Categorical variables:

Frequency table:

Categorical variable	count	Percent
Level 1	n1	P1
Level 2	n2	P2
…	…	…
total	n	100

Graph:

From values of count ⇨ bar graph

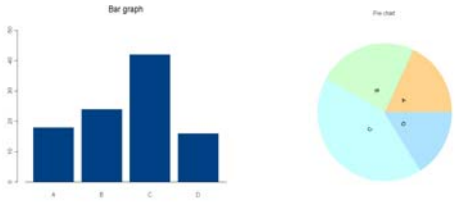
From values of percent ⇨ pie chart

Data Set 1:

C B C A C D D B D C C C A C D C C C C A C A D B B A D A C B
A D B B C B B D A D B C D D A B C B A B D B C B D A C C B C
C C A C C C D C A A C D B A B C A A C C B B B C C C A D C B
B C C C C C C C B C

Frequency: A-18, B-24, C-42, D-16

Percent: A-18%,B-24%,C-42%,D-16%



Quantitative variables:

Stemplot:

- Stem: consisting of all but the final digit
- Leaf: the final digit
- Back-to-back stemplot
- Two modifications for moderate data and many digits data

Histogram:

- Breaking the range of values of a variable into equal length intervals
- Calculating the freq or relative freq of the obs that fall into each interval

Back to back stemplot comparing the home run data:

Ruth		Maris
	0	8
52	1	346
54	2	368
9766611	3	39
944	4	
0	5	
	6	1

Stemplots of 50 shoppers data

0	399
1	13456778899
2	001234555668888
3	25699
4	1345579
5	0359
6	1
7	0
8	366
9	3

0	3
0	99
1	134
1	5677889
2	001234
2	55668888
3	2
3	5699
4	134
4	5579
5	03
5	59
6	1
6	
7	0
7	
8	3
8	66
9	3

Data Set 2:

98	94	101	98	105	76	98	87	108	109
127	105	102	85	97	94	106	86	113	113
118	95	84	98	93	100	74	99	91	104
100	76	122	128	96	100	93	111	118	100
84	91	91	95	119	107	93	111	102	98
87	93	92	112	78	110	92	94	108	88
87	107	100	82	97	118	99	112	91	81
109	93	103	118	86	78	101	103	84	80
95	102	99	104	96	98	123	100	106	93
93	119	89	99	80	97	86	99	103	95

Stemplot for Data Set 2:

Stem	Leaf
7	46688
8	0012444566677789
9	11112233333334445555667778888889999
10	00000011222333445566778899
11	0112233888899
12	0112233888899

Stem	Leaf
7	4
7	6688
8	0012444
8	566677789
9	1111223333333444
9	5555667778888999
10	0000001122233344
10	5566778899
11	0112233
11	888899
12	0112233
12	888899

An example to illustrate the leaf unit of stemplots

Data set 1

85 110 110 160 160 180 180 200 185 225 245 90 120 160 145 282 340 154 170 240

Stemplot

0	99
1	112556667889
2	03458
3	4

Leaf unit=10

Split the stem

0	99
1	112
1	556667889
2	034
2	58
3	4

Data set 2

1.81 1.82 1.26 1.22 2.02 2.40 2.56 2.32 3.20 3.14 1.60 0.85 0.90 1.60 1.50 1.56 2.30 3.60 2.50 2.76

Stemplot

0	99
1	23566688
2	0334568
3	126

Leaf unit=0.1

Split the stem

0	99
1	23
1	566688
2	0334
2	568
3	12
3	6

Examining distribution by graph (stemplot, histogram)

Overall pattern

•Shape: symmetric or skewed; major peaks

•Center: midpoint

•Spread: range

Deviation

•Outlier: individual outside overall pattern

Newcomb's measurements of the passage time of light

Table 1.1 Newcomb's measurements of the passage time of light					
28	22	36	26	28	28
26	24	32	30	27	24
33	21	36	32	31	25
24	25	28	36	27	32
34	30	25	26	26	25
-44	23	21	30	33	29
27	29	28	22	26	27
16	31	29	36	32	28
40	19	37	23	32	29
-2	24	25	27	24	16
29	20	28	27	39	23

Draw a histogram

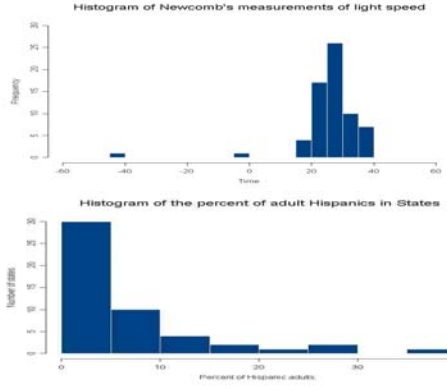
•Intervals: (-45, -40], (-40, 35], … , (30, 35], (35, 40], (40, 45]

Example 1.9 Hispanics adult population in States

Table 1.2 Percent of Hispanics in the adult population, by state (2000)					
State	Percent	State	Percent	State	Percent
AL	1.5	LA	2.4	OH	1.6
AK	3.6	ME	0.6	OK	4.3
AZ	21.3	MD	4	OR	6.5
AR	2.8	MA	5.6	PA	2.6
CA	28.1	MI	2.7	RI	7
CO	14.9	MN	2.4	SC	2.2
CT	8	MS	1.3	SD	1.2
DE	4	MO	1.8	TN	2
FL	16.1	MT	1.6	TX	28.6
GA	5	NB	4.5	UT	8.1
HI	5.7	NV	16.7	VT	0.8
ID	6.4	NH	1.4	VA	4.2
IL	10.7	NJ	12.3	WA	6
IN	3.1	NM	38.7	WV	0.6
IA	2.3	NY	13.8	WI	2.9
KS	5.8	NC	4.3	WY	5.5
KY	1.3	ND	1		

Draw a histogram

•Intervals: (0.0, 5.0], (5.0, 10.0], … , (30.0, 35.0], (35.0, 40.0]



Time plot

A plot with data against time (or the order of observations)

Light plot

Both outliers were made early

Variability is decreasing over time

Gasoline plot

It shows the increasing trend over time

Time series

•Measurements of a variable taken at regular intervals over time