Data Visualization (CSCI 627/490)

Introduction

Dr. David Koop
What is Data Visualization?
How is it different from Computer Graphics?
The purpose of computing is about insight, not numbers

— R. W. Hamming
The purpose of visualization is about insight, not pictures

— Shneiderman
Why do we visualize data? (vs. looking at tables?)
# MTA Fare Data Table

<table>
<thead>
<tr>
<th>REMOTE</th>
<th>STATION</th>
<th>FF</th>
<th>Y</th>
<th>SEN/DIS</th>
<th>7-D ATAS UNL</th>
<th>D ATAS/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>
<td></td>
</tr>
<tr>
<td>2</td>
<td>R170 14TH STREET-UNION SQUARE</td>
<td>00224603</td>
<td>00011051</td>
<td>00000827</td>
<td>00003026</td>
<td>00000660</td>
<td>00089367</td>
<td>00199841</td>
<td></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>
<td></td>
</tr>
<tr>
<td>4</td>
<td>R012 34TH STREET &amp; 6TH AVENUE</td>
<td>00188311</td>
<td>00006490</td>
<td>00000498</td>
<td>00001279</td>
<td>00003622</td>
<td>00036527</td>
<td>00067483</td>
<td></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>
<td></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>00006909</td>
<td>00058931</td>
<td>00078644</td>
<td></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>
<td></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>
<td></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>00007237</td>
<td>00037978</td>
<td>00090745</td>
<td></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>
<td></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>
<td></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>
<td></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>
<td></td>
</tr>
<tr>
<td>14</td>
<td>R031 34TH STREET &amp; 7TH AVENUE</td>
<td>00095076</td>
<td>00003990</td>
<td>00000232</td>
<td>00000727</td>
<td>00001459</td>
<td>00024284</td>
<td>00038671</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>R017 LEXINGTON AVENUE</td>
<td>00094655</td>
<td>00004688</td>
<td>00000190</td>
<td>00000833</td>
<td>00000754</td>
<td>00020183</td>
<td>00055066</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>R175 8TH AVENUE-14TH STREET</td>
<td>00094313</td>
<td>00003907</td>
<td>00000286</td>
<td>00001144</td>
<td>00000256</td>
<td>00038272</td>
<td>00074661</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>R057 BARCLAYS CENTER</td>
<td>00093804</td>
<td>00004204</td>
<td>00000454</td>
<td>00001386</td>
<td>00001491</td>
<td>00030113</td>
<td>00068119</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>R138 WEST 4TH ST-WASHINGTON SQ</td>
<td>00093562</td>
<td>00004677</td>
<td>00000251</td>
<td>00000965</td>
<td>00000127</td>
<td>00031628</td>
<td>00074458</td>
<td></td>
</tr>
</tbody>
</table>
MTA Fare Data Visualization
Why do we visualize data?

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]

D. Koop, CSCI 627/490, Fall 2023
What are the purposes for visualization?
Analysis: Subway Ridership Density
Communication: Peyton's Records

Why Peyton Manning's Record Will Be Hard to Beat

By GREGOR AISCH and KEVIN QUEALY	OCT. 19, 2014

The Broncos quarterback set the all-time N.F.L. touchdown passing record — and is still going strong.

[D. Koop, CSCI 627/490, Fall 2023]

[G. Aisch and K. Quealy, NYTimes]
Exploration <-> Communication Spectrum

Consecutive Starts by a Quarterback for a Single Team

Each streak —— shows consecutive starts by a quarterback for a single team, through most playoffs.

Only two players have longer streaks: Bart Starr (276) and Eli’s brother, Peyton (227).

Among active players, Philip Rivers (152) and Joe Flacco (106) are closest behind Eli.

Questions

Answers/Persuasion

[K. Quealy, 2013]
What types of data can we visualize?
Types of Data

- Tables
- Networks (Graphs)
- Spatial Data
  - Geography
  - Physical (e.g. Scientific, Medical)
- Text
- Sets
Where have you seen visualizations?
Books/Posters

[Rock 'N' Roll is Here to Pay, R. Garofalo, 1977 (via Tufte)]
Newspapers
What is the advantage of the second version?
Interaction
How do we create modern visualizations?
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
Tableau
Observable Plot

```javascript
seattle = FileAttachment("seattle-weather.csv").csv({typed: true})

Plot.plot({
  height: 300,
  padding: 0,
  y: {
      tickFormat: Plot.formatMonth("en", "short")
  },
  marks: [
    Plot.cell(seattle, Plot.group({fill: "max"}, {
      x: d => d.date.getUTCDate(),
      y: d => d.date.getUTCMonth(),
      fill: "temp_max",
      inset: 0.5
    }))
  ]
})
```

---

D. Koop, CSCI 627/490, Fall 2023
Best Global Brands
Value in $M; color indicates sector. Data: Interbrand

Coca-Cola 72,537
Microsoft 70,196
IBM 53,183
Intel 39,046
Nokia 38,528
GE 38,127
Ford 36,988
Disney 33,563
McDonald's 27,869
AT&T 25,548
Marlboro 22,118
Mercedes-Benz 21,104

2000

[M. Bostock]
Best Global Brands
Value in $M; color indicates sector. Data: *Interbrand*

- Coca-Cola: 72,537
- Microsoft: 70,198
- IBM: 53,183
- Intel: 39,046
- Nokia: 38,528
- GE: 38,127
- Ford: 36,988
- Disney: 33,563
- McDonald's: 27,859
- AT&T: 25,548
- Marlboro: 22,119
- Mercedes-Benz: 21,104

2000
Why do we care about the design of visualizations?
Design: Focus on only the y-axis
Design: Year on the y-axis

Year (A.D.)

Year
Design: Different y-axis

Average Annual Global Temperature in Fahrenheit
1880-2015

[S. Hayward, 2015]
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
Exams, Quizzes, and Missed Classes

• All quizzes and exams are **in-person, closed-book, and closed-notes**
• **No makeup** quizzes or exams will be given*
• You are responsible for all material presented in class and assigned readings. If you miss a lecture, be sure to obtain class notes from a classmate before the next class meeting
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 consult books, online documentation, tutorials
  - Do discuss that specific issue with a classmate
  - Do email me with specific questions

• 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?