7
Data Presentation

Evelyn Shambaugh

7.1 Selecting, assembling and presenting data

The first step in preparing a statistical report is to define the problem. What information does the user want? What information is available in the registry? The objectives and scope of the report must be defined at the outset. The request may be answered from the routine data collected in the registry. On the other hand, it may require the collection of additional data for a special study. If it is an annual report, define what types of information are to be included from the available information, such as years of diagnosis, tumour sites, age and sex distributions, and treatment and stage distributions.

Once the objectives of the report and the items of information are clearly identified, determine the cases to be included. For a count of all cancers seen at a hospital for a given year, include both alive and dead patients, including cases identified at autopsy only. For a count of all cancers in a population-based registry, count all residents of that population diagnosed with cancer during the period under study from whatever sources, including cases diagnosed on death certificate only. If planning a survival analysis, exclude those patients with cancers identified at autopsy or identified by death certificate only. For example, the population may be all patients under 15 years of age who had acute lymphocytic leukaemia diagnosed between 1 January 1990 through 31 December 1991.

For some studies it is desirable to select a random sample of cases. The most common aid to selecting a random sample is a table of random numbers (see any statistical textbook). A popular practical equivalent of random selection is systematic selection, i.e., taking every fifth patient on a list. When a systematic selection is used, make sure that the number of items between successive selections does not correspond to some recurring cycle of cases.

The presentation of the data will depend on the purpose of the study. If the purpose requires only counts or cross tabulations of the characteristics in the patient record, the data may be presented in the form of a table or a graph.

Tables versus graphs: advantages and disadvantages:

Ever since records were first kept, there has been the problem of interpretation of numerical data. Statistical tables were a big step forward for summarizing data, but graphs went even further by presenting data in visual form.

Too often data are presented in an awkward or confusing format. By following certain simple rules described in section 7.2 for tables and in section 7.3 for graphs, it should be possible to present the data with maximum effectiveness.

The question of whether to present data in the form of a table or a graph depends on the purpose. Tables have the following advantages over graphs:

- more information may be presented;
- exact values can be read from a table to retain precision;
- less work and less cost are required in the preparation;
- flexibility is maintained without distortion of data.

On the other hand, graphs have the advantage of:

- attracting attention more readily;
- showing trends or comparisons more vividly;
- being a simple and efficient method of showing observations in the past, present and future;
- providing results that are more easily remembered.

In short, one picture (graph) is worth a thousand words. However, in some studies it may be advantageous to give both the detailed table and a simple summary graph. Graphs can bring out hidden facts and relationships which stimulate analytical thinking, but tables provide the supportive details. Together they present a better balanced understanding of a study.

7.2 Preparing tables

A table is an orderly classification of facts arranged in vertical columns and horizontal rows which groups related numbers into classes. Each variable such as sex, race, age, treatment and stage of disease has a system of classification. Sex has two classes while age can have any number depending on age groupings.

Title: The title must tell as simply as possible what is in the table. It should answer the questions:

- **Who?**
  White females with breast cancer,black males with lung cancer.
- **What** are the data?
  Counts, percentage distributions, rates.
- **Where** are the data from?
  One hospital, or the entire population covered by your registry.
- **When** ?
  A particular year, time period.

Boxhead: The boxhead contains the captions or column headings. The heading of each column should contain as few words as possible, yet explain exactly what the data in the columns represent.

Stub: The row captions are known as the stub. Items in the stub should be grouped to facilitate interpretation of the data. For example, group ages into 10-year age-groups.

Footnotes: Anything in a table which cannot be understood by the reader from the title, boxhead, or stub should be explained by footnotes. The footnotes contain information on missing numbers, preliminary or revised numbers, or explanations for any unusual numbers. Definitions, abbreviations, and/or qualifications for captions or cell names, and all pertinent information should be footnoted. A footnote usually applies to a specific cell(s) within the table and a symbol, such as '*' or '#', is used to key the cell to the footnote. If several footnotes are required, it is better to use small letters rather than numbers. Footnote numbers might be confused with the numbers within the table.

Source: If data from a source outside the registry are used, the exact reference to the source should be given. For example, if comparing the registry's patient survival with the survival data from the USA SEER Program, reference the SEER data, e.g.,
Source: Axtell, L.M., Asire, A.J. & Myers, M.H.: Cancer Patient Survival, Report Number 5. DHEW Publication No. (NIH) 77-992
Denoting the source lends authenticity to the numbers and enables the reader to locate the source if further information is desired.

Tables usually are arranged so the length exceeds the width; it is generally better to use the longer wording in the stub. Important numbers to be compared should be placed in adjoining columns or rows. Time series are listed in chronological order, beginning usually with the earliest time period. Traditional listings are usually listed in that order, e.g., anatomical sites are listed in ICD-O order. For emphasis, the order may be changed to another order, such as the relative frequency of occurrence. Classifications of size are usually listed from smallest to largest.

Cross-classified tables must always account completely for the data being classified. For this reason, unimportant classes are put in a composite class labelled 'Other'. The 'Other' categories are placed to the right of the bottom of the rows or columns, respectively.

Many analytical tables contain both numbers of cases and percentage distributions. Numbers provide information on magnitude; percentage data facilitate comparisons.

Check the table to be sure that:

- it is a logical unit (separate problems call for separate tables);
- it is self-explanatory (can it stand alone if removed from its context?);
- all sources and units are specified.

Construction of tables:

In table construction, good judgement is more important than blind following of rules. Present the data to answer a definite question or phase of a question. The simplest table is a one-way classification in which one variable, for example, sex, is presented either in terms of numbers of cases or a percentage distribution or both. Table 1c below has both.

If a classification is desired according to two characteristics simultaneously, they are cross-classified in a 'two-way' table. One classification will appear horizontally and the other vertically as shown in Table 2a below:

Table 1a. Form for a ONE-way classification: Numbers of cases:

Distribution by sex of children with acute leukaemia Community hospital, 1991		
	Sex	Number of cases
	Total	50
Stub—	Male	30
	Female	20

— Title
— Box-

Table 1b. Form for a ONE-way classification: Percentage distribution

Distribution by sex of children with acute leukaemia Community hospital, 1991	
Sex	Percent
Total	100
Male	60
Female	40

Table 1c. Form for a ONE-way classification with both numbers of cases and a percentage distribution

Distribution by sex of children with acute leukaemia Community hospital, 1991		
Sex	Number of cases	Percent
Total	50	100
Male	30	60
Female	20	40

Table 2a. Form for a two-way classification

Age distribution of lung cancer patients by sex, 1980-1993			
Age	Total	Male	Female
All ages ——————————Row——————————			
<45	Cell	C	
45-54	Cell	o	
55-64	Cell	l	
65-74	Cell	u	
75 and over	Cell	m	
		n	

— Title
— Boxhead
Stub —

Footnote
Source

In this case sex is listed horizontally in the BOXHEAD and age is listed vertically in the STUB. The individual entries which are classified according to the row and column are called CELLS. Often it is desirable to include summary information in a table. For example, in Table 2a the ratio of males to females might be pertinent to the discussion and could be added to the caption entries as fol lows:

Table 2b. Age distribution of lung cancer patients by sex, 1980-1993

Age	Total	Male	Female	M/F ratio
All ages				
<45				
45–54				
55–64				
65-74				
75 & over				

When three or more classifications of the data are desired, the problem becomes more difficult. This multi-dimensional relationship must be shown on a two-dimensional sheet of paper. This can be done by subdividing each class of one or both classifications. Table 3 illustrates a 'three-way' classification.

Table 3. Form for a three-way classification

Age and sex distribution of white and black lung cancer patients for cases diagnosed in 1991				
Race and age	Total	Male	Female	M/F ratio
White:				
All ages:				
Under 45				
45-54				
55-64				
65-74				
75 & over				
Black:				
All ages:				
Under 45				
45-54				
55-64				
65-74				
75 & over				

Table 4 illustrates a 'four-way' classification

Whenever further subdivision of data leads to tables which are too complex to be read easily, it is preferable to increase the number of tables.

Table 4: Form for a four-way classification

Age and sex distribution of white and black lung cancer patients for cases diagnosed 1980-86 and 1987-93								
Race and age	Total		Male		Female		M/F Ratio	
	1980–1986	1987–1993	1980–1986	1987–1993	1980–1986	1987–1993	1980–1986	1987–1993
White:								
All ages:								
Under 45								
45-54								
55-64								
65-74								
75 & over								
Black:								
All ages:								
Under 45								
45-54								
55-64								
65-74								
75 & over								

Mutually exclusive categories:

In defining classifications, the classes should be mutually exclusive so that the limits do not overlap

In general, it is advisable to divide detailed data into a reasonable number of arbitrary classes. If the number of classes is too few, important characteristics may be concealed. If there are too many classes, there may be a confusing variation of frequencies and some classes may contain no values. A proper balance must be struck so that one neither overlooks a relationship nor creates the effect of one by chance.

The location of classes must be stated precisely to avoid ambiguity. Any of several methods of designating classes may be used depending in part on the nature of the data. The table below demonstrates four methods of designating classes of tumour size for breast cancer patients. Of the four methods, the one in Column A is the poorest, for it is ambiguous: it is not clear where a tumour of 2 cm should be counted. Column B clearly states the midpoint of each interval, but it is not clear what the limits of each class are. The class limits in Column C are appropriate for discrete data, that is, data that are recorded as whole numbers. The class limits in Column D are the most suitable for continuous data when some values could include decimal values.

Table 5: Examples of classification

Classification for tumour size (in cm)			
A	B	C	D
0–2	1	0-1	Less than 2.0
2–4	3	2-3	2.0-3.9
4–6	5	4-5	4.0-5.9
6–8	7	6-7	6.0-7.9
8–10	9	8-9	8.0-9.9
10 & over	11	10 &over	10 & over

It is highly desirable that all class intervals have the same width because equal intervals are easier to interpret. For some types of data, however, unequal intervals must be used. For example, in the classification above for the relationship between tumour size and prognosis, it is more important to have smaller interval sizes for small tumours and larger interval sizes for the larger tumours, such as:

Under 0.5 cm
0.5-0.9
1.0-1.9
2.0-2.9
3.0-3.9
4.0-4.9
5.0-9.9
10.0 & over

It often happens that what is needed is not so much the absolute number of patients which fall in each class but rather the relative number. Then, the total number of patients is 100% and the percent is the number of patients in each class divided by the total. This is known as a percentage distribution. It is illustrated in Table 6.

Table 6. Example of percentage distribution

Age distribution of acute lymphocytic leukaemia Patients: 1985–94		
Age	Male	Female
Number of cases	617	449
	Percent	
Total	100	100
0–9	55.2	53.4
10–19	14.5	13.1
20–29	4.9	4.0
30–39	4.4	4.8
40–49	2.0	3.6
50–59	5.1	5.2
60–69	5.4	7.0
70–79	6.0	6.3
80+	2.5	2.6

Source: Cancer Patient Survival, Report No. 5, 1977

7.3 Graphs: form and construction

A graph is the best medium for presenting data for quick visualization of relationships between various factors. Graphs effectively emphasize the main points in an analysis and clarify relationships which might otherwise remain elusive.

Form of graphs:

There are many types of graphs: picture graphs, maps using dots or shading, pie charts, bar graphs, and line graphs plotted

on a variety of scales. The type of graph used will depend on the data. There are four different kinds of data; their scales of measurement vary from the simple nominal to the more complex ratio scale of measurement:

(1) **Nominal** (named) data require unordered categories for such data as sex, race and blood groups, but ordered categories for data such as stage of disease where there is a logical order based on disease prognosis. Either numbers of cases or a percentage distribution is presented.

(2) **Ordinal** (ordered) data, as the name implies, require that the categories be ordered in a definite way. Examples are rating scales and performance status scales. For rating scales the data are ordered from least satisfactory to most satisfactory. For performance status scales, the physical ability of the patient is ordered from normal to decreasing ability from 100 to 0 in the Karnofsky scale.

(3) **Interval** data require equal units along the scale as, for example, temperature, but the zero point may be a different value depending on the scale. If the temperature is 50 degrees, the next question is:

'What is the scale - centigrade or Fahrenheit?'

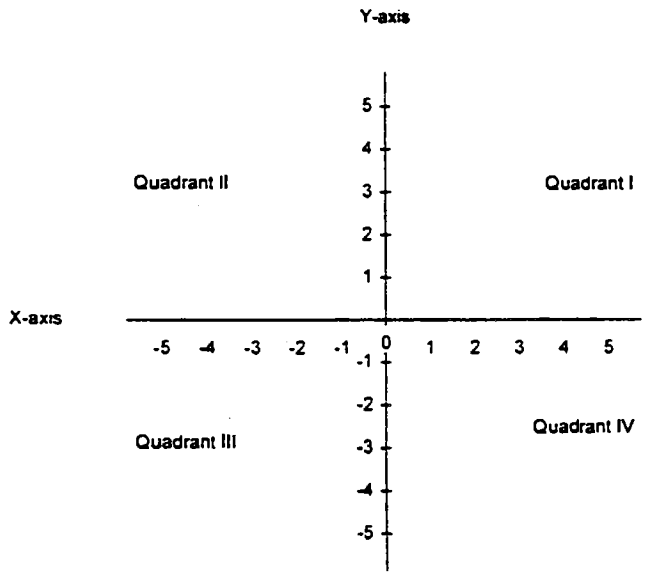
(4) Ratio data must have numerical values for computing ratios. A ratio scale has not only *equal units*, but also a *zero point* which is absolute in the sense that zero represents a complete lack of the value being measured. It allows comparisons such as, 'If A is age 20 and B is age 40, then B is twice as old as A.'

Constructing graphs:

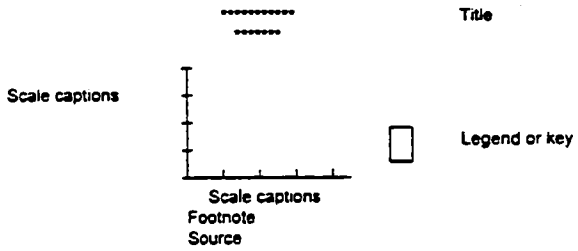
The basic form of a graph is derived by plotting numbers in relation to two axes. A scale is arranged in both directions from a zero point at the intersection of the axes. A comparison of data is shown by variation in the slopes of lines, the heights of bars, or the sizes of areas. Most graphs use positive values only, thus only the upper right-hand part of the grid (Quandrant I) is usually known. 'Tic' marks are used to indicate the grid lines in the example below. The axes are marked off in equal units and may be extended as far as necessary in any direction.

The scale of values for the X-axis is placed along the bottom of the graph from the lowest value on the left to the highest value on the right. The scale of values for the Y-axis is placed at the left of the graph from the lowest value at the bottom to the highest value at the top of the graph.

Choose the largest scale that will fit the paper and allow the graph to be centrally located on the page. Clearly label each scale and indicate the units.



Graph 1



Graph 2

TITLE: The title must tell as simply as possible what the graph shows. It should answer the same questions as the title for a table.

- **Who?**
  White females with breast cancer; black males with lung cancer.
- **What** are the data?
  Counts; percentage distributions; rates.
- **Where** are the data from?
  One hospital; the entire state.
- **When?**
  A particular year; a time period.

Legend or key: When several variables are included on the same graph, it is necessary to identify each by using a key or legend. The legend should be placed in a clear space on the face of the graph and each line identified on the graph as in the example below.

White males .....
Black males ---------
White females ...-...-...-...-
Black females —.—.—.—.

Scale captions: Scale captions are placed on both axes to identify the scale values clearly. It is essential that both the subject and the units used be identified. The caption for the horizontal scale is generally centred under the X-axis. The caption for the vertical axis is placed either at the top left of the Y-axis or along the Y-axis, whichever is the easier to read.

Footnotes: If the title, scale labels, and legend cannot explain everything in the graph, then footnotes should be used as in tables.

Source: The exact reference to the source should be given just as for tables. For example, when comparing a registry's patient load with the incidence data from the USA SEER Program, reference the SEER data as follows:

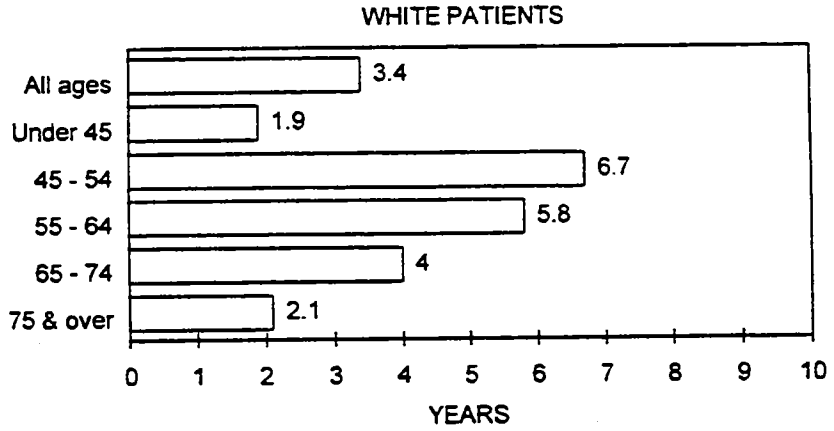
Source: Young, J.L., Percy, C.L. & Asire, A.J. Surveillance, Epidemiology, and End Results: Incidence and Mortality Data, 1973-77, NCI Monograph 57, NIH Publication No. 81-233

Remember that a graph is useless unless it is read. It has to be interesting and attractive if it is to be read.It may be advisable to include a table with each graph so that the reader may see the actual numbers on which the graph is based.

7.3.1 Bar graphs

Bar graphs are commonly used for frequencies, proportions and percentages of nominal and ordinal data. They are easy toconstruct and can be readily interpreted. Bars emphasize individual amounts in contrast to lines which emphasize general trends. Bars are effective for showing the component parts of a whole and for making comparisons between groups such as breast cancer patients by race and stage of disease. The bars may be either horizontal or vertical and may be filled in with stripes, cross-hatching, or shading to make them stand out. Because the bars represent magnitudes by their lengths, a zero line must be shown and the arithmetic scale must be used. In a simplebar graph, the spaces between the bars are usually about half of the width of each bar.

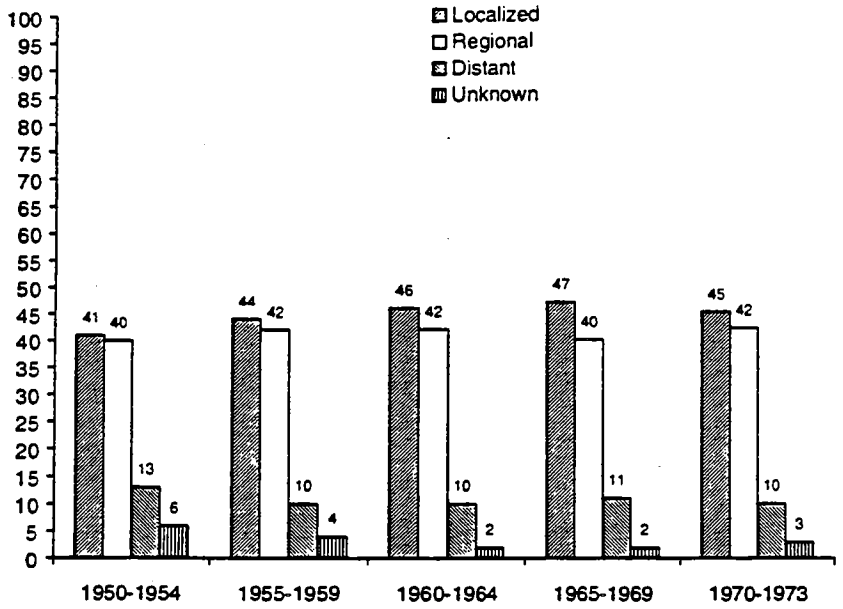
Figure 1. Observed median survival time by age for white males with cancer of the prostate diagnosed 1960-73



Source: End Results in Cancer. Report No. 5, 1977

The width of the bar for 'all ages' is the same as that for each of the specific age-groups in this graph. The values for each age-group are independent of the values for the other age-groups. In this graph, the values have been written in the bars for ease in reading.Bars may be grouped without spaces between bars to show subdivisions within several groups where the subdivisions are the same for each group. Each bar within a group is distinguished by a different texture and the same shading is used for each group. This is illustrated in Figure 2 in which trends among white female breast cancer patients are illustrated for stage of disease, i.e., localized, regional, distant and unknown, within each of five time periods.The major groups, which are separated by a space, are the five time periods 1950-54, 1955-59, 1960-64, 1965-69 and 1970-73. Stage of disease is the subdivision within each time period with no spaces between localized, regional, distant and unknown. This graph clearly indicates the percent of cases in each stage category for each time period.

Figure 2: Trends among white female breast cancer patients by stage of disease, 1950-1973, SEER Program



Example for trend data
Percent treated by

Year of diagnosis	Surgery only	Radiation only
1950–54	30	23
1955–59	22	40
1960–64	19	42
1965–70	12	54
1970–73	14	57

The graph for the trend data in the preceding table is illustrated below in Figure 3. This bar graph uses vertical bars to contrast differences in treatment over time. Again the bars can be of equal width since the percentage for each time period is independent of the percentages for the other time periods. The bars representing the two different types of treatment are of different textures so they can be readily distinguished.

7.3.2 Histogram

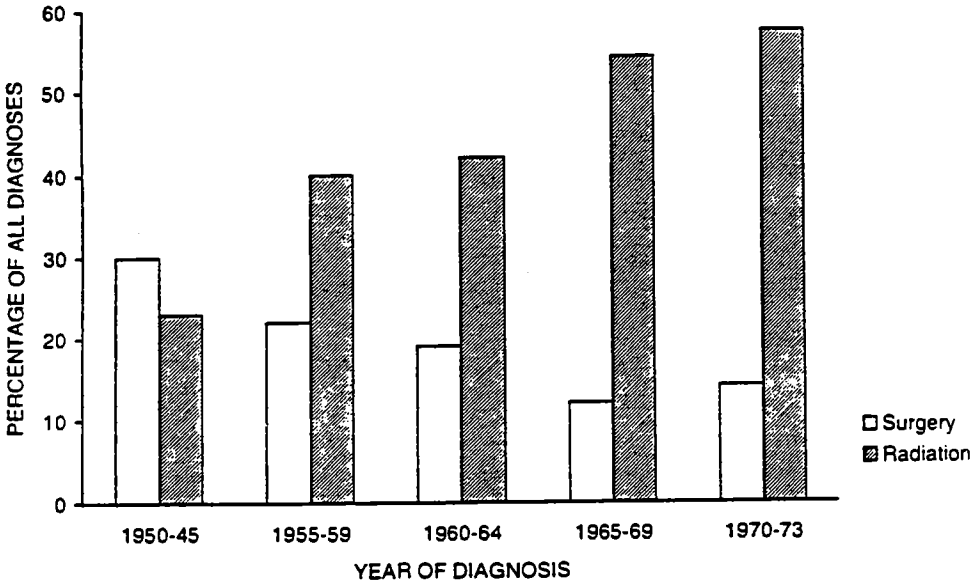
A histogram is usually the best type of graph to use when only one distribution is being represented. It is a distribution expressed either in terms of numbers or percentages. A histogram consists of a series of columns each having as its base one class interval and as its height the number of cases as the distribution in that class. In this type of graph there are no spaces between the columns. The sum of the heights of the columns represents the total number or 100% of the cases. A histogram should be

used only when the data on the horizontal or X-axis are measured on an interval or ratio scale.

In order to avoid having the figure too flat or too steep, it is usually good to arrange the scales so the width of the histogram itself is about one and two-thirds times its height (a ratio of 3:5). A column is centred around the midpoint of the class interval (see Figure 4 below).

Histograms are easy to understand. They are useful for showing differences in age distributions, as in Figure 5 which indicates that brain cancers occur with the highest frequency between 50-59 years of age. This distribution is bimodal; that is, it has two age-groups which have a higher frequency than the adjacent age-groups. The age-group with the highest frequency (50-59) is called the MODE. To obtain the percentage of cases diagnosed between two ages, e.g., between 40-79, it is necessary to add the values of the bars, including the values between these ages, i.e., 19% + 25% + 18% + 5% = 67%.

Figure 3. Trends in percentage of white oesophageal cancer patients treated by surgery only and by radiation only, 1950–1973



Source: Cancer Patient Survival, Report No. 5, 1977

Figure 4: Cases of rash illness, elementary school, Sample City, 22-23 March, 1990

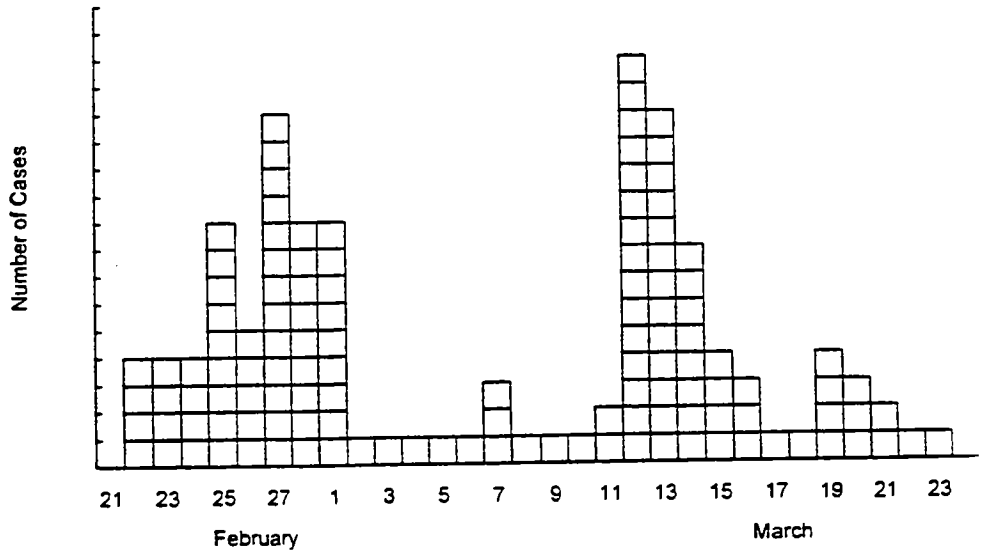
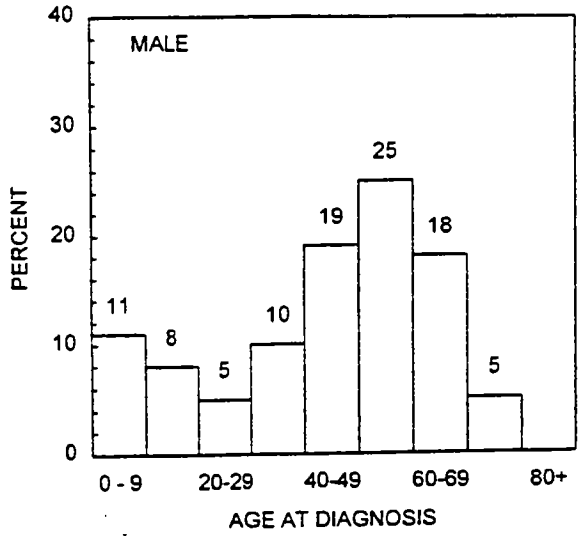


Figure 5: Age distribution of male brain cancer patients diagnosed 1955-64



Source: End Results in Cancer, Report No. 3, 1968

Width of intervals: In working with histograms it is a good idea, if possible, to use intervals of the same width, e.g., all five-pound weight intervals or all 10-year age-groups. If the intervals are not equal, but have varied interval sizes, the frequency value on the vertical scale should be adjusted for differences in interval width. It is area which represents frequency, not height of a column. Each column MUST represent the same size group since the total area represents relative differences in the frequency distribution. If all the intervals were for five years except one that was 10 years, the 10-year interval would have to be converted by dividing its percentage in half.

7.3.3 Frequency polygon

A frequency polygon may be used as an alternative to the histogram just described. Simply join the midpoints at the top of each bar in the histogram as shown in the figure below. The advantage of the frequency polygon over the histogram is that several frequency polygons can easily be plotted on the same graph for purposes of comparison. It is also easy to interpret (see Figure 6 overleaf).

As with all work with graphs, two axes (X and Y) are used. In constructing the graph of the frequency polygon, the X-axis is longer than the Y-axis; a ratio of 3 to 2 or 4 to 3 will result in a good graph. The frequency values are always placed on the Y-axis and the scores on the X-axis. Frequency values are plotted at the midpoint of each class interval as a rule.

If the numbers in different groups vary widely, it may be impractical to put them on the same graph, e.g., the distribution of patient ages in one hospital versus another. In that case convert each frequency into a percentage and plot it. The percentages are plotted in the same manner as the numbers. When comparing items of data, each line should be constructed using different types of lines of different colours. A legend should be included identifying the different lines.

7.3.4. Cumulative frequency polygon

As a further step in the analysis of the frequency distribution, the value in a series may be cumulated. The cumulative frequency for any interval is the total of the frequencies for that interval and for all lower intervals. The cumulative relative frequency is the cumulative frequency divided by the total number of observations. It is used to find percentiles of a distribution, i.e., the percent or proportion of observations less than a given value. Always plot the cumulative frequency or cumulative relative frequency against the true upper limit of each interval.

Figure 6: Number of cases of influenza-like illness by week, Sample City, 1970

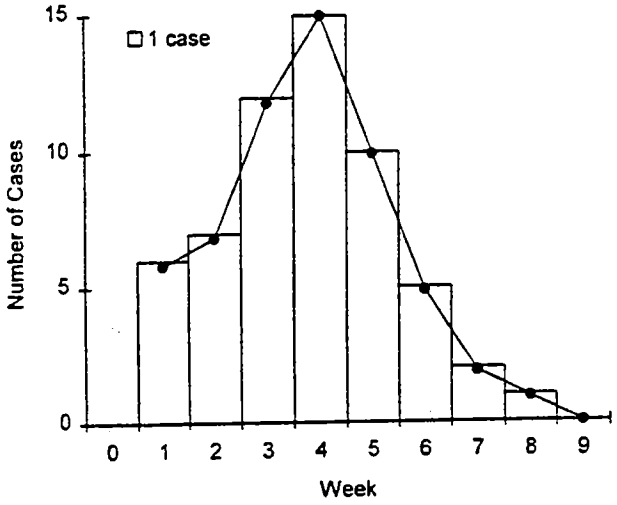
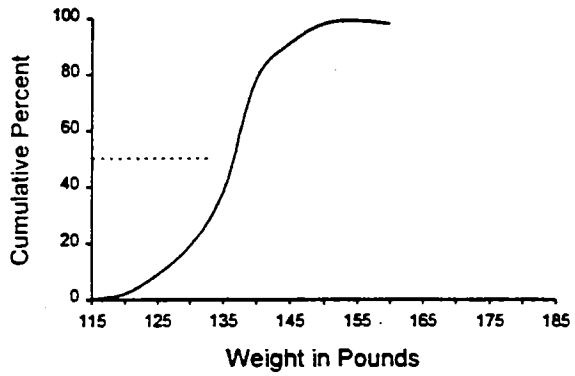


Figure 7 Female patients in a weight study at memorial hospital, 1990



In the example in Figure 7 note that, at the midpoint of the cumulative percentage curve (50th percentile) where 50% of the cases lie above the midpoint and 50% below the midpoint, the median weight is 135 lbs. The 10th percentile is seen to equal 117.5 pounds.

7.3.5. Component band graph

The component band graph is used to compare the various components of independent groups. Like a bar graph, it analyses nominal and ordinal data, but instead of bars it has bands. It can be either vertical or horizontal, whichever is easier to read.

Figure 8 illustrates how this type of graph can be used. In constructing it the length of each band is the same and each band represents 100% of the cases of a particular group.

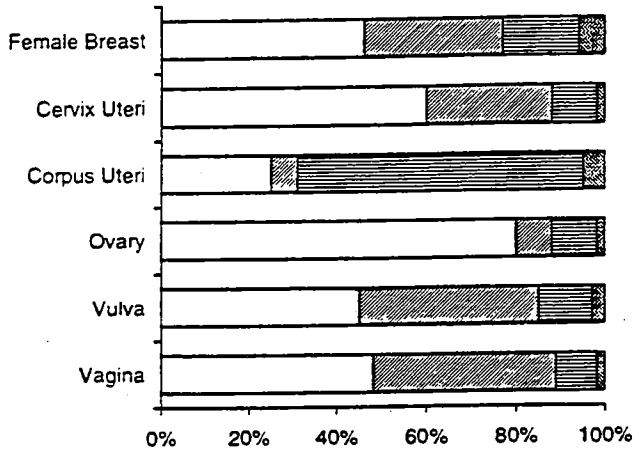
The segments in each band represent the different components of the total and are arranged in the same order in all groups. The three summary stages are arranged in the conventional order, and the length of each segment is determined according to the scale shown on the X-axis. For this graph 'localized' stage is the most important stage and, therefore, is shaded the darkest for greater emphasis. The space between the bands is usually one-half the width of the bands.

The graph for the component band graph is illustrated in Figure 8 overleaf.

Sample data for component band graph

	Localized		Regional		Distant		Unknown	
Female breast	48	+	41	+	9		2	= 100
Cervix	45	+	40	+	12	+	3	= 100
Corpus	80	+	8	+	10	+	2	= 100
Ovary	25	+	6	+	64	+	5	= 100
Vulva	60	+	28	+	10	+	2	= 100
Vagina	46	+	31	+	17	+	6	= 100

Figure 8 Stage distribution for female reproductive organs for patients diagnosed 1970-73



Source: Cancer Patient Survival, Report No. 5, 1977

7.3.6 Line graphs

There are two kinds of line graphs, arithmetical and semilogarithmic. The arithmetical measures absolute differences. It is like an automobile odometer in that it measures 'how far'. The semilogarithmic, on the other hand, measures the rate of change. Like a speedometer, it measures 'how fast'.

(1) Arithmetical line graph

An arithmetical line graph consists of a line connecting a series of points on an arithmetical scale. It should be designed toachieve simplicity without too much information on any one graph. The selection of proper scales and complete and accurate titles and legends is important. If a graph is too long and narrow, either vertically or horizontally, it has an awkward appearance and unduly exaggerates one aspect of the data. The most attractive ratio between the width and the length as a whole is between 3 to 4 and 4 to 7.

The line graph is used when there are considerable numbers of values to be plotted. It is also used when presenting continuous data. Conventionally for a time series, the horizontal scale shows the time units from left to right, while the vertical scale measures the value of the factor being shown, e.g., percent of cases classified as localized. If the coordinate lines are used to represent the time interval, the value is plotted on the coordinate line itself. If the space between any two of the coordinates represents the time period, the value is plotted at the centre of the space allotted itself. If the space between any two of the coordinates represents the time period, the value is plotted at the centre of the space allotted itself. If the space between any two of the coordinates represents the time period, the value is plotted at the centre of the space allotted.

When a frequency distribution is graphed, the coordinate lines are used to indicate the group limits. The value is plotted at a point halfway between the two coordinates at the place indicated by theappropriate value on the vertical (Y-axis) scale.

If there are several variables on the same graph, different types of lines should be used for each of the lines in order to distinguish them. This is especially important if any of the lines cross or almost touch each other. Each of these lines must be identified in the key or legend.

White males .....
Black males - - - - - - - -
White females ...-...-...-...
Black females —.—.—.—

There are two kinds of time-trend data:

- point data which are taken at a specified instant of time; and
- period data which cover an average or total over a specified period of time, such as, a year of a five-year time interval.

The vertical ticks mark instances of times and the spaces between the ticks represent periods of time. The following examples illustrate these two kinds of plotting.

When plotting the percent of the patients surviving to the end of each interval, plot the values and then connect each point plotted by a straight line. In this example the value at diagnosis (year = 0) is understood to be 100%. All three sets of values are placed on one graph in order to compare the absolute differences in survival.

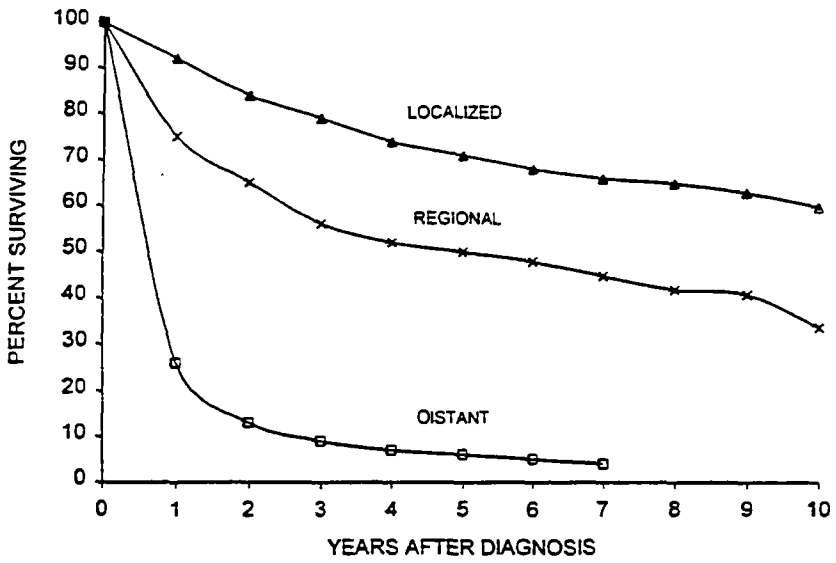
The graph for the point data in the preceding table is illustrated in Figure 9 below.

Example for point data: Relative survival rates for white patients with kidney cancer diagnosed 1960-73 by stage

Years after diagnosis	Percent surviving to end of interval		
	Localized	Regional	Distant
1	92	75	26
2	84	65	13
3	79	56	9
4	74	52	7
5	71	50	6
6	68	48	5
7	66	45	4
8	65	42	*
9	63	41	*
10	60	34	

*Too few cases to show survival rate

Figure 9 Observed survival for white kidney cancer patients 35-64 years of age, diagnosed 1960-1973



When plotting a summary statistic such as one-year and five-year survival rates for several time periods, plot the values at the mid-point value of the time periods. Example for period data.

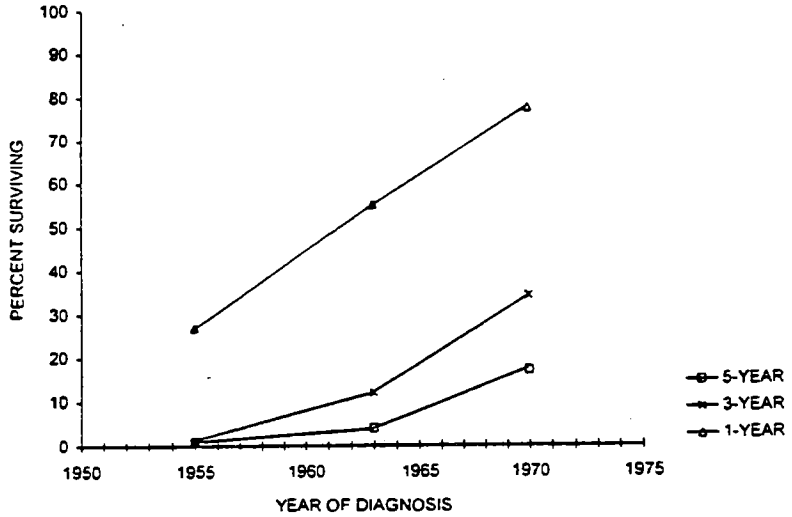
The graph for the PERIOD data in the preceding table is illustrated in Figure 10 below. Since this is plotted on an arithmetic scale, the lines represent absolute changes in the survival values.

1-year, 3-year and 5-year survival rates for children under 15 with acute lymphocytic leukaemia: 1950-59, 1960-66 and 1967-73

		Survival rate		
Year of diagnosis	Mid-point of interval	1-year	3-year	5-year
1950-59	1955	2	1	1
1960-66	1963.5	55	12	4
1967-73	1970.5	77	34	17

Source: Cancer Patient Survival, Report No. 5, 1977

Figure 10 Survival for children under 15. Acute lymphatic leukaemia



(2) Semilogarithmic line graph

Lines plotted on semilogarithmic (or semi-log) graph paper show the relative changes (rate of change) by the slope of the lines. The X-axis usually shows time and is plotted on the usual arithmetic scale. The values of the variable, usually survival rates, measured at each interval of time, are plotted on the Y-axis, which is a logarithmic scale. Logarithmic scales are scales in which the spaces between division marks are not constant but vary according to the logarithms of the numbers that are represented on the scales (instead of the numbers themselves). The steeper the line, the greater the rate of change. When values of the variable range in value between 1 and 10, a single-cycle log scale is used. For values ranging from 1 to 100, a two-cycle scale is used.

The logarithm of zero is minus infinity and, therefore, cannot be located on the scale. Each cycle begins with a power of 10, i.e., 0.1, 1, 10, 100, 1000. Distances between 2 and 4, 4 and 8, 8 and 16 (100% increases) will be the same, and distances between 2 and 3, 8 and 12, 16 and 24 (50% increases) will also be constant.

The example used in the arithmetic plot as point data would require two-cycle semilog paper since survival values range from 4% to 92% (see Figure 15 below). The values for each rate would be plotted on the vertical lines as before. The plot on the arithmetic paper

shows the absolute difference in survival for the three stage groups. For example, the absolute differences in the rates for regional and distant cases after the three years is between 41% and 45% and the lines have about the same slope. On the semilog plot the slopes for the regional and distant stages are much farther apart and the slope is much steeper for the distant cases. The steeper slope indicates that the distant cases are experiencing a higher annual mortality rate.

The slope of the line on a ratio graph indicates the percentage change between two points in time. The steeper the slope, the greater the percentage change.

A rate of change which is constant over all years of observation would plot as a straight line. In the figure for each of the stages for the first three years, the slope of the line becomes less steep indicating that the mortality rate is decreasing each year after diagnosis. From the third through the ninth year the line for regional cases is straight which indicates a constant mortality.

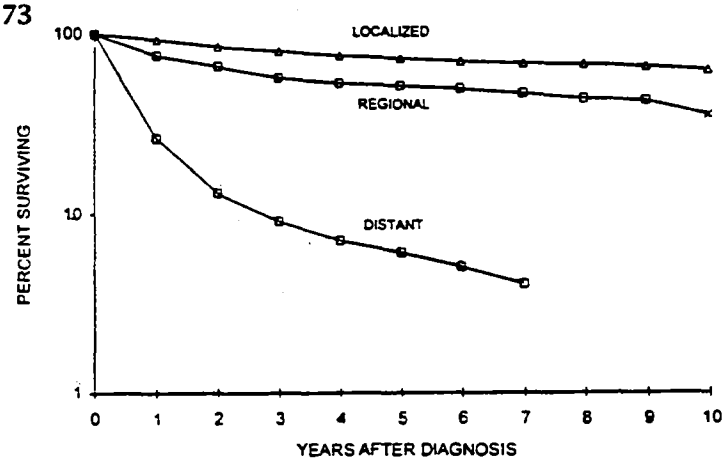
Graphs plotted on semilog scale are useful for plotting survival curves since they show rates of change. When plotting factors for which the absolute values are important, the arithmetical scale should be used.

In Figure 11, observed survival rates for kidney are shown. Compare these rates done on semilog scale with those shown in Figure 9 which are done on an arithmetic scale.

7.3.7 Geographical coordinates

A map of an area is used as a reference, and certain statistical information is superimposed upon it. Two commonly used graphs of this type are dot maps and shaded maps.

Figure 11 Observed survival for white kidney cancer patients 35-64 years of age, diagnosed 1960-1973



(1) Dot maps

Dots or coloured pins are placed in their proper locations on a map to indicate the occurrence of a particular observation at that location and, thus, give the general effect of density. Each dot represents a certain number of cases. In some areas the dot may be too close to be counted, but an impression of density can be clearly brought out. The dots may represent the number of cases for a geographical area. A better value would be the number of cases per 100 000 population. Such maps would be useful in pinpointing areas of excessive incidence which need to be investigated.

For example, influenza epidemics are plotted routinely every week by the Public Health Service in the USA. This method has been useful in pinpointing epidemics as they travel geographically.

Variations in quantities may be indicated also by varying the size, shape, and/or colour of the dot or pin. The construction of dot maps can be difficult because of the care which must be exercised in selection of the size of the dot and the quantity it is to represent. On the other hand, the pin map is flexible and quick and easy to change.

(2) Shaded maps

These maps are most often used, instead of dots, for incidence or mortality rates. In

designing a shaded map, the lightest shading should indicate the lowest rate, and the shading should increase with the darkest shading indicating the highest rate (see Figure 12).

7.3.8. Pie charts

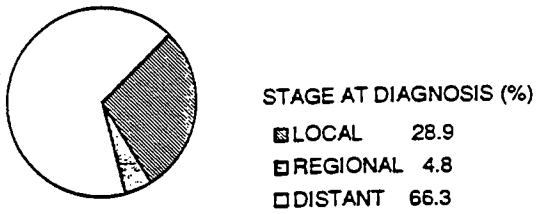
Another method of showing the component parts of the whole is to plot them on a circle (360°) called a pie chart. Each part is expressed as a percent of the total and is plotted with a protractor (1% = 3.6°) as a sector around a circle whose total circumference represents the whole of 100%.

Pie charts are constructed as follows:

- convert percents to degrees (1% = 3.6°),
- start at the 12 o'clock point, and
- plot clockwise either in order of magnitude (size) or in conventional order;
- all printing should be in a horizontal plane for ease of reading (either within the circle or outside).

Never use two pie charts to compare distributions. Pie charts are not as appropriate as are component band charts for such comparisons. A pie chart should only be used to illustrate how the whole is divided into segments; for example, stage of disease for a particular site is divided into in situ, localized, regional and distant. Extent of disease is also an example where logical or conventional order is preferred to magnitude. Start with the best prognosis and end with the worst – in situ, localized, regional and distant.

Figure 13 Stage distribution for ovarian cancer patients diagnosed 1970-7



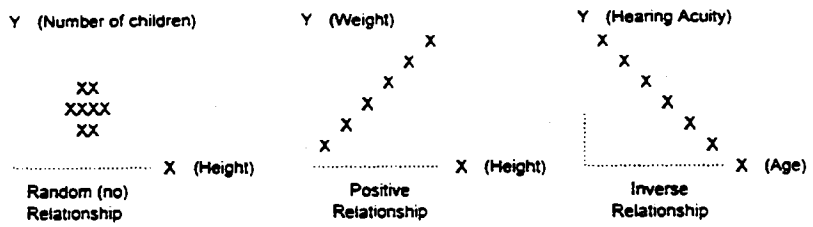
Source: Cancer Patient Survival, Report No. 5, 1977

7.3.9 Scatter diagrams

A scatter diagram is a means of presenting relationships between two variables. For example, generally one thinks of the characteristics of weight and height as being rather closely related in adults. As one increases, the other increases; the two variables are positively related. Variables can be inversely related also, for example, the older one is, the worse is one's hearing acuity.

7.3.10 Picture graphs

Identical numbers (symbols) are used to stand for a certain total number. Then the number of these numbers indicates the size of the number being illustrated. It is easy to understand, and is a fair picture as long as each figure is the same size; it is not acceptable to use different sized numbers in the same graph.



Scatter diagrams

Picture graphs

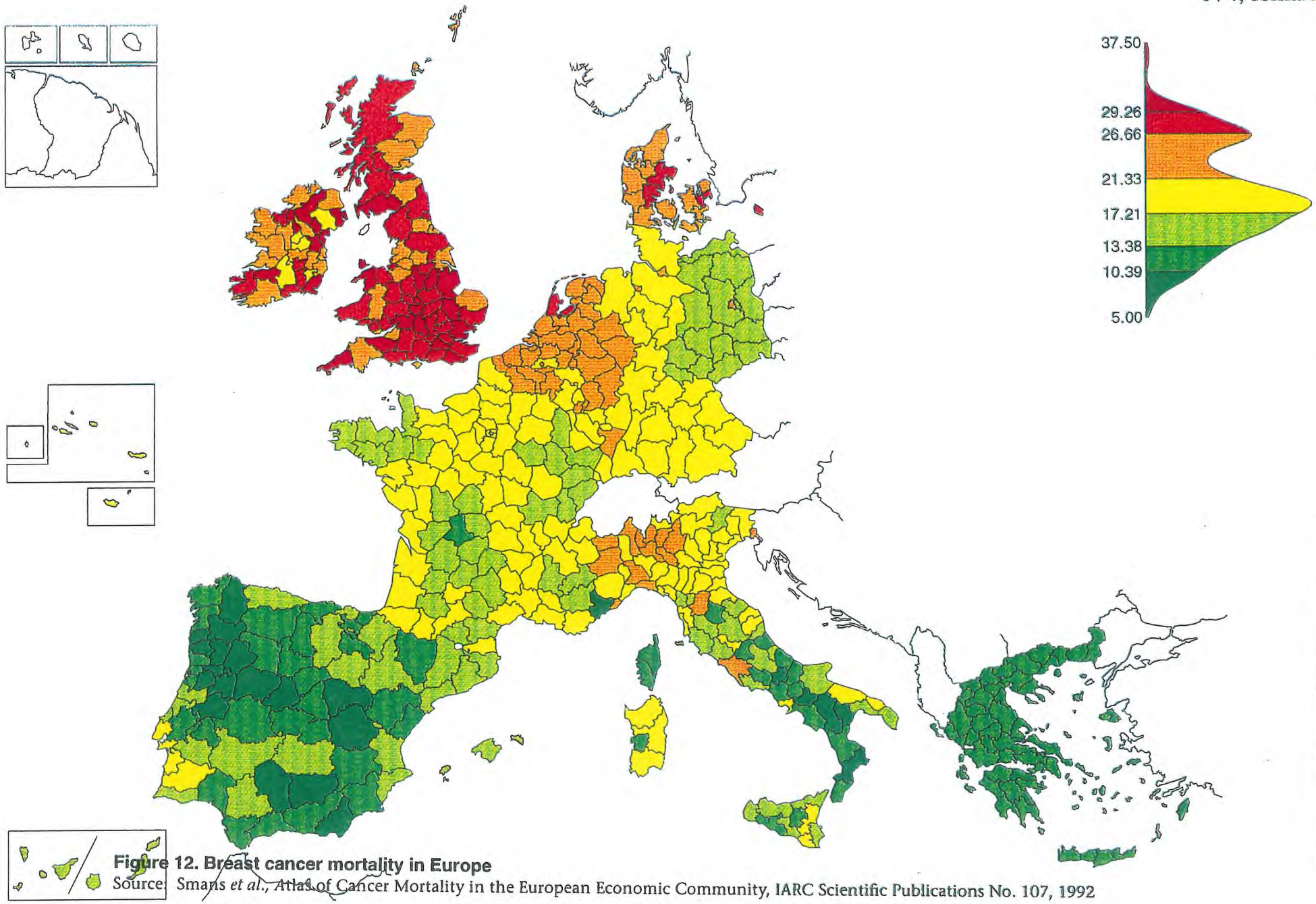


Figure 12. Breast cancer mortality in Europe
Source: Smans *et al.*, Atlas of Cancer Mortality in the European Economic Community, IARC Scientific Publications No. 107, 1992

EXERCISES ON DATA PRESENTATION

TABLES AND GRAPHS

1. Using the data in Table 1 (overleaf), make a table depicting the incidence rate of , mortality rate from and the incidence-mortality ratio of the 10 most frequently occurring cancers in the San Francisco-Oakland SMSA during 1976 (Note: the incidence-mortality ratios will have to be calculated.)
2. Using the data in Study 2, make a bar chart showing the distribution of lung cancer cases in Hawaii 1975-77 by race and sex.
3. Plot the age-specific mortality rate from lung cancer among males in the United States, 1975 (data from Study 3 below).
4. Plot the age-adjusted lung cancer mortality rates both for males and for females for the time period 1950-75 (Study 3). Repeat using a log scale for the rates.
5. Of the 610 lung cancer cases diagnosed in males in Hawaii 1975-77 (Study 2), 101 (16.6%) were localized, 149 (24.4%) were regional, 307 (50.3%) were distant and 53 (8.7%) were unstaged. Illustrate this information with a pie chart.

Study 2. A three year lung cancer study (1975 through 1977), Hawaii residents

There were 870 lung cancer cases diagnosed between 1975 and 1977. This study was done by sex, age, race, stage by sex, histology, treatment and survival by status of the disease. There were 610 males and 260 females.

Ages by sex		
Males	Age	Females
0	20-24	1
3	30-34	0
7	35-39	6
15	40-44	9
52	45-49	19
58	50-54	35
112	55-59	42
92	60-64	32
98	65-69	42
63	70-74	33
59	75-79	18
22	80-84	11
18	85-89	8
11	90+	4

Table 1. Number and age-adjusted (1970 US standard) incidence and mortality rates per 100 000 population by site, year and all races, San Francisco-Oakland SMSA (Alameda, Contra Costa, Marin, San Mateo and San Fransisco Counties), 1973–1976

	Incidence[a]						Mortality[b]					
	Both sexes		Male		Female		Both sexes		Male		Female	
	No.	Rate	No.	Rate	No.	Rate	No.	Rate	No.	Rate	No.	Rate
All sites												
1973	11 314	350.2	5333	386.9	5981	335.5	5476	170.0	2825	208.2	2651	145.8
1974	12 004	366.7	5618	401.3	6386	353.8	5476	176.5	3011	219.1	2745	149.0
1975	12 050	363.9	5503	389.9	6547	358.3	5857	177.0	3082	221.3	2775	147.9
1976	11 777	351.2	5571	390.1	6206	334.3	5790	172.2	3009	214.1	2781	145.1
1973–1976	47 145	358.0	22 025	392.0	25 120	345.5	22 879	173.9	11 927	215.7	10 952	147.0
Stomach												
1973	370	11.6*	221	16.9*	149	8.0*	283	8.8*	166	12.8*	117	6.1*
1974	358	10.9*	219	15.8*	139	7.1*	267	8.2*	156	11.6*	111	5.7*
1975	359	10.8*	211	15.2*	148	7.7*	280	8.4*	165	11.9*	115	5.9*
1976	358	10.6*	211	14.9*	147	7.2*	252	7.4*	154	11.2*	98	4.8‖
1973–1976	1445	11.0	862	15.7	583	7.5	1082	8.2	641	11.8	441	5.6
Colon												
1973	1097	34.2	505	38.1	592	32.0	564	17.5	263	20.1*	301	16.1*
1974	1114	34.1	541	39.8	573	30.4	601	18.3	287	21.5*	314	16.1*
1975	1160	35.0	515	37.6	645	33.4	611	18.3	289	21.1*	322	16.1*
1976	1189	35.3	552	39.5	637	32.5	606	17.9	286	20.8*	320	16.0*
1973–1976	4560	34.6	2113	38.7	2447	32.1	2382	18.0	1125	20.9	1257	16.2
Rectum												
1973	455	14.2	241	17.5*	214	11.6*	161	5.0*	90	6.8*	71	3.8‖
1974	472	14.4	261	18.7*	211	11.4*	159	4.9*	84	6.3*	75	3.9‖
1975	490	14.8	261	19.0*	229	11.9*	144	4.3*	68	5.1*	76	3.9‖
1976	537	16.0	305	21.4*	232	11.9*	146	4.2*	76	5.6*	70	3.3‖
1973–1976	1954	14.9	1068	19.2	886	11.7	610	4.6	318	6.0	292	3.7*
Pancreas												
1973	331	10.3*	177	12.9*	154	8.4*	317	9.9*	148	10.9*	169	9.2*
1974	360	11.0*	178	12.9*	182	9.9*	334	10.3*	166	12.0*	168	9.1*
1975	381	11.6*	206	14.6*	175	9.2*	371	11.2*	192	13.7*	179	9.4*
19763	322	9.6*	162	11.4*	160	8.1*	337	10.1*	186	13.1*	151	7.6*
1973–1976	1394	10.6	723	12.9	671	8.9	1359	10.4	692	12.4	667	8.8
Lung & bronchus												
1973	1548	48.1	1114	79.2	434	24.6	1124	35.0	819	58.7	305	17.2*
1974	1735	53.4	1224	86.1	511	28.7	1301	40.0	925	65.7	376	20.9*
1975	1695	51.5	1128	78.9	567	31.2	1283	39.0	917	64.8	366	20.0*
1976	1717	51.7	1156	80.5	561	30.7	1322	39.7	891	62.4	431	23.2
1973–1976	6695	51.2	4622	81.2	2073	28.8	5030	38.4	3552	62.9	1478	20.4
Melanoma												
1973	255	7.6*	122	8.0*	133	7.5*	67	2.0‖	37	2.6‖	30	1.7‖
1974	273	8.1*	133	8.6*	140	7.8*	49	1.5‖	27	1.8‖	22	1.2‖
1975	290	8.4*	141	8.9*	149	8.1*	54	1.6‖	29	2.0‖	25	1.4‖
1976	271	7.7*	126	7.8*	145	7.8*	55	1.6‖	35	2.4‖	20	1.0‖
1973–1976	1089	8.0	522	8.3	567	7.8	225	1.7*	128	2.2*	97	1.3‖
Breast												
1973	1641	50.4	18	1.3‖	1623	92.1	505	15.6	3	0.2‖	502	28.2
1974	1915	58.3	24	1.7‖	1891	106.1	588	17.8	3	0.2‖	585	32.2
1975	1762	53.3	7	0.5‖	1755	98.0	556	16.8	5	0.4‖	551	30.2
1976	1636	48.9	12	0.9‖	1624	89.3	515	15.3	5	0.4‖	510	27.6
1973–1976	6954	52.7	61	1.1‖	6893	96.4	2164	16.4	16	0.3‖	2148	29.5

Cervix												
1973	266	7.9*	–	–	266	14.9*	66	2.0‖	–	–	66	3.7‖
1974	215	6.4*	–	–	215	12.1*	81	2.5‖	–	–	81	4.5‖
1975	255	7.4*	–	–	255	14.0*	56	1.7‖	–	–	56	3.1‖
1976	227	6.6*	–	–	227	12.6*	68	2.0‖	–	–	68	3.7‖
1973–1976	936	7.1	–	–	963	13.4	271	2.0*	–	–	271	3.7*
Corpus												
1973	705	21.6	–	–	705	40.1	40	1.2‖	–	–	40	2.2‖
1974	729	22.1	–	–	729	41.0	36	1.1‖	–	–	36	1.9‖
1975	802	24.3	–	–	802	45.0	38	1.2‖	–	–	38	2.0‖
1976	741	22.1	–	–	741	41.0	39	1.1	–	–	39	2.0‖
1973–1976	2977	22.5	–	–	2977	41.8	153	1.2	–	–	153	2.0*
Ovary												
1973	279	8.5*	–	–	279	15.8*	161	5.0	–	–	161	9.0*
1974	298	9.1*	–	–	298	16.8*	181	5.6*	–	–	181	10.1*
1975	306	9.3*	–	–	306	17.1*	173	5.3*	–	–	173	9.5*
1976	254	7.6*	–	–	254	14.1*	193	5.7*	–	–	193	10.3*
1973–1976	1137	8.6	–	–	1137	16.0	708	5.4	–	–	708	9.7
Prostate												
1973	889	27.9	889	69.0	–	–	245	7.6*	245	19.7*	–	–
1974	938	28.8	938	71.8	–	–	272	8.3*	272	21.5*	–	–
1975	927	28.2	927	69.6	–	–	277	8.3*	277	21.4*	–	–
1976	931	27.6	931	68.9	–	–	291	8.5*	291	22.3*	–	–
1973–1976	3685	28.1	3685	69.8	–	–	1085	8.2	1085	21.2	–	–
Bladder												
1973	431	13.4	329	24.5*	102	5.5*	159	4.9*	109	8.6*	50	2.5*
1974	478	14.6	332	24.1*	146	7.7*	131	4.0*	89	6.8*	42	2.1‖
1975	560	16.9	405	29.3*	155	8.1*	156	4.6*	102	7.7*	54	2.6‖
1976	503	14.9	365	26.3*	138	7.0*	155	4.6*	106	7.9*	49	2.4‖
1973–1976	1972	15.0	1431	26.1	541	7.1	601	4.5	406	7.7	195	2.4*
Kidney												
1973	187	5.8*	121	8.5*	66	3.8‖	80	2.5‖	53	3.8‖	27	1.5‖
1974	207	6.4*	133	9.2*	74	4.1‖	97	3.0‖	65	4.7‖	32	1.7‖
1975	162	5.0*	116	7.9*	46	2.5‖	83	2.5‖	52	3.6‖	31	1.6‖
1976	222	6.8*	141	9.7*	81	4.6‖	84	2.5‖	56	3.8‖	28	1.4‖
1973–1976	778	6.0	511	8.8	267	3.8*	344	2.6*	226	4.0*	118	1.6*
Lymphomas												
1973	399	12.2*	210	14.1*	189	10.6*	215	6.6*	120	8.4*	95	5.2‖
1974	414	12.6	229	15.5*	185	10.3*	215	6.5*	118	8.2*	97	5.4‖
1975	451	13.5	241	15.8*	210	11.4*	217	6.5*	121	8.3*	96	5.0‖
1976	408	12.1	226	14.7*	182	9.7*	217	6.4*	120	8.3*	97	5.1‖
1973–1976	1672	12.6	906	15.0	766	10.5	864	6.5	479	8.3	385	5.2*
Leukaemias												
1973	305	9.8*	170	12.3*	135	7.9*	211	6.6*	116	8.6*	95	5.1‖
1974	297	9.4*	161	12.0*	136	7.6*	200	6.4*	114	8.6*	86	4.9‖
1975	283	8.8*	137	9.9*	146	8.1*	227	6.9*	116	8.3*	111	6.1*
1976	273	8.4*	155	11.3*	118	6.4*	225	6.7*	122	8.9*	103	5.4‖
1973–1976	1158	9.1	623	11.4	535	7.5	863	6.6	468	8.6	395	5.4*

a/b Source: SEER Program. Data for 1976 are provisional.
b Source: Number of deaths from National Center for Health Statistics

* Standard error of the rate is between 5 and 10
‖ Standard error of the rate is 10% or greater

Race by sex		
Males	Race	Females
219	Caucasian	95
162	Japanese	64
115	Hawaiin/Part Hawaiian	49
59	Filipino	22
34	Chinese	18
7	Korean	4
4	Puerto Rican	2
6	Others/Mixed races	4
4	No record of race	2

Stage at diagnosis by sex		
Males	Stage at diagnosis	Females
101	Local	56
	Regional:	
59	Lymph nodes	26
33	Direct extension	23
20	Both L.N. & extension	8
37	Regional NOS	15
307	Distant	115
53	Unstaged	17

Histologies

Cases	Histology
233	Squamous cell carcinomas
188	Adenocarcinomas NOS
110	Oat cell carcinoma
103	Carcinomas NOS
66	Bronchoalveolar carcinoma
55	Undifferentiated/anaplastic carcinoma
44	Large cell carcinoma
23	Adenosquamous carcinoma
13	Giant cell carcinoma
8	Mucinous adenocarcinoma
7	Papillary adenocarcinoma
6	Malignancy NOS
4	Malignant carcinoids
2	Fibrosarcomas
2	Leiomyosarcomas
1	Embryonal rhabdomyosarcoma
1	Mucoepidermoid cancer
1	Transitional cell carcinoma
1	Adenoid cystic carcinoma
1	Clear cell carcinoma
1	Papillary carcinoma NOS

First course treatment

12 cases wedge restrictions
7 cases wedge restrictions plus radiation
2 cases wedge restrictions plus radiation and chemotherapy
134 cases lobectomies
28 cases lobectomy plus radiation
3 cases lobectomy plus chemotherapy
6 cases lobectomy plus radiation and chemotherapy
4 cases lobectomy plus radiation and immunotherapy
24 cases more than 1 lobe removed
5 cases more than 1 lobe removed plus radiation
27 cases pneumonectomy
15 cases pneumonectomy plus radiation
1 case pneumonectomy plus chemotherapy
266 cases radiation only (primary or metastatic lesion)
36 cases chemotherapy only
110 cases radiation and chemotherapy
150 cases no treatment - reason:
- 99 patients expired before treatment began
- 38 patients no treatment because of age and condition
- 13 patients refused treatment
18 cases unknown if treatment done (no record of treatment or death in Hawaii)

22 cases autopsy diagnosis only

Survival by stage

Of the 870 patients diagnosed in the three year period, over one half, 422, had distant metastases at diagnosis and 287 expired with metastases within six months of diagnosis.

Stage at diagnosis:
expired patients' survival time

Local stage: *expired with metastases or recurrent carcinoma with exception of 4 cases who died post operatively.*
13 patients expired within 6 months
10 patients lived 7-12 months
12 patients lived 13-24 months
7 patients lived 25-36 months
2 patients lived 37-48 months
1 patient expired - status of disease unknown
6 patients expired - diagnosed at autopsy

Regional stage:
63 patients expired within 6 months
59 patients expired in 7-12 months
40 patients expired in 13-14 month
9 patients expired in 25-36 months
1 patient expired in 37-48 months
1 patient expired – unknown if from cancer
5 patients expired – autopsy diagnosis

Distant stage:

287 patients expired within 6 months
80 patients expired in 7-12 months
37 patients expired in 13-24 months
5 patients expired in 25-36 months
11 patients expired - diagnosed at autopsy

Unknown stage:

47 patients expired within 6 months
16 patients expired in 7-12 months
3 patients expired in 13-24 months
1 patient expired - cancer status unknown

There were 8 patients who expired with no cancer.

Status of patients alive without cancer: months of survival by stage at diagnosis

Local stage:

1 case	7-12 months
26 cases	13-24 months
28 cases	25-36 months
10 cases	37-48 months

Regional:

1 case	7-12 months
18 cases	13-24 months
11 cases	25-36 months
3 cases	37-48 months

Distant:

2 cases	13-24 months
1 case	25-36 months

Patients lost to follow-up by stage at diagnosis:

30 patients with localized disease
6 patients with regional disease
5 patients with distant disease
3 patients with unstaged disease

Study 3 Age-specific and age-adjusted lung cancer mortality rates per 100 000 population, females, United States, 1950-1975

Age	1950	1955	1960	1965	1970	1975
0-4	0.1	0.1	0.0	0.0	0.0	0.0
5-9	0.0	0.0	0.0	0.0	0.0	-
10-14	0.1	0.0	0.1	-	0.0	0.0
15-19	0.1	0.1	-	0.1	0.0	0.0
20-24	0.1	0.2	0.0	0.1	0.1	0.1
25-29	0.2	0.4	0.2	0.2	0.2	0.2
30-34	0.7	0.6	0.8	0.9	0.9	0.8
35-39	1.2	1.7	1.9	2.7	4.0	3.5
40-44	2.7	3.1	4.6	6.4	8.1	10.2
45-49	4.5	5.2	7.8	11.4	16.1	20.5
50-54	7.3	8.1	10.7	16.6	26.2	32.5
55-59	10.9	11.6	13.4	20.5	34.9	49.9
60-64	16.9	16.2	17.6	23.5	38.9	63.1
65-69	20.0	22.1	22.0	28.3	42.0	64.3
70-74	27.1	26.3	27.6	34.5	44.4	65.6
75-79	32.0	33.8	29.1	36.5	51.6	66.5
80-84	34.0	34.4	39.6	38.2	53.7	68.6
85+	28.0	30.0	38.8	39.0	50.0	66.9
1950 age-adj.	4.4	4.6	5.2	6.9	10.3	14.3

Age-specific and age-adjusted lung cancer mortality rates per 100 000 population, males, United States, 1950-1975

Age	1950	1955	1960	1965	1970	1975
0-4	0.0	0.0	0.0	0.1	0.1	-
5-9	0.1	0.0	0.0	0.0	0.0	-
10-14	0.1	0.1	0.0	0.0	0.0	-
15-19	0.2	0.1	0.1	0.1	0.1	0.0
20-24	0.2	0.2	0.1	0.2	0.3	0.1
25-29	0.4	0.7	0.7	0.4	0.6	0.4
30-34	1.8	1.8	2.2	1.9	2.2	1.9
35-39	4.2	5.1	6.1	7.8	7.9	6.9
40-44	10.6	12.1	15.3	19.8	7.9	6.9
45-49	25.1	30.9	33.3	40.2	47.3	51.6
50-54	47.3	58.7	70.2	79.4	89.6	96.2
55-59	75.0	95.0	115.1	132.9	156.4	158.1
60-64	98.6	133.2	168.5	190.9	229.1	252.7
65-69	99.0	152.1	200.1	249.0	303.7	330.9
70-74	99.4	143.7	209.9	284.0	343.8	414.8
75-79	88.8	133.3	174.9	255.1	350.6	432.3
80-84	72.6	113.1	151.3	206.8	295.5	392.3
85+	63.1	73.0	107.7	136.4	194.0	266.9
1950 age-adj.	29.5	27.8	35.3	43.4	52.8	58.8

ANSWERS TO THE EXERCISES

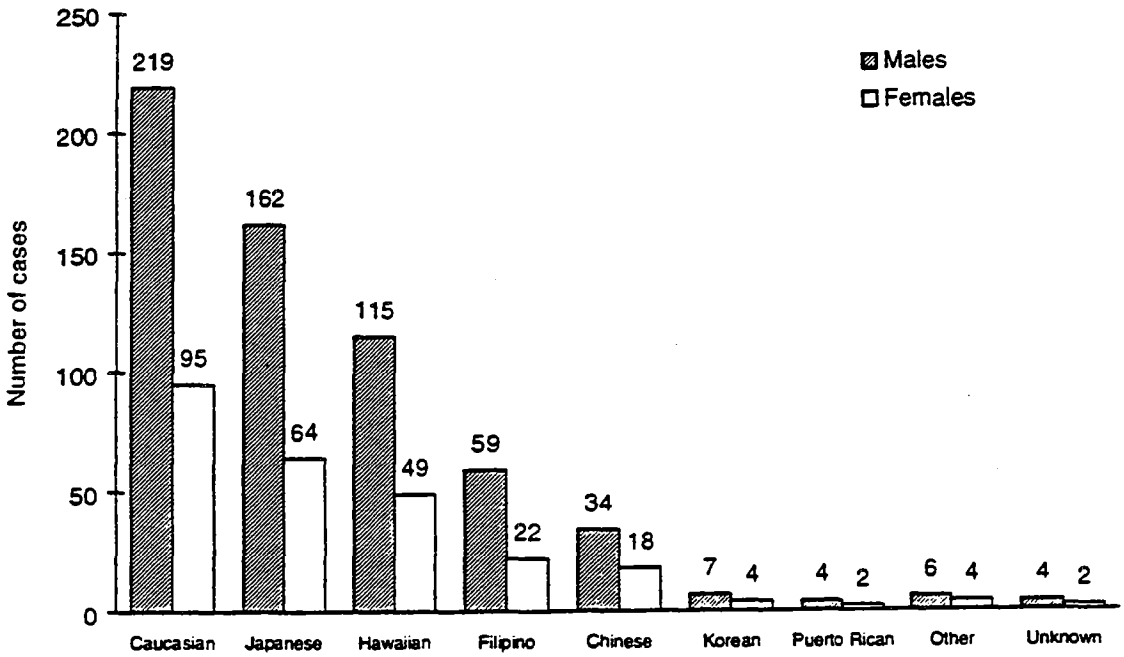
1. Age-adjusted (1970 standard) incidence and mortality rates per 100 000 population for the 10 most common sites in San Francisco - Oakland SMSA*, 1976

SITE	Incidence	Mortality	Mortality/incidence ratio
Lung and bronchus	51.7	39.7	1.30
Breast	48.9	15.3	3.20
Colon	35.3	17.9	1.97
Prostate	27.6	8.5	3.25
Corpus	22.1	1.1	20.09
Rectum	16.0	4.2	3.81
Bladder	14.9	4.6	3.24
Lymphoma	12.1	6.4	1.89
Stomach	10.6	7.4	1.43
Pancreas	9.6	10.1	0.95

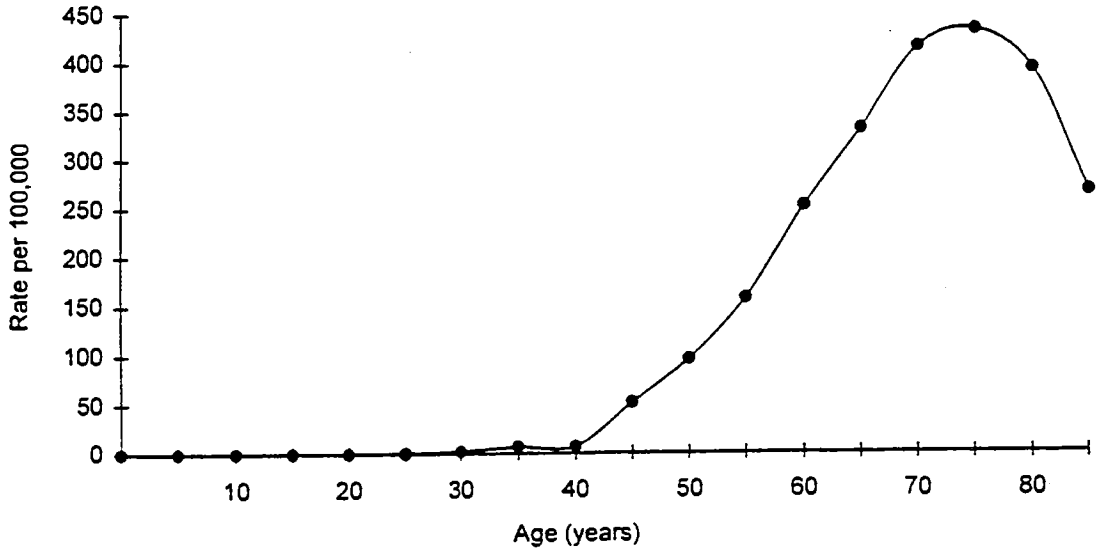
Source: Cancer Incidence and Mortality in the United States, SEER, 1973-1976. DHEW Publication (NIH) 78-1837, US Department of Health, Education, and Welfare, Public Health Service, National Institutes of Health

*SMSA: Standard Metropolitan Statistical Area

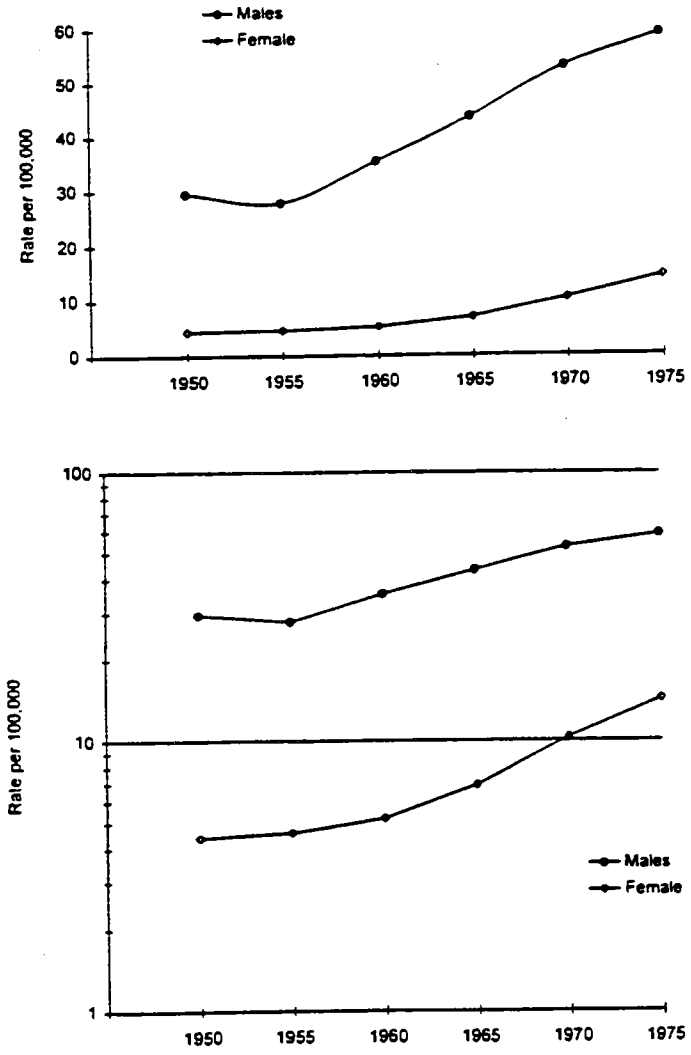
2. **Distribution of lung cancer cases in Hawaii 1975-1977, by race and sex**



3. Age-specific lung cancer mortality rates per 100 000, males, US, 1975



4. Age-adjusted lung cancer mortality rates per 100 000, males and females, 1950-1975, US



5. Stage at diagnosis for males diagnosed with lung cancer, 1975-1977, in Hawaii