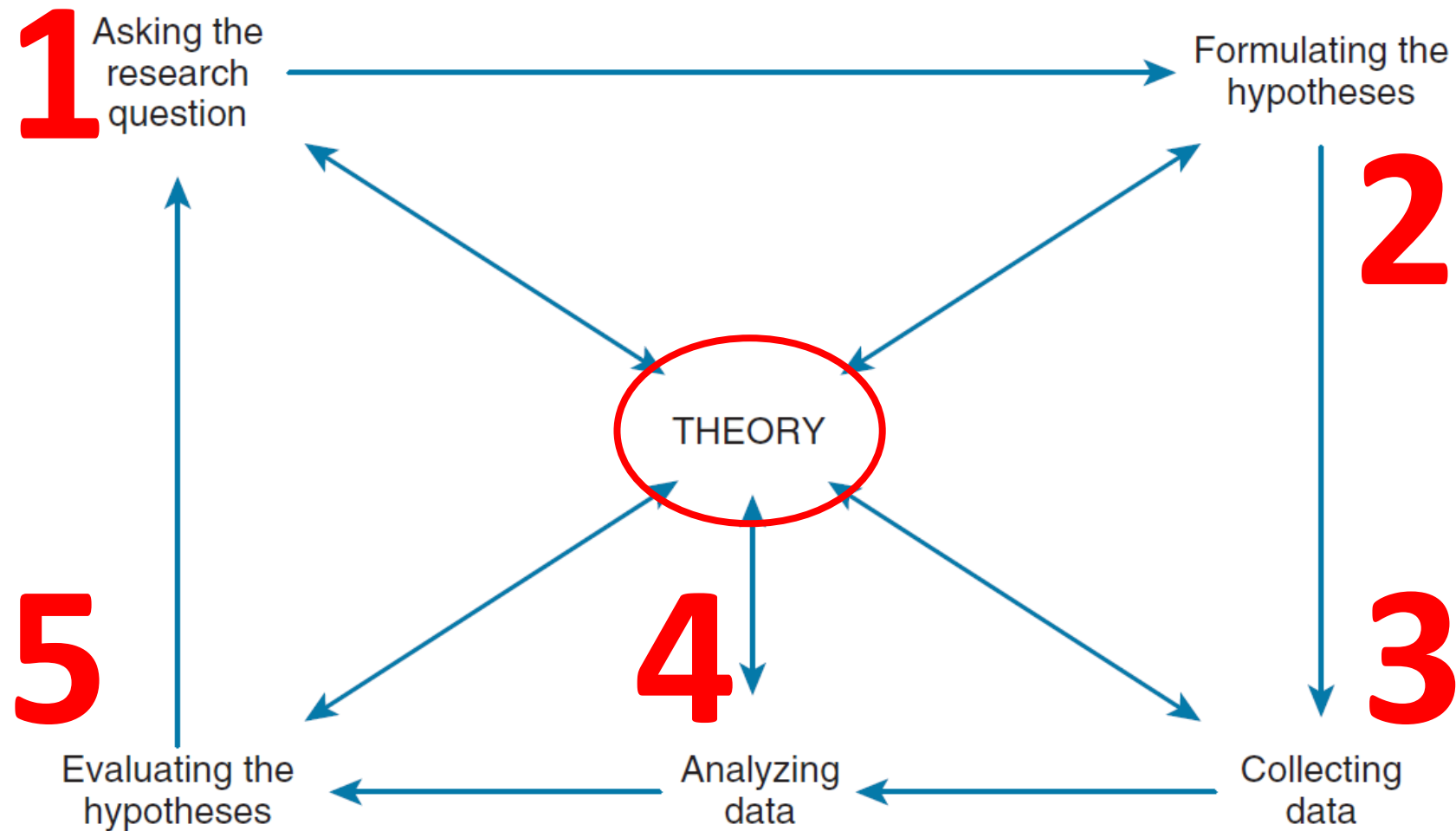
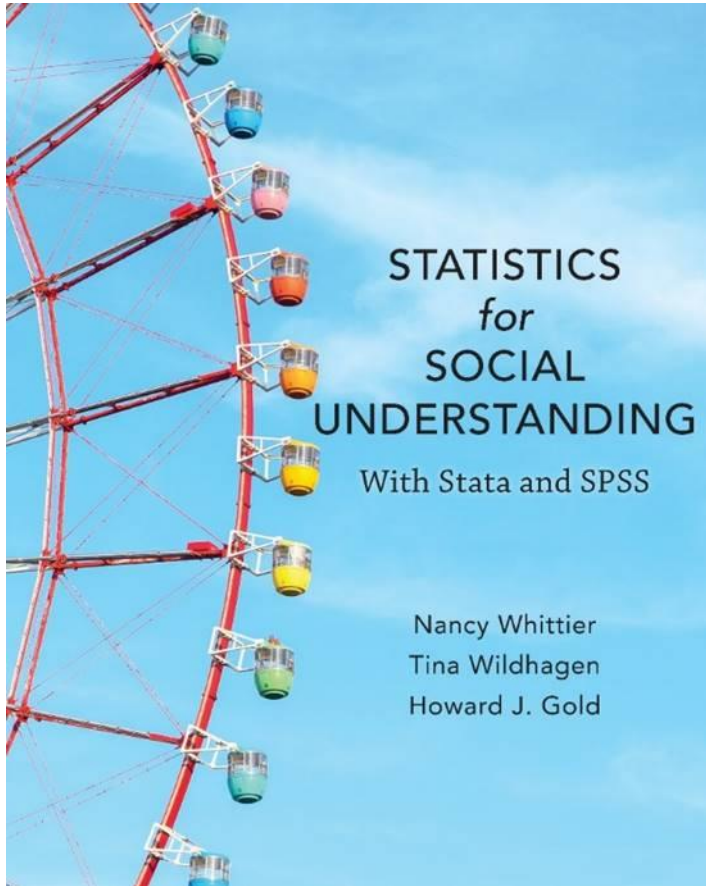


Figure 1.1 The Research Process

Where do you start?





Chapter 2

Getting to Know Your Data: Frequency Distributions and Visual Representations of Data

Frequency Distributions

- **Frequency distributions:**

- Give the number, percentage, or proportion of cases in a sample who fall into each category of a variable.
- Are **univariate statistics** because they describe one variable at a time

- **Percentages:**

- show the size of each response category relative to the overall number of cases
- By formula, percentage is: $\frac{f}{N} (100)$, where f is frequency in the category, and N is the total number of cases
- It is easier to compare the relative sizes of categories using percentages when the number of cases in the categories is very large

Proportion and Percent

$$\text{Proportion} = \left(\frac{f}{N} \right)$$

$$\text{Percentage} = \left(\frac{f}{N} \right) * 100$$

Example: Compare Two Frequency Distributions

Table 2.1 Frequency of Votes for Presidential Candidates, 2016

Candidate	Total Votes
Donald Trump	62,985,134
Hillary Clinton	65,853,652
Other	8,286,698
Total	137,125,484

Source: Dave Leip's Atlas of U.S. Presidential Elections.

Table 2.5 Votes for 2016 Presidential Candidates

Candidate	Frequency	Percent
Donald Trump	62,985,134	45.9
Hillary Clinton	65,853,652	48.0
Other	8,286,698	6.0
Total	137,125,484	100

Source: Dave Leip's Atlas of U.S. Presidential Elections.
Total percent may not sum to 100 due to rounding.

Frequency Distributions for Ordinal Variables

Table 3: “Our Grand Leader is leading us toward the Truth.”

Response	Frequency	Percentage
Strongly agree	17	28
Agree	30	50
Disagree	4	7
Strongly Disagree	9	15
<i>Total</i>	<i>60</i>	<i>100</i>

Source: Fictitious Data

Frequency Distributions for Ordinal Variables


“Our Grand Leader is leading us toward the Truth.”

Response	Frequency	Percentage
Strongly agree or Agree	47	78
Disagree or Strongly Disagree	13	22
<i>Total</i>	<i>60</i>	<i>100</i>

Source: Fictitious Data

Frequency Distributions for Interval-Ratio Variables

Age Distribution of Doomsday Group		
Response	Frequency	Percentage
18	1	2
19	7	12
20	8	13
22	10	17
23	9	15
24	7	12
27	5	8
28	10	17
31	2	3
45	1	2
<i>Total</i>	<i>60</i>	<i>100</i>

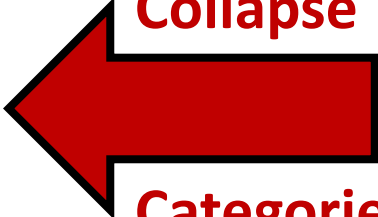
Too many

data points

Source: Fictitious Data

Frequency Distributions for Interval-Ratio Variables

Age Distribution of Doomsday Group

Response	Frequency	Percentage
18 to 22	26	43
23 to 25	16	27
26 to 30	15	25
31 to 40	2	3
41 and older	1	2
<i>Total</i>	<i>60</i>	<i>100</i>

Collapse

Categories

Source: Fictitious Data

Cumulative Percentage and Percentile

- **Cumulative percentage:**

- The total percentage of cases with a given value or below (or a given value and above)
- Example: percentage of people who have two or fewer children
- Only use ordinal or interval-ratio variables because values must be rank-ordered (either from least to most or most to least)
- $\text{Cum\%} = \frac{\text{cumf}}{N} (100)$, where “cumf” stands for cumulative frequency and N is the total number of cases

- **Percentile:**

- The position of any given case or value in the overall distribution
- Example: A child who is in the 50th percentile for height has a height for which the cumulative percentage would be 50. Of the children, 50% are below that height.

A Frequency Distribution with Cumulative Percentage

Table 2.7 Americans Who Say Life Is Exciting, Routine, or Dull (Frequency, Percentage, and Cumulative Percentage)

Rating of Life	Frequency	Percent	Cumulative Percent
Dull	93	5.0	5.0
Routine	835	44.6	49.6
Exciting	945	50.5	100
Total	1,873	100	

Source: 2016 General Social Survey.
Total percent may not sum to 100 due to rounding.

Frequency Distributions with Cumulative Percentage Ranked in Opposite Orders

Table 2.8 Number of Children, Including Cumulative Percentage

Number of Children	Frequency	Percent	Cumulative Percent
0	797	27.9	27.9
1	459	16.1	44.0
2	733	25.6	69.6
3	467	16.3	85.9
4	213	7.5	93.4
5	92	3.2	96.6
6	51	1.8	98.4
7	25	0.9	99.3
8 or more	22	0.8	100
Total	2,859	100	

Source: 2016 General Social Survey.
Total percent may not sum to 100 due to rounding.

Table 2.9 Number of Children, Including Cumulative Percentage (Reverse Order)

Number of Children	Frequency	Percent	Cumulative Percent
0	797	27.9	100
1	459	16.1	72.2
2	733	25.6	56.1
3	467	16.3	30.5
4	213	7.5	14.2
5	92	3.2	6.7
6	51	1.8	3.5
7	25	0.9	1.7
8 or more	22	0.8	0.8
Total	2,859	100	

Source: 2016 General Social Survey.
Total percent may not sum to 100 due to rounding.

Percent Change

- **Percent change:**

- The size of a change over time in comparison to the initial percent or frequency

- Percent change = $\frac{(\text{percent at time 2} - \text{percent at time 1})}{\text{percent at time 1}} (100)$

- Example: 30% of statistics students liked statistics at the beginning of a course, and 95% of them liked it at the end of the course, what is the percent change in students who like statistics?

- Percent change = $\frac{(95 - 30)}{30} (100) = 217\%$

Rates and Ratios

- **Rate:**

- The prevalence of an outcomes of interest in a given population
- The frequency of an occurrence relative to the number of times that the event could have occurred in a given group

- $$\text{Rate} = \frac{f_{\text{actual occurrences}}}{f_{\text{possible occurrences}}}$$

- Example: In 2009, number of divorces: 820,669; number of marriages: 2,080,000;

$$\text{Divorce rate} = \frac{820,669}{2,080,000} = .39, \text{ or } 39\%$$

- **Ratio:**

- Examines the size of one category relative to another
- $$\text{Ratio} = \frac{f_1}{f_2}, \text{ where } f_1 \text{ and } f_2 \text{ are the frequencies for the groups of interest}$$

- Example: In 2016, there were 125,220,000 married adults in the U.S. and 25,511,000 divorced adults; Ratio of married to divorced people:
$$\frac{125,220,000}{25,511,000} = 4.9$$

Activity: Important Rates in Our City

Working with Frequency Distributions: Missing Values and Collapsing Categories

- **Valid percent:**
 - When missing values are excluded from percentages
 - Percentages are larger because denominator is smaller
- **Collapsing categories** (called “recoding” when using software)
 - Collapse (or combine) when fine distinctions between categories are not useful
 - Collapse when there are too many categories in a variable to make sense of its frequency distribution
 - Collapsing categories can change the data’s story

Missing Values Two Ways (Percent and Valid Percent)

Table 2.10 Self-Placement on Ideological Scale

Self-Placement	Frequency	Percent	Valid Percent	Cumulative Percent
1: Extremely liberal	169	2.9	3.1	3.1
2: Liberal	603	10.2	11.2	14.3
3: Slightly liberal	631	10.7	11.7	26.0
4: Moderate	1,859	31.4	34.4	60.4
5: Slightly conservative	842	14.2	15.6	76.0
6: Conservative	1,048	17.7	19.4	95.4
7: Extremely conservative	250	4.2	4.6	100
Total, 1–7	5,402	91.3	100	
Refused	44	.7		
Don't know	18	.3		
Haven't thought much about this	452	7.6		
Total, missing	514	8.7		
Total	5,916	100		

Source: 2012 American National Election Study.
Total percent may not sum to 100 due to rounding.

Ideology Two Ways: Collapsed and Not Collapsed

Table 2.11 Self-Placement on Ideological Scale, Collapsed Version

Self-Placement	Frequency	Percent	Valid Percent	Cumulative Percent
Liberal	1,403	26.0	26.0	26.0
Moderate	1,859	34.4	34.4	60.4
Conservative	2,140	39.6	39.6	100
Total	5,402	100	100	

Source: 2012 American National Election Study.

Table 2.12 Party Identification, November 2016

Party Identification	Percent
Strong Democrat	25
Not very strong Democrat	9
Independent who leans Democrat	9
Strictly Independent	12
Independent who leans Republican	10
Not very strong Republican	6
Strong Republican	22
Other (Vol.)	5
Not sure	2
Total	100

Source: NBC News/Wall Street Journal poll.

Bar Graph

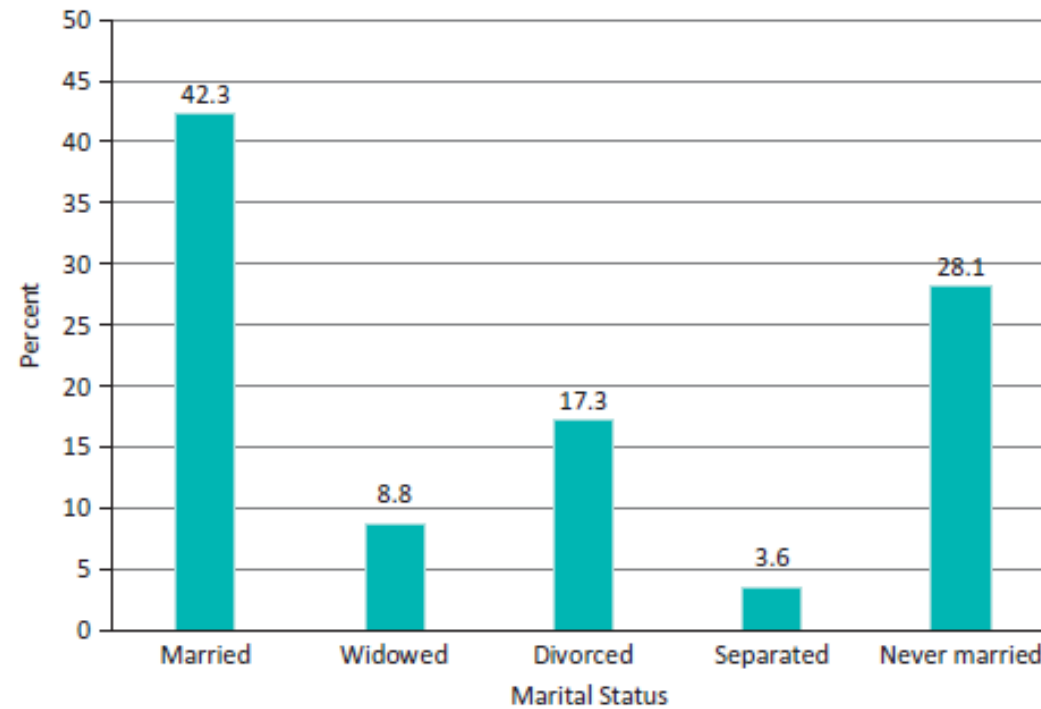


Figure 2.1 Marital Status among Americans, Bar Graph

Source: 2016 General Social Survey

Pie Chart

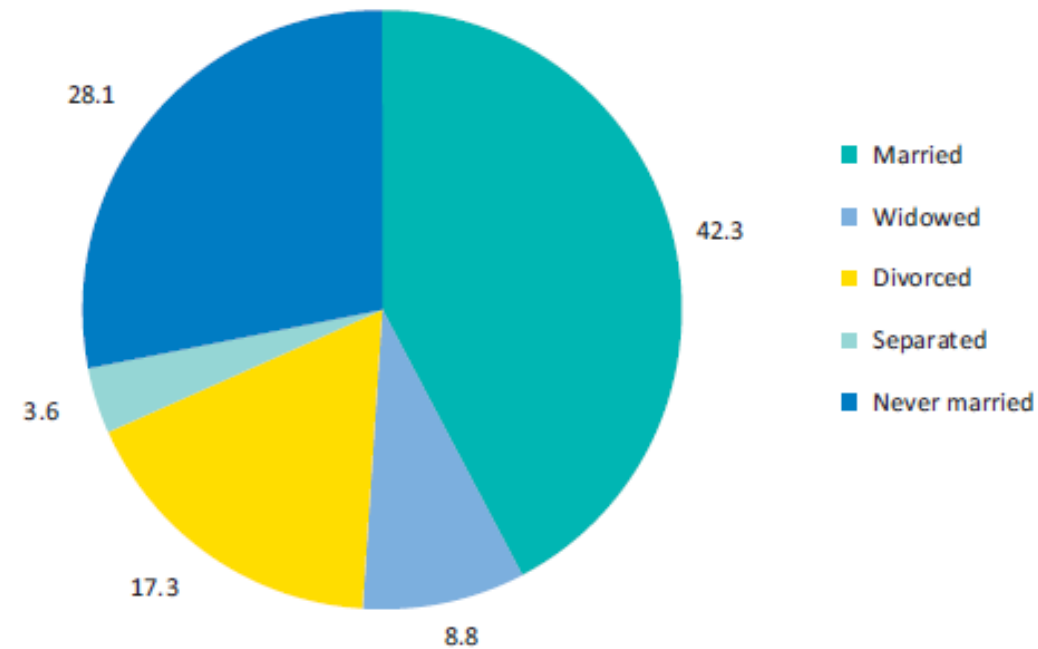


Figure 2.2 Marital Status among Americans, Pie Chart

Source: 2016 General Social Survey

Histogram and its Frequency Distribution

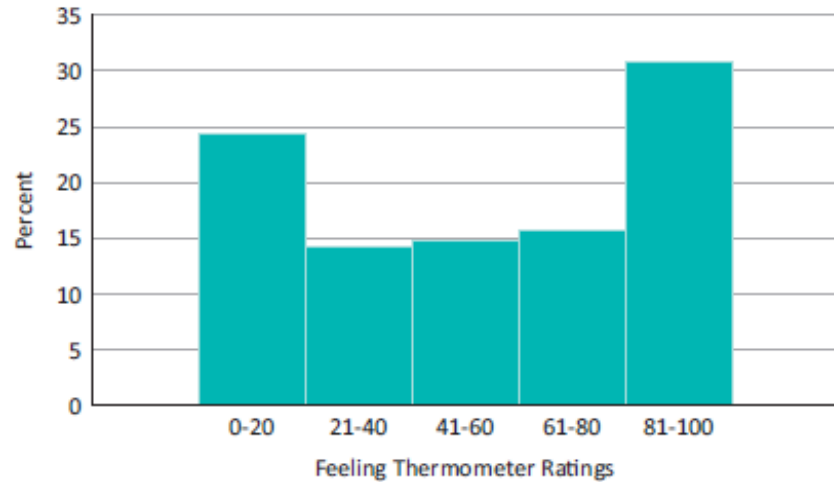


Figure 2.3 Histogram for Obama Feeling Thermometer, Recoded

Source: 2012 American National Election Study

Table 2.16 Obama Feeling Thermometer Ratings, Recoded

Rating Interval	Frequency	Percent	Cumulative Percent
Very cold (0–20)	1,435	24.3	24.3
Cold (21–40)	840	14.2	38.5
Slightly cold to slightly warm (41–60)	871	14.8	53.3
Warm (61–80)	931	15.8	69.1
Very warm (81–100)	1,820	30.9	100
Total	5,897	100	

Source: 2012 American National Election Study.

Stem-and-Leaf Plot and its Frequency Distribution

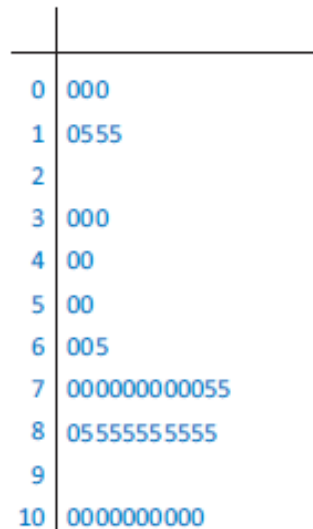


Figure 2.4 Obama Feeling Thermometer Ratings, Stem-and Leaf-Plot

Source: 2012 American National Election Study

Table 2.17 Obama Feeling Thermometer Ratings, Random Sample of Fifty

Rating	Frequency	Percent
0	3	6.0
10	1	2.0
15	3	6.0
30	3	6.0
40	2	4.0
50	2	4.0
60	2	4.0
65	1	2.0
70	10	20.0
75	2	4.0
80	1	2.0
85	10	20.0
100	10	20.0
Total	50	100

Source: 2012 American National Election Study.

Frequency Polygon (superimposed on its histogram)

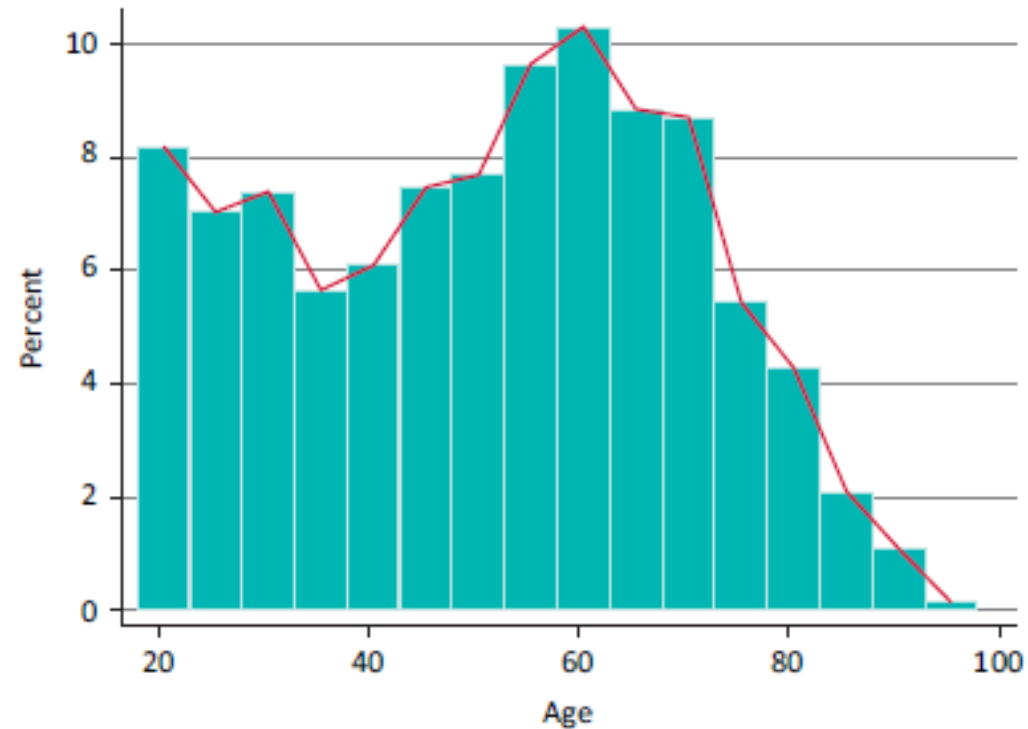


Figure 2.5 Frequency Polygon for Age, Superimposed on a Histogram for Age

Source: 2016 Pew Research Center

Copyright (C) 2020 The Rowman & Littlefield Publishing Group, Inc. All Right Reserved

Time Series Chart

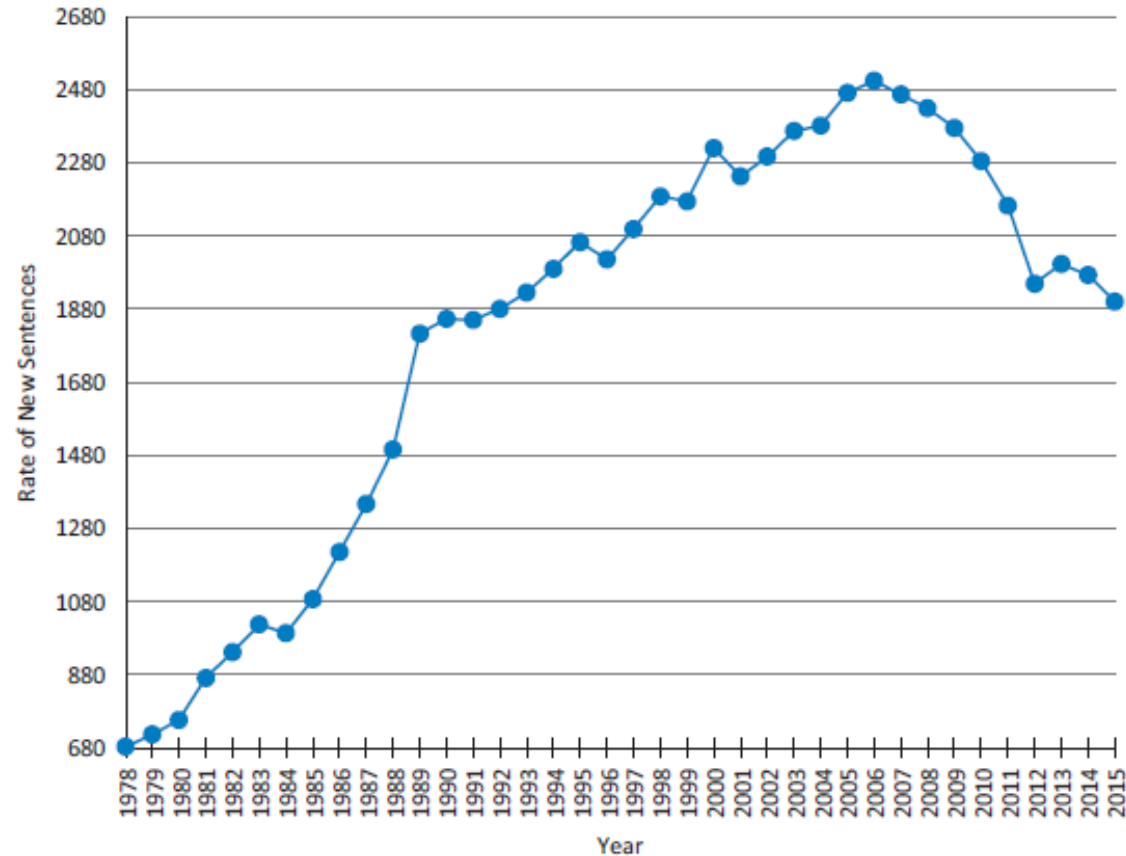


Figure 2.6 Rates of New Prison Sentences (per Million), 1978–2015

Source: Bureau of Justice Statistics, 1972–2012: *Prisoners in 2012: Trends in Admissions and Releases*, 1991–2012 (<https://www.bjs.gov/content/pub/pdf/p12tar9112.pdf>); 2013–2015: <https://www.bjs.gov/index.cfm?ty=tp&tid=131>

Copyright (C) 2020 The Rowman & Littlefield Publishing Group, Inc. All Right Reserved

Comparing Frequency Distributions of Subgroups

Table 2.18 Gay Marriage Support, by Political Party and Level of Education

	Republican			Democrat	
	Frequency	Percent		Frequency	Percent
Strongly favor	24	7.0	Strongly favor	130	30.3
Favor	80	23.4	Favor	153	35.7
Oppose	133	38.9	Oppose	84	19.5
Strongly oppose	105	30.7	Strongly oppose	62	14.5
Total	342	100	Total	429	100
	High School or Less			Bachelor's Degree or Higher	
	Frequency	Percent		Frequency	Percent
Strongly favor	51	12.4	Strongly favor	169	33.3
Favor	138	33.5	Favor	161	31.7
Oppose	129	31.3	Oppose	92	18.1
Strongly oppose	94	22.8	Strongly oppose	86	16.9
Total	412	100	Total	508	100

Source: 2016 Pew Research Center.

Comparing Bar Graphs of Subgroups

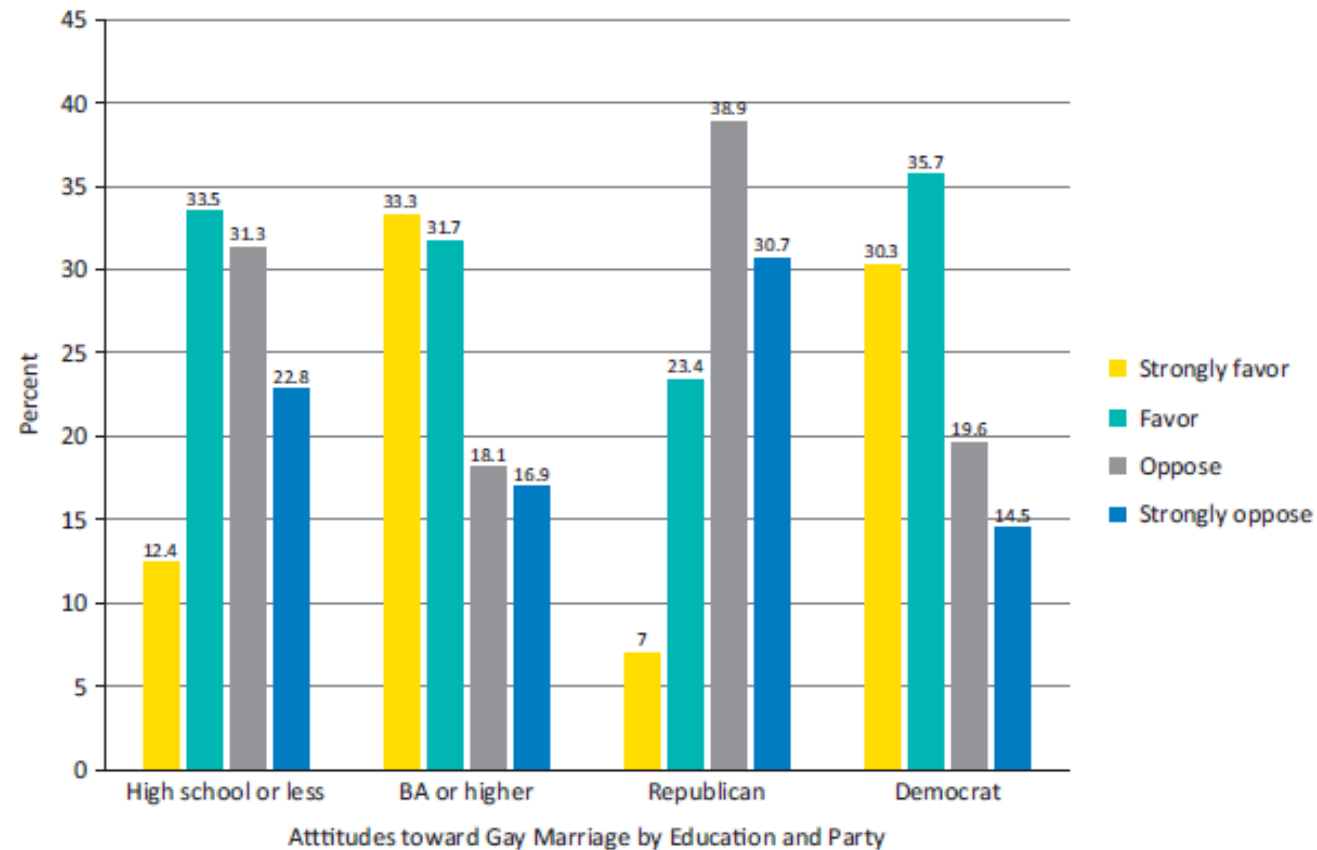


Figure 2.7 Bar Graphs of Attitudes toward Legalization of Gay Marriage, by Education and Political Affiliation

Source: 2016 Pew Research Center Copyright (C) 2020 The Rowman & Littlefield Publishing Group, Inc. All Right Reserved

Comparing Pie Charts of Subgroups

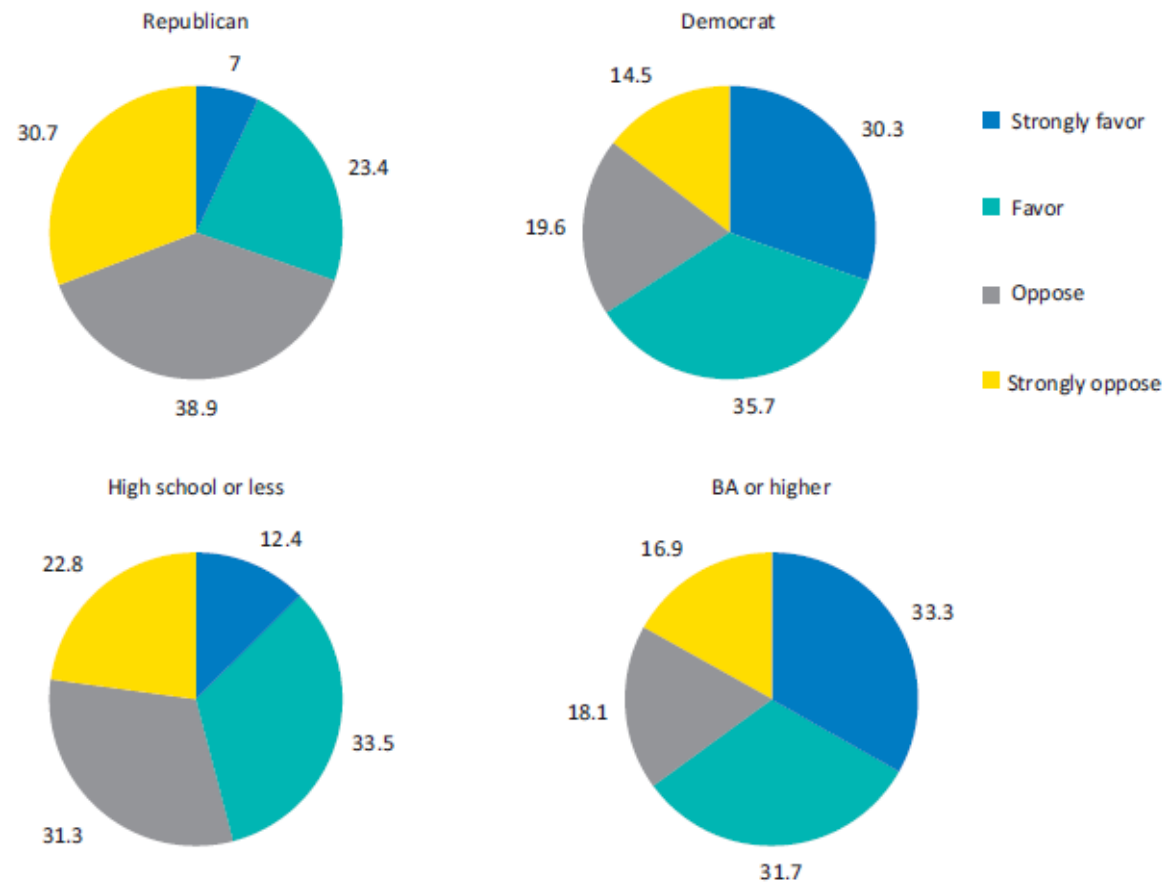


Figure 2.8 Pie Charts of Attitudes toward Legalization of Gay Marriage, by Education and Political Affiliation

Source: 2016 Pew Research Center. Copyright (C) 2020 The Rowman & Littlefield Publishing Group, Inc. All Right Reserved

Comparing Histograms of Subgroups

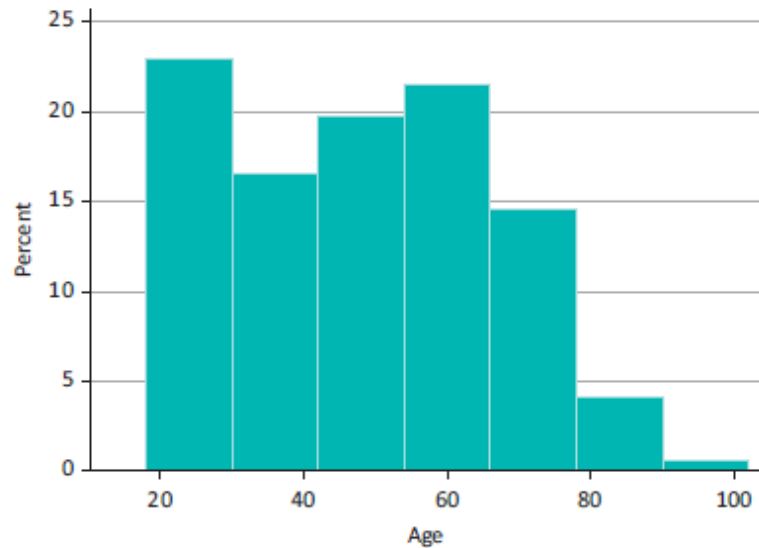


Figure 2.9a Age Distribution for Gay Marriage Supporters

Source: 2016 Pew Research Center.

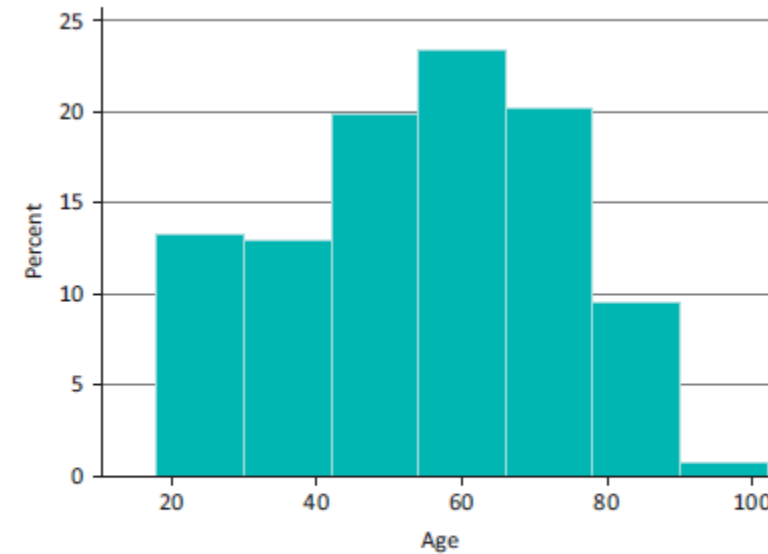


Figure 2.9b Age Distribution for Gay Marriage Opponents

Source: 2016 Pew Research Center.

Comparing Frequency Polygons of Subgroups

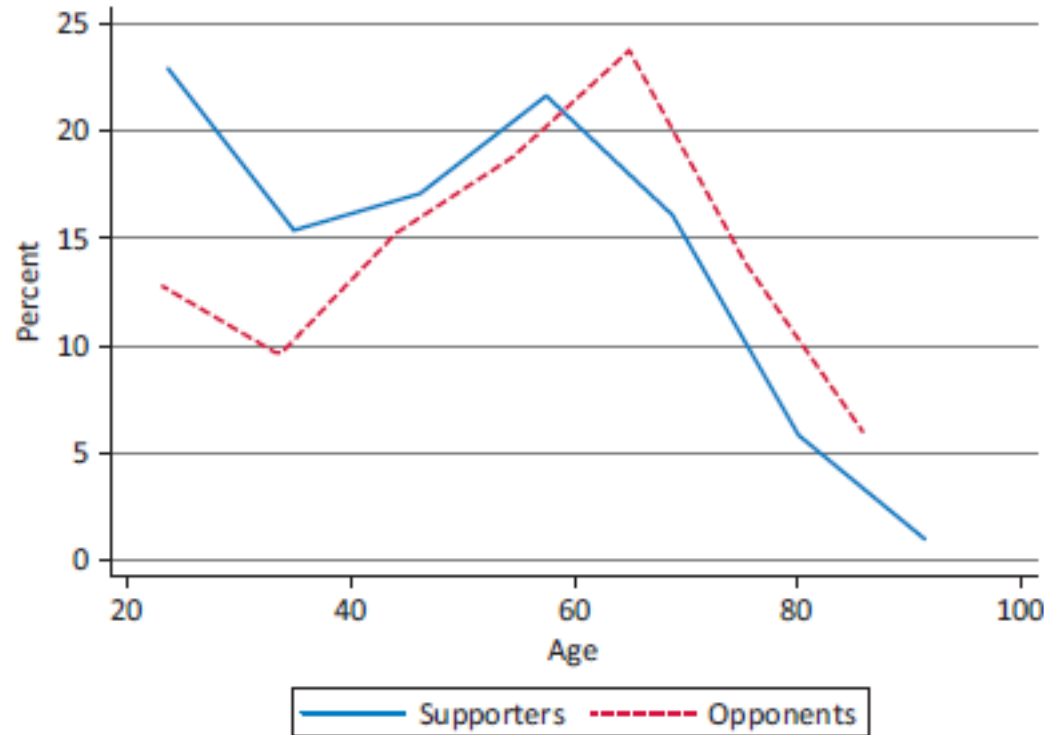


Figure 2.10 Frequency Polygons for Age, by Support for the Legalization of Gay Marriage

Source: 2016 Pew Research Center.