Storytelling with data story and visualization (Process)

Introduction to Big Data
Index

01.-Storytelling
02.-Envisioning information
   02.01 Escaping the flatland
   02.02 Direct methods
   02.02 micro/macro readings
   02.03 layering and separation
   02.04 Small multiples
   02.05 Color and information
   02.06 Narratives of space and time
03.- Steps
04.- Visualization
   04.01 Definition
   04.02 Choosing a chart
   04.02 Elements of the chart
   04.02 Delivery: infographics and dashboards
01.- Data project process

5 Ps of Big Data

Process: Remember. Cost, plan, work packages, deliverables,.....

It is a R&D project
“Every dataset, every database, every spreadsheet has a story to tell,” says Stuart Frankel, CEO of Narrative Science.

**Story** is meant to capture the key findings, features, or patterns in data, to convey what caused them where possible, and, looking forward, to spell out the implications and recommendations to the organization.

To Stephen Few, “**Data visualization** is the use of visual representations to explore, analyse, and present quantitative data. I view **storytelling** here as an additional interpretative layer, a narrative structure, on top of the data visualization: **Visualization + narrative**
02.- Envisioning Information

Escaping flatland

EVEN though we navigate daily through a perceptual world of three spatial dimensions and reason occasionally about higher dimensional arenas with mathematical ease, the world portrayed on our information displays is caught up in the two-dimensionality of the endless flatlands of paper and video screen.

- What, then, are general strategies for extending the dimensional and informational reach of display flatlands?
- And what specific techniques effectively document and envision multivariate worlds?
- Why are some performances better than others?
02.- Envisioning Information

Escaping flatland
DIRECT methods for the display of three dimensions include making models

Narratives of the universe were impressively cranked up in orreries, simulations of our solar system (as known in 1800), with planets and their satellites rotating and orbiting.
**DIRECT methods:** Particularly intriguing are stereo illustrations, which deliver vivid three-dimensional scenes by means of paired images (one for each eye), which are then fused mentally by viewers. Aerial landscapes, molecular structures, and other worldly objects are commonly portrayed.
Micro / Macro Readings :Arrays micro-details mixing into overall pattern

- At work here is a critical and effective principle of information design: **Panorama**, vista, and prospect deliver to viewers the freedom of choice that derives from an overview, a capacity to compare and sort through detail.
- And that micro-information, like smaller texture in landscape perception, provides a credible refuge where the pace of visualization is condensed, slowed, and personalized.
- Such designs can report immense detail, organizing complexity through multiple and (often) hierarchical layers of contextual reading.
Micro / Macro Readings: Arrays micro-details mixing into overall pattern

- A high-resolution aerial photograph of Senlis, one of the oldest cities in France (construction started on this Notre Dame cathedral in 1153), arrays micro-details mixing into overall pattern
04. Envisioning Information

Micro / Macro Readings: Arrays micro-details mixing into overall pattern

- These multi-layered graphs report a clouded relationship between temperature and conductivity for various elements, as measured by many different laboratories. Each set of connected points comes from a single publication, cited by an identification number.

- In this micro/macro arrangement, 4 layers of data are placed in evidence: individual points measured within each study, connected curves formed by those results, and, finally, an overall conglomeration of curves (which are compared with the standard).
Micro / Macro Readings: Arrays micro-details mixing into overall pattern

• These extraordinary statistical maps report data for thousands of tiny grid squares (1 km on a side). Below, a map of Tokyo shows population density; note smaller concentrations dotting the tracks radiating from the city, as people cluster along rail lines and station stops.
04.- Envisioning Information

Micro / Macro Readings : Arrays micro-details mixing into overall pattern

- **An additional data dimension comes from the ordering of names.** The memorial's designer, Maya Ying Lin, proposed that names be listed by date of death rather than alphabetically.
Micro / Macro Readings: Arrays micro-details mixing into overall pattern

- GRAPHICAL timetables also exemplify the multiplicity and wholeness of micro/macro design
- This graphical timetable shown here governs Japanese high-speed trains, or Shinkansen. Station stops are plotted down the side of the grid; time of day runs across the top; diagonal lines show the space-time path of each train.
Micro / Macro Readings: Arrays micro-details mixing into overall pattern

• All micro/macro designs mentioned have portrayed large quantities of data at high densities, up to thousands of bits per square centimetre and 20 million bits per page

• Such quantities are thoroughly familiar, although hardly noticed:
  • the human eye registers 150 million bits
  • Conventional large-scale topographic maps up to 150 million bits
  • RGB images are composed of three colour channels. An RGB image with 8 bits per pixel has 256 possible values for each channel, which means more than 16 million possible colour values

• High-density designs also allow viewers to select, to narrate, to recast and personalize data for their own uses
Micro / Macro Readings: Arrays micro-details mixing into overall pattern

- What about confusing clutter? Information overload? Doesn't data have to be "boiled down" and "simplified"?
  - These common questions miss the point, for the quantity of detail is an issue completely separate from the difficulty of reading.

- Clutter and confusion are failures of design, not attributes of information. Often the less complex and less subtle the line, the more ambiguous and less interesting is the reading.

False equation

Simpleness of data and design = clarity of reading
Layering and separation

• Among the most powerful devices for reducing noise and enriching the content of displays is the technique of layering and separation, visually stratifying various aspects of the data.

• What matters is the proper relationship among information layers.

• These visual relationships must be in relevant proportion and in harmony to the substance of the ideas, evidence, and data conveyed.

• INFORMATION consists of differences that make a difference. A fruitful method for the enforcement of such differences is to layer and separate data, much as is done on a high-density map.
Layering and separation

- Among the most powerful devices for reducing noise and enriching the content of displays is the technique of layering and separation. What matters is the visually stratifying various aspects of the data.
- Proper relationship among information layers.
- These visual relationships must be in relevant proportion and in harmony to the substance of the ideas, evidence, and data conveyed.
Layering and separation

- Similarly for music notation, some staff paper is better than others
- Dark grid lines are “chartjunk”
Layering and separation

- Usually this involves creating a hierarchy of visual effects, possibly matching an ordering of information content
04.- Envisioning Information

Small multiples

• At the heart of quantitative reasoning is a single question: *Compared to what?* Small multiple designs, multivariate and data bountiful, answer directly by visually enforcing comparisons of changes, of the differences among objects, of the scope of alternatives

• IN this splendid 1659 drawing by Christiaan Huygens, the inner ellipse traces Earth's yearly journey around the Sun; the larger ellipse shows Saturn's orbit, viewed from the heavens. The outermost images depict Saturn as seen through telescopes located on Earth
04.- Envisioning Information

Small multiples

- Simultaneous two-dimensional indexing of the multiplied image, **flatland within flatland**, significantly deepens displays, with little added complication in reading.

- These neurometric maps record distributions of brain electrical activity, arraying data over a matrix of color images — with frequency bands (delta, theta, alpha, and beta) sorting the columns, and individual diagnosis forming the rows.
04.- Envisioning Information

Colour and information

- IN representing and communicating information, how are we to benefit from colour's great dominion?

- Human eyes are exquisitely sensitive to colour variations: a trained colourist can distinguish among 1,000,000 colours at least when tested under contrived conditions of pairwise comparison

  - Some 20,000 colors are accessible to many viewers

- For encoding abstract information, however, more than 20 or 30 colors frequently produce not diminishing but negative returns

  - First rule: Pure, bright or very strong colours have loud, unbearable effects when they stand unrelieved over large areas adjacent to each other, but extraordinary effects can be achieved when they are used sparingly on or between dull background tones. "Noise is not music“

  - Second rule: The placing of light, bright colours mixed with white next to each other usually produces unpleasant results, especially if the colors are used for large area
04.- Envisioning Information

Color and information

- Bad example
Narratives of space and time

- MANY information displays report on the world's workaday reality of three-space and time.
- Painting four-variable narrations of space-time on to flatland combines two familiar designs, the map and the time-series

- A comprehensive narrative description of a transport system requires a record of both time and spatial experiences.
03.- Steps

- **What are you trying to achieve?**
  - What is your objective?
  - Why are you putting this presentation or report together?
  - What do you hope to achieve?

- **Who is your audience?**
  - How data literate are they?
  - How technical are they?
  - What expectations do they have?
  - What’s their level of interest and motivation? How busy are they?

- **What’s your medium?**
  - should you use a written report; a visual presentation, such as PowerPoint; a dashboard; or an infographic;
To Stephen Few, “Data visualization is the use of visual representations to explore, analyse, and present quantitative data. I view storytelling here as an additional interpretative layer, a narrative structure, on top of the data visualization: Visualization + narrative.

Assuming that an analyst has chosen appropriate metrics, appropriate dimensions (i.e., how the data has been sliced, say by month or by sales channel) and has found some interesting and significant pattern in those data, he next needs to select a way to present the data. In some cases, that might be a data table, but typically it will involve a chart.
Choosing a chart

- Selecting a chart type is central in terms of the ability to get the story across to the viewer
- **Four types of reason for that chart:**
  - **Comparisons:** For example, how cohorts compare, how something changes with time
  - **Distribution:** To show the variability of a set of data
  - **Relationships:** To show correlations or associations among variables
  - **Category composition:** To show how data is divided among two or more categories
Choosing a chart

Chart Suggestions—A Thought-Starter

Comparison

Relationship

What would you like to show?

Distribution

Composition

Changing Over Time

Static

Few Periods

Many Periods

Only Relative Differences Matter

Relative and Absolute Differences Matter

Simple Share of Total

Accumulation or Subtraction to Total

Components of Components

Stacked 100% Column Chart

Stacked Column Chart

Stacked 100% Area Chart

Stacked Area Chart

Pie Chart

Waterfall Chart

Stacked 100% Column Chart with Subcomponents
04.- Visualization

Elements of the chart

Text
- 6-12 word descriptive title
- Subtitle and/or annotations provide additional information.
- Text size is hierarchical and readable.
- Text is horizontal.
- Data are labeled directly.
- Labels are used sparingly.

Arrangement
- Proportions are accurate.
- Data are intentionally ordered.
- Axis intervals are equidistant.
- Graph is two dimensional.
- Display is free from decoration.

Color
- Color scheme is intentional.
- Color is used to highlight key pattern
- Color is legible when printed in black
- Color is legible for people with colorblindness
- Text sufficiently contrasts background

Lines
- Gridlines, if present, are muted.
- Graph does not have border line.
- Axes do not have unnecessary tick marks.
- Graph has one horizontal and one vertical line.

Overall
- Graph highlights things
- The type of graph is appropriate for data
- Graph has appropriate level of precision
- Contextualized or comparison data
- Individual chart elements work together to ensure overarching takeaway message
Focusing the message

The purpose of creating a visual is to convey a message clearly.
- You have a number of “weapons” in the arsenal, such as fonts, gridlines and orientation. Another is the use of highlight colours. One approach to focus the message and reader is to show only the data of interest

Edward Tufte coined the term chartjunk to cover these visual distractions: “Chartjunk refers to all visual elements in charts and graphs that are not necessary to comprehend the information represented on the graph, or that distract the viewer from this information

![Chartjunk Gridlines](chartjunk_gridlines.png)
In a data-driven business context, I’m not a huge fan of infographics, meant in the modern sense of flashy, colourful visuals peppered with a few facts, typically created by designers rather than analysts.”

Thus, the infographic format worked well in that situation, and it can also be useful to communicate externally to the public.
Delivery

- **Dashboards**
  Many organizations mistakenly measure their data-drivenness by the number of reports and dashboards they produce.
- **Dashboards are very useful and can support a range of activities**

We can think of dashboards as falling into **three categories**:

- **Executive or strategic**
- **Analytical**
- **Operational**