CS 8803-MDS
Human-in-the-loop Data Analytics

Lecture 4

08/31/22

Logistics

Paper signup reminder

     check your credits

Project reminder

     09/09 finalize groups

        team of 1-3, email instructor & TA

     09/21 project proposal

        more details next class

        goal: first draft of your paper introduction

COVID reminder

Today's class

[BlinkDB: Queries with Bounded Errors and Bounded Response Times on Very Large Data](#)

Author: Eric

Reviewer: Sahil

Measuring user experiences (continued)

Help us learn your names!

EuroSys 2013

APRIL 14-17, 2013
PRAGUE, CZECH REPUBLIC

Prague — St Vitus Cathedral, Prague Castle and Charles Bridge

PAPERS

MONDAY, APRIL 15, 2013

08:45 – 09:00

OPENING

09:00 – 10:30

LARGE SCALE DISTRIBUTED COMPUTATION I

- **TimeStream: Reliable Stream Computation in the Cloud**

  Zhengping Qian (Microsoft Research Asia), Yong He (South China University of Technology), Chunzhi Su, Zhuojie Wu, and Hongyu Zhu (Shanghai Jiaotong University), Taizhi Zhang (Peking University), Lidong Zhou (Microsoft Research Asia), Yuan Yu (Microsoft Research Silicon Valley), and Zheng Zhang (Microsoft Research Asia)

- **Optimus: A Dynamic Rewriting Framework for Execution Plans of Data-Parallel Computation**

  Qifa Ke, Michael Isard, and Yuan Yu (Microsoft Research Silicon Valley)

- **BlinkDB: Queries with Bounded Errors and Bounded Response Times on Very Large Data**

  Sameer Agarwal (University of California, Berkeley), Barzan Mozafari (Massachusetts Institute of Technology), Aurojit Panda (University of California, Berkeley), Henry Milner (University of California, Berkeley), Samuel Madden (Massachusetts Institute of Technology), and Ion Stoica (University of California, Berkeley)

ORGANISATION

EuroSys
ACM SIGOPS in Europe

ČVUT

TECHNISCHE
UNIVERSITÄT
DRESDEN

WE THANK OUR SUPPORTERS

4

Exercise: come up with a missing experiment

Think about a plot that you'd like the authors to produce

?

?

What you said about BlinkDB

- BlinkDB supports only aggregates involving COUNT, AVG, SUM and QUANTILE. This seems quite limited. Joins, nested queries and UDFs don't seem to be supported.

- How can this be leveraged for non-aggregation queries, for instance, limit the first fetch to 100 entries and then if the user needs more then let me query again.

What you said about BlinkDB

- BlinkDB creates multiple sets of stratified samples of the same dataset. This results in a significant storage overhead.

- It is well-known that it is hard to maintain sample for a large number of columns (curse of dimensionality). I am curious whether it is possible to maintain some stratified samples in a high-dimensional database, that is both representative and has a small size.

- Trade-offs between efficiency and the generality of the queries they support are the main areas for improvements.

- Results of the paper revealed that for a 7.5 TB dataset, the query times were much longer due to disk spillage.

- Scalability of the mixed integer linear program

What you said about BlinkDB

- BlinkDB assumes that the QCS is stable over time, it does not perform good for queries whose QCS is not covered by the query workload.

- How does the system address data consistency, i.e. how the samples will be updated as the data changes?

- How these samples may be updated overtime if the queries by user are not stationary over time.

- One "bad" sample may affect multiple queries that use that particular sample

What you said about BlinkDB

- Lack of clarity on how parameters M and K which determine the capacity for subgroups across a QCS are determined.

- Some heuristically-chosen parameters in optimizing stratified samples might be replaced by more properly learned or computed ones

- The sample selection process is not automated, so the user would have to manually create the sample set by explicitly specifying the sample ratio.

What you said about BlinkDB

- Online aggregation is one such approach and would allow BlinkDB to continuously refine the value of aggregates.

- Implementing synopses along with BlinkDB and incoporating the storage requirement of synopses into optimization formulation could be a potential improvement.

BlinkDB's Legacy

Best paper award in Eurosys 2013

    800+ citation

AQP adoption in industry:

    Facecook's Presto, Infobright's IAQ

    Ideas introduced by BlinkDB was commercialized as part of SnappyData (acquired by TIBCO)

    VerditDB: further addresses sample construction cost

Measuring the user experience

Why is this relevant?

What is user experience:

    a user is involved

    the user is interacting with a product, system, or an interface

    the users' experience is of interest, and observable/measurable

How to systematically measure and evaluate user experience?

    effectiveness, efficiency, satisfaction

Why is this relevant?

Papers with user studies:

Database Benchmarking for Supporting Real-Time Interactive Querying of Large Data

Towards Effective Foraging by Data Scientists to Find Past Analysis Choices

Lux: Always-on Visualization Recommendations for Exploratory Dataframe Workflows

Expressive Time Series Querying with Hand-Drawn Scale-Free Sketches

How to measure

Self-report
  reported directly by a participant
  subject to many biases
  common, useful and easy to administer
Behavioral
  measured through observation
  more objective representation of events
Physiological
  heart rate, skin conductance, etc.

Self-report measures

Questionnaire:

      forced choices: yes/no

      scale responses

      open-ended responses

Designing questionnaires is an art in itself

      Odd or even number of values

      How many points on a scale?

Use validated, commonly used self-report questionnaires:
      [System Usability Scale (SUS)](#)

Likert Scales

Please circle the number that represents how you feel about the computer software you have been using

I am satisfied with it
Strongly Disagree ---1---2---3---4---5---6---7--- Strongly Agree

It is simple to use
Strongly Disagree ---1---2---3---4---5---6---7--- Strongly Agree

It is fun to use
Strongly Disagree ---1---2---3---4---5---6---7--- Strongly Agree

It does everything I would expect it to do
Strongly Disagree ---1---2---3---4---5---6---7--- Strongly Agree

I don't notice any inconsistencies as I use it
Strongly Disagree ---1---2---3---4---5---6---7--- Strongly Agree

It is very user friendly
Strongly Disagree ---1---2---3---4---5---6---7--- Strongly Agree

When to collect self-reported data

the think-aloud protocol
    while the participants are interacting with the product

post-task ratings
    immediately after each task
    help pinpoint problematic tasks and interfaces

post-study ratings
    at the end of the entire session
    overall evaluation

Biases in self-reported data

Social desirability bias

      telling you the answer you want to hear

How to mitigate

      leaving the room when the user fills out the survey

      making the survey itself anonymous

Behavioral Measures

less obtrusive to participants, often more quantitative in nature

Direct Observation

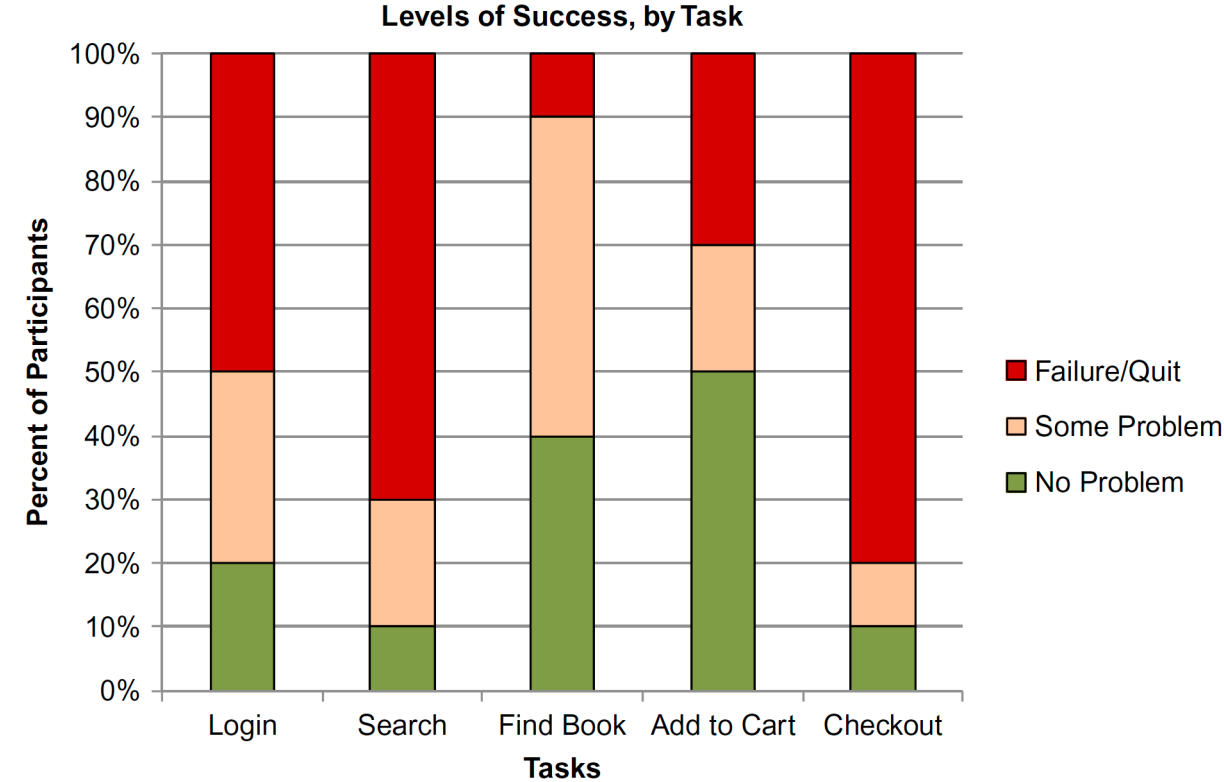   researcher observes and records participant activity

Indirect Observation

     video recorded and analyzed later

     participant records activity (e.g., diary entries)

How to measure: behavioral
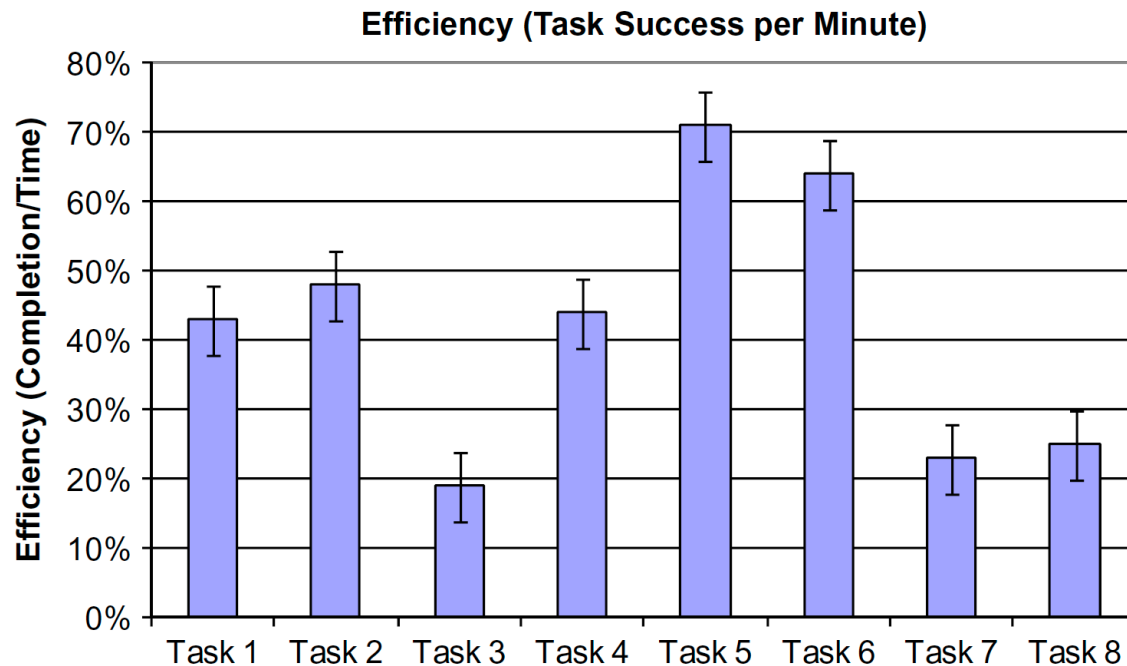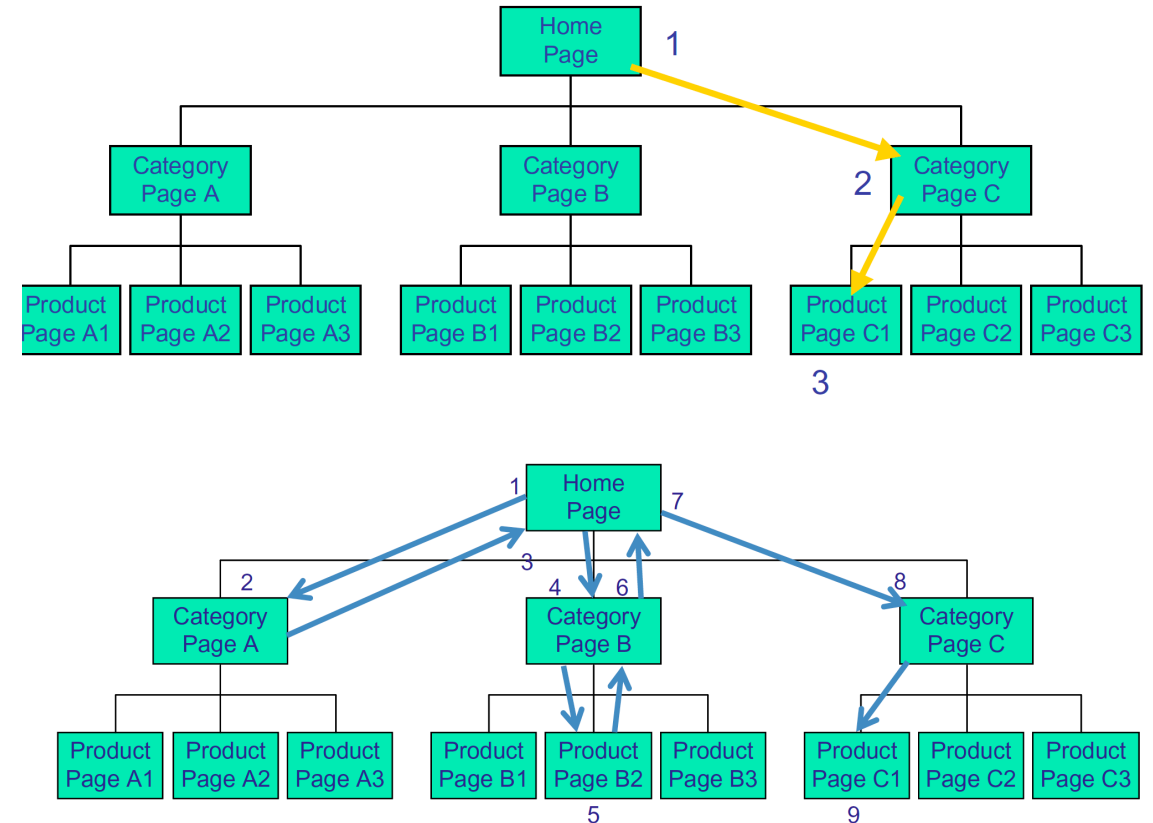

Percent Correct, by Task
(Error bars represent the 90% confidence interval)

Levels of Success, by Task
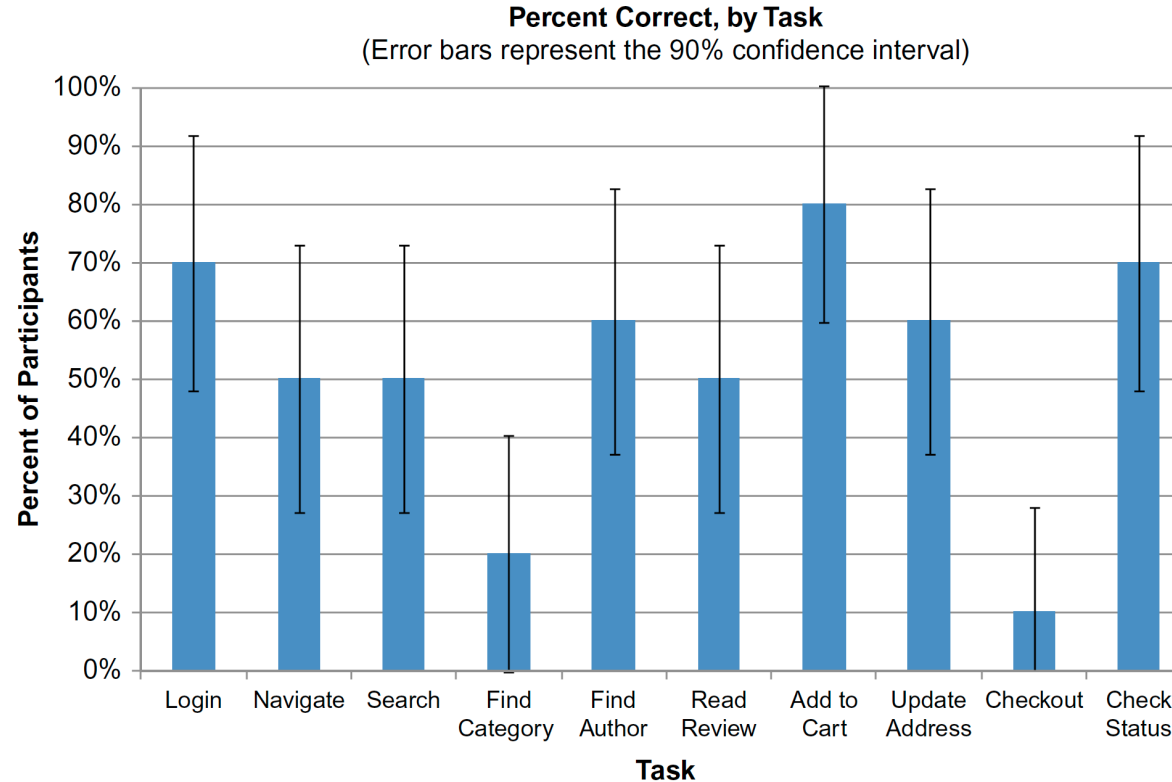
Binary success

Levels of success

Source: Albert, Bill, and Tom Tullis. *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Newnes, 2013.
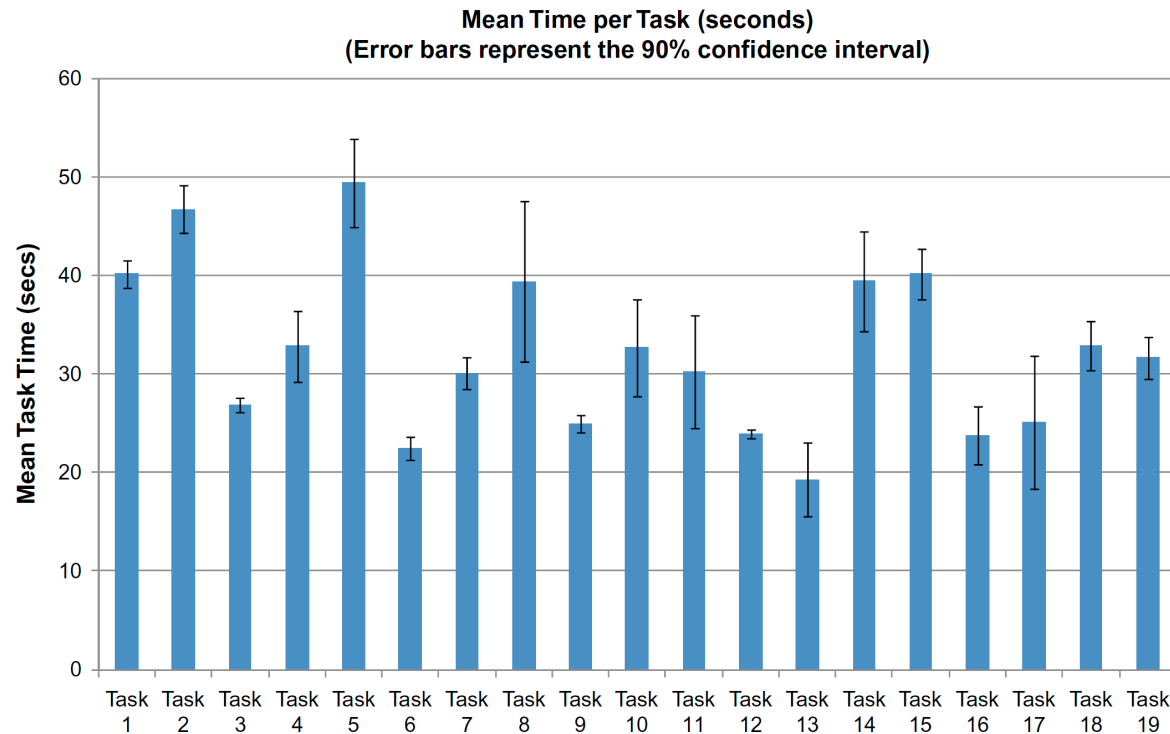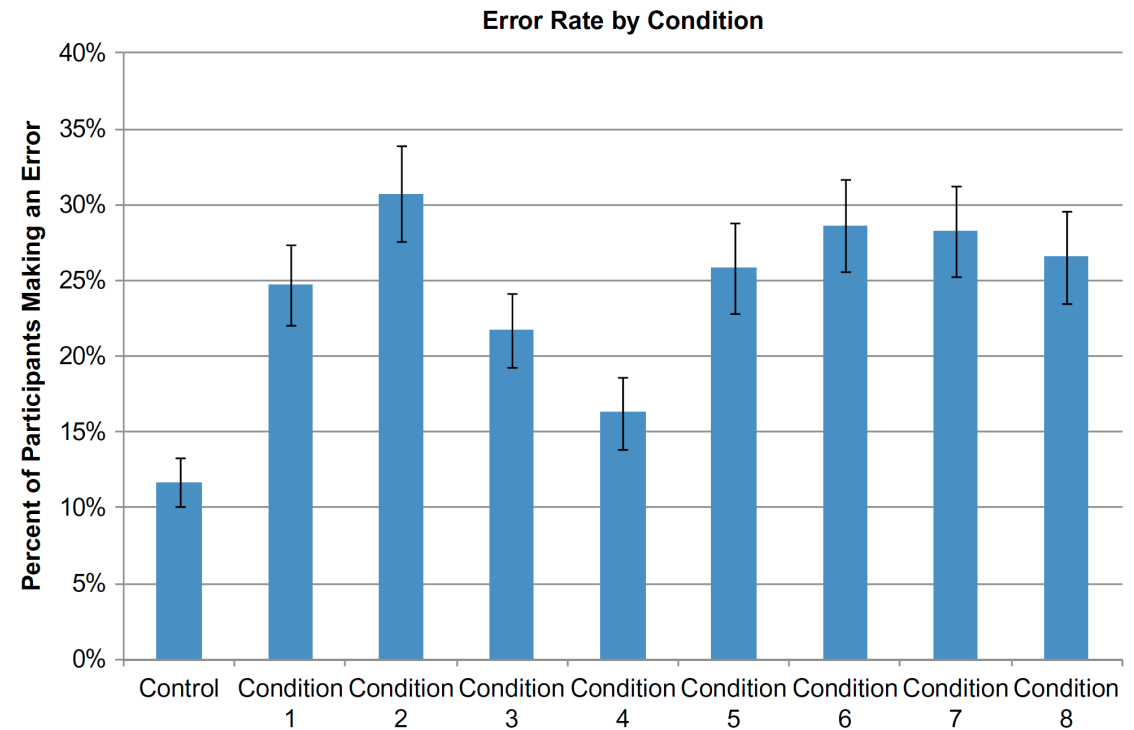
How to measure: behavioral

Efficiency

actions

Source: Albert, Bill, and Tom Tullis. *Measuring the user experience: collecting, analyzing, and presenting usability metrics*. Newnes, 2013.

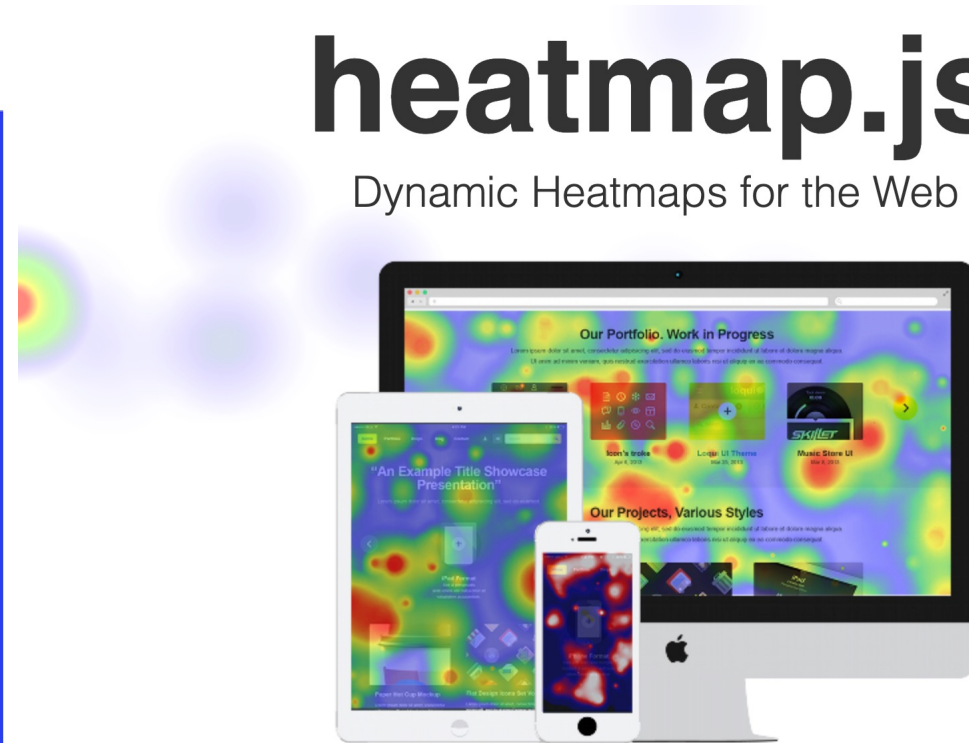# How to measure: behavioral

Time on Task

Errors

How to measure: behavioral

Engagement patterns

heatmap.js is a lightweight, easy to use JavaScript library to help you visualize your three dimensional data!

https://www.patrick-wied.at/static/heatmapjs/

IRB: Institutional Review Board

https://oria.gatech.edu/irb

Experiments conducted at universities require ethical oversight

Consider: risk to participants, data privacy etc.

Protocols must be reviewed and approved by IRB

Study might quality for Exempt Review (but still need to apply)

Your tasks for next class

NO class on Monday!

[AQP++: Connecting Approximate Query Processing With Aggregate Precomputation for Interactive Analytics](#)
        Authors: Gaurav, Sankalp
        Reviewer: Cangdi
        Practitioner: Aniruddha

Project reminder
        09/09 finalize groups
        office hours 11-12 Thursday and Friday