

# Data Visualization in Practice

Rami Sayar - @ramisayar

Technical Evangelist

Microsoft Canada

# Agenda

- Quick Introduction to Data Visualization
- Five Data Visualization Principles & Best Practices to Follow
- How Successful Visualizations are Designed
- Tips and Mistakes Learned with D3.js

# Assumptions

- Working knowledge of JavaScript, D3.js and HTML5.
- Basic understanding of statistics.



code. ship. repeat.

# **What is Data Visualization?**

Data visualization helps people understand data through visual display.

# Why we do Data Visualization?

*Communicating  
knowledge clearly and  
efficiently.*

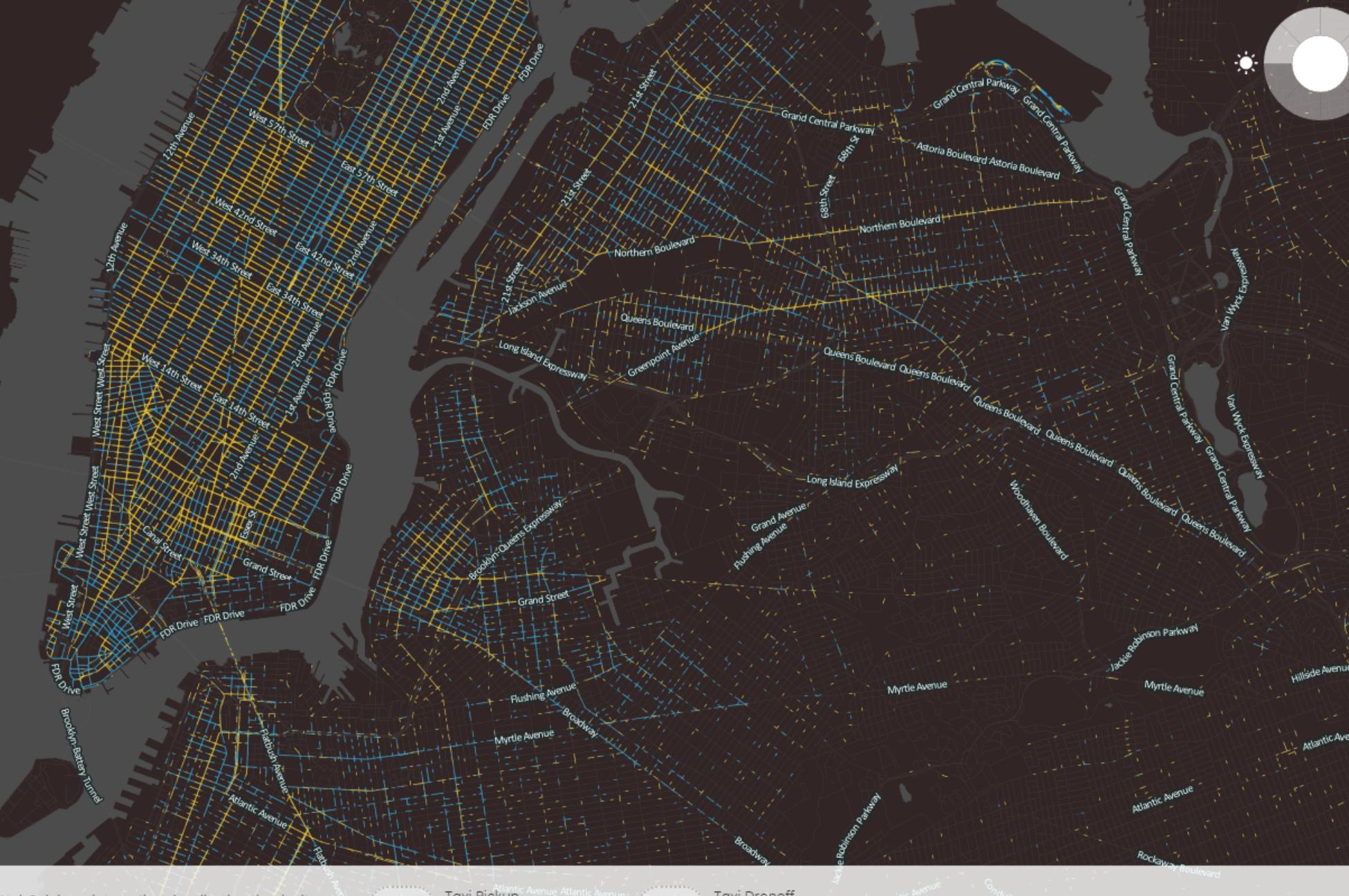
*Displaying data to  
understand cause and  
effect.*

**Both Art and Science.**

# hubcab

MIT  
senseable  
city lab:::  
Audi

HubCab is an interactive visualization that invites you to explore the ways in which over 170 million taxi trips connect the City of New York in a given year. [Show me how it works.](#)

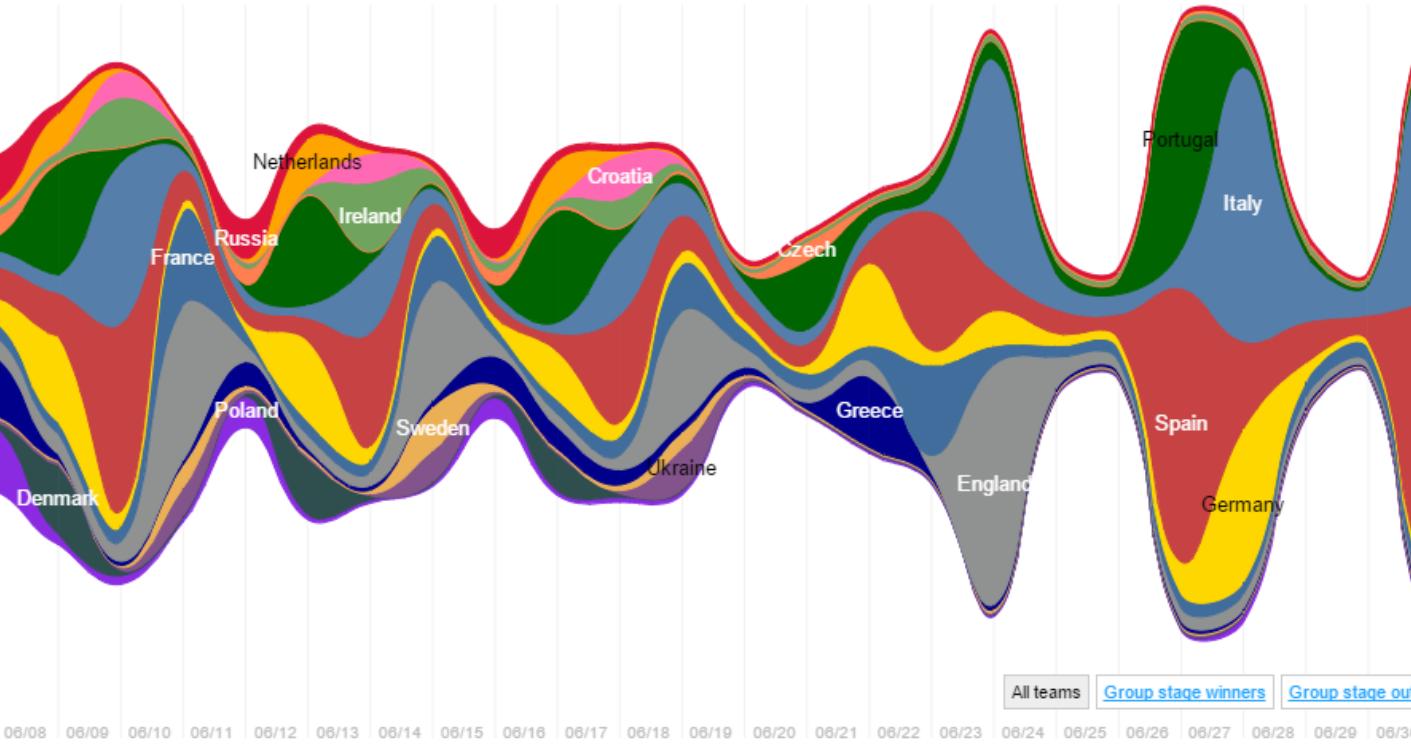




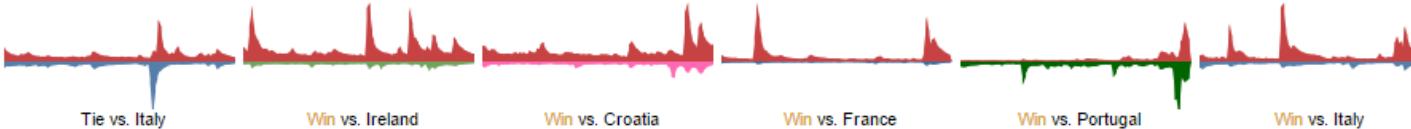
# #Euro2012

A summary for the action on Twitter during the European football tournament. [Tweet](#) 1,810

The Streamgraph below shows volume of Tweets during the #Euro2012 period. Click on a team's name to see details.



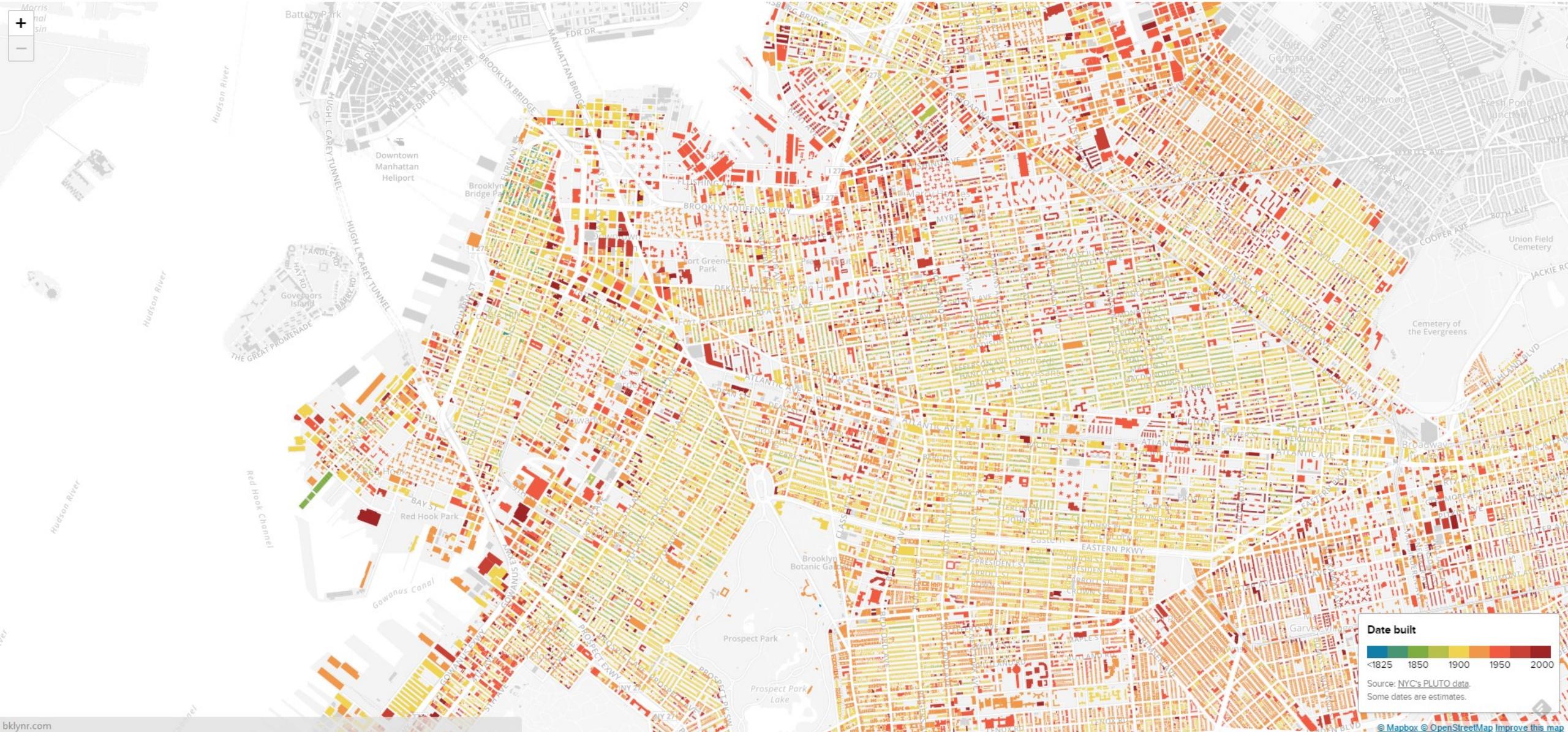
## ■ SPAIN



## ■ ITALY



# Block by Block, Brooklyn's Past and Present



# Los Angeles Times

INTERACTIVE MAP

## L.A. street quality grades

An analysis by the [Los Angeles Times Data Desk](#)

Enter an address in the Los Angeles city limits

Explore pavement quality ratings for each of the 68,000 street segments in L.A., graded from A to F.

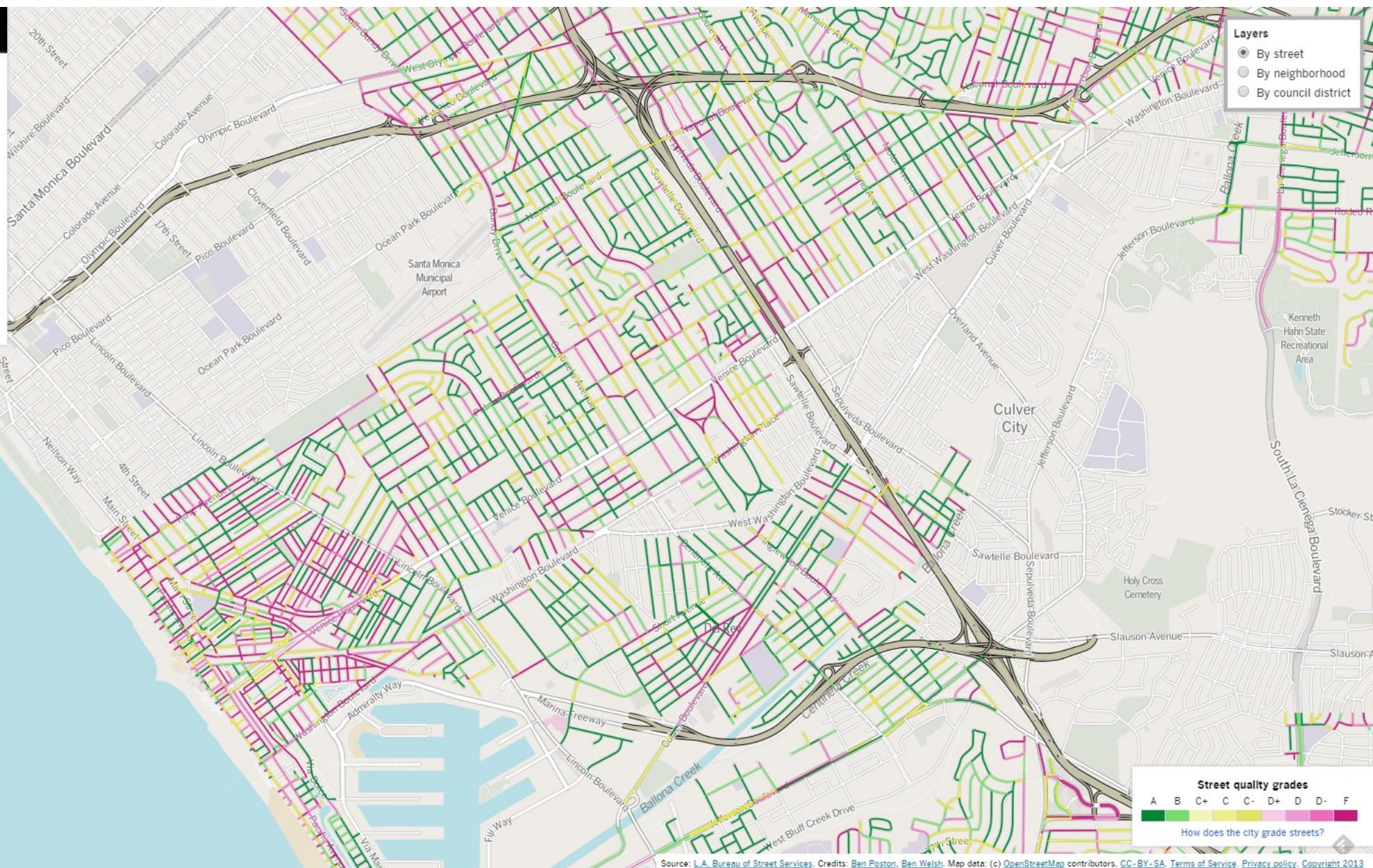
MORE: [Read the story](#) | [How we made this map](#)

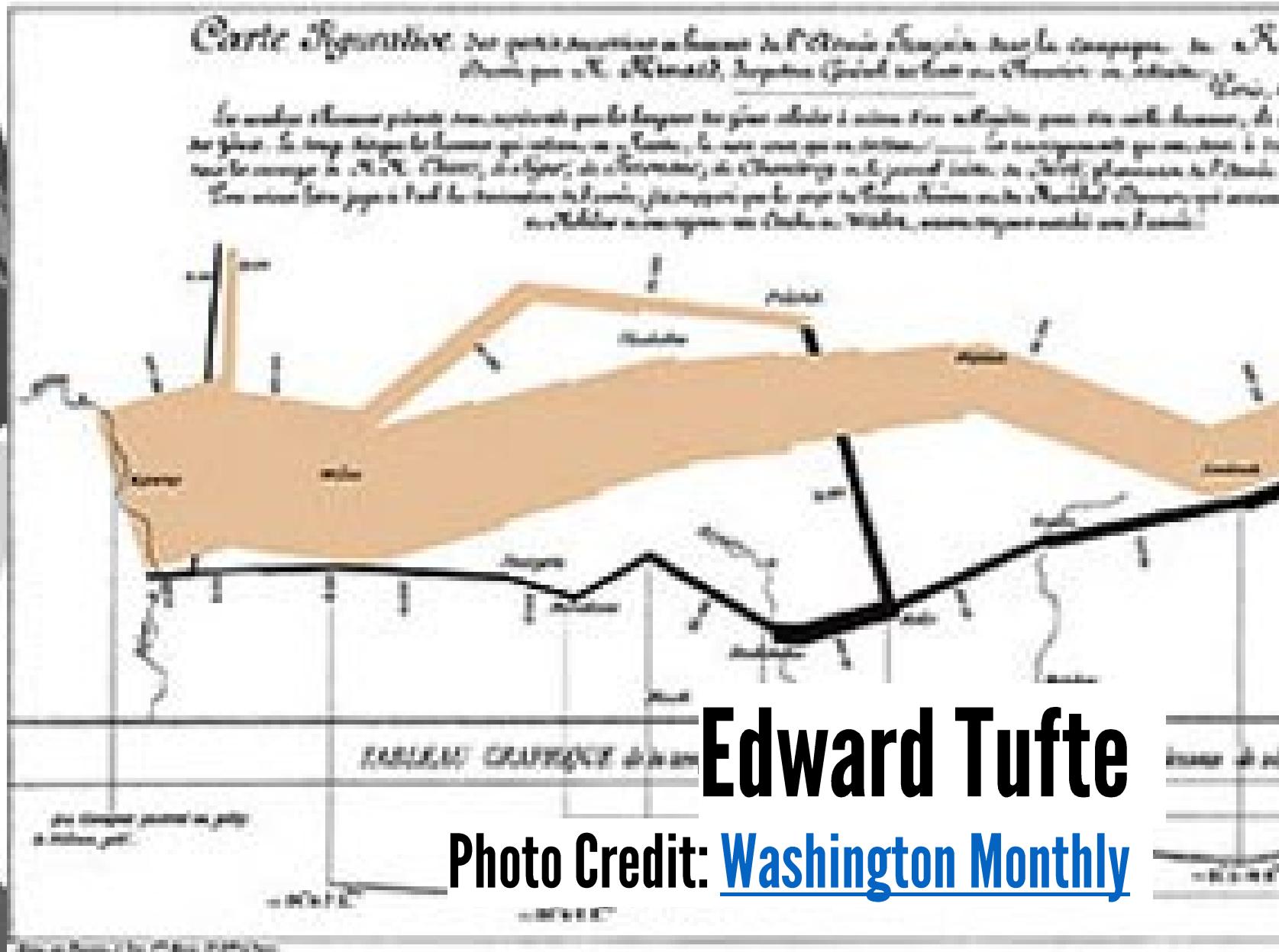
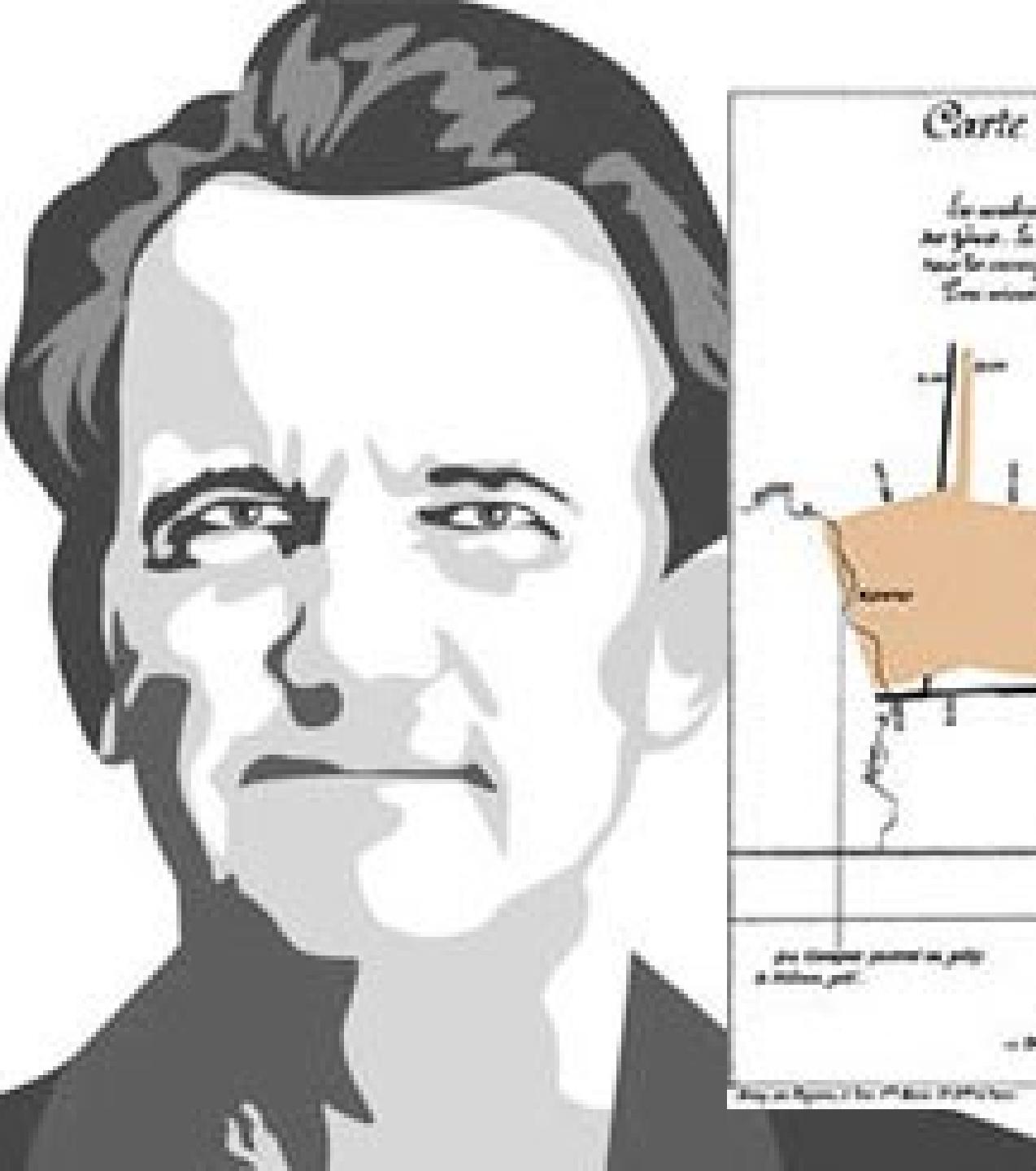
advertisement



Los Angeles Times

Bite Nite





# What is not Data Visualization?

“The purpose of visualization is insight, not pictures”  
Ben Schneiderman

A.K.A. It's not about aesthetics.

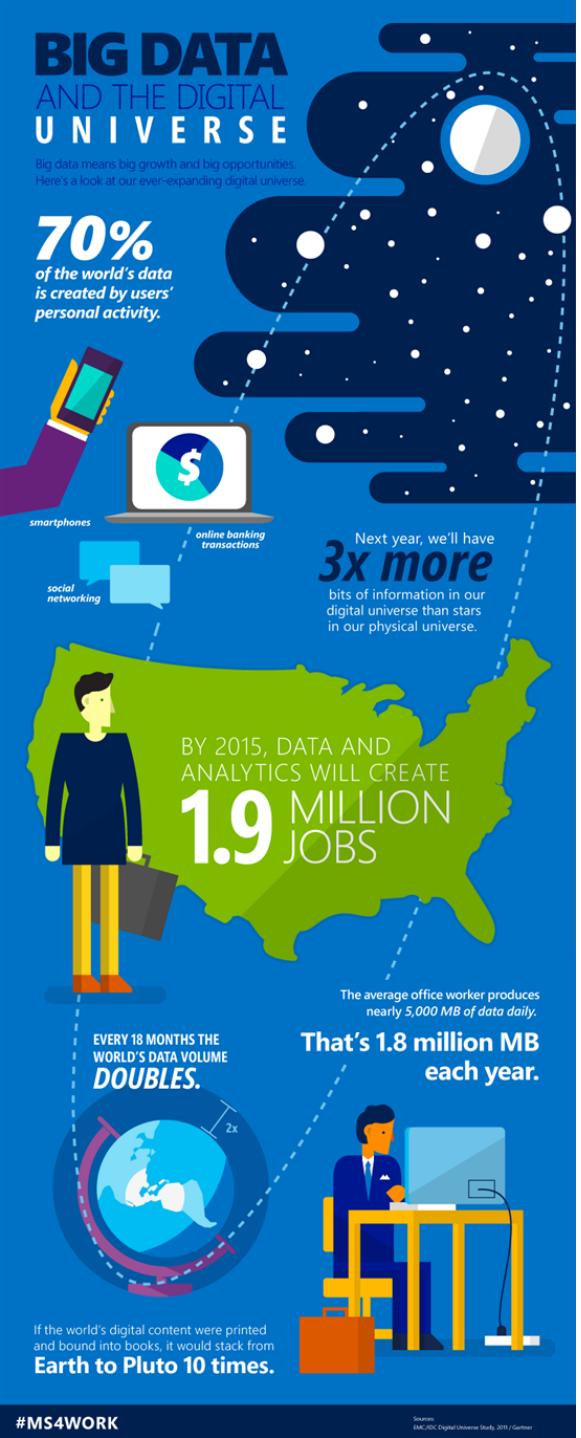
“#DataVisualization is not about creating infographics for your marketing department.  
- @RamiSayar #bigdata #dataviz”

I dare you to tweet that.

**Why should you care?**

**Besides that I am going to show you how to add real-time data visualizations to your front-end web apps and share some of the principles you need to create effective data visualizations...  
Here's why you really should care and listen up and tweet #FITCspotlight @ramisayar.**

# **Big Data**



Off course, I have  
to show you an  
infographic!  
Sorta...

LINK:

<http://blogs.microsoft.com/work/2014/02/12/big-data-and-the-digital-universe>

# **Data Science and Data Scientists**

Booming Field – Potentially Creating 1.9 Million Jobs.

# The BigData-Startups Open Source Landscape 2.0

<h3>Data Analysis &amp; Platforms</h3>	<h3>Databases / Data warehousing</h3>	<h3>In-Memory Computing</h3>
<h3>Business Intelligence</h3>	<h3>Data Mining</h3>	<h3>Programming</h3>
<h3>KeyValue</h3>	<h3>Document Store</h3>	<h3>Data aggregation</h3>
<h3>Object databases</h3>	<h3>Graph databases</h3>	<h3>Operational</h3>
	<h3>Multimodel</h3>	<h3>Multidimensional</h3>

**Photo Credit:** Datafloq

# **Big Data Visualization in Practice with D3.js**

# **Five Best Practices and Principles for Big Data Visualization.**

Let's start with this + some sample D3.js code.

# **Principle 1 - Context is King**

The context in which data is visually placed impacts the knowledge that can be gleaned or communicated.

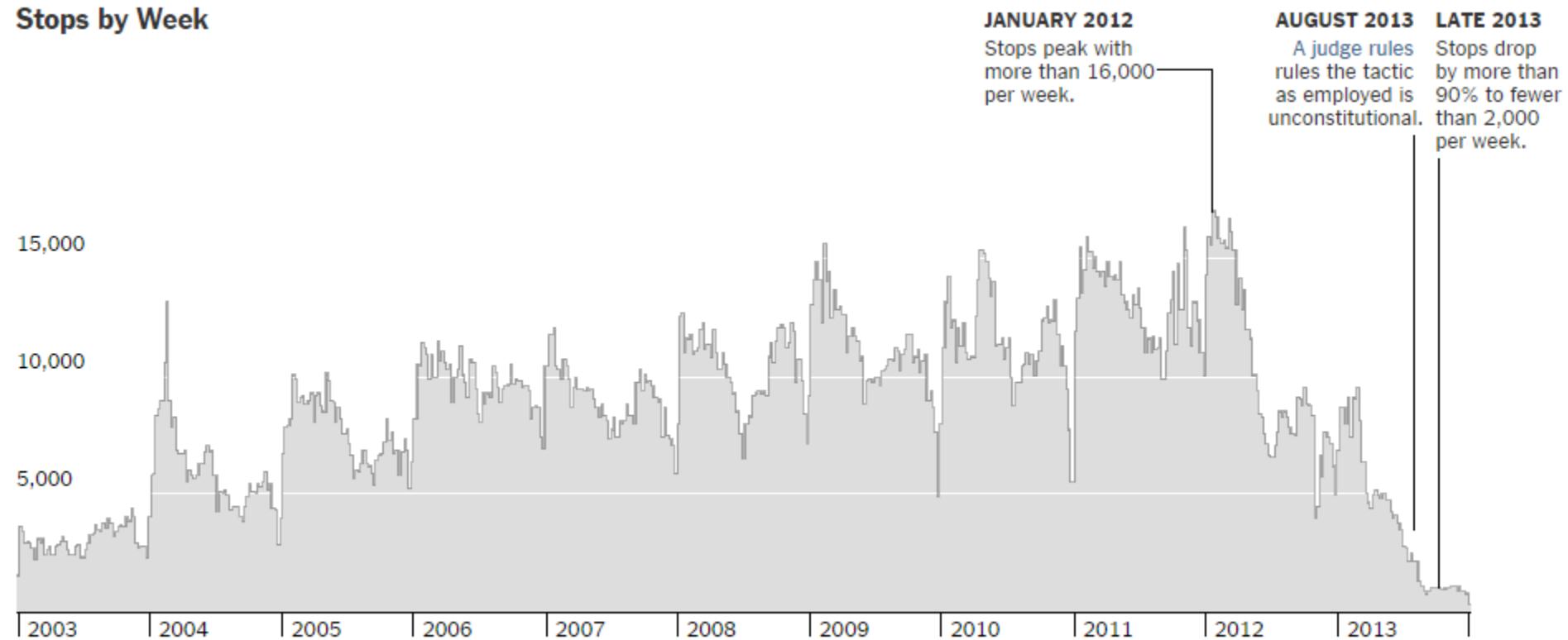
# Principle 1 - Context is King

Example: ['Stop-and-Frisk' Is All but Gone From New York](#) by Mike Bostock

- Controversial policing tactic that involved stopping and frisking people for what police deemed “suspicious behavior”.
- Report data composed of detailed info such as location, time, etc...

# Principle 1 - Context is King

Example: 'Stop-and-Frisk' Is All but Gone From New York by Mike Bostock  
Data Visualization Attempt 1:



Data Visualization Attempt 2

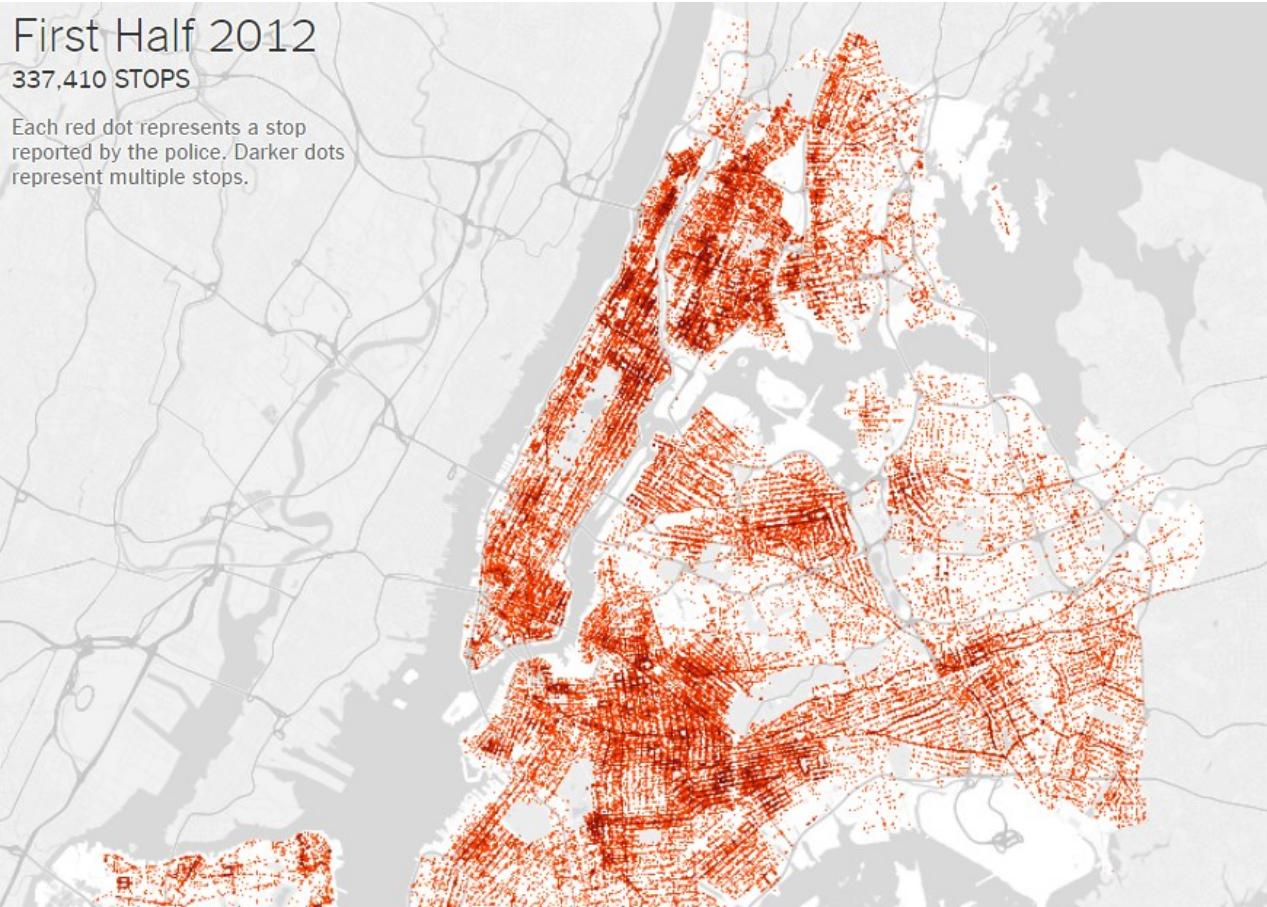
**Principle 1 – Context is King**

# Data Visualization Attempt 3

First Half 2012

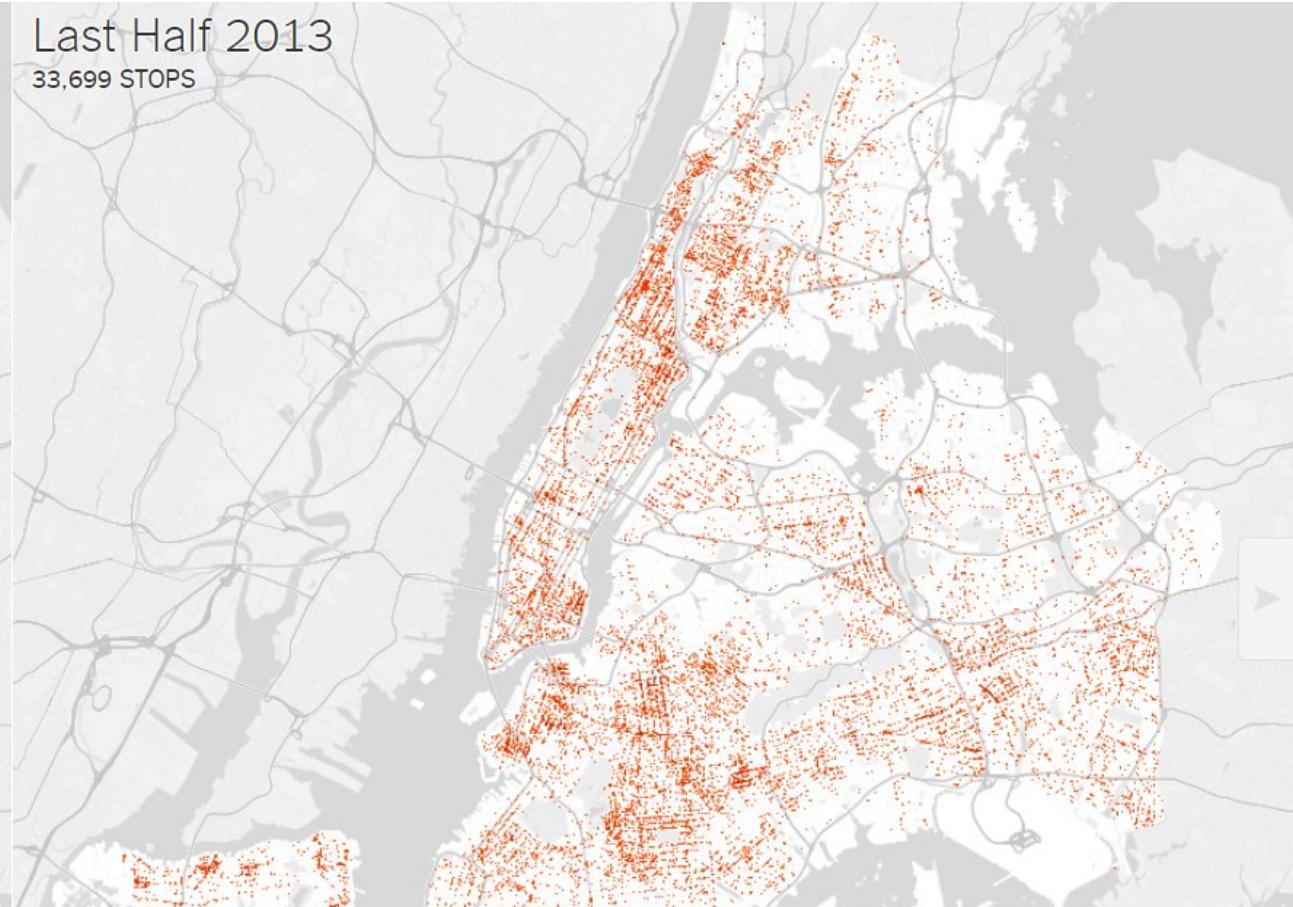
337,410 STOPS

Each red dot represents a stop reported by the police. Darker dots represent multiple stops.



Last Half 2013

33,699 STOPS



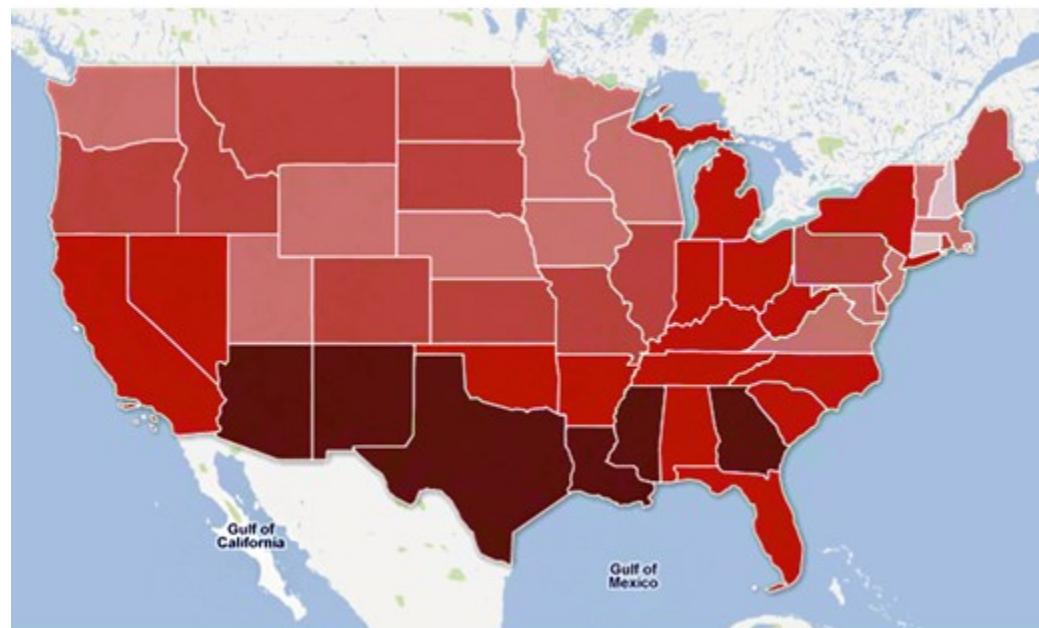
**Principle 1 - Context is King**

Data Visualization Attempt 4

# Principle 1 – Context is King

# Principle 1 - Context is King

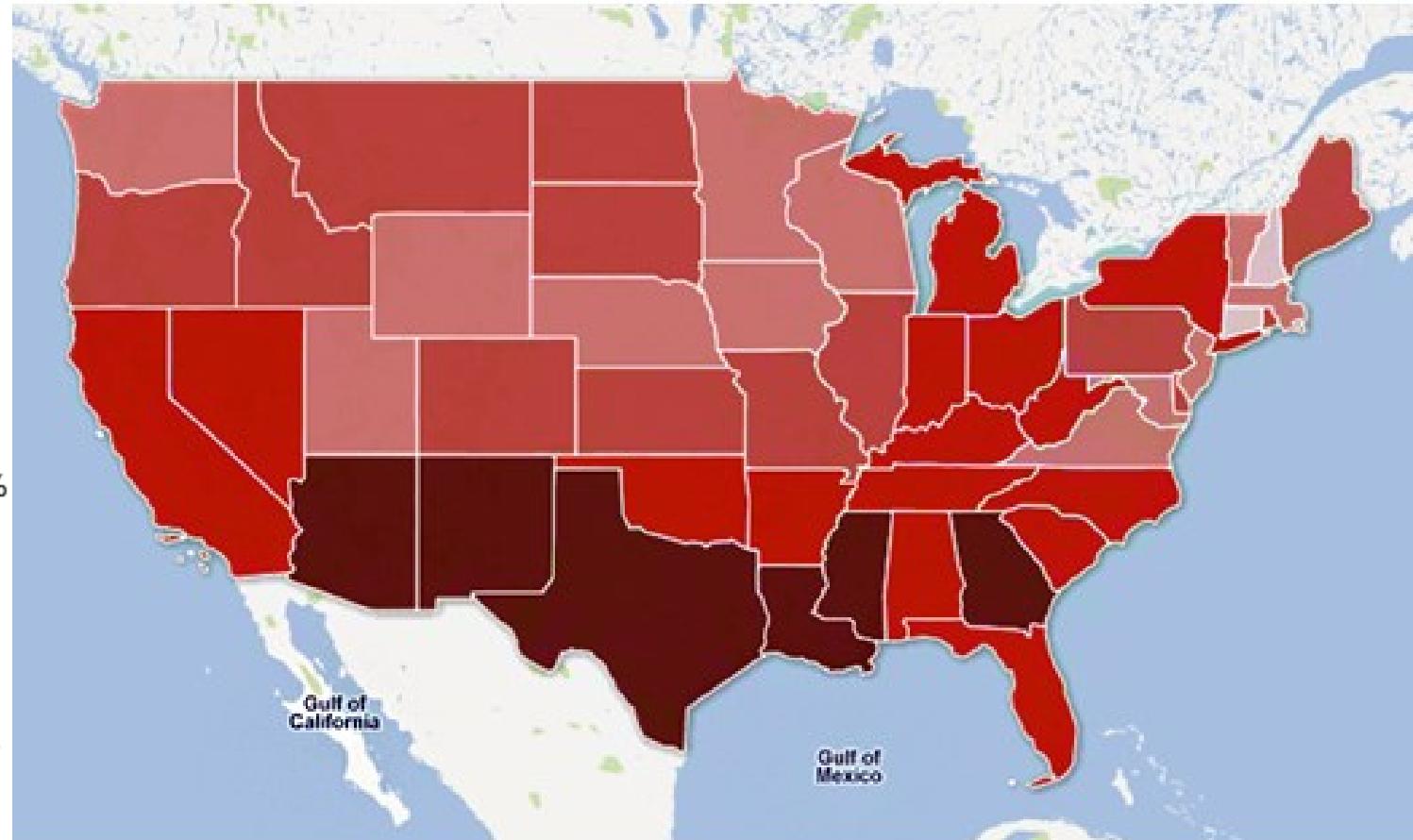
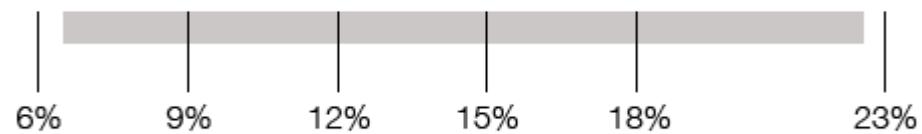
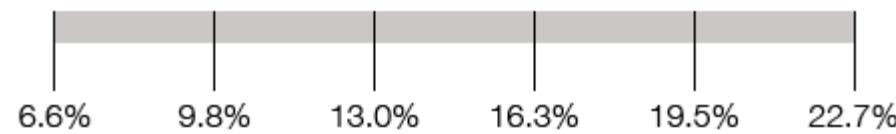
Example: [Take Care of your Choropleth Maps](#) by Gregor Aisch  
Guardian data blog published a US poverty map...



# Principle 1 – Context is King

Example: [Take Care of your Choropleth Maps](#) by Gregor Aisch

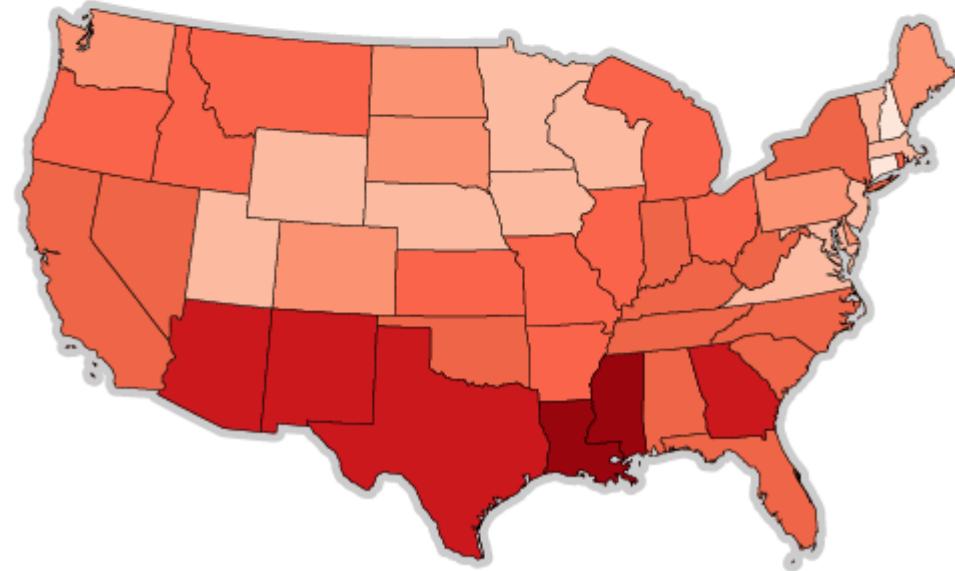
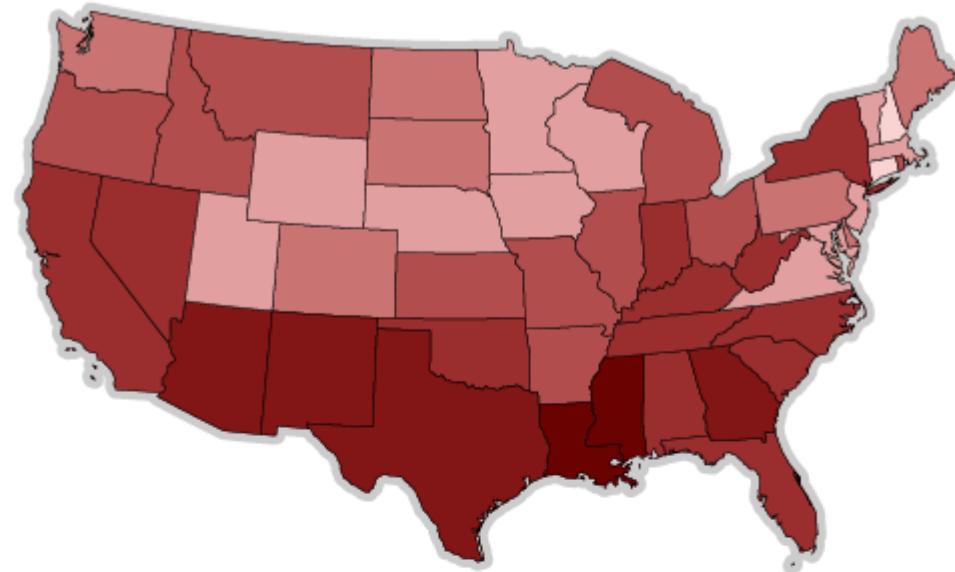
Don't mess around with your class limits.



# Principle 1 – Context is King

Example: [Take Care of your Choropleth Maps](#) by Gregor Aisch

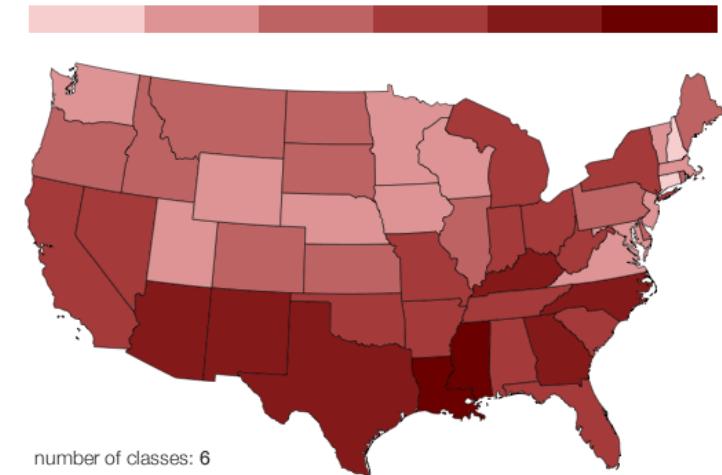
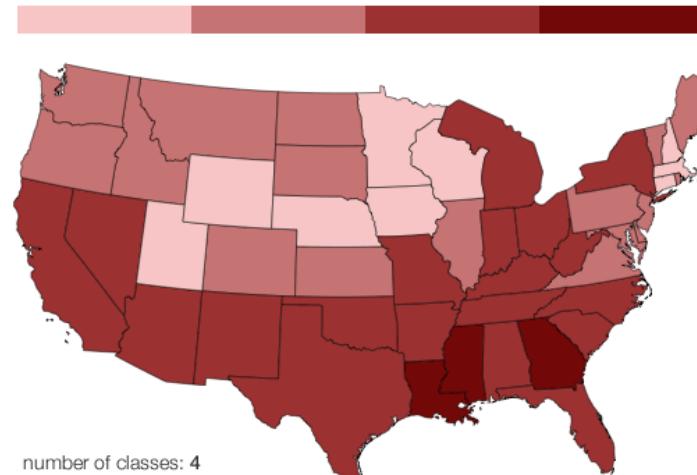
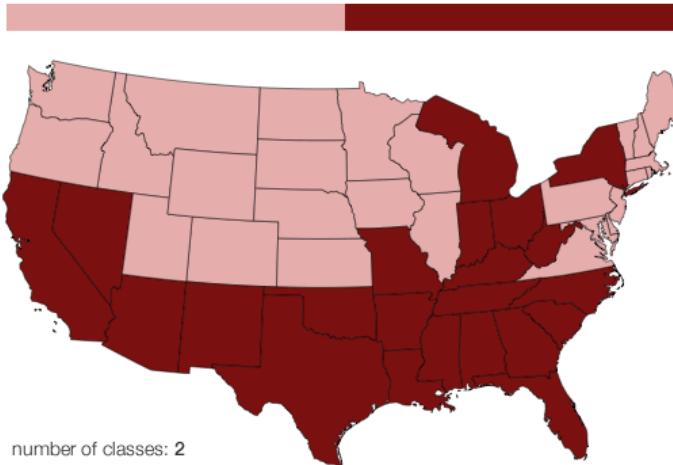
Don't mess around with your class colors.



# Principle 1 - Context is King

Example: [Take Care of your Choropleth Maps](#) by Gregor Aisch

- Place data in comparison for context, e.g. class count.



# Principle 1 - Context is King

## Mapping with D3.js

- D3.js includes routines for handling geographic information.
- GeoJSON is the geographic data file of choice. Fairly complicated process to convert primary mapping information to GeoJSON.
  - Instead use: [johan/world.geo.json](#)
- TopoJSON is an extension of GeoJSON that encodes topology instead of geometrics -> smaller data files.

# Principle 1 - Context is King

## Mapping with D3.js

- D3.js includes routines for handling geographic information.
- GeoJSON is the geographic data file of choice. Fairly complicated process to convert primary mapping information to GeoJSON.
  - Use: [johan/world.geo.json](#)
- TopoJSON is an extension of GeoJSON that encodes topology instead of geometrics -> smaller data files.
  - Use: [mbostock/world-atlas](#)
  - Use: [mbostock/us-atlas](#)

# Principle 1 - Context is King

Mapping with D3.js

Demo: [Area Choropleth](#)

Demo: [U.S. TopoJSON](#)

# **Principle 1 - Context is King**

- The context in which data is visually placed impacts the knowledge that can be gleaned or communicated.
- Enforce the right comparisons for the context.
- Many problems are multivariate (i.e. multiple variables) and that needs to be recognized in the data visualization.

# **Principle 2 - Visualizations Must Match Data**

The knowledge communicated through visualizations must match the underlying data.

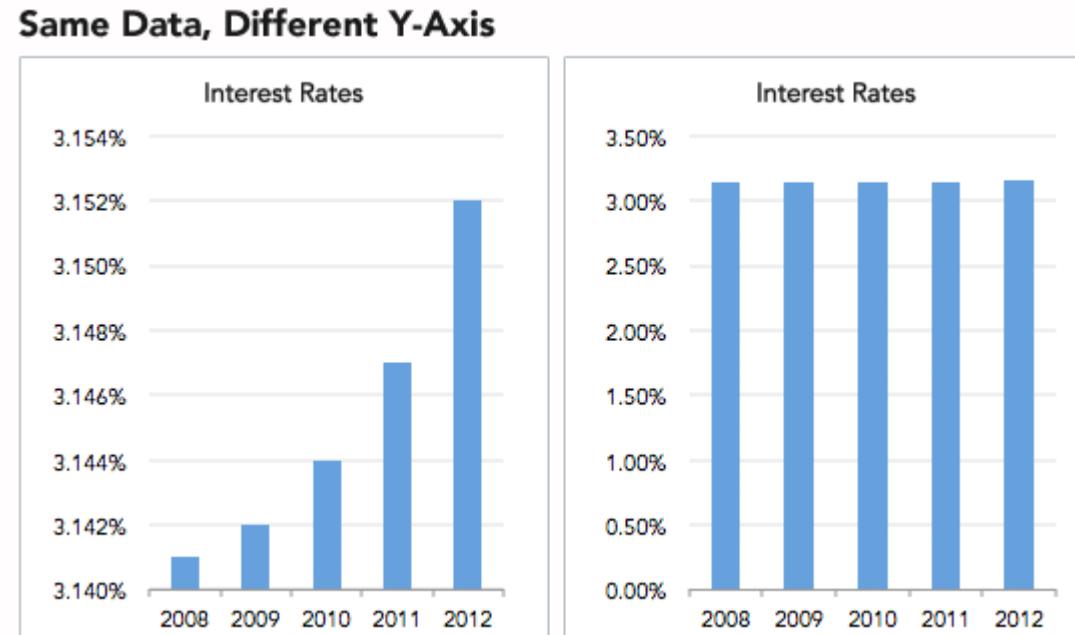
# **Or Anti-Principle 1 – How to use Data Visualization to Mask, Cheat and Lie.**

Because I know that's what some of you are thinking...

# Principle 2 – Visualizations Must Match Data

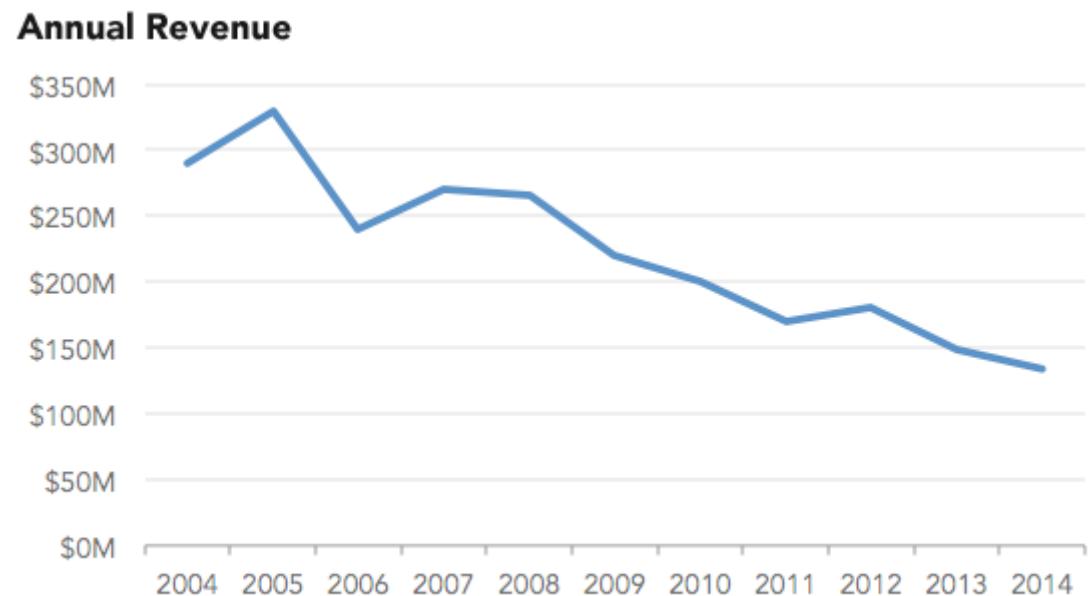
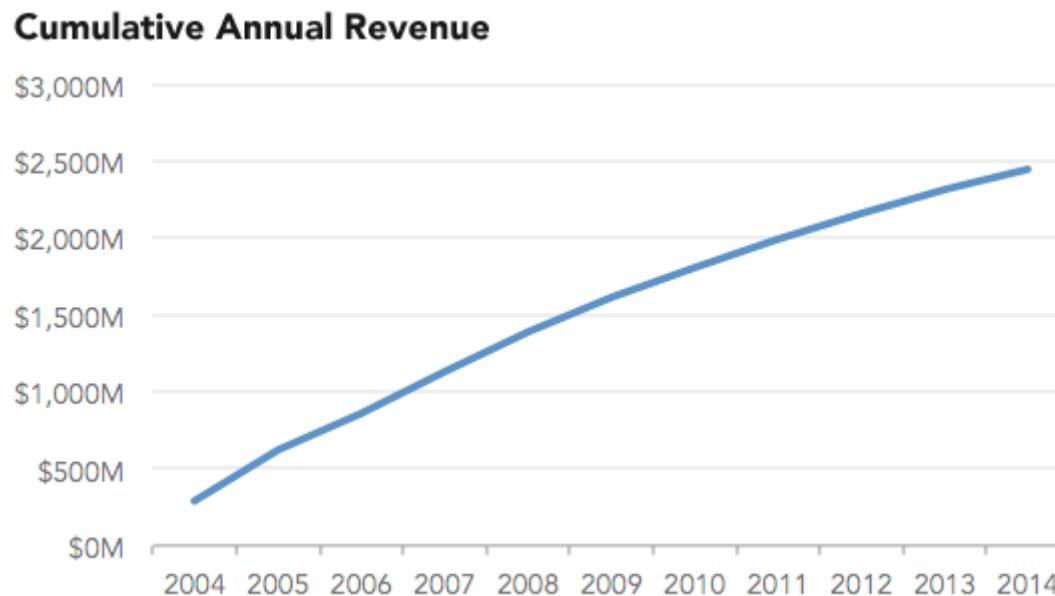
- Fairly obvious point but most people miss this surprisingly... or they do it knowingly...

Example: Truncated Y-Axis from [How to Lie with Data Visualization](#) by Ravi Parikh



# Principle 2 – Visualizations Must Match Data

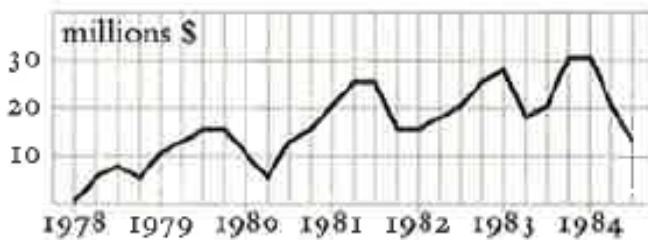
Example: Cumulative Graphs from [How to Lie with Data Visualization](#) by Ravi Parikh



# Principle 2 – Visualizations Must Match Data

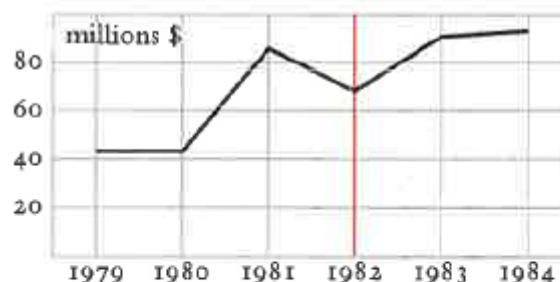
Example: Effects of Temporal Aggregation from Modern Visual Evidence by Gregory Joseph, discovered in Visual Explanations by Edward Tufte.

Quarterly Revenue



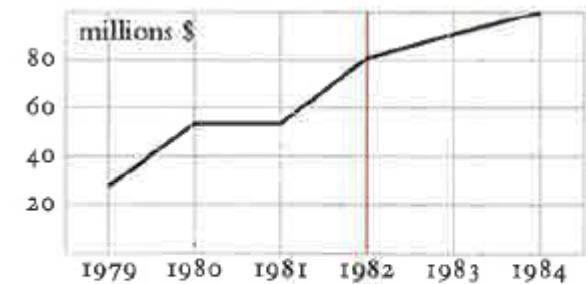
Above, this chart shows *quarterly* revenue data in a financial graphic for a legal case. Several dips in revenue are visible.

Revenue by Fiscal Year (July 1 to June 30)



Aggregating the quarterly data into years, this chart above shows revenue by *fiscal year* (beginning July 1, ending June 30). Note the dip in 1982, the basis of a claim for damages.

Revenue by Calendar Year



Shown above are the same quarterly revenue data added up into *calendar years*. The 1982 dip has vanished.

# Principle 2 – Visualizations Must Match Data

## Demo: Crossfilter

Crossfilter is a JavaScript library for exploring large multivariate datasets in the browser. Crossfilter supports extremely fast (<30ms) interaction with coordinated views, even with datasets containing a million or more records.

# Principle 2 – Visualizations Must Match Data

- The knowledge communicated through visualizations must match the underlying data.
- Follow convention in how you model your data and axis.
- The objective of data visualization is communicate information to the viewer, misleading by deception or confusion (even accidentally) will not serve your purpose.

# **Principle 3 - Escape Flatland**

Although devices present data in two dimensions, this desolate flatland can be escaped.

# Principle 3 – Escape Flatland

Example: Using Three Dimensions to Show Skyline Changes.

[The Bloomberg Years: Reshaping New York by NYTimes](#)

[Demo](#)

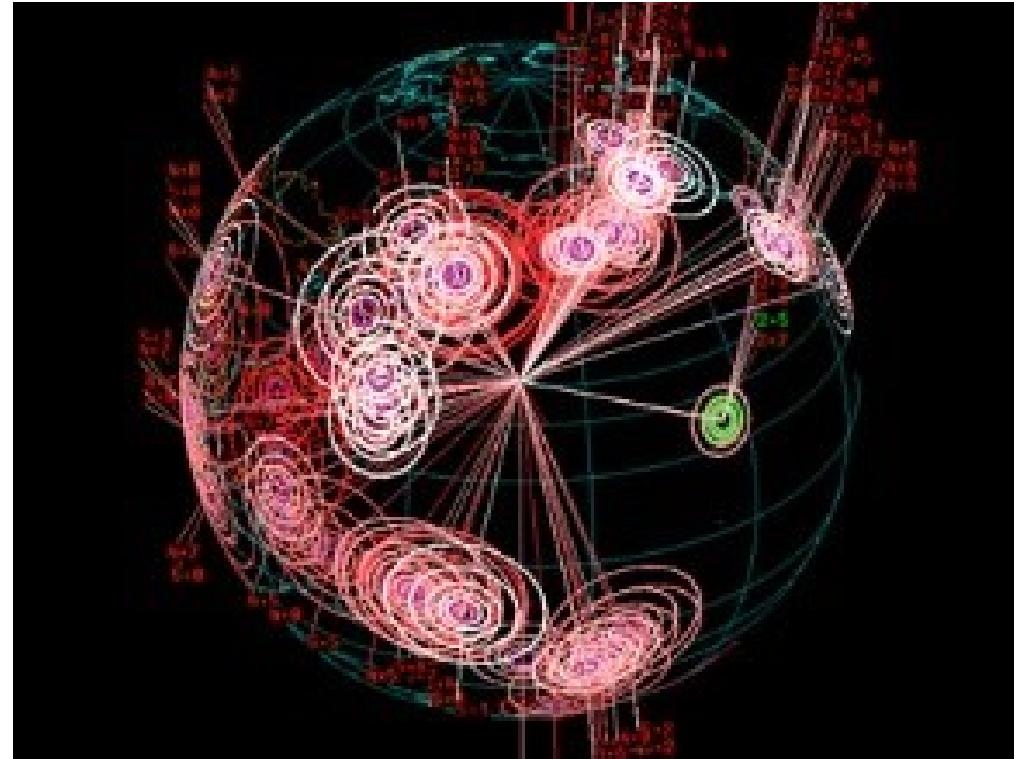


Example: [Visualizing Friendships](#) by Paul Butler  
**facebook**

December 2010

# Principle 3 – Escape Flatland

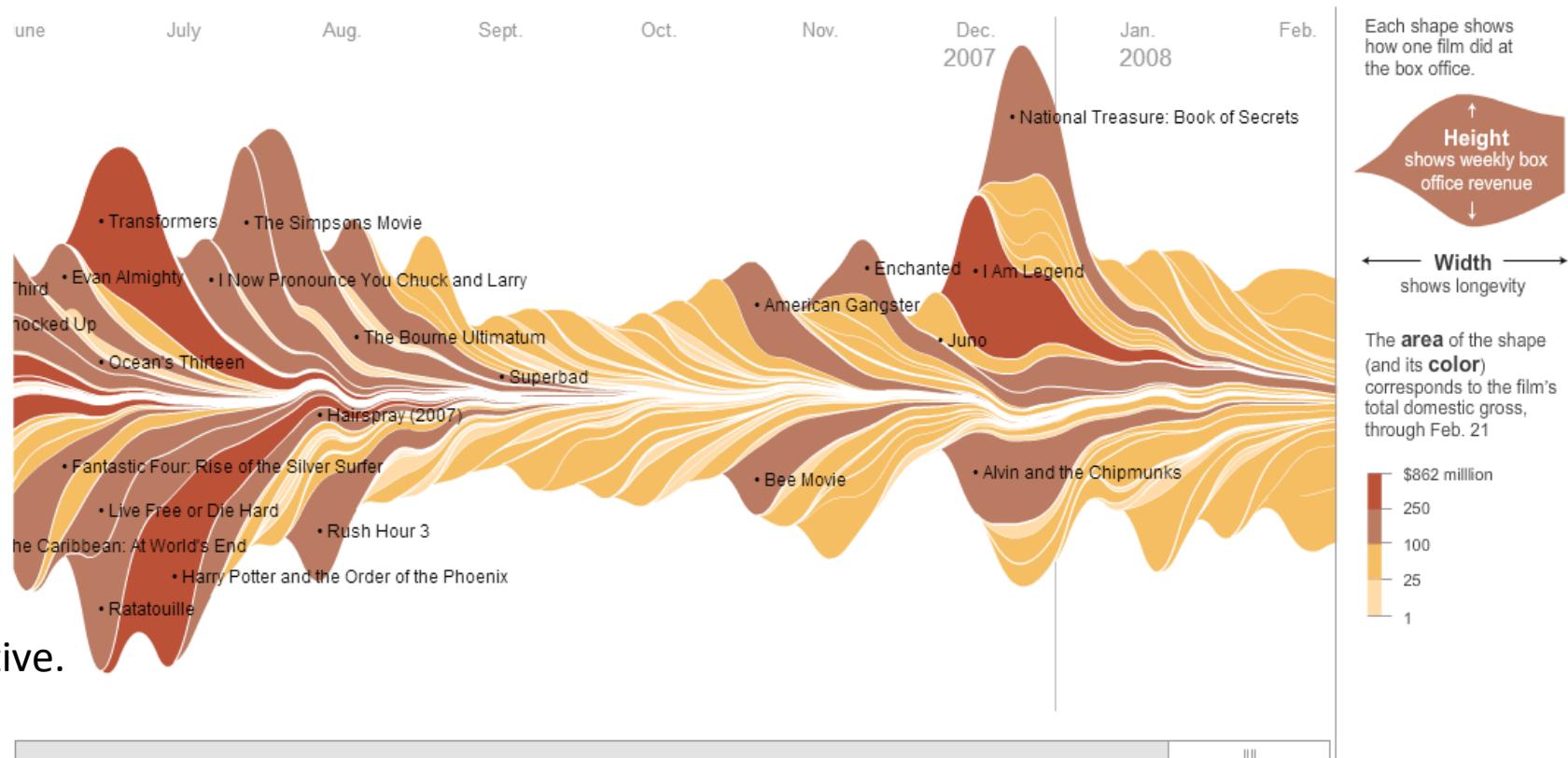
Example: Using Three Dimensions to Convey Geographic Information at a Global Scale



# Principle 3 – Escape Flatland

Using three dimensions incorrectly adds no value or obfuscates the data visualization.

Example: [The Ebb and Flow of Movies: Box Office Receipts by NYTimes](#).



# Principle 3 – Escape Flatland

## UK Temperature History

Explore the temperature in the United Kingdom since 1910 by scrolling the page up and down. Visit years marked in red or blue that have significant weather events. You can also sort by year, maximum, minimum or mean temperature

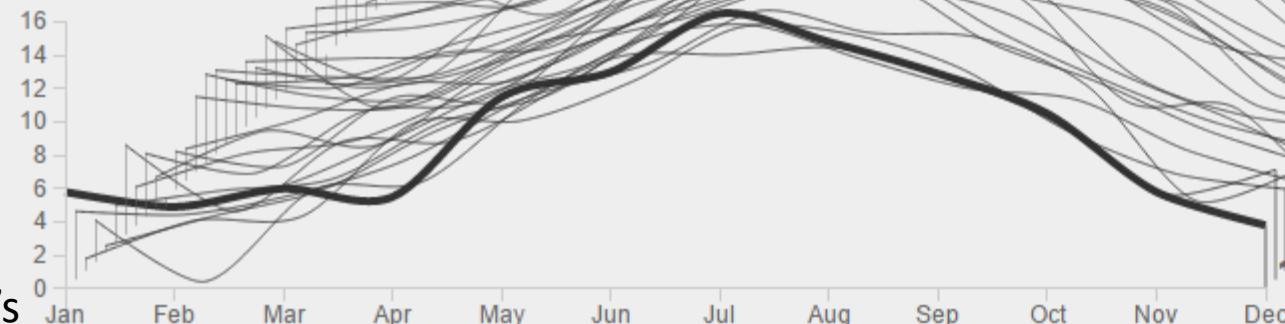
Sort by: Year / Maximum / Minimum / Mean

**1989**

Hottest month: 16.5°C

Coolest month: 3.8°C

Year average: 9.3°C



[Link: UK Temperature History](#)

**1989**

**1983**

**1976**

**1963**

**1947**

**1911**

Yes I know, it's  
meant to be interactive.

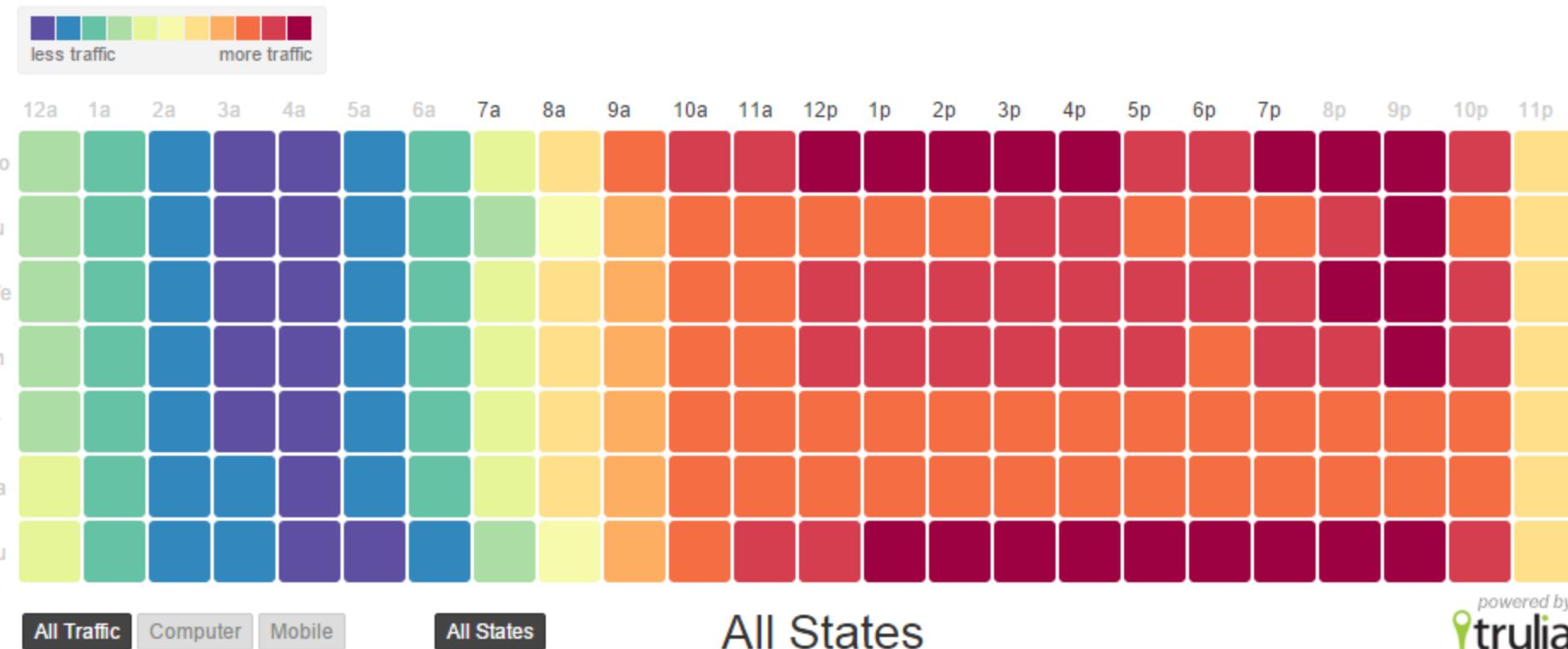
# **Principle 4 – Show Your Work**

Aggregating details can reveal the knowledge present in data.

# Principle 4 – Show Your Work

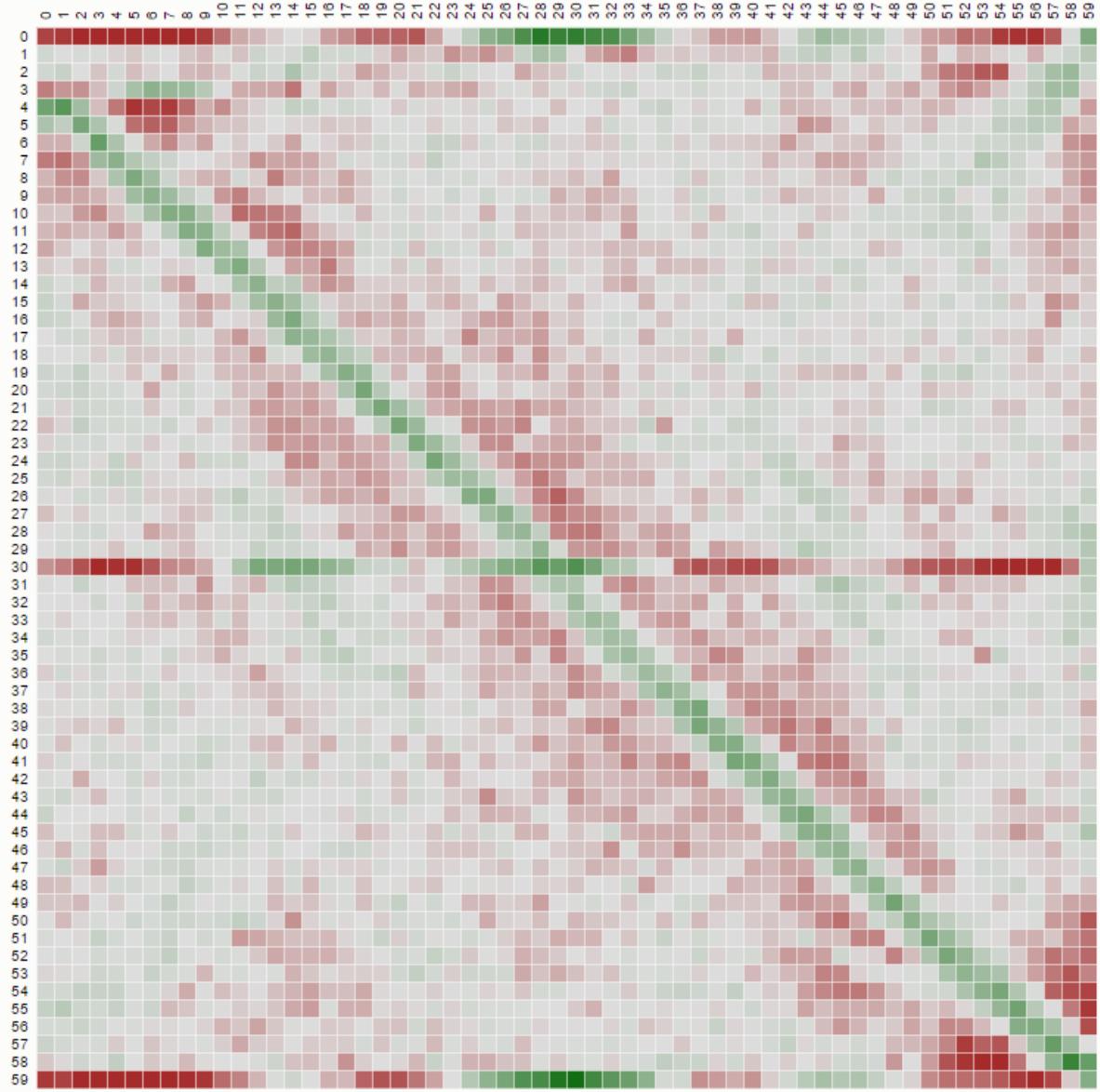
On occasion, showing all the data points in one big visualization can communicate knowledge clearly.

Example: [House Hunting All Day, Every Day](#) by Trulia Trends



# Will It Shuffle?

Or, why random comparators are bad (in addition to being slow).



## Principle 4 – Show Your Work

Example: [Will it Shuffle?](#) By Mike Bostock

Run SHUFFLE DEMO.

## Principle 4 – Show Your Work

Example: [Life Expectancy](#) by Nathan Yau

# Life Expectancy

People are living longer around the world, some more so than others. Select a region (as defined by World Bank) below to compare, or roll over to the graph to highlight countries. [Read more...](#)

[East Asia and Pacific](#) [South Asia](#) [Europe and Central Asia](#) [Middle East and North Africa](#) [Sub-Saharan Africa](#) [Latin America and Caribbean](#) [North America](#)

### WORLD

The average life expectancy in the world in 2009 was 69 years.



Source: The World Bank; Graphic by: [Nathan Yau](#)

# **Principle 5 - Insights from Hay Bales**

Layering and parallelizing data visualizations can reveal insights but be careful not to form haystacks.

# Principle 5 – Insights from Hay Bales

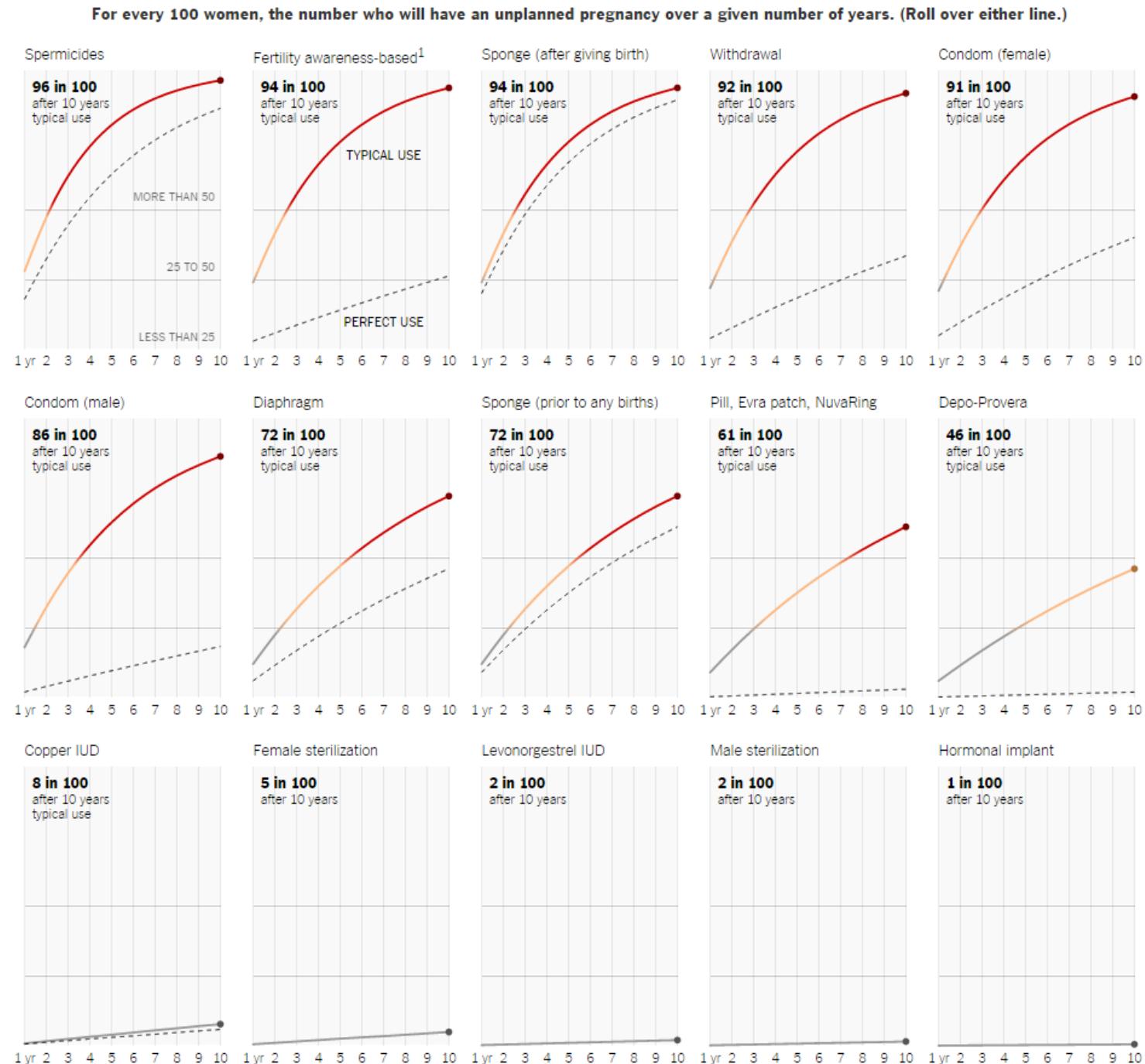


# Principle 5 – Insights from Hay Bales

- Layering data on a common X Axis maximizes visualization of coincidence and anomalies. Best use for time series data.
  - Use Color and Transparency.
- Parallelizing data is as powerful as layering data to show significant differences between multiple data sets.

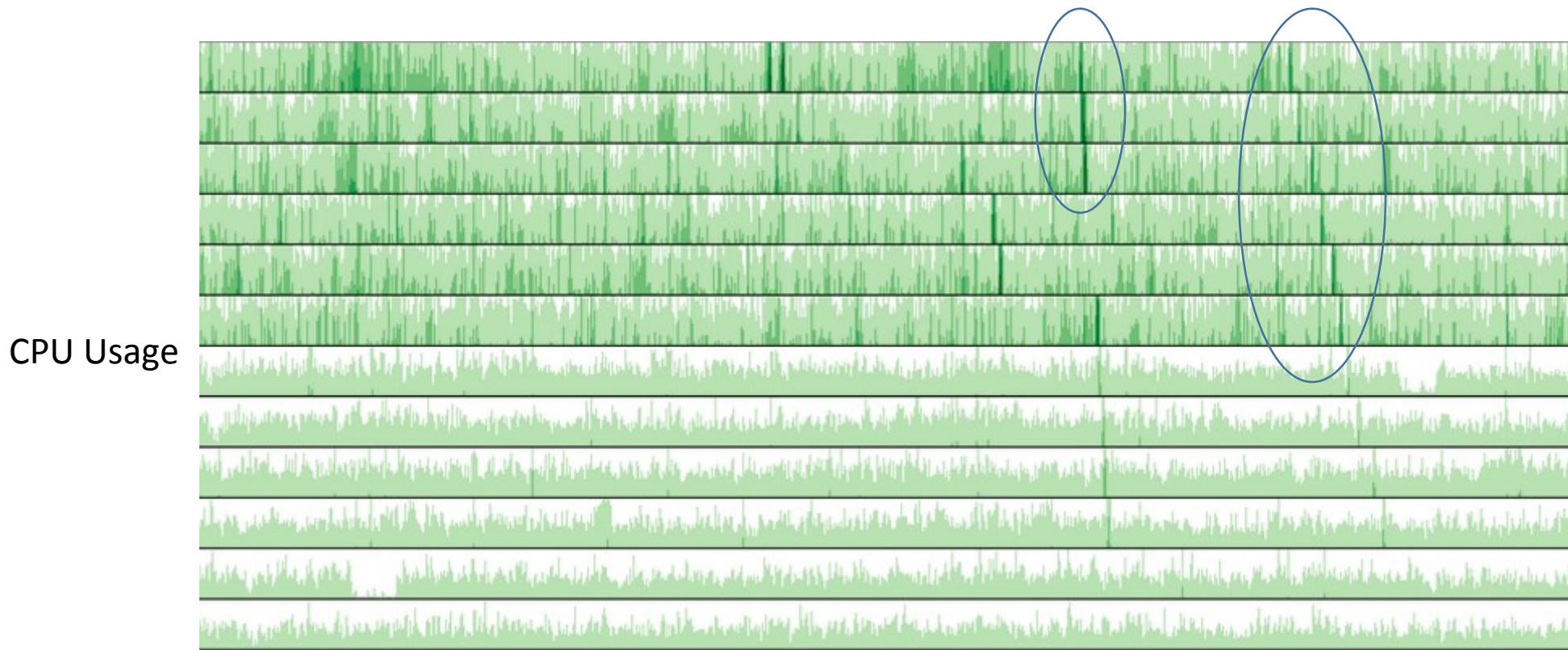
# Principle 5 – Insights from Hay Bales

# How Likely Is It That Birth Control Could Let You Down? By Gregor Aisch and Bill Marsh.



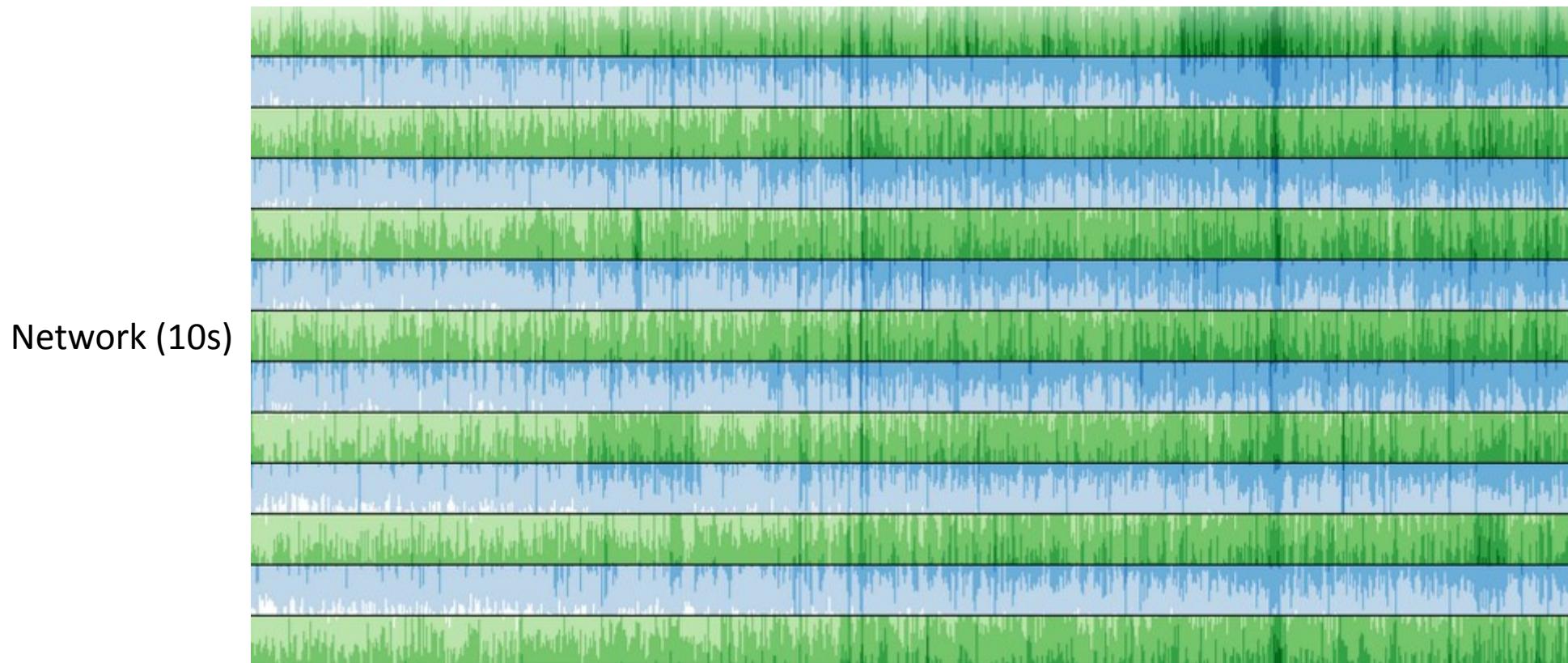
# Principle 5 – Insights from Hay Bales

Example: [Introduction to Cubism](#) by Mike Bostock



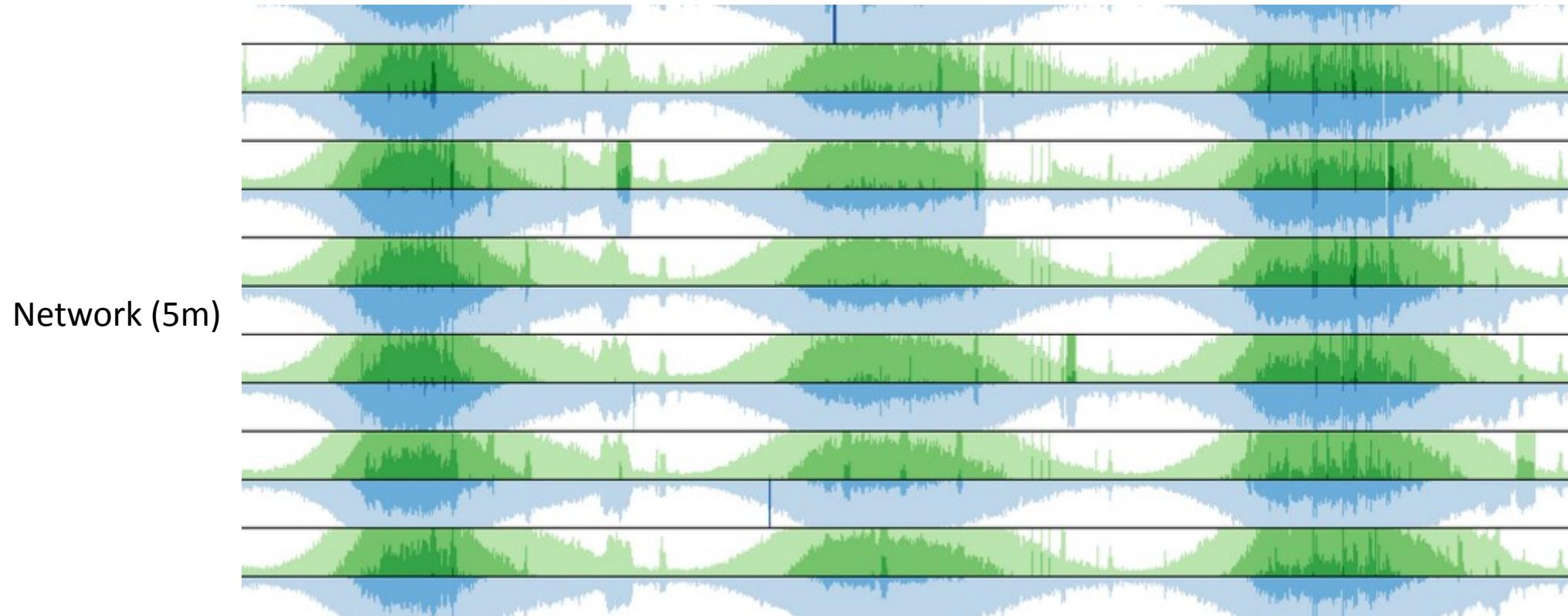
# Principle 5 – Insights from Hay Bales

Example: [Introduction to Cubism](#) by Mike Bostock



# Principle 5 – Insights from Hay Bales

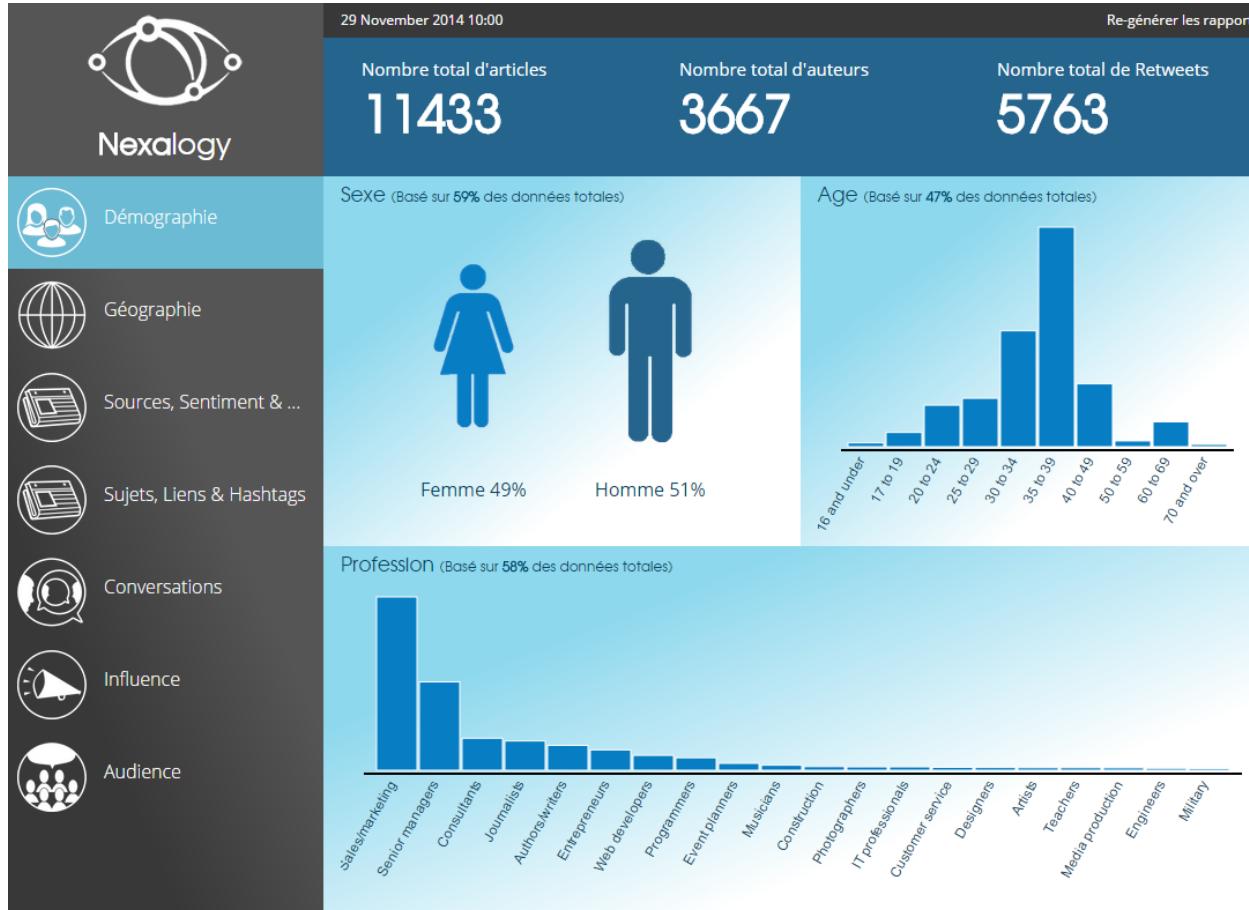
Example: [Introduction to Cubism](#) by Mike Bostock



# Principle 5 – Insights from Hay Bales

Demo Cubism Sample from [github.com/xaranke/cubism-intro](https://github.com/xaranke/cubism-intro)

# Case Study - Nexalogy



# Books + Websites + Conferences

## Books

- Anything by Edward Tufte
- Knowledge/Information is Beautiful
- Designing News by Franchi

## Conferences

- Visualized in NYC

## Websites

- [InformationisBeautiful.net](http://InformationisBeautiful.net)
- [FlowingData.com](http://FlowingData.com)
- [Visualizing.org](http://Visualizing.org)
- [Information Aesthetics](http://InformationAesthetics.com)
- [VisualComplexity.com](http://VisualComplexity.com)
- [Mike Bostock](http://MikeBostock.com)
- [D3.js Gallery](http://D3.jsGallery.com)
- [Driven-by-data.net](http://Driven-by-data.net)

# People to Follow on Twitter

- **Edward Tufte** - Data Scientist & Professor of PoliSci, Statistics, CS at Yale + Princeton for 33 years - @EdwardTufte
- **Mike Bostock** – Graphics Editor at NYTimes + Creator of D3.js @mbostock
- **Jonathan Corum** - Information Designer and NYTimes Science Graphics Editor - @13pt
- **Bret Victor** – Purveyor of Impossible Dreams - @worrydream
- **Gregor Aisch** – Graphics Editor at NYTimes + Creator of kartograph.js & datawrapper.js - @driven\_by\_data

# What did we learn?

- Introduction to Data Visualization
- Five Data Visualization Principles & Best Practices to Follow
  - Context is king.
  - Visualizations must match the data.
  - Escape flatland when useful.
  - Aggregating details can reveal the knowledge present in data.
  - Layering and parallelizing data visualizations can reveal insights but be careful not to form haystacks.
- How Successful Visualizations are Designed

# Thank You! Questions?

Follow @ramisayar



©2013 Microsoft Corporation. All rights reserved. Microsoft, Windows, Office, Azure, System Center, Dynamics and other product names are or may be registered trademarks and/or trademarks in the U.S. and/or other countries. The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation. MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.