2

Displaying Data

2.1 The most appropriate graphs:
- **a.** Two histograms.
- **b.** Scatter diagram.
- **c.** Time series graph.
- **d.** Scatter diagram.
- **e.** Two histograms.

2.2 The most appropriate graphs are:
- **a.** Time series graph.
- **b.** Scatter diagram.
- **c.** Scatter diagram.
- **d.** Scatter diagram.
- **e.** Two histograms.

2.3 The most appropriate graphs are:
- **a.** Scatter diagram.
- **b.** Scatter diagram.
- **c.** Two histograms.
- **d.** Time series graph.
- **e.** Scatter diagram.

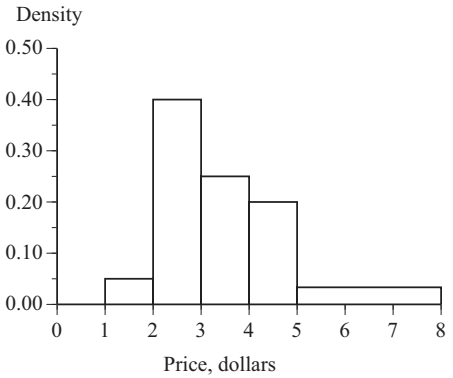
2.4 The most appropriate graphs are:
- **a.** Time series graph.
- **b.** Scatter diagram.
- **c.** Scatter diagram.
- **d.** Two histograms.
- **e.** Two histograms.

2.5 Here are the calculations:

Interval	Number of Countries	Relative Frequency	Histogram Bar Height
1–2	1	0.05	0.05
2–3	8	0.40	0.40
3–4	5	0.25	0.25
4–5	4	0.20	0.20
5–8	2	0.10	0.10/3
Total	20	1.00	

Copyright © 2015 Gary Smith. Published by Elsevier Inc. All rights reserved.

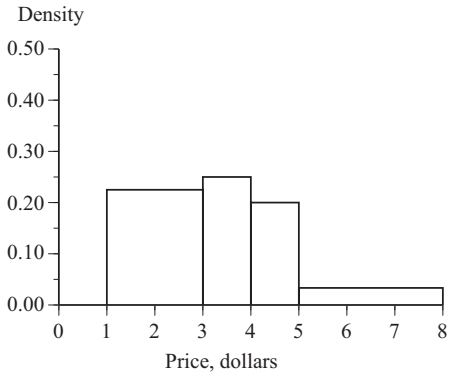
Here is the histogram:



2.6 Here are the calculations:

Interval	Number of Countries	Relative Frequency	Histogram Bar Height
1–3	9	0.45	0.45/2
3–4	5	0.25	0.25
4–5	4	0.20	0.20
5–8	2	0.10	0.10/3
Total	20	1.00	

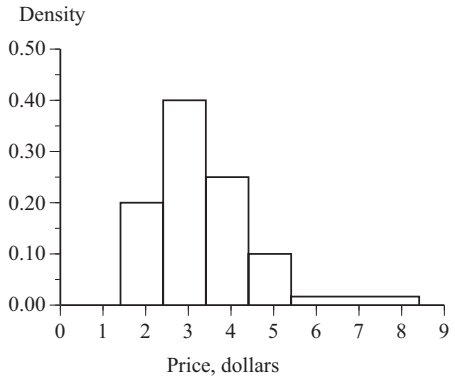
Here is the histogram:



2.7 Here are the calculations:

Interval	Number of Countries	Relative Frequency	Histogram Bar Height
1.41–2.41	4	0.20	0.20
2.41–3.41	8	0.40	0.40
3.41–4.41	5	0.25	0.25
4.41–5.41	2	0.10	0.10
5.41–8.41	1	0.05	0.05/3
Total	20	1.00	

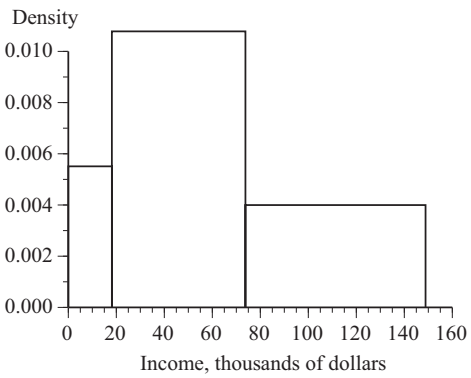
Here is the histogram:



2.8 Here are the data, collected by tax bracket:

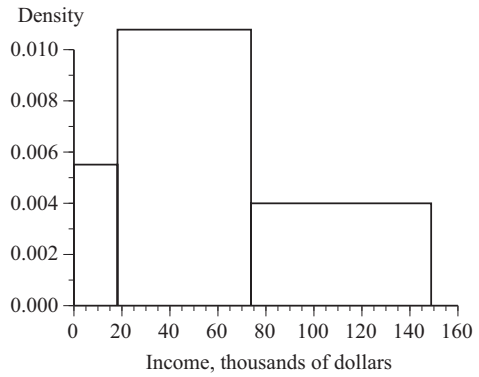
Taxable Income Bracket (thousands of dollars)	Observations	Frequency	Frequency/Interval Width
\$0 to \$18.150	1	0.1	0.1/18.15
\$18.151 to \$73.800	6	0.6	0.6/(73.80 − 18.15)
\$73.801 to \$148.850	3	0.3	0.3/(148.85 − 73.80)
\$148.851 to \$226.850	0	0	
\$226.851 to \$405.100	0	0	
\$405.101 to \$457,600	0	0	
\$457.601 to no limit	0	0	
Total	10	1.0	

Here is a histogram:



2.9 Here are the data, collected by tax bracket:

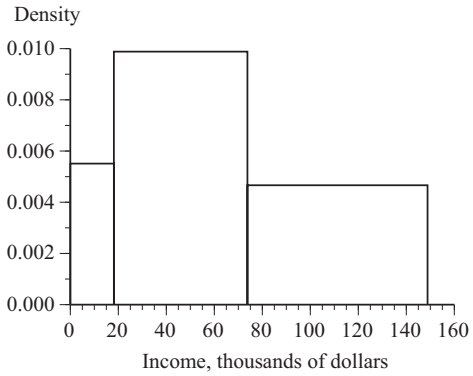
Taxable Income Bracket (thousands of dollars)	Observations	Frequency	Frequency/Interval Width
$0 to $18.150	1	0.1	0.1/18.15
$18.151 to $73.800	6	0.6	0.6/(73.80 − 18.15)
$73.801 to $148.850	3	0.3	0.3/(148.85 − 73.80)
$148.851 to $226.850	0		
$226.851 to $405.100	0		
$405.101 to $457,600	0		
$457.601 to no limit	0		
Total	10	1.0	



2.10 Here are the data, collected by tax bracket:

Taxable Income Bracket (thousands of dollars)	Observations	Frequency	Frequency/Interval Width
$0 to $18.150	2	0.10	0.10/18.15
$18.151 to $73.800	11	0.55	0.55/(73.80 − 18.15)
$73.801 to $148.850	7	0.35	0.35/(148.85 − 73.80)
$148.851 to $226.850	0		
$226.851 to $405.100	0		
$405.101 to $457,600	0		
$457.601 to no limit	0		
Total	20	1.0	

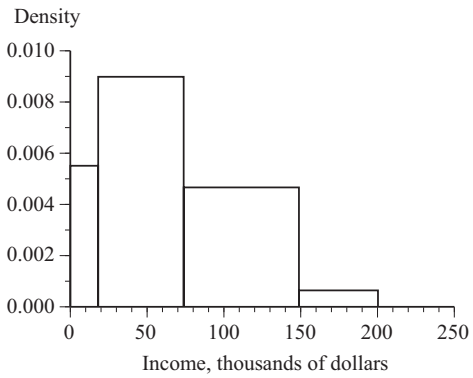
Here is a histogram:



2.11 Here are the data, collected by tax bracket:

Taxable Income Bracket (thousands of dollars)	Observations	Frequency	Frequency/Interval Width
\$0 to \$18.150	2	0.10	0.10/18.15
\$18.151 to \$73.800	10	0.50	0.50/(73.80 − 18.15)
\$73.801 to \$148.850	7	0.35	0.35/(148.85 − 73.80)
\$148.851 to \$226.850	1	0.05	0.05/(226.85 − 148.85)
\$226.851 to \$405.100	0		
\$405.101 to \$457,600	0		
\$457.601 to no limit	0		
Total	20	1.0	

Here is a histogram:



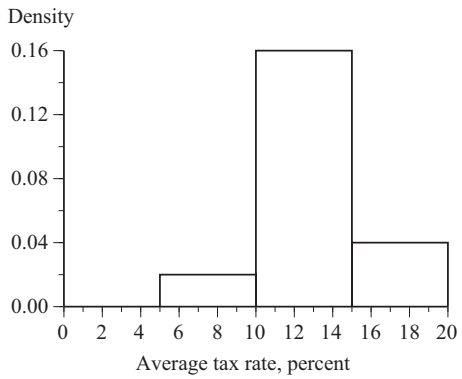
2.12 Here are the taxes and average tax rates:

Income	Tax ($)	Tax (percent)
10	1.000	10.00
20	2.093	10.99
30	3.593	12.33
40	5.093	12.99
50	6.593	13.40
60	8.093	13.66
70	9.593	14.55
80	11.713	14.64
90	14.213	15.79
100	16.713	16.71

Here are the histogram calculations:

Interval	Number of Couples	Relative Frequency	Histogram Bar Height
0–5			
5–10	1	0.10	0.10/5 = 0.02
10–15	7	0.70	0.70/5 = 0.16
15–20	2	0.20	0.20/5 = 0.04
20–25			
Total	10	1.00	

Here is the histogram:



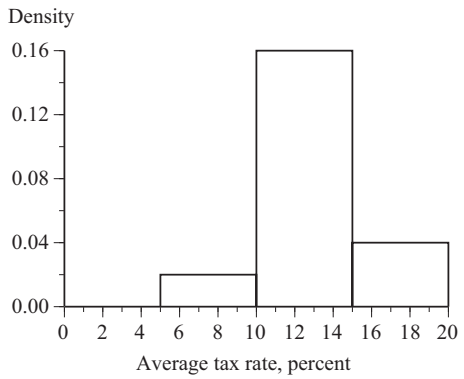
2.13 Here are the taxes and average tax rates:

Income	Tax ($)	Tax (percent)
15	1.343	8.95
23	2.543	11.05
32	3.893	12.16
37	4.643	12.55
48	6.293	13.11
48	6.293	13.11
75	10.463	13.95
80	11.713	14.64
110	19.213	17.47
130	24.213	18.63

Here are the histogram calculations:

Interval	Number of Couples	Relative Frequency	Histogram Bar Height
0–5			
5–10	1	0.10	0.10/5 = 0.02
10–15	7	0.70	0.70/5 = 0.16
15–20	7	0.20	0.20/5 = 0.04
20–25			
Total	10	1.00	

Here is the histogram:



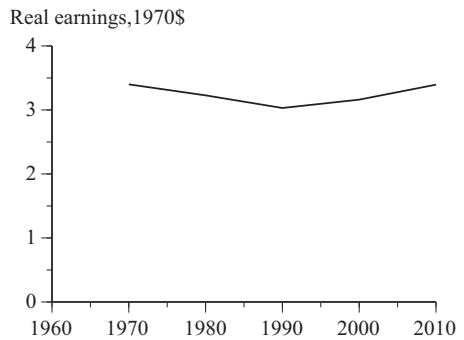
2.14 Answers will vary, but the total area should equal 1.

2.15 Answers will vary, but the total area should equal 1.

2.16 Real earnings are calculated by dividing by the CPI. The real earnings below are in 1970 dollars, calculated by multiplying nominal earnings by 116.3/CPI:

	Nominal Earnings	CPI	Real Earnings
1970	3.40	116.3	3.40
1980	6.85	246.8	3.23
1990	10.20	391.4	3.03
2000	14.02	515.8	3.16
2010	19.07	653.2	3.40

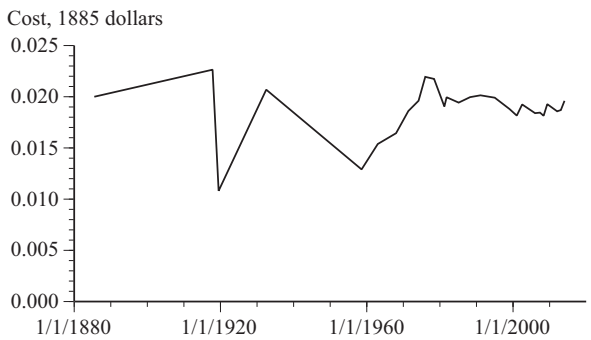
Here is a time series graph:



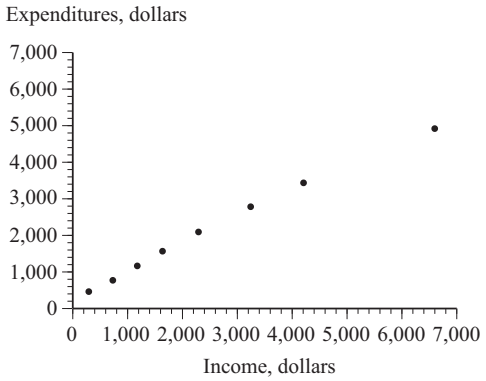
2.17 The real cost is calculated by dividing the postage cost by the CPI. I then multiplied by 4.0 to make the real cost in 1885 dollars: postage(4.0/CPI). For other base years, the graph would look the same, but the numbers on the vertical axis would be different.

	Postage	CPI	Real Cost
7/1/1885	0.02	4.00	0.0200
11/3/1917	0.03	5.30	0.0226
7/1/1919	0.02	7.40	0.0108
7/6/1932	0.03	5.80	0.0207
8/1/1958	0.04	12.40	0.0129
1/7/1963	0.05	13.00	0.0154
1/7/1968	0.06	14.60	0.0164
5/16/1971	0.08	17.20	0.0186
3/2/1974	0.10	20.40	0.0196
12/31/1975	0.13	23.70	0.0219
5/29/1978	0.15	27.60	0.0217
3/22/1981	0.18	37.80	0.0190
11/1/1981	0.20	40.10	0.0200
2/17/1985	0.22	45.30	0.0194
4/3/1988	0.25	50.10	0.0200
2/3/1991	0.29	57.60	0.0201
1/1/1995	0.32	64.30	0.0199
1/10/1999	0.33	70.20	0.0188
1/7/2001	0.34	74.90	0.0182
6/30/2002	0.37	76.90	0.0192
1/8/2006	0.39	84.80	0.0184
5/14/2007	0.41	88.90	0.0184
5/12/2008	0.42	92.60	0.0181
5/11/2009	0.44	91.30	0.0193
1/22/2012	0.45	96.90	0.0186
1/27/2013	0.46	98.40	0.0187
1/26/2014	0.49	100.00	0.0196

Here is a time series graph:



2.18 There seems to be a close positive relationship, with maybe a little tailing off at higher income levels:



2.19 **a.** Positively related because of genes that are passed down from the father (explanatory variable) to the son (dependent variable).
b. Positively related because the mother and child grow old together. The mother s age is the explanatory variable.
c. Negatively related because of hair loss as men age. Age is the explanatory variable.
d. Positively related. Large parents tend to have large children.
e. Positively related because people who are older are more likely to die in the next year, and insurance companies therefore charge more for life insurance. Age is the explanatory variable.

2.20 **a.** Positively related because children grow during their rst year. Age is the explanatory variable.
b. Positively related because breast cancer, like many diseases, is more common as people get older. Age is the explanatory variable.
c. Essentially unrelated. After a person is fully grown, the passage of time has little effect on height. (The elderly do tend to get smaller, but not between the ages 30 and 40.)
d. Negatively related because it takes more energy to move heavier cars. Weight is the explanatory variable.

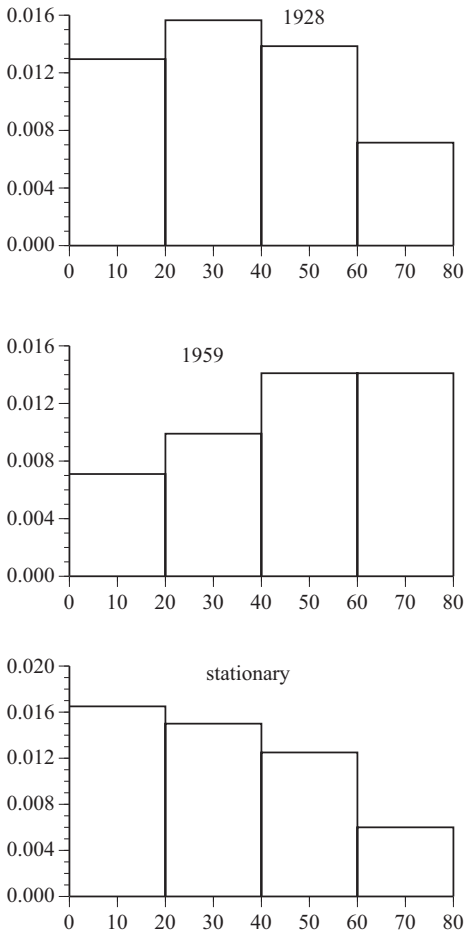
2.21 Because these are percentages, we can divide by 100 to obtain the relative frequency, and then calculate the histogram heights by dividing each relative

frequency by the constant interval width of 20 (I assume that age 19, for example, includes the full year that the person is 19):

Histogram Heights

Age	1928	1959	Stationary Model
0–19	0.0130	0.0071	0.0165
20–39	0.0157	0.0099	0.0150
40–59	0.0139	0.0141	0.0125
60–80	0.0072	0.0141	0.0060

The three histograms show that the 1928 age distribution was closer to that for a stationary population than was the 1959 age distribution, which was skewed toward older ages.

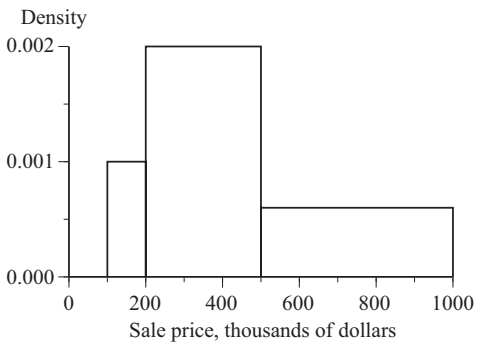


2.22 The histogram bar heights are each relative frequency by the interval width:

Interval ($1000)	Number of Houses	Relative Frequency	Histogram Bar Height
75–100	10	0.1667	0.1667/25
100–150	30	0.5000	0.5000/50
150–200	20	0.3333	0.3333/50
Total	60	1.00	

2.23 The histogram bar heights are determined by dividing each relative frequency by the interval width:

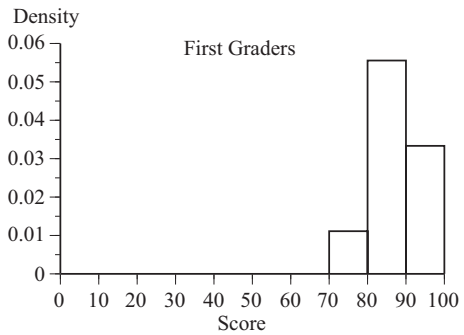
Interval ($1000)	Number of Houses	Relative Frequency	Histogram Bar Height
100–200	10	0.10	0.10/100
200–500	60	0.60	0.60/300
500–1000	30	0.30	0.30/500
Total	100	1.00	

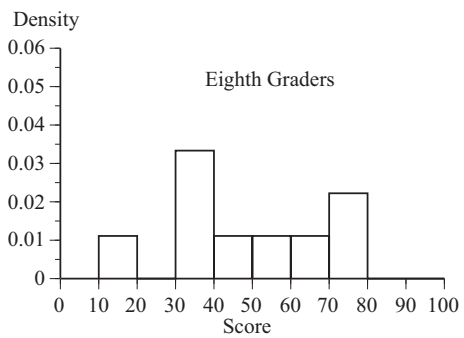


2.24 Here are the data grouped into intervals:

Score	Students	First Grade Frequency	Height	Students	Eighth Grade Frequency	Height
10–19				1	1/9	(1/9)/10
20–29						
30–39				3	3/9	(3/9)/10
40–49				1	1/9	(1/9)/10
50–59				1	1/9	(1/9)/10
60–69				1	1/9	(1/9)/10
70–79	1	1/9	(1/9)/10	2	2/9	(2/9)/10
80–89	5	5/9	(5/9)/10			
90–99	3	3/9	(3/9)/10			
Total	9	1.0		9	1.0	

Their scores are very compact and high in first grade, but lower and more dispersed in eighth grade:





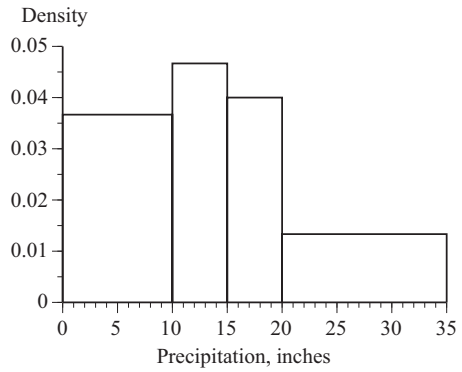
2.25 Arranging the 30 observations in order:

34.04	30.57	26.81	26.33	26.32	23.66	18.00	17.45	17.00	16.69
16.54	15.37	14.97	14.41	12.91	12.31	11.01	10.92	10.70	9.98
9.26	9.11	8.92	8.90	7.98	7.58	6.54	6.49	5.83	4.56

Here are the observed frequencies and histogram heights:

Interval (inches)	Number of Years	Relative Frequency	Histogram Bar Height
0–9.99	11	11/30	(11/30)/10 = 0.0367
10–14.99	7	7/30	(7/30)/5 = 0.0467
15–19.99	6	6/30	(6/30)/5 = 0.0400
20–34.99	6	6/30	(6/30)/15 = 0.0133
Total	30	1.0	

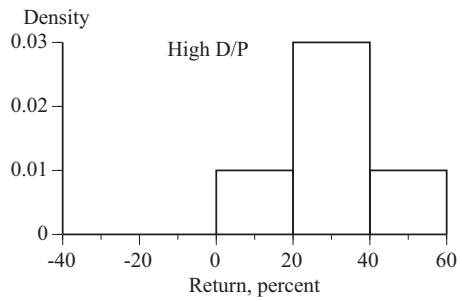
Here is a histogram:

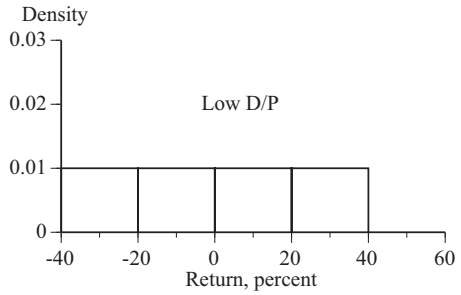


2.26 Here are the data grouped into intervals:

Return	Stocks	High D/P Frequency	Height	Stocks	Low D/P Frequency	Height
−40 to −20				2	2/10	(2/10)/20
−20 to 0				2	2/10	(2/10)/20
0 to 20	2	2/10	(2/10)/20	2	2/10	(2/10)/20
20 to 40	6	6/10	(6/10)/20	2	2/10	(2/10)/20
40 to 60	2	2/10	(2/10)/20	2	2/10	(2/10)/20
Total	10	1.0		10	1.0	

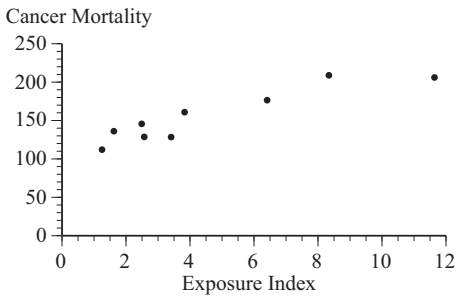
Here are the histograms:



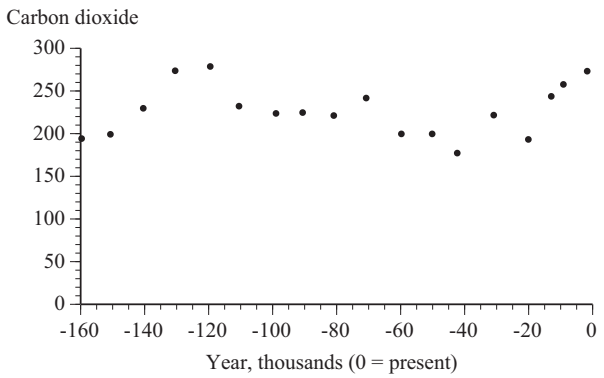


The histograms show high D/P stocks to generally have higher returns than low D/P stocks and to have considerably less variation in returns.

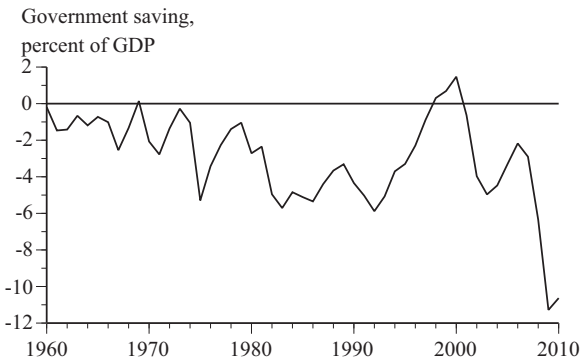
2.27 Exposure is the explanatory variable and cancer mortality is the dependent variable. There appears to be a positive relationship.



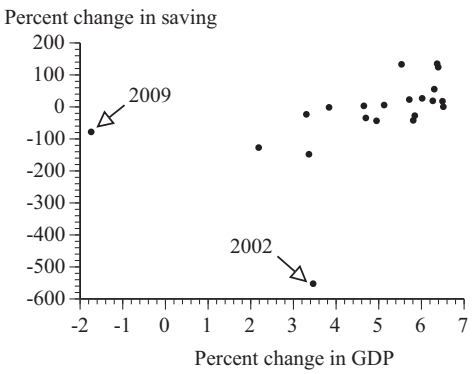
2.28 A time series graph shows considerable variation over time, with an upswing during the past 20,000 years from a relatively low point, and comparable to an earlier upswing that happened 150,000 years ago.



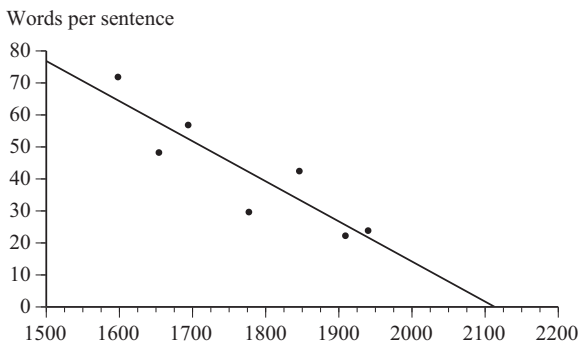
2.29 The US federal government has run a de cit almost every year since 1960, with a brief reversal of the long-run trend in the 1990s and a plunging de cit in 2008 2010.



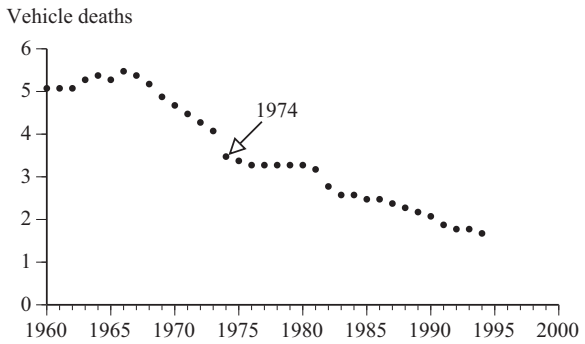
2.30 There are outliers in 2002 and 2009 but, overall, there is a positive relationship: when GDP increases, government saving tends to increase (the de cit declines) because government tax revenue increases when GDP increases.



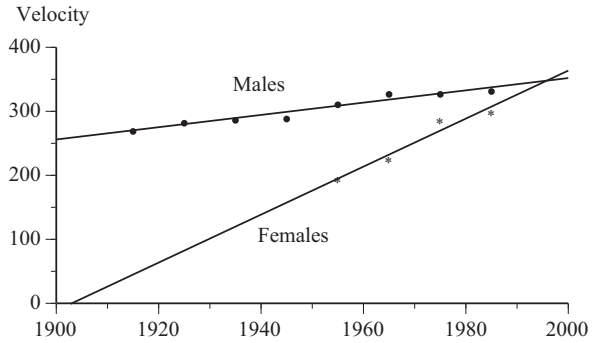
2.31 This is an example of incautious extrapolation since a straight line t to these data predicts negative words per sentence in 2200:



2.32 There was a sharp drop in 1974 but, otherwise, the data show a continuation of the long-term downward trend in US motor vehicle deaths per 100 million miles driven:

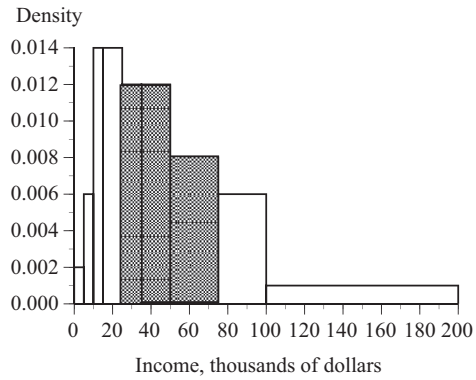


2.33 The lines cross in around 1995; the predicted female velocity in 1900 is negative!



2.34 The width of the bars does not re ect the differing widths of the income intervals. Also the middle three bars are labeled as the middle class, but this is not the middle of the income distribution: 45 percent of the population have higher incomes than this and only 11 percent have lower incomes. A better de nition of the middle class would be the middle 50 percent of the population, with incomes ranging from $25,000 to $75,000.

Not asked: Here is a properly drawn histogram:



2.35 Histogram bars cannot overlap.

2.36 The total area of the bars is not 1.

2.37 The total area of the bars is not 1.

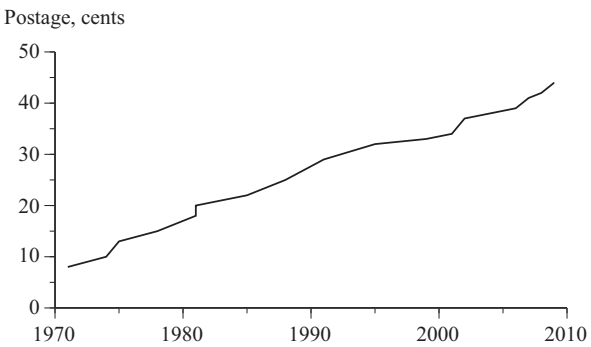
2.38 The total area of the bars is not 1.

2.39 In a histogram, density is not the rainfall for each year divided by the total rainfall for all 100 years. For a histogram, we should divide the 100 annual observations into a small number of categories, such as 0 10 inches, 10 20 inches, and so on. The density is the fraction of the total number of years with rainfall in that interval, divided by the interval width.

2.40 The omission of zero from the vertical axis magni es the index s zigs and zags.

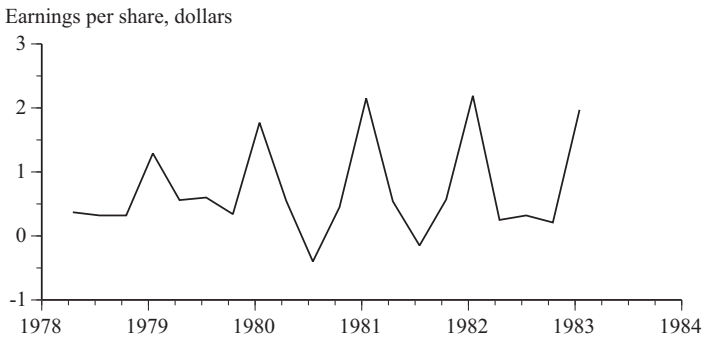
2.41 The units on the horizontal axis are not consistent, for example, the rst interval is 3 years, the next is 1 year. The intervals re ect the years when postage rates were increased, not the annual passage of time.

Not asked: A correctly drawn graph is:



2.42 It seems that these are the life expectancies of 400 people, but they are the age of the oldest person they know, not a typical person, let alone themselves. Plus, especially in a small town, they may know similar people. Imagine that one person is 110 and everyone in the town knows her. There would be 400 stickers at 110. Finally, it does not address the question of living longer at all because there are no data for earlier years.

2.43 These time series data are not graphed in historical order. Instead, the first five observations are from the first quarters of these 5 years; the next five observations are from the second quarters of these 5 years. A correctly drawn graph shows a pronounced seasonal pattern, with earnings rising sharply in each December 1 to February 28 quarter.

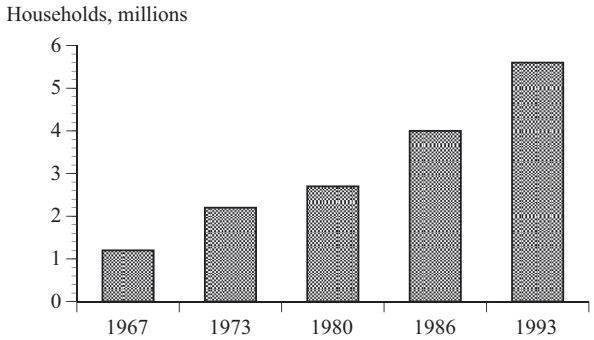


2.44 The data for 2011 are only for the first half of the year. If the number of passengers in the second half is as high as in the first half of the year (and this seems likely because of the traditional travel for Thanksgiving, Christmas, and vacationing in warmer climates), the total passenger miles for all of 2011 will be substantially higher than in 2009 and 2010.

2.45 In order to avoid distorting the image of the dollar bill, the height and width were both reduced by 50 percent, but this makes the area covered by the 2010 bill only 25 percent of the area covered by the 1986 bill. We look at the relative areas and think that the purchasing power fell by 75 percent, when it only fell by 50 percent.

2.46 It looks like a relatively sharp increase between 1980 and 1993, but then we notice that the years on the horizontal axis are not evenly spaced. There are 6 years between 1967 and 1973 and 7 years between 1973 and 1980, but 13 years between 1980 and 1993. If the number of

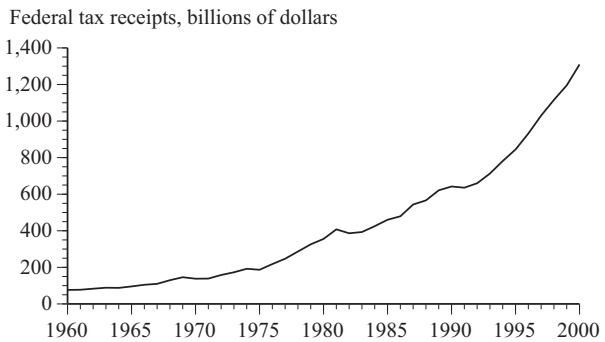
families in 1986 or 1987 is halfway between the number in 1980 and 1993, it would look like a simple linear progression, rather than an acceleration:



To put things into perspective, we should de ate these data by the price level and the total number of families. With prices and the total number of families increasing, it is not surprising that the number of families earning $100,000 or more is increasing.

2.47 Our attention is drawn to the area occupied by the city pictures, not their heights. In order to avoid distorting the pictures, the height and width of the Cairo picture are both twice the height and width of the Berlin picture, and the height and width of the Shanghai picture are both twice the height and width of the Cairo picture. Thus, the Cairo area is four times the Berlin area and the Shanghai area is four times the Cairo, even though the populations are twice as large, not four times as large.

2.48 The axes are reversed. Readers who do not look too carefully and are accustomed to seeing time on the horizontal axis will mistakenly conclude that taxes have leveled off. If time is put on the horizontal axis, as is customary, a very different picture emerges with taxes accelerating, rather than leveling off:



In addition, the data should be adjusted for the growth of the population and the increase in the general price level.

2.49 This pie chart has four slices when there are ve categories, and the slices are not labeled so we cannot tell which slice goes with which label. In addition, the gure label says, Poll Shows Little Change in Safety, but: (a) the numbers are from a poll regarding perceived safety, not actual safety data; and (b) the poll seems to be about current safety perceptions rather than the change in safety.

2.50 The break in the vertical axis is similar to omitting the origin: it allows us to see the difference more clearly, but it exaggerates the size of the differences. The height of the 2010 bar is approximately six times the height of the 2008 bar, but the numbers on the vertical axis show that sales were only 2 percent higher.