

IDI – Quantitative and Qualitative Methods for Human-Subject Experiments

Motivation

- Measuring the response of humans to different experiments is the only way to:
 - Evaluate how humans perceive, manipulate, reason with applications or webpages
 - Measure utility of applications and webpages
- Key issue in software development
- It is important to do **before** launching any product!



Outline

- *Motivation*
- **Validity of experiments**
- Experiment design
- Data analysis



Validity of experiments

- Experimental Validity
 - Does experiment really measure what we want it to measure?
 - Do our results really mean what we think (and hope) they mean?
 - Are our results reliable?
 - If we run the experiment again, will we get the same results?
 - Will others get the same results?



Validity of experiments

- Experimental variables: **Independent Variables**
 - What the experiment is studying
 - Systematically varied by experiment
 - Example: stereopsis, at the levels of stereo, mono



Validity of experiments

- Experimental variables: **Dependent Variables**
 - What the experiment measures
 - Assume dependent variables will be affected by independent variables
 - Must be measurable quantities
 - Time, task completion counts, error counts, survey answers, scores, etc.
 - Example: VR navigation performance, in total time; number of errors...



Validity of experiments

- Experimental variables:
 - Independent variables can vary in two ways
 - *Between-subjects*: each subject sees a different level of the variable
 - Example: 1/2 of subjects see stereo, 1/2 see mono
 - *Within-subjects*: each subject sees all levels of the variable
 - Example: each subject sees both stereo and mono



Validity of experiments

- Experimental variables: **Confounding factors** (or confounding variables)
 - Factors that are not being studied, but will still affect experiment
 - Example: stereo condition less bright than mono condition
 - Important to predict and control confounding factors, or experimental validity will suffer
 - E. g.: Mono vs stereo and brightness



Outline

- *Motivation*
- *Validity of experiments*
- **Experiment design**
- Data analysis



Experiment design

- To **avoid skewing effects**, experiments must be designed carefully
 - *E. g.: Learning a technique*
 - After N repetitions of the same experiment, the user will go fast to solve the same problem
 - *E. g.: Suffering fatigue*
 - After N repetitions, if the task requires physical effort, the performance may suffer



Experiment design

- Counterbalancing design:
 - Avoid learning/fatigue effects by randomizing the tasks
 - Randomizing does not necessarily mean random,
 - but sorting adequately users and conditions (systematic variation)



Experiment design

- Let's imagine we have 10 subjects and we want to test solving the same task (e. g. buying a book) using two different websites:

Subjects	First shopping	Second shopping
1, 3, 5, 7, 9	Website A	Website B
2, 4, 6, 8, 10	Website B	Website A



Experiment design

- Let's imagine we want to test solving the same task (e. g. buying a book) using three different devices (desktop, tablet, and mobile).
 - We will have the following conditions:

Device	Website	
	Website A	Website B
Smartphone		
Tablet		
Desktop		



Experiment design

- Say that we want each user to perform each task 4 times
 - We will have $3 \text{ (devices)} \times 2 \text{ (websites)} \times 4 \text{ (repetitions)} = 24 \text{ tasks}$
 - Note that this grows with a factorial explosion!!!
- To ensure reliability, those tests must be performed in the adequate order
 - Different for each subject



Experiment design

- Latin squares :
 - Tabular expression of systematic variations
 - Can be used to adequately sort experimental tasks
 - Counterbalances to avoid confounding factors
 - Within-subjects variables: control fatigue and learning effects
 - Between-subjects variables: control other factors that change with time (e. g. network speed, cache contents)



Experiment design

- Latin squares. Properties:
 - Every level appears in the every position the same number of times
 - Every level is followed by every other level
 - Every level is preceded by every other level



Experiment design

- Latin squares. Examples:

2x2

1	2
2	1

6x3

1	2	3
2	3	1
3	1	2

1	3	2
2	1	3
3	2	1

4x4

1	2	3	4
2	4	1	3
3	1	4	2
4	3	2	1



Experiment design

- Studying the previous example (3 devices) x 2 websites with 4 repetitions:
 - Form a Cartesian product of latin squares:
 - 6 conditions (3 devices) x 2 conditions (2 webs)
 - This will counterbalance properly a group of 12 subjects



Experiment design

Subject	Presentation order
1	1A, 1B, 2A, 2B, 3A, 3B
2	1B, 1A, 2B, 2A, 3B, 3A
3	2A, 2B, 3A, 3B, 1A, 1B
4	2B, 2A, 3B, 3A, 1B, 1A
5	3A, 3B, 1A, 1B, 2A, 2B
6	3B, 3A, 1B, 1A, 2B, 2A
7	1A, 1B, 3A, 3B, 2A, 2B
8	1B, 1A, 3B, 3A, 2B, 2A
9	2A, 2B, 1A, 1B, 3A, 3B
10	2B, 2A, 1B, 1A, 3B, 3A
11	3A, 3B, 2A, 2B, 1A, 1B
12	3B, 3A, 2B, 2A, 1B, 1A



Experiment design

- More examples in the PDF.



Outline

- *Motivation*
- *Validity of experiments*
- *Experiment design*
- **Data analysis**

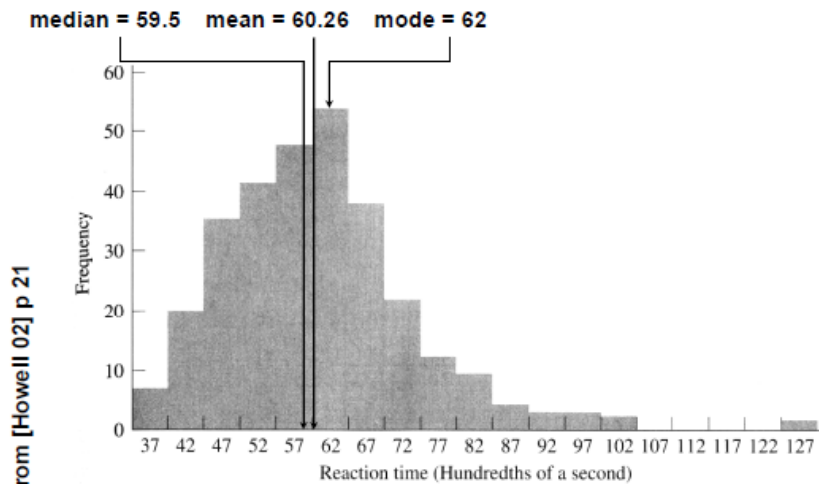


Data analysis

- *Descriptive statistics:*
 - Describe and explore data
 - All types of graphs, histograms...
 - Understand data distribution
 - Start to think of significance tests
- *Inferential statistics:*
 - Detect relationships in data
 - Significance tests
 - Infer population characteristics from sample characteristics



Data analysis



Data analysis

Mean:

$$\bar{X} = \frac{\sum X}{N}$$

Mean absolute deviation:

$$\text{m.a.d.} = \frac{\sum |X - \bar{X}|}{N}$$

Variance:

$$s^2 = \frac{\sum (X - \bar{X})^2}{N - 1}$$

Standard deviation:

$$s = \sqrt{\frac{\sum (X - \bar{X})^2}{N - 1}}$$

$$\sigma^2 = \frac{\sum (X - \mu)^2}{N}$$

- Standard deviation uses same units as samples and mean.
- Calculation of population variance σ^2 is theoretical, because μ almost never known and the population size N would be very large (perhaps infinity).

Data analysis

- Hypothesis testing, analysis of variance
 - Read in the PDF, your notes of statistics previous courses...



Adequate data representation

- The objective of a chart is to help user understand data



Adequate data representation

- Basic principles:
 - Avoid Pie charts
 - Avoid 3D projections of charts
 - Keep a high data to chart ratio
 - Use the appropriate graph for the appropriate purpose
 - And NEVER use a pie chart!



Adequate data representation

- Types of graphs
 - Trend graphs
 - Relative size graphs
 - Composition graphs
- Chartjunk: Unnecessary or confusing visual elements in charts and graphs



34

Adequate data representation

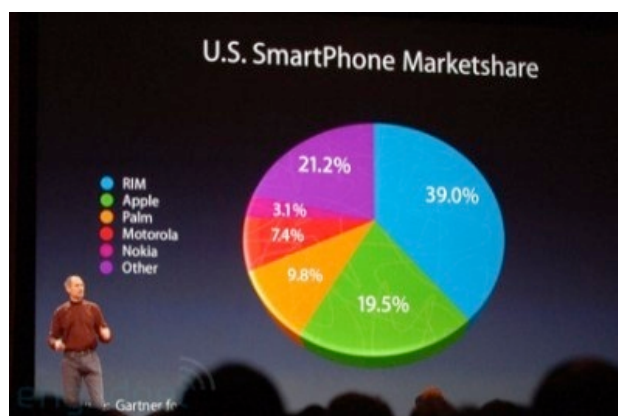
- Typical problems
 - Wrong graph type
 - Missing information (graph title, scale, labels,...)
 - Inconsistent scale (changes in the scale)
 - Misplaced zero point
 - Poor chart effects (ducks, shadows...)
 - Confusing of area and length
 - No adjustment for inflation
 - Too much precision
 - Poor ink-data balance



35

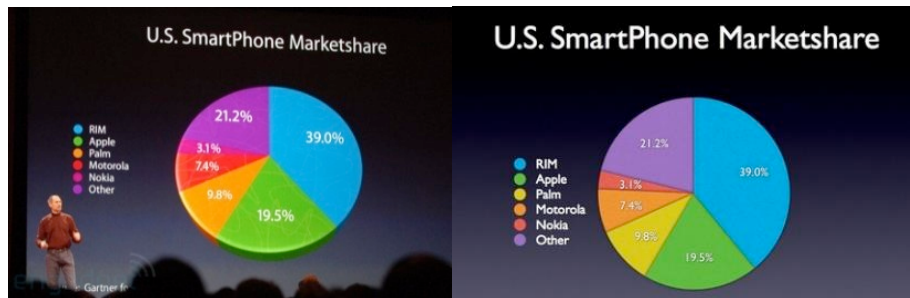
Adequate data representation

- Chartjunk



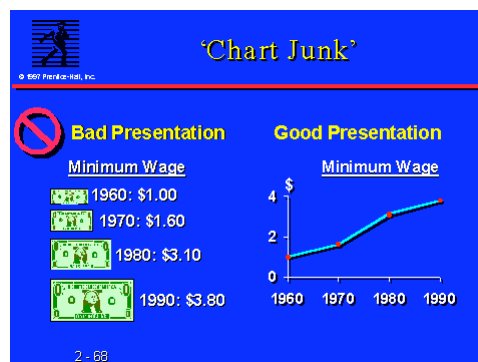
Adequate data representation

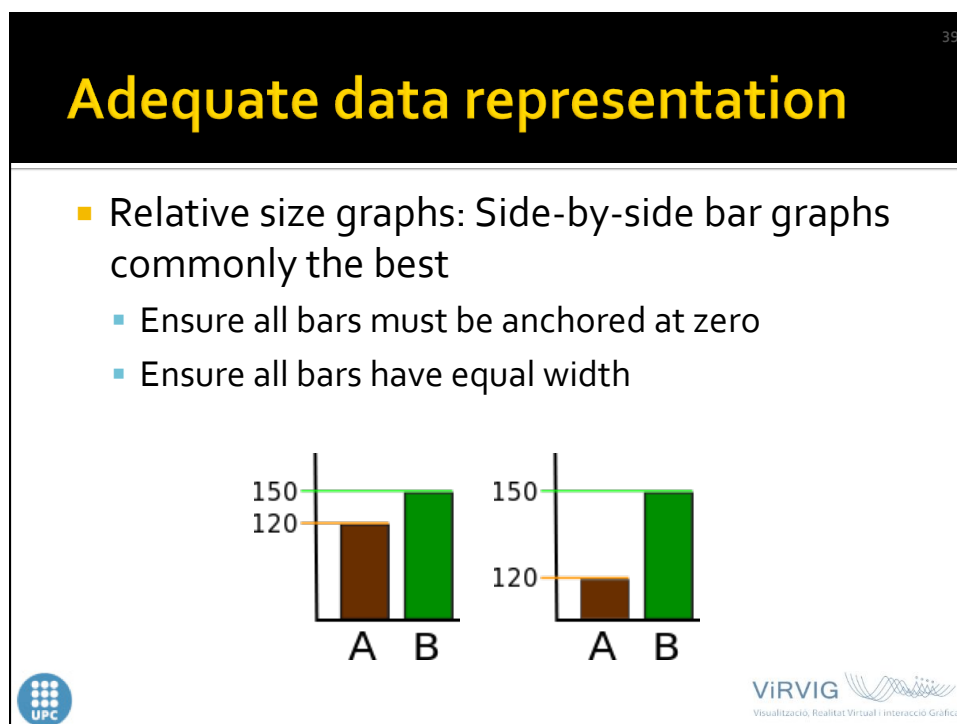
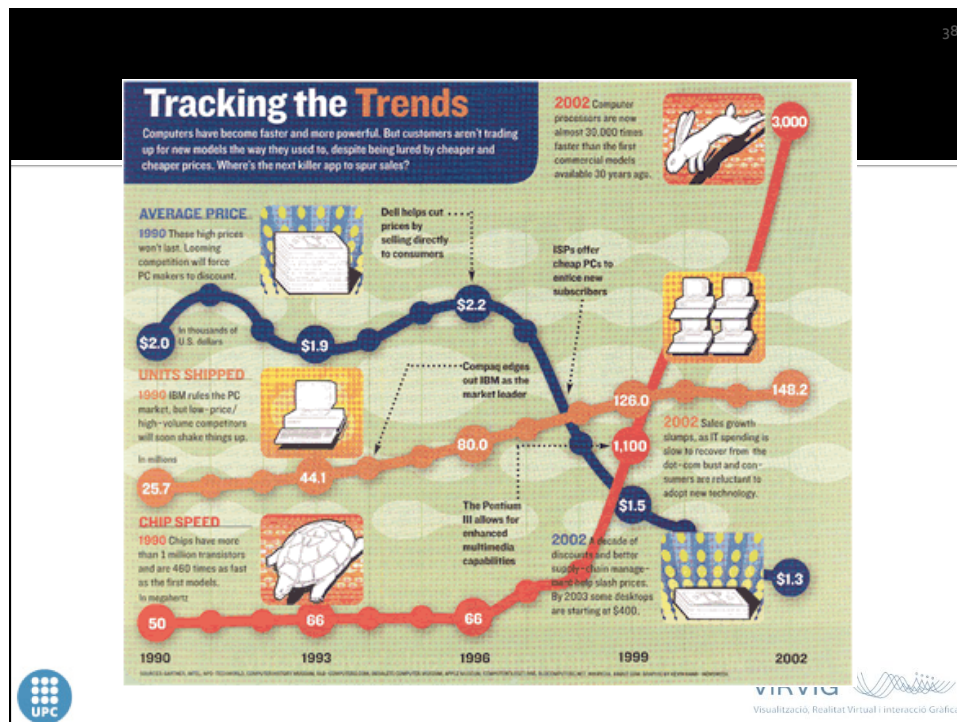
■ Chartjunk



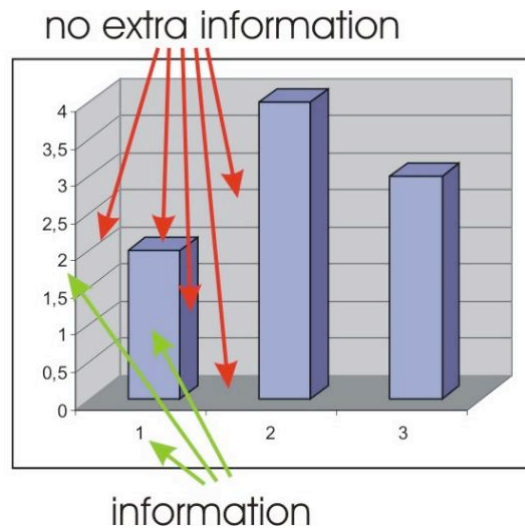
Adequate data representation

- Trend graphs: Time series
 - Line chart often better than bars





Adequate data representation



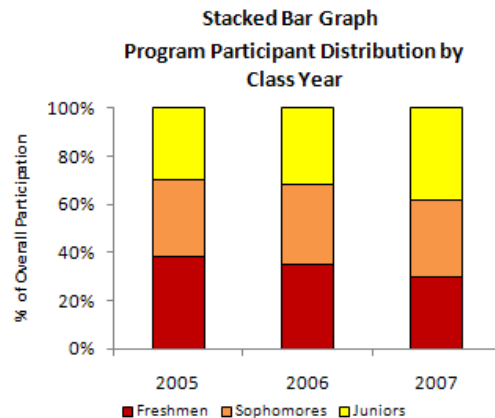
Adequate data representation

- Composition graphs
 - Pie-charts are inadequate:
 - People not able to compare angles properly
 - Better use segmented bar-chart where the bar (that stretches from 0 to 100%) is segmented into pieces.
 - Most important segments at the top or the bottom



Adequate data representation

- Composition graphs



Adequate data representation

- Make sure that the graph is complete. All axes must be labelled. There should be a title on the graph



Adequate data representation

■ Data-Ink ratio:

$$\text{Data-ink ratio} = \frac{\text{Data-ink}}{\text{Total ink used to print the graphic}}$$

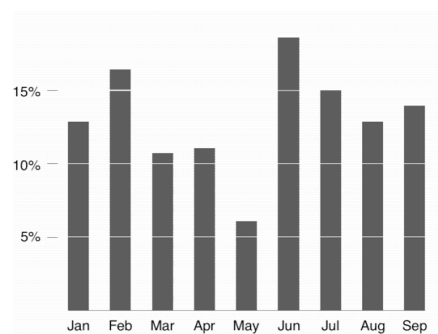
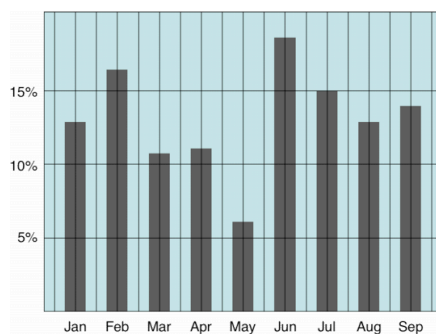
= proportion of a graphic's ink devoted to the non-redundant display of data-information

= 1.0 - proportion of a graphic that can be erased



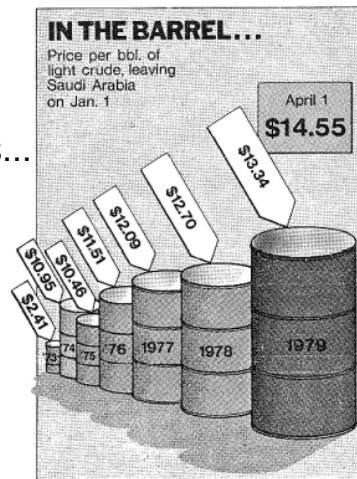
Adequate data representation

■ Data-Ink ratio:



Adequate data representation

- Not all examples are good
 - Be fair!!!
 - Tufte has plenty of examples...



VIRVIG
Visualització, Realitat Virtual i Interacció Gràfica

Adequate data representation

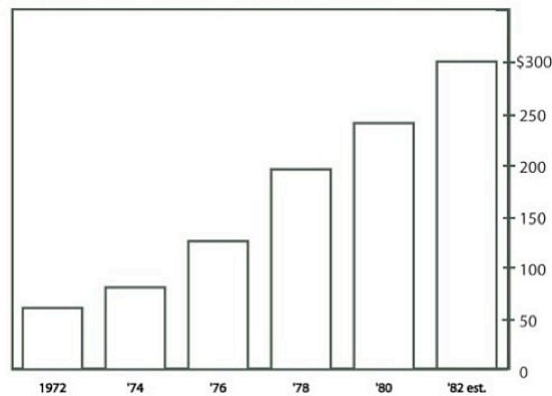


VIRVIG
Visualització, Realitat Virtual i Interacció Gràfica

Adequate data representation



MONSTROUS COSTS
Total House and Senate campaign expenditures, in millions

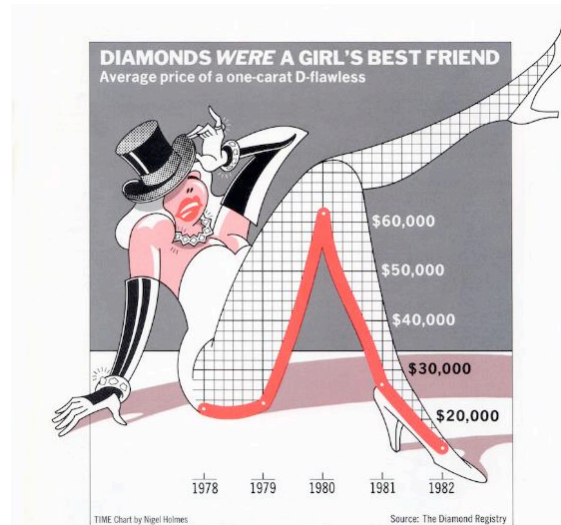


iRVIG
Visualització, Realitat Virtual i Interacció Gràfica

Adequate data representation

49

DIAMONDS WERE A GIRL'S BEST FRIEND
Average price of a one-carat D-flawless



TIME Chart by Nigel Holmes

Source: The Diamond Registry



RVIG
Visualització, Realitat Virtual i Interacció Gràfica