Classification and Tabulation of Data
History of Graphical Development

- First geographic maps were drawn on clay tablets.
- 17th Century: combined map skills and statistical skills to construct maps.
- Trade winds and monsoons on a world map.
- Chart patterns of disease.
- Later sophistication showed distribution of 1.3 million galaxies.
“Graphical excellence consists of the efficient communication of complex quantitative ideas.”
Presentation Topics

- Organizing Numerical Data:
  - The Ordered Array and Stem-leaf Display
- Tabulating and Graphing Numerical Data:
  - Frequency Distributions: Tables, Histograms, Polygons
  - Cumulative Distributions: Tables, the Ogive
Presentation Topics (continued)

- Tabulating and Graphing Univariate Categorical Data:
  - The Summary Table
  - Bar and Pie Charts, the Pareto Diagram

- Tabulating and Graphing Bivariate Categorical Data:
  - Contingency Tables
  - Side by Side Bar charts

- Graphical Excellence and Common Errors in Presenting Data
“At their best, graphics are instruments for reasoning about quantitative information.”
Organizing Numerical Data

Numerical Data

Ordered Array
21, 24, 24, 26, 27, 27, 30, 32, 38, 41

Stem and Leaf Display
2 144677
3 028
4 1

Frequency Distributions
Cumulative Distributions

Histograms
Ogive

Tables
Polygons
Organizing Numerical Data:

• Data in **Raw** form (as collected):
  24, 26, 24, 21, 27, 27, 30, 41, 32, 38

• Date **Ordered** from **Smallest to Largest**:
  21, 24, 24, 26, 27, 27, 30, 32, 38, 41

• Stem and Leaf display:
“Design is choice.”
Tabulating and Graphing Numerical Data

Numerical Data

Ordered Array
21, 24, 24, 26, 27, 27, 30, 32, 38, 41

Stem and Leaf Display
2 144677
3 0 28
4 1

Frequency Distributions
Cumulative Distributions

Histograms

Ogive

Tables

Polygons
Tabulating Numerical Data: Frequency Distributions

Data in ordered array:
12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

<table>
<thead>
<tr>
<th>Class</th>
<th>Frequency</th>
<th>Relative Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 but under 20</td>
<td>3</td>
<td>.15</td>
<td>15</td>
</tr>
<tr>
<td>20 but under 30</td>
<td>6</td>
<td>.30</td>
<td>30</td>
</tr>
<tr>
<td>30 but under 40</td>
<td>5</td>
<td>.25</td>
<td>25</td>
</tr>
<tr>
<td>40 but under 50</td>
<td>4</td>
<td>.20</td>
<td>20</td>
</tr>
<tr>
<td>50 but under 60</td>
<td>2</td>
<td>.10</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
<td>1</td>
<td>100</td>
</tr>
</tbody>
</table>
Graphing Numerical Data: The Histogram

Data in ordered array:
12, 13, 17, 21, 24, 24, 26, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

Class Midpoints

No Gaps Between Bars

Class Midpoints
Graphing Numerical Data: The Frequency Polygon

Data in ordered array:
12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

Class Midpoints
Tabulating Numerical Data: Cumulative Frequency

Data in ordered array:

12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

<table>
<thead>
<tr>
<th>Class</th>
<th>Cumulative Frequency</th>
<th>Cumulative % Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 but under 20</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>20 but under 30</td>
<td>9</td>
<td>45</td>
</tr>
<tr>
<td>30 but under 40</td>
<td>14</td>
<td>70</td>
</tr>
<tr>
<td>40 but under 50</td>
<td>18</td>
<td>90</td>
</tr>
<tr>
<td>50 but under 60</td>
<td>20</td>
<td>100</td>
</tr>
</tbody>
</table>
Graphing Numerical Data: The Ogive (Cumulative % Polygon)

Data in ordered array:
12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58
Tabulating and Graphing Categorical Data: Univariate Data

Categorical Data

Tabulating Data
- The Summary Table

Graphing Data
- Pie Charts
- Bar Charts
- Pareto Diagram
## Summary Table
(University Revenues)

<table>
<thead>
<tr>
<th>Revenue Category</th>
<th>Amount</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(in thousands $)</td>
<td></td>
</tr>
<tr>
<td>Patient Services</td>
<td>46.5</td>
<td>42.27</td>
</tr>
<tr>
<td>Tuition/fees</td>
<td>32</td>
<td>29.09</td>
</tr>
<tr>
<td>Appropriations</td>
<td>15.5</td>
<td>14.09</td>
</tr>
<tr>
<td>Grants/Contracts</td>
<td>16</td>
<td>14.55</td>
</tr>
<tr>
<td>Total</td>
<td>110</td>
<td>100</td>
</tr>
</tbody>
</table>

Variables are Categorical.
Graphing Categorical Data: Univariate Data

Categorical Data

Tabulating Data
The Summary Table

Graphing Data

Pie Charts

Bar Charts

Pareto Diagram
Bar Chart

Enrollment Summary

1st Prof  Grad  Unclass  Seniors  Juniors  Sophomores  Freshmen

0  500  1000  1500  2000  2500  3000
Students by Classification

- Seniors 15%
- Sophomores 14%
- Juniors 29%
- Freshmen 42%

Percentages are rounded to the nearest percent.
Pareto Diagram

Axis for bar chart shows % in each category

Axis for line graph shows cumulative %
Tabulating and Graphing Bivaria Categorical Data

• Contingency Tables
• Side by Side Charts
Tabulating Categorical Data: Bivariate Data

Contingency Table: Enrollment by College

<table>
<thead>
<tr>
<th>Enrollment Category</th>
<th>A&amp;S</th>
<th>BUS</th>
<th>NRS</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freshmen</td>
<td>46</td>
<td>55</td>
<td>27</td>
<td>128</td>
</tr>
<tr>
<td>Sophomores</td>
<td>32</td>
<td>44</td>
<td>19</td>
<td>95</td>
</tr>
<tr>
<td>Juniors</td>
<td>15</td>
<td>20</td>
<td>13</td>
<td>48</td>
</tr>
<tr>
<td>Seniors</td>
<td>16</td>
<td>28</td>
<td>7</td>
<td>51</td>
</tr>
<tr>
<td>Total</td>
<td>109</td>
<td>147</td>
<td>66</td>
<td>322</td>
</tr>
</tbody>
</table>
Graphing Categorical Data: Bivariate Data

Side by Side Chart

1st Prof
Grad
Unclass
Seniors
Juniors
Sophomores
Freshmen

0 200 400 600 800 1000 1200

A&S  AH  NS
Principles of Graphical Excellence

- Well designed presentation of data that provides:
  - Substance
  - Statistics
  - Design
- Communicates complex ideas with clarity, precision and efficiency
- Gives the largest number of ideas in the most efficient manner
- Almost always involves several dimensions
- Requires telling the truth about the data
Data-Ink Ratio

Data information
Total ink used to print the graphic
“Much of twentieth-century thinking about statistical graphics has been preoccupied with the question of how some amateurish chart might fool a naive viewer.”
Errors in Presenting Data

- Using ‘chart junk’
- No relative basis
- In comparing data
- Batches
- Compressing the
- Vertical axis
- No zero point on the
- Vertical axis
‘Chart Junk’

Bad Presentation

Minimum Wage

- 1960: $1.00
- 1970: $1.60
- 1980: $3.10
- 1990: $3.80

Good Presentation

Minimum Wage

- 1960: $1.00
- 1970: $1.60
- 1980: $3.10
- 1990: $3.80
Lie Factor

Size of effect shown in graphic

Size of effect in data
No Relative Basis

Bad Presentation

A’s received by students.

FR = Freshmen,  SO = Sophomore,  JR = Junior,  SR = Senior
Compressing Vertical Axis

Bad Presentation

Good Presentation

Quarterly Income

$0\quad Q1 \quad Q2 \quad Q3 \quad Q4$

$200$

$0$

$100$

$25$

$50$

$0$

$100$

$200$

Quarterly Income

$0\quad Q1 \quad Q2 \quad Q3 \quad Q4$

$25$

$50$

$0$
No Zero Point on Vertical Axis

Bad Presentation
Monthly Expenses

Good Presentation
Monthly Expenses

Graphing the first six months of sales.
No Zero Point on Vertical Axis

Bad Presentation

Monthly Expenses

Good Presentation

Monthly Expenses

Graphing the first six months of sales.
Main defense of the lying graphic....

“Well, at least it was approximately correct, we were just trying to show the general direction of change.”
Presentation Summary

- Organized Numerical Data:
  - The Ordered Array and Stem-leaf Display
- Tabulated and Graphed Numerical Data
  - Frequency Distributions: Tables, Histograms, Polygons
  - Cumulative Distributions: Tables, the Ogive
Presentation Summary
(continued)

• Tabulated and Graphed Univariate Categorical Data:
  • The Summary Table
  • Bar and Pie Charts, the Pareto diagram

• Tabulated and Graphed Bivariate Categorical Data:
  • Contingency Tables
  • Side by Side charts

• Discussed Graphical Excellence and Common Errors in Presenting Data
“There remain, however, many other consideration in the design of statistical graphics – not only of efficiency, but also of complexity, structure, density, and even beauty.”