Lesson 3

Summarizing Nominal and Ordinal Data

Lesson 3 Overview

- Lesson 2 covered summary statistics, graphs and tables for numerical data
- The other main measurement scale is Categorical data which includes both nominal (no order) and ordinal (ordered) data
- Summary statistics and graphs for categorical data are covered in Lesson 3.

Outline

- Proportions
- Rates
  - Mortality Rates
  - Morbidity Rates
  - Change rates
  - Follow-up rates
  - Adjusting Rates (skip) – read in your book if you are interested: pages 44-46
- Graphs for Proportions and Rates

Review: Categorical (Qualitative) Data

- Data that can be classified as belonging to a distinct number of categories is called categorical data

  - Ordinal – data classified into categories that have a natural ordering
  - Nominal- data classified into categories that do not have a natural order
  - Binary – data classified into one of 2 possible categories

Proportions

- A proportion is the number of observations with the characteristic of interest divided by the total number of observations
- The proportion is represented by the letter *p*
- The number of observations with the event or characteristic of interest = x
- The total number of observations = n

$$p = \frac{x}{n}$$

$$percent = \frac{x}{n}(100)$$

Properties of Proportion

- The numerator is a subset of the denominator so the numerator is always less than or equal to the denominator

- Proportions are always between 0 and 1
  - Or between 0% and 100% if expressed as a percent

Proportions Example

- Data are from 34 patients with acquired hemophilia (Table 3-16 in text)

	Hematuria		
	No	**Yes**	**Total**
Men	**9**	**6**	**15**
Women	**12**	**7**	**19**
Total	**21**	**13**	**34**

- From the data calculate the proportion
- of all patients with hematuria 13 / 34 = 0.38
- of men with hematuria 6 / 15 = 0.40
- of women with hematuria 7/19 = 0.37

Calculating Proportions

- The first step in calculating proportions is to make sure you have the correct denominator
- In the previous example, the word ‘of’ identifies the group in the denominator. The number of subjects in the identified group is the number in the denominator

- *Work through the proportions calculations on the Lesson 3 Practice Exercises*

Other summary statistics for categorical data

- Besides proportions, categorical data can be summarized with rates
- Several different rates will be defined in this lesson
- Rates are usually not proportions because, in most cases, the numerator is not a subset of the denominator.

Rates

- Rates are usually measured over an interval of time. The numerator represents the number of events over a specified period of time.

- Rates used to summarize the number of events in a population are often expressed as the number of events per 1000 (or 10,000 or million) persons.

Choosing the base

- Choose a base that results in a rate that is easy to interpret
  - Example: 20/ 100,000 is easier to interpret than 0.2 /1000

- Rates with the same base can easily be compared. Rates with different bases cannot be directly compared.

Example: Heart Disease Death Rate

Minnesota has the lowest heart disease death rate of any state in the nation, according to a report by the American Heart Association.

In 2004, Minnesota had 90 coronary heart disease deaths per 100,000 people, compared to the national average of 150 per 100,000, according to the report released in Dec. 2007

- The Minnesota rate can be directly compared to the national rate because the same base (100,000) was used.

Some Common Epidemiology Rates

- Morbidity rates – measure health status of a population
  - Prevalence (Not really a rate)
  - Incidence
- Mortality rates
  - Crude rate
  - Sex-specific mortality rates
  - Cause-specific mortality rates
  - Adjusted mortality rates

Prevalence

Disease Prevalence =

$$\frac{\text{Number of diseased persons at a given time}}{\text{Total number of persons at risk for the disease at that time}}$$

Multiply the calculated prevalence by a base (such as 100,000) to report prevalence as ‘per 100,000’ or some other meaningful base.

■ *Find the appropriate base for each prevalence on the Lesson 3 Practice Exercises*

Incidence Rate

Disease Incidence =

$$\frac{\text{Number of “new cases” in a period of time}}{\text{Number of individuals at risk for the disease (at the beginning of the time period).}}$$

Incidence is calculated over a specified period of time which should always be referenced.

Annual incidence rates are calculated over a 1 year period.

Prevalence and Incidence

- Prevalence
  - Represents disease status of population at one point in time – similar to data collected from a cross-sectional study
  - Actually a proportion rather than a rate
  - Includes those who have had the disease for an extended period of time in addition to new cases
  - Will be larger than incidence for chronic diseases
- Incidence
  - Represents disease status over a period of time - similar to data from a longitudinal study
  - Only includes ‘new’ cases
  - Will be larger than prevalence for diseases with short duration

Mortality Rates

Crude Mortality Rate = the number of deaths from any cause over a specified time interval divided by the total population of interest.

For annual crude mortality rates, the total population is estimated at the midpoint of the year.

$$\text{Annual Crude Mortality Rate} = \frac{\text{Number of deaths in a calendar year}}{\text{Population at midpoint of the year}}$$

Other Mortality rates

- Sex-specific mortality rates
  - Calculate crude mortality rates separately for men and women

- Cause-specific mortality rates
  - Number of deaths from a specific disease over a specified time period divided by the population of interest
  - Tracked over time to measure disease trends

- Infant Mortality rate
  - Number of infants who die before 1 year divided by number of live births
  - Usually represented as ‘per 1000 live births’

Adjusted Rates

- Crude rates are useful to describe a single population

- *Comparisons of crude rates are often invalid because the populations may be different with respect to important characteristics (ie. age, gender, race) that affect mortality or morbidity.*

- Adjusted rates can be used to compare rates for two populations.

Two more rates

- Two more rates that are often used in reporting health data or study results are
  - Change rate
  - Follow-up rate
- These are described in the next few slides. There are change rate and follow-up rate exercises on the Lesson 3 Practice Exercises

Change Rate

- Change rates are used to describe the change in some measure over a period of time.

$$\text{change rate (\%)} = \frac{\text{new value - old value}}{\text{old value}} \text{ X } 100$$

- Example: In 2002 there were 741,253 food shelf visits in the Twin Cities. In 2006 there were 898,311 visits. Calculate the change rate for visits to food shelves from 2002 to 2006:

$$\left(\frac{898{,}311 - 741{,}253}{741{,}253}\right)100 = 21.2\%$$

Source: Star Tribune, Dec. 2007

Change Rates: Some Notes

- The change rate is NOT a proportion.
  - The numerator is not a subset of the denominator
  - Unlike proportions, which can only take on values between 0 and 1 (or 0 % and 100%), the change rate can take on values > 100% or < 0% (negative values).
- A positive change rate indicates an increase over time, a negative change rate indicates a decrease over time

- *Calculate change rates on the Lesson 3 Practice Exercises*

Follow-up Rate

- General Definition:

$$\frac{\text{Number of events}}{\text{Total person-time of observation}}$$

- Events can be:
  - Death
  - Disease
  - Relapse
  - Accidents, etc.
- Person-time can be:
  - Years
  - Months
  - Treatment months (number of months subjects are in treatment)

Person-time

- Person-time: the sum of the amount of time each individual is observed while
  - *free of disease* if study outcome is *disease*

    or
  - *alive* if study outcome is *death.*

- Each subject may contribute a different amount of person-time in a study.

Person-time

- Follow-up time begins when a subject is enrolled in the study
  - *Not all subjects enter the study at the same time*
- Follow-up time ends with one of three events:
  - *Subject is still alive (or disease free) on the analysis date.*
  - *Subject died (or diagnosed) on a known date within the study period.*
  - *Subject was lost to follow-up after a certain date.*

Person-time Example

Cohort	Subjects	Follow-up time (years)	Outcome (at study end)
Exposed to Asbestos	A	4.2	Dead
	B	3.8	Dead
	C	10	Alive
Not exposed to Asbestos	D	4	Alive
	E	2.8	Alive
	F	2.2	Dead

Follow-up death rate for 'Exposed to Asbestos':

2 deaths / (4.2 + 3.8 + 10) person-years:

2 deaths / 18 person-years = 1 death / 9 person-years

Follow-up death rate for 'Not Exposed to Asbestos':

1 death / (4 + 2.8 + 2.2) person-years:

1 death / 9 person-years

Follow-up rates: Notes

- Follow-up rates allow for more valid comparison between groups when follow-up time is not the same in both (or all) groups.
  - Notice in the example that the follow-up death rate is the same for both groups even though there were 2 deaths in one group and only one death in the other group.
- Can be expressed as per 100, 1000, 100,000, etc. person-years, person-months, treatment months
  - When comparing follow-up rates, make sure the reference value in the denominator is the same.

Displaying Proportions and Rates Graphically

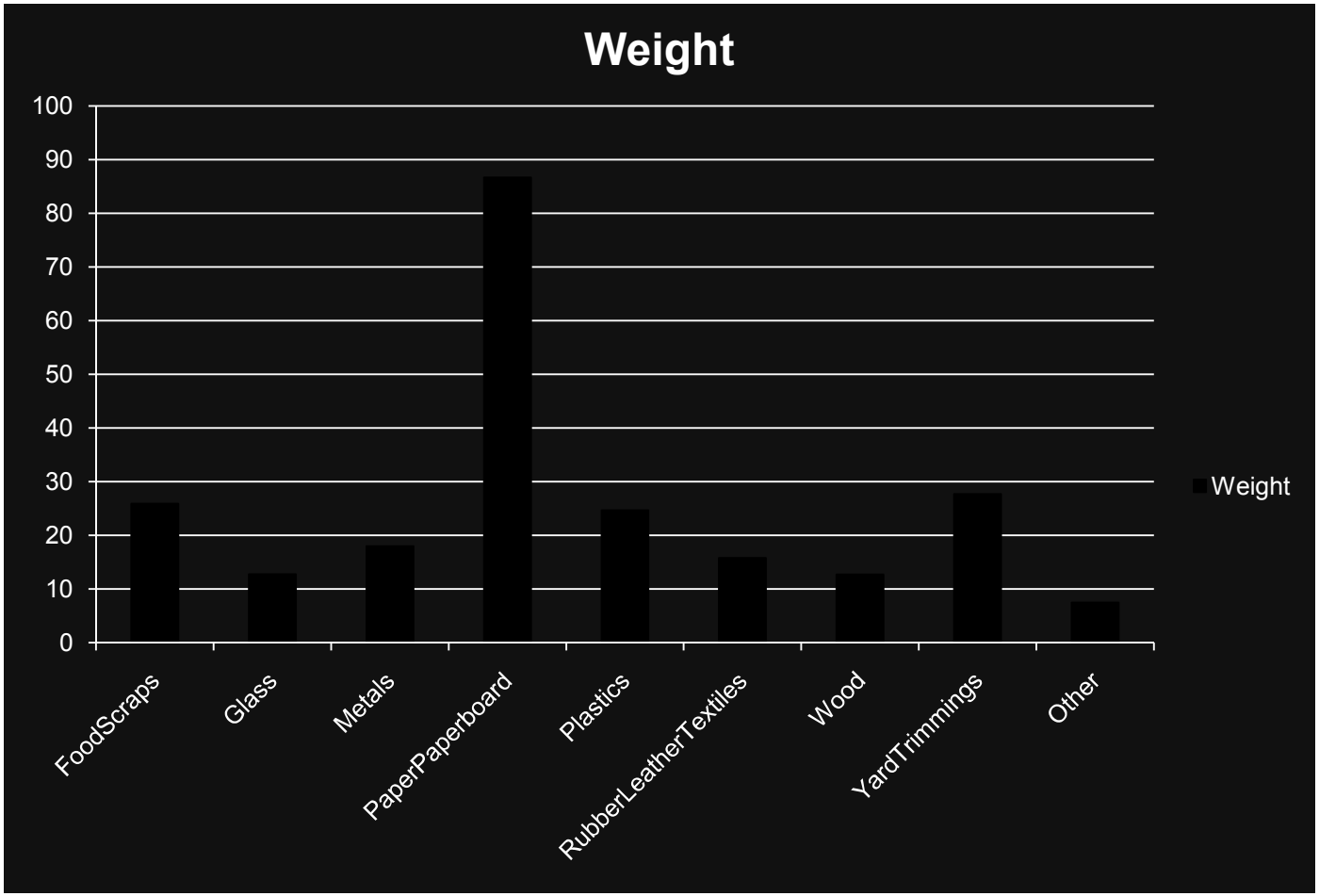
- Proportions can be graphically represented by
  - Bar Chart – more often used in scientific research
  - Pie Chart – more often used in the media

- Rates are often presented in line graphs where the x-axis is time and the line illustrates the change in rates over time.

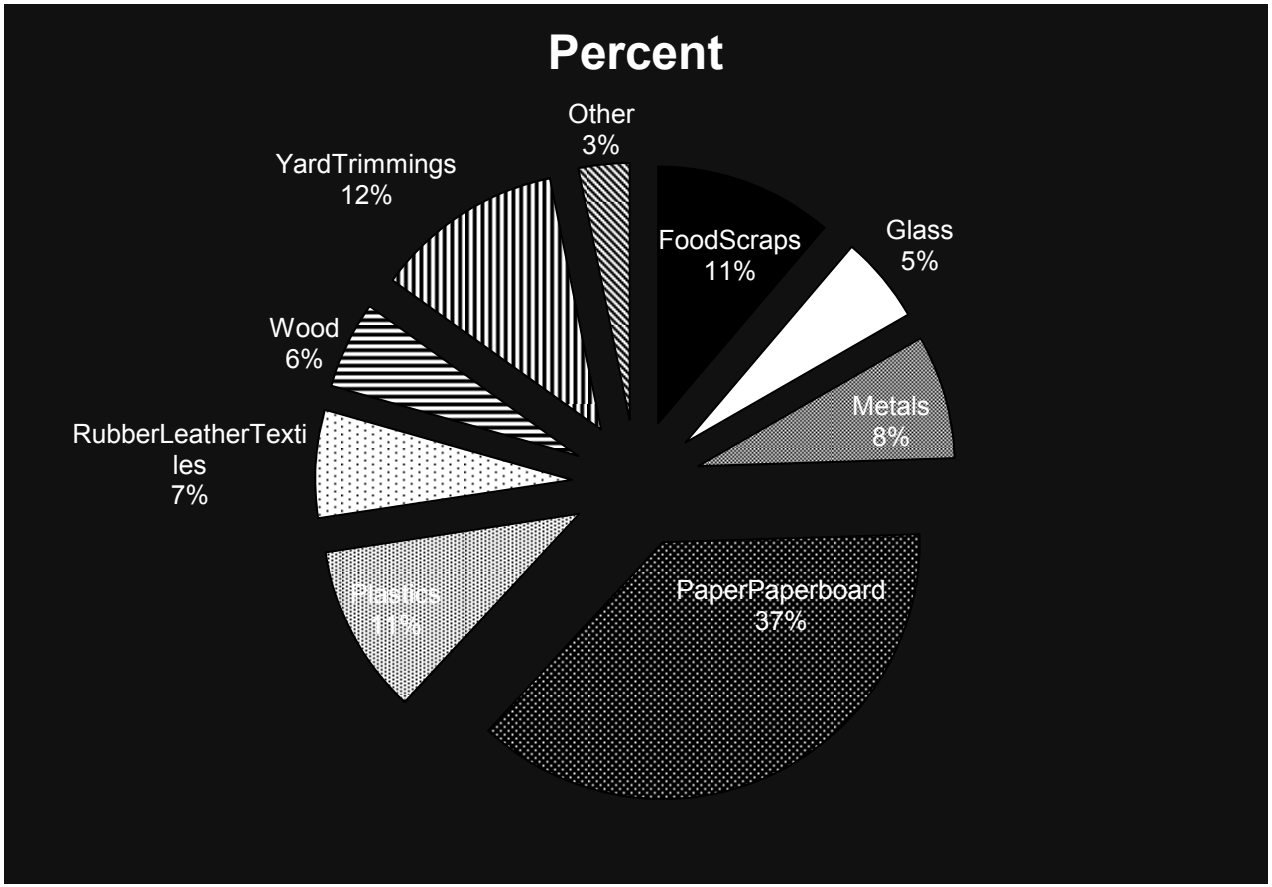
Garbage Content

Material	Weight	Percent
FoodScraps	25.9	11.2
Glass	12.8	5.5
Metals	18	7.8
PaperPaperboard	86.7	37.4
Plastics	24.7	10.7
RubberLeatherTextiles	15.8	6.8
Wood	12.7	5.5
YardTrimmings	27.7	11.9
Other	7.5	3.2

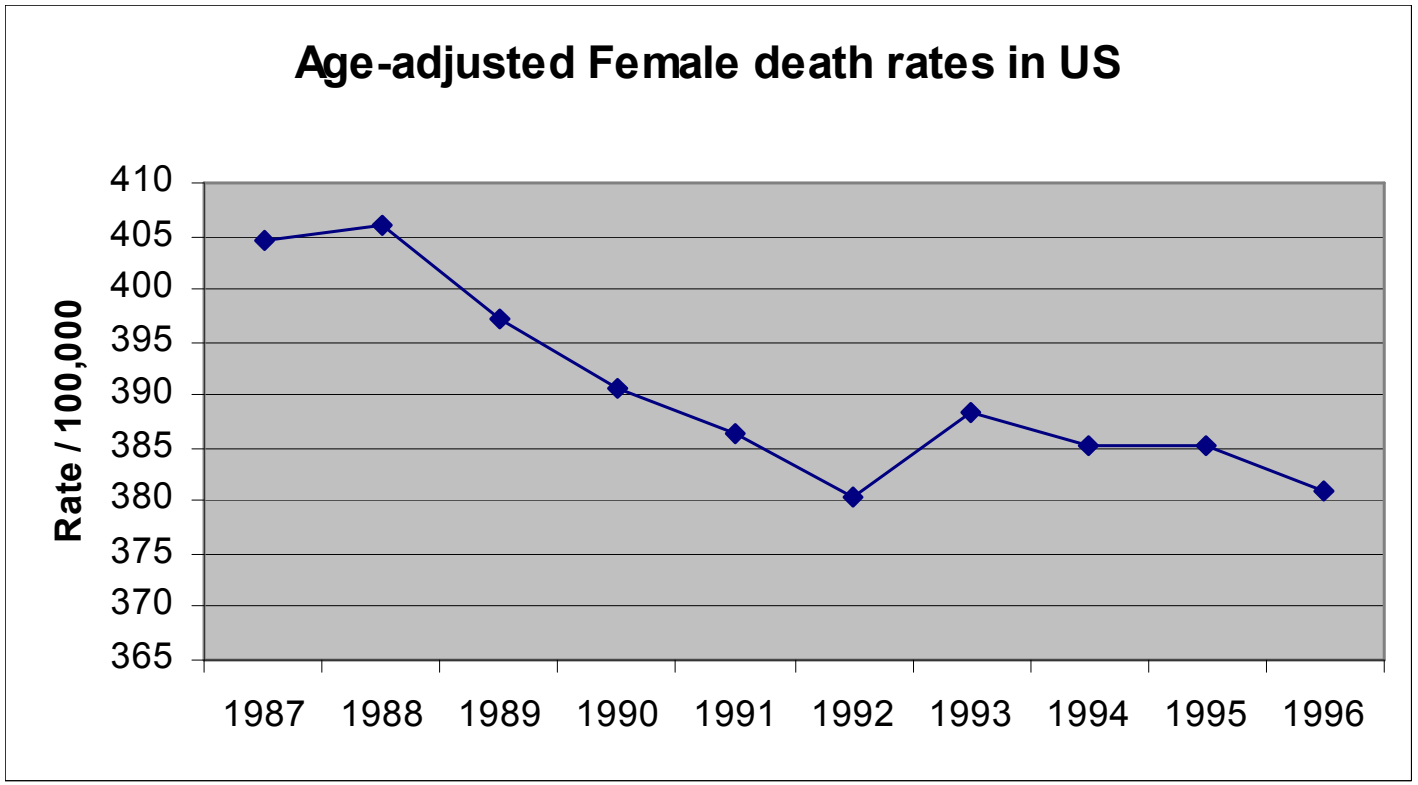
Bar Chart of Garbage Weight



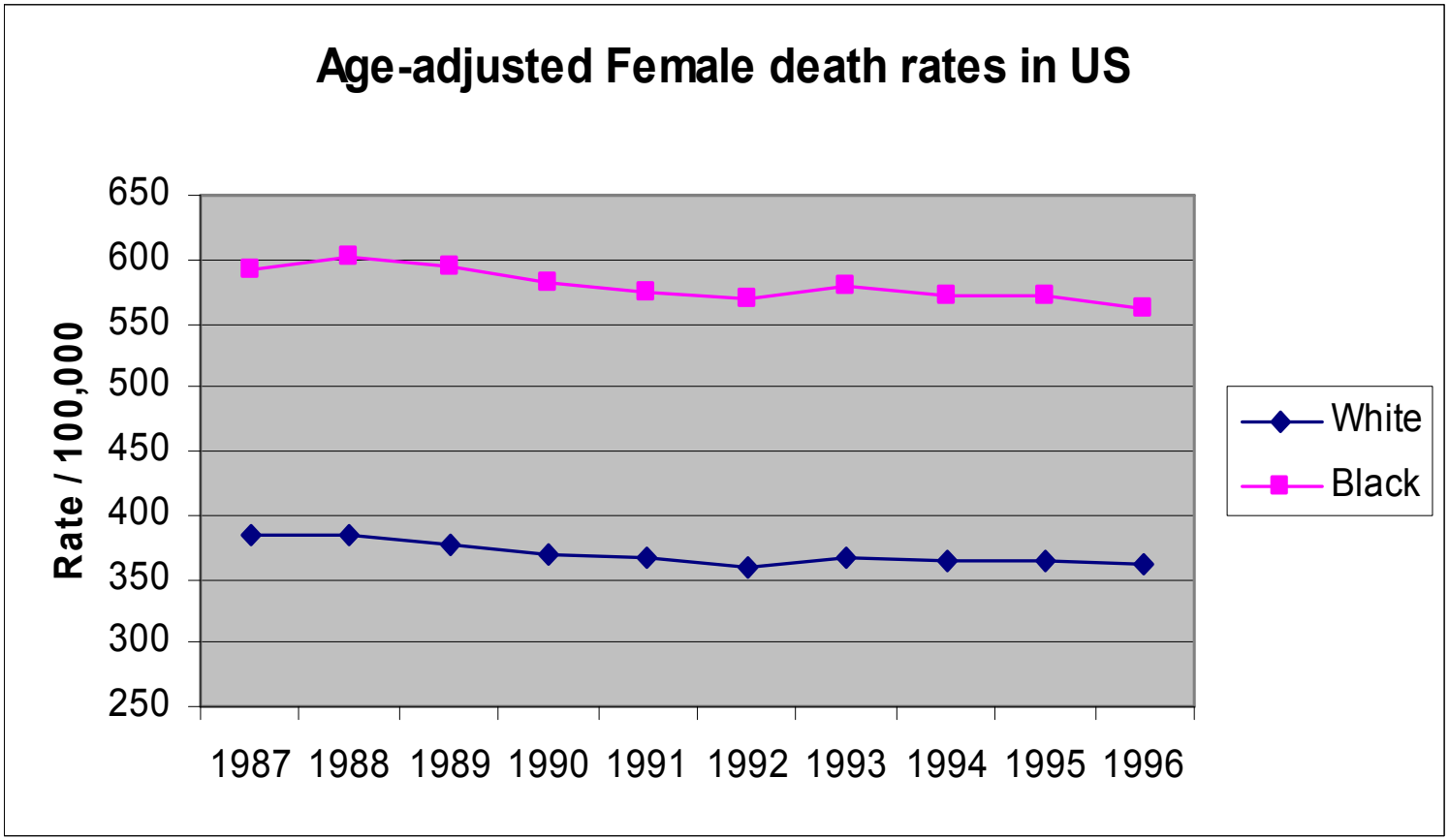
Pie Chart of Garbage Percent



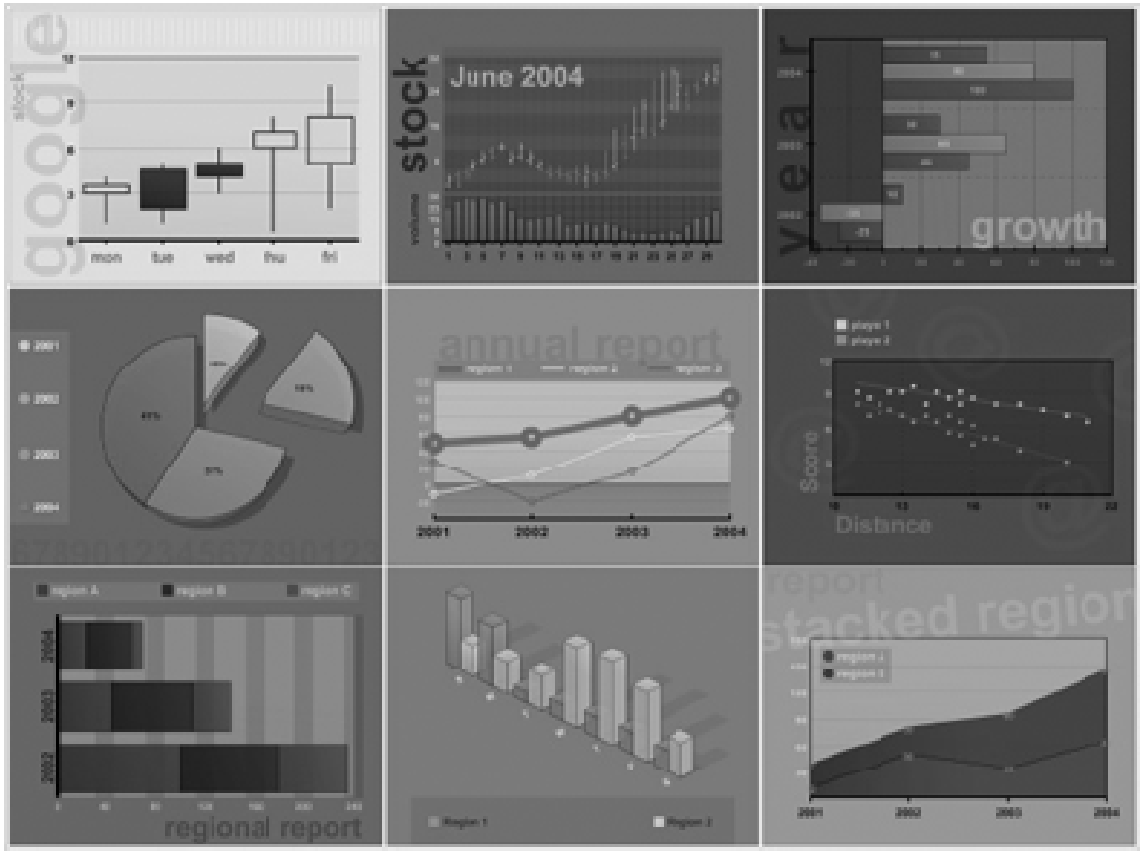
Line graph of death rates for data in Table 3-10 (pg 43)



Line graph to compare rates
Table 3-10, pg. 43



Exploratory Data Analysis with Pictures – What to use ?



Every day in 2006, an average of 101 Minnesotans died. These averages included:

25 cancer deaths

21 heart disease-related deaths

8 respiratory deaths

6 stroke deaths

3 diabetic deaths

4 Alzheimer´s deaths

27 deaths from other natural causes

8 injury/accident/violent death

Source: 2006 Minnesota Health Statistics Annual Summary, Minnesota Department of Health, Dec. 2007

What type of diagram would represent this information best?

1. Stem & Leaf Plot
2. Time Chart
3. Bar Chart
4. Histogram

Every day in 2006, an average of 101 Minnesotans died. You have time from diagnosis until death for the 25 cancer deaths

What type of diagram would represent this information best?

1. Stem & Leaf Plot
2. Pie Chart
3. Bar Chart
4. Admissions Chart

Every day in 2006, many Minnesotans visited the emergency room at HCMC. You have time in the emergency waiting room for 2,000 HCMC ER visits.

What type of diagram would represent this information best?

1. Stem & Leaf Plot
2. Pie Chart
3. Bar Chart
4. Histogram

Readings and Assignments

Reading

- Chapter 3 pgs. 40 - 47

Assignments

- Complete Homework 1 (Problems 5-7) and submit by due date.
- Lesson 3 Excel Module
- Lesson 3 Practice Exercises