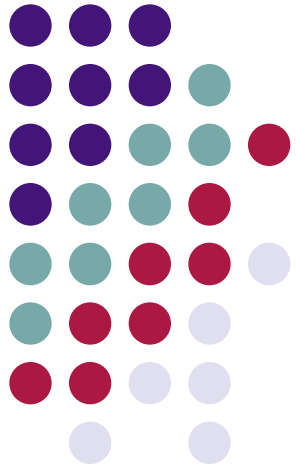
Visualizing Data for Information

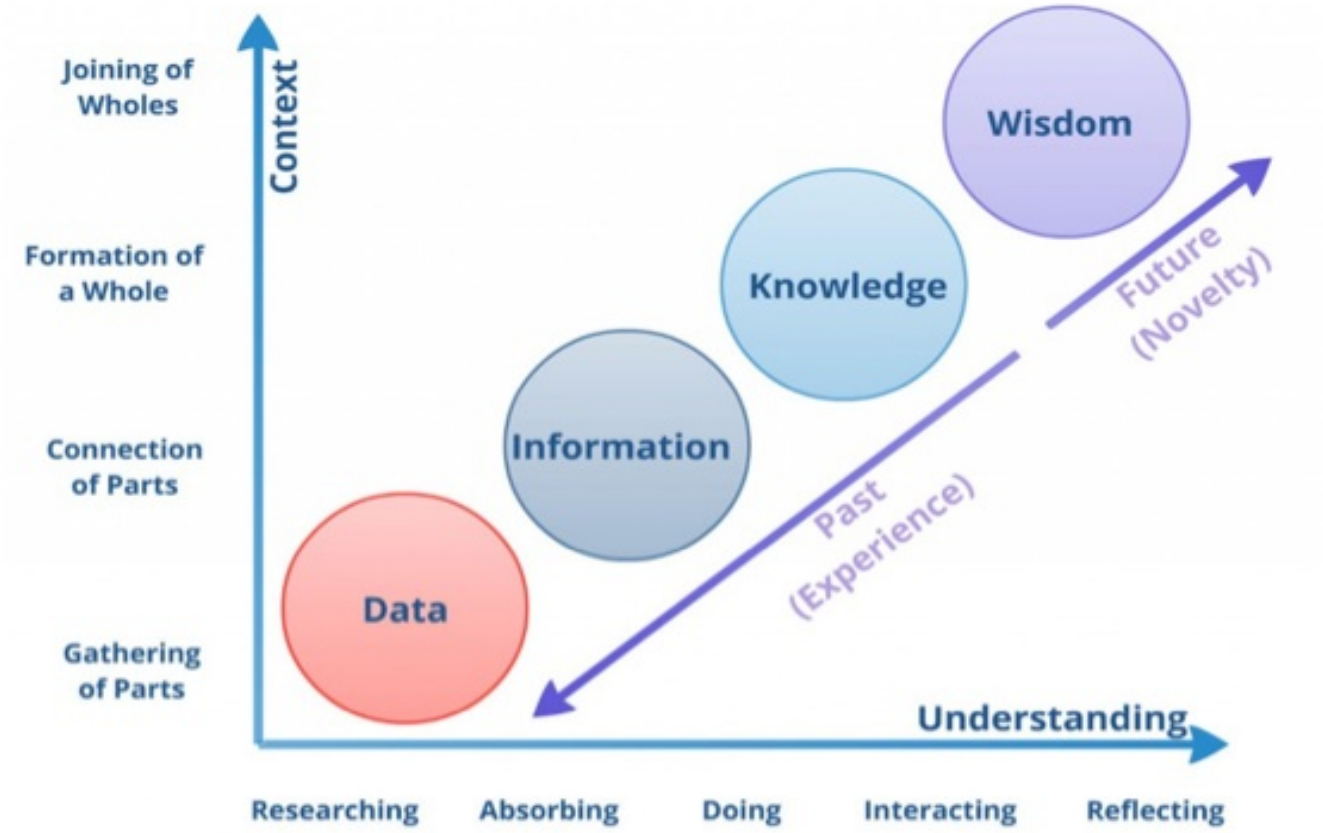
Dr. Ingrid Hsieh-Yee
Dept. of Library & Information Science
Catholic University of America
VLACRL Session
2016 Virginia Library Association
Hot Springs, Virginia



Session Objectives

- To clarify the relationship between data analysis and data visualization
- To highlight data visualization issues
- To introduce data visualization tools (readily available, free, fairly easy to use)

Value of Data Analysis

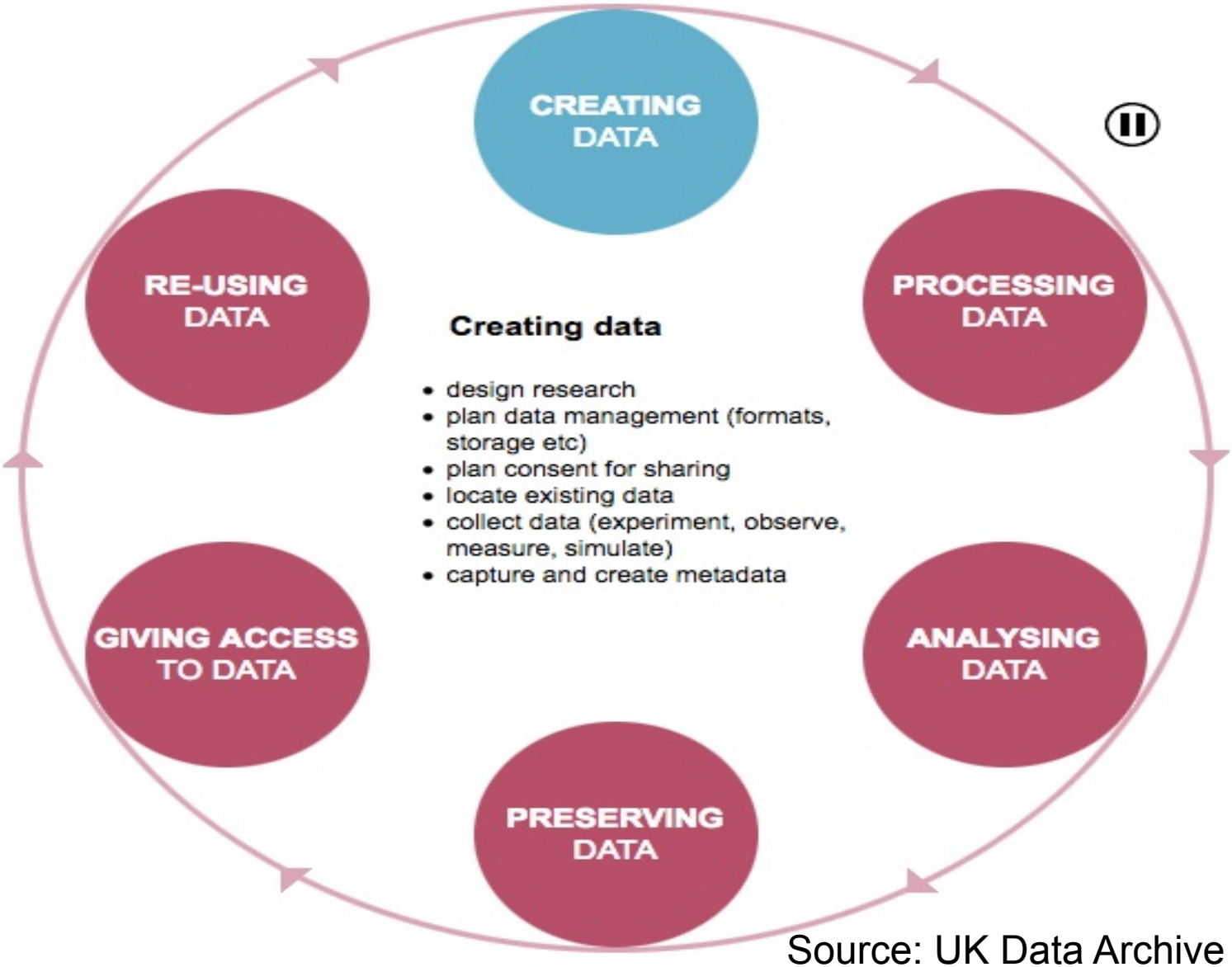


This diagram is adapted from a classic paper by Russ Ackoff "From Data to Wisdom", Journal of Applied Systems Analysis, Volume 16, 1989 p 3-9. Ackoff originally had Understanding as one of the circles, but subsequent authors have pointed out that it makes more sense as one of two dimensions you move along as you make sense of the data, the other being 'context' or 'connectedness'.

Visualization Begins with Data Analysis

- What are the questions?
- What types of data can address the questions?
- How will the data be collected?
- How will the data be processed before analysis?
- What types of analysis will be performed to answer the questions?

Research Data Lifecycle



Source: UK Data Archive

Sample Data Relationships

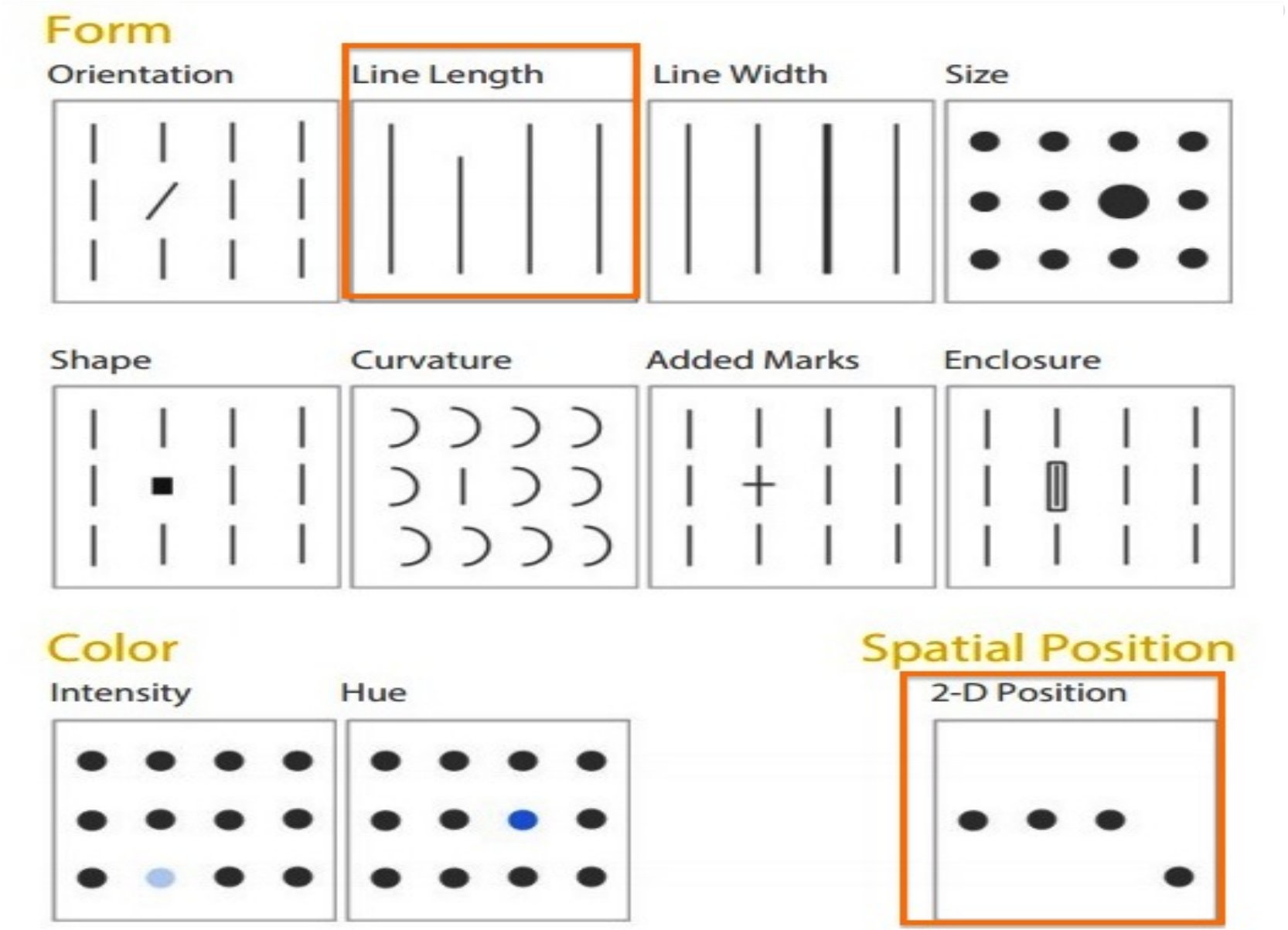
- Comparison (bar, line, stacked)

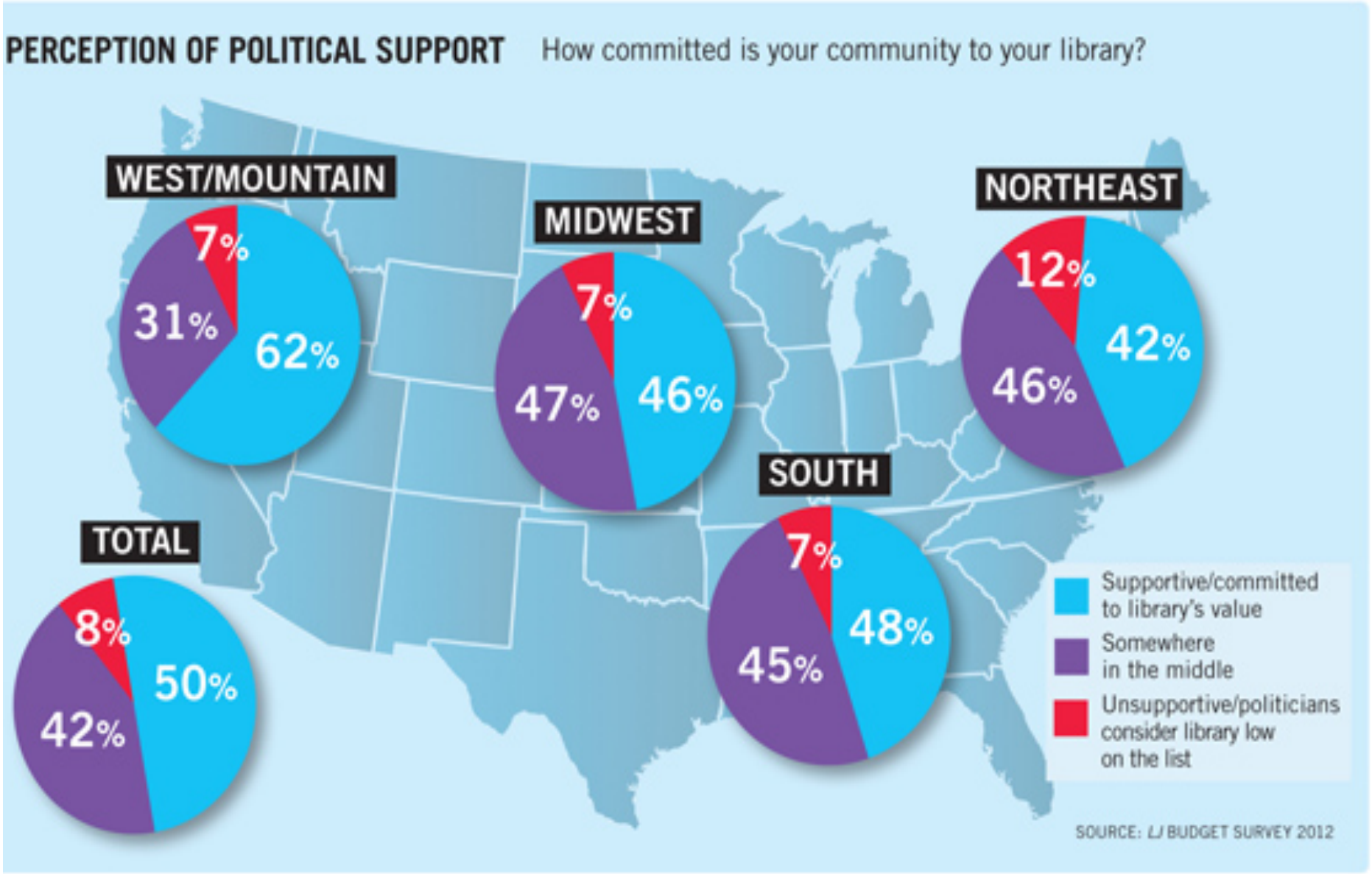
  E.g. Information use behavior of groups of users

  E.g.Library budget for print and nonprint resources
- Composition (pie, area)

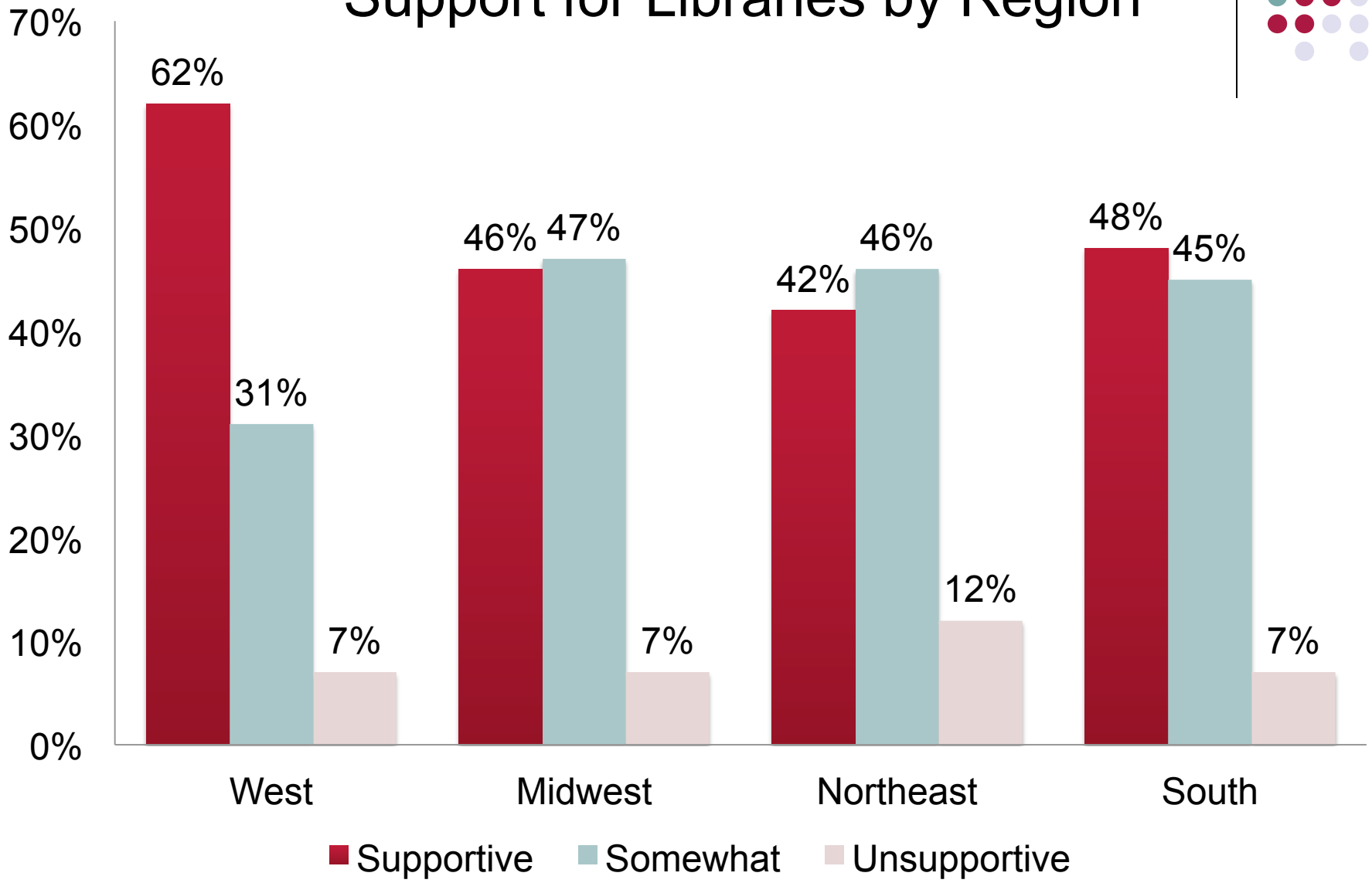
  E.g. Types of employees
- Distribution (pie, scatter, histogram) e.g. library spending
- Relationship (bubble)
- Process (organization, Gannt)

Visual Perception: Pre-attentive



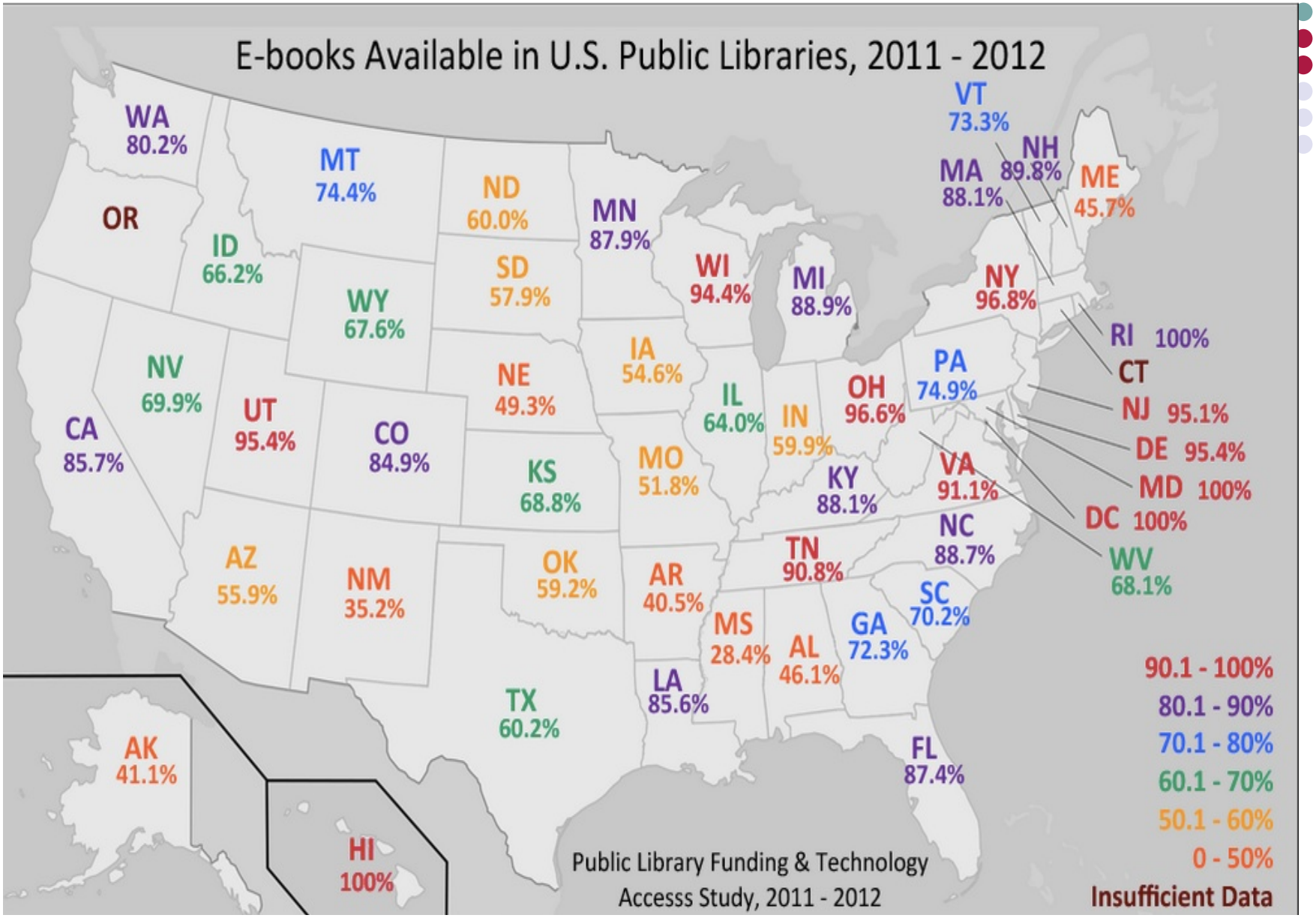

PERCEPTION OF POLITICAL SUPPORT
How committed is your community to your library?
WEST/MOUNTAIN
7%
31%
62%
MIDWEST
7%
47%
46%
NORTHEAST
12%
42%
46%
SOUTH
7%
48%
45%
TOTAL
8%
50%
42%
Supportive/committed to library's value
Somewhere in the middle
Unsupportive/politicians consider library low on the list
SOURCE: LJ BUDGET SURVEY 2012

Support for Libraries by Region



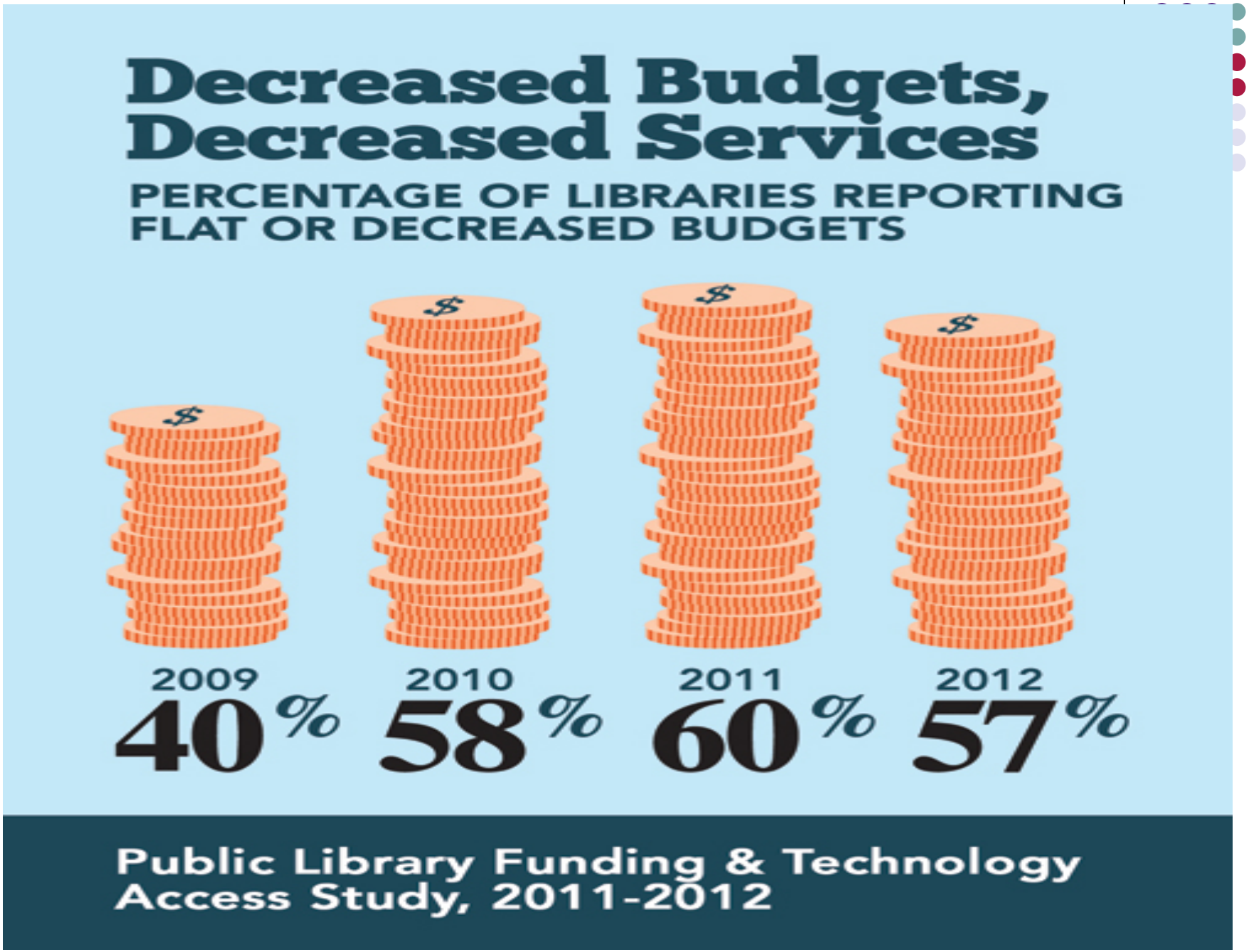
How These Charts Compare

- Clear labels
- Color contrast (hue, intensity)
- Line length
- Charts begin at the same level “0”
- Bars are proportional


E-books Available in U.S. Public Libraries, 2011 - 2012
WA 80.2%
OR
CA 85.7%
ID 66.2%
NV 69.9%
MT 74.4%
WY 67.6%
UT 95.4%
AZ 55.9%
CO 84.9%
NM 35.2%
ND 60.0%
SD 57.9%
NE 49.3%
KS 68.8%
OK 59.2%
TX 60.2%
MN 87.9%
IA 54.6%
MO 51.8%
AR 40.5%
LA 85.6%
WI 94.4%
IL 64.0%
MS 28.4%
MI 88.9%
IN 59.9%
KY 88.1%
TN 90.8%
AL 46.1%
OH 96.6%
GA 72.3%
FL 87.4%
SC 70.2%
NC 88.7%
VA 91.1%
WV 68.1%
PA 74.9%
NY 96.8%
VT 73.3%
NH 89.8%
MA 88.1%
ME 45.7%
RI 100%
CT
NJ 95.1%
DE 95.4%
MD 100%
DC 100%
AK 41.1%
HI 100%
90.1 - 100%
80.1 - 90%
70.1 - 80%
60.1 - 70%
50.1 - 60%
0 - 50%
Insufficient Data
Public Library Funding & Technology Accesss Study, 2011 - 2012

Strengths and Limitations of the E-Book Map

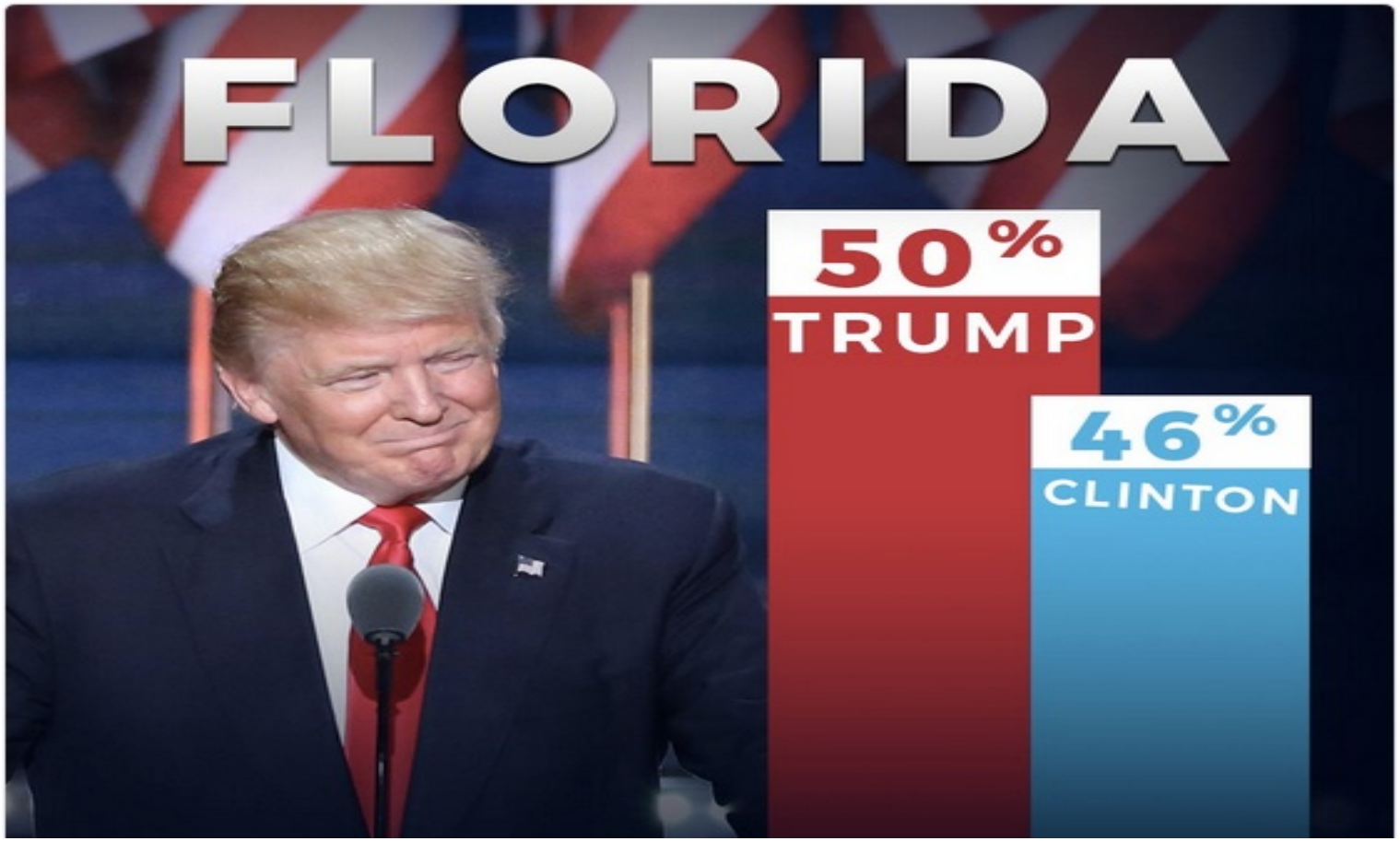
- What is your first impression? Why?
- The map conveys the message of distribution
- What is this map good for?


Decreased Budgets, Decreased Services
PERCENTAGE OF LIBRARIES REPORTING FLAT OR DECREASED BUDGETS
2009
40%
2010
58%
2011
60%
2012
57%
Public Library Funding & Technology Access Study, 2011-2012

Official Team Trump
@TeamTrump

Follow

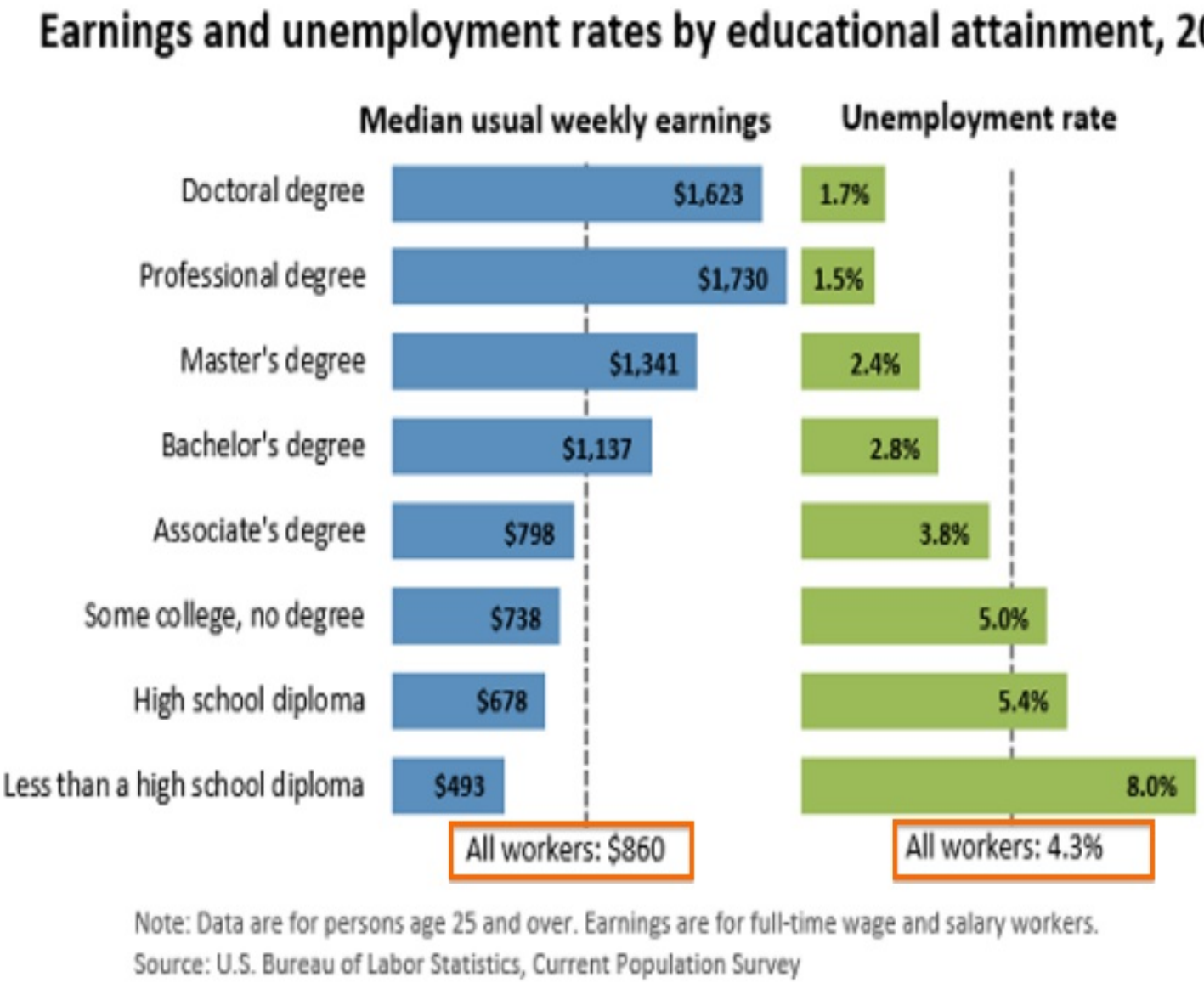
.@realDonaldTrump will be in FL tonight - where he is WINNING! Get tix to Miami rally here: donaldjtrump.com/schedule/regis …



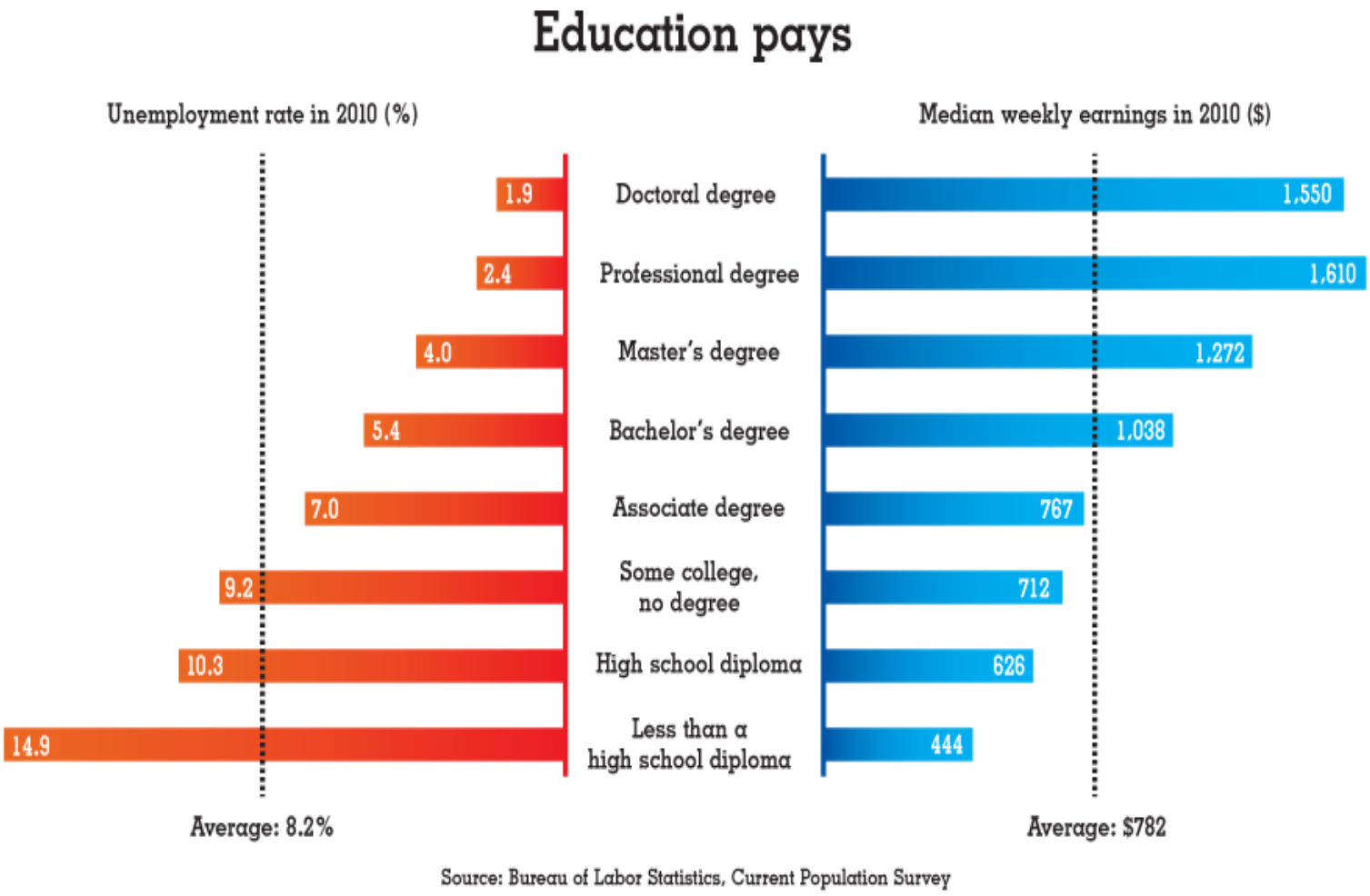
Earnings and unemployment rates by educational attainment, 2015

Education attained	Unemployment rate in 2015 (Percent)	Median weekly earnings in 2015
Doctoral degree	1.7	$1,623
Professional degree	1.5	1,730
Master's degree	2.4	1,341
Bachelor's degree	2.8	1,137
Associate's degree	3.8	798
Some college, no degree	5.0	738
High school diploma	5.4	678
Less than a high school diploma	8.0	493
All workers	4.3	860

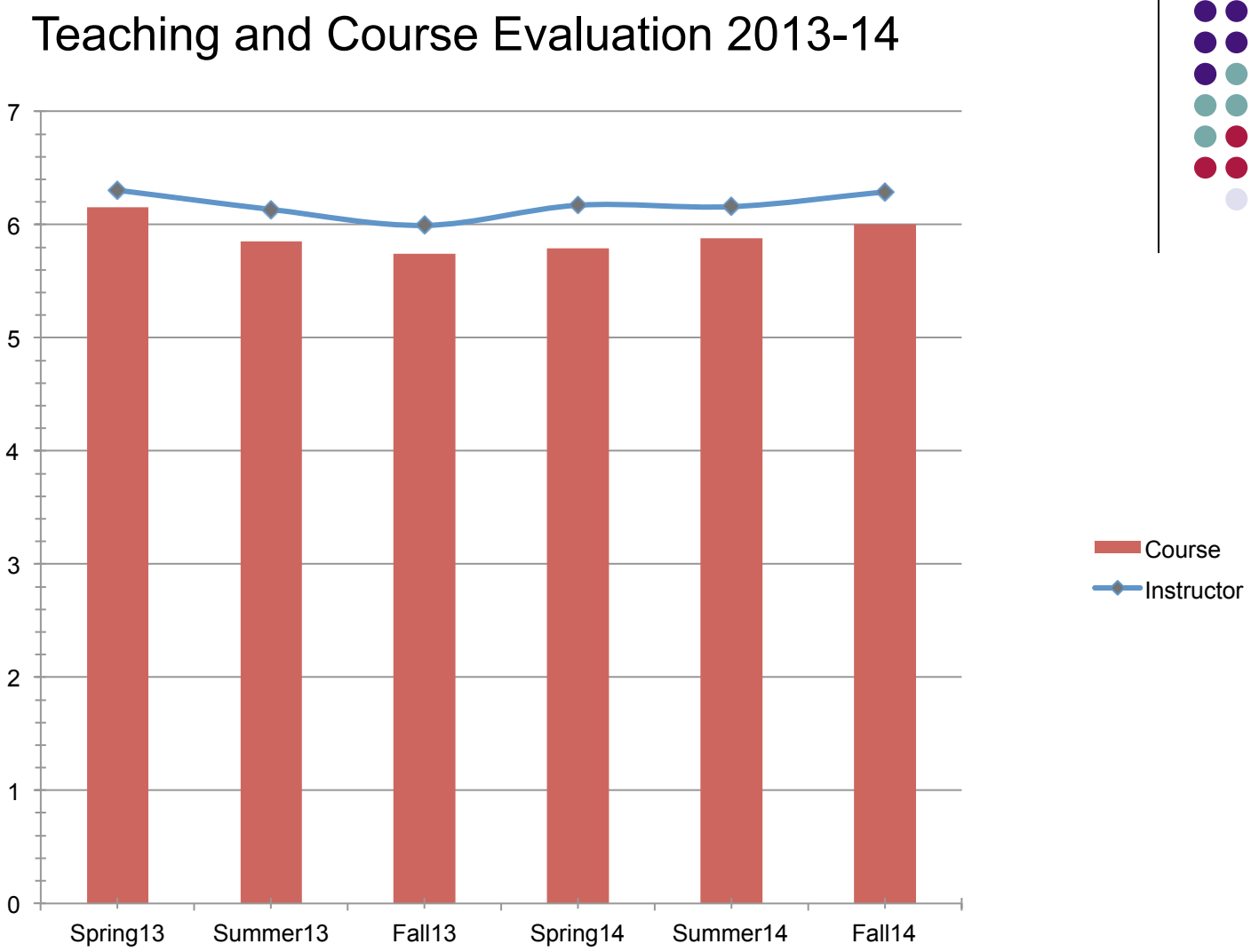
Note: Data are for persons age 25 and over. Earnings are for full-time wage and salary workers.
Source: Current Population Survey, U.S. Department of Labor, U.S. Bureau of Labor Statistics



Visualize data with reference points

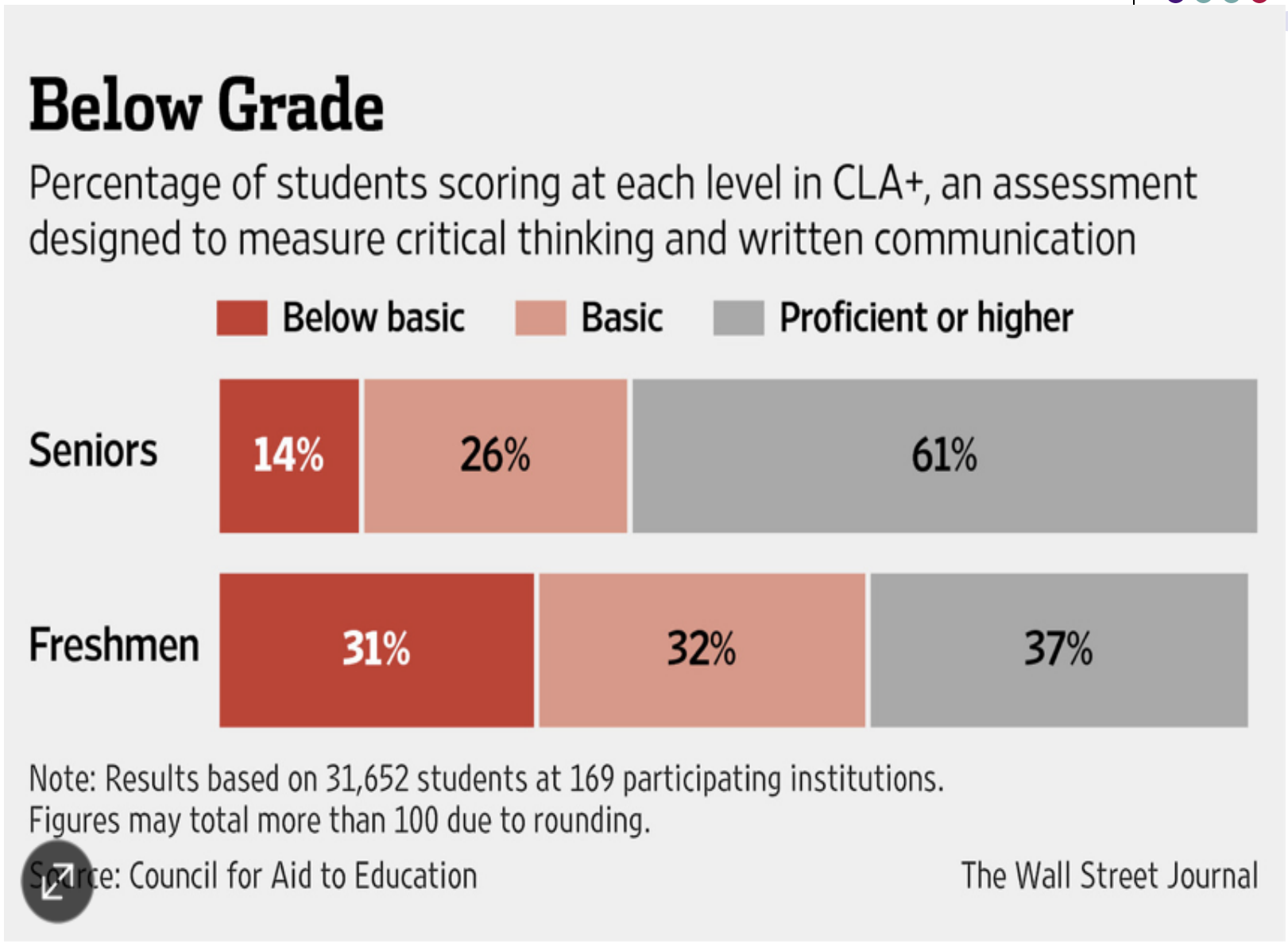


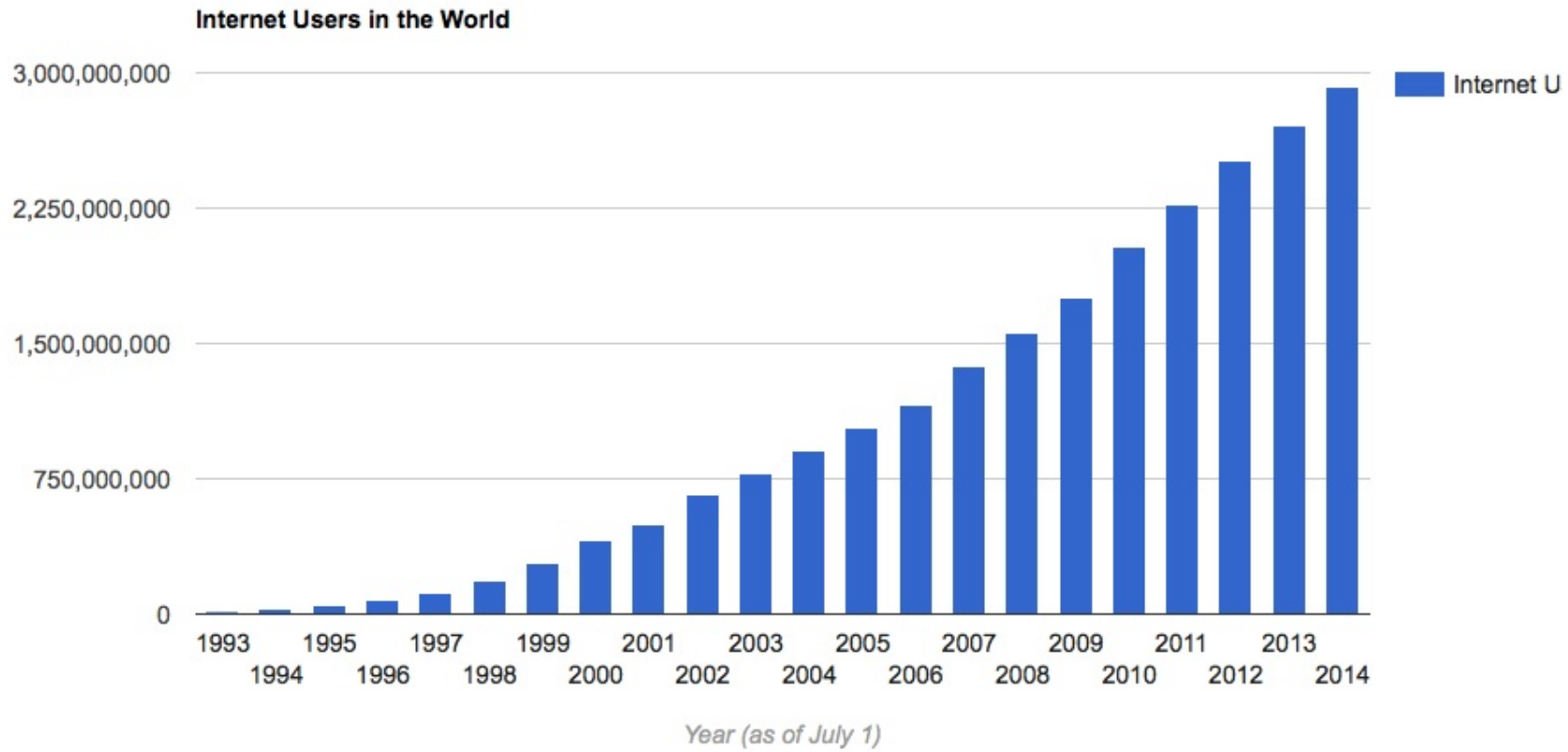
Is this chart better? Why?



More than one variable presented. Different styles combined to convey the message

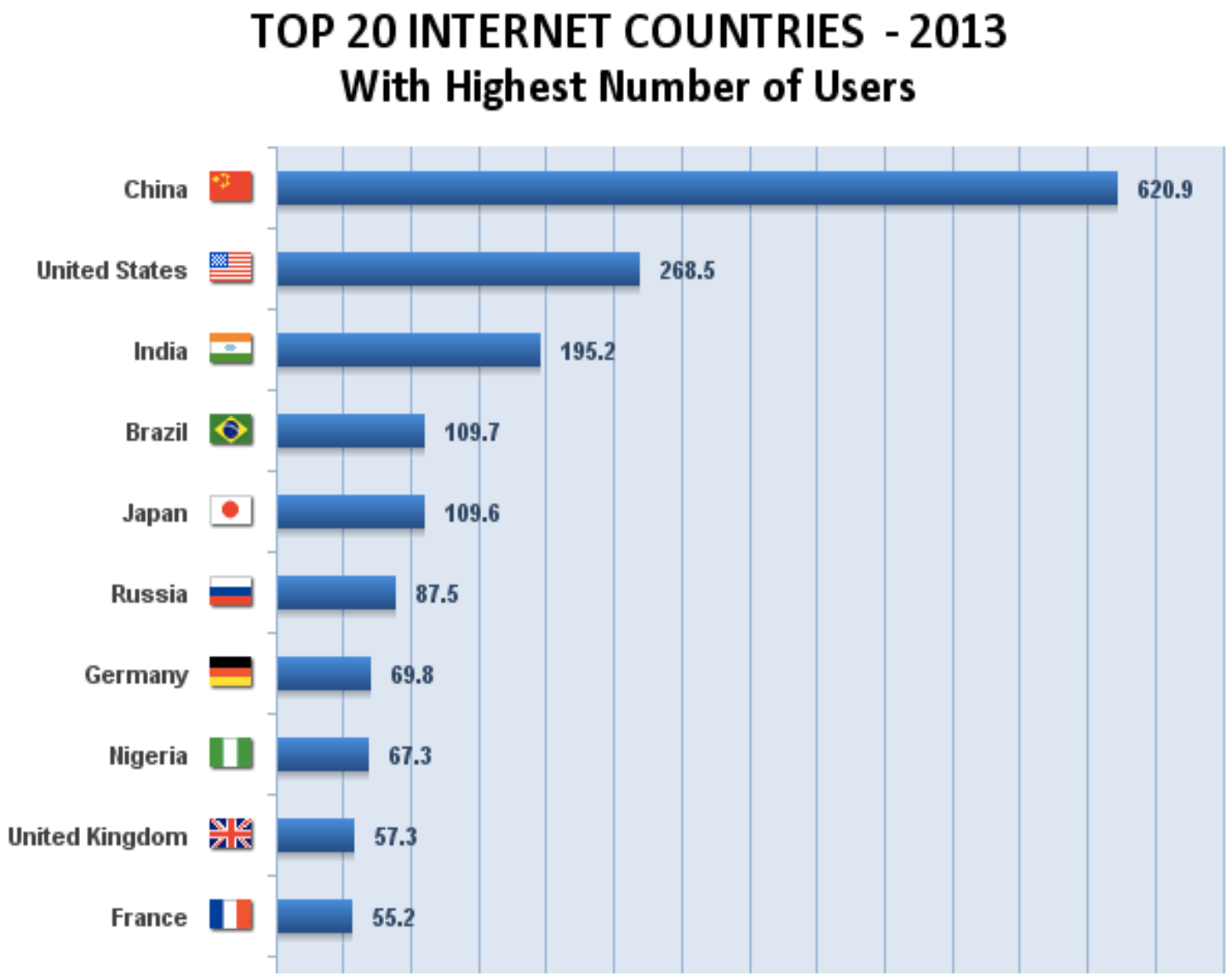
Group comparison by one variable



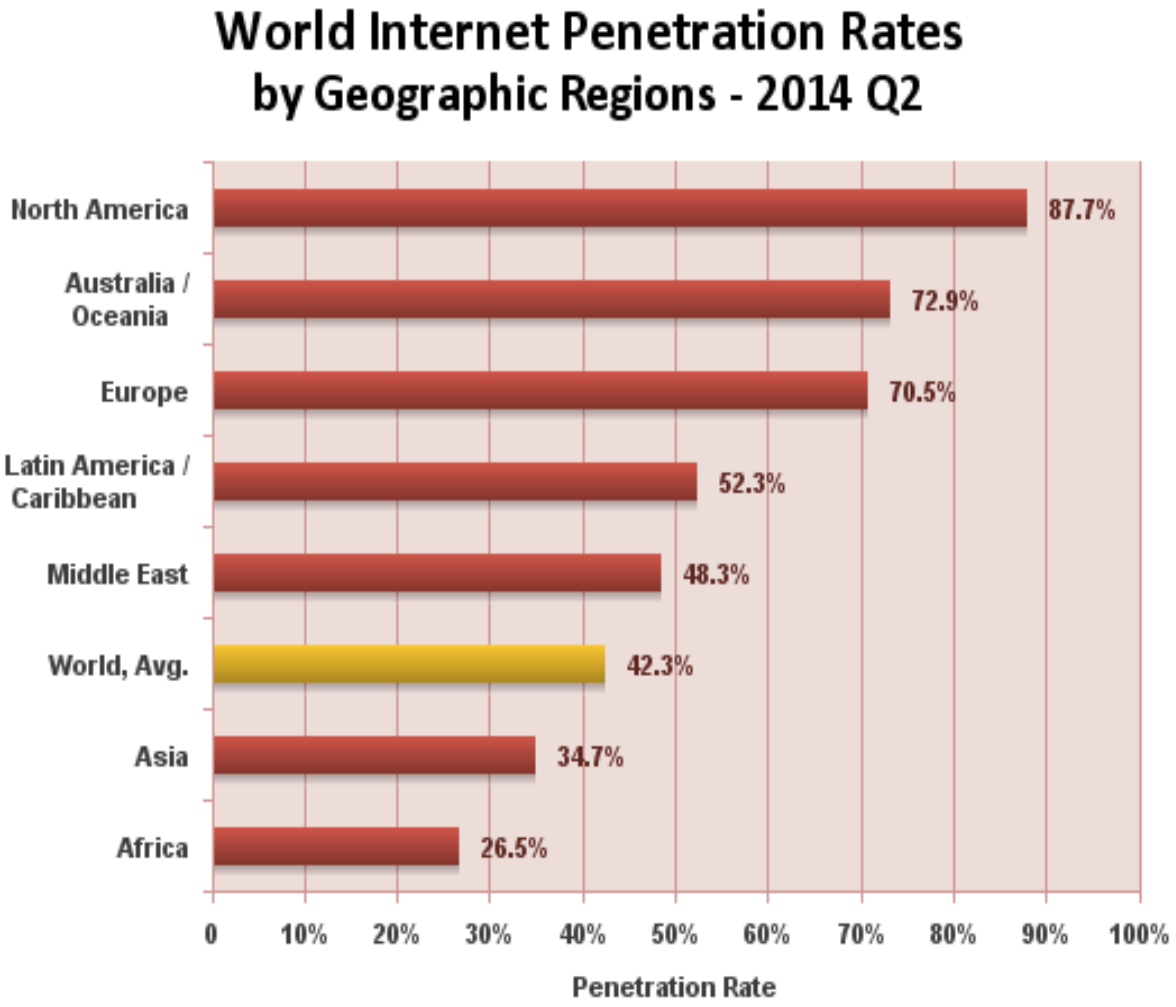


Year (July 1)	Internet Users	Users Growth	World Population	Population Growth	Penetration (% of Pop. with Internet)
2014*	**2,925,249,355**	7.9%	7,243,784,121	1.14%	**40.4%**
2013	**2,712,239,573**	8.0%	7,162,119,430	1.16%	**37.9%**
2012	**2,511,615,523**	10.5%	7,080,072,420	1.17%	**35.5%**
2011	**2,272,463,038**	11.7%	6,997,998,760	1.18%	**32.5%**
2010	**2,034,259,368**	16.1%	6,916,183,480	1.19%	**29.4%**
2009	**1,752,333,178**	12.2%	6,834,721,930	1.20%	**25.6%**

Comparison across many groups

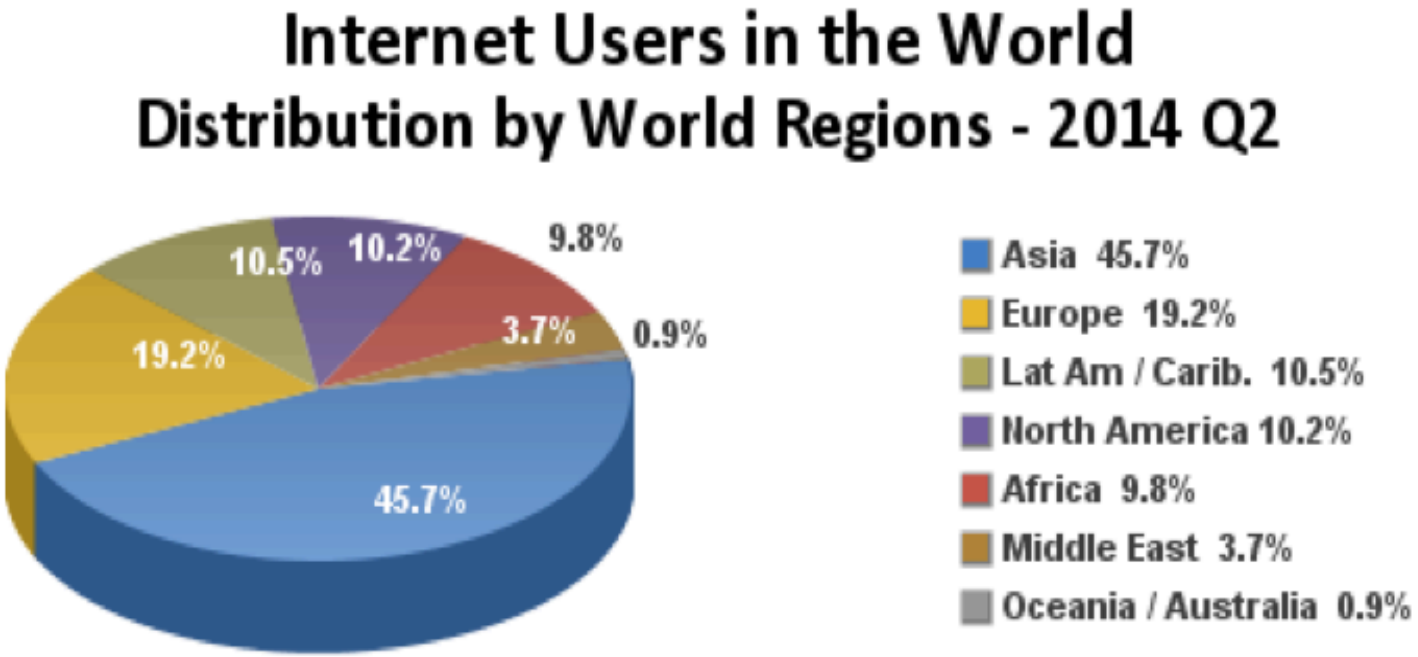


Different perspectives reveal different pictures



Source: Internet World Stats - www.internetworldststs.com/stats.htm
Penetration Rates are based on a world population of 7,182,406,565 and 3,035,749,340 estimated Internet users on June 30, 2014.

Use pie charts for distribution



Source: Internet World Stats - www.internetworldstats.com/stats.htm
Basis: 3,035,749,340 Internet users on June 30, 2014

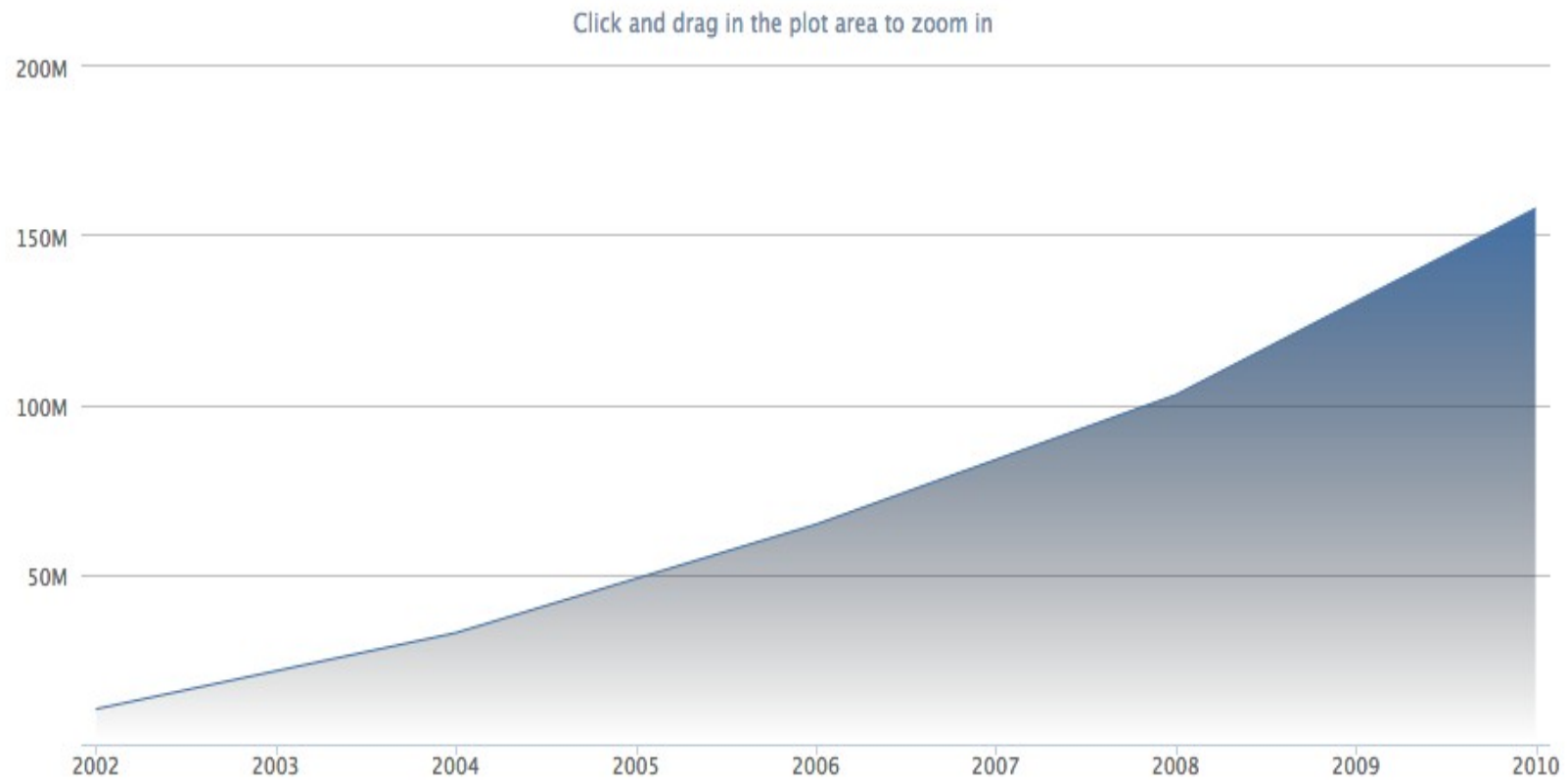
Distribution and Changes over time

DATA-PLANET™
A UNIVERSE OF DATA

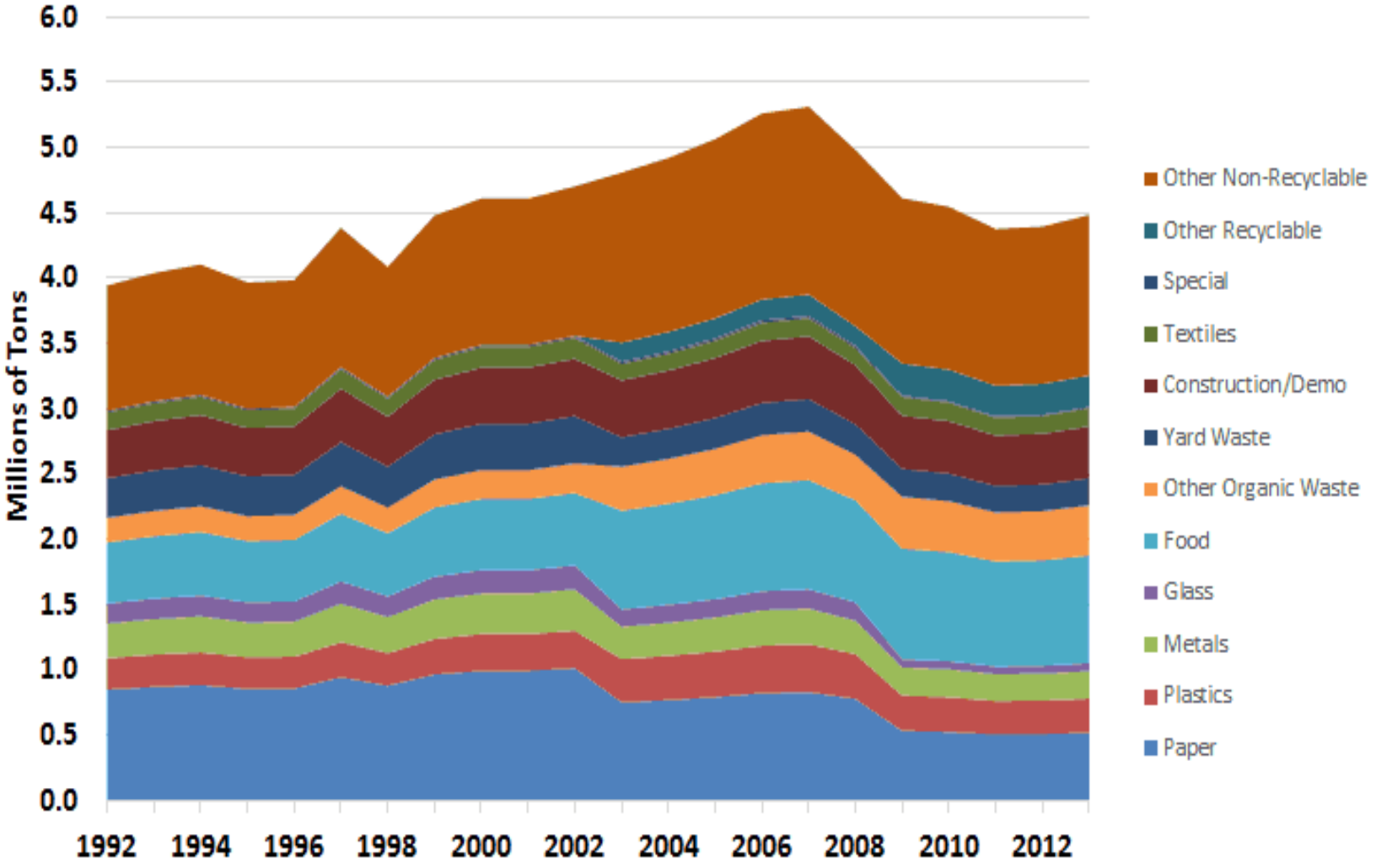
Areas and shades indicate growth

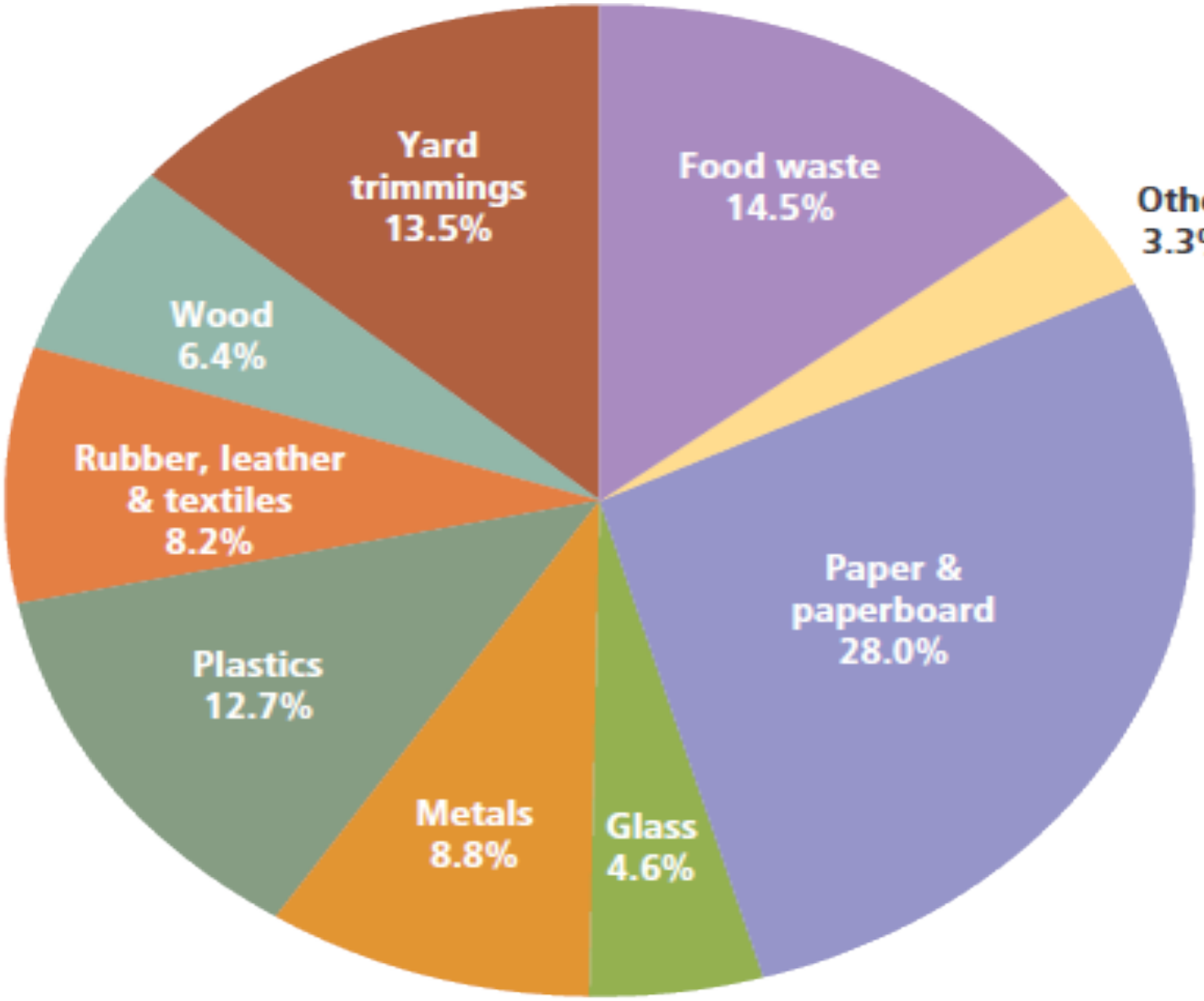
E-Books- Held at End of FY from the **Academic Library Statistics: United States** Dataset shown as **count**
Country: **USA**

Trend 2002-2010 | Ranking of States (2010) | Map of States (2010) | Ranking of Institutions (2010) | Map of Institutions (2010) | Most Recent Data (2010)



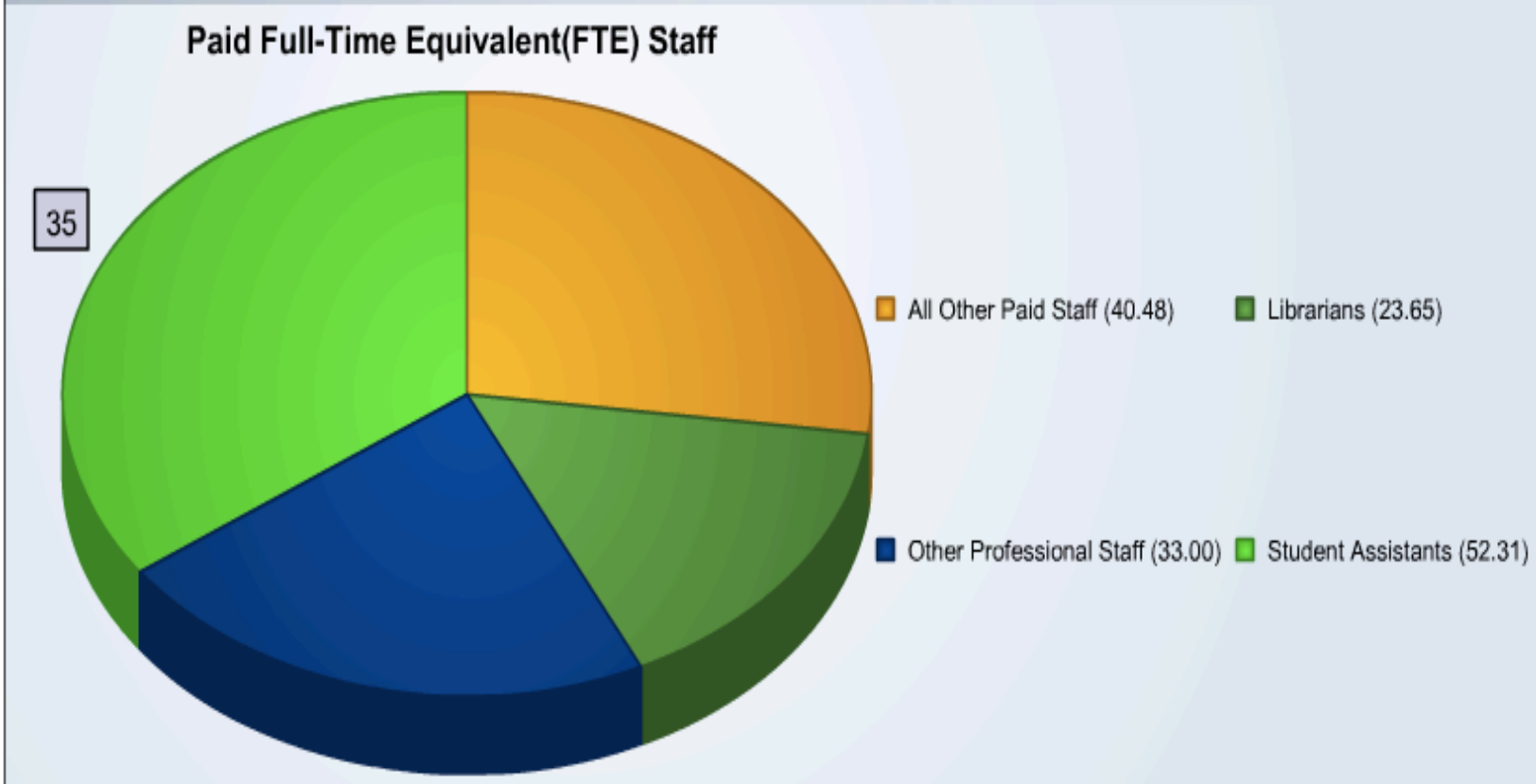
Disposed Material in MSW in Washington: 1992-2013





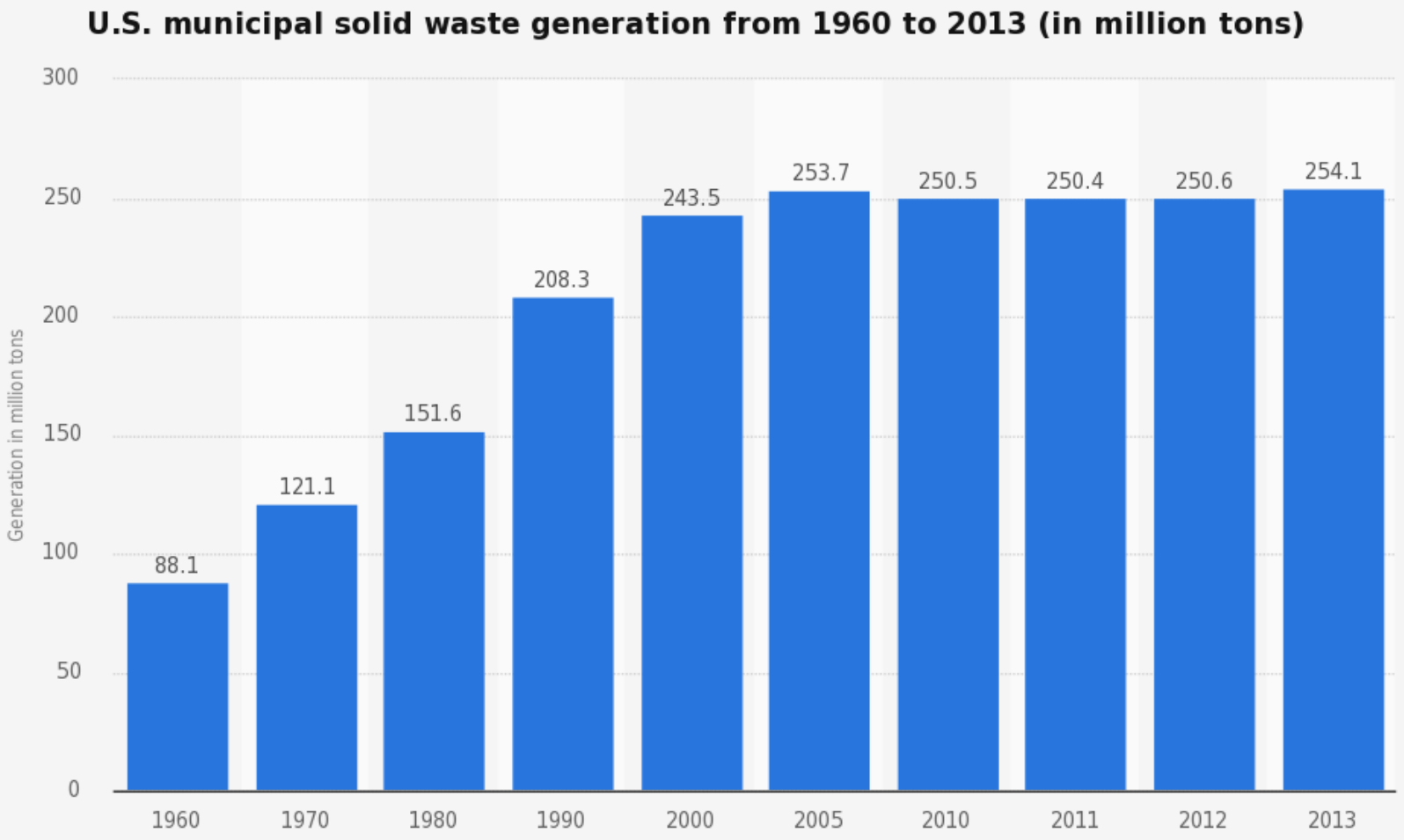
Municipal Solid Waste by Type

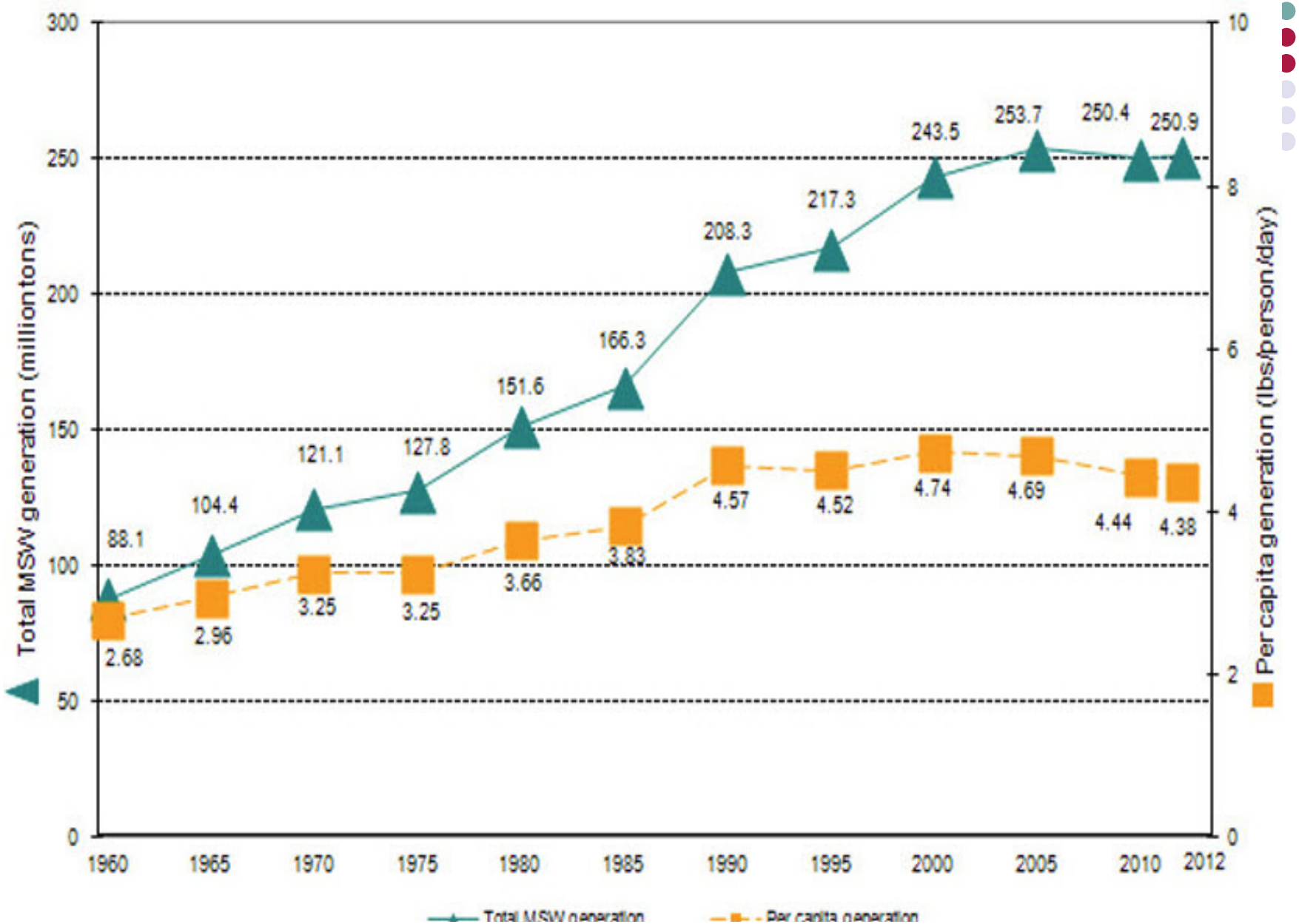
Percentages of Paid Full-Time Equivalent(FTE) Staff, American University, DC: 2012

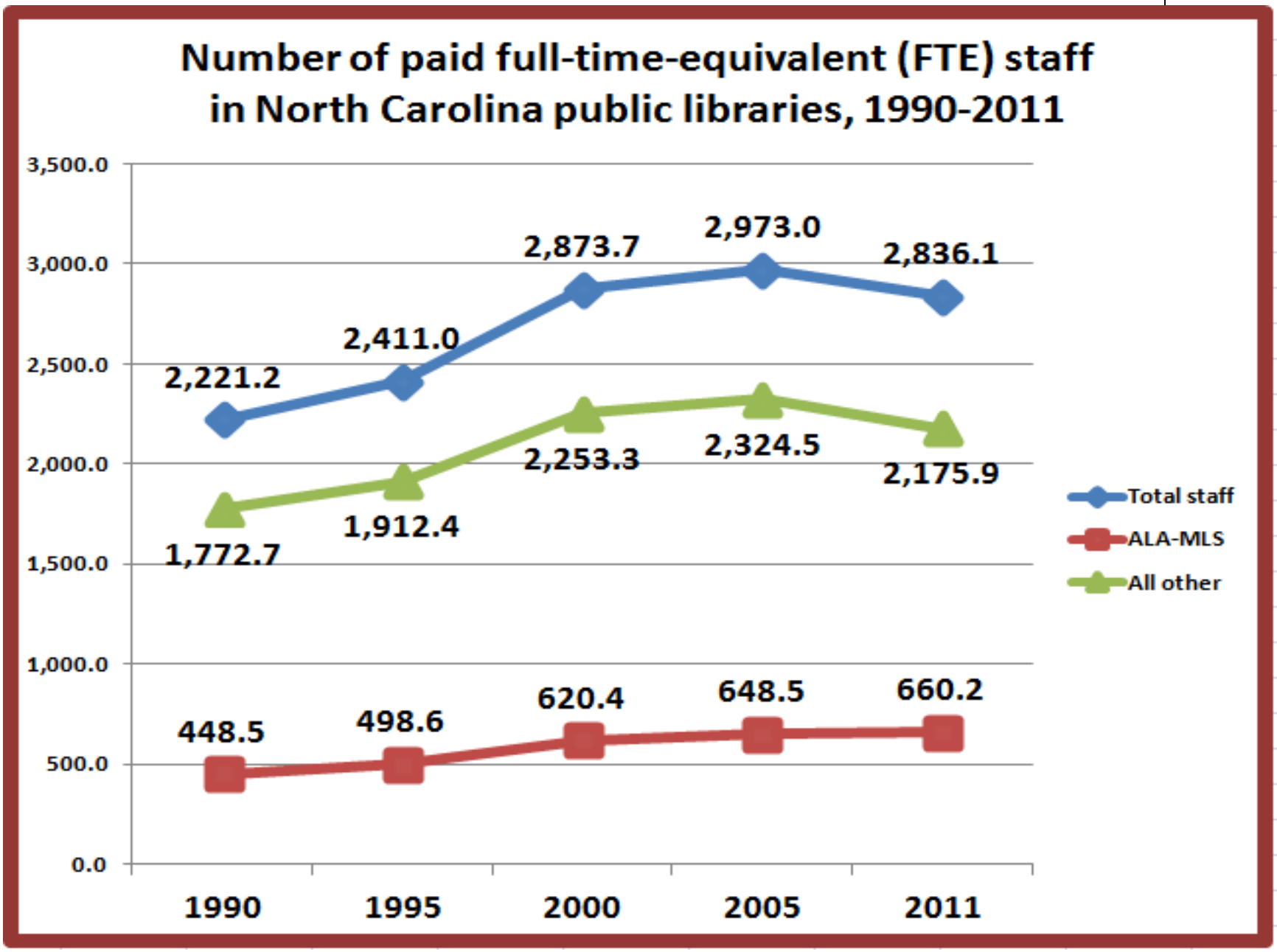


SOURCE: U.S. Department of Education, National Center of Statistics, Academic Libraries Survey (ALS), fiscal years 2012,

Increase over time

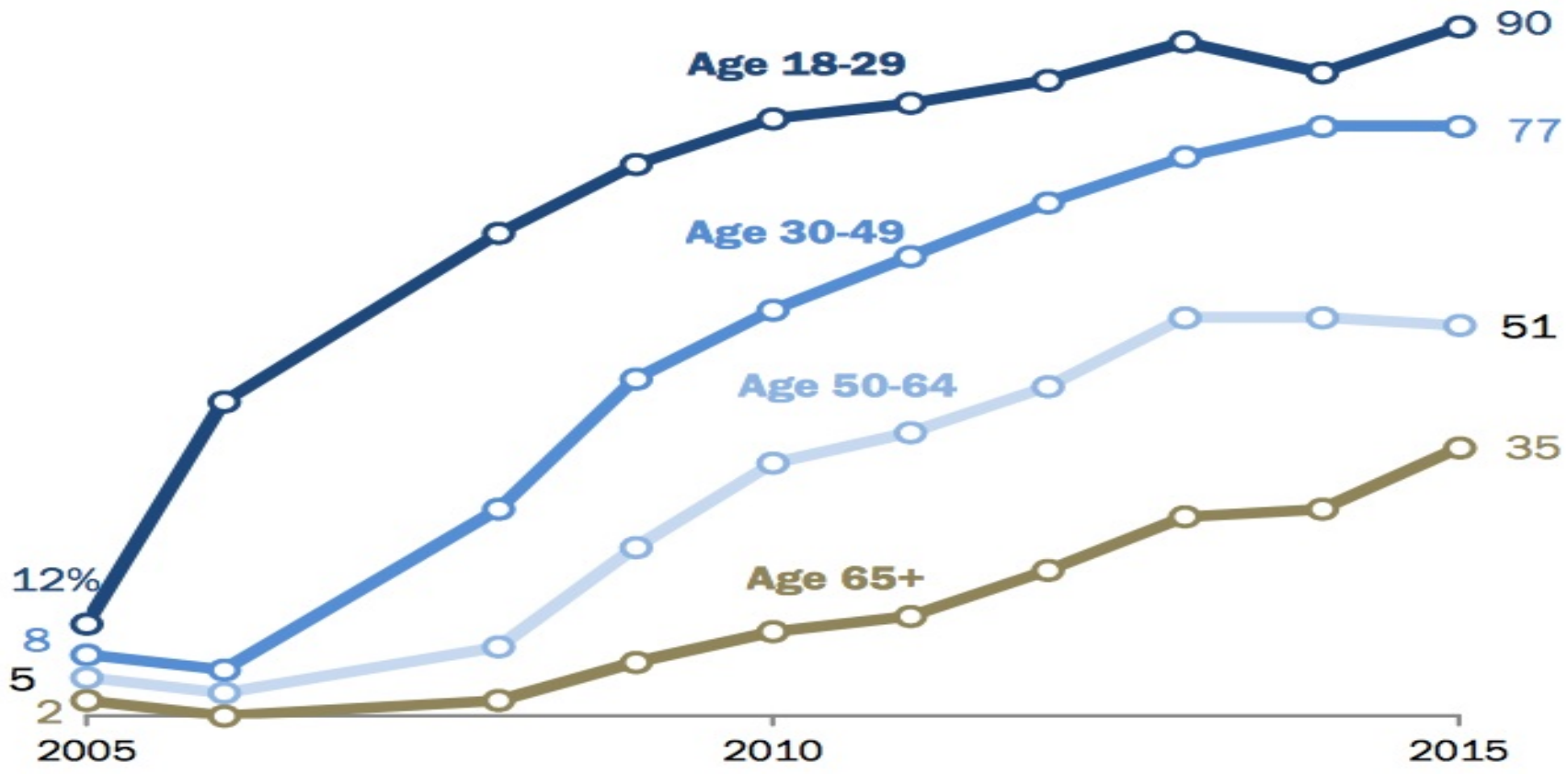



Total MSW generation (million tons)
Per capita generation (lbs/person/day)
300
250
200
150
100
50
0
10
8
6
4
2
0
88.1
104.4
121.1
127.8
151.6
166.3
208.3
217.3
243.5
253.7
250.4
250.9
2.68
2.96
3.25
3.25
3.66
3.83
4.57
4.52
4.74
4.69
4.44
4.38
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2012
Total MSW generation
Per capita generation


Number of paid full-time-equivalent (FTE) staff in North Carolina public libraries, 1990-2011
3,500.0
3,000.0
2,500.0
2,000.0
1,500.0
1,000.0
500.0
0.0
2,221.2
2,411.0
2,873.7
2,973.0
2,836.1
1,772.7
1,912.4
2,253.3
2,324.5
2,175.9
448.5
498.6
620.4
648.5
660.2
1990
1995
2000
2005
2011
Total staff
ALA-MLS
All other

Young Adults Still Are the Most Likely to Use Social Media

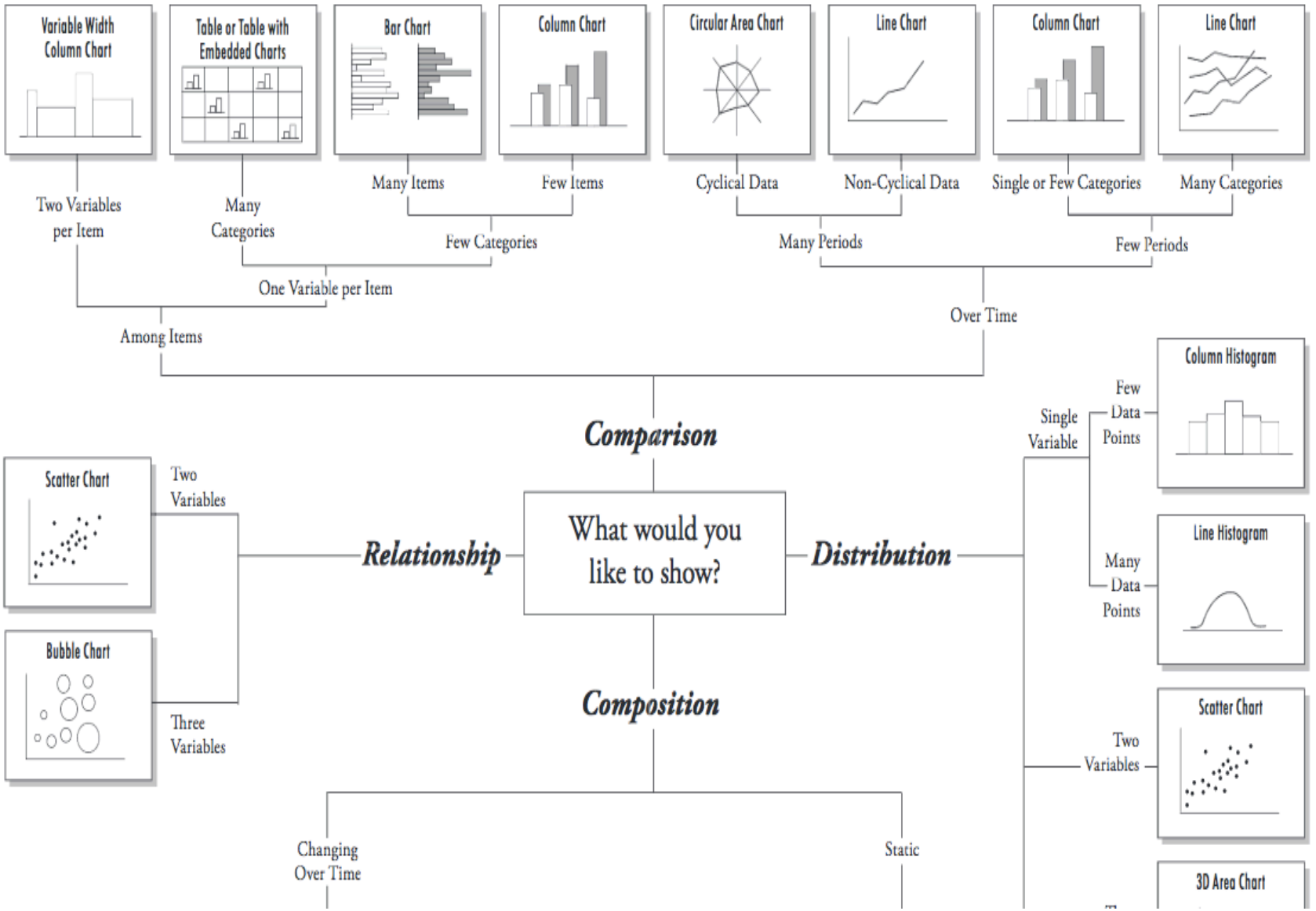
Among all American adults, % who use social networking sites, by age



Source: Pew Research Center surveys, 2005-2006, 2008-2015. No data are available for 2007

Free Visualization Resources

Chart Suggestions—A Thought-Starter



Click these tabs to get help, view examples, or use templates

Help | Examples

Graphs and charts are great because they communicate information visually. For this reason, graphs are often used in newspapers, magazines and businesses around the world.

NCES constantly uses graphs and charts in our publications and on the web. Sometimes, complicated information is difficult to understand and needs an illustration. Graphs or charts can help impress people by getting your point across quickly and visually.

Here you will find five different graphs and charts for you to consider. Not sure about which graph to use? Confused between bar graphs and pie charts? Read our:

Create A Graph Tutorial

Design | Data | Labels | Preview | Print / Save

Bar

Line | **Area** | **Pie** | **XY**

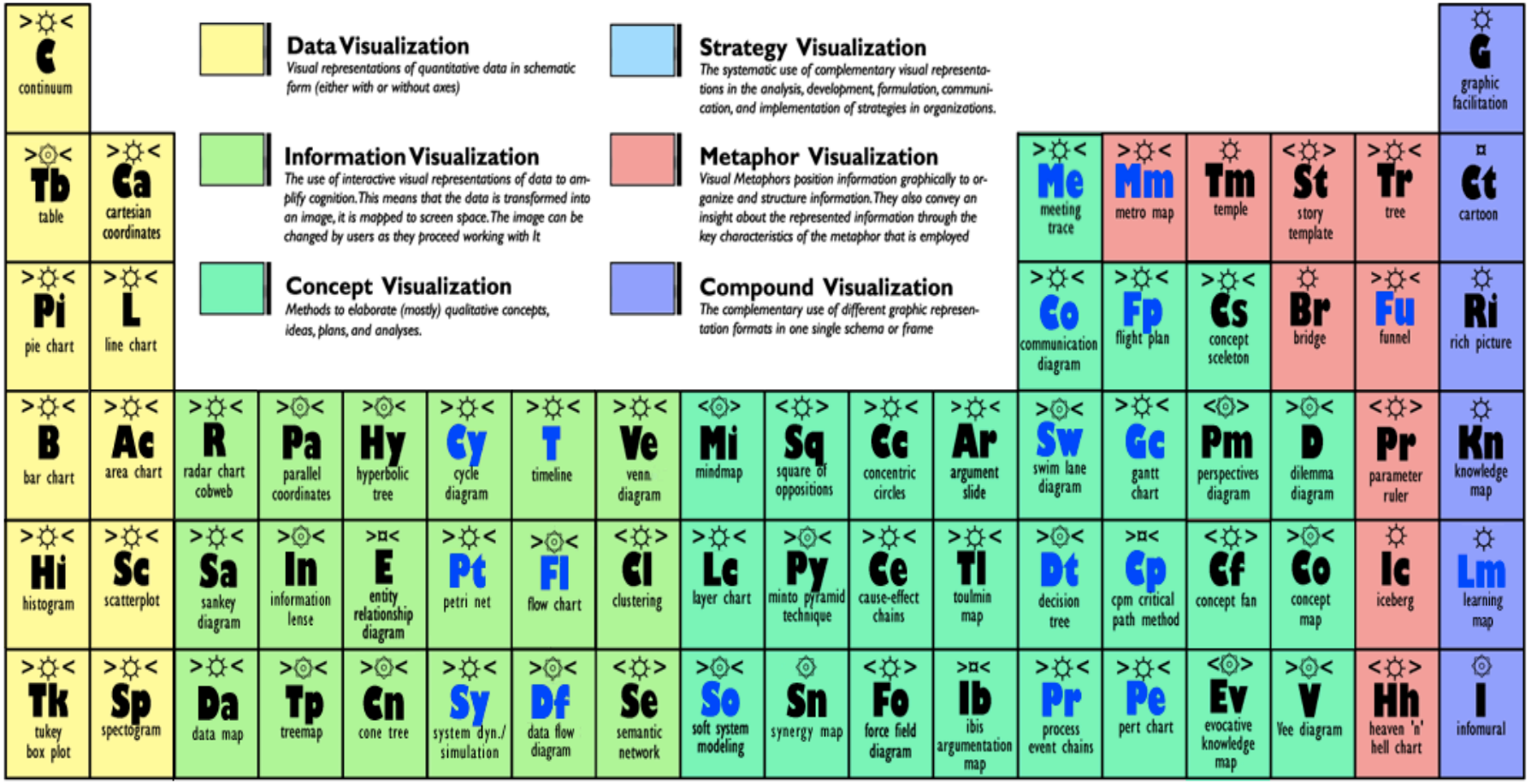
Direction: Vertical | Horizontal | True 3D | Stacked | Stacked

Shape: Rectangle | Cylinder | Triangle | Rhombus | Pattern

Style: Background Color: Grid Color: Grid Lines: 6

Appearance: 2D Legend: position right

A PERIODIC TABLE OF VISUALIZATION METHODS

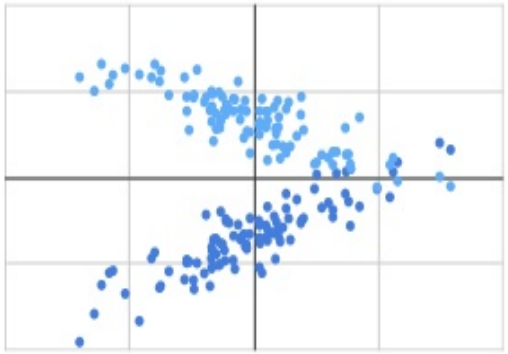


Note: Depending on your location and connection speed it can take some time to load a pop-up picture.

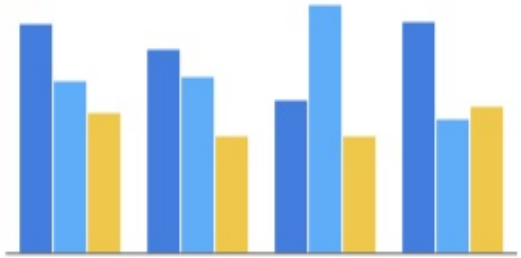
version 1.5

Geo Chart

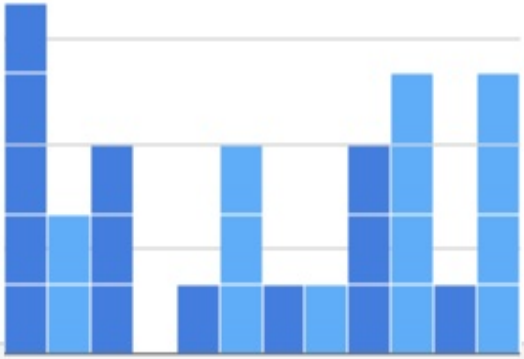
Scatter Chart



Column Chart

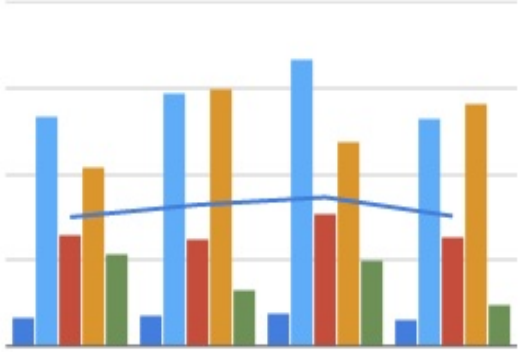


Histogram



Bar Chart

Combo Chart

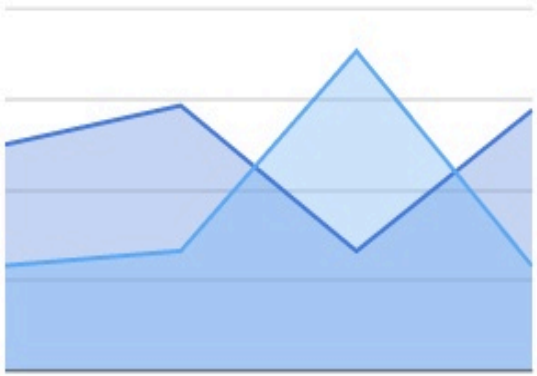


Area Chart

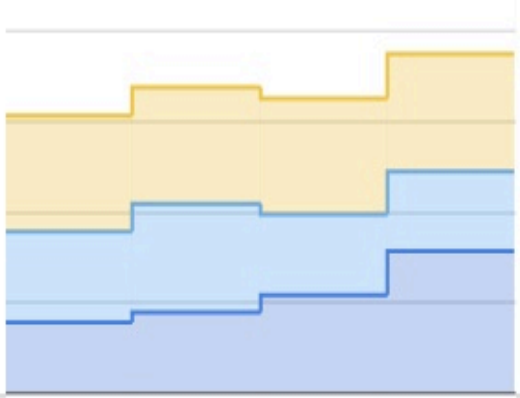
Stepped Area Chart

Line Chart

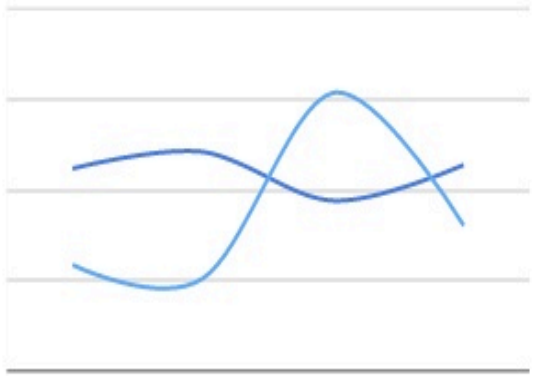
Area Chart



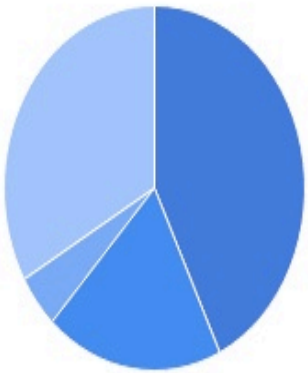
Stepped Area Chart



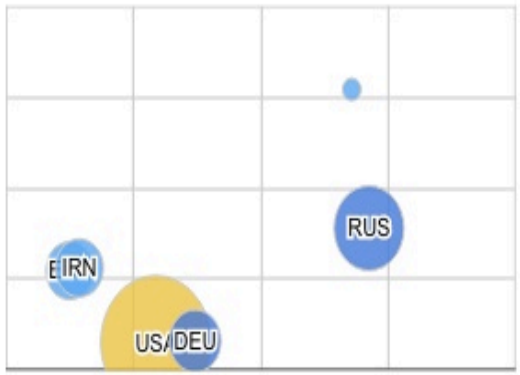
Line Chart



Pie Chart



Bubble Chart



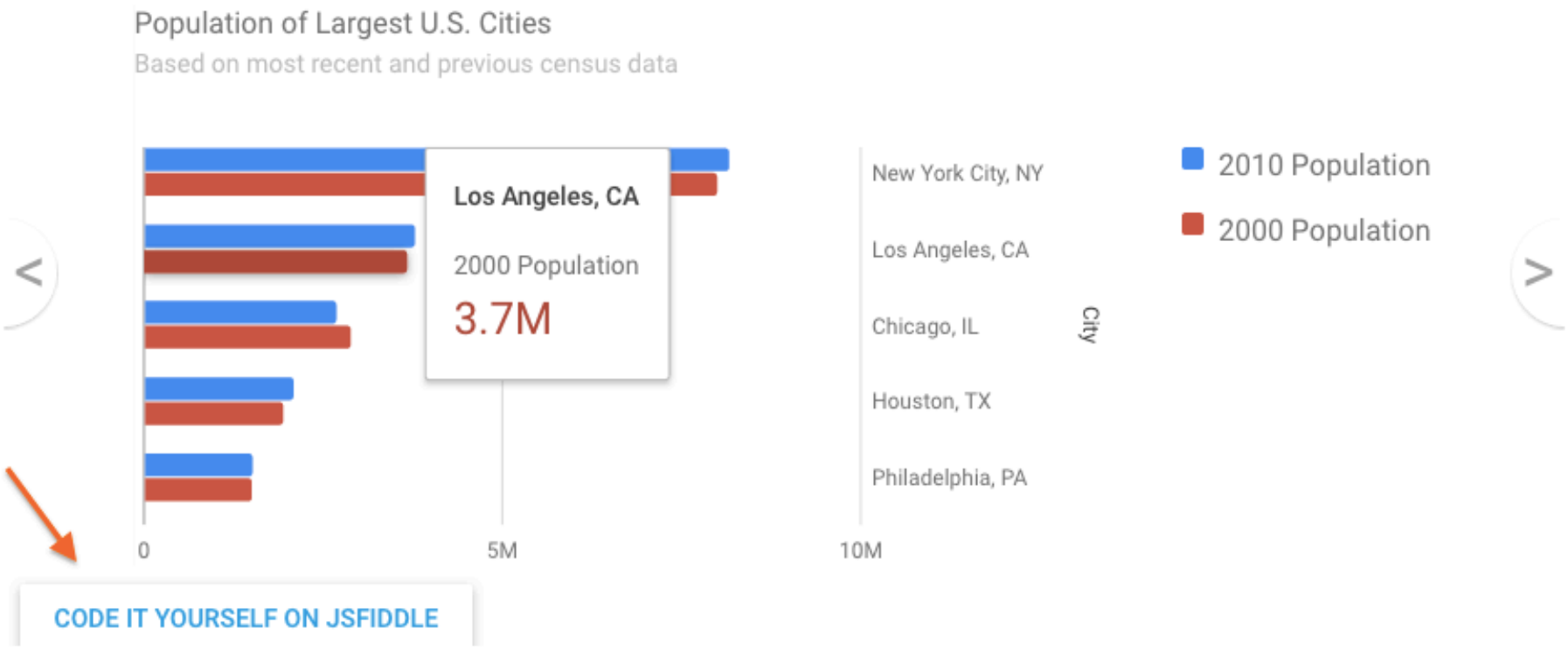
Donut Chart

Org Chart

Treemap

Table

Right Y-axis


Population of Largest U.S. Cities
Based on most recent and previous census data
Los Angeles, CA
2000 Population
3.7M
New York City, NY
Los Angeles, CA
Chicago, IL
Houston, TX
Philadelphia, PA
City
2010 Population
2000 Population
0
5M
10M
CODE IT YOURSELF ON JSFIDDLE

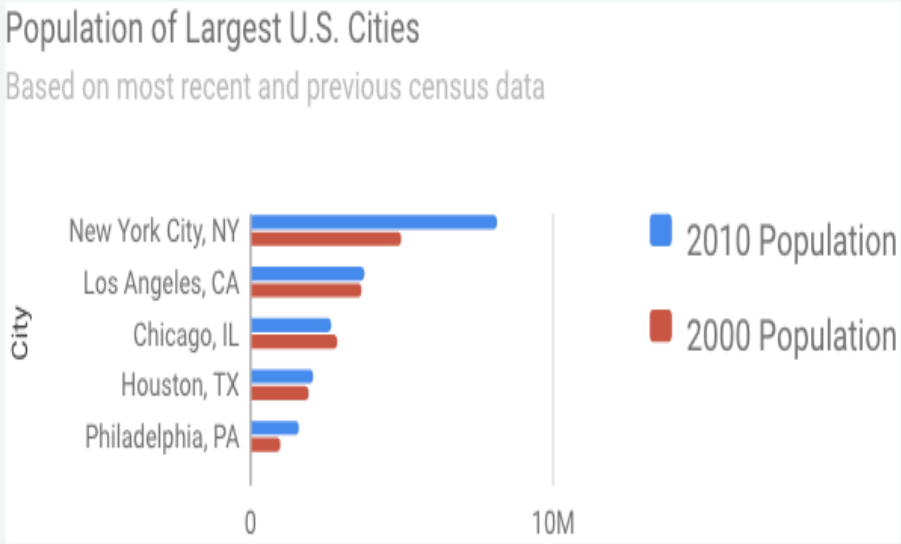
```html
<script type="text/javascript" src="https://www.gstatic.com
/charts/loader.js"></script>
<div id="chart_div"></div>
```

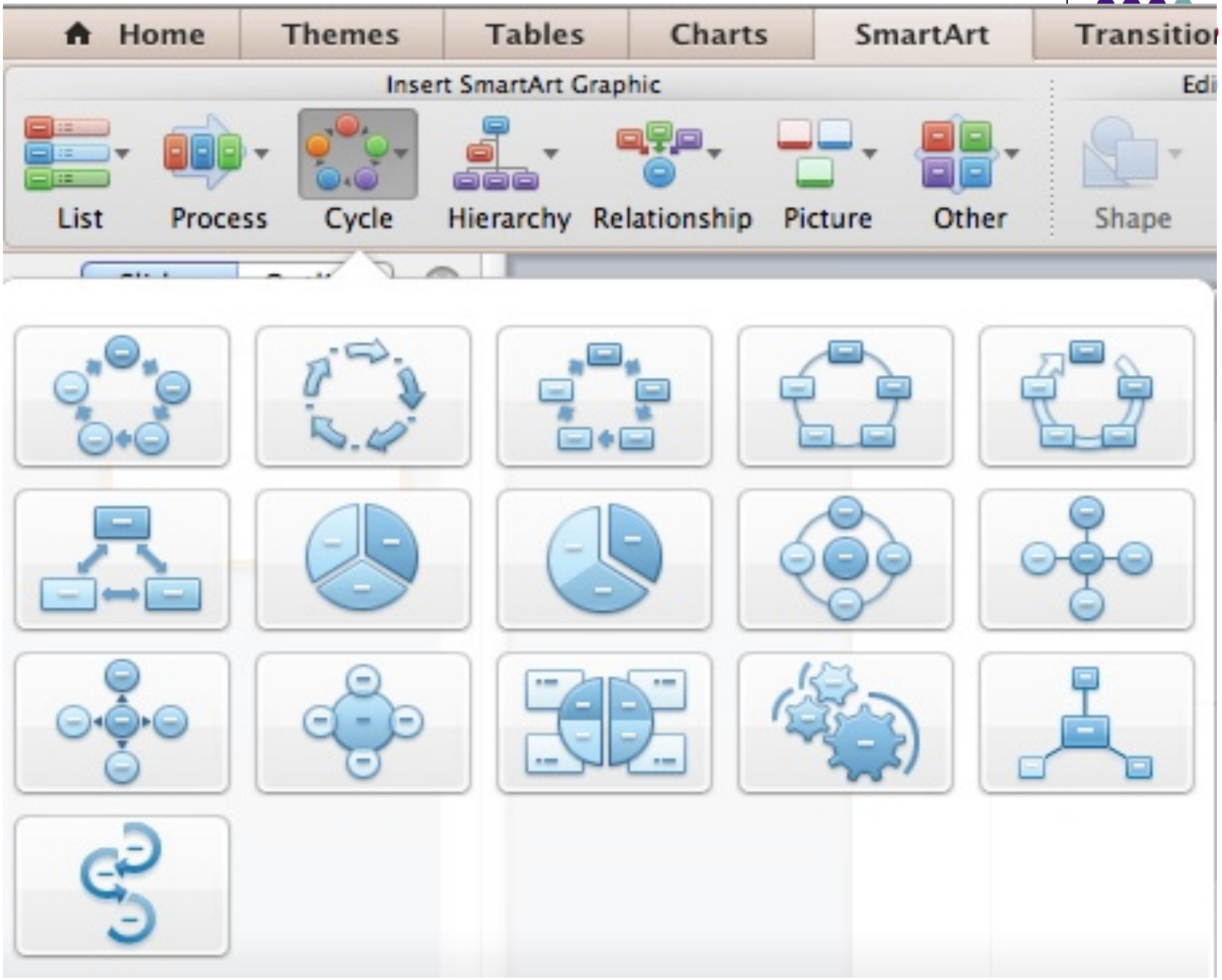
HTML

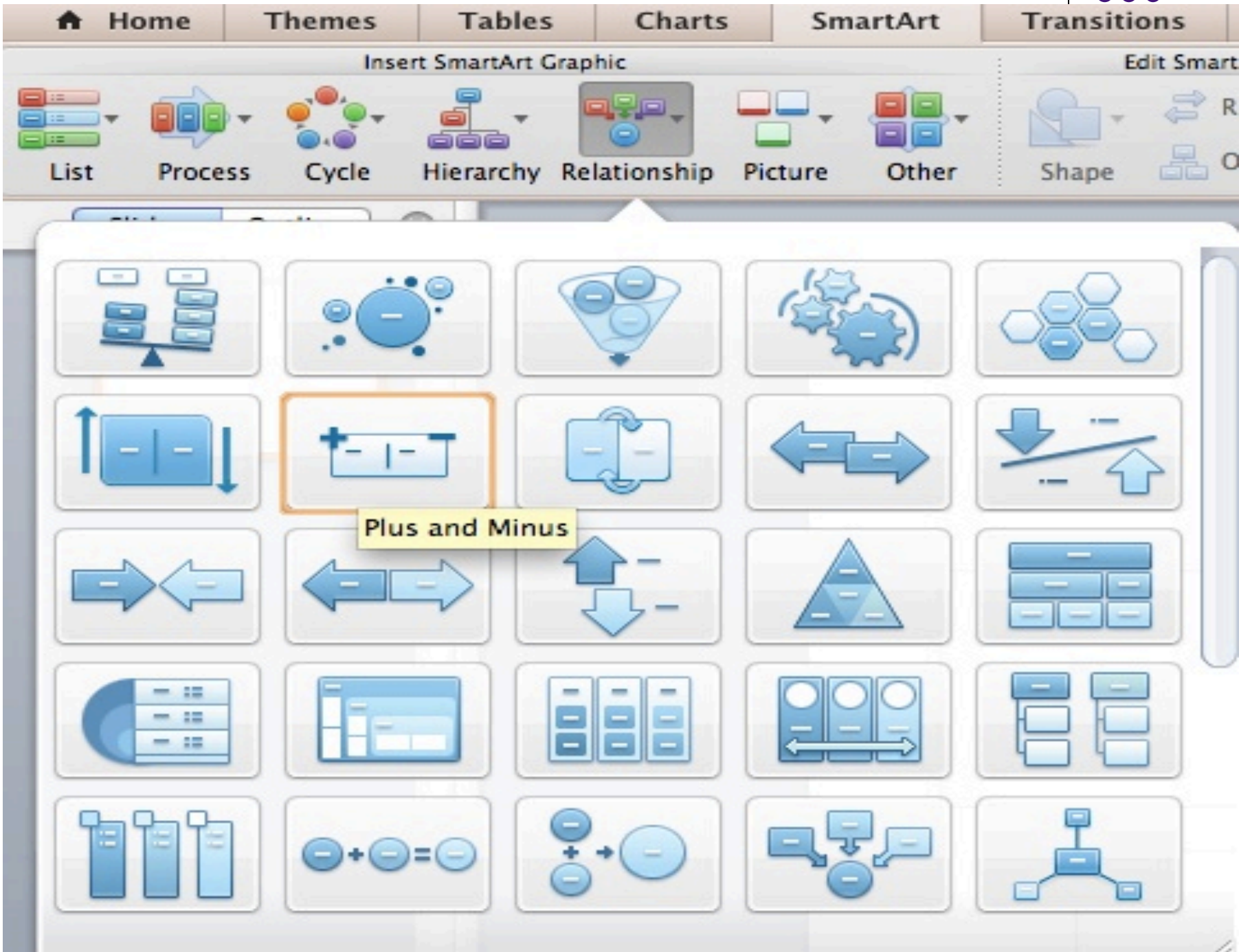
CSS

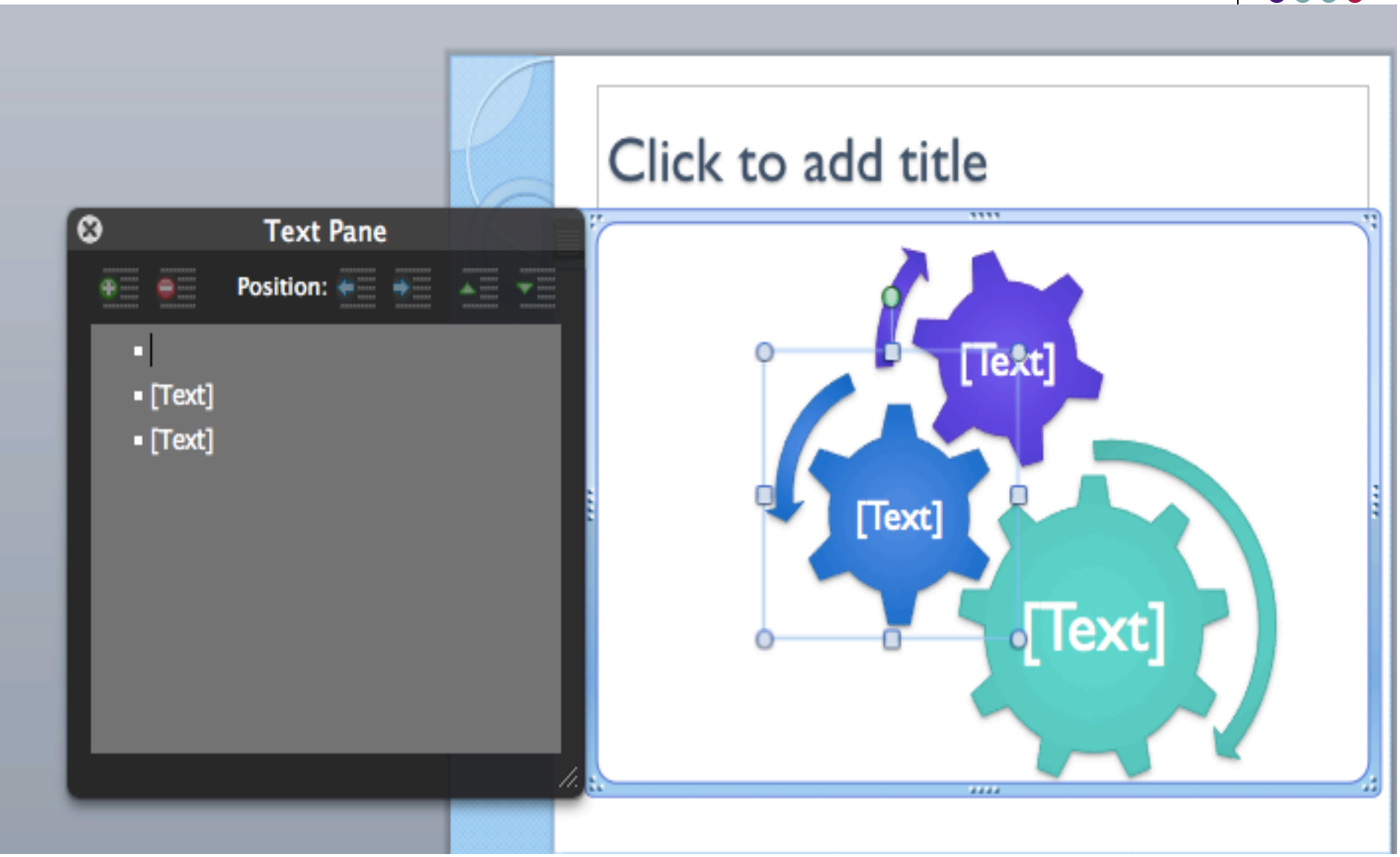
```javascript
google.charts.load('current', {packages: ['corechart', 'bar']});
google.charts.setOnLoadCallback(drawTitleSubtitle);

function drawTitleSubtitle() {
      var data = google.visualization.arrayToDataTable([
        ['City', '2010 Population', '2000 Population'],
        ['New York City, NY', 8175000, 5008000],
        ['Los Angeles, CA', 3792000, 3694000],
        ['Chicago, IL', 2695000, 2896000],
        ['Houston, TX', 2099000, 1953000],
        ['Philadelphia, PA', 1626000, 1017000]
      ]);
```

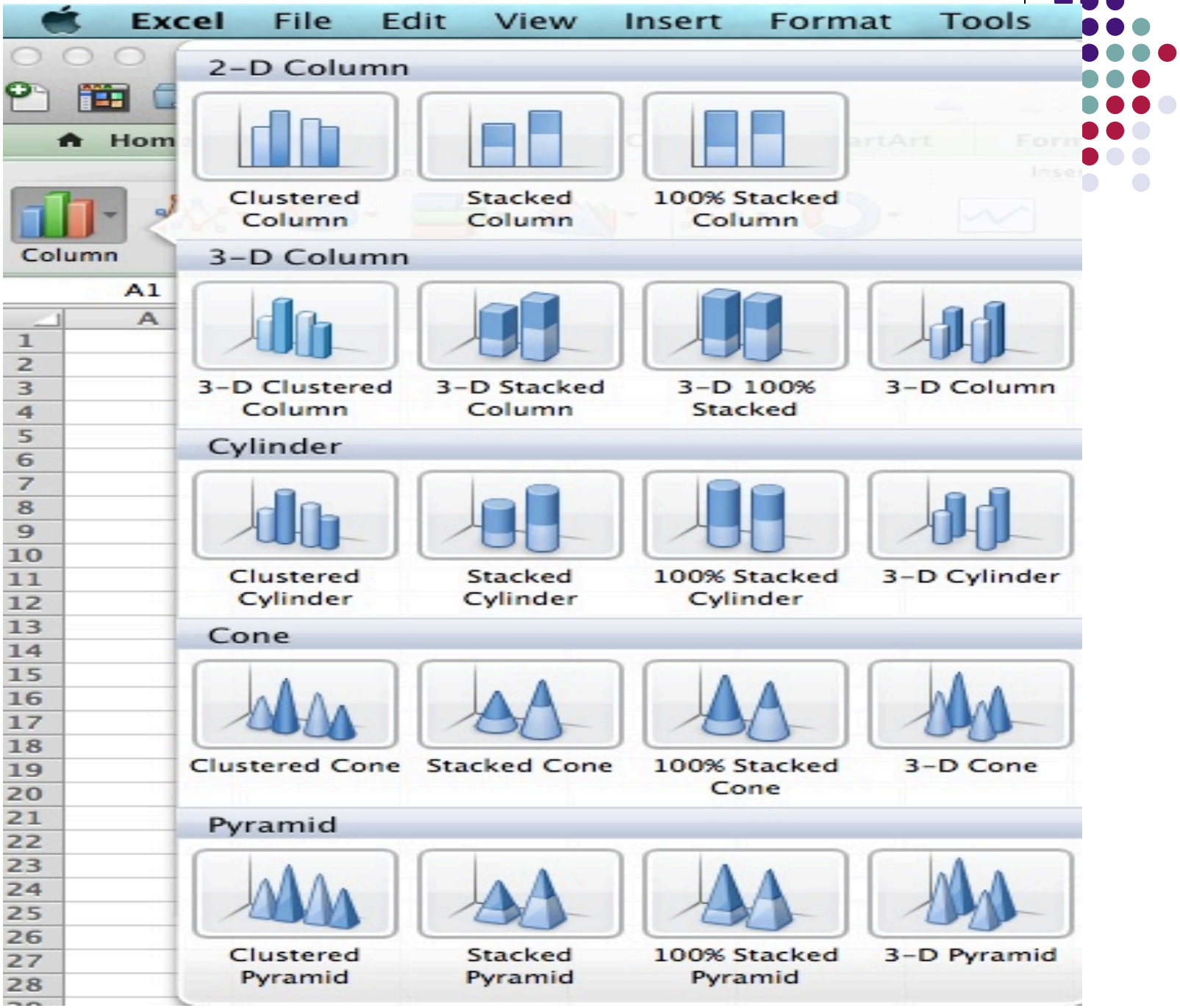
JAVASCRIPT

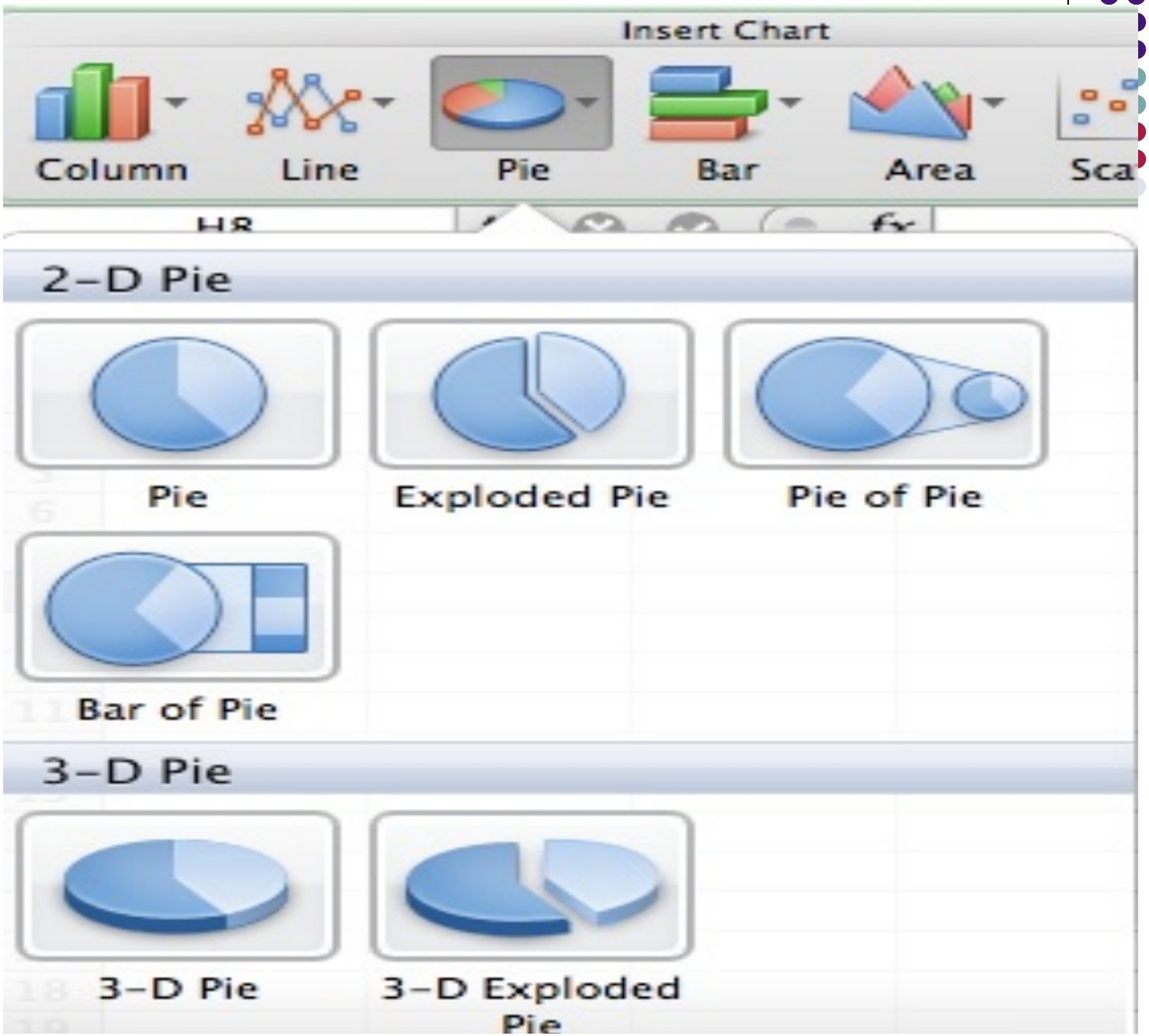


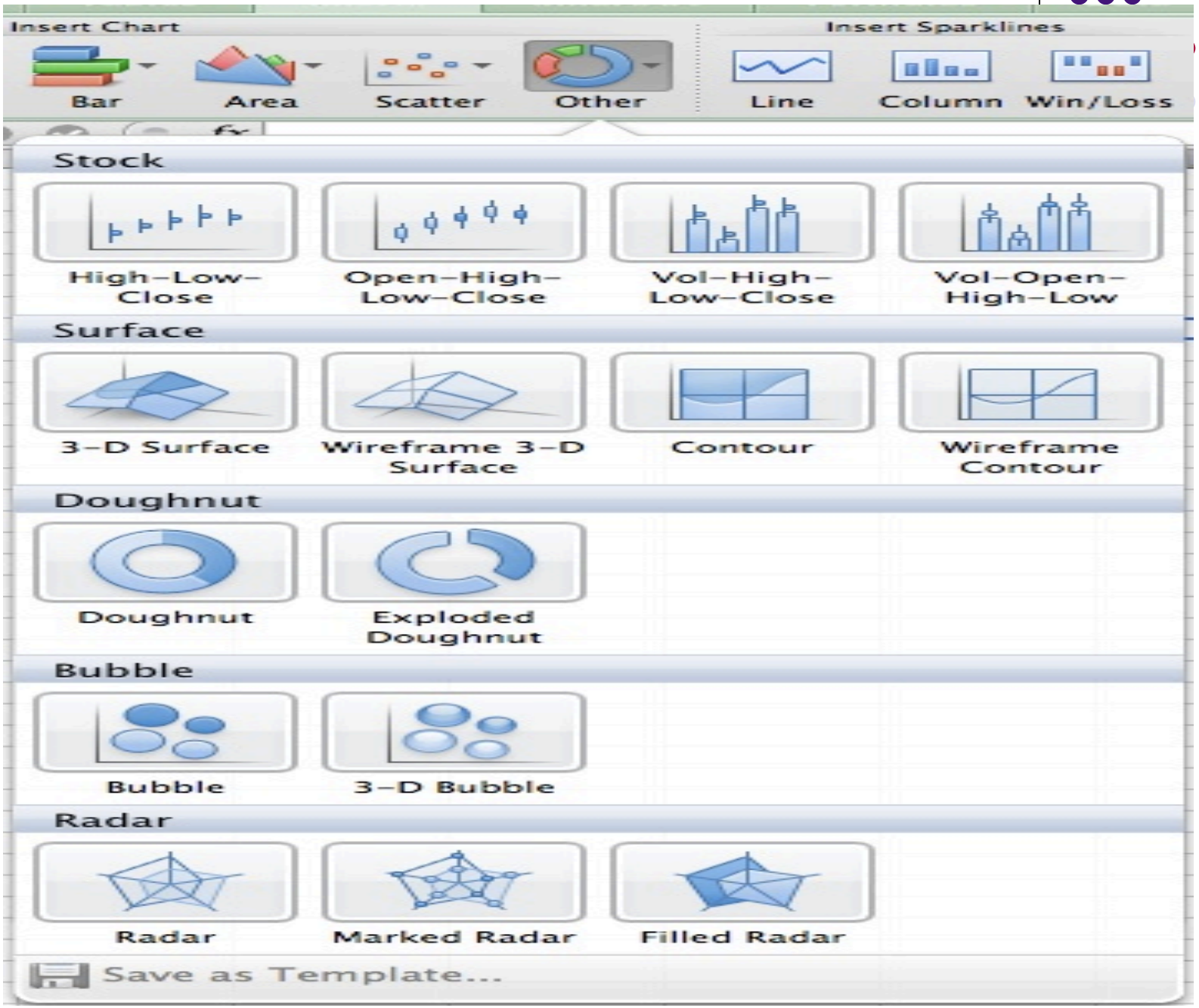

Home
Themes
Tables
Charts
SmartArt
Transitio
Insert SmartArt Graphic
Edi
List
Process
Cycle
Hierarchy
Relationship
Picture
Other
Shape


Home
Themes
Tables
Charts
SmartArt
Transitions
Insert SmartArt Graphic
Edit Smart
List
Process
Cycle
Hierarchy
Relationship
Picture
Other
Shape
Plus and Minus

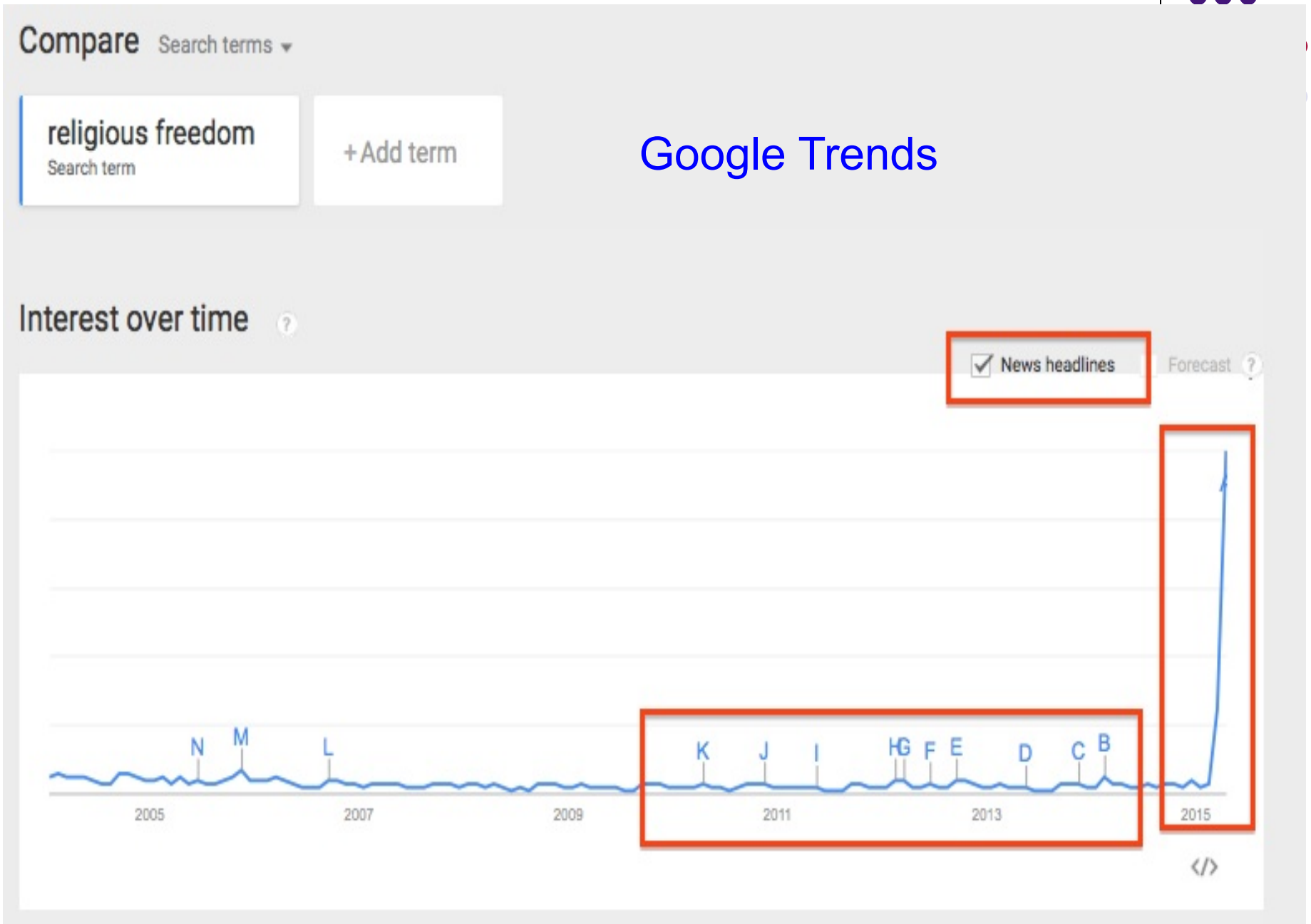

Click to add title
Text Pane
Position:
[Text]
[Text]
[Text]
[Text]
[Text]


Excel File Edit View Insert Format Tools
Hom
Column
A1
A
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
2-D Column
Clustered Column
Stacked Column
100% Stacked Column
3-D Column
3-D Clustered Column
3-D Stacked Column
3-D 100% Stacked
3-D Column
Cylinder
Clustered Cylinder
Stacked Cylinder
100% Stacked Cylinder
3-D Cylinder
Cone
Clustered Cone
Stacked Cone
100% Stacked Cone
3-D Cone
Pyramid
Clustered Pyramid
Stacked Pyramid
100% Stacked Pyramid
3-D Pyramid

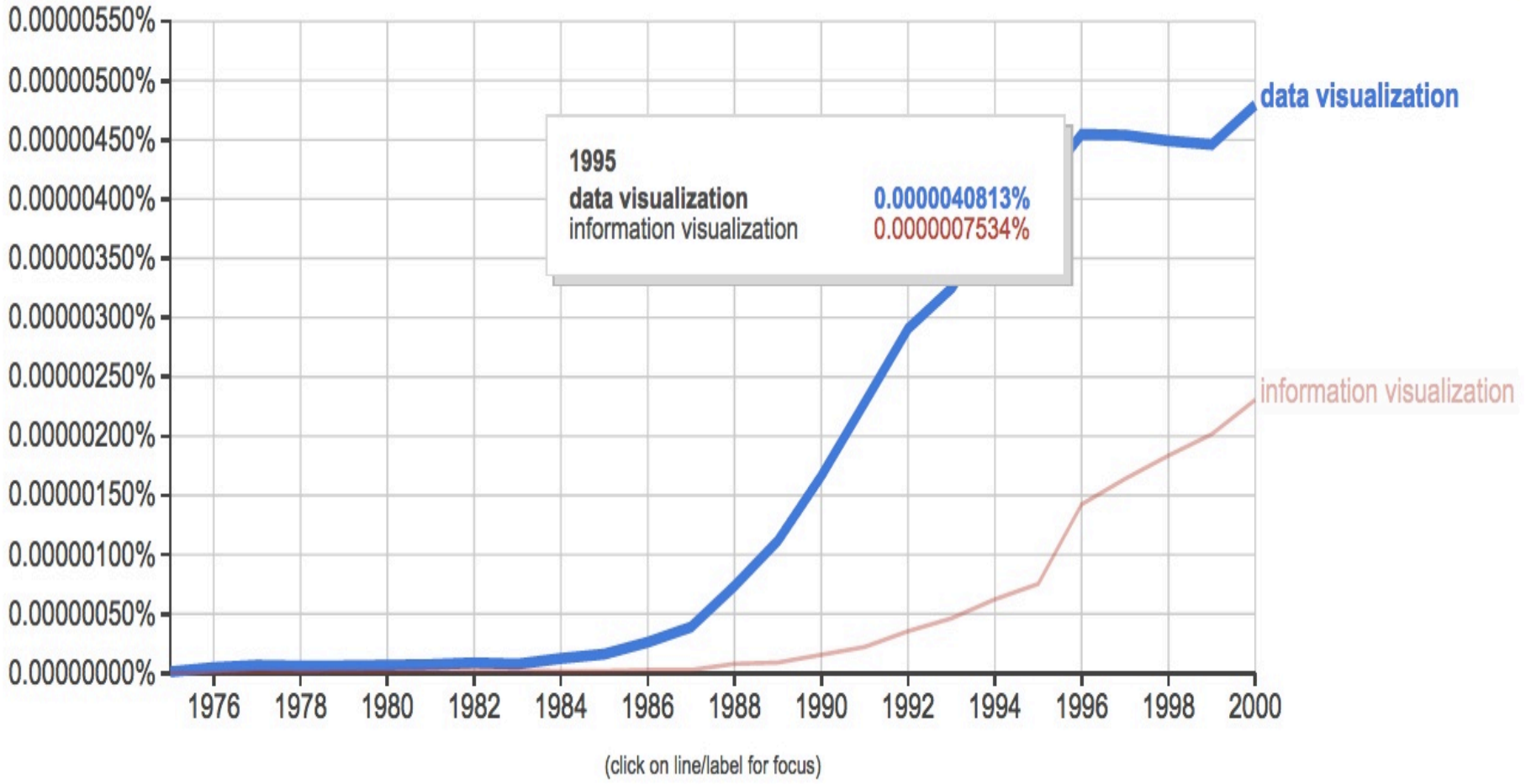

Insert Chart
Column
Line
Pie
Bar
Area
Sca
2-D Pie
Pie
Exploded Pie
Pie of Pie
Bar of Pie
3-D Pie
3-D Pie
3-D Exploded Pie


Insert Chart
Bar
Area
Scatter
Other
Insert Sparklines
Line
Column
Win/Loss
Stock
High–Low–Close
Open–High–Low–Close
Vol–High–Low–Close
Vol–Open–High–Low
Surface
3–D Surface
Wireframe 3–D Surface
Contour
Wireframe Contour
Doughnut
Doughnut
Exploded Doughnut
Bubble
Bubble
3–D Bubble
Radar
Radar
Marked Radar
Filled Radar
Save as Template...

Advances in Data Visualization


Compare
Search terms
religious freedom
Search term
+ Add term
Google Trends
Interest over time
News headlines
Forecast
N
M
L
K
J
I
H
G
F
E
D
C
B
2005
2007
2009
2011
2013
2015

Interest in Data and Information Visualization 1975-2000



Google Ngram

Workplace Digital Health Is Associated with Improved Cardiovascular Risk Factors in a Frequency-Dependent Fashion: A Large Prospective Observational Cohort Study

R. Jay Widmer, Thomas G. Allison, Brendie Keane, Anthony Dallas, Kent R. Bailey, Lilach O. Lerman, Amir Lerman

Published: April 19, 2016 • http://dx.doi.org/10.1371/journal.pone.0152657

Viewed

PLOS

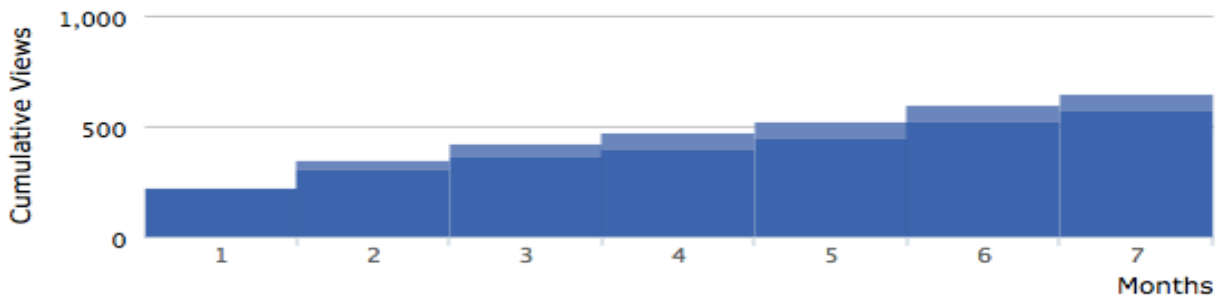
Total Article Views

642

Apr 19, 2016 (publication date) through Oct 21, 2016 *

	HTML Page Views	PDF Downloads	XML Downloads	Totals
PLOS	469	98	0	567
PMC	56	19	n.a.	75
Totals	525	117	0	642

22.29 % of article views led to PDF downloads



*Although we update our data on a daily basis, there may be a 48-hour delay before the most recent numbers are available. PMC data is posted on a monthly basis and will be made available once received.

Grab

WHERE WE CONNECT

AGES 18-24

TMI? Nearly a third (32%) of people aged 18-24 use social networking in the bathroom.

AGES 25-34

51%

More than half of people aged 25-34 use social networking in the office, more than any other age group.

Source: Nielsen 2012

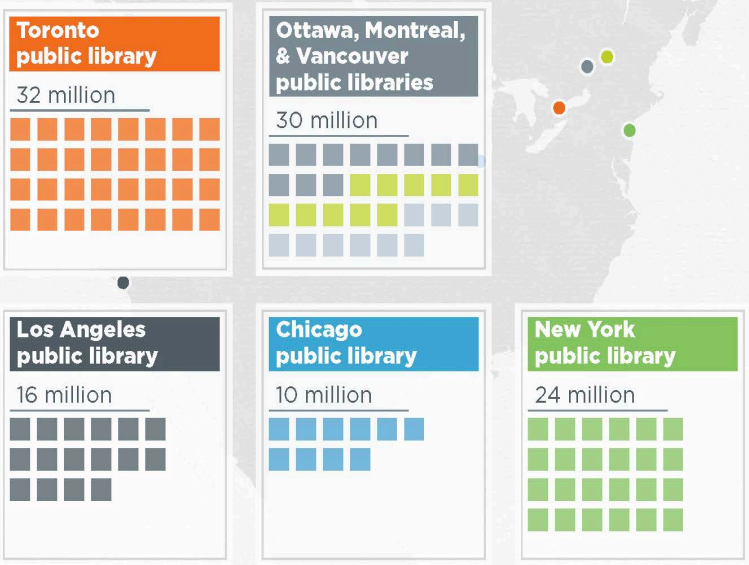
Toronto public libraries are busy.

LEARN MORE Read *The Great Equalizer: The Case For Investing in the Toronto Public Library* at **www.policyalternatives.ca/ontario**

CCPA
CANADIAN CENTRE for POLICY ALTERNATIVES ONTARIO OFFICE

In fact, they're the busiest in North America, and the second busiest in the world.

Total circulation is higher in the Toronto public libraries than the Los Angeles, Chicago, and even New York public libraries. It's higher than the circulation of the Ottawa, Montreal, and Vancouver public libraries put together.



Toronto public libraries are even more popular than the city's top 10 entertainment draws.

More people visit the Toronto public library every year than visit 10 of the most popular attractions in the Toronto area combined.

This includes the Air Canada Centre, Rogers Centre, CN Tower, Royal Ontario Museum, Canada's Wonderland, Toronto Zoo, Art Gallery of Ontario, Ontario Science Centre, CNE, and the Toronto International Film Festival.

PIWIK

Dashboard All Websites English

WEBSITE: VIRTUAL-DRUM... 2016-10-23 ALL VISITS DASHBOARD

YOU ARE VIEWING THE DEMO OF PIWIK

- Dashboard
- Visitors
- Actions
  - Pages
  - Entry pages
  - Exit pages
  - Page titles
  - Site Search
  - Outlinks
  - Downloads
  - Events

Visitors in Real-time

DATE	VISITS	ACTIONS
Last 24 hours	34	85
Last 30 minutes	2	4

Monday, October 24, - 11:11:05 *(8s)*

Direct Entry

Actions:

Monday, October 24, - 10:43:31 *(6 min 44s)*

from demo.piwik.org

Actions:

Monday, October 24, - 10:36:20 *(2 min 6s)*

Visits Over Time

Visitor Map

World-Wide Visits

Professional Services for Piwik

Did you know you can adjust the look and feel of Piwik to your brand, and even replace "Piwik" with your product name? Try the White Label product!

Referrers Overview

REFERRER TYPE	VISITS	▼ UNIQUE VISITORS
Websites	7	6

Resources

- Few, S. (2009). *Now you see it: Simple visualization techniques for quantitative analysis*. Oakland, CA: Analytics Press.
- Google Chart Gallery. https://developers.google.com/chart/interactive/docs/gallery
- Google Trend. https://www.google.com/trends/
- How to select charts for your data. Retrieved from

http://extremepresentation.typepad.com/files/choosing-a-good-chart-09.pdf

- Orcutt, D. (Ed.) (2010). *Library data: Empowering practice and persuation*. Santa Barbara, CA: Libraries Unlimited.
- A periodic table of visualization methods. Retrieved from http://www.visual-literacy.org/periodic_table/periodic_table.html
- Phelps, M. (Sept./Oct. 2012). Visualization tools for turning information into insights. *Information Today.* Retrieved from

http://www.infotoday.com/online/sep12/Phelps--Visualization-Tools-for-Turning-Information-Into-Insights.shtml

- Piwik. https://piwik.org/
- UK Data Archive. Research Data Life Cycle. Retrieved from http://www.data-archive.ac.uk/create-manage/life-cycle
- "Create a Graph. U.S. Dept. of Education, National Center for Education Statistics. Retrieved from http://nces.ed.gov/nceskids/createagraph/
- Ware, C. (2004). *Information visualization: Perception for design* (2nd ed.). San Francisco: Morgan Kaufmann.

Thank you!

Contact

Ingrid Hsieh-Yee

Professor

DLIS

Catholic University of America

Washington, DC 20064

hsiehyee@cua.edu