

# Statistical Graphics for Visual Data Exploration

Cheng Peng

## Table of Contents

1. General Introduction to Data Visualization .....	2
1.1. Data and Information – basic description.....	2
1.2. Values of Data and Information Visualization .....	2
1.3. The Power of Visual Presentation of Information .....	2
2. Data Visualization Process .....	3
3. Gestalt Principles of Visualization.....	3
4. Good and Bad Graphics.....	5
6. Visualizing Spatial Data .....	10
7. Guidelines for Using Charts and Graphs .....	10
7.1. RELEVANT TERMS .....	10
7.2. WHY USE DIFFERENT FONTS, COLORS, AND GRAPHICS?.....	11
7.3. WHAT ARE SOME CONSIDERATIONS WHEN USING FONTS?.....	11
7.4. WHAT ARE SOME CONSIDERATIONS WHEN USING COLOR? .....	11
7.5. GENERAL GRAPHICS GUIDELINES.....	11
8. EIGHT GRAPHICS AND WHEN TO USE THEM .....	12
8.1. Tables .....	12
8.2. Pie Charts .....	12
8.3. Bar Charts.....	12
8.4. Line Charts.....	12
8.5. Schedule Charts .....	13
8.6. Flowcharts.....	13
8.7. Organizational Charts.....	13
8.8. Technical Drawings .....	13
9. WHAT TO AVOID WHEN USING GRAPHICS.....	14

# 1. General Introduction to Data Visualization

## 1.1. Data and Information – basic description

- **Data** is the raw material that can be processed by any computing machine. Data can be represented in the form of numbers, text, images, videos, audio, etc.
- **Information** is data that has been converted into a more useful or intelligible form. It is the set of data that has been organized for direct utilization by mankind, as information helps human beings in their decision-making process.

## 1.2. Values of Data and Information Visualization

- Methods of Information Recording
  1. Documentation forms
  2. Photographs,
  3. Video or Audio Recordings
- Why Analyze Data?
  1. Develop and assess hypotheses
  2. Explore patterns and discover the unknown
- Information Communication
  1. Explain and persuade
  2. Share and inspire
- Methods of Visual Communication
  1. One or more symbolic languages – letterforms, numerals, graphics
  2. Integrate the visual (left) and verbal (right) sides of the brain

## 1.3. The Power of Visual Presentation of Information

- Studies show that people remember: (1). 10% of what they hear; (2). 20% of what they read; (3). 80% of what they see and do.
- MIT neuroscientists find the brain can identify images seen for as little as 13 milliseconds. [<http://news.mit.edu/2014/in-the-blink-of-an-eye-0116>]
- 93% of most engaging posts on Facebook are images – Harris Eisenberg, Executive Vice President. [<https://www.searchenginepeople.com/blog/925-image-stats.html>]
- 30% of our brains are engaged in the task of processing visual information, as compared with only 8% for touch and 3% for hearing. [<http://discovermagazine.com/1993/jun/thevisionthingma227>]

## 2. Data Visualization Process

- **Acquire:** Obtain the data from different data sources and integrate them.
- **Parse:** The amount of data could be large and messy. It is necessary to put the data into a structure. This structure will make it easier to extract relevant information to visualize.
- **Filter:** After putting the data into a structure, we must filter out the data that is not necessary for data visualization. For example, if you are doing a specific gender visualization, you would have to remove the genders you did not want in your data set.
- **Mine:** Apply methods from statistics or data mining to discover patterns or place the data in the mathematical context. The focus will be on basic statistics in the beginning. This step helps get a basic understanding of the data before doing the representational step.
- **Represent:** Choose a basic visual model, such as a bar graph, list, map, or tree diagram.
- **Refine:** Improve the basic representation to make it clearer and more visually engaging. Applying basic color theory and basic graphic design theory will make visuals more effective.
- **Interact:** Add methods for manipulating the data or controlling what features are visible.

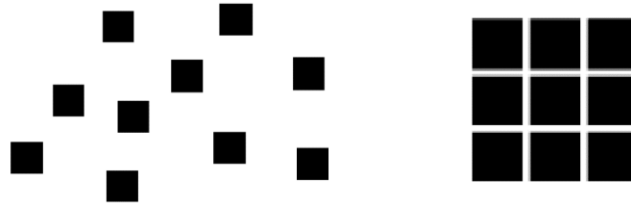
## 3. Gestalt Principles of Visualization

- According to **Gestalt theory**, the human mind attempts to recognize objects as a whole before examining individual parts. Information that is unrelated in size, shape, orientation, etc. will appear chaotic and unorganized to a viewer.
- **Grouping** - Elements that are alike one perceptually join to create whole units or figures. If a shape is repeated often enough it creates a pattern or recognizable form/figure.

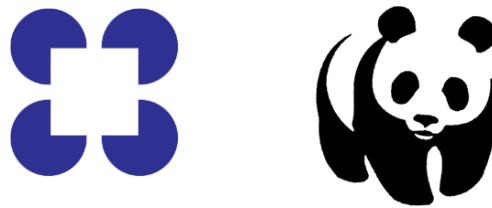


Methods for grouping include orientation, shape, value (color), size, subject matter, etc.

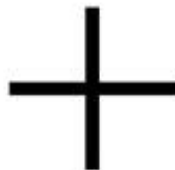
- **Proximity** - When individual elements are placed close together, they tend to be perceived as a single group.



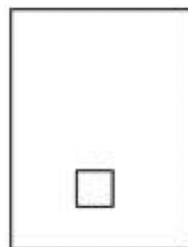
- **Closure** - The innate perceptual tendency for a viewer to perceive multiple objects as a group or totality. The ability to close “gaps” and create “wholes” out of dis-continuous lines, forms, or masses. Closure occurs when an object is incomplete, or space is not completely enclosed. If enough of the shape is indicated, people perceive the whole by filling in the missing information.



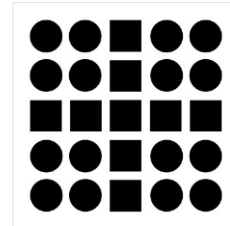
- **Continuation** - Continuation occurs when the eye is compelled to move through one object and continue to another object. The principle of continuity predicts the preference for continuous figures.



- **Containment** - A unifying force created by the outer edge of a composition or by a boundary within the composition. Containers encourage the viewer to make connections among visual units and add definition to the negative space around a positive space.



- **Repetition** - The Gestalt Principle that states we look for a pattern, and that pattern repeated over and over becomes salient in a composition. It moves forward in space, helping to define the composition as a whole.



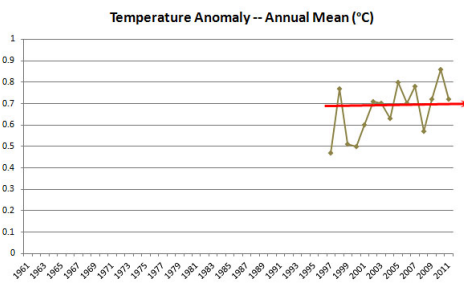
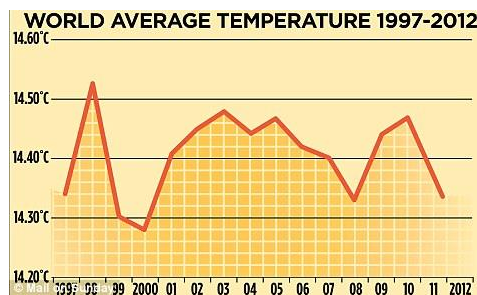
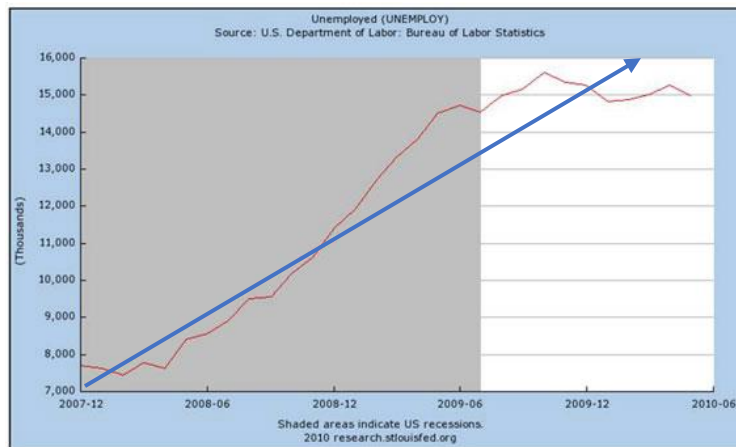
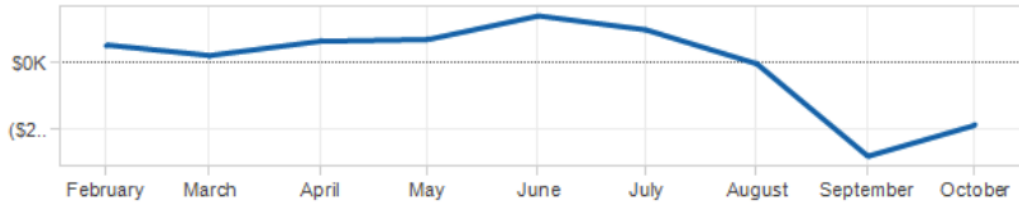
- Edward Tufte's Visual Design Principles

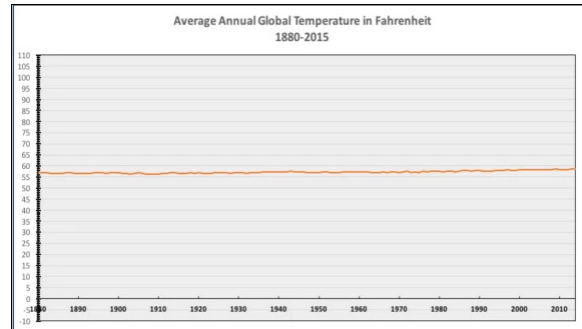
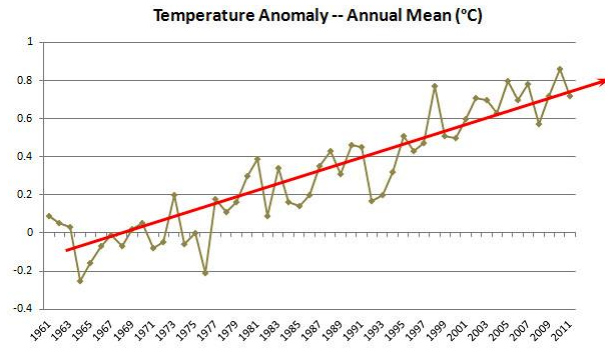


## 4. Good and Bad Graphics

- Graphical Integrity – Examples of Bad and Good Visuals



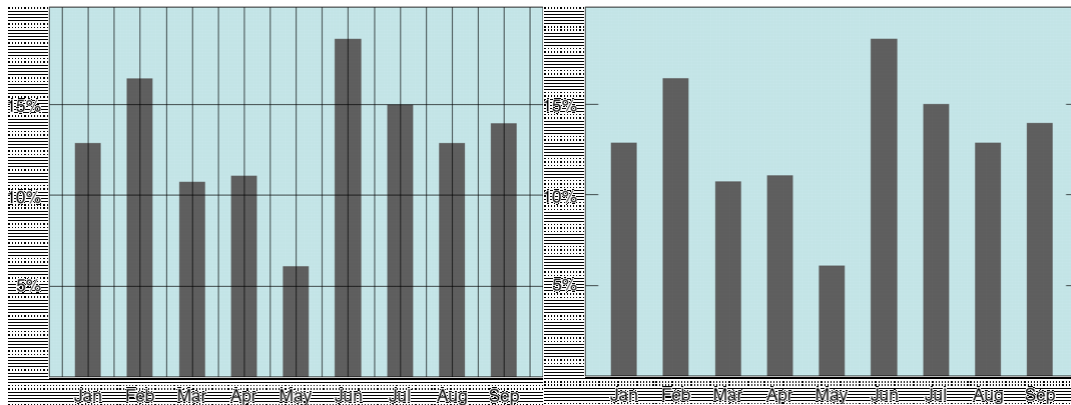


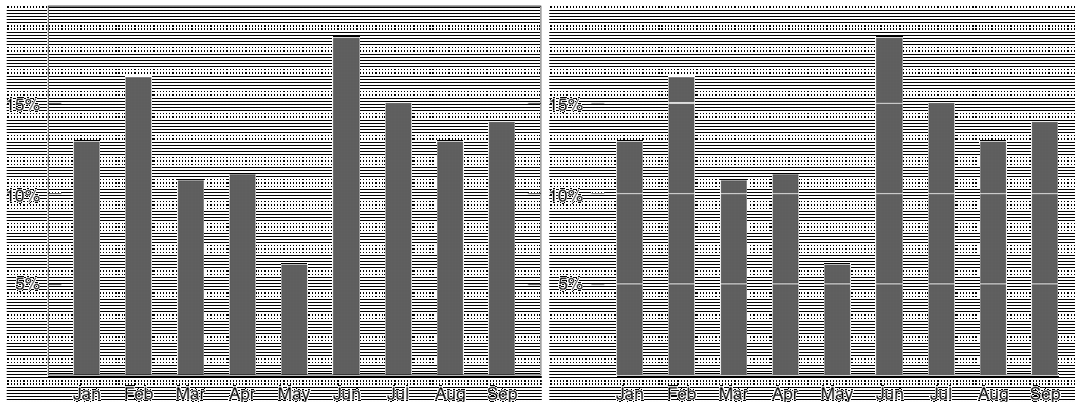


- Tell the Truth!

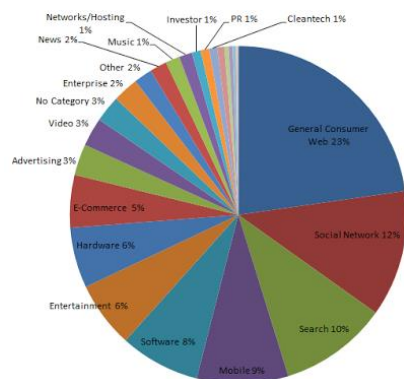
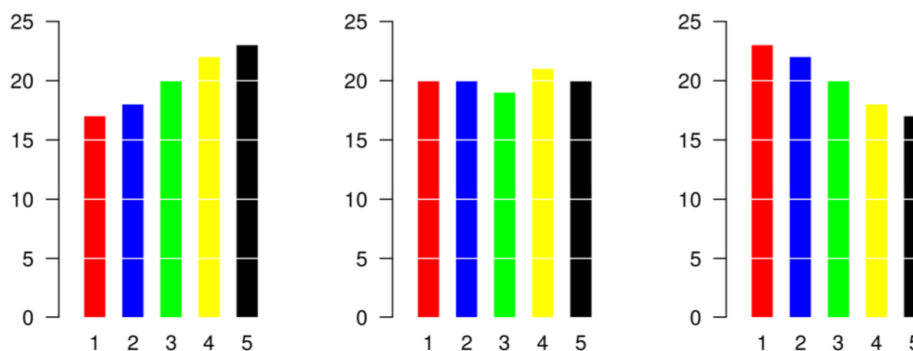
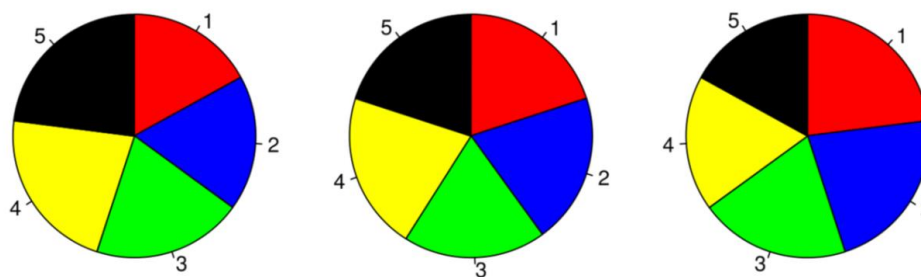
The representation of numbers ... should be directly proportional to the numerical quantities measured. — Edward Tufte 1983

1. Lie Factor
2. Avoid Distortion
3. Avoid Chart Junks - Unnecessary visual elements in charts that distract the viewer from the information

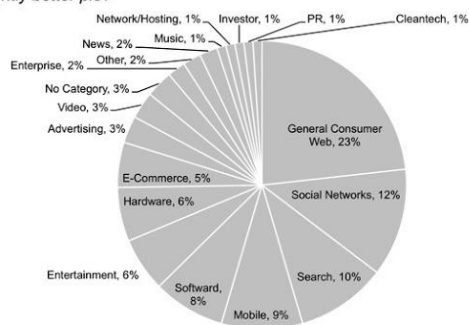




## 1. Misleading Representations – Some Examples on Pie Charts



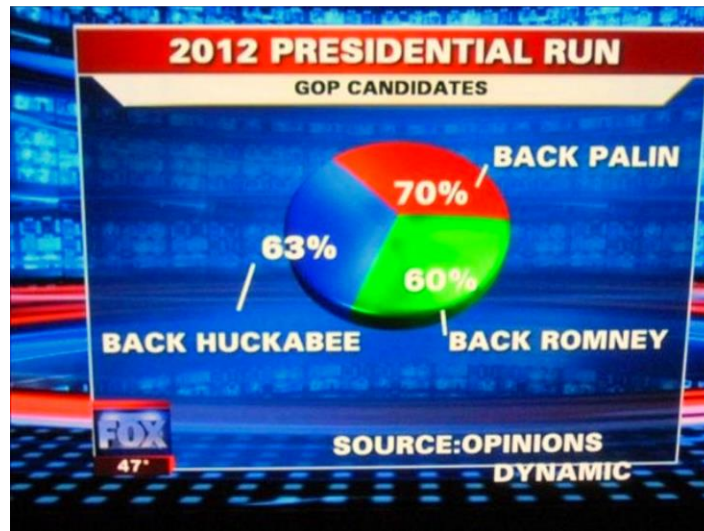
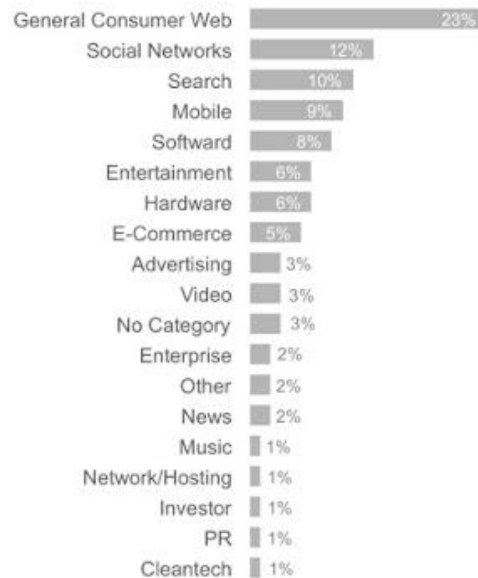
**TechCrunch Coverage: 2005 - 2011**  
A slightly better pie?





### TechCrunch Coverage: 2005 - 2011

*Bars are best!*



## 2. How to Design Good Pie Charts - Some Rules of Thumb

1. You're comparing the parts to the whole.
2. There's a small number of slices.
3. Sort the values.
4. Start at 12 o'clock.
5. Use alternatives (e.g., bar chart)

Source: <https://eagereyes.org/pie-charts>

**A [good] picture is worth a thousand words. But a bad picture can waste more than a thousand words!**

## 6. Visualizing Spatial Data

- **Key Components of a Story Map**

1. **Story:** The concept or message that a map is intended to communicate
2. **Supplemental Text:** Brief and to the point
3. **Spatial Data:** Only the locations and attributes that support your story
4. **Cartography:** Attractive and understandable representation of your spatial data
5. **Supporting Content:** Popups, graphs/charts, and images...not too much!
6. **User Experience:** Design and presentation of interactive functionality

- **Story Mapping Process**

1. **Planning:** Identify the story you want to tell, storyboard it, and identify each data set you have and will need
2. **Gathering:** Pull together all resources you need before building a map
3. **Browse:** Look at how others tell their stories and identify what can be done in-house and where you may need help
4. **Build:** put the pieces together, test, and publish

## 7. Guidelines for Using Charts and Graphs

<http://www2.sandhills.edu/academic-departments/english/wordguide/chartadvice.html>

### 7.1. RELEVANT TERMS

- Graphics, illustrations, visual aids: parts of documents or oral presentations that are not text.
- Tables: illustrations that place numbers or words in columns, rows, or both.
- Figures: all graphics other than tables, including charts, maps, and photos.
- Charts: figures that display data in visual forms, such as line graphs.
- Technical drawing: a type of figure that represents a physical object.

## 7.2. WHY USE DIFFERENT FONTS, COLORS, AND GRAPHICS?

1. Graphics make concepts easier to understand.
2. Color, fonts, and graphics can help the reader comprehend an idea.
3. Graphics support and emphasize ideas.
4. Graphics generate interest and attract attention.
5. Graphics are important and powerful when integrated with the text.

## 7.3. WHAT ARE SOME CONSIDERATIONS WHEN USING FONTS?

Many fonts are available today. If you are unfamiliar with this term, consult your printer manual. Fonts are generally grouped into two categories: *serif* (with "feet") and *sans serif* (without "feet"). Some fonts are inappropriate for professional writing, but there are many options. these guidelines will help you make the right choices:

1. Find out what the reader or company wants or expects.
2. Consider what will be clear and readable.
3. Take into account the available space.
4. Think about the purpose of the document.
5. Consider the tone you want to use.

## 7.4. WHAT ARE SOME CONSIDERATIONS WHEN USING COLOR?

Color is expected by today's readers, but it is more expensive and can not always be justified. Take into account the following guidelines when deciding to use color:

1. Colors should properly relate to the topic.
2. Colors should enhance the company logo.
3. Dark or textured backgrounds should be used sparingly.
4. All colors should be tested as to what they will look like when produced in the final report form.

## 7.5. GENERAL GRAPHICS GUIDELINES

You should observe these guidelines when using any graphic:

1. Know the purpose of the graphic.
2. Check to see that the data are correct.
3. Always refer to graphics in the text.
4. Consider where to place the graphic in the text.
5. Place graphics vertically.
6. Keep graphics simple and uncluttered.

7. Place titles, source documentation, etc., with the graphic.

## 8. EIGHT GRAPHICS AND WHEN TO USE THEM

**Note:** Consult the External Links section for weblinks to sites that contain examples of all of the following graphics.

### 8.1. Tables

- Informal table: data in rows or columns.
  - Formal table: data in a grid with horizontal rows and vertical columns.
1. Use informal tables to extend or expand your text.
  2. Use formal tables for complicated data separated from the text.
  3. Use white space.
  4. Place titles, headings, etc., were needed to explain data.
  5. Be especially careful with financial data.

### 8.2. Pie Charts

This graphic shows relationships between the parts and the whole.

1. Limit pie "slices" to no more than 6 or 7.
2. Work clockwise from largest to smallest "slice."
3. Use pie charts for money and percentages.
4. Keep it simple.
5. Label carefully.

### 8.3. Bar Charts

This graphic shows simple comparisons, especially changes in quantity.

1. Limit the number of bars.
2. Be sure comparisons are clear.
3. Adjust bar widths and space between them to make them equal.
4. Arrange the order of bars carefully.
5. Make creative choices.

### 8.4. Line Charts

This graphic is used to show trends or changes over time, such as price changes.

1. Show trends with line charts.
2. Place line charts where they can get attention.
3. Make line charts that are accurate and clear.
4. Avoid putting numbers on the line chart itself.
5. Do not place too many lines on the chart.

#### 8.5. Schedule Charts

This graphic is a special kind of chart showing when certain activities will be accomplished. Tasks and times are highlighted and also mentioned in the text.

1. Only the main activities should be included.
2. Activities should be listed in sequence.
3. Labels should run in the same direction.
4. New formats should be devised as needed.
5. A realistic schedule should be used.

#### 8.6. Flowcharts

This graphic depicts a process, usually with boxes and shapes that represent activities.

1. An overall view of the process is all that is needed—not every detail.
2. The number of shapes should be restricted.
3. A legend should be provided when needed.
4. The sequence of steps should run from top to bottom or from left to right.
5. All shapes should be clearly labeled.

#### 8.7. Organizational Charts

This graphic shows how a company or organization is set up.

1. Use rectangles connected by lines to represent top-level positions in the organization.
2. Use dotted or solid lines to join boxes.
3. Show mid-level and low-level positions using a circular design.
4. Use varied shapes with care.
5. Use creativity.

#### 8.8. Technical Drawings

This graphic accompanies instructions, sales documents, etc. Technical drawings may be more useful than photographs because they show very specific views.

1. Select the proper amount of detail.
2. Label parts carefully.
3. Select the best view.
4. Use a legend when there are many parts.

## 9. WHAT TO AVOID WHEN USING GRAPHICS

It is possible to distort or otherwise misuse graphics. Following are some common ways to misrepresent data:

1. Don't use a bar chart improperly; be sure the height of lines corresponds to the actual increase or decrease.
2. Don't place data such as expenses in a confusing sequence.
3. Don't misuse a pie chart by omitting percentages or moving in a direction other than clockwise/largest to smallest or including too many "slices."