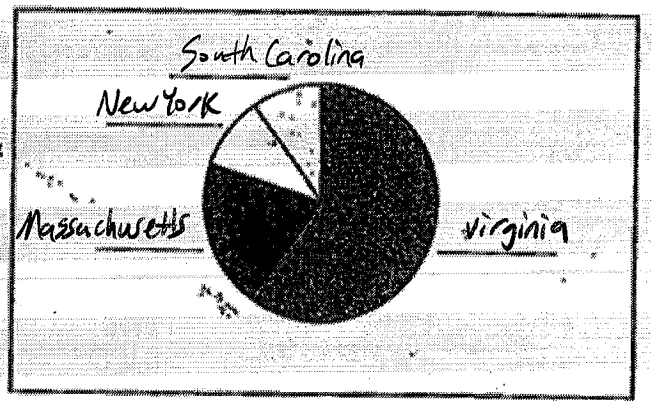
AP Statistics Ch 3 - Extra Practice **Name:** Solutions **Per:** ____

Below is some information about the first ten United States Presidents.

Name	Political Party	Age at Inauguration	Age at Death	State of Birth
George Washington	Federalist	57	67	Virginia
John Adams	Federalist	61	90	Massachusetts
Thomas Jefferson	Democratic-Republican	57	83	Virginia
James Madison	Democratic-Republican	57	85	Virginia
James Monroe	Democratic-Republican	58	73	Virginia
John Quincy Adams	Democratic-Republican	57	80	Massachusetts
Andrew Jackson	Democrat	61	78	South Carolina
Martin Van Buren	Democrat	54	79	New York
William H. Harrison	Whig	68	68	Virginia
John Tyler	Whig	51	71	Virginia

#1. **Identify the variables that were recorded, and indicate whether each one is categorical or quantitative.** political party (categorical) Age at Inauguration (quantitative) State of Birth (categorical) Age at death (quantitative)

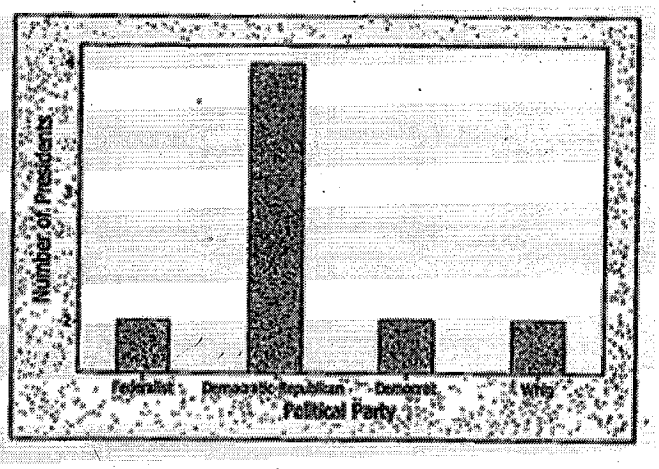
#2. **Here is a pie chart for the distribution of the variable "State of birth." Fill in the blanks with the appropriate values of the variable.**



#3. **Below is a bar graph of the number of presidents of each political party. What is wrong with the way information is presented in this graph?**

the baseline is not zero, which makes everything except democratic-republican look smaller.

(This is another misleading graphing technique, along with area principle)



You suspect that there is a relationship between teenagers' preference in movies and their preference in pizza. You ask 110 students at your school to choose between three movies and three pizza types. Here are your results.

		Pizza Preference			
		Pepperoni	Meatball	Mushroom	
Movie Preference	Men in Black	20	15	10	45
	The Big Lebowski	8	16	11	35
	Monsters, Inc.	15	2	13	30
		43	33	34	110

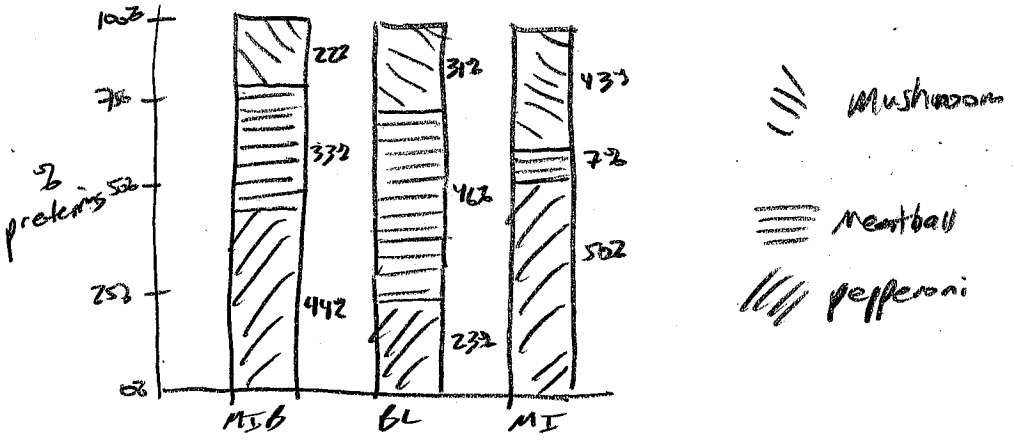
#4. Write the marginal distribution of movie preference (in counts and in percents).

Men in Black	Big Lebowski	Monsters Inc	
45	35	30	/110
(41%)	(32%)	(27%)	

#5. Write the conditional distribution of pizza preference for each movie preference (in counts and in percents).

	Pepperoni	meatball	mushroom	
for Men in Black:	20 (44%)	15 (33%)	10 (22%)	/45
for Big Lebowski:	8 (23%)	16 (46%)	11 (31%)	/35
for Monsters Inc:	15 (50%)	2 (7%)	13 (43%)	/30

#6. Sketch side-by-side segmented bar graphs for the two conditional distributions in #5.



#7. Write a few sentences summarizing what the segmented bar graphs reveal about the association between these variables.

The conditional distributions of pizza preference are quite different for the different movies:
- Pepperoni: 44% MIB vs 23% BL vs 50% MI
- Meatball: 33% MIB vs 46% BL vs 7% MI
- Mushroom: 22% MIB vs 31% BL vs 43% MI

With differences in percentage as high as 26% (meatball), this data suggest that pizza preference depends upon movie preference or is not independent of movie preference or there is an association

18. Politics. Students in an Intro Stats course were asked to describe their politics as "Liberal," "Moderate," or "Conservative." Here are the results:

		Politics			
		L	M	C	Total
Sex	Female	35	36	6	77
	Male	50	44	21	115
	Total	85	80	27	192

a) What percent of the class is male?
b) What percent of the class considers themselves to be "Conservative"?
c) What percent of the males in the class consider themselves to be "Conservative"?
d) What percent of all students in the class are males who consider themselves to be "Conservative"?
e) What percent of all females in the class are "Liberals"?
f) What percent of all males in the class are "Liberals"?
g) Do *politics* and *sex* appear to be independent?

a) $P(\text{male}) = \frac{115}{192} = 60\%$

b) $P(\text{conservative}) = \frac{27}{192} = 14\%$

c) $P(\text{conservative} \mid \text{male}) = \frac{21}{115} = 18\%$ (like a conditional probability)

d) $P(\text{male} \underset{\text{"AND"}}{\cap} \text{conservative}) = \frac{21}{192} = 11\%$

e) $P(\text{liberal} \mid \text{female}) = \frac{35}{77} = 45.5\%$

f) $P(\text{liberal} \mid \text{male}) = \frac{50}{115} = 43.5\%$

g) $P(\text{conservative} \mid \text{male}) = \frac{21}{115} = 18\%$

$P(\text{conservative} \mid \text{female}) = \frac{6}{77} = 8\%$

10% difference is not that large, so it appears politics and sex are independent.

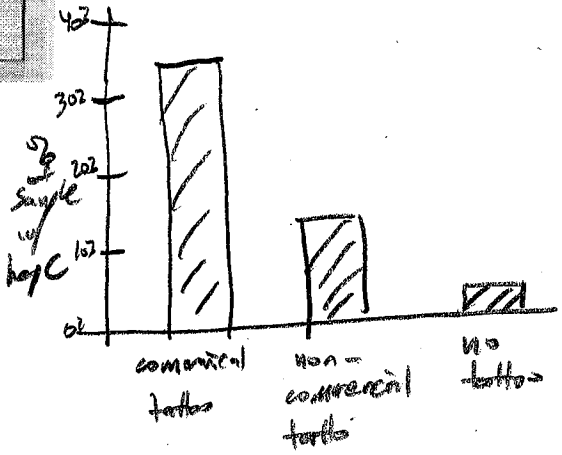
26. Tattoos. A study by the University of Texas Southwestern Medical Center examined 626 people to see if there was an increased risk of contracting hepatitis C associated with having a tattoo. If the subject had a tattoo, researchers asked whether it had been done in a commercial tattoo parlor or elsewhere. Write a brief description of the association between tattooing and hepatitis C, including an appropriate graphical display.

	Tattoo done in commercial parlor	Tattoo done elsewhere	No tattoo
Has hepatitis C	17	8	18
No hepatitis C	35	53	495

$P(\text{hep C} \mid \text{commercial}) = \frac{17}{52} = 32.7\%$

$P(\text{hep C} \mid \text{non commercial}) = \frac{8}{61} = 13.1$

$P(\text{hep C} \mid \text{no tattoo}) = \frac{18}{513} = 3.5\%$



There is a very large difference in percentage of people with hepatitis C in the different categories, with people having no tattoos with lowest (3.5%) and people w/ ~~tattoos~~ from commercial parlors highest (32.7%) There appears to be a strong association between having a tattoo (and where you get it) and hepatitis C.

37. Hospitals. Most patients who undergo surgery make routine recoveries and are discharged as planned. Others suffer excessive bleeding, infection, or other postsurgical complications and have their discharges from the hospital delayed. Suppose your city has a large hospital and a small hospital, each performing major and minor surgeries. You collect data to see how many surgical patients have their discharges delayed by postsurgical complications, and find the results shown in the following table.

	Discharge Delayed	
	Large hospital	Small hospital
Major surgery	120 of 800	10 of 50
Minor surgery	10 of 200	20 of 250

a) Overall, for what percent of patients was discharge delayed?

b) Were the percentages different for major and minor surgery?

c) Overall, what were the discharge delay rates at each hospital?

d) What were the delay rates at each hospital for each kind of surgery?

e) The small hospital advertises that it has a lower rate of postsurgical complications. Do you agree?

f) Explain, in your own words, why this confusion occurs.

a) $P(\text{delayed}) = \frac{160}{1300} = 12.3\%$

b) $P(\text{delayed} \mid \text{major}) = \frac{120}{850} = 15.3\%$

$P(\text{delayed} \mid \text{minor}) = \frac{30}{450} = 6.7\%$

yes

c) $P(\text{delayed} \mid \text{large}) = \frac{130}{1000} = 13\%$

$P(\text{delayed} \mid \text{small}) = \frac{30}{300} = 10\%$

d) Large: $P(\text{delay} \mid \text{major}) = \frac{120}{800} = 15\%$

$P(\text{delay} \mid \text{minor}) = \frac{10}{200} = 5\%$

Small: $P(\text{delay} \mid \text{major}) = \frac{10}{50} = 20\%$

$P(\text{delay} \mid \text{minor}) = \frac{20}{250} = 8\%$

e) No, for both major and minor surgeries, the small hospital's delay rate is higher than at the large hospital.

f) This is due to Simpson's Paradox where the reverse conclusion is drawn when data from multiple categories is aggregated. The smaller hospital overall rate is lower because they mostly perform minor surgeries which generally have lower delay rates.