HCI and Design

Topics for today

- Quantitative data
  - Instrumentation and logging
  - Quantitative surveys

- AB testing

- Activity

NO CLASS ON TUESDAY!

Quantitative vs. Qualitative

| Quantitative	Qualitative
1. Explanation through numbers	1. Explanation through words
2. Objective	2. Subjective
3. Deductive reasoning	3. Inductive reasoning
4. Predefined variables and measurement	4. Creativity, extraneous variables
5. Data collection before analysis	5. Data collection and analysis intertwined
6. Cause and effect relationships	6. Description, meaning

Instrumentation / logging

Instrument the software to record all interactions of the participant with the system

- key presses
- mouse movements
- screen touches / clicks
- reading time
- eye tracking
- etc.

The details of each interaction are recorded in a log and time stamped

Data can be analyzed to understand performance, cognitive load, user behavior, reactions to variations in task, system usage over time, and more.

Example: Log for clicking UI targets
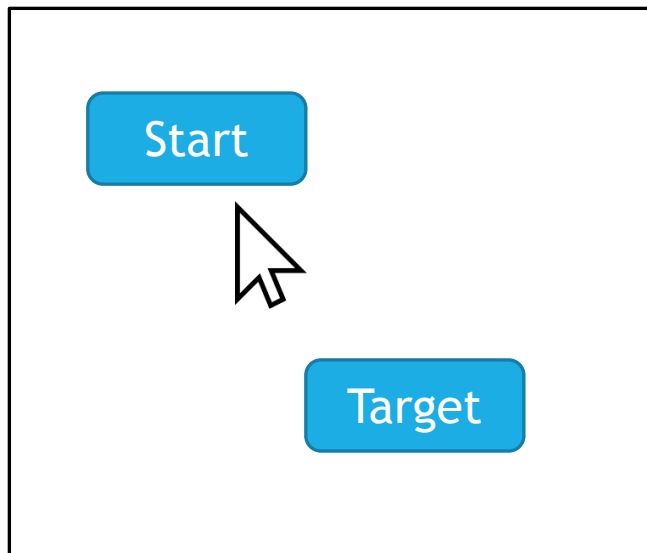

```
----------------History--------------
---Click!---
At 1352079613888: START Button is hit.
---Click!---
---Click!---
At 1352079660558: HI Button is hit.
It takes 46670 to click the button.

---Click!---
At 1352079666294: START Button is hit.
---Click!---
At 1352079666941: HI Button is hit.
It takes 647 to click the button.

--------------------Data analysis:----------------
For Button 1
Distance 228
Size 50
Time 46670 ms

For Button 6
Distance 907
Size 110
Time 647 ms
```

Instrumentation / logging

- Useful for understanding interaction.

- Useful for web studies.

- Testing layouts, displays, GUI interfaces etc.

- Understanding users' reactions to variations in tasks and layout combinations.

- Testing the efficiency of the interface and its ease of use.

- Measuring levels completed in games, time to fill out forms, time to complete tasks, etc.

- Measuring features most used, most clicked on, etc.
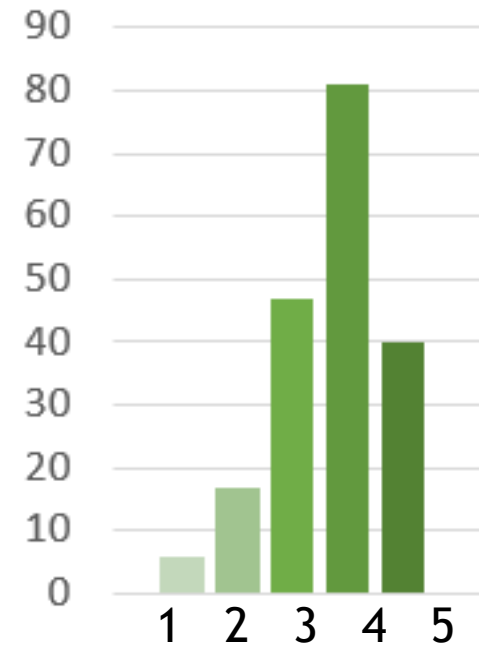
- Measuring speed and performance, etc.

Quantitative surveys

Numeric survey responses (e.g., how many times you bought a product)

Likert scale responses
◦ Odd (neutrality allowed) or even (forced choice) number of choices allowed

Statistics calculated for each question
◦ Mean, median, standard deviation, N-per-answer, etc.

Descriptive Statistics

	Household income in thousands
Mean	59.59
Std. Deviation	67.130
N	1500
Median	40.00
Minimum	12
Maximum	1,079

Topics for today

- Quantitative data
  - Instrumentation and logging
  - Quantitative surveys

- AB testing

- Activity

What is AB Testing?

A practical way to test different designs ☺️

Also called "split testing" or "bucket testing"

A/B testing splits live traffic into two (or more) parts:
◦ Some users see the standard design ("A")
◦ Some users see an alternative design ("B")
◦ After collecting enough data, the one with the better performance "wins"

Multivariate testing
◦ More than 2 variations
◦ A/B/C/D...

What is AB Testing?

Empirical, data-driven method
- Real site, authentic users

Controlled experiment
- Split users into (at least) two groups: Control and test(s)

Cookies can ensure an individual gets consistent version.

A metric
- Overall evaluation criteria (OEC) or key performance indicator (KPI)
- e.g., click-through rate, purchases

50 % visitors see variation A → Variation A → 23% conversion

50 % visitors see variation B → Variation B → 11% conversion
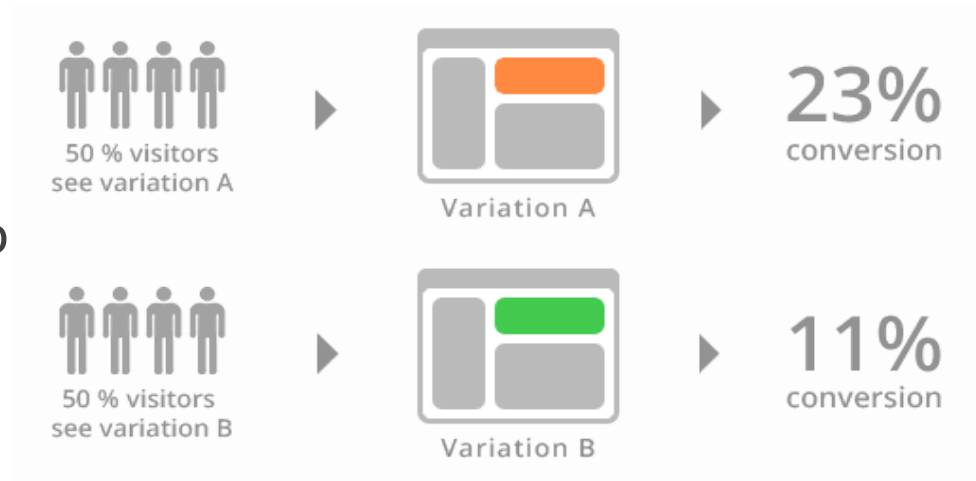
Process

Come up with a driving question:
- e.g. Can I increase click through rates from my home page

Formulate a hypothesis:
- e.g. Underlined links will outperform my current links

Calculate time to run test:
- Tools available to help you do this, e.g. [A/B Test Duration Calculator](#).

Run the test:
- X% of users get underlined links
- How to decide X?

Review results

Some examples…

Example 1: Sponsor a Child website

https://whichtestwon.com/test/with-sliders-or-without/

Key Performance Indicator (KPI):
- clicks on the sponsorship "Call To Action" button

Difference between versions:
- Version A: Sponsor a child page formatted with sliders
- Version B: Radically redesigned page without sliders

A

B

And the winner is…

Version B
- ◦ 60.4% increase in clicks on the sponsorship button!
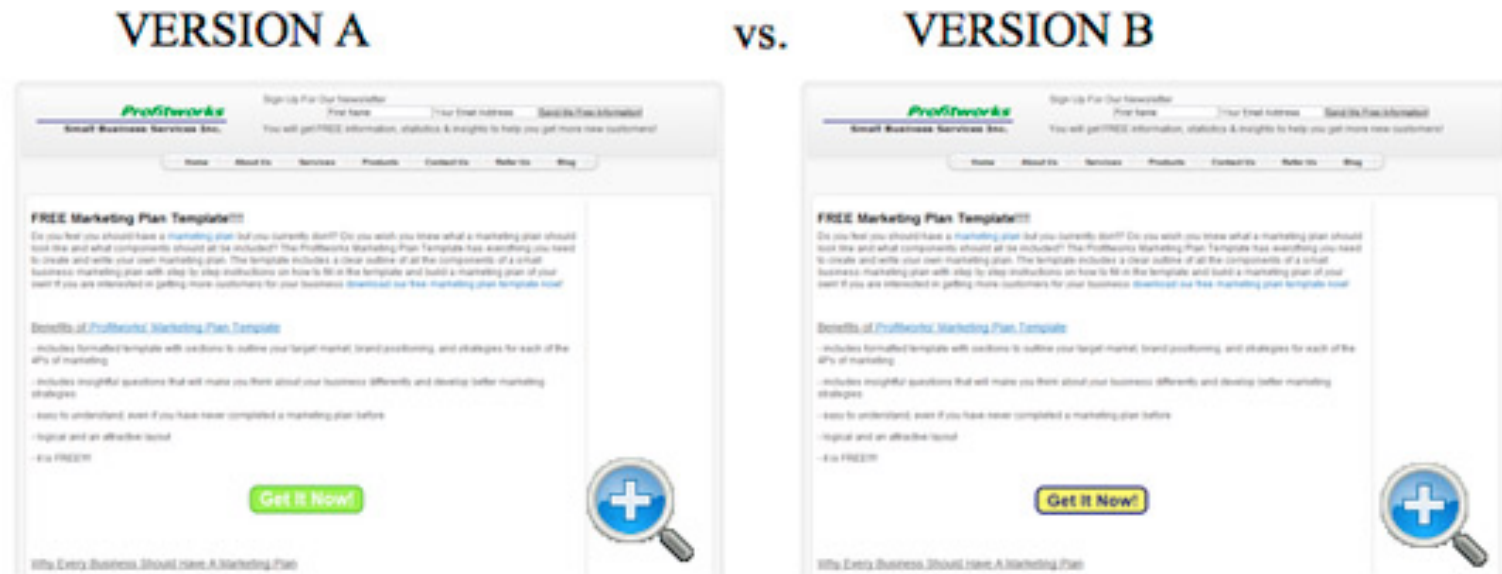
Used:
- ◦ Optimizely
- ◦ 14 days 8,600 visitors.

Post-hoc rationalization
- ◦ story narrative
- ◦ updated visual flow
- ◦ clickable colored boxes were interactive

Example 2: Button Color Test Which Page Drove More Clicks?

Green is an affirming color that signifies positive action. BUT… it's been used with white text which completely washes the button out.

Version B's yellow and black button may be ugly but it is clear and led to a 14.5% increase in conversions."

Example 3: Final Page in a 4 Step Sequence

A

B

Example 3: Final Page in a 4 Step Sequence

Both pages are well-designed and both, seemingly, doing a good job

Version B
◦ video center piece, well-structured; it should convert

Version A had 439% higher performance.

Post-hoc rationalization
◦ text within the image brings focus
◦ form has fewer fields and looks shorter
◦ images give the page a more authentic and trustworthy feel

Benefits of AB testing

Compared with other methods, A/B testing has a few huge benefits:

1. It measures the actual behavior of customers in real-world conditions.

2. It can measure very small performance differences with high statistical significance because you can throw tons of traffic at each design.

3. It can resolve trade-offs between conflicting designs, guidelines, or qualitative usability findings.

4. It's cheap: once you've created the design alternatives you simply put both of them on the server and employ a tiny bit of software to randomly serve each new user one version or the other.

5. It's low risk: if one of the alternatives performs badly, you can simply turn it off or try again.

Limitations of AB testing

1. A/B testing can only be used for projects that have one clear, all-important goal (KPI).

2. The goal must be measurable by computer through user actions.

3. A/B testing provides data only on the element you're testing.

4. A/B testing only works for fully implemented designs. It's cheap to test a design once it's up and running, but implementation can take a long time.

5. Encourages a focus on short-term (and short-sighted) improvements.

6. No Behavioral Insights: you don't necessarily know *why* you get the measured results.

Limitations of AB testing

Final Page in a 4 Step Sequence

Post-hoc rationalization: text within the image brings focus, form has fewer fields, looks shorter, images give the page a more authentic/trustworthy feel.
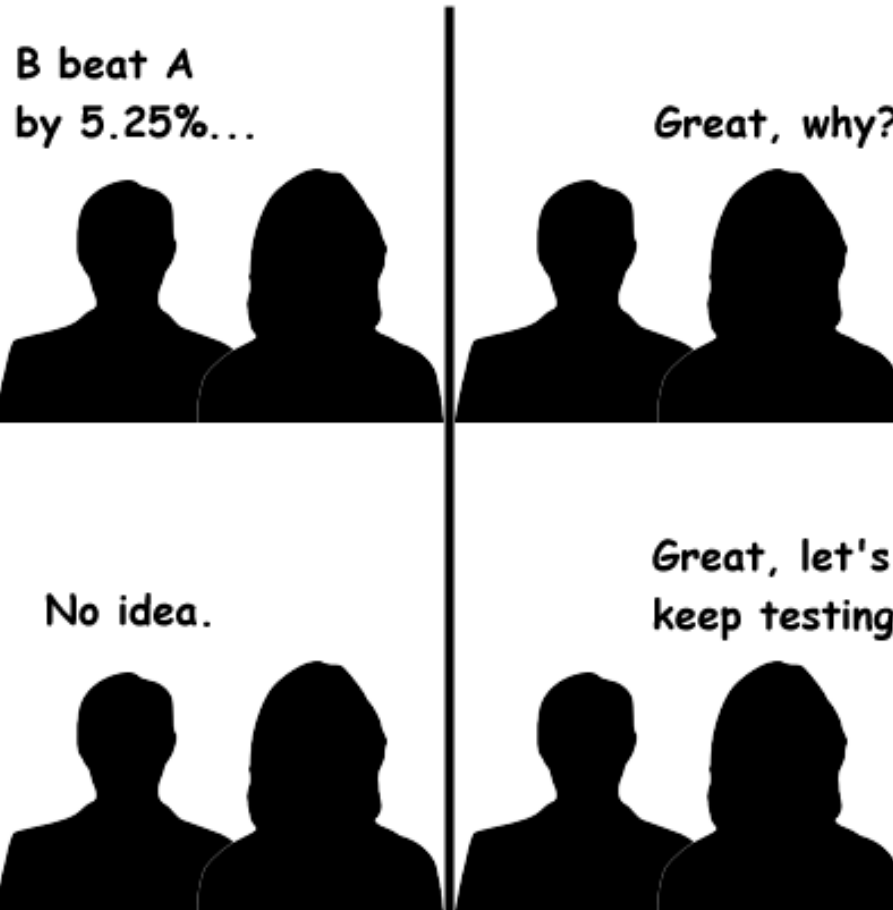
Don't actually know why!

Take-aways

A/B testing can be useful and it definitely has its place
- Easy, cheap, fast, can lead to significant results.

Don't EVER rely solely on A/B testing

It's ALWAYS a good idea to combine methods
- Qualitative methods
- Other quantitative methods
  - Logging / instrumentation
- More…

Resources for doing AB testing

There are dozens of A/B consultants, tools, tips and tricks. Here are a few links:

Robin Johnson, writing in Optimizely, gives you 71 different ideas on things to test, including website copy, visual elements and just plain common sense things such as whether you have a positive or negative spin on what you are trying to sell.

Uri Bar-Joseph, writing in SearchEngineWatch, has eight rules for A/B testing, including focusing on one variable, choosing your groups randomly and measuring the results carefully.

Visual Website Optimizer (vwo.com) is testing software for marketers (i.e. if you don't want to do it yourself).

For some fun, take a look at these 12 results that were somewhat counter-intuitive from WhichTestWon.com, a commercial A/B testing provider.

Activity

1. Work in pairs.

2. Create an A/B experiment for improving a commercial product of your choice (existing app or website)
   ◦ Come up with a driving question – write it down
   ◦ Formulate a hypothesis – write it down
   ◦ Design an experiment to test your hypothesis (e.g. make different variants, define your "key performance indicator")
   ◦ Explain your choices: why did you pick these variants?
   ◦ How long would you run your experiment for and why?
   ◦ What results would you hope to see?

3. (If you're done)

   Discuss your AB experiment another team. Critique!