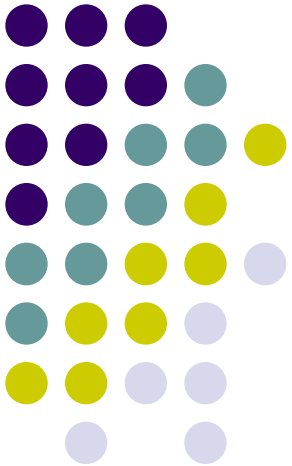
Categorical Data

Frequency Distribution

Bar Graph

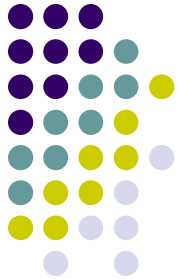
Pi Charts



Warm Up

- 20 volunteers agree to an experiment to test if cars get better gas mileage with premium instead of regular unleaded gasoline.
- Experimental Units: **20 Cars**
- Factor(s): **Type of Gasoline**
- Treatment(s): **1) Premium Gas 2) Regular Gas**
- Response Variable: **Compare Gas Mileage**

Section 1.1
Analyzing Categorical Data

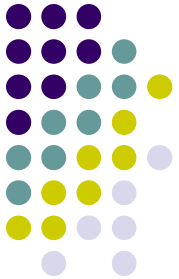


Learning Objectives

After this section, you should be able to…

- ✓ CONSTRUCT and INTERPRET bar graphs and pie charts
- ✓ RECOGNIZE “good” and “bad” graphs
- ✓ CONSTRUCT and INTERPRET two-way tables
- ✓ DESCRIBE relationships between two categorical variables
- ✓ ORGANIZE statistical problems

Categorical Data



- The values for a categorical variable are the labels that you attach to it.

- The distribution will show either the count or the percent of individuals who fall into each category.

- **Categorical Variables** place individuals into one of several groups or categories
  - The values of a categorical variable are labels for the different categories
  - The distribution of a categorical variable lists the **count** or **percent** of individuals who fall into each category.

Example, page 8

Frequency Table

Format	Count of Stations
Adult Contemporary	1556
Adult Standards	1196
Contemporary Hit	569
Country	2066
News/Talk	2179
Oldies	1060
Religious	2014
Rock	869
Spanish Language	750
Other Formats	1579
Total	**13838**

Relative Frequency Table

Format	Percent of Stations
Adult Contemporary	11.2
Adult Standards	8.6
Contemporary Hit	4.1
Country	14.9
News/Talk	15.7
Oldies	7.7
Religious	14.6
Rock	6.3
Spanish Language	5.4
Other Formats	11.4
Total	**99.9**

Variable

Values

Count

Percent

Consistency

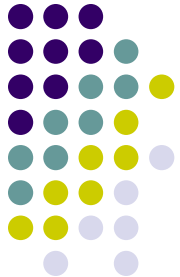
- It's always a good idea to check for consistency.

- Add the counts – make sure it adds to 13,838.

- Add the percents – should add to 100%.

  (Why 99.9%? Roundoff error: the effect of roundoff results)

Frequency Distribution - a listing of how many are in each category.



Article reported on the physical activity patterns in urban women.

W – Walking

T – weight training

C – cycling

G – gardening

A - aerobics

W	T	A	W	G
T	W	W	C	W
T	W	A	T	T
W	G	W	W	C
A	W	A	W	W
W	T	W	W	T

Category	Frequency

Relative Frequency Distribution

Category	Frequency	Relative Frequency
Walking	15	$\frac{15}{30} = 0.5$
Weight Training	7	$\frac{7}{30} = 0.23333$
Cycling	2	$\frac{2}{30} = 0.06667$
Gardening	2	$\frac{2}{30} = 0.06667$
Aerobics	4	$\frac{4}{30} = 0.13333$

Calculator: Put frequency in L1. Go to top of L2 and type L1/Sum(L1).

2nd Stat
Math
Sum

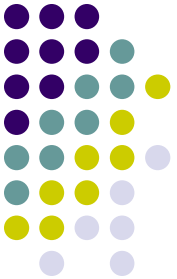
Let's make a Bar Graph

Category	Frequency
Walking	15
Weight Training	7
Cycling	2
Gardening	2
Aerobics	4

There's a special name for Bar Graphs done in DESCENDING ORDER:

Pareto Chart

What type of car do you drive?



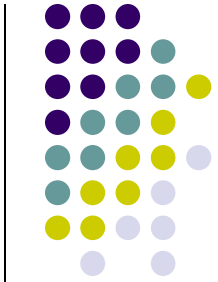
Pie Chart – Circle Graph

Favorite Class	Frequency	Percentage	Degree
Math	66		
English	51		
Science	45		
History	38		

n = 200

GDC: Frequencies in L1
L2: L1/Sum(L1) x 100
L3: L1/Sum(L1) x 360

Picking the Best Graph

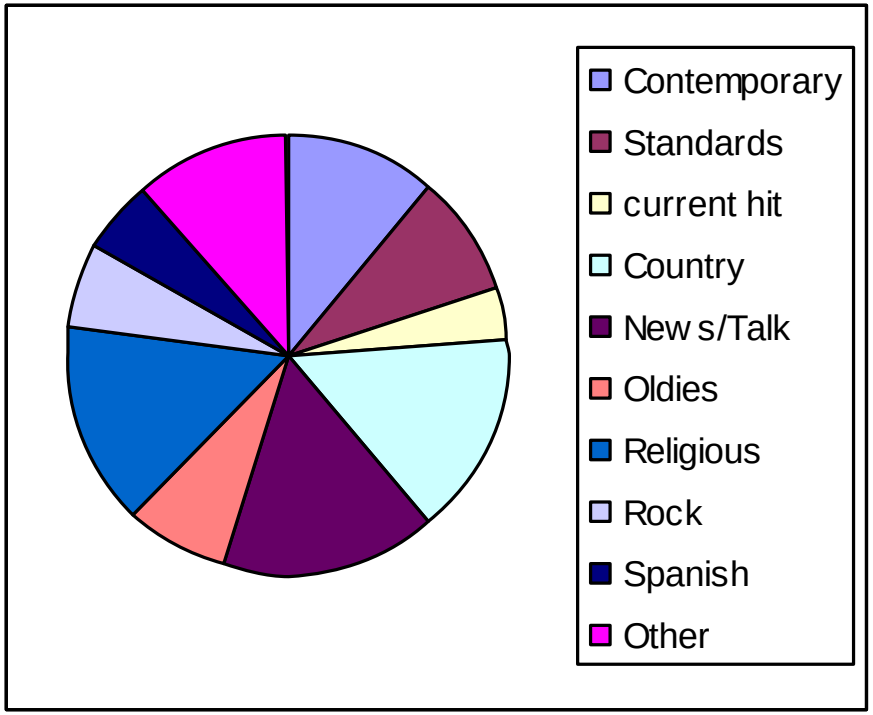


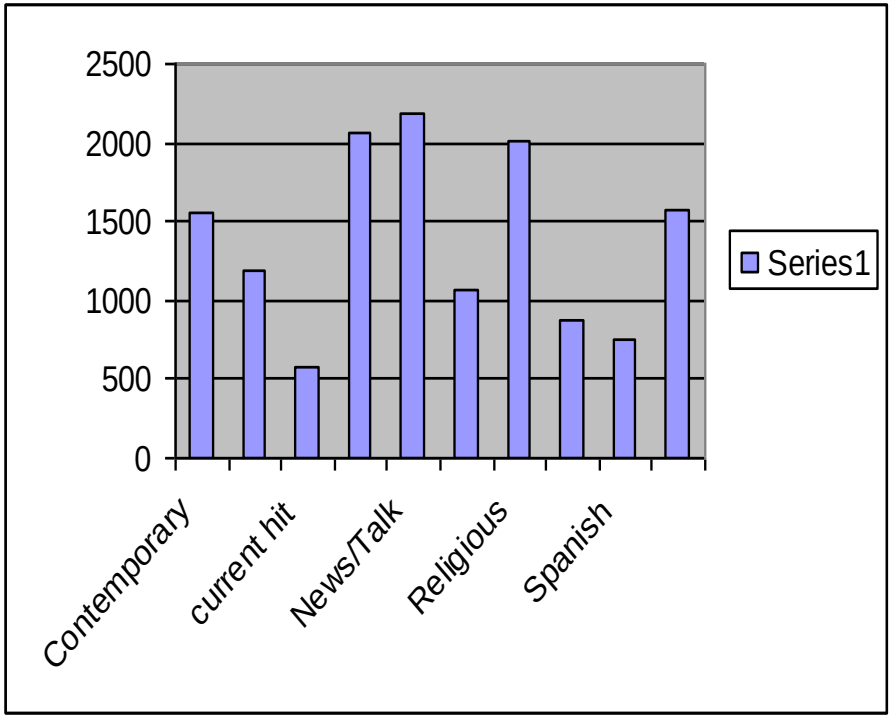
- Do the data tell you what you want to know?
- Does the graph show what you wanted it to show?
- Always think about whether the data you have help answer your questions!

Distribution of Radio Station Formats

Radio Format	# Stations	% of Stations
Contemporary	1556	11.2
Standards	1196	8.6
current hit	569	4.1
Country	2066	14.9
News/Talk	2179	15.7
Oldies	1060	7.7
Religious	2014	14.6
Rock	869	6.3
Spanish	750	5.4
Other	1579	11.4
Total	13838	99.9

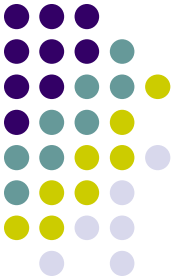
I want to buy radio time to advertise my web site for downloading MP3 music. Which graph is more helpful?





You are more interested in counting listeners – NOT stations. Thus these graphs are not good for our purpose.

Which graph should I use?



- Use a pie graph only when you want to emphasize the category's relation to the whole.

- Otherwise…use a bar graph
  - They are easier to make
  - They are easier to read
  - It's easier to compare categories

Who owns an MP3 Player

Age Group	% own MP3
12 to 17	54
18 to 24	30
25 to 34	30
35 to 54	13
55 and older	5

Make a well labeled bar graph.

Describe what you see.

Who owns an MP3 Player

Age Group	% own MP3
12 to 17	54
18 to 24	30
25 to 34	30
35 to 54	13
55 and older	5

Would it be appropriate to make a pie chart for these data?

It is not appropriate to use a pie chart for this data since each percent in the table refers to a different age group, **not to parts of a single whole.**

Graphs: Good or Bad

- Beware of pictographs….they can appear to be larger than they really are.
  - Page 11 in book

- Watch the scales…it can give a distorted impression of the relative percents.
  - Page 12 in book.

Activity

- You will be given data from our survey yesterday.

- You need to choose the best graph to represent the data you have….
  - Dot Plot
  - Bar Graph
  - Pie Graph

Homework

- Page 22 (9, 11, 13, 15, 17, 18)