Almost to the more fun parts!

Unit 2 – Graphical Summaries, Day 2

Your Ahead-of-the-game Professor Colton

LU 2, Day 2 - Outline

Graphs for Categorical Data

- Frequency Tables for Categorical Variables
- Pie Charts
- Bar Graphs
- Pareto Chars
- Pictograms
- Key Elements to a Good Graph

Frequency Tables for Categorical Variables

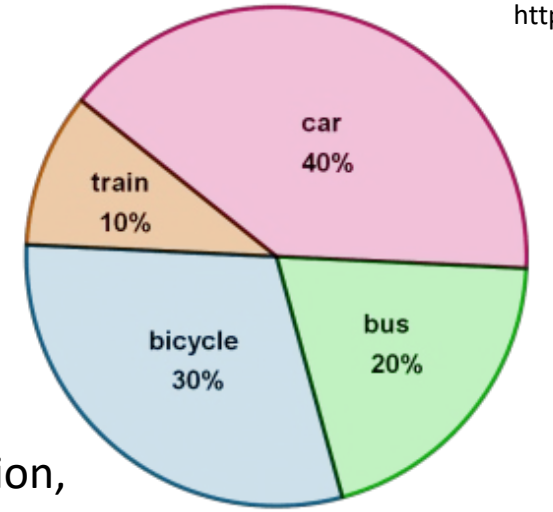
Example: Computer Sales

- The following table summarizes the type of computer sold to the last 50 customers at a computer retailer.
- The variable being studied, type of computer, is categorical because the values of the variable are categories (desktop, laptop, notebook, and tablet).

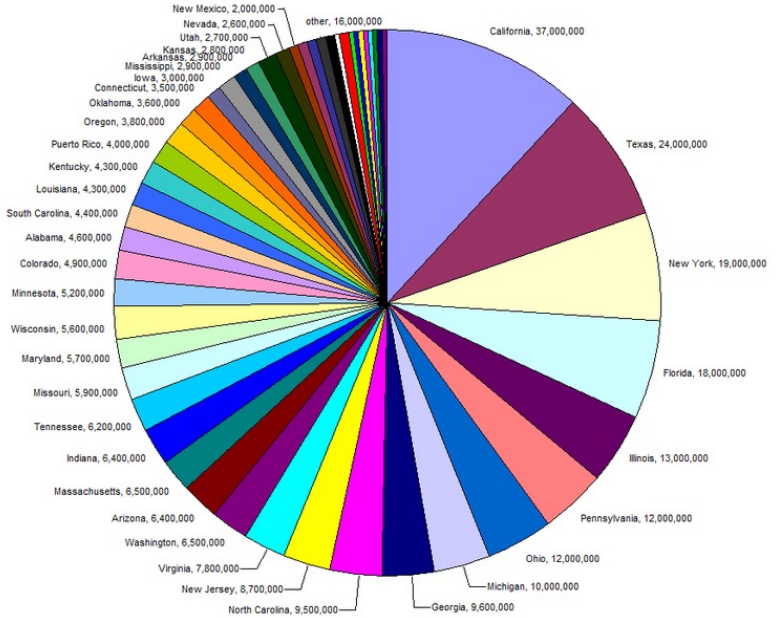
Type of Computer	Frequency	Percent
Desktop	11	11/50 = 22%
Laptop	23	23/50 = 46%
Notebook	9	9/50 = 18%
Tablet	7	7/50 = 14%

- To visually summarize the distribution of a categorical variable, there are two main types of graphs that can be drawn.
  - **Pie charts and Bar graphs**.

Pie Charts



https://www.mathemania.com/lesson/pie-chart/



https://commons.wikimedia.org/wiki/File:Pie_chart_of_US_population_by_state.png

Pie Chart

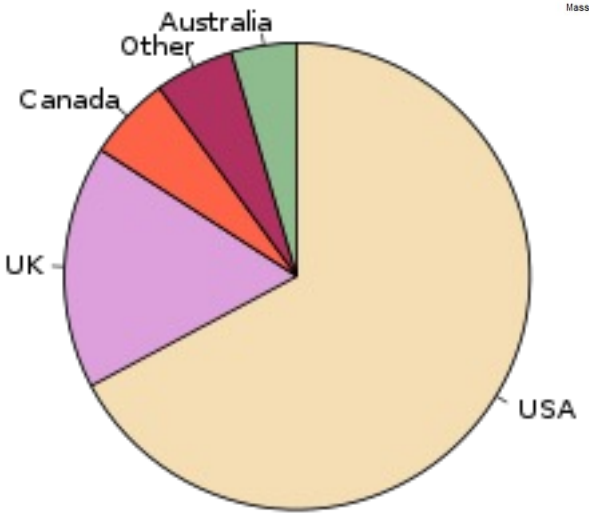
- Shows **one qualitative** variable.
- Separates parts of a whole into categories.
  - Each piece displays the category, the proportion, and usually the amount.
- The parts should all add up to 100%.

Advantage

- Simple and common.

Disadvantage

- Harder to compare area than heights!
- Not helpful when there are lots of categories!
  - Can make an ‘Other’ category and combine ones with smaller proportions, but be careful!



https://en.wikipedia.org/wiki/Pie_chart

Bar Graphs

Bar Graph

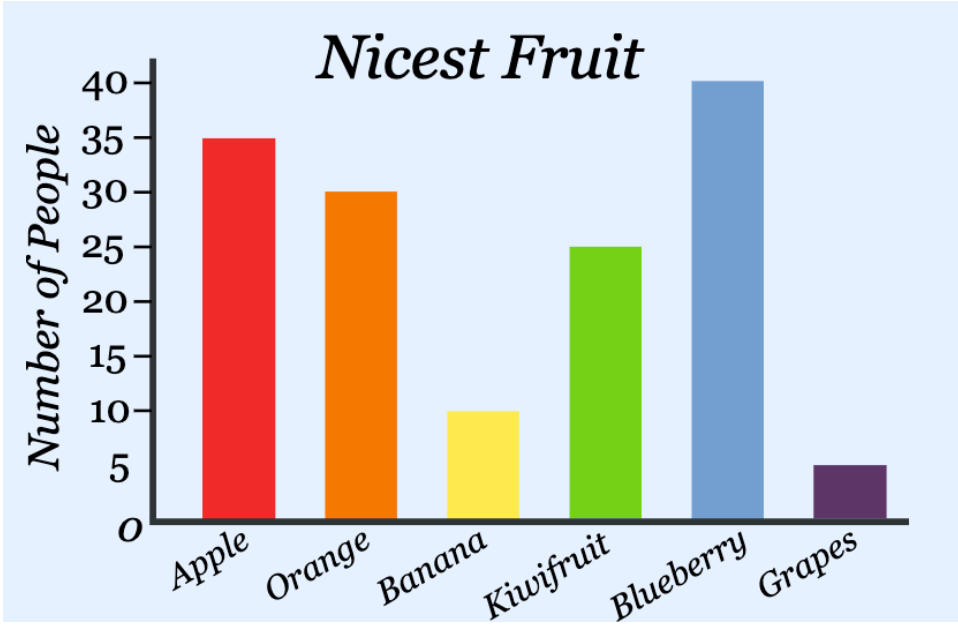
- Shows **one qualitative** variable.
- Simple and common and easy to read!
- Shows differences in frequencies well and the mode (most common group) well!

Features

- Plots vertical bar for each category.
- Height is the **frequency** or **relative frequency**.
  - Sum of the heights equals the sample size.

Differences from a Histogram

- GAPS between bars.
  - X-axis is NOT a number scale, so this makes sense
- SHAPE is NOT a thing with bar graphs.
  - Order of categories is MEANINGLESS.
  - Appearances obviously will change as reorder cats.

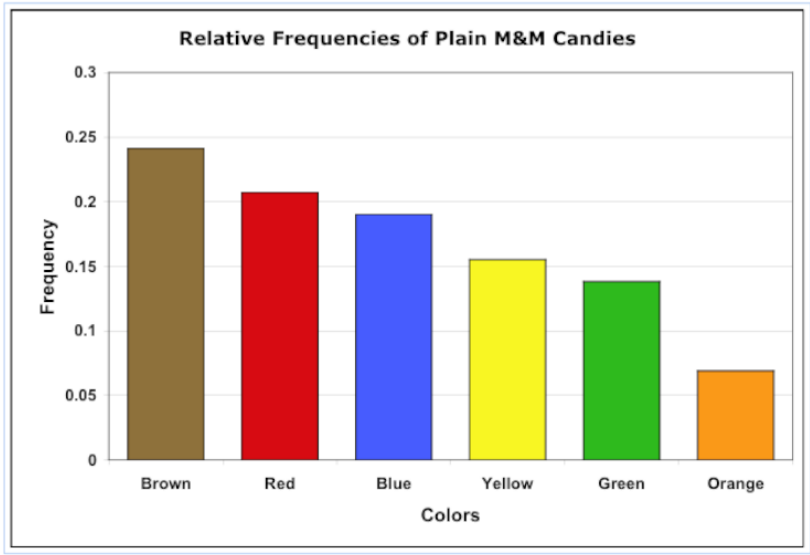


https://www.mathsisfun.com/data/bar-graphs.html

Pareto Charts

Pareto Chart

- Just a bar graph with the categories arranged in decreasing order!
  - Tallest (largest frequency) on the left and then the next tallest to the right and so on…
- Shows which categories / areas are more important.

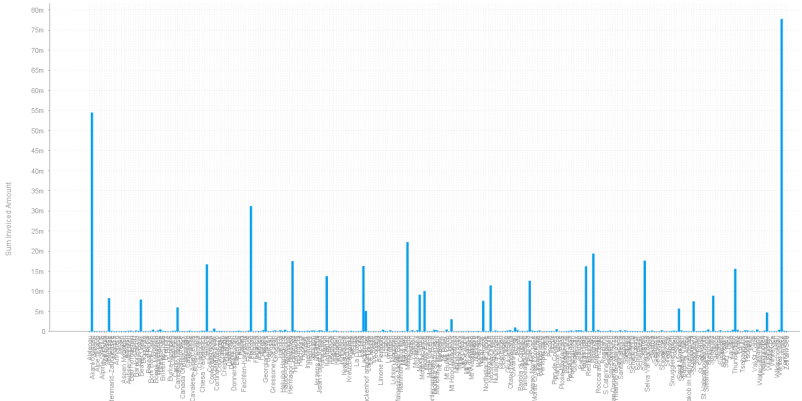


https://www.process.st/pareto-chart/

Bad Bar Graphs

Bad Bar Graphs

- Too many cats

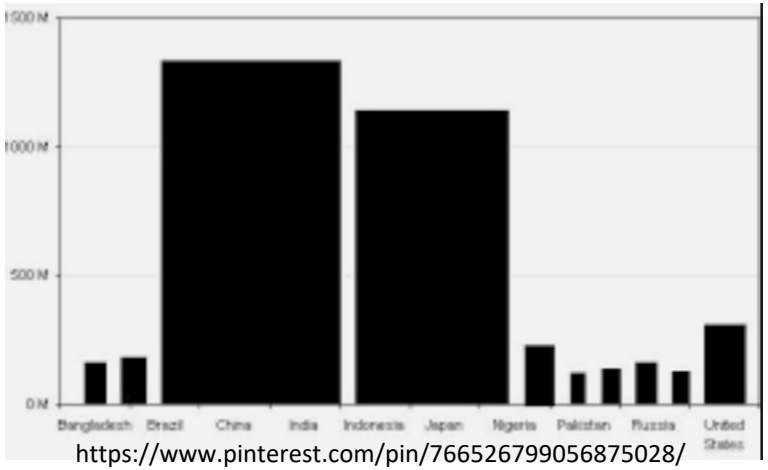


http://www.deepdive.co.za/index.php/item/28-solved-charts-displaying-too-many-categories.html

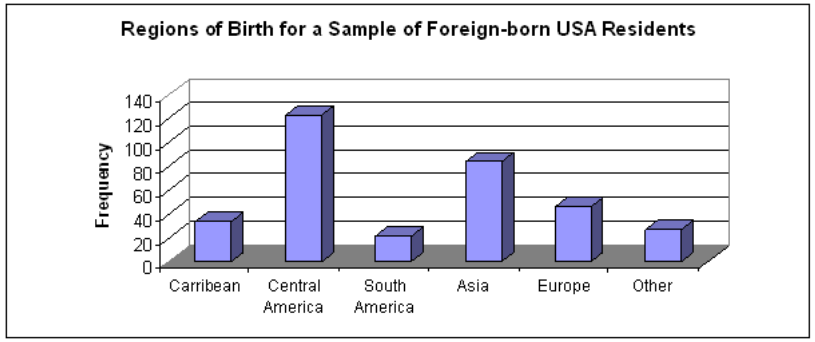
Misleading Bar Graphs

- Very common….
- Examples:
  - Truncating
  - Unnecessary modifications

https://www.redw.com/blog/2020/01/27/an-introduction-to-misleading-charts-when-good-data-is-plotted-badly/



https://www.pinterest.com/pin/766526799056875028/

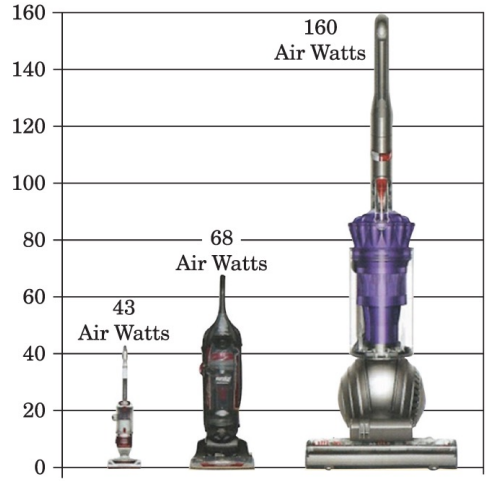


https://www2.southeastern.edu/Academics/Faculty/dgurney/Math241/StatTopics/GraphGuide2.htm

Pictograms

Pictograms

- The **pictogram** replaces the bars of the bar graph with pictures to provide visual interest.
- However, since both the height and width change, the viewer focuses on the area of the picture which can be misleading.
  - This violates the **area principle**
- For this reason, we typically approach the pictogram with caution (i.e. don't make them).



Moore/Notz, *Statistics: Concepts and Controversies*, 10e, © 2020 W. H. Freeman and Company

Canadian Coins

- By the numbers, the value of the Canadian dollar is in 1980 ($1.00) is double of 1995 ($0.50).
- By the looks, the 1980 coin is roughly 4 times larger than the 1995.
- **Very misleading**.

Key Elements to a Good Graph

1) Use Labels and Legends
   - Label axes, give units, provide source of data.

2) Make Data Stand Out
   - Don’t add too much clutter.

3) Pay Attention to What the Eye Sees
   - Avoid pictograms, pay attention to scales.

PROBLEM SESSION!!!!!!!!!!

Problem #1

As part of the human resource group of your company you are asked to summarize the educational levels of the 512 employees in your division. From company records you find that 164 have no college degree (None), 42 have an associate's degree (AA), 225 have a bachelor's degree (BA), 52 have a master's degree (MA), and 29 have PhDs. For the educational level of your division:

a) Make a frequency table.

b) Make a relative frequency table.

Problem #3

As part of the human resource group of your company you are asked to summarize the educational levels of the 512 employees in your division. From company records you find that 164 have no college degree (None), 42 have an associate's degree (AA), 225 have a bachelor's degree (BA), 52 have a master's degree (MA), and 29 have PhDs. For the educational level of your division:

a) Make a bar chart using counts on the y-axis.

b) Make a relative frequency bar chart using percentages on the y-axis.

c) Make a pie chart.

Problem #5

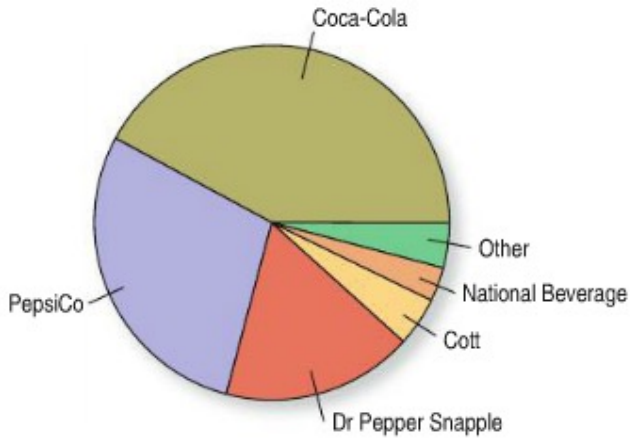
As part of the human resource group of your company you are asked to summarize the educational levels of the 512 employees in your division. From company records you find that 164 have no college degree (None), 42 have an associate's degree (AA), 225 have a bachelor's degree (BA), 52 have a master's degree (MA), and 29 have PhDs. For the educational level of your division:

a) Write two to four sentences summarizing the distribution.

b) What conclusions, if any, could you make about the educational level at other companies?

Problem #15

An article in *The Wall Street Journal* (March 18, 2011) reported the 2010 U.S. market share of leading sellers of carbonated soft drinks, summarized in the foll

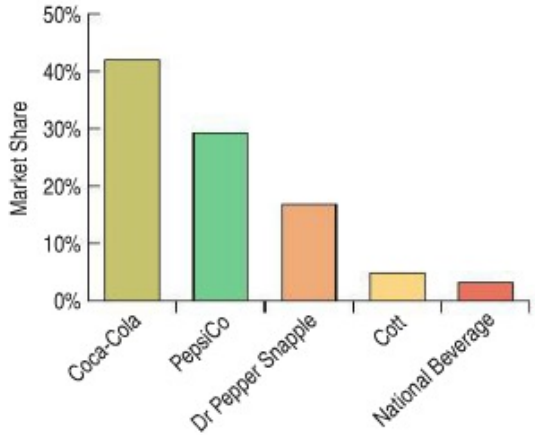


a) Is this an appropriate display for these data?

b) Which company had the largest share of the market?

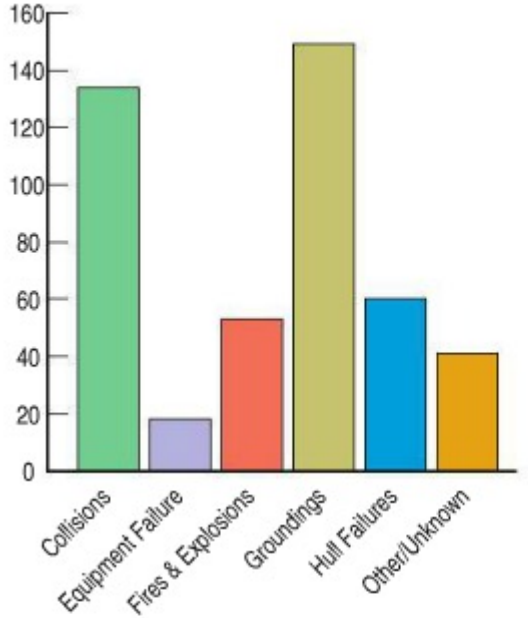
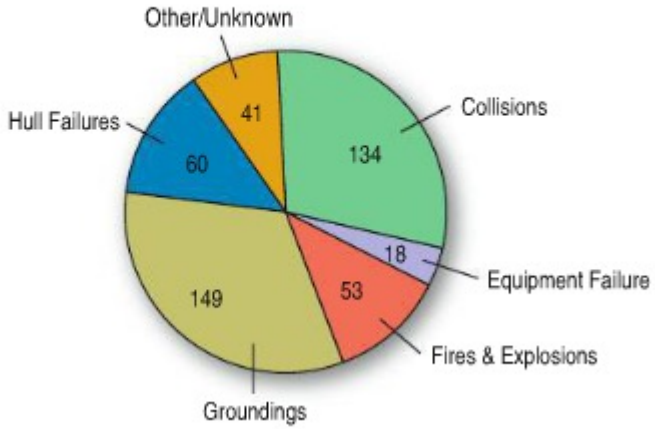
Problem #17

An article in *The Wall Street Journal* (March 18, 2011) reported the 2010 U.S. market share of leading sellers of carbonated soft drinks, summarized in the following bar chart:



a) Compared to the pie chart, which is better for displaying the relative portions of market share?

b) What is missing from this display that might make it somewhat misleading?

Problem #25

Data from the International Tanker Owners Pollution Federation Limited (www.itopf.com) gives the cause of spillage for 455 large oil tanker accidents from 1970 to 2012. Write a brief report interpreting what the displays show. Is a pie chart an appropriate display for these data? Why or why not?