Summarize Categorical Variables with

Counts—how many in each category ?

- convert to percent

Ex1

Frequency distribution

|  |  |  |  |
| --- | --- | --- | --- |
| Hair colour | count | Relative |  |
| Black | 23 | 23/36=64% |  |
| Brown | 7 | 7/36=19% |  |
| Red | 0 | 0/36=0% |  |
| Blond | 4 | 4/36=11% |  |
| Other | 2 | 2/36=6% |  |
| Total | Total 36 | Total 100% |  |

Relative frequency distribution

Principle

1 Small number of data set (table probably better)

2 ink of the graph

3 goal of the graph-store not to info

4 everything needs to be proper labeled

Graph types:

Dotplot:

-- Use horizontal axis for your values

-- put one dot above each data value

Things to look for in a graph

Distribution

Location – center or average

--median(middle valve)/mean ( usual average)

Spread – how narrow/wide around center

--range/interquartile range/standard deviation

Shape

--symmetric or not?

--skew or direction

--bell-shape, uniform

Outliers: unusual values

Stem-and-leaf plot

Each measurement breaks into stem and leaf

Histogram

-bar of graph of clumped quantitative data

Ex

|  |  |  |
| --- | --- | --- |
| Category | Count |  |
| 360-369 | 2 |  |
| 370-379 | 2 |  |
| 380-389 | 4 |  |
| 390-399 | 4 |  |
| 400-409 | 5 |  |
| 410-419 | 1 |  |
| 420-429 | 1 |  |
| 430-439 | 1 |  |

-split interval into equal-sized subintervals (about6-15)

-count # in each

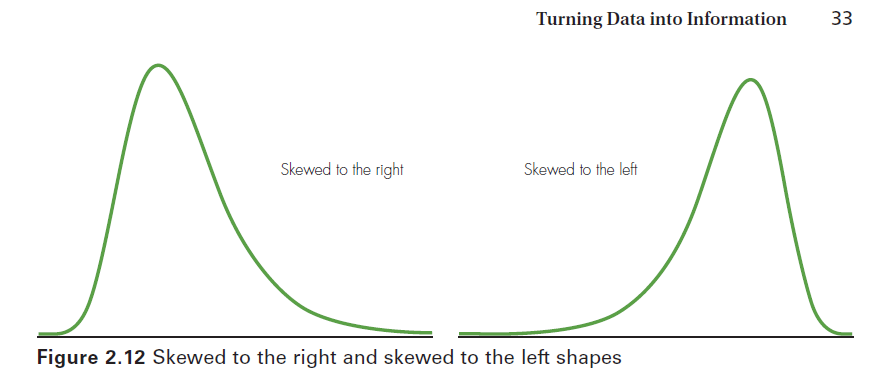
-make bar graph

Describing shape

Symmetric or not?

Sit the same(ish) on both sides

Skewed—more spread out on one side



A symmetric graph could be bell-shaped :

-normal distribution

-distribution

Unimodal

How many peaks?

Bimodal

Multimodal

