Segmentifier: Interactively Refining Clickstream Data into Actionable Segments

Kimberly Dextras-Romagnino
Supervisor: Tamara Munzner
University of British Columbia
April 18th, 2018

Introduction: *E-commerce*

Collaborators

- Build mobile apps for large e-commerce companies
- Understand the importance of good websites on revenue

E-commerce Goals

- **Increase traffic**
  - number of users on a site
- **Reduce abandonment**
  - number of users leaving the site
- **Increase consumer engagement**
  - time users spend on the site
  - chances that a user returns to the site
- **Increase conversion rate**
  - odds a user purchases

How did we help them solve these goals?

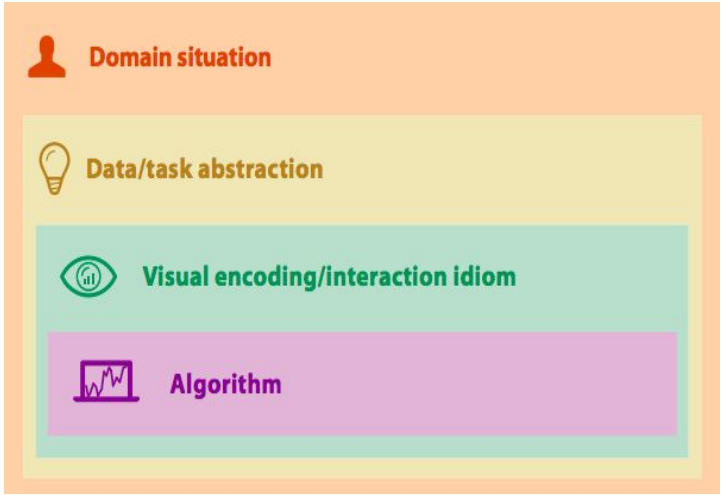
Research Method

Contributions

Research Method: Applied Visualisation Research

Nested Model for Visualization Design and Evaluation:
[Munzner 2009]

- **Domain Situation:**
  - Characterize domain problem
- **Data/task abstraction:**
  - Translate problem into abstract data types and user tasks
- **Visual encoding/Interaction Idiom:**
  - Design visual encoding and interaction techniques that map to abstract data and task
- **Algorithm:**
  - Create algorithms to execute techniques efficiently



Research Method: Design Study Methodology

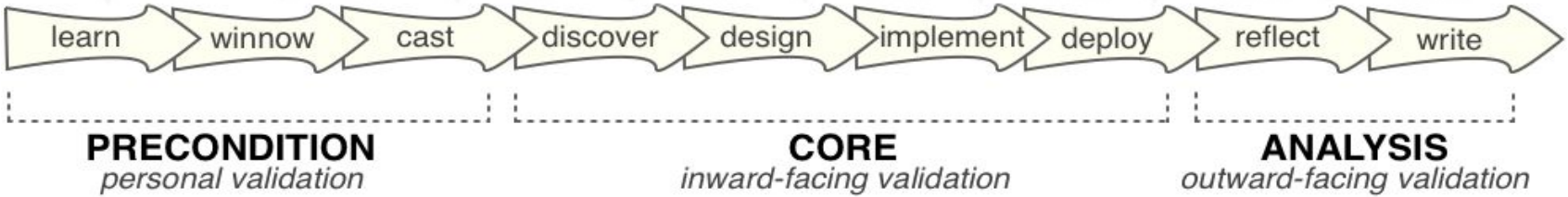
Design Study:

- A project
- **Specific real-world problem**
- **Design** a visualization system
- **Validate** the design
- **Reflect** about lessons learned

9 stage Framework

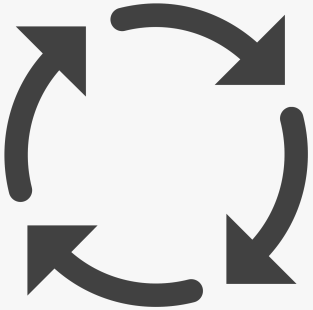
[Sedlmair et al. 2012]

- **Precondition:** Learn, Winnow, Cast
- **Core:** Discover, Design, Implement, Deploy
- **Analysis:** Reflect, Write



Research Method: Mobify

- **Pre-condition** Phase
  - Period of 5 months
  - Met with 12 employees
- **Core** Phase
  - Data and Task Abstraction
  - Design interface
  - Implement interface
- **Analysis** Phase
  - Formulate Framework
  - Write Paper/Thesis



Research Method: Contributions

- ➢ Thorough **characterization of task and data abstraction** for clickstream data analysis
  - **Clickstream Segment Analysis Framework** abstracts iterative process
    - View, Refine, Record, Export, Abandon, Conclude

- ➢ **Segmentifier: novel analytics interface** for refining data segments and viewing characteristics before downstream fine-grained analysis
  - Rich set of views showing both *derived attributes* and *raw sequence details*
  - *Filtering* and *Partitioning* through visual queries
    - Quantitative attributes
    - Custom sequences of events aggregated according to a novel three-level hierarchy
  - Detailed glyph based *visual history* of the automatically recorded refinement process showing the provenance of each segment in terms of its analysis path

- ➢ Preliminary **evidence of utility** from:
  - *Usage Scenario* with real world data
  - *Case Study* with industry analyst

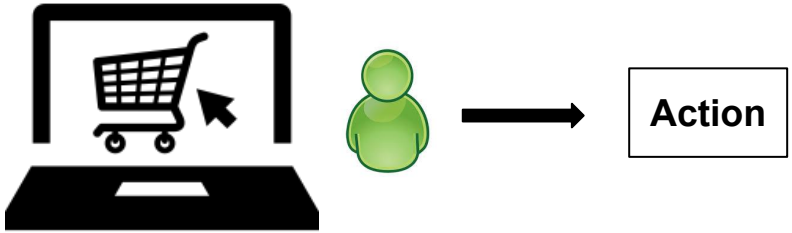
What are the **Data and Task Abstractions** for *Clickstream Data Analysis?*

Clickstream Data

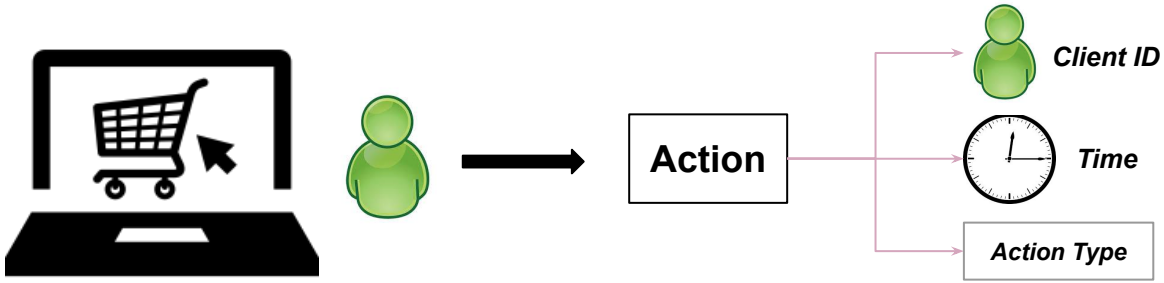
Clickstream Analysis Tasks

What is ***Clickstream Data***?

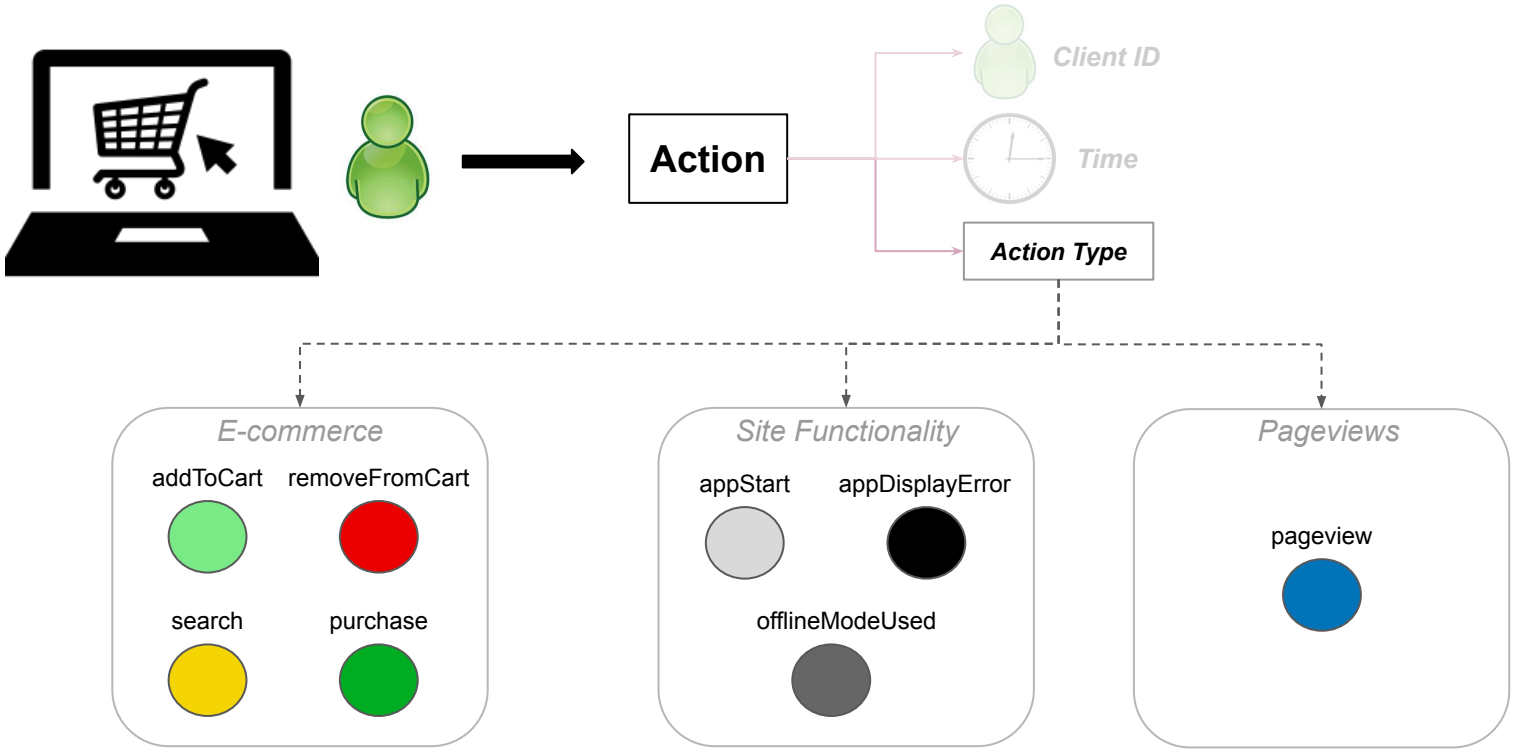
Data: *Actions*



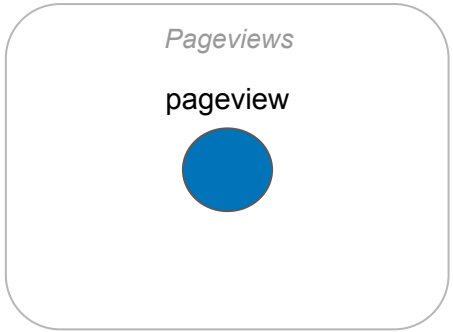
Data: *Action Attributes*



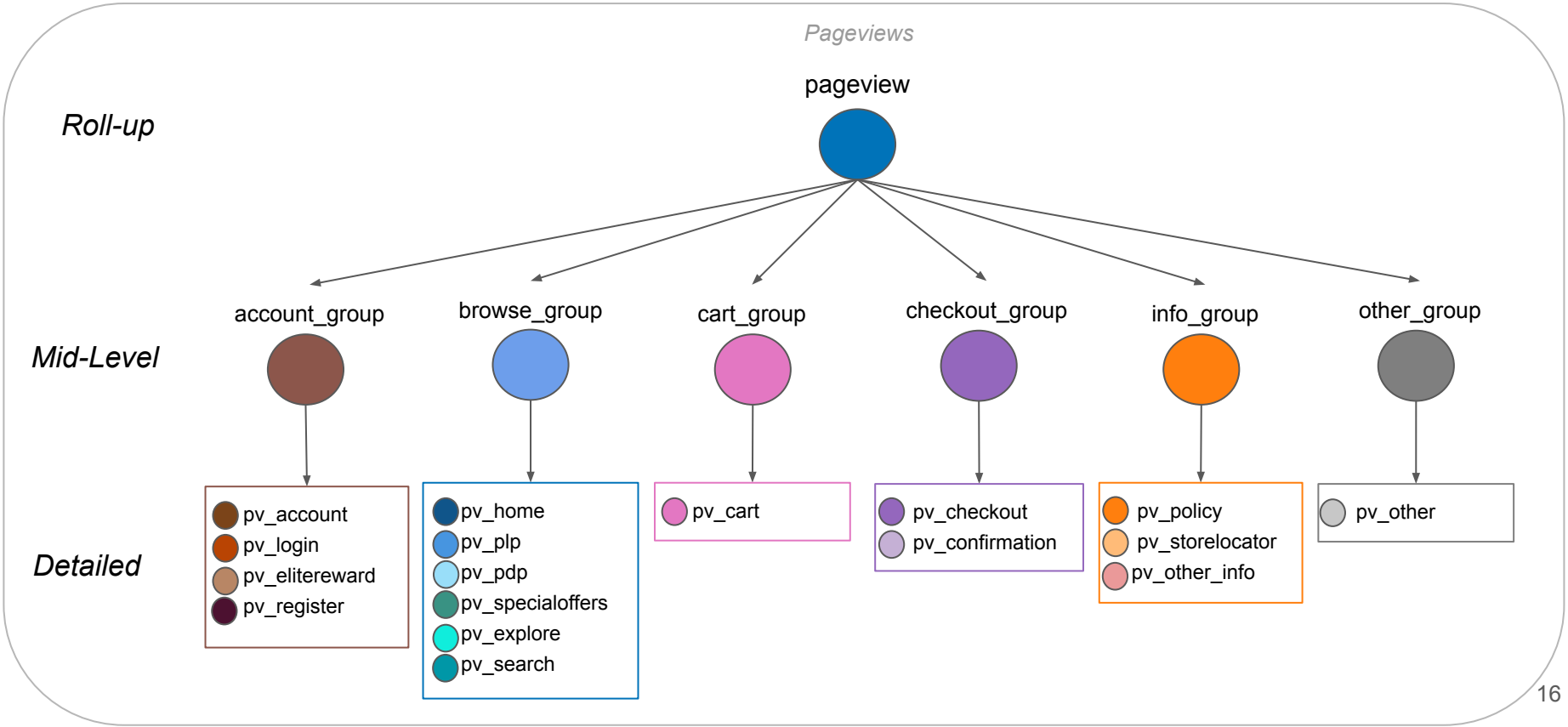
Data: *Action Types*



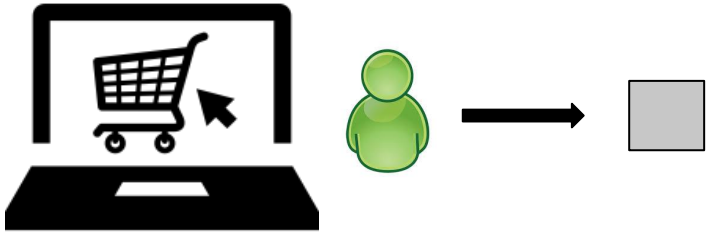
Action Hierarchy

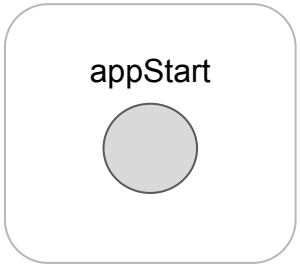


Action Hierarchy

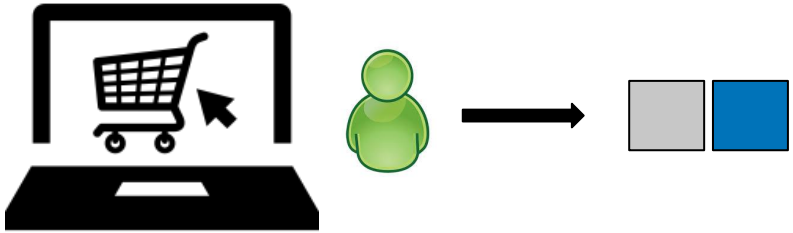


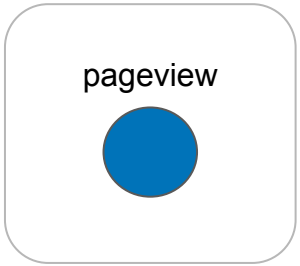
Data: *Sequences*



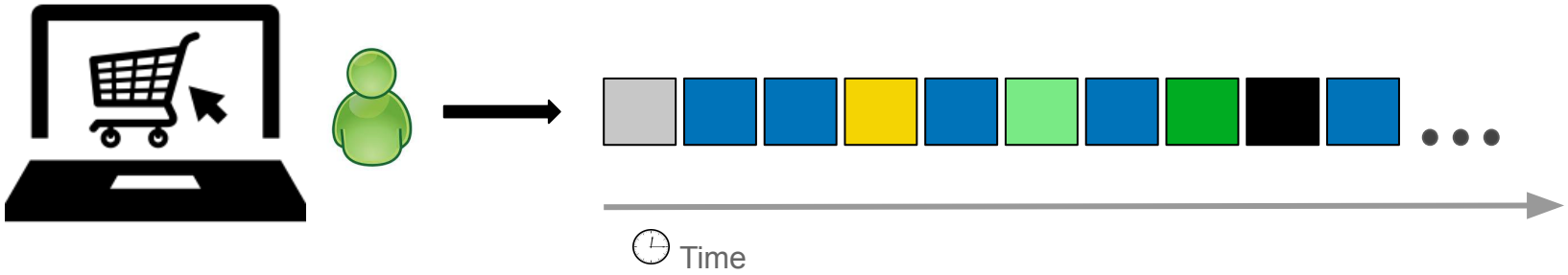


Data: *Sequences*



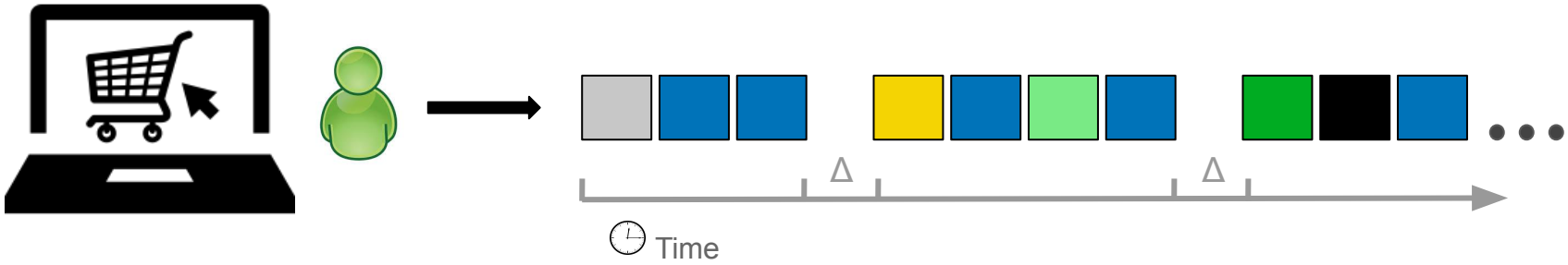


Data: *Client Sequences*



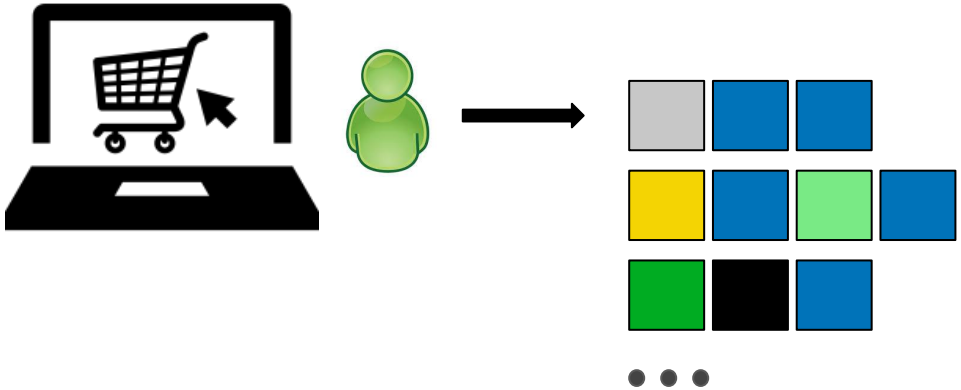
***Client Sequences*:** all actions performed by a single user

Data: *Session Sequences*



***Session Sequences*:** all actions performed by a single user within a defined amount of time (Δ) from each other. Δ is usually 30 min.

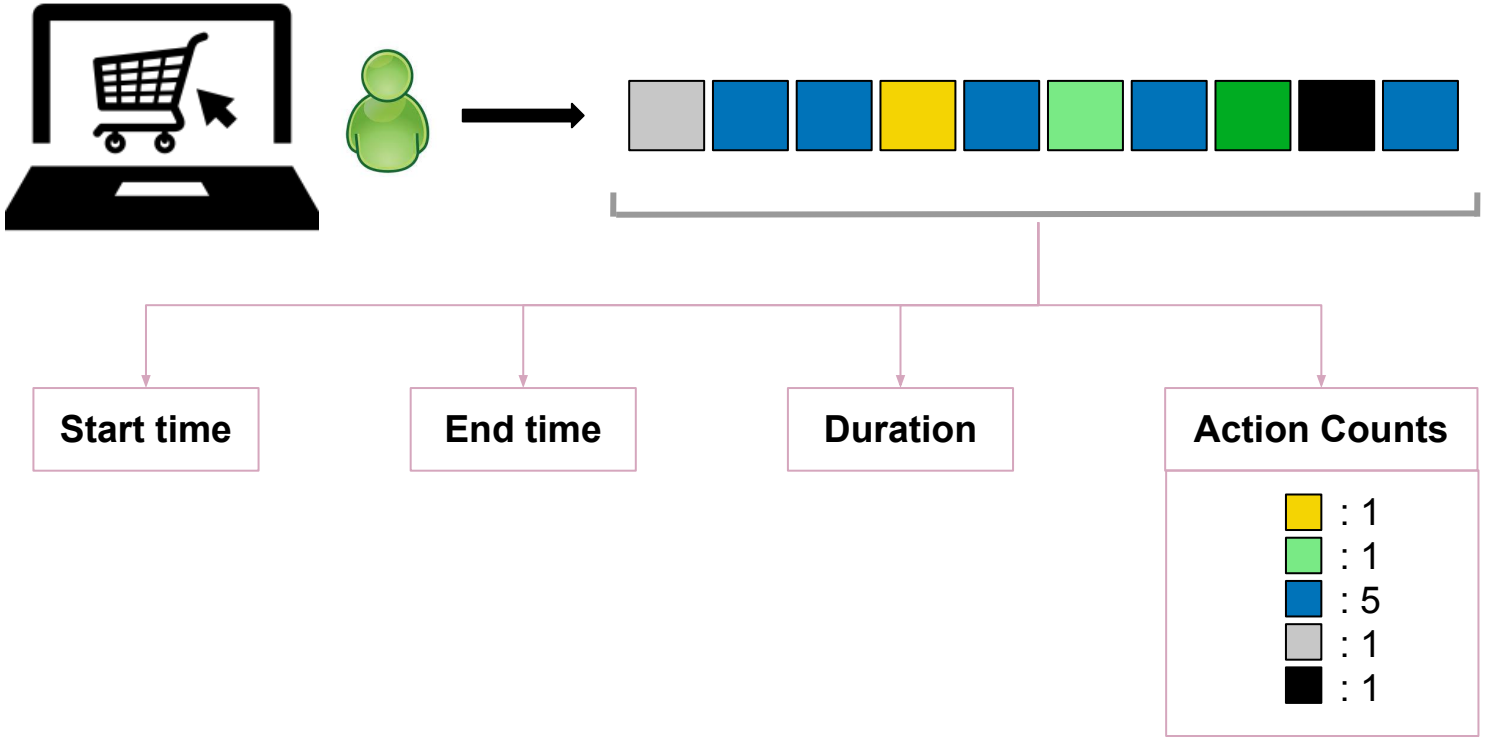
Data: *Session Sequences*



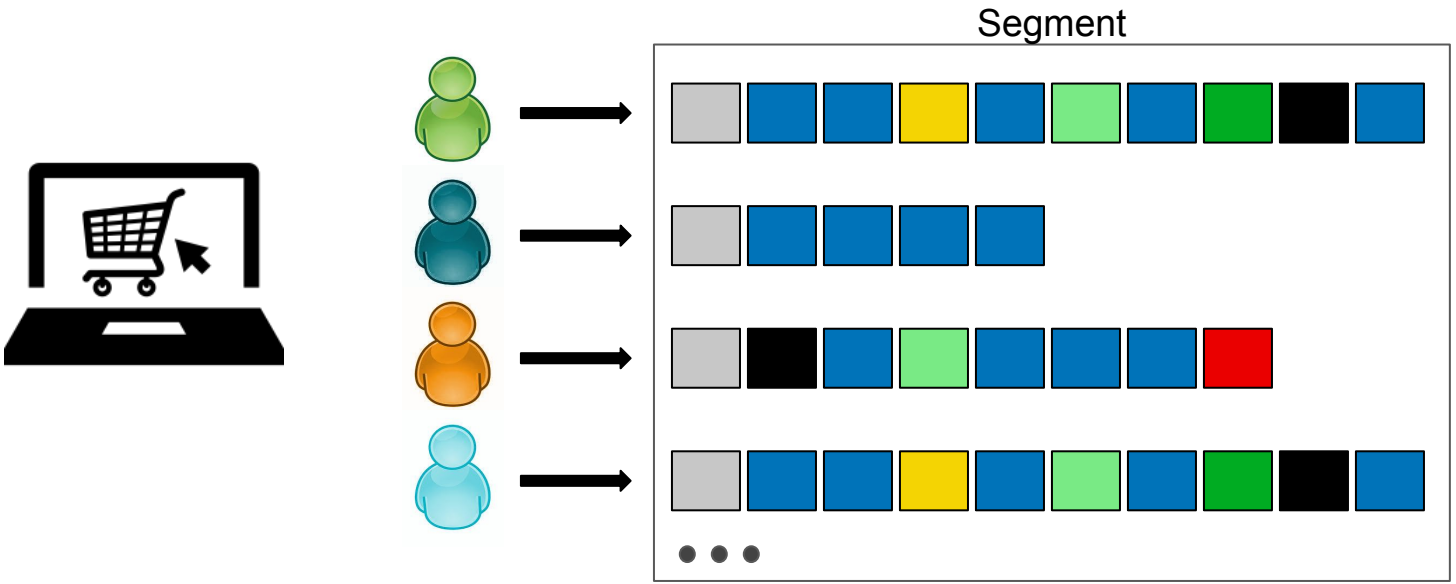
Multiple Sequences
per user

***Session Sequences*:** all actions performed by a single user within a defined amount of time (Δ) from each other. Δ is usually 30 min.

Data: *Sequence Attributes*

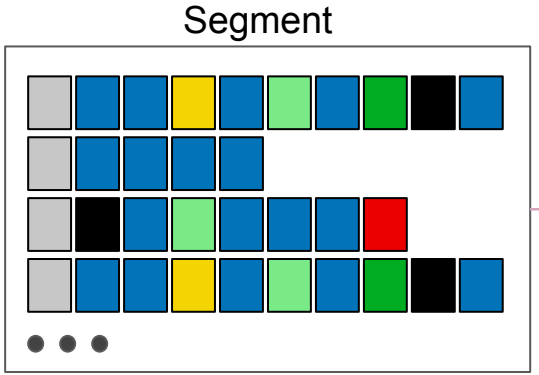


Data: *Segments*



***Segment*:** any set of sequences

Data: *Segment Attributes*



Size

Counts of sequences:
Absolute, Relative

Sequence Related

Sequence Distributions:
Start Time, Duration, Action Counts

Action Related

Action Distributions:
Action Transitions:
action before, action after

What are
Clickstream Data Analysis Tasks?

Tasks: Actionable Results

Actionable Result: result or insight found through analysis that can be acted on

Result ⇒ *Action*

Actionable Results

Identify successful trends	⇒	Optimize
Identify problems	⇒	Fix/Improve
Identify groups of common behavior	⇒	Personalize experience
Identify site metrics/benchmarks	⇒	Keep track of state of website

Tasks: Actionable Results

Actionable Result: result or insight found through analysis that can be acted on

Result ⇒ *Action*

Actionable Results

Identify successful trends	⇒	**Optimize**
Identify problems	⇒	Fix/Improve
Identify groups of common behavior	⇒	Personalize experience
Identify site metrics/benchmarks	⇒	Keep track of state of website

Domain-Specific Questions

How many users purchase? What path did they choose?

Tasks: Actionable Results

Actionable Result: result or insight found through analysis that can be acted on

Result ⇒ *Action*

Actionable Results

Identify successful trends	⇒	Optimize
Identify problems	**⇒**	**Fix/Improve**
Identify groups of common behavior	⇒	Personalize experience
Identify site metrics/benchmarks	⇒	Keep track of state of website

Domain-Specific Questions

How many bounce (exit after viewing one page)?

Tasks: Actionable Results

Actionable Result: result or insight found through analysis that can be acted on

Result ⇒ *Action*

Actionable Results

Identify successful trends	⇒	Optimize
Identify problems	⇒	Fix/Improve
Identify groups of common behavior	**⇒**	**Personalize experience**
Identify site metrics/benchmarks	⇒	Keep track of state of website

Domain-Specific Questions

Can you classify different types of buying behaviors?

Tasks: Actionable Results

Actionable Result: result or insight found through analysis that can be acted on

Result ⇒ *Action*

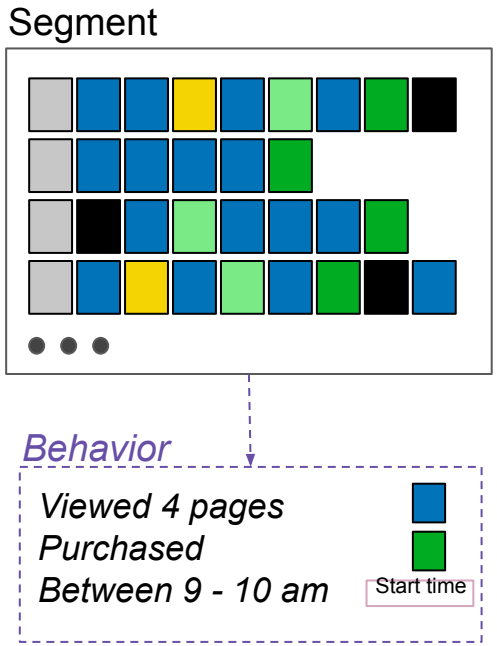
Actionable Results

Identify successful trends	⇒	Optimize
Identify problems	⇒	Fix/Improve
Identify groups of common behavior	⇒	Personalize experience
Identify site metrics/benchmarks	**⇒**	**Keep track of state of website**

Domain-Specific Questions

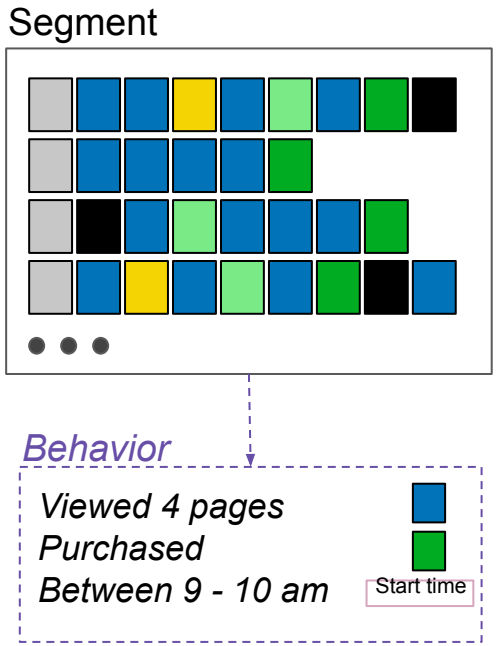
What is the average number of sessions in a month? Was this month abnormal?

Tasks: Segment Behavior



Behavior: set of attribute constraints

Tasks: Segment Behavior



Behavior: set of attribute constraints

- **Expected**
  Users add to cart before purchasing
- **Unexpected**
  No purchases on a certain month
- **Favorable**
  Purchased
- **Unfavorable**
  Bounced

Tasks: Task Abstraction

Identify Tasks

*Discovering new interesting **behaviors***

- *Identify* new interesting *behaviors*:
  - *Is there a way to distinguish a set of sequences representing behaviors?*
- *Identify* the cause or effect of a *behavior*:
  - *What behaviors trigger/follow behavior X?*
- *Identify* more fine-grained *behaviors* from an initial behavior:
  - *Can sequences that follow behavior X be described by more specific behavior Y?*

Verify Tasks

*Interesting **behaviors** known prior to analysis*

- *Verify* existence of a *behavior*
  - *Do any sequences follow behavior X?*
- *Verify* amount of support for a *behavior*
  - How many sequences follow behavior X?
- *Verify* if a *behavior* causes another *behavior*
  - *Do sequences that follow behavior X also follow behavior Y?*

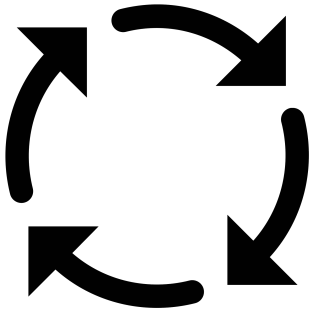
How to solve these goals with **Visual Analytics**?

Visual Analytics

Other Related Work

Our Solution

Why Visual Analytics?



- Automation would be nice...
  - Put data in, action results appear
- … but it is not realistic
  - Many possible questions, data-driven interplay between finding answers and generating new questions
- Human-in-the-loop visual data analysis
  - Integrate computing power of machine with intuition of domain experts

What **Visual Analytics Systems** exist for **Clickstream Data Analysis**?

Related Work

Two major categories:

Technique

- Focuses on the development of specific techniques
- Addresses one specific task
- Examples:
  - Clustering
  - Pattern mining

Exploratory

- Closest work to ours
- Focuses on exploring data
- Overview of data
- Major emphasis on viewing sequences

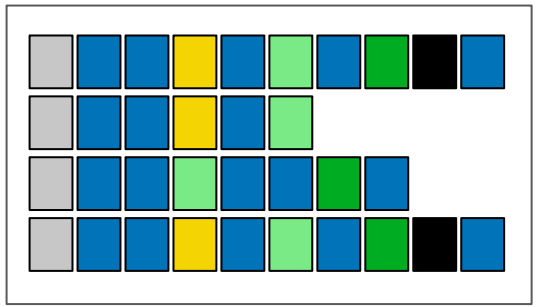
Related Work: Problems

Technique

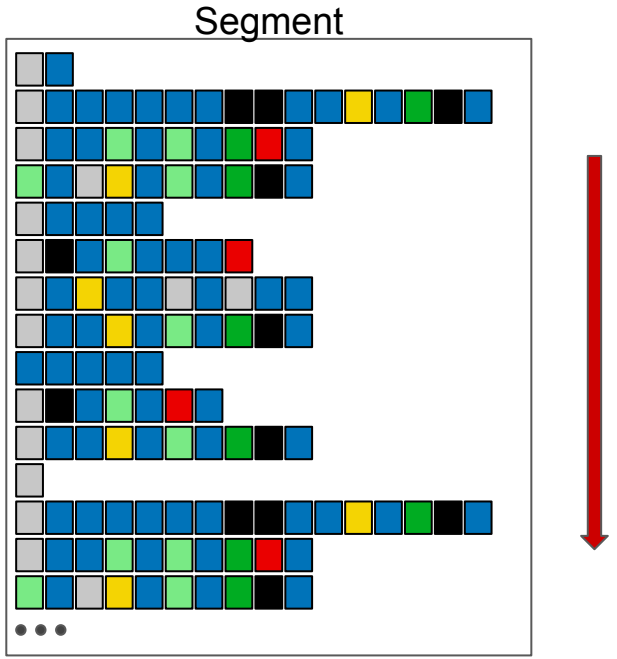
Exploratory

Evaluated using small, clean datasets

Segment

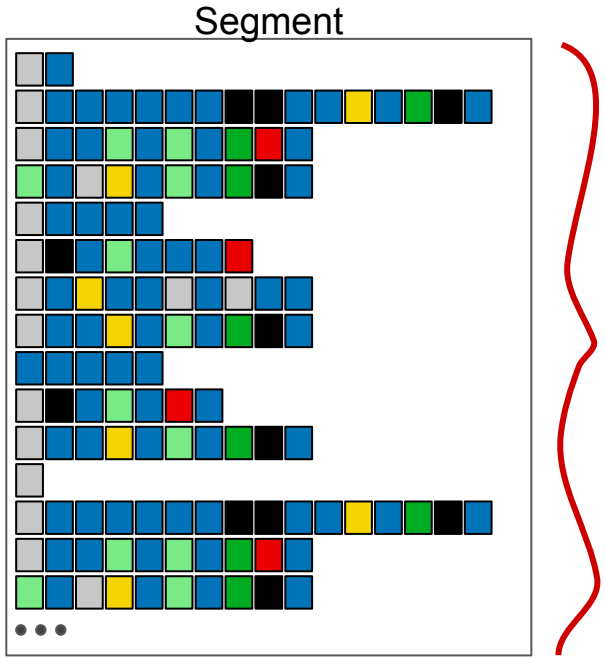


Real-world Clickstream Data



Scale is huge

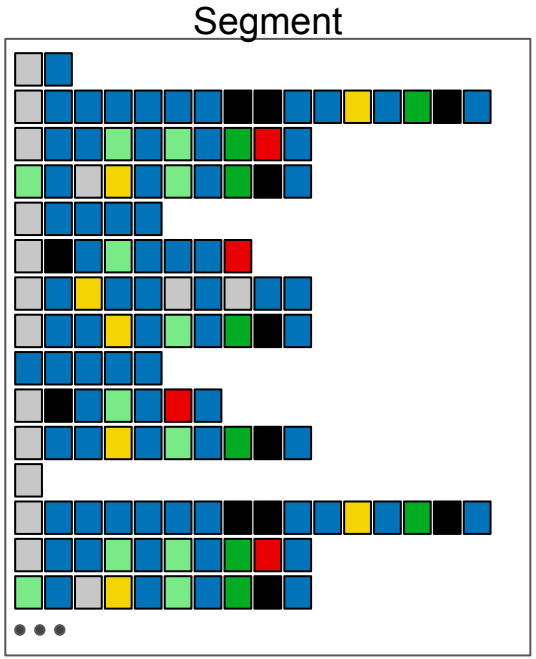
Real-world Clickstream Data



Scale is huge

Variability is high

Real-world Clickstream Data



Scale is huge

Variability is high

Most work **fails** when applied to real-world data.

Technique

Most techniques have data requirements to work effectively

Exploratory

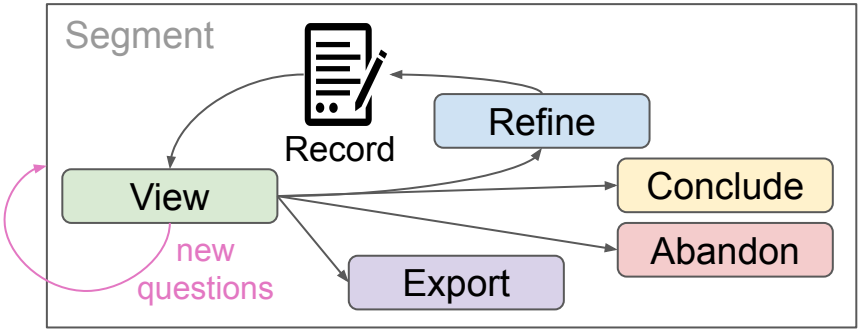
Most focus on analyzing sequences.
Too many to view at once.

Our Solution

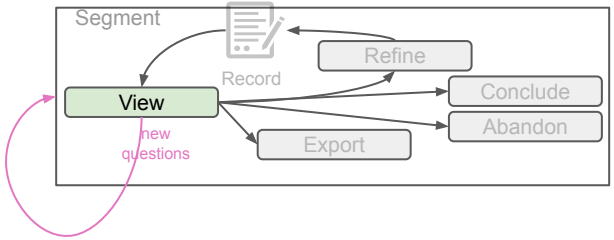
Clickstream Segment Analysis Framework

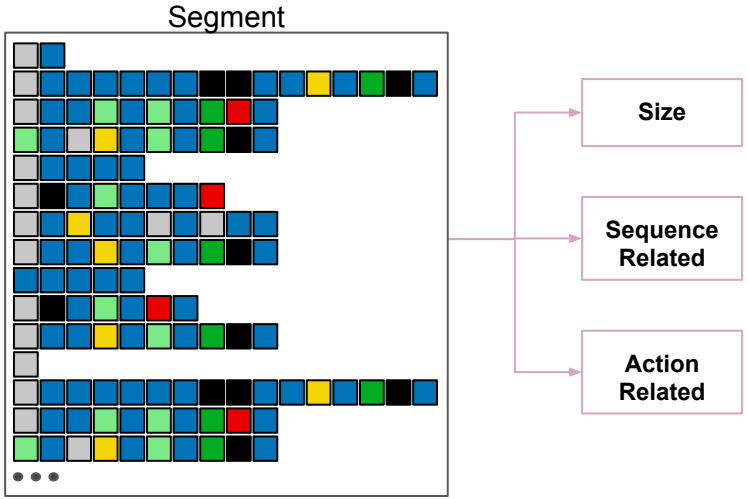
General idea:

Combine domain knowledge with computational support to iteratively view and refine data into segments that lead to **actionable results** or more effective **downstream analysis**



Clickstream Segment Analysis Framework

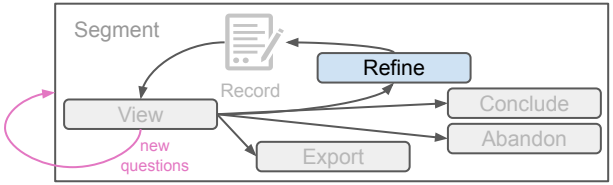


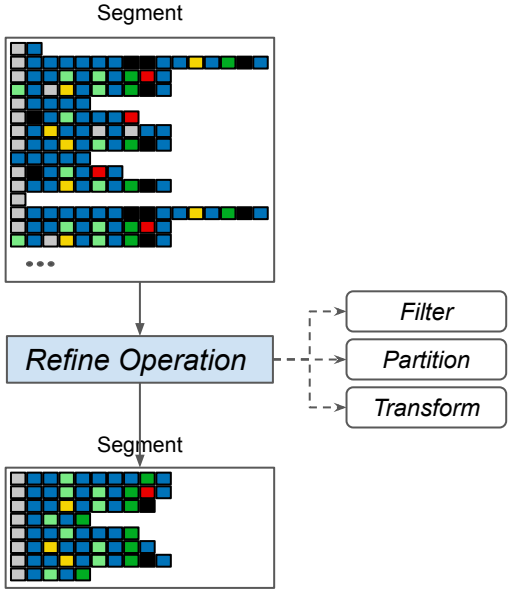


View

- Gives Insight into underlying data of segment
  - Sequences
  - Segment Attributes
- Leads to:
  - Actionable answers
  - New *questions*
  - New ways on how to *refine*
  - Whether segment should be *abandoned*
  - Whether segment should be *exported*

Clickstream Segment Analysis Framework

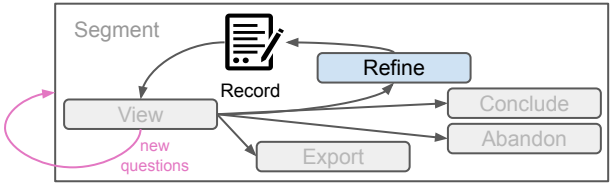


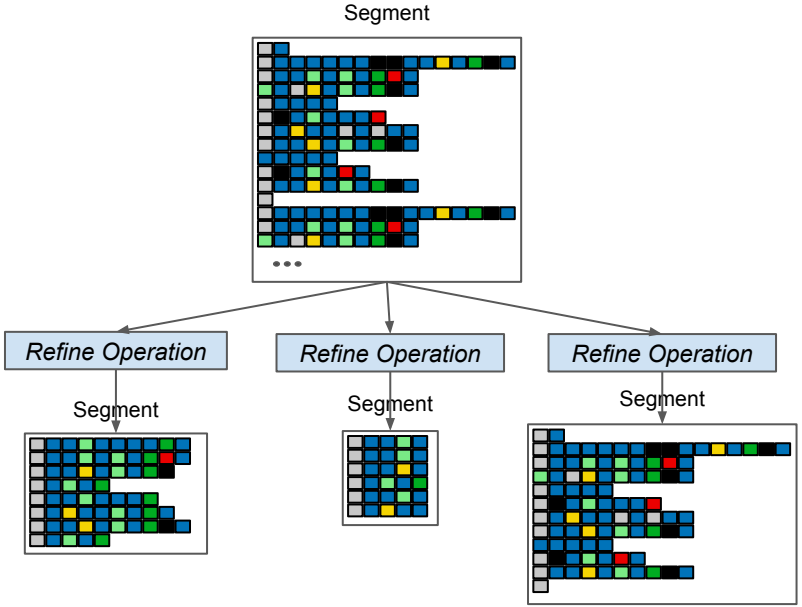


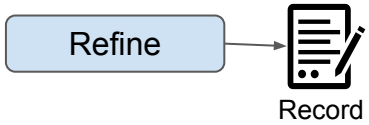
Refine

- Apply operation to create new segments
- Type of Refinements
  - *Filter*
  - *Partition*
  - *Transform*

Clickstream Segment Analysis Framework

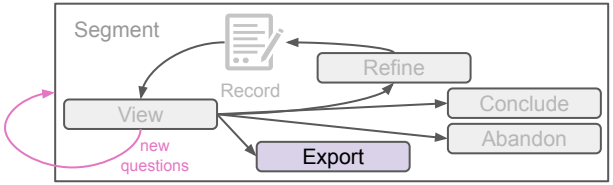


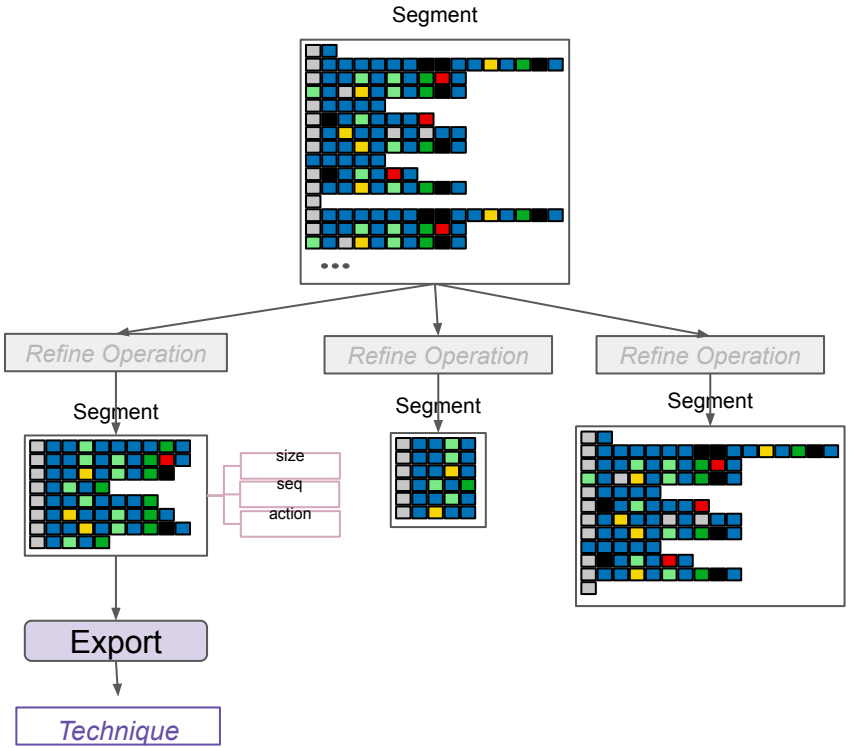




- Record all refinement steps automatically
- Keep track of questions
- Ability to create and view multiple segments from the same segment

Clickstream Segment Analysis Framework

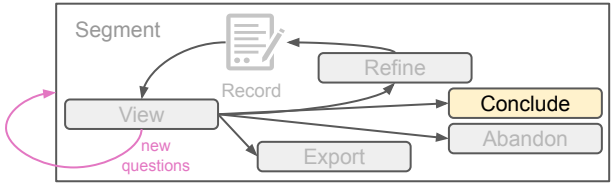


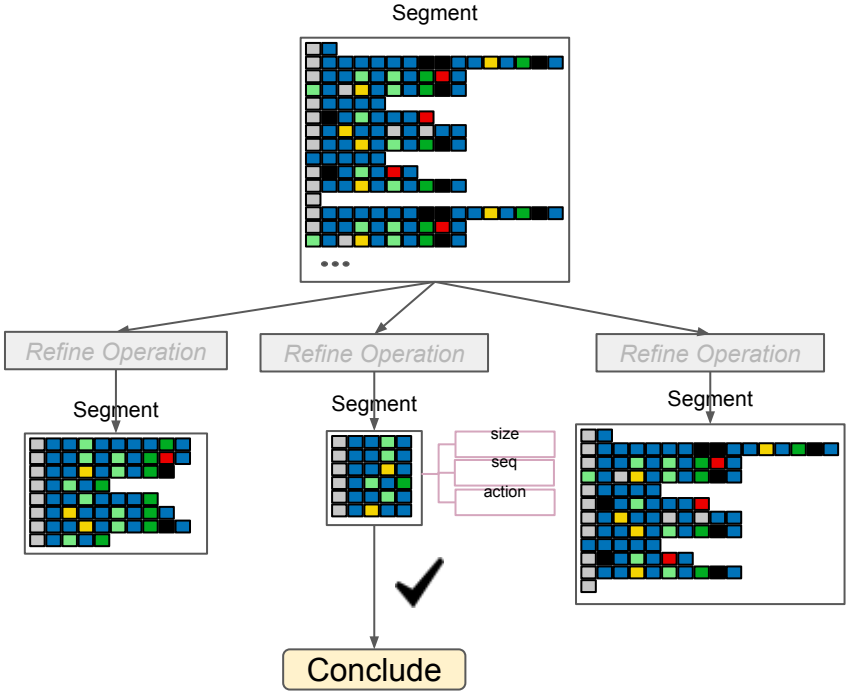


Export

- Export refined segments for further downstream analysis, such as applying techniques:
  - Pattern mining
  - Clustering
- Each technique:
  - Answers a specific *task*
  - Has *data requirements* for effective results

Clickstream Segment Analysis Framework

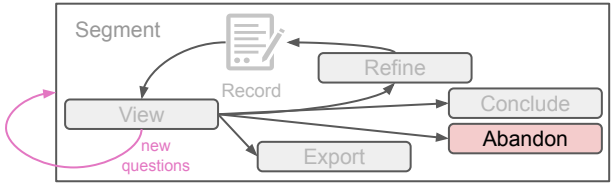


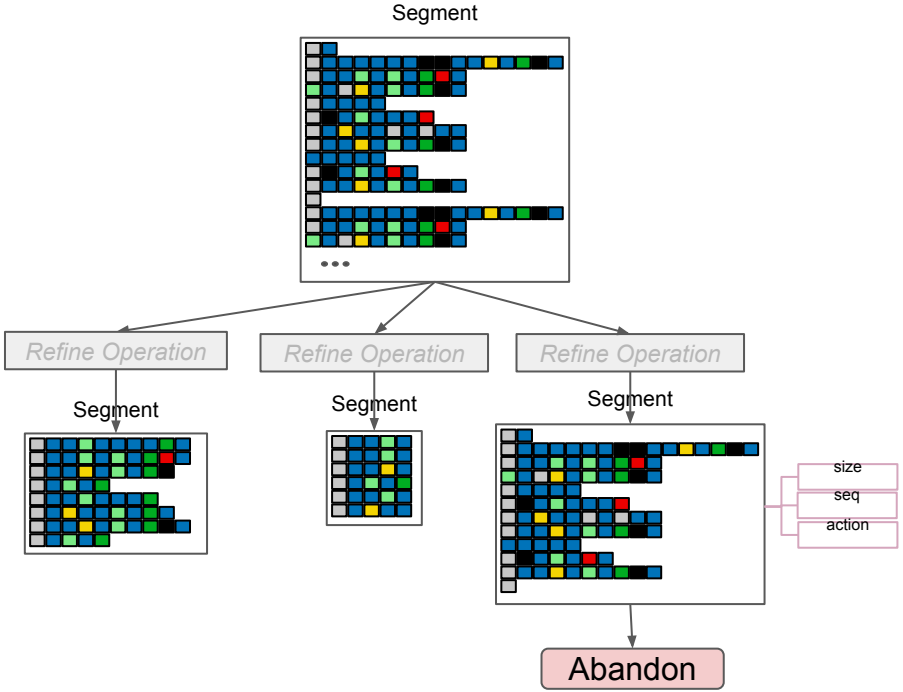


Conclude

- Discover action result by *viewing* segment

Clickstream Segment Analysis Framework



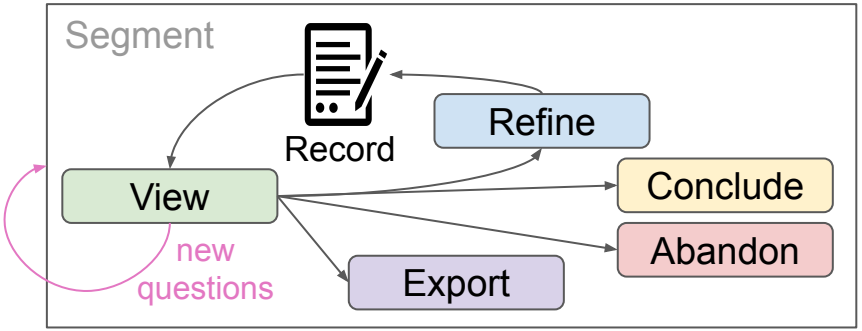


Abandon

- By *viewing* the segment, analyst *abandons* if:
  - No actionable results
  - No further ways to *refine*
  - Not suitable for *export*

Clickstream Segment Analysis Framework

- Take a *giant, noisy dataset* and refine it into *small, clean segments* appropriate for each *task*
- Bridge the gap between *real-world data* and other techniques
- Encapsulates the design rationale of **Segmentifier**



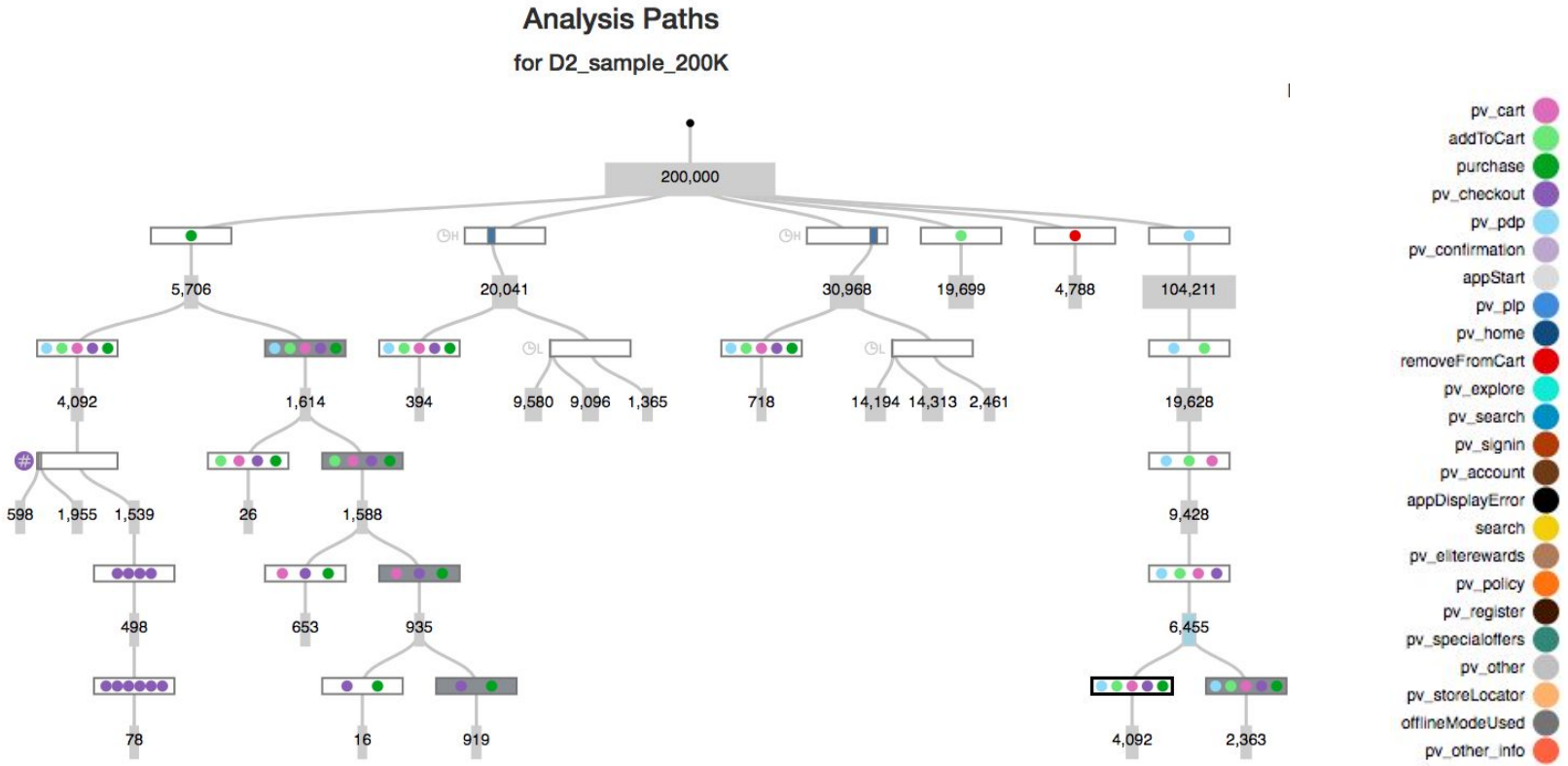
The Segmentifier Interface

Results

Case Study #1

- 2 hour chauffeured analysis
- With Mobify data analyst
- Purpose:
  - One month post launch report
  - Discover actionable insights and improvements for customer
- Data
  - Session sequences
  - 200K sequences

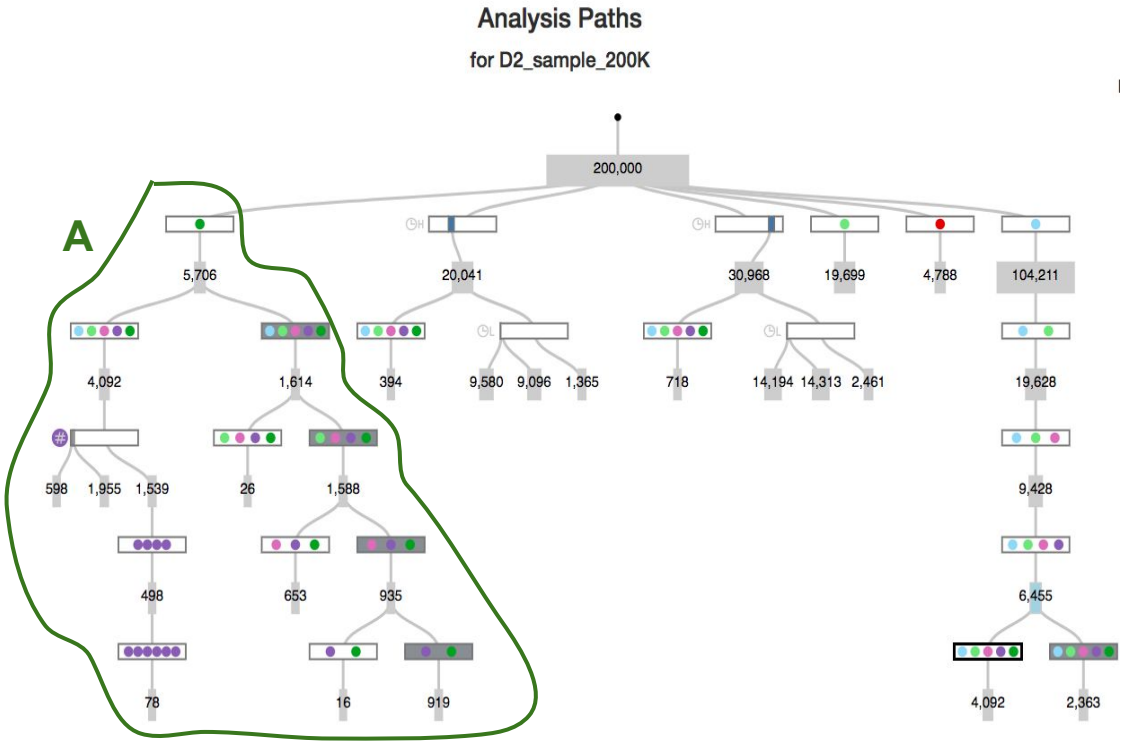
Case Study #1


Analysis Paths
for D2_sample_200K
200,000
5,706
20,041
30,968
19,699
4,788
104,211
4,092
1,614
394
9,580
9,096
1,365
718
14,194
14,313
2,461
19,628
598
1,955
1,539
26
1,588
9,428
498
653
935
6,455
78
16
919
4,092
2,363
pv_cart
addToCart
purchase
pv_checkout
pv_pdp
pv_confirmation
appStart
pv_plp
pv_home
removeFromCart
pv_explore
pv_search
pv_signin
pv_account
appDisplayError
search
pv_eliterewards
pv_policy
pv_register
pv_specialoffers
pv_other
pv_storeLocator
offlineModeUsed
pv_other_info

Case Study #1: *Analysis A*

A Analyze Purchasing Behavior

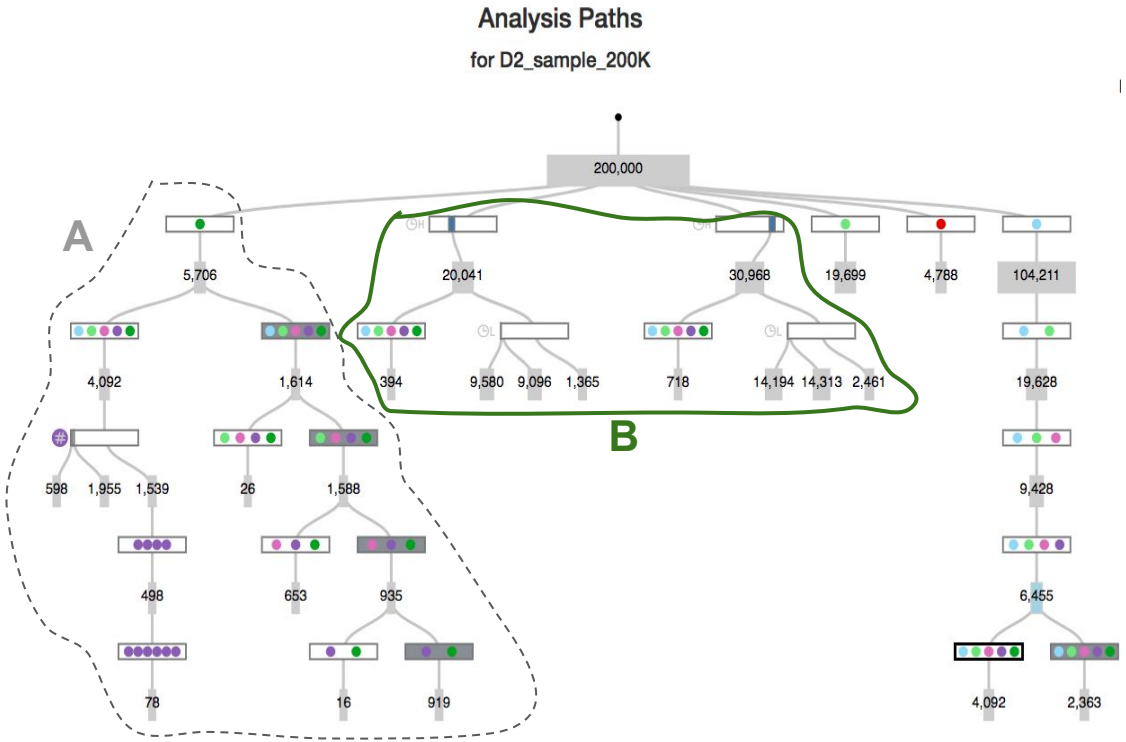
- 12% of sessions contain more checkout pages than necessary
- 30% of users actually exit the site and return later to complete their purchase



Case Study #1: *Analysis B*

B Compare Morning vs Night

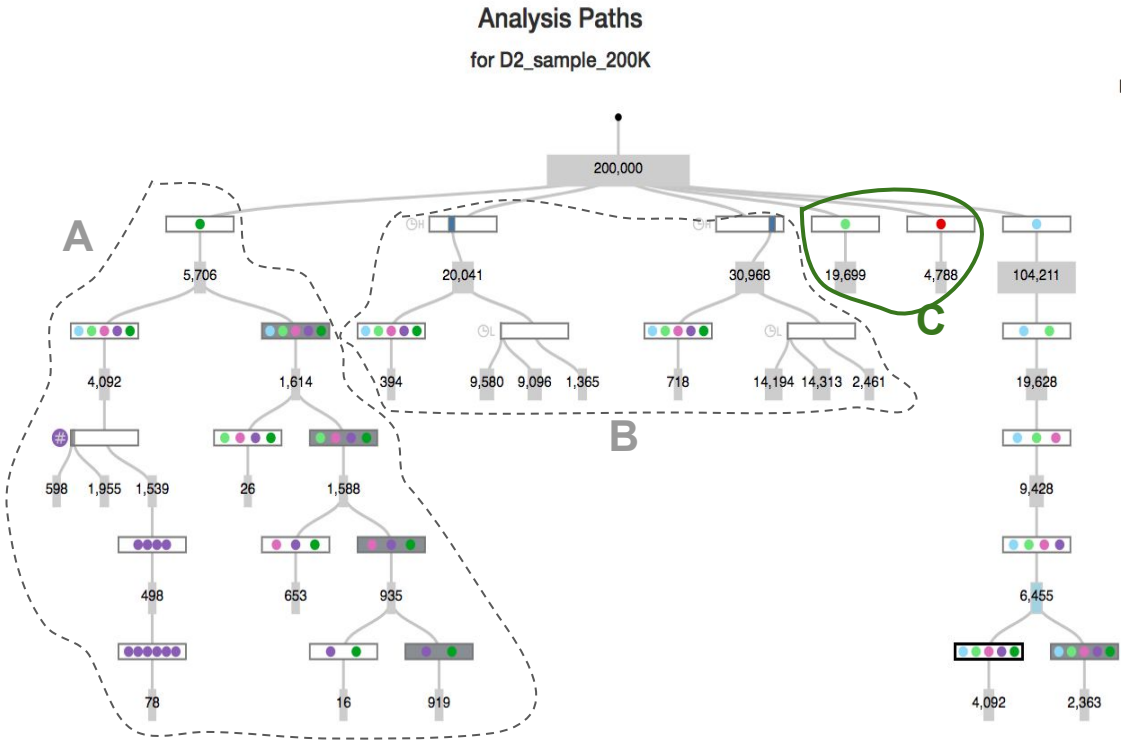
- No significant difference for percentage of sessions that contain full purchasing funnel
- No significant difference for number of actions



Case Study #1: *Analysis C*

C Analyze add and remove from cart behavior

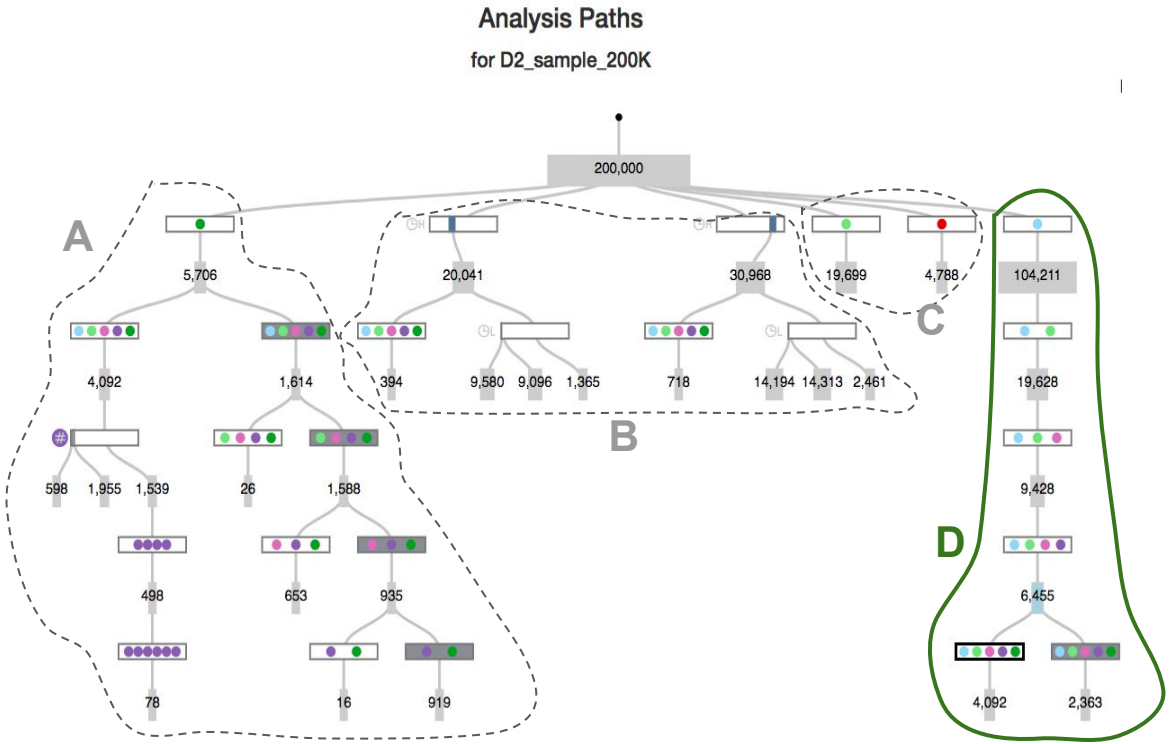
- No insight for add to cart behavior
- 30% of users who removed from cart exited the session and most likely did not come back



Case Study #1: *Analysis D*

D	Analyze purchasing funnel

- 20% of people who get to checkout will not end up purchasing



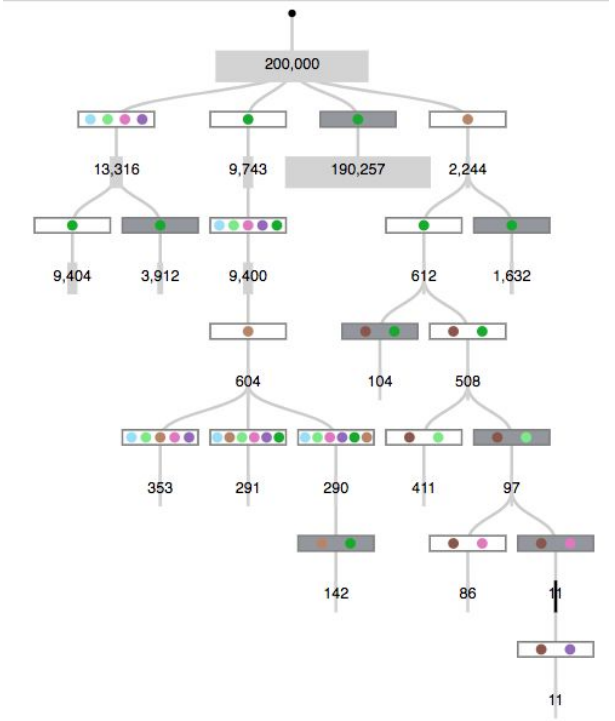
Case Study #2

- 2 hour chauffeured analysis
- With Mobify data analyst
- Purpose:
  - Revisit some questions from last analysis using **client sequences**
- Data
  - **Client sequences**
    - Much longer
    - Capture longitudinal behavior
  - 200K sequences

Case Study #2

Summary of Insights

- 25% who remove from cart at checkout stage, exit and never purchase
- appStart action triggered before cart page
- Awards page analysis:
  - 1% signed up
  - 27% purchased
  - Longer sequences

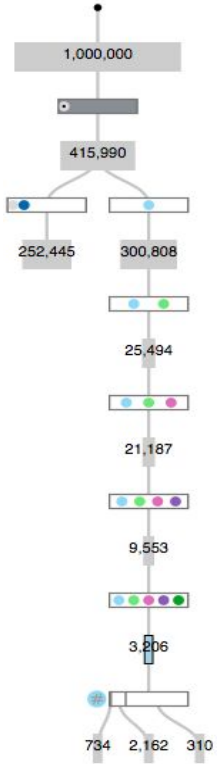


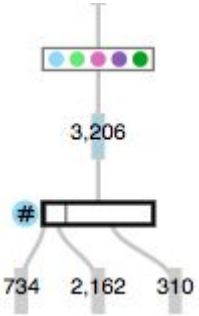
Discussion + Future Work

Discussion + Future Work

- Understandable segments:
  - Each possible refinement operation corresponds to one attribute constraint
  - In contrast to clustering, pattern mining that have uninterpretable results for this scale of noisy data

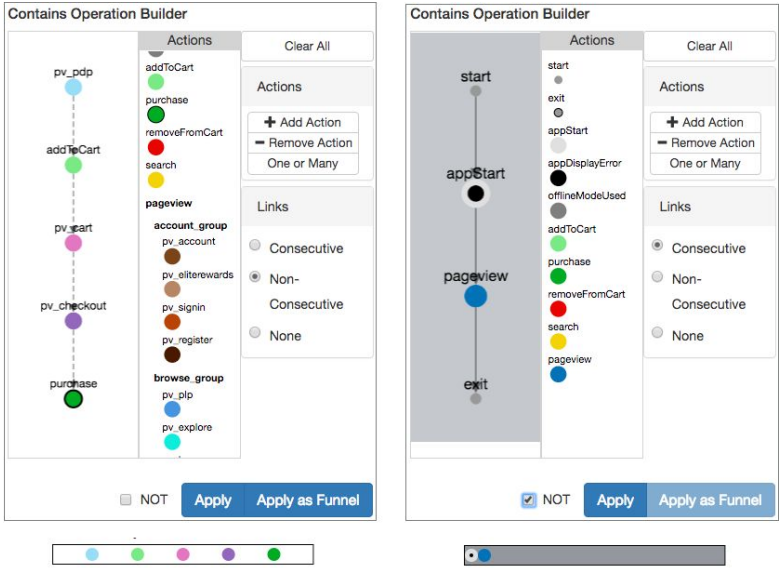
- Segmentifier explicitly supports refinement through both filtering and partitioning.
  - Encourages subsequent analysis
  - Allows comparison
  - Future comparison work





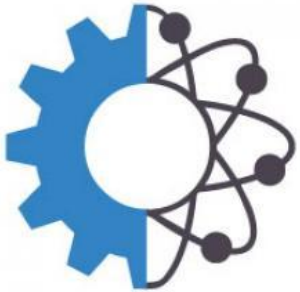
Discussion + Future Work

- Trade-off between power and simplicity for this application context
  - Actions Operation Builder: regex with glyphs
  - Previous work:
    - Full support of regex
    - Difficult for non-programmers
  - Our design: deliberately less powerful so usable by non-technical analysts



Discussion + Future Work

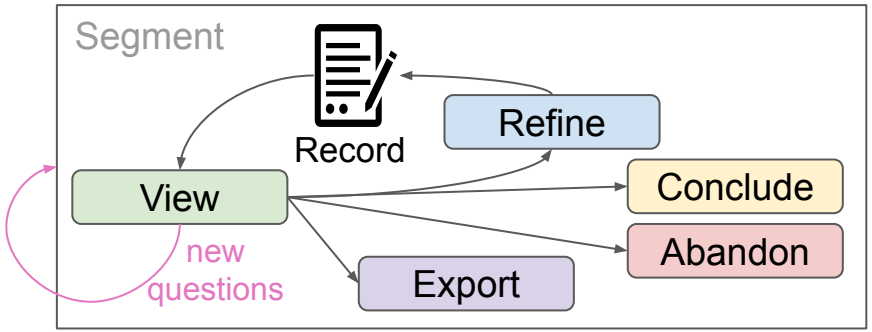
- Focus on agile and iterative development of design
  - Modest engineering effort to achieve base level of usability to test design concept
    - Loading times
    - Processing time
  - Goal:
    - Proof of concept that design works for target tasks
    - Not (premature) engineering optimization
  - Future work:
    - Engineering optimization for this final design



Conclusions

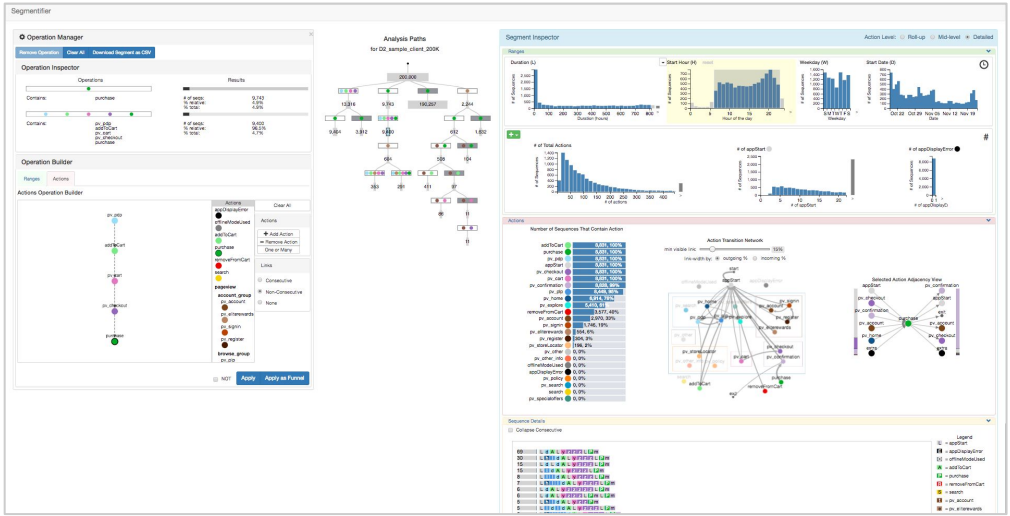
- Thorough **characterization of task and data abstraction** for clickstream data analysis

Clickstream Segment Analysis Framework



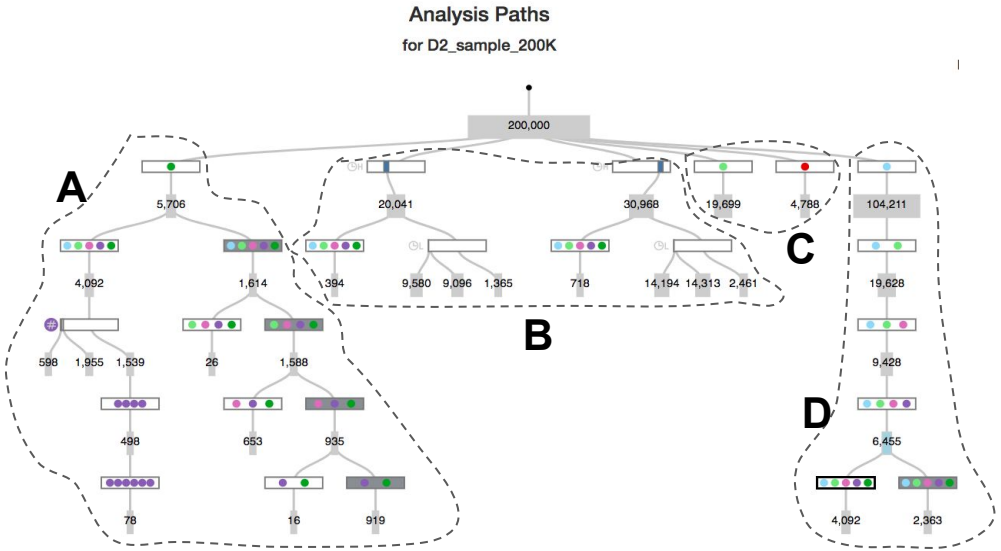
Conclusions

- Thorough **characterization of task and data abstraction** for clickstream data analysis
- **Segmentifier: novel analytics interface** for refining data segments and viewing characteristics before downstream fine-grained analysis



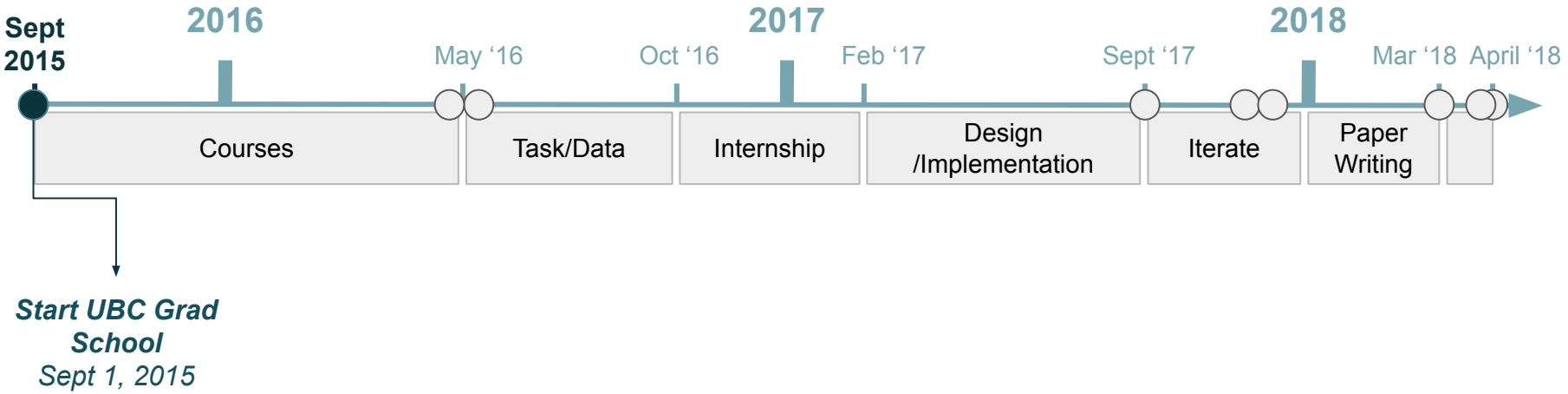
Conclusions

- Thorough **characterization of task and data abstraction** for clickstream data analysis
- **Segmentifier: novel analytics interface** for refining data segments and viewing characteristics before downstream fine-grained analysis
- Preliminary **evidence of utility**

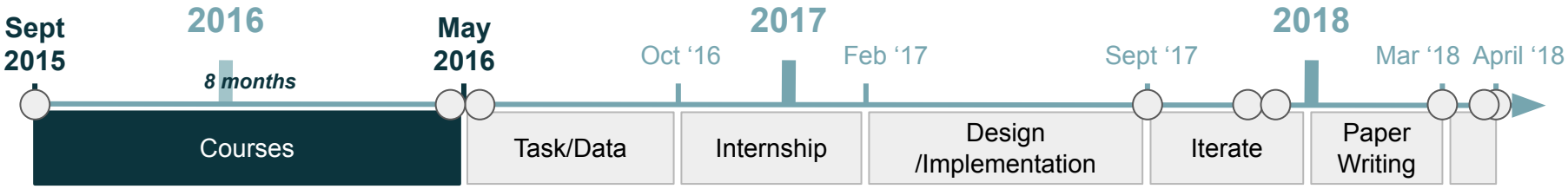


Master Thesis Timeline

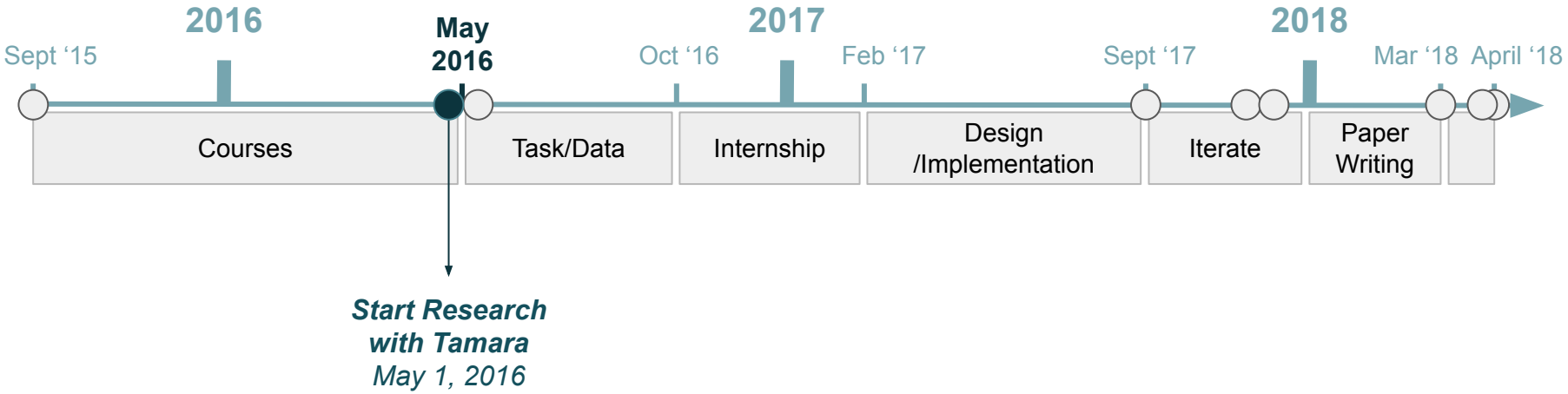
Master Thesis Timeline



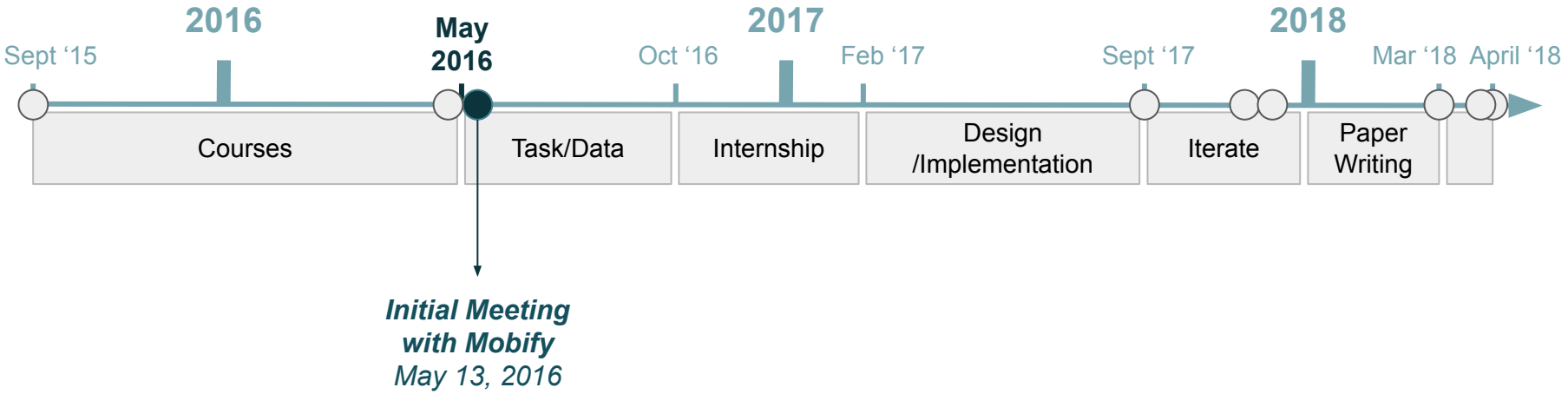
Master Thesis Timeline



Master Thesis Timeline



Master Thesis Timeline



Master Thesis Timeline

Task and Data

- *Winnow, discover* phase
- A lot of initial meetings
  - 12 employees
  - 3 different teams (Design, Data Insights, Data Engineering)
- Very general goals
  - No clear everyday tasks
  - No clear idea of solution
- Originally three different tasks
  - Messaging
  - Load times
  - **User behavior**

Master Thesis Timeline

Internship

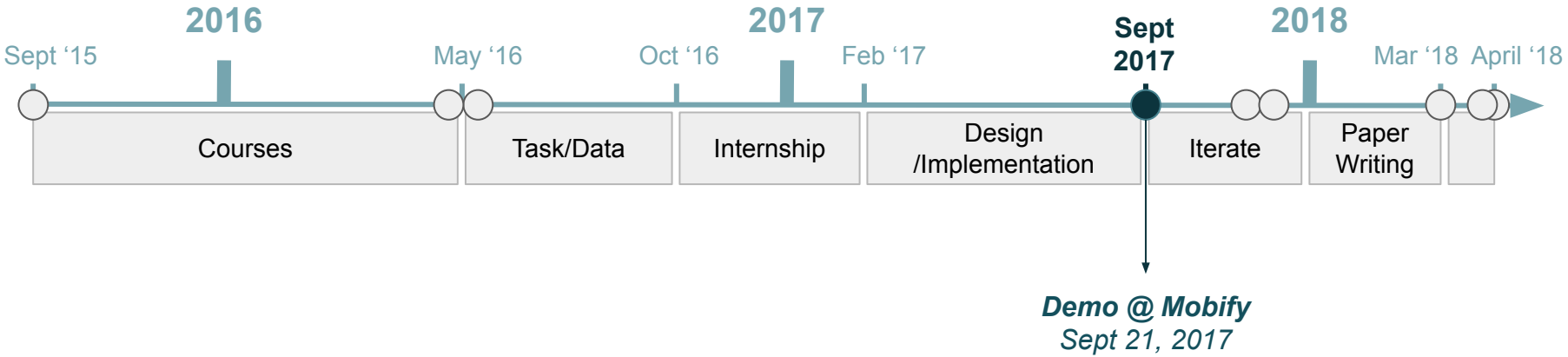
- 4 months @ Mobify
- Completely unrelated to thesis work
  - In contrast to original plan
  - Side effect of Mitacs funding

Master Thesis Timeline

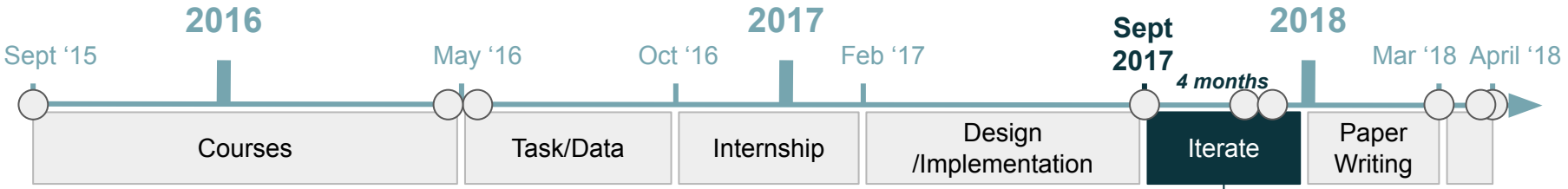
Design/Implementation

- **Design**
  - Data and task abstractions
  - Mapping to Design elements
  - Many meetings with Tamara
- **Implementation**
  - No previous web dev experience

Master Thesis Timeline



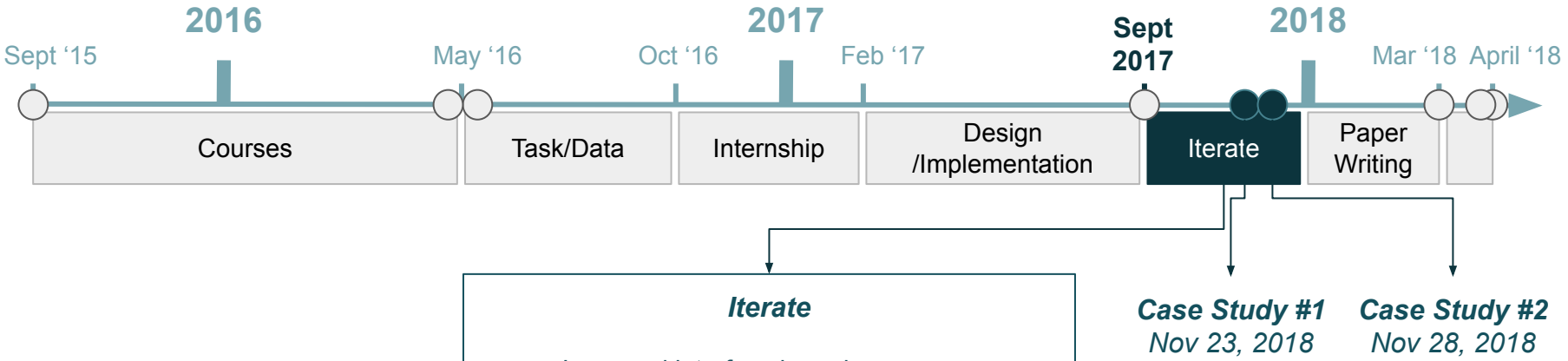
Master Thesis Timeline



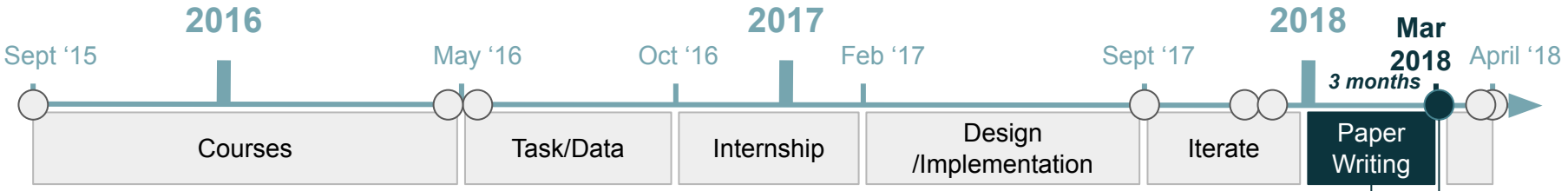
Iterate

- Improved interface based on observations/feedback from ***Demo***

Master Thesis Timeline



Master Thesis Timeline

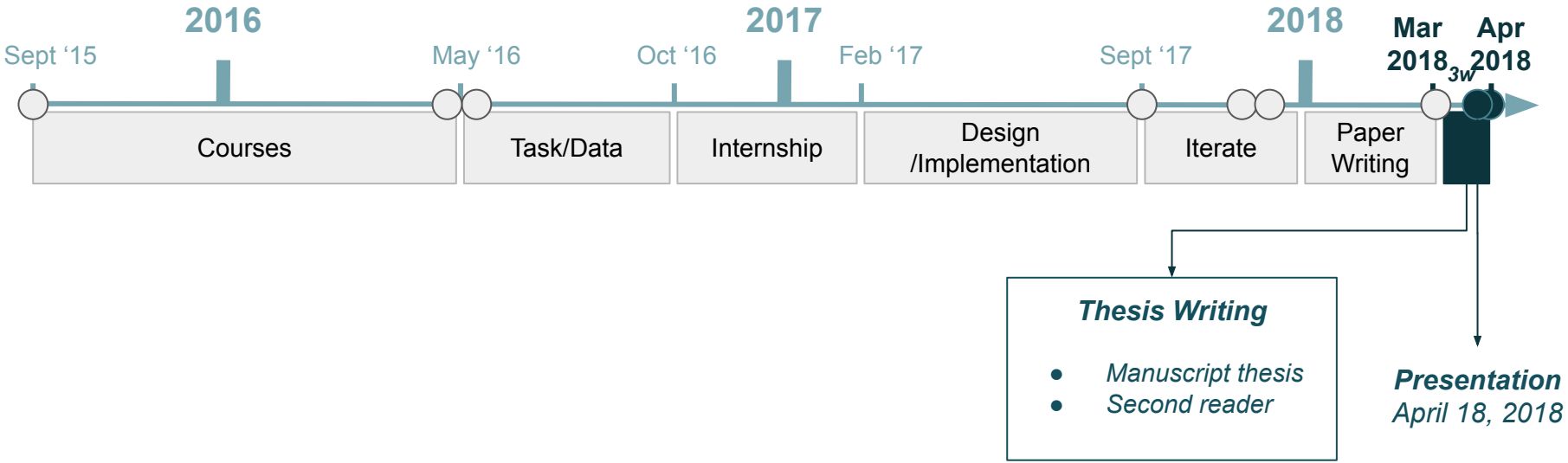


Paper Writing

- *No deployment*
  - *Lack of analyst time inside company*
- *Framework articulation*
- *Pre-paper talk*
  Jan 17th, 2018
- *Paper draft*
  March 7th, 2018

Paper Deadline
March 31, 2018

Master Thesis Timeline



Thank you

Questions?

Extra Slides