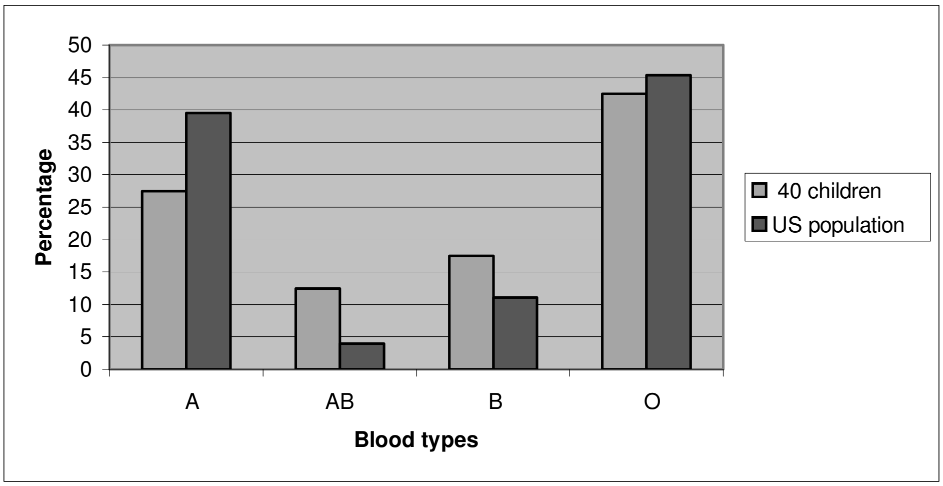
Chapter 2
How to Describe and Summarize data

2.1

Blood Type	Freq.	40 children (%)	US pop. (%)
A	11	27.50	39.5
AB	5	12.50	4.0
B	7	17.50	11.1
O	17	42.50	45.4
	n = 40		

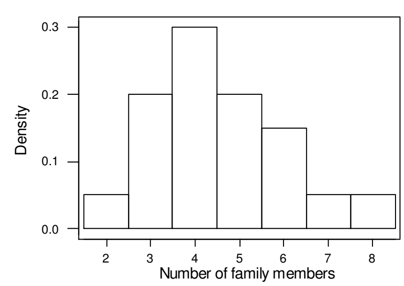


2.3 Sturge's rule, $2^{k-1} \approx n$, can be given in the form of a table.

n	1	2	4	8	16	32	64	128
k	1	2	3	4	5	6	7	8

Since n = 20, $k \approx 6$. From the data, we find 2 is the minimum and 8 is the maximum.
The range is 8 – 2 = 6.
Thus, the class interval can be 1, since 6/6 =1.

Family members	Count	Percent
2	1	5.00
3	4	20.00
4	6	30.00
5	4	20.00
6	3	15.00
7	1	5.00
8	1	5.00
	n = 20	

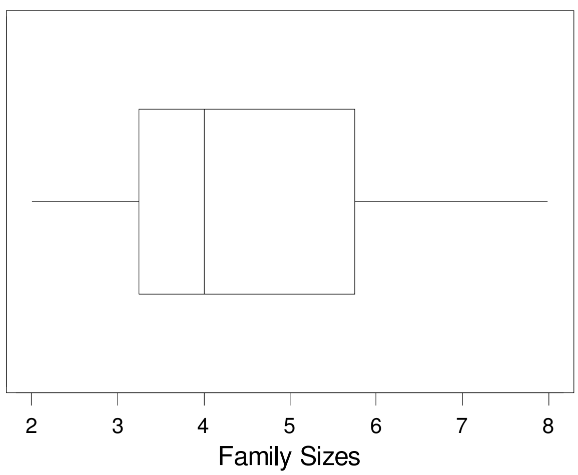


2.5 Stem-and-leaf plot of family size n = 20

```
2 |  0
3 |  0000
4 |  000000
5 |  0000
6 |  00
7 |  0
8 |  0
```

Five-number summary: 2 3.5 4 5.5 8

2.7

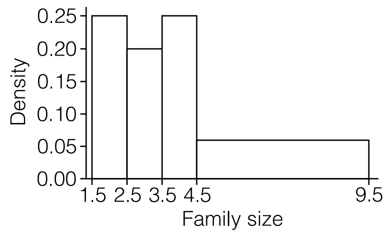


2.9

Class	x	f	xf
[–1,1)	0	1	0
[1,3)	2	2	4
[3,4)	3.5	4	14
[4,8)	6	1	6
Total		n = 8	24

$\overline{x} = \Sigma\, xf/n = 24/8 = 3$

2.11 **a.**



Notice that the last class is four times wider than the others.

b.

Class	x	f	xf
2	2	15	30
3	3	12	36
4	4	15	60
5 – 9	7	18	126
Total		n = 60	252

$\bar{x} = \Sigma$ xf/n = 252/60 = 4.20

c. Since there are 15 families of size 2, there are 30 people who come from families of size 2. Since there are 12 families of size 3, there are 36 people who come from families of size 3, and so on.

Class	x	f	xf
2	2	30	60
3	3	36	108
4	4	60	240
5 – 9	7	126	882
Total		n = 252	1290

$\bar{x} = \Sigma$ xf/n = 1290/252 = 5.12

The mean family size of the 60 families is 4.20. Polling the 252 people will give you a larger mean of 5.12 because families with more people are polled more often than small families. This gives the perception that there are many large families.

2.13 $\bar{x} = (2.4 + 2.2 + 2.4 + \ldots + 3.4)/10 = 23/10 = 2.3$

$s^2 = [(2.4 - 2.3)^2 + (2.2 - 2.3)^2 + (2.4 - 2.3)^2 + \ldots (3.4 - 2.3)^2]/9 = 4.56/9 = .5067$

$s = \sqrt{.5067} = .7118$

2.15

Class	x	f	xf	$x - \bar{x}$	$f(x - \bar{x})^2$
[−1,1)	0	1	0	−3	9
[1,3)	2	2	4	−1	2
[3,4)	3.5	4	14	+.5	1
[4,8)	6	1	6	+3	9
Total		n = 8	24		21

$s^2 = 21/7 = 3$ $\quad s = \sqrt{3} = 1.7321 \quad SE = s/\sqrt{n} = \sqrt{3}/\sqrt{8} = 0.6124$

2.17 $\bar{x} = (2 + 8 + 4 + \ldots + 6)/20 = 91/20 = 4.55$

$s^2 = [(2 - 4.55)^2 + (8 - 4.55)^2 + (4 - 4.55)^2 + \ldots 6 - 4.55)^2]/19 = 42.978/19 = 2.262$

$s = \sqrt{2.262} = 1.504$

2.19 **a.** $\bar{x} = 6.994$, $\quad s = 2.139$, $\quad SE = 2.139/\sqrt{16} = .535$

b. 6.99 1bs, 2.139 1bs

c. .535 lbs

2.21

x	f	fx	f(x−xbar)^2	
65	4	260	69.950	
67	7	469	33.322	
69	9	621	0.297	
71	9	639	29.753	
73	3	219	43.736	
75	1	75	33.851	
	sum:	33	2283	210.910

$\bar{x} = 69.1818$

$s^2 = 210.91/32 = 6.59094$

$s = \sqrt{6.59094} = 2.56728$

2.23

x_p	Percentile
4.5	.50
?	.80
6.5	.875

Since the function H is a straight line between 4.5 and 6.5, the slope of the line is $\frac{.875 - .50}{6.5 - 4.5} = .1875$.

As p increases from .50 to .80, x_p will increase from 4.5 to $x_{.80}$.

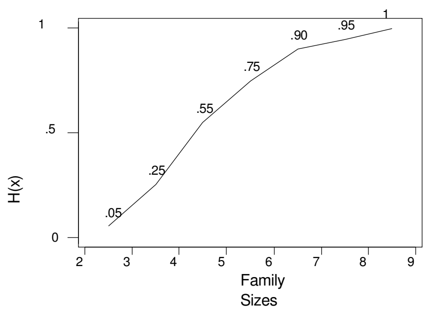
$$.80 - .50 = .1875(x_{.80} - 4.5)$$

By solving for $x_{.80}$, we have $x_{.80} = 4.5 + (.8 - .5)/.1875 = 6.1$

Since the function H is a straight line around $x_{.30}$, it follows that $.30 - .50 = .1875(x_{.30} - 4.5)$

By solving for $x_{.30}$, we have $x_{.30} = 4.5 + (.30 - .50)/.1875 = 3.433$

2.25



Reading from the ogive, we have $x_{.25} = 3.5$

2.27

lower	upper	income	f	f*income
0	5000	2500	79	197500
5000	10000	7500	103	772500
10000	12500	11250	20	225000
12500	15000	13750	14	192500
15000	17500	16250	26	422500
17500	20000	18750	38	712500
20000	22500	21250	52	1105000
22500	25000	23750	39	926250
25000	27500	26250	78	2047500
27500	30000	28750	54	1552500
30000	32500	31250	74	2312500
32500	35000	33750	29	978750
35000	37500	36250	101	3661250
37500	40000	38750	36	1395000
40000	42500	41250	78	3217500
42500	45000	43750	63	2756250
45000	47500	46250	140	6475000
47500	50000	48750	123	5996250
50000	55000	52500	250	13125000
55000	60000	57500	147	8452500
60000	75000	67500	366	24705000
75000	100000	87500	483	42262500
100000	125000	112500	434	48825000
125000	150000	137500	260	35750000
150000	300000	225000	1468	330300000
		sum:	4555	538366250

The mean = 538366250/4555 = $118192.

For some other upper bound b, the mean becomes

$$118192 + \frac{b - 30000}{2} * \frac{1468}{4555} = 69849.52 + .1611416 * b$$

For b = $342,248.56, the mean becomes $125,000.

For b = $497,391.61, the mean becomes $150,000

2.29 The median occurs at y = 4555/2 = 2277.5 and x = $94,022. This can be computed as follows. The cumulative frequency at $75,000 is F = (79 + 103 + 20 + … + 366) = 1910. The cumulative frequency at $100,000 is F = 1910 + 483 = 2393. The median occurs at the cumulative frequency of 4555/2 = 2277.5. So the median occurs at

$$x_{.50} = 75000 + \frac{2277.5 - 1910}{2393 - 1910}(100000 - 75000) = 94022$$

2.33 **Note:** $\int_a^b I_j(x)dx = a_j - a_{j-1} = w_j$

So $\int_a^b h(x)dx = \int_a^b \sum_{j=1}^{k} I_j(x)\frac{f_j}{n \cdot w_j}dx = \sum_{j=1}^{k}\frac{f_j}{n \cdot w_j}\int_a^b I_j(x)dx = \sum_{j=1}^{k}\frac{f_j}{n} = 1$

2.35 The cumulative distribution function is constant on intervals between observed values. So the derivative is zero on these open intervals. The derivative does not exist at the n observed values.

2.39 **a.**

Class	x	f
[7.5, 12.5)	10	2
[12.5, 17.5)	15	4
[17.5, 22.5)	20	7
[22.5, 27.5)	25	2

b. 20

c. 18

d. 4.55

e. 9.9 13.55 17.7 19.55 24.5

f. IQR = 19.55 – 13.55 = 6.0

2.41 **a.**

Class	x	f	fx
[64,66)	65	12	780
[66,68)	67	13	871
[68,70)	69	5	345
[70,72)	71	2	142
[72.74)	73	1	73

b. 67

c. 67.0

d. 2.06

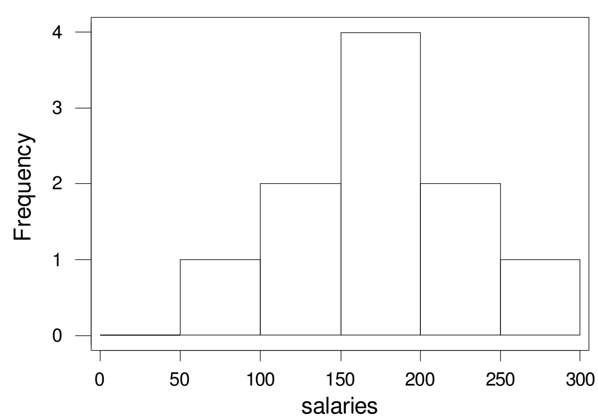
e. 64.3 65.5 66.8 67.8 72.2

f. 2.3

2.43 **a.**

Class	x	f	xf
[0.00, 50.00)	25	0	0
[50.00,100.00)	75	1	75
[100.00, 150.00)	125	2	250
[150.00,200.00)	175	4	700
[200.00,250.00)	225	2	450
[250.00,300.00)	275	1	275

b.



c. $\overline{x} = 175$, $s = 57.77$

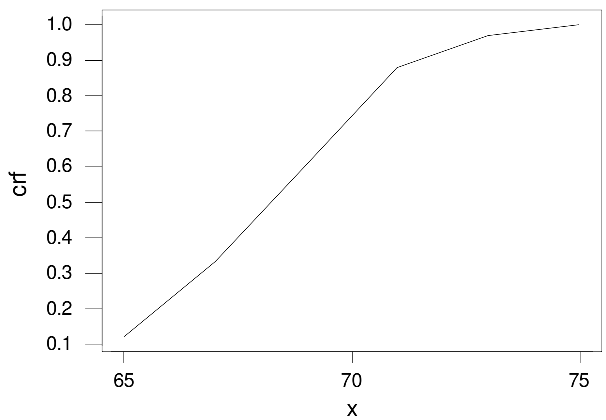
d. $\overline{x} = 173.1$, $s = 53.2$

e. 175.00

2.45 Approx. 68% of the histogram should lie between $\bar{x} - s = 80$ and $\bar{x} + s = 120$.
Approx. 95% of the histogram should lie between $\bar{x} - 2s = 60$ and $\bar{x} + 2s = 140$.
The third scale is the only one that satisfies these two conditions.

2.47 The mean and standard deviation increase to 17.27 and 33.45, respectively. The median and IQR are resistant to errors because they remain the same.

2.49 **a.**



b. $$x_{.60} = 67 + (69 - 67)\frac{33 * .6 - 11}{20 - 11} = 67 + (69 - 67)\frac{19.8 - 11}{20 - 11} = 69 \approx 70 \text{ inches}$$