Chapter 1

Picturing Distributions with Graphs

Distribution

- Tells what values a variable takes and how often it takes these values
- Can be a table, graph, or function

Displaying Distributions

- Categorical variables
  - Pie charts
  - Bar graphs
- Quantitative variables
  - Histograms
  - *You don't have to know Stemplots (stem-and-leaf plots)*

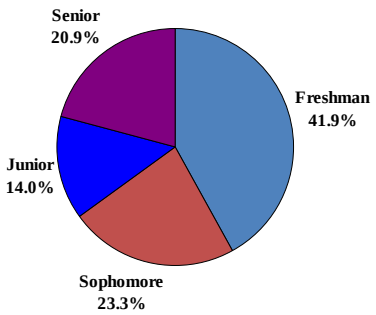
Example: Class Make-up on First Day

Year	***Count***	***Percent***
1	18	41.9%
2	10	23.3%
3	6	14.0%
4	9	20.9%
Total	43	100.1%

"Count" is the # of students in that year

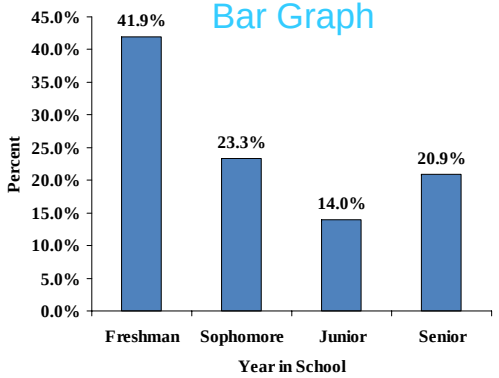
Class Make-up on First Day

Pie Chart



Class Make-up on First Day

Bar Graph



Pie Charts vs. Bar Graphs

Pie Charts:

- better for comparing group sizes to the total
  (e.g. “does a group contains more than half of the total number of individuals?”)

Bar Graphs:

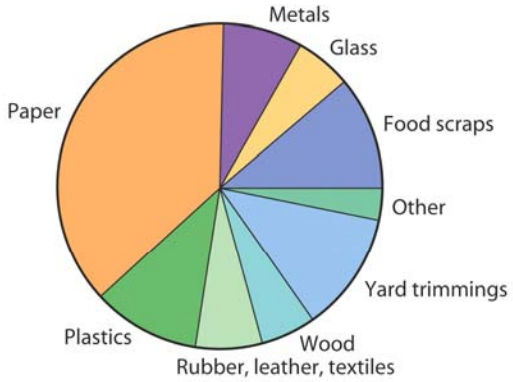
- better for comparing group sizes to each other
- (e.g. “are there more second-year or third-year students in a class?”)

Example: U.S. Solid Waste (2000)

Material	Weight (million tons)	Percent of total
Food scraps	25.9	11.2 %
Glass	12.8	5.5 %
Metals	18.0	7.8 %
Paper, paperboard	86.7	37.4 %
Plastics	24.7	10.7 %
Rubber, leather, textiles	15.8	6.8 %
Wood	12.7	5.5 %
Yard trimmings	27.7	11.9 %
Other	7.5	3.2 %
Total	**231.9**	**100.0 %**

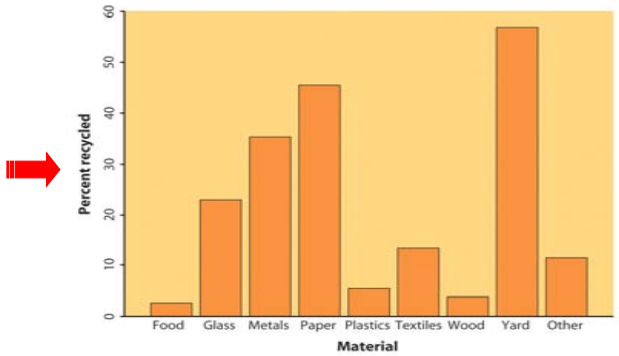
Example: U.S. Solid Waste (2000)

Pie Chart



Example: U.S. Solid Waste (2000)

Bar Graph



WAIT!! Compare Data Table on Slide 8 with Bar Graph on Slide 10. They don't match!!

- Data Table shows biggest group is "paper, paperboard" at 37.4% of total weight, but:
- Bar Graph shows Yard Waste has biggest percentage, almost 60%
- Reason: The bar graph is describing a completely different thing, the % of waste in each category that is recycled
- **A Reminder: Read the axis labels**

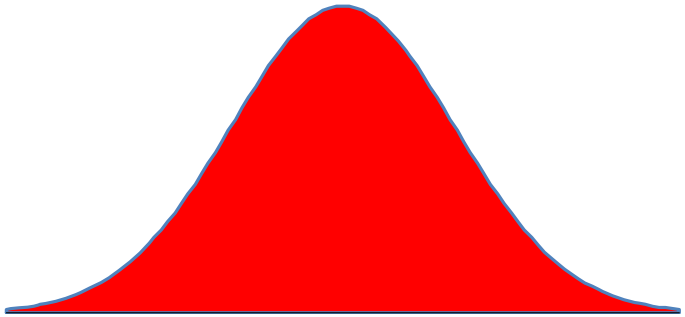
Examining the Distribution of Quantitative Data

- Overall pattern of graph
- Deviations from overall pattern
- Shape of the data
- Center of the data
- Spread of the data (Variation)
- Outliers

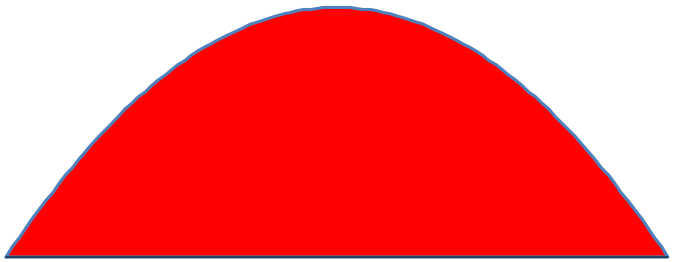
Shape of the Data

- Symmetric
  - bell shaped
  - other symmetric shapes
- Asymmetric
  - right skewed
  - left skewed

Symmetric Bell-Shaped

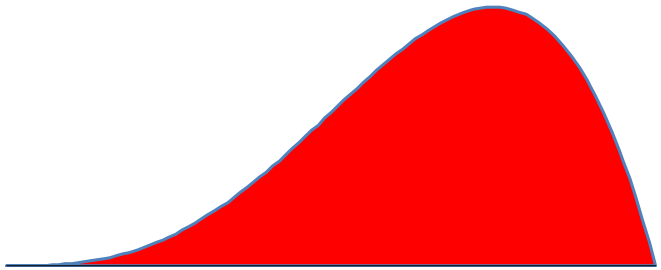


Symmetric Mound-Shaped

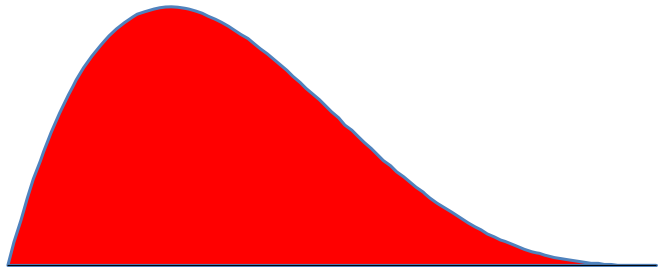


Symmetric Uniform

Asymmetric
Skewed to the Left



Asymmetric
Skewed to the Right



Outliers

- Extreme values that fall outside the overall pattern
  - May occur naturally
  - May occur due to error in recording
  - May occur due to error in measuring
  - Observational unit may be fundamentally different

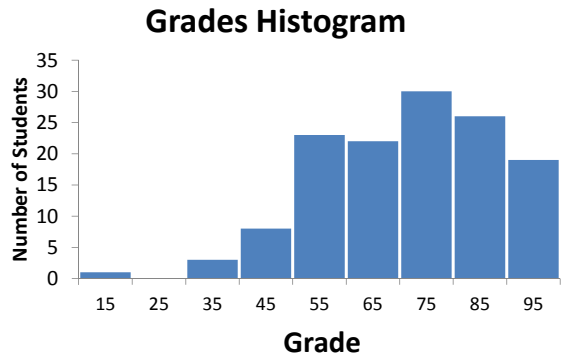
Histograms

- For quantitative variables that take many values
- Divide the possible values into *class intervals* (we will only consider equal widths)
- Count how many observations fall in each interval (may change to percent of total obs'ns)
- Draw histogram (like a bar graph, but with less space between bars)

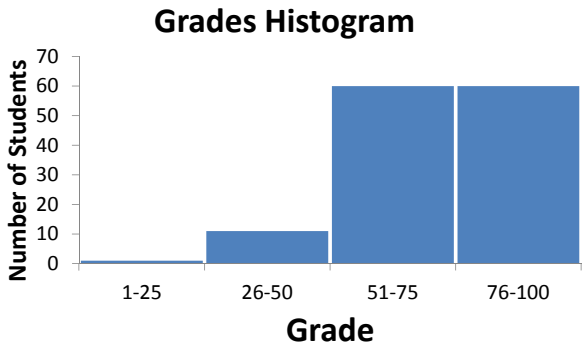
Histograms: Class Intervals

- How many intervals?
  - One rule is to calculate the square root of the sample size, and round up. (e.g. if n=132, $\sqrt{n}$ = 11.5, could use 11 or 12 groups)
- Size of intervals?
  - Divide range of data (max–min) by number of intervals desired, and round to convenient number
- Pick intervals so each observation can only fall in exactly one interval (no overlap)

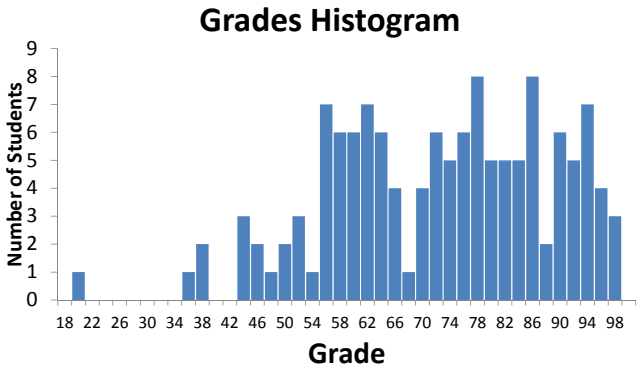
Histogram using preferred choice of bins



Histogram using too few bins



Histogram using too many bins



reasoning for histogram preference

preferred histogram:

- square root rule for bins suggested 12 bins, but:
- choose 9 bins instead so the interval width is 10, a round number, easier to interpret

too few bins:

- the pattern is clear, but too much information is lost

too many bins:

- contains the most details about the original data, but is harder to see the pattern

Time Plots

- A time plot shows behavior over time.
- Time is always on the horizontal axis, and the variable being measured is on the vertical axis.
- Look for an overall pattern (*trend*), and deviations from this trend. Connecting the data points by lines may emphasize this trend.
- Look for patterns that repeat at known regular intervals (seasonal variations).

Average Tuition (Public vs. Private)