Big Data for Labour Market Intelligence

Capacity development programme 2022

Module 1: Technical training

Session 10

Use of raw data from ETF databases: procedures and technical aspects to create new dashboards.

Speaker: Anna Clara Gatti

29/11/2022

European Training Foundation

Challenges of Data Presentation
Key points

Which Users and navigation patterns?

How data and users interact with the context?

How data must be integrated?

How to manage data governance?

Presentation Area

Presentation Area

- Data Discovery Capabilities
- Publish, Share and Collaborate
- Self-service analytics and BI
- Embedded Advanced Analytics
- Visual-Based Investigation
- Systems of Insight
- Spatial location intelligence
- Graph Discovery and investigation
- Data storytelling

Evolution

Augmented analytics

Augmented analytics is the next wave of disruption in the data and analytics market. It uses machine learning (ML) and AI techniques to transform how analytics content is developed, consumed and shared.

Continuous intelligence

Continuous intelligence is a design pattern in which real-time analytics are integrated within a business operation, processing current and historical data to prescribe actions in response to events.

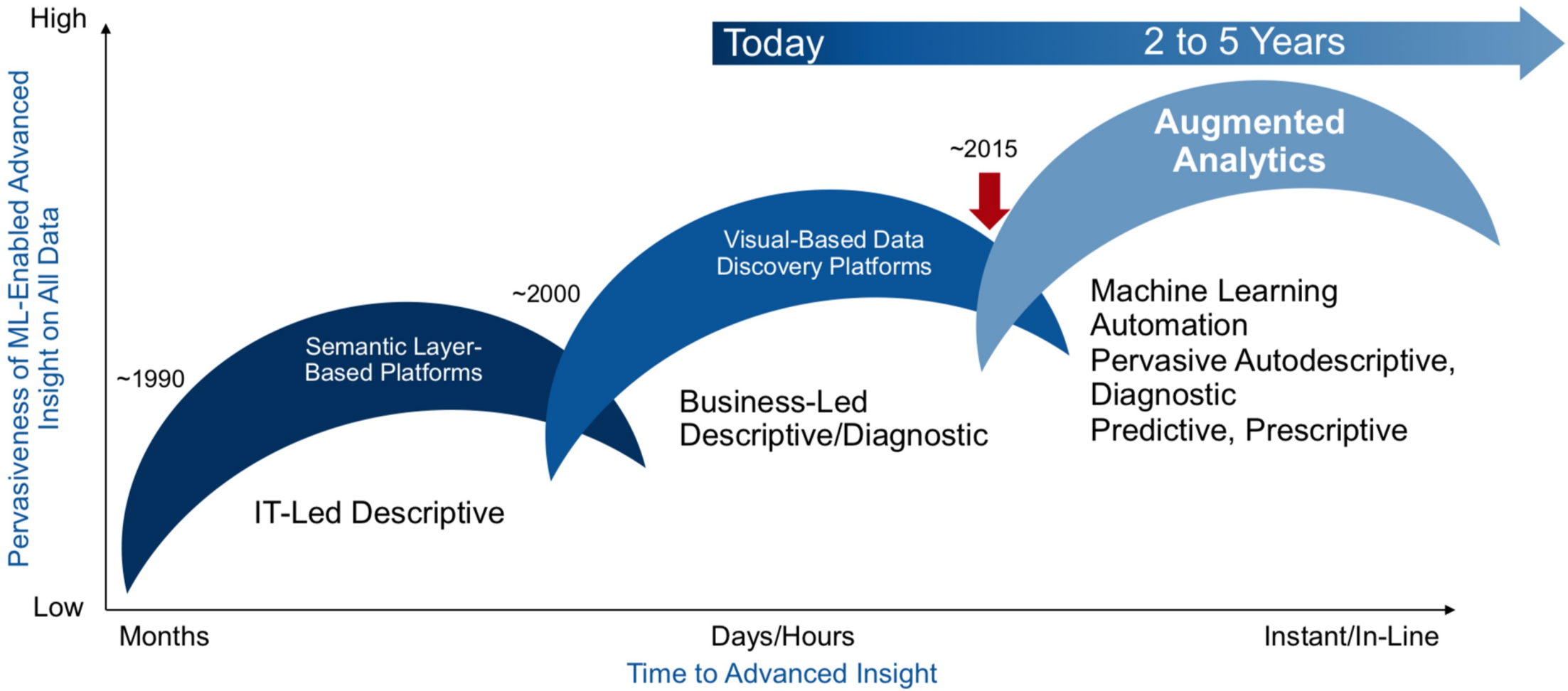
Explainable AI

Explainable AI in data science and ML platforms, auto-generates an explanation of models in terms of accuracy, attributes, model statistics and features in natural language.

Blockchain

The core value proposition of blockchain, and distributed ledger technologies, is providing decentralized trust across a network of untrusted participants. The potential ramifications for analytics use cases are significant, especially those leveraging participant relationships and interactions.

The Future of Data & Analytics Is Augmented Analytics



Gartner

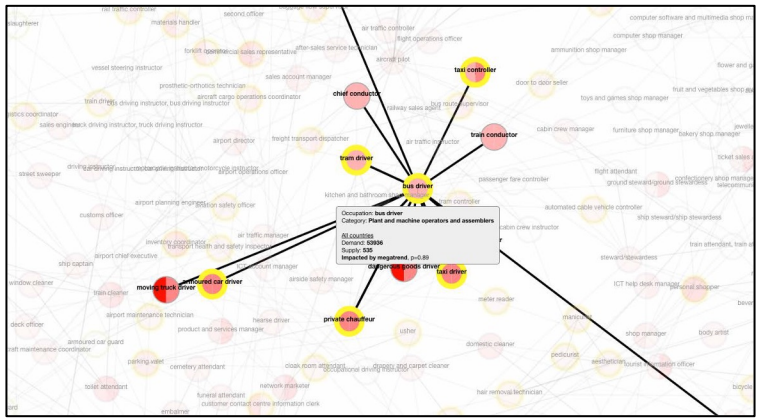
Identify users and navigation patterns

Data Scientists and Analysts

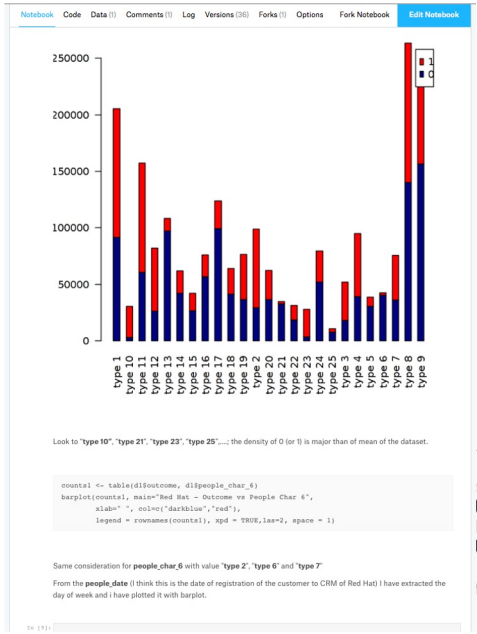
Decision Makers and Business Users

Data Scientists and Analysts

Notebook



An occupation graph (from 2017 Big Data Hackathon – Estonia Team)



Critical Capabilities for Business Intelligence and Analytics Platforms

- Core capabilities from traditional BI vendors have largely caught up to *data discovery* vendors who initially disrupted this market, although differences remain at the subcriteria level and in the degree of excellence exhibited.
- The next wave of disruption in the form of *smart data discovery* has begun, with larger vendors innovating first or acquiring startups.
- Although this is a crowded market, *significant differences remain in functionality*, and in *which products are most appropriate for a given use case*.

Source: Critical Capabilities for Business Intelligence and Analytics Platforms, Gartner, Mar 2018

Evaluated solutions for Decision Makers

Microsoft Power BI

Pentaho

Tableau

Evaluated solutions for Data Scientist

databricks

Magic Quadrant for Analytics and Business Intelligence Platforms

Source: Magic Quadrant for Analytics and Business Intelligence Platforms, Gartner, Feb 2021

ETF Working together Learning for life
European Training Foundation

Gartner on Tableau

- Strengths
  - Intuitive interactive visual exploration and dashboards
  - Number of data sources
  - Mobile support
  - Responsive design
- Areas of Improvement
  - Gaps in enterprise features and collaboration
  - Limited scale and variable performance

Source: Critical Capabilities for Business Intelligence and Analytics Platforms, Gartner, Mar 2018

Time to market

- Depending on the selected platform different time to market should be taken into account
- Start up
  - Tableau is currently in use in the project
    - » No training period
    - » No migration issues
- Development of new solutions
  - Tableau: fast development and release
- Maintenance (corrective and evolutionary)
  - Tableau: fast

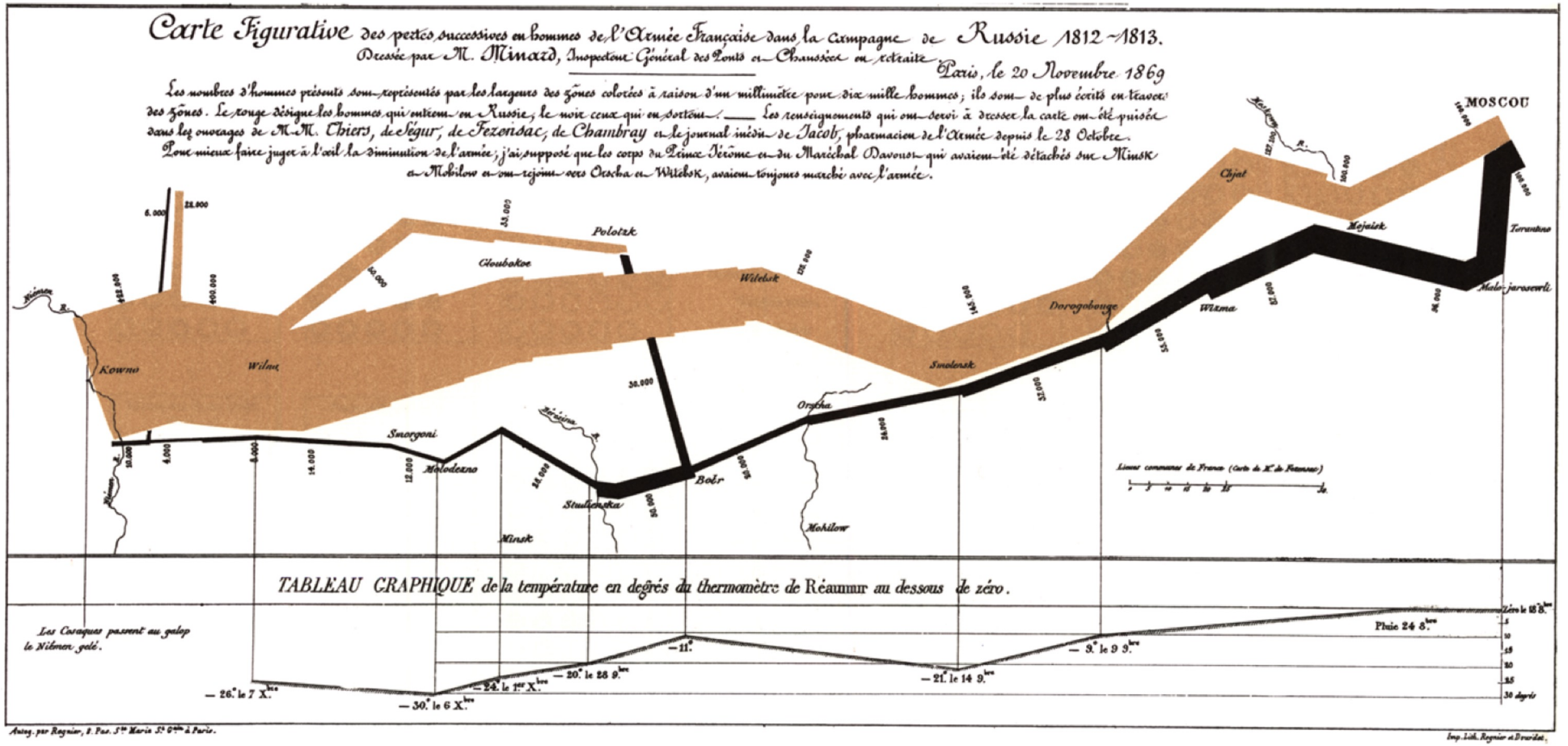
Tableau

- Pros
  - Reference point in data visualization
  - Fast development, update and maintenance of solutions
  - Good integration with data sources
  - Data discovery
- Cons
  - Just introduced integration with external components (but it's the only tool providing it)
  - Variable performance with big volumes of data

Source: Critical Capabilities for Business Intelligence and Analytics Platforms, Gartner, Mar 2018

Visualization is the highest bandwidth channel into the human brain

Data visualisation



Minard (1869)

Visualization is the highest bandwidth channel into the human brain

Data Visualization

1 - Measure

2 - Understand

3 - Describe and strategize

4 - Re-think

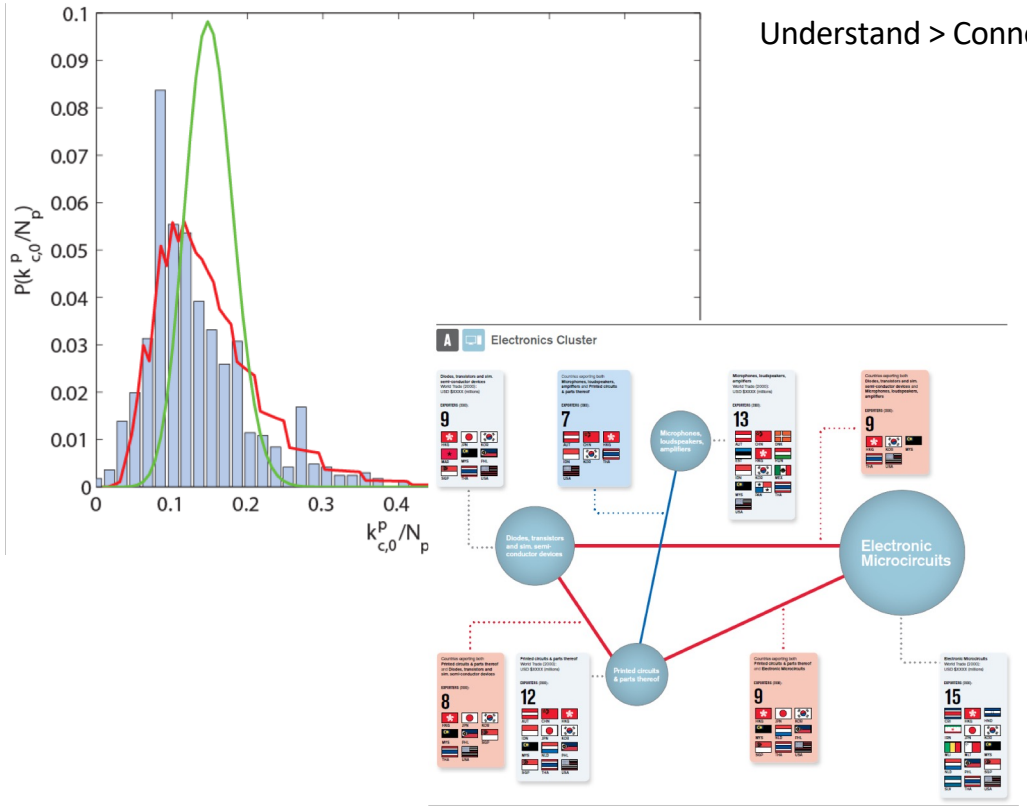
Measure

Understand and measure the complexity and the dimensions of the data

Intuition…

Understand

Demonstrate what dimension is a fundamental determinant of result



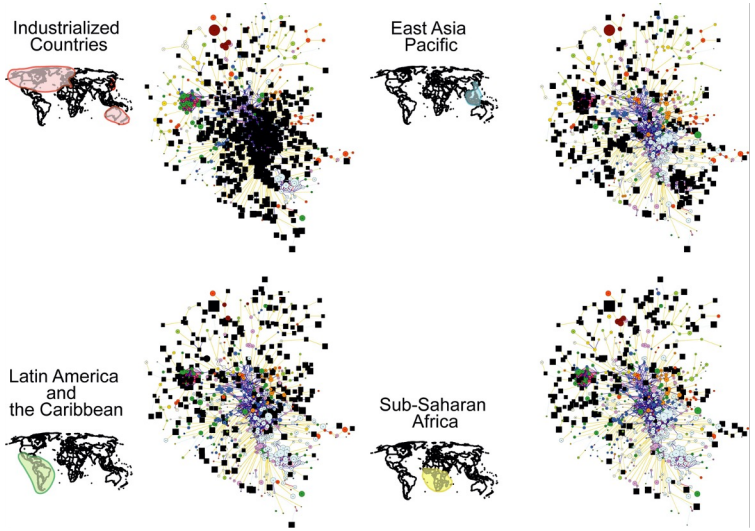
Understand > Connect > Theorize

ETF Working together Learning for life

European Training Foundation

Describe

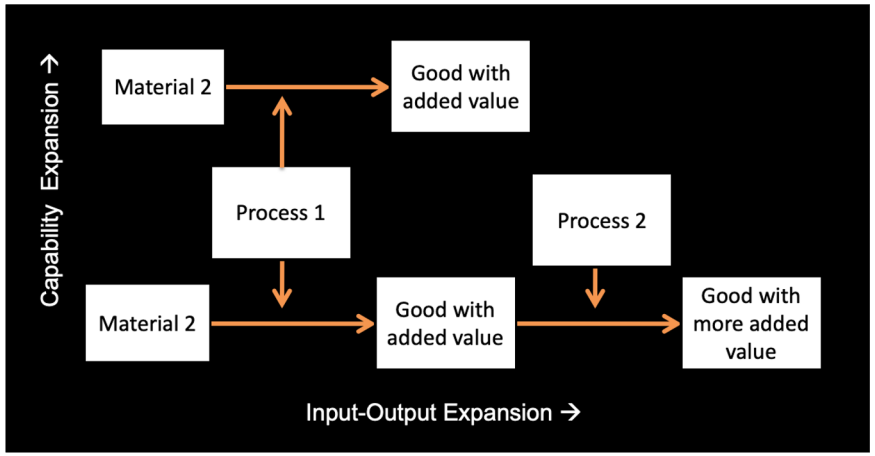
Describe how the result have been developed, what is the relation between the dimension and our theory



Patterns

Re-Think
Chalk in our hands

Rephrase the initial question, what is the relevant patterns? what is the lesson learned and how apply it to improve the context?



Explain

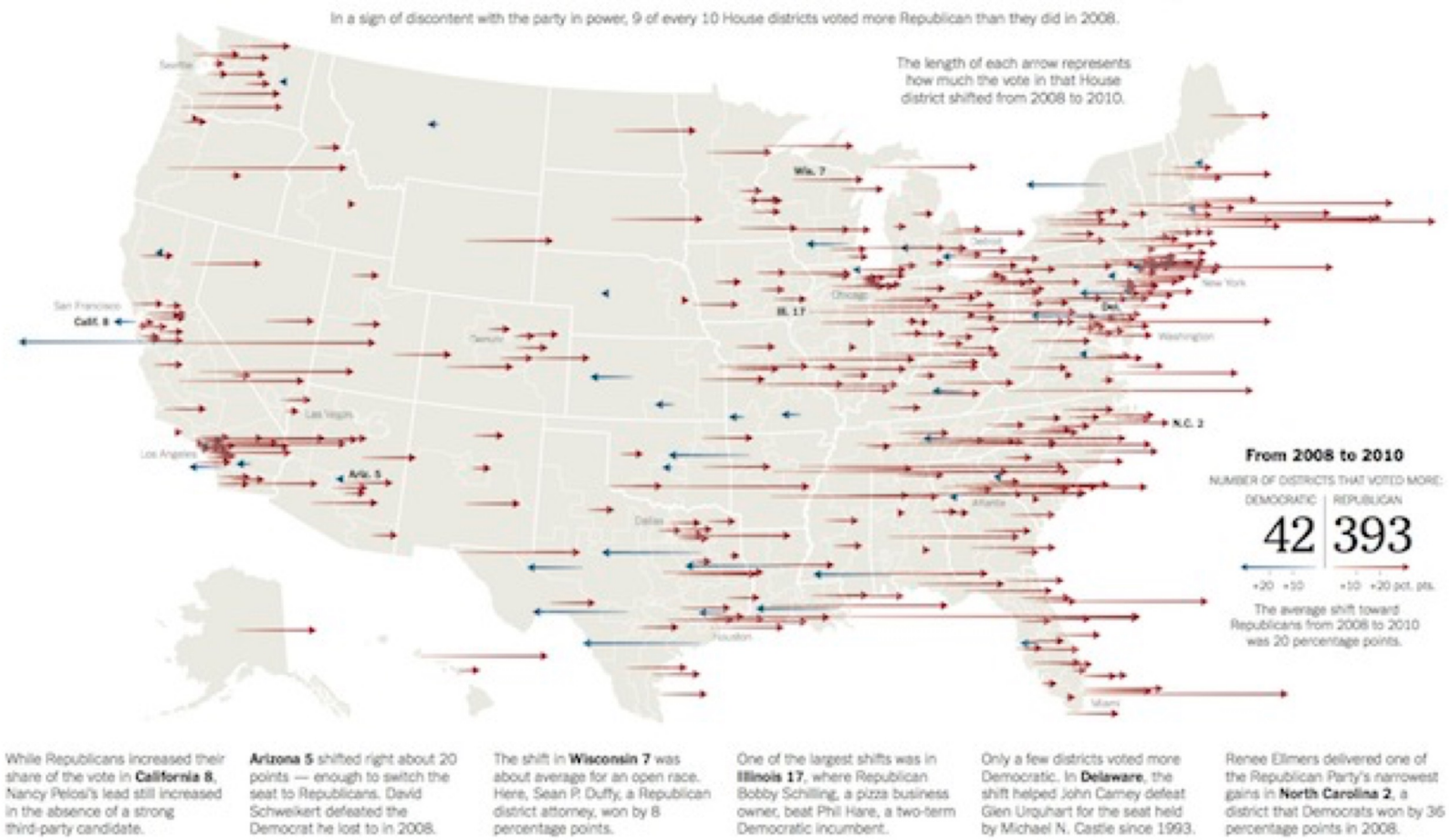
ETF Working together Learning for life
European Training Foundation

LATCH

- Location
- Alphabet
- Time
- Category
- Hierachy

NEW YORK TIMES (2010)

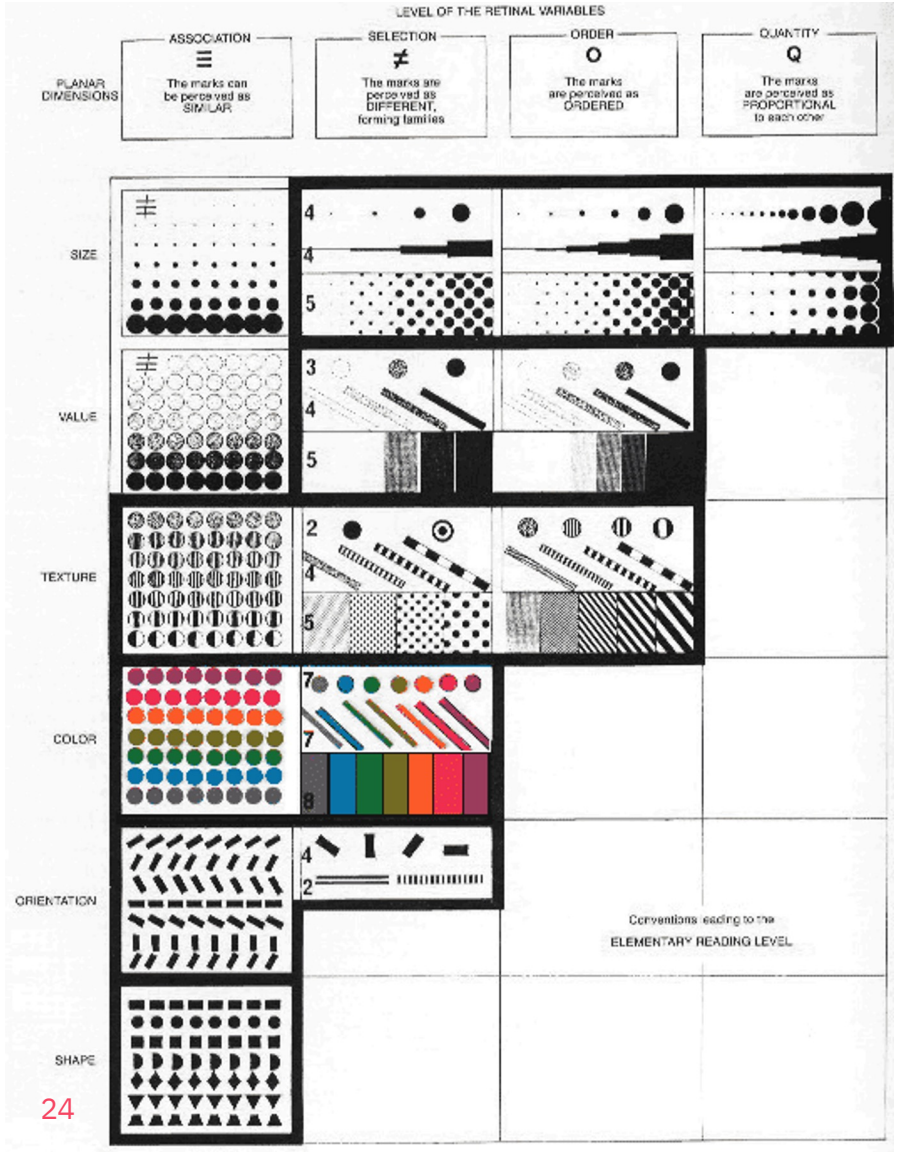
This news graphic by The New York Times shows discontent with the political party in power. Nearly all districts voted more Republican (red arrows) in 2010 than they did in 2008.



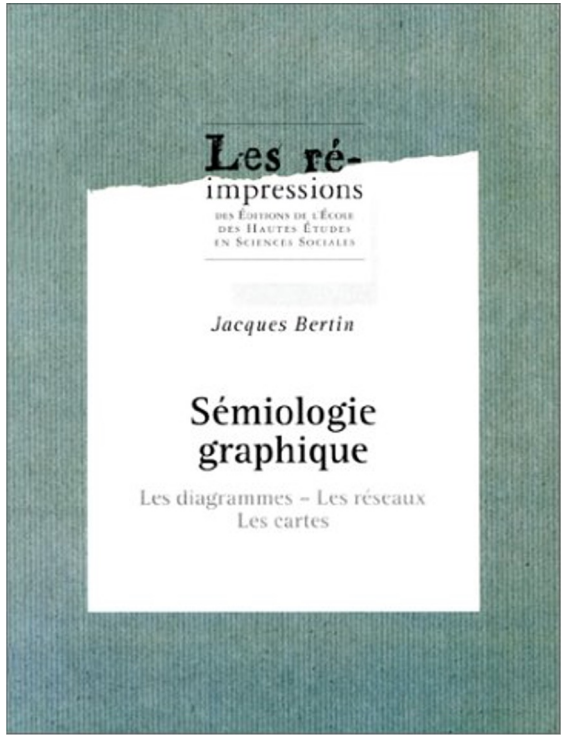
Working together Learning for life

Training Foundation

Semiologie graphique



1) **Size**
2) **Value/Opacity**
3) **Texture**
4) **Color**
5) **Orientation**
6) **Shape**



Bertin, 1967
cartografo

Live session with Tableau (Public) and ETF dataset