

Visualizing Collections of Internet Archives

John Berlin, Joel Rodriguez, Slobodan Milanko

Motivation

- Current solutions are limiting; small focus
 - Unable to effectively retrieve mementos without knowing exact URI-R
- Archivers are unable to visually experience collections
 - Lack of resource evolution
 - Lack of comparison
 - Lack of statistics
- User driven, clutter free, design for moderate collections
- Archivers desire to visually share collections with each other

Dataset Description

- We're abstracting the user's archive collection as a table
- Key attribute (URI-R) that maps to categorical or ordered values
 - Archiver - Categorical
 - Tags - Categorical
 - Number of archives - Quantitative
 - URI-M - Categorical
 - Time/date - Ordered
- The table approach
 - Facilitates ordering
 - Filtering of data

Abstract Tasks

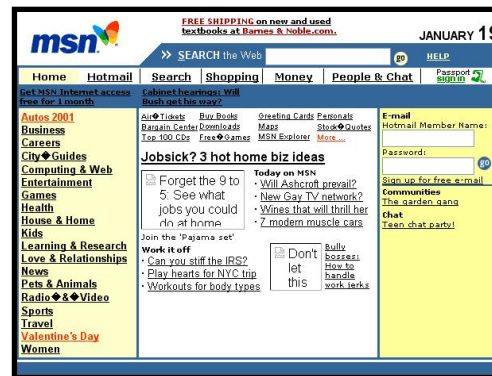
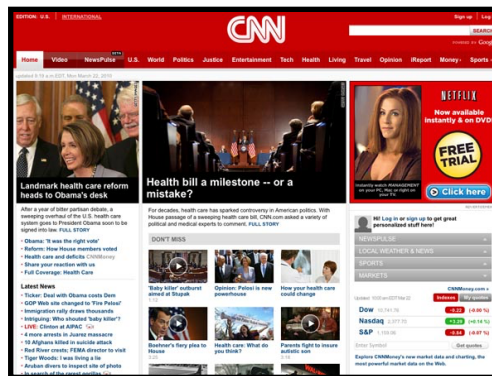
- Lookup, browse, explore and locate can be achieved easily with the table abstraction
- Users start with an overview first, zoom and filter later, and retrieve details on demand
- Users will be able to identify, compare and summarize different archive records
- Analysis through discovery, presentation, and enjoyment
- Observe interesting characteristics, trends and relationships

Potential Design

Mementos

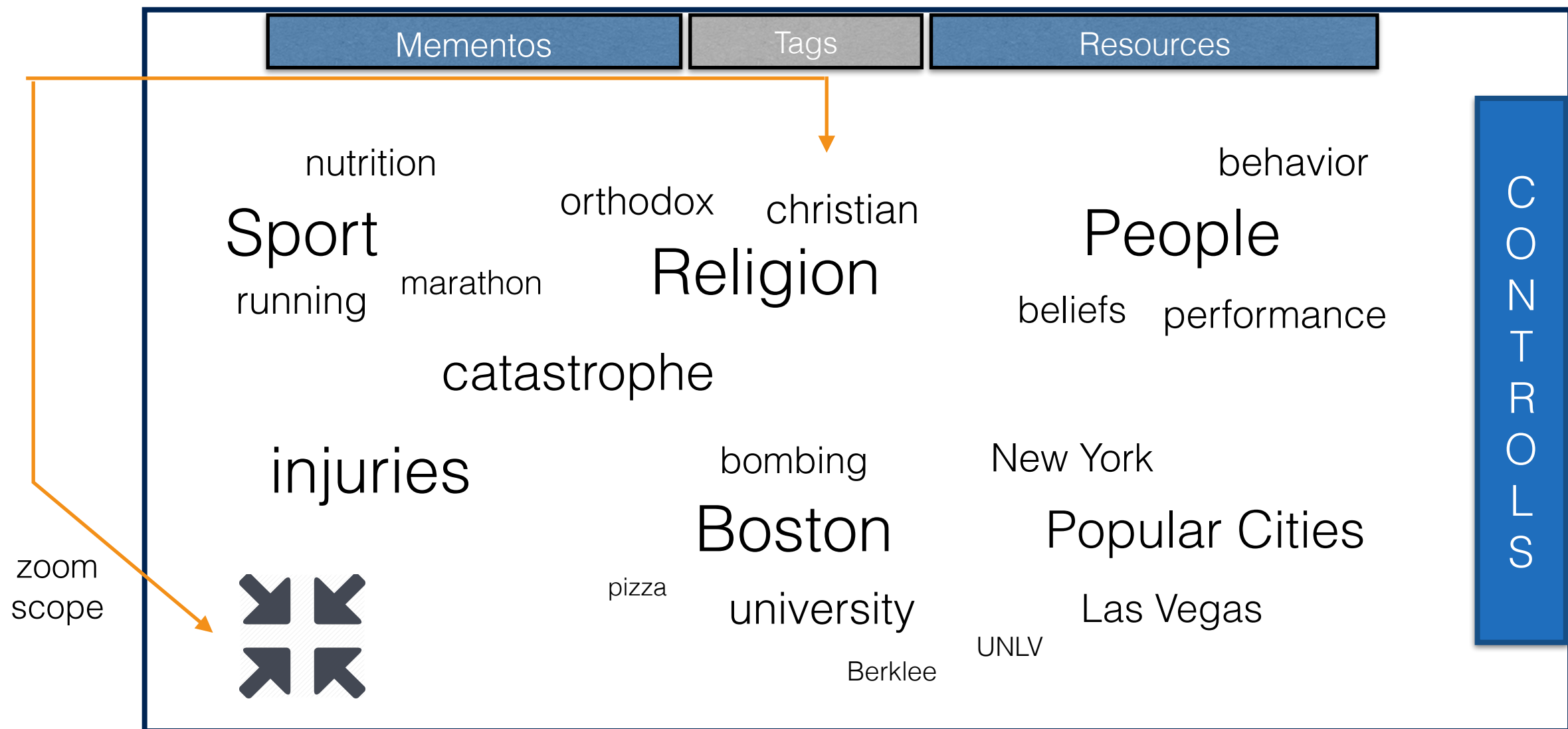
Tags

Resources

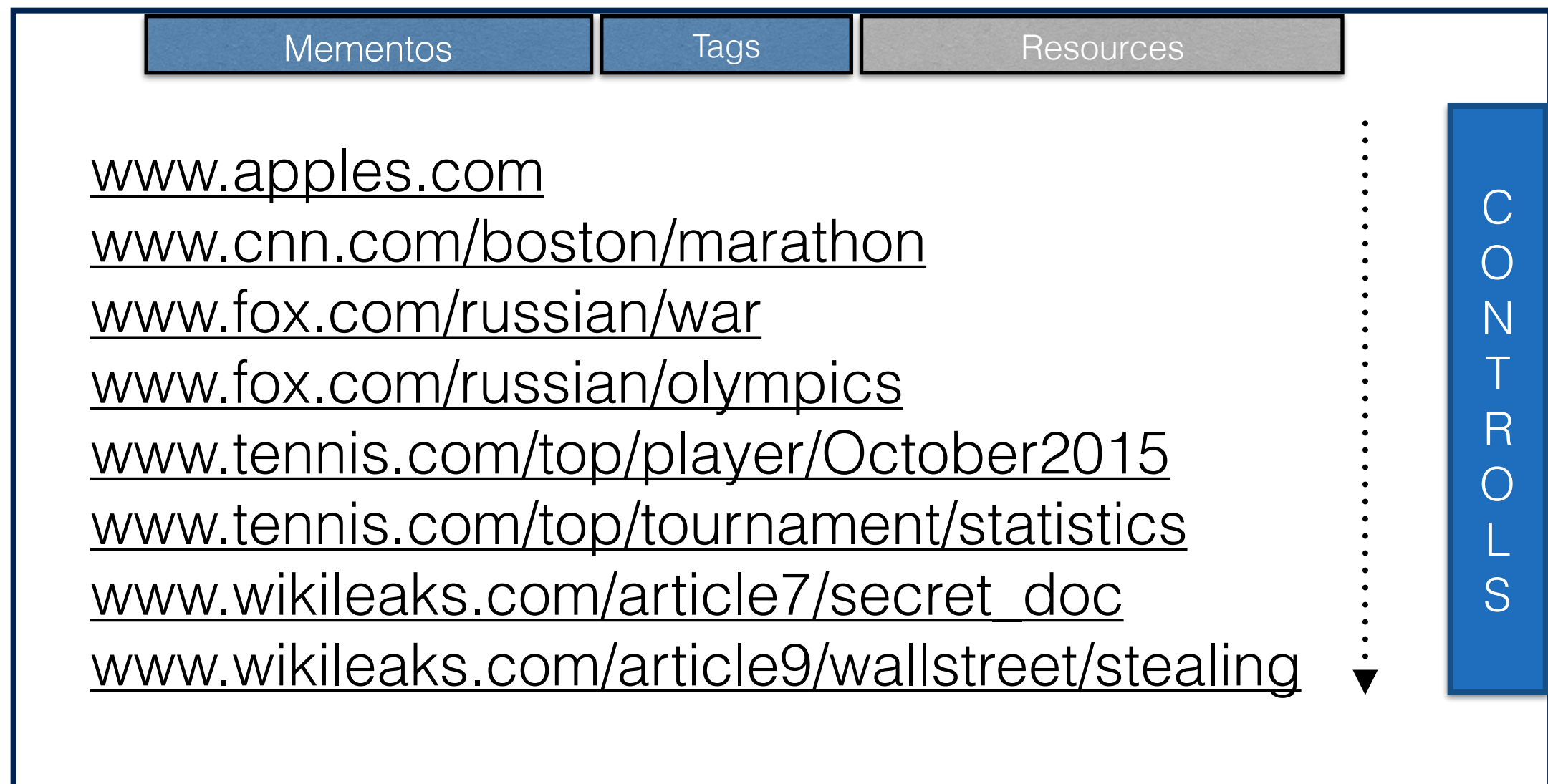


CONTROLS

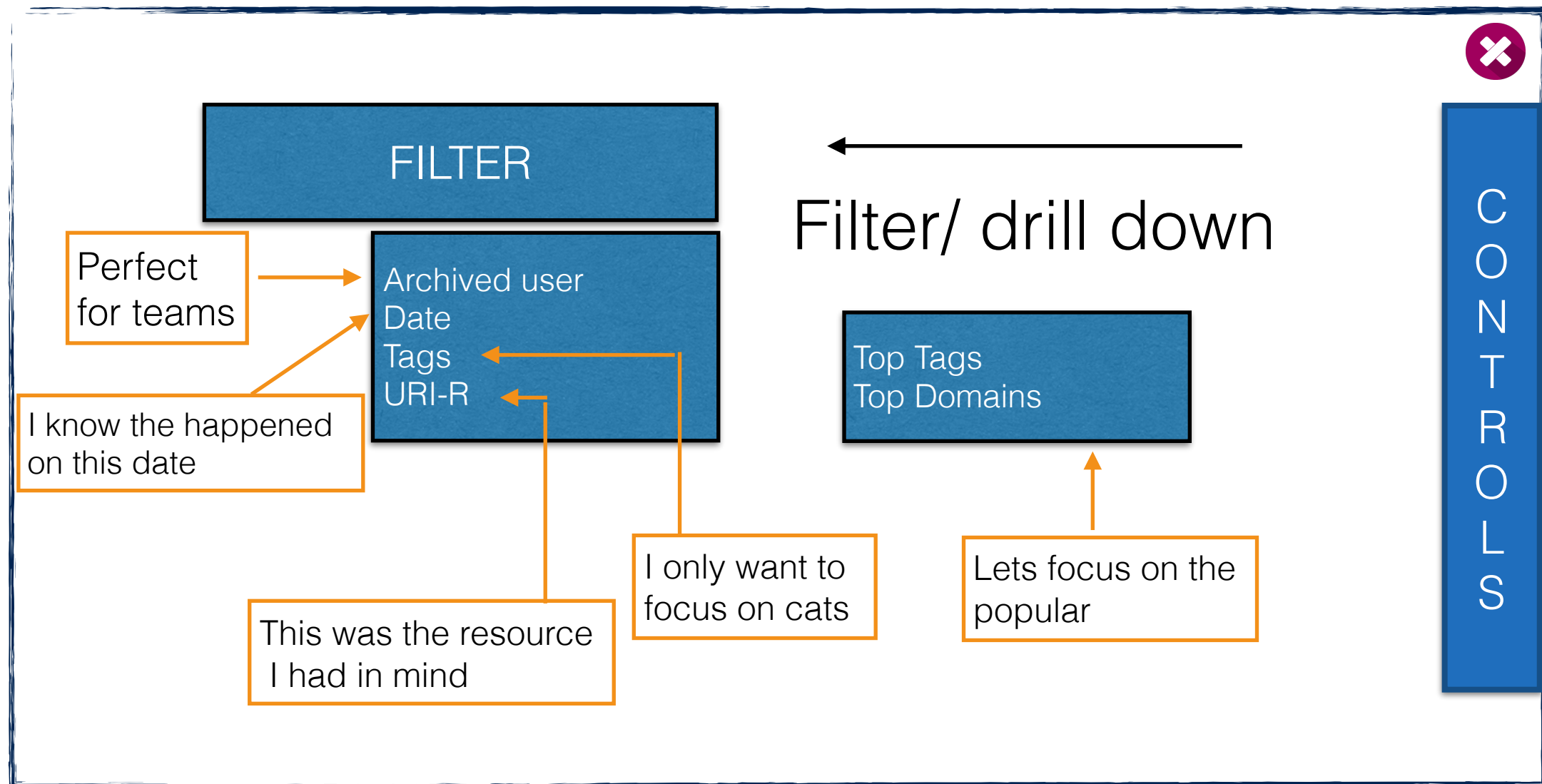
Continued: Potential Design



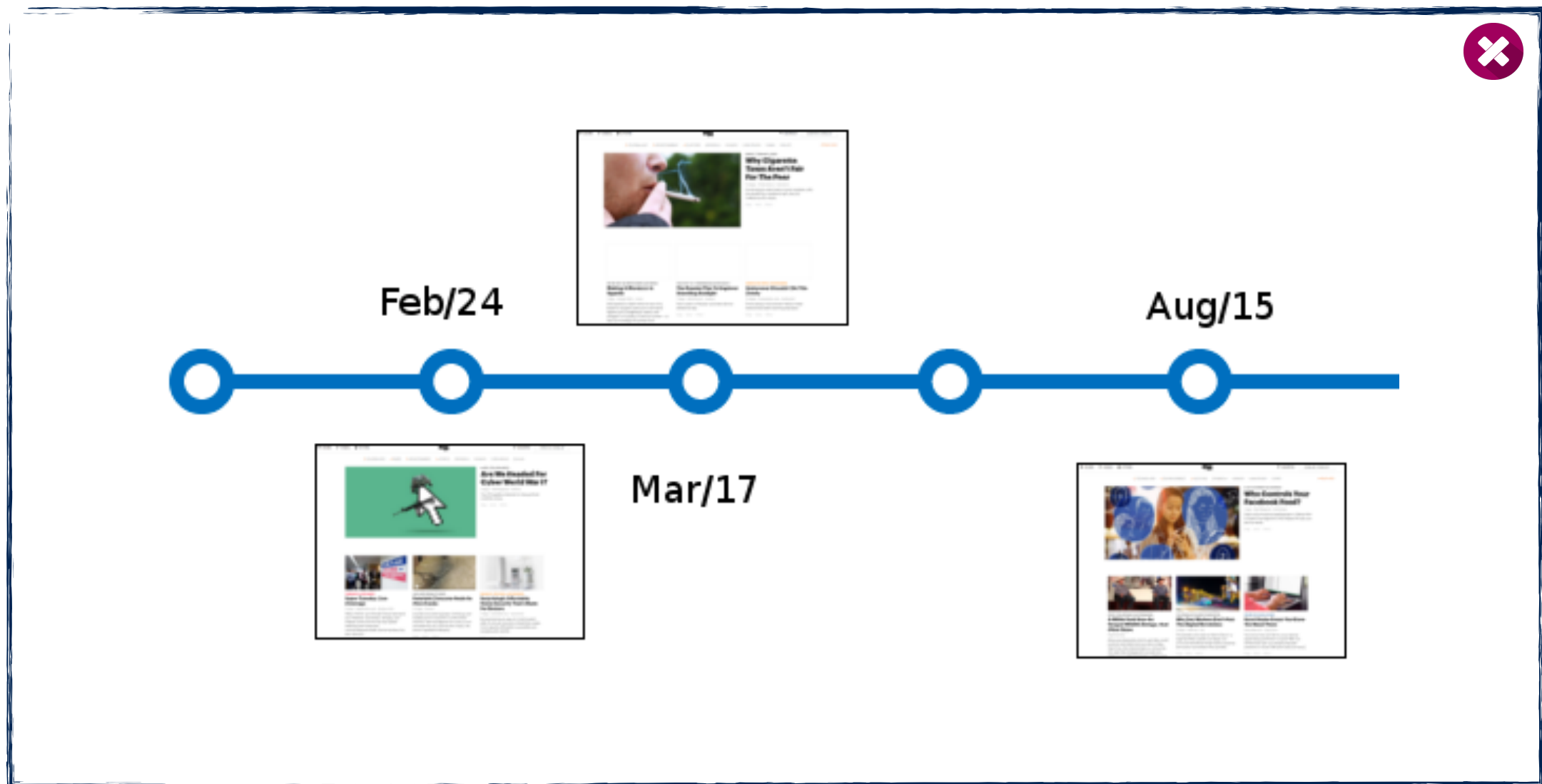
Continued: Potential Design



User Driven Control

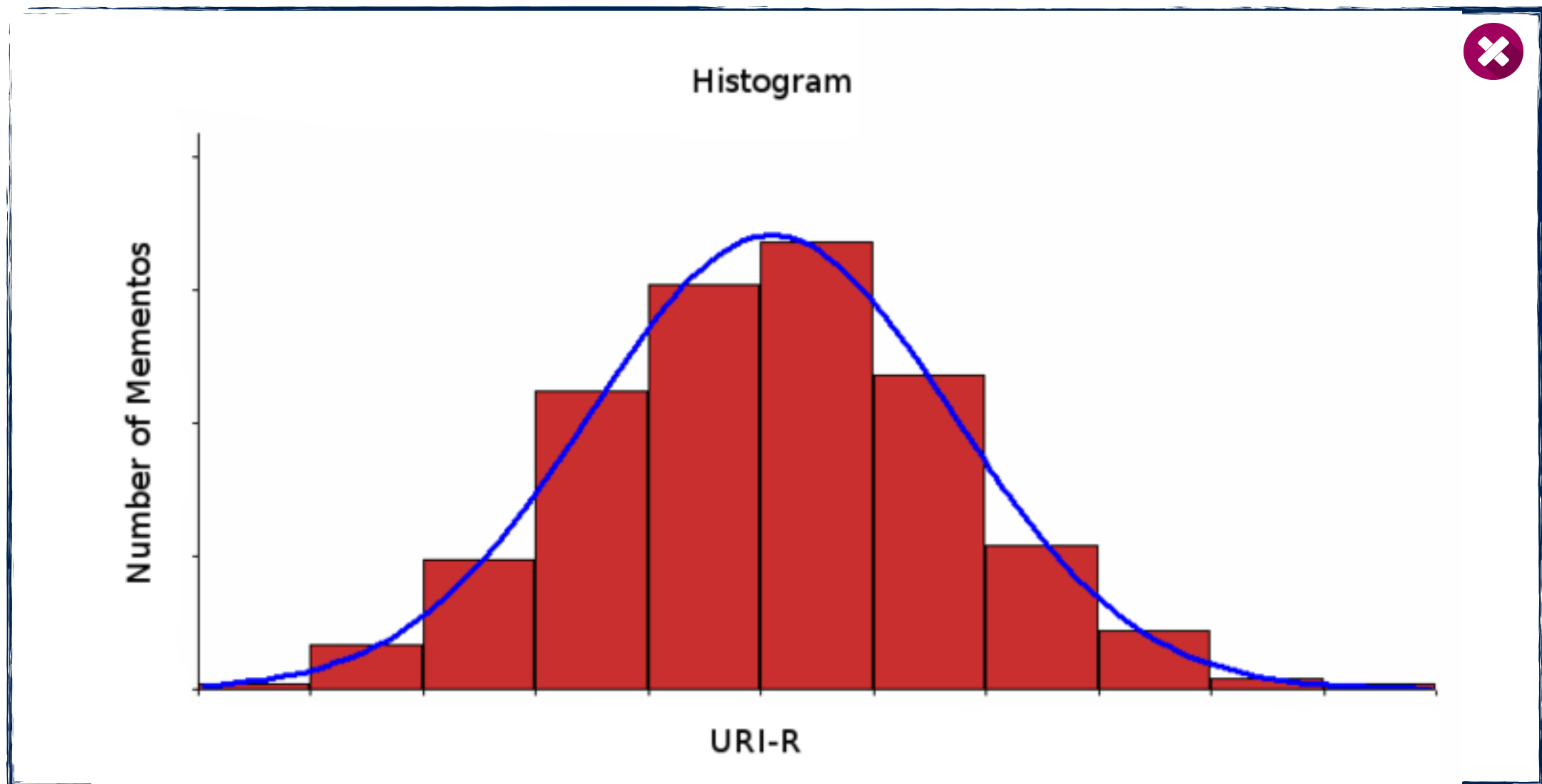


Details



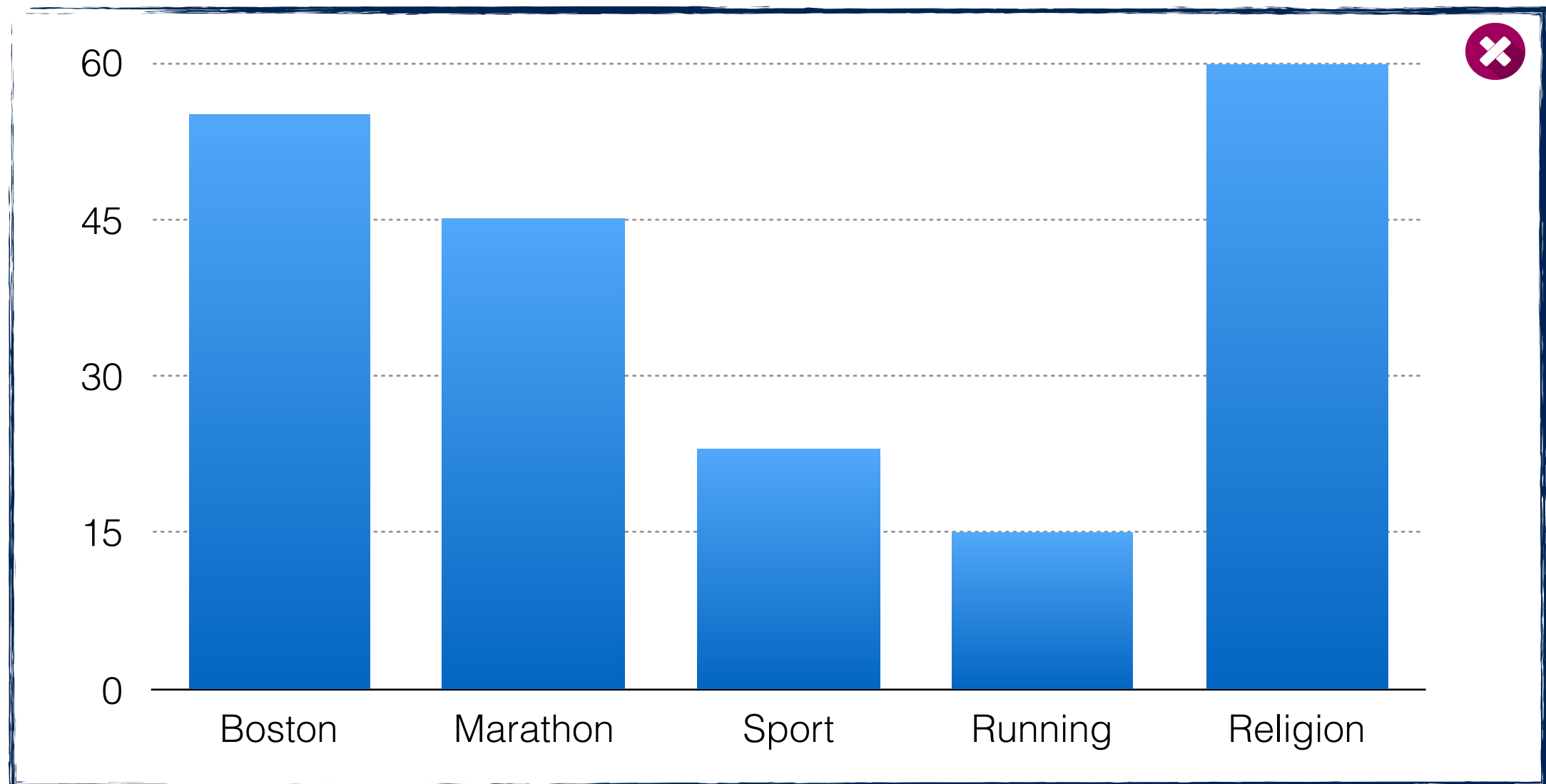
Resource evolution over time

Details: Continued



Counts of mementos per resource

Details: Continued



Show top tags or top domains

Needed Tools

- Mink, WARCCreate
 - Create sample archives for local testing
- Wayback Machine (Local/Remote)
 - Get an understanding for look of archives
- WAIL
 - Local archiving and hosting
- Google API Spreadsheet

Group Roles

- Collaborative goal: Develop a working product
 - Connect to spreadsheet and retrieve data
 - Create operations:
 - Sorting, filtering, grouping, searching
 - Implement views and controls to call operations

Group Roles: Continued

- Entire Group
 - Test - Ensure visualization accepts dynamic data
 - Survey for feedback
 - Create demo
- Dan & Joel
 - Create final presentation
- John
 - Final presenter of solution

Estimation

- 4-6 weeks to develop
- 1 week to demo and test
- Weekly feedback to ensure satisfaction