IDI – Quantitative and Qualitative Methods for Human-Subject Experiments

Outline

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

UPC

MOVING

Outline

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

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
  - Occur at different levels
    - Example: stereopsis, at the levels of stereo, mono
  - Systematically varied by experiment

UPC MOVING

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…

UPC MOVING

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

UPC MOVING

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

UPC MOVING

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:

	Website	
Device	Website A	Website B
Smartphone		
Tablet		
Desktop		

MOVING

Experiment design

- Say that we want each user to perform each task 4 times
  - We will have 3 (devices) x 2 (websites) x 4 (repetitions) = 24 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

MOVING

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

MOVING

Experiment design

- Studying the previous example (3 devices) x 2 websites with 4 repetitions:
  - Form a Cartesian product of latin squares:
    - 6x3 (devices) x 2x2 (conditions)
    - This will counterbalance properly a group of 12 subjects

MOVING

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

MOVING

Experiment design

- More examples in the PDF.

MOVING

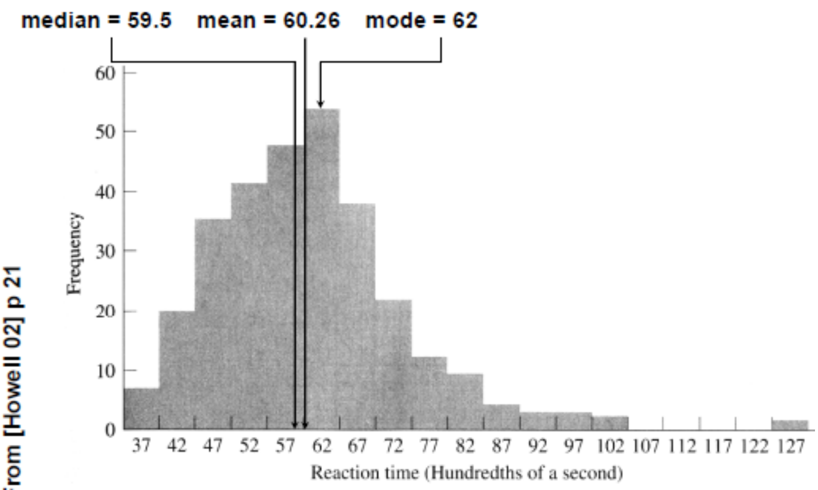
Outline

- *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



From [Howell 02] p 21

Data analysis

Mean:

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

Mean absolute deviation:

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

Variance:

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

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

Standard deviation:

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

- 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!

UPC

MOVING

Adequate data representation

- Types of graphs
  - Trend graphs
  - Relative size graphs
  - Composition graphs

- Chartjunk: Unnecessary or confusing visual elements in charts and graphs

UPC

MOVING

35

Adequate data representation

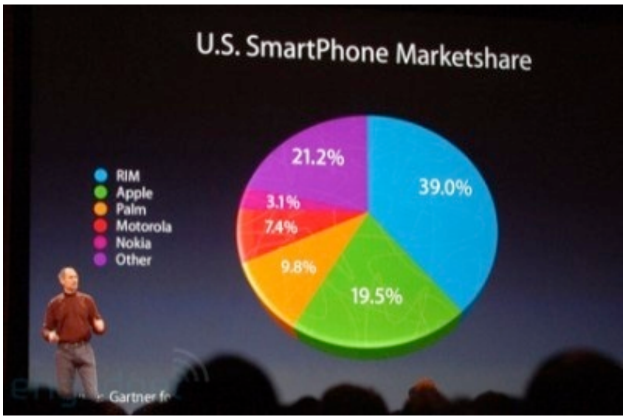
- Typical problems
  - Wrong graph type
  - Missing information (graph title, scale, labels,…)
  - Inconsistent scale (changes in the scale)
  - Misplaced zero point
  - Poor chart effects (ducks, shadows…)
  - Confusing of area and length
  - No adjustment for inflation
  - Too much precision
  - Poor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

MOVING

36

Adequate data representation

- Chartjunk

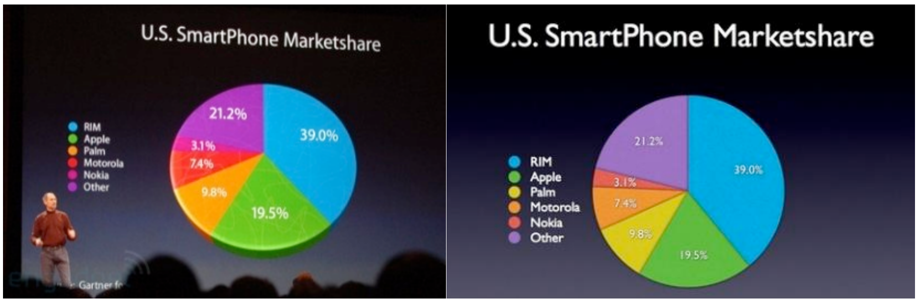


MOVING

37

Adequate data representation

- Chartjunk

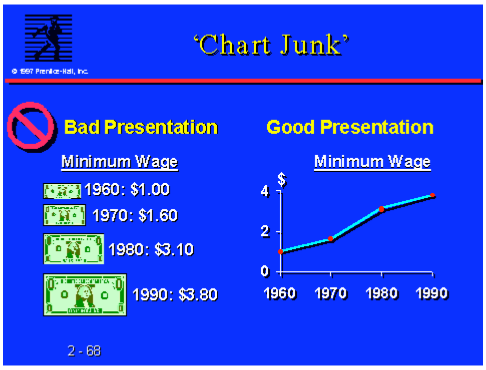


MOVING

38

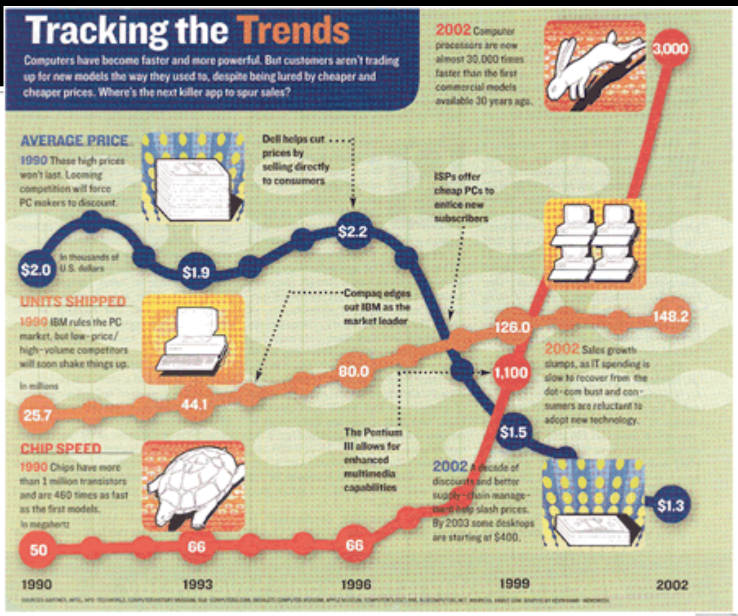
Adequate data representation

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



MOVING

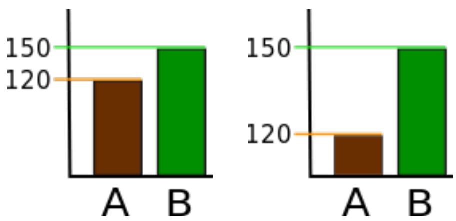
39



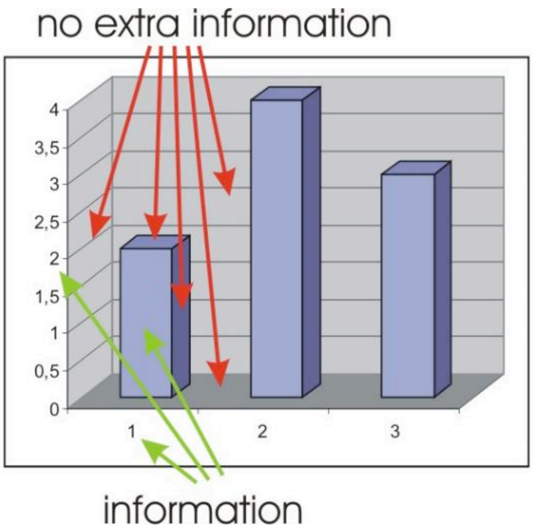
40

Adequate data representation

- Relative size graphs: Side-by-side bar graphs commonly the best
  - Ensure all bars must be anchored at zero
  - Ensure all bars have equal width



Adequate data representation



UPC

MOVING

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 streches from 0 to 100%) is segmented into pieces.
    - Most important segments at the top or the bottom

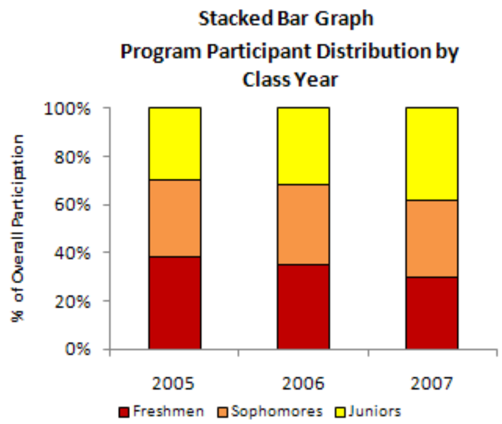
UPC

MOVING

43

Adequate data representation

- Composition graphs



UPC

MOVING

44

Adequate data representation

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

UPC

MOVING

45

Adequate data representation

- Data-Ink ratio:

Data-ink ratio = $\dfrac{\textbf{Data-ink}}{\textbf{Total ink used to print the graphic}}$

= **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**

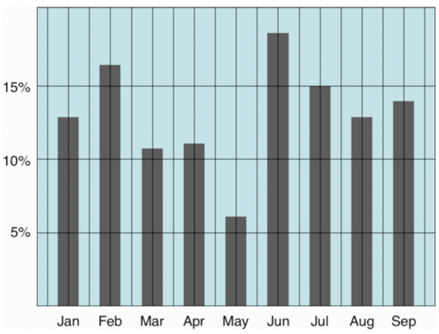
UPC

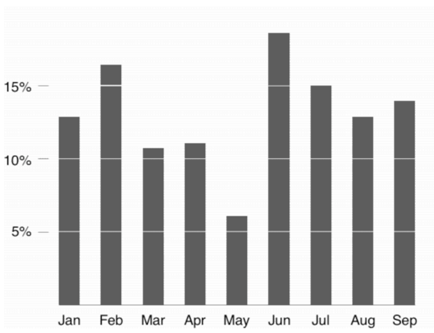
MOVING

46

Adequate data representation

- Data-Ink ratio:



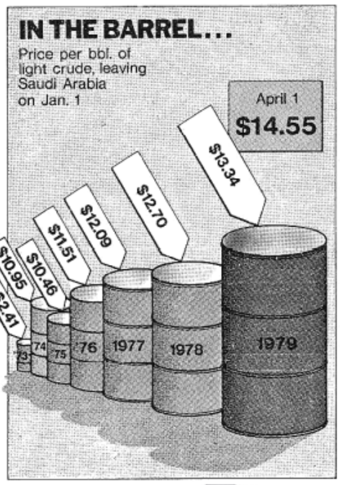


UPC

MOVING

Adequate data representation

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

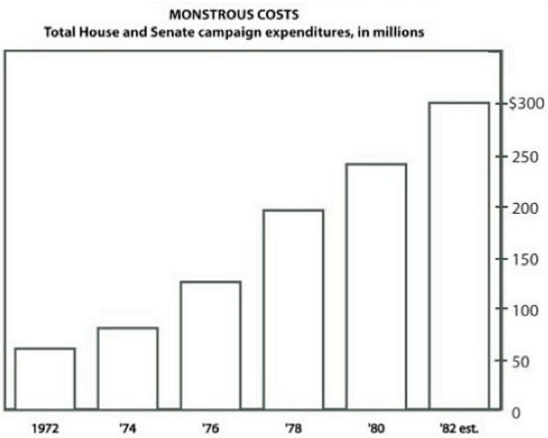


MOVING

Adequate data representation

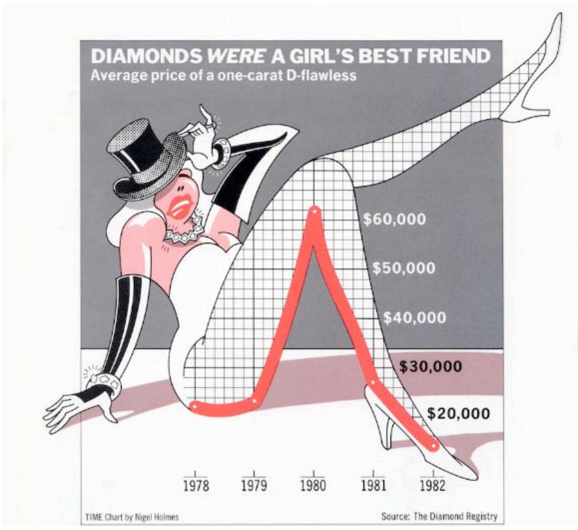
MOVING

Adequate data representation



50

Adequate data representation

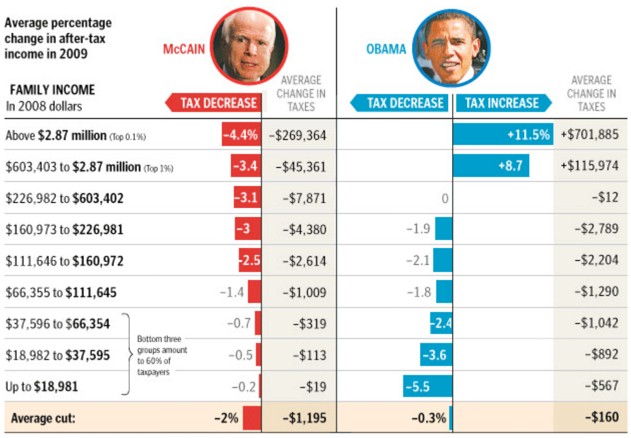


Pere-Pau Vázquez

IDI – Quantitative and Qualitative Methods for Human-Subject Experiments

52

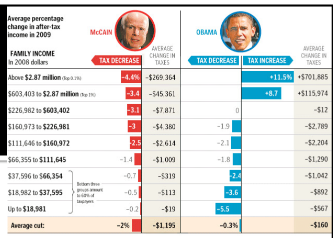
Chartjunk



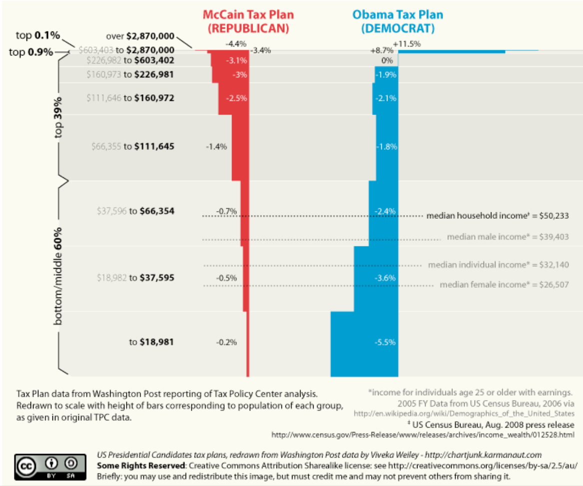
Average percentage change in after-tax income in 2009 / FAMILY INCOME In 2008 dollars	McCAIN TAX DECREASE	McCAIN AVERAGE CHANGE IN TAXES	OBAMA TAX DECREASE	OBAMA TAX INCREASE	OBAMA AVERAGE CHANGE IN TAXES
Above $2.87 million (Top 0.1%)	-4.4%	-$269,364		+11.5%	+$701,885
$603,403 to $2.87 million (Top 1%)	-3.4	-$45,361		+8.7	+$115,974
$226,982 to $603,402	-3.1	-$7,871	0		-$12
$160,973 to $226,981	-3	-$4,380	-1.9		-$2,789
$111,646 to $160,972	-2.5	-$2,614	-2.1		-$2,204
$66,355 to $111,645	-1.4	-$1,009	-1.8		-$1,290
$37,596 to $66,354	-0.7	-$319	-2.4		-$1,042
$18,982 to $37,595	-0.5	-$113	-3.6		-$892
Up to $18,981	-0.2	-$19	-5.5		-$567
Average cut:	-2%	-$1,195	-0.3%		-$160

Bottom three groups amount to 60% of taxpayers

Chartjunk



- Corrected version



MOVING