

VISUAL ANALYTICS

APPLICATION AREAS

LECTURE 8

Petra Isenberg

TOC

- Business Intelligence
- Legal Matters
- Work at Aviz

ANALYSIS IN THE BUSINESS SECTOR

BUSINESS INTELLIGENCE

“An interactive process for exploring and analyzing structured and domain-specific information to discern trends or patterns, thereby deriving insights and drawing conclusions.

The business intelligence process includes communicating findings and effecting change.”
Gartner

→ there are many more definitions and they don't all say the same thing

BUSINESS INTELLIGENCE

- **Typical Goals:**
 - increase profitability
 - decrease costs
 - improve customer relationship management
 - decrease risks (e.g. credit risk analysis)
- **Main goal: aid in making decisions**

ANALYSIS IN ENTERPRISES IN GENERAL

- business analysts
 - data analysts
 - data scientists
- typical user population for analysis & visualization tools

Sean Kandel, Andreas Paepcke, Joseph M. Hellerstein, Jeffrey Heer: Enterprise Data Analysis and Visualization: An Interview Study. IEEE Trans. Vis. Comput. Graph. 18(12): 2917-2926 (2012)

A STUDY

researchers conducted semi-structured interviews in enterprises:

- 35 analysts (26 male)
- 25 organizations: healthcare, retail, finance, social networking, insurance, ...

Sean Kandel, Andreas Paepcke, Joseph M. Hellerstein, Jeffrey Heer: Enterprise Data Analysis and Visualization: An Interview Study. IEEE Trans. Vis. Comput. Graph. 18(12): 2917-2926 (2012)

QUESTIONS

- What tasks do analysts perform?
- What kinds of data sources and formats do they work with?
- What tools do they regularly use and how do they use them?
- How do analysts vary in terms of programming proficiency?
- How do analysts vary in terms of statistical proficiency?
- What are the “results” of analysis?
- What happens to these results “downstream”?
- What are recurring bottlenecks and pain points?
- How important is scalability?
- How important is sociability?
- What is the relationship between analysts and other business units?
- Where are analysts situated within their corporate hierarchy?

THE ANALYST

Three archetypes found

1) hackers

- most comfortable manipulating data
- used 3+ programming languages (R, Python, SQL, ...)
- complex workflows
- work quite isolated (don't need a lot of help)
- used visualizations: Tableau, Excel, PPT, D3, ...

THE ANALYST

Three archetypes found

2) scripters

- most analysis done in R/Matlab
- not versed in custom operations (parsing, scraping)
- generally worked on data from data warehouse (with help from IT staff)
- applied models to data
- did exploratory analysis with visualization

THE ANALYST

Three archetypes found

3) application user

- performed most operation in spreadsheet or analysis app (SAS/JMP, SPSS, ...)
- needed help preparing data
- typically worked on smaller datasets

ANALYSTS

[illegible]

IT TEAM

- Crucial role in process
 - getting, maintaining, accessing data
 - operationalize workflows
 - provide documentation

OTHER FINDINGS

- data:
 - stored in variety of repositories and formats
- consumers of analysis:
 - many different departments, also other analysts
 - static reports shared most commonly, sometimes dynamic dashboards

OTHER FINDINGS

- collaboration
 - work on analysis task mostly done individually
 - resources shared, however: data, scripts, results, documentation

REPORTED CHALLENGES

- discovery
 - where is my data?
 - what does my data mean? (unclear field names, missing units, timezones, ...)

REPORTED CHALLENGES

- wrangling
 - processing semi-structured data
 - data integration from multiple sources
 - advanced aggregation and filtering

I spend more than half of my time integrating, cleansing and transforming data without doing any actual analysis. Most of the time I'm lucky if I get to do any analysis. Most of the time once you transform the data you just do an average... the insights can be scarily obvious. It's fun when you get to do something somewhat analytical.

REPORTED CHALLENGES

- profiling
 - assessing & improving data quality
(missing data, wrong formats, ...)
 - checking assumptions
(data distribution, semantics of data, ...)

REPORTED CHALLENGES

- modeling
 - finding the right features to analyze
 - scale of data
 - visualization of statistical models missing

REPORTED CHALLENGES

- reporting
 - how to document assumptions
 - flexibility in reports missing

REPORTED CHALLENGES

- workflows are non-linear
 - moving data between tools necessary (creates formatting issues)
 - creating repeatable, reliable, and scalable workflows

BUSINESS ANALYSIS TOOLS

a research view

TOOLS PREDECESSORS

- Management Information Systems (MIS)
- Management Decision Systems (MDS)
- Decision Support Tools (DSS)
- Executive Information Systems (EIS)
- Analysis Information Systems (AIS)
- ...

TOOLS

Warning: I haven't tried any of them, so can't make recommendations

- Professional
 - SAP Business Intelligence
 - IBM Cognos
 - ...
- Open Source
 - SpagoBI
 - Pentaho
 - ...
- Research
 - see next slides

DOTLINK360

Basole, R.C.; Clear, T.; Mengdie Hu; Mehrotra, H.; Stasko, J., "Understanding Interfirm Relationships in Business Ecosystems with Interactive Visualization," *Visualization and Computer Graphics, IEEE Transactions on* , vol.19, no.12, pp.2526,2535, Dec. 2013
doi: 10.1109/TVCG.2013.209

2013

dotlink360

Understanding Interfirm Relationships in Business Ecosystems with Interactive Visualization



Rahul C. Basole | Trustin Clear | Mengdie Hu | Harshit Mehrotra | John Stasko

contact: basole@gatech.edu



BUSINESS ECOSYSTEM INTELLIGENCE

- market analysts
 - understand competitive trends/strategies/threats/opportunities
- executives
 - identify strategic collaborations & customers, find areas for innovation
- venture capitalists
 - identify investment opportunities, see how they fit in the business landscape

DOTLINK360



DESIGN PROCESS

- field study of analysts
- derive set of design requirements
- develop initial version
- user testing
- redesign

DATA SOURCES

- Thomson Reuters SDC Platinum database
 - commercial database
 - ca. 700,000 global alliances, agreements, joint-ventures, since 1985
- Capital IQ Compustat
 - contains e.g. quarterly financial and accounting data for active and inactive publicly listed companies

DATA TRANSFORMATION

- turned data into a network
 - nodes = companies
 - edges = agreements
 - nodes have multiple attributes
 - edges have multiple types
 - +temporal data
- time-varying multivariate network

DESIGN REQUIREMENTS

- Field study
 - online survey + interviews
 - 24 senior industry individuals (market analysts, executives, venture capitalists)
 - each >10 years of experience

DESIGN REQUIREMENTS

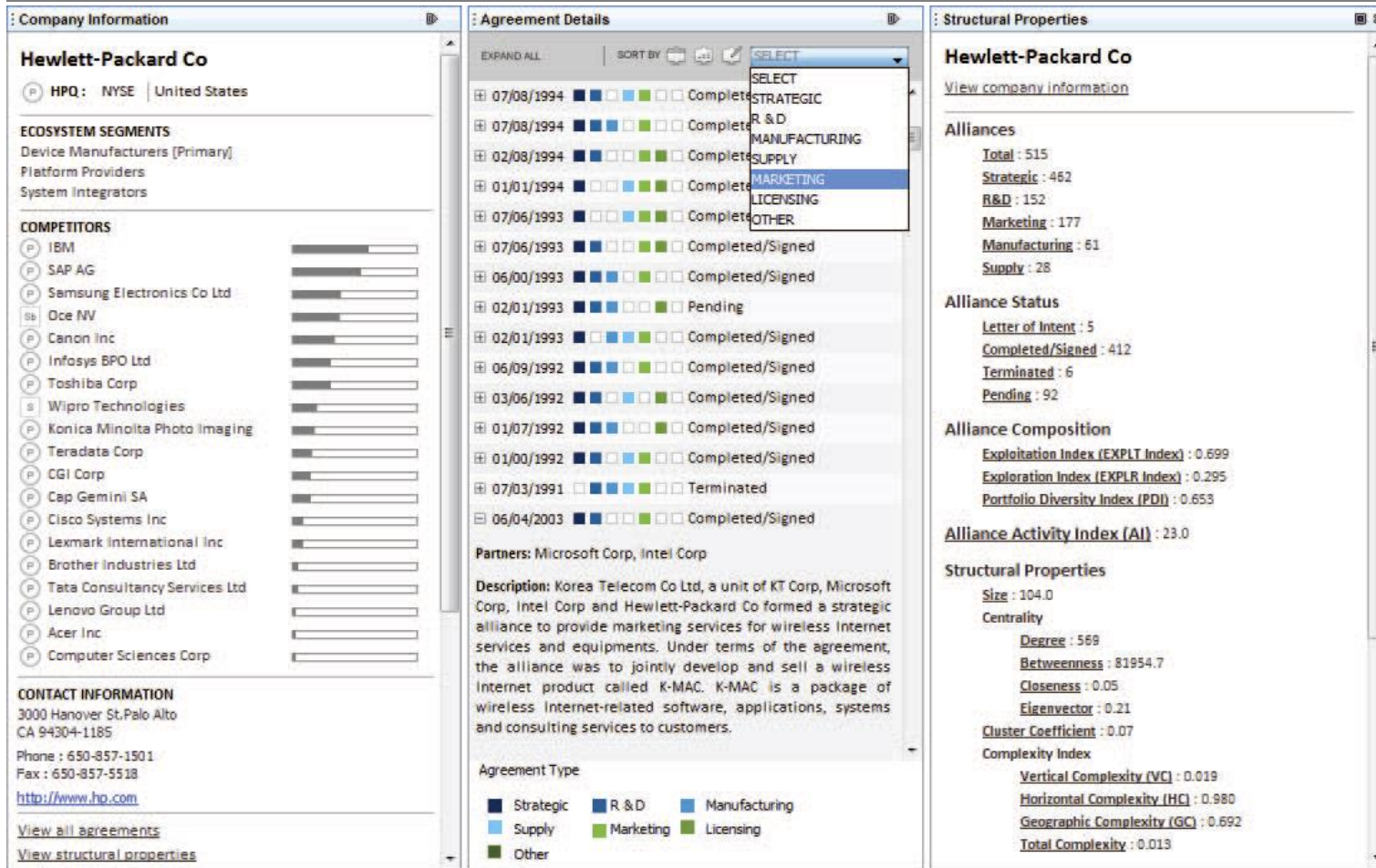
top-down & bottom-up examination of an ecosystem are critical

“it is helpful to have a birds-eye view of the ecosystem, but at the same time [the system] should be able to allow users to drill down into individual companies and segments.”

DESIGN REQUIREMENTS

- understanding interfirm connectivity, composition, and temporality is vital
- comparative perspectives drive insight
- first: communicate agreement summaries (structural information) → then: offer details
- provide a familiar metaphor while supporting direct and prompt interaction, not complex queries and commands
- add common network-related analysis tasks (see the InfoVis lecture on graphs and networks)

MULTIPLE VIEWS



details

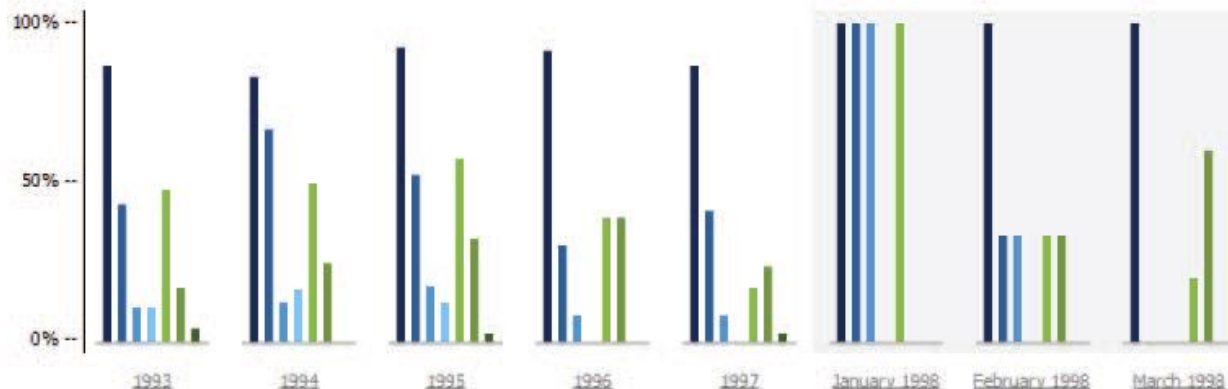
MULTIPLE VIEWS

Hewlett-Packard Co

Agreement Type



timeline



MULTIPLE VIEWS

portfolio overview



Ecosystem Segments



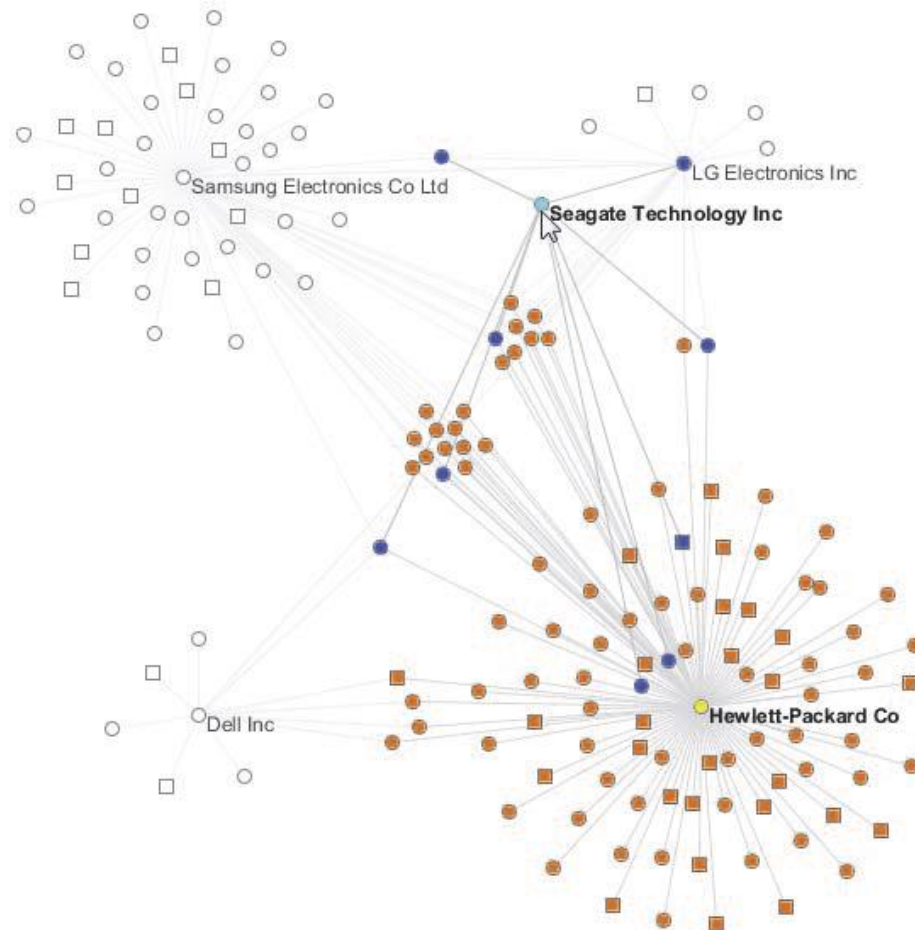
Companies



Countries and Regions



MULTIPLE VIEWS



comparison

... and many more

- financial institutions have obligation to discover suspicious financial transactions
 - can be fined if not found or shut down
- many transactions are purely digital & banks often are only the middle “man”
- large banks need to monitor hundreds of thousands of transactions per day

WIREVIS

Remco Chang, Mohammad Ghoniem, Robert Kosara, William Ribarsky, Jing Yang,
Evan Suma, Caroline Ziemkiewicz, Daniel Kern, Agus Sudjianto

IEEE Visual Analytics Science and Technology (VAST) 2007.

WireVis

***Visualization of Categorical, Time-Varying Data
From Financial Transactions***

UNC Charlotte

Remco Chang Mohammad Ghoniem

Robert Kosara William Ribarsky

Jing Yang Evan Suma Caroline Ziemkiewicz

Bank of America

Daniel Kern Agus Sudjianto

<https://www.youtube.com/watch?v=RPKcrLQBqiE>

WIREVIS

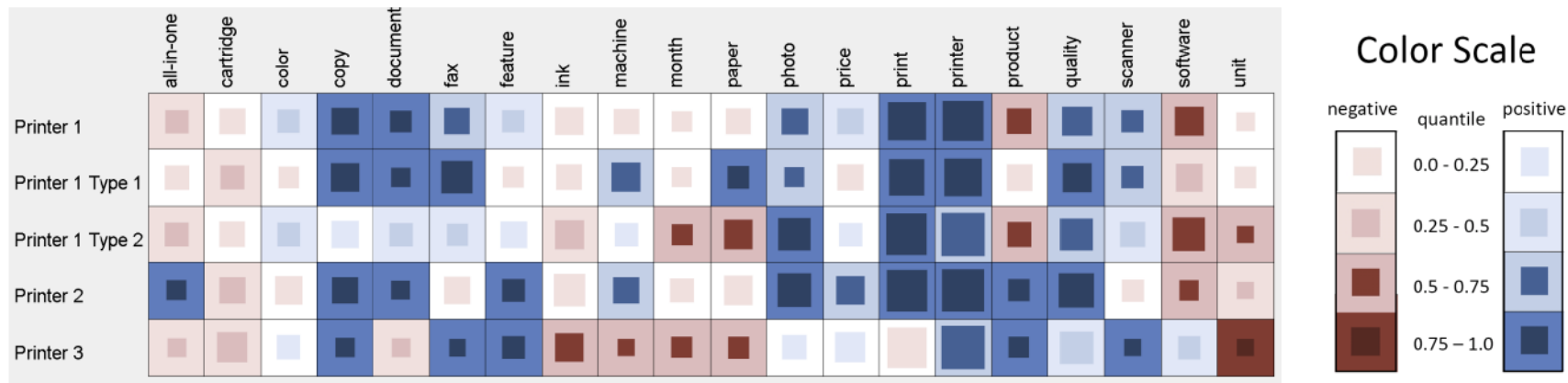
goals

- provide overview of transactions over any period of time → apply to large data
- allow identification of patterns over time and keywords
- replace blind queries with in-place analysis
- provide search-by-example technique

WIREVIS

- wire transaction data
 - semi-structured data record
 - fixed data (sender, receiver, etc.)
 - optional free text
 - temporal, categorical, quantitative
- current (2007) procedure
 - filter based on risk assessment (e.g. money exceed threshold, sender/receiver is a high-risk country or organization)
 - look at spreadsheets

VISUAL ANALYSIS OF CUSTOMER FEEDBACK



Oelke, D.; Ming Hao; Rohrdantz, C.; Keim, D.A.; Dayal, U.; Haug, L.; Janetzko, H., "Visual opinion analysis of customer feedback data," *Visual Analytics Science and Technology*, 2009. VAST 2009. IEEE Symposium on , vol., no., pp.187,194, 12-13 Oct. 2009
doi: 10.1109/VAST.2009.5333919

VISUAL ANALYSIS OF CUSTOMER FEEDBACK

2

ATTRIBUTE	POSITIVE	NEGATIVE
document	0.0%	8.06%
fax	0.0%	17.74%
ink cartridge	0.0%	33.87%
machine	0.0%	32.25%
photo	0.0%	11.29%
scanner	0.0%	9.67%

6

ATTRIBUTE	POSITIVE	NEGATIVE
document	0.0%	33.33%
fax	0.0%	16.66%
ink cartridge	0.0%	66.66%
machine	100.0%	0.0%
photo	16.66%	16.66%
scanner	33.33%	16.66%

4

ATTRIBUTE	POSITIVE	NEGATIVE
document	0.0%	45.45%
fax	0.0%	18.18%
ink cartridge	54.54%	0.0%
machine	36.36%	0.0%
photo	0.0%	36.36%
scanner	9.09%	0.0%

5

ATTRIBUTE	POSITIVE	NEGATIVE
document	0.0%	10.0%
fax	40.0%	0.0%
ink cartridge	0.0%	50.0%
machine	0.0%	40.0%
photo	20.0%	0.0%
scanner	40.0%	0.0%

7

ATTRIBUTE	POSITIVE	NEGATIVE
document	71.42%	0.0%
fax	14.28%	0.0%
ink cartridge	0.0%	14.28%
machine	0.0%	71.42%
photo	14.28%	14.28%
scanner	42.85%	14.28%

1

ATTRIBUTE	POSITIVE	NEGATIVE
document	16.01%	0.0%
fax	32.3%	0.0%
ink cartridge	27.52%	0.0%
machine	25.84%	0.0%
photo	24.15%	0.0%
scanner	20.22%	0.0%

3

ATTRIBUTE	POSITIVE	NEGATIVE
document	10.0%	10.0%
fax	60.0%	0.0%
ink cartridge	50.0%	0.0%
machine	30.0%	0.0%
photo	30.0%	20.0%
scanner	0.0%	80.0%

LEGAL MATTERS

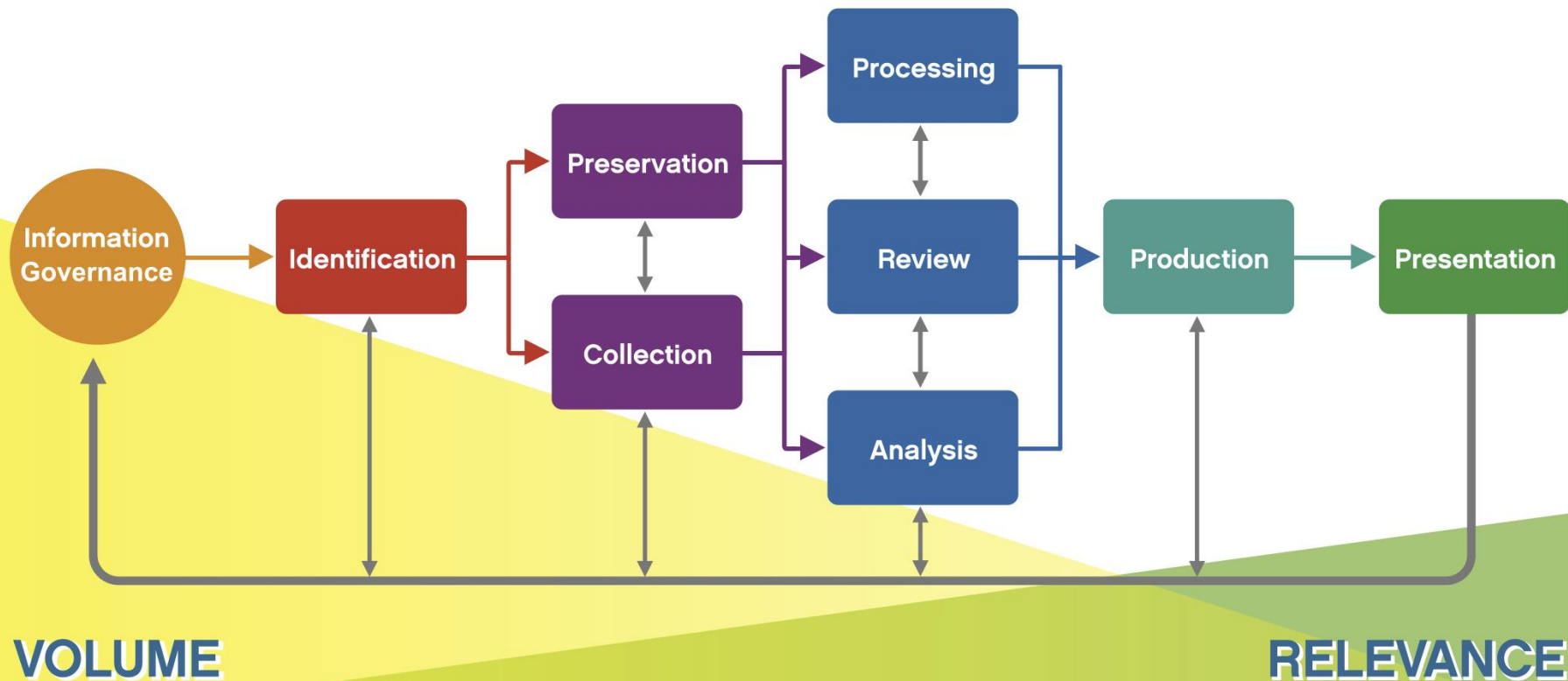
or

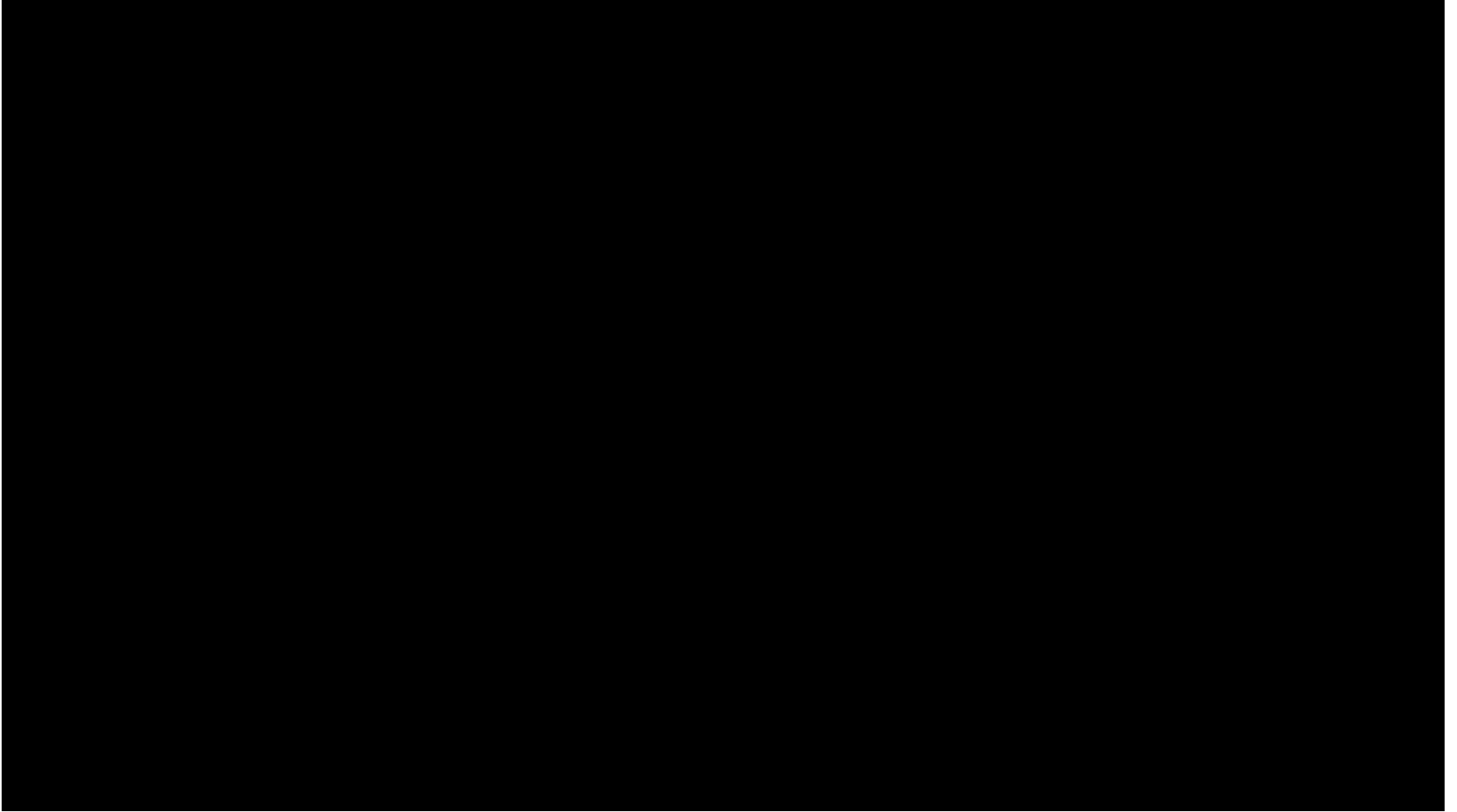
working with large corpora of electronic texts

E-DISCOVERY (MOSTLY USA)

- legal electronic document discovery
 - for use in law suits
 - regulatory information requests
 - investigations, audits, freedom of information act, ...
- documents typically given to court & opponent
- if complaint received, corporations have to produce all related information

Electronic Discovery Reference Model

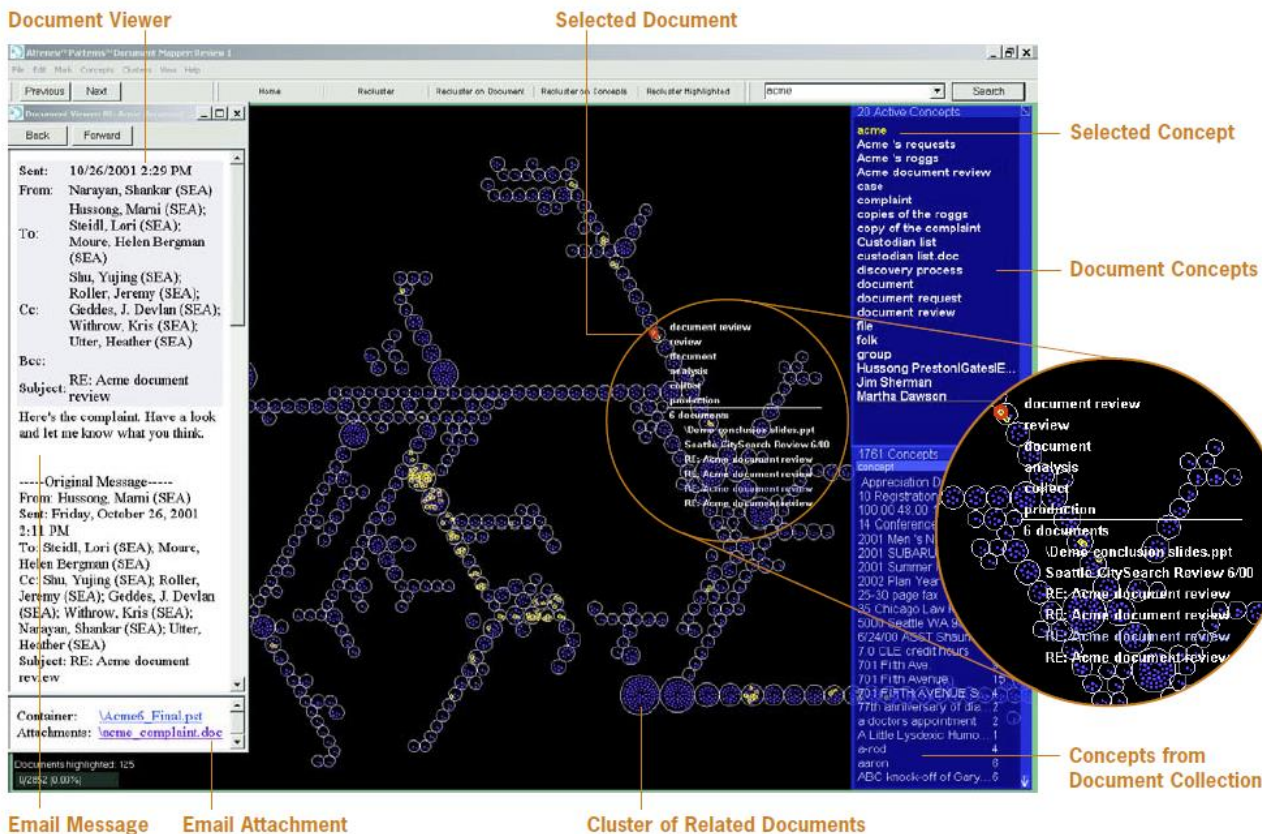




https://www.youtube.com/watch?v=eo03DWk4_IU

EXAMPLE TOOLS

Commercial: Attenex Patterns (now FTI Ringtail)



SUCCESS STORIES OF VA TOOL

- Lovells (6th largest law firm worldwide) was investigating 35Gb of email data
 - traditional method: estimated cost: 1 year, 4-5 million \$US
 - with VA software: three months, cost 1 million \$US
- JMBM law firm represented spinal surgeon & inventor in intellectual property suit
 - 20 claims, 50 million pages of documents, contracts, patents, etc.
 - with VA software (used on 7 computers), 44 million pages sifted through in 4 months
 - important doc found, surgeon won, received \$570 million in compensation

Figure 1. Magic Quadrant for E-Discovery Software



Source: Gartner (June 2014)

RESEARCH

- working on this level in research is difficult
 - man power for providing support
 - “research aspect” needs most attention
 - adoption of prototypes unlikely when stakes are high (\$\$\$)
- thus, next: smaller research projects in regards to document analysis

PARALLEL TAG CLOUDS

C. Collins, F. B. Viégas, and M. Wattenberg, “Parallel Tag Clouds to Explore and Analyze Facted Text Corpora,” in *Proc. of the IEEE Symp. on Visual Analytics Science and Technology (VAST)*, 2009.

Parallel Tag Clouds to Explore
and Analyze Faceted Text Corpora

Christopher Collins
Fernanda B. Viégas
Martin Wattenberg

FACET ATLAS

Nan Cao; Jimeng Sun; Yu-Ru Lin; Gotz, D.; Shixia Liu; Huamin Qu,
"FacetAtlas: Multifaceted Visualization for Rich Text Corpora,"
Visualization and Computer Graphics, IEEE Transactions on , vol.16,
no.6, pp.1172,1181, Nov.-Dec. 2010
doi: 10.1109/TVCG.2010.154

FacetAtlas: Multifaceted Visualization for Rich Text Corpora

InfoVis 2010

**NanCao, Jimeng Sun, Yu-Ru Lin, David Gotz,
Shixia Liu, Huamin Qu**

CAMBIERA

Collaborative Brushing and Linking
for Co-located Visual Analytics
of Document Collections

Petra Isenberg & Danyel Fisher
Microsoft Research

VISUAL ANALYTICS @ AVIZ

www.aviz.fr

WORK AT AVIZ

- We often work non-applied (on generalizable conceptual problems)
 - collaboration
 - perception
 - interaction
- Some application areas:
 - biology
 - digital humanities
 - Wikipedia
 - sports

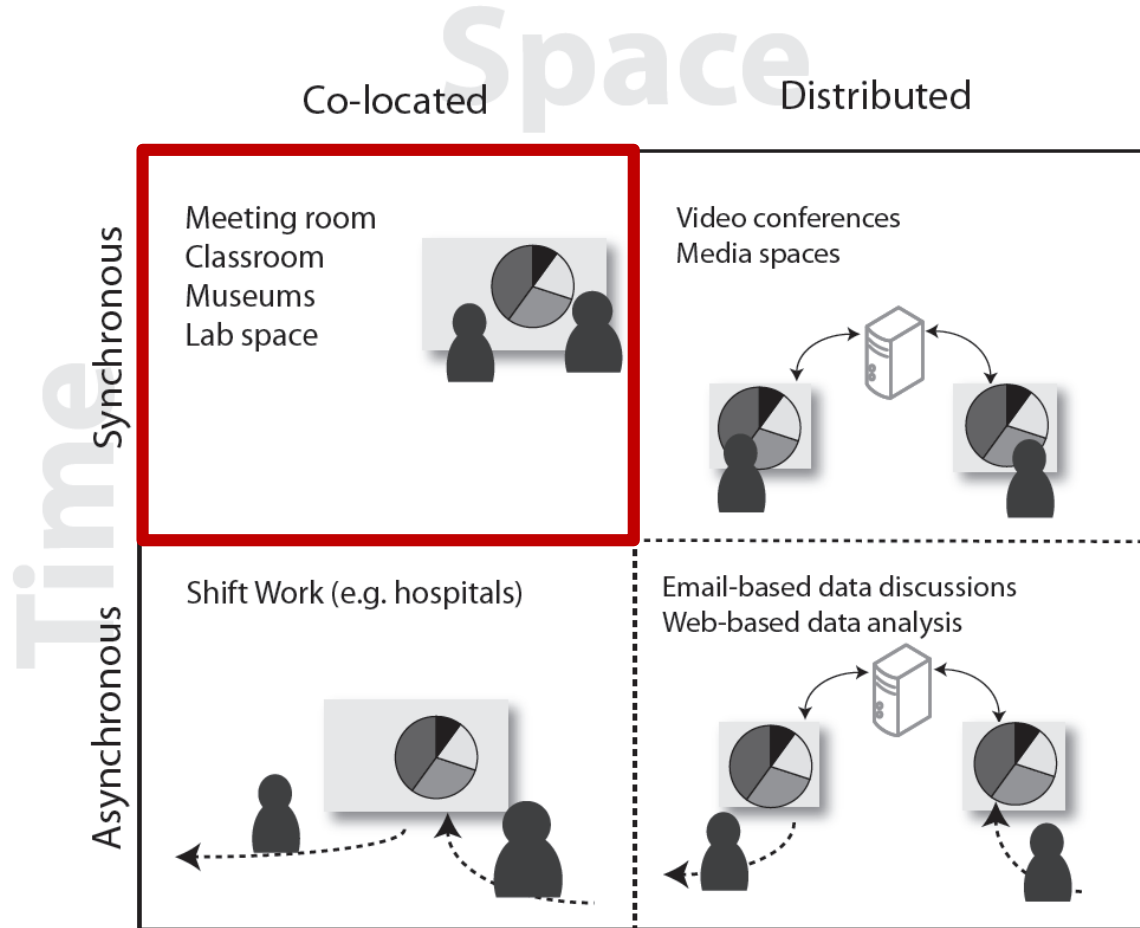
GENERAL CONCEPT

Collaboration

COLLABORATIVE ANALYSIS

- share task load
- share expertise
- joint problem solving



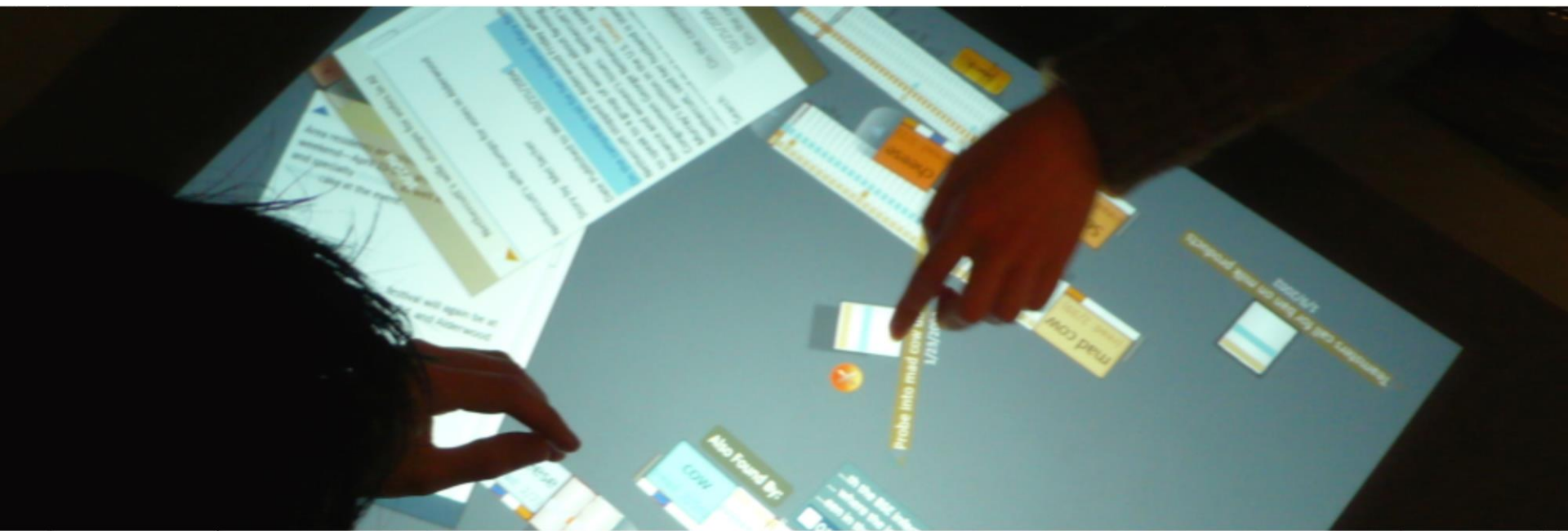


CHALLENGES

collaborative data analysis is still poorly understood

- collaborative analysis & sharing?
- collaborative space usage?
- task coordination?
- collaborative sensemaking?
- when & where?

CAMBIERA



EXPLORATORY STUDY

understand digital tabletop as data analysis environment

- **task coordination**
- **patterns of collaboration**
- **system design implications**

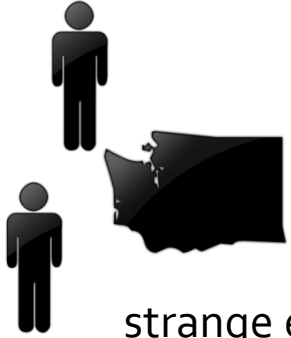
STUDY DESIGN

- participants:
 - 15 pairs
 - team members knew each other
 - master's or more advanced degree
 - enjoy puzzles / mysteries
 - age: 25 - 55
- data collected:
 - notes, video data, interaction logs, questionnaire, interview data
 - 2-pass video coding + visualization of log files

246 documents



e-coli contamination



strange events
Alderwood, WA



car accident



poisoned apples



terrorism in south america

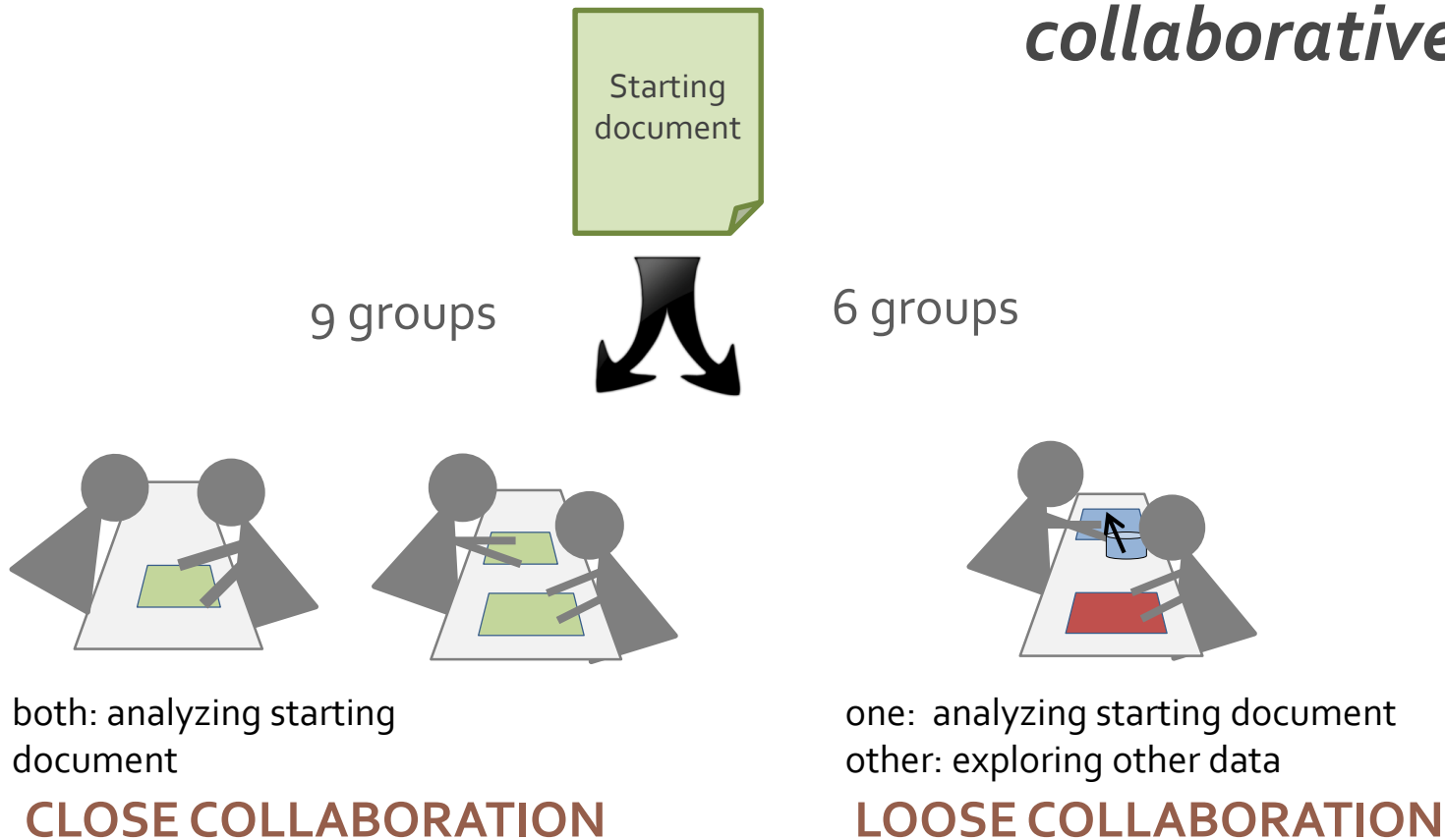
More information:

@ Visual Analytics Benchmark Repository

<http://hcil.cs.umd.edu/localphp/hcil/vast/archive>

RESULT

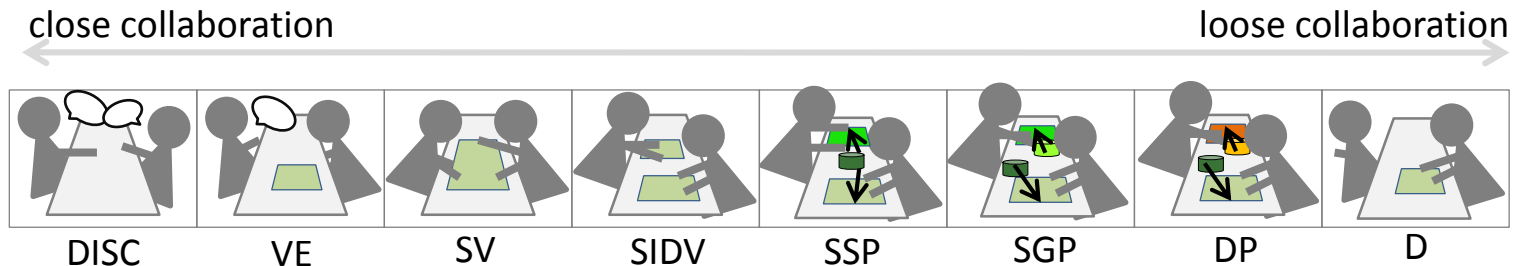
*solving the
problem
collaboratively*



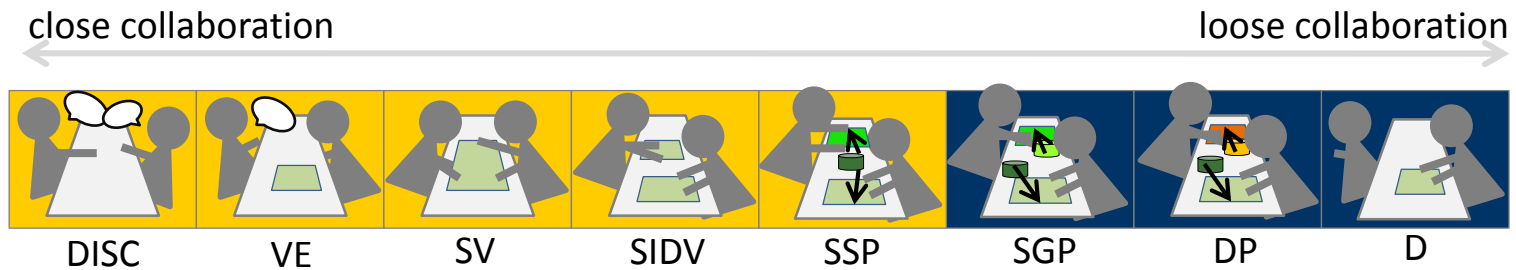
RESULTS

information sharing and collaboration

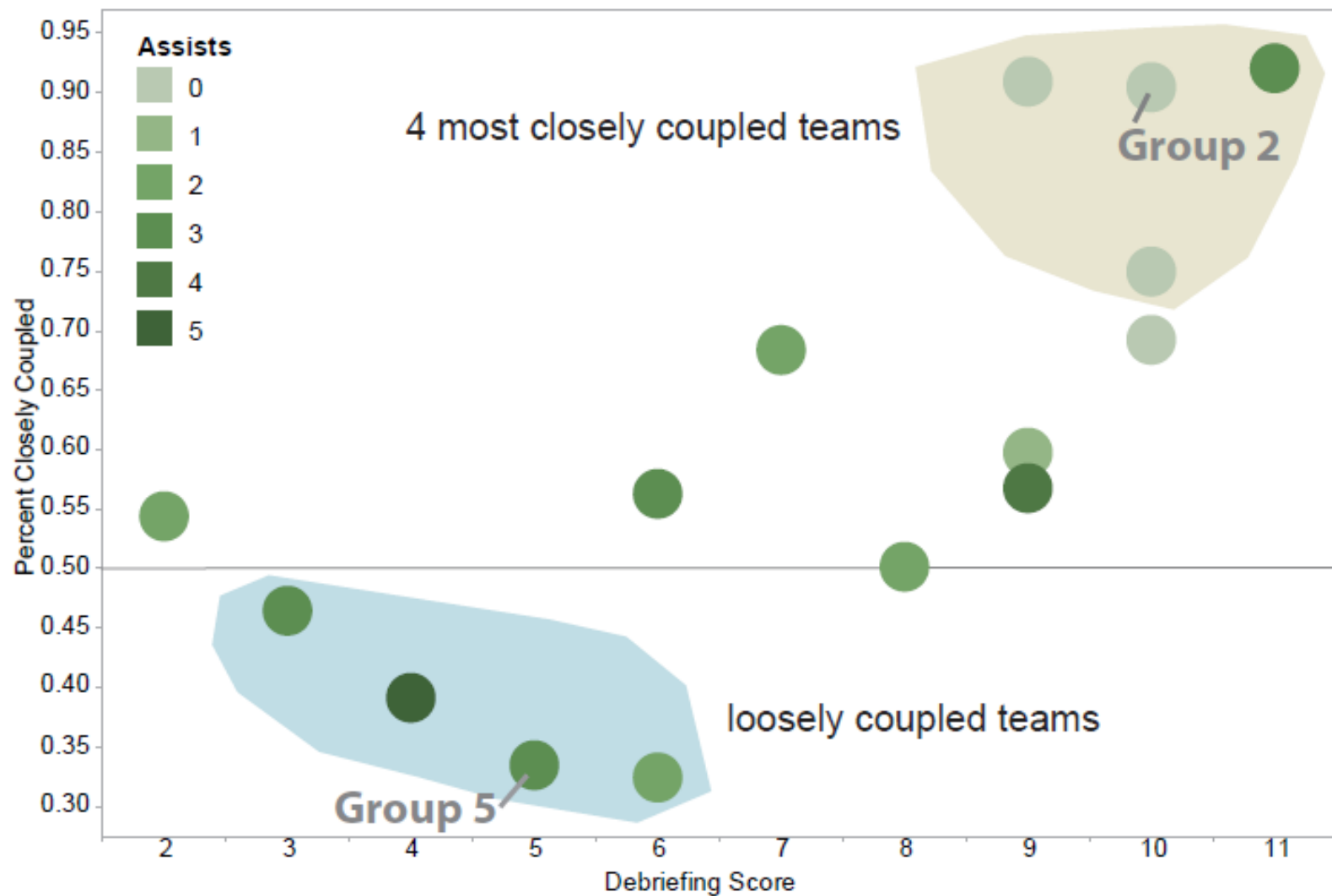
- different collaboration styles adopted
- influenced what data/views were shared
- allowed flexible investigation based on emerging information



PROCESS



Pair	% C.C.	Coupling Styles
1	92%	
15	91%	
2	90%	
12	75%	
10	69%	
13	68%	
14	60%	
11	57%	
9	56%	
6	54%	
4	50%	
3	46%	
8	39%	
5	33%	
7	32%	



IMPLICATIONS FOR DESIGN

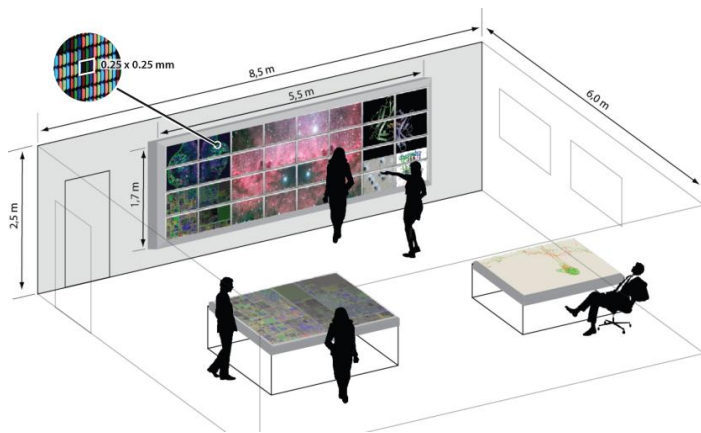
- design for transient behavior
 - strategies change & interfaces need to accommodate
 - design system features to support different styles
- encourage closely coupled work
 - awareness features possibly not strong enough for loosely-coupled teams
 - make common information even more obvious

GENERAL CONCEPT

perception

HYBRID IMAGE VISUALIZATION

Petra Isenberg, Pierre Dragicevic, Wesley Willett, Anastasia Bezerianos, and Jean-Daniel Fekete. Hybrid-Image Visualization for Large Viewing Environments. IEEE Transactions on Visualization and Computer Graphics, 19(12):2346–2355, December 2013.

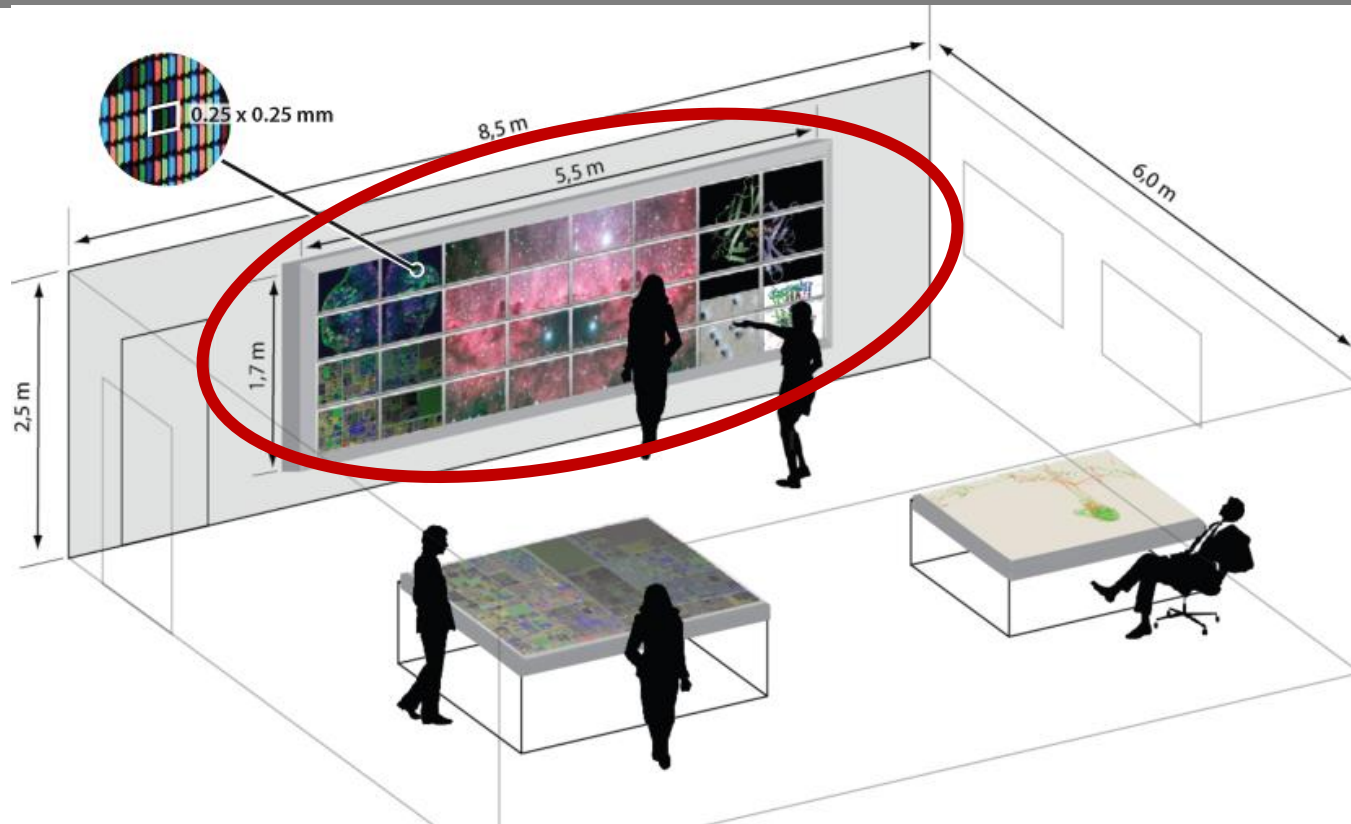


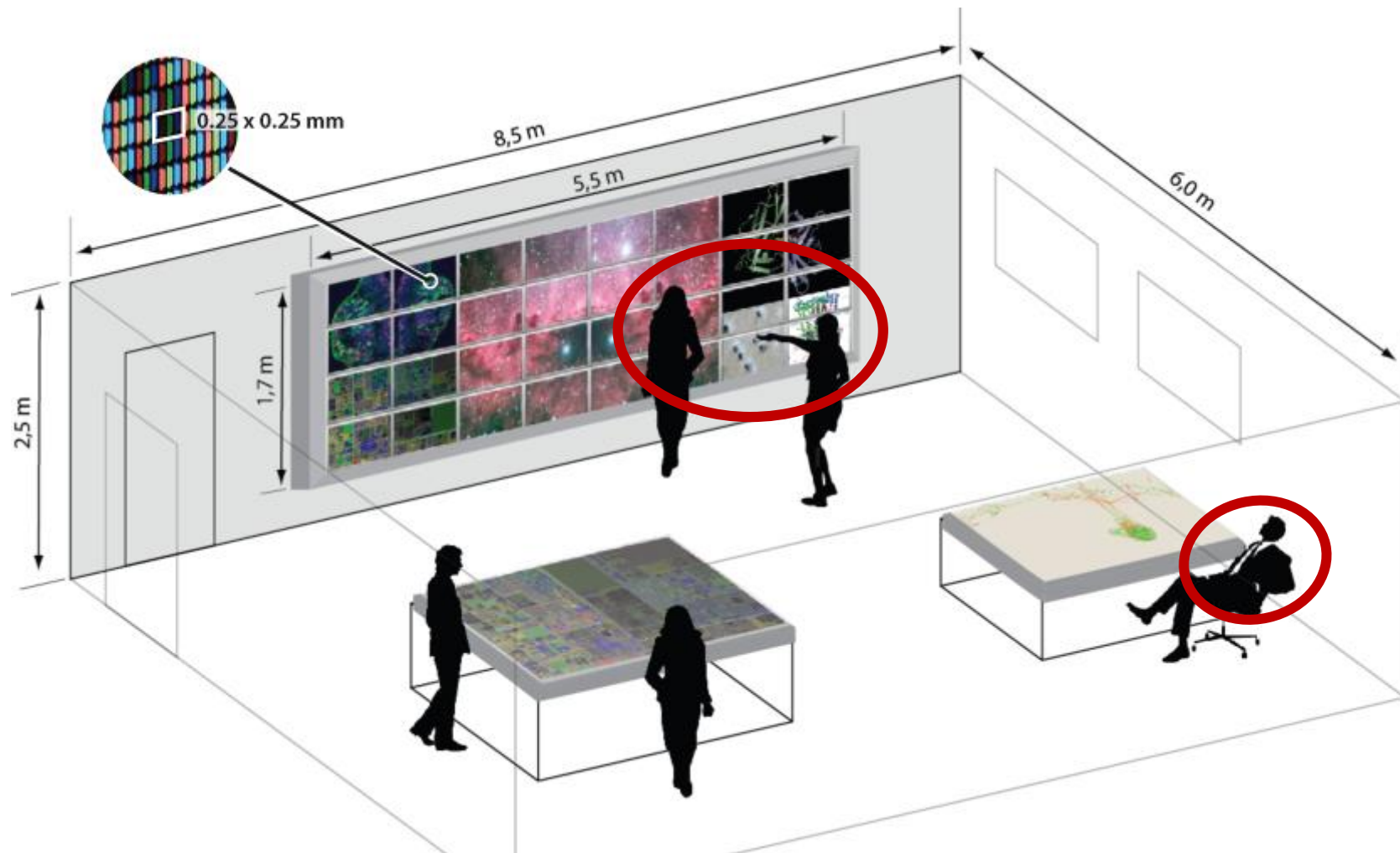
large displays
meeting rooms / war rooms
emergency response rooms



public viewing
open spaces
museums

SCENARIO: WILD DISPLAY ENVIRONMENT

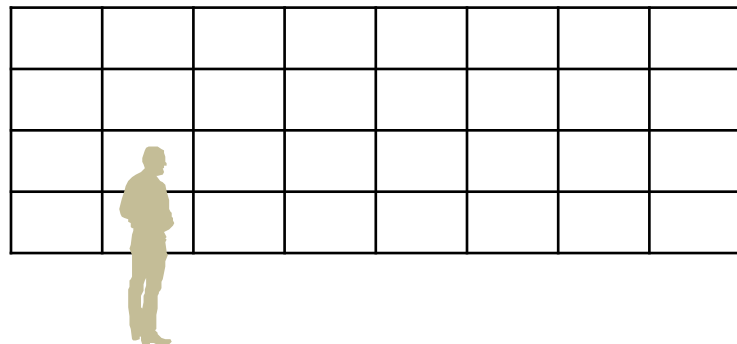




EXAMPLE DATA

23 years of daily temperatures for 32 cities

Station	Country	Station ID	SOUID	Date	Temperature (decidegrees Celsius)	Temperature Quality
BERLIN-DAHLEM	GERMANY	41	133	1990-01-01	-21	0
BERLIN-DAHLEM	GERMANY	41	133	1990-01-02	-13	0
BERLIN-DAHLEM	GERMANY	41	133	1990-01-03	-2	0
BERLIN-DAHLEM	GERMANY	41	133	1990-01-04	-12	0
BERLIN-DAHLEM	GERMANY	41	133	1990-01-05	0	0
BERLIN-DAHLEM	GERMANY	41	133	1990-01-06	-13	0
BERLIN-DAHLEM	GERMANY	41	133	1990-01-07	-40	0
BERLIN-DAHLEM	GERMANY	41	133	1990-01-08	6	0
BERLIN-DAHLEM	GERMANY	41	133	1990-01-09	35	0
BERLIN-DAHLEM	GERMANY	41	133	1990-01-10	44	0
BERLIN-DAHLEM	GERMANY	41	133	1990-01-11	48	0
BERLIN-DAHLEM	GERMANY	41	133	1990-01-12	60	0
BERLIN-DAHLEM	GERMANY	41	133	1990-01-13	22	0
BERLIN-DAHLEM	GERMANY	41	133	1990-01-14	30	0
BERLIN-DAHLEM	GERMANY	41	133	1990-01-15	42	0



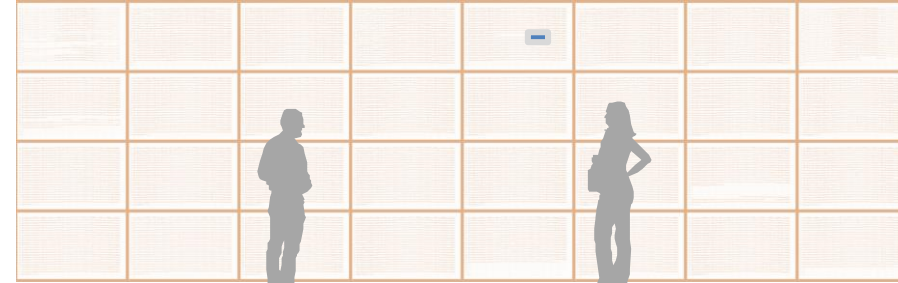
NAÏVE APPROACH

Optimize for one specific viewing distance



overviews from far

Problem:
detailed information lost



detail from close

Problem:
mental aggregation, comparison

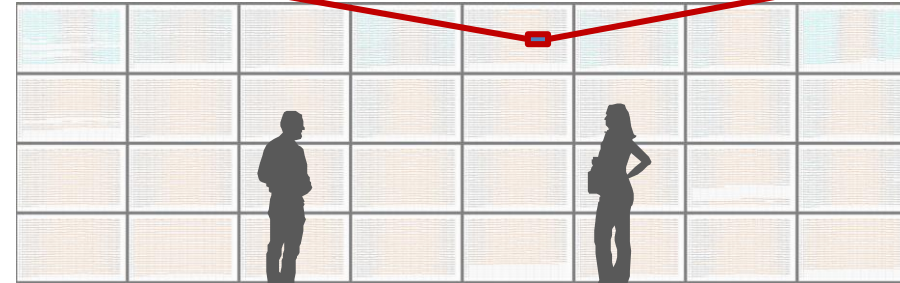
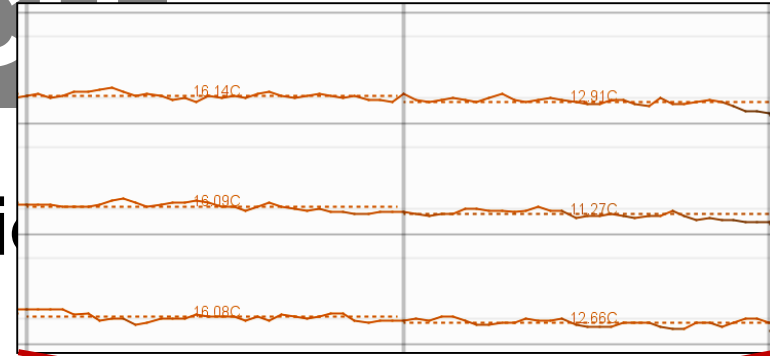
NAÏVE APPROACH

Optimize for one specific view



overviews from far

Problem:
detailed information lost

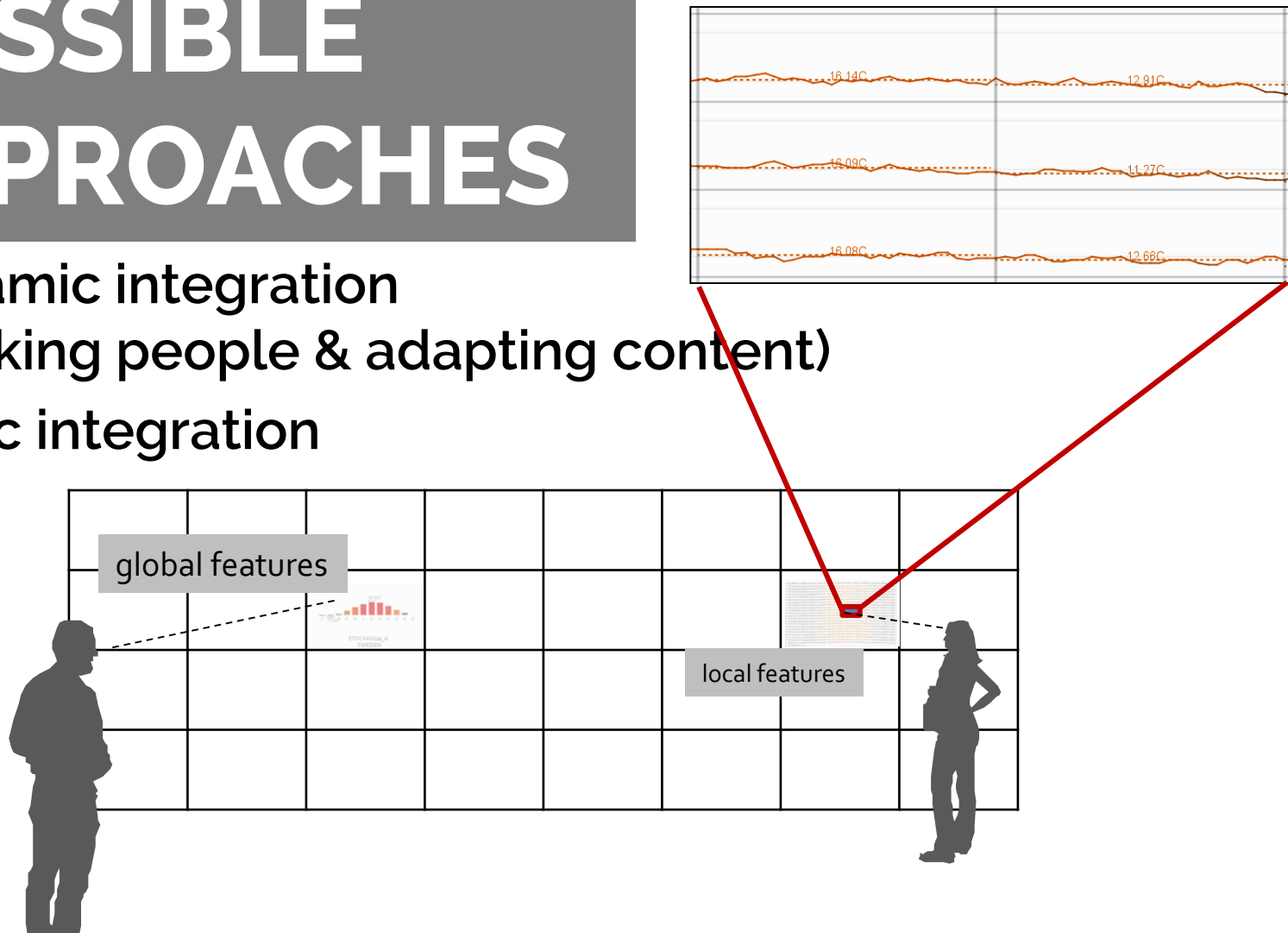


detail from close

Problem:
mental aggregation, comparison

POSSIBLE APPROACHES

- Dynamic integration
(tracking people & adapting content)
- Static integration

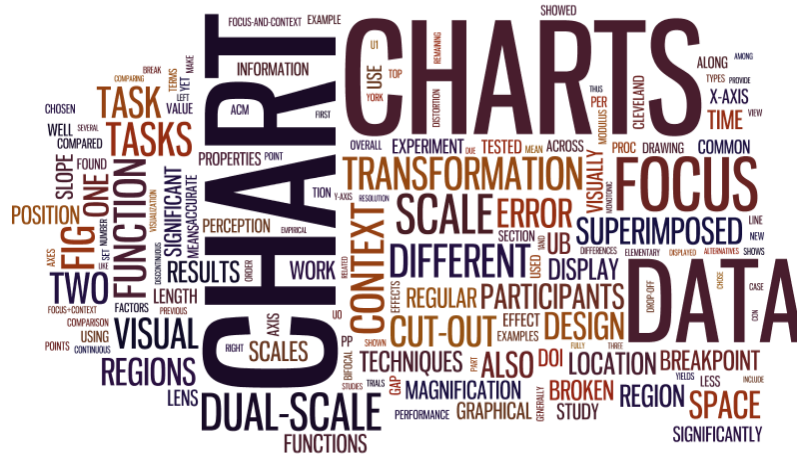


STATIC APPROACHES

Juxtaposition:
global & local features next to each other



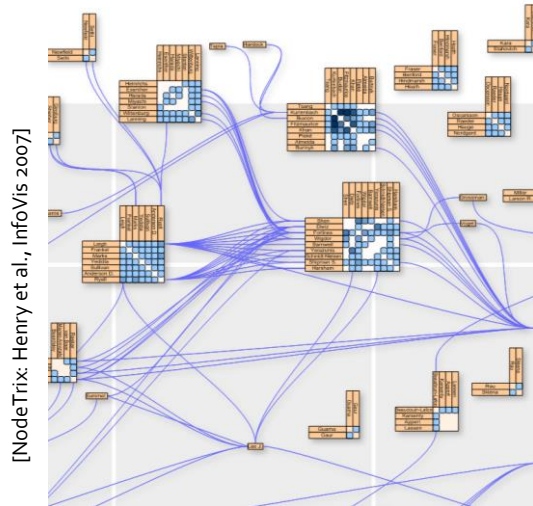
Infographic Layout



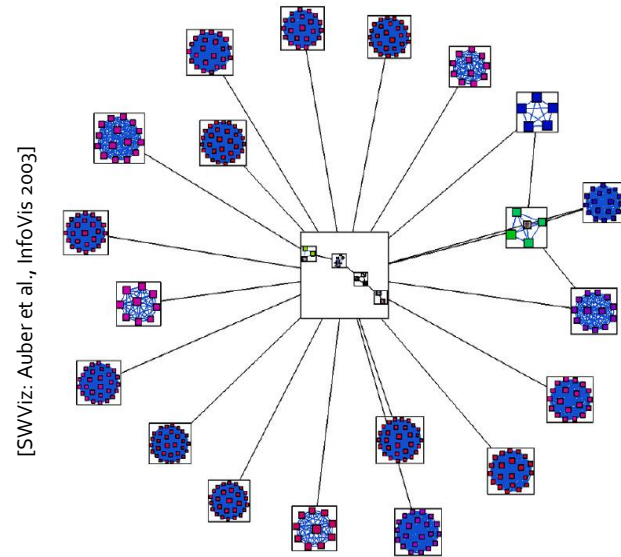
Large Tag Clouds

STATIC APPROACHES

Nesting:
local inside global features



NodeTrix

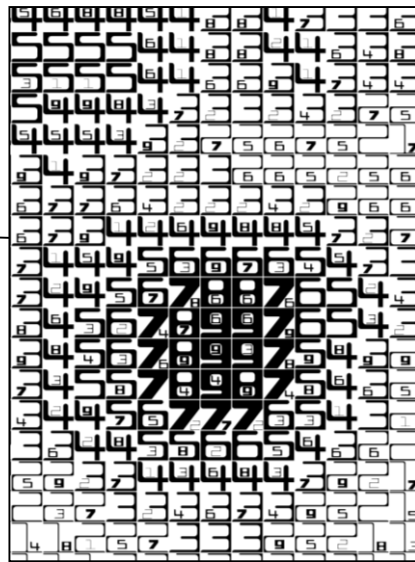
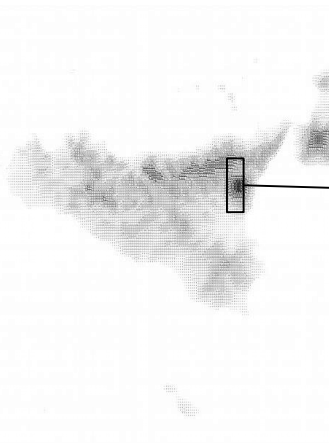


SWViz

STATIC APPROACHES

Visual aggregation:
local forms global features

[FatFonts: Nacenta et al., AVI 2012]



FatFonts



Arcimboldo Paintings, Calligrams

STATIC APPROACHES

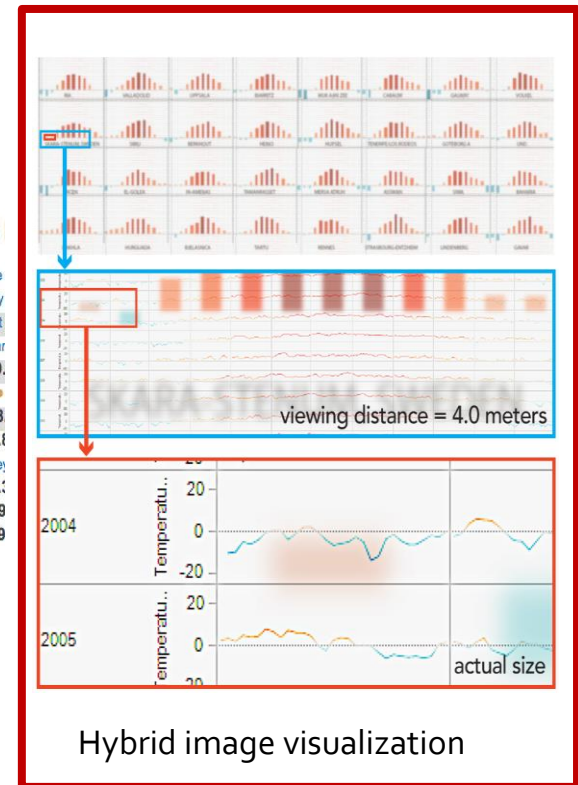
Blending:
local & global features overlap



Map regions over map detail

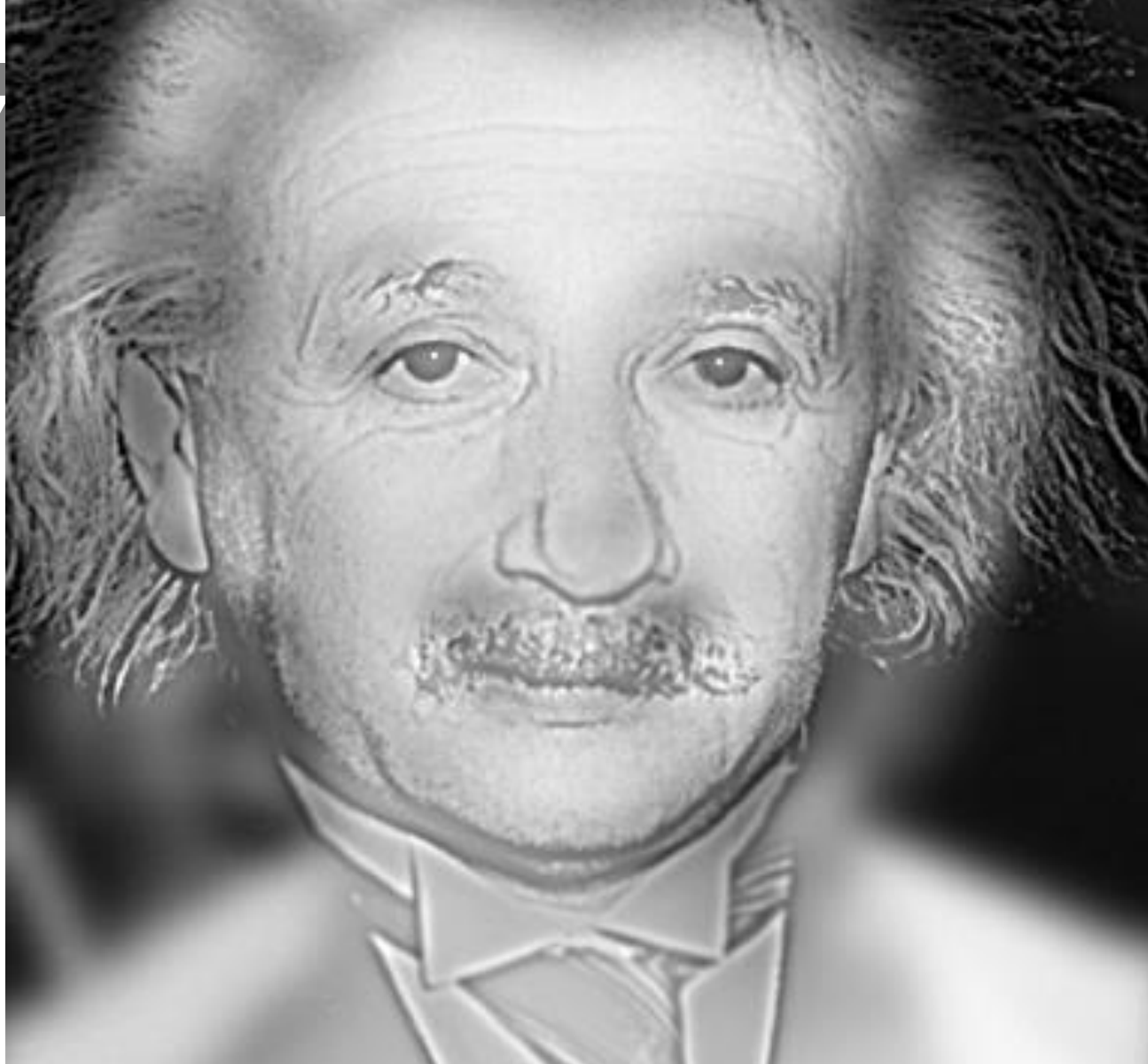


Label detail over bar chart



Hybrid image visualization

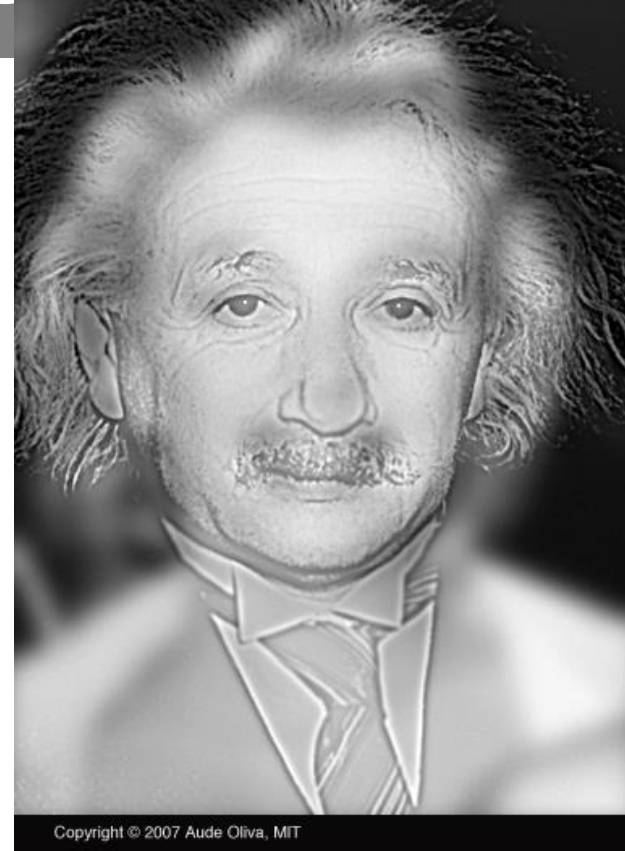
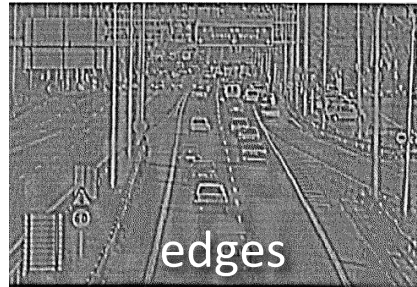
HY



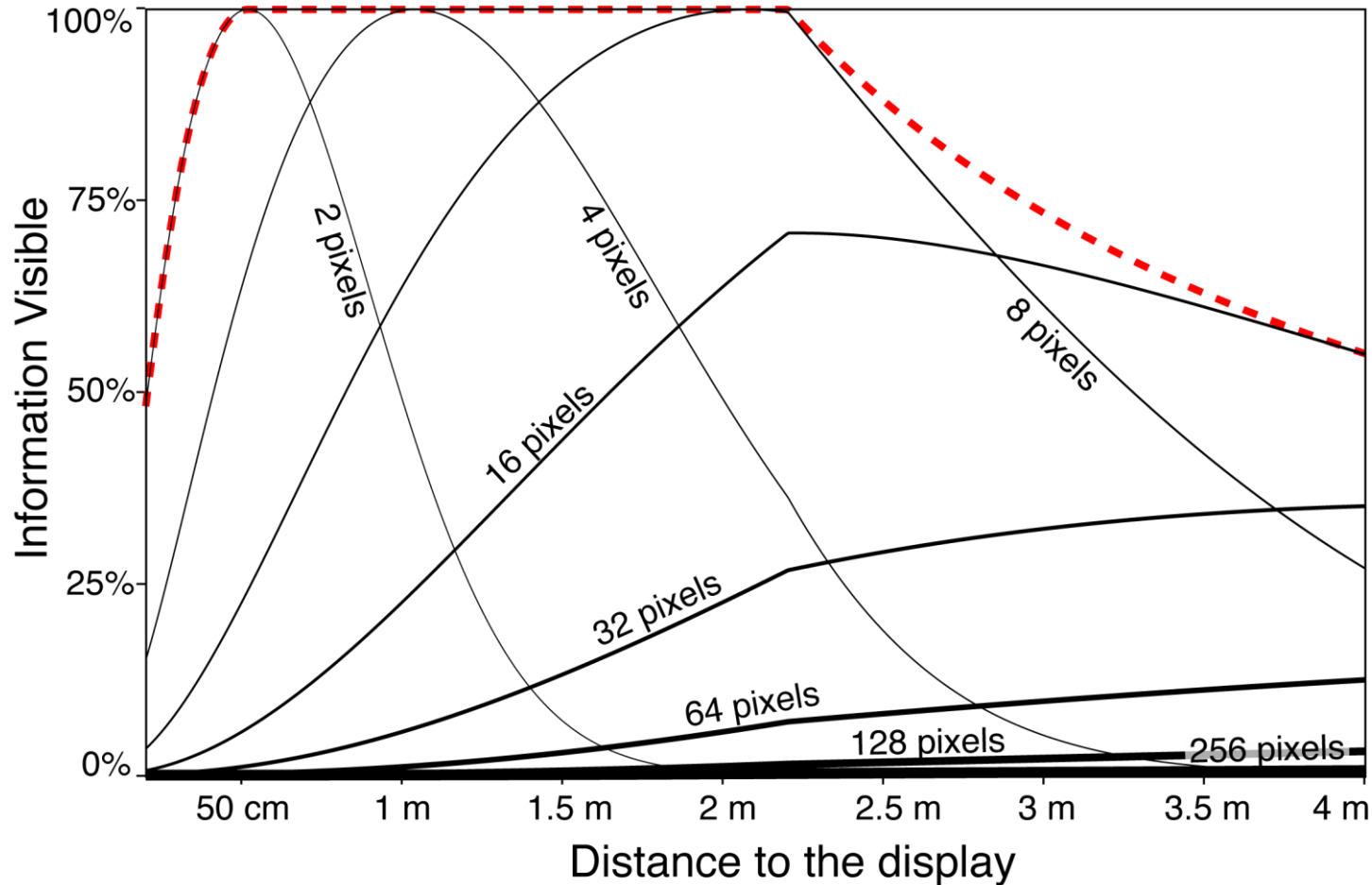
HYBRID IMAGES



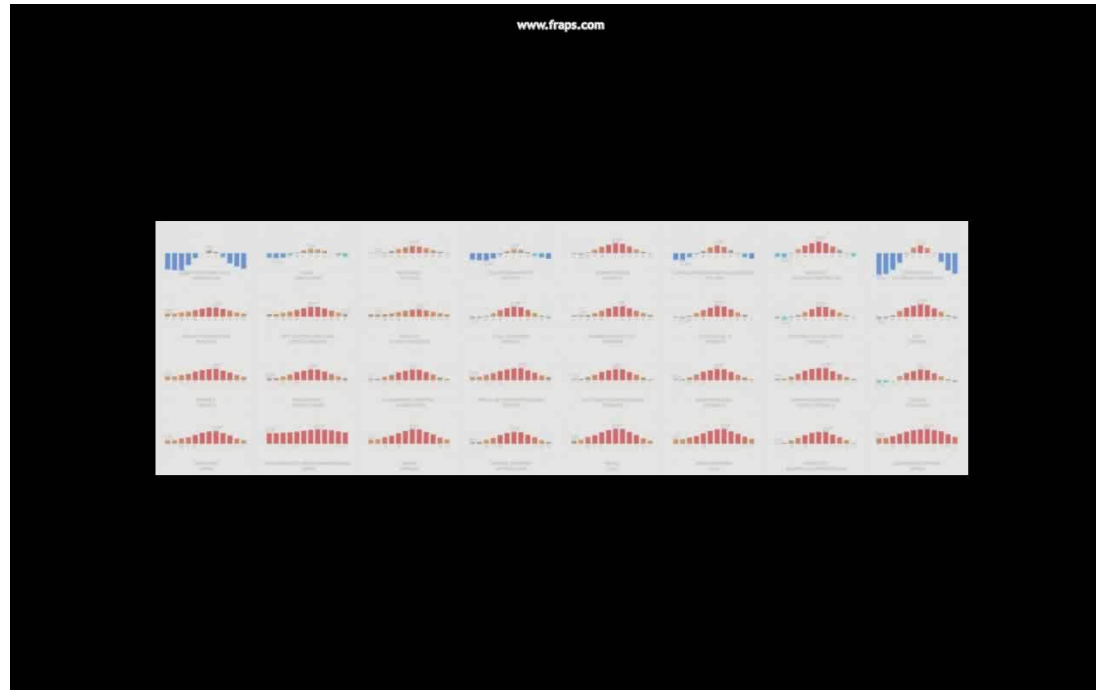
+



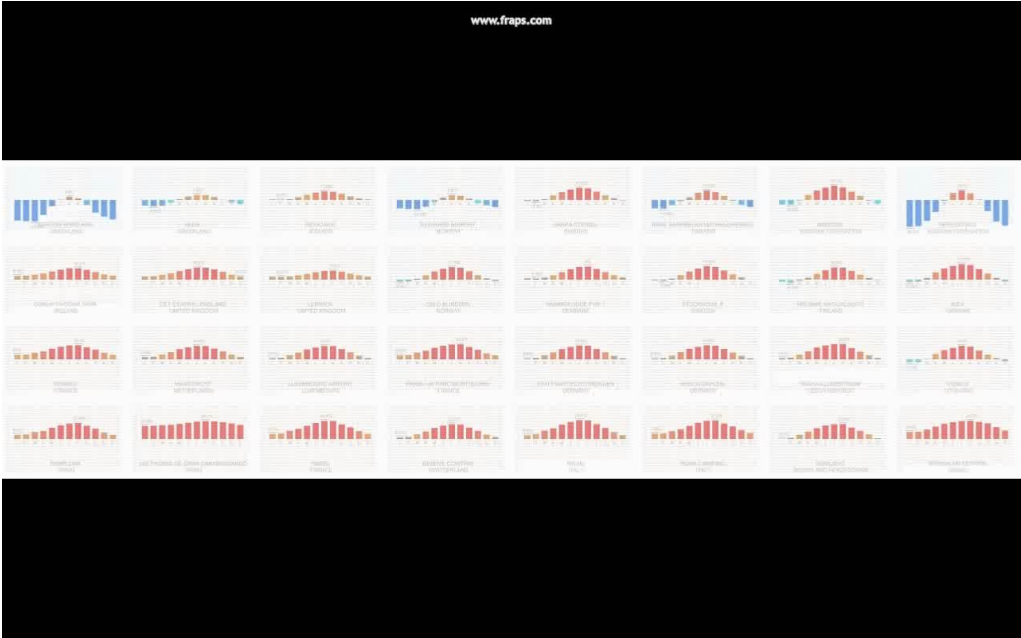
[VICTIM OF TIME: THE THEORY]



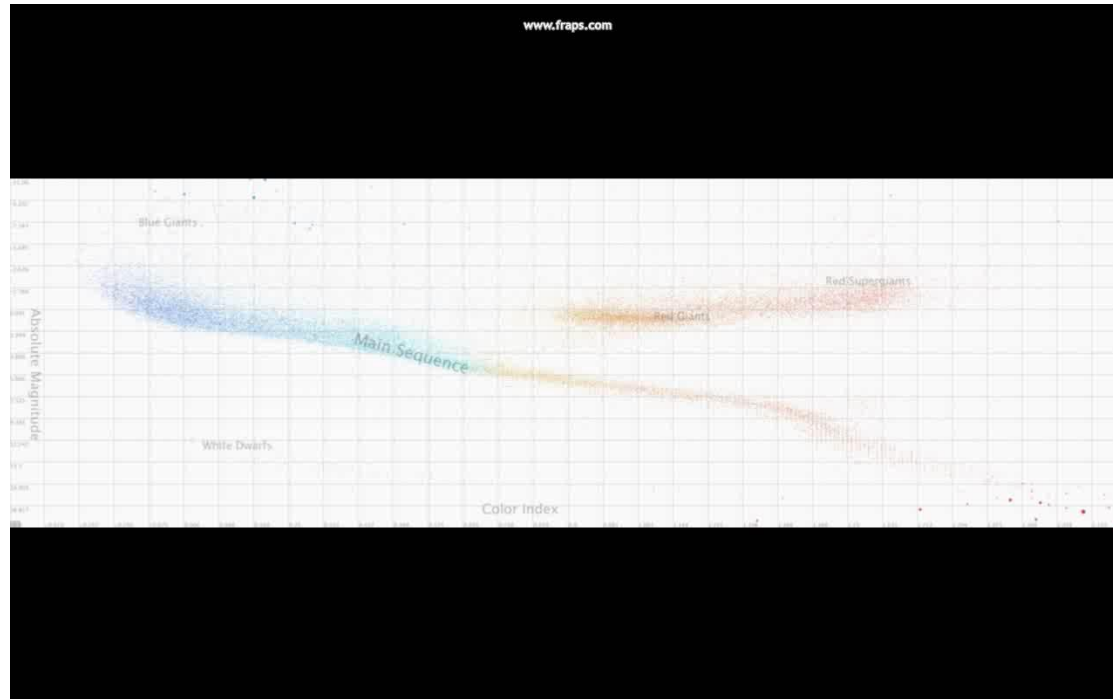
HYBRID IMAGE VISUALIZATION - VIDEO



ALPHA BLENDING ONLY - VIDEO



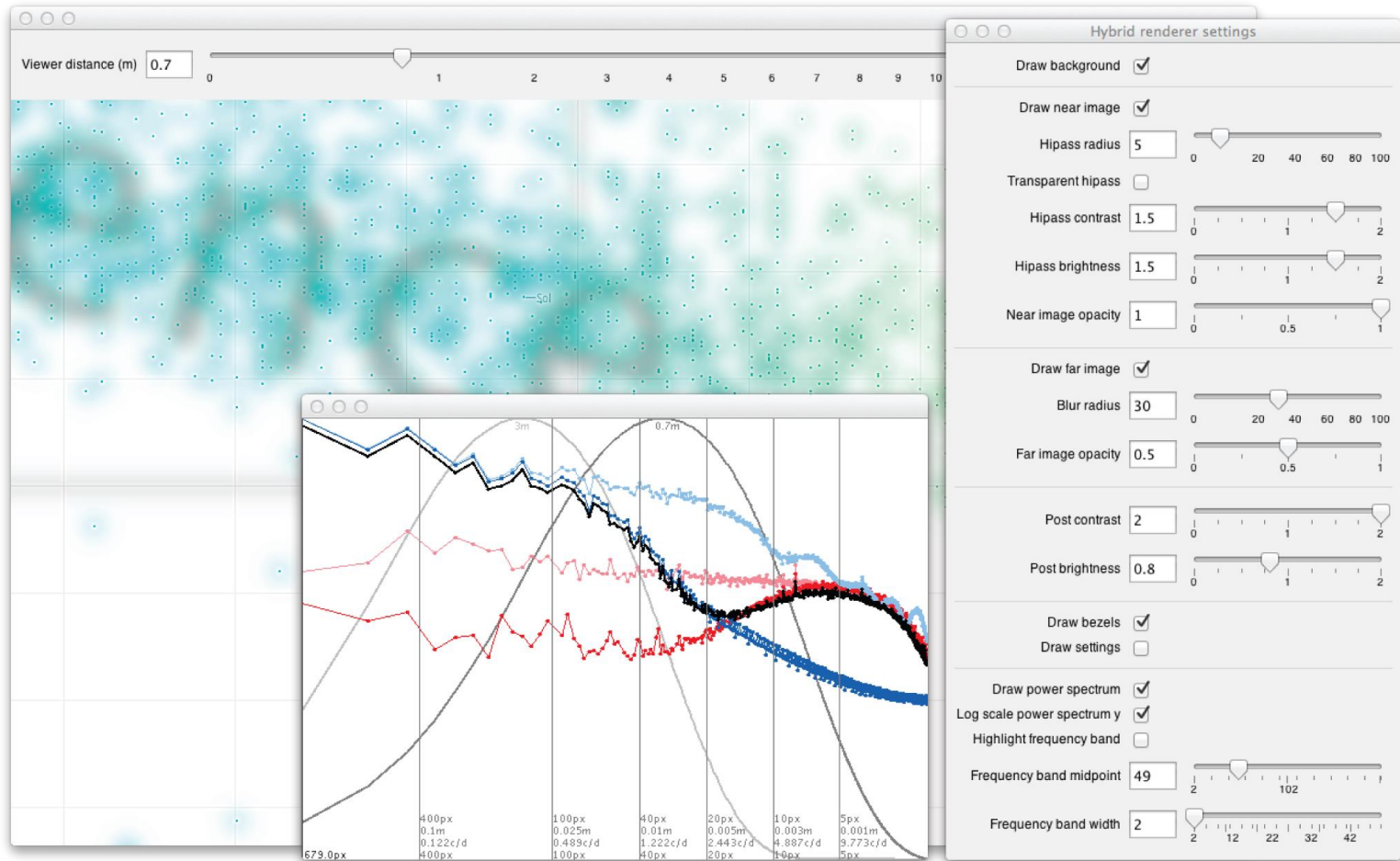
HERTZSPRUNG RUSSELL DIAGRAM – VIDEO



TREEMAP - VIDEO



HYBRID IMAGE MAKER



APPLICATIONS @ AVIZ

SOCCER STORIES



SoccerStories

A kick-off for Soccer Visual Analysis


www.aviz.fr/soccerstories

Charles Perin
Romain Vuillemot
Jean-Daniel Fekete

INRIA & Univ. Paris-Sud
INRIA
INRIA

BUBBLE-TV

DIRECT 



@LeGrandWebze
legrandwebze
Faut-il garder François Rollin ?
Jouez !! #IlReste #IlPart #cQui
#change2ton #lgw

Hashtag	Count
#IlReste	41
#IlPart	13
#cQui	10
#change2ton	13


LEGRANDWEBZE.COM

POLEMICTWEET



Accueil
Programme
A propos

日本語
Français
English

Intégrer



Clay Shirky : « Personne n'est titulaire du code source de la démocratie »

Journaliste américain spécialiste des nouvelles technologies de l'information et de la communication - NTIC, Clay Shirky est aussi consultant, écrivain, professeur.

Diplômé de l'Université de New York il écrit et enseigne, entre autres, sur les effets de corrélation de la topologie sociale et technologique de réseau, où comment nos réseaux forment la culture et inversement.

Dernier ouvrages :

- [2010 : Cognitive Surplus: Creativity and Generosity in a Connected Age](#)
- [2008 : Here Comes Everybody: The Power of Organizing Without Organizations](#)

Source : [Wikipedia.fr](#)

Plus d'information : [www.shirky.com](#)

Crédit photo : [CC BY NA - Pop!Tech 2008](#)

Liste des Annotations

Mots-clés



Question 4 : format and media (46:44 - 50:47)



Créé par : admin

Description : lien entre les types de médias et les organisation politique question de joaquin keller

Mots-clés :

consumers democracy fascism Internet media
printing democracy radio revolution share television

MUCH MORE AT

<http://www.aviz.fr> → projects