



**NANYANG  
TECHNOLOGICAL  
UNIVERSITY**  
SINGAPORE

# Structured Data in Practice

**Sourav SEN GUPTA**  
Lecturer, SCSE, NTU





## Data Science

# Common Data Types

## Two Primary Data Types

### Structured Data

Highly Organized, Easy to Analyze  
Numeric/Factor, Time Series, Network

### Unstructured Data

Highly Unorganized and Contextual  
Text, Image, Voice, Videos

TV	Radio	Newspaper	Sales
230.1	37.8	69.2	22.1
44.5	39.3	45.1	10.4
17.2	45.9	69.3	9.3
151.5	41.3	58.5	18.5
180.8	10.8	58.4	12.9
8.7	48.9	75.0	7.2
57.5	32.8	23.5	11.8
8.6	2.1	1.0	4.8
199.8	2.6	21.2	10.6
66.1	5.8	24.2	8.6
214.7	24.0	4.0	17.4
23.8	35.1	65.9	9.2

Data Science

## Structured Data

### Numeric Data

Highly Organized Data

Clearly Defined Variables

Easy to Mine and Analyze

Numeric Continuous Variables

#### Example Source

- Spreadsheets (Excel, CSV)
- Standard SQL Databases
- Sensors and Devices

Advertising dataset from ISL by James et al.

3

Safety	Doors	Seats	Condition
high	4	2	unacc
med	5more	more	good
high	5more	more	vgood
high	2	2	unacc
high	2	2	unacc
low	4	more	acc
med	5more	2	unacc
high	4	4	acc
med	2	more	acc
high	4	2	unacc
low	3	4	unacc
high	3	4	unacc

## Data Science

# Structured Data

## Categorical Data

Highly Organized Data  
Clearly Defined Variables  
Easy to Mine and Analyze  
Factor/Level/Class Variables

### Example Source

- Spreadsheets (Excel, CSV)
- Standard SQL Databases
- Sensors and Devices

Car Evaluation dataset from ISL by James et al.

Price	Doors	Seats	Condition
230.1	4	2	unacc
44.5	5more	more	good
17.2	5more	more	vgood
151.5	2	2	unacc
180.8	2	2	unacc
8.7	4	more	acc
57.5	5more	2	unacc
8.6	4	4	acc
199.8	2	more	acc
66.1	4	2	unacc
214.7	3	4	unacc
23.8	3	4	unacc

## Data Science Structured Data

### Mixed Data

Highly Organized Data  
Clearly Defined Variables  
Easy to Mine and Analyze  
Numeric and Categorical

#### Example Source

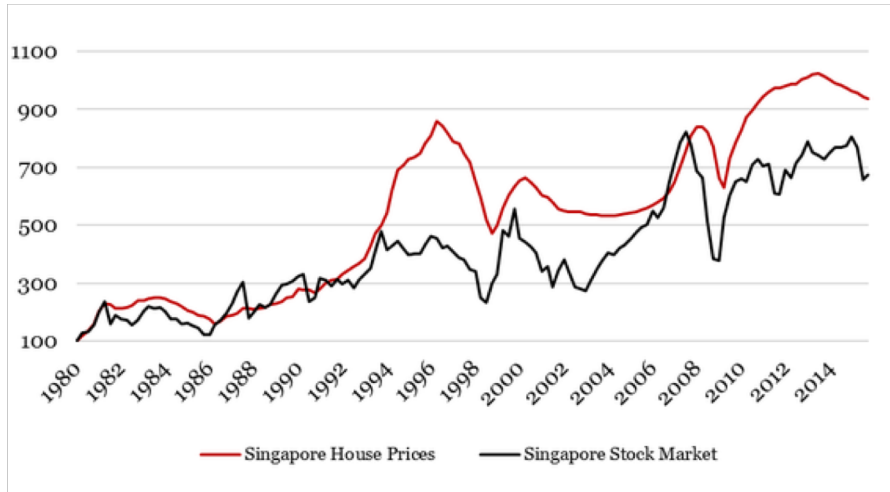
- Spreadsheets (Excel, CSV)
- Standard SQL Databases
- Sensors and Devices

Car Evaluation dataset from ISL by James et al.

# Data Science

## Structured Data

### Time Series Data



Highly Organized Data  
Clearly Defined Time Axis  
Easy to Mine and Analyze  
Numeric with Timestamps

#### Example Source

- Stock and Equity Markets
- Weather Data over Time
- Prices and Promotions

House Prices vs. Stock Data from Bloomberg

# Data Science

## Structured Data

### Network Data

Highly Organized Nodes  
Clearly Defined Links/Edges  
Easy to Mine and Analyze  
Nodes and Connections

#### Example Source

- Social Networks and Web
- Transport Networks (MRT)
- Financial Transactions



Singapore MRT Network from MRT Website



Sample  
COLLECTION



Practical  
MOTIVATION

Data  
PREPARATION



Problem  
FORMULATION

Exploratory  
ANALYSIS



Statistical  
DESCRIPTION

Analytic  
VISUALIZATION



Pattern  
RECOGNITION

Algorithmic  
OPTIMIZATION



Machine  
LEARNING

Information  
PRESENTATION



Statistical  
INFERENCE

Ethical  
CONSIDERATION



Intelligent  
DECISION

## Data Science Pipeline

### Data Acquisition and Preparation

What is the type of acquired Data?

How to prepare the acquired Data?

How to analyze the acquired Data?

## How to intelligently handle relevant Data?