

1. A frequency distribution of qualitative data is a listing of the distinct values and their frequencies.
2. The ratio of the frequency to the total number of observations is called relative frequency.
3. The relative frequency is always between 0 and 1, for every category. Sum of relative frequencies of all categories is always equal to 1.
4. To create frequency distribution in google sheet, use pivot table option.
5. A pie chart is a circle divided into pieces/wedges proportional to the relative frequencies of the qualitative data.
6. A bar chart displays the distinct values of the qualitative data on a horizontal axis and the relative frequencies (or frequencies or percentages) of those values on a vertical axis. The frequency/relative frequency of each distinct value is represented by a vertical bar whose height is equal to the frequency/relative frequency of that value. The bars should be positioned such that they do not touch each other.
7. Order of categories doesn't matter in bar chart.
8. All bars in a bar chart should have equal width.
9. Bar charts can be vertical (column chart) or horizontal.
10. When trying to compare parts of a whole, use pie-chart. When trying to compare things between different groups, use bar-chart.
11. Pareto charts are bar charts, where categories appear in the descending order of their relative frequencies. Sometimes, they are used to represent categories in their ascending order too.
12. Pareto charts cannot be used when the plotted variable is ordinal. In this case, always order the bars by rank of the variable (For example, Small, Medium, Large)
13. If there are too many groups to be represented in a bar chart, represent the top few groups in the first few bars, and then group the rest as 'Others' and represent in a single bar.
14. Know your purpose and choose table/graph appropriately.
15. Always label your charts to avoid ambiguity.
16. Display of data must obey a fundamental rule called the *area principle*, which says that the area occupied by a part of the graph should correspond to the amount of data it represents. Violations of the area principle are a common way to mislead with statistics.
17. Another common violation is when the baseline of a bar chart is not at zero. Such graphs are called *truncated* graphs. Truncated graphs should introduce a y-axis break, to make this meaningful.
18. Final violation is due to the rounding-off errors, due to which the relative frequencies do not add up to 1.
19. When the data type is nominal, we can use mode to describe it. When the data type is ordinal, we can use mode/median to describe. These are collectively called measures of central tendency.
20. Mode of a categorical variable is the most common category, the category with the highest frequency. The mode labels the longest bar in a bar chart, widest slice in a pie chart. In a pareto chart, the mode is the first category shown.
21. Some datasets might be bi-modal or even multi-modal, depending on how many modes they have.
22. Median of an ordinal variable is the category of the middle observation of the sorted values. If there are an even number of observations, choose the category on either side of the middle of the sorted list as the median.