



Displaying Public Health Data

- Introduction to tables and graphs
- Constructing epidemiological tables
- Constructing epidemiological graphs
- Other forms of data display
- Applying computer technology to data management

Christopher W. Blackwell, Ph.D., ARNP-C
Assistant Professor, College of Nursing
University of Central Florida

NGR 7642: Epidemiology Principles in Advanced Practice Nursing

Introduction to Tables and Graphs

- Data can be displayed as:
 - Nominal: No intrinsic order and differences between variables have no meaning (sex, race, exposure as yes/no)
 - Ordinal: Do have intrinsic order; but differences again have no meaning (low, medium, high; age ranges)
 - Discrete: Values are integers (number of ill persons exposed to a risk factor)
 - Continuous: Value in any range (amt. of time between meal being served and onset of GI s/s; infant morality rate)
 - Answer ?s helps to determine what type of data you want to display:
 - Are you showing a change from past patters in the data?
 - Are you showing a difference in geographic area or some predetermined risk factor?
 - What interpretation do you want the reader to reach?



Introduction to Tables and Graphs

- To begin analyzing data:
 - Start by examining records (eg: constructing a line listing)
 - Next, summarize the data as a table
 - Next, if necessary, graph or chart the data to assess for trends and variations in trends
- Tables:
 - A table is a set of data arranged in rows and columns.
 - Almost any quantitative information can be organized into a table.
 - Useful for demonstrating patterns, exceptions, differences, and other relationships.
 - Serve as the basis for preparing additional visual displays of data, such as graphs and charts, in which some of the details may be lost.
 - Tables designed to present data to others should be as simple as possible.
 - Two or three small tables, each focusing on a different aspect of the data, are easier to understand than a single large table that contains many details or variables.
 - PUBLICATION TIP: A table in a printed publication should be self-explanatory.
 - If a table is taken out of its original context, it should still convey all the information necessary for the reader to understand the data



Introduction to Tables and Graphs

- **More About Constructing Tables:**
 - Use a clear and concise title that describes person, place and time — what, where, and when — of the data in the table. Precede the title with a table number (use APA format)
 - Label each row and each column and include the units of measurement for the data (for example, years, mm Hg, mg/dl, rate per 100,000)
 - Show totals for rows and columns, where appropriate. If you show percentages (%), also give their total (always 100)
 - Identify missing or unknown data either within the table (for example, Table 4.11) or in a footnote below the table
 - Explain any codes, abbreviations, or symbols in a footnote (for example, Syphilis P&S = primary and secondary syphilis)
 - Note exclusions in a footnote (e.g., 1 case and 2 controls with unknown family history were excluded from this analysis)
 - Note the source of the data below the table or in a footnote if the data are not original



Introduction to Tables and Graphs

- APA Format for Tables:

Table #: *Title of the Table*

<u>Descriptor</u>	<u>Variable 1</u>	<u>Variable 2</u>	<u>Variable 3</u>
Data	Data	Data	Data
Data	Data	Data	Data
Data	Data	Data	Data



Introduction to Tables and Graphs

- **One-variable tables:**

- In descriptive epidemiology, the most basic table is a simple frequency distribution with only one variable, such as Table 4.1a
- Displays number of reported syphilis cases in the United States in 2002 by age group.² (Frequency distributions are discussed in Lesson 2)
- In this type of frequency distribution table, the first column shows the values or categories of the variable represented by the data, such as age or sex. The second column shows the number of persons or events that fall into each category
- In constructing any table, the choice of columns results from the interpretation to be made
- In Table 4.1a, the point the analyst wishes to make is the role of age as a risk factor of syphilis
- Thus, age group is chosen as column 1 and case count as column 2



Introduction to Tables and Graphs

- Often, an additional column lists the percentage of persons or events in each category (see Table 4.1b)
- The percentages shown in Table 4.1b actually add up to 99.9% rather than 100.0% due to rounding to one decimal place
- Rounding that results in totals of 99.9% or 100.1% is common in tables that show percentages
- Nonetheless, the total percentage should be displayed as 100.0%, and a footnote explaining that the difference is due to rounding should be included
- The addition of percent to a table shows the relative burden of illness; for example, in Table 4.1b, we see that the largest contribution to illness for any single age category is from 35–39 year olds
- The subsequent addition of cumulative percent (e.g., Table 4.1c) allows the public health analyst to illustrate the impact of a targeted intervention
- Here, any intervention effective at preventing syphilis among young people and young adults (under age 35) would prevent almost half of the cases in this population



Introduction to Tables and Graphs

- The one-variable table can be further modified to show cumulative frequency and/or cumulative percentage, as in Table 4.1c
- From this table, you can see at a glance that 46.7% of the primary and secondary syphilis cases occurred in persons younger than age 35 years, meaning that over half of the syphilis cases occurred in persons age 35 years or older
- Note that the choice of age groupings will affect the interpretation of your data



Introduction to Tables and Graphs

Table 4.1a: Reported Cases of Primary and Secondary Syphilis by Age—United States, 2002

<u>Age Group (years)</u>	<u>Number of Cases</u>
<14	21
15–19	351
20–24	842
25–29	895
30–34	1,097
35–39	1,367
40–44	1,023
45–54	982
≥55	284
Total	6,862

Data Source: Centers for Disease Control and Prevention (2003). *Sexually Transmitted Disease Surveillance 2002*. Atlanta: U.S. Department of Health and Human Services.



Introduction to Tables and Graphs

Table 4.1b: Reported Cases of Primary and Secondary Syphilis by Age—United States, 2002

<u>Age Group (years)</u>	<u>CASES</u>	
	<u>Number</u>	<u>Percent</u>
<14	21	0.3
15–19	351	5.1
20–24	842	12.3
25–29	895	13.0
30–34	1,097	16.0
35–39	1,367	19.9
40–44	1,023	14.9
45–54	982	14.3
≥55	284	4.1
Total	6,862	100.0*

Actual total of percentages for this table is 99.9% and does not add to 100.0% due to rounding error.

Data Source: Centers for Disease Control and Prevention. (2003). *Sexually Transmitted Disease Surveillance 2002*. Atlanta: U.S. Department of Health and Human Services.



Introduction to Tables and Graphs

Table 4.1c: *Reported Cases of Primary and Secondary Syphilis by Age—United States, 2002*

<u>Age Group (years)</u>	<u>CASES</u>		
	<u>Number</u>	<u>Percent</u>	<u>Cumulative Percent</u>
<14	21	0.3	0.3
15–19	351	5.1	5.4
20–24	842	12.3	17.7
25–29	895	13.0	30.7
30–34	1,097	16.0	46.7
35–39	1,367	19.9	66.6
40–44	1,023	14.9	81.6
45–54	982	14.3	95.9
≥55	284	4.1	100.0
Total	6,862	100.0 *	100.0 *

* Percentages do not add to 100.0% due to rounding error.

Data Source: Centers for Disease Control and Prevention. (2003). *Sexually Transmitted Disease Surveillance 2002*. Atlanta: U.S. Department of Health and Human Services.



Introduction to Tables and Graphs

- Two- and Three-Variable Tables:
 - Data can be cross-tabulated to show counts by an additional variable
 - A two-variable table w/ data categorized jointly by those two variables is a Contingency Table:
 - Contingency Tables are also called “Two by Twos” and are useful in comparing persons with to without an exposure and those with and without a disease:

Table 4.3: *Generator Location and Risk of CO Poisoning after an Ice Storm—Maine 1998*

		Number of		Total
		Cases	Controls	
Generator Location	Inside Home or Attached Structure	23	23	46
	Outside Home	4	139	143
Total		27	162	189



Introduction to Tables and Graphs

- This is the generic form of the 2x2 Table:

Table 4.3: *Generator Location and Risk of CO Poisoning after an Ice Storm—Maine 1998*

	Ill	Well	Total	Attack Rate (Risk)
Exposed	a	b	$a + b = H_1$	$a / a + b$
Not-Exposed	c	d	$c + d = H_0$	$v / c + d$
Total	$a + c = V_1$	$b + d = V_0$	T	V_1 / T



Introduction to Tables and Graphs

- When presenting information in-print or projection, limit the variables to 1-2 unless a third indicates an interactive effect
- Because three-way tables are often difficult to understand, only use them when you can devote a lot of time to explanation
- Table 4.5 in the CDC manual depicts this:
 - When do you think this might be essential?
 - Let's practice: Exercise 4.1:

The data in Table 4.6 describe characteristics of the 38 persons who ate food at or from a church supper in Texas in August 2001. Fifteen of these persons later developed botulism 5A. Construct a table of the illness (botulism) by age group. Use botulism status (yes/no) as the column labels and age groups as the row labels.



Introduction to Tables and Graphs

The data in Table 4.6 describe characteristics of the 38 persons who ate food at or from a church supper in Texas in August 2001. Fifteen of these persons later developed botulism 5A. Construct a table of the illness (botulism) by age group. Use botulism status (yes/no) as the column labels and age groups as the row labels:

Table A: *Botulism by Age Group , TX Church Supper Outbreak, 2001*

Age Group (Years)	Botulism Status	
	Yes	No
≤9	2	2
10–19	1	1
20–29	2	2
30–39	0	2
40–49	4	4
50–59	3	4
60–69	1	5
70–79	2	3
≥80	0	0
Total	15	23



Introduction to Tables and Graphs

- Exercise 4.1 (ctd.): Construct a 2 x 2 of botulism to exposure to chicken:

Table B: *Botulism Status by Exposure to Chicken**, TX Church Supper Outbreak, 2001

		Botulism?		Total
		Yes	No	
Ate Chicken?	Yes	8	11	19
	No	4	12	16
Total		12	23	35

*- 3 cases were included who did not eat chicken.



Introduction to Tables and Graphs

- Exercise 4.1 (ctd.): Construct a 2 x 2 of botulism to exposure to chili:

Table C: *Botulism Status by Exposure to Chili**, TX Church Supper Outbreak, 2001*

		Botulism?		Total
		Yes	No	
Ate Chili?	Yes	14	8	22
	No	0	15	15
Total		14	23	37

*- Includes one botulism case without chili exposure



Introduction to Tables and Graphs

- Exercise 4.1 (ctd.): Construct a 2 x 2 of botulism to exposure to chili:

Table D: *Botulism Cases/Controls by Exposure to Chili and Leftover Chili TX Church Supper Outbreak, 2001**

		Ate left over chili?		Total
		Yes	No	
Ate Chili?	Yes	1/1	13/7	22
	No	0/1	0/14	15
Total		3	34	37

*- Includes one botulism case without initial chili exposure

- In the cells of this table, the first number is the cases who ate chili and developed botulism; the second are those who ate leftover and got it



Introduction to Tables and Graphs

- Tables of Statistical Measures other than Frequency:
 - Cells could display averages, rates, relative risks, or other epidemiological measures
 - Headings must clearly ID what data are presented
- Composite Tables:
 - To conserve space in reports/manuscripts, composite tables, which consist of several tables are frequently employed
 - This type of table should not be confused with three-way tables; Instead, several one-variable tables have been consolidated, so comparison between variables isn't feasible
 - This is more descriptive in nature/design



Introduction to Tables and Graphs

- Example of Composite Table:

Table 4.8: *Confirmed CO Poisonings ID from 4 Hospitals—Maine, Jan 1998*

Characteristic	Cases	
	Number	Percent
Total Cases	100	100
Sex (female)	59	59
Age (years)		
0-3	5	5
4-12	17	17
13-18	9	9
19-64	52	52
≥ 65	17	17
Smokers	20	20
Disposition		
Released from ED	83	83
Hospital Admit	11	5
Transferred	5	5
Died	1	1



Introduction to Tables and Graphs

- Table Shells:
 - Useful when outlining how data will be analyzed
 - Tables that are complete except for the data; they show titles, headings, and categories
 - Keep in mind: when designing table shells with continuous variables (eg: age), create more categories which you may later use
 - These shells usually follow a sequence of: descriptive → analytic
 - 1st and 2nd tables usually cover the clinical features and sample demographics; next, the association between interest points are depicted
 - The following examples are Table Shells designed to depict results from a case-control study of Fx in an elderly community, where researchers were studying the effects of activity on fall-related Fx



Introduction to Tables and Graphs

Table 4.9: *Anatomic Site of Fall-related Fx Sustained by Participants, SAFE Study—Miami, 1987-1989*

Fx Site	Number	(Percent)
Skull	_____	()
Spine	_____	()
Clavicle	_____	()
Scapula	_____	()
Humerus	_____	()
Radius	_____	()
Hand Bones	_____	()
Ribs/Sternum	_____	()
Pelvis	_____	()
etc.	_____	()



Introduction to Tables and Graphs

Table 4.9b: *Selected Characteristics of Participants, SAFE Study—Miami 1987-1989*

		Cases		Controls	
		Number	(percent)	Number	(percent)
Age	65–74	—	()	—	()
	75–84	—	()	—	()
	≥85	—	()	—	()
Sex	Male	—	()	—	()
	Female	—	()	—	()
Race	White	—	()	—	()
	Black	—	()	—	()
	Other	—	()	—	()
	Unknown	—	()	—	()
Ethnicity	Hispanic	—	()	—	()
	Non-Hispanic	—	()	—	()
	Unknown	—	()	—	()
Hours/day spent on feet	<1	—	()	—	()
	2–4	—	()	—	()
	5–7	—	()	—	()
	>8	—	()	—	()
Smoking status	Never smoked	—	()	—	()
	Former smoker	—	()	—	()
	Current smoker	—	()	—	()
	Unknown	—	()	—	()
Alcohol use (drinks / week)	None	—	()	—	()
	<1	—	()	—	()
	1–3	—	()	—	()
	>4	—	()	—	()
	Unknown	—	()	—	()



Introduction to Tables and Graphs

- Creating Class Intervals:
 - While the information presented thus far is effective for simple categorical intervals, most variables are not this simple;
 - Most variables look at more complex data (BP measurements, days of hospitalization, etc.)
 - For these type of variable, we can create class intervals
 - Class intervals should be mutually exclusive and exhaustive; data should not overlap and should follow conventional rounding rules:
 - If a fraction is $\geq .5$, round it up; if $< .5$, round down.
 - Begin with larger sized class intervals; you can always combine categories later
 - Make sure your intervals make sense and are biologically logic:
 - For example, when assessing pediatric mortality data, you would want to keep age intervals at:
 - 0-12 mos. (possible institutional influences)
 - 1-5 yrs. (possible external to institutional influences)
 - 5-10 yrs. (possible risks in school settings)



Introduction to Tables and Graphs

- CDC Age Classifications:
 - < 1: Infants
 - 1-4: Toddlers
 - 5-14: Adolescents
 - 15-24: Teen/Young Adult
 - 25-44: Adults
 - 45-64: Older Adults
 - ≥ 65 : Elderly
- A natural baseline group should be kept as a distinct category; this is usual those who have not had an exposure (non-smokers)
- Make sure your numerators and denominators match for the appropriate data (for example: if using CDC data regarding age, use the CDC Age Classifications)
- Include a category for “unknown” or “unstated”



Introduction to Tables and Graphs

- CDC Age Groupings for Different Conditions:

Table 4.10 Age Groupings Used for Different Conditions, as Reported in Surveillance Summaries, CDC, 2003

Overweight In Adults ⁷	Traumatic Brain Injury ⁸	Pregnancy-Related Mortality ⁹	HIV/AIDS ¹⁰	Vaccine Adverse Events ¹¹
18–24 years	<4 years	≤19 years	<13 years	<1 year
25–34	5–14	20–24	13–14	1–6
35–44	15–19	25–29	15–24	7–17
45–54	20–24	30–34	25–34	18–64
55–64	25–34	35–39	35–44	≥65
65–74	35–44	≥40	45–54	
≥75	45–64		55–64	
	≥65		≥65	
Total	Total	Total	Total	Total



Introduction to Tables and Graphs

- Strategy 1: Divide the Data into Groups of Similar Size:
 - Keep your class intervals with the same number of obs.:
 - For example, divide incidence rates of lung CA by state, each w/ 10-12 obs.:

Table 4.11 Rates of Lung Cancer in Men, 2001 by State (and the District of Columbia)

Rate	Number of States in the US	Cumulative Frequency
22.1–48.3	11	11
48.4–53.3	11	22
53.4–58.7	12	34
58.8–73.3	10	44
Missing data	7	51

Data Source: U.S. Cancer Statistics Working Group. United States Cancer Statistics: 2002 Incidence and Mortality. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute; 2005.



Introduction to Tables and Graphs

- Strategy 2: Base Intervals on Mean and SD:
 - With this strategy, you can create 3, 4, or 6 class intervals
 - First, calculate the Mean and SD of the data distribution
 - Next, use the mean + or – different multiple of the SD to establish the upper levels
 - For example: You are preparing a study assessing readiness of your State health departments to emergency response; each department can score from 0-100 (100 highest); the range of the departments are from 19-82 with a mean of 50 with a SD of 10; Here's how to establish the 6 data intervals:
 - Upper limit of interval 6 = maximum value = 82
 - Upper limit of interval 5 = $50 + 20 = 70$
 - Upper limit of interval 4 = $50 + 10 = 60$
 - Upper limit of interval 3 = 50
 - Upper limit of interval 2 = $50 - 10 = 40$
 - Upper limit of interval 1 = $50 - 20 = 30$
 - Lower limit of interval 1 = 19
 - Next, define your intervals:

Interval 6 = 71–82	Interval 3 = 41–50
Interval 5 = 61–70	Interval 2 = 31–40
Interval 4 = 51–60	Interval 1 = 19–30



Introduction to Tables and Graphs

- Strategy 3: Divide the Range into Equal Class Intervals:
 - Simplest and most common method
 - To use equal class intervals:
 - Find the range (largest value – smallest value)
 - Choose between 4 and 8 class intervals for tables; for graphs and maps, select between 3 and 6 class intervals
 - Divide the range by the number of class intervals you want
 - Minimum value will be the lowest level, continue to add the calculated value (above) until you've reached your highest value
 - Example: Calculate the Class intervals for a table depicting men > age 40 screened for prostate CA in the US (including DC and PR):
 - » 52 observations (50 states + 2 = 52)
 - » 5 class intervals; Range = $64.9 - 40.0 = 24.9$
 - » $24.9 / 5 = 4.98$: Intervals: 40.0 - 44.9 45.0 - 49.9
50.0 – 54.9 55.0- 59.9
60.0 – 64.9



Introduction to Tables and Graphs

- Let's practice each strategy:

EXAMPLE: Creating Class Interval Categories

Use each strategy to create four class interval categories by using the lung cancer mortality rates shown in Table 4.13.

Table 4.13 Age-adjusted Lung Cancer Death Rates per 100,000 population, in Rank Order by State—United States, 2000

Rank	State	Rate per 100,000	Rank	State	Rate per 100,000
1	Kentucky	116.1	26	Florida	75.3
2	Mississippi	111.7	27	Kansas	74.5
3	West Virginia	104.1	28	Massachusetts	73.6
4	Tennessee	103.4	29	Alaska	72.9
5	Alabama	100.8	30	Oregon	72.7
6	Louisiana	99.2	31	New Hampshire	71.2
7	Arkansas	99.1	32	New Jersey	71.2
8	North Carolina	94.6	33	Washington	71.2
9	Georgia	93.2	34	Vermont	70.2
10	South Carolina	92.4	35	South Dakota	68.1
11	Indiana	91.6	36	Wisconsin	67.0
12	Oklahoma	89.4	37	Montana	66.5
13	Missouri	88.5	38	Connecticut	66.4
14	Ohio	85.6	39	New York	66.2
15	Virginia	83.0	40	Nebraska	65.6
16	Maine	80.2	41	North Dakota	64.9
17	Illinois	80.0	42	Wyoming	64.4
18	Texas	79.3	43	Arizona	62.0
19	Maryland	79.2	44	Minnesota	60.7
20	Nevada	78.7	45	California	60.1
21	Delaware	78.2	46	Idaho	59.7
22	Rhode Island	77.9	47	New Mexico	52.3
23	Iowa	77.0	48	Colorado	52.1
24	Michigan	76.7	49	Hawaii	49.8
25	Pennsylvania	76.5	50	Utah	39.7
Total			United States		76.9

Data Source: Stewart SL, King JB, Thompson TD, Friedman C, Wingo PA. Cancer Mortality—United States, 1990–2000. In: Surveillance Summaries, June 4, 2004. MMWR 2004;53 (No. SS-3):23–30.



Introduction to Tables and Graphs

- **Strategy 1: Divide the data into groups of similar size:**
 - (Note: If the states in Table 4.13 had been listed alphabetically rather than in rank order, the first step would have been to sort the data into rank order by rate. Fortunately, this has already been done.)
- 1. Divide the list into four equal sized groups of places:
 - $50 \text{ states} / 4 = 12.5 \text{ states per group}$. Because states can't be cut in half, use two groups of 12 states and two groups of 13 states. Missouri (#13) could go into either the first or second group and Connecticut (#38) could go into either third or fourth group. Arbitrarily putting Missouri in the second category and Connecticut into the third results in the following groups:
 - a. Kentucky through Oklahoma (States 1–12)
 - b. Missouri through Pennsylvania (States 13–25)
 - c. Florida through Connecticut (States 26–38)
 - d. New York through Utah (States 39–50)
- 2. Identify the rate for the first and last state in each group:
 - a. Oklahoma through Kentucky 89.4–116.1
 - b. Pennsylvania through Missouri 76.5–88.5
 - c. Connecticut through Florida 66.4–75.3
 - d. Utah through New York 39.7–66.2
- 3. Adjust the limits of each interval so no gap exists between the end of one class interval and beginning of the next. Deciding how to adjust the limits is somewhat arbitrary — you could split the difference, or use a convenient round number:
 - a. Oklahoma through Kentucky 89.0–116.1
 - b. Pennsylvania through Missouri 76.0–88.9
 - c. Connecticut through Florida 66.3–75.9
 - d. Utah through New York 39.7–66.2



Introduction to Tables and Graphs

- **Strategy 2: Base intervals on mean and standard deviation:**
- 1. Calculate the mean and standard deviation (see Lesson 2 for instructions in calculating these measures.): Mean = 77.1; Standard deviation = 16.1
- 2. Find the upper limits of four intervals:
 - a. Upper limit of interval 4 = maximum value = 116.1
 - b. Upper limit of interval 3 = mean + 1 standard deviation = $77.1 + 16.1 = 93.2$
 - c. Upper limit of interval 2 = mean = 77.1
 - d. Upper limit of interval 1 = mean – 1 standard deviation = $77.1 - 16.1 = 61.0$
 - e. Lower limit of interval 1 = minimum value = 39.7
- 3. Select the lower limit for each upper limit to define four full intervals. Specify the states that fall into each interval (Note: To place the states with the highest rates first, reverse the order of the intervals):
 - a. North Carolina through Kentucky (8 states) 93.3–116.1
 - b. Rhode Island through Georgia (14 states) 77.1–93.2
 - c. Arizona through Iowa (21 states) 61.1–77.1 (why can 77.1 be used twice?)
 - d. Utah through Minnesota (7 states) 39.7–61.0



Introduction to Tables and Graphs

- **Strategy 3: Divide the range into equal class intervals:**
- 1. Divide the range from zero (or the minimum value) to the maximum by 4:
 - $(116.1 - 39.7) / 4 = 76.4 / 4 = 19.1$
- 2. Use multiples of 19.1 to create four categories, starting with 39.7:
 - 39.7 through $(39.7 + 19.1) = 39.7$ through 58.8
 - 58.9 through $(39.7 + [2 \times 19.1]) = 58.9$ through 77.9
 - 78.0 through $(39.7 + [3 \times 19.1]) = 78.0$ through 97.0
 - 97.1 through $(39.7 + [4 \times 19.1]) = 97.1$ through 116.1
- 3. Final categories:
 - a. Arkansas through Kentucky (7 states) 97.1–116.1
 - b. Delaware through North Carolina (14 states) 78.0–97.0
 - c. Idaho through Rhode Island (25 states) 58.9–77.9
 - d. Utah through New Mexico (4 states) 39.7–58.8
- 4. Alternatively, since 19.1 is close to 20, multiples of 20 might be used to create the four categories that might look cleaner. For example, the final categories could look like:
 - a. Arkansas through Kentucky (7 states) 97.0–116.9
 - b. Iowa through North Carolina (16 states) 77.0–96.9
 - c. Idaho through Michigan (23 states) 57.0–76.9
 - d. Utah through New Mexico (4 states) 37.0–56.9
- OR
 - a. Alabama through Kentucky (5 states) 100.0–119.9
 - b. Illinois through Louisiana (12 states) 80.0–99.9
 - c. California through Texas (28 states) 60.0–79.9
 - d. Utah through Idaho (5 states) 39.7–59.9



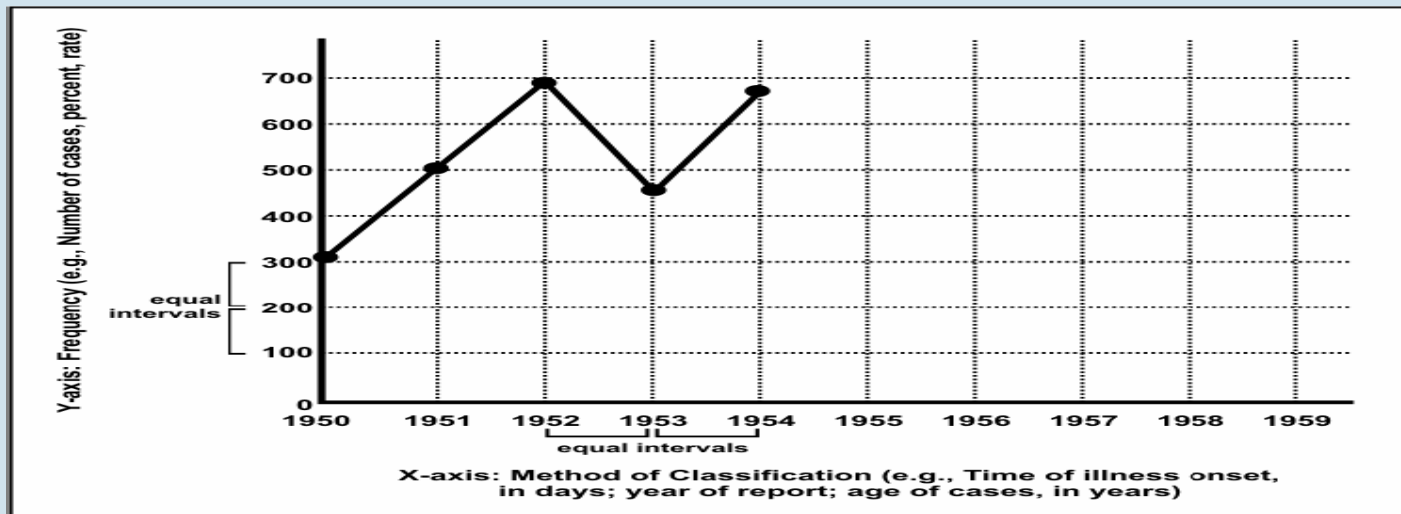
Graphs

- Displays numeric data in visual form
- Displays patterns, trends, aberrations, similarities, and data differences not readily seen in tables
- Some best practices:
 - Ensure a graph can stand alone by clearly labeling title, sources, axes, scales, and legends
 - Clearly ID variables portrayed (legends or keys), including units of measure
 - Minimize number of lines on a graph
 - Display frequency on the vertical (Y) axis, starting at 0, and classification variable on the horizontal (X) axis
 - Ensure that scales for each axis are appropriate for presented data
 - Define abbreviations/ symbols
 - Specify excluded data
- X Axis shows values of the independent variable, such as time or age group
- Y Axis shows the dependent variable, usually # of cases or rates of disease
- Line graphs are appropriate at showing continuous variable data



Graphs

- Let's Plot:
 - **Scenario:** Table 4.14 shows the number of measles cases by year of report from 1950 to 2003. The number of measles cases in years 1950 through 1954 has been plotted in Figure 4.1, below.
 - The independent variable, years, is shown on the horizontal axis. The dependent variable, number of cases, is shown on the vertical axis.
 - A grid is included in Figure 4.1 to illustrate how points are plotted. For example, to plot the point on the graph for the number of cases in 1953, draw a line up from 1953, and then draw a line from 449 cases to the right. The point where these lines intersect is the point for 1953 on the graph:



Graphs

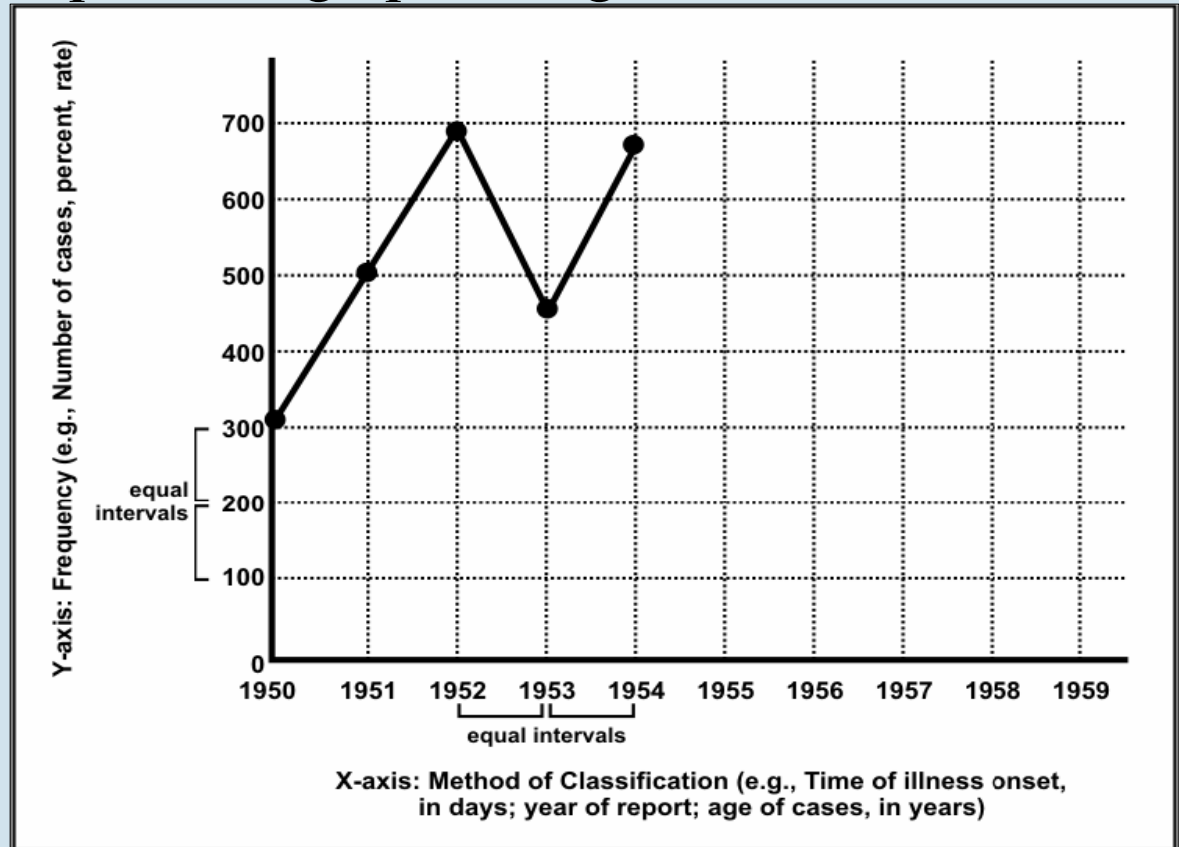
- Let's Plot:

- Your Turn:** Use the data in Table 4.14 to plot the points for 1955 to 1959 and complete the graph in Figure 4.1:

Table 4.14 Number of Reports

Year	Cases
1950	319,000
1951	530,000
1952	683,000
1953	449,000
1954	683,000
1955	555,000
1956	612,000
1957	487,000
1958	763,000
1959	406,000
1960	442,000
1961	424,000
1962	482,000
1963	385,000
1964	458,000
1965	262,000
1966	204,000
1967	62,705
1968	22,231
1969	25,826

Data Sources: Centers for Disease Control, *Morbidity and Mortality Weekly Report*, 1989;38(No. 54).
Centers for Disease Control and Prevention
Centers for Disease Control and Prevention



College of Nursing



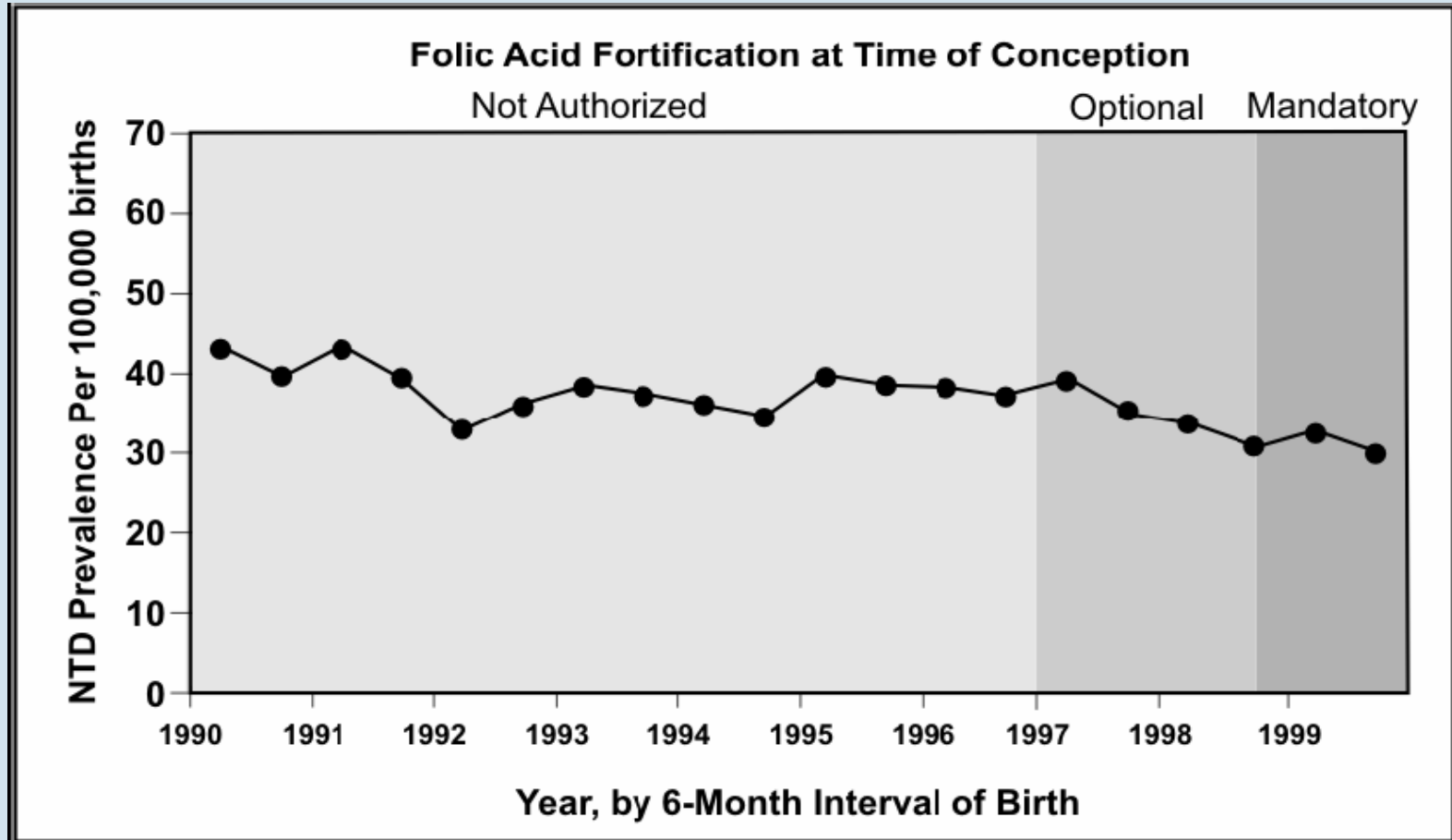
Graphs

- Arithmetic-Scale Line Graphs:
 - Shows patterns in trends over some variable, often time
 - Useful at showing changes over a period of time and is method of choice for plotting rates over time
 - A set distance along any axis represents the same quantity anywhere on that axis
 - Thus an arithmetic scale line graph is one in which equal distances along either the x or y- axis portray equal values
 - Arithmetic-scale line graphs can display numbers, rates, proportions, or other quantitative measures on the y-axis
 - Generally, the x-axis for these graphs is used to portray the time period of data occurrence, collection, or reporting (e.g., days, weeks, months, or years)
 - Thus, these graphs are primarily used to portray an overall trend over time, rather than an analysis of particular observations (single data points)



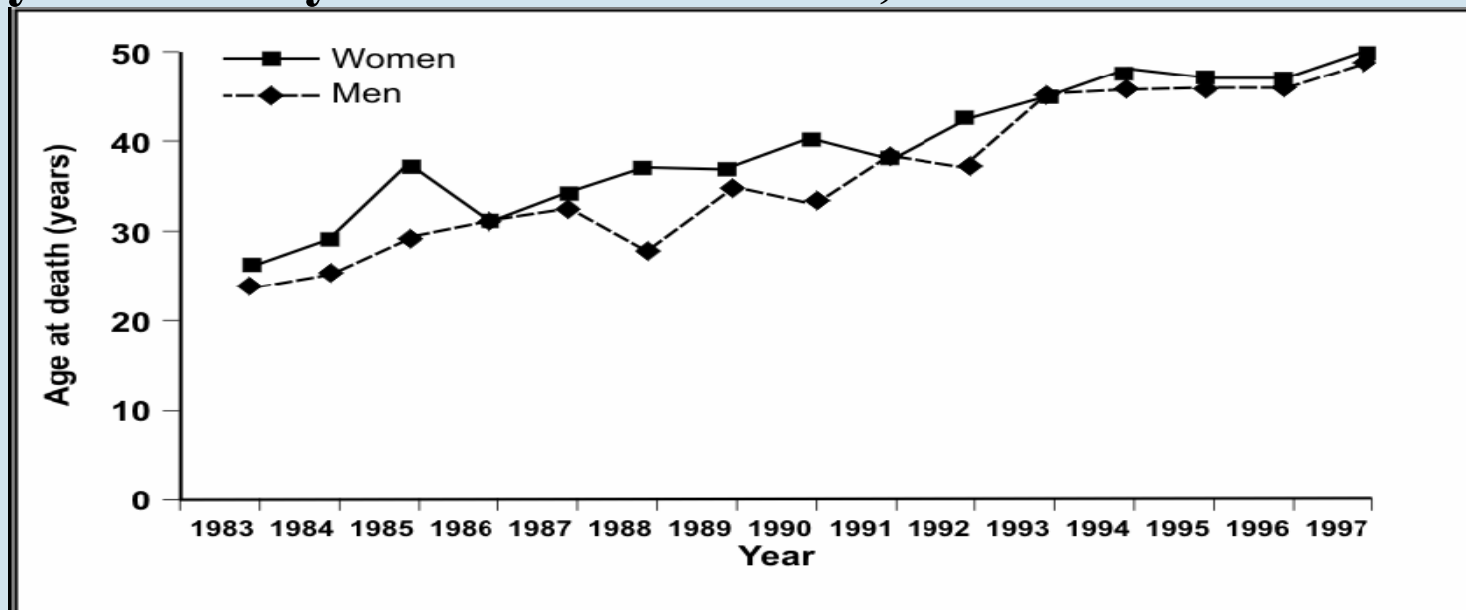
Graphs

- **Figure 4.2 Trends in Neural Tube Defects (Anencephaly and Spina Bifida) Among All Births, 45 States and District of Columbia, 1990– 1999**



Graphs

- Figure 4.3 shows another example of an arithmetic-scale line graph
 - Here the y-axis is a calculated variable, median age at death of people born with Down's syndrome from 1983–1997
 - Here also, we see the value of showing two data series on one graph; we can compare the mortality risk for males and females
- **Figure 4.3 Median Age at Death of People with Down's Syndrome by Sex—United States, 1983–1997**



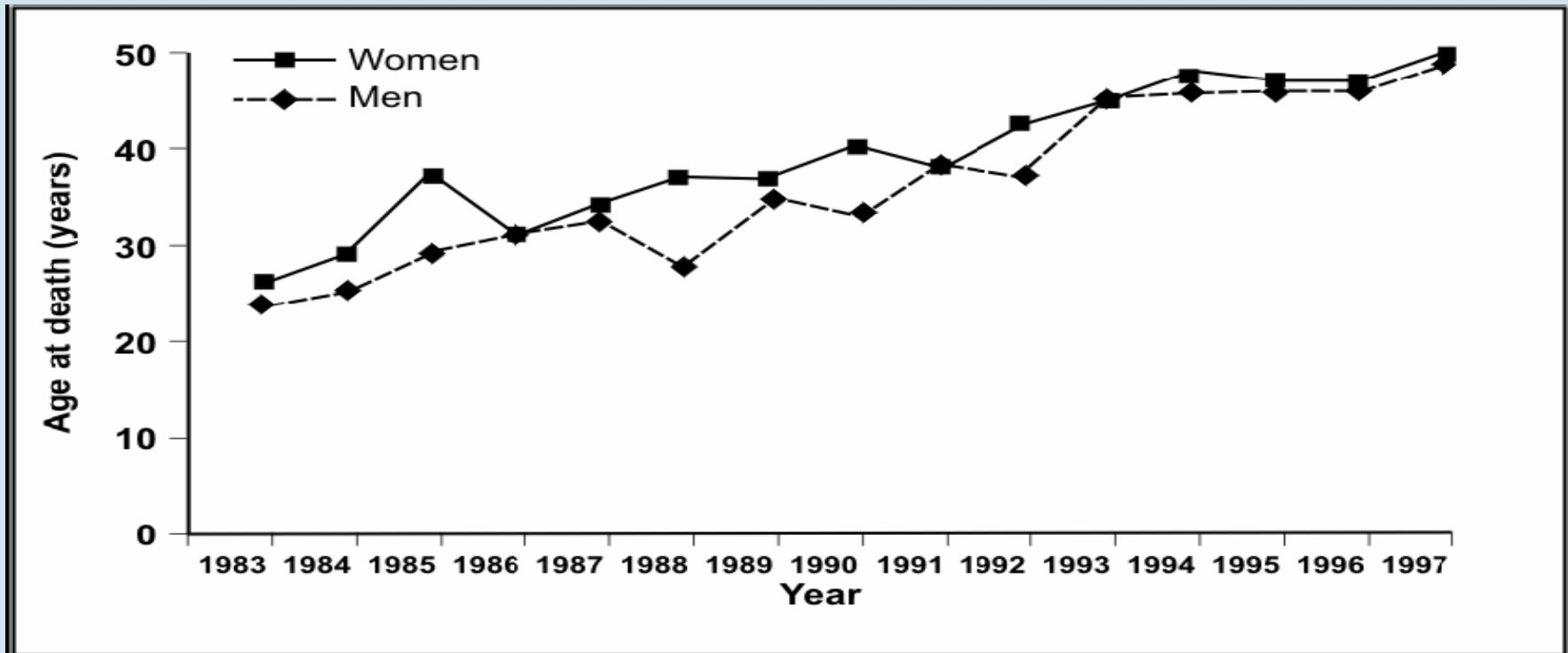
Graphs

- The following steps are recommended for creating a scale for the y-axis:
 - Make the length of the y-axis shorter than the x-axis so that your graph is horizontal or “landscape.” A 5:3 ratio is often recommended for the length of the x-axis to y-axis
 - Always start the y-axis with 0. While this recommendation is not followed in all fields, it is the standard practice in epidemiology
 - Determine the range of values you need to show on the y-axis by identifying the largest value you need to graph on the y-axis and rounding that figure off to a slightly larger number
 - For example, the largest y-value in Figure 4.3 is 49 years in 1997, so the scale on the y-axis goes up to 50. If median age continues to increase and exceeds 50 in future years, a future graph will have to extend the scale on the y-axis to 60 years
 - Space the tick marks and their labels to describe the data in sufficient detail for your purposes. In Figure 4.3, five intervals of 10 years each were considered adequate to give the reader a good sense of the data points and pattern



Graphs

- **Figure 4.3 Median Age at Death of People with Down's Syndrome by Sex—United States, 1983–1997**



Graphs



Exercise 4.3

Using the data on measles rates (per 100,000) from 1955 to 2002 in Table 4.15:

A. Construct an arithmetic-scale line graph of rate by year. Use intervals on the y-axis that are appropriate for the range of data you are graphing.

B. Construct a separate arithmetic-scale line graph of the measles rates from 1985 to 2002. Use intervals on the y-axis that are appropriate for the range of data you are graphing.

Graph paper is provided at the end of this lesson.

Table 4.15 Rate (per 100,000 Population) of Reported Measles Cases by Year of Report—United States, 1955–2002

Year	Rate per 100,000	Year	Rate per 100,000	Year	Rate per 100,000
1955	336.3	1971	36.5	1987	1.5
1956	364.1	1972	15.5	1988	1.4
1957	283.4	1973	12.7	1989	7.3
1958	438.2	1974	10.5	1990	11.2
1959	229.3	1975	11.4	1991	3.8
1960	246.3	1976	19.2	1992	0.9
1961	231.6	1977	26.5	1993	0.1
1962	259.0	1978	12.3	1994	0.4
1963	204.2	1979	6.2	1995	0.1
1964	239.4	1980	6.0	1996	0.2
1965	135.1	1981	1.4	1997	0.06
1966	104.2	1982	0.7	1998	0.04
1967	31.7	1983	0.6	1999	0.04
1968	11.1	1984	1.1	2000	0.03
1969	12.8	1985	1.2	2001	0.04
1970	23.2	1986	2.6	2002	0.02

Data Sources: Centers for Disease Control. Summary of notifiable diseases—United States, 1989. *MMWR* 1989;38(No. 54).
Centers for Disease Control and Prevention. Summary of notifiable diseases—United States, 2002. Published April 30, 2004 for *MMWR* 2002;51(No. 53).



Graphs

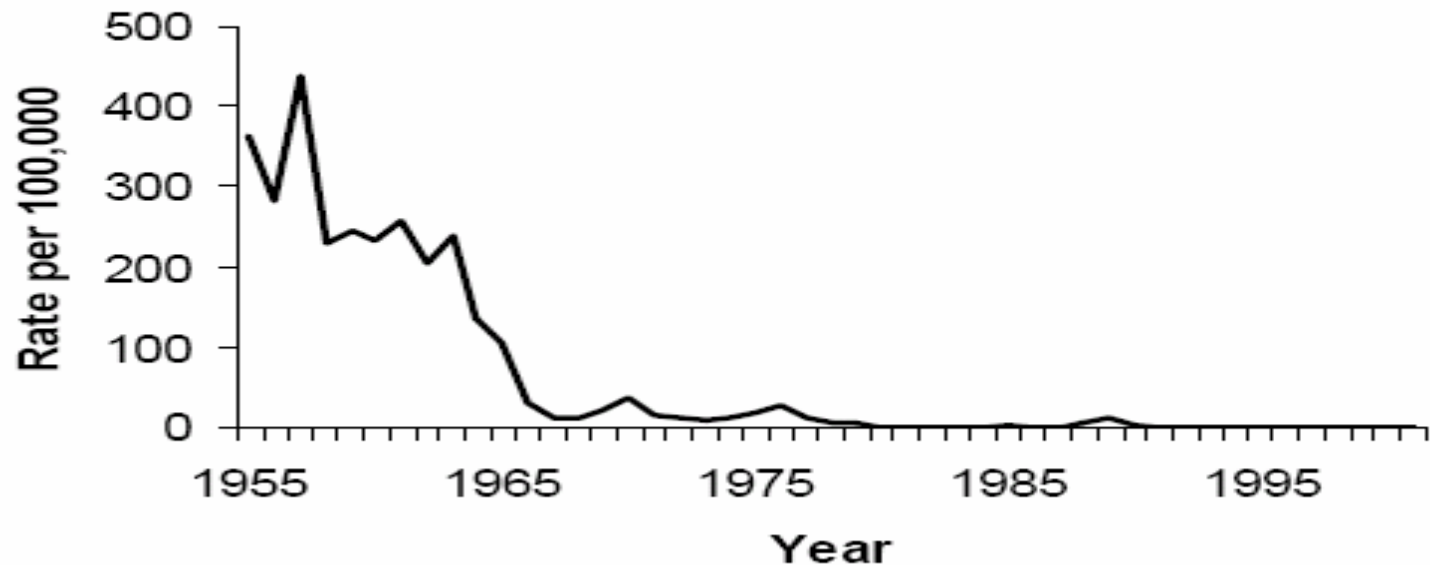


Exercise 4.3

Using the data on measles rates (per 100,000) from 1955 to 2002 in Table 4.15:

A. Construct an arithmetic-scale line graph of rate by year. Use intervals on the y-axis that are appropriate for the range of data you are graphing.

Rate (per 100,000 Population) of Reported Measles Cases by Year of Report—United States, 1955–2002

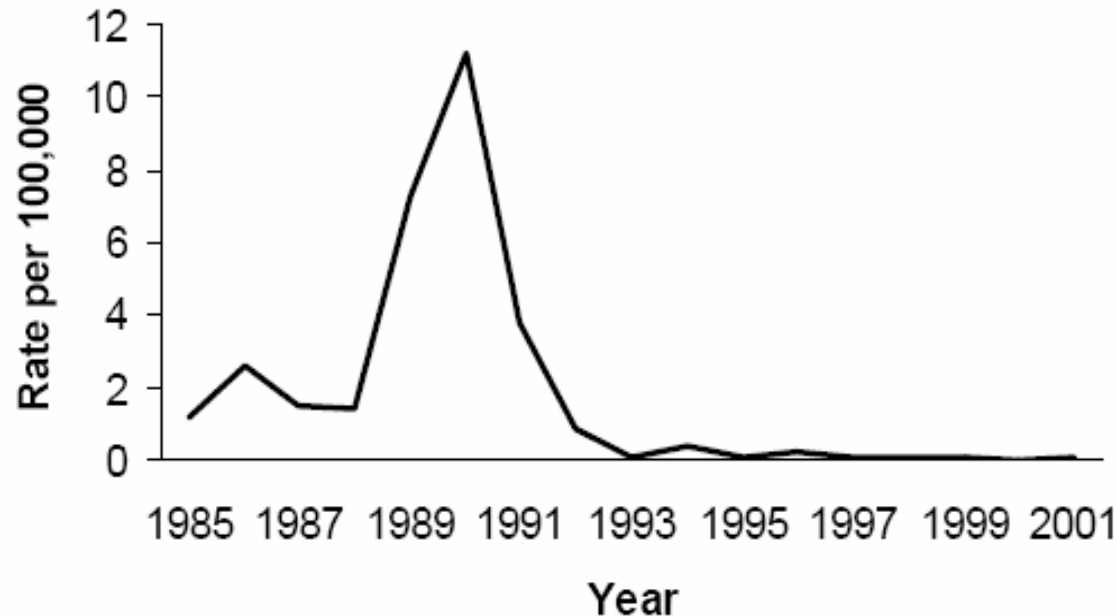


Graphs

PART B

Highest rate between 1985 and 2002 was 11.2 per 100,000 in 1990), so maximum on y-axis should be 12 per 100,000.

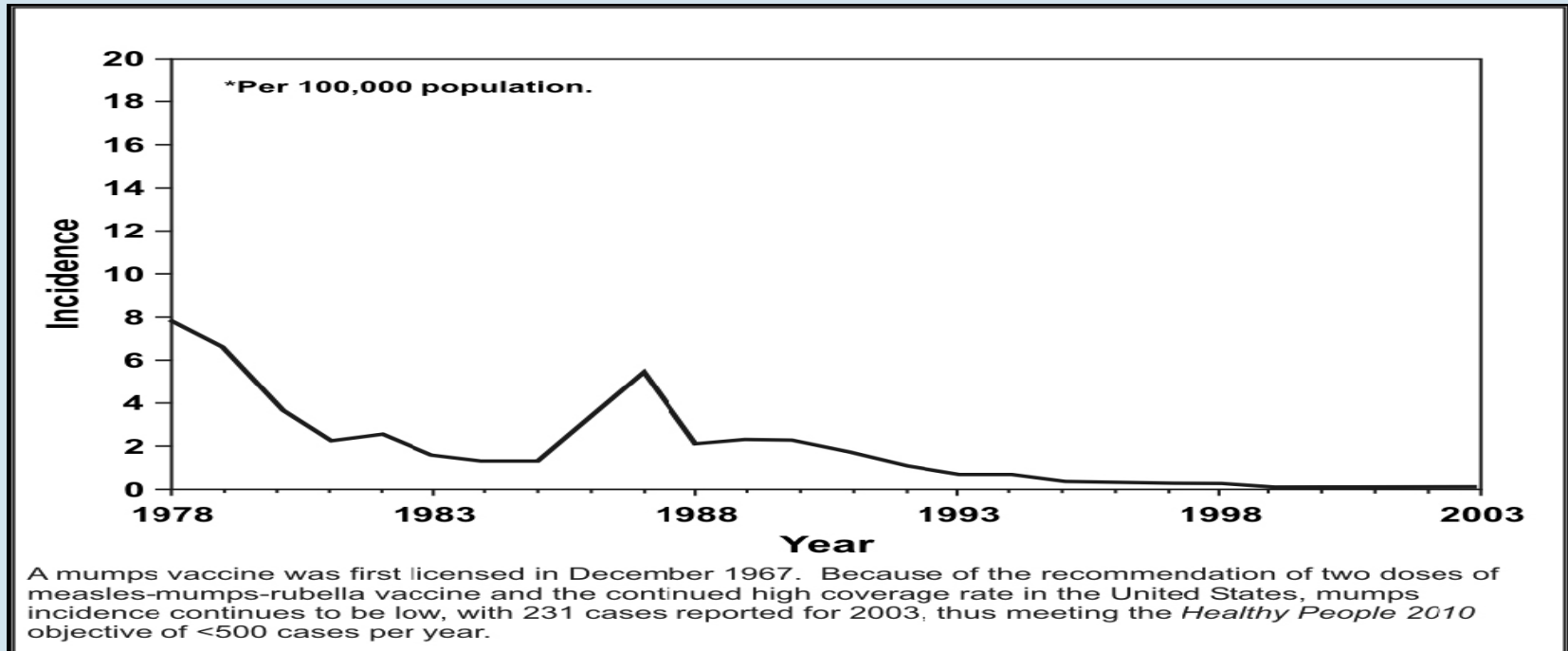
**Rate (per 100,000 Population) of Reported Measles Cases
by Year of Report—United States, 1985–2002**



Graphs

- Semilogarithmic- (Semi-Log) Scale Line Graphs:
 - Useful at showing a variable with a wide range of values
 - The x-axis uses the usual arithmetic-scale, but the y-axis is measured on a logarithmic rather than an arithmetic scale
 - As a result, the distance from 1 to 10 on the y- axis is the same as the distance from 10 to 100 or 100 to 1,000:

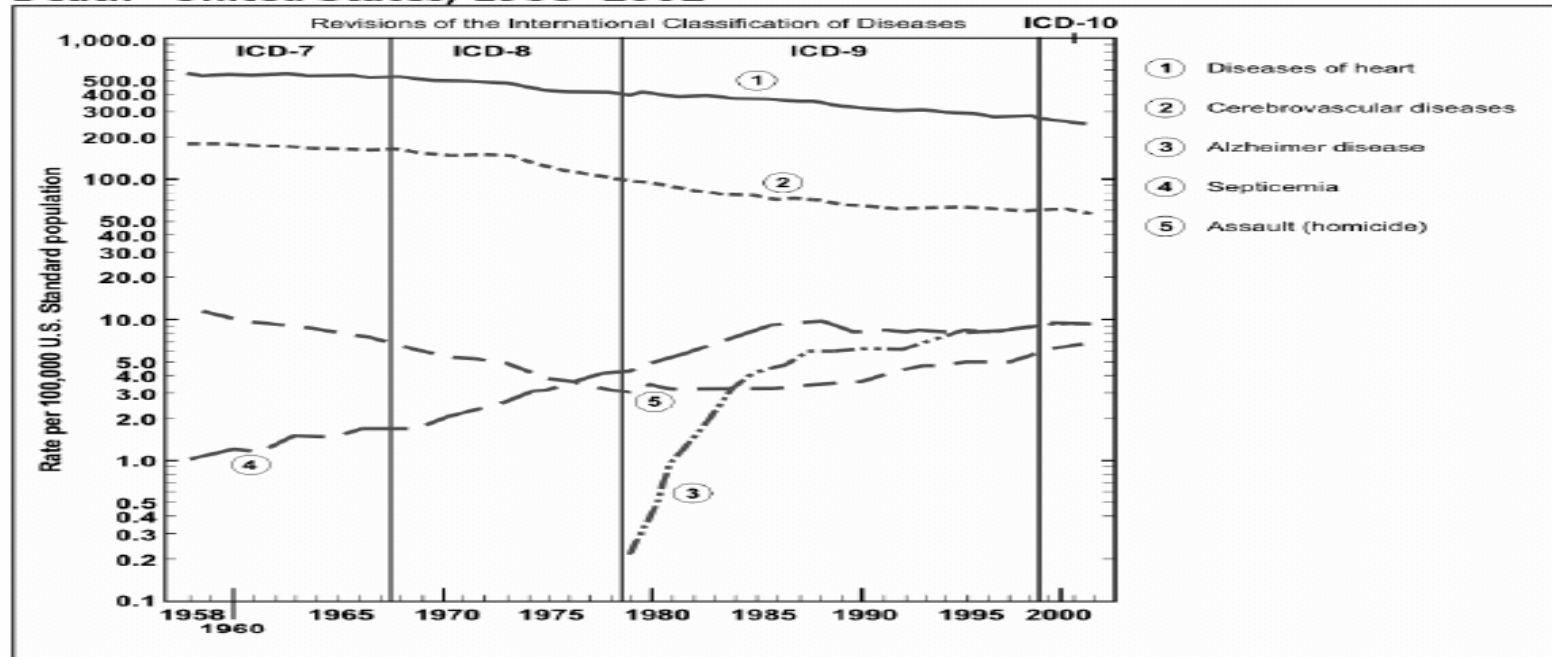
Figure 4.4 Mumps by Year—United States, 1978–2003



Graphs

- Another use for the semi-log graph is when you are interested in portraying the relative **rate of change** of several series, rather than the absolute value
- Figure 4.5 shows this application:

Figure 4.5 Age-adjusted Death Rates for 5 of the 15 Leading Causes of Death—United States, 1958–2002

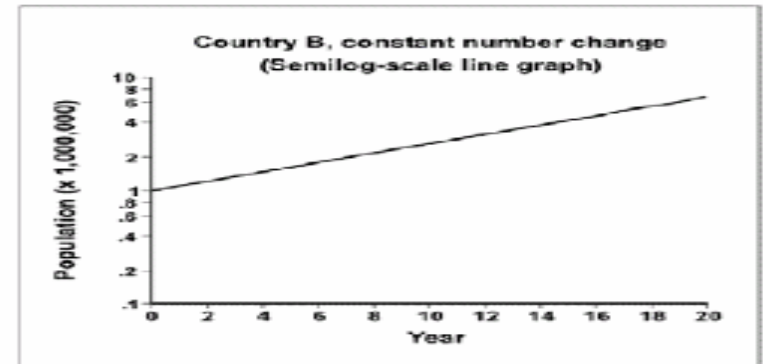
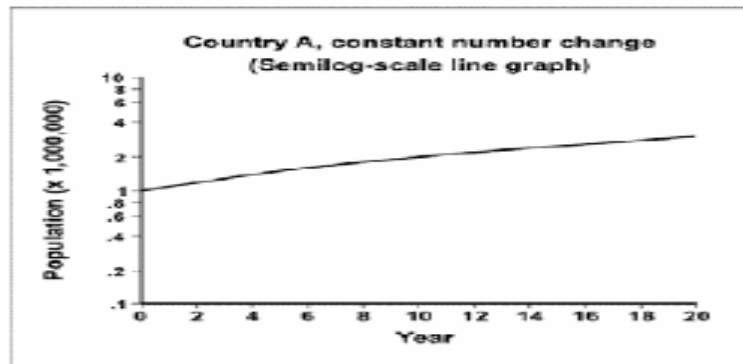
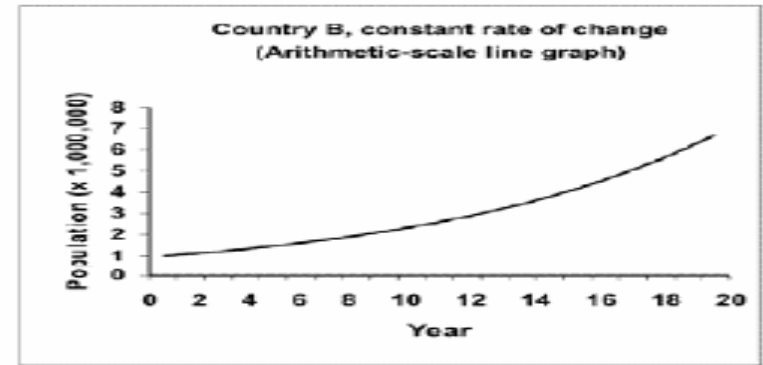
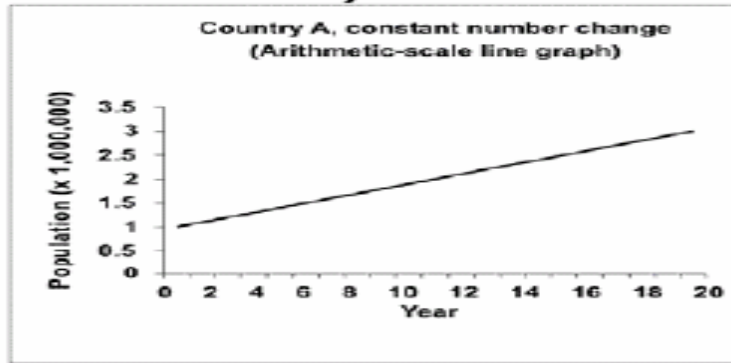


Adapted from: Kochanek KD, Murphy SL, Anderson RN, Scott C. Deaths: final data for 2002. National vital statistics report; vol 53, no 5. Hyattsville, Maryland: National Center for Health Statistics, 2004. p. 9.



Graphs

Figure 4.6 Comparison of Arithmetic-scale Line Graph and Semilogarithmic-scale Line Graph for Hypothetical Country A (Constant Increase in Number of People) and Country B (Constant Increase in Rate of Growth)



Graphs

- Consequently, a semilog-scale line graph has the following features:
 - The slope of the line indicates the rate of increase or decrease.
 - A straight line indicates a constant rate (not amount) of increase or decrease in the values
 - A horizontal line indicates no change
 - Two or more lines following parallel paths show identical rates of change
 - Semilog graph paper is available commercially, and most include at least three cycles.



Graphs

- Histograms:

Figure 4.7a Number of Cases of *Salmonella* Enteritidis Among Party Attendees by Date and Time of Onset—Chicago, Illinois, February 2000

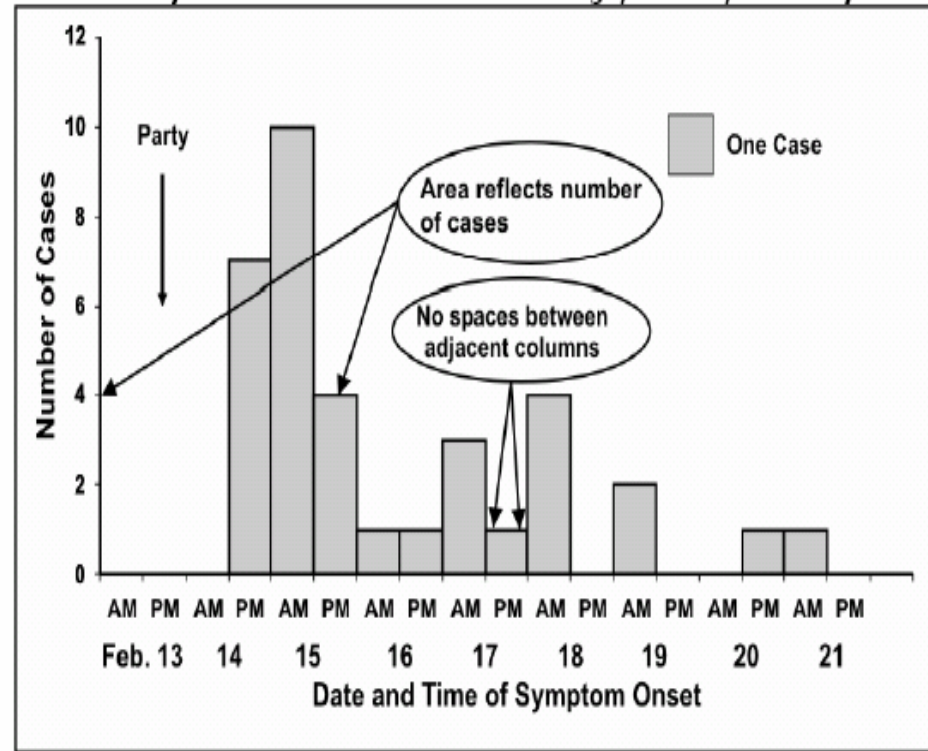
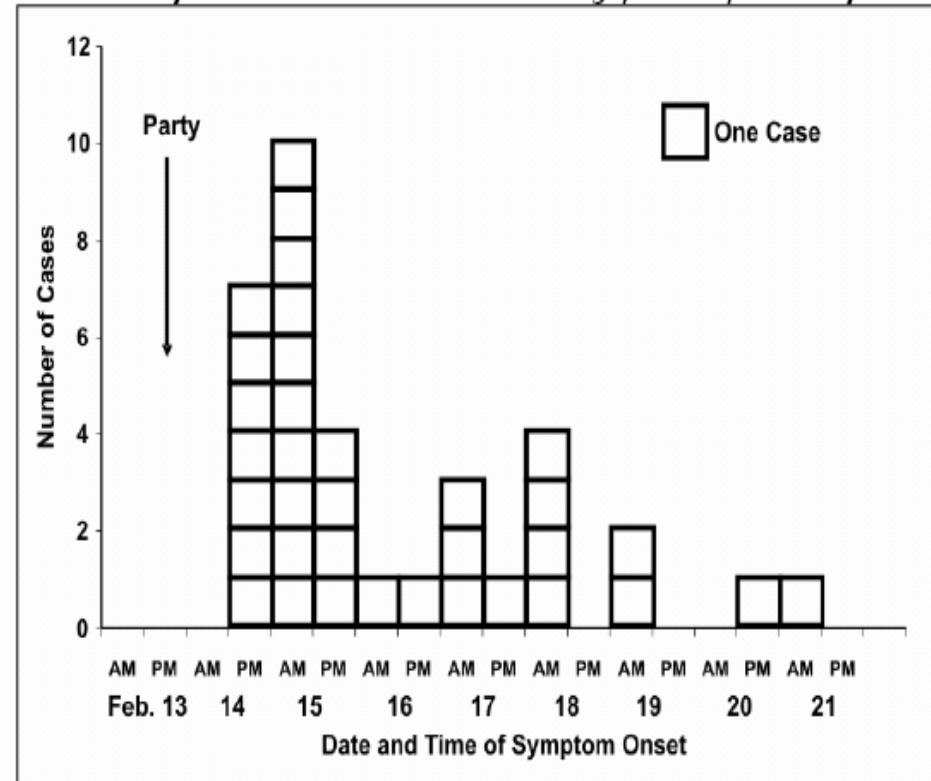


Figure 4.7b Number of Cases of *Salmonella* Enteritidis Among Party Attendees by Date and Time of Onset—Chicago, Illinois, February 2000



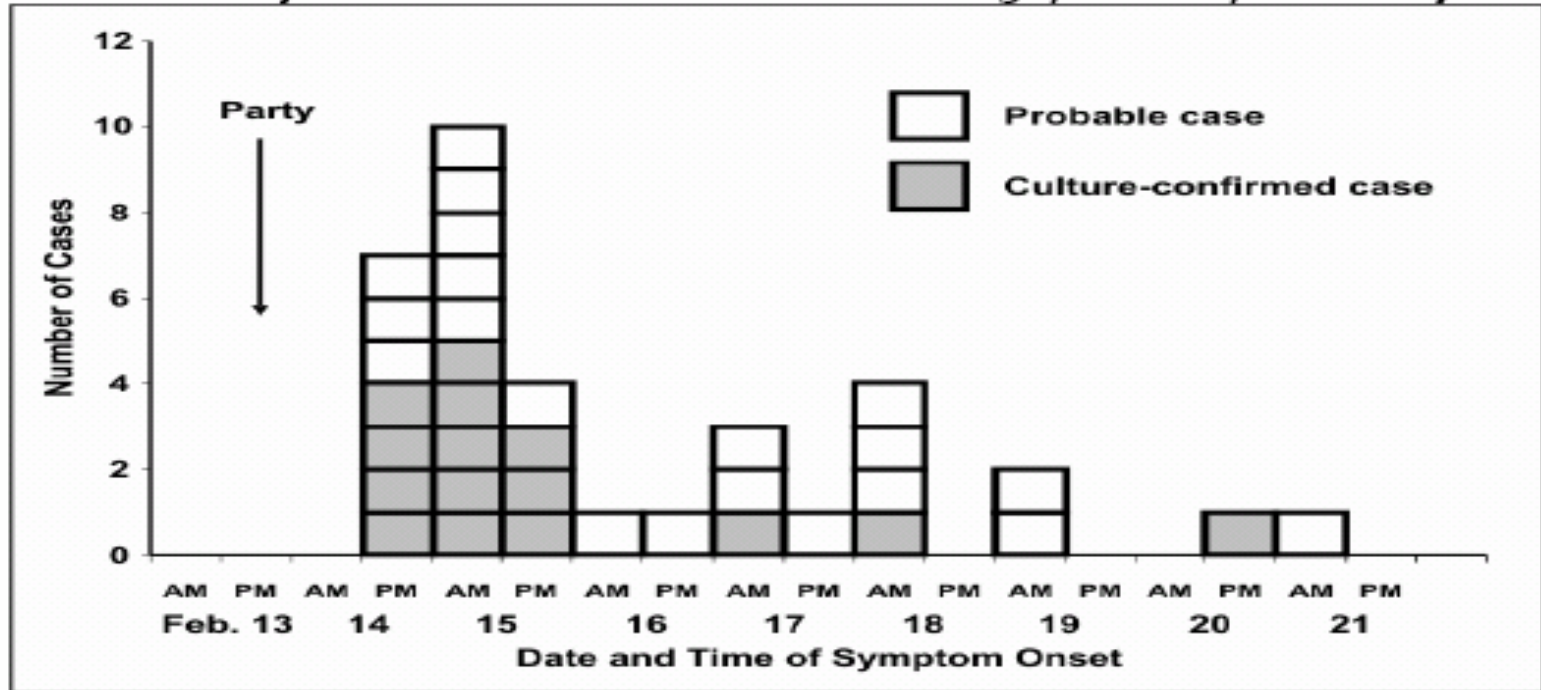
Source: Cortese M, Gerber S, Jones E, Fernandez J. A *Salmonella* Enteritidis outbreak in Chicago. Presented at the Eastern Regional Epidemic Intelligence Service Conference, March 23, 2000, Boston, Massachusetts.

Source: Cortese M, Gerber S, Jones E, Fernandez J. A *Salmonella* Enteritidis outbreak in Chicago. Presented at the Eastern Regional Epidemic Intelligence Service Conference, March 23, 2000, Boston, Massachusetts.



Graphs

Figure 4.7c Number of Cases of *Salmonella* Enteritidis Among Party Attendees by Date and Time of Onset—Chicago, Illinois, February 2000



Source: Cortese M, Gerber S, Jones E, Fernandez J. A *Salmonella* Enteritidis outbreak in Chicago. Presented at the Eastern Regional Epidemic Intelligence Service Conference, March 23, 2000, Boston, Massachusetts.



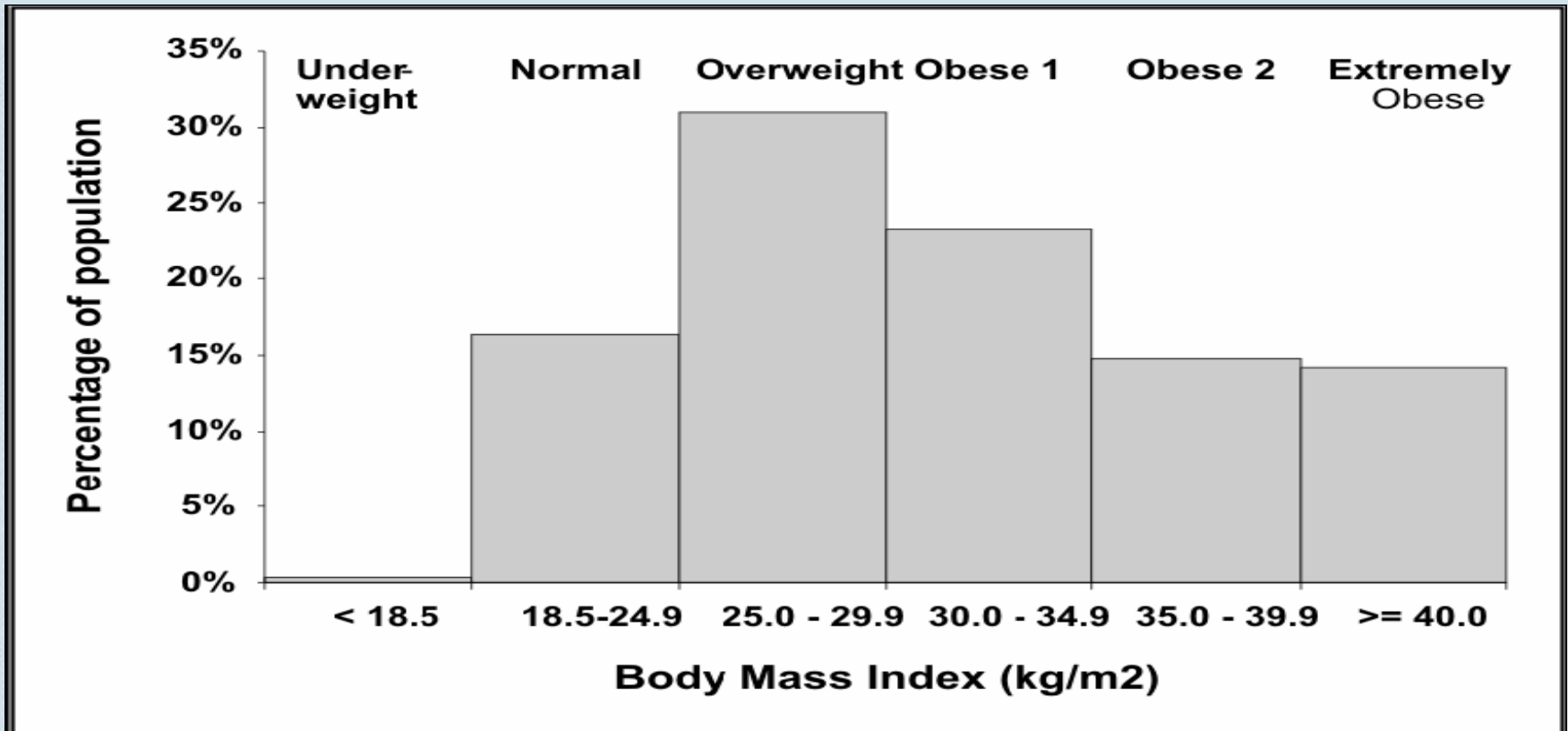
Graphs

- Histograms:
 - Conventionally, the numbers on the x-axis are centered between the tick marks of the appropriate interval
 - The interval of time should be appropriate for the disease in question, the duration of the outbreak, and the purpose of the graph
 - If the purpose is to show the temporal relationship between time of exposure and onset of disease, then a widely accepted rule of thumb is to use intervals approximately one-fourth (or between one-eighth and one-third) of the incubation period of the disease shown
 - The incubation period for salmonellosis is usually 12–36 hours, so the x-axis of this epidemic curve has 12-hour intervals (AM and PM interval classes are in-terms of 12-hours)
 - The most common choice for the x-axis variable in field epidemiology is calendar time, as shown in Figures 4.7a–c
 - However, age, cholesterol level or another continuous-scale variable may be used on the x-axis of an epidemic curve
 - Some histograms, particularly those that are drawn as stacks of squares, include a box that indicates how many cases are represented by each square
 - While a square usually represents one case in a relatively small outbreak, a square may represent five or ten cases in a relatively large outbreak.



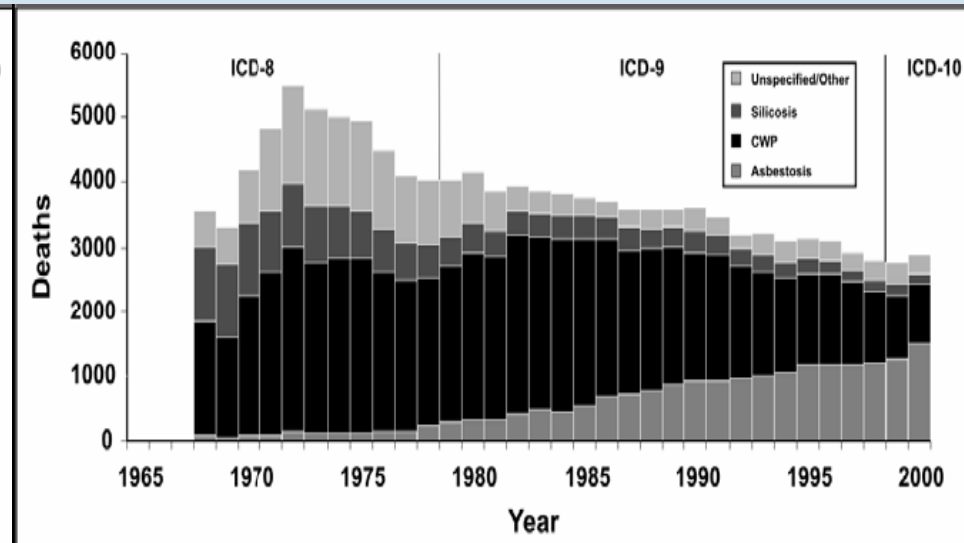
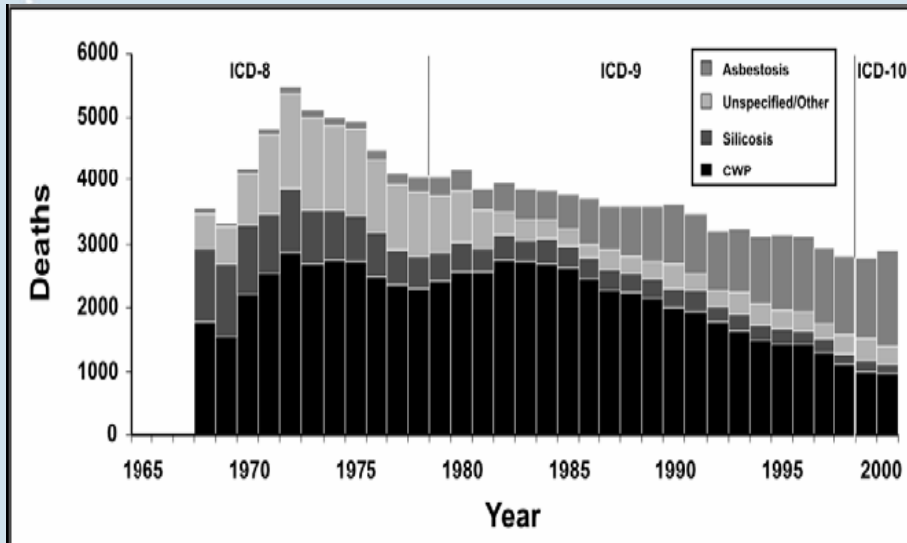
Graphs

Figure 4.8 Distribution of Body Mass Index Among Adults with Diagnosed Diabetes—United States, 1999–2002



Graphs

- The component of most interest should always be put at the bottom because the upper component usually has a jagged baseline that may make comparison difficult
- Consider the data on pneumoconiosis in Figure 4.9a (Left Graph)
- The graph clearly displays a gradual decline in deaths from all pneumoconiosis between 1972 and 1999. It appears that deaths from asbestosis (top subgroup in Figure 4.9a–) went against the overall trend, by increasing over the same period
- However, Figure 4.9b (Right Graph) makes this point more clearly by placing asbestosis along the baseline:



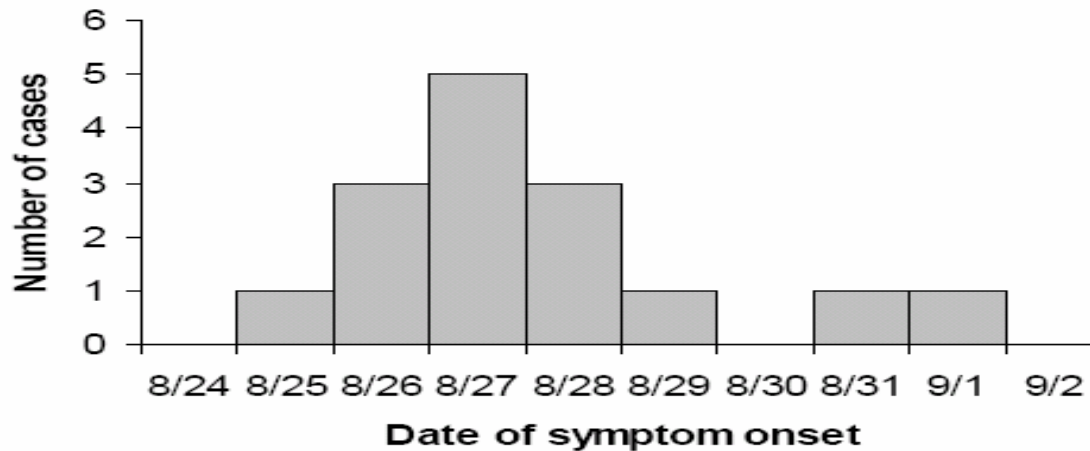
Graphs

- **Exercise 4.4:**

- *Using the botulism data presented in Exercise 4.1, draw an epidemic curve. Then use this epidemic curve to describe this outbreak as if you were speaking over the telephone to someone who cannot see the graph. Graph paper is provided at the end of this lesson.*

Exercise 4.4

**Number of Cases of Botulism by Date of Onset of Symptoms,
Texas Church Supper Outbreak, 2001**



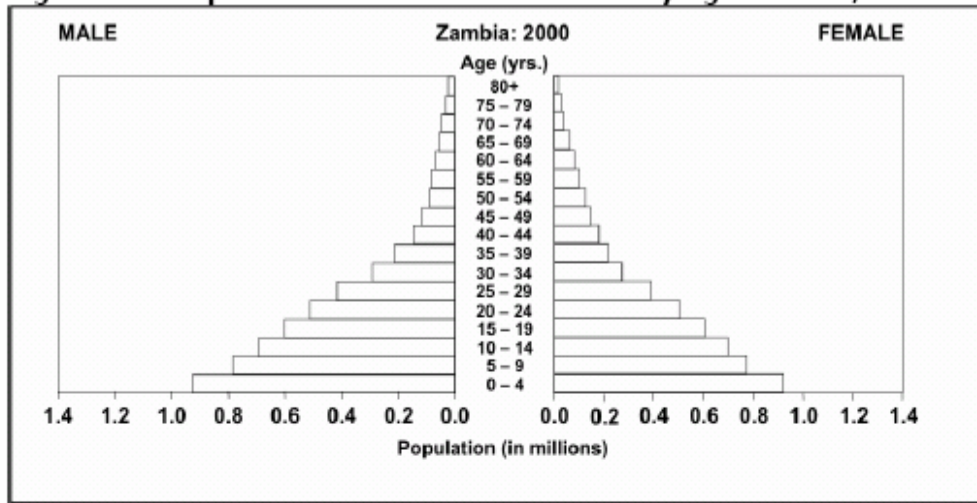
The first case occurs on August 25, rises to a peak two days later on August 27, then declines symmetrically to 1 case on August 29. A late case occurs on August 31 and September 1.



Graphs

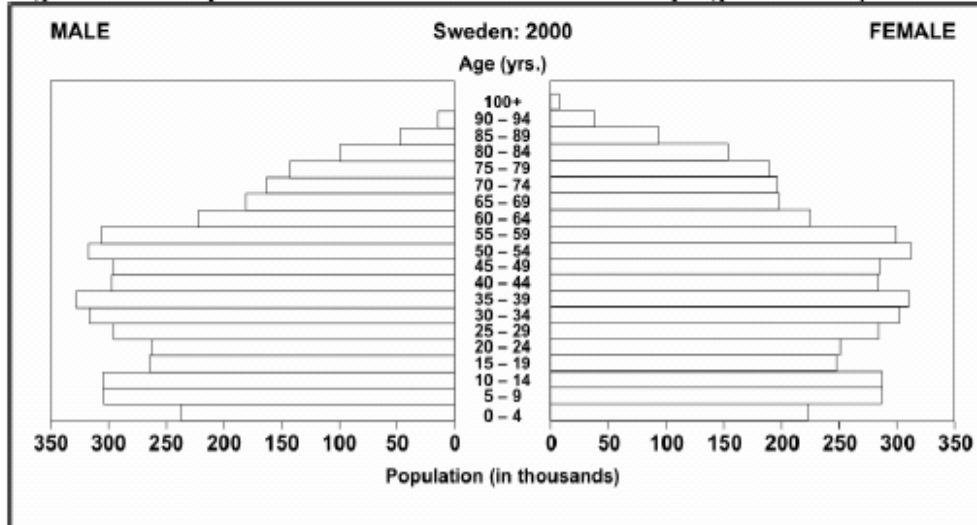
- Population Pyramid:
 - A population pyramid displays the count or percentage of a population by age and sex. It does so by using two histograms — most often one for females and one for males, each by age group — turned sideways so the bars are horizontal, and placed base to base (Figures 4.10 and 4.11)
 - Notice the overall pyramidal shape of the population distribution of a developing country with many births, relatively high infant mortality, and relatively low life expectancy (Figure 4.10)
 - Compare that with the shape of the population distribution of a more developed country with fewer births, lower infant mortality, and higher life expectancy (Figure 4.11)

Figure 4.10 Population Distribution of Zambia by Age and Sex, 2000



Source: U.S. Census Bureau [Internet]. Washington, DC: IDB Population Pyramids [cited 2004 Sep 10]. Available from: <http://www.census.gov/ipc/www/idbpyr.html>.

Figure 4.11 Population Distribution of Sweden by Age and Sex, 1997



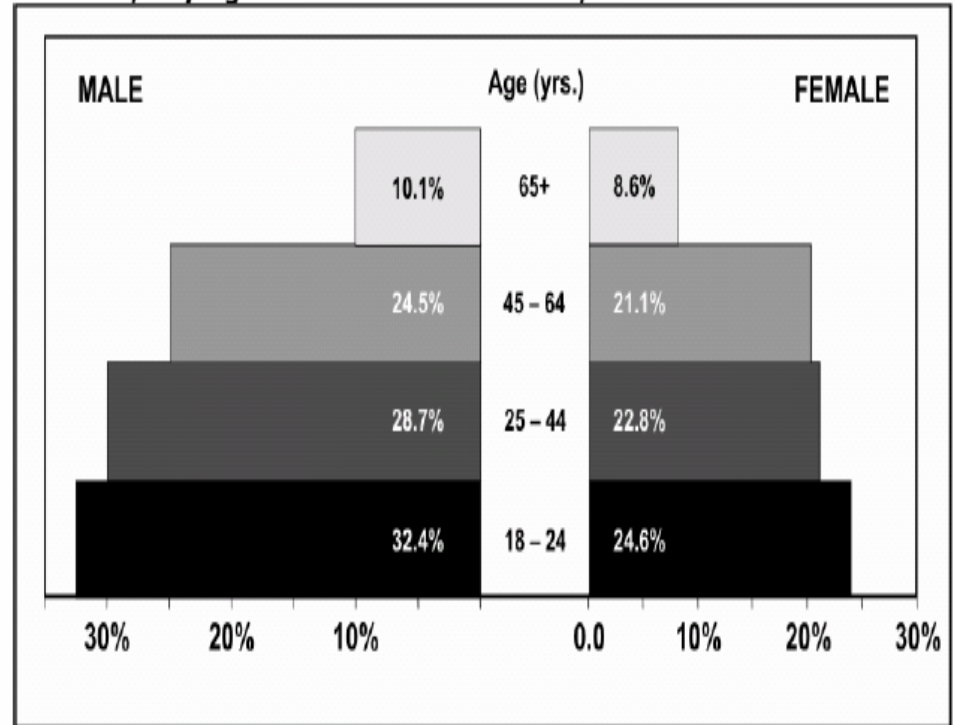
Source: U.S. Census Bureau [Internet]. Washington, DC: IDB Population Pyramids [cited 2004 Sep 10]. Available from: <http://www.census.gov/ipc/www/idbpyr.html>.



Graphs

- While population pyramids are used most often to display the distribution of a national population, they can also be used to display other data such as disease or a health characteristic by age and sex
- For example, smoking prevalence by age and sex is shown in Figure 4.12
- This pyramid clearly shows that, at every age, females are less likely to be current smokers than males

Figure 4.12 Percentage of Persons ≥ 18 Years Who Were Current Smokers,* by Age and Sex—United States, 2002



Answer "yes" to both questions: "Do you now smoke cigarettes everyday or some days?" and "Have you smoked at least 100 cigarettes in your entire life?"

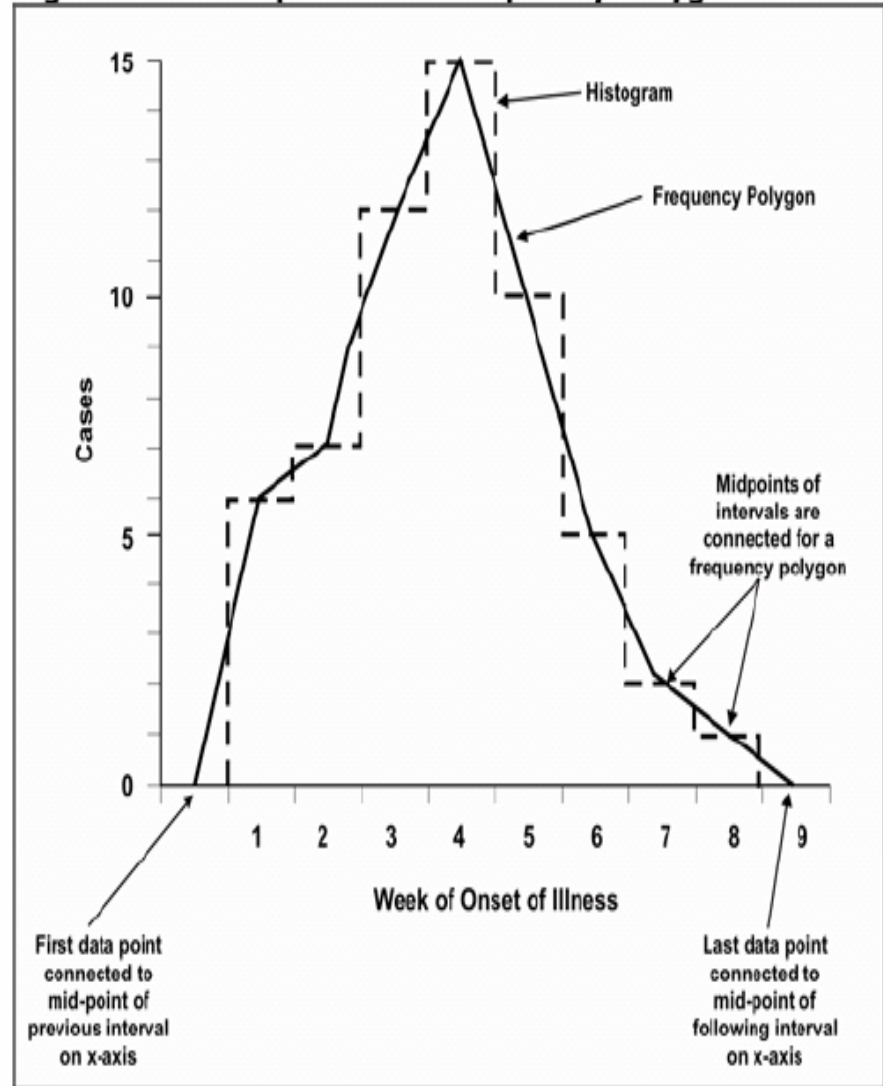
Data Source: Centers for Disease Control and Prevention. Cigarette smoking among adults—United States, 2002. *MMWR* 2004;53:427–31.



Graphs

- Frequency Polygons:
 - A frequency polygon, like a histogram, is the graph of a frequency distribution
 - In a frequency polygon, the number of observations within an interval is marked with a single point placed at the midpoint of the interval
 - Each point is then connected to the next with a straight line
 - Figure 4.13 shows an example of a frequency polygon over the outline of a histogram for the same data.
 - This graph makes it easy to identify the peak of the epidemic (4 weeks)

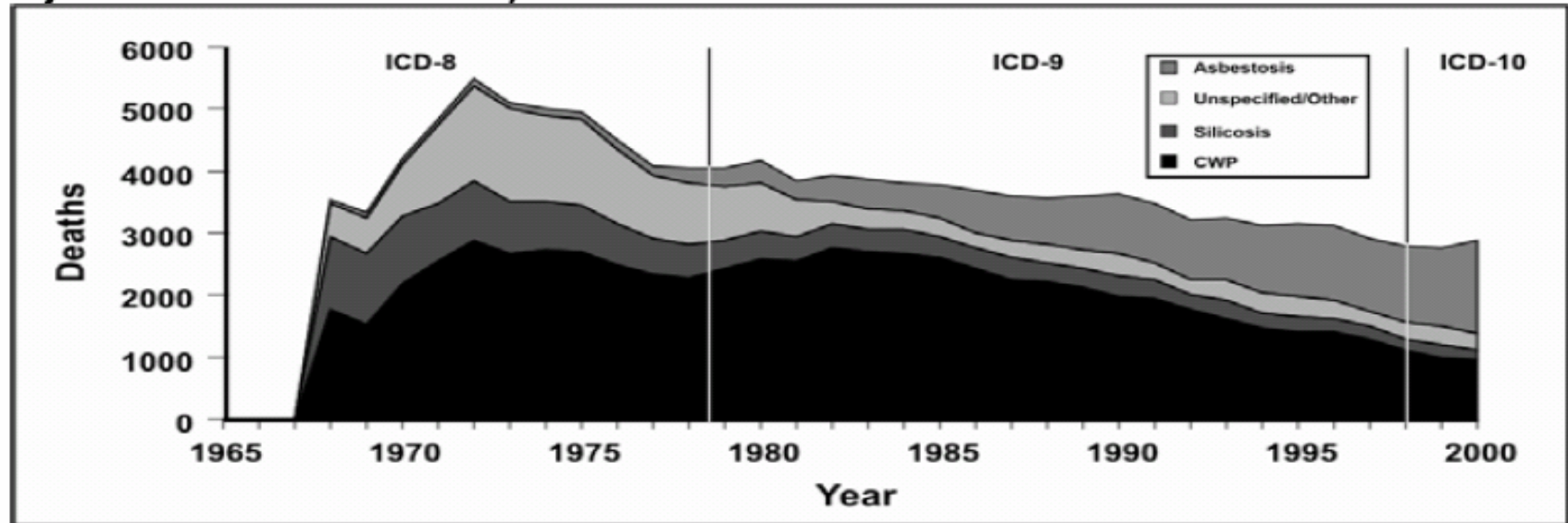
Figure 4.13 Comparison of Frequency Polygon and Histogram



Graphs

- A frequency polygon contains the same area under the line as does a histogram of the same data. Indeed, the data that were displayed as a histogram in Figure 4.9a are displayed as a frequency polygon in Figure 4.14:

Figure 4.14 Number of Deaths with Any Death Certificate Mention of Asbestosis, Coal Worker's Pneumoconiosis (CWP), Silicosis, and Unspecified/Other Pneumoconiosis Among Persons Aged ≥ 15 Years, by Year—United States, 1968–2000



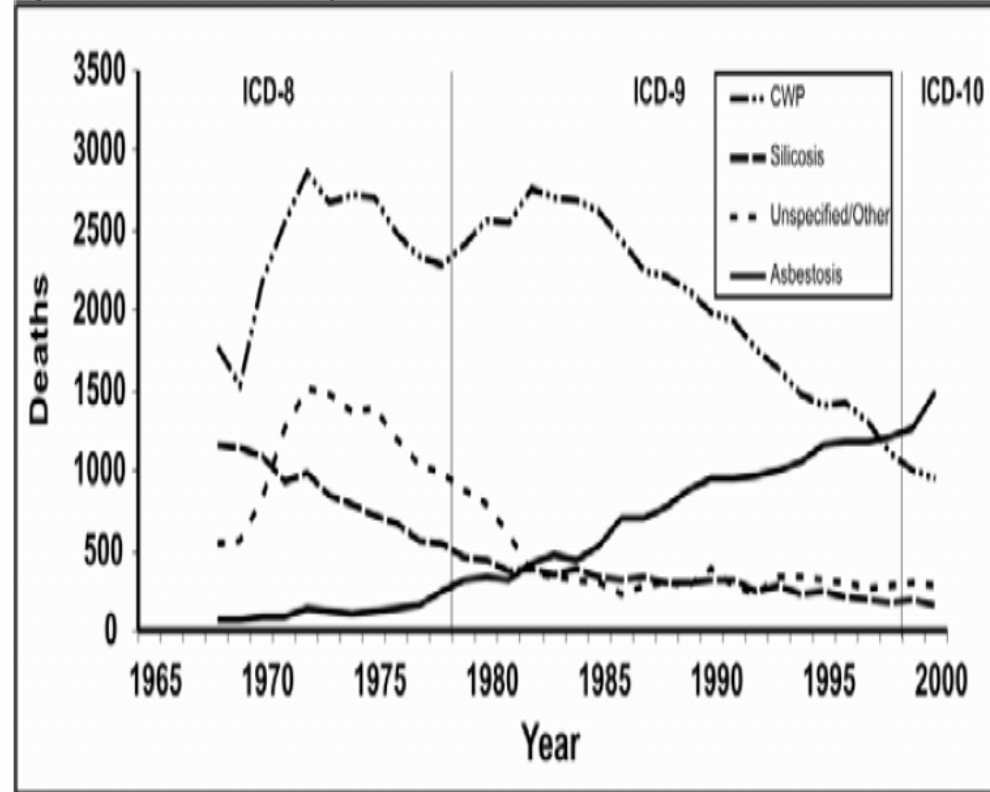
Data Source: Centers for Disease Control and Prevention. Changing patterns of pneumoconiosis mortality—United States, 1968-2000. MMWR 2004;53:627–31.



Graphs

- A frequency polygon differs from an arithmetic-scale line graph in several ways. A frequency polygon (or histogram) is used to display the entire frequency distribution (counts) of a continuous variable
- An arithmetic-scale line graph is used to plot a series of observed data points (counts or rates), usually over time
- A frequency polygon must be closed at both ends because the area under the curve is representative of the data; an arithmetic-scale line graph simply plots the data points
- Compare the pneumoconiosis mortality data displayed as a frequency polygon in Figure 4.14 and as a line graph in Figure 4.15

Figure 4.15 Number of Deaths with Any Death Certificate Mention of Asbestosis, Coal Worker's Pneumoconiosis (CWP), Silicosis, and Unspecified/Other Pneumoconiosis Among Persons Aged ≥ 15 Years, by Year—United States, 1968–2000



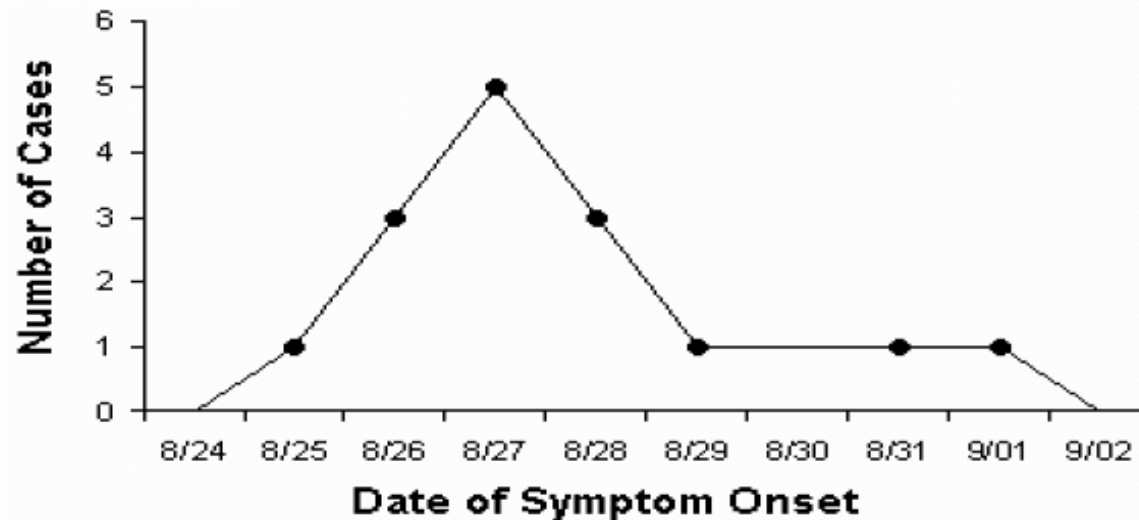
Data Source: Centers for Disease Control and Prevention. Changing patterns of pneumoconiosis mortality—United States, 1968-2000. MMWR 2004;53:627-31.



Graphs

- **Exercise 4.5:**
 - Consider the epidemic curve constructed for Exercise 4.4
 - Prepare a frequency polygon for these same data
 - Compare the interpretations of the two graphs:

**Number of Cases of Botulism by Date of Onset of Symptoms,
Texas Church Supper Outbreak, 2001**



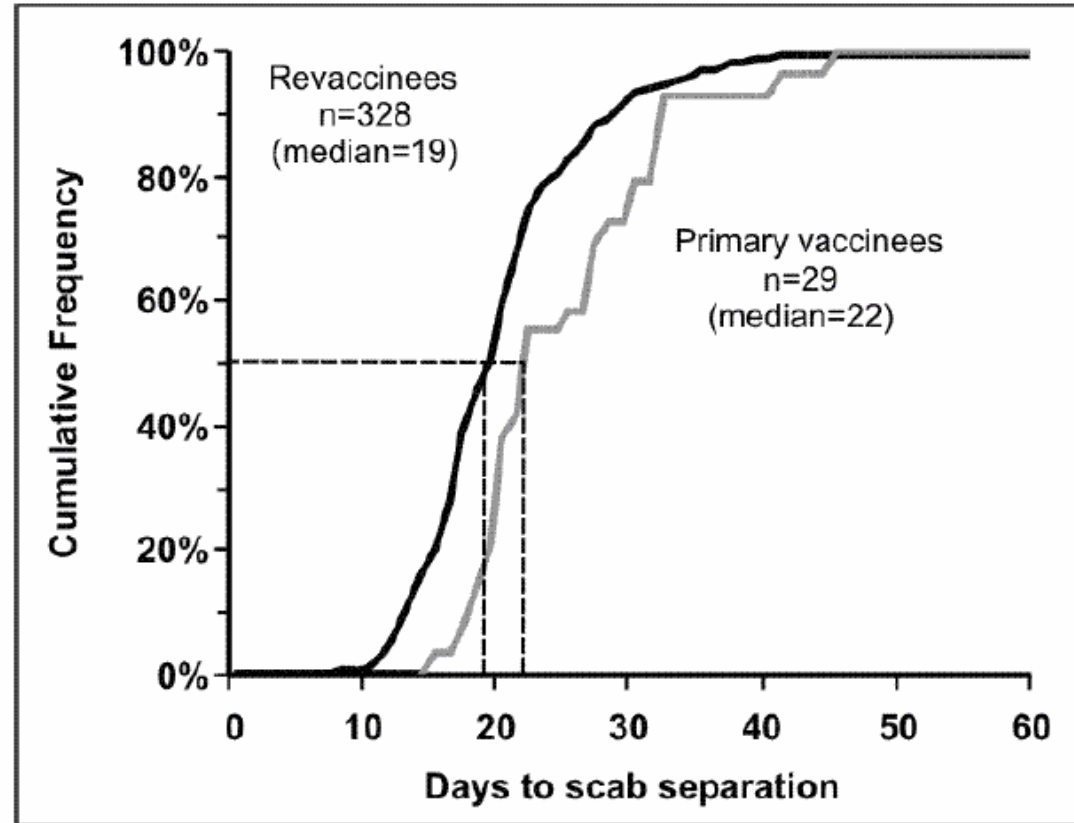
The area under the line in this frequency polygon is the same as the area in the answer to Exercise 4.4 The peak of the epidemic (8/27) is easier to identify.



Graphs

- Cumulative Frequency and Survival Curves:
 - Plots the cumulative, rather than actual, frequency distribution of a variable
 - Useful at identifying medians, quartiles, and other %
 - The x-axis records the class intervals, while the y-axis shows the cumulative frequency either on an absolute scale (e.g., number of cases) or, more commonly, as percentages from 0% to 100%
 - The median (50% or half-way point) can be found by drawing a horizontal line from the 50% tick mark on the y-axis to the cumulative frequency curve, then drawing a vertical line from that spot down to the x-axis

Figure 4.16 Days to Smallpox Vaccination Scab Separation Among Primary Vaccinees (n=29) and Revaccinees (n=328)—West Virginia, 2003



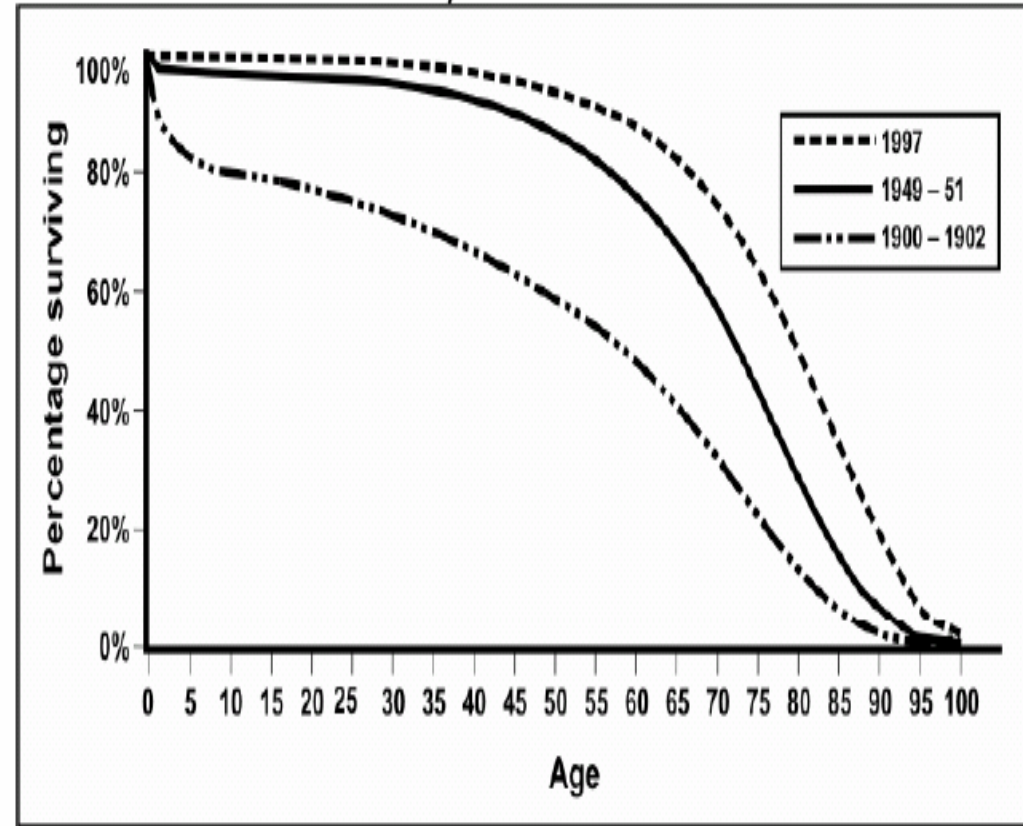
Source: Kaydos-Daniels S, Bixler D, Colsher P, Haddy L. Symptoms following smallpox vaccination—West Virginia, 2003. Presented at 53rd Annual Epidemic Intelligence Service Conference, April 19-23, 2004, Atlanta, Georgia.



Graphs

- A survival curve can be used with follow-up studies to display the proportion of one or more groups still alive at different time periods.
- Similar to the axes of the cumulative frequency curve, the x-axis records the time periods, and the y-axis shows percentages, from 0% to 100%, still alive
- The most striking difference is in the plotted curves themselves
- While a cumulative frequency starts at zero in the lower left corner of the graph and approaches 100% in the upper right corner, a survival curve begins at 100% in the upper left corner and proceeds toward the lower right corner as members of the group die
- The survival curve in Figure 4.17 shows the difference in survival in the early 1900s, mid-1900s, and late 1900s

Figure 4.17 Percent Surviving by Age in Death-registration States, 1900–1902 and United States, 1949–1951 and 1997



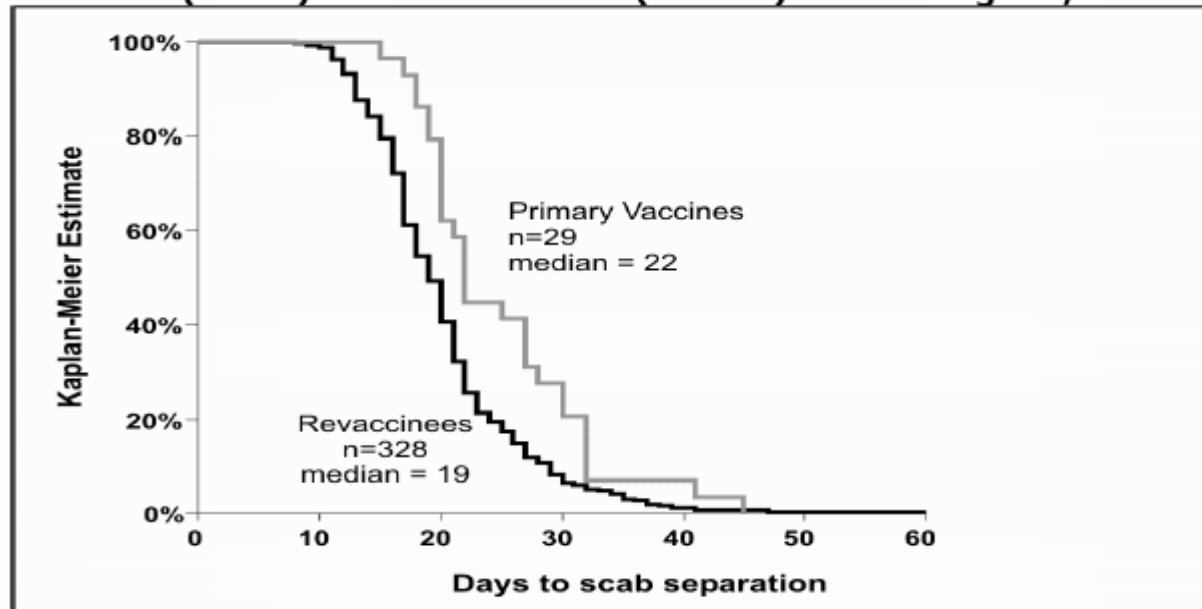
Source: Anderson RN. *United States life tables, 1997*. National vital statistics reports; vol 47, no. 28. Hyattsville, Maryland: National Center for Health Statistics, 1999.



Graphs

- Note that the smallpox scab separation data plotted as a cumulative frequency graph in Figure 4.16 can be plotted as a smallpox scab survival curve, as shown in Figure 4.18:

Figure 4.18 "Survival" of Smallpox Vaccination Scabs Among Primary Vaccines (n=29) and Revaccinees (n=328)—West Virginia, 2003



Source: Kaydos-Daniels S, Bixler D, Colsher P, Haddy L. Symptoms following smallpox vaccination—West Virginia, 2003. Presented at 53rd Annual Epidemic Intelligence Service Conference, April 19-23, 2004, Atlanta, Georgia.



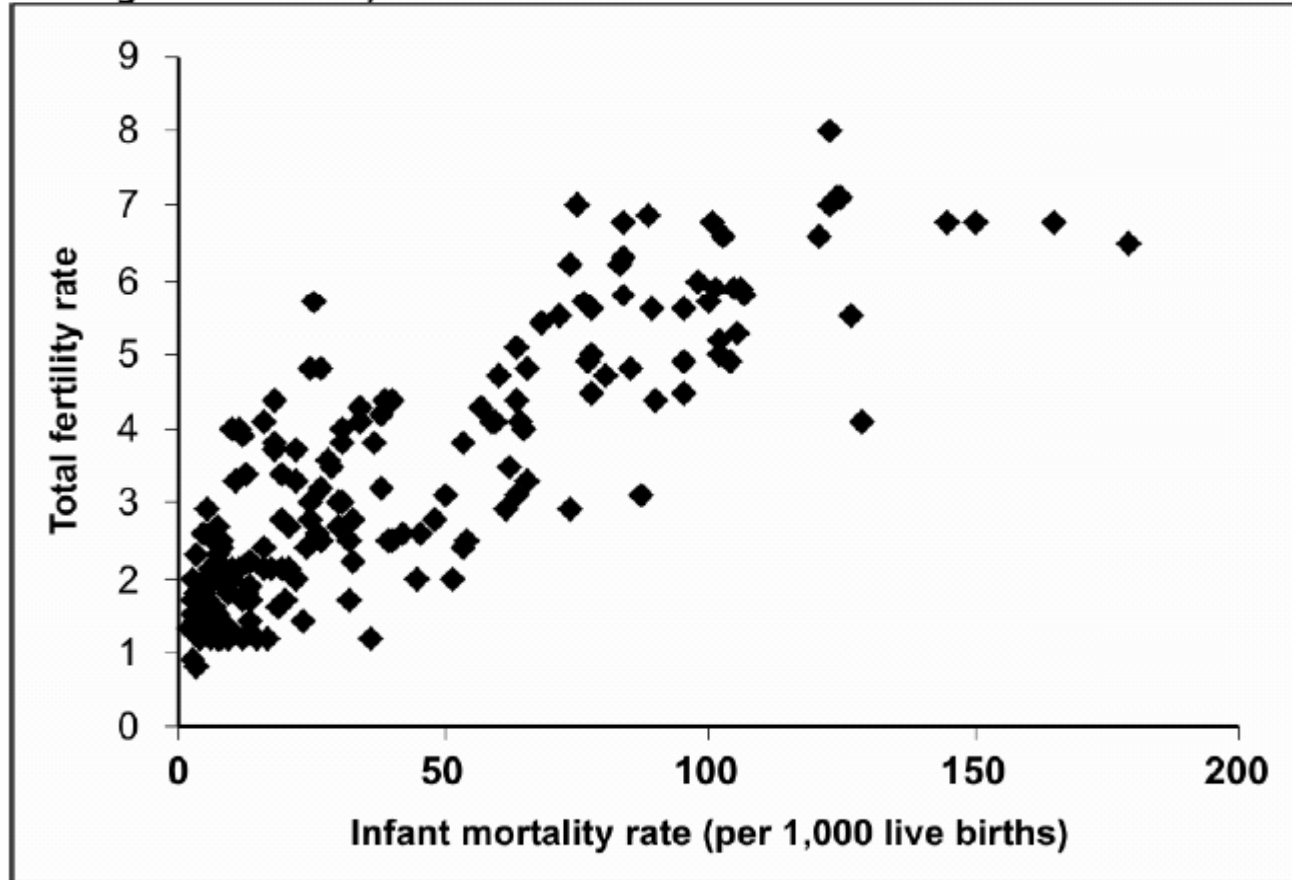
Other Data Displays

- Scatter Diagrams:
 - Graph that portrays the relationship between two continuous variables, with the x-axis representing one variable and the y-axis representing the other
 - To create a scatter diagram you must have a pair of values (one for each variable) for each person, group, country, or other entity in the data set, one value for each variable. A point is placed on the graph where the two values intersect
 - To interpret a scatter diagram, look at the overall pattern made by the plotted points. A fairly compact pattern of points from the lower left to the upper right indicates a positive correlation, in which one variable increases as the other increases
 - A compact pattern from the upper left to lower right indicates a negative or inverse correlation, in which one variable decreases as the other increases
 - Widely scattered points or a relatively flat pattern indicates little correlation
 - Statistical tools such as linear regression can be applied to such data to quantify the correlation between variables in a scatter diagram
 - Similarly, scatter diagrams often display correlations that may provoke intriguing hypotheses about causal relationships, but additional investigation is almost always needed before any causal hypotheses should be accepted



Other Data Displays

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>.



Other Data Displays

- Bar Charts:
 - Uses bars of equal width to display comparative data
 - Comparison of categories is based on the fact that the length of the bar is proportional to the frequency of the event in that category
 - Therefore, breaks in the scale could cause the data to be misinterpreted and should not be used in bar charts
 - Bars for different categories are separated by spaces (unlike the bars in a histogram)
 - The bar chart can be portrayed with the bars either vertical or horizontal (This choice is usually made based on the length of text labels — long labels fit better on a horizontal chart than a vertical one)
 - The bars are usually arranged in ascending or descending length, or in some other systematic order dictated by any intrinsic order of the categories
 - Appropriate data for bar charts include discrete data (e.g., race or cause of death) or variables treated as though they were discrete (age groups)
 - (Recall that a histogram shows frequency of a continuous variable, such as dates of onset of symptoms)



Other Data Displays

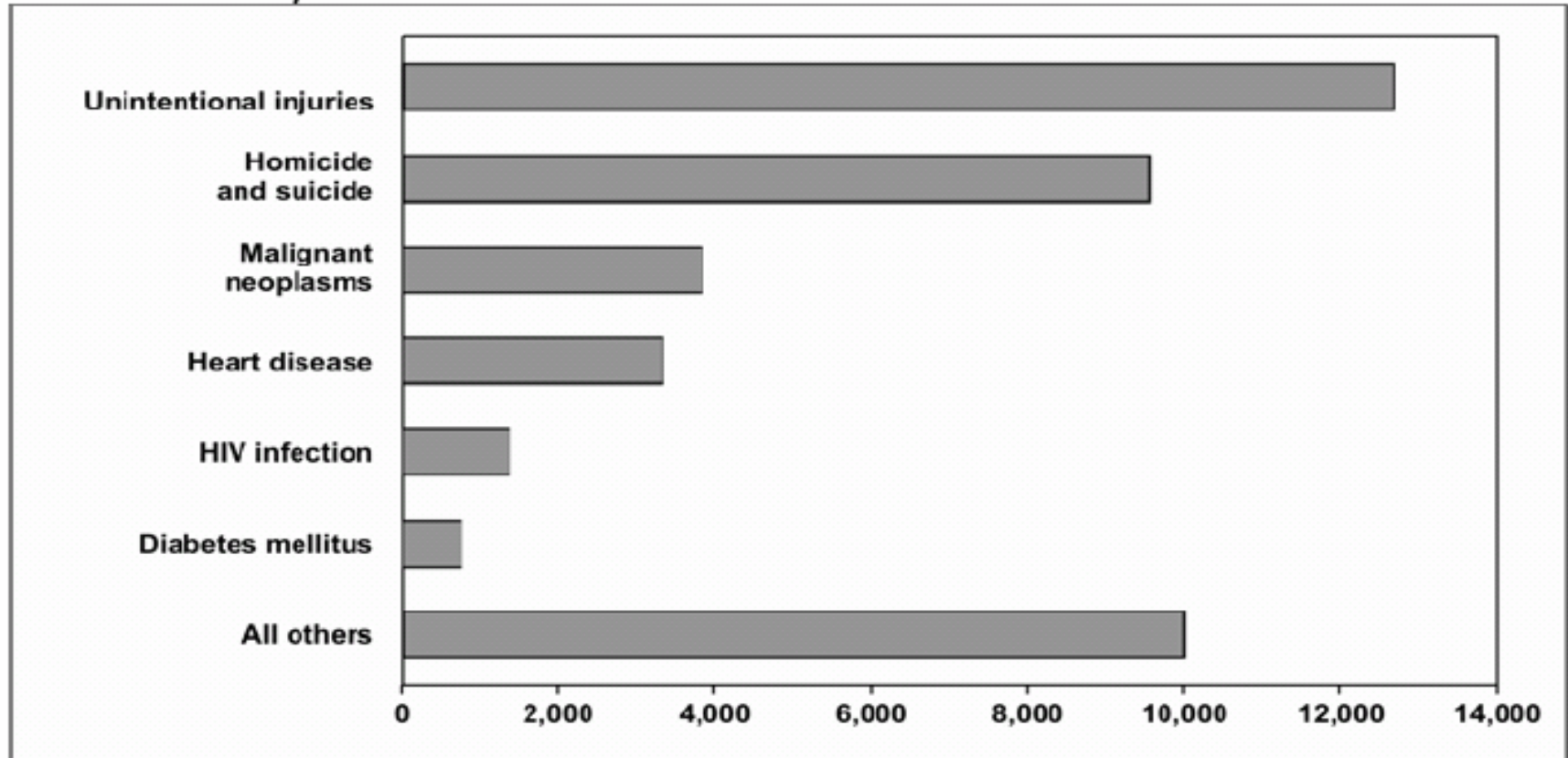
More About Constructing Bar Charts

- Arrange the categories that define the bars or groups of bars in a natural order, such as alphabetical or increasing age, or in an order that will produce increasing or decreasing bar lengths.
- Choose whether to display the bars vertically or horizontally.
- Make all of the bars the same width.
- Make the length of bars in proportion to the frequency of the event. Do not use a scale break, because the reader could easily misinterpret the relative size of different categories.
- Show no more than five bars within a group of bars, if possible.
- Leave a space between adjacent groups of bars but not between bars within a group (see Figure 4.22).
- Within a group, code different variables by differences in bar color, shading, cross hatching, etc. and include a legend that interprets your code.



Other Data Displays

Figure 4.20 Number of Deaths by Cause Among 25–34 Year Olds—United States, 2003



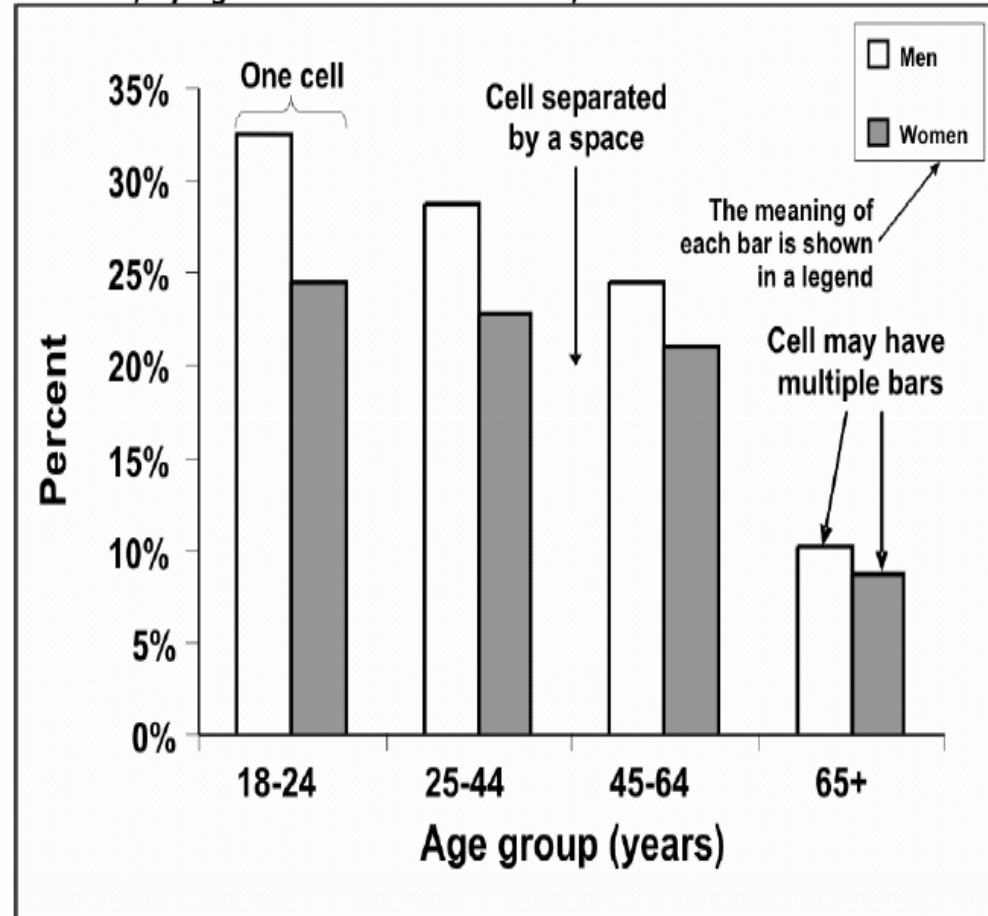
Data Source: Web-based Injury Statistics Query and Reporting System (WISQARS) [online database] Atlanta; National Center for Injury Prevention and Control. [cited 2006 Feb 15]. Available from: <http://www.cdc.gov/ncipc/wisqars>.



Other Data Displays

- Grouped Bar Charts:
 - Used to illustrate data from two-variable or three-variable tables
 - A grouped bar chart is particularly useful when you want to compare the subgroups within a group
 - Bars within a group are adjoining
 - The bars should be illustrated distinctively and described in a legend
 - Consider the data for Figure 4.12 Current smokers by age and sex:
 - Each bar grouping represents an age group
 - Within the group, separate bars are used to represent data for males and females
 - This shows graphically that regardless of age, men are more likely to be current smokers than are women, but that difference declines with age

Figure 4.21 Percentage of Persons Aged ≥ 18 Years Who Were Current Smokers, by Age and Sex—United States, 2002



Data Source: Centers for Disease Control and Prevention. Cigarette smoking among adults—United States, 2002. *MMWR* 2004;53:427–31.



Other Data Displays

Figure 4.22a Number of Deaths by Cause Among 25–34 Year Olds—United States, 1997 and 2003

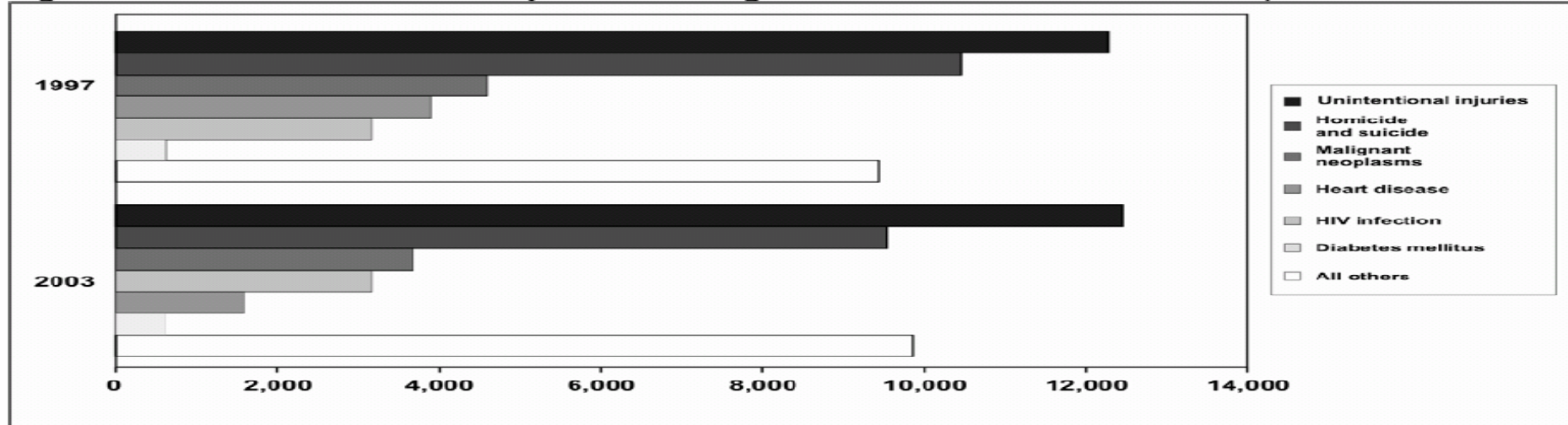
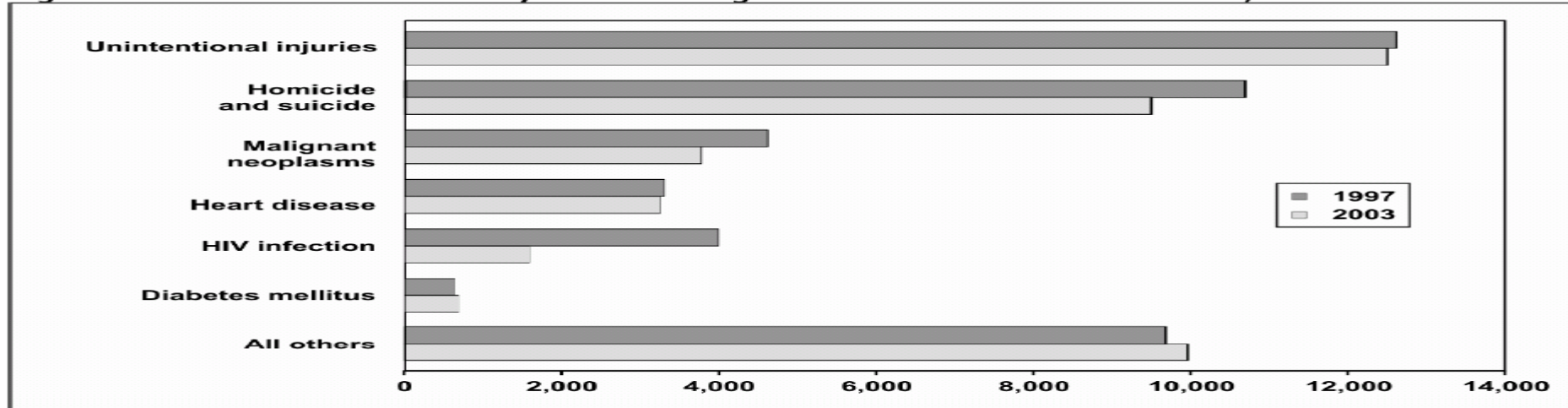


Figure 4.22b Number of Deaths by Cause Among 25–34 Year Olds—United States, 1997 and 2003



Data Source: Web-based Injury Statistics Query and Reporting System (WISQARS) [online database] Atlanta; National Center for Injury Prevention and Control. [cited 2006 Feb 15]. Available from: <http://www.cdc.gov/ncipc/wisqars>.



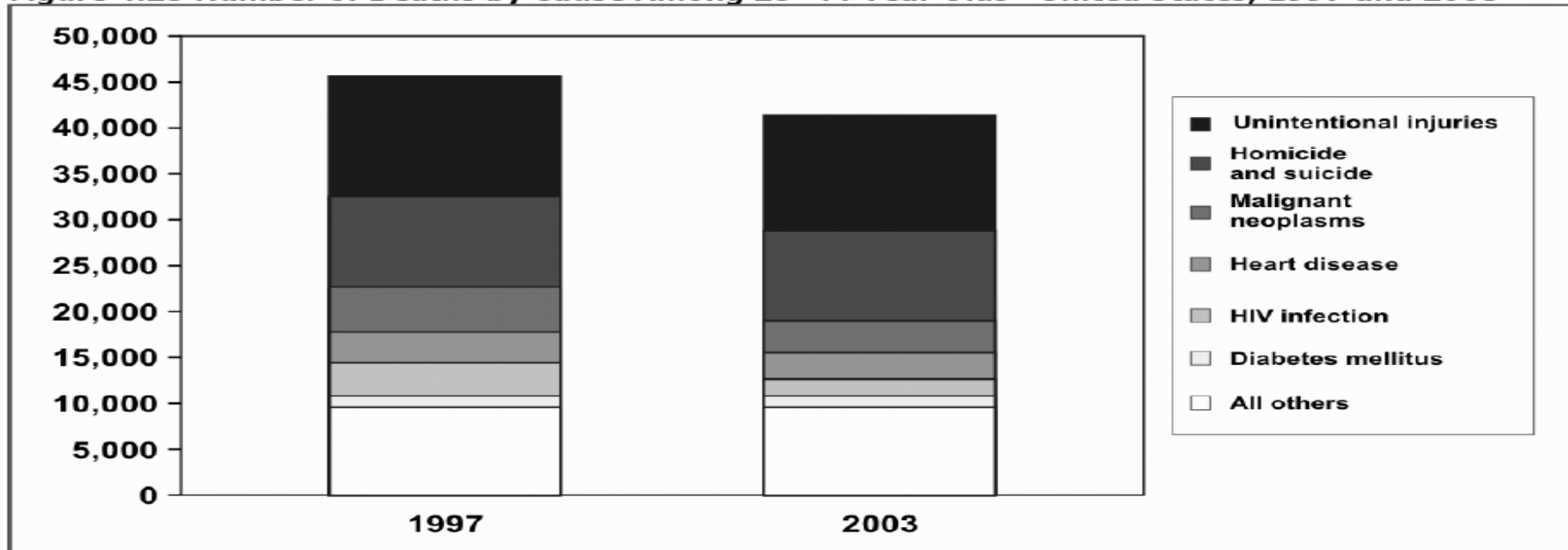
College of Nursing



Other Data Displays

- Stacked Bar Charts:
 - Stacks the subgroups of the 2nd variable into a single bar of the 1st variable
 - Good at comparing variable patterns but poor at reflecting sample size

Figure 4.23 Number of Deaths by Cause Among 25–44 Year Olds—United States, 1997 and 2003



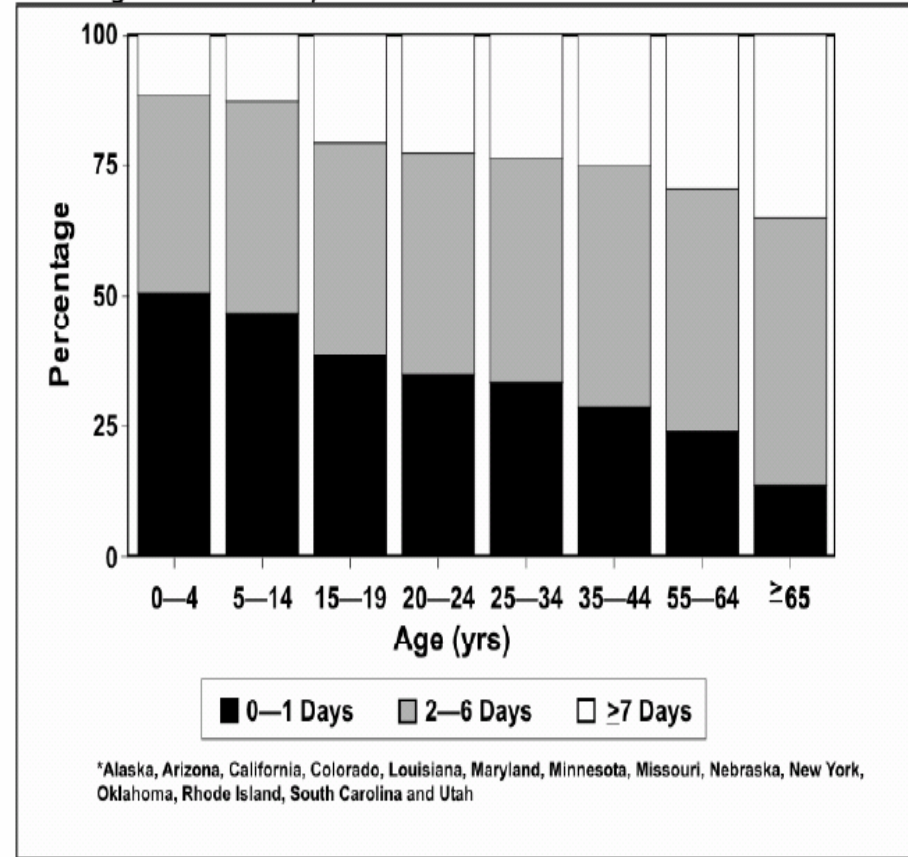
Data Source: Web-based Injury Statistics Query and Reporting System (WISQARS) [online database] Atlanta; National Center for Injury Prevention and Control. [cited 2006 Feb 15]. Available from: <http://www.cdc.gov/ncipc/wisqars>.



Other Data Displays

- 100% Component Bar Charts:
 - Like stacked bar charts; but show in terms of percentage of 100%
 - Useful at comparing contribution of different subgroups within categories of main variable

Figure 4.24 Length of Hospital Stay for Traumatic Brain Injury-related Discharges—14 States*, 1997



*Alaska, Arizona, California, Colorado, Louisiana, Maryland, Minnesota, Missouri, Nebraska, New York, Oklahoma, Rhode Island, South Carolina and Utah

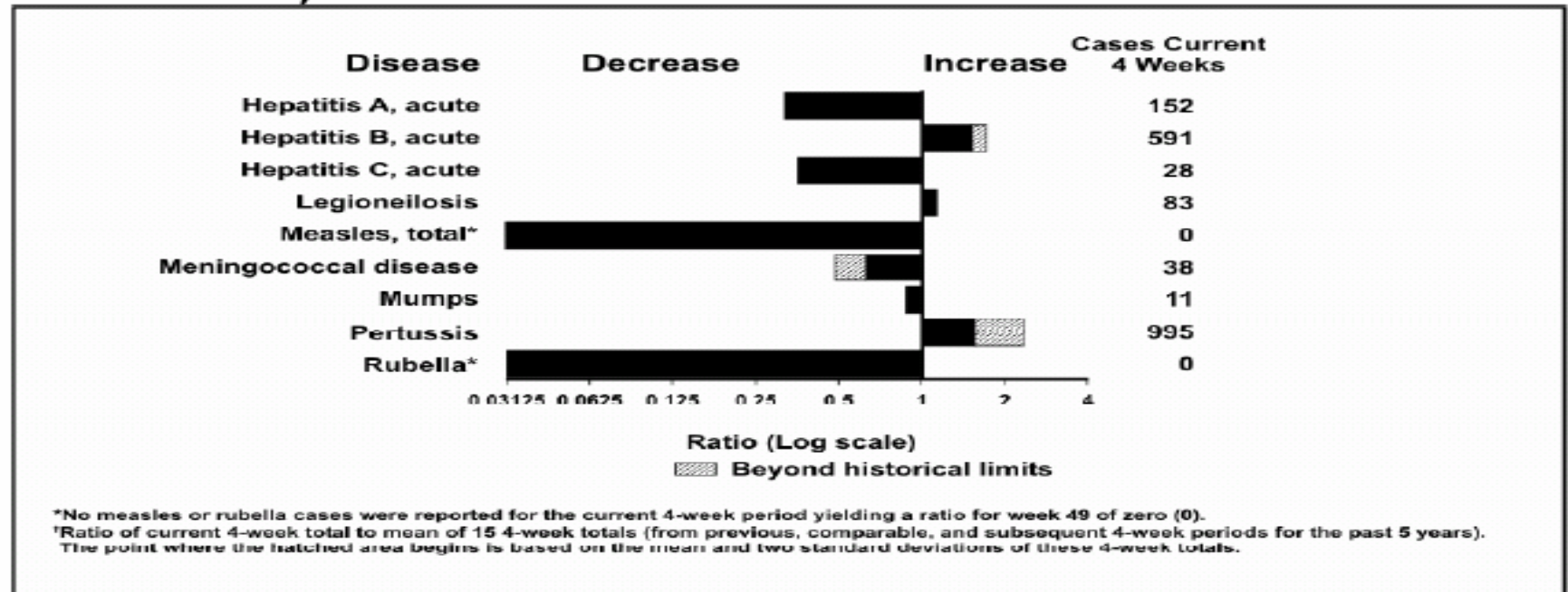
Source: Langlois JA, Kegler SR, Butler JA, Gotsch KE, Johnson RL, Reichard AA, et al. Traumatic brain injury-related hospital discharges: results from a 14-state surveillance system. In: Surveillance Summaries, June 27, 2003. MMWR 2003;52(No. SS-04):1-18.



Other Data Displays

- Deviation Bar Charts
 - Shows trends both above and below a baseline

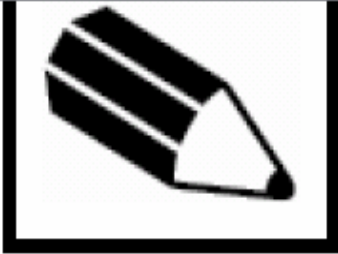
Figure 4.25 Comparison of Current Four-week Totals with Historical Data for Selected Notifiable Diseases—United States, 4-weeks Ending December 11, 2004



Source: Centers for Disease Control and Prevention. Figure 1. Selected notifiable disease reports, United States, comparison of provisional 4-week totals ending December 11, 2004, with historical data. *MMWR* 2004;53:1161.



Other Data Displays



Exercise 4.6

Use the data in Table 4.17 to draw a stacked bar chart, a grouped bar chart, and a 100% component bar chart to illustrate the differences in the age distribution of syphilis cases among white males, white females, black males, and black females. What information is best conveyed by each chart? Graph paper is provided at the end of this lesson.

Table 4.17 Number of Reported Cases of Primary and Secondary Syphilis, by Age Group, Among Non-Hispanic Black and White Men and Women—United States, 2002

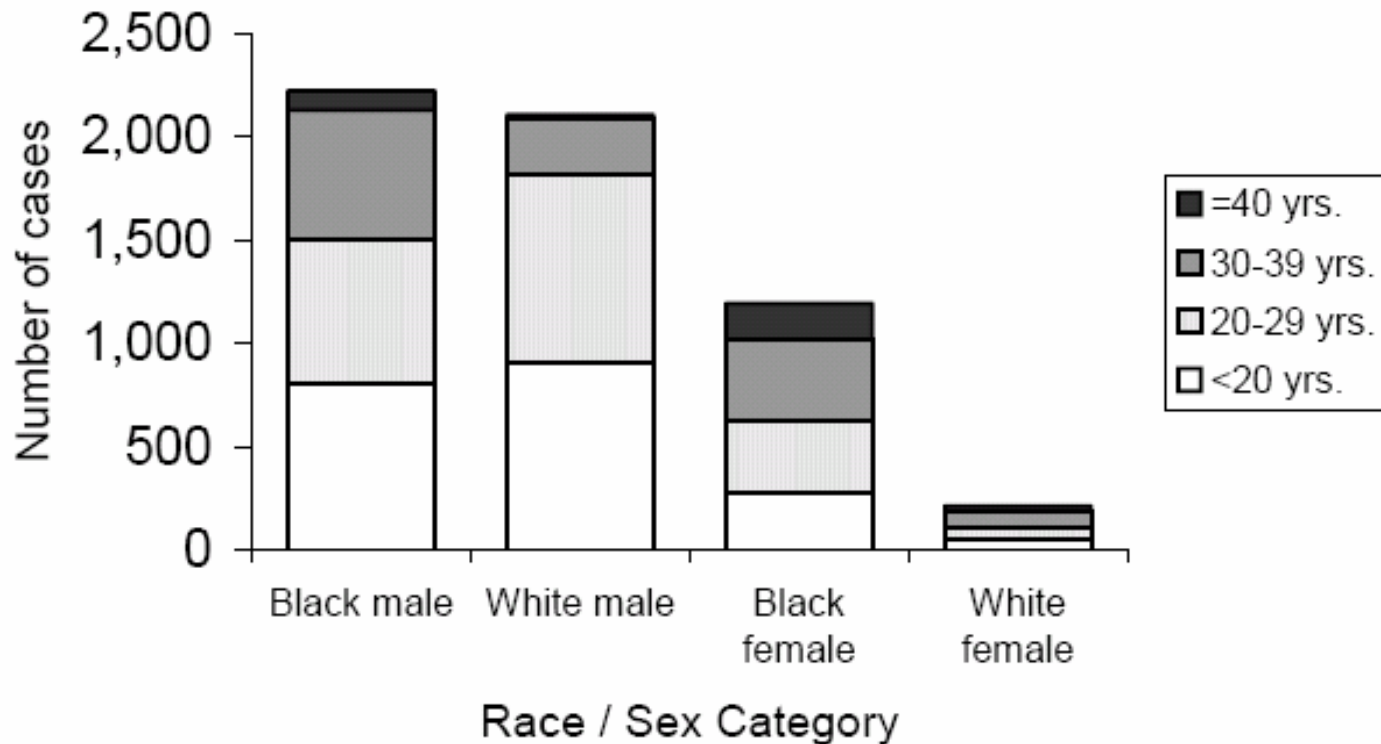
Age Group (Years)	Black Men	White Men	Black Women	White Women
≥40	804	905	277	50
30-39	695	914	349	66
20-29	635	277	396	76
<20	92	12	173	25

Data Source: Centers for Disease Control and Prevention. Sexually Transmitted Disease Surveillance 2002. Atlanta: U.S. Department of Health and Human Services; 2003.



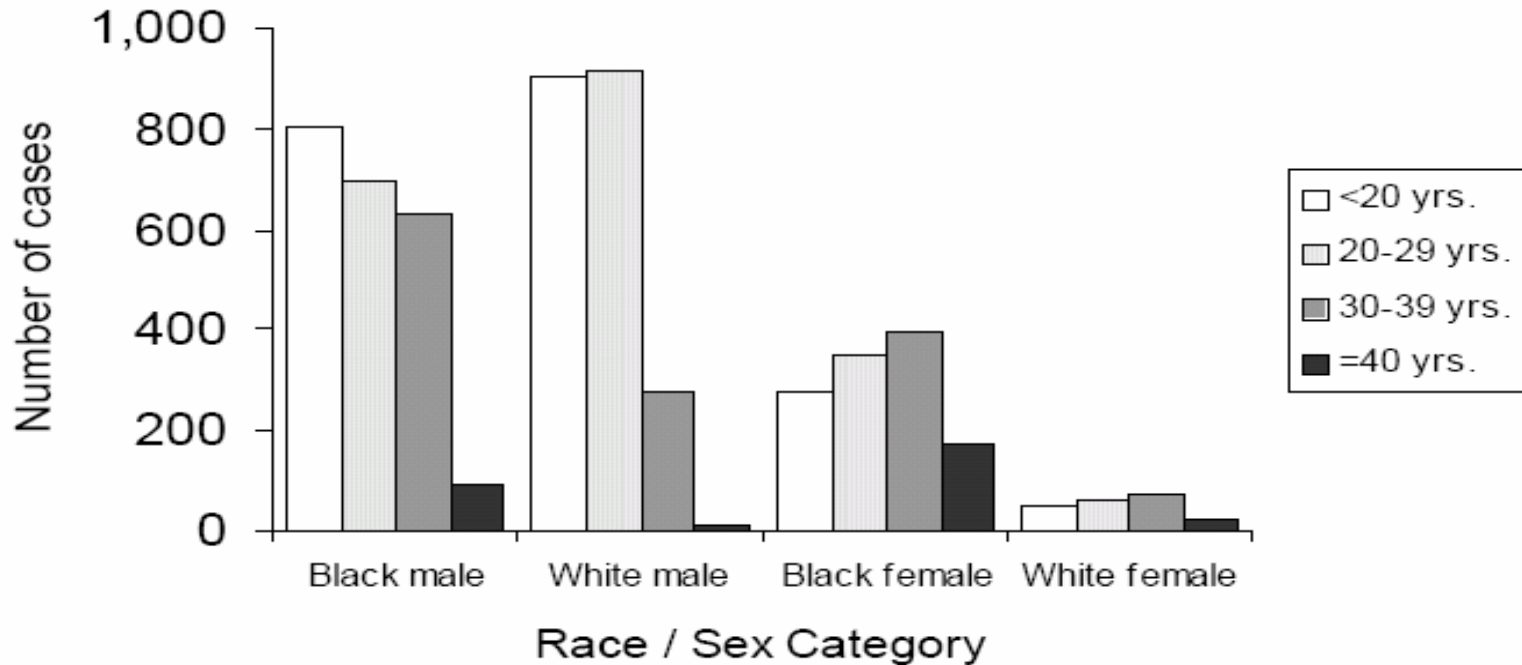
Other Data Displays

Number of Reported Cases of Primary and Secondary Syphilis, by Age Group, Among Non-Hispanic Black and White Men and Women—United States, 2002 (Stacked Bar Chart)



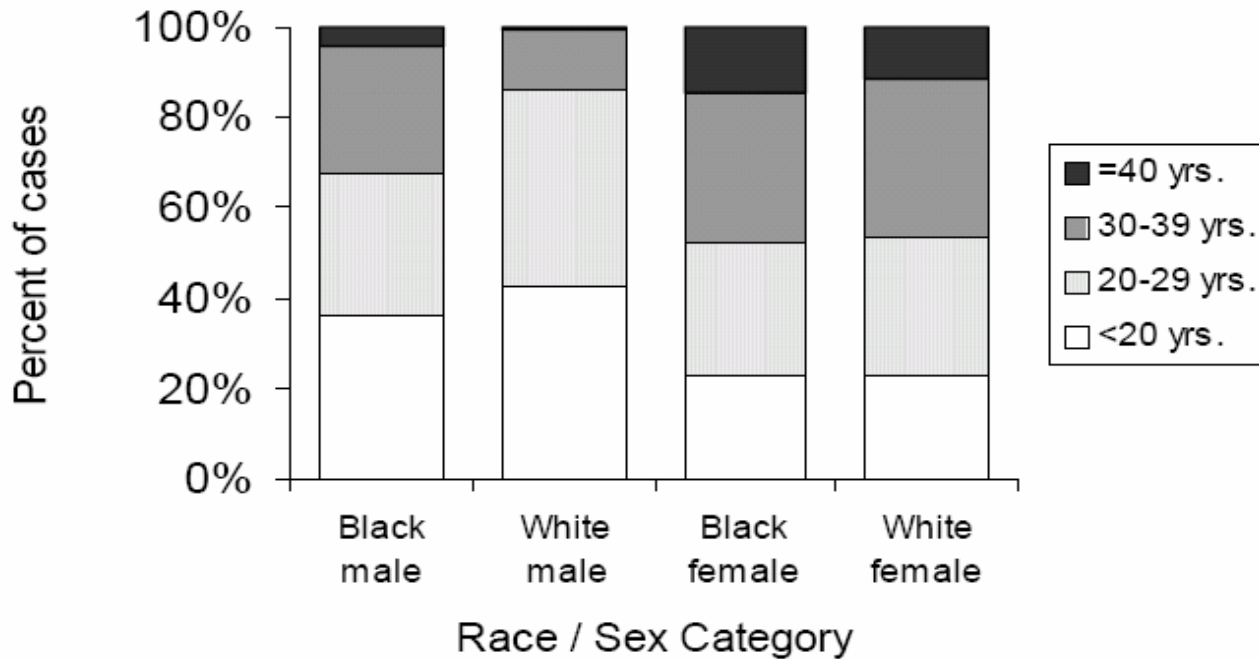
Other Data Displays

Number of Reported Cases of Primary and Secondary Syphilis, by Age Group, Among Non-Hispanic Black and White Men and Women—United States, 2002 (Grouped Bar Chart)



Other Data Displays

Percent of Reported Cases of Primary and Secondary Syphilis, by Age Group, Among Non-Hispanic Black and White Men and Women—United States, 2002 (100% Component Bar Chart)



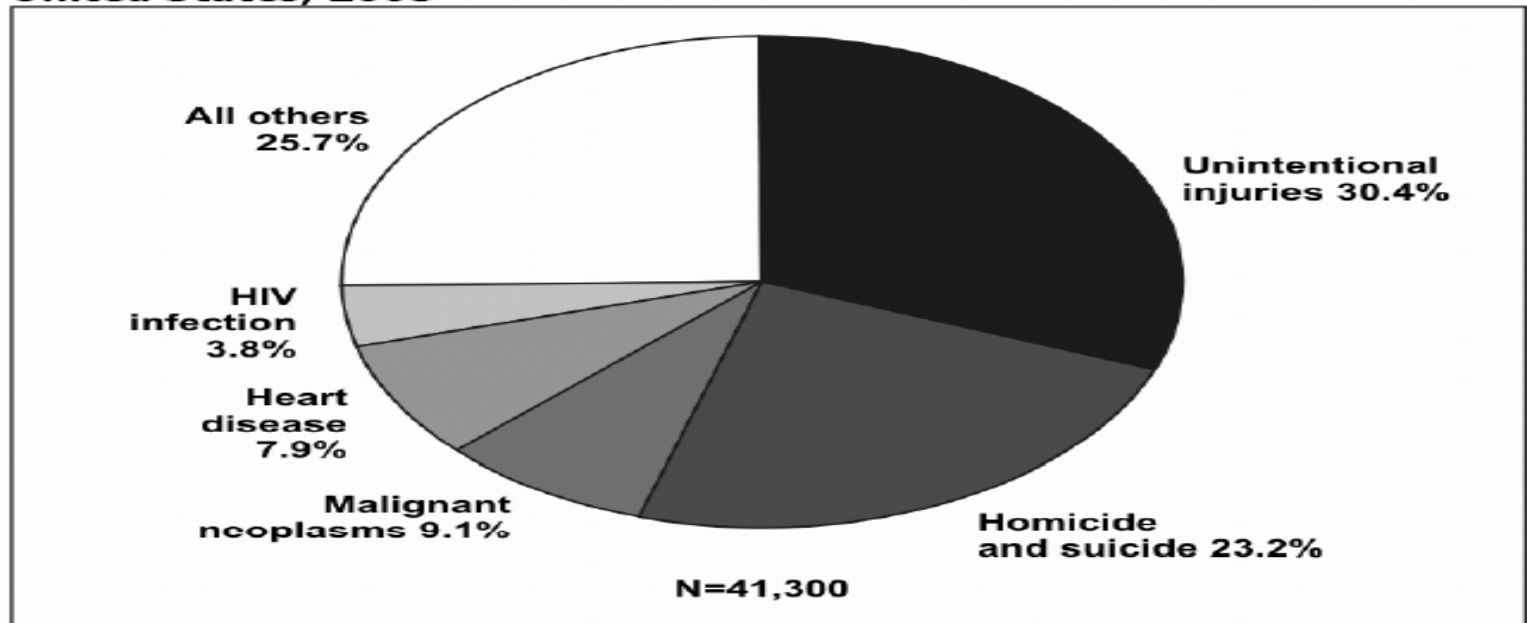
Source: Centers for Disease Control and Prevention. Sexually Transmitted Disease Surveillance 2002. Atlanta, Georgia. U.S. Department of Health and Human Services; 2003.



Other Data Displays

- Pie Charts:
 - A pie chart is a simple, easily understood chart in which the size of the “slices” or wedges shows the proportional contribution of each component part
 - Pie charts are useful for showing the proportions of a single variable’s frequency distribution

Figure 4.26 Number of Deaths by Cause Among 25–34 Year Olds—United States, 2003



Data Source: Web-based Injury Statistics Query and Reporting System (WISQARS) [online database] Atlanta; National Center for Injury Prevention and Control. [cited 2006 Feb 15]. Available from: <http://www.cdc.gov/ncipc/wisqars>.



Other Data Displays

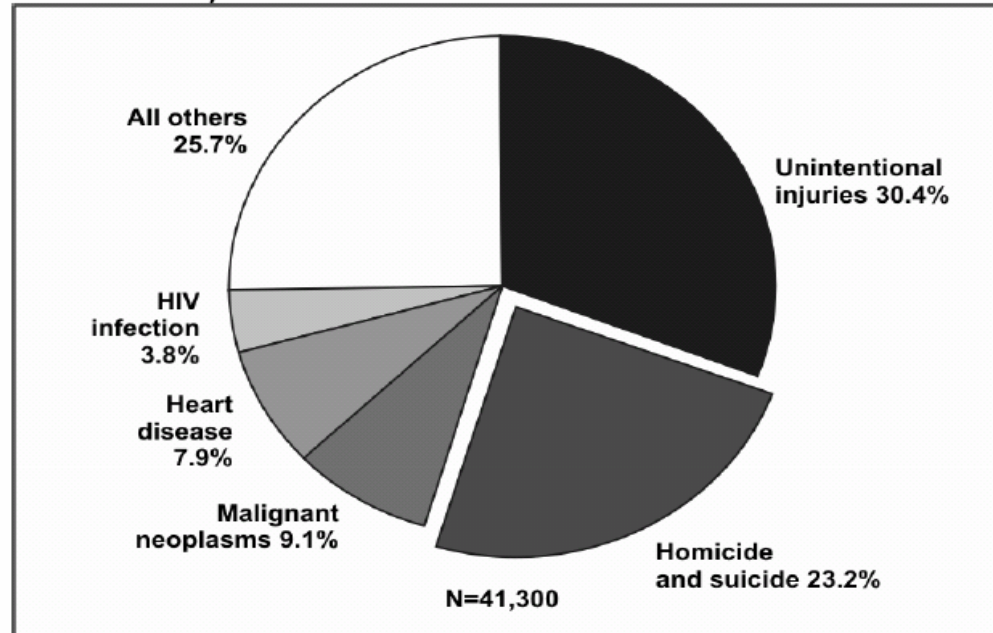
More About Constructing Pie Charts

- Conventionally, pie charts begin at 12 o'clock.
- The wedges should be labeled and arranged from largest to smallest, proceeding clockwise, although the "other" or "unknown" may be last.
- Shading may be used to distinguish between slices but is not always necessary.
- Because the eye cannot accurately gauge the area of the slices, the chart should indicate what percentage each slice represents either inside or near each slice.

-Many computer programs allow one or more slices to "explode" or be pulled out of the pie

-In general, this technique should be limited to situations when you want to place special emphasis on one wedge, particularly when additional detail is provided about that wedge

Figure 4.27 Number of Deaths by Cause Among 25–34 Year Olds—United States, 2003



Data Source: Web-based Injury Statistics Query and Reporting System (WISQARS) [online database] Atlanta; National Center for Injury Prevention and Control. [cited 2006 Feb 15]. Available from: <http://www.cdc.gov/ncipc/wisqars>.



Other Data Displays

Figure 4.28a Number of Deaths by Cause Among 25–34 and 35–44 Year Olds—United States, 2003

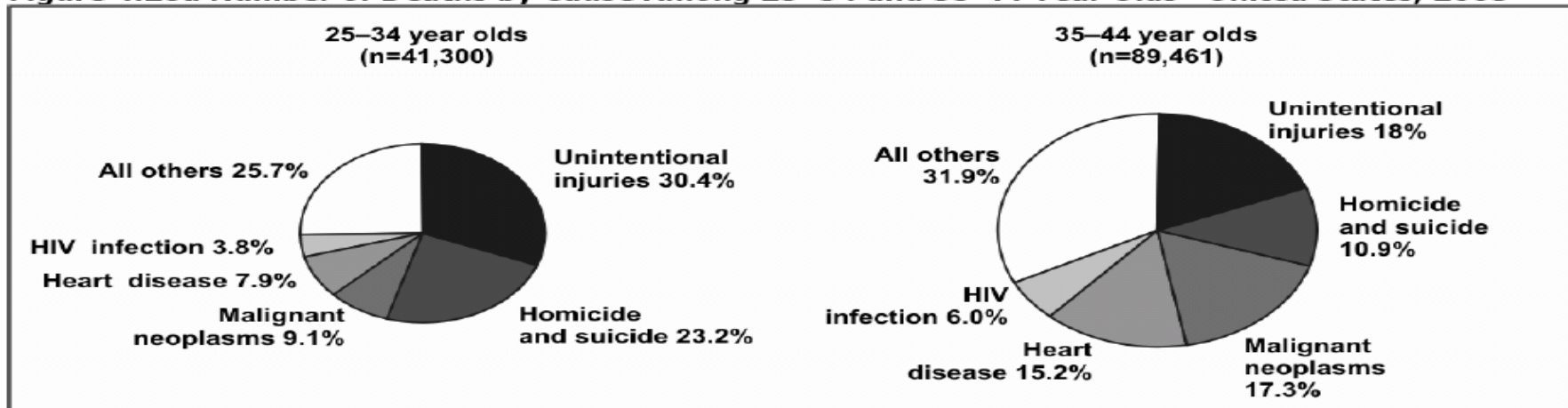
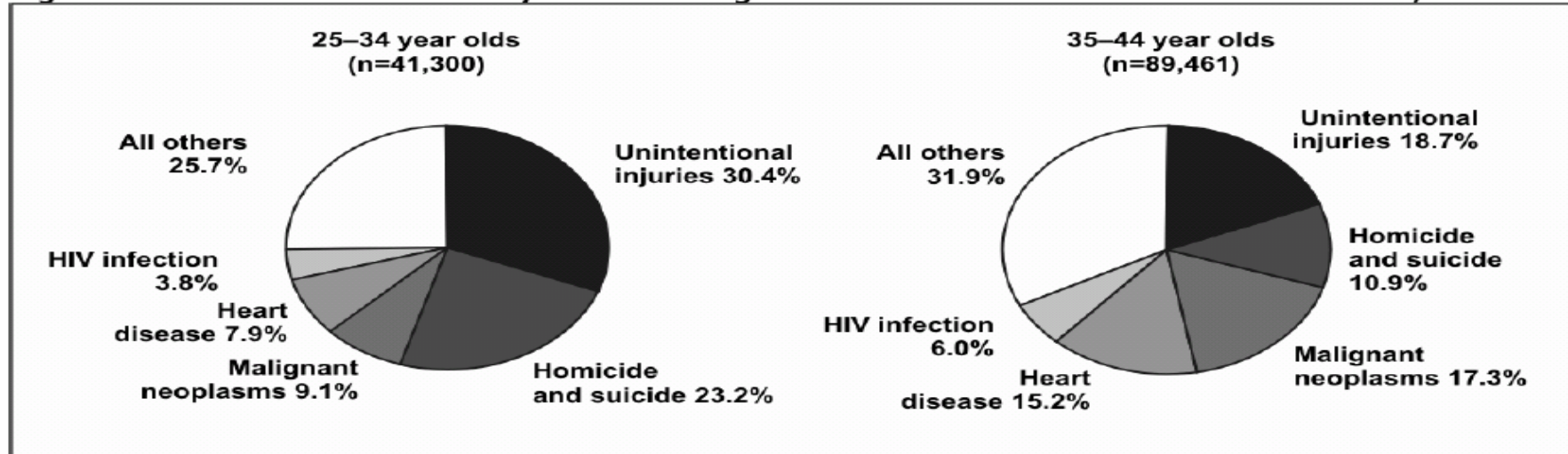


Figure 4.28b Number of Deaths by Cause Among 25–34 and 35–44 Year Olds—United States, 2003



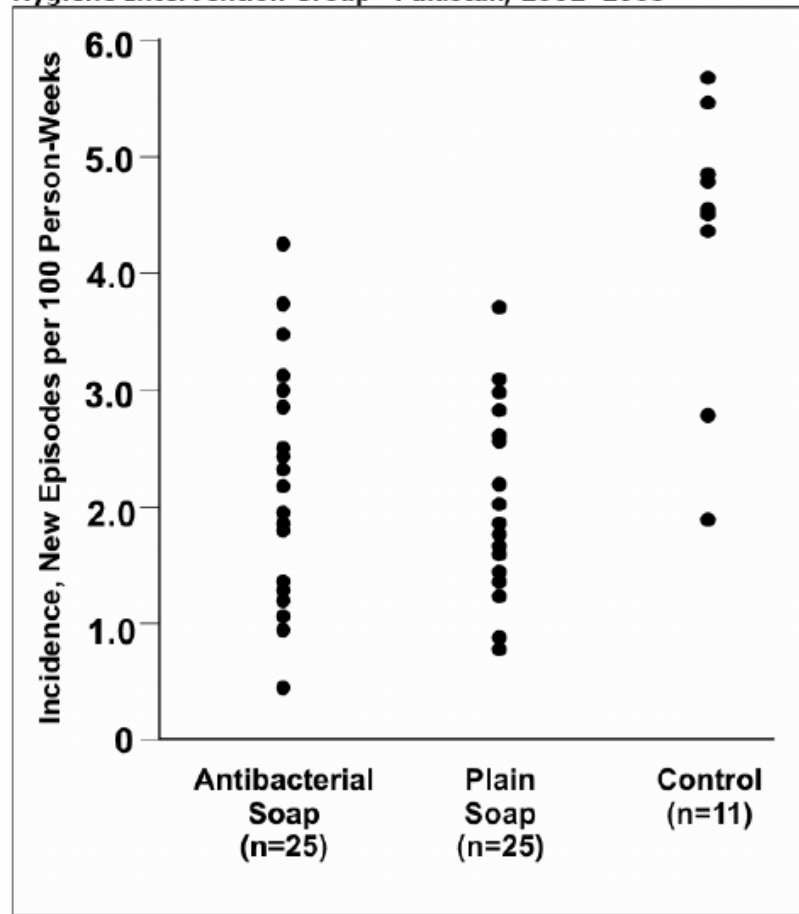
Data Source: Web-based Injury Statistics Query and Reporting System (WISQARS) [online database] Atlanta; National Center for Injury Prevention and Control. [cited 2006 Feb 15]. Available from: <http://www.cdc.gov/ncipc/wisqars>.



Other Data Displays

- Dot/Box Plots:
 - Uses dots to show the relationship between a categorical variable on the x-axis and a continuous variable on the y-axis
 - A dot is positioned at the appropriate place for each observation
 - The dot plot displays not only the clustering and spread of observations for each category of the x-axis variable but also differences in the patterns between categories
 - In Figure 4.29 the villages using either antibacterial soap or plain soap have lower incidence rates of diarrhea than do the control (no soap) villages

Figure 4.29 Incidence of Childhood Diarrhea in Each Neighborhood by Hygiene Intervention Group—Pakistan, 2002–2003



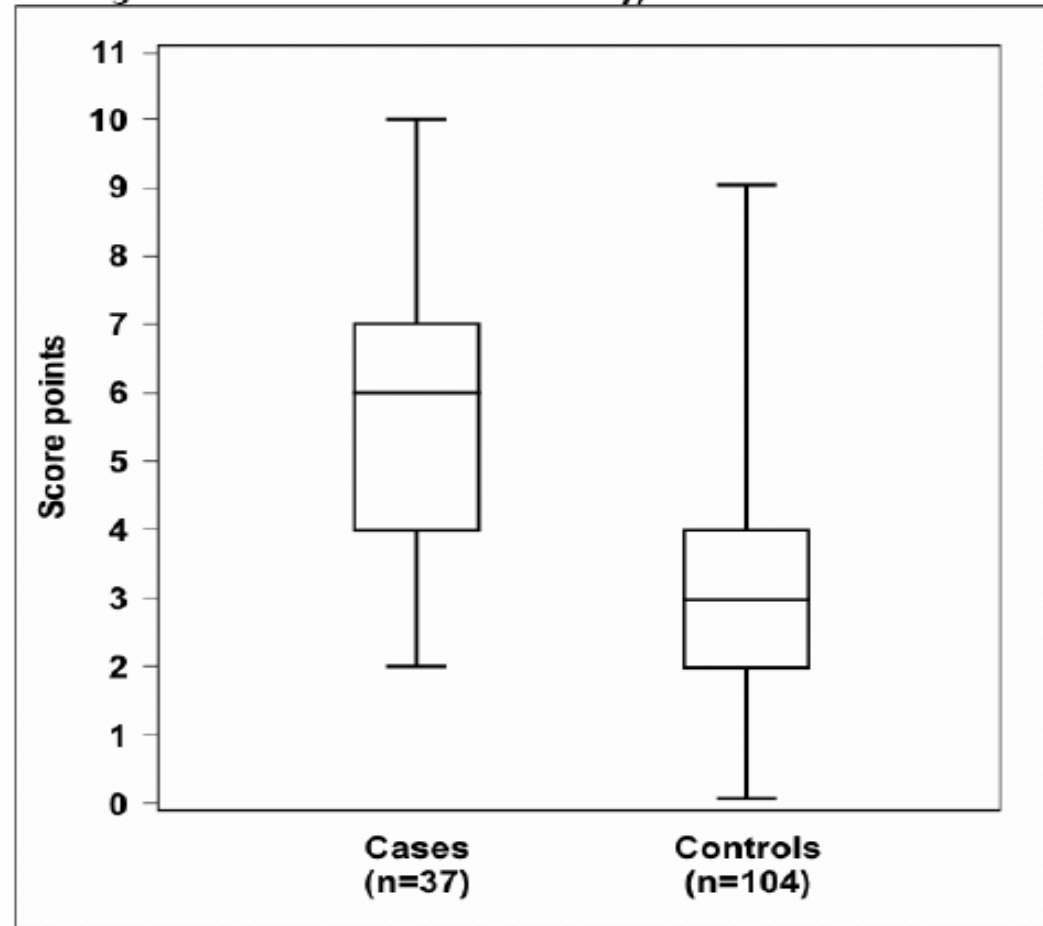
Source: Luby SP, Agboatwalla M, Painter J, Altaf A, Billhimer WL, Hoekstra RM. Effect of intensive handwashing promotion on childhood diarrhea in high-risk communities in Pakistan: a randomized controlled trial. *JAMA* 2004;291:2547–54.



Other Data Displays

- In a box plot, data are summarized by using “box-and whiskers”
- The “box” represents values of the middle 50% (or interquartile range) of the data points, and the “whiskers” extend to the minimum and maximum values that the data assume
- The median is usually marked with a horizontal line inside the box
- As a result, you can use a box plot to show and compare the central location (median), dispersion (interquartile range and range), and skewness (indicated by a median line not centered in the box)

Figure 4.30 Risk Score for Alveolar Echinococcosis Among Cases and Controls—Germany, 1999–2000

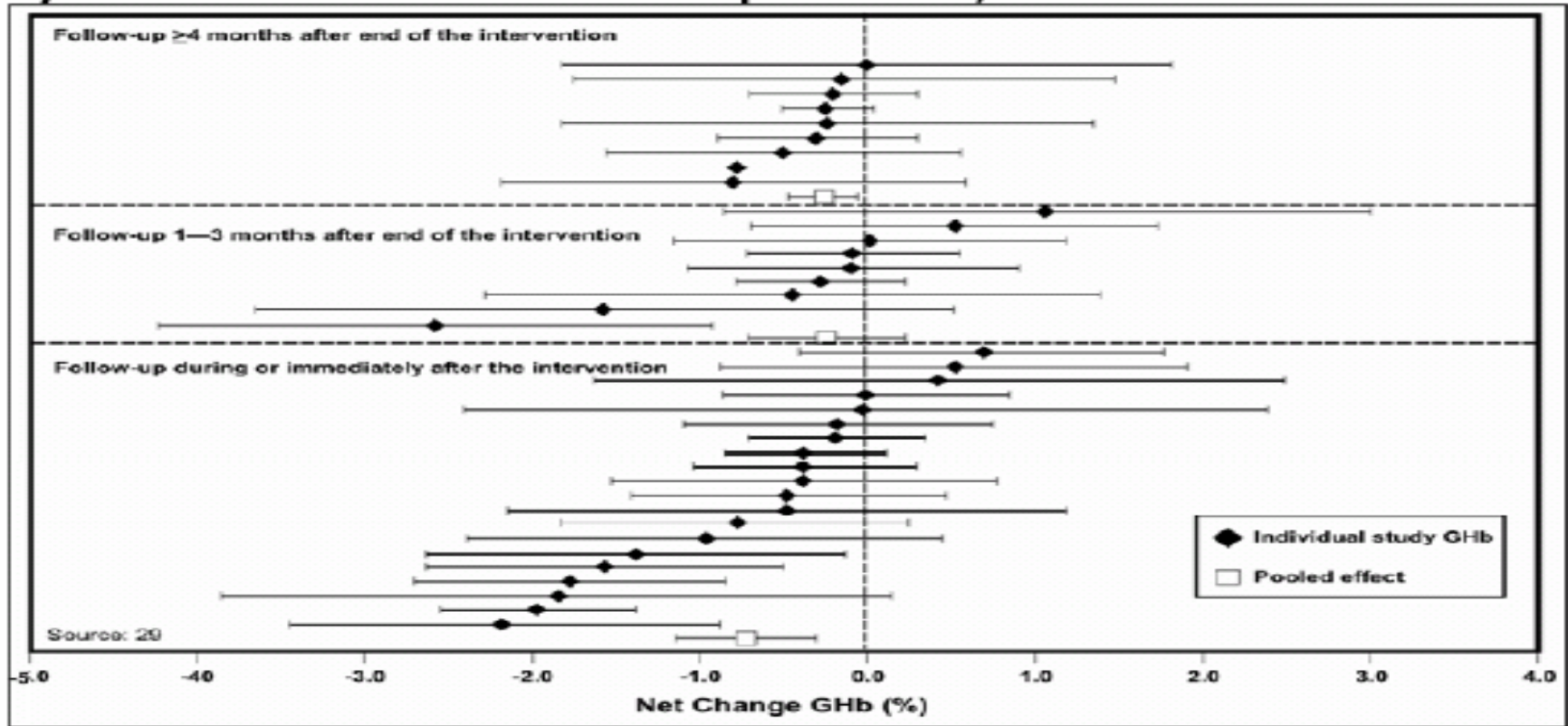


Adapted from: Kern P, Ammon A, Kron M, Sinn G, Sander S, Petersen LR, et al. Risk factors for alveolar echinococcosis in humans. *Emerg Infect Dis* 2004;10:2089-93.



Other Data Displays

Figure 4.31 Net Change in Glycohemoglobin (GHb) Following Self-management Education Intervention for Adults with Type 2 Diabetes, by Different Studies and Follow-up Intervals, 1980–1999

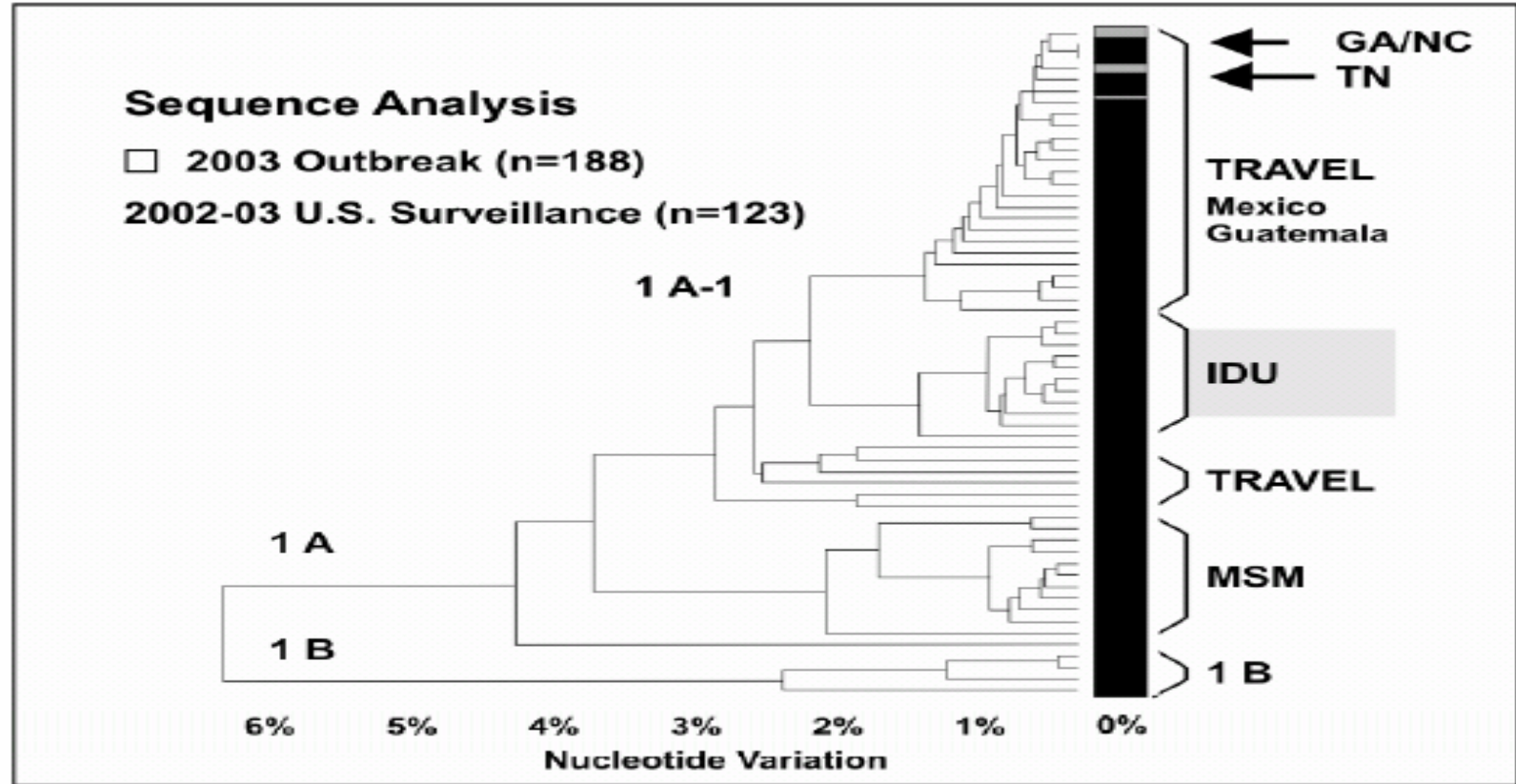


Source: Norris SL, Lau J, Smith SJ, Schmid CH, Engelgau MM. Self-management education for adults with type 2 diabetes. *Diabetes Care* 2002;25:1159–71.



Other Data Displays

Figure 4.32 Comparison of Genetic Sequences of Hepatitis A Virus Isolates from Outbreaks in Georgia, North Carolina, and Tennessee in 2003 with Isolates from National Surveillance



Source: Amon JJ, Devasia R, Guoliang X, Vaughan G, Gabel J, MacDonald P, et al. Multiple hepatitis A outbreaks associated with green onions among restaurant patrons—Tennessee, Georgia, and North Carolina, 2003. Presented at 53rd Annual Epidemic Intelligence Service Conference, April 19-23, 2004, Atlanta, Georgia.



Other Data Displays

- Decision Trees:
 - Branching chart that represents the logical sequence or pathway of a clinical or public health decision
 - Decision analysis is a systematic method for making decisions when outcomes are uncertain
 - The basic building blocks of a decision analysis are (1) decisions, (2) outcomes, and (3) probabilities
 - A decision is a choice made by a person, group, or organization to select a course of action from among a set of mutually exclusive alternatives
 - The decision maker compares expected outcomes of available alternatives and chooses the best among them
 - This choice is represented by a decision node, a square, with branches representing the choices in the decision-tree diagram
 - For example, after receiving information that a person has a family history of a disease (colorectal cancer for this example), that person may decide (choose) to seek medical advice or choose not to do so
 - **Outcomes** are the chance events that occur in response to a decision
 - Outcomes can be intermediate or final. Intermediate outcomes are followed by more decisions or chance events



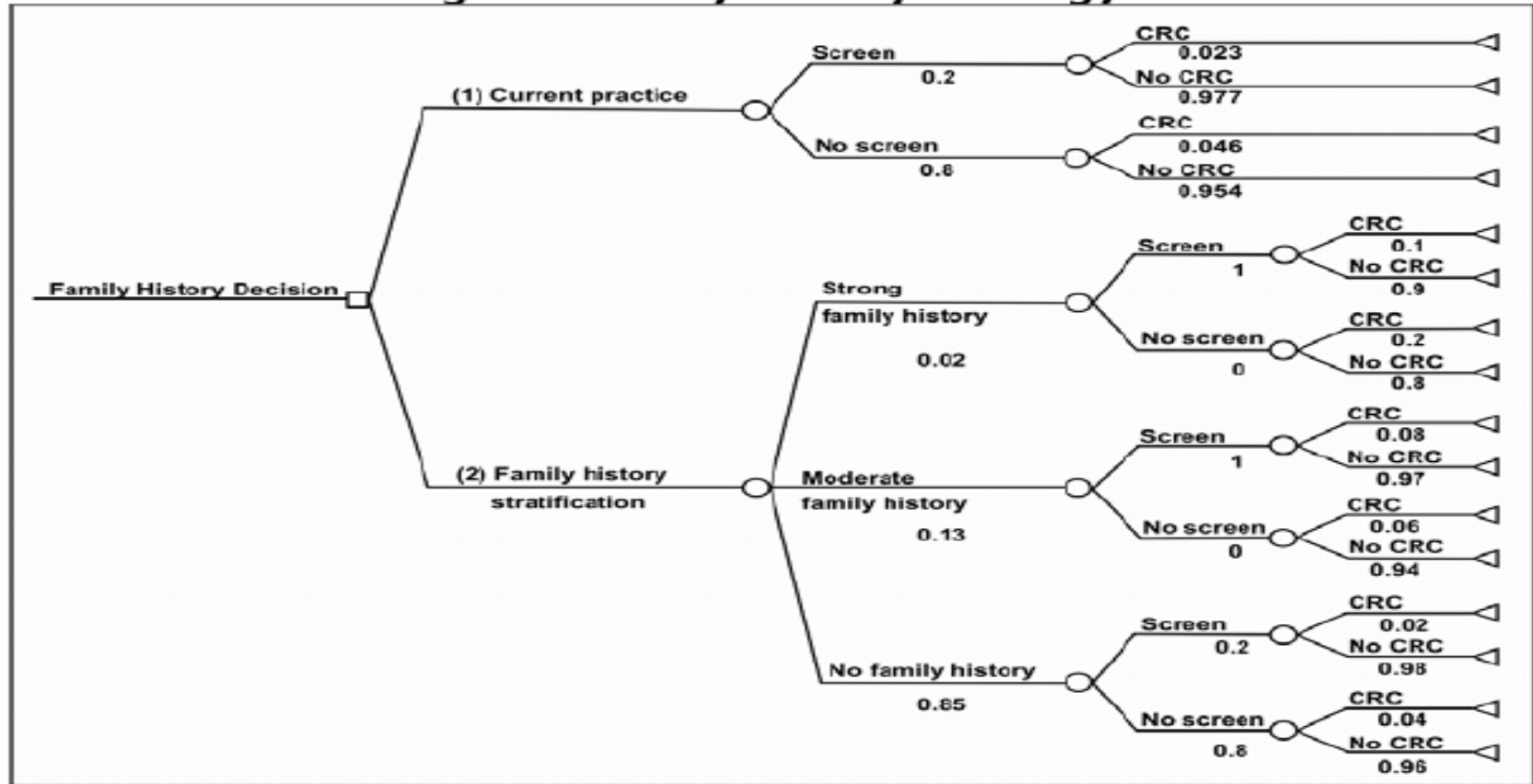
Other Data Displays

- For example, if a person decides to seek medical care for colorectal cancer screening, depending on the findings (outcomes) of the screening, his or her ARNP may advise diet or more frequent screenings; some combination of these two; or treatment
- From the person's perspective, this is a chance outcome; from a health-care provider's perspective, it is a decision
- Whether an outcome is intermediate or final may depend on the context of the decision problem
- For example, colorectal cancer screening may be the final outcome in a decision analysis focusing on colorectal cancer as the health condition of interest, but it may be an intermediate outcome in a decision analysis focusing on more invasive cancer treatment
- In a decision tree, outcomes follow a chance node, a circle, with branches representing different outcomes that occur by chance, one and only one of which occurs
- Each chance outcome has a probability by which it can occur written below the branch in a decision-tree diagram
- The sum of probabilities for all outcomes that can occur at a chance node is one
- The building blocks of decision analysis — decisions, outcomes, and probabilities — can be used to represent and examine complex decision problems



Other Data Displays

Figure 4.33 Decision Tree Comparing Colorectal Screening Current Practice with a Targeted Family History Strategy



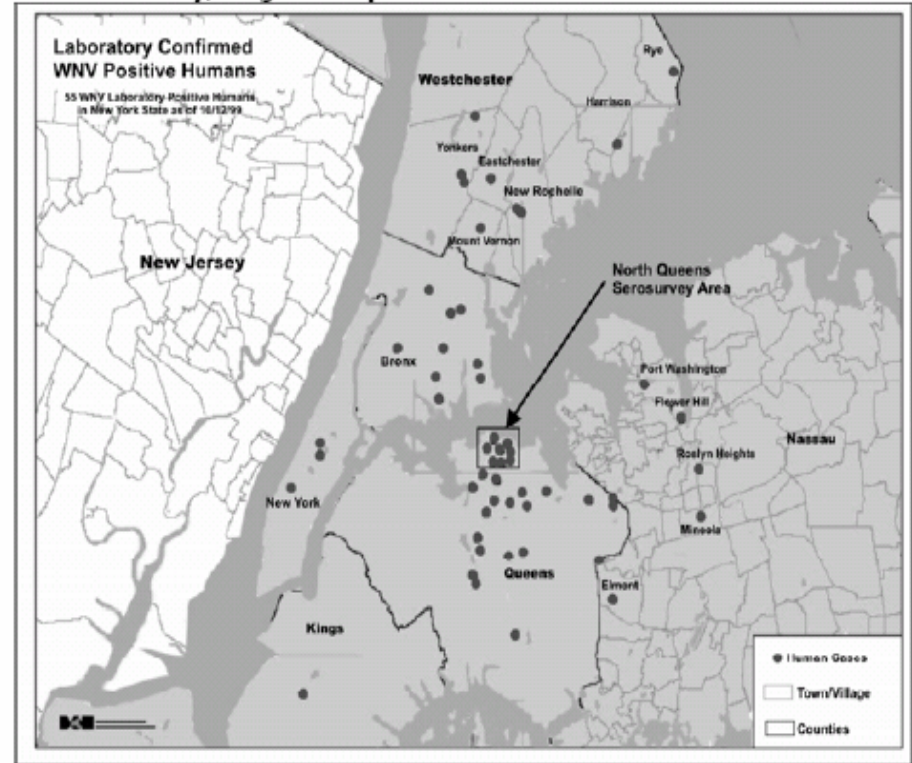
Source: Tyagi A, Morris J. Using decision analytic methods to assess the utility of family history tools. *Am J Prev Med* 2003;24:199–207.



Other Data Displays

- Maps:
 - Used to show the geographic location of events or attributes
 - Two types of maps commonly used in field epidemiology are spot maps and area maps
 - Spot maps use dots or other symbols to show where each case-patient lived or was exposed
 - Figure 4.34 is a spot map of the residences of persons with West Nile Virus encephalitis during the outbreak in the New York City area in 1999:

Figure 4.34 Laboratory-confirmed Cases of West Nile Virus Disease—New York City, August–September 1999



Source: Nash D, Mostashari F, Murray K, et al. Recognition of an outbreak of West Nile Virus disease. Presented at 49th Annual Epidemic Intelligence Service Conference, April 10–14, 2000, Atlanta, Georgia.



Other Data Displays

- A spot map is useful for showing the geographic distribution of cases, but because it does not take the size of the population at risk into account a spot map does not show risk of disease
- Even when a spot map shows a large number of dots in the same area, the risk of acquiring disease may not be particularly high if that area is densely populated

More About Constructing Maps

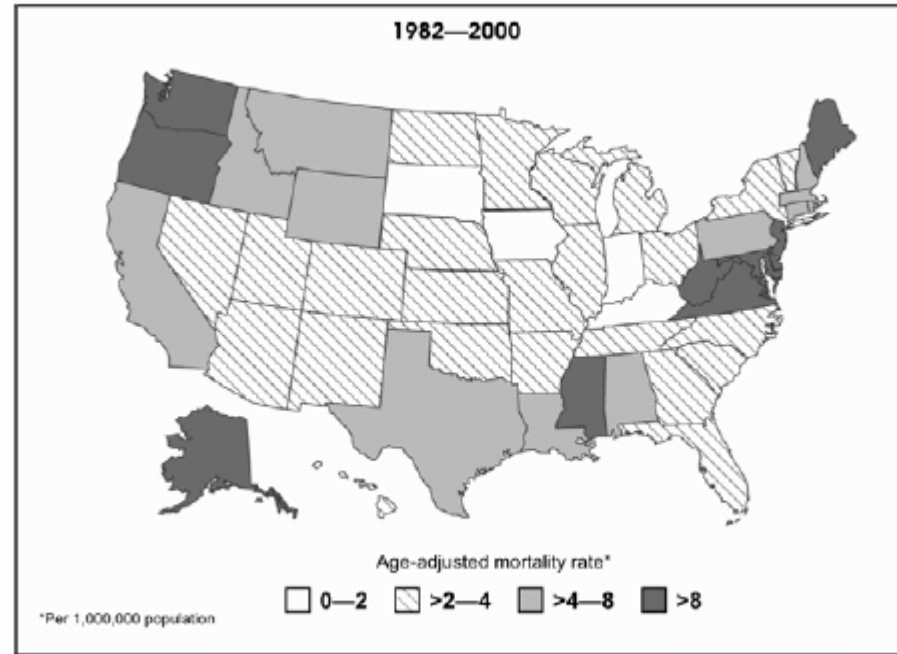
- Excellent examples of the use of maps to display public health data are available in these selected publications:
- Atlas of United States Mortality, U. S. Department of Health and Human Services, Centers for Disease Control and Prevention, Hyattsville, MD, 1996 (DHHS Publication No. (PHS) 97-1015)
- Atlas of AIDS. Matthew Smallman-Raynor, Andrew Cliff, and Peter Haggett. Blackwell Publishers, Oxford, UK, 1992
- An Historical Geography of a Major Human Viral Disease: From Global Expansion to Local Retreat, 1840-1990. Andrew Cliff, Peter Haggett, Matthew Smallman-Raynor. Blackwell Publishers, Oxford, UK, 1988



Other Data Displays

- An area map, also called a choropleth map, can be used to show:
 - rates of disease or other health conditions in different areas by using different shades or colors (Figure 4.35)
 - When choosing shades or colors for each category, ensure that the intensity of shade or color reflects increasing disease burden.
 - In Figure 4.35, as mortality rates increase, the shading becomes darker.

Figure 4.35 Mortality Rates (per 100,000) for Asbestosis by State—United States, 1982–2000



Source: Centers for Disease Control and Prevention. Changing patterns of pneumoconiosis mortality—United States, 1968–2000. *MMWR* 2004;53:627–31.

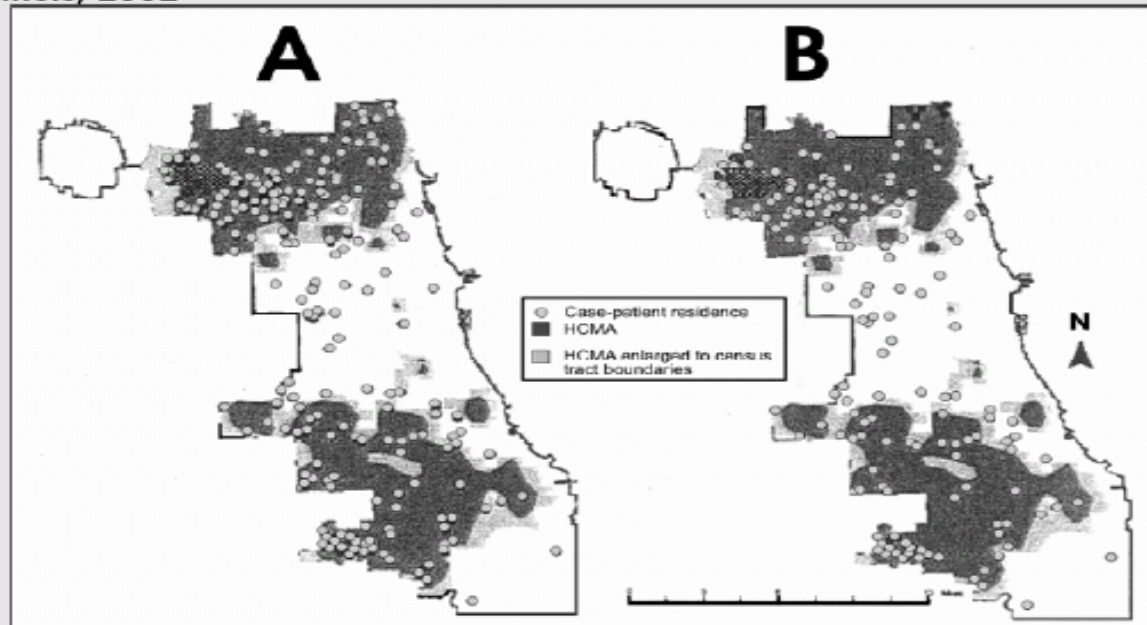


Other Data Displays

More About Geographic Information Systems (GIS)

A geographic information system is a computer system for the input, editing, storage, retrieval, analysis, synthesis, and output of location-based information.²² In public health, GIS may use geographic distribution of cases or risk factors, health service availability or utilization, presence of insect vectors, environmental factors, and other location-based variables. GIS can be particularly effective when layers of information or different types of information about place are combined to identify or clarify geographic relationships. For example, in Figure 4.36, human cases of West Nile virus are shown as dots superimposed over areas of high crow mortality within the Chicago city limits.

Figure 4.36 High Crow-mortality Areas (HCMAs) and Reported Residences of A) West Nile Virus (WNV)-infected Case-patients, or B) WNV Meningoencephalitis Case-patients (WNV Fever Cases Excluded)—Chicago, Illinois, 2002



Source: Watson JT, Jones RC, Gibbs K, Paul W. Dead crow reports and location of human West Nile virus cases, Chicago, 2002. *Emerg Infect Dis* 2004;10:938–40.



College of Nursing



Using Computer Technology

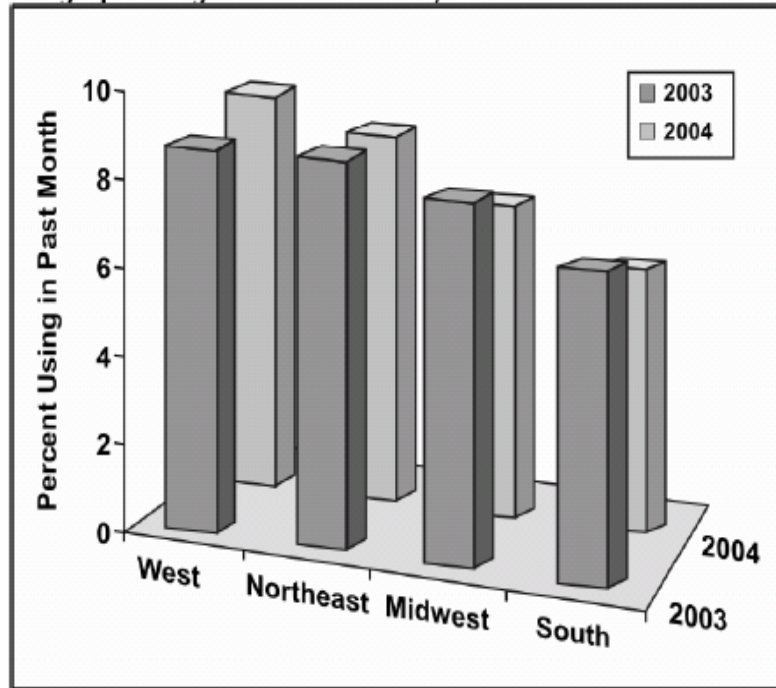
- Many computer software packages are available to create tables and graphs
- Most of these packages are quite useful, particularly in allowing the user to redraw a graph with only a few keystrokes
- With these packages, you can now quickly and easily draw a number of graphs of different types and see for yourself which one best illustrates the point you wish to make when you present your data
- On the other hand, these packages tend to have default values that differ from standard epidemiologic practice
- Do not let the software package dictate the appearance of the graph
- Remember the adage:
 - **let the computer do the work, but you still must do the thinking!**
- Keep in mind the primary purpose of the graph — to communicate information to others
- For example, many packages can draw bar charts and pie charts that appear three-dimensional
- Will a three-dimensional chart communicate the information better than a two-dimensional one?



Using Computer Technology

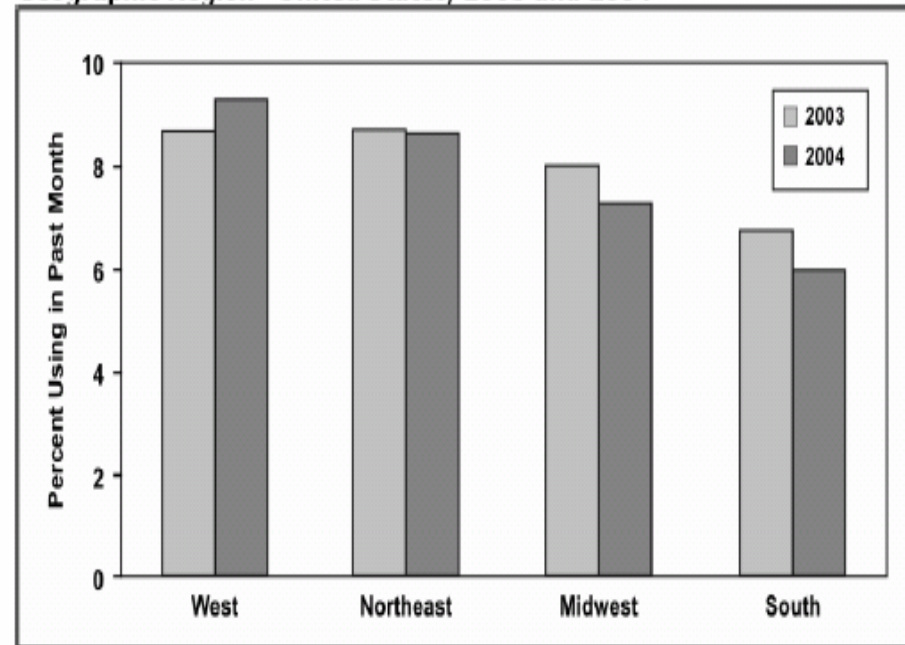
- Compare and contrast the effectiveness of Figure 4.37a and 4.37b in communicating information:

Figure 4.37a Past Month Marijuana Use Among Youths Aged 12–17, by Geographic Region—United States, 2003 and 2004



Data Source: Substance Abuse and Mental Health Services Administration. (2005). Results from the 2004 National Survey on Drug Use and Health: National Findings (Office of Applied Studies, NSDUH Series H-28, DHHS Publication No. SMA 05-4062). Rockville, MD.

Figure 4.37b Past Month Marijuana Use Among Youths Aged 12–17, by Geographic Region—United States, 2003 and 2004



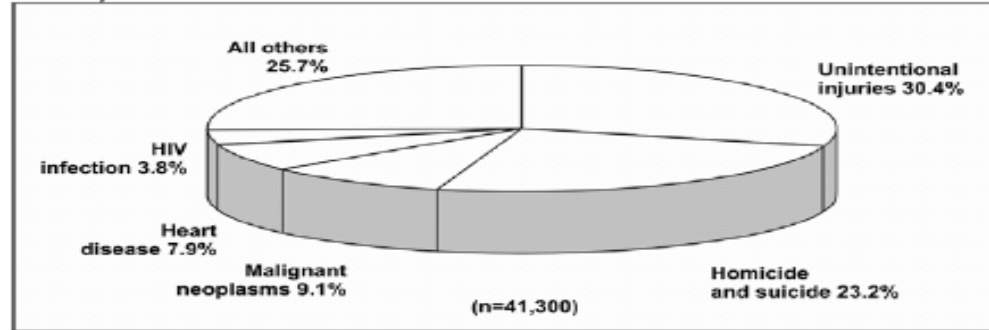
Data Source: Substance Abuse and Mental Health Services Administration. (2005). Results from the 2004 National Survey on Drug Use and Health: National Findings (Office of Applied Studies, NSDUH Series H-28, DHHS Publication No. SMA 05-4062). Rockville, MD.



Using Computer Technology

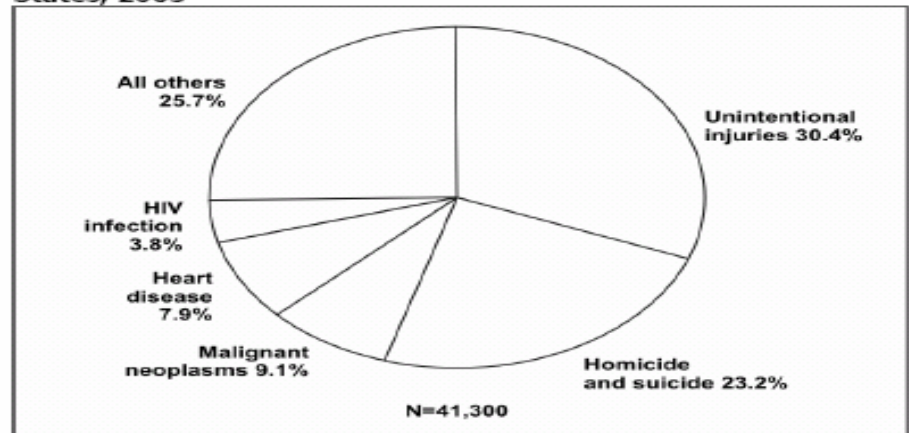
- Similarly, does the three-dimensional pie chart in Figure 4.38a provide any more information than the two-dimensional chart in Figure 4.38b?
- The relative sizes of the components may be difficult to judge because of the tilting in the three-dimensional version.
- From Figure 4.38a, can you tell whether the wedge for heart disease is larger, smaller, or about the same as the wedge for malignant neoplasms?
- Now look at Figure 4.38b. The wedge for malignant neoplasms is larger.
- Remember that communicating the names and relative sizes of the components (wedges) is the primary purpose of a pie chart.
- Keep the number of dimensions as small as possible to clearly convey the important points, and avoid using gimmicks that do not add information.

Figure 4.38a Leading Causes of Death in 25–34 Year Olds—United States, 2003



Data Source: Web-based Injury Statistics Query and Reporting System (WISQARS) [online database] Atlanta; National Center for Injury Prevention and Control. [cited 2006 Feb 15]. Available from: <http://www.cdc.gov/ncipc/wisqars>.

Figure 4.38b Leading Causes of Death in 25–34 Year Olds—United States, 2003



Data Source: Web-based Injury Statistics Query and Reporting System (WISQARS) [online database] Atlanta; National Center for Injury Prevention and Control. [cited 2006 Feb 15]. Available from: <http://www.cdc.gov/ncipc/wisqars>.



Summary

- Much work has been done on other graphical methods of presentation
- One of the more creative is face plots
- Originally developed by Chernoff, these give a way to display n variables on a two-dimensional surface
- For instance, suppose you have several variables (x , y , z , etc.) that you have collected on each of n people, and for purposes of this illustration, suppose each variable can have one of 10 possible values
- We can let x be eyebrow slant, y be eye size, z be nose length, etc
- The figures below show faces produced using 10 characteristics – head eccentricity, eye size, eye spacing, eye eccentricity, pupil size, eyebrow slant, nose size, mouth shape, mouth size, and mouth opening) – each assigned one of 10 possible values

Figure 4.39 Example of Face Plot Faces Produced Using 10 Characteristics



Source: Weisstein, Eric W. Chernoff Face. From MathWorld--A Wolfram Web Resource.
<http://mathworld.wolfram.com/ChernoffFace.html>.



Summary

- To convey the messages of epidemiologic findings, you must first select the best illustration method
- Tables are commonly used to display numbers, rates, proportions, and cumulative percents
- Because tables are intended to communicate information, most tables should have no more than two variables and no more than eight categories (class intervals) for any variable
- Printed tables should be properly titled, labeled, and referenced; that is, they should be able to stand alone if separated from the text
- Tables can be used with either nominal or continuous ordinal data
- Nominal variables such as sex and state of residence have obvious categories
- For continuous variables that do not have obvious categories, class intervals must be created. For some diseases, standard class intervals for age have been adopted
- Otherwise a variety of methods are available for establishing reasonable class intervals
- These include class intervals with an equal number of people or observations in each; class intervals with a constant width; and class intervals based on the mean and standard deviation
- Graphs can visually communicate data rapidly
- Arithmetic-scale line graphs have traditionally been used to show trends in disease rates over time
- Semilogarithmic-scale line graphs are preferred when the disease rates vary over two or more orders of magnitude



Summary

- Histograms and frequency polygons are used to display frequency distributions
- A special type of histogram known as an epidemic curve shows the number of cases by time of onset of illness or time of diagnosis during an epidemic period
- The cases may be represented by squares that are stacked to form the columns of the histogram; the squares may be shaded to distinguish important characteristics of cases, such as fatal outcome
- Simple bar charts and pie charts are used to display the frequency distribution of a single variable
- Grouped and stacked bar charts can display two or even three variables
- Spot maps pinpoint the location of each case or event
- An area map uses shading or coloring to show different levels of disease numbers or rates in different areas
- The final pages of this lesson provide guidance in the selection of illustration methods and construction of tables and graphs
- When using each of these methods, it is important to remember their purpose: to summarize and to communicate
- Even the best method must be constructed properly or the message will be lost. Glitzy and colorful are not necessarily better; sometimes less is more!

