# Data and Experiments

# Collecting data – for evaluations or experiments

Five P's:

**Performance** – response time, errors

**Preference** – preference in a comparative task

**Perception** – free-form opinions

**Process** – actions performed in doing  a task

**Product** – artefact created by the participant

Purchase, H.C. Experimental Human-Computer Interaction, 2012, CUP

# Data Collection Methods

- Questionnaire
- Interview
- Answers on a task sheet
- Observational notes
- Automatic collection
  - answers
  - time taken
  - action log (internal system logs, video/audio of users, screen recordings)
- Artefact storage

| data | questionnaire | interview | task sheet | observation | automatic collection | artefact storage |
|---|---|---|---|---|---|---|
| Performance | | | x | x | x | x |
| Preference | x | x | | x | | |
| Perception | x | x | | x | | |
| Process | | | | x | x | x |
| Product | | | | x | | x |

| record | questionnaire | interview | task sheet | observation | automatic collection | artefact storage |
|---|---|---|---|---|---|---|
| evaluator | | x | | x | x | |
| user | x | | x | | | x |

| questionnaire | interview | task sheet | observation | automatic collection | artefact storage |
|---|---|---|---|---|---|
| easy to administer | not always easy to conduct | easy to administer | obvious thing to do | easy to collect | (maybe) easy to collect |
| difficult to design | difficult to design | based on task | difficult to gather systematic data | tempting to collect more than is needed | being unambiguous is crucial |
| typically quick | takes time | takes as long as the task itself | takes as long as the task itself | takes as long as the task itself | takes as long as the task itself |
| difficult to get detailed qualitative answers | can elicit detailed qualitative answers | quantitative responses to direct task-based questions | can be qualitative and/or quantitative | quantitative only; time consuming to analyse | qualitative product data |
| easy to get large numbers of participants | difficult to get participants | availability of participants will depend recruitment to do the task | availability of participants will depend recruitment to do the task | availability of participants will depend recruitment to do the task | availability of participants will depend recruitment to do the task |

# Terminology

subjective/ objective

quantitative/ qualitative

preference/ performance

open/ closed questions

formative/ summative

evaluation/ experiment

# ‘Evaluation’ *vs* ‘Experiment’

- qualitative *vs* quantitative
- formative *vs* summative

- small numbers *vs* large numbers
- visualisation tool *vs* visualisation method/idea
- large task *vs* small task
- lengthy *vs* brief

# Experiments

- **Hypothesis-driven**, or driven by a **Research Question**
- In IV, seeking verification that a proposed tool/ method is 'good'

- The experiment is defined by
  - the independent variable (what is being tested)
  - conditions (different values of the independent variable)
  - control variables (stay the same)
  - random variables (that you don't care about)
  - dependent variables (the data you collect)
  - tasks (what the participants are asked to do)

Experimental factors are things that can change during an experiment, e.g.

- the data used in a visualisation
- number of menu options available
- the colour palette used
- the tasks being performed by the participants
- the participants themselves
- temperature of the room
- the computer used

Some of these you want to change, some you want to keep the same, some you have no control over, some you don't care about

# Variables (or ‘factors’)

## Independent variables

Factors you want to change, because they are related to your RQ

Defined by a set of ‘conditions’

## Control variables (not really ‘variables’)

Factors you need to make sure do not change

## Random variables

Factors whose values you allow to change randomly

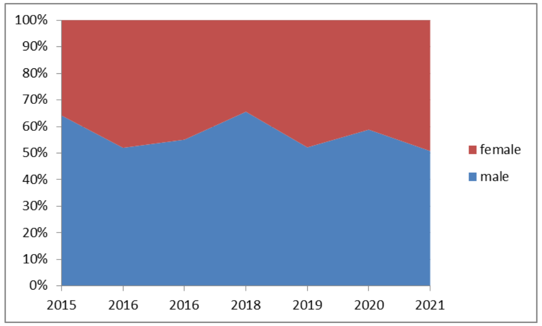
## Confounding factors

Factors that changed during the experiment (although you wish they hadn’t)

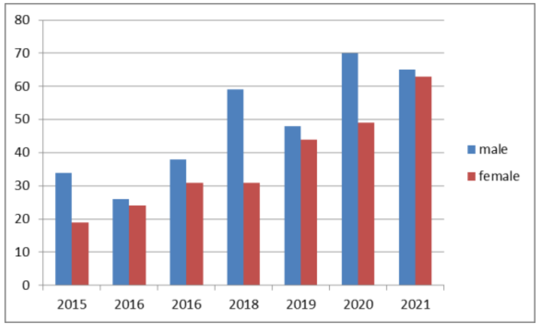
## Dependent variables

The data that you collect

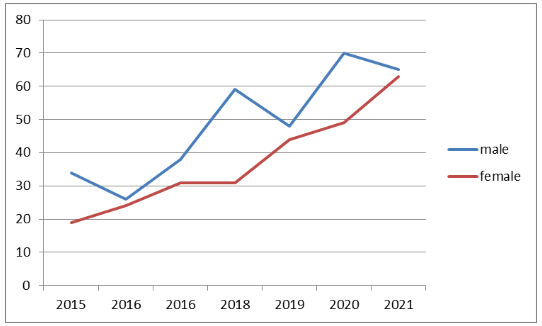
# Which data chart best shows the trends in club membership, per gender?




A




B




C

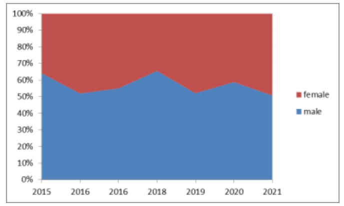
We would like to show that:

*a change in the values of the **independent variable** will result in a change in the values of the **dependent variable***

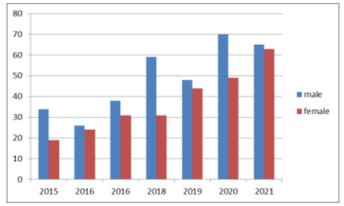
And we want to be sure that:

*any change in the value of the **dependent variable** is a result of the change in the value of the **independent variable** (rather than a result of change in the value of any other factors)*

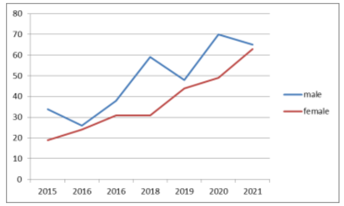
# RQ: *which data chart is best?*




- independent: type of data chart
- conditions:

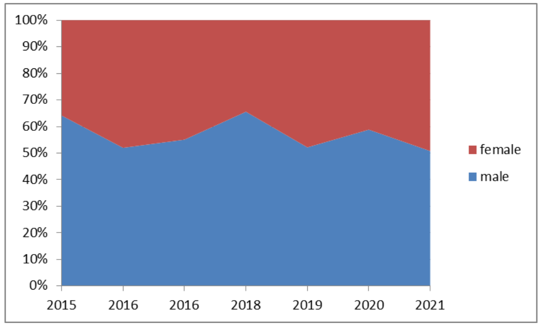



  - A: 100% stacked area chart
  - B: clustered column chart
  - C: line chart

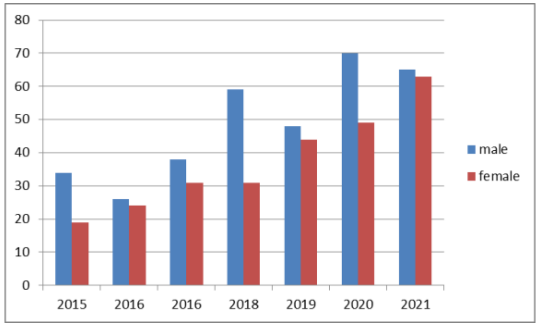



- control: the data sets presented
- control: the tasks (values to be read off the charts)
- random: size of computer screen
- random: age and gender of participants
- random: club membership; running ability
- dependent: accuracy in reading the values

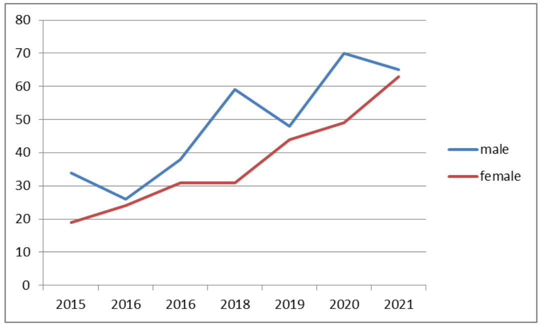
# Which data chart best shows the trends in club membership, per gender?




A

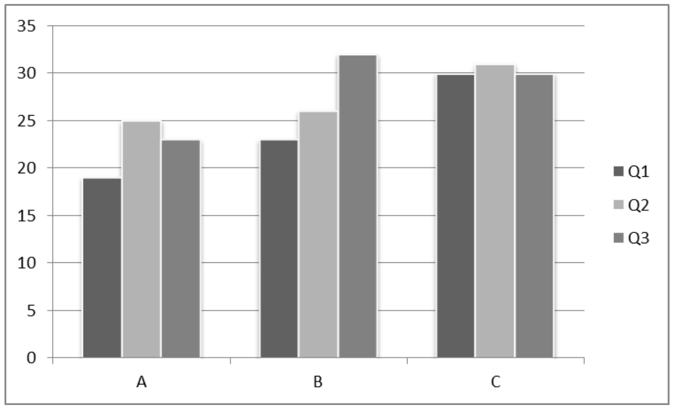


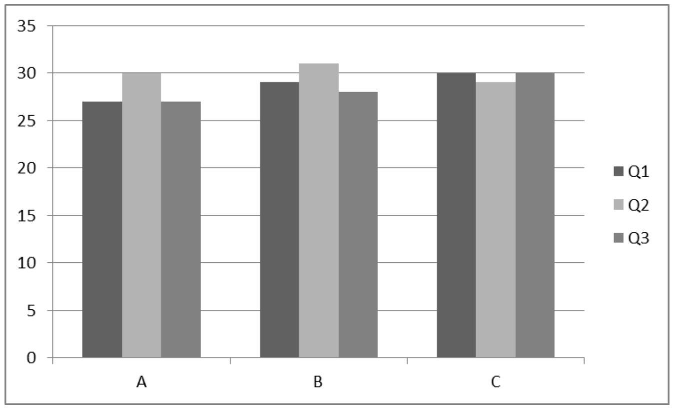

B




C

number of participants getting the answer correct in three questions (out of 32)



# Confounding factors

- Factors that changed during the experiment…
- … but only discovered after completion of the experiment

# Confounding factors

- Factors that changed during the experiment…
- … but only discovered after completion of the experiment

- These variables may have affected the value of the dependent variable…

# Confounding factors

- Factors that changed during the experiment…
- … but only discovered after completion of the experiment

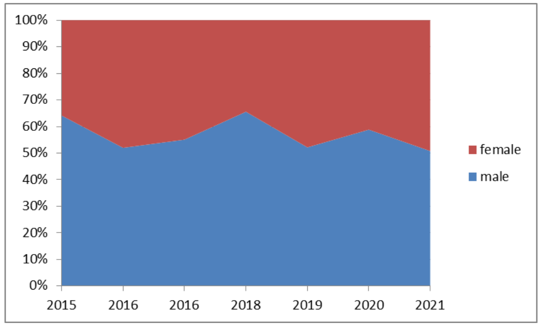
- These variables may have affected the value of the dependent variable…
- … and may have done so in way that means we cannot be sure that it was change in the value of the independent variable that affected the value of the dependent variable…

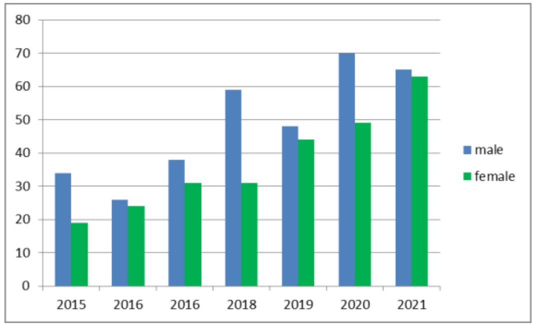
# Confounding factors

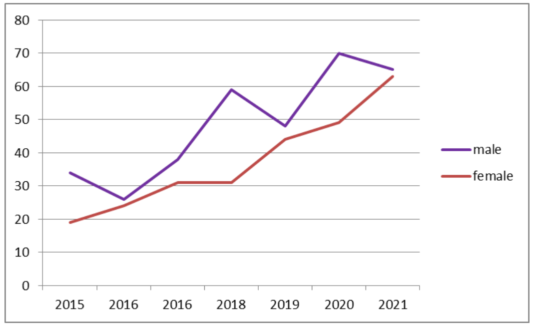
- Factors that changed during the experiment…
- … but only discovered after completion of the experiment

- These variables may have affected the value of the dependent variable…
- … and may have done so in way that means we cannot be sure that it was change in the value of the independent variable that affected the value of the dependent variable…
- … since it may have been the change in the confounding factor instead

# Which data chart best shows the trends in club membership, per gender?



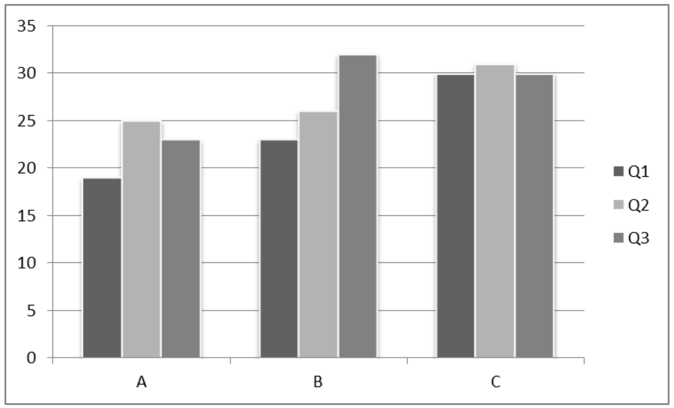



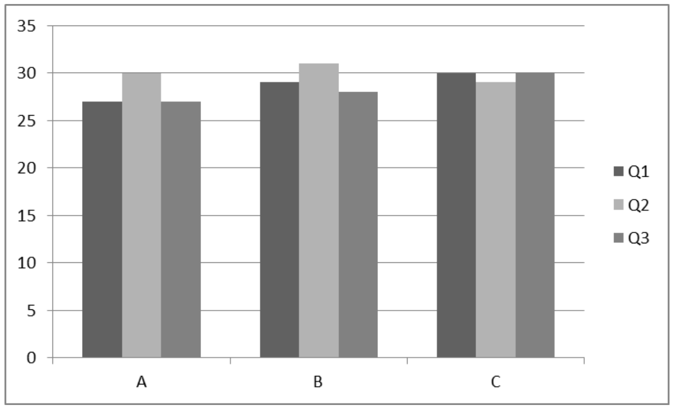

A

B

C

number of participants getting the answer correct in three questions (out of 32)



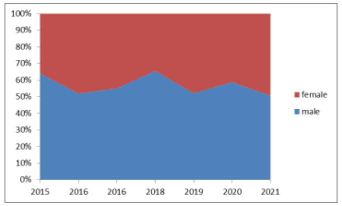
# Control variables

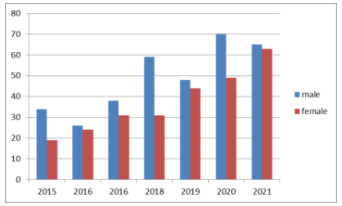
- Kept constant throughout the experiment
- If constant, these variables cannot affect the value of the dependent variable…

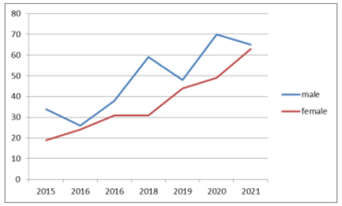
# Control variables

- Kept constant throughout the experiment
- If constant, these variables cannot affect the value of the dependent variable…
- … thus ensuring that we can be sure that it is the value of the independent variable that affects the value of the dependent variable

# RQ: *which data chart is best?*

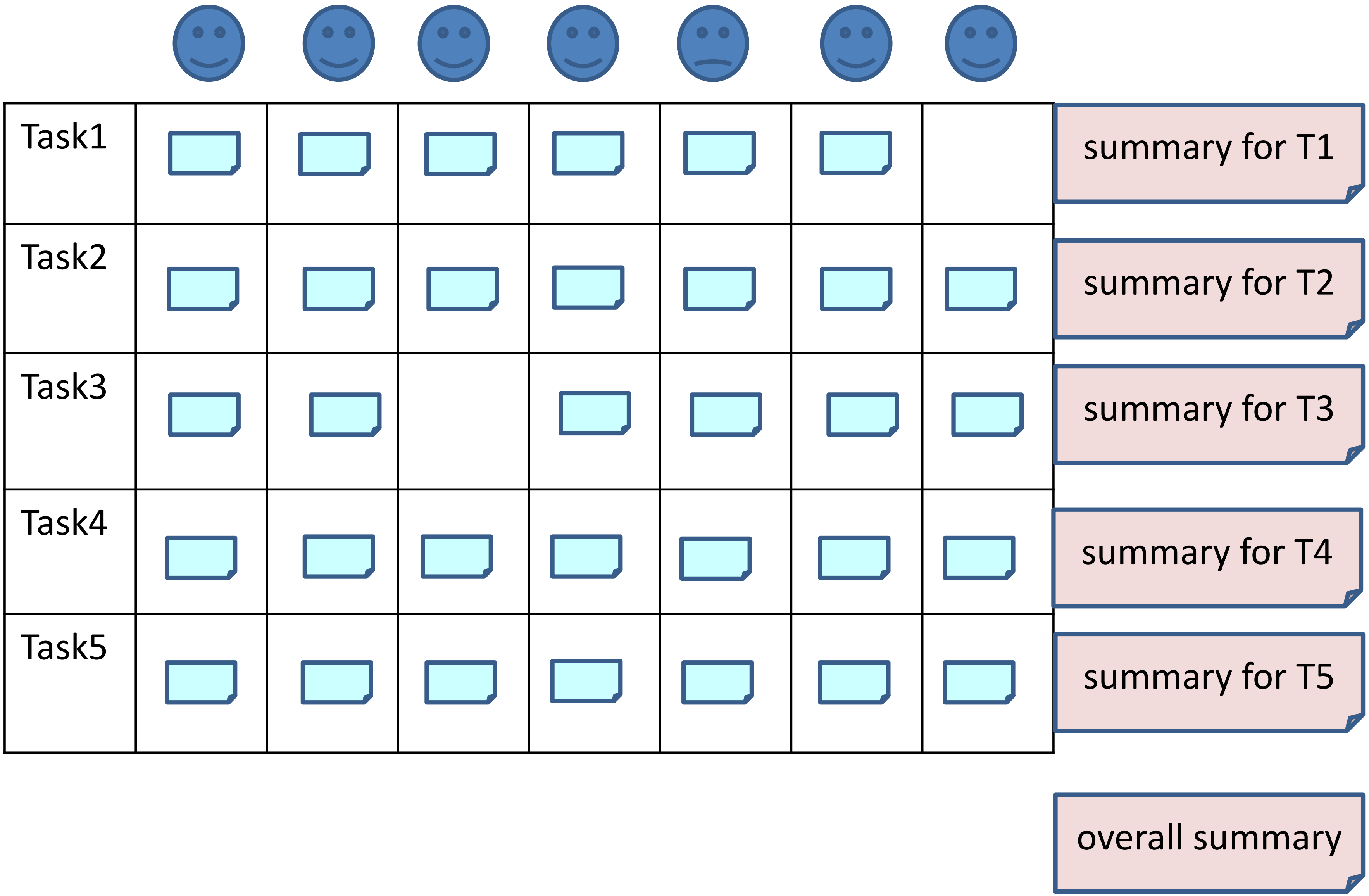



- independent: type of data chart
- conditions:
  - A: 100% stacked area chart
  - B: clustered column chart
  - C: line chart
- control: the data sets presented
- control: the tasks (values to be read off the charts)
- control: colours used in each chart
- random: size of computer screen
- random: age and gender of participants
- random: club membership; running ability
- dependent: accuracy in reading the values

# Data analysis overview

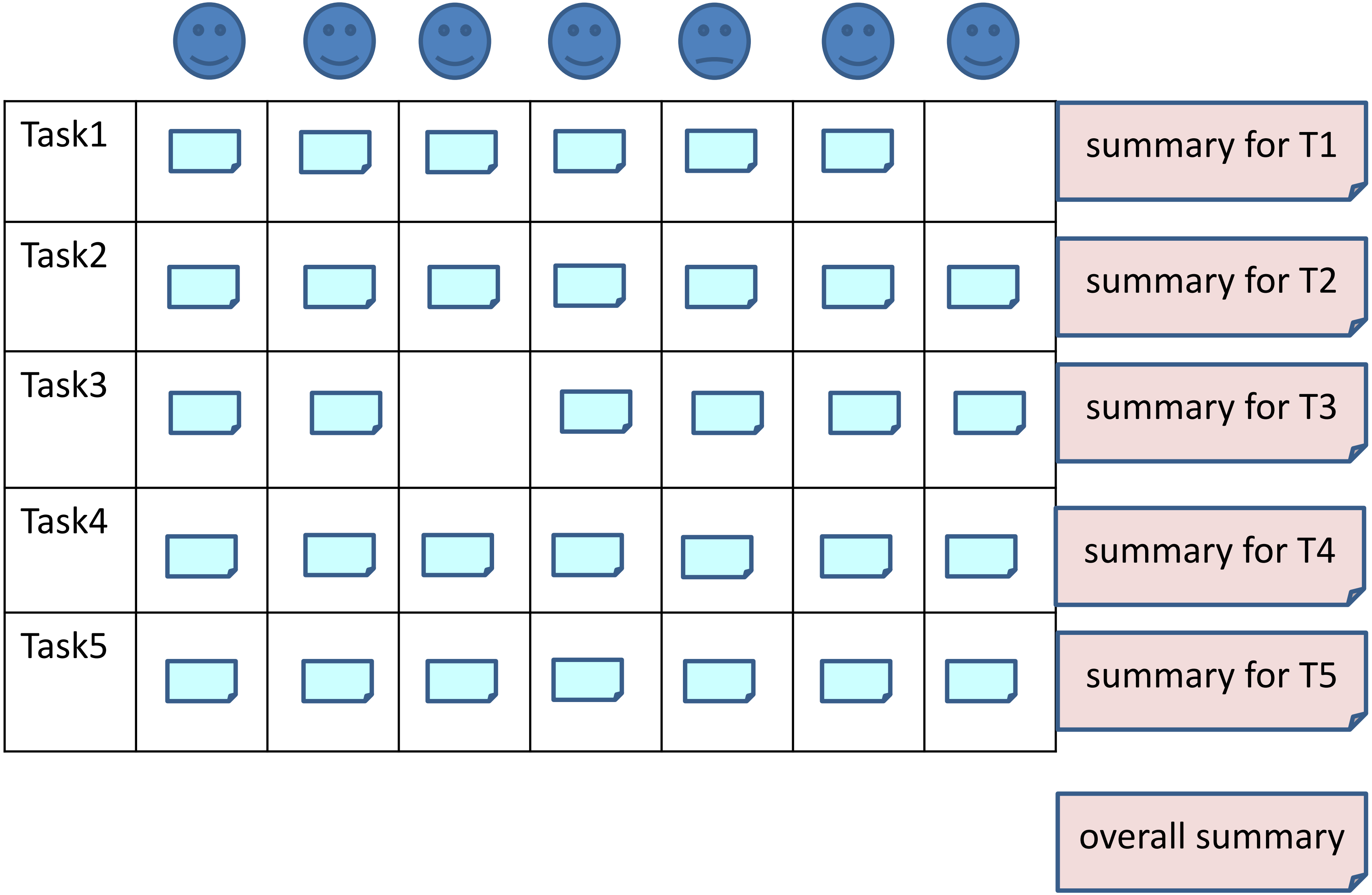


# Quantitative data analysis overview

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Task1 | 4 | 6 | 7 | 8 | 4 | 3 | | 5.33 | 53.33% |
| Task2 | 56 | 23 | 98 | 3 | 10 | 67 | 55 | 44.57 | 44.57% |
| Task3 | 7 | 8 | | 8 | 8 | 8 | 7 | 7.67 | 76.67% |
| Task4 | 10 | 12 | 15 | 14 | 13 | 16 | 20 | 14.29 | 71.43% |
| Task5 | 115 | 128 | 109 | 98 | 133 | 87 | 101 | 110.14 | 73.43% |

63.89%

# Qualitative data analysis overview



# Qualitative Data Analysis

- Analysing data from, e.g.:
  - Interviews
  - Focus groups
  - Open-ended question responses in a questionnaire
- Transcribe any audio/video data
  - Familiarise yourself with *all* the data
- Inductive coding
  - Find the themes in the data, in a structured and rigorous way
- Long-established classic textbook
  - *Thematic Analysis: A Practical Guide*, Braun and Clarke

# Qualitative Data Analysis: Inductive Approaches

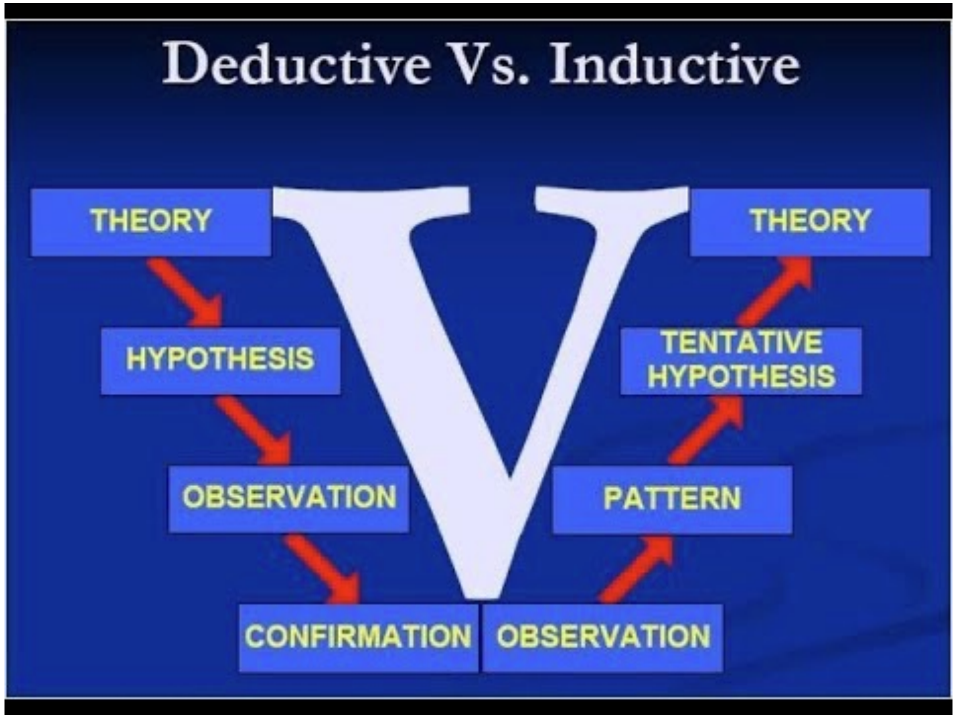
Thematic analysis

One form: grounded theory

- No preconceived theories that bias
- Open mind to what may be in data
- ‘Theory’ is what eventually ‘emerges’ from the process

Verify coding with multiple coders at different stages:

1. Open coding
2. Axial coding
3. Selective coding



# Stage 1: Open coding

Identify distinct pieces of information

- Assign an open code
- In-vivo coding: use participants' own words to define codes

Size/scope of pieces are determined by the coder's interpretive process

Participant: I think teens use drugs as a release from their parents [rebellious act]. Well, I don't know. I can only talk for myself. For me, it was an experience [experience] [in vivo code]. You hear a lot about drugs [drug talk]. You hear they are bad for you ["negative connotation" to the drug talk]. There is a lot of them around [available supply]. You just get into them because they're accessible [easy access] and because it's kind of a new thing [novel experience]. It's cool! You know, it's something that is bad for you, taboo, a "no" [negative connection]. Everyone is against it [adult negative stance]. If you are a teenager, the first thing you are going to do is try them [challenge the adult negative stance].

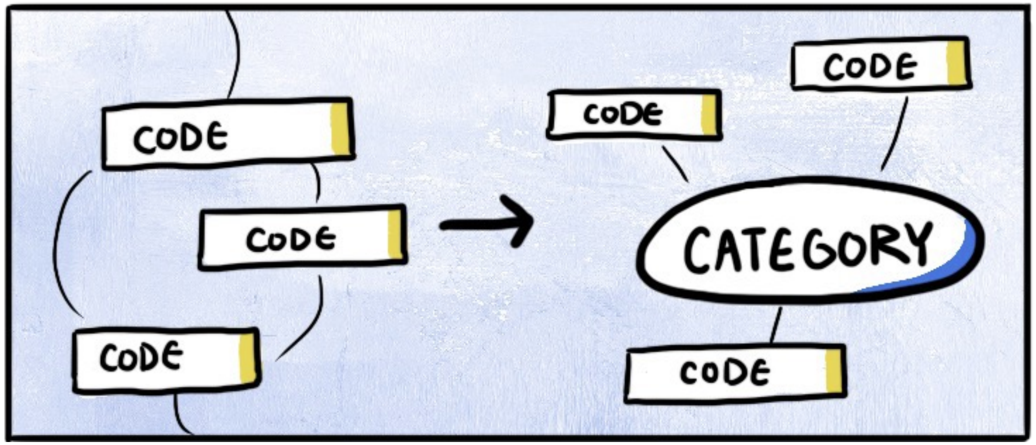
# Stage 2: Axial coding

Organise open codes into sets of concepts/categories

Think about the relationships between concepts

Concepts and categories don't have to all be at the same level of specificity, or need the same number of codes assigned to them

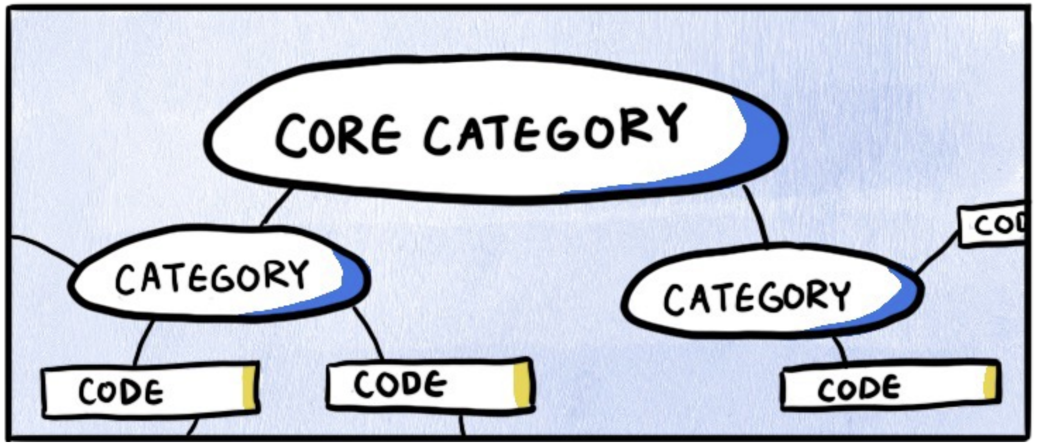
- Be driven by what is in the data, not what you think of as 'even' or ideas of symmetry/balance…



# Stage 3: Selective coding

Combining concepts and categories, to form larger concept framework

Re-code original transcripts using this new framework



# Reporting Results from Qualitative Data Analysis

First, set out the coding framework

Then if, for example, a thematic analysis uncovers 5 main themes

- 5 main sections in your report, each explaining one theme
- Use participant quotes as evidence, perhaps also discussion of how often or in what contexts instances of the theme happened

Then, discuss overall or higher-level findings

- Put this in the context of related work, e.g. reinforce/constrast other studies, consider different factors that might influence the data and the participants
- If appropriate, discuss implications for future system design requirements

***Data extract:***

| | |
|---|---|
| • *Too cluttered* | *[001]* |
| • *I'm loving the bright colours* | *[002]* |
| • *Things way too close together* | *[001]* |
| • *Not enough space* | *[001]* |
| • *Too garish* | *[003]* |
| • *I like the busy-ness* | *[004] [006]* |
| • *Very squished up* | *[001]* |
| • *Wonderfully active and lively* | *[005] [006]* |
| • *Bright and lively* | *[002] [005] [006]* |
| • *Colourful* | *[002]* |

(…more data items)

**Codes (or "themes"):**

001: cluttered

002: bright colours (+ve)

003: bright colours (-ve)

004: busy (+ve)

005: lively (+ve)

006 = 004+005: lively (+ve)

**Code count:**

001 [44]

002 [32]

003 [5]

006 [12]

Most participants (44/70) complained that the screen was too cluttered, while 12 of the other participants liked the "busy-ness" and activity that the display depicted. The brightness and colours were appreciated (32/70), although a minority (5/70) said that the colours were too bright and "made my eyes hurt".

# Number of participants

Formative evaluations:

Jakob Nielsen (2000), "*Why You Only Need to Test with 5 Users*"

"Elaborate usability tests are a waste of resources. The best results come from testing no more than 5 users and running as many small tests as you can afford."

Summative evaluations:

Probably more than this, e.g. 10, to be sure of  the validity and generalisability of the 'proof of worth' outcome

Experiments:

For statistical analysis, even more; more than 30 is an oft-quoted minimum

(https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/)

# Key principles of evaluations & experiments

Participant time is your most precious resource

Pilot: try out your evaluation/experimental methodology in advance to make sure it will produce the data needed

– phrasing of tasks
– adequacy of training or tutoring, before starting tasks
– expected duration of experiment session (too long/short?)

Don’t start until you are sure that the data you will collect will be useful to you; create ‘dummy’ data to see if it makes any sense

Participants should never be confused as to what to do

Get as much as you can out of the participant in each session

# A Word of Warning

Quantitative analysis is NOT better than qualitative analysis

- Too often people are biased towards 'clean' quant models, but they forget the subjectivity/variability/errors of parameters, metrics, models, etc.
- It's common in qualitative analysis to get multiple people to code the same data independently, and then compare. It's rare to get multiple people to do quant analysis on the same data, independently, and then compare.

Each can be done well or badly, but neither is better than the other

- Rigour of good measurement and modelling is similar in level to the rigour of good coding, concept formation, and theory building

Ideally, do both: Use quant to learn *what* people do, and qual to learn *why*

# Data and Experiments