

Session 5

Organizing database

ARTNeT Consultant

Witada Anukoonwattaka, PhD

Thammasat University, Thailand

witada@econ.tu.ac.th



Asia-Pacific Research and Training Network on Trade

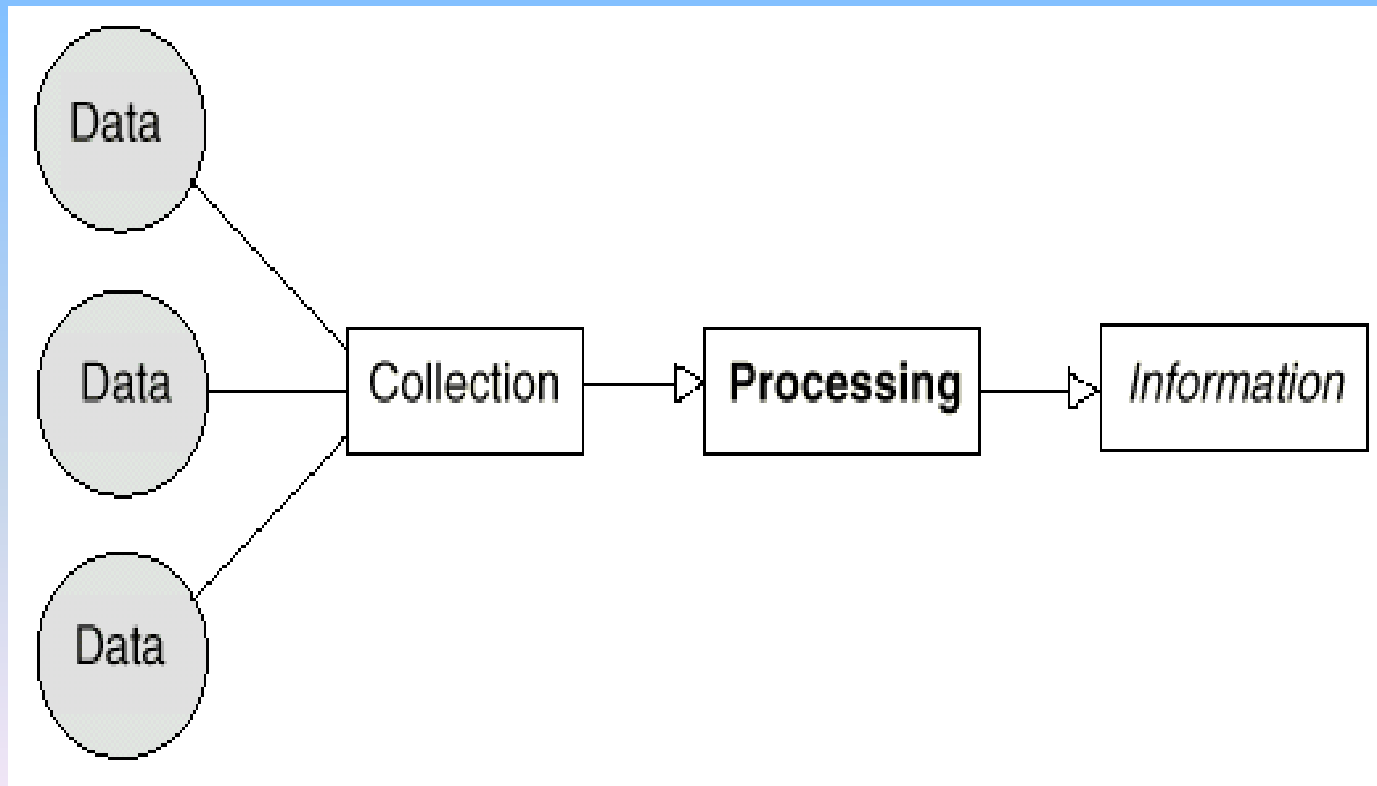
www.artnetontrade.org

Outline

- Data Processing
 - Data Coding
 - Data Input
 - Data Editing
 - Data Manipulation
- Scale of Data
 - Nominal
 - Ordinal
 - Numerical

Data Processing

- The procedures for turning data into information



Data Coding

<u>Data</u>	<u>Variable</u>	<u>Qualitative or Quantitative</u>
Gender of the worker	Gender	Qualitative
Age of the worker	Age	Quantitative
Speed of assembly	Speed	Quantitative
Country of production	Country	Qualitative

<u>Variable</u>	<u>Response</u>	<u>Code</u>
Gender	Female, Male	F, M
Age	___years	
Speed	___units per hr	
Country	Country names	J, T, C

[Asia-Pacific Research and Training Network on Trade](http://www.artnetontrade.org)

Why is coding important ?

- Coding makes it easier to do data entry
- Coding makes it possible to do quantitative analysis on the qualitative data

E.g.

- How many percent of the sample is female?
 - The share of “F” in value list of the gender variable.

Data Editing

- To check for accuracy and eliminate problems that can produce disorganized or incorrect information

Do you see any inaccurate response in this survey form?

34A.	ON WHICH DAYS DID <u>Person 1</u> WORK LAST WEEK (IN ALL JOBS)?
Yes	MON <input checked="" type="checkbox"/> TUES <input checked="" type="checkbox"/> WED <input checked="" type="checkbox"/> THU <input checked="" type="checkbox"/> FRI <input checked="" type="checkbox"/> SAT <input checked="" type="checkbox"/> SUN <input checked="" type="checkbox"/>
No	MON <input type="checkbox"/> TUES <input type="checkbox"/> WED <input type="checkbox"/> THU <input type="checkbox"/> FRI <input type="checkbox"/> SAT <input type="checkbox"/> SUN <input type="checkbox"/>
34B.	DID <u>Person 1</u> HAVE ANY TIME OFF FROM <u>this</u> JOB(S) ON THOSE DAYS?
Yes	<input type="checkbox"/>
No	<input checked="" type="checkbox"/>
34C.	DID <u>Person 1</u> WORK ANY PAID OR UNPAID OVERTIME ON ANY DAY LAST WEEK?
Yes	<input checked="" type="checkbox"/>
No	<input type="checkbox"/>
Other	<input type="checkbox"/>
34D.	HOW MANY HOURS DID <u>Person 1</u> ACTUALLY WORK LAST WEEK (LESS THE TIME OFF) (BUT) (COUNTING THE OVERTIME)?
35 hours or more	<input type="checkbox"/>
1 - 34 hours	<input type="checkbox"/>
Less than 1 hour/ no hours	<input checked="" type="checkbox"/>

Data Manipulation

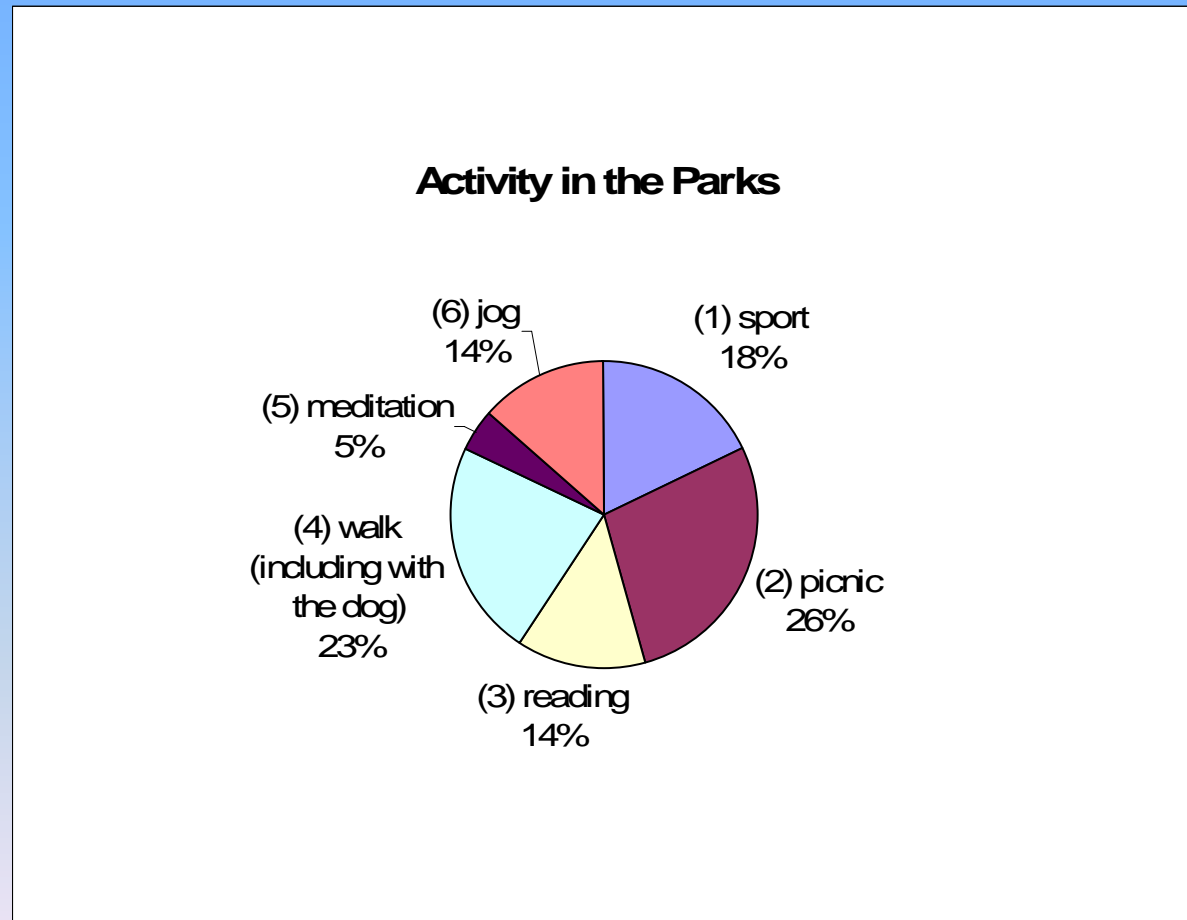
Using computer software to produce the desired output.

- *Spreadsheets* are used to create formulas that automatically add columns or rows of figures, calculate means and perform statistical analyses.
- *Databases* are electronic filing cabinets: systematically storing data for easy access to produce summaries and reports. A database program should be able to store, retrieve, sort, and statistically analyze data.
- *Charts* can be created from a table of numbers and displayed in a number of ways, to show the significance of a selection of data. Bar, line, pie and other types of charts can be generated and manipulated to advantage.

Example of Output: Descriptive Statistics

Statistics	<i>Activity</i>							
	<i>Worker</i>	<i>Industry</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>Patent</i>	<i>Foreign</i>
Mean	58.75	2	0.33	0.50	0.25	0.42	1	0.58
Standard Error	13.50	0.28	0.14	0.15	0.13	0.15	0.33	0.15
Median	45	2	0	0.5	0	0	1	1
Mode	30	2	0	1	0	0	2	1
Standard Deviation	46.76	0.95	0.49	0.52	0.45	0.51	1.13	0.51
Minimum	15	1	0	0	0	0	-1	0
Maximum	180	4	1	1	1	1	2	1
Sum	705	24	4	6	3	5	12	7
Count	12	12	12	12	12	12	12	12

Example of Output: Chart



Scale of Data

- Nominal Scale
- Ordinal Scale
- Numerical Scale
 - Interval
 - Ratio

Nominal Scale

- You can count but not order or measure the nominal variable.

Variables

Responses

Country: () Laos () Vietnam () Cambodia

Gender: () Male () Female

Examples of Coding Scheme:

Country: L= Laos V= Vietnam C= Cambodia

Gender: 0 = Male 1= Female

Note: Number in the code does not have 'quantitative' meaning.

Ordinal Scale

- You can rank or count but not measure the ordinal variable.

Variable

Responses

Quality: () Good () Average () Poor

Examples of coding scheme

3 = Good 2= Average 1= Poor

1 = Good 0= Average -1= Poor

Note: Number in the code has order meaning, but not value meaning.

Numerical Scale

- You can rank, count, and measure the numerical variable.

Variables

Responses

Speed of production: _____ units per hour

Export ratio: () below 25% () 25-50%
() between 50-75% () above 75%

Coding Scheme?

The scale of the numerical variable has a meaning.

Numerical Scale: Interval

- In interval scale the “interval” between value *does* have meaning, but “zero” *doesn't* have meaning
 - For example, the temperature from 30C-40C is the same as the distance from 70C-80C. The interval between values is interpretable.

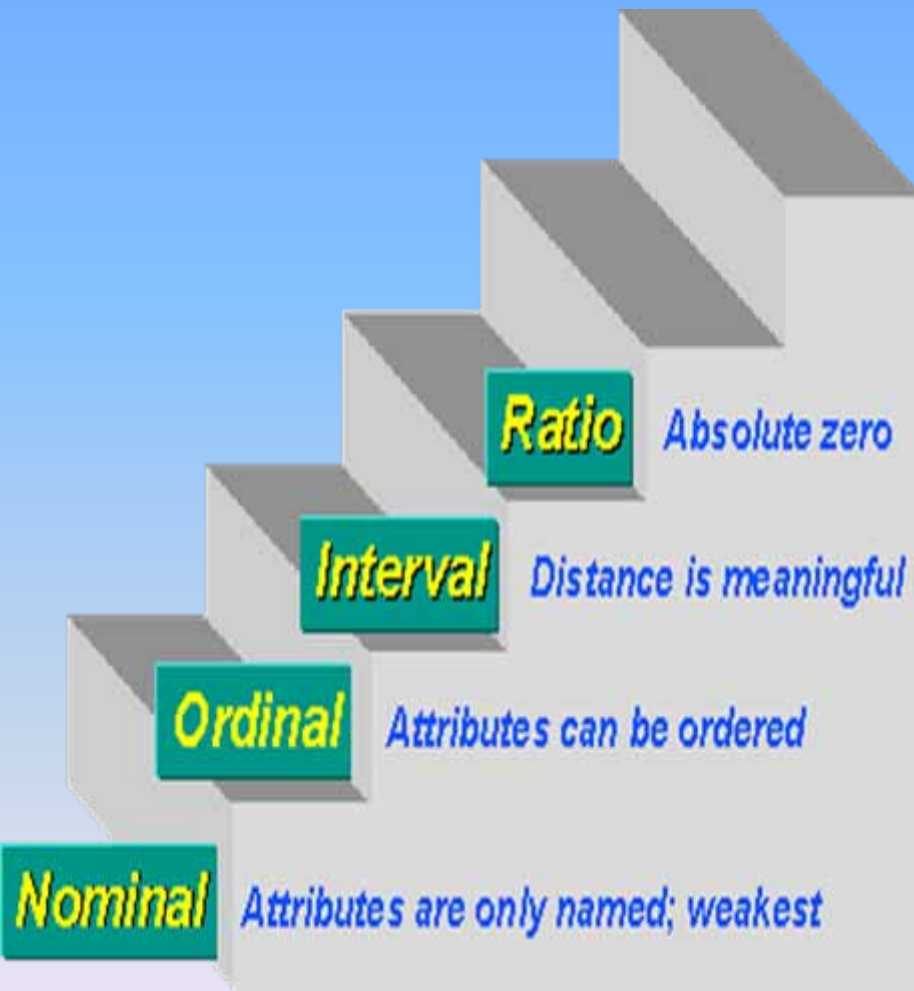
Numerical Scale: Ratio

- In ratio scale “zero” is meaningful. You can construct a fraction (or ratio) with a ratio variable.
 - For example: the export volume.
 - Is it meaningful if the export volume is zero in the third quarter of this year?
 - What does zero Celsius mean?

Examples: Interval and Ratio

- “The current temperature in Bangkok is 30C, while in Tokyo it is 15C.” Is it correct to say that Bangkok is twice as warm as Tokyo?
- Is it meaningful to say that “...we had twice as many clients in the past six months as we did in the previous six months.” ?

Analytical Quality of Data Scaling



There is a hierarchy implied in the level of measurement idea.

At lower levels of measurement, assumptions tend to be less restrictive and data analyses tend to be less sensitive.

At each level up the hierarchy, the current level includes all of the qualities of the one below it and adds something new.

In general, it is desirable to have a higher level of measurement (e.g., interval or ratio) rather than a lower one (nominal or ordinal).

Exercise 4: Data Processing

From Exercise 2.2, We did a firm survey and got the following data.

- *Firm* is the observation number of the questionnaire. One questionnaire was filled out for one firm.
- *Worker* is the number of workers in the firm. (What scale it is?)
- *Industry* consists of 4 choices of industry:
 - Agriculture, Manufacturing, Mining, Services.
 - What scale is it?
- *Activity* consists of 4 choices of trading activity. One firm can have several activities :
 - Export inputs, export finished goods, import inputs, import finished goods.
 - What scale is it?
- *AFTA* has 5 values about opinions towards AFTA:
 - 2= strongly disagree, -1= disagree, 0= indifference, 1= agree, 2=strongly agree.
 - What scale is it?
- *Foreign investment* has dichotomous responses (Yes or No) about the existence of foreign investment in the firm.
 - What scale is it?

The survey on 12 sample firms gives the raw data as follows

Firm	Worker	Industry	Activity	AFTA	Foreign Investment
1	30	1	1, 2, 3	0	N
2	30	3	2,3	1	Y
3	60	2	1, 2	2	Y
4	45	1	2,4	-1	N
5	30	1	1	1	N
6	60	2	2	2	Y
7	30	3	4	1	N
8	45	2	3, 4	-1	N
9	15	1	4	1	Y
10	60	2	2	2	Y
11	180	4	1, 2, 3, 4	2	Y
12	120	2	1,2,4	2	Y

Please code the raw data and fill in the formatted data table (provided on the next page).

Note: We usually convert our questionnaire data into *numbers, because most statistical software only accepts numerical data.*

Formatted Data Sheet

Firm	Worker	Industry	Activity				AFTA	Foreign
			1	2	3	4		
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								