1.3 Graphical Summaries

e.g. Here are average commute times (in minutes) for the 50 states and DC:

State	Time	State	Time	State	Time	State	Time
AL	22.7	IN	21.2	NE	16.5	SC	23.0
AK	18.9	IA	18.1	NV	21.8	SD	15.2
AZ	23.4	KS	17.5	NH	24.6	TN	23.4
AR	19.9	KY	22.1	NJ	28.5	TX	23.7
CA	26.5	LA	23.3	NM	19.4	UT	19.7
CO	22.9	ME	22.6	NY	30.4	VT	20.3
CT	23.6	MD	30.2	NC	23.2	VA	25.8
DE	22.5	MA	26.0	ND	15.4	WA	24.8
FL	24.8	MI	22.7	OH	22.1	WV	24.7
GA	26.1	MN	21.7	OK	19.1	WI	20.4
HI	24.5	MS	21.6	OR	21.0	WY	17.5
ID	19.5	MO	23.3	PA	23.8	DC	28.4
IL	27.0	MT	16.9	RI	21.8		

Here are the same data, sorted:

15.2 15.4 16.5 16.9 17.5 17.5 18.1 18.9 19.1 19.4

19.5 19.7 19.9 20.3 20.4 21.0 21.2 21.6 21.7 21.8

21.8 22.1 22.1 22.5 22.6 22.7 22.7 22.9 23.0 23.2

23.3 23.3 23.4 23.4 23.6 23.7 23.8 24.5 24.6 24.7

24.8 24.8 25.8 26.0 26.1 26.5 27.0 28.4 28.5 30.2

30.4

Stem-and-Leaf Plot

To make a *stem-and-leaf plot*:

1. separate each observation into
   - *stem*: _____ (or two)
   - *leaf*: _____
2. write stems in a sorted column left of a vertical line
3. for each stem, write its leaves in increasing order to the right

e.g. Make a stem plot of commute times.

Dotplot

To make a *dotplot*:

1. draw a number line

2. draw a dot for each datum above its value on the number line

e.g. Make a dotplot of these data: 4 5 3 4 4 2 5 8 3 4.

Histogram

To make a *histogram*:

1. make a *frequency table*:

    - find min and max
    - choose $\approx$ 5-15 equal-length *class intervals* covering [min, max]
    - count #points in each interval (include _____, exclude _____)
    - (optional) calculate *relative frequency* = (class frequency) / (total #observations)

2. above each interval, draw a rectangle whose height indicates its count (or relative frequency)

e.g. Make a histogram of commute times.

Properties of Histograms

A histogram is

- *symmetric* if its right half is _____ of its left half
- *skewed right* if its right half of values extend _____ than its left half
- *skewed left* if its left half of values extend _____ than its right half
- *unimodal* if it has only _____
- *bimodal* if it has _____ (often indicating two _____)
- divided into equal-area halves by _____
- balanced on _____ (which is pulled toward _____ of a skewed histogram)

Boxplot

To make a *boxplot*:

1. draw a vertical scale
2. draw horizontal lines at _____, _____, and _____
3. draw vertical lines to complete a box
4. find the *interquartile range*, $IQR =$ _____, a measure of spread spanning center _____ of data
5. call points outside $[Q_1 - 1.5 \times IQR, Q_3 + 1.5 \times IQR]$ _____ and plot each one
6. draw lines from box to largest non-outlier and from box to smallest non-outlier

Boxplots are useful for side-by-side comparison of mutliple samples.

e.g. Make a boxplot of commute times, given $Q_1 = 19.9, M = 22.7$, and $Q_3 = 24.6$.