

Introduction to Data Visualization and Infographics

Jane Zhao

Digital Media Commons

Fondren Library

Photo: Qiwei Li

Student survey, Spring 2015

Undergraduate

Graduate

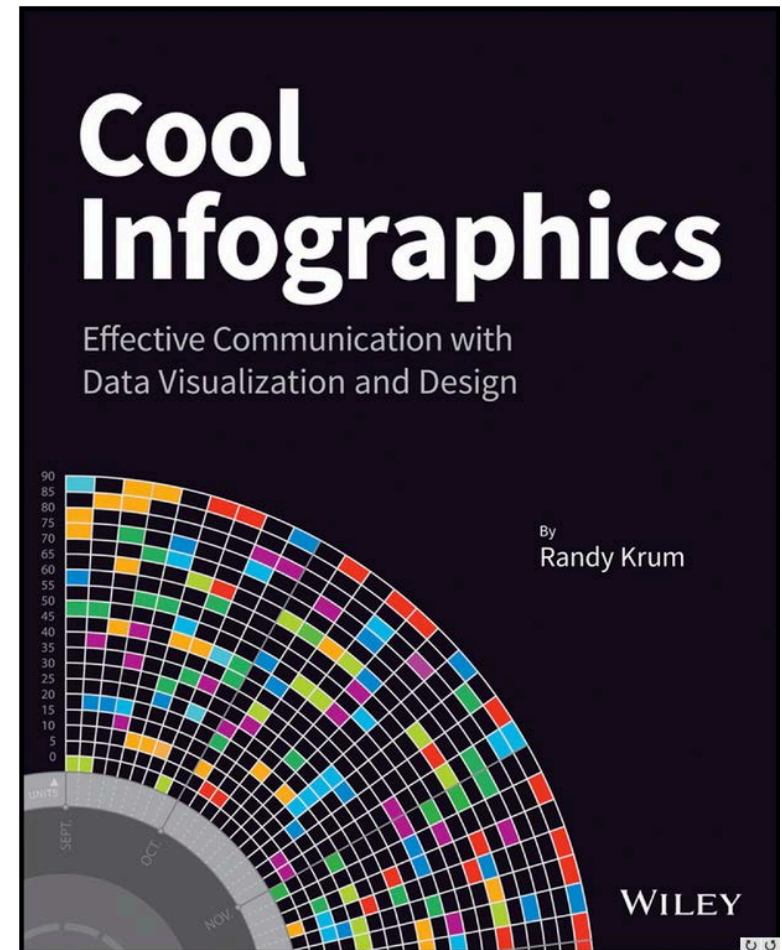
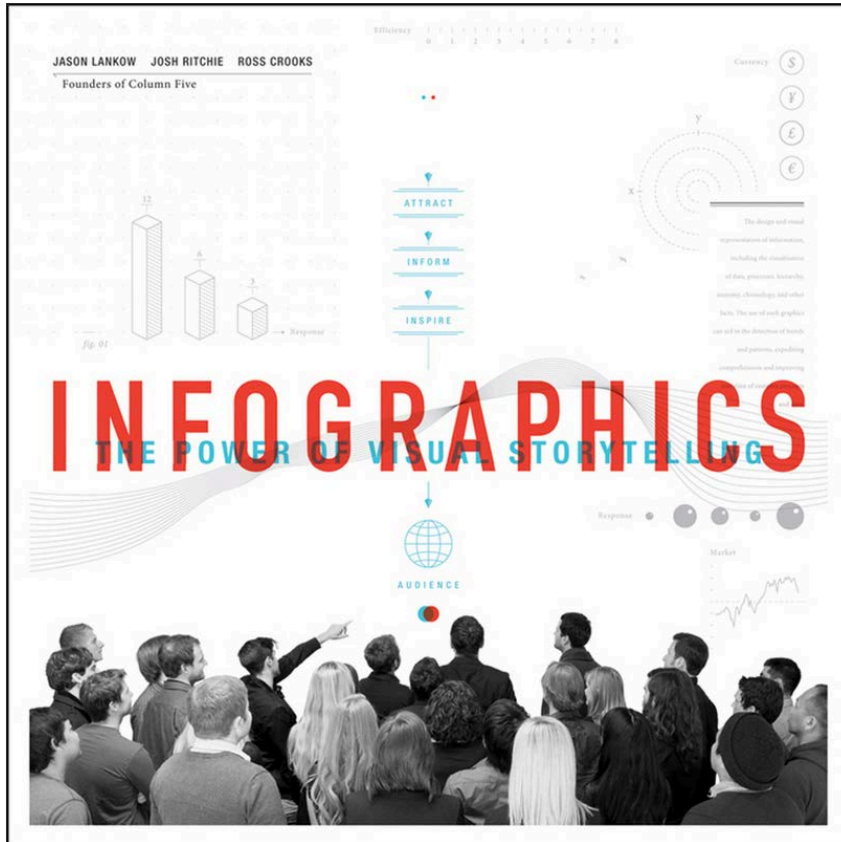
9. Would you be interested in taking a library short course on:

#	Answer	Response	%
1	Zotero	162	21%
2	Mendeley	82	10%
3	EndNote	145	18%
4	GIS	223	28%
5	Visualizing data	270	34%
6	Creating infographics	292	37%
7	Digital storytelling	181	23%
8	Library research methods	232	29%
9	Navigating the library website	115	15%
10	Specific database(s) - (please specify)	16	2%
11	Other (please specify)	19	2%

9. Would you be interested in taking a library short course on:

#	Answer	Response	%
1	Zotero	178	24%
2	Mendeley	168	23%
3	EndNote	245	34%
4	GIS	166	23%
5	Visualizing data	302	41%
6	Creating infographics	183	25%
7	Digital storytelling	132	18%
8	Library research methods	166	23%
9	Navigating the library website	85	12%
10	Specific database(s) - (please specify)	11	2%
11	Other (please specify)	21	3%

Books used



Objectives

- **Learn** the basic concepts of data visualization and infographics as well as the best practices of information design
- **Be aware** the handy tools for creating Infographics and Data Visualization

Outline

- **What** is Infographics? What is Data Visualization?
- **Why** do they work?
- **What** makes a good Infographic?
- **Information** design best practices.
- **Tools** for creating Infographics and Data Visualization.
- **Creating** a simple graphics with PowerPoint
- **Visualizing** a small set of data with Excel and Google Chart

INFOGRAPHICS AND DATA VISULIZATION



Check out some
examples...

What is Infographics?



An Example of Infographics



Discover



Learn



Create



Publish

Your projects, our passion!

An Example of Infographics

1ST FLOOR CONSTRUCTION



An Example of Infographics

“Life of **Outsiders** in Rio de Janeiro during the Time of the King”

Ariel Guerrero-Stewart, Ericka Howard, Tracey Franklin
History 251 Rice University
Spring 2012

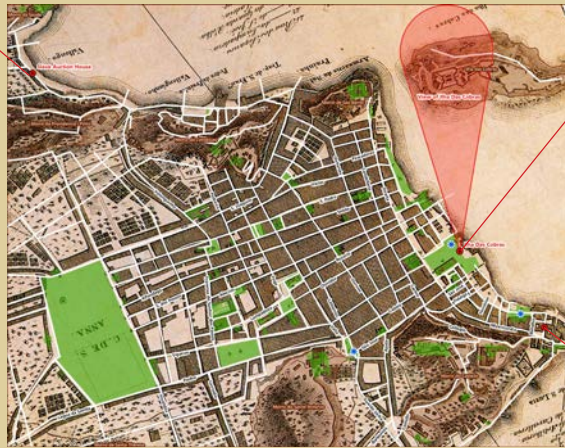
In the nineteenth century there was a mainstream Brazilian society and then those who were outsiders to it. This map shows some of the places where these outcasts, including gypsies, prisoners, and freed slaves would have visited during their time in Rio. While other groups were confined in one specific area these people were distributed throughout the city.



Debret, Three female gypsies outside a slave auction house



Steinmann, Ilha das Cobras, an island near the coastland of Rio



Debret, Freed black slaves running a barbershop

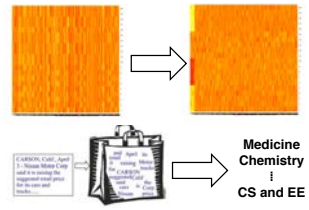


Debret, Prisoner begging for alms on the eve of Pentecost

An Example of Infographics

INTRODUCTIONS

- An explosion of data for which the dimension p is considerably larger than the sample size n , i.e., $p \gg n$;
- A challenge to uncover the group structure of the observations and to determine the discriminating variables;
- A lot of well studies on continuous and Gaussian-distributed large-scale data, e.g., DNA microarray;
- A call for Bayesian method on non-negative count data, e.g., next-generation sequencing (RNA-Seq) and bag-of-word data;
- Difficulties: the number of groups, normalization, variability modeling.



STATISTICAL MODELS AND MCMC ALGORITHMS

Data and Parameter Specification

- **Observable Data**
 - X is a set of n p -dimensional observations from K populations;
 - Each element $x_{ij} \in \mathbb{Z}^+$ is a nonnegative count number.

$$X_{n \times p} = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{pmatrix}$$

- **Parameters for Clustering Observations**
 - n samples are from a mixture of K Poisson distributions:

$$f(X_i | W, \theta) = \sum_{k=1}^K w_k f(X_i | \theta_k);$$
 - The size of each component $(n_1 \ n_2 \ \dots \ n_k)$ follows a multinomial distribution with parameter n and $(w_1 \ w_2 \ \dots \ w_k)$ with conjugate prior $(\alpha \ \alpha \ \dots \ \alpha)$;
 - A latent n -vector is introduced to identify the cluster, where $z_i = k$ indicates i -th observation belongs to k -th component:

$$Z = (z_1 \ z_2 \ \dots \ z_n).$$

- **Parameters for Identifying Discriminating Variables**
 - Not all the variables provide information about group structure and some even obscure the recovery of the true structure;
 - A latent p -vector is introduced to identify the most discriminating variables, where $\gamma_j = 1$ indicates j -th variable is informative:

$$\Gamma = (\gamma_1 \ \gamma_2 \ \dots \ \gamma_p);$$
 - Assume γ_j s are independent Bernoulli random variables with parameter ω , that is, $|\Gamma| = \sum_{j=1}^p \gamma_j \sim \text{Binomial}(p, \omega)$.

- **Parameters for Modeling Heterogeneity**
 - The variation in the number of counts per sample is very high, e.g., different RNA samples may be sequenced to different depths.
 - A n -vector is introduced to model the unobserved heterogeneity with prior $\text{Gamma}(1/\sigma^2, 1/\sigma^2)$:

$$S = (s_1 \ s_2 \ \dots \ s_n).$$

Hierarchical Framework

- **Data Likelihood**
 - We assume

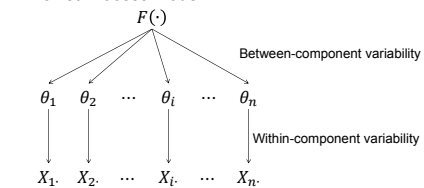
$$x_{ij} \sim \begin{cases} \text{Poisson}(s_i \theta_k) & \text{if } z_i = k, \gamma_j = 1 \\ \text{Poisson}(s_i \theta_0) & \text{if } \gamma_j = 0 \end{cases}$$
 - Data likelihood of each observation:

$$f(X_i | z_i = k, s_i, \Gamma, \theta_k, \theta_0) = \prod_{j=1}^p e^{-s_i d_{ij}(\Gamma)} \theta_0^{h_{i0}} e^{-s_i d_{ik}(\Gamma)} \theta_k^{h_{ik}};$$
 - Data likelihood of each variable:

$$f(X_j | Z, S, \gamma_j, \theta, \theta_0) = \begin{cases} \prod_{i=1}^n s_i^{x_{ij}} e^{-|S| \theta_0} \theta_0^{v_j} & \text{if } \gamma_j = 1 \\ \prod_{i=1}^n s_i^{x_{ij}} \prod_{k=1}^K e^{-|S| \theta_k} \theta_k^{v_{jk}} & \text{if } \gamma_j = 0 \end{cases}$$
 - $h_{i0} = \sum_{j: \gamma_j=0} x_{ij}$, $h_{ik} = \sum_{j: \gamma_j=1} x_{ij}$, $v_j = \sum_{i=1}^n x_{ij}$, $v_{jk} = \sum_{i: z_i=k} x_{ij}$;
 - Full data likelihood:

$$f(X | Z, S, \Gamma, \theta, \theta_0) = \prod_{i=1}^n f(X_i | z_i = k, s_i, \Gamma, \theta_k, \theta_0).$$

Dirichlet Process Model



- **Motivation:**
 - The number of component K is unknown;
 - The prior distribution of θ_k is unknown;
 - Each θ_k shares a common but completely unknown $F(\cdot)$;
 - The prior of $F(\cdot)$ is $DP(F_0, \alpha)$, where α is a weighting factor that characterizes how close $F(\cdot)$ is to the shape of F_0 :

$$F_0 \sim \text{Gamma}(a, b);$$
 - Integrating over $F(\cdot)$, we obtain

$$\theta_k | \theta_{-k} \sim \frac{1}{K-1+\alpha} \sum_{m=1, m \neq k} \delta(d_m) + \frac{\alpha}{K-1+\alpha} F_0;$$
 - If θ only takes on K distinctive values, then we have a mixture of the smooth measure F_0 and the K point masses;
 - Any observations θ_k and θ_m that have the same value are defined as being in the same cluster.

Posterior and Full Conditionals

- **Posterior:**

$$\pi(Z, S, \Gamma, \theta, \theta_0 | X) \propto f(X | Z, S, \Gamma, \theta, \theta_0) \pi(Z) \pi(S) \pi(\Gamma) \pi(\theta) \pi(\theta_0);$$
- **Full conditionals**

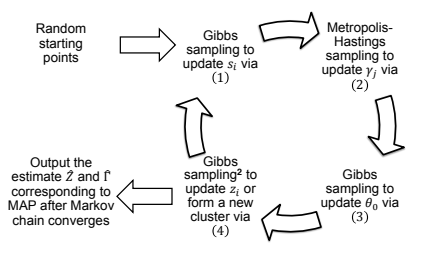
$$s_i | z_i, \Gamma, \theta_k, \theta_0, X_i \sim \text{Gamma}\left(\frac{1}{\sigma^2} + h_{i0} \frac{1}{\sigma^2} + \theta_0(p - |\Gamma|) + \theta_k |\Gamma|\right); \quad (1)$$

$$\pi(\gamma_j | Z, S, \theta, \theta_0, X_j) \propto f(X_j | Z, S, \gamma_j, \theta, \theta_0) \pi(\gamma_j); \quad (2)$$

$$\theta_0 | z_i, \Gamma, \theta_k, \theta_0, X_i \sim \text{Gamma}(a + \sum_{i: z_i=k} h_{i0}, b + |\Gamma| |S_k|); \quad (3)$$

$$\pi(z_i = k | z_{-i}, s_i, \Gamma, \theta, \theta_0, X_i) \propto f(X_i | z_i = k, s_i, \Gamma, \theta_k, \theta_0) \pi(\theta_k | \theta_{-k}); \quad (4)$$

MCMC Algorithms

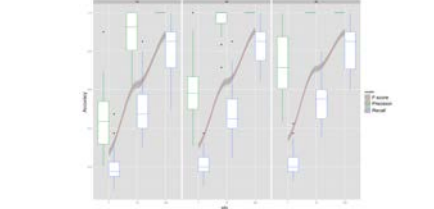


RESULTS AND DISCUSSION

Evaluation with Synthetic Data

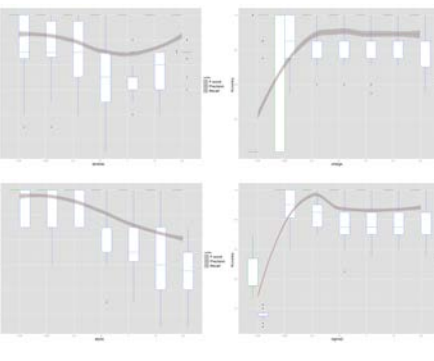
- **Synthetic Data Generating**
 - 20 observations and 1000 variables, of which $|\gamma|$ are discriminant:

$$x_{ij} | \gamma_j = 1 \sim I_{(1 \leq i \leq 5)} \text{Poisson}(s_i \theta_1) + I_{(6 \leq i \leq 10)} \text{Poisson}(s_i \theta_2) + I_{(11 \leq i \leq 15)} \text{Poisson}(s_i \theta_3) + I_{(16 \leq i \leq 20)} \text{Poisson}(s_i \theta_4);$$
 - $\theta_0 = 10$ and to model overdispersion, we set $\theta_k \sim \text{Gamma}(\psi, \psi/d_k)$, $d_1 = 80$, $d_2 = 40$, $d_3 = 60$, and $d_4 = 100$.
- **Statistical Performance**
 - Precision = $\frac{\# \text{ of true pairwise relationship that are correctly estimated}}{\# \text{ of all pairwise in estimated } Z}$;
 - Recall = $\frac{\# \text{ of true pairwise relationship that are correctly estimated}}{\# \text{ of all pairwise in true } Z}$;
 - F score = $\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$;
 - Program settings: $a = 0.01$, $b = a/\bar{X}$, $c = 1/\sigma^2$, $\alpha = 1$, $\omega = 0.01$.



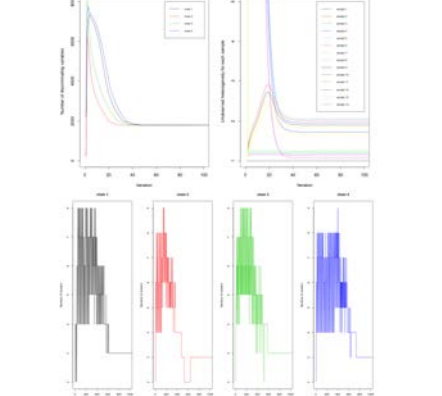
Sensitivity Analysis

- We set $|\gamma| = 40$ and $\psi = 10$;
- Program settings: $a = 0.01$, $b = a/\bar{X}$, $c = 1/\sigma^2$, $\alpha = 1$, $\omega = 0.01$.



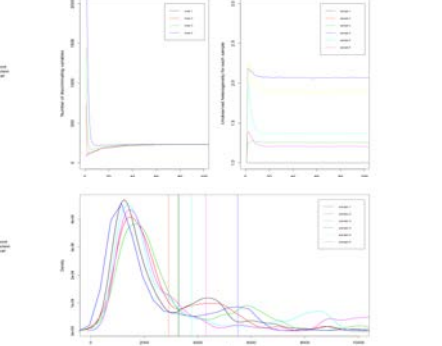
Experiment on Real Data

- **Liver and Kidney RNA-Seq Data Set³**
 - 22,925 genes, 7 replicates from a liver sample and 7 replicates from a kidney sample, each from a single human male.



Yeast (*Saccharomyces cerevisiae*) RNA-Seq Data Set⁴

- 6,874 genes, 3 replicates from random hexamer (RH) library preparation protocols and 3 replicates from oligo (DT) ones.



CONCLUSION

- Proposed a fully Bayesian method for simultaneously clustering high-dimensional data and selecting the variables that best discriminate the different groups on Poisson model;
- Formulated the clustering problem in terms of Poisson mixture model via Dirichlet process with unknown K ;
- Evaluated the MCMC algorithms on both simulated and real data and provided recommendations for priors;
- To extent Poisson model to negative binomial model and to model variance shrinkage more elaborate.

Reference

- J. Li, D. M. ...
- R. M. Neal, ...
- J. Marion, ...
- U. Nagalaksh...

Definition of Infographics

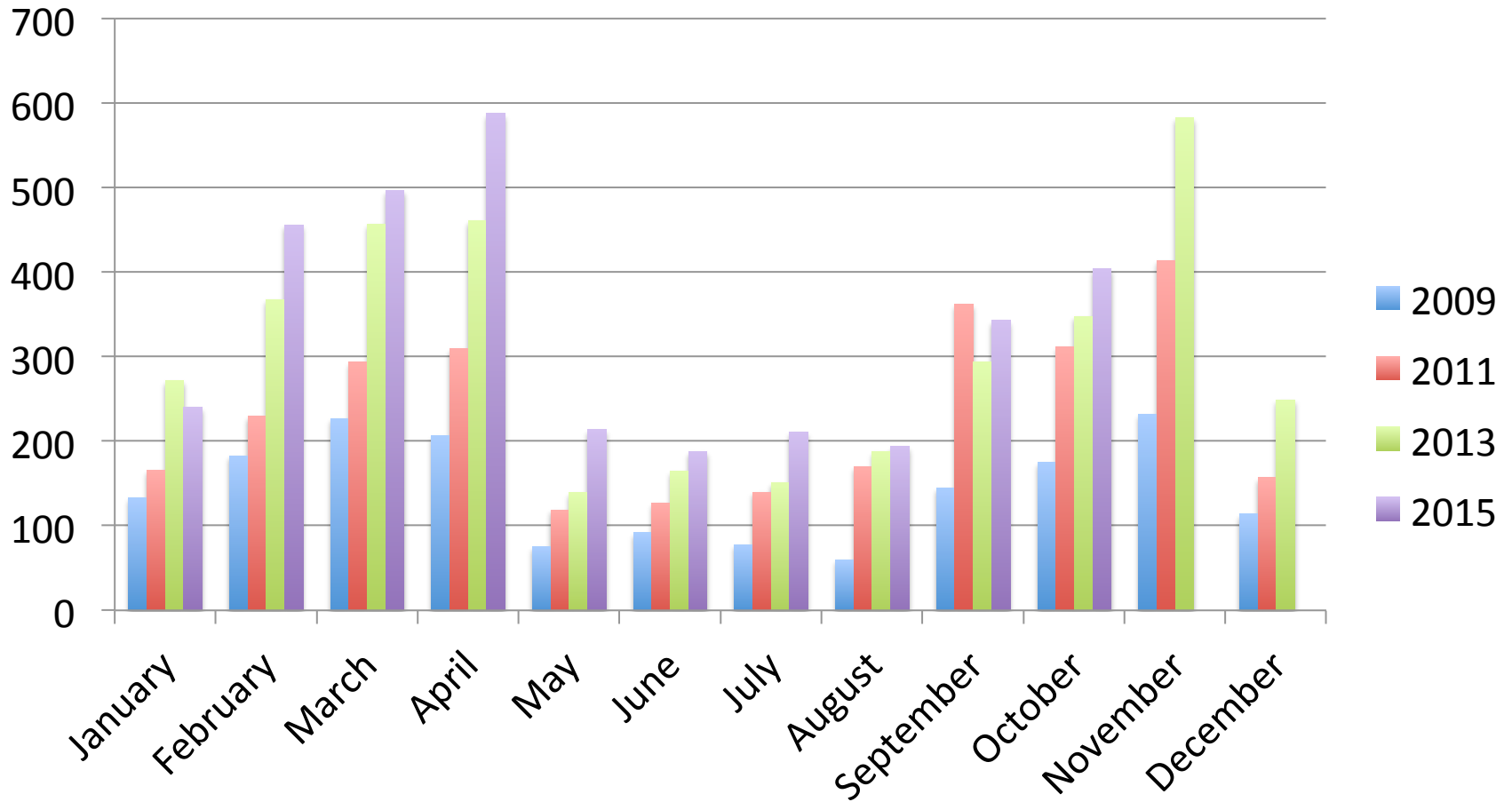
- Infographic is an abbreviation of "information graphic".
- It combines data visualizations, illustrations, text and images together into a format that tells a complete story (Krum, 2014, p. 6).
- It uses visual cues to communicate information.



Check out some
examples...

What is Data Visualization?

DMC Equipment Circulation Statistics



An Example of Data Visualization

Definition of Data Visualization

- Data visualization is a visual representation of data or the practice of visualizing data.
- Common forms include pie charts, bar graphs, line charts, and so forth.
- It is a powerful tool that designers often use to help tell their story visually in an infographic (Krum, 2014, p. 6).

Data Visualization is a Separate Design Element Used in the Design of Infographics.



RICE

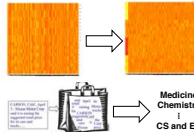
Bayesian Clustering and Variable Selection of High-Dimensional Count Data

Qiwei Li and Marina Vannucci

Department of Statistics, Rice University, Houston, Texas

INTRODUCTIONS

- An explosion of data for which the dimension p is considerably larger than the sample size n , i.e., $p \gg n$.
- A challenge to uncover the group structure of the observations and to determine the discriminating variables;
- A lot of well studies on continuous and Gaussian-distributed large-scale data, e.g., DNA microarray;
- A call for Bayesian method on non-negative count data, e.g., next-generation sequencing (RNA-Seq) and bag-of-word data;
- Difficulties: the number of groups, normalization, variability modeling.



STATISTICAL MODELS AND MCMC ALGORITHMS

Data and Parameter Specification

- Observable Data
 - X is a set of n p -dimensional observations from K populations;
 - Each element $x_{ij} \in \mathbb{Z}^+$ is an nonnegative count number:

$$X_{n \times p} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}$$

Parameters for Clustering Observations

- n samples are from a mixture of K Poisson distributions:

$$f(X; \theta) = \sum_{k=1}^K w_k f(X; \theta_k);$$
- The size of each component ($\theta_1, \theta_2, \dots, \theta_K$) follows a multinomial distribution with parameter π and (π_1, w_1, \dots, w_K) with conjugate prior ($\alpha, \alpha, \dots, \alpha$);
- A latent π -vector is introduced to identify the cluster, where $\pi_i = k$ indicates i -th observation belongs to k -th component:

$$Z = (\pi_1, \pi_2, \dots, \pi_n).$$

Parameters for Identifying Discriminating Variables

- Not all the variables provide information about group structure and some even obscure the recovery of the true structure;
- A latent γ -vector is introduced to identify the most discriminating variables, where $\gamma_j = 1$ indicates j -th variable is informative:

$$\gamma = (\gamma_1, \gamma_2, \dots, \gamma_p).$$

Parameters for Modeling Heterogeneity

- The variation in the number of counts per sample is very high, e.g., different RNA samples may be sequenced to different depths;
- A ν -vector is introduced to model the unobserved heterogeneity with prior $\text{Gamma}(1/\sigma^2, 1/\sigma^2)$:

$$S = (\nu_1, \nu_2, \dots, \nu_n).$$

Hierarchical Framework

Data Likelihood

- We assume

$$x_{ij} \sim \begin{cases} \text{Poisson}(s_i \theta_{kj}) & \text{if } \pi_i = k, \gamma_j = 1 \\ \text{Poisson}(s_i \theta_{kj}) & \text{if } \gamma_j = 0 \end{cases}$$

Data likelihood of each observation:

$$f(x_{ij} | z_i, s_i, \Gamma, \theta_{kj}, \gamma_j) = \prod_{j=1}^p \frac{e^{-s_i \theta_{kj}} (s_i \theta_{kj})^{x_{ij}}}{x_{ij}!} \mathbb{1}_{\gamma_j = 1} + \prod_{j=1}^p \frac{e^{-s_i \theta_{kj}}}{x_{ij}!} \mathbb{1}_{\gamma_j = 0}$$

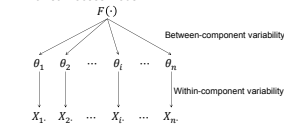
Data likelihood of each variable:

$$f(x_{ij} | z_i, s_i, \gamma_j, \theta_{kj}) = \begin{cases} \prod_{i=1}^n \frac{s_i^{x_{ij}} e^{-s_i \theta_{kj}}}{x_{ij}!} & \text{if } \gamma_j = 1 \\ \prod_{i=1}^n \frac{e^{-s_i \theta_{kj}}}{x_{ij}!} & \text{if } \gamma_j = 0 \end{cases}$$

Full data likelihood:

$$f(X | Z, S, \Gamma, \theta_{kj}) = \prod_{i=1}^n f(x_{ij} | z_i, s_i, \Gamma, \theta_{kj})$$

Dirichlet Process Model



- Motivation:
 - The number of component K is unknown;
 - The prior distribution of θ_k is unknown;
 - Each θ_k shares a common but completely unknown $F(\cdot)$;
 - The prior of $F(\cdot)$ is $DP(\xi, \alpha)$, where α is a weighting factor that characterizes how close $F(\cdot)$ is to the shape of F_0 :

$$F_0 = \text{Gamma}(a, b);$$

Integrating over $F(\cdot)$, we obtain

$$\theta_k | \theta_{-k} \sim \frac{1}{K-1 + \alpha} \sum_{j=1}^{K-1} \delta(d_{kj}) + \frac{\alpha}{K-1 + \alpha} F_0;$$

- If only takes on K distinctive values, then we have a mixture of the smooth measure F_0 and the K point masses;
- Any observations θ_k and θ_m that have the same value are defined as being in the same cluster;

Posterior and Full Conditionals

- Posterior:

$$\pi(Z, S, \Gamma, \theta_{kj} | X) \propto f(X | Z, S, \Gamma, \theta_{kj}) \pi(Z) \pi(S) \pi(\Gamma) \pi(\theta_{kj});$$
- Full conditionals

$$s_i | z_i, \Gamma, \theta_{kj}, X_i \sim \text{Gamma}\left(\frac{x_{ij}}{s_i} + \alpha, \frac{1}{s_i} + \theta_{kj}(s_i - 1) + \theta_{kj}\right); \quad (1)$$

$$\pi(\gamma_j | Z, S, \theta_{kj}, X_j) \propto f(X_j | Z, S, \gamma_j, \theta_{kj}) \pi(\gamma_j); \quad (2)$$

$$\theta_{kj} | z_i, \Gamma, \theta_{kj}, X_i \sim \text{Gamma}\left(\alpha + \sum_{i=1}^n \mathbb{1}_{z_i=k} x_{ij}, b + \Gamma \mathbb{1}[S_i]\right); \quad (3)$$

$$\pi(z_i = k | Z_{-i}, s_i, \Gamma, \theta_{kj}, X_i) \propto f(X_i | z_i = k, s_i, \Gamma, \theta_{kj}) \pi(\theta_{kj} | z_i = k); \quad (4)$$

MCMC Algorithms



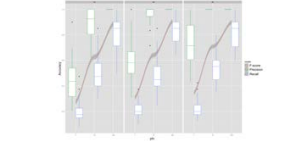
RESULTS AND DISCUSSION

Evaluation with Synthetic Data

- Synthetic Data Generating
 - 20 observations and 1000 variables, of which γ_j are discriminant:

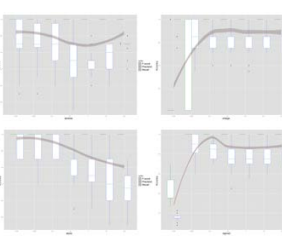
$$x_{ij} | \gamma_j = 1 \sim \text{Poisson}(s_i \theta_{kj}) + \text{Poisson}(s_i \theta_{kj}) + \text{Poisson}(s_i \theta_{kj}) + \text{Poisson}(s_i \theta_{kj});$$
 - $\theta_k = 10$ and to model overdispersion, we set

$$\theta_k \sim \text{Gamma}(\psi, \theta) / \theta_k, \psi = 30, \theta_k = 40, \theta_k = 60, \text{ and } \theta_k = 100.$$
- Statistical Performance
 - Precision = # of true pairwise relationship that are correctly estimated, # of all pairwise in estimated Z
 - Recall = # of true pairwise relationship that are correctly estimated, # of all pairwise in true Z
 - F score = $\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$
 - Program settings: $a = 0.01, b = a/K, c = 1/\sigma^2, \alpha = 1, \omega = 0.01.$



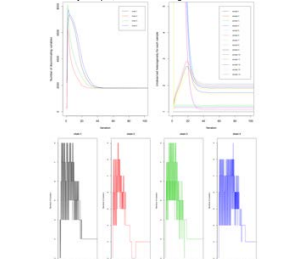
Sensitivity Analysis

- We set $\gamma_j = 10$ and $\theta_j = 10$;
- Program settings: $a = 0.01, b = a/K, c = 1/\sigma^2, \alpha = 1, \omega = 0.01.$



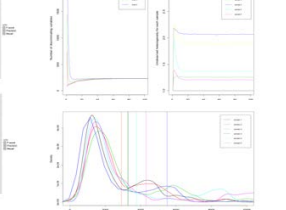
Experiment on Real Data

- Liver and Kidney RNA-Seq Data Set¹
 - 22,925 genes, 7 replicates from a liver sample and 7 replicates from a kidney sample, each from a single human male.



Yeast (Saccharomyces cerevisiae) RNA-Seq Data Set⁴

- 6,874 genes, 3 replicates from random hexamer (RH) library preparation protocols and 3 replicates from oligo (DT) ones.

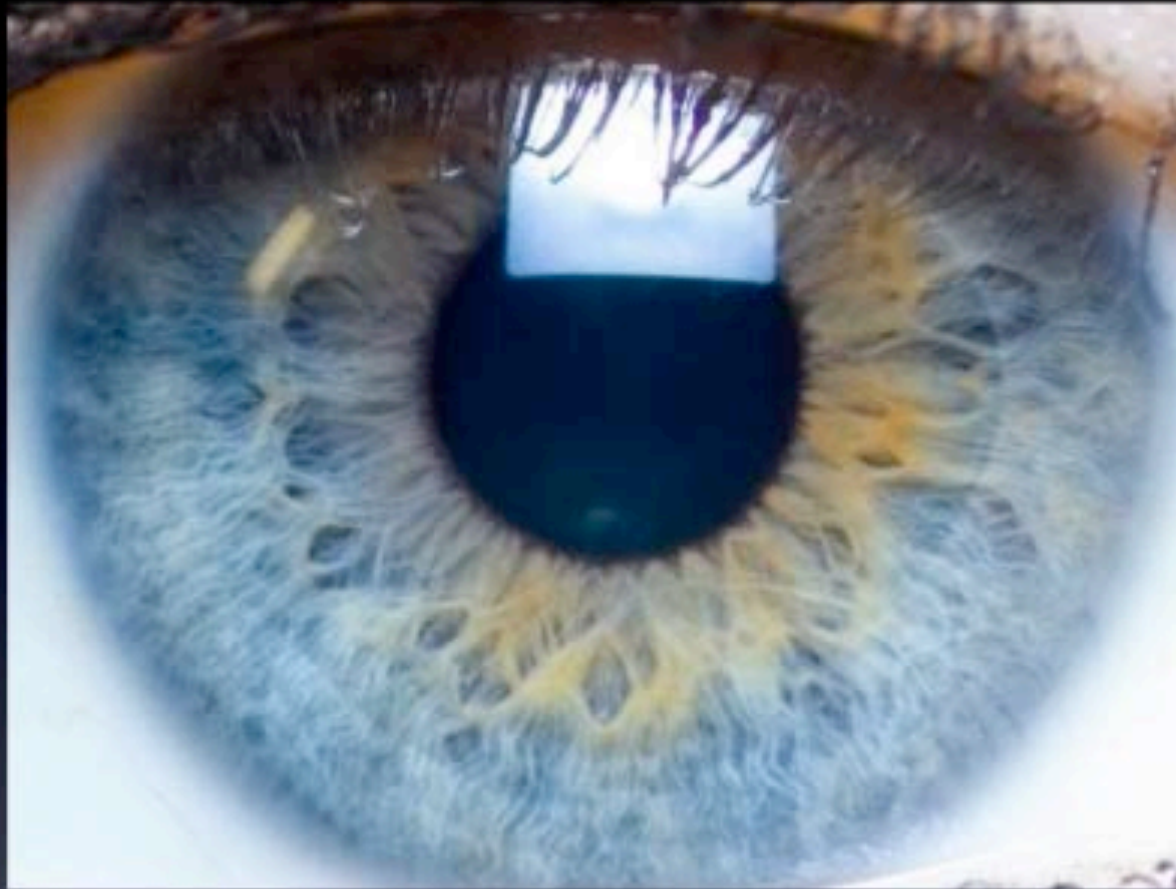


CONCLUSION

- Proposed a fully Bayesian method for simultaneously clustering high-dimensional data and selecting the variables that best discriminate the different groups on Poisson model;
- Formulated the clustering problem in terms of Poisson mixture model via Dirichlet process with unknown K ;
- Evaluated the MCMC algorithms on both simulated and real data and provided recommendations for priors;
- To extent Poisson model to negative binomial model and to model variance shrinkage more elaborate.


References:
 1. Li, D. M., Witton, I. M., Johnstone, and R. Tibshirani, "Normalization, Testing, and False Discovery Rate Estimation for RNA-Seqing Data," *Bioinformatics*, 2012, Volume 13, Issue 3, pp. 523-538.
 2. M. Neal, "Metropolis Chain Sampling Methods for Dirichlet Process Mixture Models," *Journal of Computational and Graphical Statistics*, 2000, Volume 9, Issue 2, pp. 249-265.
 3. M. Neale, C. Hanson, B. Hahn, B. Stephens, and Y. Geiss, "RNA-Seq: An Assessment of Technical Reproducibility and Comparison with Core Expression Array," *Genome Research*, 2008, Volume 18.
 4. Nagabathini, Z., Wang, K., Warren, C., Shou, D., Raha, M., Gerstein, and M. Snyder, "The Transcriptional Landscape of the Yeast Genome defined by RNA Sequencing," *Science*, Volume 302.

WHY INFOGRAPHICS WORK?



80% of the brain is dedicated to
visual processing

University of Rochester, 2004

A close-up photograph of a lioness looking directly at the camera through a dense field of tall, dry, golden-brown grass. The lioness's face is partially obscured by the blades of grass, and its eyes are focused on the viewer.

The human brain is a
pattern recognition machine

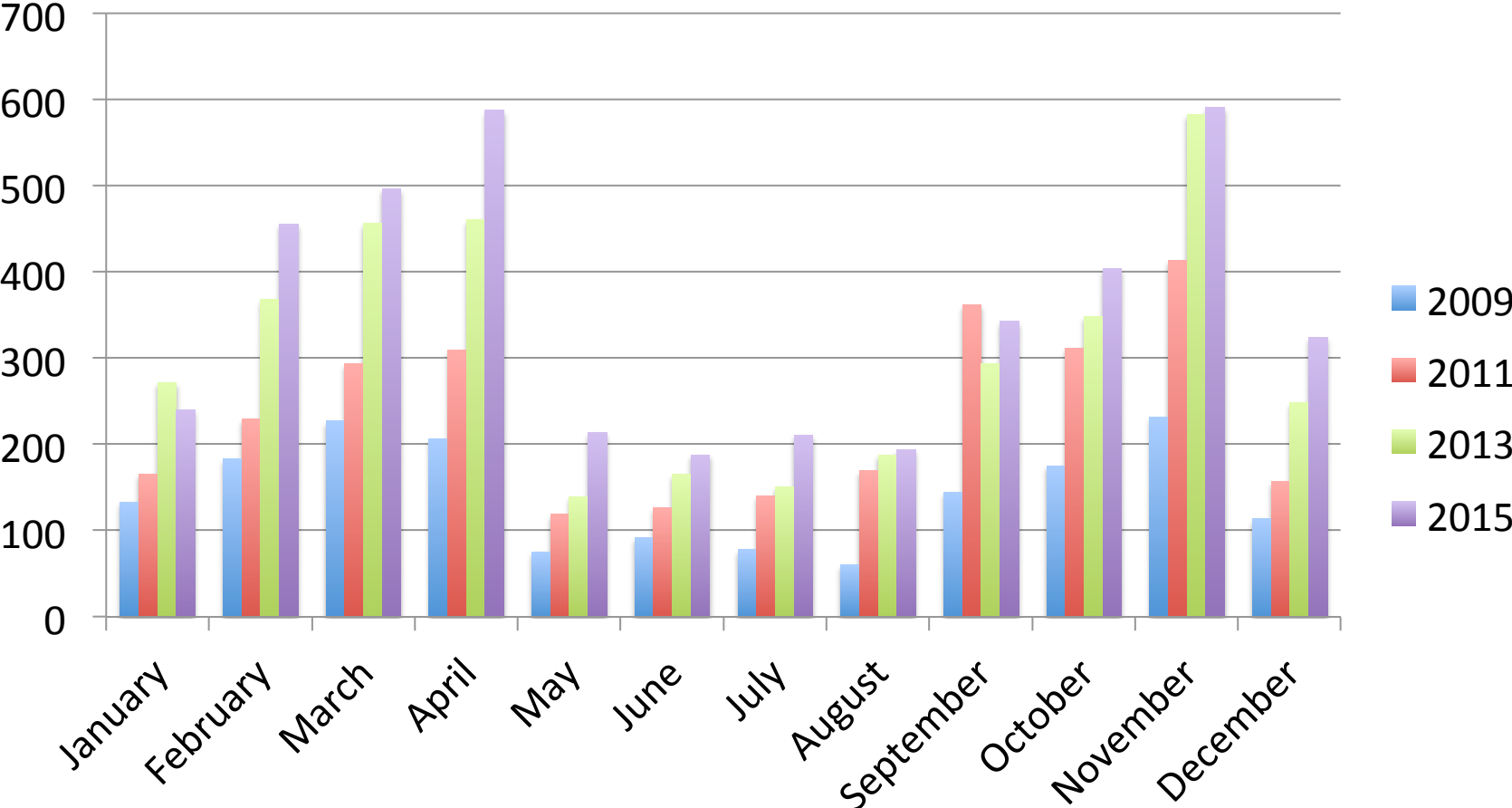
A Table of Data, Hard to See its Pattern and Trend.

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

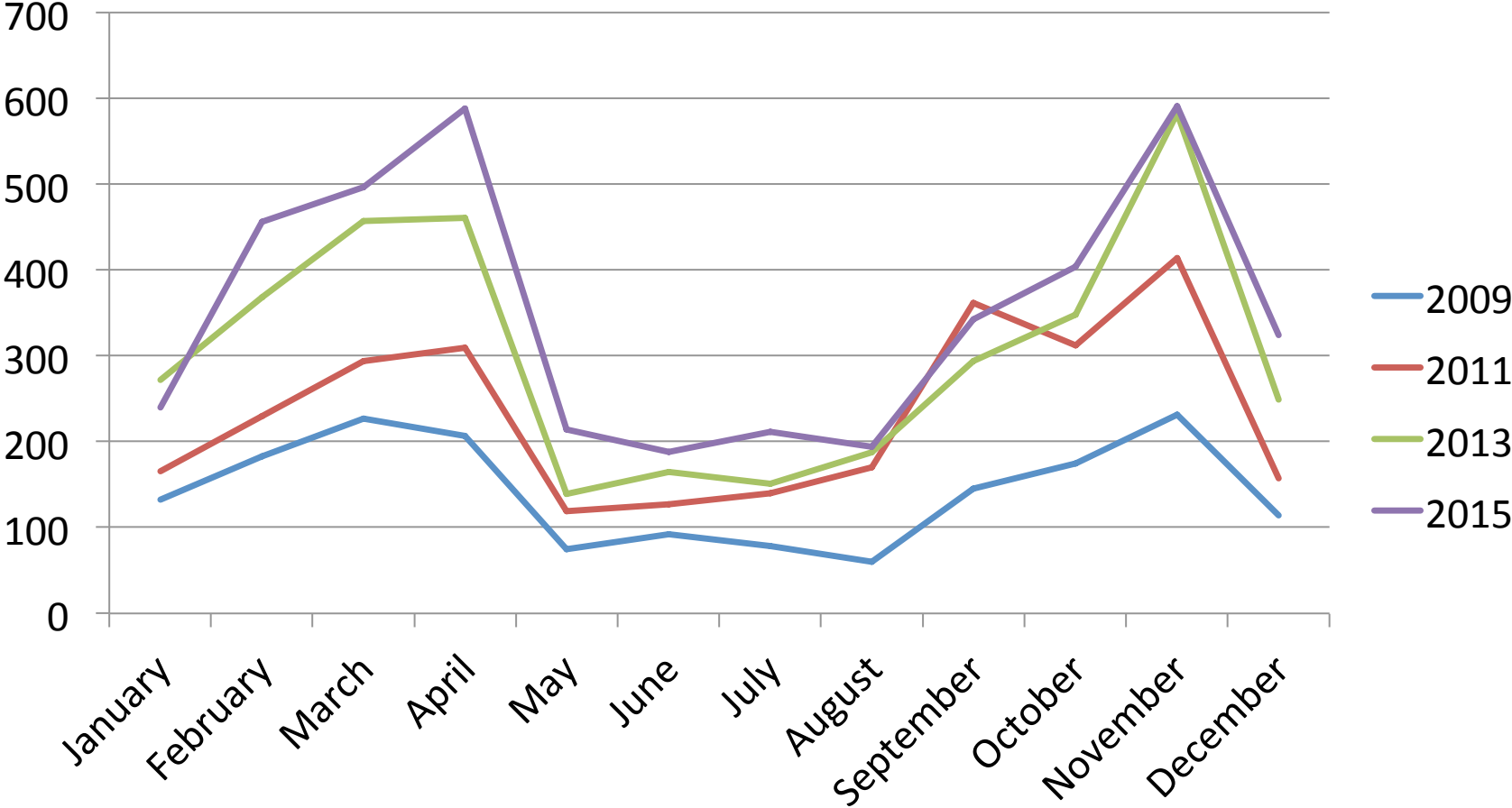
Convert the Data to a Bar Chart, Easy to See the Pattern.

DMC Equipment Circulation Statistics



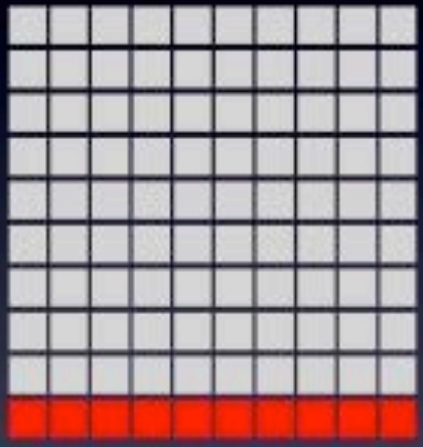
Convert the Data to a Line Chart, Easy to See the Trend.

DMC Equipment Circulation Statistics



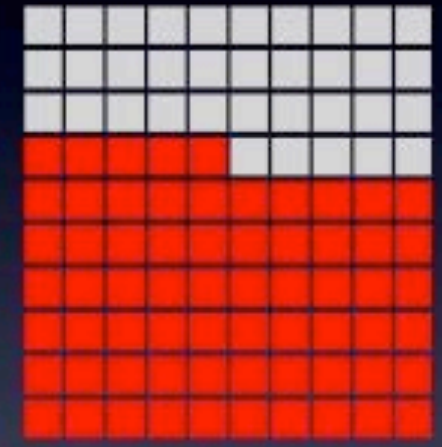
Picture Superiority Effect

Memory retention after 3 days



10%

Text or Audio Only



65%

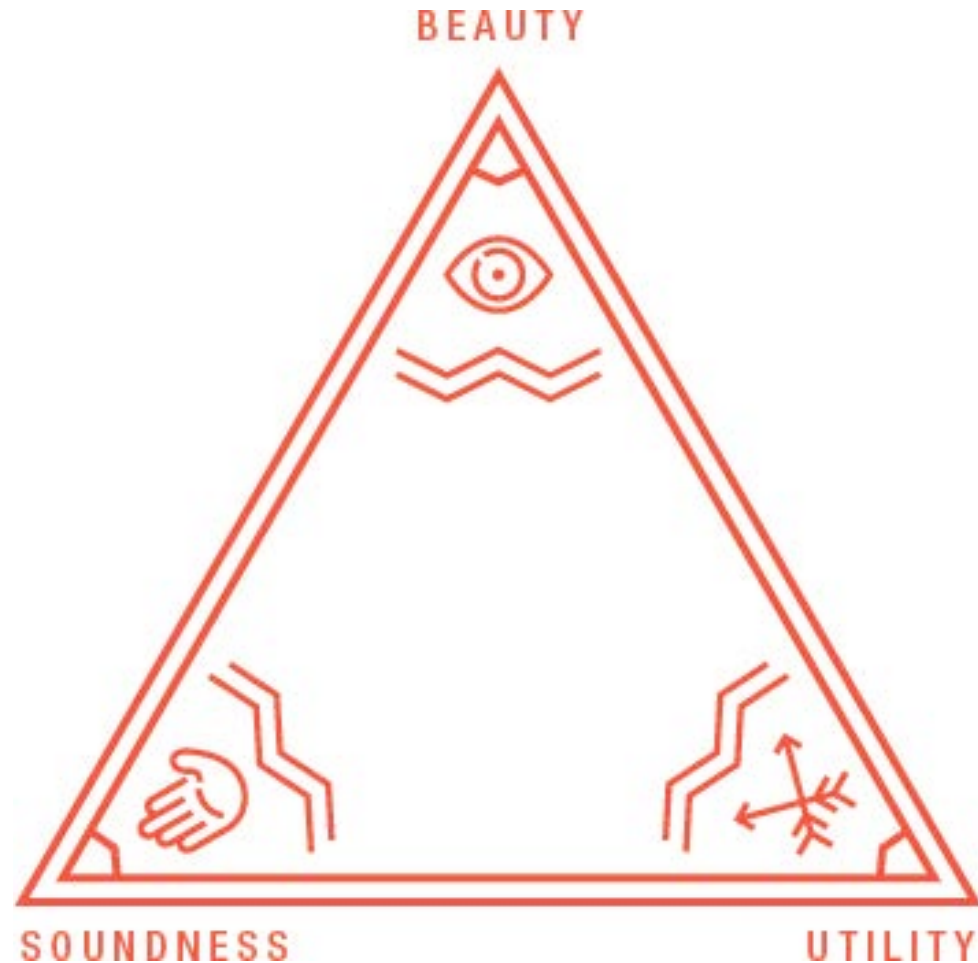
Text + Picture

“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

**WHAT MAKES A GOOD
INFOGRAPHIC?**

A good infographic has all three:



Lankow, J., Ritchie, J., & Crooks, R. (2012). *Infographics [electronic resource]: the power of visual storytelling*. ©2012. P198

INFORMATION DESIGN BEST PRACTICES

1

Keep it simple



“Simplicity means the achievement of maximum effect with minimum means.”

- Dr. Koichi Kawana – artist, designer, and architect

A Data Visualization is

DATA



SORTED



ARRANGED



PRESENTED
VISUALLY




Hot Butter Studio © 2012 www.hotbutterstudio.com @HOTBUTTERSTUDIO



PHOTOGRAPHY BY BRANDON ROSCEN PHOTOGRAPHY WWW.BRANDONROSCEN.COM #BRANDANROSCEN

2

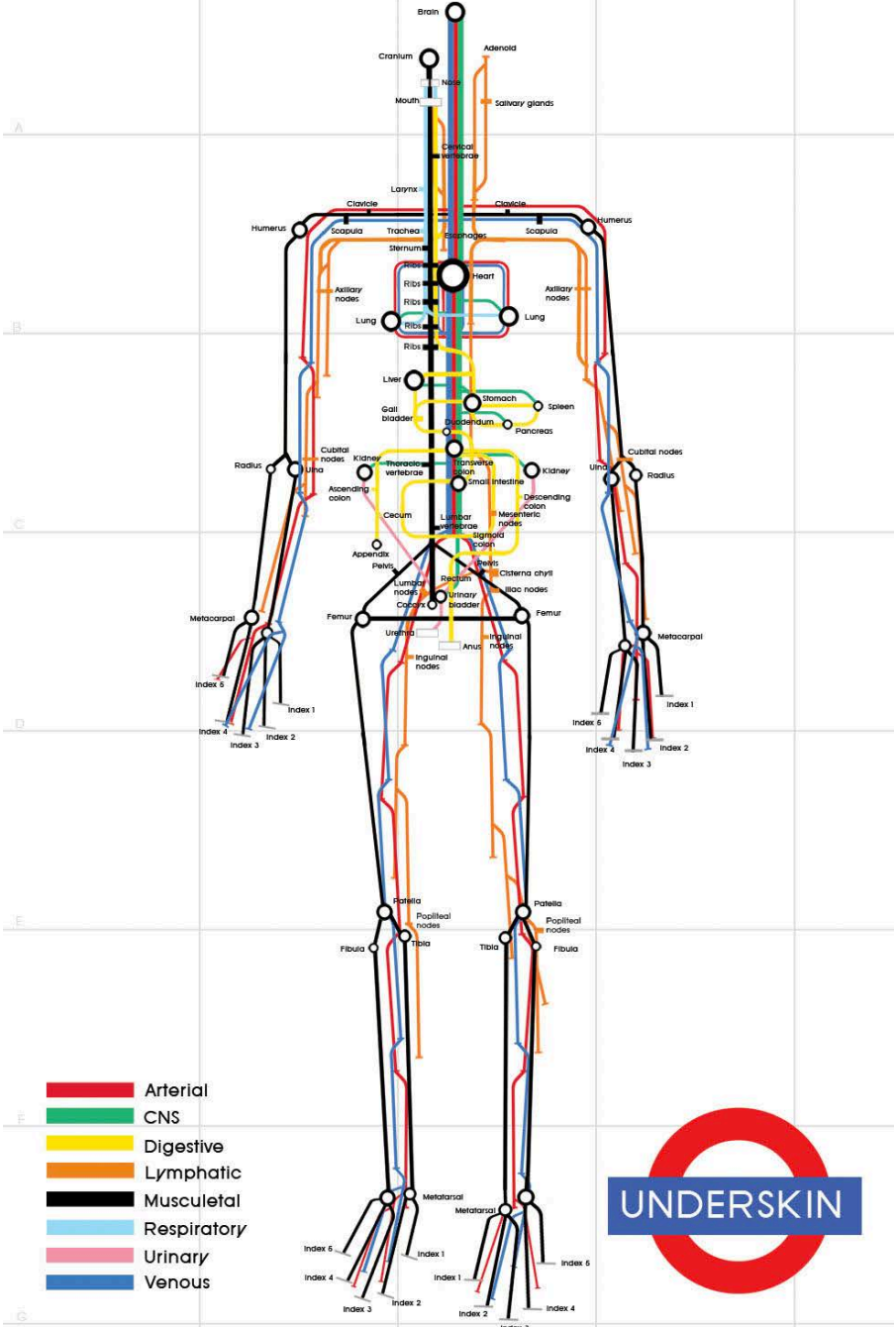
Use a simple text message
combined with a relevant image.



Getting information off the
Internet is like taking a
drink from a fire hydrant.

Mitchell Kapor

Make it unique!



A good infographic leaves you feeling informed or delighted.

- Krum, Randy, Cool Infographics, P52

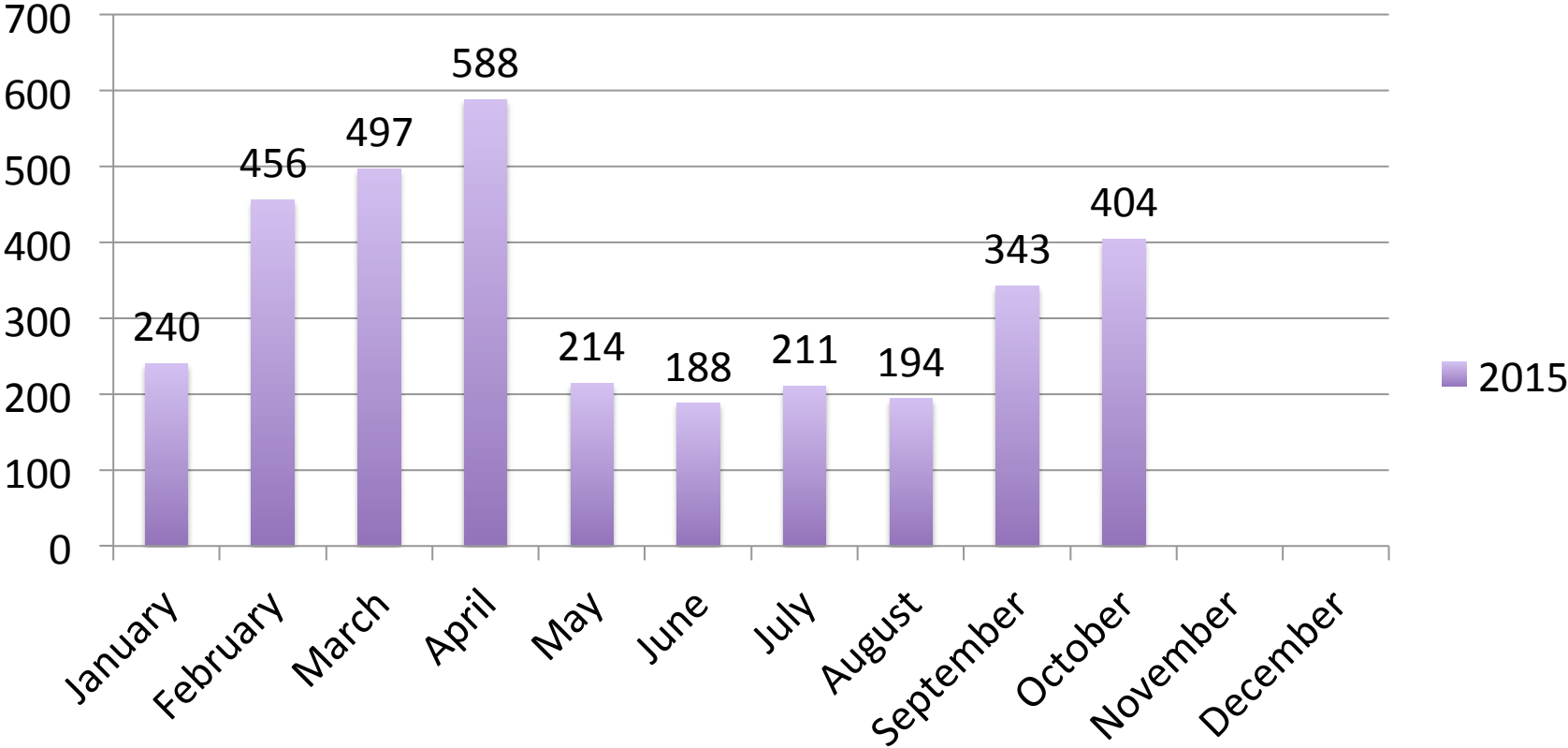
DATA VISUALIZATION BEST PRACTICES

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

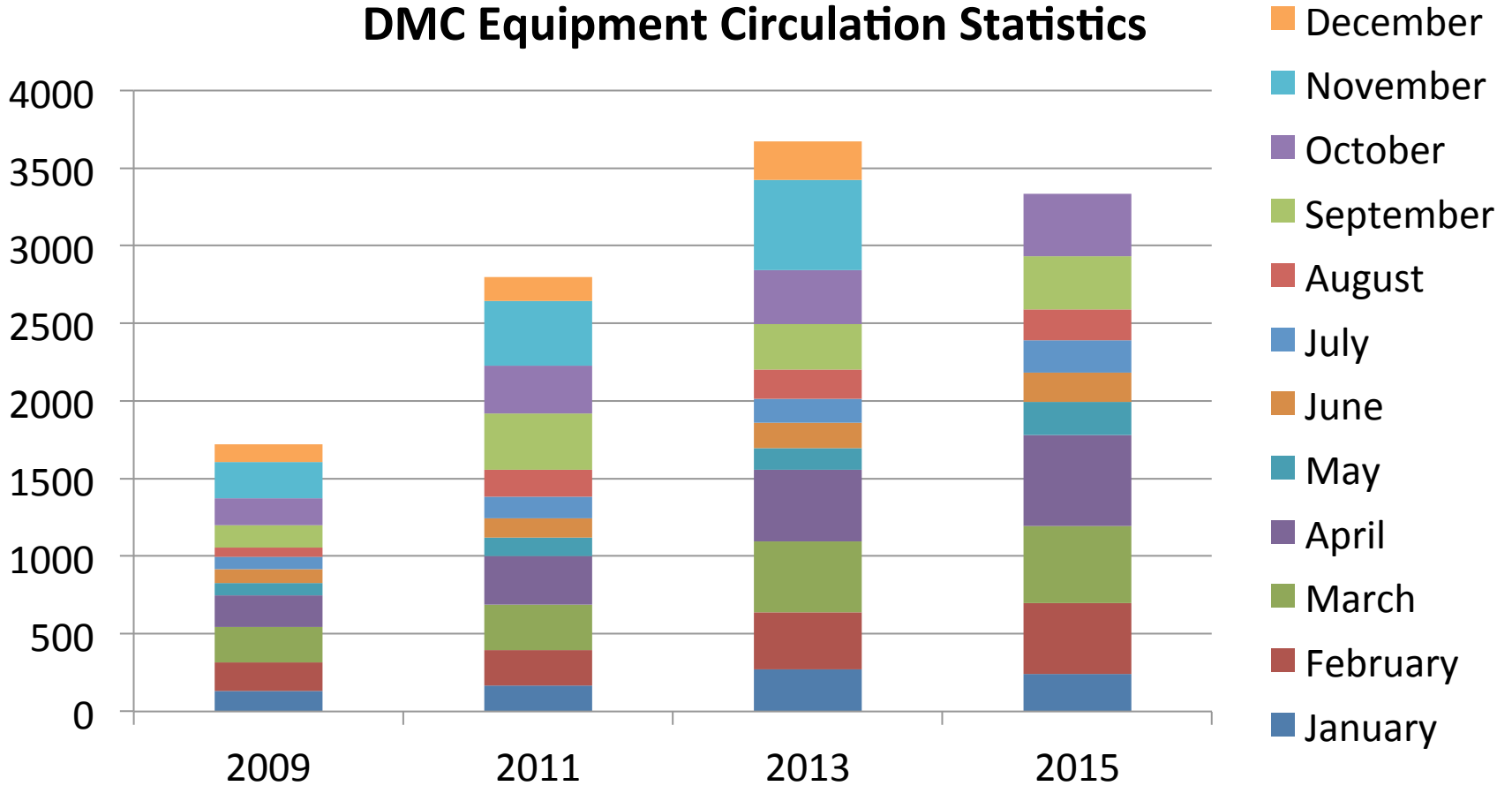


Avoid 3-D Bar Chart



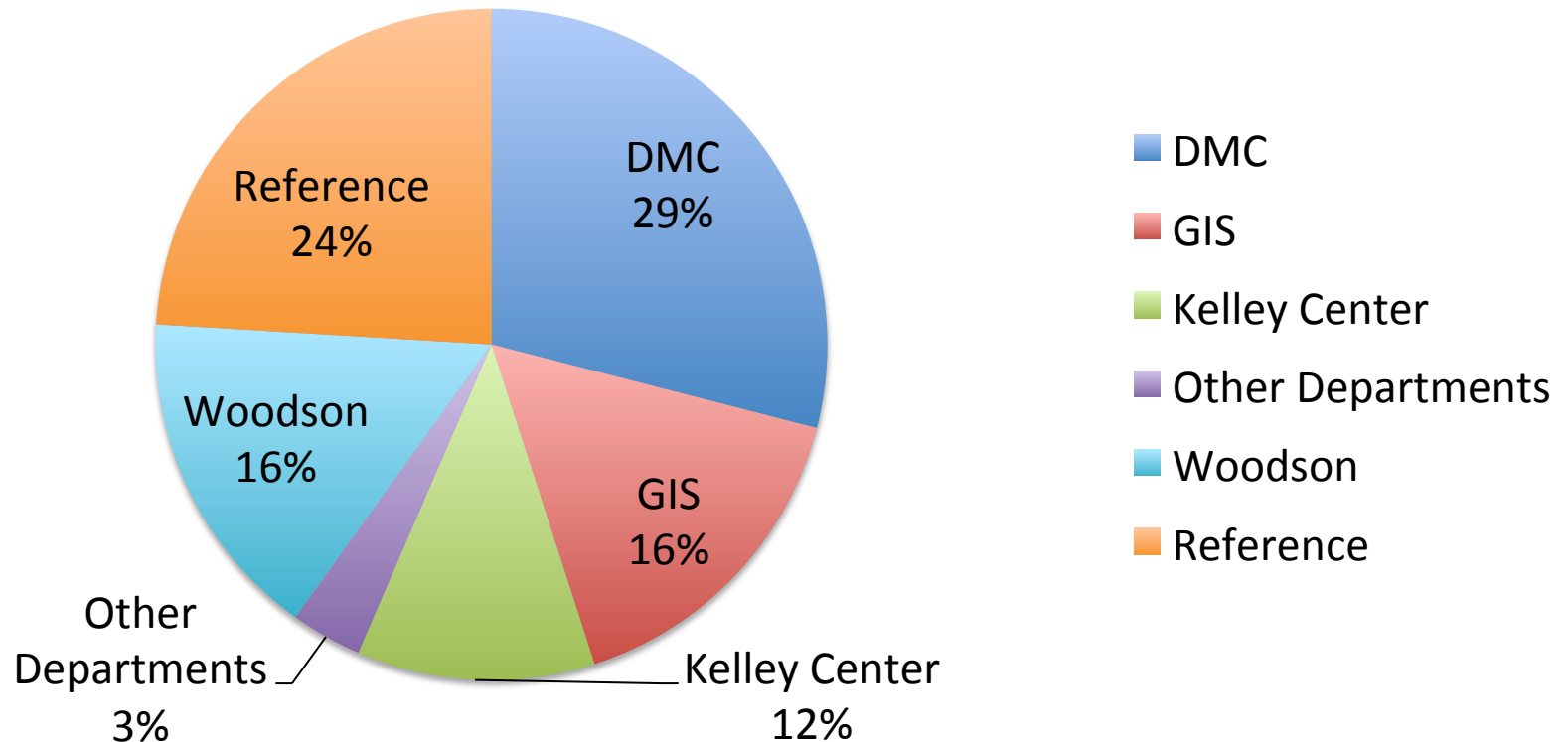
Stacked Bar Chart for Multiple Part-to-Whole Relationships

DMC Equipment Circulation Statistics



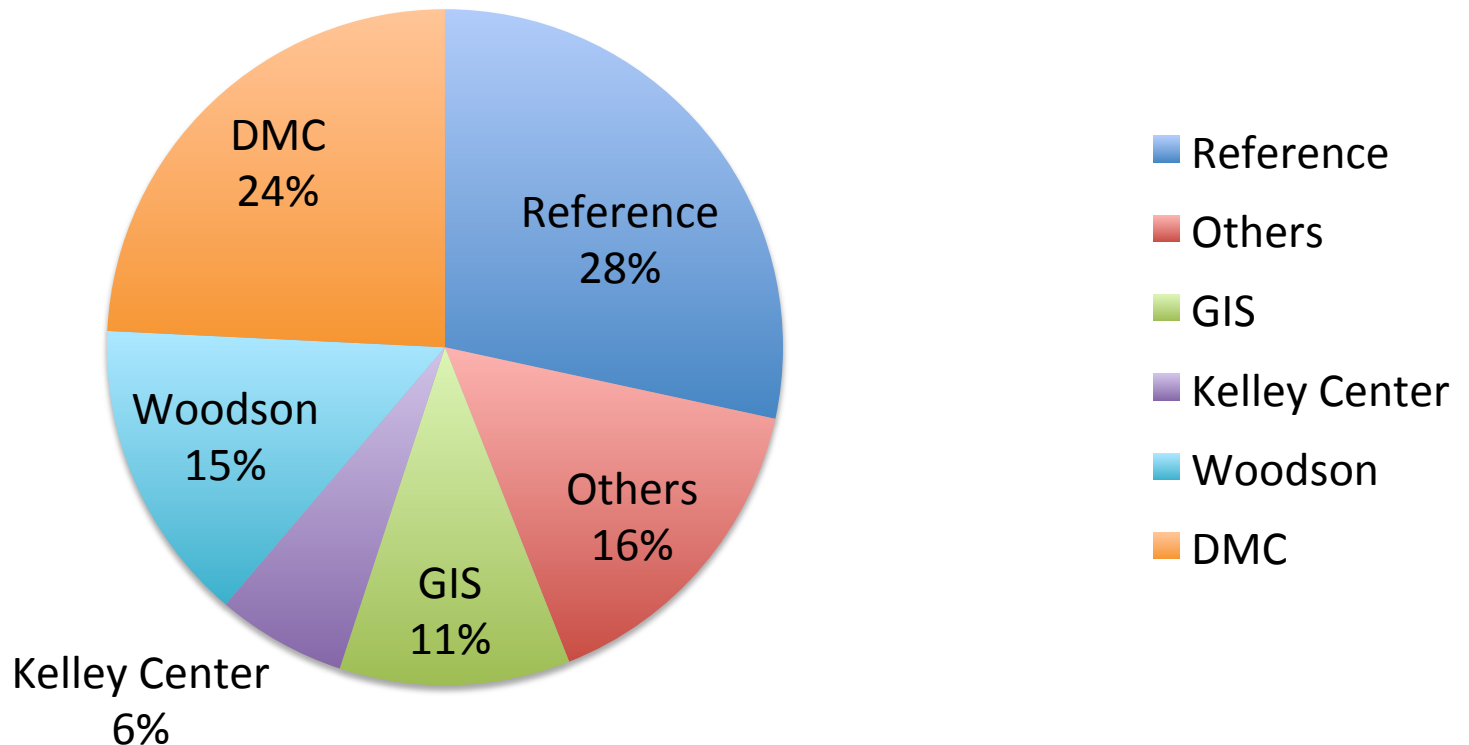
Pie Chart for Part-to-Whole Comparisons

2014-2015 Library Instruction Session Statistics

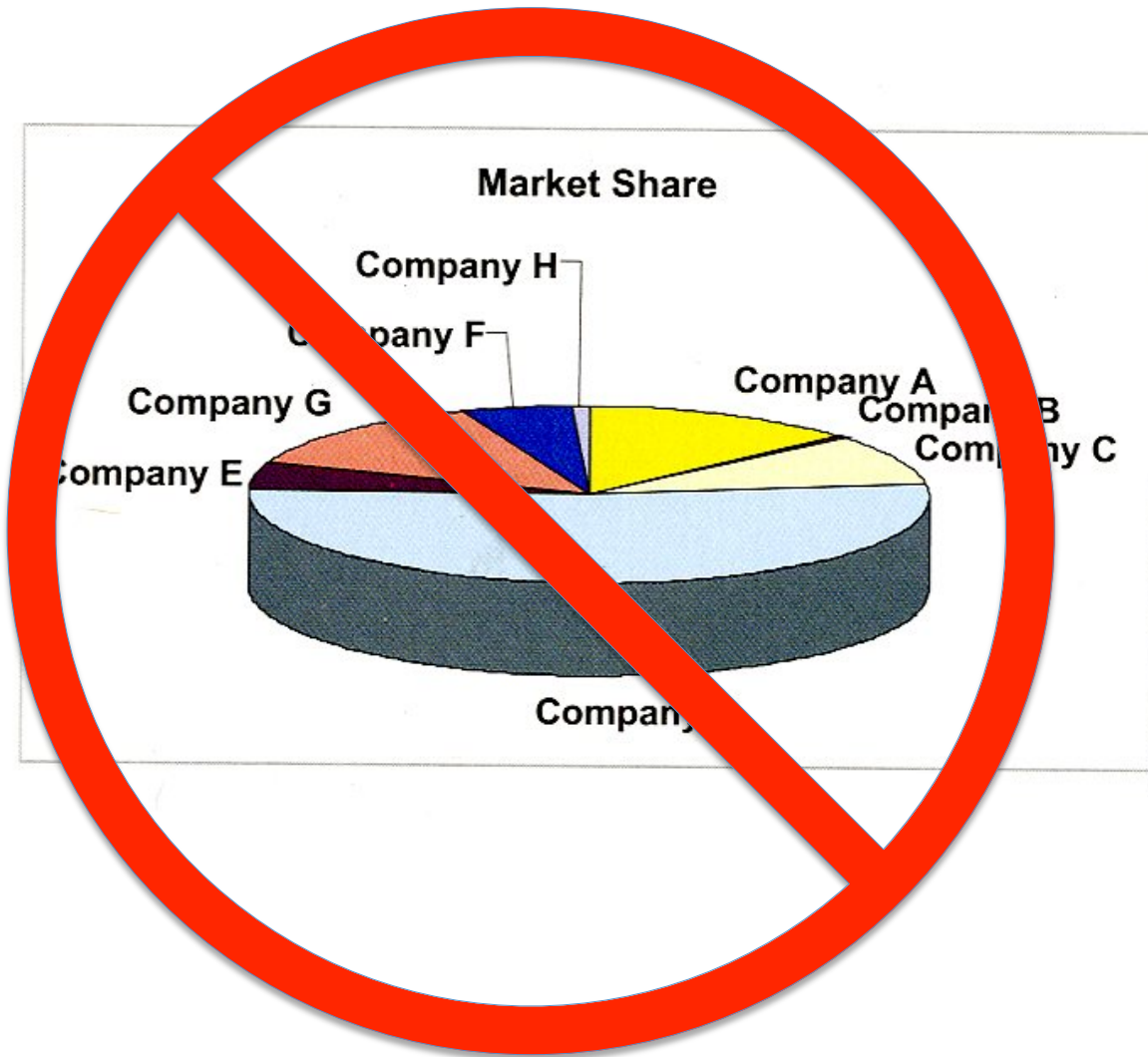


Pie Chart for Part-to-Whole Comparisons

2014-2015 People Trained by Library Instruction Sessions

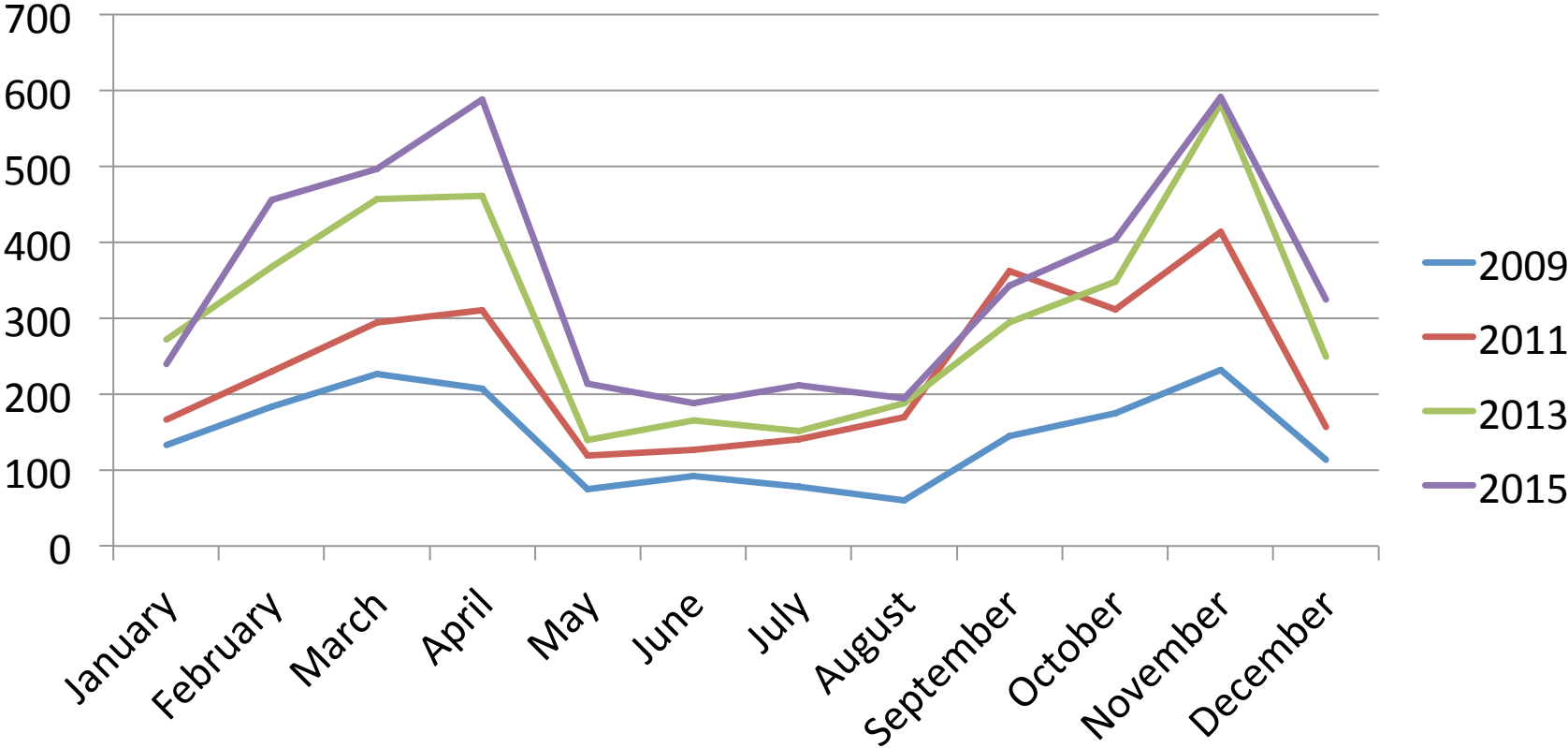


Avoid 3-D Pie Chart



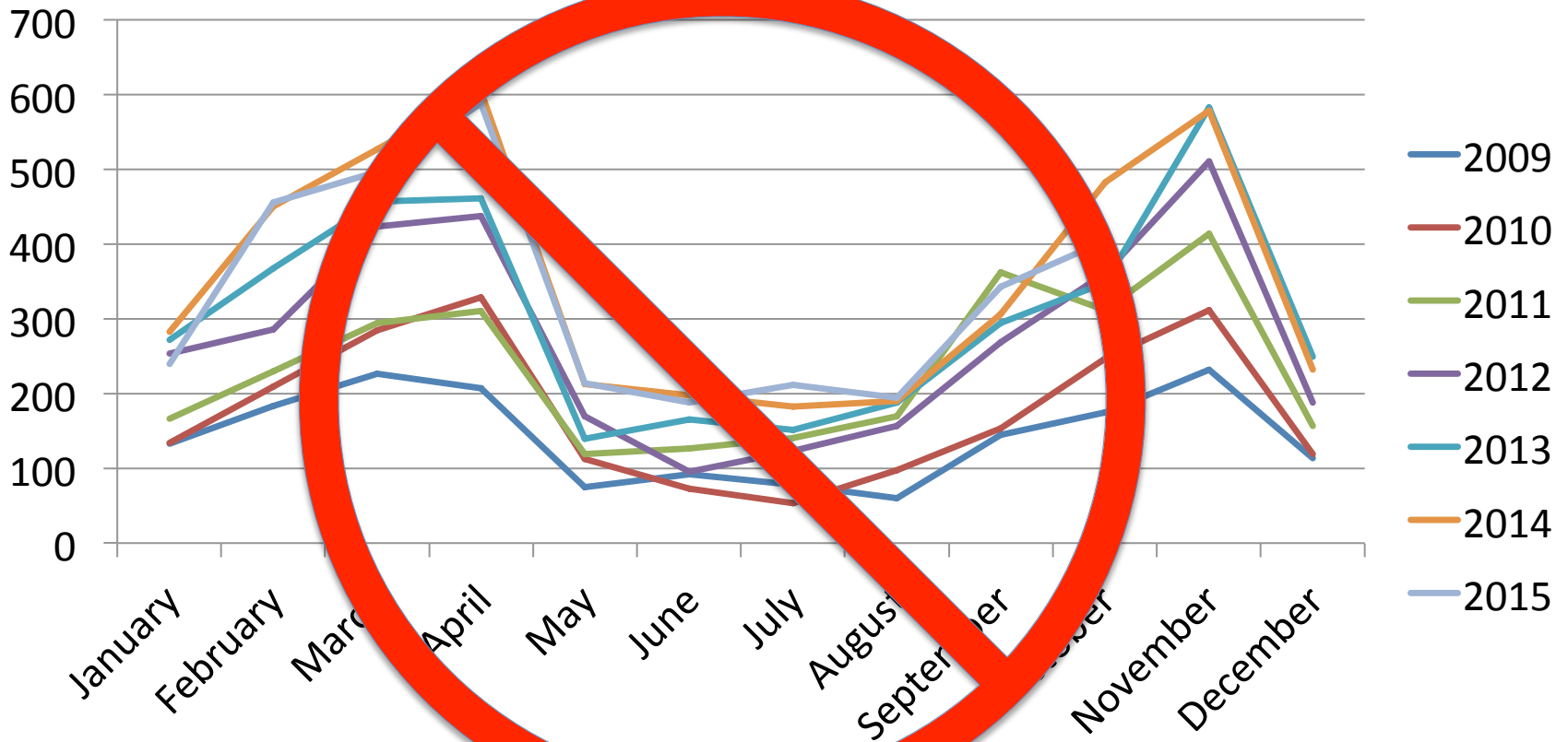
Line Chart for Time Series

DMC Equipment Circulation Statistics



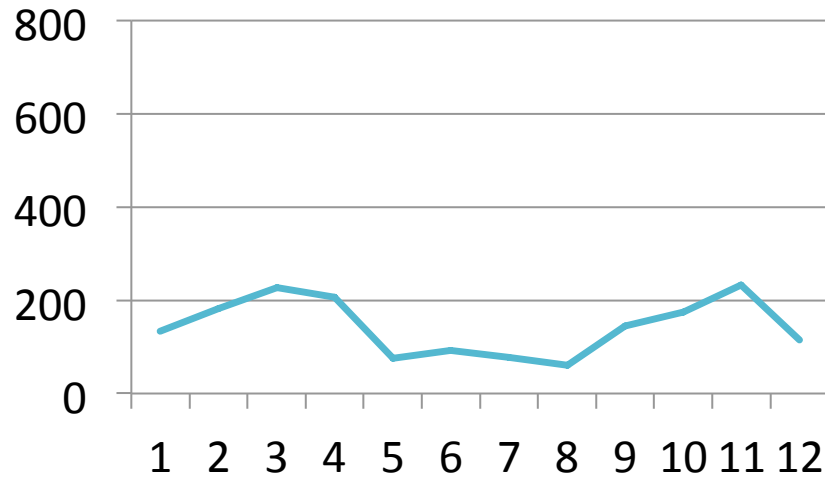
Keep the Line Chart to Four or Fewer. Otherwise the Chart is Too Busy!

DMC Equipment Circulation Statistics

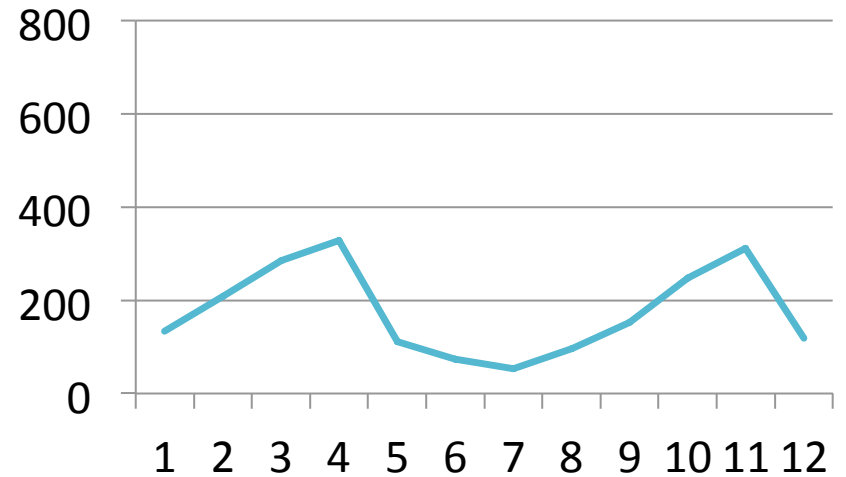


Use the Practice of Paneling and a Constant Scale for Consistency if You have More Than Four Lines.

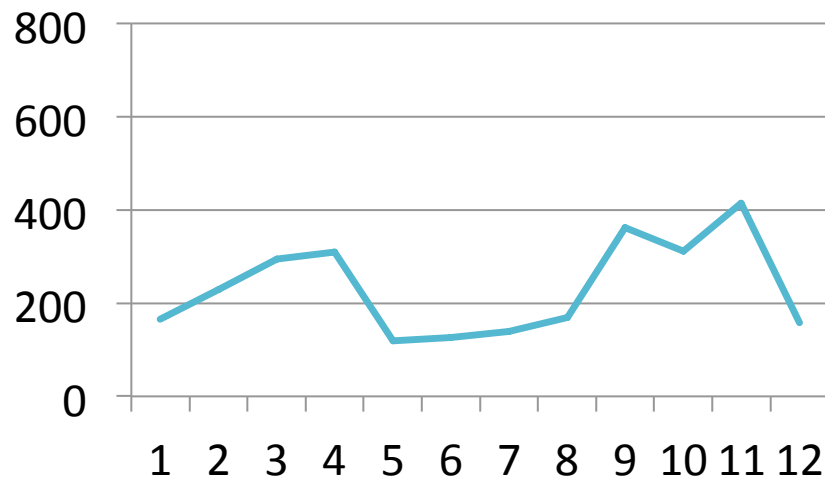
2009



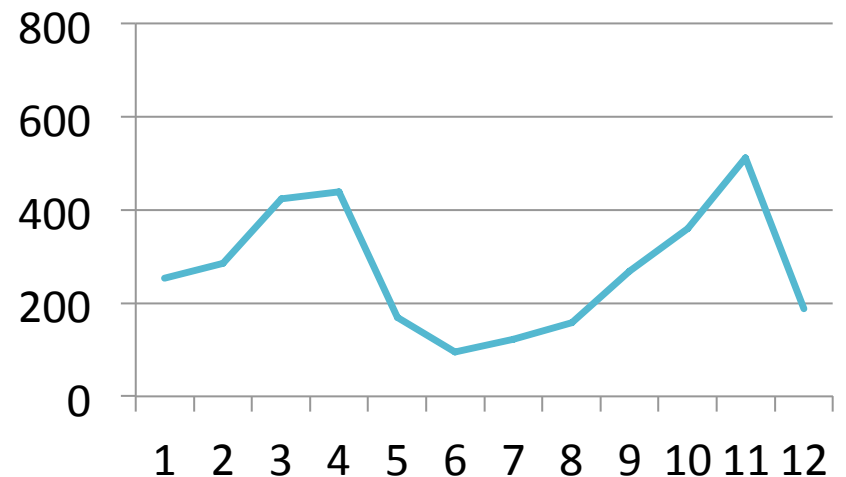
2010



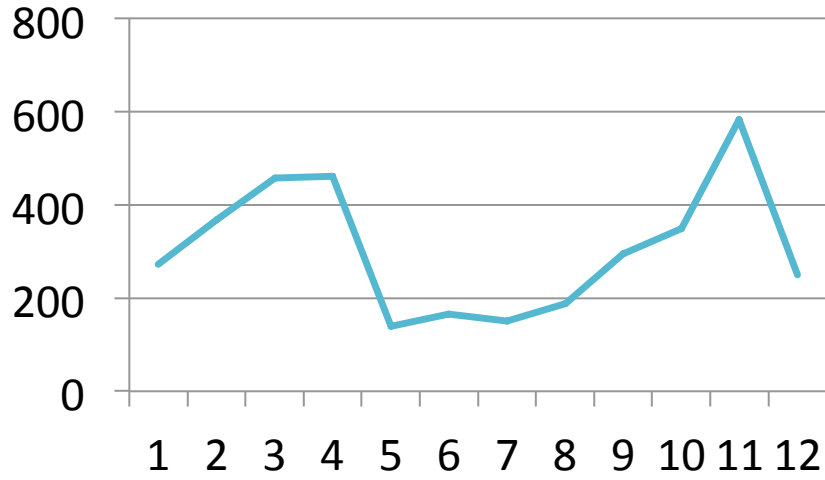
2011



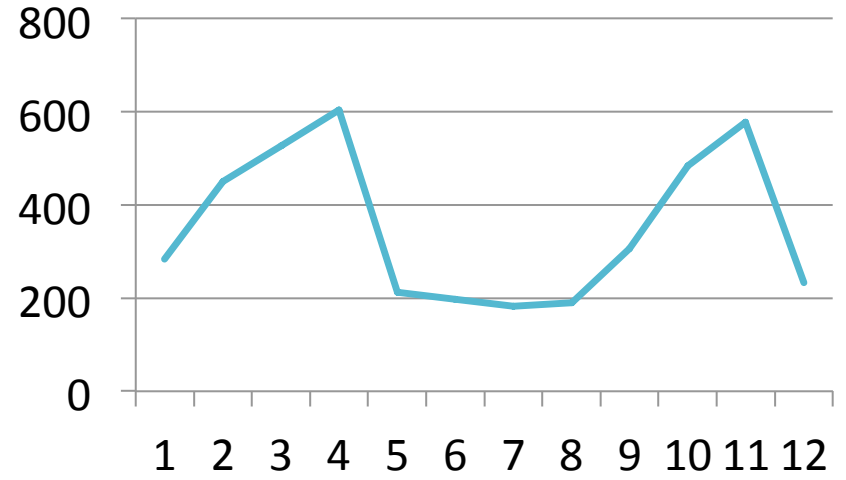
2012



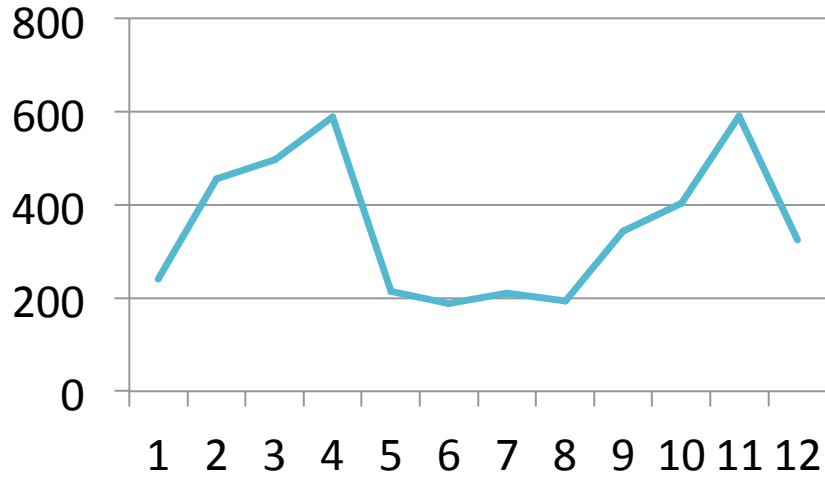
2013



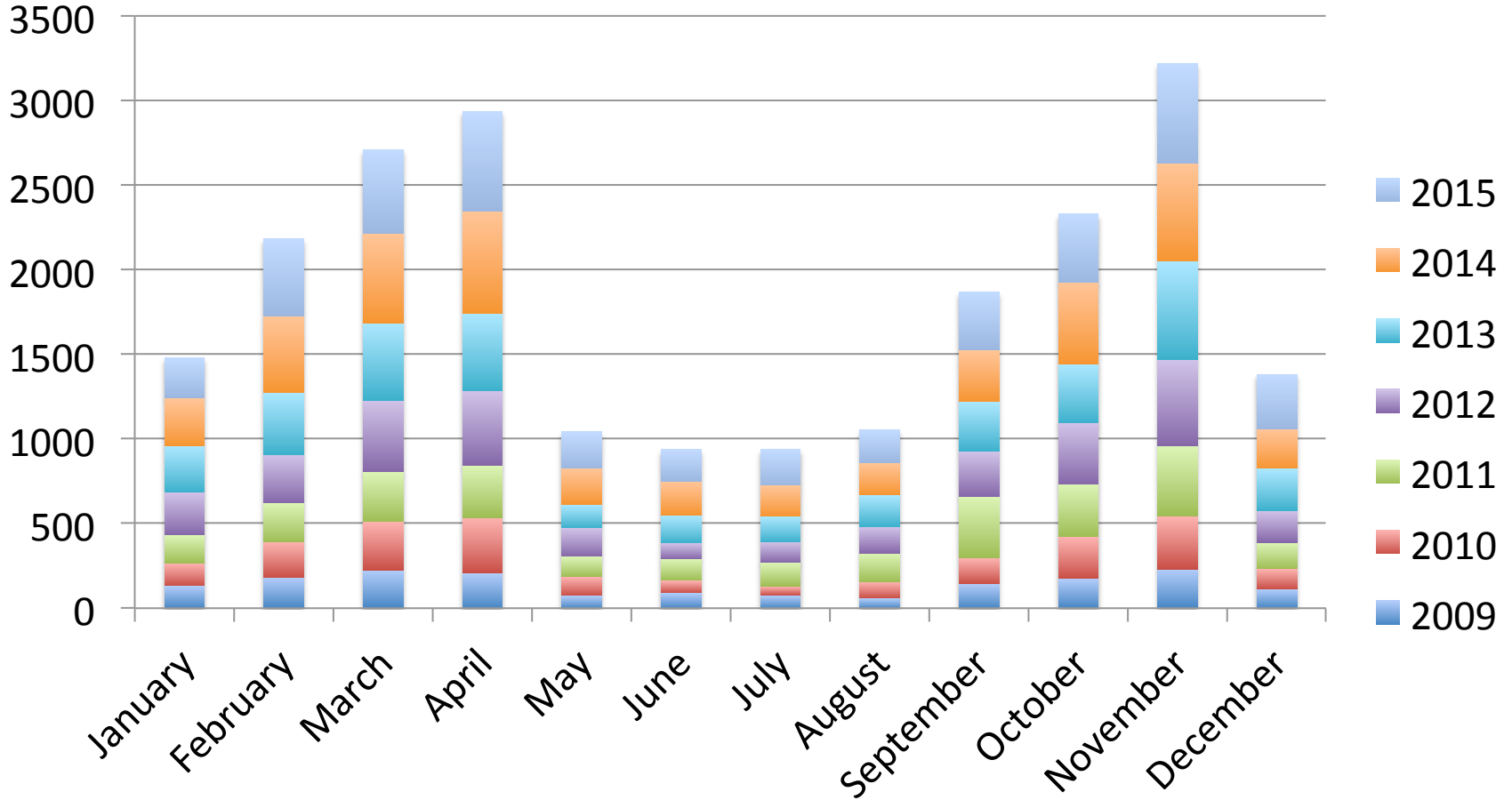
2014



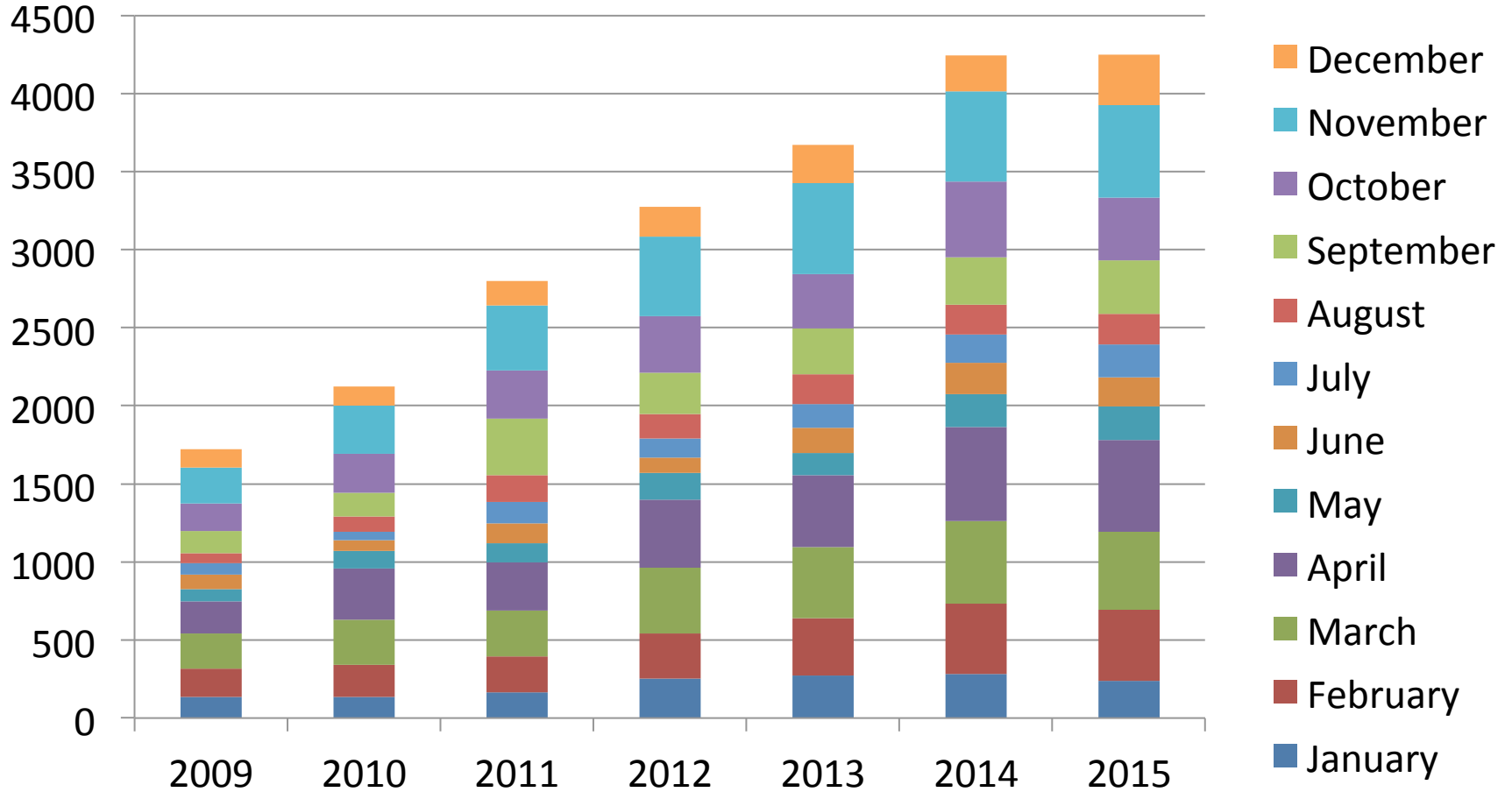
2015



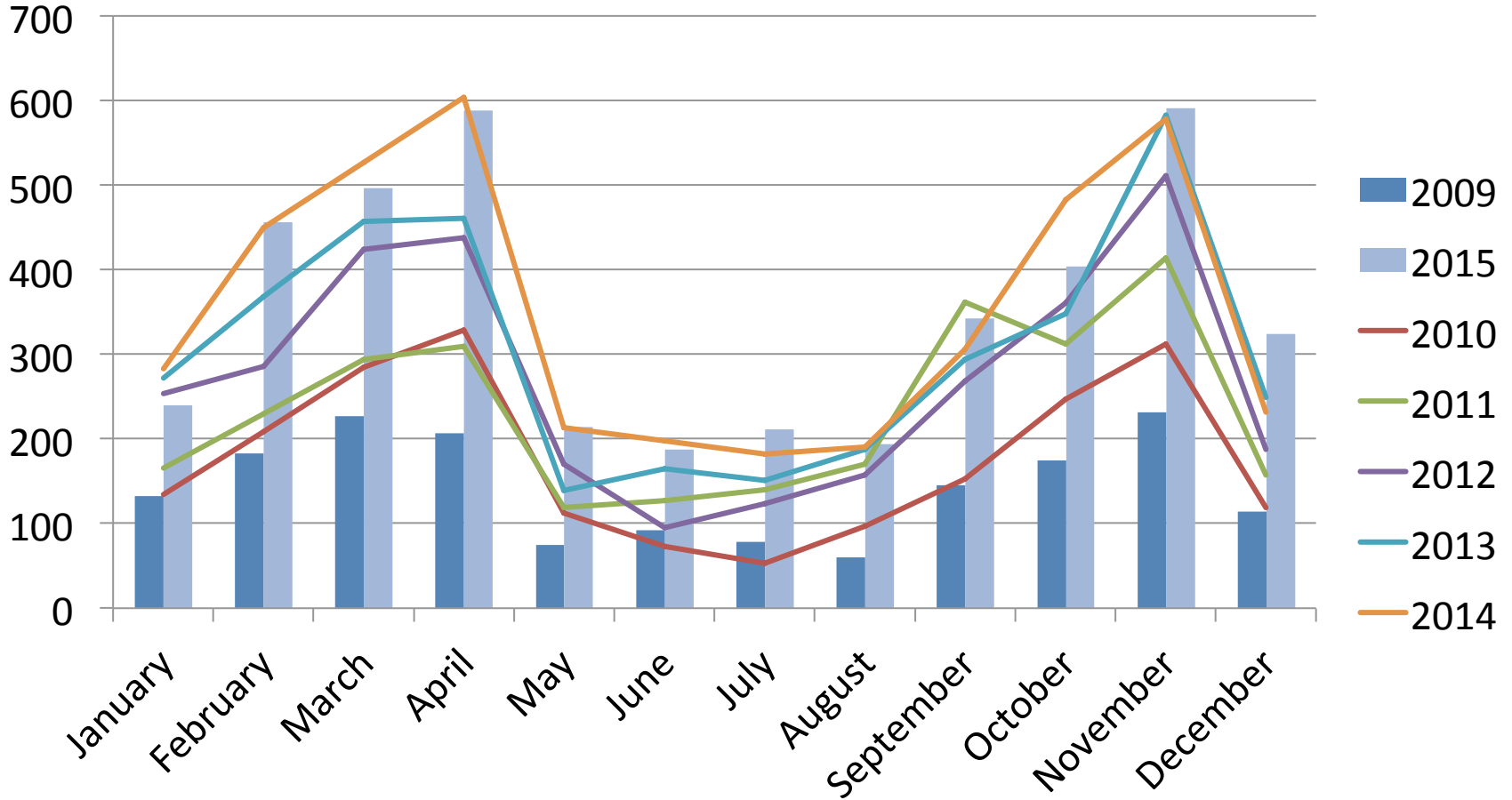
Stacked Bar Chart for Multiple Part-to-whole Relationships



Stacked Bar Chart for Multiple Part-to-whole Relationships



A Mix of Bar Chart and Line Chart



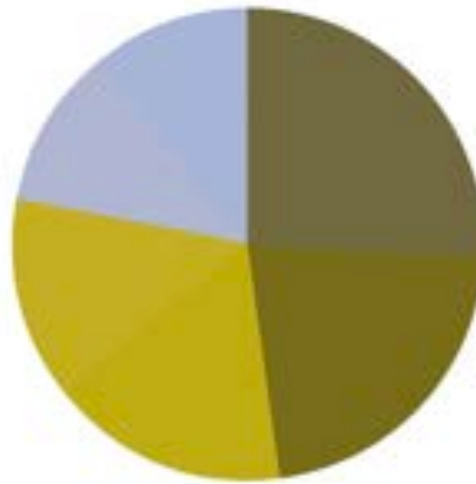
Make sure your infographic is complaint with Color Universal Design (CUG), which means **the graphical information is conveyed accurately to people with various types of color vision, including people with color blindness.**

Adjust hue or color brightness

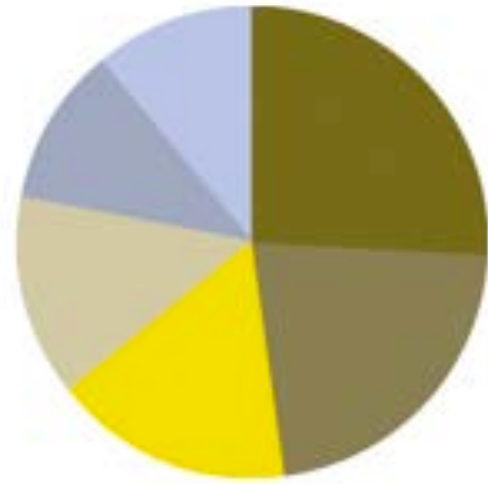
to make color-blind friendly color schemes



A



B

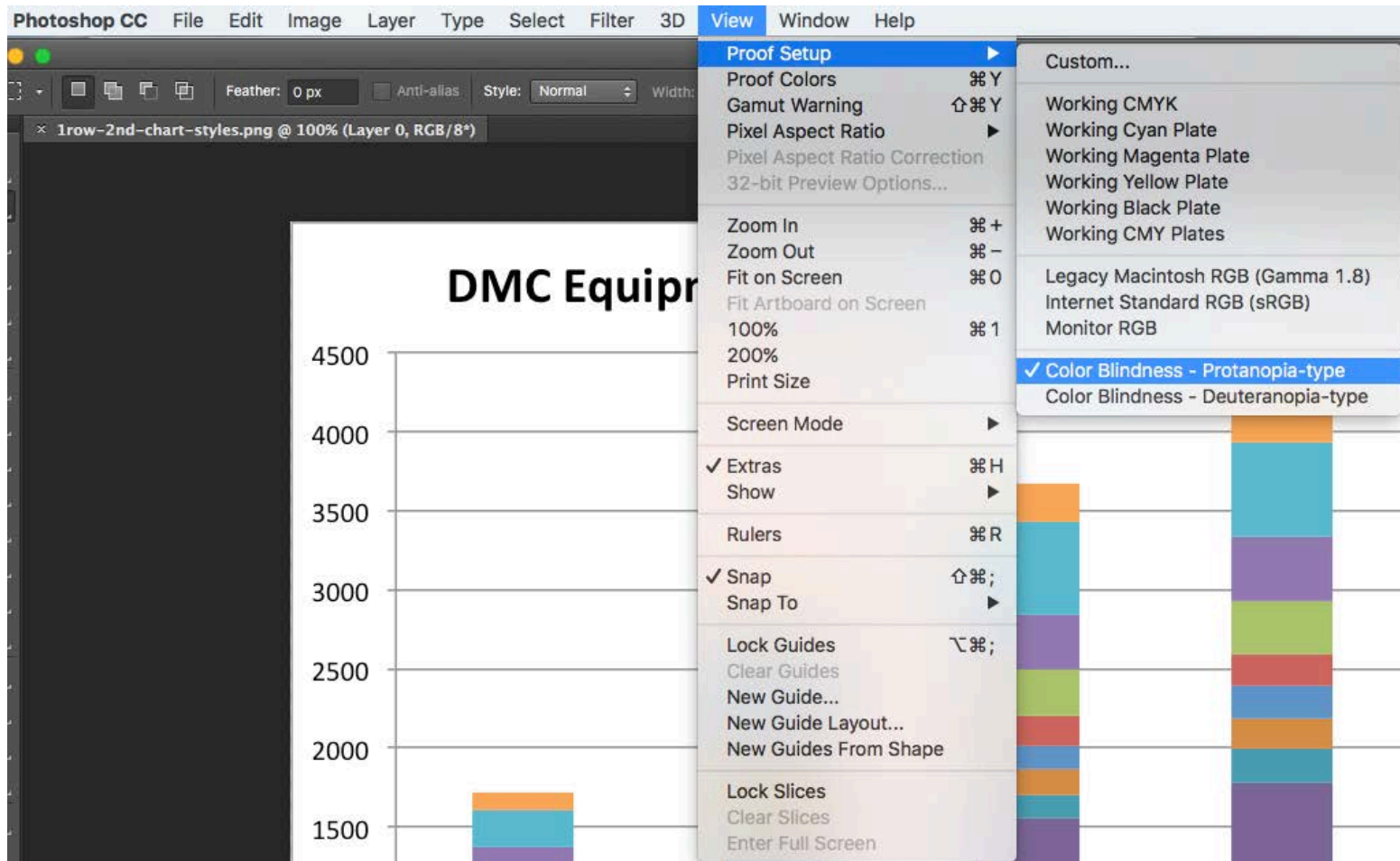


C

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

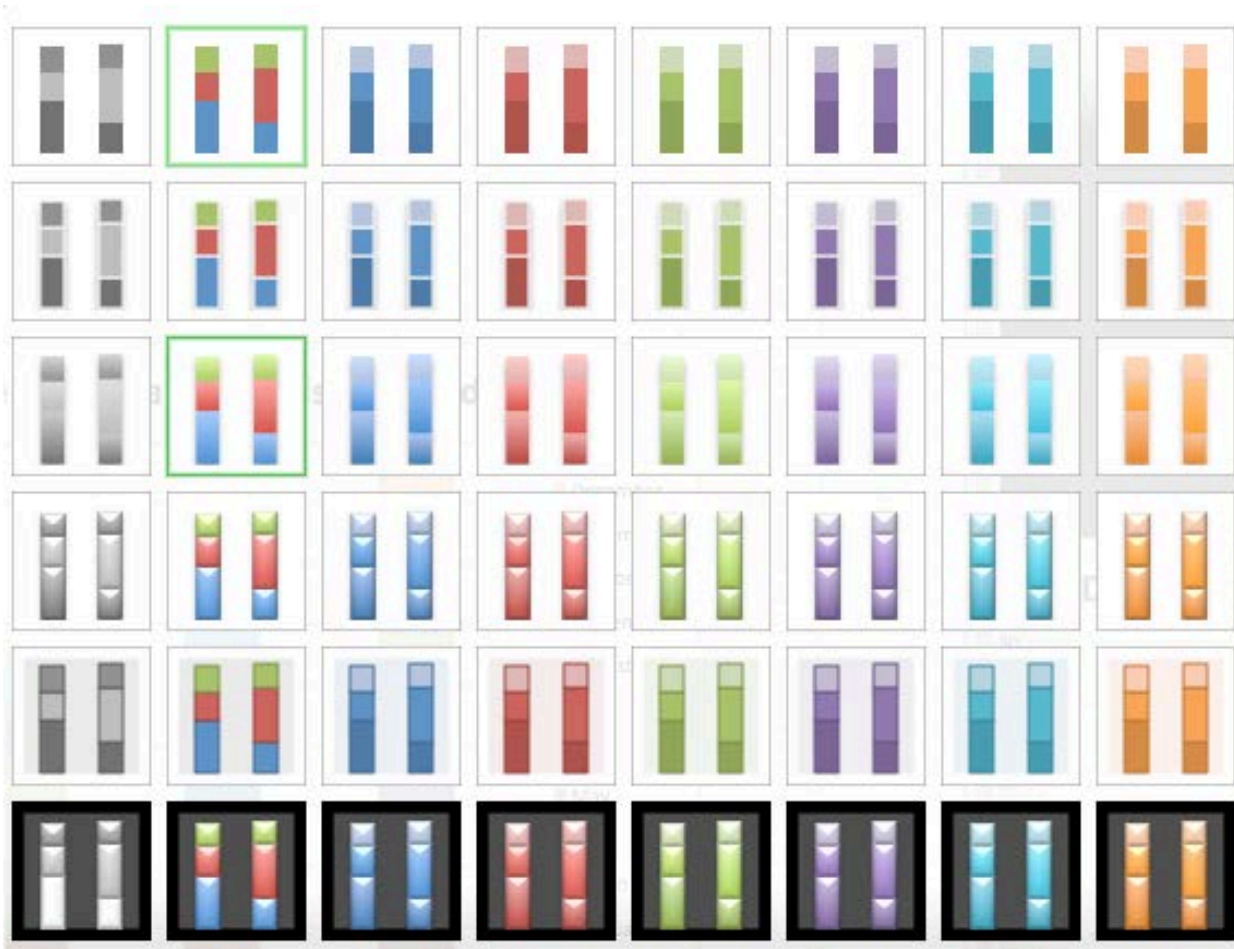
EXPORT
RGB

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

axismaps

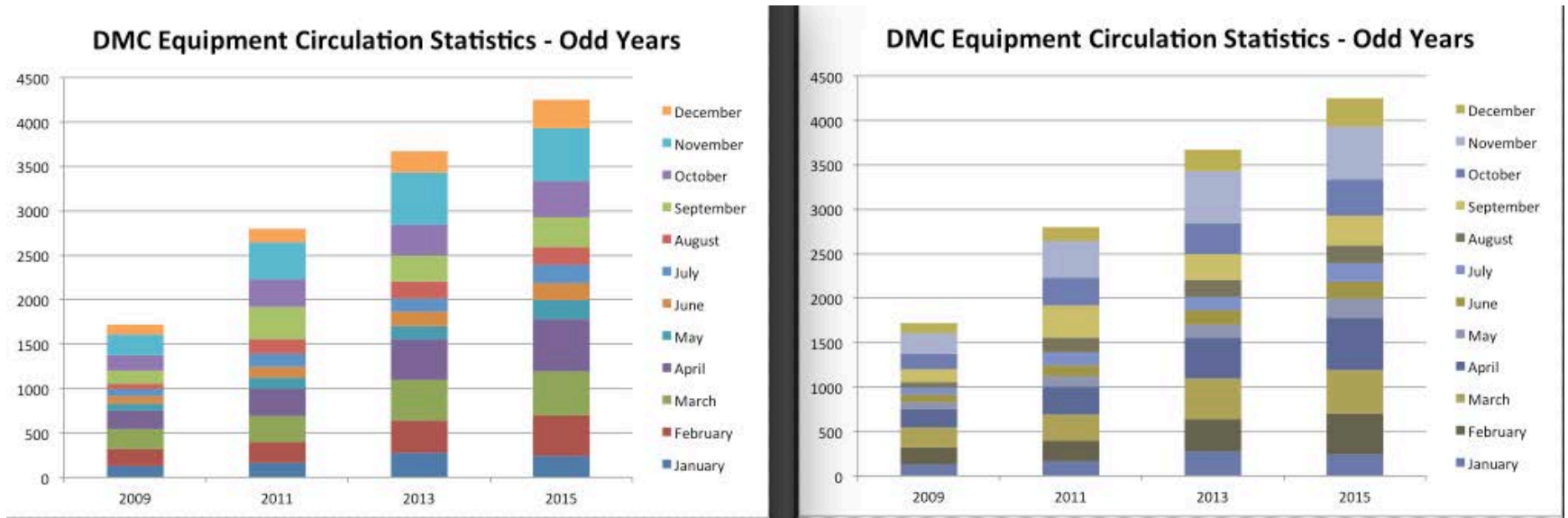
<http://colorbrewer2.org/> – Color Advice for Cartography

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



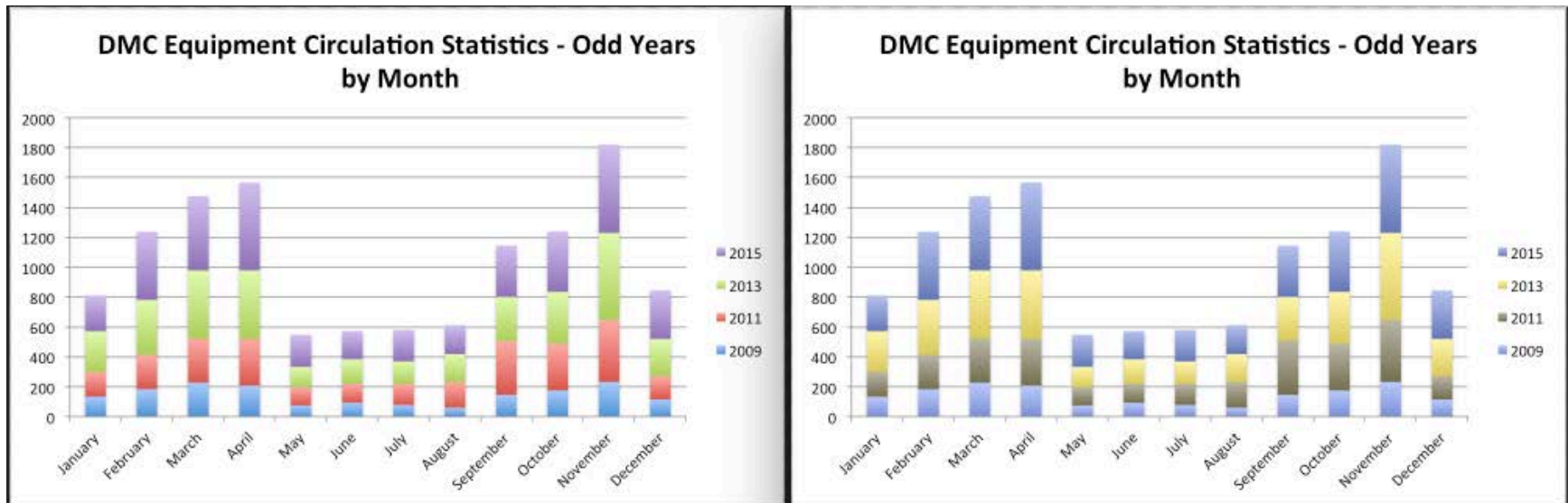
Left: Excel chart style
– 1st row 2nd one

Right: after turning on color-blindness tool in Photoshop



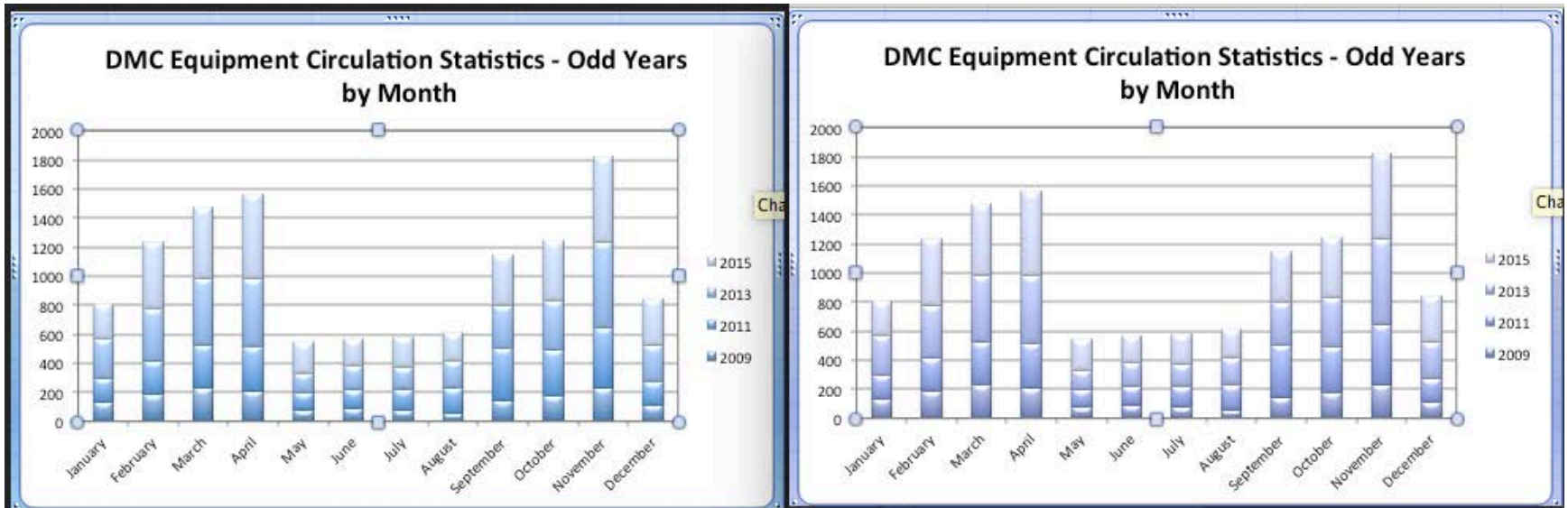
Left: Excel chart style
– 3rd row 2nd one

Right: after turning on color-blindness tool in Photoshop



Left: Excel chart style
– 4th row 3rd 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!

TOOLS FOR CREATING INFOGRAPHICS AND DATA VISUALIZATION

Desktop Tools – Vector Graphics

PowerPoint	Excel	Adobe Illustrator	Adobe InDesign
			
Gephi(free)	OmniGraffle	InkScape(free)	
			

Our Mission



The DMC supports the creation and use of multimedia in education, scholarship, and creative expression. Working toward this end,

we provide services that include hands-on training, assistance with digital projects, and access to the essential tools for creating digital resources such as digital video and audio, images and animations, infographics, PowerPoint presentations, web pages, and more.

DMC Offers Hands-on Training on Media Editing and Assistance with Various Digital Projects

1. Help with using DMC equipment
2. Demonstration of DMC equipment
3. Assistance on video/audio editing, and graphics creation
4. Consultation on patron's project
5. Short courses for using digital tools



DMC Provides Access to the Essential Tools and Facilities for Creating Digital Media

1. Poster printing
2. Skyping/Podcasting
3. Equipment available for checking out
4. Lecture/interview recording
5. Photo taking
6. iMovie, Final Cut Pro, Photoshop, Illustrator, InDesign, and more



Your Projects,
Our Passion!

Our mission

The DMC supports the creation and use of multimedia in education, scholarship, and creative expression. Working toward this end, we provide services that include hands-on training, assistance with digital projects, and access to the essential tools for creating digital resources such as digital video and audio, images and animations, PowerPoint presentations, e-books, and more.

DMC Offers Hands-on Training on Media Editing and Assistance with Various Digital Projects

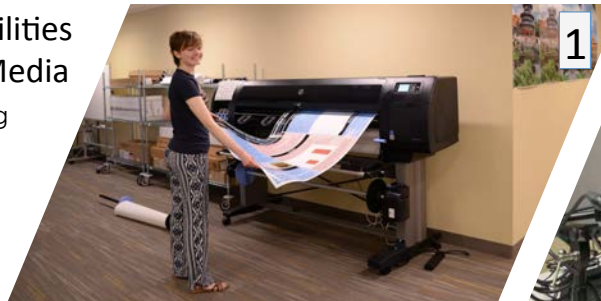
1. Help with using DMC equipment
2. Demonstration of DMC equipment
3. Assistance on video/audio editing, and graphics creation
4. Consultation on patron's project
5. Short courses for using digital tools



DMC Provides Access to the Essential Tools and Facilities for Creating Digital Media





1. Poster printing
2. Skyping/Podcasting
3. Equipment available for checking out
4. Lecture/interview recording
5. Photo taking

6. iMovie, Final Cut Pro, Photoshop, Illustrator, InDesign, and more



Your Projects, Our Passion!

Desktop Tools – Image Editing

Adobe Photoshop	Gimp(free)
 The Adobe Photoshop icon, featuring a blue square with the letters 'Ps' in a lighter blue font.	 The GIMP icon, which is a grey, cartoonish creature with large eyes and a pencil in its mouth.
Pixelmator	Acorn
 The Pixelmator icon, showing a blue pen nib resting on a square photograph of a sunset over palm trees.	 The Acorn icon, depicting a realistic brown acorn with its cap.

Gephi

- <https://dhs.stanford.edu/tools/maps-graphs-and-workshops/>

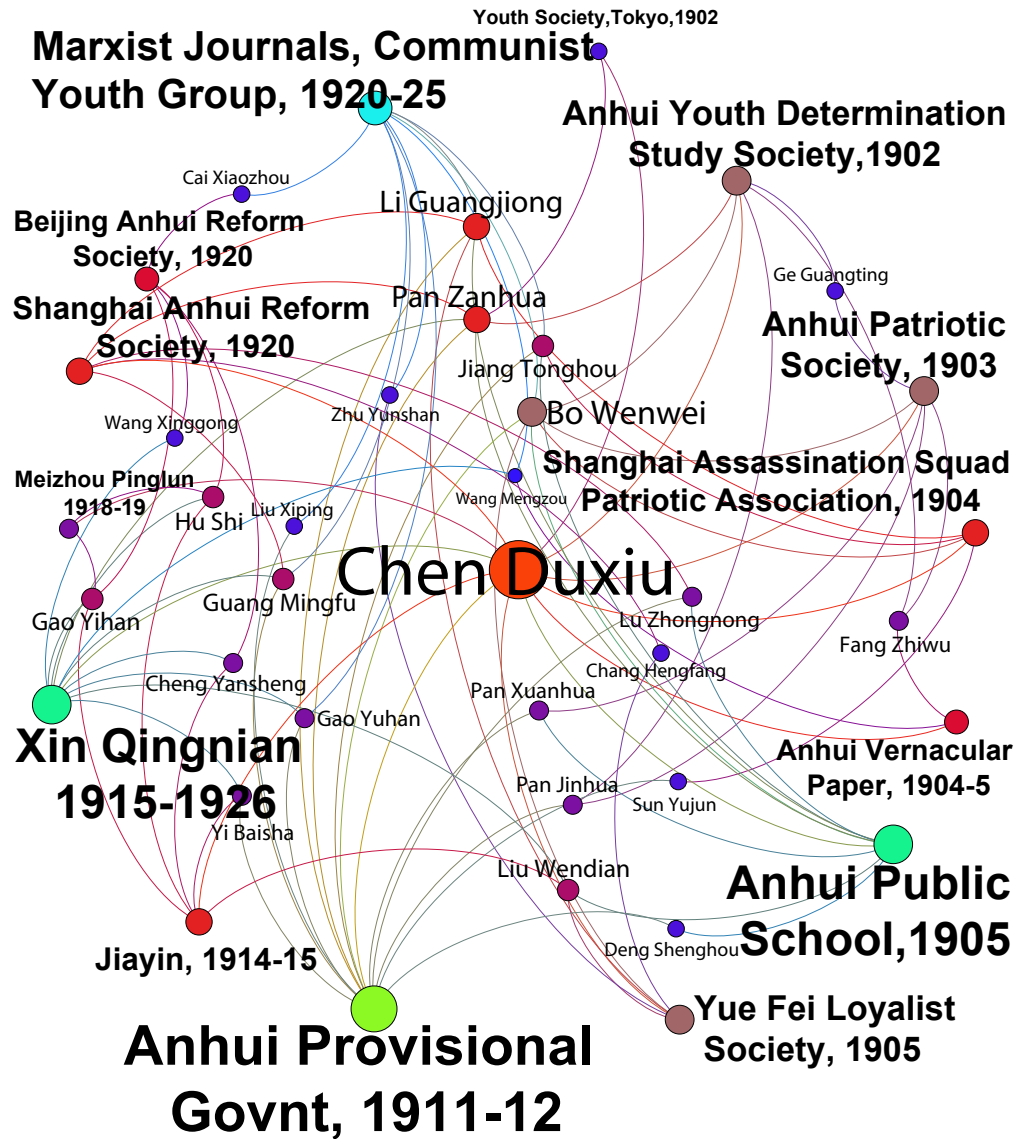


Diagram 2 - Anhui Co-Proprietary in Chen's Networks

Online Tools

- **Wordle.net** <http://www.wordle.net/>
- **Google Chart** <https://developers.google.com/chart/>
- **Tableau Public** <https://public.tableau.com/s>

Online Infographics Resources

Periodic Table of Visualization Methods

C continuum	Data Visualization Visual representations of quantitative data in schematic form (either with or without axes)		Strategy Visualization The systematic use of complementary visual representations in the analysis, development, formulation, communication, and implementation of strategies in organizations.		G graphic facilitation															
Tb table	Ca cartesian coordinates	Information Visualization The use of interactive visual representations of data to amplify cognition. This means that the data is transformed into an image, it is mapped to screen space. The image can be changed by users as they proceed working with it.		Metaphor Visualization Visual Metaphors position information graphically to organize and structure information. They also convey an insight about the represented information through the key characteristics of the metaphor that is employed																
Pi pie chart	L line chart	Concept Visualization Methods to elaborate (mostly) qualitative concepts, ideas, plans, and analyses.		Compound Visualization The complementary use of different graphic representation formats in one single schema or frame																
B bar chart	Ac area chart	R radar chart	Pa parallel coordinates	Hy hyperbolic tree	Cy cycle diagram	T timeline	Ve vena diagram	Mi mindmap	Sq square of oppositions	Cc concentric circles	Ar argument slide	Sw swim lane diagram	Gc gant chart	Pm perspectives diagram	D dilemma diagram	Pr parameter ruler	Kn knowledge map			
Hi histogram	Sc scatterplot	Sa sankey diagram	In information lense	E entity relationship diagram	Pt petri net	Fl flow chart	Cl clustering	Le layer chart	Py pyramid	Ce ceiling	Tl tall	Dt dot plot	Cp cpm critical path method	Cf concept fan	Co concept map	Ic iceberg	Lm learning map			
Th tukey box plot	Sp spectrogram	Da data map	Tp treemap	Cn cone tree	Sy system dyn./ simulation	Df data flow diagram	Se semantic network	So soft system modeling	Me meeting trace	Mm metro map	Tm temple	St story template	Tr tree	Ct cartoon	Co communication diagram	Fp flight plan	Cs concept skeleton	Br bridge	Fu funnel	Ri rich picture

Cy Process Visualization

Hy Structure Visualization

Overview
 Detail

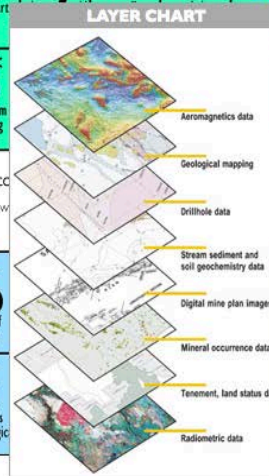
Detail AND Overview

Divergent thinking

Convergent thinking

Note: Depending on your location and context, you may want to click on a cell to load a pop-up picture.
© Ralph Lengler & Martin J. Eppler, www.

Su supply demand curve	Pe performance charting	St strategy map	Oc organisation chart	Ho house of quality	Po porter's five forces	S s-cycle	Sm stakeholder map	Is ishikawa diagram	Tc technology roadmap
Ed edgeworth box	Pf portfolio diagram	Sg strategic game board	Mz mintzberg's organigraph	Z zwicky's morphologic box	Vc value chain	Hy hype-cycle	Sr stakeholder rating map	Ta taps	Sd spray diagram

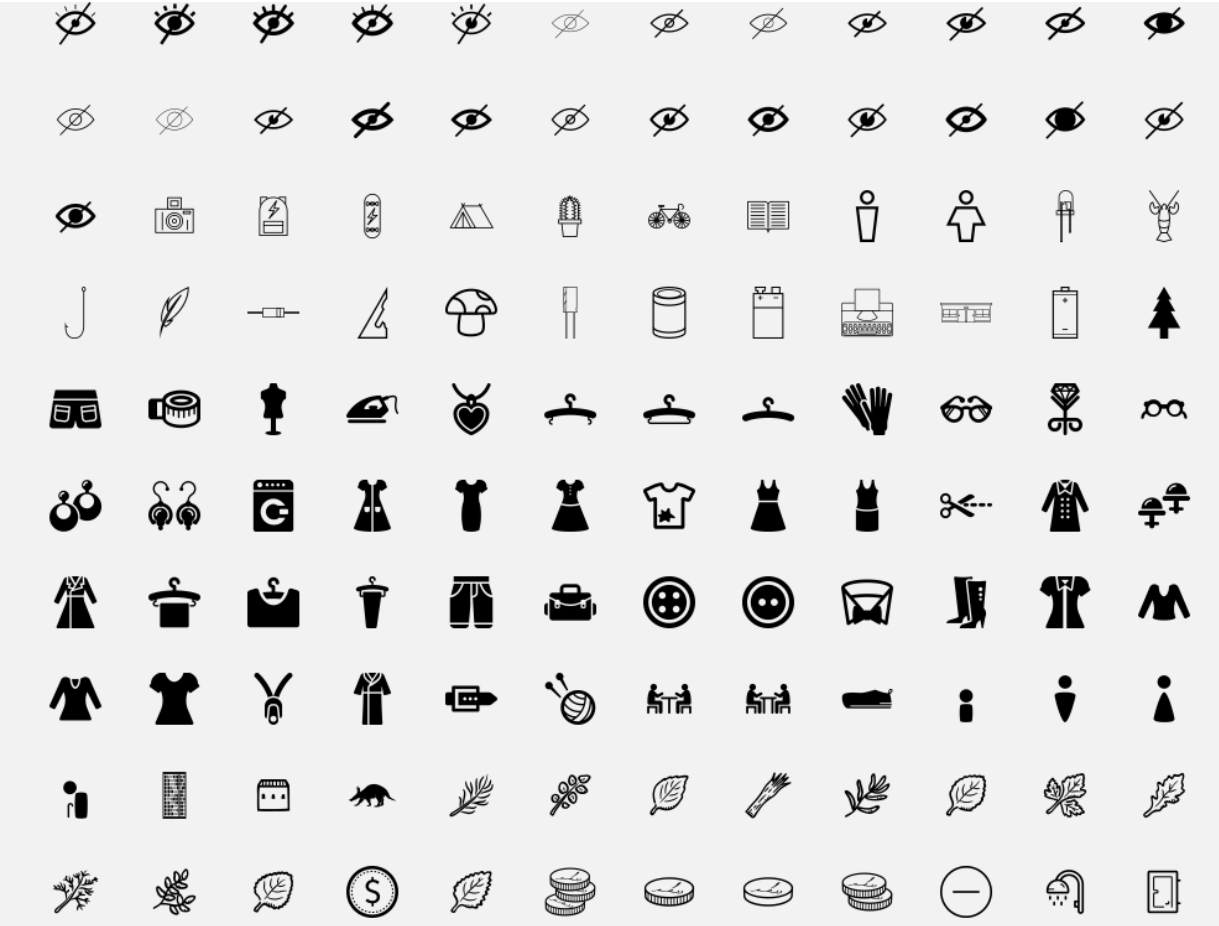


load a pop-up picture.

version 1.5

Online Infographics Resources

The Noun Project



Online Infographics Resources

- [Periodic Table of Visualization Methods](#)
- [The Noun Project](#)
- [22 free tools for data visualization and analysis](#)
- [infographics world](#)
- Datavisualization.ch Selected Tools
<http://selection.datavisualization.ch/>

More Sample Infographics

- Cool infographics

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

- Edward Tufte

<http://www.edwardtufte.com/tufte/posters>

- Information is beautiful

<http://www.informationisbeautiful.net/>

by David McCandless, an author and designer.

Data Sources

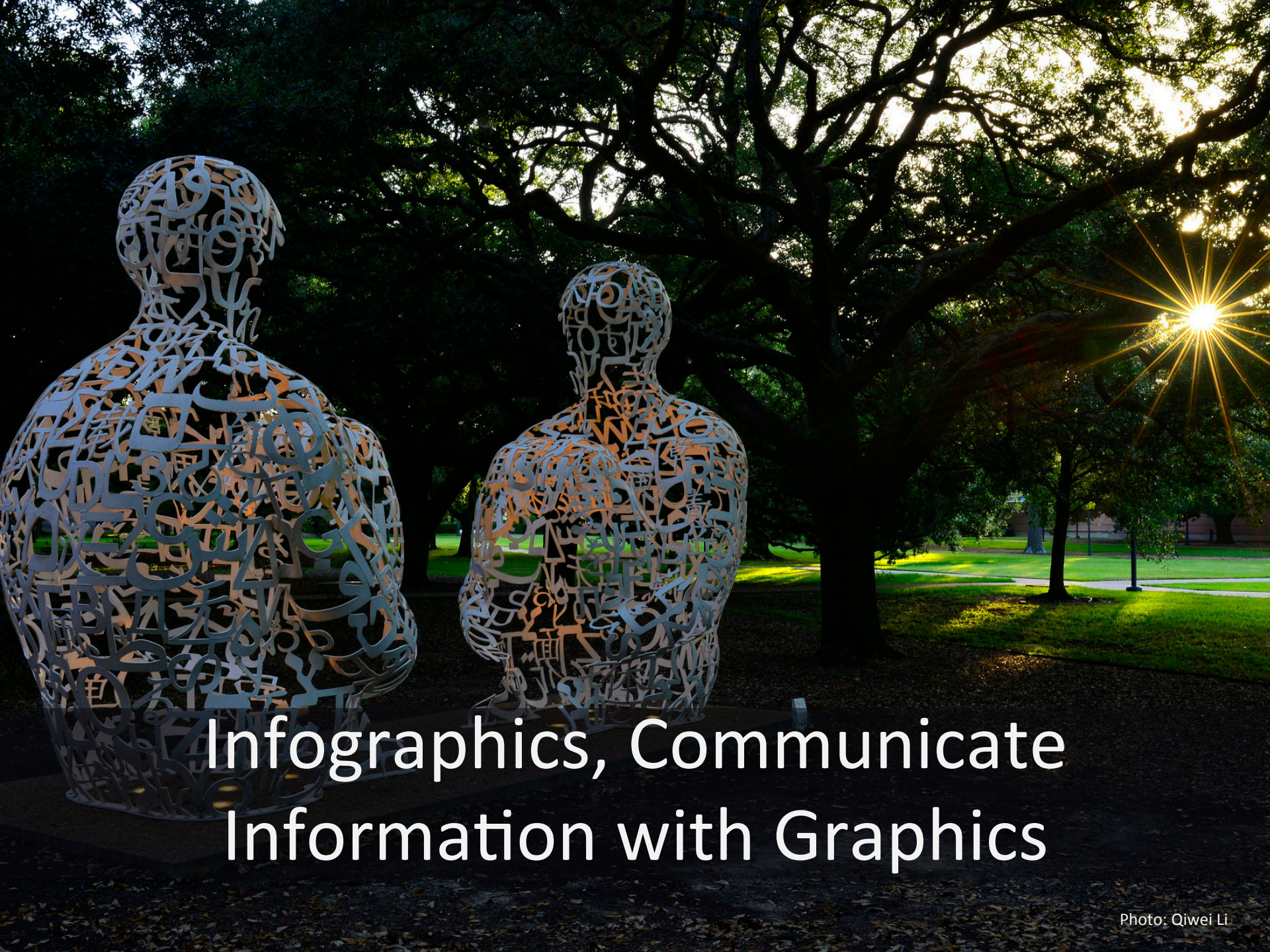
- **data.gov**
<http://www.data.gov/>.
- FactBrowser
<http://www.factbrowser.com/>
- Google Public Data
<http://www.google.com/publicdata/directory>
- Wolfram Alpha
<http://www.wolframalpha.com/>
- Wikipedia
https://en.wikipedia.org/wiki/Main_Page

On Campus Resources

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

Summary

- A good infographic should be useful, sound, and beautiful.
- Best practices
 - Information Design
 - Keep it simple
 - Use a simple text combined with a relevant image
 - Make it unique
 - Data Visualization
 - Bar Chart – for ranking and time series, starting with zero baseline, avoid 3-D
 - Pie Chart – for part-to-whole comparisons, limiting to 5 slices, avoid 3-D
 - Line Chart – for time series, limiting to 4 or less
- Use color schemes that are color-blind friendly
- Tools
 - PowerPoint, Excel, Illustrator, InDesign, Gephi, Photoshop, Gimp
 - Wordle, Google Chart, Tableau Public



Infographics, Communicate Information with Graphics