Data Visualization Best Practices

Spring, 2018

GSA

D2D
DATA TO DECISIONS

Course Outline

- Why visualize
- Good practice guidelines
- Basic visualizations
- Composition rules
- Interactivity

GSA

D2D
DATA TO DECISIONS

Why Visualize?

- Data visualization shifts the balance from cognition (thinking) to perception (seeing)
- Seeing is handled by the visual cortex located in the rear of the brain, is extremely fast and efficient
- We see immediately, with little effort
- Thinking is handled primarily by the cerebral cortex in the front of the brain, is much slower and requires bigger effort
- Images are processed faster than audio or text
- Our brains have powerful image processing which works unconsciously

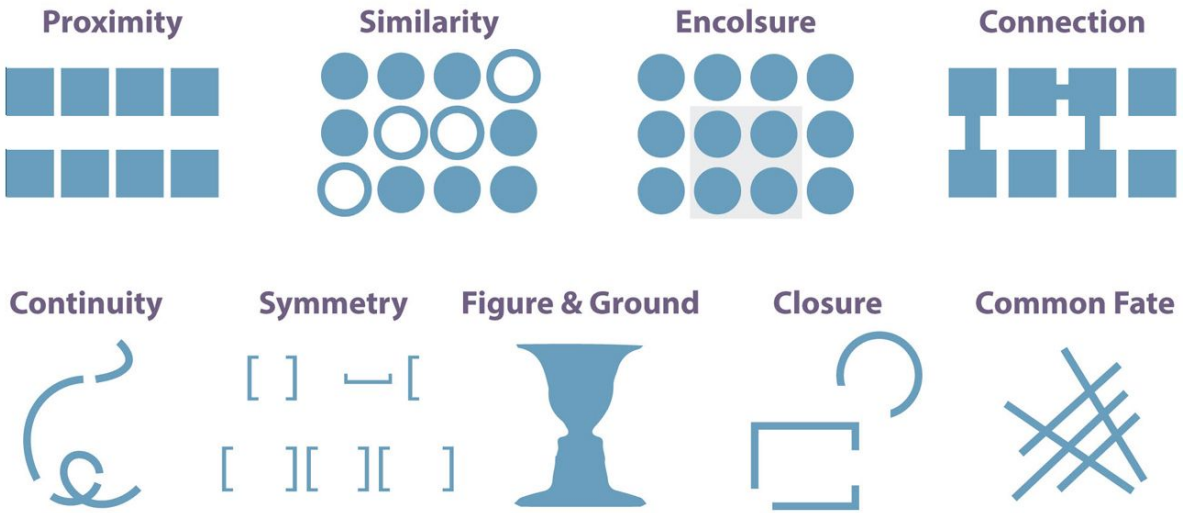
GSA

D2D
DATA TO DECISIONS

Gestalt Principles of Grouping

- a philosophy of mind developed by the Berlin School of Experimental Psychology, "Gestalt" is German for shape, form

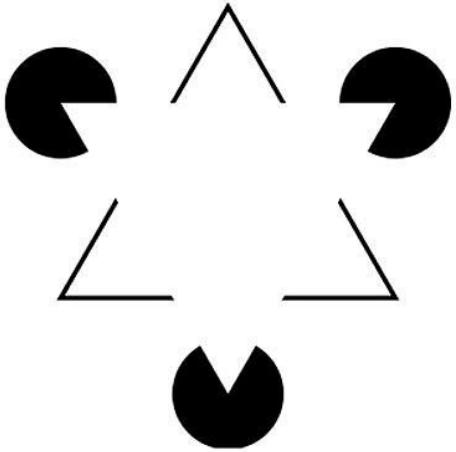
- **Proximity: close together objects are perceived as a group**
- **Similarity: object with same attributes (color, size) are perceived as a group**
- **Enclosure: objects within a boundary are perceived as a group**
- **Closure: open structures are perceived as closed if there is a way to interpret as such**
- **Continuity: aligned objects are perceived as a group**
- **Connection: connected objects are perceived as a group**

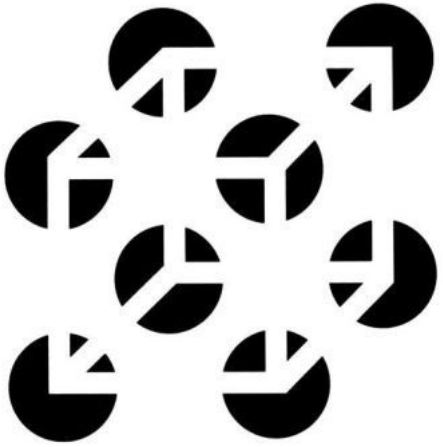


GSA

D2D
DATA TO DECISIONS

Exploit Ability of the Human Brain to Infer Simplest Solution for Incomplete Visual Information

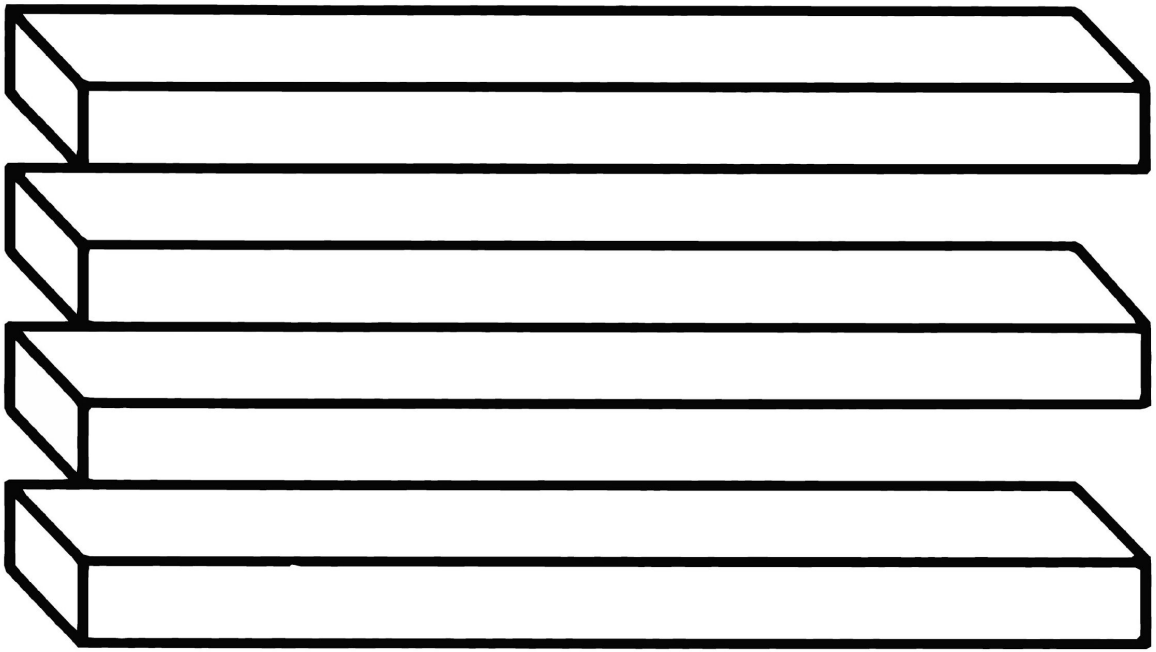




GSA

D2D
DATA TO DECISIONS

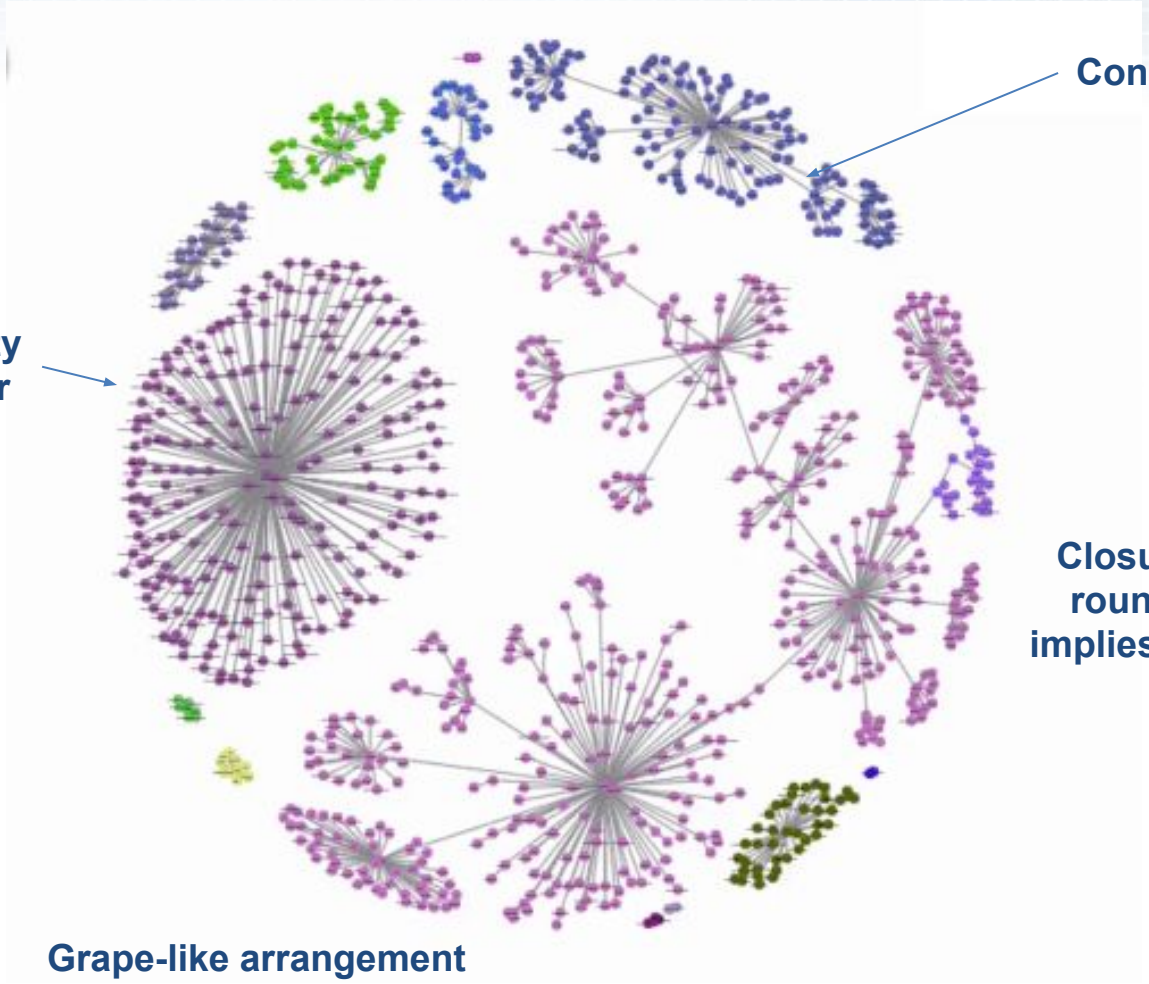
An eye can be fooled



GSA

D2D
DATA TO DECISIONS

Nice Picture Makes You Interested

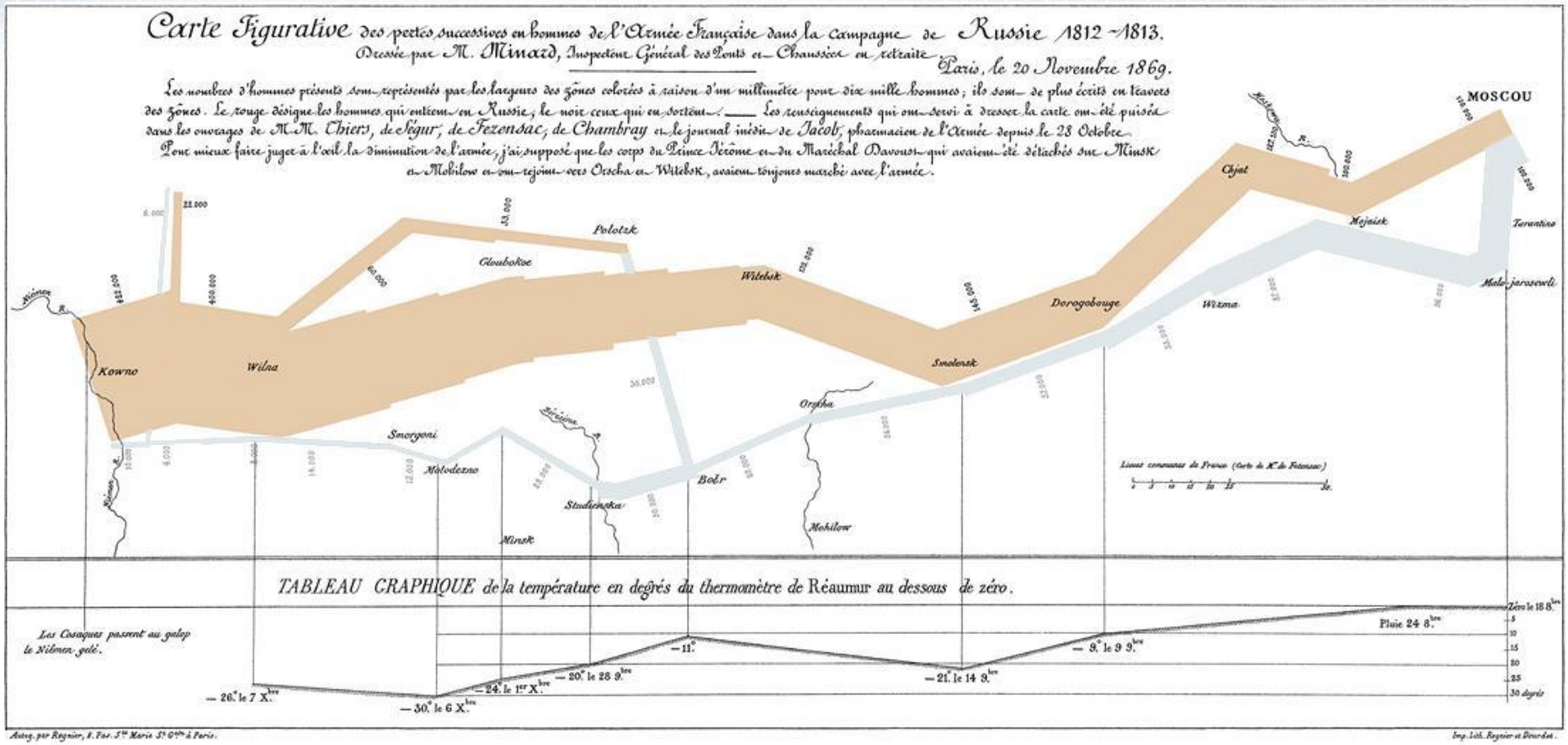


GSA

D2D
DATA TO DECISIONS

Classic Visualization

Charles Minard: Napoleon's Russian Campaign of 1812



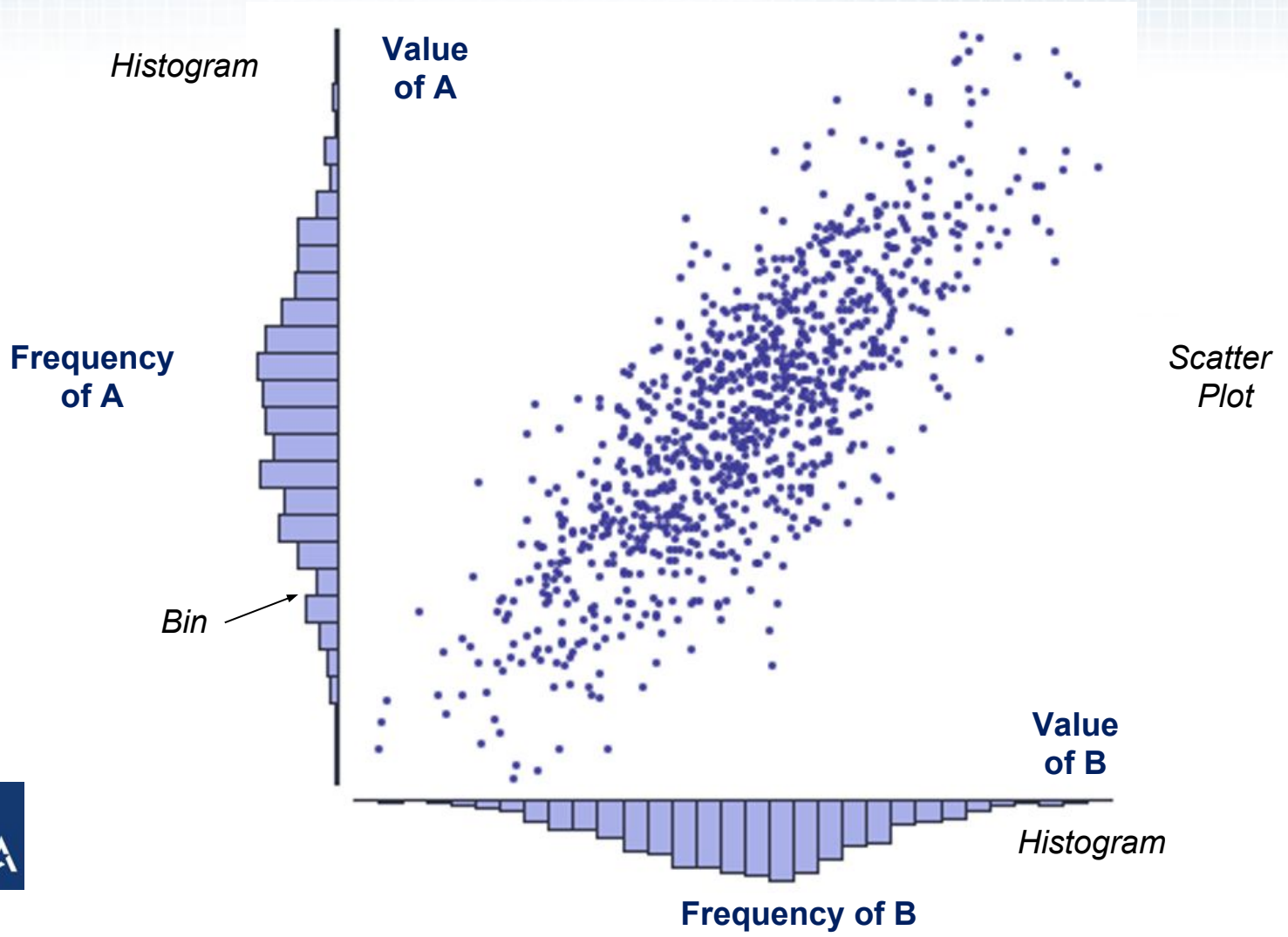
The width of the lines corresponds to the number of troops
The beige color is for the direction to Moscow
The grey color is for the return

Edward Tufte's Visualization Goals

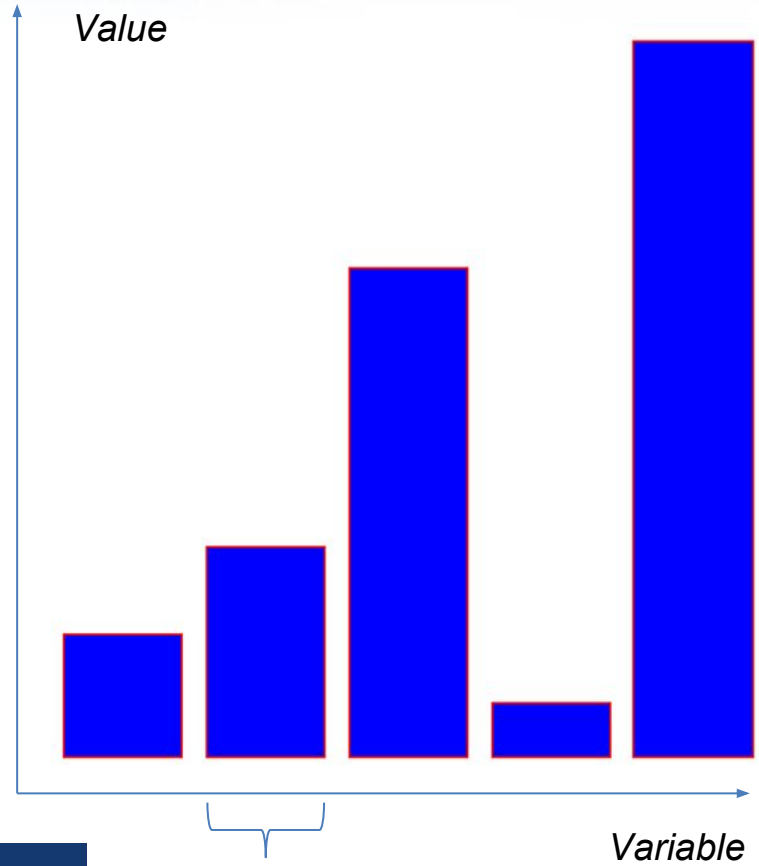
- **Explore**
- **Compare**
- **See trends**
- **Relate part to whole**
- **Rank**
- **Correlate**
- **Show distribution**

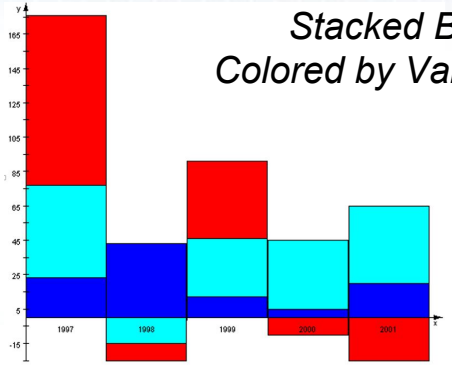
*Edward Tufte, The Visual Display of Quantitative Information

Data Exploration
Scatter Plot and Histogram



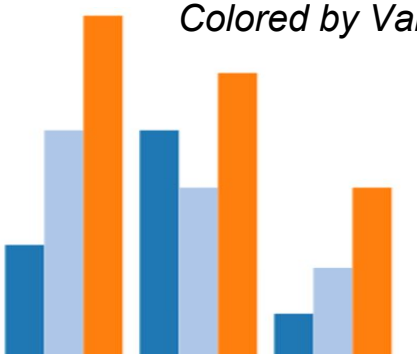
Data Comparison
Bar Charts





Stacked Bar Chart
Colored by Value Attributes

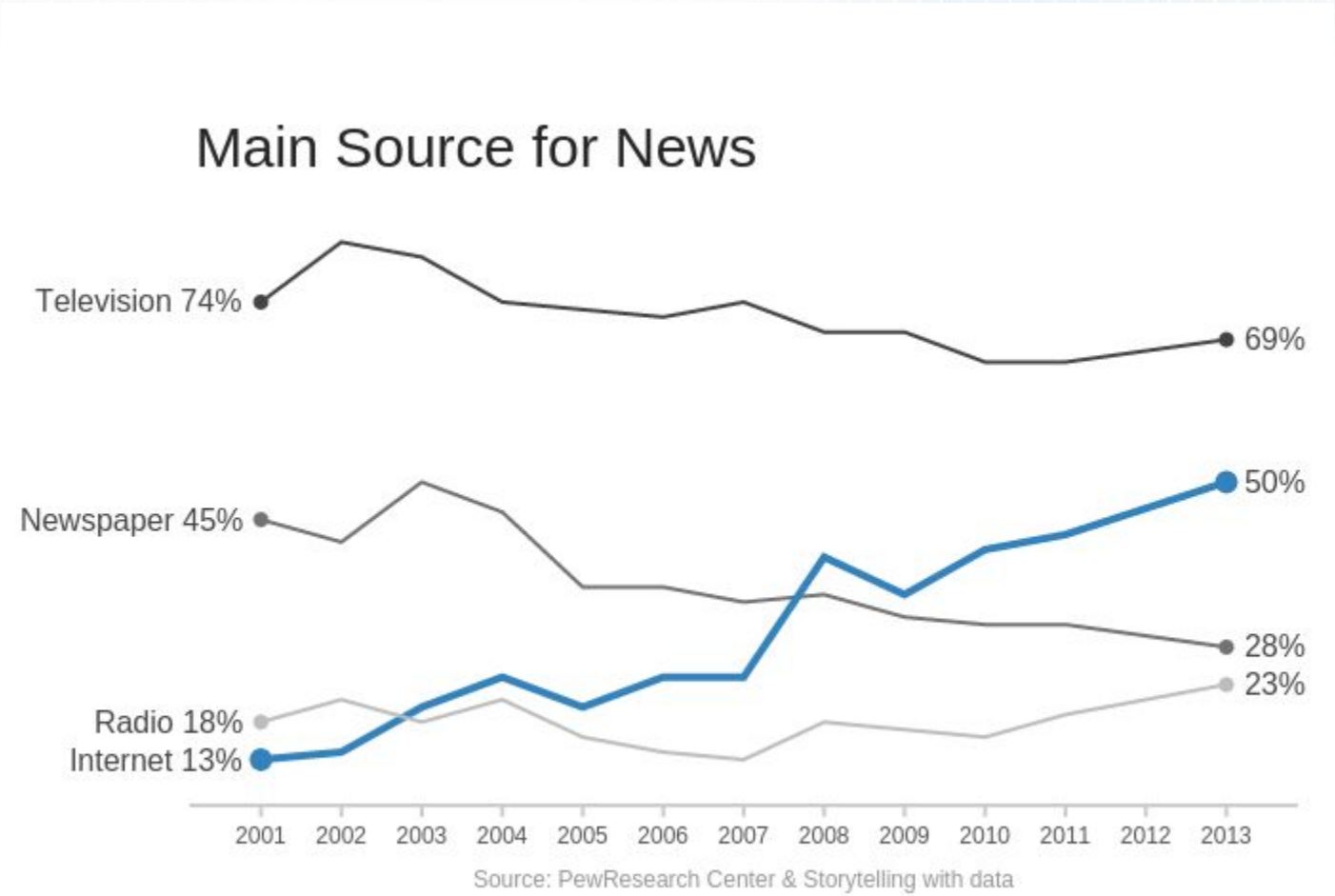
Clustered Bar Chart
Colored by Value Attributes



GSA

D2D
DATA TO DECISIONS

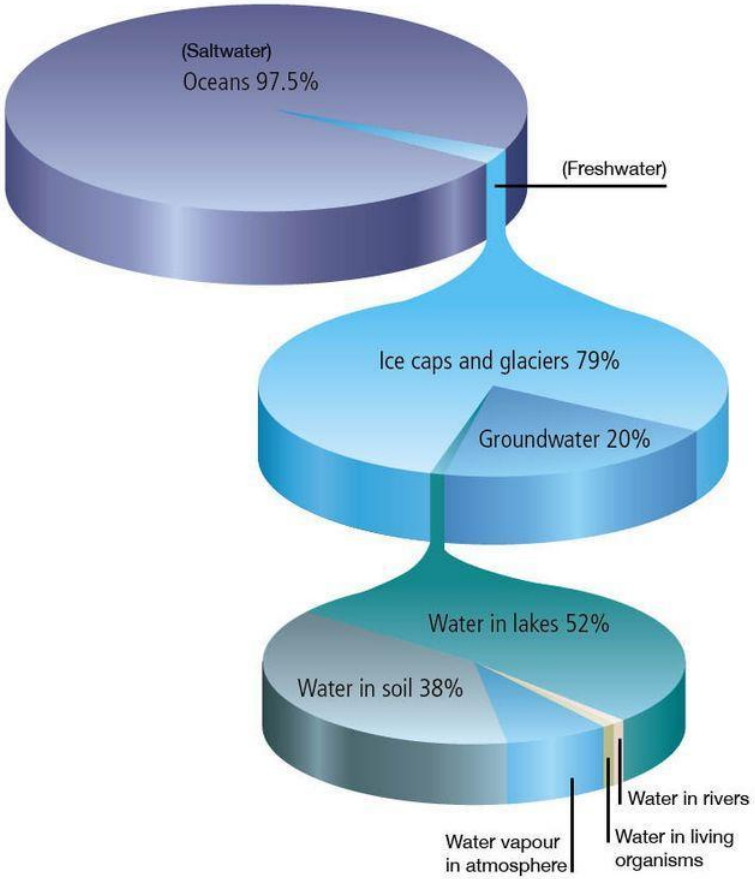
Data Trends
Line Charts

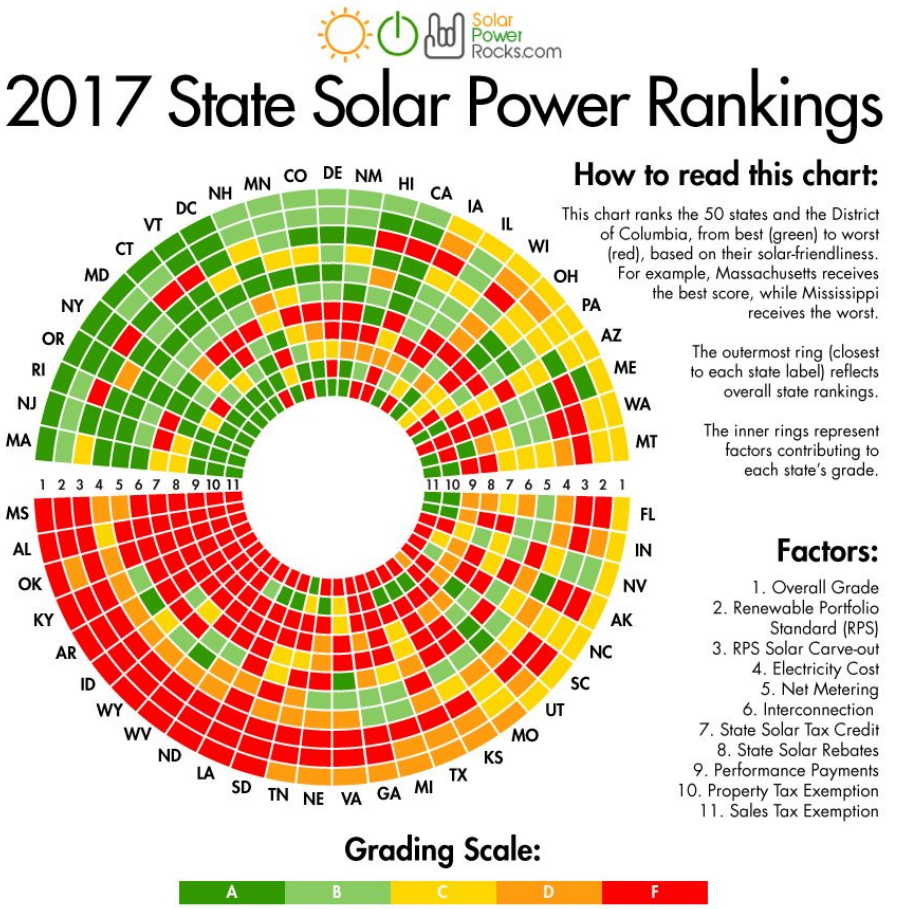


GSA

D2D
DATA TO DECISIONS

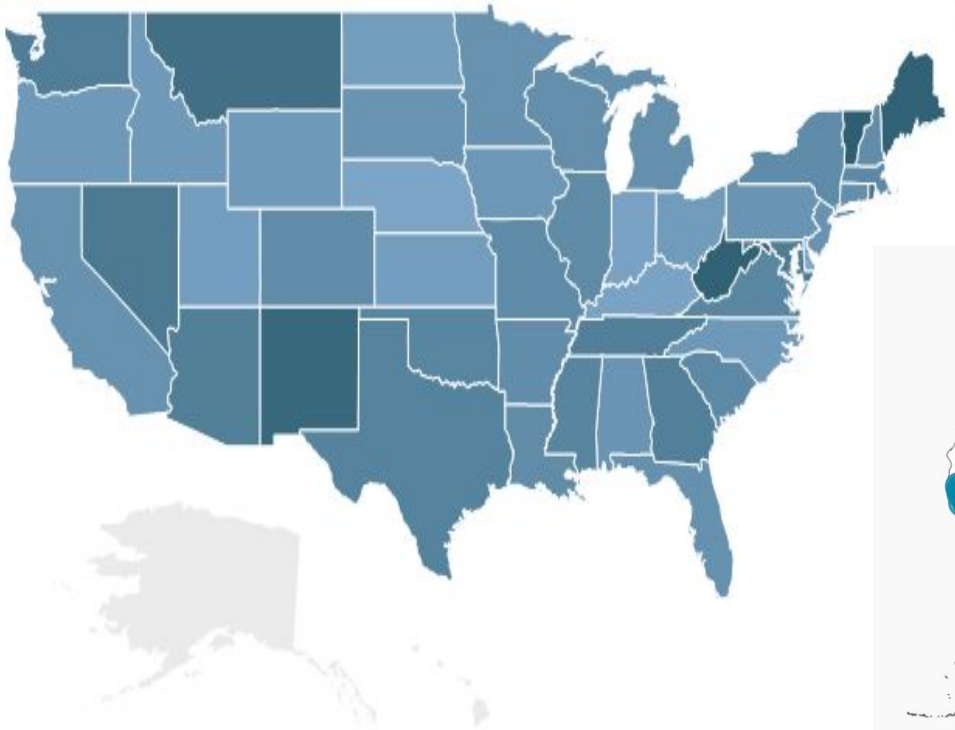
Relate Part to Whole
Pie Charts





No explanations needed!

Maps



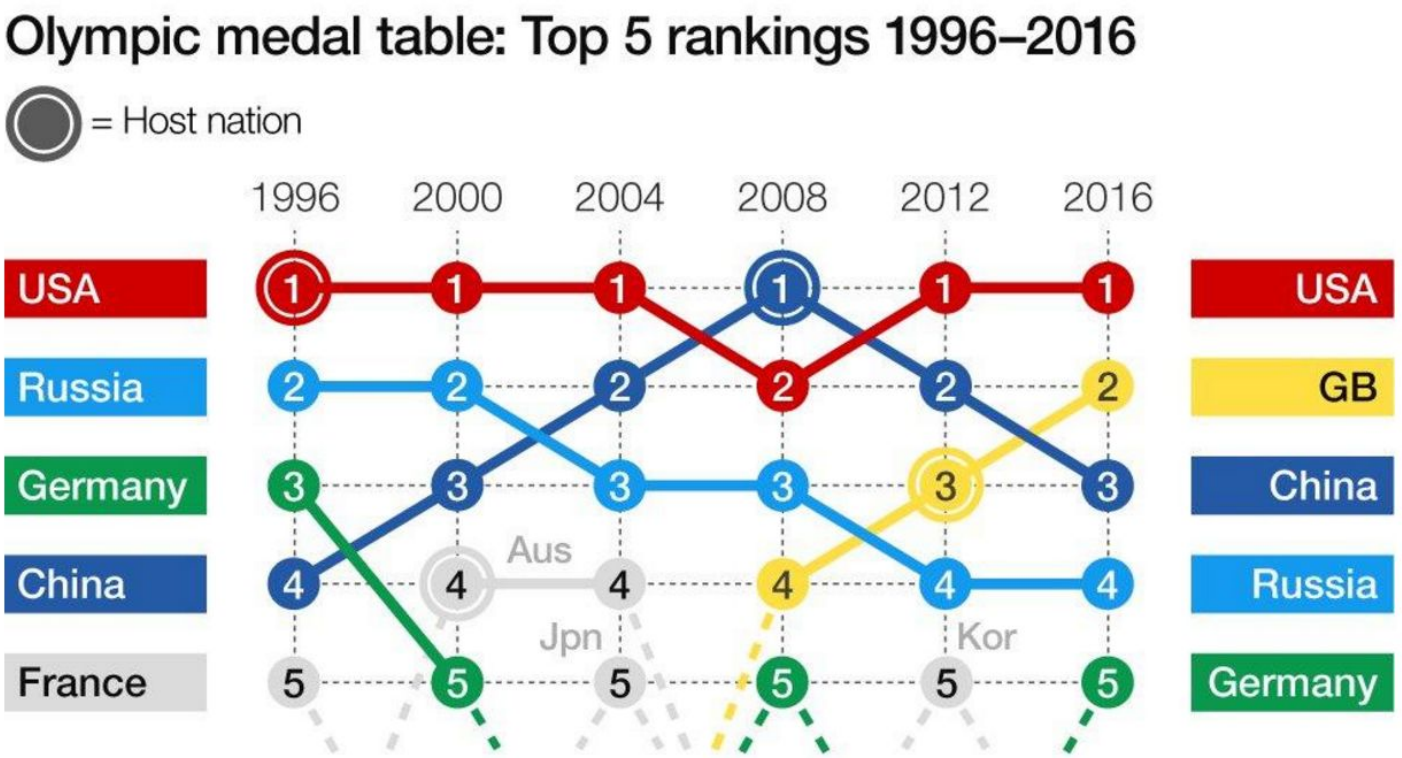
Heat Map
State borders polygons
Color density
proportional to value

Bubble Markers
Pinpointed to coordinates
Size and density
proportional to values

GSA

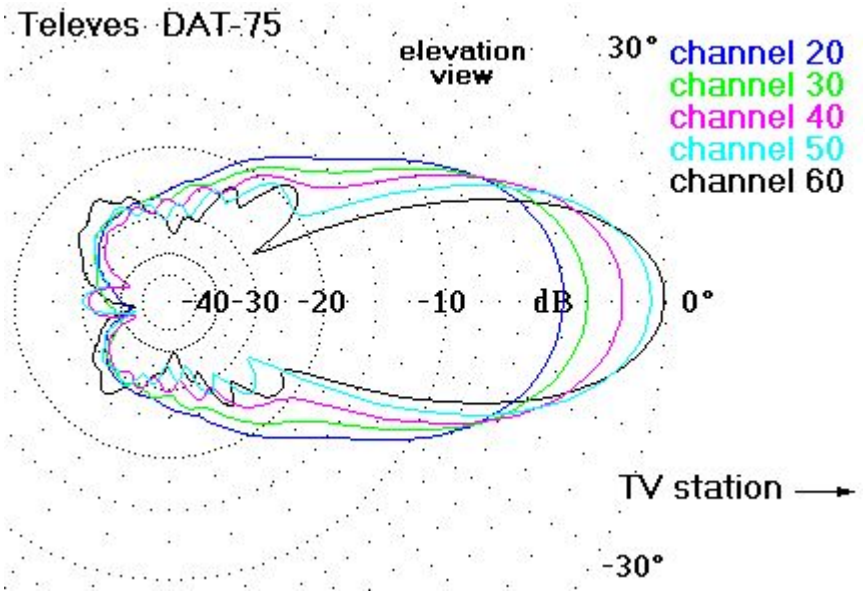
D2D
DATA TO DECISIONS

Ranking Charts

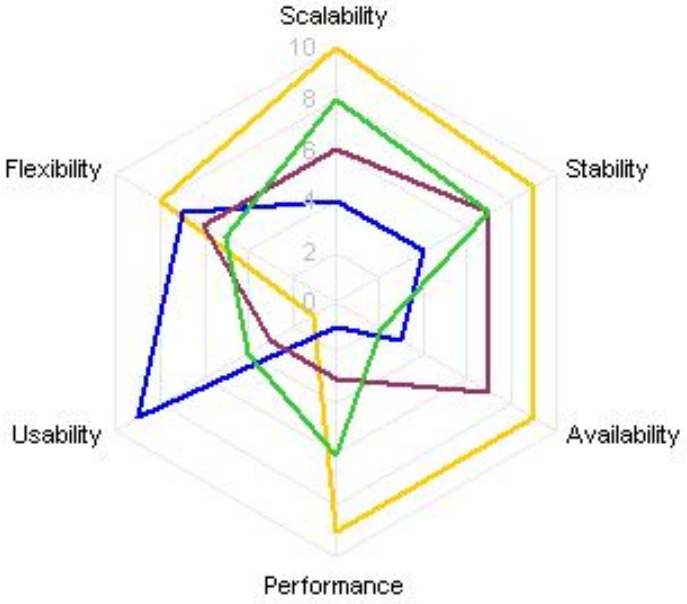


Radar Charts

Antenna Diagram
Attenuation
Vs. Direction



Spider Diagram
Quick way to assess on multiple parameters

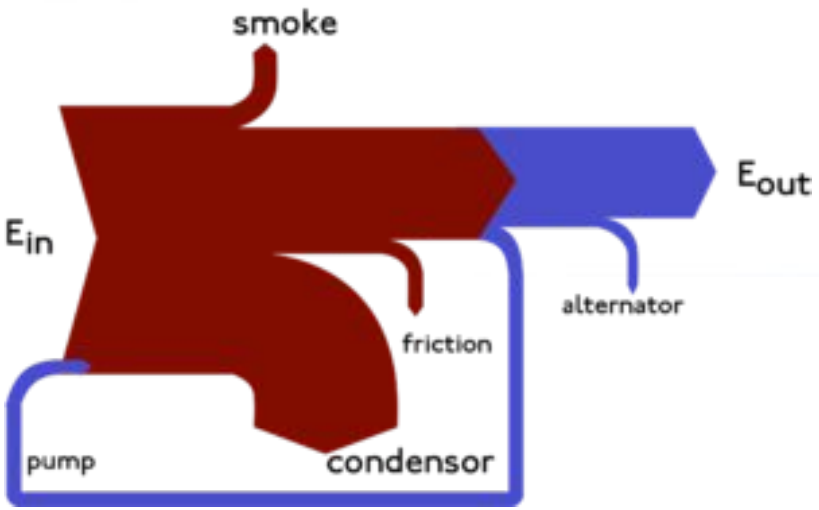


GSA

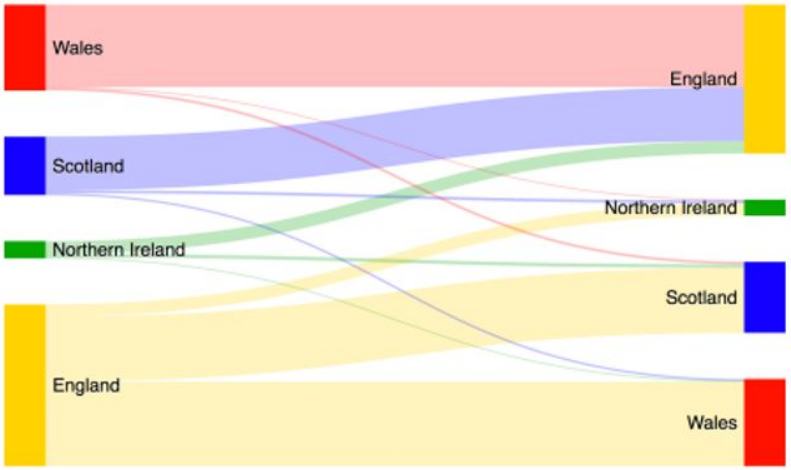
D2D
DATA TO DECISIONS

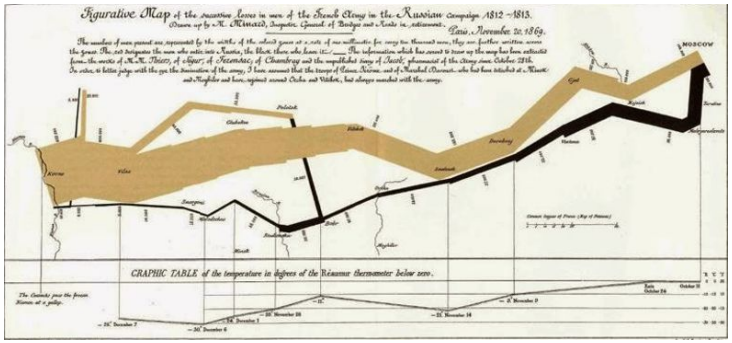
Sankey Diagram

Steam Engine Diagram



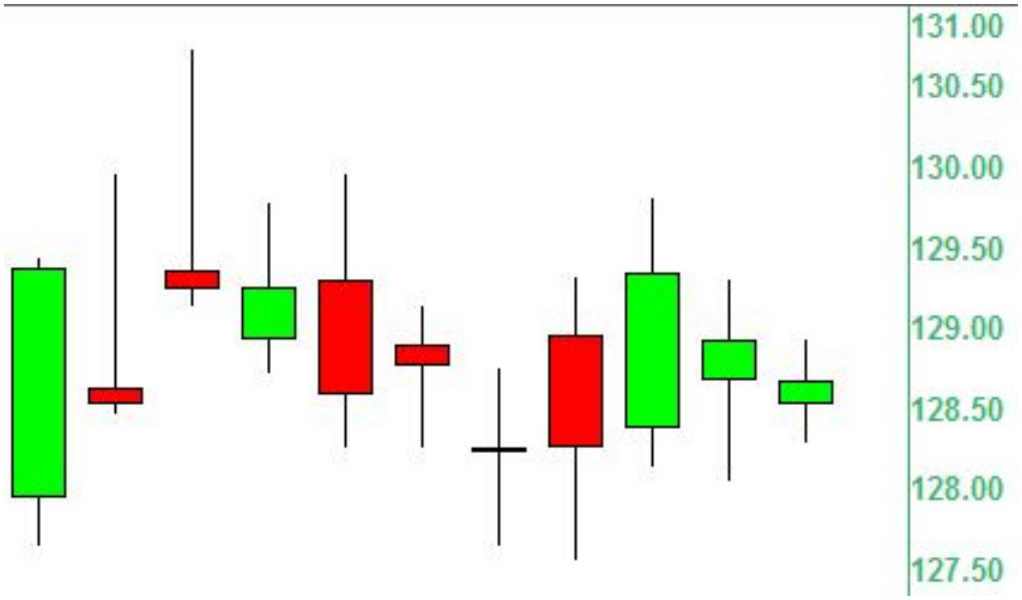
Migration Between UK Countries

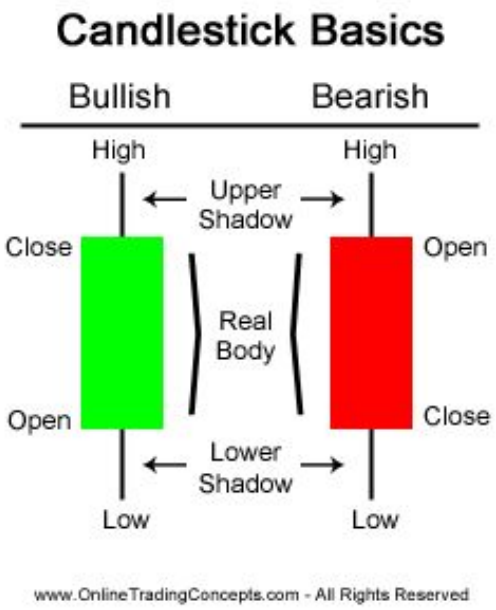




Remember Charles Minard?

Special: Candlestick Diagram
(Shows Statistics: Range 1, Range 2, Gradient)



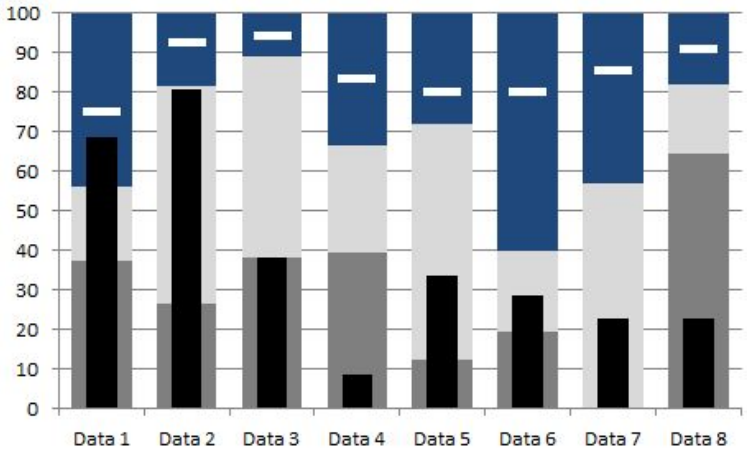


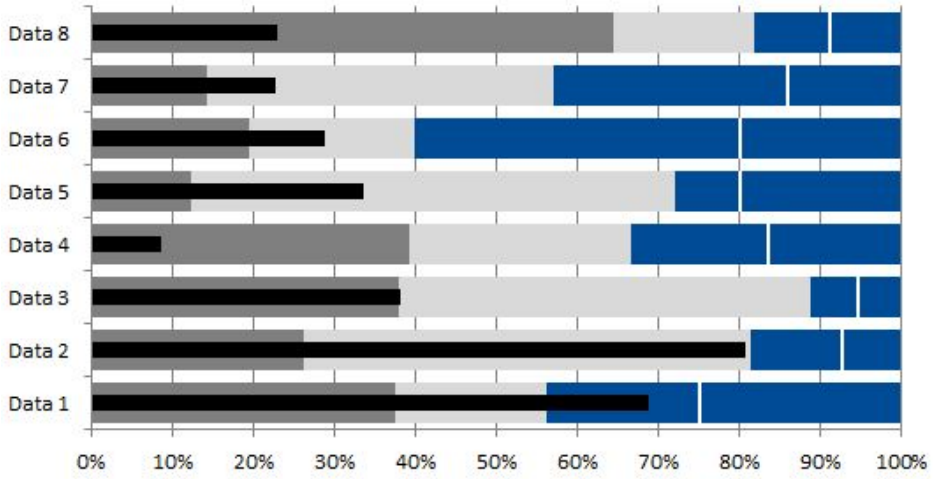
GSA

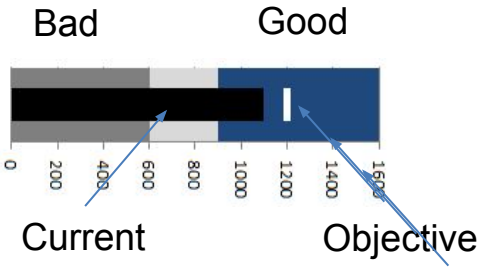
D2D
DATA TO DECISIONS

Special: Bullet Chart

Compares current value to target, previous, and objective



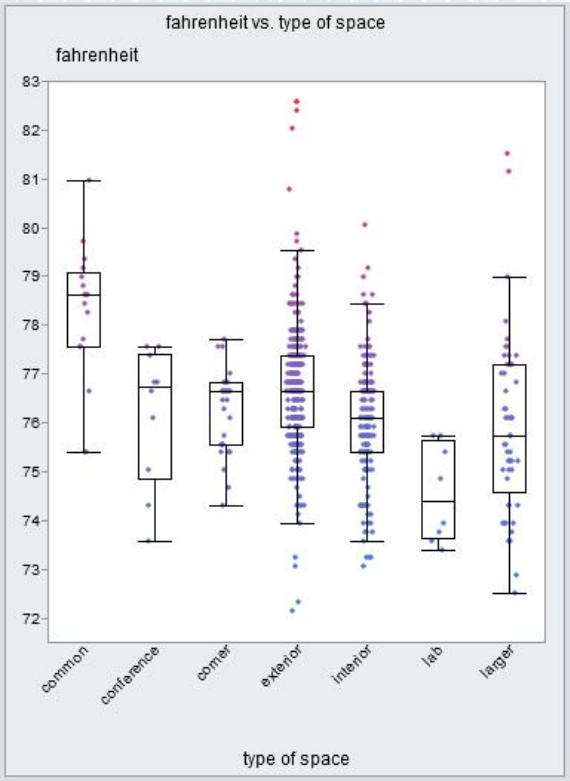


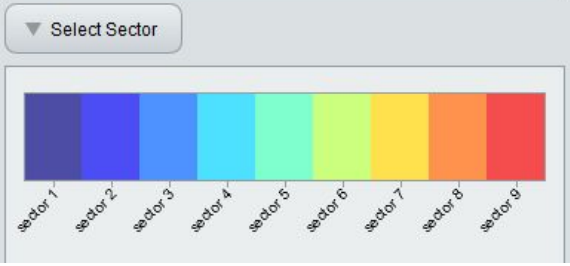


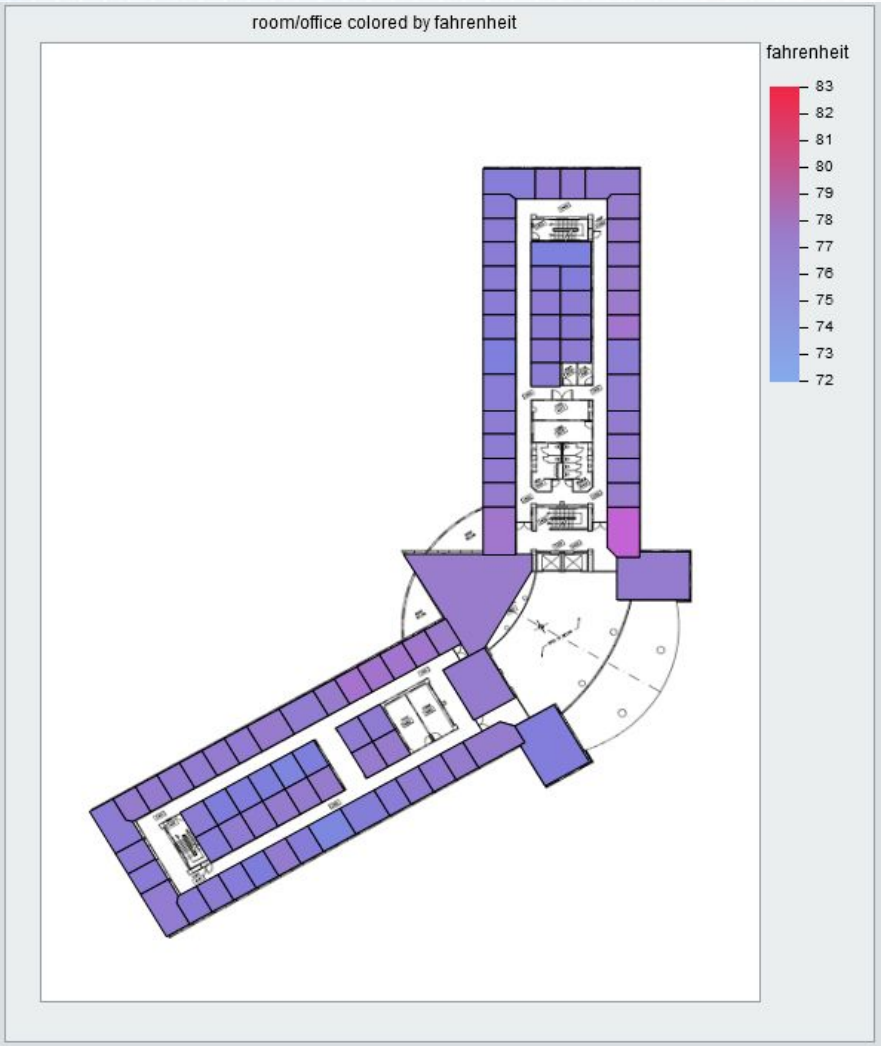
GSA

D2D
DATA TO DECISIONS

Building Air Temperature





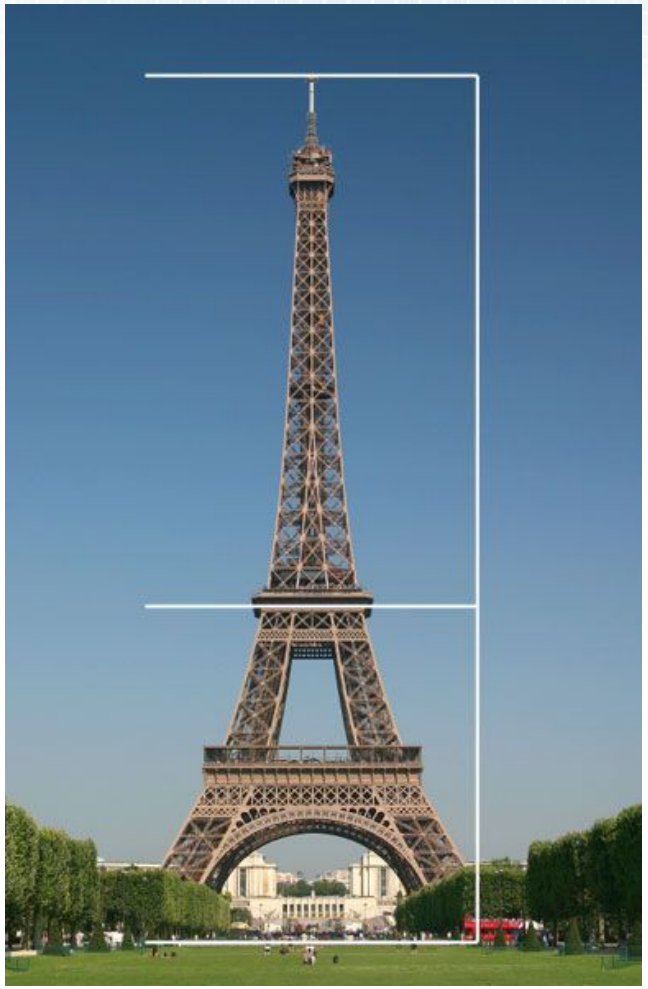


GSA

D2D
DATA TO DECISIONS

Composition Rules

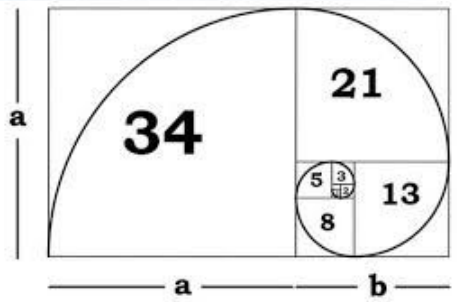
- The visualization shall please the eye
- People spend more time with something they like
- People are more forgiving for what they like
- The eye likes
  - ➢ Symmetries, balance
  - ➢ Golden ratio (Phi = 1.618…)



GSA

D2D
DATA TO DECISIONS

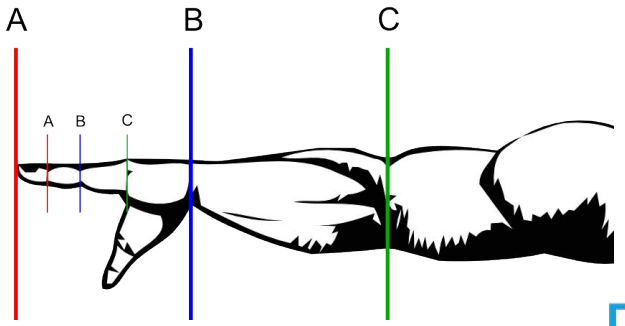
Golden Ratio (Fibonacci) Spiral (1,2,3,5,8,13,21…)



GSA

D2D
DATA TO DECISIONS

Golden Ratio in Nature



GSA

D2D
DATA TO DECISIONS

Brain Wave Frequencies

The classical frequency bands of the EEG can indeed be described as a geometric series with a ratio between neighboring frequencies approximating F = 1.618

Frequency band		Frequency subband		Peak	Period
Name	[Hz]	name	[Hz]	[Hz]	[ms]
delta	1.5–4	delta1	1–2	1.5	667
		delta2	2–3	2.5	400
theta	4–10	theta1	3–5	4	250
		theta2	5–8	6.5	154
alpha	8–12	alpha	8–12	10	100
beta	10–30	beta1	12–20	16	62.5
		beta2	20–30	25	40
gamma	30–80	gamma1	30–50	40	25
		gamma2	50–80	65	15
fast	80–200	ripples1	80–120	100	10
ripples		ripples2	120–200	160	6.25

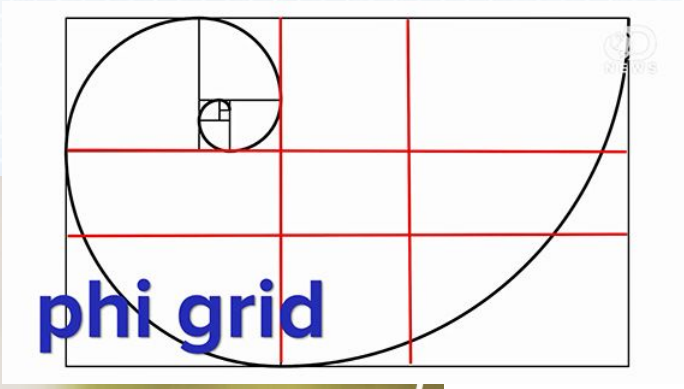
1,2,3,5,8 - Fibonacci Sequence => Golden Ratio = n/(n-1) = 1.618

http://www.i-sis.org.uk/Golden_Music_of_the_Brain.php

GSA

D2D
DATA TO DECISIONS

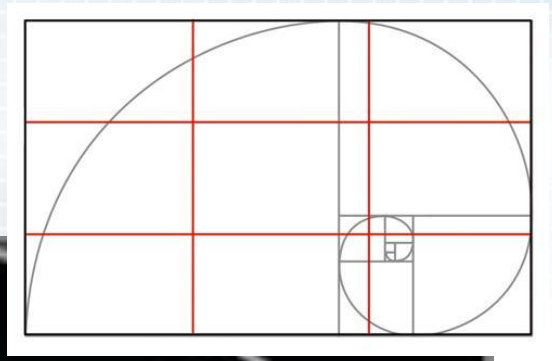
Phi Grid



GSA

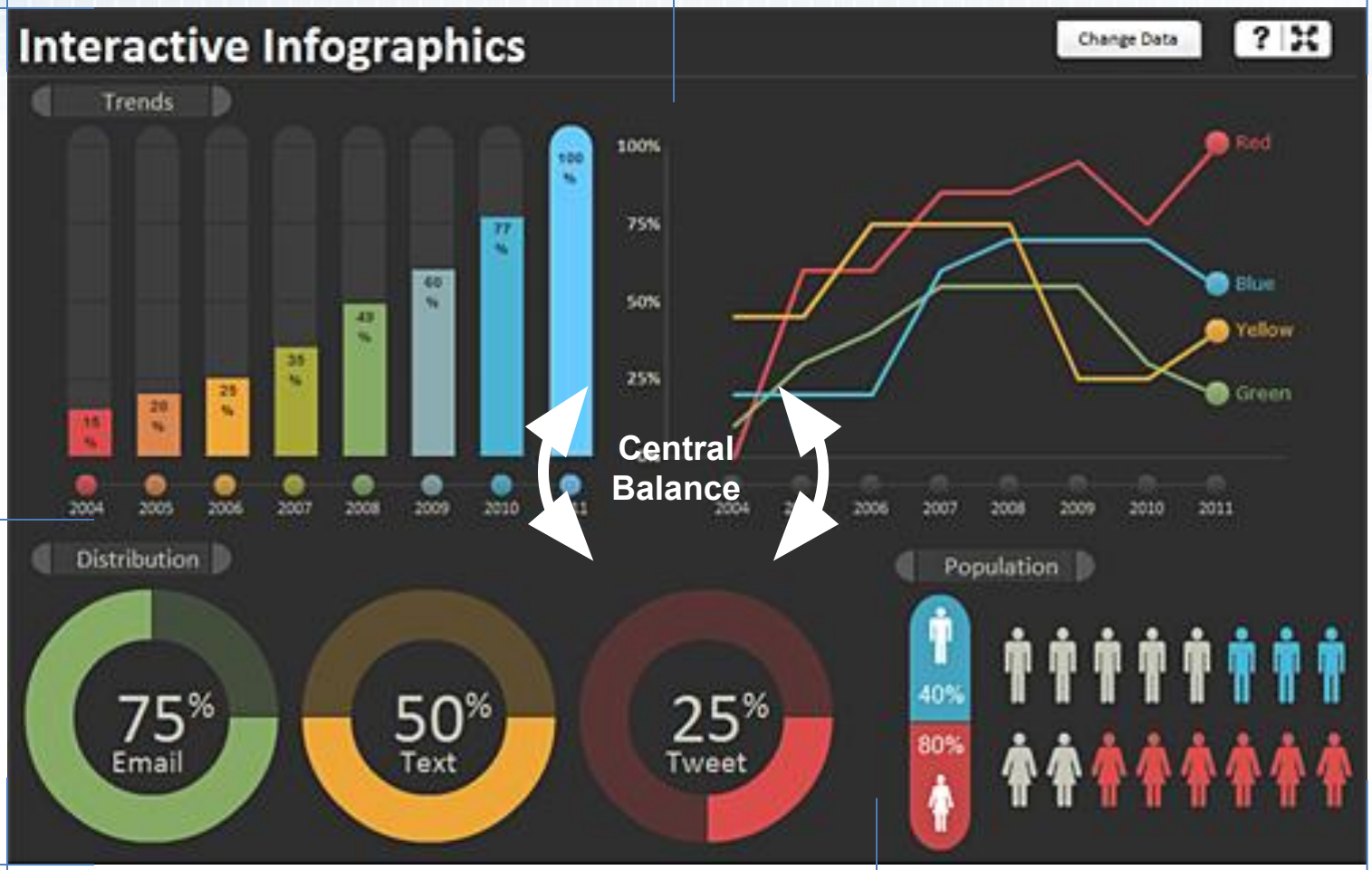
D2D
DATA TO DECISIONS

Rule of Thirds
Phi Grid for “Dummies”



When quick action is needed – like when making a photo - use “Rule of Thirds”

Dashboard Composition Example



GSA

D2D
DATA TO DECISIONS

Colors

- **"…the first principle in bringing color to information: above all, do no harm."**
  (Envisioning Information, Edward Tufte, Graphics Press, 1990)
- **Poor choice of colors will obscure, muddle, confuse or irritate**
- **If possible, use color palettes provided by visualization tool**
- **If not, create customer-matching palettes**
  - ➢ **Use customer logos or webpages**
  - ➢ **Extract color codes**
  - ➢ **Use online tools**

GSA

D2D
DATA TO DECISIONS

Color Theory

- First color circle was developed by Isaac Newton in 1666
- Analogues colors are adjacent
- Complementary colors are opposite

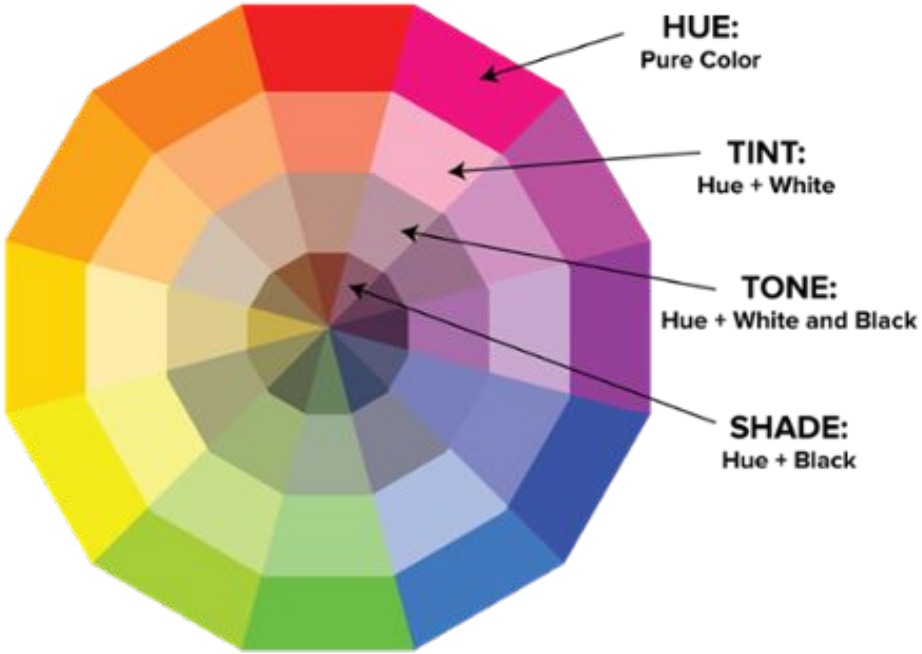
1776
Harris

TODAY

1810
Goethe

https://monoskop.org/images/4/46/Itten_Johannes_The_Elements_of_Color.pdf

Essential Color Terminology



GSA

D2D
DATA TO DECISIONS

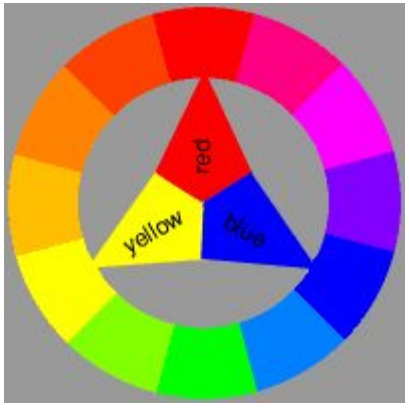
Why Printers and Monitors Use Different Primary Colors?

- You can get (almost) any color from either RGB or RYB (or other primary set)

Note: magenta, yellow, cyan cannot make black, hence there is black cartridge in printers

- You can mix colors by adding or subtracting (light)
- When you paint, you subtract light by adding another paint (reduce the amount of light going through – light is "subtracted")
- Subtractive color wheel gives black when all colors are added
- Additive color wheel gives white when all colors are added

Color wheel for painting



Color wheel for printing

Color wheel for projecting

GSA

D2D
DATA TO DECISIONS

Color Codes

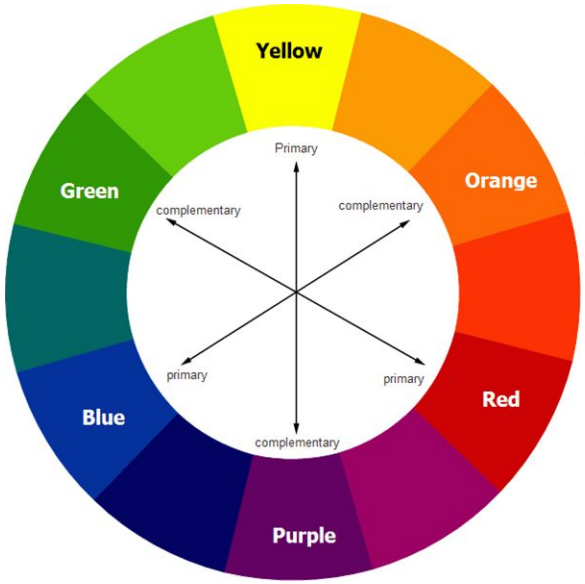
Hex Color Codes - Websites	RGB Color Codes - Screens
#FFFFFF	255,255,255
#000000	0,0,0,
#XXXXXX	XXX,XXX,XXX

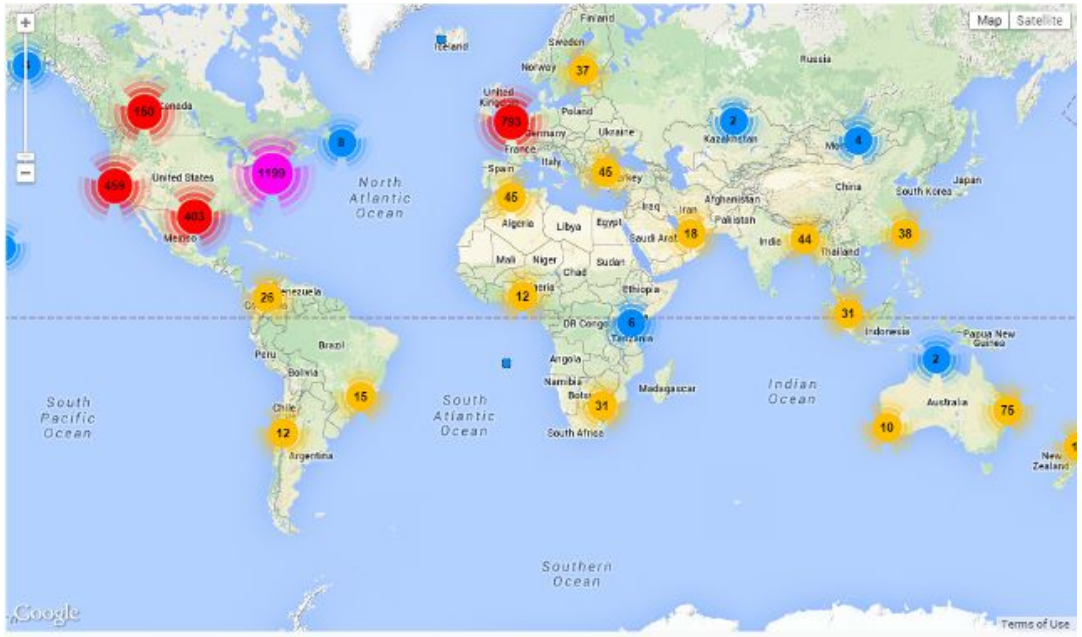
#9400D3	RGB 148, 0, 211
#4B0082	RGB 75, 0, 130
#0000FF	RGB 0, 0, 255
#00FF00	RGB 0, 255, 0
#FFFF00	RGB 255, 255, 0
#FF7F00	RGB 255, 127, 0
#FF0000	RGB 255, 0 , 0

https://htmlcolorcodes.com/
https://www.rapidtables.com/web/color/RGB_Color.html

Contrast Colors

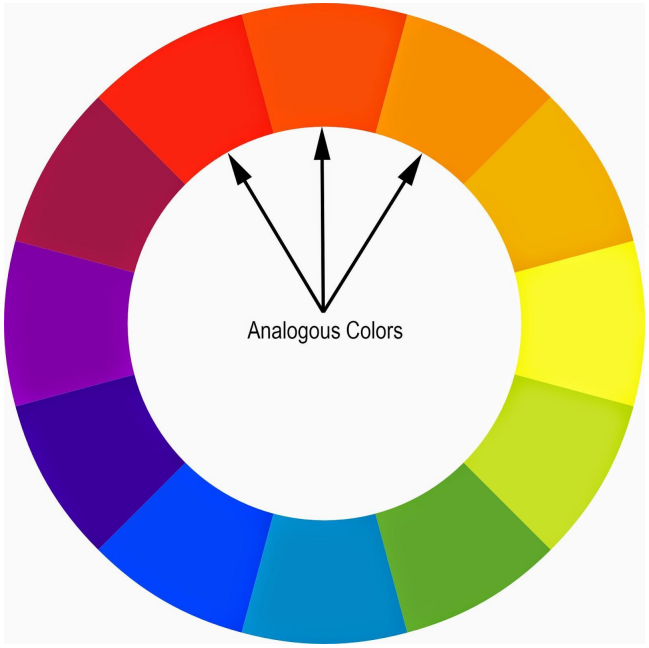
- Contrast colors draw attention
- Make things distinct
- Use color wheels to pick complementary contrast color

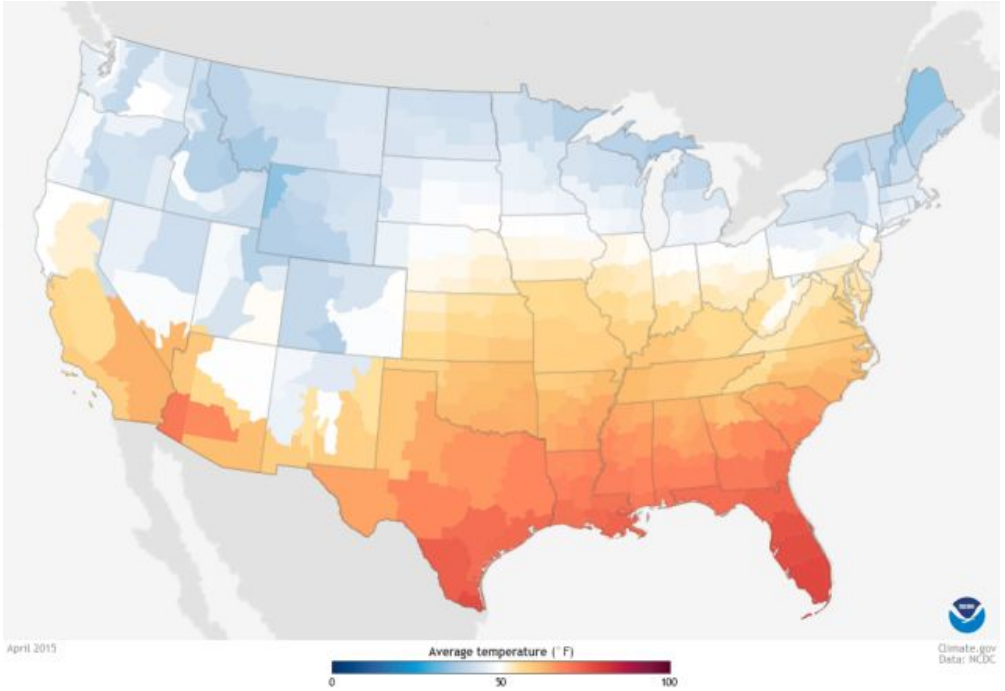




Analogous Colors

- Analogous colors suggest grouping
- Use them to show gradual changes



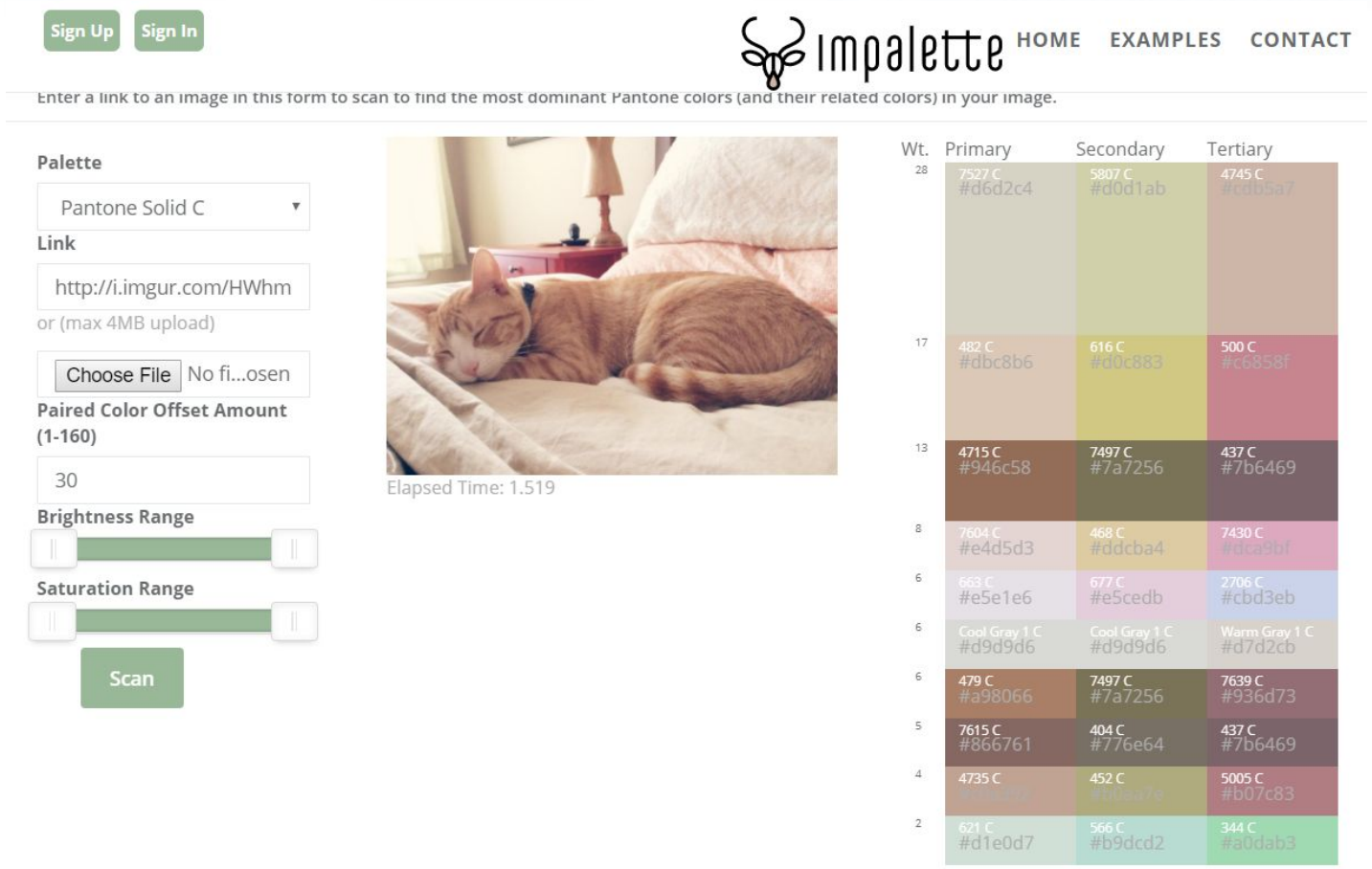


GSA

D2D
DATA TO DECISIONS

Match Customer Colors

- **Use online tools to create customer-matching palettes**
  e.g.: https://www.impalette.com/



GSA

D2D
DATA TO DECISIONS

Interactivity

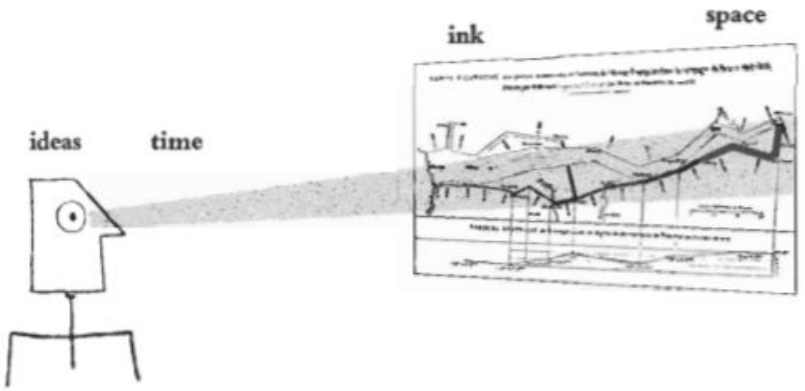
- **Interactivity improves usability when**
  - ➢ **Intuitive**
  - ➢ **Forgiving (e.g., can get back to where you where)**
- **Most common interactivity features**
  - ➢ **Sort**
    - **E.g.: descending, ascending, alphabetically, top 10, etc.**
  - ➢ **Filter**
    - **Exclude variables based on a parameter**
  - ➢ **Drill down**
    - **Path values from one visualization to another**
    - **Works best with established hierarchies**

GSA

D2D
DATA TO DECISIONS

Edward Tufte's Principles of Graphical Excellence*

- **Complex ideas communicated with**
  - ➢ **Clarity**
  - ➢ **Precision**
  - ➢ **Efficiency**
- **Greatest number of ideas in shortest time with least ink in the smallest space**



*Edward Tufte, The Visual Display of Quantitative Information

GSA

D2D
DATA TO DECISIONS

Edward Tufte's Visualization Guidelines*

- Show the data
- Induce the viewer to think over the substance, not graphic design, technology, etc.
- Do not distort what the data have to say
- Present many numbers in small space
- Make large data sets coherent
- Encourage the eye to compare different pieces of data
- Reveal the data at several layers of detail, from broad to fine
- Serve a clear purpose: description, explorations, tabulation, or decoration
- Be closely integrated with statistical and verbal descriptions of a data set

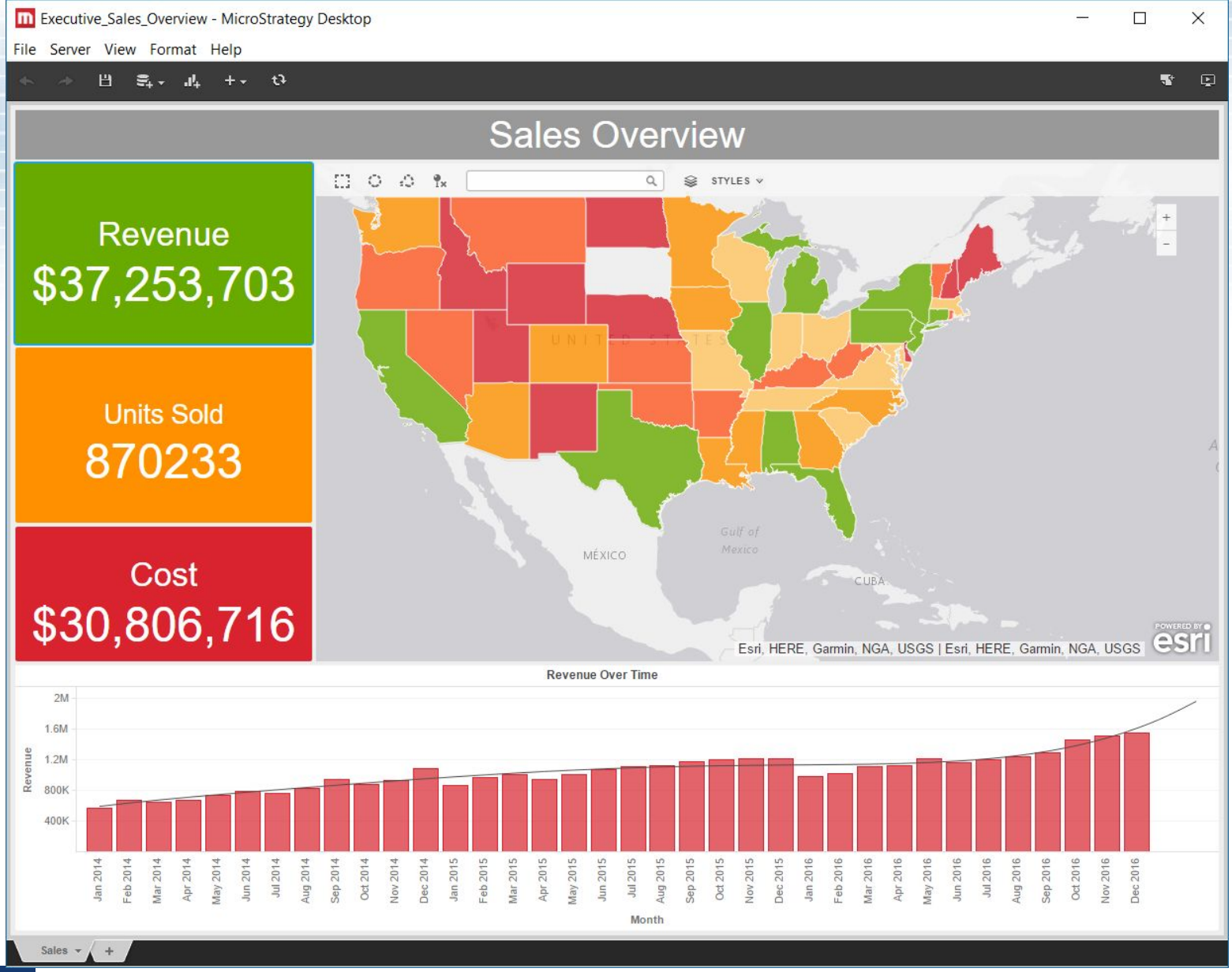
*Edward Tufte, The Visual Display of Quantitative Information

MicroStrategy Examples

- MicroStrategy Desktop is free
- Create data extracts in R
- Add data to MSTR desktop
- Create dashboards
- Save, and export as PDF
- https://www.microstrategy.com/us/desktop

GSA

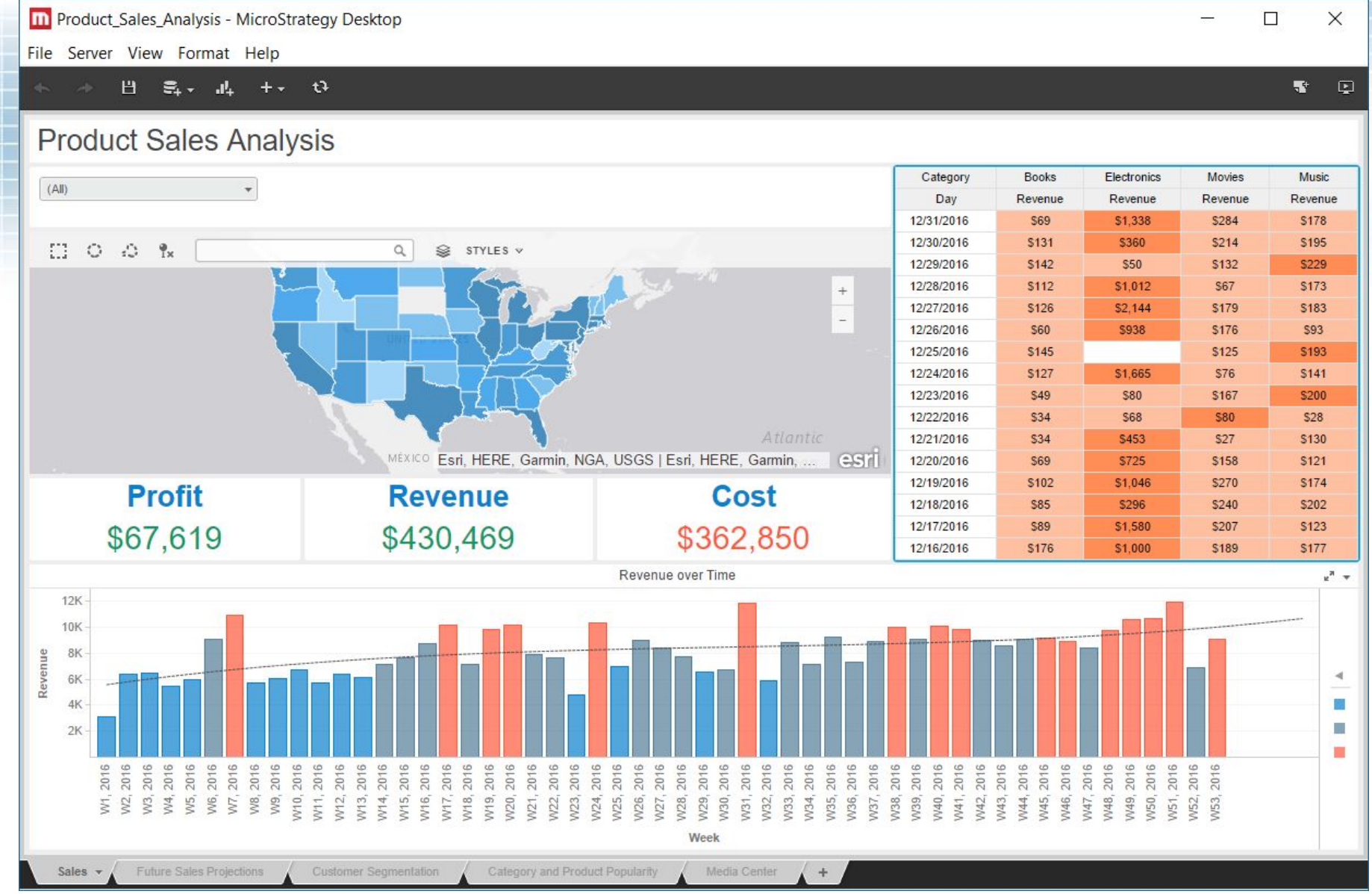
D2D
DATA TO DECISIONS



GSA

D2D
DATA TO DECISIONS


Slot_Machine_Performance - MicroStrategy Desktop
File Server View Format Help
Slot Hold by Slot Machine Type
25 $543,914.67
10 $556,867.08
100 $659,916.97
2 $1,095,205.81
0.01 $1,311,697.13
0.5 $1,739,362.83
5 $1,928,089.90
1 $3,026,255.88
0.05 $7,034,573.60
0.25 $12,004,078.92
2M 4M 6M 8M 10M 12M
Slot Hold
Percentage of Slot Hold
0% - 20%
20% - 40%
40% - 60%
60% - 80%
80% - 100%
Key Performance Indicator Review
Jackpot Amount
520K 440K 360K
Slot Hold
800K 720K 640K 560K
13.2M 13.4M 13.6M 13.8M 14M 14.2M
Coin In
Slot Machine Daily Activity
Jac... | Slot... | Mov... | Coi...
4.5M 4M 3.5M 3M 2.5M 2M 1.5M 1M 500K
$3,378,262.01 $2,639,526.70 $3,284,692.75 $2,562,975.77 $3,126,594.72 $2,574,149.29 $3,415,368.16 $3,456,858.92 $2,438,601.86 $3,467,685.38 $2,473,612.05 $3,699,486.65 $2,465,281.14 $3,540,138.88 $3,153,928.05 $2,668,002.70 $3,521,786.57 $2,744,534.80 $3,456,423.69 $2,612,861.34 $3,596,009.64 $2,430,896.41 $3,510,976.35 $2,647,827.53 $3,530,190.66 $2,582,386.29 $3,591,160.
$114,516.36 $161,862.78 $74,073.48 $106,421.11 $168,860.24 $90,924.05 $185,666.25 $204,382.82 $147,937.55 $124,636.31 $107,992.31 $109,802.86 $136,210.93 $191,741.55 $158,101.89 $146,750.12 $76,441.36
6/28/2015 6/29/2015 6/30/2015 7/1/2015 7/2/2015 7/3/2015 7/4/2015 7/5/2015 7/6/2015 7/7/2015 7/8/2015 7/9/2015 7/10/2015 7/11/2015 7/12/2015 7/13/2015 7/14/2015 7/15/2015 7/16/2015 7/17/2015 7/18/2015 7/19/2015 7/20/2015 7/21/2015 7/22/2015 7/23/2015 7/24/2015 7/25/2015 7/26/2015 7/27/2015 7/28/2015 7/29/2015 7/30/2015 7/31/2015 8/1/2015 8/2/2015 8/3/2015 8/4/2015 8/5/2015 8/6/2015 8/7/2015 8/8/2015 8/9/2015 8/10/2015 8/11/2015 8/12/2015 8/13/2015 8/14/2015 8/15/2015 8/16/2015 8/17/2015 8/18/2015 8/19/2015 8/20/2015 8/21/2015 8/22/2015 8/23/2015 8/24/2015 8/25/2015 8/26/2015 8/27/2015 8/28/2015 8/29/2015 8/30/2015 8/31/2015 9/1/2015 9/2/2015 9/3/2015 9/4/2015 9/5/2015
Casino Overview
Slot Machine and Games



Category	Books	Electronics	Movies	Music
Day	Revenue	Revenue	Revenue	Revenue
12/31/2016	$69	$1,338	$284	$178
12/30/2016	$131	$360	$214	$195
12/29/2016	$142	$50	$132	$229
12/28/2016	$112	$1,012	$67	$173
12/27/2016	$126	$2,144	$179	$183
12/26/2016	$60	$938	$176	$93
12/25/2016	$145		$125	$193
12/24/2016	$127	$1,665	$76	$141
12/23/2016	$49	$80	$167	$200
12/22/2016	$34	$68	$80	$28
12/21/2016	$34	$453	$27	$130
12/20/2016	$69	$725	$158	$121
12/19/2016	$102	$1,046	$270	$174
12/18/2016	$85	$296	$240	$202
12/17/2016	$89	$1,580	$207	$123
12/16/2016	$176	$1,000	$189	$177

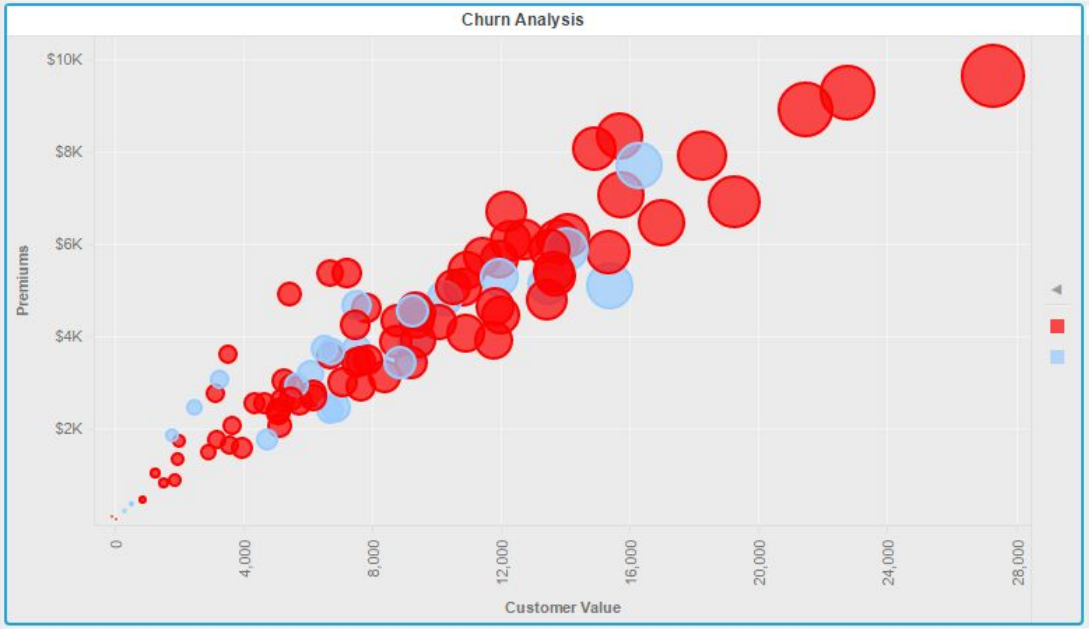
GSA

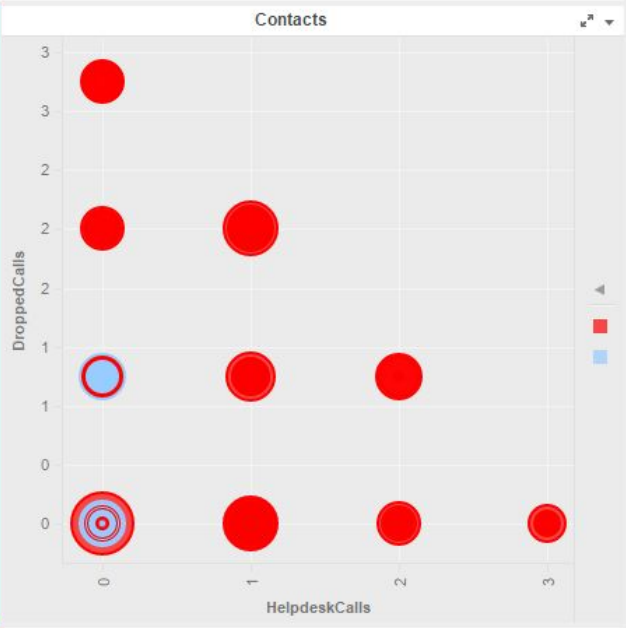
D2D
DATA TO DECISIONS

Insurance_Churn_Prediction_Analysis - MicroStrategy Desktop

File Server View Format Help

(All) Life Non-Life





Policyholders

Policyholder		Category	Subcategory	Household Count	Income Bracket	Powiat	Churn	Premiums	Loss	Customer Value	HelpdeskCalls	DroppedCalls
Total							**75**	**$392,047**	**$332,414**	**852,338**	**81**	**82**
Agan	Regina	Non-Life	Travel	6	70K-80K	Swietochlowice	1	$9,632	$7,983	27,297	0	0
Agranov	Myra	Non-Life	Travel	5	70K-80K	Myslowice	1	$2,527	$2,222	4,364	1	0
Ali	Zachary	Life	Whole Maxim	5	70K-80K	wodzislawski	1	$3,095	$2,506	8,409	1	2
Archuleta	Otha	Life	Universal 131	6	70K-80K	Ruda Slaska	1	$2,881	$2,411	7,667	0	3

Churn Premiums Sales Rep +

GSA

D2D DATA TO DECISIONS

Q & A