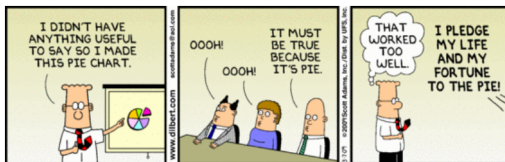


# Data Visualization

BUS 230: Business and Economic Research and  
Communication

- Purpose of graphs and charts is to show a picture that can enhance a message, or quickly communicate a message, as compared to reporting descriptive statistics.
- Keep charts as simple as possible. Unnecessary ink like fancy formatting, pictures, clip art, etc., can distract an audience.
- Make sure charts communicate an *honest message*.
- We'll review some common chart types:
  - Pie charts
  - Bar charts
  - Line plots
  - Area charts
  - Scatter plots

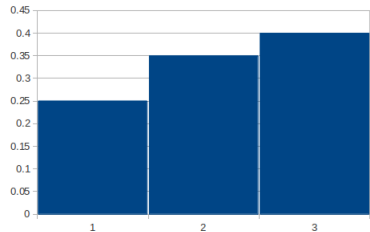
- Designed to relative sizes of categories which are part of a whole (percentages).
- Best when there are only a few categories.
- One problem with pie charts: human brain and eyes are not good at estimating or comparing angles.



## Really Stupid

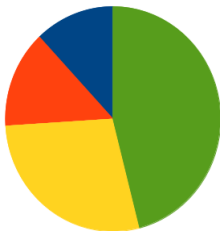


## Bar Charts Make for Easier for Comparisons



## Comparing pie graphs to each other is nearly impossible.

2000 Average Distribution of Cost For 1 Gallon Gasoline



- Distribution Costs
- Refining Costs
- Federal and State Taxes
- Crude Oil

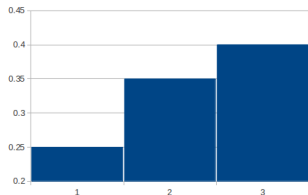
2004 Average Distribution of Cost For 1 Gallon Gasoline



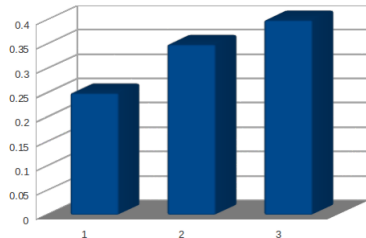
- Distribution Costs
- Refining Costs
- Federal and State Taxes
- Crude Oil

- Useful for making comparisons between groups.
- Can be useful for a small number, or a large number of groups.
- Does not require all parts add up to 100%.
- Smart bar charts:
  - NO 3-D!!
  - Minimal gaps between bars make for easier comparisons (not the Excel default!).
  - Begin vertical axis at 0 (not the Excel default!). Best with *ratio* data for each category.
  - *If it makes sense*, order items from smallest to largest.
  - Use differences in color only if it corresponds to differences in meaning or emphasis.

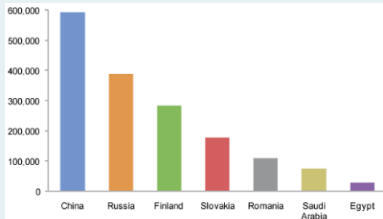
## Vertical Axis Misrepresentation

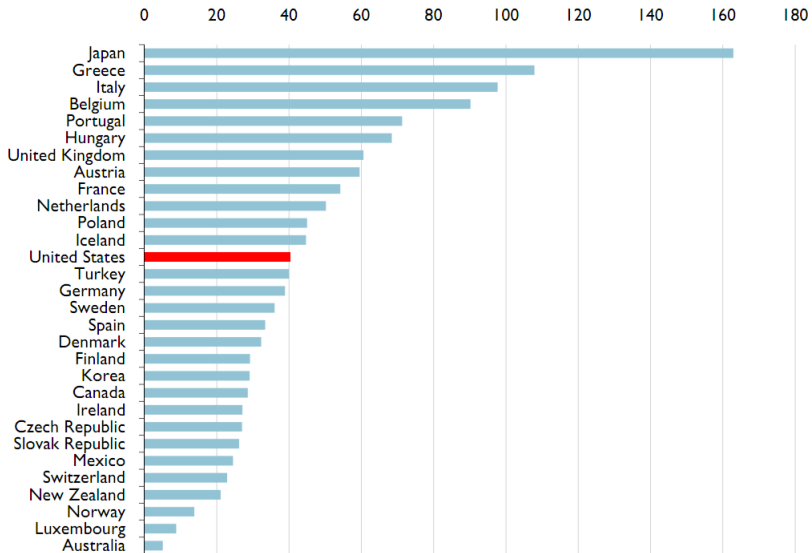


## 3-D Makes Comparison More Difficult



## Colorful Bars Distract





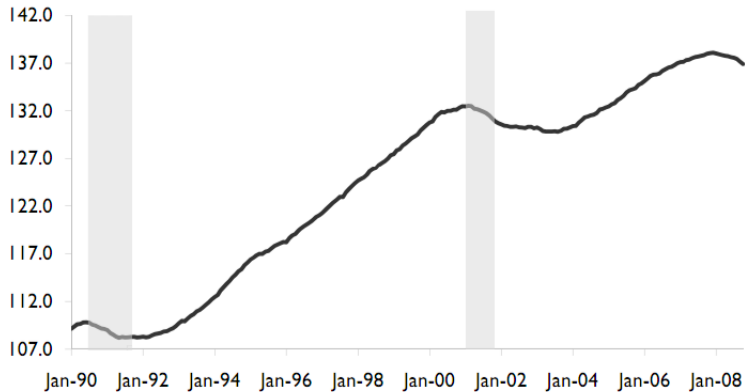
2008 Debt to GDP Ratio for OECD



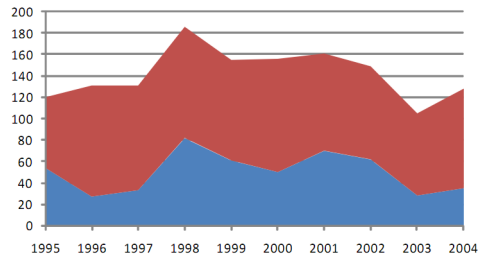


- Best with a single variable, measured over time.
- Also works well with a relative frequency of a single response category, measured over time.

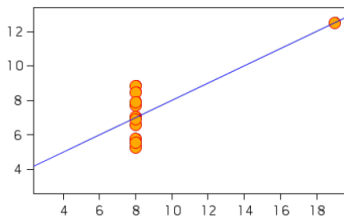
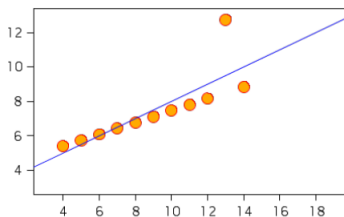
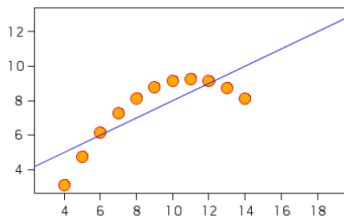
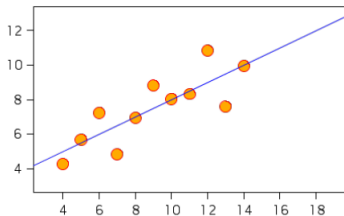
U.S. Payroll Employment: Total Nonagricultural: SA, Thousands of Persons



- An area chart is a line chart with the area underneath shaded.
- It is best with two lines in which one line represents a variable that is a subset of the other.
- Example: Total retail sales and Durable Goods sales.

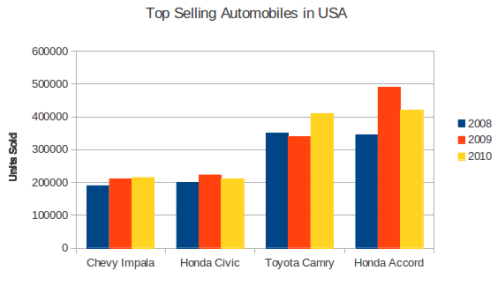


- Scatter plots are useful for showing the association for two different ratio/interval data.
- Complement a Pearson or Spearman correlation coefficient.
- Illustrate additional detail besides the strength of the relationship.

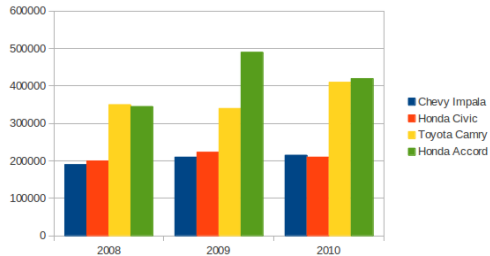


All of these sets of data have the same Pearson Correlation = 0.816.

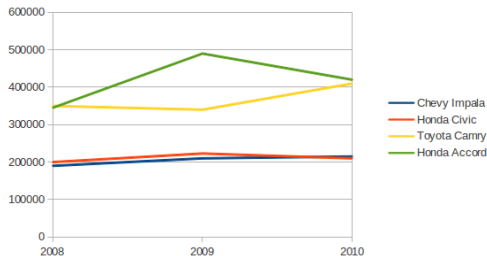
- Multiple-bar chart can illustrate measures of multiple categories.
- Can make comparisons on sales of each car between the three years.
- Can make comparisons between each car, for a given year.
  - This is more difficult. Why?



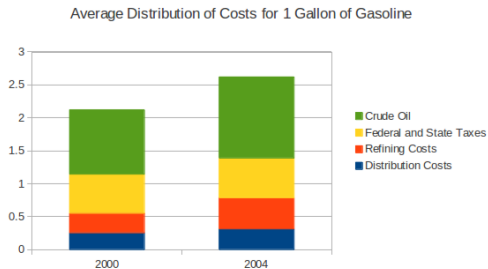
- This one is easier for make comparisons between cars.
- Even worse though for changes over time.



- A line graph effectively communicates movement over time.
- Comparing the height of the lines effectively communicates differences between cars.



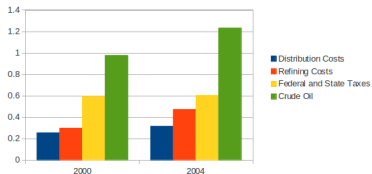
- Similar to a Multiple Bar Chart, except bars are stacked on top of one another, instead of placed next to one another.
- Difficult to make the following comparisons:
  - Relative costs of each category within a single year.
  - Relative costs of each category between 2000 and 2004.





## Actual Costs in Dollars

Average Distribution of Costs for 1 Gallon of Gasoline



## Percentage of Costs

Percentage Distribution of Costs for Gasoline

