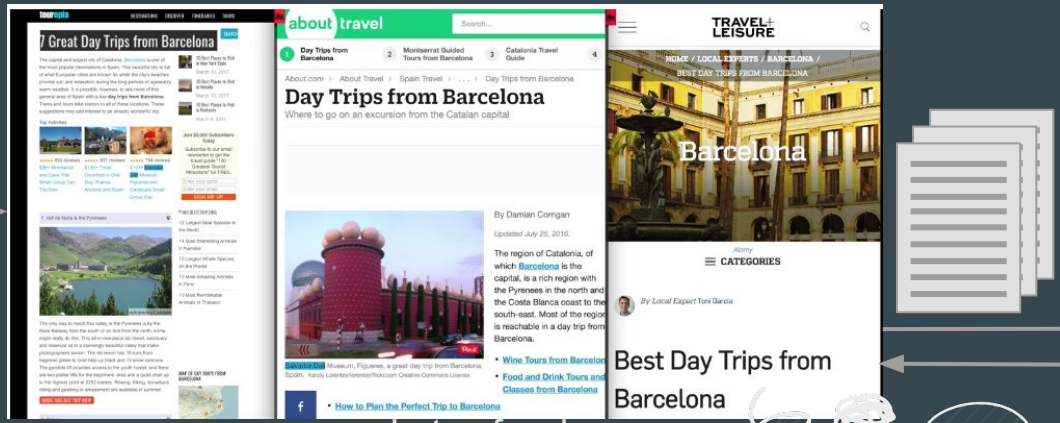
# Understanding Search Landscapes

Joseph Chee Chang
Nathan Hahn
Jade Yu-Ju Chang

# Overview and Motivation

Motivation: Existing search interfaces require individuals to manually manage, cross reference, and summarize information

Day Trips from Barcelona



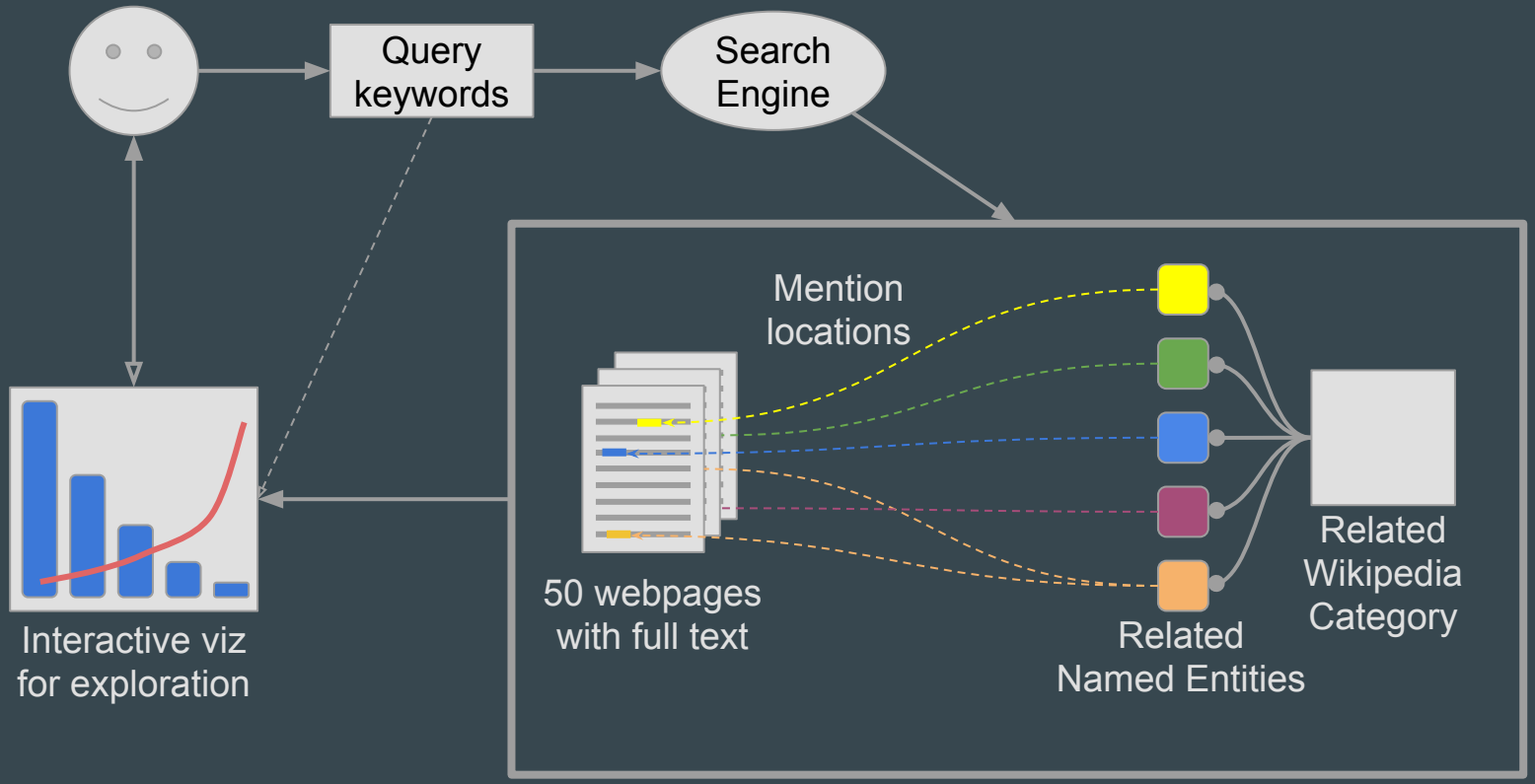

Lots of webpages

Cross referencing multiple webpages and requerying

# Questions

- What is a possible better version of displaying search result?
  - If we could aggregate those search results?
  - If we could help people to find what they are interested in faster?
  - If we could provide an exploratory visualization that people could quickly make sense of a large set of webpages?

# Data


Query keywords
Search Engine
Mention locations
50 webpages with full text
Related Named Entities
Related Wikipedia Category
Interactive viz for exploration

# Framing Searches

- Planning a trip to Barcelona Spain
  - What activities should I do
  - How do I get there
  - Where should I stay
- 2016 Presidential Election
  - Multiple categories of things: PEOPLE, PARTY, STATES, TIME
- Harry Potter
  - Complex plotlines with many characters
- ...

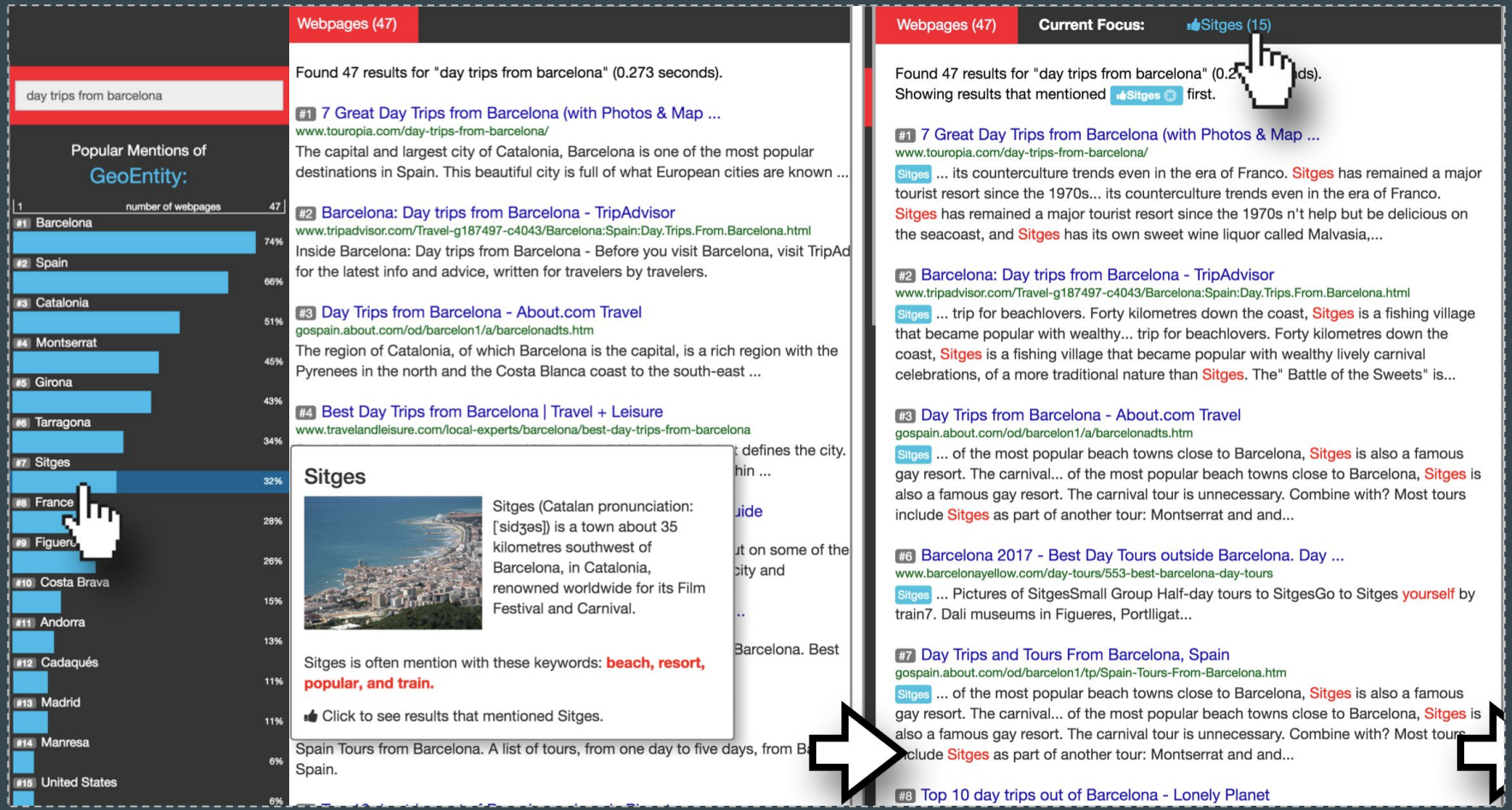
# Related Work

- Exploratory Search
  - Exploratory search: from finding to understanding (2006) by  Gary Marchionini
  - Concrete information need is not always present and information seekers usually engage in learning and investigation strategies instead of plain lookup of documents.
- Information Foraging
  - Information Foraging (2009) by Peter Pirolli
  - Information seekers tend to switch between topics and sources to gain information faster
- Faceted Search
  - User Interfaces for Search (2009) by Marti Hearst
  - Clustering versus faceted categories for information exploration (2006) by Marti Hearst
  - Metadata and facets can give exploratory searchers a good overview and drill-down mechanisms

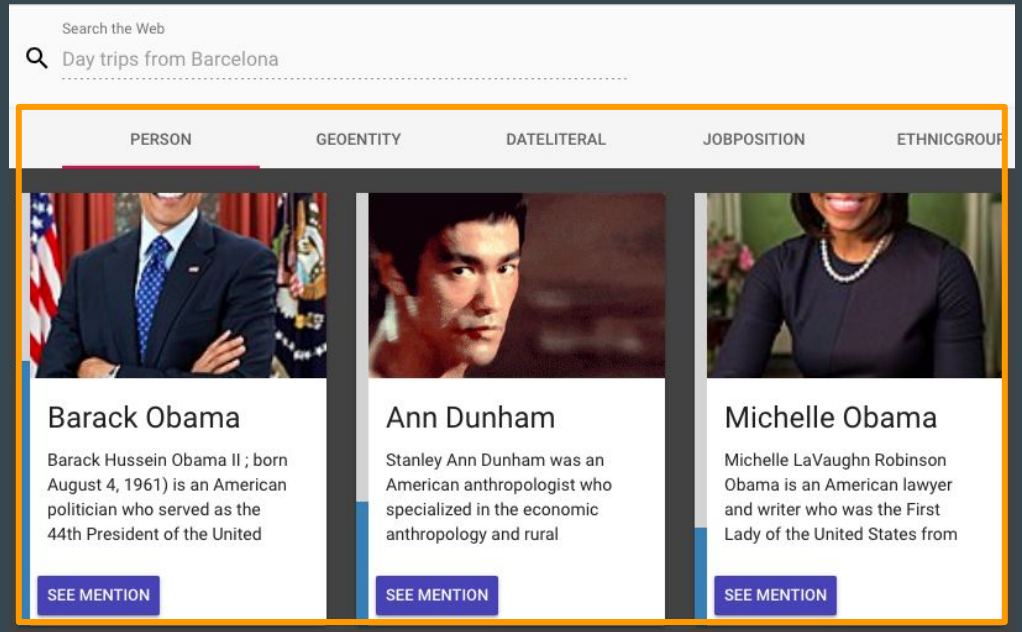
# Exploratory Data Analysis

- Grouping results in different categories only makes sense when the categories and subcategories/entities are appropriate.
  - GeoEntity is appropriate for trip planning, yet might not so appropriate for treatment options seeking
- Frequency of entity mention highly related to importance of the result web pages
- The existing subcategories don't include subjective judgments yet subjective judgments could be useful in some kind of searches
  - When people search treatment options for cancer, they not only want to know options, but also want to know if there is negative side effect
  - When people plan for a trip, they want to know if one activity/place is fun to visit too
- In the existing interface, it's not intuitive enough to switch between subcategories/entities

# Initial version of Interface

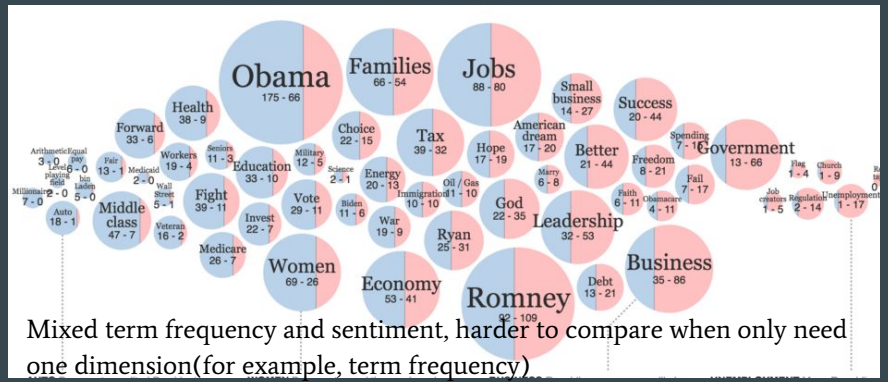


# Design Evolution - Entity Cards

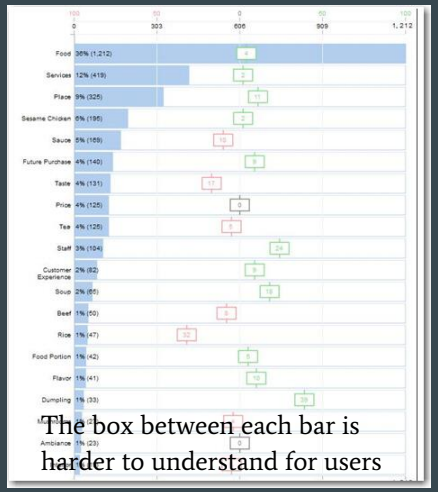



Shift the focus to the aggregated data to show the data better and better integrate verbal description

# Design Evolution - Design choices considered for sentiment

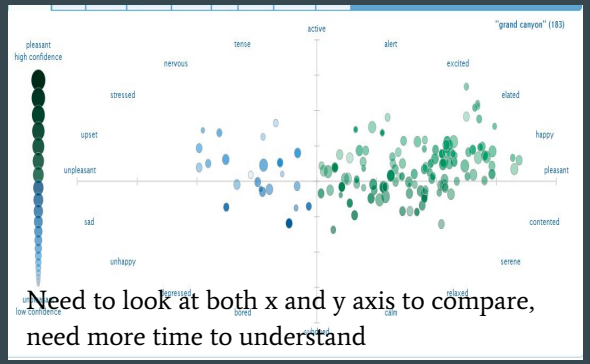



Mixed term frequency and sentiment, harder to compare when only need one dimension(for example, term frequency)

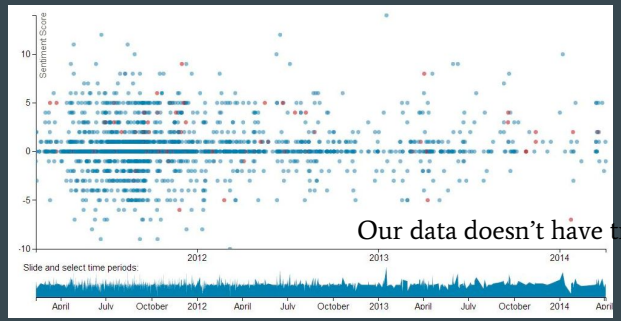


The box between each bar is harder to understand for users

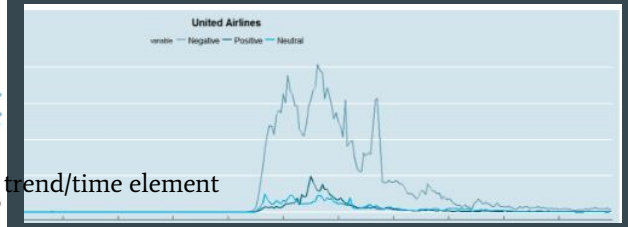
- Easy to compare
- Negative and positive is intuitive in context of sentiment analysis




Need to look at both x and y axis to compare, need more time to understand




Our data doesn't have trend/time element

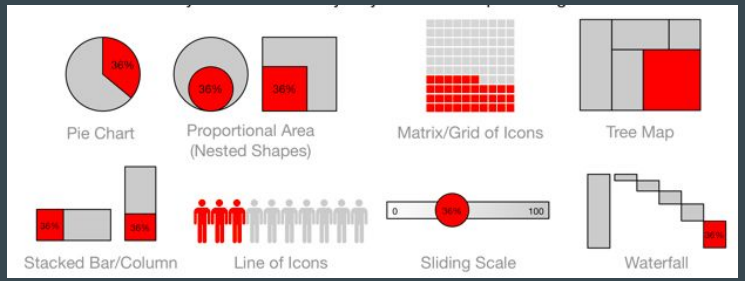


# Design Evolution - Sentimental Analysis

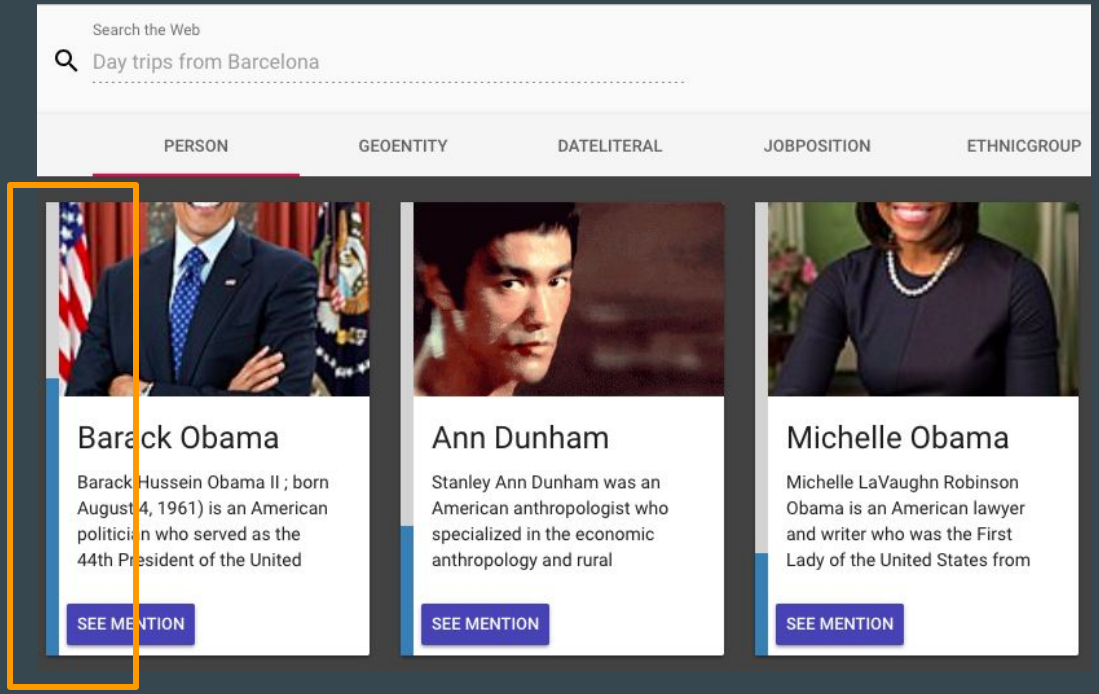
- Show sentiment analysis to present positive/negative judgments
- Show frequency of terms to allow user quickly know if the entity is relevant
- Present and reveal the data at several layers of detail

# Design Evolution - Frequency bar for each entity/term freq

Percentage frequency bar:
- encourage people to compare different cards
- present the number in a smaller space




Source: http://coolinfographics.com/blog/2016/1/7/three-simple-resolutions-to-design-better-dataviz.html



# Implementation

Demo link: http://joseph.nlpweb.org/SearchScapeVIz/

Video link: https://bit.ly/2I5ApUj

# Evaluation

- By using our visualizations, users can explore a large set of webpages easier and faster
- The visualization does present a large set of webpages in a concise way, and users could explore multiple layers of the result and understand it
- Possible improvement:
  - Refine entity list and categories
  - Test with users

# Overview and Motivation

- Goal: Allow individuals to gain an overview of a large set of webpages. Understand the popularity and distribution of information across those pages.
  - *Medical* - common symptoms of a disease, treatment options
  - *Trip Planning* - Common or unique activities, interesting locations to visit
  - ***Shopping*** - Common models / popularity of models, Features to be on the lookout for

- Tone: Reading
- Experience: Exhibitory / Exploratory

# Features

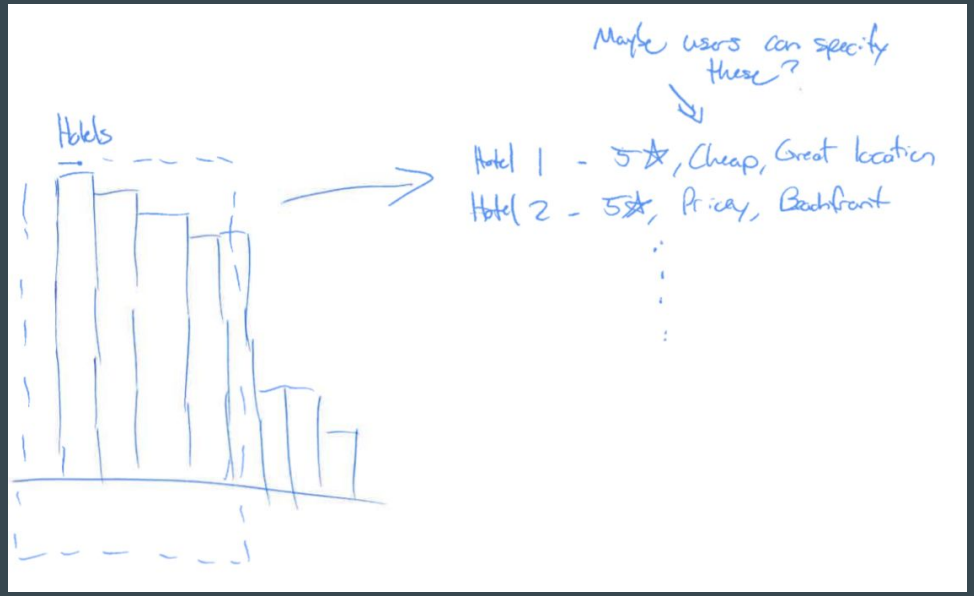
Overview

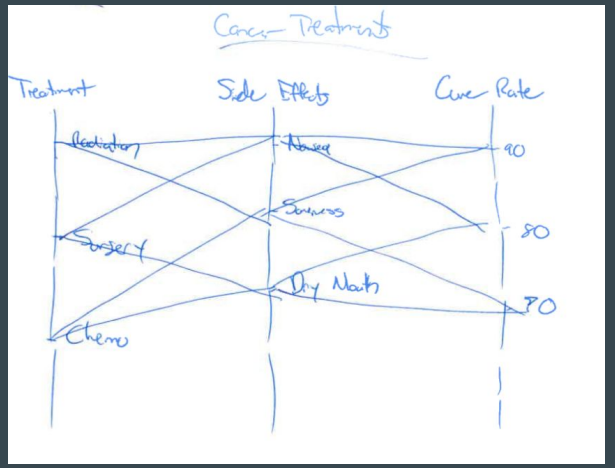
- Discover unknown topics (like what else do I have to consider while beekeeping)
- Some sort of graphical overview of distribution / popularity
- See what sources talk about what topics

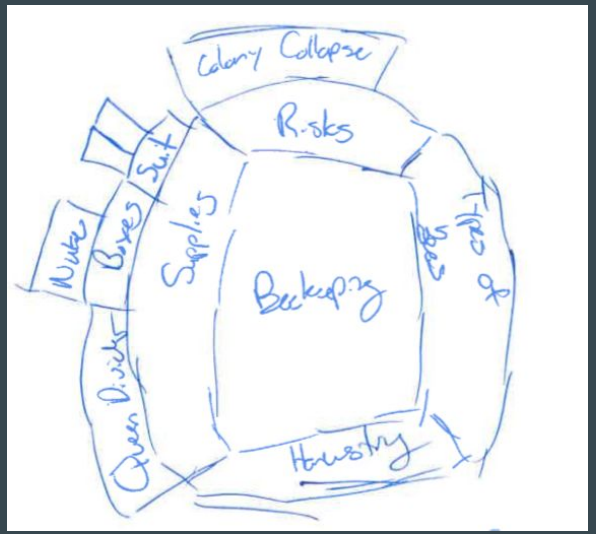
Comparisons

- What of these items is unique? Which are common?
- What are the features of these items (and how do they compare)

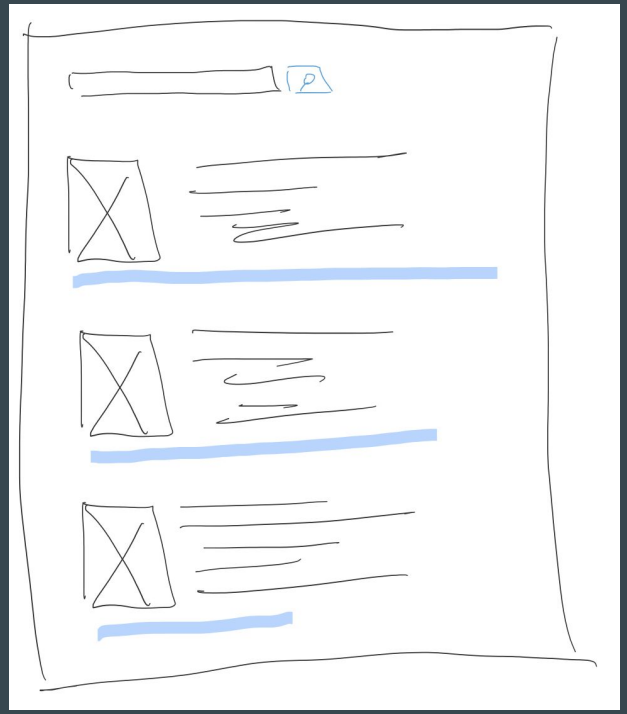
# Visual Designs - Potential Interfaces

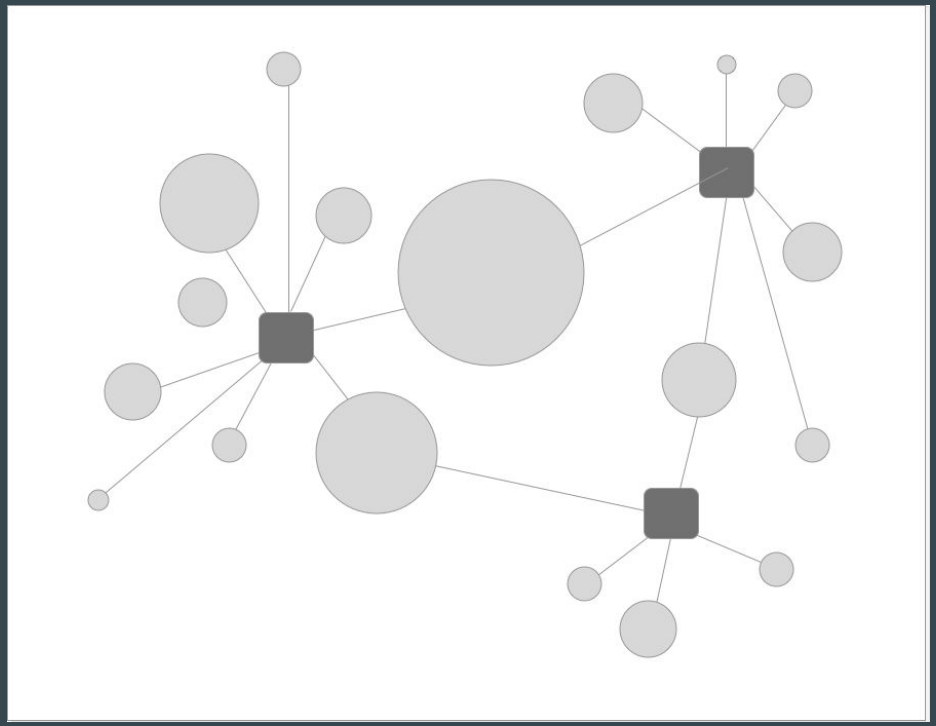


# Visual Designs - Potential Interfaces





# Project Schedule

April 10th - 15th -- HiFi Designs

April 15th - 20th -- Initial Implementation

April 20th - 26th -- Conference Travel

April 26th - May 1 - Feedback + Revisions