



# Module #9a: Text Visualization

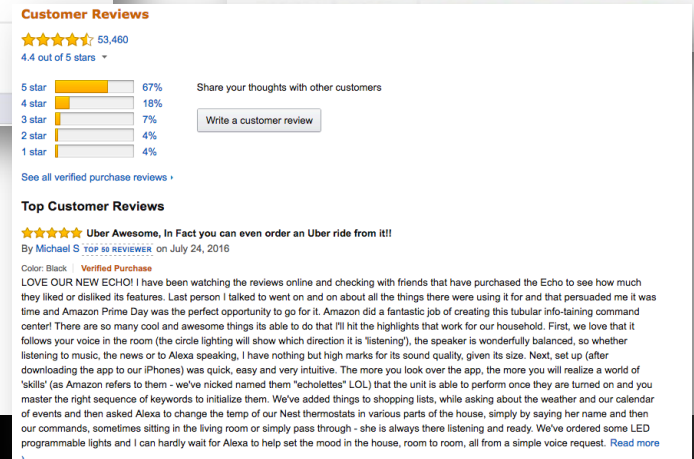
# Objectives

- Discuss challenges of visualizing unstructured data and large collection of text and documents
- Explain different techniques for representing words and concepts in a document
- Talk about the benefits and limitations of word clouds
- Discuss
  1. Text as Data
  2. Visualizing Document Content
  3. Evolving Documents
  4. Visualizing Conversation
  5. Document Collections



# Unstructured Text

- *Unstructured data*: information that either does not have a pre-defined data model or is not organized in a pre-defined manner.
- Documents are used as the primary communication and information artifact in our lives
  - Emails, letters, reviews, blogs, etc..





How data visualization can help users better understand and gather information from unstructured text and document collections?



# Why Visualize Text?

1. **Understanding / Summarization** – get an overview of the document
2. **Grouping** – cluster for overview or classification
3. **Comparison** – compare document collections, or inspect evolution of collection over time
4. **Correlation** – compare patterns in text to those in other data, e.g., correlate with social network



# Text Analytics: Our Understanding

## What is in my text?

<b>What's inside the Old Bailey Data?</b>	<b>What are the major causes of injuries?</b>	<b>What did my customers say about my hotels?</b>
197,000+ crime reports	100,000+ patient emergency room records	3000+ customer- posted reviews

## Information Understanding: Text Summarization



# Text Analytics: Our Understanding

## What is in my text?

<b>Which hotel features do my customers like/dislike</b>	<b>How customers' sentiment have changed toward my hotels</b>	<b>How do customers feel about my new product launch</b>
3000+ customer reviews	3000+ customer-posted reviews	thousands of e-opinion postings

## Insight Discovery: Sentiment Analysis



# Text Analytics: Our Understanding

## What is in my text?

<b>What's the correlation between crime and year:</b>  197,000+ crime reports	<b>What are the correlations of patient gender and the cause of injury</b>  70,000+ patient emergency room records	<b>Compare the customer's attitude toward our product vs our competitors</b>  3000+ customer-posted reviews
---	--	---

**Decision Making and Problem Solving: Text Analysis**





# Text Visualization Challenges

- High Dimensionality
  - Huge amounts of complex information
  - Understanding the meaning of free text is hard
- Context & Semantics
  - Provide relevant context to aid understanding. Show (or provide access to) the source text.
- Modeling Abstraction
  - Determine your analysis task.
  - Understand abstraction of your language models. Match analysis task with appropriate tools and models.



# Visualization Techniques

1. Text as Data
2. Visualizing Document Content
3. Evolving Documents
4. Visualizing Conversation
5. Document Collections