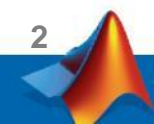


Data Analytics Workshop

Terasoft
Application Engineer
Jeffrey Liu

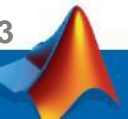
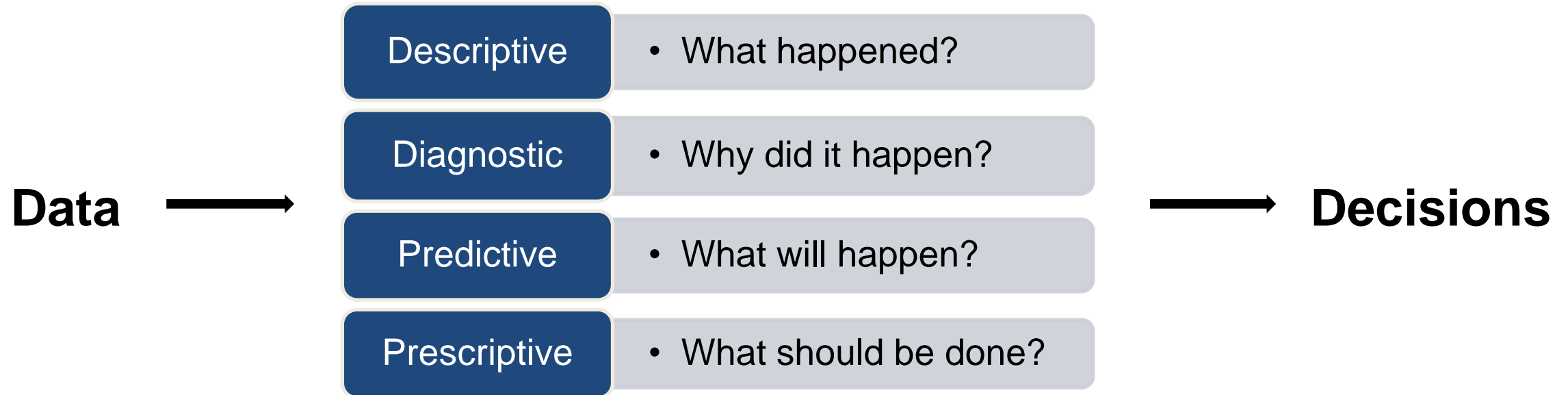
Outlines

- Introducing the Workflow of Data Analytics
 - From Data to actionable insights
- Data import and preprocessing
 - Import data from external files
 - Understanding the data and preprocessing
 - Merging tables (exercise 1)
 - Locating and removing missing data (exercise 2)
 - Grouping statistics (exercise 3)
- Train a prediction model
 - Classification Learner App (exercise 4)
 - Hand write code

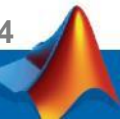
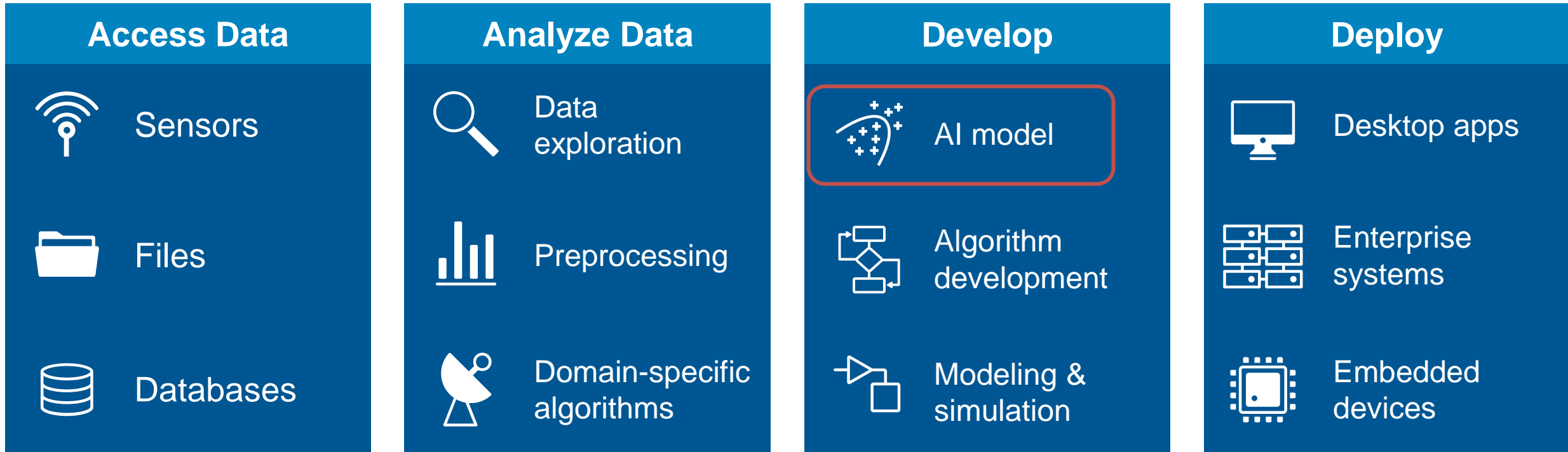


What is Data Analytics?

Turn large volumes of complex data into actionable information



Basic Workflow in most of AI / Data Analytics Journey

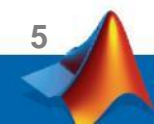


What is Table?

	A	B	C	D	E	F	G	H	I	J	K	L
1	Date	AAL	ABF	ADM	ADN	AGK	ARM	AV	AZN	BA	BAB	BARC
2	1/3/2006	2241.76	840.5	466.75	132.75	295.74	121.25	708	2870	386.5	233	617.5
3	1/4/2006	2251.65	857.5	481	133	291.1	121.5	726	2900	395.75	235	627
4	1/5/2006	2193.41	855	477.25	135	292.9	121.5	725.5	2896	393.5	233.5	621.5
5	1/6/2006	2235.17	857	477.25	135.5	289.03	124.5	733.5	2900	408.5	238	627.5
6	1/9/2006	2212.09	862	478.25	138	283.1	123.75	723.5	2878	405	233.5	628
7	1/10/2006	2141.76	858.5	473.25	138.25	281.29	124.75	715	2835	401.25	236.25	619
8	1/11/2006	2168.13	855	495	138.5	283.61	129.25	721	2864	406.75	239.75	626.5
9	1/12/2006	2197.8	838	503	140.75	284.13	129.75	7				628.5
10	1/13/2006	2187.91	825	503	141.5	280.77	131.5	72				621
11	1/16/2006	2192.31	825	498.5	144.75	287.74	131.25	7				620.5
12	1/17/2006	2212.09	820.5	494	145	278.71	130	7				609
13	1/18/2006	2178.02	807.5	494	144.5	279.74	127.75	7				601.5
14	1/19/2006	2205.5	802.5	506	150	284.13	131	7				605
15	1/20/2006	2215.39	826.5	505.75	153	290.84	129	710	2672	421	231.25	597.5
16	1/23/2006	2205.5	820	508.75	159.75	287.48	129.25	707.5	2646	421.5	228.5	596
17	1/24/2006	2182.42	816.5	500.75	162.75	289.03	126.75	702	2609	418.75	229.75	587.5
18	1/25/2006	2278.02	825.5	512	165	289.55	128.75	706	2673	418	238	587.5
19	1/26/2006	2330.77	830	511.5	163.75	289.03	133.25	719	2677	413.25	236	601
20	1/27/2006	2403.3	832.5	520	167.5	293.93	138.25	733	2721	426	244.5	612.5
21	1/30/2006	2379.12	824					725	2704	416.75	248.75	601
22	1/31/2006	2373.63	825					720.5	2722	417	247.75	601
23	2/1/2006	2418.68	827					727	2734	432.5	246.5	614.5
24	2/2/2006	2381.32	830					718	2638	426	248	611


**Each row
is a set of
observations**


**Each column is a
named variable**



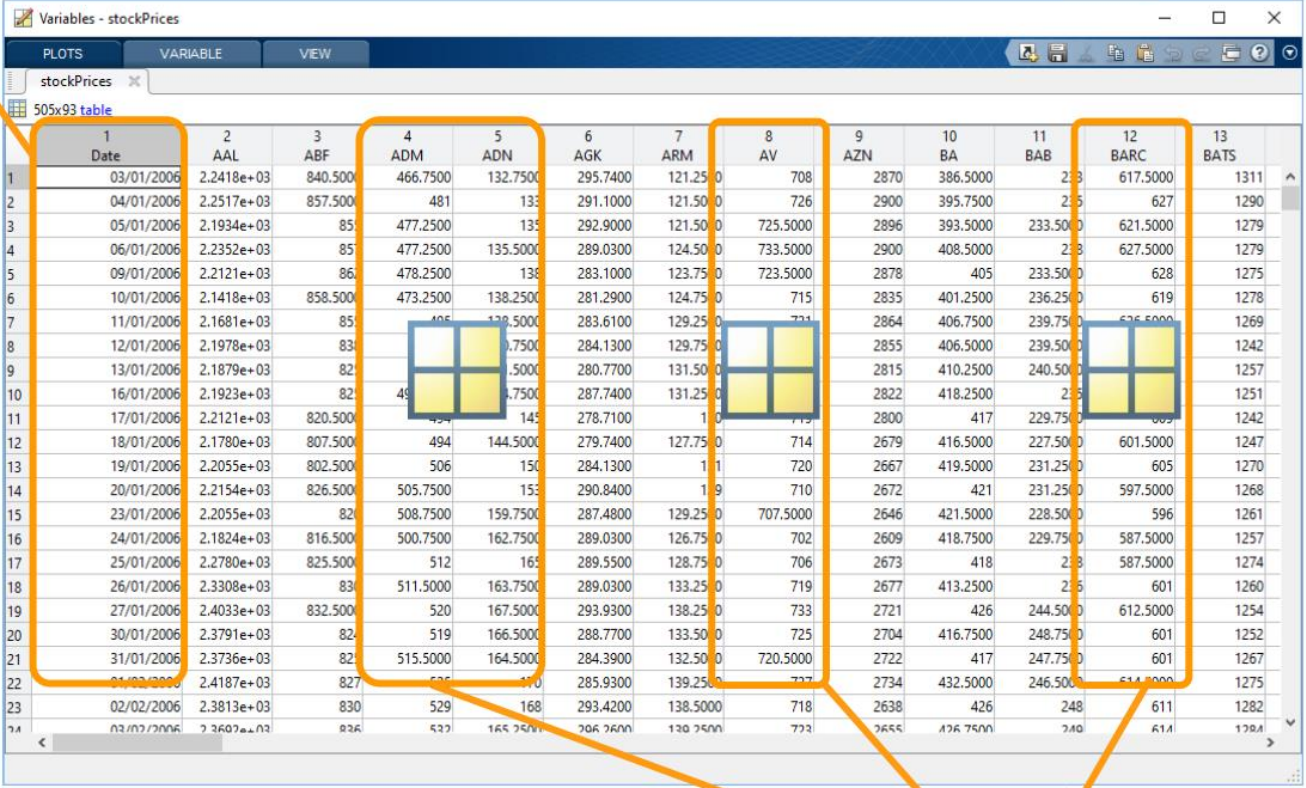
Extracting Data From Tables

`dates = stockPrices.Date;`

 **505-by-1**

 **505-by-4**

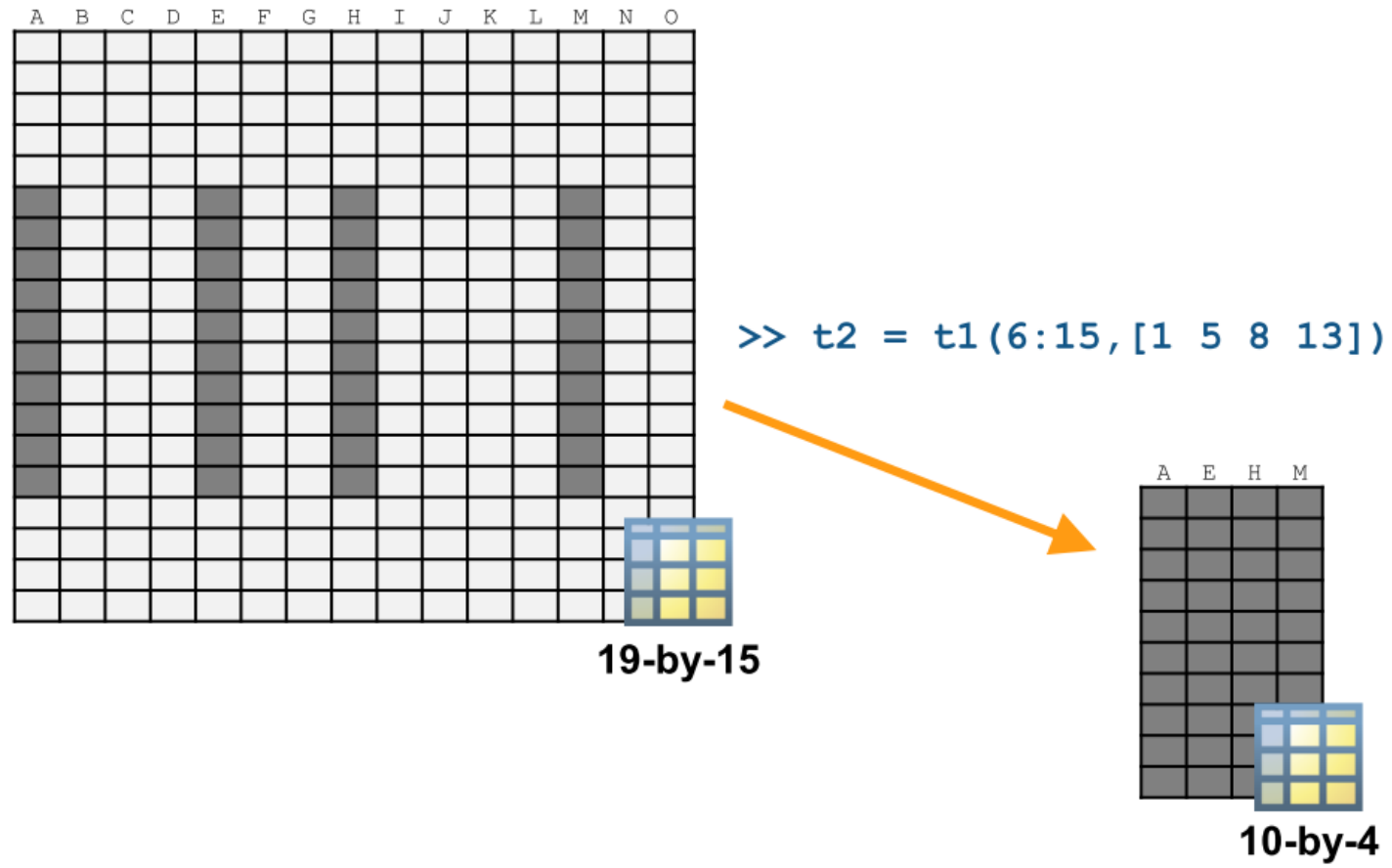
505-by-93



1	2	3	4	5	6	7	8	9	10	11	12	13
Date	AAL	ABF	ADM	ADN	AGK	ARM	AV	AZN	BA	BAB	BARC	BATS
03/01/2006	2.2418e+03	840.500	466.7500	132.7500	295.7400	121.25 0	708	2870	386.5000	233	617.5000	1311
04/01/2006	2.2517e+03	857.500	481	133	291.1000	121.50 0	726	2900	395.7500	235	627	1290
05/01/2006	2.1934e+03	85	477.2500	133	292.9000	121.50 0	725.5000	2896	393.5000	233.50 0	621.5000	1279
06/01/2006	2.2352e+03	85	477.2500	135.5000	289.0300	124.50 0	733.5000	2900	408.5000	233	627.5000	1279
09/01/2006	2.2121e+03	86	478.2500	138	283.1000	123.75 0	723.5000	2878	405	233.50 0	628	1275
10/01/2006	2.1418e+03	858.500	473.2500	138.2500	281.2900	124.75 0	715	2835	401.2500	236.25 0	619	1278
11/01/2006	2.1681e+03	85	405	138.5000	283.6100	129.25 0	733	2864	406.7500	239.75 0	616.5000	1269
12/01/2006	2.1978e+03	83	405	138.5000	284.1300	129.75 0	733	2855	406.5000	239.50 0	616.5000	1242
13/01/2006	2.1879e+03	82	405	138.5000	280.7700	131.50 0	733	2815	410.2500	240.50 0	616.5000	1257
16/01/2006	2.1923e+03	82	405	138.5000	287.7400	131.25 0	733	2822	418.2500	235	616.5000	1251
17/01/2006	2.2121e+03	820.500	494	145	278.7100	131.25 0	713	2800	417	229.75 0	609	1242
18/01/2006	2.1780e+03	807.500	494	144.5000	279.7400	127.75 0	714	2679	416.5000	227.50 0	601.5000	1247
19/01/2006	2.2055e+03	802.500	506	150	284.1300	131.25 0	720	2667	419.5000	231.25 0	605	1270
20/01/2006	2.2154e+03	826.500	505.7500	153	290.8400	131.25 0	710	2672	421	231.25 0	597.5000	1268
23/01/2006	2.2055e+03	82	508.7500	159.7500	287.4800	129.25 0	707.5000	2646	421.5000	228.50 0	596	1261
24/01/2006	2.1824e+03	816.500	500.7500	162.7500	289.0300	126.75 0	702	2609	418.7500	229.75 0	587.5000	1257
25/01/2006	2.2780e+03	825.500	512	165	289.5500	128.75 0	706	2673	418	233	587.5000	1274
26/01/2006	2.3308e+03	83	511.5000	163.7500	289.0300	133.25 0	719	2677	413.2500	235	601	1260
27/01/2006	2.4033e+03	832.500	520	167.5000	293.9300	138.25 0	733	2721	426	244.50 0	612.5000	1254
30/01/2006	2.3791e+03	82	519	166.5000	288.7700	133.50 0	725	2704	416.7500	248.75 0	601	1252
31/01/2006	2.3736e+03	82	515.5000	164.5000	284.3900	132.50 0	720.5000	2722	417	247.75 0	601	1267
01/02/2006	2.4187e+03	827	525	170	285.9300	139.25 0	733	2734	432.5000	246.50 0	614.5000	1275
02/02/2006	2.3813e+03	830	529	168	293.4200	138.5000	718	2638	426	248	611	1282
03/02/2006	2.3607e+03	826	523	165	296.7600	139.25 0	723	2655	426.7500	248	614	1284

`finance = stockPrices{:, [4, 5, 8, 12]};`

Indexing into Tables



INTRO to the Table Data Type

“Working with Financial Data in MATLAB”

Variables - MidCap

MidCap

314x145 table

	1 Dates	2 ABCO	3 ACGL	4 ACHC	5 ACIW	6 AFG	7 AGCO	8 AIRM	9 AIT	10 AJG	11 ALGN	12 ALKS
1	733409	30.5900	23.1100	10.4000	16.2800	24.1400	67.6600	15.2300	22.9100	17.6700	16.5000	14.7900
2	733414	31.3000	23.7600	10.2000	14.3200	24.6400	63.3500	15.0700	21.7600	17.9800	14.7700	15.3800
3	733421	31.2200	22.4400	10.6000	12.5600	23.0300	57.4000	13.9300	20.8100	17.7000	13.4000	14.2000
4	733429	30.6500	22.5600	10.1200	13.3700	23.0700	56.3200	13.9900	24.2700	18.6000	13.8700	13.6100
5	733435	32.1200	23.4600	9.8000	16.1700	24.3500	61.7100	14.8000	26.3100	19.7300	12.0600	13.7000
6	733442	27.5000	23.2800	10.7200	15.2300	23.6400	56.3300	14.4100	24.6600	18.8200	13.0100	13.5700
7	733449	27.9600	23.6100	11	15.6700	23.7600	64.0400	13.7300	24.1200	18.1500	13.0200	12.8900
8	733457	27.9200	23.3400	11	16.5600	23.4200	64.2000	12.6100	25.7900	18.2400	12.8200	13.0400
9	733463	27.7800	22.8300	11.2000	17.6600	22.3800	64.4000	12.8400	24.0900	17.9400	12.3500	12.9400
10	733470	27.2300	22.7100	11.2000	19.6300	21.7700	62.0400	12.7700	24.3600	18.0800	11.1000	11.2600
11	733477	26.2300	22.1600	10.8000	18.0300	21.4400	59.2100	13.8400	23.1800	18.2300	10.9200	11.1200
12	733484	27.1700	22.5600	11.0400	18.0200	21.0100	53.0700	14.6700	25.5400	18.8400	11.2000	11.0000

Workspace

Name

MidCap

Command History

plot(Pr
plot(Pr
pointfi
boxplot
glyphpl
quiver(
stem3(P
plot3(P

Command Window

New to MATLAB? Watch this [Video](#), see [Examples](#), or read [Getting Started](#).

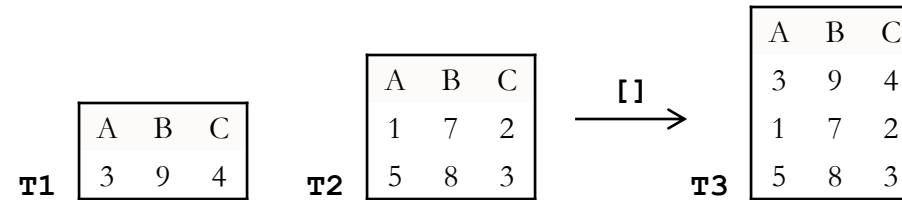
fx >>

```
>> doc readtable  
>> T = table()  
>> methods(T)  
>> doc table
```

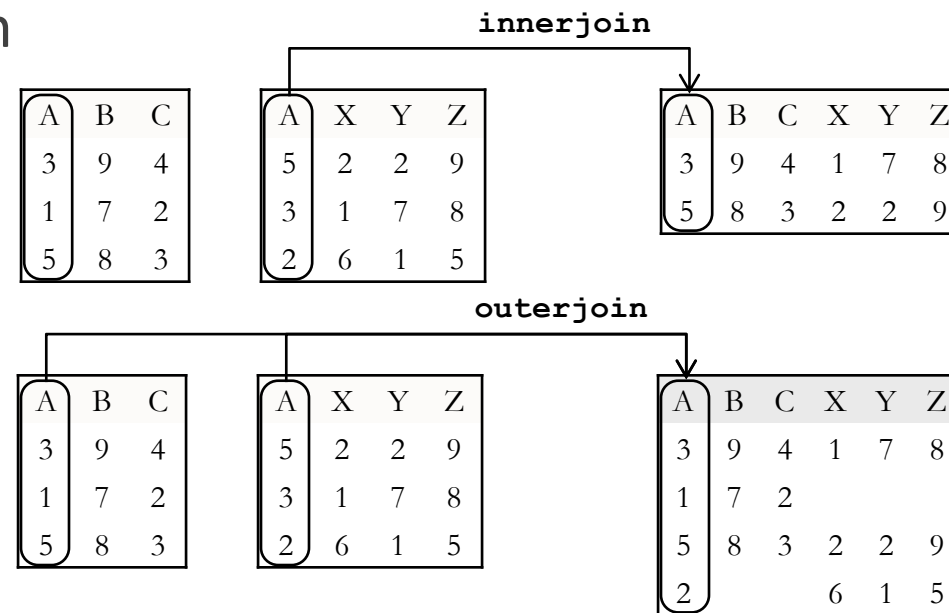
- Tables are data containers that allow for working with mixed datatypes
- **Methods:** join, innerjoin, outerjoin, write, sortrows, width, height, etc.

Merging Data

- $T3 = [T1; T2];$



- Join, innerjoin, outerjoin

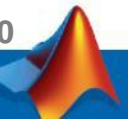


Exercise 1

- 將join_ex.xlsx中的兩張表格(TableA, TableB)依照”Date”欄位左右合併。
- Use “innerjoin”, “outerjoin” to combine “table” variables
 - Import files with table array

```
>> Variables = readrable(date, data, 'var_name');
```
 - Merge two tables
 - innerjoin, outerjoin

```
>> Combined_table = outerjoin(TblA, TblB, 'Keys', 'Date', 'MergeKeys', true);
```



Dealing with Missing Values

3
NaN
NaN
2
4
NaN
1
6

Ignore

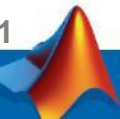
3
2
4
1
6



Delete

3
2.67
2.33
2
4
2.50
1
6

Replace



Avoiding NaNs in Calculations

- Several functions are designed to ignore NaNs in calculations

`nancov`
`nanmax`
`nanmean`

`nanmedian`
`nanmin`
`nanstd`

`nansum`
`nanvar`

- if a column contains all NaNs, ignoring them will result in applying the desired function to an empty array.

3	2	4	NaN
1	NaN	2	1
NaN	NaN	6	1

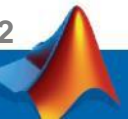
↓ `nanmean`

2	2	4	1
---	---	---	---

3	NaN	4	NaN
1	NaN	2	1
NaN	NaN	6	1

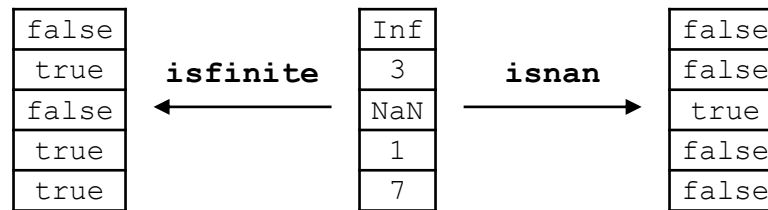
↓ `nanmean`

2	NaN	4	1
---	-----	---	---



Locating Missing(Other) Values

- MATLAB provides numerous “is*” functions that take an array as input and return a logical output that signifies if the input has a certain characteristic.



- You can use logical indexing to remove elements from an array.

```
>> idx = isnan(x);
```

```
>> x(idx) = [];
```

3
1
7

x



3	NaN	4	NaN
1	NaN	2	1
NaN	NaN	6	1

3	NaN	4	NaN
1	NaN	2	1
NaN	NaN	6	1

Numerical comparison

```
>> x == NaN
```

Not a Number!

F	F	F	F
F	F	F	F
F	F	F	F



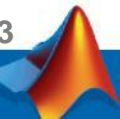
```
>> isnan(x)
```

F	T	F	T
F	T	F	F
T	T	F	F



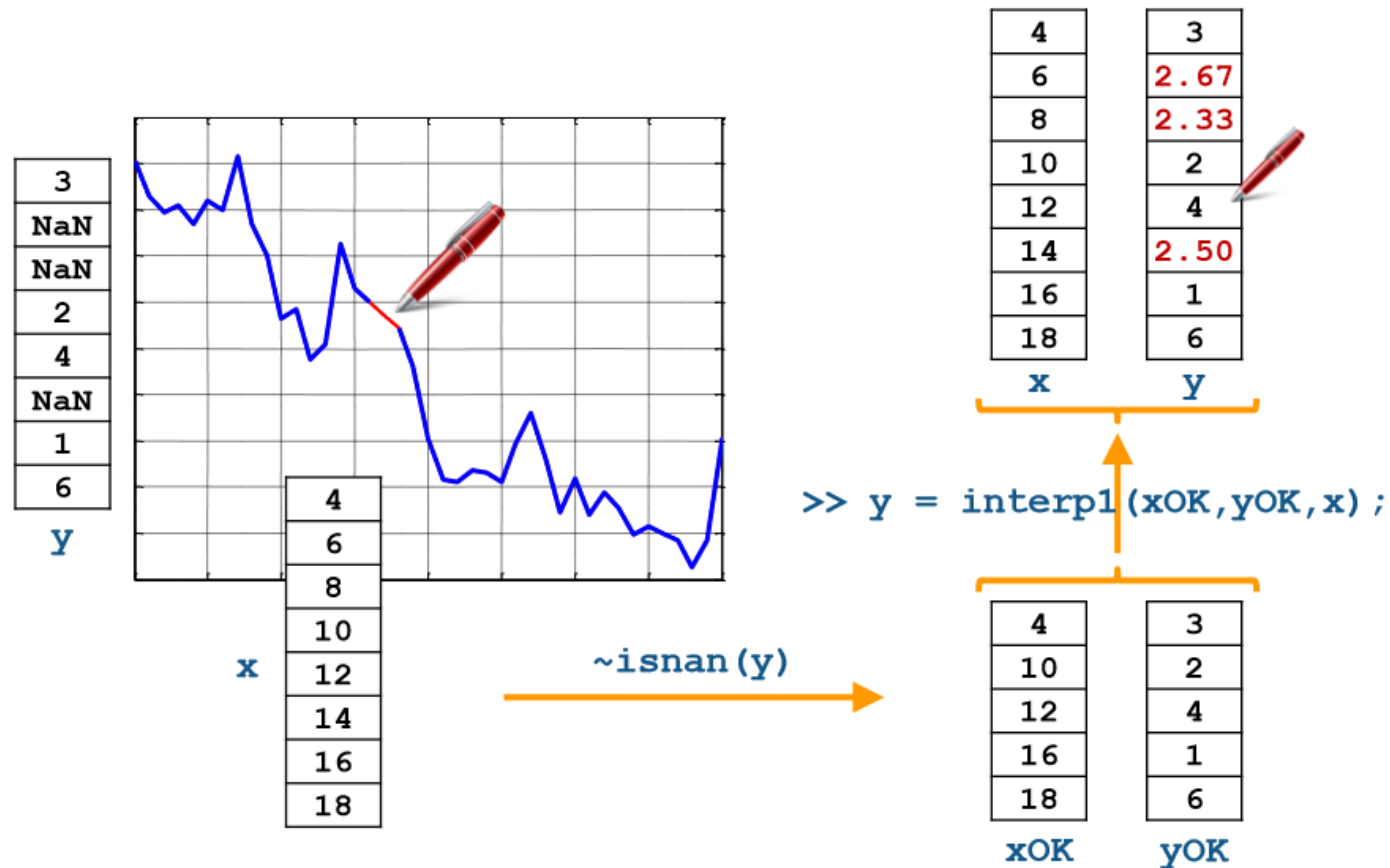
```
>> all(isnan(x))
```

F	T	F	F
---	---	---	---



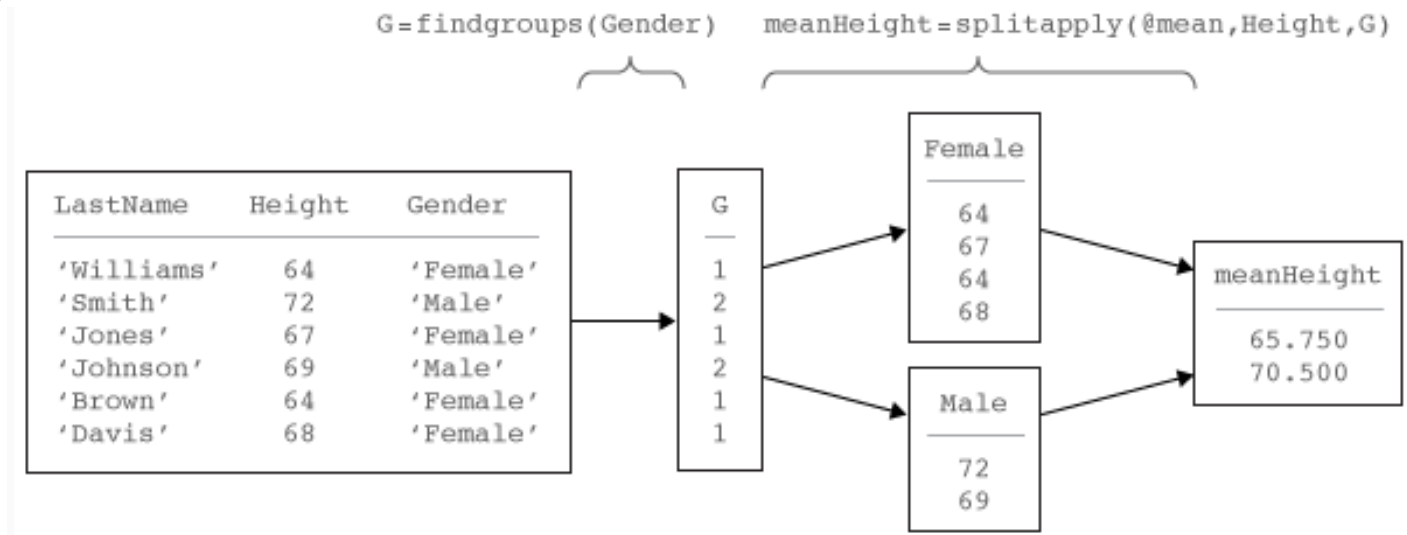
Exercise 2

- 依照下圖宣告變數並得到內插法後結果“y”



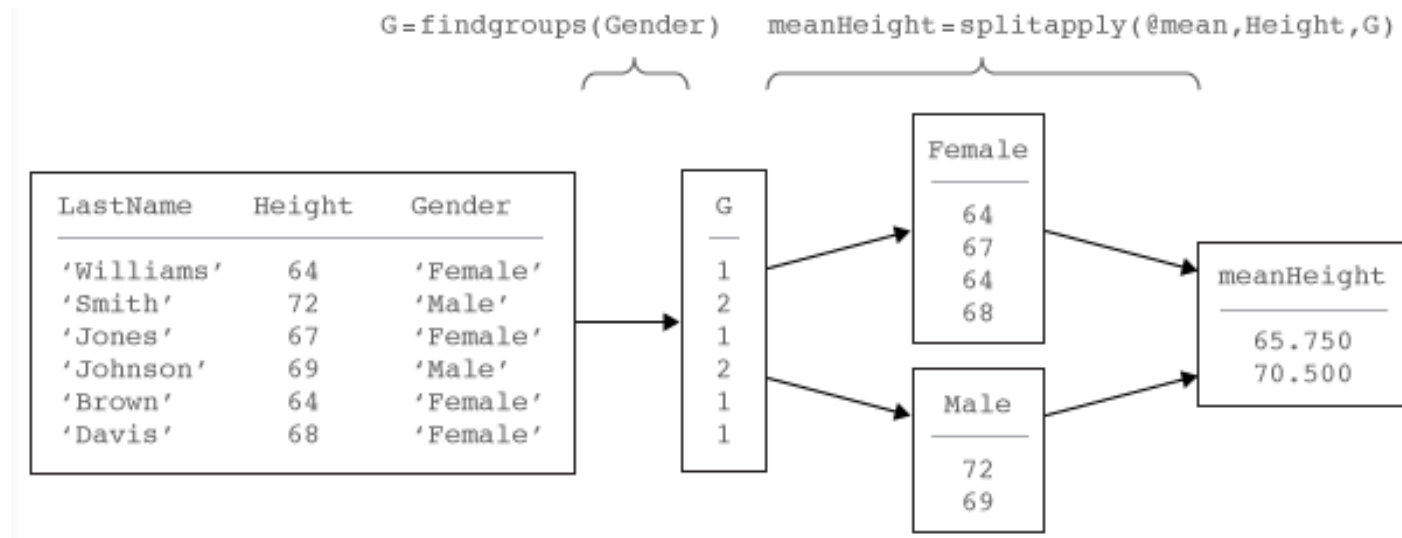
Grouping statistics

- Function 'grpstats' : Summary statistics organized by group
 - Syntax : `statarray = grpstats(tbl,groupvar,whichstats)`
tbl : data in table or dataset array
groupvar : Column name for grouping in tbl
whichstats : Types of summary statistics (numel, std, max, sum...)
- New Split-apply workflow

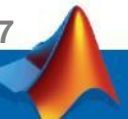
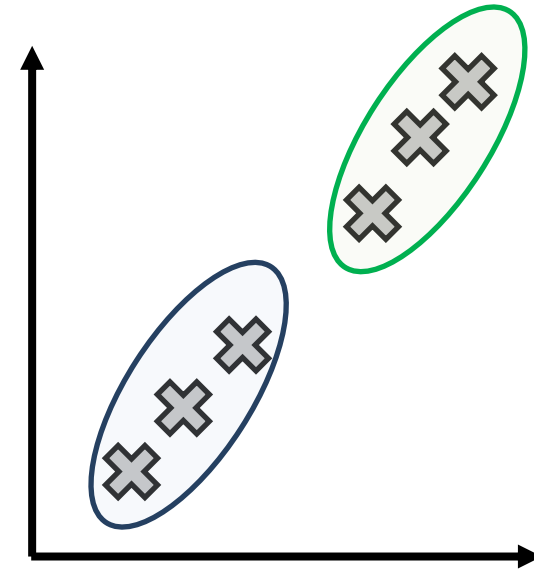
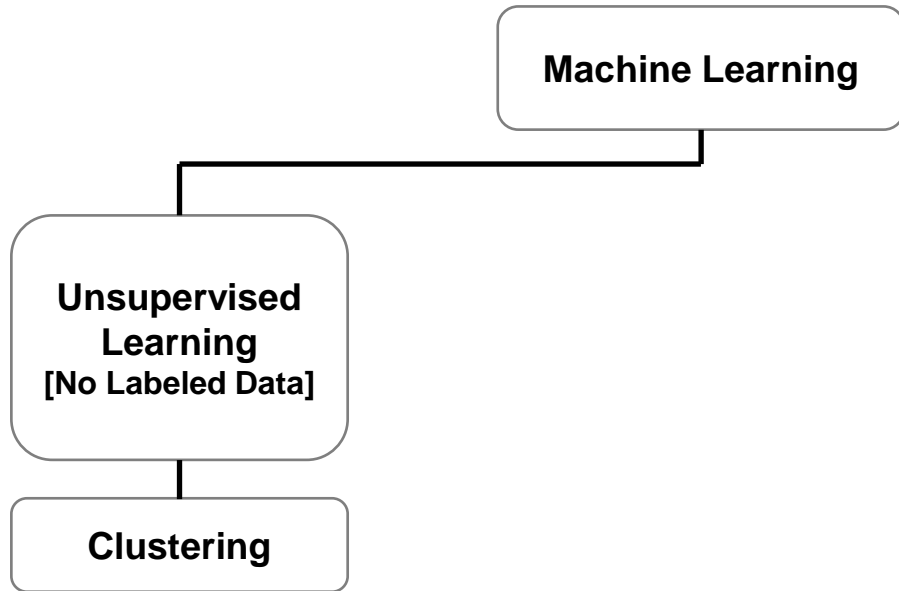


Exercise 3

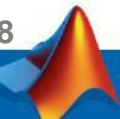
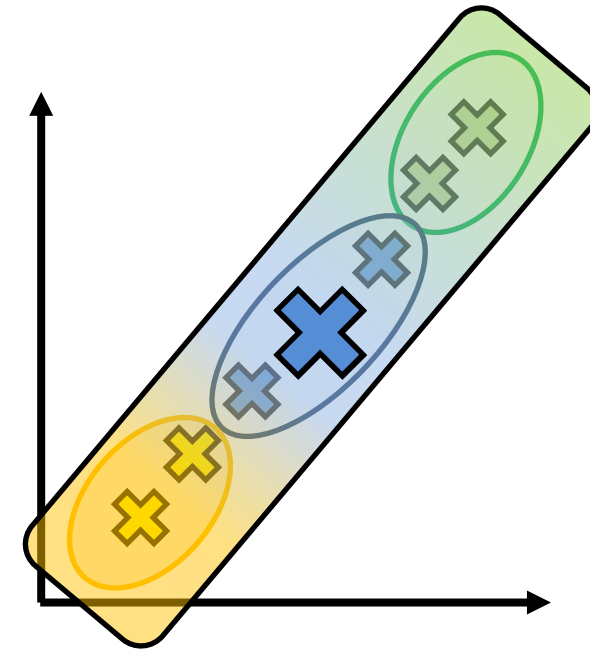
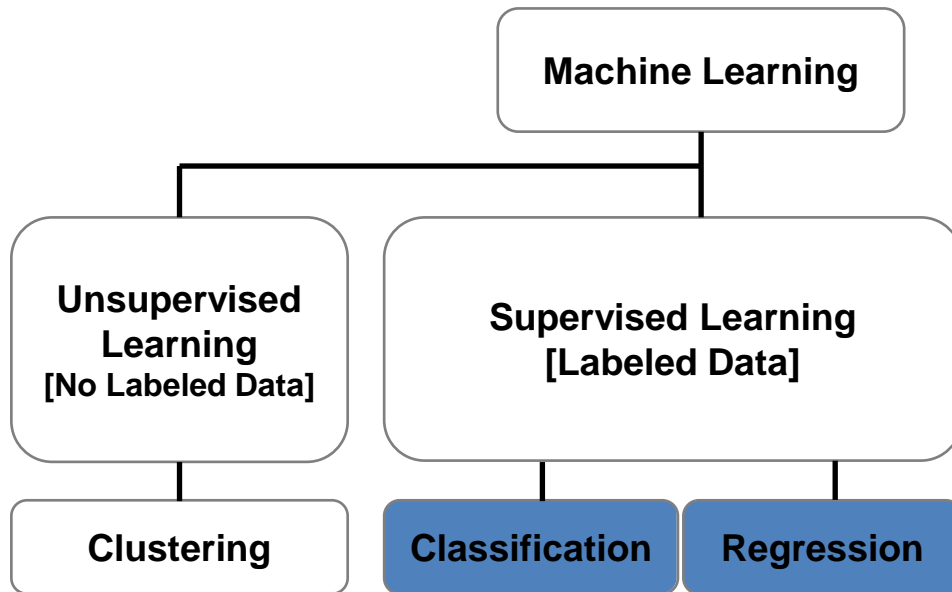
- 在bank-full.csv中依照三種類別(工作、婚姻、購買與否)區分客戶，並計算類別中的平均收入



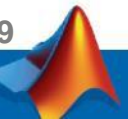
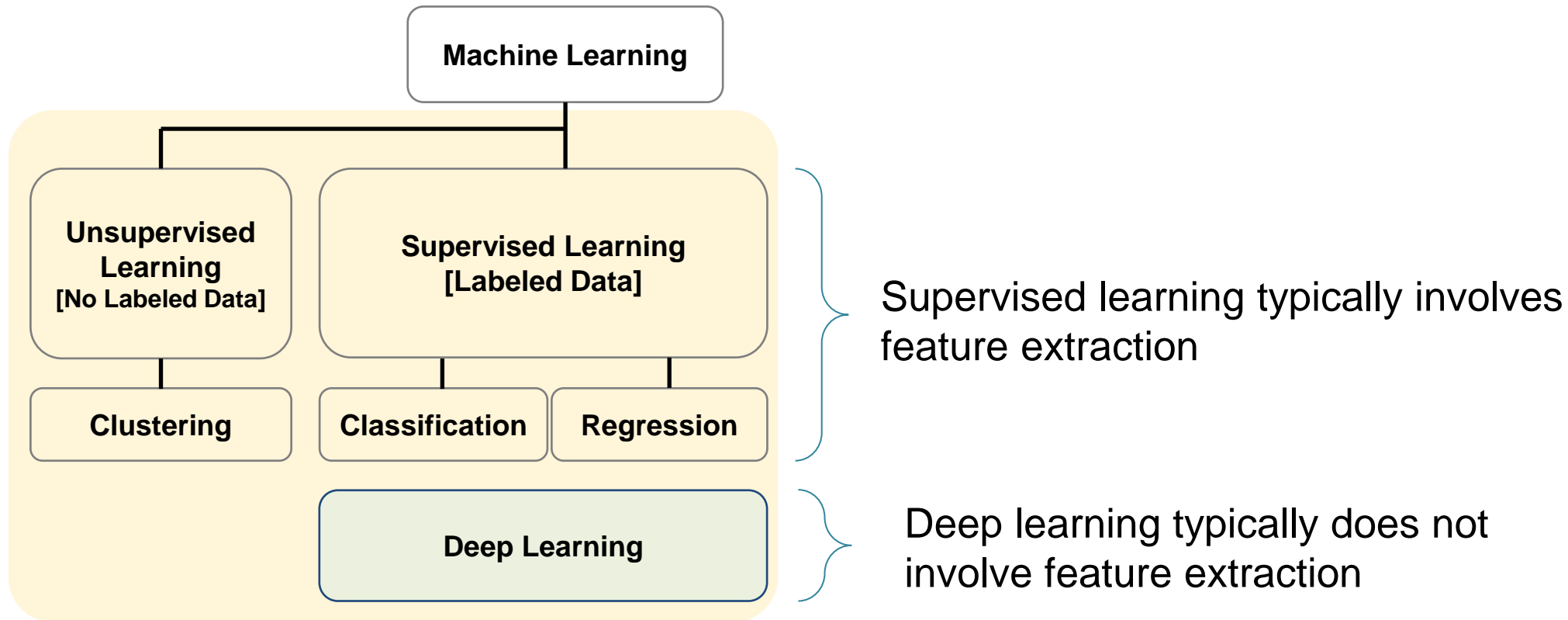
Machine Learning, Deep Learning, and Reinforcement Learning



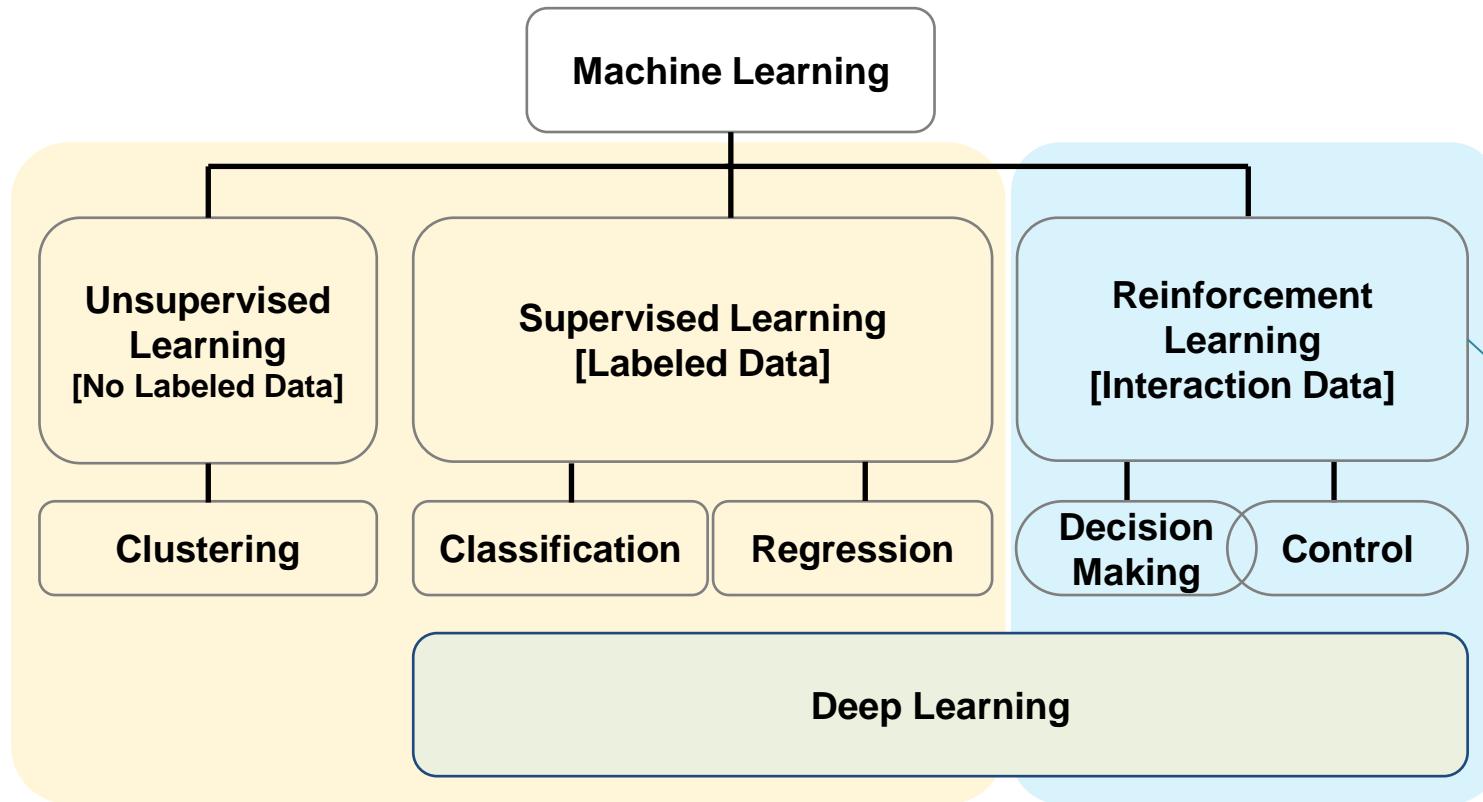
Machine Learning, Deep Learning, and Reinforcement Learning



Machine Learning, Deep Learning, and Reinforcement Learning

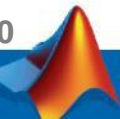


Machine Learning, Deep Learning, and Reinforcement Learning

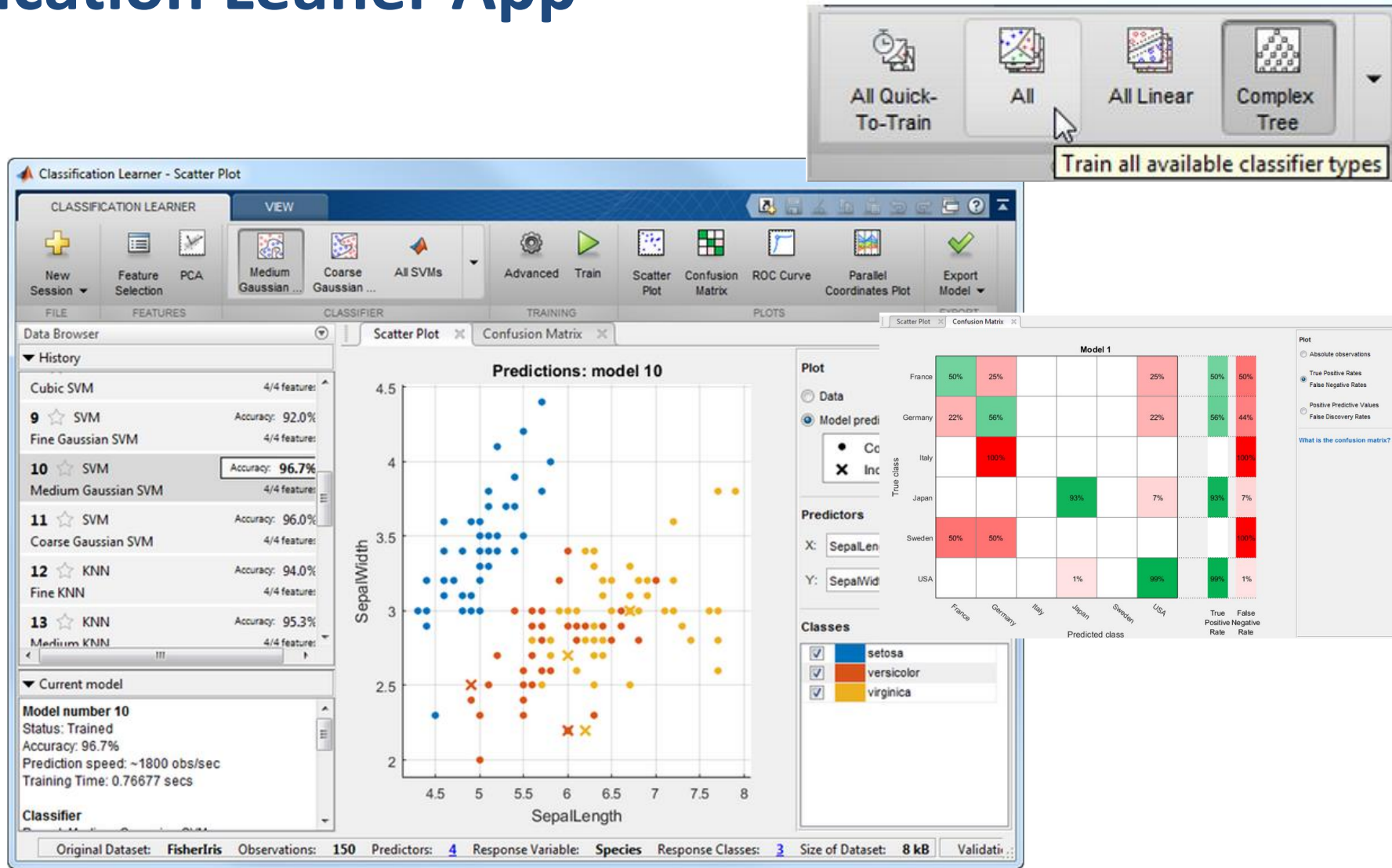


Reinforcement learning:

- Learning through trial & error [*interaction*]
- Complex problems typically need deep learning [*Deep Reinforcement Learning*]
- It's about learning a **behavior** or accomplishing a **task**



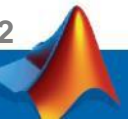
Classification Learner App



Exercise 4

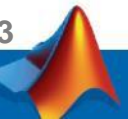
- 使用bankfull.csv中的資料，將最後一個欄位y(是否購買)當作被預測變數，並利用Classification Learner訓練出一個機器學習模型。

```
>> Data = readtable('bankfull.csv');  
>> ClassificationLearner  
>> ...  
>> y_hat = predict(myModel, Data);
```



Other Recourses

- Free online courses
 - MATLAB Onramp
 - Machine Learning Onramp
 - Deep Learning Onramp
 - Image Processing Onramp
 - ...
- FUJ Campus Wide License installation



Thanks for your attention

