# Chapter 10

# Graphs, Good and Bad

Data from categorical variable and quantitative variable are described in both a graphical and tabular form. Data for *categorical* variable organized into one of several groups (categories) and can only be counted. Data from *quantitative* variable can be added, subtracted, multiplied and divided. Distribution tables, bar graphs (Pareto charts), pie charts and line graphs are discussed in this chapter. *Distribution table* of variable give the values of variable and how often these values occur. *Line graph* is a plot of variable, along y-axis, versus time, along x-axis.

**Exercise 10.1 (Graphs, Good and Bad)**

1. *Bar graph and pie chart: patient health.* Health of twenty patients in a high blood pressure study are:

   good, good, fair, poor, bad, poor, great, fair, good, good,
   good, fair, fair, fair, good, poor, poor, bad, good, good.

   Distribution table, bar graph, Pareto chart and pie charts for data given below.

| category | number of patients | proportion of patients |
|---|---|---|
| bad | 2 | $\frac{2}{20} = 0.10$ |
| poor | 4 | $\frac{4}{20} = 0.20$ |
| fair | 5 | $\frac{5}{20} = 0.25$ |
| good | 8 | $\frac{8}{20} = 0.40$ |
| great | 1 | $\frac{1}{20} = 0.05$ |
| total | 20 | 1.0 |

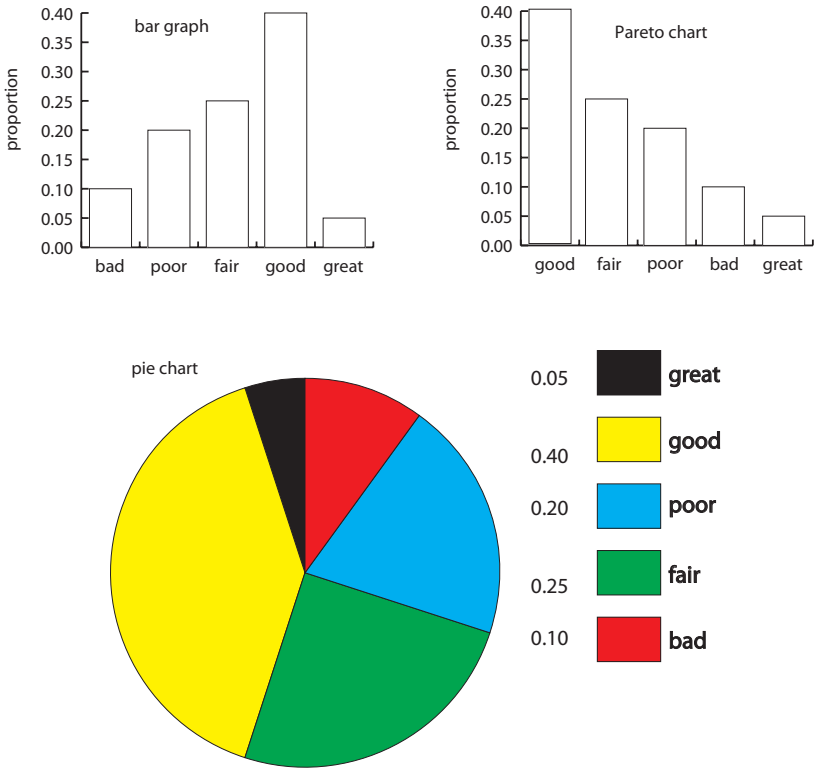


Figure 10.1 (Bar graph, Pareto chart and pie chart for patient health)

(a) This data is **categorical** / **quantitative** because data grouped into five categories: bad, poor, fair, good and great.

(b) Of 20 patients, **2** / **4** / **5** / **8** are in fair health or a proportion of $\frac{5}{20} = 0.25$.

(c) Height of each vertical bar in bar graph corresponds to proportion for each category. For example, vertical bar for "good" category has a height (or proportion) of (choose one) **0.30** / **0.35** / **0.40**.

(d) Adding heights of all vertical bars in five categories together, we get (choose one) **0.40** / **0.75** / **1.00**.

(e) Pareto chart is a bar graph where bars are arranged left to right in **decreasing** / **increasing** order.

(f) **True** / **False**. Another possible variation of a bar graph has *number* rather than proportion along y-axis. Heights of this version of bar graph do *not* necessarily add to one.

(g) **True** / **False** *Width* of each vertical bar has *no* meaning.

(h) Angle spanned by each wedge in pie chart is **smaller than** / **in proportion to** / **larger than** size of category. Wedges *must* add to a "whole" in pie chart since wedge angles add to $360^o$.

2. *Side-by-side bar graphs: comparing patients' health.* Health of twenty patients in a high blood pressure study are compared in 2001 and 2009.

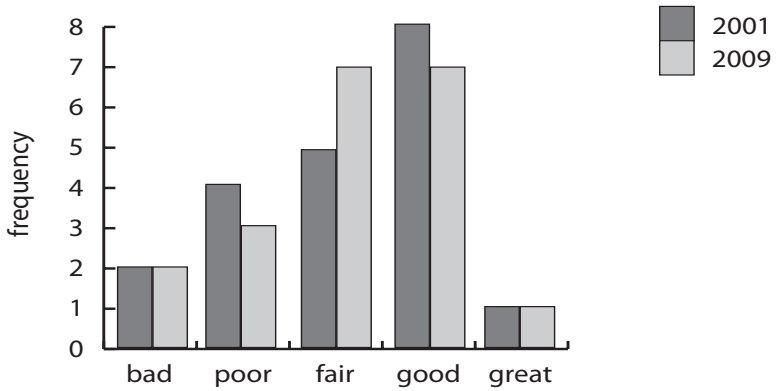


Figure 10.2 (Side–by–side bar graph)

(a) Number of patients in poor health in 2009: **3** / **4** / **5** / **8**.

(b) What category of health most improved from 2001 to 2009?
**bad** / **poor** / **fair** / **good** / **great**

(c) *Percentage* of patients in fair health in 2009: **25%** / **30%** / **35%** / **40%**.

3. *Line graph: stock prices over time.* Two stocks are compared over year period.

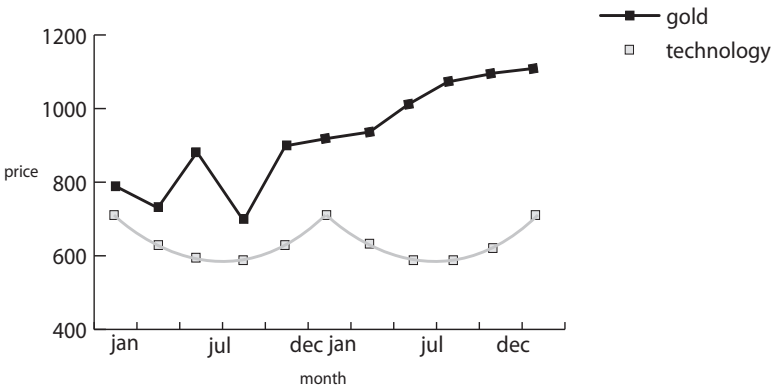


Figure 10.3 (Line graph of two stocks)

(a) *Overall trend.* Trend in gold stock is
**increasing** / **steady** / **decreasing** / **both decreasing and increasing**.

(b) *Overall trend.* Trend in technology stock is
**increasing** / **steady** / **decreasing** / **both decreasing and increasing**.

(c) *Deviations.* (Sharp) deviations in gold stock occur
**beginning** / **middle** / **end** / **throughout** / **no where in** time period.

(d) *Deviations.* (Sharp) deviations in technology stock occur **beginning / middle / end / throughout / no where in** time period.

(e) *Seasonal variation.* Seasonal variation in gold stock occur **beginning / middle / end / throughout / no where in** time period.

(f) *Seasonal variation.* Seasonal variation in technology stock occur **beginning / middle / end / throughout / no where in** time period.

4. *Graphical misrepresentations: unequal widths.*

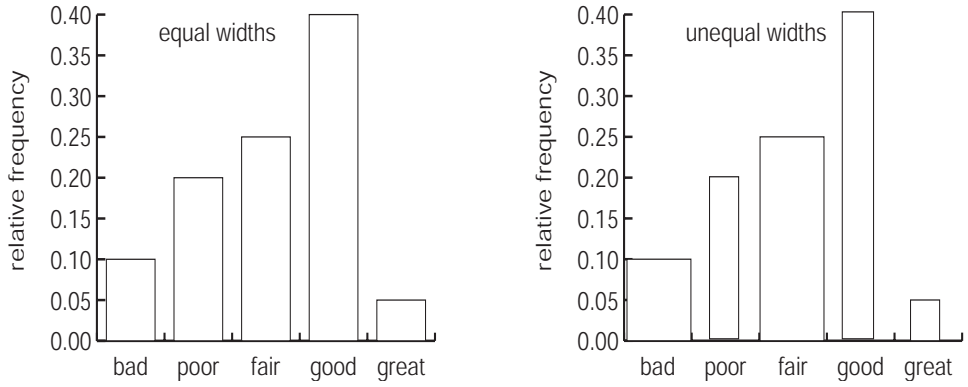


Figure 11.4 (Graphical misrepresentations: unequal widths)

Bar graph on right possibly misleading because it seems "bad" and "fair" health occur **less frequently than / as frequently as / more frequently than** other categories.

5. *Graphical misrepresentations: truncated and adjusted scale.*

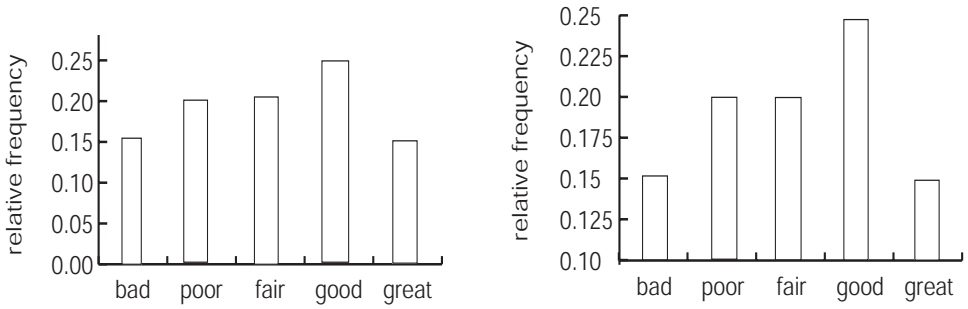


Figure 11.5 (Graphical misrepresentations: truncated and adjusted scale)

Bar graph on right possibly misleading because it seems **greater / same / lesser** difference between categories.

# Chapter 11

# Displaying Distributions with Graphs

Data from quantitative variable is described in both a graphical form, in particular *histograms* and *stemplots*, and in a tabular form. We look at overall pattern and for *deviations* such as *outliers* in these graphs. Overall pattern can be described using *shape*, *center* and *spread*. Some graphs have simple shapes which are *symmetric*, *(right or left) skew*, or *uniform*.

**Exercise 11.1 (Displaying Distributions with Graphs)**

1. *Histograms for quantitative data.*

   (a) *Patient Ages.* Distribution table and histogram for patient ages, treated as continuous data, are given below.

   32, 37, 39, 40, 41, 41, 41, 42, 42, 43,
   44, 45, 45, 45, 46, 47, 47, 49, 50, 51

| class | count | proportion |
|---|---|---|
| 30 to 34 | 1 | $\frac{1}{20} = 0.05$ |
| 35 to 39 | 2 | 0.10 |
| 40 to 44 | 8 | 0.40 |
| 45 to 49 | 7 | 0.35 |
| 50 to 54 | 2 | 0.10 |

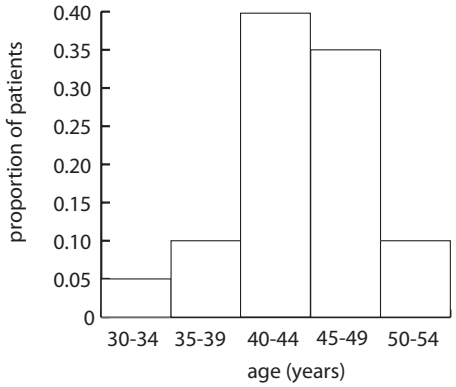


Figure 11.1 (Histogram for patient ages)

i. Number of classes is (circle one) **3** / **4** / **5** / **6**.

ii. First class is (circle one) **30 to 34** / **35 to 39** / **40 to 44**.

iii. *Lower* class limit of first class is (circle one) **30** / **34** / **35** / **39**.

iv. *Upper* class limit of first class is (circle one) **30** / **34** / **35** / **39**.

v. *Width* of first class is $35 - 30 =$ (circle one) **3** / **4** / **5** / **6** years.

vi. *Class width* equals (choose two)
**difference between consecutive lower class limits**
**difference between consecutive upper class limits**
**difference between upper and lower class limits**

vii. Number of patients in 30-34 age class is (circle one) **1** / **2** / **3** / **4**.

viii. Percentage of patients in 30-34 age class: **5%** / **10%** / **35%** / **40%**.

ix. *Center* of histogram roughly: **40** / **45** / **55** / **60** years.

x. *Spread* of histogram roughly: **35 − 55** / **30 − 60** / **30 − 55** years.

(b) *pH levels.* Consider distribution and histogram of 28 pH levels of soil data.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 4.3 | 5 | 5.9 | 6.5 | 7.6 | 7.7 | 7.7 | 8.2 | 8.3 | 9.5 |
| 10.4 | 10.4 | 10.5 | 10.8 | 11.5 | 12 | 12 | 12.3 | 12.6 | 12.6 |
| 13 | 13.1 | 13.2 | 13.5 | 13.6 | 14.1 | 14.1 | 15.1 | | |

| class | frequency | relative frequency |
|---|---|---|
| 4–5.9 | 3 | $\frac{3}{28} \approx 0.107$ |
| 6–7.9 | 4 | 0.143 |
| 8–9.9 | 3 | 0.107 |
| 10–11.9 | 5 | 0.179 |
| 12–13.9 | _____ | _____ |
| 14–15.9 | _____ | _____ |

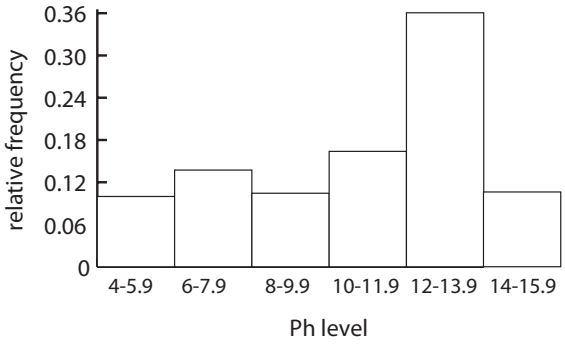


Figure 11.2 (Histogram for pH Level Data)

i. Fill in blanks in distribution table. Hint: 28 readings total.

ii. Number of classes is (circle one) **3** / **4** / **5** / **6**.

iii. Width of each class is (circle one) **2** / **3** / **4** / **5** pH.

iv. Most frequent pH reading is
**8 − 9.9** / **10 − 11.9** / **12 − 13.9** / **14 − 15.9**.

v. *Center* of histogram roughly: **10** / **11** / **12** / **13** pH.

vi. *Spread* of histogram roughly: **4 − 16** / **5 − 15** / **6 − 14** pH.

2. *Stemplots.*

(a) *Patient Ages.* Stemplot for patient ages given below.

32, 37, 39, 40, 41, 41, 41, 42, 42, 43,
44, 45, 45, 45, 46, 47, 47, 49, 50, 51

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 2 | 7 | 9* | | | | | | | | | | | | | |
| 4 | 0 | 1 | 1 | 1** | 2 | 2 | 3 | 4 | 5 | 5 | 5 | 6 | 7 | 7 | 9 | stem: 10s |
| 5 | 0 | 1 | | | | | | | | | | | | | | leaf: 1s |

i. Starred number, **9***, represents age (circle one) **39** / **93** / **9**.
Double–starred number, 1**, represents age (circle one) **41** / **14** / **1**.

ii. Numbers left of double line (in first column) are called **stems** / **leaves**; numbers to right are called (circle one) **stems** / **leaves**.

iii. Starred number 9* is a leaf of stem (circle one) **3** / **4** / **5**.

iv. **True** / **False** Note to right of stem-and-leaf plot specifies numbers used as stems are "tens" (or "10s") and numbers used as leaves are "ones" (or "1s"). So, for instance, stem "3" represents 3x10 = 30 and leaf "2" represents 1x2 = 2.

v. Stemplot ordered where, in first stem, for example, 32 followed by (circle one) **37** / **39** / **40**.

vi. *Center* of stemplot: **30**s / **40**s / **50**s.

(b) *Split stemplots: patient ages.* Sometimes, to spread data out, stems are *split* as, for example, in following table.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 3 | 2 | | | | | | | | |
| 3 | 7 | 9 | | | | | | | |
| 4 | 0 | 1 | 1 | 1 | 2 | 2 | 3 | 4 | |
| 4 | 5 | 5 | 5 | 6 | 7 | 7 | 9 | | |
| 5 | 0 | 1 | | | | | | | stem: 10s |
| 5 | | | | | | | | | leaf: 1s |

i. **True** / **False** Low stem 3 contains one half of leaves, 0, 1, 2, 3 or 4; high stem 3 contains other half of leaves, 5, 6, 7, 8 and 9.

ii. **True** / **False** Stemplots can have stems split not only twice, but also three or more times. Splitting each stem three times might, say, consist of a low stem which contains leaves 0, 1 and 2; a middle stem with leaves 3, 4, 5 and 6 and a high stem with leaves 7, 8 and 9.

iii. A stemplot with 10s as stems can be split at most (circle one) **5** / **7** / **10** / **100** times.

iv. **True** / **False** Although no one "best" way of constructing a stemplot, most consist of 5 to 20 stems.

v. *Spread* of stemplot: **32 − 51** / **30 − 50** / **29 − 49**.

(c) *Back–to–back plots.* Back–to–back stemplot compares "body temperature (before taking tablets)" to "body temperature (after taking tablets)".

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| body | | | | | | 4 | 97L | | | | | | | | body |
| temperature | | | | | | 8 | 97H | | | | | | | | temperature |
| before | | | | | | | 98L | 1 | | | | | | | after |
| tablet | | | | | 8 | 5 | 98H | 5 | | | | | | | tablet |
| | | | | | | 0 | 99L | | | | | | | | |
| | | | | | | 9 | 99H | 6 | 8 | | | | | | |
| | | | 3 | 2 | 0 | 0 | 100L | 0 | 0 | 1 | 1 | 2 | 3 | 4 | |
| | 9 | 9 | 8 | 7 | 7 | 6 | 100H | 7 | 8 | 8 | 8 | 9 | 9 | | stem: 1s |
| | | | | 1 | 1 | 0 | 101L | 0 | 1 | 4 | | | | | leaf: 0.1s |
| | | | | | | 5 | 101H | | | | | | | | outlier: 96.2 |

i. Before and after body temperatures appear centered around (circle one) **90** / **100** / **110** Also, "shape" of both plots seem about the same.

ii. Before body temperatures **higher than** / **about same as** / **lower than** after body temperatures.

iii. Based on back–to–back stemplot, it (circle one) **does** / **does not** appear as though drug is lowering high body temperature.

iv. Stem–and–leaf plots are most useful for (circle one)

A. categorical data

B. small quantitative data sets

C. large quantitative data sets

3. *Shapes of graphs for quantitative data.*
Common shapes of *quantitative* histograms given below. Identify previous histograms or stemplots as one of these shapes.

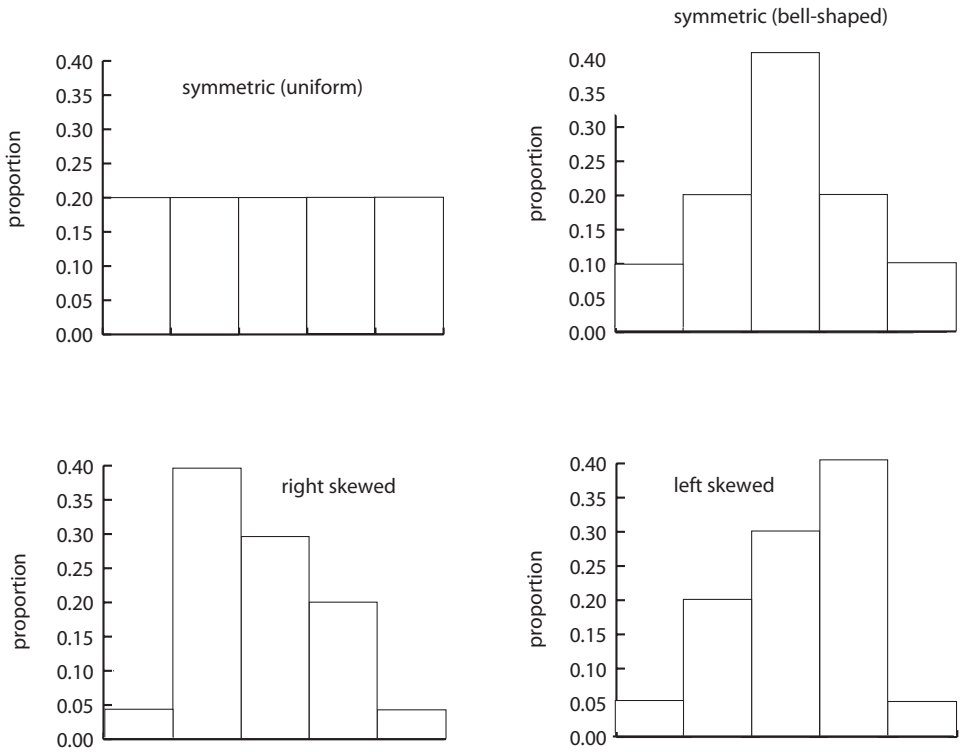


Figure 11.3 (Shapes of graphs)

(a) Figure 11.1 Histogram for patient ages is roughly
**uniform** / **bell-shaped** / **right-skewed** / **left-skewed** / **none**.

(b) Figure 11.2 Histogram for pH Level data is roughly
**uniform** / **bell-shaped** / **right-skewed** / **left-skewed** / **none**.

(c) Stem-and-leaf and split stem-and-leaf of patient ages are both roughly
**uniform** / **bell-shaped** / **right-skewed** / **left-skewed** / **none**.

(d) Back-to-back stem-and-leaf of body temperature data are both roughly
**uniform** / **bell-shaped** / **right-skewed** / **left-skewed** / **none**.

Rotate stem-and-leaf $90^o$ counterclockwise so higher temperatures to right, then check shape.