

Data Visualization: Strategies, Tips, and Tools

February 10th, 2016

Matt Jansen

Digital Research Services

UNC- Chapel Hill

University Library

<http://library.unc.edu/hub/>



UNC
UNIVERSITY LIBRARIES

Agenda

- Why visualize? Why not?
- Visualization Principles
 - Human Perception
 - Preattentive Processing
 - Reducing Distractions
 - Color
 - Tips and Tricks
- Tool Landscape
 - Spreadsheets
 - In-browser tools
 - Business Intelligence
 - Coding
 - Design
- Q&A

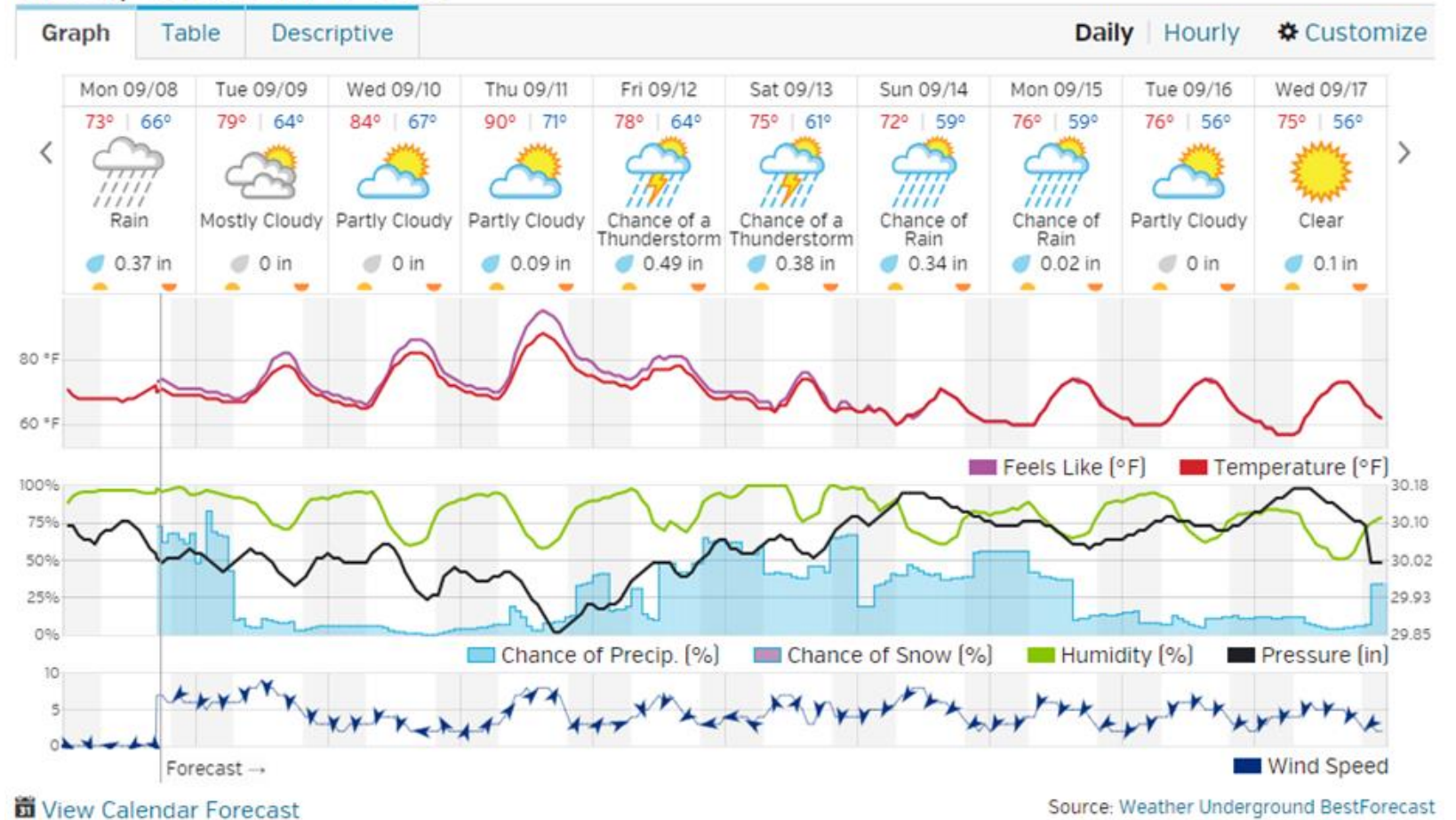
About Me

- University of North Carolina at Chapel Hill
 - Data Analyst at University Libraries
- Background
 - Social Sciences and Statistics
- Supporting visualization and data analysis
 - Across disciplines and experience levels
 - Workshops
 - Consultations



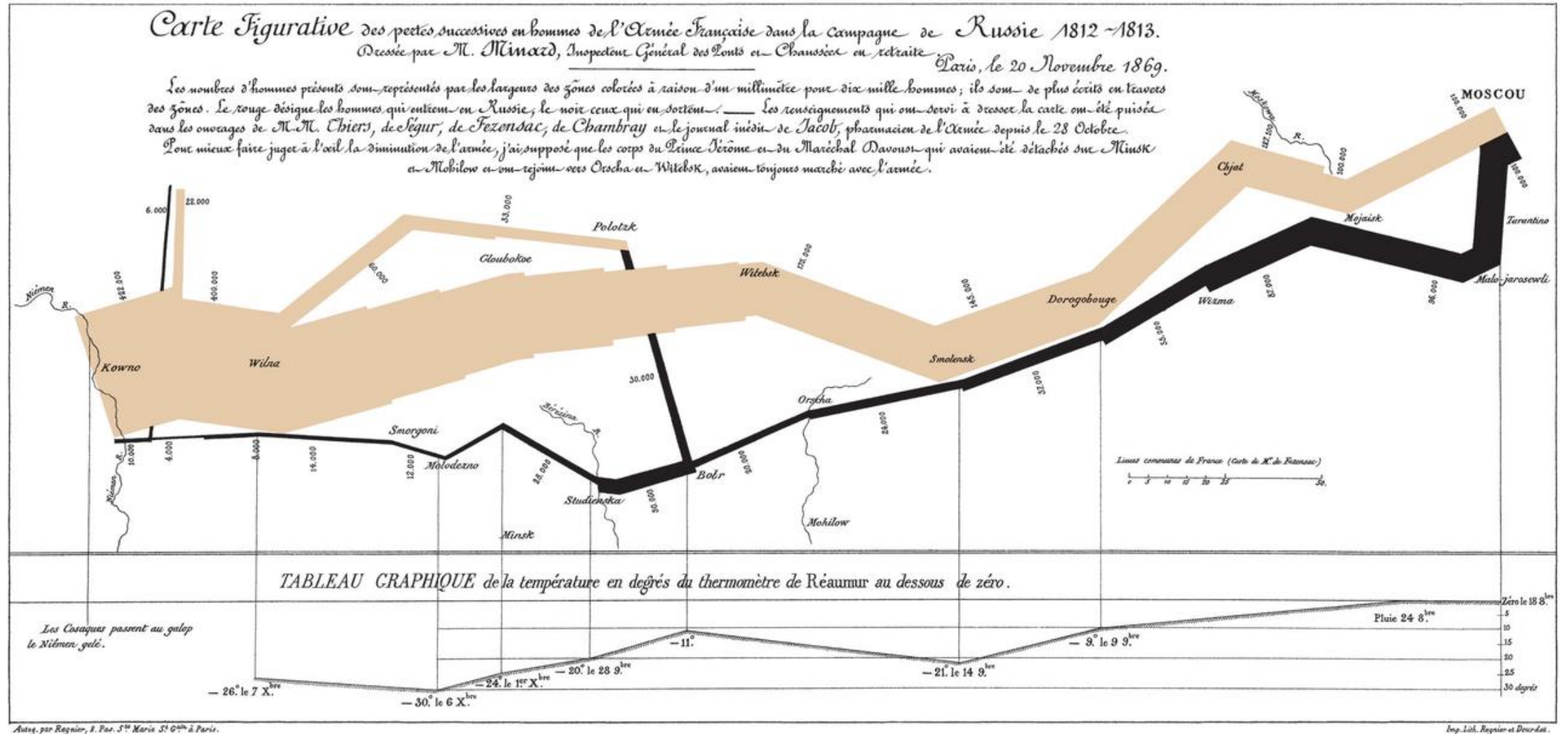
Why Visualize: Analysis and Dashboards

10-Day Weather Forecast



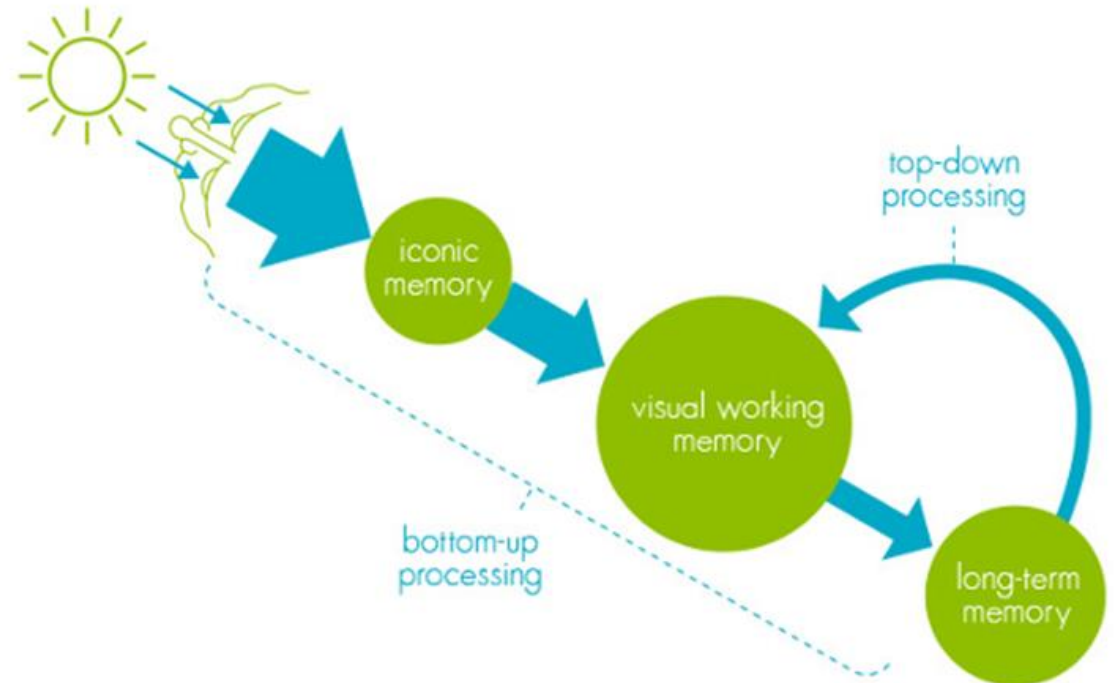
<http://www.wunderground.com/>

Why Visualize: Storytelling



Why Visualize: Human Cognitive Limitations

- Visual working memory is **small**
 - Numbers
 - Patterns



Why **not** a visualization?

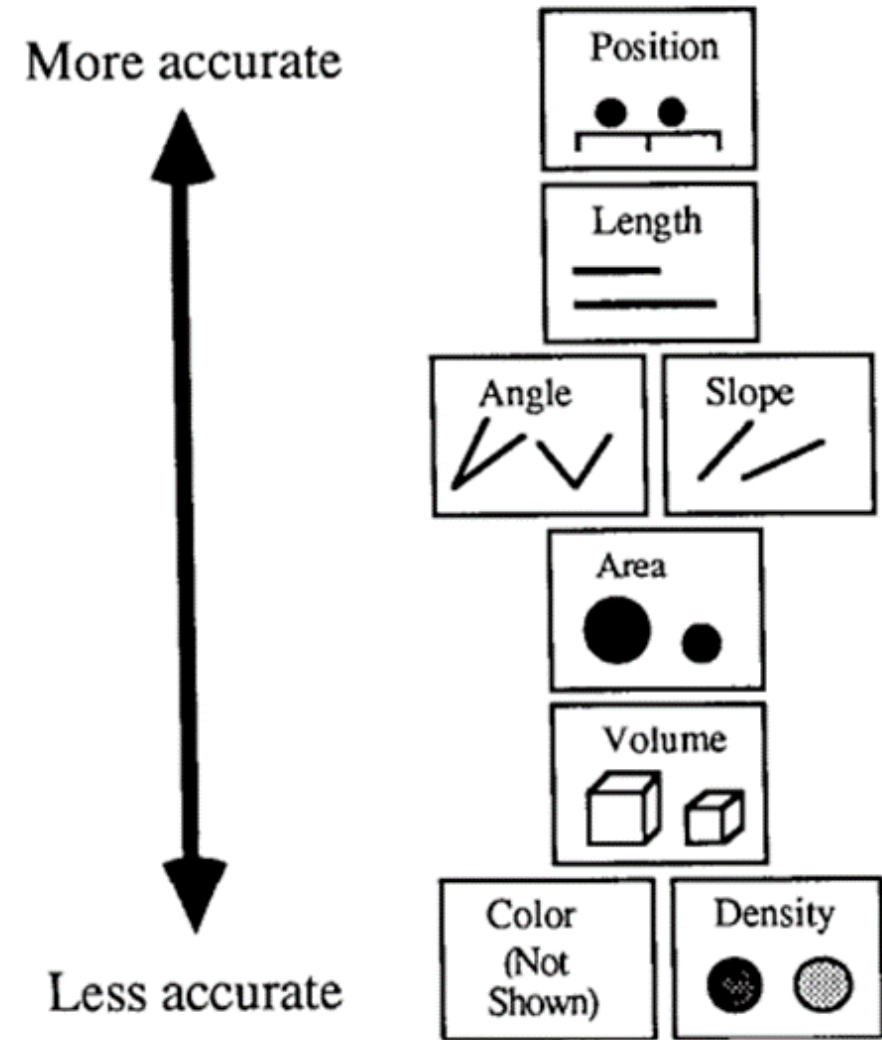
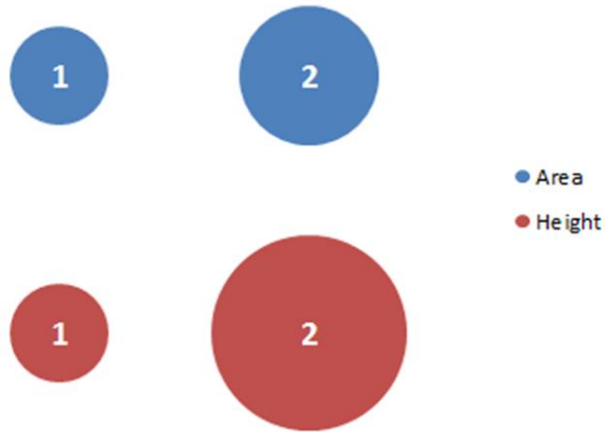
- Does it beat a table?
 - e.g. Only a few values visualized
- Does it beat text?
- Does it oversimplify the issue?
 - Textual context
 - Statistics
- Models never tell the whole story

Basic Principles for Visualization

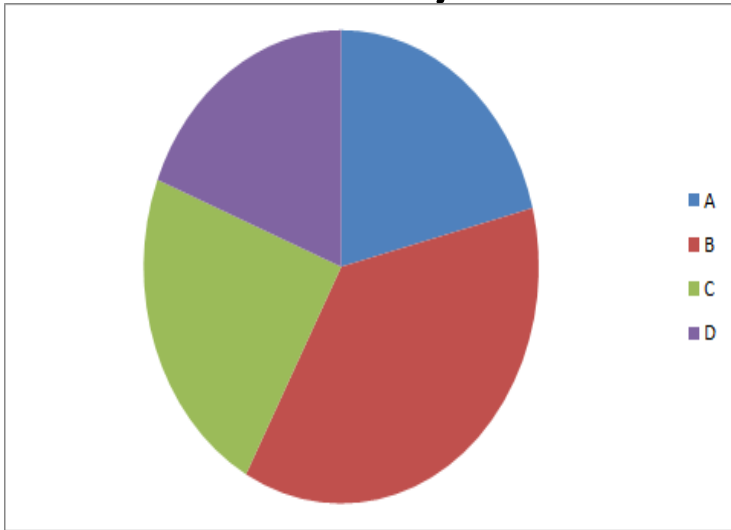
- Accuracy and Human Perception
- Preattentive Processing, or 'Pop!'
- Minimizing Distractions
- Colors

Encoding data for accuracy

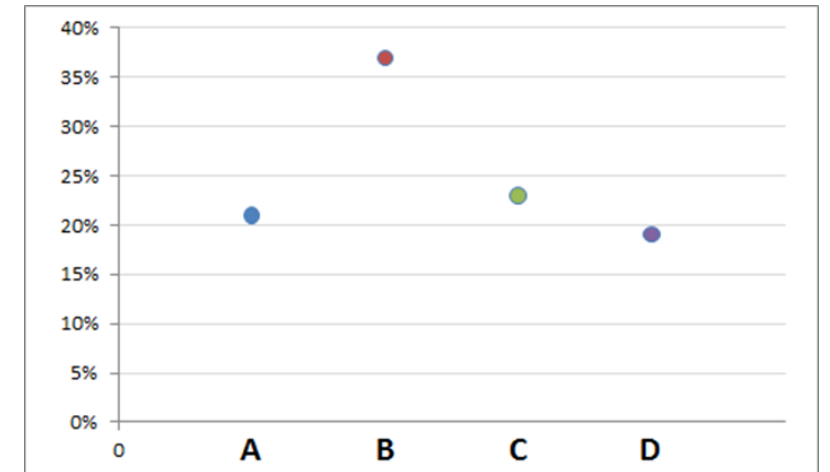
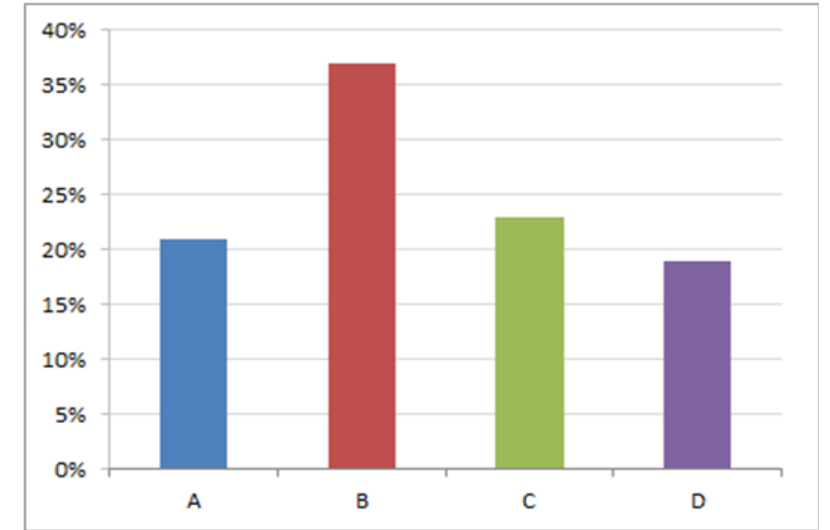
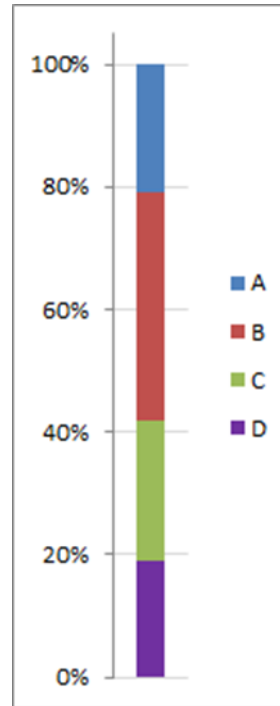
- Maximizing accuracy may **not** be your only goal
- More dimensions = Less accuracy
 - Length vs. Area
 - Area vs. Volume (avoid 3D effects)



Accuracy



Category	Percentage
A	21.0%
B	37.0%
C	23.0%
D	19.0%



Bonus: 45 ways to communicate two quantities:

<http://blog.visual.ly/45-ways-to-communicate-two-quantities/>

Preattentive Processing

- What 'pops' out in a visualization
- Survival instincts

**981879263837498197949613897461394497
873218498762161799546213254989796531
859129939549719819295198197354687929**

How many sevens appear above?

Preattentive processing

981879263837498197949613897461394497
873218498762161799546213254989796531
859129939549719819295198197354687929

How many sevens appear above?

9818**7**926383**7**49819**7**94961389**7**46139449**7**
8**7**3218498**7**62161**7**99546213254989**7**96531
859129939549**7**1981929519819**7**35468**7**929

Preattentive processing

981879263837498197949613897461394497
873218498762161799546213254989796531
859129939549719819295198197354687929

How many sevens appear above?

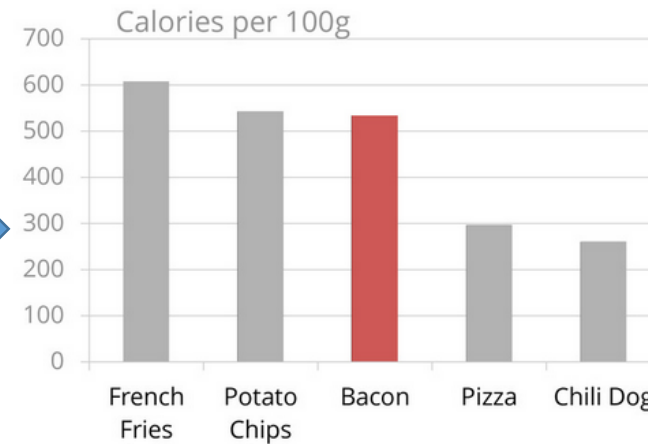
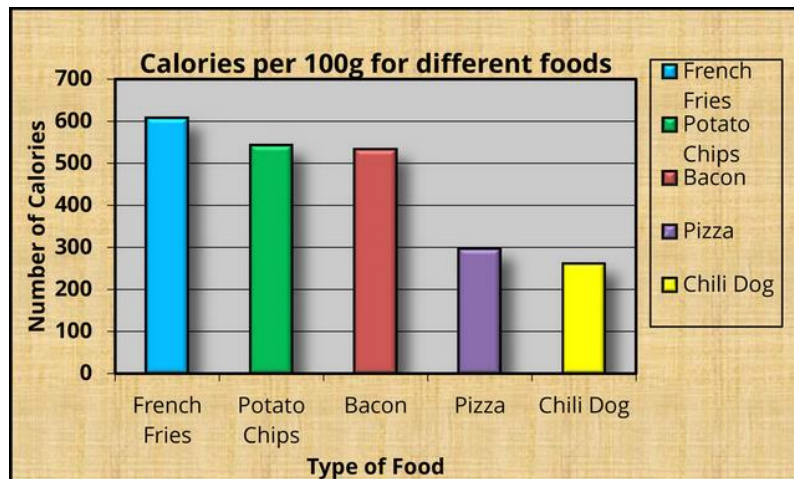
9818**7**926383**7**49819**7**94961389**7**46139449**7**
8**7**3218498**7**62161**7**99546213254989**7**96531
859129939549**7**1981929519819**7**35468**7**929

9818**7**926383**7**49819**7**94961389**7**46139449**7**
8**7**3218498**7**62161**7**99546213254989**7**96531
859129939549**7**1981929519819**7**35468**7**929

Distractions: Data-ink Ratio

- Edward Tufte
- Within reason, maximize:

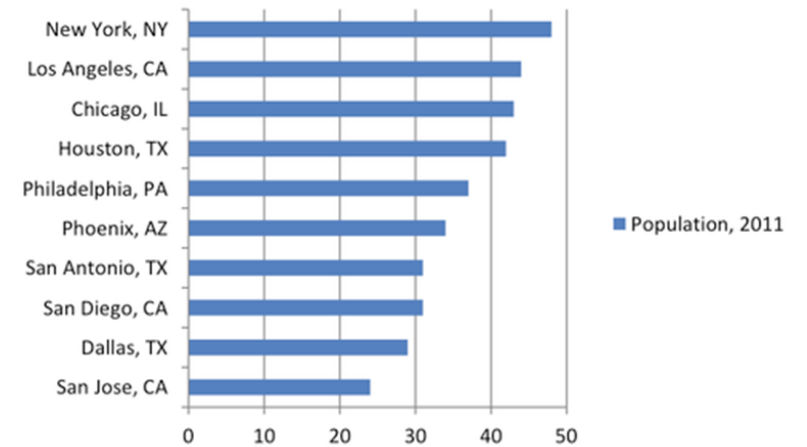
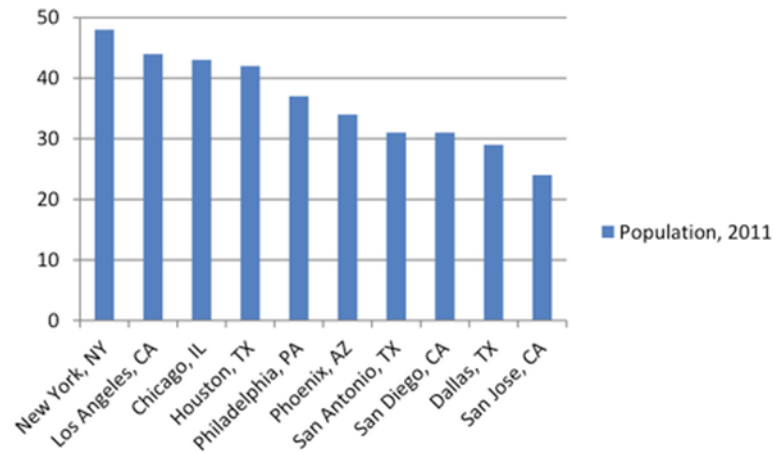
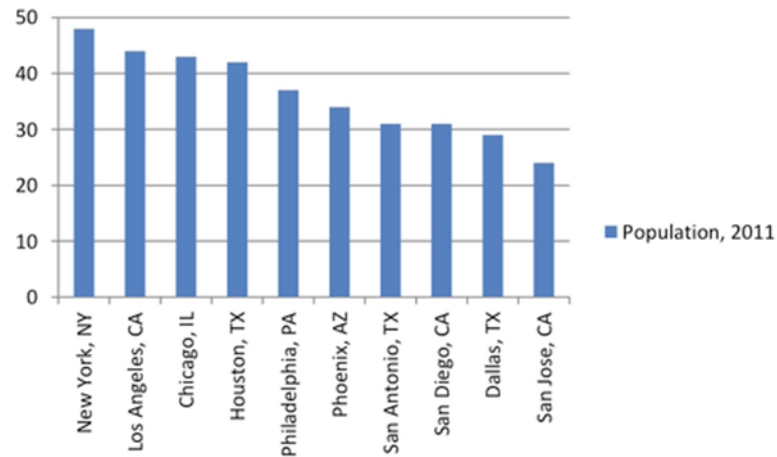
$$\frac{\text{Ink used to encode data values}}{\text{Total ink in visualization}}$$



See the entire progression from the colorful, decorative chart on the left to the simpler chart above, and beyond at:
<https://speakerdeck.com/cherdarchuk/remove-to-improve-the-data-ink-ratio>

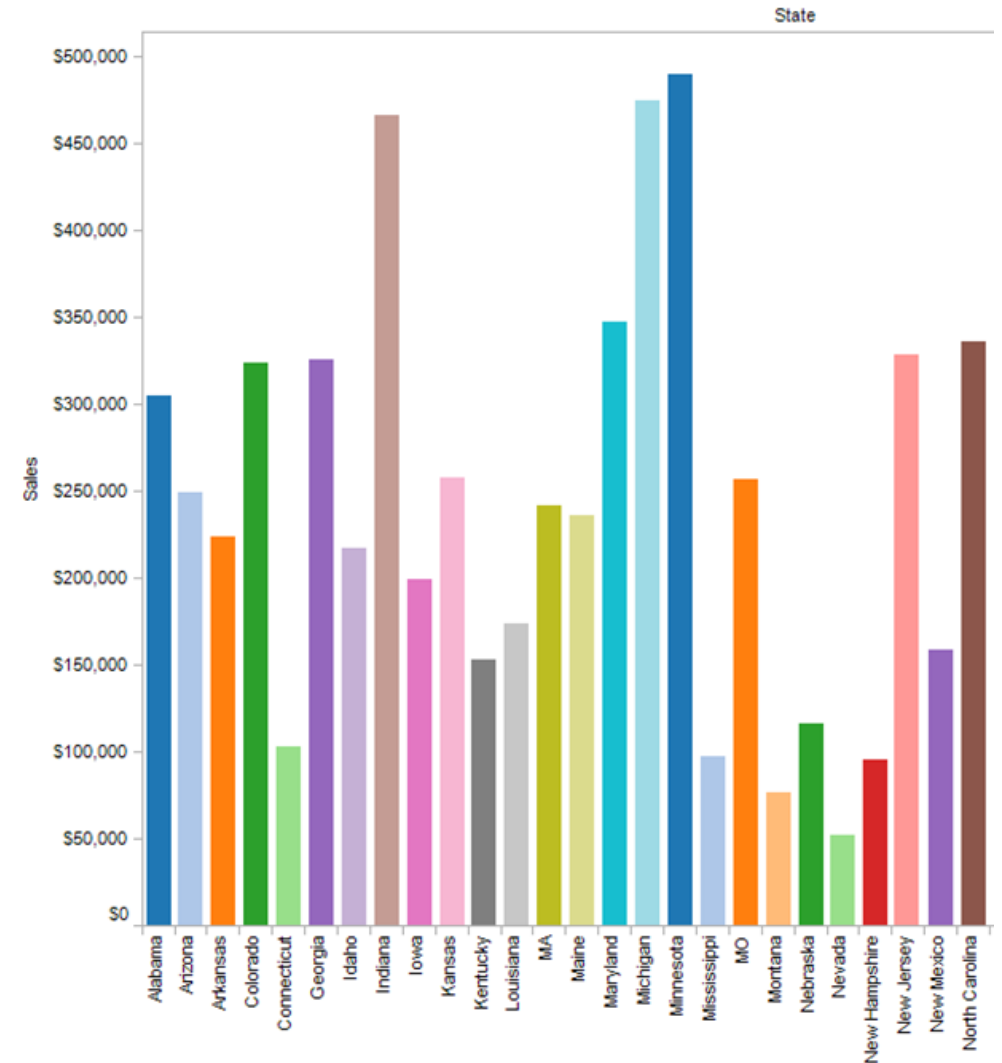
Distractions: Text Orientation

- Text is often one of the most important elements of a visualization
- Text should be as easily readable as possible



Color

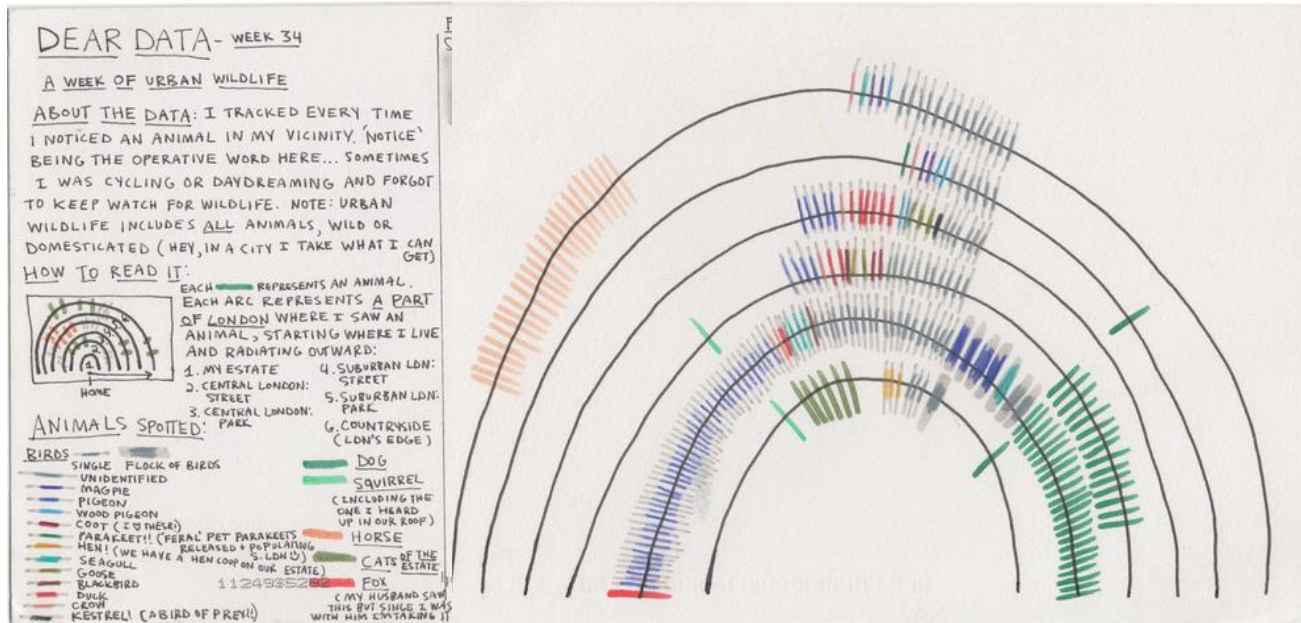
- Don't use Red-Green palettes
 - Indistinguishable to most common forms of colorblindness
- “Get it right in black and white”
- Color is often unnecessary or distracting
- Choosing a palette:
 - <http://colorbrewer2.org/>



Tips and Tricks

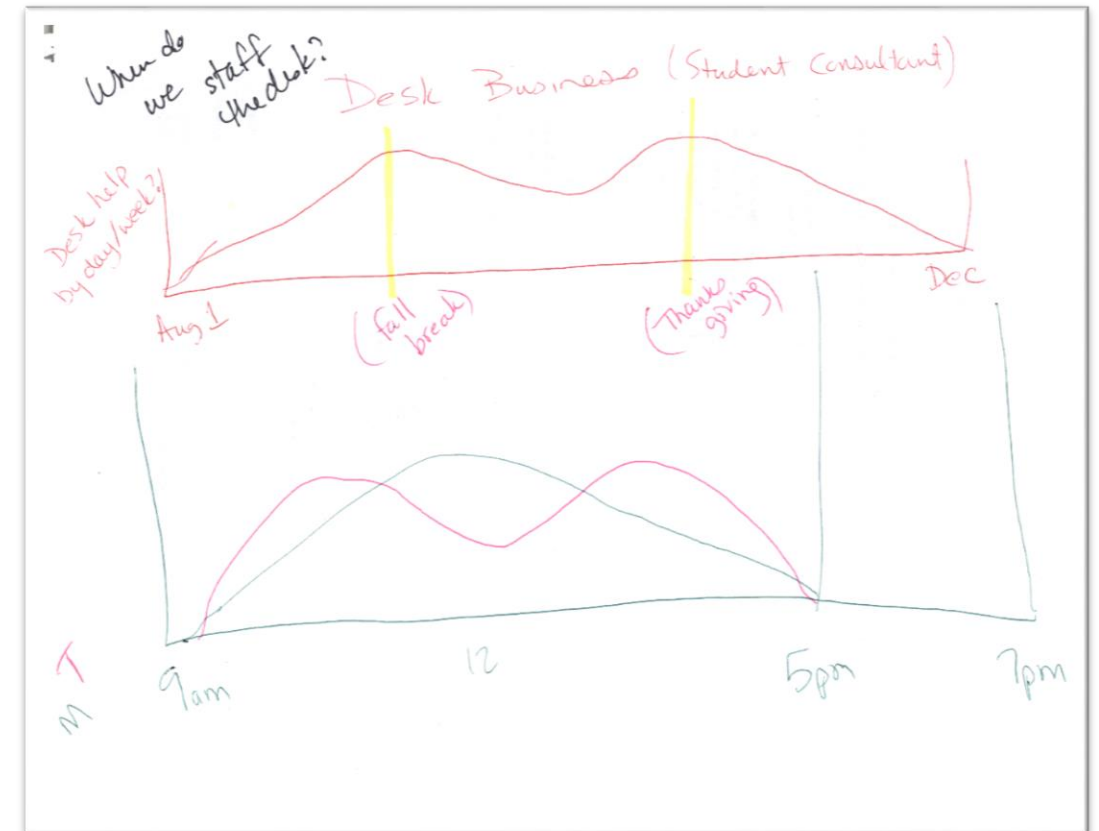
- Draw pictures!
- Squint Test
- Small Multiples

Draw Pictures!



Dear Data project

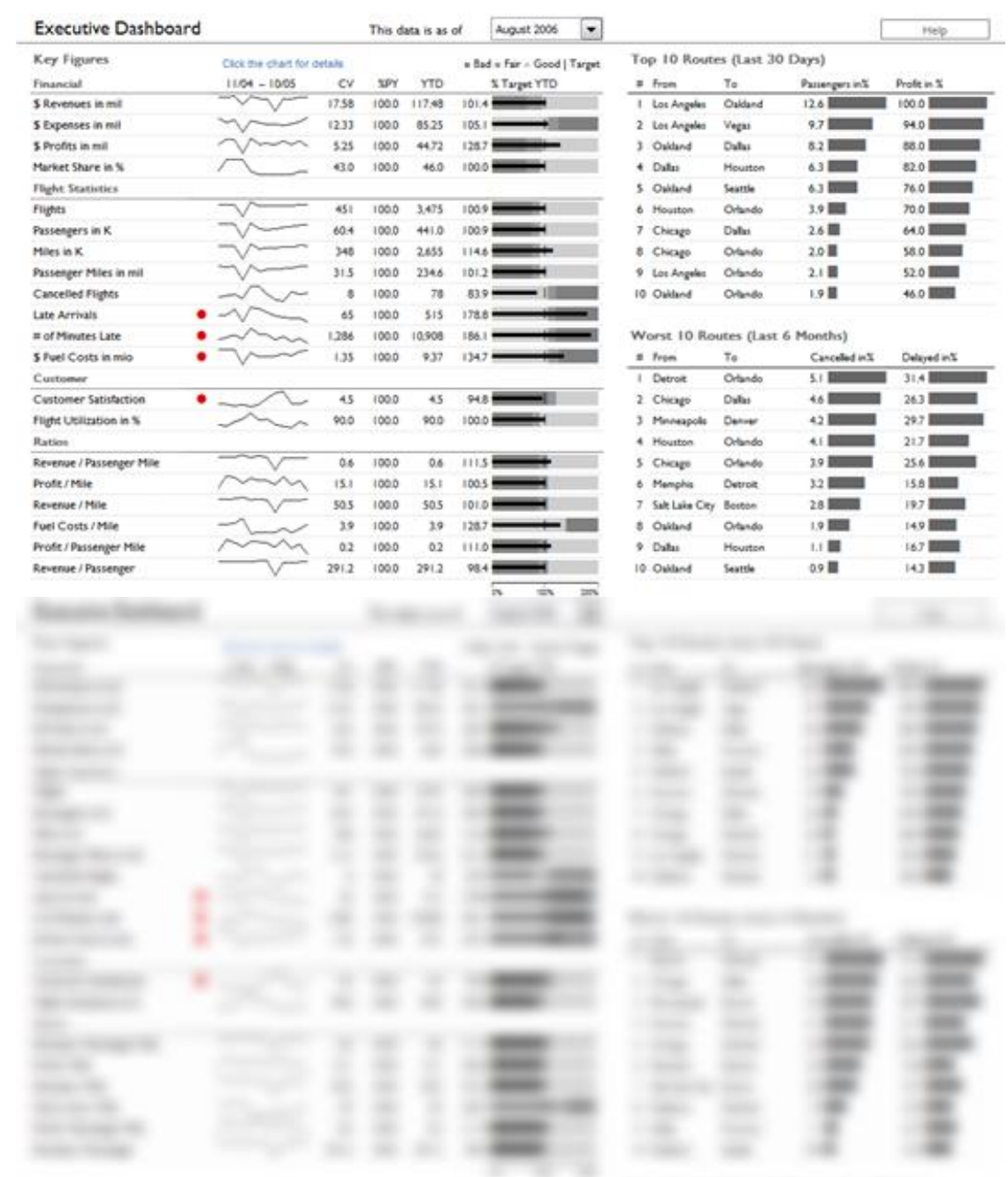
<http://www.dear-data.com/week-34-a-week-of-urban-wildlife>



Designing a visualization of library desk traffic at UNC

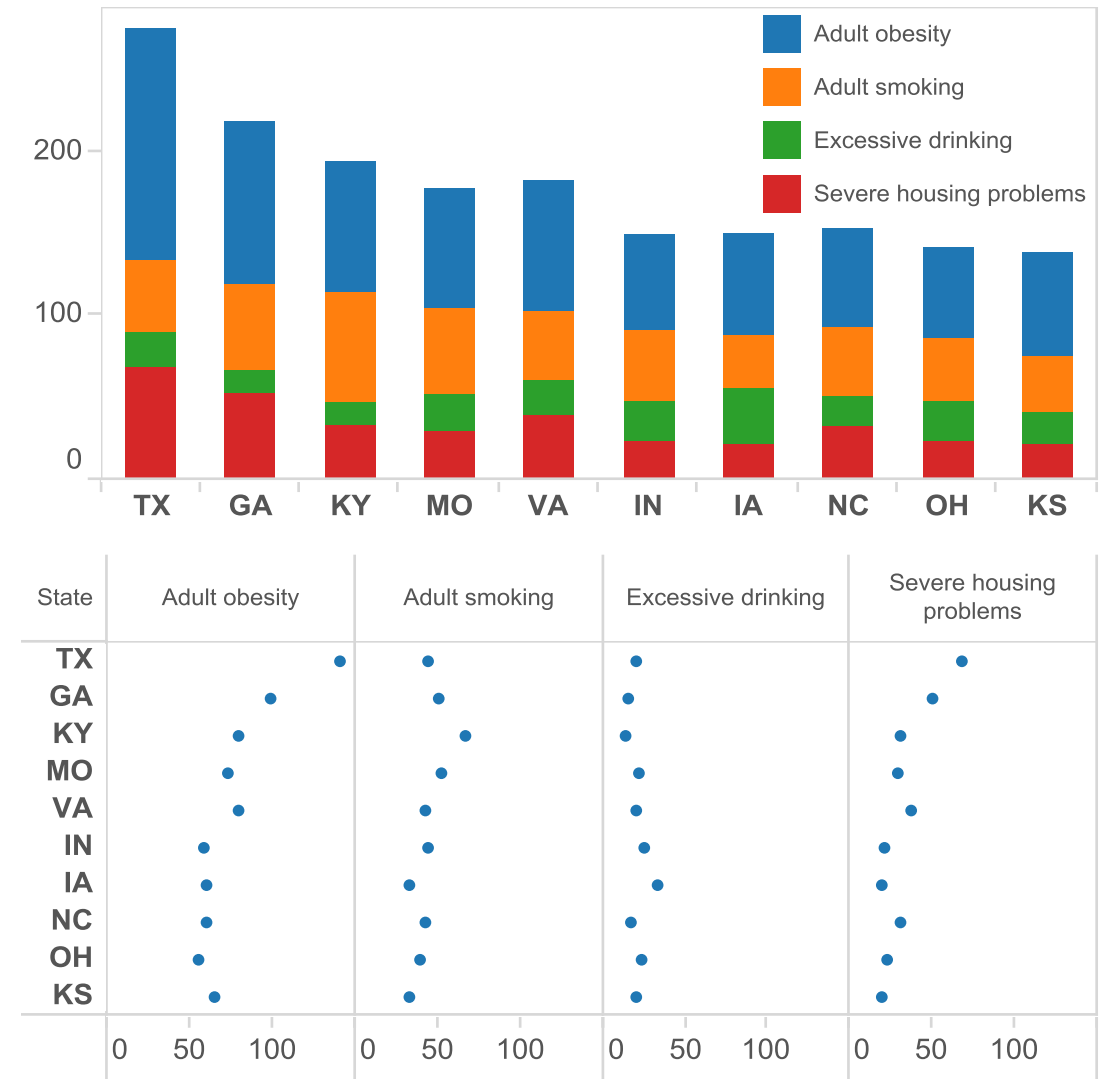
Squint Test

- Preattentive Processing
- Easy way to test what elements of a visualization ‘pop’ and what draws the eye of a new viewer



Small Multiples

- What you **can** fit in one visualization vs. what you **should**
- Usually, the simpler the better
- Alternatively, remove unnecessary variables



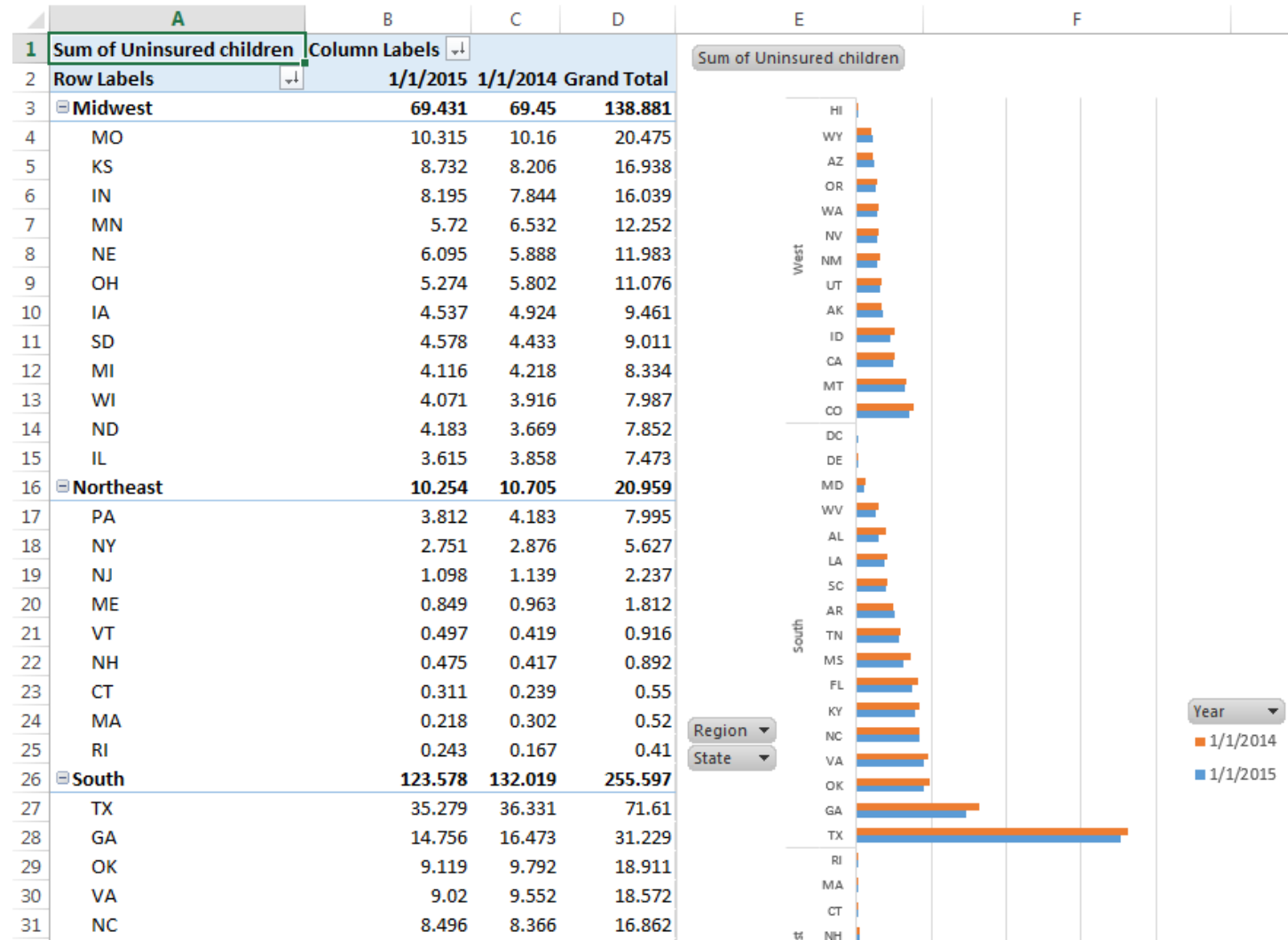
Tool Landscape

- Spreadsheets
- In-browser tools
- Business Intelligence Tools
- Coding
- Design

Spreadsheets

e.g. **Microsoft Excel**, LibreOffice, Open Office

- Pros:
 - You probably already have it
 - Your data probably passes through it already
 - Secure
 - Already integrated in workflows
- Cons:
 - Software not primarily designed for visualization
 - Static and local



In-Browser

e.g. **Plot.ly**, Datawrapper, Raw, Timeline.js

- Pros:
 - Often easiest, most accessible, quickest
 - Often free or cheap
 - Many tools available
- Cons:
 - Most subject to change (or disappearance)
 - Inflexibility
 - Specialized functionality
 - Strict data format needs
 - Dependence on other software
 - Too many options



Plot.ly

In-Browser: Mapping

e.g. **ArcGIS Online**, CartoDB

- Pros:
 - Most flexible mapping tools
 - Includes other map-oriented functionality
- Cons:
 - Full benefits require a more advanced tool
 - ArcMap
 - QGIS
 - No other visualization types available



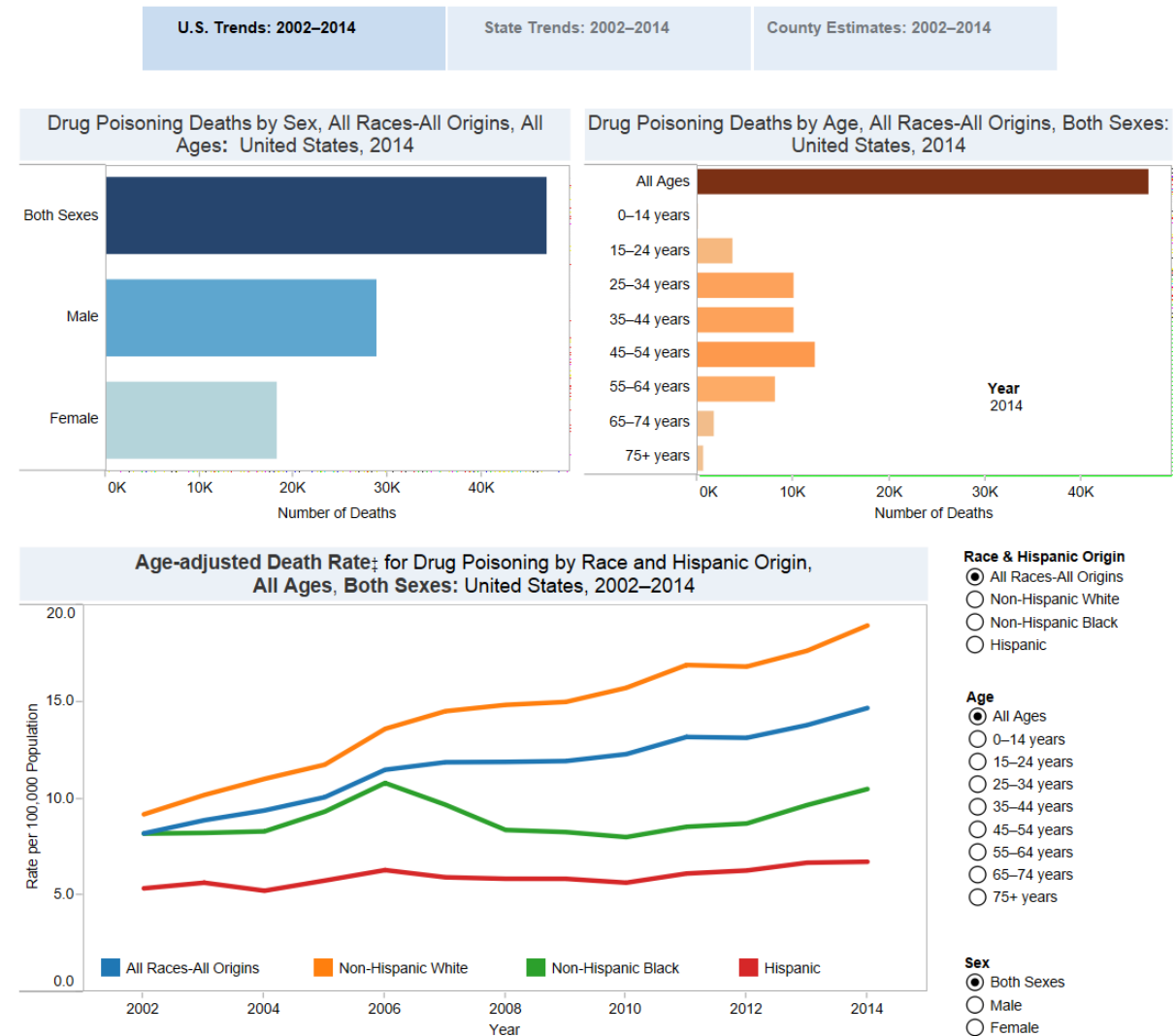
ArcGIS Online

Business Intelligence

e.g. **Tableau**, Qlik, SAS Visual Analytics

- Pros:
 - Flexible, but don't require much if any coding
 - Point and click interfaces
 - Good support/frequent updates
 - Some free public options
- Cons:
 - Most expensive
 - IT support for large implementations
 - Business-oriented user communities

Drug Poisoning Mortality: United States, 2002–2014



Designed by B. Bastian, L. Rossen & Y. Chong. SOURCE: CDC/NCHS, National Vital Statistics System.

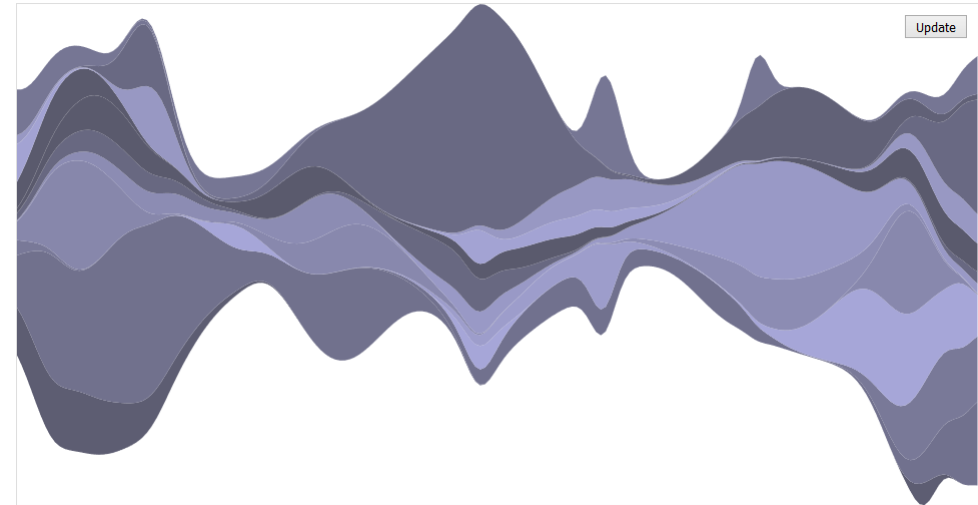
<https://public.tableau.com/s/gallery/fatal-drug-overdose-rates-united-states>

Tableau

Coding

e.g. **JavaScript(D3.js)**, R(ggplot), Python

- Pros:
 - Generally Free
 - If you have the time to learn it
 - Most flexible and powerful
- Cons:
 - Multiple languages necessary
 - Need to hire developer(s)
 - Time-intensive



```
<!DOCTYPE html>
<meta charset="utf-8">
<title>Streamgraph</title>
<style>

body {
  font-family: "Helvetica Neue", Helvetica, Arial, sans-serif;
  margin: auto;
  position: relative;
  width: 960px;
}

button {
  position: absolute;
  right: 10px;
  top: 10px;
}

</style>
<button onclick="transition()">Update</button>
<script src="//d3js.org/d3.v3.min.js"></script>
<script>

var n = 20, // number of layers
    m = 200, // number of samples per layer
    stack = d3.layout.stack().offset("wiggle"),
    layers0 = stack(d3.range(n).map(function() { return bumpLayer(m); })),
    layers1 = stack(d3.range(n).map(function() { return bumpLayer(m); }));
```

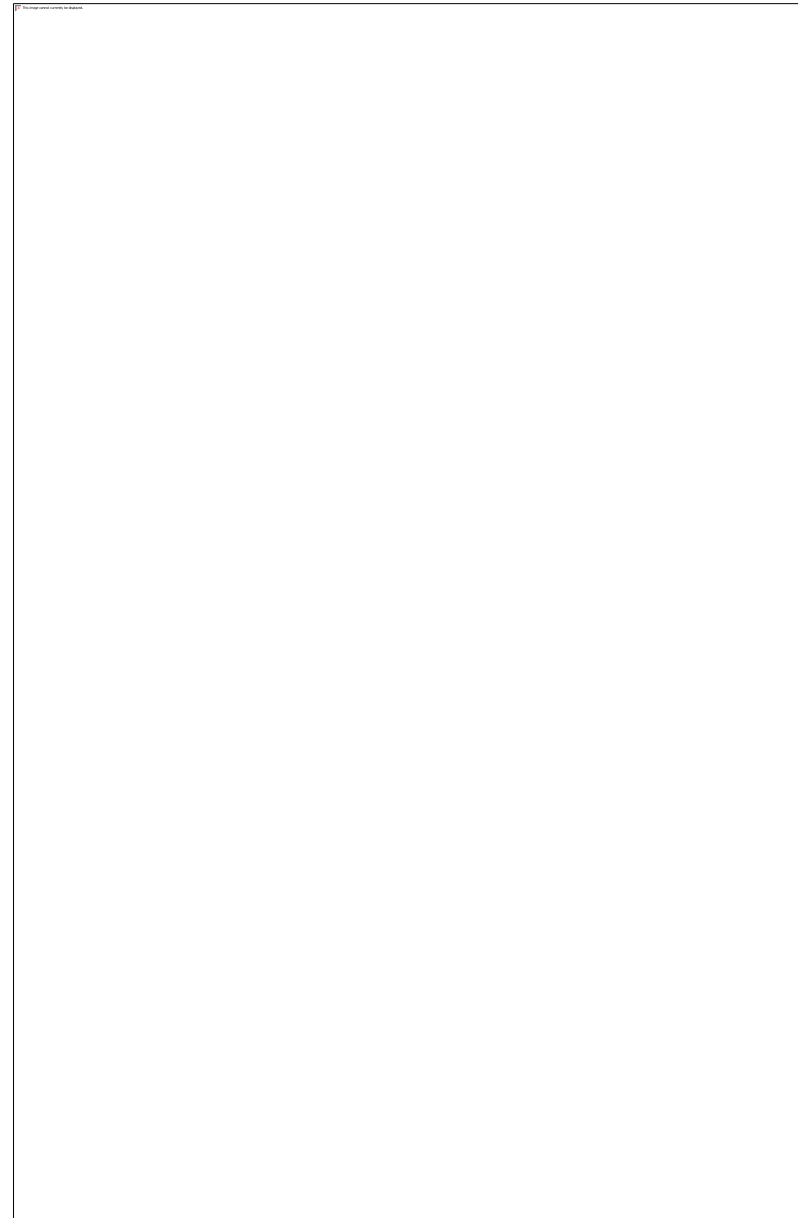
<http://bl.ocks.org/mbostock/4060954>

D3.js

Design

e.g. **Adobe Creative Suite**, Inkscape

- Pros:
 - Most aesthetically oriented
 - Can be combined with other tools
- Cons:
 - Expensive
 - Not data-oriented
 - 'Infographic effect'
 - Static



<http://icharts.net/blogs/2013/spotlight-interview-unique-approach-infographics-journalism-alberto-cairo>

Adobe Illustrator

Conclusions

- Easier tools come at a cost
- Time/Price:
 - Fancier is often more expensive
 - Are the free tools really free?
- Ever-growing spectrum
- Combining tools and knowing your data
- How much flexibility and fine tuning do you need?

Questions?

mtjansen@email.unc.edu

Learn more: Tools

- Tool lists:
 - <http://dirtdirectory.org/>
 - <http://selection.datavisualization.ch/>
- Map Galleries
 - CartoDB:
<https://cartodb.com/gallery/>
 - ArcGIS Online:
<http://www.arcgis.com/home/gallery.html#c=esri&t=maps&o=avgrating>
- Sample Galleries
 - D3:
<https://github.com/mbostock/d3/wiki/Gallery>
 - Tableau:
<https://public.tableau.com/s/gallery>
 - Plot.ly
<https://plot.ly/feed/>

Learn more:

Theory

Edward Tufte:

The Visual Display of Quantitative Information (2001)

Visual Explanations (1997)

Envisioning Information (1990)

Colin Ware:

Information Visualization:

Perception for Design (2004)

Practice

Stephen Few:

Show Me the Numbers (2004)

Information Dashboard Design (2006)

Now You See It (2009)

Alberto Cairo:

The Functional Art (2012)