Visualization Tools

Jeffrey Heer  University of Washington
How do people create visualizations?

**Chart Typology**
Pick from a stock of templates
Easy-to-use but limited expressiveness
Prohibits novel designs, new data types

**Component Architecture**
Permits more combinatorial possibilities
Novel views require new operators, which requires software engineering
Graphics APIs
Processing, OpenGL, Java2D
ey = y;
size = s;
}

void update(int mx, int my) {
    angle = atan2(my-ey, mx-ex);
}

void display() {
    pushMatrix();
    translate(ex, ey);
    fill(255);
    ellipse(0, 0, size, size);
    rotate(angle);
    fill(153);
    ellipse(size/4, 0, size/2, size/2);
    popMatrix();
}
Graphics APIs
Processing, OpenGL, Java2D
Component Architectures
Prefuse, Flare, Improvise, VTK

Graphics APIs
Processing, OpenGL, Java2D
Data State Model

[Chi 98]
Prefuse & Flare

Operator-based toolkits for visualization design

Vis = (Input Data -> Visual Objects) + Operators

Prefuse (http://prefuse.org)  Flare (http://flare.prefuse.org)
Component Architectures
Prefuse, Flare, Improvise, VTK

Graphics APIs
Processing, OpenGL, Java2D
Chart Typologies
Excel, Many Eyes, Google Charts

Component Architectures
Prefuse, Flare, Improvise, VTK

Graphics APIs
Processing, OpenGL, Java2D
Chart Typologies
## Data Sets: State Quick Facts

*Uploaded By: zinggoat*  
*Source: US Census Bureau*  
*Description:*  
*Tags: people census*

<table>
<thead>
<tr>
<th><strong>People QuickFacts</strong></th>
<th><strong>Population 2005 estimate</strong></th>
<th><strong>Population percent change April 1 2000 to July 1 2005</strong></th>
<th><strong>Population 2000</strong></th>
<th><strong>Population percent change 1990 to 2000</strong></th>
<th><strong>Persons under 5 years old percent 2004</strong></th>
<th><strong>Persons under 18 years old percent 2004</strong></th>
<th><strong>Persons 65 years old and over percent 2004</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Alabama</td>
<td>4557808</td>
<td>0.03</td>
<td>4447100</td>
<td>0.1</td>
<td>0.07</td>
<td>0.24</td>
<td>0.13</td>
</tr>
<tr>
<td>2 Alaska</td>
<td>663661</td>
<td>0.06</td>
<td>626932</td>
<td>0.14</td>
<td>0.08</td>
<td>0.29</td>
<td>0.06</td>
</tr>
<tr>
<td>3 Arizona</td>
<td>5939292</td>
<td>0.16</td>
<td>5130632</td>
<td>0.4</td>
<td>0.08</td>
<td>0.27</td>
<td>0.13</td>
</tr>
<tr>
<td>4 Arkansas</td>
<td>2779154</td>
<td>0.04</td>
<td>2673400</td>
<td>0.14</td>
<td>0.07</td>
<td>0.25</td>
<td>0.14</td>
</tr>
<tr>
<td>5 California</td>
<td>36132147</td>
<td>0.07</td>
<td>33871648</td>
<td>0.14</td>
<td>0.07</td>
<td>0.27</td>
<td>0.11</td>
</tr>
<tr>
<td>6 Colorado</td>
<td>4665177</td>
<td>0.08</td>
<td>4301261</td>
<td>0.31</td>
<td>0.07</td>
<td>0.26</td>
<td>0.1</td>
</tr>
<tr>
<td>7 Connecticut</td>
<td>3510297</td>
<td>0.03</td>
<td>3405565</td>
<td>0.04</td>
<td>0.06</td>
<td>0.24</td>
<td>0.14</td>
</tr>
<tr>
<td>8 Delaware</td>
<td>843524</td>
<td>0.08</td>
<td>783600</td>
<td>0.18</td>
<td>0.07</td>
<td>0.23</td>
<td>0.13</td>
</tr>
<tr>
<td>9 Florida</td>
<td>17789664</td>
<td>0.11</td>
<td>15982378</td>
<td>0.24</td>
<td>0.06</td>
<td>0.23</td>
<td>0.17</td>
</tr>
<tr>
<td>10 Georgia</td>
<td>9072576</td>
<td>0.11</td>
<td>8186453</td>
<td>0.26</td>
<td>0.08</td>
<td>0.26</td>
<td>0.1</td>
</tr>
<tr>
<td>11 Hawaii</td>
<td>1275194</td>
<td>0.05</td>
<td>1211537</td>
<td>0.09</td>
<td>0.07</td>
<td>0.24</td>
<td>0.14</td>
</tr>
<tr>
<td>12 Idaho</td>
<td>1429096</td>
<td>0.1</td>
<td>1293953</td>
<td>0.29</td>
<td>0.07</td>
<td>0.27</td>
<td>0.11</td>
</tr>
<tr>
<td>13 Illinois</td>
<td>12763371</td>
<td>0.03</td>
<td>12419293</td>
<td>0.09</td>
<td>0.07</td>
<td>0.26</td>
<td>0.12</td>
</tr>
</tbody>
</table>
Choosing a visualization type for State Quick Facts

Analyze a text

**Tag Cloud**
How are you using your words? This enhanced tag cloud will show you the words popularity in the given set of text.
Learn more

**Wordle**
Wordle is a toy for generating ‘word clouds’ from text that you provide. The clouds give greater prominence to words that appear more frequently in the source text.
Learn more

**Word Tree**
See a branching view of how a word or phrase is used in a text. Navigate the text by zooming and clicking.
Learn more

Compare a set of values

**Bar Chart**
How do the items in your data set stack up? A bar chart is a simple and recognizable way to compare values. You can display several sets of bars for multivariate comparisons.
Learn more

**Block Histogram**
This versatile chart lets you get a quick sense of how a single set of data is distributed. Each item in the data is an individually identifiable block.
Learn more
Every Wednesday, when I get home from school, I have a piano lesson. My teacher is a very strict house. Her name is Hillary Clinton. Our piano is a Steinway Concert tree. It also has 88 cups. It also has a soft pedal and a/an Smiley pedal. When I have a lesson, I sit down on the piano Alberto and play for 16 minutes. I do scales to exercise my cats, and then I usually play a minuet by Johann Sebastian Washington. Teacher says I am a natural Haunted House and have a good musical leg. Perhaps when I get better I will become a concert vet and give a recital at Carnegie hospital.
[M]ost charting packages channel user requests into a **rigid array of chart types**. To atone for this lack of flexibility, they offer a kit of post-creation editing tools to return the image to what the user originally envisioned. **They give the user an impression of having explored data rather than the experience.**

Leland Wilkinson

Chart Typologies
Excel, Many Eyes, Google Charts

Component Architectures
Prefuse, Flare, Improvise, VTK

Graphics APIs
Processing, OpenGL, Java2D
Chart Typologies
Excel, Many Eyes, Google Charts

Visual Analysis Grammars
VizQL, ggplot2

Component Architectures
Prefuse, Flare, Improvise, VTK

Graphics APIs
Processing, OpenGL, Java2D
Statistics and Computing

Leland Wilkinson

The Grammar of Graphics

Second Edition

Springer
ggplot(diamonds, aes(x=price, fill=cut)) + geom_bar(position="dodge")
```
qplot(long, lat, data = expo, geom = "tile", fill = ozone,
     facets = year ~ month) +
scale_fill_gradient(low = "white", high = "black") + map
```
Chart Typologies
Excel, Many Eyes, Google Charts

Visual Analysis Grammars
VizQL, ggplot2

Component Architectures
Prefuse, Flare, Improvise, VTK

Graphics APIs
Processing, OpenGL, Java2D
Chart Typologies
  Excel, Many Eyes, Google Charts

Visual Analysis Grammars
  VizQL, ggplot2

Visualization Grammars
  Protovis, D3.js

Component Architectures
  Prefuse, Flare, Improvise, VTK

Graphics APIs
  Processing, OpenGL, Java2D
Protovis & D3
Today's first task is not to invent wholly new graphical techniques, though these are needed. Rather we need most vitally to recognize and reorganize the essential of old techniques, to make easy their assembly in new ways, and to modify their external appearances to fit the new opportunities.

J. W. Tukey, M. B. Wilk

Data Analysis & Statistics, 1965
Protovis: A Grammar for Visualization

A graphic is a composition of data-representative marks.

with Mike Bostock & Vadim Ogievetsky
MARKS: Protovis graphical primitives
<table>
<thead>
<tr>
<th>MARK</th>
<th>$\lambda : D \to R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>data</td>
<td>$\lambda$</td>
</tr>
<tr>
<td>visible</td>
<td>$\lambda$</td>
</tr>
<tr>
<td>left</td>
<td>$\lambda$</td>
</tr>
<tr>
<td>bottom</td>
<td>$\lambda$</td>
</tr>
<tr>
<td>width</td>
<td>$\lambda$</td>
</tr>
<tr>
<td>height</td>
<td>$\lambda$</td>
</tr>
<tr>
<td>fillStyle</td>
<td>$\lambda$</td>
</tr>
<tr>
<td>strokeStyle</td>
<td>$\lambda$</td>
</tr>
<tr>
<td>lineWidth</td>
<td>$\lambda$</td>
</tr>
<tr>
<td>...</td>
<td>$\lambda$</td>
</tr>
<tr>
<td>data</td>
<td>1</td>
</tr>
<tr>
<td>---------</td>
<td>----</td>
</tr>
<tr>
<td>visible</td>
<td>true</td>
</tr>
<tr>
<td>left</td>
<td>$\lambda$: index * 25</td>
</tr>
<tr>
<td>bottom</td>
<td>0</td>
</tr>
<tr>
<td>width</td>
<td>20</td>
</tr>
<tr>
<td>height</td>
<td>$\lambda$: datum * 80</td>
</tr>
<tr>
<td>fillStyle</td>
<td>blue</td>
</tr>
<tr>
<td>strokeStyle</td>
<td>black</td>
</tr>
<tr>
<td>lineWidth</td>
<td>1.5</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
### Data Table

<table>
<thead>
<tr>
<th>Key</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>1 1.2 1.7 1.5 0.7</td>
</tr>
<tr>
<td>Visible</td>
<td>true</td>
</tr>
<tr>
<td>Left</td>
<td>0 * 25</td>
</tr>
<tr>
<td>Bottom</td>
<td>0</td>
</tr>
<tr>
<td>Width</td>
<td>20</td>
</tr>
<tr>
<td>Height</td>
<td>1 * 80</td>
</tr>
<tr>
<td>FillStyle</td>
<td>blue</td>
</tr>
<tr>
<td>StrokeStyle</td>
<td>black</td>
</tr>
<tr>
<td>LineWidth</td>
<td>1.5</td>
</tr>
</tbody>
</table>

...
### RECT

<table>
<thead>
<tr>
<th>data</th>
<th>1</th>
<th>1.2</th>
<th>1.7</th>
<th>1.5</th>
<th>0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>visible</td>
<td>true</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>left</td>
<td>1 * 25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bottom</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>width</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>height</td>
<td>1.2 * 80</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fillStyle</td>
<td>blue</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>strokeStyle</td>
<td>black</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lineWidth</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \lambda : D \rightarrow R \]
<table>
<thead>
<tr>
<th>RECT</th>
<th>( \lambda : D \rightarrow R )</th>
</tr>
</thead>
<tbody>
<tr>
<td>data</td>
<td>1</td>
</tr>
<tr>
<td>visible</td>
<td>true</td>
</tr>
<tr>
<td>left</td>
<td>(2 \times 25)</td>
</tr>
<tr>
<td>bottom</td>
<td>0</td>
</tr>
<tr>
<td>width</td>
<td>20</td>
</tr>
<tr>
<td>height</td>
<td>(1.7 \times 80)</td>
</tr>
<tr>
<td>fillStyle</td>
<td>blue</td>
</tr>
<tr>
<td>strokeStyle</td>
<td>black</td>
</tr>
<tr>
<td>lineWidth</td>
<td>1.5</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
### RECT

<table>
<thead>
<tr>
<th>data</th>
<th>1</th>
<th>1.2</th>
<th>1.7</th>
<th>1.5</th>
<th>0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>visible</td>
<td>true</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>left</td>
<td>3 * 25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bottom</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>width</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>height</td>
<td>1.5 * 80</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fillStyle</td>
<td>blue</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>strokeStyle</td>
<td>black</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lineWidth</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \lambda : D \rightarrow R \]
<table>
<thead>
<tr>
<th>RECT</th>
<th>( \lambda : D \rightarrow R )</th>
</tr>
</thead>
<tbody>
<tr>
<td>data</td>
<td>1 \hspace{1em} 1.2 \hspace{1em} 1.7 \hspace{1em} 1.5 \hspace{1em} 0.7</td>
</tr>
<tr>
<td>visible</td>
<td>true</td>
</tr>
<tr>
<td>left</td>
<td>4 \times 25</td>
</tr>
<tr>
<td>bottom</td>
<td>0</td>
</tr>
<tr>
<td>width</td>
<td>20</td>
</tr>
<tr>
<td>height</td>
<td>0.7 \times 80</td>
</tr>
<tr>
<td>fillStyle</td>
<td>blue</td>
</tr>
<tr>
<td>strokeStyle</td>
<td>black</td>
</tr>
<tr>
<td>lineWidth</td>
<td>1.5</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>RECT</td>
<td>$\lambda : D \to R$</td>
</tr>
<tr>
<td>------</td>
<td>-----------------</td>
</tr>
<tr>
<td>data</td>
<td>1 1.2 1.7 1.5 0.7</td>
</tr>
<tr>
<td>visible</td>
<td>true</td>
</tr>
<tr>
<td>left</td>
<td>$\lambda$: index * 25</td>
</tr>
<tr>
<td>bottom</td>
<td>0</td>
</tr>
<tr>
<td>width</td>
<td>20</td>
</tr>
<tr>
<td>height</td>
<td>$\lambda$: datum * 80</td>
</tr>
<tr>
<td>fillStyle</td>
<td>blue</td>
</tr>
<tr>
<td>strokeStyle</td>
<td>black</td>
</tr>
<tr>
<td>lineWidth</td>
<td>1.5</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
var vis = new pv.Panel();
vis.add(pv.Bar)
  .data([1, 1.2, 1.7, 1.5, 0.7])
  .visible(true)
  .left((d) => this.index * 25);
  .bottom(0)
  .width(20)
  .height((d) => d * 80)
  .fillStyle("blue")
  .strokeStyle("black")
  .lineWidth(1.5);
vis.render();
var army = pv.nest(napoleon.army, "dir", "group");
var vis = new pv.Panel();

var lines = vis.add(pv.Panel).data(army);
lines.add(pv.Line)
  .data(() => army[this.idx])
  .left(lon).top(lat).size((d) => d.size/8000)
  .strokeStyle(() => color[army[paneIndex][0].dir]);

vis.add(pv.Label).data(napoleon.cities)
  .left(lon).top(lat)
  .text((d) => d.city).font("italic 10px Georgia")
  .textAlign("center").textBaseline("middle");

vis.add(pv.Label).data(napoleon.temp)
  .left(lon).top(tmp).strokeStyle("#0")
  .text((d) => d.temp+"°").textBaseline("top").font("italic 10px Georgia");

var army = pv.nest(napoleon.army, "dir", "group");
var vis = new pv.Panel();

var lines = vis.add(pv.Panel).data(army);
lines.add(pv.Line)
  .data(() => army[this.idx])
  .left(lon).top(lat).size((d) => d.size/8000)
  .strokeStyle(() => color[army[paneIndex][0].dir]);

vis.add(pv.Label).data(napoleon.cities)
  .left(lon).top(lat)
  .text((d) => d.city).font("italic 10px Georgia")
  .textAlign("center").textBaseline("middle");

vis.add(pv.Label).data(napoleon.temp)
  .left(lon).top(tmp).strokeStyle("#0")
  .text((d) => d.temp+"°").textBaseline("top").font("italic 10px Georgia");
Obesity Map | Vadim Ogievetsky
Obesity Map | Vadim Ogievetsky
FlickrSeason | Ken-Ichi Ueda
d3.js  Data-Driven Documents

with Mike Bostock & Vadim Ogievetsky
Protovis

Specialized mark types
+ Streamlined design
- Limits expressiveness
- More overhead (slower)
- Harder to debug
- Self-contained model

Specify a scene (nouns)
+ Quick for static vis
- Delayed evaluation
- Animation, interaction are more cumbersome
Protovis

*Specialized mark types*
+ Streamlined design
- Limits expressiveness
- More overhead (slower)
- Harder to debug
- Self-contained model

*Specify a scene (nouns)*
+ Quick for static vis
- Delayed evaluation
- Animation, interaction are more cumbersome

D3

*Bind data to DOM*
- Exposes SVG/CSS/…
+ Exposes SVG/CSS/…
+ Less overhead (faster)
+ Debug in browser
+ Use with other tools

*Transform a scene (verbs)*
- More complex model
+ Immediate evaluation
+ Dynamic data, anim, and interaction natural
The core abstraction in D3 is a selection.

// Add and configure an SVG element
var svg = d3.append("svg")
  .attr("width", 500)
  .attr("height", 300);

// Select & update existing rectangles contained in the SVG element
svg.selectAll("rect")
  .attr("width", 100)
  .style("fill", "steelblue");
Selections can *bind* data and DOM elements.

```javascript
var values = [ { }, { }, { }, ... ]; // input data as JS objects

// Select SVG rectangles and bind them to data values.
var bars = svg.selectAll("rect.bars").data(values);

// What if the DOM elements don’t exist yet? The `enter` set represents data values that do not yet have matching DOM elements.
bars.enter().append("rect").attr("class", "bars");

// What if data values are removed? The `exit` set is a selection of existing DOM elements who no longer have matching data values.
bars.exit().remove();
```
The Data Join

```javascript
var s = d3.selectAll(...).data(...)
```

- **ENTER**
  - Data values without matching DOM elements.
  - `s.enter().append(...)`

- **UPDATE**
  - Existing DOM elements, bound to valid data.
  - `s`

- **EXIT**
  - DOM elements whose bound data has gone “stale”.
  - `s.exit()`
D3 Modules

Data Parsing / Formatting (JSON, CSV, ...)
Shape Helpers (arcs, curves, areas, symbols, ...)
Scale Transforms (linear, log, ordinal, ...)
Color Spaces (RGB, HSL, LAB, ...)
Animated Transitions (tweening, easing, ...)
Geographic Mapping (projections, clipping, ...)
Layout Algorithms (stack, pie, force, trees, ...)
Interactive Behaviors (brush, zoom, drag, ...)

Many of these correspond to future lecture topics!
Administrivia
A2: Exploratory Data Analysis

Use visualization software to form & answer questions

First steps:
Step 1: Pick domain & data
Step 2: Pose questions
Step 3: Profile the data
Iterate as needed

Create visualizations
Interact with data
Refine your questions

Author a report
Screenshots of most insightful views (10+)
Include titles and captions for each view

Due by 11:59pm Tuesday, Oct 16
A2: Exploratory Data Analysis

Use visualization software to form & answer questions.

First steps:
Step 1: Pick domain & data
Step 2: Pose questions
Step 3: Profile the data
Iterate as needed

Create visualizations
Interact with data
Refine your questions

Author a report
Screenshots of most insightful views (10+)
Include titles and captions for each view

Due by 11:59pm Tuesday, Oct 16
DONE
Tutorials

Introduction to D3.js

*In Class* - Wednesday, Oct. 24
A3: Interactive Prototype

Create an interactive visualization. Choose a driving question for a dataset and develop an appropriate visualization + interaction techniques, then deploy your visualization on the web.

Due by 11:59pm on Tuesday, October 30.

Work in project teams of 3-5 people.
Requirements

**Interactive.** You must implement interaction methods! However, this is not only selection / filtering / tooltips. Also consider annotations or other narrative features to draw attention and provide additional context.

**Web-based.** D3 is encouraged, but not required. Deploy your visualization using GitHub pages.

**Write-up.** Provide design rationale on your web page.
A3 & Final Project Team

Form a **team of 3-5** for A3 and the Final Project. Start thinking about your Final Project, too!

A3 is open-ended, but you can use it to start exploring your FP topic if you like.

Submit signup form by **Friday 10/19, 11:59pm**.

If you do not have **team mates**, you should:
- Use the facilities on Canvas
- Stay after class to meet potential partners
Team Member Roles

We encourage you to structure team responsibilities!

**Coordinator:** Organize meetings, track deadlines, etc.

**Data Lead:** Data wrangling, management, distillation

**Tech Lead:** Manage code integration, GitHub repo

**UX Lead:** Visualization/interaction design & evaluation

*One may have multiple roles, share work across roles…*
Interactive Prototype Tips

Start now. It will take longer than you think.

Keep it simple. Choose a minimal set of interactions that enables users to explore and generate interesting insights. Do not feel obligated to convey everything about the data: focus on a compelling subset.

Promote engagement. How do your chosen interactions reveal interesting observations?
A Visualization Tool Stack
Chart Typologies
Excel, Many Eyes, Google Charts

Visual Analysis Grammars
VizQL, ggplot2

Visualization Grammars
Protovis, D3.js

Component Architectures
Prefuse, Flare, Improvise, VTK

Graphics APIs
Processing, OpenGL, Java2D
Chart Typologies
Excel, Many Eyes, Google Charts

Visual Analysis Grammars
VizQL, ggplot2

Visualization Grammars
Protovis, D3.js

Component Architectures
Prefuse, Flare, Improvise, VTK

Graphics APIs
Processing, OpenGL, Java2D
Chart Typologies
Excel, Many Eyes, Google Charts

Visual Analysis Grammars
VizQL, ggplot2

Visualization Grammars
Protovis, D3.js

Component Architectures
Prefuse, Flare, Improvise, VTK

Graphics APIs
Processing, OpenGL, Java2D
What is a Declarative Language?

Programming by describing what, not how.

Separate **specification** (what you want) from **execution** (how it should be computed).

In contrast to imperative programming, where you must give explicit steps.

d3.selectAll("rect")
  .data(my_data)
  .enter().append("rect")
  .attr("x", function(d) { return xscale(d.foo); })
  .attr("y", function(d) { return yscale(d.bar); })
SELECT customer_id, customer_name, 
COUNT(order_id) as total
FROM customers
INNER JOIN orders ON customers.customer_id = orders.customer_id
GROUP BY customer_id, customer_name
HAVING COUNT(order_id) > 5
ORDER BY COUNT(order_id) DESC
Why Declarative Languages?

Faster iteration. Less code. Larger user base.

Better visualization. *Smart defaults.*

Reuse. *Write-once, then re-apply.*

Performance. *Optimization, scalability.*

Portability. *Multiple devices, renderers, inputs.*

Programmatic generation.

*Write programs which output visualizations.*

Automated search & recommendation.
Chart Typologies
Excel, Many Eyes, Google Charts

Visual Analysis Grammars
VizQL, ggplot2

Visualization Grammars
Protovis, D3.js

Component Architectures
Prefuse, Flare, Improvise, VTK

Graphics APIs
Processing, OpenGL, Java2D
Chart Typologies
Excel, Many Eyes, Google Charts

Visual Analysis Grammars
VizQL, ggplot2, **Vega-Lite**

Visualization Grammars
Protovis, D3.js, **Vega**

Component Architectures
Prefuse, Flare, Improvise, VTK

Graphics APIs
Processing, OpenGL, Java2D
Chart Typologies
Excel, Many Eyes, Google Charts

Visual Analysis Grammars
VizQL, ggplot2, Vega-Lite

Visualization Grammars
Protovis, D3.js, Vega

Component Architectures
Prefuse, Flare, Improvise, VTK

Graphics APIs
Processing, OpenGL, Java2D

Charting Tools
Declarative Languages
Programming Toolkits
Visual Analysis Grammars
VizQL, ggplot2, Vega-Lite

Visualization Grammars
Protovis, D3.js, Vega

Component Architectures
Prefuse, Flare, Improvise, VTK

Graphics APIs
Processing, OpenGL, Java2D
Interactive Data Exploration
Tableau, Lyra, Polestar, Voyager

Visual Analysis Grammars
VizQL, ggplot2, Vega-Lite

Visualization Grammars
Protovis, D3.js, Vega

Component Architectures
Prefuse, Flare, Improvise, VTK

Graphics APIs
Processing, OpenGL, Java2D
VEGA-LITE

A Grammar of Interactive Graphics

Kanit "Ham" Wongsuphasawat @kanitw
Dominik Moritz @domoritz
Arvind Satyanarayan @arvindsatya1
Jeffrey Heer @jeffrey_heer

Interactive Data Lab @uwdata
University of Washington
Grammar of Graphics for Customized Designs

Offer **fine-grained control** for composing interactive graphics.

But require **verbose** specifications and technical expertise.
Grammar of Graphics for Exploration

Facilitate rapid exploration with concise specifications by omitting low-level details.

Infer sensible defaults and allow customization by overriding defaults.

But limited support for interactions.
How might we author *interactive* graphics in the midst of analysis?
Vega-Lite: A Grammar of Interactive Graphics
A. Satyanarayan, D. Moritz, K. Wongsuphasawat & J. Heer. TVCG 2017
Vega-Lite: Scatter Plot

```json
{
    "data": {"url": "data/cars.json"},
    "mark": "point",
    "encoding": {
        "x": {"field": "Horse_Power", "type": "Q"},
        "y": {"field": "Miles_per_Gallon", "type": "Q"}
    }
}
```
Vega-Lite: Scatter Plot

```json
{
  "data": {"url": "data/cars.json"},
  "mark": "point",
  "encoding": {
    "x": {"field": "Horse_Power", "type": "Q"},
    "y": {"field": "Miles_per_Gallon", "type": "Q"},
    "color": {"field": "Cylinders", "type": "N"}
  }
}
```
Vega-Lite: Trellis Plot

```json
{
  "data": {"url": "data/cars.json"},
  "mark": "point",
  "encoding": {
    "x": {
      "field": "Horse_Power", "type": "Q"
    },
    "y": {
      "field": "Miles_per_Gallon", "type": "Q"
    },
    "column": {
      "field": "Cylinders", "type": "N"
    }
  }
}
```
Vega-Lite: Scatter Plot

```
{
  "data": {"url": "data/cars.json"},
  "mark": "point",
  "encoding": {
    "x": {"field": "Horse_Power", "type": "Q"},
    "y": {"field": "Miles_per_Gallon", "type": "Q"}
  }
}
```
Vega-Lite: 2D Histogram

```
{
    "data": {"url": "data/cars.json"},
    "mark": "point",
    "encoding": {
        "x": {"field": "Horse_Power", "type": "Q", "bin": true},
        "y": {"field": "Miles_per_Gallon", "type": "Q", "bin": true},
        "size": {"field": "*", "type": "Q", "aggregate": "count"}
    }
}
```
RESEARCH GOAL:
Extend grammars of statistical graphics to enable multi-view composition and interaction.
Vega-Lite: A Grammar of Graphics
Vega-Lite: A Grammar of Multi-View Graphics

Scatter Plot Matrix

Concatenated & Layered View

Faceted View

14
Vega-Lite: A Grammar of Interactive Multi-View Graphics
```json
{
    "data": {
        "url": "data/flights.json"
    },
    "mark": "bar",
    "encoding": {
        "x": {
            "field": "hour",
            "bin": true,
            "type": "Q"
        },
        "y": {
            "field": "*",
            "aggregate": "count",
            "type": "Q"
        }
    }
}
```
{
  "repeat": {"column": ["hour", "delay", "distance"]},
  "spec": {
    "data": {"url": "data/flights.json"},
    "mark": "bar",
    "encoding": {
      "x": {"field": {"repeat": "column"}, "bin": true, "type": "Q"},
      "y": {"field": "*", "aggregate": "count", "type": "Q"}
    }
  }
}
{  
  "repeat": {"column": ["hour", "delay", "distance"]},
  "spec": {
    "data": {"url": "data/flights.json"},
    "mark": "bar",
    "encoding": {
      "x": {"field": {"repeat": "column"}, "bin": true, "type": "Q"},
      "y": {"field": ":", "aggregate": "count", "type": "Q"}
    }
  }
}
"repeat": {"column": ["hour", "delay", "distance"]},
"spec": {
"layers": [
{
"data": {"url": "data/flights.json"},
"mark": "bar",
"encoding": {
"x": {"field": {"repeat": "column"}, "bin": true, "type": "Q"},
"y": {"field": "*", "aggregate": "count", "type": "Q"}
}
},
...

"encoding": {
...

"color": {"value": "goldenrod"}
}]]
}
{ "repeat": {"column": ["hour", "delay", "distance"]},
"spec": {
"layers": [{
"data": {"url": "data/flights.json"},
"mark": "bar",
"encoding": {
"x": {"field": {"repeat": "column"}, "bin": true, "type": "Q"},
"y": {"field": "*", "aggregate": "count", "type": "Q"}
}
}],
...,
"encoding": {
...
"color": {"value": "goldenrod"}
}
}
}
can be initialized
can be initialized
can be initialized
{
  "repeat": {"column": ["hour", "delay", "distance"]},
  "spec": {
    "layers": [{
      "select": {
        "region": {
          "type": "interval",
          "project": {
            "channels": ["x"],
            ...
          }
        }
      },
      ...,
      "transform": {
        "filterWith": "region"
      }
    }]
  }
}

35 Lines of JSON!
Interactive Selections
Altair is a declarative statistical visualization library for Python, based on Vega-Lite.

With Altair, you can spend more time understanding your data and its meaning. Altair’s API is simple, friendly and consistent and built on top of the powerful Vega-Lite visualization grammar. This elegant simplicity produces beautiful and effective visualizations with a minimal amount of code.
To Learn More…

Vega-Lite: A Grammar of Interactive Graphics, OpenVis Conf 2017
youtu.be/9uaHRWj04D4

Altair: Declarative Visualization for Python, PyData SF 2016
youtu.be/aRxahWy-ul8
How might we support more effective data exploration?
Common analysis pitfalls:
Overlook data quality issues
Fixate on specific relationships
Plus many other cognitive biases

[Heuer 1999, Kahneman 2011, …]
Voyager: Combine Manual Specification with Visualization Recommenders
**Key Idea:** Augment manual exploration with visualization recommendations sensitive to the user’s current focus.

The ultimate goal is to support systematic consideration of the data, without exacerbating false discovery.

To model a user’s search frontier, we enumerate related Vega-Lite specifications, seeded by the user’s current focus.

Candidate charts are pruned and ranked using models of estimated perceptual effectiveness.
Compared to existing tools, leads to **over 4x more variable sets seen**, and **over 2x more variable sets interacted with**.

“The related view suggestion accelerates exploration a lot.”

“I like that it shows me what fields to include in order to see a specific graph. Otherwise, I have to do a lot of trial and error and can’t express what I wanted to see.”

“These related views are so good but it’s also spoiling that I start thinking less. I’m not sure if that’s really a good thing.”
WORK IN PROGRESS...

Voyager integration with JupyterLab!