FREQUENCY BAR GRAPHS AND HISTOGRAMS:
ANSWERS AND DISCUSSION

1. (a) About **11-12%**

   (b) About **15%**

   (c) About 18% (36-40) + 17% (31-35) + 10% (-30) = **45%**

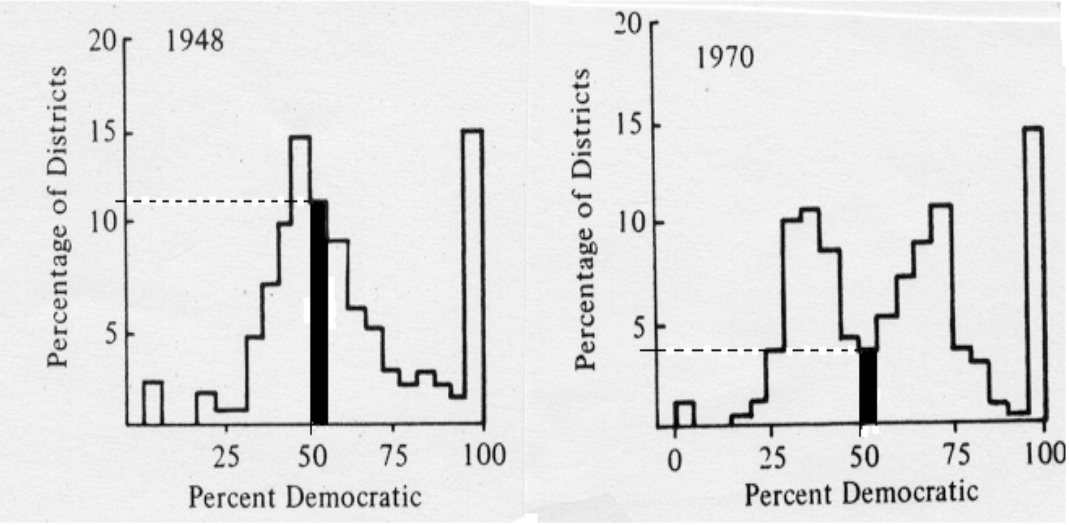
   (d) About 12% (36-40) + 8% (31-35) + 2% (-30) = **22%**

   (e) About 7% (61-65) + 2% (65-) = **9%**

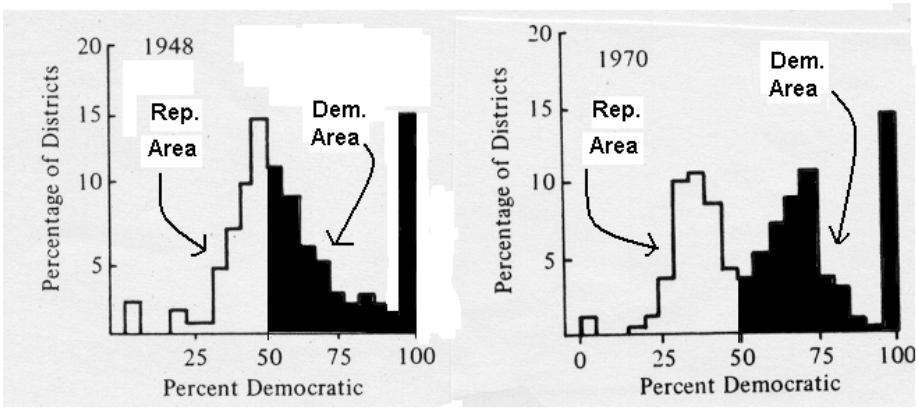
   (f) About 11% (61-65) + 8% (65-) = **19%**

   (g) Clearly the age distribution *shifted upward* substantially over the 20-year period. (Every measure of central tendency has increased.) The earlier faculty age distribution probably was the atypical one historically, as it reflected the great expansion of higher education in this country in the 1960s and the relatively vast hiring of new (mostly young) faculty members in that decade. Twenty years later that "junior faculty boom" became, inevitably, a "middle-aged faculty boom."

2A. About **12%** in 1948, about **4%** in 1970. (See heights of darkened bars below. Note that this is an equal-interval histogram, so the area of each bar is determined entirely by its height, and it is possible to provide a vertical frequency scale.)



2B. Democrats won every district in which Percent Democratic exceeded 50%; Republicans won every district in which Percent Democratic fell below 50%. If we compare the "Democratic area" of the histogram (darkened below) with the "Republican area" (undarkened), it is reasonably evident that the Democratic area is larger. In particular, in both years the Democrats appear at least to match the Republicans in districts won even before we consider the approximately 15% of the districts in which Democrats were apparently unchallenged (got 95-100% of the vote).

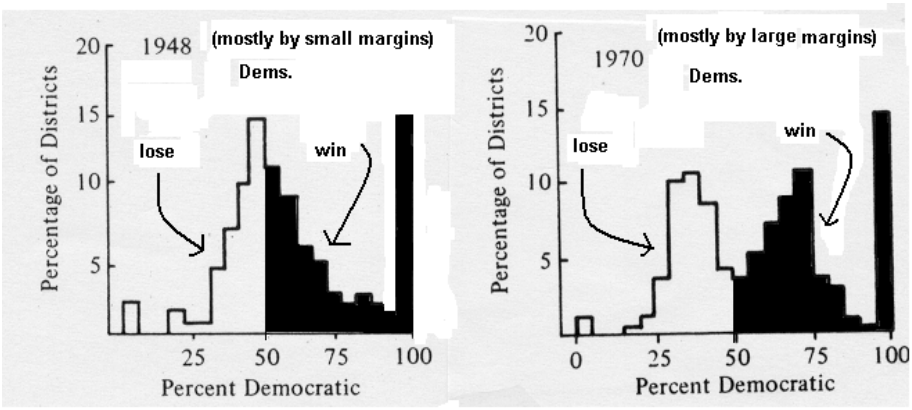


To make a more precise comparison, we can read (off of the vertical relative frequency scale) the height of each of the 20 individual bars and then add the heights of all bars in the Republican and Democratic areas respectively.

		Relative Frequencies			
	Percent Dem.	1948		1970	
	0- 5	2.5%		1.5%	
	5-10	0.0%		0.0%	
Rep.	10-15	0.0%		0.0%	
	15-20	2.0%		0.5%	
Area	20-25	1.0%	Total	1.5%	Total
	25-30	1.0%	42.5%	4.0%	40.5%
	30-35	5.0%		10.0%	
	35-40	7.0%		10.5%	
	40-45	10.0%		8.0%	
	45-50	15.0%		4.5%	
	50-55	12.0%		4.0%	
	55-60	9.0%		5.0%	
	60-65	6.0%		7.0%	
Dem.	65-70	5.0%	Total	9.0%	Total
	70-75	2.5%	57.5%	10.5%	59.5%
Area	75-80	2.0%		4.0%	
	80-85	2.5%		3.5%	
	85-90	2.0%		1.0%	
	90-95	1.5%		0.5%	
	95-100	15.0%		15.0%	
		100.0%		100.0%	

It should be noted, however, that these histograms cannot answer the question completely, because they depict only the population of congressional districts *with incumbents running for re-election* (and exclude the several dozen "open seats" each year). In fact (as you can verify by checking appropriate records), the Democrats won 263 House seats in 1948 (60.6% of the two-party total) and 254 in 1972 (58.5%).

2C. Leaving aside the 15% of seats that the Democrats won each year with 95-100% of the votes (Democratic incumbents without Republican challengers), the remaining frequency distribution in 1948 is an approximately *normal* ("bell-shaped" around a single peak) with most cases clustered in the middle and their frequency declining as move away from the center. But by 1970, the distribution has two distinct peaks, one in the 30-45% Democratic range and the other in the 65-75% Democratic range. Put substantively, in both years, Democrats won a majority of the contests, but in 1948 they won contested seats mostly by small margins and in 1970 mostly by large margins. At the same time, while only a minority of contests in 1970 were won by Republicans, those Republicans that did win won mostly by large margins, while in 1948 most Republican victories also were narrow. (These "most famous" histograms were early evidence of the *increasing electoral safety of Congressional incumbents*, both Republican and Democrats.)

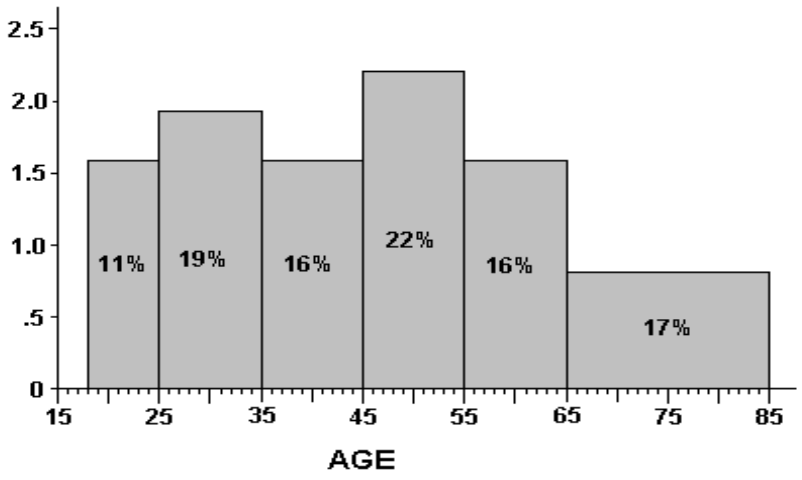


3. If the hypothesis that lack of competition for House seats is mainly due to gerrymandering were true, the Presidential vote in Congressional Districts should be equally non-competitive, i.e., the two histograms should look about the same. But they look very different. The 2000 House vote histogram is bimodal and looks very much like the 1970 House vote histogram in the previous question, except that the greater part of the area is now on the left (Republican) side of 50% mark (and now there are fewer Democrats and more Republicans without challengers). On the other hand, the 2000 Presidential vote histogram (like the 1948 Congressional one) is much more concentrated in the more or less competitive range (with an interesting asymmetry on either side of its peak).

4. See the histogram below, which is derived from the table immediately below. The *width* of each bar corresponds to the *width of the class interval* it represents. The *area* of each bar corresponds to the *frequency* of cases within that class interval. Thus the *height* of each bar corresponds to the "density" of cases within the interval, i.e., the frequency (area) divided by the width of the interval. DISPLAY SPSS FREQUENCIES ROUNDING OF CASE NUMBERS

V60 (AGE)

Class Interval	Abs. Freq.	Rel. Freq.	Adj.Rel. Freq.	Freq/Width = Height ("Density")
18-24	259	11.1	11.2	11.2 / 7 = 1.6
25-34	428	18.4	18.6	18.6 / 10 = 1.9
35-44	368	15.8	16.0	16.0 / 10 = 1.6
45-54	501	21.6	21.7	21.7 / 10 = 2.2
55-64	358	15.4	15.5	15.5 / 10 = 1.6
65-85	391	16.8	17.0	17.0 / 20 = 0.85
NA	20	.9	MISSING	
TOTAL	2325	100.0	100.0	



Note 1. Every respondent in the sample was coded in some age category, including respondents between ages 24 and 25, 34 and 35, etc. Presumably the "17-24" category actually includes all respondents who have not yet observed their 25th birthdays, and likewise for the other categories. This assumption is reflected in drawing the class intervals in the horizontal scale above. In any event, there should be no space between the bars/rectangles, which should "touch."

Note 2. The lowest age category is shown in the Codebook as 17-24. Since the NES is a sample of the *voting age population* and the minimum voting age is 18, I assume that age data

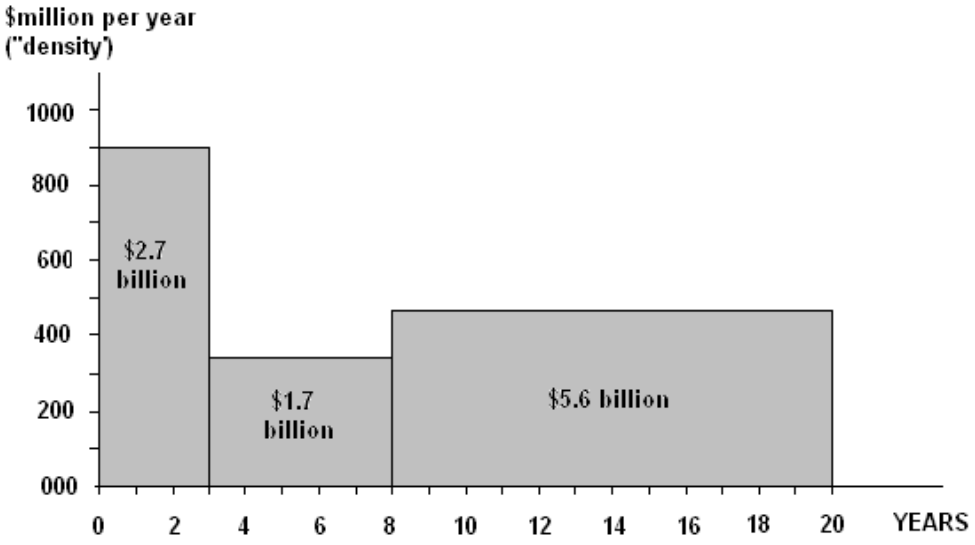
was collected in the pre-election interview (of the two-wave NES panel survey) and included a handful of respondents who would turn 18 between the time of the interview and election day. Thus any "17 year-old" respondents were actually just a few weeks short of 18. On this basis, I have drawn the lowest class interval in the histogram as 18-25. If you used 17-25, the base of your rectangle would be a just a bit wider and its height would be just bit a bit less (11.2 / 8 = 1.4), but its area, of course, should be just the same.

Note 3. Note that the first and third age categories have bars of the same height even though the first category includes a smaller percent of the population; this is because the age intervals are of different width. The third and fifth categories have bars of the same height because they representation the same percent of the population and the intervals are of the same width.

Note 4. You could just as well determine heights by dividing *absolute frequencies* (or unadjusted relative frequencies) by the width of the interval. The numerical magnitude of the tick marks on the vertical scale you would use in drawing the histogram would change, but the shape of the histogram and the information it conveys would not be affected.

Note 5. Remember that the vertical scale is "scaffolding" — it useful in constructing the histogram but should thereafter be removed.

5. Years 1-3: \$2.7 billion / 3 years = \$900 million per year = } average rate
Years 4-8: \$1.7 billion / 5 years = \$340 million per year = } (or "density")
Years 9-20: \$5.6 billion / 12 years = \$467 million per year = } of spending



6. a. About **2/3** ($10K line is a bit over 2 inches high, $20K line about 3 inches high)

   b. About **1/4** ($60K line is about half an inch high, $30K line about 2 inches high)

   c. **15-25%** (area to left of $15K, as percent of total area under curve)

   d. **40-50%** (area to right of $30K)

   e. **65-75%** (area between $10K and $40K; maybe 10% to left and 20% to right, so about 70% in between)

   f. Very approximately **$65,000** (area to right of $65K equals about 10% of total area under the curve)

   g. Probably between **$25,000 and $30,000** (equal areas on either side)