Center for Business Intelligence

4.0 Visualization

Eugene Rex L. Jalao, Ph.D.

Associate Professor

Department Industrial Engineering and Operations Research

University of the Philippines Diliman

@thephdataminer

Module 1 of the Business Intelligence and Analytics Track of UP NEC and the UP Center of Business Intelligence

Module 1 Outline

1. Intro to Business Intelligence
   – Case Study on Selecting BI Projects
2. Data Warehousing
   – Case Study on Data Extraction and Report Generation
3. Descriptive Analytics
   – Case Study on Data Analysis
4. **Visualization**
   – **Case Study on Dashboard Design**
5. Classification Analysis
   – Case Study on Classification Analysis
6. Regression and Time Series Analysis
   – Case Study on Regression and Time Series Analysis
7. Unsupervised Learning and Modern Data Mining
   – Case Study on Text Mining
8. Optimization for BI

Outline for this Session

- What is Visualization?
- Types of Visualizations
- Visual Design Principles
- Design Optimization
- Case Study

What is Visualization?

Definition 4.1: Visualization

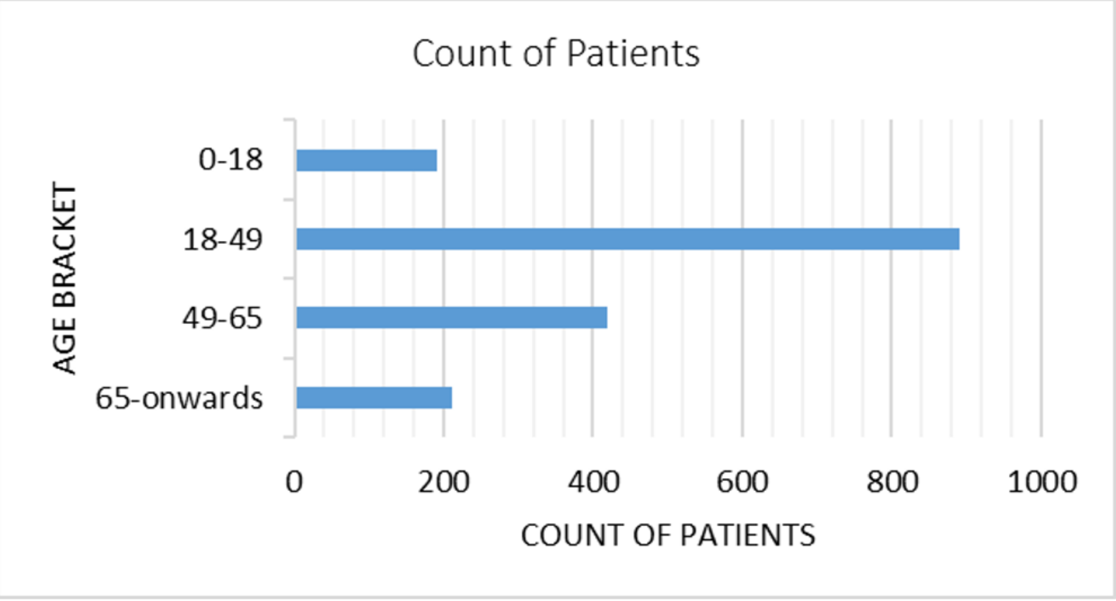
- Visualization is the presentation of **information** using spatial or graphical representations, for the purposes of: comparison facilitation, recognition of patterns and general decision making.
- “Transformation of the symbolic into the geometric” (McCormick et al., 1987)

What is Visualization?

- Problem
  - Big datasets: How to **understand** them?
- Solution
  - Take better advantage of human **perceptual system**
  - Convert information into a graphical representation.
- Issues
  - How to convert abstract information into **graphical form**?
  - Do visualizations do a better job than other methods?

What is Visualization?

Age Bracket	Count of Patients
0-18	190
18-49	890
49-65	420
65-onwards	212



What is Visualization?

- More specifically, visualization should:
  - Make large datasets **coherent**

    (Present huge amounts of information compactly)
  - Present information from various **viewpoints**
  - Present information at **several levels** of detail

    (from overviews to fine structure)
  - Support visual **comparisons**
  - Tell **stories** about the data

What is Visualization?

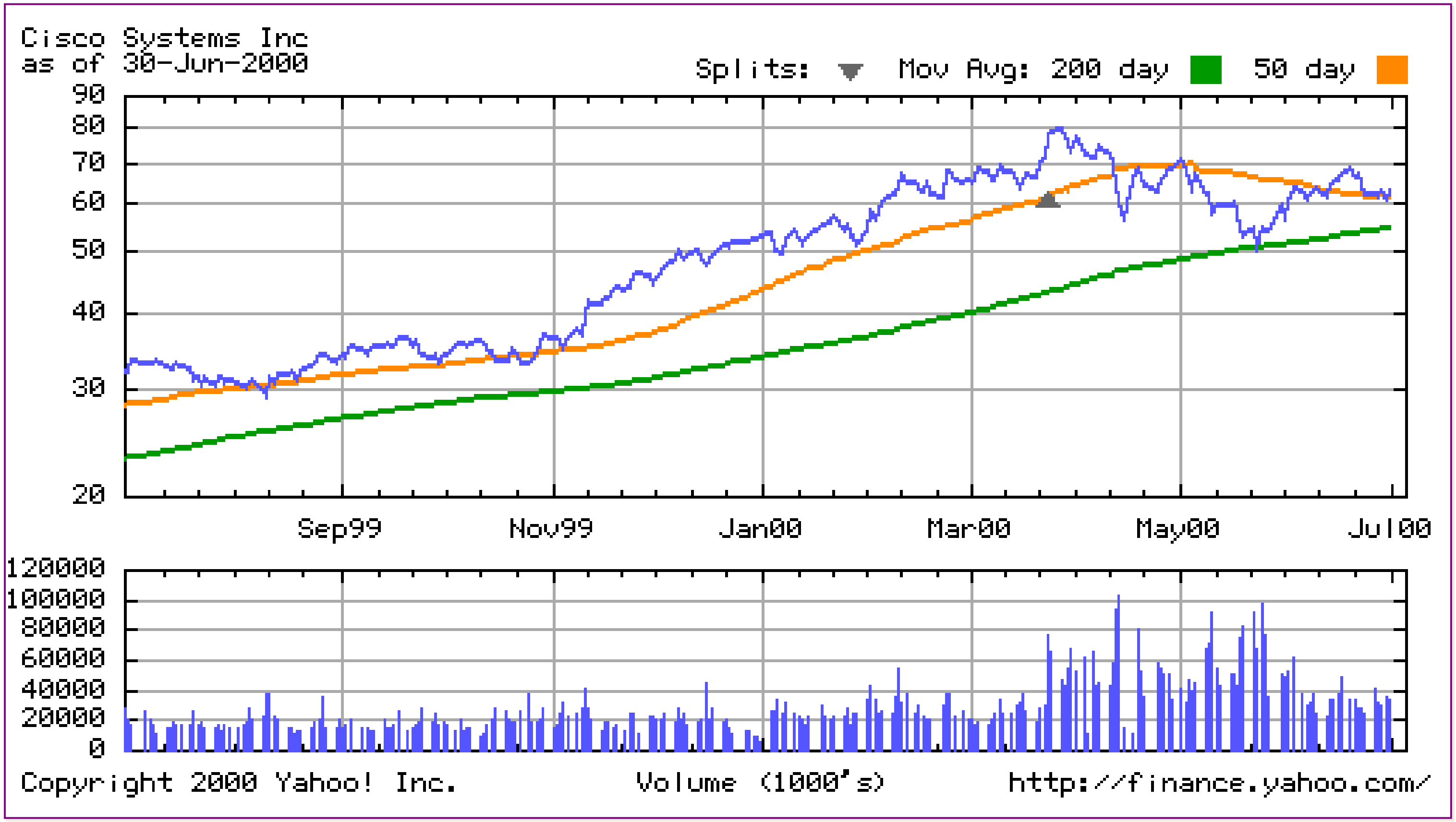
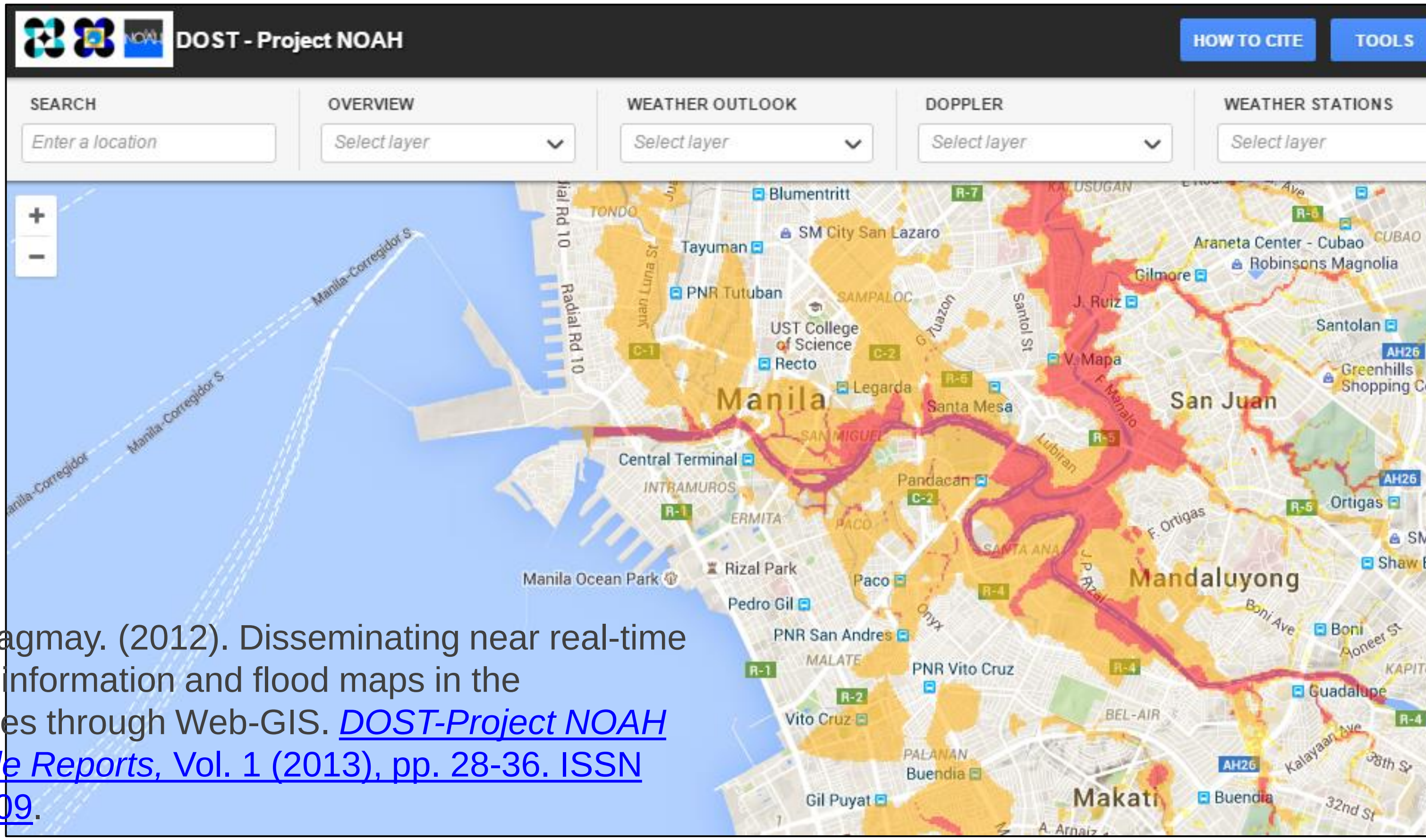


Figure 4.1: Time Series Visualization

What is Visualization?



A.M.F. Lagmay. (2012). Disseminating near real-time hazards information and flood maps in the Philippines through Web-GIS. [*DOST-Project NOAH Open-File Reports*, Vol. 1 (2013), pp. 28-36. ISSN 2362 7409].

Figure 4.2: Project NOAH

What is Visualization?

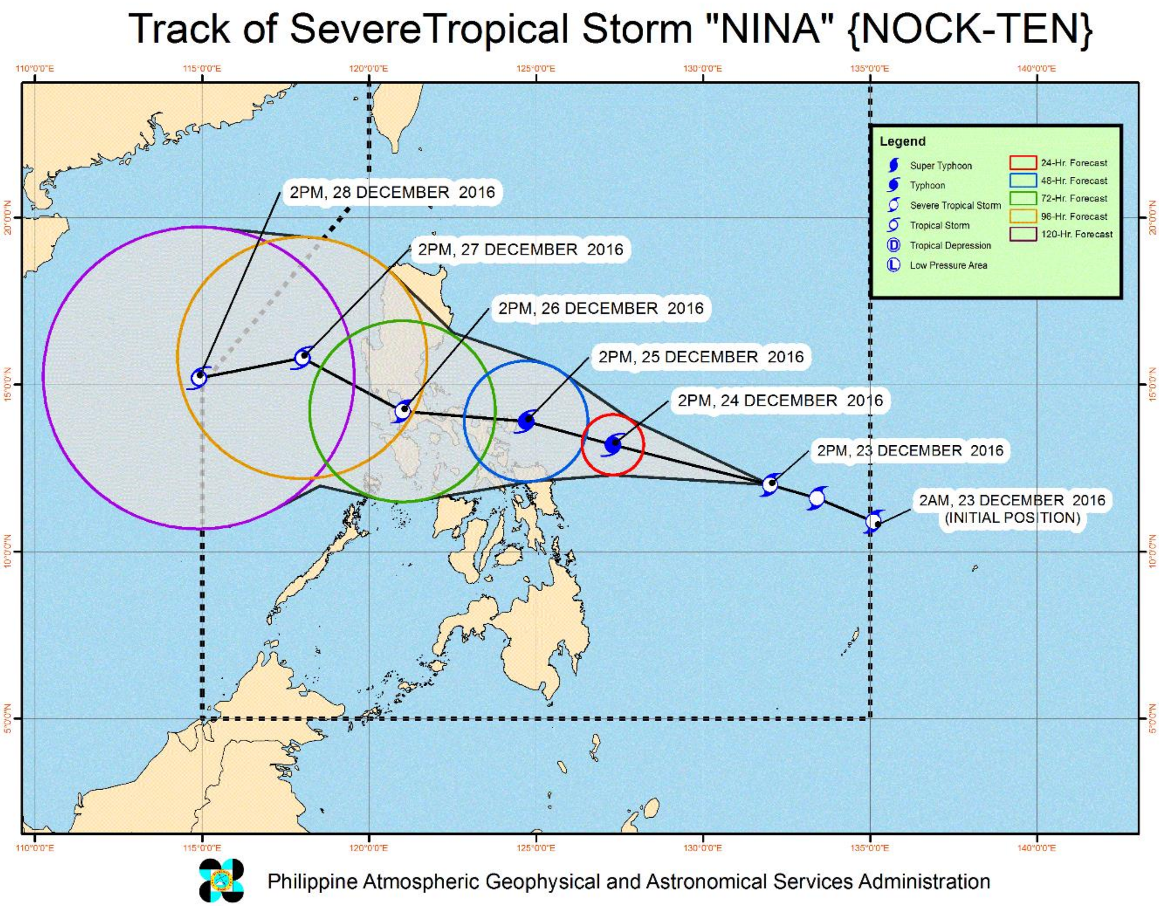


Figure 4.3: Nina Track December 2016 from PAGASA

What is Visualization?

- Mystery: what is causing a cholera epidemic in London in 1854?

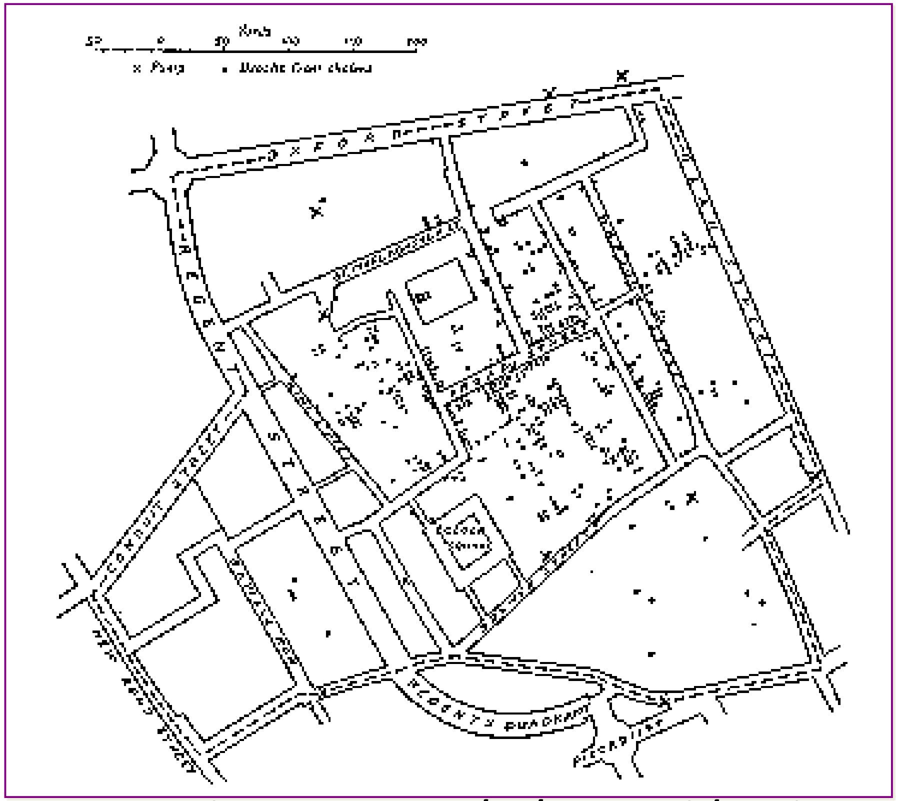
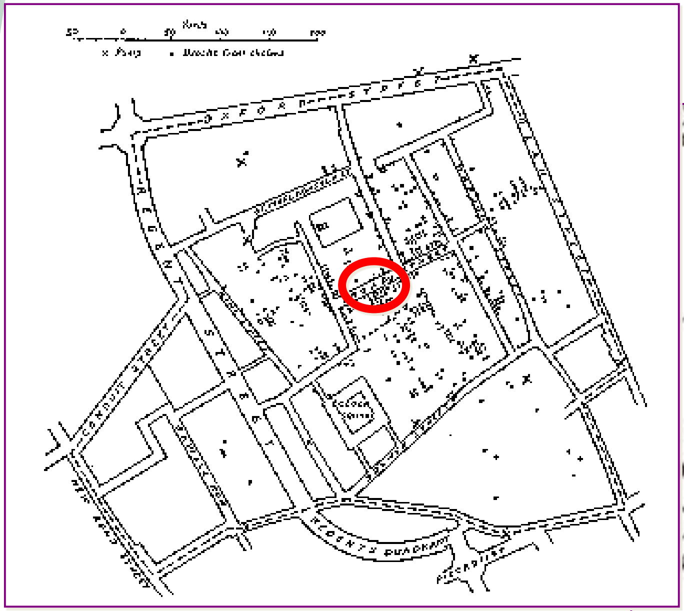


Figure 4.4a: Cholera Epidemic

Illustration of John Snow's deduction that a cholera epidemic was caused by a bad water pump, circa 1854.

Horizontal lines indicate locations of deaths.

What is Visualization?



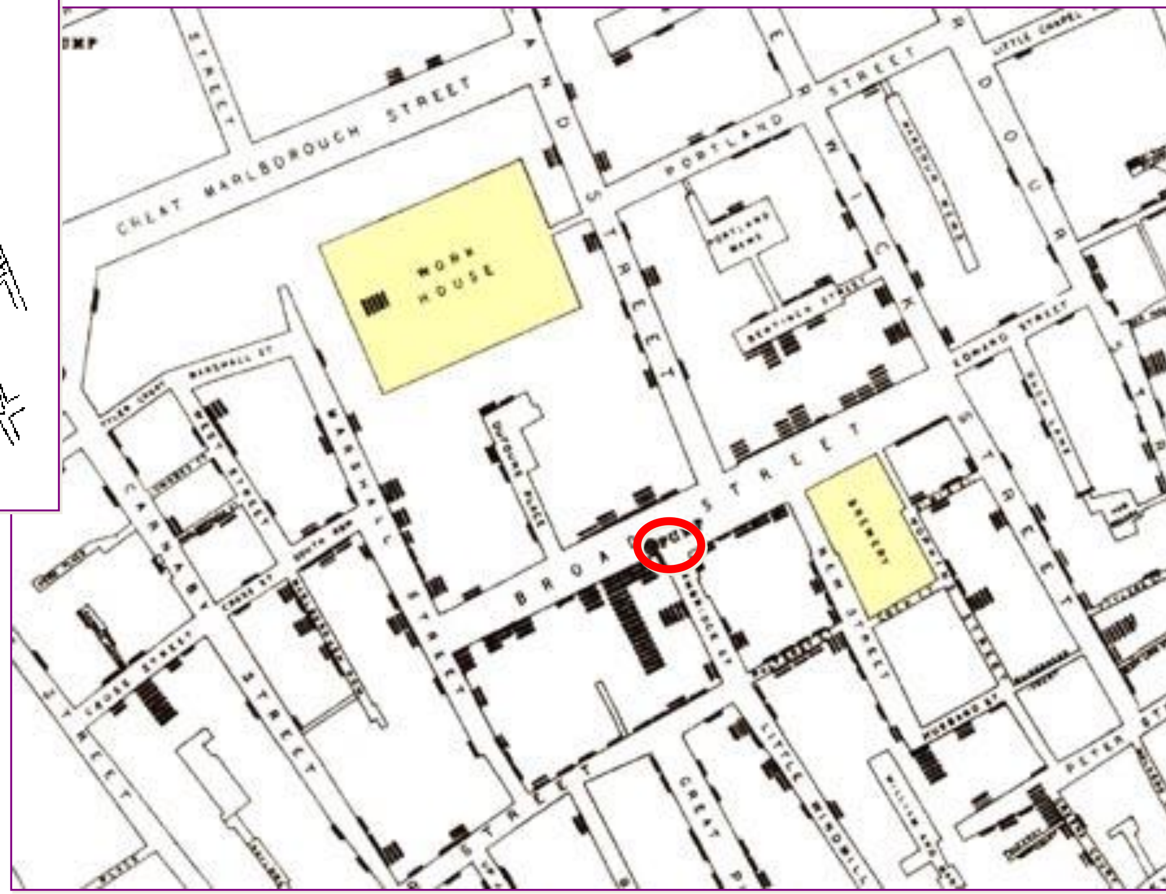


Figure 4.4b: Cholera Epidemic

What is Visualization?

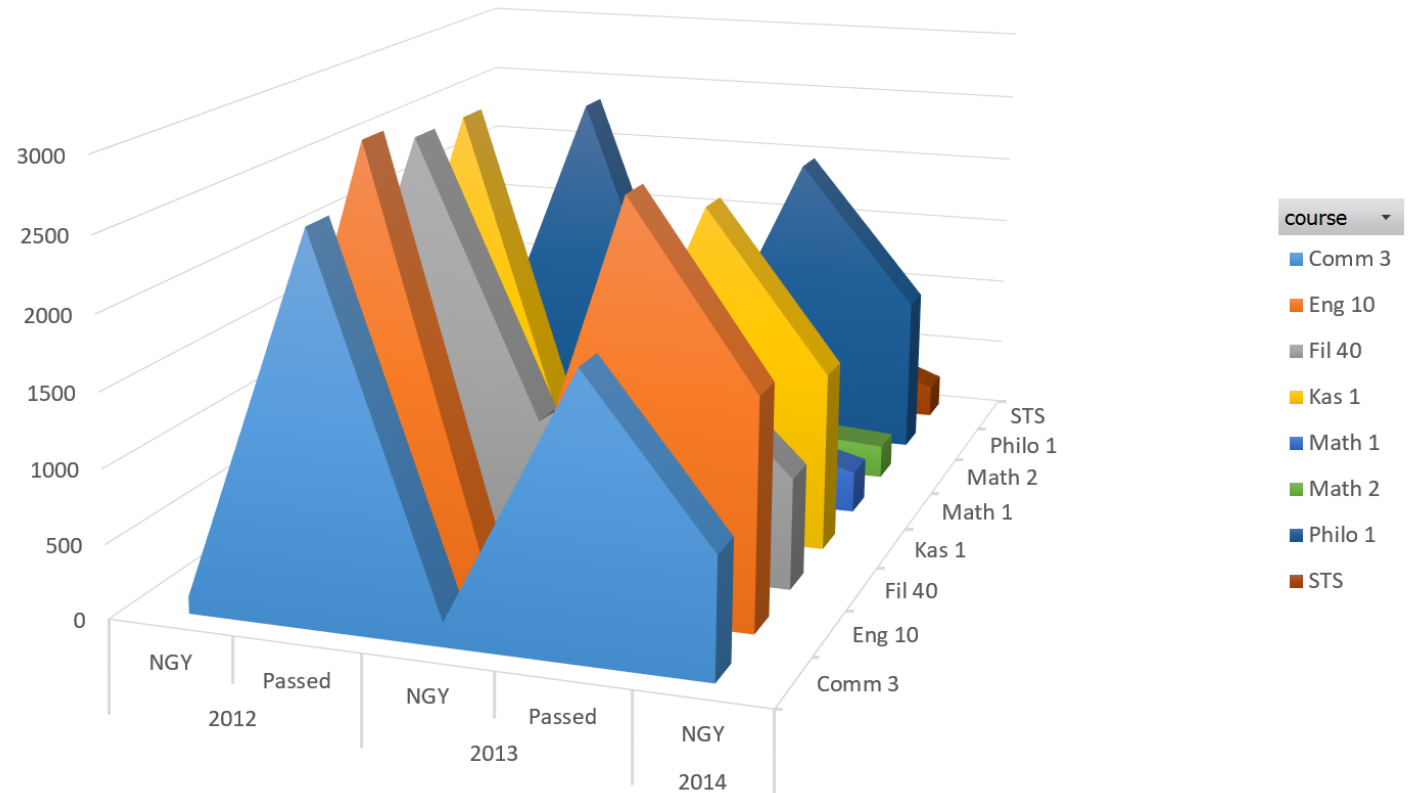
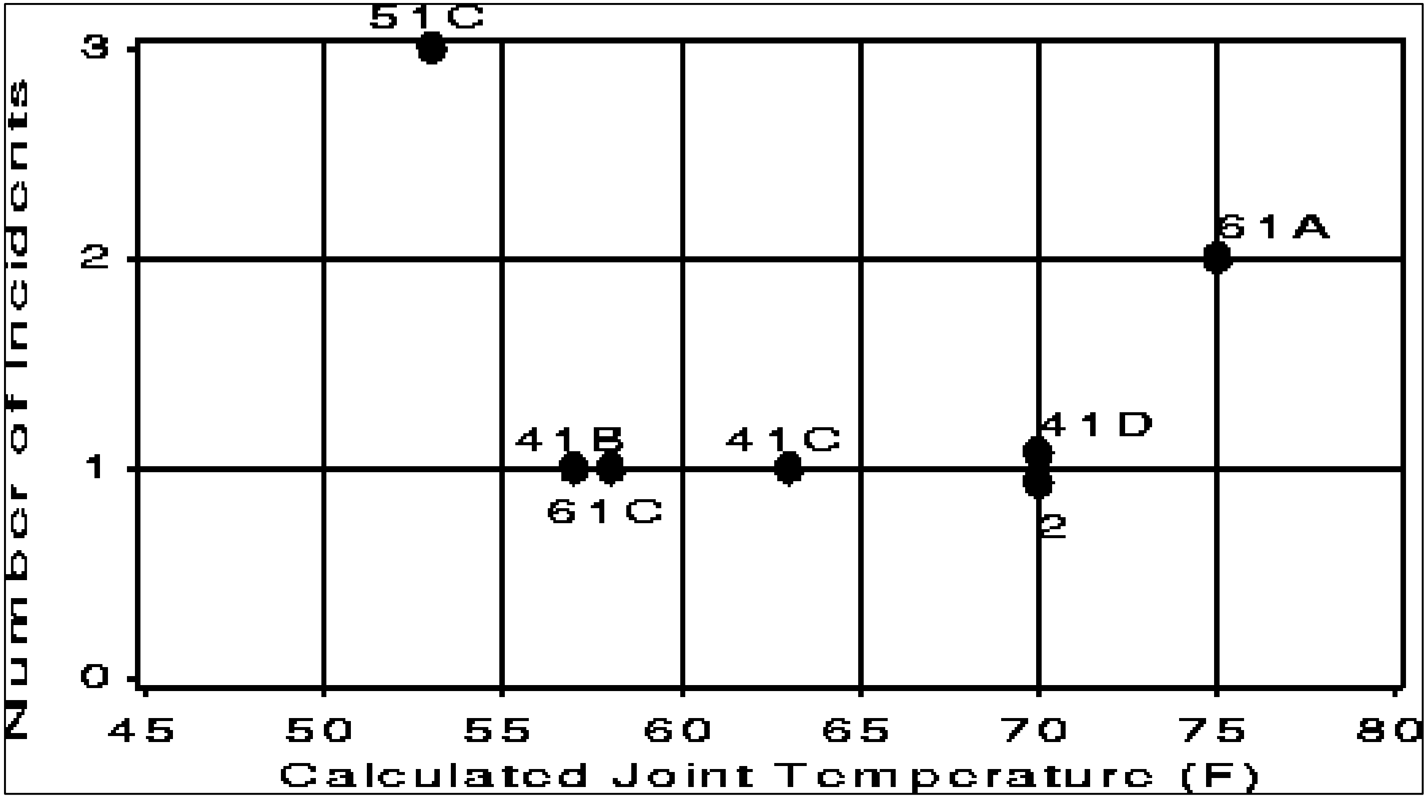


Figure 4.5: Visualization Failure UP Diliman RGEP Courses

eljalao@up.edu.ph

What is Visualization?



http://www.math.yorku.ca/SCS/Gallery/

Figure 4.6: Visualization Failure for O-Ring Damage

Advantages of Visualization

- Use the eye for **pattern recognition** since people are good at
  - scanning
  - recognizing
  - remembering images
- Graphical elements facilitate **comparisons** via
  - length
  - shape
  - orientation
- **Animation** shows changes across time
- **Color** helps make distinctions
- **Aesthetics** help maintain interest

Advantages of Visualization

- Two Different Types of Visualization
  - Explore/Calculate
    - Analyze
    - Reason about Information
  - Communicate
    - Explain
    - Make Decisions
    - Reason about Information

Outline for this Session

- What is Visualization?
- **Types of Visualizations**
- Visual Design Principles
- Design Optimization
- Case Study

Types of Visualizations

Definition 4.2: Graph

- A medium of visualization to communicate information with a requirement of required at least two scales where values are associated by symmetric “paired with” relations
  - Examples: scatter-plot, bar-chart, layer-graph

Types of Visualizations

- Anatomy of a Graph
  - **Framework**
    - sets the stage
    - kinds of measurements, scale, ...
  - **Content**
    - marks
    - point symbols, lines, areas, bars, …
  - **Labels**
    - title, axes, tic marks, ...

eljalao@up.edu.ph

Types of Visualizations

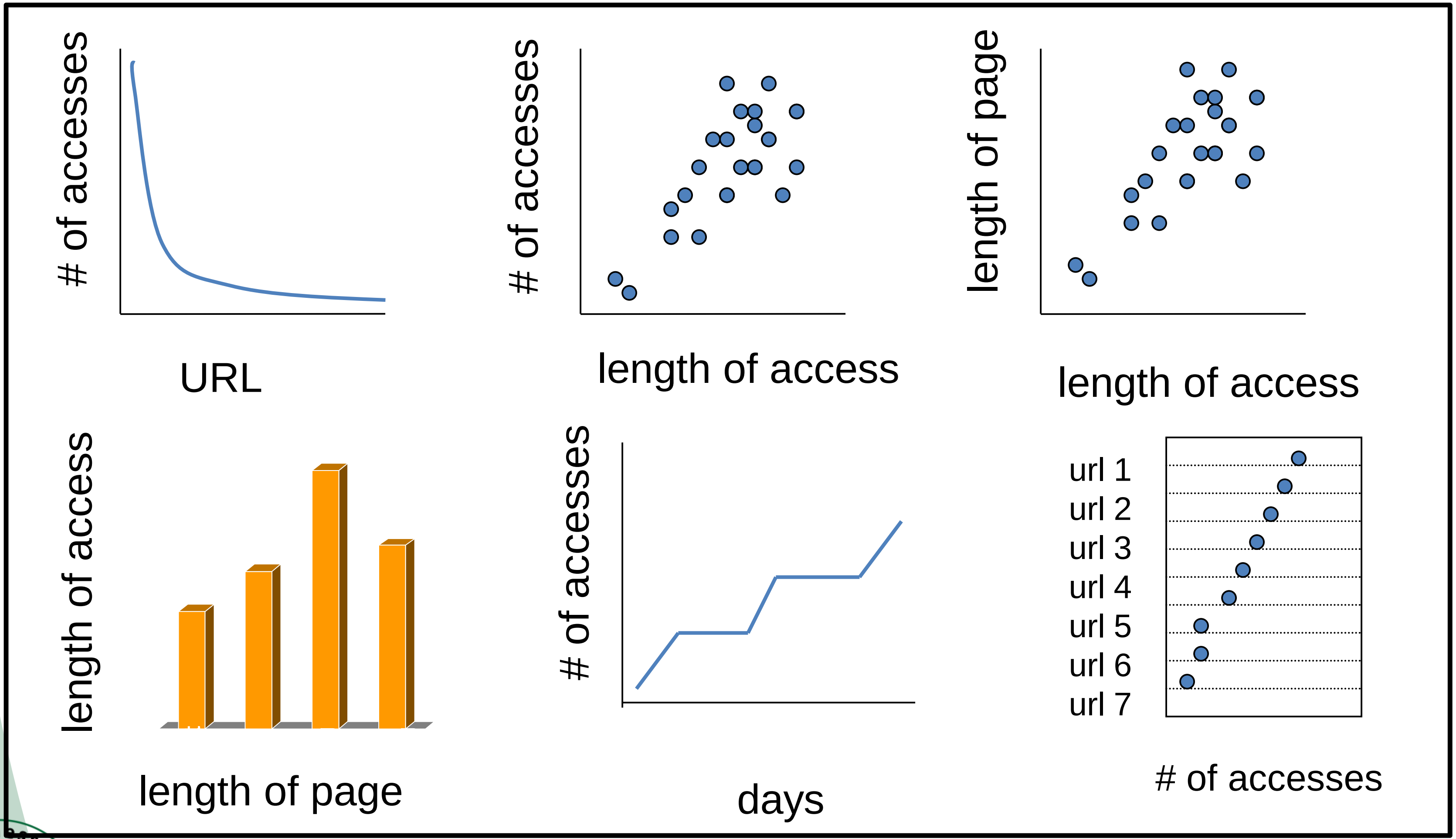


Figure 4.10: Common Graphs Types

Graphs

Definition 4.3: Line Graph

- A line graph displays a quantitative variable of interest against variable usually an interval or a time span. It also shows trends and relationships between the two variables (when grouped with other lines).

Graphs

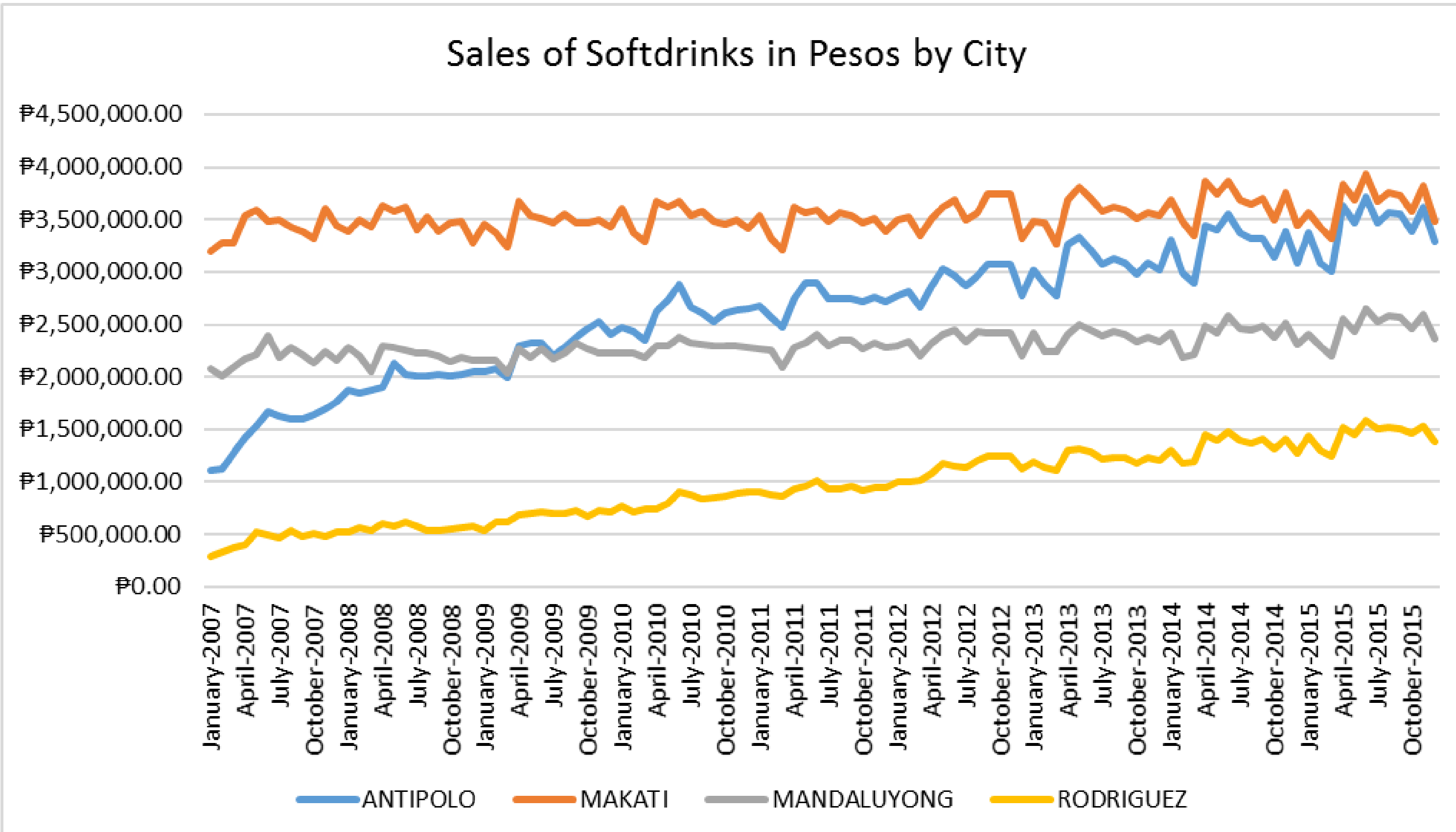


Figure 4.11: Line Graph

eljalao@up.edu.ph

Graphs

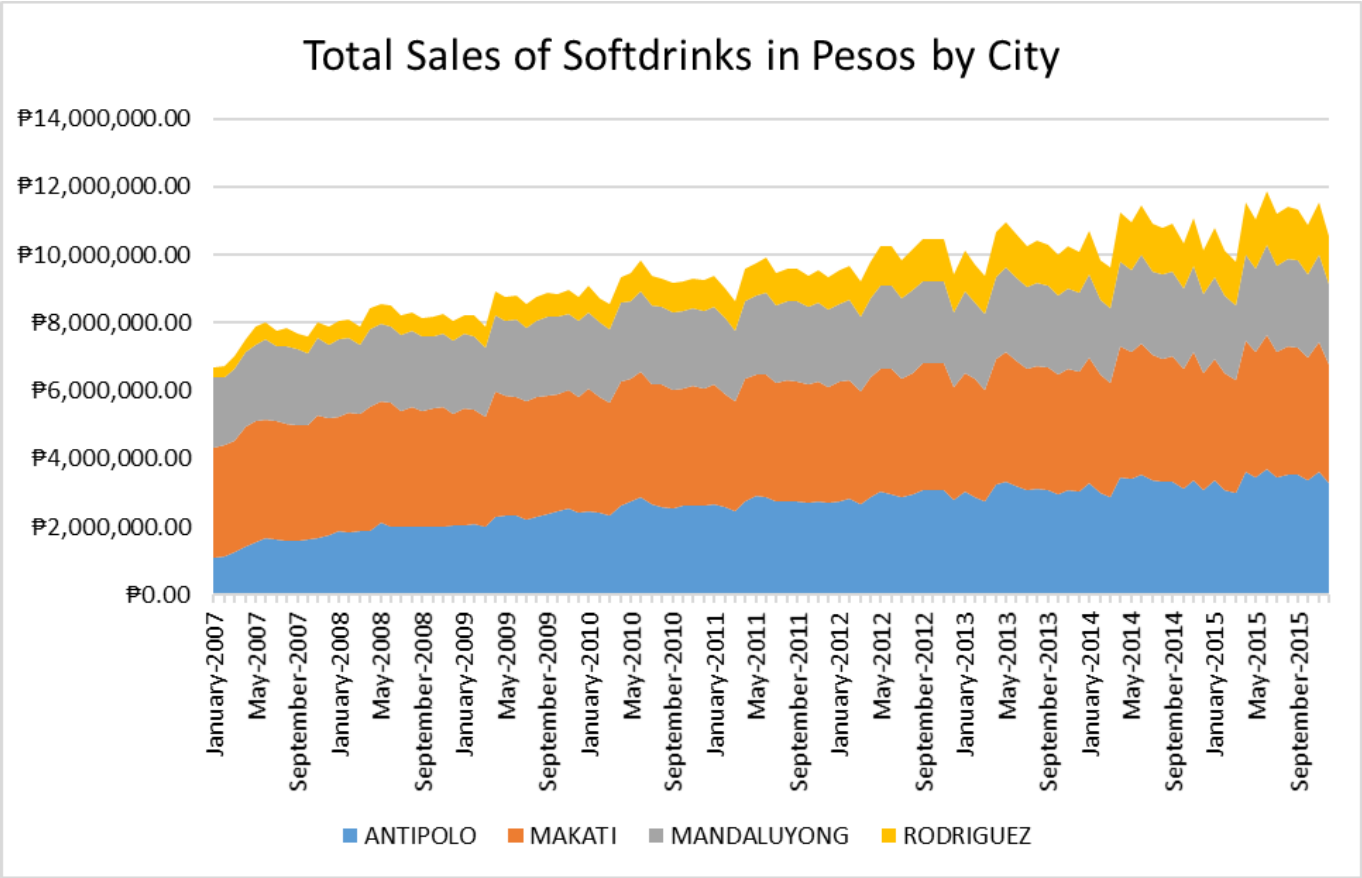


Figure 4.12: Area Graph

Graphs

Definition 4.4: Bar Graph

- To compare the values of numeric variables at a single point in time using bar symbols
- Usually bar charts show the relationship between a single quantitative variable versus a single qualitative variable

Graphs

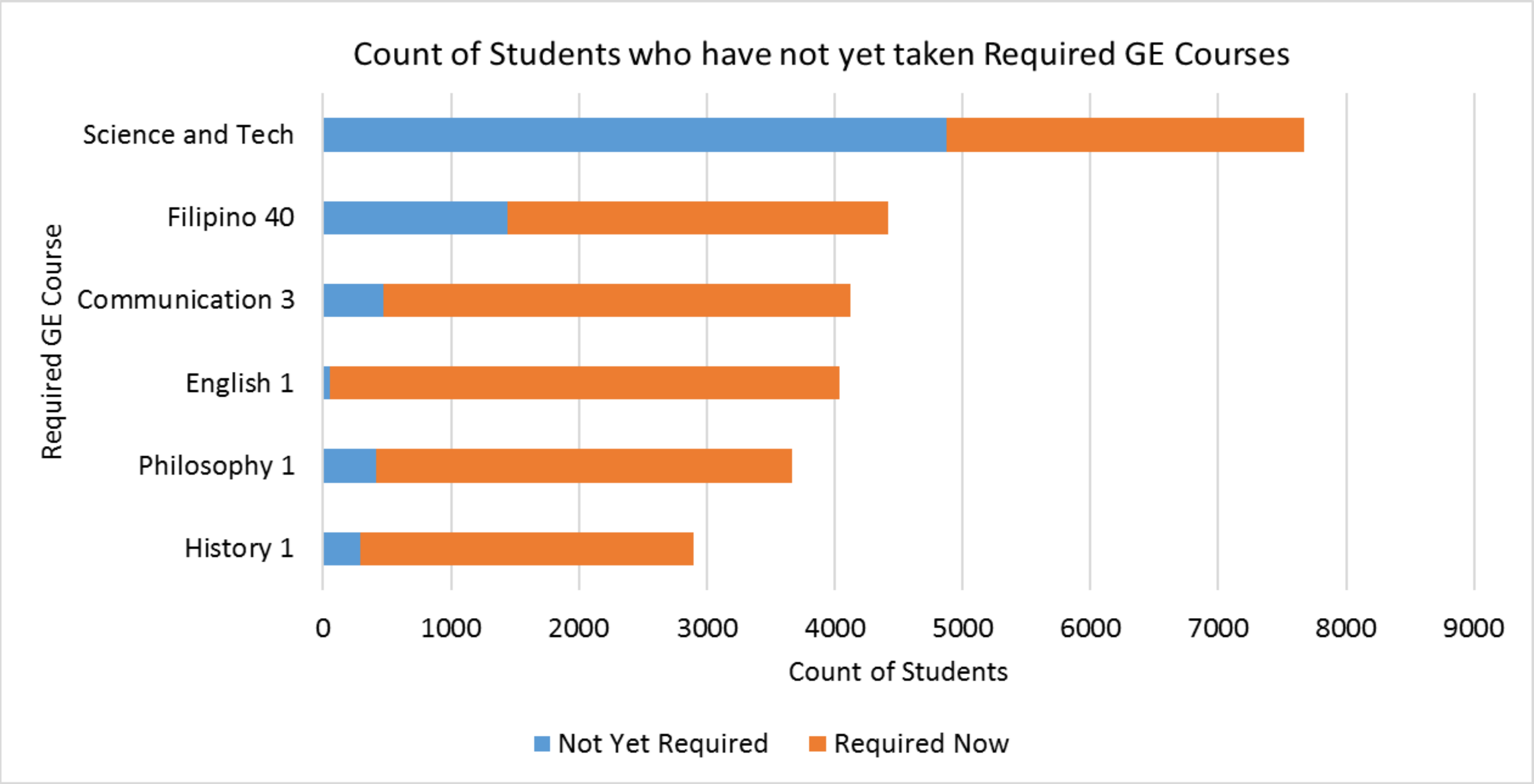


Figure 4.13: Stacked Bar Graph

Graphs

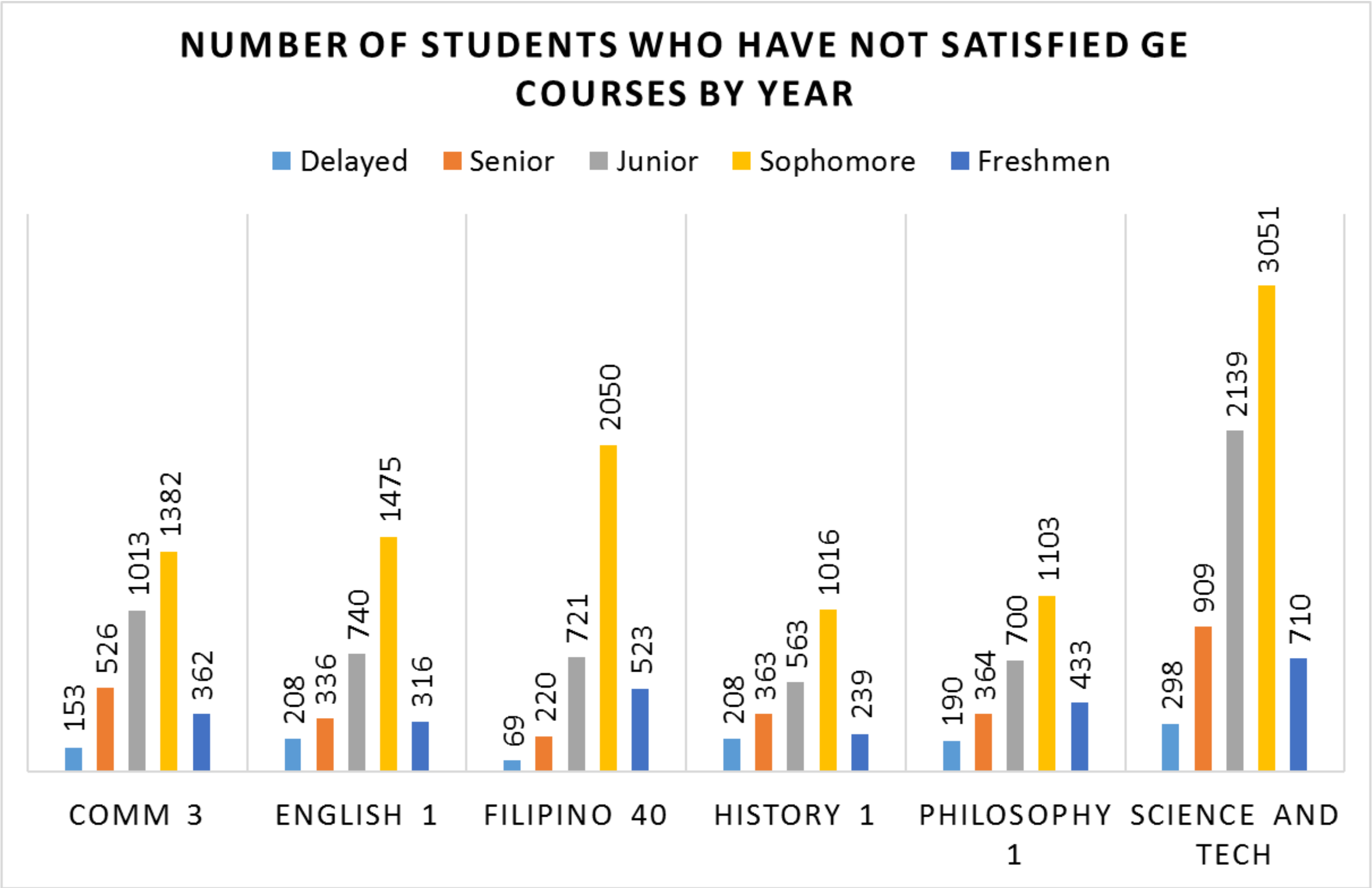


Figure 4.14: Stacked Bar Graph

Graphs

Definition 4.5: Scatter Graph

- Scatter plots are descriptive plots that draws individual data points into a two dimensional plot.
- Attributes values determine the position of each point

Graphs

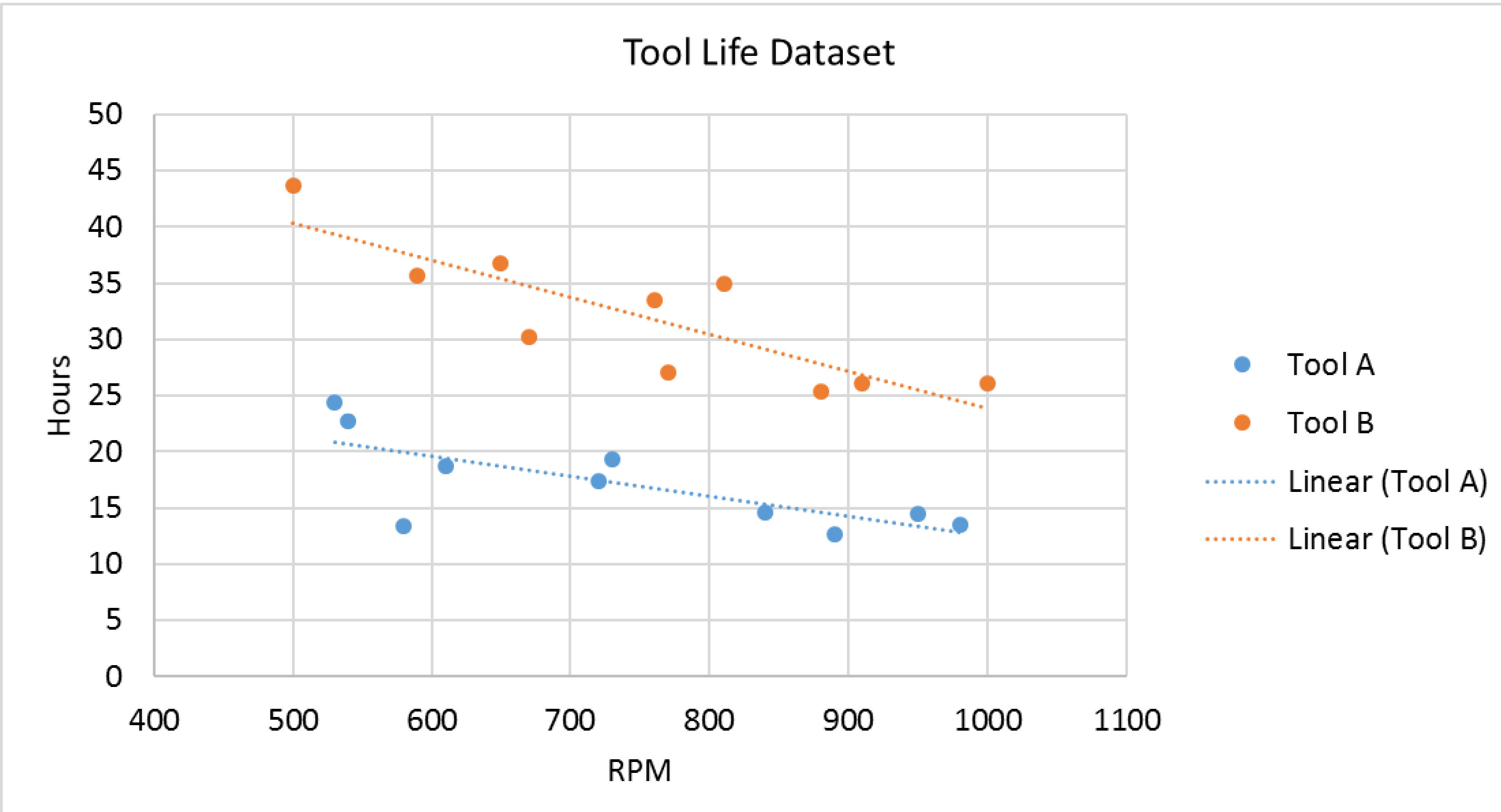


Figure 4.15: Stacked Bar Graph

Graphs

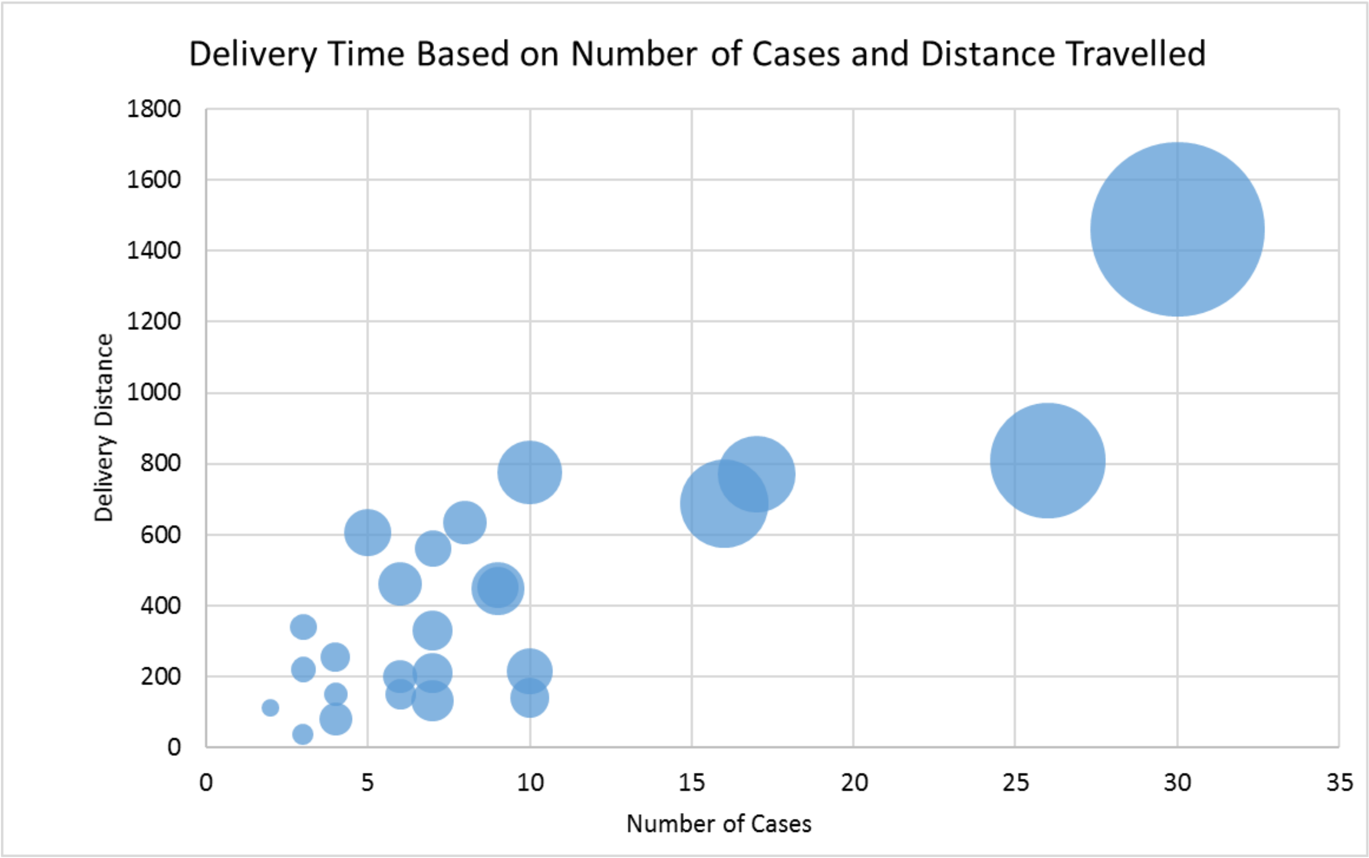


Figure 4.16: Bubble Chart

Graphs

Definition 4.6: Pie Chart

- Pie Charts help present proportions and percentages between categorical values, by dividing a circle into proportional slices.
- Each circle arc length represents a proportion of each categorical value, while the full 360-degree circle represents the total sum of all the data, which is equal to 100%.

Graphs

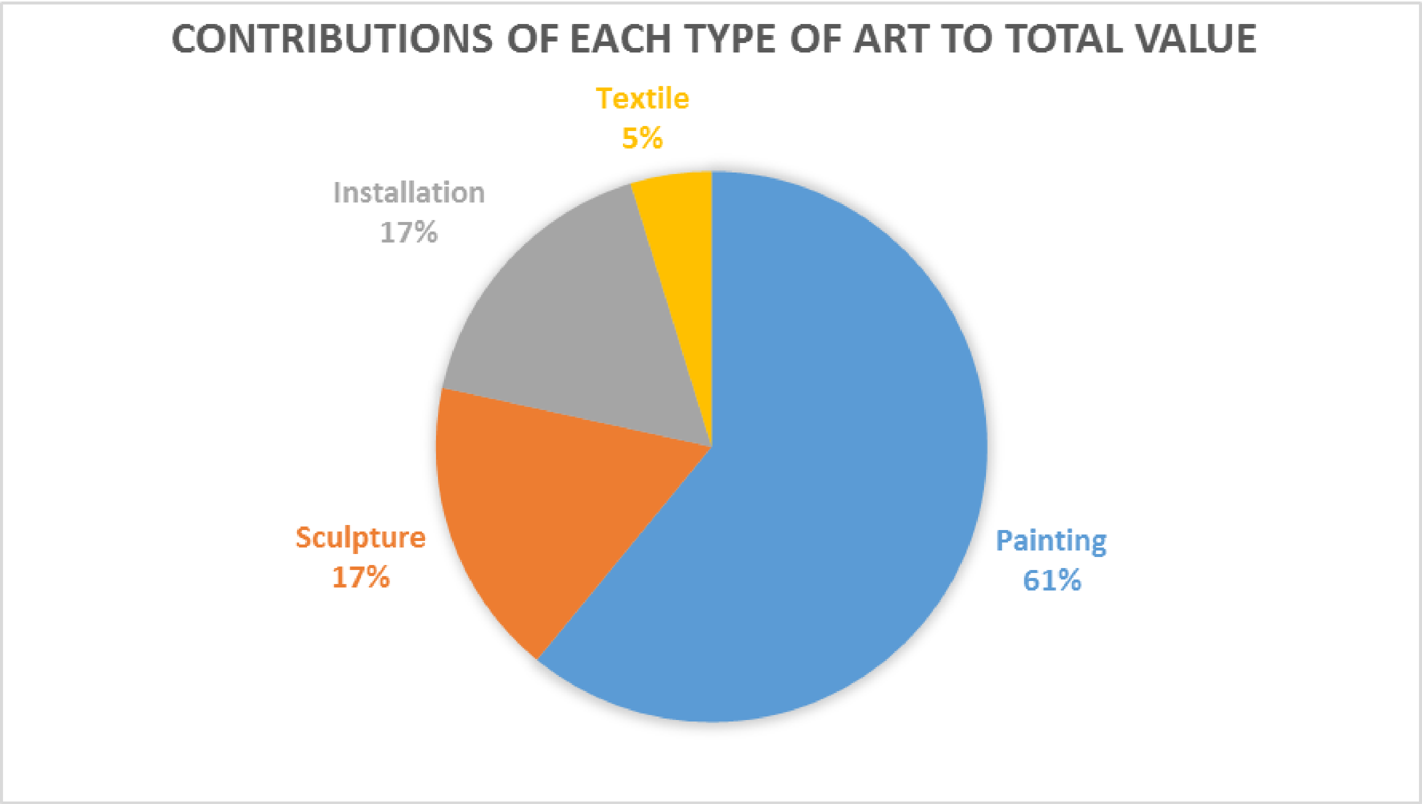


Figure 4.17: Pie Chart

Graphs

Definition 4.7: Histograms

- Is a chart that shows the distribution or frequency of occurrence of values of a single variable.
- We usually divide the values into bins and show a bar plot of the number of objects in each bin.
- The height of each bar indicates the number or frequency of the number of rows or objects in that bin

Graphs

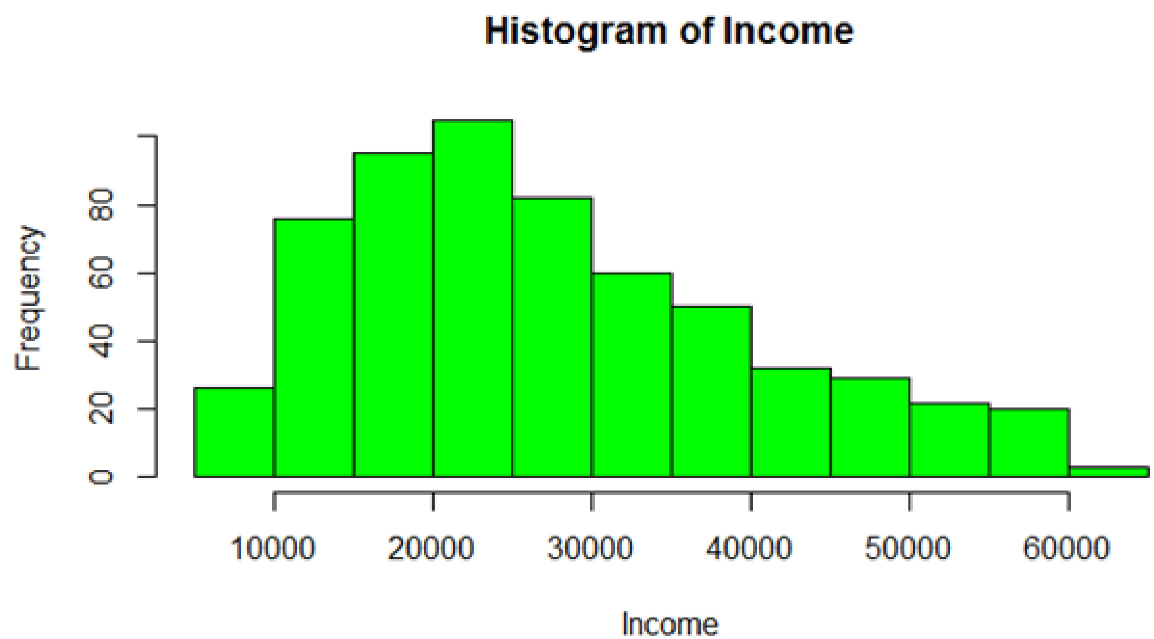


Figure 4.18: Histogram

Graphs

Definition 4.8: Box Plots

- A box plot is a convenient way of visually displaying the distribution of groups of numerical data through their quartiles.
- The box plot figure contains lines extending parallel from the boxes are known as the "whiskers", which are used to indicate the distribution below the first and above the third quartiles

eljalao@up.edu.ph

Graphs

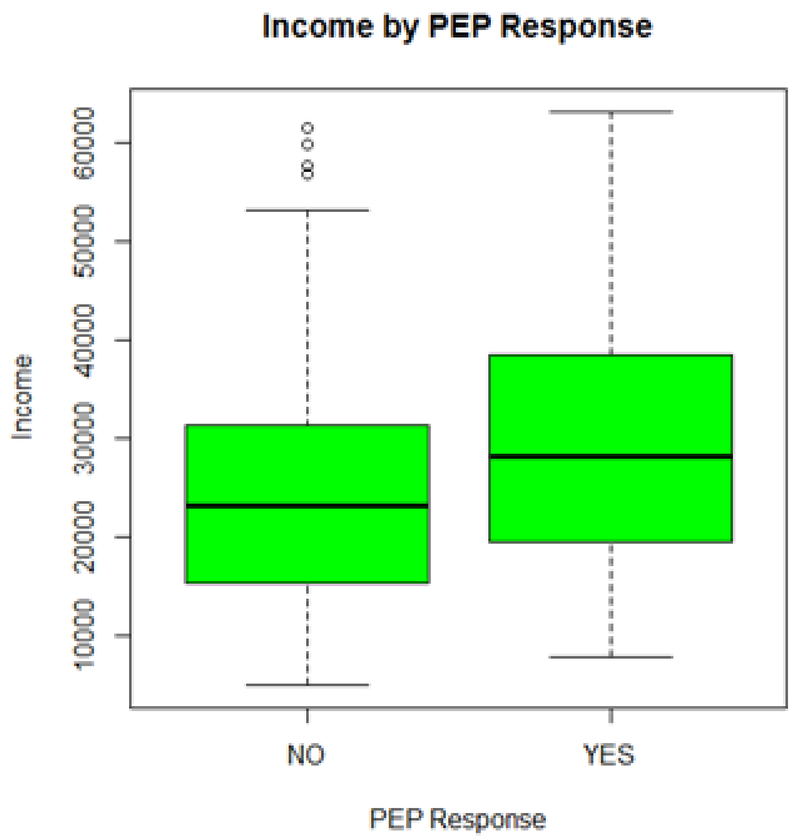


Figure 4.19: Box Plot

Graphs

Definition 4.9: Flow Charts

- Flow charts used to show the sequential steps of a process or procedure.
- These map out a process using a series of connected standard symbols, which makes the process easy to comprehend for other people.
- Flow charts are useful for explaining the intricacies and decisions of a complex process.

Graphs

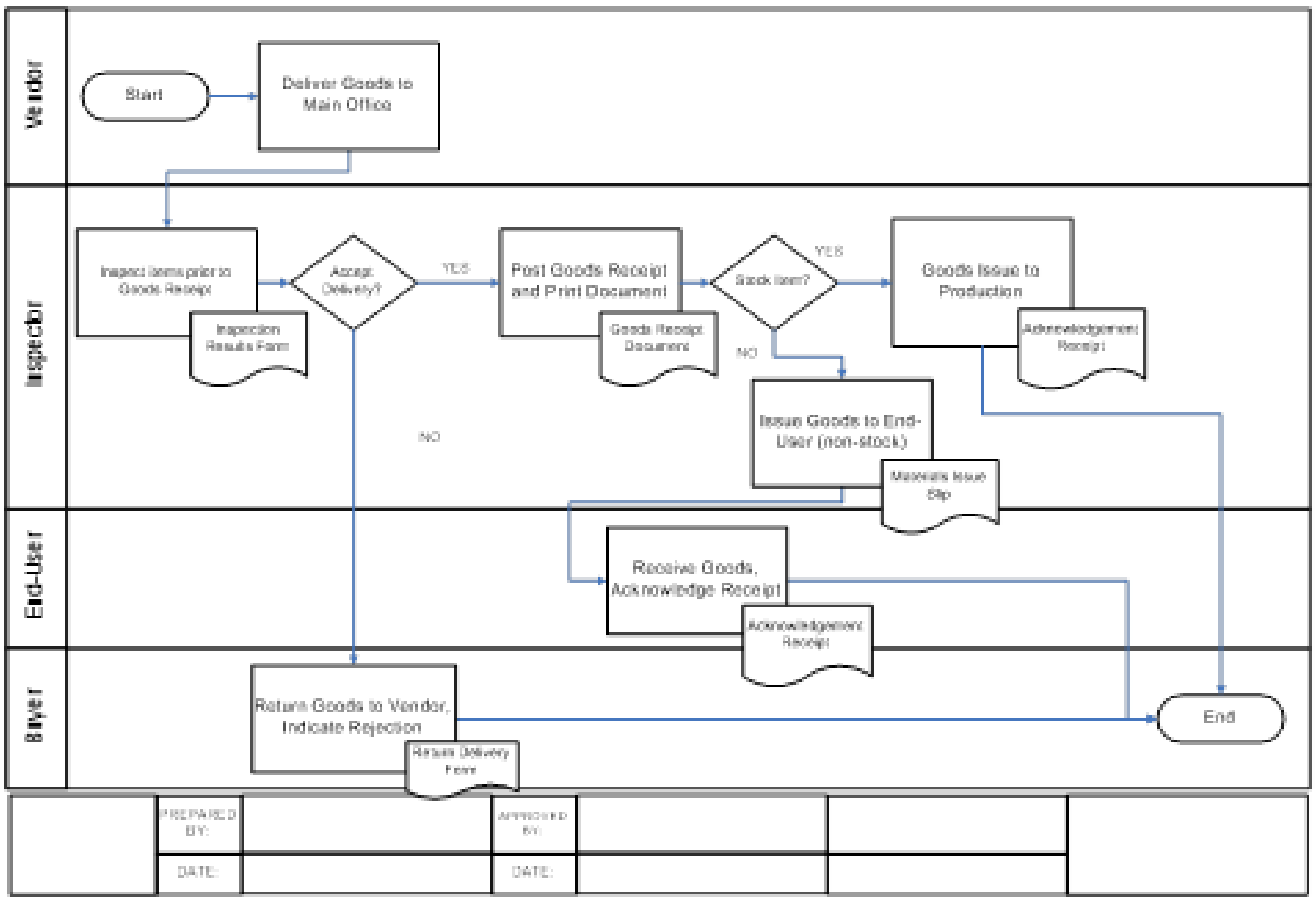


Figure 4.20: Box Plot

Graphs

Definition 4.10: Network Diagrams

- A network diagram is a type of visualization shows how elements are interconnected with nodes / vertices and link lines to represent their connections.
- These help illuminate the type of relationships between groups of entities.
- Network Diagrams have a limited data presentation capacity and start to become hard to interpret when there are too many nodes illustrated in a single diagram.

Graphs

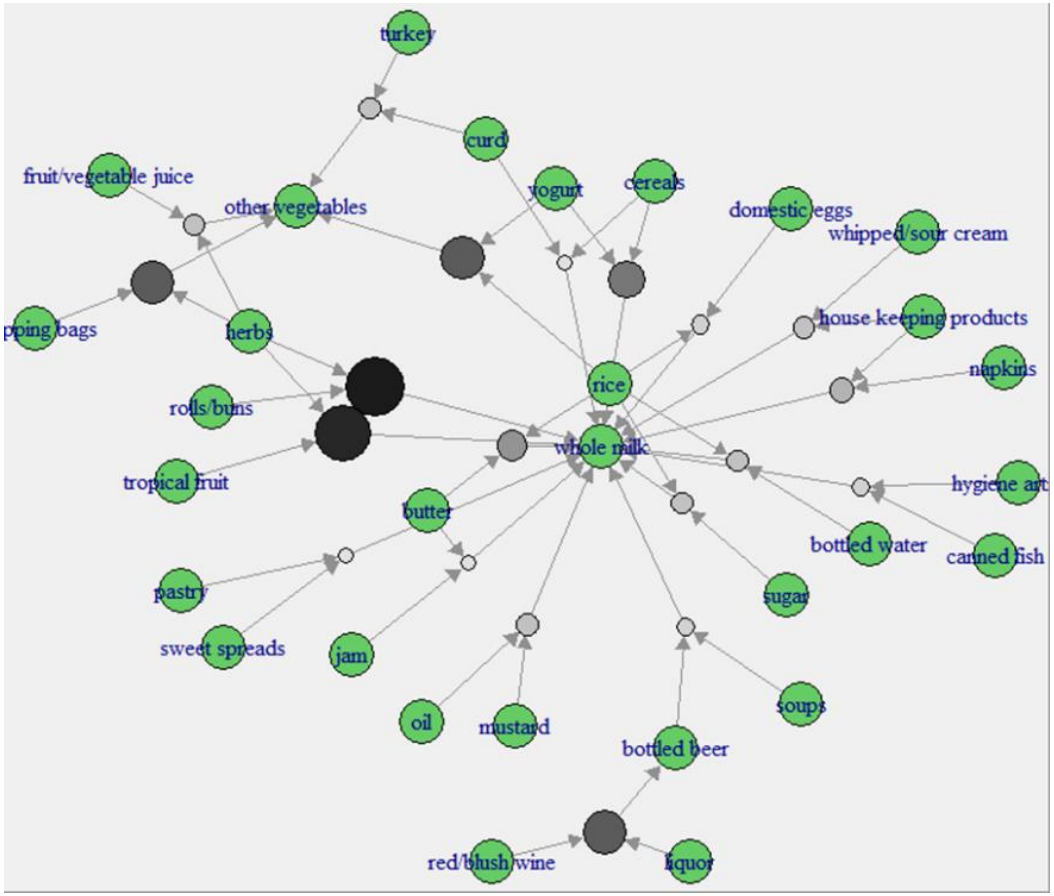


Figure 4.21: A Network Diagram

Graphs

Definition 4.11: Tree Maps

- Tree maps visualize the hierarchical structure of a dataset by displaying quantities for each category via area size.
- Each category value is assigned a rectangle area with their subcategory rectangles nested inside of it wherein all subcategories account for 100% of their parent.

Graphs

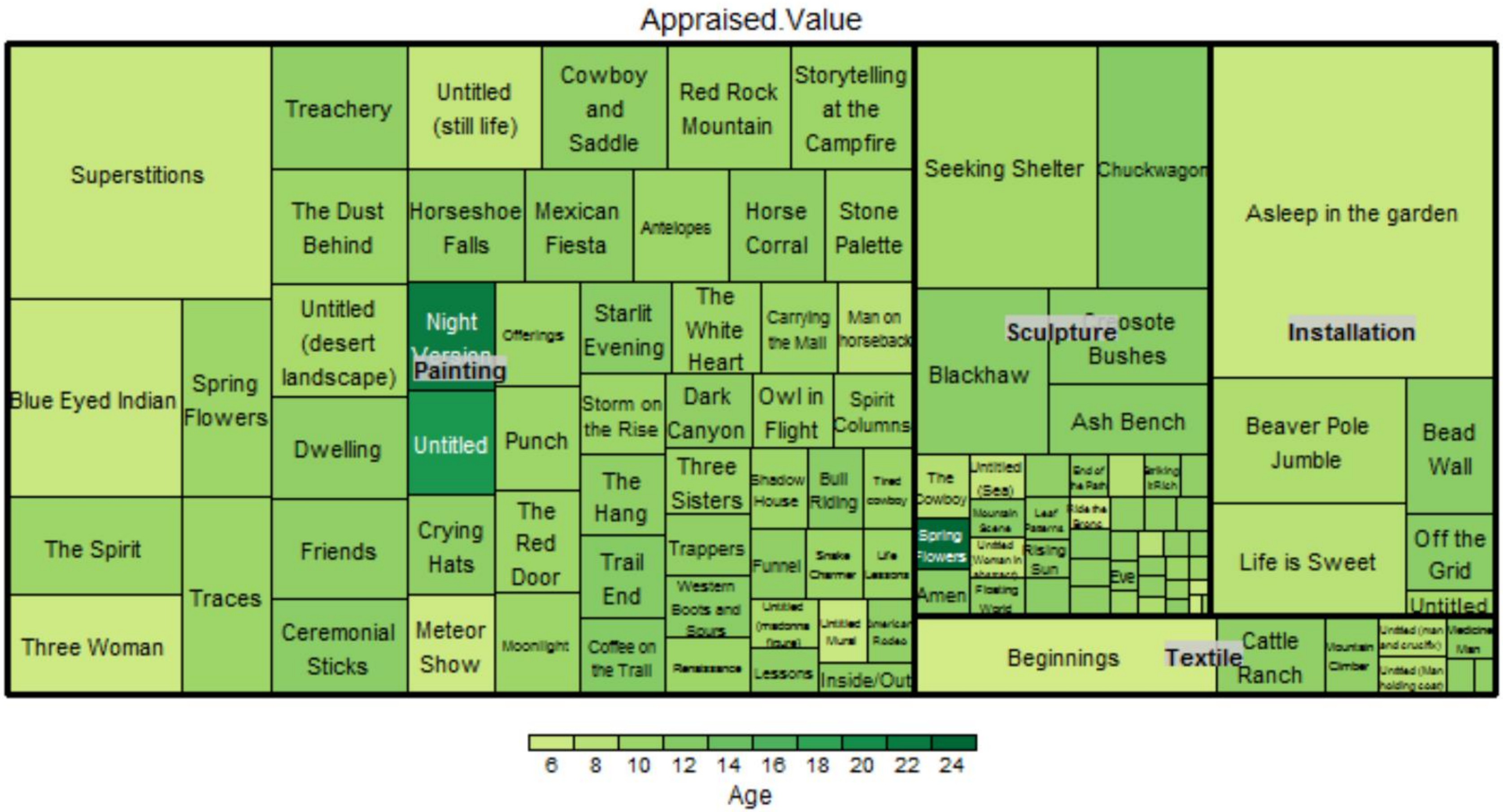


Figure 4.22: An Example of a Tree Map

Graphs

Definition 4.12: Choropleth Maps

- Choropleth Maps present divided geographical areas or regions that are colored, or patterned in relation to a variable in question.
- This provides a way to visualize values of the variable in question over a geographical area, which can show variation or patterns across the displayed location.

Graphs

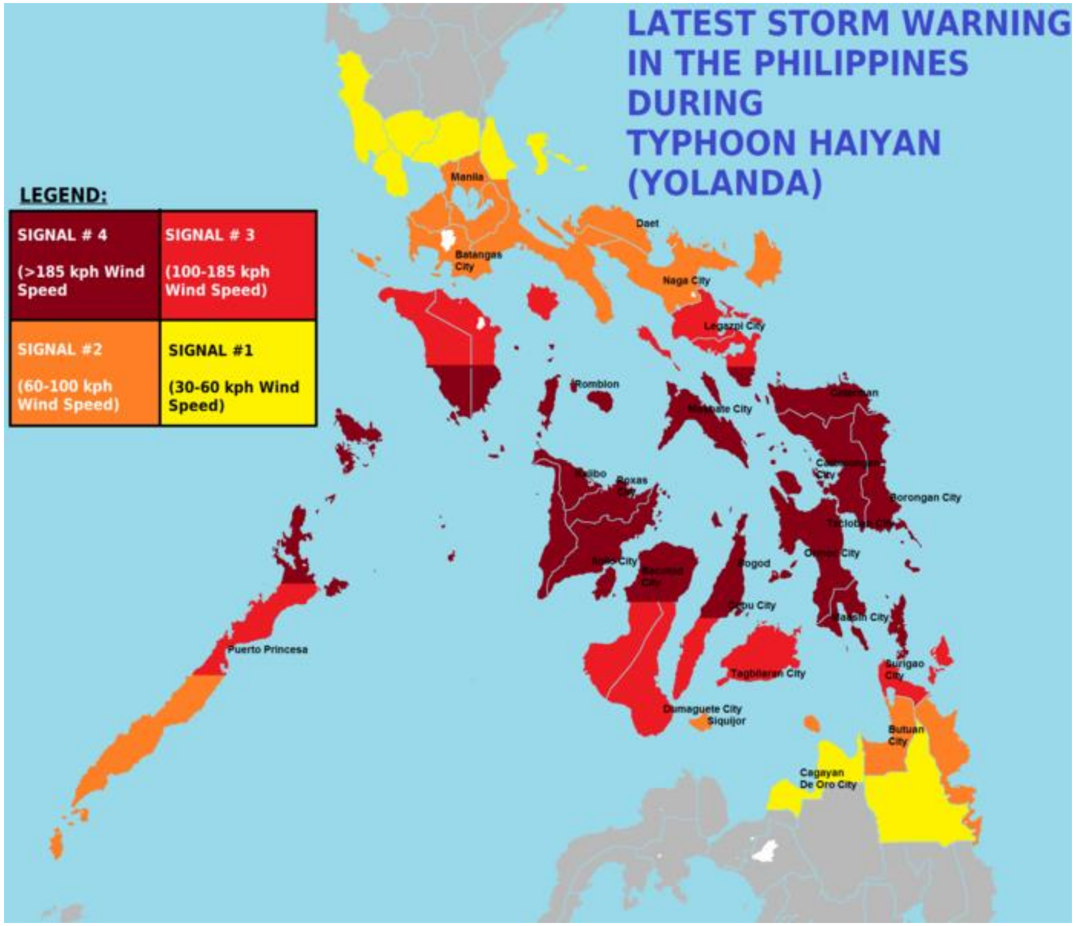


Figure 4.23: Choropleth Map

Visual Design Principles

Definition 4.13: Charts

- A medium of visualization that shows a structure that relates entities to one another
  - lines and relative position serve as links
  - Examples:
    family tree,
    flow chart,
    network diagram

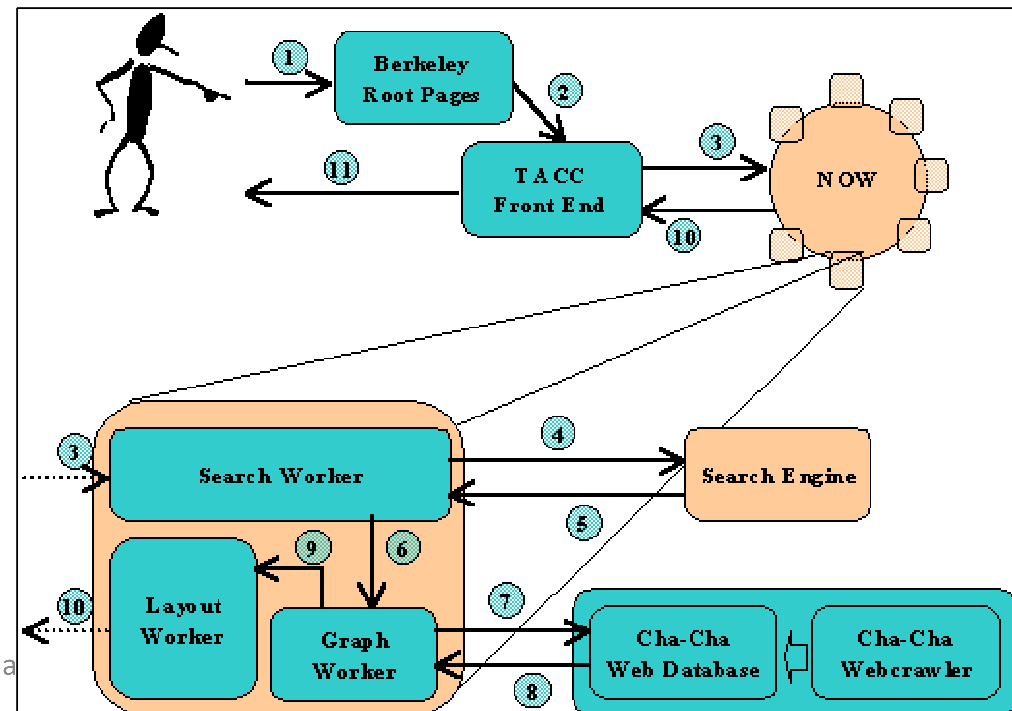


Figure 4.24: Sample Chart

Visual Design Principles

Definition 4.14: Diagrams

- A medium of visualization that shows schematic pictures of objects or entities
  - parts are symbolic (unlike photographs)
  - Examples:
    how-to illustrations,
    figures in a manual

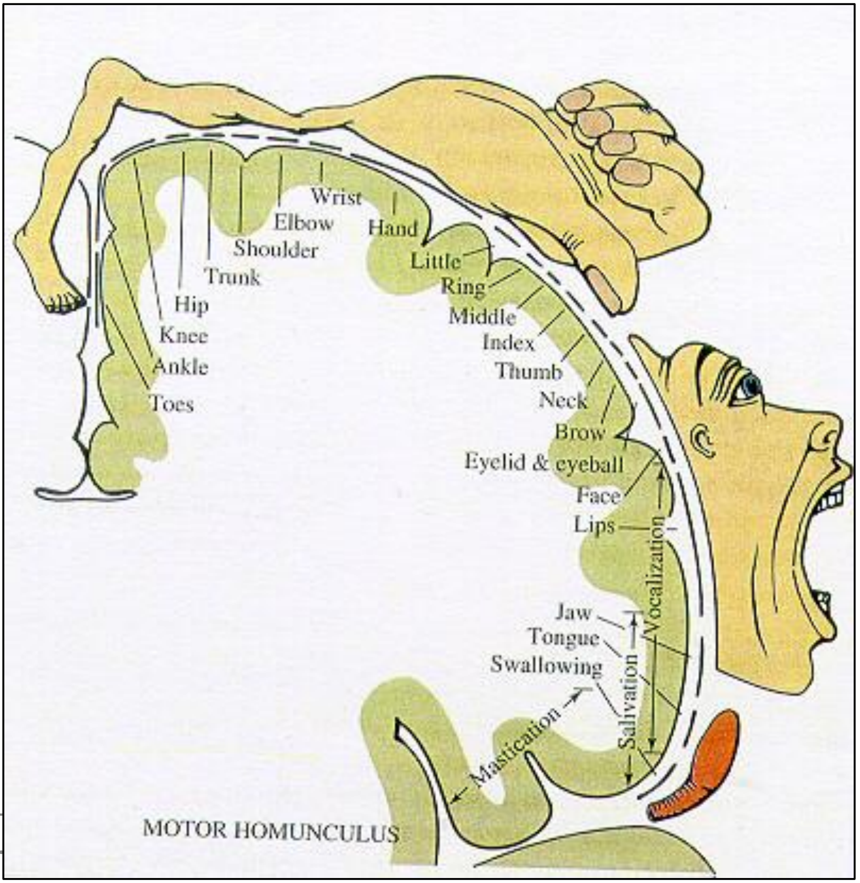


Figure 4.25: Diagram

Outline for this Session

- What is Visualization?
- Types of Visualizations
- **Visual Design Principles**
- Design Optimization
- Case Study

Visual Design Principles

- Classifying Visual Representations
  - Study by: Lohse et al. 1994
  - Participants sorted 60 visualizations into categories
    - Others assigned labels from Likert scales
    - Experimenters clustered the results various ways.

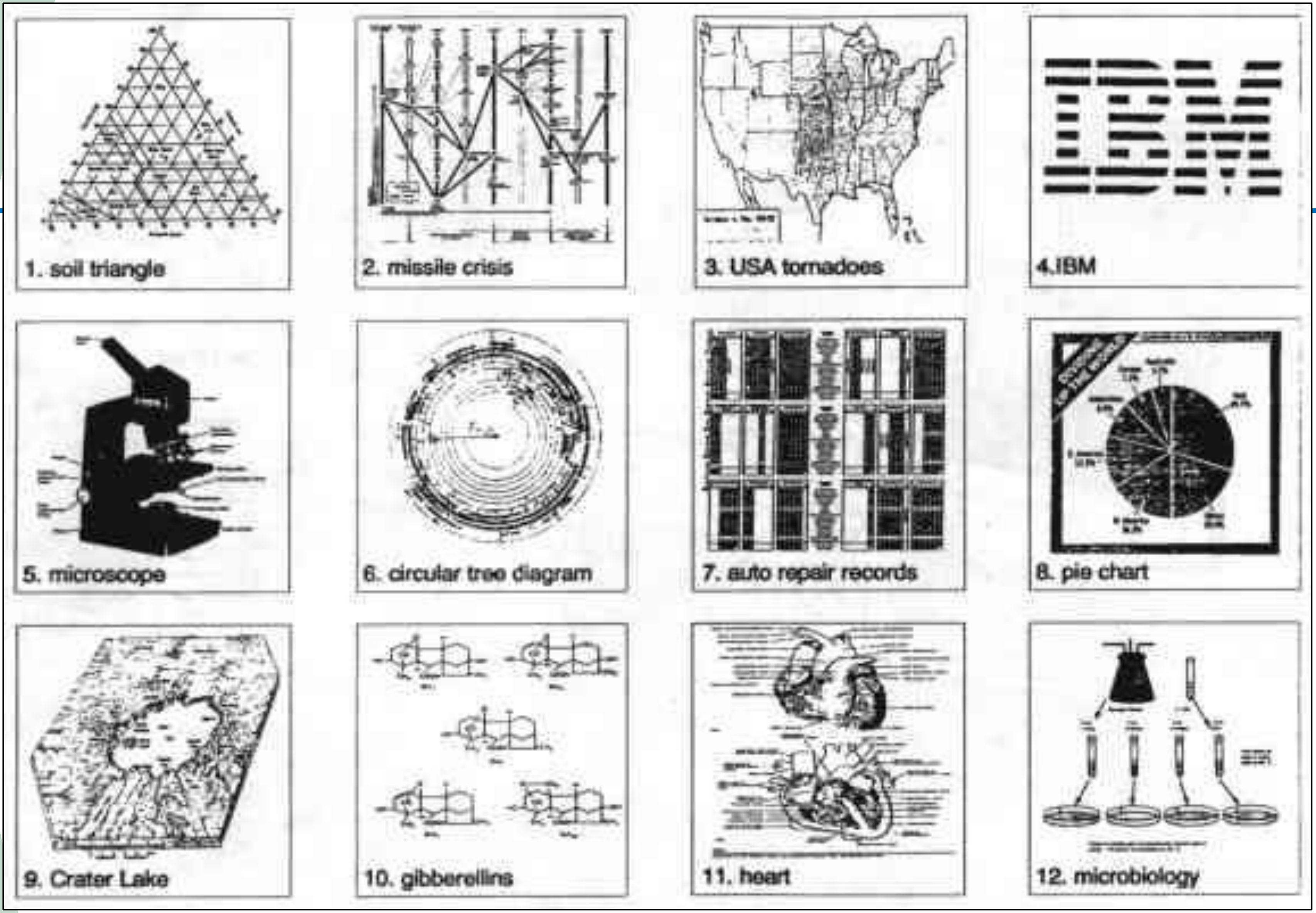


Figure 4.26a: Subset of Example Visual Representations From Lohse et al. 94

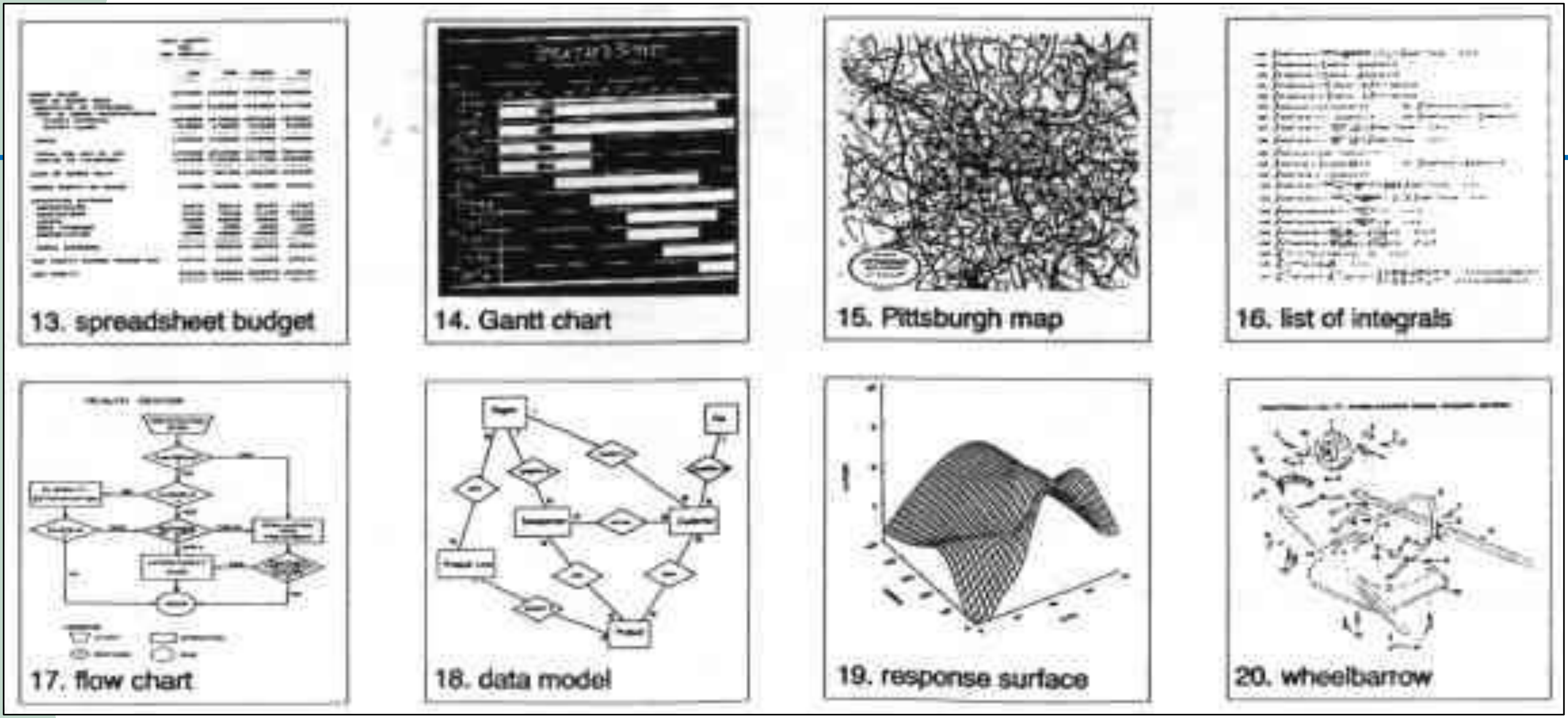


Figure 4.26b: Subset of Example Visual Representations From Lohse et al. 94

Visual Design Principles

- Photorealistic images were least informative
  - Echos results in icon studies – better to use less complex, more schematic images
- Graphs and tables are the most self-similar categories
- Temporal data more difficult to show than cyclic data
  - Recommend using animation for temporal data

Visual Design Principles

- Visual Properties
  - Preattentive Processing
  - Accuracy of Interpretation of Visual Properties
  - Illusions and the Relation to Graphical Integrity

Visual Design Principles

Definition 4.15: Preattention

- Visual information is processed without the need of **focusing attention**
- Important for design of visualizations
  - what can be perceived immediately
  - what properties are good discriminators
- < 200–250 ms qualifies as pre-attentive
  - eye movements take at least 200ms
  - yet certain processing can be **done very quickly**, implying low-level processing in parallel
- If a decision takes a fixed amount of time regardless of the number of distractors, it is considered to be preattentive

Visual Design Principles

- Viewer can rapidly and accurately determine whether the target (red circle) is present or absent. Difference detected in color.

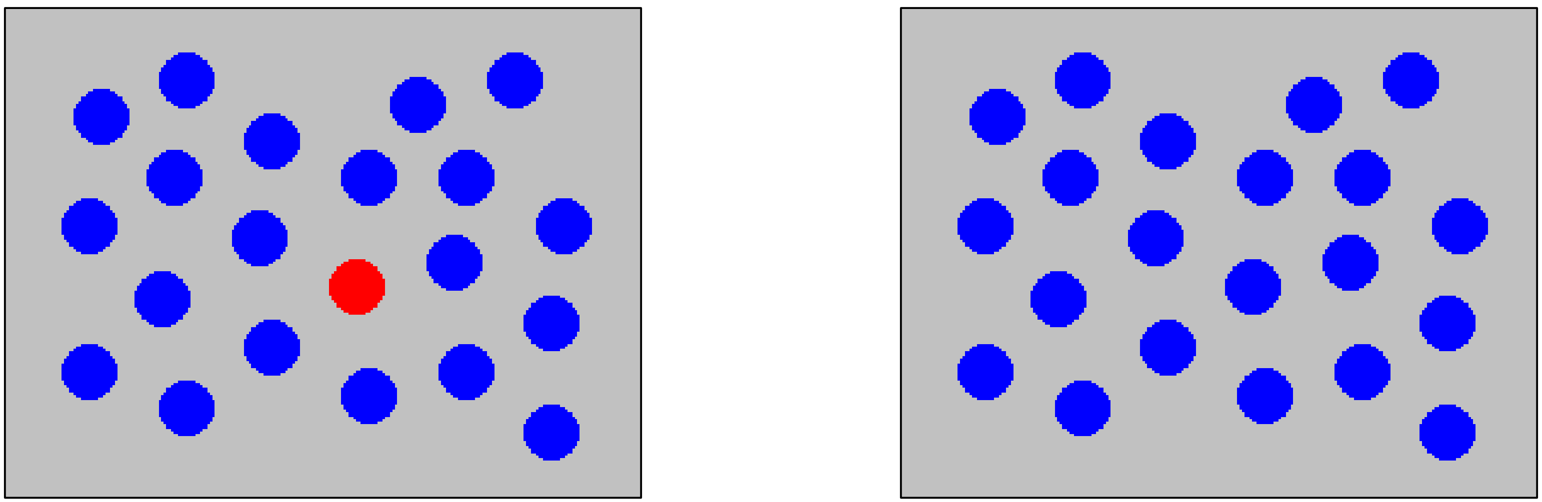


Figure 4.16: Color Selection Preattention

Visual Design Principles

- Viewer can rapidly and accurately determine whether the target (red circle) is present or absent. Difference detected in form (curvature)

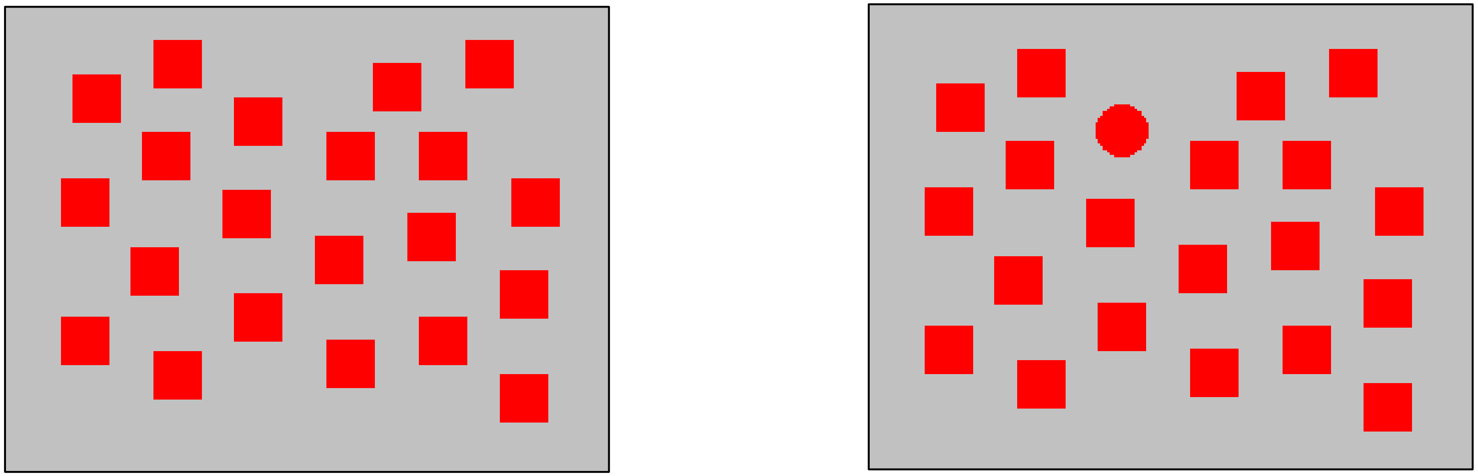


Figure 4.17: Shape Selection Preattention

Visual Design Principles

- Viewer ***cannot*** rapidly and accurately determine whether the target (red circle) is present or absent when target has two or more features, each of which are present in the distractors. Viewer must search sequentially.

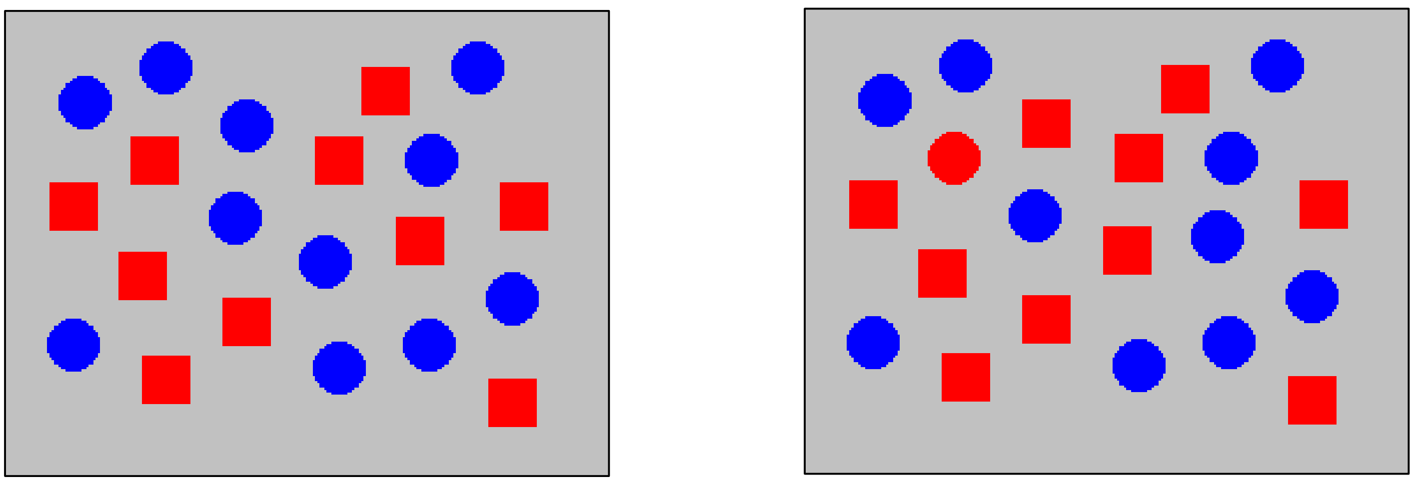
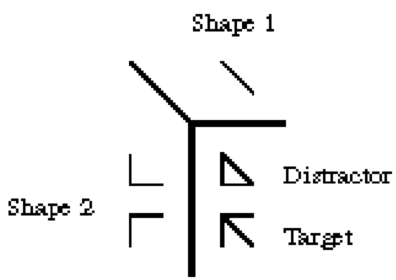


Figure 4.18: Conjunction of Features

Visual Design Principles

- Target has a **unique feature** with respect to distractors (open sides) and so the group can be detected preattentively.



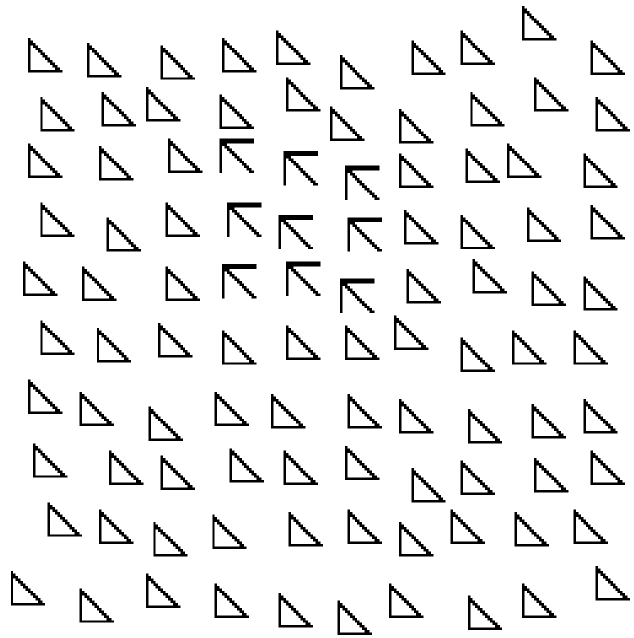


Figure 4.19: Emergent Features

Visual Design Principles

- Target does **not** have a unique feature with respect to distractors and so the group **cannot** be detected preattentively.

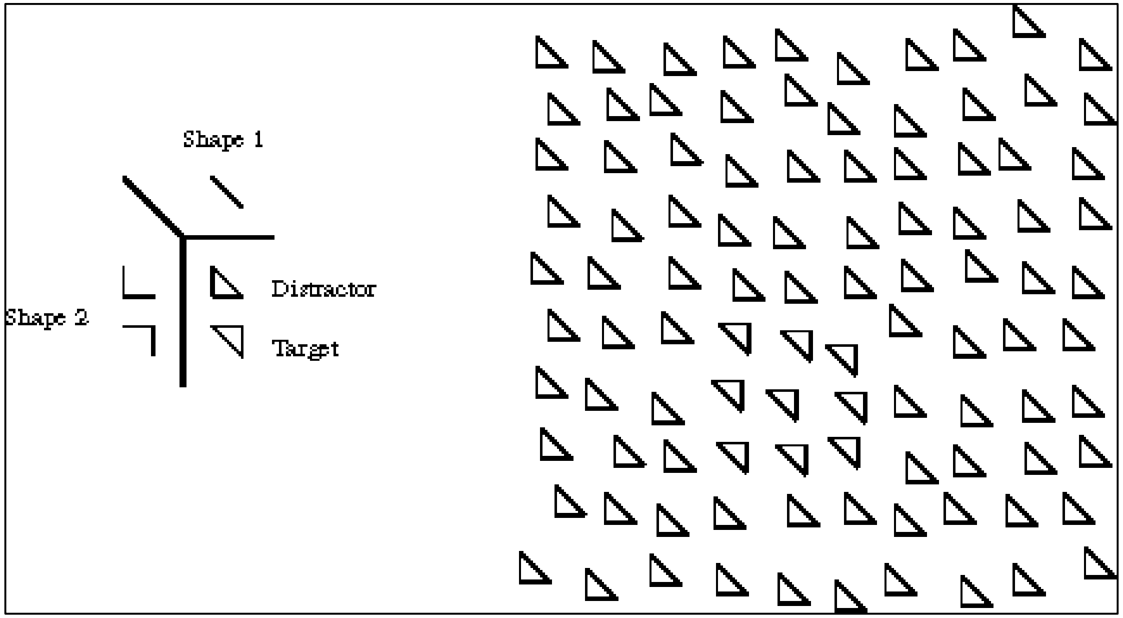


Figure 4.20: Emergent Features

Visual Design Principles

- Some properties are **asymmetric**
  - a sloped line among vertical lines is preattentive
  - a vertical line among sloped ones is not

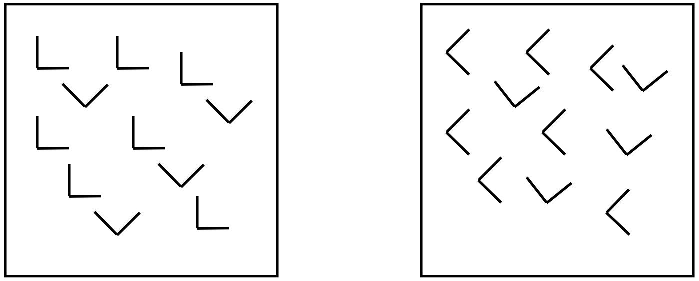


Figure 4.20: Asymmetric Preattentive Properties

Visual Design Principles

- Block Text is NOT Preattentive

SUBJECT PUNCHED QUICKLY OXIDIZED TCEJBUS DEHCNUP YLKCIUQ DEZIDIXO
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC
GOVERNS PRECISE EXAMPLE MERCURY SNREVOG ESICERP ELPMAXE YRUCREM
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM
GOVERNS PRECISE EXAMPLE MERCURY SNREVOG ESICERP ELPMAXE YRUCREM
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC
SUBJECT PUNCHED QUICKLY OXIDIZED TCEJBUS DEHCNUP YLKCIUQ DEZIDIXO
CERTAIN QUICKLY PUNCHED METHODS NIATREC YLKCIUQ DEHCNUP SDOHTEM
SCIENCE ENGLISH RECORDS COLUMNS ECNEICS HSILGNE SDROCER SNMULOC

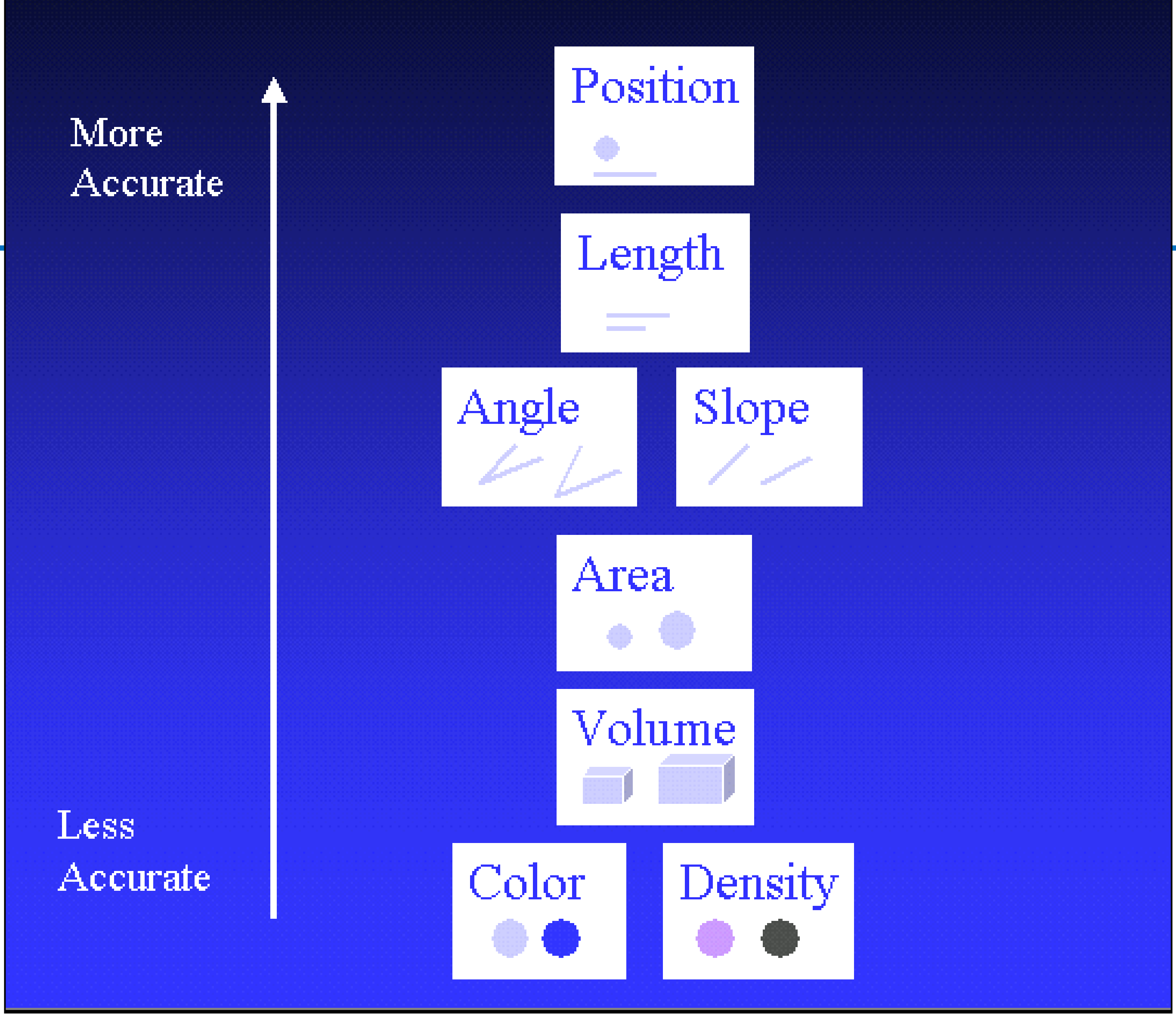


Figure 4.21: Accuracy Ranking of Quantitative Perceptual Tasks Estimated

Visual Design Principles

- Visual Illusions
  - People don’t perceive length, area, angle, brightness the way they **“should”**
  - Some illusions have been reclassified as systematic **perceptual errors**
    - e.g., brightness contrasts (grey square on white background vs. on black background)
    - partly due to increase in our understanding of the relevant parts of the visual system
  - Nevertheless, the visual system does some really unexpected things

Visual Design Principles

- Illusions of Linear Extent
  - Mueller-Lyon (off by 25-30%)

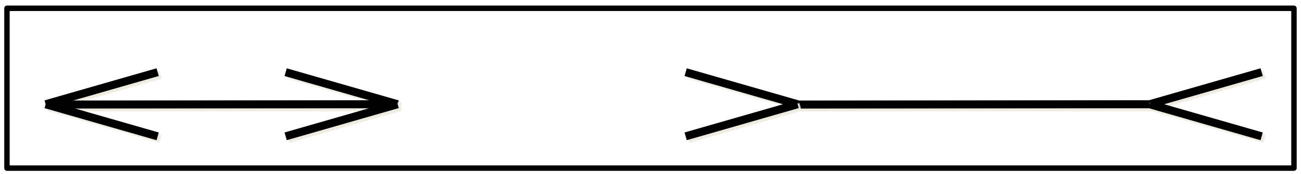


Figure 4.22: Example of a Mueller-Lyon Illusion

  - Delboeuf Illusion

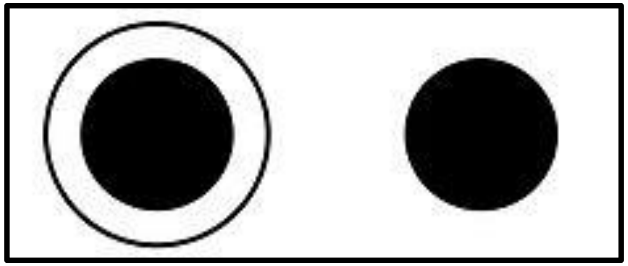


Figure 4.23: Example of a Delboeuf Illusion

Illusions of Area

- Ebbinghaus Illusion

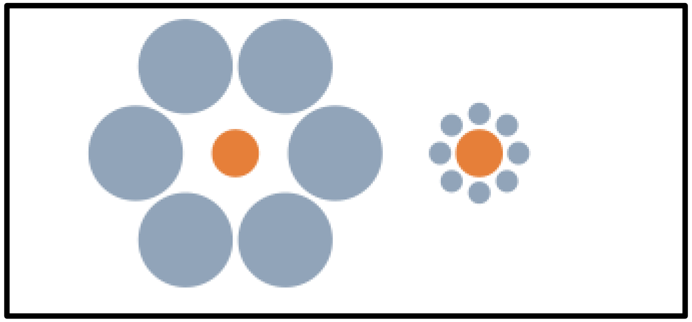


Figure 4.24: Example of a Ebbinghaus Illusion

Outline for this Session

- What is Visualization?
- Types of Visualizations
- Visual Design Principles
- **Design Optimization**
- Case Study

Design Optimization

- Tufte's Principles of Graphical excellence
  - is the **well-designed presentation** of interesting data – a matter of substance, of statistics, and of design
  - consists of **complex ideas** communicated with clarity, precision and efficiency
  - is that which gives to the viewer the **greatest number of ideas** in the shortest time with the **least ink** in the **smallest space**
  - requires telling the truth about the data

Design Optimization

- Tufte's Principles of Graphical excellence
  - Use multifunctioning **graphical elements**
  - Use small **multiples**
  - Show mechanism, process, dynamics, and causality
  - High data **density**
    - Number of items/area of graphic
    - This is controversial
      - White space thought to contribute to good visual design
      - Tufte's book itself has lots of white space

Design Optimization

- Tufte's Graphical Integrity
  - Some lapses intentional, some not

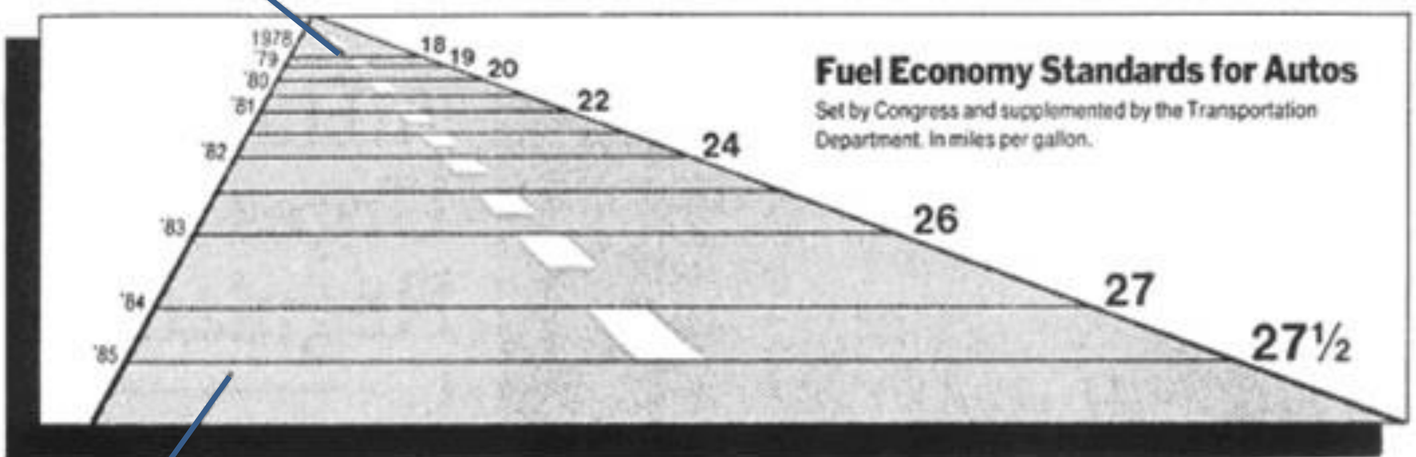
$$Lie\ Factor = \frac{size\ of\ effect\ in\ graph}{size\ of\ effect} \tag{4.1}$$

  - Misleading uses of area and perspective
  - Lie Factor should have a value between 0.95 and 1.05.
  - If the value is less or greater, it indicates a substantial (and often intended) distortion

Design Optimization

Example 4.1: Calculating Lie Factor

Line representing 18 mpg is 0.6 inches long



[Tufte, 1991]

Line representing 27.5 mpg is 5.3 inches long

Design Optimization

Example 4.1 (Cont.): Calculating Lie Factor

$$Lie\ Factor = \frac{size\ of\ effect\ in\ graph}{size\ of\ effect}$$

$$Lie\ Factor = \frac{\frac{5.3 - 0.6}{0.6}}{\frac{27.5 - 18}{18}} = 14.8$$

Design Optimization

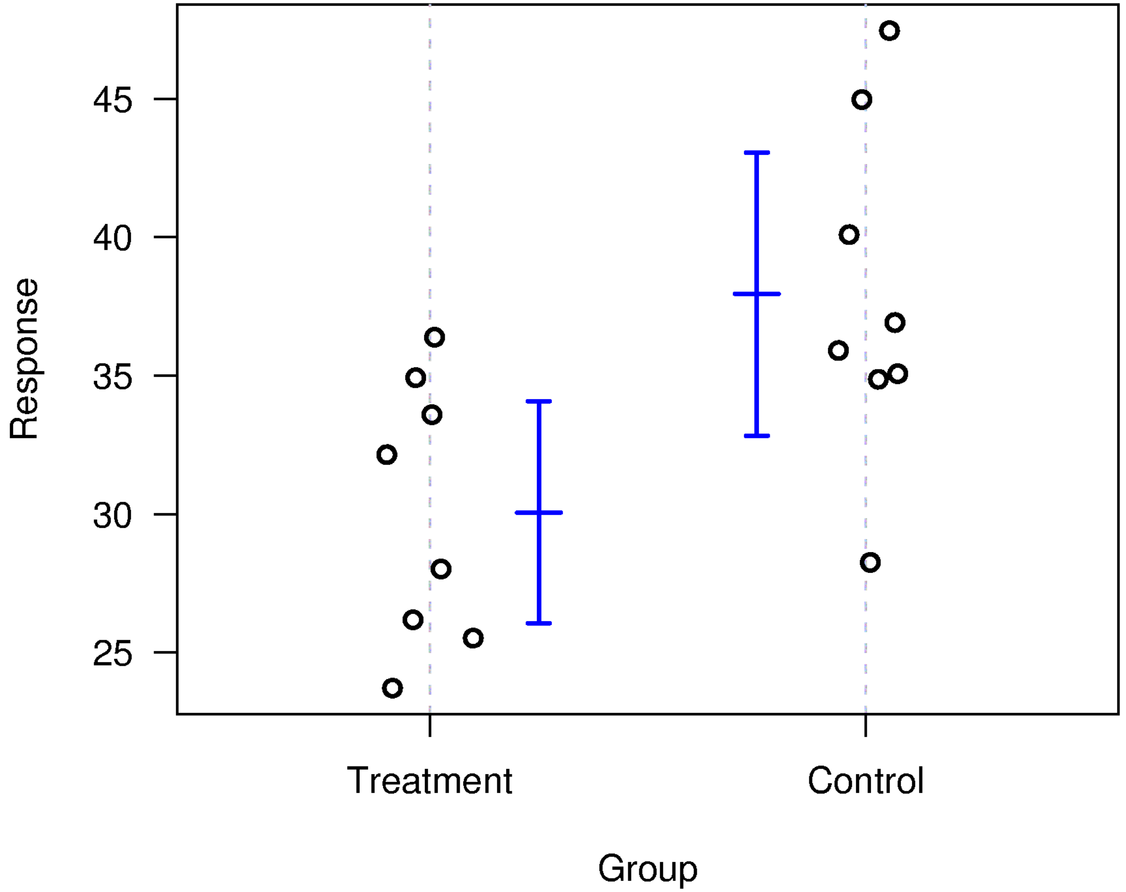
- Tufte's Principle of Data Ink Maximization
  - Goal: maximize ratio of "**data ink**" to total ink
    - draw viewers' attention to the substance of the graphic
    - the role of redundancy
    - principles of editing and redesign

$$Data\ Ink\ Ratio = \frac{data\ ink}{total\ ink\ used} \tag{4.2}$$

  - **Good visualization** should include as much as possible only data-Ink.
  - Non-Data-Ink is to be deleted everywhere where possible.
  - Avoids drawing the attention of viewers of the data presentation to irrelevant elements.

Design Optimization

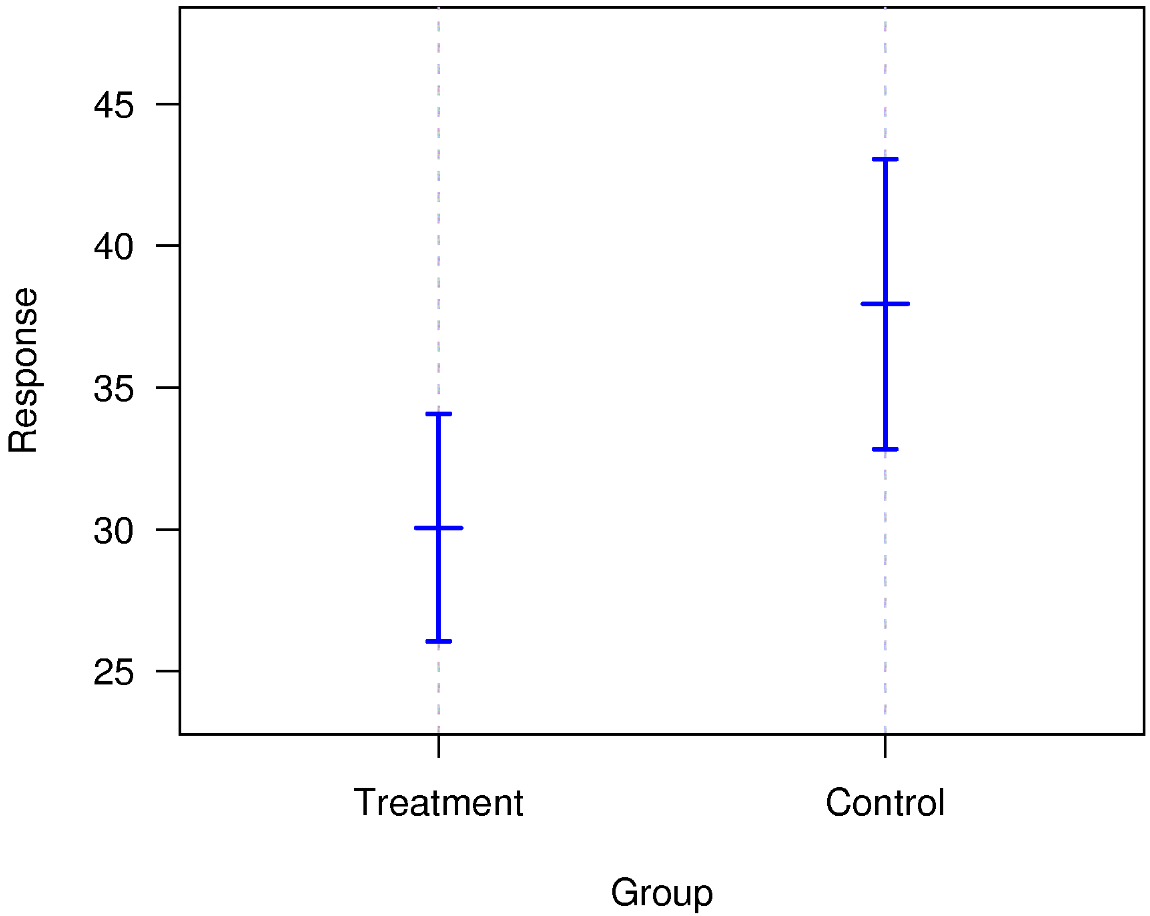
Example 4.2: Chart Ink



[Karl Broman]

Design Optimization

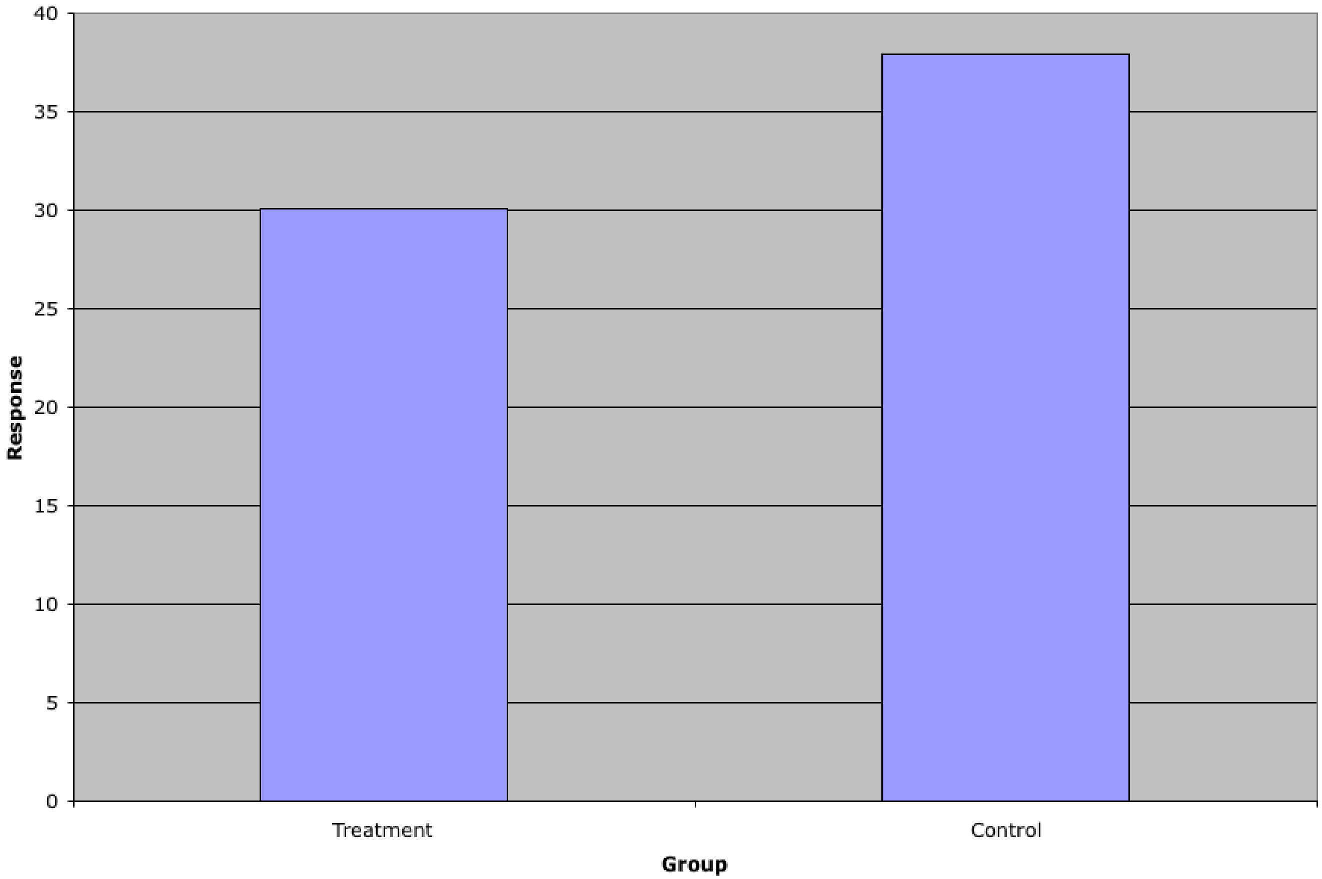
Example 4.2 (Cont.): Chart Ink



[Karl Broman]

Design Optimization

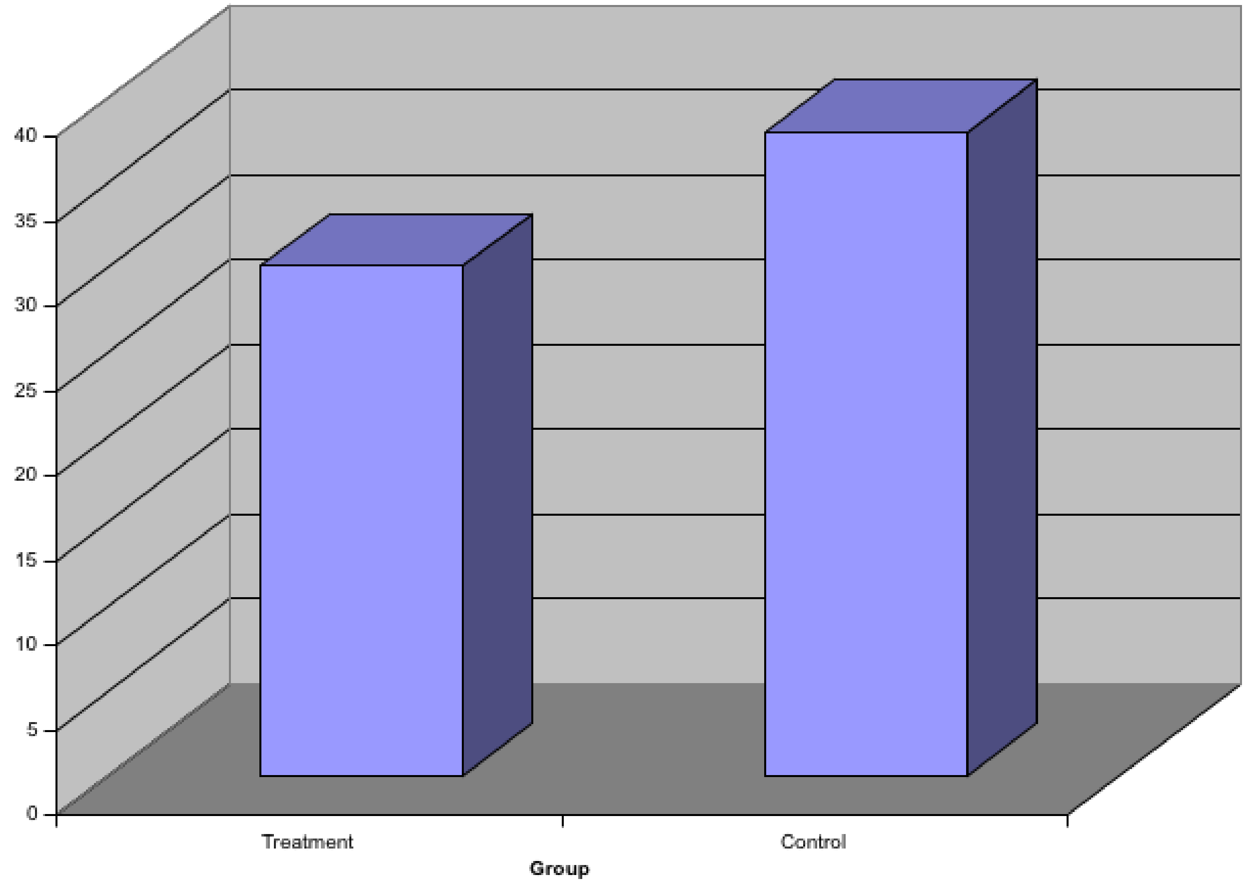
Example 4.3: Chart Ink



[Karl Broman]

Design Optimization

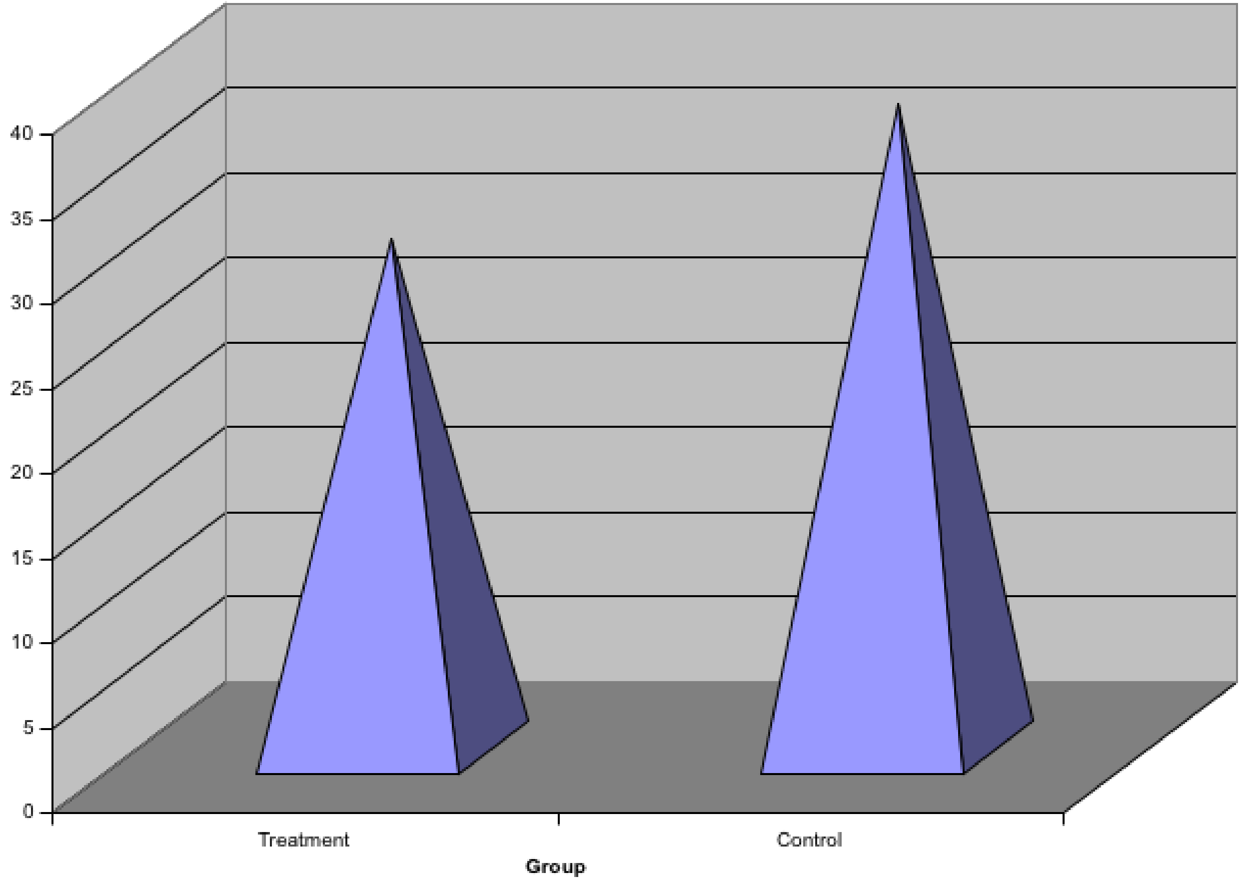
Example 4.3 (Cont.): Chart Ink



[Karl Broman]

Design Optimization

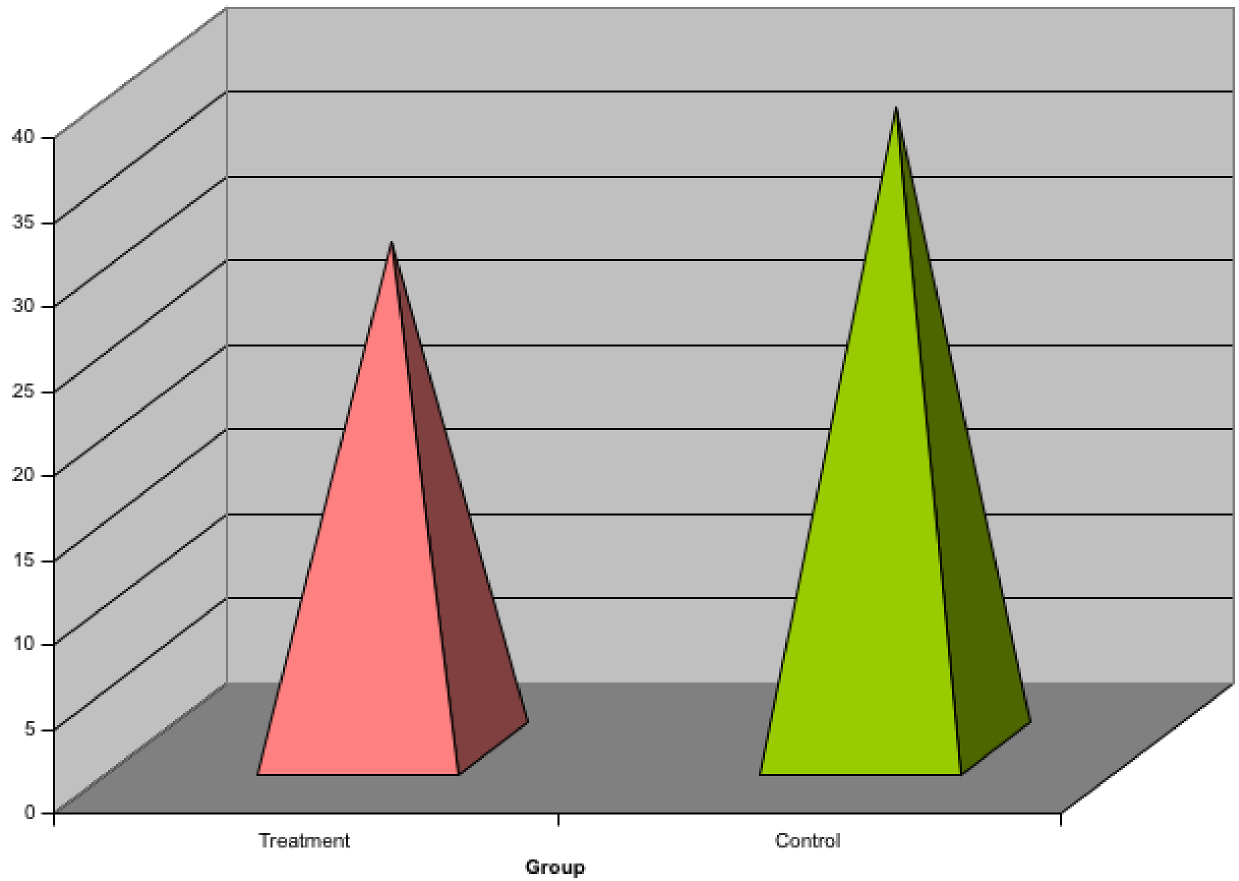
Example 4.3 (Cont.): Chart Ink



[Karl Broman]

Design Optimization

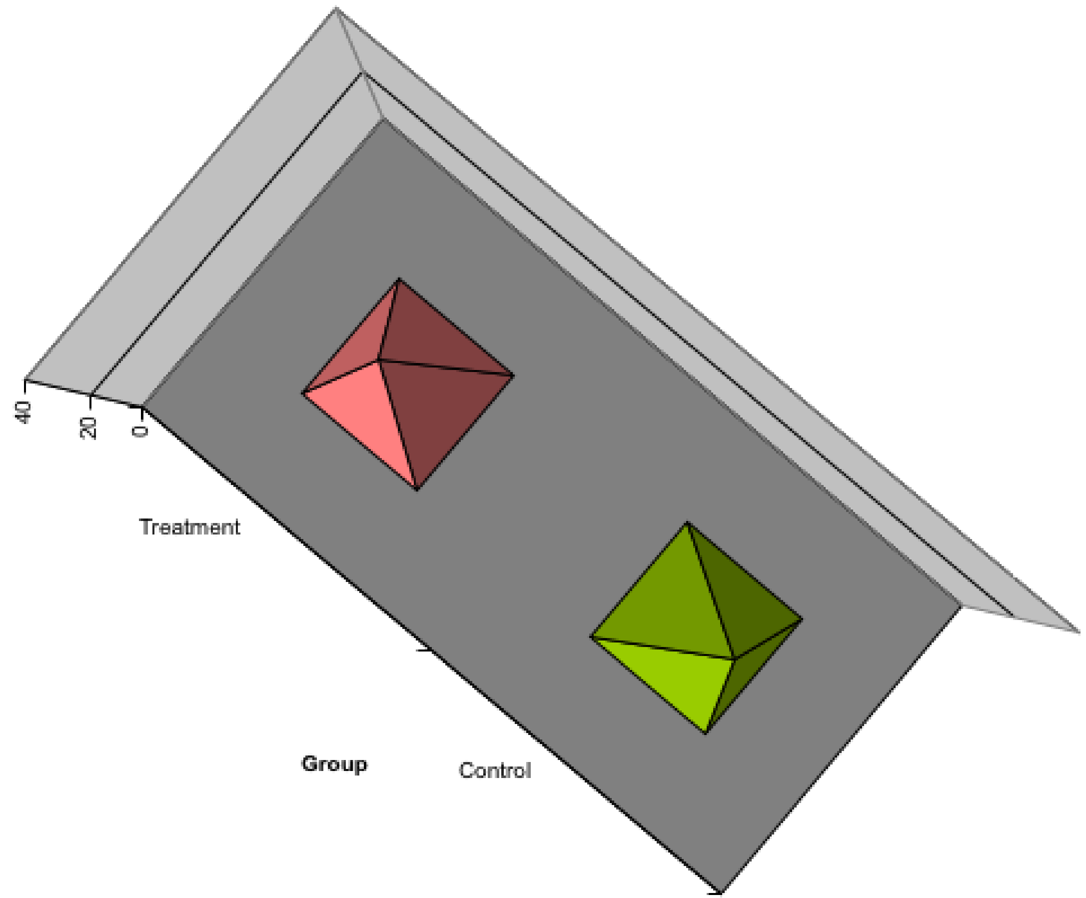
Example 4.3 (Cont.): Chart Ink



[Karl Broman]

Design Optimization

Example 4.3 (Cont.): Chart Ink



[Karl Broman]

Design Optimization

Example 4.4: Chart Ink

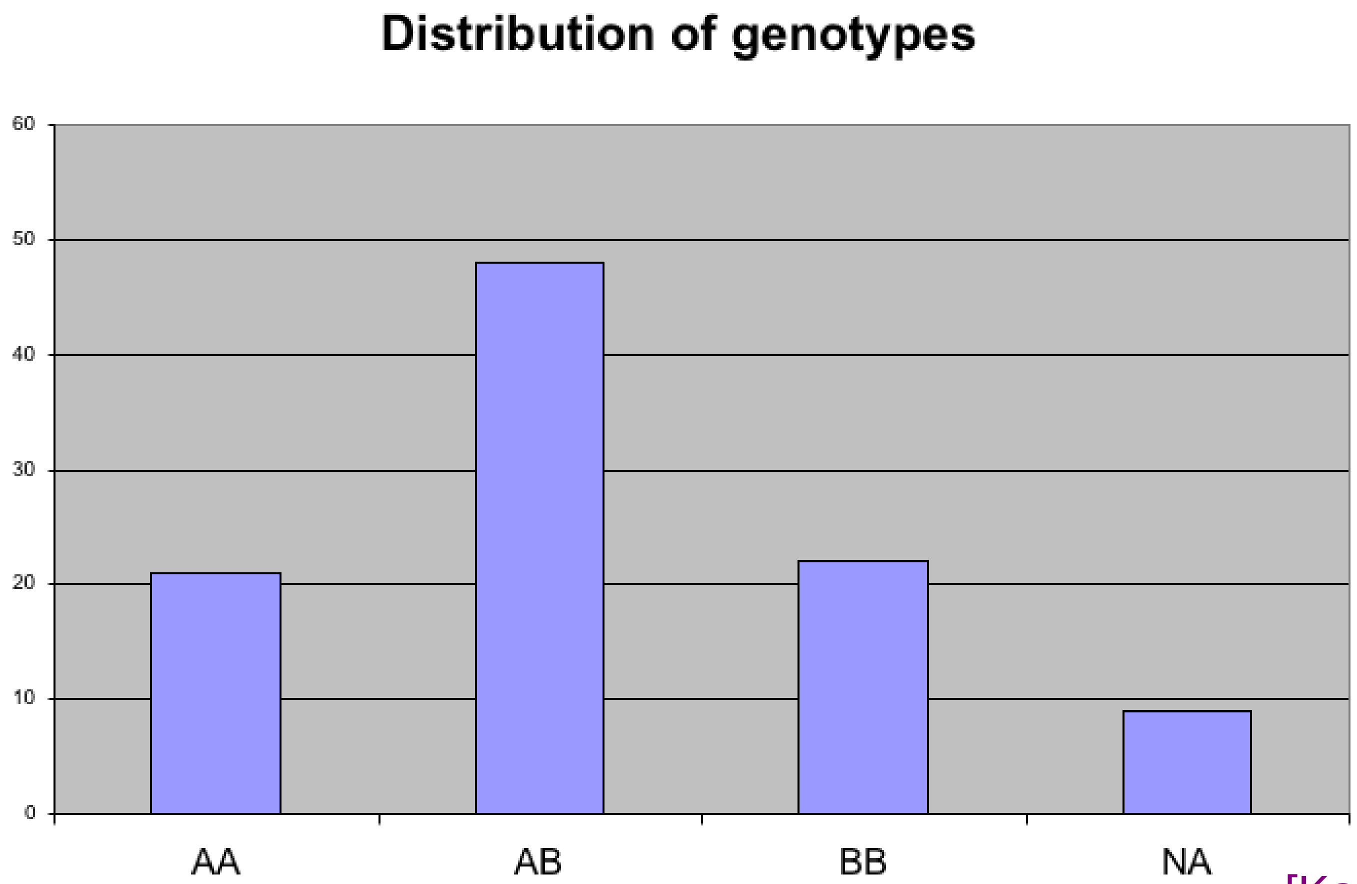
Distribution of genotypes

AA	21%
AB	48%
BB	22%
missing	9%

[Karl Broman]

Design Optimization

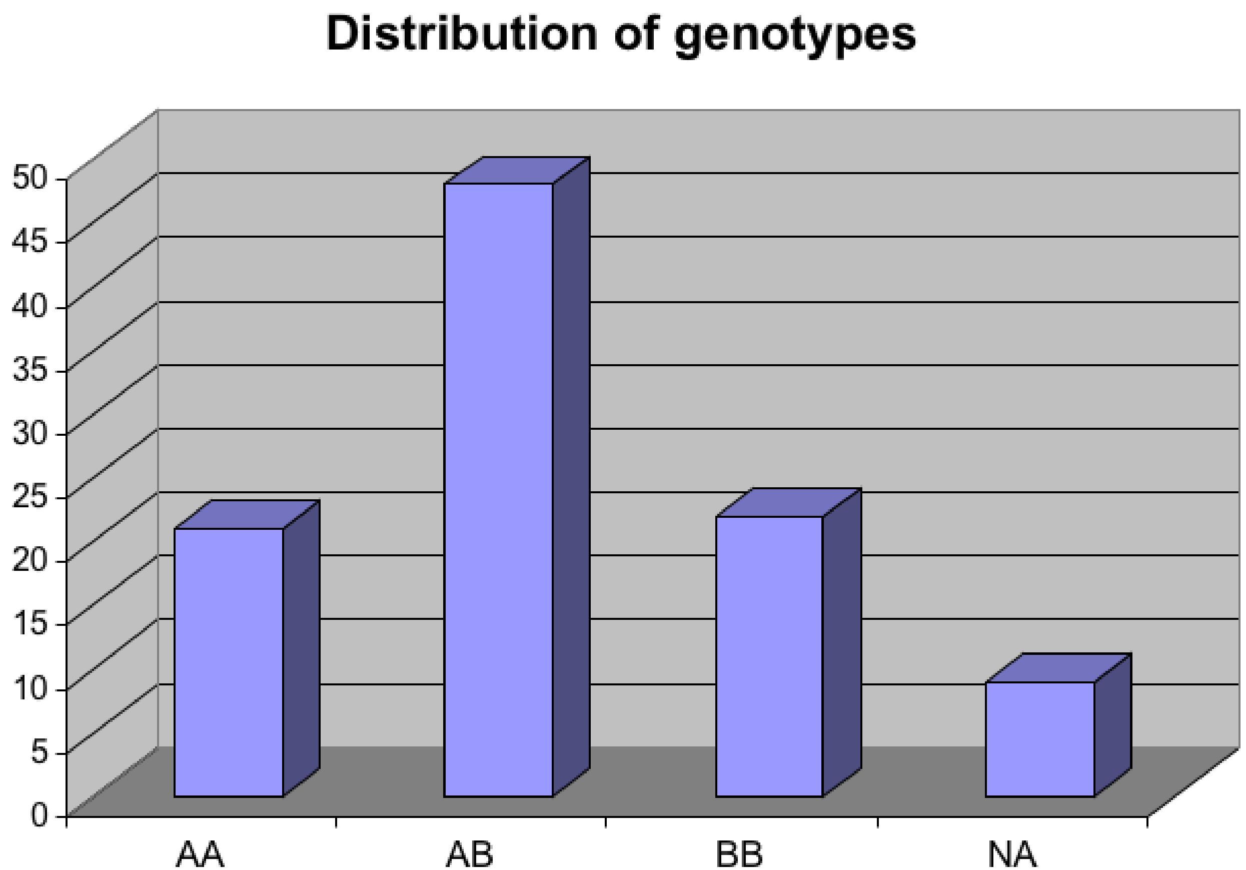
Example 4.4 (Cont.): Chart Ink



[Karl Broman]

Design Optimization

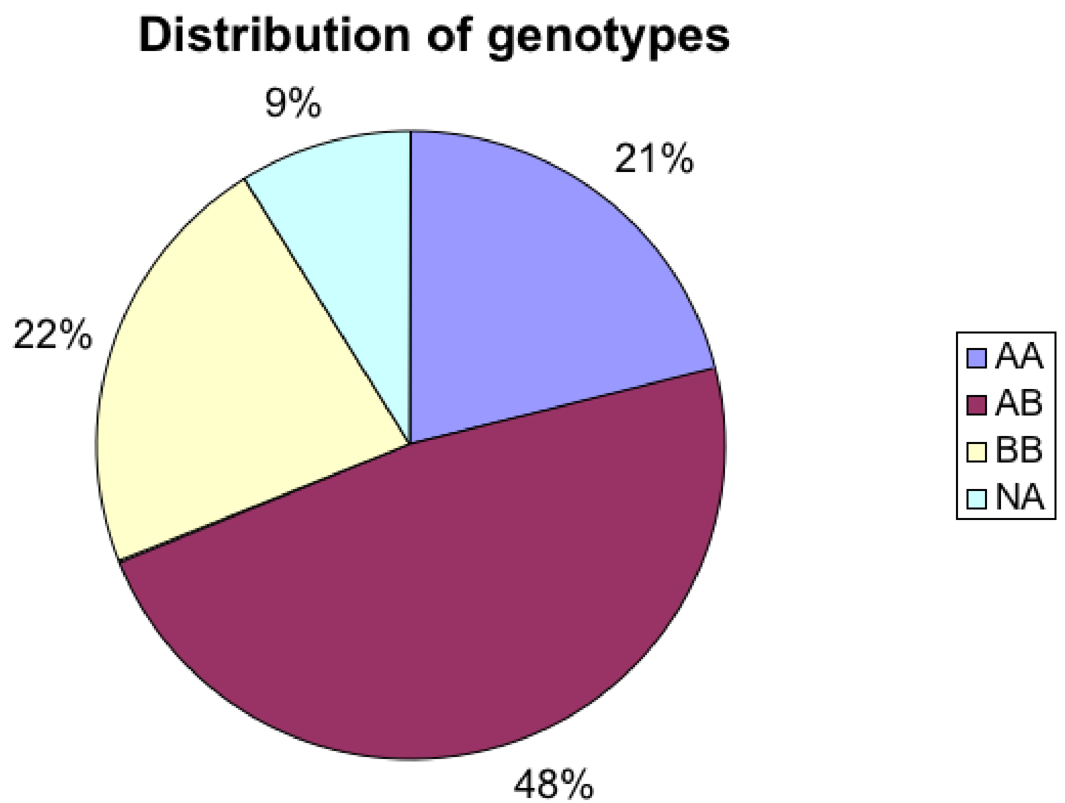
Example 4.4 (Cont.): Chart Ink



[Karl Broman]

Design Optimization

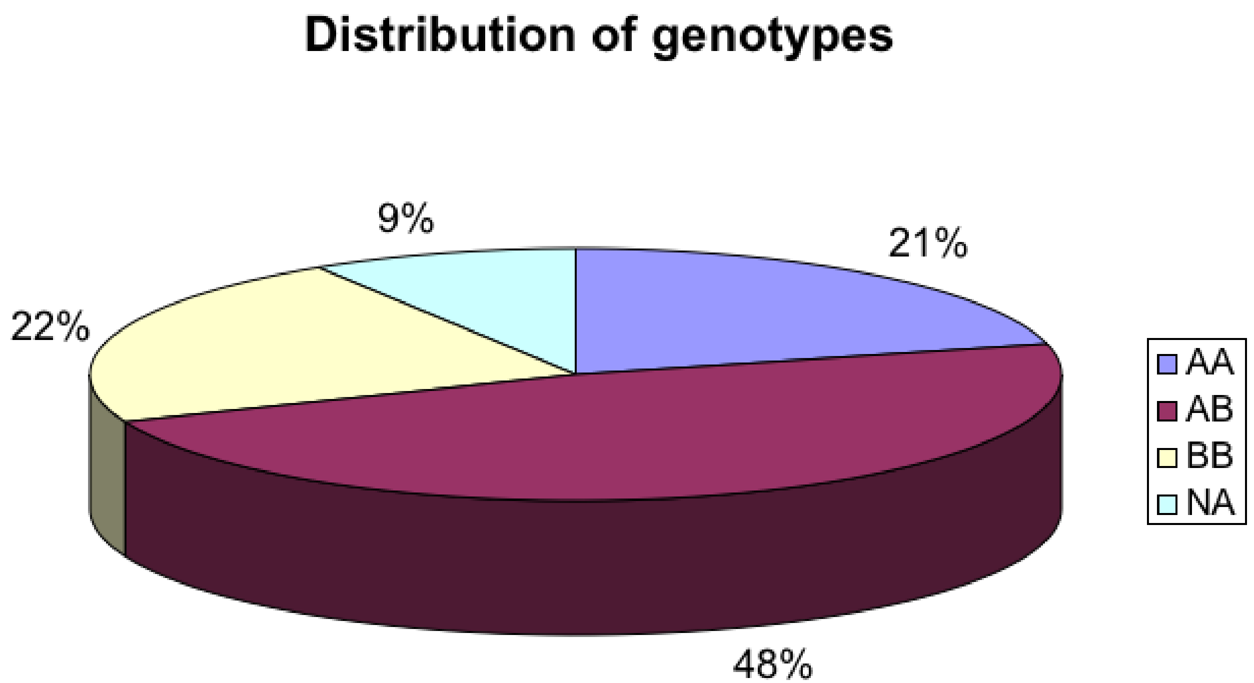
Example 4.4 (Cont.): Chart Ink



[Karl Broman]

Design Optimization

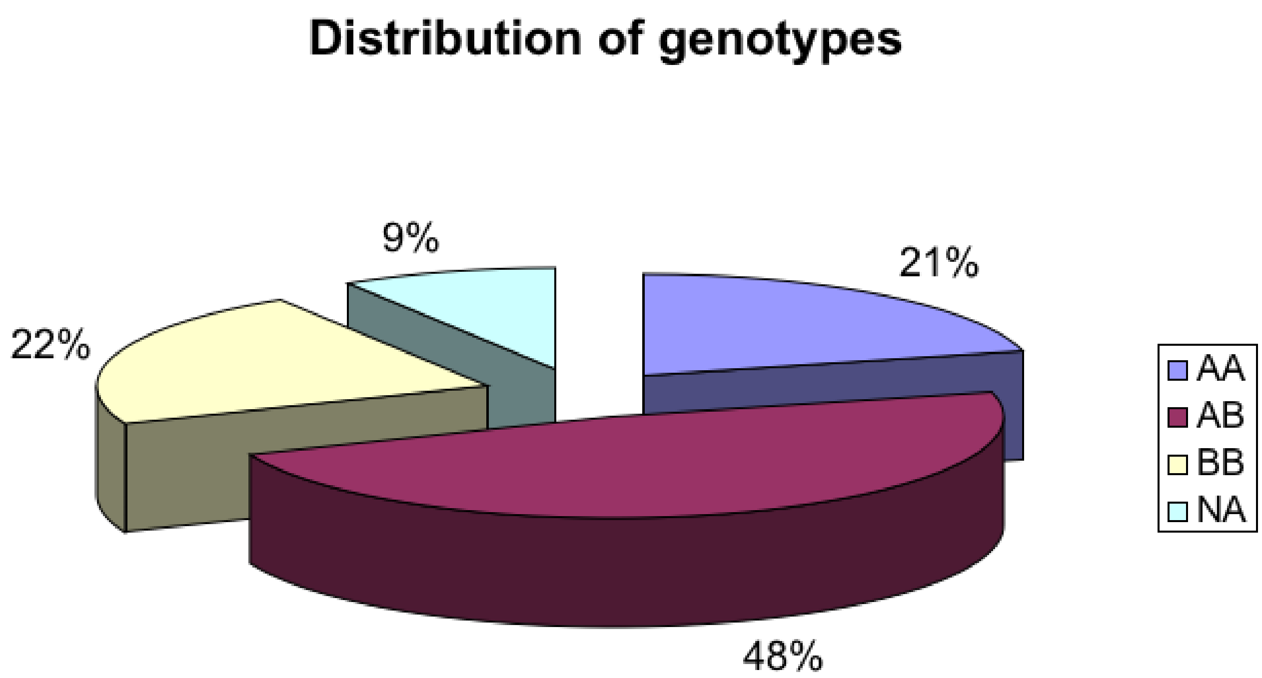
Example 4.4 (Cont.): Chart Ink



[Karl Broman]

Design Optimization

Example 4.4 (Cont.): Chart Ink



[Karl Broman]

Outline for this Session

- What is Visualization?
- Types of Visualizations
- Visual Design Principles
- Design Optimization
- **Case Study**

References

- Data Visualization: COS 323 - Computing for the Physical and Social Sciences
- http://www.cs.princeton.edu/courses/archive/spr06/cos323/notes/lecture12_vis/cos323_s06_lecture12_vis.ppt
- Preattentive processing slides from Healey: https://www.csc.ncsu.edu/faculty/healey/PP/index.html
- Kosslyn, S. M., C. F. Chabris, and D. P. Baker. "Neural network models as evidence for different types of visual representation". Cognitive Science 19 (1995): , 19, 575-579. Print.
- Lohse, G L; Biolsi, K; Walker, N and H H Rueter, A Classification of Visual Representations, CACM, Vol. 37, No. 12, pp 36-49, 1994
- Card, Stuart K., and Jock Mackinlay. "The structure of the information visualization design space." Information Visualization, 1997. Proceedings., IEEE Symposium on. IEEE, 1997.
- http://www.infovis-wiki.net/index.php/Data-Ink_Ratio