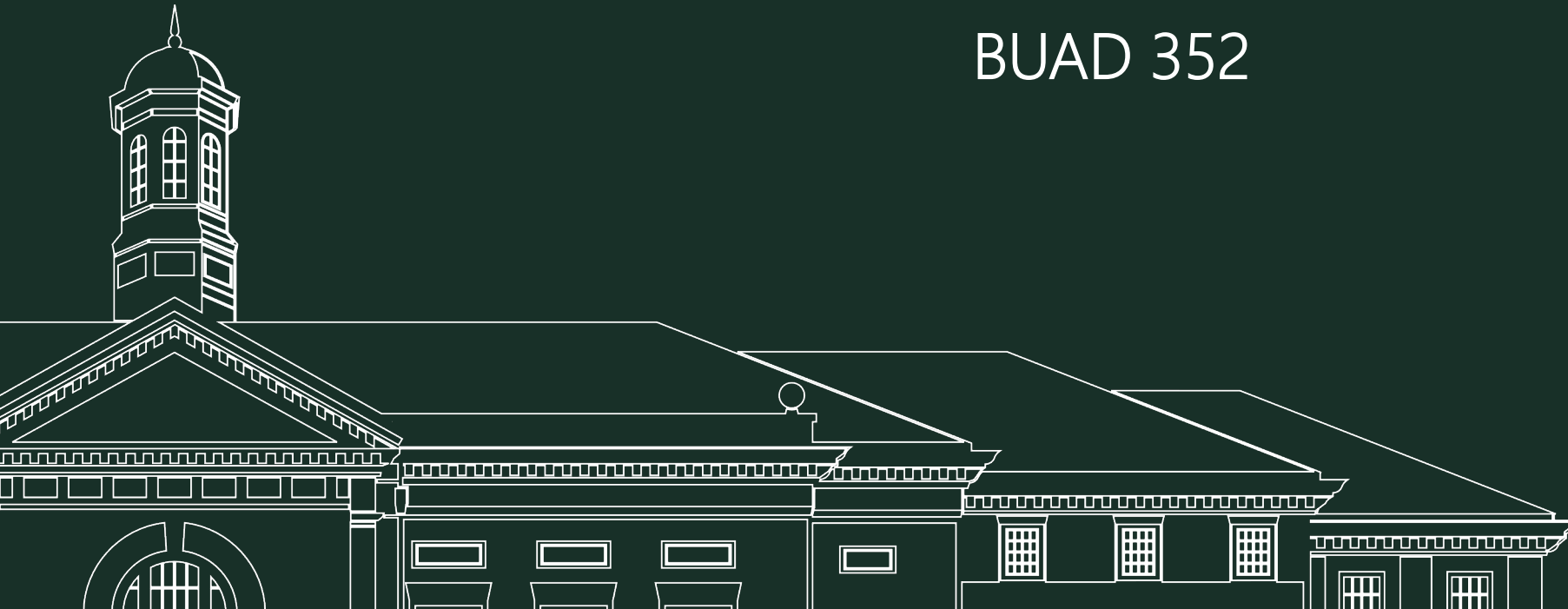




Raymond A. Mason
School of Business
WILLIAM & MARY

Data Visualization and Modeling

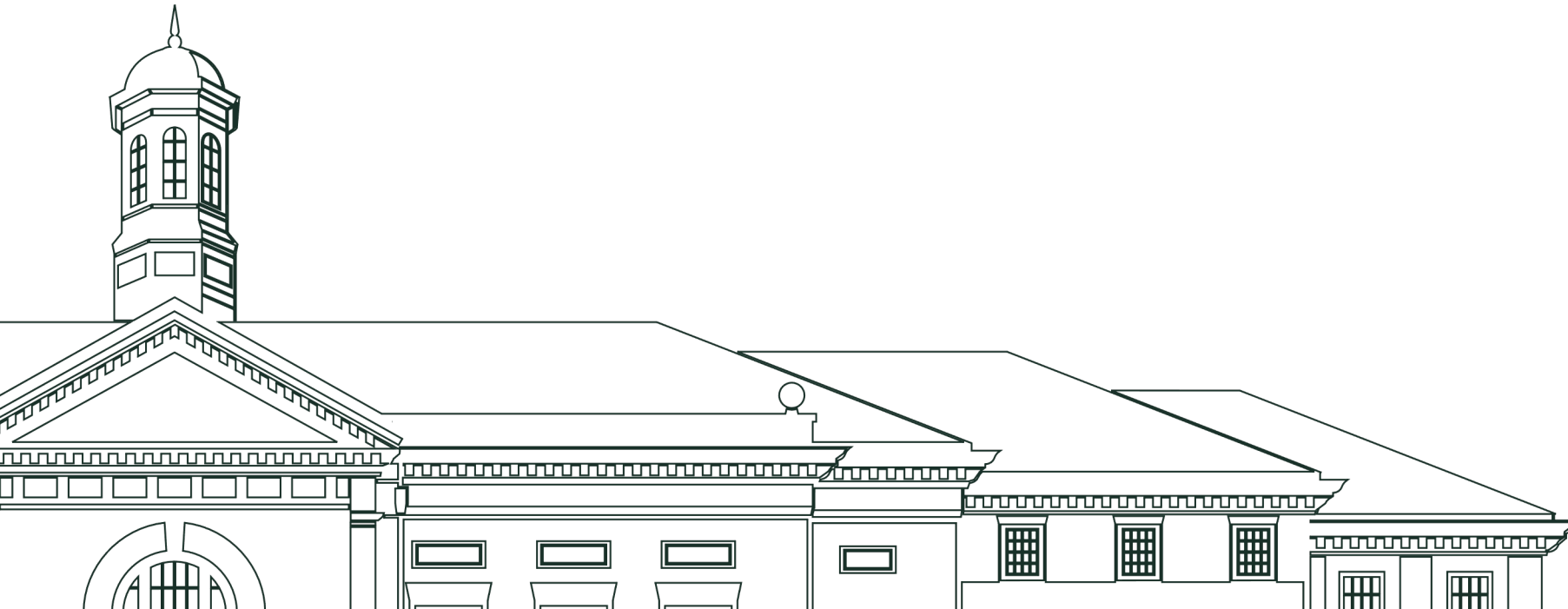
BUAD 352





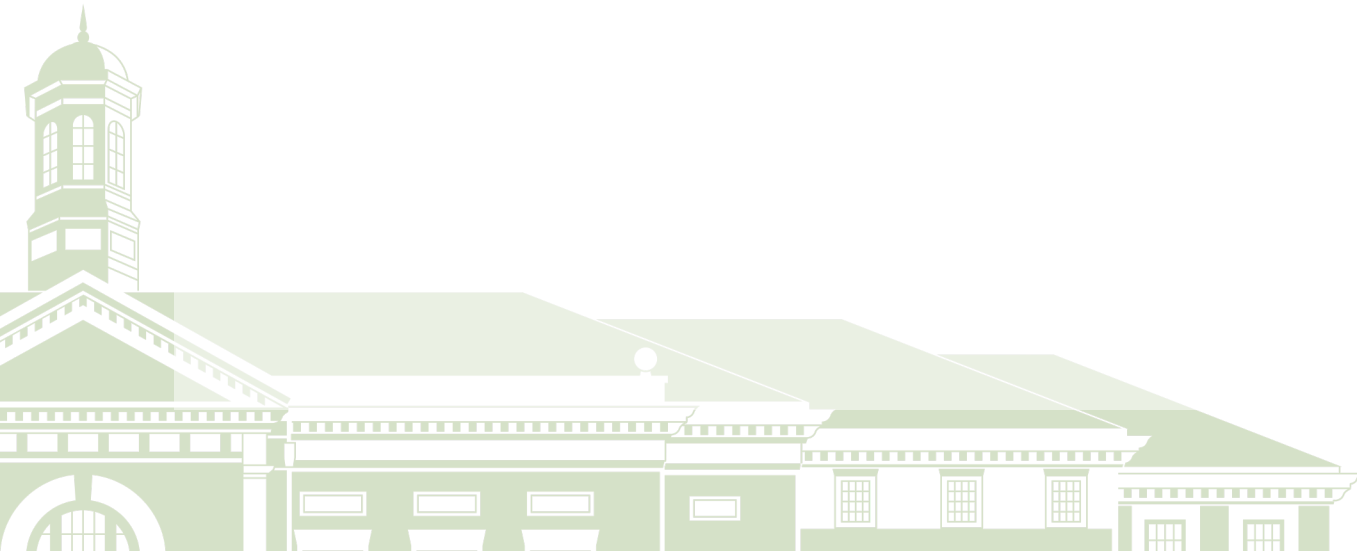
Raymond A. Mason
School of Business
WILLIAM & MARY

Data Visualization



Data Visualization

Data Visualization: The process of converting **raw data** into easily understood pictures of information that enable **fast** and **effective decisions**.



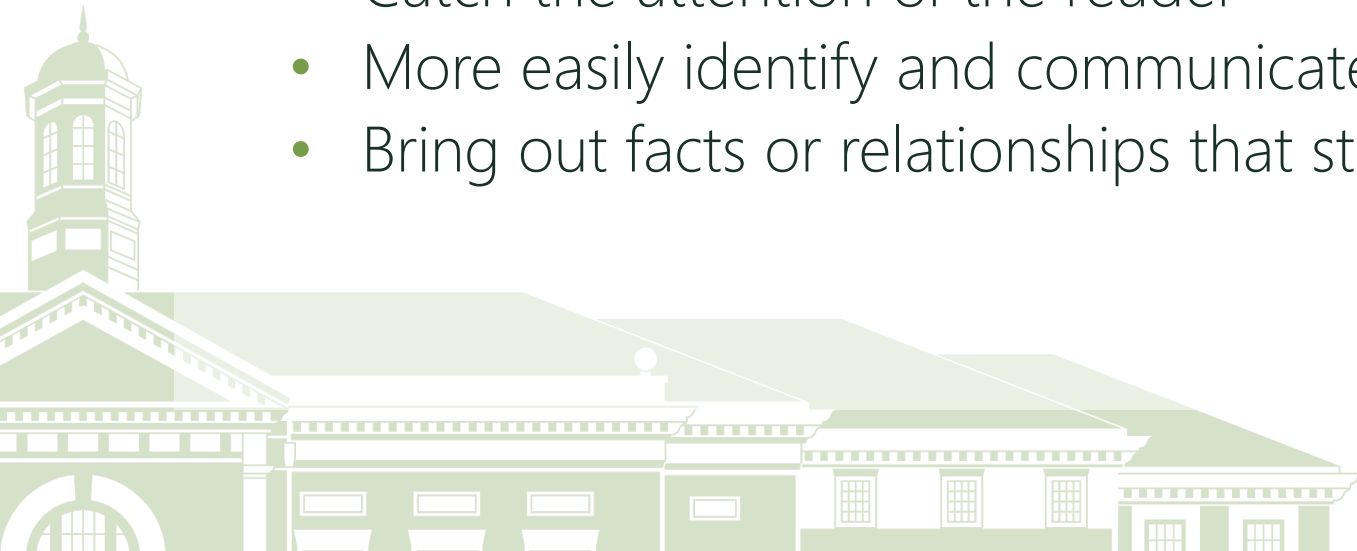
Why Graphs?

Tables have several advantages over **graphs**:

- Present more information than a graph
- Display the exact values
- Require less work to create

Graphs also have advantages over **tables**:

- Catch the attention of the reader
- More easily identify and communicate patterns
- Bring out facts or relationships that stimulate thinking



Why Graphs?

Four Series have the same:

Average of $x = 9$

Average of $y = 7.5$

Sample variance of $x = 11$

Sample variance of $y = 4.12$

Correlation between x and $y = 0.816$

OLS line is $y = 3 + 0.5x$

Are these series the same?

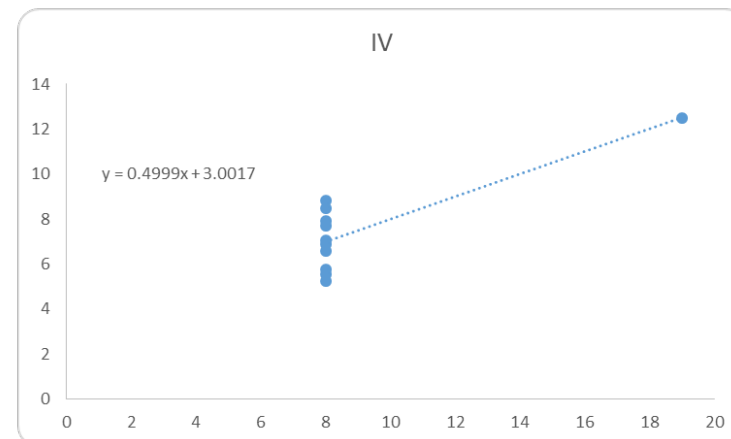
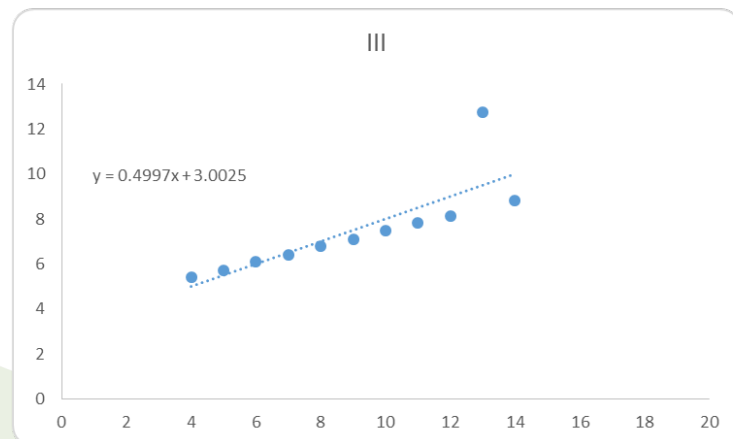
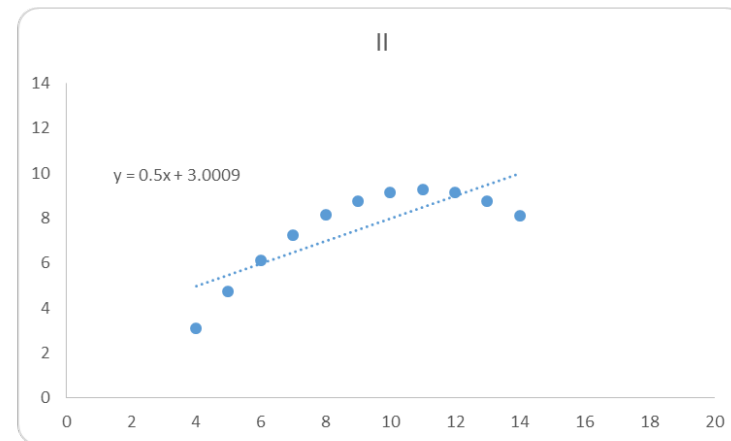
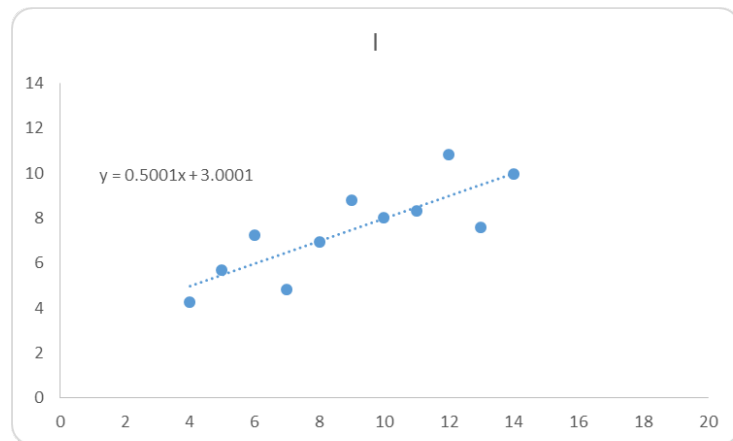
I		II		III		IV	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Why Graphs?

Patterns help us model the data.

Understand the

process that governs the data.



Citation: Anscombe, F.J. (1973).
"Graphs in Statistical Analysis".
American Statistician 27 (1): 17–
21

Why Graphs?

The **first step** to understanding your data should always be to **examine it visually**.

Visualization can play a critical role in helping you figure out what is **interesting** about your data.

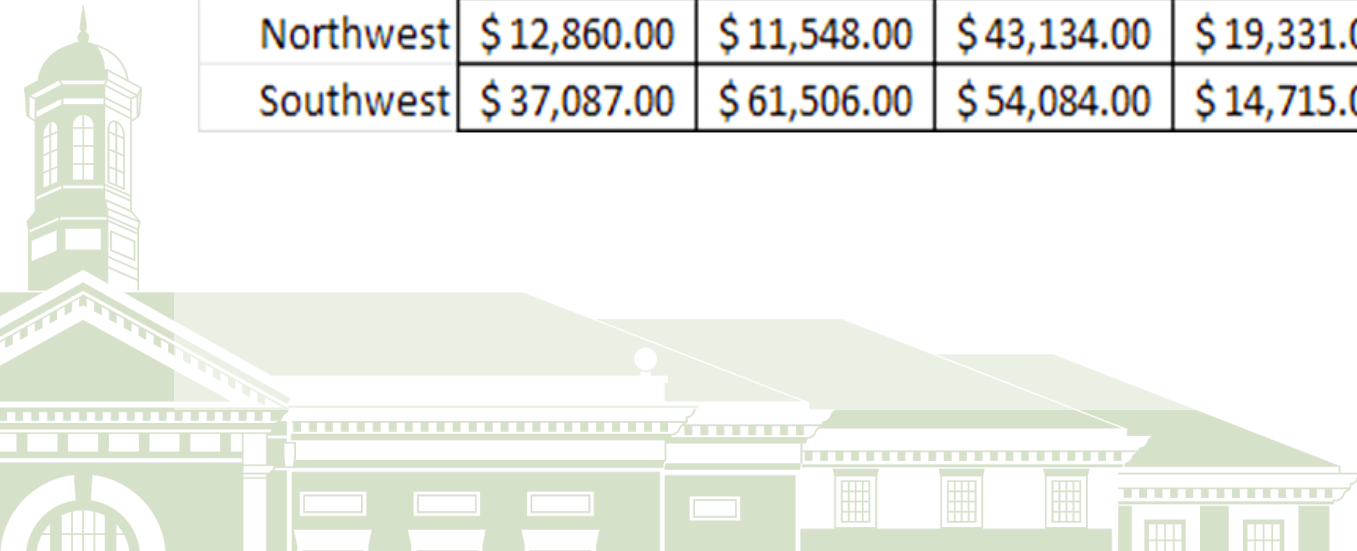
Pictures are retained at much higher rates than word (picture superiority effect).



The Value of Highlighting

Which are the 5 highest revenue product-region combinations last month?

	Product A	Product B	Product C	Product D	Product E	Product F	Product G	Product H
Northeast	\$ 15,749.00	\$ 40,195.00	\$ 15,472.00	\$ 63,029.00	\$ 8,509.00	\$ 42,987.00	\$ 27,778.00	\$ 12,510.00
Southeast	\$ 48,044.00	\$ 20,741.00	\$ 40,643.00	\$ 15,687.00	\$ 12,342.00	\$ 23,297.00	\$ 10,401.00	\$ 10,522.00
Central	\$ 5,240.00	\$ 45,296.00	\$ 16,114.00	\$ 63,359.00	\$ 58,198.00	\$ 24,191.00	\$ 46,826.00	\$ 50,278.00
Northwest	\$ 12,860.00	\$ 11,548.00	\$ 43,134.00	\$ 19,331.00	\$ 60,563.00	\$ 51,475.00	\$ 28,954.00	\$ 14,405.00
Southwest	\$ 37,087.00	\$ 61,506.00	\$ 54,084.00	\$ 14,715.00	\$ 17,811.00	\$ 32,814.00	\$ 47,853.00	\$ 44,639.00



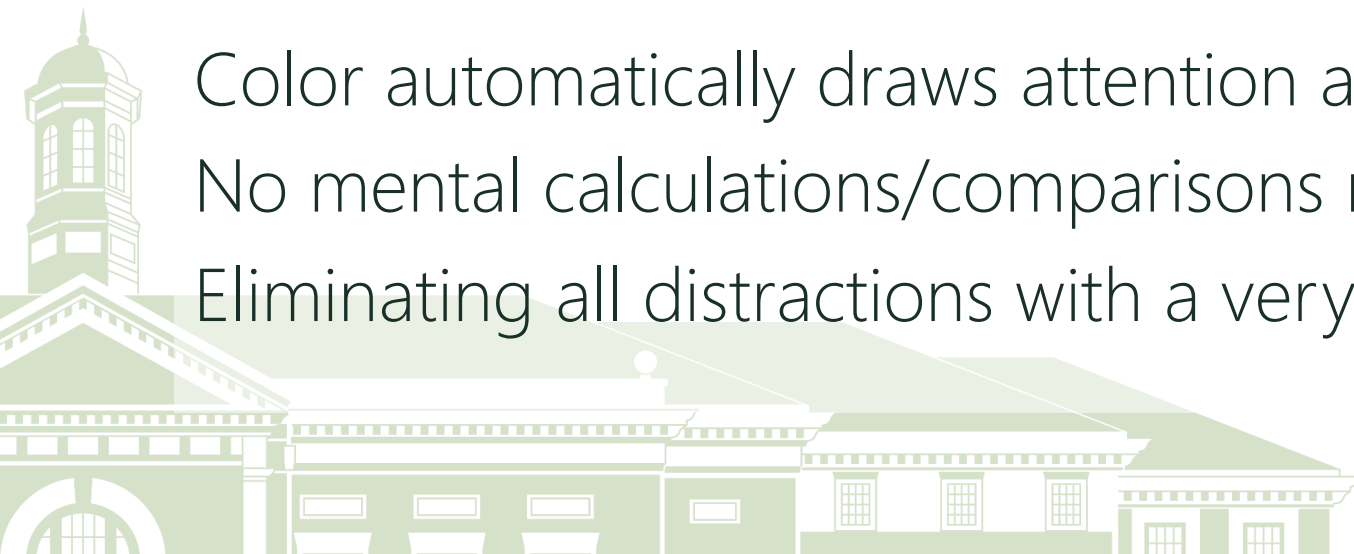
The Value of Highlighting

	Product A	Product B	Product C	Product D	Product E	Product F	Product G	Product H
Northeast	\$ 15,749.00	\$ 40,195.00	\$ 15,472.00	\$ 63,029.00	\$ 8,509.00	\$ 42,987.00	\$ 27,778.00	\$ 12,510.00
Southeast	\$ 48,044.00	\$ 20,741.00	\$ 40,643.00	\$ 15,687.00	\$ 12,342.00	\$ 23,297.00	\$ 10,401.00	\$ 10,522.00
Central	\$ 5,240.00	\$ 45,296.00	\$ 16,114.00	\$ 63,359.00	\$ 58,198.00	\$ 24,191.00	\$ 46,826.00	\$ 50,278.00
Northwest	\$ 12,860.00	\$ 11,548.00	\$ 43,134.00	\$ 19,331.00	\$ 60,563.00	\$ 51,475.00	\$ 28,954.00	\$ 14,405.00
Southwest	\$ 37,087.00	\$ 61,506.00	\$ 54,084.00	\$ 14,715.00	\$ 17,811.00	\$ 32,814.00	\$ 47,853.00	\$ 44,639.00

Color automatically draws attention and creates focus.

No mental calculations/comparisons needed.

Eliminating all distractions with a very simple viz can be effective.



The Value of Highlighting

Find the two p's:

Fjhvjfsjbfskjkjbfvkjvkjhefvkjvbejvhtehvekjbkjhvefjkhfekjhvefkjhfvkjhefkjhepfeiljtkljegfvljveffjefkjfnvkjhefvjhfkjhpfe
eljnkhfevkjhvefjkheffevfhvfekhfekjjhefefvekjefhkjhveefjvhtuejktnvphkjrejohrevreojhvrerverejkhrerevjkrreervr
eeveervukhreverjknvrerveljhrejlhrevjlhervjlhervjohrveenrjnerjnvejlrekjvneljrnvfjfhnrnvsgdbdgbdbgsngnhr

Highlighting makes them easier to find:

Fjhvjfsjbfskjkjbfvkjvkjhefvkjvbejvhtehvekjbkjhvefjkhfekjhvefkjhfvkjhefkjhe

p

feiljtkljegfvljveffjefkjfnvkjhefvjhfkjhpfe
eljnkhfevkjhvefjkheffevfhvfekhfekjjhefefvekjefhkjhveefjvhtuejktnv

p

hkjrejohrevreojhvrerverejkhrerevjkrreervr
eeveervukhreverjknvrerveljhrejlhrevjlhervjlhervjohrveenrjnerjnvejlrekjvneljrnvfjfhnrnvsgbgbbrbrgbbgbsbrgg

Makes sure you are truthful!

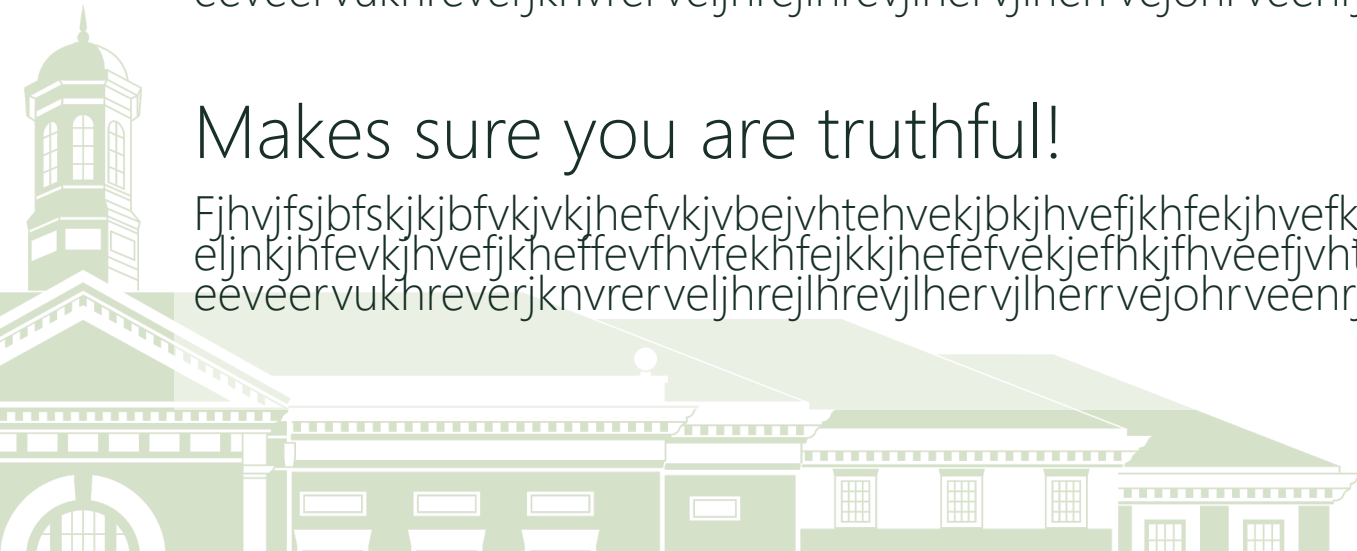
Fjhvjfsjbfskjkjbfvkjvkjhefvkjvbejvhtehvekjbkjhvefjkhfekjhvefkjhfvkjhefkjhe

p

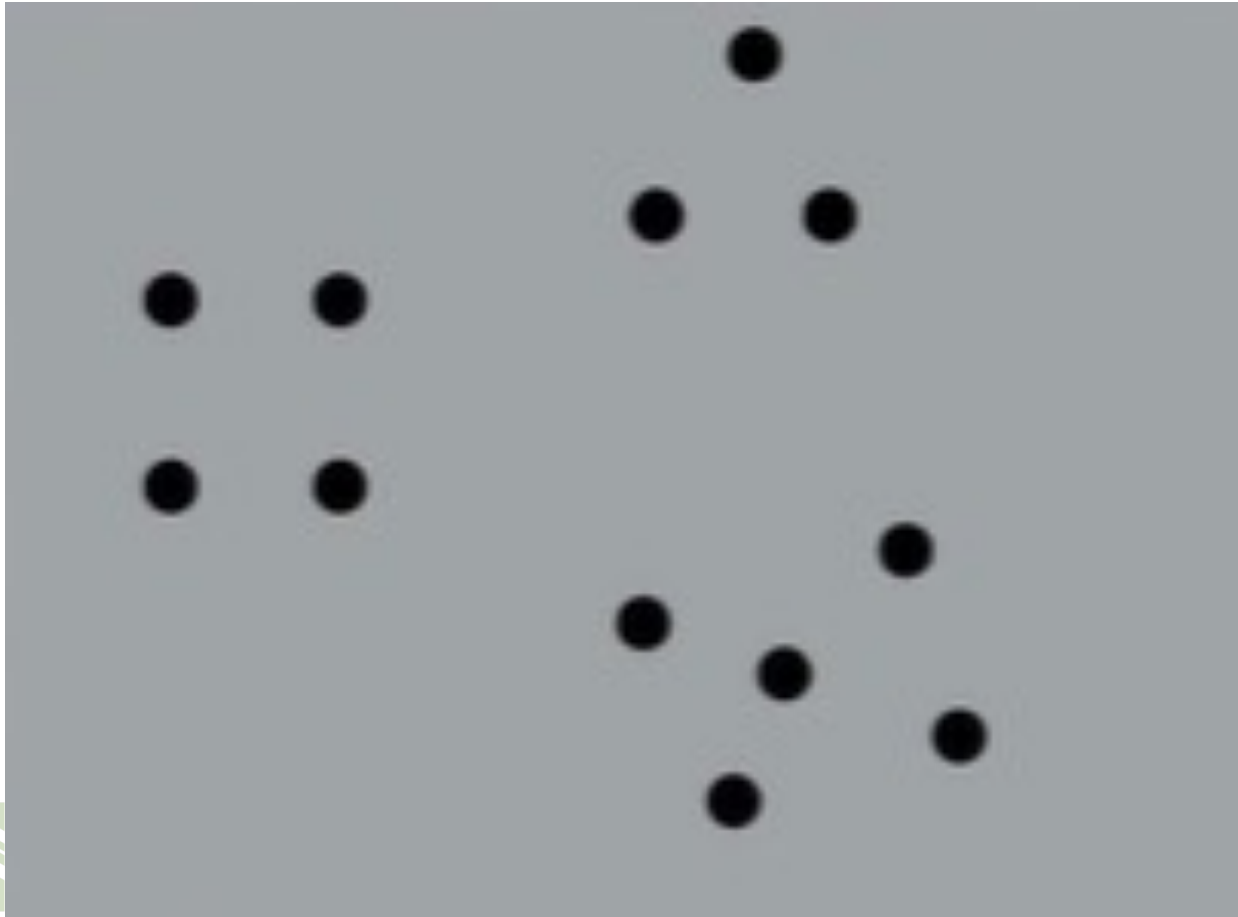
feiljtkljegfvljveffjefkjfnvkjhefvjhfkjhpfe
eljnkhfevkjhvefjkheffevfhvfekhfekjjhefefvekjefhkjhveefjvhtuejktnv

p

hkjrejohrevreojhvrerverejkhrerevjkrreervr
eeveervukhreverjknvrerveljhrejlhrevjlhervjlhervjohrveenrjnerjnvejlrekjvneljrnvfjfhnrnvjyrthmilthgbwbwbgsgge



The Value of Spacing



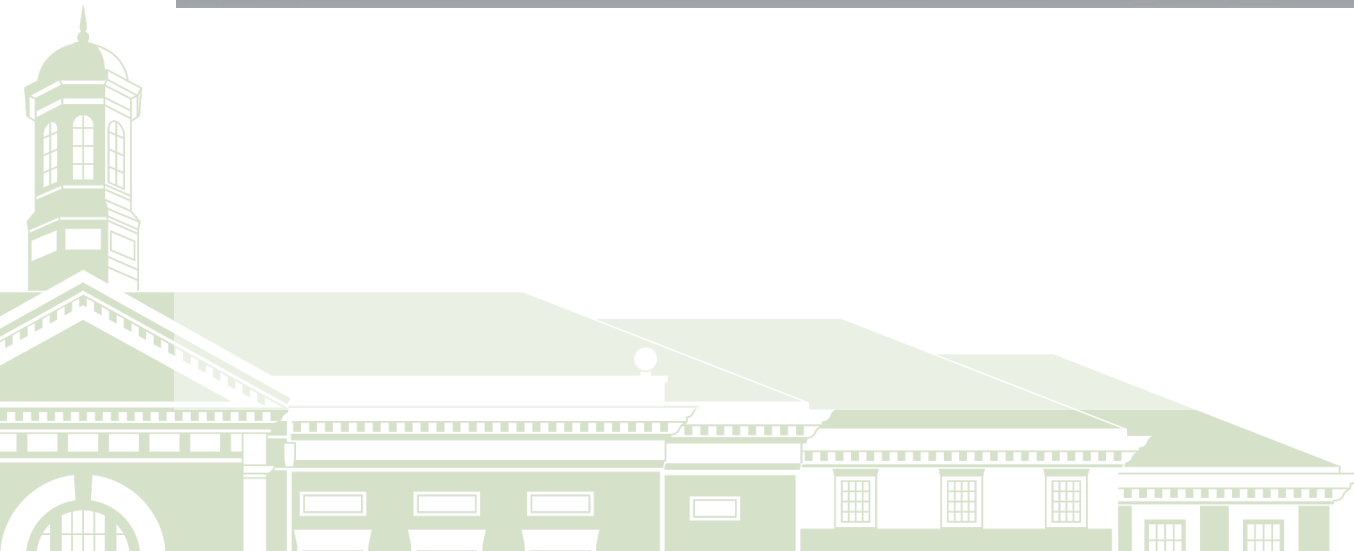
Which dots are related?
How many groups do we have?



The Value of Spacing



Do you see rows or columns?



Gestalt Principles



Good Figure

Objects grouped together tend to be perceived as a single figure. Tendency to simplify.



Proximity

Objects tend to be grouped together if they are close to each other.



Similarity

Objects tend to be grouped together if they are similar.



Continuation

When there is an intersection between two or more objects, people tend to perceive each object as a single uninterrupted object.



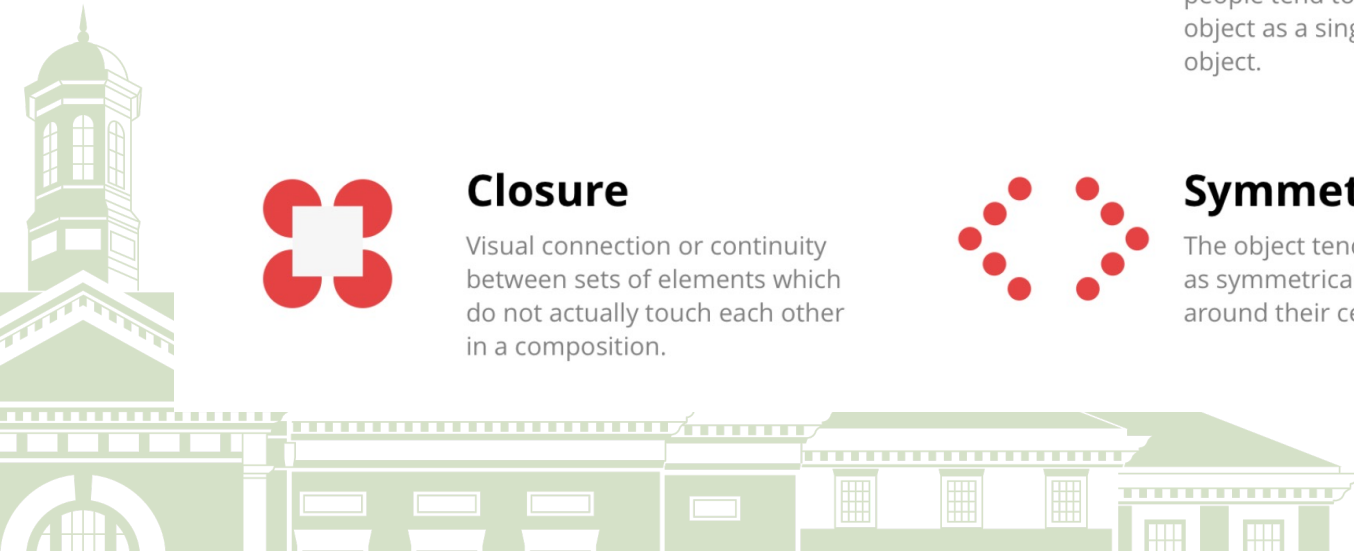
Closure

Visual connection or continuity between sets of elements which do not actually touch each other in a composition.



Symmetry

The object tend to be perceived as symmetrical shapes that form around their center.



Using Color to Highlight

Earthtones

Gentle browns, blues. Calming, sinks into the page



Cool

Soothing, restful, calm



Unnatural colors

Alarming, unnerving. draws attention.



Warm

Optimistic, active, vivid



Increasing color intensity

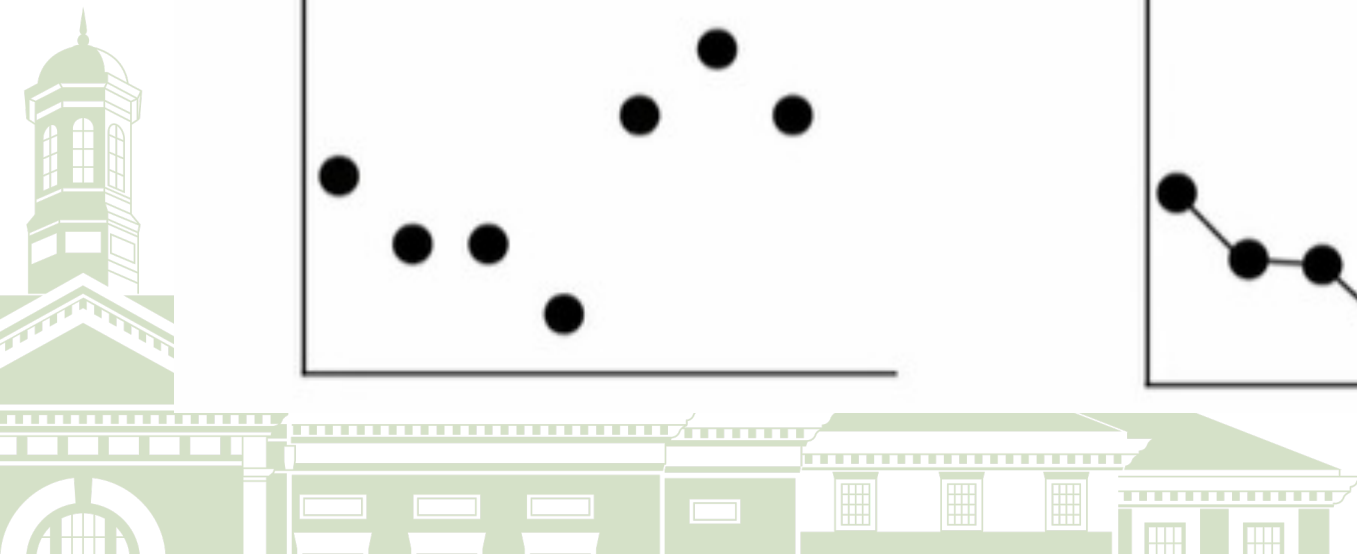
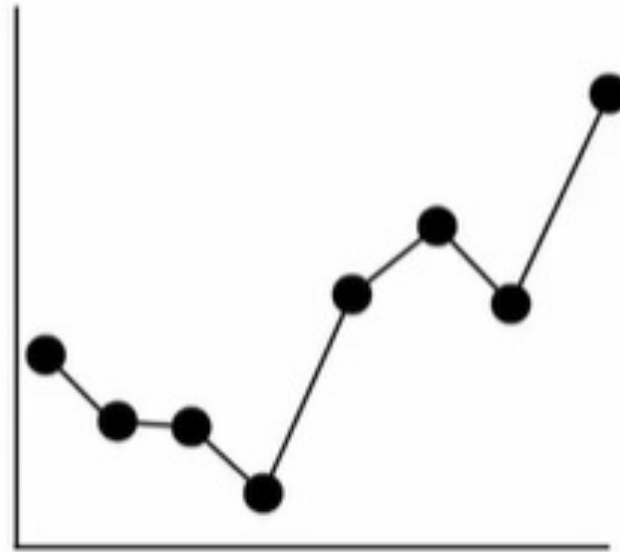
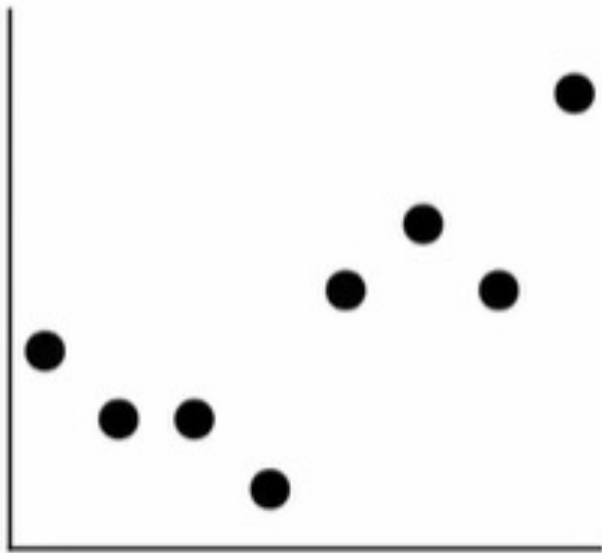
Increasing saturation and brightness draws the eye and means the point is more important



Source: *Juice Analytics Whitepaper (part 3)*

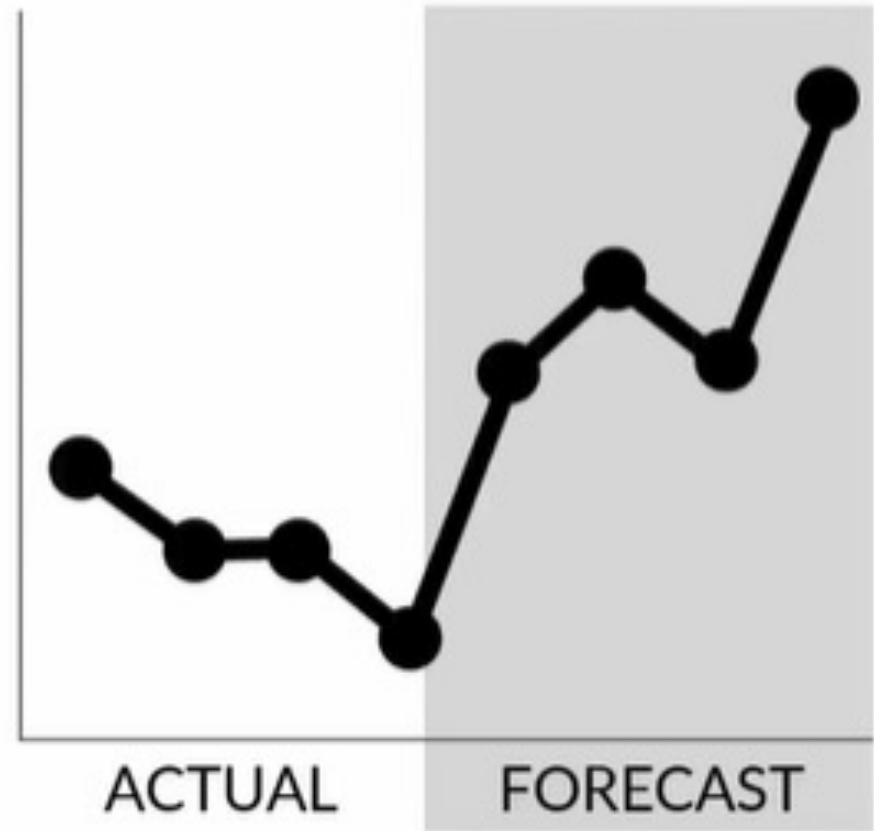
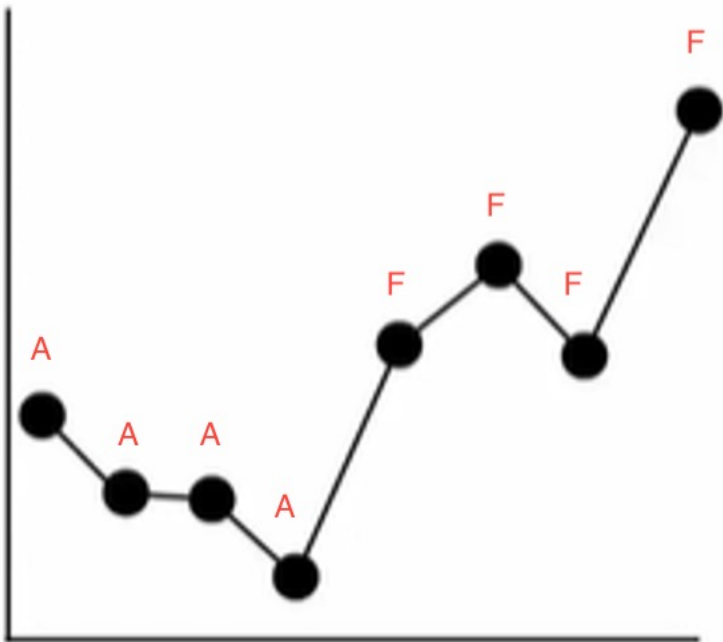
Application

Which is better for time-series data?

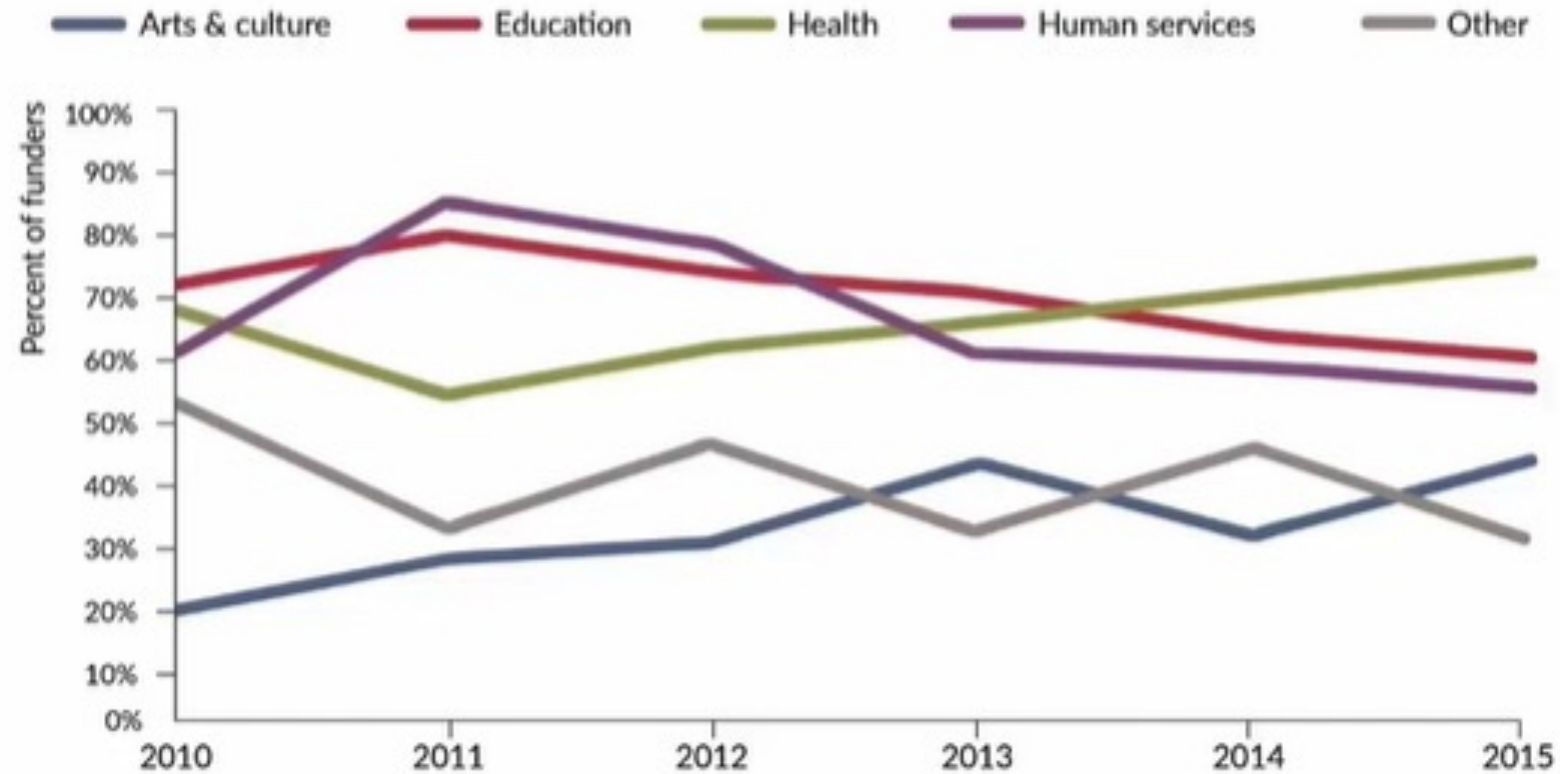


Application

How did we create difference between forecast and actual?



What is this graph's purpose?



Guidelines: Schaffer 4 C's of Data Visualization

Clear - easily seen; sharply defined

- who's the audience? what's the message?
- clarity more important than aesthetics

Clean - thorough; complete; unadulterated

- labels, axis, gridlines, formatting, right chart type, color choice, etc.

Concise - brief but comprehensive

- not minimalist but not verbose

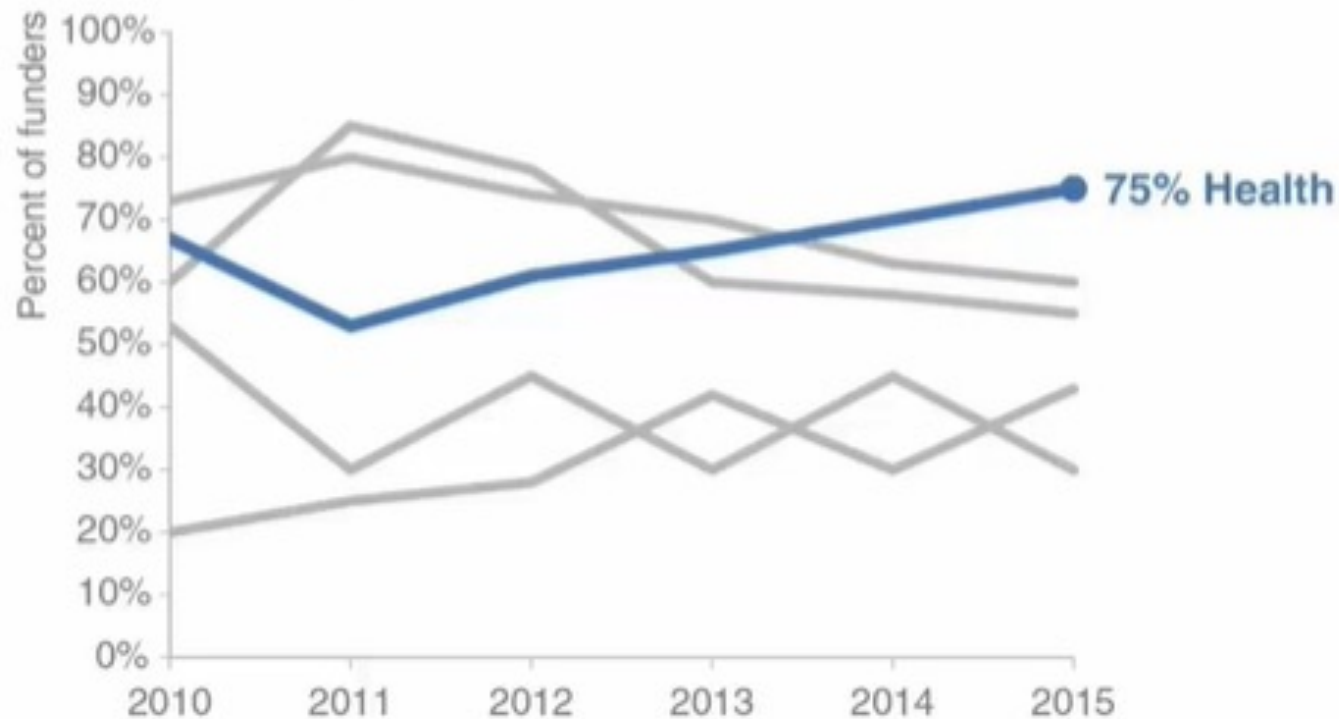
Captivating - to attract and hold by beauty or excellence

- does it capture attention? is it interesting? does it tell the story?



Improve Visualization

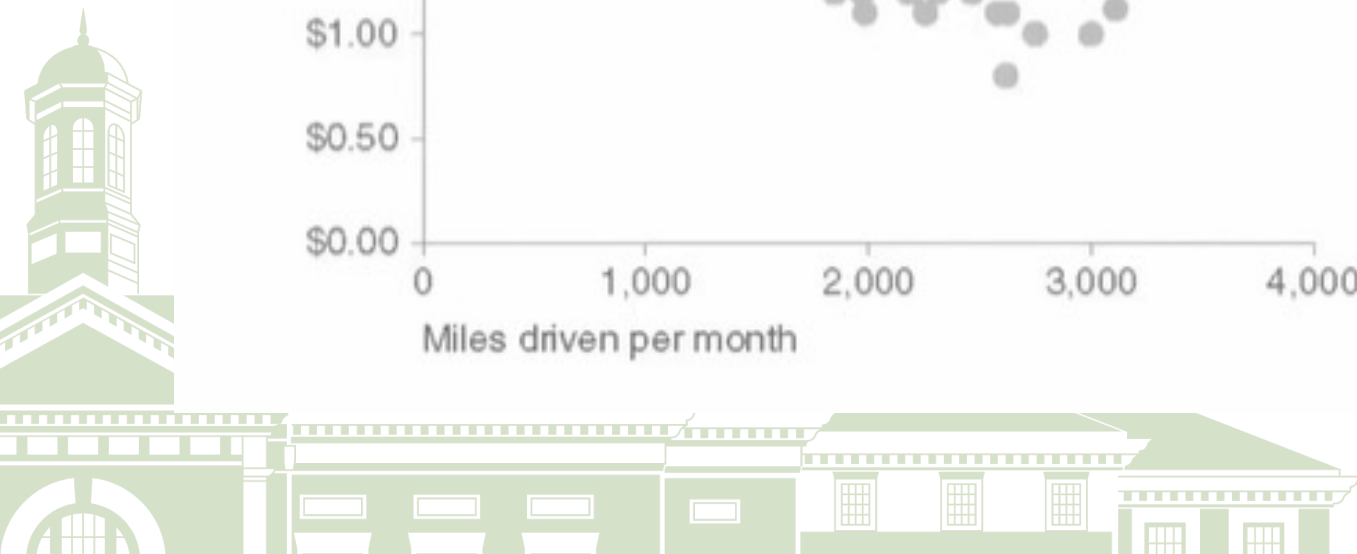
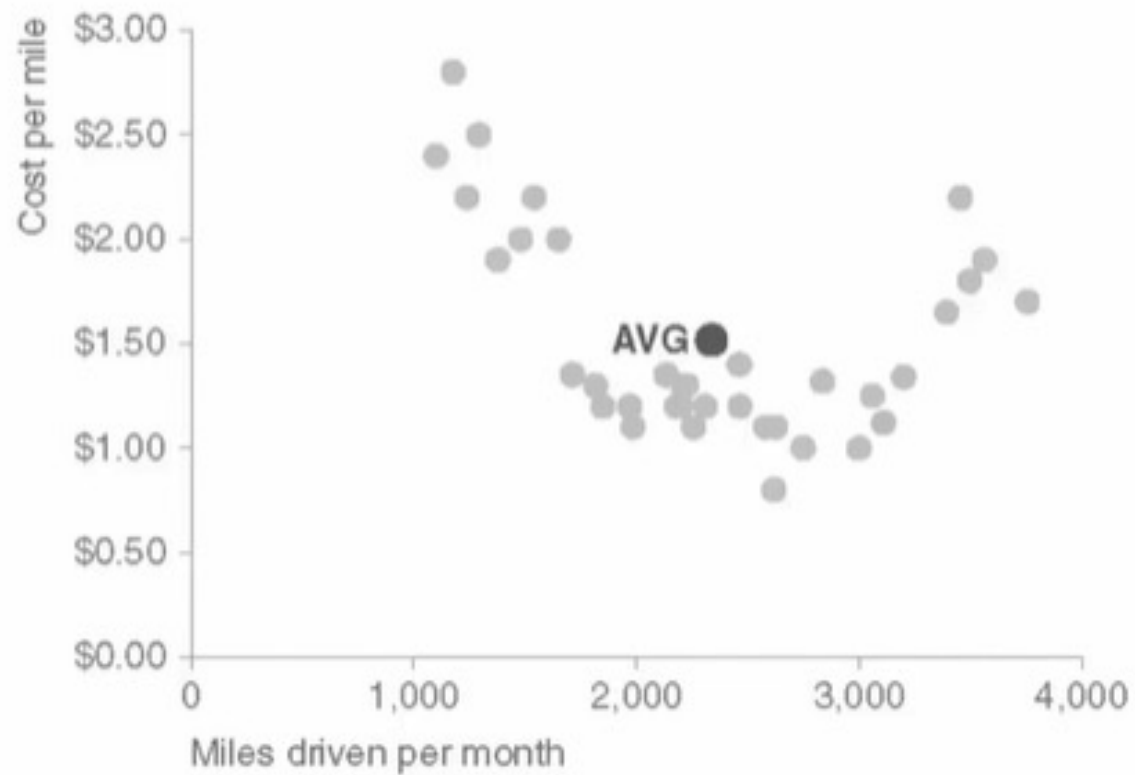
Types of non-profits supported by area funders



Data is self-reported by funders; percents sum to greater than 100 because respondents can make multiple selections.

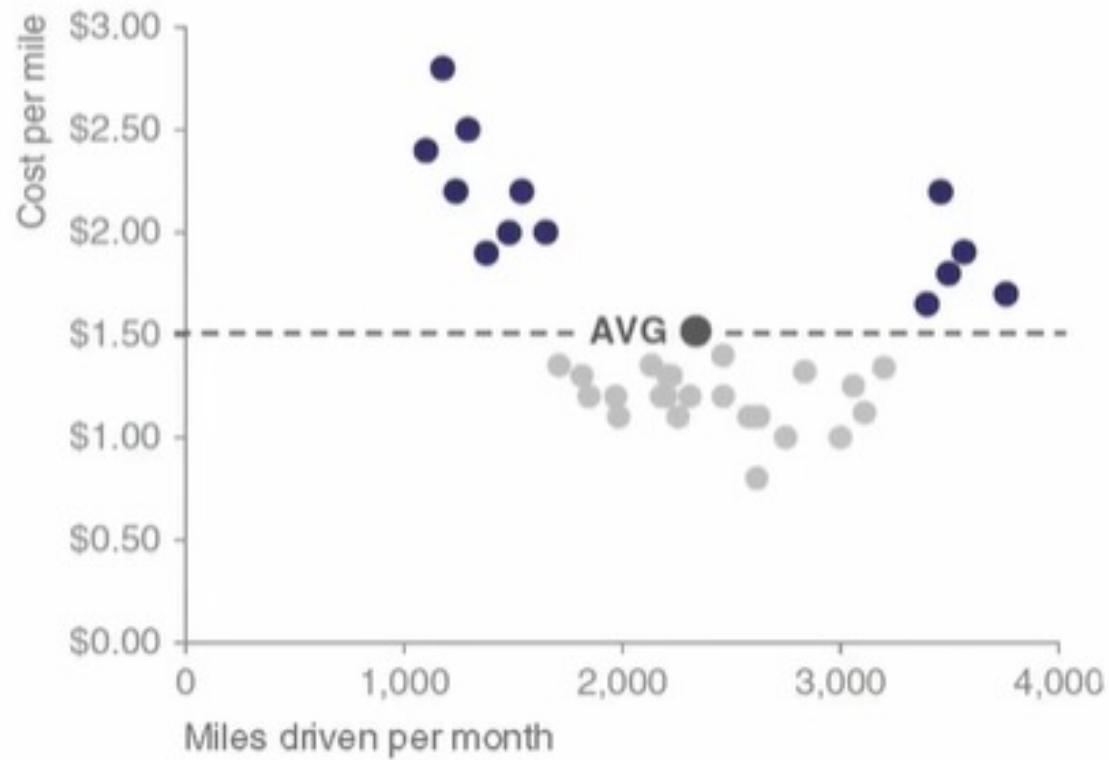
Visualization

Cost per mile by miles driven

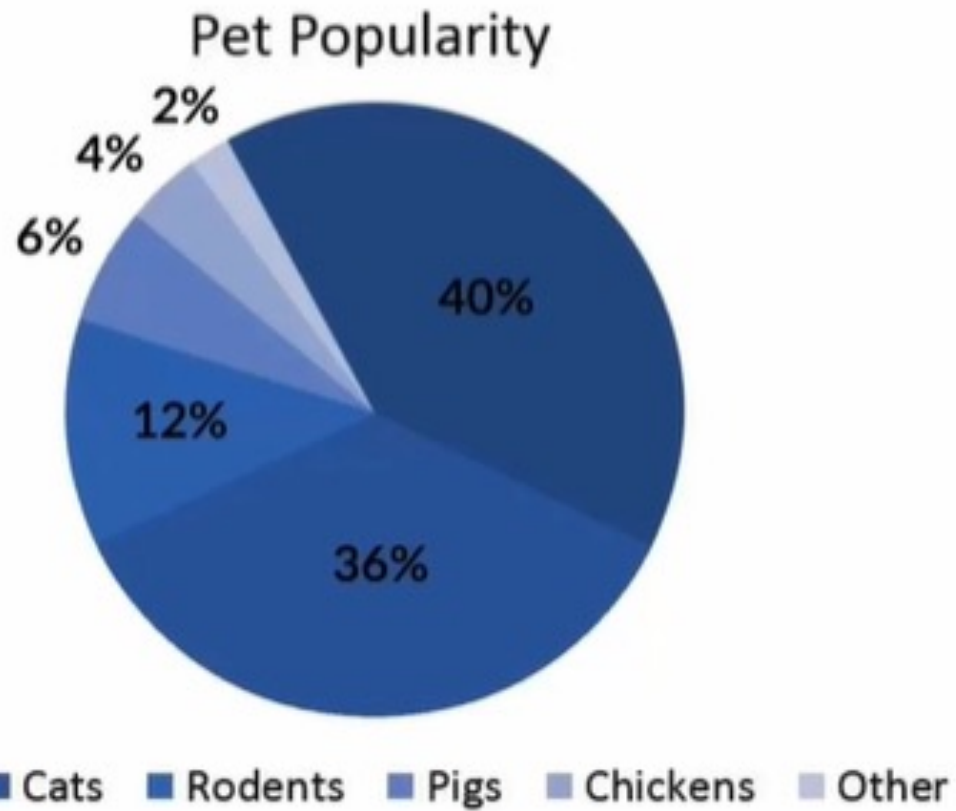


Improved Visualization

Cost per mile by miles driven

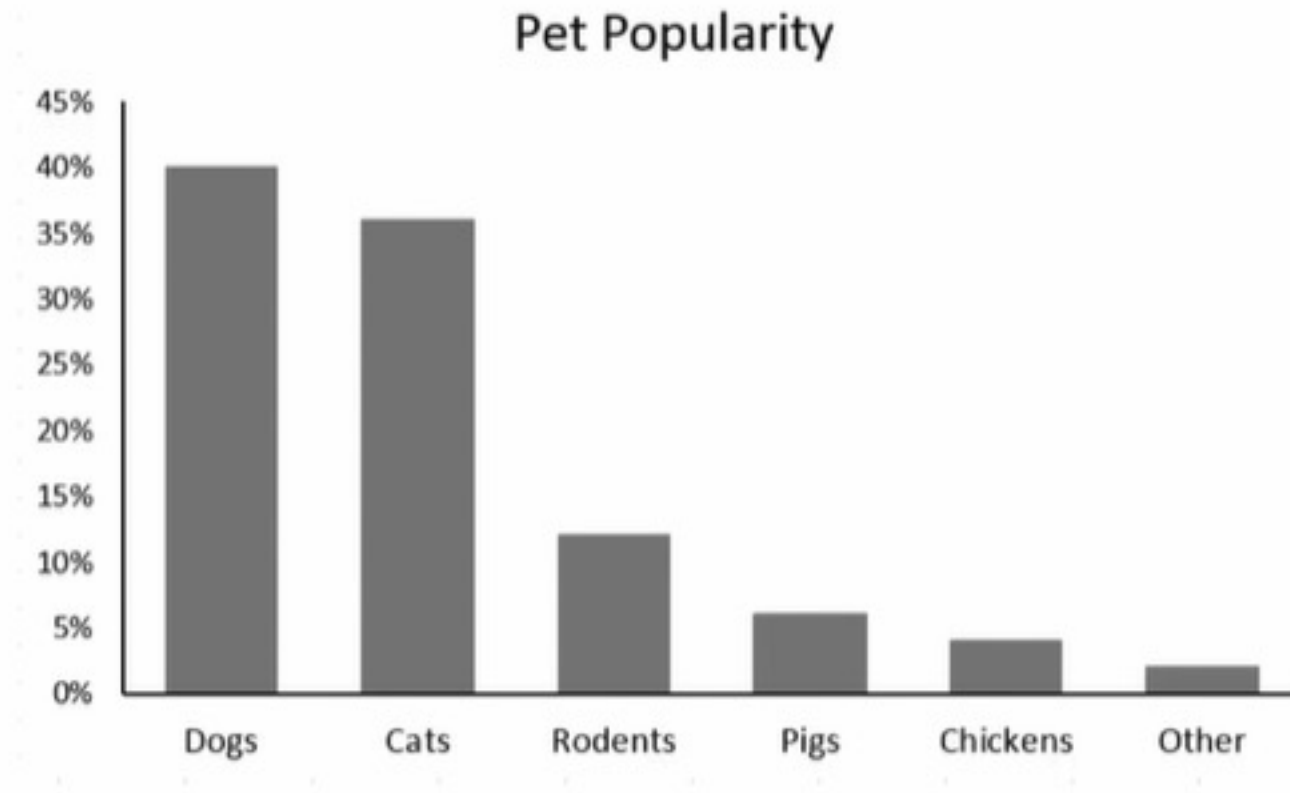


Pie Charts



What is the third most popular pet?

Pie Charts



Guidelines: Tufte's Five Laws of Data Ink

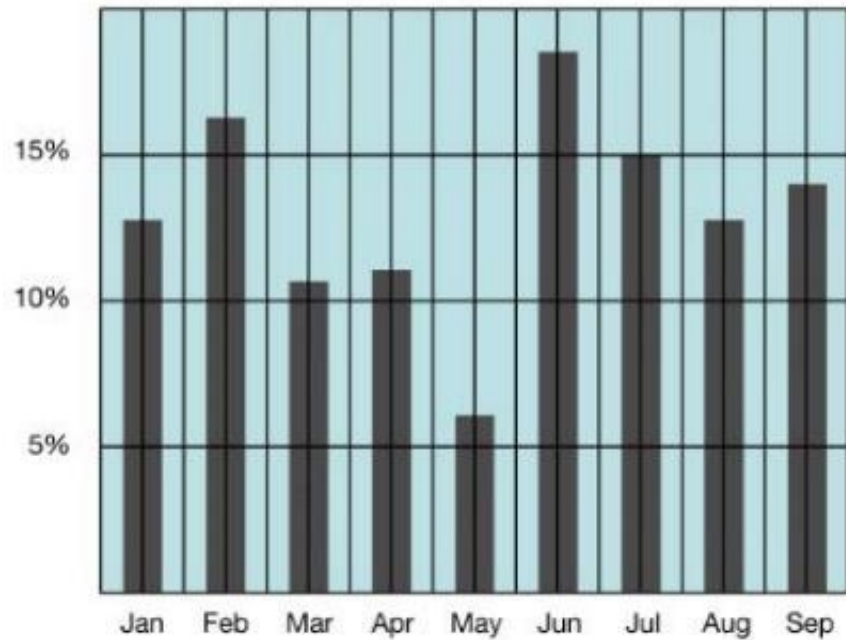
1. Above all else show the data
2. Maximize the data-ink ratio
3. Erase non-data-ink
4. Erase redundant data-ink
5. Revise and edit

"It seems that perfection is reached not when there is nothing to add, but when there is nothing left to take away."

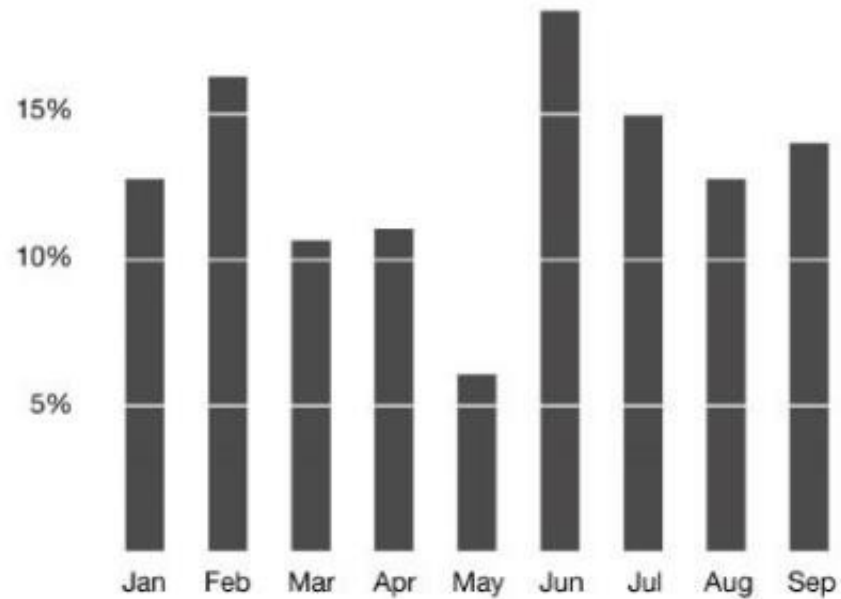
- Antoine de Saint Exupéry

Guidelines: Tufte's Five Laws of Data Ink

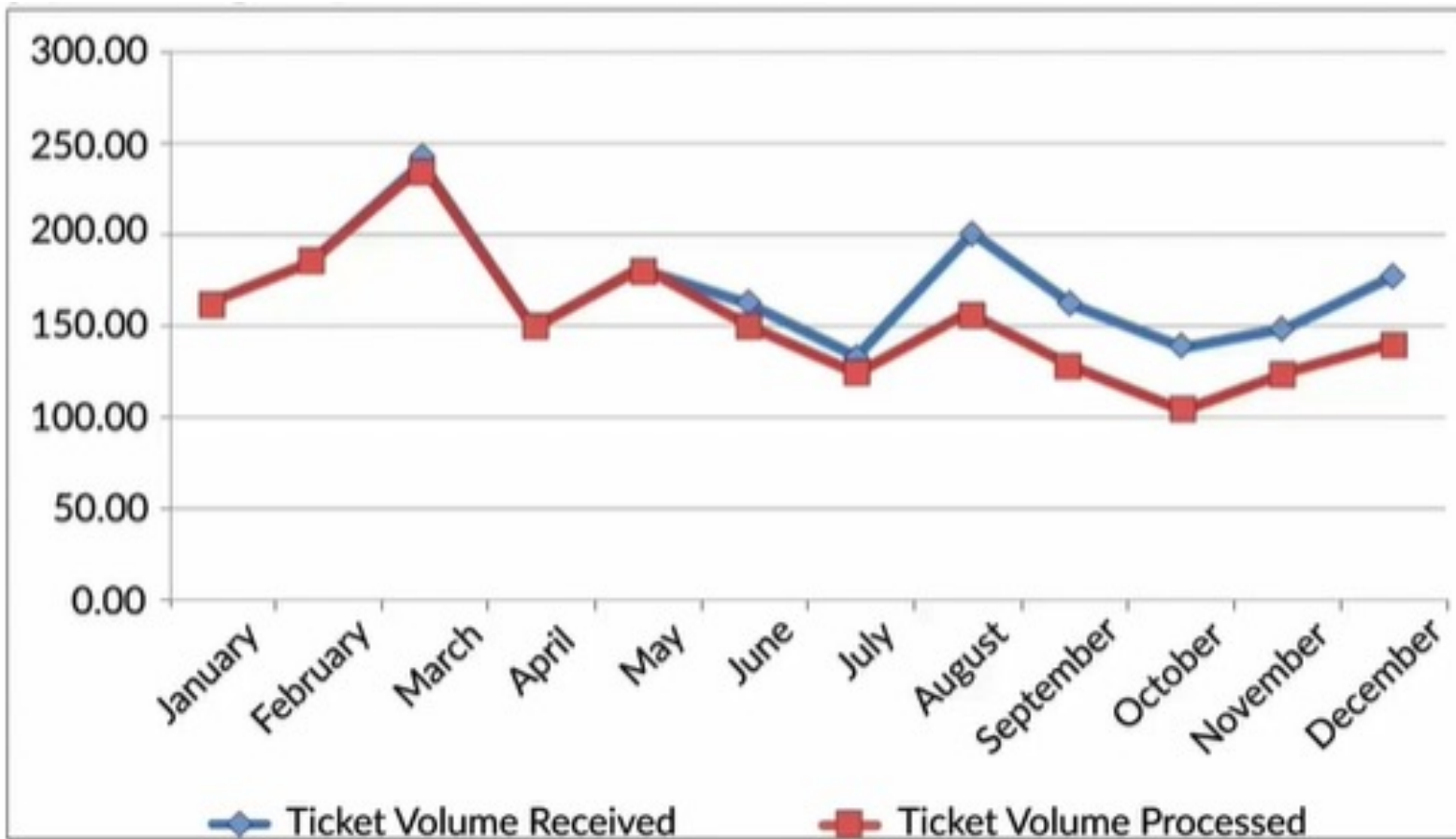
LOW INK-DATA RATIO



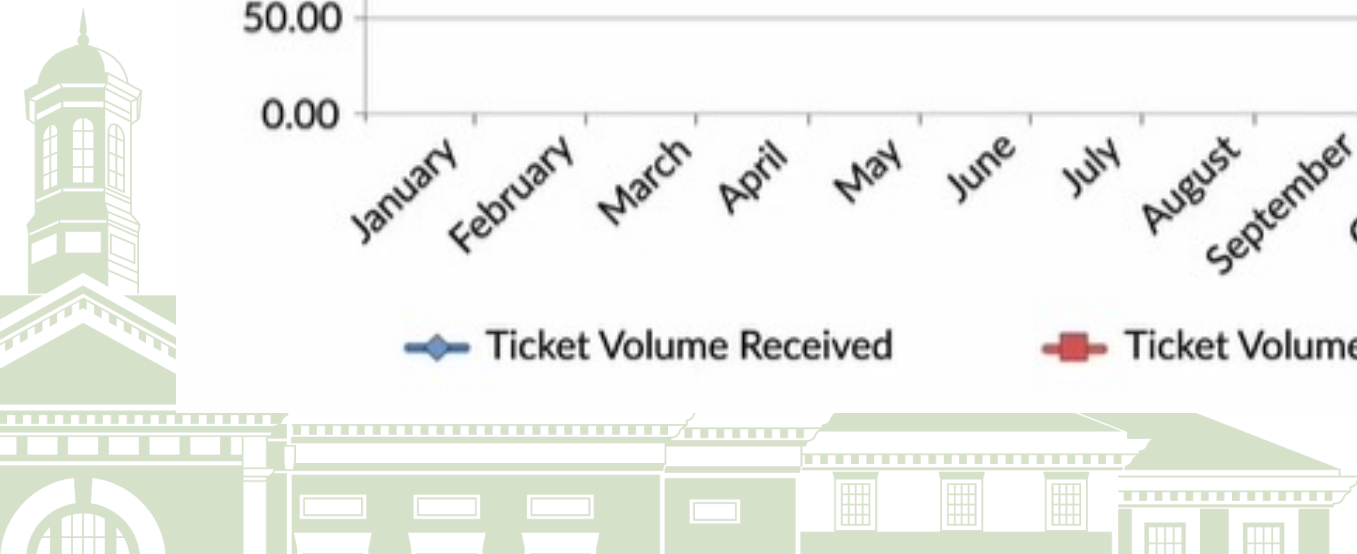
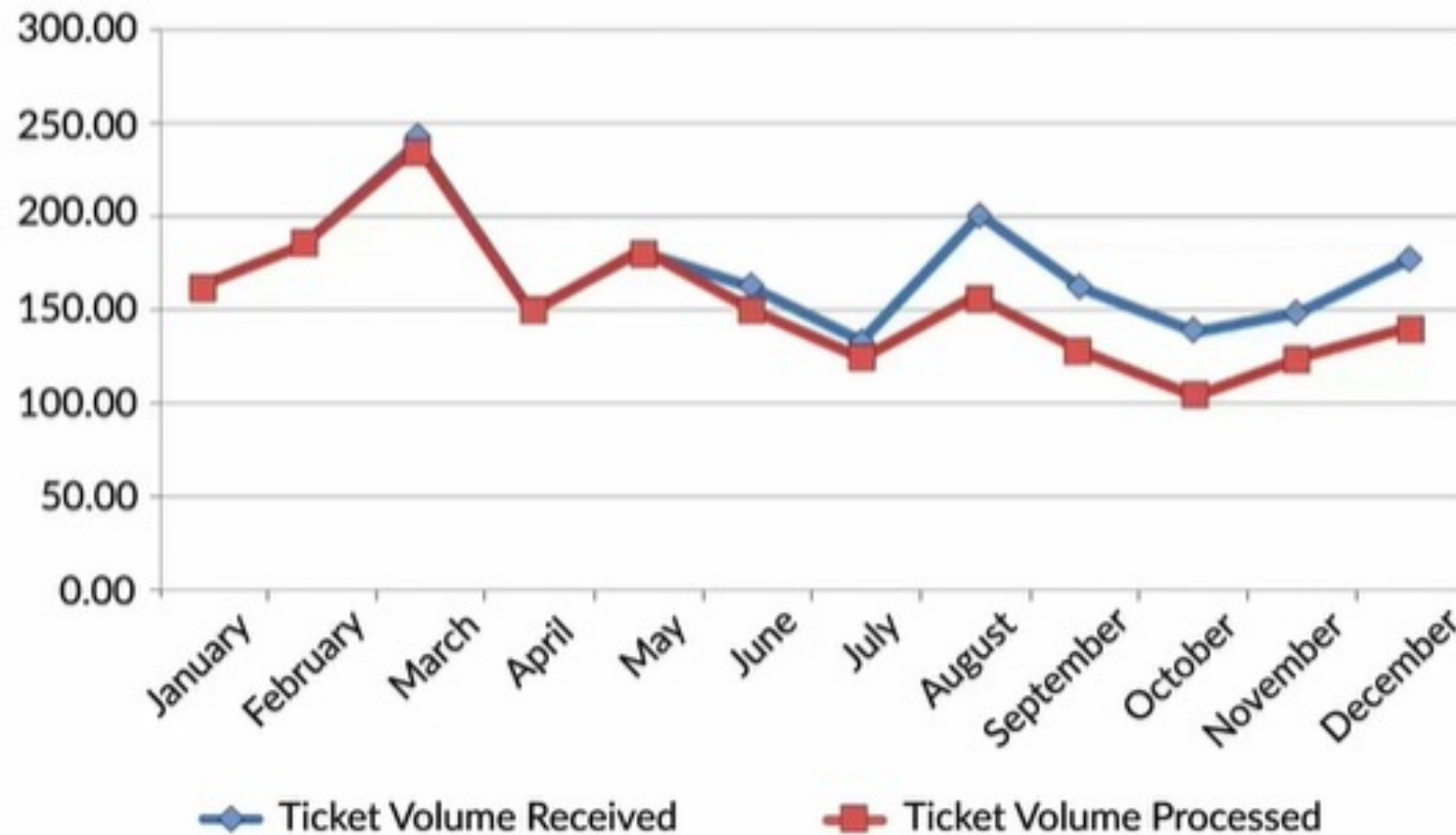
HIGH INK-DATA RATIO



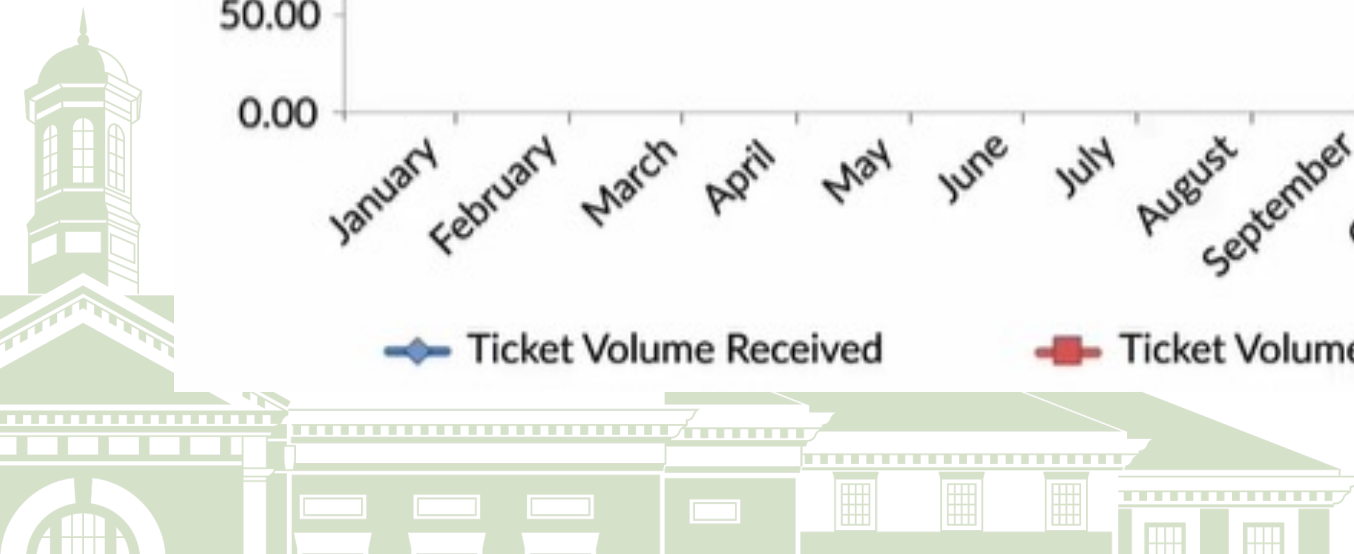
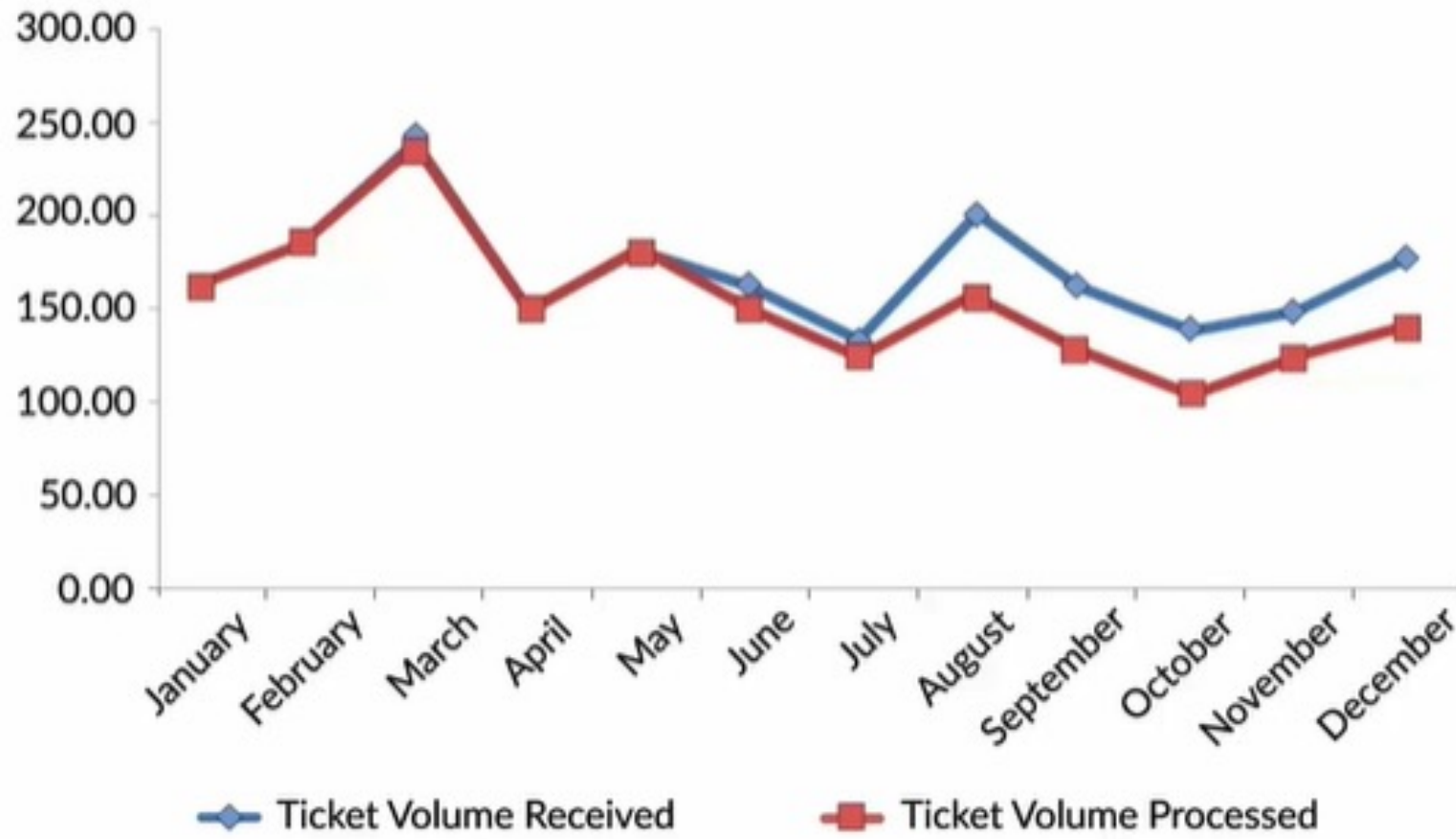
Guidelines: Tufte's Five Laws of Data Ink



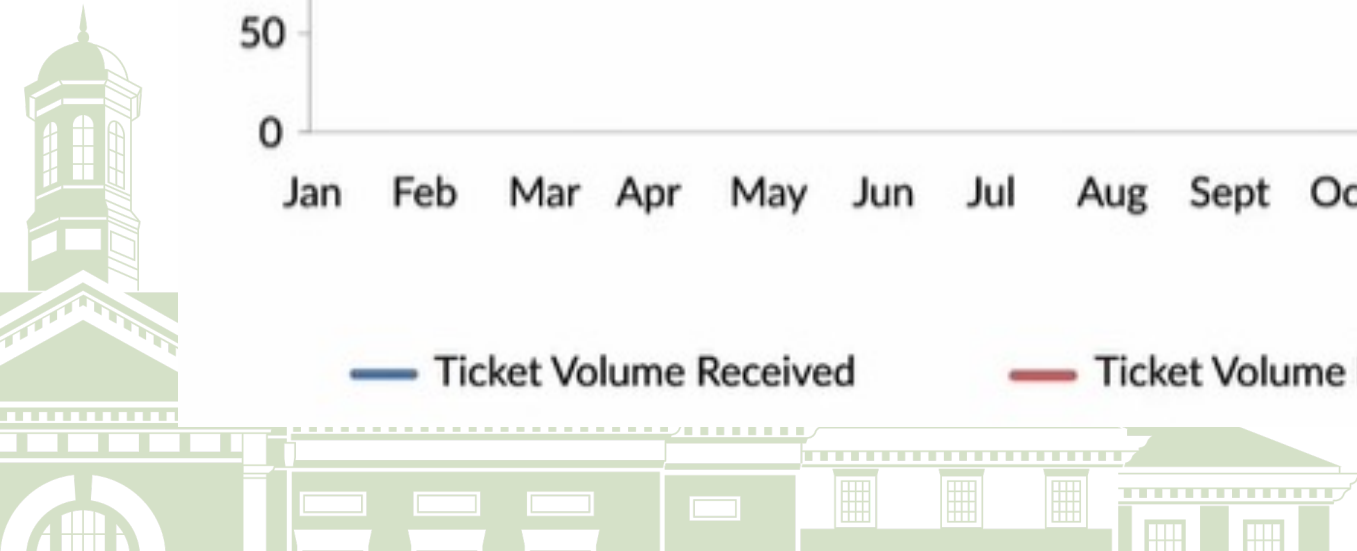
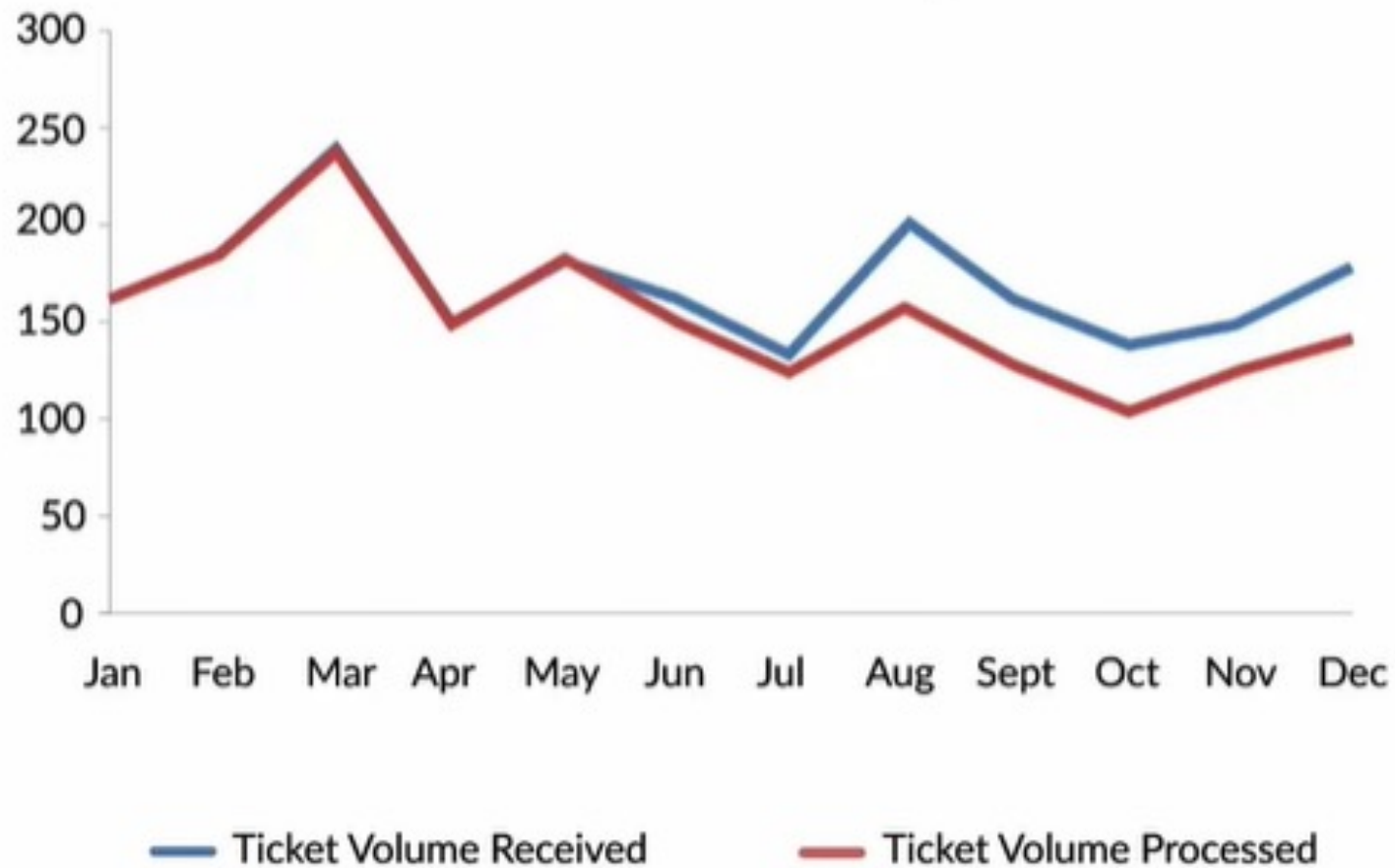
Remove graph border



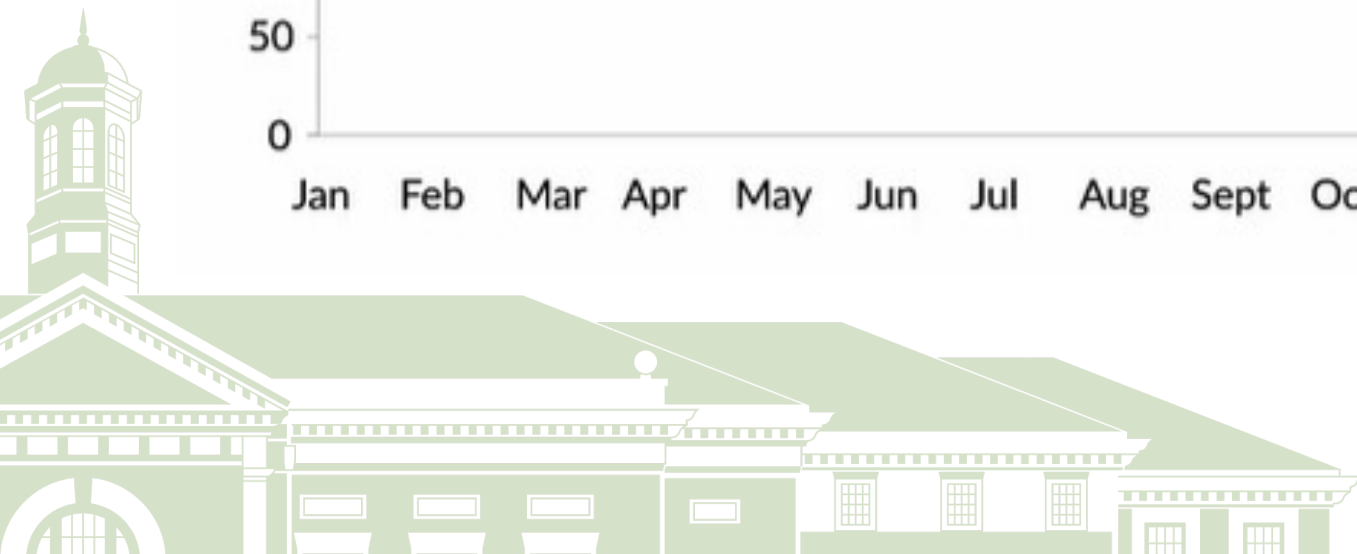
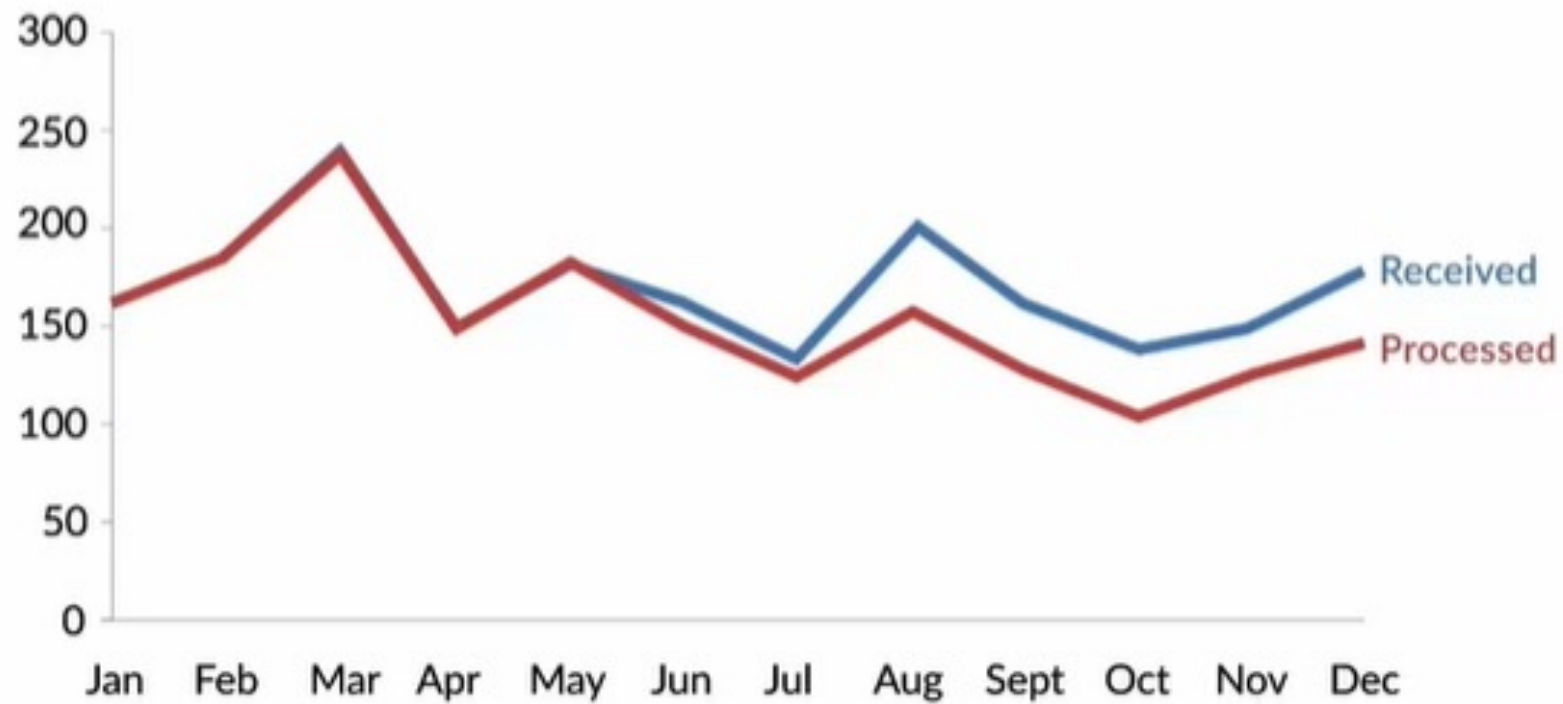
Remove gridlines



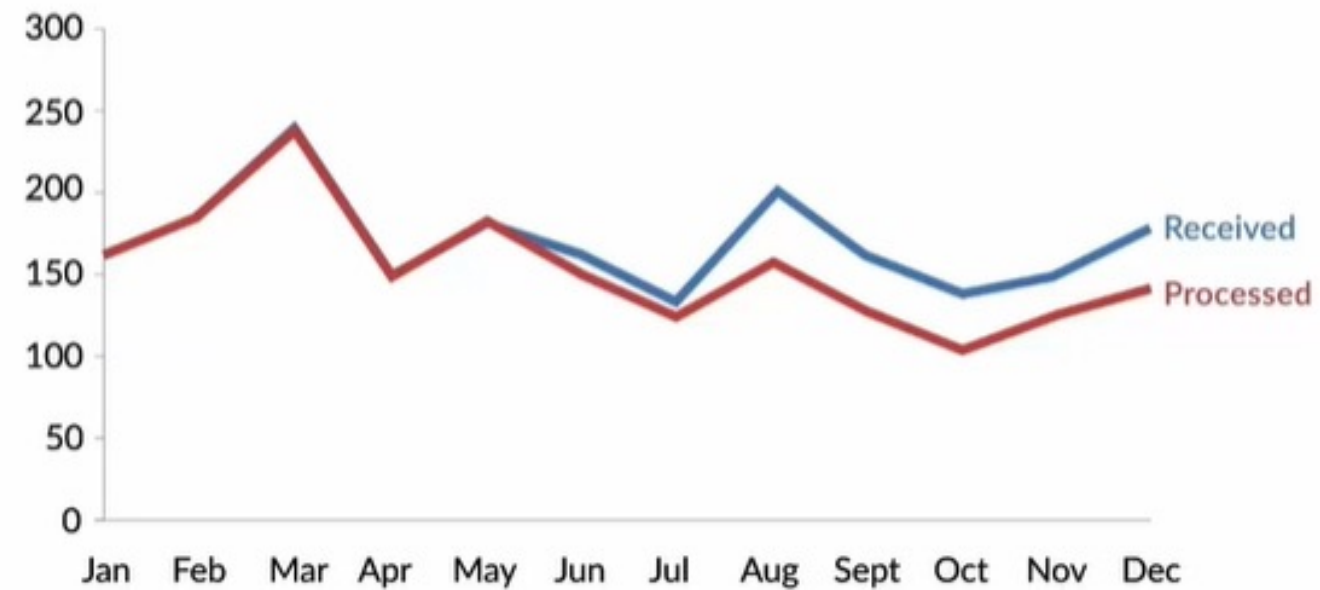
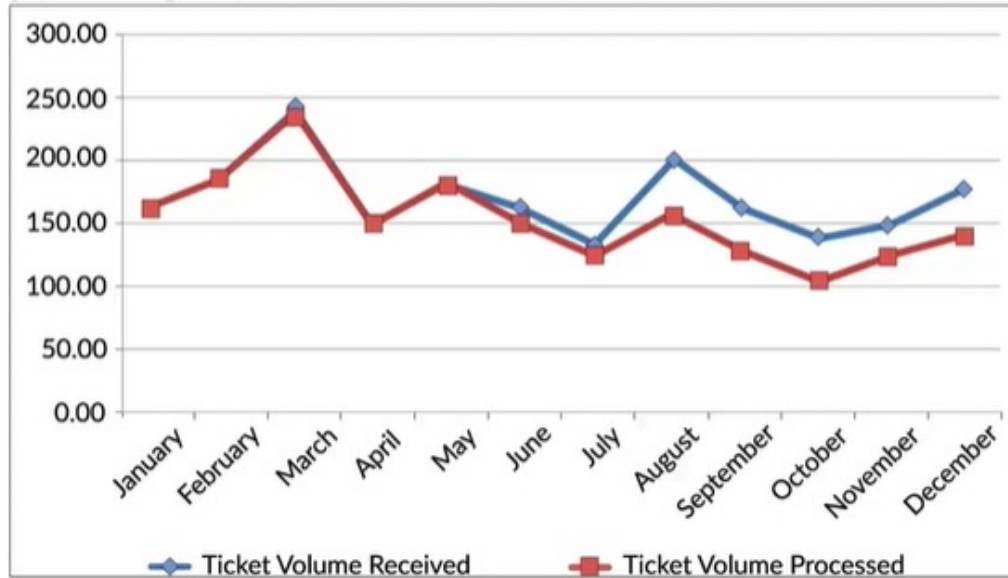
Abbreviate Months, Remove Points



Put Legend on Graph



Before and After



Stephen Few Guide To Visualization

The **effectiveness** of a data visualization is determined by its **creator's intentions** and its **audience's needs**.

What is your role in the organization? expertise?

Who is your audience? What are their needs and expectations?

The visualization should **inform** people to the degree that will **serve their needs** as well as possible.

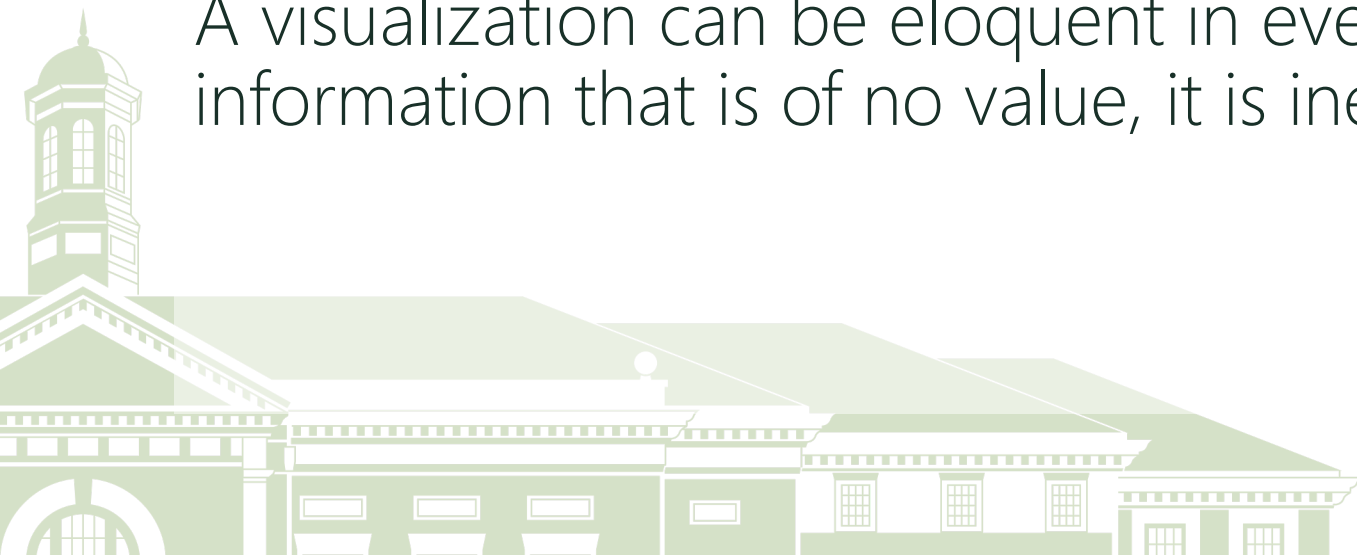


Stephen Few's Guide To Visualization

Usefulness

Data visualization is of little value if it helps people understand something that doesn't matter to them.

A visualization can be eloquent in every way, but if it communicates information that is of no value, it is ineffective.



Stephen Few's Guide To Visualization

Completeness

An effective data visualization includes the information that's needed to produce the intended level of understanding, but not more.

The context that's needed to understand the information has been provided.

Context is usually provided in the form of comparisons, such as comparisons to targets, measures of the norm, and historical values.





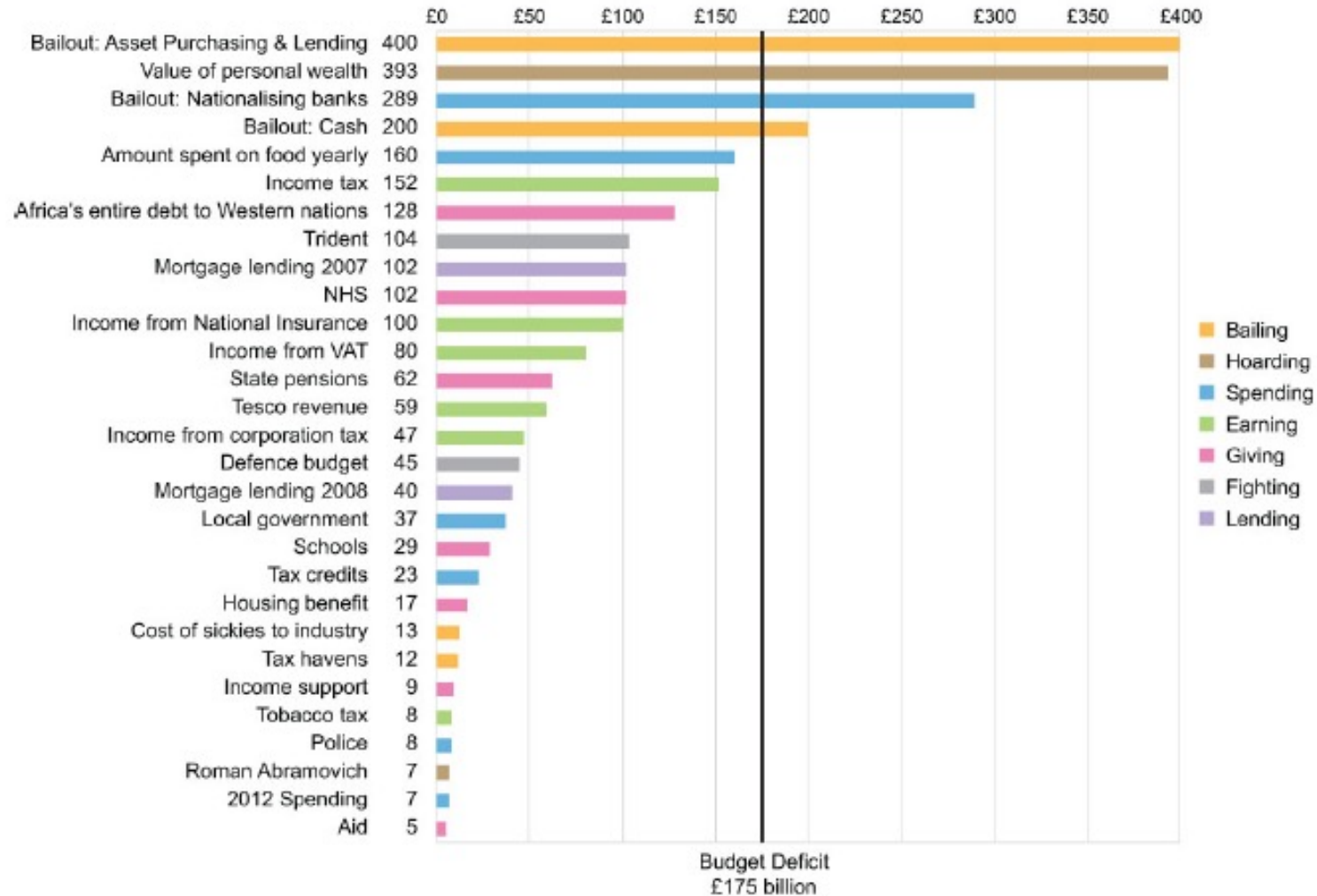
The Billion Pound-O-Gram

David McCandless / InformationIsBeautiful.net

■ Giving
 ■ Spending
 ■ Fighting
 ■ Hoarding
 ■ Lending
 ■ Bailing
 ■ Earning

Source: UK Treasury, Guardian

The UK Budget Deficit Compared to Other Big Amounts of Money (Billions of Pounds)

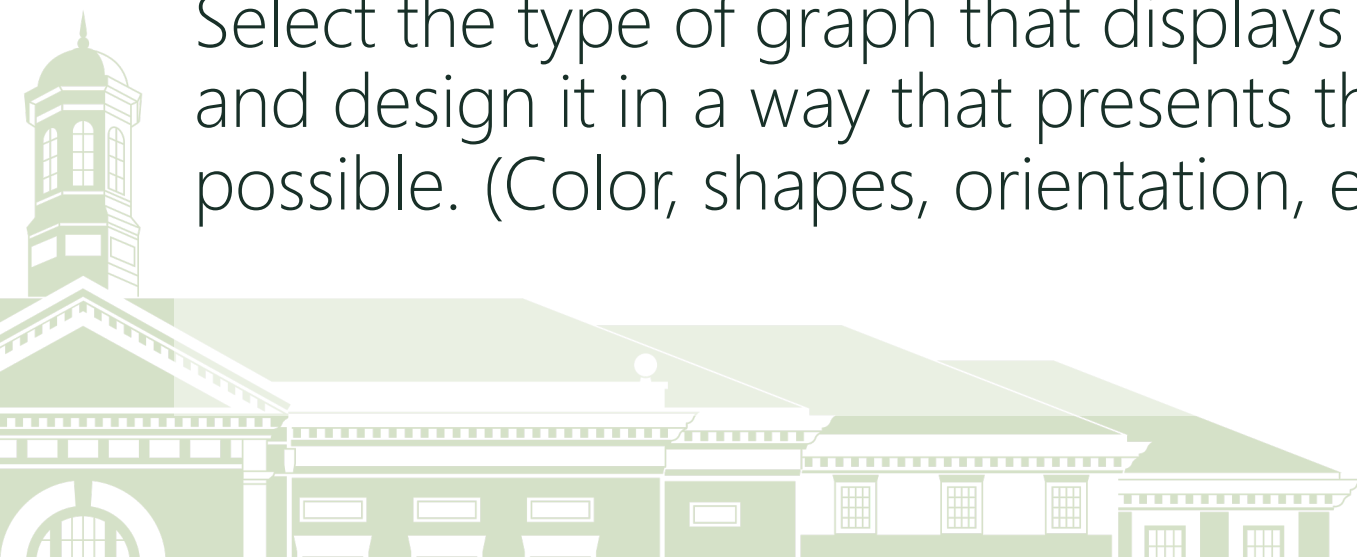


Stephen Few's Guide To Visualization

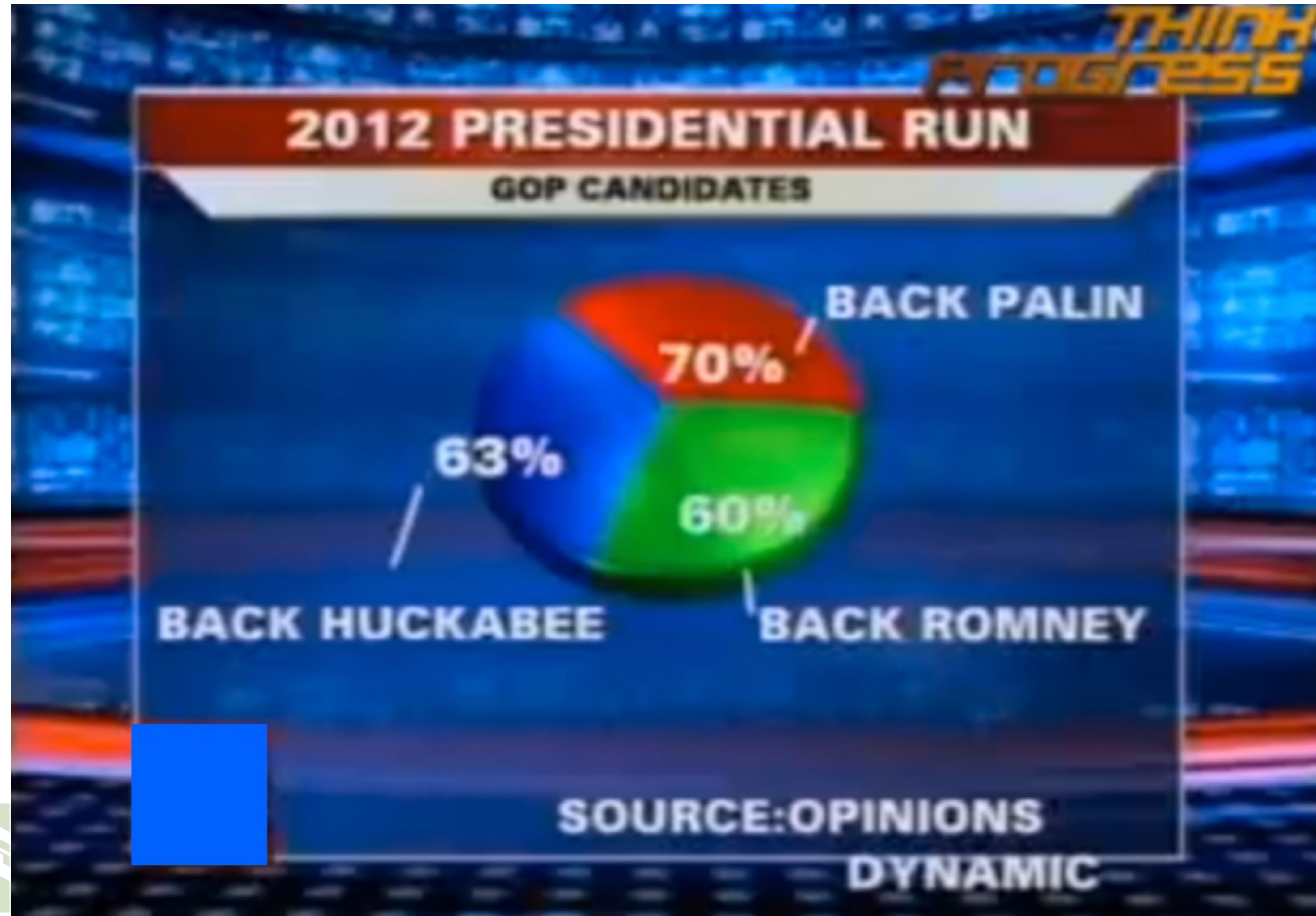
Perceptibility

The information must be displayed in a manner that the human eye and brain can perceive with minimal effort and appropriate precision.

Select the type of graph that displays the information most effectively and design it in a way that presents the information as clearly as possible. (Color, shapes, orientation, etc.)



Be Careful...



Stephen Few's Guide To Visualization

Truthfulness

The visualization should be both accurate and valid.

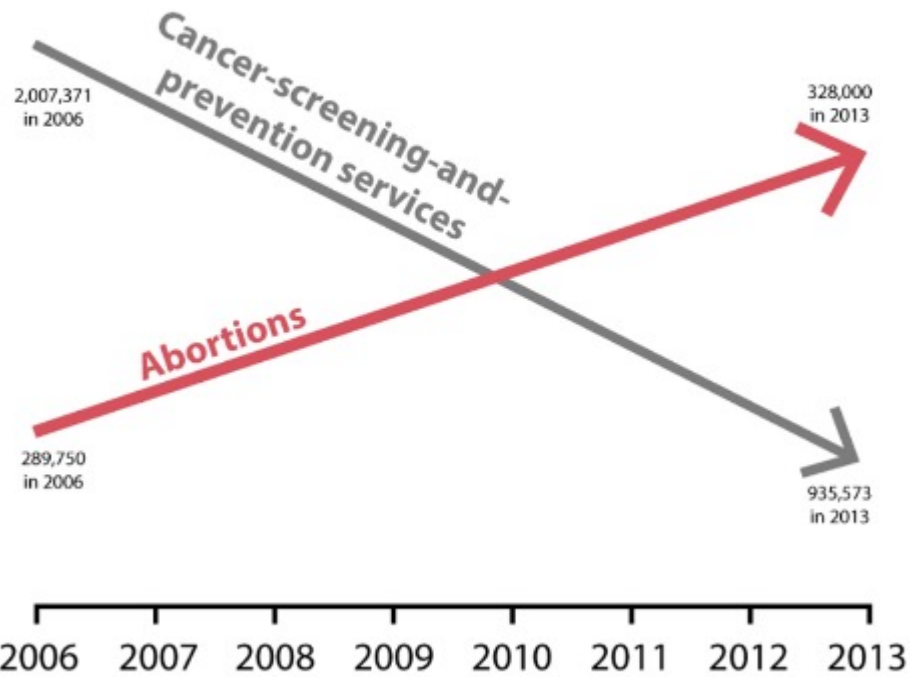
Accuracy is a measure of reliability and appropriate precision.

Validity indicates how well something represents what it claims.



Be Careful...

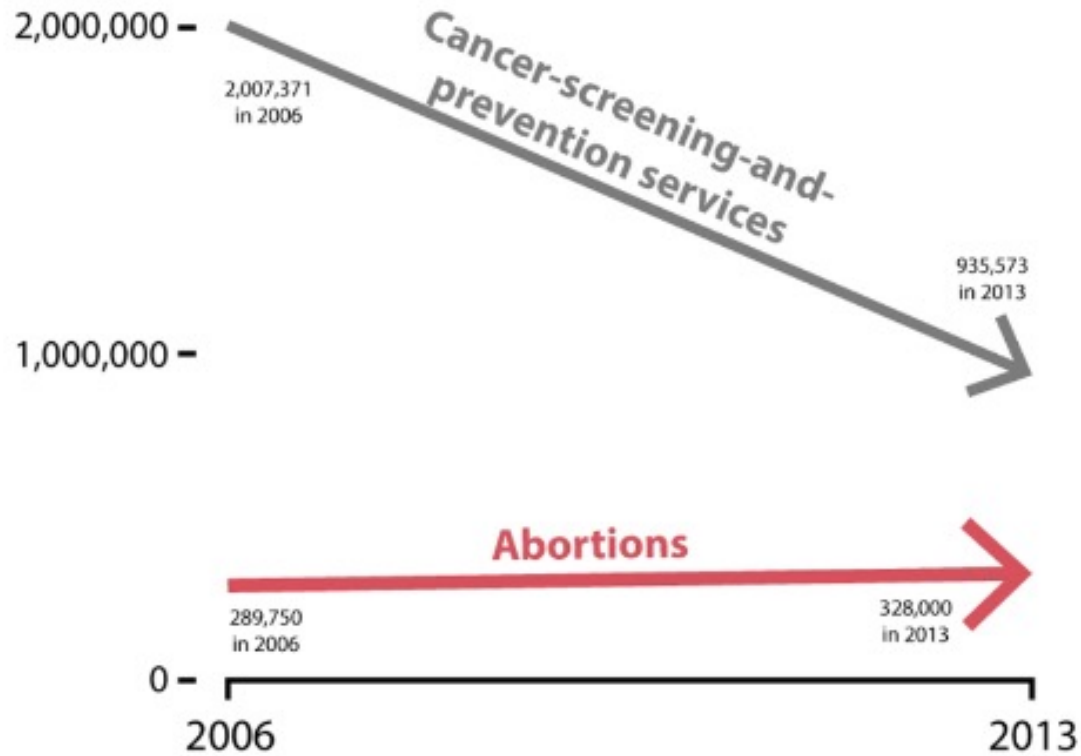
Planned Parenthood Federation of America:
Abortions up—life-saving procedures down



(Source: Americans United for Life)

Be Careful...

Planned Parenthood Federation of America



Stephen Few's Guide To Visualization

Intuitiveness

A data visualization is intuitive to the degree that it is familiar and easily understood.

Is your audience familiar with the type of visualization? Does the type of visualization need explain? Is there a better visualization type that achieves the same goal?



Stephen Few's Guide To Visualization

Aesthetics

The aesthetic quality of a data visualization can range from ugly to beautiful.

Ugly: a display that no one would choose to view for long.

Beautiful: One that would invite even those who care little for the information to take perhaps delight in learning about it.



Stephen Few's Guide To Visualization

Engagement

Engagement: quality that can be achieved by various means, including but not limited to aesthetics, that invites the audience to examine the information.



Informative

Usefulness

Useless

Very useful

Completeness

No relevant data

All relevant data

Perceptibility

Unclear and difficult

Clear and easy

Truthfulness

Inaccurate and/or invalid

Accurate and valid

Intuitiveness

Unfamiliar; difficult to understand

Familiar; easy to understand

Emotive

Aesthetics

Ugly

Pleasing to the eye

Beautiful

Engagement

Distracts from data

Neutral

Draws one into the data

Rate some visualizations that you find here:

<https://public.tableau.com/en-us/gallery/?tab=viz-of-the-day&type=viz-of-the-day>

