

Data Analysis 1

Week 1 (16:00, 15/09/2021)

Vivatsathorn Thitasirivit

Student. Faculty of Engineering, Chulalongkorn University.



DA101 Course Syllabus

Required: Basic Mathematical Knowledge: Algebra, Geometry and Statistics
Basic Knowledge on Physics, Astrophysics, Astronomy and Science
Basic Knowledge on Mechanical and Electrical Engineering
Basic Programming Skills

Languages: Thai (Main), English

Curriculum: Fundamental and Advanced Statistics with Methodology, Software Usages and Applications in Scientific and Engineering Studies

Outlines:

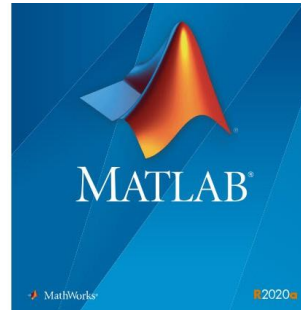
1. Types of Data and Data Acquisition (3/0)
2. Mathematics (8/4)
3. Scientific Experimentation (2/2)
4. Data Analysis (2/2)

Min. Hrs.: 23

Necessary Equipment/Software



Scientific Calculator



MATRIX LABORATORY



Microsoft Excel



Python



Graphing Software: Desmos

Week 1 Outlines

Types of Data and Data Acquisition

- Ways of thinking about data
- Types of data
- Data strategies
- Data framework
- Data acquisition

Assignments

(none)

Google Classroom



3mxy7qn

1. Ways of thinking about data

What do analysts do?

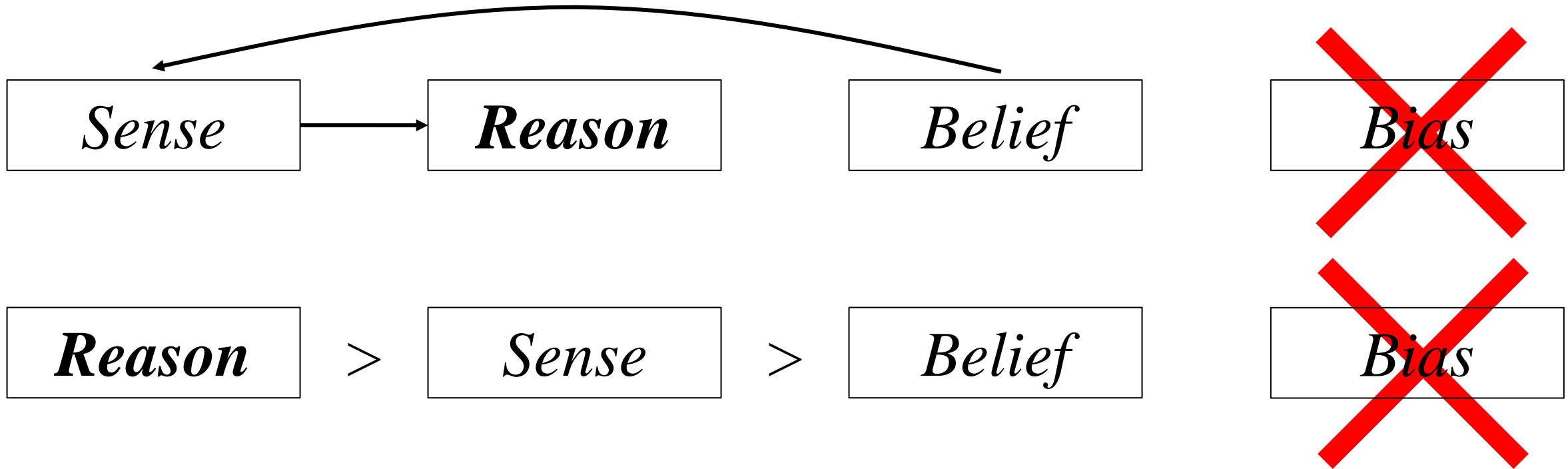
- Do the mission planning
- Design overall process and subsystems

- Perform overall and detailed calculation
- Perform mathematical/graphical approaches to problems
- Perform possibility check based on theory and statistics

- Visualize information
- Minimize errors
- Perfect the minor details which no one (<1%) looks at

1. Ways of thinking about data

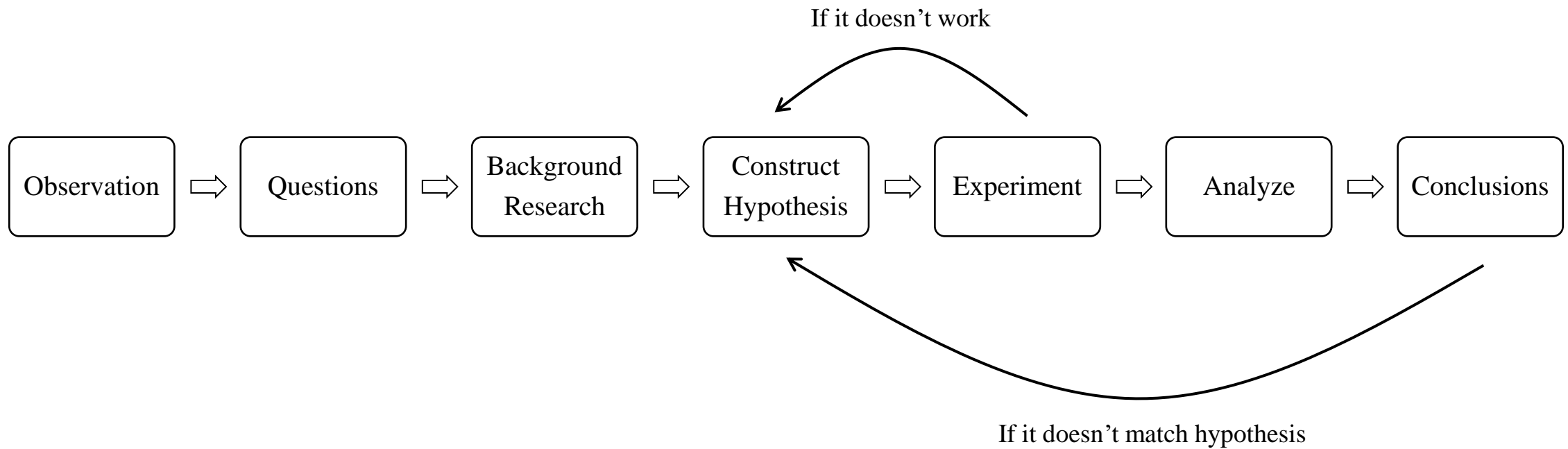
Key to be an analyst



1. Ways of thinking about data

The Methodology

Scientific Method



1. Ways of thinking about data

Definition of Data

data noun, plural in form but singular or plural in construction, often attributive

 Save Word

da·ta | \ 'dā-tə , 'da-  also 'dä-  \

Definition of *data*

1 : factual information (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation

// the *data* is plentiful and easily available

— H. A. Gleason, Jr.

// comprehensive *data* on economic growth have been published

— N. H. Jacoby

Merriam Webster Dictionary—*Definition of “data”*

1. Ways of thinking about data

What can be Data?

“ล่าสุด ณ วันที่ 13 กันยายน 2564 มียอดผู้เสียชีวิตจากโควิดทั้งสิ้น 14,353 คน”

“โทรศัพท์ iPhone รุ่นแรกเปิดตัวเมื่อปี ค.ศ. 2007”

“มดดำมีน้ำหนักเฉลี่ย 5 มิลลิกรัม”

“สัรถยนต์ในไทยที่นิยมมากที่สุดคือ สีขาว (32%)”

“หนี้สาธารณะไทยมีแนวโน้มที่จะมากขึ้นเรื่อย ๆ”

1. Ways of thinking about data

What can be Data?

Timestamp *Data type* *Number*
“ล่าสุด ณ วันที่ 13 กันยายน 2564 มียอดผู้เสียชีวิตจากโควิดทั้งสิ้น 14,353 คน”

Date Information
“โทรศัพท์ iPhone รุ่นแรกเปิดตัวเมื่อปี ค.ศ. 2007”

Representation
“มดดำมีน้ำหนักเฉลี่ย 5 มิลลิกรัม”

Number

“สัรถยนต์ในไทยที่นิยมมากที่สุดคือ สีขาว (32%)”

Trend *Tendency*
“หนี้สาธารณะไทยมีแนวโน้มที่จะมากขึ้นเรื่อย ๆ”

1. Ways of thinking about data

Data Interpretation

8,485,690 คน

ข้อมูลดิบ
(Raw data)

คิดเป็น

12.82%

ของ ...

ข้อมูลที่ผ่านการประมวลผล
(Processed data)

1. Ways of thinking about data

Data Interpretation

3	2	3
2	0	1
1	1	0

Matrix-defined

3,2,3
2,0,1
1,1,0

Comma-delimited

can be either

3	2	3
2	0	1
1	1	0

	Apple	Orange	Banana
Somchai	3	2	3
Somsri	2	0	1
Sompong	1	1	0

ข้อมูลดิบ
(Raw data)

ข้อมูลที่ผ่านการประมวลผล
(Processed data)

2. Types of Data

Quantity & Quality

Quantity

Quality

Quantitative

Qualitative

Quantify

Qualify

Quantification

Qualification

2. Types of Data

Quantity & Quality

	ชาย (คน)	หญิง (คน)	สัดส่วน (ร้อยละ)
0-14 ปี	5,512,470	5,209,331	16.20
15-24 ปี	4,333,814	4,151,876	12.82
25-54 ปี	14,584,071	14,963,073	44.64
55-64 ปี	3,891,464	4,466,079	12.63
65 ปีขึ้นไป	3,409,472	4,485,814	11.93

DOI.go.th—Thailand Population in 2563

ข้อมูลเชิงปริมาณ
(Quantitative data)

ข้อมูลแบบผสม
(Mixed record)

ข้อมูลเชิงคุณภาพ
(Qualitative data)

	สมหมาย	สมหญิง
เพศกำเนิด	ชาย	หญิง
อายุ	19	37
สุขภาพ	ดีมาก	ดี
สีม่านตา	ดำ	น้ำตาล
สีที่ชอบ	เทอร์คอยซ์	ลาเวนเดอร์

Som Mai and Som Ying Sample Data

2. Types of Data

Qualitative Data



Wikipedia—*Bangkok in 2010*



CINE—*Photogrammetry*

ภาพถ่าย (Photograph/Picture/Image)

ภาพนิ่ง (Still image)

ภาพเคลื่อนไหว (Videography)

ภาพช่วงคลื่นต่าง ๆ (Multispectral P.) →

ภาพ 3 มิติ

วัตถุ 3 มิติ (3D Object)

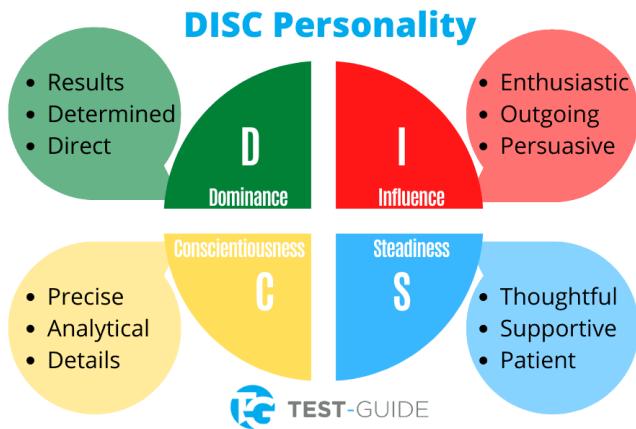
เซตภาพรังวัดรูปแบบต่าง ๆ (Photogrammetry)

Radar และ LiDAR



2. Types of Data

Qualitative Data



แบบสำรวจให้คะแนนหรือแบ่งเกณฑ์

ให้คะแนนเป็นช่วงตัวเลข

แบ่งเกณฑ์ด้วยช่วงตัวเลข*

ใช่/ไม่ใช่

แบบสำรวจปลายเปิด

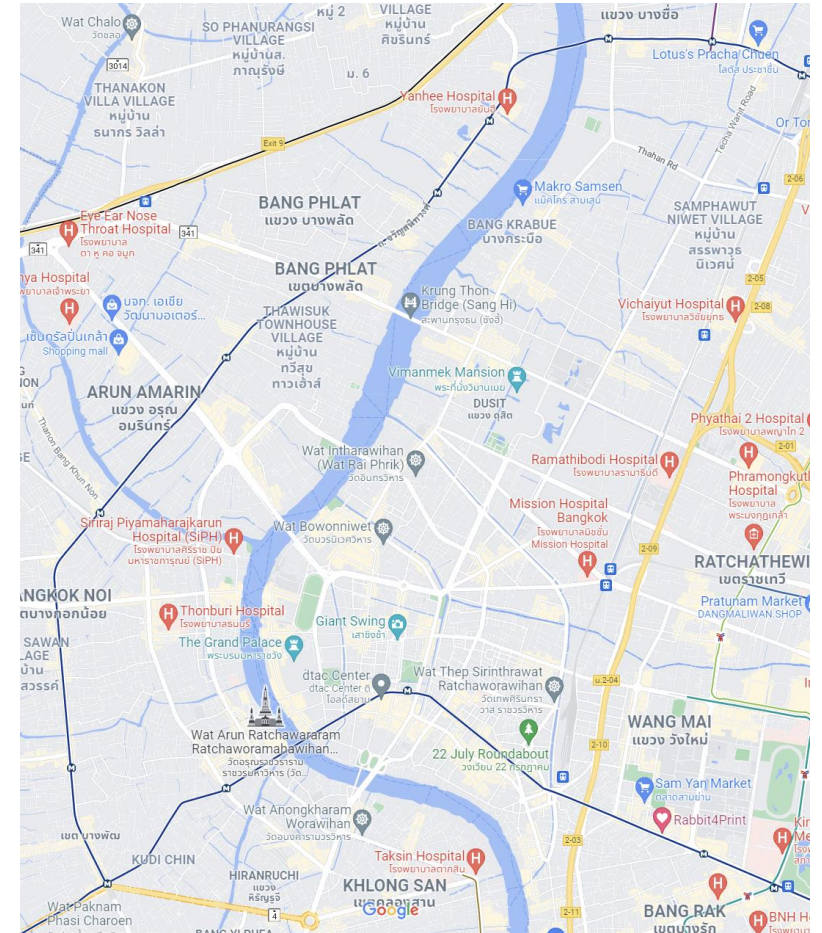
แผนที่ แผนผัง เส้นทางการเดินทาง

แผนที่แบบต่าง ๆ

แผนผังสถานที่ เช่น แปลนบ้าน

เส้นทางที่ใช้มาจาก P_1 มา P_2

อื่น ๆ อีกมากมาย



2. Types of Data

Quantitative Data

*ข้อมูลที่เป็นตัวเลขอะไรก็ได้ที่นำมาคำนวณออกมาเป็นตัวเลขได้
โดยผลลัพธ์ไม่ออกมาเป็นข้อมูลเชิงคุณภาพ*

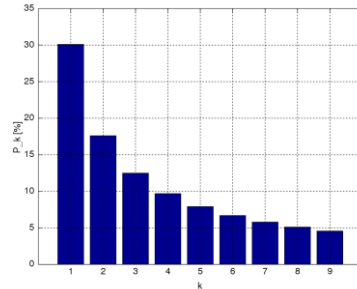
2. Types of Data

Analysis Approach

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

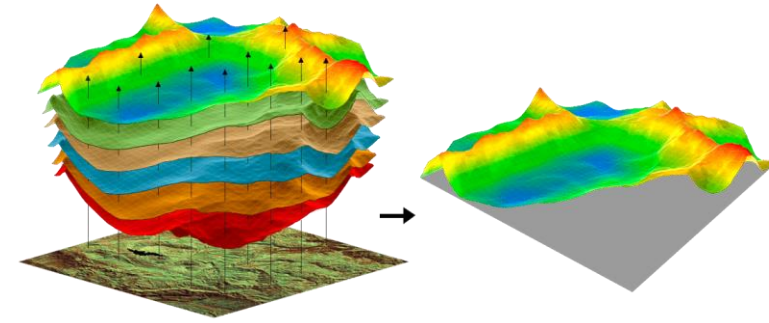
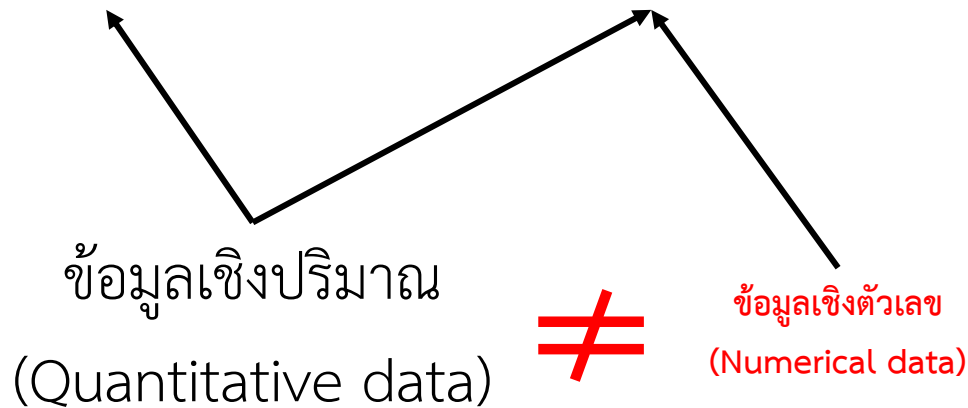
Bayes' Theorem

การวิเคราะห์เชิงปริมาณ
(Quantitative analysis)



Benford's Law

การวิเคราะห์เชิงตัวเลข
(Numerical analysis)



- การวิเคราะห์เชิงพื้นที่ (Spatial analysis)
- การแปลตีความภาพด้วยสายตา (Visual Analysis)
- การวิเคราะห์ข้อมูลภาพด้วยคอมพิวเตอร์ (Image Processing)
- การวิเคราะห์ลักษณะคลื่น (Waveform Analysis)
- กรณีศึกษา (Case study)
- การบรรยายคุณลักษณะ (Characteristic Narrative)

อื่น ๆ

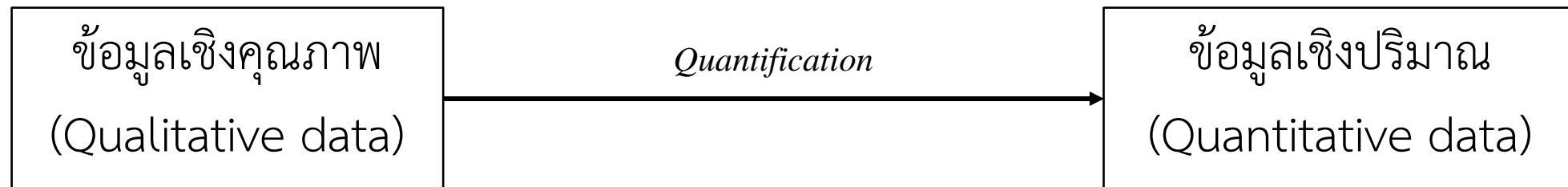
การวิเคราะห์เชิงคุณภาพ
(Qualitative analysis)



ข้อมูลเชิงคุณภาพ
(Qualitative data)

2. Types of Data

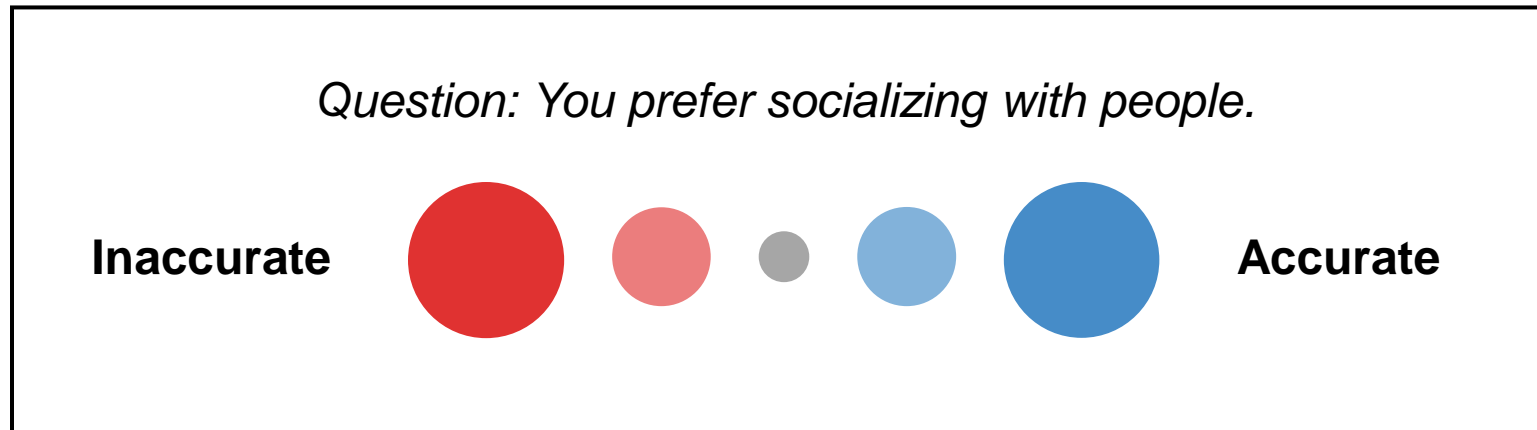
Analysis Approach



2. Types of Data

Analysis Approach

Quantification Example

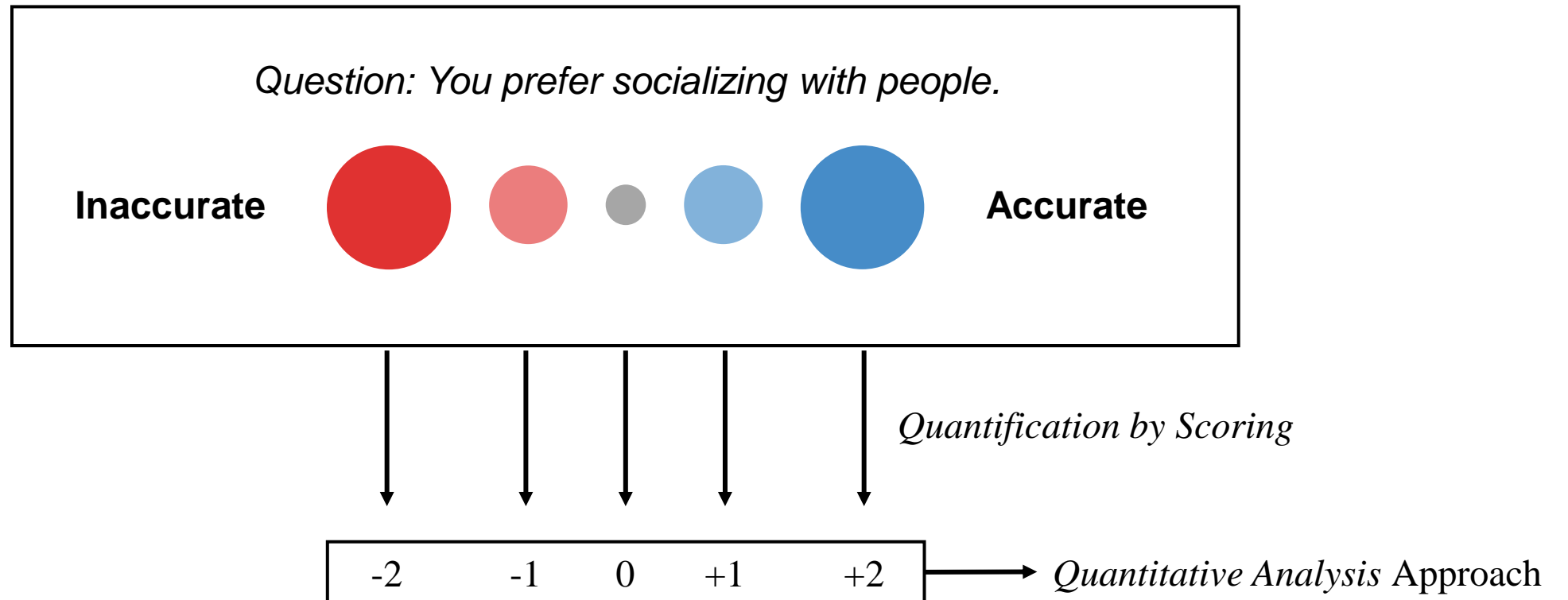


2. Types of Data

Analysis Approach

Quantification Example: Scoring

Qualitative Data describing test taker's sociability



2. Types of Data

Analysis Approach

Quantification Example

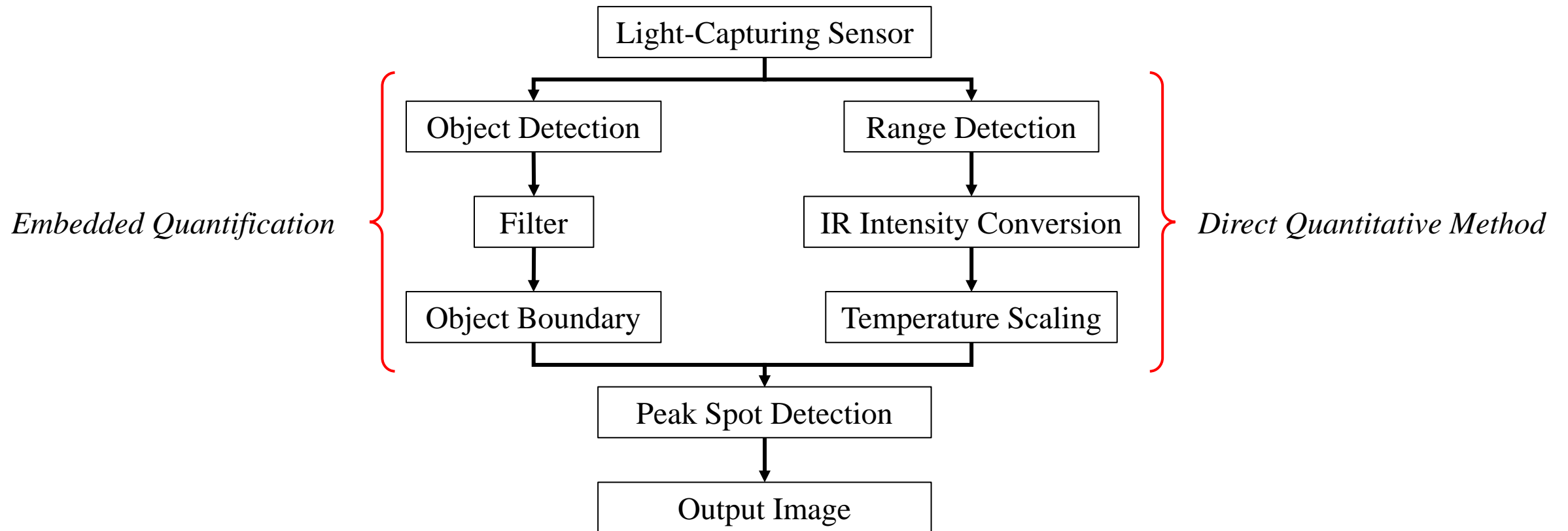


Oasis Scientific Inc.—*Fever Screening Thermal Camera*

2. Types of Data

Analysis Approach

Quantification Example: Image Processing



3. Data Strategies

For Quantitative Data

Visualization

Exploratory Analysis

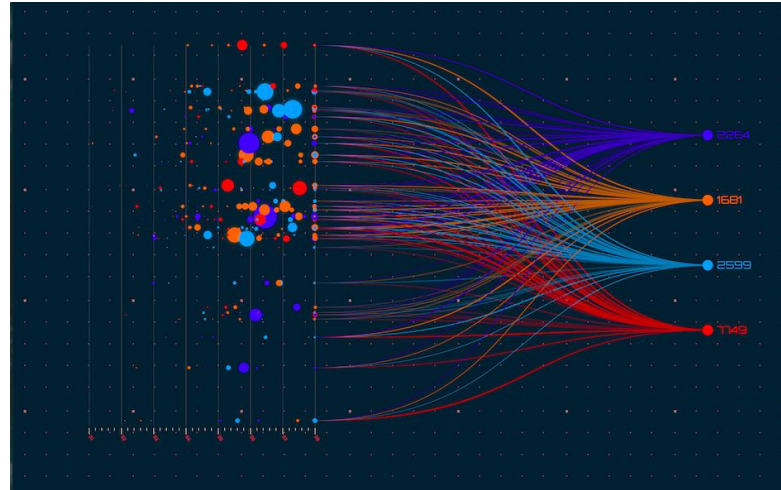
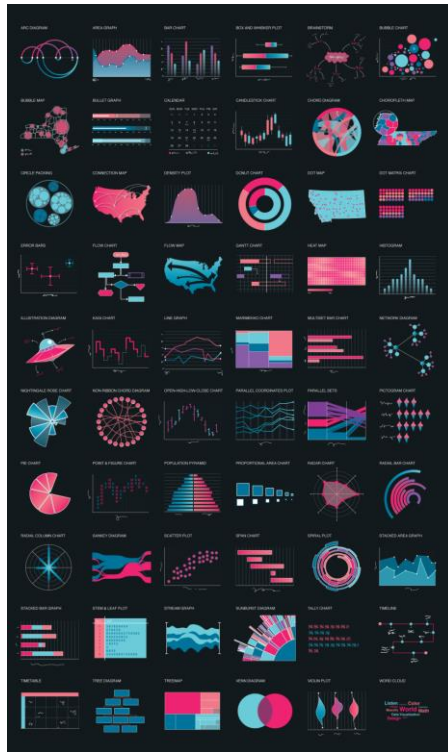
Trend Analysis

Estimation

3. Data Strategies

For Quantitative Data

Visualization



References
*finereport, ourgreenfish, treehousetechnology,
Olivia Davis*

3. Data Strategies

For Quantitative Data

Visualization

Chart Types:

I. Data Comparison Chart

- Comparison
- Composition
- Distribution

II. Data Reduction Chart

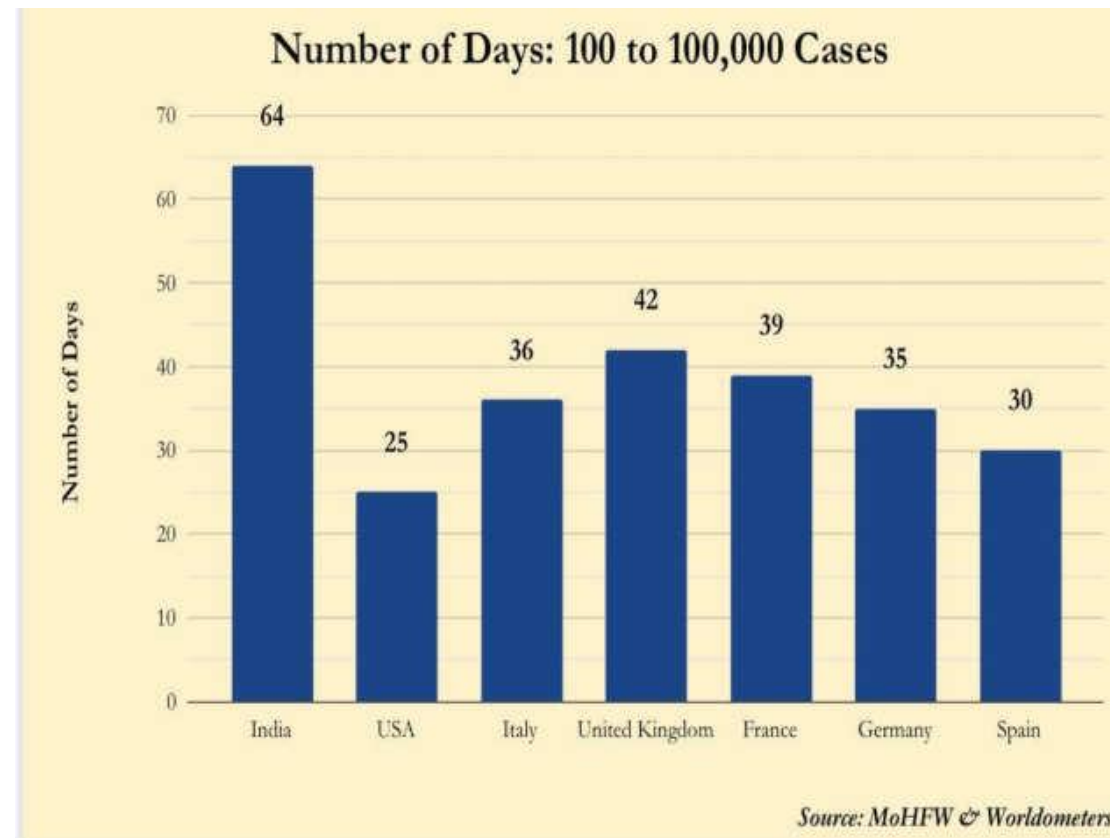
- Evolution
- Relationship (Correlation)
- Profiling (Relative examination)

3. Data Strategies

For Quantitative Data

Visualization—Data Comparison Chart: Comparison

Bar Chart

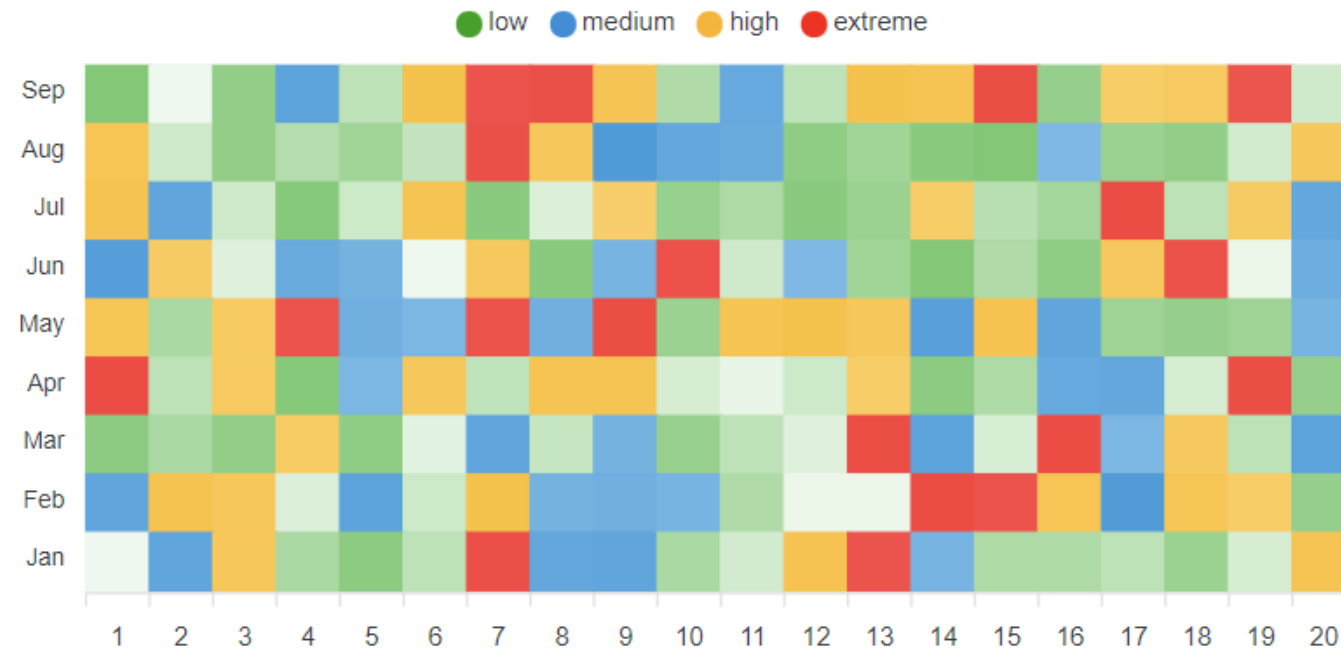


3. Data Strategies

For Quantitative Data

Visualization—Data Comparison Chart: Comparison

Heat map



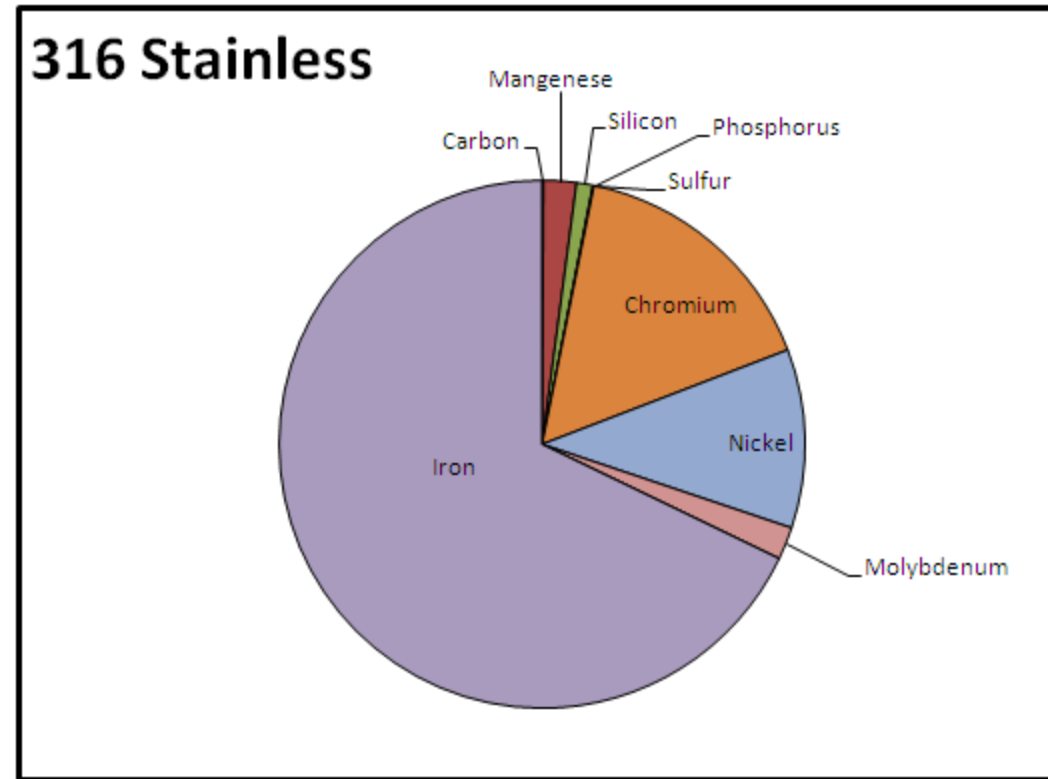
ApexCharts

3. Data Strategies

For Quantitative Data

Visualization—Data Comparison Chart: Composition

Pie/M.D. Pie



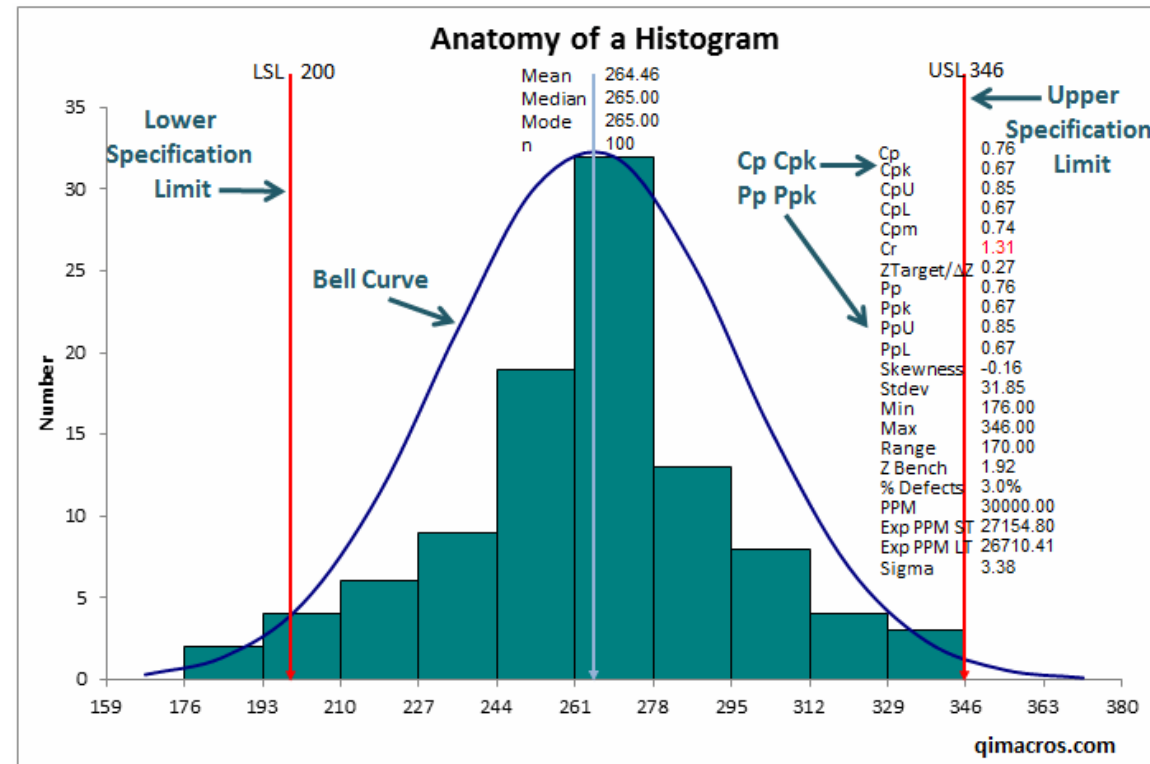
Connected Discourse

3. Data Strategies

For Quantitative Data

Visualization—Data Comparison Chart: Distribution

Histogram



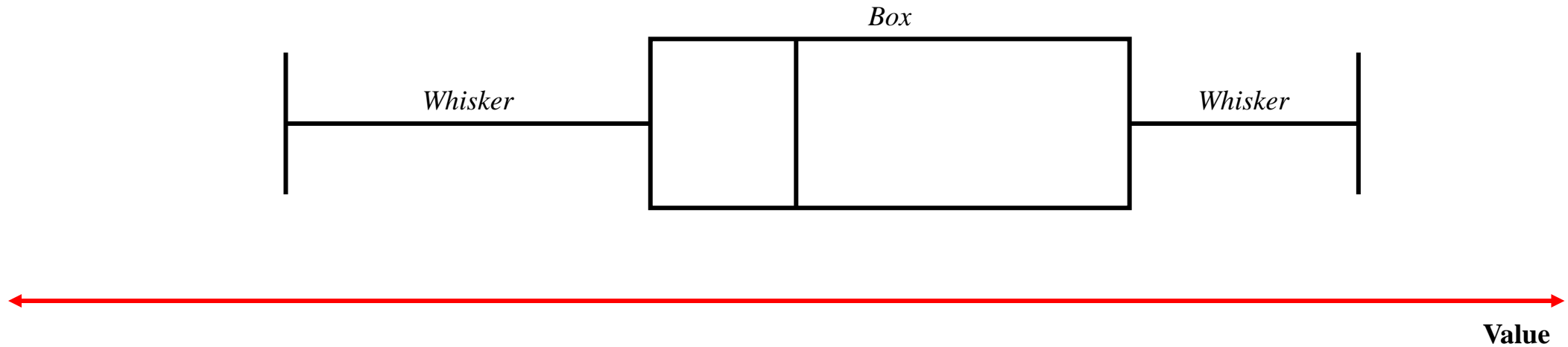
QIMacros.com

3. Data Strategies

For Quantitative Data

Visualization—Data Comparison Chart: Distribution

Box and Whisker Plot

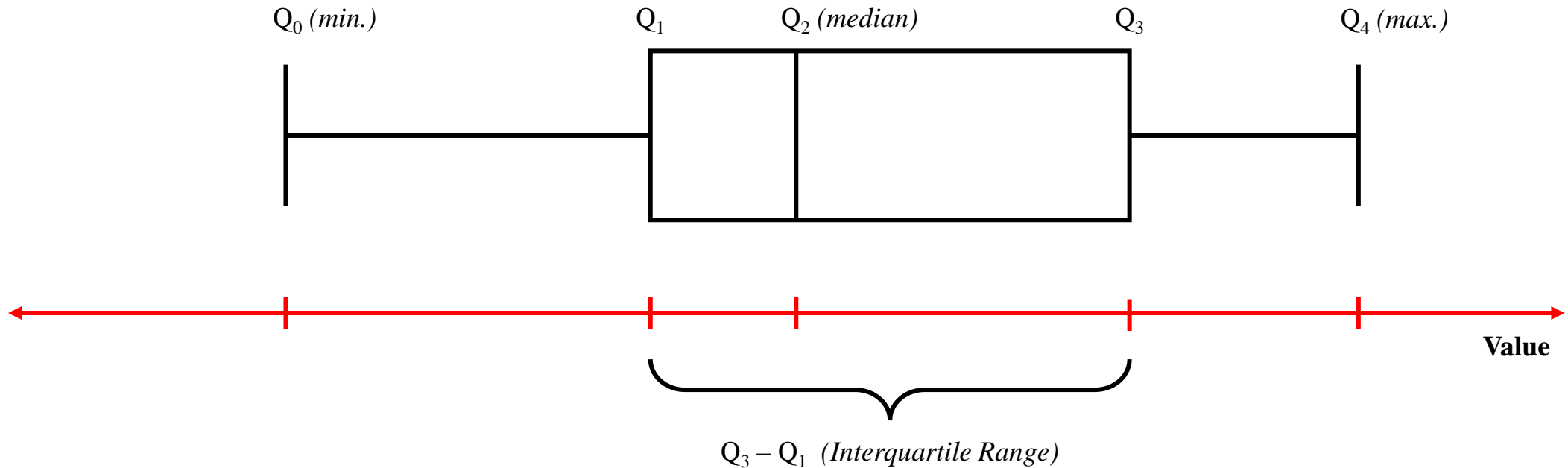


3. Data Strategies

For Quantitative Data

Visualization—Data Comparison Chart: Distribution

Box and Whisker Plot

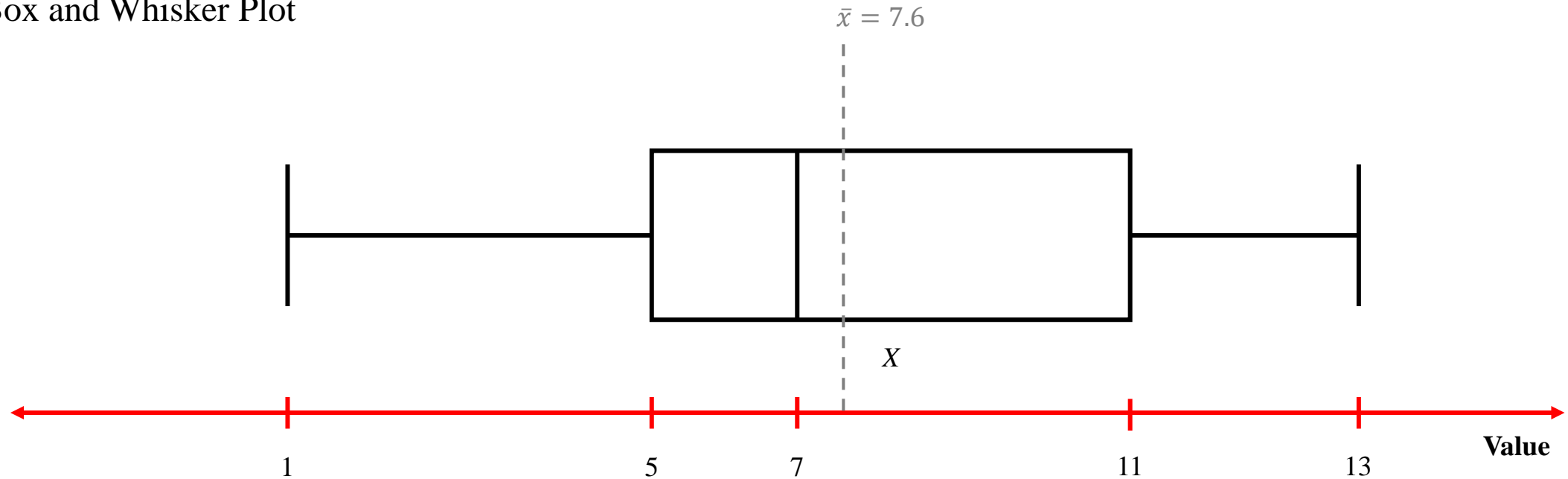


3. Data Strategies

For Quantitative Data

Visualization—Data Comparison Chart: Distribution

Box and Whisker Plot



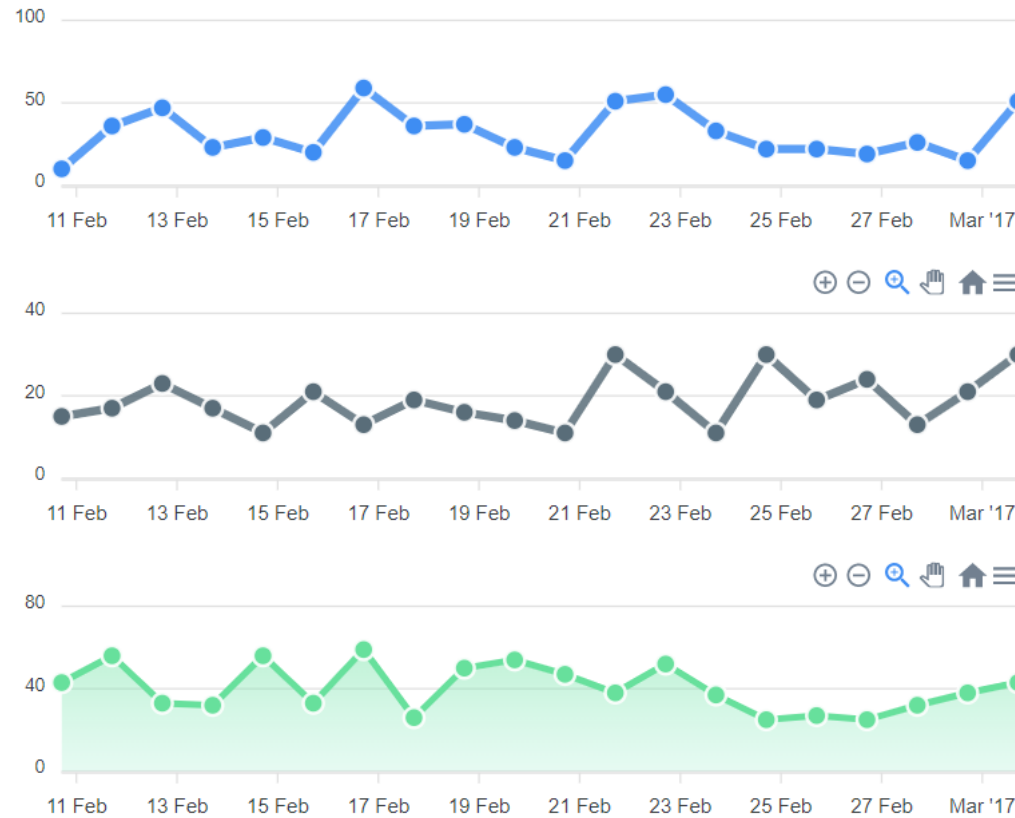
Data X: 1, 5, 6, 7, 10, 11, 13

3. Data Strategies

For Quantitative Data

Visualization—Data Reduction Chart: Evolution

Dot/Line/Area/Combo



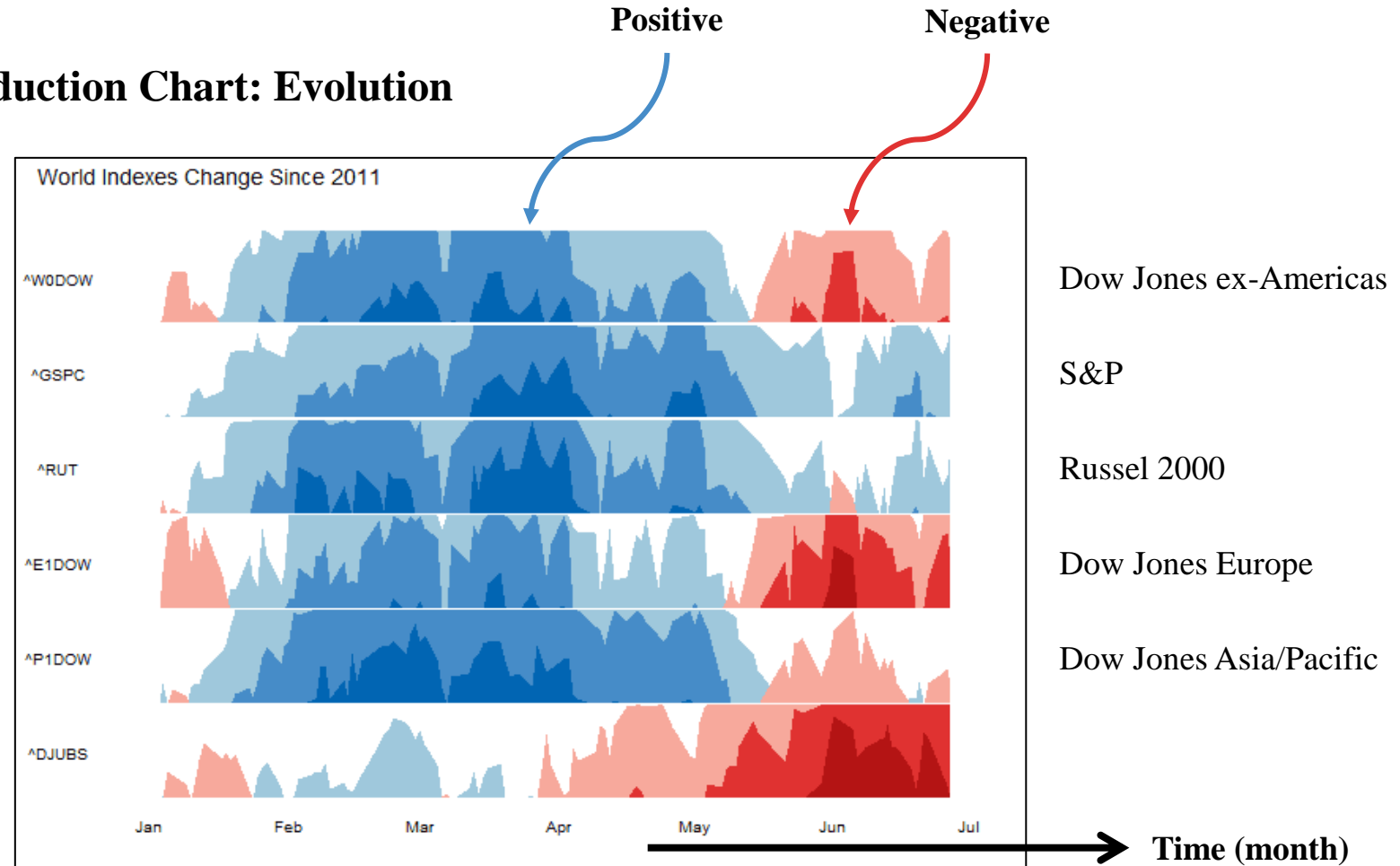
ApexCharts

3. Data Strategies

For Quantitative Data

Visualization—Data Reduction Chart: Evolution

Horizon



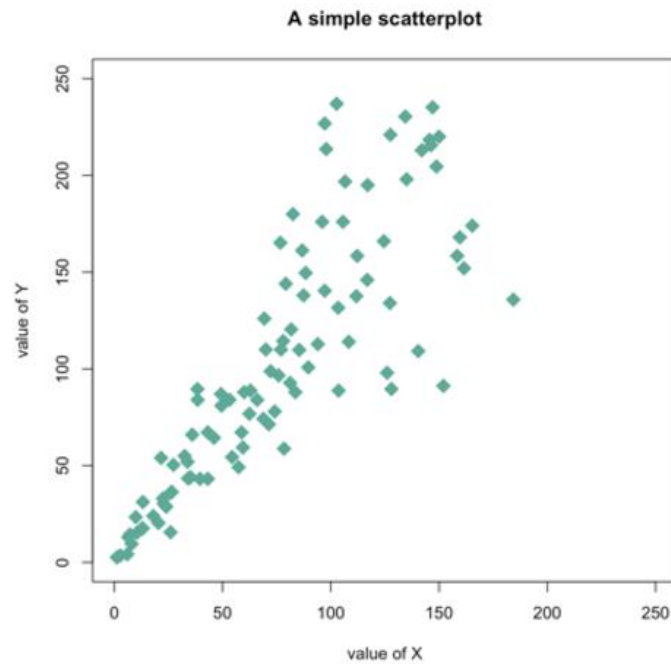
Timely Portfolio

3. Data Strategies

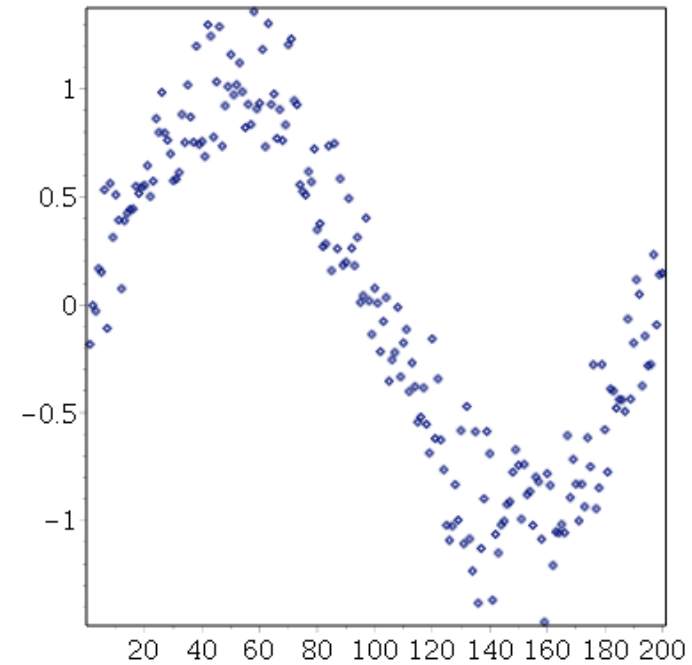
For Quantitative Data

Visualization—Data Reduction Chart: Relationship

Scatterplot



The R Graph Gallery



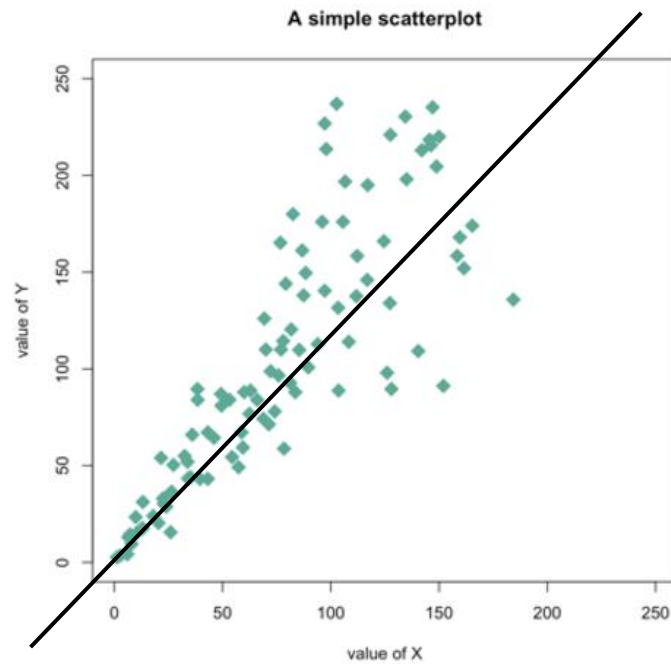
Maple Help

3. Data Strategies

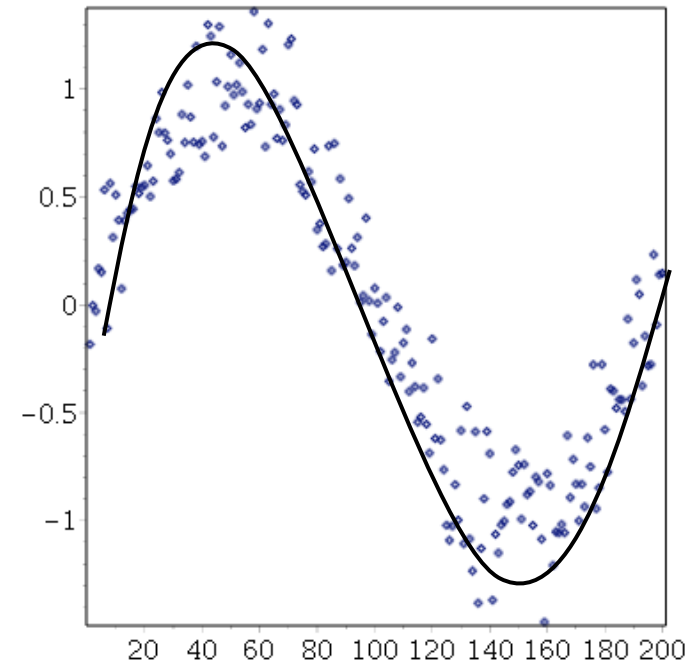
For Quantitative Data

Visualization—Data Reduction Chart: Relationship

Scatterplot
(*Reg.*)



The R Graph Gallery



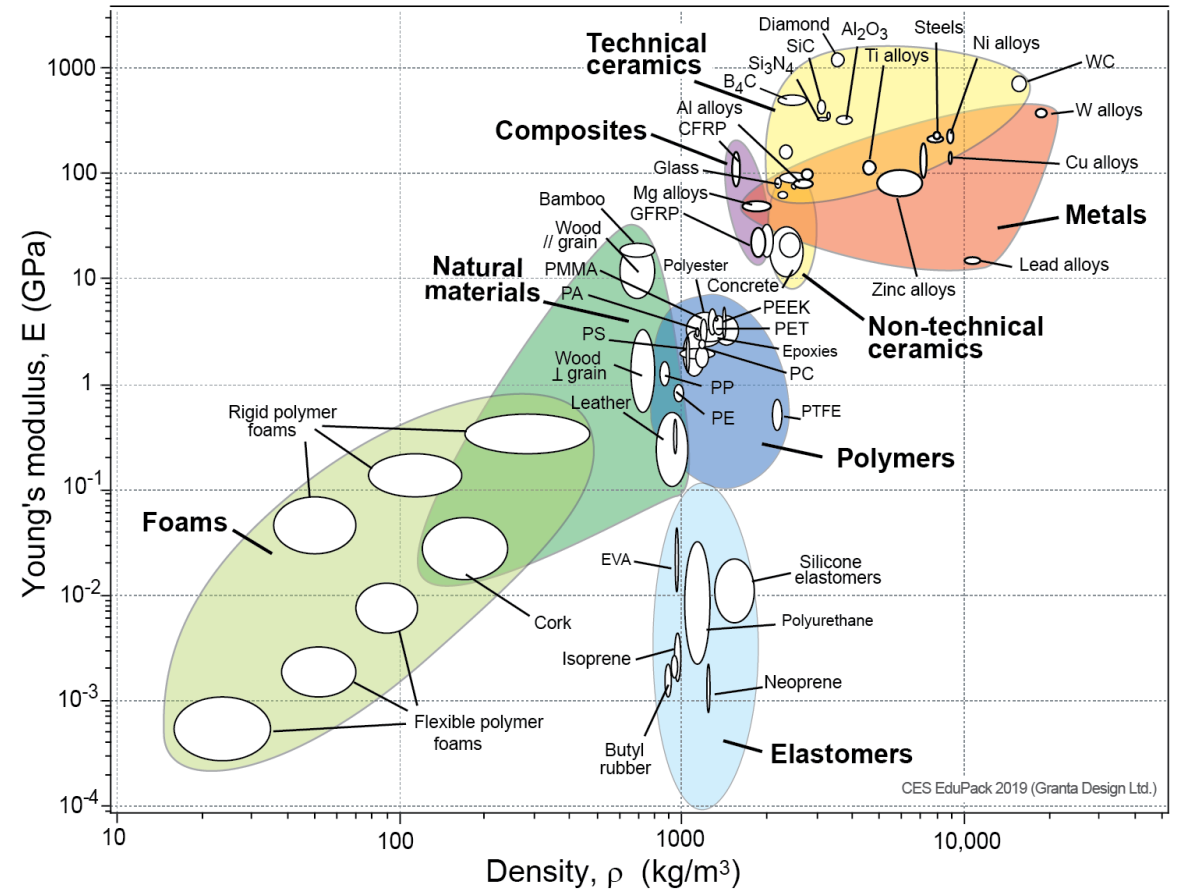
Maple Help

3. Data Strategies

For Quantitative Data

Visualization—Data Reduction Chart: Relationship

Bubble



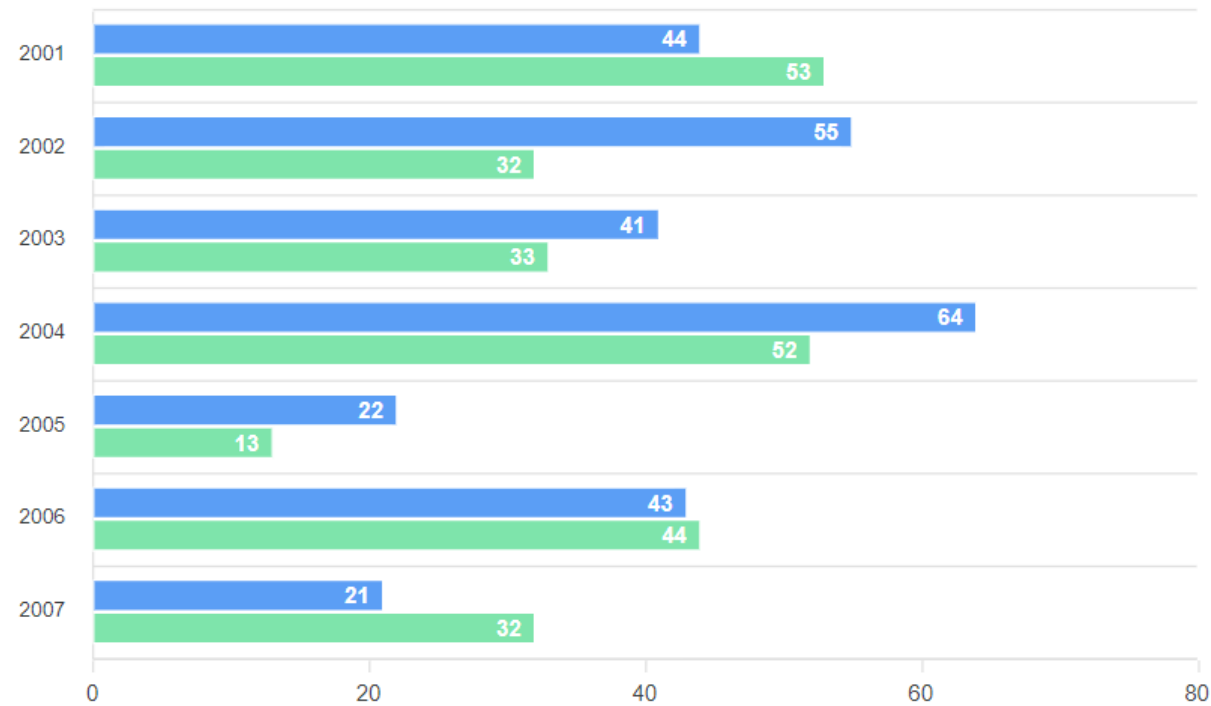
Ansys Granta—Material Chart

3. Data Strategies

For Quantitative Data

Visualization—Data Reduction Chart: Profiling

Grouped Bars



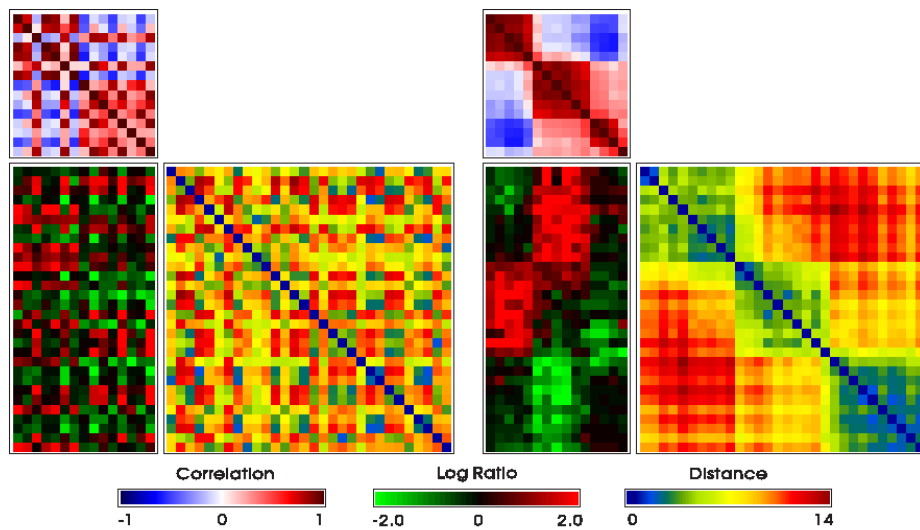
ApexCharts

3. Data Strategies

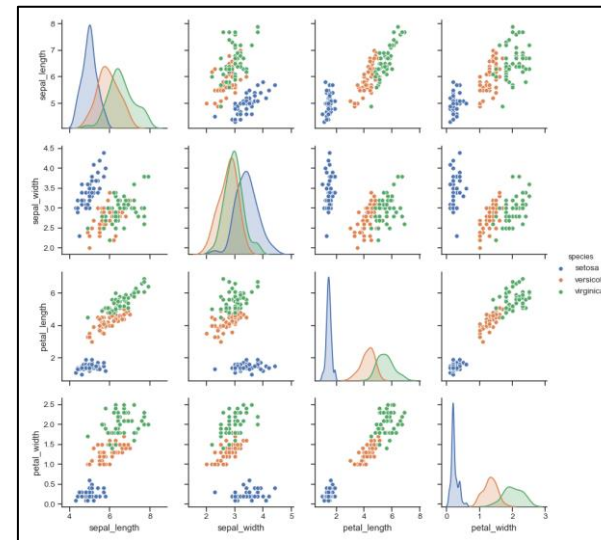
For Quantitative Data

Visualization—Data Reduction Chart: Profiling

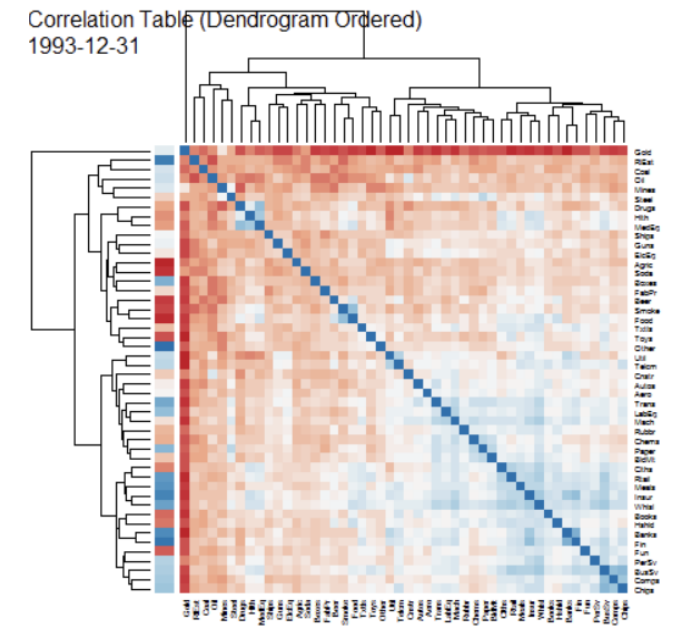
Matrix Visualization/Matrix of Visualizations



Han Ming *et al.*—*Matrix Visualization*



Raghavan—*Scatter Matrix*



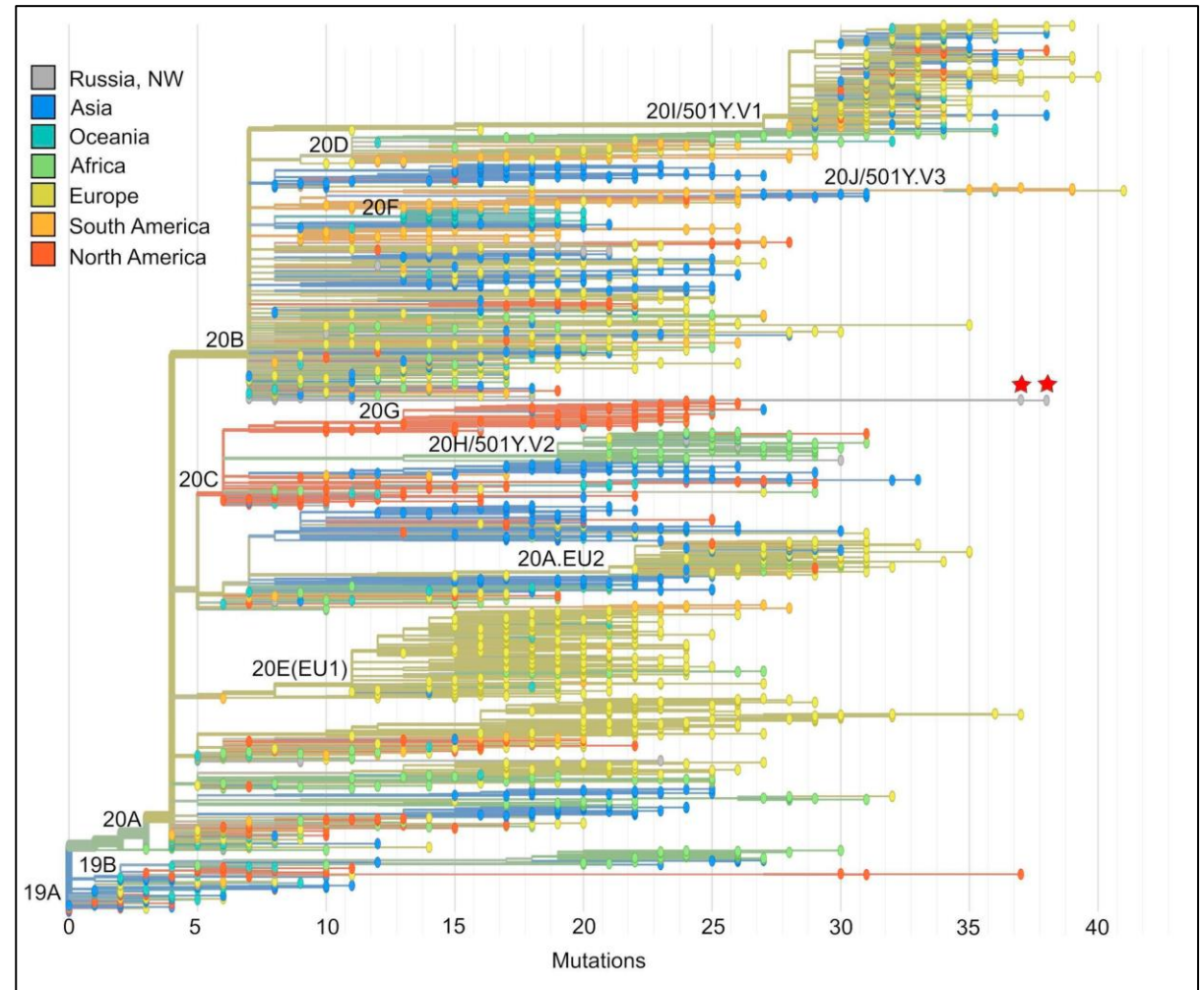
Timely Portfolio—*Correlation-Dendrogram Table*

3. Data Strategies

For Quantitative Data

Visualization—Data Reduction Chart: Evolution

Phylogenetic Tree



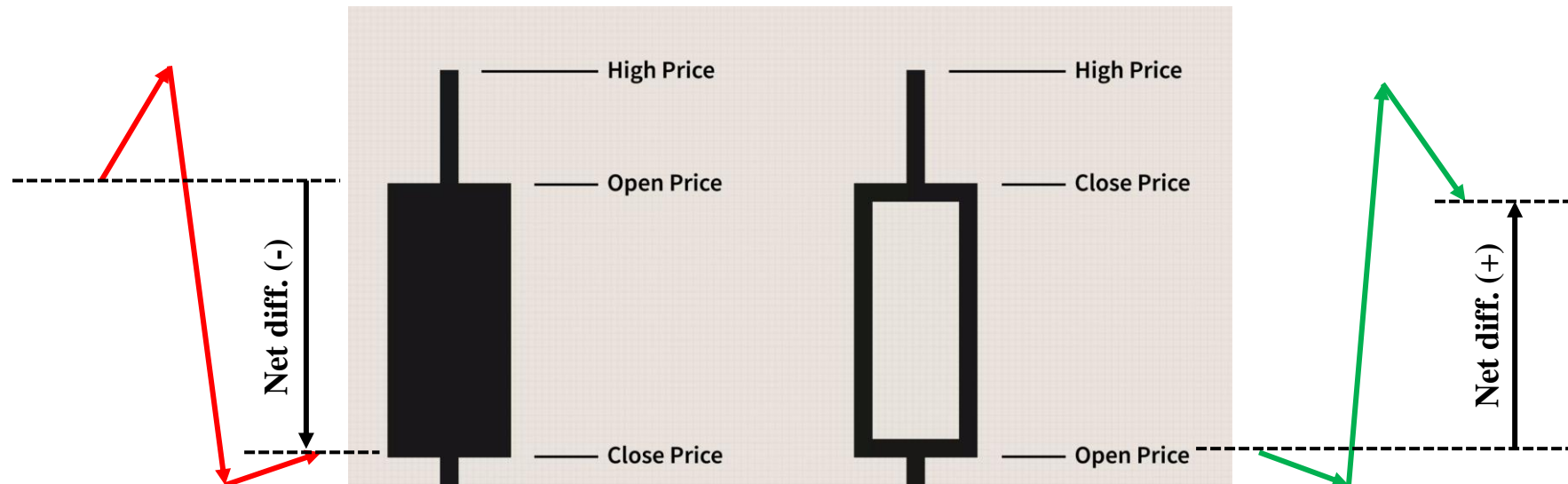
Anna et al.—SARS-CoV-2 Phylogenetic Tree (2020)

3. Data Strategies

For Quantitative Data

Visualization—etc.

Candlestick



Investopedia—*Candlestick Chart*

3. Data Strategies

For Quantitative Data

Visualization—etc.

Candlestick

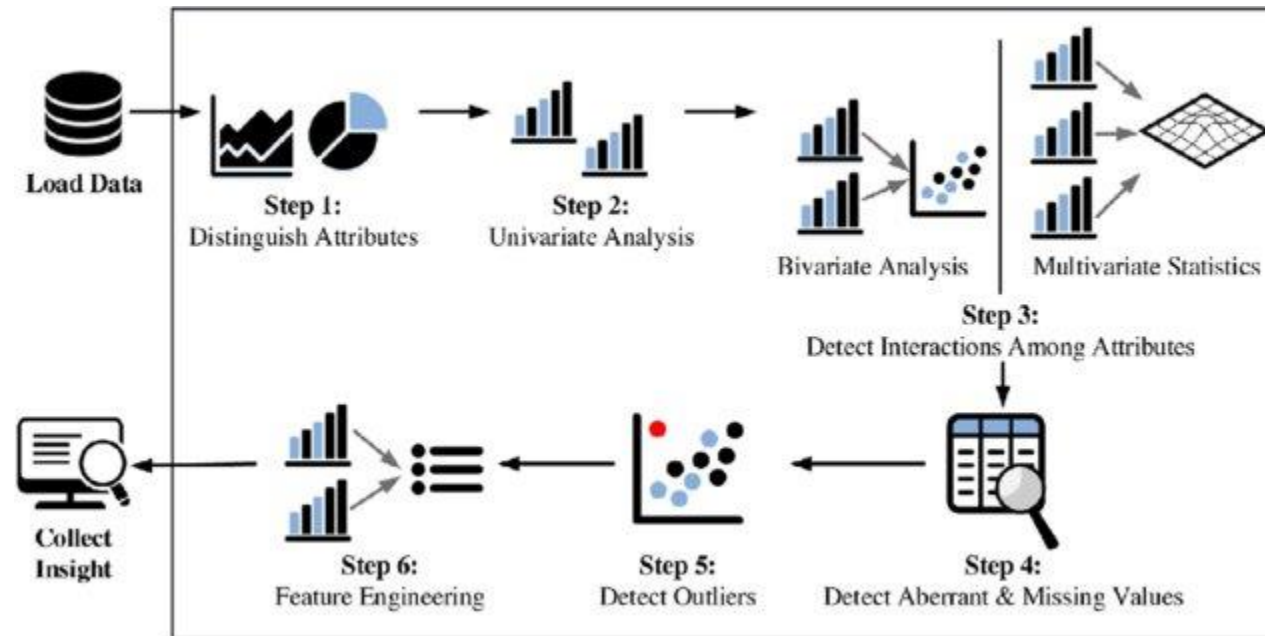


ApexCharts

3. Data Strategies

For Quantitative Data

Exploratory Data Analysis (EDA)

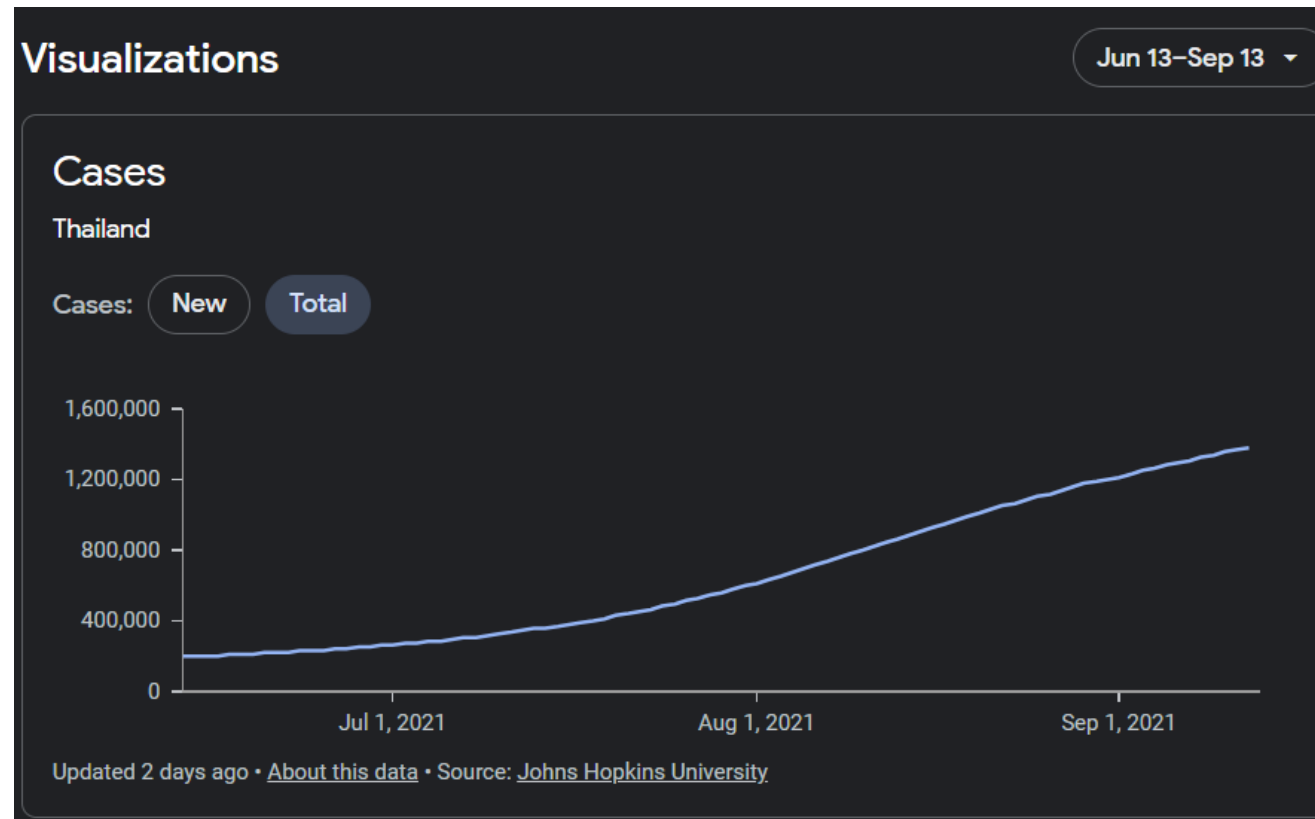


Aindrila *et al.*—The fundamental steps of the exploratory data analysis process.

3. Data Strategies

For Quantitative Data

Trend Analysis



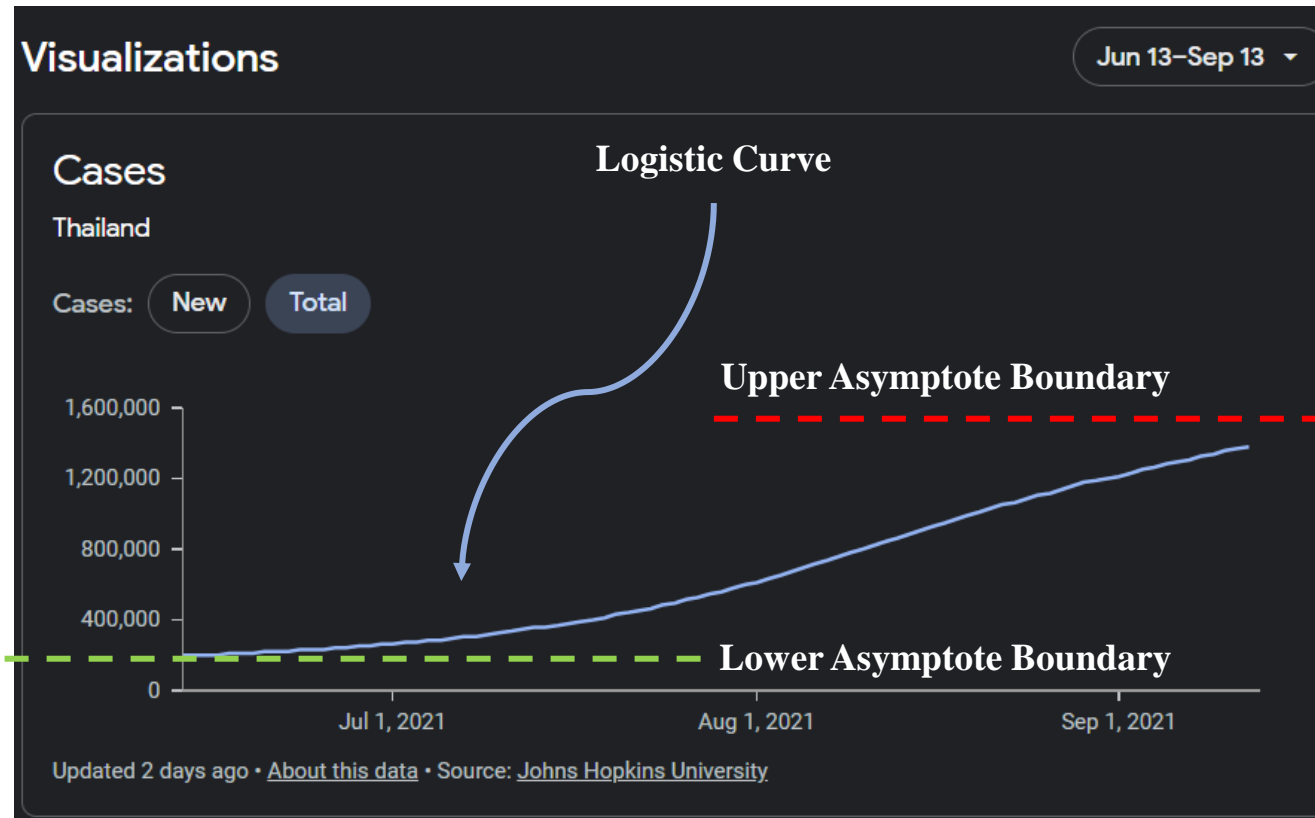
Google Src.—*Thailand COVID-19 Cases (13/06/2021–13/09/2021)*

3. Data Strategies

For Quantitative Data

Trend Analysis

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}}$$



Google Src.—*Thailand COVID-19 Total Cases (13/06/2021–13/09/2021)*

3. Data Strategies

For Quantitative Data

Trend Analysis

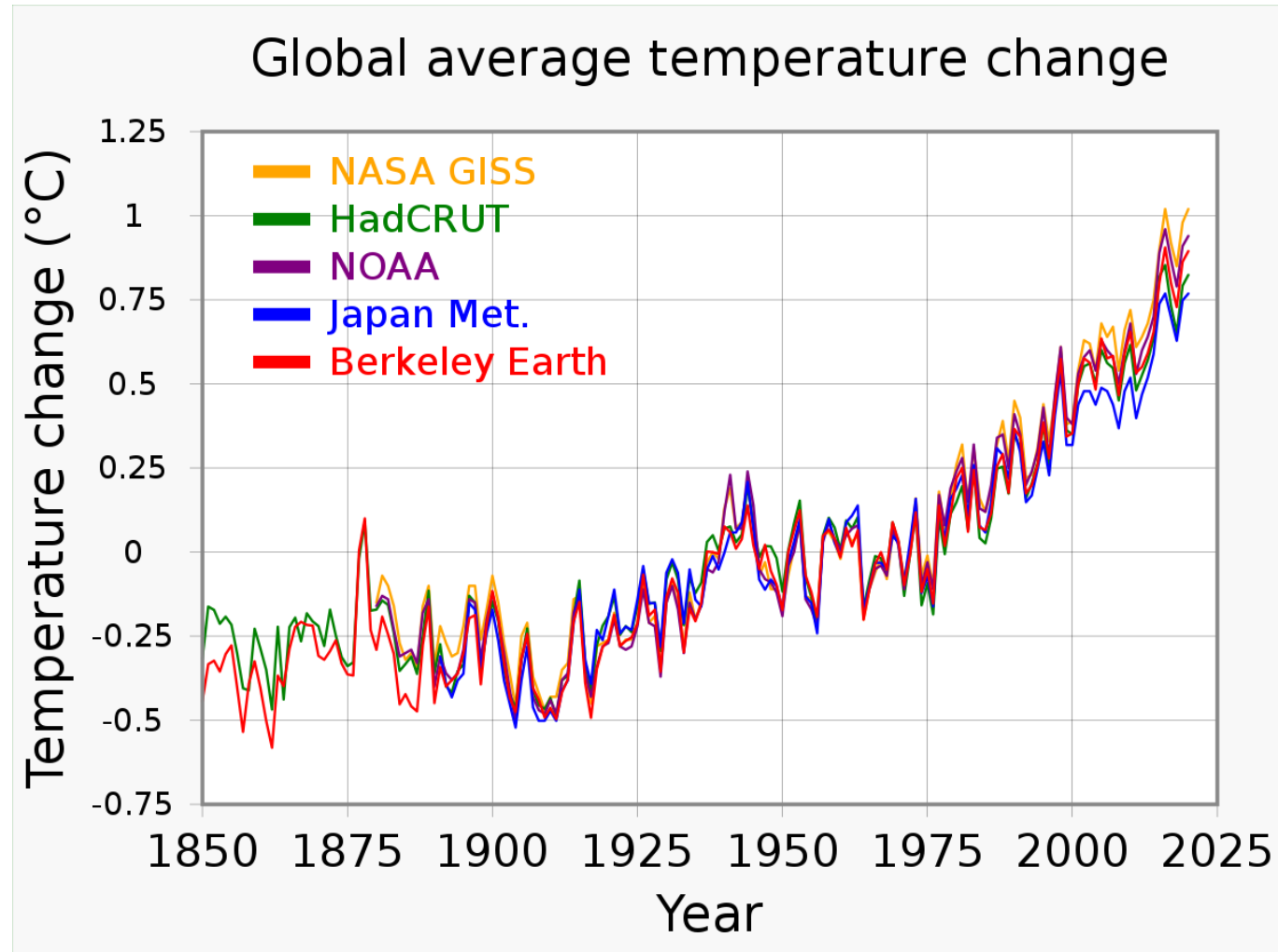


Google Src.—*Thailand COVID-19 Total Cases (13/06/2021–13/09/2021)*

3. Data Strategies

For Quantitative Data

Trend Analysis

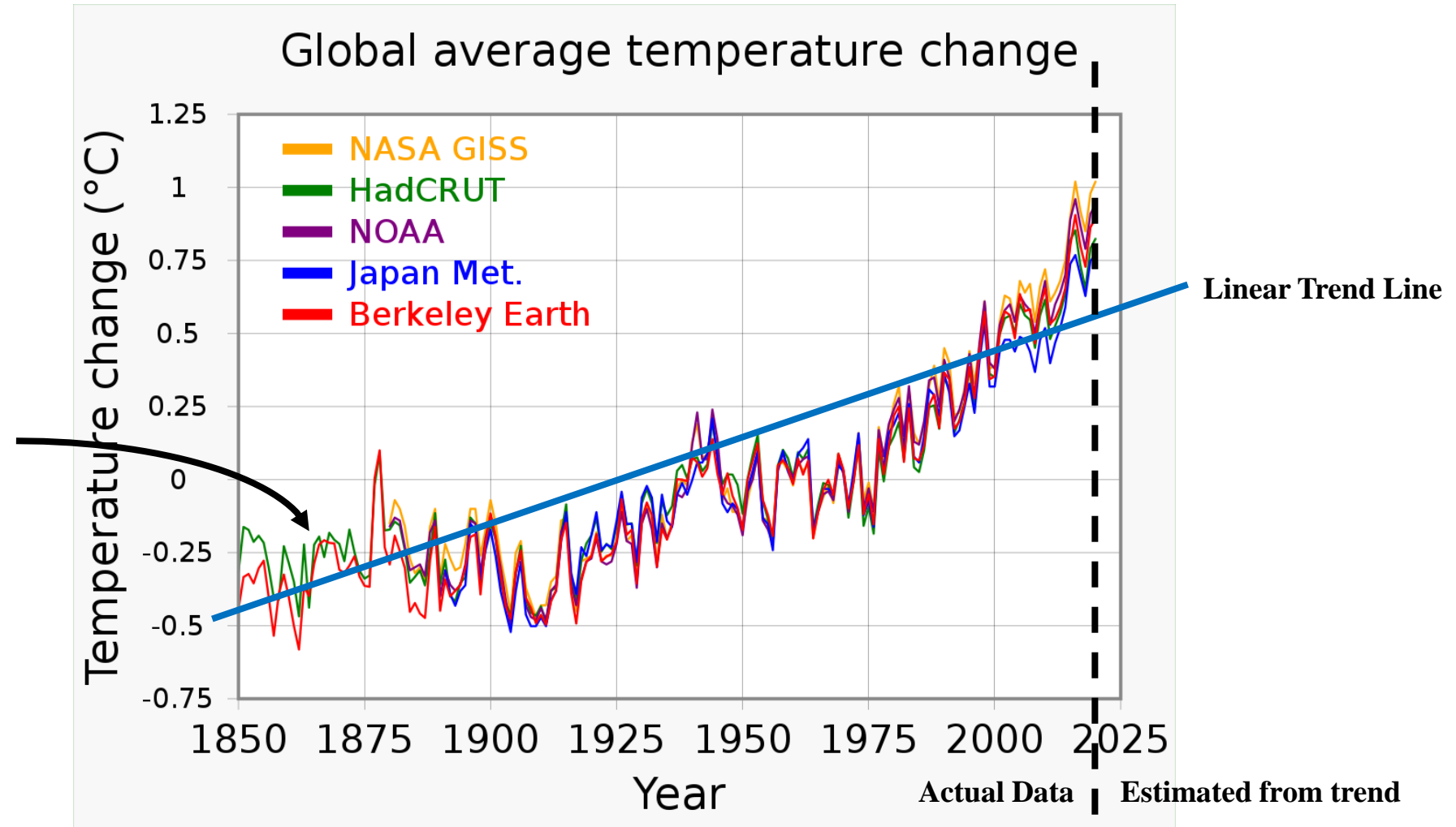


3. Data Strategies

For Quantitative Data

Trend Analysis

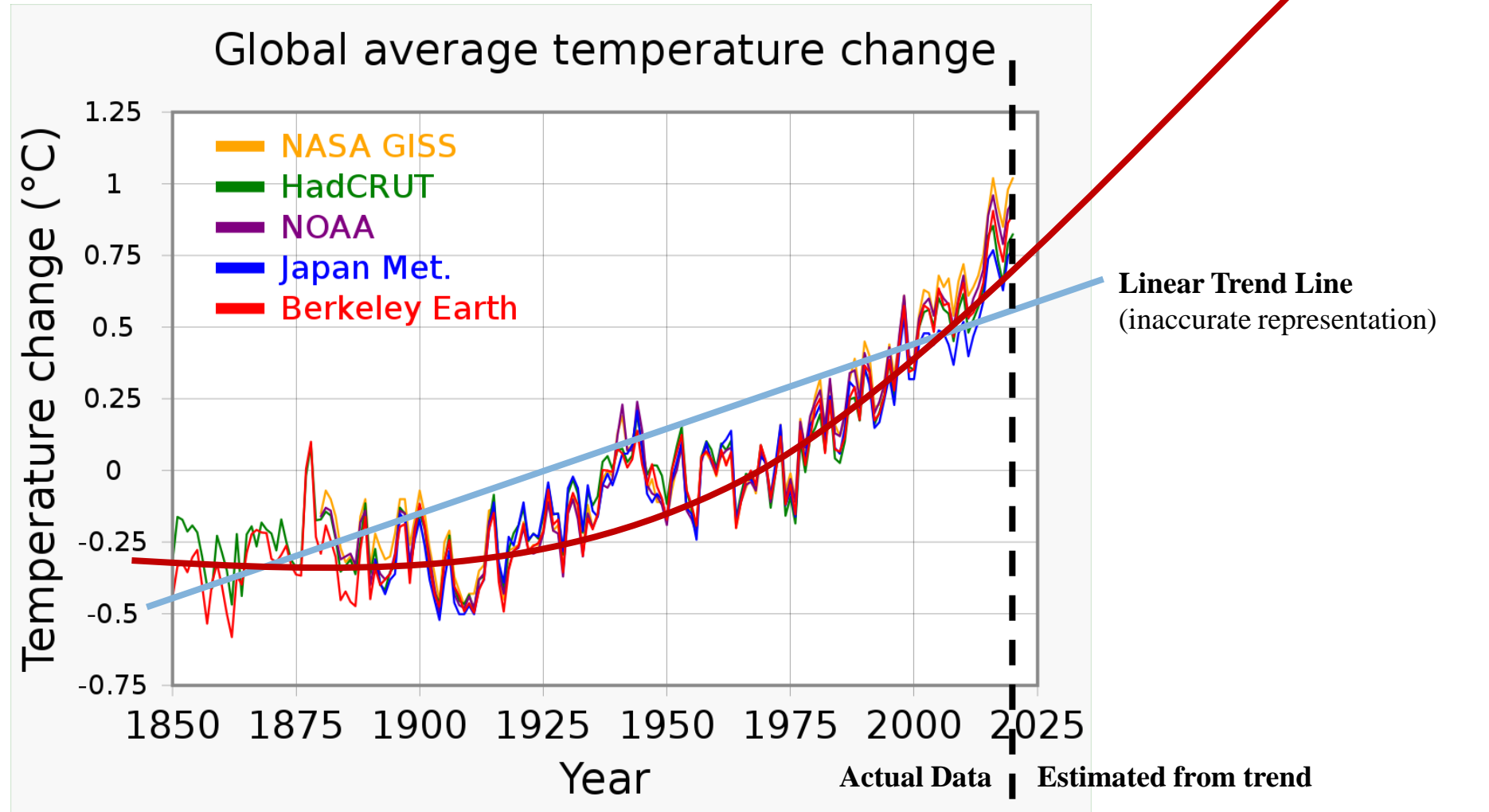
Aka. Regression Line



3. Data Strategies

For Quantitative Data

Trend Analysis



3. Data Strategies

For Quantitative Data

Those randomness are usually described by theories: *Brownian Motion* and *Stochastic Process*.

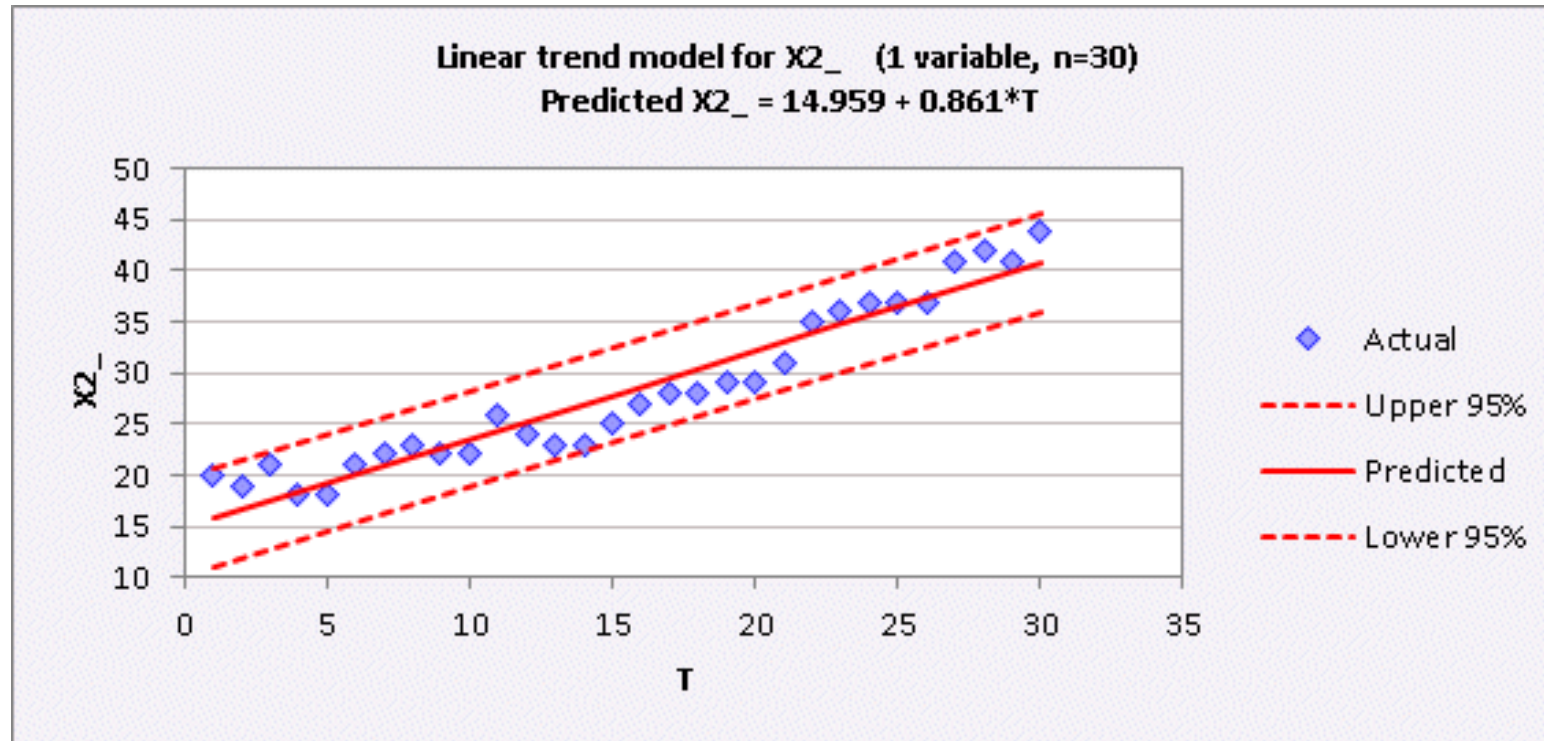
Trend Analysis



3. Data Strategies

For Quantitative Data

Estimation



Duke People—Variables (T,X2_) Prediction

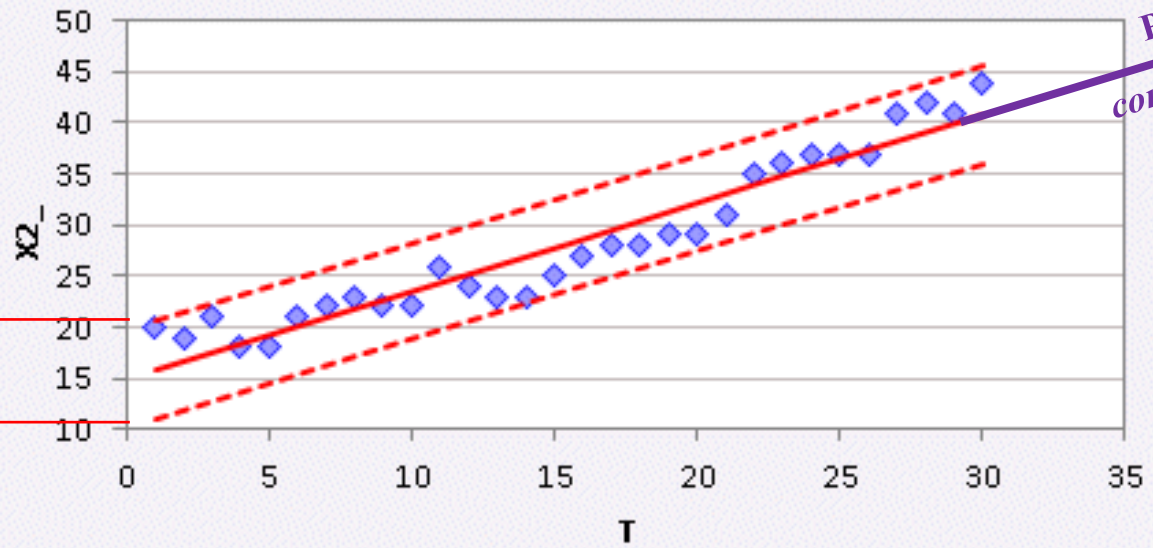
3. Data Strategies

For Quantitative Data

Estimation

Next-point Prediction
(by trend analysis
and estimation)

Linear trend model for X2_ (1 variable, n=30)
Predicted X2_ = 14.959 + 0.861*T



Maximum and Minimum Line
(±Δy)

Duke People—Variables ($x=T$, $y=X2_$) Prediction

4. Data Framework

Linear Framework

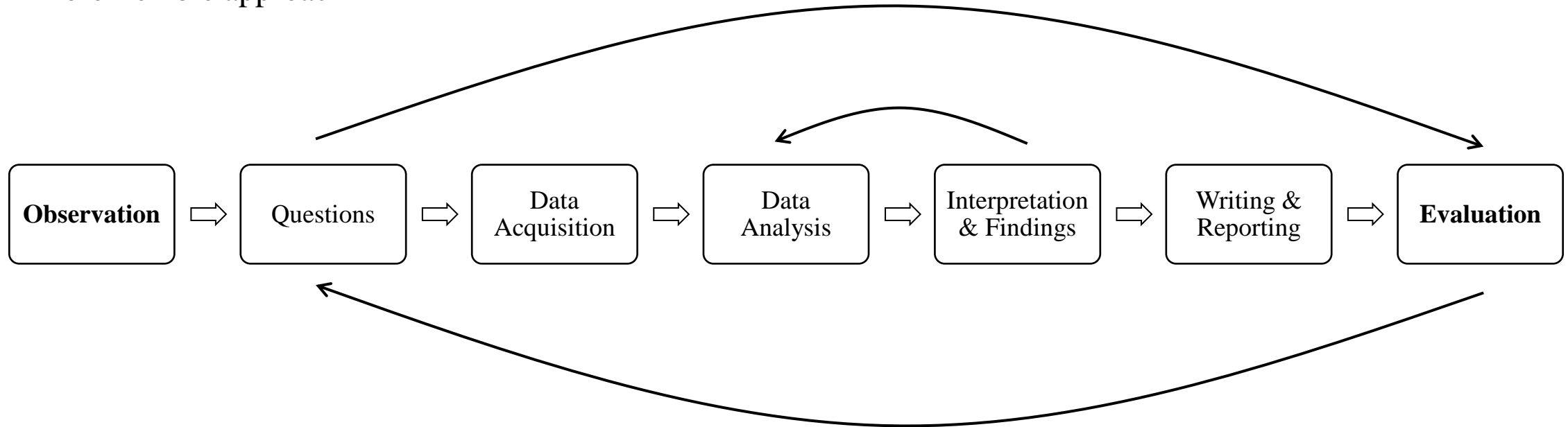
Strict approach, ideal but not flexible



4. Data Framework

Cyclical Framework

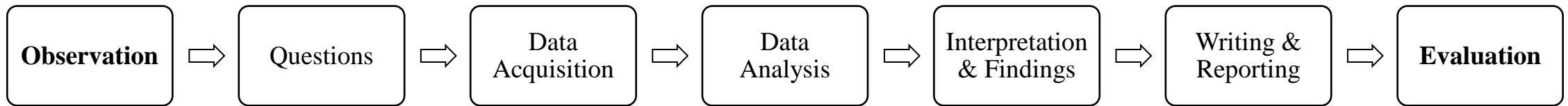
More flexible approach



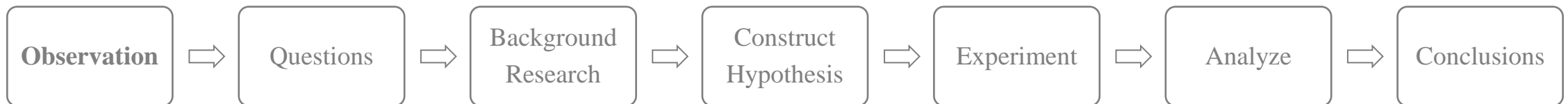
4. Data Framework

Similarity of Methodologies

Data Framework กระบวนการทางวิทยาการข้อมูล



Scientific Method กระบวนการทางวิทยาศาสตร์



5. Data Acquisition

Research

Finding Data from Existing Results and Raw Data published by authors

Reliability of Data

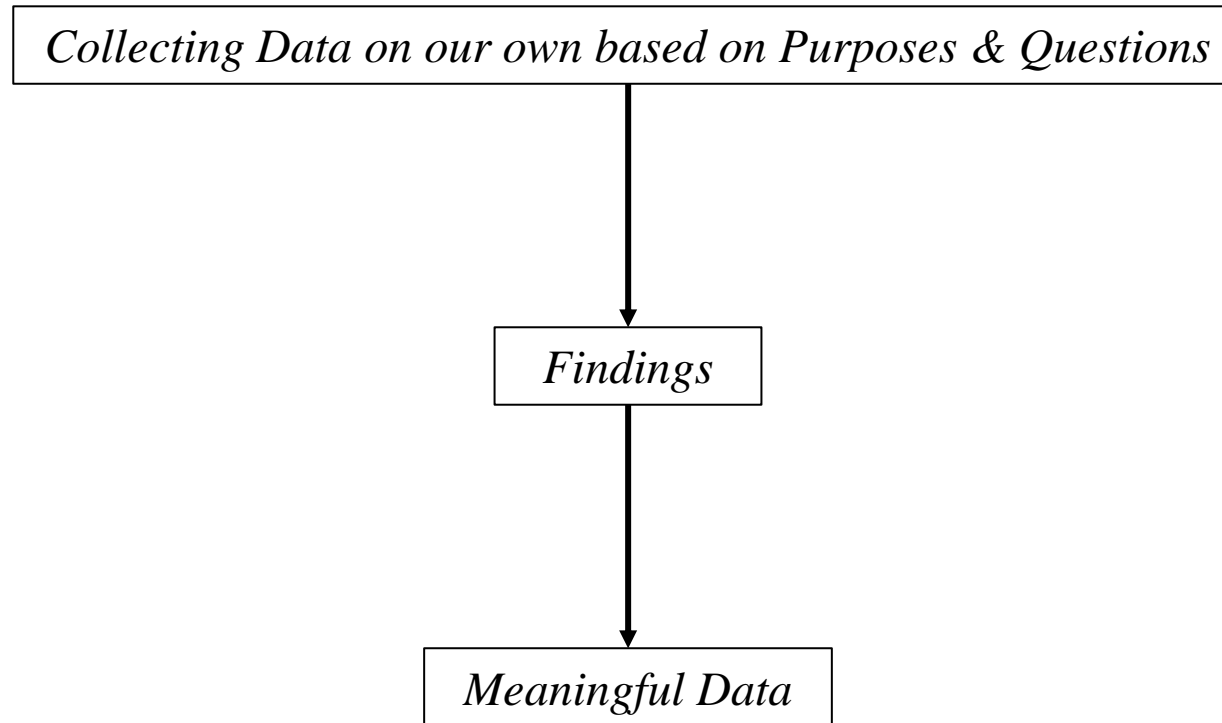
- Number of Samples
- Academic Index Ranking of the Source
- Citations
- Error/Uncertainty/Deviation Evaluation
- Significant Figures of Data

What might be extracted?

- Results of source
- Some of needed data
- Formula used
- Figures and Tables

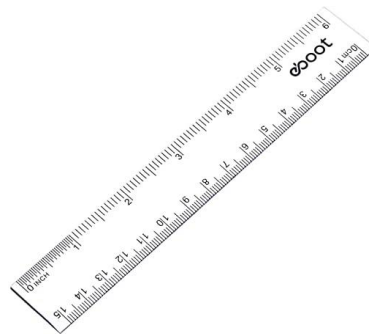
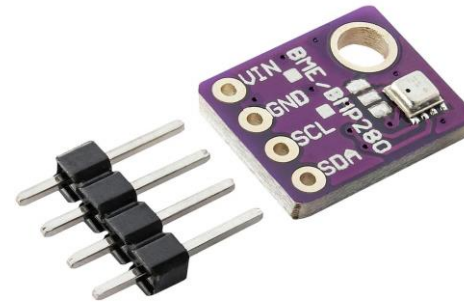
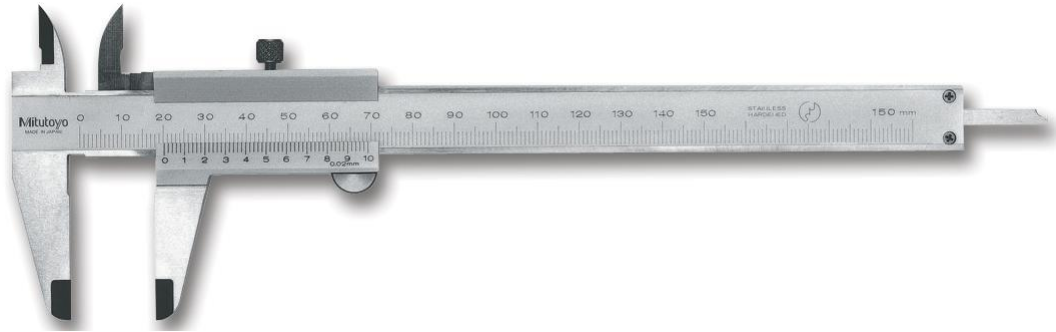
5. Data Acquisition

Survey and Exploration



5. Data Acquisition

Measurement



5. Data Acquisition

Measurement

