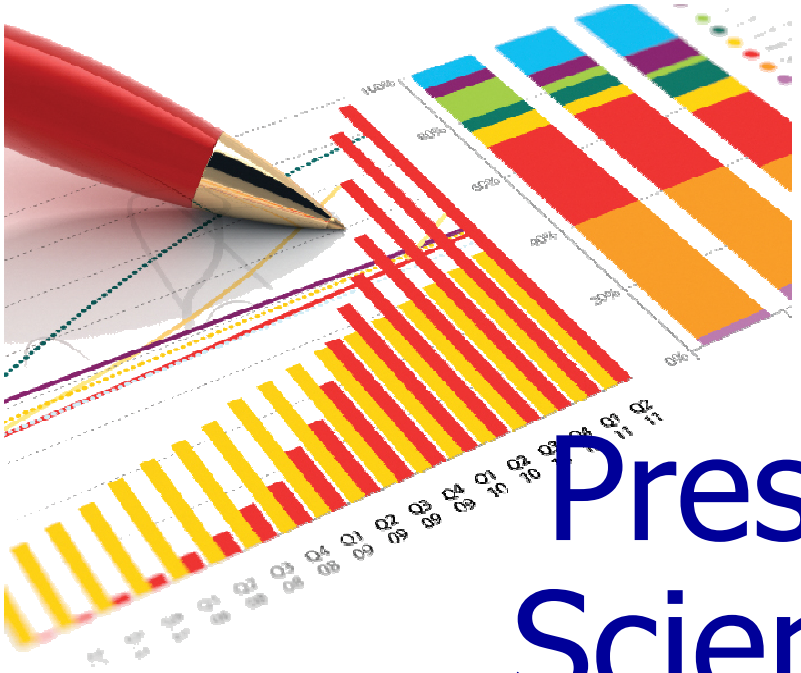
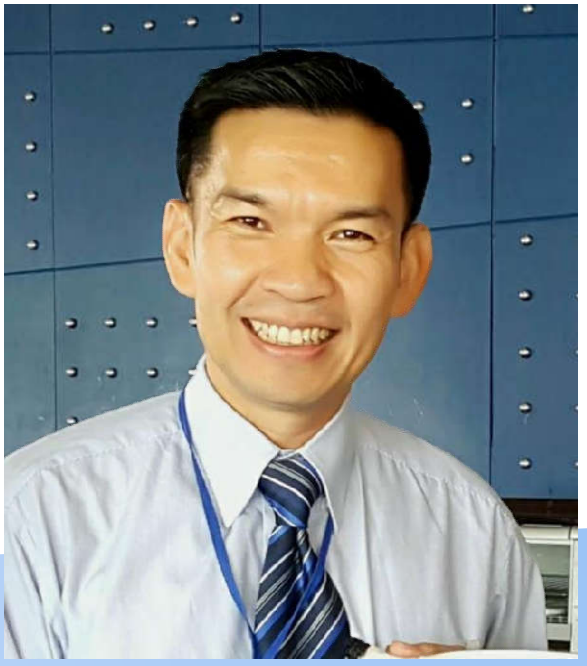


Presenting of Scientific Data

Chawetsan Namwat

MD., MPH., FETP.

Director, FETP Thailand



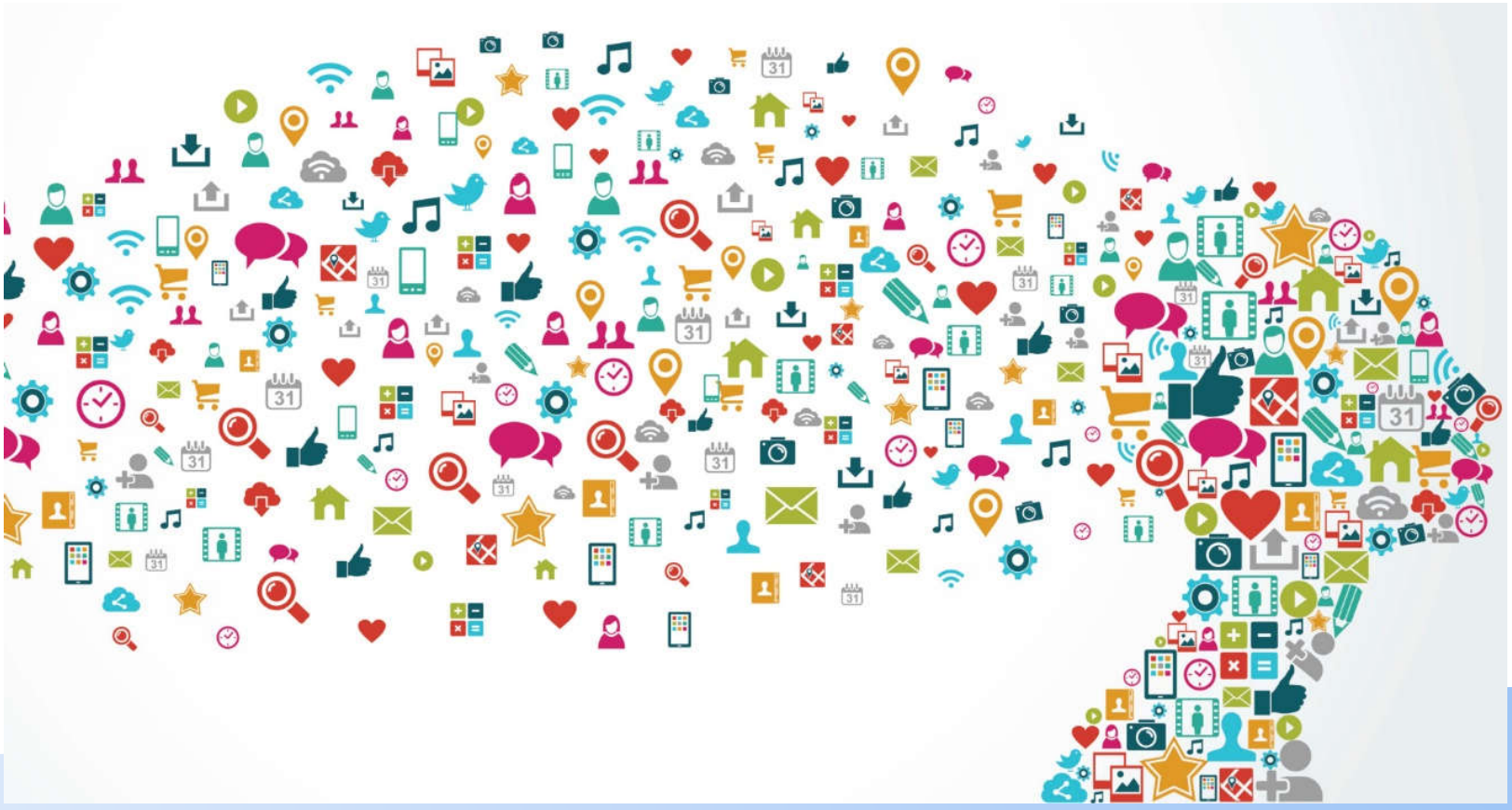
Why data presentation important

- How you can summarize data contained thousands of records in the annual report
- How you present your study in 15 minutes
- How you understand data your network collected and sent to you

Note: Keep alert! I will random your number (by excel) to interpret the presentation.

Another end of data presentation: Data visualization / Infographic

- https://www.slideshare.net/ethos3/the-secret-to-memorable-data-visualizations



Outline

- Table
- Graph & Chart : how to use & common mistakes
  - Line
  - Bar
  - Histogram : Epidemic Curve
  - Pie
  - Maps
  - Survival curve

Why tables & graphs

- In case of small amount of data, line listing review is enough for summarizing. → use table

ID #	Last name	myalgia	fever	facial edema	eosinophil count	Physician diagnosis	Lab confirm	Classification
1	Abels	yes	yes	no	495	trichinosis	yes	----------
2	Baker	yes	yes	yes	pending	trichinosis ?	pending	----------
3	Corey	yes	yes	no	1,100	trichinosis	pending	----------
4	Dale	yes	no	no	2,050	EMS ?	pending	----------
5	Ring	yes	no	no	600	trichinosis	not done	----------

- More complex data → graphs, charts

TABLE. Clinical features, treatment, and outcomes in five patients with laboratory-confirmed influenza A (H5N1), by sex and age of patient — Thailand, 2004

Sex	Age (yrs)	Signs and symptoms on admission*	Subsequent complications	Initial investigative findings	Treatment and outcome
Male	7	Fever, cough, sore throat for 6 days. Dyspnea on day 6; CXR† bilateral interstitial infiltrates.	Respiratory failure on day 10; cardiac failure, pneumothorax, ARDS§, gastrointestinal bleeding.	Leukocytes: 4,100/μL Lymphocytes: 1,440/μL Platelets: 304,000/μL AST¶: 120, ALT**: 52	Oseltamivir on days 18–22. Died on day 29.
Male	6	Fever, cough, rhinorrhea for 5 days. Dyspnea on day 6; CXR patchy infiltrates in right lower lobe.	Respiratory failure on day 8; hepatitis, ARDS.	Leukocytes: 1,200/μL Lymphocytes: 624 /μL Platelets: 89,000/μL AST: 790, ALT: 150 Proteinuria: ≥3	Oseltamivir on days 18–20. Died on day 20.
Male	6	Fever, cough, rhinorrhea, sore throat for 4 days. Dyspnea on day 5; CXR multifocal patchy infiltrates.	Respiratory failure on day 6; pneumothorax, ARDS.	Leukocytes: 2,200/μL Lymphocytes: 638/μL Platelets: 150,000/μL AST: 175, ALT: 43	Died on day 18.
Female	58	Fever, cough, sore throat, myalgia for 2 days. Dyspnea on day 2; CXR multifocal patchy infiltrates.	Respiratory failure on day 4; cardiac failure, renal failure, ARDS.	Leukocytes: 5,680/μL Lymphocytes: 454/μL Platelets: 185,000/μL BUN††: 39 mg/dL Creatinine: 2.3 mg/dL	Died on day 8.
Male	6	Fever, cough, sore throat, myalgia for 4 days. Dyspnea on day 5; CXR multifocal patchy infiltrates.	Respiratory failure on day 5; cardiac failure, renal failure, ARDS.	Leukocytes: 2,900/μL Lymphocytes: 696/μL Platelets: 87,000/μL AST: 280, ALT: 50 BUN: 54 mg/dL Creatinine: 4.6 mg/dL	Oseltamivir on days 5–8. Died on day 8.

* No patients had an underlying illness reported.
† Chest radiograph.
§ Acute respiratory distress syndrome.
¶ Aspartate aminotransferase.
** Alanine aminotransferase.
†† Blood urea nitrogen.

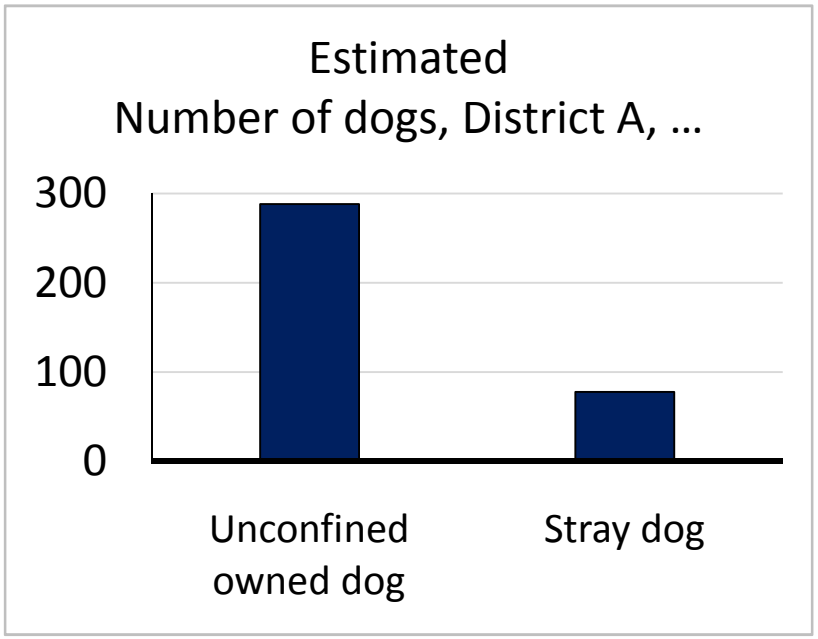
Source: https://www.cdc.gov/mmwr/PDF/wk/mm5305.pdf

Graph & Chart

- Helpful tools to aid in verifying and analyzing the data
- Visualizing: A picture is worth a thousand words.

Estimated number of dogs in District A, Thailand, 2010

Dog type	Estimated Number
Unconfined owned dog	288
Stray dog	78



Data source: Wongphruksasoong V, Santayakorn S, Sitthi W, Ardkham B, Pisek S, Srisai P, et al. Census versus capture-recapture method to estimate dog population in Lumlukka District, Phathum Thani Province, Thailand, 2010. OSIR. 2016 Mar;9(1):15-20. <http://www.osirjournal.net/issue.php?id=93>.

Before constructing any display of epidemiologic data

- Determine the point to be conveyed, first.
  - Are you highlighting a changing pattern from the past?
  - Are you showing a difference in incidence by geographic area or by some predetermined risk factor?
  - What is the interpretation you want the reader to reach?
- Your answers to these questions will help in determining the choice of display.

Tables

- A table is a set of data arranged in rows and columns.
- Serve as the basis for preparing graphs & charts
- A table in a printed publication should be self-explanatory

One-variable table

Table 1. Reported cases of influenza by age, Thailand, 2015

Age	Cases	
	Number	Percentage
0 - 27 d	27	0.03
28d - 11mo	2,402	3.08
1-4 yr	14,235	18.27
5-9 yr	12,368	15.87
10-14 yr	7,829	10.05
15-24 yr	8,890	11.41
25-34 yr	9,139	11.73
35-44 yr	8,044	10.32
45-54 yr	6,360	8.16
55-64 yr	4,630	5.94
65+ yr	3,995	5.13
Unknown	7	0.01
Total	77,926	100.00

What

When

Where

decimals, and alignment

Sum of the column

Source: report 506, BOE, DDC, MOPH, Thailand

One-variable tables with cumulative percent

Table 2. Reported cases of influenza by age, Thailand, 2015

Age	Cases		
	Number	Percentage	Cumulative %
0 - 27 d	27	0.03	0.03
28d - 11mo	2,402	3.08	3.11
1-4 yr	14,235	18.27	21.38
5-9 yr	12,368	15.87	37.25
10-14 yr	7,829	10.05	47.3
15-24 yr	8,890	11.41	58.71
25-34 yr	9,139	11.73	70.44
35-44 yr	8,044	10.32	80.76
45-54 yr	6,360	8.16	88.92
55-64 yr	4,630	5.94	94.86
65+ yr	3,995	5.13	99.99
Unknown	7	0.01	100.00
Total	77,926	100.00	

Source: report 506, BOE, DDC, MOPH, Thailand

Two-variable table

Table 3. Reported cases of influenza by age and sex, Thailand, 1 Jan-18 June 2018

Variable 1

Variable 2

Age	Number of cases by sex		
	Male	Female	Total
0 - 27 d	13	14	27
28d - 11mo	1,189	914	2,103
1-4 yr	5,833	5,022	10,855
5-9 yr	4,714	4,241	8,955
10-14 yr	2,513	2,044	4,557
15-24 yr	2,517	2,948	5,465
25-34 yr	2,462	3,827	6,289
35-44 yr	2,575	3,631	6,206
45-54 yr	1,891	2,793	4,684
55-64 yr	1,474	2,403	3,877
65+ yr	1,321	1,785	3,106
Unknown	0	3	3
Total	26,502	29,625	56,127

Source: report 506, BOE, DDC, MOPH, Thailand

Two-variable table

Table 3. Dog type and vaccination coverage in urban district A, Thailand, 2010

Row %

Dog type	Vaccination history in previous year			
	Vaccinated	Non-vacc	Total	%vacc
Confined	488	31	519	94.0
Unconfined	109	55	164	66.5
Stray	24	29	53	45.3
Total	621	115	736	84.4

Note: unknown history excluded

Data source: Wongphruksasoong V, Santayakorn S, Sitthi W, Ardkham B, Pisek S, Srisai P, et al. Census versus capture-recapture method to estimate dog population in Lumlukka District, Phathum Thani Province, Thailand, 2010. OSIR. 2016 Mar;9(1):15-20. <http://www.osirjournal.net/issue.php?id=93>.

- **Three-variable table**:

  three-way tables are often hard to understand, they should be used only when ample explanation and discussion is possible.

Pitfalls in constructing a table

- Unclear title, & label
- Confused meaning of “zero” e.g.
  “0”=No case, “ - ” =No report received
- Bad alignment of decimal
- Too much decimals

Composite Table

Table 1. Baseline Characteristics of the Subjects (Modified Intention-to-Treat Population).

Variable	Vaccine (N = 8197)	Placebo (N = 8198)	All Subjects (N = 16,395)
	number (percent)		
Sex			
Male	5033 (61.4)	5031 (61.4)	10,064 (61.4)
Female	3164 (38.6)	3167 (38.6)	6,331 (38.6)
Age group			
≤20 yr	2297 (28.0)	2246 (27.4)	4,543 (27.7)
21–25 yr	3633 (44.3)	3708 (45.2)	7,341 (44.8)
≥26 yr	2267 (27.7)	2244 (27.4)	4,511 (27.5)
Province			
Chon Buri	4107 (50.1)	4107 (50.1)	8,214 (50.1)
Rayong	4090 (49.9)	4091 (49.9)	8,181 (49.9)
Marital status			
Single	(40.9)	3338 (40.7)	6,691 (40.8)
Married	(50.1)	4169 (50.9)	8,279 (50.5)
Divorced	(7.3)	541 (6.6)	1,143 (7.0)
Widowed	(0.6)	64 (0.8)	114 (0.7)
Separate	(1.0)	86 (1.0)	168 (1.0)

Vaccination with ALVAC and AIDSVAX to Prevent HIV-1 Infection in Thailand
S. Rerks-Ngarm, et. al
N Engl J Med 2009;361.

Composite Table

Table 2. Rate of HIV Infection and Vaccine Efficacy, According to Selected Baseline Variables (Modified Intention-to-Treat Population).

Variable	Vaccine (N=8197)				Placebo (N=8198)				Vaccine Efficacy
	No. Evaluated	No. with Infection	No. of Person-Years	Rate	No. Evaluated	No. with Infection	No. of Person-Years	Rate	
				no./person-yr				*no./person-yr*	*% (95% CI)*
All subjects	7960	51	26,507	0.192	7988	74	26,478	0.279	31.2 (1.7 to 51.8)
Sex									
Male	4875	32	16,221	0.197	4885	43	16,179	0.266	25.8 (–17.3 to 53.0)
Female	3085	19	10,286	0.185	3103	31	10,300	0.301	38.6 (–8.6 to 65.3)
Age group									
≤20 yr	2228	12	7,358	0.163	2185	11	7,216	0.152	7.1 (–143.0 to 52.7)
21–25 yr	3517	20	11,713	0.171	3610	40	11,946	0.335	49 (12.8 to 70.2)
≥26 yr	2215	19	7,437	0.255	2193	23	7,316	0.314	18.7 (–49.3 to 55.7)
Living with partner									
Yes	4017	19	13,466	0.141	4083	34	13,612	0.25	43.5 (1.0 to 67.8)
No	3943	32	13,041	0.245	3905	40	12,866	0.311	21 (–25.7 to 50.4)
Risk group									
Low					37	29	12,798	0.227	40.4 (–8.5 to 67.2)
Medium					22	22	7,353	0.299	47.6 (–6.0 to 74.0)
High					29	23	6,327	0.364	3.7 (–72.7 to 46.3)

Vaccination with ALVAC and AIDSVAX to Prevent HIV-1 Infection in Thailand
S. Rerks-Ngarm, et. al
N Engl J Med 2009;361.

Age group commonly used in US CDC

Table 4.8
Some standard groupings used at CDC

Notifiable diseases	Pneumonia & influenza mortality	Final mortality statistics	HIV/AIDS
<1 year	<28 days	<1 year	<5 years
1-4	28 days-<1 year	1-4	5-12
5-9	1-14	5-14	13-19
10-14	15-24	15-24	20-24
15-19	25-44	25-34	25-29
20-24	45-64	35-44	30-34
25-29	65-74	45-54	35-39
30-39	75-84	55-64	40-44
40-49	≥85	65-74	45-49
50-59	Unknown	75-84	50-54
≥60		≥85	55-59
Age not stated		Not stated	60-64
			≥65
Total	**Total**	**Total**	**Total**

Knowledge check

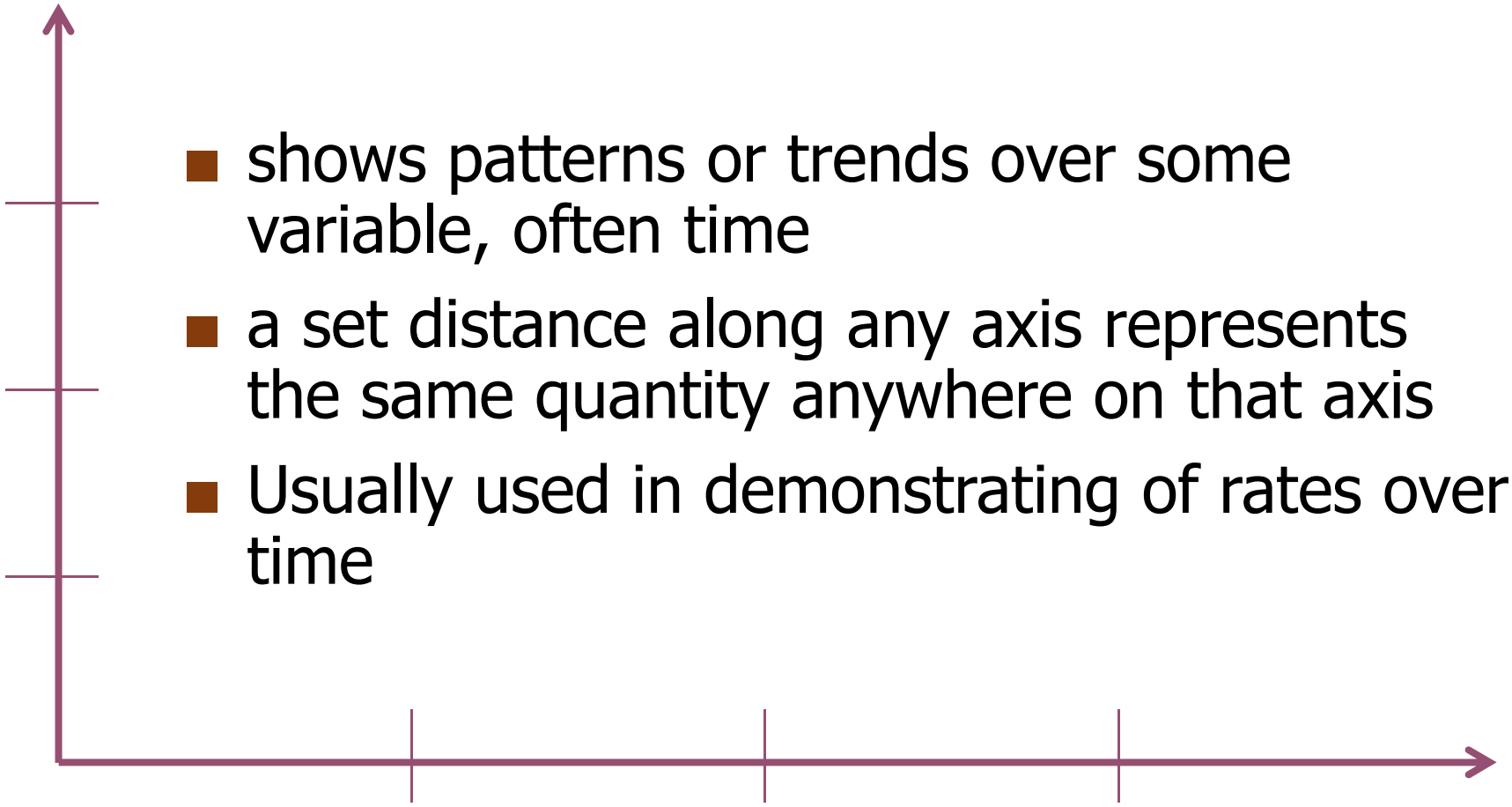
- Which is correct about composite table?
  A. Good for powerpoint presentation
  B. Commonly used for journal article
  C. Line listing
  D. Use for mapping
- In table display, which is not a common mistake?
  A. Decimal use : bad alignment, too much decimal
  B. Unclear title
  C. Inappropriate color use
  D. All of the above

Graphs

- Choose the right graph for the right data
- Use one graph to get one idea across
- Eliminate distracting, non-essential elements.

  e.g. secondary axis, gridlines, 3-D effects, bordering lines, etc.

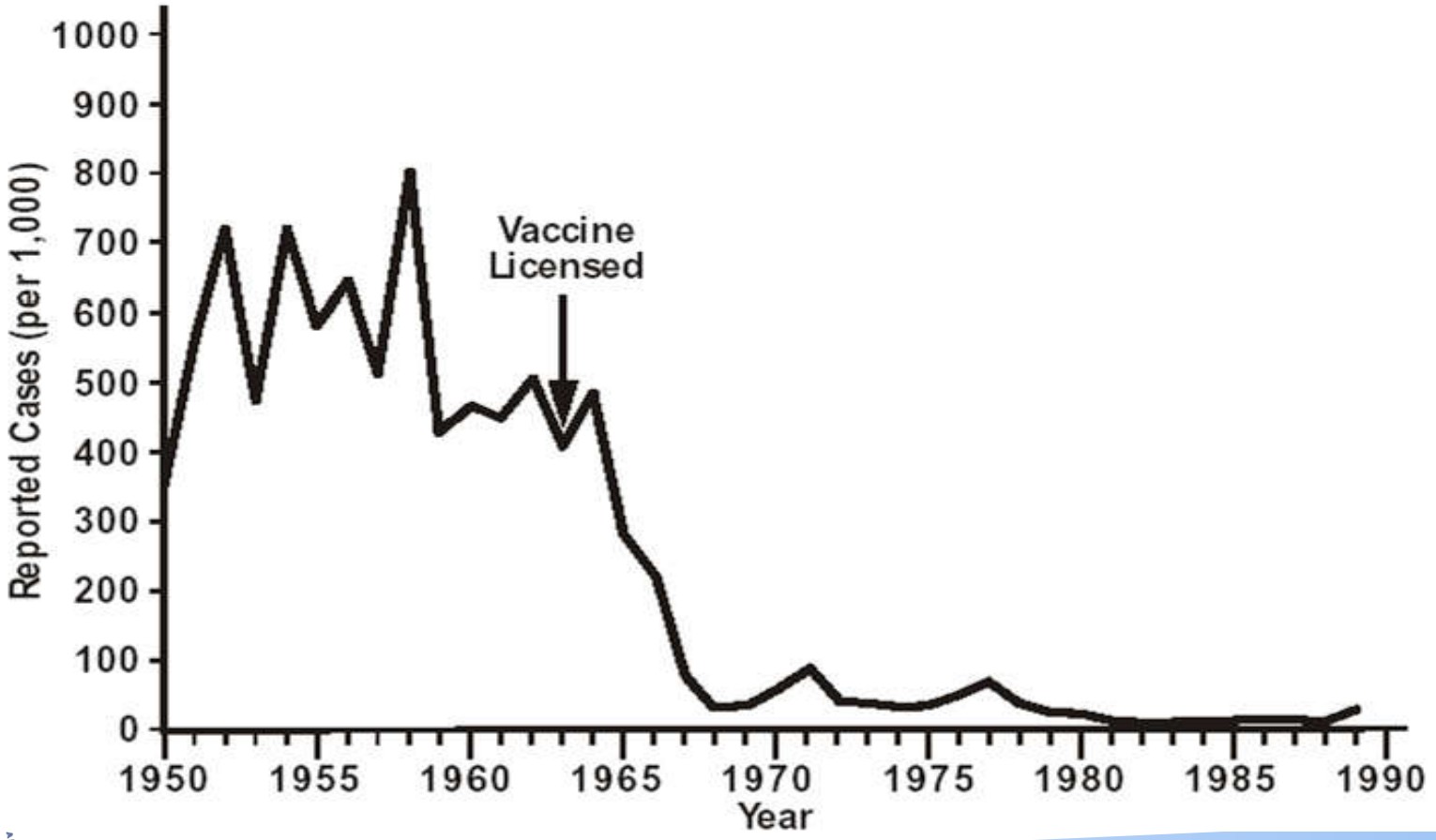
Arithmetic-scale Line Graph



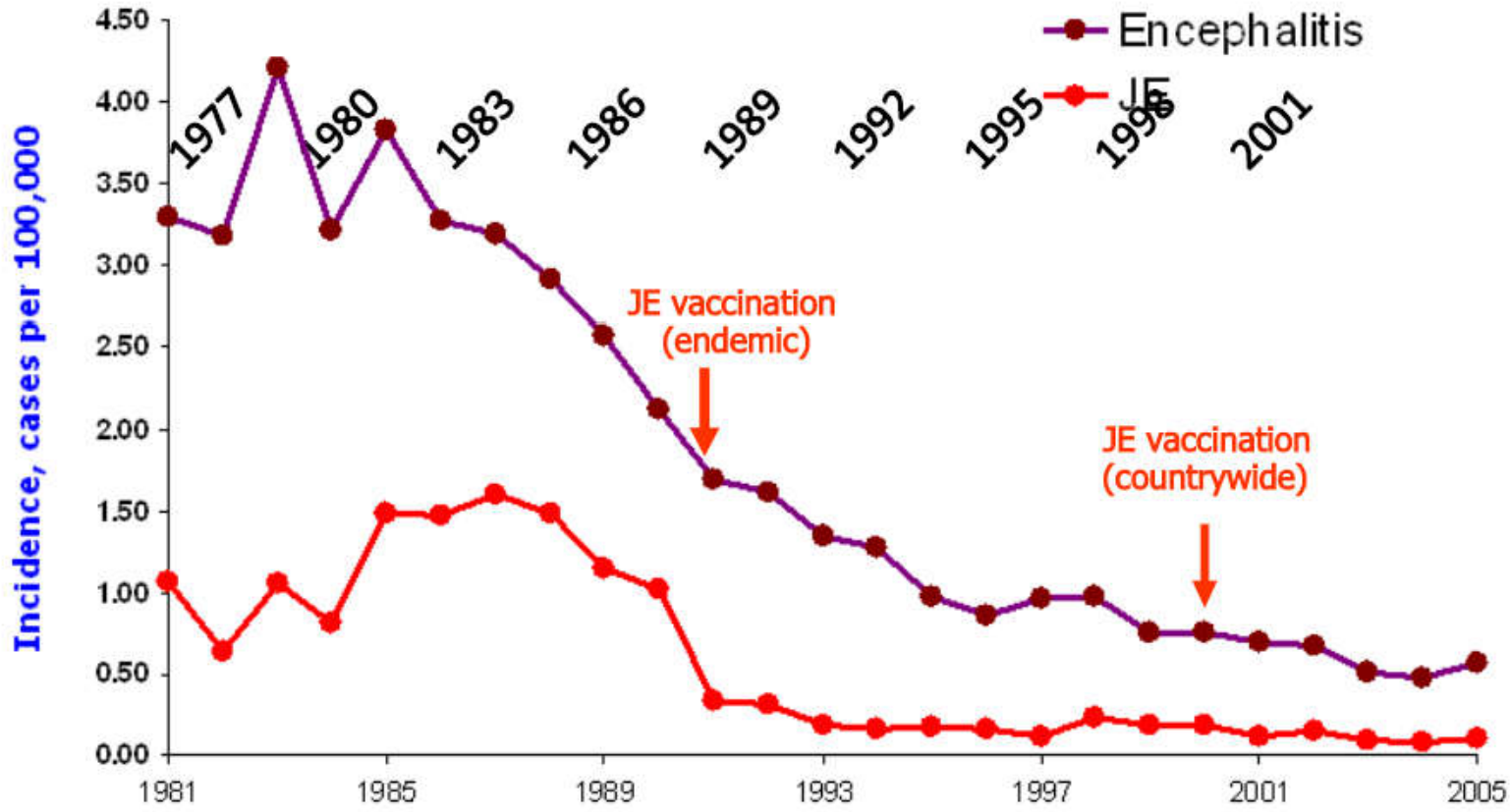
- shows patterns or trends over some variable, often time
- a set distance along any axis represents the same quantity anywhere on that axis
- Usually used in demonstrating of rates over time

Arithmetic line graph

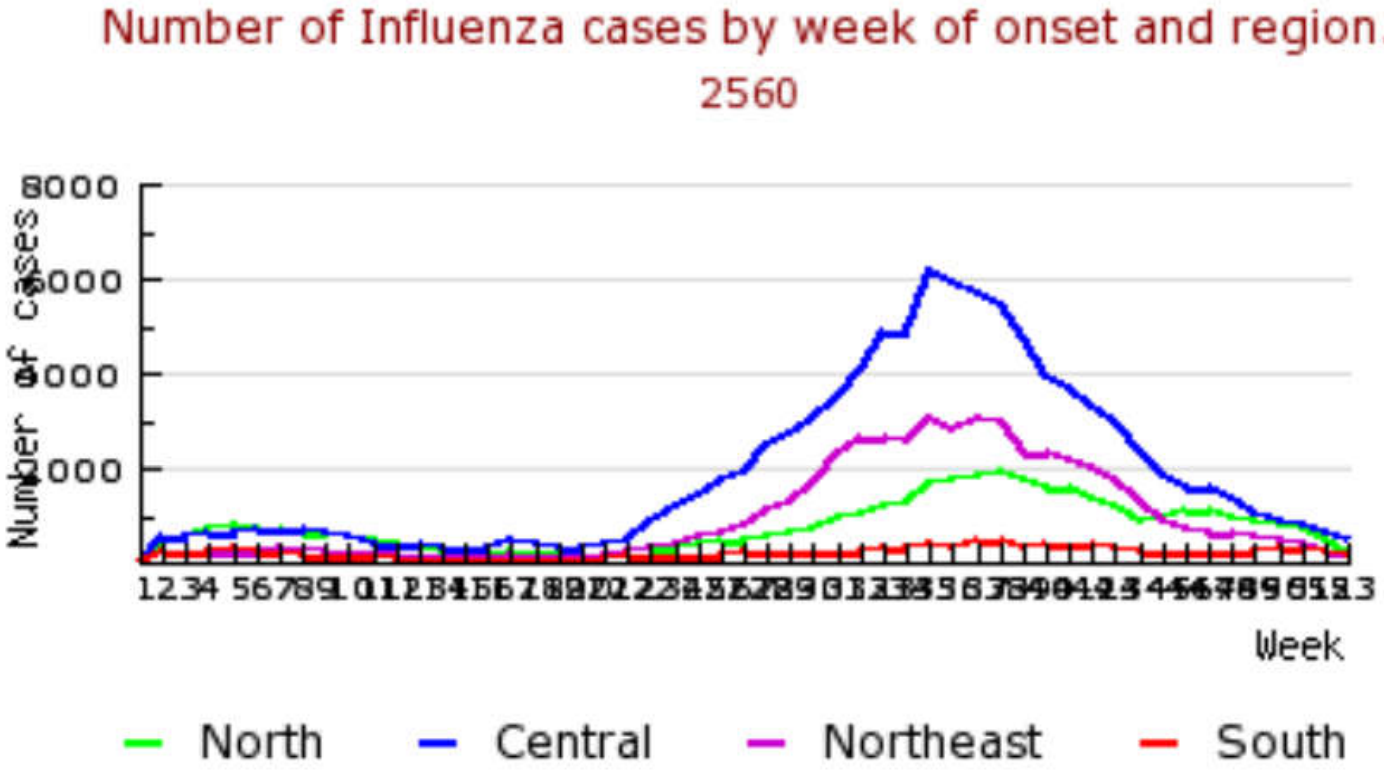
Figure 4.3
Example of arithmetic-scale line graph:
Measles (rubeola) by year of report, United States, 1950-1989



Encephalitis and JE incidence trend, Thailand, 1981-2005



Arithmetic line graph



© Bureau of Epidemiology, Department of Disease Control, Ministry of Public Health

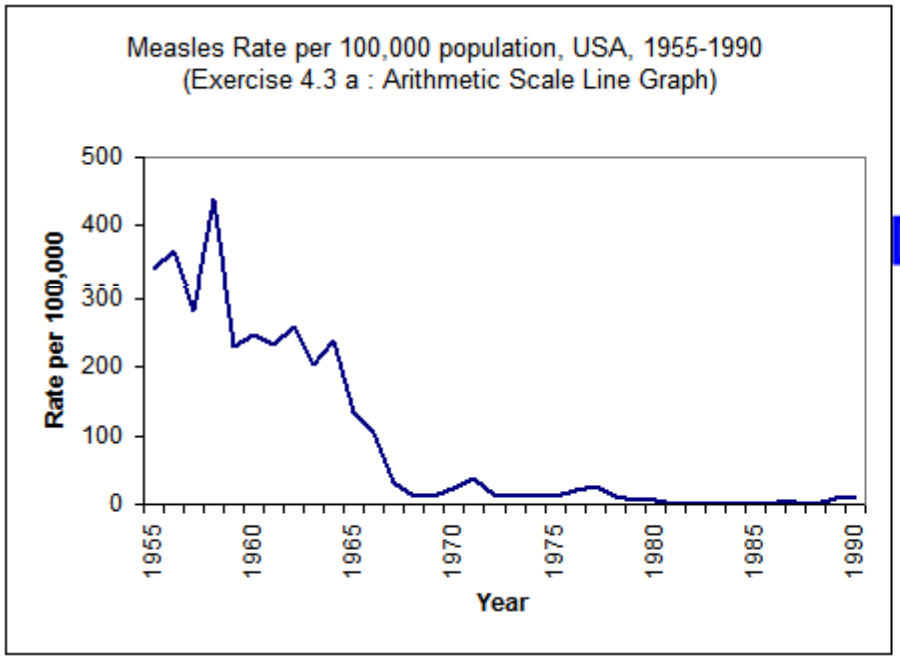
- Comparison of multiple data series
- Ratio of height : width approx. 3:5

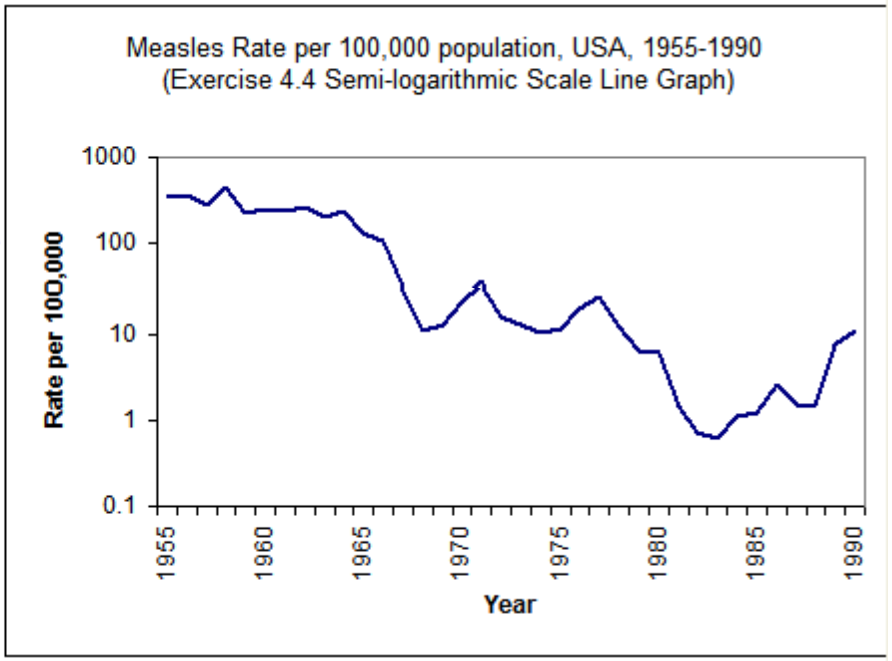
Semi-logarithmic scale line graph

- Y-axis: logarithmic scale
- X-axis: arithmetic scale

- Display the wide range of Y-axis value usually max : min >100 : 1
- Distance on Y-axis:
  1 to 10 = 10 to 100

Semi-logarithmic scale line graph





Histogram

- A histogram is a graph of the frequency distribution of a continuous variable, based on class intervals
- No space between columns
- Example: Epidemic Curve

  X-axis : Onset date (or time)

  Y-axis: Number of case

Figure 4.8
Example of histogram: Reported cholesterol levels among 4,462 men, Men's Health Study, United States, 1985-1986

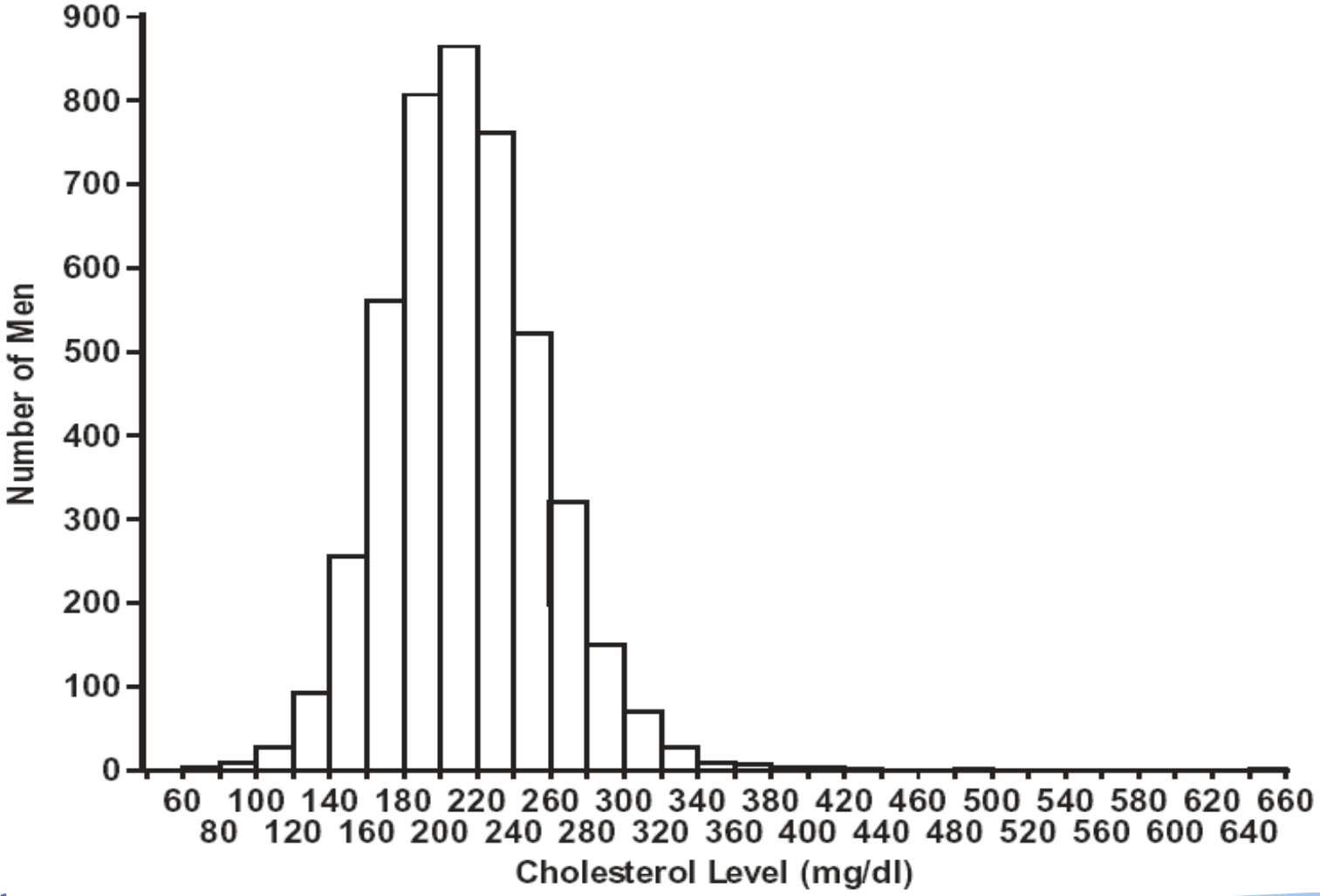
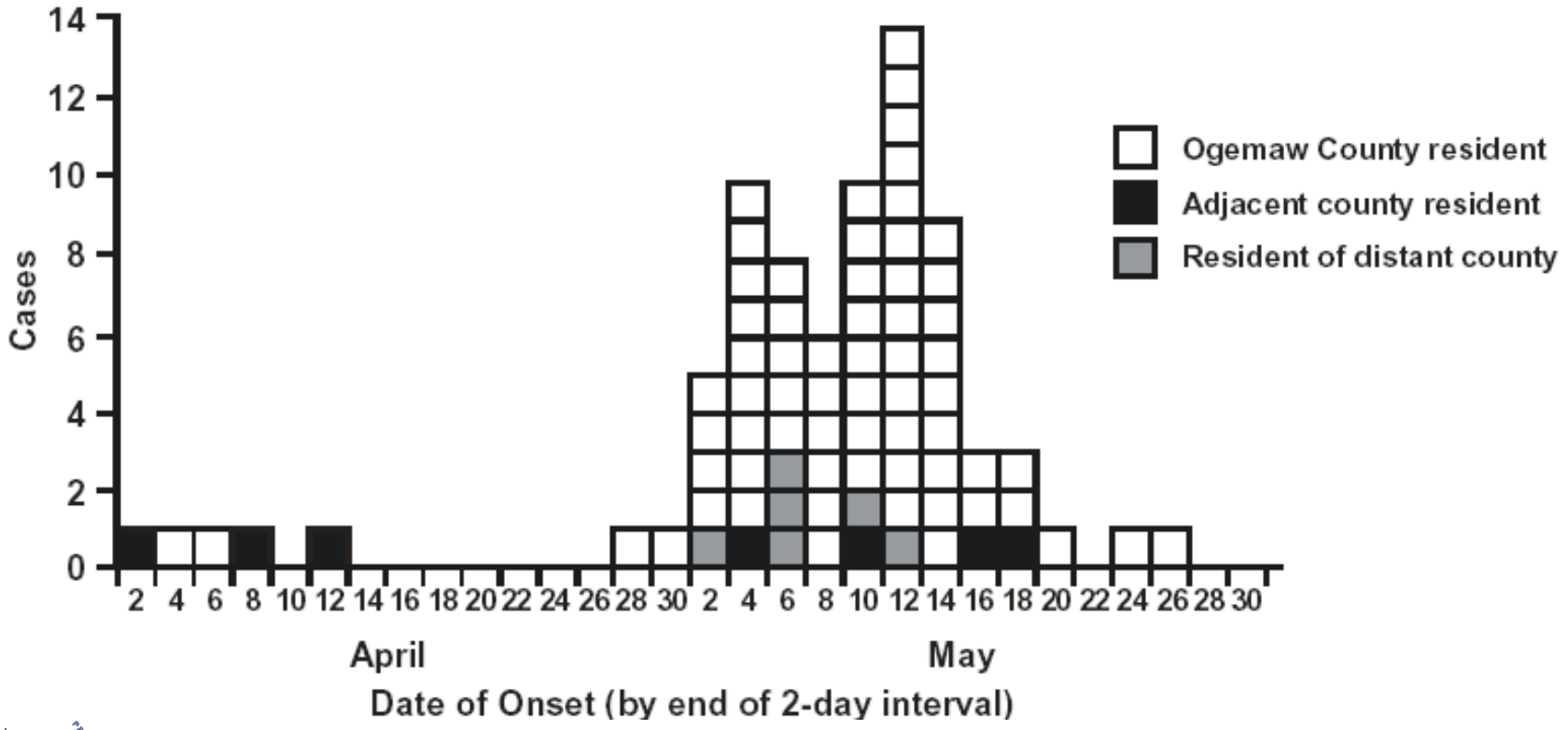
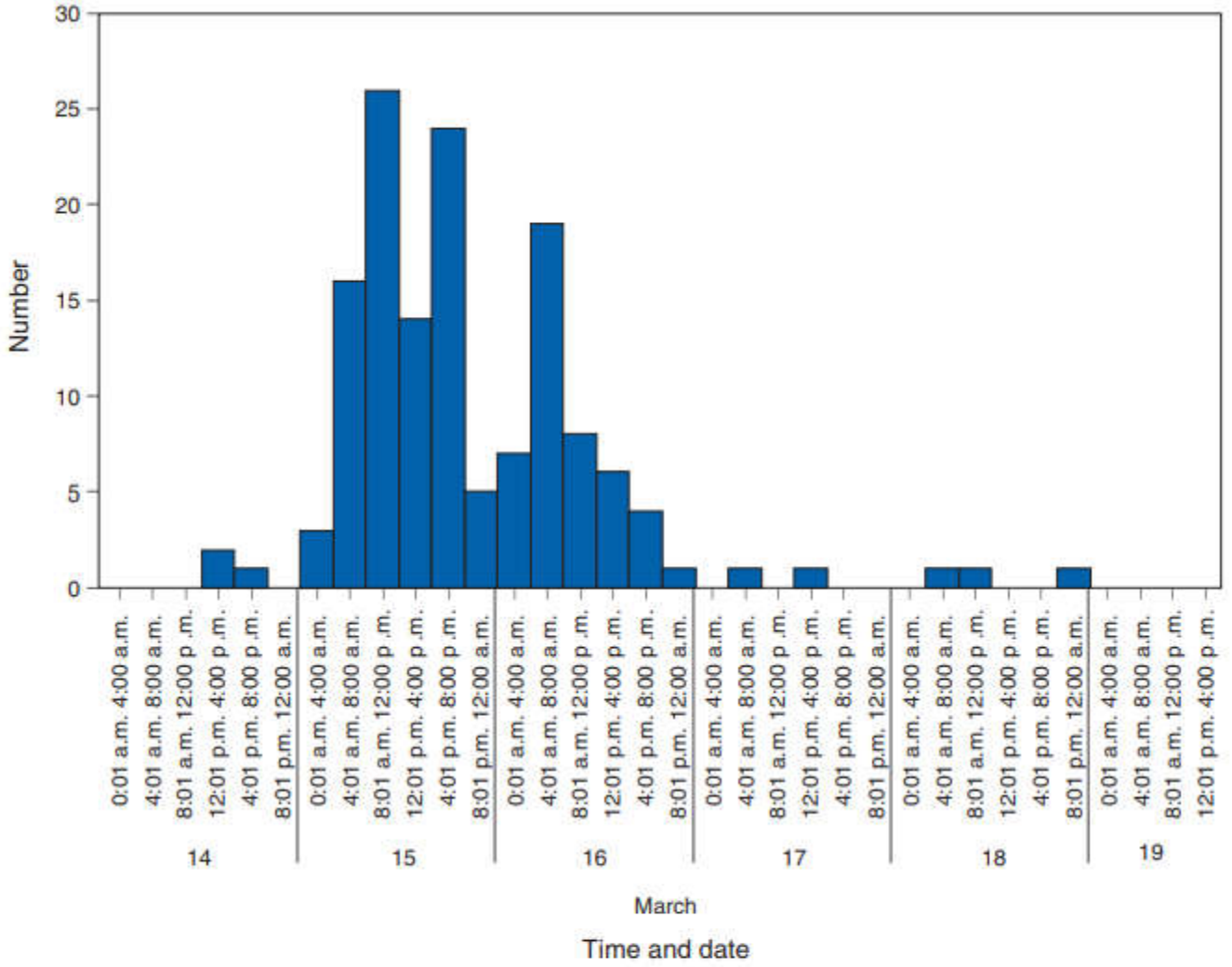


Figure 4.9
Example of histogram:
Number of reported cases of hepatitis A
by date of onset and residency status, Ogemaw County, April-May 1968



Epidemic curve of botulism outbreak, by time and date of symptom onset* — Nan province, Thailand, March 2006

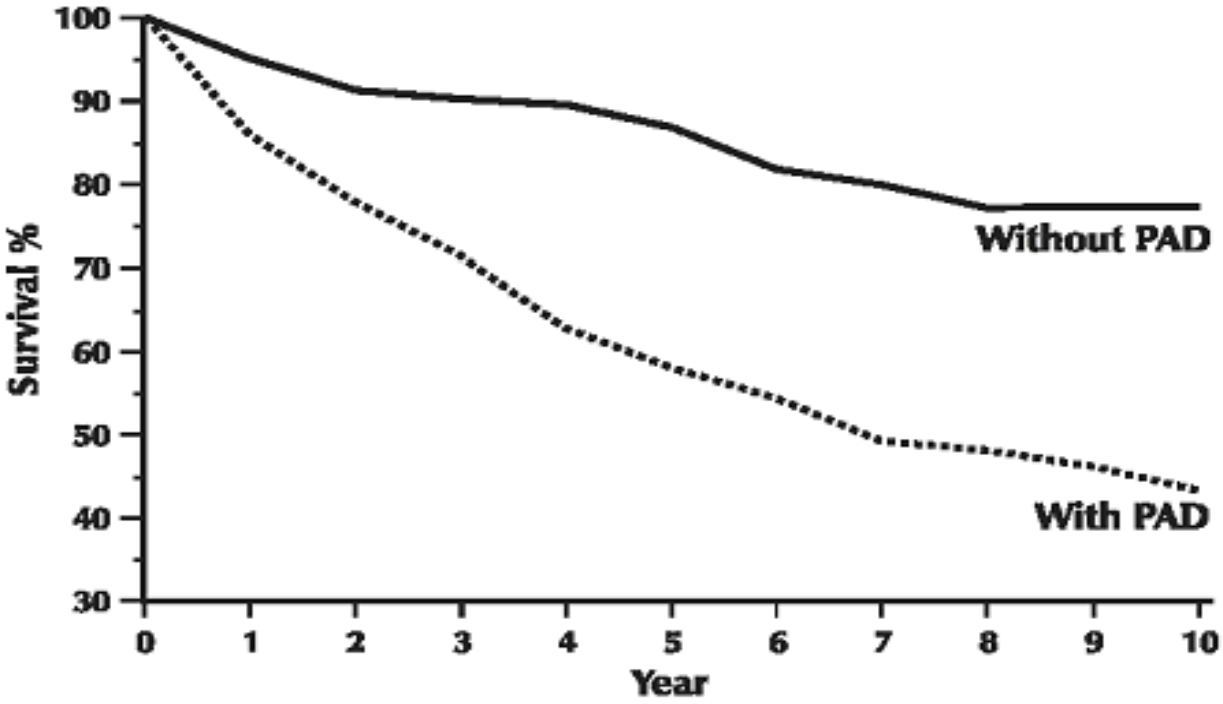


*Among patients for whom data regarding time of symptom onset were available (n = 145).

Source: https://www.cdc.gov/mmwr/PDF/wk/mm5514.pdf

Cumulative frequency and Survival curves

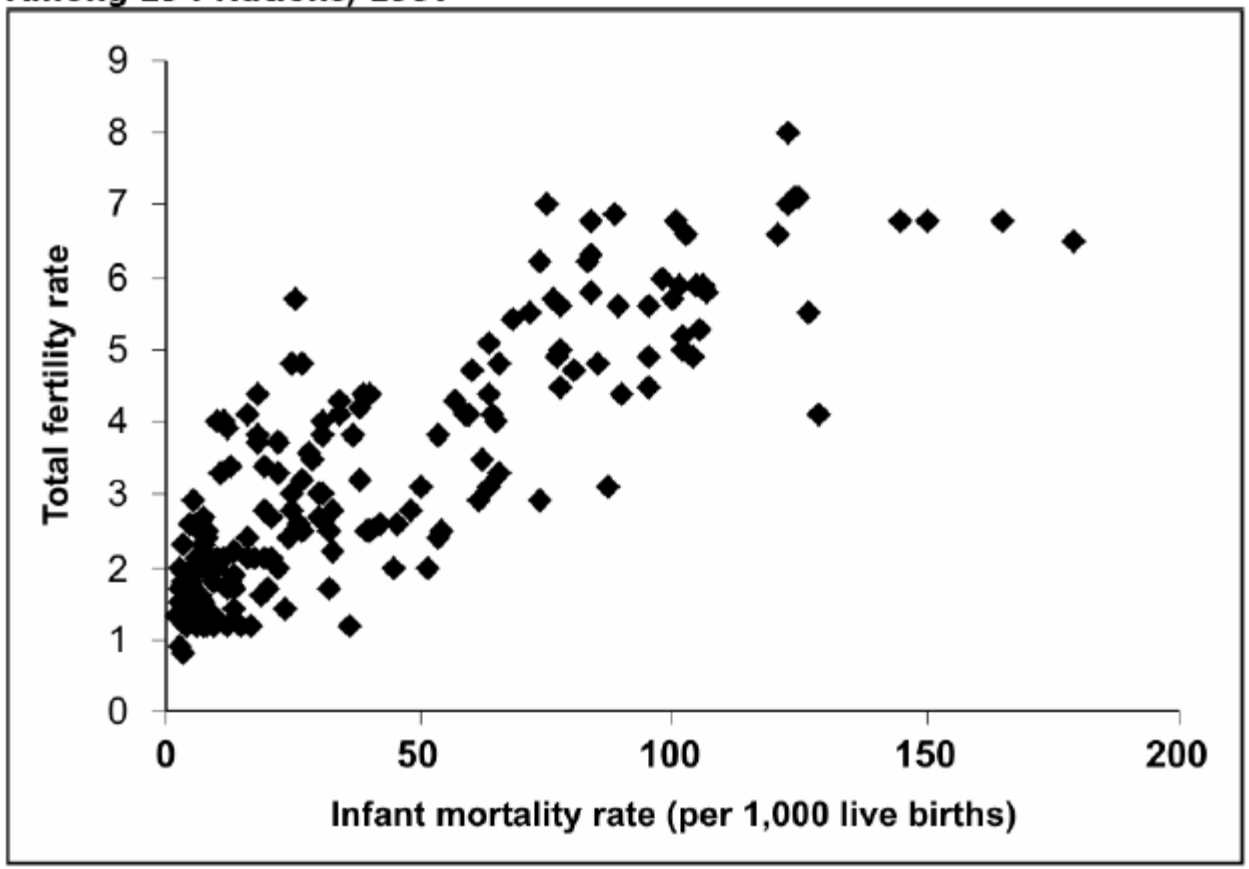
Figure 4.15
Survival curves for a cohort of patients with peripheral arterial disease (PAD) (n=482) and without PAD (n=262)
Pittsburgh, Pennsylvania, 1977-1985



- Find 5-year survivals among each group
- Number of years at 50% survival

Scatter diagrams or Scattergram

Figure 4.19 Correlation of Infant Mortality Rate and Total Fertility Rate Among 194 Nations, 1997

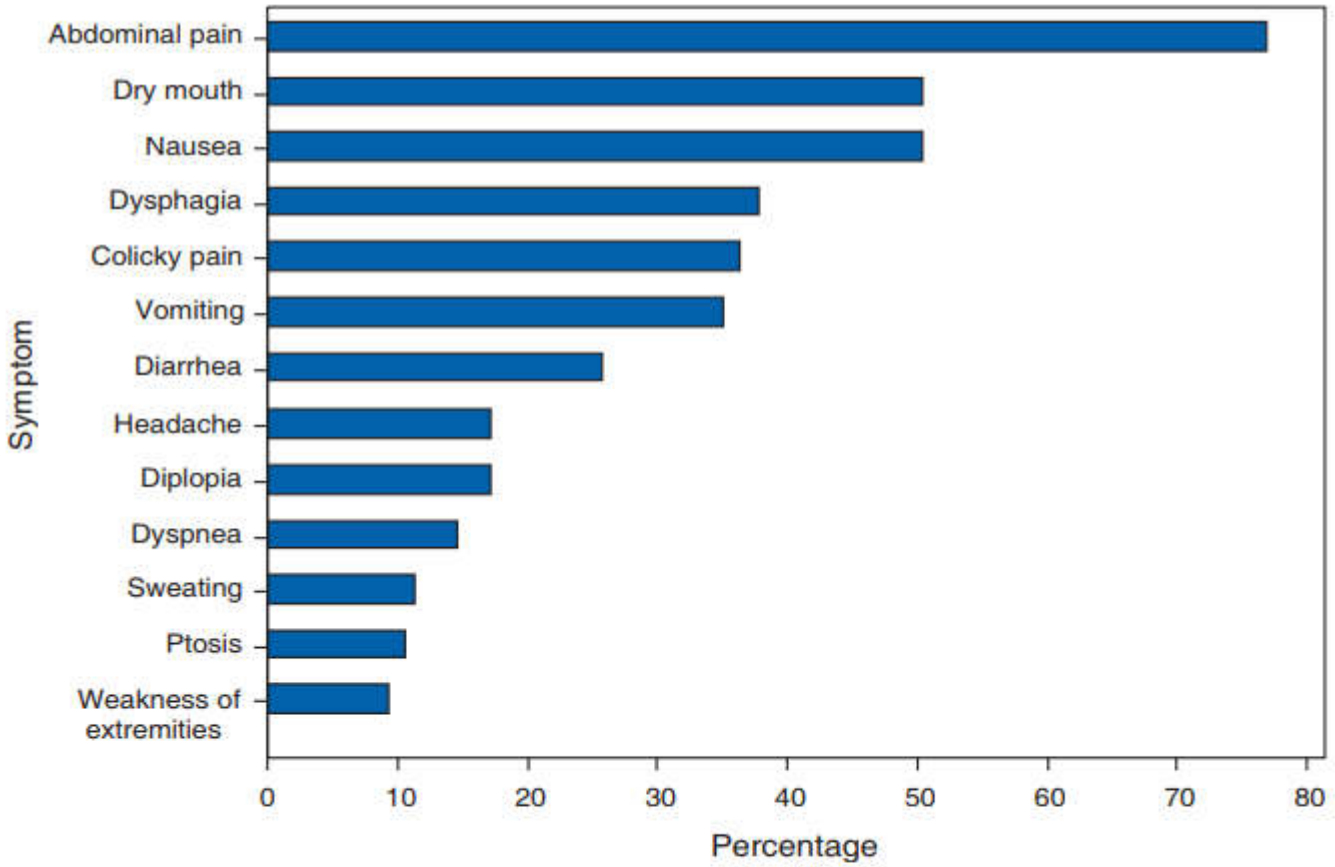


Data Source: Population Reference Bureau [Internet]. Datafinder [cited 2004 Dec 13]. Available from: http://www.prb.org/datafind/datafinder7.htm.

Ecological data
One dot = one nation

Bar chart

FIGURE 2. Percentage of patients* with symptoms of food poisoning caused by botulism, by symptom — Nan province, Thailand, March 2006



*Includes 141 hospitalized patients and 10 persons treated as outpatients who were queried about their symptoms (n = 151).

Source: https://www.cdc.gov/mmwr/PDF/wk/mm5514.pdf

Please interpret this table

Factors studied for association with HIV infection among female prostitutes in Chiang Mai, Thailand, August 1989

Variables	HIV-	HIV+
Frequency of intercourse per day in last week		
<=1	45	2
1.5-3.0	57	29
3.5-6.0	15	10
>6	27	44
Total	144	85

Source: modified from
"Risk factors for HIV among prostitutes in Chiangmai, Thailand." Taweesap Siraprapasiri, et al., AIDS 1991, 5:59-582

Please interpret this table

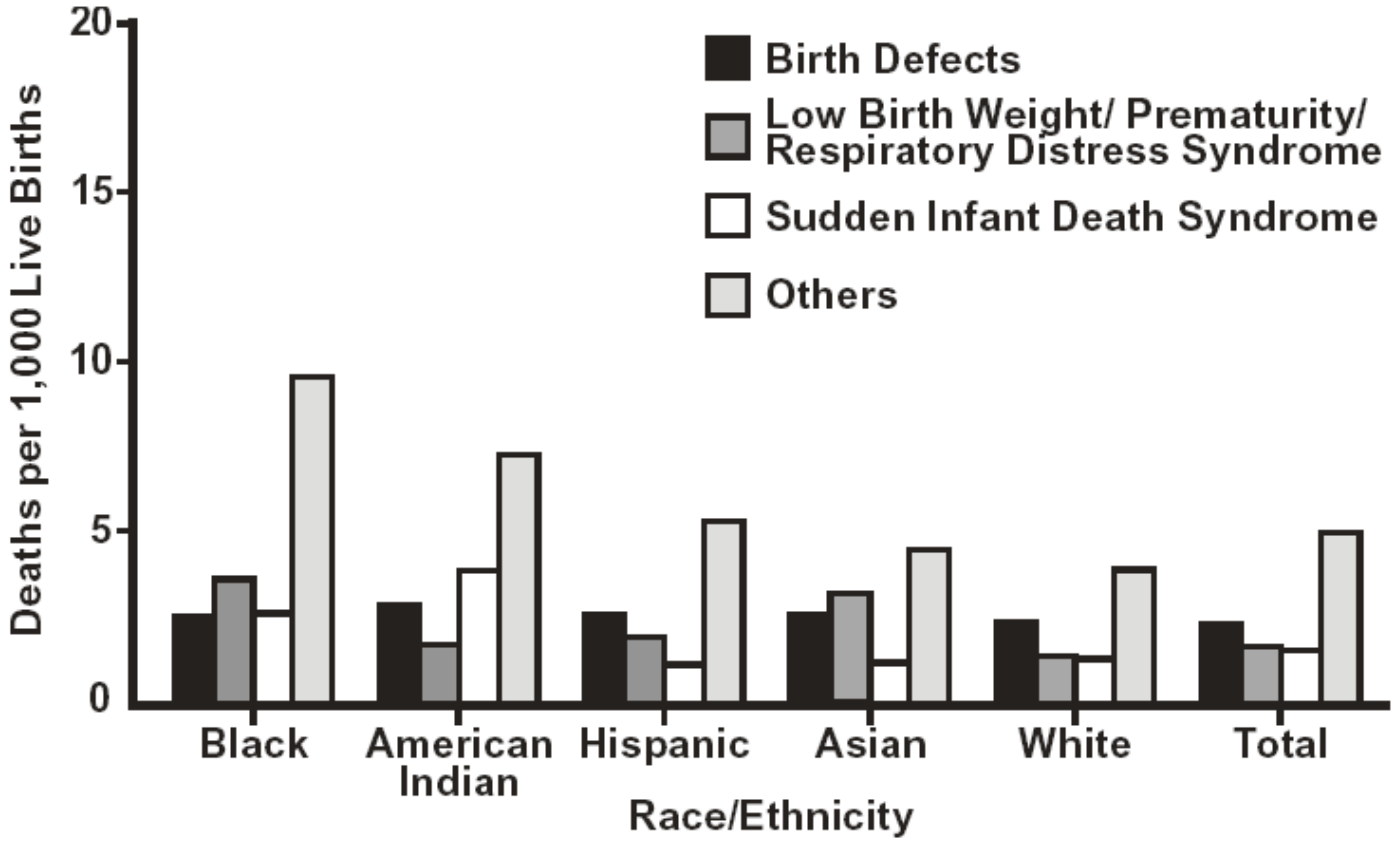
Factors studied for association with HIV infection among female prostitutes in Chiang Mai, Thailand, August 1989

Variables	HIV-	HIV+	Crude OR	(95% CI)	P-value
Frequency of intercourse per day in last week					
<=1	45	2	1		
1.5-3.0	57	29	11.4	(2.6-102.7)	
3.5-6.0	15	10	15.0	(2.6-149.1)	
>6	27	44	36.7	(8.2-326.1)	<.001
Total	144	85			
Charge for sex (Baht)					
>150	59	4			
51-150	47	13			
30-50	45	70			
Total	151	87			

Source: modified from
"Risk factors for HIV among prostitutes in Chiangmai, Thailand." Taweesap Siraprapasiri, et al., AIDS 1991, 5:59-582

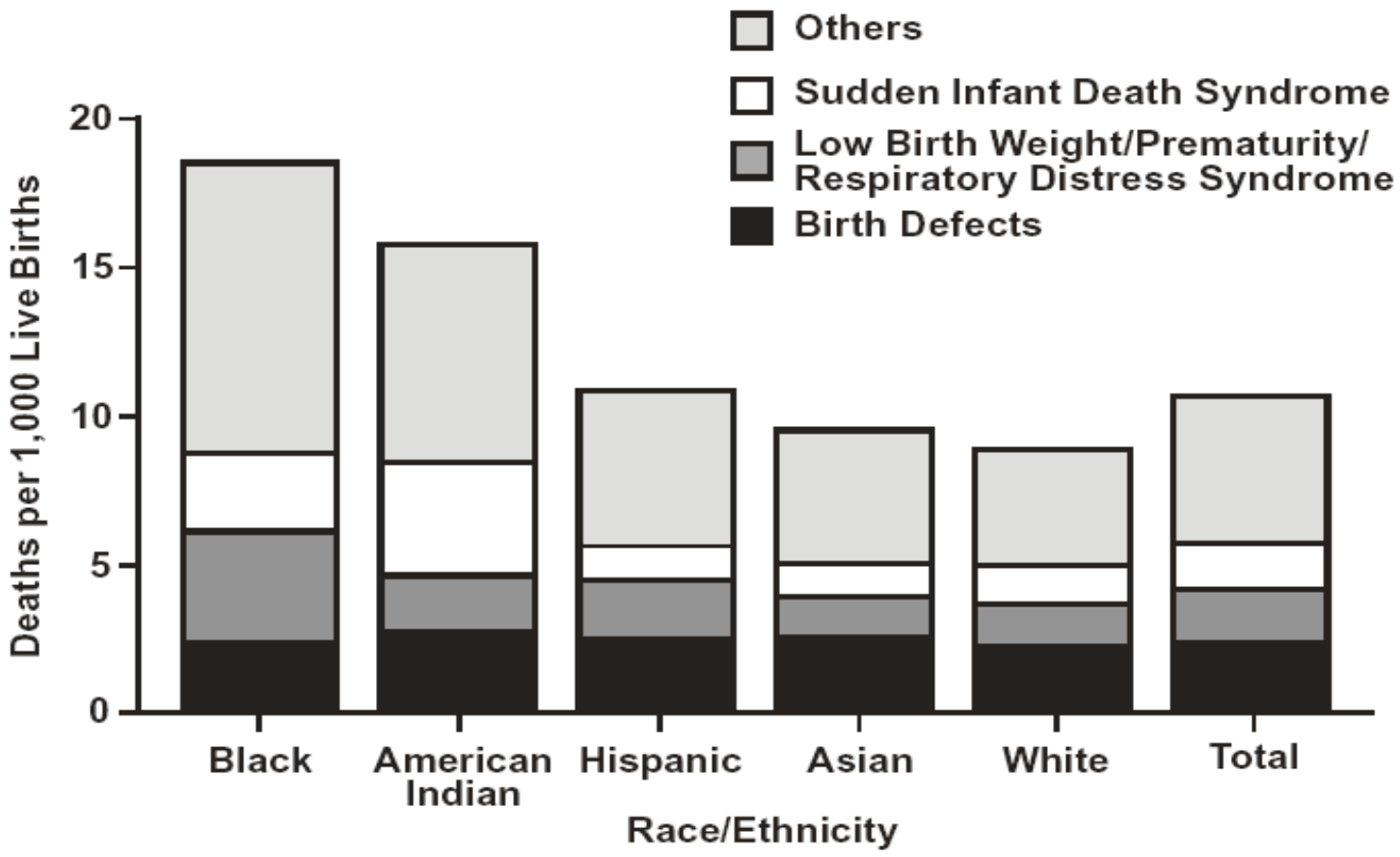
Grouped Bar Charts

Figure 4.18
Underlying cause of infant mortality among racial/ethnic groups, United States, 1983

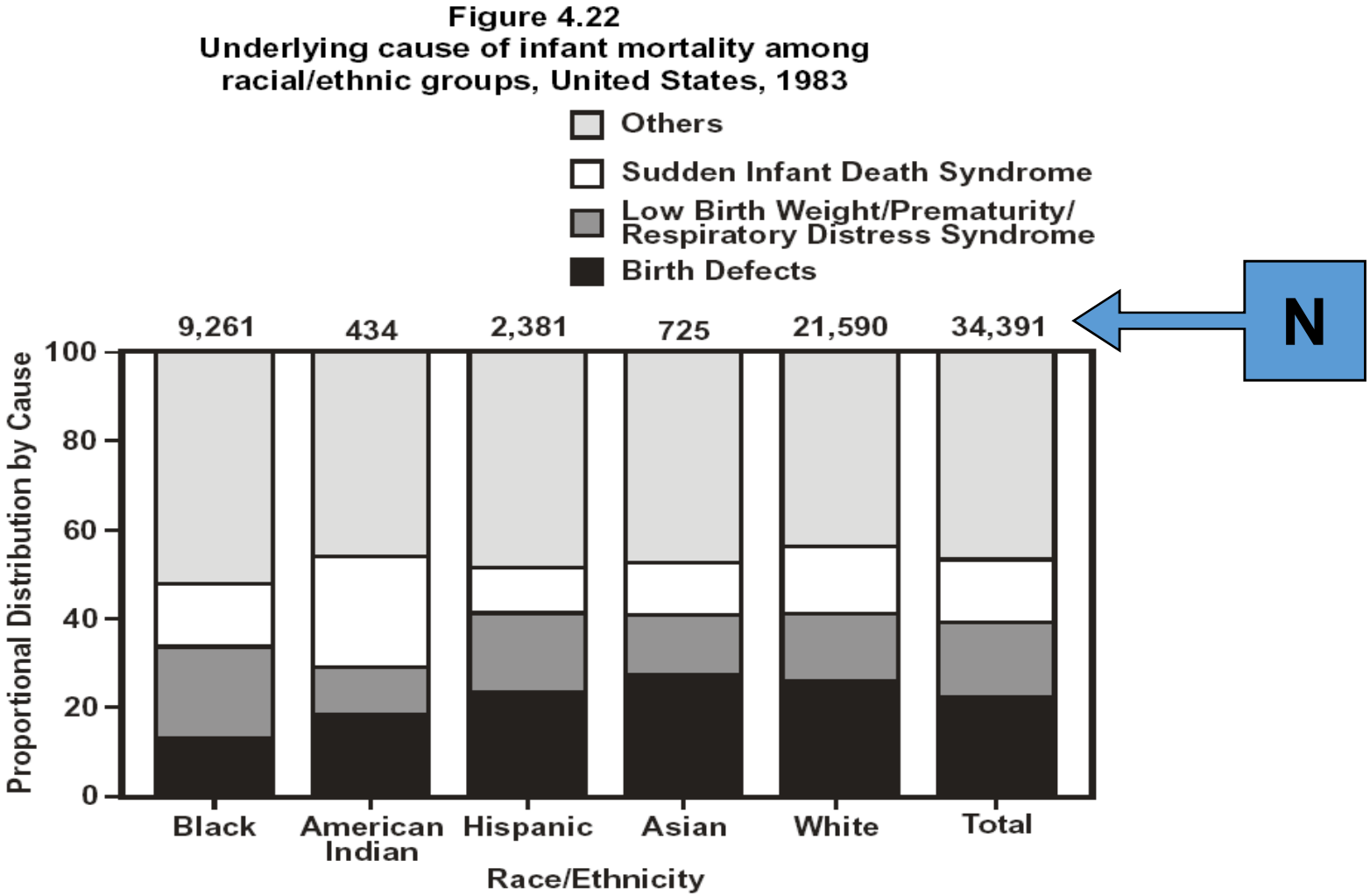


Stacked Bar Charts

Figure 4.20
Underlying causes of infant mortality among racial/ethnic groups, United States, 1983



100% Component Bar Charts



- In MS excel, changing chart format is very easy.

Combo: Bar and Line graph

Compare the heights of bars and show the total in line

Graph title with explanation!

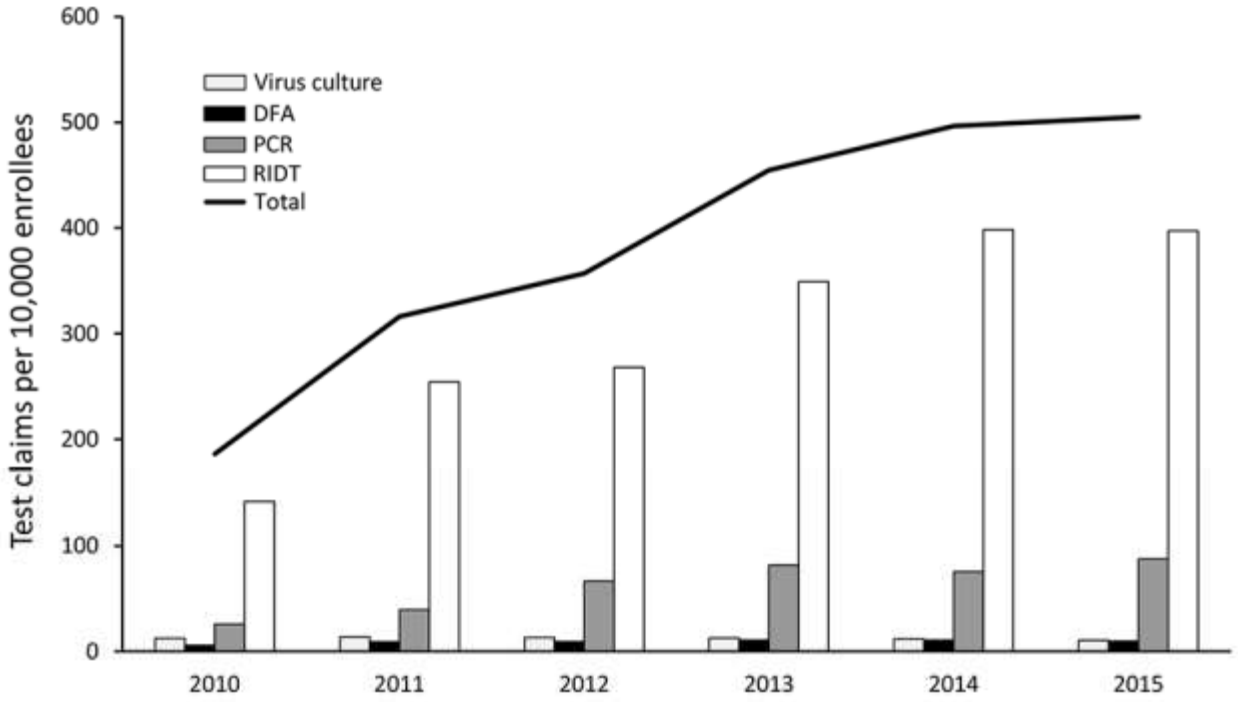
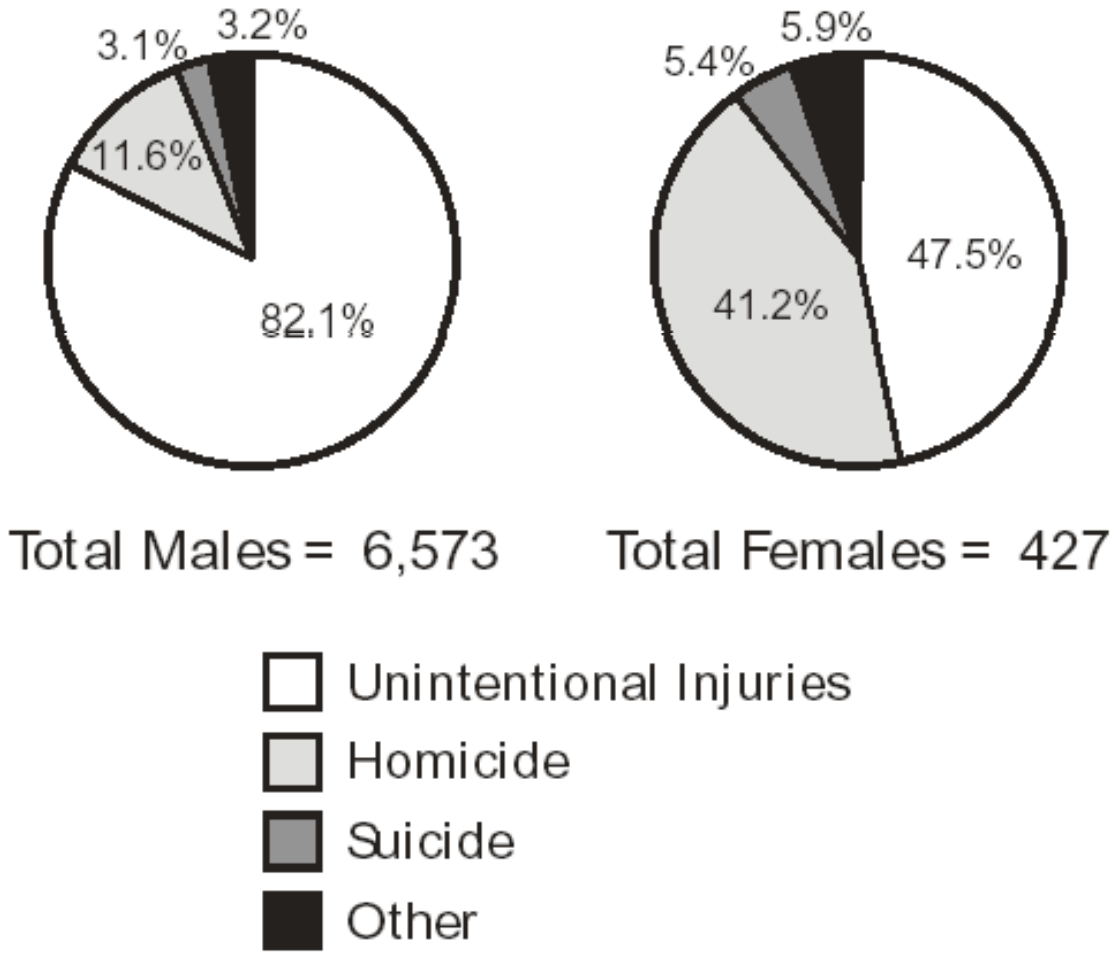


Figure 2. Number of influenza test claims per 10,000 enrollees in Truven Health Analytics' Database 2010–2015, demonstrating that the total number of influenza tests has consistently increased, with RIDTs comprising the largest percentage of tests. DFA, direct fluorescent antibody test; RIDT, rapid influenza diagnostic test.

Source: EID Vol. 24, No. 7, July 2018 DOI: https://doi.org/10.3201/eid2407.180028
https://wwwnc.cdc.gov/eid/article/24/7/pdfs/18-0028.pdf

Pie Charts

Figure 4.23
Manner of traumatic deaths for male and female workers in the United States, 1980-1985



Pie Charts

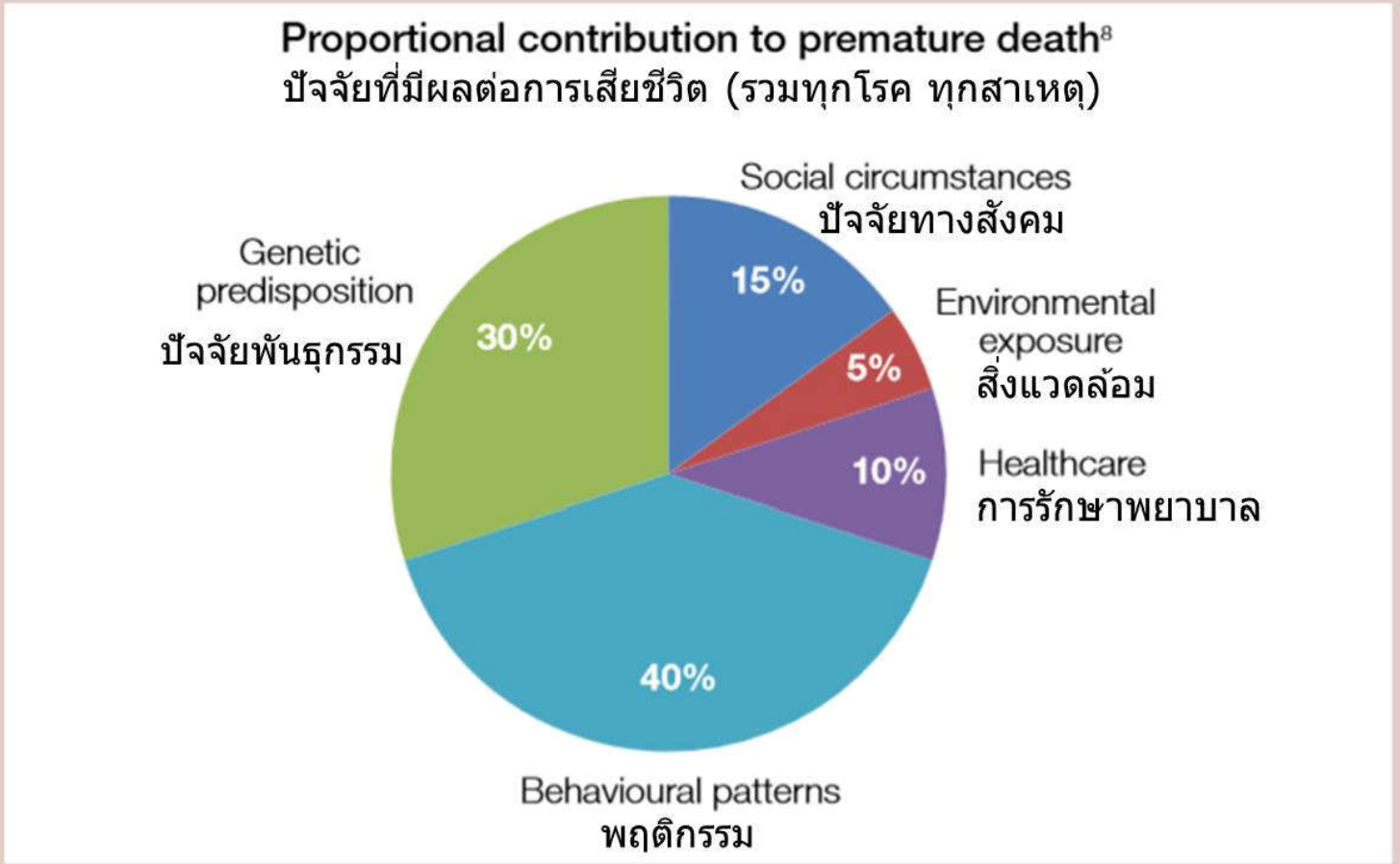


Figure 1 In the US, McGinnis et al show how healthcare plays an important though proportionately small role in preventing early deaths. Similar studies have supported these findings in the UK. Improving how we live our lives offers far greater opportunity for improving health.

Source: "From evidence into action: opportunities to protect and improve the nation's health." Public health England 2014

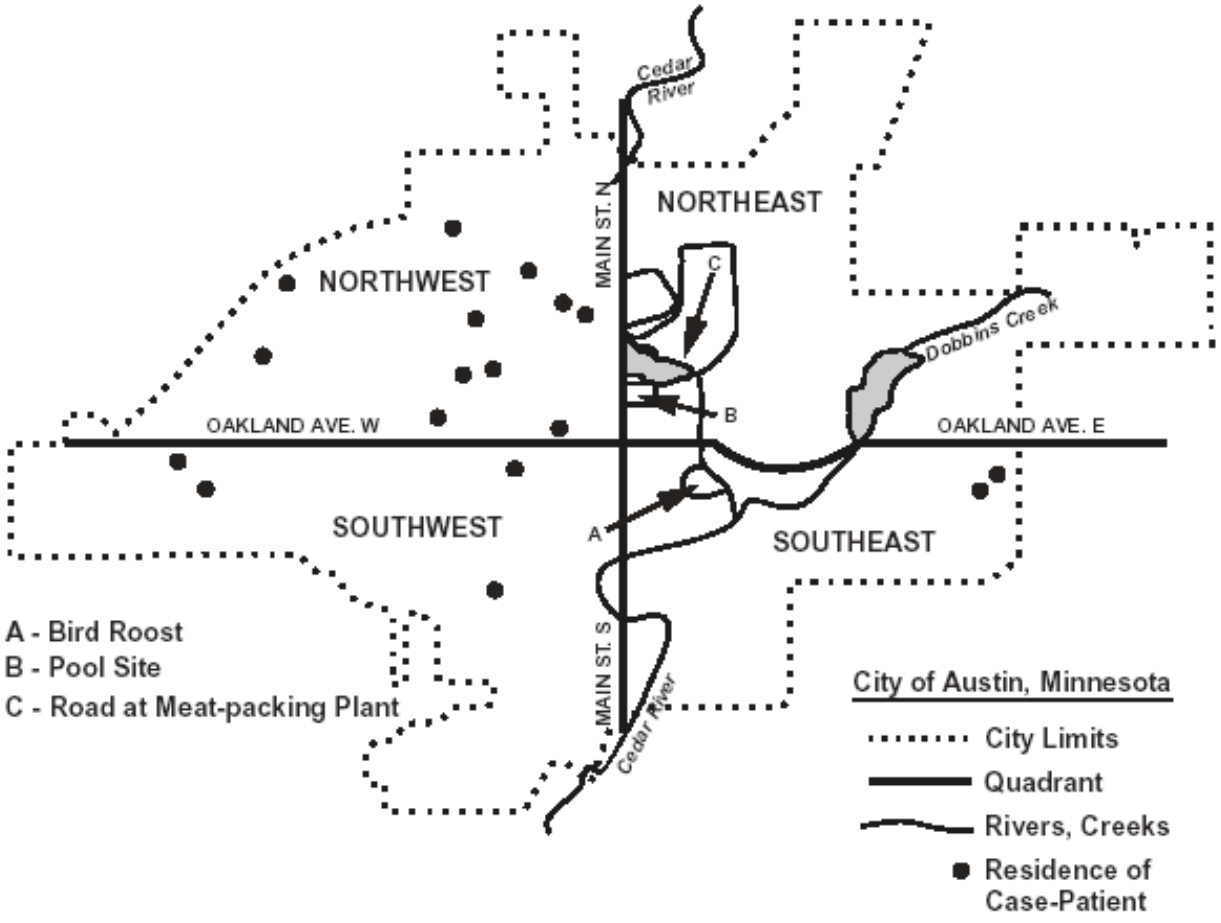
Maps

- Display data by geographic area
- Two types: spot map and area map

Data demonstration	**Spot Map**	**Area Map**
Number of cases	Yes	Yes
Case rate	No	Yes

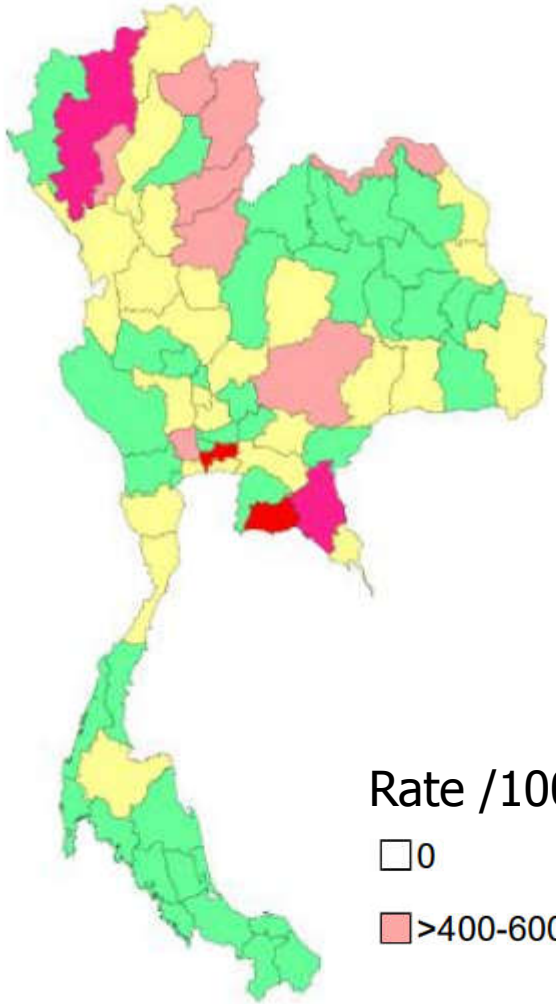
Spot Map

Figure 4.24
Example of spot map: Histoplasmosis by residence
Austin, Minnesota, October-November 1984



Area Map

Influenza morbidity rate
Thailand, 2017



0 1-200 >200-400
>400-600 >600-800 >800-1000

Area Map : combined with charts

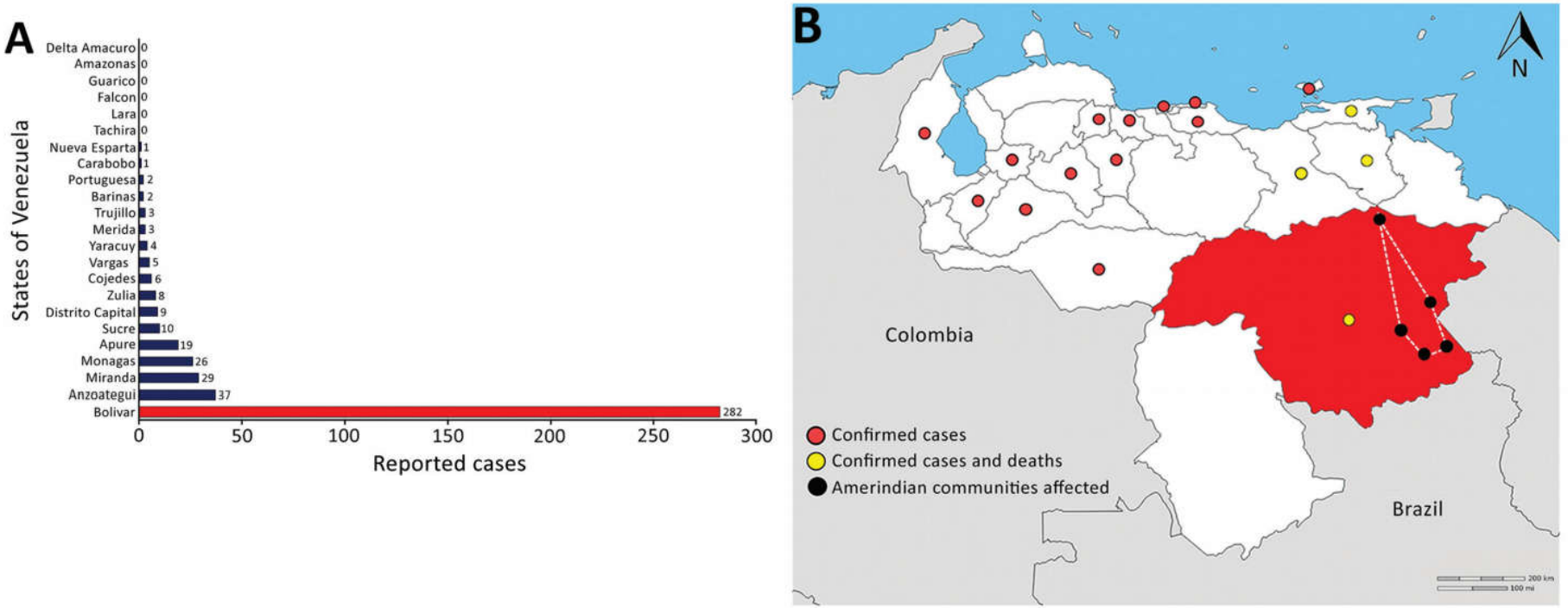
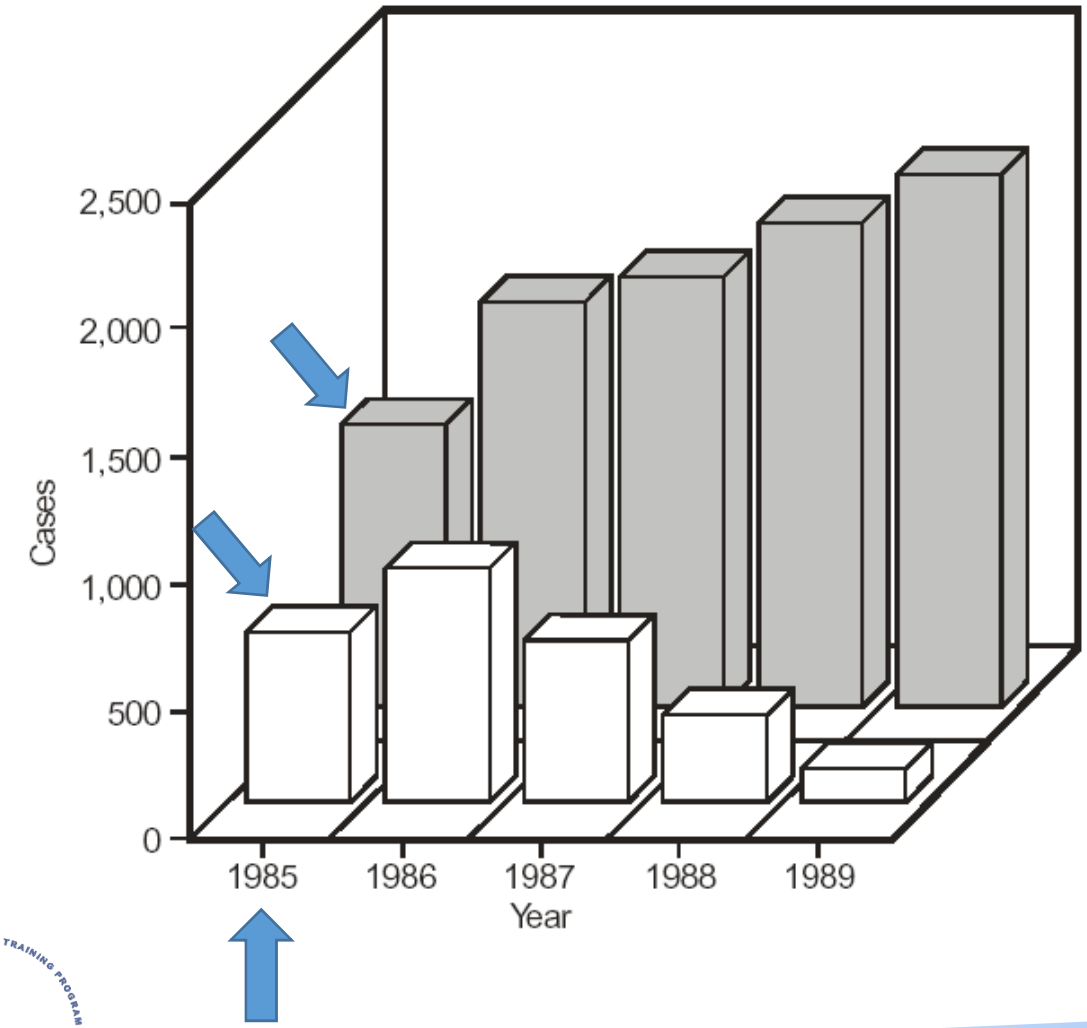


Figure 2. Suspected and confirmed diphtheria cases and deaths, by state, Venezuela, 2016–2017. The highest number of cases occurred in the state where Amerindians reside (Bolivar, red). A) Number of suspected cases of diphtheria reported from week 28 of 2016 through week 24 of 2017, by state. B) Location of confirmed cases and deaths, Venezuela, 2017. The affected Amerindian communities reside in the area within the dotted line. Map obtained from d-maps (http://d-maps.com/carte. php?num_car=4080&lang=es).

https://wwwnc.cdc.gov/eid/article/24/7/pdfs/17-1712.pdf

Pitfalls: 3-Dimension

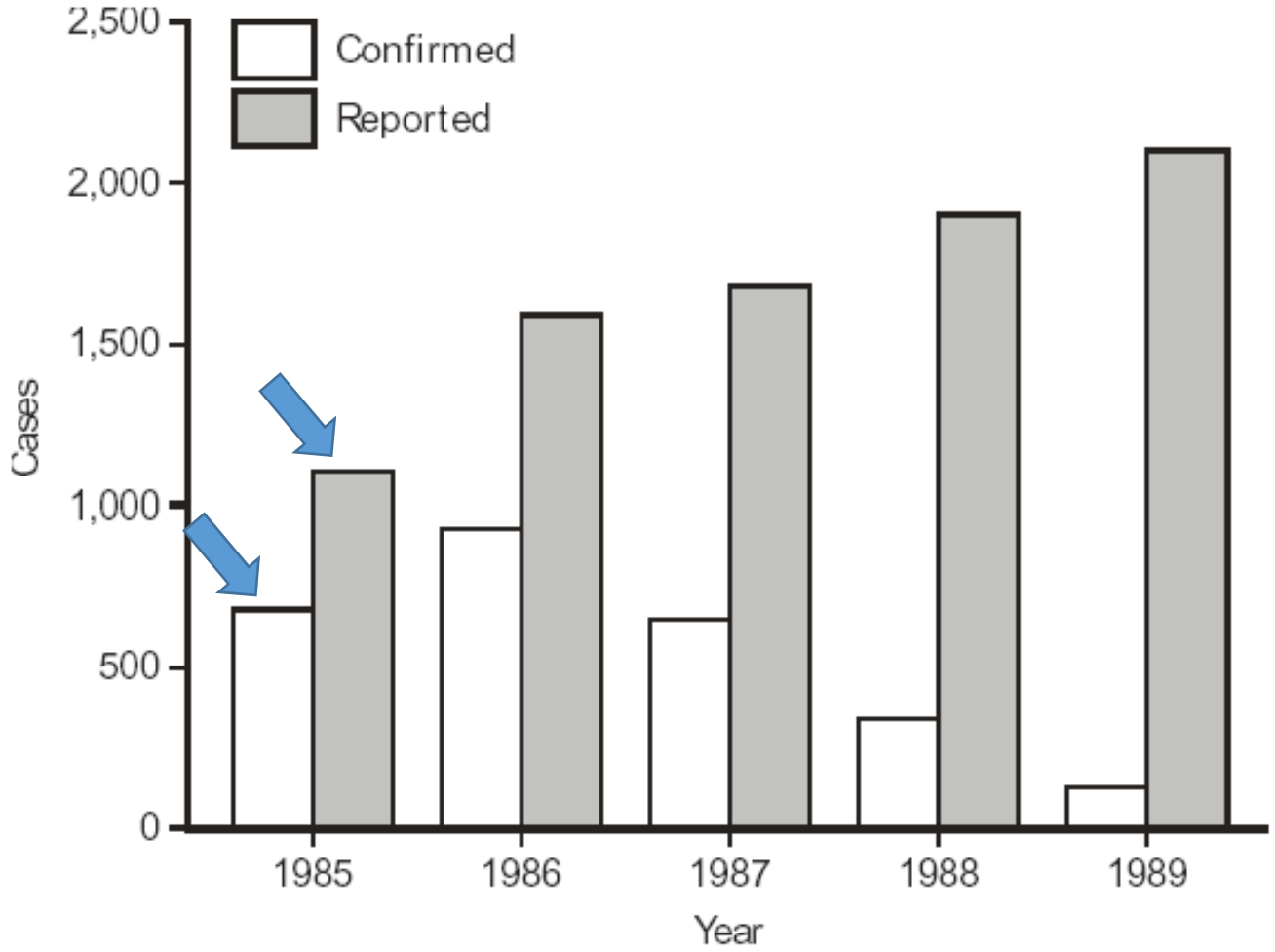
Figure 4.28b
Example of three-dimensional bar chart:
Reported and confirmed polio cases by year, the Americas, 1985-1989



- In 1985 What is the height ratio of the white bar to the grey bar?

  Is it ½?

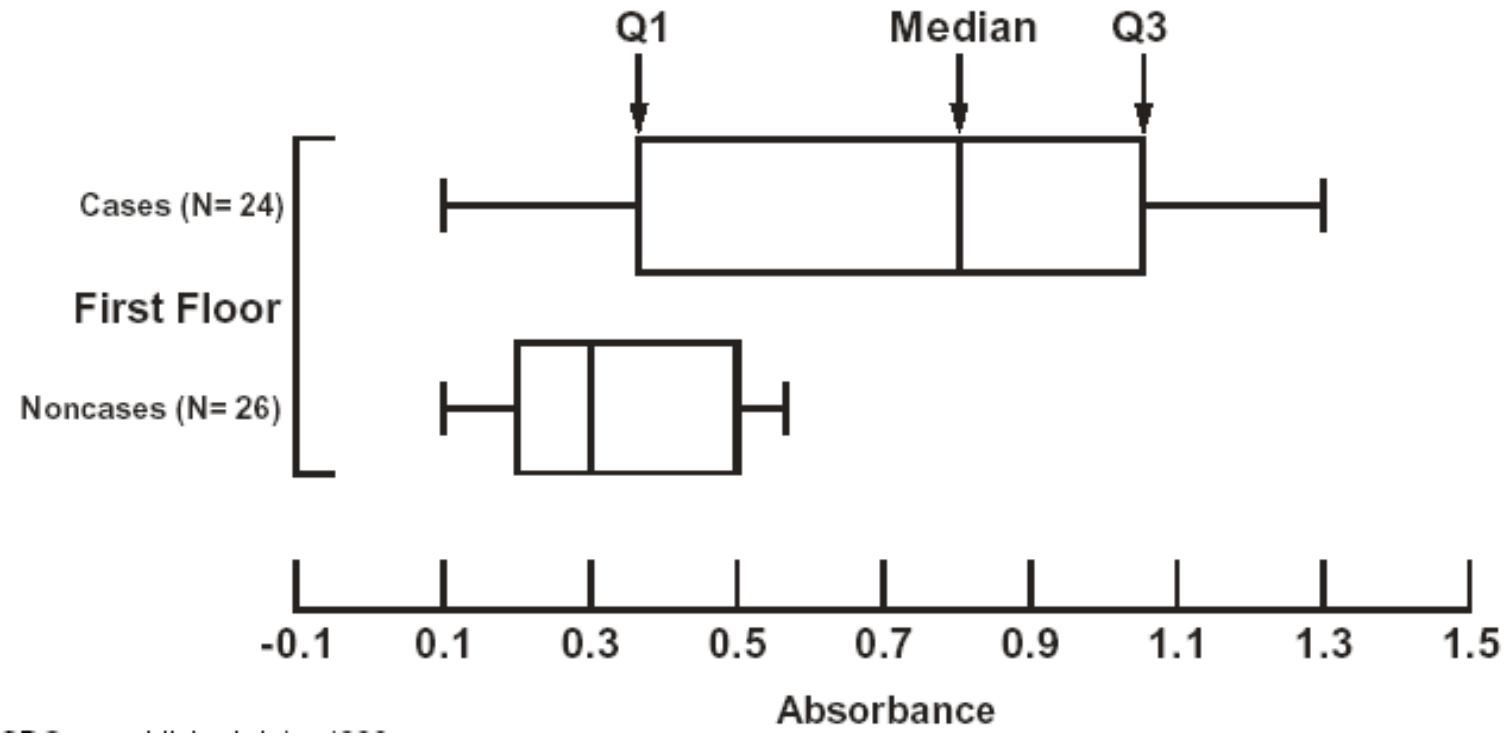
Figure 4.28a
Example of two-dimensional bar chart:
Reported and confirmed polio cases by year, the Americas, 1985-1989



- Accurately interpretation in 2-D presentation → 2/3

Box Plots

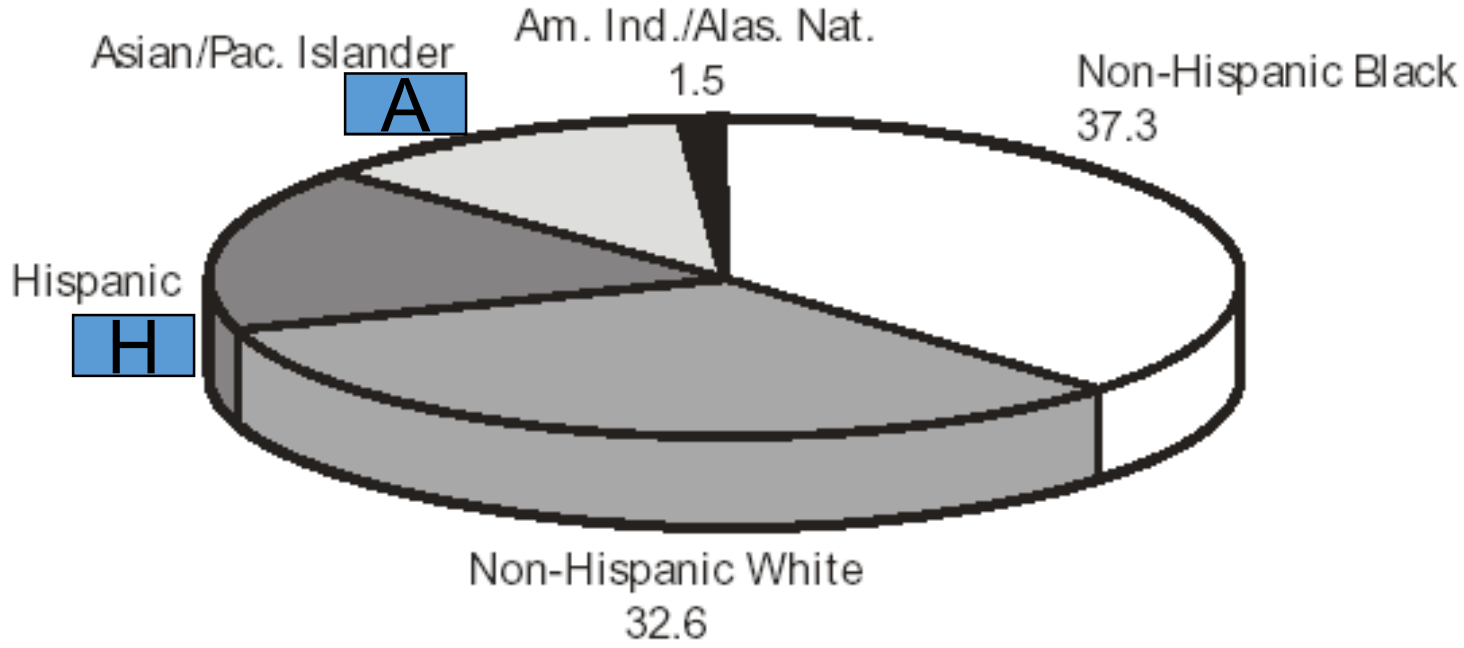
Figure 4.27
Example of box plot: Results of indirect ELISA for IgG antibodies to parainfluenza type I virus in convalescent phase serum specimens from cases to noncases, Baltimore County, Maryland, January 1990



- Display the distribution of data
- Commonly used in some statistics programs

Compare "A" and "H" area: which one is bigger?

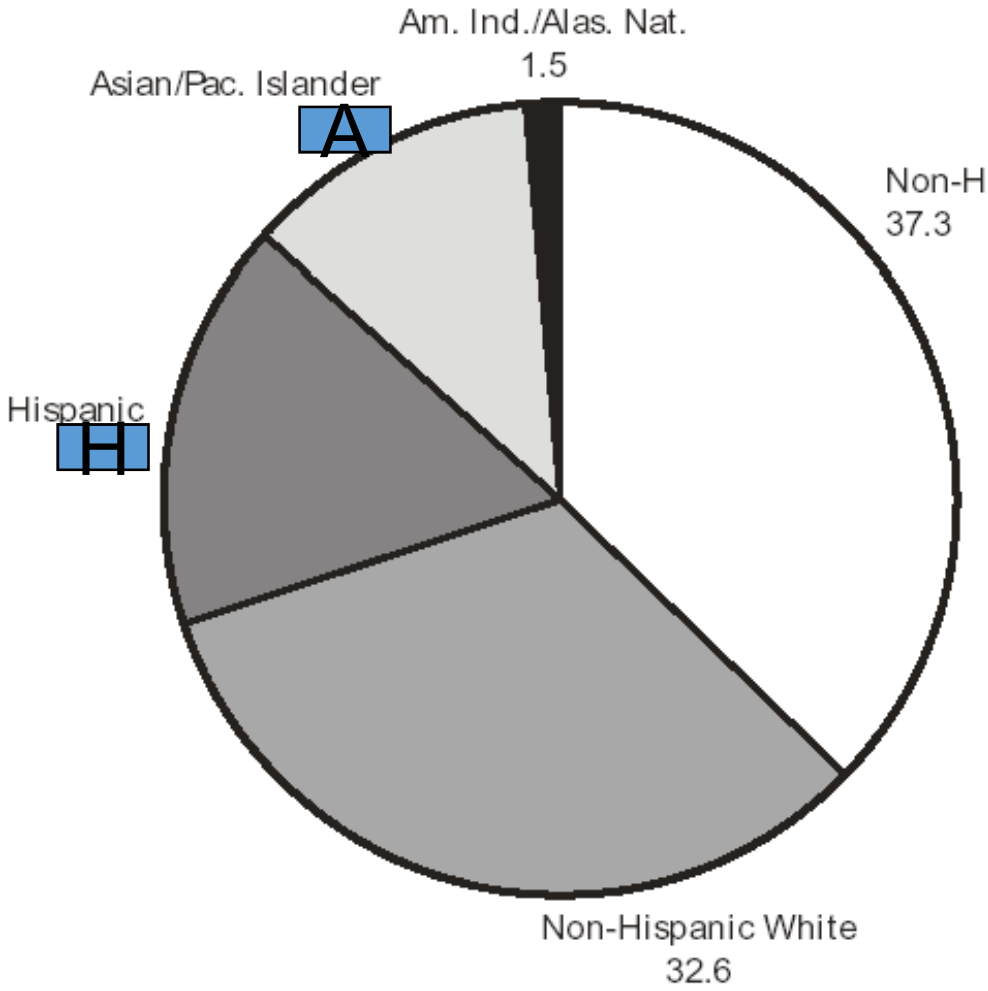
Figure 4.29b
Example of three-dimensional pie chart:
Percentage of tuberculosis cases by race and ethnicity,
United States, 1989 (n = 23,495)



It looks like A ≥ H

Compare "A" and "H" area: which one is bigger?

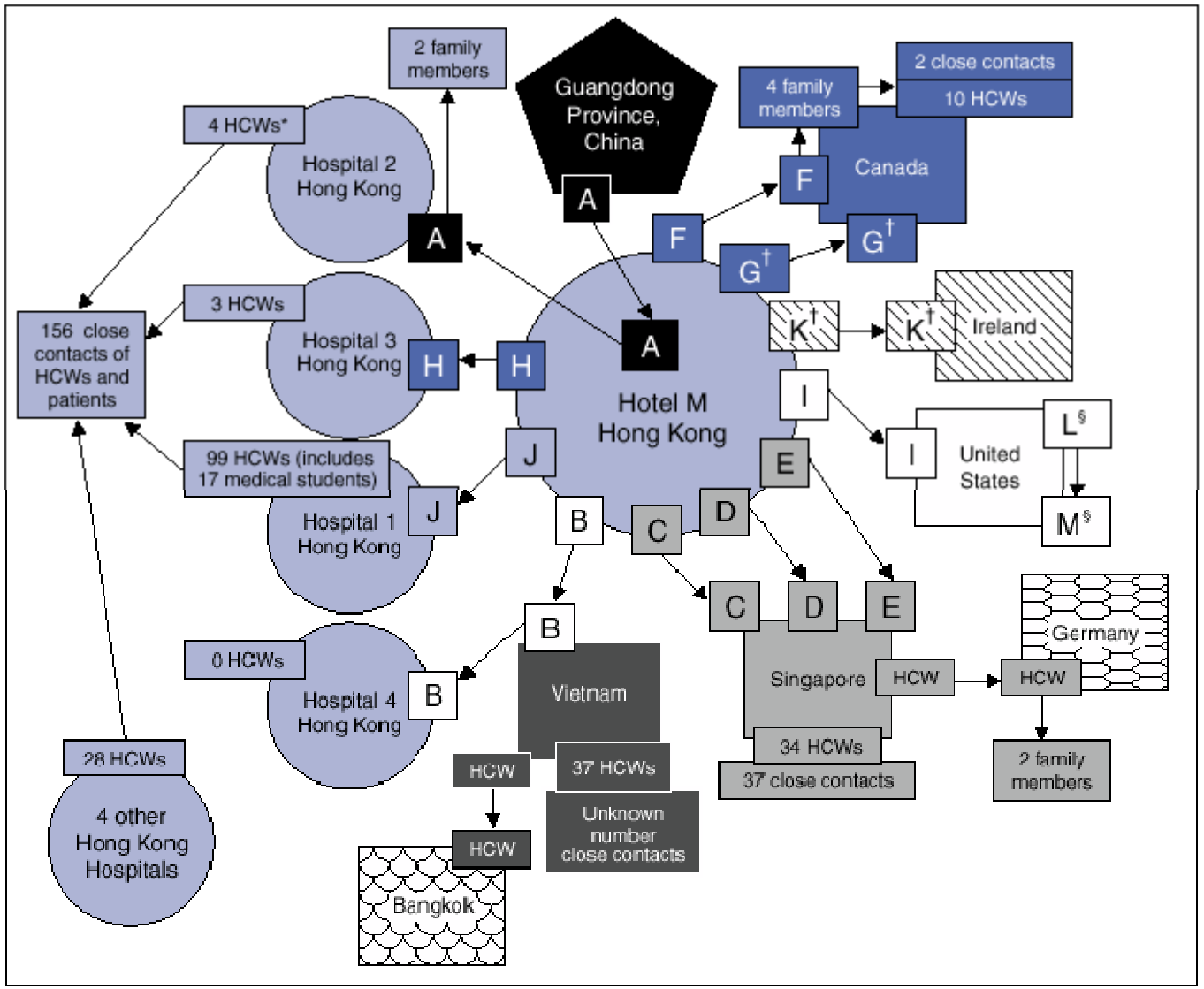
Figure 4.29a
Example of two-dimensional pie chart:
Percentage of tuberculosis cases by race and ethnicity,
United States, 1989 (n = 23,495)



Correctly interpreted in 2-D pie chart:
H >A

Diagram : SARS 2003 in HK

FIGURE 1. Chain of transmission among guests at Hotel M — Hong Kong, 2003



* Health-care workers.
† All guests except G and K stayed on the 9th floor of the hotel. Guest G stayed on the 14th floor, and Guest K stayed on the 11th floor.
§ Guests L and M (spouses) were not at Hotel M during the same time as index Guest A but were at the hotel during the same times as Guests G, H, and I, who were ill during this period.

source : https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5212a1.htm

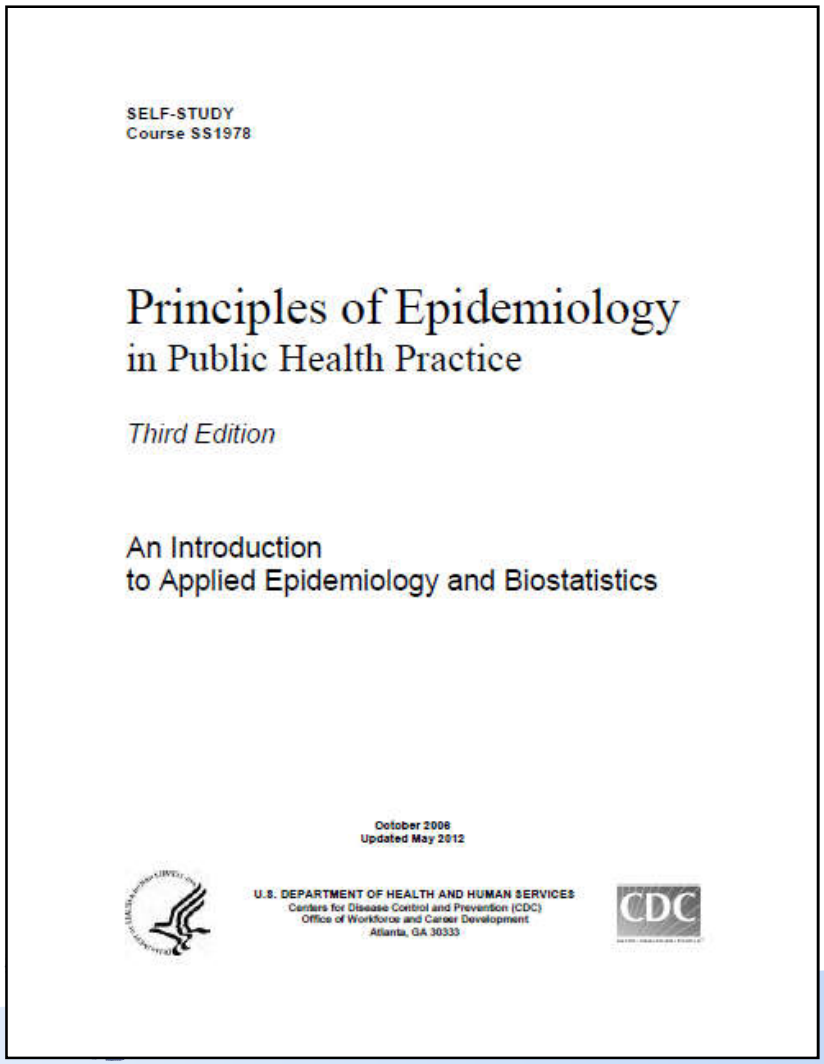
Excel practice

- Line graph
- Bar chart: vertical, horizontal
- Histogram, Epidemic curve
- Pie chart
- Etc.

- Copy from excel -> paste special as picture**

Conclusion in your own words

Thank you



Suggested reading:

- Lesson Four: Displaying Public Health Data
- Outbreak investigation and surveillance report

  www.osirjournal.net

  www.cdc.gov/mmwr
- https://www.cdc.gov/mmwr/PDF/wk/mm4821.pdf
- https://www.cdc.gov/mmwr/PDF/wk/mm5514.pdf