



Mahidol University
IT Management

หลักสูตรอบรมเชิงปฏิบัติการ หลักการวิทยาการข้อมูลสำหรับข้อมูลสุขภาพประชาชน *แนะนำวิทยาการข้อมูล (Data Science)*

23 – 26 พฤศจิกายน 2563

กลุ่มสาขาวิชาเทคโนโลยีการจัดการระบบสารสนเทศ

คณะวิศวกรรมศาสตร์ มหาวิทยาลัยมหิดล



+ a b l e a u



กำหนดการ



Mahidol University
IT Management

- ข้อมูลและรูปแบบของข้อมูล
- รูปแบบวิเคราะห์ข้อมูล
- หลักการคุณภาพข้อมูล
- การนำเสนอข้อมูล



<http://zotar.com>

ผศ.ดร.วิชิตรัตน์ ธรรมบุษดี



- อาจารย์ประจำหลักสูตร IT Management คณะวิศวกรรมศาสตร์ มหาวิทยาลัยมหิดล
- หัวหน้าโครงการฝึกอบรม Datalent Team
- วิทยากรและที่ปรึกษาด้าน Data Governance ทั้งหน่วยงานภาครัฐและเอกชน
- วิทยากรและที่ปรึกษาด้าน Data Science and Data Analytics

- พ.ศ. 2563 – ปัจจุบัน ที่ปรึกษาธรรมาภิบาลข้อมูลและการคุ้มครองข้อมูลส่วนบุคคล
คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล
- พ.ศ. 2562 – ปัจจุบัน คณะกรรมการจัดทำนโยบายคุ้มครองข้อมูลส่วนบุคคล
มหาวิทยาลัยมหิดล
- พ.ศ. 2562 – ปัจจุบัน คณะกรรมการจัดทำร่างมาตรฐาน ข้อกำหนด และหลักเกณฑ์ ภายใต้อพพร.ดิจิทัล
สำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน)
- พ.ศ. 2561 คณะกรรมการศึกษากระบวนการธรรมาภิบาลข้อมูลและการเปิดเผยข้อมูลดิจิทัล
สำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน)

itm.egmu
itm.eg.mahidol.ac.th



ข้อมูลและรูปแบบของข้อมูล

ข้อมูล (Data)

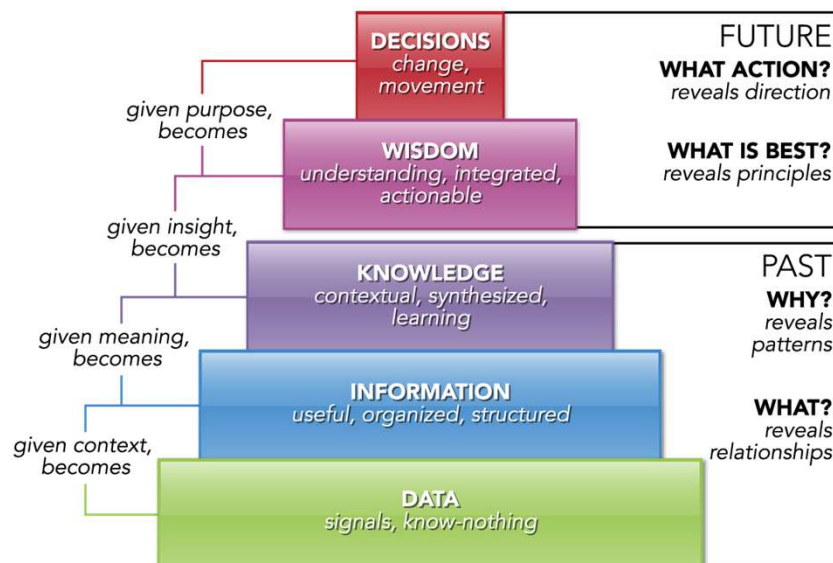


- Data is an ASSET:

- **ข้อมูล** ถูกจัดอยู่ในประเภทของทรัพย์สินที่มีค่ามากขององค์กร เนื่องจากข้อมูลที่อยู่ในองค์กรนั้น ๆ สามารถบ่งชี้ จุดเด่น จุดด้อย ความชำนาญ และความสามารถในการแข่งขัน นอกเหนือจากทรัพยากรทางการเงิน และทรัพยากรมนุษย์ ที่ถูกจัดเป็นทรัพย์สินภายในองค์กร ตั้งแต่ในอดีต จนถึงปัจจุบัน
- **ข้อมูลภาครัฐ** ถือเป็นทรัพย์สินมีค่าที่จะสามารถนำไปใช้เป็นพื้นฐานในการพัฒนาประเทศในทิศทางต่าง ๆ ได้

5

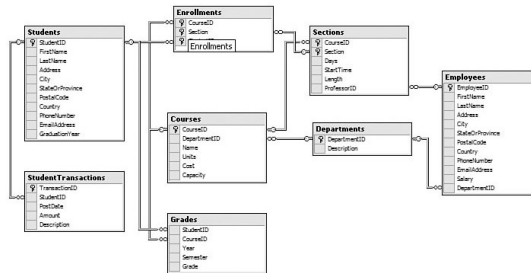
DIKW Pyramid



6

ประเภทของข้อมูล

- ข้อมูลที่มีโครงสร้าง (Structured Data)
 - เป็นข้อมูลที่มีการนิยามโครงสร้างของข้อมูลไว้ ซึ่งเป็นโครงสร้างที่ทำให้ง่ายต่อการค้นหา เช่น
 - ฐานข้อมูล SQL
 - Comma-Separated Values – CSV
 - ข้อมูลจากเซ็นเซอร์
 - ข้อมูลจาก Web Log
 - ข้อมูลจากจุดขายสินค้า
 - ข้อมูลทางการเงิน

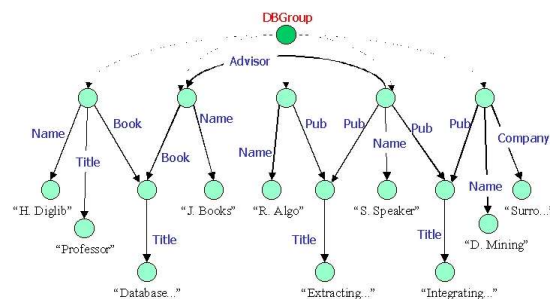


0.103	0.176	0.387	0.300	0.379
0.333	0.384	0.564	0.587	0.857
0.421	0.309	0.654	0.729	0.228
0.266	0.750	1.056	0.936	0.911
0.225	0.326	0.643	0.337	0.721
0.187	0.586	0.529	0.340	0.829
0.153	0.485	0.560	0.428	0.628

7

ประเภทของข้อมูล

- ข้อมูลกึ่งมีโครงสร้าง (Semi-Structured Data)
 - เป็นข้อมูลที่มีการนิยามโครงสร้างของข้อมูลไว้ แต่มีโครงสร้างเป็นแบบลำดับชั้น (Hierarchy) เช่น
 - Extensible Markup Language – XML
 - JavaScript Object Notation – JSON



8

ประเภทของข้อมูล

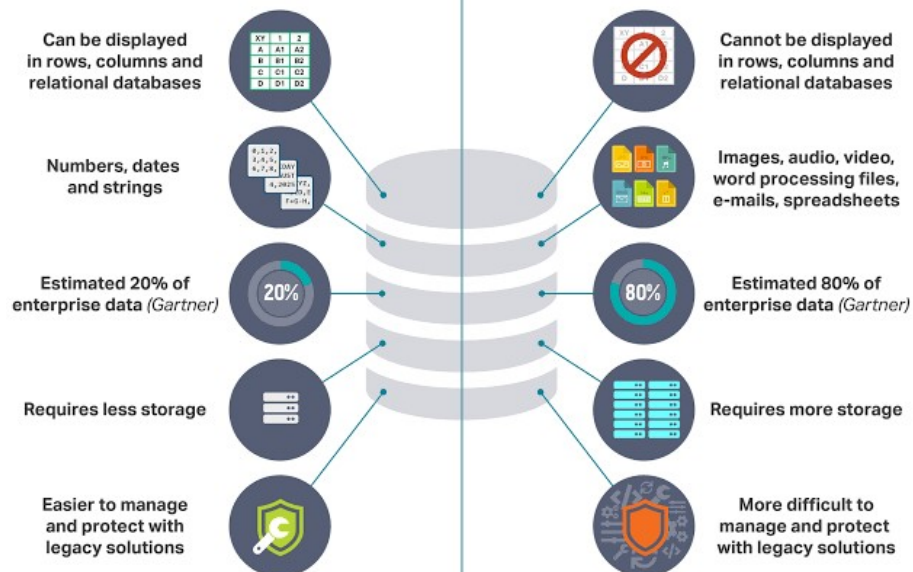
• ข้อมูลแบบไม่มีโครงสร้าง (Unstructured Data)

- เป็นข้อมูลที่ไม่ได้มีการนิยามโครงสร้างของข้อมูลไว้ มักจะอยู่ในรูปแบบ เช่น
- กระดาษ
- ข้อความ
- รูปภาพ
- เสียง
- วิดีทัศน์



9

Structured Data VS Unstructured Data



10

ชุดข้อมูล

ชุดข้อมูล คือ ข้อมูลที่มีการรวบรวมไว้ โดยปกติอยู่ในรูปแบบของตารางข้อมูล ซึ่งการรวบรวมข้อมูลนี้มาจากหลายแหล่ง และนำข้อมูลมาจัดเป็นชุดให้ถูกต้องตามลักษณะโครงสร้างข้อมูลที่กำหนดไว้

5 ฟิลด์

3 แถว

ชื่อ	นามสกุล	เพศ	อายุ	ระดับการศึกษา
วิชัย	ใจดี	ชาย	๒๖	ปริญญาตรี
พิเชษฐ์	วิเศษศิลป์	ชาย	๒๘	ปริญญาโท
วิมลมาส	วงศ์สกุล	หญิง	๒๕	ปริญญาตรี

11

มูลค่าของข้อมูล

Cost

ต้นทุนในการจัดหาและจัดเก็บข้อมูล
ต้นทุนในการหาข้อมูลมาแทนที่ข้อมูลที่สูญหายไป
ผลกระทบต่อการดำเนินงานเมื่อข้อมูลสูญหาย
ต้นทุนในการลดความเสี่ยงที่เกี่ยวข้องกับข้อมูล
ต้นทุนในการปรับปรุงข้อมูลให้มีคุณภาพ

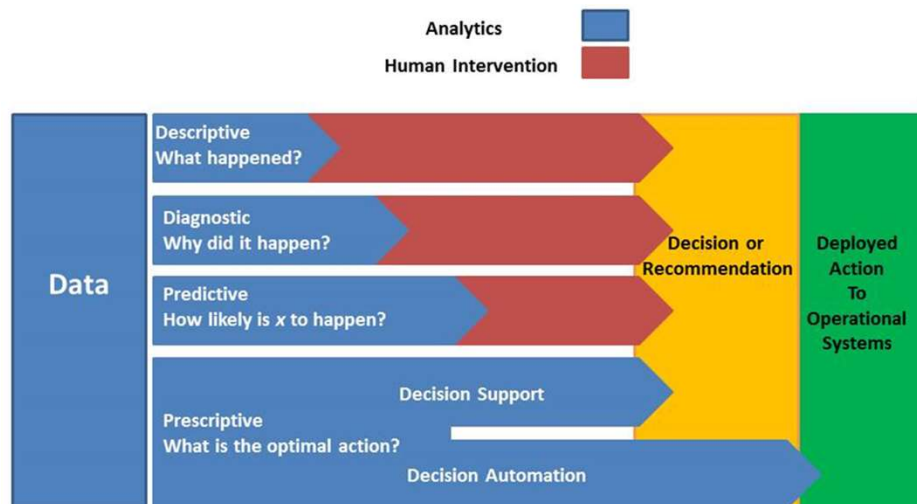
Benefit

ประโยชน์จากข้อมูลที่มีคุณภาพ
รายได้จากการขายข้อมูล
รายได้จากการสร้างนวัตกรรมจากข้อมูล
การเพิ่มประสิทธิภาพการทำงานในองค์กร
การเพิ่มขีดความสามารถในการแข่งขัน

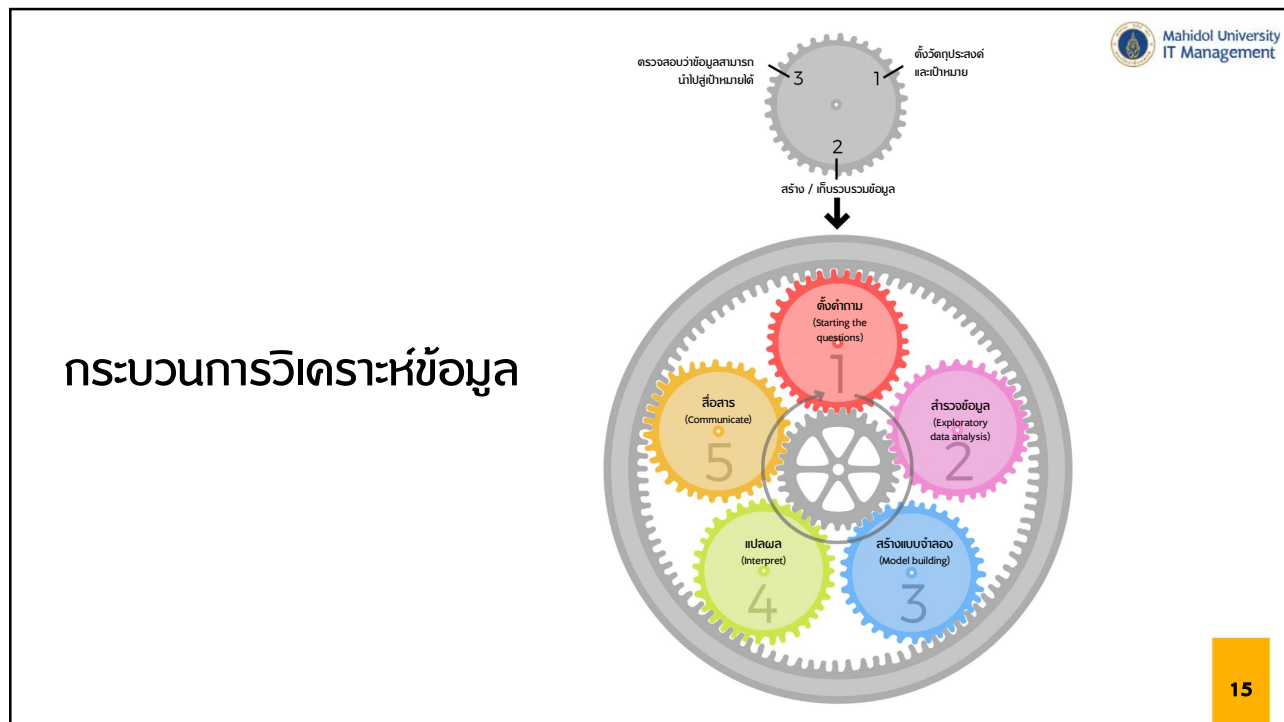
12

รูปแบบของการวิเคราะห์ข้อมูล

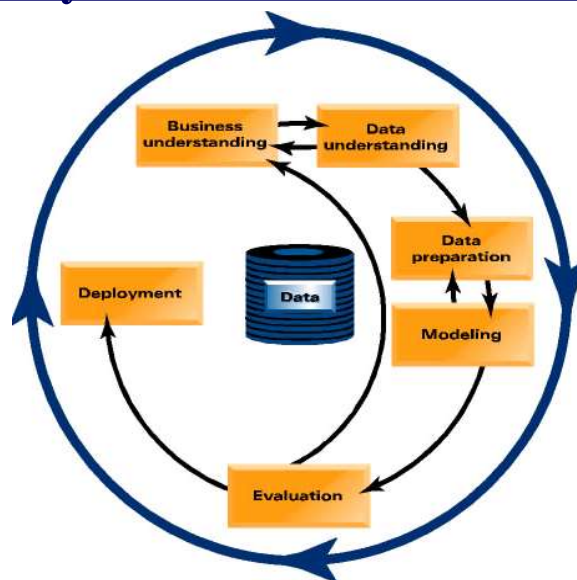
Type of Data Analytics



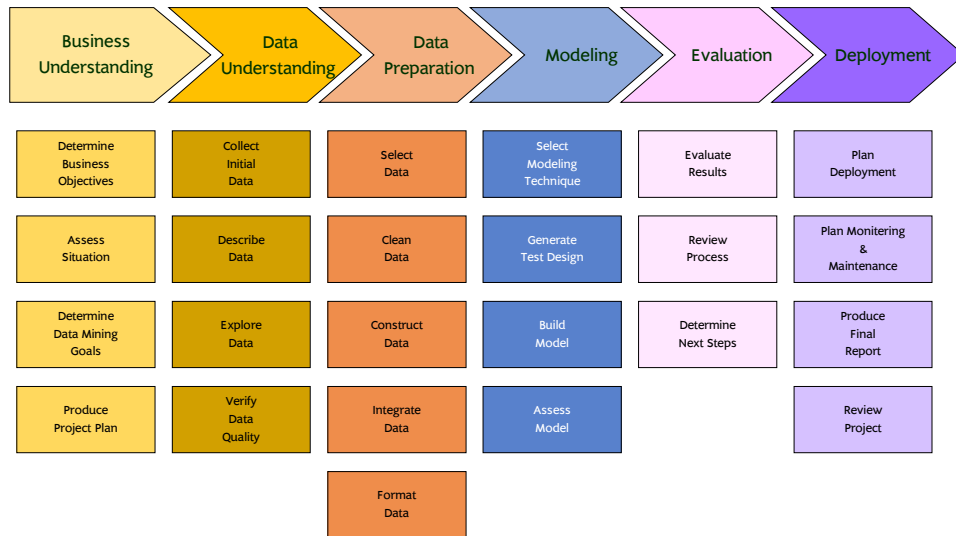
After Gartner (September 2013)



Cross-Industry Standard Process for Data Mining

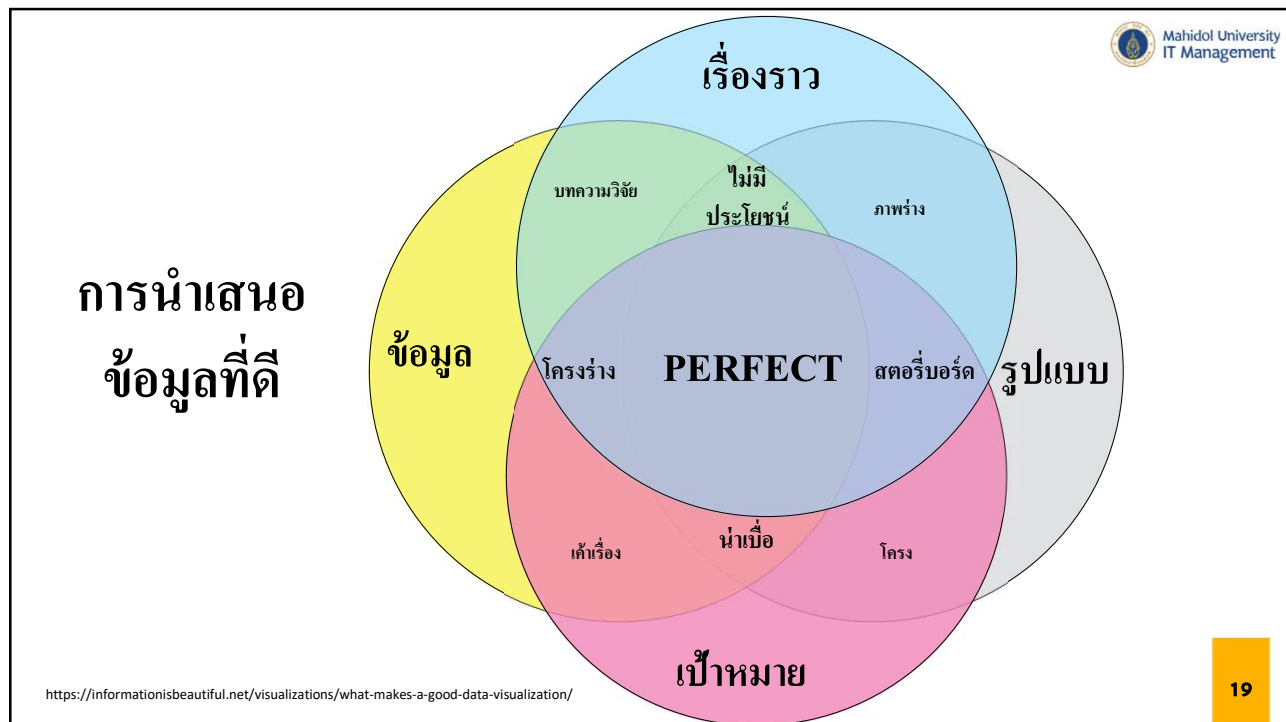


CRISP-DM



17

การนำเสนอข้อมูลด้วยภาพ (Data visualization)

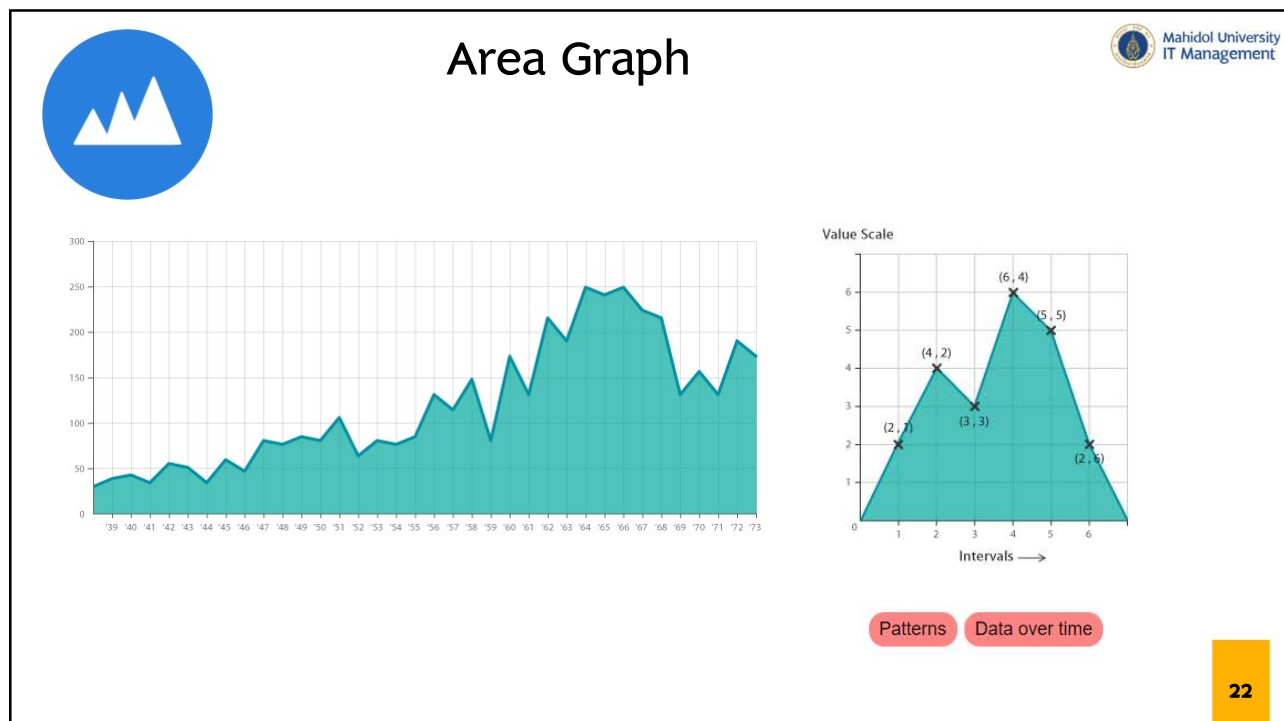
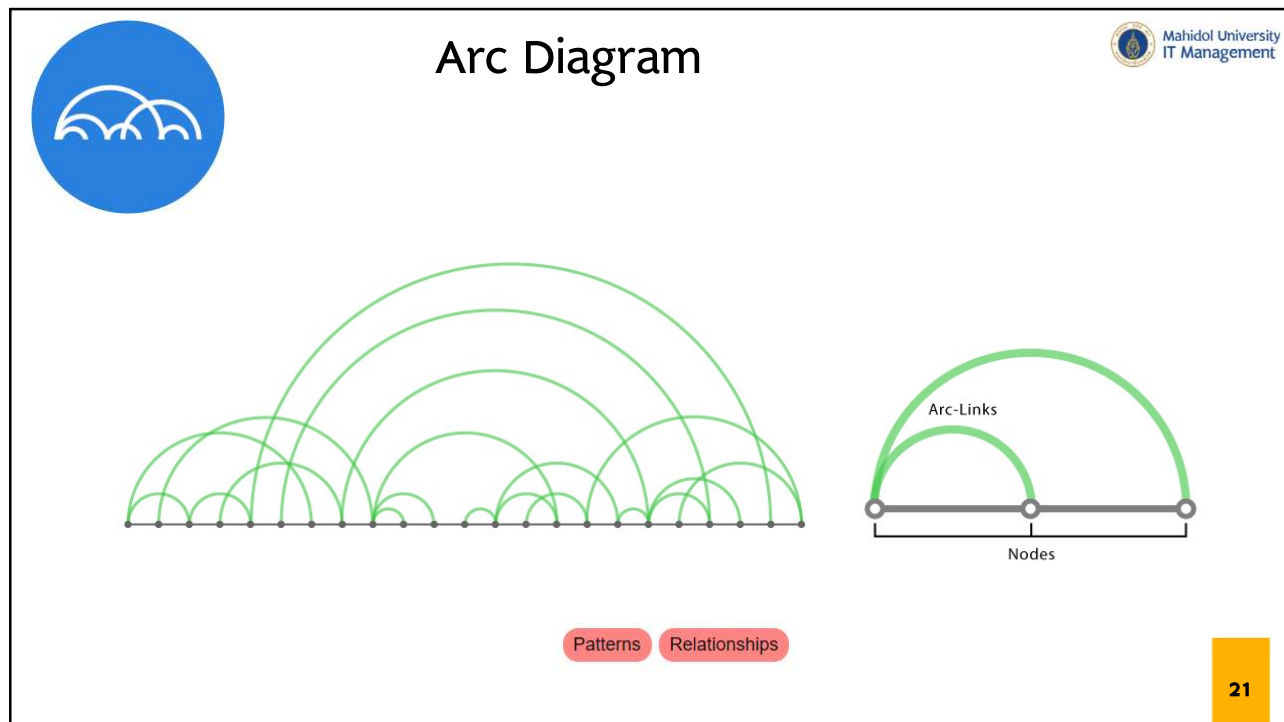


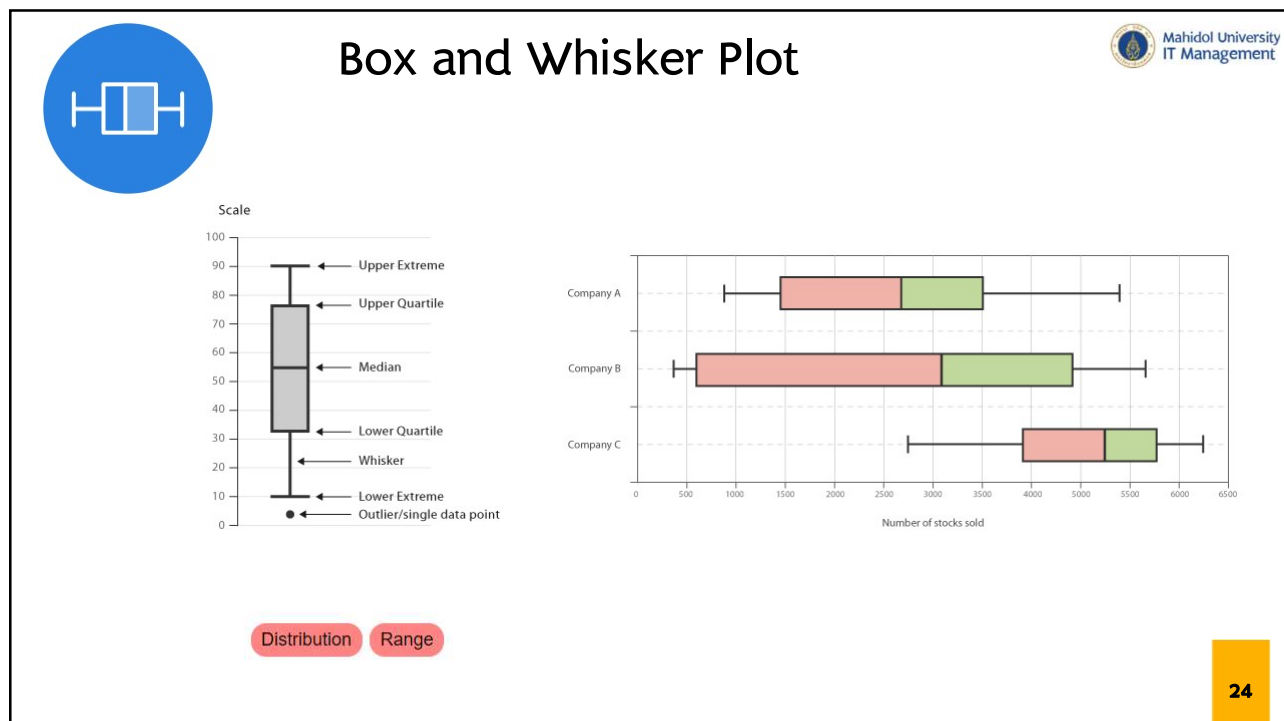
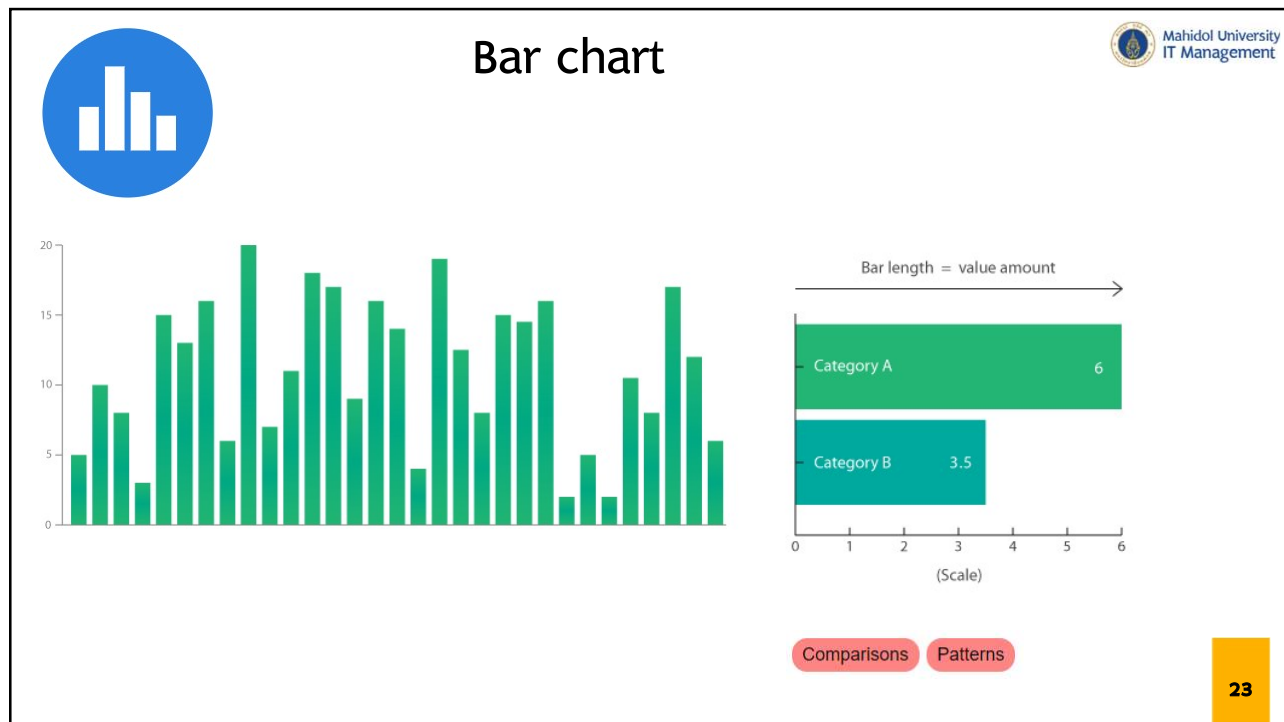
Mahidol University
IT Management

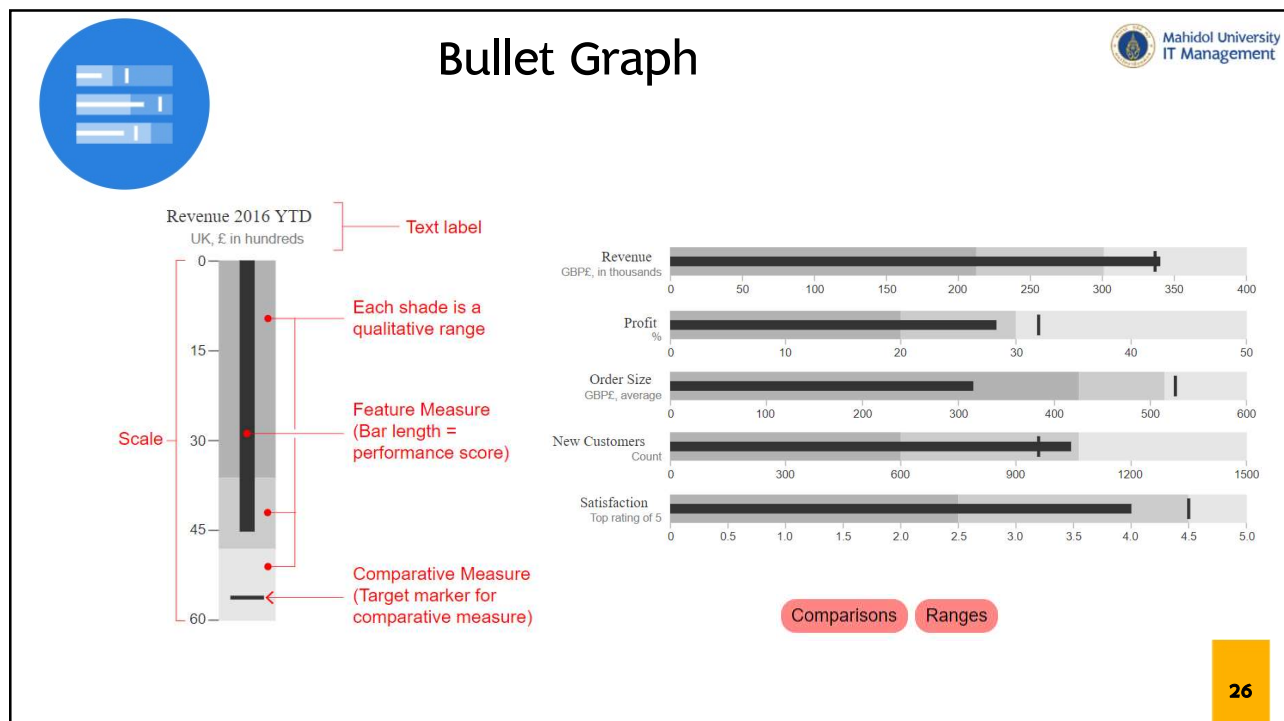
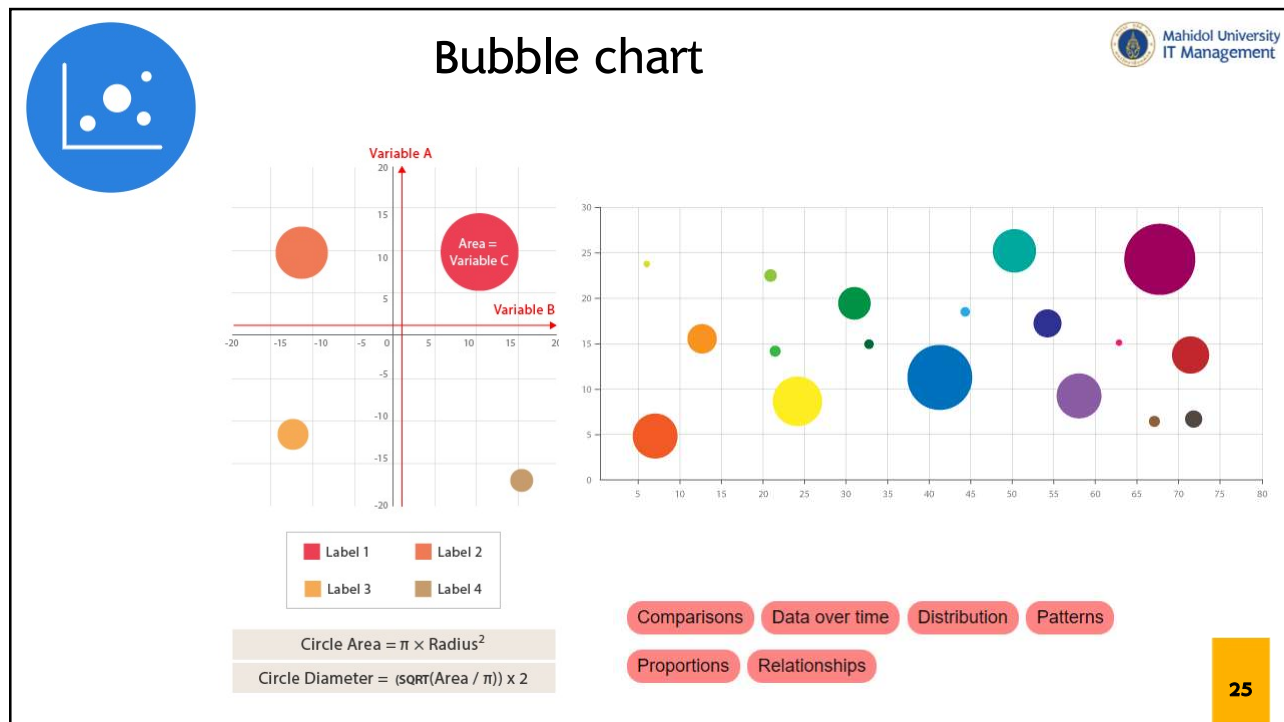
Data Visualization Catalogue

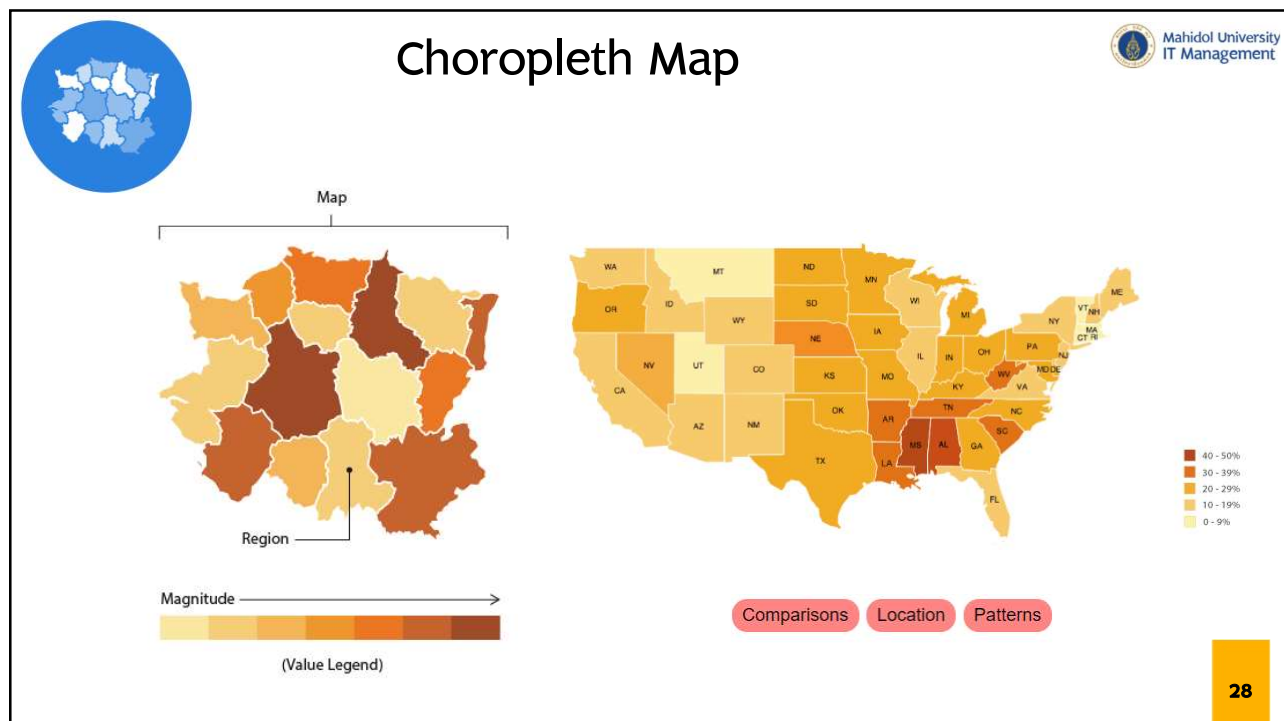
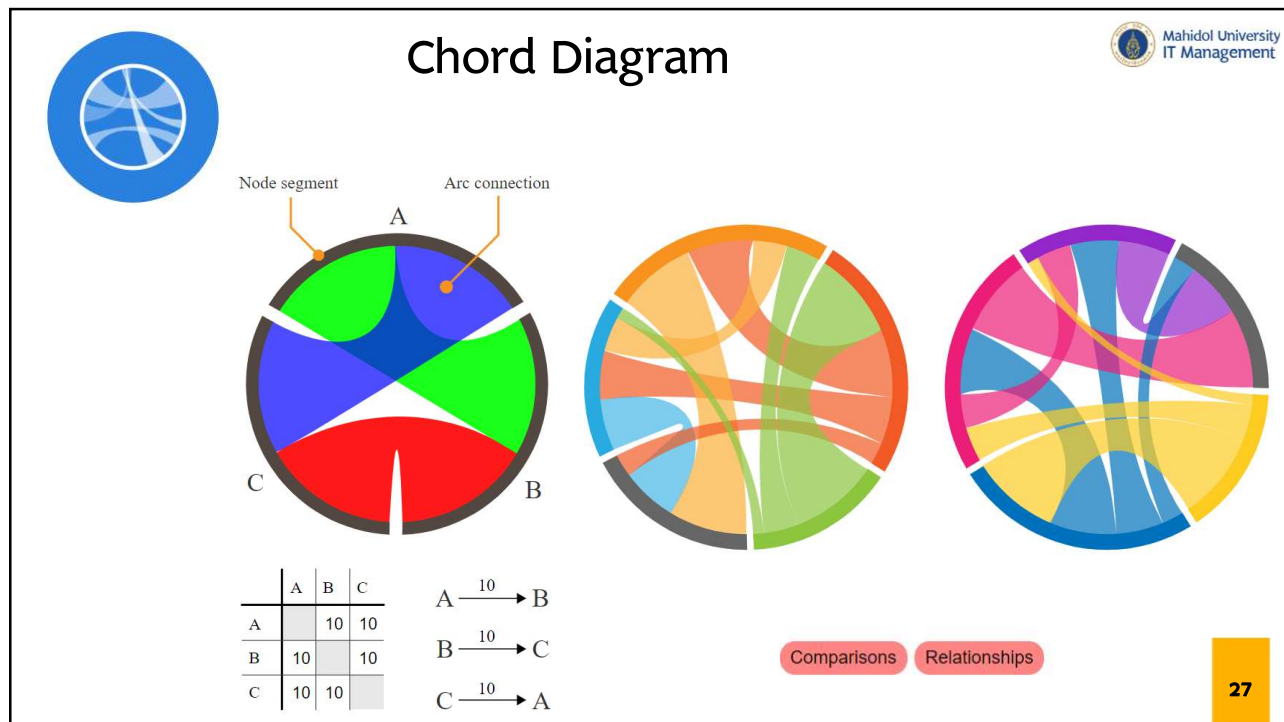
<https://datavizcatalogue.com/>

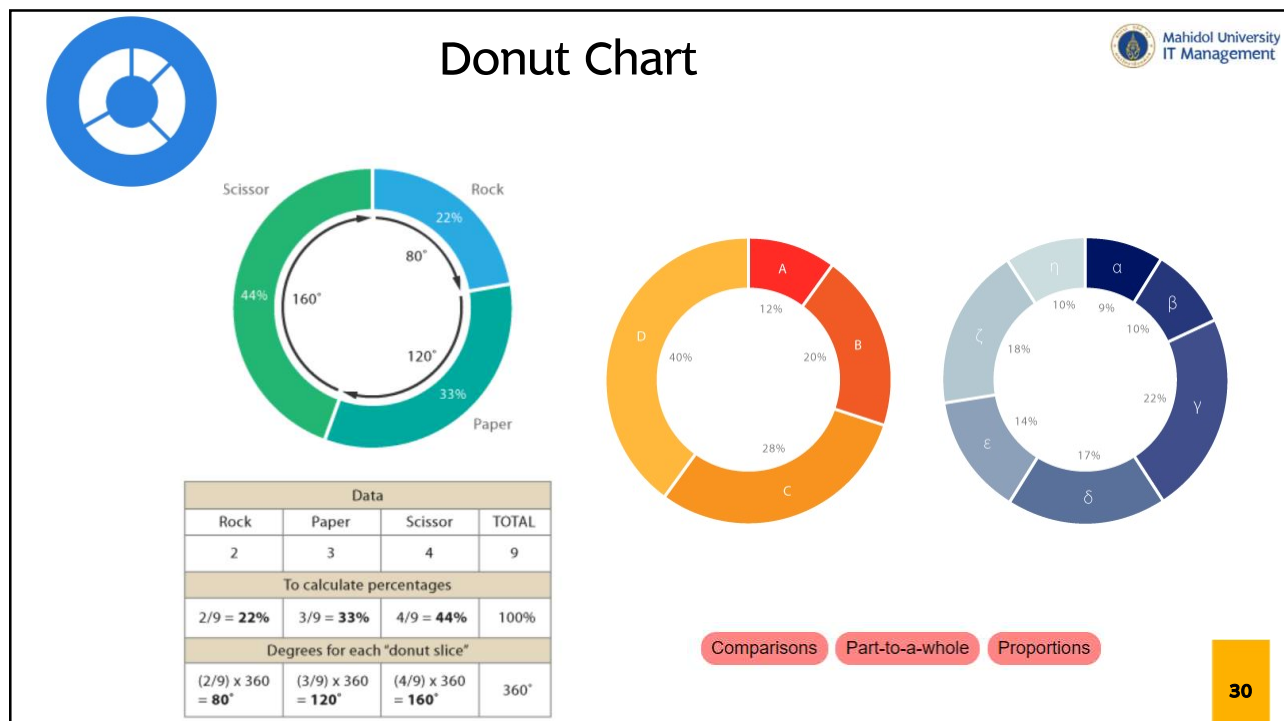
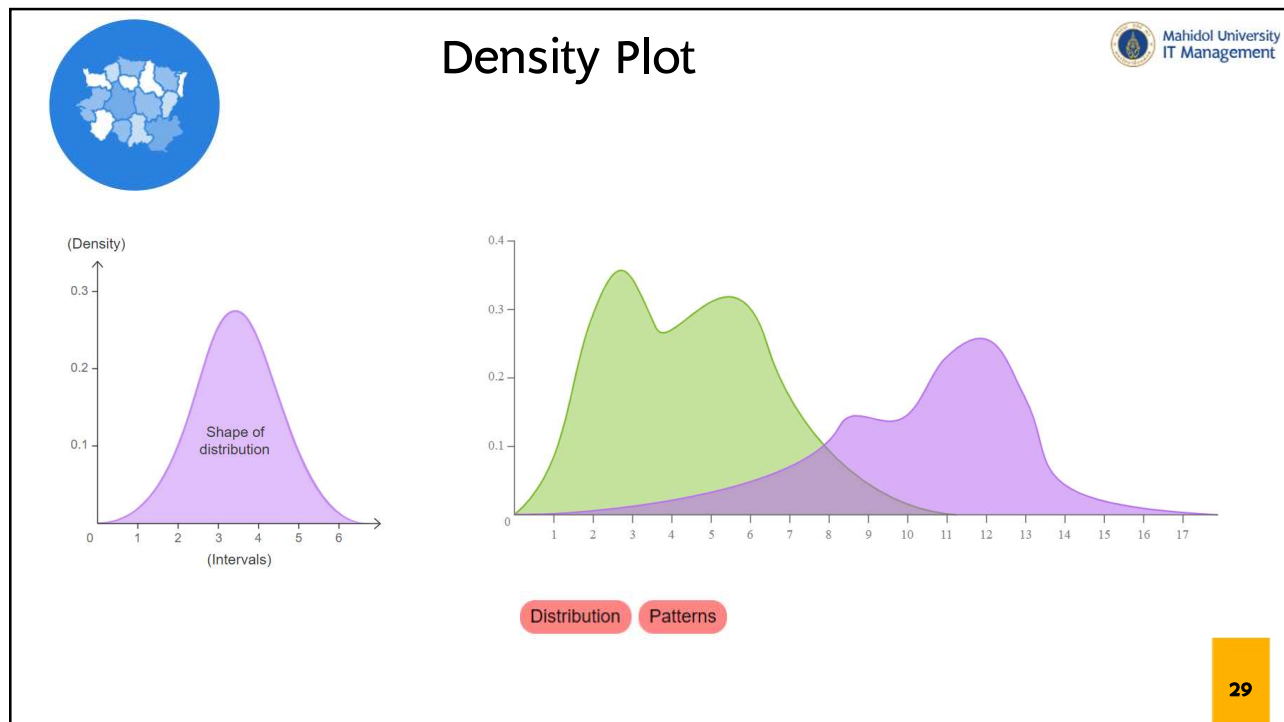
20

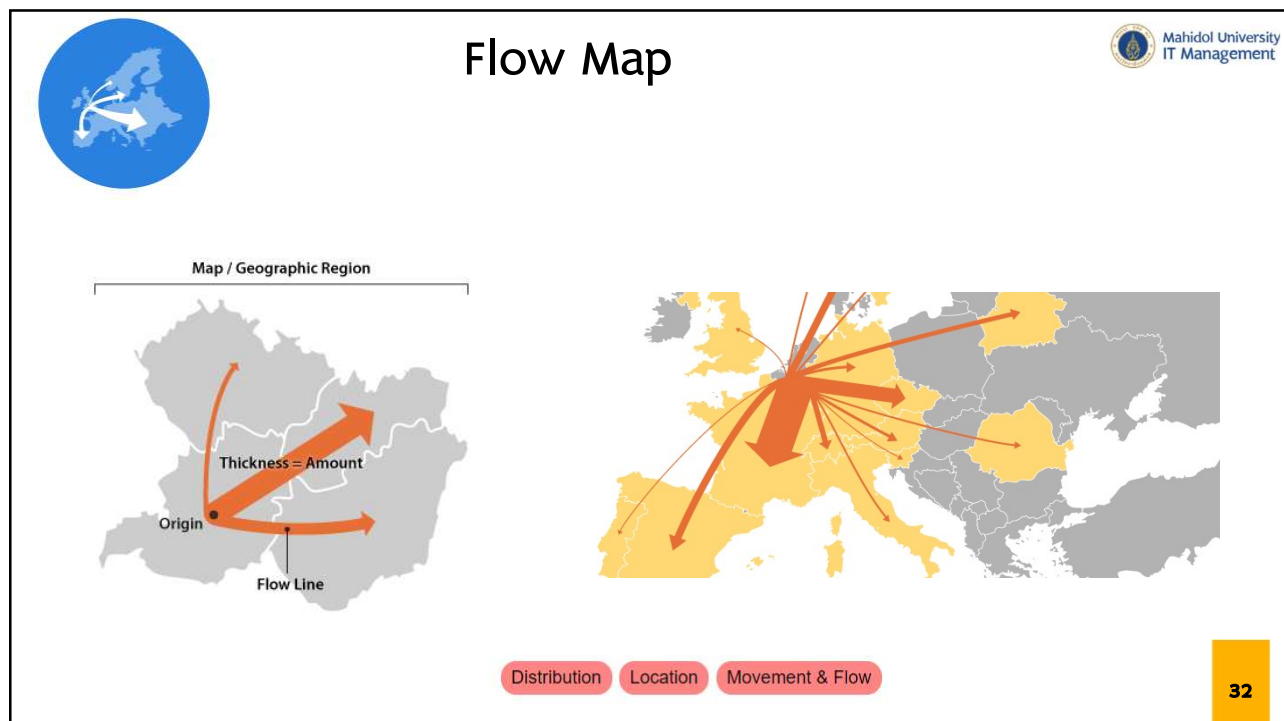
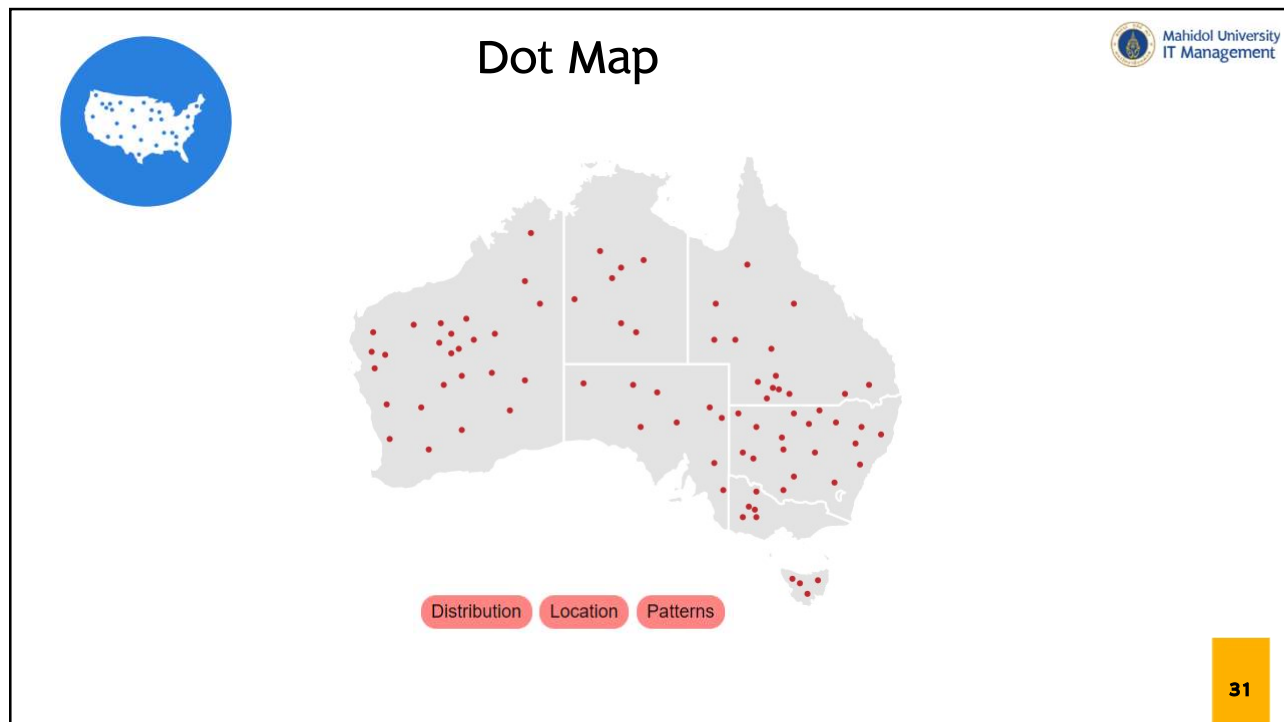








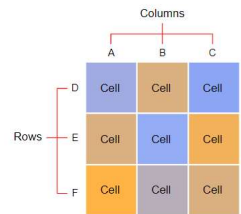






Heatmap (Matrix)

Heatmap using numerical data:



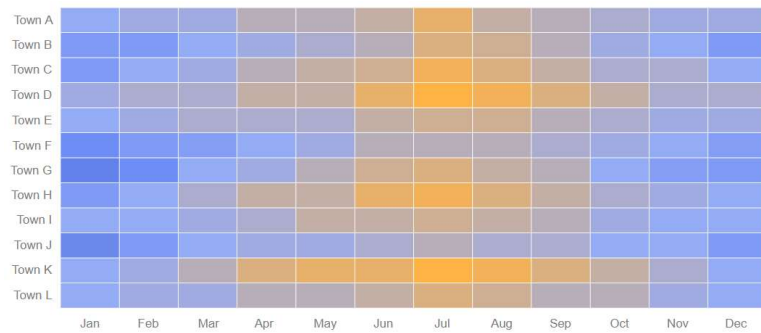
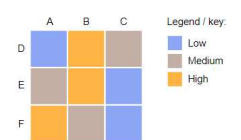
Value scale for determining cell colouring:

0 50 100

Alternative value scale broken into ranges:

0 - 20 21 - 40 41 - 60 61 - 80 81 - 100

Heatmap using categorical data:

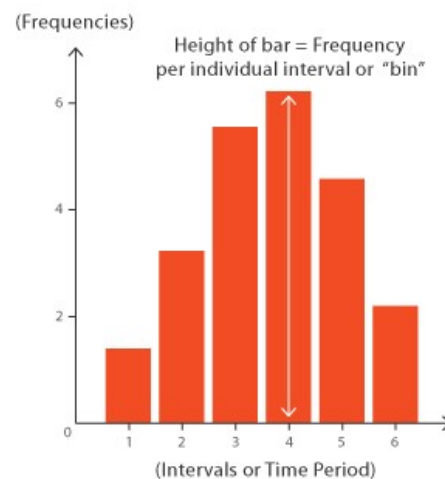


Comparisons Data over time Patterns Relationships

33



Histogram



Comparisons Data over Time Distribution Patterns

Range

34

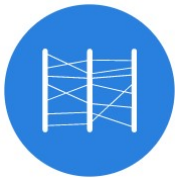


Line Graph

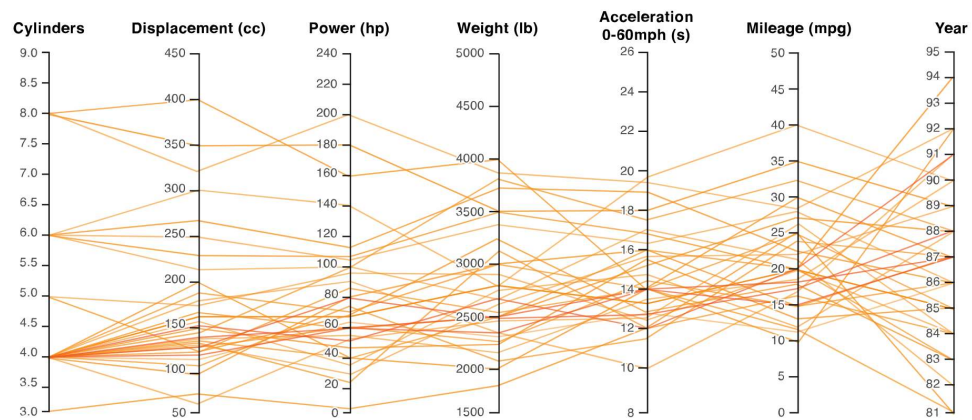


Patterns Data over time

35



Parallel Coordinates Plot

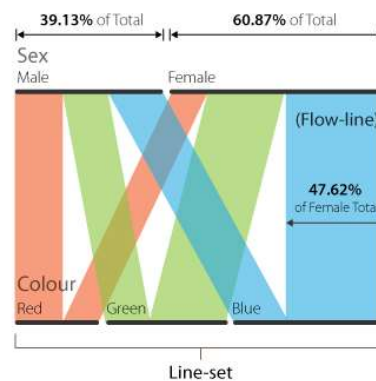


Comparisons Relationships Patterns

36



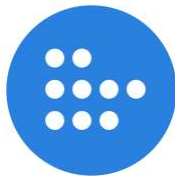
Parallel Sets



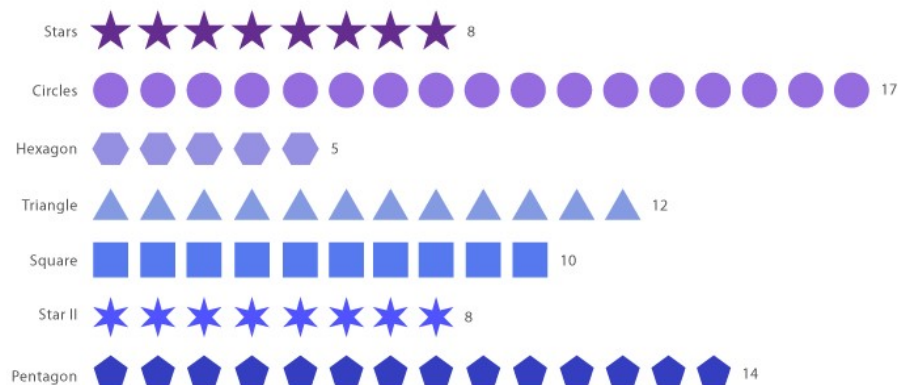
Comparisons Distribution Flow Processes & Methods
Proportions

Sex	Colour	Count	% of Sex TOTAL
Male	Red	35	32.41
	Green	33	30.56
	Blue	40	37.04
Female	Red	28	16.67
	Green	60	35.71
	Blue	80	47.62

37

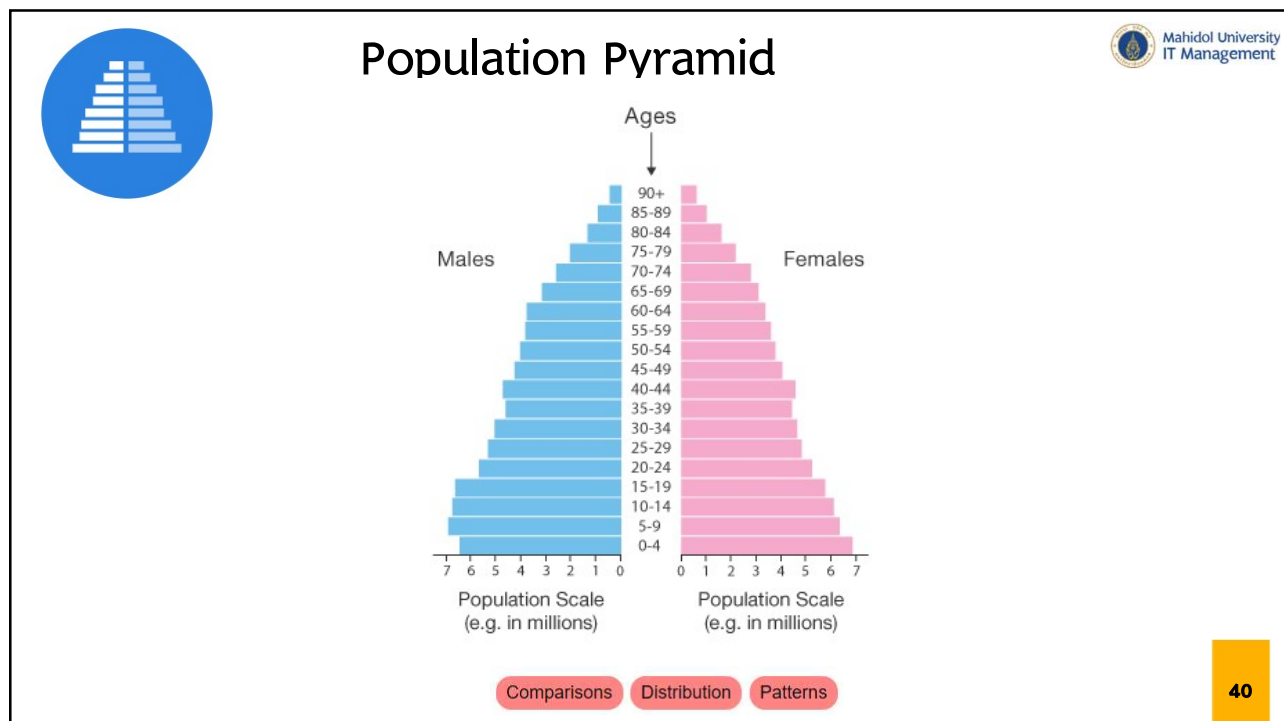
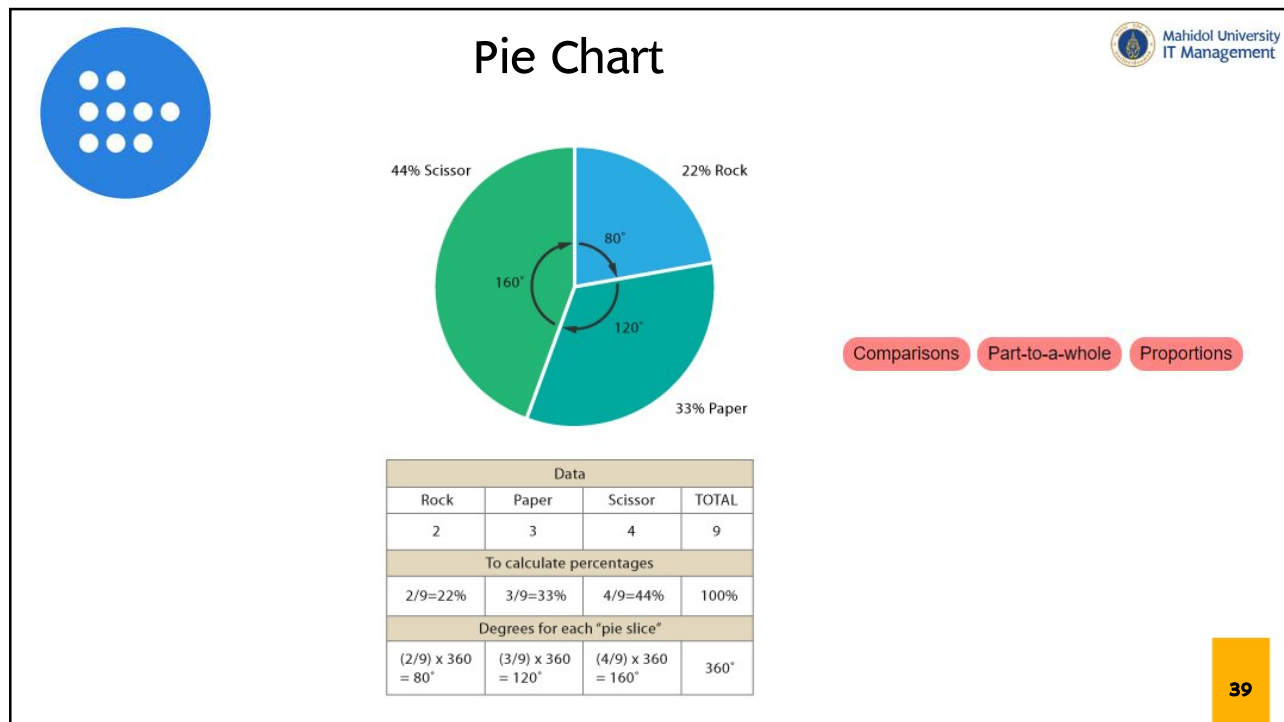


Pictogram Chart



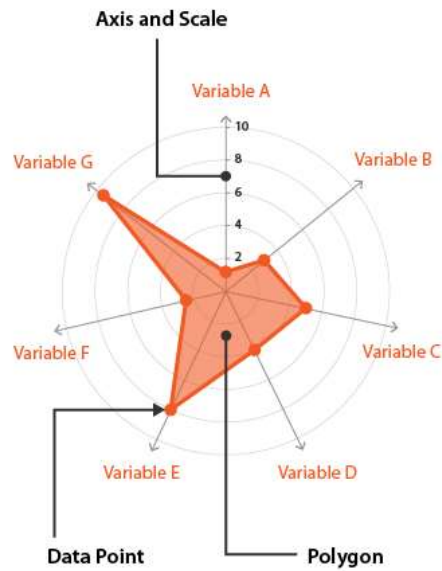
Comparisons Distribution

38





Radar Chart

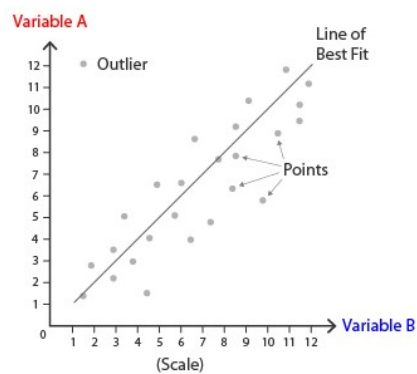


Comparisons Relationships Patterns

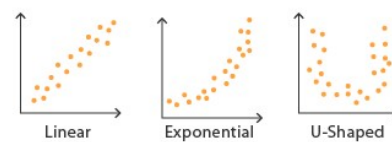
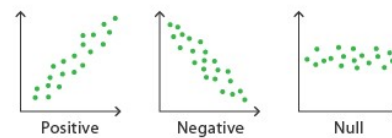
41



Scatterplot



Patterns Relationships



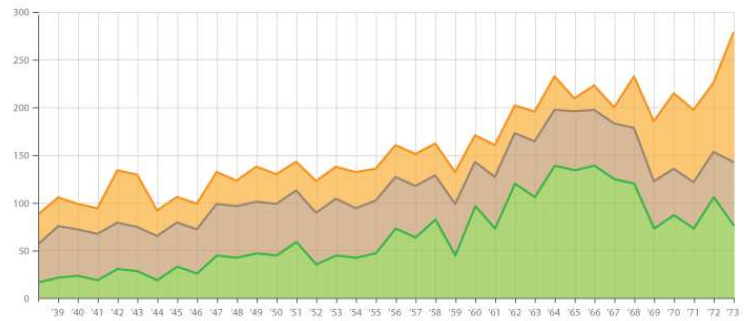
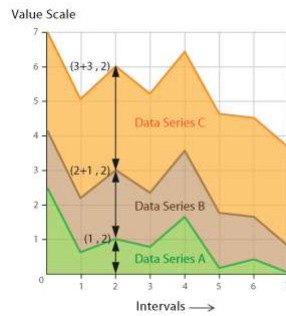
Correlation Strength:



42



Stacked Area Graph

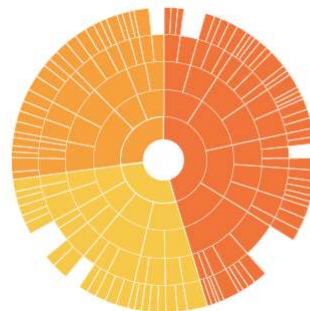
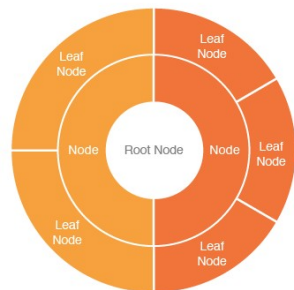
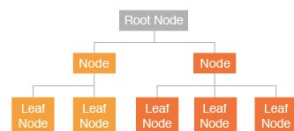


Comparisons Data over time Patterns

43



Sunburst Diagram



Hierarchy Part-to-a-whole

44



Word Cloud

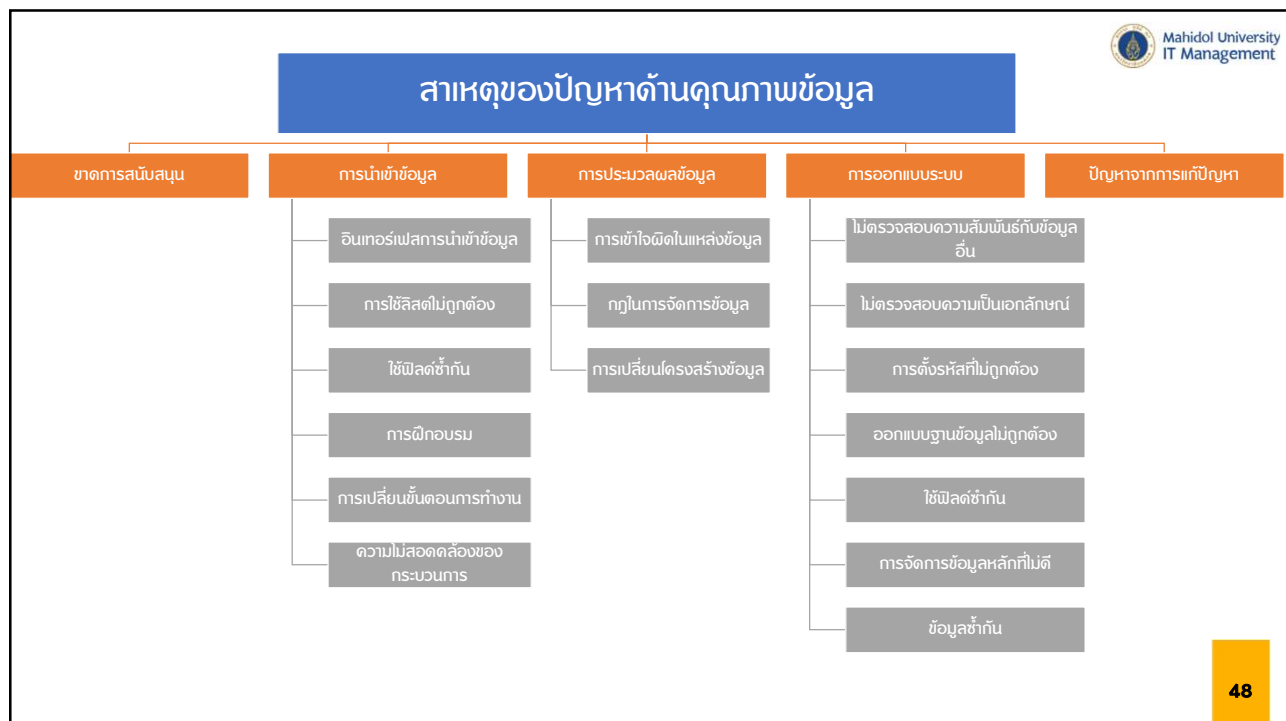


Analysing Text Distribution / Frequency Proportions

23

หลักการคุณภาพข้อมูล

47

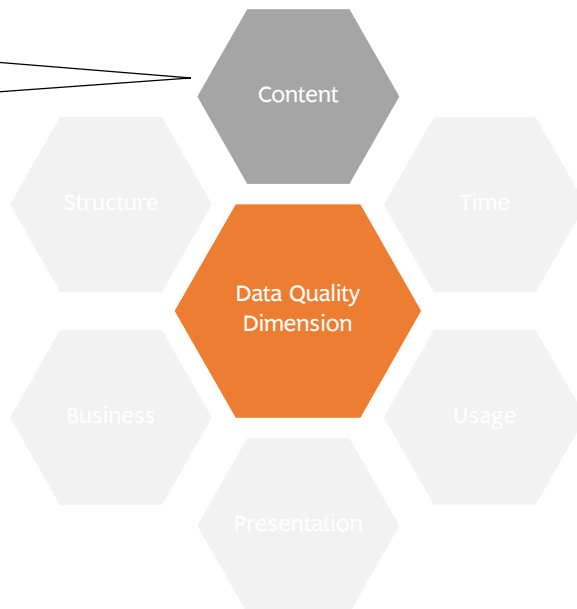


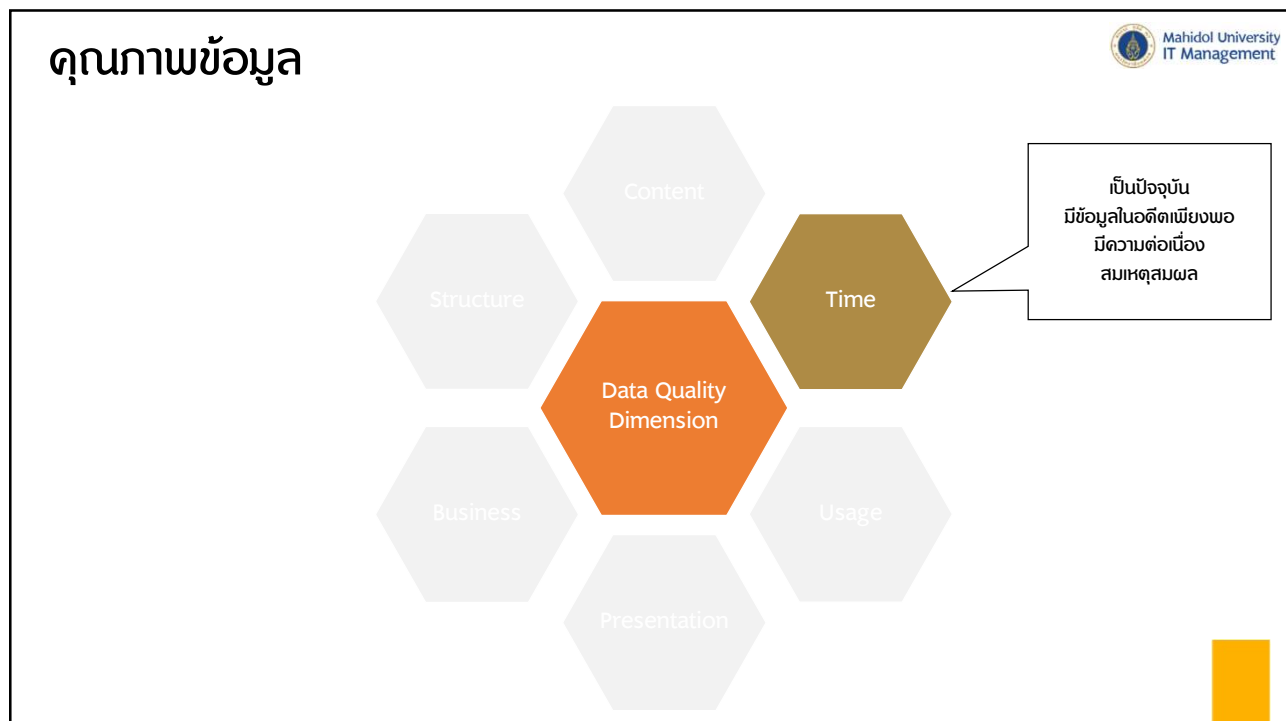
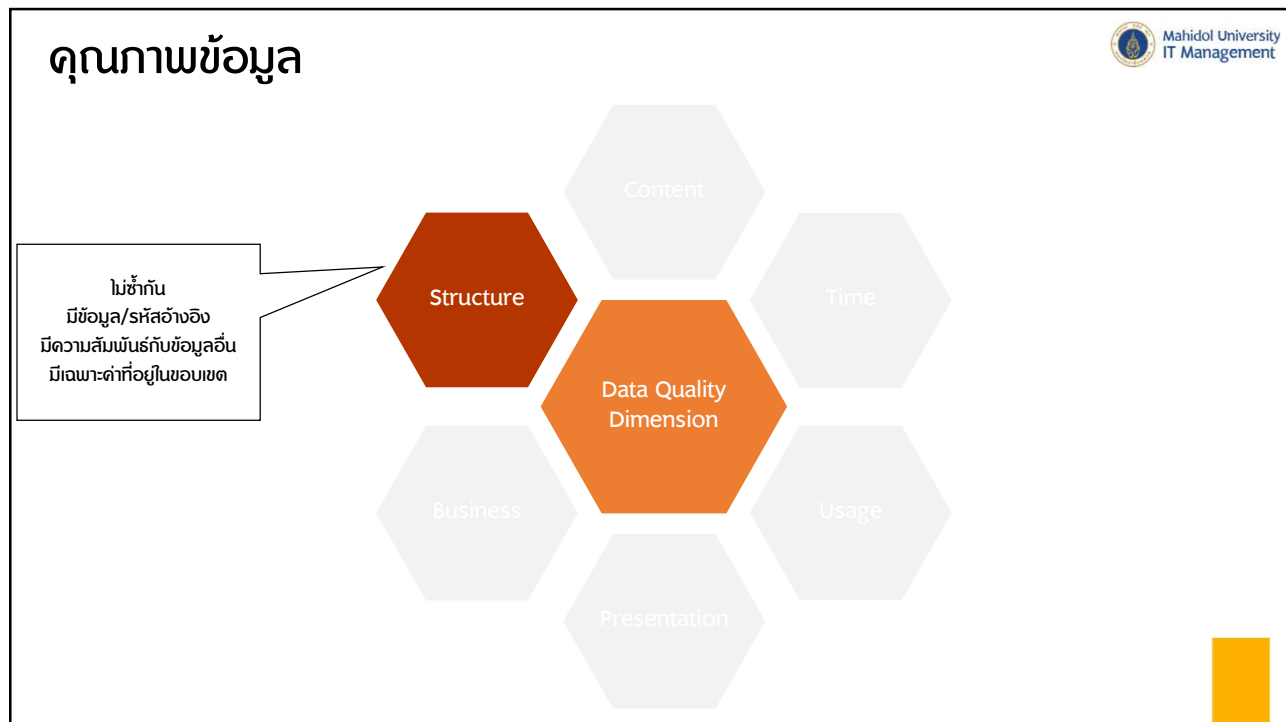
คุณภาพข้อมูล

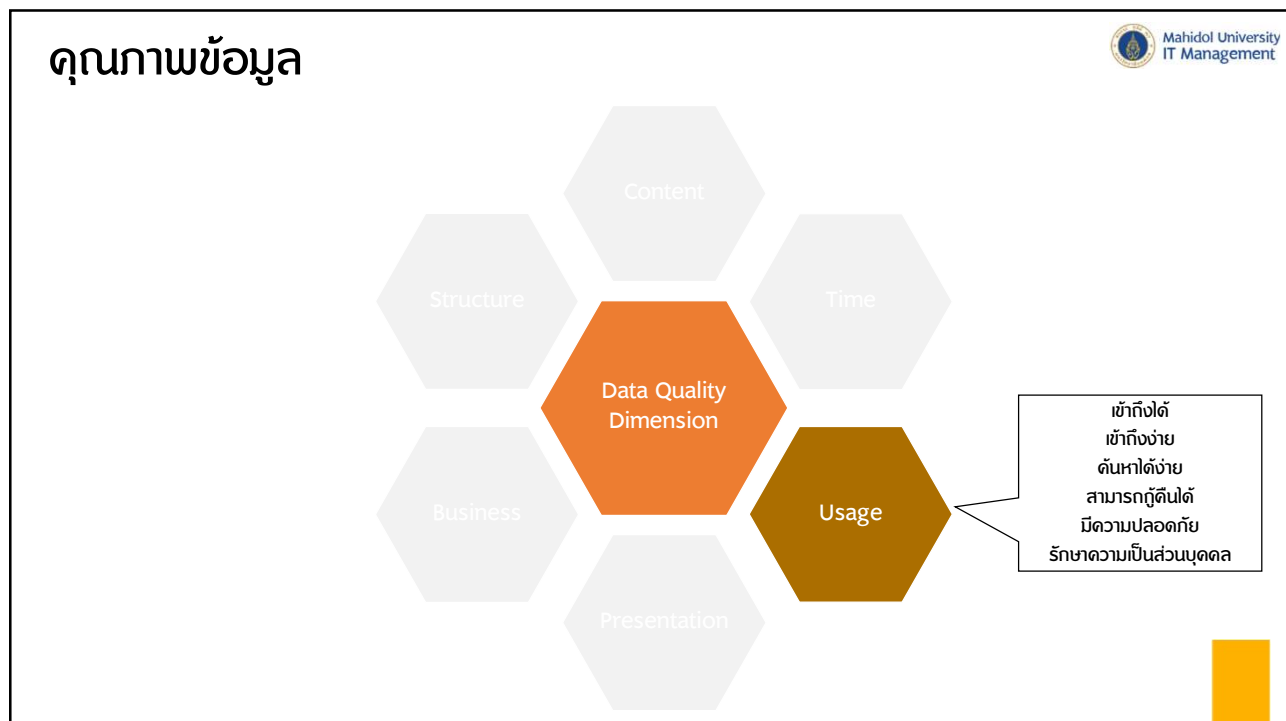
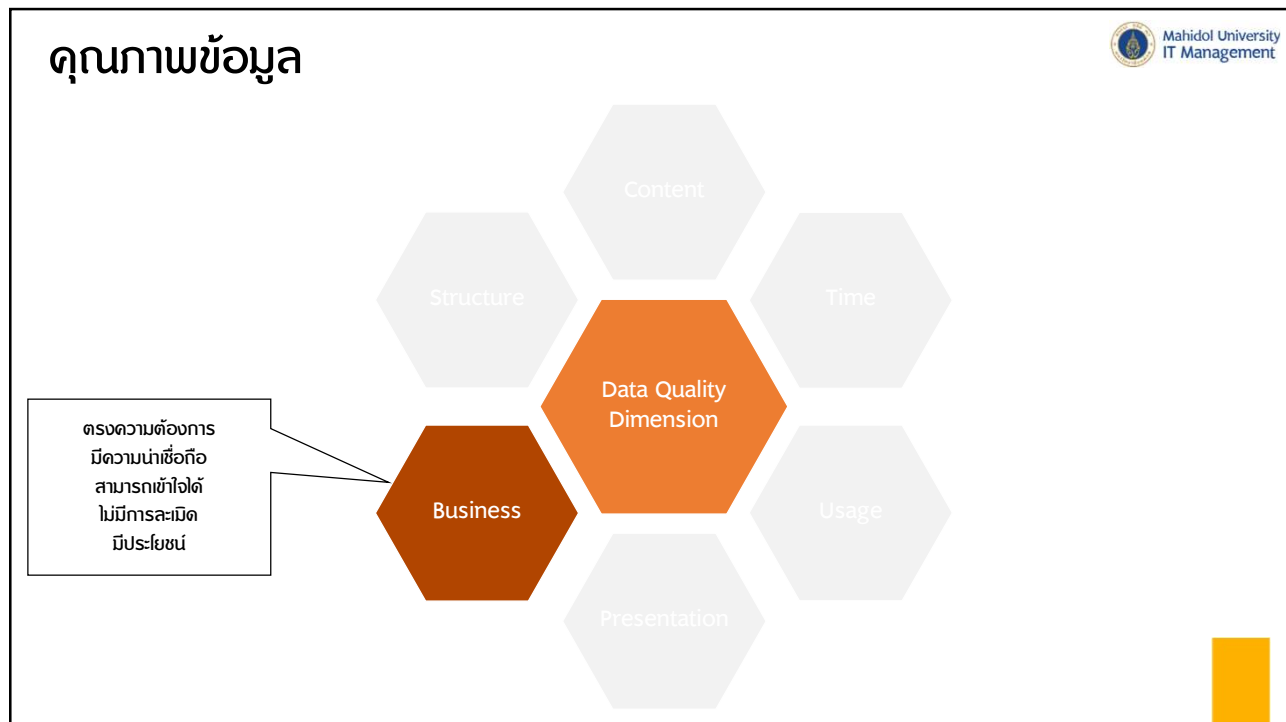


คุณภาพข้อมูล

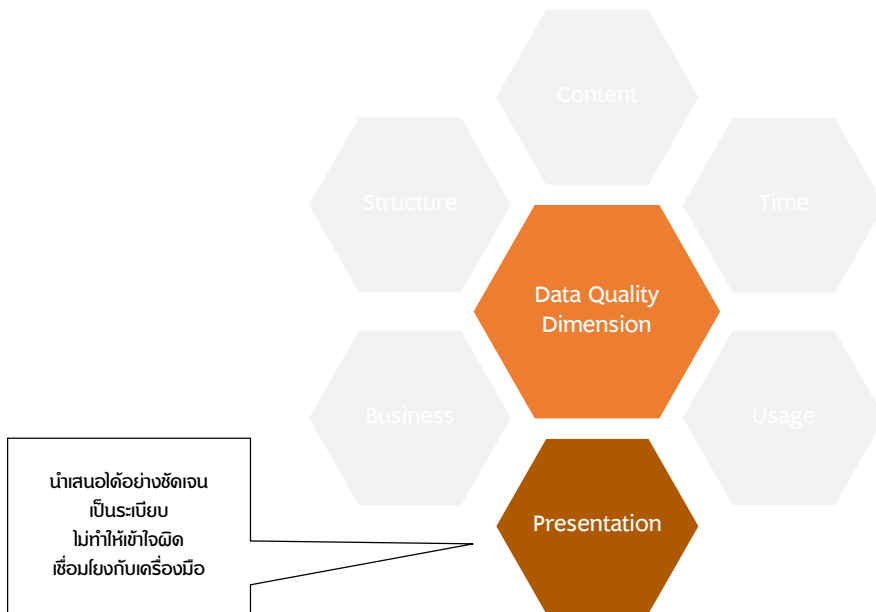
ถูกต้อง
ครบถ้วน
มีความเฉพาะเจาะจง
มีรายละเอียดเพียงพอ
ไม่ขัดแย้งกับข้อมูลอื่น







คุณภาพข้อมูล



Data Quality Business Rules

Definitional conformance	
Value presence and record completeness	
Format compliance	
Value domain membership	
Range conformance	
Mapping conformance	
Consistency rules	
Accuracy verification	
Uniqueness verification	
Timeliness validation	

Mahidol University
IT Management

Sales	Quantity	Profit	Discount	Region	State	Row ID	Order ID	Segment	Customer ID	Customer Name	Ship Mode	Order Date	Ship Date
18.648	7	-12.432	0.7	South	North Carolina	231	US-2015-156216	Corporate	EA-14035	Erin Ashbrook	Standard Class	09/13/2015, 12:00:00 AM	09/17/2015, 12:00:00 AM
176.384	2	22.298	0.2	South	Florida	315	CA-2015-167850	Corporate	AG-10525	Andy Gerboode	Standard Class	08/09/2015, 12:00:00 AM	08/16/2015, 12:00:00 AM
15.552	3	5.4422	0.2	South	Florida	316	CA-2015-167850	Corporate	AG-10525	Andy Gerboode	Standard Class	08/09/2015, 12:00:00 AM	08/16/2015, 12:00:00 AM
39.072	6	9.768	0.2	South	North Carolina	404	CA-2015-155208	Corporate	SP-20650	Stephanie Phelps	Standard Class	04/16/2015, 12:00:00 AM	04/20/2015, 12:00:00 AM
10.368	2	3.6288	0.2	South	North Carolina	705	CA-2015-138527	Corporate	BN-11470	Brad Norvell	Standard Class	09/12/2015, 12:00:00 AM	09/17/2015, 12:00:00 AM
166.84	5	18.7655	0.2	South	North Carolina	706	CA-2015-138527	Corporate	BN-11470	Brad Norvell	Standard Class	09/12/2015, 12:00:00 AM	09/17/2015, 12:00:00 AM
15.216	1	2.2824	0.2	South	North Carolina	707	CA-2015-138527	Corporate	BN-11470	Brad Norvell	Standard Class	09/12/2015, 12:00:00 AM	09/17/2015, 12:00:00 AM
11.36	2	5.3392	0	South	Louisiana	764	CA-2015-162775	Corporate	CS-12250	Chris Salesnick	Second Class	01/13/2015, 12:00:00 AM	01/15/2015, 12:00:00 AM

57

Mahidol University
IT Management

Kahoot!

•QRCODE

58



Mahidol University
Wisdom of the Land