UPY BIS universities

UNIVERSIDAD POLITÉCNICA DE YUCATÁN

Visual Modeling for Information

Storytelling

M. Sc. Jorge J. Pedrozo Romero

Data Engineering Program

9^{th} quadrimester September - December (2025)

Chapter 1: The Importance of Context

Core Principle

Success in data visualization starts with understanding context, not with creating visuals.

Exploratory vs. Explanatory Analysis

Exploratory Analysis

- What you do to understand data and find insights
- Like hunting for pearls in oysters (opening 100 oysters to find 2 pearls)
- Internal process of discovery

Explanatory Analysis

- Communicating specific findings to an audience
- Showing only the pearls, not all 100 oysters
- **Focus of effective communication**

Common Mistake: Showing all exploratory work instead of curated insights

The Who-What-How Framework

WHO: Your Audience

Key Principles:

- Be specific about your audience
- Avoid general groups like "all stakeholders"
- Identify the decision maker
- Consider your relationship with the audience
- Establish credibility if needed

Important: Different audiences require different communications

WHAT: Desired Action

Critical Questions:

- What do you need your audience to know or do?
- Why should they care?

Best Practices:

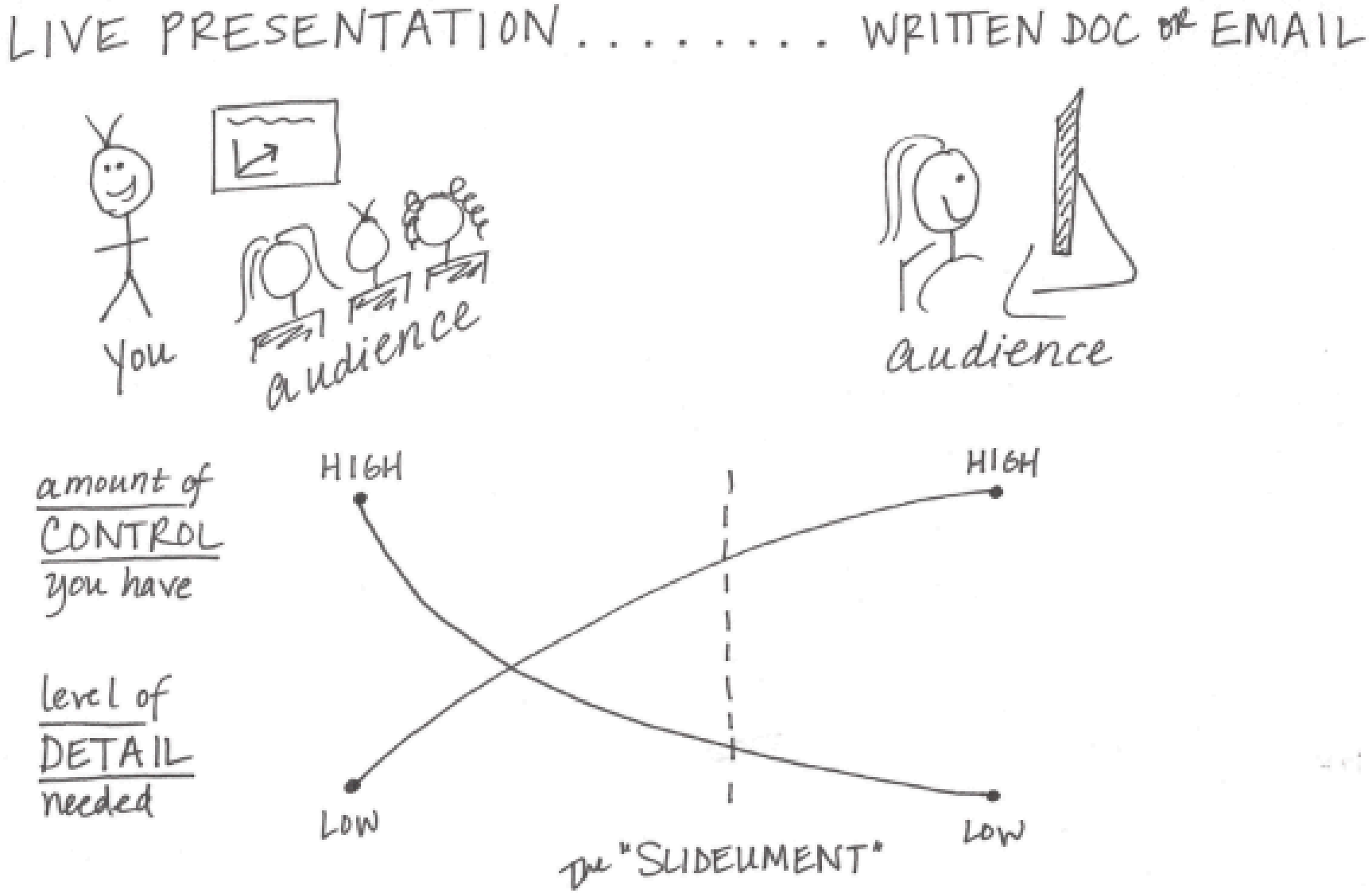
- Take a confident stance with recommendations
- Don't just present data—drive action
- If explicit recommendations aren't appropriate, suggest next steps
- Prompt conversation focused on action

HOW: Data as Evidence

Only after answering WHO and WHAT:

- Determine what data supports your story
- Data becomes supporting evidence, not the main focus
- Include context and opposing data when relevant

Communication mechanism continuum



Live Presentation	← →	Written Document
Full control		Audience control
Sparse detail (you're there to explain)		High detail required
Practice extensively		Address all questions proactively

The "Slideument" Problem: Trying to create one document for both needs—challenging but sometimes necessary

Key Tools for Clarity

The 3-Minute Story

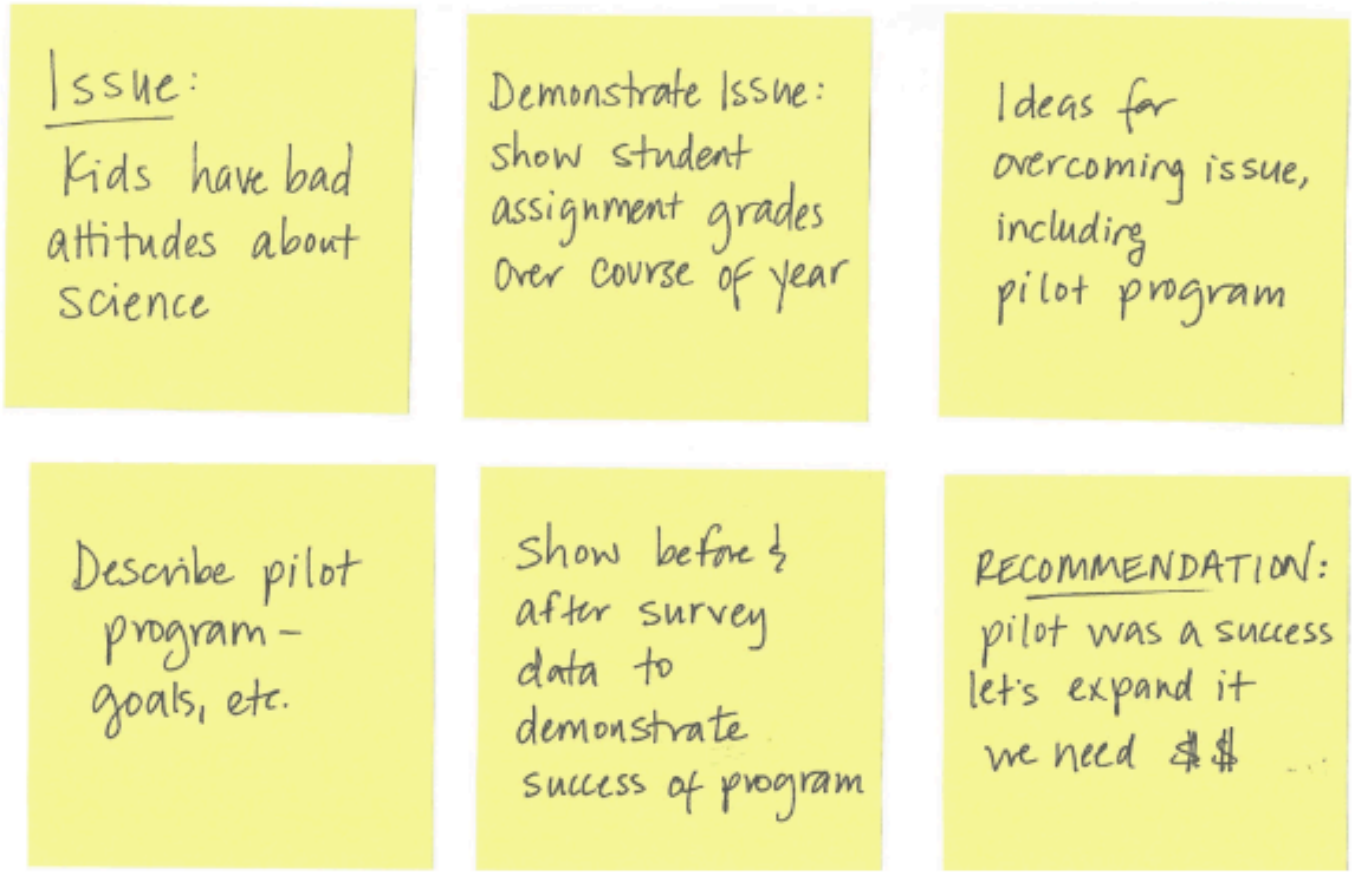
- If you had only 3 minutes, what would you say?
- Ensures you can articulate your story without slides
- Useful for impromptu situations

The Big Idea

Nancy Duarte's concept with three components:

1. Articulates your unique point of view
2. Conveys what's at stake
3. Is a complete sentence

Example



"The pilot summer learning program successfully improved students' perceptions of science; we recommend continuing it and request budget approval."

Storyboarding: The Critical Planning Step

Why It Matters:

- Single most important upfront activity
- Creates visual outline of content
- Establishes structure before detailed work begins

Best Practice:

- Start LOW TECH (whiteboard, Post-it notes, paper)
- Don't start with presentation software
- Avoid premature attachment to content
- Easy to rearrange and iterate

Questions for Consulting on Context

When creating communication for others, ask:

- What background information is essential?
- Who is the decision maker and what do we know about them?
- What biases might they have?
- What data strengthens our case?
- Where are the risks?
- What does success look like?
- **If you had one sentence, what would you say?**

Key Takeaways

1. Understand context before creating visualizations
2. Focus on explanatory, not exploratory analysis
3. Be specific about audience (WHO)
4. Drive toward action (WHAT)
5. Use data as supporting evidence (HOW)
6. Articulate your Big Idea clearly
7. Storyboard before building content
8. Taking time upfront saves time overall

Chapter 2: Choosing an Effective Visual

Core Principle

A handful of visual types will meet the majority of your needs—focus on mastering the basics.

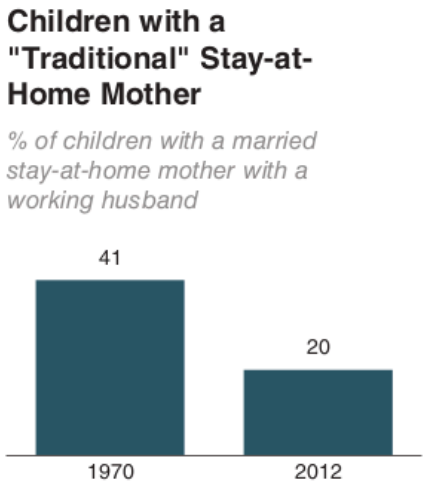
The Most Commonly Used Visuals (91% of use cases)

1. Simple text
2. Tables
3. Heatmaps
4. Scatterplots
5. Line graphs
6. Slopegraphs
7. Vertical bar charts
8. Horizontal bar charts
9. Stacked bar charts (vertical & horizontal)
10. Waterfall charts
11. Square area graphs

Simple Text

When to use: You have just one or two numbers to share

Key principle: Make the number itself prominent—don't hide it in a table or graph



VS

20%

of children had a **traditional stay-at-home mom** in 2012, compared to 41% in 1970

Example

- (**BAD IDEA**) Before: Graph with two data points requiring labels and axes
- (**GOOD IDEA**) After: "20% of children had a traditional stay-at-home mom in 2012, compared to 41% in 1970"

Caution: When reducing multiple numbers to a single metric (like percent change), consider what context may be lost.

Tables

When to Use Tables

- Communicating to mixed audiences (each person looks for their specific row)
- Multiple different units of measure
- When precise values matter more than trends

How People Interact with Tables

- **Verbal system**: We read tables (rows and columns)
- **Physical engagement**: Often use finger to track position

Design Best Practices

- Let design fade into background—data takes center stage
- Use light borders or white space instead of heavy borders
- Avoid heavy shading that competes for attention

Avoid in live presentations: Audience reads silently and you lose their attention. Consider using appendix with references instead.

Heatmaps

What it is: Tabular data visualization using color saturation to show relative magnitude

Advantages:

- Reduces mental processing required
- Uses color saturation as visual cues
- Quickly identifies highest/lowest values
- Eyes and brain target points of interest faster

Critical requirement: Always include a legend to help readers interpret the data

Graphs: General Principles

Why Graphs Work

- **Visual system**: Faster at processing information than verbal system
- Well-designed graphs communicate information more quickly than tables

The Four Main Categories

1. **Points** (Scatterplots)
2. **Lines** (Line graphs, Slopegraphs)
3. **Bars** (Vertical, Horizontal, Stacked, Waterfall)
4. **Area** (Square area graphs)

Points: Scatterplots

When to use: Showing relationship between two variables

How it works:

- Encodes data on both x-axis and y-axis simultaneously
- Reveals whether and what relationship exists between variables

Use case example: Cost per mile vs. miles driven for bus fleet management

Lines

Standard Line Graph

Best for: Continuous data, especially time series

Key requirements:

- Use consistent time intervals on x-axis
- Can show single series, two series, or multiple series
- Physically connected points imply connection between them

Important: Don't mix time intervals (e.g., decades then suddenly years)

Advanced technique: Show average within a range (min, average, max)

Slopegraph

When to use: Two time periods or points of comparison showing relative changes

What it shows:

- Absolute values (the points)
- Rate of change (the slope/direction)
- Increases and decreases across categories

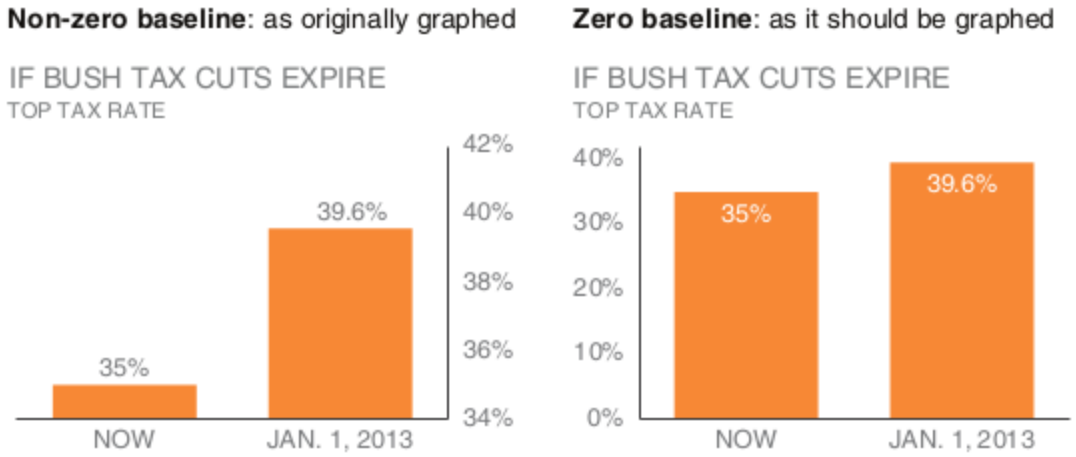
Advantage: Packs in lots of information intuitively—no explanation needed for what "rate of change" means

Bars: The Workhorse of Data Visualization

Why Bars Are Excellent

- Common = less learning curve for audience
- Eyes easily compare endpoints
- Quick identification of largest, smallest, incremental differences
- Should be leveraged because they're common, not avoided

CRITICAL RULE: Bar Charts MUST Have Zero Baseline



Why: Eyes compare relative endpoints—without zero baseline, you get false visual comparisons

Real example: Fox News tax rate chart

- Non-zero baseline showed 460% visual increase
- Zero baseline showed actual 13% increase

Ethics matter: Manipulating scale to reinforce your point is misleading and destroys credibility

Bar Width Guidelines

- Bars should be wider than white space between them
- Not so wide that audience compares areas instead of lengths
- Think "Goldilocks": not too thin, not too thick, just right

Bar Chart Varieties

1. Vertical Bar Chart (Column Chart)

- Can be single, two, or multiple series
- Visual grouping happens based on spacing
- Order of categorization matters for comparison

2. Stacked Vertical Bar Chart

Use for: Comparing totals across categories + seeing subcomponents

Limitation: Hard to compare subcomponents beyond the bottom series (inconsistent baseline)

Formats:

- Absolute numbers
- 100% stacked (each column sums to 100%)

3. Waterfall Chart

Use for:

- Pulling apart stacked bar components one at a time
- Showing starting point → changes → ending point

Structure: Beginning value + Additions - Deductions = Ending value

4. Horizontal Bar Chart

The author's favorite for categorical data

Why it's superior:

- Extremely easy to read
- Perfect for long category names (text reads left-to-right)
- Natural reading pattern: category names before data (no eye darting)
- Follows natural z-pattern of information processing

Logical ordering principle: Be thoughtful about category order

- Use natural ordering when available (age groups, time periods)
- Otherwise order by data values (descending/ascending)
- Put most important category at top

5. Stacked Horizontal Bar Chart

Best for: Survey data on Likert scales (Strongly Disagree → Strongly Agree)

Advantage: Consistent baselines on both left and right allow easy comparison

Area: Square Area Graphs

When to use: Numbers of vastly different magnitudes

Why: Two dimensions (height + width) allow more compact display than bars

Example: Interview funnel (100 phone screens → 25 onsite interviews → 9 offers)

General rule: Avoid most area graphs—human eyes don't accurately attribute quantitative value to 2D space

Axis Labels vs. Data Labels

Decision framework:

- Want big-picture trends? → Preserve axis (make it grey)
- Specific values important? → Label data points directly and omit axis
- Avoid redundant information

To Be AVOIDED

1. Pie Charts Are Evil

Problems:

- Human eyes can't ascribe quantitative value to 2D space
- Hard to tell which segment is bigger when sizes are close
- Can't judge "by how much" one is bigger
- 3D and perspective make interpretation even worse

Solution: Use horizontal bar charts instead

- Easy to see largest segment
- Easy to see incremental differences
- Clear visual comparison with aligned baseline

Special case: If you must show "parts of a whole," show that segments sum to 100%

2. Donut Charts

Problem: Asking audience to compare arc lengths—even worse than comparing angles in pie charts

Solution: Don't use them

3. Never Use 3D

The golden rule: NEVER USE 3D

Only exception: Actually plotting a third dimension (even then, be extremely careful)

Problems 3D causes:

- Skews numbers
- Makes data difficult or impossible to interpret
- Introduces unnecessary elements (side panels, floor panels)
- Graphing applications plot values in confusing ways (invisible tangent planes)

4. Secondary Y-Axis

Problem: Takes time to understand which data reads against which axis

Better alternatives:

1. Label data points directly (no second axis needed)
2. Pull graphs apart vertically with separate y-axes on left
3. If necessary, link axis to data through color (not recommended—color has better uses)

Consideration: Displaying two datasets against same axis can imply relationship that may not exist

Testing Your Visual

The best test: Show it to a friend or colleague

Ask them to articulate:

- Where they focus
- What they see
- What observations they make
- What questions they have

This reveals whether your visual hits the mark or needs changes.

Key Decision Framework

"What is the right graph for my situation?"

Answer: Whatever will be easiest for your audience to read

Process:

1. Clearly articulate what you need audience to know (from Chapter 1)
2. Choose visual display that makes this clear
3. Test with a colleague
4. Iterate based on feedback

Key Takeaways

1. Master the basics—they cover 91% of use cases
2. Simple text for one or two numbers
3. Tables for mixed audiences and precise values
4. Graphs for faster visual processing
5. Horizontal bars are excellent for categorical data
6. Bar charts MUST have zero baseline
7. **NEVER**: pie charts, donut charts, 3D
8. Avoid secondary y-axis
9. Test your visuals with colleagues
10. Choose visuals based on audience needs, not personal preference