



*Anna T. Waggener, Ph.D.*

**Institutional Assessment**

**United States Army War College**

**Principles of Graphical Excellence**

*Best Paper: ALAIR April 5–6, 2001*

*AIR: June 2-5, 2002, Toronto*

*Focus-IR, February 21, 2003*

---

# The Visual Display of Quantitative Information

Leading authority: *Edward R. Tufte*

# History of Graphical Development

---

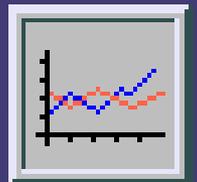
- First geographic maps were drawn on clay tablets.
- 17<sup>th</sup> Century: combined map skills and statistical skills to construct maps.
- Trade winds and monsoons on a world map.
- Chart patterns of disease.
- Later sophistication showed distribution of 1.3 million galaxies.

---

*“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**

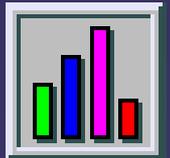


# Presentation Topics

*(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

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

### 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:

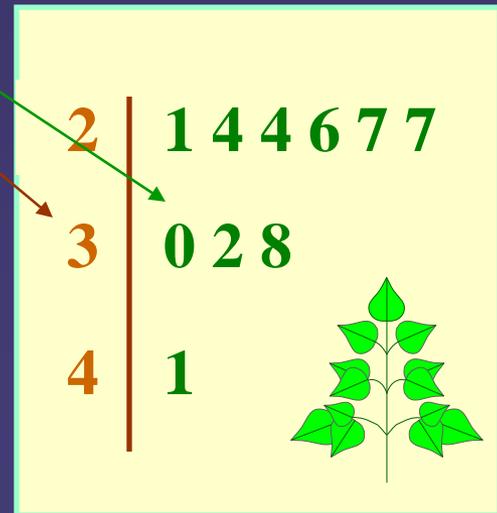
- **Data in *Raw* form (as collected):**

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**

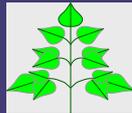
41, 24, 32, 26, 27, 27, 30, 24, 38, 21

**Ordered Array**

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

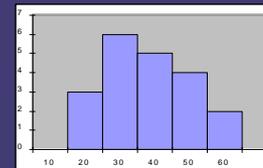
**Stem and Leaf Display**

2	144677
3	028
4	1



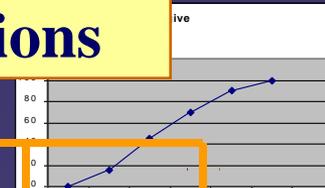
**Frequency Distributions  
Cumulative Distributions**

**Histograms**



**Tables**

**Ogive**



**Polygons**

# Tabulating Numerical Data: Frequency Distributions

*(continued)*

Data in ordered array:

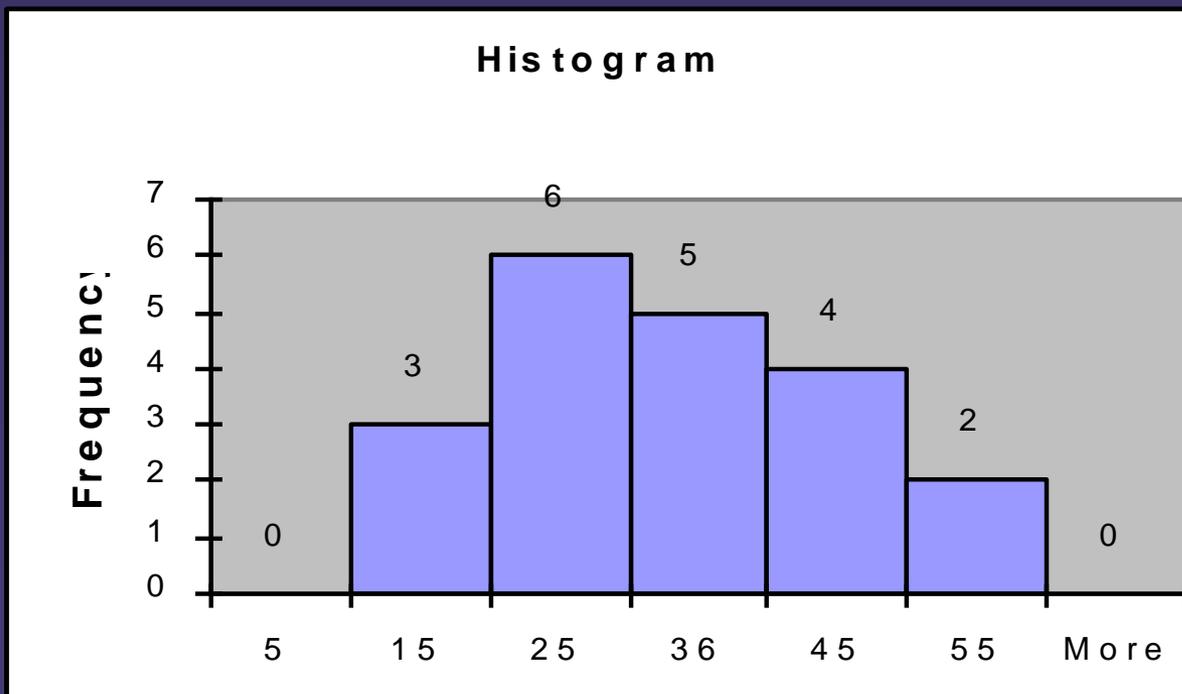
12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

<b>Class</b>	<b>Frequency</b>	<b>Relative Frequency</b>	<b>Percentage</b>
10 but under 20	3	.15	15
20 but under 30	6	.30	30
30 but under 40	5	.25	25
40 but under 50	4	.20	20
50 but under 60	2	.10	10
<b>Total</b>	<b>20</b>	<b>1</b>	<b>100</b>

# Graphing Numerical Data: The Histogram

Data in ordered array:

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



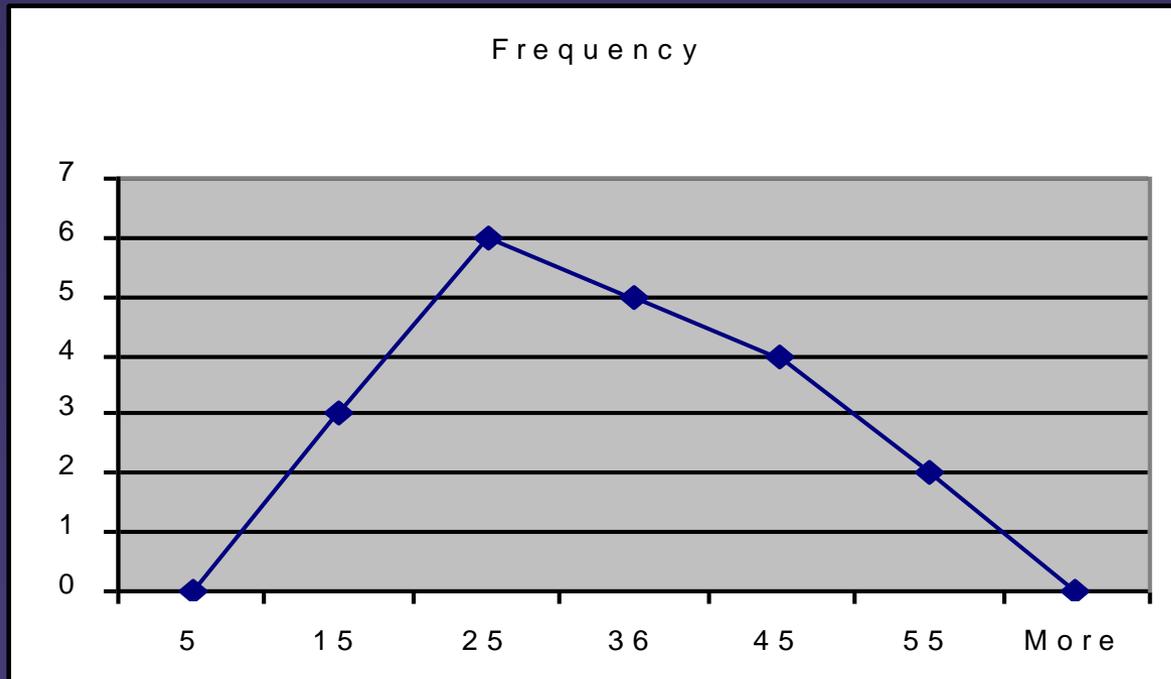
**No Gaps  
Between  
Bars**

**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:

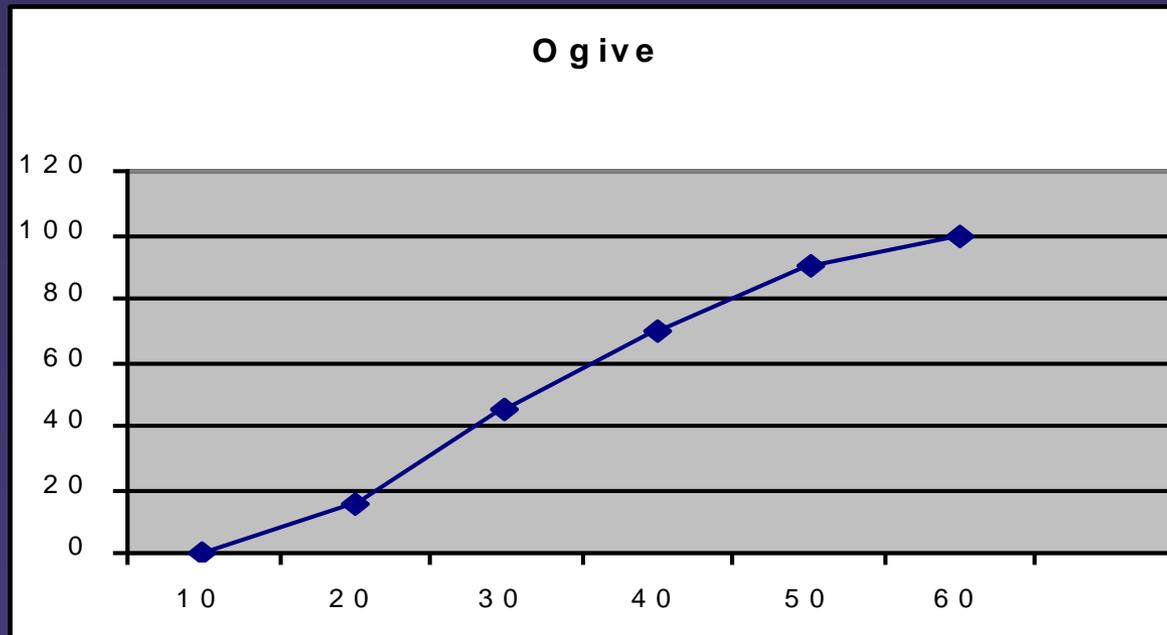
12, 13, 17, 21, 24, 24, 26, 27, 27, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58

<b>Class</b>	<b>Cumulative Frequency</b>	<b>Cumulative % Frequency</b>
<b>10 but under 20</b>	<b>3</b>	<b>15</b>
<b>20 but under 30</b>	<b>9</b>	<b>45</b>
<b>30 but under 40</b>	<b>14</b>	<b>70</b>
<b>40 but under 50</b>	<b>18</b>	<b>90</b>
<b>50 but under 60</b>	<b>20</b>	<b>100</b>

# 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**

```
graph TD; A[Categorical Data] --> B[Tabulating Data]; A --> C[Graphing Data]; B --- D["The Summary Table"]; C --> E[Bar Charts]; C --> F[Pie Charts]; C --> G[Pareto Diagram];
```

**Tabulating Data**

The **Summary Table**

**Graphing Data**

**Pie Charts**

**Bar Charts**

**Pareto Diagram**

# Summary Table

## (University Revenues)

Revenue Category	Amount (in thousands \$)	Percentage
Patient Services	46.5	42.27
Tuition/fees	32	29.09
Appropriations	15.5	14.09
Grants/Contracts	16	14.55
<b>Total</b>	<b>110</b>	<b>100</b>

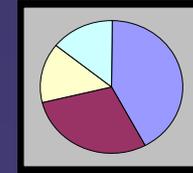
Variables are Categorical.

# Graphing Categorical Data: Univariate Data

## Categorical Data

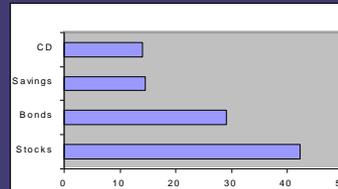
**Tabulating Data**  
**The Summary Table**

## Graphing Data

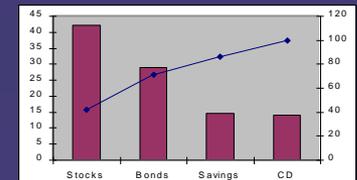


**Pie Charts**

**Bar Charts**

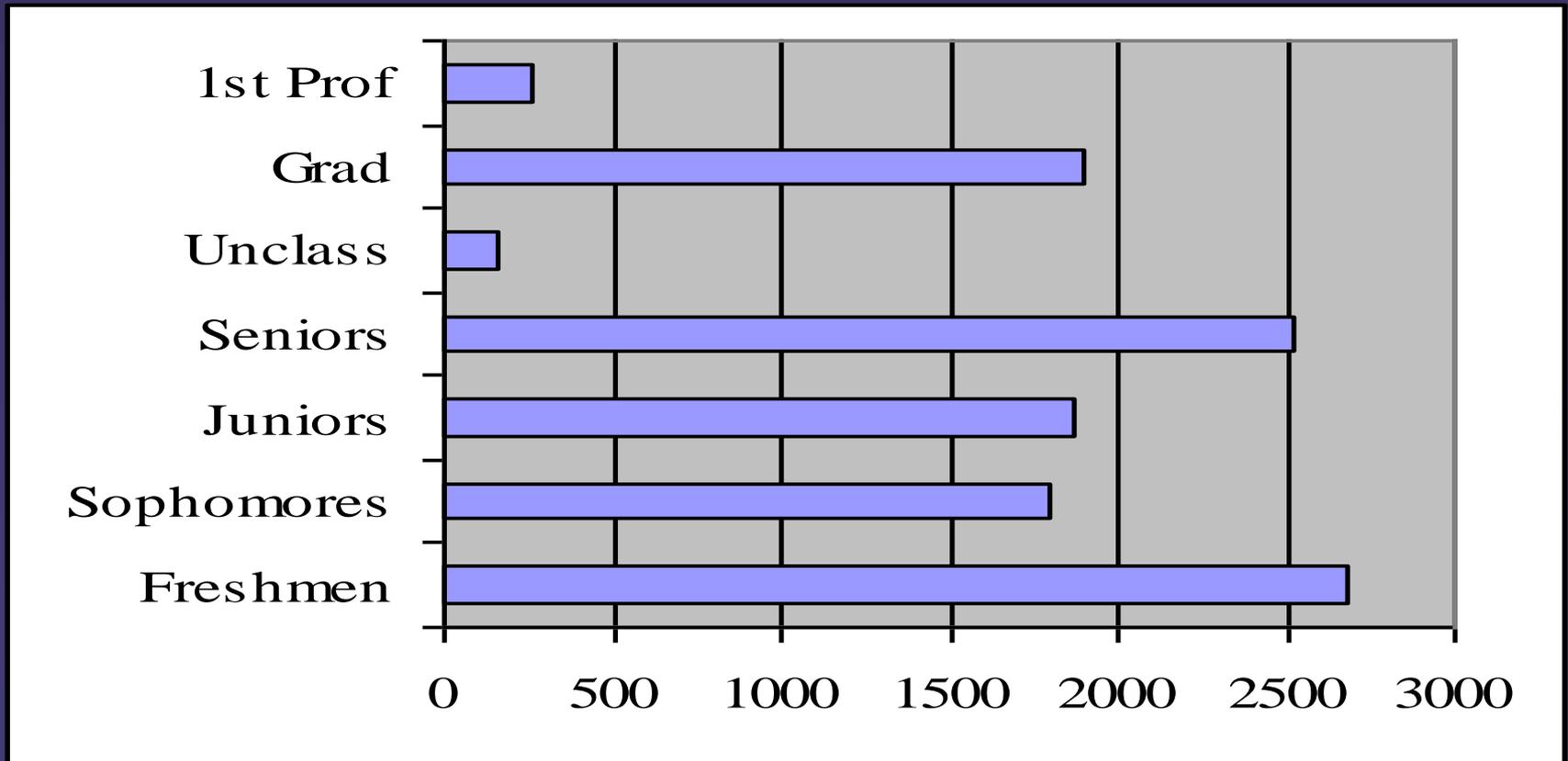


**Pareto Diagram**



# Bar Chart

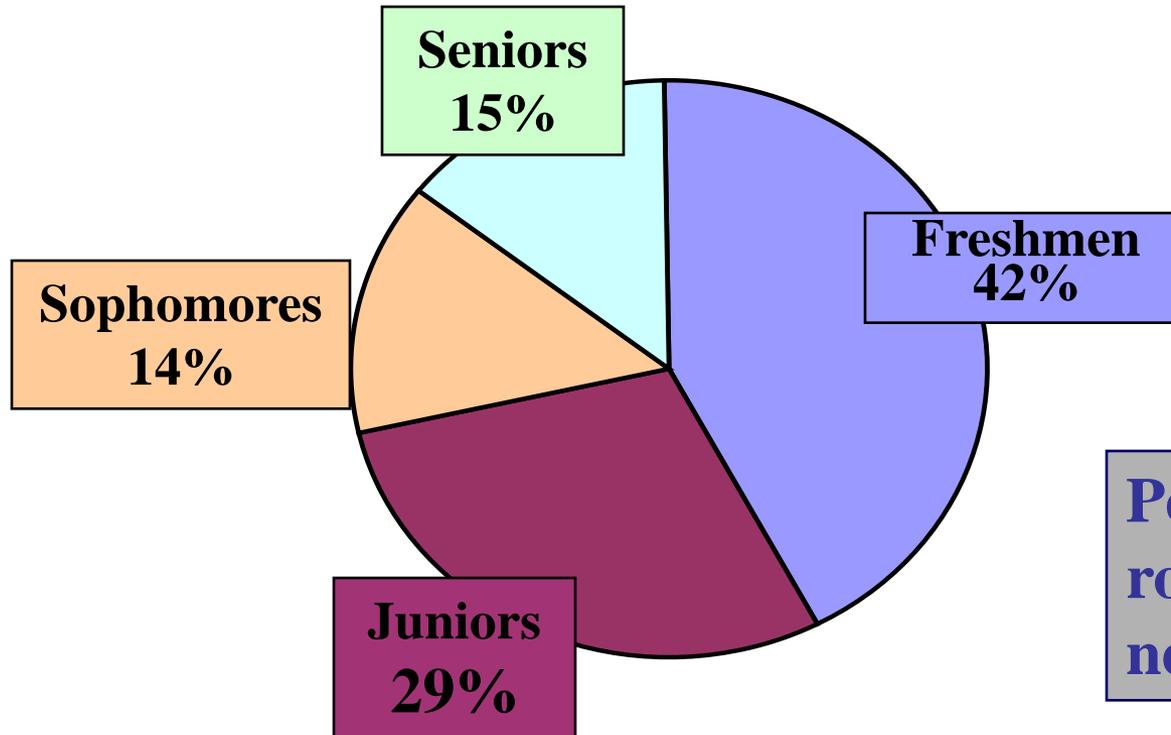
## Enrollment Summary



# Pie Chart

(for a factbook)

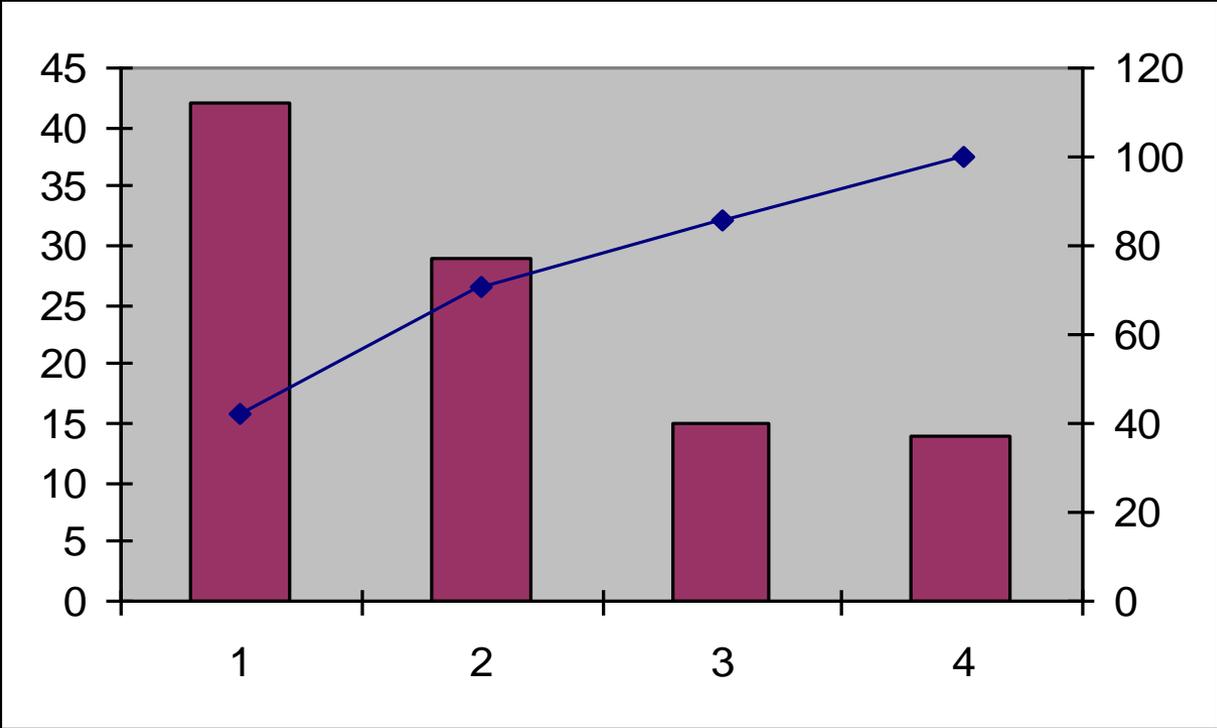
**Students by Classification**



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 Bivariate Categorical Data

---

- **Contingency Tables**
- **Side by Side Charts**

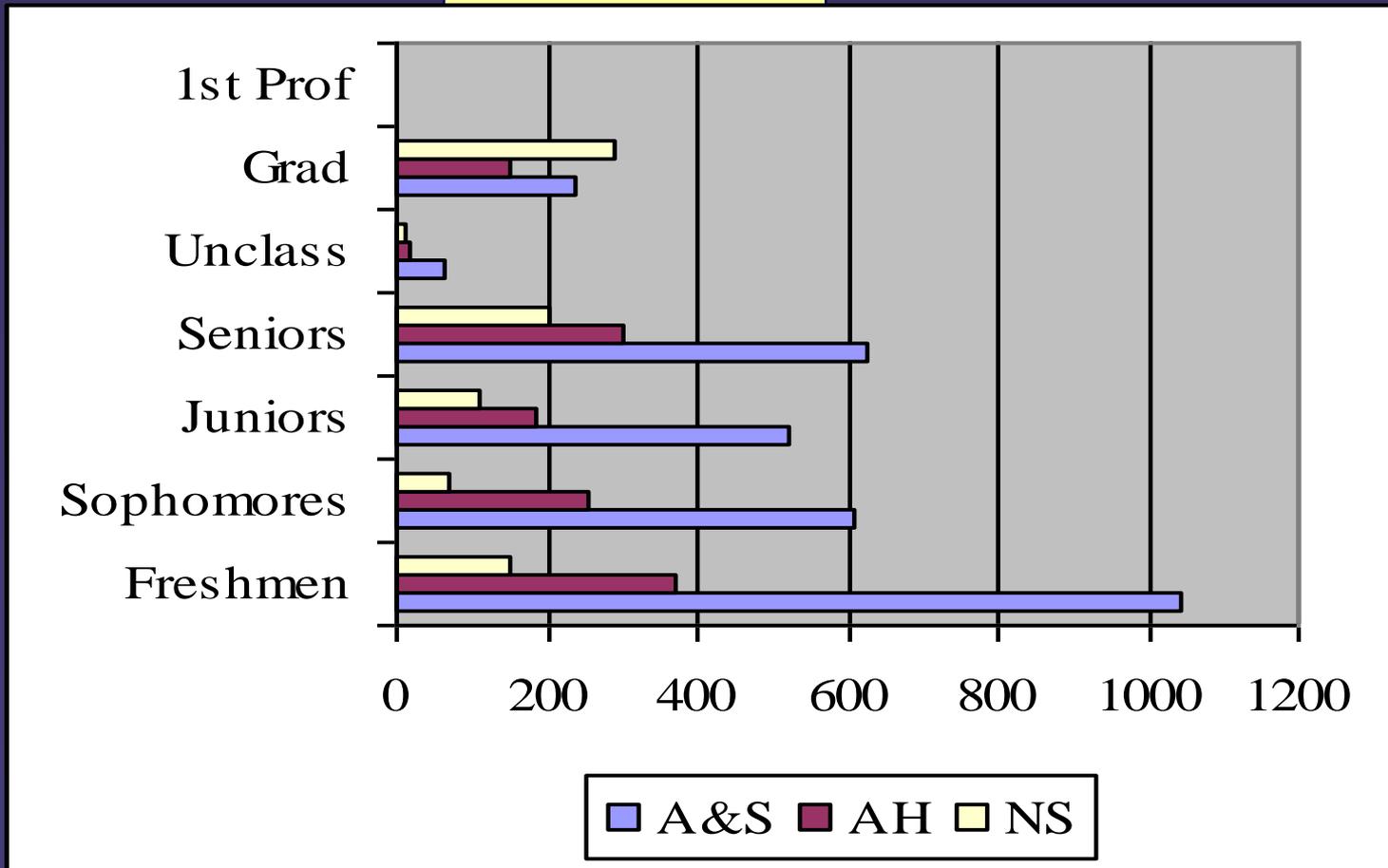
# Tabulating Categorical Data: Bivariate Data

## Contingency Table: Enrollment by College

<b>Enrollment Category</b>	<b>A&amp;S</b>	<b>BUS</b>	<b>NRS</b>	<b>Total</b>
Freshmen	46	55	27	<b>128</b>
Sophomores	32	44	19	<b>95</b>
Juniors	15	20	13	<b>48</b>
Seniors	16	28	7	<b>51</b>
<b>Total</b>	<b>109</b>	<b>147</b>	<b>66</b>	<b>322</b>

# Graphing Categorical Data: Bivariate Data

## Side by Side Chart



# 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

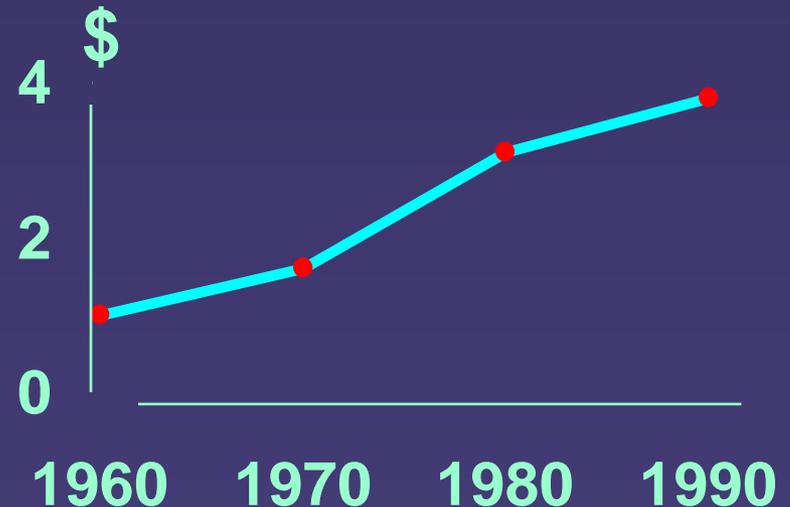


1990: \$3.80



## Good Presentation

### Minimum Wage



# Lie Factor

---

Size of effect shown in graphic  
Size of effect in data

# No Relative Basis



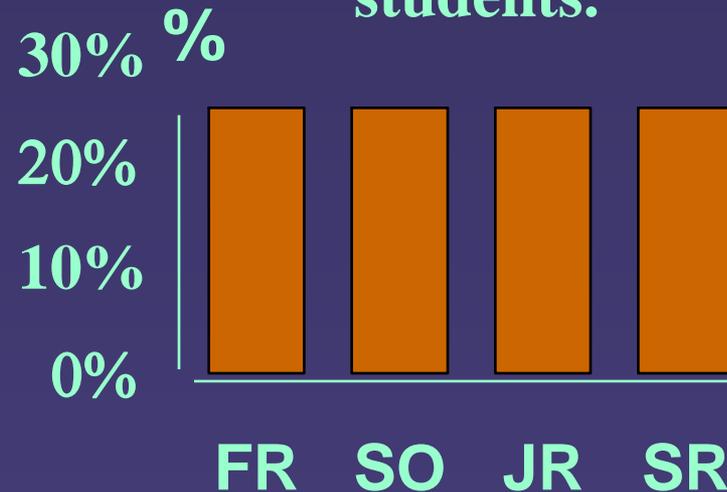
## Bad Presentation

A's received by students.



## Good Presentation

A's received by students.



FR = Freshmen, SO = Sophomore, JR = Junior, SR = Senior

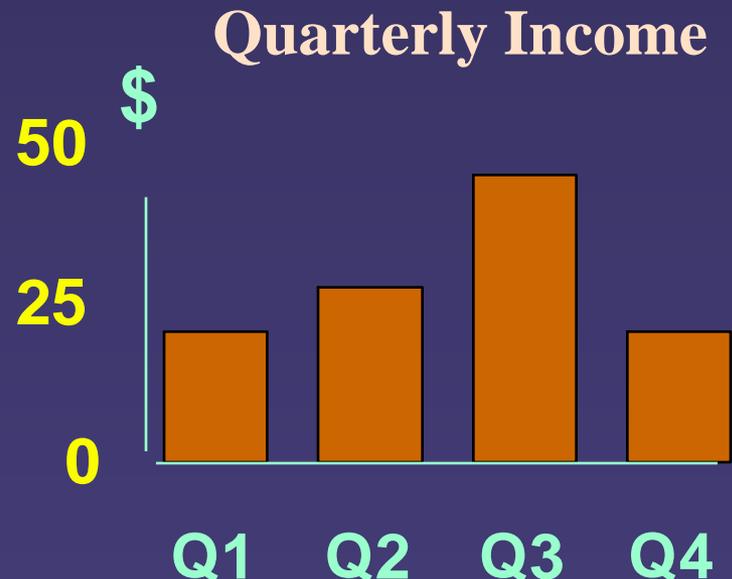
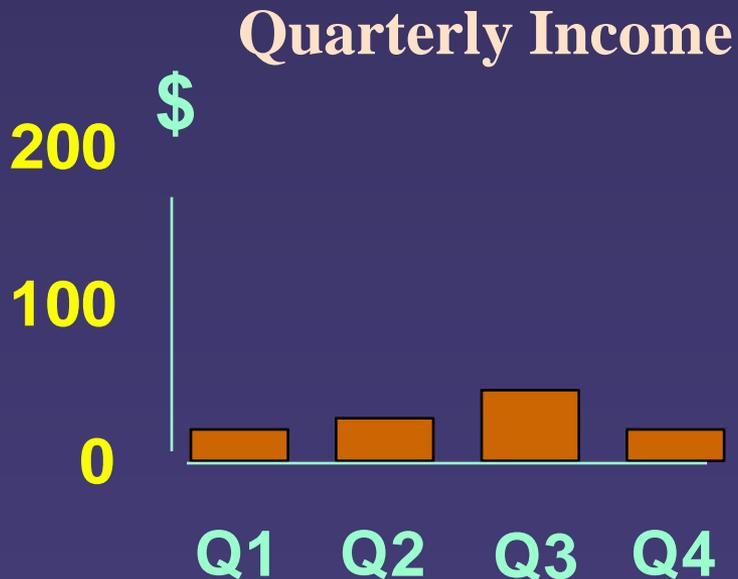
# Compressing Vertical Axis



**Bad Presentation**



**Good Presentation**



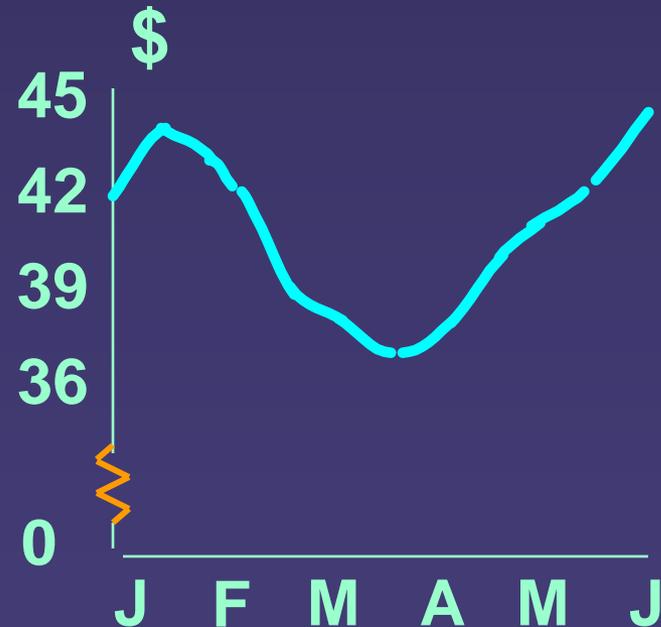
# No Zero Point on Vertical Axis



**Bad Presentation**  
Monthly Expenses



✓ **Good Presentation**  
Monthly Expenses



Graphing the first six months of sales.

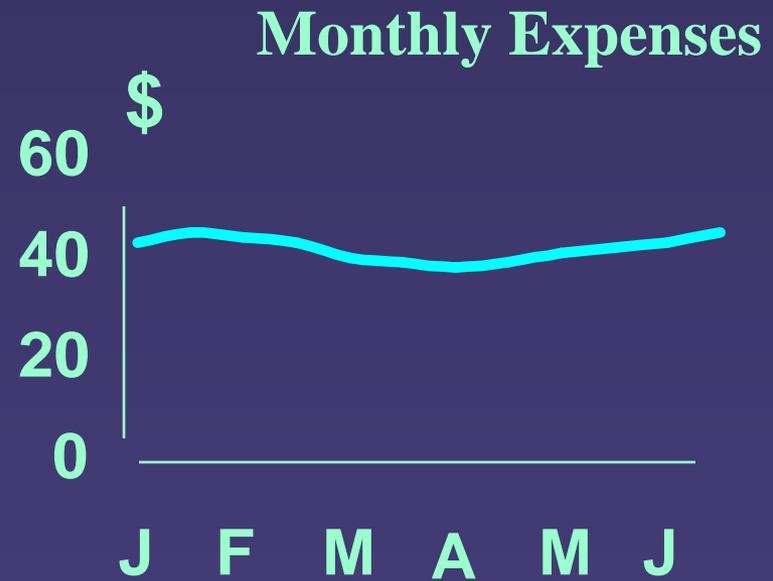
# No Zero Point on Vertical Axis



**Bad Presentation**



**Good Presentation**



Graphing the first six months of sales.

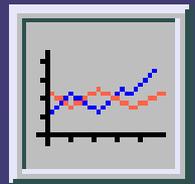
# Main defense of the lying graphic....

---

*“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**

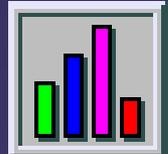


# Presentation Summary

*(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 considerations in the design of statistical graphics – not only of efficiency, but also of complexity, structure, density, and even beauty.”*

---

*“Without data, it is anyone’s  
opinion.”*

*Author unknown*