Lesson 9

Visualization 2

Kush Kulshrestha

Why do we even need Viz?

There could be datasets that appear to be similar when using typical summary statistics, yet tell different stories when graphed.
Example, each dataset consists of eleven *(x,y)* pairs as follows:

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Why do we even need Viz?

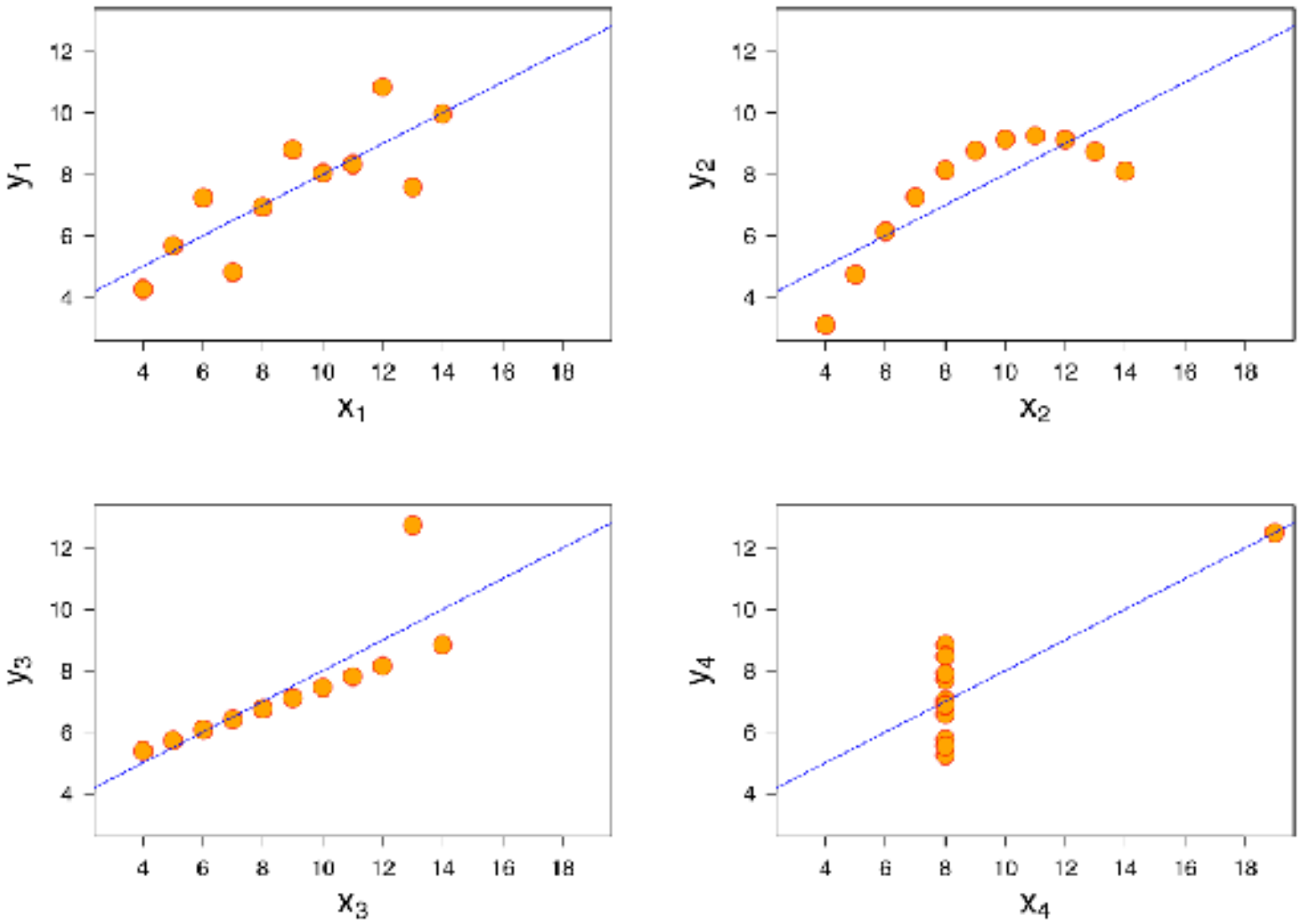
Observations:

- The average *x* value is 9 for each dataset.
- The average *y* value is 7.50 for each dataset
- The variance for *x* is 11 and the variance for *y* is 4.12
- The correlation between *x* and *y* is 0.816 for each dataset

So far these four datasets appear to be pretty similar. But when we plot these four data sets on an x/y coordinate plane, we get the following results:

Why do we even need Viz?

Now we see the real relationships in the datasets start to emerge

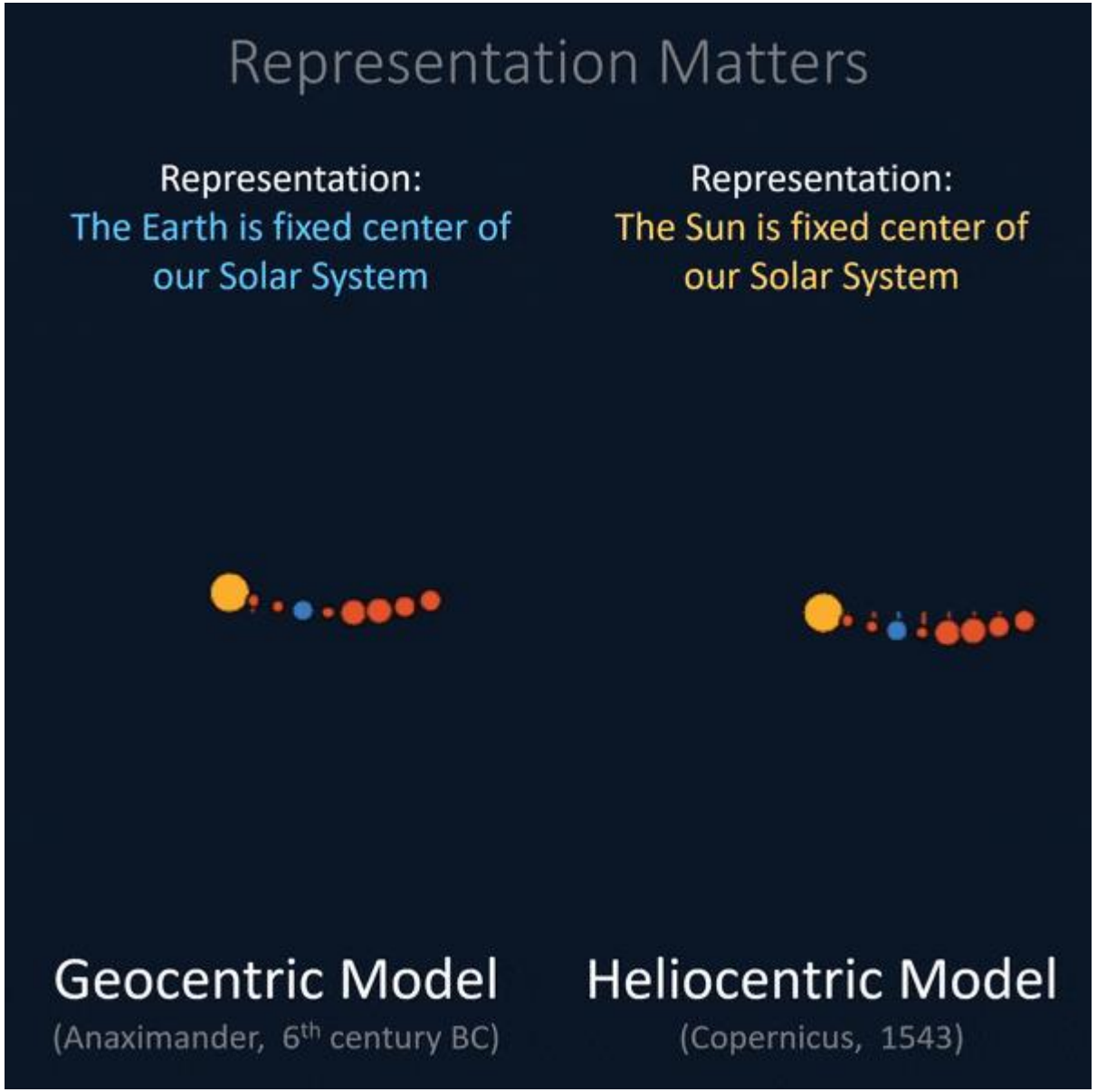


Why do we even need Viz?

- Dataset I consists of a set of points that appear to follow a rough linear relationship with some variance.
- Dataset II fits a neat curve but doesn't follow a linear relationship (maybe it's quadratic?).
- Dataset III looks like a tight linear relationship between *x* and *y*, except for one large outlier.
- Dataset IV looks like *x* remains constant, except for one outlier as well.

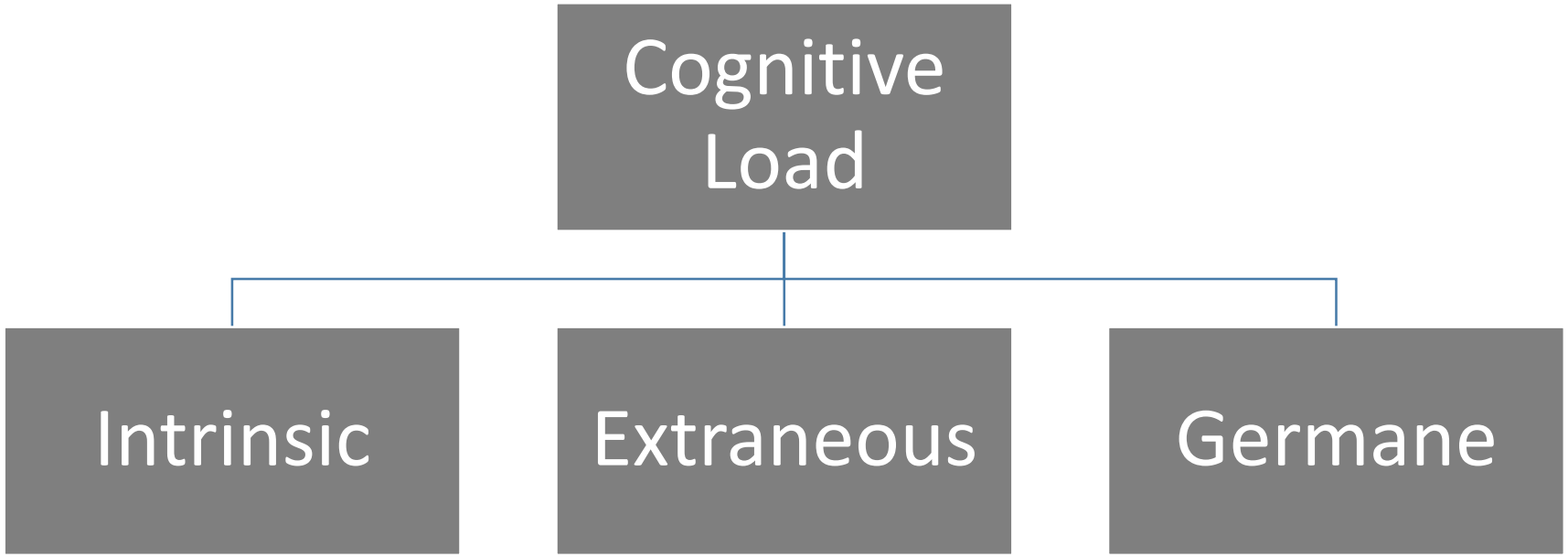
Computing summary statistics or staring at the data wouldn't have told us any of these stories. Instead, it's important to visualize the data to get a clear picture of what's going on.

Representation Matters



Cognitive Load and Clutter

Cognitive Load is the amount of mental effort required to interpret information.

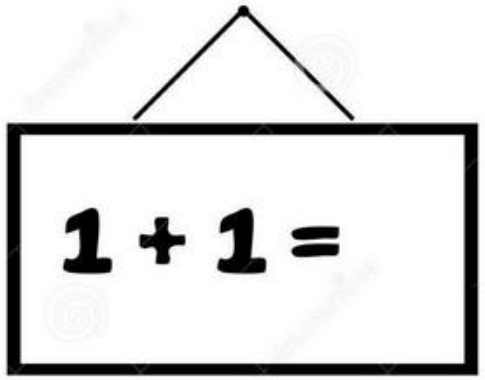


Cognitive Load and Clutter

Intrinsic Cognitive Load

Amount of memory we need to understand an idea.

Adding 2+2 is easier, doing difficult math division is difficult. If we don't concentrate in the latter, we end up spending more time.



$$\det\begin{pmatrix} 1 & x_1 & \dots & x_1^{n-1} \\ 1 & x_2 & \dots & x_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \dots & x_n^{n-1} \end{pmatrix} = \prod_{1 \le i < j \le n} (x_j - x_i)$$

Cognitive Load and Clutter

Extraneous Cognitive Load

- It relates to how information is presented.
- It is the amount of extra brain power you need to understand a poorly designed Viz.
- Poor design requires more effort to identify problems and create a mental image.

Cognitive Load and Clutter

Germane Cognitive Load

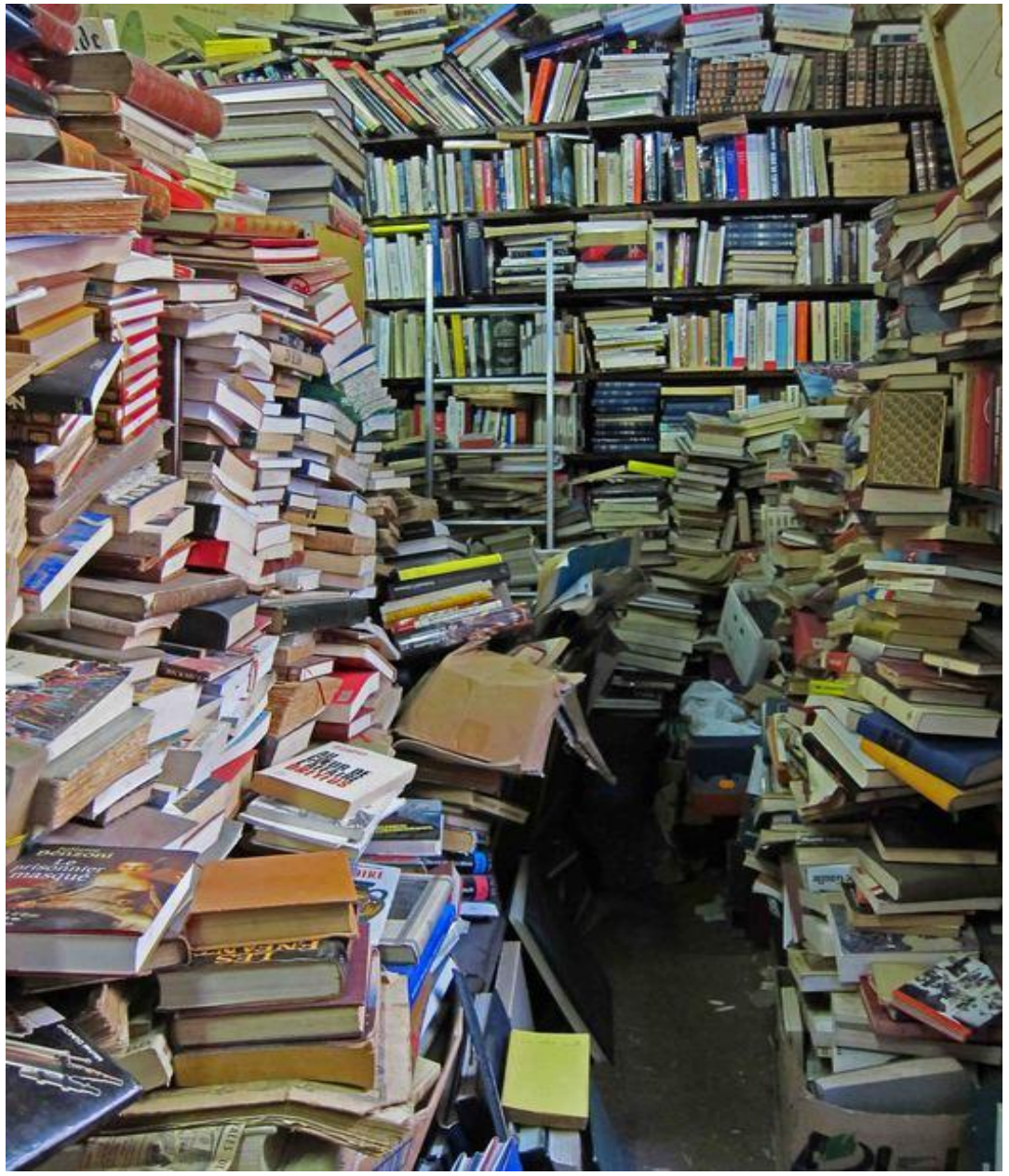
It is a way for the brain to look for the patterns to develop context.

Cognitive Load and Clutter

Clutter is all the things you remove while still preserving key ideas.

Reduce clutter to minimize user's cognitive load

Less clutter = More effective Visualization



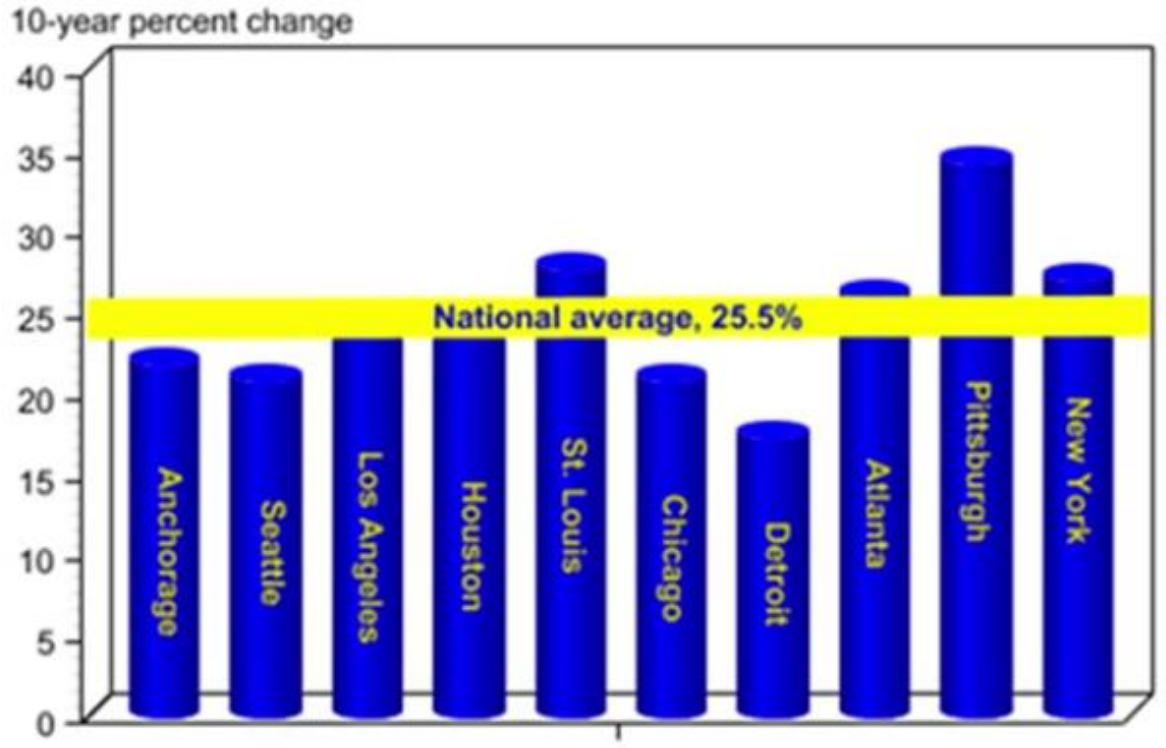
Cognitive Load and Clutter

Report of USDA's (US Department of Agriculture) Economic Research Department

This is a regular report put out regularly by renowned statisticians.

Retail food price inflation by Metropolitan Statistical Area (MSA), 2006-15

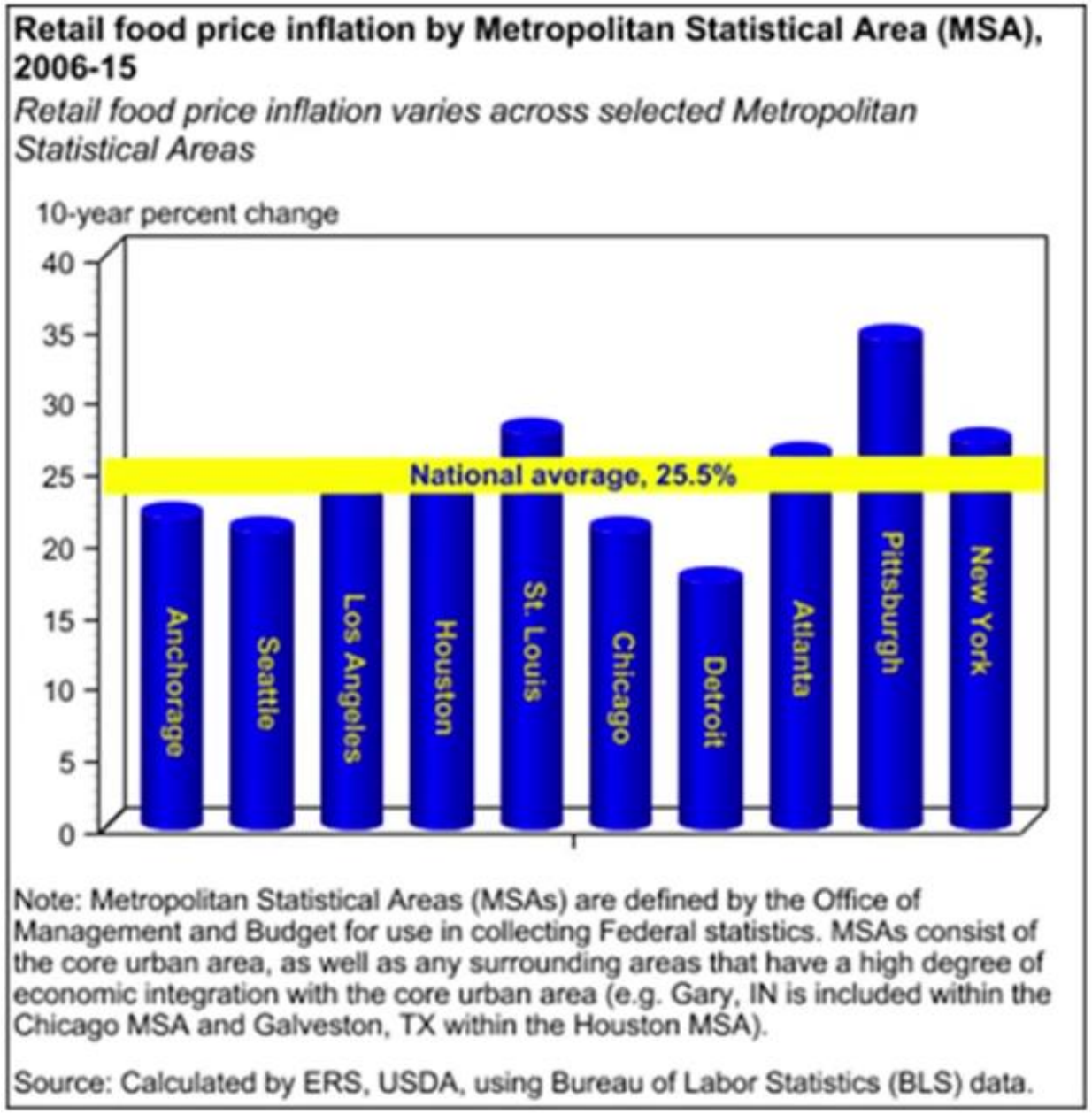
Retail food price inflation varies across selected Metropolitan Statistical Areas



Note: Metropolitan Statistical Areas (MSAs) are defined by the Office of Management and Budget for use in collecting Federal statistics. MSAs consist of the core urban area, as well as any surrounding areas that have a high degree of economic integration with the core urban area (e.g. Gary, IN is included within the Chicago MSA and Galveston, TX within the Houston MSA).

Source: Calculated by ERS, USDA, using Bureau of Labor Statistics (BLS) data.

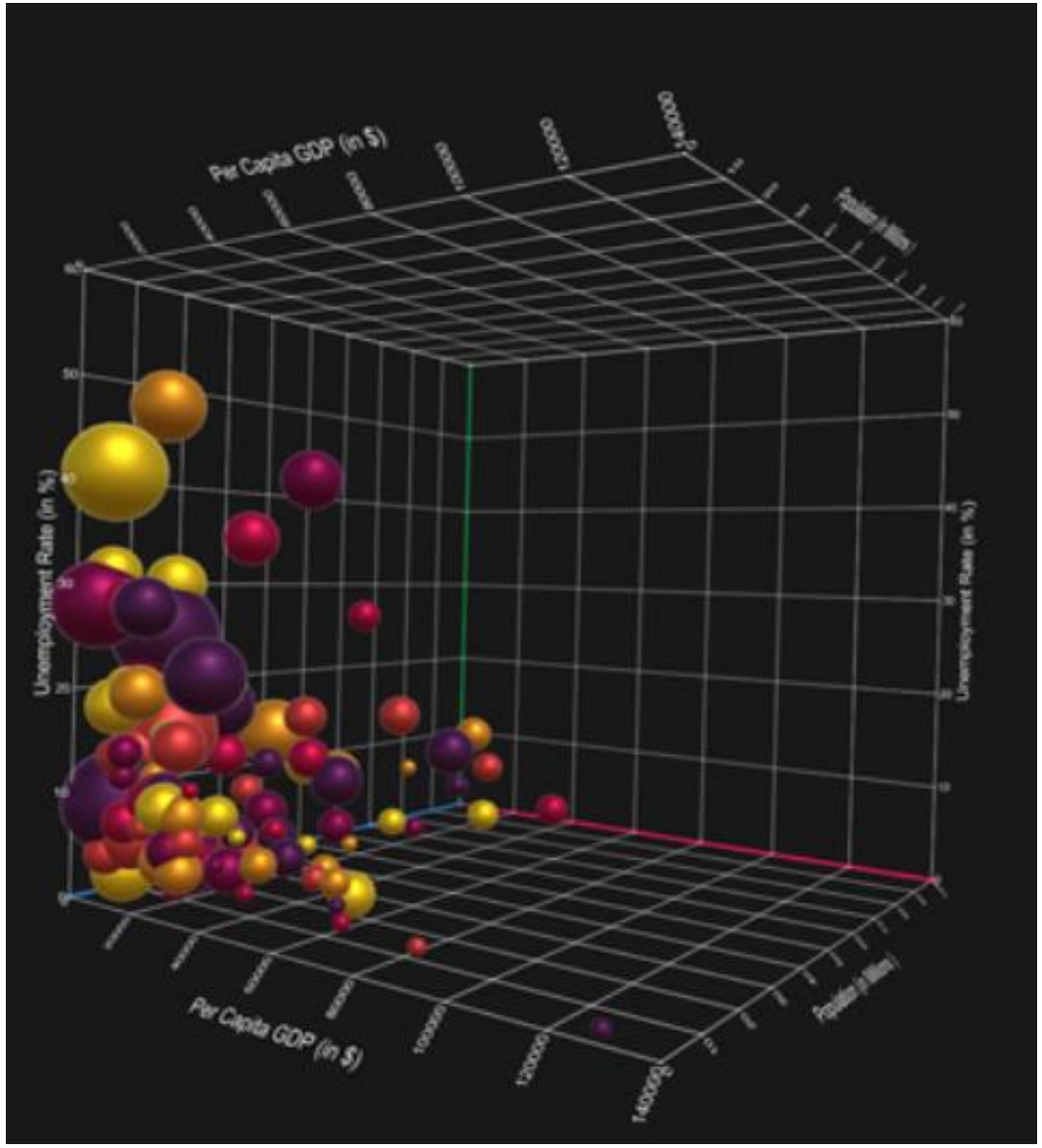
Cognitive Load and Clutter



Things to be removed:

- 3 D effect
- Overuse of bright colors. They are drawing attention to no specific place.
- Yellow ticker as a threshold!
- No sorting of the data?
- What is x-axis for? Unhelpful axis

Cognitive Load and Clutter



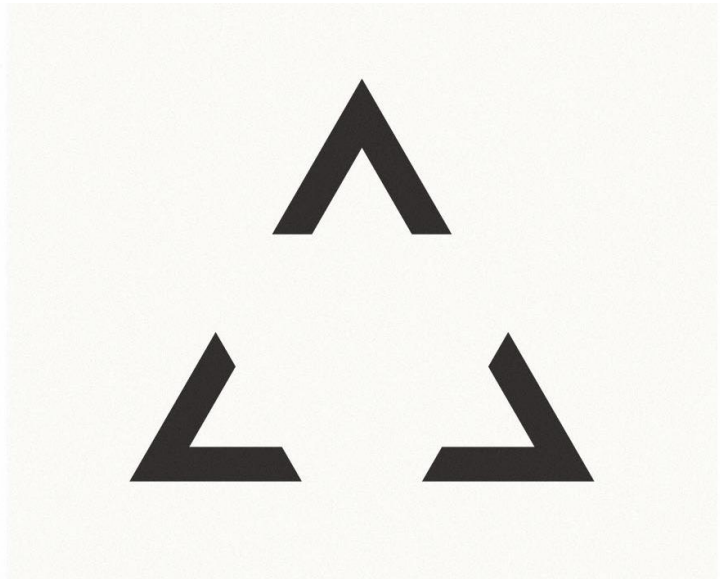
Why not to use 3-D?

- 3-D doesn't improve a Visualization
- Skews information (remember Steve Jobs?)
- Adds confusion

Cognitive Load and Clutter

Some other tips:

- Use currency sign with the number
- Always use percent sign when drawing percentages.
- Commas with number, whenever number is bigger than 9999
- Use consistent scientific notation if numbers are too big.



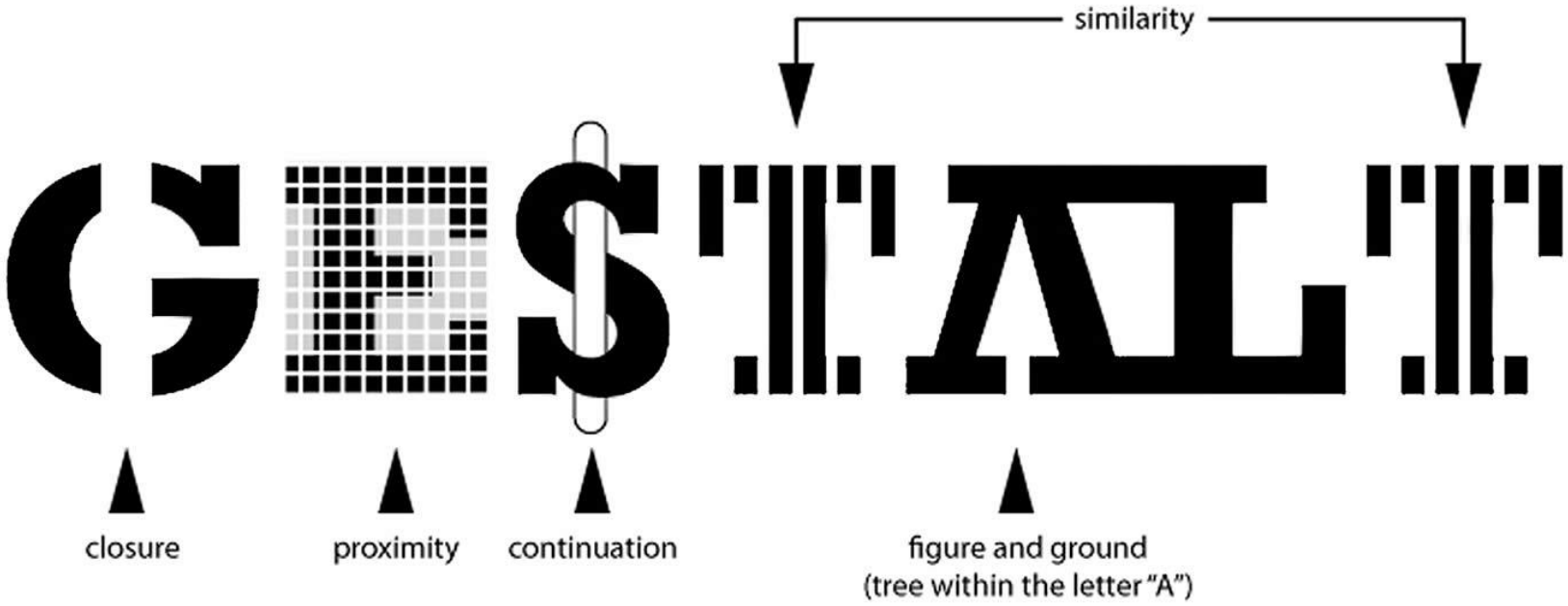
Gestalt Principles

Gestalt Principles

Gestalt refers to the patterns that you perceive when presented with a few graphical elements or why we perceive things the way we do.

These rules are important when it comes to creating data visualizations to ensure efficient, clean, informative and correct information transfer.

Understanding gestalt is useful for improving traditional charts like bar charts and line charts.



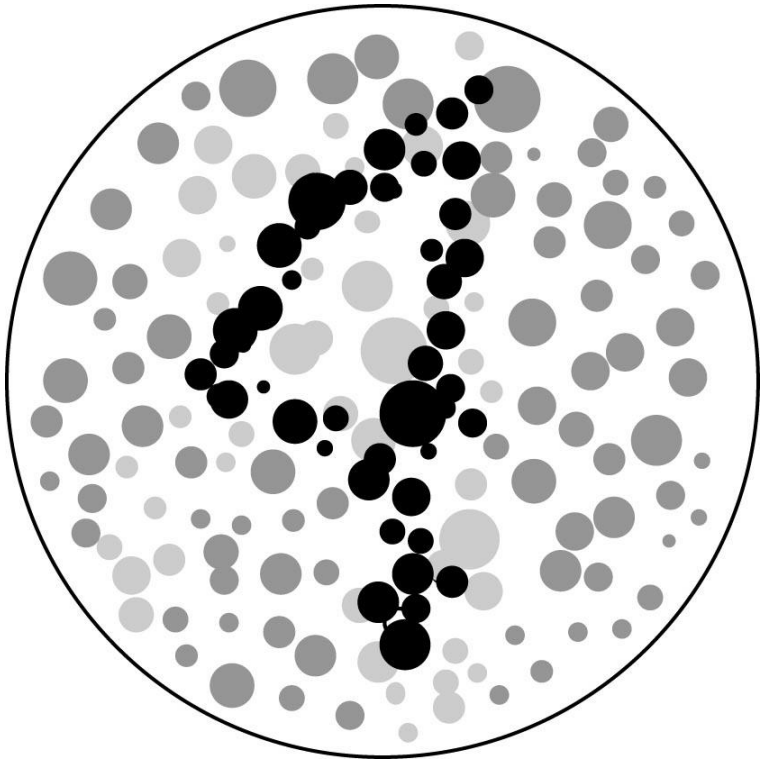
Gestalt Principles - Similarity

Items that are similar are grouped together.

Visual elements similar in shape, size, colour, proximity, motion, and direction are perceived as part of a group.

Example:

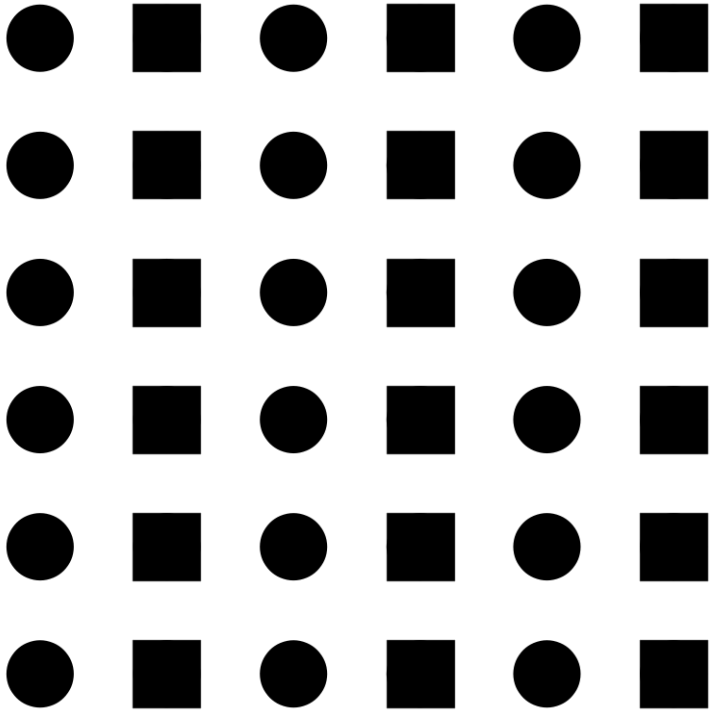
We seek patterns, easily grouping the black dots into a number 4



Gestalt Principles - Similarity

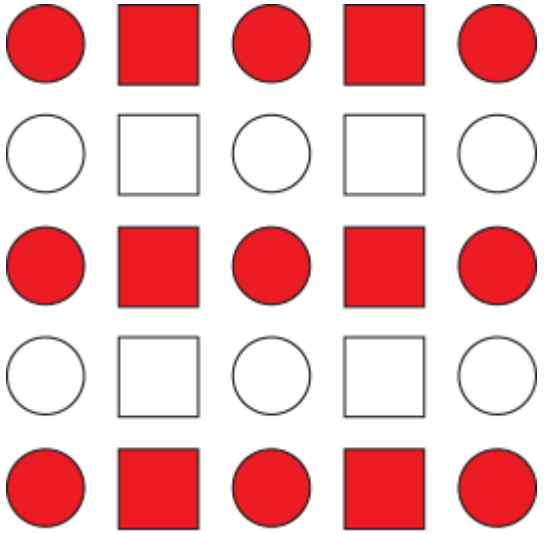
Example:

We perceive vertical groups of squares and circles.



Example:

We perceive horizontal groups of same red/white color elements.



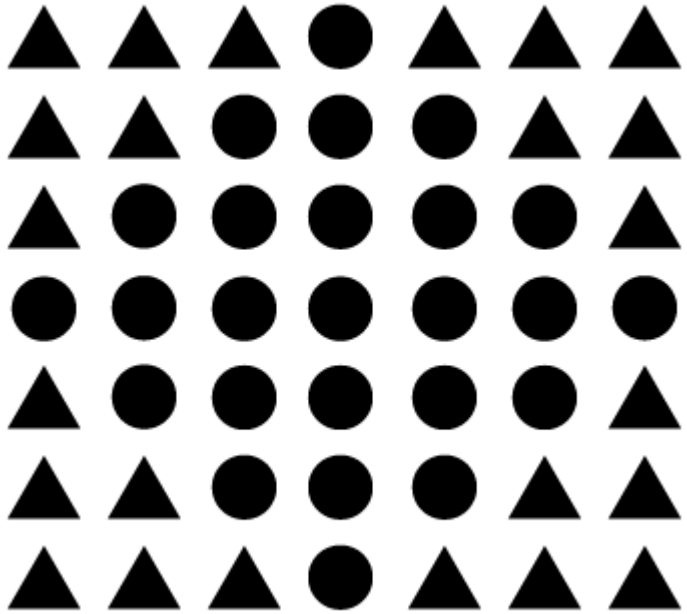
Gestalt Principles - Similarity

Example:

We perceive green/grey colored squares as a group.

Example:

A complex example of groups.

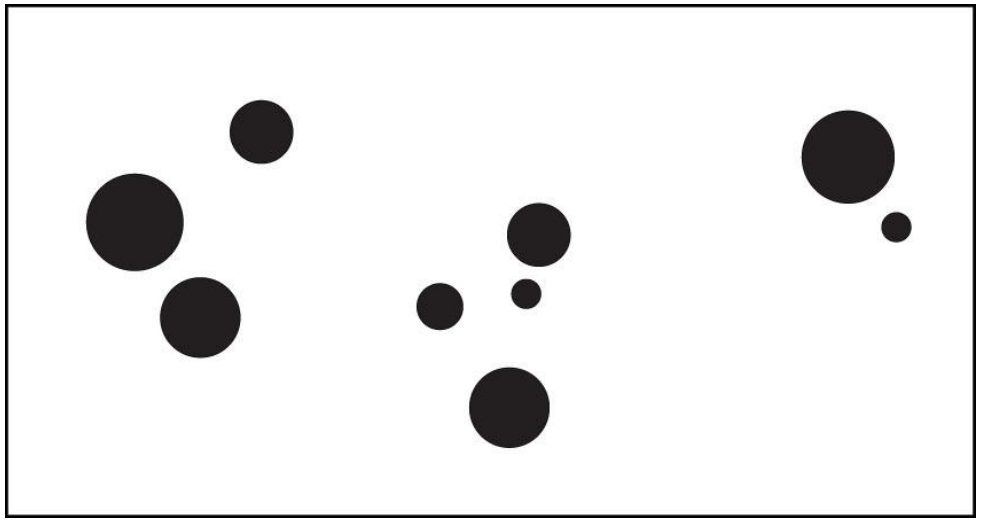


Gestalt Principles - Proximity

Objects that are close (spatially) are grouped together. The closer items are the more likely the perception they are a group.

Example:

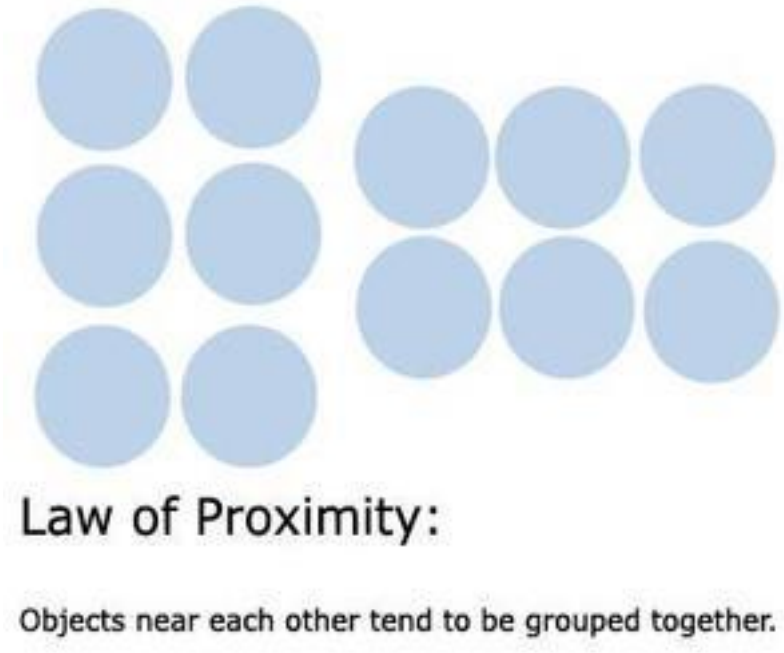
Instead of counting numerous distinct circles we see three “groups”.



Gestalt Principles - Proximity

Example:

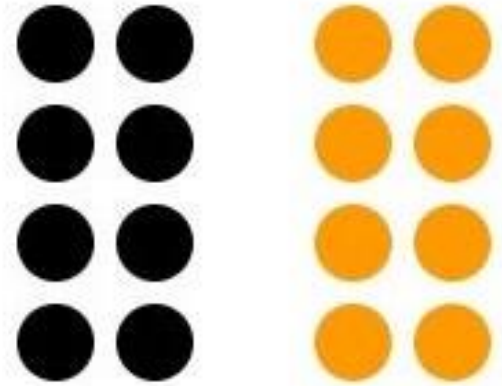
The circles on the left appear to be grouped in vertical columns, while those on the right appear to be grouped in the horizontal rows.



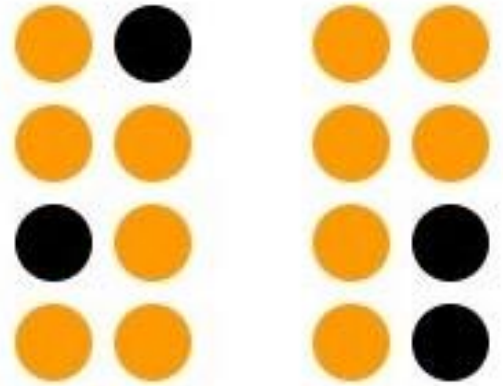
Gestalt Principles - Proximity

Example:

These groups appear to be separated by color or contrast.



Proximity overpowers other signals of distinction

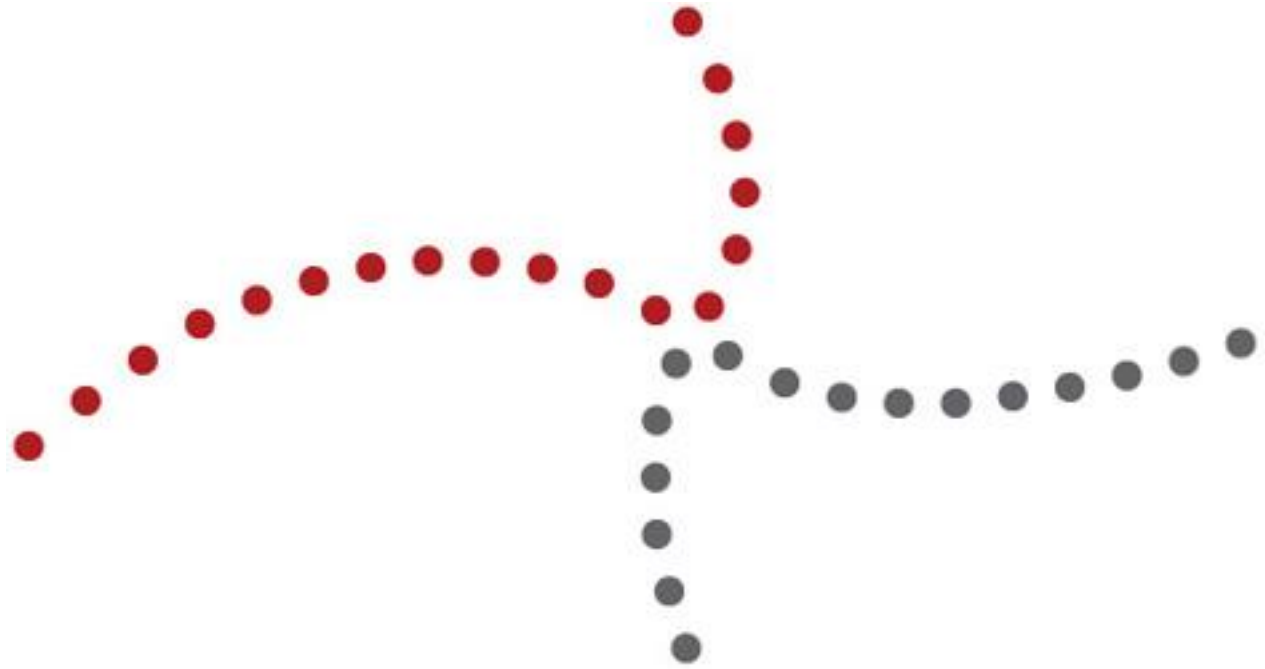


Gestalt Principles - Continuity

Lines are seen as following smoothest path.

Example:

The eye tends to want to follow the straight line from one end of this figure to the other, and the curved line from the top to the bottom, even when the lines change colour midway through



Gestalt Principles - Continuity

We perceive objects as belonging together, as part of a single whole, if they are aligned with one another or appear to form a continuation of one another.

Example:

We tend to see the individual lines as a continuation of one another, more as a dashed line than separate lines.

Gestalt Principles - Continuity

Things that are aligned with one another appear to belong to the same group. In the table in Figure 4-17, it is obvious which items are division names and which are department names, based on their distinct alignment. Divisions, departments, and headcounts are clearly grouped, without any need for vertical grid lines to delineate them. Even though the division and department columns overlap with no white space in between, their distinct alignment alone makes them easy to distinguish.

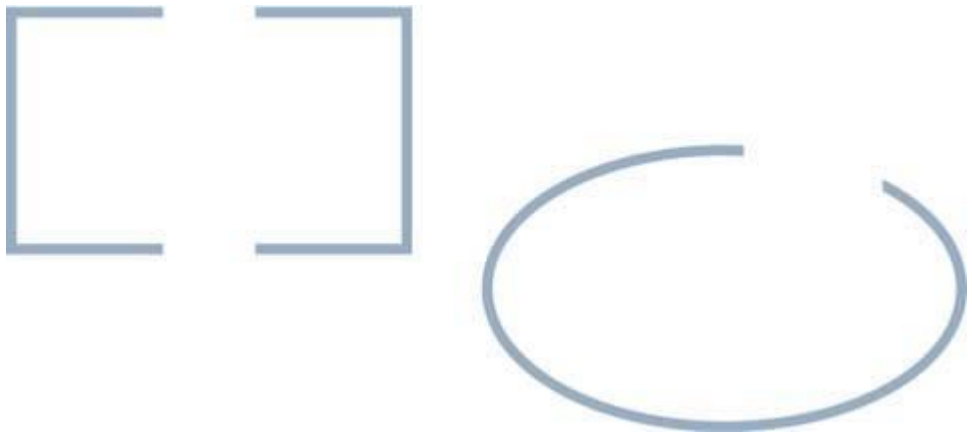
Division/Department	Headcount
G&A	
Finance	15
Purchasing	5
Information Systems	17
Sales	
Field Sales	47
Sales Operations	10
Engineering	
Product Development	22
Product Marketing	5

Gestalt Principles - Closure

Humans have a keen dislike for loose ends. The principle of closure asserts that we perceive open structures as closed, complete, and regular whenever there is a way that we can reasonably do so.

Example:

The Gestalt principle of closure explains why we see these as closed shapes, despite the fact that they are not finished.

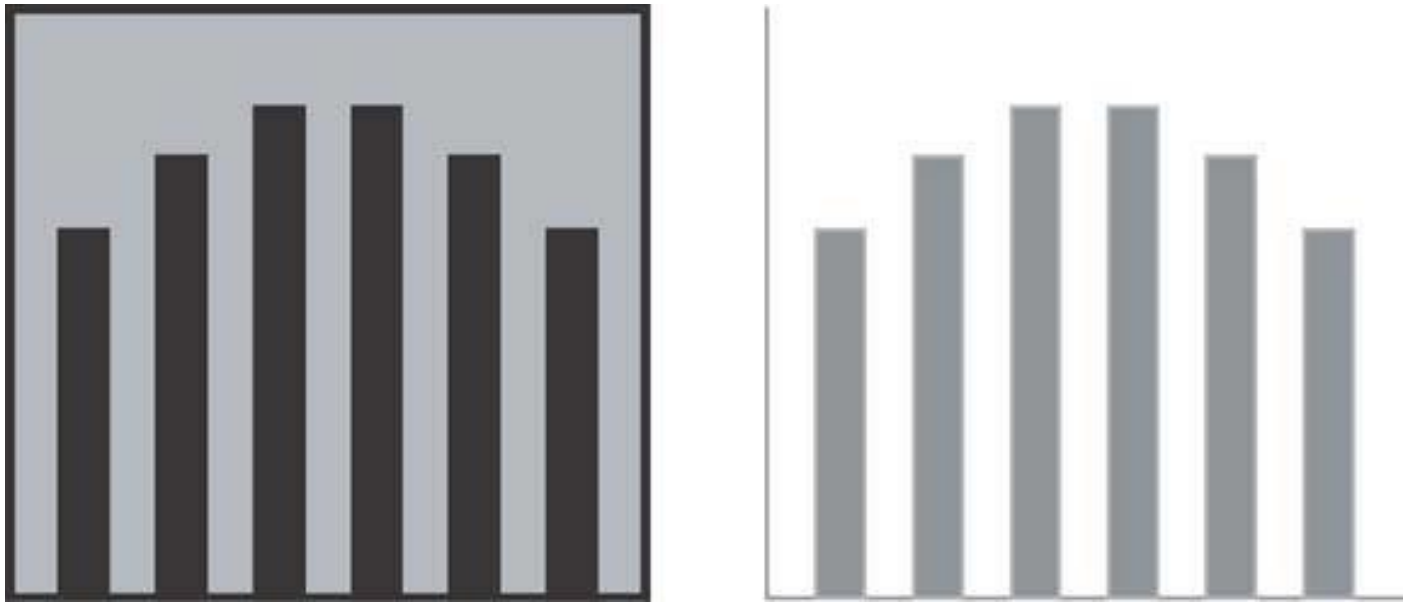


Gestalt Principles - Closure

Example:

We can apply this tendency to perceive whole structures in dashboards, especially in the design of graphs. For example, we can group objects (points, lines, or bars in a graph, etc.) into visual regions without the use of complete borders or background colours to define the space.

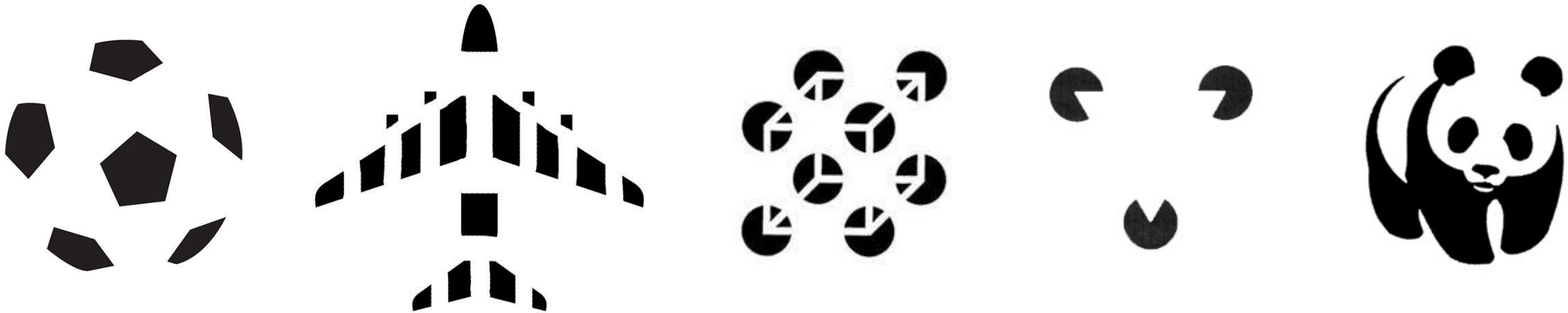
This is good and required so that we can reduce the clutter.



The Gestalt principle of closure also explains why only two axes, rather than full enclosure, are required on a graph to define the space in which the data appears.

Gestalt Principles - Closure

More Example:

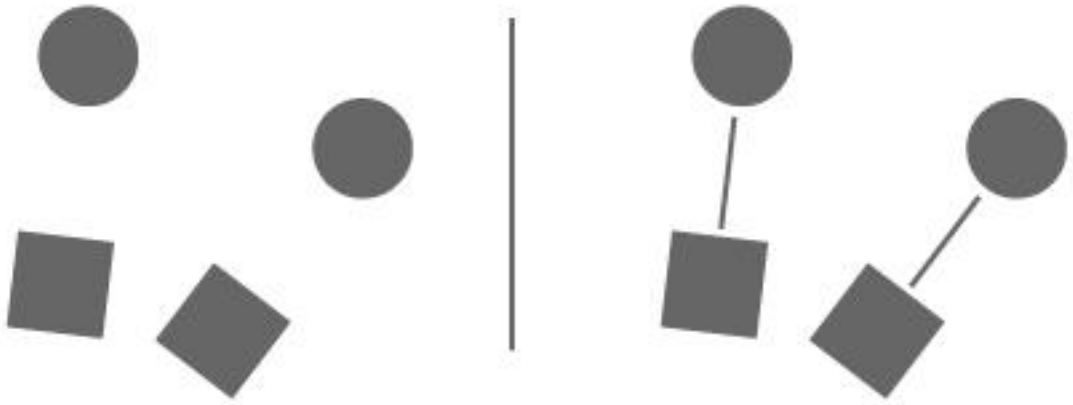


Gestalt Principles - Connectedness

Elements that are visually connected are perceived as more related than the elements with no connection.

Example:

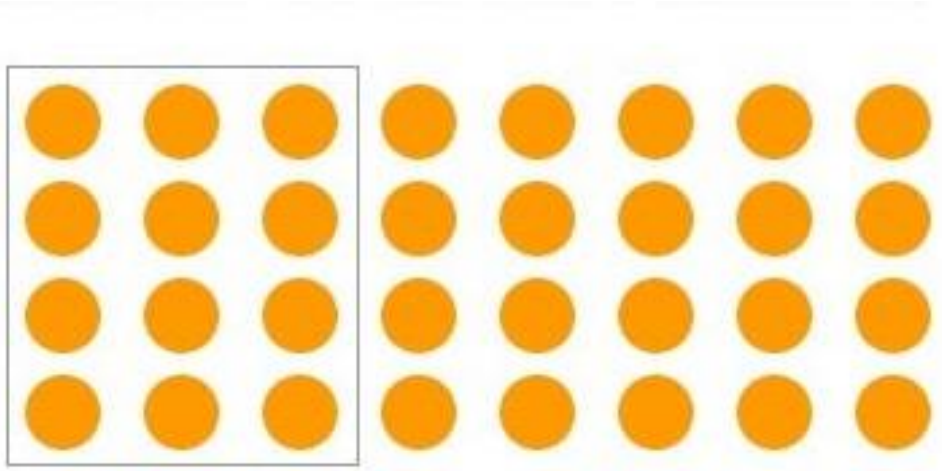
The Gestalt principle of closure explains why we see these as closed shapes, despite the fact that they are not finished.



Gestalt Principles - Connectedness

Example:

Even though the spacing and the color is consistent within this collection of elements, those inside of the connecting lines are perceived to be more related than the rest.

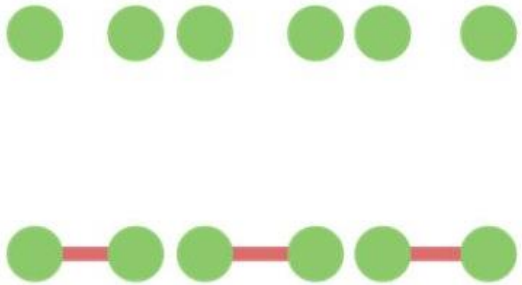


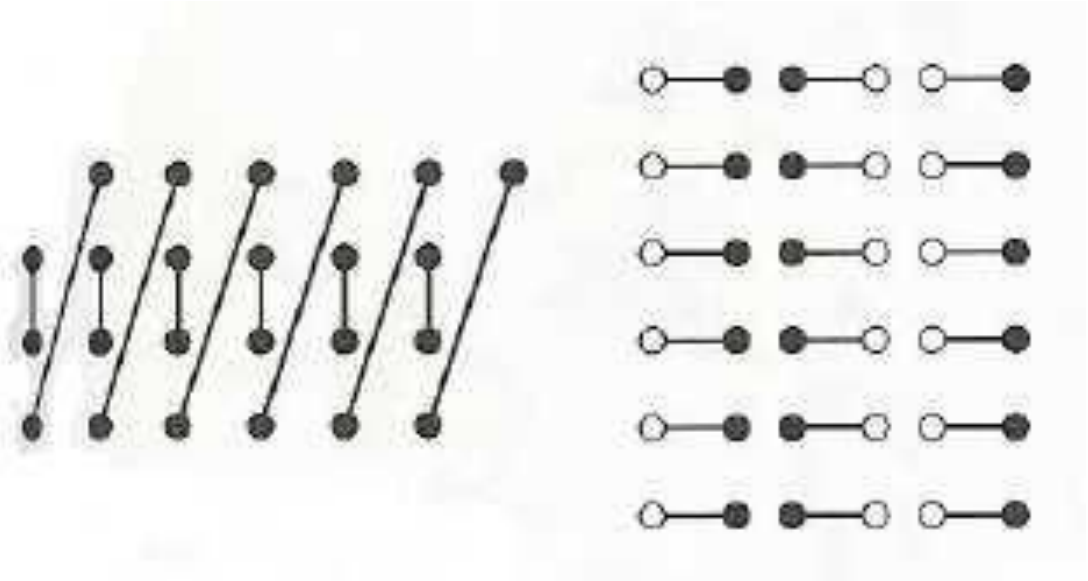
As are the ones connected by the lines..

Gestalt Principles - Connectedness

Example:

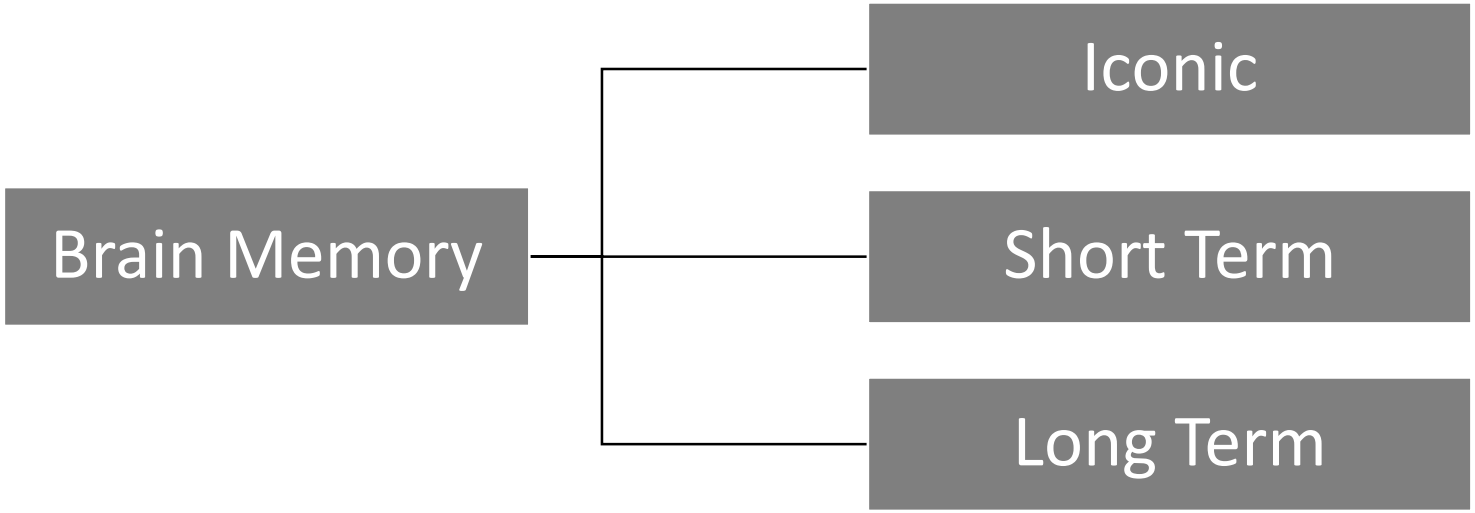
And it is more powerful than proximity and similarity.





Pre-Attentive Attributes of a Visualization

Three types of memories we can influence:



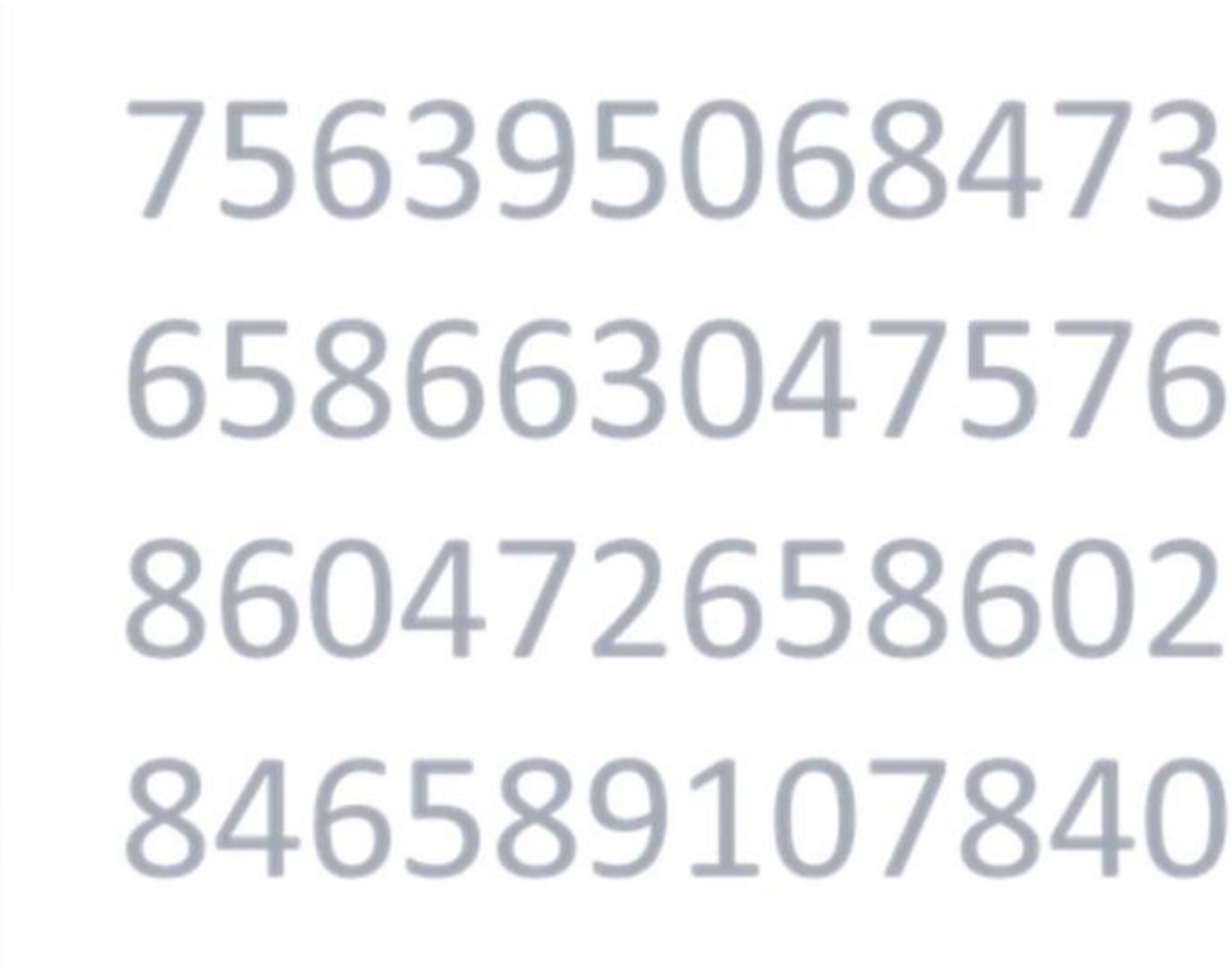
We want to get something in readers brain without them even to think about it.

Why?

This decreases the cognitive load while conveying the information so that there is not much work to be done by the brain to understand.

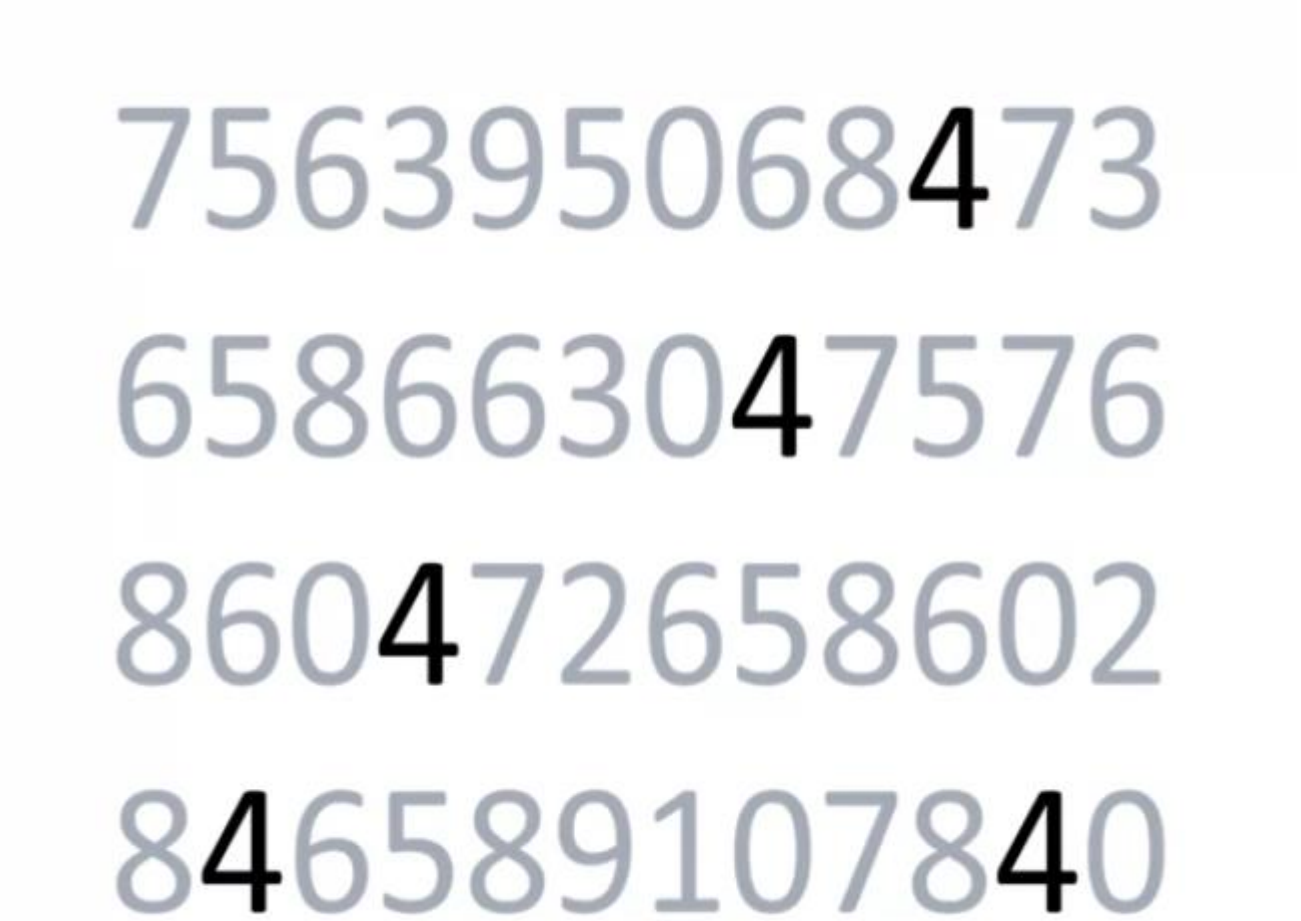
Pre-Attentive Attributes of a Visualization

Example: Count the total number of 4's in the figure.



Pre-Attentive Attributes of a Visualization

Example: Now count the total number of 4's in the figure. See the difference?



Pre-Attentive Attributes of a Visualization

Good visualization allow users to see what we want them to see before they know that they have seen it.

Pre-Attentive Attributes of a Visualization

So, how do I do all that?

Pre-Attentive Attributes of a Visualization

Change one of these to **focus** on users attention:

- Si**Z**e
- Color
- Orientation
- $h@pe
- Line com_{position}
- Enclosure
- Intensity
- Position

Pre-Attentive Attributes of a Visualization