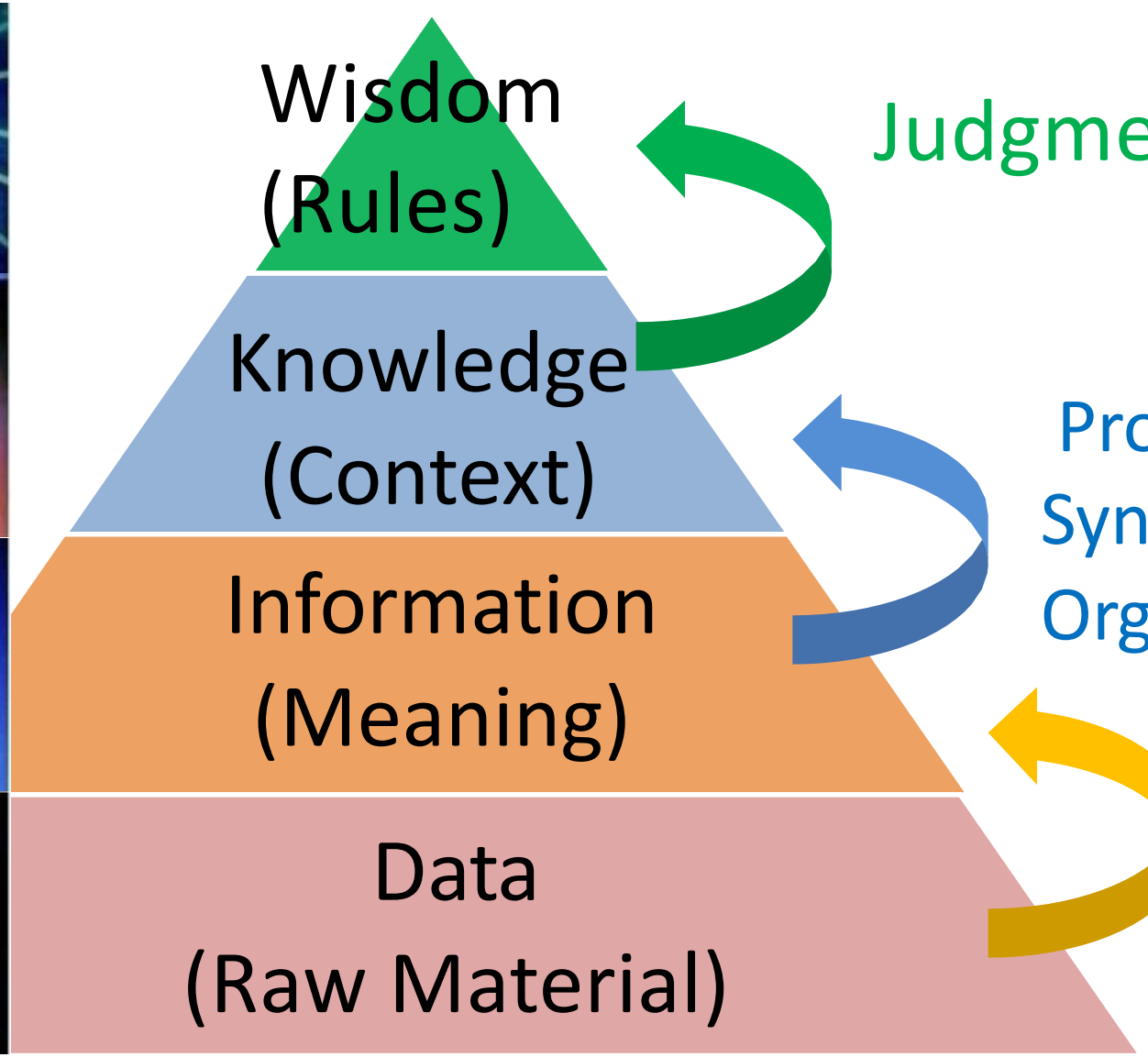
Data-driven Public Health
(การขับเคลื่อนงานด้านสาธารณสุขด้วยข้อมูล)

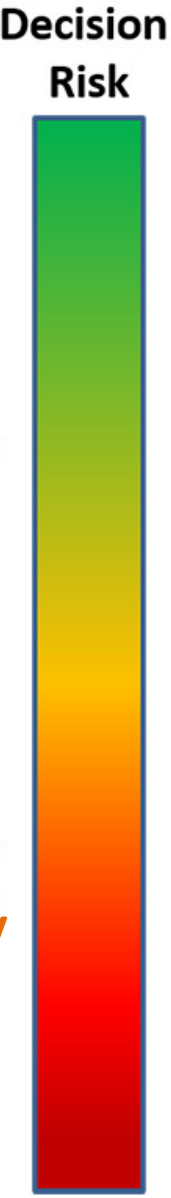
Suphanat Wongsanuphat, MD, MS (Health Informatics)
Division of Epidemiology, Department of Disease Control, Thai Ministry of Public Health

Outline of This Talk

- Brief Principle of Data and Data Science (DS)
- Big Data in Public Health
- Data Science Application to drive Public Health
  - Predictive Modeling
  - AI and Machine Learning
  - Data Engineering
- Challenges and Risks

What is Data?


Wisdom
(Rules)
Judgment
Knowledge
(Context)
Processing/
Synthesis/
Organization
Information
(Meaning)
Contextualization/
Interpretation
Data
(Raw Material)


Decision
Risk

What is Data?

Wisdom

Knowledge

Information

Data

Applied: DDC should investigate this outbreak.

Context: Possibly nosocomial COVID-19 outbreak

Meaning: 38 URI patient from IPD unit in 1 hos

Raw: 38

Data Science

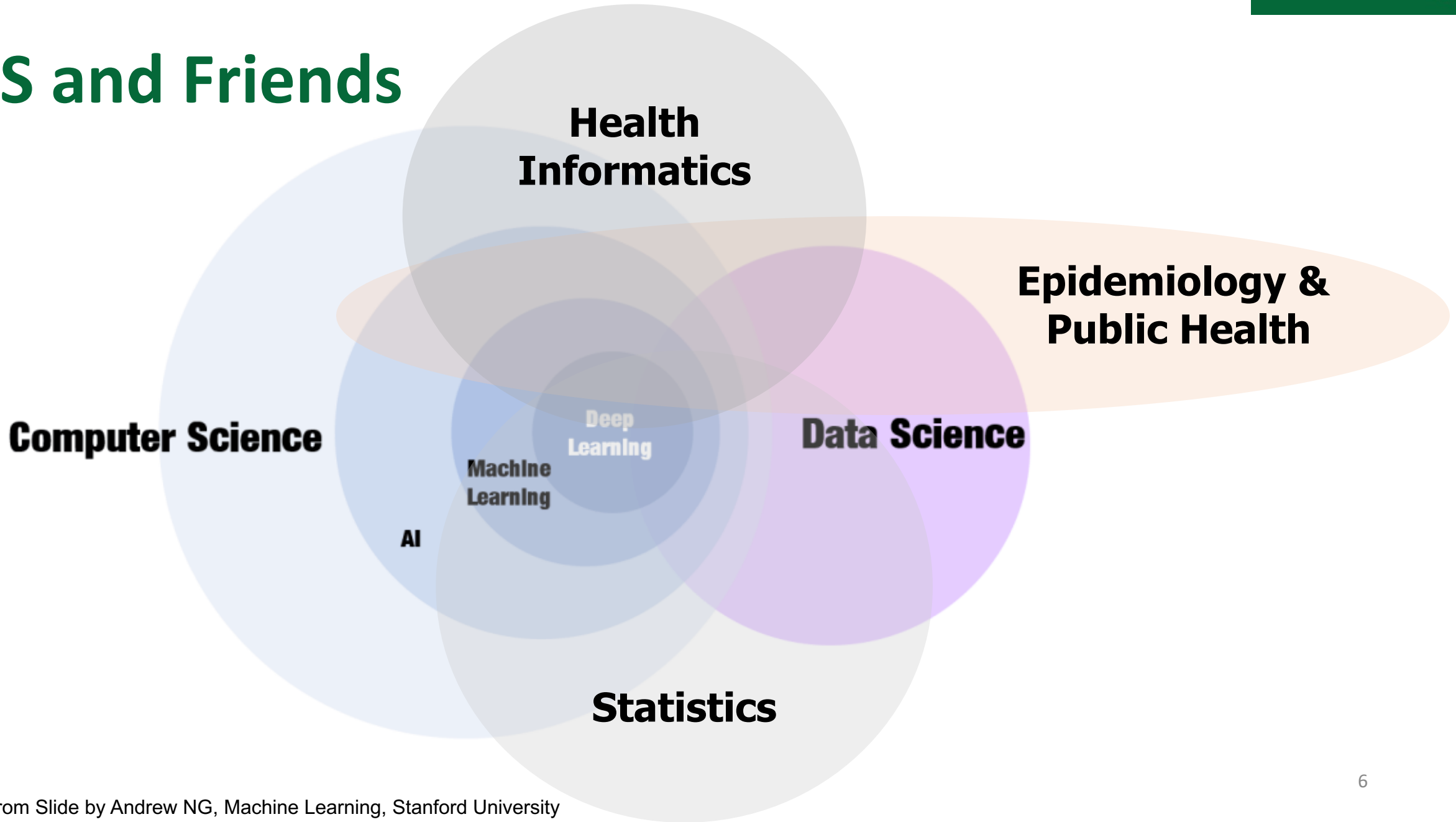
What is Data Science?

"Data science is an inter-disciplinary field that uses ***scientific methods, processes, algorithms and systems*** to ***extract knowledge and insights*** from many structural and unstructured data.[1,2]

"Data science is related to **big data**, **data mining**, **machine learning**, and **deep learning**."

1. *Dhar, V. (2013). "Data science and prediction". Communications of the ACM. **56**(12): 64–73. doi:10.1145/2500499. Archived from the original on 9 November 2014. Retrieved 2 September 2015.*
2. Jeff Leek (12 December 2013). "The key word in "Data Science" is not Data, it is Science". Simply Statistics. Archived from the original on 2 January 2014. Retrieved 1 January 2014.

DS and Friends

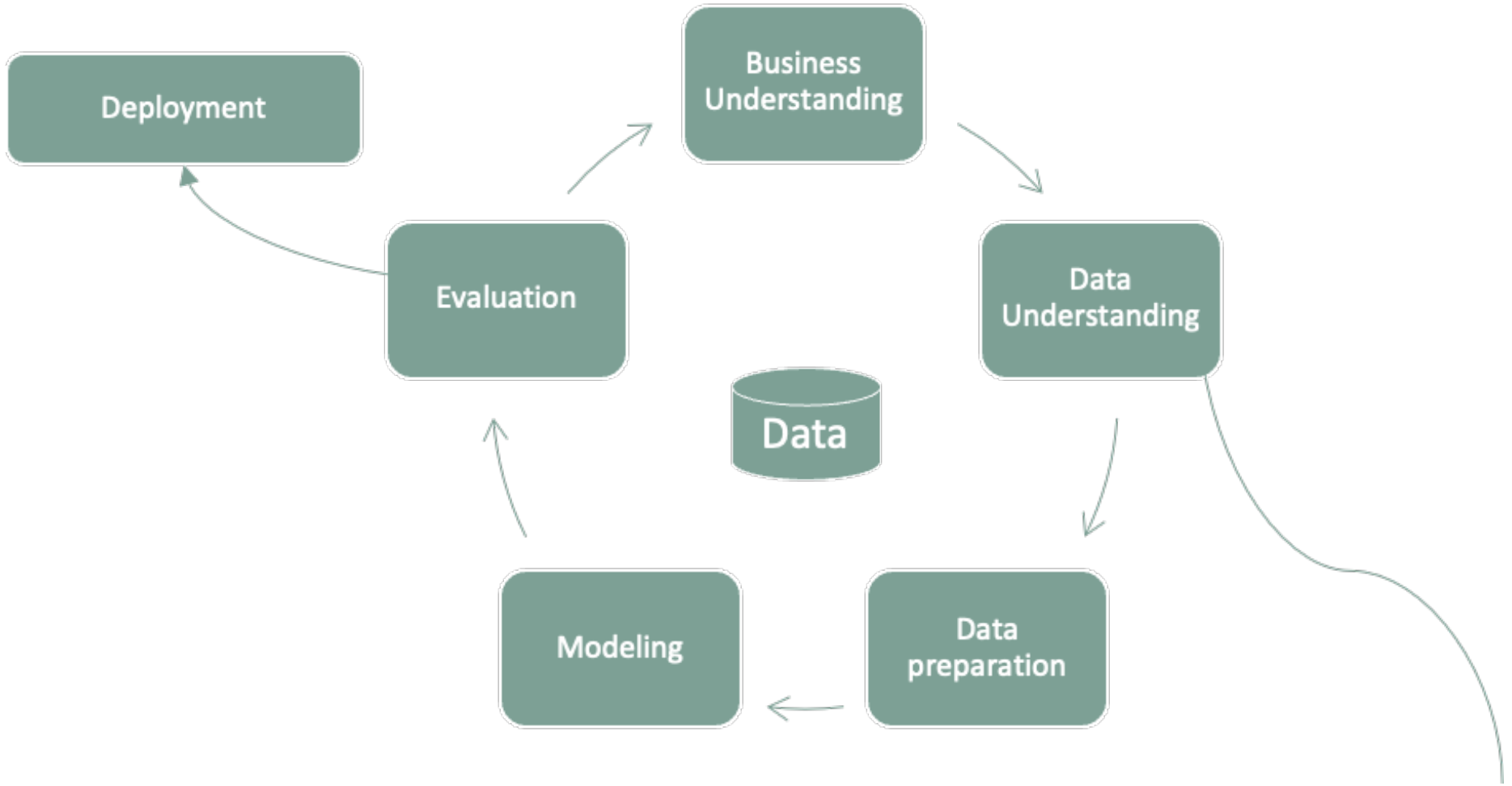


Adapted from Slide by Andrew NG, Machine Learning, Stanford University

Data Science Cycle

Start Here

What is your business?
What are your tasks/jobs?
What are your questions?

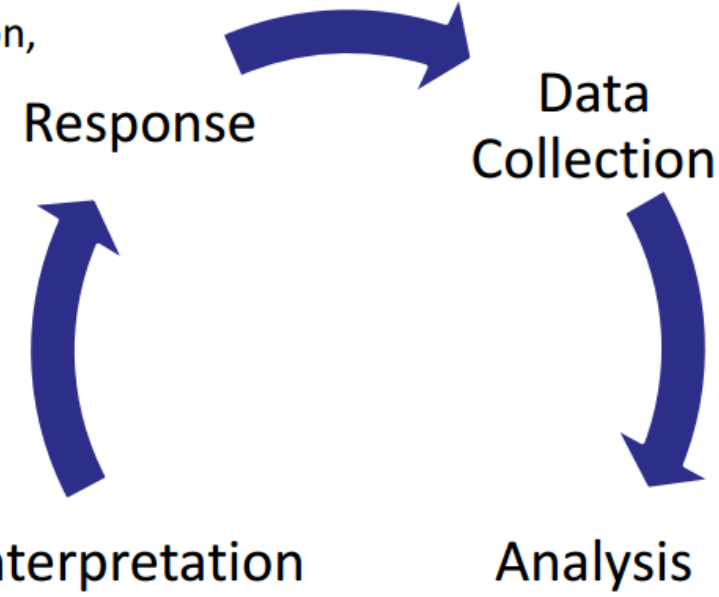


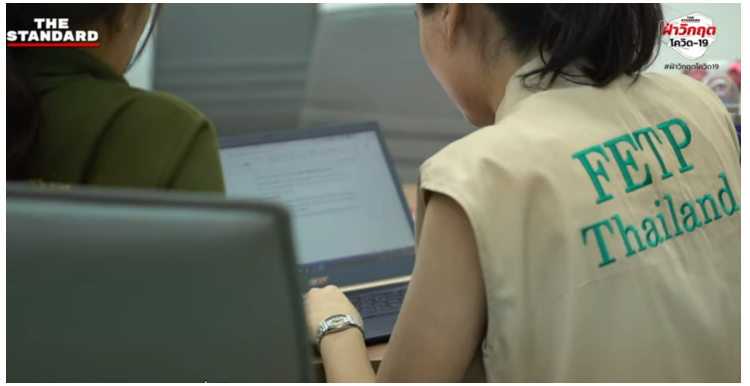
https://mustafaserdarkonca.medium.com/data-science-life-cycle-e4d74afe4bf5

Surveillance Cycle

- Dissemination: reports, alerts, publications
- Action: investigation, policy, research

- Case definitions
- Sampling frame
- Collection tools
- Data transfer



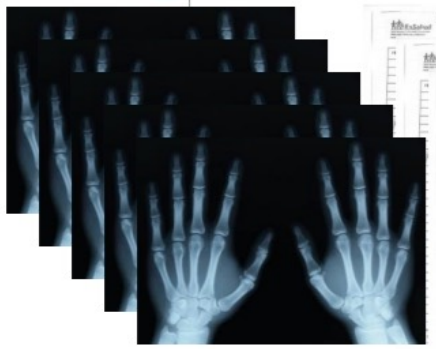


- Trends/comparisons
- Hypotheses
- Aberration detection

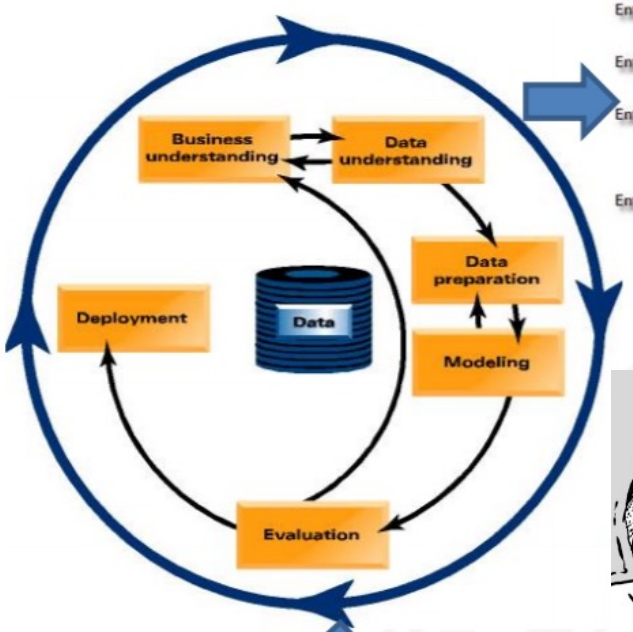
- Person, place, time
- Counts/rates
- Time series

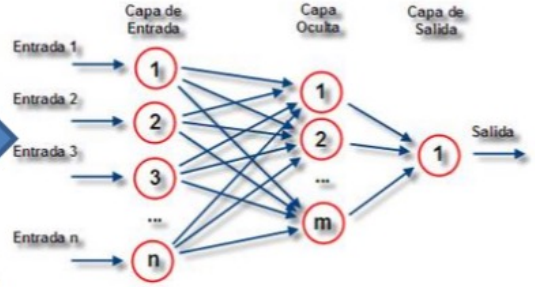
1. Implementation of Early Warning and Response with a focus on Event-Based Surveillance Interim Version. 2014 [cited 2018 Nov 30]; Available from: www.who.int/about/licensing/copyright_form/en/index.html
2. https://www.youtube.com/watch?v=Zy6Yjd7N2Xo

DS Framework



Data Lake





Process → Integrate → Feature extraction → Data Warehouse

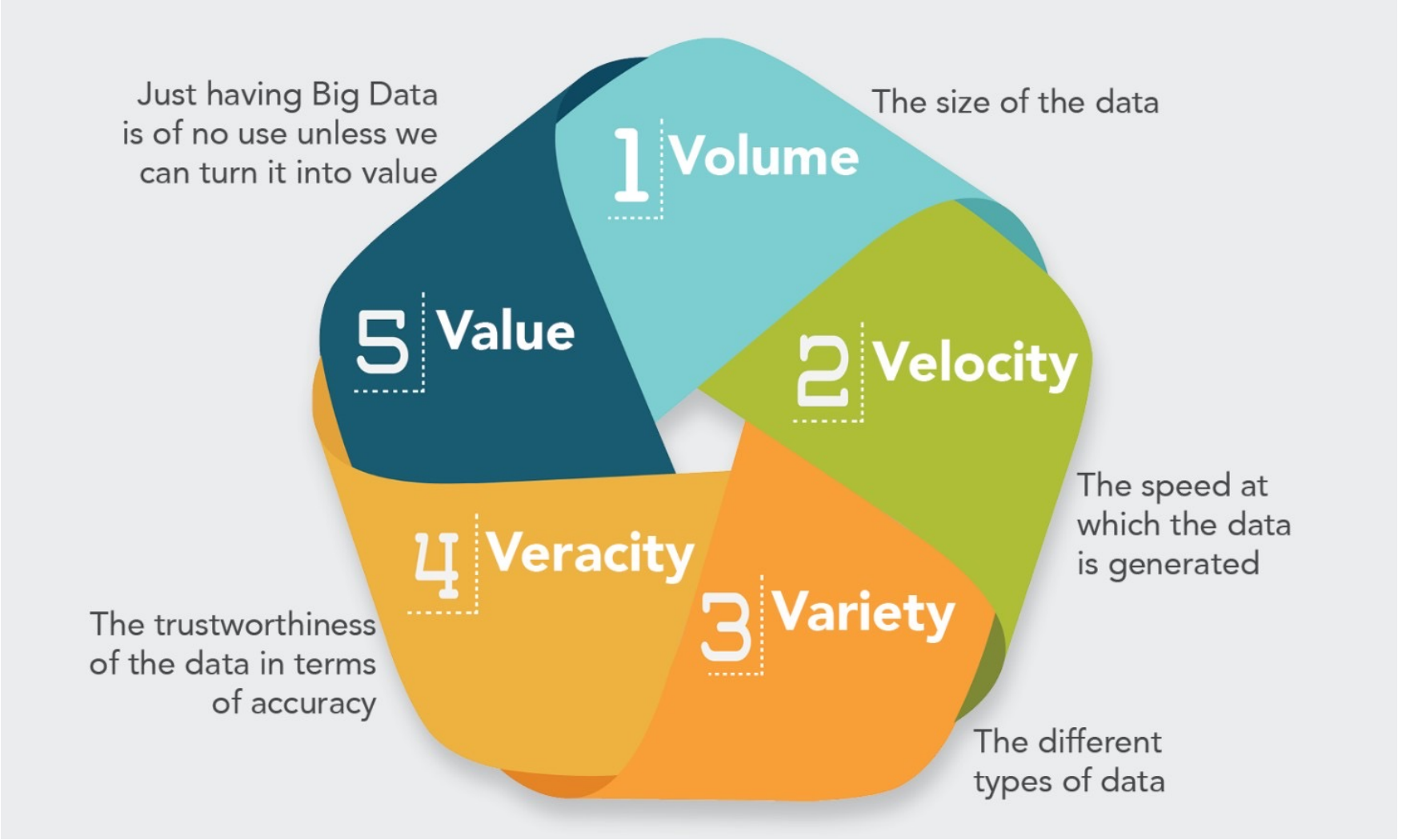
Big Data in Healthcare: motivation, current state and specific use cases, Alejandro Rodríguez González Centro de Tecnología Biomédica Universidad Politécnica de Madrid

Big Data in Public Health

"**Big data** is data sets that are ***too large or complex to be dealt with by traditional data-processing*** application software."

https://medium.com/@amritabaranwal3011/top-eight-big-data-tools-d0f59c885ec9

The 5Vs of Big Data



Do you consider Public Health Data as Big Data?

5Vs: Volume

Total: 11,289

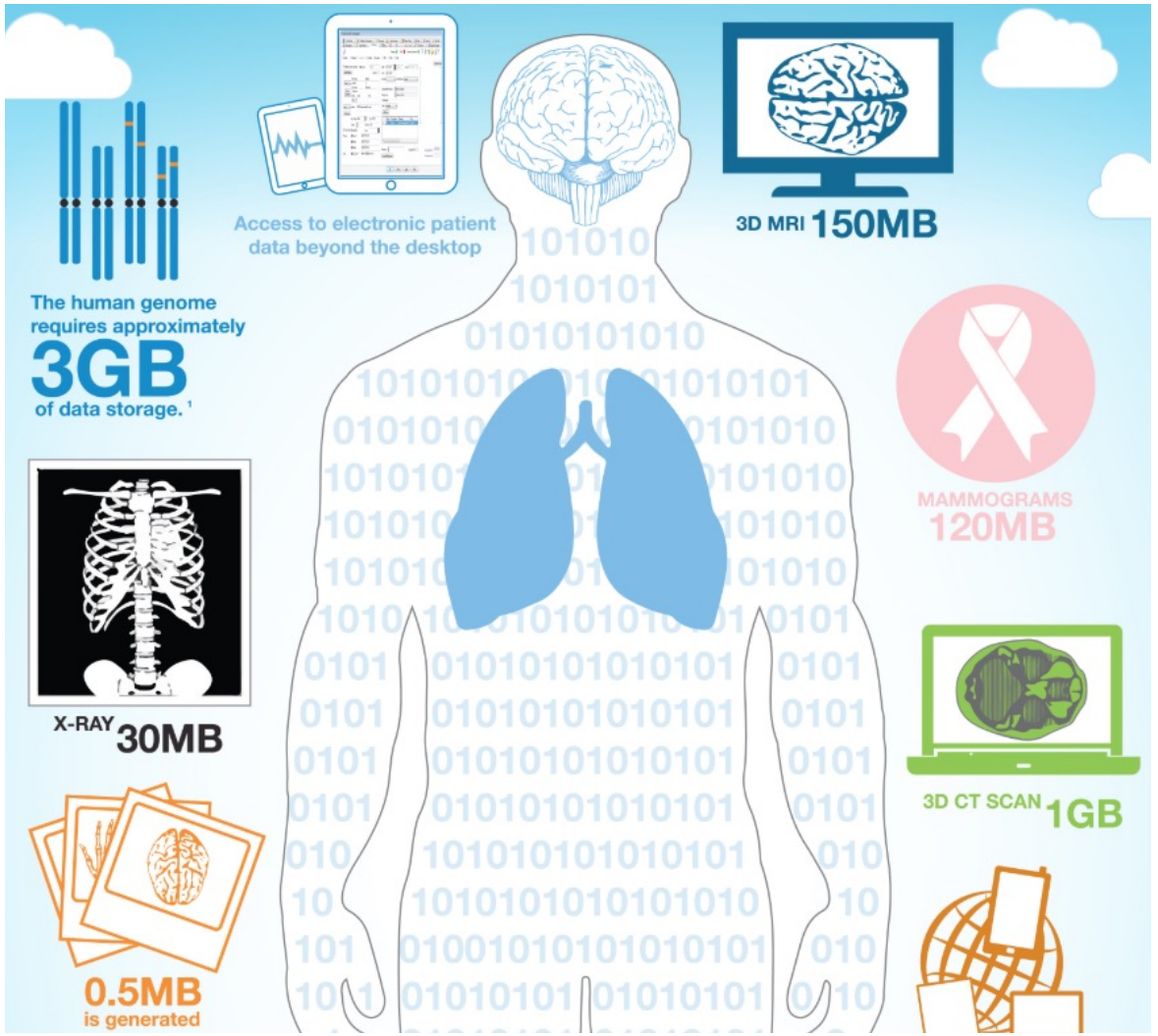
- Village: 9,899
- Public: 1,043
- Private: 347

>200 million

Patient visits / Year

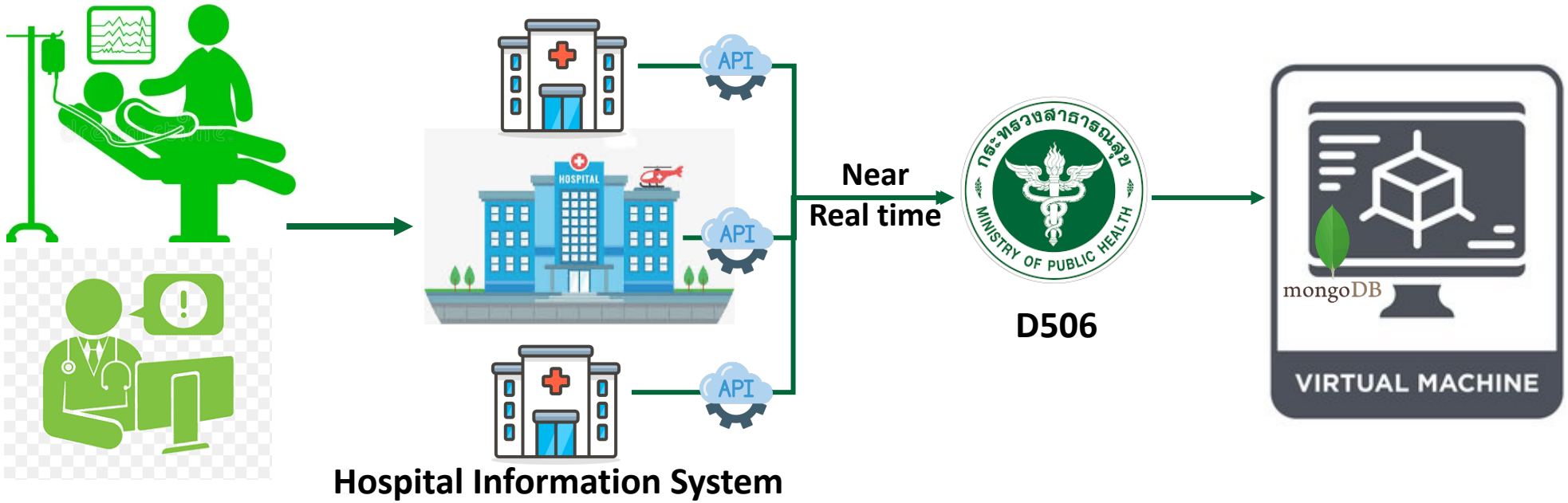
>300 million

Records

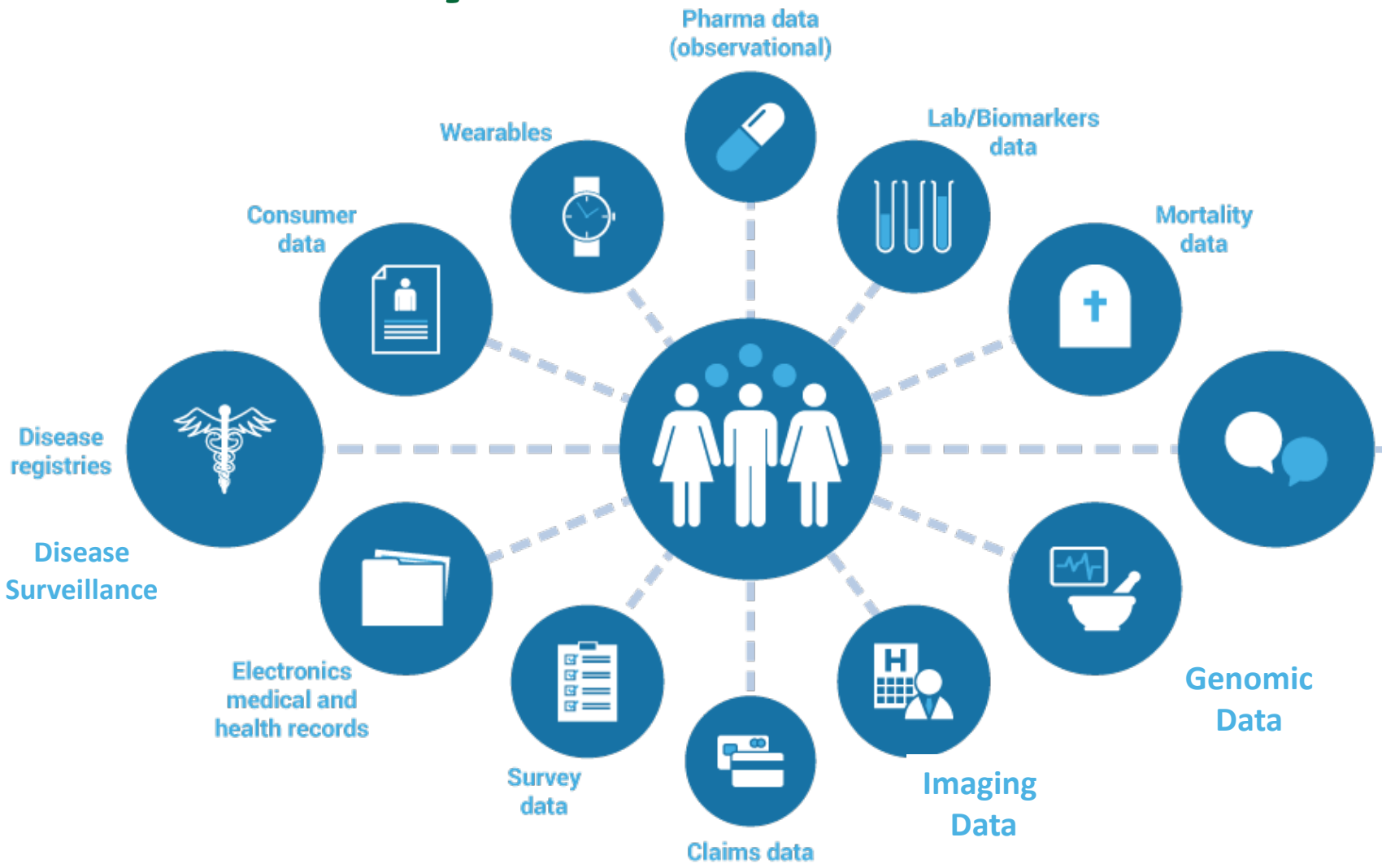


http://medcitynews.com/2013/03/the-body-in-bytes-medical-images-as-a-source-of-healthcare-big-data-infographic/

5Vs: Velocity (Speed)



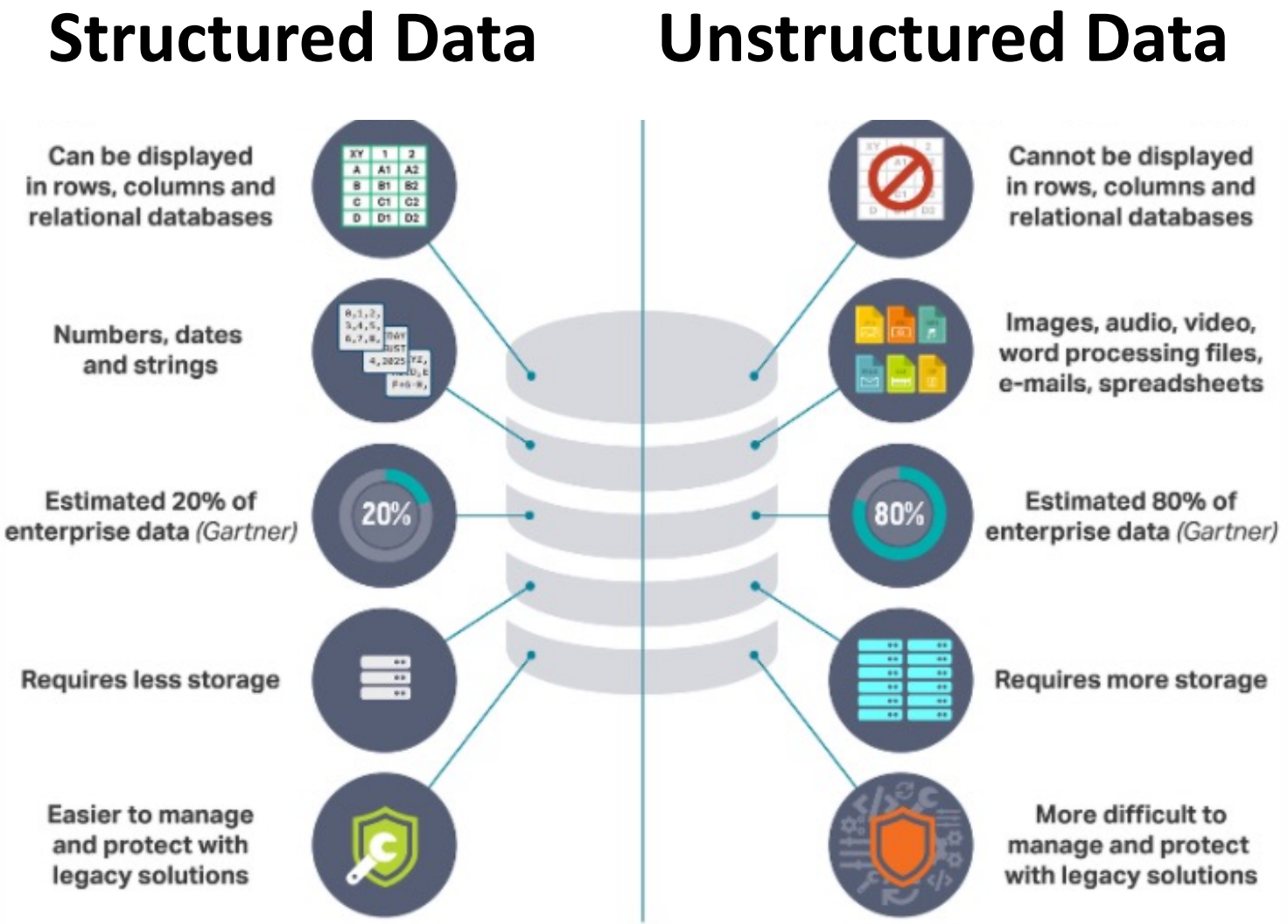
5Vs: Variety



- Animal Data
- Vector Data
- Environmental Data
- School/Work Absenteeism
- Local News
- Online Search
- Social Media Data
- Transportation
- Waste and Water
- Financial Data
- Police Record

https://myassignmenthelpoz.com/blog/sit718-real-world-analytics-assignment-solution-task-3/

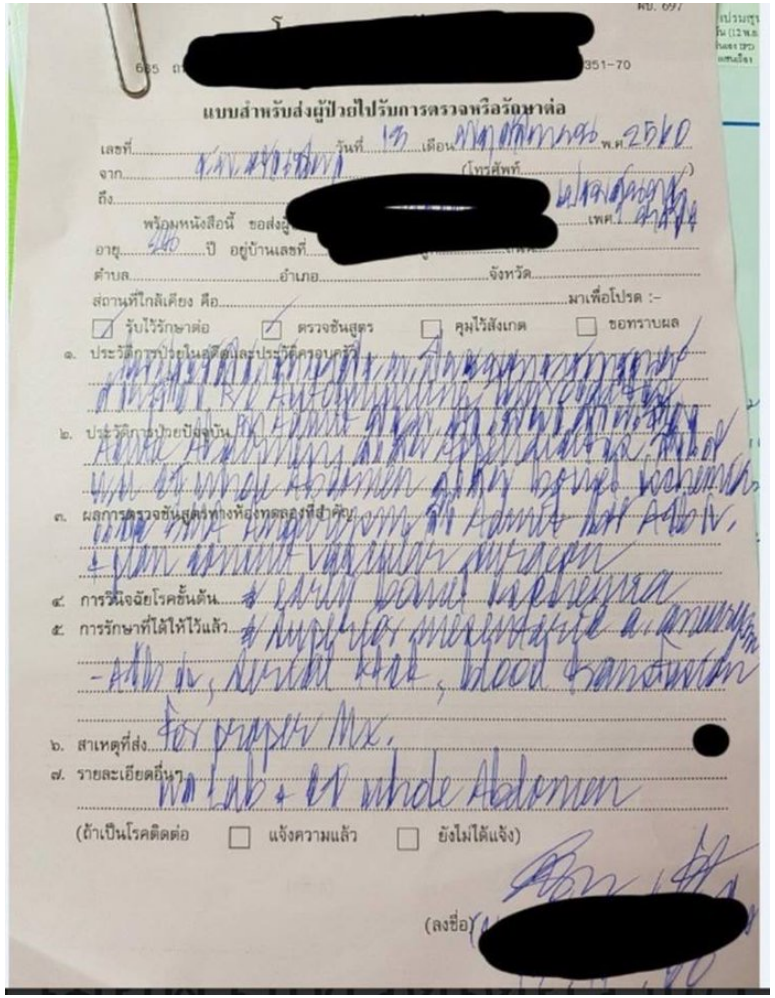
5Vs: Variety



- **Structured data**
  - Database
  - Data Table
- **Unstructured data**
  - Free Text
  - Image(s)
  - Video
  - Audio / Voice / Speech
  - Motion
  - Signal
  - Rhythm
  - Geolocation (semi-str.)

https://lawtomated.com/structured-data-vs-unstructured-data-what-are-they-and-why-care/

Unstructured Data



OCR*

Pt's SOB and DOE are ↓ed. AF w/ VSS. CXR: LLL ASD s̄ Δ. WBC 11K; S/B Cx → GPC c/w PC w/o GNR, will d/c cef. → PCN

This translates into:

The patient's shortness of breath and dyspnea on exertion are diminished. He is afebrile, with stable vital signs. A repeat chest x-ray continues to reveal left lower lobe air space disease without any change from his admission film. His white blood cell count is now 11,000. Because his sputum and blood cultures demonstrated gram-positive cocci (consistent with pneumococcus) without gram-negative rods, we will discontinue his cefuroxime therapy and institute treatment with penicillin instead.[94]

***OCR** = Optical Character Recognition

Fine PL. The Wards: An Introduction to Clinical Clerkships. Boston, Mass: Little Brown, 1994

Unstructured Data

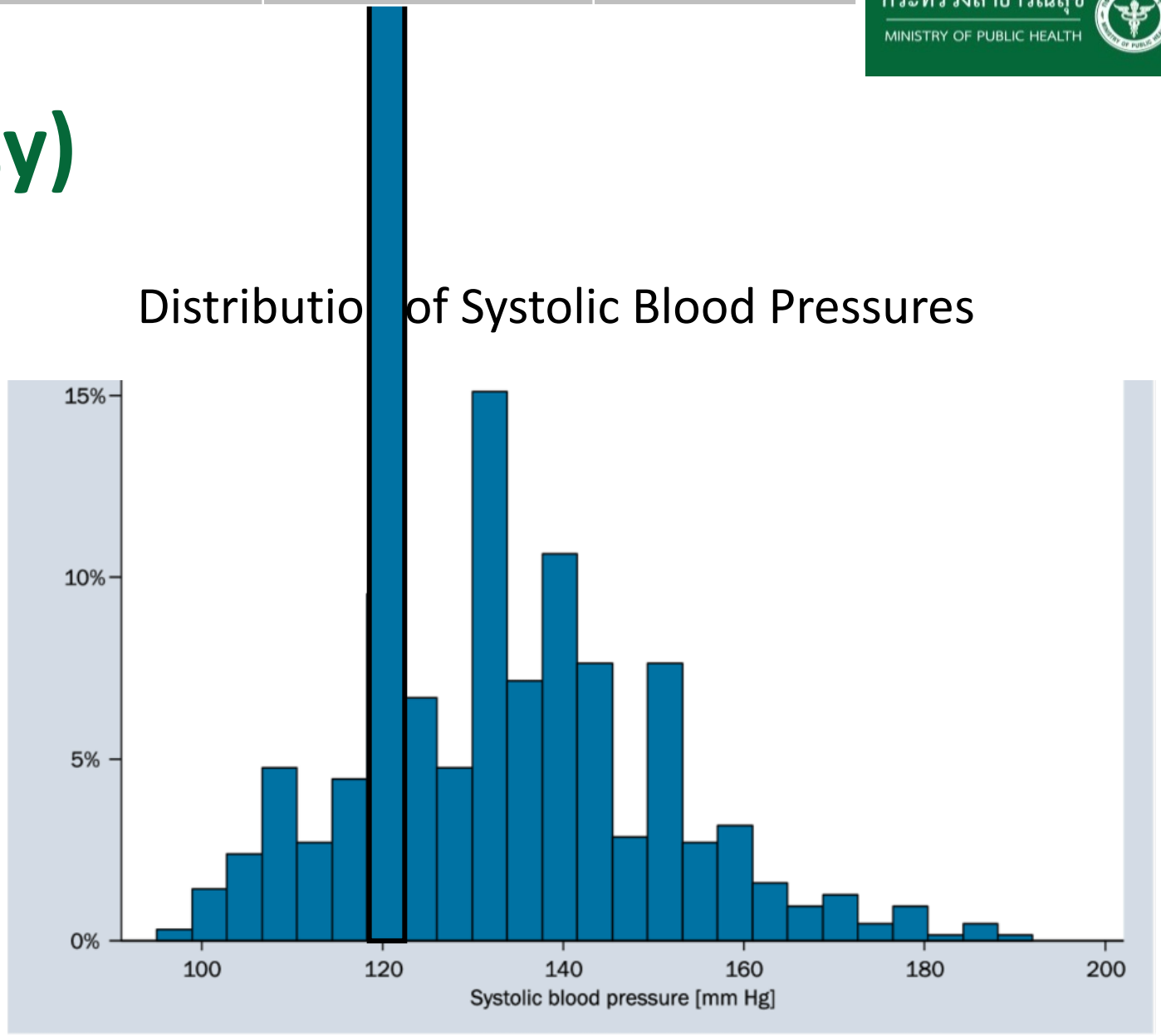
Grayscale Images

- Pixels → number grid
- Pixel brightness → grid value

```
[[  2  2  1 37  1 10 66 60 77 94 78 69 64 23 12 45 28 45 ]
 [ 58  1  9 13 17 29 56 72 65 64 59 58 39 18 15 12  7  1 ]
 [ 71 49 53 38 30 41 73 73 80 71 69 69 72 45 45 49 36 59 ]
 [ 88 60 73 50 59 59 54 51 71 81 69 50 54 75 56 61 80 67 ]
 [ 94 91 86 59 65 57 57 52 64 88 66 56 55 54 70 64 109 114 ]
 [ 94 95 84 74 70 41 48 55 74 85 84 60 50 46 70 82 92 122 ]
 [ 85 85 95 83 54 37 59 60 84 97 82 50 38 44 56 92 111 112 ]
 [ 81 87 94 92 54 54 56 54 79 96 79 48 36 44 62 103 107 145 ]
 [ 67 83 91 87 60 59 61 71 91 108 86 65 53 40 63 101 110 121 ]
 [ 49 73 88 72 66 73 78 84 107 120 102 71 57 39 56 89 114 103 ]
 [ 31 61 84 65 73 80 92 103 117 128 114 76 66 57 52 89 111 91 ]
 [  6 51 82 84 92 90 92 114 128 135 122 109 73 69 69 84 109 66 ]
 [  2 44 72 87 95 104 113 124 138 141 130 122 96 77 68 76 104 10 ]
 [  0 37 74 84 102 113 115 131 146 146 133 124 113 94 83 96 90 1 ]
 [  0 33 67 90 113 126 130 140 148 147 136 130 117 95 91 81 71 1 ]
 [  0 33 68 98 122 139 141 144 153 149 135 127 122 108 96 76 65 1 ]
 [  0 36 81 105 127 144 151 151 155 149 125 114 113 121 105 76 49 1 ]
 [  0 39 90 114 131 151 155 157 161 153 122 96 102 107 110 66 50 1 ]]
```

AI in Public Health, Stanford (Coursera)

5Vs: Veracity (Accuracy)



กระทรวงสาธารณสุข
MINISTRY OF PUBLIC HEALTH

D506 Big Data

Outbreak Verification
43 Files,Lab, Vaccine

Digital 506

ArcGIS
VIRTUAL MACHINE

Analytical Dashboard

Routine Analysis

Ad-hoc Analysis/
Research

Data Sharing

Public Communication

Routine Report

Publication

Policy

ข้อมูลการสำรวจ

กรมอุตุนิยมวิทยา
อุณหภูมิ น้ำฝน ความชื้น

กรมควบคุมมลพิษ
มลภาวะทางสิ่งแวดล้อม

กรมปศุสัตว์
โรคในสัตว์

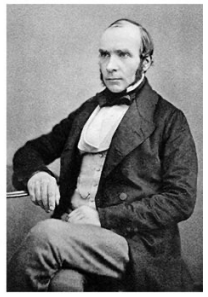
Data Science to drive Public Health Decision Making

กระทรวงสาธารณสุข
MINISTRY OF PUBLIC HEALTH

From intuition to machine learning

Intuition

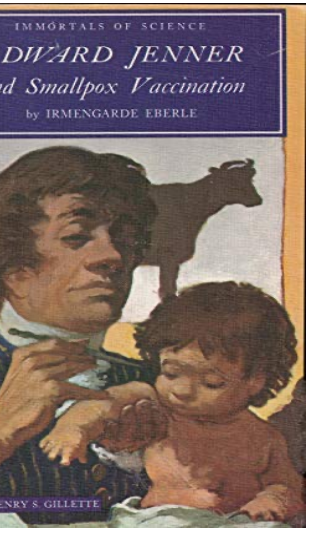
Using experience and judgement to predict outcomes



Statistical programming languages

Writing code to construct statistical models

1800s

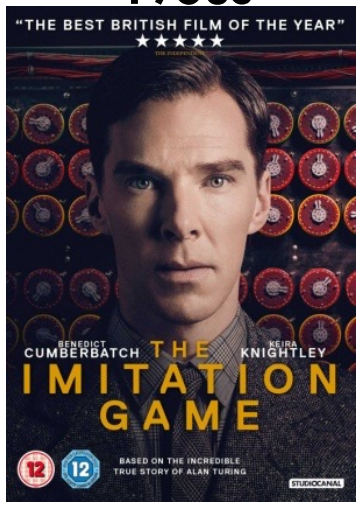


Manual analysis

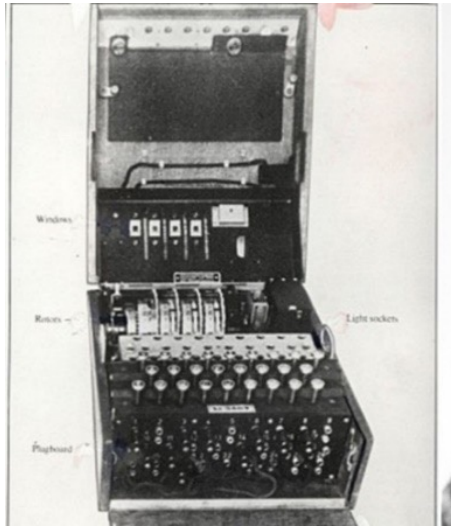
Manual calculations to predict outcomes

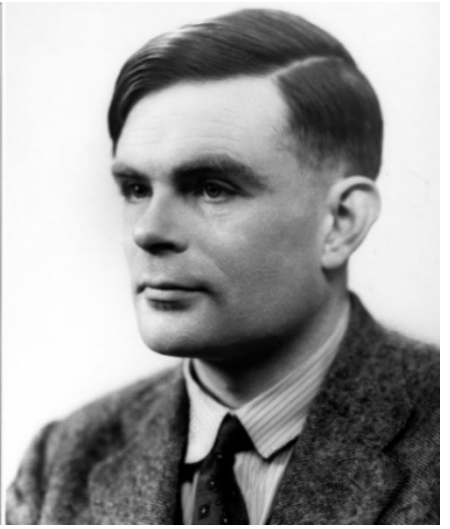
	Number of houses.	Deaths from Cholera.	Deaths in each 10,000 houses.
Southwark and Vauxhall Company	40,046	1,263	315
Lambeth Company . . .	26,107	98	37
Rest of London	256,423	1,422	59

1950s

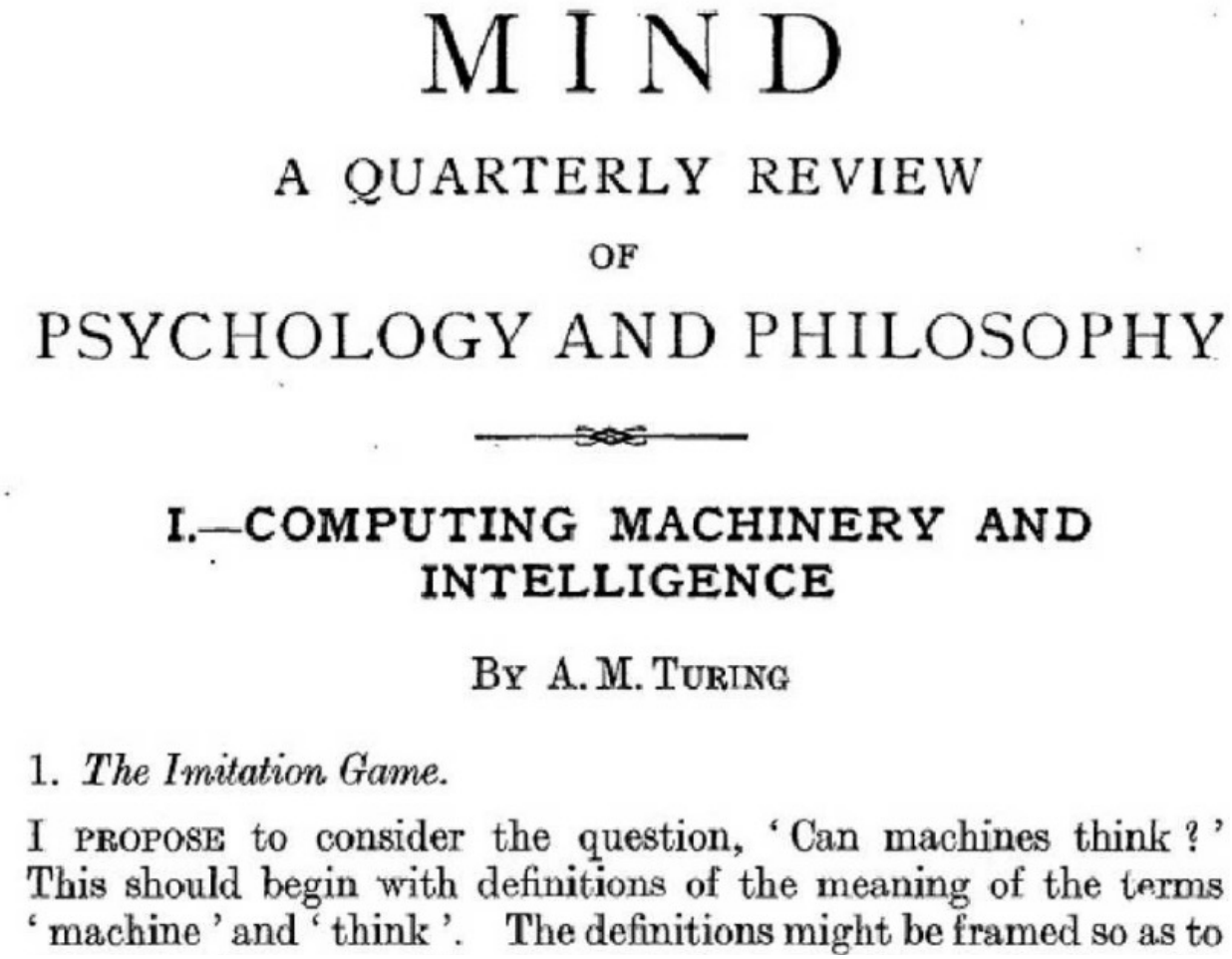


159 MMM patterns





Brief History of Data Analysis



MIND

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

By A. M. Turing

1. *The Imitation Game.*

I propose to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to

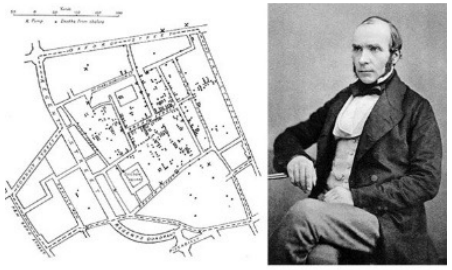
" I propose to consider the question **'Can machines think?'** "

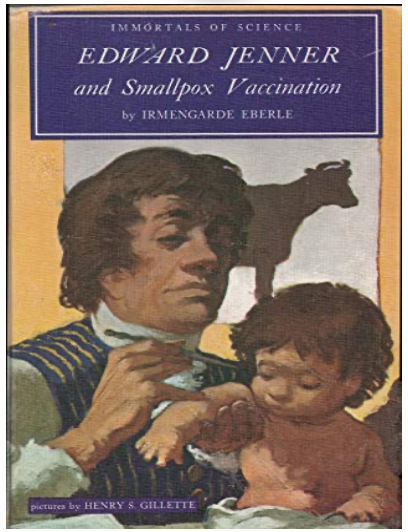
-Alan Turing-
-1950-

From intuition to machine learning

Intuition

Using experience and judgement to predict outcomes





Early 1900s

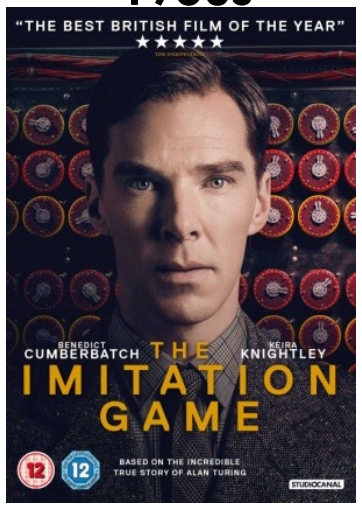
Manual analysis

Manual calculations to predict outcomes

Statistical programming languages

Writing code to construct statistical models

1950s



1990s

Visual statistical software

Drag and drop workflows with menu driven commands to set up and statistical analysis

STATA

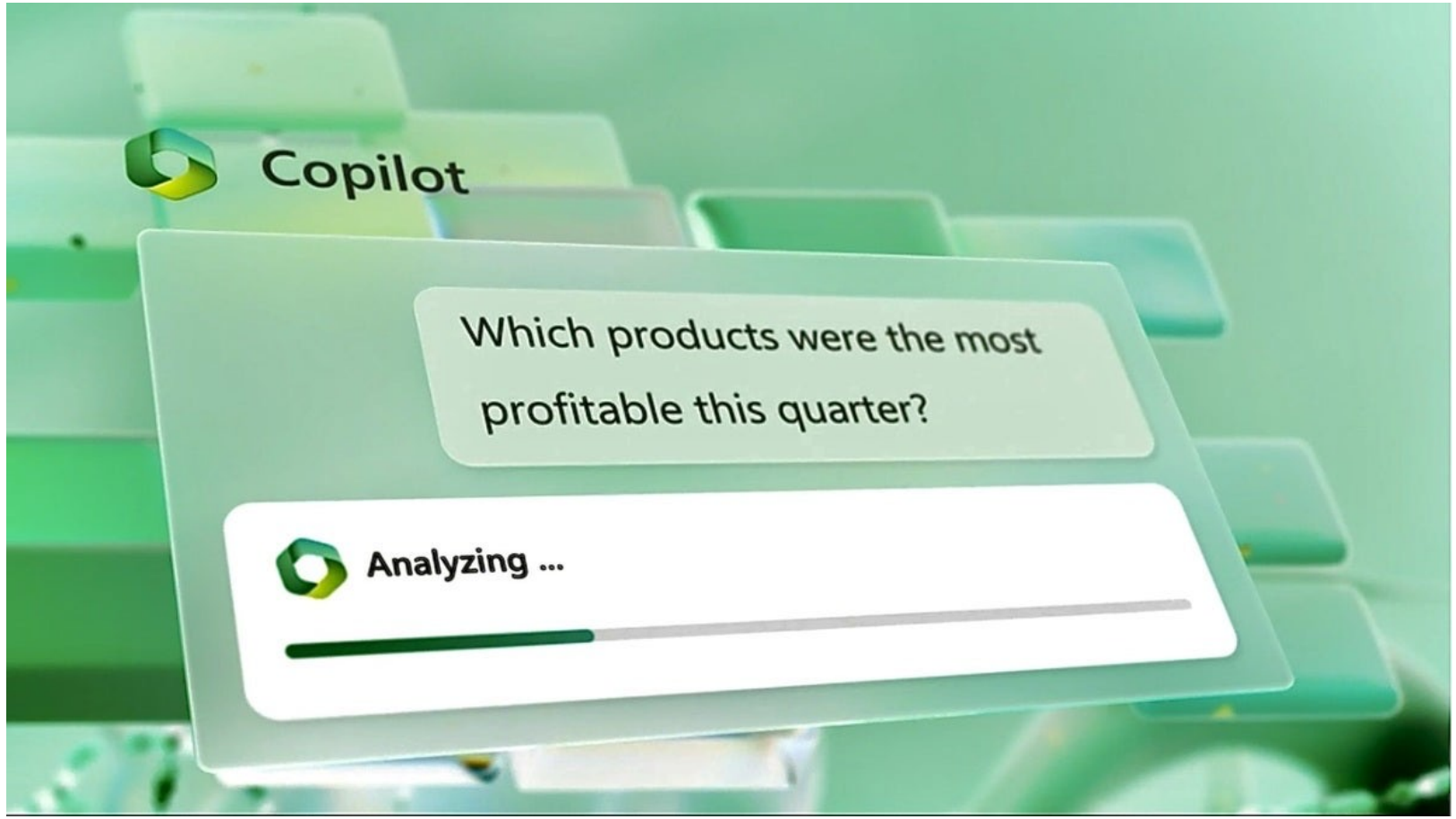
Now

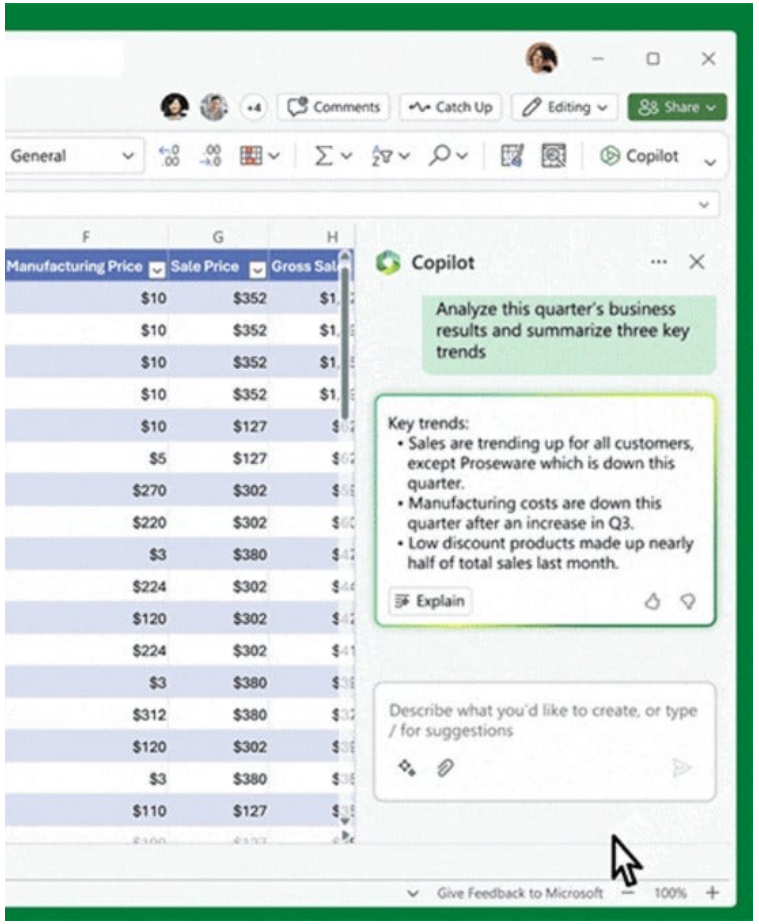
Automated machine learning

The software knows how to analyze your data and does it for you

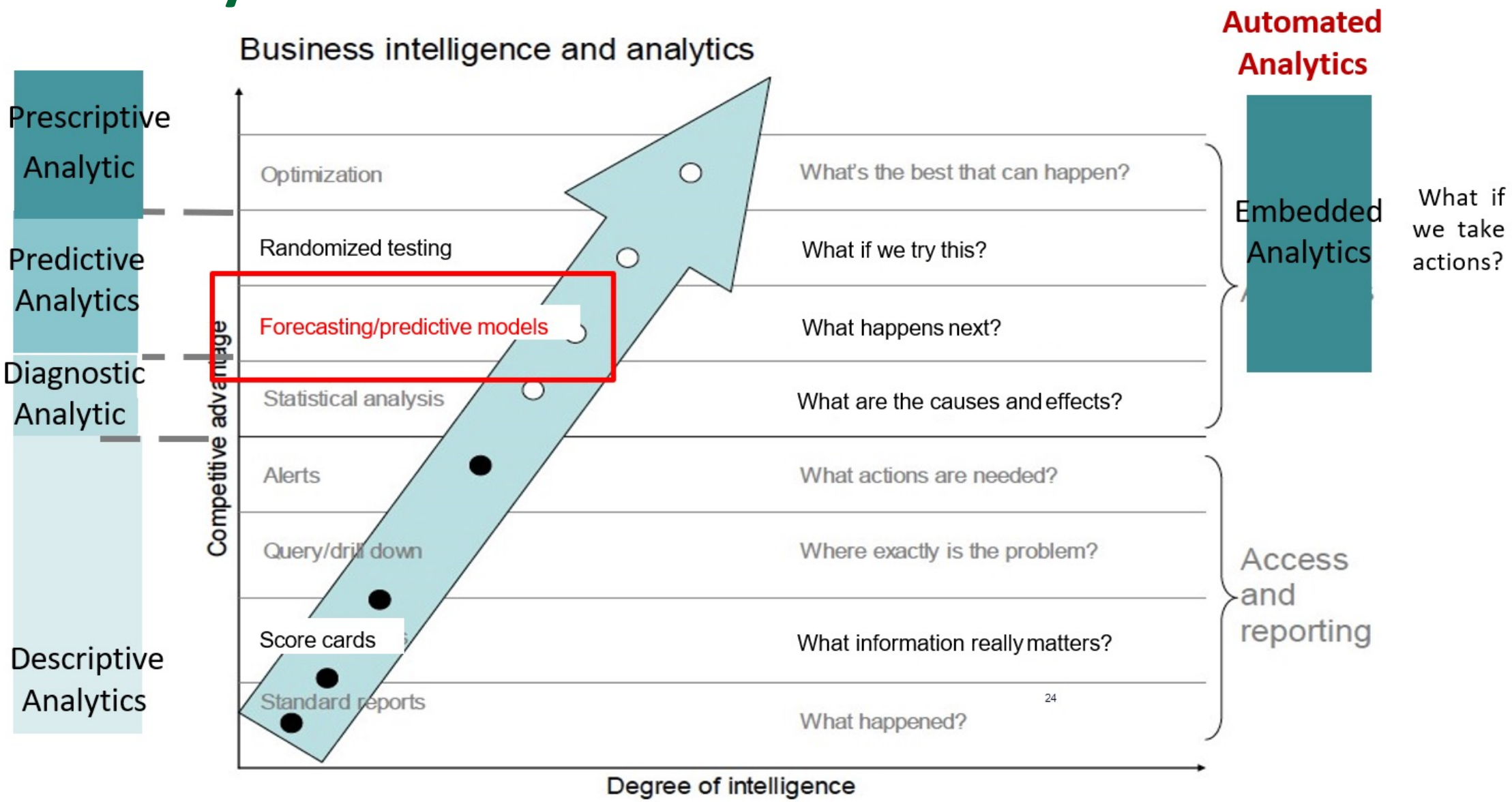
Copilot

Automated Machine Learning: 365 Copilot





Data Analytics



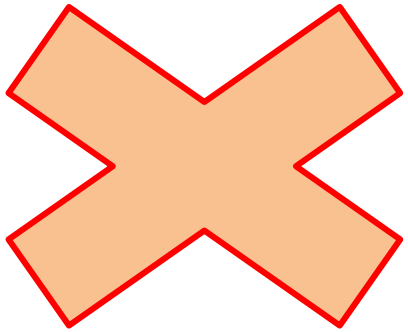
Source: Thomas H. Davenport (2016)

Common Data Analytics in Public Health

- Epidemiological Methods
- Statistical Methods
- Predictive Modeling
  - Time Series Forecasting
  - Mathematical Modeling
  - Decision Analysis and Cost Effectiveness Analysis
- Social Network Analysis
- GIS and Spatial Data Analysis
- AI and Machine Learning

This Talk

How do We Make Decisions?



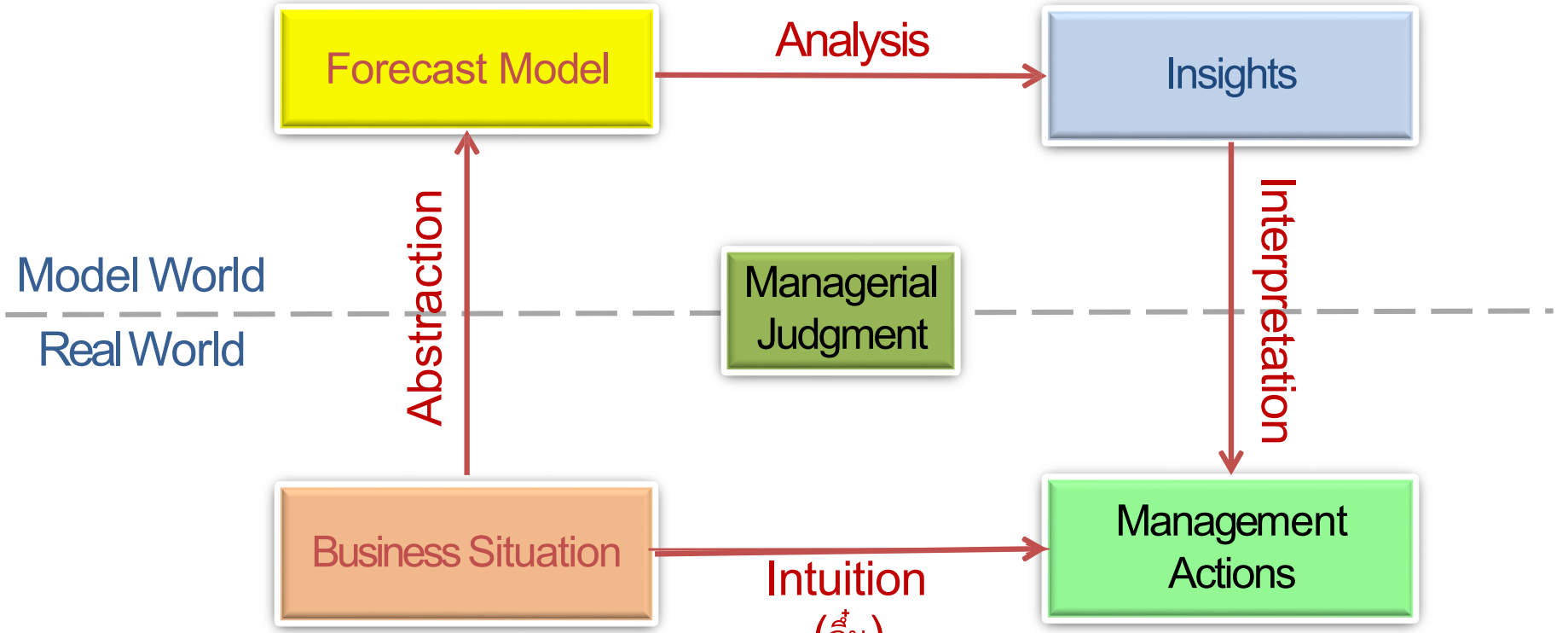
Trust Your Gut (กึ๋น)!

*40% major decision making are based on **gut feeling**!

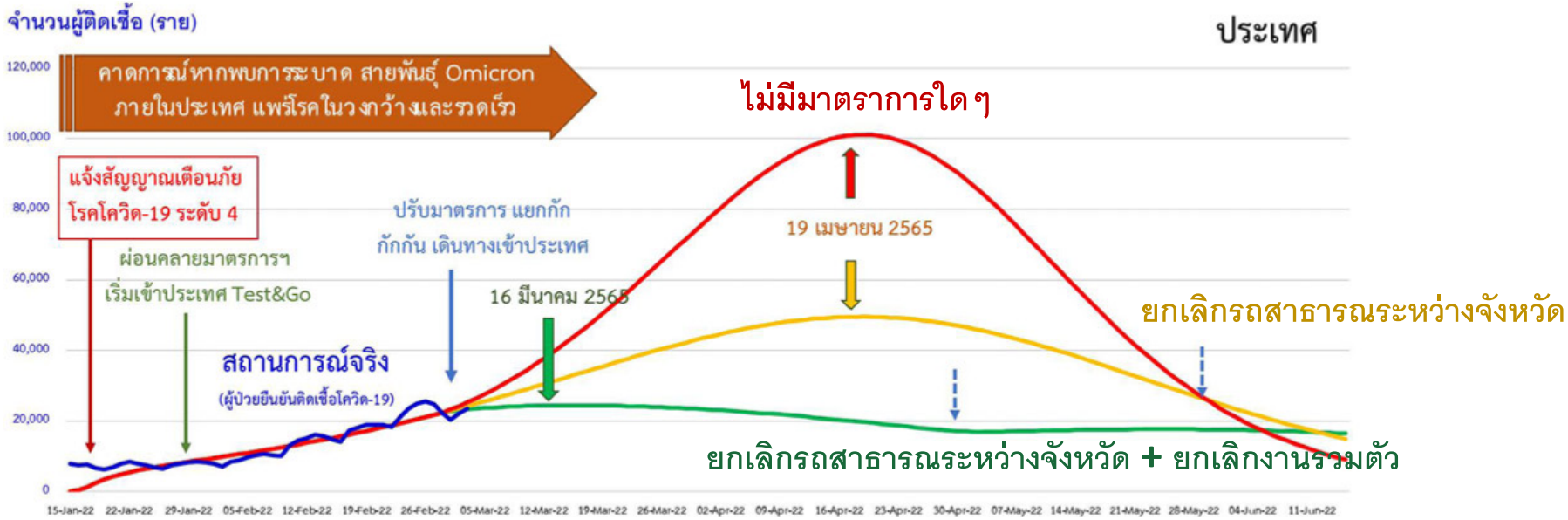
* Accenture 2008

Data Science to drive Public Health Policy

จำลองสถานการณ์การระบาดกรณีที่มีการ Lockdown vs ไม่ Lockdown ช่วงสงกรานต์



Disease Forecasting



Before modelling, we need to think of…

Disease Importance
(ความสำคัญของโรค)

- Frequency
- Severity
- Impact (QOL, Social, economic)
- Public/political concern

Modifiability
(ความเป็นไปได้ในการปรับเปลี่ยน)

- Preventable disease
- Available effective intervention
- Acceptable cost

Predictability
(ความเป็นไปได้ในการพยากรณ์)

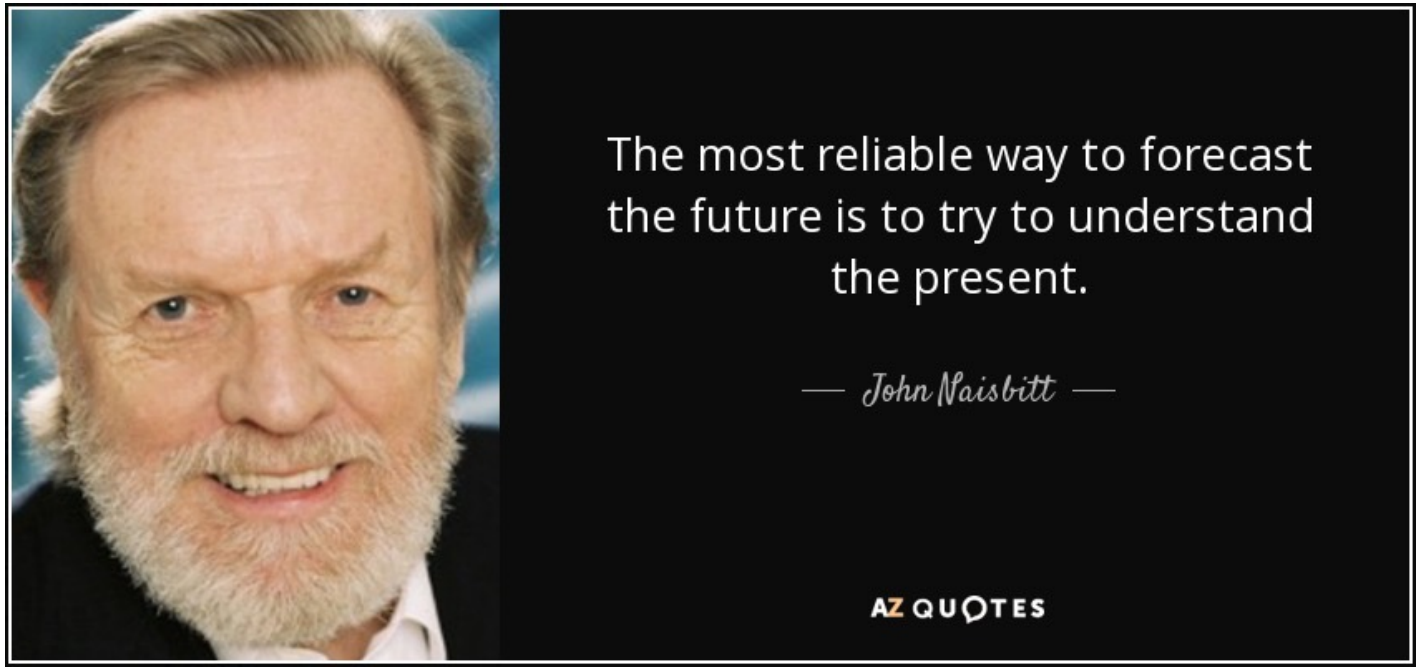
- Available data
- Available technique
- Available technology
- Available experts and experiences (both content and technique)

Time Series Forecasting

Time Series Forecasting

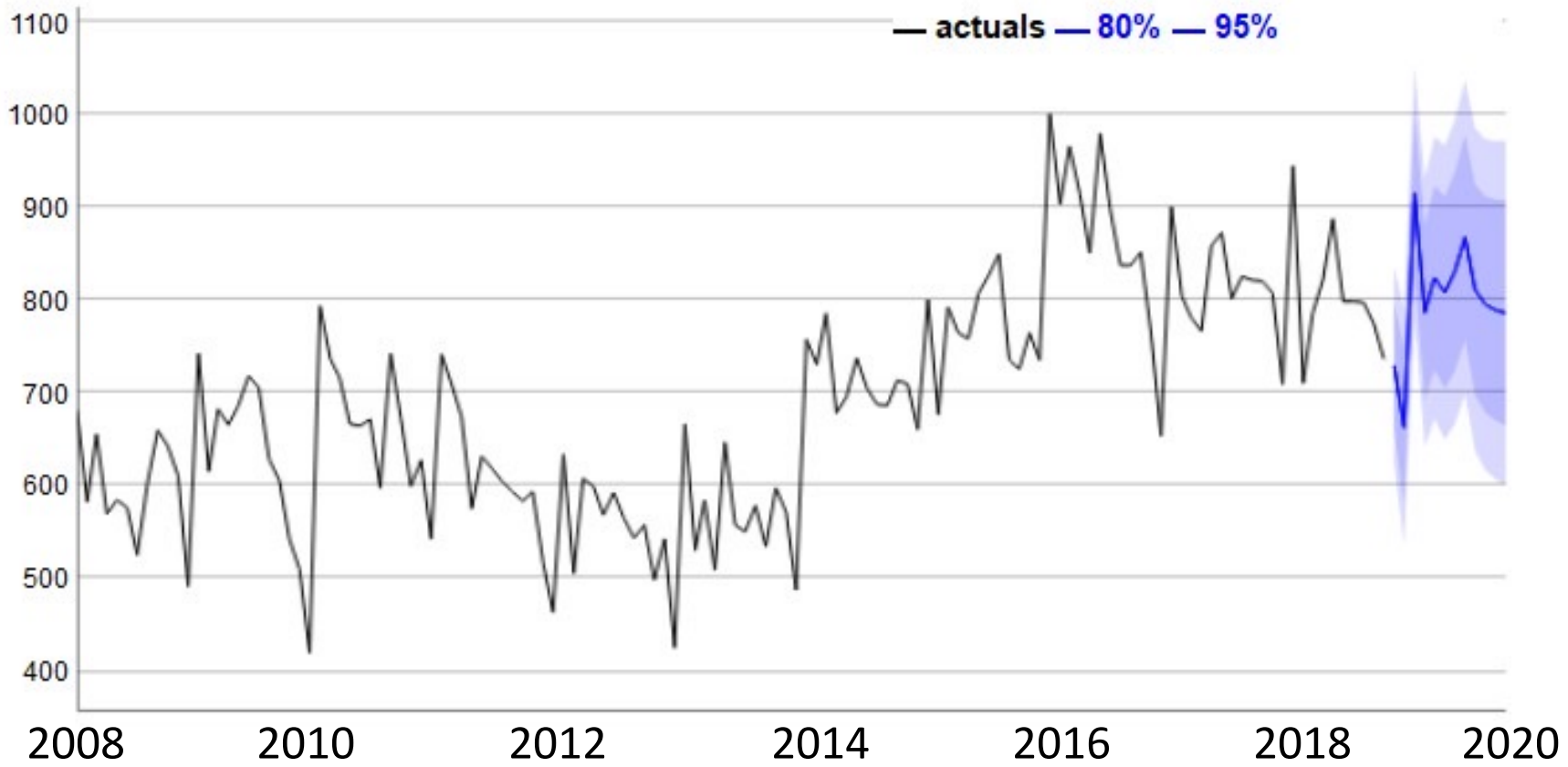
- **Forecasting** is the process of making **predictions of the future based on past and present data** and most commonly by analysis of trends.

การใช้อดีตและปัจจุบันทำนายอนาคต



Prediction of the incidence

Number of GC cases during 2008-2019 and **prediction in 2020**



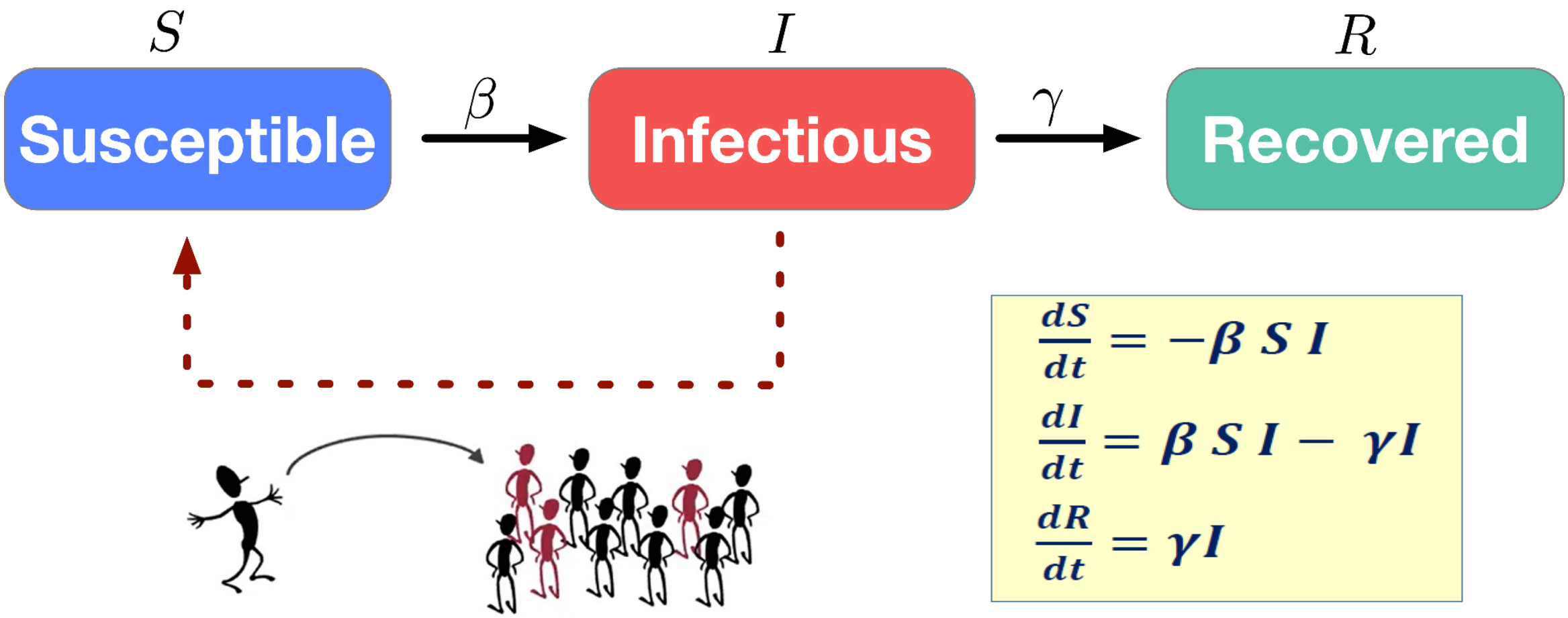
Source: R506, DDC

Mathematical Modeling

Mathematical Modeling

- An important tool in public health research and policy-making.
- Using mathematical equations and statistical methods to
  - simulate the spread of disease and effect of control measure,
  - predict future outbreaks,
  - design and evaluate the effectiveness of interventions.

Compartmental Modeling

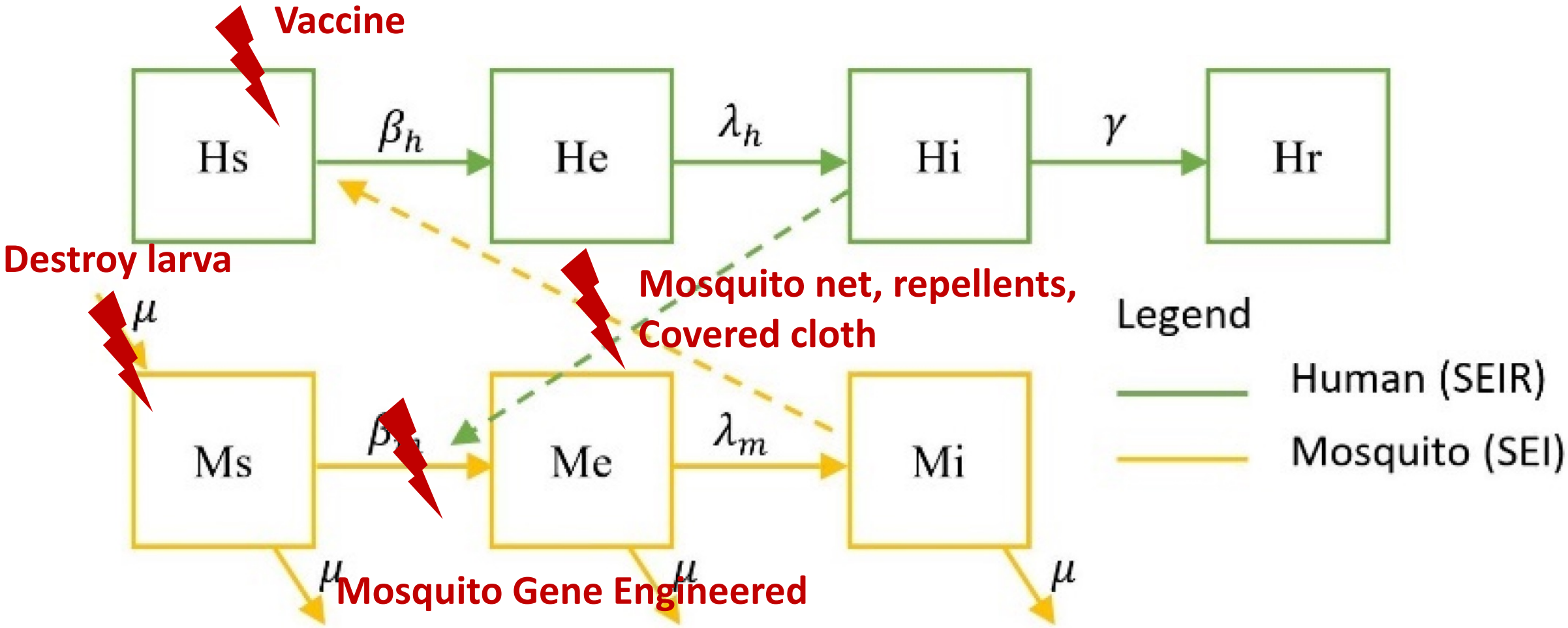


$$\frac{dS}{dt} = -\beta S I$$

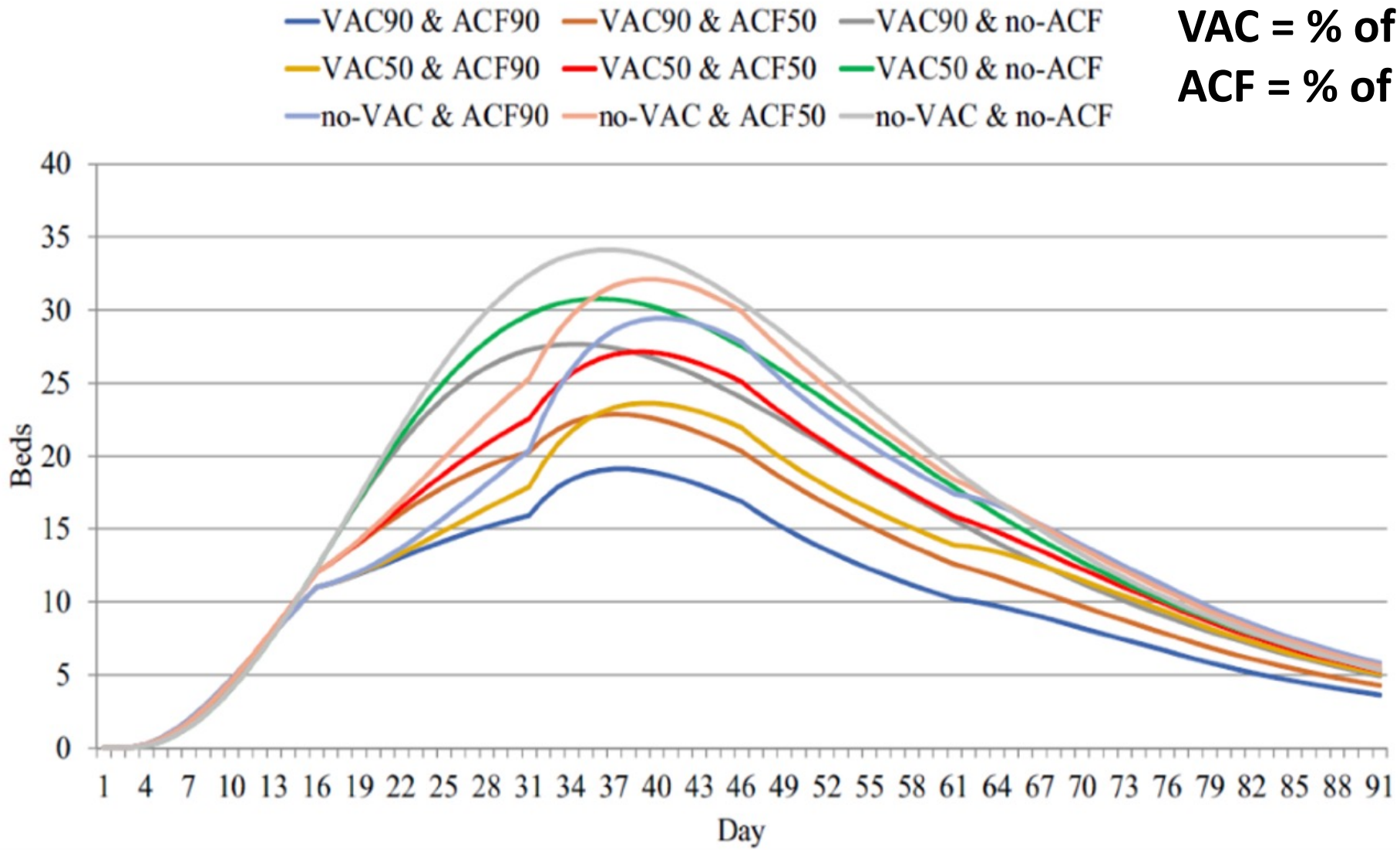
$$\frac{dI}{dt} = \beta S I - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

Compartmental Modeling with Intervention

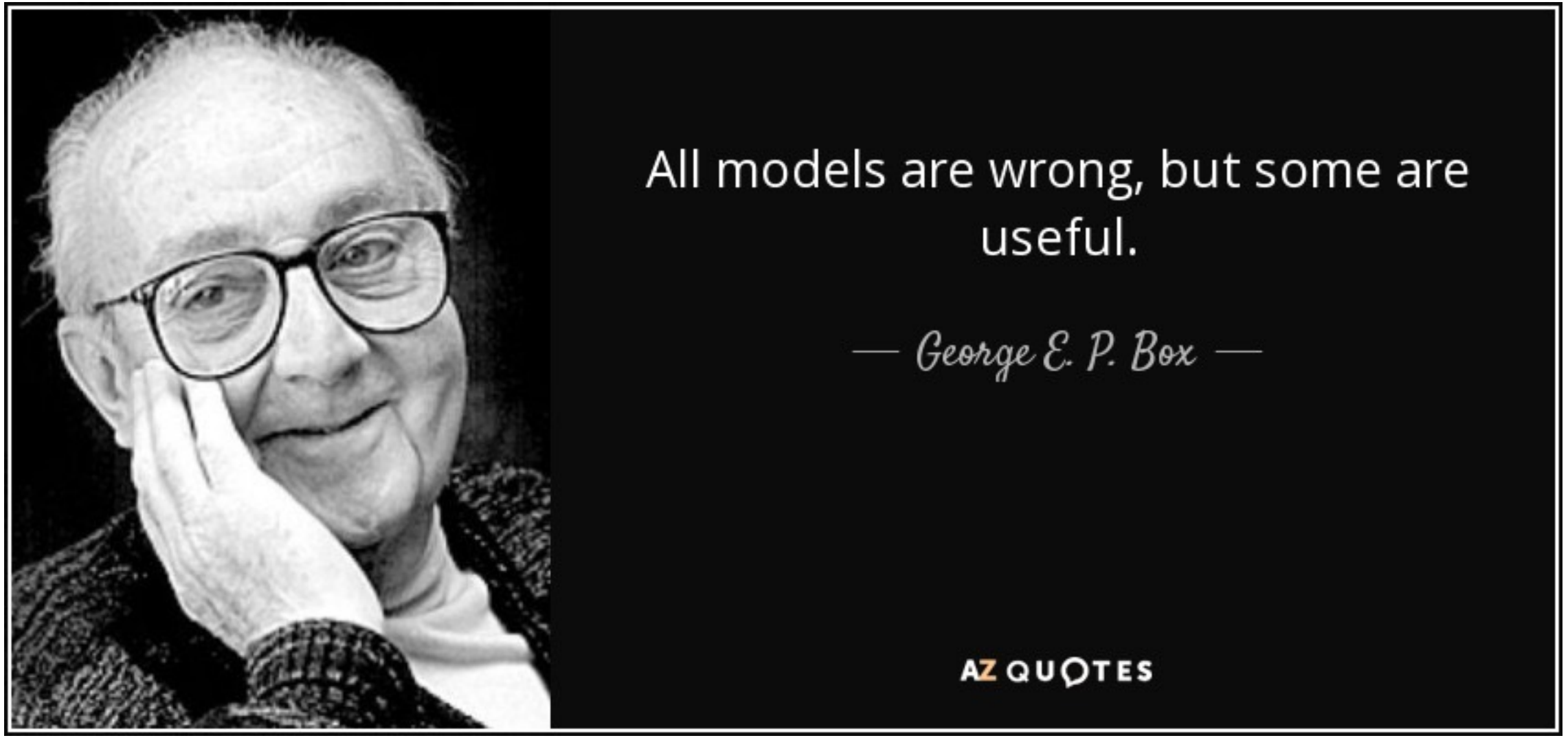


Compartmental Modeling with Intervention



VAC = % of Vaccine

ACF = % of Mosquito net


All models are wrong, but some are useful.
— George E. P. Box —
AZ QUOTES

GIS and Spatial Data Analysis

Spatial Data Analysis

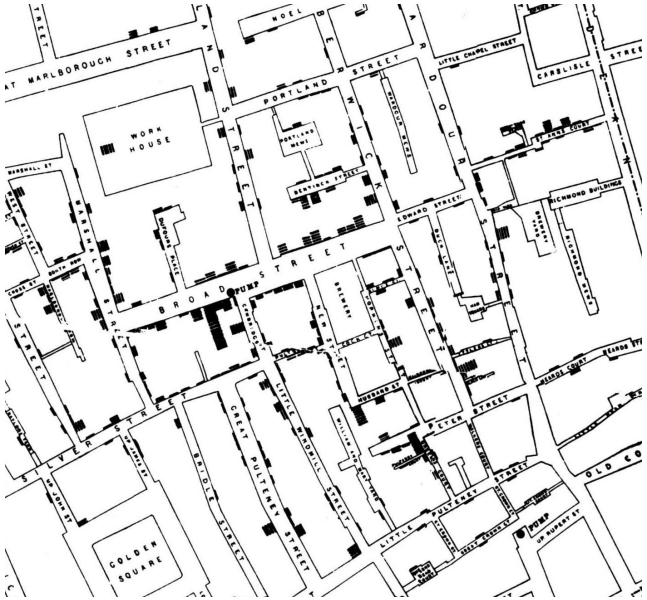
Spatial Data ⟷ **GIS** ⟷ **Spatial Statistics**

- **Spatial data** is when the "***where***" in addition to the "what" is observed and reported
  - Data that contains a **spatial reference**
    - Address (house number, district, province
    - Coordinate system (latitude, longtitude)
    - Spot on map

John Snow and the Broad Street Pump

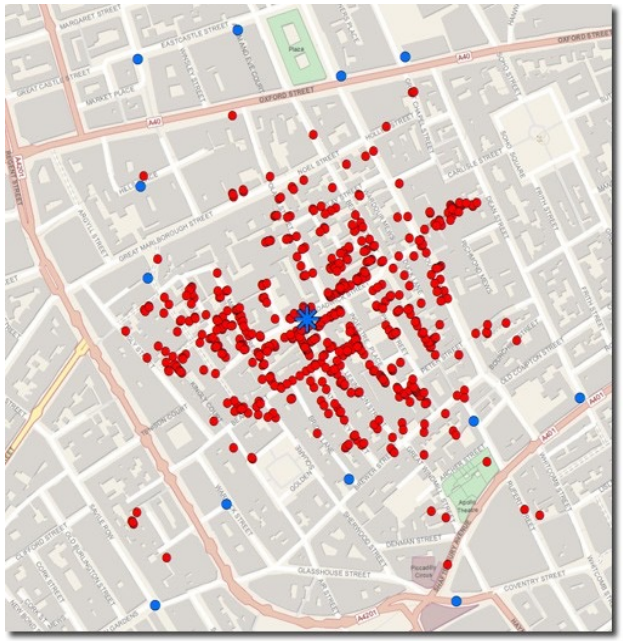
Spatial Data

Location of Cholera Death in London



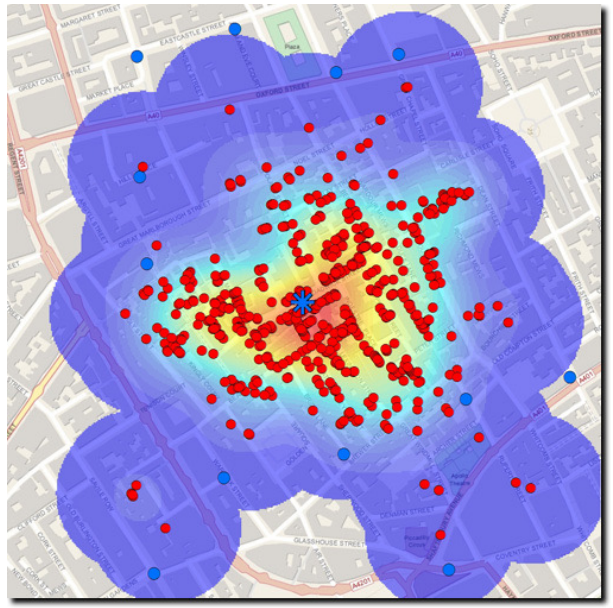
GIS

Cholera Death mapping



Spatial Statistics

Cholera Death hotspot

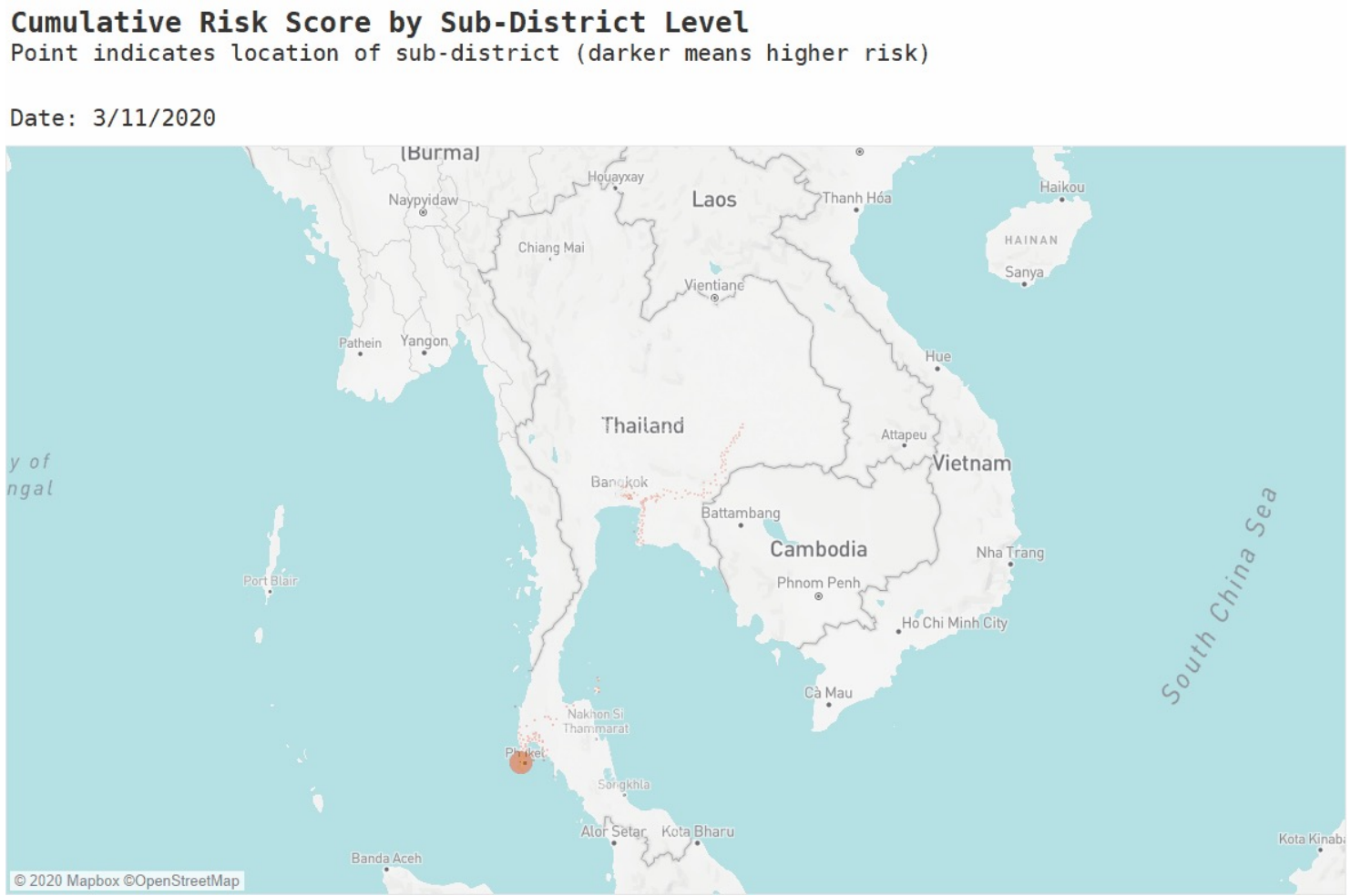


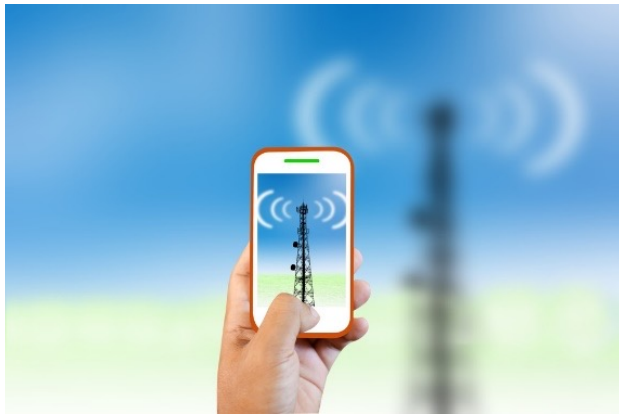
Georeferencing

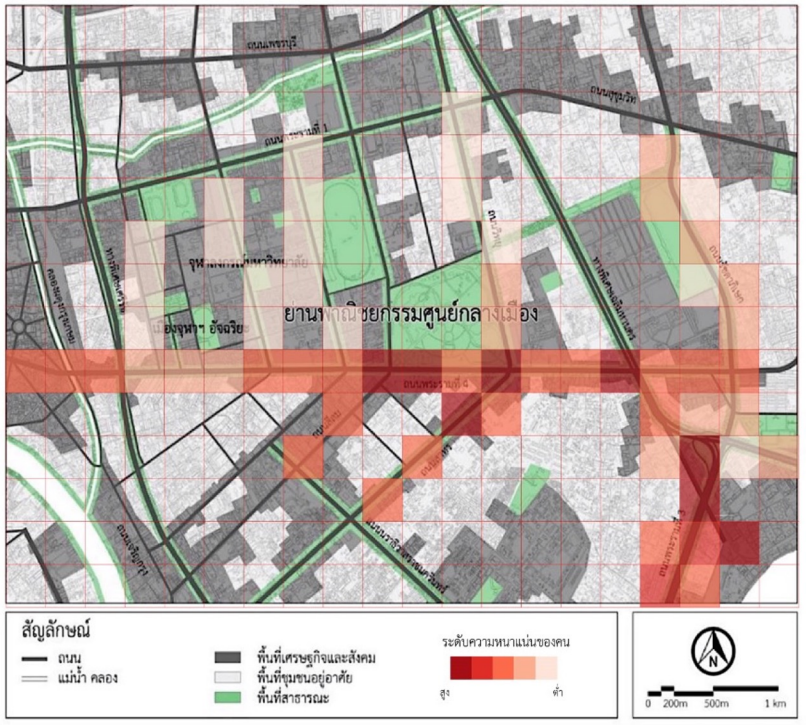
Spatial Intensity

GIS and Movement Tracking during Pandemic

Movement of Bang La, Phuket visitors after March 11



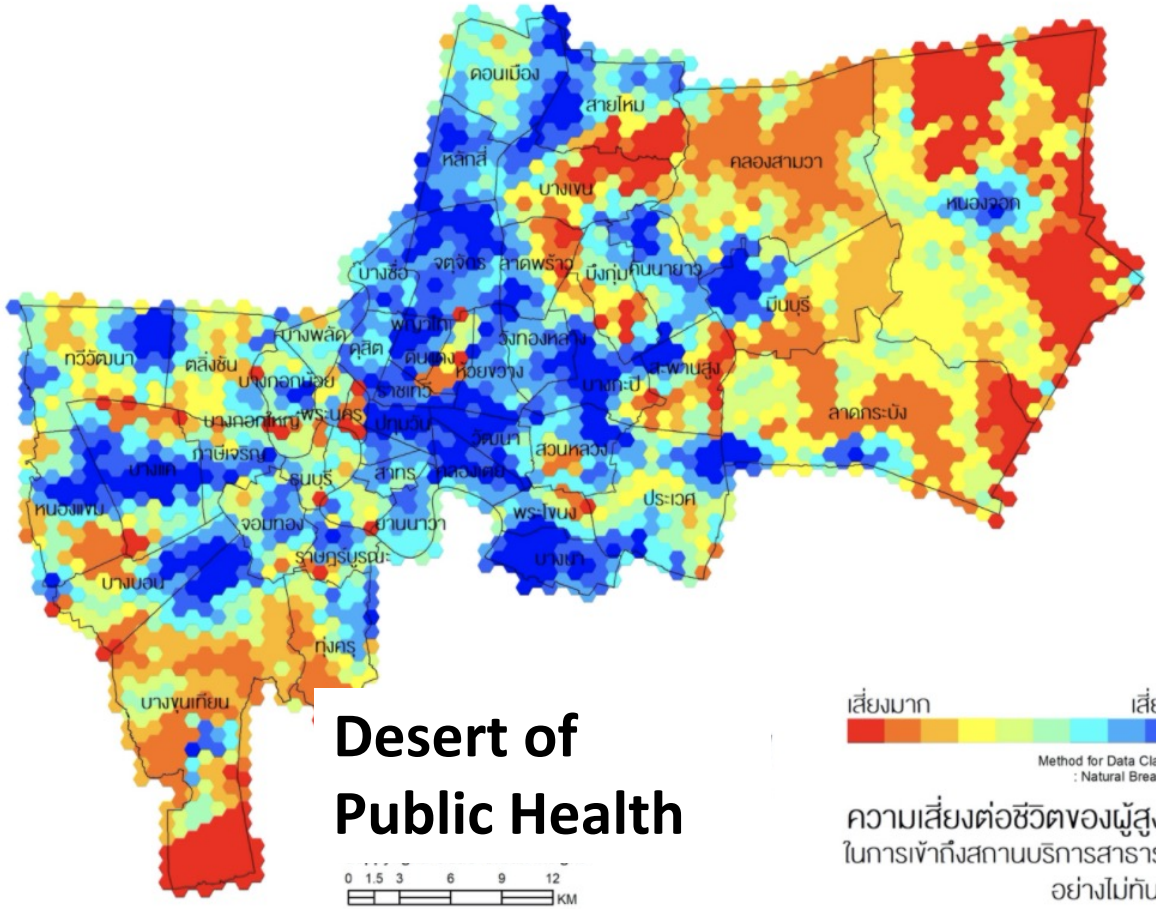




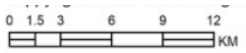
Source: True Analytics Business Unit, Operation team, DPHI, TDG

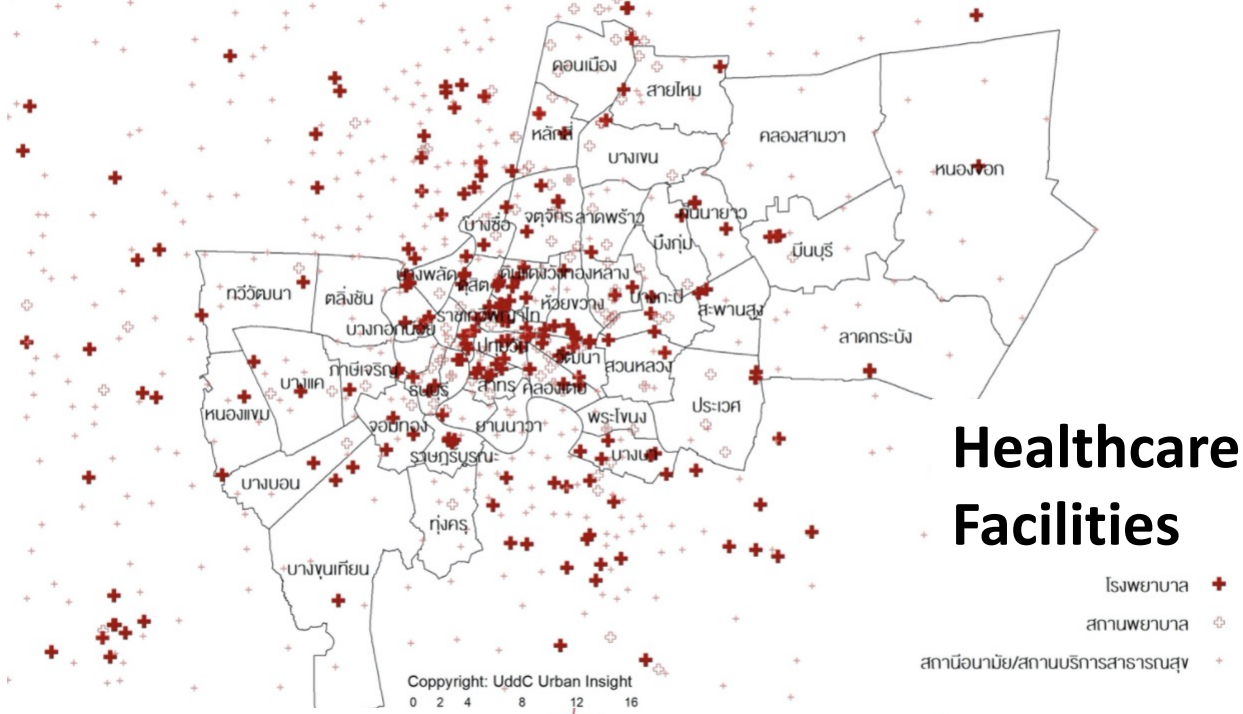
Health Facilities Accessibility among Eldery in Bangkok

ว่าน ฉันทวิลาสวงศ์ | อดิศักดิ์ กันทะเมืองลี้

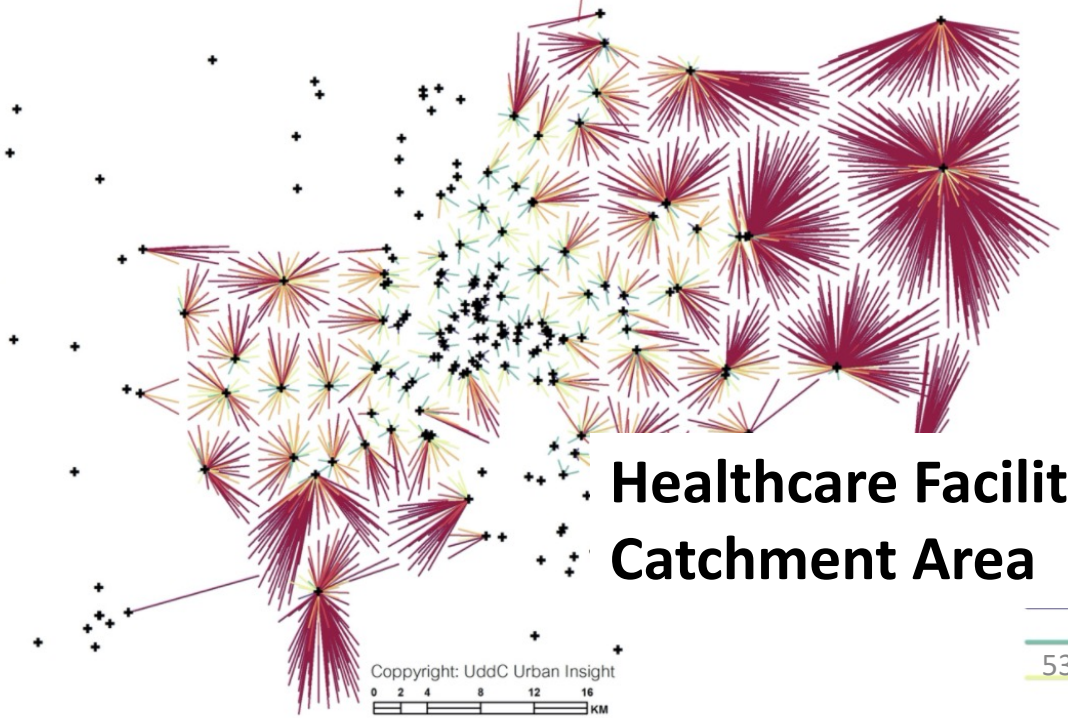


Desert of Public Health





Healthcare Facilities



Healthcare Facilities Catchment Area

RESEARCH ARTICLE

Large scale detailed mapping of dengue vector breeding sites using street view images

Peter Haddawy[1,2]*, Poom Wettayakorn[1], Boonpakorn Nonthaleerak[1], Myat Su Yin[1], Anuwat Wiratsudakul[3], Johannes Schöning[4], Yongjua Laosiritaworn[5], Klestia Balla[6], Sirinut Euaungkanakul[1], Papichaya Quengdaeng[1], Kittipop Choknitipakin[1], Siripong Traivijitkhun[1], Benyarut Erawan[1], Thansuda Kraisang[1]

1 Faculty of ICT, Mahidol University, Salaya, Thailand, **2** Bremen Spatial Cognition Center, University of Bremen, Bremen, Germany, **3** Faculty of Veterinary Science, Mahidol University, Salaya, Thailand, **4** University of Bremen, Bremen, Germany, **5** Ministry of Public Health, Bangkok, Thailand, **6** Computer Science Department, School of Science and Technology, University of Camerino, Camerino, Italy

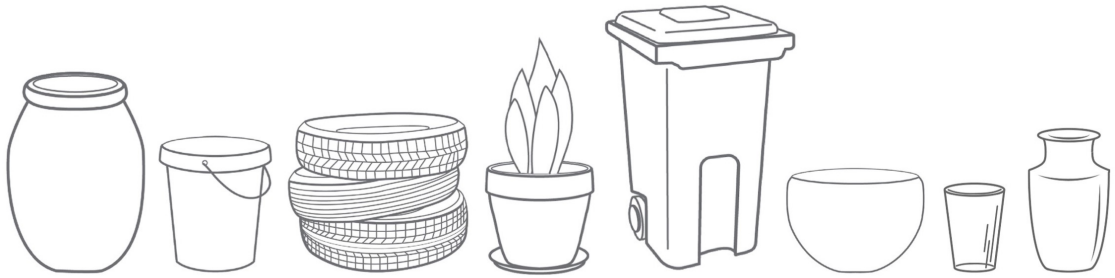
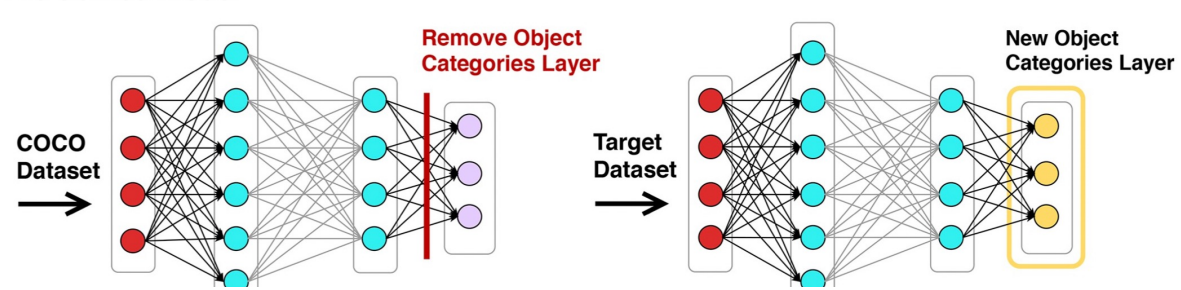


Fig 1. Common outdoor dengue vector breeding sites in Thailand (from left to right): large jar, bucket, old tire, potted plant, bin, ceramic bowl, cup, vase.



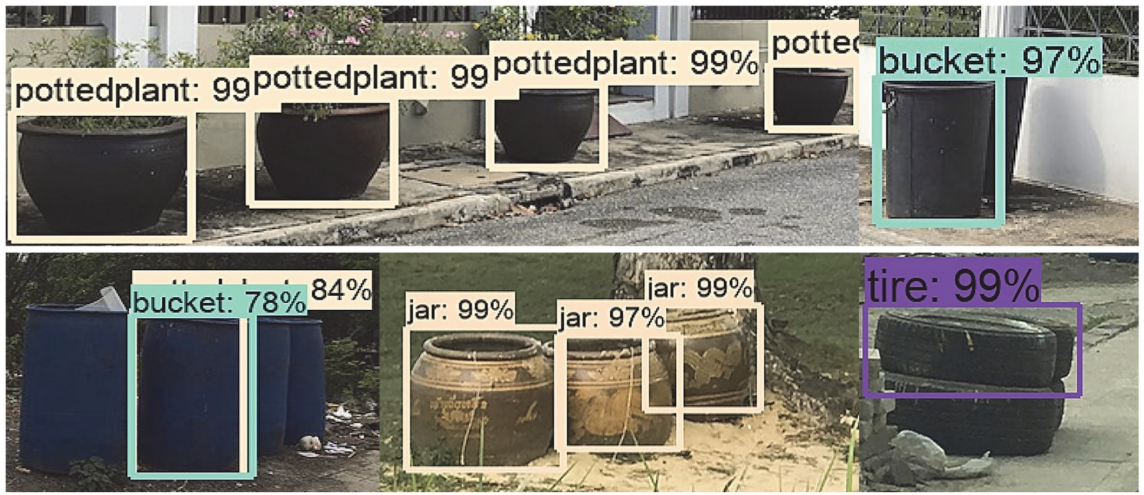
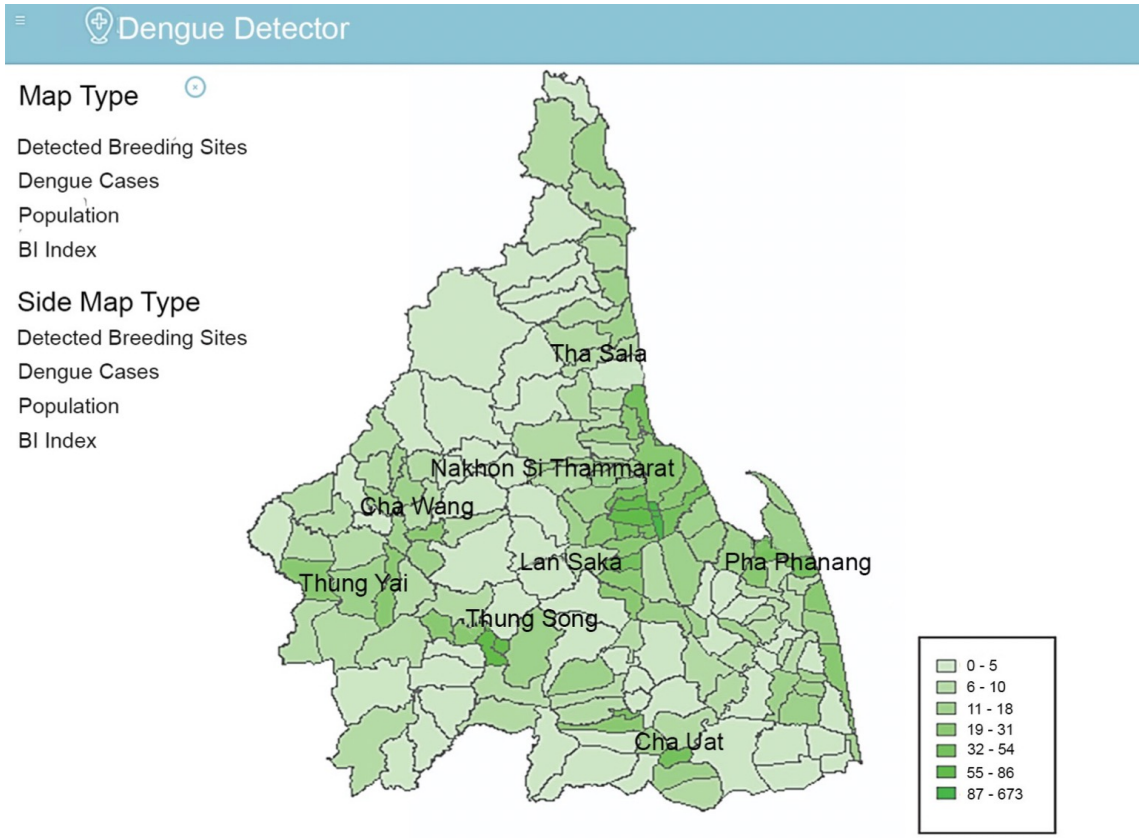
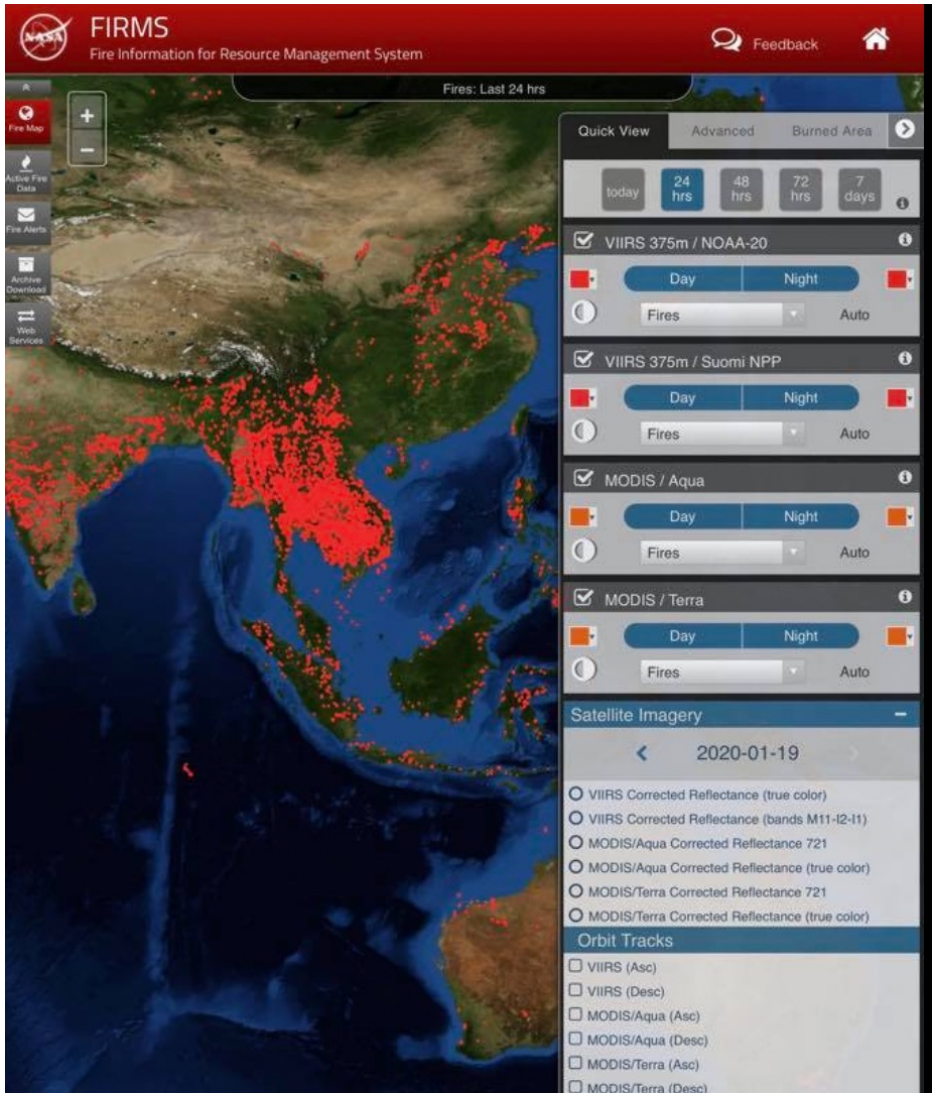


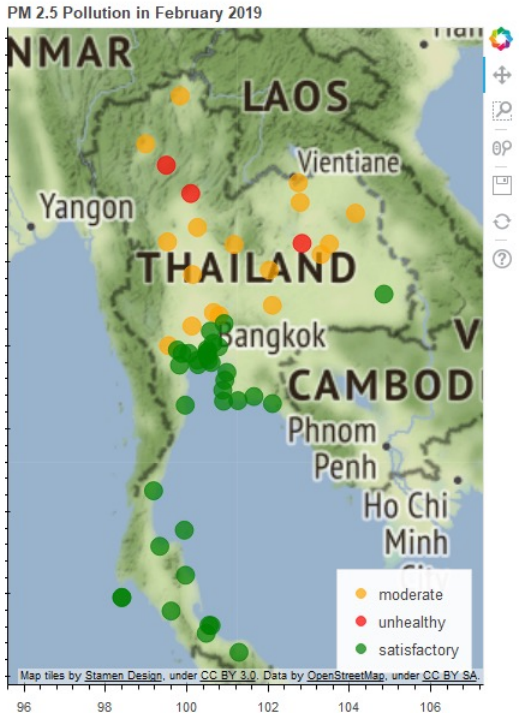
Fig 3. Examples of containers detected by using Faster R-CNN with new transferred categories.

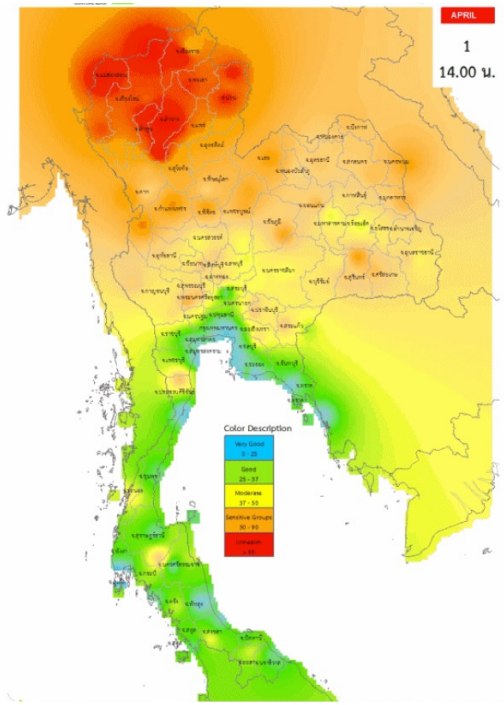


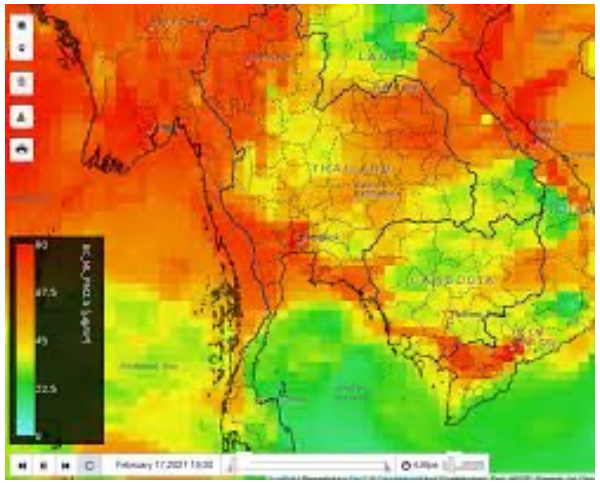
https://journals.plos.org/plosntds/article/file?id=10.1371/journal.pntd.0007555&type=printable

PM2.5


FIRMS
Fire Information for Resource Management System
Feedback
Fires: Last 24 hrs
Quick View
Advanced
Burned Area
today
24 hrs
48 hrs
72 hrs
7 days
VIIRS 375m / NOAA-20
Day
Night
Fires
Auto
VIIRS 375m / Suomi NPP
MODIS / Aqua
MODIS / Terra
Satellite Imagery
2020-01-19
VIIRS Corrected Reflectance (true color)
VIIRS Corrected Reflectance (bands M11-I2-I1)
MODIS/Aqua Corrected Reflectance 721
MODIS/Aqua Corrected Reflectance (true color)
MODIS/Terra Corrected Reflectance 721
MODIS/Terra Corrected Reflectance (true color)
Orbit Tracks
VIIRS (Asc)
VIIRS (Desc)
MODIS/Aqua (Asc)
MODIS/Aqua (Desc)
MODIS/Terra (Asc)
MODIS/Terra (Desc)


PM 2.5 Pollution in February 2019
LAOS
Vientiane
Yangon
THAILAND
Bangkok
Phnom Penh
Ho Chi Minh
moderate
unhealthy
satisfactory
Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under CC BY SA.
96
98
100
102
104
106


APRIL
1
14.00 น.
Color Description



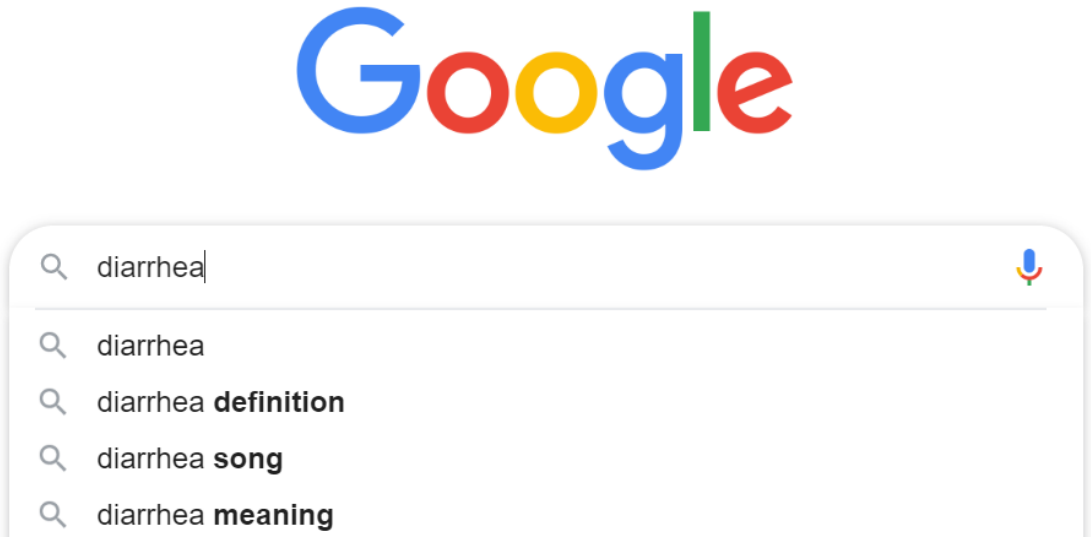
Outbreak detection with GIS data

ARTICLE OPEN

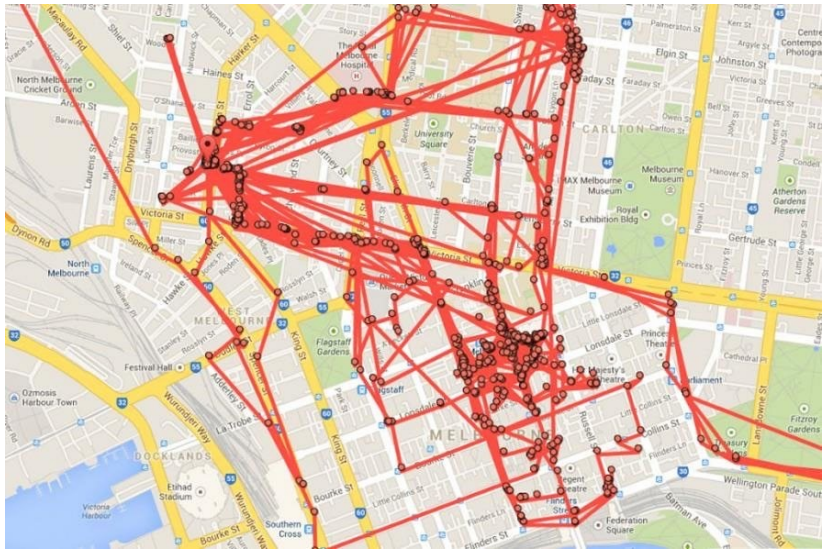
Machine-learned epidemiology: real-time detection of foodborne illness at scale

Adam Sadilek[1], Stephanie Caty[2], Lauren DiPrete[3], Raed Mansour[4], Tom Schenk Jr[5], Mark Bergtholdt[3], Ashish Jha[2,6], Prem Ramaswami[1] and Evgeniy Gabrilovich[1]

Anonymized web search data



Anonymized location data

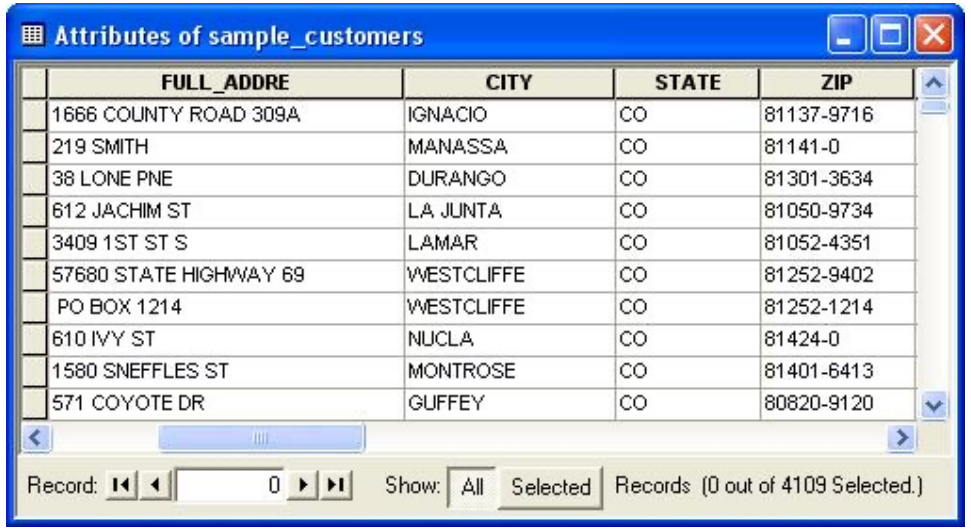


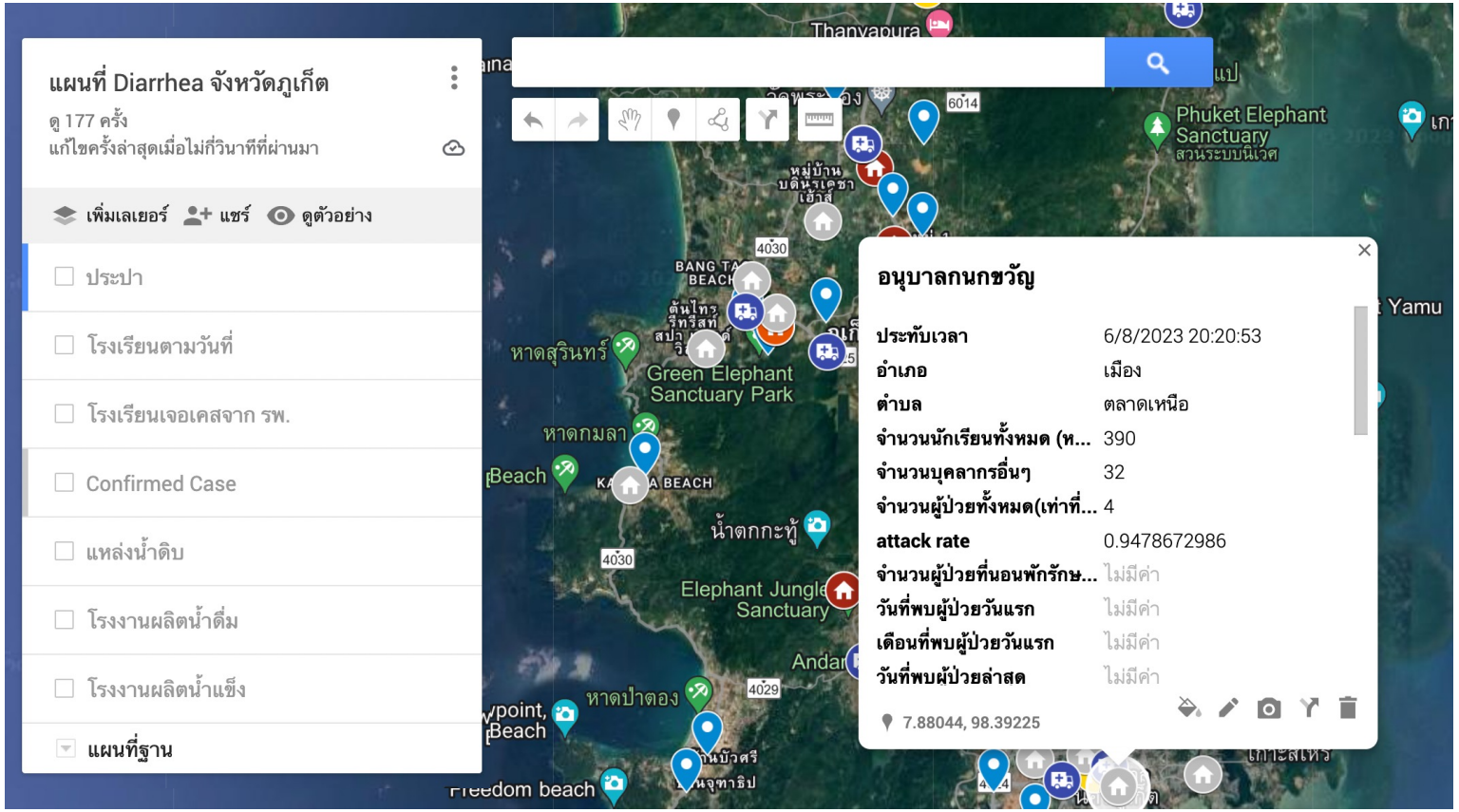
กระทรวงสาธารณสุข
MINISTRY OF PUBLIC HEALTH

Google My Map for Outbreak Investigation

Google My Maps

แผนที่ Diarrhea จังหวัดภูเก็ต - Google My Maps



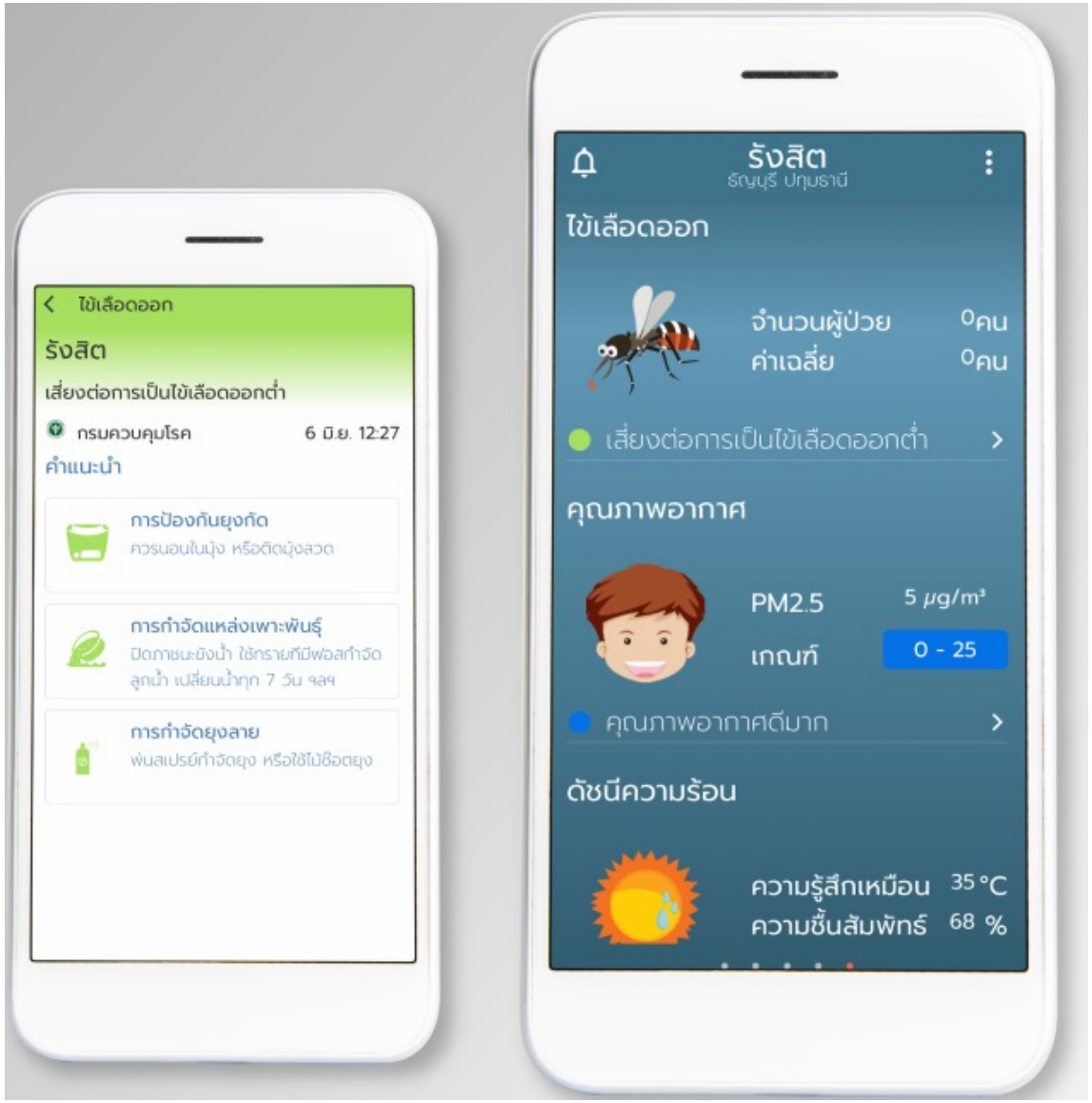


How To Get The Geolocation Of A Place In Google Forms?

Updated on September 3, 2020 by Neeraj Agarwal

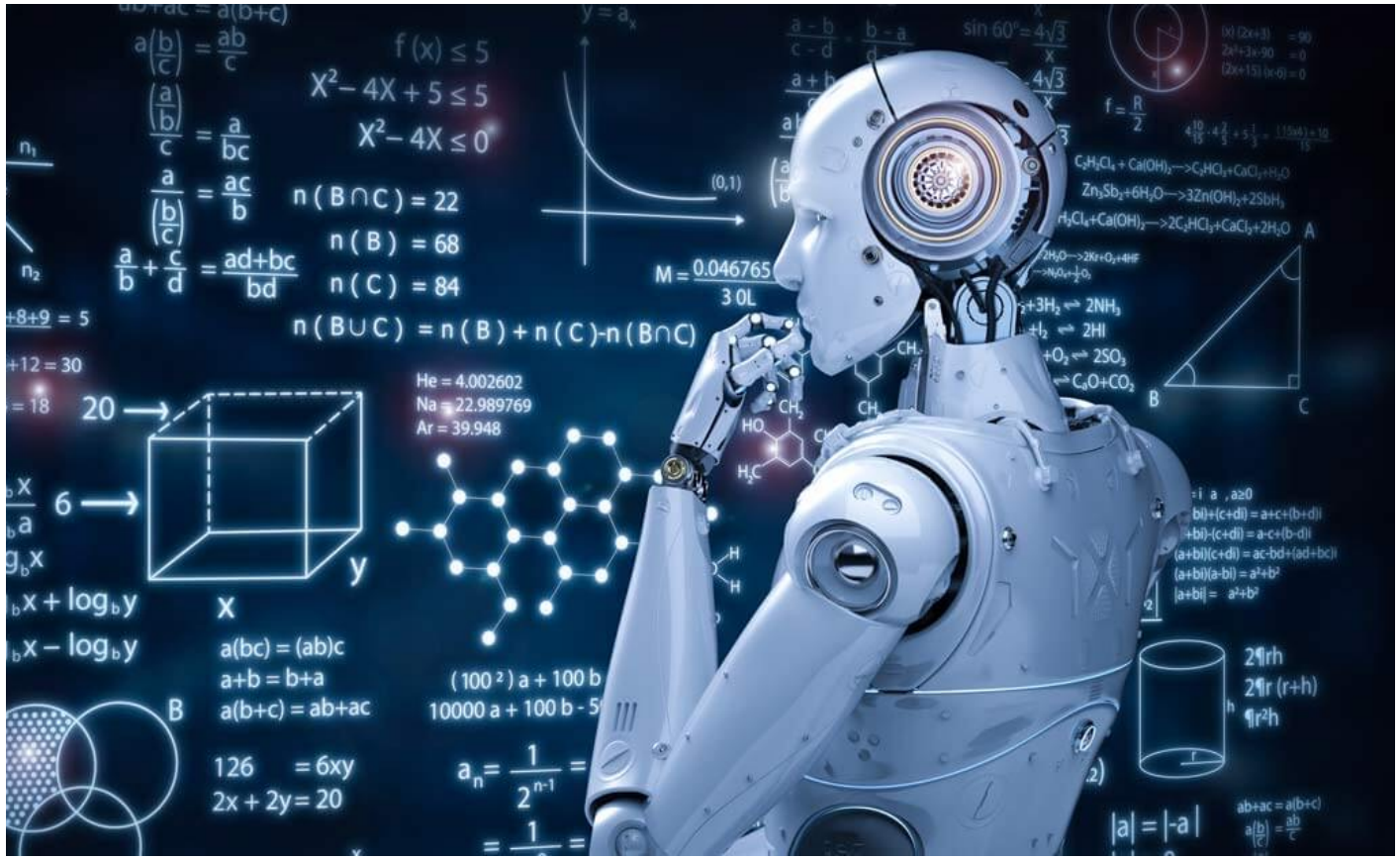
กระทรวงสาธารณสุข
MINISTRY OF PUBLIC HEALTH

Precision Health Risk Communication with GIS data

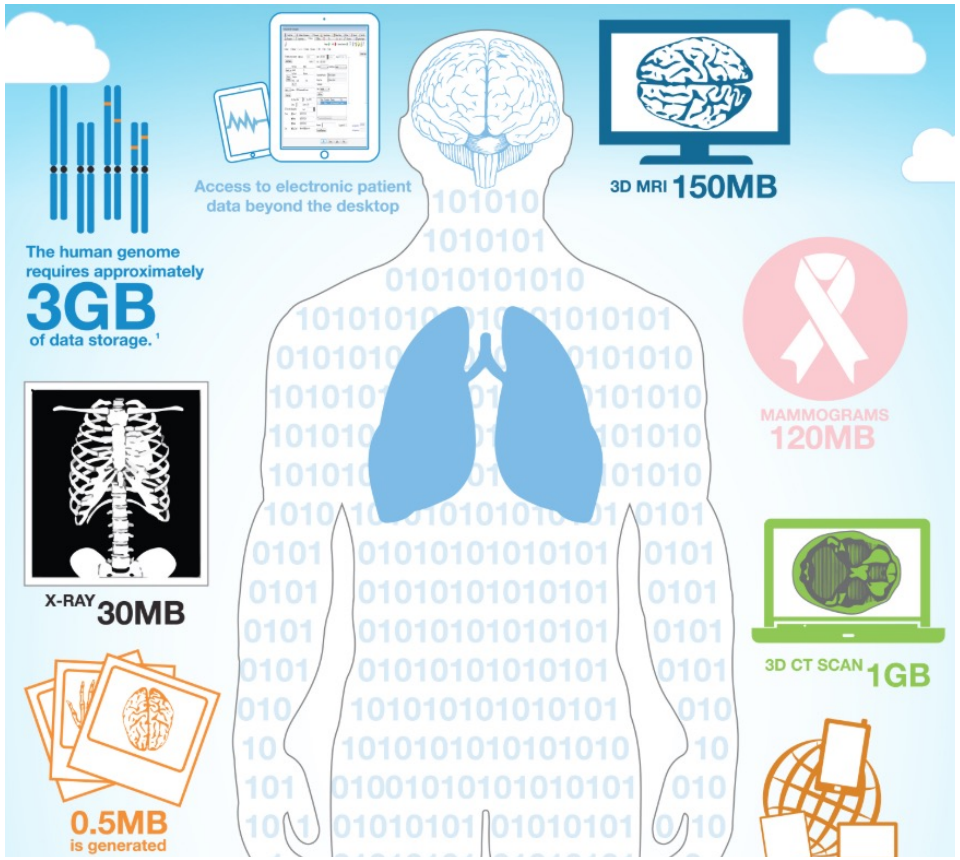


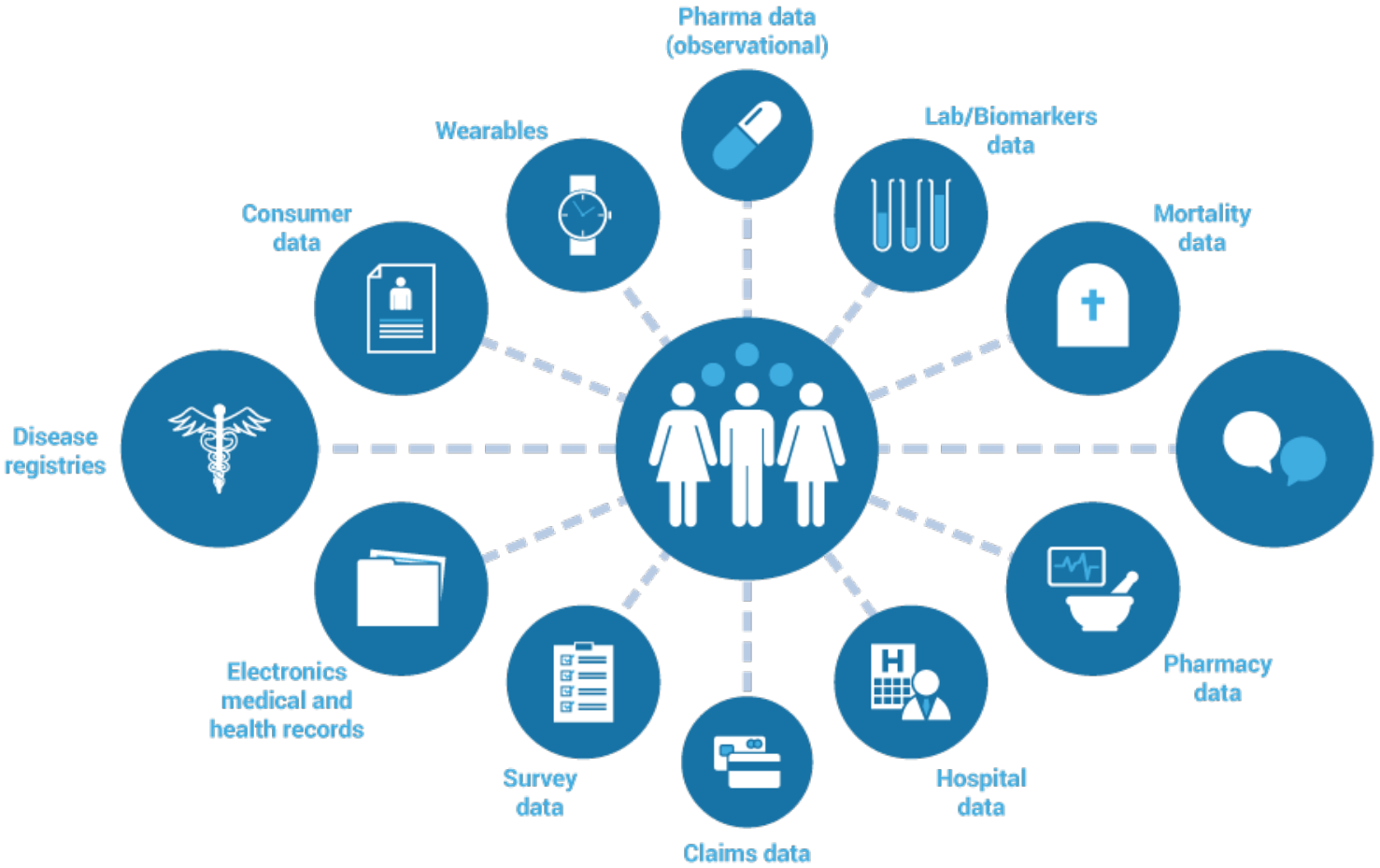
AI & ML in Public Health

Why AI and Machine Learning in Public Health?

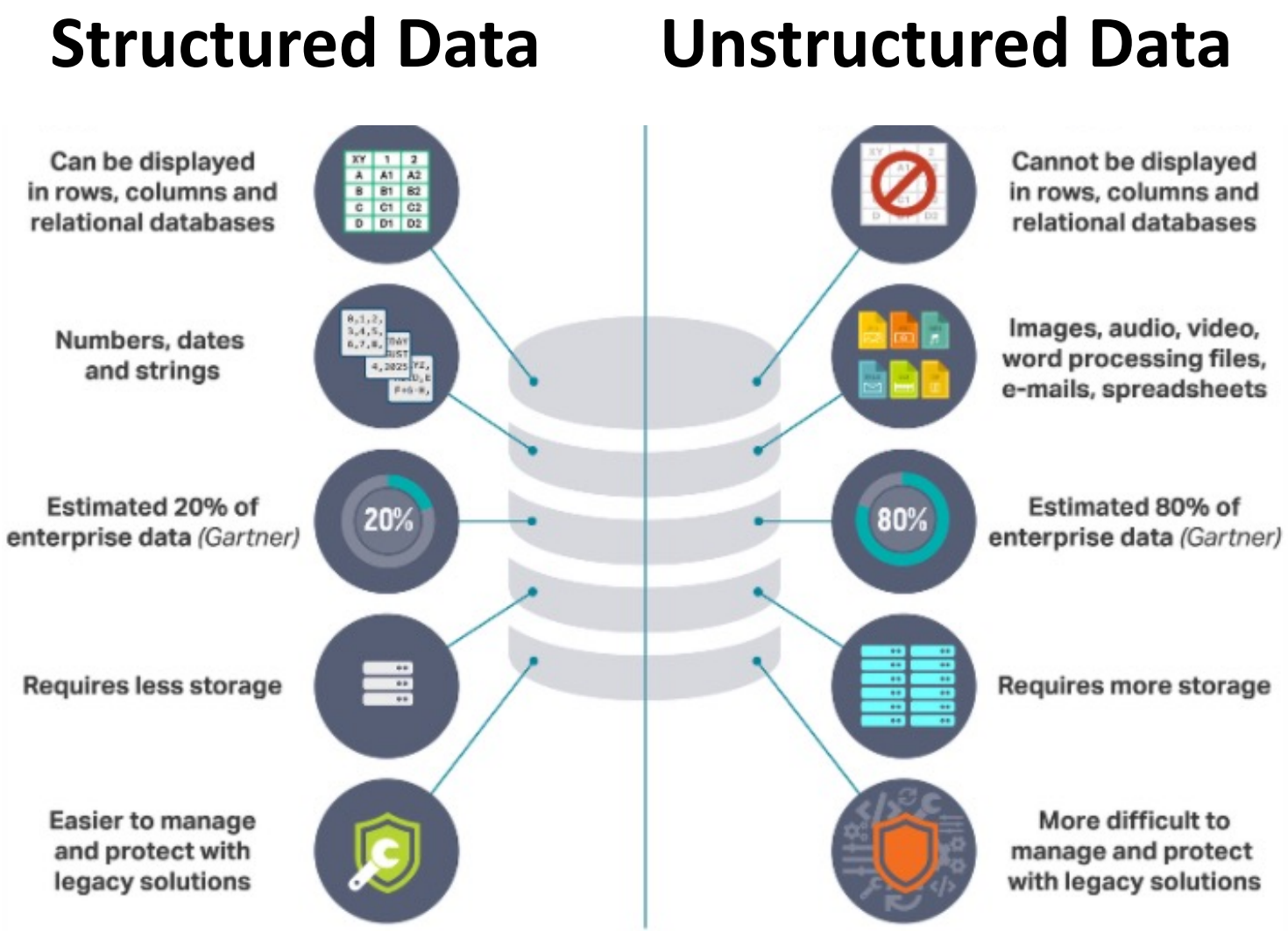


Public Health Data is Big Data





Complexity of Unstructured data



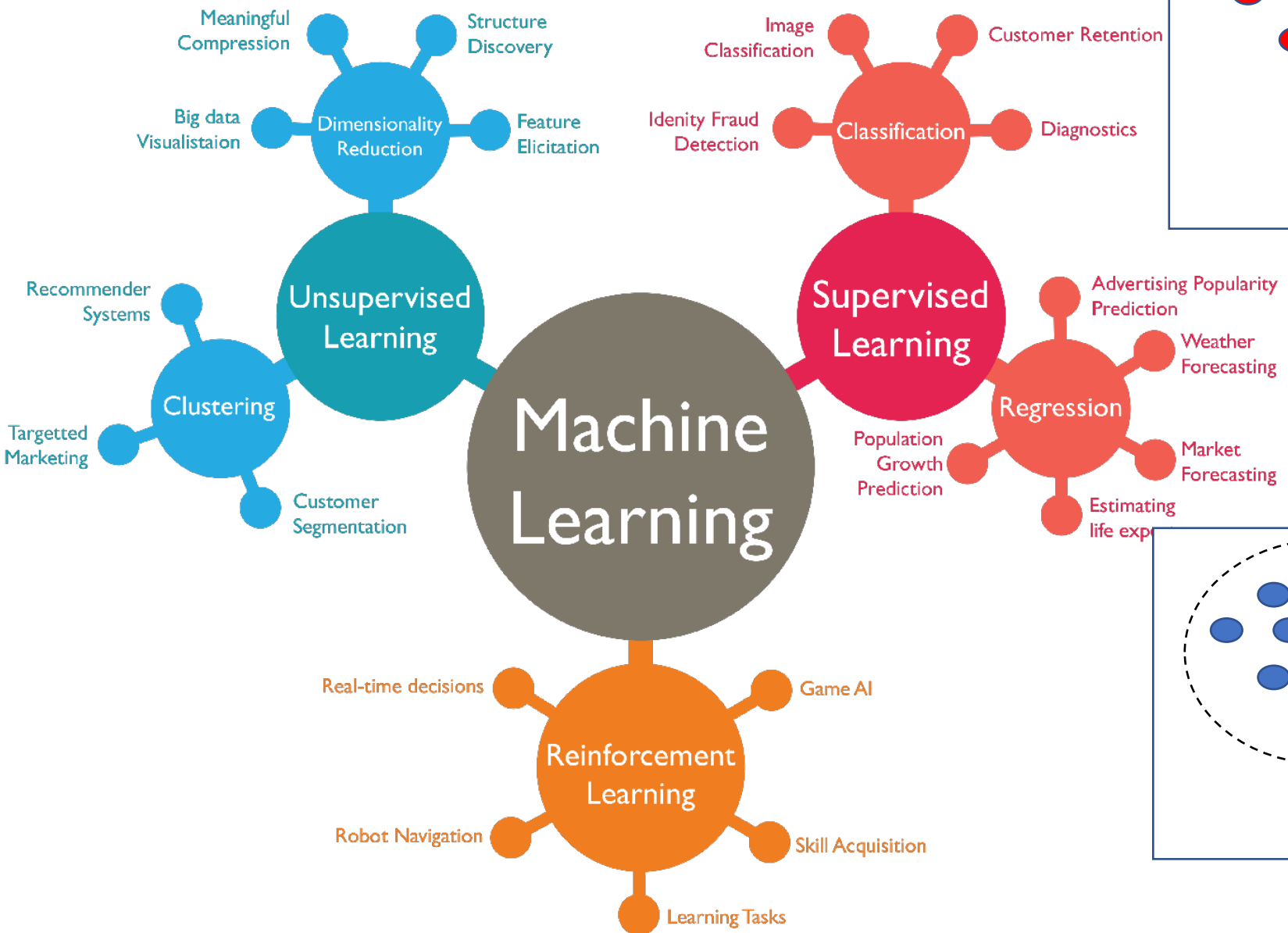
- **Structured data**
  - Database
  - Data Table
- **Unstructured data**
  - Free Text
  - Image(s)
  - Video
  - Audio / Voice / Speech
  - Motion
  - Signal
  - Rhythm
  - Geolocation (semi-str.)

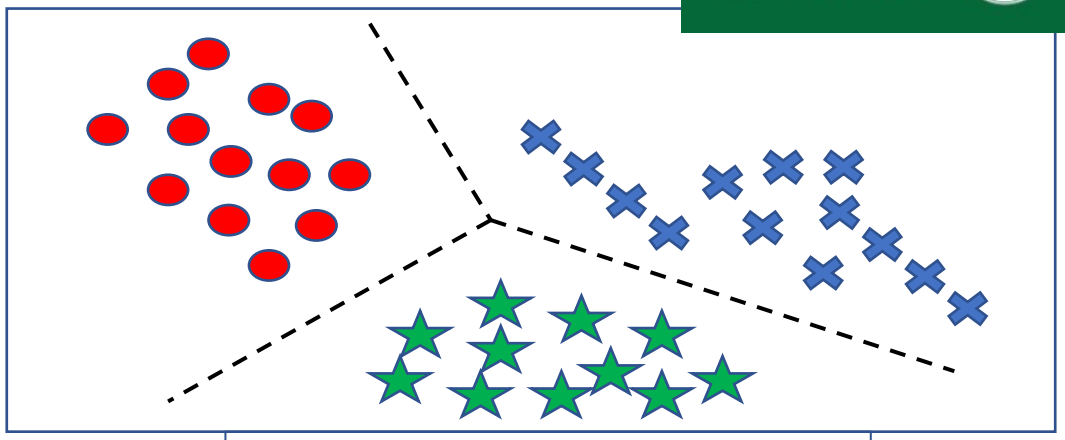
https://lawtomated.com/structured-data-vs-unstructured-data-what-are-they-and-why-care/

Need of Data Driven Decision Making

From Evidence Based Medicine to Data-driven Medicine

Machine Learning



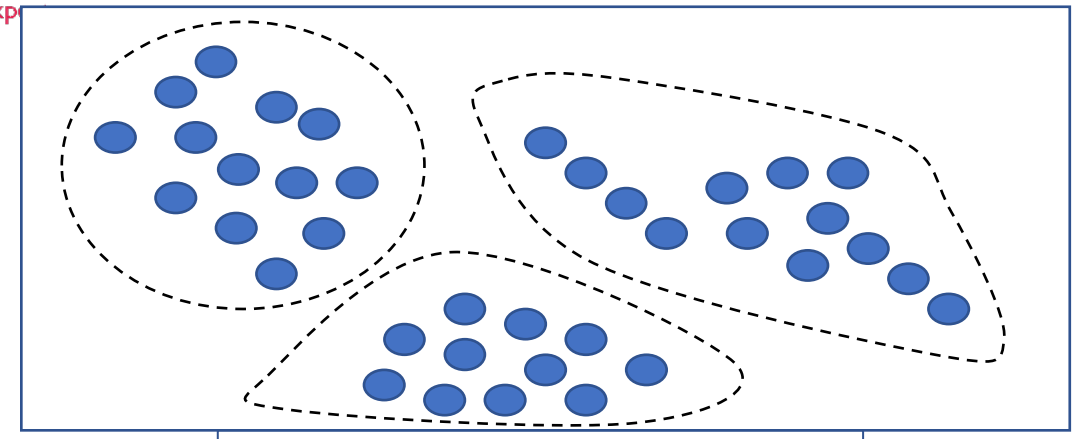


Supervised learning

Data are **labeled**

- Sick
- Borderline
- Healthy

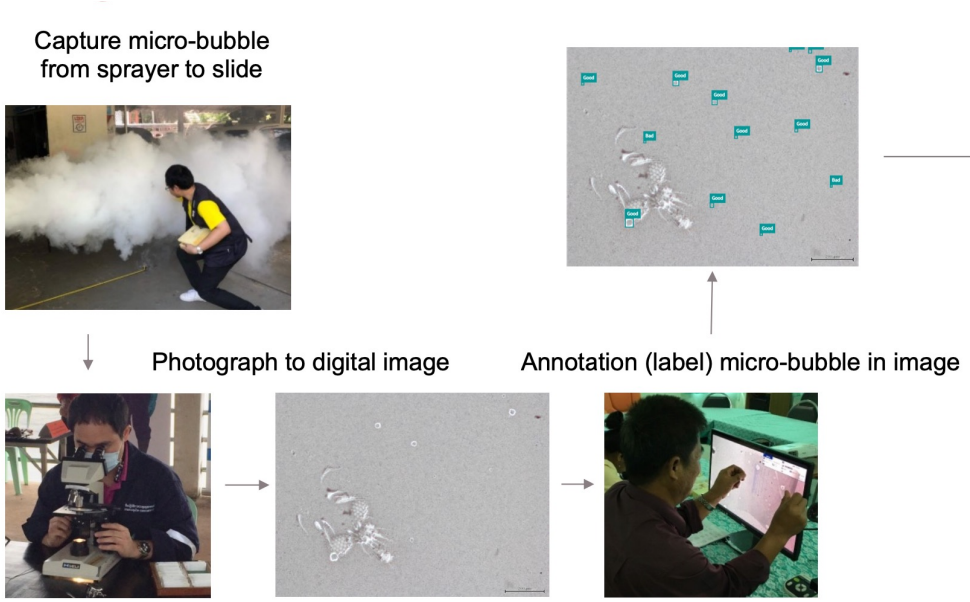
Data are **not labeled**

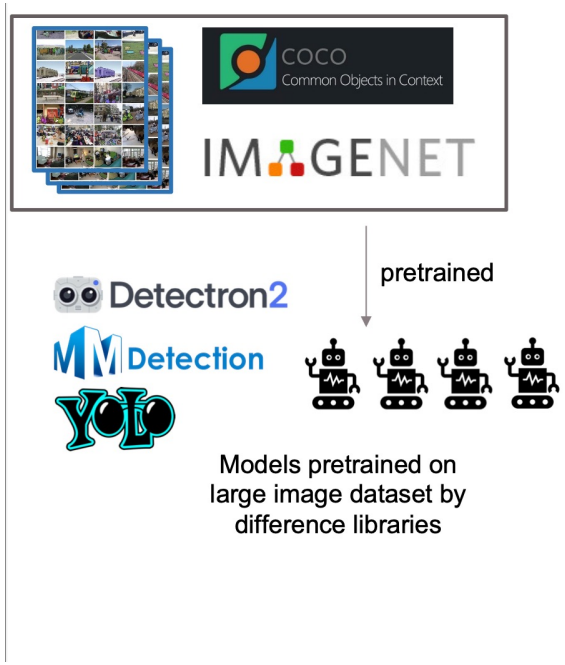


Unsupervised learning

Machine learning models for micro-bubble image detection in insecticide sprayer quality control: addressing class and scale imbalance.

Suppasit Srisaeng, Pongsakorn Sadakorn





Model	Back bone	Library	Epoch	Unqualified micro-bubble				Qualified micro-bubble				mAP	mAR	Time / image
				PPV	Sens	Max F1	AP	PPV	Sens	Max F1	AP			
Retina Net	X-101		12	4	91	7	31.2	31	99	47	70.5	50.8	69.2	0.14 s
			12	2	23	4	8.0	16	94	28	31.8	16.3	33.2	0.24 s
Faster-RCNN	R-50		12	34	30	32	17.7	83	95	88	68.8	43.2	50.5	0.10 s
	X-101		12	38	32	35	26.5	96	76	85	73.3	49.9	56.9	0.26 s
			12	37	46	41	23.9	89	29	44	23.8	24.0	29.8	0.10 s
YOLO v7			300	54	42	47	31.0	93	95	83	73.8	52.4		0.08 s
Swin Transformer v2	RCNN		12	27	42	33	20.0	85	28	42	22.7	21.4	27.2	0.26 s
FocalNet	RCNN		12	14	43	21	13.5	80	28	42	21.3	17.4	25.8	0.25 s
Intern Image	Cascade -RCNN		12	18	50	26	25.5	84	34	49	28.4	21.6	26.7	0.42 s

กระทรวงสาธารณสุข
MINISTRY OF PUBLIC HEALTH

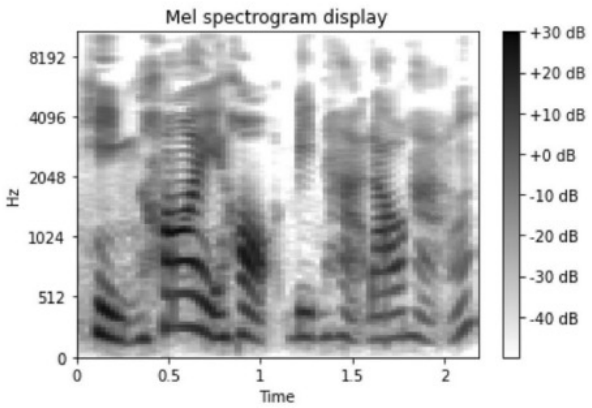
Voice Processing

Check for updates

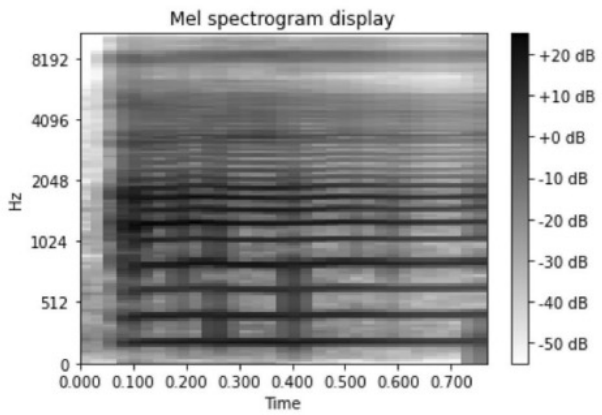
OPEN

Identifying individuals with recent COVID-19 through voice classification using deep learning

Pichatorn Suppakitjanusant[1], Somnuek Sungkanuparph[1], Thananya Wongsinin[1], Sirapong Virapongsiri[1], Nittaya Kasemkosin[2], Laor Chailurkit[3] & Boonsong Ongphiphadhanakul[3]✉



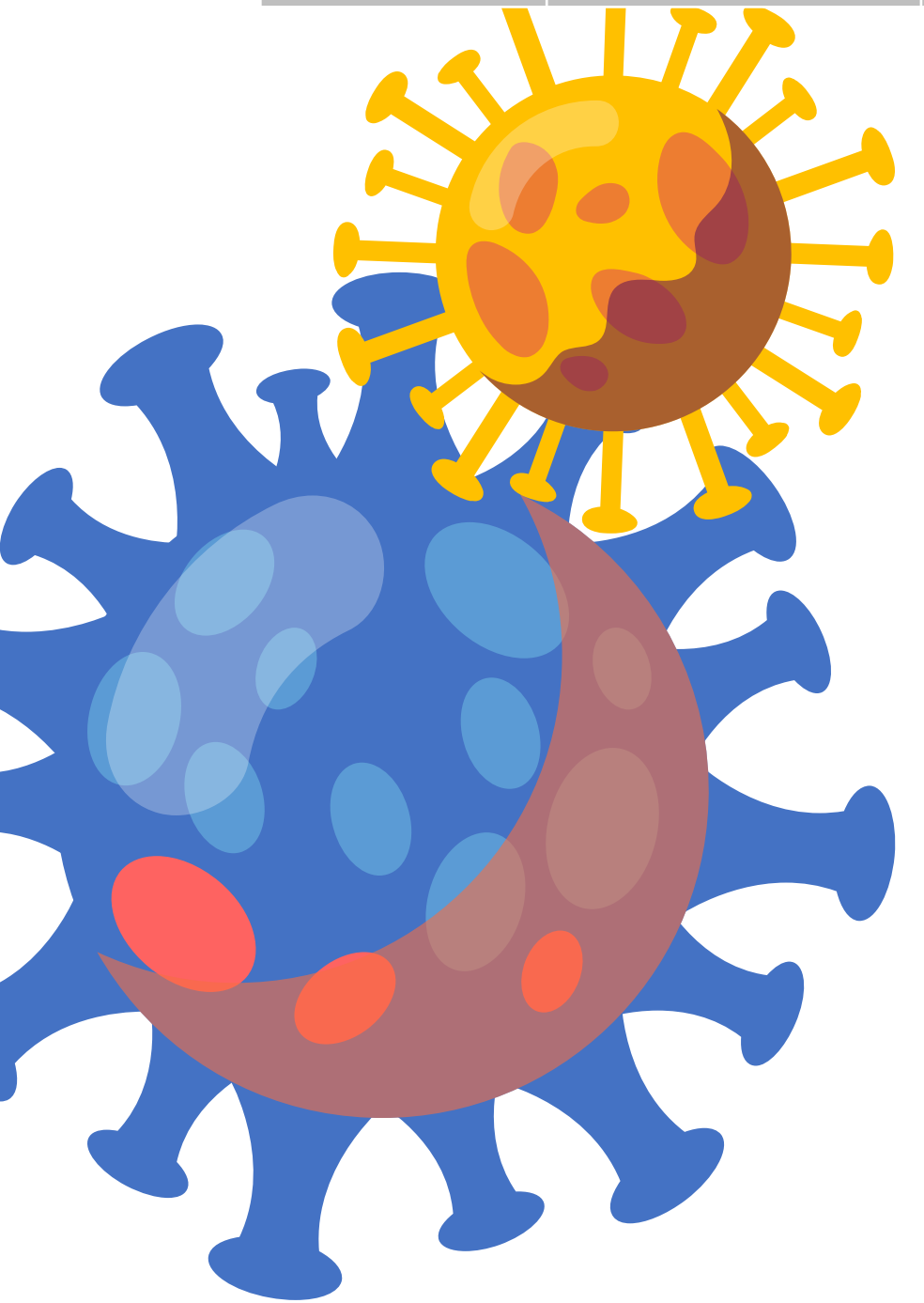
Polysyllabic sound



Ah' sound

	Accuracy	Sensitivity	Specificity	PPV	NPV	Kappa
A						
Fold 1	0.77	1.00	0.44	0.72	1.00	0.48
Fold 2	0.82	0.96	0.57	0.80	0.89	0.58
Fold 3	0.85	0.97	0.50	0.85	0.83	0.54
	0.81 ± 0.04	0.98 ± 0.02	0.50 ± 0.07	0.79 ± 0.07	0.91 ± 0.09	0.53 ± 0.05
B						
Fold 1	0.74	0.78	0.69	0.78	0.69	0.47
Fold 2	0.82	0.88	0.71	0.85	0.77	0.60
Fold 3	0.77	0.76	0.80	0.92	0.53	0.48
	0.78 ± 0.03	0.81 ± 0.05	0.73 ± 0.05	0.85 ± 0.05	0.66 ± 0.10	0.52 ± 0.26
C						
Fold 1	0.71	1.00	0.27	0.68	1.00	0.31
Fold 2	0.56	0.72	0.29	0.64	0.36	0.01
Fold 3	0.73	0.82	0.44	0.82	0.44	0.27
	0.67 ± 0.07	0.85 ± 0.12	0.33 ± 0.08	0.71 ± 0.08	0.60 ± 0.28	0.19 ± 0.13

กระทรวงสาธารณสุข
MINISTRY OF PUBLIC HEALTH



An algorithm to classify Twitter Comments on COVID-19 Vaccine Side Effects:

Using social media to assess public opinion on vaccine hesitancy

Suphanat Wongsanuphat

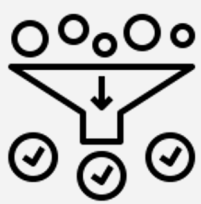
Development of ML: Twitter Mining

1

Data Extraction

Extract data from Twitter API Using keyword ***"fever"***

2



Preprocessing

- Manually **labeled** data (Related/not related to VSE).

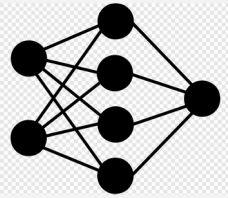
- Preprocessing data by removing hashtag, url and stopword.

3

Feature Extraction

Including; Bag of word, Part of Speech tag (POS), and Embedded "nnlm-en-dim128"

4



Classification

Including; **Supervised ML** (Naive Bayes, Decision Tree, SVM, Regression) and Feed Forward **Neural Network model** (FFNN)

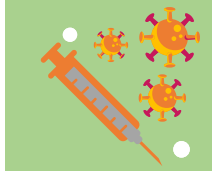
5

Evaluation

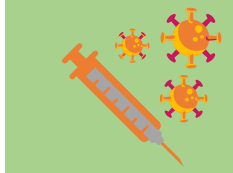
Performed 5-fold cross **validation** and measure precision, recall and F-score (Macro vs Micro average)

Development of ML: Twitter Mining

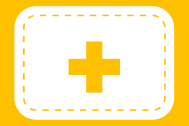
@TeddyPersians ***Pfizer***!! Fully ***vaccinated second shot*** had ***me*** down for a day with **fever** and body aches but I felt like a million bucks



Yup second ***Moderna*** yesterday. ***Feeling*** medium crappy, aches and pains body wide, sore arm, and **fever.** Imma live but ugg! https://t.co/nNuurVJrYK

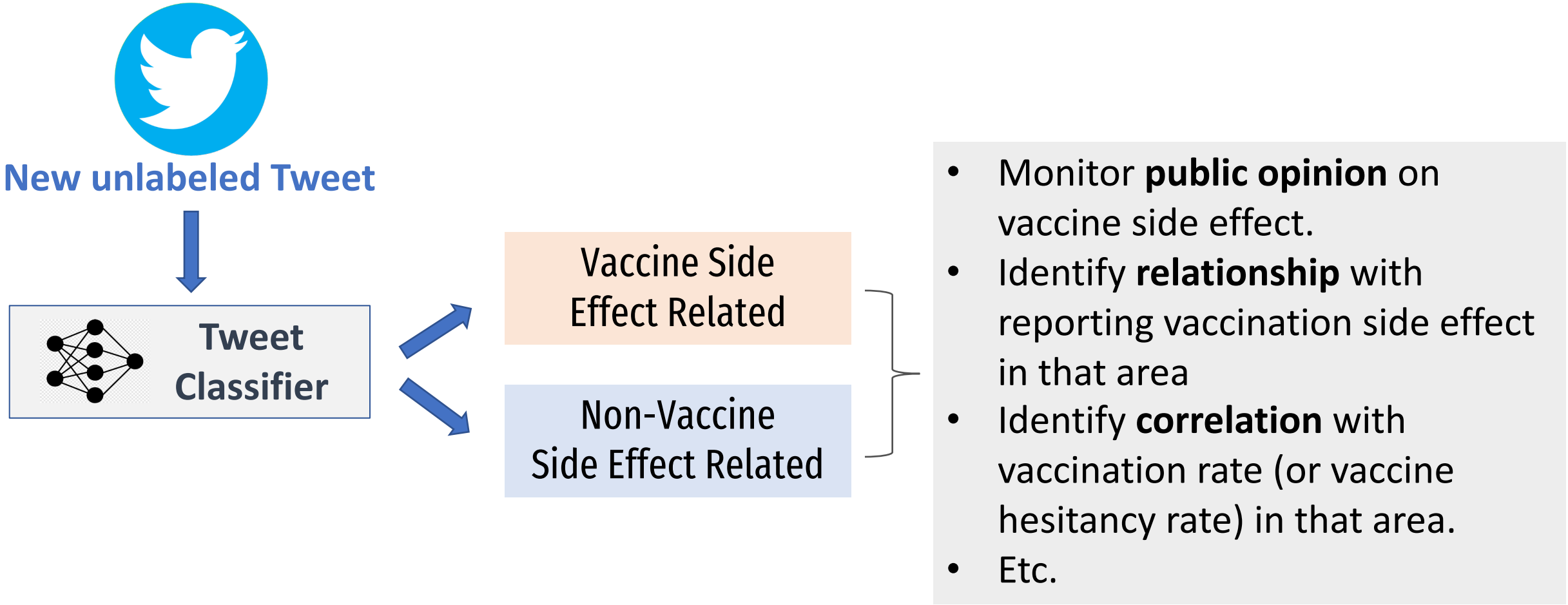


@TheIndy400 Nothing. However, a ***few friend got*** a **fever** the night after, then they were fine upon waking up next day.



@BeeGees @KassaiKarim My life in Hawaii, it was when the "Saturday Night **Fever** album was released. @***BeeGees***

Development of ML: Twitter Mining



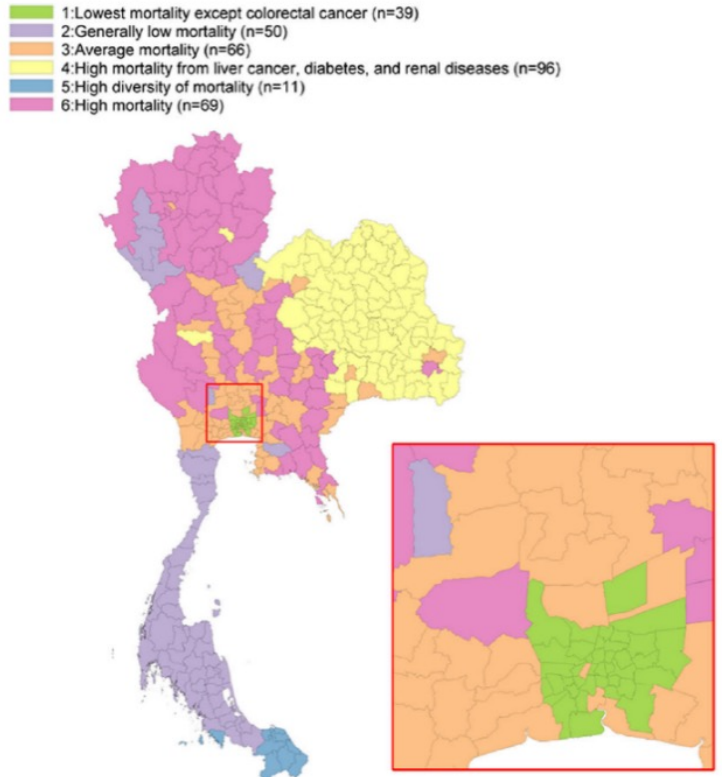
- Monitor **public opinion** on vaccine side effect.
- Identify **relationship** with reporting vaccination side effect in that area
- Identify **correlation** with vaccination rate (or vaccine hesitancy rate) in that area.
- Etc.

กระทรวงสาธารณสุข
MINISTRY OF PUBLIC HEALTH

Unsupervised ML: Cluster Analysis

Area-level socioeconomic deprivation and mortality differentials in Thailand: results from principal component analysis and cluster analysis

Suchunya Aungkulanon[1,2], Viroj Tangcharoensathien[2], Kenji Shibuya[3], Kanitta Bundhamcharoen[2] and Virasakdi Chongsuvivatwong[1*]



Mortality cluster

Disease	1:Lowest mortality except colorectal cancer	2:Generally low mortality	3:Average mortality	4:High mortality from liver cancer, diabetes, and renal diseases	5:High diversity of mortality	6:High mortality
Traffic injury	-1.314	0.301	0.01	-0.144	-0.236	0.231
Liver and intrahepatic bile ducts cancer	-0.48	-1.007	-0.421	0.529	-1.762	-0.138
Assualt	-1.057	0.75	-0.288	-0.862	1.65	-0.166
Pneumonia	0.232	-0.35	0.271	-0.526	-0.757	0.117
Self harm	-1.374	0.031	-0.171	-0.337	-1.363	0.474
Renal diseases	-0.595	-0.498	-0.343	0.405	-0.842	-0.058
Unspecified	-0.533	-0.036	-0.084	0.094	0.286	0.063
Sepsis	0.149	-0.329	0.185	-0.355	-0.531	0.117
Ischaemic heart diseases	0.215	-0.003	0.183	-0.543	0.009	-0.06
Liver diseases	-0.322	-0.46	0.105	-0.015	-1.281	0.159
Diabetes	-0.517	-0.751	-0.592	0.502	0.231	-0.531
Cerebrovascular diseases	0.025	-0.268	0.171	-0.307	-0.55	0.15
Colon and rectal cancer	0.546	-0.365	0.296	-0.655	-0.815	-0.333
Chronic lower respiratory diseases	-0.48	0.022	-0.292	-0.507	0.742	0.243
Drowning	-0.682	-0.49	0.145	-0.023	-0.602	0.163
Remainder of diseases of the digestive system	-0.352	-0.448	0.036	-0.125	-0.607	0.249
Lung cancer	0.188	-0.298	-0.068	-0.197	-1.095	0.099
Other external causes	-0.061	-0.324	0.148	-0.101	-0.655	0.065
Remainder of diseases of the nervous system	-0.692	-0.121	-0.193	-0.204	0.54	0.171
Hypertensive diseases	-0.327	-0.704	-0.37	-0.475	1.217	0.02
Remainder of diseases of the respiratory system	-0.468	-0.148	0.007	-0.02	-0.372	0.114
Human immunodeficiency virus [HIV] disease	-0.352	-0.004	-0.019	-0.393	-0.432	0.208
Remainder of certain infectious and parasitic diseases	-0.193	-0.394	-0.099	-0.082	-0.909	0.135
Lip, oral cavity and pharynx cancer	0.047	-0.191	0.243	-0.301	-0.745	-0.152
Tuberculosis	-0.314	-0.485	0.052	-0.016	0.036	-0.038
Remainder of diseases of the genitourinary system	-0.155	-0.25	-0.192	-0.372	0.558	0.14
Certain conditions originating in the perinatal period	-0.167	-0.307	0.092	-0.486	0.309	-0.487
Cancer of meninges, brain and other parts of central nervous system	-0.312	-0.091	-0.291	0.079	-0.792	-0.075
Remainder of malignant neoplasms	-0.069	-0.072	-0.141	0.066	-0.446	-0.047
Remainder of diseases of the circulatory system	-0.1	-0.216	-0.123	-0.088	0.415	-0.005
Leukemia	0.037	-0.157	-0.046	-0.122	-0.424	-0.139

-1 0 1

Disease specific average LogSMR

กระทรวงสาธารณสุข
MINISTRY OF PUBLIC HEALTH

แหล่งข่าว Online

Google Alerts

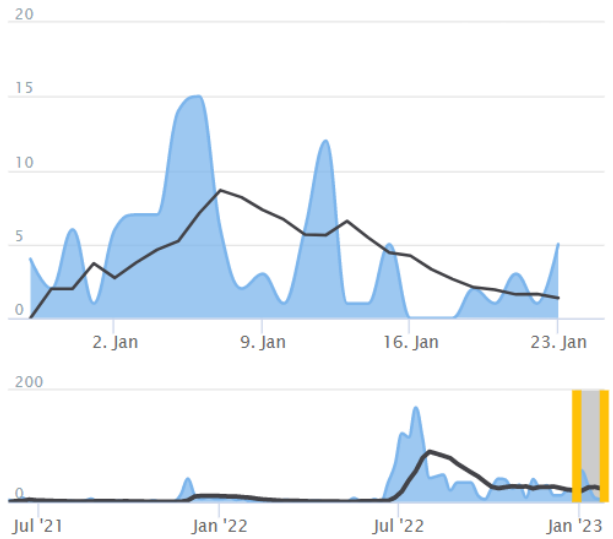
EIOS EPIDEMIC INTELLIGENCE FROM OPEN SOURCES

Monitoring | Documents | Dashboards | Training | Help and Feedback | Suphanat Wongsanuphat Logout

Search text within articles...

TIME PERIOD

Period: (Import Date) Last 28 Days



Last - 28 + Days until

now

From: 28/12/2022 00:00

Board: COVID Modified

Categories: Coronavirus Variant Delta, Coronavirus Variant Omicron Mentioned countries/territories: Thailand Filter definition

111 TOTAL ARTICLES

1 of 3

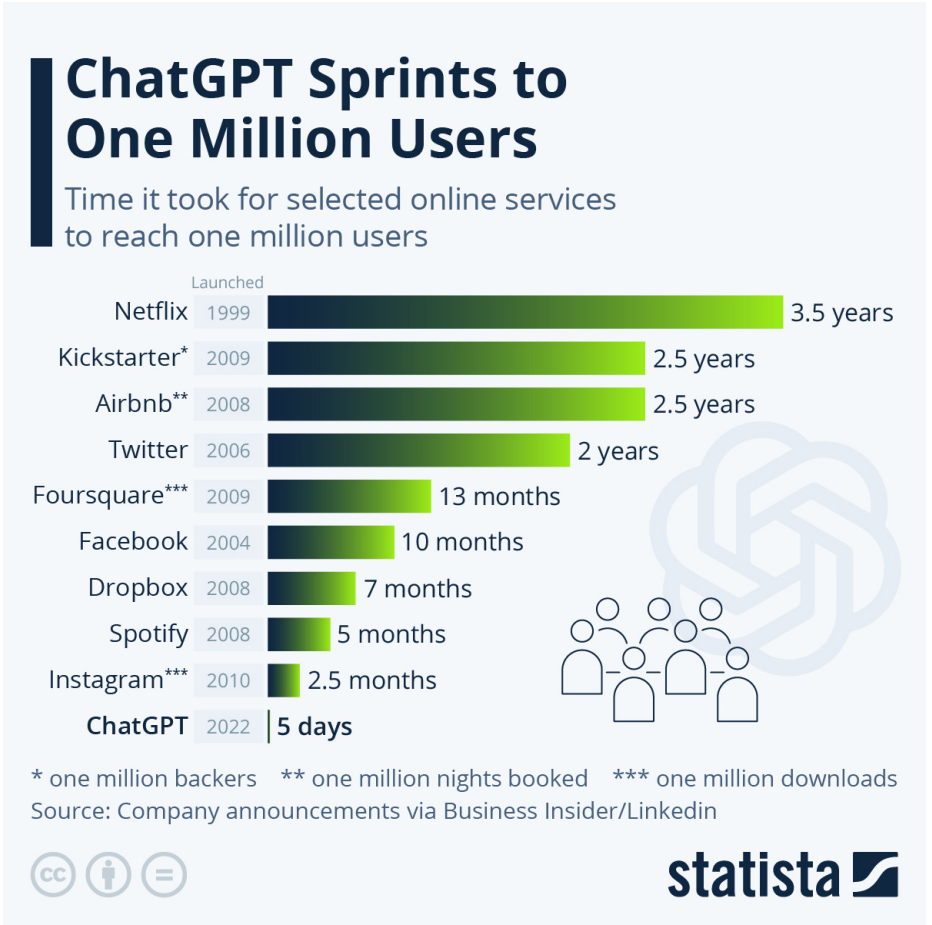
Any article

Import Date

Facebook-sorrayuth91... 23 Jan 21:25 UTC Thai	กรมวิทยาศาสตร์การแพทย์ เผยตรวจ RT-PCR ก่อนขึ้นเครื่อง พบโควิดกว่า 300 ราย เป็นคนจีน 10 กว่าคน หรือ	Coronavirus Variant Omicron	Thailand
hfocus 23 Jan 11:26 UTC Thai	สถานการณ์โควิด-19 หลังไทยรับ นทท.จีน ไม่น่าห่วง! ย้ำ! ยังไม่พบสายพันธุ์ XBB.1.5 ในประเทศ	Coronavirus Variant Omicron	Thailand
matichon 23 Jan 10:09 UTC Thai	สธ.สุ่มตรวจโควิดนักท่องเที่ยวเข้าไทยเจอ 5 ราย ขาออกผลบวกกว่า 400 คน	Coronavirus Variant Omicron	Thailand
bangkokbiznews 23 Jan 09:25 UTC Thai	"โควิด-19"อัตราการติดเชื้อเป็นไปตามคาด หลังไทยรับจีนเปิดประเทศ 23 ม.ค. 2566 | 14:35	Coronavirus Variant Omicron	Thailand
manager-th 23 Jan 05:53 UTC Thai	"หมอธีระ"คาดโควิดสายพันธุ์ XBB.1.5 จะเป็นสายพันธุ์หลักในสหรัฐฯ-ยุโรปภายใน ก.พ.นี้	Coronavirus Variant Omicron	Thailand
bangkokbiznews	สาธารณสุข ศูนย์จีโนมฯ เผยโควิด-19 ที่ระบาดใน	Coronavirus Variant	Thailand

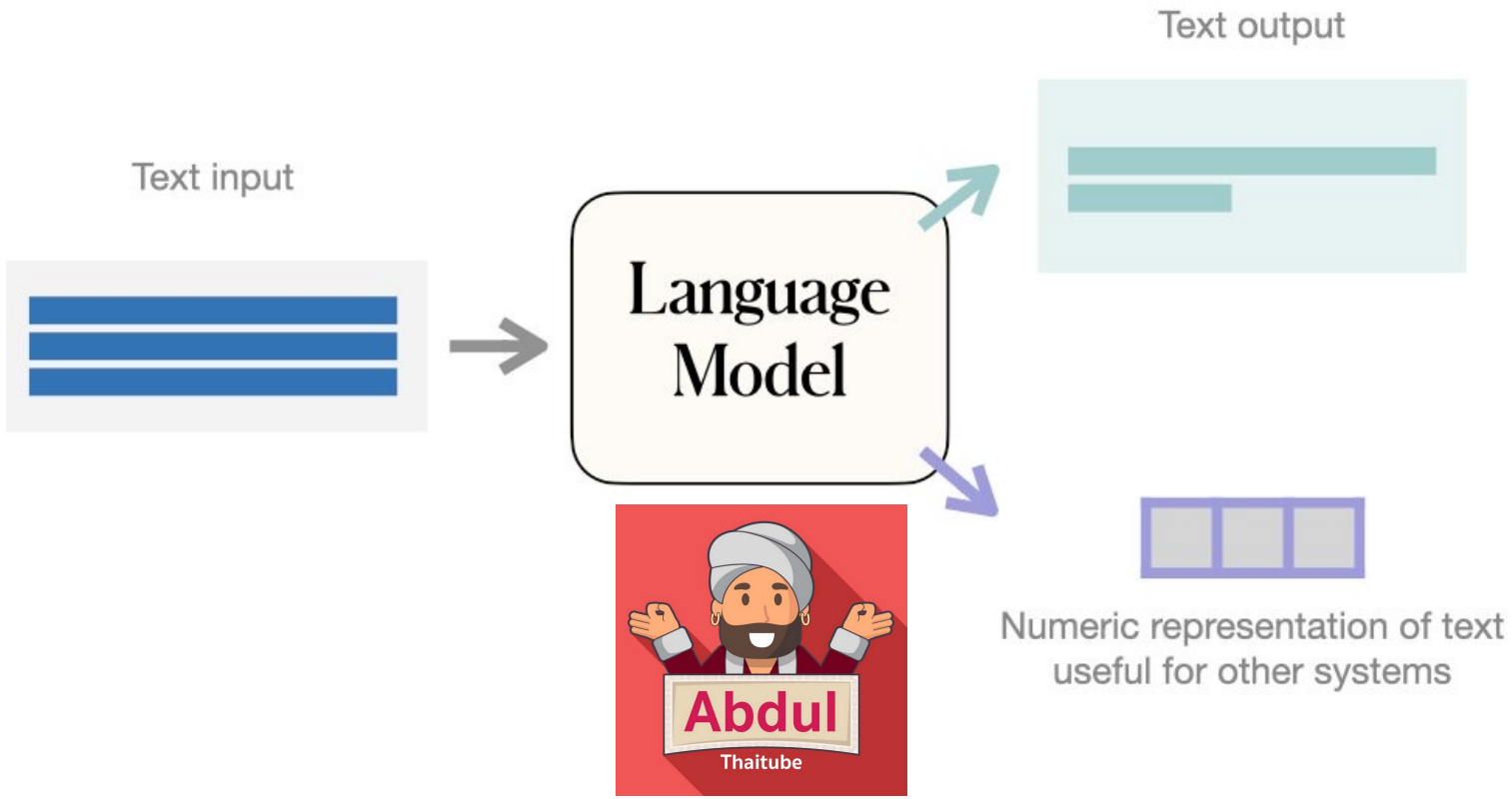
ChatGPT

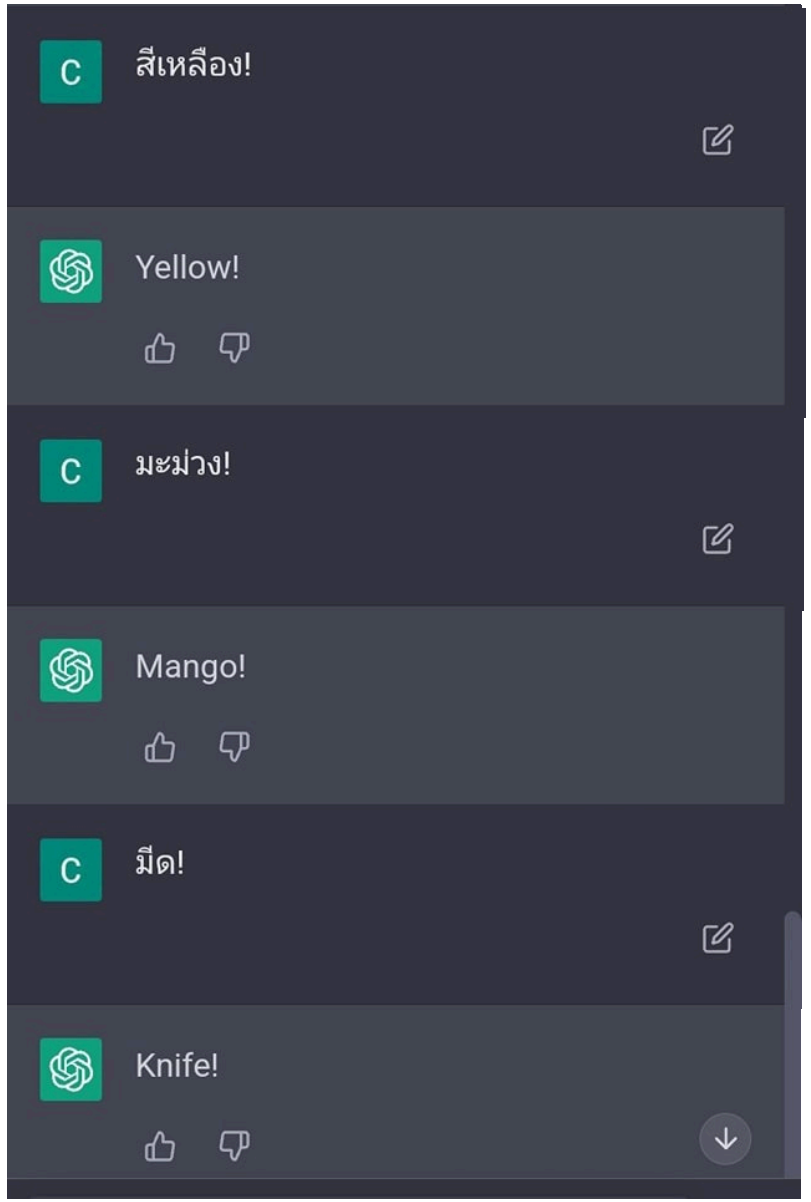
- **GPT** = **Generative** Pre-trained Transformer
- **Generative** AI (**“สร้างใหม่”**)
- **Large Language Model**



https://www.atriainnovation.com/en/how-does-chat-gpt-work/

Large Language Model





https://www.atriainnovation.com/en/how-does-chat-gpt-work/

Chat GPT for Public Health Officer and Epidemiologist

- **Providing information**: ***I can*** answer questions and provide information on various topics related to epidemiology, such as infectious disease outbreaks, transmission dynamics, risk factors, preventive measures, and more.
- **Data analysis**: ***I can*** assist with data analysis tasks, such as statistical analyses and machine learning algorithms, to help epidemiologists make sense of complex datasets.
- **Literature review: *I can*** help epidemiologists stay up-to-date with the latest research and publications in their field by summarizing key findings.
- **Improving Writing: *I can*** check your grammar and expand your vocabulary
- **Modeling**: ***I can*** help epidemiologists build and test mathematical models of disease transmission to inform public health policy decisions.
- **Educational purposes**: ***I can*** also be used for educational purposes, such as teaching and training students and researchers in epidemiology and related fields.

Source: ChatGPT

Data Engineering

iNCD – DOE - DDC

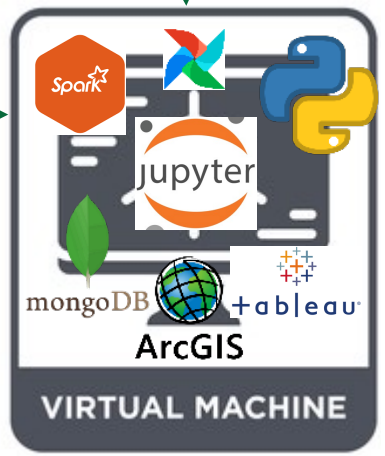
D506 Data Lake

Outbreak Verification
43 Files, Lab, Vaccine

API

API

Digital 506



API API API API

ข้อมูลการสำรวจ

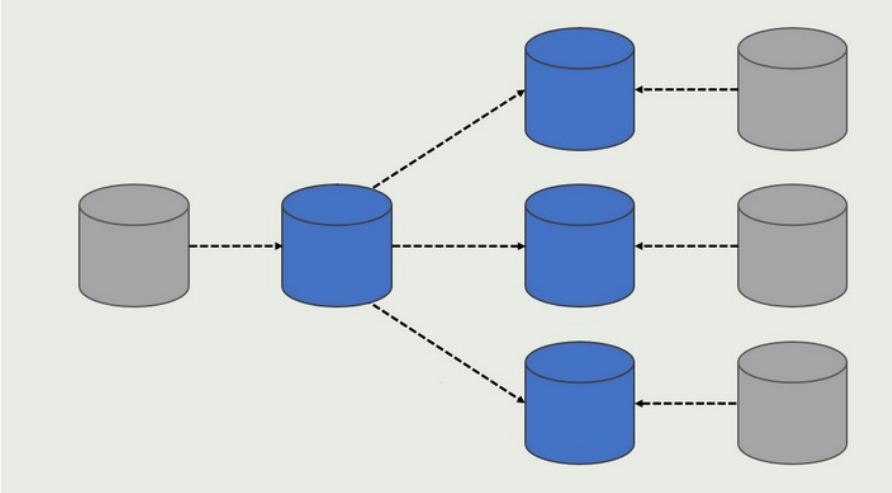
กรมอุตุนิยมวิทยา
อุณหภูมิ น้ำฝน ความชื้น

กรมควบคุมมลพิษ
มลภาวะทางสิ่งแวดล้อม

กรมปศุสัตว์
โรคในสัตว์

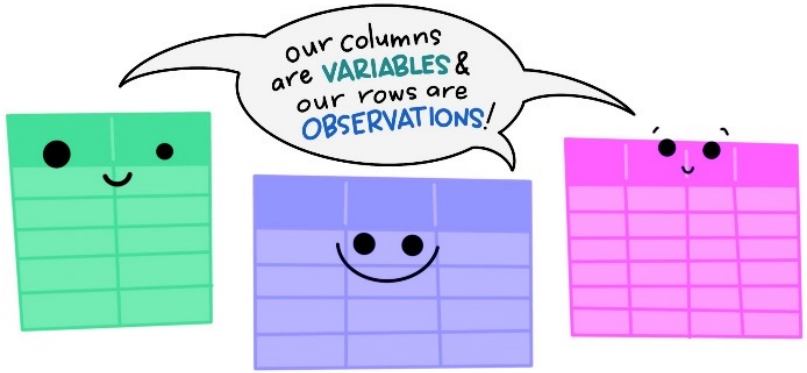
Real World Data is Dirty!

Multiple Databases

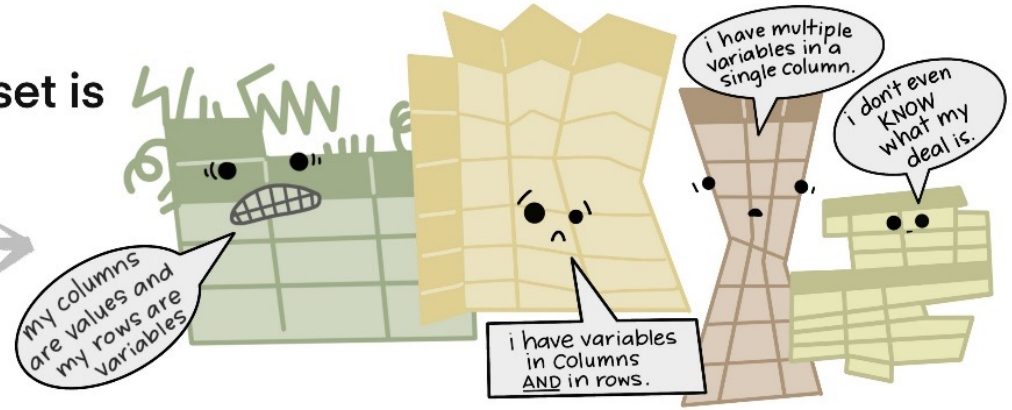


Dirty Data

The standard structure of tidy data means that "tidy datasets are all alike..."



"...but every messy dataset is messy in its own way."
–HADLEY WICKHAM



Data Engineer (and Data Cleansing) to rescuer!

Dirty Data: Lab result (No Data Standardization

- Detected
- Not Detected

LOINC®
Logical Observation Identifiers Names and Codes

- Not
- Neg
- ดูผลในหน้า Image
- อยู่ใน Image
- Inconclusive
- ผลตรวจฯมาแล้วในImage
- ผลการตรวจรูปภาพ
- colab
- vtm
- ผลมาแล้ว
- Inconclu
- อ่านผลใน Lab Image
- รอผล
- See lab results at Image mode
- ส่ง ซีบี เมดดิคอล แอนด์ แล็บ สหคลินิก
- Nagative
- ส่งผลให้ SRRT แล้ว
- ตรวจสอบแล้ว
- Out lab
- PCR for COVID-19
- 03102022
- รางงานผลแล้ว
- รายงานผลแล้ว
- รายงายผลแล้ว
- รายงาสนผลแล้ว
- รางงานผลแล้ว
- รายงานผลแล้ว
- รายงายผลแล้ว
- รายงาสนผลแล้ว
- ดูในImage
- ดูในImage
- ทำที่ER
- ^see image
- PDF
- ผลใน Image

Dirty Data

หมายเหตุ Dead from COVID-19 เสีย ชีวิตวันที่ 05/02/2023 (DCM, CKD State 3, HT, DM, DLP, Stroke) (แถลง

หมายเหตุ วันที่เสียชีวิต 30 ธันวาคม 2565 Died from COVID-19

หมายเหตุ เสียชีวิต 15 ม.ค. 2566 Die from COVID-19

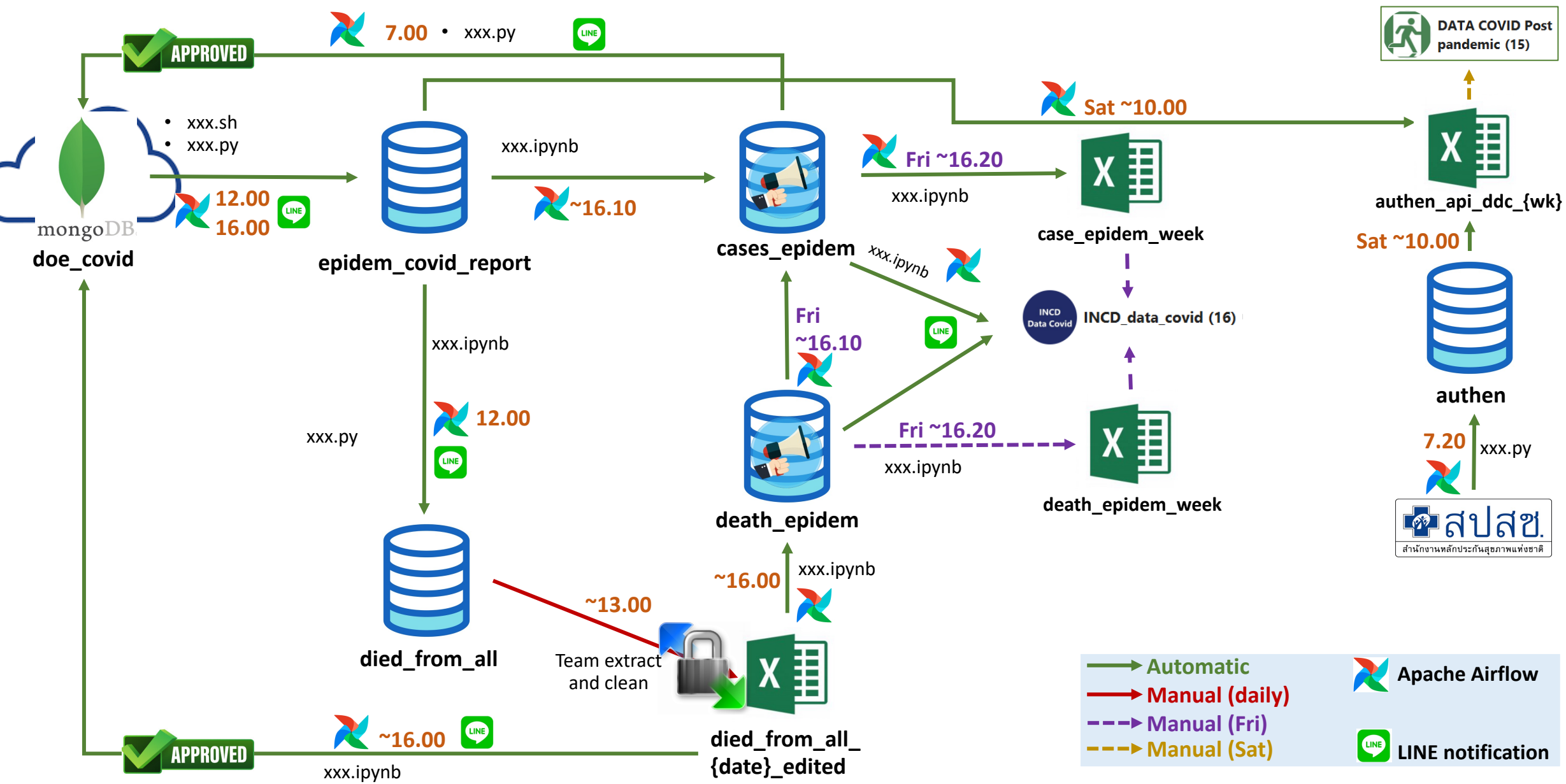
หมายเหตุ Died from COVID-19 เสียชีวิต วันที่ 5/01/2023 (COVID-19 Pneumonia with Chronic Obstructive Pulmonary

หมายเหตุ เสียชีวิตวันที่ 28/12/2022 died from covid-19 with pneumonia with NF

Data Transformation

```
months_abbr = ['มค', 'กพ', 'มีค', 'เมย', 'พค', 'มิย', 'กค', 'สค', 'กย', 'ตค', 'พย', 'ธค']
months_abbr = dict(zip(months_abbr, [str(i) for i in range(1, 13)] ))
months = ['มกราคม', 'กุมภาพันธ์', 'มีนาคม', 'เมษายน', 'พฤษภาคม', 'มิถุนายน', 'กรกฎาคม', 'สิงหาคม', 'กันยายน', 'ตุลาคม', 'พฤศจิกายน',
'ธันวาคม']
months = dict(zip(months, [str(i) for i in range(1, 13)] ))
repl_year = {"2565": "2022", "65": "2022","66":"2023", "2566":"2023"}
announced['comment_newdate'] = announced['comment'].str.replace(".", "", regex=False).replace({**months, **months_abbr},
regex=True).replace(repl_year, regex=True)
announced['date_death1'] = announced['comment_newdate'].str.extract("([0-9]+/[0-9]+/[0-9]+)")
announced['date_death1'] = pd.to_datetime(announced['date_death1'], format='%d/%m/%Y', errors='coerce')
announced['date_death2'] = announced['comment_newdate'].str.extract("([0-9]+ [0-9]+ [0-9]+)")
announced['date_death2'] = pd.to_datetime(announced['date_death2'], format='%d %m %Y', errors='coerce')
announced['date_death3'] = announced['comment_newdate'].str.extract("([0-9]+ [0-9]+)")[0]
announced['date_death3'] = pd.to_datetime(announced['date_death3'], format='%d %m%Y', errors='coerce')
announced['date_death4'] = announced['comment_newdate'].str.extract("([0-9]+/[0-9]+/[0-9]+)")
announced['date_death4'] = pd.to_datetime(announced['date_death4'], format='%d/%m/%y', errors='coerce')
announced['date_death'] = np.where(announced['date_death1'].notna(), announced['date_death1'], announced['date_death2'])
announced['date_death'] = np.where(announced['date_death'].isna(), announced['date_death3'], announced['date_death'])
announced['date_death'] = np.where(announced['date_death'].isna(), announced['date_death4'], announced['date_death'])
```

กระทรวงสาธารณสุข
MINISTRY OF PUBLIC HEALTH



Batch Scheduler

Airflow DAGs Datasets Security Browse Admin Docs 12:22 UTC AA

DAGs

All 52 | Active 8 | Paused 44

Filter DAGs by tag | Search DAGs | Auto-refresh

DAG	Owner	Runs	Schedule	Last Run	Next Run	Rece
api_authen_saturday	airflow	8 1	00 0 * * 6	2023-02-25, 00:00:00	2023-03-04, 00:00:00	
authen	airflow	49	00 0 * * 1-4,6,7	2023-03-02, 00:00:00	2023-03-04, 00:00:00	
case_approve	airflow	67 3	20 0 * * *	2023-03-03, 00:20:00	2023-03-04, 00:20:00	
covid_etl_1	airflow	52 6	00 9 * * 1-4,6,7	2023-03-02, 09:00:00	2023-03-04, 09:00:00	
covid_etl_dead	airflow	56 7	00 5 * * *	2023-03-03, 05:00:00	2023-03-04, 05:00:00	
coward	airflow	32 8	00 0 * * *	2023-03-03, 00:00:00	2023-03-04, 00:00:00	

Sort by asc next_dagrun.

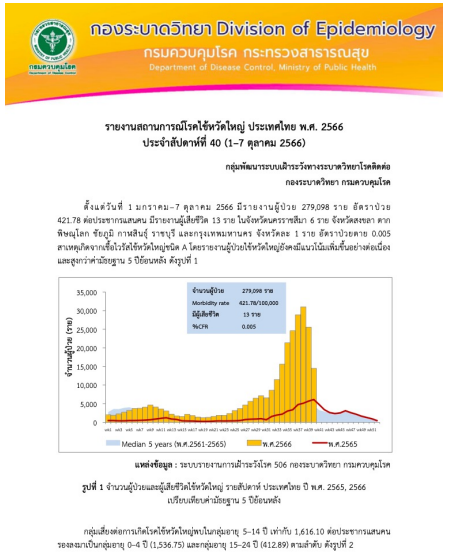
Weekly Disease Situation Report

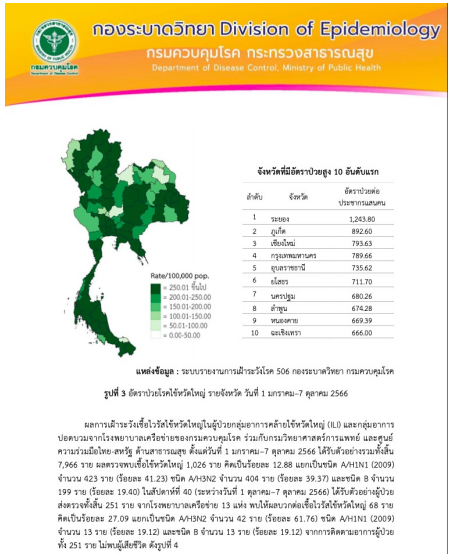
R506

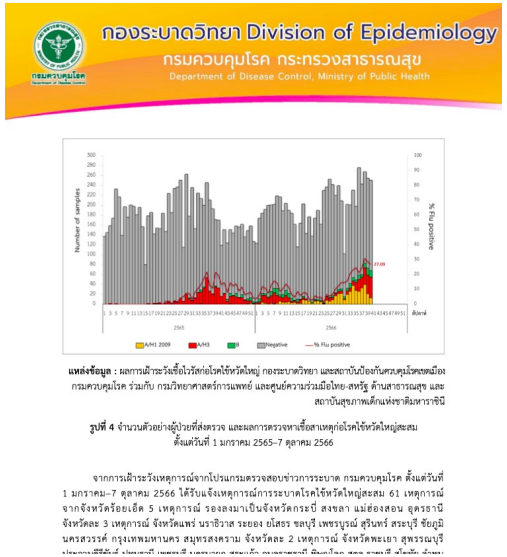
Lab

Event Based

~10-12 hrs/wk







~10-12 hrs *52 wks = 22 - 24 days/year

กระทรวงสาธารณสุข
MINISTRY OF PUBLIC HEALTH

Digitalization

R506

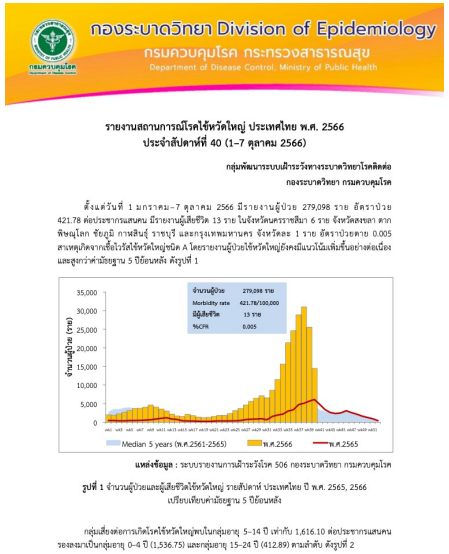
Lab

Event Based

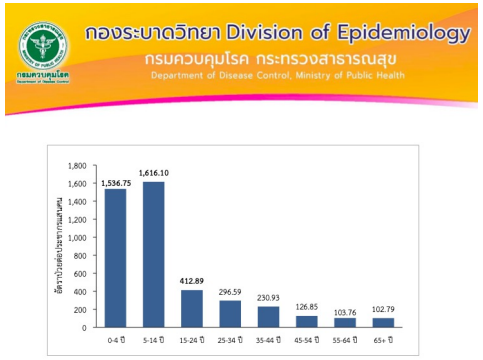
auto

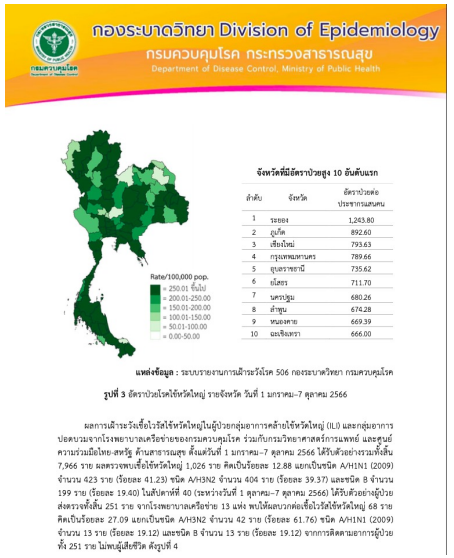
python

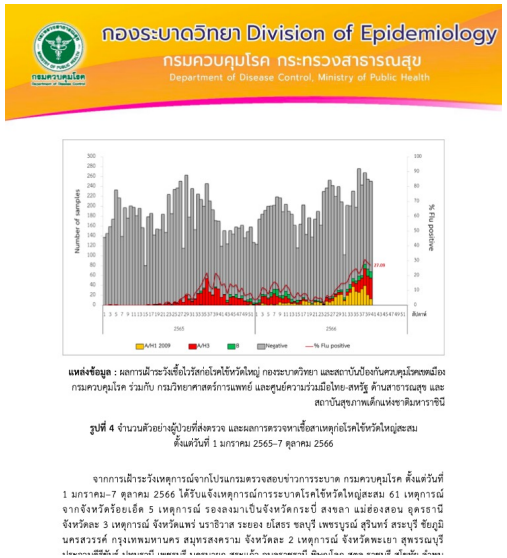
~ 15 min


กองระบาดวิทยา Division of Epidemiology
กรมควบคุมโรค กระทรวงสาธารณสุข
Department of Disease Control, Ministry of Public Health

รายงานสถานการณ์โรคไข้หวัดใหญ่ ประเทศไทย พ.ศ. 2566
ประจำสัปดาห์ที่ 40 (1–7 ตุลาคม 2566)

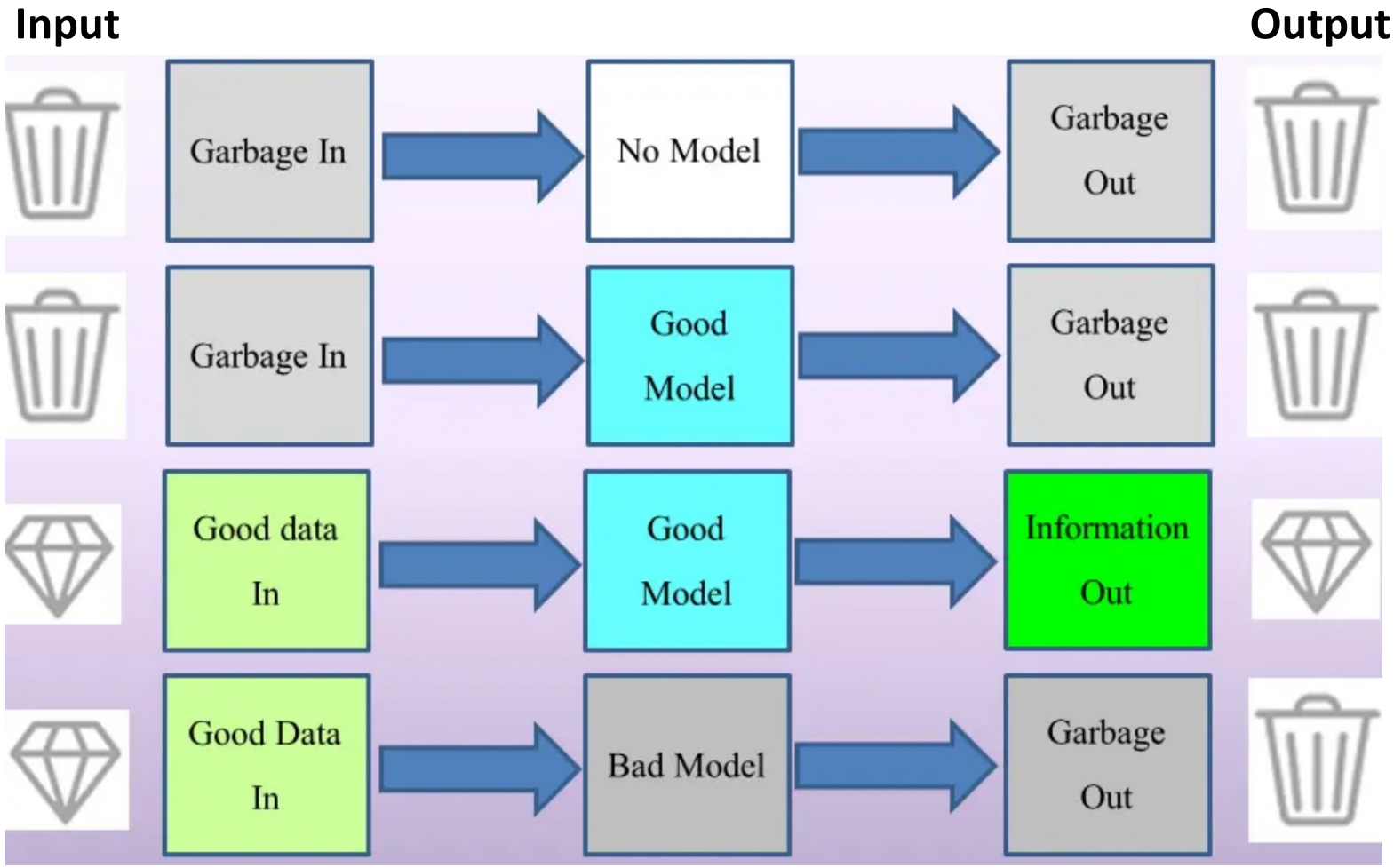

กองระบาดวิทยา Division of Epidemiology
กรมควบคุมโรค กระทรวงสาธารณสุข
Department of Disease Control, Ministry of Public Health


กองระบาดวิทยา Division of Epidemiology
กรมควบคุมโรค กระทรวงสาธารณสุข
Department of Disease Control, Ministry of Public Health

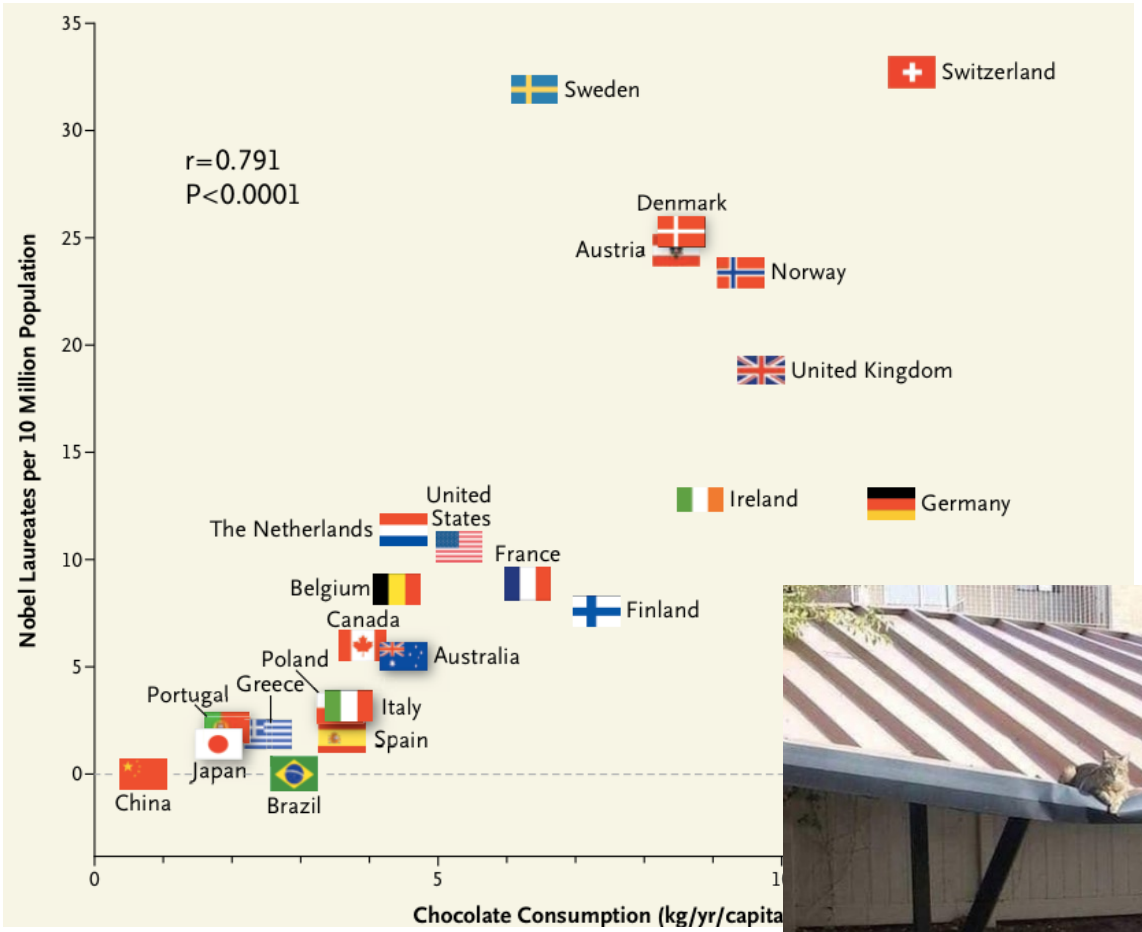

กองระบาดวิทยา Division of Epidemiology
กรมควบคุมโรค กระทรวงสาธารณสุข
Department of Disease Control, Ministry of Public Health

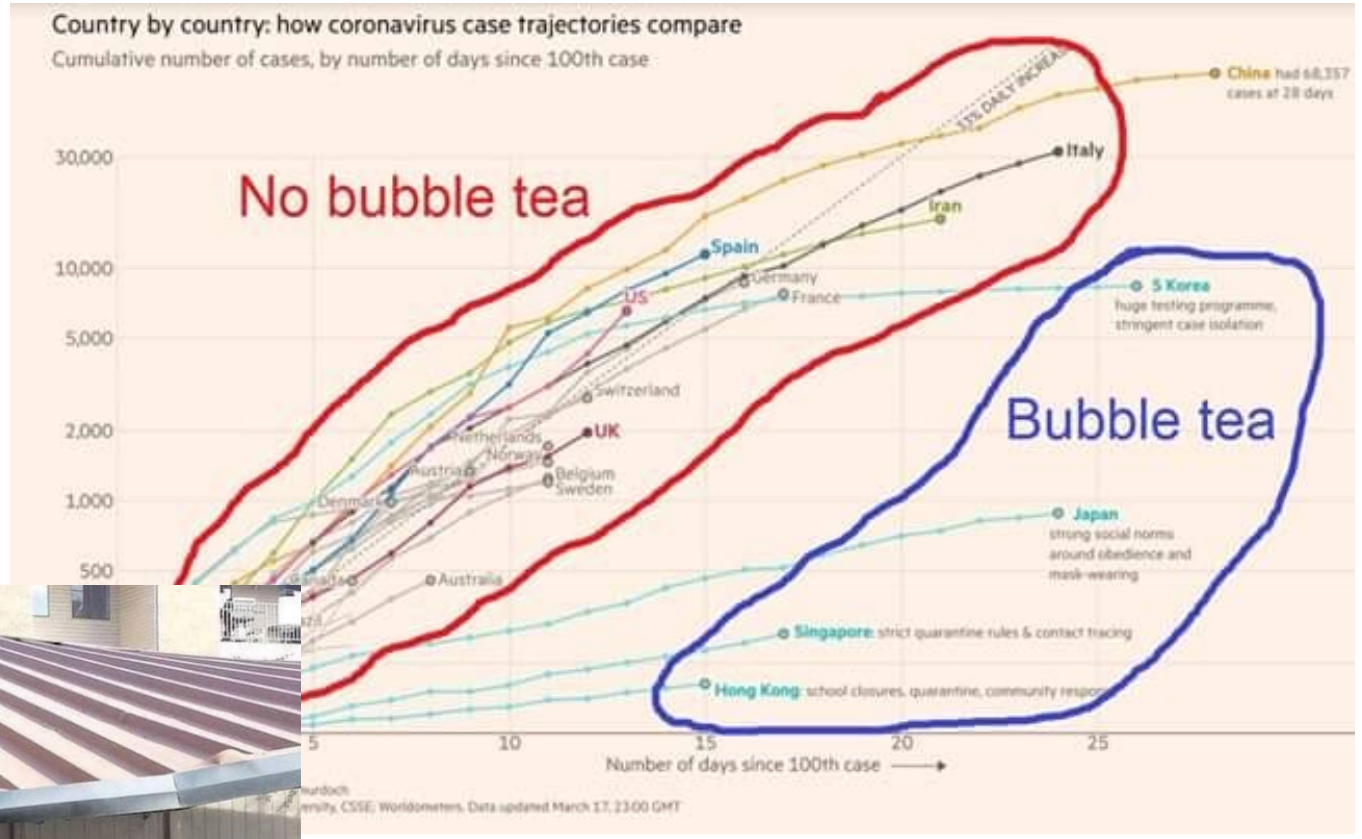
Challenges and Risks

Garbage in Garbage out

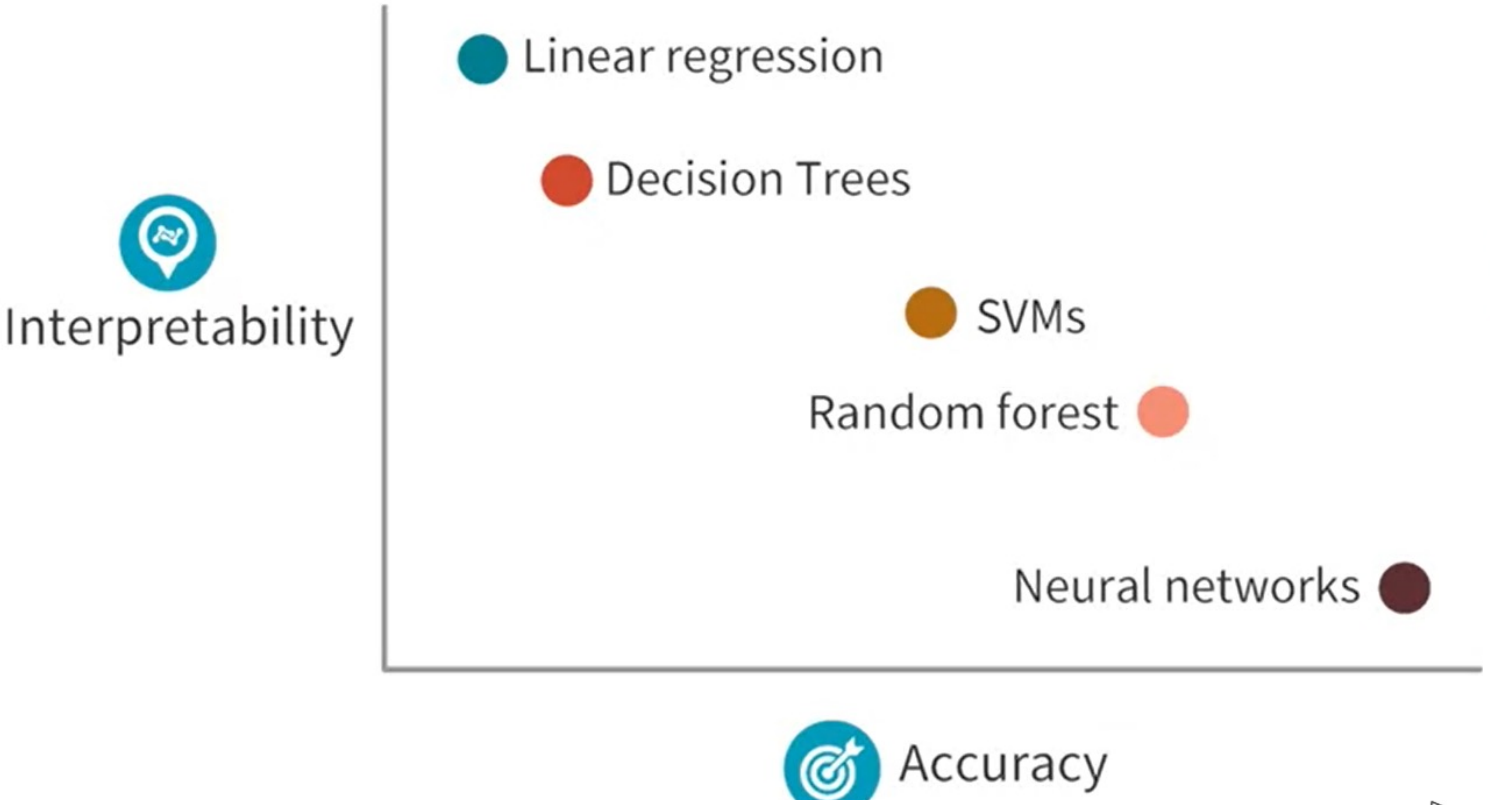


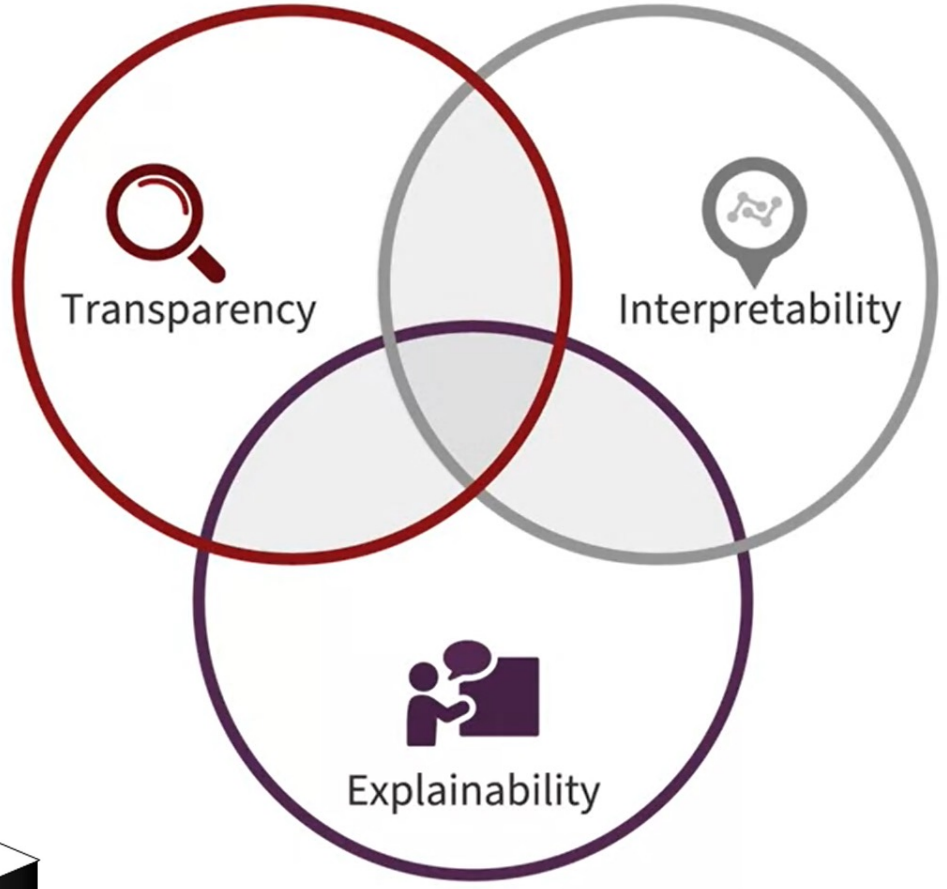
Correlation is not Causation!

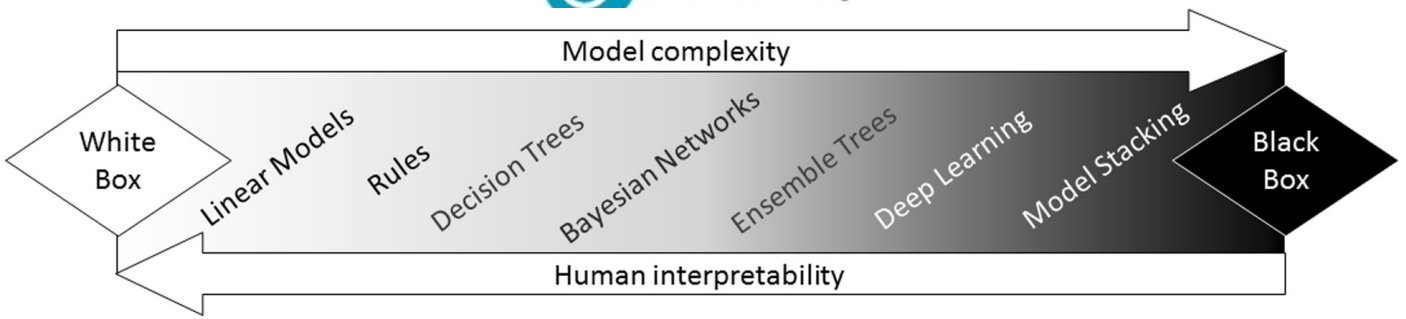


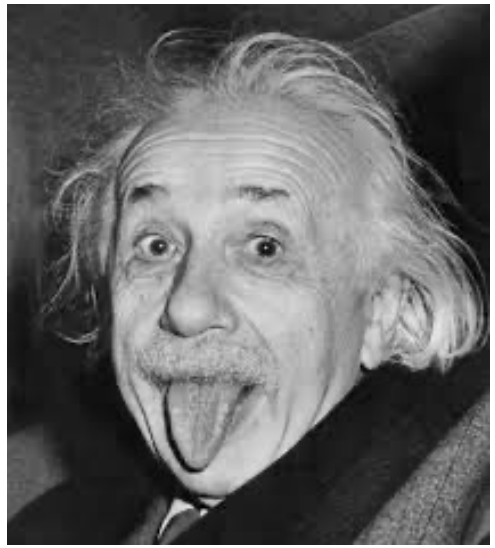


Accuracy vs Interpretability







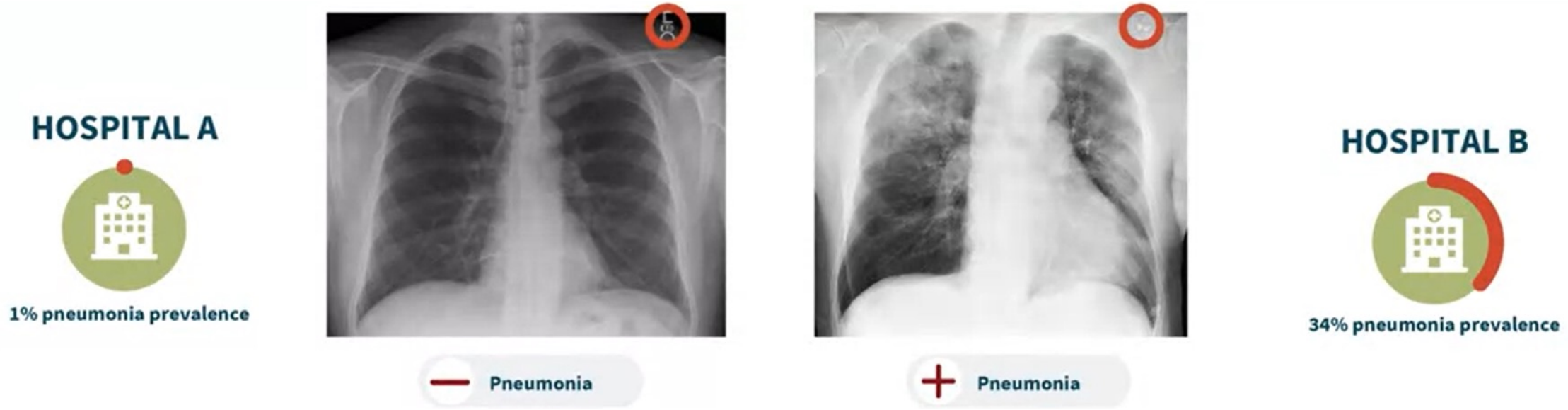


"Everything should be made **as simple as possible**, but **not simpler**."

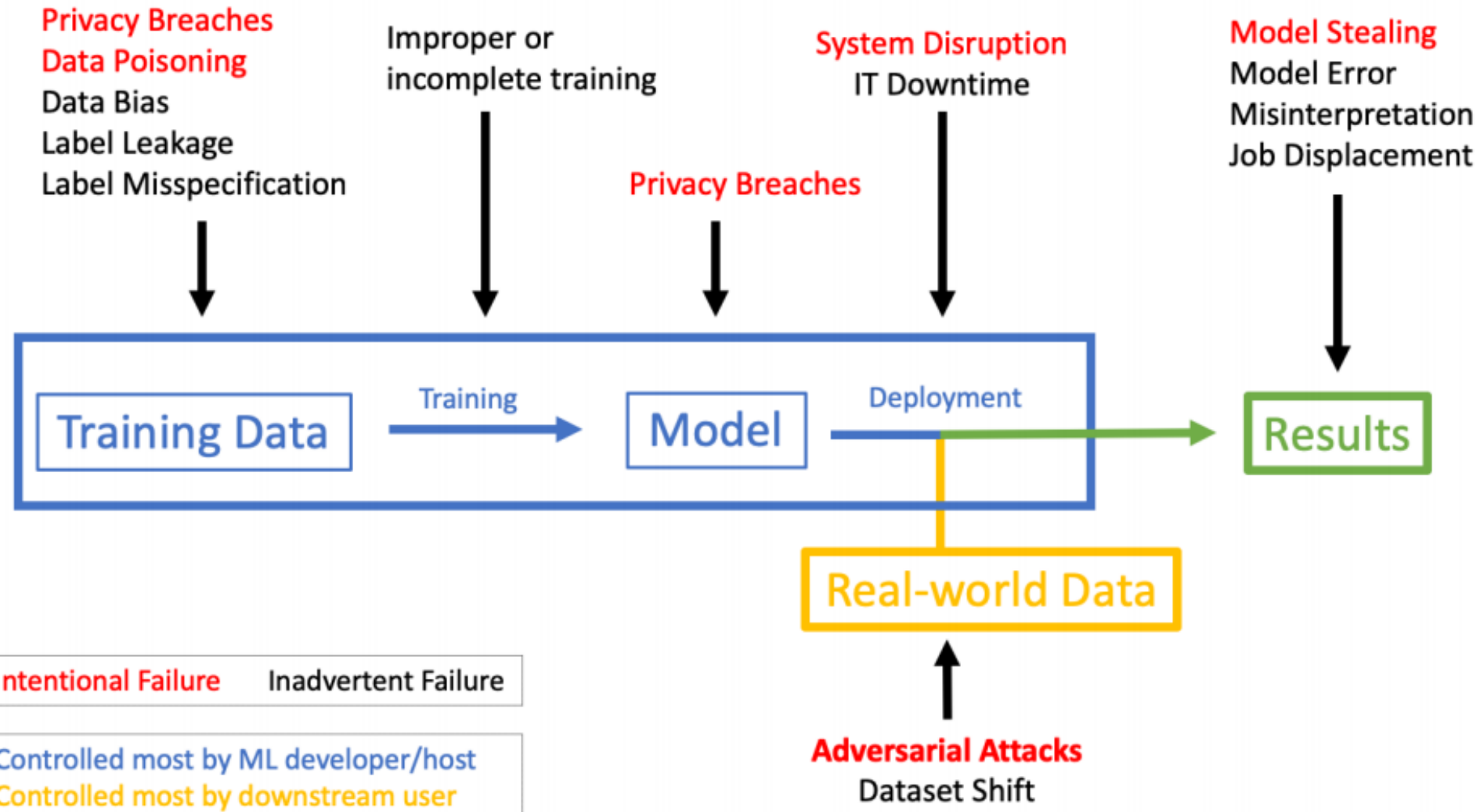
- ***Albert Einstein***

(Parsimonious Model)

Black Box

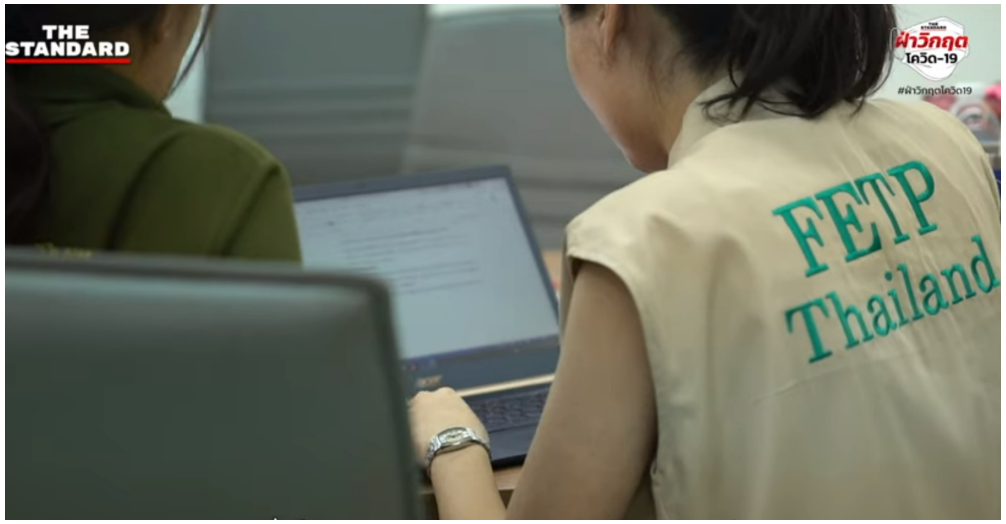

HOSPITAL A
1% pneumonia prevalence
Pneumonia
Pneumonia
HOSPITAL B
34% pneumonia prevalence

Other Challenges and Risks

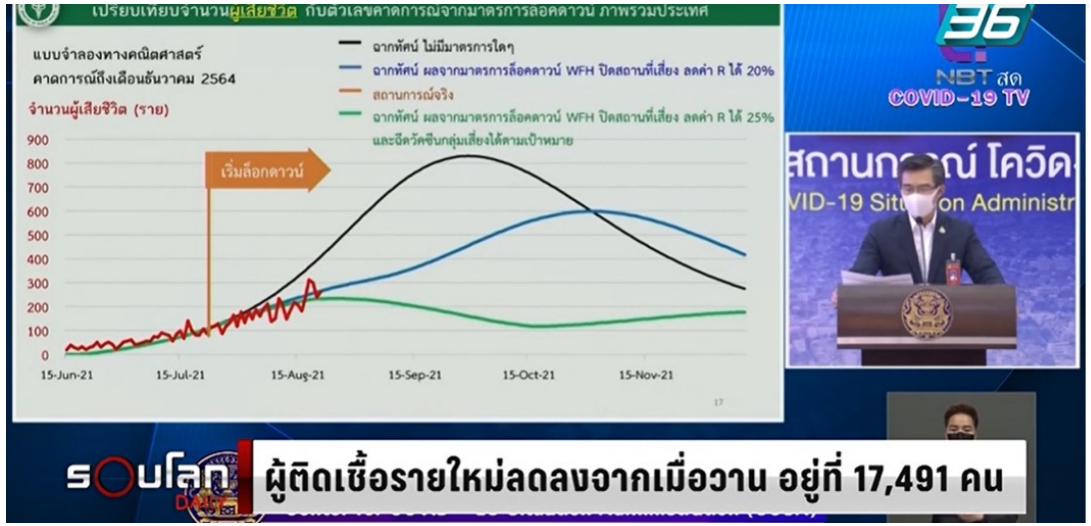


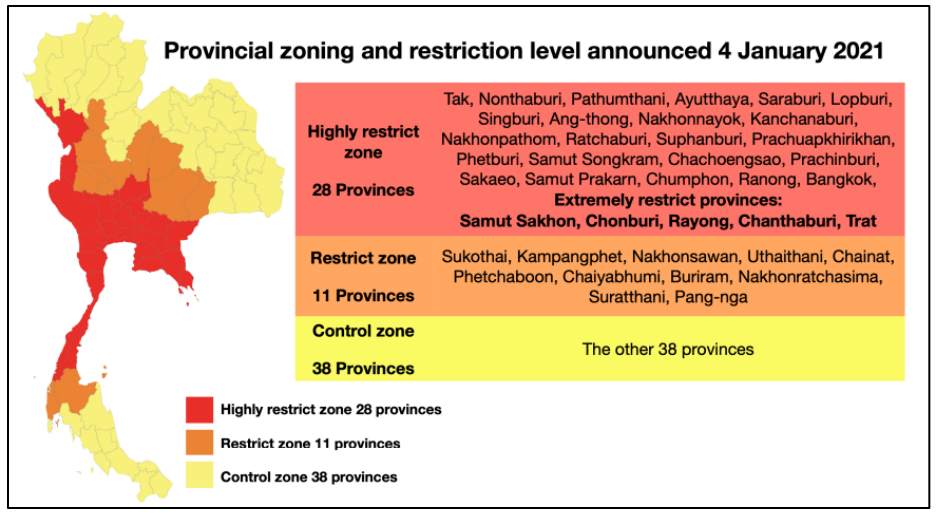
Samuel G. Finlayson, Adversarial Attacks Against Medical Deep Learning Systems

Data to Action



Data to Policy



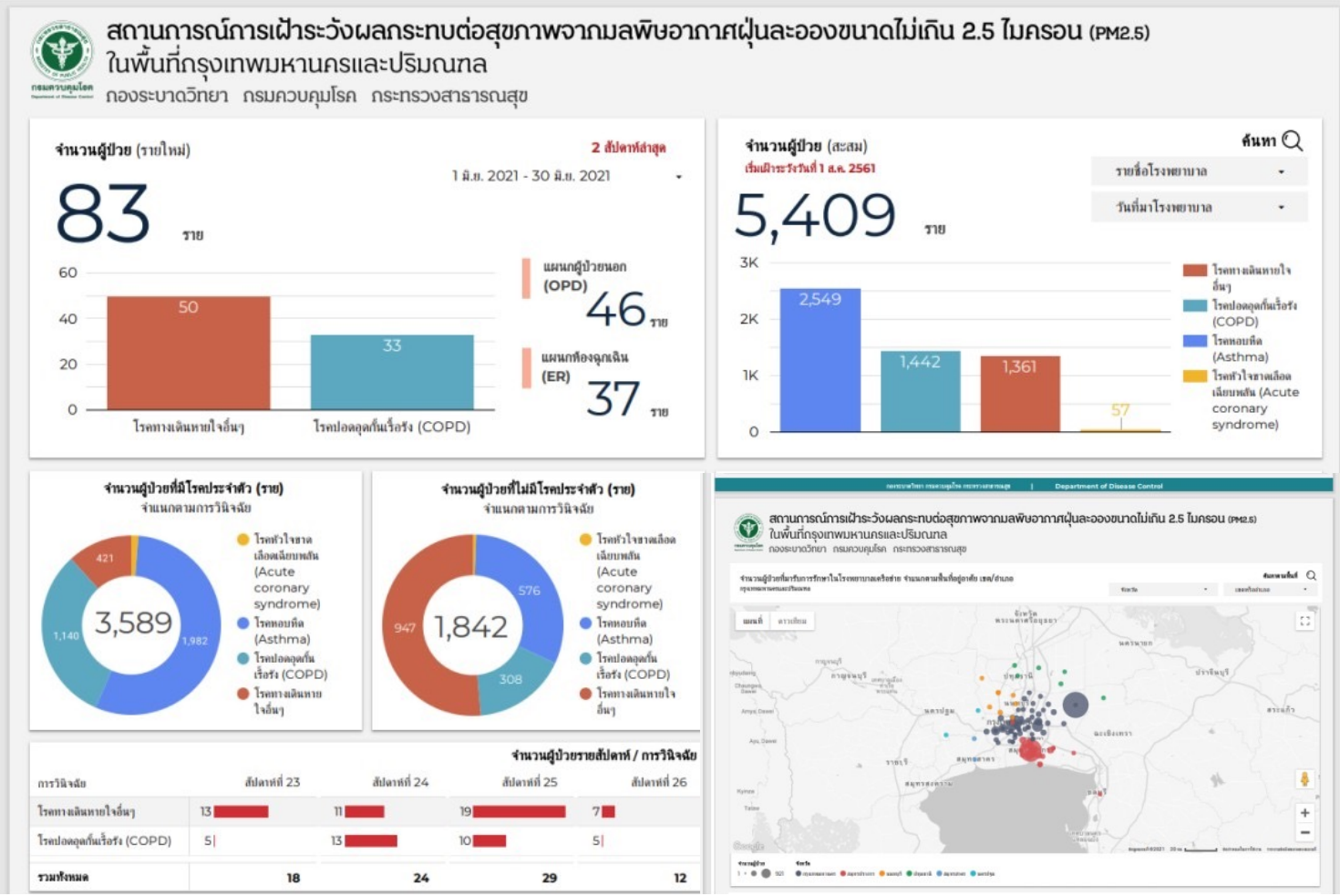


Rajatanavin N, Tuangratananon T, Suphanchaimat R, Tangcharoensathien V. Responding to the COVID-19 second wave in Thailand by diversifying and adapting lessons from the first wave. BMJ Global Health. 2021;6(7).

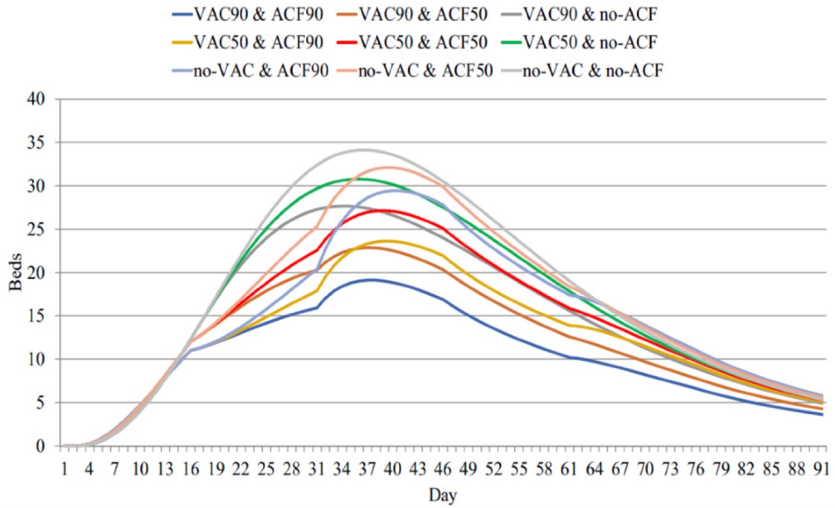
	Ban of working at office	Domestic traveling restriction	Ban of Alcohol selling	Curfew	Closure of: School	Restaurant	Markets/ department store	Meeting/ Seminar	Pub/ Night Club	Massage parlour	Spa/ massage	Fitness/ gym/ stadium	Other services business
The first wave	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Extremely restrict zone (5 provinces)		✓			✓	✓ Close at 21.00/ No alcohol drinking			✓	✓	✓	✓	✓
Highly restrict zone (23 provinces)					On site, Online, On air teaching	✓ Close at 23.00/ No alcohol drinking		< 100 partici pants	✓	✓	Customer Limitation	No audience	
Restrict zone (11 provinces)						✓ Close at 23.00/ No alcohol drinking		< 300 partici pants					
Control zone (38 zones)						✓ Close at 24.00							

Figure 5 Provincial zoning and restriction level announced 4 January 2021 and comparison of social measures used in the first wave and specific zones during the second wave. Five extremely restrict provinces are Samut Sakhon, Chonburi, Rayong, Chanthaburi and Trat.

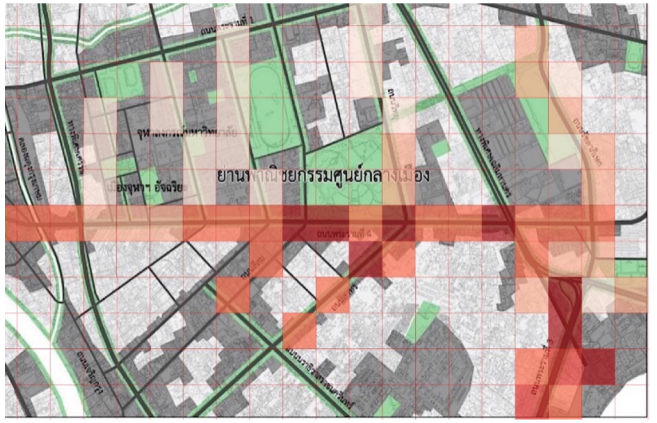
Data for Decision Making & Communication (2024-2026)



Descriptive Analytics

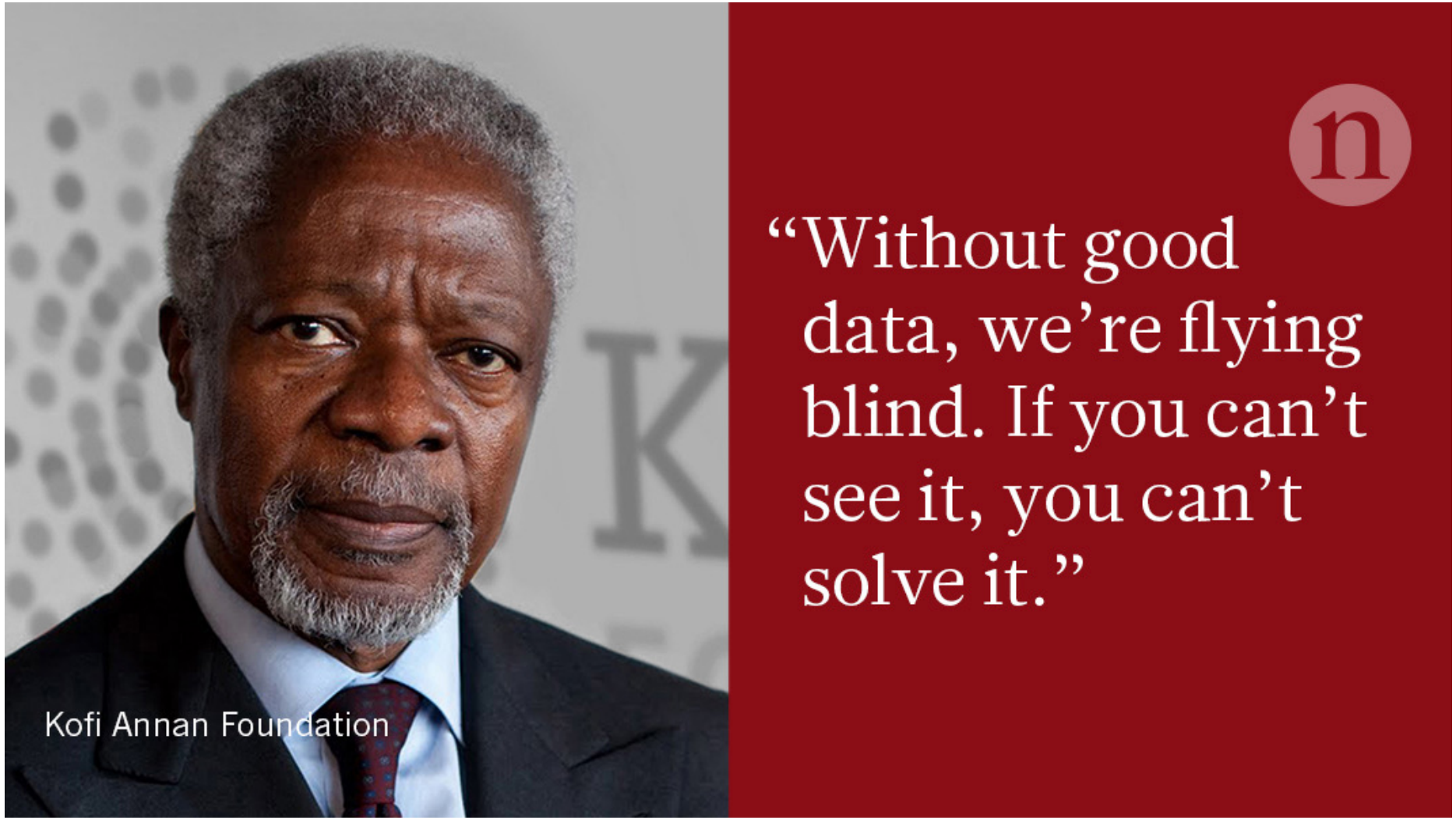


Predictive Analytics



Prescriptive Analytics

Final Thoughts



Thank you