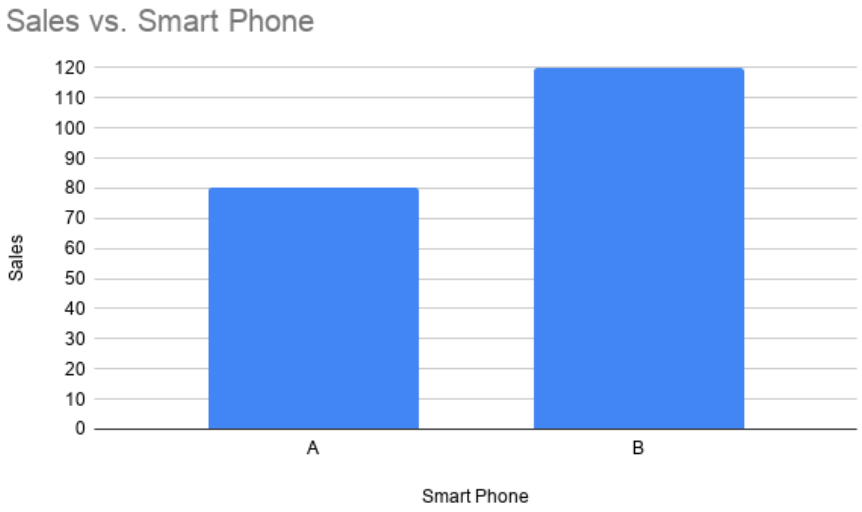
Statistics for Data Science - 1

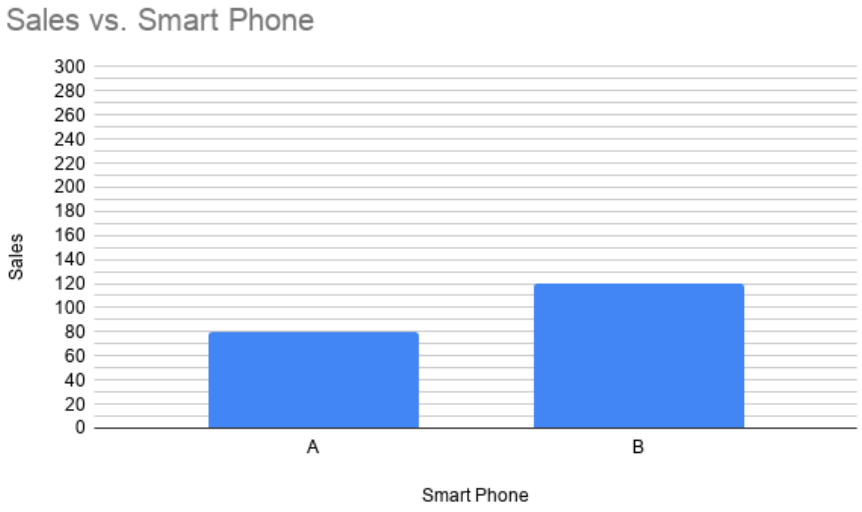
FAQ

Week 2

1. **When to use pie chart and bar chart?**
   Use a bar chart to show the frequencies of a categorical variable. Order the categories either alphabetically or by size. The bars can be oriented either horizontally or vertically.
   Use a pie chart to show the proportions of a categorical variable. Arrange the slices (if you can) to make differences in the sizes more recognizable. A pie chart is a good way to show that one category makes up more than half of the total.
   Consider this situation: A manager has partitioned the company's sales into five cities: Gurgaon, Pune, Mumbai, New Delhi, and Chennai. What graph would you use to make these points in a presentation for management?
   a. A figure that shows that slightly more than half of all sales are made in the Chennai city.
   b. A figure that shows that sales topped 12 crores in every city.
   Use a pie chart for part a. Pie charts emphasize the breakdown of the total into pieces and make it easy to see that more than half of the total sales are in the Chennai city. For part b, use a bar chart because bar charts show the values rather than the relative sizes. Every bar would be long enough to reach past a grid line at 12 crores.

2. **What do you mean by manipulation of y-axis?**
Expanding or compressing the scale on a graph, that can make changes in the data seem less significant than they actually are, is known as the manipulation of y-axis. For example if we represent the number of sales of smart phone A and B of a local shop, from the first figure we are getting the information that a significant amount of sales is





being done of both the smart phones but from the second figure it seems that the sales is very low of the smart phone A and B. So, the second graph is misleading because it has manipulated y-axis.

3. **What do you mean by truncated graph?**
   Omitting baselines, or the axis of a graph, is one of the most common ways data is manipulated in graphs, known as a truncated graph. This misleading tactic is frequently used to make one group look better than another.

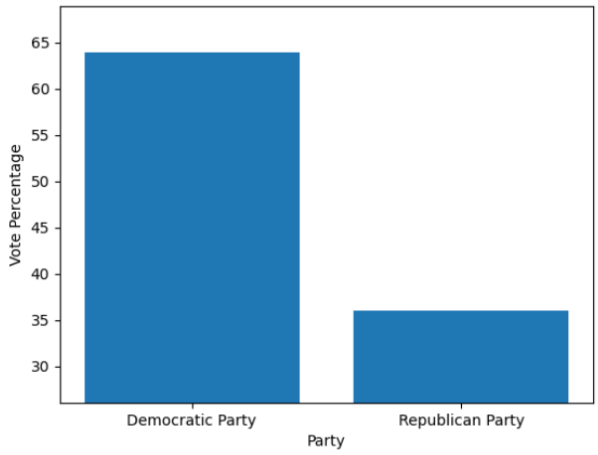


Figure 2.4.5: Share of the votes in an election in the United States of America

For example if we refer the image from AQ 2.4.5, from the length of the bar we observe that Republic party voting percentage is less than half of the Democratic party but if consider the actual number this is not the case.