

The Art of Data Presentation

Raj Jain

Washington University in Saint Louis

Saint Louis, MO 63130

Jain@cse.wustl.edu

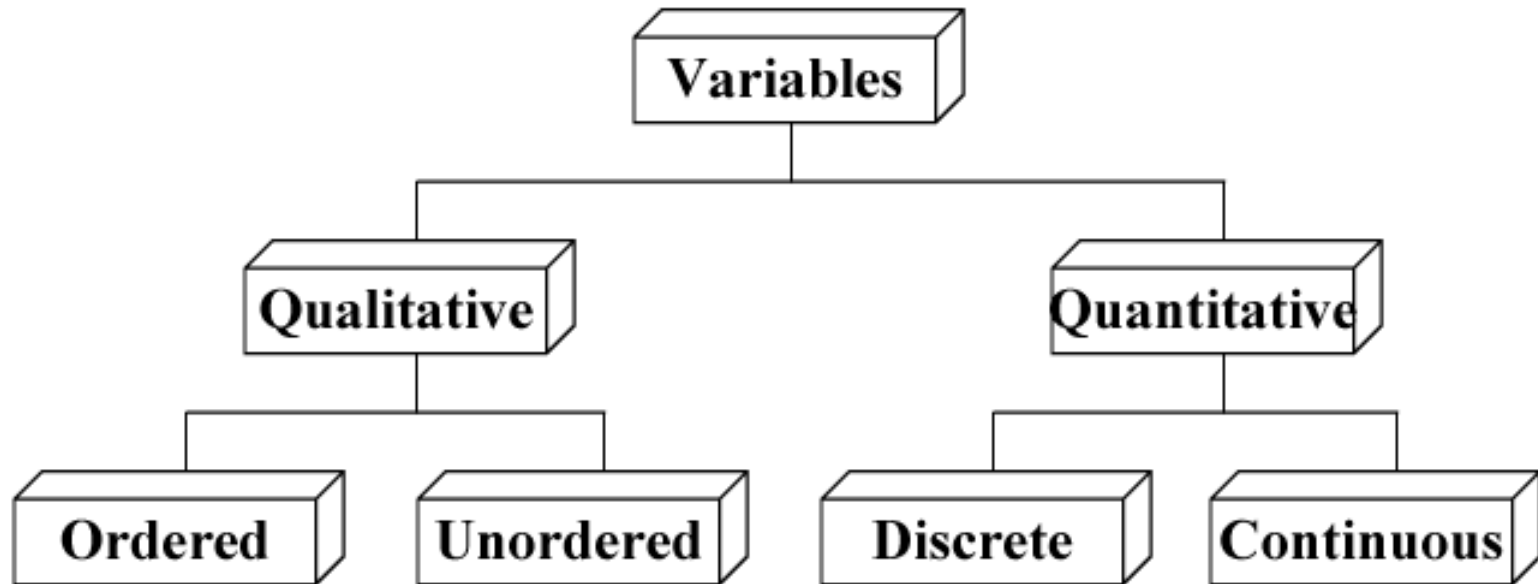
These slides are available on-line at:

<http://www.cse.wustl.edu/~jain/cse567-17/>



- ❑ Types of Variables
- ❑ Guidelines for Preparing Good Charts
- ❑ Common Mistakes in Preparing Charts
- ❑ Pictorial Games
- ❑ Decision Maker's Games

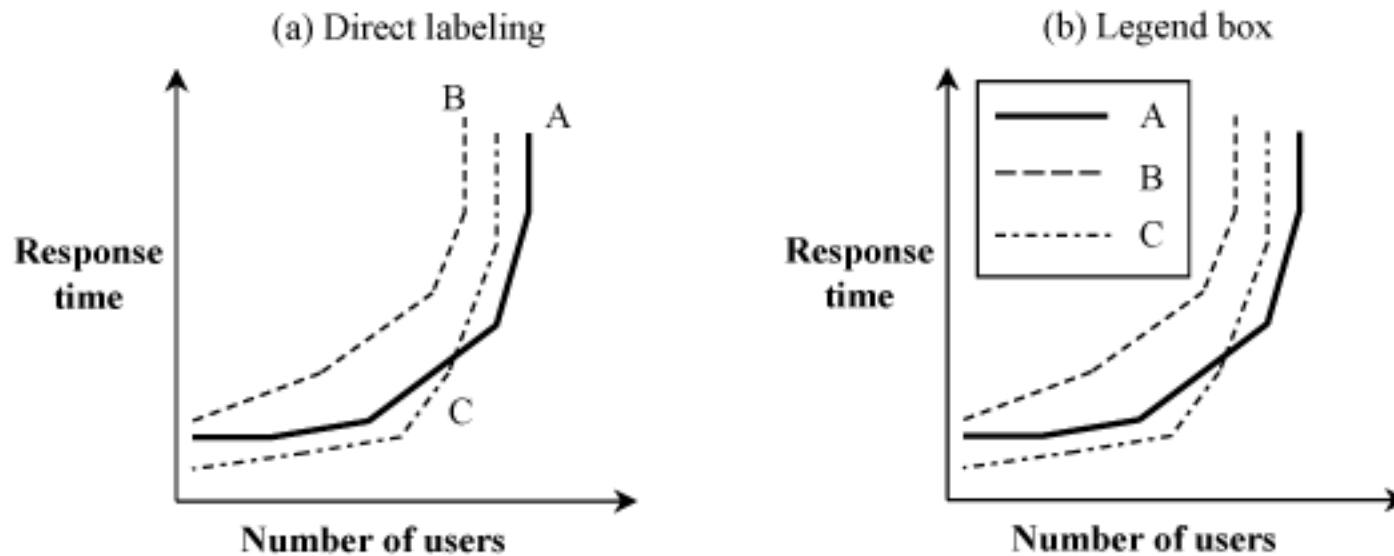
Types of Variables



- ❑ Type of computer: Super computer, minicomputer, microcomputer
- ❑ Type of Workload: Scientific, engineering, educational
- ❑ Number of processors
- ❑ Response time of system

Guidelines for Preparing Good Charts

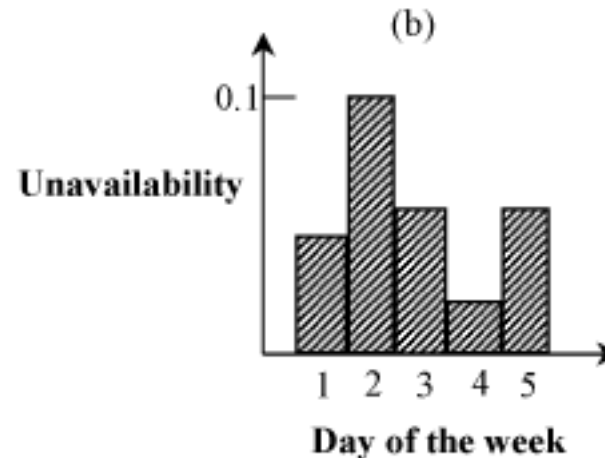
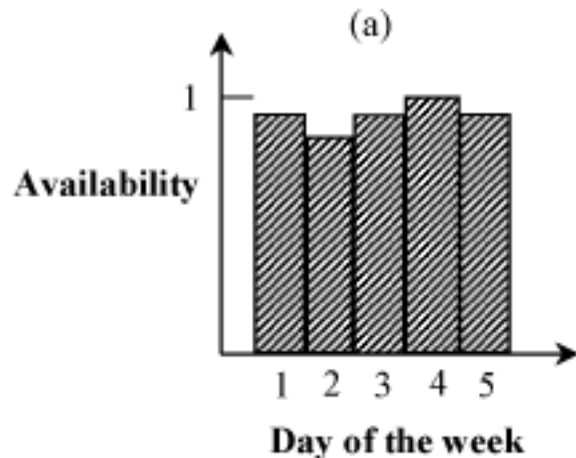
- Require minimum effort from the reader
Direct labeling vs. legend box



- Maximize Information: Words in place of symbols
Clearly label the axes

Guidelines (cont)

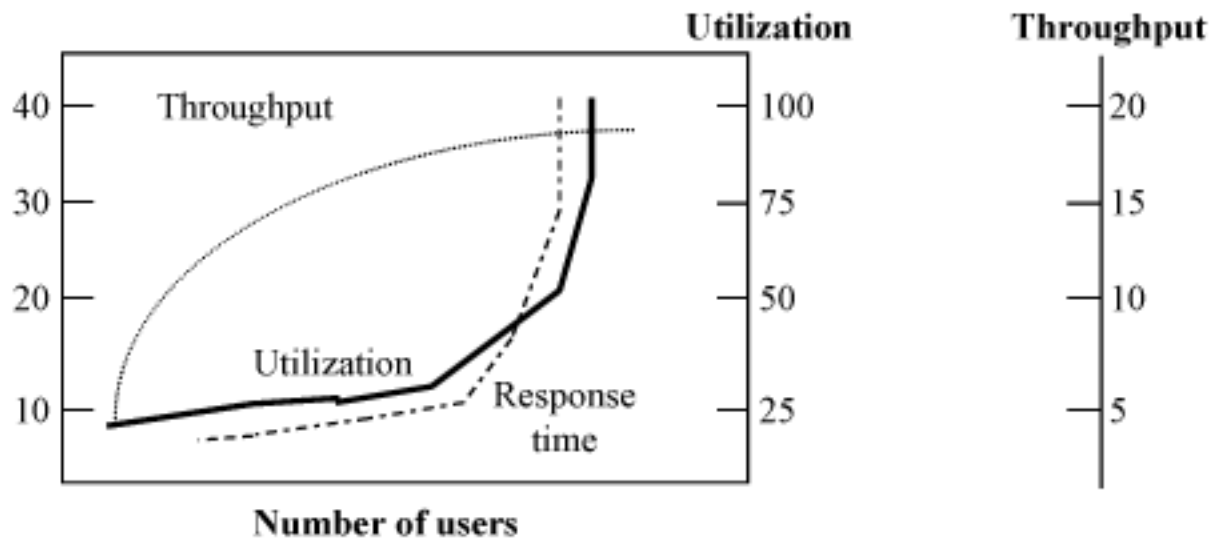
- ❑ Minimize Ink: No grid lines, more details



- ❑ Use Commonly accepted practices: origin at (0,0)
Independent variable (cause) along x axis, linear scales, increasing scales, equal divisions
- ❑ Avoid ambiguity: Show coordinate axes, scale divisions, origin. Identify individual curves and bars.
- ❑ See checklist in Box 10.1

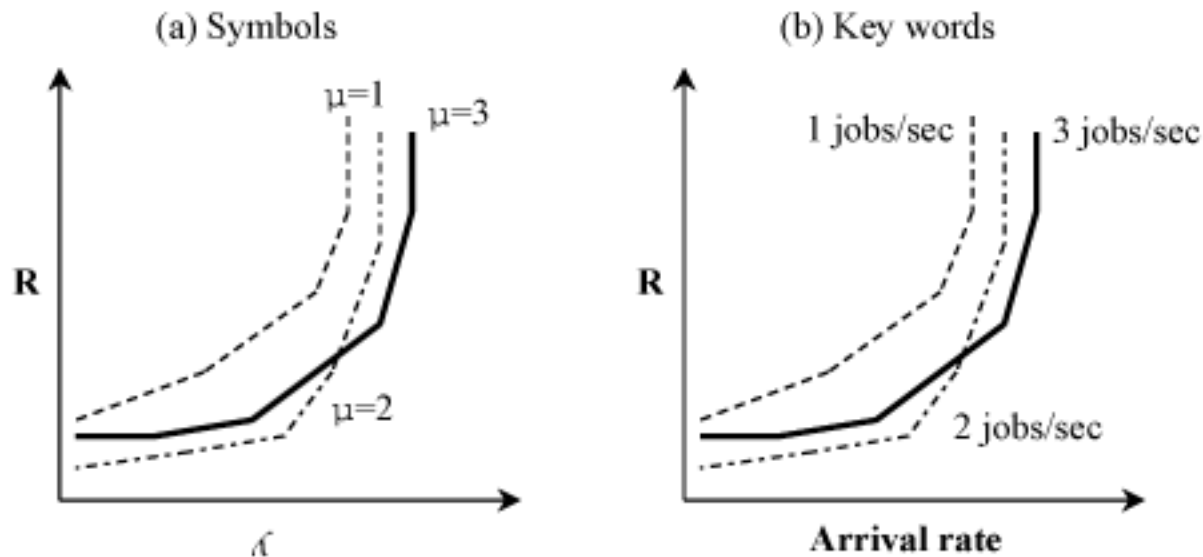
Common Mistakes in Preparing Charts

- ❑ Presenting too many alternatives on a single chart
Max 5 to 7 messages \Rightarrow Max 6 curves in a line charts, no more than 10 bars in a bar chart, max 8 components in a pie chart
- ❑ Presenting many y variables on a single chart



Common Mistakes in Charts (Cont)

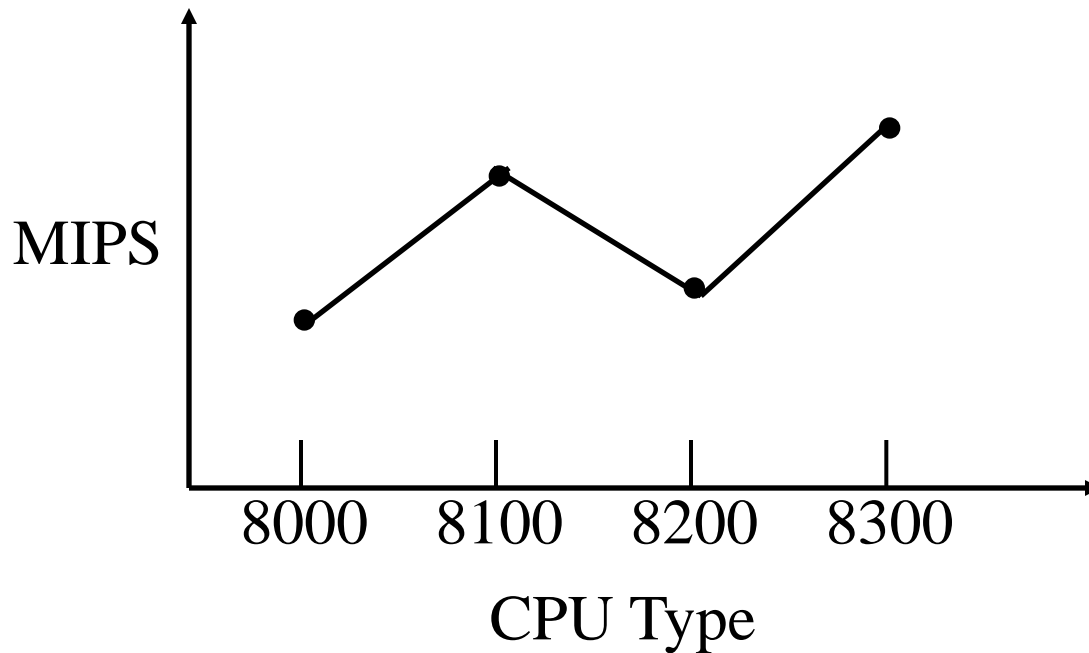
- ❑ Using symbols in place of text



- ❑ Placing extraneous information on the chart: grid lines, granularity of the grid lines
- ❑ Selecting scale ranges improperly: automatic selection by programs may not be appropriate

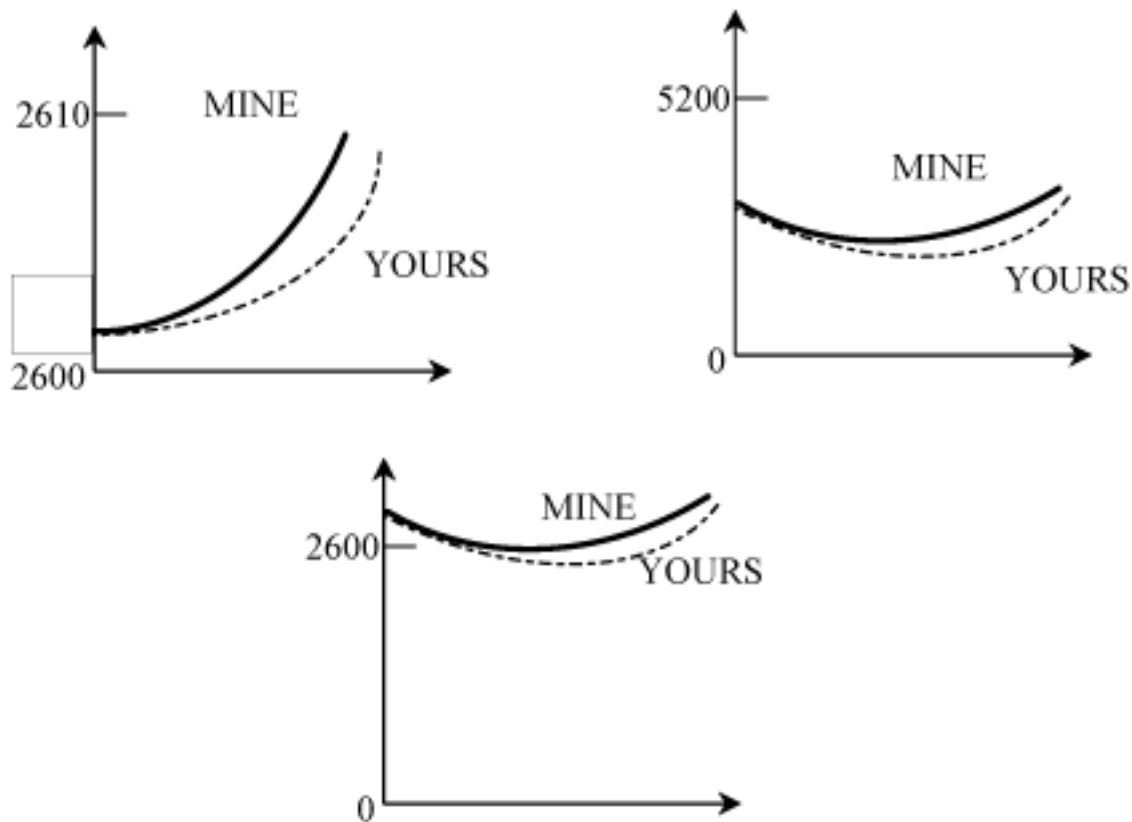
Common Mistakes in Charts (Cont)

- ❑ Using a line chart in place of column chart:
Line \Rightarrow Continuity



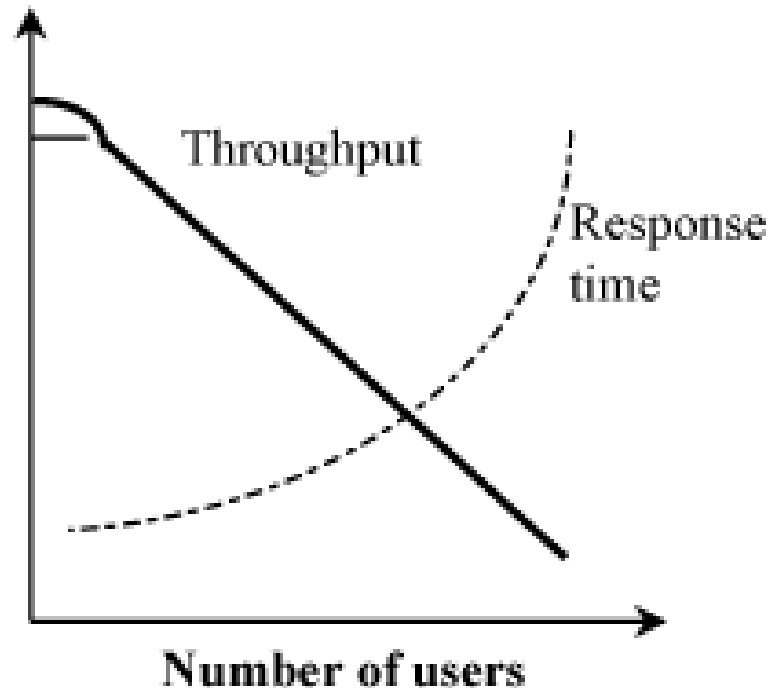
Pictorial Games

- Using non-zero origins to emphasize the difference
Three quarter high-rule \Rightarrow height/width $> 3/4$



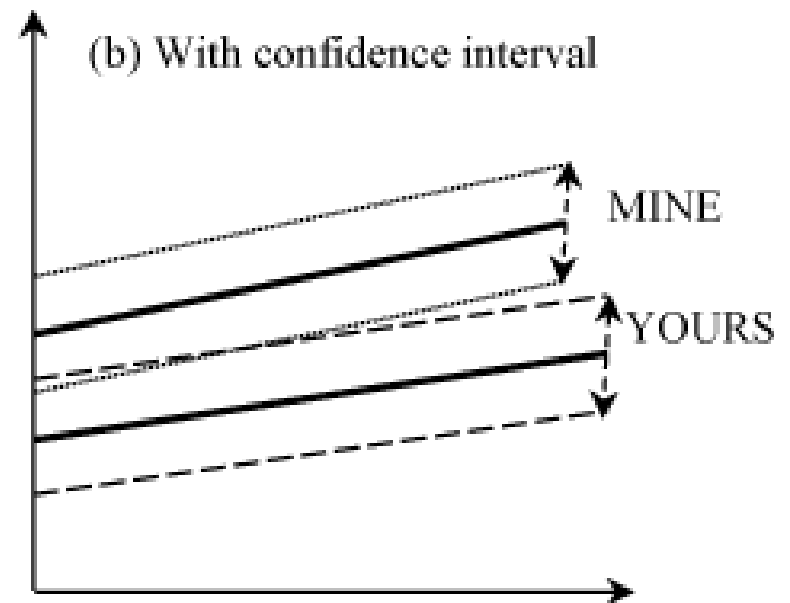
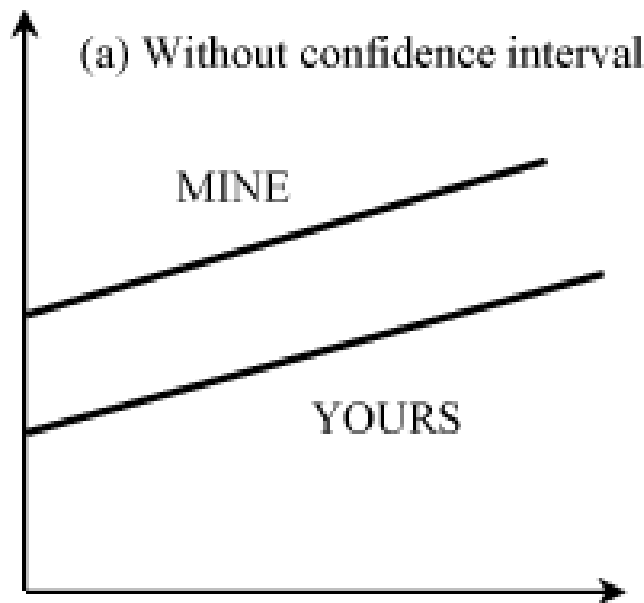
Pictorial Games (Cont)

- Using double-whammy graph for dramatization
Using related metrics



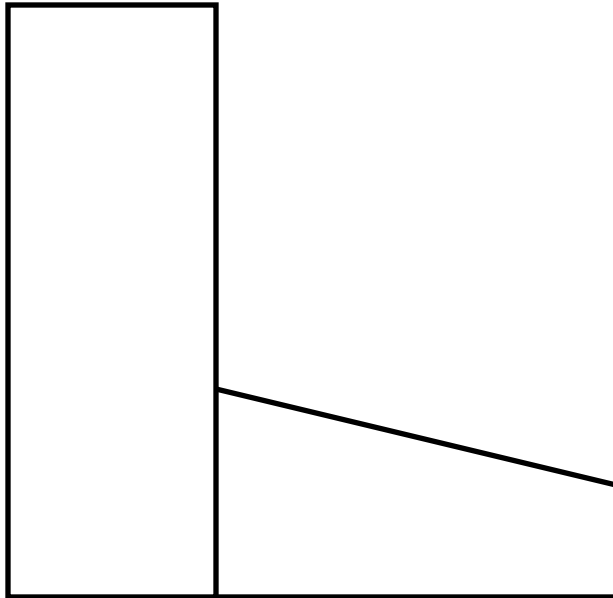
Pictorial Games (Cont)

- Plotting random quantities without showing confidence intervals

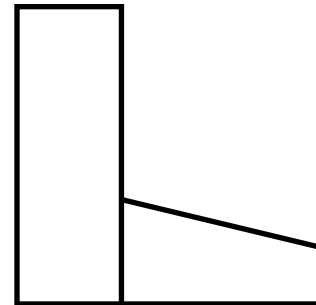


Pictorial Games (Cont)

- Pictograms scaled by height



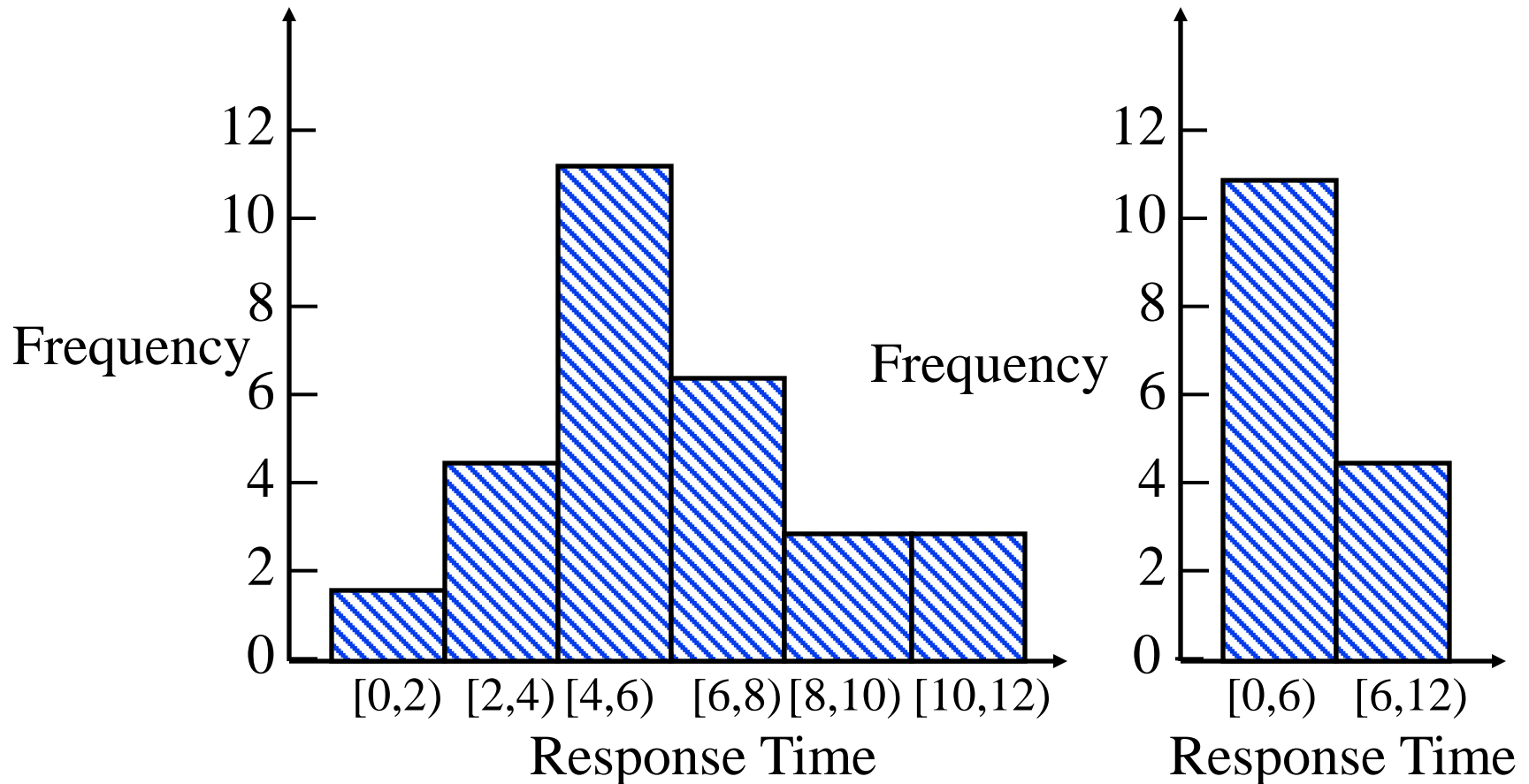
Mine
Performance = 2



Yours
Performance = 1

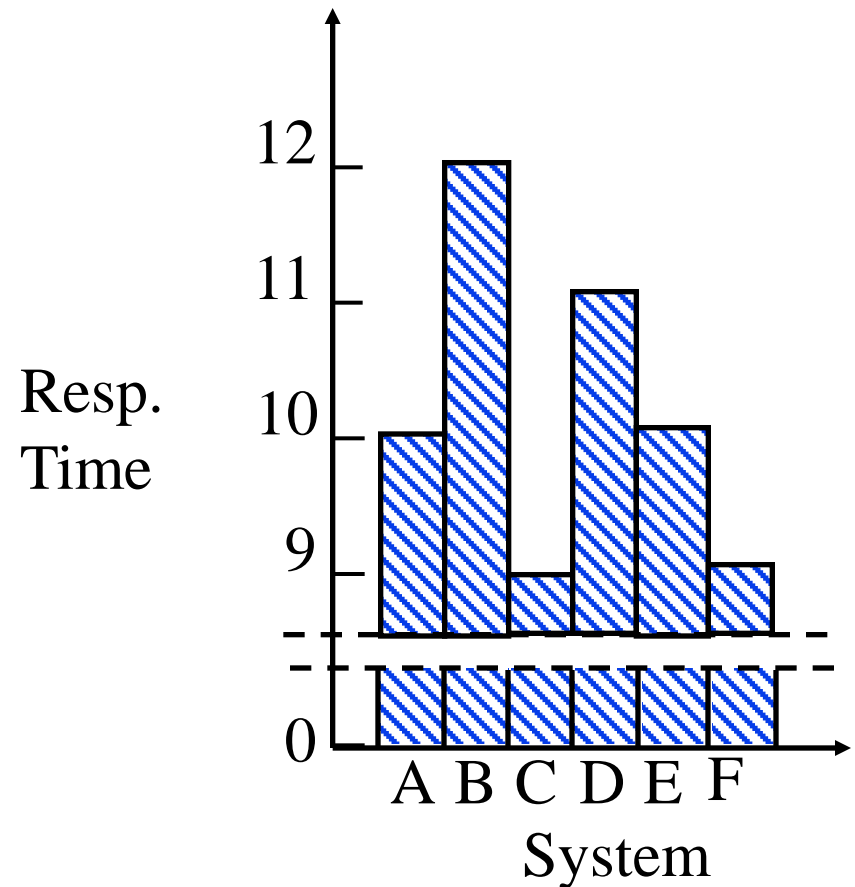
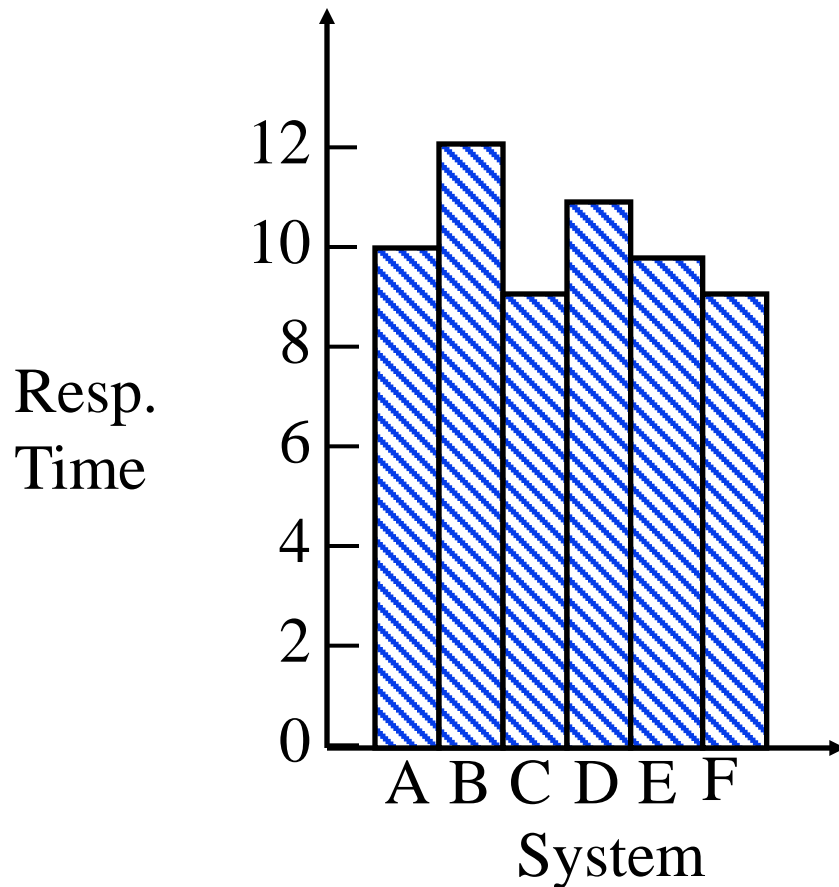
Pictorial Games (Cont)

- Using inappropriate cell size in histograms

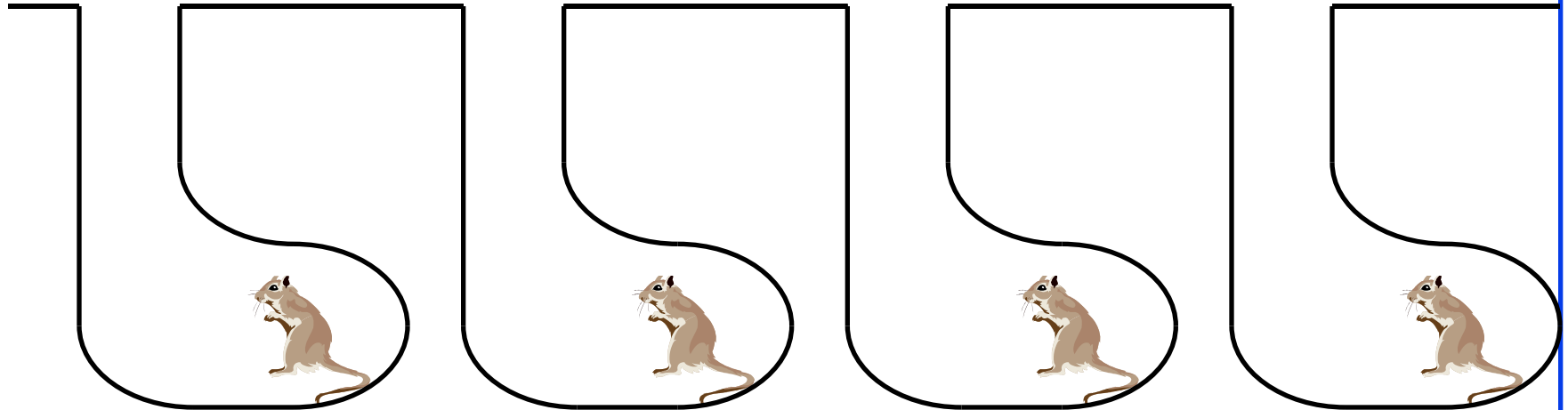


Pictorial Games (Cont)

- Using broken scales in column charts



Performance Analysis Rat Holes



Workload

Metrics

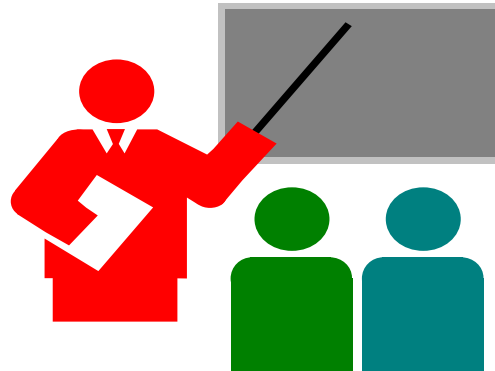
Configuration Details

Reasons for not Accepting an Analysis

- ❑ This needs more analysis.
- ❑ You need a better understanding of the workload.
- ❑ It improves performance only for long IOs/packets/jobs/files, and most of the IOs/packets/jobs/files are short.
- ❑ It improves performance only for short IOs/packets/jobs/files, but who cares for the performance of short IOs/packets/jobs/files, its the long ones that impact the system.
- ❑ It needs too much memory/CPU/bandwidth and memory/CPU/bandwidth isn't free.
- ❑ It only saves us memory/CPU/bandwidth and memory/CPU/bandwidth is cheap.

See Box 10.2 on page 162 of the book for a complete list

Summary



1. Qualitative/quantitative, ordered/unordered, discrete/continuous variables
2. Good charts should require minimum effort from the reader and provide maximum information with minimum ink
3. Use no more than 5-6 curves, select ranges properly, Three-quarter high rule
4. Workload, metrics, configuration, and details can always be challenged. Should be carefully selected.

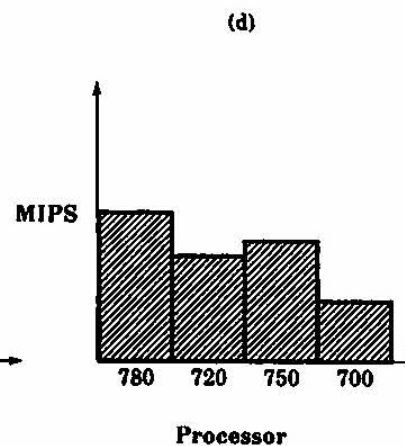
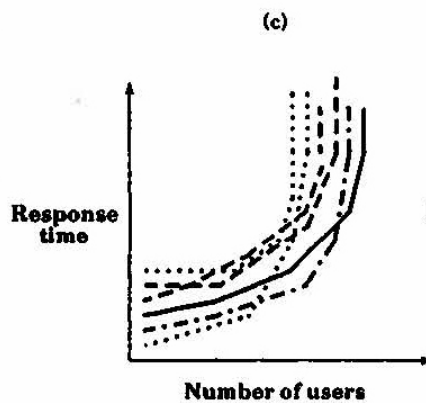
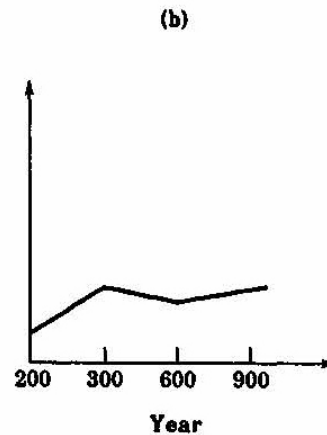
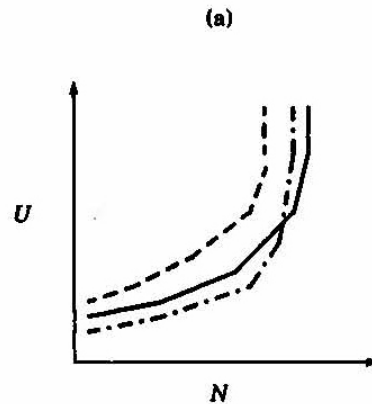
Exercise 10.1

What type of chart (line or bar) would you use to plot:

- a. CPU usage for 12 months of the year
- b. CPU usage as a function of time in months
- c. Number of I/O's to three disk drives: A, B, and C
- d. Number of I/O's as a function of number of disk drives in a system

Homework 10: Exercise 10.2

- List the problems with the following charts



Scan This to Download These Slides



Raj Jain

<http://rajjain.com>

Related Modules



CSE567M: Computer Systems Analysis (Spring 2013),

https://www.youtube.com/playlist?list=PLjGG94etKypJEKjNAa1n_1X0bWWNyZcof

CSE473S: Introduction to Computer Networks (Fall 2011),

https://www.youtube.com/playlist?list=PLjGG94etKypJWOSPMh8Azcg5e_10TiDw



Wireless and Mobile Networking (Spring 2016),

https://www.youtube.com/playlist?list=PLjGG94etKypKeb0nzyN9tSs_HCd5c4wXF

CSE571S: Network Security (Fall 2011),

<https://www.youtube.com/playlist?list=PLjGG94etKypKvzfVtutHcPFJXumyyg93u>



Video Podcasts of Prof. Raj Jain's Lectures,

<https://www.youtube.com/channel/UCN4-5wzNP9-ruOzQMs-8NUw>