

OVERCOMING CHARTAPHOBIA

Moving Your Organization Toward Interesting & Enlightening Data
Visualization

Meagan Longoria

GETTING STARTED

- Slides are on my blog (or the SQL Saturday site):
<http://datasavvy.wordpress.com/presentations>
- Feel free to share questions and comments throughout the presentation



ABOUT ME

- Meagan Longoria
- Solutions Consultant with BlueGranite
- Blog: <http://datasavvy.wordpress.com>
- Twitter: [@mmarie](https://twitter.com/mmarie)
- LinkedIn: www.linkedin.com/in/meaganlongoria/

HOW OFTEN HAVE YOU HAD TO BUILD THIS?

0,014	0,929	0,069	4,858	0,124	8,786
0,015	1,000	0,07	4,929	0,125	8,858
0,016	1,072	0,071	5,001	0,126	8,929
0,017	1,144	0,072	5,072	0,127	9,001
0,018	1,215	0,073	5,144	0,128	9,072
0,019	1,286	0,074	5,215	0,129	9,144
0,02	1,358	0,075	5,286	0,13	9,215
0,021	1,429	0,076	5,358	0,131	9,286
0,022	1,501	0,077	5,429	0,132	9,358
0,023	1,572	0,078	5,501	0,133	9,429
0,024	1,644	0,079	5,572	0,134	9,501
0,025	1,715	0,08	5,644	0,135	9,572
0,026	1,786	0,081	5,715	0,136	9,644
0,027	1,858	0,082	5,786	0,137	9,715
0,028	1,929	0,083	5,858	0,138	9,786
0,029	2,000	0,084	5,929	0,139	9,858
0,03	2,072	0,085	6,000	0,14	9,929
0,031	2,144	0,086	6,072	0,141	10,000
0,032	2,215	0,087	6,144		
0,033	2,286	0,088	6,215		
0,034	2,358	0,089	6,286		
0,035	2,429	0,09	6,358		
0,036	2,501	0,091	6,429		
0,037	2,572	0,092	6,501		
0,038	2,644	0,093	6,572		
0,039	2,715	0,094	6,644		
0,04	2,786	0,095	6,715		
0,041	2,858	0,096	6,786		
0,042	2,929	0,097	6,858		
0,043	3,001	0,098	6,929		
0,044	3,072	0,099	7,001		
0,045	3,144	0,1	7,072		

(Plain Tabular Report)

OR THIS?





Mark Vaillancourt
@markvsq1



Following

Chartaphobia: When business insist on relying solely on numbers arranged in tables to track performance. [#DataViz](#)
[#Epidemic](#)



RETWEETS
2

FAVORITE
1



1:40 PM - 20 Feb 2015



Reply to @markvsq1



Meagan Longoria @mmarie · Feb 20

@markvsq1 Yes! I'm going to start calling it Chartaphobia from now on. Love it.



Mark Vaillancourt @markvsq1 · Feb 20

@mmarie I'm looking at a few dozen spreadsheets for a client: ALL MASSIVE TABLES OF NUMBERS. Not ONE chart to actually SEE the story.



WHAT'S THE BIG DEAL?

You are doing yourself and your customers (internal or external) a disservice if you always just make a table of data.

Our (developers') outputs are decision-makers' inputs – and **their** outputs are what ultimately matter. – Rob Collie (Power Pivot Pro)

The greatest value of a picture is when it forces us to notice what we never expected to see. - John Tukey.

“Don't let your design choices be happenstance; rather, they should be the result of explicit decisions” - Cole Nussbaumer Knaflic

WHY DO WE STRUGGLE WITH DATA VIZ?

Lack of appropriate data

Reports as intermediate steps

Poor presentation that makes it difficult to gain insight and take action

Poor presentation that discourages engagement

SNOW CHOLERA MAP



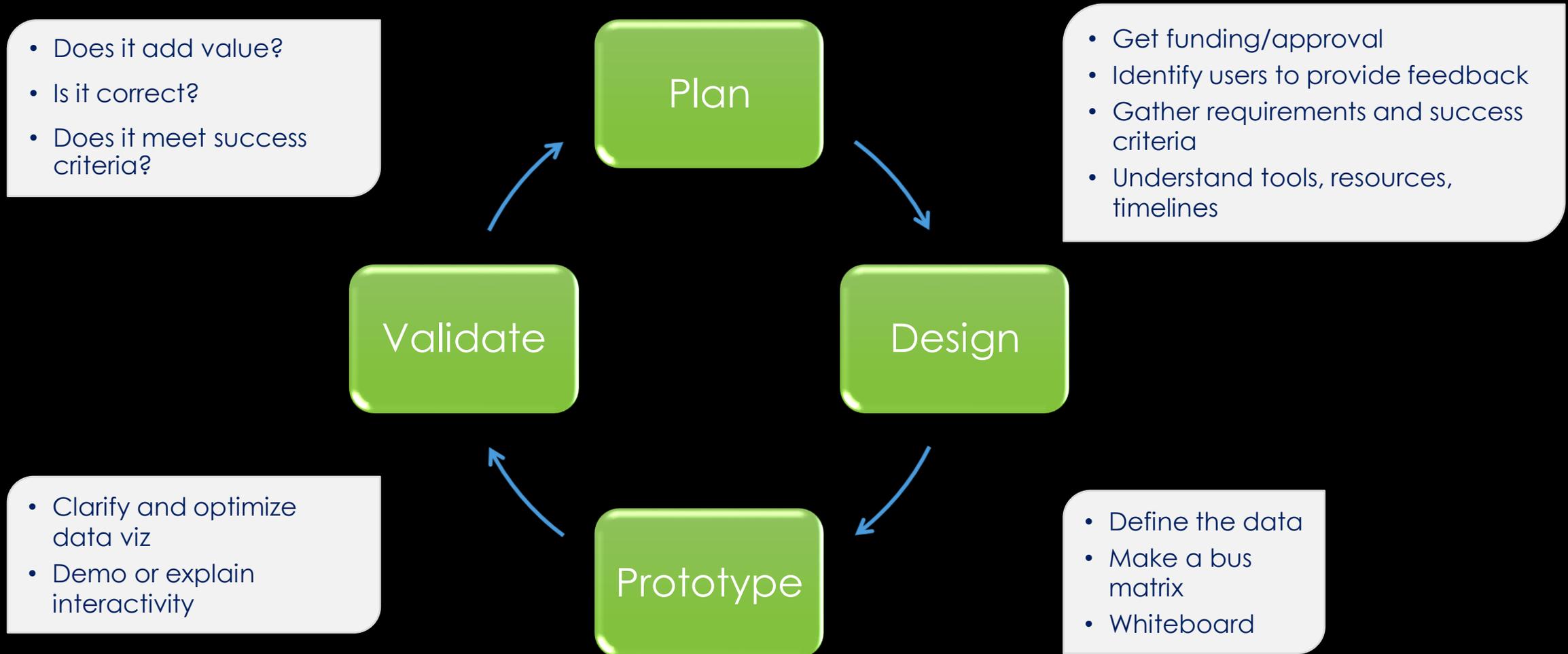
GET THE SCOOP

- Who is your audience? (Executives? Analysts? Website users?)
 - Helps determine needs, priorities, and level of detail
- What metrics are important? What is the dimensionality?
- Is the report operational, analytical, or a mix of both?

YOU'LL NEVER GUESS WHAT HAPPENS NEXT

- What do they do with the data/information?
 - Sometimes reports are step 1 in a process. What comes next?
- Borrow a page from the 5 Whys
 - We're developers. We automate repetitive and tedious things.
 - Can we add more value? Add predictive or prescriptive capabilities? Push alerts?

BEGIN THE PROCESS



HOW DO WE DO IT?

1. Understand the context and craft your message
2. Choose an appropriate visual display
3. Eliminate clutter
4. Focus attention where you want it

(First 4 steps from Storytelling With Data book)

CONTEXT/MESSAGE

- Who is your audience?
- What do you want your audience to know or do?
- How can you use data to help make your point?

“Know that even if you highlight or recommend the wrong thing, it prompts the right sort of conversation focused on action... If you simply present data, it's easy for your audience to say ‘Oh, that's interesting’ and move on to the next thing. But if you ask for action, your audience has to make a decision whether to comply or not.” - Cole Nussbaumer Knaflic

DESIGN: DATA DEFINITION

Dimension information	
Name	a simple concise name to help users identify it
Definition	a non-technical definition that explains what is being measured
Data Source(s)	the source system that contains the underlying data (not necessarily where you will query it)
Important/Commonly Used Attributes	the field the user slices by or pivots on to get useful information
Time reference	indicates if the dimension data is as originally entered, current state, or historical (type 0, 1, or 2)

KPI/Metric definition	
Name	a simple concise name to help users identify it
Definition	a non-technical definition that explains what is being measured
Related Business Objective	ties the KPI to an overall business objective so users understand the reason for the measurement
Unit Type	count, dollar amount, percent, etc.
Frequency of Measurement	indicates how often the underlying data is gathered and over what period of time
Data Source(s)	the source system that contains the underlying data (not necessarily where you will query it)
Target Goal	the goal against which performance is measured
Calculation	a technical definition of the metric/KPI (could be similar to an Excel formula)
Statuses	explanation of the possible states (red/yellow/green, 1 - 5, etc.)
Related Metrics or KPIs	any other KPIs that are related as a parent, child, or sibling of the metric/KPI

DESIGN: BUS MATRIX

- Beware of inferred relationships in flat data sets
- Think like the user in terms of relationships and business attributes

Facts/Dimensions	Date	Item	Vendor	Distr Center	Shipper	Store	Customer	Promo
Purchase Orders	x	x	x	x				
Distribution Center Delivery	x	x	x	x	x			
Distribution Center Inventory	x	x		x				
Store Deliveries	x	x		x	x	x		
Store Inventory	x	x				x		
Store Sales	x	x				x	x	x
Returns	x	x				x	x	x

DESIGN: VIZ DESIGN DOCUMENTATION

No long requirements docs. Life is too short and the business changes too quickly.

Viz Metadata	
Report Title	a simple, concise name to help users identify the viz
Originator	person who owned the previous tabular report (from a business perspective)
Purpose	reason why user would consume the report/viz
Activity That Triggers the Report	ex: user visits page on website, financial quarter close process, etc.
Type of Data	ex: headcount by department, dollars spent by category, etc.
Filter Criteria	ex: year and week, department, category
Drillthrough or Linking	explain any drillthrough action or links to external resources
Used By	audience that consumes the report
How Often Report Is Used	ex: weekly, ad hoc, monthly
How Report is Accessed and/or Distributed	ex: web site, email alerts, mobile app, Tableau Server, etc.
Data Security	ex: only managers can access the report, managers only see their department data, users only see data for their account, etc.
Source System(s)	source of data that is queried by report
Printable	indicates whether users should be able to print the report
Exportable	indicates whether users should be able to export the report to another format

Data Definition
Insert the dimension and metric boxes here.

Filters Definition
Define the field(s) and values used in the filters of the viz. Ex: Weeks in the current fiscal year, sorted chronologically descending.

Functional Requirements
Provide details on interactivity, drillthrough, push notifications, printing, and exporting that isn't built in to the tool that will be used to build the viz

Data Rules
Explain any logic that is built into the queries or reports (In SQL, this might be case statements or filters in the where clause)

Examples and Screenshots
Include a link or screenshot of the original tabular report and any helpful materials to design the new report

DESIGN: WHITEBOARD

- Are you making a single chart, a report, or a dashboard?
- Come up with high-level ideas and place them on your canvas
Mobile Reports Developer (formerly Datazen)
- What specific items of information should be displayed? What does each of these items tell you, and why is that important? At what level of summary or detail should the information be expressed?
- Use your Data Definitions and Bus Matrix as a catalog!

DESIGN QUESTIONS

- Which items of information are most important for achieving your objectives?
- What are the logical groupings that could be used to organize items of information on the dashboard? In which of these groups does each item belong?
- What are the most useful comparisons that will allow you to see these items of information in meaningful context?

(From Stephen Few's *Information Dashboard Design*)

DESIGN: CHART TYPES

You can also categorize charts into these types (Andy Kirk):

CATEGORICAL	Comparing categories and distributions of quantitative values
HIERARCHICAL	Charting part-to-whole relationships and hierarchies
RELATIONAL	Graphing relationships to explore correlations and connections
TEMPORAL	Showing trends and activities over time
SPATIAL	Mapping spatial patterns through overlays and distortions

(From Data Visualization: A Handbook for Data Driven Design)

DESIGN: CHOOSE THE RIGHT CHART

"What is the right graph for my situation?

...whatever is easiest for your audience to read." - Cole Nussbaumer Knaflic

No chart is evil, they just have different roles & limitations – Andy Kirk

perceptual edge

Graph Selection Matrix

Graph	Relationship	Time Series Values display how something changed through time (yearly, monthly, etc.)	Ranking Values are ordered by size (descending or ascending)	Part-to-Whole Values represent parts (ratios) of a whole (for example, regional portions of total sales)	Deviation The difference between two sets of values (for example, the variance between actual and budgeted expenses)	Distribution Counts of values per interval along a quantitative scale from lowest to highest (for example, counts of people in an organization by age intervals of 10 years each)	Correlation Comparison of two paired sets of values (for example, the heights and weights of several people) to determine if there is a relationship between them	Nominal Comparison A simple comparison of values for a set of unordered items (for example, products or regions)
Bar Graph (vertical)		Yes (to feature individual values and support their comparisons; quantitative scale must begin at zero)	Yes (quantitative scale must begin at zero)	Yes (quantitative scale must begin at zero)	Yes (quantitative scale must begin at zero)	Yes (quantitative scale must begin at zero)		Yes (quantitative scale must begin at zero)
Bar Graph (horizontal)			Yes (quantitative scale must begin at zero)	Yes (quantitative scale must begin at zero)	Yes (quantitative scale must begin at zero)	Yes (quantitative scale must begin at zero)		Yes (quantitative scale must begin at zero)
Line Graph		Yes (to feature overall trends and patterns and support their comparisons)			Yes (only when also featuring a time series or single distribution)	Yes (to feature the overall shape of the distribution)		
Dot Plot (vertical)		Yes (when you do not have a value for every interval of time)	Yes					Yes
Dot Plot (horizontal)			Yes					Yes
Strip Plot (single)						Yes (when you want to see each value)		
Strip Plot (multiple)		Yes (only when also featuring distributions)				Yes (when comparing multiple distributions and you want to see each value)		
Scatter Plot							Yes	
Box Plot (vertical)		Yes (only when also featuring distributions)				Yes (when comparing multiple distributions)		
Box Plot (horizontal)						Yes (when comparing multiple distributions)		

CHECK YOURSELF

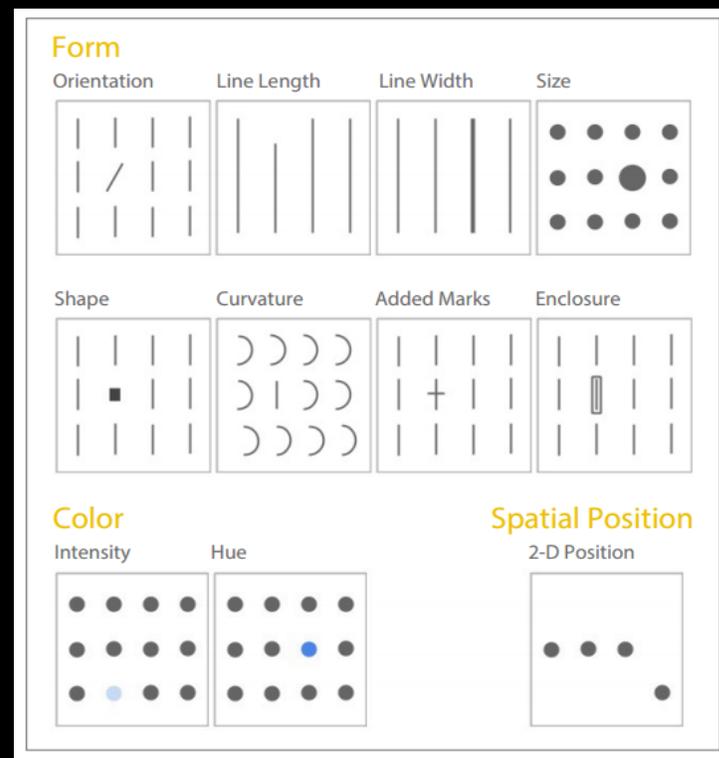
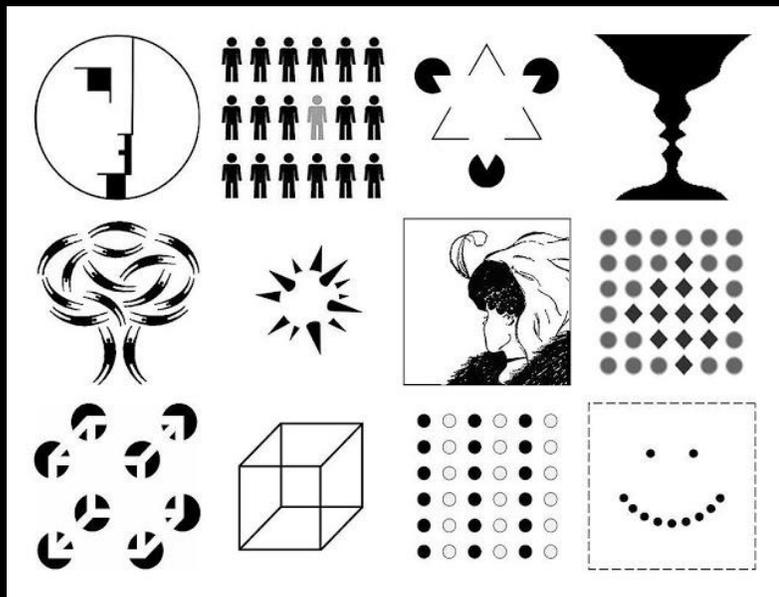
- Are the groupings of information obvious?
- Are the key metrics being featured adequately?
- Can you easily spot the items that need attention?
- Is enough information being displayed about the items that need attention to decide whether you must respond by taking action?

PROTOTYPE: TECHNICAL ITEMS TO CONSIDER

- Average or minimum screen resolution/size of users
- Capabilities of the reporting tool
- Method of access (mobile?)

OPTIMIZE YOUR DATA VIZ

- Limits of working memory: 3 chunks at a time
- Encoding data for rapid perception using preattentive attributes
- Gestalt principles of visual perception

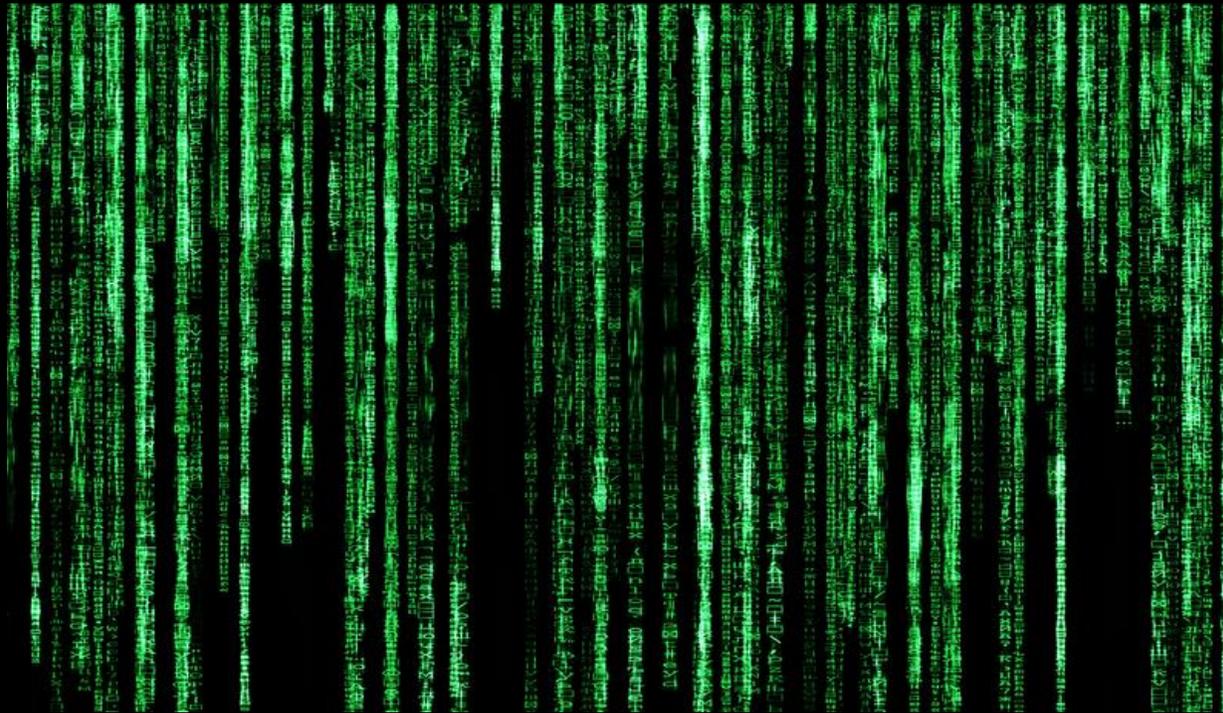


MORE TIPS

- If you don't want to include large graphs, consider bullet graphs and sparklines to provide visual context 
- Use enough descriptive text to provide necessary context
- Put supplementary information within reach
- Reduce information to what's essential
- Make the experience aesthetically pleasing

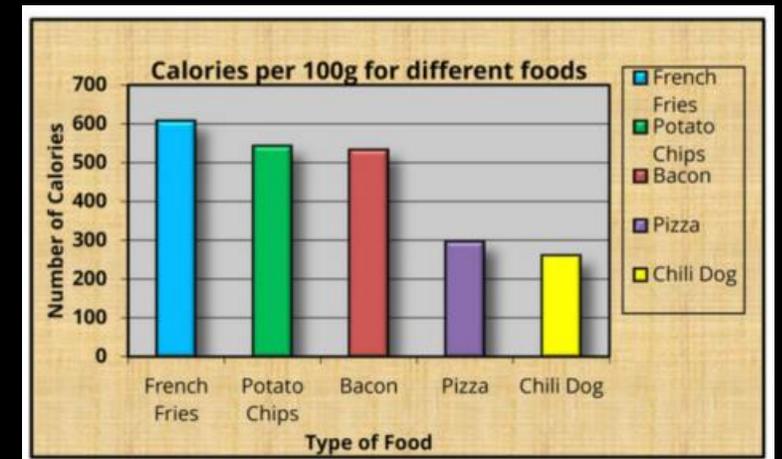
A NOTE ON BIG DATA

Kirk: Visualizing big data isn't a data problem, it's a summarization problem. You've only got so many pixels on the screen. Summarize and then add interactivity to explore more detail.



ELIMINATE CLUTTER

- Data-Ink Ratio (Edward Tufte/Stephen Few)
- Signal to Noise Ratio (Nancy Duarte)
- Chartjunk (Edward Tufte)



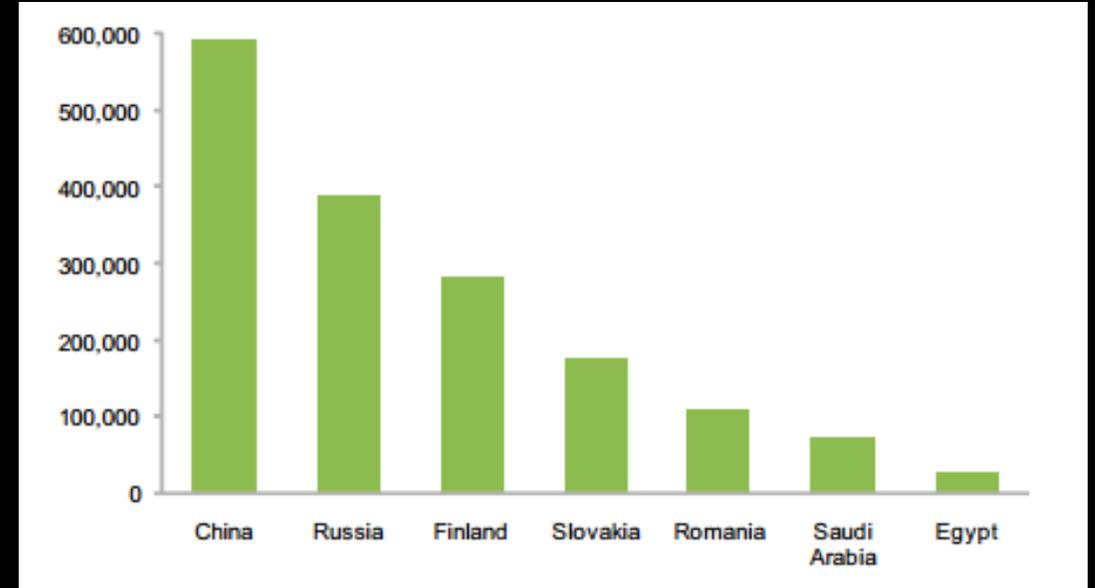
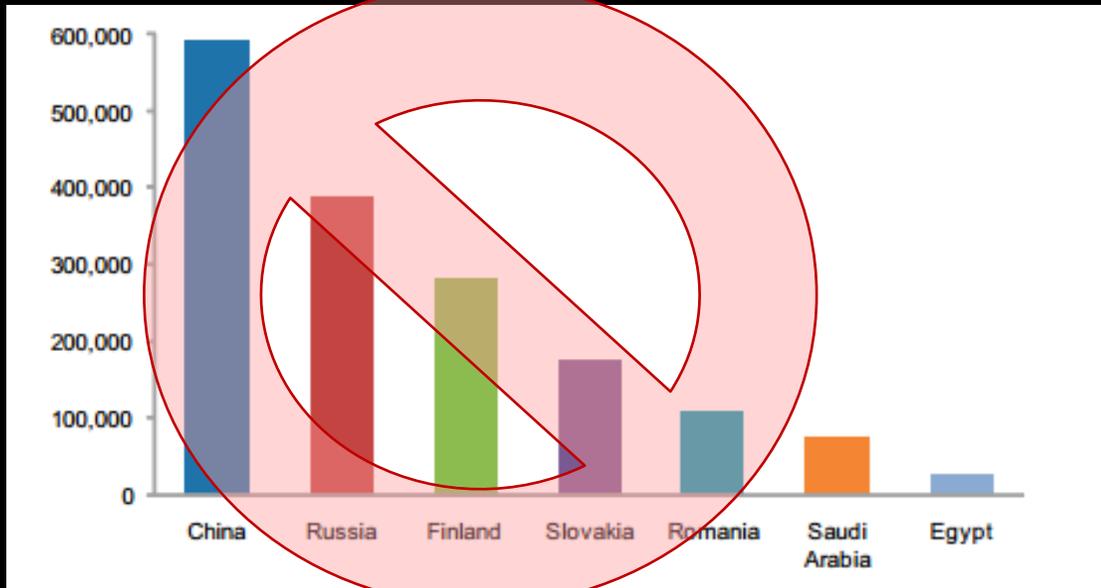
<https://darkhorseanalytics.com/blog/data-looks-better-naked/>

Few: Maximize data-ink ratio

Kirk: Balance data-ink ratio, maximize reward/effort

CHOOSE APPROPRIATE & MEANINGFUL COLORS

Changing colors indicates a difference

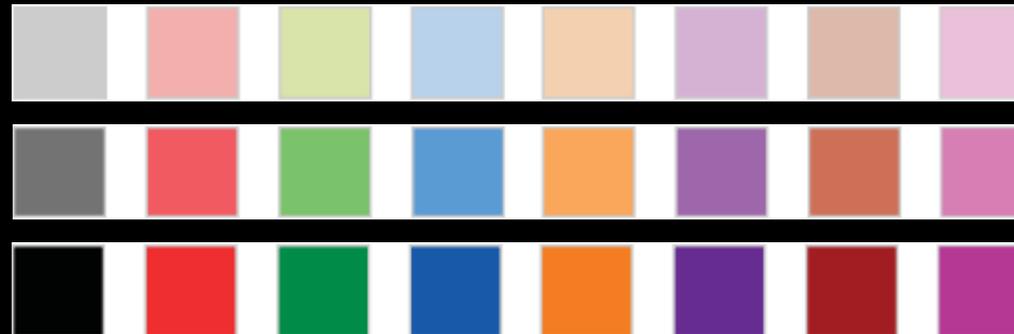


COLOR PALETTES

- Categorical
- Sequential
- Diverging



- Light
- Medium
- Dark



Great tool: <http://highintegritydesign.com/tools/tinter-shader/>

ETTES

- Cate
-
-
- M
- Dark

PLEASE STOP SCREAMING AT ME!

THE SQUINT TEST

- Shrink things down and/or half close your eyes to see what colored properties are most prominent and visible.
- Are those the right ones?



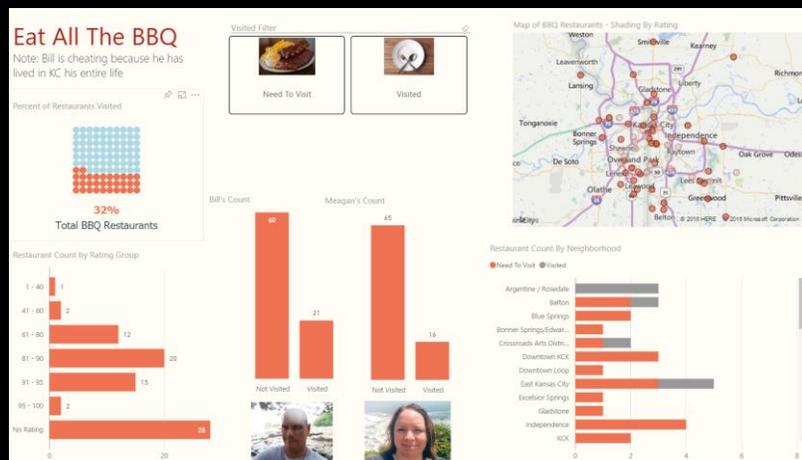
COLOR VISION DEFICIENCY

Color Vision Deficiency affects 1 in 12 men and 1 in 200 women.

Red-green color blindness is most common.

Blue and orange are good options for safe colors.

Use <http://www.color-blindness.com/coblis-color-blindness-simulator/> to test your viz.

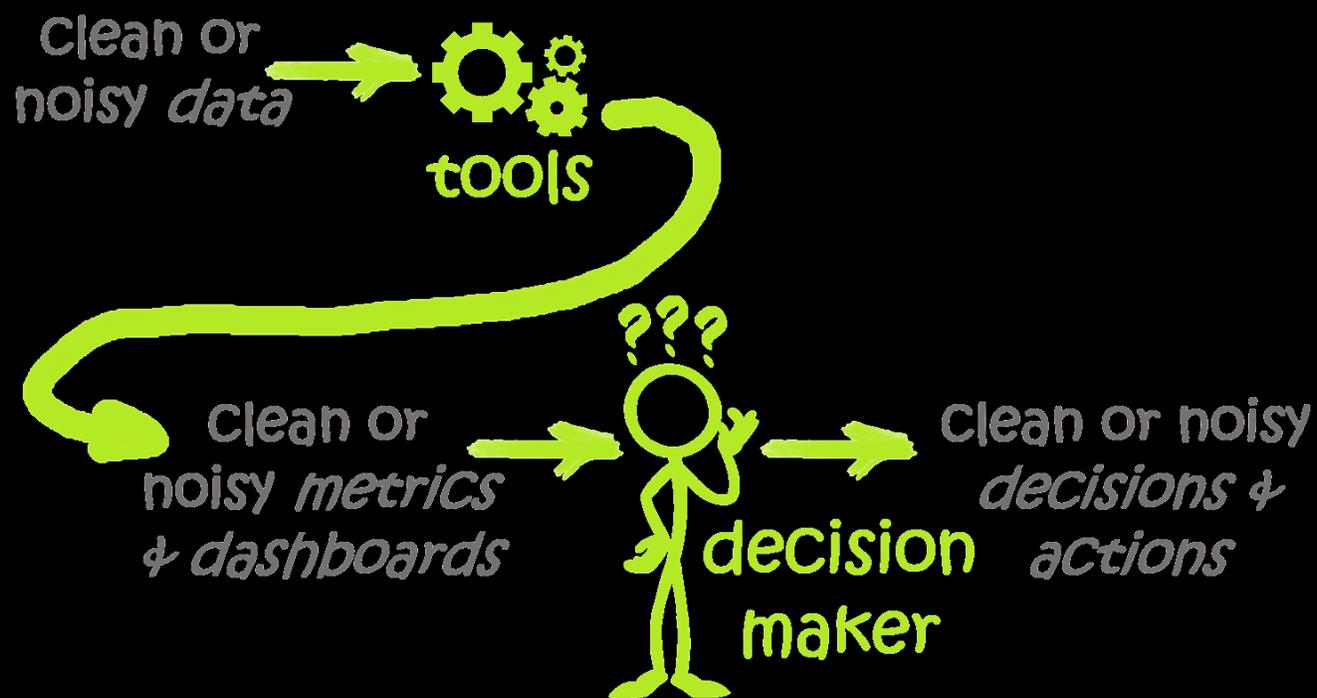


REMEMBER THIS

- Unless you are the main user, you are not building this data viz for you. Build your data viz to provide the most to your users.
- Do not overpromise with your prototype and under deliver with your final product.
 - If you've already decided upon a technology, don't promise features (chart types, interactivity) that can't be made with it.
- The one true measure of success is adoption/usage.
 - Getting people to engage is sometimes as important as building the cognitively most valid method. – Andy Cotgreave

REMEMBER THIS

Our (developers') outputs are decision-makers' inputs – and **their** outputs are what ultimately matter.



MY BBQ DASHBOARD

Chartaphobia V2 VISIT REPORT

Eat All The BBQ

Note: Bill is cheating because he has lived in KC his entire life

Percent of Restaurants Visited

Visited Filter:

Total BBQ Restaurants: 32%

Restaurant Count by Rating Group:

Rating Group	Count
1 - 40	1
41 - 60	2
61 - 80	12
81 - 90	20
91 - 95	15
95 - 100	2
No Rating	28

Bill's Count: Not Visited: 60, Visited: 21

Meagan's Count: Not Visited: 65, Visited: 16

Bill's Restaurants To Visit

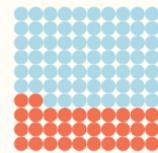
Restaurant	Distance Fro...	Rating
Fritz's Meat &...	1.76	0
Driftwood BBQ	1.83	0
The Stack Bar...	2.75	0
Jon Russell's ...	3.02	92
The Burnt En...	4.73	78
Rock & Brews	4.90	69

Meagan's Restaurants To Visit

Eat All The BBQ

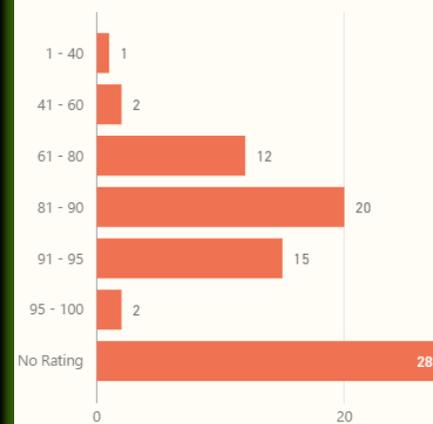
Note: Bill is cheating because he has lived in KC his entire life

Percent of Restaurants Visited



32%
Total BBQ Restaurants

Restaurant Count by Rating Group

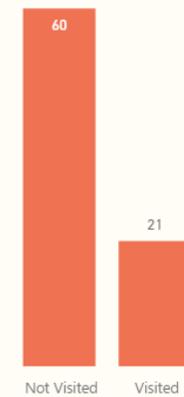


Visited Filter

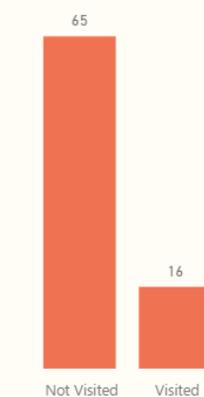
Need To Visit

Visited

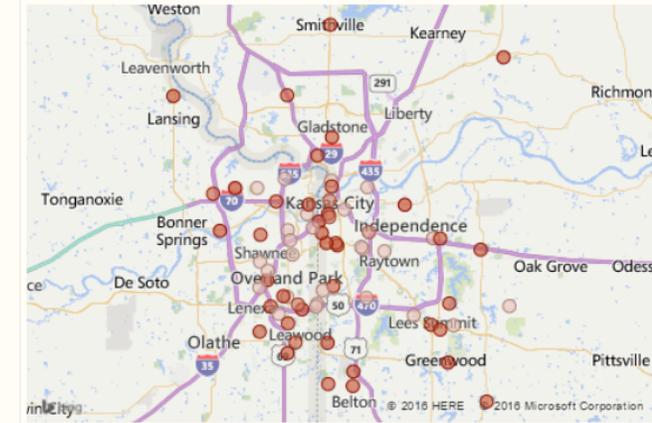
Bill's Count



Meagan's Count

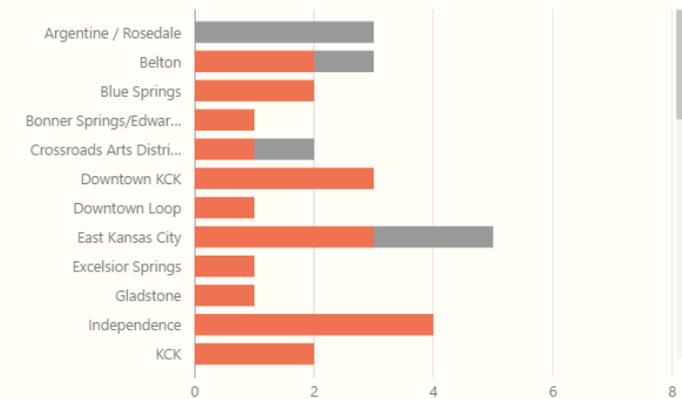


Map of BBQ Restaurants



Restaurant Count By Neighborhood

● Need To Visit ● Visited



LINKS FOR FURTHER LEARNING

- Storytelling With Data: <http://www.storytellingwithdata.com/>
- Stephen Few/Perceptual Edge blog: <http://www.perceptualedge.com/blog/>
- Paul Turley - Transforming Reporting Requirements Into a Visual Masterpiece: <https://www.youtube.com/watch?v=7c1hjdEzNfQ>
- Preattentive Features and Tasks video: <https://www.youtube.com/watch?v=wnvoZxe95bo>
- Data Viz Done Right: <http://www.datavizdoneright.com/>
- Power Pivot Pro: A New Take on “Data Quality?”: <http://www.powerpivotpro.com/2015/06/a-new-take-on-data-quality/> Improving Data Viz Effectiveness: <http://www.blue-granite.com/blog/improving-data-visualization-effectiveness>
- Andy Kirk: Separating Myth From Truth in Data Visualization: <https://www.brighttalk.com/webcast/9059/193677>
- Gestalt Principles Composition Image By Impronta (Own work) [CC BY-SA 3.0 (<http://creativecommons.org/licenses/by-sa/3.0>)], via Wikimedia Commons: http://commons.wikimedia.org/wiki/File:Gestalt_Principles_Composition.jpg

QUESTIONS & FINAL COMMENTS

Feel free to contact me with questions or feedback.

Meagan Longoria

Solutions Consultant

Blog: datasavvy.wordpress.com

Twitter: [@mmarioe](https://twitter.com/mmarioe)

LinkedIn: www.linkedin.com/in/meaganlongoria/

