Variables

- A **variable** is an empirical measurement of a characteristic.
- Every variable has one name and at least two possible values.
  - Name: region. Values: South, Northeast, Midwest, West
  - Name: income. Values: $42,000; $42,001; $16,152

3 Levels of Precision

- Three levels of measurement or “degrees of precision”
  - 1. Interval level variables have values that are the most precise (eg. Income in dollars).
  - 2. Ordinal-level variables have values that are somewhat less precise (eg. approval of disapproval same-sex marriage).
  - 3. Nominal-level variables have the least precise values (eg., gender: male and female) .

Interval Level Variables

- Have values that communicate the exact amount of the characteristic being measured
- Use a widely-recognized scale
- Comparing two units of analysis
  - classify into different values of the characteristic
  - rank by relative amount of the characteristic
  - compute the exact difference between units
- For interval variables, the numeric codes convey the exact quantities of the measured characteristic.
  - Person A is 21 years old. Person B is 22. The values are the ages, in years. The difference in ages is exactly 1 year.

Ordinal Level Variables

- Have values that communicate the relative amount of the characteristic being measured.
  - may use a scale but the scale measures relative amount, not absolute amount
- Comparing units of analysis
  - classify into different values on the characteristic
  - rank by relative amount of the characteristic
- For ordinal variables, the numeric codes represent relative amounts, not absolute amounts.
  - Person A “approves” of same-sex marriage (numeric code 1). Person B “disapproves” (numeric code 2). Person B has *relatively* more disapproval than Person A. But the difference is not 1 unit of approval.

Nominal Level Variables

- Have values that communicate differences
  - may be coded with numbers but the numbers just differentiate between units
- Comparing units
  - Classify into different values on the characteristic
  - Person A is male (numeric code 1). Person B is female (numeric code 2).
- The codes simply represent differences. The numbers have no inherent meaning.
  - not exact amounts of gender
  - not relative amounts of gender
  - only differences in gender

Ordinal scales

- A common practice: summing ordinal items into an *additive* scale
  - Likert scales the most common
  - all items in the scale must measure the same concept
- The similarity of the questions ensures that only one concept is measured.
- The greater the number of questions, the more reliable the scale.

Describing variables

- Any variable can be described by its:
  1. Central tendency
     - “average” value
     - value that best typifies the variable
  2. Dispersion
     - extent to which the units of analysis are distributed across the values of the variable

Measures of Central Tendency

- Mode: The "common-most" value
  - may be used with any level of measurement
  - for nominal variables, the only measure that may be used
- Median: The "middle-most" value
  - 50% of cases fall above / 50% fall below the median
  - the same as the 50^{th} percentile
  - for ordinal or interval variables
- Mean: The arithmetic average
  - add up all the values and divide by the number of cases
  - for interval variables only

Dispersion

- a neglected aspect of description
- In political discussion, we use dispersion to describe variables.
  - "polarization," "consensus," "equality"
- A variable has **maximum dispersion** if the cases are spread evenly across all values of the variable.
- A variable has **minimum dispersion** if the cases all fall into one value of the variable.

Frequency distributions

- 3 (sometimes 4) columns
- Column 1: labels of the variable's values
- Column 2: raw number of cases in each value
- Column 3: percentage of cases in each value
- Column 4: cumulative percentages showing the percentage of cases in or below each value
  - Ordinal and interval variables only

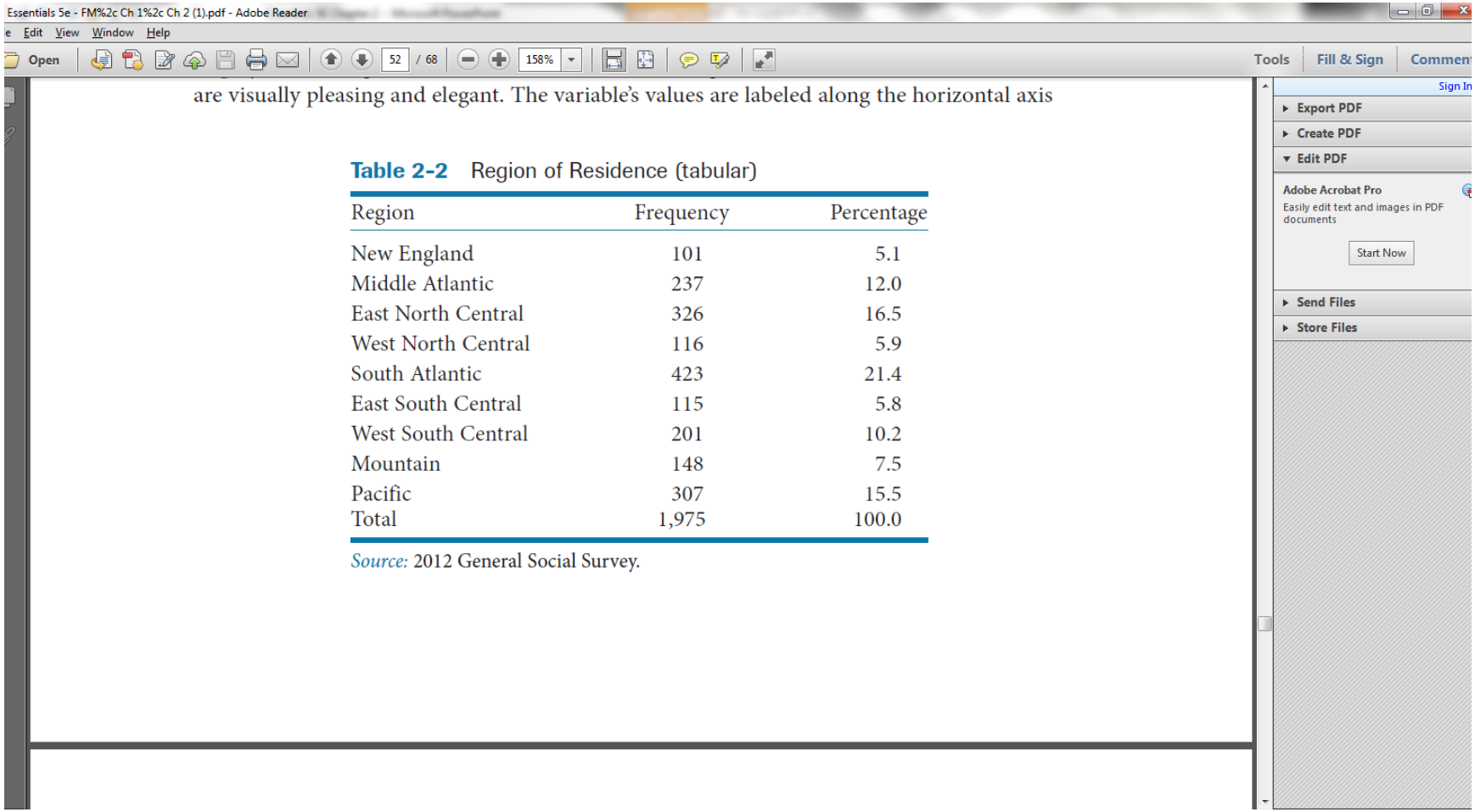
Graphs: Bar charts

- The variable's values appear on the horizontal axis.
- The percentage of cases in each value are recorded on the vertical axis.
- for variables at all levels of measurement

Examples

- Nominal
  - region of residence
- Ordinal
  - religious attendance
  - ideological self-placement
- Interval
  - TV hours
  - support for spending on government programs
  - feeling thermometers

Region of residence (Nominal)



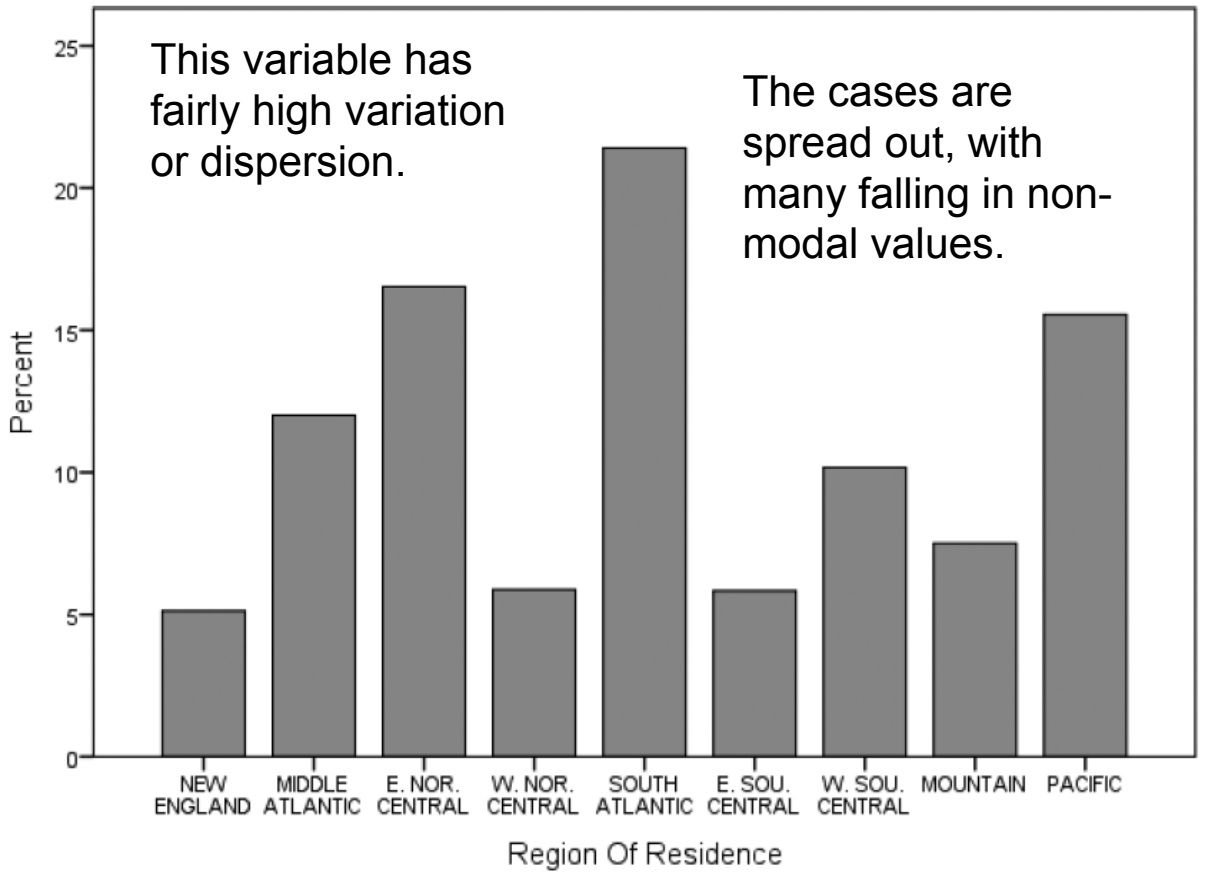
are visually pleasing and elegant. The variable's values are labeled along the horizontal axis

Table 2-2 Region of Residence (tabular)

Region	Frequency	Percentage
New England	101	5.1
Middle Atlantic	237	12.0
East North Central	326	16.5
West North Central	116	5.9
South Atlantic	423	21.4
East South Central	115	5.8
West South Central	201	10.2
Mountain	148	7.5
Pacific	307	15.5
Total	1,975	100.0

Source: 2012 General Social Survey.

Figure 2-2 Region of Residence (graphic)



Religious attendance (Ordinal)

Table 2-3 Attendance at Religious Services (tabular)

Attendance	Frequency	Valid percentage	Cumulative percentage
Never or less than once a year	595	30.3	30.3
Once a year	256	13.0	43.3
Several times a year	213	10.8	54.1
Once a month	133	6.8	60.9
2–3 times a month	174	8.9	69.7
Nearly every week	79	4.0	73.8
Every week or more	516	26.2	100.0
Total	1,967	100.0	

Bimodal

Source: 2012 General Social Survey.

Note: Question: "How often do you attend religious services?"

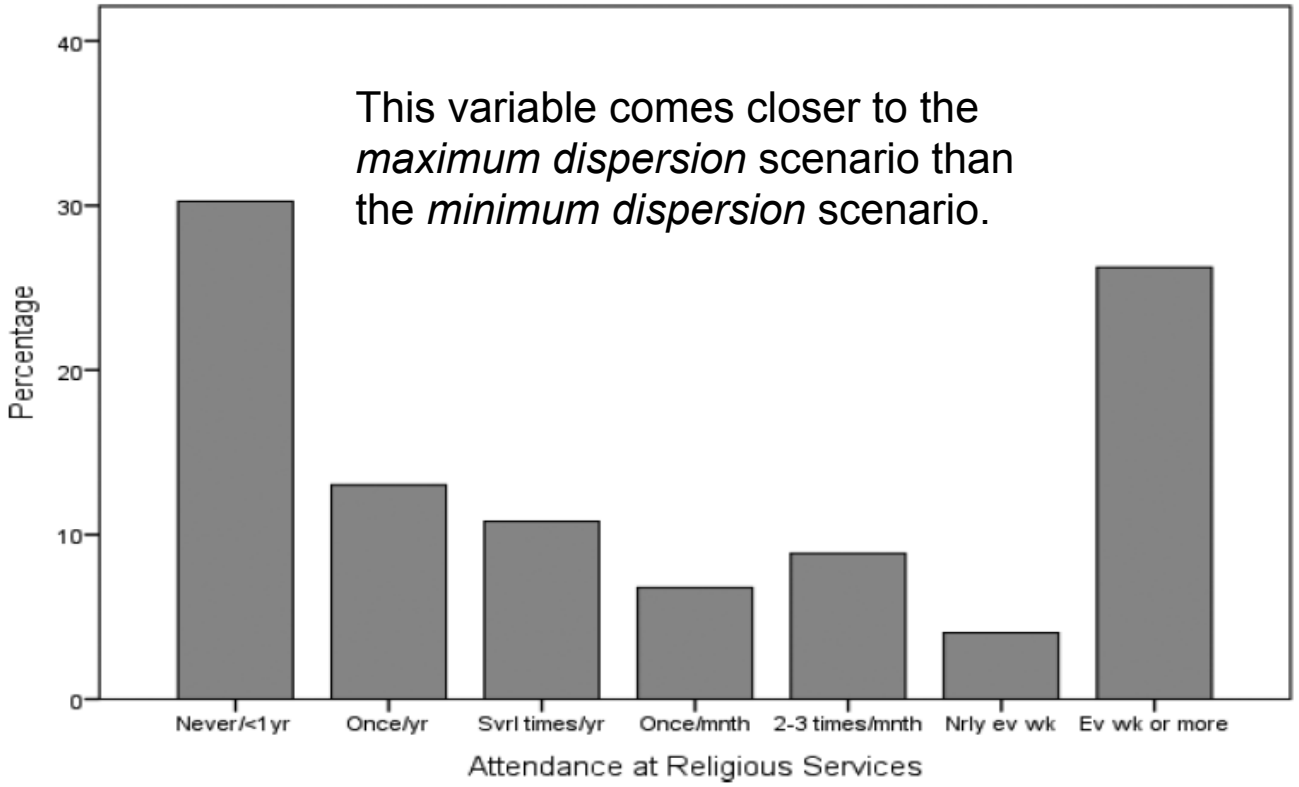
Table 2-3 Attendance at Religious Services (tabular)

Attendance	Frequency	Valid percentage	Cumulative percentage
Never or less than once a year	595	30.3	30.3
Once a year	256	13.0	43.3
Several times a year ← Median	213	10.8	54.1
Once a month	133	6.8	60.9
2–3 times a month	174	8.9	69.7
Nearly every week	79	4.0	73.8
Every week or more	516	26.2	100.0
Total	1,967	100.0	

Source: 2012 General Social Survey.

Note: Question: "How often do you attend religious services?"

Figure 2-3 Attendance at Religious Services (graphic)



Source: 2012 General Social Survey.

Note: Question: "How often do you attend religious services?"

Ideological Self-Placement (Ordinal)

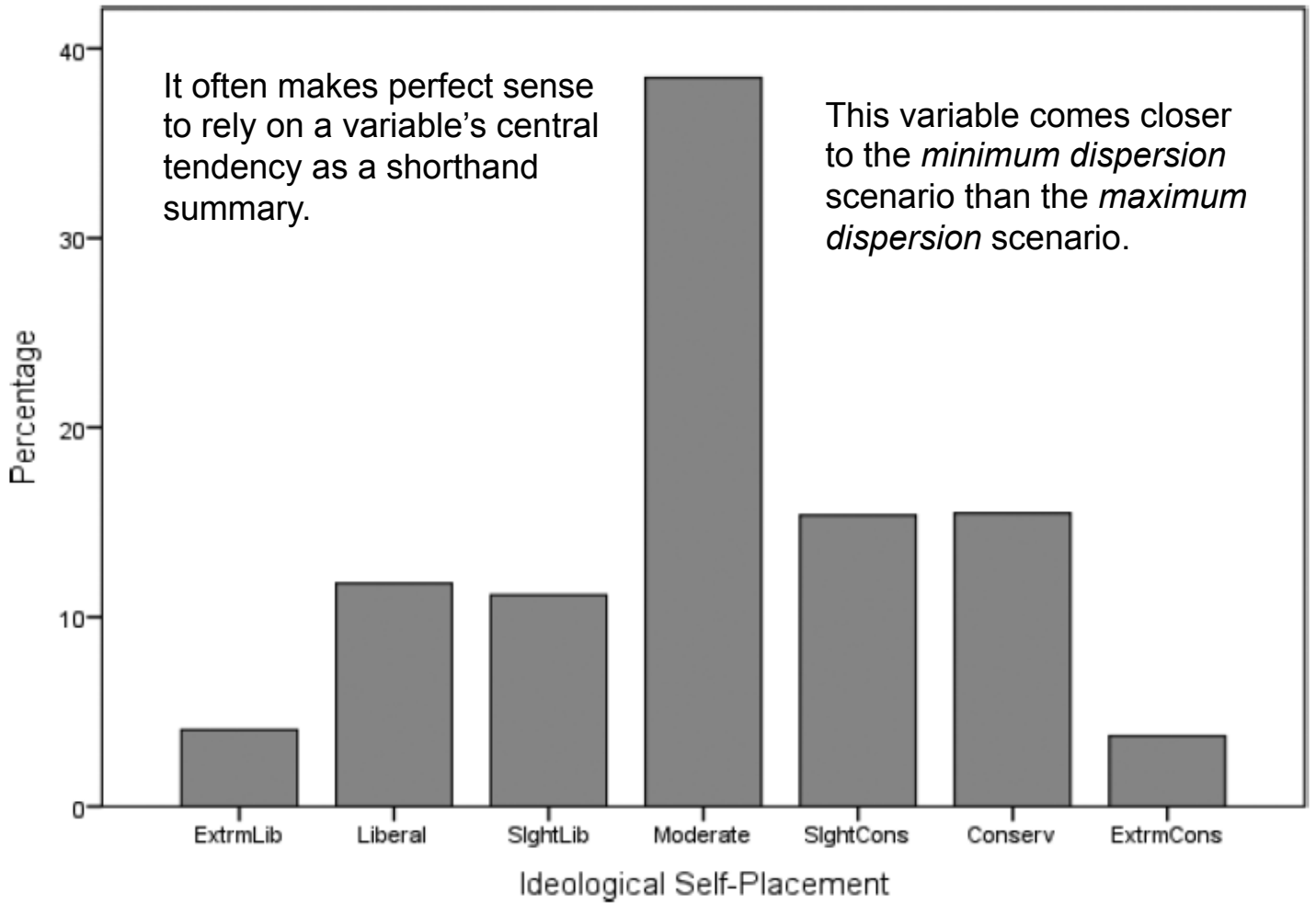
Table 2-4 Ideological Self-Placement (tabular)

Ideology	Frequency	Percentage	Cumulative percentage
Extremely liberal	76	4.0	4.0
Liberal	220	11.8	15.8
Slightly liberal	209	11.2	27.0
Moderate ← Mode = "Moderate" Median = "Moderate"	720	38.5	65.4
Slightly conservative	288	15.4	80.8
Conservative	290	15.5	96.3
Extremely conservative	70	3.7	100.0
Total	1,873	100.0	

Source: 2012 General Social Survey.

Note: Question: "We hear a lot of talk these days about liberals and conservatives. I'm going to show you a 7-point scale on which the political views that people might hold are arranged from extremely liberal (point 1) to extremely conservative (point 7). Where would you place yourself on this scale?"

Figure 2-4 Ideological Self-Placement (graphic)



Source: 2012 General Social Survey.

TV hours (Interval)

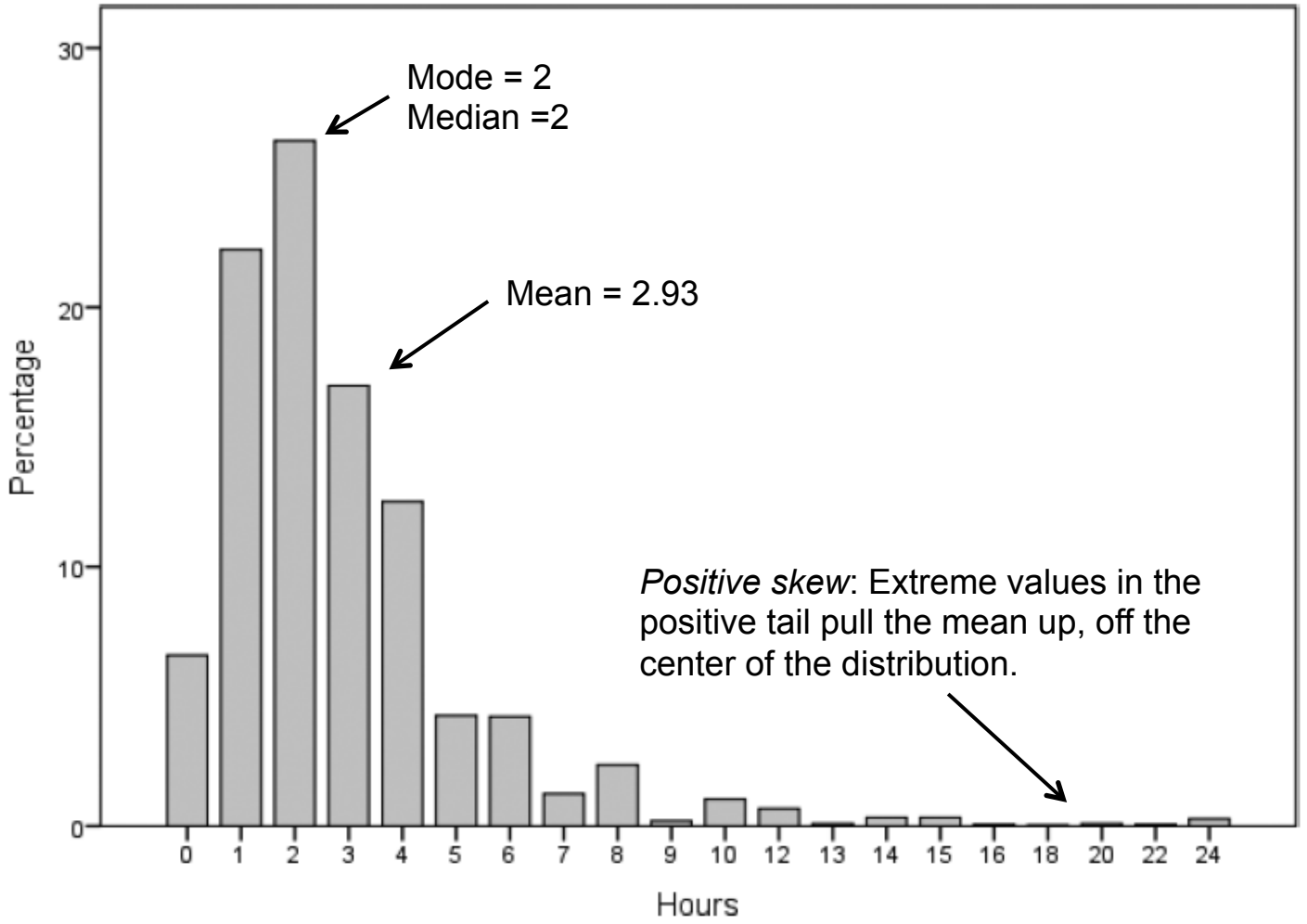
Table 2-5 Hours Watching TV (tabular)

Hours	Frequency	Percentage	Cumulative percentage
0	86	6.6	6.6
1	291	22.2	28.8
2 ← Mode = 2 Median = 2	346	26.4	55.2
3	222	17.0	72.2
4	164	12.5	84.7
5	56	4.3	89.0
6	55	4.2	93.2
7	16	1.2	94.5
8	31	2.4	96.8
9	3	.2	97.0
10	14	1.0	98.0
12	9	.7	98.7
13	1	.1	98.8
14	4	.3	99.1
15	4	.3	99.5
16	1	.1	99.5
18	0	.0	99.6
20	1	.1	99.7
22	1	.1	99.7
24	4	.3	100.0
Total	1,309	100.0	

Source: 2012 General Social Survey.

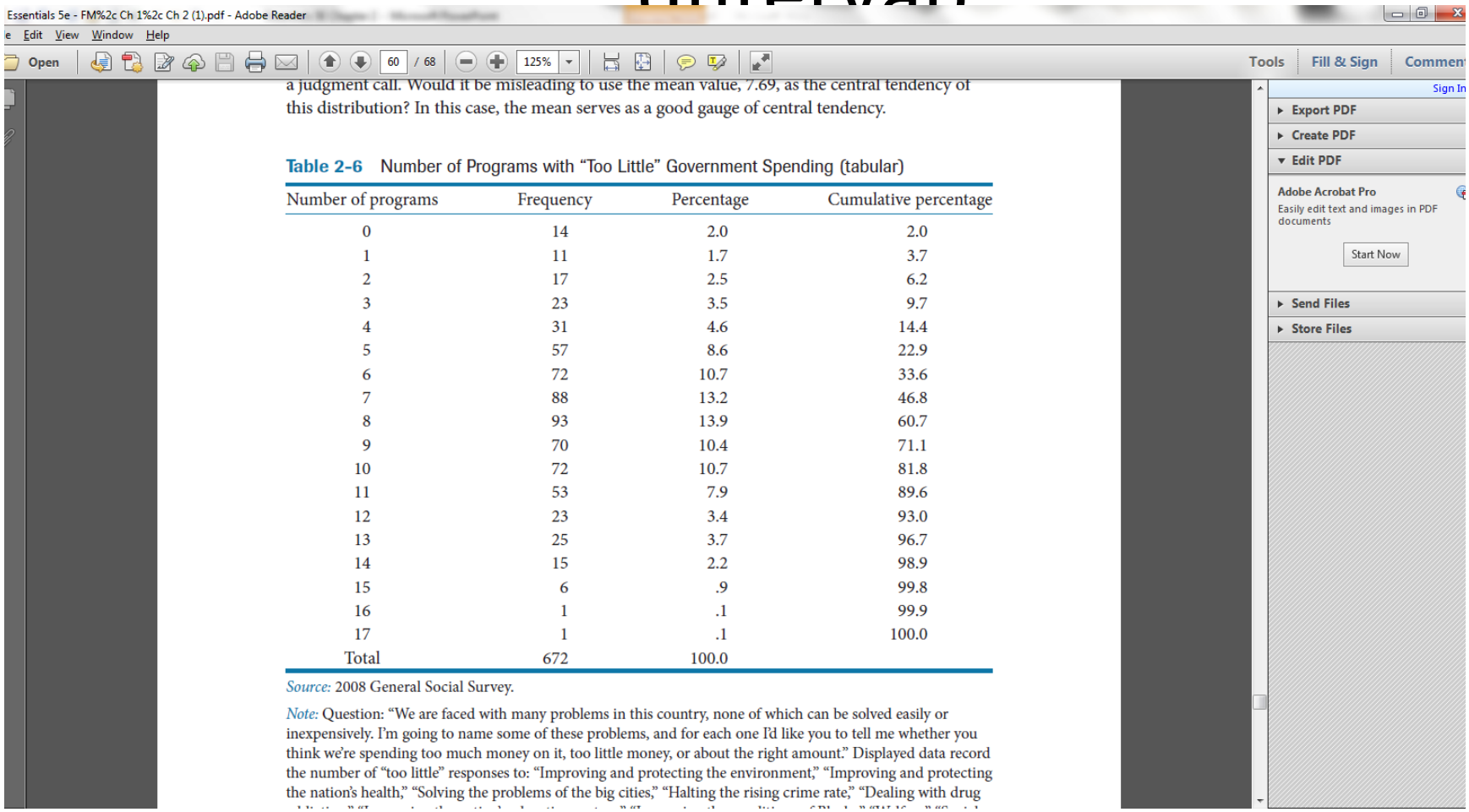
Note: Question: "On the average day, about how many hours do you personally watch television?

Figure 2-5 Hours Watching TV (graphic)



Source: 2012 General Social Survey.

Support for spending on 17 programs (Interval)



a judgment call. Would it be misleading to use the mean value, 7.69, as the central tendency of this distribution? In this case, the mean serves as a good gauge of central tendency.

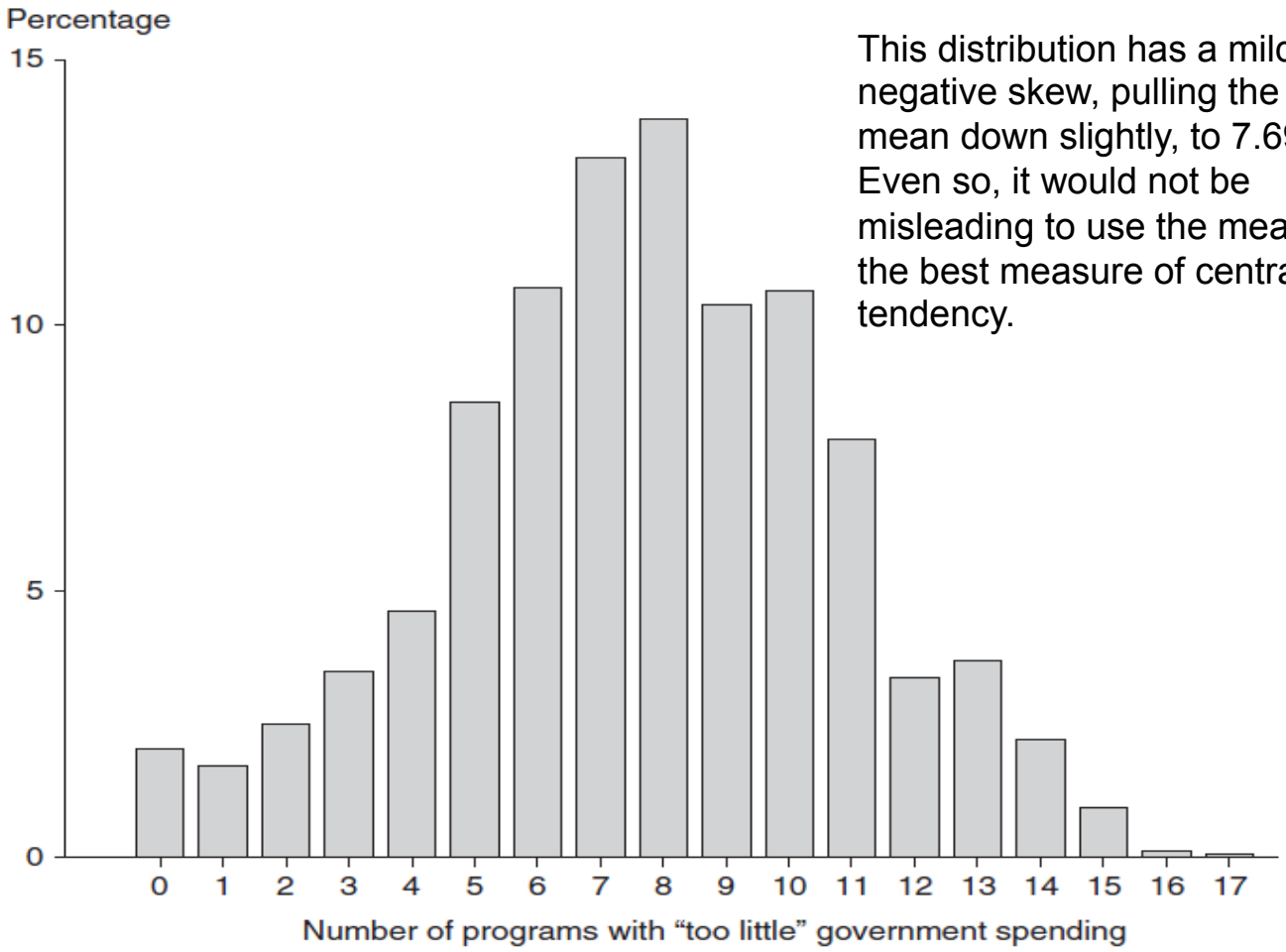
Table 2-6 Number of Programs with "Too Little" Government Spending (tabular)

Number of programs	Frequency	Percentage	Cumulative percentage
0	14	2.0	2.0
1	11	1.7	3.7
2	17	2.5	6.2
3	23	3.5	9.7
4	31	4.6	14.4
5	57	8.6	22.9
6	72	10.7	33.6
7	88	13.2	46.8
8	93	13.9	60.7
9	70	10.4	71.1
10	72	10.7	81.8
11	53	7.9	89.6
12	23	3.4	93.0
13	25	3.7	96.7
14	15	2.2	98.9
15	6	.9	99.8
16	1	.1	99.9
17	1	.1	100.0
Total	672	100.0	

Source: 2008 General Social Survey.

Note: Question: "We are faced with many problems in this country, none of which can be solved easily or inexpensively. I'm going to name some of these problems, and for each one I'd like you to tell me whether you think we're spending too much money on it, too little money, or about the right amount." Displayed data record the number of "too little" responses to: "Improving and protecting the environment," "Improving and protecting the nation's health," "Solving the problems of the big cities," "Halting the rising crime rate," "Dealing with drug

Figure 2-6 Number of Programs with "Too Little" Government Spending (graphic)



This distribution has a mild negative skew, pulling the mean down slightly, to 7.69. Even so, it would not be misleading to use the mean as the best measure of central tendency.

Source: 2008 General Social Survey.

Interquartile range

- non-statistical measure of dispersion for interval (and ordinal) variables
- especially informative when two distributions are being compared.
- Defined as the range of a variable's values that defines the "middle half " of a distribution:
  - between the upper boundary of the lowest quartile (which is the same as the 25th percentile) and the lower boundary of the upper quartile (the 75th percentile)
- graphically depicted by a **box plot**

Feeling thermometers (Interval)

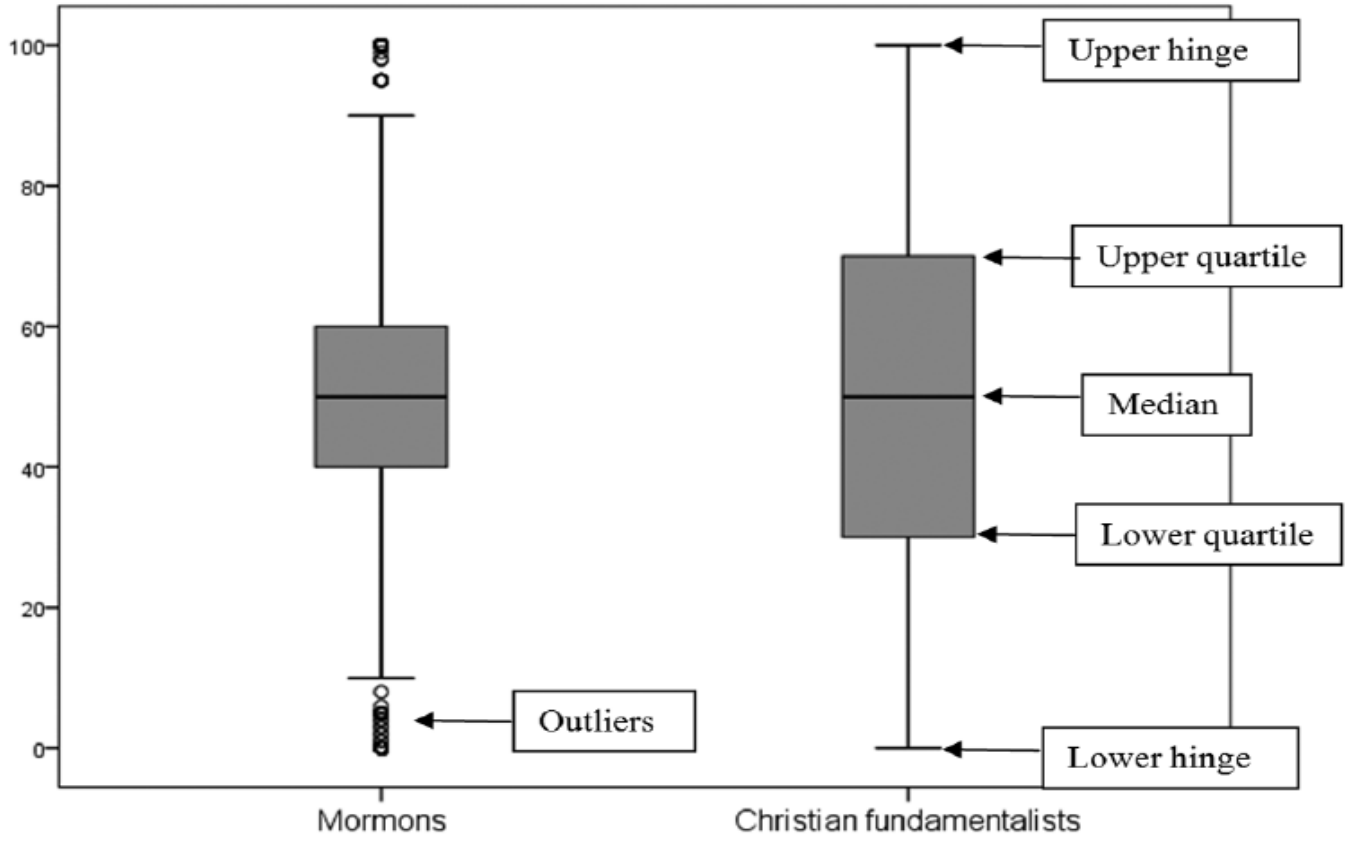
Table 2-7 Summary Information for Two Interval Variables

	Mormons	Christian fundamentalists
Mean	50.5	48.9
Median	50	50
Mode	50	50
Percentiles:		
25	40	30
50	50	50
75	60	70

Source: 2012 American National Election Study.

Box plots

Figure 2–7 Box Plots of Two Interval Variables



Source: 2012 American National Election Study.

Note: Figure annotations based on Robert I. Kabacoff, *R in Action: Data Analysis and Graphics with R* (Shelter Island, N.Y.: Manning, 2011), 133.