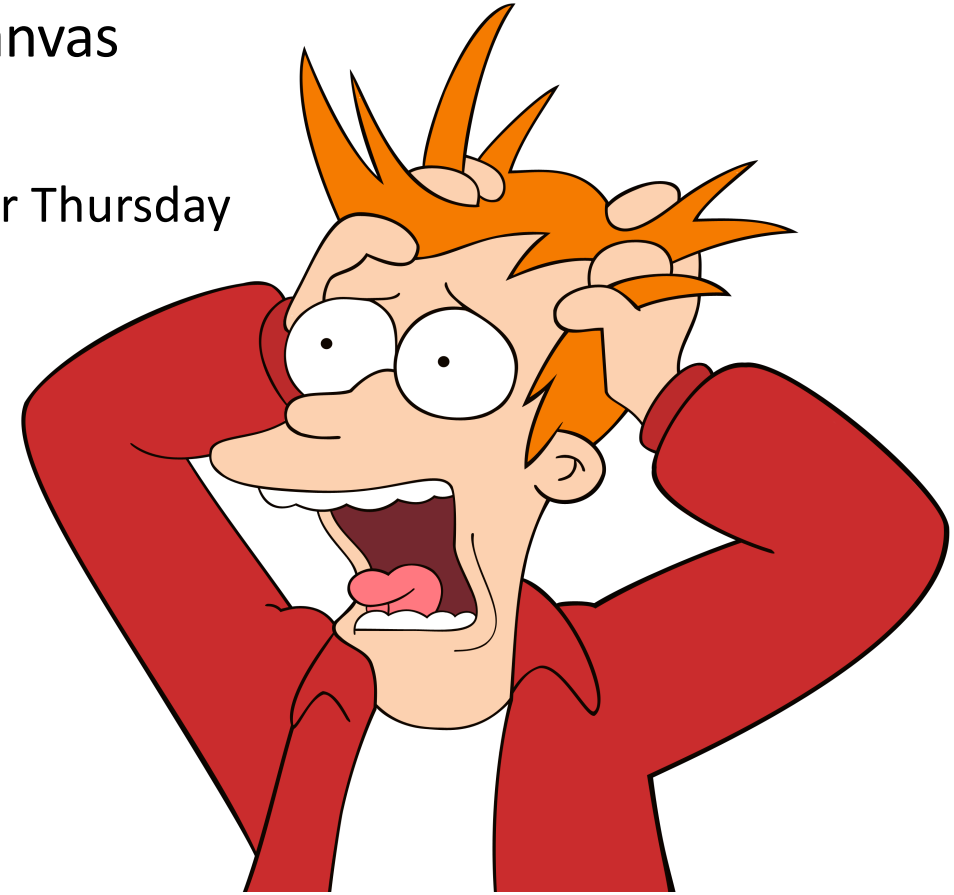
Intro to Experimental Design

Dr Theodoros Georgiou

Before we start

- Coursework specification is now on Canvas
- Group Schedule now on Canvas
  - Check if you are scheduled for Monday or Thursday for week 3!



In this lecture…

- Why do we run Experiments?
- Why to collect data?
- What is an experiment/study
- Types of data
  - Objective vs Subjective
  - Quantitative vs Qualitative
- What is a variable in an experiment
- Different types of quantitative data

Why do we run Experiments?

- I am a Computer Scientist!
  - I'm not a physics or a chemistry student!!
  - Why do I need to run experiments?!?!?
  - Why should I collect data?

- If you don’t, how do you know your design is any good?

You are not your users

- Remember previous lecture?
  - You are not your users
  - And your users are not you.
- Experiments allow you to see and evaluate your system from your **actual** users' perspective.

Empirically driven Design

- Experimental design
  - How to design experiments
- How to use empirical methods to *inform* and *evaluate* design

Motivating Example

- Imagine you have a company that builds a touch screen self checkout scanner

Example – why do we collect data?

- You want to convince Tesco to switch to your machine
- How do you show them yours is better than their existing machine?
- You need to run an experiment and collect empirical data!
  - One way would be to recruit some usability experts
    - Run an experiment
    - Prove empirically (i.e. with data) that your system is e.g. faster, better, more accurate, and easier to use.
    - This will show your results in a way that is quantifiable and can support your claims.

What is an experiment?

- Scientific study methodology
  - Allows conclusions about causality
- Hypothesis(es)
  - Statement to be proven
- Independent variable(s)
  - What you manipulate or change
  - This is the part of your experiment you have control over.
- Dependent variable(s)
  - What you measure – data you are collecting

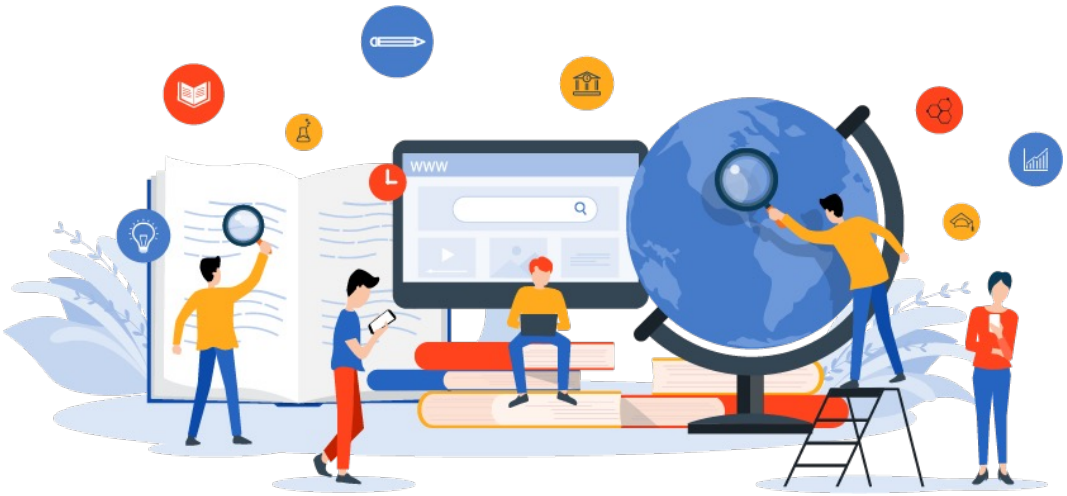
Hypothesis

- Don’t over complicate an experiment
- Home in on one or two statements you would like to prove
  - E.g. the machine is faster to use than Tesco’s current one
  - E.g. the machine is more accurate than the current one
- The more hypotheses, the harder it is to design a well-controlled experiment to test them all.

- More on Hypotheses later on..

Experiments...things to think about

- What are we evaluating?
- Narrow it down
  - Figure out what you are comparing
  - Keep everything else the same
- Who are your participants going to be?
  - Students?
  - Older people?
  - Or a mix?
- Consider the procedure.
  - What order should they do things in?
  - How long it will take etc.



Isolating your variables

- Which is better?
- Careful.....what are you testing?
  - Interactive interface?
  - Screen size?
  - Bagging area?
  - Voice?
- Things you are not testing for should remain the same.

- Isolate what you want to study - control for everything else

Example conclusions

- My self-checkout machine is better because:
  - 80% of people said they were highly satisfied with the checkout process
  - 90% of the participants said they would use it in the future compared to 60% for Tesco's
  - The checkout process is 50% faster *
  - There were 25% fewer unrecognised scans *

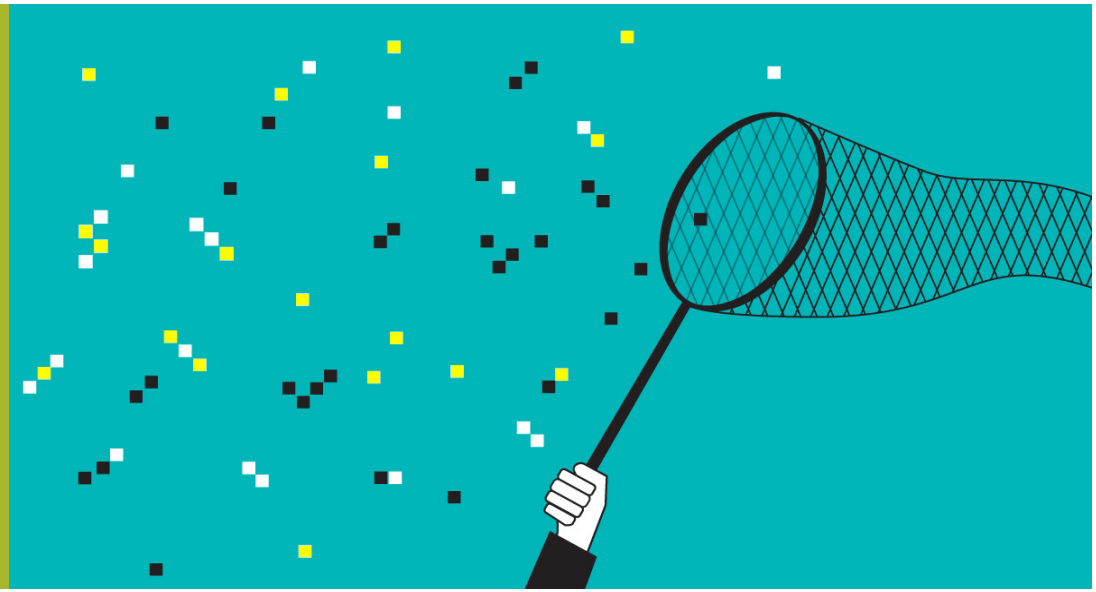
- * These differences are statistically significant

Observational vs Experimental

- *Observational:*
  - *The independent variable is not controlled by the researcher*
  - Observing children on the playground, you see that they are eating skittles. Later they behave aggressively
- *Experimental:*
  - *Researchers control (and test) using the independent variable*
  - Give 1 group of children a packet of Skittles and another group an apple and see which group is more aggressive

Data

- Whether you are comparing two systems OR
- Doing a usability test for one system
  - ...you need to collect data

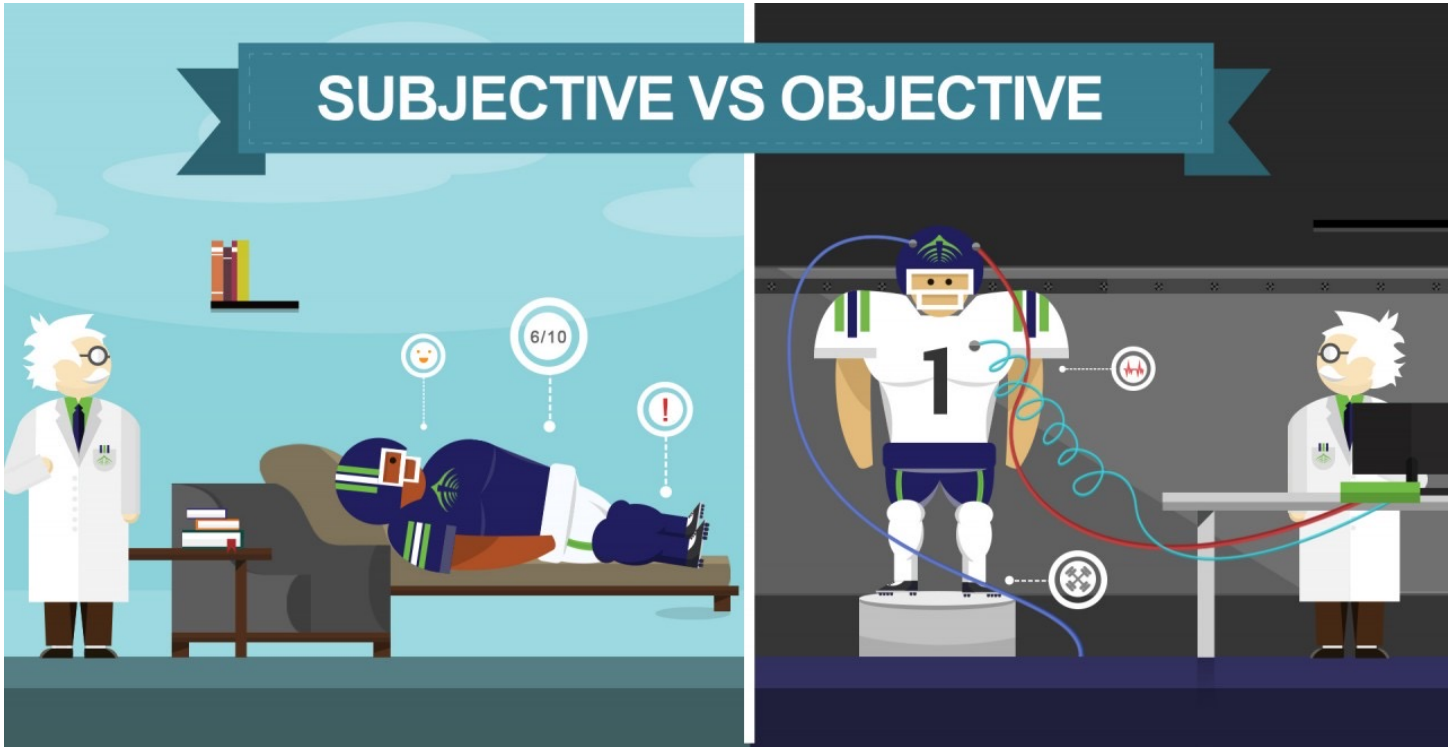


Data

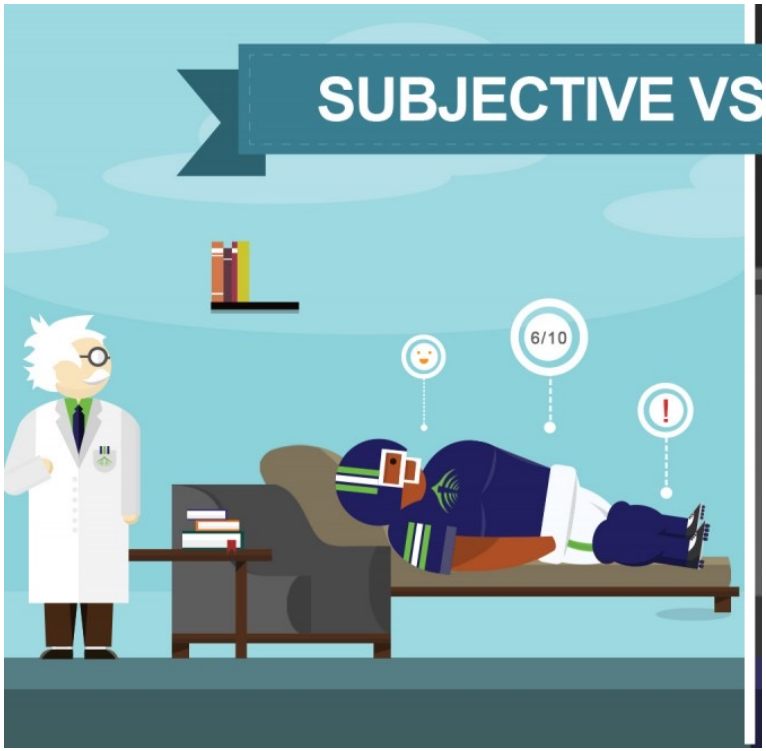
- But why….?

- Gives you evidence for your claims
- Types of data
  - Subjective vs Objective
  - Quantitative vs Qualitative

Subjective vs Objective



Subjective Data



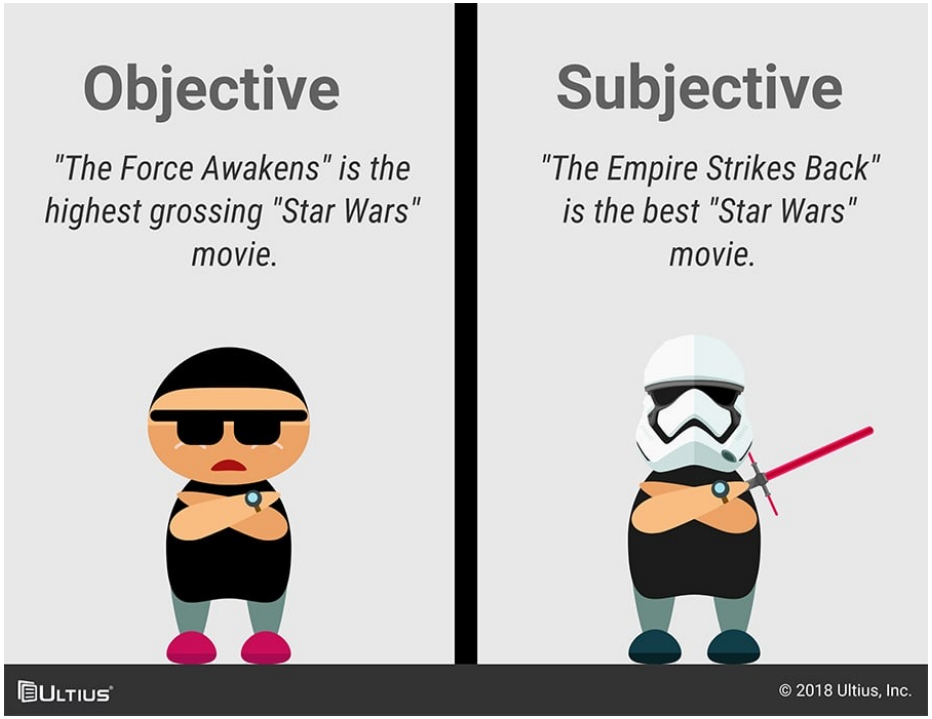
- Subjective data is data that is collected or obtained via personal interactions,
  - i.e., talking, sharing, explaining, etc.
- It is collected to make an assumption about what the fact might be, what event might have occurred, what calculations may have to be done, etc.
- This data varies from one person to another, with every situation, every minute.
- It cannot be declared as the truth, as it evolves from so many sources with different inputs.

Objective Data

- Objective Data is the physical data we can observe using our senses.
- Objective findings come in either a measurement or a direct observation.
- Objective data cannot be argued, as it is measured and observed through tests, and experiments
- Objective data is often easier for many to discern, as there is no gray area.

Subjective vs Objective

- Through **subjective data**, an idea/assumption can be formed.
- This can then be verified via gathering **objective data**.



Quantitative and Qualitative Data

- Quantitative data is data that is in the form of numbers
  - Or that can easily be translated into numbers.
    - Number of years an interviewee has
    - The number of projects the department handles
    - The numbers of seconds/minutes it takes to complete a task
- Qualitative data is not expressed in numbers or numerical forms
  - Descriptions, quotes from interviews, images, etc.

Quantitative and Qualitative Data

- Incorrectly assumed that certain forms of data gathering techniques can *only* result in quantitative data and others *only* in qualitative.
- For example:
  - A questionnaire may contain ask for the participant's age, and how many software packages they use in their day to day life (Quantitative), but may also have comment fields do describe an aspect of their life (Qualitative).
  - In observation studies, quantitative data may include the number of people in the study, hours each participant spends trying to work out a solution, while researcher notes interactions between team members, and opinions of participants (Qualitative).

Quantitative and Qualitative Data

- In summary:
- Quantitative data analysis uses numerical methods to ascertain magnitudes, amount, or size of something.
  - Average person is 170 cm, weighs 56.5kg, and is 45years old

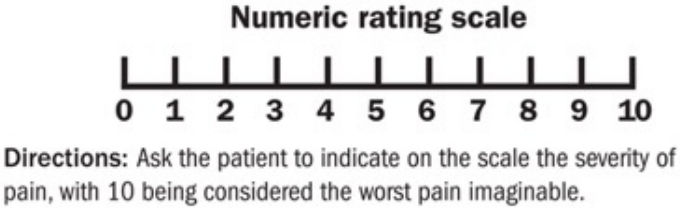
- Qualitative data analysis focuses on the nature of something and can be represented by themes, patterns, and stories.
  - Average person is tall, thin, and middle aged.

Subjective Quantitative

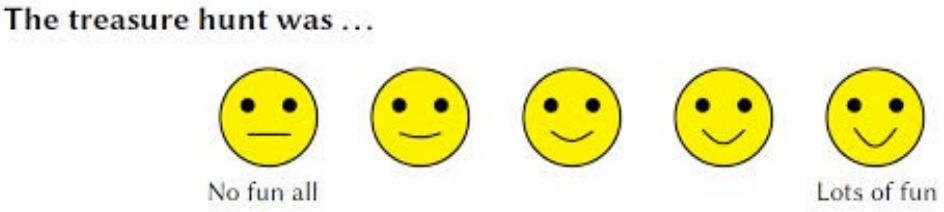
- Likert scale (words + optional numbers)

"I like broccoli"	Agree	1	2	3	4	5	Disagree

- Rating (numbers)



- Smileyometer

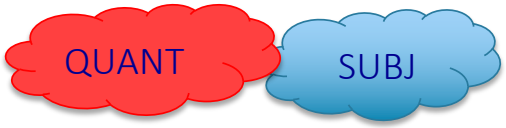


Example

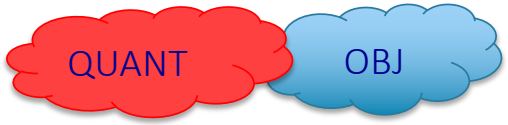
- A group of researchers want to test whether jetlag affects people's ability to drive safely
- They have 2 groups: one group is jet lagged- the other is not. Each participant drives in a driving simulator
- The hypothesis is that the the jet lagged group will perform worse

Example

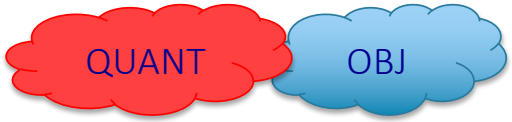
- Questionnaire asking questions how they thought they drove (using a 1-5 point Likert scale)



- Heart-rate of the user while driving

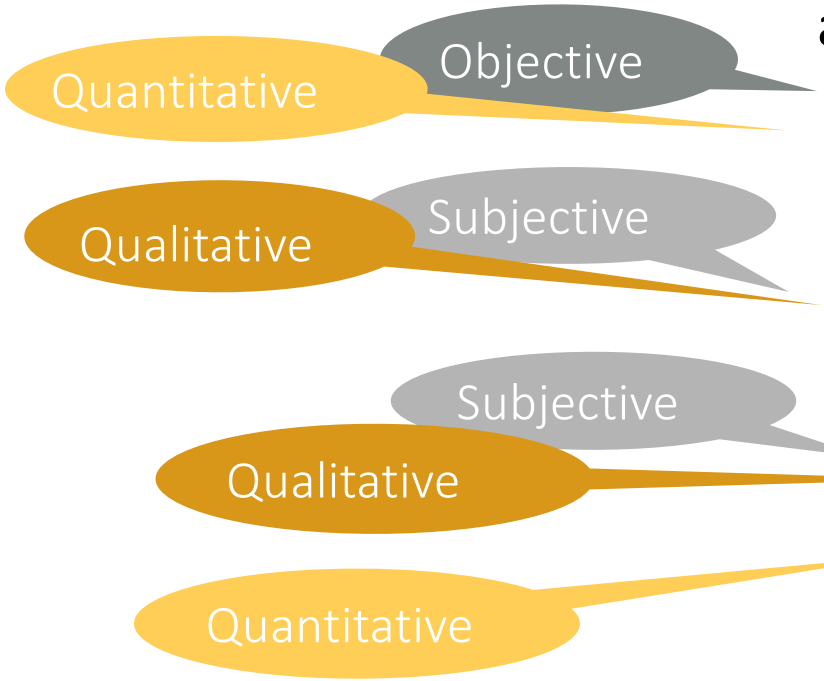


- Number of driving mistakes

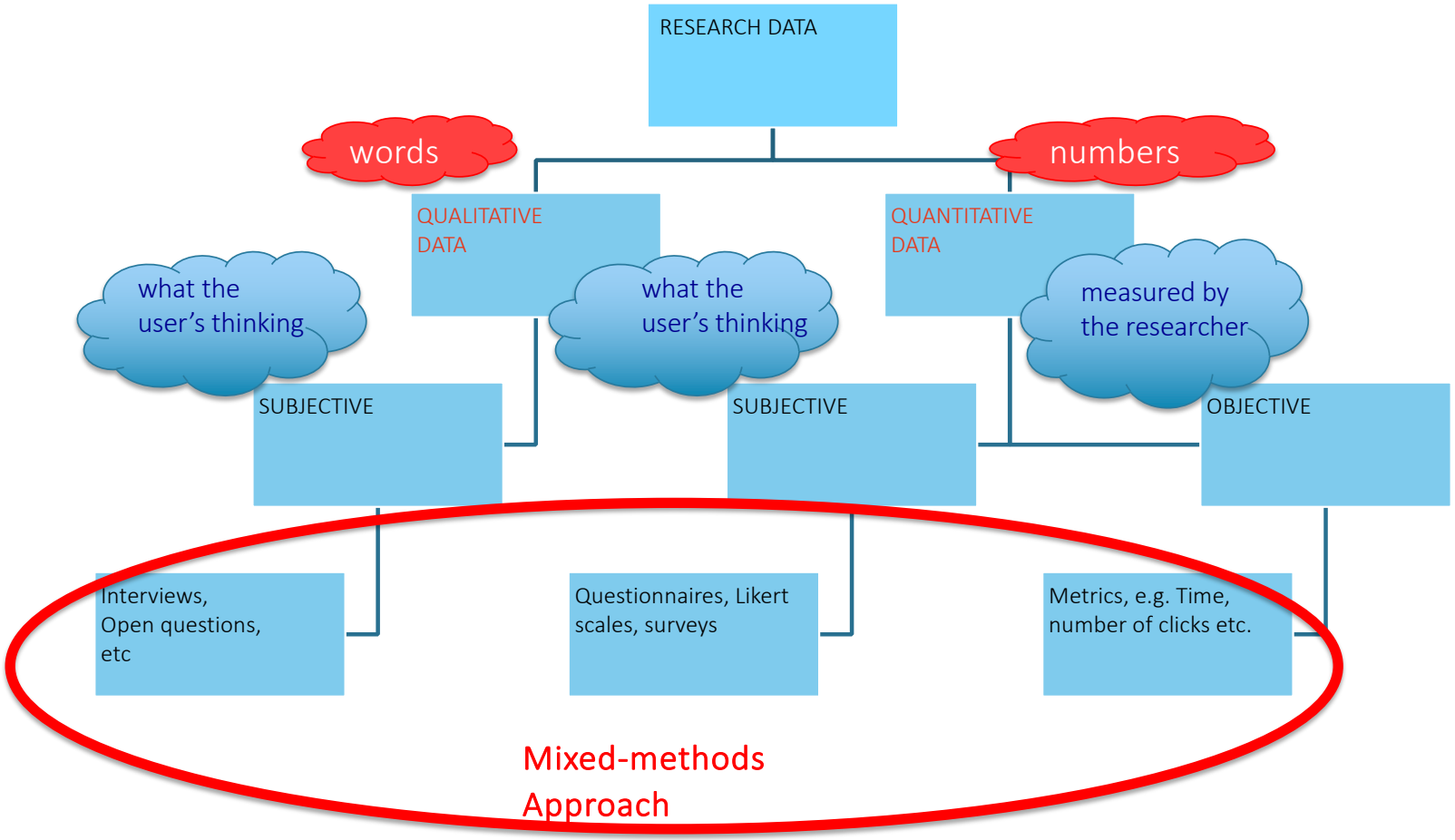


  - What about a qualitative measure?

Example



- Take (ideally) target users
- Give them a representative task and:
  - Observe them doing the task and collect metrics
  - Ask them to explain interface aspects in their own words to gauge validity of intended functions
  - Ask their opinion on the interface
    - Open text boxes
    - Likert/rating scale questions

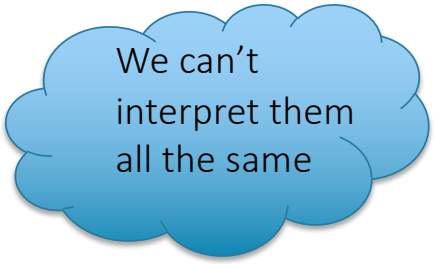

RESEARCH DATA
words
numbers
QUALITATIVE DATA
QUANTITATIVE DATA
what the user's thinking
what the user's thinking
measured by the researcher
SUBJECTIVE
SUBJECTIVE
OBJECTIVE
Interviews, Open questions, etc
Questionnaires, Likert scales, surveys
Metrics, e.g. Time, number of clicks etc.
Mixed-methods Approach

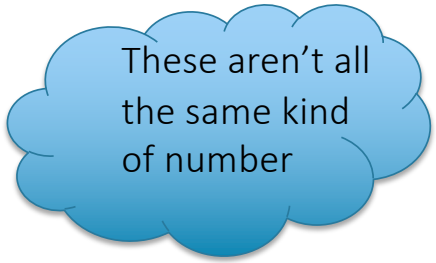
Quantitative Data

- A closer look at numbers

Different types of Quantitative data

- Numbers
  - Likert scales
  - Binary: would you use this in the future? Yes/No
  - Ranking: which one do you prefer
  - Number of clicks
  - Time on task





Different types of numbers

- Why is 3 on these scales different to counting 3 clicks on a website?

- Please rate your impression of the robot on these scales:

• Fake	1	2	3	4	5	Natural
• Machinelike	1	2	3	4	5	Humanlike
• Unconscious	1	2	3	4	5	Conscious
• Artificial	1	2	3	4	5	Lifelike

Why bother?

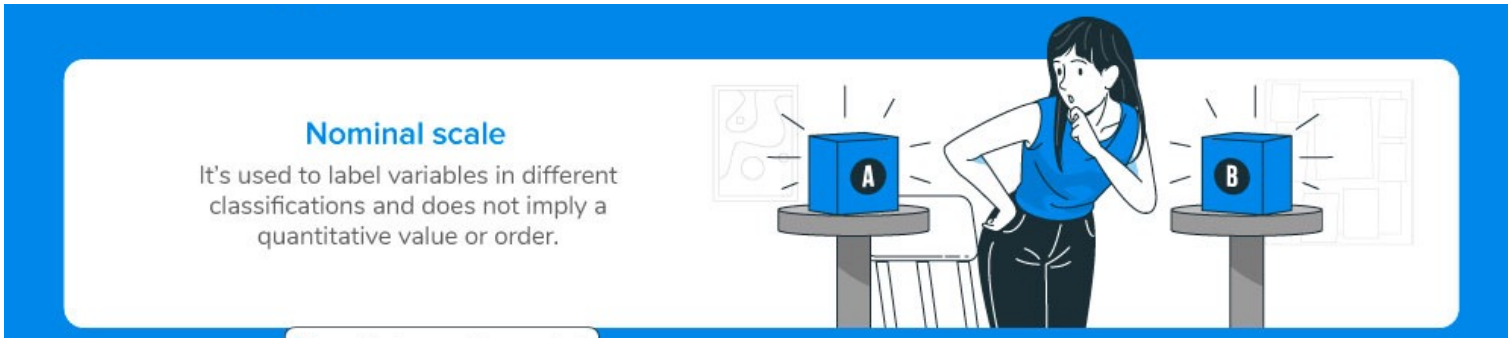
- It matters when it comes to analysing our data and doing statistical tests

Variables

- int x = 0;
  - No, not that one…. 😐
- Anything that changes:
  - e.g. Children's anxiety while watching horror movies
  - e.g. Nervousness while sitting exams
- Quantitative
  - Represented as a number
- We are interested in relationships between variables and what causes change

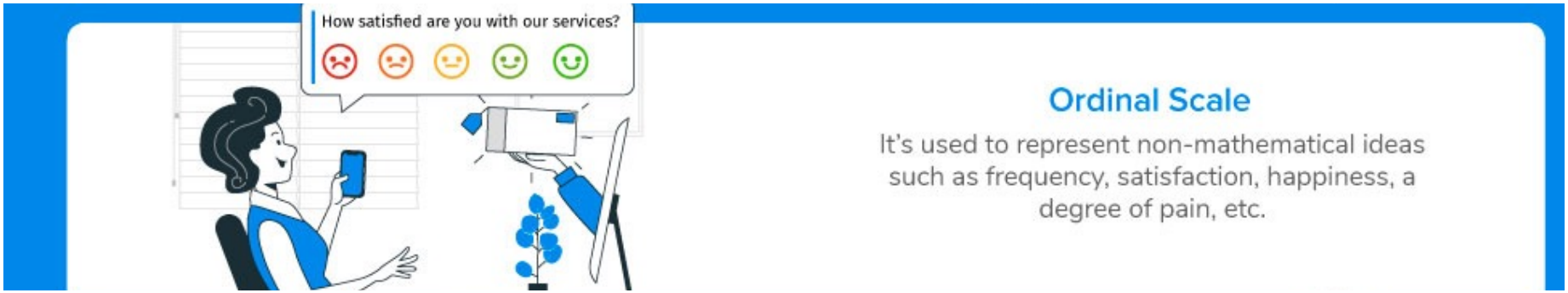
Nominal and Ordinal Data/Scale

- “Nominal" and “Ordinal” refer to different types of categorisable data.
- Nominal data involves naming or identifying data; because the word "nominal" shares a Latin root with the word "name" and has a similar sound, nominal data's function is easy to remember.
- Nominal data assigns names to each data point without placing it in some sort of order. For example, the results of a test could be each classified nominally as a "pass" or "fail”.



Nominal and Ordinal Data/Scale

- Ordinal data involves placing information into an order, and "ordinal" and "order" sound alike, making the function of ordinal data also easy to remember.
- Ordinal data groups data according to some sort of ranking system: it orders the data. For example, test results could be grouped in descending order by grade: A, B, C, D, E and F.



Nominal and Ordinal Data/Scale

- Why is this differentiation critical?

- Knowing if your data is nominal or ordinal helps you decide how to use the data; how to determine what statistical analysis to apply to a data set based on whether it is nominal or ordinal.

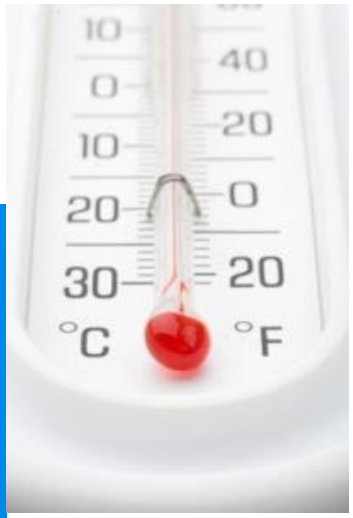
- Ways of labelling data in statistics are called "scales"; along with nominal and ordinal scales are interval and ratio scales.

Interval Scale

- Data which has been measured using a scale with ***equal intervals*** on it.
- It is a numerical scale where the order of the variables is known as well as the *difference between these variables*.
- Variables that have familiar, constant, and computable differences are classified using the Interval scale.

Interval Scale

- No pre-decided starting point or a true zero value.
- Difference between 20°C and 27°C is the same as the difference between 85°C and 92°C (both 7°C).
  - However, 40°C is not twice as hot as 20°C.
- Intervals have no absolute zero, e.g. 0°C doesn't mean "no temperature"
- Celsius is an Interval scale, Kelvin is not. Why?



Ratio Scales

- Ratio scales must have a point zero characteristic
  - But no negative values
- Timing is ratio: no-one will ever do something in a negative number of seconds
- Weight: there is zero weight but not negative
- Makes it possible to compare.

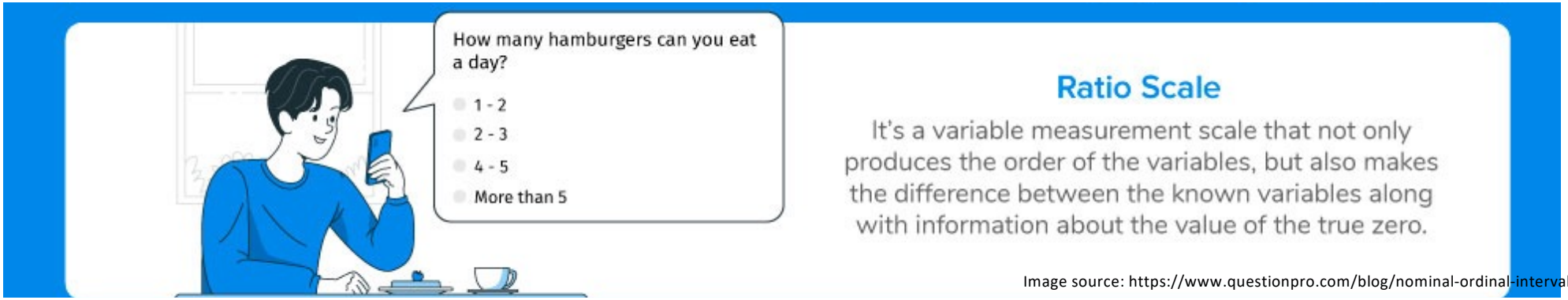


Image source: https://www.questionpro.com/blog/nominal-ordinal-interval-ratio/

Summary

- Why empirical approach is useful
- Data
  - Subjective (user reported) vs Objective (experimenter measured)
  - Quantitative (numbers) vs Qualitative (words)
- Variables
  - Nominal, ordinal, interval, ratio