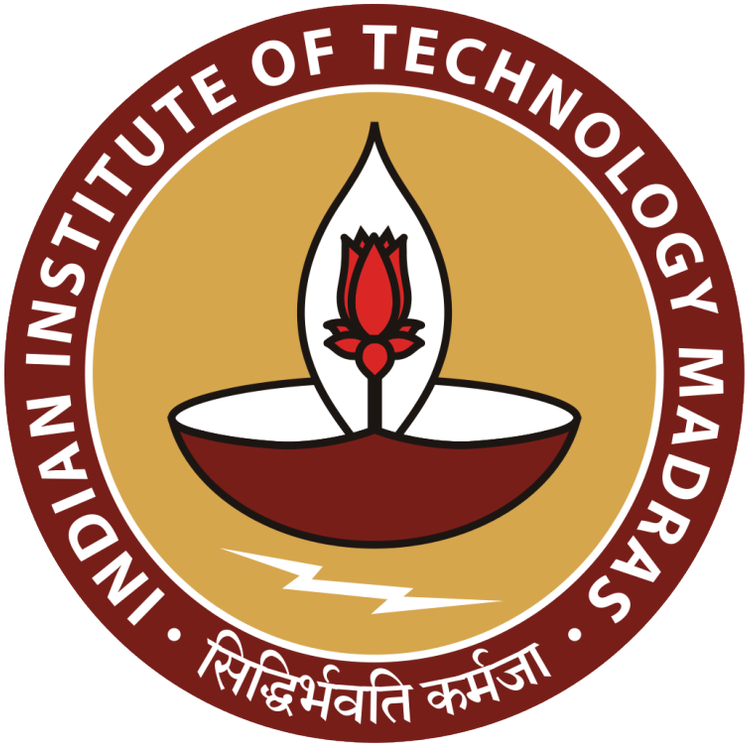


IIT Madras

ONLINE DEGREE

Statistics for Data Science -1

Describing Categorical Data- Single Variable

Usha Mohan

Indian Institute of Technology Madras

Know your purpose

- Have a purpose for every table or graph you create
  - Choose the table/graph to serve the purpose.

Know your purpose

- Have a purpose for every table or graph you create
  - Choose the table/graph to serve the purpose.
- Pie charts are best to use when you are trying to compare parts of a whole.

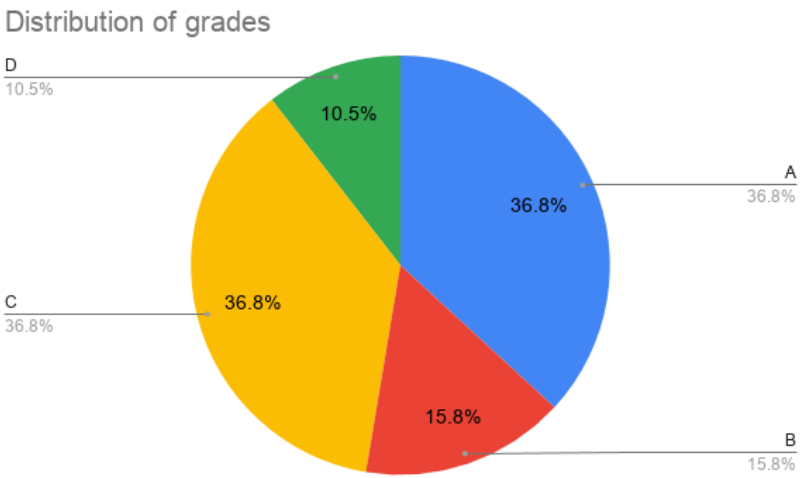
Know your purpose

- Have a purpose for every table or graph you create
  - Choose the table/graph to serve the purpose.
- Pie charts are best to use when you are trying to compare parts of a whole.
- Bar graphs are used to compare things between different groups.

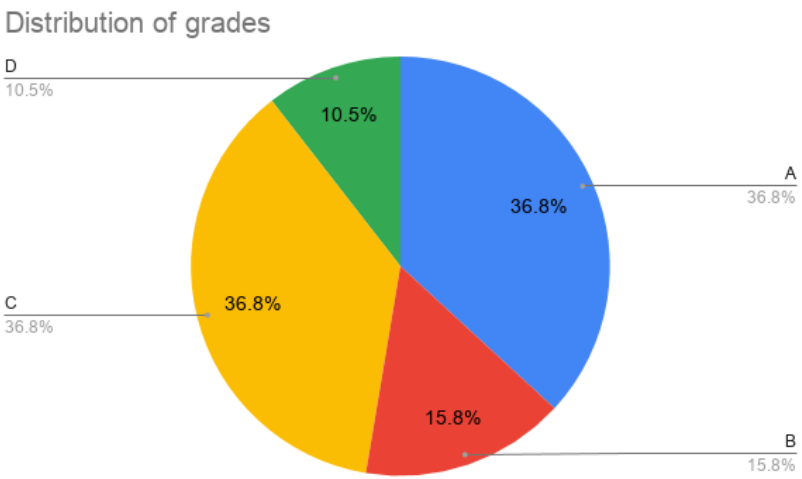
Label your data

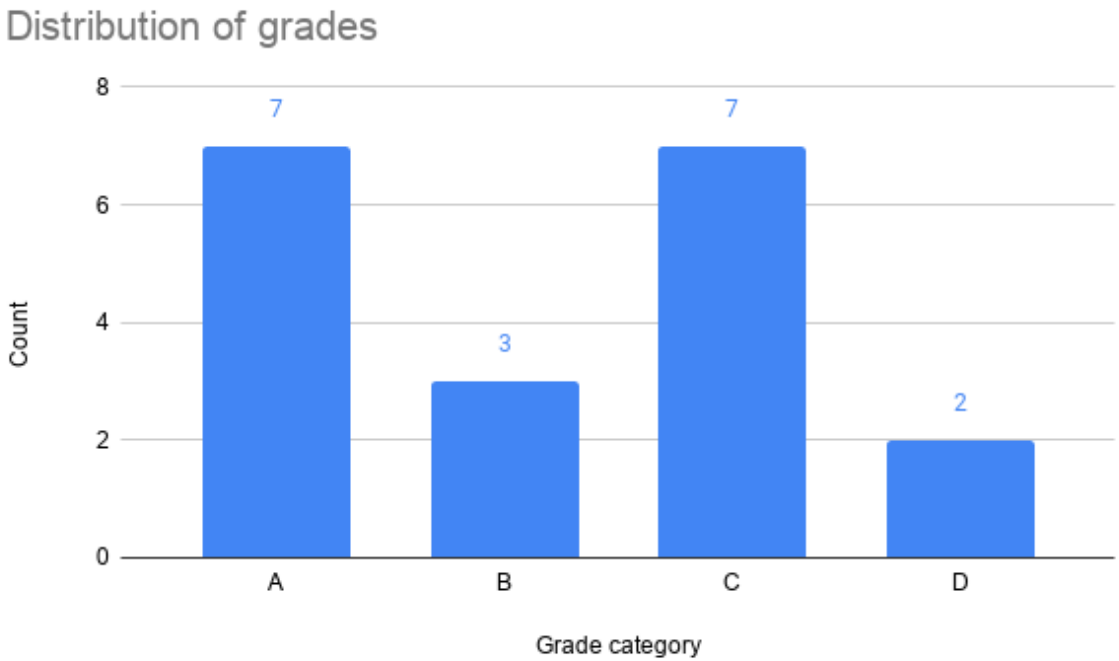
- Label your chart to show the categories and indicate whether some have been combined or omitted.
- Name the bars in a bar chart.
- Name the slices in a pie chart.
- If you have omitted some of the cases, make sure the label of the plot defines the collection that is summarized.

Label your data

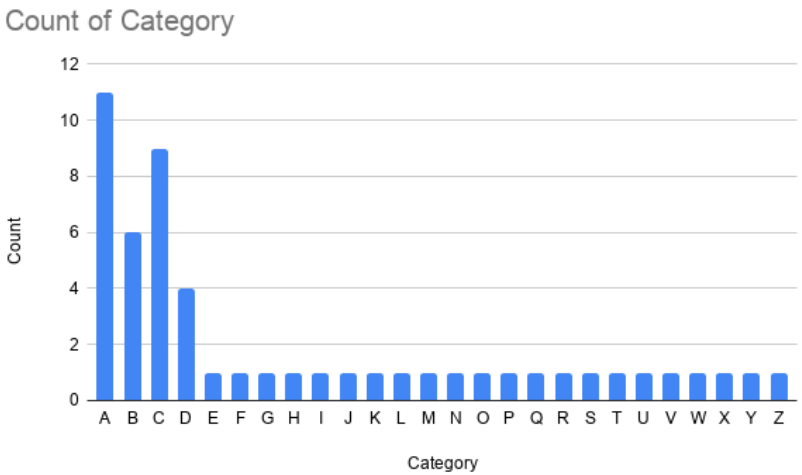


Label your data

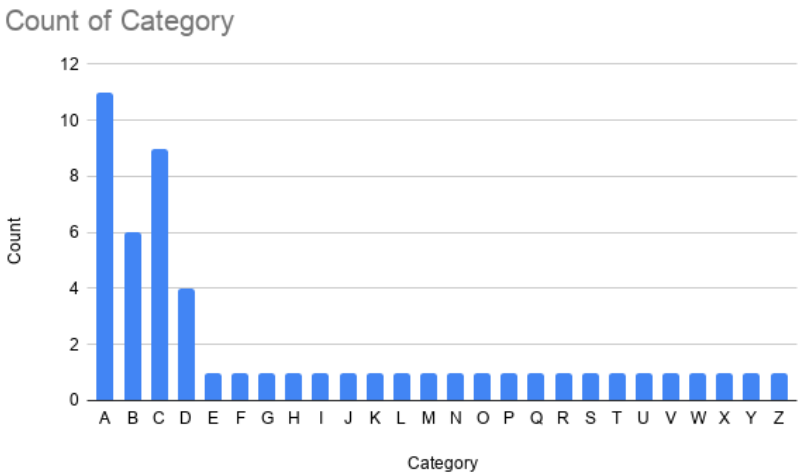




Many categories

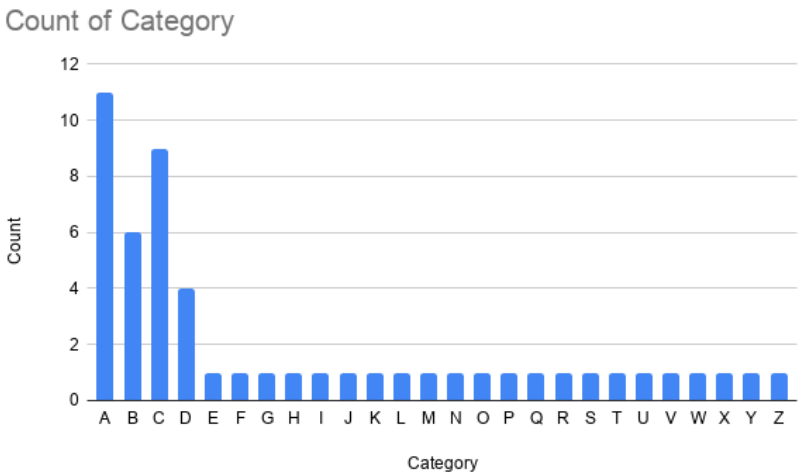


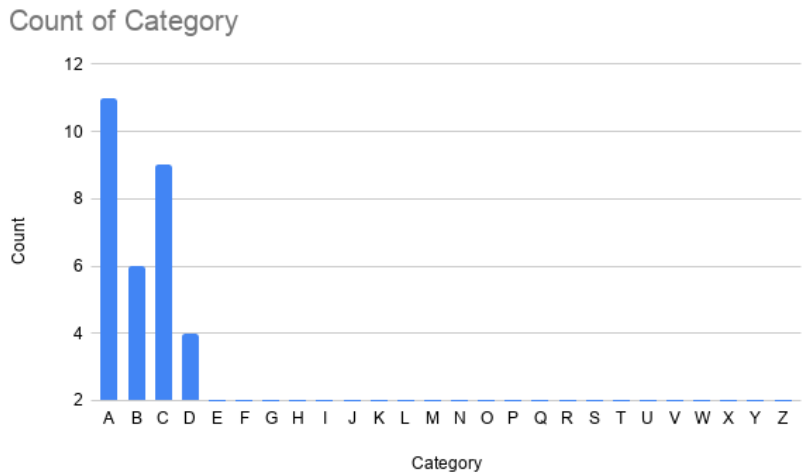
Many categories



A bar chart or pie chart with too many categories might conceal the more important categories. In some case, grouping other categories together might be done.

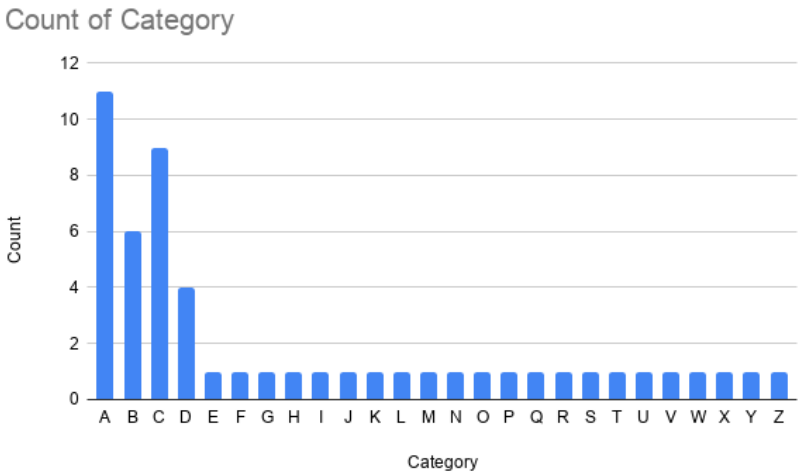
Many categories

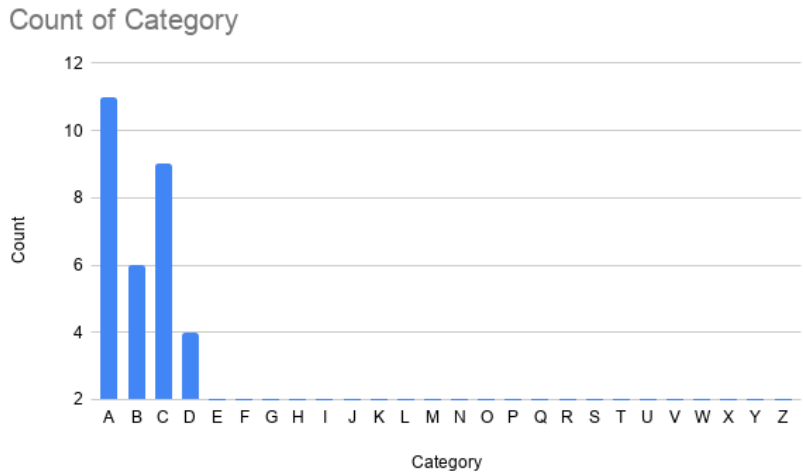




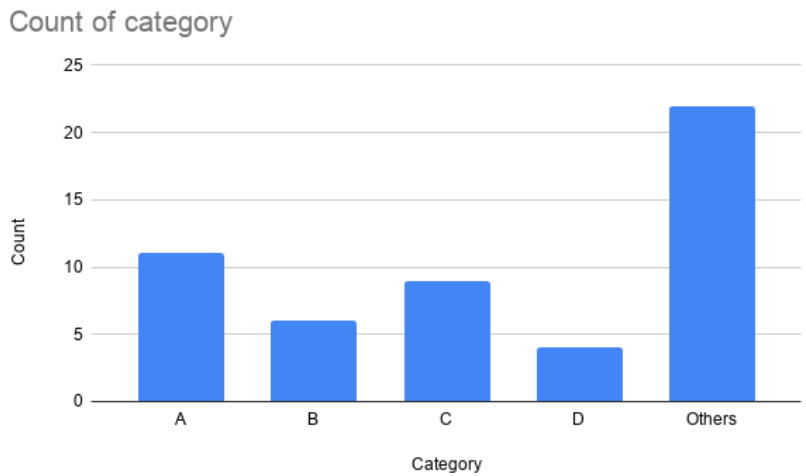
A bar chart or pie chart with too many categories might conceal the more important categories. In some case, grouping other categories together might be done.

Many categories





A bar chart or pie chart with too many categories might conceal the more important categories. In some case, grouping other categories together might be done.



The area principle

- Displays of data must obey a fundamental rule called the area principle[1].

[1] Stine, Robert, and Dean Foster. Statistics for Business: Decision Making and. Addison-Wesley, 2011.

The area principle

- Displays of data must obey a fundamental rule called the area principle[1].
- The area principle says that the area occupied by a part of the graph should correspond to the amount of data it represents.

[1] Stine, Robert, and Dean Foster. Statistics for Business: Decision Making and. Addison-Wesley, 2011.

The area principle

- Displays of data must obey a fundamental rule called the area principle[1].
- The area principle 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.

[1] Stine, Robert, and Dean Foster. Statistics for Business: Decision Making and. Addison-Wesley, 2011.

Misleading graphs: violating area principle

- Decorated graphics: Charts decorated to attract attention often violate the area principle[2]

[2] Stine, Robert, and Dean Foster. Statistics for Business: Decision Making and. Addison-Wesley, 2011.

Misleading graphs: violating area principle

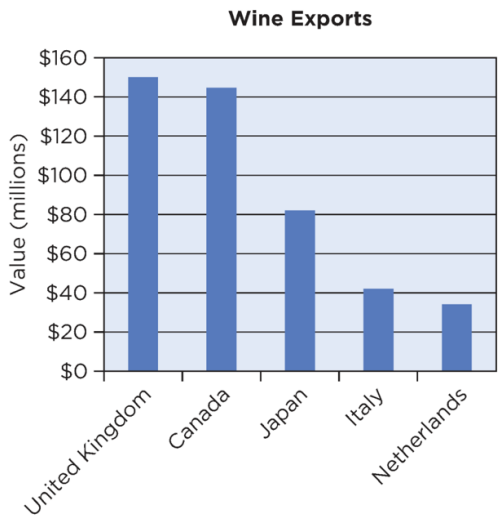
- Decorated graphics: Charts decorated to attract attention often violate the area principle[2]

- No baseline and the chart shows bottles on top of labeled boxes of various sizes and shapes.

[2] Stine, Robert, and Dean Foster. Statistics for Business: Decision Making and. Addison-Wesley, 2011.

Misleading graphs: violating area principle

- Decorated graphics: Charts decorated to attract attention often violate the area principle[2]



- No baseline and the chart shows bottles on top of labeled boxes of various sizes and shapes.
- Obeys area principle and accurate

[2] Stine, Robert, and Dean Foster. Statistics for Business: Decision Making and. Addison-Wesley, 2011.

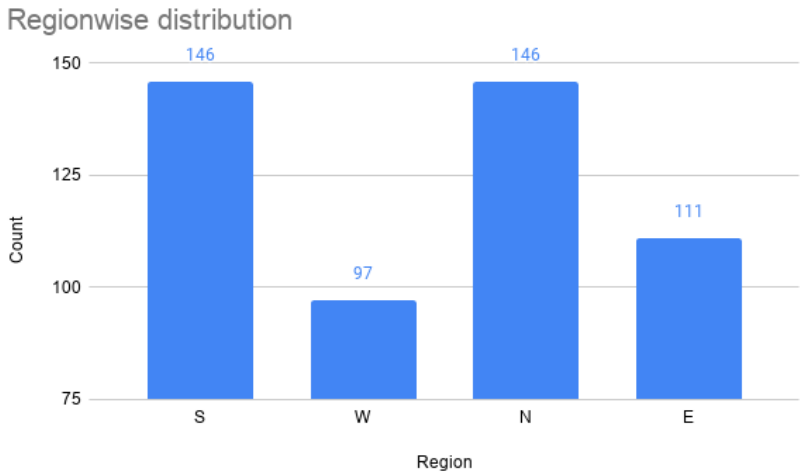
Misleading graphs: truncated graphs

Misleading graphs: truncated graphs

- Another common violation is when the baseline of a bar chart is not at zero.

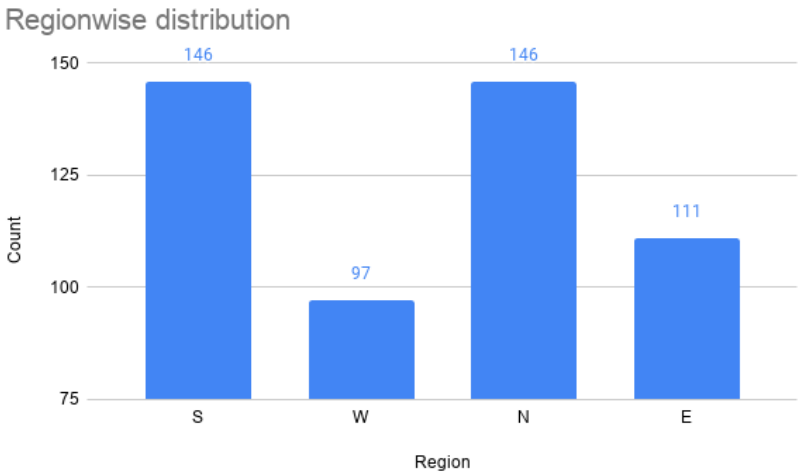
Misleading graphs: truncated graphs

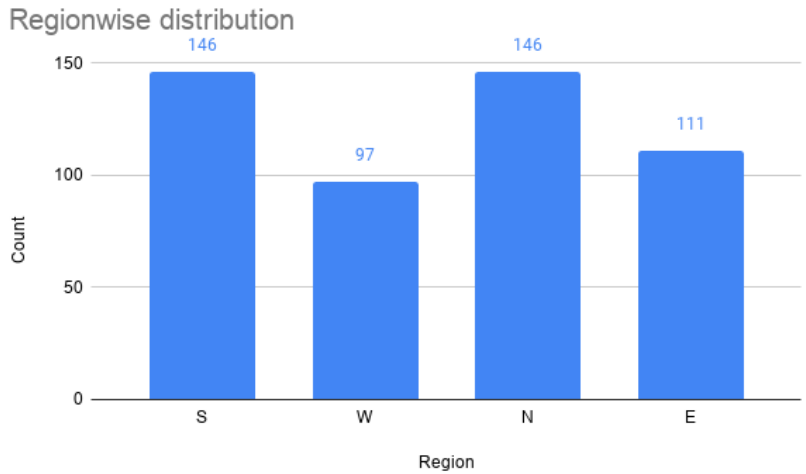
- Another common violation is when the baseline of a bar chart is not at zero.



Misleading graphs: truncated graphs

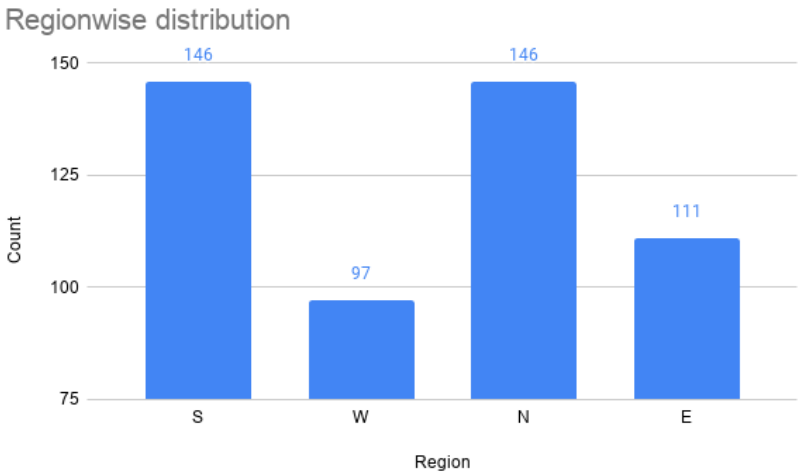
- Another common violation is when the baseline of a bar chart is not at zero.

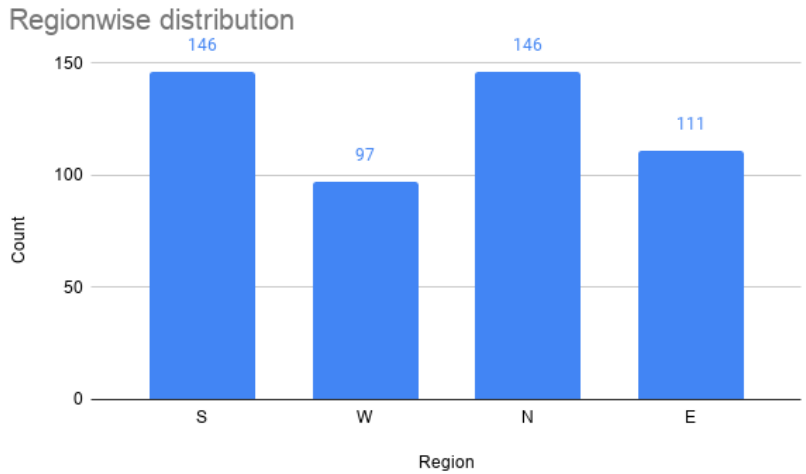




Misleading graphs: truncated graphs

- Another common violation is when the baseline of a bar chart is not at zero.





Left graph exaggerates the number coming from the South and North. Graph on right shows same data with the baseline at zero.

Indicating a y- axis break

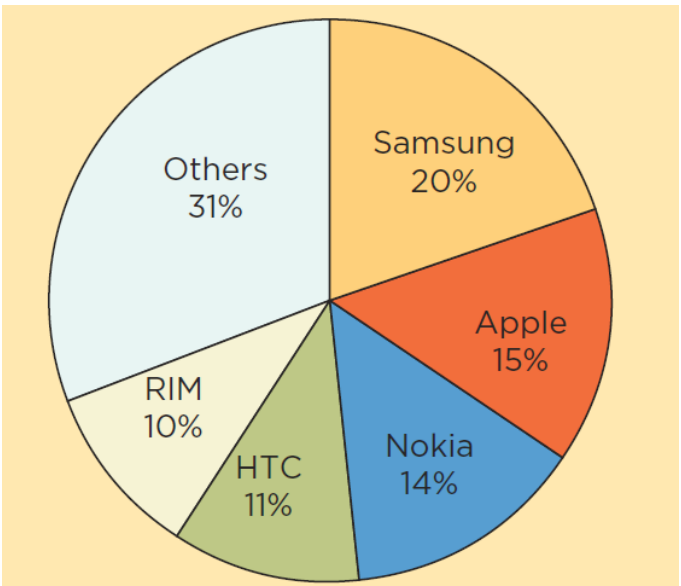
Round-off errors

- Important to check for round-off errors.

[3] Stine, Robert, and Dean Foster. Statistics for Business: Decision Making and. Addison-Wesley, 2011.

Round-off errors

- Important to check for round-off errors.
- When table entries are percentages or proportions, the total may sum to a value slightly different from 100% or 1. This might result in a pie chart where the total does not add up[3].



[3] Stine, Robert, and Dean Foster. Statistics for Business: Decision Making and. Addison-Wesley, 2011.

Sectional summary

1. Know your purpose and choose table/graph appropriately
2. Label your charts
3. Handle multiple categories appropriately.
4. Respect area principle
   - 4.1 Avoid overly decorated graphs
   - 4.2 Avoid truncated graphs- use special symbols to indicate vertical axis has been modified.
   - 4.3 Check for round-off errors