Data Visualization (CSCI 627/490)

Definition & Web Programming

Dr. David Koop
The purpose of visualization is about insight, not pictures

—Shneiderman
Exploration <-> Communication Spectrum

Consecutive Starts by a Quarterback for a Single Team

Questions

Answers/Persuasion

[K. Quealy, 2013]
Static Visualization

A Detailed Political Geography of the U.S.

And What It Means for 2016

What Donald Trump Sees in This Map

What Hillary Clinton Sees in This Map

[NYTimes]
Interactive Visualization

An Extremely Detailed Map of the 2020 Election
BY ALICE PARK, CHARLIE MARTZ, KINSEY TAYLOR AND MILES WATKINS

This map has detailed data from 2,523 of 3,143 counties in 47 states, representing 80% of all votes cast. It was last updated on March 30.

2020 results
Change from 2016

Dane County, Wis.
CANDIDATE
VOTES PCT. 2020 MARGIN
Joseph R. Biden 260,121 75% +53
Donald J. Trump 78,794 23%

© Mapbox © OpenStreetMap Improve this map

[NYTimes]
Administrivia

• Course Web Site
• Syllabus
  - Plagiarism
  - Accommodations
• Textbook:
  - Required: Munzner (VAD)
• Assignments
• Exams: Midterm (Oct. 23) and Final (Dec. 11)
Administrivia

- Undergraduate (CSCI 490) and Graduate (CSCI 627)
  - Graduate: Extra reading, exam questions, project emphasis
- Research Topics:
  - Also investigate some topics in depth
  - Research papers as assigned reading (CSCI 627)
- Project: Create an interactive visualization (or vis research)
  - Design
  - Data analysis
  - Insight
  - **Presentations**: Last week of class
Office Hours & Communication

- Office hours will be held **in person**
- Scheduled office hours are open to all students
  - M: 1:45-3pm, W: 10:45am-12pm, or by appointment
- You **do not** need an appointment to stop in during scheduled office hours
- If you need an appointment outside of those times, please email me with details about what you wish to discuss
- Many questions can be answered via email. **Please consider writing an email before scheduling a meeting.**
- **Do not send me screenshots of code!** (send code or Observable links)
Do not cheat!
Do not plagiarize

- It is **Academic Misconduct**
- Do your own work, do not copy anyone else's work, text, sentences, …
  - Anyone = another student, an internet source, book, blog, …
- Never quote text unless there is a specific need.
  - Usually, only famous quotes or very specific definitions
  - "I think there is a world market for maybe five computers."
    —Thomas Watson (1874-1956), Chairman of IBM, 1943
- **Cite** sources that back up your claims or reflect the origin of an idea
  - Vertex cover is an NP-Complete problem [1]. …
Do not cheat

• Cheating on assignments, projects, and exams is not allowed
• You will receive a zero for any assignment/exam/etc. where cheating has occurred
• You will fail the course if you cheat more than once
• Misconduct is reported through the university's system
• You may discuss problems and approaches with other students
• You may not copy or transcribe code from another source (includes generative AI)
Do ask questions!
Do ask questions

• If you are stuck on a specific issue with an assignment:
  - Do email me with **specific** questions
  - Do consult books, online documentation, tutorials
  - Do discuss that specific issue with a classmate

• If you are asked about a question:
  - Do not share your code
  - If the questioner is trying to cheat, walk away
  - If you see an obvious mistake, kindly point it out
  - Suggest a specific function or library that may be useful
Questions?
Assignment 1

• To be released soon…
• Write HTML, CSS, and SVG
• Text markup and styling
• Drawing markup and styling
• Data Visualization using Observable Plot
Definition of Visualization

“Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.”

— T. Munzner
Definition

“Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.”
Definition

“Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.”
Definition

"Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively."
Definition

“Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.”
Why People?

• Certain tasks can be totally **automated**
  - Statistical computations
  - Machine learning algorithms
  - We don’t need visualization for these tasks (although perhaps for debugging them…)

• Analysis problems are often **ill-specified**
  - What is the correct question?
  - Exploit human visual system, pattern detection capabilities
  - Goal may be an automated solution or a visual analysis system

• Presentation
Why Computers?

[Cerebral, Barsky et al., 2007]
Why Computers?

[Cerebral, Barsky et al., 2007]
Resource Limitations

- Memory and space constraints
- How many pixels do I have?
- Information Density

![Diagram of tree structures](image-url)
Definition

“Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively”
Why do we visualize data?

**Total Bandwidth**
(millions of bits per second)

- Sight
  - 1250 MB/s
  - Same bandwidth as a computer network
- Touch
  - 125 MB/s
  - USB Key
- Hearing
  - 12.5 MB/s
  - Hard Disk
- Smell
  - 12.5 MB/s
  - Taste

Figures are richer; provide more information with less clutter and in less space. Figures provide the gestalt effect: they give an overview; make structure more visible.

Figures are more accessible, easier to understand, faster to grasp, more comprehensible, more memorable, more fun, and less formal.

List adapted from: [Stasko et al. 1998]

[via A. Lex] [T. Norretranders]
## Why Visual?

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>II</td>
<td>III</td>
<td>IV</td>
<td></td>
</tr>
<tr>
<td>x</td>
<td>y</td>
<td>x</td>
<td>y</td>
<td>x</td>
</tr>
<tr>
<td>10.0</td>
<td>8.04</td>
<td>10.0</td>
<td>9.14</td>
<td>10.0</td>
</tr>
<tr>
<td>8.0</td>
<td>6.95</td>
<td>8.0</td>
<td>8.14</td>
<td>8.0</td>
</tr>
<tr>
<td>13.0</td>
<td>7.58</td>
<td>13.0</td>
<td>8.74</td>
<td>13.0</td>
</tr>
<tr>
<td>9.0</td>
<td>8.81</td>
<td>9.0</td>
<td>8.77</td>
<td>9.0</td>
</tr>
<tr>
<td>11.0</td>
<td>8.33</td>
<td>11.0</td>
<td>9.26</td>
<td>11.0</td>
</tr>
<tr>
<td>14.0</td>
<td>9.96</td>
<td>14.0</td>
<td>8.10</td>
<td>14.0</td>
</tr>
<tr>
<td>6.0</td>
<td>7.24</td>
<td>6.0</td>
<td>6.13</td>
<td>6.0</td>
</tr>
<tr>
<td>4.0</td>
<td>4.26</td>
<td>4.0</td>
<td>3.10</td>
<td>4.0</td>
</tr>
<tr>
<td>12.0</td>
<td>10.84</td>
<td>12.0</td>
<td>9.13</td>
<td>12.0</td>
</tr>
<tr>
<td>7.0</td>
<td>4.82</td>
<td>7.0</td>
<td>7.26</td>
<td>7.0</td>
</tr>
<tr>
<td>5.0</td>
<td>5.68</td>
<td>5.0</td>
<td>4.74</td>
<td>5.0</td>
</tr>
</tbody>
</table>

[F. J. Anscombe]
## Why Visual?

<table>
<thead>
<tr>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
<td>x</td>
<td>y</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>10.0</td>
<td>8.04</td>
<td>10.0</td>
<td>9.14</td>
</tr>
<tr>
<td>8.0</td>
<td>6.95</td>
<td>8.0</td>
<td>8.14</td>
</tr>
<tr>
<td>13.0</td>
<td>7.58</td>
<td>13.0</td>
<td>8.74</td>
</tr>
<tr>
<td>9.0</td>
<td>8.81</td>
<td>9.0</td>
<td>8.77</td>
</tr>
<tr>
<td>11.0</td>
<td>8.33</td>
<td>11.0</td>
<td>9.26</td>
</tr>
<tr>
<td>14.0</td>
<td>9.96</td>
<td>14.0</td>
<td>8.10</td>
</tr>
<tr>
<td>6.0</td>
<td>7.24</td>
<td>6.0</td>
<td>6.13</td>
</tr>
<tr>
<td>4.0</td>
<td>4.26</td>
<td>4.0</td>
<td>3.10</td>
</tr>
<tr>
<td>12.0</td>
<td>10.84</td>
<td>12.0</td>
<td>9.13</td>
</tr>
<tr>
<td>7.0</td>
<td>4.82</td>
<td>7.0</td>
<td>7.26</td>
</tr>
<tr>
<td>5.0</td>
<td>5.68</td>
<td>5.0</td>
<td>4.74</td>
</tr>
</tbody>
</table>

### Summary Statistics:

- **Mean of x**: 9
- **Variance of x**: 11
- **Mean of y**: 7.50
- **Variance of y**: 4.122
- **Correlation**: 0.816

[F. J. Anscombe]
Why Visual?

\[ y_1 = 3 \times x_1 \]

\[ y_2 = \frac{x_2}{2} \]

\[ y_3 = \frac{x_3}{2} + 3 \]

\[ y_4 = \frac{x_4}{2} + 3 \]

[F. J. Anscombe]
Why Visual?

<table>
<thead>
<tr>
<th>Mean of x</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance of x</td>
<td>11</td>
</tr>
<tr>
<td>Mean of y</td>
<td>7.50</td>
</tr>
<tr>
<td>Variance of y</td>
<td>4.122</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.816</td>
</tr>
</tbody>
</table>

[F. J. Anscombe]
Visual Pop-out
Visual Pop-out
Visual Pop-out
Visual Perception Limitations
Animation Can Help

![Graphs showing data points over time with animation concept]

A. Barrett via J. Cherdarchuk
Animation Can Help

[D. Koop, CSCI 627/490, Fall 2023]
But…
But...
Change & Inattentional Blindness Not Uncommon
Change & Inattentional Blindness Not Uncommon
Other Human Limitations

- Visual working memory is **small**
- **Change blindness**: A failure to notice a change in our view
- **Inattentional blindness**: A failure to notice something else going on in our view while focusing on a particular task
- "The goal of vision is not to build a complete photograph or model of the world in your mind. The goal of vision is to make sense of the meaning of the world around you." - D. Simons
Definition

“Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively”
Design Process: Iteration
Design Process: Iteration

<table>
<thead>
<tr>
<th>Team</th>
<th>Player 1</th>
<th>Player 2</th>
<th>Player 3</th>
<th>Player 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York Giants</td>
<td>Eli Manning</td>
<td>Andrew Luck</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indianapolis Colts</td>
<td>Peyton Manning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>San Diego Chargers</td>
<td>Drew Brees</td>
<td>Drew Brees</td>
<td>Philip Rivers</td>
<td></td>
</tr>
<tr>
<td>Baltimore Ravens</td>
<td>Kyle Boller</td>
<td>Steve McNair</td>
<td>Joe Flacco</td>
<td></td>
</tr>
<tr>
<td>New England Patriots</td>
<td>Tom Brady</td>
<td>Matt Cassel</td>
<td>Tom Brady</td>
<td></td>
</tr>
<tr>
<td>Green Bay Packers</td>
<td>Brett Favre</td>
<td>Aaron Rodgers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Orleans Saints</td>
<td>Aaron Brooks</td>
<td>Drew Brees</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atlanta Falcons</td>
<td>Michael Vick</td>
<td>Michael Vick</td>
<td>Matt Ryan</td>
<td></td>
</tr>
<tr>
<td>New York Jets</td>
<td></td>
<td>Chad Pennington</td>
<td>Brett Favre</td>
<td>Mark Sanchez</td>
</tr>
<tr>
<td>Cincinnati Bengals</td>
<td>Carson Paln</td>
<td>Carson Palmer</td>
<td>Ryan Fitzp</td>
<td>Andy Dalton</td>
</tr>
<tr>
<td>Houston Texans</td>
<td>David Carr</td>
<td>Matt Schaub</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carolina Panthers</td>
<td>Jake Delhomme</td>
<td>Jake Delhomme</td>
<td>Cam Newton</td>
<td></td>
</tr>
<tr>
<td>Denver Broncos</td>
<td>Jake Plummer</td>
<td>Jay Cutler</td>
<td>Kyle Orton</td>
<td>Tim Tebow</td>
</tr>
<tr>
<td>Arizona Cardinals</td>
<td></td>
<td></td>
<td>Matt Leina</td>
<td>Kurt Warner</td>
</tr>
<tr>
<td>Jacksonville Jaguars</td>
<td>Byron Leftwich</td>
<td>David Garrard</td>
<td>David Garrard</td>
<td>Blaine Gabbert</td>
</tr>
<tr>
<td>Detroit Lions</td>
<td>Joey Harrington</td>
<td>Jon Kinta</td>
<td></td>
<td>Matthew Stafford</td>
</tr>
<tr>
<td>Tampa Bay Buccaneers</td>
<td></td>
<td>Chris Simms</td>
<td>Bruce Gra</td>
<td>Josh Freeman</td>
</tr>
<tr>
<td>Dallas Cowboys</td>
<td>Drew Bledsoe</td>
<td>Tony Romo</td>
<td>Tony Romo</td>
<td></td>
</tr>
</tbody>
</table>

[K. Quealy, 2013]
Design Process: Iteration

Each streak shows consecutive starts by a quarterback for a single team. Streaks include playoffs.

Only two players have longer streaks: Brett Favre (275) and Eli’s brother, Peyton (227).

Among active players, Philip Rivers (122) and Joe Flacco (96) are closest behind Eli.

Find a quarterback
Eli Manning (149)

[K. Quealy, 2013]
Design Example

Each solid circle represents a bee species active in Carlinville, Ill., in both the late 1800s and 2010.

Hatching represents a bee species active in the 1800s but now locally extinct.

The spot where each block rests on the circle indicates one of 26 plant species frequented by these bees.

In the 1880s scientists observed the following about the bee-plant encounters:

- Present
- Frequent
- Abundant

Studies in 2009 and 2010 showed many bee-plant interactions had changed:

- Lost
- Persisted
- New

M. Stefaner, 2013 [M. Stefaner, 2013]
Impact of Design Choices: y-axis scale

Average Annual Global Temperature in Fahrenheit
1880-2015

[S. Hayward, 2015]
Impact of Design Choices: y-axis scale

Average Annual Global Temperature in Fahrenheit
1880-2015

[S. Hayward, 2015]
Impact of Design Choices: color
Impact of Design Choices: color

Red, yellow, blue

Purple, orange do not exist!
Analyzing Visualizations

Why?
How?
What?

Why?
How?
What?

Why?
How?
What?

Dependency

[Munzner (ill. Maguire), 2014]
Data Exploration through Visualization
Transportation Data - NYC MTA
## MTA Fare Data Exploration

<table>
<thead>
<tr>
<th>REMOTE</th>
<th>STATION</th>
<th>FF</th>
<th>SEN/DIS</th>
<th>7-D AFAS UNL</th>
<th>D AFAS/RMF</th>
<th>JOINT RR TKT</th>
<th>7-D UNL</th>
<th>30-D UNL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>R011 42ND STREET &amp; 8TH AVENUE</td>
<td>00228985</td>
<td>00008471</td>
<td>00000441</td>
<td>00001455</td>
<td>00000134</td>
<td>00033341</td>
<td>00071255</td>
</tr>
<tr>
<td>2</td>
<td>R170 14TH STREET-UNION SQUARE</td>
<td>00224603</td>
<td>0011051</td>
<td>00000827</td>
<td>00003026</td>
<td>00000660</td>
<td>00089367</td>
<td>00199841</td>
</tr>
<tr>
<td>3</td>
<td>R046 42ND STREET &amp; GRAND CENTRAL</td>
<td>00207758</td>
<td>00007908</td>
<td>00000323</td>
<td>00001183</td>
<td>00003001</td>
<td>00040759</td>
<td>00096613</td>
</tr>
<tr>
<td>4</td>
<td>R012 34TH STREET &amp; 8TH AVENUE</td>
<td>00188311</td>
<td>00006490</td>
<td>00000498</td>
<td>00001279</td>
<td>00003622</td>
<td>00035527</td>
<td>00067483</td>
</tr>
<tr>
<td>5</td>
<td>R293 34TH STREET - PENN STATION</td>
<td>00168768</td>
<td>00006155</td>
<td>00000523</td>
<td>00001065</td>
<td>00005031</td>
<td>00030645</td>
<td>00054376</td>
</tr>
<tr>
<td>6</td>
<td>R033 42ND STREET/TIMES SQUARE</td>
<td>00159382</td>
<td>00005945</td>
<td>00000378</td>
<td>00001205</td>
<td>00000690</td>
<td>00058931</td>
<td>00078644</td>
</tr>
<tr>
<td>7</td>
<td>R022 34TH STREET &amp; 6TH AVENUE</td>
<td>00156008</td>
<td>00006276</td>
<td>00000487</td>
<td>00001543</td>
<td>00000712</td>
<td>00058910</td>
<td>00110466</td>
</tr>
<tr>
<td>8</td>
<td>R084 59TH STREET/COLUMBUS CIRCLE</td>
<td>00155262</td>
<td>00009484</td>
<td>00000589</td>
<td>00002071</td>
<td>00000542</td>
<td>00053397</td>
<td>00113966</td>
</tr>
<tr>
<td>9</td>
<td>R020 47-50 STREETS/ROCKEFELLER</td>
<td>00143500</td>
<td>00006402</td>
<td>00000384</td>
<td>00001159</td>
<td>00000723</td>
<td>00037978</td>
<td>00090745</td>
</tr>
<tr>
<td>10</td>
<td>R179 86TH STREET-LEXINGTON AVE</td>
<td>00142169</td>
<td>00010367</td>
<td>00000470</td>
<td>00001839</td>
<td>00000271</td>
<td>00050328</td>
<td>00125250</td>
</tr>
<tr>
<td>11</td>
<td>R023 34TH STREET &amp; 6TH AVENUE</td>
<td>00134052</td>
<td>00005005</td>
<td>00000348</td>
<td>00001112</td>
<td>00000649</td>
<td>00031531</td>
<td>00075040</td>
</tr>
<tr>
<td>12</td>
<td>R029 PARK PLACE</td>
<td>00121614</td>
<td>00004311</td>
<td>00000287</td>
<td>00000931</td>
<td>00000792</td>
<td>00025404</td>
<td>00065362</td>
</tr>
<tr>
<td>13</td>
<td>R047 42ND STREET &amp; GRAND CENTRAL</td>
<td>00100742</td>
<td>00004273</td>
<td>00000185</td>
<td>00000704</td>
<td>00001241</td>
<td>00022808</td>
<td>00068216</td>
</tr>
</tbody>
</table>
MTA Fare Data Exploration
MTA Fare Data Exploration
MTA Fare Data Exploration
MTA Fare Data Exploration

![Bar Chart](chart.png)

- **East 161st Street and River Avenue**

- **Date**
  - 08-02
  - 08-09
  - 08-16
  - 08-23
  - 08-30
  - 09-06
  - 09-13
  - 09-20
  - 09-27
  - 10-04
  - 10-11
  - 10-18
  - 10-25
  - 11-01

- **Full Fares Purchased**
  - 0
  - 10000
  - 20000
  - 30000
  - 40000
  - 50000
  - 60000
  - 70000
  - 80000
  - 90000
  - 100000
  - 110000
  - 120000
MTA Fare Data Exploration

East 161st Street and River Avenue

Full Fares Purchased

Date

08-02 08-09 08-16 08-23 08-30 09-06 09-13 09-20 09-27 10-04 10-11 10-18 10-25 11-01

New York Yankees

AUGUST

Full Fares Purchased

Date

2013 REGULAR SEASON SCHEDULE
Tools

- Desktop Applications:
  - Excel (see excelcharts.com)
  - Tableau
  - ...
- Grammars:
  - Vega-Lite
- Programming Frameworks
  - d3.js
  - Observable Plot, plot.ly, deck.gl
  - ...

- Tradeoffs
  - Speed
  - Customization
  - Understanding
  - Dissemination
Programming

• "Programming is blindly manipulating symbols." - B. Victor
• "Code is often the best tool we have because it is the most general tool we have; code has almost unlimited expressiveness" - M. Bostock

• You will write code in this class
  - Your assignments will involve code
  - Your project will involve code

• JavaScript is the language of the Web
  - Somewhat forgiving, not always the easiest to debug
  - Lots of references out there
  - A quickly-changing environment of frameworks
What languages do we use on the Web?
Languages of the Web

- HTML
- CSS
- SVG
- JavaScript
  - Versions of Javascript: ES6, ES2015, ES2020…
  - Specific frameworks: react, jQuery, bootstrap, D3
Web Programming Tools

• Basic: Text editor and Modern Browser
• Developer Tools: Built in to browsers (e.g. Chrome Developer Tools)
• Web Environments: Observable, CodePen, JSFiddle, etc.
• IDEs: WebStorm, VSCode
Observable

- Observable is a platform that allows users to create notebooks using JavaScript, Markdown, and HTML
- Strong support of data visualization (company and community interests)
- Introduction: A Taste of Observable
- Type markup/code, "execute" the cell, and results appear above the code
- Pin the cell to keep the code visible
- Can choose the type of cell (JavaScript, Markdown, or HTML)
- Can create an output (variable) in each cell that can be used in other cells
- Content is all global scope!
Other Platforms

• CodePen
• deno.land