Session 72, Data Visualization as a Tool in Healthcare Data Analytics

2019 SOA Health Meeting

ELENA BLACK, PHD, CFA, FSA, EA, MAAA, FCA

YI-LING LIN, FSA, MAAA, FCA

Session 072, Data Visualization as a Tool in Healthcare Data Analytics

June 25, 2019

SOCIETY OF ACTUARIES®

SOCIETY OF ACTUARIES
Antitrust Compliance Guidelines

Active participation in the Society of Actuaries is an important aspect of membership. While the positive contributions of professional societies and associations are well-recognized and encouraged, association activities are vulnerable to close antitrust scrutiny. By their very nature, associations bring together industry competitors and other market participants.

The United States antitrust laws aim to protect consumers by preserving the free economy and prohibiting anti-competitive business practices; they promote competition. There are both state and federal antitrust laws, although state antitrust laws closely follow federal law. The Sherman Act, is the primary U.S. antitrust law pertaining to association activities. The Sherman Act prohibits every contract, combination or conspiracy that places an unreasonable restraint on trade. There are, however, some activities that are illegal under all circumstances, such as price fixing, market allocation and collusive bidding.

There is no safe harbor under the antitrust law for professional association activities. Therefore, association meeting participants should refrain from discussing any activity that could potentially be construed as having an anti-competitive effect. Discussions relating to product or service pricing, market allocations, membership restrictions, product standardization or other conditions on trade could arguably be perceived as a restraint on trade and may expose the SOA and its members to antitrust enforcement procedures.

While participating in all SOA in person meetings, webinars, teleconferences or side discussions, you should avoid discussing competitively sensitive information with competitors and follow these guidelines:

- **Do not** discuss prices for services or products or anything else that might affect prices
- **Do not** discuss what you or other entities plan to do in a particular geographic or product markets or with particular customers.
- **Do not** speak on behalf of the SOA or any of its committees unless specifically authorized to do so.
- **Do** leave a meeting where any anticompetitive pricing or market allocation discussion occurs.
- **Do** alert SOA staff and/or legal counsel to any concerning discussions
- **Do** consult with legal counsel before raising any matter or making a statement that may involve competitively sensitive information.

Adherence to these guidelines involves not only avoidance of antitrust violations, but avoidance of behavior which might be so construed. These guidelines only provide an overview of prohibited activities. SOA legal counsel reviews meeting agenda and materials as deemed appropriate and any discussion that departs from the formal agenda should be scrutinized carefully. Antitrust compliance is everyone's responsibility; however, please seek legal counsel if you have any questions or concerns.

Presentation Disclaimer

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the participants individually and, unless expressly stated to the contrary, are not the opinion or position of the Society of Actuaries, its cosponsors or its committees. The Society of Actuaries does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented. Attendees should note that the sessions are audio-recorded and may be published in various media, including print, audio and video formats without further notice.

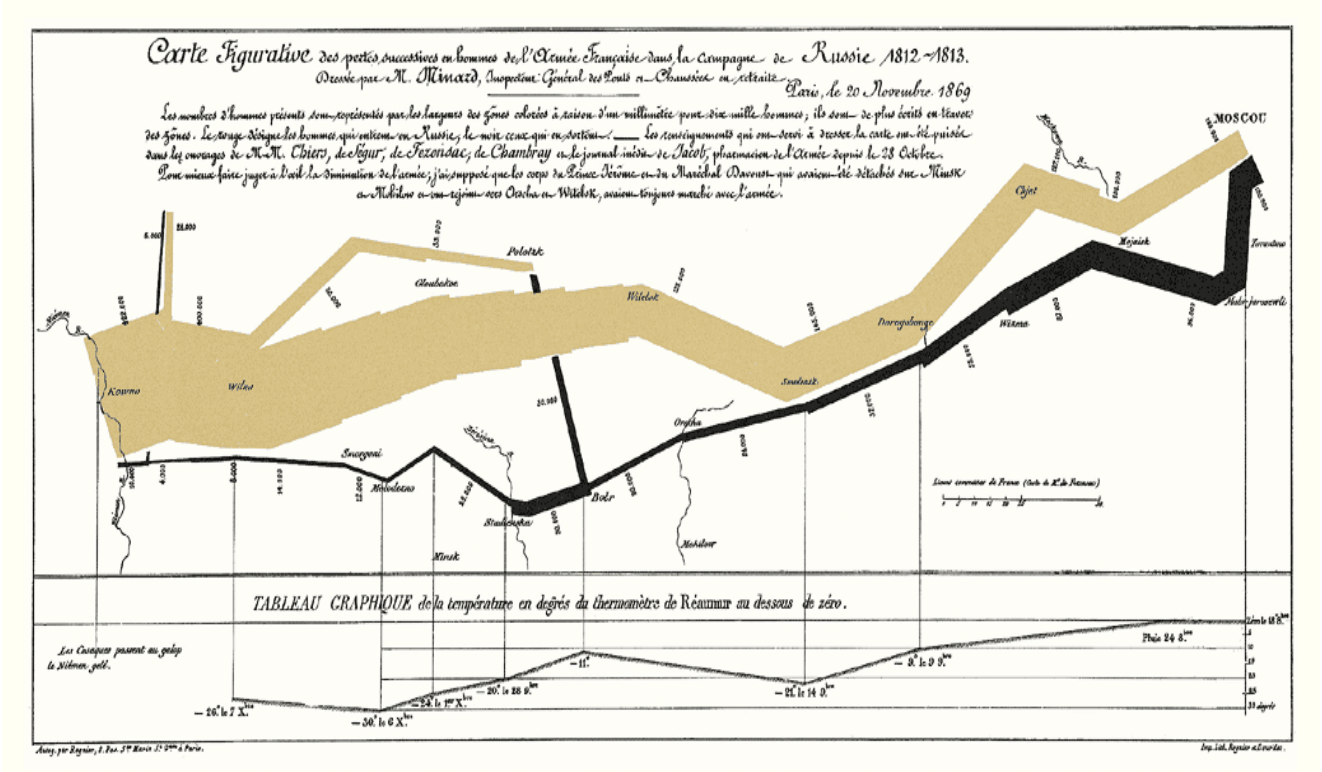
Primary goal of data visualization is storytelling….

Storytelling with data at it’s finest

SOCIETY OF ACTUARIES®

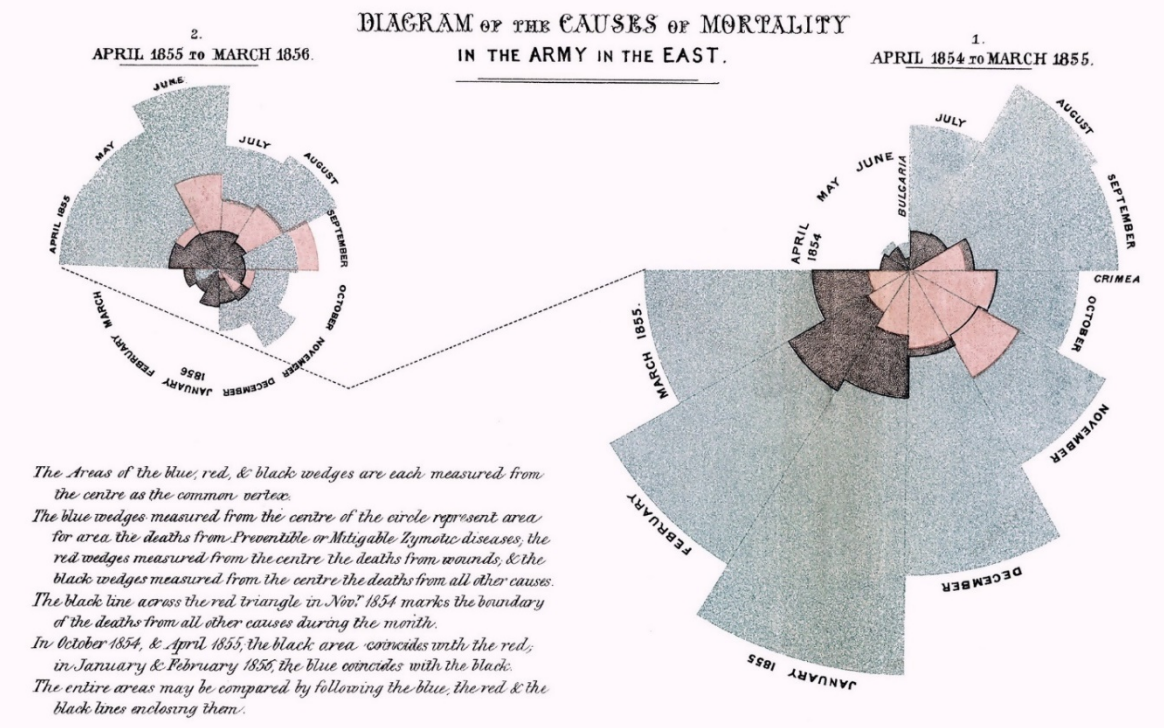
Napoleon’s March to Moscow 1812

Minard’s work is an anti-war poster…



Focus on the human cost of war is subtly reinforced by his choice of content… The word “Napoleon” does not appear on the map of Napoleon’s march…

Causes of Mortality in the Crimean War



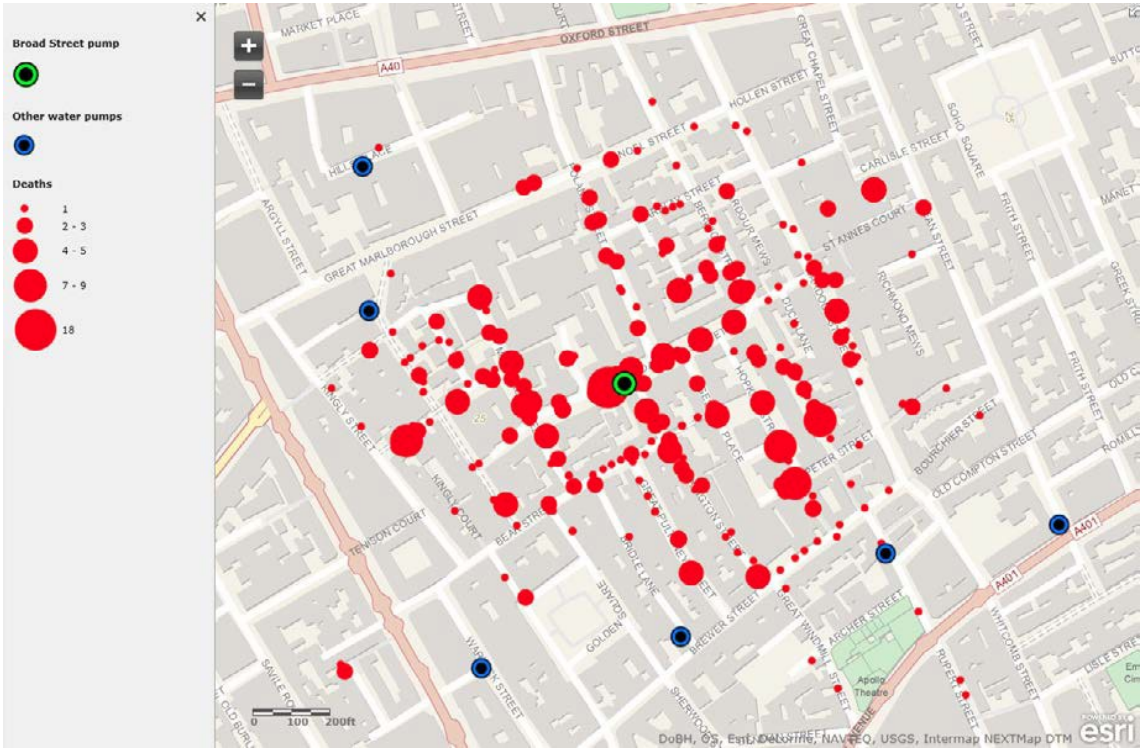
Statistics and sanitary reform

- Florence Nightingale (1858)
- Pioneer in visual presentation of information and statistical graphics
- “Nightingale rose diagram” (equivalent to modern circular histogram)
- Sources of patient mortality in military field hospital she managed—reveals that majority of deaths are due to poor hospital practices
- Report nature and magnitude of conditions of medical care to members of Parliament and civil servants (unlikely audience for traditional statistical reports)

https://en.wikipedia.org/wiki/Florence_Nightingale#Crimean_War

Cholera Outbreak in London 1854

Early use of graphics in data analytics and hypothesis testing…



- Use of visualization in diagnostic data analytics…?
- Hypothesis testing…?
- Theory depiction to convince others…?
- Modern version of Dr. Snow's map to identify cause of the outbreak…

Visualization – indispensable tool in data analysis….

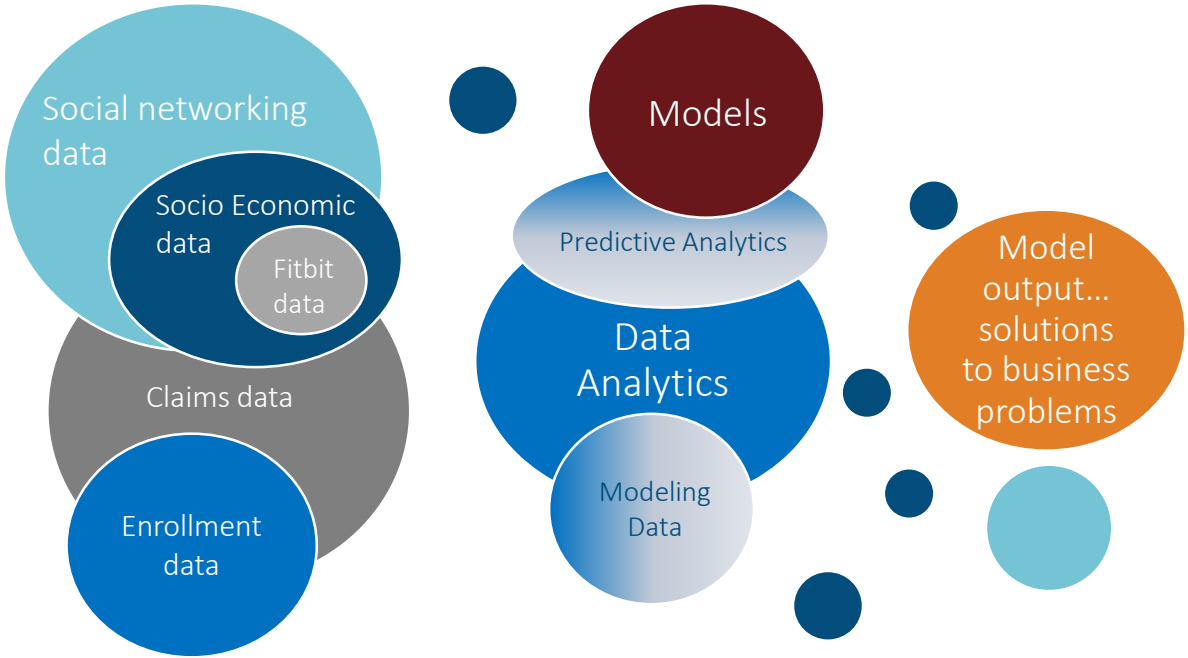
Data visuals as tools in data analytics

Let the data tell the story…

SOCIETY OF ACTUARIES®

Storytelling with Data

Typically data visualization discussions focus on telling story you already know…



Spectrum of storytelling:

- Integrity/Quality of Data – data preprocessing
- Data exploration and feature engineering
- Model validation and evaluation
- Understanding results
- Relaying modeling results to different audiences (your clients and friends!)

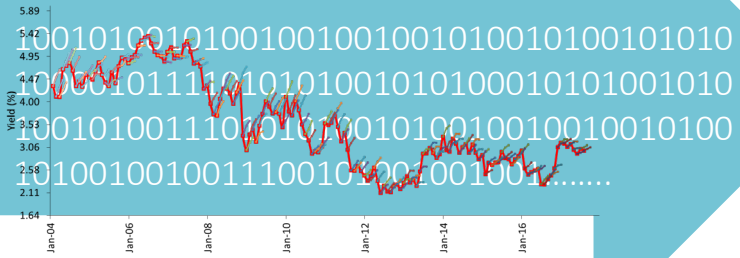
Spectrum of storytelling – data visualization at different levels and purpose when working with data

Two Types of Data Visualization

Exploratory

Understanding data & letting it tell the story

Visualization tools



Explorer

Aha!

Explanatory

110101010010100
100100100101001
01001001001111...

Presenter

Message from our analysis to others

Key Point

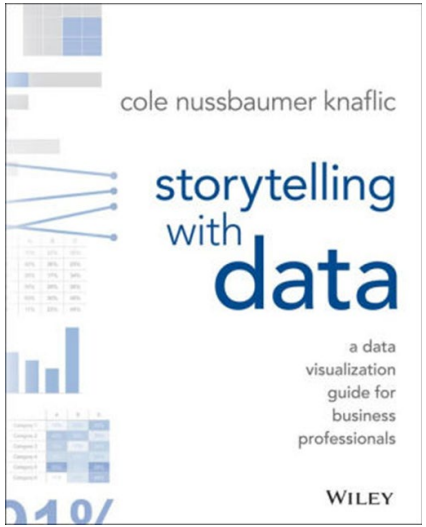
Audience

I get it!

An Explanatory Data Visualization Guide

Six key lessons for explanatory data visualization…

1. Understand the context
2. Choose an appropriate visual display
3. Eliminate clutter
4. Focus audience attention
5. Think like a designer
6. Tell a story



We focus on exploratory visualization… rules maybe somewhat different…

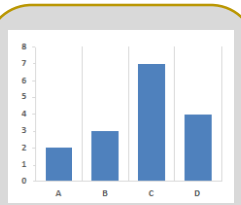
Spectrum of Data Analytics

Data visualization is a powerful tool at each stage:
Aha! moments and stories to tell…

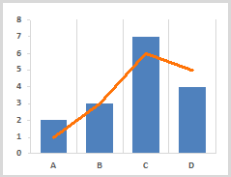
Typically not part of analytics process… but with Big Data…maybe it is!!!!!

Data preprocessing

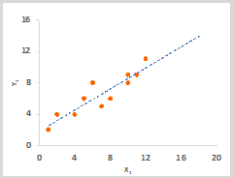
Value towards business solutions



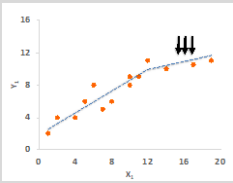
What happened?



Why did it happen?



What will happen?



How do I influence what will happen?

Descriptive

Diagnostic

Predictive

Prescriptive
…or decision-informing

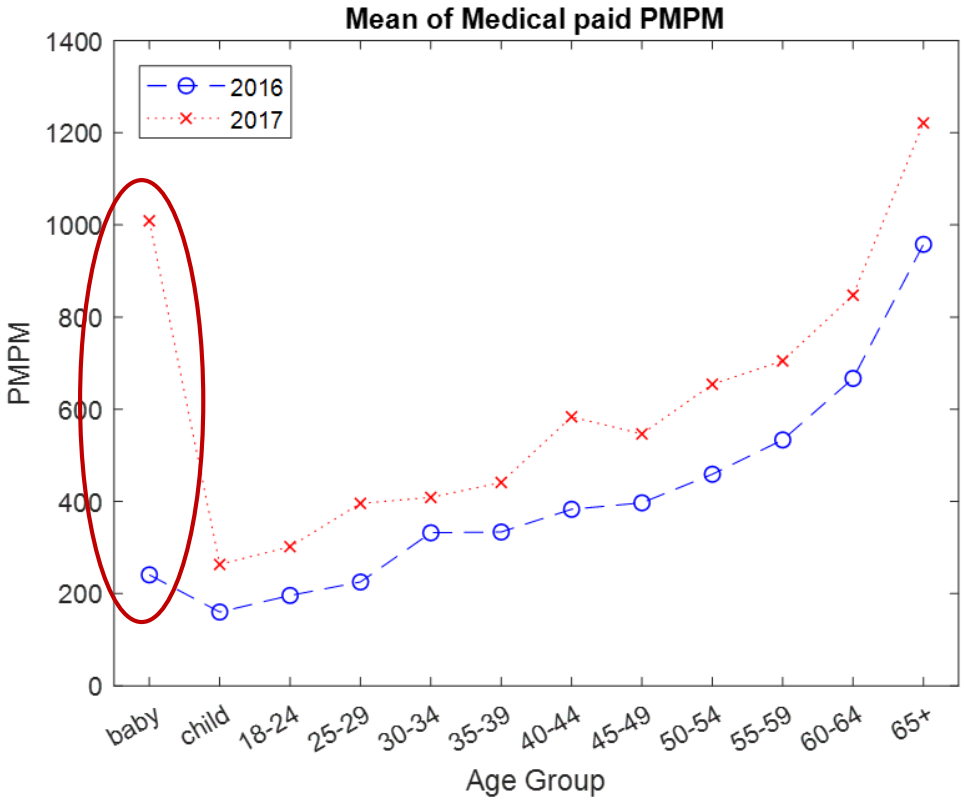
Analytical sophistication

Spectrum of data analytics: hindsight to insight to foresight

Adapted from Gartner's Data Analytics Maturity Model

Preprocessing and Data Validation

Reasonability and quality of data, potential missing values, outliers and more….



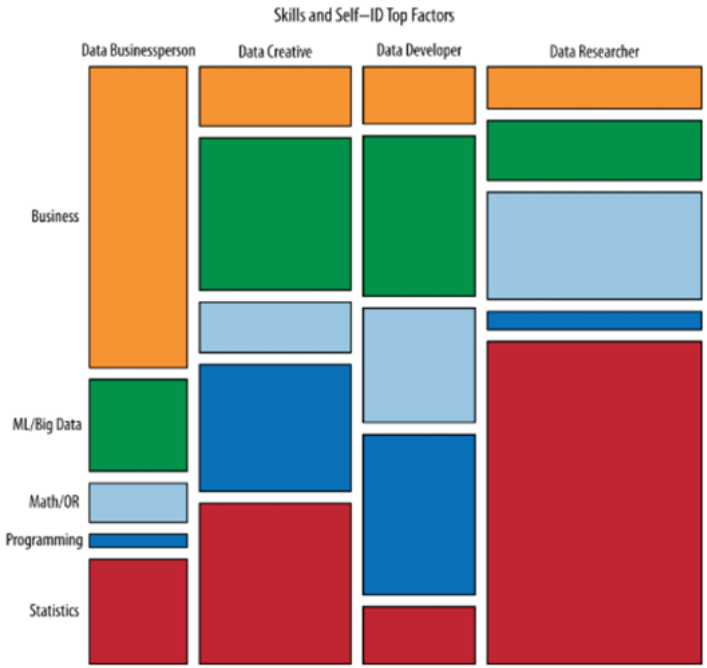
- Bad data?
- Missing data?
- Reasonable patterns?
- Outliers – real or mistakes?
- Processing, connecting and reconciling of complicated data: setting check points

…and much more…

Data Exploration & Feature Engineering

Investigating relationships and trends… before diving into modeling…

Categorical versus Categorical: Example of Mosaic Plot



https://jeremiahstanghini.com/2017/07/30/what-is-data-science/

		Input/Feature	
		Categorical	Numeric
Target	Categorical	Mosaic plots	Box plots
	Numeric	Density plots	Scatter

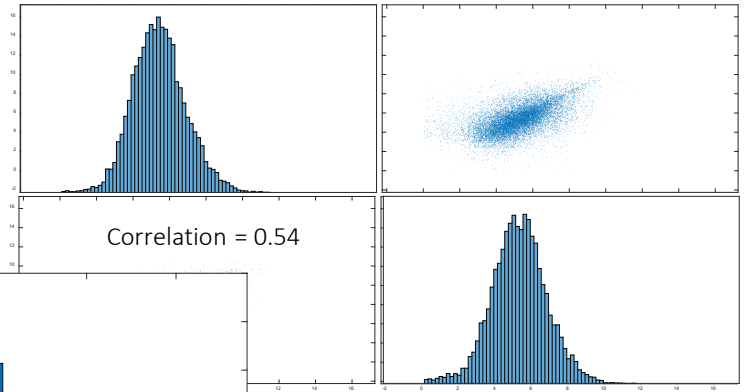
- Relationships between variables
- Relationships between target and predictor variables
- Hypothesis testing
- Identifying patterns and trends
- Feature generation
- “Aha” moments…
- Data possibly revealing new stories…
- But also story to tell on what happened… to clients and friends!

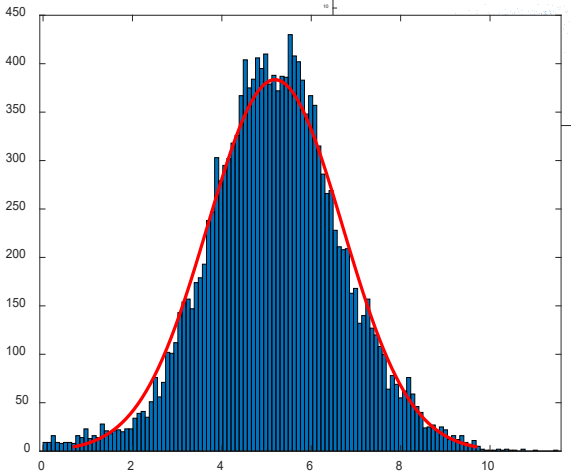
Trend and Relationship Identification

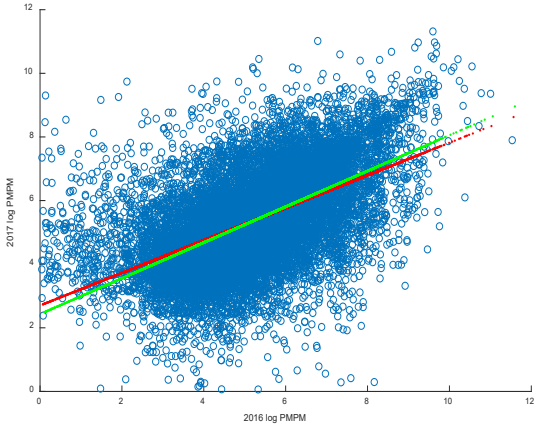
Histograms and scatter plots…. Tools to trends and relationships

Plot matrix:

- Histograms on diagonal
- Pairwise scatter plots
- Correlations in lower triangle







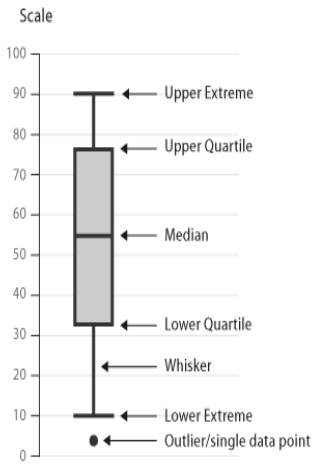
Claim costs are lognormally distributed; prior and current somewhat linearly correlated

Case study for illustration purposes only

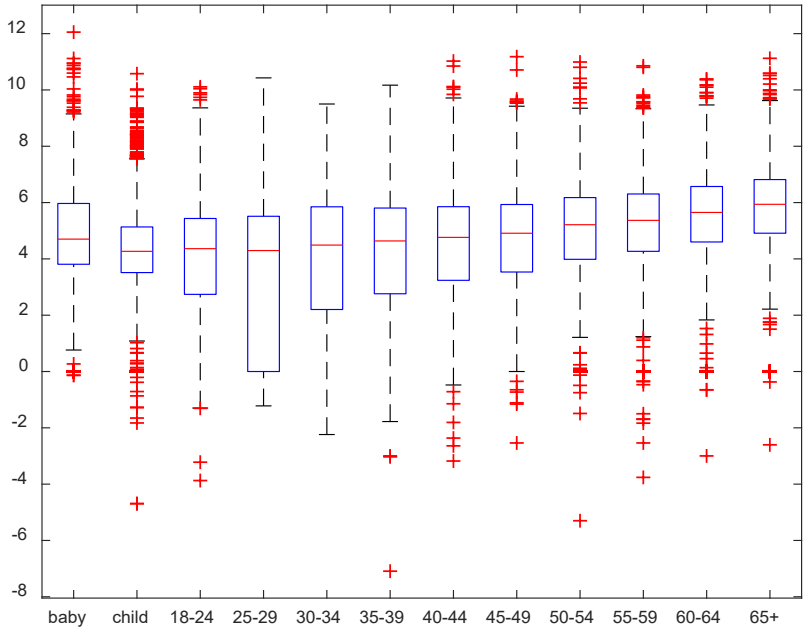
Box Plot to Visualize Distributions

Box plots provide for a good way to visualize and compare distributions

Reading Box Plots



Log PMPM distributions by age group

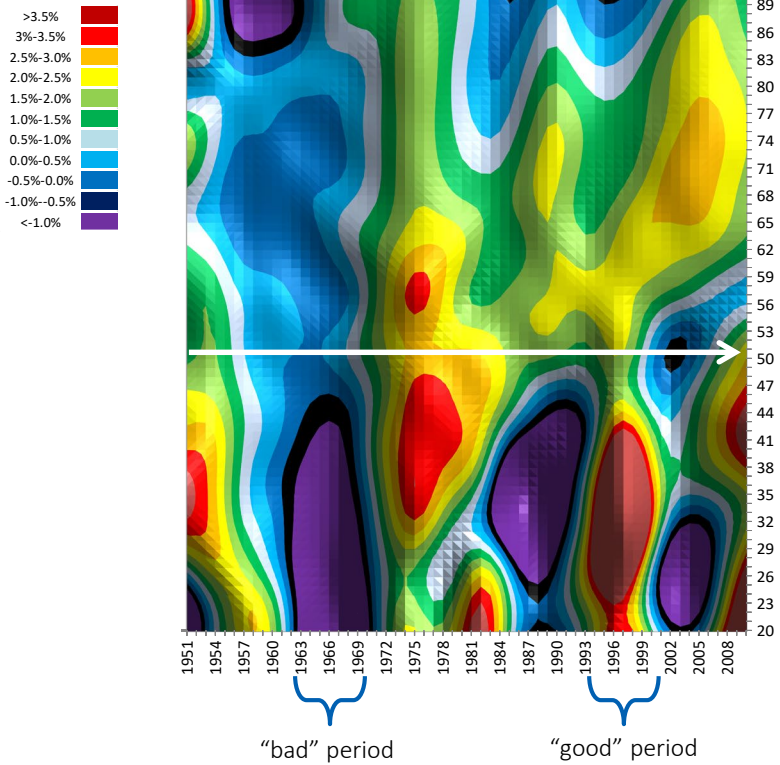


- Box plots (aka “box and whisker” plots): visualizing distribution, central value, variability, and outliers
- Median, 25th and 75th percentiles form the box
- Whiskers extends to “extreme” values not considered outliers
- Outliers are identified as single points
- Tool to compare distributions by categorical variable

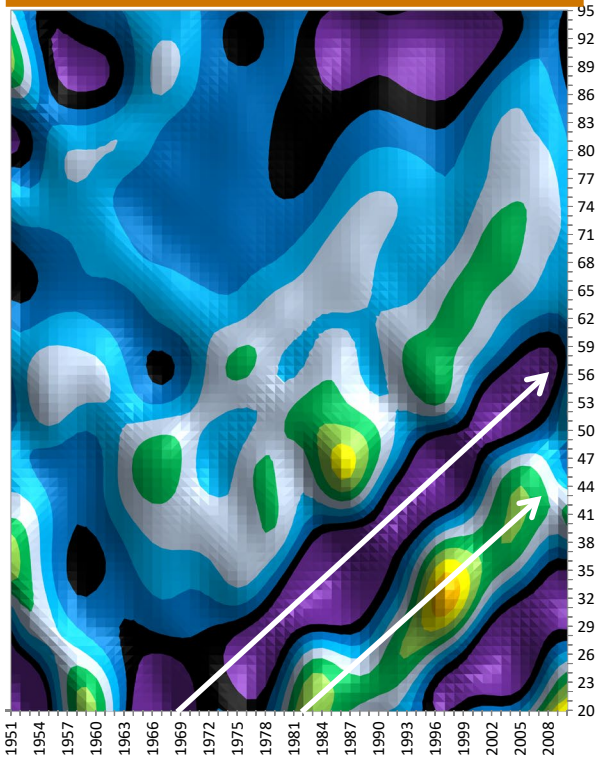
Case study for illustration purposes only

Hypothesis Testing: US Mortality Improvement

Male Improvement Rates 1951-2010



Male Cohort Component 1951-2010



- Period effects - vertical "stripes" (improved road conditions, medical advances, epidemics)
- Age effect - horizontal patterns
- Cohort effect - diagonal patterns (smoking, obesity)
- These effects overlap: hard to see on total MI rates' map

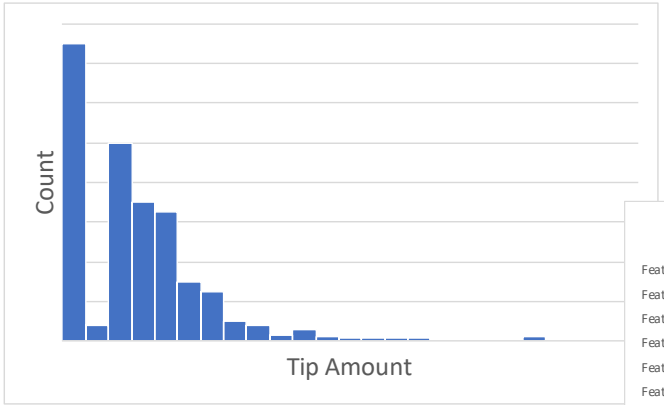
Age group younger than 50 → more pronounced period effect

Case study for illustration purposes only

“Taxi Rides and Tips” Story

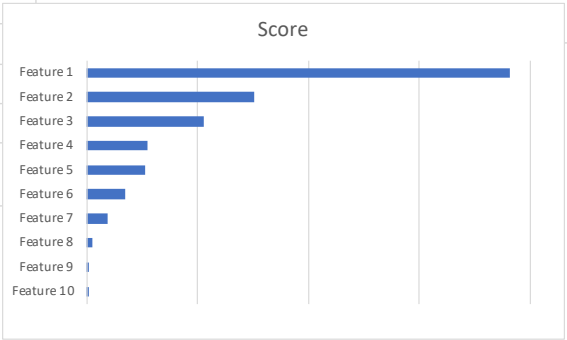
Instructive example of power of visuals…

NYC taxi trip and fare information: original tip distribution

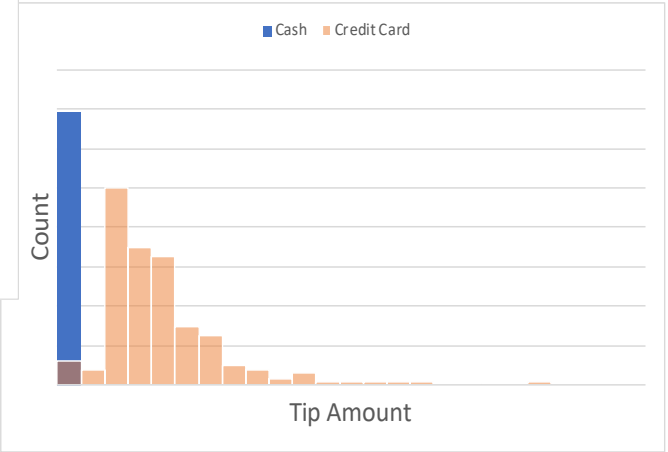


Too-good-to-be-true scenario …. Is probably not true…

Feature Importance Graph



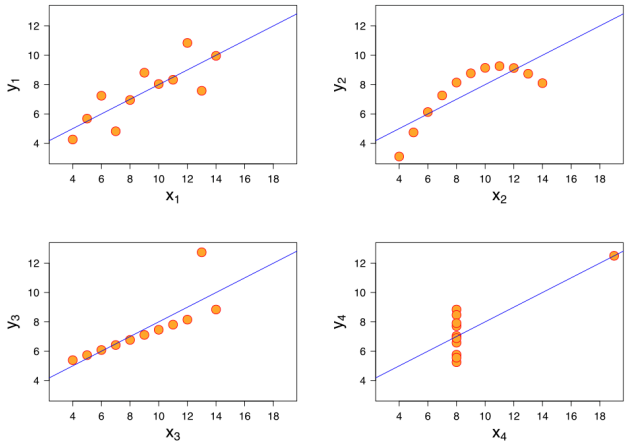
Overlaid histograms of tip distribution by payment type



- Detailed NYC taxi trip records from every taxi trip in 2013
- Many categorical and numeric variables
- Classification Problem: predicting if there would be a tip…

Model Evaluation and Validation

Model evaluation is an important part of any modeling project



Cautionary tale!

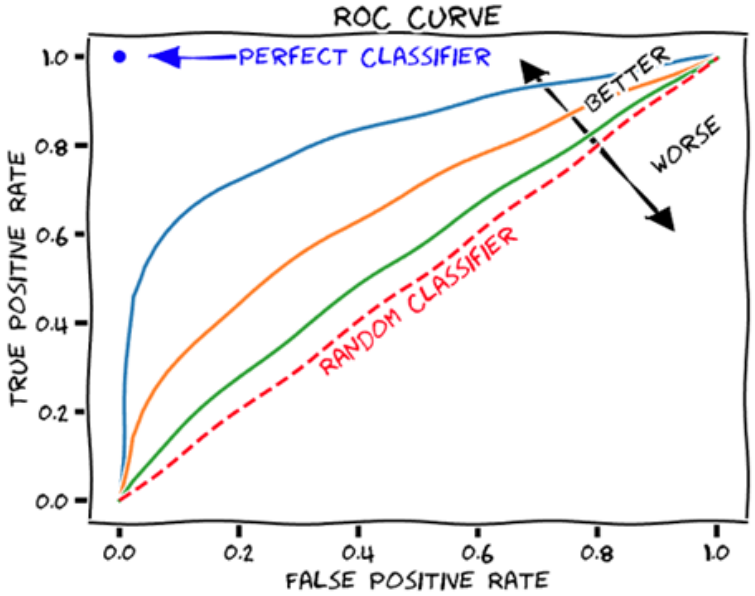
Famous Anscombe's quartet: all four datasets have the same statistical properties, including R squared=0.67, means and variance of x and y, correlation and linear regression model: y=3+0.5x

- Appropriate and consistent with purpose
- Visualization used to validate, evaluate, compare, understand modeling results
- Examples of statistical criteria/metrics
  - ✓ Standard statistical measures (R squared, RMSE, MAE, etc.)
  - ✓ Predictive Ratios (E/A) on groups of interest (e.g. Diagnostic groups or age groups)
  - ✓ Tolerance curves
  - ✓ Sensitivity and specificity (confusion matrix)
  - ✓ ROC curves
  - ✓ Comparison with naïve and standard models

  …and many more

Evaluation of Classifier: ROC Curve

Visualizing model performance, selecting threshold and comparing models



- Example of insurance application: binary classifier
  - ✓ Smoker status identification for life underwriting triage
- ROC = Receiver Operator Curve
  - ✓ Trade-off determined by “probability threshold”
  - ✓ “Elbow” point but also relative importance by type of error
  - ✓ Area under curve (AUC) – comparing models
  - ✓ Sensitivity and specificity (confusion matrix)
  - ✓ Accuracy of classifier – indifferent to types of errors

Questions?

SOCIETY OF ACTUARIES®

SOCIETY OF ACTUARIES®

Data Visualization as a Tool in Healthcare Data Analytics

TRACY ROTH, MBA

Session 072, 2019 SOA Health Meeting

June 25, 2019

SOCIETY OF ACTUARIES®

SOCIETY OF ACTUARIES
Antitrust Compliance Guidelines

Active participation in the Society of Actuaries is an important aspect of membership. While the positive contributions of professional societies and associations are well-recognized and encouraged, association activities are vulnerable to close antitrust scrutiny. By their very nature, associations bring together industry competitors and other market participants.

The United States antitrust laws aim to protect consumers by preserving the free economy and prohibiting anti-competitive business practices; they promote competition. There are both state and federal antitrust laws, although state antitrust laws closely follow federal law. The Sherman Act, is the primary U.S. antitrust law pertaining to association activities. The Sherman Act prohibits every contract, combination or conspiracy that places an unreasonable restraint on trade. There are, however, some activities that are illegal under all circumstances, such as price fixing, market allocation and collusive bidding.

There is no safe harbor under the antitrust law for professional association activities. Therefore, association meeting participants should refrain from discussing any activity that could potentially be construed as having an anti-competitive effect. Discussions relating to product or service pricing, market allocations, membership restrictions, product standardization or other conditions on trade could arguably be perceived as a restraint on trade and may expose the SOA and its members to antitrust enforcement procedures.

While participating in all SOA in person meetings, webinars, teleconferences or side discussions, you should avoid discussing competitively sensitive information with competitors and follow these guidelines:

- **Do not** discuss prices for services or products or anything else that might affect prices
- **Do not** discuss what you or other entities plan to do in a particular geographic or product markets or with particular customers.
- **Do not** speak on behalf of the SOA or any of its committees unless specifically authorized to do so.
- **Do** leave a meeting where any anticompetitive pricing or market allocation discussion occurs.
- **Do** alert SOA staff and/or legal counsel to any concerning discussions
- **Do** consult with legal counsel before raising any matter or making a statement that may involve competitively sensitive information.

Adherence to these guidelines involves not only avoidance of antitrust violations, but avoidance of behavior which might be so construed. These guidelines only provide an overview of prohibited activities. SOA legal counsel reviews meeting agenda and materials as deemed appropriate and any discussion that departs from the formal agenda should be scrutinized carefully. Antitrust compliance is everyone's responsibility; however, please seek legal counsel if you have any questions or concerns.

Presentation Disclaimer

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the participants individually and, unless expressly stated to the contrary, are not the opinion or position of the Society of Actuaries, its cosponsors or its committees. The Society of Actuaries does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented. Attendees should note that the sessions are audio-recorded and may be published in various media, including print, audio and video formats without further notice.

Data engineering, the key to accurate and efficient visualizations

Architecting the Framework

SOCIETY OF ACTUARIES®

Architecting the Framework

Design Precedes Visualization

- Establishing purpose – descriptive, diagnostic, predictive, or prescriptive

- Identifying content and methods needed – accessible, accurate, timely, focused

- Engineering data – summary to drill

- Knowing the user – depth of knowledge in content and tool, determines selection

Architecting the Framework

Identifying the Content – Accessible, Accurate, Timely, Focused

Foundation

- Eligibility
- Medical Claims
- Pharmacy Claims
- Provider Demographics

Status Quo

- Lab Data
- State Immunization Data
- Gaps in Care
- Program Participation
- Risk Scoring (Diagnoses)

Best in Class

- ADT
- Quality and Practice Patterns of Providers
- Social Determinants of Health
- Stage of Chronic Disease
- Safety Data
- Anticipation of Health Need

Architecting the Framework

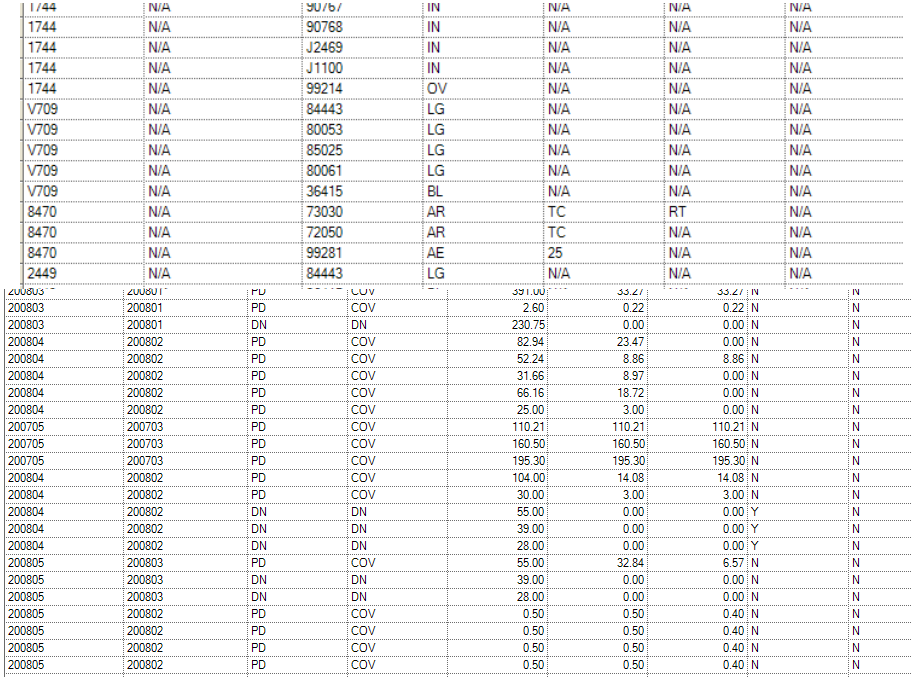
Engineering the Data – Definitions

- Detail drill vs efficiency
- Refine definitions according to strategy e.g. “engagement”, “low acuity”
- Setting targets and using benchmarks effectively

Architecting the Framework

Engineering the Data – Structure

- Level of data needed
- Detail drill vs efficiency tradeoff
- Blending content for dynamic rate calculation
- Binary, categorical, discrete, continuous…



1744	N/A	90767	IN	N/A	N/A	N/A		
1744	N/A	90768	IN	N/A	N/A	N/A		
1744	N/A	J2469	IN	N/A	N/A	N/A		
1744	N/A	J1100	IN	N/A	N/A	N/A		
1744	N/A	99214	OV	N/A	N/A	N/A		
V709	N/A	84443	LG	N/A	N/A	N/A		
V709	N/A	80053	LG	N/A	N/A	N/A		
V709	N/A	85025	LG	N/A	N/A	N/A		
V709	N/A	80061	LG	N/A	N/A	N/A		
V709	N/A	36415	BL	N/A	N/A	N/A		
8470	N/A	73030	AR	TC	RT	N/A		
8470	N/A	72050	AR	TC	N/A	N/A		
8470	N/A	99281	AE	25	N/A	N/A		
2449	N/A	84443	LG	N/A	N/A	N/A		
200803	200801	PD	COV	391.00	33.27	33.27	N	N
200803	200801	PD	COV	2.60	0.22	0.22	N	N
200803	200801	DN	DN	230.75	0.00	0.00	N	N
200804	200802	PD	COV	82.94	23.47	0.00	N	N
200804	200802	PD	COV	52.24	8.86	8.86	N	N
200804	200802	PD	COV	31.66	8.97	0.00	N	N
200804	200802	PD	COV	66.16	18.72	0.00	N	N
200804	200802	PD	COV	25.00	3.00	0.00	N	N
200705	200703	PD	COV	110.21	110.21	110.21	N	N
200705	200703	PD	COV	160.50	160.50	160.50	N	N
200705	200703	PD	COV	195.30	195.30	195.30	N	N
200804	200802	PD	COV	104.00	14.08	14.08	N	N
200804	200802	PD	COV	30.00	3.00	3.00	N	N
200804	200802	DN	DN	55.00	0.00	0.00	Y	N
200804	200802	DN	DN	39.00	0.00	0.00	Y	N
200804	200802	DN	DN	28.00	0.00	0.00	Y	N
200805	200803	PD	COV	55.00	32.84	6.57	N	N
200805	200803	DN	DN	39.00	0.00	0.00	N	N
200805	200803	DN	DN	28.00	0.00	0.00	N	N
200805	200802	PD	COV	0.50	0.50	0.40	N	N
200805	200802	PD	COV	0.50	0.50	0.40	N	N
200805	200802	PD	COV	0.50	0.50	0.40	N	N
200805	200802	PD	COV	0.50	0.50	0.40	N	N

Enabling the Strategy

SOCIETY OF ACTUARIES®

Enabling the Strategy

Design Precedes Visualization

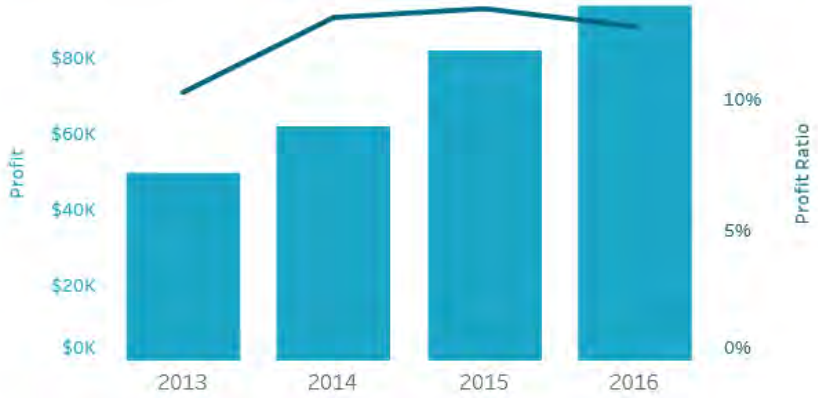
- Establishing purpose – descriptive, diagnostic, predictive, or prescriptive
- Identifying content needed – accessible, accurate, timely, focused
- Engineering data – summary to drill
- Knowing the user – depth of knowledge in content and tool, determines selection

Enabling the Strategy

Know the Need, Know the User, Create the Experience

COLUMN + LINE TRENDLINE

A good way of showing the relationship over time between an amount (columns) and a rate (line).



CIRCLE TIMELINE

Good for showing discrete values of varying size across multiple categories (i.e., earthquakes by continent).

Enabling the Strategy

Know the Need, Know the User, Create the Experience

RADAR

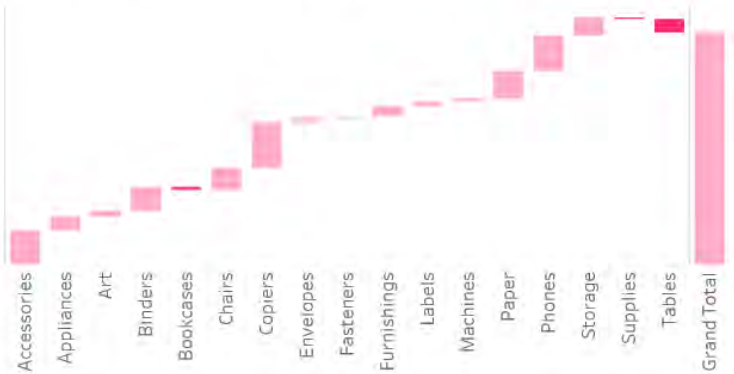
POLAR CHART, SPIDER CHART, START CHART, OR WEB CHART

A space-efficient way of showing value of multiple variables -- but make sure they are organized in a way that makes sense to the reader.

WATERFALL CHART

BRIDGE, FLYING BRICKS CHART, AND MARIO CHART

Can be useful for showing part-to-whole relationships where some of the components are negative.



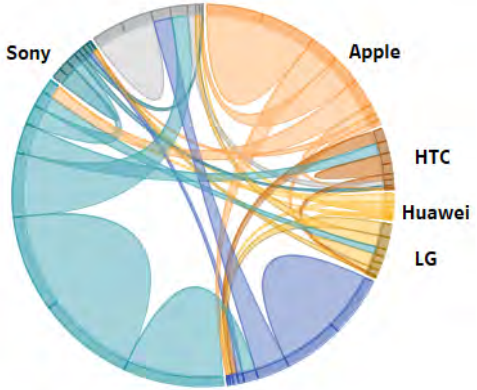
Enabling the Strategy

Know the Need, Know the User, Create the Experience

CHORD

RADIAL NETWORK DIAGRAM

A complex but powerful diagram which can illustrate 2-way flows (and net winner) in a matrix.



CHORD DIAGRAM: https://bit.ly/2BiOZQS

FAN CHART (PROJECTIONS)

Use to show the uncertainty in future projections -- usually this grows the further forward to projection.

Enabling the Strategy

Know the Need, Know the User, Create the Experience

- Consistency in Design – left to right, filters, conditional highlighting
- Descriptions over Codes – where user is not familiar with codes
- Clarity of Visual Purpose – focused insights
- Recurrence – tracking mechanisms and reforecasting, user guides

Questions?

SOCIETY OF ACTUARIES®


SOCIETY OF
ACTUARIES®