

Introduction to Data Visualization

Jane Zhao

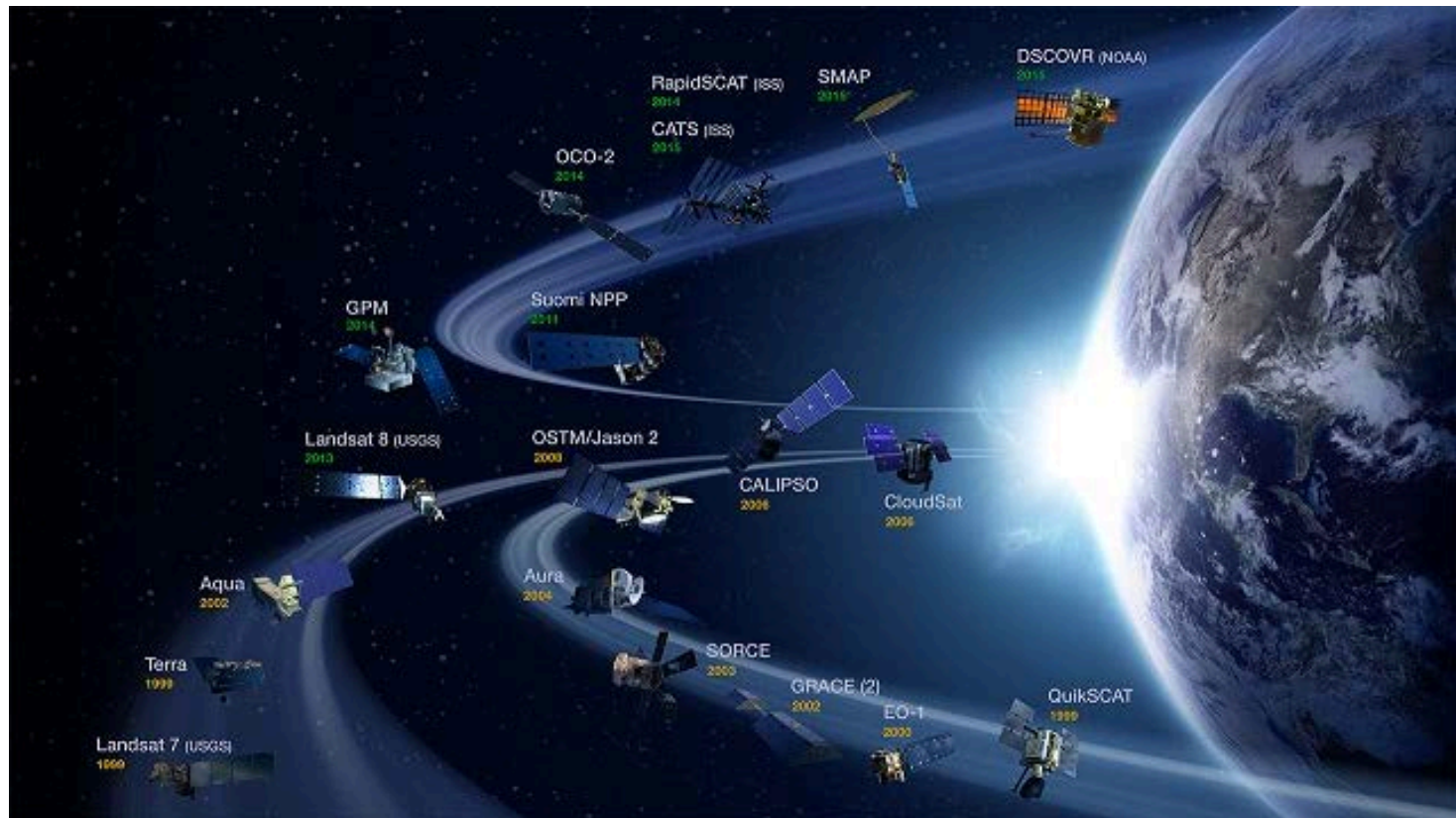
Digital Media Commons

Fondren Library

Objectives

- **Discuss** the needs of data visualization
- **Identify** software tools for data visualizations
- **Overview** the design guidelines for data visualization
- **See and learn** how data visualization works with Excel and Tableau

Data is generated everywhere and every day.



NASA Earth observation data are collected from a fleet of orbiting satellites as well as through airborne and field campaigns. NASA graphic. NASA graphic <https://earthdata.nasa.gov/eosdis-role-in-bedi>

Data is generated faster than ever
before.

Data is generated faster than ever before, from personal data perspective.

Floppy Disk,
1.44 MB, 1987



Zip Drive,
100-250
MB, 1994

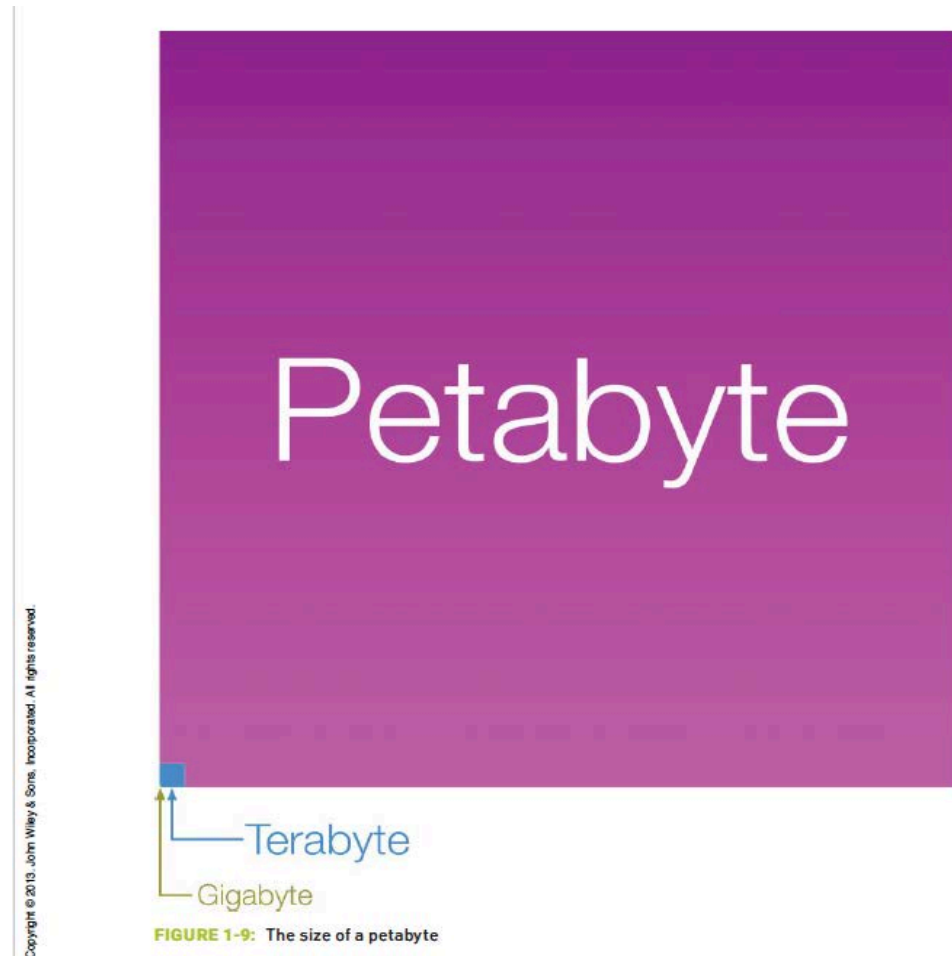


USB Drive,
8-256 GB, 2016
/8000 – 256,000 MB,
2016



Wikipedia, https://en.wikipedia.org/wiki/Floppy_disk ,
https://en.wikipedia.org/wiki/Zip_drive, https://en.wikipedia.org/wiki/USB_flash_drive

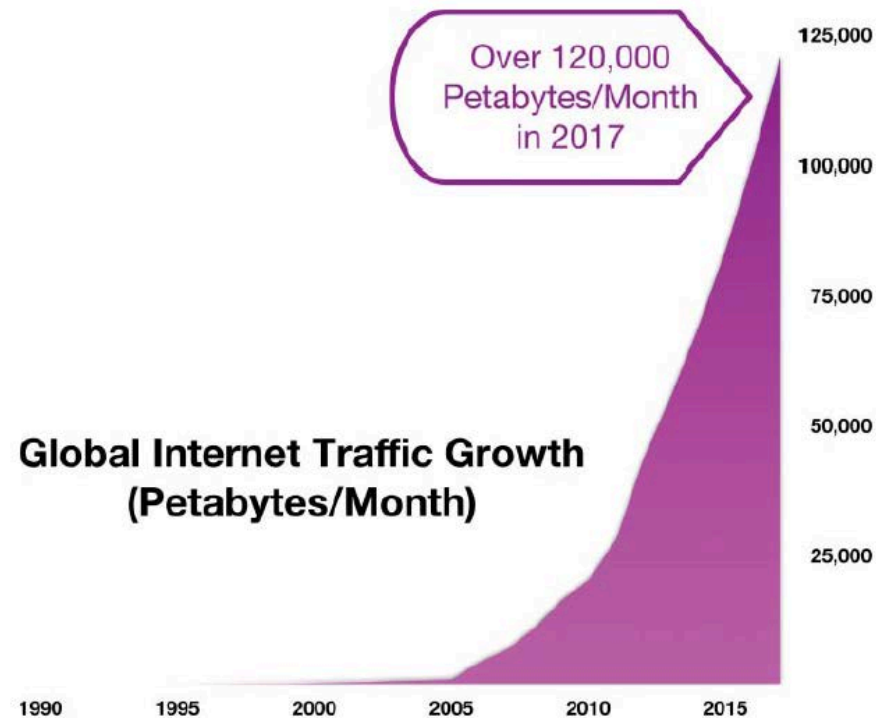
Data is generated faster than ever before,
from business and scientific data perspective.



Krum, Randy. Cool Infographics : Effective Communication with Data Visualization and Design, John Wiley & Sons, Incorporated, 2013.

Estimate the sheer magnitude of information by measuring the amount of data that moves across the Internet.

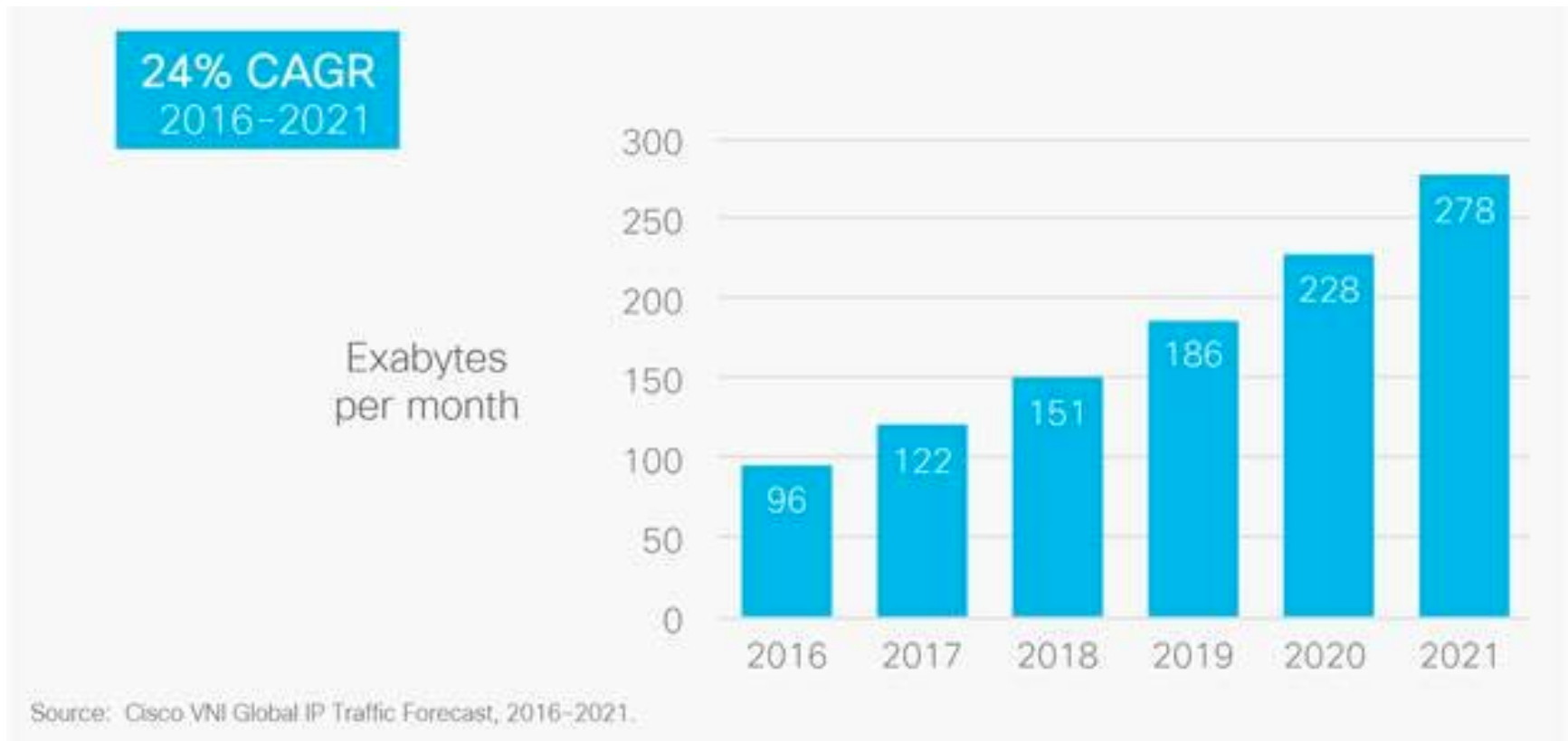
FIGURE 1-10:
Global Internet
traffic growth



Visual Networking Index, Cisco, 2013

Krum, Randy. Cool Infographics : Effective Communication with Data Visualization and Design, John Wiley & Sons, Incorporated, 2013.

Cisco VNI forecasts 278 Exabyte per month of IP traffic by 2021



Cisco,

<https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.html>, 6/17/2017

We need an effective way to
understand data.

Data visualization helps people understand data and see the insights intuitively.

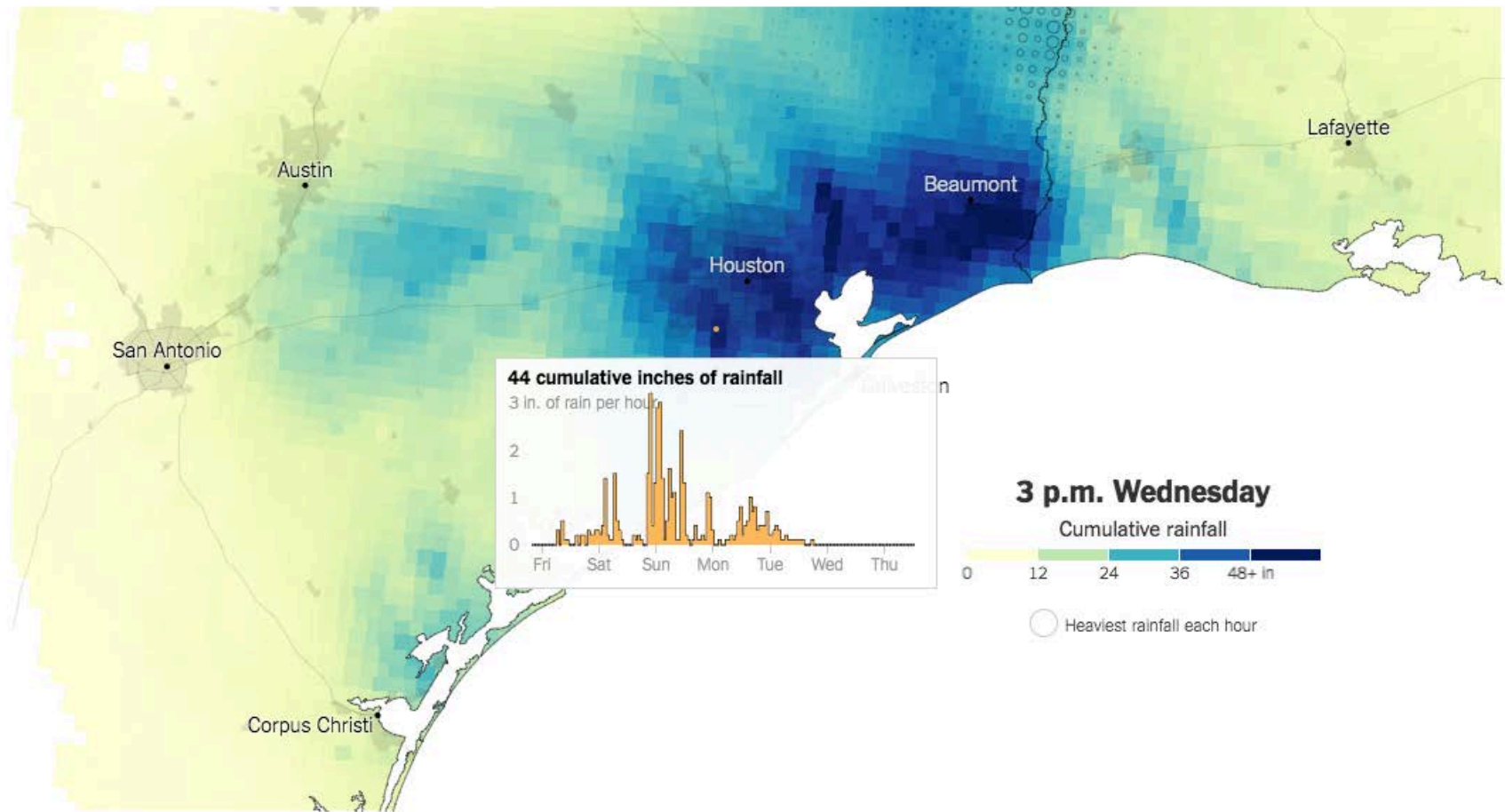
Meanings of Data Visualization

Task	Goal
Exploration: Searching for significant facts	Discovery
Sensemaking: Examining and making sense of data	Understanding
Communication: Conveying information to others	Informed decisions

Data visualization is the use of visual representations to explore, make sense of, and communicate data. Stephen Few, Visual Business Intelligence,

<https://www.perceptualedge.com/blog/?p=1897>

A data visualization helps people intuitively see how much rainfall Houston received from Hurricane Harvey.

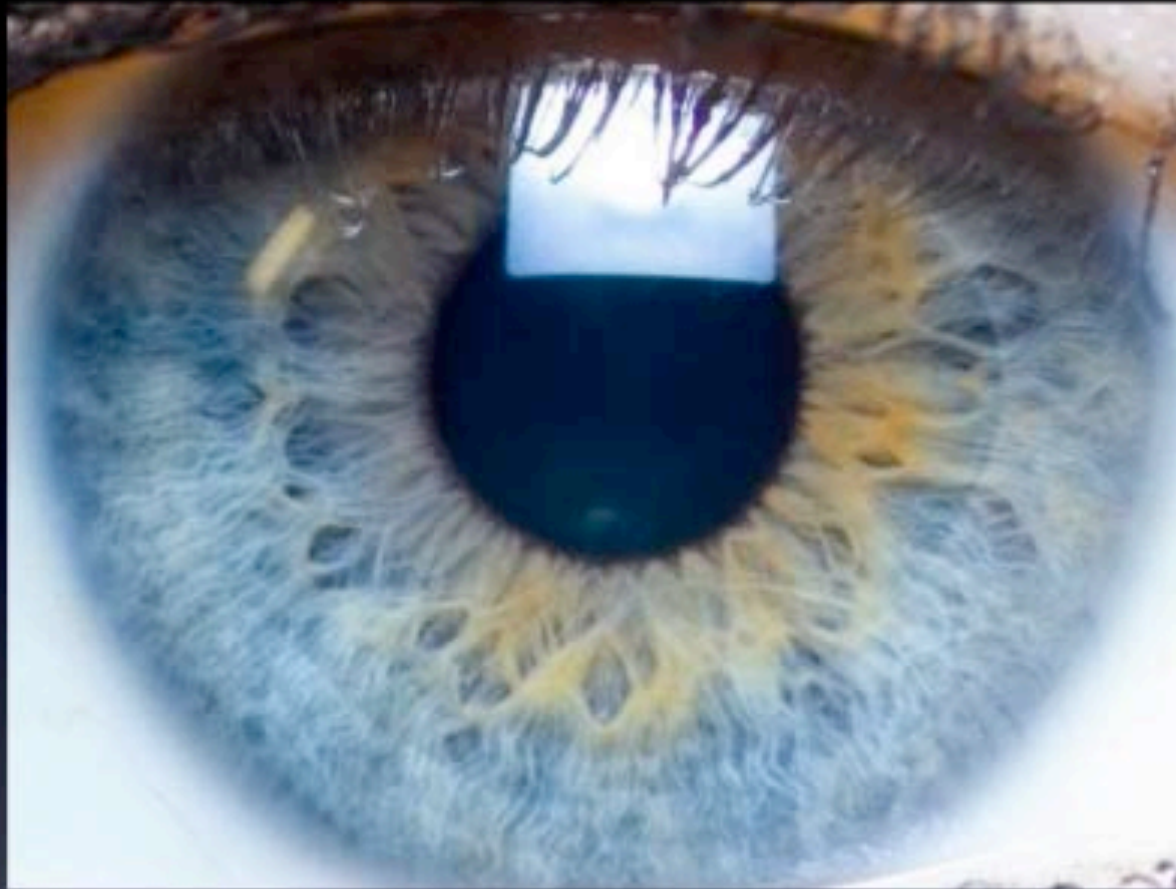


Source: National Weather Service

Maps: Tracking Harvey's Destructive Path Through Texas and Louisiana, New York Times, <https://www.nytimes.com/interactive/2017/08/24/us/hurricane-harvey-texas.html>

The Science of Data Visualization

WHY DATA VISUALIZATION WORKS?



80% of the brain is dedicated to
visual processing

University of Rochester, 2004



The human brain is a
pattern recognition machine

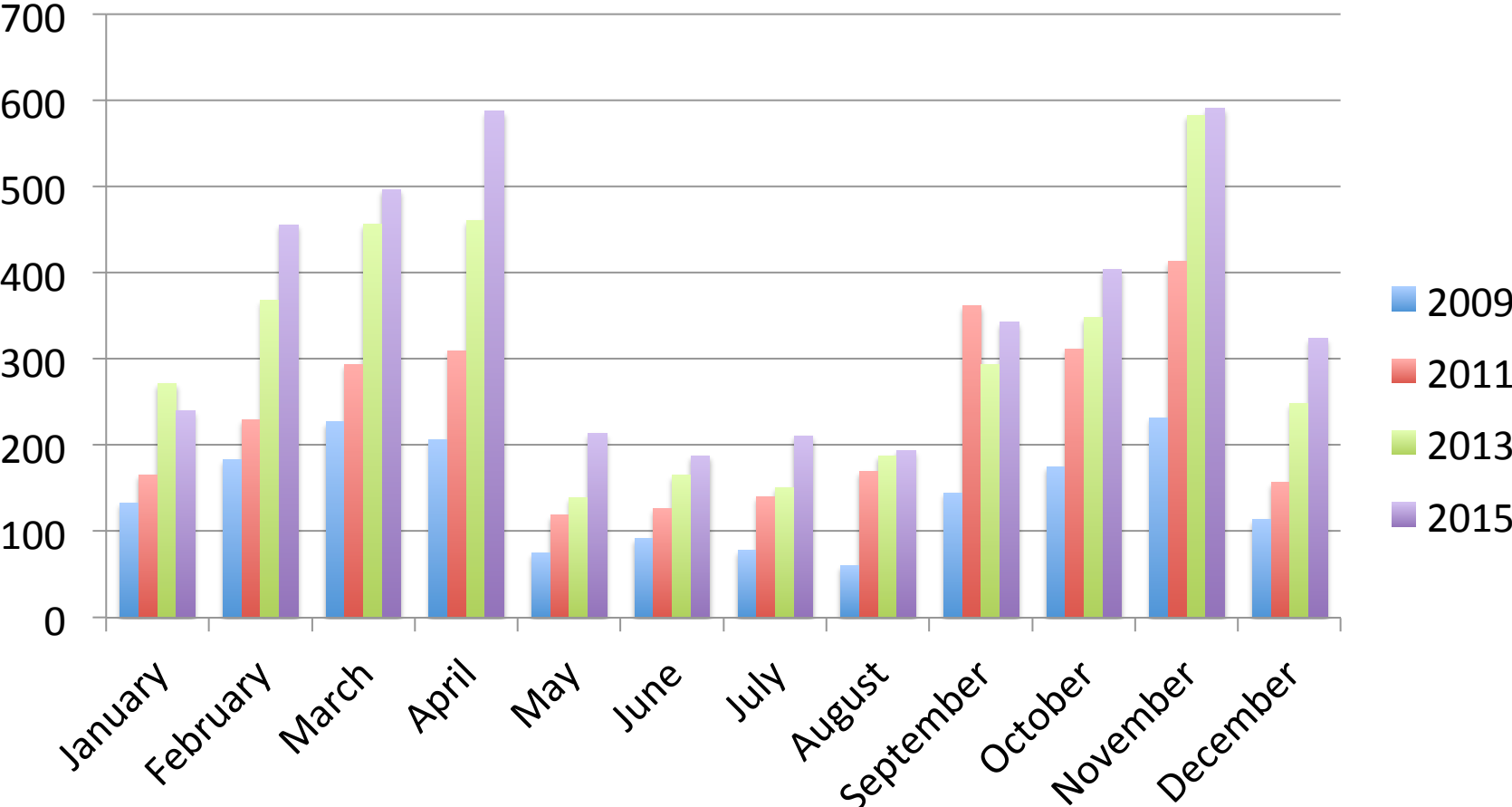
A Table of Data, Hard to See its Pattern and Trend.

DMC Equipment Circulation Statistics

Month	2009	2011	2013	2015
January	133	166	272	240
February	183	230	368	456
March	227	294	457	497
April	207	310	461	588
May	75	119	139	214
June	92	127	165	188
July	78	140	151	211
August	60	170	188	194
September	145	362	294	343
October	175	312	348	404
November	232	414	583	591
December	114	157	249	324

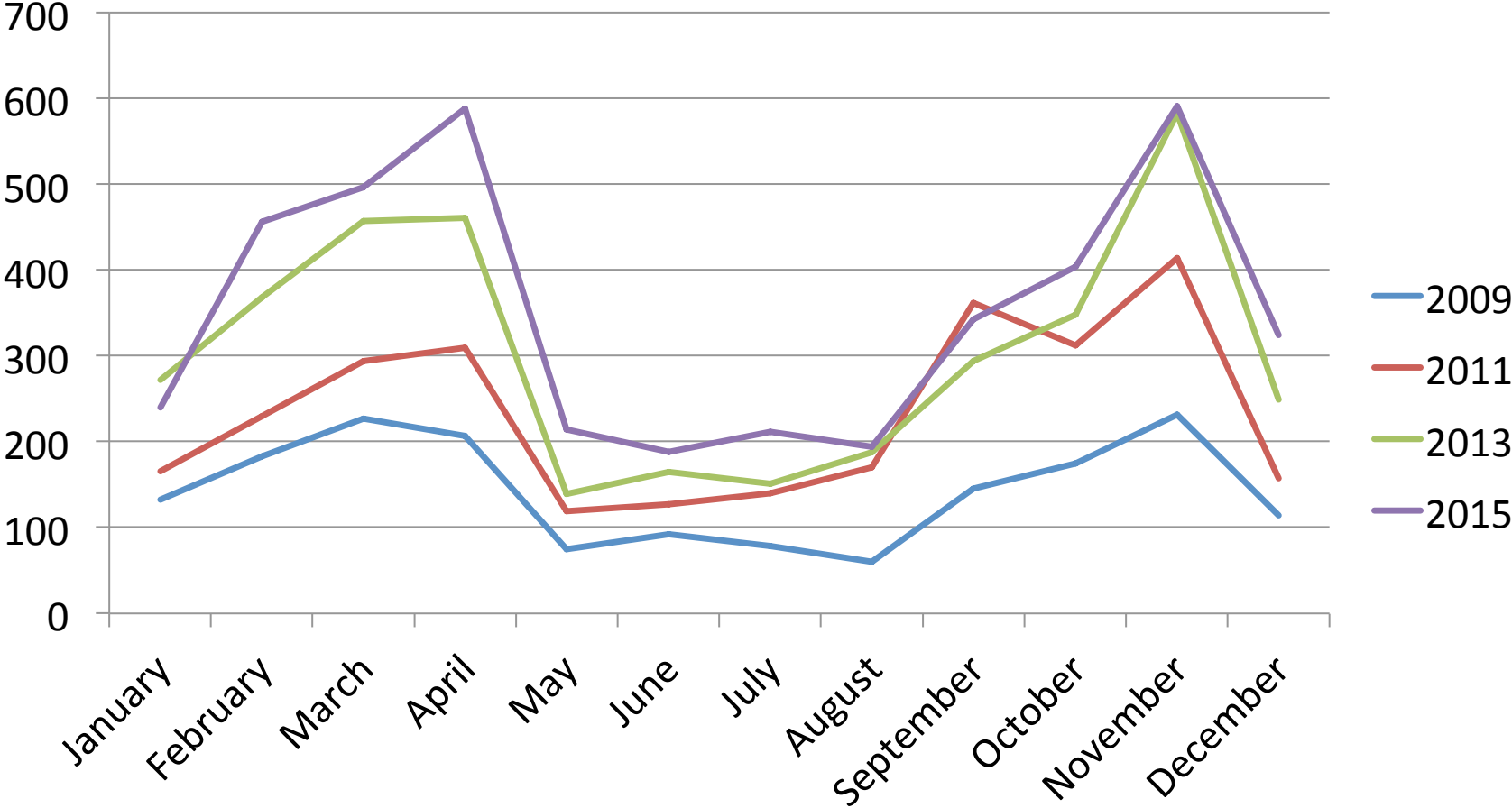
Convert the Data to a Bar Chart, Easy to See the Pattern.

DMC Equipment Circulation Statistics



Convert the Data to a Line Chart, Easy to See the Trend.

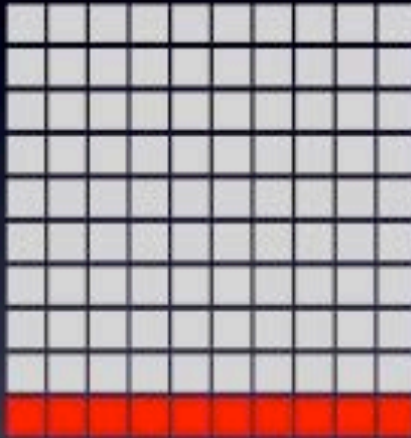
DMC Equipment Circulation Statistics



III.

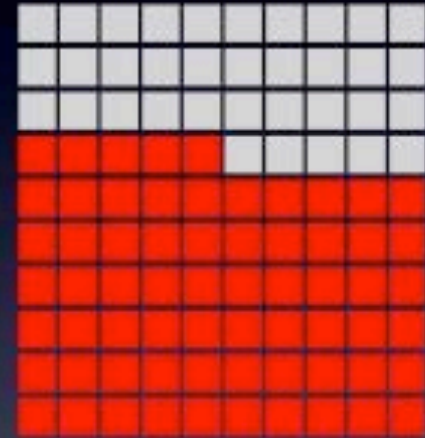
Picture Superiority Effect

Memory retention after 3 days



10%

Text or Audio Only



65%

Text + Picture

“Of all methods for analyzing and communicating statistical information, well-designed data graphics are usually the simplest and at the same time the most powerful.”

Eward Tufte, Yale Professor

DATA VISUALIZATION EXAMPLES AND TOOLS

What is Visualization?

Visualization is the transformation of data or information into pictures.

- *Schroeder et al. The Visualization Toolkit, 4th ed. 2006*

Type of Data Visualization Based on Data Sources

- **Scientific visualization**
Data from sciences and engineering field such as seismic data, medical data, simulation and measurement data from physics, chemistry, geo-science, medical-biological, climate, oceanography, energy, ...
- **Information visualization**
Data from financial, marketing, business, HR, statistical, social media, political, ...

Information Visualization examples

- Netflix Queues (NYTimes)

<http://www.nytimes.com/interactive/2010/01/10/nyregion/20100110-netflix-map.html?ref=nyregion>

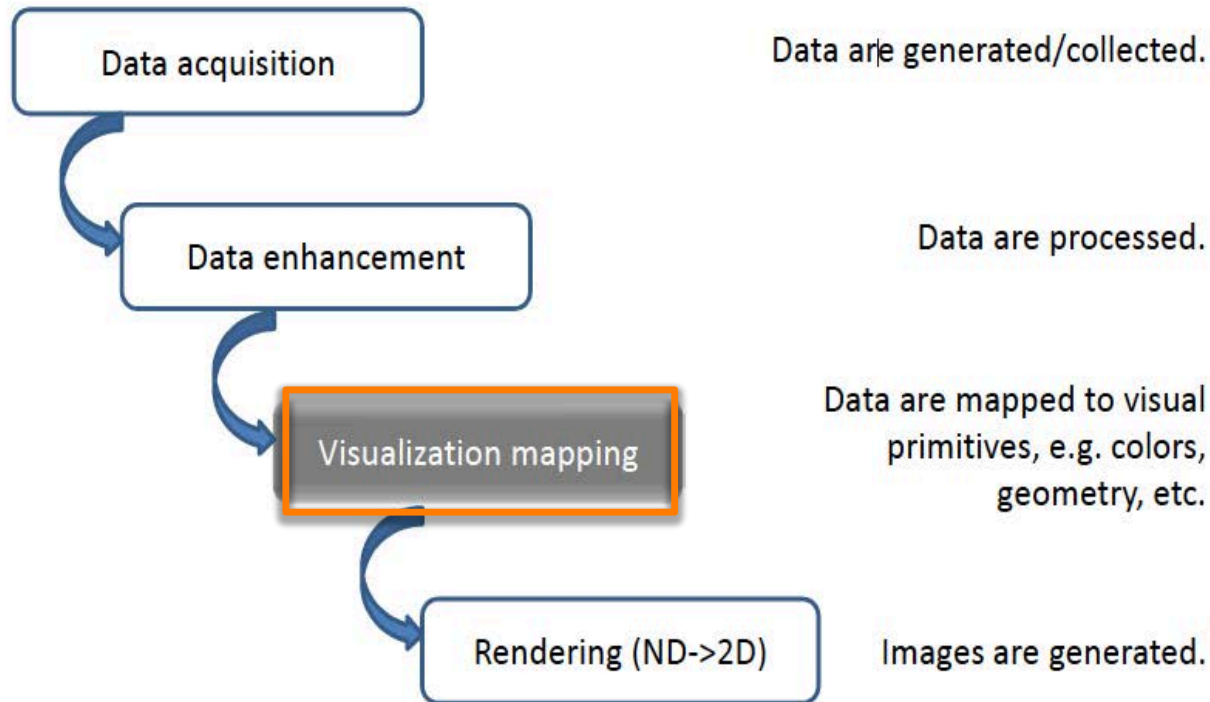
- Unemployment Visualization (NYTimes)

<http://www.nytimes.com/interactive/2009/11/06/business/economy/unemployment-lines.html>

- Interesting Examples from Fernanda Martin and Viegas Wattenberg

<http://hint.fm/>

A Visualization Pipeline



Business Intelligent System Five Layers by Microsoft

Presentation Layer (e.g. data visualization tools such as Tableau)

Analytical Layer

Data Storage

Transform Data

Data source layer

Data Visualization Tools

- Microsoft Excel
 - Pivot table
 - PowerPivot – an add-in for Excel
- Python or R
- Tableau
- Software supported at Rice Visualization Center <http://viz.blogs.rice.edu/software/>

Types of Data Visualization Graphs

- Static
 - Rice V2C2: <https://v2c2.rice.edu/v2c/research>
- Interactive
 - Working with data in a very visual dynamic way
 - Making working with data an easy and intuitive way
 - Tableau Gallery
<https://www.tableau.com/solutions/gallery>

DATA VISUALIZATION GUIDELINES

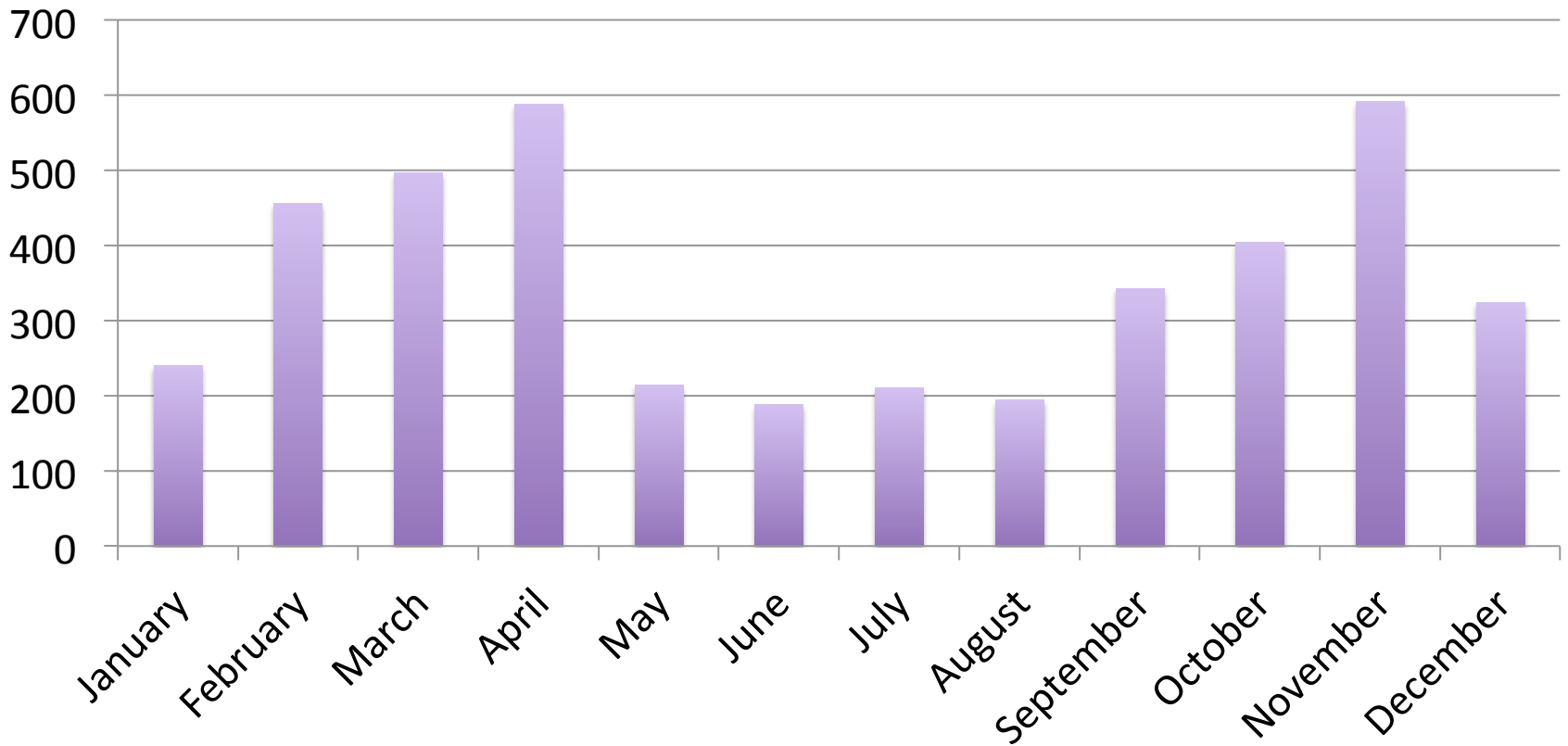
I. Start with a question

II. Choose the appropriate graph based on the relationship type!

- Graphs are a representation of the relationships in quantitative information.
- The graph type chosen is based on the type of relationship.
- Different types of graphs can display some types of relationships better than others.
- When there are a number of acceptable options, choose the graph that you think is the best (most effective) way to convey your message to your audience (Lankow et al., 2012, p. 213).

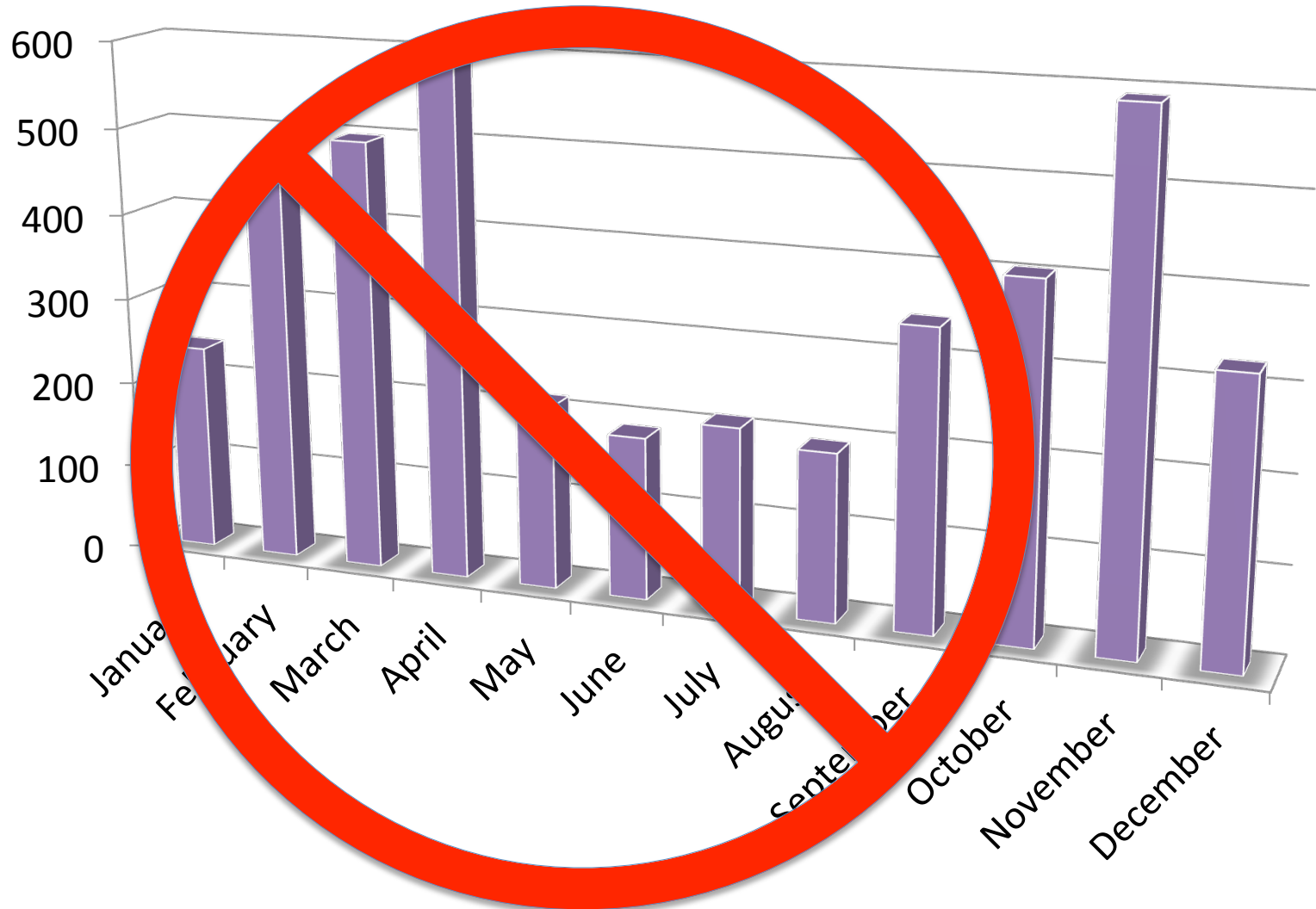
Bar Chart for Ranking or Time Series

2015 DMC Equipment Circulation Statistics



Avoid 3-D Bar Chart

2015 DMC Equipment Circulation Statistics

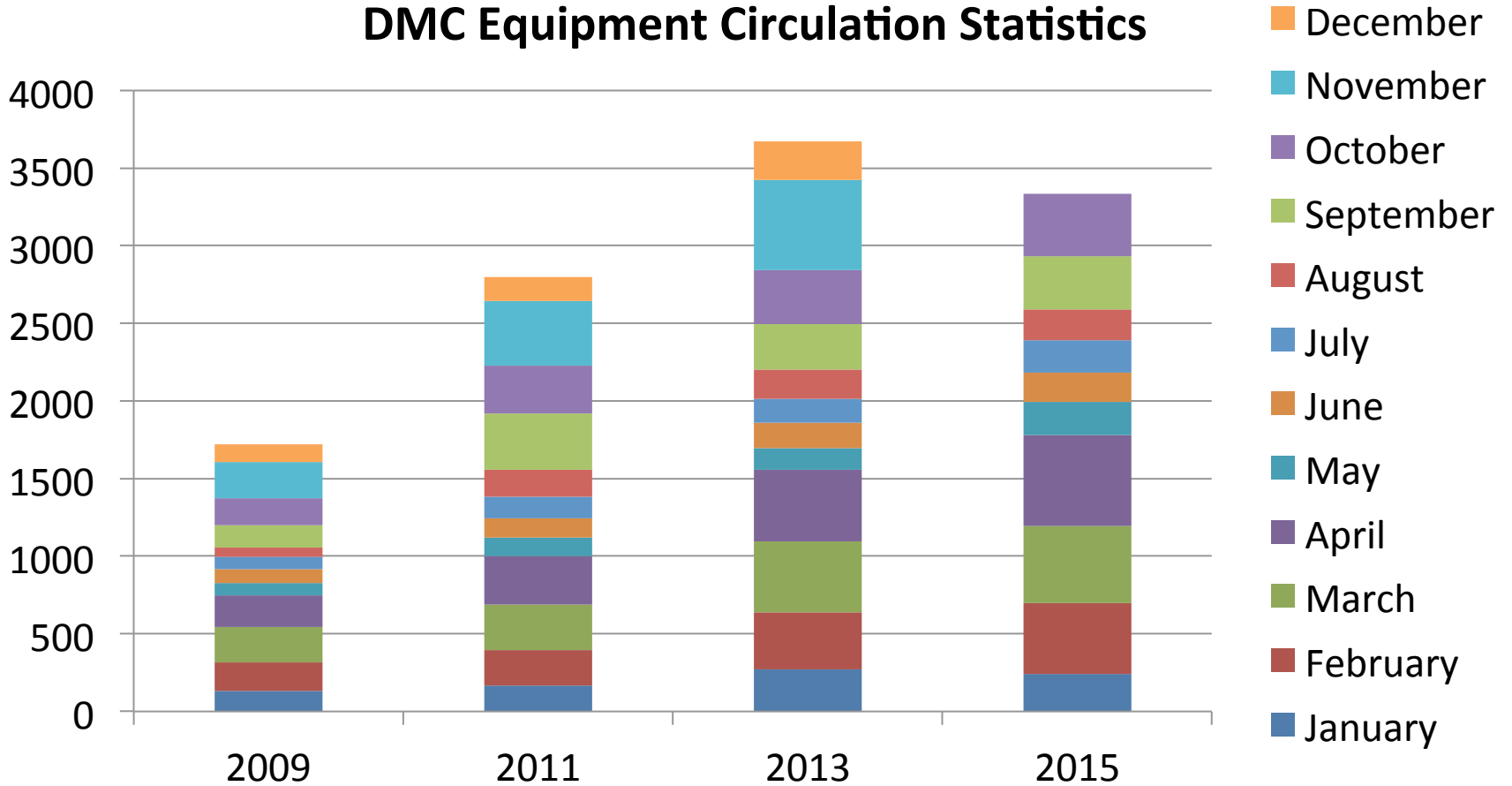


Avoid 3-D Bar Chart



Stacked Bar Chart for Multiple Part-to-Whole Relationships

DMC Equipment Circulation Statistics



Formatting Bar Graphs

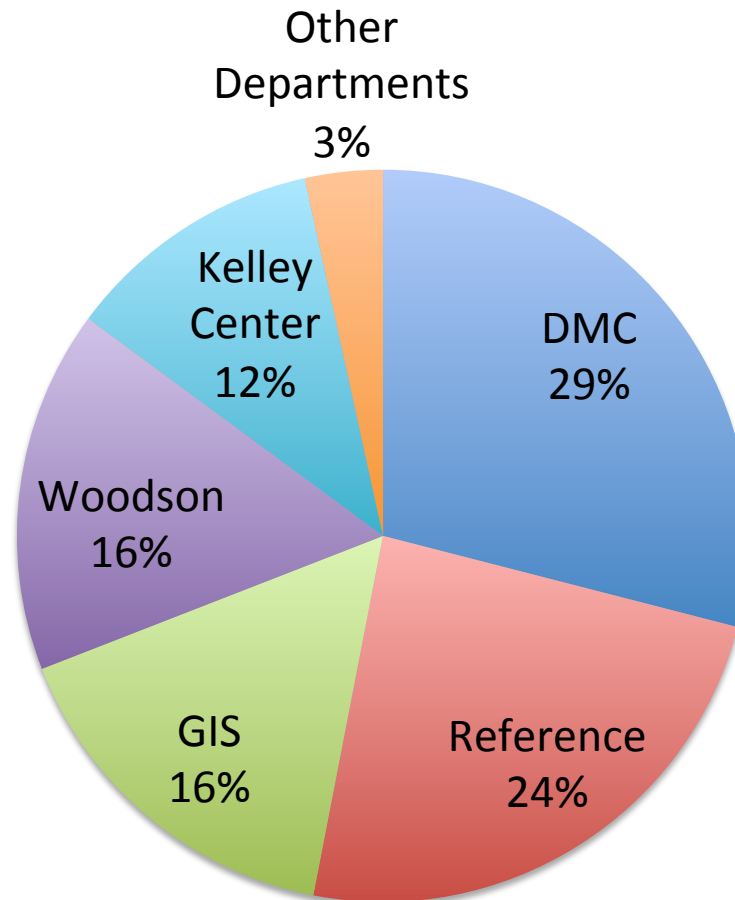
<https://www.coursera.org/learn/oral-communication/lecture/8NDyY/how-to-show-complex-data> - Oral Communication for Engineering Leaders on Coursera

- *Beata Krupa, Lecturer in Professional and Engineering Communication, Rice Center for Engineering Leadership.*

More Bar Chart Design Examples

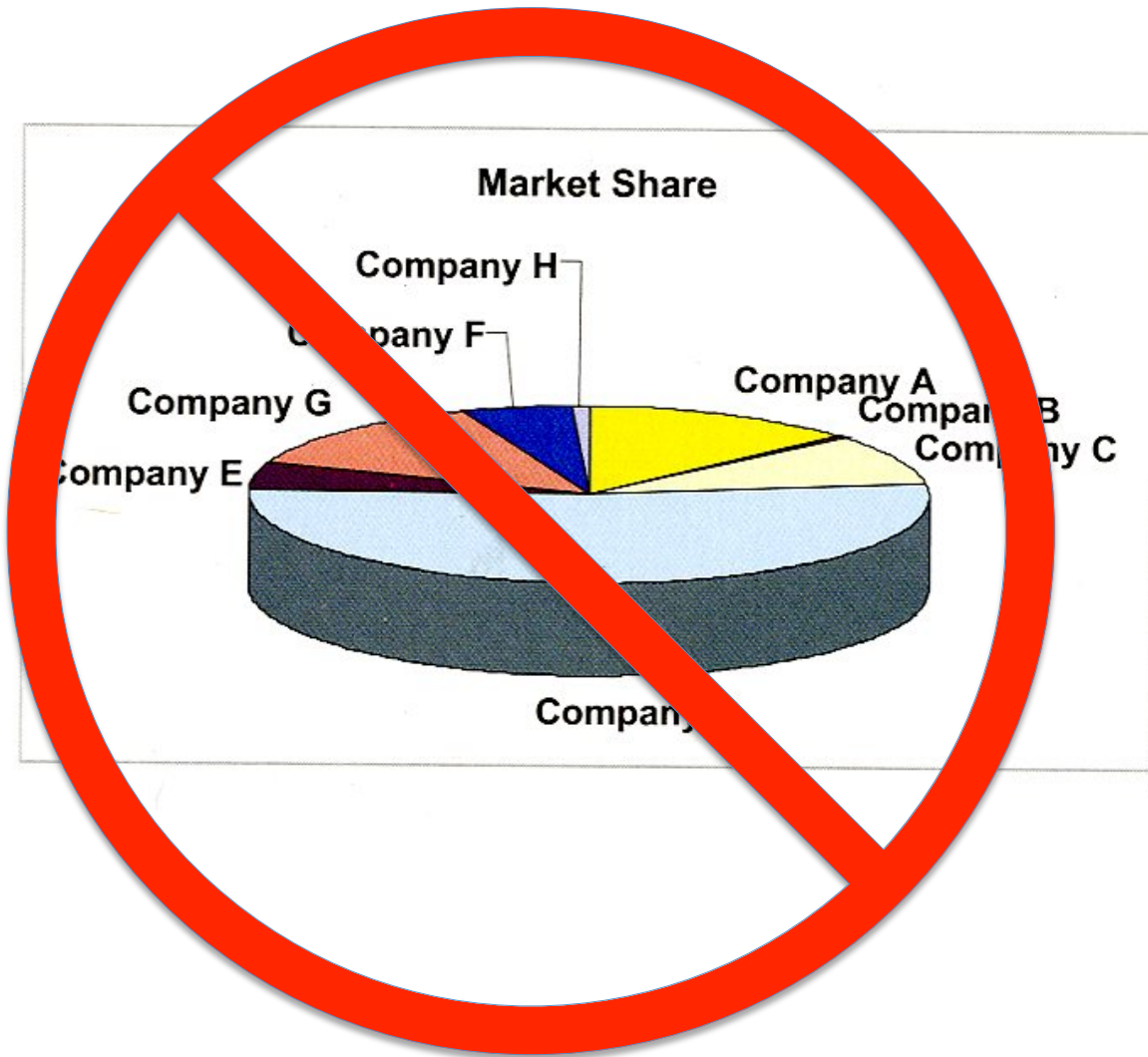
- <http://www.perceptualedge.com/example17.php>
- <http://www.perceptualedge.com/example13.php>
 - Stephen Few, <http://www.perceptualedge.com/>

Pie Chart for Part-to-Whole Comparisons



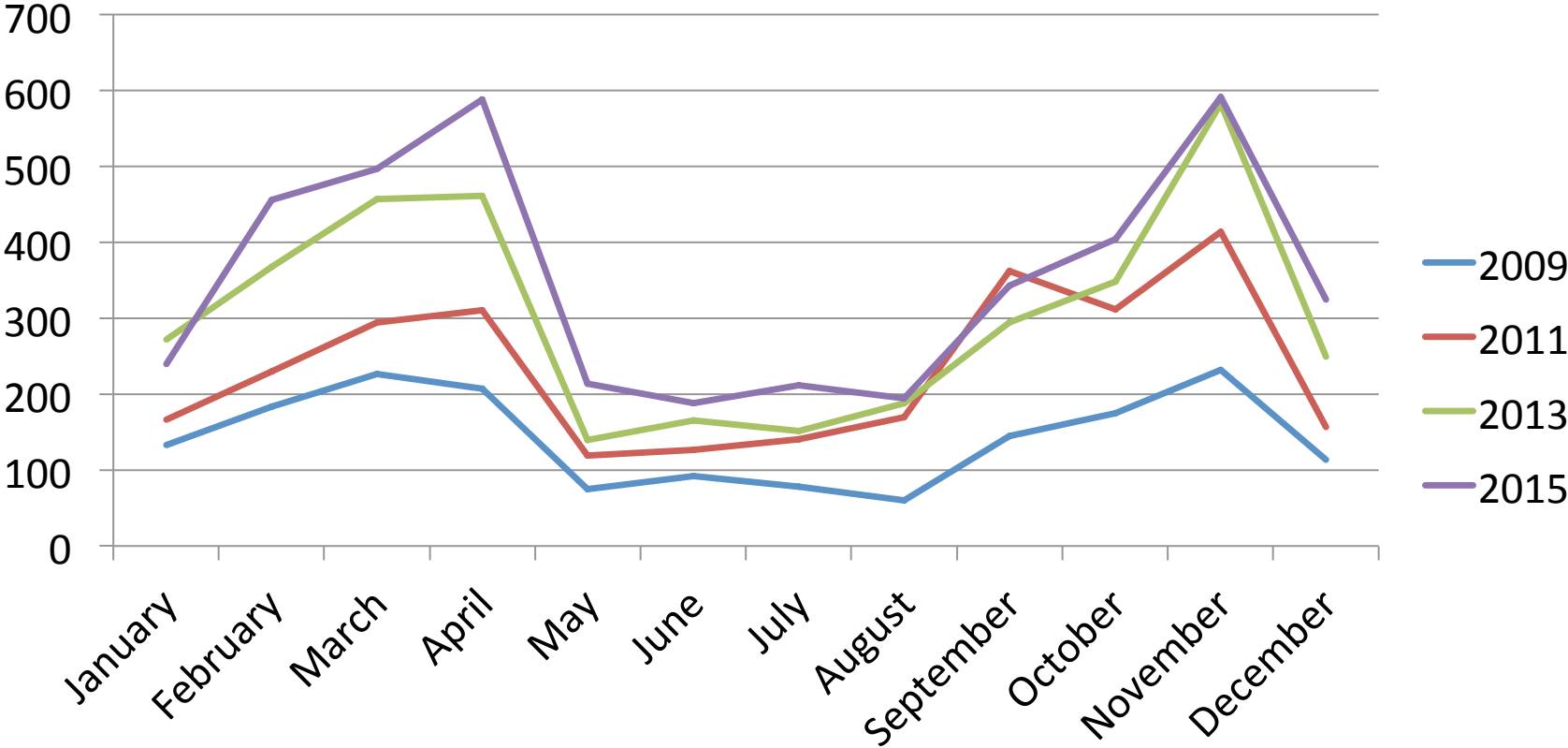
2014-2015 Library Instruction Session Statistics

Avoid 3-D Pie Chart



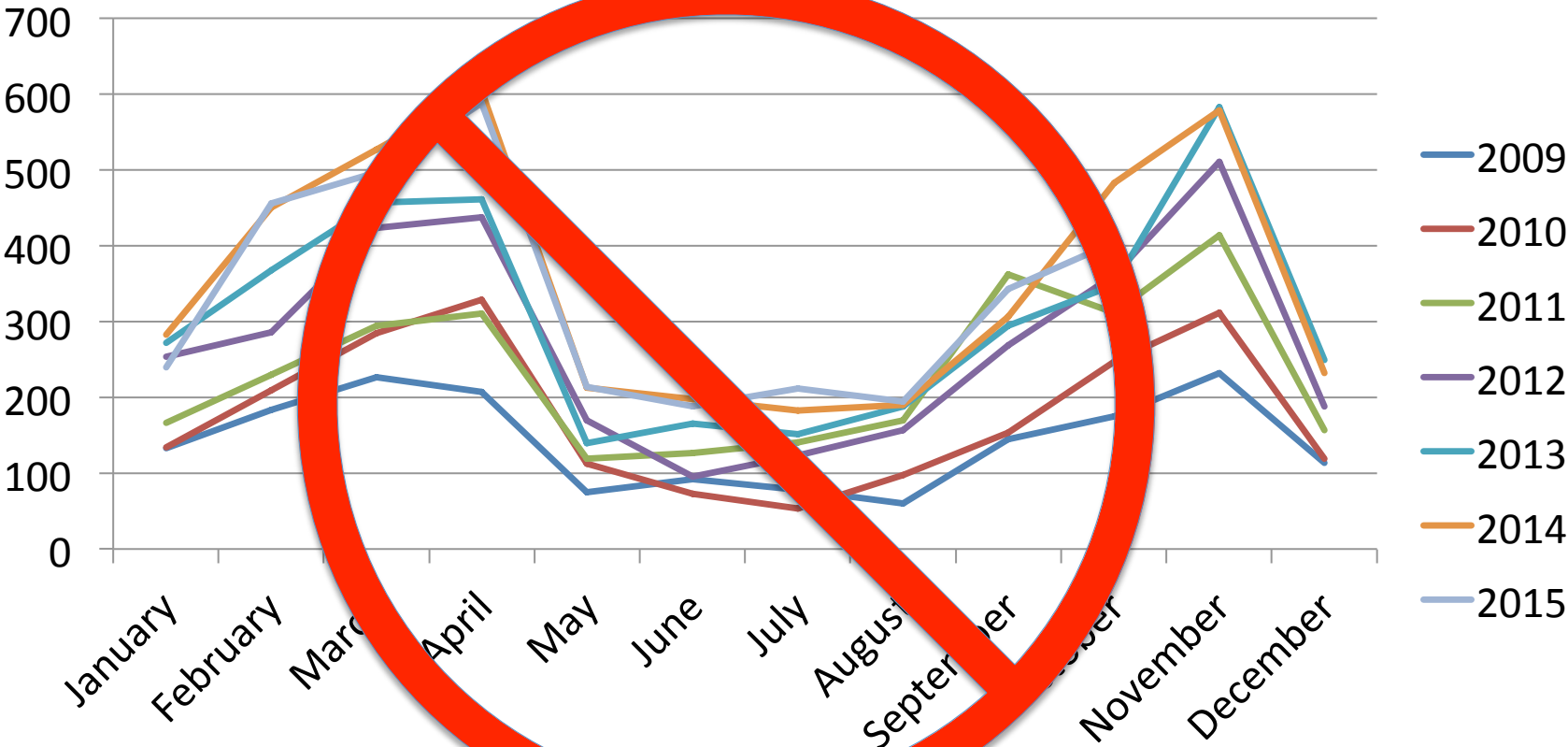
Line Chart for Time Series

DMC Equipment Circulation Statistics



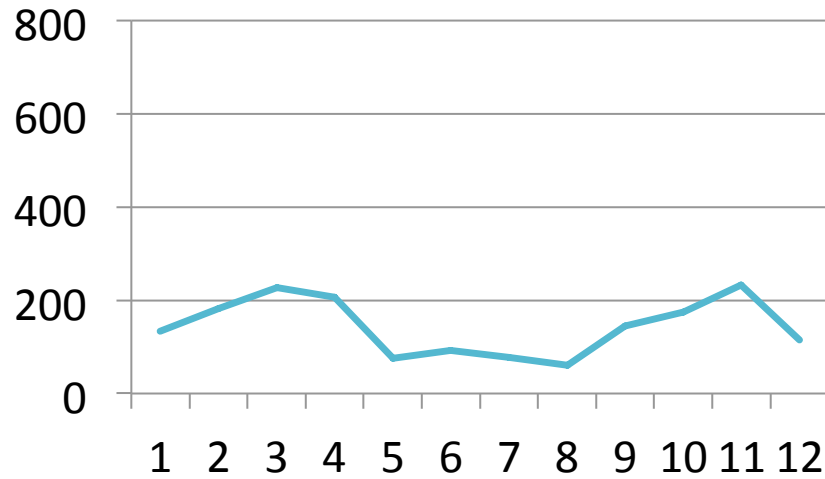
Keep the Line Chart to Four or Fewer. Otherwise the Chart is Too Busy!

DMC Equipment Circulation Statistics

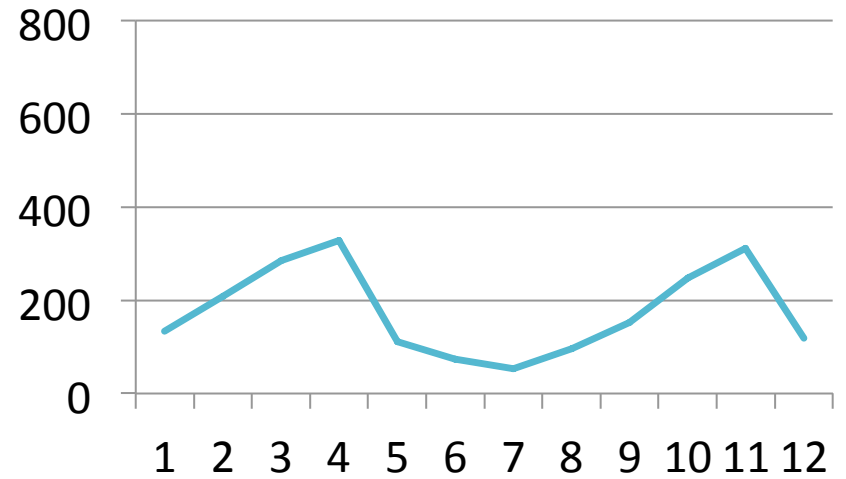


Use the Practice of Paneling and a Constant Scale for Consistency if You have More Than Four Lines.

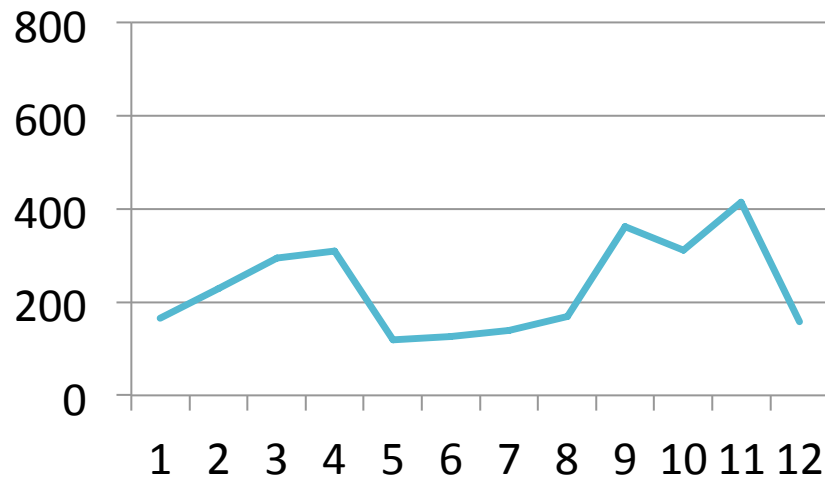
2009



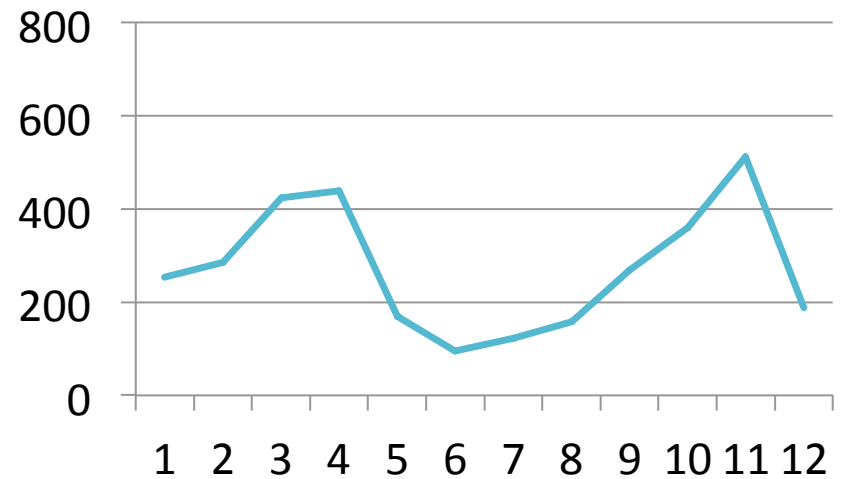
2010



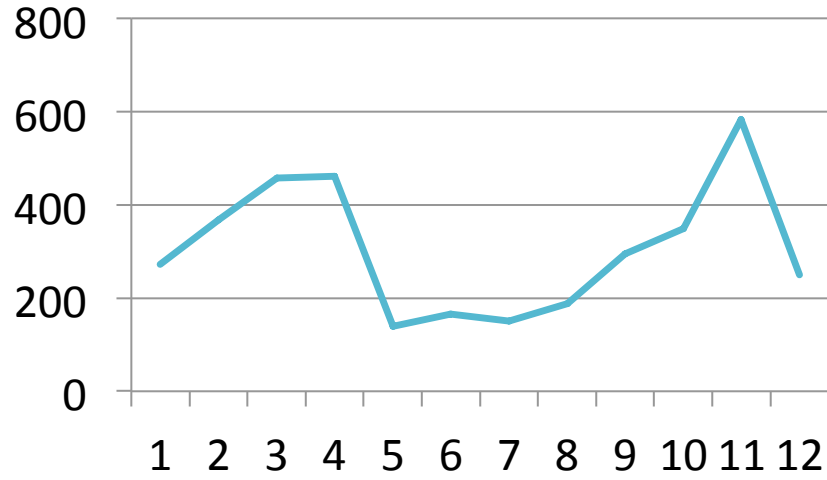
2011



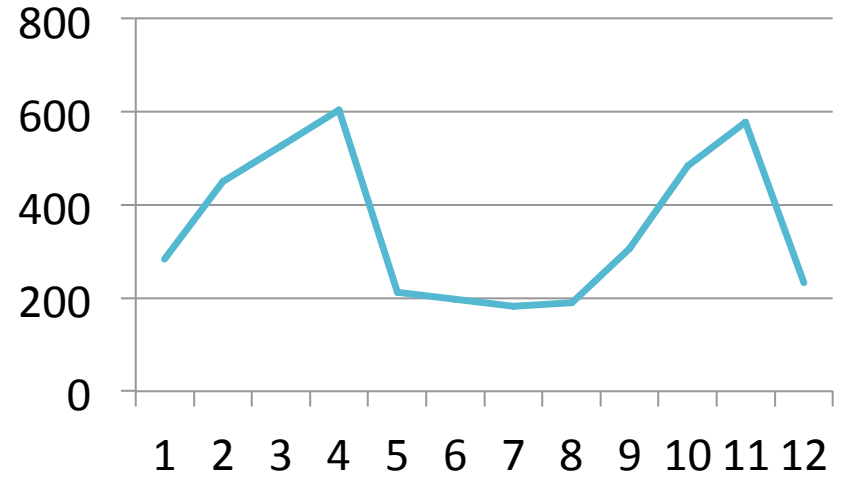
2012



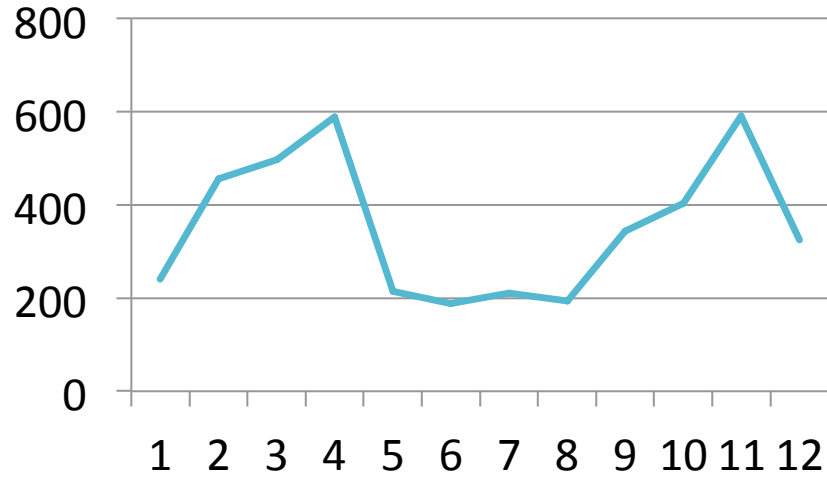
2013



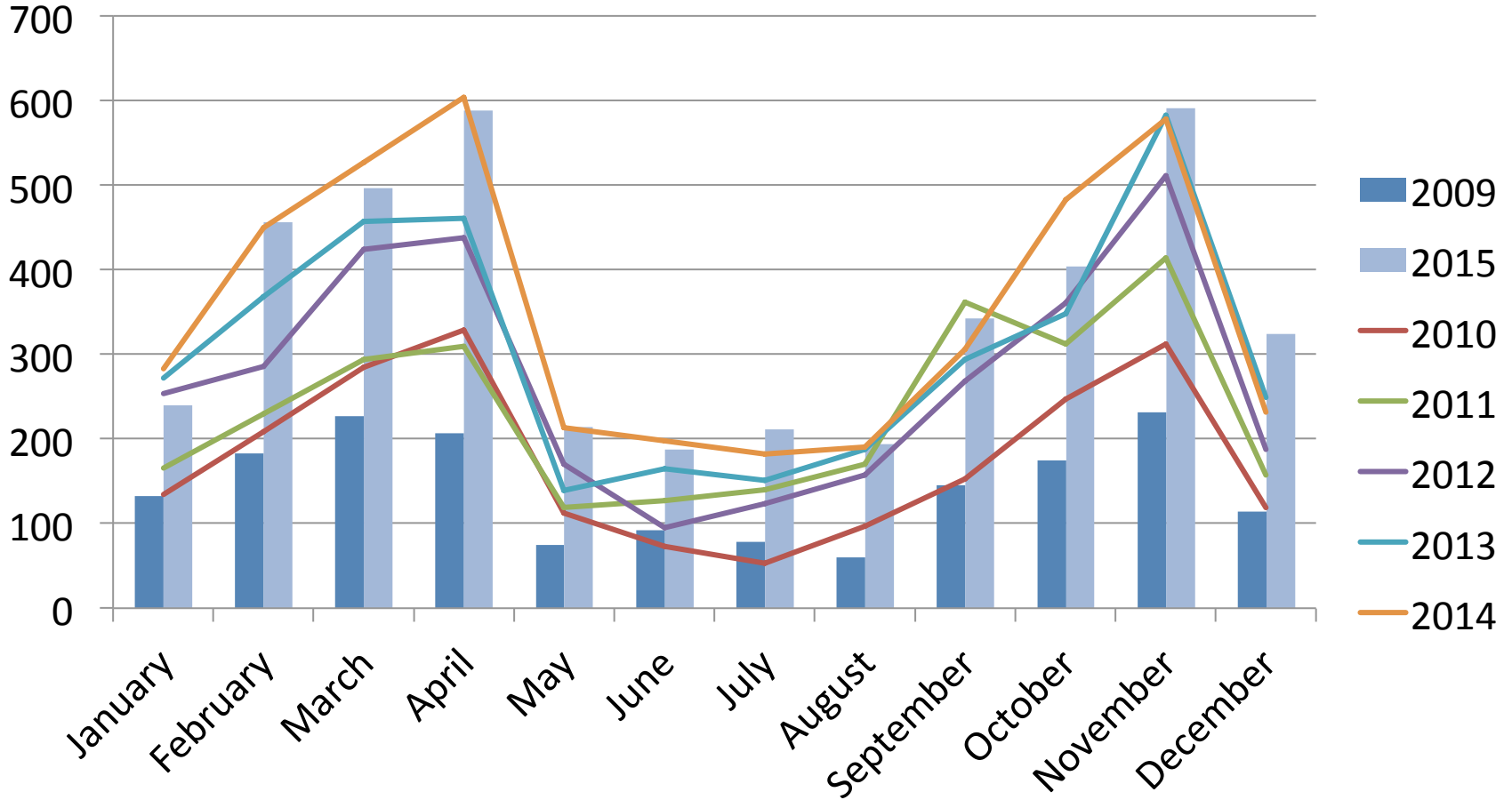
2014



2015



A Mix of Bar Chart and Line Chart



Formatting Line Graphs

<https://www.coursera.org/learn/oral-communication/lecture/8NDyY/how-to-show-complex-data> - Oral Communication for Engineering Leaders on Coursera

- *Beata Krupa, Lecturer in Professional and Engineering Communication, Rice Center for Engineering Leadership.*

More Line Chart Design Examples

- <http://www.perceptualedge.com/example6.php>
- <http://www.perceptualedge.com/example3.php>
 - Stephen Few, <http://www.perceptualedge.com/>

III. Articulate the conclusions

- “Data Visualization is not really about data. It is about **the meaning of data.**”
- “To communicate your data effectively you first need **to articulate the conclusions** you want your audience to adopt.”
 - Nancy Duarte. Slideology 2008

Design Guidelines for Data Visualization

- Overview
- Zoom and Filter
- Details on demand

- *The Visual Information-Seeking Mantra [Shneiderman, 1996]*

Design Guidelines for Data Visualization

- Design the view
- Design the control panel to allow users to pull up with what they want

Be complaint with Color Universal Design (CUG)

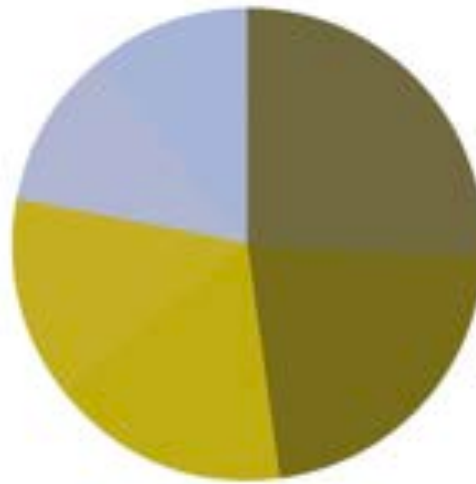
Make sure your infographic is complaint with Color Universal Design (CUG), which means **the graphical information is conveyed accurately to people with various types of color vision, including people with color blindness.**

Adjust hue or color brightness

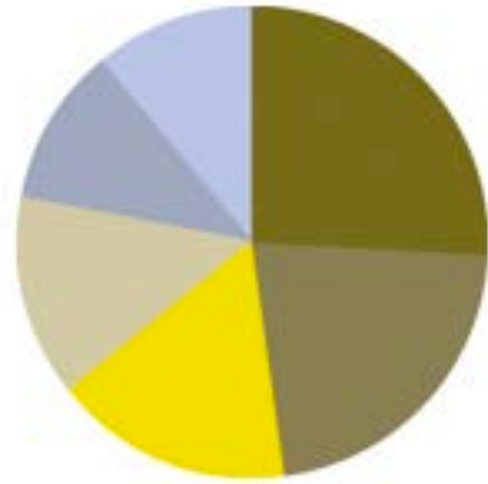
to make color-blind friendly color schemes



A



B

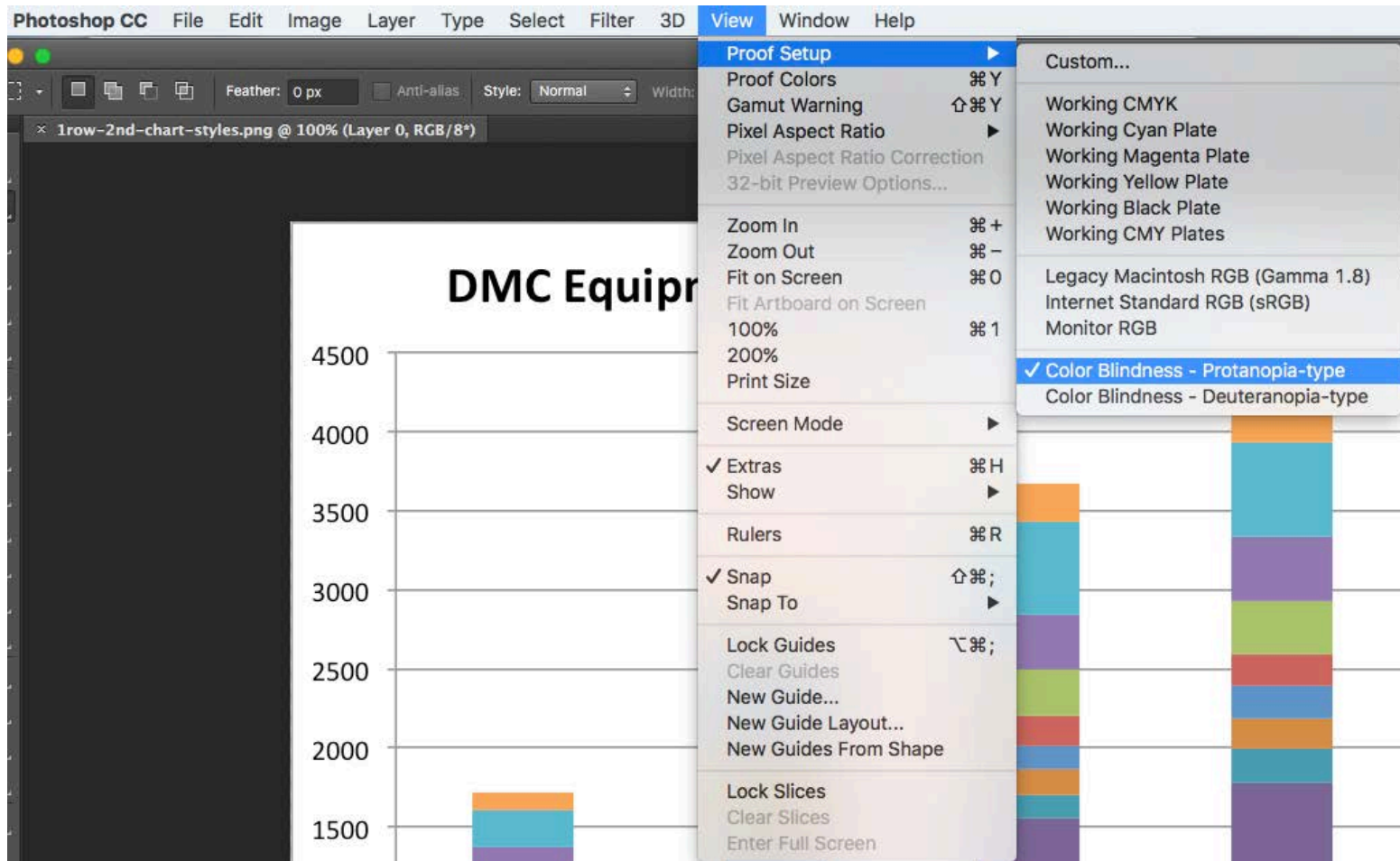


C

Adjusting design for color blindness

A. Original image B. Color-blind proof C. Optimized design

Use Photoshop/Illustrator to Proof Colors



Choice of colors for color-blind readers

- Tips from Edward Tufte website

http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0000HT

Set of colors that is unambiguous both to colorblinds and non-colorblinds

Original	Simulation				Hue	for Photoshop, Illustrator, Freehand, etc.		for Word, Power Point, Canvas, etc.
	Protan	Deutan	Tritan			C,M,Y,K (%)	R,G,B (0-255)	R,G,B (%)
1				Black	-°	(0,0,0,100)	(0,0,0)	(0,0,0)
2				Orange	41°	(0,50,100,0)	(230,159,0)	(90,60,0)
3				Sky Blue	202°	(80,0,0,0)	(86,180,233)	(35,70,90)
4				bluish Green	164°	(97,0,75,0)	(0,158,115)	(0,60,50)
5				Yellow	56°	(10,5,90,0)	(240,228,66)	(95,90,25)
6				Blue	202°	(100,50,0,0)	(0,114,178)	(0,45,70)
7				Vermilion	27°	(0,80,100,0)	(213,94,0)	(80,40,0)
8				reddish Purple	326°	(10,70,0,0)	(204,121,167)	(80,60,70)

Fig. 16 Colorblind barrier-free color pallet

Use Color Brewer as a Reference to Create Color-blind Friendly Color Scheme

Number of data classes: 8

Nature of your data:
 sequential diverging qualitative

Pick a color scheme:

Only show:
 colorblind safe
 print friendly
 photocopy safe

Context:
 roads
 cities
 borders

Background:
 solid color
 terrain

color transparency

8-class BrBG

140,81,10
191,129,45
223,194,125
246,232,195
199,234,229
128,205,193
53,151,143
1,102,94

EXPORT

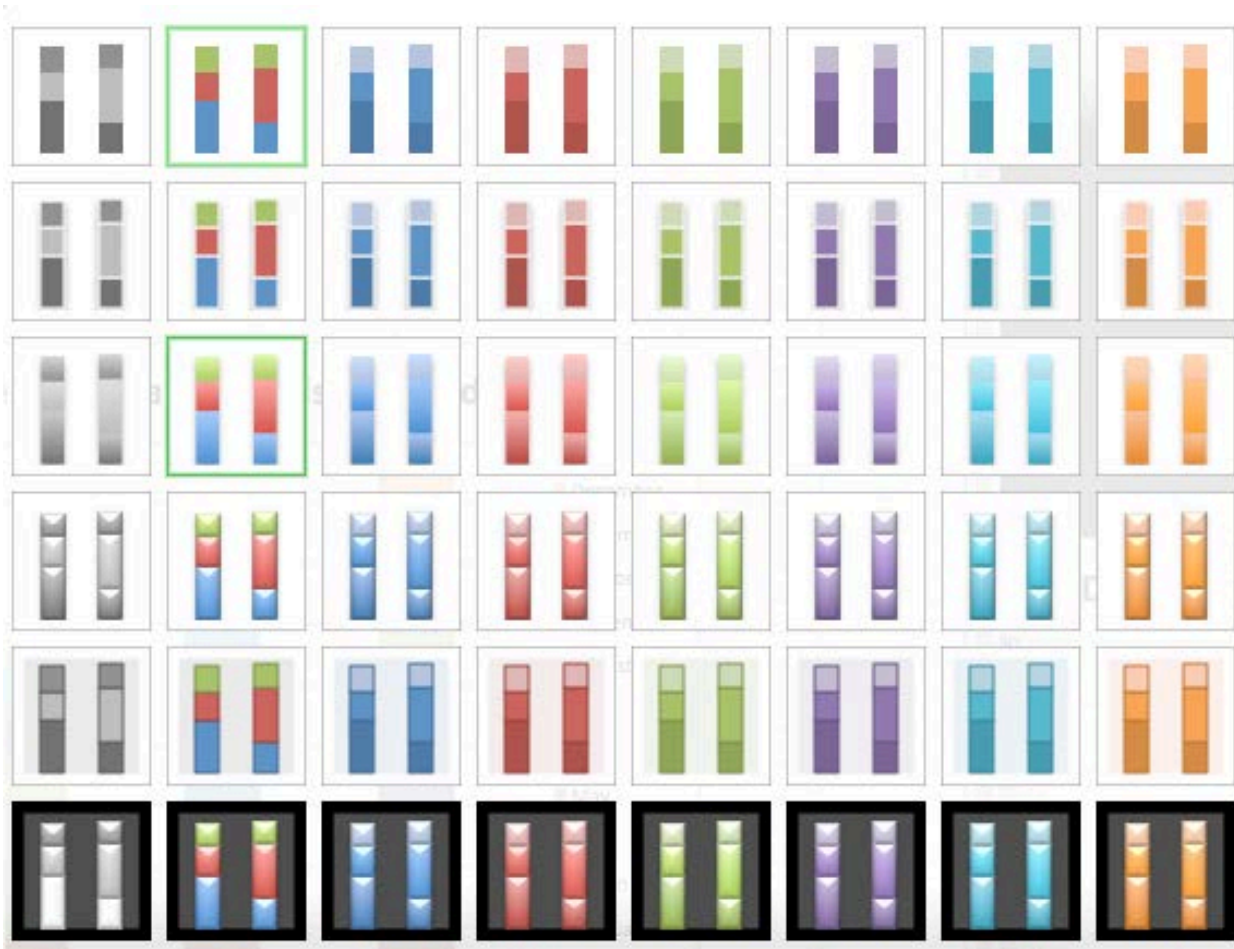
RGB

© Cynthia Brewer, Mark Harrower and The Pennsylvania State University
Support
Back to Flash version
Back to ColorBrewer 1.0

axismaps

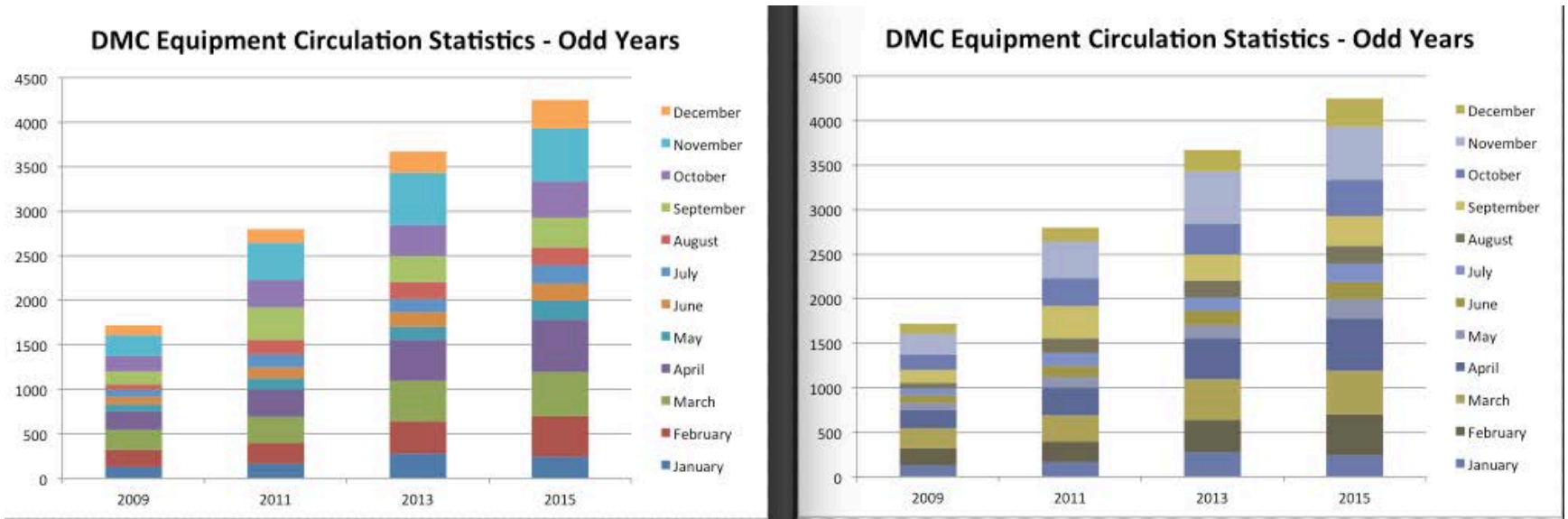
<http://colorbrewer2.org/> – Color Advice for Cartography

Do Excel's built-in chart styles pass color-blind test?



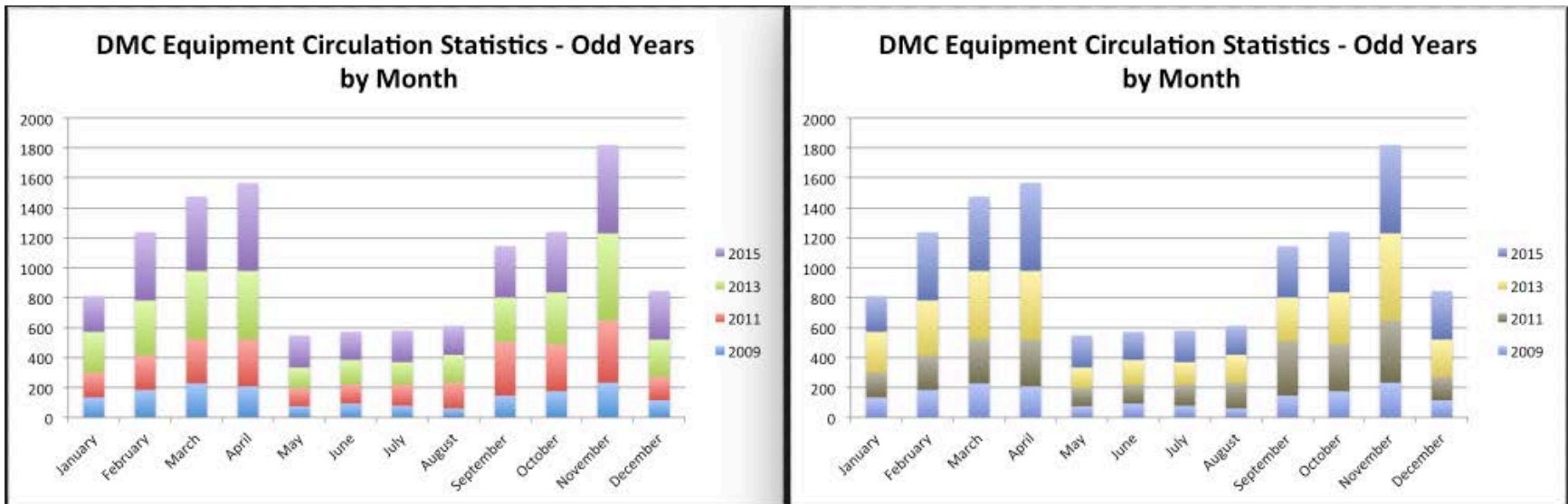
Left: Excel chart style
– 1st row 2nd one

Right: after turning on color-blindness tool in Photoshop



Left: Excel chart style
– 3rd row 2nd one

Right: after turning on color-blindness tool in Photoshop



My finding: It is safe to use Excel's built-in chart styles to create color-blind friendly color schemes!

Data Sources

- **data.gov** - The home of the U.S. Government's open data.
<http://www.data.gov/>.
- FactBrowser
<http://www.factbrowser.com/>
- Google Public Data
<http://www.google.com/publicdata/directory>
- Wolfram Alpha
<http://www.wolframalpha.com/>
- Wikipedia
https://en.wikipedia.org/wiki/Main_Page
- Tableau Sample Data Sets - A changing sample of datasets for use in teaching and learning.
https://public.tableau.com/s/resources?qt-overview_resources=1 - qt-overview_resources

Some Datasets for Teaching and Learning

- Awesome Public Datasets - This curated list of datasets is arranged by discipline; the majority of the datasets are free. <https://github.com/caesar0301/awesome-public-datasets/blob/master/README.rst>
- FedStats - This site provides access to the full range of official statistical information produced by the U.S. Government without having to know in advance which Federal agency produces which particular statistic. <https://fedstats.sites.usa.gov/>
- Raleigh Open Data - Data portal containing City of Raleigh datasets with API access in a variety of categories based on services that the City manages such as crime, building construction, road closures.
- NOAA Storm Events Database - The Storm Events Database contains records documenting significant weather phenomena and meteorological events. <https://www.ncdc.noaa.gov/stormevents/>
- CORGIS Datasets Project - Real-world datasets for subjects such as politics, education, literature, and construction. <https://think.cs.vt.edu/corgis/>

List is from NCSU libraries website:

<https://www.lib.ncsu.edu/teaching-and-learning-datasets>

On Campus Resources

- [Rice Visualization Center](#)
- [Kelly Center for Government Information, Data, and Geospatial Services](#)
- [GIS/Data Center](#)

Summary

- Data visualization helps us **see insights of data intuitively**.
- **Data Visualization Guidelines**
 - Start with a question
 - Choose the right type of graph/chart/map
 - Articulate the conclusion
- **Use color schemes** that are color-blind friendly