DATA VISUALIZATION 101
The Art & Science of Communicating Data
Holly Goodson Rubio
Director of Institutional Research at Georgia Regents University

http://www.goodgirlgraphics.com

Timeline

Director of Institutional Research (CURRENT)
Georgia Regents University (formerly MCG)

Education

MS'12
Southern Polytechnic State University
Information Design + Communication

ABJ'94
The University of Georgia
Journalism

Job Duties

- Management
- Chief Data Officer
- Research and Analysis
- Information Design
Communicating with Data (History, Why? Benefits)

Data Basics (What? Types? Relationships?)

Find the Story in the Data
  o Identify the Questions (Question, Data, Source)
  o Sources of Data (Original, Comprehensive, Reliable)
  o Analyze the Data (Trends, Correlations, Outliers)
  o Craft the Story (Outline & Arrange Content)
  o Visualize the Data (Charts, Best, Tools, Learn)

Key Takeaways
Data is one of the most powerful tools to communicate a message or support an argument. But those numbers alone can make a bigger impact if your audience can easily interpret them. That’s why data visualization is key.
History of Data Design

William Playfair (inventor of line, bar and pie charts) made this graph in 1821 showing the price of wheat, weekly wages and the reigning monarch for 250 years (1565-1820).

John Snow’s map of Cholera outbreaks in London 1854 led to the discovery of one water source that caused the outbreak.
History of Data Design

Charles Minard’s 1869 chart of Napoleon’s march to Russia (and back)
Why Visualize Data?

OUR BRAINS LOVE VISUAL CONTENT.

Visual content is one of the most powerful communication mediums because our brains process visuals more efficiently than text alone.

Time it takes for our brains to process visual cues: ¼ second

Average consumer attention span: 8 seconds
Why Visualize Data?

Crafted with intent, understanding and a solid framework, visual content succeeds in three core areas of value:

- **APPEAL**
- **COMPREHENSION**
- **RETENTION**

Humans are biologically wired to process the world visually, which is why making content more visual increases its impact and efficacy.
Why Visualize Data?

APPEAL

Visually engaging content is more attractive. Well-designed information piques interest before information is actually processed.

A University of Saskatchewan study surveyed participants who viewed a visually embellished chart found it more attractive and more memorable than a plain chart.
Why Visualize Data?

COMPREHENSION

Our brains are pre-wired to automatically interpret relationships between objects, allowing for instant comprehension with minimal effort.

Notice how quickly your eye finds variations in these attributes.

The brain also interprets visuals faster than it can process the language associated with them. Which sign is easier to process?
Why Visualize Data?

RETENTION

- Our visual system is able to make rapid connections to already-stored information in our memories, helping cement concepts in our brains.
- A design with an aesthetic and emotional appeal also prompts a qualitative connection, making retention more likely.
Benefits of Data Visualization

Data is powerful ammunition, but it needs to be visualized to really hit the target. By presenting your data in a visually engaging way, you enhance your story.

ADDRESSES CREDIBILITY
When you use data to support your claims, you demonstrate that you’ve done your research.

ENCOURAGES ENGAGEMENT
Clean, compelling and credibly sourced visuals draw your audience into the content, encouraging them to explore the data.

INCREASES IMPACT
Letting your audience really see the data—and what it means—increases comprehension and retention, making a more lasting impact on your audience.
Benefits of Data Visualization

WHICH MESSAGE IS STRONGER?

Both infographics attempt to show the amount of taxes paid by various income groups. Which graphic makes the data easier to understand? Which provides more context? Which appears more credible?
Benefits of Data Visualization

This is not data visualization. Information is presented but not visualized, weakening the impact. Not always bad, but relying on only language misses an opportunity for more context.

This graphic is an example of how to present a complex data set in an easy to understand format. Provides deeper insight into the larger story.
**When to Use Data Visualization?**

Data visualization is an effective tool to enhance your research or communication. Whether providing more clarity or supporting your argument, it helps your audience make better-informed decisions.

**WHY TO USE**

- **ANSWER QUESTIONS**
  How many Web visitors last month?

- **ADD OBJECTIVITY TO AN ARGUMENT**
  This is why we should invest.

- **DISCOVER NEW INSIGHTS**
  The majority of visitors clicked the calendar.

- **COMMUNICATE QUICKLY AND EFFECTIVELY**
  Traffic to our website is up.

**WAYS TO USE**

- **MARKETING CONTENT**
  Public-facing content for thought leadership or promotion such as blogs, white papers, infographics

- **REPORTING**
  Data conveying results of performance or research findings presented in a variety of formats.

- **INTERNAL COMMUNICATION**
  Information delivered to leaders that influence planning and decision-making such as reports and dashboards.
THERE ARE TWO TYPES OF DATA:

**QUANTITATIVE**
This is all about the numbers—things that can be measured and counted; all values are numerical (e.g., number of M&Ms in a bag).

**QUALITATIVE**
These are things that you can observe, but you can’t measure (e.g., color of M&Ms in a bag).

**DISCRETE**
Data that is based on counts and can only be certain values, such as whole numbers (e.g., the number of M&Ms in a bag).

**CONTINUOUS**
Data that can take any value in a range and would require infinite specificity to be accurate (e.g., height, time, temperature).
Types of Data

KNOW YOUR DATA

Before understanding visualizations, you must understand the types of data that can be visualized and their relationships to each other.

- **QUANTITATIVE**
  Data that can be counted or measured, all values are numerical. Example: Enrollment

- **CONTINUOUS**
  Data that is measured and has a value within a range. Example: Rainfall in a year.

- **CATEGORICAL**
  Data that can be sorted according to group or category. Example: Types of products sold.

- **DISCRETE**
  Numerical data that has a finite number of possible values. Example: Number of employees in the office.
Types of Data

CROSS SECTIONAL
The sample of elements is measured only once. This shows you a snapshot of variables at a point in time.

LONGITUDINAL
The data sample is measured repeatedly over time.
Data Relationships

**NOMINAL COMPARISON**
Simple comparison of the quantitative values of subcategories.

**TIME-SERIES**
Tracks changes in values of a consistent metric over time.

**CORRELATION**
Two or more variables that may demonstrate a positive or negative correlation to each other.

**RANKING**
How two or more values compare to each other in relative magnitude.
Data Relationships

VARIABILITY/STANDARD DEVIATION
This examines how data points relate to each other, particularly how far any given data point differs from the mean. Example: Cycling on a rainy day vs. a regular day.

MEAN
The sum of all variables divided by the number of variables.

PART-TO-WHOLE
This shows a subset of data compared to the larger whole. This is used to show things like proportion or percentages.

DISTRIBUTION
This shows data distribution, often around a central value.

OUTLIERS
A variable that is an abnormal distance from other variables.

RANGE
The difference between the highest and lowest variables in your data set.

Chart Types for Part-to-Whole

STACKED BAR CHART
STACKED AREA CHART
PIE CHART
TREE MAP
1. IDENTIFY THE QUESTIONS  
2. SOURCES OF DATA  
3. ANALYZE THE DATA  
4. CRAFT THE STORY  
5. VISUALIZE THE DATA  

FINDING THE STORY IN THE DATA
First things first, identify a question you want to answer. You may have an assumption or argument that as you dig into your data, you may find an entirely new or unexpected story.

**WHAT QUESTION ARE YOU LOOKING TO ANSWER?**

**WHAT DATA DO YOU NEED TO ANSWER YOUR QUESTION?**

**WHERE IS THE DATA MOST LIKELY TO BE FOUND?**
Sources of Data

Focus on collecting a data set(s) that will provide a solid framework to help answer your question. A good data source is...

- **ORIGINAL**
  - Stick to primary sources
  - Friends don’t let friends use Wikipedia; check the references
  - Tell the truth with your data

- **COMPREHENSIVE**
  - Questions about your data? Your audience will too!
  - Look at the larger set of data so you won’t miss a trend or another side to the story
Sources of Data

**CURRENT**
- Data changes at an astounding rate; 1 or 2 years old at most
- Always use the most recent version of available data
- Include the date of the data if possible

**RELIABLE**
- Verify the source is relevant, legitimate and as unbiased as possible
- Strong sources include government agencies such as Department of Labor or Census Bureau and academic publications
- Avoid data that is anecdotal or based on hearsay
Analyze the Data

The most powerful communication happens when data shows—not just tells—the story. Find the similarities and differences within your data set by searching for patterns or insights using:

- TRENDS
- CORRELATIONS
- OUTLIERS

Example: Ice cream sales over time
Example: Ice cream sales vs. temperature
Example: Ice cream sales in an unusual region
Trends

Look for consistent patterns and relationships between data points or data sets.
Correlations

Some elements may have an obvious relationship, some will be revealed through data exploration.

**REVENUE, BY PRODUCT FAMILY**
Outliers

Unusual or irregular occurrences often have an intriguing story
Craft the Story

Your data analysis will inform the story you want to tell. Your goal is to shape a concise and coherent narrative around the data to compel your audience to take some sort of action.

- **OUTLINE**
  - Identify the flow of the narrative
  - Map your data to support each point
  - Consider your audience
  - Does the data support the argument?
  - What is the takeaway?

- **ARRANGE CONTENT**
  - Guide the reader through the data in a logical flow
  - Data should help build the narrative as you go
Craft the Story

- **PROVIDE CONTEXT**
  - Data is supposed to clarify, not confuse
  - Provide context for the data

- **HIGHLIGHT NOTABLE INSIGHTS**
  - Key points need attention
  - Map your data to support each point
  - Always consider your audience

- **TRIM WHERE NECESSARY**
  - Don’t overwhelm with data
  - Only include what is necessary and relevant
Visualize the Data

Well-designed data visualization can enhance your story by making data beautiful and easy to understand. Poorly visualized data can be misinterpreted.

WHEN GOOD DATA GOES BAD?
Just because data is visualized doesn’t mean it’s good! Can you spot the mistakes?

See that impressive growth on the left? It’s actually misleading. When visualized with a zero baseline on the right, that growth isn’t quite so impressive.
Guide to Chart Types

- Bar Chart
- Pie Chart
- Line Chart
- Area Chart
- Scatter Plot
- Bubble Chart
- Heat Map
Bar Chart

Bar charts are very versatile. They are best used to show change over time, compare different categories, or compare parts of a whole.

**PAGE VIEWS, BY MONTH**

**CONTENT PUBLISHED, BY CATEGORY**

**VERTICAL (COLUMN CHART)**
Best used for chronological data (time-series should always run left to right), or when visualizing negative values below the x-axis.

**HORIZONTAL**
Best used for data with long category labels.
Bar Chart

MONTHLY TRAFFIC, BY SOURCE

PERCENTAGE OF CONTENT PUBLISHED, BY MONTH

STACKED
Best used when there is a need to compare multiple part-to-whole relationships. These can use discrete or continuous data, oriented either vertically or horizontally.

100% STACKED
Best used when the total value of each category is unimportant and percentage distribution of subcategories is the primary message.
Bar Chart Best Practices

**Use Horizontal Labels**
Avoid steep diagonal or vertical type, as it can be difficult to read.

**Space Bars Appropriately**
Space between bars should be ½ bar width.

**Start the Y-Axis Value at 0**
Starting at a value above zero truncates the bars and doesn’t accurately reflect the full value.

**Use Consistent Colors**
Use one color for bar charts. You may use an accent color to highlight a significant data point.

**Order Data Appropriately**
Order categories alphabetically, sequentially, or by value.
Pie Chart

Pie charts are best used for making part-to-whole comparisons. They are most impactful with a small data set limiting to 5 slices preferably.

**STANDARD**
Used to show part-to-whole relationships.

**DONUT**
Stylistic variation that enables the inclusion of a total value or design element in the center.
Pie Chart Best Practices

**Visualize No More Than 5 Categories Per Chart**
It is difficult to differentiate between small values; depicting too many slices decreases the impact of the visualization. If needed, you can group smaller values into an "other" or "miscellaneous" category, but make sure it does not hide interesting or significant information.

**Order Slices Correctly**
There are two ways to order sections, both of which are meant to aid comprehension:

- **Option 1**: Place the largest section at 12 o’clock, going clockwise. Place the second largest section at 12 o’clock, going counterclockwise. The remaining sections can be placed below, continuing counterclockwise.
- **Option 2**: Start the largest section at 12 o’clock, going clockwise. Place remaining sections in descending order, going clockwise.

**Don’t Use Multiple Pie Charts For Comparison**
Slice sizes are very difficult to compare side-by-side. Use a stacked bar chart instead.

**Make Sure All Data Adds Up To 100%**
Verify that values total 100% and that pie slices are sized proportionate to their corresponding value.
Line Chart

Line charts are used to show time-series relationships with continuous data.

DIRECT MARKETING VIEWS, BY DATE
**Line Chart Best Practices**

**INCLUDE A ZERO BASELINE IF POSSIBLE**
Although a line chart does not have to start at a zero baseline, it should be included if possible. If relatively small fluctuations in data are meaningful (e.g., in stock market data), you may truncate the scale to showcase these variances.

**DON'T PLOT MORE THAN 4 LINES**
If you need to display more, break them out into separate charts for better comparison.

**USE SOLID LINES ONLY**
Dashed and dotted lines can be distracting.

**LABEL THE LINES DIRECTLY**
This lets readers quickly identify lines and corresponding labels instead of referencing a legend.

**USE THE RIGHT HEIGHT**
Plot all data points so that the line chart takes up approximately two-thirds of the y-axis' total scale.
Area Chart

Area charts depict a time-series relationship, but they are different than line charts in that they can represent volume.

**Area Chart**
Best used to show or compare a quantitative progression over time.

**Stacked Area**
Best used to visualize part-to-whole relationships, helping show how each category contributes to the cumulative total.

**100% Stacked Area**
Best used to show distribution of categories as part of a whole, where the cumulative total is unimportant.
Area Chart Best Practices

DON'T DISPLAY MORE THAN 4 DATA CATEGORIES
Too many will result in a cluttered visual that is difficult to decipher.

USE TRANSPARENT COLORS
In standard area charts, ensure data isn’t obscured in the background by ordering thoughtfully and using transparency.

DON'T USE AREA CHARTS TO DISPLAY DISCRETE DATA
The connected lines imply intermediate values, which only exist with continuous data.

MAKE IT EASY TO READ
In stacked area charts, arrange data to position categories with highly variable data on the top of the chart and low variability on the bottom.

START Y-AXIS VALUE AT 0
Starting the axis above zero truncates the visualization of values.
Scatter Plot

Scatter plots show the relationship between items based on two sets of variables. They are best used to show correlation in a large amount of data.

REVENUE, BY PRODUCT FAMILY

![Scatter Plot Diagram](image-url)
Scatter Plot Best Practices

START Y-AXIS VALUE AT 0
Starting the axis above zero truncates the visualization of values.

INCLUDE MORE VARIABLES
Use size and dot color to encode additional data variables.

USE TREND LINES
These help draw correlation between the variables to show trends.

DON'T COMPARE MORE THAN 2 TREND LINES
Too many lines make data difficult to interpret.
Bubble charts are good for displaying nominal comparisons or ranking relationships.

**Bubble Plot**
This is a scatter plot with bubbles, best used to display an additional variable.

**Make Sure Labels Are Visible**
All labels should be unobstructed and easily identified with the corresponding bubble.

**Size Bubbles Appropriately**
Bubbles should be scaled according to area, not diameter.

**Don’t Use Odd Shapes**
Avoid adding too much detail or using shapes that are not entirely circular; this can lead to interpretation issues.
Heat Map

Heat maps display categorical data, using intensity of color to represent values of geographic areas or data tables.

STATES WITH NEW SERVICE CONTRACTS
Heat Map Best Practices

**USE A SIMPLE MAP OUTLINE**
These lines are meant to frame the data, not distract.

**SELECT COLORS APPROPRIATELY**
Some colors stand out more than others, giving unnecessary weight to that data. Instead, use a single color with varying shade or a spectrum between two analogous colors to show intensity. Also remember to intuitively code color intensity according to values.

**USE PATTERNS SPARINGLY**
A pattern overlay that indicates a second variable is acceptable, but using multiple is overwhelming and distracting.

**CHOOSE APPROPRIATE DATA RANGES**
Select 3-5 numerical ranges that enable fairly even distribution of data between them. Use +/- signs to extend high and low ranges.
The main purpose of data visualization is to provide a visual reference for comparing two or more values. Be sure that you choose the right chart to maximize meaning.
Data Design Best Practices

VISUALIZE DATA IN A WAY THAT IS EASY FOR YOUR AUDIENCE TO COMPARE VALUES

Including data adds strength to your presentation, but too much data can be overwhelming and confusing.
Data Design Best Practices

ACCURATELY PRESENT THE DATA.

Don’t skew data with design. Be wary of visualizations that can actually deceive the eye, like 3D charts or odd shapes. Don’t distract from the story.
Data Design Best Practices

ORDER CATEGORIES LOGICALLY

Intuitive ordering helps aid comprehension by shortening the time it takes to become oriented. Arrange in descending order.
Data Design Best Practices

USE COLOR, ICONOGRAPHY, AND IMAGERY APPROPRIATELY

Back up your numbers with powerful images to make your audience feel your message.
**Data Design Best Practices**

**DON’T DISTRACT FROM THE DATA WITH CHART JUNK**

Unnecessary elements can detract from the data’s message. Avoid illustrations, drop shadows, or ornamentations.
Data Design Do’s and Don’ts

1. DO use one color to represent each category.
2. DO order data sets using logical heirarchy.
3. DO use callouts to highlight important or interesting information.
4. DO visualize data in a way that is easy for readers to compare values.
5. DO use icons to enhance comprehension and reduce unnecessary labeling.
6. DON'T use high contrast color combinations such as red/green or blue/yellow.
7. DON'T use 3D charts. They can skew perception of the visualization.
8. DON'T add chart junk. Unnecessary illustrations, drop shadows, or ornamentations distract from the data.
9. DON'T use more than 6 colors in a single layout.
10. DON'T use distracting fonts or elements (such as bold, italic, or underlined text).
Data Tools

Tableau – the greatest visualization tool of all time! (tableausoftware.com)
Data Tools

Microsoft Excel Power View
Data Viz Inspiration

Hans Rosling’s Gapminder – gapminder.org
Data Viz Inspiration

Visually Community – visual.ly
Data Viz Inspiration

“Show Me the Numbers is not just a book on statistical graphics. It is the book on statistical graphics ...A true masterpiece.”

--Alberto Cairo, author of The Functional Art
Data Viz Inspiration

“A commonality between Science and Art is trying to see profoundly – to develop strategies of seeing and showing.”

--Edward Tufte, Information Design Giant
Look for opportunities to enhance all your communication, whether internal or external, with data visualization.

Start with your question, then look for the data that can help answer or give you more insight into your problem.

Work with a solid data set that comes from a relevant source.

Craft a cohesive narrative around your key data point to deliver your message.

Follow data visualization best practices to ensure your information makes the best impact.

KEY TAKEAWAYS
Presentation Sources

- Thanks to my heroes Edward Tufte and Stephen Few for changing my life and career
- Thanks to Visage for some of the awesome graphics and content
- Thanks to Dona Wong, editor at the Wall Street Journal and author of *Guide to Information Graphics* for content
- Thanks to Tableau for content, inspiration, promotion of Visual Analytics and for just being the best tool invented since the toaster
Holly Goodson Rubio

Email: hgoodson@gru.edu
Web: http://www.gru.edu/ie/ir/
Twitter: http://twitter.com/hollyvann
Linked in: http://linkedin/hollyvann

THANK YOU!