Data Management and Visualization

Dr. Ashish Kumar Jha

Session 5

Principles of Visualization

Agenda

What is role of visualization in analytics?

Why is it important?

Visualization before analytics?

Visualization principles

Big Data Visualization

Analytics and Visualization

Data → Information → Visualization → Story → Learning → Action

Descriptive – Explorative – Predictive – Prescriptive

Goldfish Effect

Perception and Cognition

Visual perception constitutes about 70% of the total human perceptual capabilities

2/3 of the cerebral cortex is indirectly "involved" in vision

It takes about 0.15 seconds from the moment light hits the retina to when the earliest recognition of basic object identity can occur

Visualization

12323456584035762856305624284537956341385634648438563427495543850352413957264656247629 4725194027848

12323456584035762856305624284537956341385634648438563427495543850352413957264656247629472519402784 8

Visualization

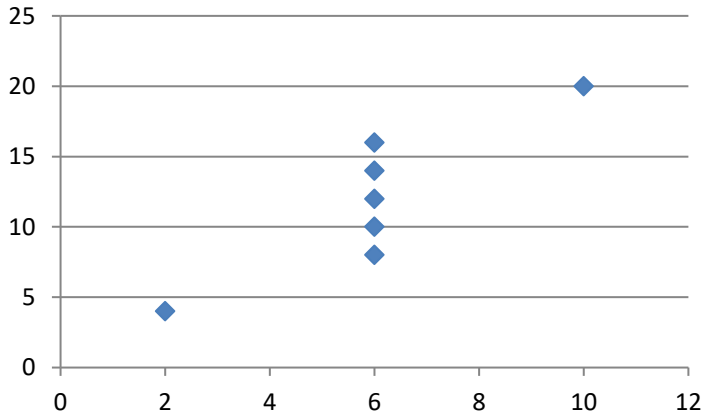
12323456584035762856305624284537956341385634648438563427495543850352413957264656247629472519402784 8

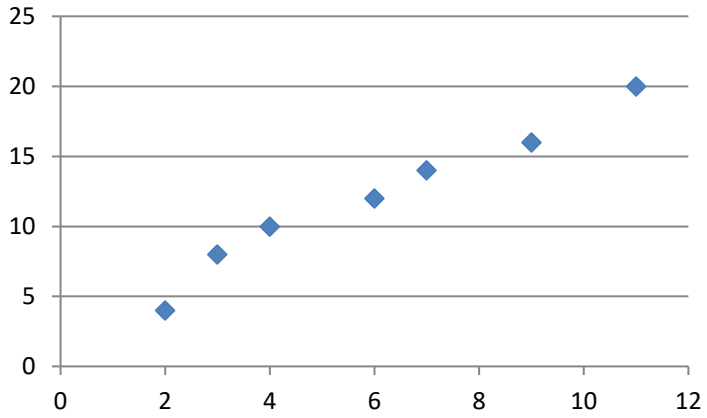
No. of 3= 13

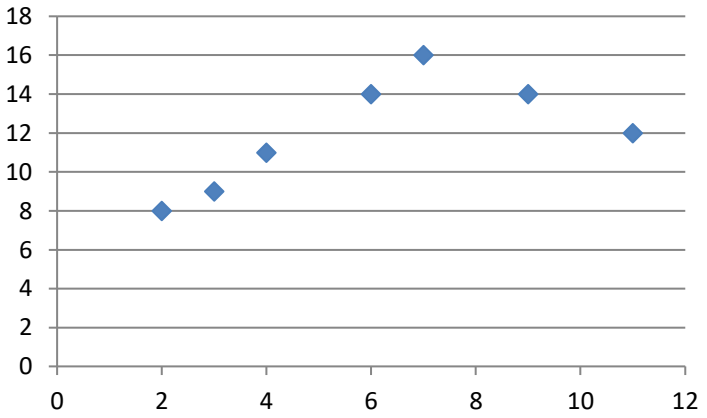
12323456584035762856305624284537956341385634648438563427495543850352413957264656247629472519402784 8

Visualization

	Company 1	Company 2	Company 3
Average Experience of Employees	6	6	6
Average Salary	12	12	12





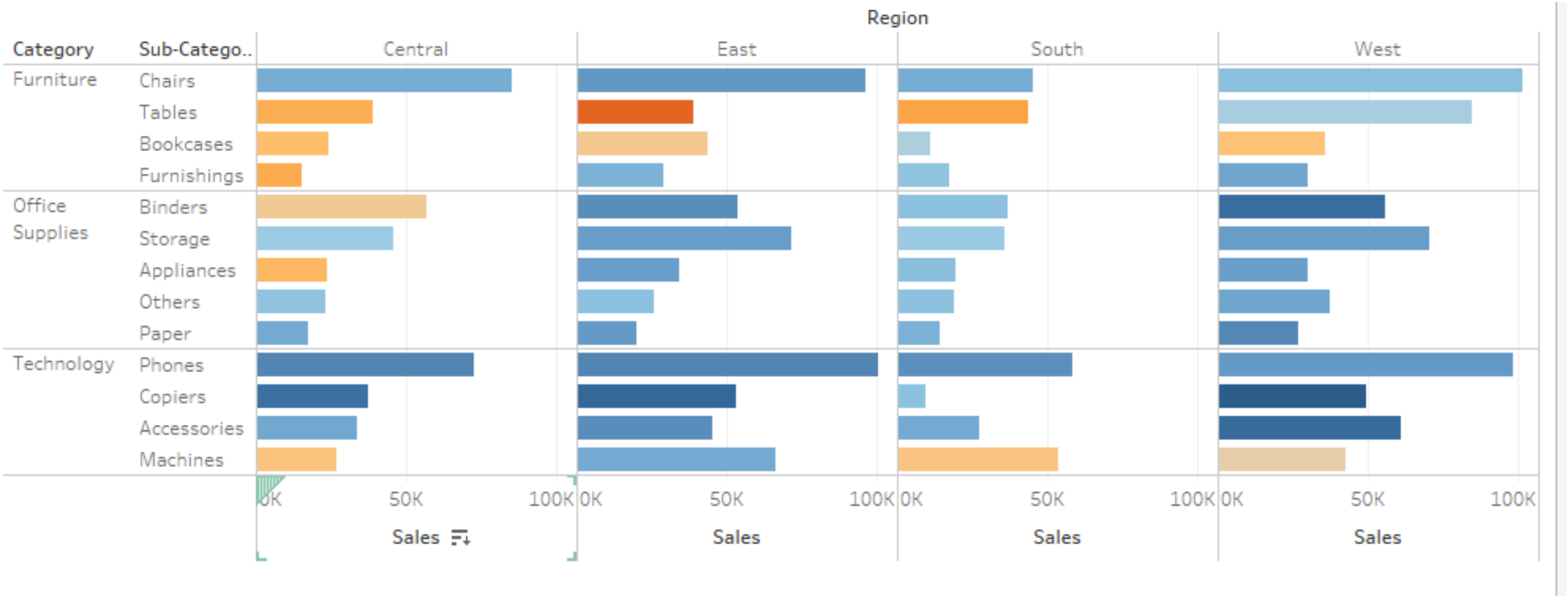


		Region			
Category	Sub-Catego..	Central	East	South	West
Furniture	Chairs	85,231	96,261	45,176	101,781
	Tables	39,155	39,140	43,916	84,755
	Bookcases	24,157	43,819	10,899	36,004
	Furnishings	15,254	29,071	17,307	30,073
Office Supplies	Binders	56,923	53,498	37,030	55,961
	Storage	45,930	71,613	35,768	70,533
	Appliances	23,582	34,188	19,525	30,236
	Others	23,099	26,044	19,177	37,459
	Paper	17,492	20,173	14,151	26,664
Technology	Phones	72,403	100,615	58,304	98,684
	Copiers	37,260	53,219	9,300	49,749
	Accessories	33,956	45,033	27,277	61,114
	Machines	26,797	66,106	53,891	42,444

Visualization

		Region			
Category	Sub-Catego..	Central	East	South	West
Furniture	Chairs	✓ 85,231	✓ 96,261	✓ 45,176	! 101,781
	Tables	✗ 39,155	✗ 39,140	✗ 43,916	✗ 84,755
	Bookcases	✗ 24,157	✗ 43,819	✗ 10,899	✗ 36,004
	Furnishings	✗ 15,254	✓ 29,071	! 17,307	✓ 30,073
Office Supplies	Binders	✗ 56,923	✓ 53,498	! 37,030	✓ 55,961
	Storage	✗ 45,930	✓ 71,613	✗ 35,768	✓ 70,533
	Appliances	✗ 23,582	✓ 34,188	! 19,525	✓ 30,236
	Others	! 23,099	! 26,044	! 19,177	✓ 37,459
	Paper	✓ 17,492	✓ 20,173	✓ 14,151	✓ 26,664
Technology	Phones	✓ 72,403	✓100,615	✓ 58,304	✓ 98,684
	Copiers	✓ 37,260	✓ 53,219	! 9,300	✓ 49,749
	Accessories	✓ 33,956	✓ 45,033	✓ 27,277	✓ 61,114
	Machines	✗ 26,797	✓ 66,106	✗ 53,891	✗ 42,444

Visualization



Visualization

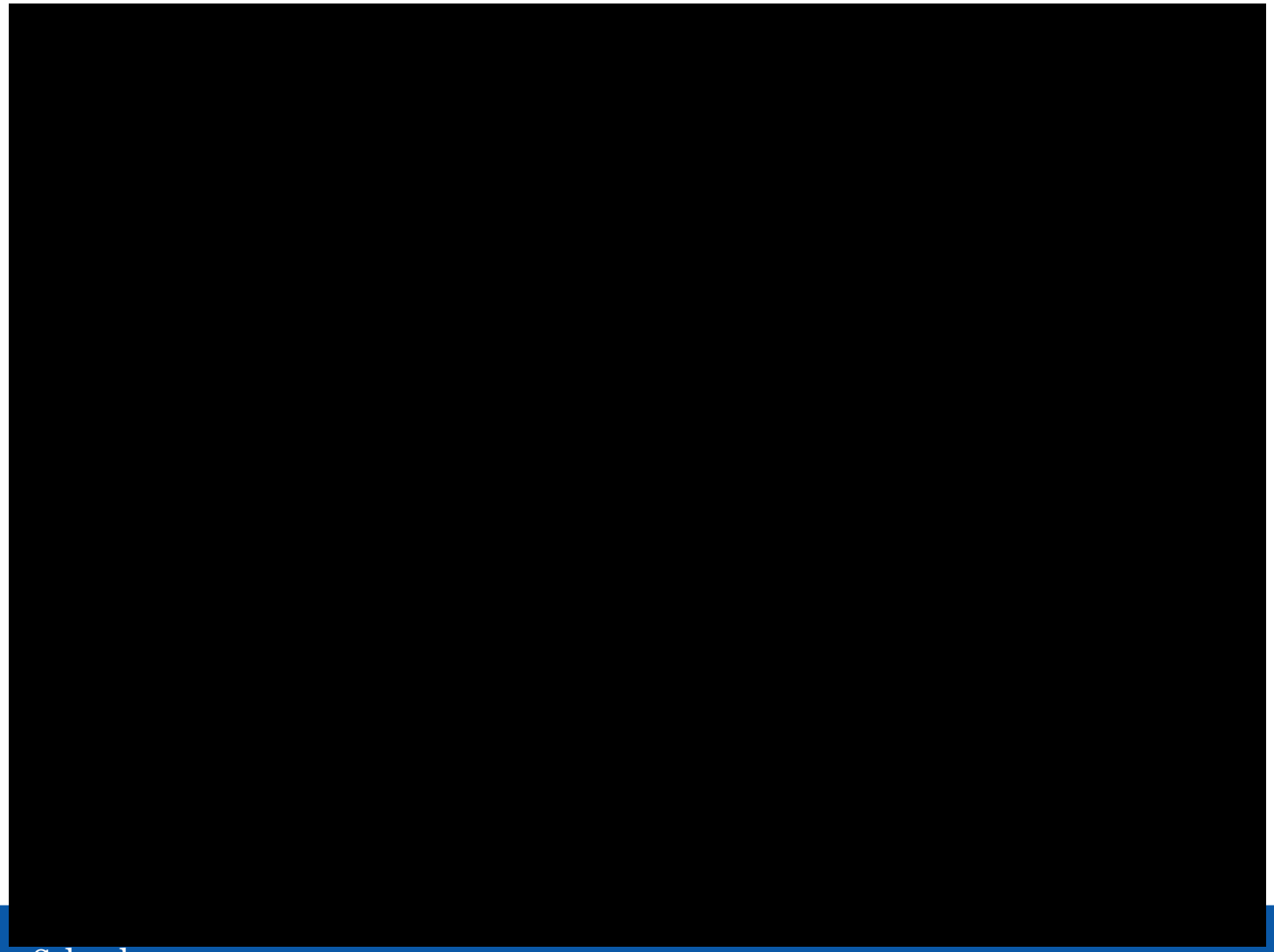
Pre-attentive attributes

Why do our brains like looking at *things* in a certain way?

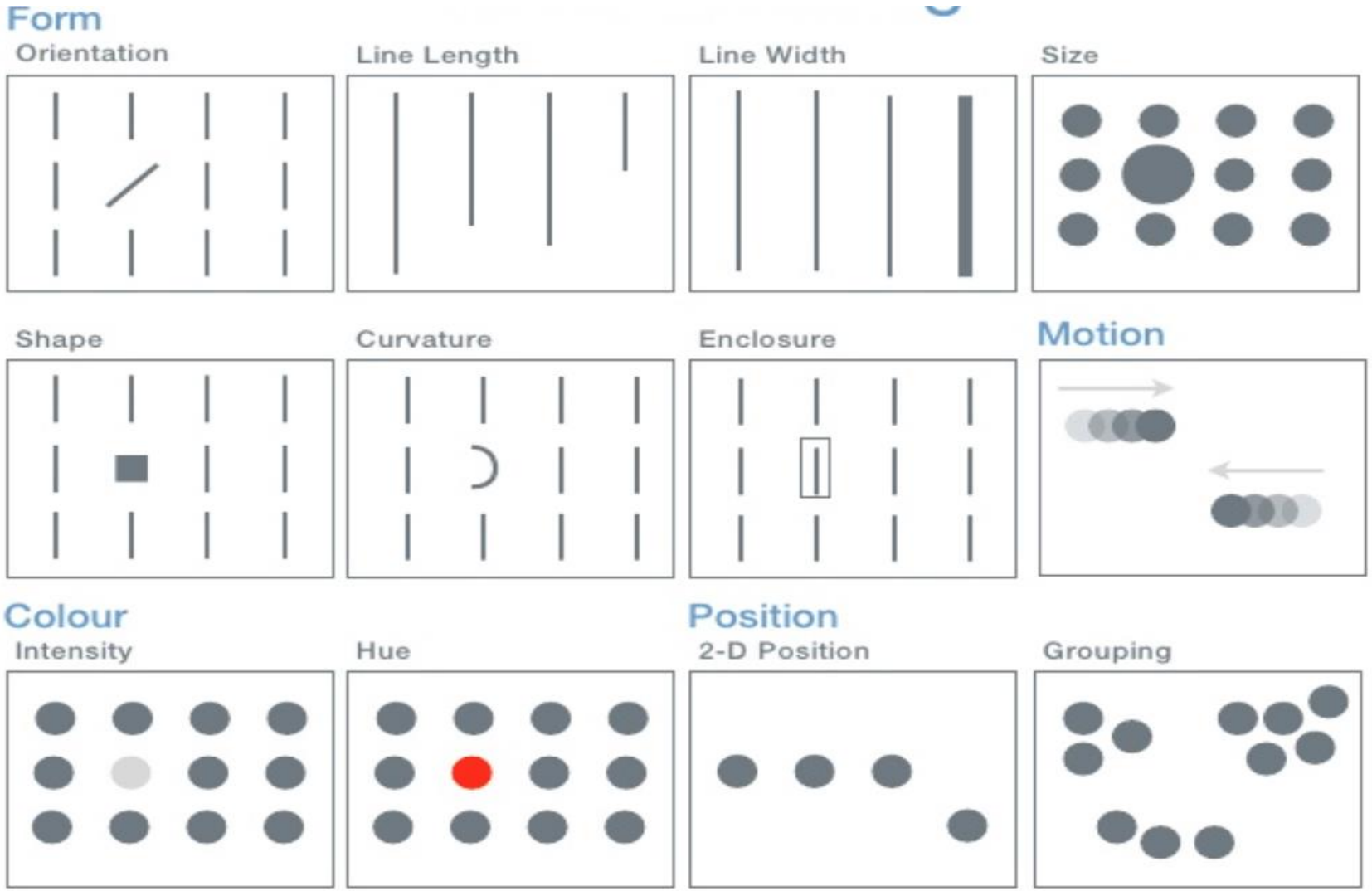
Visual attributes that we notice before we are aware of the message.

Take advantage of perceptual psychology and cognition to represent data to be smarter.

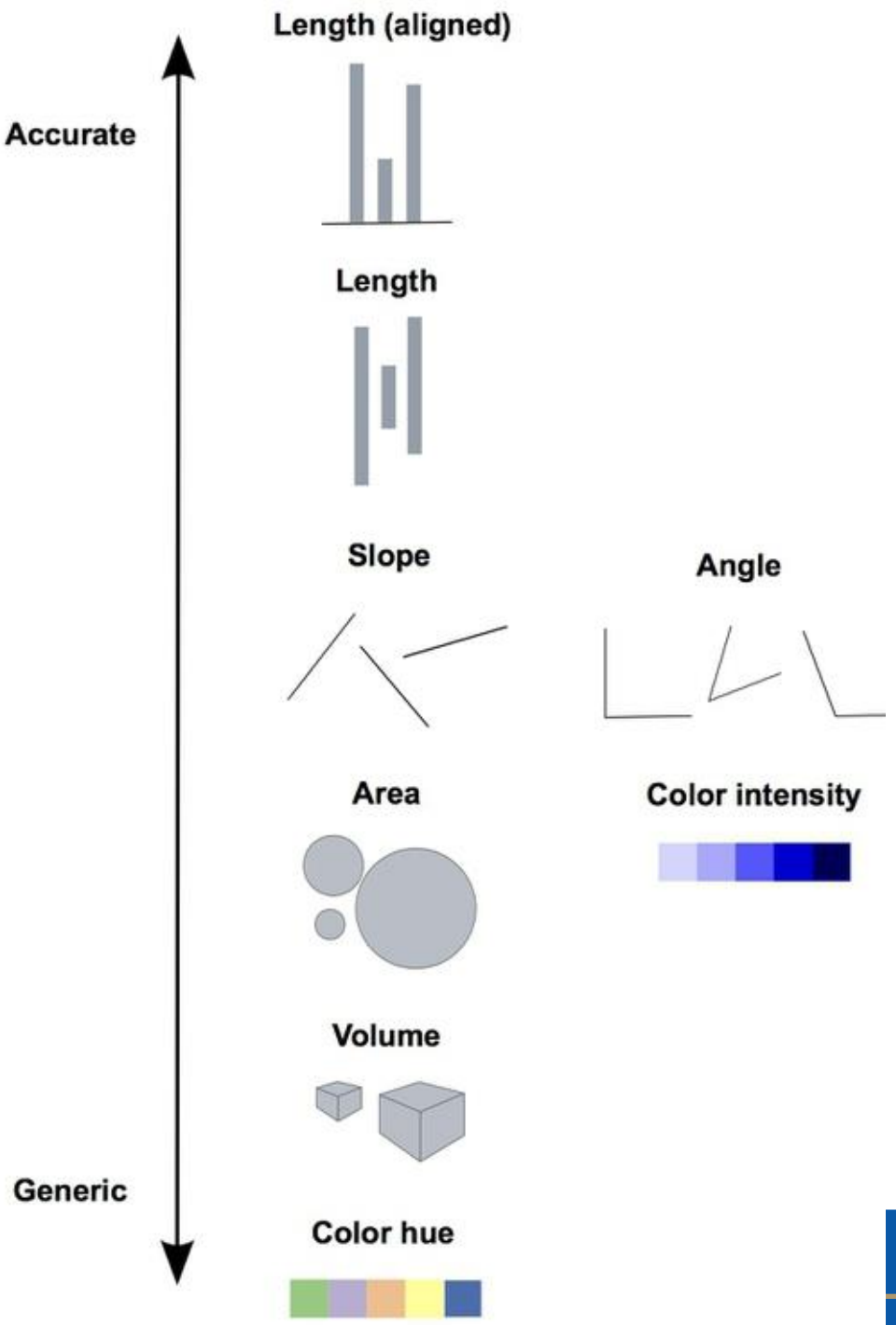
Pre-attentive attributes



Pre-attentive attributes



Source: FlinkLabs Visualizing Marathon



Hierarchy

This perceptual hierarchy of visual cues is important

Revealed by statisticians , William Cleveland and Robert McGill

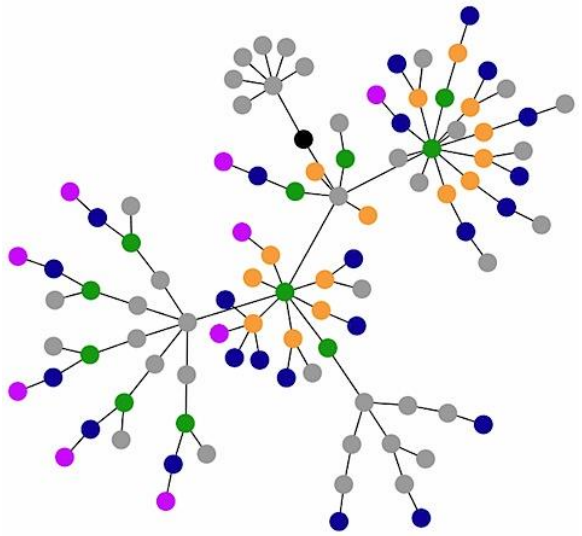
When making comparisons with continuous variables, aim to use cues near the top of the scale wherever possible

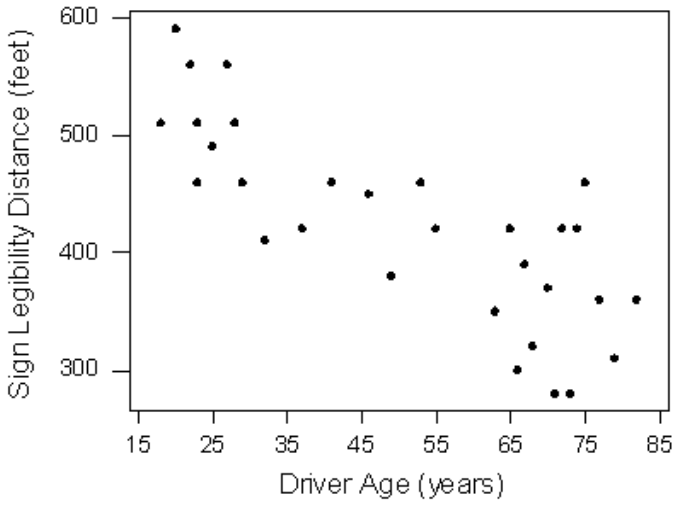
Data Types

Nominal	Ordinal	Quantifiable
Male	Gold	10 Kg
Female	Silver	21 Kg
	Bronze	30 Kg
Asia	Low	$10000
Europe	Medium	$20000
Australia	High	$30000

Data Types - Attributes

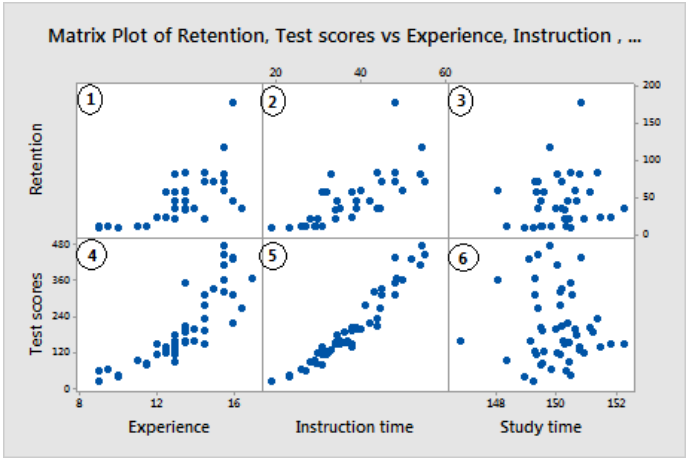
Nominal	Ordinal	Quantifiable
Position	Position	Position
Hue	Hue	Limited Hue
Shape	Shape	Intensity
Cluster	Intensity	Size
Enclosure	Size	Orientation
		Length
		Width

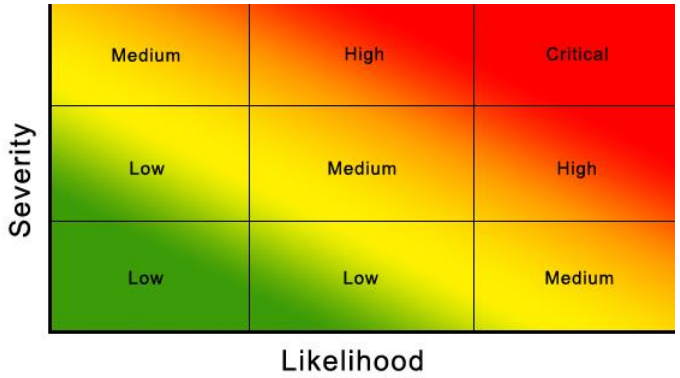




Data Representations

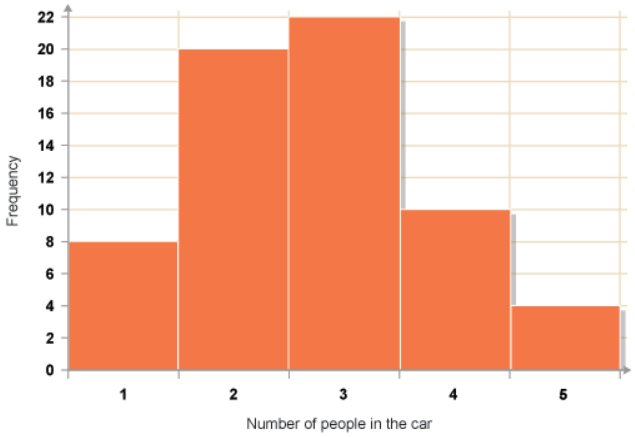
Relationship among data points

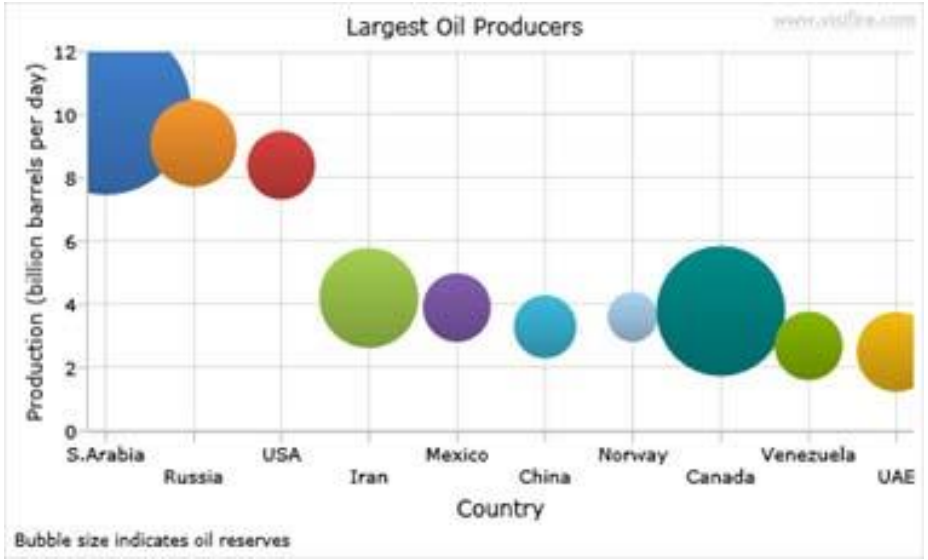


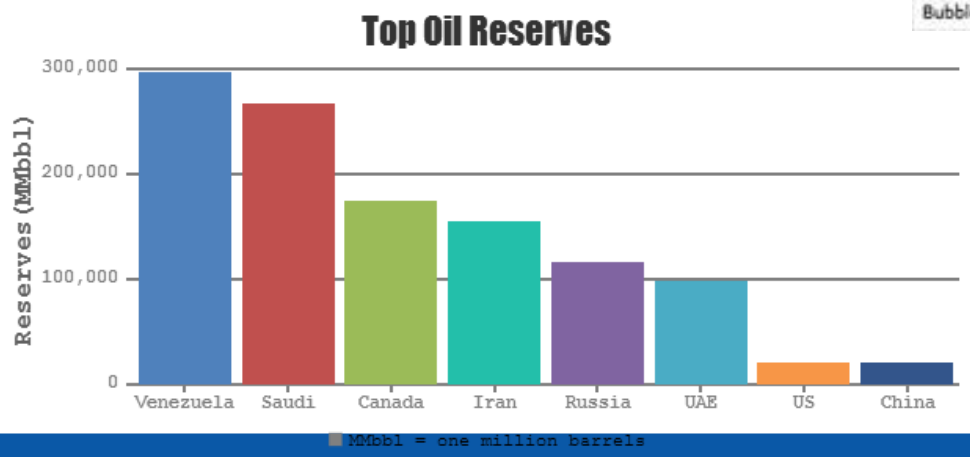


Data Representations

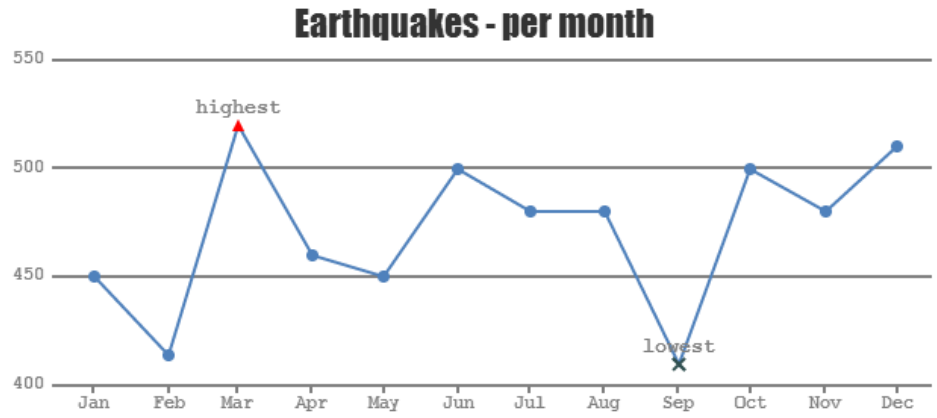
Comparing set of values



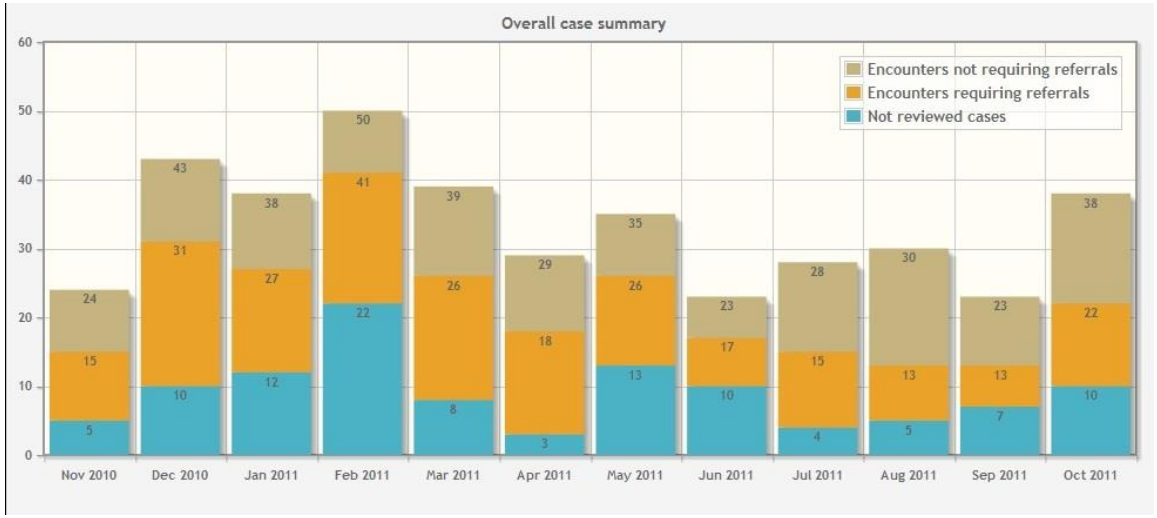




Data Representations

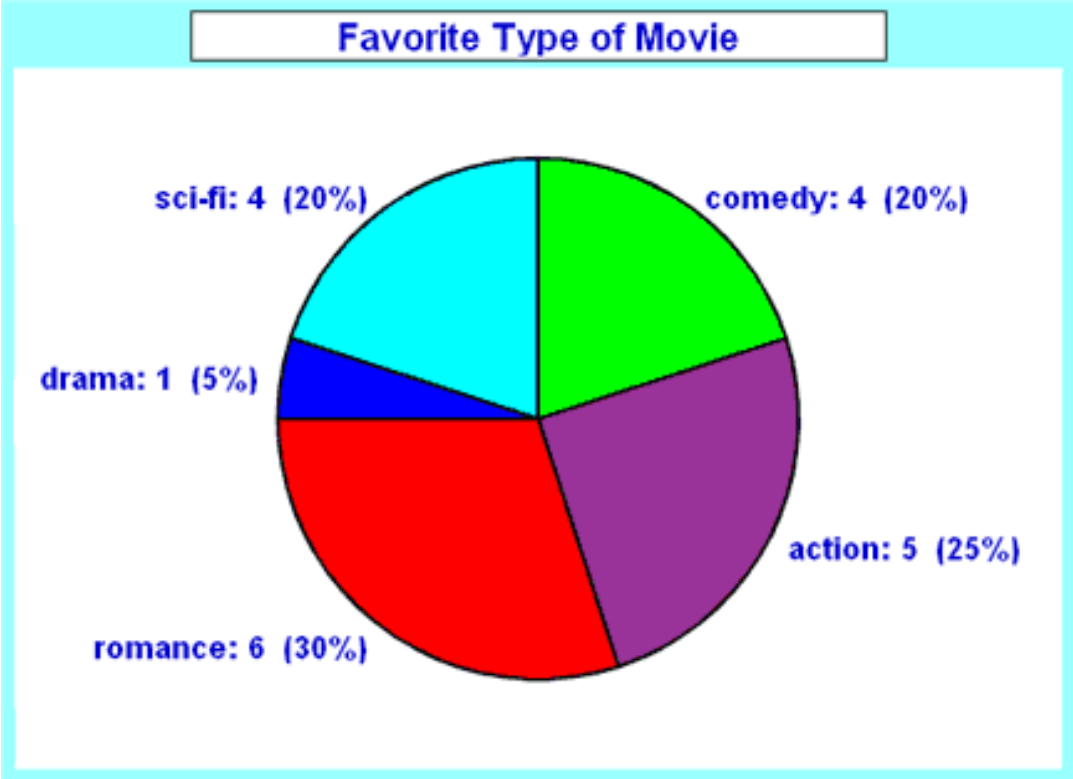


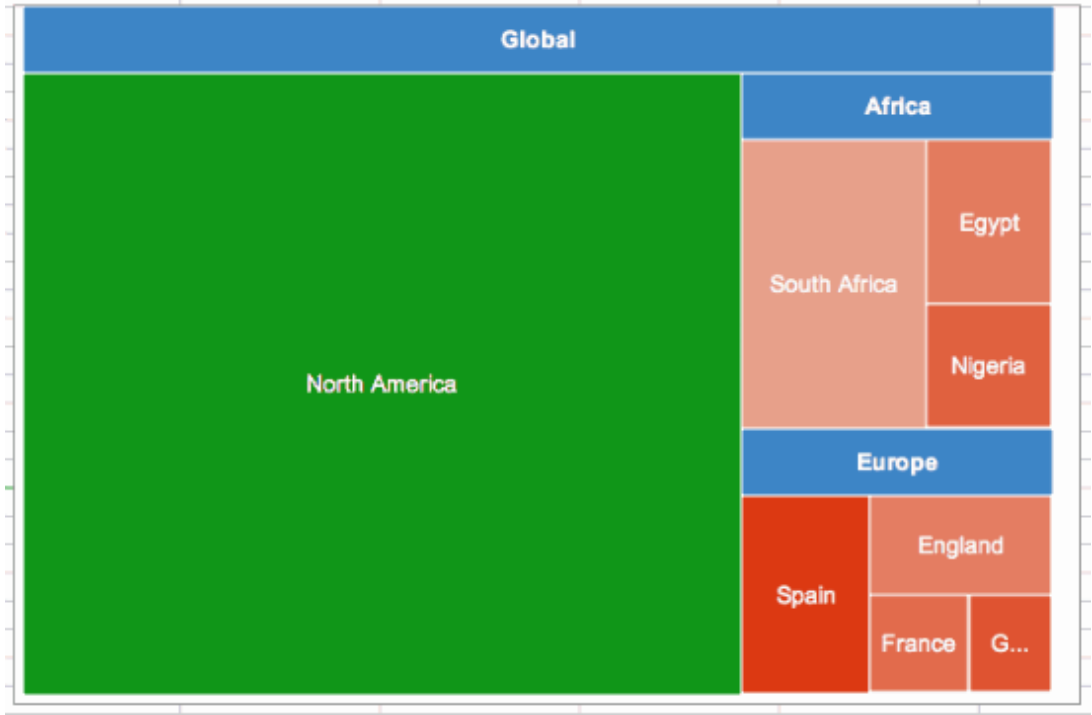
Change of one value with respect to other



Composition of parts

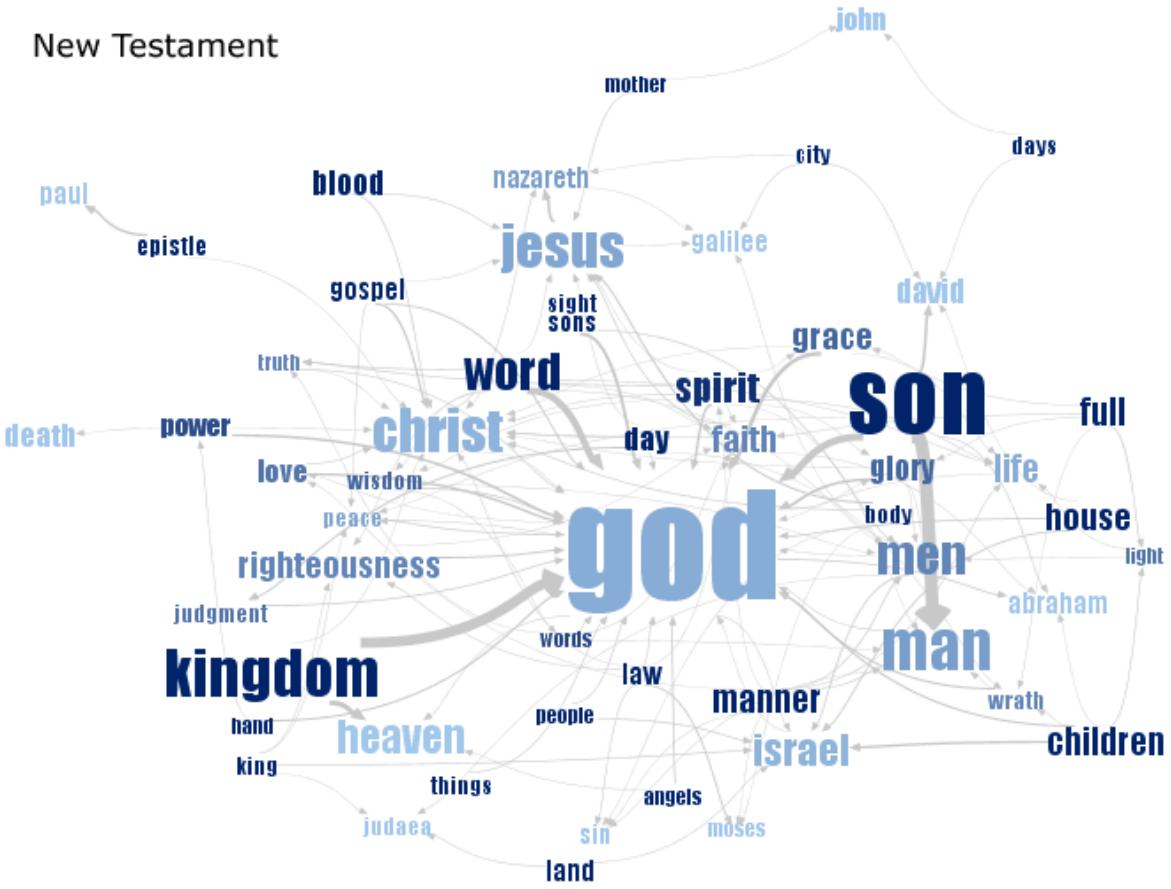
Data Representations

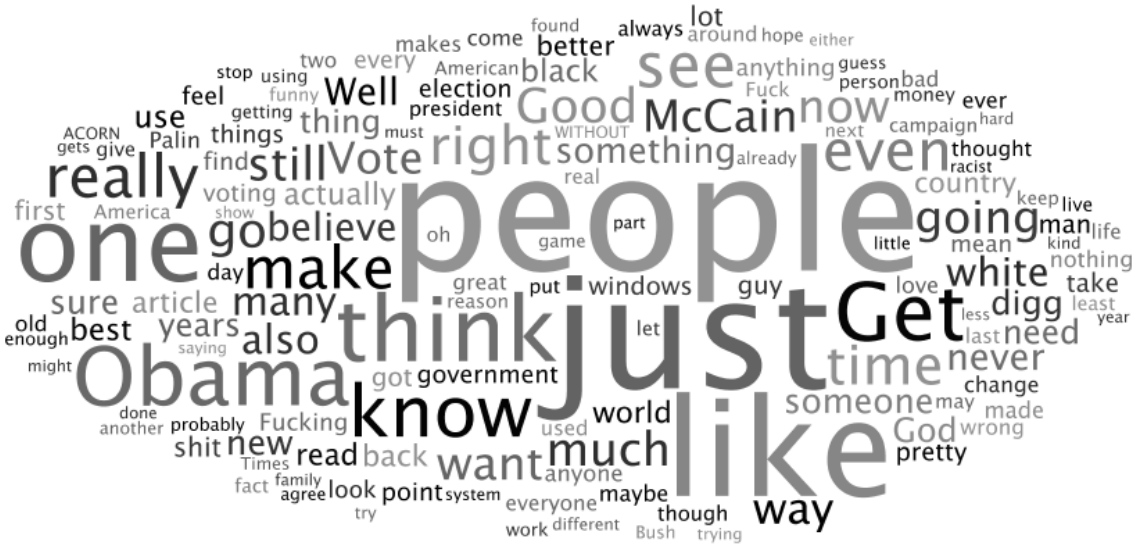




Data Representations

Text Data





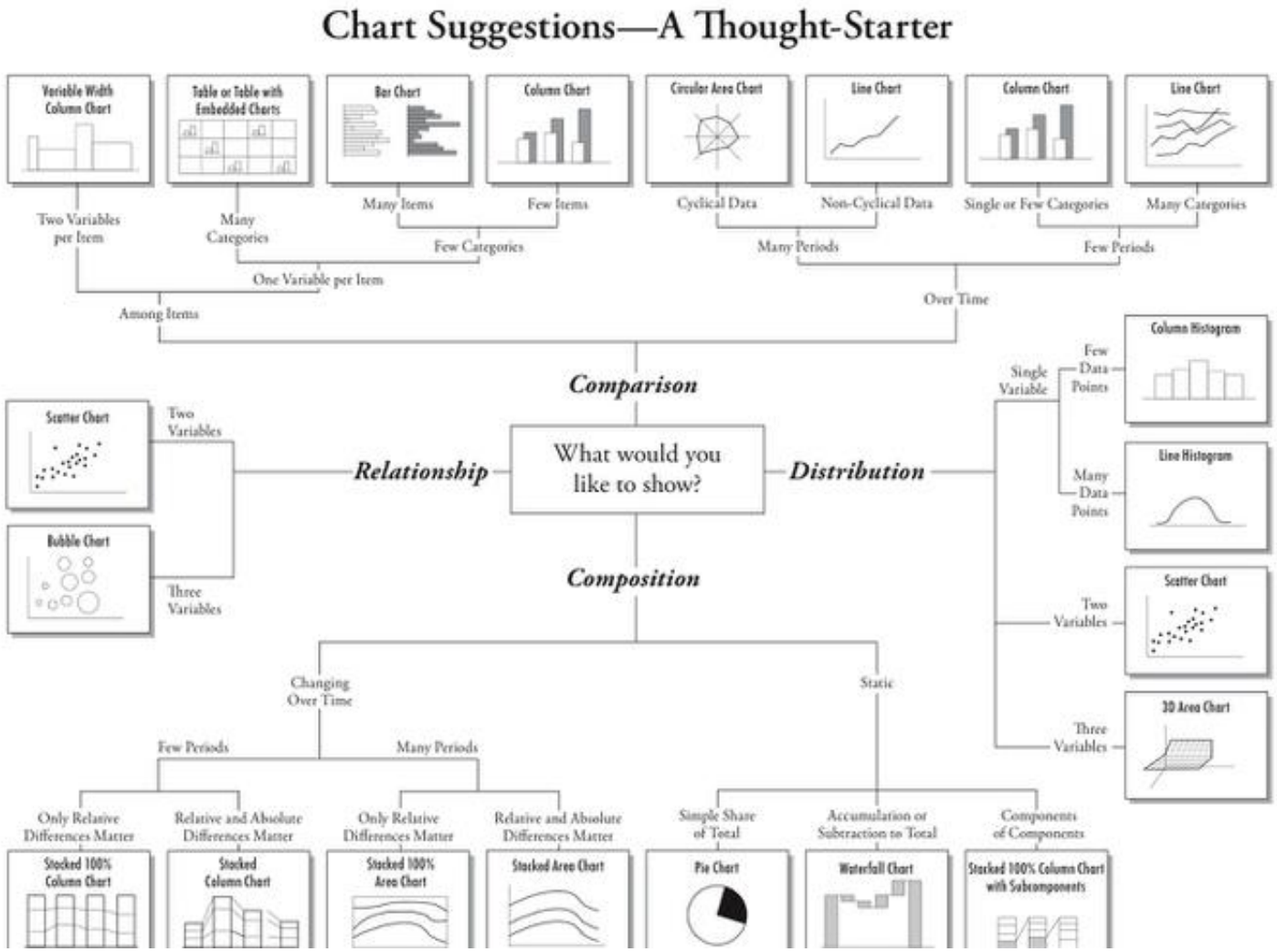


Chart options

Source: Dr. Andrew Abela, Chart Suggestions, a Thought Starter

Visual Analytics

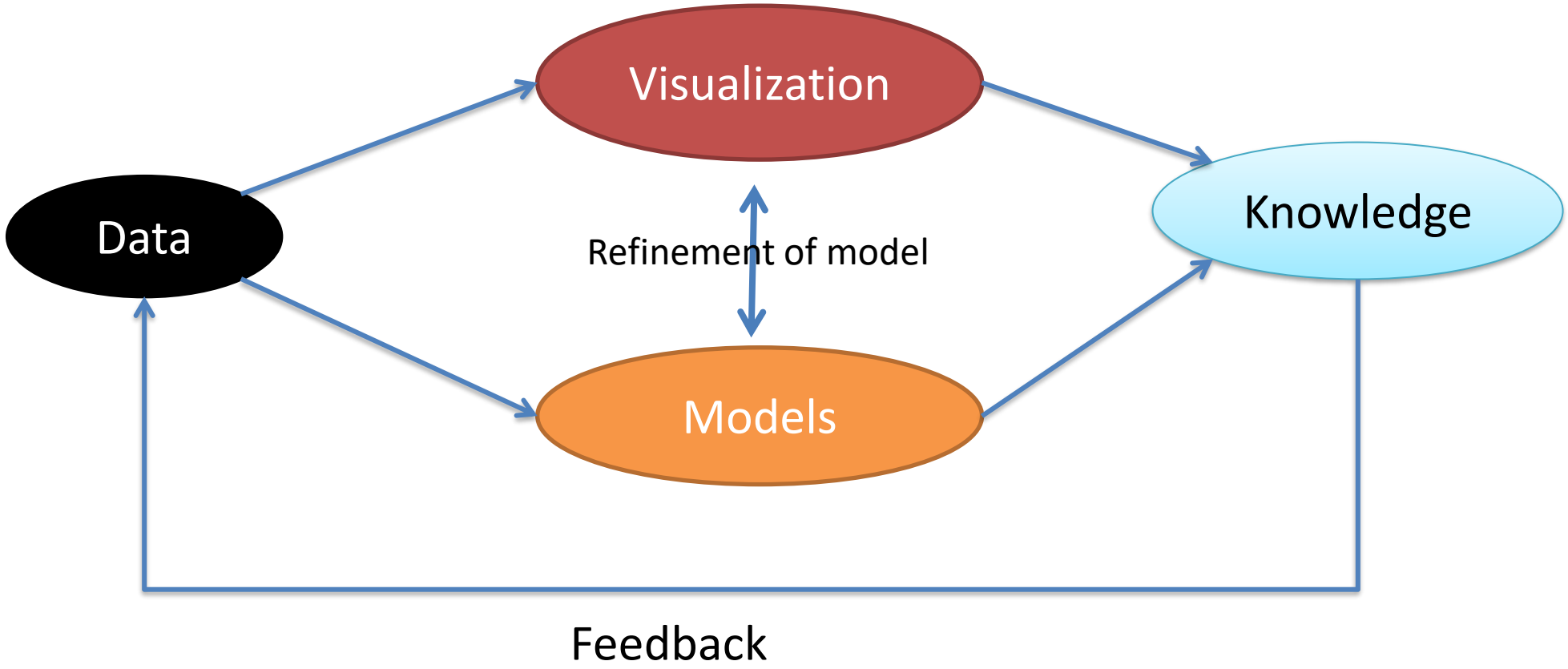
"Visual Analytics is the science of analytical reasoning supported by interactive visual interfaces."

- Thomas, J., Cook, K.: Illuminating the Path: Research and Development Agenda for Visual Analytics. IEEE-Press

"It involves representation of data to exploit our visual perception abilities in order to amplify cognition"

- Andy Kirk : Data Visualization: A successful Design Process

Visual Analytics



Visual Analytics

Explanation vs. Exploration enabling

Representational primacy" versus "Analytic primacy"

- Telling truth about data vs. providing analytically useful visualizations

Traditional thinking of Analytics is closer to "Hypothesis testing"

- By stating hypotheses up front, limit variables and sharpens thinking, more precise measurement
- Too far from reality, initial hypotheses bias toward finding evidence to support it

Visual Analytics- Exploratory Data Analysis

- Find the interesting things this way, we now have computational capabilities to do them
- Not generalizable, everything is a special case, detecting statistical replationships does not infer cause and effect

Visual Analytics

Best Principles of Visualization – "*Tell a Story with Data*"

- Visualization should be Incremental
- Visualization should be expressive
- Visualization should be direct

Popular Visual Analytics Tools

- Qlikview
- Tableau
- SAS Visual Analytics
- Tibco Spotfire
- Microsoft – Power Pivot and Power BI

Tableau

Pointers for tableau

- Tableau your data
- Tableau Public
- Tableau resources online

Big Data Visualization

How is it different?

Big Data & visualization

The V's of big data – Volume, Variety, Velocity

Visualization more important to mash up disparate data sources to create custom analytical views

Visualization is the "front end" of big data

Designing a new visualization with efficient indexing is not easy in big data

Big Data & visualization

Myths

- All data must be visualized
- Only good data should be visualized
- Visualization will always manifest the right decision or action
- Visualization will lead to certainty.

Challenges with Big data

- Diversity
- Heterogeneity
- Scalability
- Dynamics

Big Data & visualization

Tree Map

Circle packing

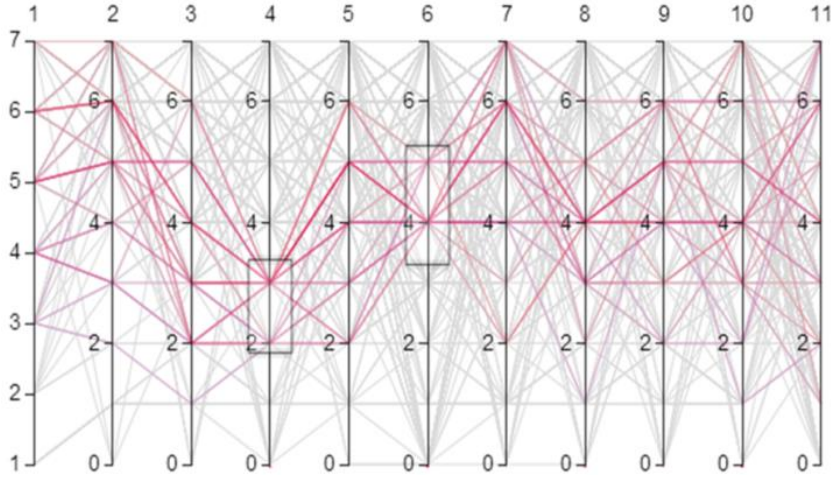
Sunburst diagram

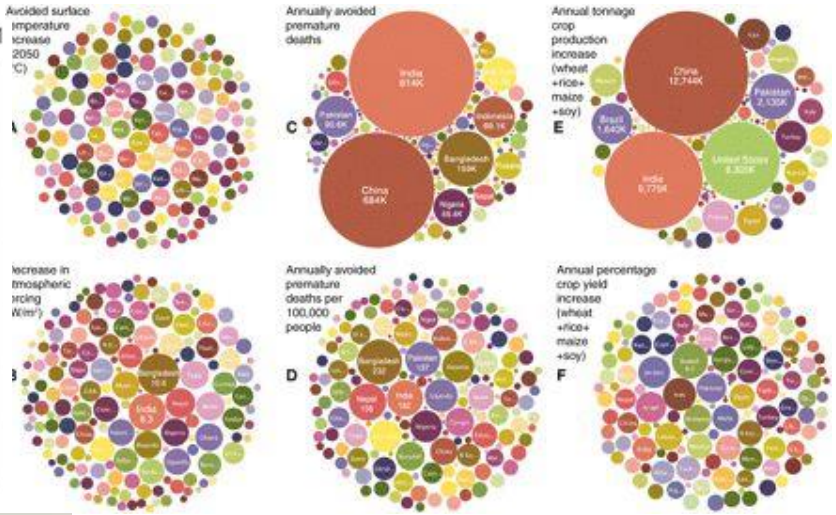
Parallel Coordinates

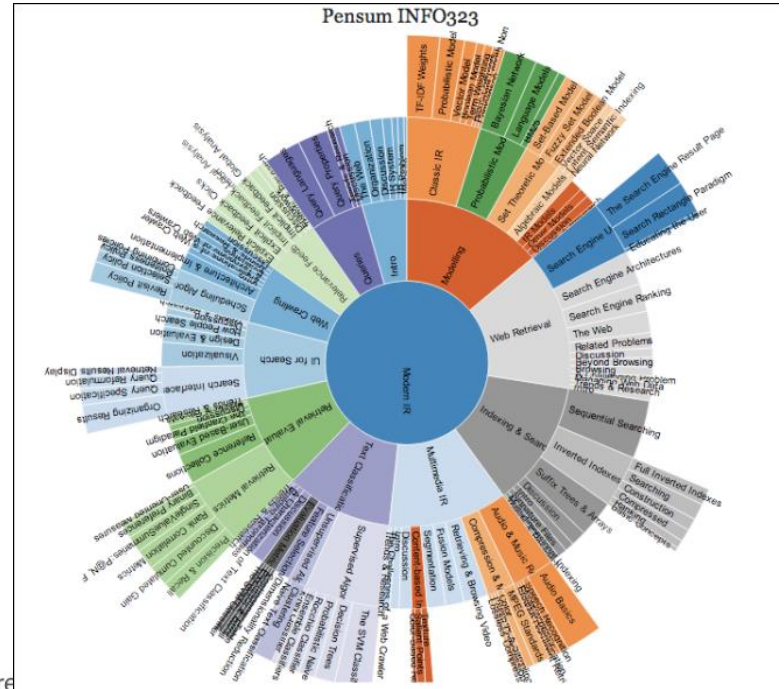
Streamgraph

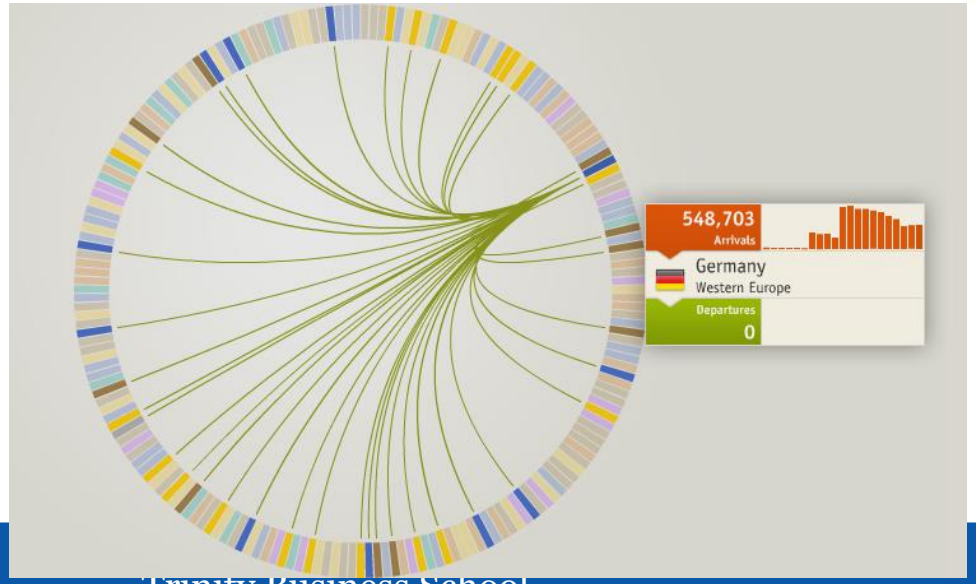
Circular network diagram

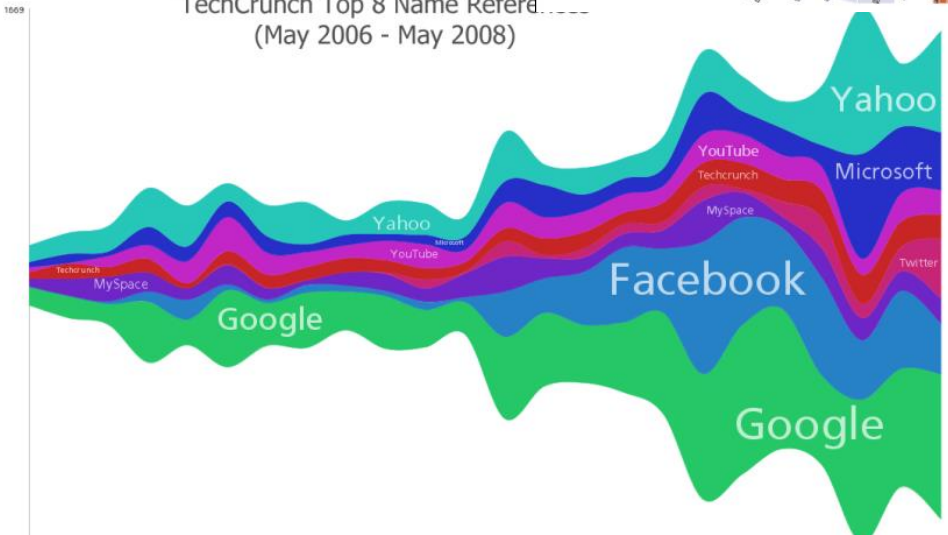
Big Data & visualization









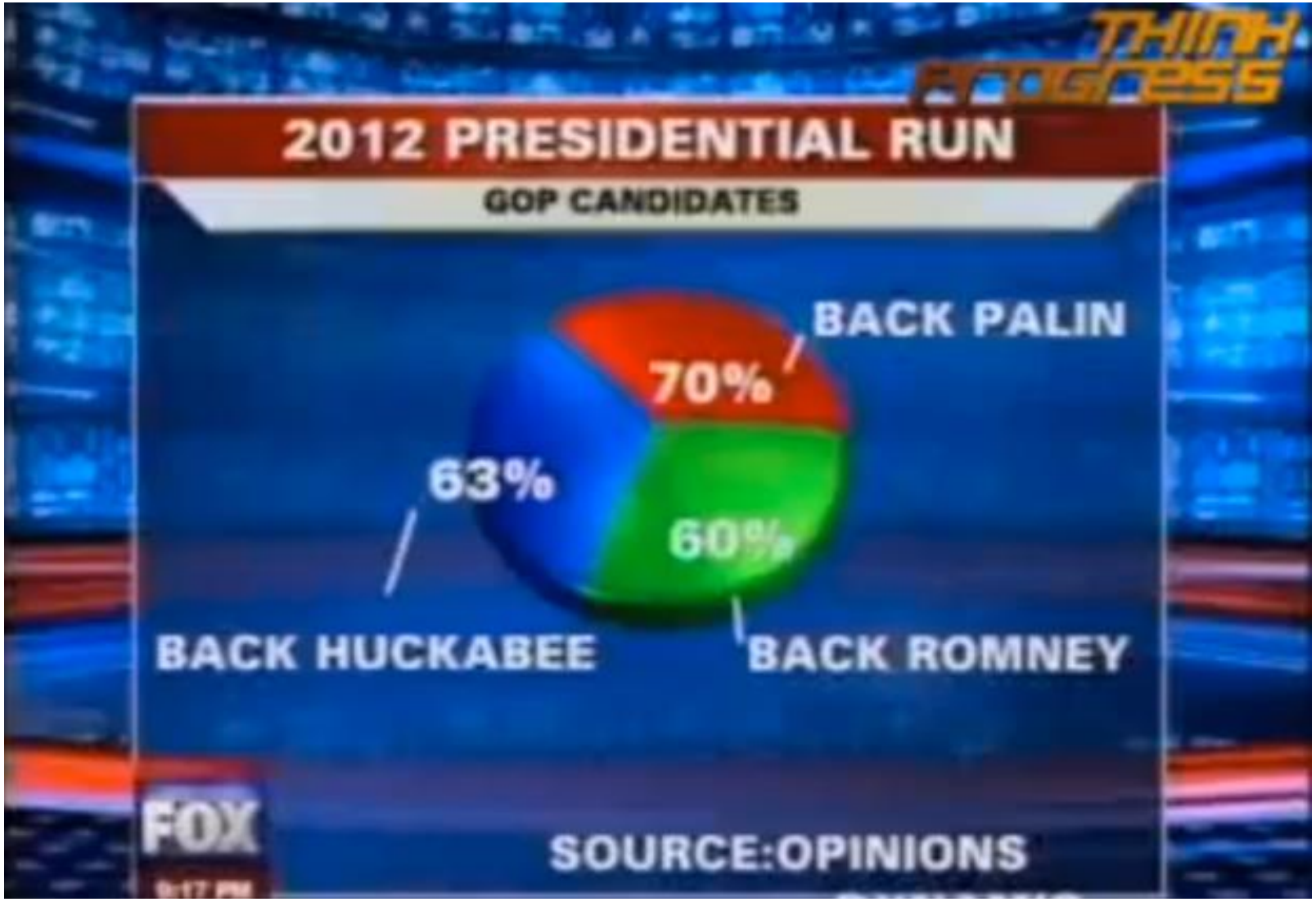


Big Data & visualization

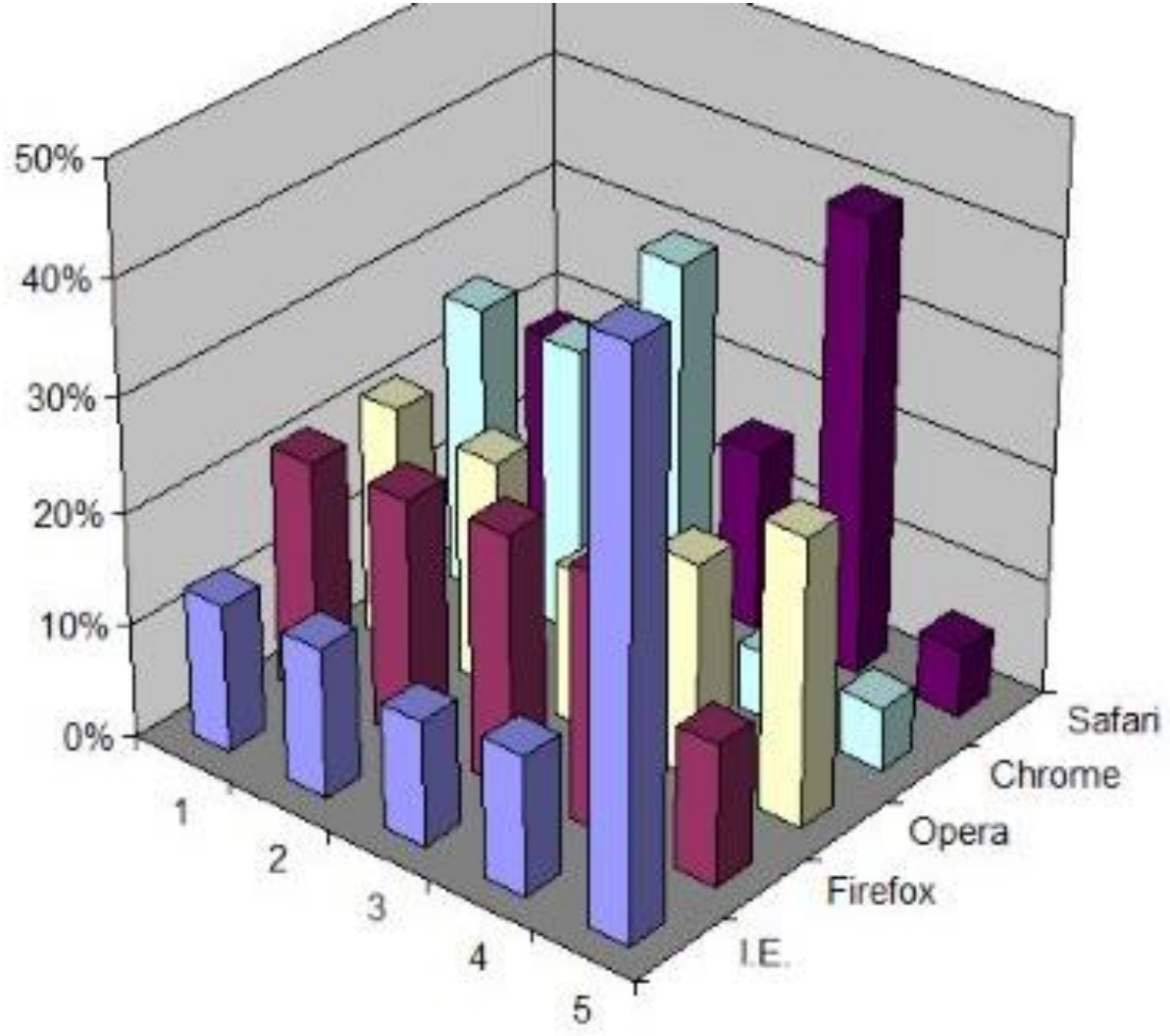
Method name	Large data volume	Data variety	Data dynamics
Treemap	+	-	-
Circle packing	+	-	-
Sunburst	+	-	+
Parallel coordinates	+	+	+
Streamgraph	+	-	+
Circular network diagram	+	+	-

Method name	Big data class
Treemap	Can be applied only to hierarchical data
Circle packing	Can be applied only to hierarchical data
Sunburst	Volume + Velocity
Parallel coordinates	Volume + Velocity + Variety
Streamgraph	Volume + Velocity
Circular network diagram	Volume + Variety

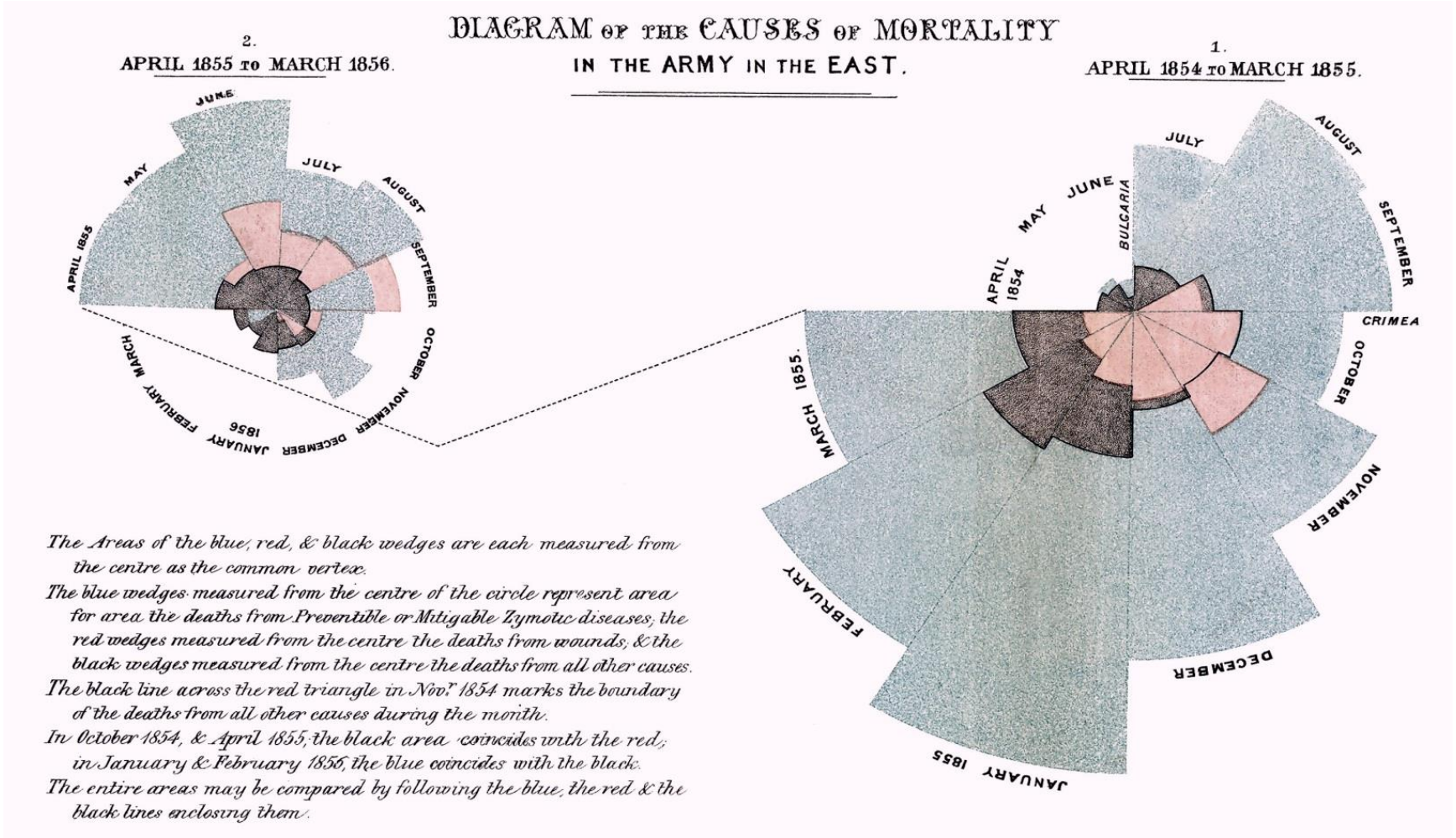
Best Practices in Data Visualization



Weird Visualizations

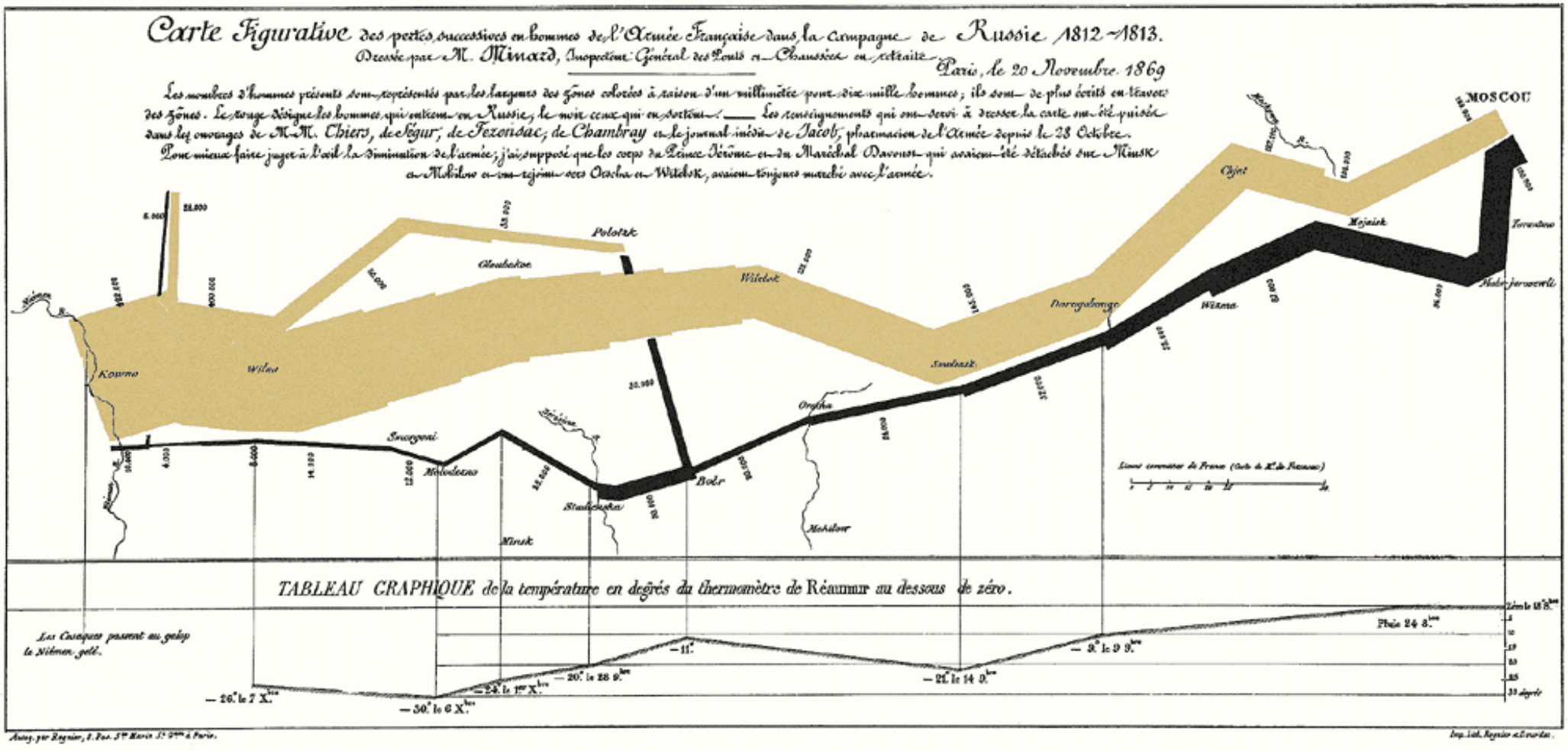


Weird Visualization

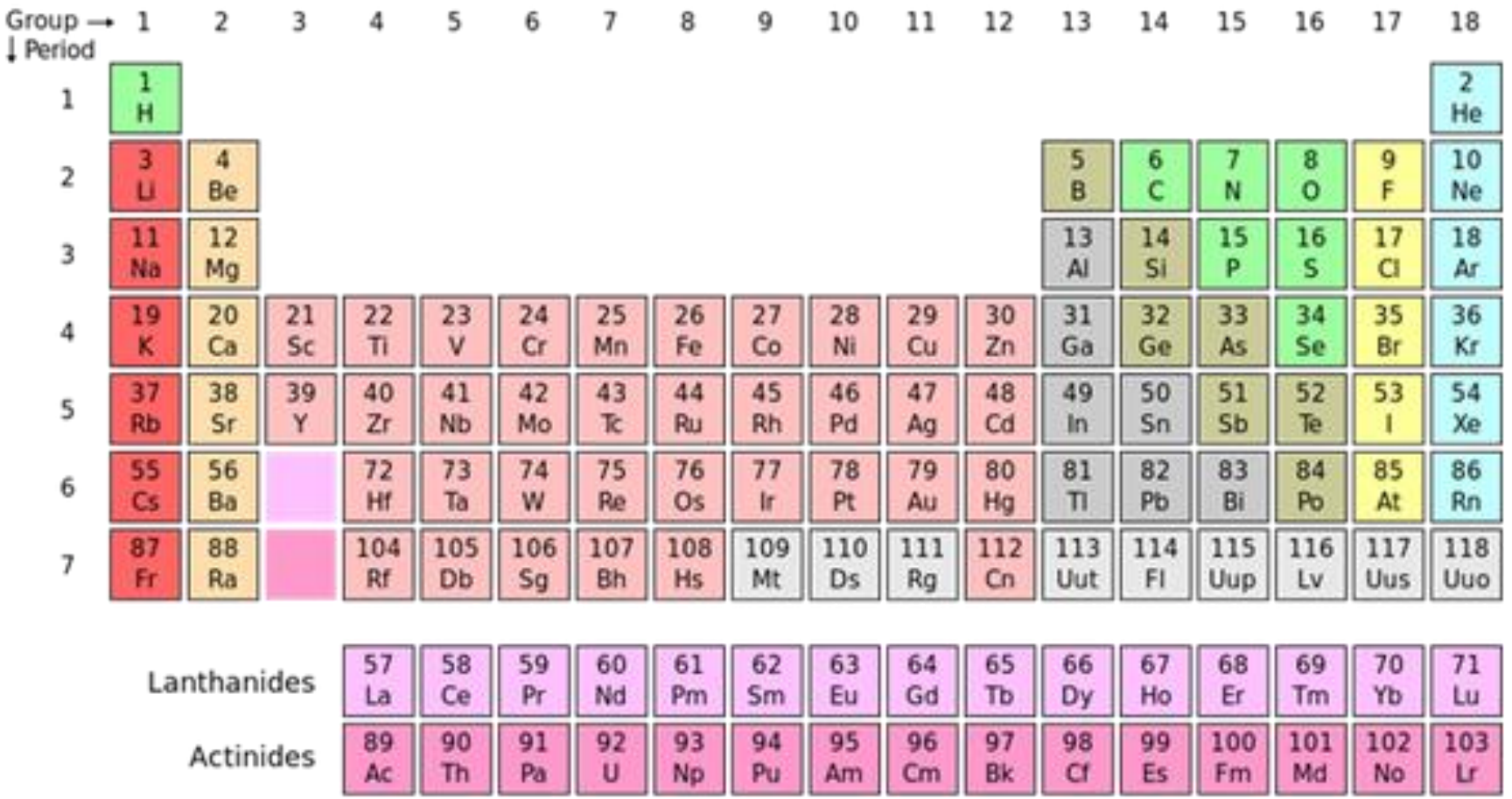


Great visualizations

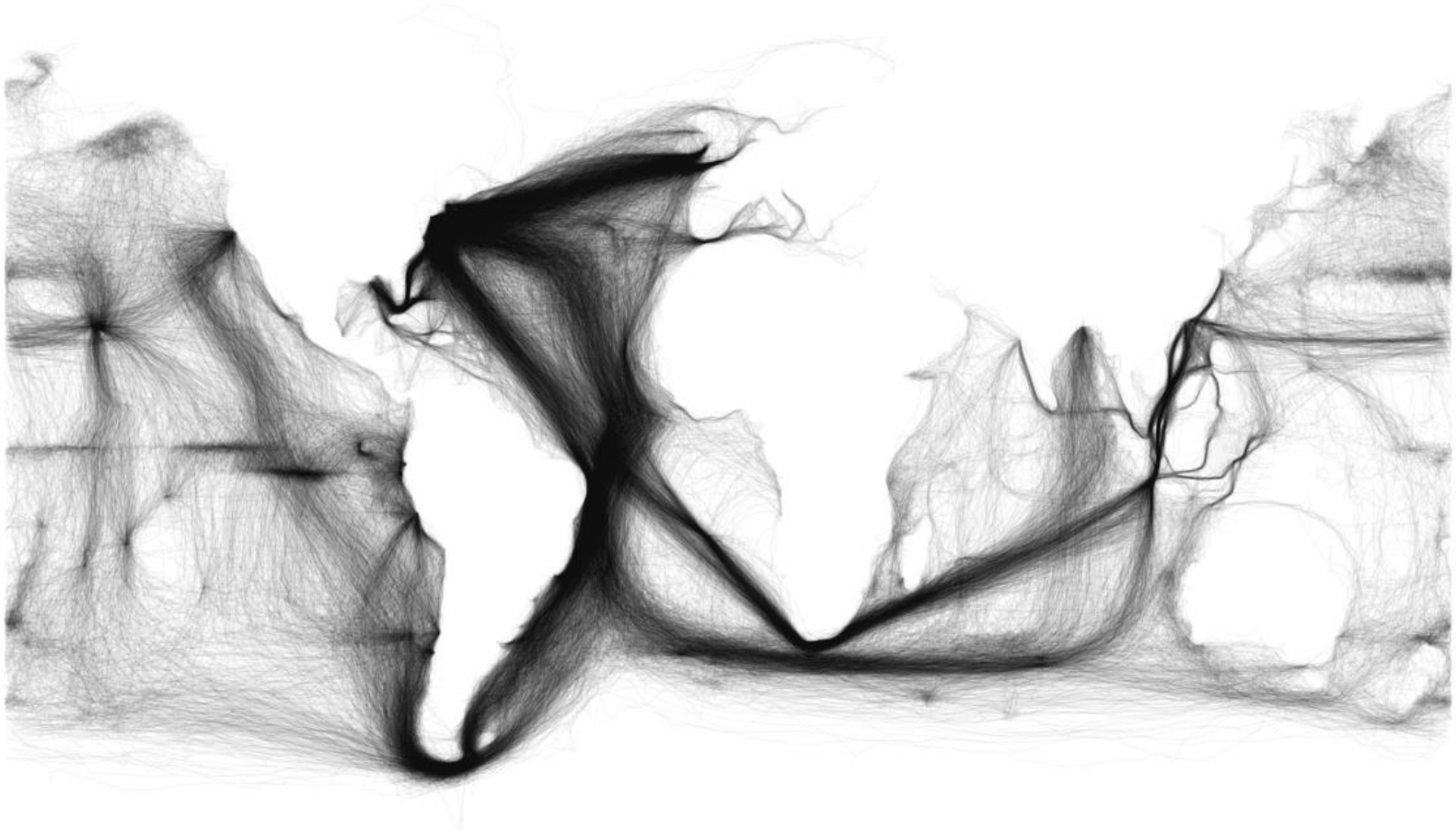
Great visualizations



Great visualizations

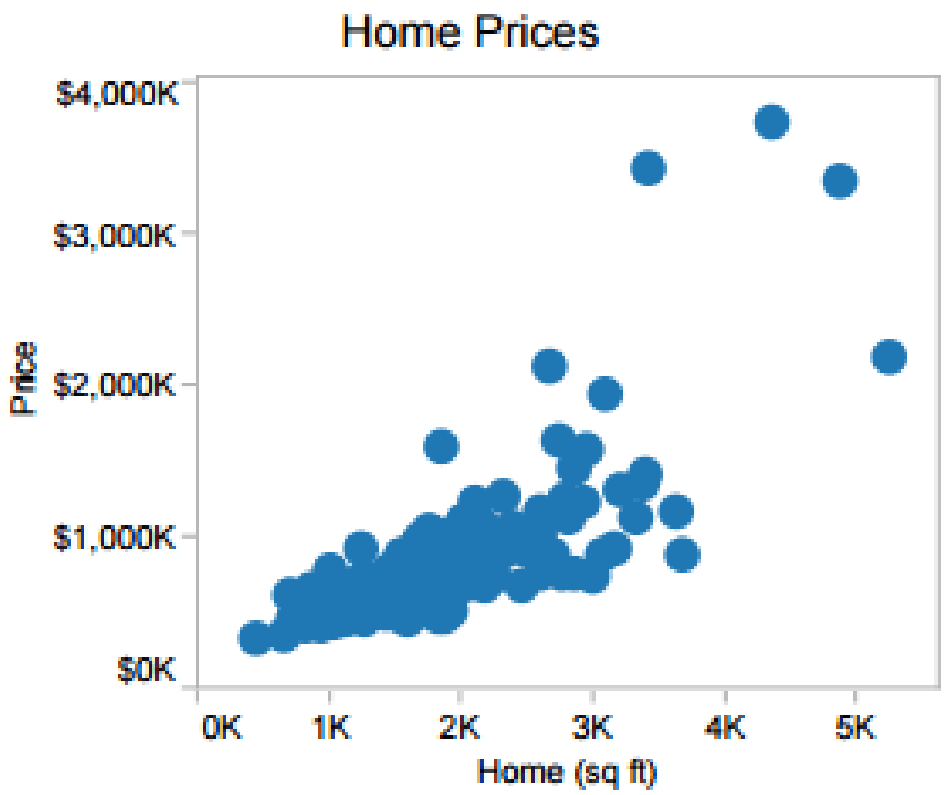


Great visualizations

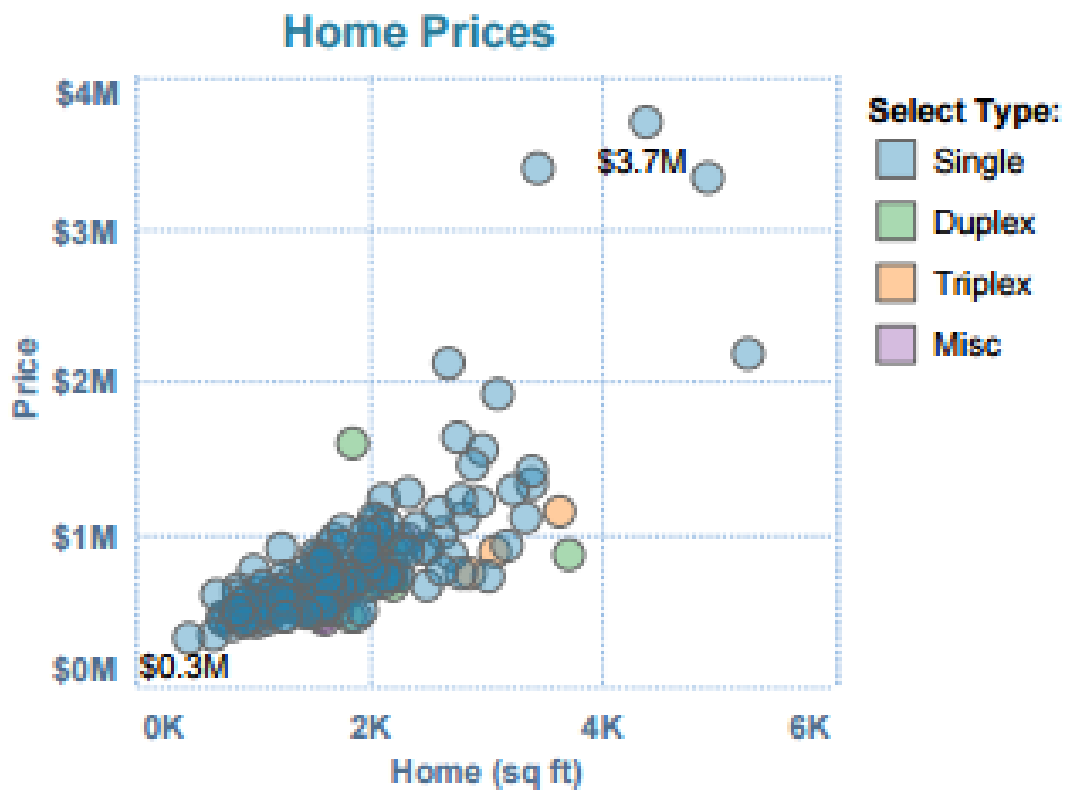


Great visualizations

Good visualization



Great visualization



Best Practices

Best Practices

Create visualization for the highest priority audience

Resist the temptation to create a dashboard that meets the needs of every single stakeholder

Provide Context: Always present performance measured against clear goals

Dashboard content has to be *snackable*

Best Practices

5 ‘W’ of visualization

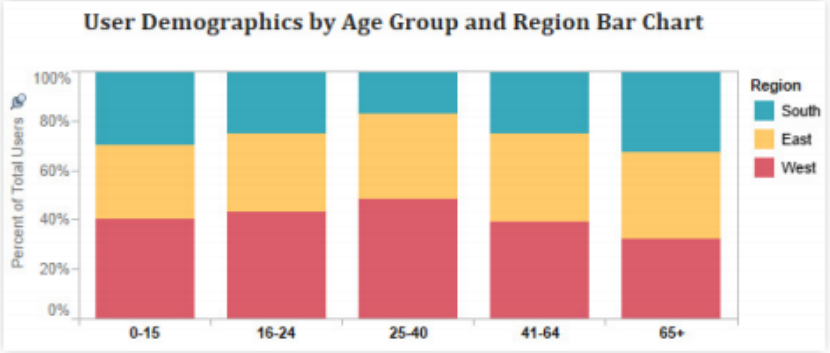
- *What is your audience*
- *What audience wants to know*
- *What answer do you have*
- *What are next set of questions from your answers*
- *What is the conversation resulting in*

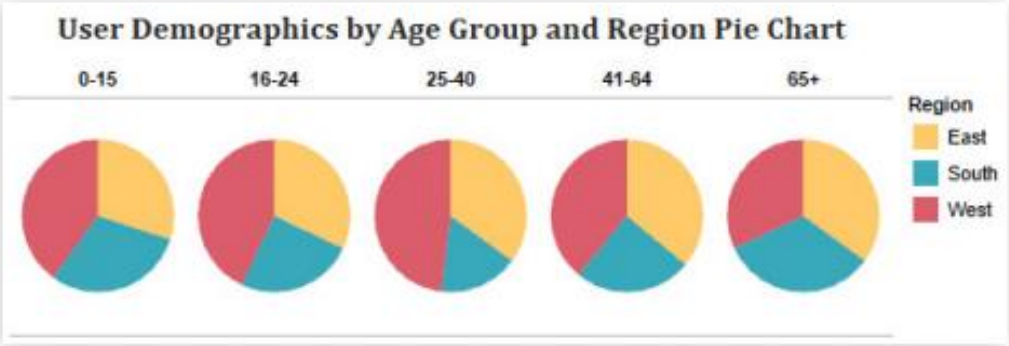
Best Practices

Use appropriate charts

Always avoid Pie charts

- Human visual perception not best suited for estimating areas
- Pie chart works best on maps with pie size indicated





Best Practices

In multiple variable visualizations

- Put most important data on X-Y axis
- Use other markers for secondary information

Don’t overload charts

- Break heavily condensed views in multiple small views

Limit number of colors and shapes in one view

- Put together there should not be more than 7-10 different perceptible items in a single view (Color plus shape)

Best Practices- Dashboard

Most important view in top or left corner

For connected views, the final view should be on right or in bottom

Ideal number of views in a dashboard – 3 to 4

Use highlighting and filters in dashboard for higher impact

Scrollbars are a bad idea – always

Best Practices- Design

Do not use more than 2 color palettes

Be sensitive to color blind people

Use non-overlapping palettes

It is difficult to quickly identify more than 12 different colors

Best Practices- Design

Preferable fonts

- Terbuchet MS or Verdana for tables and numbers
- Arial/Gerogia/Tahoma/Times New Roman/Lucida Sans

Avoid Calibri and Cambria

Avoid more than 2-3 different fonts in a single visualization

More modify more than 1 attribute of continuous fonts