    0    0    43.945    2368.2    0    0    128    33.926    1    42188
    NaT    119.98    59    91.818    91.511    6000    2    0    1    0    0    43.945    2363.3    0    0    128    33.926    1    42188
    NaT    119.98    60    91.818    91.521    6000    2    0    1    0    0    48.828    2368.2    0    0    128    33.926    1    42188
```



データアクセス

読み込んだデータ



flightData1Hz 4636x89 table

時間 × センサ

データ.変数名

flightData1Hz.FIRE_1

変数型でソート

flightData1Hz{: ,vartype("numeric")}

```
ans = 4636x88
r 106 ×
    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
    ⋮
```



タイムテーブルへの変換

タイムテーブルへ変換

```
t1hz = table2timetable(flightData1Hz)
```

t1hz = 4636x88 timetable

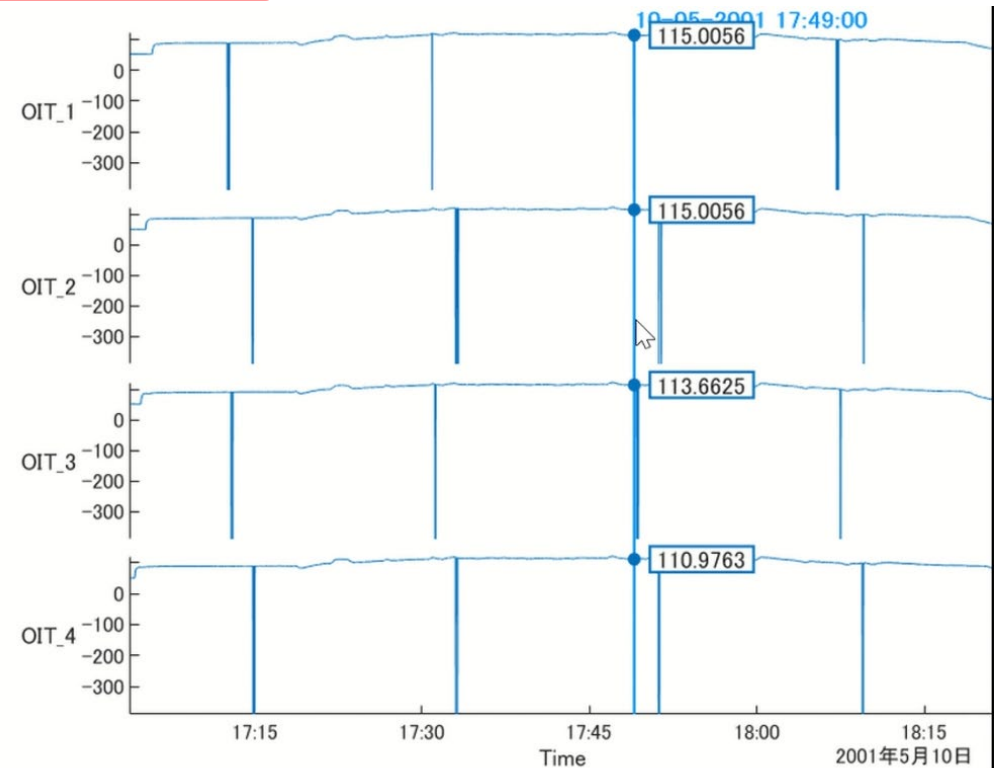
	Time	ABRK	ACMT	AIL_1
1	10-May-20...	119.9836	59	91.8789
2	10-May-20...	119.9836	59	91.8994
3	10-May-20...	119.9836	60	91.8994
4	10-May-20...	119.9836	59	91.8994
5	10-May-20...	119.9836	60	91.8789
6	10-May-20...	119.9836	60	91.9607
7	10-May-20...	119.9836	59	91.9198
8	10-May-20...	119.9836	60	91.8789
9	10-May-20...	119.9836	59	91.8175

時間によるデータアクセス

```
starttime = datetime(2001,5,10,17,30,0);
endtime = datetime(2001,5,10,17,30,5);
t1hz(timerange(starttime,endtime),:)
```

可視化

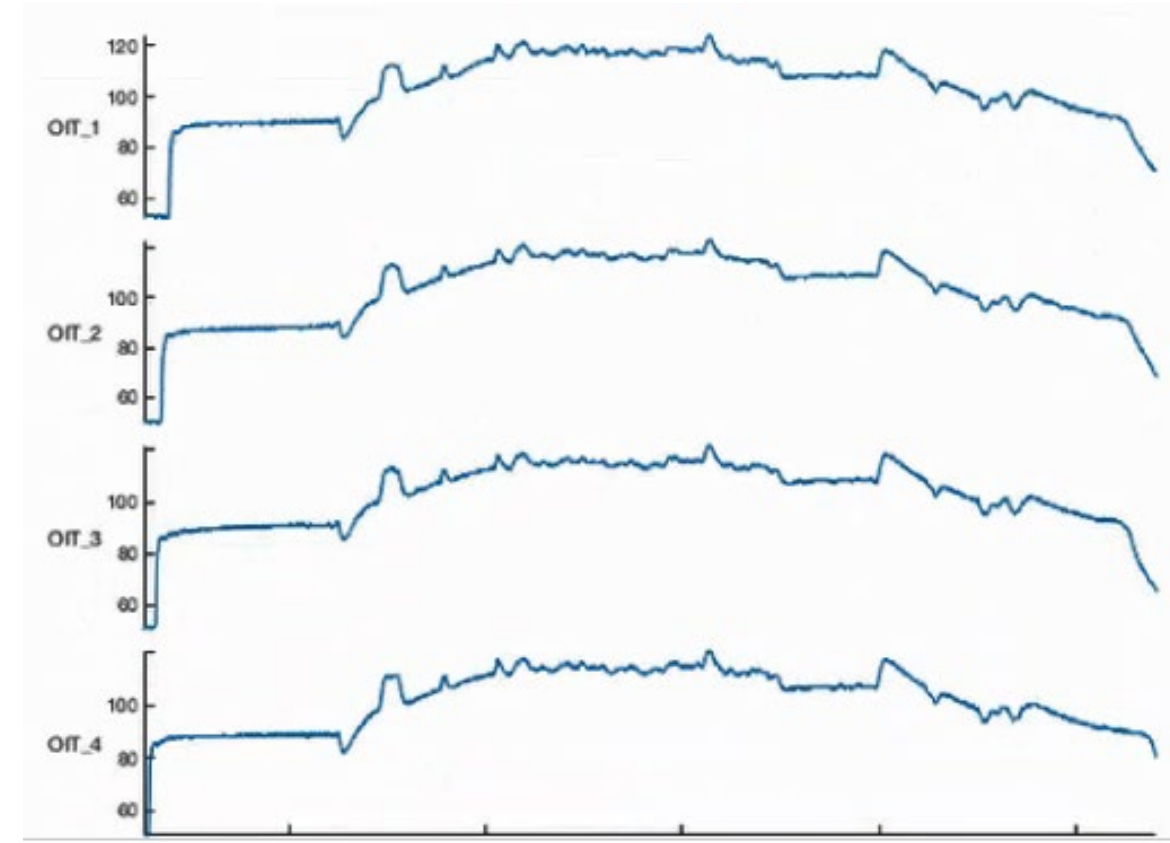
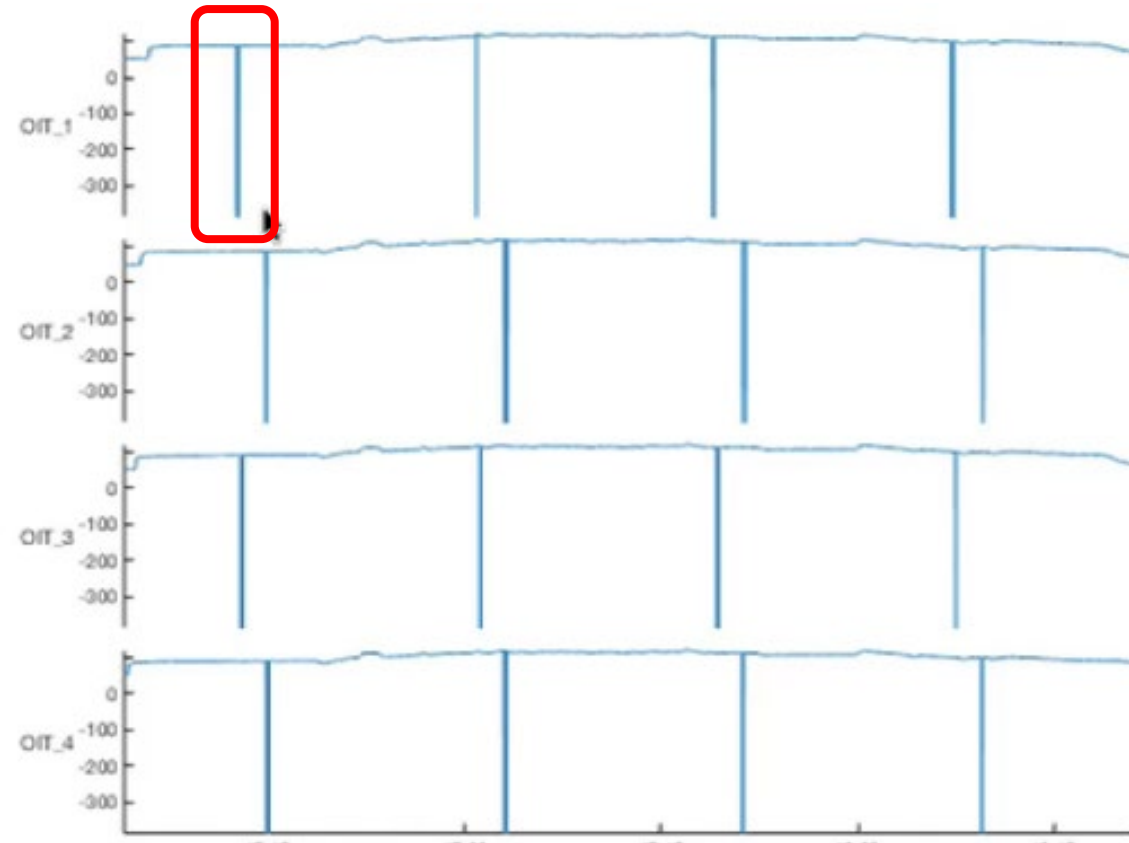
```
stackedplot(t1hz,{'OIT_1','OIT_2','OIT_3','OIT_4'});
```





異常値の削除・補間処理

```
filloutliers()
```

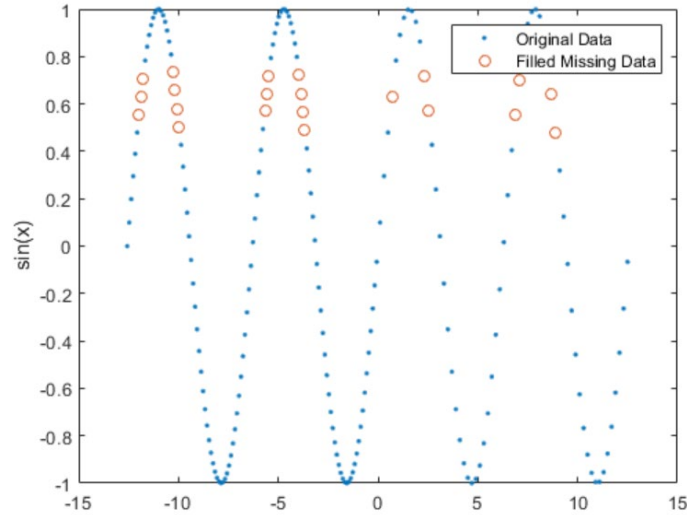




便利な前処理関数

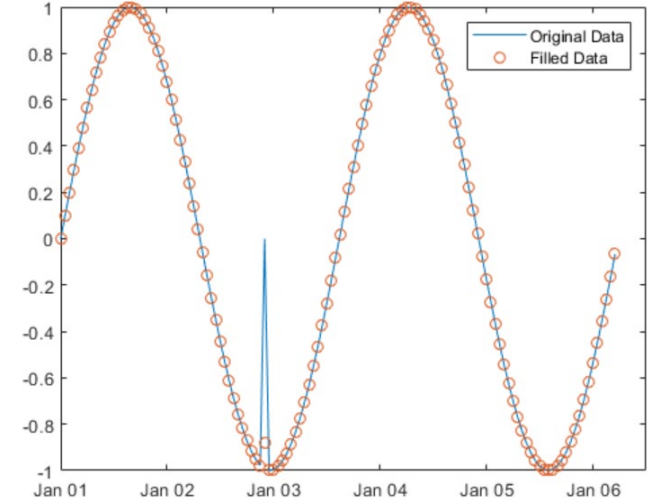
欠損値

ismissing
rmmissing
fillmissing



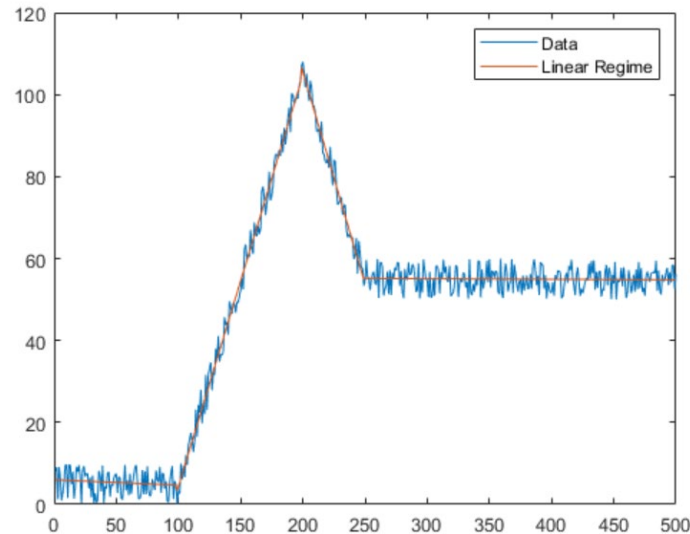
外れ値

isoutlier
rmoutliers
filloutliers



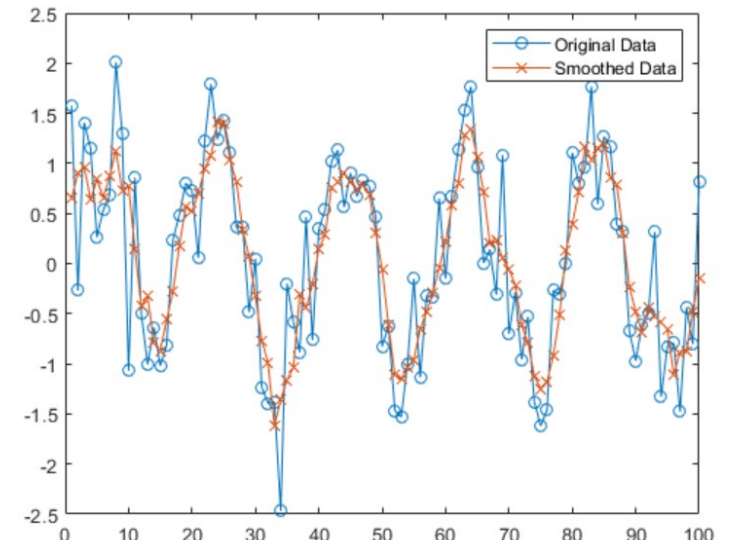
急激な変化点

ischange



平滑化

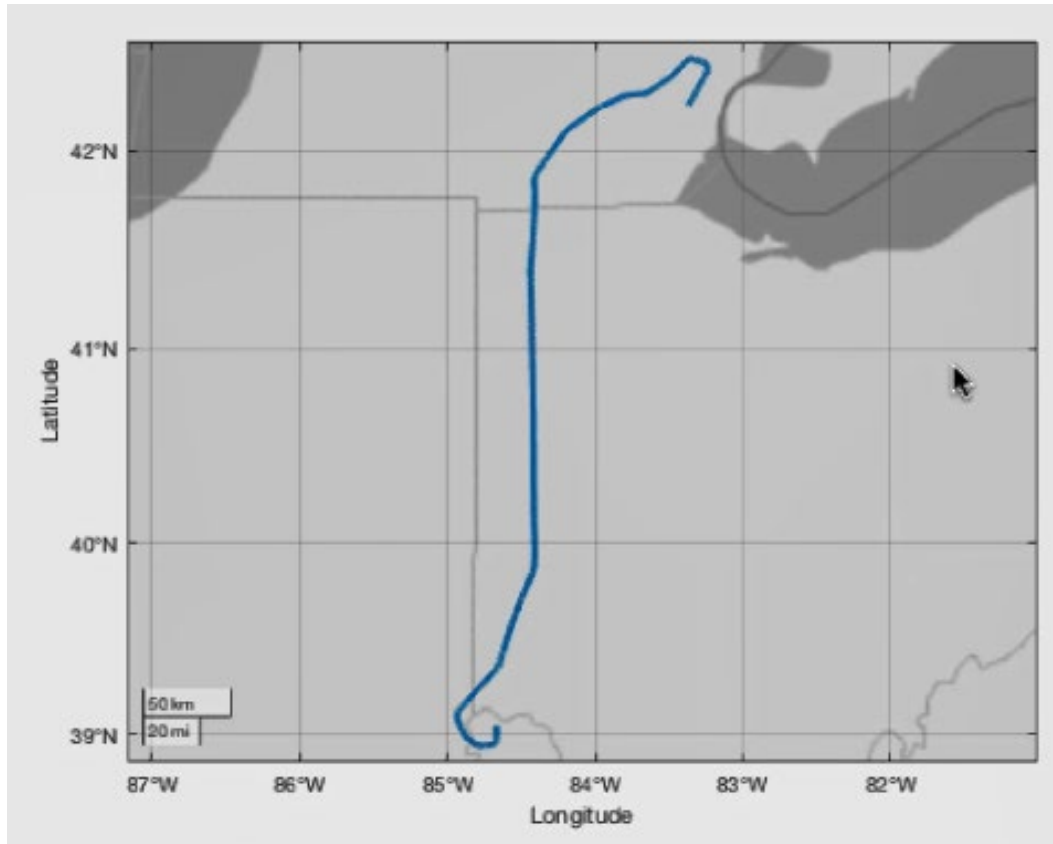
smoothdata



可視化

地図上に軌跡をプロット

```
geoplot(t.LATP(inFlight),t.LONP(inFlight),'LineWidth',3);
```



geoplot

geoscatter

geobubble

geodensityplot

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



ポイントまとめ



GUI で簡単操作・操作と等価の関数も1クリックで作成



行列数を使わないデータアクセス・データソート



時系列解析に便利なタイムテーブル



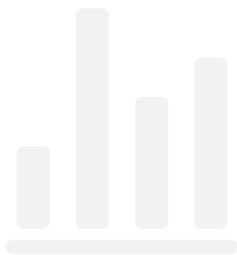
データの前処理は関数1行で実施



地理プロットで簡単可視化

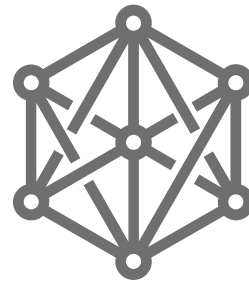
① アドホックな解析

- データの読み込み
- データのアクセス
- 前処理
- 可視化



② ビッグデータ解析

- ビッグデータの読み込み
- ビッグデータの解析・深掘り



フライトデータ概要

- 35 機
- 180,000 飛行
- 容量 300 GB
- 出典:
 - NASA Dash Link: Sample Flight Data
 - <https://c3.nasa.gov/dashlink/projects/85/>





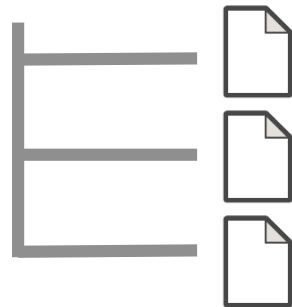
ビッグデータの読み取り

大規模なデータ集合の読取り

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



× 実ファイル
○ リポジトリ



⇒ PC の RAM メモリ以上の
データ読み取りが可能

大規模なデータを 1 変数で管理

```
t = read(ds);
```

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



データストアのカスタマイズ (カスタムデータストア)

自作のカスタムデータストアでデータストア読み込み

R2019a

```
>> ds = flightDataStore('/Volumes/Ext/Data/FlightData/675/')
```

```
ds =
```

```
flightDataStore with properties:
```

```
    CurrentFileIndex: 1
```

```
    NumberOfFiles: 5700
```

```
    VariableNames: {1×186 cell}
```

```
    SelectedVariableNames: {1×186 cell}
```

```
    AddFlightIdentifier: 0
```

オリジナルフォーマットのビッグデータ解析にも対応可能



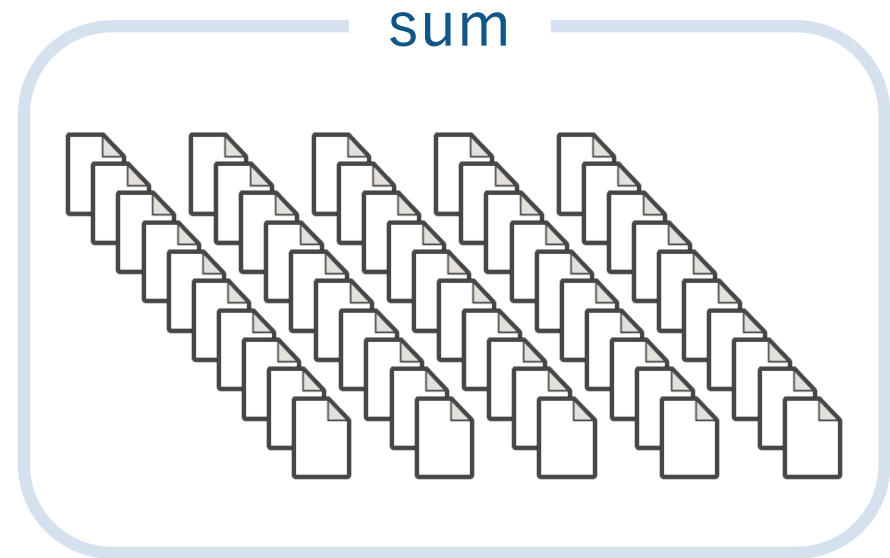
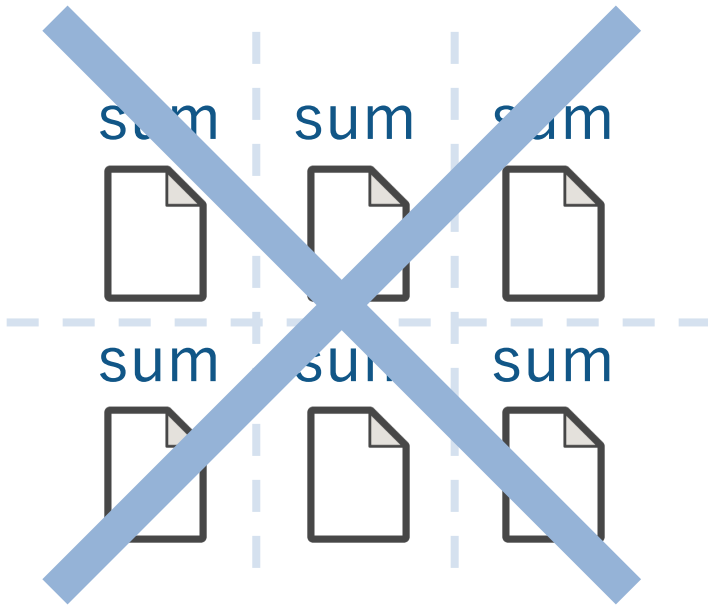
ポイントまとめ



データストアにより RAM サイズ以上の大容量データを一括読取り



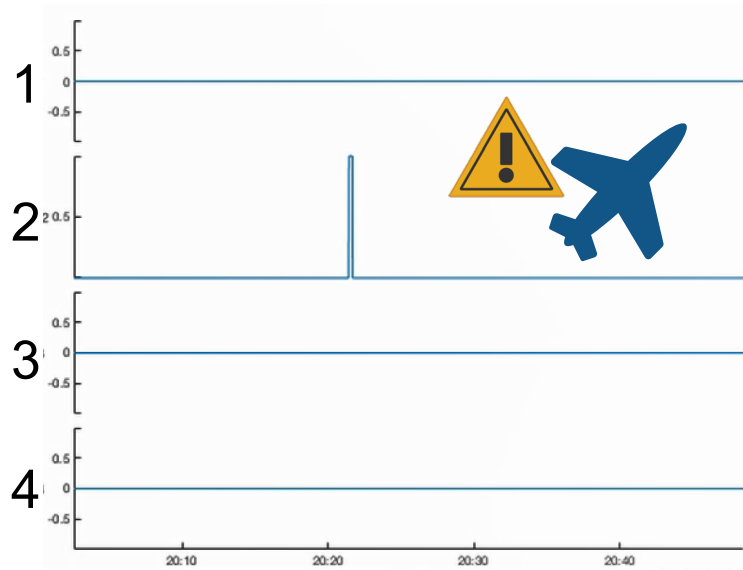
独自のファイルもデータストア化し一元操作



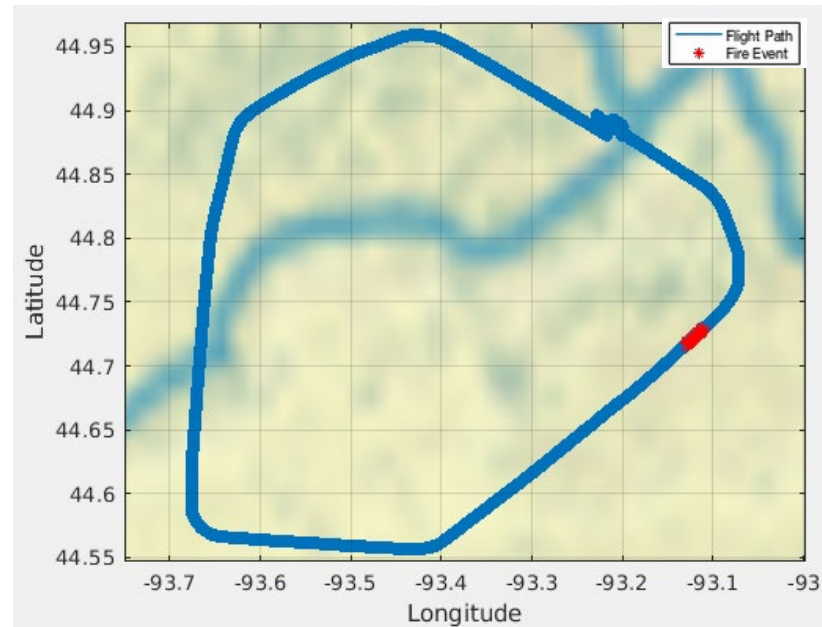


ビッグデータ活用例 – 事象の深掘り

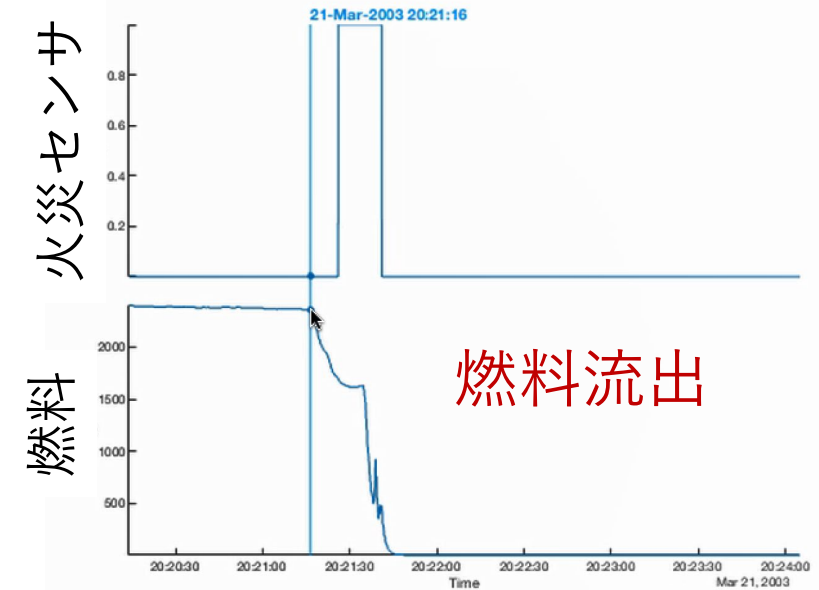
センサ2で火災を検知



フライト軌跡の確認



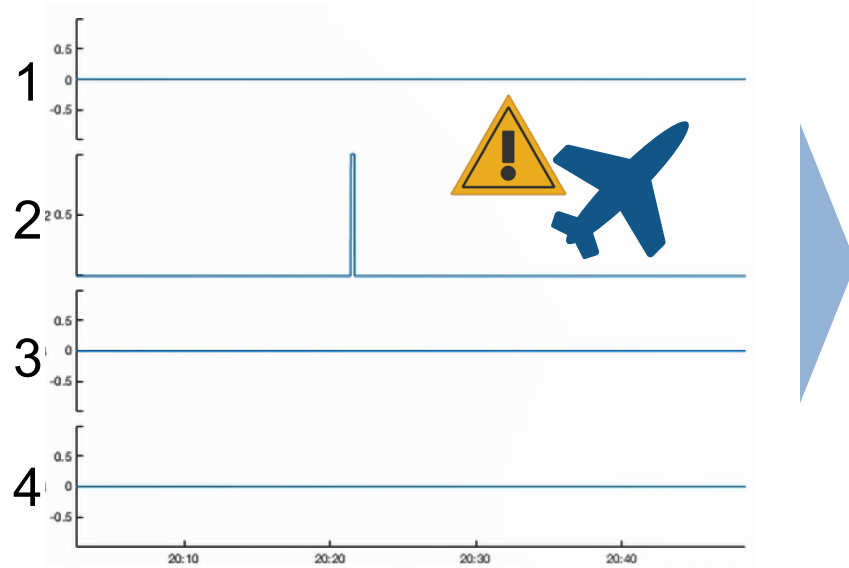
燃料データの確認



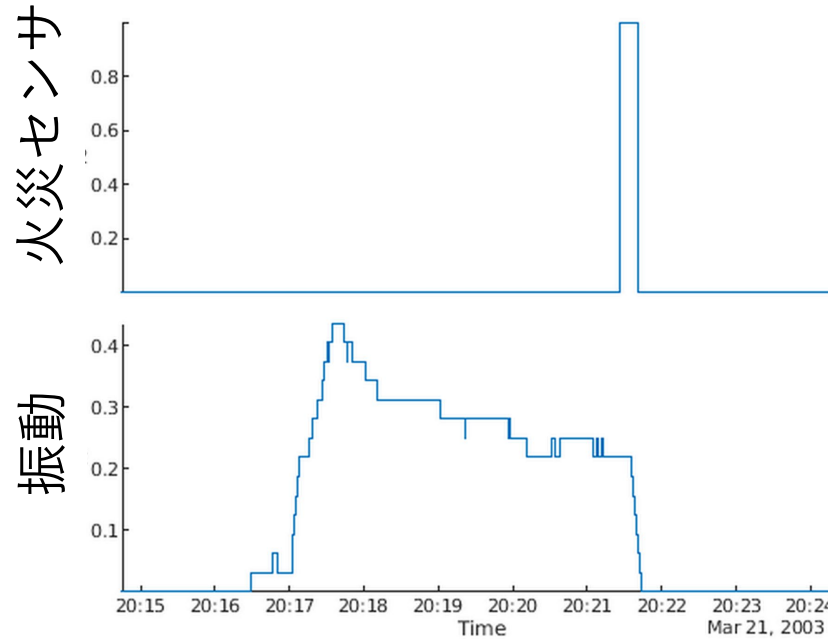


ビッグデータ活用例 – 事象の深掘り

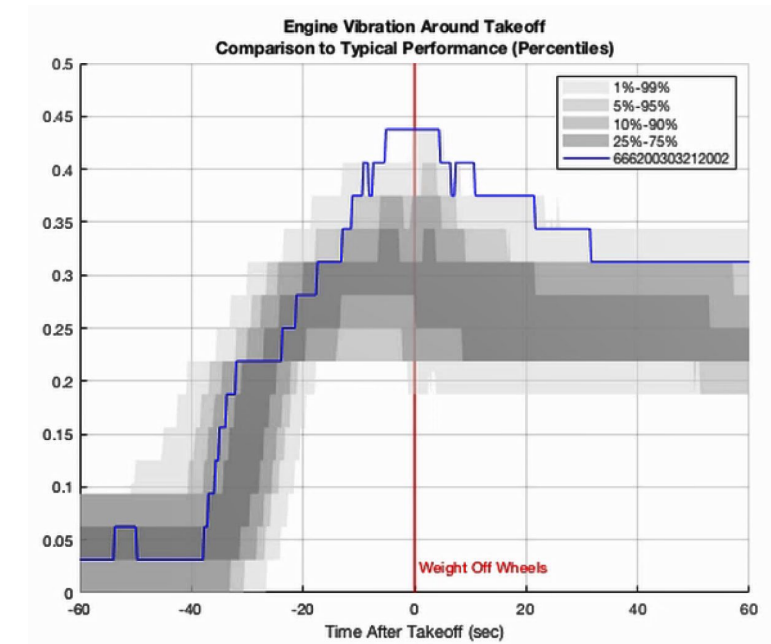
センサ2で火災を検知



振動データの確認



振動確率分布と比較

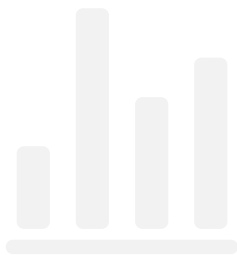


- 火災の原因をデータから検討
- 再発防止に役立つ知見を得る

振動による燃料流出が火災の原因か？

①アドホックな解析

- データの読み込み
- データのアクセス
- 前処理
- 可視化



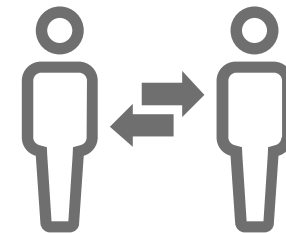
②ビッグデータ解析

- ビッグデータの読み込み
- ビッグデータの解析・掘り出し



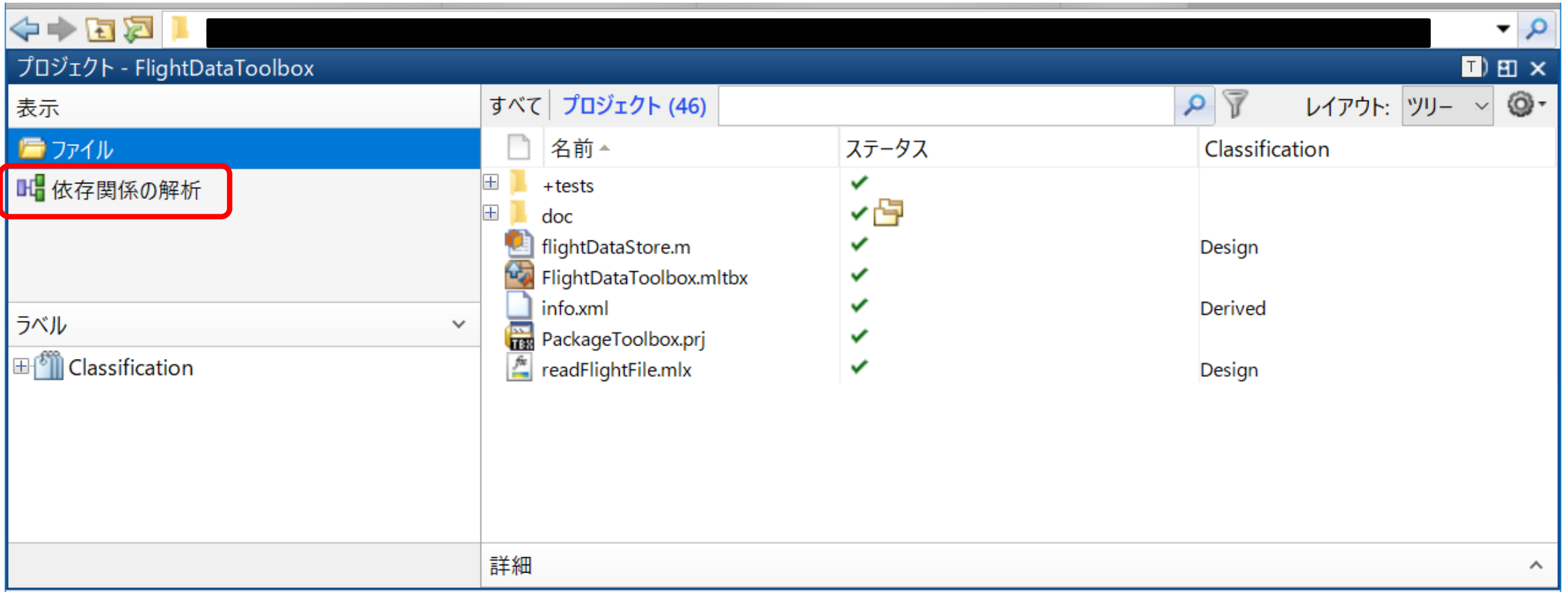
③プロジェクト化

- プロジェクトの作成・共有
- 大人数で解析



 プロジェクト化 **R2019a**

必要なファイルをまとめてプロジェクト化 → ファイルの一元管理



プロジェクト - FlightDataToolbox

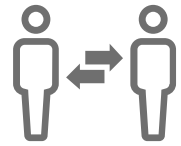
表示: すべて | プロジェクト (46) | レイアウト: ツリー

名前 ^	ステータス	Classification
+ tests	✓	
doc	✓	
flightDataStore.m	✓	Design
FlightDataToolbox.mltbx	✓	
info.xml	✓	Derived
PackageToolbox.prj	✓	
readFlightFile.mlx	✓	Design

ラベル: Classification

依存関係の解析

詳細



プロジェクトによる業務一元化 お役立ち機能 R2019a

ファイルの依存関係を可視化 → プロジェクトの簡単解読

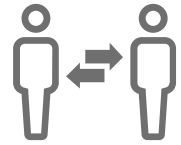
The screenshot displays the MATLAB R2019a interface with the 'Dependency Analysis' tool active. The main window shows a project named 'Airframe Example' with a dependency graph. The graph illustrates the relationships between various files, including MATLAB files, Simulink models, and C source files. The 'Dependency Type' legend indicates the following categories:

- Function (6)
- Class (1)
- Simulink Model (7)
- Simulink Library (1)
- C Source (1)
- Coder TLC Source (1)
- MEX-file (1)
- Simulink Data Dictionary (3)

The 'Impact View' on the right provides a summary of the project's dependencies:

- Airframe Example**
- Root: /.../airframe
- Analyzed: 3/6/19, 8:0...
- Products (3)
- MATLAB 9.7
- Simulink 9.3
- Simulink Coder 9.1
- External Toolboxes (1)
- C/C++ Library Header
- Problems (1)
- Not in project (1)

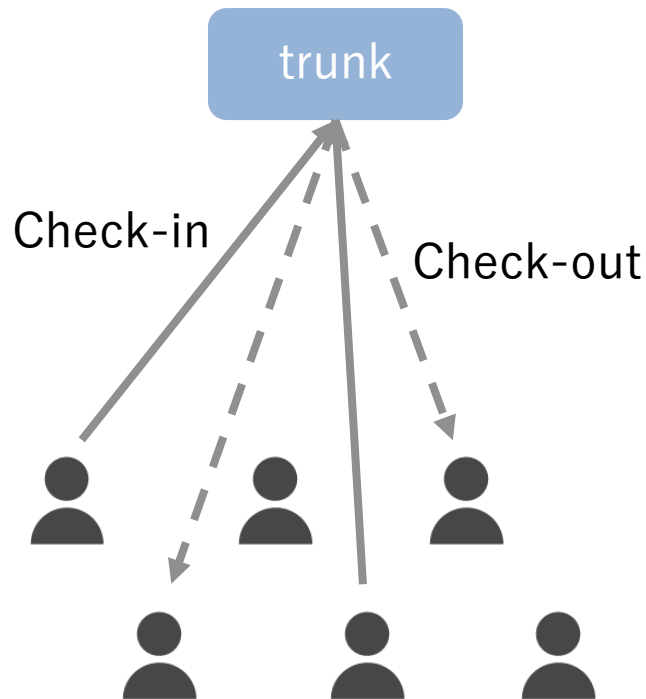
The interface also shows a 'Project - FlightDataToolbox' window with a 'Dependency Analysis' view, highlighting the 'flightDataStore.m' file and its dependencies on other files like 'testFlightDatastore.m', 'FleetWideFuelEconomy...', and 'SlicingFlightDataExample...'.



プロジェクトによる業務一元化 お役立ち機能

ソース管理 → 複数人での作業支援・バージョン管理・差分表示

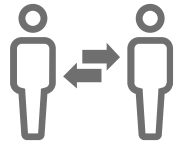
Subversion (SVN) Git との連携



```

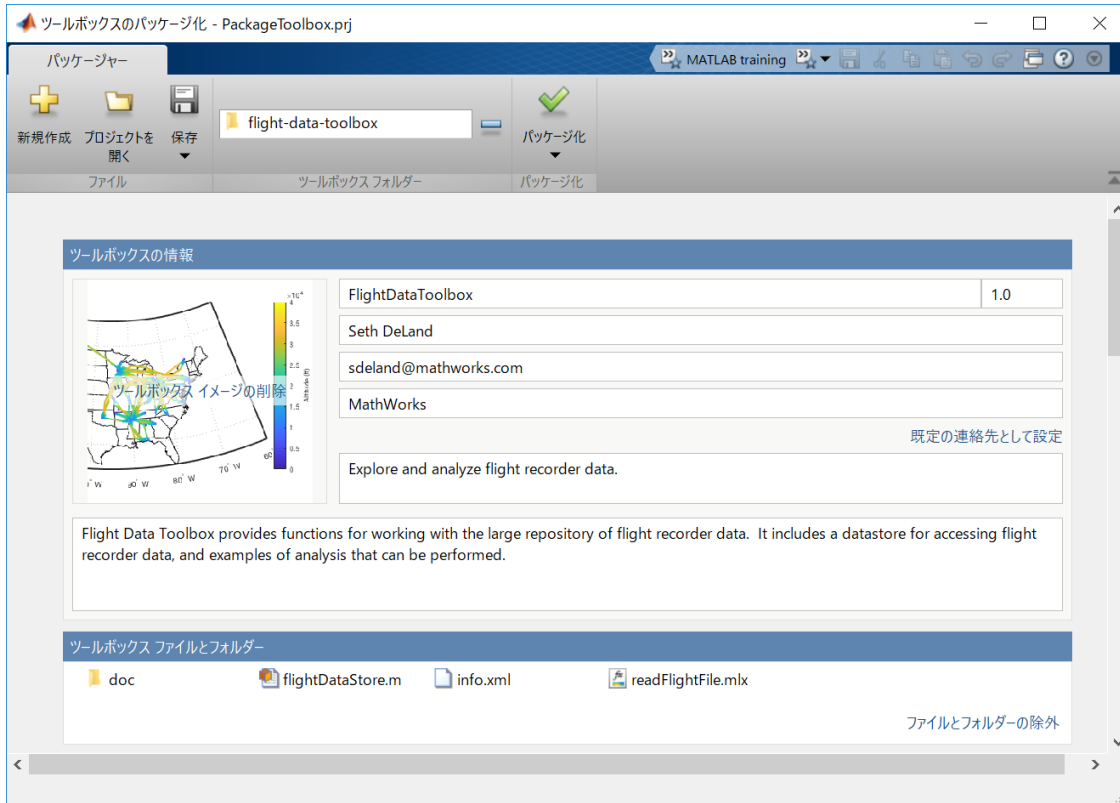
23 methods                                     . methods                                     23
24                                             .                                             24
25 function fds = flightDataStore(location)    . function fds = flightDataStore(location)    25
-                                             > %FLIGHTDATASTORE Dastore for reading flight data 26
-                                             > % DS = FLIGHTDATASTORE(LOCATION) creates a flightDataStore DS base 27
-                                             > % of the data. LOCATION is a folder that contains .MAT files 28
-                                             > % that contain flight data. Alternatively, LOCATION can be 29
-                                             > % an individual .MAT file of flight data. If LOCATION is a 30
-                                             > % folder, .MAT files are searched for recursively (there can 31
-                                             > % be multiple levels of folders). 32
-                                             > % 33
-                                             > % flightDataStore supports the following functions: 34
-                                             > % 35
-                                             > % read - Read one file's worth of data. 36
-                                             > % readall - Read all of the data. 37
-                                             > % hasdata - Returns true if there is more data. 38
-                                             > % reset - Reset the flightDataStore to the start of the data 39
-                                             > % signalMetaInfo - Return meta-data about each signal 40
-                                             > % 41
-                                             > % Example: 42
-                                             > % Create a flightDataStore 43
-                                             > % ds = flightDataStore('DataFolder') 44
-                                             > % We are only interested in the Altitude data 45
-                                             > % ds.SelectedVariableNames = {'ALT'} 46
-                                             > % Find the maximum altitude achieved 47
-                                             > % maxAlt = 0; 48
-                                             > % while hasdata(ds) 49
-                                             > %     tt = read(ds); 50
-                                             > %     maxAlt = max(maxAlt,max(tt.ALT)); 51
-                                             > % end 52
-                                             > % maxAlt 53
26                                             .                                             54
27 % Use the DsFileSet class to find and keep track of files . 55
28 fds.FileSet = matlab.io.datastore.DsFileSet(location,... . fds.FileSet = matlab.io.datastore.DsFileSet(location,... 56
[279 unmodified lines hidden]

```



検証済みツールとして共有 R2019a

ツールボックス



ツール説明書

Using a FlightDataStore

場所: sers/mtaguchi/AppData/Roaming/MathWorks/MATLAB%20Add-Ons/Toolboxes/FlightDataToolb

Using a FlightDataStore

All the Flight Data

```
ds = flightDataStore('/Users/sdeland/MATLAB/projects/FlightDataToolbox/+tests/sampleData/')

ds =
    flightDataStore with properties:
        CurrentFileIndex: 1
        NumberOfFiles: 6
        VariableNames: {1x186 cell}
        SelectedVariableNames: {1x186 cell}
        AddFlightIdentifier: 0
```

Data for a Subset of Signals

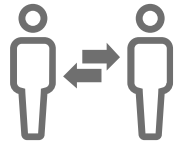
```
signalMetaInfo(ds)
```

ans = 186x5 table

	Signals	Rate	Units	Description	Alpha
1	'ABRK'	1.0000	'DEG'	'AIRBRAKE POS...	'ABRK'
2	'ACID'	0.2500	"	'AIRCRAFT NUM...	'ACID'
3	'ACMT'	1.0000	"	'ACMS TIMING ...	'ACMT'
4	'AIL_1'	1.0000	'DEG'	'AILERON POSI...	'AIL_1'
5	'AIL_2'	1.0000	'DEG'	'AILERON POSI...	'AIL_2'
6	'ALT'	4.0000	'FEET'	'PRESSURE ALT...	'ALT'
7	'ALTR'	4.0000	'FT/MIN'	'ALTITUDE RATE'	'ALTR'
8	'ALTS'	1.0000	'FEET'	'SELECTED ALT...	'ALTS'
9	'AOA1'	4.0000	'DEG'	'ANGLE OF ATT...	'AOA1'
10	'AOA2'	4.0000	'DEG'	'ANGLE OF ATT...	'AOA2'

```
ds.SelectedVariableNames = {'LATP', 'LONP'}
```

```
ds =
    flightDataStore with properties:
```



ポイントまとめ



MATLAB プロジェクトによるファイルの一元管理



プロジェクト可視化機能による簡単解読



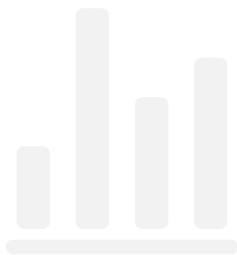
ソース管理機能連携により複数人で作業が可能



ツールボックス化し配布・さらなる協業を加速

①アドホックな解析

- データの読み込み
- データのアクセス
- 前処理
- 可視化



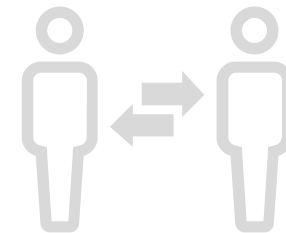
②ビッグデータ解析

- ビッグデータの読み込み
- ビッグデータの解析・掘り

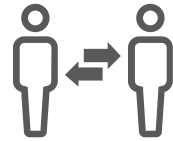


③プロジェクト化

- プロジェクトの作成・共有
- 大人数で解析



全体のまとめ



- データ解析のカギを握る前処理 MATLAB で効率良く実施可能
- MATLAB は大容量データ解析に対応、ビッグデータ活用の橋渡し
- プロジェクト化により組織横断プロジェクトへ拡大

本日はご紹介したツール - MATLAB **R2019a**