

CSE/ISE 332

Introduction to Visualization

Visual Analytics & The Visual Sense Making Process

Klaus Mueller

Computer Science Department
Stony Brook University

Lecture	Topic	Projects
1	Intro and logistics	
2	Basic visualizations and tasks, data types, examples, ethical considerations	
3	Data preparation (cleaning, imputation, data set integration)	Project #1 out
4	AI-assisted coding for VIS applications (design, debugging, refactoring)	
5	Big data and data reduction (distance/sim metrics, intro to clustering)	
6	High-D data: concept, subspaces, dimension reduction, PCA	
7	Cluster analysis: hierarchical, density, model, embedding, temporal	Project #2(a) out
8	Perception and cognition (human visual system, color, contrast)	
9	Visual design and aesthetics	
10	Visualization of multivariate and high-D data: linear methods, projections	
11	Vis. of multivariate and high-D data: non-linear methods, embeddings	Project #2(b) out
12	Visualization and AI: mutual support and capabilities (VIS4AI, AI4VIS)	
13	Principles of interaction: drive what is visualized, analyzed & how (HCI4VIS)	
14	Midterm recap	
15	Midterm	
16	Visual analytics, human-centered AI, mixed-initiative & collaborative VA	Capstone project out
17	Visualization system design & evaluation: nested model, study designs	
18	Visualization of hierarchical data	
19	Visualization of maps and data with geo-reference	
20	Vis. of time-varying, time-series, streaming data, progressive visualization	
21	Critiquing visualization: Tufte's views, integrity, uncertainty, responsibility	
22	Design of effective infographics	
23	Foundations scientific & medical visualization, computer graphics	
24	Intro to volume rendering	
25	Scientific visualization	
26	Capstone project demos (4-5 minutes + 2 minutes Q&A)	All capstone materials due
27	Capstone project demos (4-5 minutes + 2 minutes Q&A)	
28	Capstone project demos (4-5 minutes + 2 minutes Q&A)	
Final	Final exam	

Why Visual Analytics?

Big Data

12+ TBs
of tweet data
every day



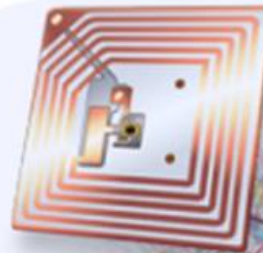
? TBs of
data every day



25+ TBs of
log data every day



30 billion RFID
tags today
(1.3B in 2005)



4.6 billion
camera
phones
world wide



100s of millions
of GPS
enabled
devices
sold
annually



76 million smart
meters in 2009...
200M by 2014

2+ billion
people on
the Web
by end
2011



Visual Analytics



Problems With Scalability

Must be scalable to

- number of data points
- number of dimensions
- data sources
- diversity of data sources (heterogeneity)
- number of users
- diversity of users and tasks
- quality of the data

Visual Analytics comes to the rescue...

The Goal of Visualization

Ease understanding of the data by providing an effective visual representation

Amplify Perception

Detect the Expected, Discover the Unexpected™

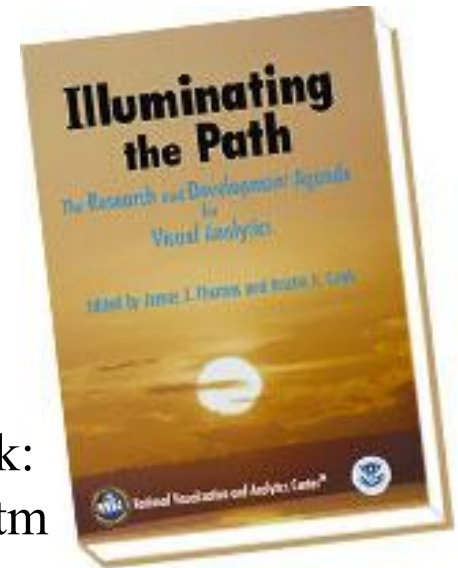
What is Visual Analytics

Visualization plus...

- interaction (HCI)
- data processing (analytics)
- story telling
- scientific approach

Agenda setting book:

<http://nvac.pnl.gov/agenda.stm>



but also...

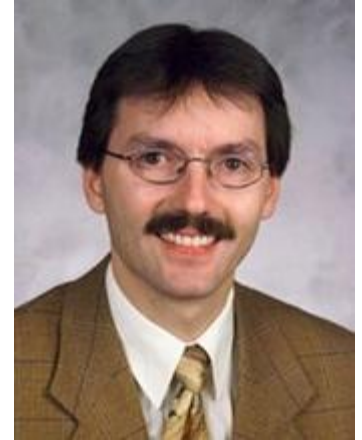
- intelligent computing (AI, machine learning)
- behavioral psychology (cognitive science, human factors)

Visual Analytics is the science of analytical reasoning supported by a highly interactive visual interface

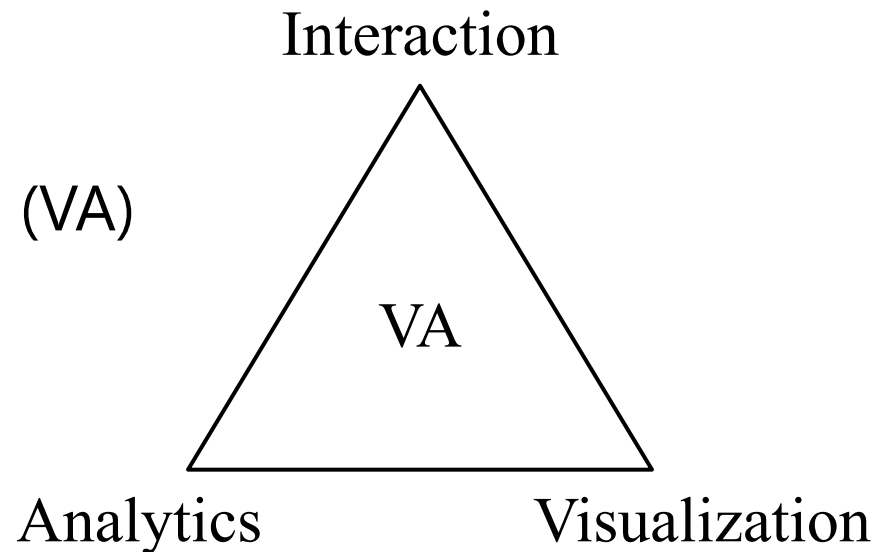
Visual Analytics Paradigm

The Daniel Keim Mantra of Visual Analytics

"Analyze First - Show the Important –
Zoom, Filter and Analyze Further -
Details on Demand"



The triangle of Visual Analytics (VA)



Intelligence Analysis

Intelligence analysis is challenging

Huge amounts of data

Many data type: text, images, video, sensor data, etc.

Uncertainty

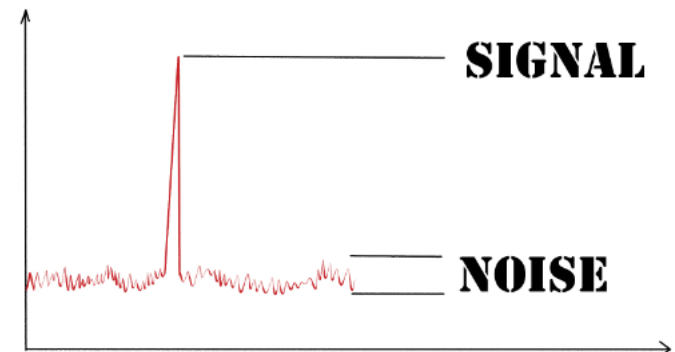
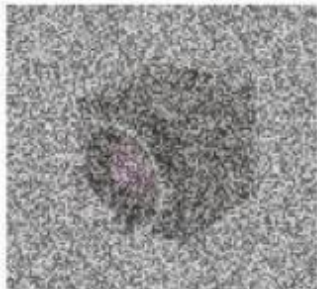
Contradictions

Omissions

Low signal to noise ratio (SNR)

Lower Signal to Noise Ratio

Higher Signal to Noise Ratio



Use of Visualization

Visual perception

- high bandwidth
- fast screening of a lot of data
- pattern recognition
- higher-level cognition

Interaction

- direct manipulation
- two-way communication

Recall intro lecture on the human visual system...

Use of Visualization

Visual perception

- high bandwidth
- fast screening of a lot of data
- pattern recognition
- higher-level cognition

Interaction

- direct manipulation
- two-way communication

Recall intro lecture on the human visual system...

But... humans are imperfect

Humans Are Imperfect

Humans tend to overlook/ignore non-focus (and unexpected) objects even when very close and obvious

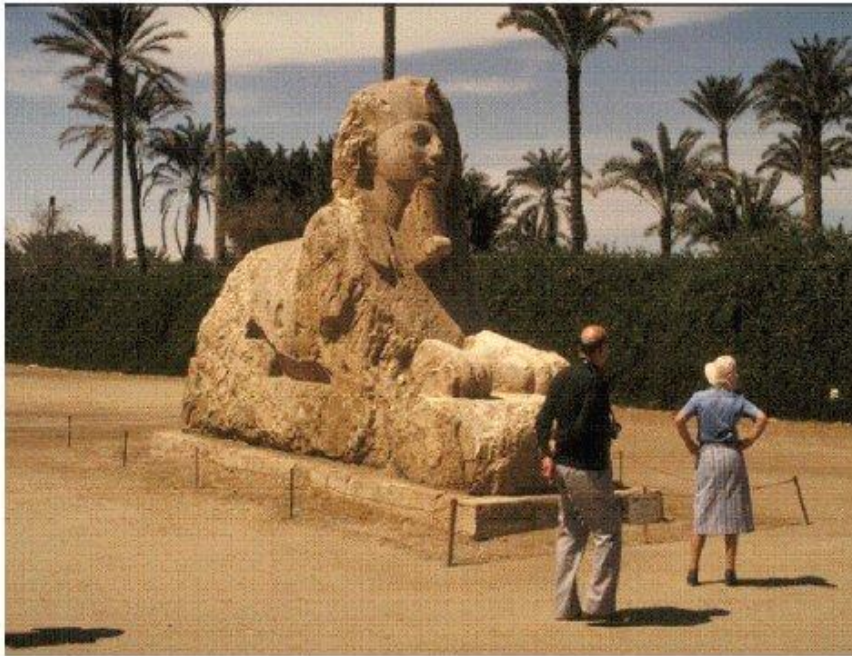
- note the Visual Analytics slogan: *Detect the Unexpected*

Humans also have limited working memory

- fine details are quickly forgotten when focus changes
- big effect in animated or interactive visualizations
- need to preserve temporal context

Example #1

Spot a difference?



This is called change blindness

Example #2

In this video you will do some counting.

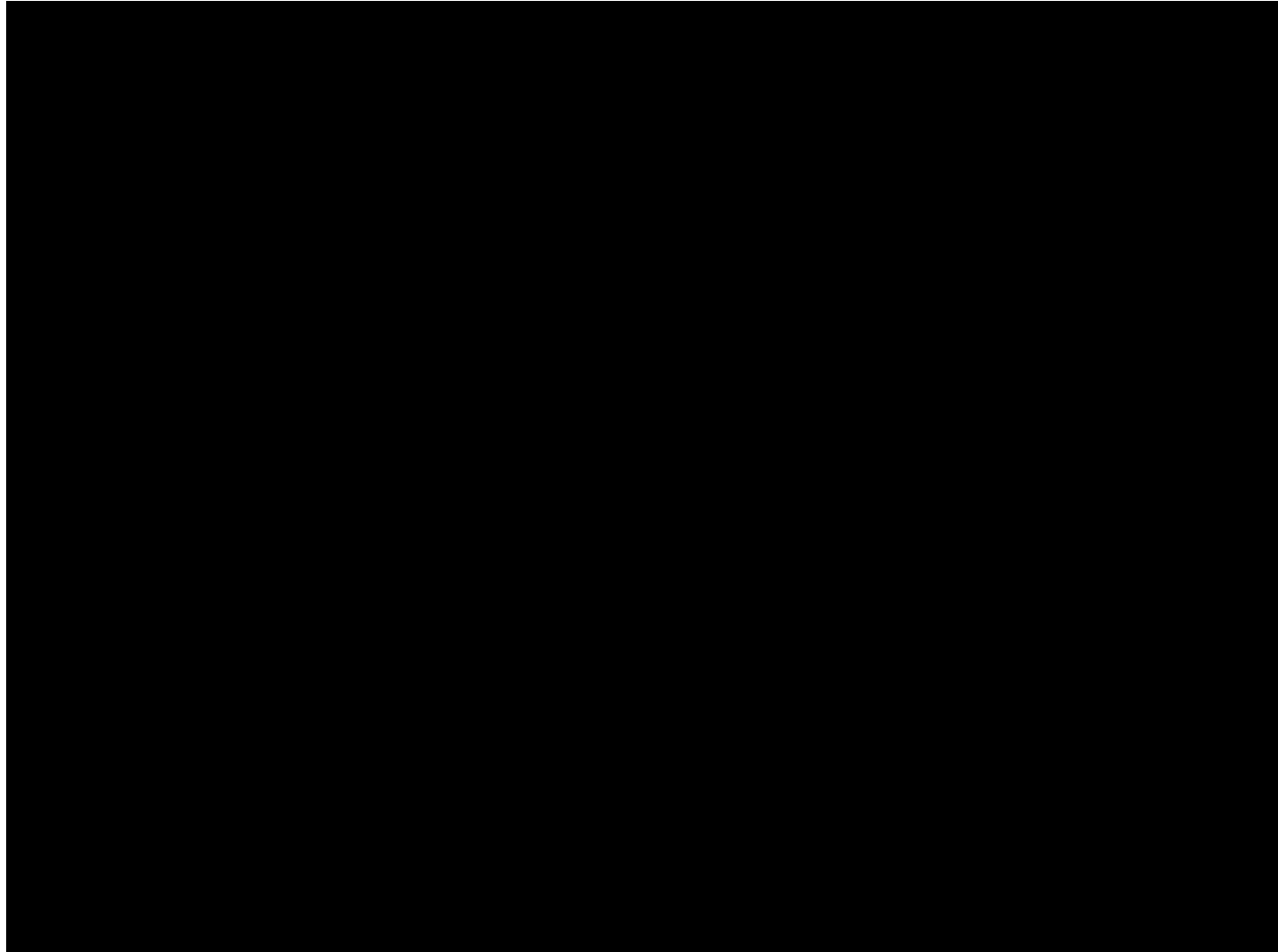
It is very important that you get the right number!

Ready?

[YouTube](#)

Video by Dan Simons (U Illinois)

Attention Experiment



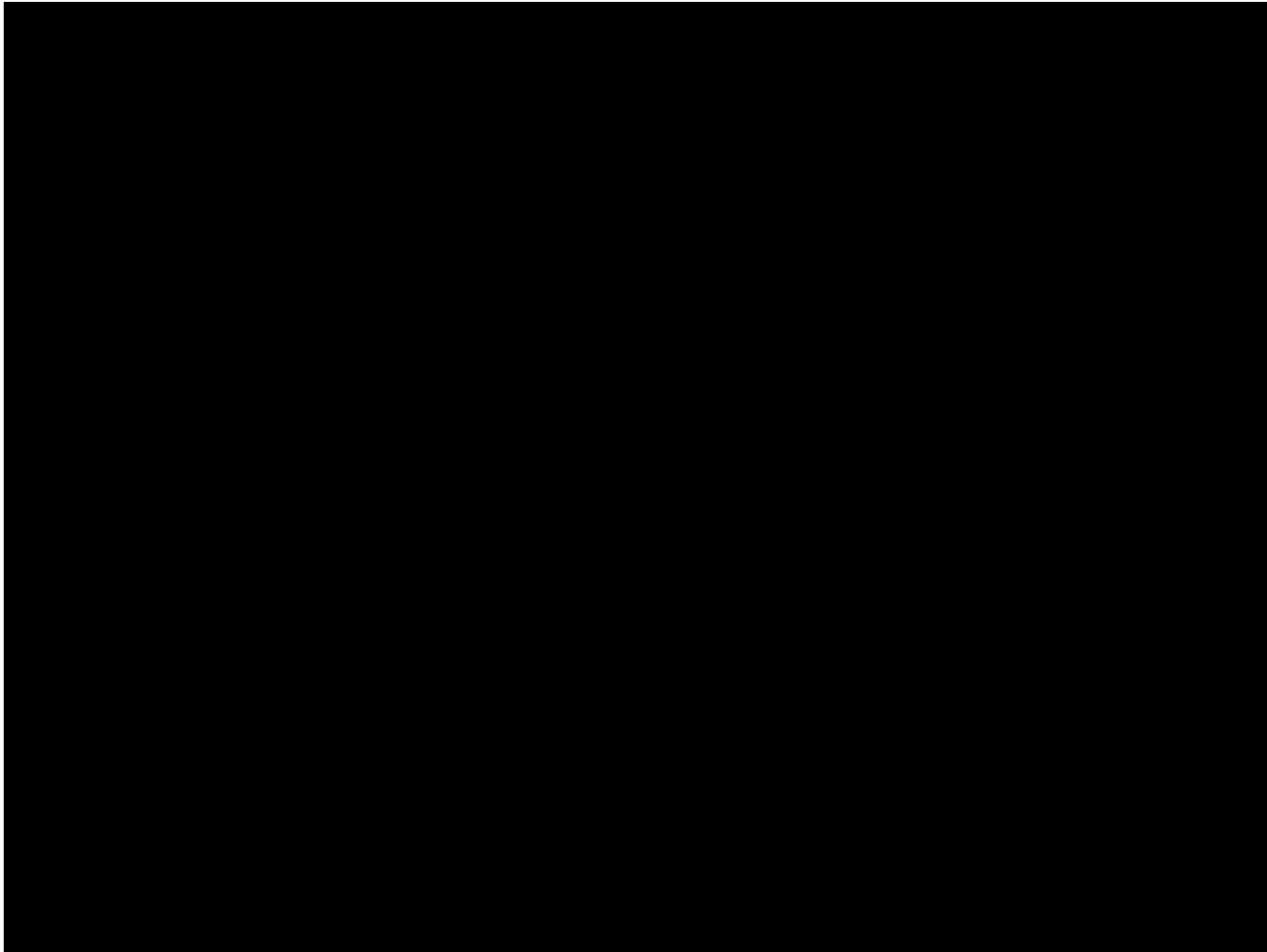
Example #3

Another distraction experiment

[YouTube](#)

Video by Dan Simons (U Illinois)

Distraction Experiment



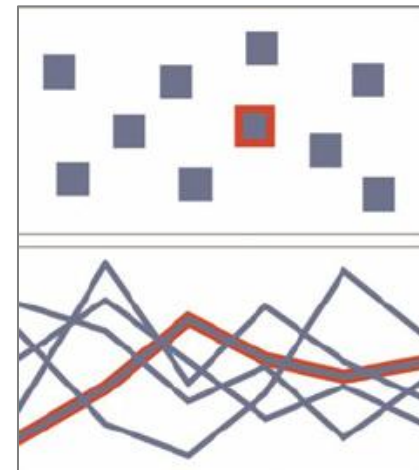
Change Blindness

Thoroughly studied by Dan Simons (U Illinois)

- see <http://www.dansimons.com/index.html>

Visual Analytics tools

- help human analysts cope with insufficient memory
 - visualizations externalize memory
 - allow humans to perform *visual queries* (see C. Ware book)
- help human analysts deal with change blindness
 - analytics can detect changes
 - visualization can highlight/emphasize these changes
- we have seen many visual tools this semester
 - this lecture is more about strategy building



Human Limitations

The Magic Number Seven

- ± 2 : the number of things most people can keep in working memory at one time
- [famous paper](#) by George A. Miller (1956)
- channel capacity 2.5 bits
- applies to letters, sounds, shapes, colors, etc.
- causes problems for complicated analysis
- reduce the problem by chunking
- words (vs. letters), bytes (vs. bits), clusters (vs. points), categories (vs. individual elements()), ...
- hierarchical decomposition, multi level of detail



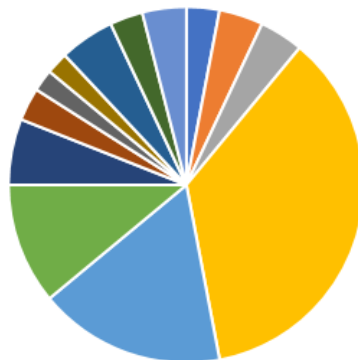
Human Limitations

The Magic Number Seven (7) for visualization

- not more than 7 segments in a pie chart
- not more than 7 colors in a line chart
- and so on

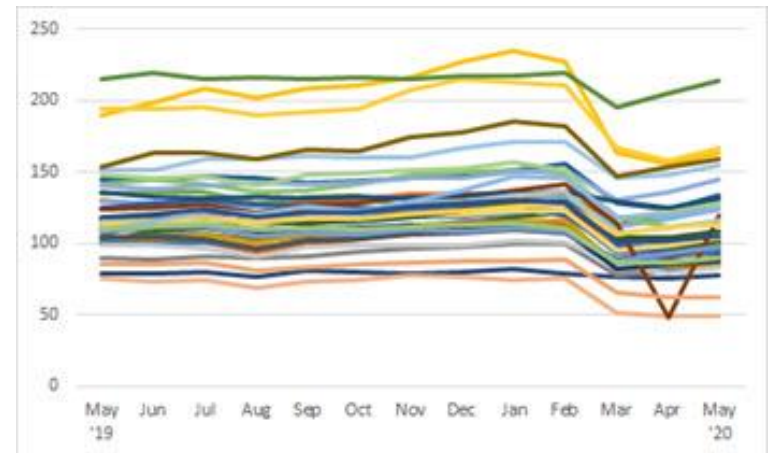
Number of developers per language

no

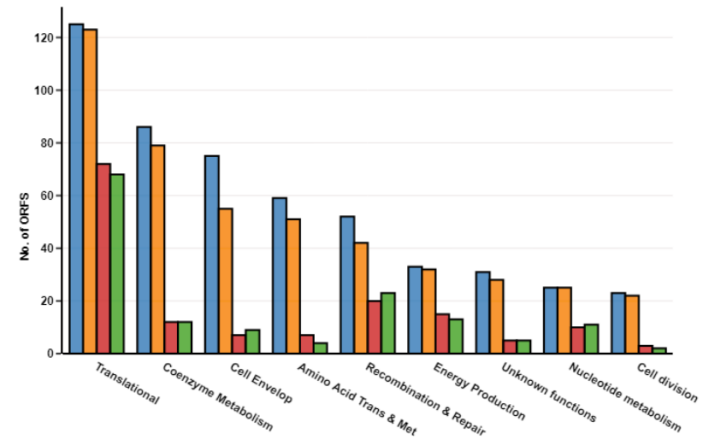
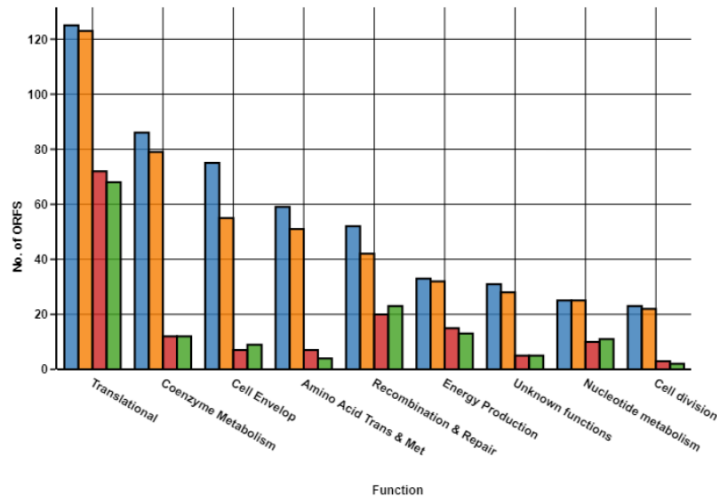


■ SAS ■ R ■ Python ■ Java ■ C ■ C# ■ Javascript ■ ODI ■ Weka ■ Hyperion ■ SSIS ■ Talend ■ Pentaho

no



Visual Complexity



do I need these grid lines?

maybe not

"Perfection is achieved not when there is nothing more to add, but when there is nothing left to take away."

— Antoine de Saint-Exupery.

Strategies for Dealing with Complexity

Decomposition

- decompose a complex problem into simpler problems
- get your thinking straight in these simpler problems

Externalization

- get the decomposed problem out of your head and down on paper or on a computer screen in some simplified form
- show the main variables, parameters, or elements of the problem and how they relate to each other

Recall principles of information visualization

- overview and detail
- focus and context
- analyze, filter, zoom,...

200 Years Ago... Ben Franklin's Letter



Mentioned his method of solving decision problems

Why is the decision problem so difficult?

- folks cannot keep all pros and cons in mind at the same time

Solution?

- write down all the pros and cons onto paper in some visible, shorthand form
- allows you make a global judgment effectively

The Sense-Making Loop

Support visualization with computations for data processing

Form a loop: visualize - refine

Gather (forage) information

Re-represent

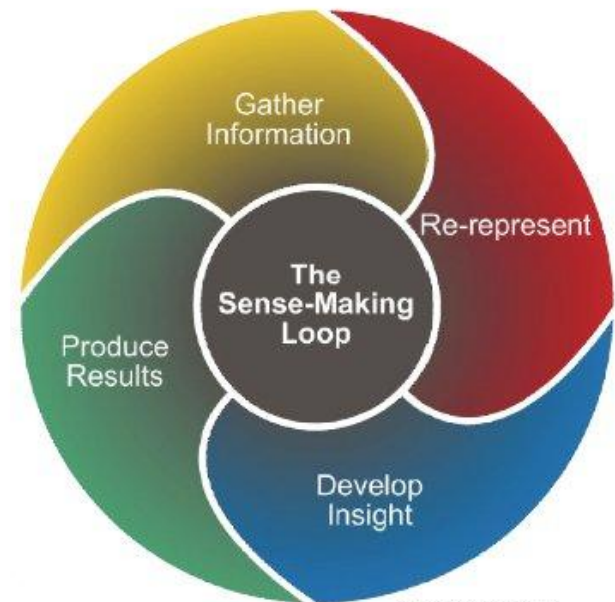
- choose a form that aids analysis

Develop insight

- through manipulation of representations

Produce results

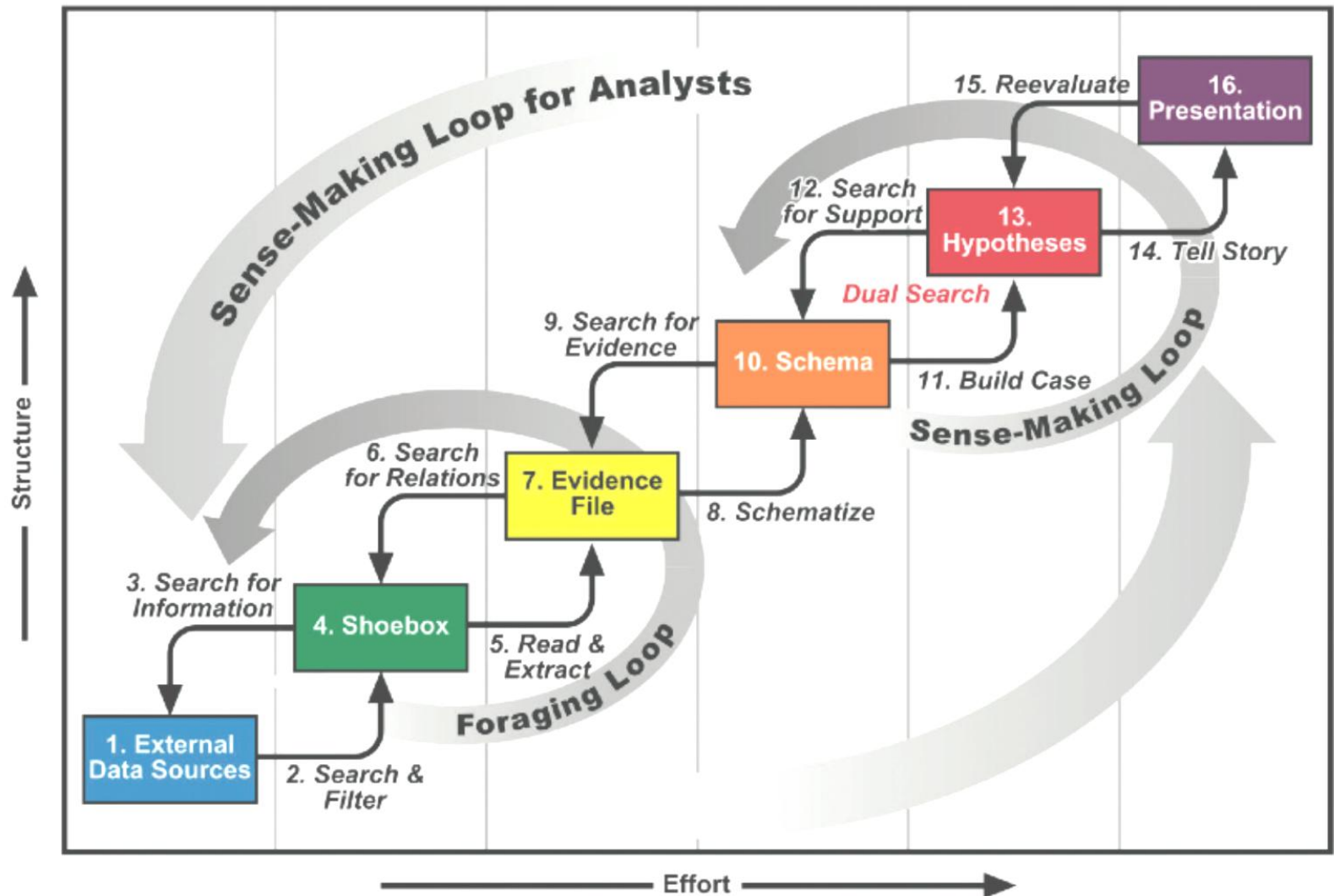
- “product”



Visual Analytics

<http://www.viscenter.uncc.edu/courses/visanalytics.html>

Nominal Sense-Making Process



Use Visualizations to Evoke The Right
Thoughts

How Many 9s Do You See?

3 3 0 3 0 1 8 7 6 8 2 1 4 0 3 8 3 7 7 2 0 5 2 3 2 7 0 2 0
7 1 4 6 0 2 1 3 2 7 6 0 2 5 6 3 2 5 7 6 3 3 0 2 0 3 0 7 2
8 7 5 7 2 8 3 8 7 7 8 2 0 7 7 5 2 3 1 1 5 6 3 8 4 7 8 2 0
0 5 0 5 1 6 1 7 5 6 8 0 4 4 6 7 4 7 1 4 0 0 8 4 4 3 0 3 2
2 4 3 1 3 5 4 9 5 0 7 6 0 7 4 3 1 8 2 7 3 4 6 0 2 4 8 2 3
8 6 2 2 6 5 4 6 7 0 7 6 0 0 3 9 0 2 4 7 1 7 2 3 3 5 8 7 0
0 8 4 5 1 3 1 7 6 4 5 4 1 2 4 5 3 3 5 4 9 6 7 7 6 3 4 2 5
4 7 7 0 2 2 0 1 1 7 7 7 0 2 6 6 4 7 5 8 6 1 4 3 7 8 5 4 6
4 3 6 6 4 6 6 2 8 4 8 5 3 7 8 8 1 3 8 5 4 5 7 4 0 3 2 8 4
5 5 0 3 5 3 5 3 8 3 2 3 8 2 3 1 6 2 7 2 4 6 3 6 4 4 3 2 5
4 4 0 2 1 7 2 4 4 7 4 1 9 2 4 5 2 5 0 4 0 0 5 3 6 3 3 6 7
7 4 6 6 8 7 5 7 9 2 0 2 8 8 8 8 3 2 4 2 6 4 0 4 6 3 7 2 1
0 1 7 1 5 9 1 4 2 8 7 3 7 1 4 5 1 8 7 8 0 5 1 7 0 5 8 8 1
2 8 5 2 1 2 8 7 7 6 2 5 6 2 6 4 1 5 1 6 1 2 1 1 0 5 6 4 0
2 1 1 7 7 2 0 0 1 8 7 0 2 9 0 2 8 5 7 8 4 6 0 6 5 0 7 1 2
0 5 2 4 1 5 3 3 1 5 5 1 4 0 1 6 4 3 3 9 8 8 3 4 6 8 4 8 6
7 3 7 5 2 4 0 2 7 6 3 8 5 5 4 5 8 8 7 5 5 6 5 6 7 9 7 7 4
0 3 2 8 1 4 4 6 0 8 2 3 0 1 3 4 6 2 0 5 7 7 3 6 1 8 7 3 5
4 4 8 3 3 3 5 0 1 0 3 8 6 3 2 0 5 0 6 1 3 3 4 3 6 1 5 8 6
1 0 2 2 7 6 3 3 0 8 8 0 3 1 8 8 1 2 1 7 5 2 9 3 5 8 3 2 5

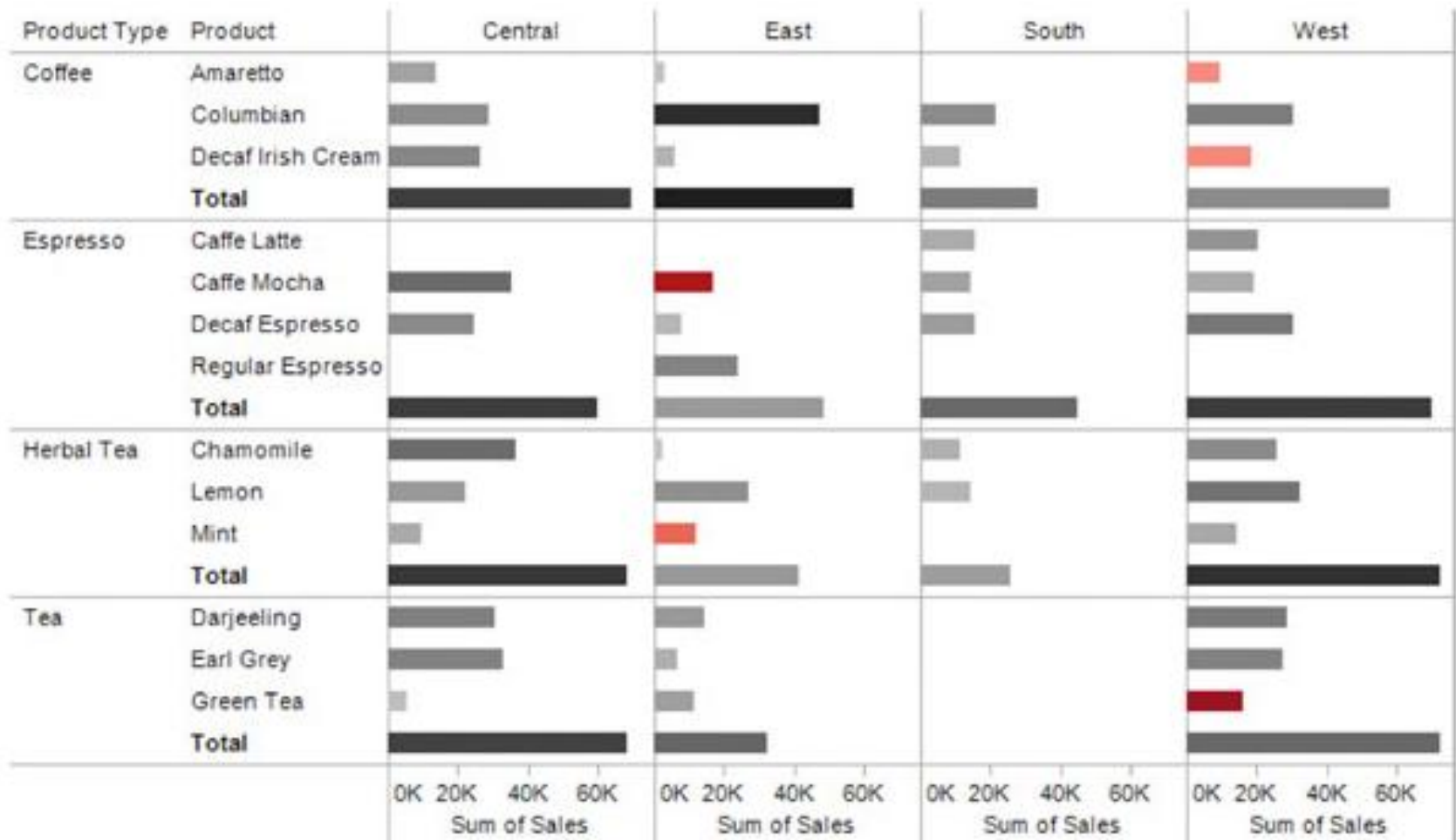
How Many 9s Do You See?

3	3	0	3	0	1	8	7	6	8	2	1	4	0	3	8	3	7	7	2	0	5	2	3	2	7	0	2	0
7	1	4	6	0	2	1	3	2	7	6	0	2	5	6	3	2	5	7	6	3	3	0	2	0	3	0	7	2
8	7	5	7	2	8	3	8	7	7	8	2	0	7	7	5	2	3	1	1	5	6	3	8	4	7	8	2	0
0	5	0	5	1	6	1	7	5	6	8	0	4	4	6	7	4	7	1	4	0	0	8	4	4	3	0	3	2
2	4	3	1	3	5	4	9	5	0	7	6	0	7	4	3	1	8	2	7	3	4	6	0	2	4	8	2	3
8	6	2	2	6	5	4	6	7	0	7	6	0	0	3	9	0	2	4	7	1	7	2	3	3	5	8	7	0
0	8	4	5	1	3	1	7	6	4	5	4	1	2	4	5	3	3	5	4	9	6	7	7	6	3	4	2	5
4	7	7	0	2	2	0	1	1	7	7	7	0	2	6	6	4	7	5	8	6	1	4	3	7	8	5	4	6
4	3	6	6	4	6	6	2	8	4	8	5	3	7	8	8	1	3	8	5	4	5	7	4	0	3	2	8	4
5	5	0	3	5	3	5	3	8	3	2	3	8	2	3	1	6	2	7	2	4	6	3	6	4	4	3	2	5
4	4	0	2	1	7	2	4	4	7	4	1	9	2	4	5	2	5	0	4	0	0	5	3	6	3	3	6	7
7	4	6	6	8	7	5	7	9	2	0	2	8	8	8	8	3	2	4	2	6	4	0	4	6	3	7	2	1
0	1	7	1	5	9	1	4	2	8	7	3	7	1	4	5	1	8	7	8	0	5	1	7	0	5	8	8	1
2	8	5	2	1	2	8	7	7	6	2	5	6	2	6	4	1	5	1	6	1	2	1	1	0	5	6	4	0
2	1	1	7	7	2	0	0	1	8	7	0	2	9	0	2	8	5	7	8	4	6	0	6	5	0	7	1	2
0	5	2	4	1	5	3	3	1	5	5	1	4	0	1	6	4	3	3	9	8	8	3	4	6	8	4	8	6
7	3	7	5	2	4	0	2	7	6	3	8	5	5	4	5	8	8	7	5	5	6	5	6	7	9	7	7	4
0	3	2	8	1	4	4	6	0	8	2	3	0	1	3	4	6	2	0	5	7	7	3	6	1	8	7	3	5
4	4	8	3	3	3	5	0	1	0	3	8	6	3	2	0	5	0	6	1	3	3	4	3	6	1	5	8	6
1	0	2	2	7	6	3	3	0	8	8	0	3	1	8	8	1	2	1	7	5	2	9	3	5	8	3	2	5

Who has the best profit and who has the worst sales?

Product Type	Product	Central		East		South		West	
		Sum of Profit	Sum of Sales	Sum of Profit	Sum of Sales	Sum of Profit	Sum of Sales	Sum of Profit	Sum of Sales
Coffee	Amaretto	\$5,105	\$14,011	\$1,009	\$2,993			(\$1,225)	\$9,265
	Columbian	\$8,528	\$28,913	\$27,253	\$47,386	\$8,767	\$21,664	\$11,253	\$30,357
	Decaf Irish Cream	\$9,632	\$26,155	\$2,727	\$6,261	\$2,933	\$11,592	(\$1,305)	\$18,235
	Total	\$23,265	\$69,080	\$30,989	\$56,640	\$11,700	\$33,256	\$8,724	\$57,856
Espresso	Caffe Latte					\$3,872	\$15,442	\$7,502	\$20,458
	Caffe Mocha	\$14,640	\$35,218	(\$6,230)	\$16,646	\$5,201	\$14,163	\$4,064	\$18,876
	Decaf Espresso	\$8,860	\$24,485	\$2,410	\$7,722	\$5,930	\$15,384	\$12,302	\$30,578
	Regular Espresso			\$10,062	\$24,036				
	Total	\$23,500	\$59,703	\$6,242	\$48,405	\$15,003	\$44,989	\$23,868	\$69,911
Herbal Tea	Chamomile	\$14,434	\$36,570	\$765	\$2,194	\$3,180	\$11,186	\$8,852	\$25,632
	Lemon	\$6,251	\$21,978	\$7,901	\$27,176	\$2,593	\$14,497	\$13,120	\$32,274
	Mint	\$4,069	\$9,337	(\$2,242)	\$11,992			\$4,330	\$14,380
	Total	\$24,754	\$67,885	\$6,424	\$41,362	\$5,774	\$25,683	\$26,301	\$72,285
Tea	Darjeeling	\$10,772	\$30,289	\$6,497	\$14,096			\$11,780	\$28,769
	Earl Grey	\$10,331	\$32,881	\$3,405	\$6,505			\$10,425	\$27,387
	Green Tea	\$1,227	\$5,211	\$5,654	\$11,571			(\$7,109)	\$16,063
	Total	\$22,330	\$68,380	\$15,557	\$32,172			\$15,097	\$72,220

Who has the best profit and who has the worst sales?



Do The Right Analytics, Don't Just
Visualize Data

Doubling down on states for strong growth

Maria
Senior Sales Analyst
March 15th, 2012

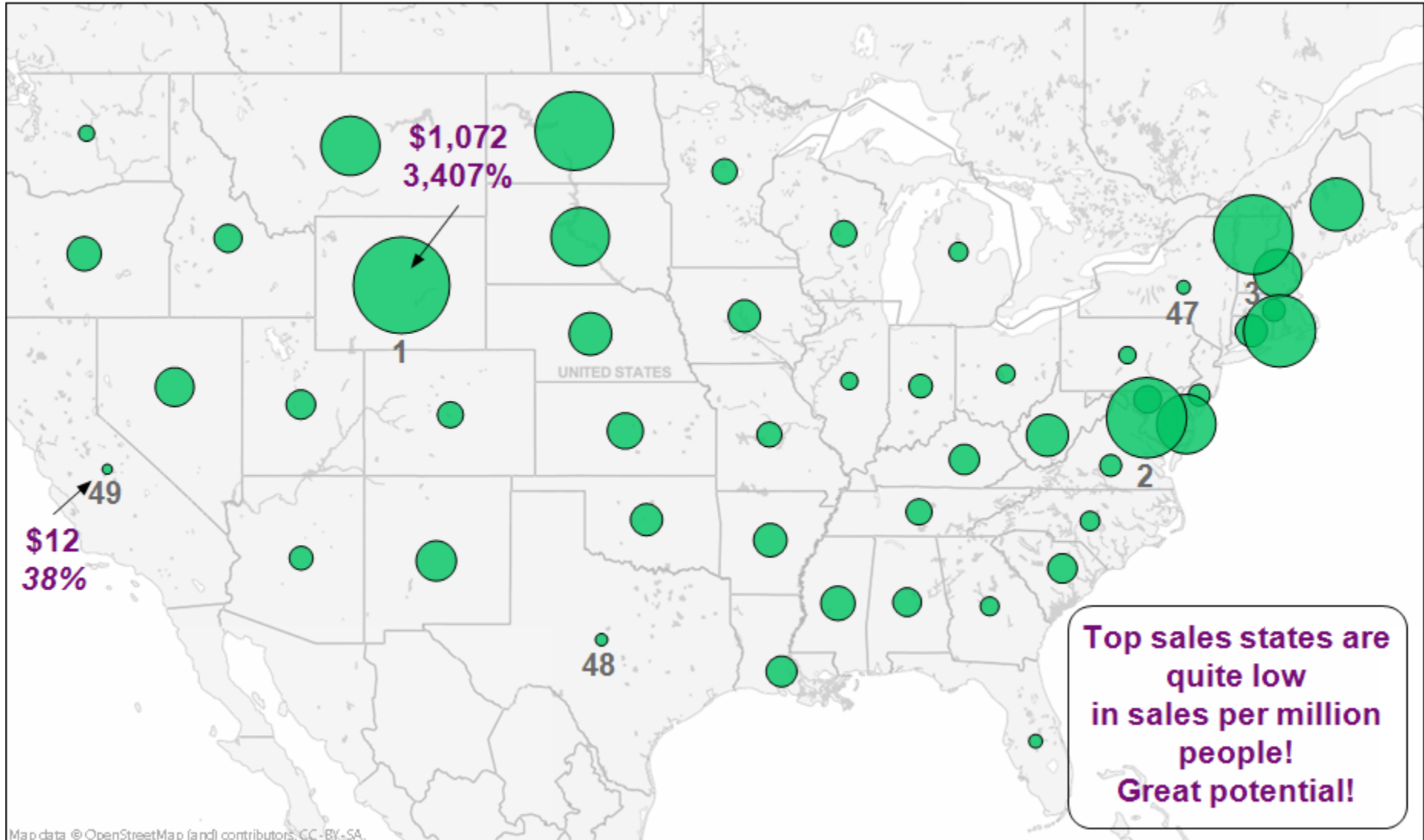
Today's question

In which states should we invest additional marketing dollars during the upcoming campaign?

Based upon sales growth potential...

Sales per State/Capita

2011 sales per million residents by state (top/bottom 3 labeled)



Potential sales by state???

+Is there a better metric?

+The emphasis is on **potential**

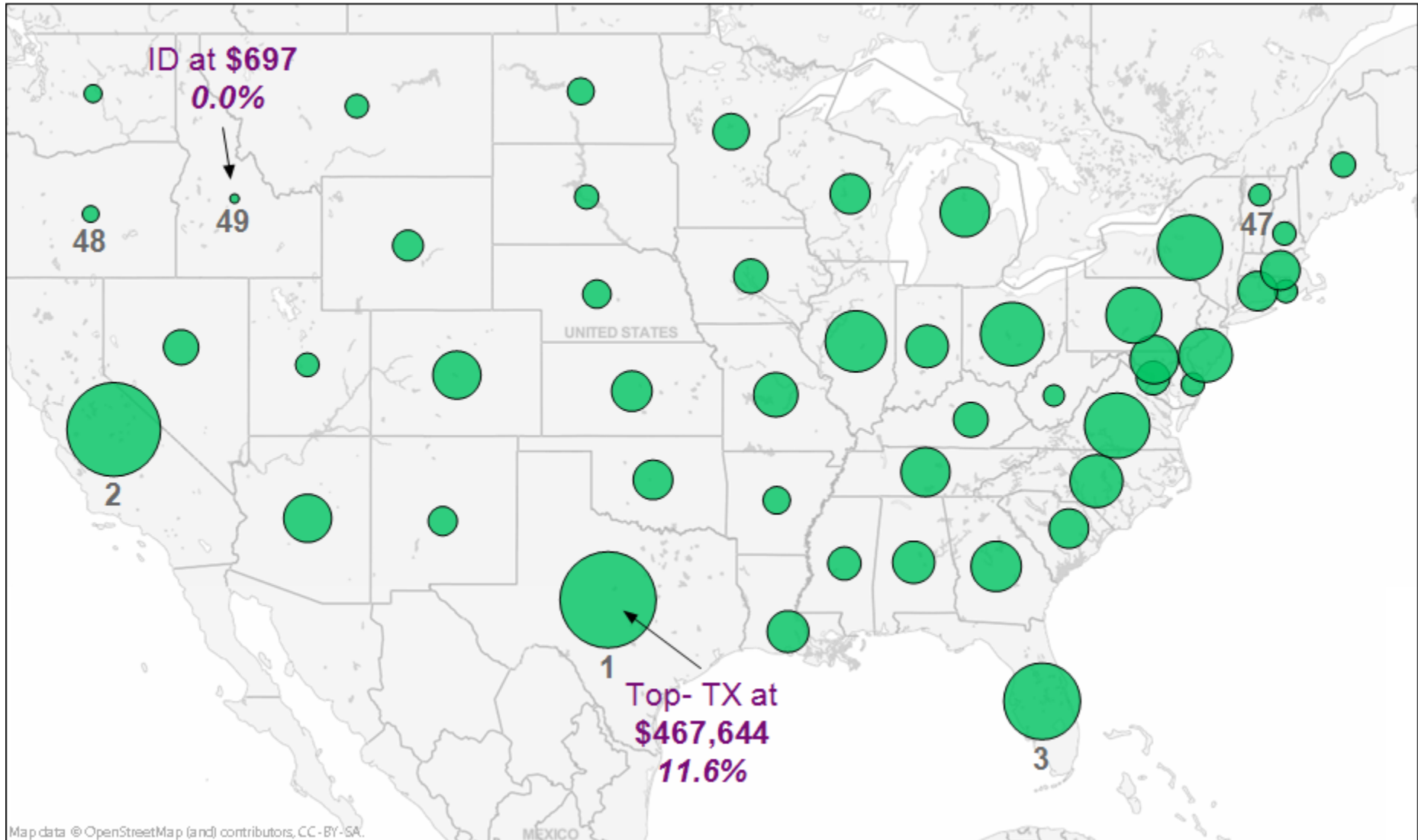
Average sale per capita for top states

multiplied by

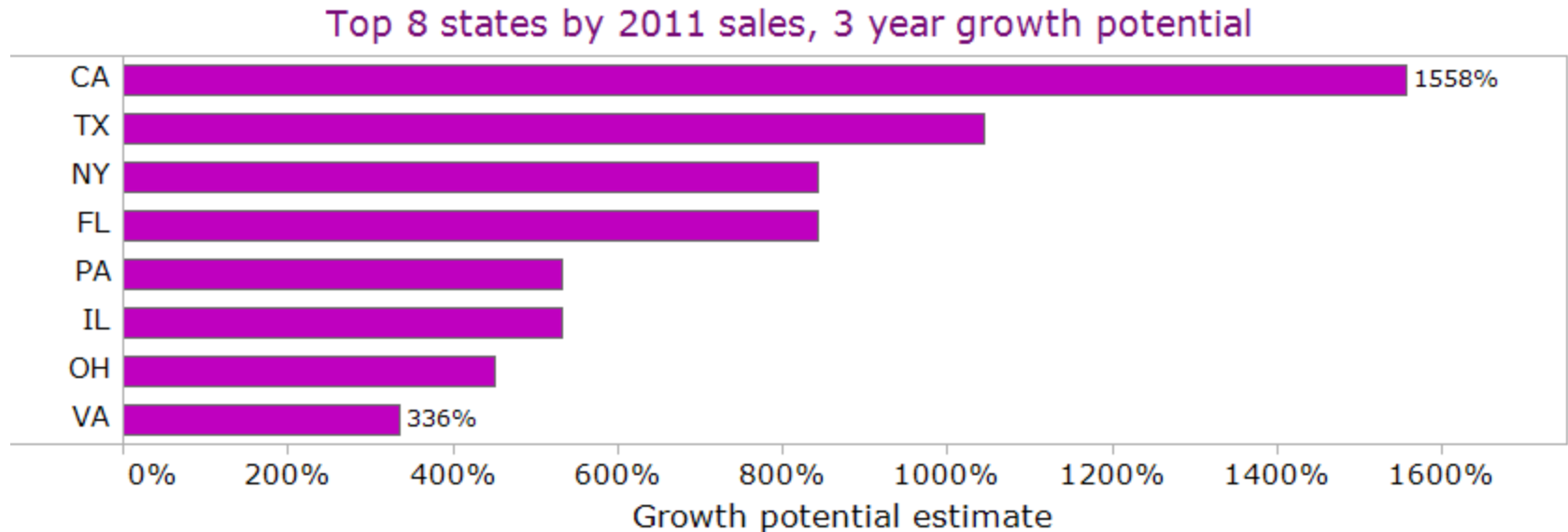
Current population of top sales states

Sales per State/Capita x Capita

2011 sales per state (top/bottom 3 labeled)



Highest growth potential in top 8

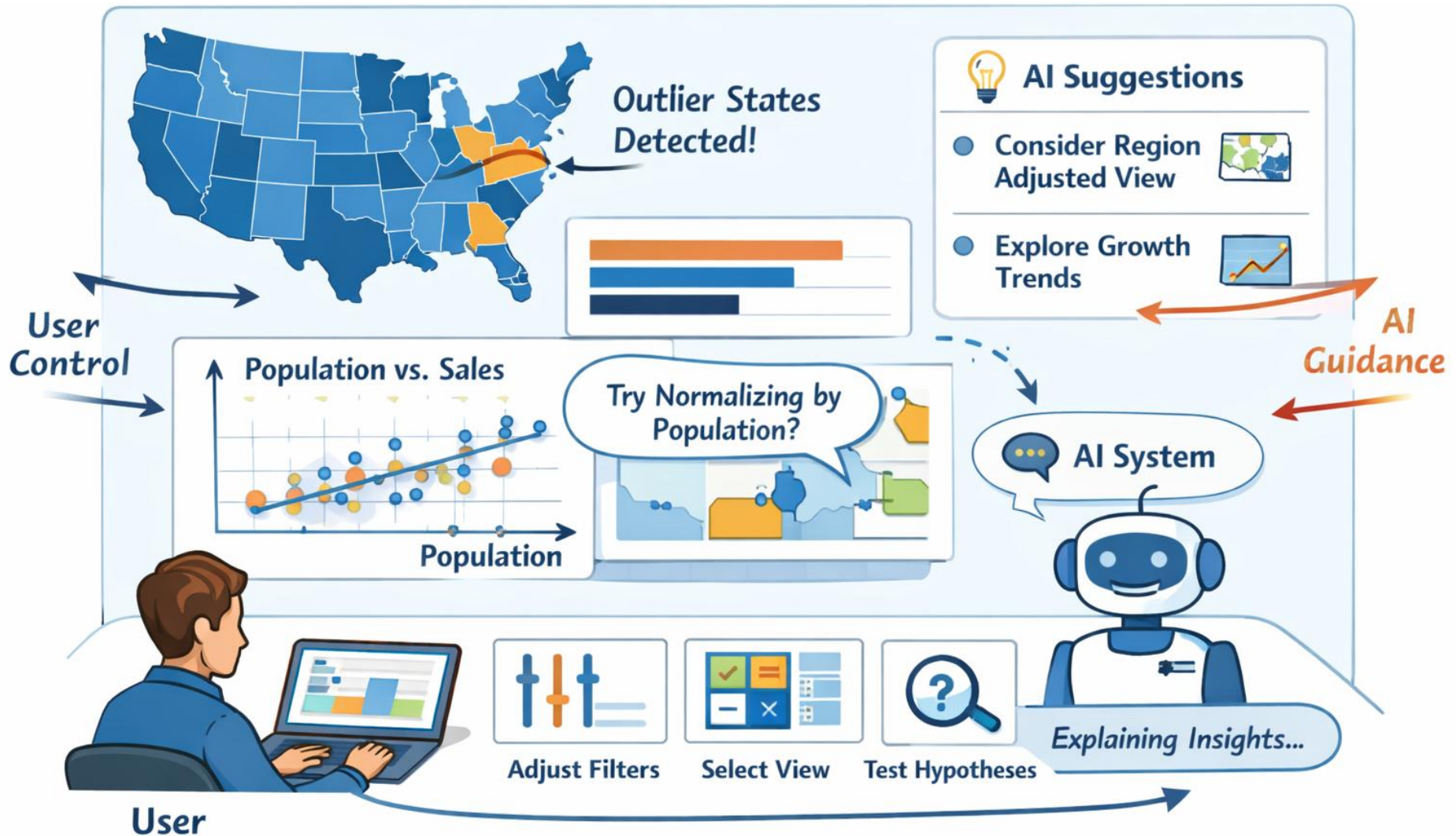


+ If we were to pick just one state,
California has the greatest potential

+ The next tier is Texas, New York & Florida

Visual Analytics in the Era of Human-Centered AI

Human-Centered (AI-Powered) VA



What is Human-Centered VA

Let's have a look at a specific example

Exploratory Data Analysis Dashboard

- Task: design a visualization system to help analysts explore a large dataset (e.g., sales, public health, or crime data)

Three options

- System A — “Visualization-only” (not human-centered AI)
- System B — “Auto-AI dashboard” (also not human-centered)
- System C — Human-Centered Visual Analytics

System A – Visualization-only

Shows:

- scatterplots
- bar charts
- maps

User must:

- manually explore
- try filters
- guess what to look for

Problems:

user may **miss important patterns**

suffers from:

- change blindness
- limited attention

no guidance

System B –Auto-AI dashboard

System automatically:

- highlights one “important” pattern
- e.g., “Sales are highest in California”

No explanation

No alternatives

No user control

Problem:

- **overly directive**
- hides other insights
- user becomes passive

System C – Human-Centered VA

Visualization + AI suggestions:

The system:

- highlights: “States with unusually high growth (outliers)”
- suggests: “Try normalizing by population”
- automatically shows: alternative views (map ↔ scatterplot ↔ ranking)

The user can:

- accept / reject suggestions
- adjust parameters
- explore other hypotheses

The system also:

- explains: *why* something is highlighted
- exposes: uncertainty, competing patterns

Take-Aways

Human-centered visual analytics does not just show data and does not just show answers—it guides exploration while keeping the human in control.

Addresses human limitations

- AI points to relevant patterns
- visualization externalizes memory

Supports sensemaking loop

- user iterates: view → suggestion → refine → insight

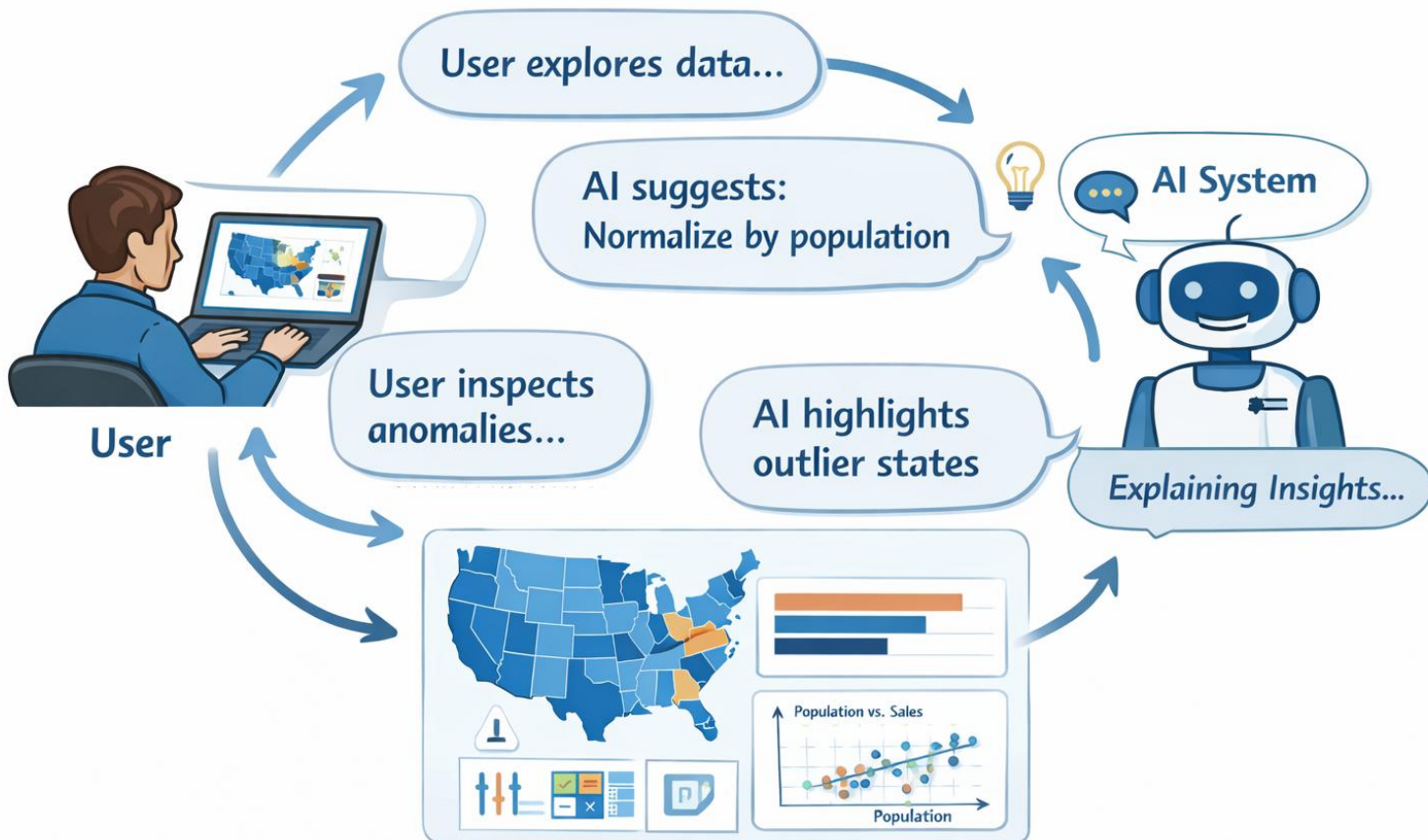
Encourages multiple hypotheses (ACH!)

- system suggests different perspectives
- user evaluates them

Finally – Mixed Initiative VA

Mixed Initiative Visual Analytics

Human and AI take turns steering the analysis



Collaborative VA

Collaborative VA

Supports joint sensemaking across analysts

- Combines diverse perspectives and expertise
- Insight evolves through alignment and negotiation
- Collaboration is an ongoing, dynamic process
- Can be remote or local on a large, shared display

Three phases:

- Phase 1: Individual sensemaking and hypothesis formation
- Phase 2: Sharing and integration of perspectives
- Phase 3: Collaborative reasoning

S. Brennan, K. Mueller, G. Zelinsky, IV Ramakrishnan, D. Warren, A. Kaufman, Toward a Multi-Analyst, Collaborative Framework for Visual Analytics Framework, *IEEE Symposium on Visual Analytics Science and Technology*, 2006 ([link](#))

Three Phases

Phase 1:

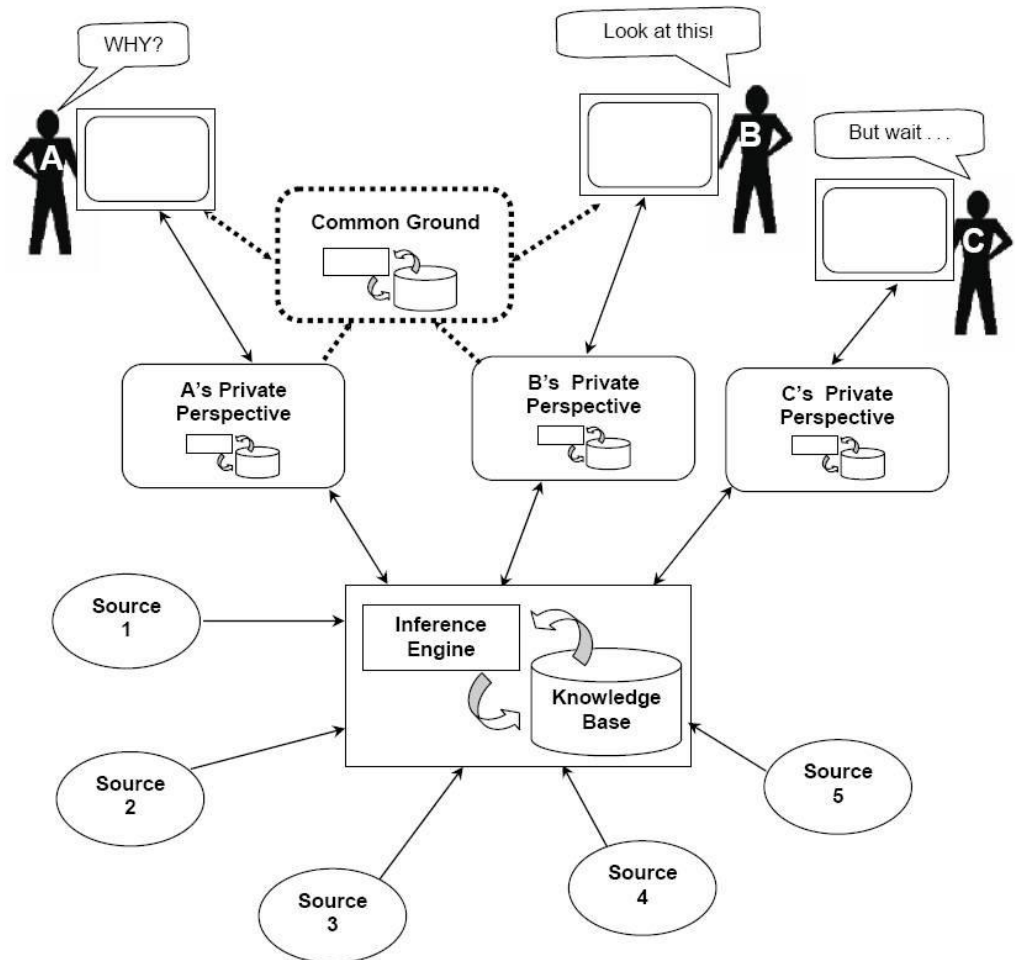
Individual sensemaking and hypothesis formation

Phase 2:

Sharing and integration of perspectives

Phase 3:

Collaborative reasoning



Example: Collaborative VA to Mitigate Bias of AI Decision Models

Problem: Fairness is not one thing

- Fairness in decision-making is critical
- But... Stakeholders have different, often conflicting fairness goals
- No single agreed-upon definition of fairness
→ Hard to reach consensus

Limitation of current approaches

- enforce one fairness standard
- ignore competing perspectives

Collaborative VA can facilitate

- Fairness as negotiated entity with all stakeholders at the table

Solution: FairPlay

FairPlay: Collaborative Fairness Analysis

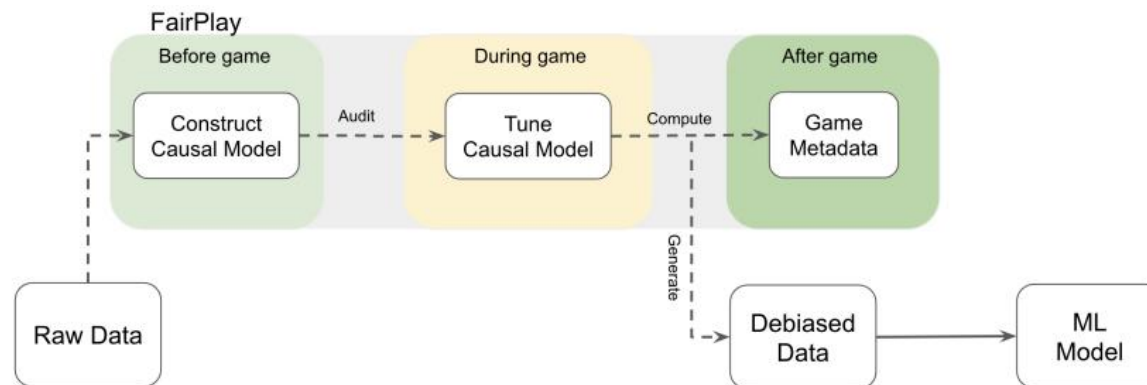
- Web-based system for collaborative dataset debiasing

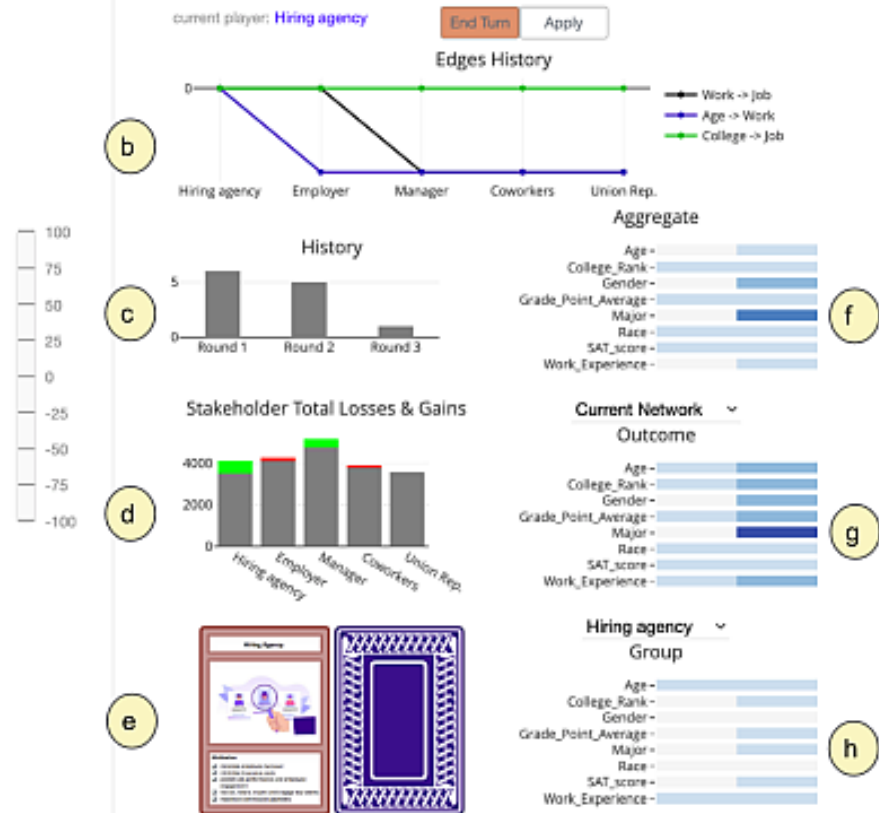
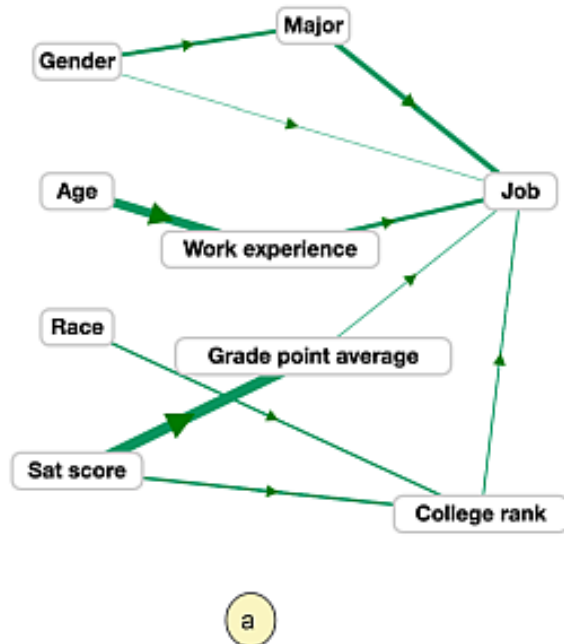
Stakeholders interact to

- adjust data used for AI model training by editing a causal network
- observe impact on allocation of resources among groups (here, jobs)
- explore fairness metrics

System enables

- comparison of perspectives among stakeholders
- structured negotiation by turn taking in the network editing
- movement toward consensus (users typically reach agreement in ~5 rounds)





(a) causal network link editor, (b) edge history chart, (c) aggregate edge history chart, (d) stakeholder total loss and gain chart, (e) active stakeholder card stack, (f) aggregate attribute disparity chart, (g) attribute outcome chart, (h) stakeholder attribute priority chart.

Provenance

Provenance: Tracking the Analysis Process

Record of how insight was obtained

- filters and edits applied
- views explored
- hypotheses tested
- decisions made

Why it matters

- supports memory (humans forget)
- enables reproducibility
- explains *how* a conclusion was reached

In collaborative settings

- tracks who did what
- captures evolving viewpoints
- supports alignment and accountability
- Example: FairPlay's dashboard supports provenance in several ways

Provenance is the memory of sensemaking