Data Visualization (CIS 490/680)

D3

Dr. David Koop
Dataset Types

- **Tables**
  - Attributes (columns)
  - Items (rows)
  - Cell containing value

- **Networks**
  - Link
  - Node (item)

- **Fields (Continuous)**
  - Grid of positions
  - Cell
  - Attributes (columns)
  - Value in cell

- **Geometry (Spatial)**
  - Position

- **Multidimensional Table**
  - Key 1
  - Key 2
  - Attributes
  - Value in cell

- **Trees**
song structure and singing

Joe Carmanica recently wrote about this trend for the New York Times, arguing that it was led by Drake, who popularized the rapping-and-singing formula over the past decade.

A better benchmark for Lil Uzi Vert's word count (2,556) might be those of pop artists, such as Beyonce (2,433 words) or even one of his major influences: Marilyn Manson (2,466 words).

There are also genre-bending artists. If Childish Gambino's Awaken, My Love! is less hip hop in the traditional '90s boom-bap sense, is it fair to compare it to vocabulary-dense Wu-Tang albums? Genre matters in vocabulary calculations—check out the chart below, which takes 500 random samples of 35,000 words from rock, country, and hip hop.

In short, if artists depart from hip-hop song structure, we'd expect their vocabulary to go down in the number of unique words. That said, the results are still directionally interesting. Of the 150 artists in the dataset, let's take a look at who is on top.

# of Unique Words Used in 500 Random Samples of 35,000 Lyrics from Country, Rock, Hip Hop

Sets & Lists

Raw Lyrics Data via John W. Miller

[M. Daniels, 2019]
Categorial, Ordinal, and Quantitative

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>S</th>
<th>T</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>10/14/06</td>
<td>5-Low</td>
<td>Large Box</td>
<td>0.8</td>
<td>10/21/06</td>
</tr>
<tr>
<td>6</td>
<td>2/21/08</td>
<td>4-Not Specified</td>
<td>Small Pack</td>
<td>0.55</td>
<td>2/22/08</td>
</tr>
<tr>
<td>32</td>
<td>7/16/07</td>
<td>2-High</td>
<td>Small Pack</td>
<td>0.79</td>
<td>7/17/07</td>
</tr>
<tr>
<td>32</td>
<td>7/16/07</td>
<td>2-High</td>
<td>Jumbo Box</td>
<td>0.72</td>
<td>7/17/07</td>
</tr>
<tr>
<td>32</td>
<td>7/16/07</td>
<td>2-High</td>
<td>Medium Box</td>
<td>0.6</td>
<td>7/18/07</td>
</tr>
<tr>
<td>32</td>
<td>7/16/07</td>
<td>2-High</td>
<td>Medium Box</td>
<td>0.65</td>
<td>7/18/07</td>
</tr>
<tr>
<td>35</td>
<td>10/23/07</td>
<td>4-Not Specified</td>
<td>Wrap Bag</td>
<td>0.52</td>
<td>10/24/07</td>
</tr>
<tr>
<td>35</td>
<td>10/23/07</td>
<td>4-Not Specified</td>
<td>Small Box</td>
<td>0.58</td>
<td>10/25/07</td>
</tr>
<tr>
<td>36</td>
<td>11/3/07</td>
<td>1-Urgent</td>
<td>Small Box</td>
<td>0.55</td>
<td>11/3/07</td>
</tr>
<tr>
<td>65</td>
<td>3/18/07</td>
<td>1-Urgent</td>
<td>Small Pack</td>
<td>0.49</td>
<td>3/19/07</td>
</tr>
<tr>
<td>66</td>
<td>1/20/05</td>
<td>5-Low</td>
<td>Wrap Bag</td>
<td>0.56</td>
<td>1/20/05</td>
</tr>
<tr>
<td>69</td>
<td>6/4/05</td>
<td>4-Not Specified</td>
<td>Small Pack</td>
<td>0.44</td>
<td>6/6/05</td>
</tr>
<tr>
<td>69</td>
<td>6/4/05</td>
<td>4-Not Specified</td>
<td>Small Pack</td>
<td>0.6</td>
<td>6/6/05</td>
</tr>
<tr>
<td>70</td>
<td>12/18/06</td>
<td>5-Low</td>
<td>Wrap Bag</td>
<td>0.59</td>
<td>12/23/06</td>
</tr>
<tr>
<td>70</td>
<td>12/18/06</td>
<td>5-Low</td>
<td>Small Box</td>
<td>0.55</td>
<td>12/23/06</td>
</tr>
<tr>
<td>96</td>
<td>4/17/05</td>
<td>2-High</td>
<td>Small Box</td>
<td>0.55</td>
<td>4/19/05</td>
</tr>
<tr>
<td>97</td>
<td>1/29/06</td>
<td>3-Medium</td>
<td>Small Box</td>
<td>0.38</td>
<td>1/30/06</td>
</tr>
<tr>
<td>129</td>
<td>11/19/08</td>
<td>5-Low</td>
<td>Small Box</td>
<td>0.37</td>
<td>11/28/08</td>
</tr>
<tr>
<td>130</td>
<td>5/8/08</td>
<td>2-High</td>
<td>Medium Box</td>
<td>0.38</td>
<td>5/9/08</td>
</tr>
<tr>
<td>130</td>
<td>5/8/08</td>
<td>2-High</td>
<td>Small Box</td>
<td>0.38</td>
<td>5/10/08</td>
</tr>
<tr>
<td>130</td>
<td>5/8/08</td>
<td>2-High</td>
<td>Small Box</td>
<td>0.6</td>
<td>5/11/08</td>
</tr>
<tr>
<td>132</td>
<td>6/11/06</td>
<td>3-Medium</td>
<td>Medium Box</td>
<td>0.6</td>
<td>6/12/06</td>
</tr>
<tr>
<td>132</td>
<td>6/11/06</td>
<td>3-Medium</td>
<td>Jumbo Box</td>
<td>0.69</td>
<td>6/14/06</td>
</tr>
<tr>
<td>134</td>
<td>5/1/08</td>
<td>4-Not Specified</td>
<td>Large Box</td>
<td>0.82</td>
<td>5/3/08</td>
</tr>
<tr>
<td>135</td>
<td>10/21/07</td>
<td>4-Not Specified</td>
<td>Small Pack</td>
<td>0.64</td>
<td>10/23/07</td>
</tr>
<tr>
<td>166</td>
<td>9/12/07</td>
<td>2-High</td>
<td>Small Box</td>
<td>0.55</td>
<td>9/14/07</td>
</tr>
<tr>
<td>193</td>
<td>8/8/06</td>
<td>1-Urgent</td>
<td>Medium Box</td>
<td>0.57</td>
<td>8/10/06</td>
</tr>
<tr>
<td>194</td>
<td>4/5/08</td>
<td>3-Medium</td>
<td>Wrap Bag</td>
<td>0.42</td>
<td>4/7/08</td>
</tr>
</tbody>
</table>
**Ordering Direction**

- **Sequential**
- **Diverging**
- **Cyclic**

---

[Munzner (ill. Maguire), 2014; Rogowitz & Treinish, 1998; Weber et al., 2001]
Tasks

What?

Why?

How?

### Actions

- **Analyze**
  - Consume
    - Discover
    - Present
    - Enjoy
  - Produce
    - Annotate
    - Record
    - Derive

### Why?

- **All Data**
  - Trends
  - Outliers
  - Features

- **Attributes**
  - One
    - Distribution
    - Extremes
  - Many
    - Dependency
    - Correlation
    - Similarity

### Search

<table>
<thead>
<tr>
<th>Location known</th>
<th>Target known</th>
<th>Target unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location known</td>
<td><strong>Lookup</strong></td>
<td><strong>Browse</strong></td>
</tr>
<tr>
<td>Location unknown</td>
<td><strong>Locate</strong></td>
<td><strong>Explore</strong></td>
</tr>
</tbody>
</table>

### Query

- **Identify**
- **Compare**
- **Summarize**

### Network Data

- **Topology**
  - Paths

### Spatial Data

- **Shape**

[Munzner (ill. Maguire), 2014]
Actions: Analyze

- **Consume**
  - Discover

- **Present**

- **Enjoy**

- **Produce**
  - Annotate
  - Record
  - Derive

[Munzner (ill. Maguire), 2014]
Visualization for Consumption

Discover

Present

Enjoy

[M. Stefaner, M. Wattenberg]
Memorability

Figure 6: Policy shifts and improvements to mobile health practices to accommodate issues of emergency service and farmer health.

[M. Borkin et al., InfoVis 2015]
Memorability of Visualizations

- S. Few: "Visualizations don’t need to be designed for memorability – they need to be designed for comprehension. For most visualizations, the comprehension that they provide need only last until the decision that it informs is made. Usually, that is only a matter of seconds."

- B. Jones (paraphrased): People make decisions using visualizations but this isn't instantaneous like robots or algorithms; they often chew on a decision for a while

- R. Kosara: there are cases where people benefit from remembering a visualization (e.g. health-related visualization)

- Are there tradeoffs between the characteristics?
Visualization for Production

- Generate new material
- Annotate:
  - Add more to a visualization
  - Usually associated with text, but can be graphical
- Record:
  - Persist visualizations for historical record
  - Provenance (graphical histories): how did I get here?
- Derive (Transform):
  - Create new data
  - Create derived attributes (e.g. mathematical operations, aggregation)
Visualization for Production: Derived Data

Original Data

Derived Data

trade balance = exports − imports

[Munzner (ill. Maguire), 2014]
Actions: Search

- What does a user know?
  - Lookup: check bearings
  - Locate: find on a map
  - Browse: what’s nearby
  - Explore: where to go
  - Patterns

<table>
<thead>
<tr>
<th>Location known</th>
<th>Target known</th>
<th>Target unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lookup</td>
<td>Browse</td>
</tr>
<tr>
<td>Location unknown</td>
<td>Locate</td>
<td>Explore</td>
</tr>
</tbody>
</table>

[Munzner (ill. Maguire), 2014]
Actions: Query

- Identify: characteristics or references
- Compare: similarities and differences
- Summarize: overview of everything

- Number of targets: One, Some (Often 2), or All
- Identify: characteristics or references
- Compare: similarities and differences
- Summarize: overview of everything

[Munzner (ill. Maguire), 2014]
Assignment 2

- Link
- Three parts: table, horizontal bar chart, vertical bar chart
  - data processing
  - highlighting (CS 680)
- Vertical chart can be tricky
- Start early!
- Questions?

![Bar Chart Example]

# Inspections

0

Jan 2017

Month

Jun 2019
Roadmap

- **What?** → Data
  - Types
  - Semantics

- **Why?** → Tasks
  - Actions
  - Targets

- **How** → Vis Idioms/Techniques
  - Data Representation
  - Visual Encoding
  - Interaction Encoding
Analysis Example: Different “Idioms”

[SpaceTree, Grosjean et al.]

[TreeJuxtaposer, Munzner et al.]
“Idiom” Comparison

SpaceTree

TreeJuxtaposer

Actions
- Present
- Locate
- Identify

Targets
- Path between two nodes

Why?
- Encode
- Navigate
- Select
- Filter
- Aggregate

SpaceTree

TreeJuxtaposer

How?
- Encode
- Navigate
- Select
- Arrange

[Munzner (ill. Maguire), 2014]
Analysis Example: Derivation

- Strahler number
  - centrality metric for trees/networks
  - derived quantitative attribute
  - draw top 5K of 500K for good skeleton

d3.js

Data-Driven Documents
Data-Driven Documents (D3)

- Open-Source JavaScript Library
- [http://d3js.org/](http://d3js.org/)
- Original Authors: Mike Bostock, Vadim Ogievetsky, and Jeff Heer
- Focus on Web standards, customization, and usability
- Grew from work on Protovis: more standard, more interactive
- By nature, a **low-level** library; you have control over all elements and styles
- A top project on GitHub (over 85,000 stars as of Sept. 2019)
- Lots of impressive examples
  - Bostock was a New York Times Graphics Editor
  - [https://bost.ocks.org/mike/](https://bost.ocks.org/mike/) and [https://observablehq.com/@mbostock](https://observablehq.com/@mbostock)
D3 Key Features

- Supports data as a core piece of Web elements
  - Loading data
  - Dealing with changing data (joins, enter/update/exit)
  - Correspondence between data and DOM elements
- Selections (similar to CSS) that allow greater manipulation
- Method Chaining
- Integrated layout algorithms, axes calculations, etc.
- Focus on interaction support
  - Straightforward support for transitions
  - Event handling support for user-initiated changes
D3 Introduction

• Ogievetsky has put together a nice set of interactive examples that show off the major features of D3

• http://dakoop.github.io/IntroD3/
  - (Updated from original for D3 v5 with new joins)

• https://beta.observablehq.com/@dakoop/d3-intro

• Other references:
  - Murrary’s book on Interactive Data Visualization for the Web
  - The D3 website: d3js.org
  - Ros's Slides on v4: https://iros.github.io/d3-v4-whats-new/
D3 Data Joins

- Two groups: data and visual elements
- Three parts of the join between them: enter, update, and exit
- enter: \( s.\text{enter}() \), update: \( s \), exit: \( s.\text{exit}() \)
Merge vs. Join

- Merge creates a new selection that includes the items from both selections
  - If you want to update all elements (including those just added via enter), use merge!
  - Useful when enter+update have similar transitions

- Join allows you to modify different parts of the selection in a single statement
  - Also will create the final selection
  - Does enter+append and exit+remove automatically
  - Pass functions to modify the enter, update, and exit parts of the selection
Transitions

- Nested transitions (those that "hang off" of a parent transition) follow immediately after the parent transition