

VU LEARNING HUB

vulearninghub.xyz

HIGHLIGHTED HANDOUT

CS630

Data Visualization and Digital
Storytelling

CS630 by VU Learning HUB

Prepared for VU students • Free study resources

<https://vulearninghub.xyz>

CS630

Data Visualization and Digital Storytelling

HANDOUT

Data visualization turns raw data into visual insight, and digital storytelling turns insight into influence. This handout covers visual perception and design principles, chart types and their correct use, tools and techniques for building visualizations, and the narrative frameworks that turn data into compelling stories.

Table of Contents

Module 1	Foundations of Data Visualization
Module 2	Visual Perception and Design Principles
Module 3	Chart Types and Choosing the Right Visual
Module 4	Data Preparation and Storytelling with Data Framework
Module 5	Tools for Data Visualization
Module 6	Digital Storytelling — Narrative Structures
Module 7	Multimedia Elements and Production
Module 8	Case Studies and Evaluation

MODULE 1: FOUNDATIONS OF DATA VISUALIZATION

Learning Objectives

- Define data visualization and explain its purpose
- Describe the data-to-insight pipeline
- Explain why visualization exploits human visual perception
- Distinguish exploratory from explanatory visualization

What is Data Visualization?

Data visualization is the graphical representation of data and information to reveal patterns, trends, and outliers that are invisible in raw tables. Its purpose is not decoration but **cognition**: the human visual system can process enormous amounts of visual information in parallel, so a well-designed chart lets a viewer grasp in seconds what would take minutes to extract from numbers.

The standard pipeline runs: **data acquisition** (collect/clean data) → **mapping** (encode data variables into visual variables like position, length, color, size) → **rendering** (draw the graphic) → **perception and interpretation** (the viewer reads meaning). Every design decision lives in the mapping step, which is why design principles matter so much.

Exploratory vs. Explanatory Analysis

Exploratory visualization is for the analyst: rapid, rough charts to discover what's in the data (histograms, scatter matrices, box plots). **Explanatory** visualization is for an audience: deliberate, polished visuals that communicate a specific finding. Confusing the two is a classic error — showing your exploratory scratch work to decision-makers instead of a focused explanatory story.

A Brief History

Visualization is old: William Playfair invented the bar chart, line chart, and pie chart in the late 1700s; Florence Nightingale's rose diagrams persuaded policymakers about sanitation; John Snow's cholera map of London (1854) traced an outbreak to a water pump. The modern field was shaped by Jacques Bertin's *Semiology of Graphics* (1967), Edward Tufte's work on graphical integrity, and the interactive revolution of the web era (D3.js, Tableau).

Key Terms & Definitions

Data visualization	Graphical encoding of data to reveal patterns and communicate findings.
Visual variables	Position, length, angle, area, color, size, shape — the channels that carry data.
Exploratory analysis	Visualization to discover patterns for oneself.
Explanatory analysis	Visualization to communicate a finding to an audience.

Example: Anscombe's Quartet

Four datasets with identical means, variances, and correlations look the same in a summary table — but their scatter plots reveal completely different structures (a linear trend, a curve, an outlier-driven line, a vertical cluster). Anscombe's quartet (1973) is the canonical proof that you must visualize data, not just summarize it.

Module Summary

- Visualization serves cognition: it lets the visual system detect patterns fast.
- Pipeline: acquire → map data to visual variables → render → interpret.
- Exploratory charts are for discovery; explanatory visuals are for communication.
- Historical roots: Playfair, Nightingale, Snow, Bertin, Tufte.

Exam Pointers (Very Important)

- ★ Define data visualization and state its primary purpose.
- ★ Name the four stages of the visualization pipeline.
- ★ Exploratory vs. explanatory: definitions and an example of each.
- ★ What does Anscombe's quartet demonstrate?

MODULE 2: VISUAL PERCEPTION AND DESIGN PRINCIPLES

Learning Objectives

- Explain pre-attentive processing and its design implications
- Apply Gestalt principles to chart design
- Use color, contrast, and hierarchy effectively
- Avoid common perceptual pitfalls (chartjunk, misleading scales)

Pre-attentive Processing

Certain visual features are processed **pre-attentively** — in under 250 milliseconds, before conscious attention: color, size, orientation, position, and motion. Designers exploit this with **pre-attentive attributes** to make key points pop: highlighting one bar in a contrasting color instantly draws the eye to it. The lesson: use strong visual contrast sparingly and deliberately, so the audience's attention lands exactly where the story needs it.

Gestalt Principles

The **Gestalt principles** describe how viewers group visual elements: **proximity** (near items are seen as related), **similarity** (same color/shape implies grouping), **enclosure** (boxes group items), **closure** (we complete incomplete shapes), **continuity** (aligned elements read as a path), and **connection** (lines link related items). Applied: group related chart elements close together, use consistent color for one category, and align elements to clean grids.

Color, Hierarchy, and Integrity

Use color with intent: **sequential** palettes (light-to-dark single hue) for ordered data, **diverging** palettes for data with a meaningful midpoint, **categorical** palettes (distinct hues) for categories — and always **check colorblind safety** (avoid red-green pairings). Establish **visual hierarchy**: the most important elements get the most visual weight. Edward Tufte's **data-ink ratio** principle says to maximize ink devoted to data and erase **chartjunk** (3-D effects, heavy gridlines, decorative clip art). Above all, preserve **graphical integrity**: never truncate a bar chart's y-axis, never distort with 3-D, never cherry-pick scales — a visualization must not lie.

Key Terms & Definitions

Pre-attentive attribute	Visual feature (color, size) processed in <250ms; used to direct attention.
Gestalt principles	Proximity, similarity, enclosure, closure, continuity, connection — laws of visual grouping.
Data-ink ratio	Tufte's principle: maximize ink used for data, minimize non-data ink.
Chartjunk	Non-data decoration that distracts from the message.

Example: Fixing a Bad Chart

A 3-D exploding pie chart with 12 slices, rainbow colors, and a truncated legend fails on every principle: 3-D distorts angles (integrity), 12 slices exceed working memory and categorical color limits (perception), and the explosion effect is pure chartjunk. The fix: a horizontal bar chart sorted descending, one accent color on the key bar, direct labels instead of a legend, and a clear title stating the takeaway.

Module Summary

- Pre-attentive attributes (color, size) direct attention in milliseconds — use sparingly.
- Gestalt principles govern grouping: proximity, similarity, enclosure, closure, continuity, connection.
- Match palette type to data type; design for colorblind viewers.
- Maximize the data-ink ratio; never mislead with truncated axes or 3-D distortion.

Exam Pointers (Very Important)

- ★ What is pre-attentive processing? Give two design applications.
- ★ State four Gestalt principles with a chart-design example each.
- ★ When do you use sequential, diverging, and categorical palettes?
- ★ What is chartjunk? Give three examples and their fixes.

MODULE 3: CHART TYPES AND CHOOSING THE RIGHT VISUAL

Learning Objectives

- Select the correct chart for a given data relationship
- Explain when to use bars, lines, scatter, histograms, and maps
- Identify misused chart types and their replacements
- Design effective tables

The Core Chart Vocabulary

Bar charts compare quantities across categories — encode with length/position, always start the axis at zero. Use horizontal bars when category labels are long. **Line charts** show trends over continuous time; avoid them for categorical x-axes. **Scatter plots** reveal relationships between two quantitative variables (correlation, clusters, outliers); add trend lines cautiously. **Histograms** show the distribution of one quantitative variable (bin carefully — too few bins hide shape, too many add noise). **Box plots** summarize distributions for comparison across groups (median, quartiles, outliers).

Pie charts are almost always the wrong choice: humans judge angles and areas poorly. Replace with bar charts except for 2-3 part-to-whole shares. **Heatmaps** show patterns in matrices (e.g., correlations).

Maps (choropleth, bubble) add geographic context but require normalized data (rates, not raw counts) to avoid population bias. **Tables** are best when precise values matter — format them well with alignment, banding, and sparklines.

Matching Chart to Question

Choose by the question: comparing categories → bar chart; change over time → line chart; relationship → scatter plot; distribution → histogram/box plot; part-to-whole → stacked bar (or pie for 2-3 parts); geographic → map; precise lookup → table. A useful discipline: write the one-sentence takeaway first, then pick the chart that makes that sentence visually obvious. If you can't state the takeaway, you need exploratory analysis, not an explanatory chart.

Small Multiples and Dashboards

Small multiples — the same chart repeated for each category in a grid — let viewers compare patterns across groups far better than overloaded single charts or 3-D. **Dashboards** combine multiple visuals to monitor a system; effective dashboards follow the same hierarchy principles (key metric prominent, supporting detail subordinate), avoid clutter, and update from live data. Every dashboard element must earn its place by answering a decision question.

Key Terms & Definitions

Bar chart	Category comparison via bar length; axis must start at zero.
Scatter plot	Two-variable relationship display; shows correlation, clusters, outliers.
Small multiples	Grid of identical charts for cross-group comparison.

Choropleth

Map coloring regions by data value; normalize by population/area.

Example: Chart Selection

Question: 'Did our sales grow over the last 12 months?' → line chart (time trend). 'Which region sold most?' → sorted bar chart. 'Do discounts correlate with volume?' → scatter plot. 'What share of revenue is each product?' → 100% stacked bar. 'Where are customers concentrated?' → bubble map with per-capita normalization.

Module Summary

- Bars compare, lines trend, scatters relate, histograms/box plots distribute, maps locate, tables give precision.
- Avoid pie charts (except 2-3 parts) and 3-D effects.
- State the takeaway sentence first, then choose the chart.
- Small multiples beat overloaded charts for comparisons; dashboards must serve decisions.

Exam Pointers (Very Important)

- ★ For each question type, name the best chart.
- ★ Why must bar charts start at zero? Why are pie charts discouraged?
- ★ What is the population bias in maps and how is it fixed?
- ★ Explain small multiples with an example.

MODULE 4: DATA PREPARATION AND STORYTELLING WITH DATA FRAMEWORK

Learning Objectives

- Describe data cleaning steps before visualization
- Apply the storytelling-with-data framework (Cole Nussbaumer Knafllic)
- Declutter visuals and focus attention
- Write effective titles and annotations

Prepare the Data First

No visualization survives bad data. Standard preparation: handle **missing values** (decide: drop, impute, or flag — never silently ignore), fix **data types** (dates as dates, numbers as numbers), remove **duplicates**, standardize **categories** (e.g., 'USA' vs 'U.S.A.'), treat **outliers** deliberately (verify, then keep/annotate or exclude with disclosure), and **aggregate** to the right grain for the question. Document every transformation — reproducibility is part of integrity.

The Storytelling Framework

Cole Nussbaumer Knafllic's *Storytelling with Data* framework: (1) **understand the context** — who is the audience, what do they need to do, how will you communicate; (2) **choose an effective visual**; (3) **declutter** — remove every element that doesn't add information (gridlines, borders, redundant labels, default colors); (4) **focus attention** — use pre-attentive attributes (one accent color, bold, size) to guide the eye through the intended reading order; (5) **think like a designer** — alignment, white space, affordances; (6) **tell a story** — structure with a beginning (plot/setup), middle (tension/conflict in the data), and end (resolution/call to action).

Titles and Annotations

Replace generic titles ('Sales 2024') with **assertive titles** stating the takeaway ('Sales grew 23% after the March campaign'). Use **annotations** to explain anomalies directly on the chart ('dip = supply disruption') rather than forcing viewers to guess. Label data directly where possible instead of legends (reduces eye travel). Order information in the sequence the audience should process it — top-left to bottom-right in most layouts.

Key Terms & Definitions

Decluttering	Removing non-informative visual elements to strengthen the message.
Assertive title	A title stating the conclusion, not just the topic.
Annotation	Direct on-chart explanation of points of interest.
Narrative arc	Beginning-middle-end structure applied to data stories.

Example: From Cluttered to Clear

Before: default Excel column chart, rainbow bars, heavy gridlines, legend, rotated labels, title 'Q1-Q4 Revenue'. After: horizontal bars sorted descending, single grey with one orange accent on Q4, gridlines removed, values labeled at bar ends, assertive title 'Q4 revenue doubled after the pricing change', annotation on Q2's dip ('one-time refund spike'). Same data — completely different comprehension.

Module Summary

- Clean data first: missing values, types, duplicates, categories, outliers, aggregation.
- Framework: context → right visual → declutter → focus attention → design thinking → story.
- Use assertive titles and on-chart annotations; label directly instead of legends.
- Structure the narrative: setup, tension, resolution with a call to action.

Exam Pointers (Very Important)

- ★ List five data-preparation steps before visualizing.
- ★ Name the six lessons of the storytelling-with-data framework.
- ★ What is an assertive title? Rewrite a weak title as assertive.
- ★ How do annotations and direct labeling improve comprehension?

MODULE 5: TOOLS FOR DATA VISUALIZATION

Learning Objectives

- Compare BI tools, code libraries, and design tools
- Describe the strengths of Tableau/Power BI, Python/R, and D3.js
- Choose an appropriate tool for a given task and audience

The Tool Landscape

BI platforms (Tableau, Power BI, Looker) offer drag-and-drop dashboards, fast prototyping, and business-user accessibility; trade-offs are licensing cost and limited customization. **Code libraries** give full control and reproducibility: **Python** (Matplotlib, Seaborn, Plotly, Altair) and **R** (ggplot2, Shiny) dominate data-science workflows; Plotly/Altair add interactivity. **D3.js** (JavaScript) builds bespoke, web-native interactive visualizations — maximum flexibility, steepest learning curve. **Design tools** (Illustrator, Figma, Canva, Datawrapper/Flourish) polish explanatory graphics for publication.

Choosing Tools

Match the tool to the job: quick business dashboard → Power BI/Tableau; reproducible analysis visual → Python/R; custom interactive web story → D3.js or Plotly; infographic for publication → Datawrapper/Flourish plus a design tool. Consider the audience's access (can they run code?), the update frequency (live dashboard vs. one-off), and your team's skills. The tool never substitutes for design thinking — a bad chart in Tableau is still a bad chart.

Interactivity Done Right

Interactivity (filters, tooltips, drill-downs, brushing) lets audiences explore — but every interactive element must serve the story, not showcase technique. Guidelines: make the default view already tell the story (don't hide the insight behind clicks), keep interactions discoverable and consistent, ensure tooltips add context (not raw dumps), and design for the device (touch targets on mobile). **Scrollytelling** — visuals that update as the reader scrolls — is a powerful digital-storytelling format when the narrative drives the interaction.

Key Terms & Definitions

BI platform	Business intelligence tool for dashboards (Tableau, Power BI).
ggplot2 / Matplotlib	Grammar-of-graphics plotting libraries in R and Python.
D3.js	JavaScript library for custom web-based data visualization.
Scrollytelling	Narrative format where visuals change as the reader scrolls.

Module Summary

- BI tools: fast dashboards; code libraries: control and reproducibility; D3.js: bespoke interactivity; design tools: polish.

- Choose by task, audience, update needs, and team skill.
- Interactivity must serve the story; the default view should already communicate the insight.
- Scrollytelling merges narrative flow with interactive visuals.

Exam Pointers (Very Important)

- ★ Compare Tableau, Python/Matplotlib, and D3.js on ease, control, and use case.
- ★ When would you choose Power BI over Python for a project?
- ★ Three guidelines for effective interactivity.

MODULE 6: DIGITAL STORYTELLING — NARRATIVE STRUCTURES

Learning Objectives

- Define digital storytelling and its elements
- Apply narrative arcs (Freytag, hero's journey, in-medias-res) to data stories
- Structure stories for different audiences and formats
- Balance data, narrative, and visuals

What is Digital Storytelling?

Digital storytelling combines narrative with digital media — text, visuals, audio, video, interactivity — to communicate ideas memorably. Stories work because human memory is narrative-shaped: we remember a beginning-middle-end with tension and resolution far better than isolated facts. In data storytelling, the three ingredients are **data** (the evidence), **narrative** (the arc that gives it meaning), and **visuals** (the channel that makes it graspable). Remove any one and the story collapses into either a dry report or an unsubstantiated anecdote.

Narrative Arcs for Data

Freytag's pyramid (exposition → rising action → climax → falling action → resolution) maps neatly onto data stories: setup the context, build tension with the problem in the data, reach the key insight (climax), then resolve with recommendations. The **hero's journey** variant casts the audience or protagonist facing a challenge the data helps overcome. **In-medias-res** opens with the striking finding ('We are losing \$2M a year to churn') then rewinds to explain — powerful for executive audiences with short attention. Nancy Duarte's '**what is / what could be**' structure alternates the painful present with the possible future, ending in a call to action — the classic persuasive pitch arc.

Adapting to Audience and Format

Executives want the **BLUF** (bottom line up front): recommendation first, evidence after. Technical audiences want methodology and caveats. Public audiences want human stakes and plain language. Formats differ: a **live presentation** needs one idea per slide and spoken narrative; a **written report/dashboard** must be self-contained since you won't be there to explain; **video** adds voice and motion; **social media** demands the hook in the first seconds. Always define the single action you want the audience to take — a story without a call to action is entertainment, not communication.

Key Terms & Definitions

Digital storytelling	Narrative combined with digital media to communicate memorably.
Freytag's pyramid	Exposition, rising action, climax, falling action, resolution.
BLUF	Bottom line up front; recommendation-first structure for executives.
Call to action	The specific decision or behavior the story asks of its audience.

Example: Duarte's Arc in a Churn Story

'What is': we lose 18% of subscribers yearly (\$2M). 'What could be': cutting churn to 10% adds \$900K annually. The data shows 60% of churners never used feature X (tension). Climax: onboarding emails lifting feature-X adoption cut churn by a third in the pilot. New bliss: full rollout. Call to action: approve the \$50K rollout budget. — Data, narrative, visuals, and a decision in one arc.

Module Summary

- Data storytelling = data + narrative + visuals; stories are remembered better than facts.
- Arcs: Freytag's pyramid, hero's journey, in-medias-res, Duarte's 'what is / what could be'.
- Adapt structure to audience: BLUF for executives, methodology for technical, human stakes for public.
- Every data story needs a clear call to action.

Exam Pointers (Very Important)

- ★ What are the three ingredients of data storytelling?
- ★ Map Freytag's pyramid onto a data-story example.
- ★ Explain Duarte's 'what is / what could be' structure.
- ★ How does story structure differ for executives vs. technical audiences?

MODULE 7: MULTIMEDIA ELEMENTS AND PRODUCTION

Learning Objectives

- Select appropriate media (images, audio, video, animation) for a story
- Apply basic principles of visual and audio design
- Plan a digital story: storyboard, script, production
- Respect copyright, privacy, and accessibility

Choosing Media

Each medium has strengths: **photography/images** create emotional immediacy and human connection; **video** shows process and change over time; **audio/voiceover** adds personality and guides attention; **animation/motion graphics** explain mechanisms and transitions in data beautifully; **text** carries precise definitions and takeaways. Combine deliberately — media should reinforce the narrative, not decorate it. A common failure is 'PowerPoint karaoke': reading text slides aloud while unrelated stock photos play.

Production Workflow

Professional digital stories follow a pipeline: (1) **define the message and audience**; (2) **research and gather** data, assets, interviews; (3) **script** the narrative (write the words you'd speak — conversational, short sentences); (4) **storyboard** — sketch each scene/shot with its visual, audio, and timing; (5) **produce** — record, design, animate; (6) **edit and review** — cut ruthlessly, check pacing (a scene change every few seconds holds attention), verify facts; (7) **publish and promote** on the right channel with the right format. Storyboarding before production saves enormous rework.

Ethics, Copyright, and Accessibility

Use only media you have rights to (licensed, Creative Commons with attribution, or original); **cite data sources** on every visual. Protect **privacy**: anonymize personal data, get consent for identifiable images/interviews. Design for **accessibility**: captions on video, alt text on images, sufficient color contrast, readable fonts — accessible stories reach wider audiences and are often legally required. Finally, the storyteller's integrity rule: never manipulate media or data to mislead; disclose methods and limitations.

Key Terms & Definitions

Storyboard	Scene-by-scene plan of visuals, audio, and timing before production.
Scrollytelling / motion graphics	Formats blending narrative with animated data visuals.
Alt text / captions	Accessibility essentials for images and video.
Creative Commons	Licensing framework allowing reuse with conditions (e.g., attribution).

Module Summary

- Match medium to purpose: images for emotion, video for process, audio for guidance, animation for mechanism.
- Pipeline: message → research → script → storyboard → produce → edit → publish.
- Respect copyright, cite sources, protect privacy, design for accessibility.
- Never manipulate media or data to mislead.

Exam Pointers (Very Important)

- ★ Name four media types and the storytelling job each does best.
- ★ List the digital-story production pipeline in order.
- ★ What accessibility provisions should every digital story include?
- ★ Why must data sources be cited on visuals?

MODULE 8: CASE STUDIES AND EVALUATION

Learning Objectives

- Analyze celebrated data stories and what makes them work
- Evaluate a visualization against design principles
- Plan measurement of a story's impact
- Avoid manipulation and misinformation pitfalls

What Great Data Stories Do

Celebrated examples share traits: **Hans Rosling's** animated Gapminder talks made global development data joyful and surprising (strong narrative + motion). The **New York Times** interactive graphics (e.g., election needles, 'The Ebb and Flow of Movies') pair rigorous data with elegant interactivity. **Florence Nightingale's** rose diagram changed policy with one image. Common DNA: a **surprising, important insight**; a **clear narrative arc**; **visuals matched to the message**; and a **call to action**. Study them actively: cover the title, guess the story from the visual, then check.

Critiquing Visualizations

Evaluate any visual with a checklist: (1) Is the **takeaway stated** (assertive title)? (2) Is the **chart type appropriate** to the data relationship? (3) Is the **encoding honest** (zero baselines for bars, proportional areas, no 3-D distortion)? (4) Is it **decluttered** with clear hierarchy? (5) Is **color purposeful** and accessible? (6) Are **sources and methods noted**? (7) Does it serve the **audience's decision**? Apply this to your own work before publishing — and to others' work before believing it.

Impact, Manipulation, and Misinformation

Measure impact by the **decision or behavior change**, not just views: did the dashboard change operations? Did the story move the budget decision? Common manipulation tactics to recognize and never use: truncated axes, cherry-picked time ranges, dual axes with mismatched scales, area distortions, and 'chartjunk' that implies false precision. In the misinformation age, the data storyteller's credibility is the asset — **show your data and methods**, acknowledge uncertainty (error bars, confidence intervals), and correct errors publicly. Trust compounds; one manipulated chart destroys it.

Key Terms & Definitions

Gapminder	Rosling's animated bubble charts of global development indicators.
Graphical integrity	Tufte's standard: visuals must not distort the underlying data.
Cherry-picking	Selecting only favorable data ranges; a form of visual dishonesty.
Uncertainty visualization	Showing error bars/confidence intervals; honesty about what data can't say.

Example: Critique Practice

A viral social-media chart shows crime 'exploding' with a y-axis starting at 4,900 instead of 0, making a 2% rise look like a 10x surge. Checklist verdict: fails honesty (truncated axis), fails takeaway (title implies catastrophe the data doesn't support), fails sourcing (no source listed). The ethical redesign: zero baseline, full time range, labeled source — the rise becomes what it is: modest.

Module Summary

- Great data stories: surprising insight + clear arc + matched visuals + call to action.
- Critique checklist: takeaway, chart type, honesty, declutter, color, sourcing, audience fit.
- Measure impact by decisions changed, not views.
- Never manipulate; show methods, uncertainty, and sources — credibility is the asset.

Exam Pointers (Very Important)

- ★ Name two celebrated data stories and what made each effective.
- ★ Apply the 7-point critique checklist to a given chart.
- ★ List four visual manipulation tactics and how to spot each.
- ★ How should a storyteller handle uncertainty and errors?