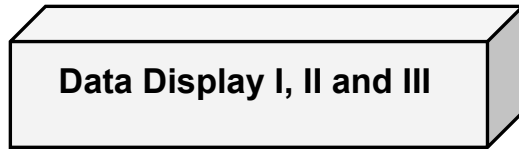


Lecture Notes for C03, C04, and C05



Highlights: Quiz 1 on Wednesday, 1/16

HWK due on Friday, 1/18

C03,04,05-Data Display 1

Announcements

- Daytime Office hours to be announced soon
- Evening help sessions start in Bashinsky 6
- First homework to be turned in: Friday 1/18
Follow Format template or lose points
- Quiz 1: online between Wednesday 1/16, 11 a.m. to Thursday 1/17, 11 a.m. Follow on screen instructions carefully!
- Do not printout lecture notes until announced in class
- Homework is shown in the lecture notes

C03,04,05-Data Display 2

Print

Section No.
here in bold

Format of work to be turned in

PRINT

Last name: _____ First name _____
Homework for (class no.) _____ Due in (class no.) _____
Class Time: _____ Seat Number: _____ Today's Date: _____

Print

First letter of
your last name in
big/ bold here

Write your answers below the line

Adhere to the page limit

Type everything in Word

Copy-paste what you need to from Excel

Be brief but explain your answers

Do the work yourself

Go to web site and download this form!

C03,04,05-Data Display 3

STATISTICS is the science of

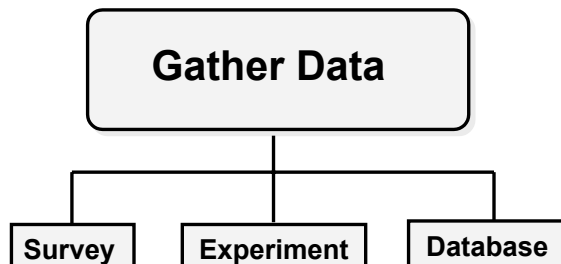
- collecting,
- organizing,
- summarizing, and
- interpreting...

DATA
for decision-making purposes.

C03,04,05-Data Display 5

Must make a decision!

Where do you start?



C03,04,05-Data Display 7

Obtaining Data Files from Outlook

- Data files from me are in the following location:

Public Folders

All Public Folders

Student Folders

Class Related Folders

Management Science and Statistics Faculty

Dr. Chakraborti's Classes

ST 260 Statistical Data Analysis

Dr. Mansfield

Johnson & Wichern Data Sets

Other Data Sets

- There may be more than one version (.txt, .xls).
- Click the desired posting to select it
- If .xls file, double click to open it in Excel
- Otherwise File → Save Attachments...

C03,04,05-Data Display 8

After we collect data, then what?

Organize / Summarize
Data

Graphical

Numerical

A Picture is worth a
A

words!

Graphs

- Histogram
 - Stem-and-leaf plot
- for a _____
variable
- Bar Chart
 - Pie Chart
- for a _____
variable

Graphs

- Histogram
 - Stem-and-leaf plot
- for a _single_
quantitative
variable
- Bar Chart
 - Pie Chart
- for a _single_
qualitative
variable

Graphs are used

To display distribution of a variable

shows:

- the type of variable
- possible values or categories
- # of occurrences (frequency)
- other features

Student Survey Data

id	height	shoesize	hairs	gender	major	minor
1	65	10	M	F	Marketing	
2	68	10	M	F	Accounting	Education
3	64	7	F	F	Psychology	
4	73	12	M	F	BSB	
5	72	11	F	F	Engineering	
6	75	17	M	F	International Business Management	
7	68	10	F	F	Marketing	
8	66	8	F	F	Marketing	
9	65	8	F	F	Marketing	
10	72	10	F	F	Marketing	
11	59	7	F	F	Accounting	
12	74	11	F	F	Undecided	
13	64	8	F	F	Accounting	
14	72	13	F	F	Accounting	
15	66	8	F	F	Accounting	
16	65	8	F	F	Undecided	
17	62	8	F	F	Mathematics	
18	66	8	F	F	Accounting	
19	64	8	F	F	Finance	
20	68	8	F	F	Marketing	
21	68	10	F	F	Accounting	
22	64	8	F	F	Health Care Management	
23	68	10	F	F	International Economics	
24	67	8	F	F	Marketing	
25	67	8	F	F	Business Management	
26	68	8	F	F	Accounting	
27	64	8	F	F	Marketing	
28	68	10	F	F	Secondary Education	
29	68	10	F	F	BSB	
30	68	10	F	F	Marketing	
31	67	10	F	F	Finance	
32	67	10	F	F	Marketing	
33	68	10	F	F	Health Care Management	
34	67	10	F	F	Finance	
35	67	10	F	F	Marketing	
36	68	10	F	F	Public Relations	
37	68	10	F	F	Business Management	
38	67	10	F	F	Marketing	
39	68	10	F	F	Marketing	
40	68	10	F	F	Marketing	
41	68	10	F	F	Marketing	
42	68	10	F	F	Business Management	
43	68	10	F	F	Marketing	
44	68	10	F	F	Marketing	
45	68	10	F	F	Accounting	
46	68	10	F	F	Finance	
47	68	10	F	F	Finance	
48	68	10	F	F	Business Management	

Data, by
itself, is not
information

Graphs

- Histogram
 - Stem-and-leaf plot
- for a single quantitative variable

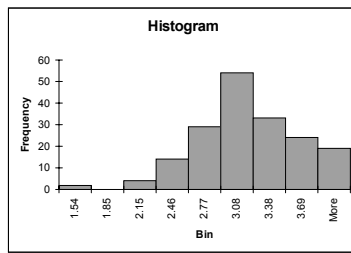
Histogram in Excel

- Enter the data or import it (from outlook)! Always examine the data!
- Use *tools -- data analysis -- histogram*
or use the *Kadd plug-in*
(Kadd is available only in the lab)
- Some editing of graph is necessary!
- Consult web site for more help!

Bin	Frequency
1.54	2
1.85	0
2.15	4
2.46	14
2.77	29
3.08	54
3.38	33
3.69	24
More	19
total	179

↑
Frequency distribution

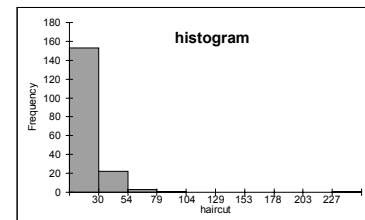
Student Survey Data section 1:
Variable: GPA
Using Excel



Lower End	Upper End	Frequency
	30	153
30	54	22
54	79	3
79	104	1
104	129	0
129	153	0
153	178	0
178	203	0
203	227	0
227		1

↑
Frequency distribution

Student Survey Data section 1:
Variable: haircut
Using Excel AND Kadd

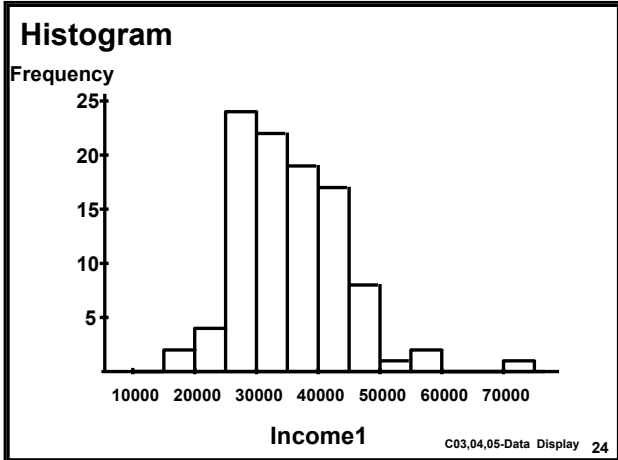


In Excel:
--tools--data analysis--
histogram (choose data)-- check
chart output



Histogram

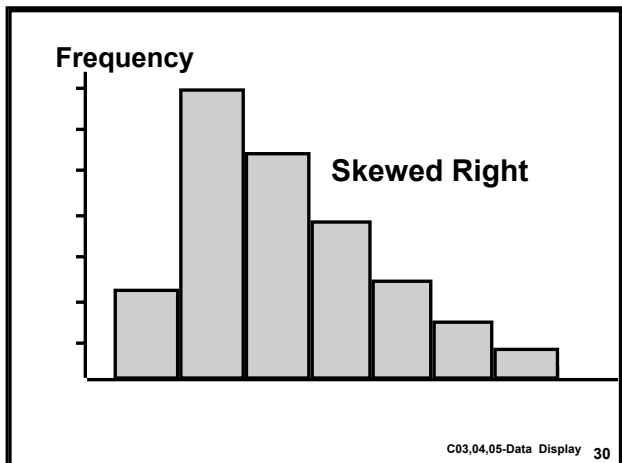
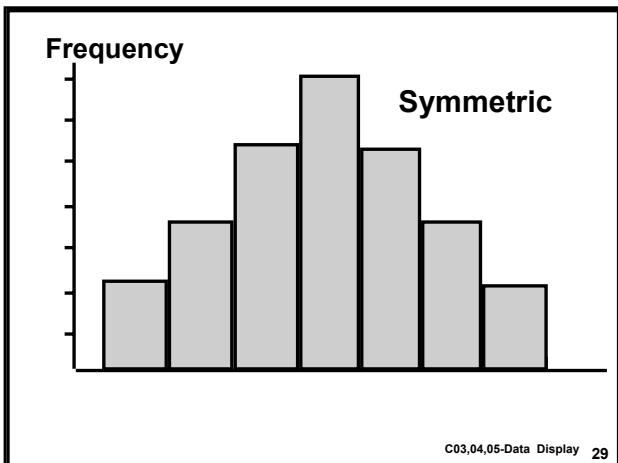
- Frequency or Relative Frequency
- Breaks the range of data into consecutive class-intervals (or bins) and shows the count or percent of data that fall into each interval
- Height of the bar shows the frequency or the relative frequency (=frequency/total number)
- Bins are usually of equal width
- Use relative frequency histograms to compare distributions with different no. of observations

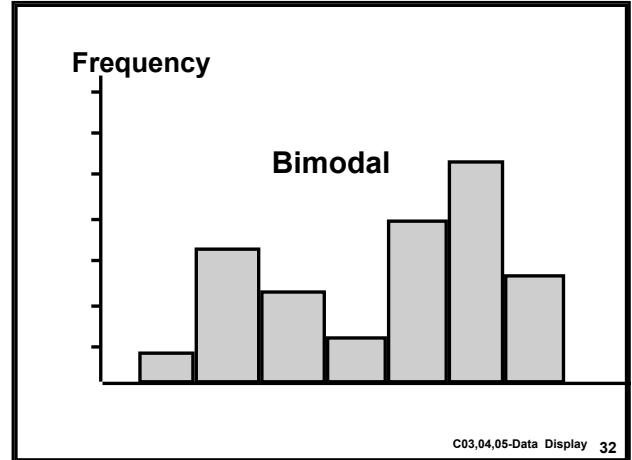
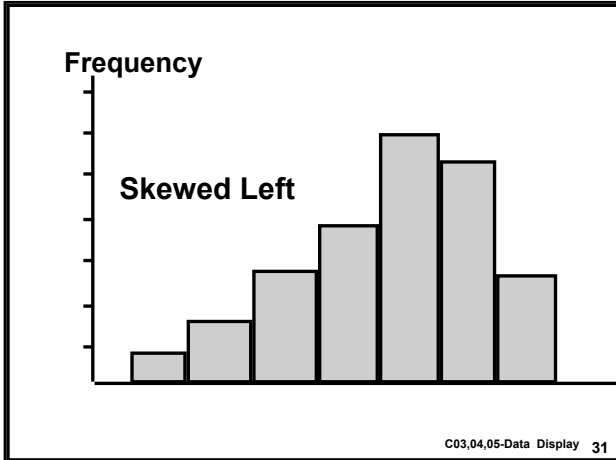


- ### Key Features of Data Distributions to look for:
- Shape
 - Typical Value
 - Spread
 - Outliers
- S,T,S,O**
- C03,04,05-Data Display 25

- ### Shapes of data distributions
- Symmetric
 - Skewed _____
 - Skewed _____
 - Bimodal (multi-modal)
- C03,04,05-Data Display 26

- ### Shapes of data distributions
- Symmetric
 - Skewed right
 - Skewed left
 - Bimodal (multi-modal)
- C03,04,05-Data Display 28





Additional Excel Notes

- Experiment with different sets of class intervals to find the histogram that best describes the data distribution.
- Always title your graphs and label the axes with meaningful variable names.
- Use cut-and-paste to paste Excel graphs into Word documents for final reports.

C03,04,05-Data Display 33

Homework for C03 (not to be turned in)

DO:

1. Draw a histogram for any appropriate variable (you choose) in fam100.xls data file (in outlook folder: see previous notes) using EXCEL. Use default no. of classes. Title your graph, label axes, etc. for readability. Write some brief comments on what you see. Find the *smallest, largest, and typical value(s)* for your variable.

What variable did you choose? Why? What does your histogram tell you? Write a few brief comments addressing S, T, S, O....

READ: Notes for C03 and section 2.4 from the book for next class

C03,04,05-Data Display 34

Stem-and-Leaf Plot *For a single quantitative variable*

- Example: Suppose data are
1.5, 1.9, 1.6, 2.3, 2.1, 2.3, 2.8, 2.7, 2.9, 3.3, 3.7, 5.2, 5.3, 5.5, 9.3

Step 1 Arrange data from lowest to highest

=> 1.5,1.6,1.9,2.1,2.3,....,9.3

C03,04,05-Data Display 36

Step 2

Stem:

to the left e.g. 1, 2, 3

Leaf:

to the right e.g. 5,6,
#’s are 1.5,1.6,1.9, ...

=> leaf unit = .1

1 | 5 means value = 15 x .1 = 1.5

C03,04,05-Data Display 37

Stem-and-Leaf Plot

```

1  569
2  133789
3  37
4
5  235
6
7
8
9  3
    
```

C03,04,05-Data Display 38

Stem-and-Leaf Plot

```

1  569
2  133789
3  37
4
5  235
6
7
8
9  3
    
```

1.5 is the lowest

Leaf unit = 0.1

Stems are integers written vertically in an ascending order

Leaves are integers written horizontally in an ascending order

Note: we write all numbers between 1 and 9 as stems even though we don't have all of them; those are missing their leaves

9.3 is the highest

C03,04,05-Data Display 39

The graphic shows:

- Smallest value = 1.5
- Largest value = 9.3
- Range = $9.3 - 1.5 = 7.8$ (S)
- Largest value far from rest (O?)
- Six numbers are with stem 2; most data around 2 (T = 2.3)
- shape--skewed right (S)

• Remarks about #'s with 2 or more digits after the decimal e.g. prob 2.10; drop all but the first digit after the decimal... C03,04,05-Data Display 40

Stem-and-Leaf Notes

- The “leaf” for a data value is always the first digit to the right of the dividing line.
- The “stem” for a data value is all digits to the left of the dividing line.
- The “leaf” is always one digit, but a “stem” can be more than one digit; if the data have 2 or more digits after the decimal, use only the first digit for the “leaf”
- e.g. 2.42 -- 2 is stem, 4 is leaf

C03,04,05-Data Display 41

Stem-and-Leaf Plot of Income data

Stem-and-leaf plot of Income1 N = 100
Leaf Unit = 1000

```

2  1 | 58
6  2 | 0223
30 2 | 5555666677777888999999
(22) 3 | 000111122222222333344
48 3 | 5555556666777788889
29 4 | 00000001111223444
12 4 | 56667889
4  5 | 0
3  5 | 57
1  6 |
1  6 |
1  7 | 2
    
```

C03,04,05-Data Display 43

What do we see?

- 1|5 means 15,000
- Lowest income = \$15,000
- Highest income = \$72,000
- Most incomes are between \$25,000 and \$49,000
- Highest income is far from the rest -- distribution has a long right-tail or “skewed” right

C03,04,05-Data Display 44

Homework for C04

(to be turned-in in C05: 1/18/02)

DO:

- Using data in problem 2.10 in the textbook:
- 1. (a) Draw a stem-and-leaf plot by hand for Diesel; use leaf unit =.10. Identify variable (Quan or Qual) and type (Discrete, Cont, etc.); Find: smallest, largest and the typical fuel cost for diesel trucks.
- (b) Studying the plot, briefly comment on the distribution of fuel cost for diesel trucks (S, T, S, O?). How could this information be useful? Write a few sentences!

NOTE: Use the homework template from web site to turn-in your solution! Save the template, open it in Word, leave spaces where you draw the plots by hand but you must type your answers. Must be your own work or penalties can apply. Bring your solution to class C05 to be turned in; You might be called to discuss your answer. Page limit: one

READ: Notes for C03-C05 and section 2.3 and 2.4 for next class

Graphs

- Bar Chart
 - Pie Chart
- } → for a single qualitative variable



Rule of Thumb:

A variable that classifies people or things, is qualitative or categorical.

Ex: sex, political affiliation, race, country of origin.

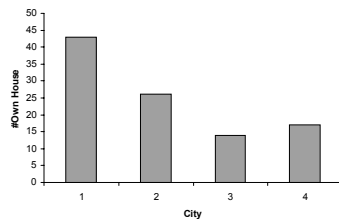
Student Survey Data

sex	height	shoesize	haircut	intended major	smoke	job
0	70	10	\$ 12.00	Marketing	1	0
1	68	10	\$ 35.00	Secondary Education	1	1
1	64	7.5	\$ 25.00	Undecided	1	0
0	73	12	\$ 15.00	MIS	0	0
0	72	11	\$ 12.00	Civil Engineering	0	0
0	75	12	\$ 15.00	International Business Management	1	1
0	69	9	\$ 14.00	Marketing	1	0
0	69	8.5	\$ 14.00	MIS	0	1
1	65	6.5	\$ 95.00	Marketing	0	0
0	72	11	\$ 20.00	Economics	1	0
0	68	10	\$ 25.00	Accounting	0	0
0	74	11.5	\$ 7.00	Undecided	0	0
1	64	8	\$ 35.00	Accounting	0	1
0	72	13	\$ 25.00	Accounting	1	0
0	68	9	\$ 15.00	Accounting	1	1
1	65	8	\$ 36.00	Undecided	0	1
1	62	8.5	\$ 5.00	Mathematics	0	0
1	66	8.5	\$ 30.00	Accounting	1	0
1	64	6	\$ 13.00	Finance	1	1
1	69	8	\$ 30.00	Marketing	1	0

Single qualitative variable

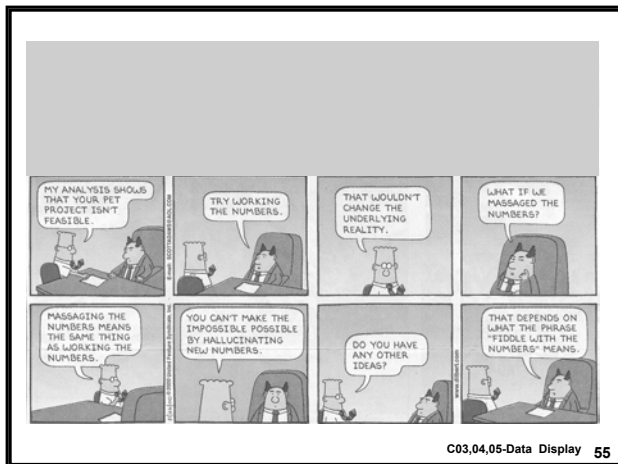
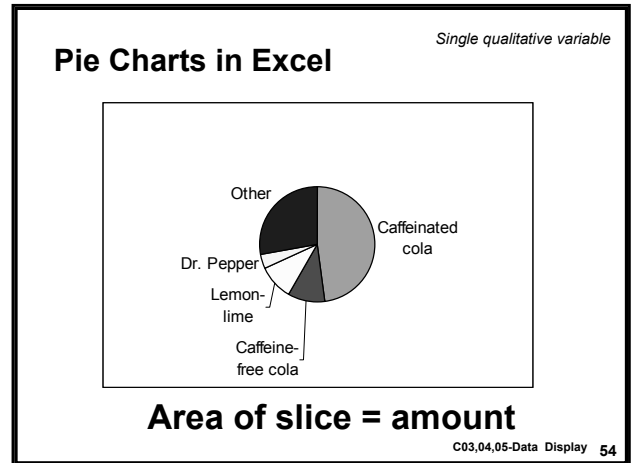
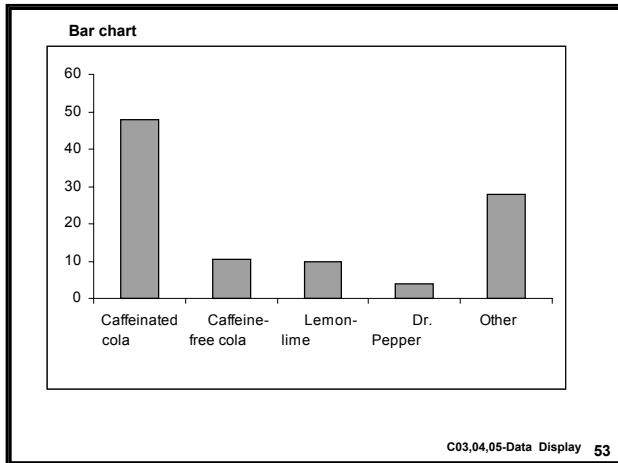
Bar Chart

Sum of Own?	Total
City	
1	35
2	14
3	9
4	7
Grand Total	65



Market share of beverages

<u>type</u>	<u>market share (%)</u>
Caffeinated cola	48.0
Caffeine-free cola	10.4
Lemon-lime	9.8
Dr. Pepper	3.9
Other	27.9

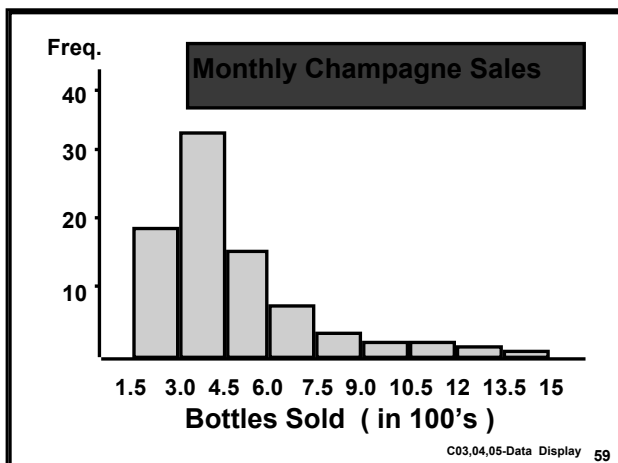


James Bartel runs a liquor store. How many bottles of champagne should he have in inventory each month?

First step: Gather data

Monthly records for number of bottles sold over last seven years.

C03,04,05-Data Display 58



Recall what to look for in a data distribution:

- Shape
- Typical Value
- Spread
- Outliers

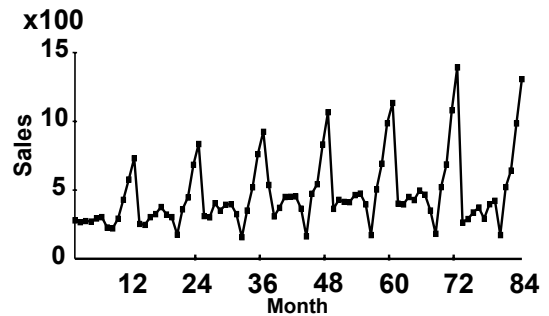
C03,04,05-Data Display 60

Graphs

- Time series plot

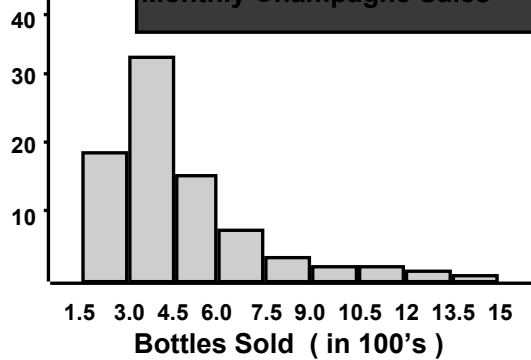
for a single quantitative variable observed over time

Monthly Champagne Sales



Freq.

Monthly Champagne Sales



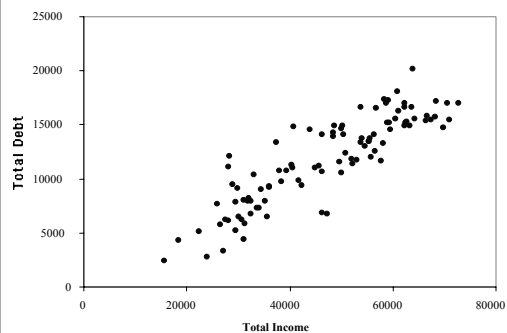
Graphs

- Time series plot

- Scatterplot

→ for two quantitative variables

Scatterplot of Total Debt vs Total Income



Graphs

- Two-way Table

- Stacked Bar Chart

- 3-D Bar Chart

→ for two categorical variables

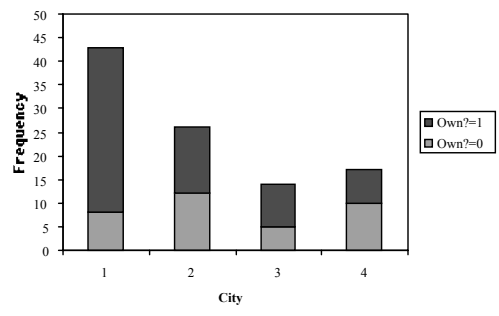
Use Excel

Two-Way Table

Count of City	City				
Own?					

C03,04,05-Data Display 68

Stacked Bar Chart of Own? by City



C03,04,05-Data Display 69

Two-Way Tables, Stacked Bar Charts, and 3-D Bar Charts in Excel

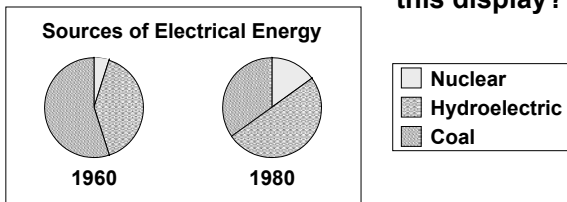
.

C03,04,05-Data Display 71

**Watch Out
for
Bad or Misleading
Data
Displays!**

C03,04,05-Data Display 76

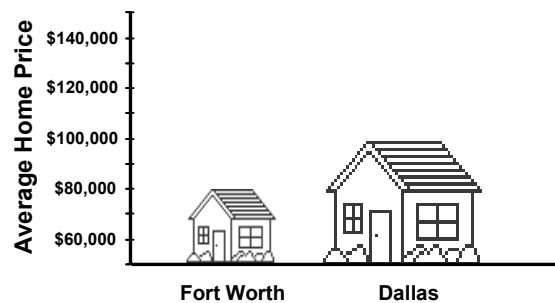
What conclusions do you draw from this display?



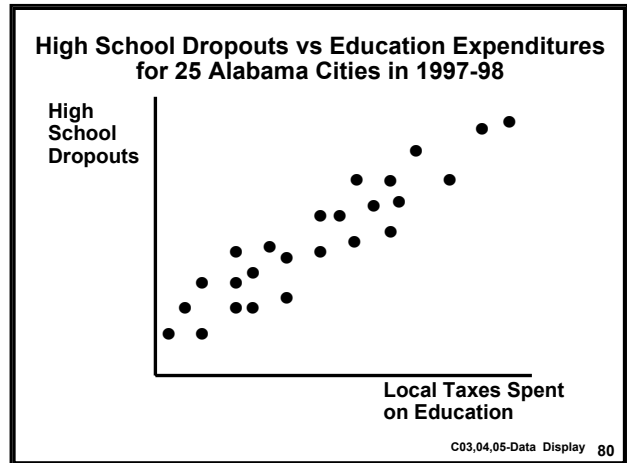
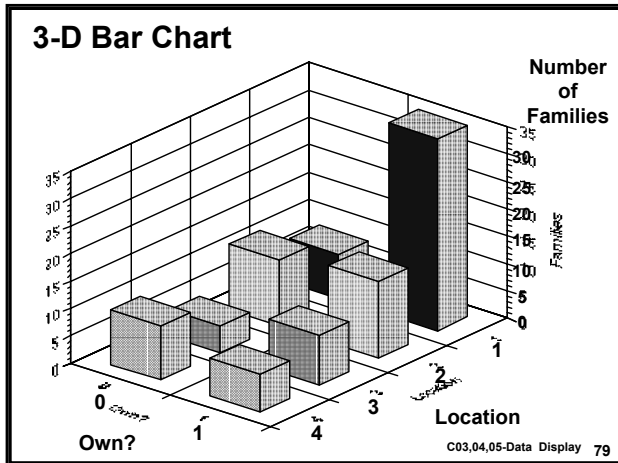
C03,04,05-Data Display 77

Pictogram

Average Home Prices in DFW Metroplex



C03,04,05-Data Display 78



Other Examples of Misleading Statistics

- Chevy Ad
- Chart of Income1 Mean vs City in Minitab
- Safety Institute (DateLine report)
 - Off-center collision testing
 - Auto manufacturers: “This type of collision occurs in fewer than 1% of all car accidents.”
 - Rebuttal: “But 40% of these accidents lead to death. Much higher than for head-on collisions.”

C03,04,05-Data Display 81

Homework for C05 (Not to be turned-in)

1. This is a good point in time to review what we have done so far. *You may begin by studying the questions on various tests, quizzes (see web site) and homework. Can you answer them? Do you feel comfortable with these concepts?*
 For example, to review graphical displays for data, you may begin with answering these questions:
 Why do we use a display? What are some displays out there?
 When do we use which display? Does it matter?
 What does each of the displays show us? S, T, S, O?
 Are there situations where one display might be “better”? For example, do you know when we use a time series plot?
 Given raw data would you be able to draw these displays, by hand, by Excel?
2. Read Examples 2.4 on page 46 and 2.7 on page 51 of your book. Learn terms: relative frequency, density, peaks in a histogram, patterns in the data.
3. READ: Notes for C06 and 2.4 for next class/Print formulas

C03,04,05-Data Display 82