

Finding Data from Secondary Sources

Using existing data carefully

การใช้ข้อมูลที่มีอยู่แล้วอย่างรอบคอบ

Main idea:

Secondary data is useful, but we must check it before using it.

ข้อมูลทุติยภูมิมีประโยชน์ แต่เราต้องตรวจสอบก่อนนำไปใช้

Today we will learn how to:

-  find data
-  check the source
-  understand file types
-  identify attributes
-  notice missing information



1. Find



2. Check Source



3. File Type



4. Attributes



5. Missing Data

Quick Review: Primary vs Secondary Data

Where did the data come from?

ข้อมูลมาจากไหน?

Primary Data

ข้อมูลปฐมภูมิ

Collected directly
by you.



Secondary Data

ข้อมูลทุติยภูมิ

Collected by
someone else.



Which one are we focusing on today?

What Are Secondary Sources?

Data that already exists

ข้อมูลที่มีอยู่แล้ว



แหล่งข้อมูลทุติยภูมิ คือ แหล่งที่เผยแพร่ข้อมูลที่มีอยู่แล้ว

Why Use Secondary Data?

It can save time, but it has limits.

ช่วยประหยัดเวลา แต่มีข้อจำกัด

Benefits

- ✓ faster to get
- ✓ cheaper
- ✓ already organized
- ✓ useful for comparison

Limits

- ⚠ may be outdated
- ⚠ may be incomplete
- ⚠ may not match our question
- ⚠ may be difficult to edit

Common Online Data Formats

Data can be published in different ways.

ข้อมูลออนไลน์มีหลายรูปแบบ



**XLSX /
spreadsheet**



CSV file



PDF file



Website Table



API



Report Page



รูปแบบไฟล์มีผลต่อการนำข้อมูลไปใช้

File format affects how we can use the data.

Which File Is Easier to Use?

Some files are better for analysis.

บางไฟล์เหมาะกับการวิเคราะห์มากกว่า



Easy to Analyze



- CSV



- XLSX



Easy to Read



- PDF



- report page



Needs Programming



- API



- online data



Key idea: Choose a format that matches your task.

เลือกรูปแบบไฟล์ให้เหมาะกับงาน

Reports and Website Tables

Useful, but often already summarized.

มีประโยชน์ แต่อาจเป็นข้อมูลสรุปแล้ว



Report



Website Table

Year	Deaths	Injuries	Total Accidents
2021	12,345	78,910	91,255
2022	11,876	74,512	86,388
2023	11,334	72,105	83,439

What we see

- ✓ totals
- ✓ averages
- ✓ summaries
- ✓ rankings

What may be missing

- ✗ raw details
- ✗ causes
- ✗ individual records
- ✗ full context



Example: A road accident report may show deaths and injuries, but not the full causes.

Raw Data vs Summarized Data

Raw data gives more details.

ข้อมูลดิบให้รายละเอียดมากกว่า

Raw Data

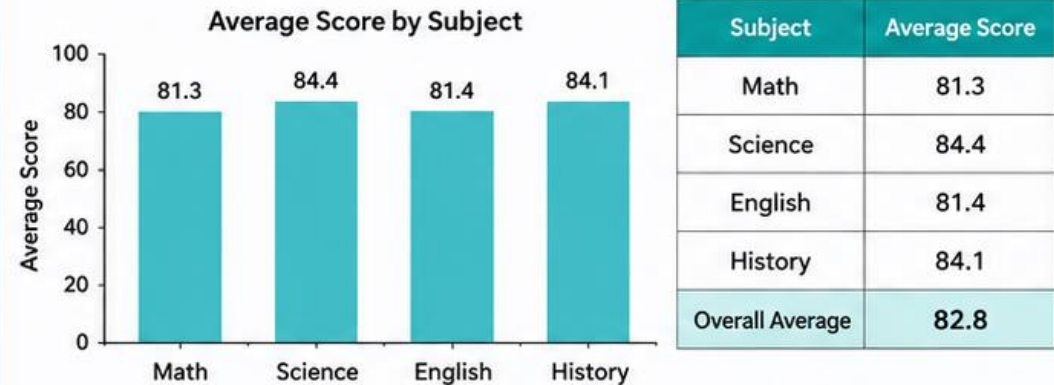
- individual records
- more details
- easier to analyze deeply

ID	Student	Math	Science	English	History	Total
1	Ava	88	92	85	90	355
2	Ben	76	81	79	84	320
3	Chai	91	89	93	87	360
4	Dara	68	74	71	77	290
5	Emma	95	94	90	92	371
6	Finn	82	85	80	86	333
7	Gina	73	78	76	75	302
8	Hugo	90	92	88	91	361
9	Ivy	66	72	69	74	281
10	Jake	84	87	83	85	339

ข้อมูลดิบ (รายละเอียดรายบุคคล)

Summarized Data

- totals or averages
- easier to read
- less detailed



ข้อมูลสรุป (ค่าเฉลี่ยหรือผลรวม)



Which one is better for making new graphs?

แบบไหนเหมาะกว่าสำหรับการสร้างกราฟใหม่?

Data Search Scenario

Start with a clear question.

เริ่มจากคำถามที่ชัดเจน

Scenario:

We want to study:

Population income in Thailand

รายได้ของประชากรในประเทศไทย



Who might collect this data?



Where could we find it?



Is it published online?



Is the source trustworthy?

Where Can We Find This Data?

Official sources are usually better.

แหล่งข้อมูลทางการมักน่าเชื่อถือกว่า

Most
Trustworthy



Less
Ideal



Random Websites



Start with official sources when accuracy matters.

เมื่อต้องการความถูกต้อง ให้เริ่มจากแหล่งข้อมูลทางการ



Guided Search Task

Find a dataset and inspect it.

ค้นหาชุดข้อมูลและตรวจสอบ



Task

Search for a dataset about:

Population income in Thailand

รายได้ของประชากรในประเทศไทย



Complete

- ☐ What website did you use?
- ☐ Who published the data?
- ☐ What file type is available?
- ☐ What years are included?
- ☐ What attributes can you see?

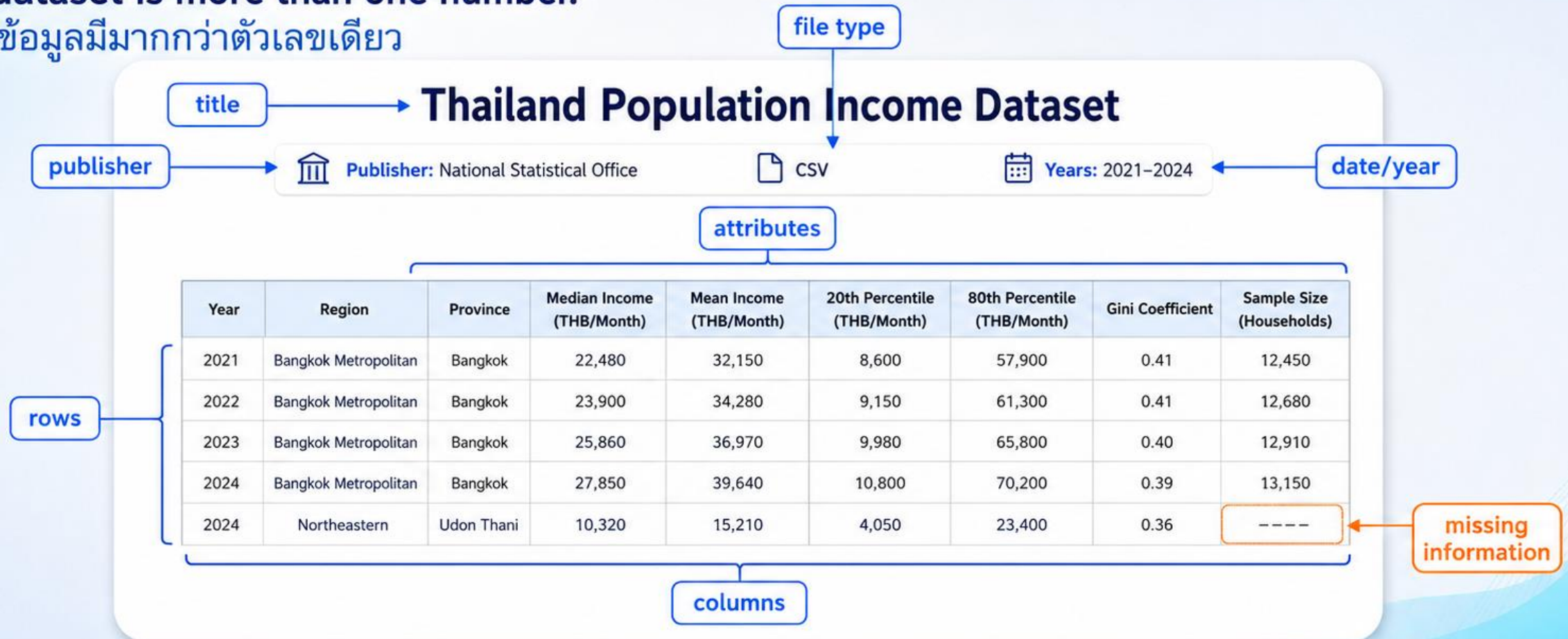


Pair Work: **8 minutes**

What Did We Find?

A dataset is more than one number.

ชุดข้อมูลมีมากกว่าตัวเลขเดียว



Is this dataset useful for our question?

ชุดข้อมูลนี้เหมาะกับคำถามของเราหรือไม่?

What Are Attributes?

Attributes describe each data item.

คุณลักษณะใช้อธิบายข้อมูลแต่ละรายการ



Attributes are the properties we record.

คุณลักษณะ คือ สิ่งที่เราใช้บันทึกข้อมูล



Name	Gender	Age	Height	Weight	Income	Expenses
นภัสสร	หญิง	16	158	48	12,000	8,500
กฤษฎี	ชาย	17	170	62	15,000	10,200
พิมพ์ชนก	หญิง	15	160	50	10,000	7,300
ชินวัตร	ชาย	16	165	55	11,000	7,800
ธาริน	หญิง	17	162	52	13,000	9,200
ปณณวิชญ์	ชาย	15	168	60	12,000	8,900



An **attribute** is a property of a person, object, place, or event.

คุณลักษณะ คือ คุณสมบัติของคน
สิ่งของ สถานที่ หรือเหตุการณ์



Name
ชื่อ



Gender
เพศ



Age
อายุ



Height
ส่วนสูง (ซม.)



Weight
น้ำหนัก (กก.)



Income
รายได้ (บาท)



Expenses
ค่าใช้จ่าย (บาท)

Attributes in a Student Dataset

Each column tells us something.

แต่ละคอลัมน์บอกข้อมูลบางอย่าง



Attributes = column names

คุณลักษณะ = ชื่อคอลัมน์

Student	Age	Height	Screen Time	Sleep Time
Student A	16	165 cm	4 h	7 h
Student B	17	170 cm	6 h	6 h
Student C	16	158 cm	3 h	8 h
Student D	17	172 cm	5 h	7 h



1. Which attributes are quantitative?

คุณลักษณะใดเป็นข้อมูลเชิงปริมาณ?



2. Which attributes help answer the question?

คุณลักษณะใดช่วยตอบคำถาม?



3. What other attribute might we need?

เรายังต้องการคุณลักษณะอื่นอะไรอีก?

What Information Is Missing?

A dataset may not answer every question.

ชุดข้อมูลอาจตอบคำถามไม่ได้ทั้งหมด



Example question:

**Is income
related to age?**



Dataset has:

- province
- year
- average income

province	year	average income
Bangkok	2022	325,000
Chiang Mai	2022	210,000
Khon Kaen	2022	195,000
...	...	...



Missing:

- average age

province	year	average income



We need another dataset or more attributes.

เราต้องหาข้อมูลเพิ่มเติมหรือคุณลักษณะเพิ่มเติม

Check the Source Before Using Data

Not all online data is trustworthy.

ข้อมูลออนไลน์ไม่น่าเชื่อถือทั้งหมด

Source Reliability Checklist



1 Who published it?



2 When was it updated?



3 Is it official?



4 Is the data complete?



5 Can we download it?



6 Does it match our question?



Useful data must be trustworthy.

ข้อมูลที่ใช้ได้ต้องน่าเชื่อถือ

Mini Activity: Evaluate a Dataset

Is this data useful?

ข้อมูลนี้ใช้ได้หรือไม่?

 Work with a partner.



Your Dataset



dataset.csv

CSV File



Inspect this dataset carefully.

Look closely at the data before you analyze.



Choose one dataset and answer:

- 1 What is the dataset about?
- 2 Who published it?
- 3 What file type is it?
- 4 What attributes does it include?
- 5 What information is missing?
- 6 Would you use it? Why?



Pair Work: **8 minutes**

What Did We Learn Today?

Finding data is only the first step.

การค้นหาคือข้อมูลเป็นเพียงขั้นตอนแรก

1



Secondary data already exists.

2



Data can come from files, reports, websites, and databases.

3



File type affects how we can use data.

4



Attributes describe each data item.

5



Some datasets are missing important information.

6



We must check the source before using data.



Question



Search



Source



File Type



Attributes



Check



Use

Exit Ticket

Before you leave...

ก่อนจบคาบเรียน...



Answer in your notebook:

- 1 What is secondary data?
- 2 Give one example of an online secondary source.
- 3 Why can PDF data be difficult to analyze?
- 4 What is an attribute in a dataset?
- 5 Why should we check the source before using data?



Good data must match the question.

ข้อมูลที่ดีต้องตรงกับคำถาม

