CSE 564
Introduction to Visualization

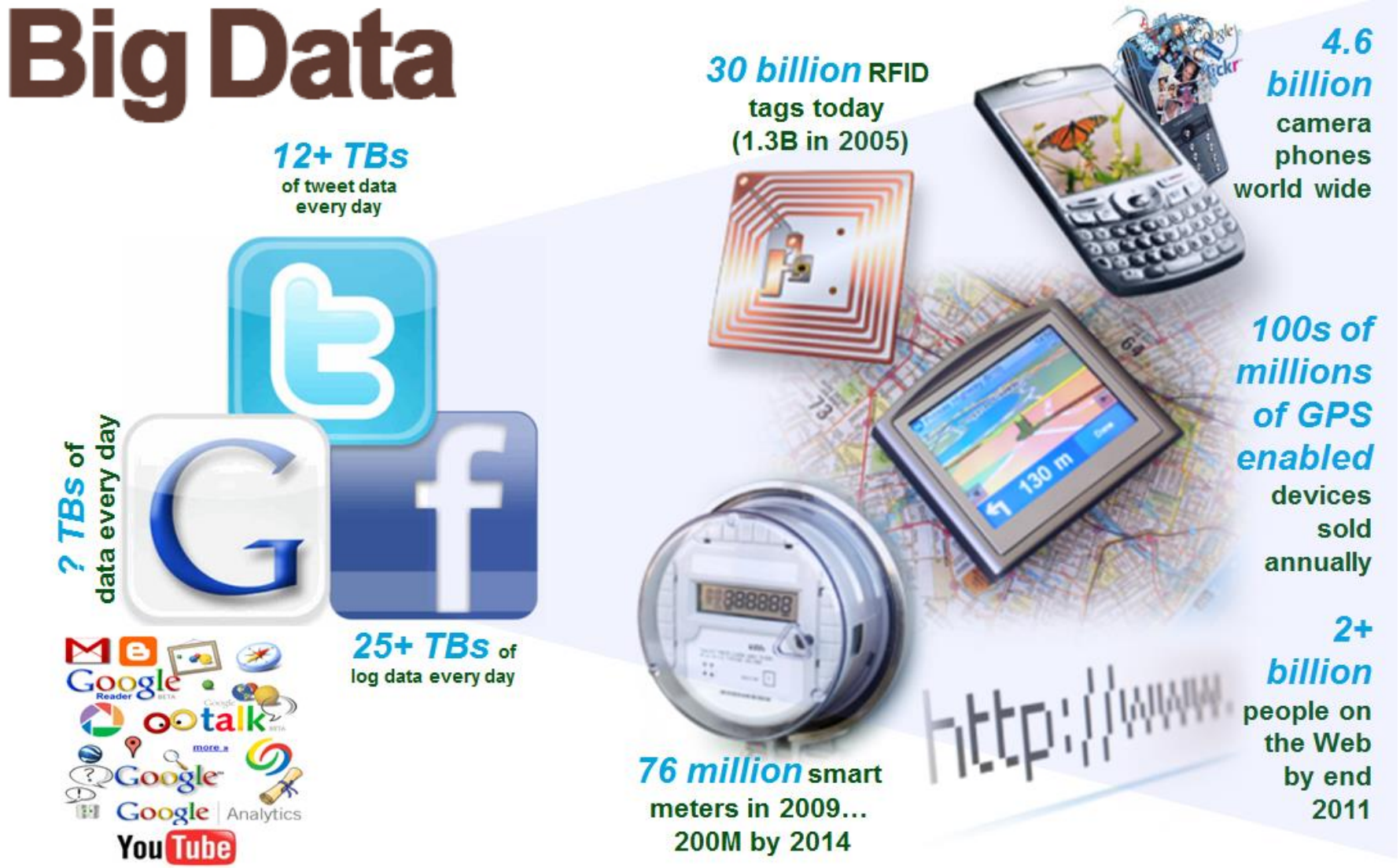
Visual Analytics & The Visual Sense Making Process

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)	
4	AI-assisted coding for VIS applications (design, debugging, refactoring)	Project #1 out
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	
8	Perception and cognition (human visual system, color, contrast)	Project #2(a) out
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	
12	Visualization and AI: mutual support and capabilities (VIS4AI, AI4VIS)	Project #2(b) out
13	Principles of interaction: drive what is visualized, analyzed & how (HCI4VIS)	
14	Visual analytics, human-centered AI, mixed-initiative & collaborative VA	
15	Midterm #1 (tentative date)	
16	VA system design and evaluation, the nested model, uncertainty	
17	Midterm #1 discussion (tentative date)	Final proj. proposal call out
18	Visualization of hierarchical data	
19	Visualization of maps and data with geo-reference	
20	Visualization of graphs, networks (incl. derivation of causal networks)	Final project proposal due
21	Vis. of time-varying, time-series, streaming data, progressive visualization	
22	Visualization of text, LLMs, and semantic data	
23	Ed Tufte revisited: principles, critiques and limits, responsible visualization	
24	Design of effective infographics	Final proj. prelim report due
25	Foundations scientific and medical visualization, intro to volume rendering	
26	Scientific visualization	Bonus project out (Vol Ren)
27	Story telling with data, data journalism	
28	Midterm #2 (tentative date)	
Final	Final project demo on zoom (public)	All final proj. materials due

Why Visual Analytics?



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

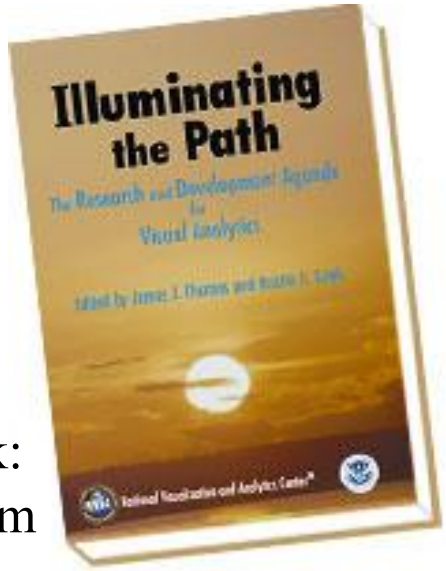
Amplify Perception

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)

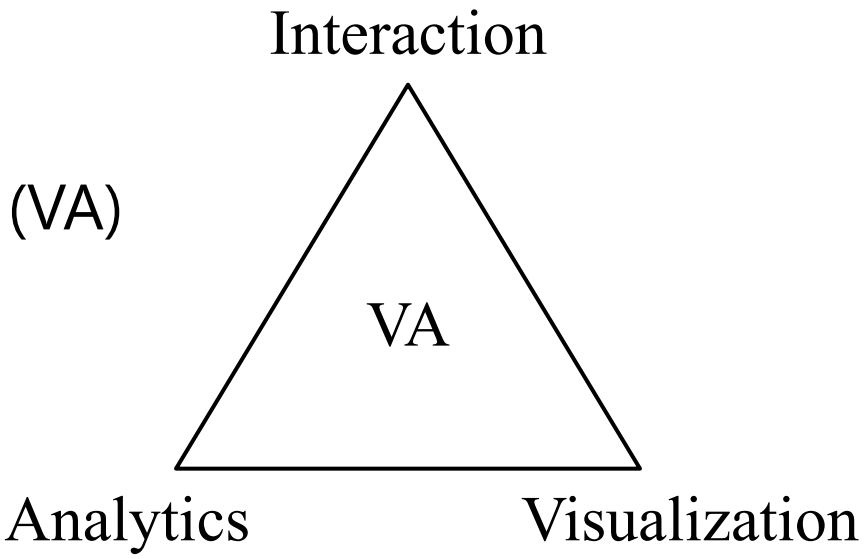
Visual Analytics is the science of analytical reasoning supported by a highly interactive visual interface

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

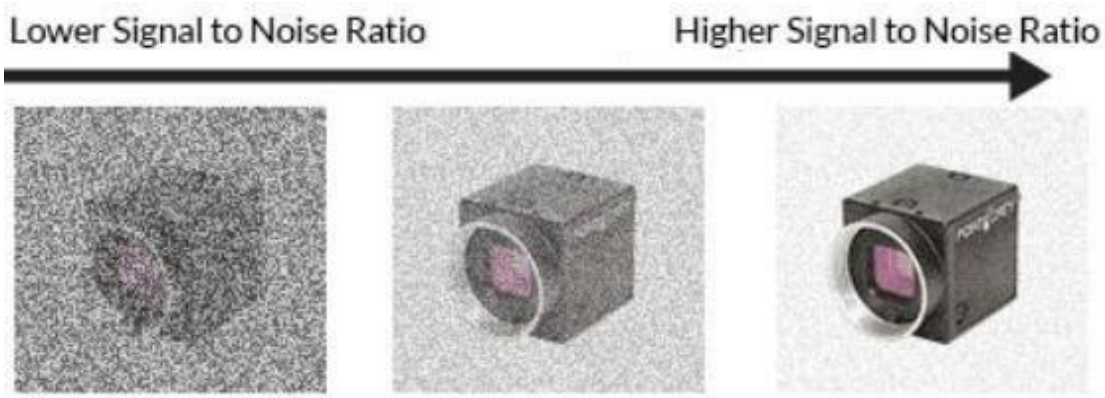
Many data type: text, images, video, sensor data, etc.

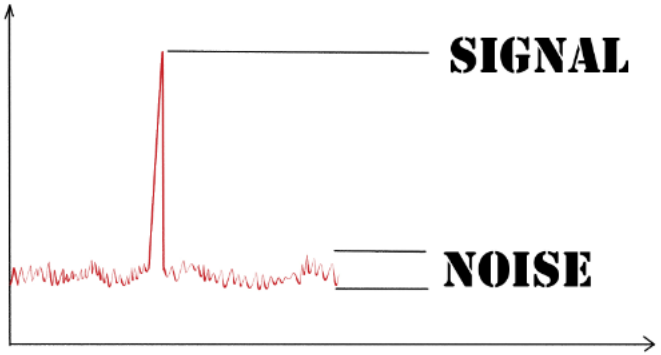
Uncertainty

Contradictions

Omissions

Low signal to noise ratio (SNR)





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

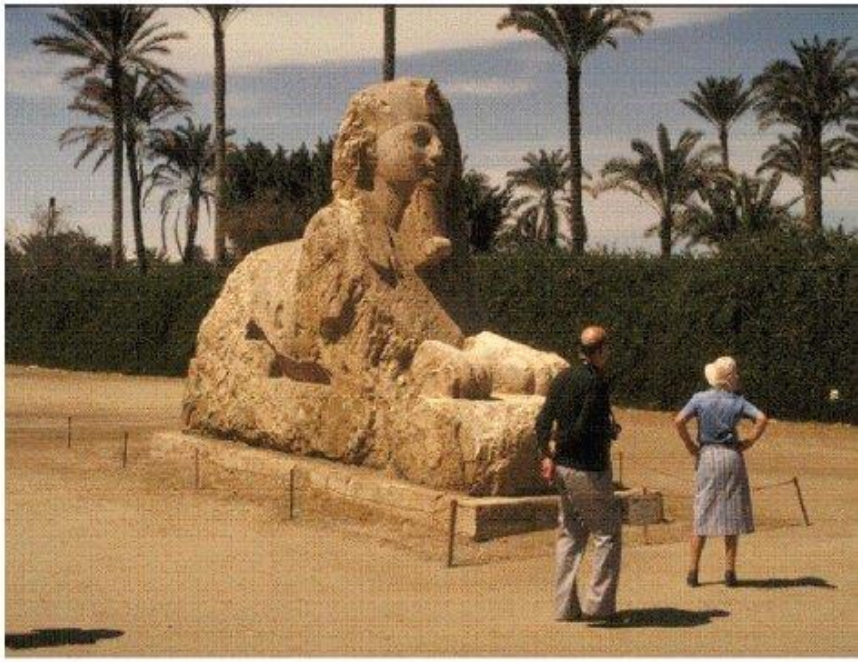
- note the Visual Analytics slogan: *Detect the Unexpected*

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.

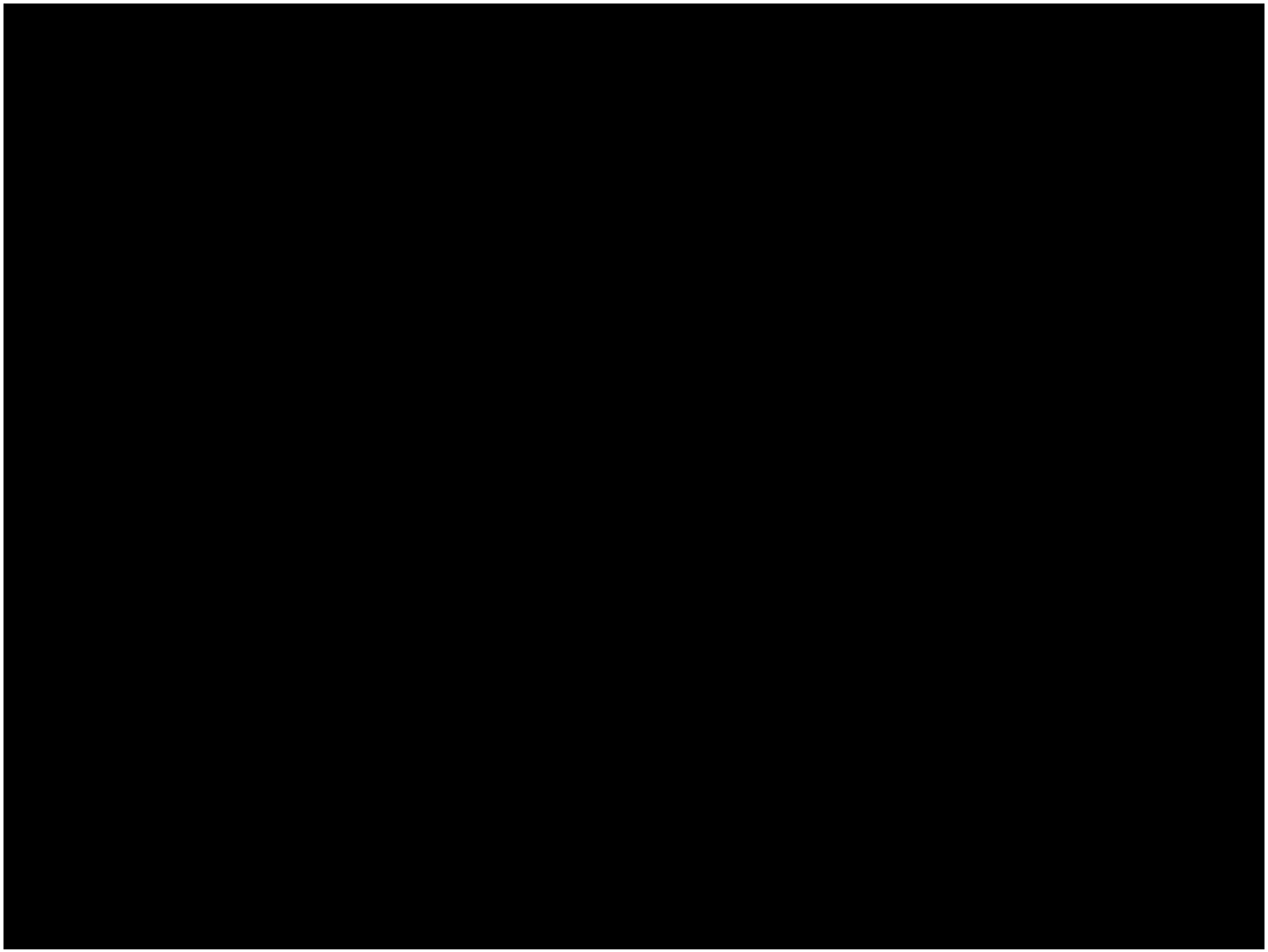
It is very important that you get the right number!

Ready?

YouTube

Video by Dan Simons (U Illinois)

Attention Experiment



Example #3

Another distraction experiment

YouTube

Video by Dan Simons (U Illinois)

Distraction Experiment

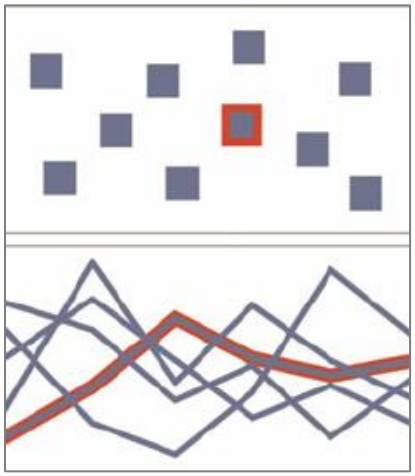
Change Blindness

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

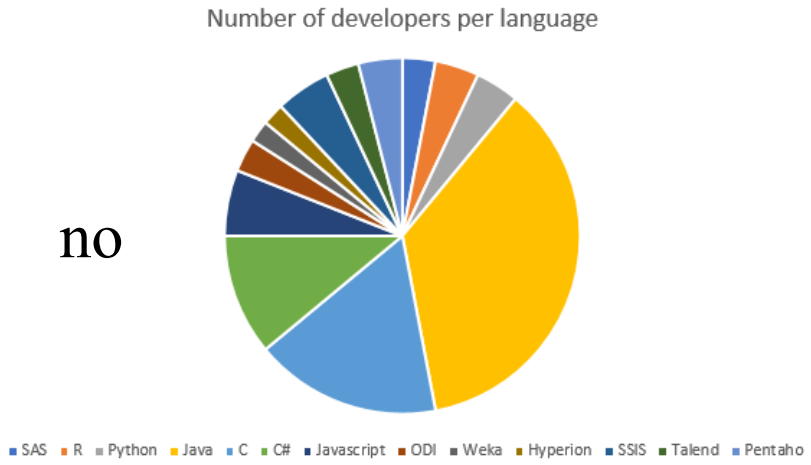
- reduce the problem by chunking
- words (vs. letters), bytes (vs. bits), clusters (vs. points), categories (vs. individual elements(), ….
- hierarchical decomposition, multi level of detail

Human Limitations

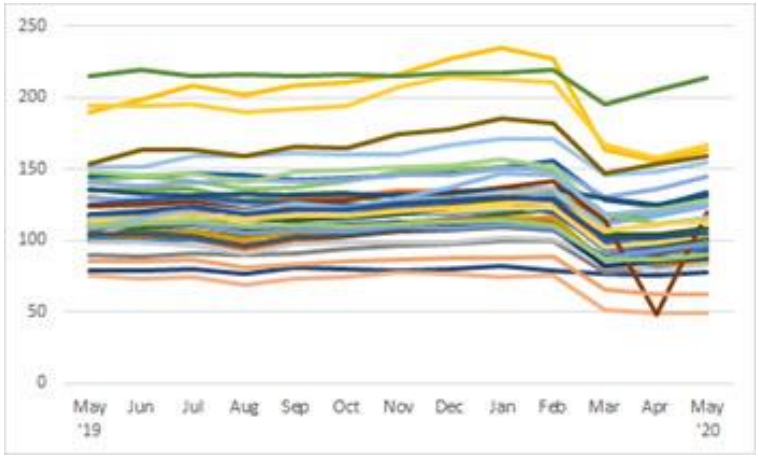
The Magic Number Seven (7) for visualization

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

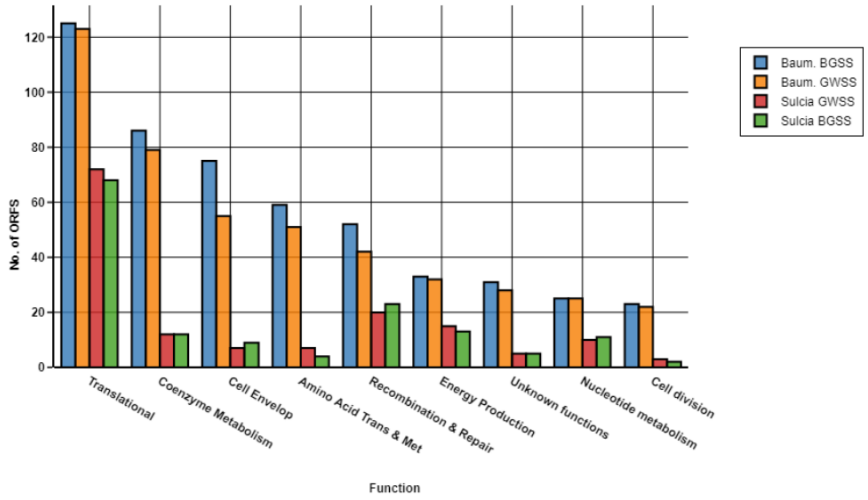
no



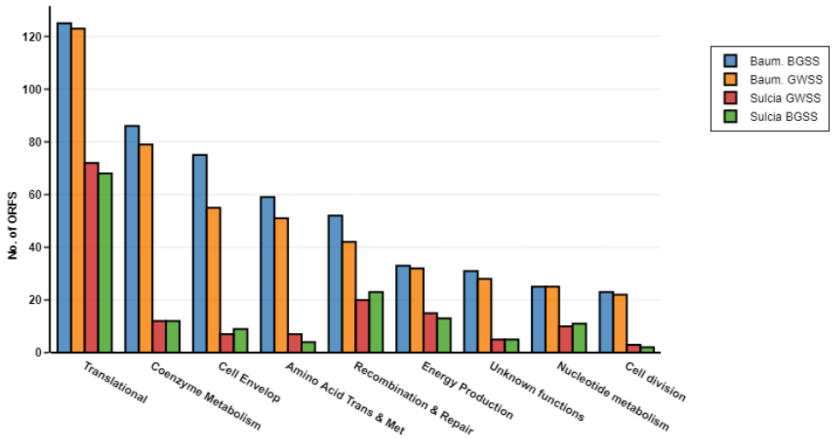
no



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

#1 List the important attributes you want to maximize

Price

Maintenance Cost

Styling

Gas Mileage

Comfort

Handling

#2 Quantify the Relative Importance of Each attribute

Price	30%
Operating Cost	10%
Styling	20%
Comfort	20%
Handling	15%
Safety	5%
Total	100%

#3 Judge Each Car How it Values on Each of these Attributes

	% Value	Car 1	Car 2	Car 3
Price	30%	3.5%	3.0%	3.5%
Operating Cost	10%	3.5%	2.0%	4.5%
Styling	20%	2.5%	4.5%	3.0%
Comfort	20%	4.0%	2.5%	3.5%
Handling	15%	3.0%	4.0%	3.0%
Safety	5%	3.5%	2.5%	4%

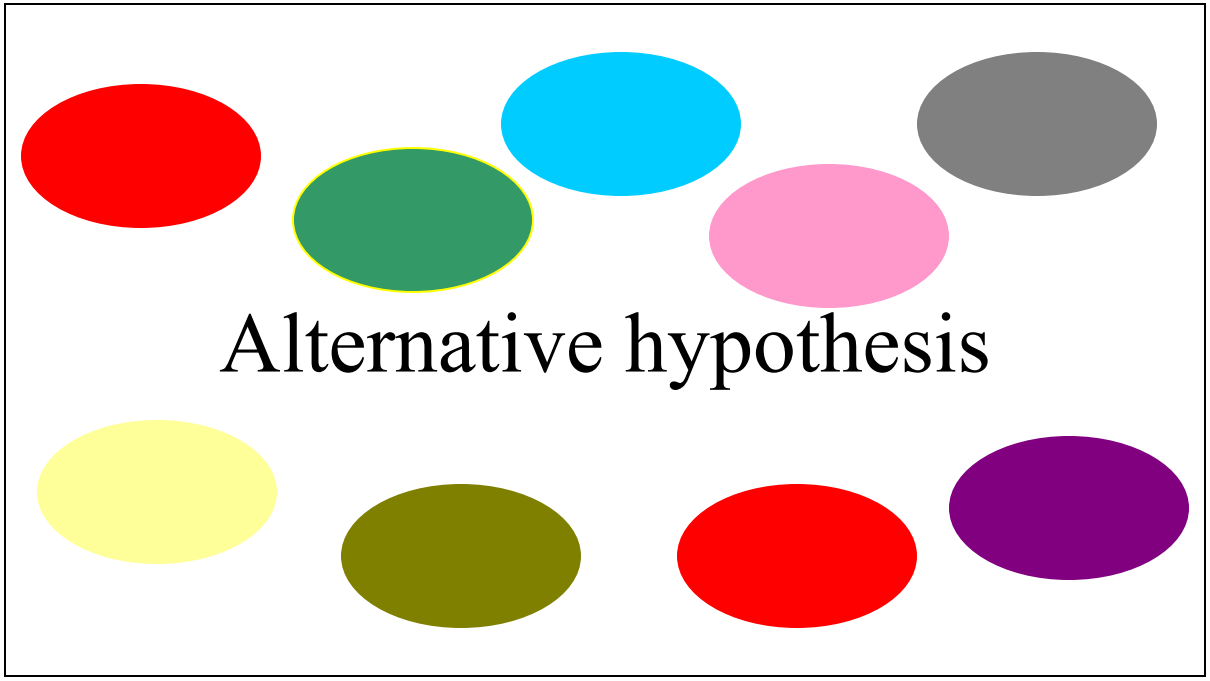
#4 Multiply the Overall Attribute Value by the Car's Attribute Value

	% Value	Car 1	Car 2	Car 3
Price	30%	105	90	105
Operating Cost	10%	35	20	45
Styling	20%	50	90	60
Comfort	20%	80	50	70
Handling	15%	45	60	45
Safety	5%	17.5	12.5	20
Totals		332.5	322.5	345

More Formally: Problem Description

When working on difficult intelligence issues

- which is the correct explanation?
- which is the most likely outcome?



Analysis of Competing Hypotheses

Analysis of Competing Hypotheses = ACH

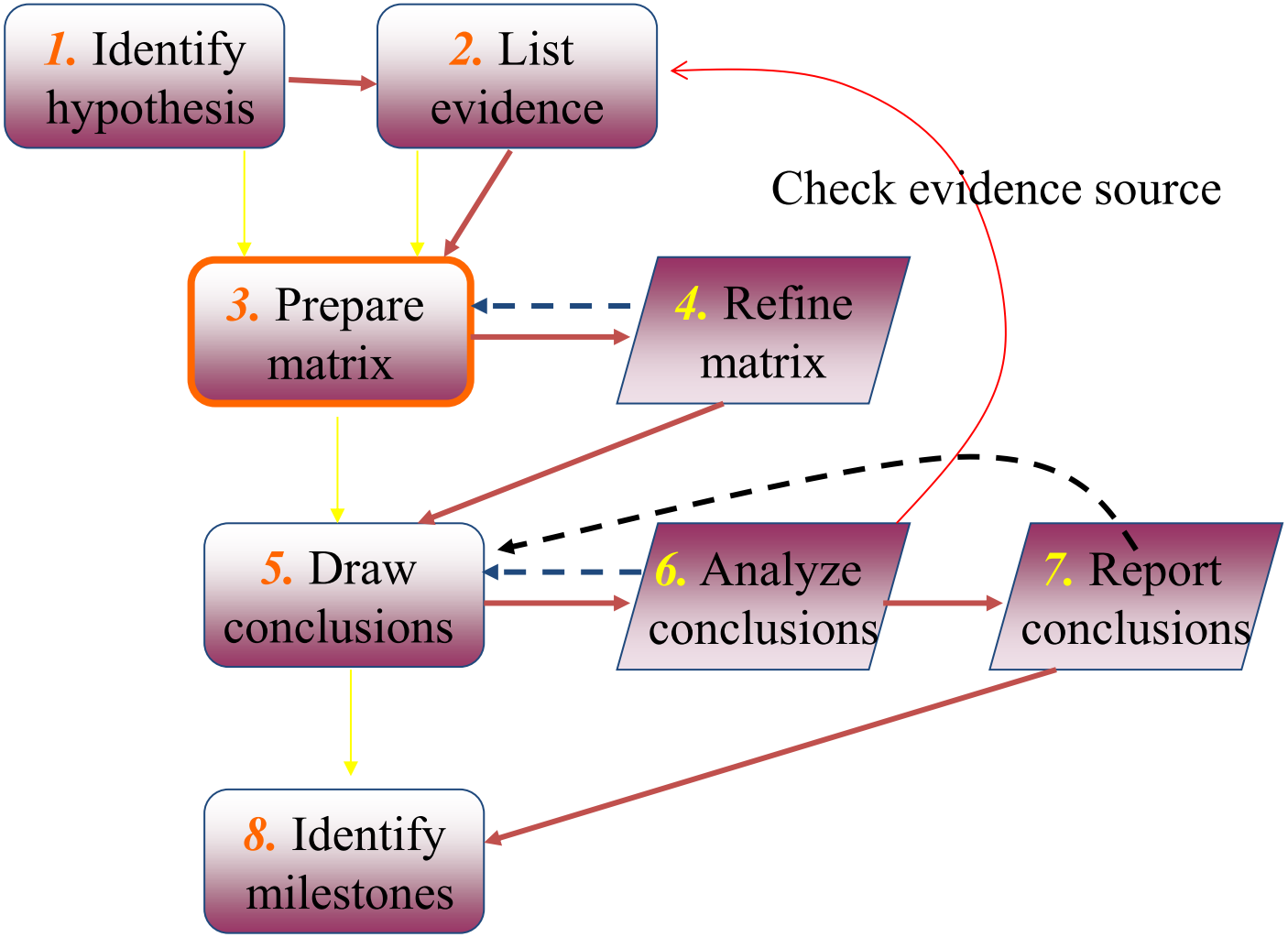
Used to

- aid judgment on important issues
- minimize cognitive limitations

Basic insights from

- cognitive psychology
- decision analysis
- scientific method

Eight-Steps of ACH



Step 1: Identify Hypotheses

Hypothesis generation *vs.* hypothesis evaluation

- generation: bring together all possibilities
- evaluation: focus on each of them and rule out from weak to strong

Disproved *vs.* unproven

- for a disproved hypothesis there is positive evidence that it is wrong
- for an unproven hypothesis, there is no evidence that it is correct

Step 2: List Evidence

Don't limit to the evidences current available

For each hypothesis, list supporting and contradicting factors

Absence and presence of evidence

- for example: Did the dog bark in the night?

  no. nobody heard it barked (absence)

Question:
Will Iraq Retaliate for a US Bombing?

H1: Iraq will not retaliate

H2: It will sponsor some minor terrorist actions.

H3: Iraq is planning a major terrorist attack, perhaps against one or more CIA installations

	H1	H2	H3
E1. Saddam public statement of intent not to retaliate.	+	+	+
E2. Absence of terrorist offensive during the 1991 Gulf War.	+	+	−
E3. Assumption that Iraq would not want to provoke another US attack.	+	+	−
E4. Increase in frequency/length of monitored Iraqi agent radio broadcasts.	−	+	+
E5. Iraqi embassies instructed to take increased security precautions.	−	+	+
E6. Assumption that failure to retaliate would be unacceptable loss of face for Saddam.	− −	+	+

Step 3: Prepare Matrix

	H1	H2	H3
E1. Saddam public statement of intent not to retaliate.	+	+	+
E2. Absence of terrorist offensive during the 1991 Gulf War.	+	+	−
E3. Assumption that Iraq would not want to provoke another US attack.	+	+	−
E4. Increase in frequency/length of monitored Iraqi agent radio broadcasts.	−	+	+
E5. Iraqi embassies instructed to take increased security precautions.	−	+	+
E6. Assumption that failure to retaliate would be unacceptable loss of face for Saddam.	− −	+	+

Step 4: Refine Matrix

Diagnostic value – likeliness of hypothesis

- high-temp indicates sickness, but can't determine which illness

Reconsider the hypotheses

- add, or need finer distinction
- combine

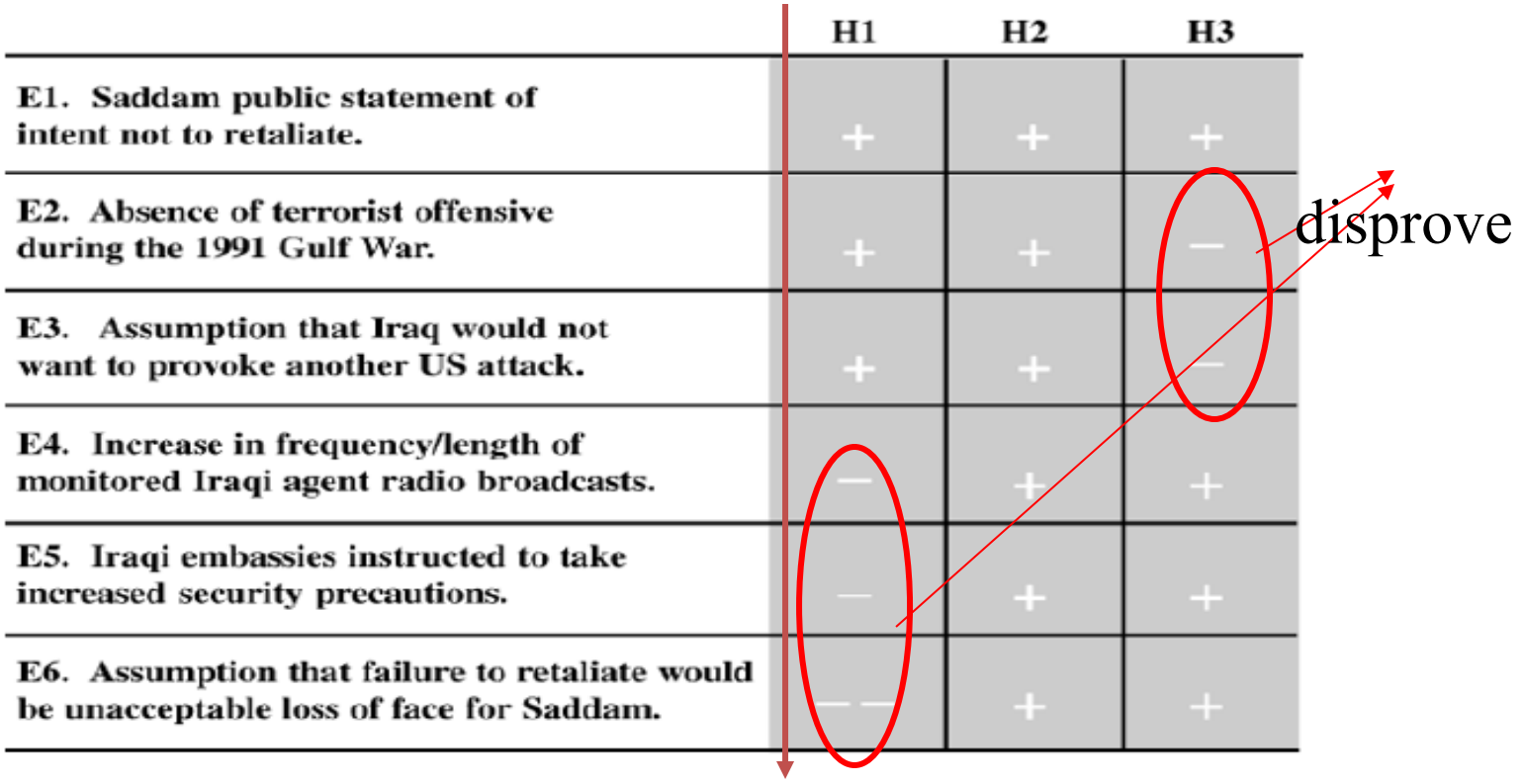
Reconsider the evidences

- put in missing factors
- delete evidence that have no diagnostic value

Step 5: Draw Conclusions

Work down the matrix, looking at each hypothesis

Proceed by trying to disprove the hypotheses rather than prove them



Step 6: Analyze Conclusions

Analyze how sensitive your conclusion is to a few critical items of evidence

- the consequences if the evidence were wrong
- check the original source

Step 7: Report Conclusions

Decision-maker needs to make decisions on the basis of a full set of alternative possibilities

The importance is on eliminating not confirming!

Discuss the relative likelihood of all the hypotheses

The Sense-Making Loop

Support visualization with computations for data processing

Form a loop: visualize - refine

Gather (forage) information

Re-represent

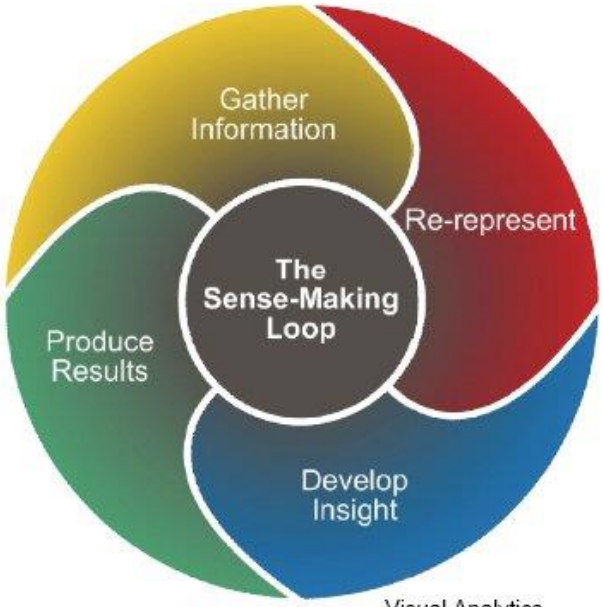
- choose a form that aids analysis

Develop insight

- through manipulation of representations

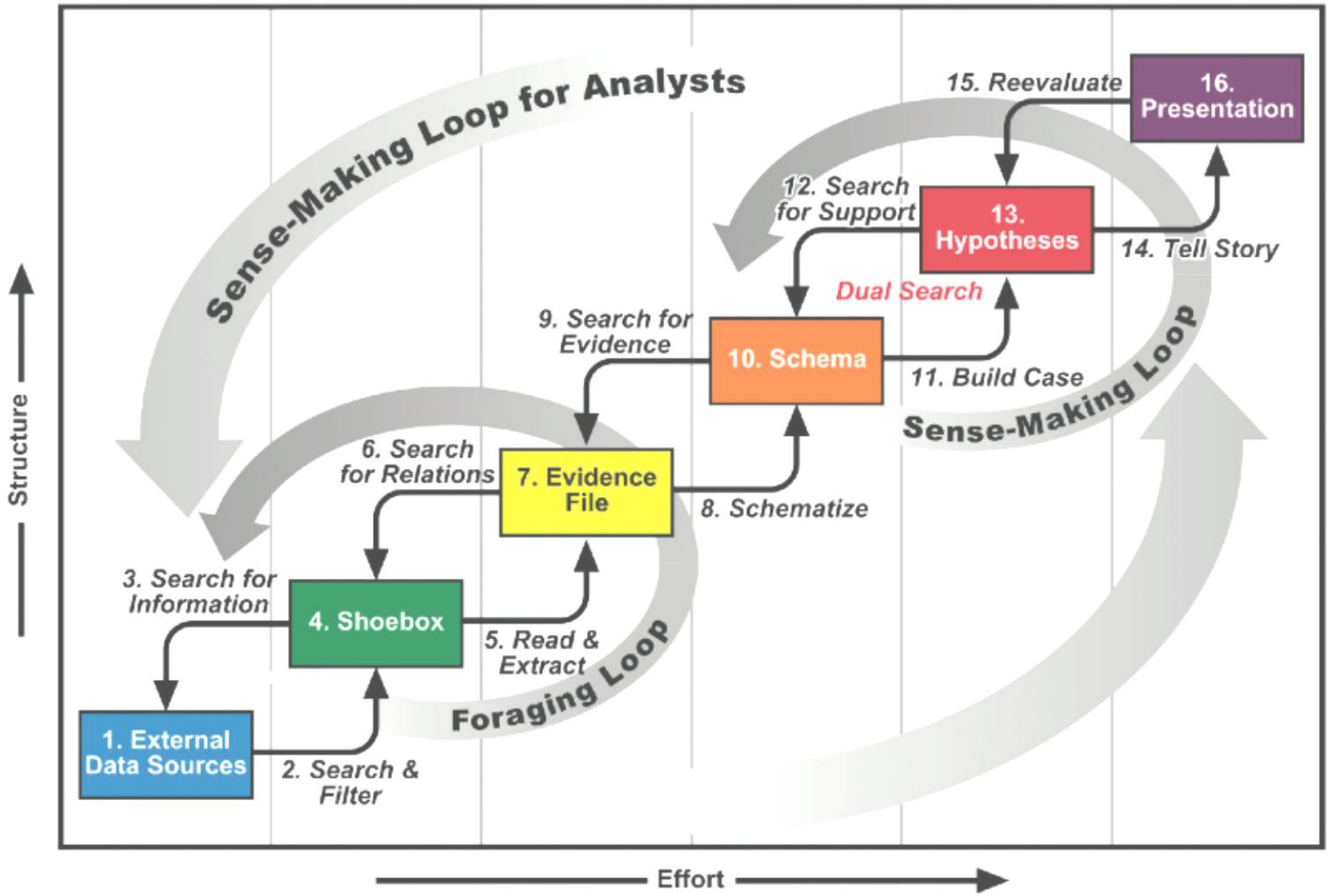
Produce results

- "product"



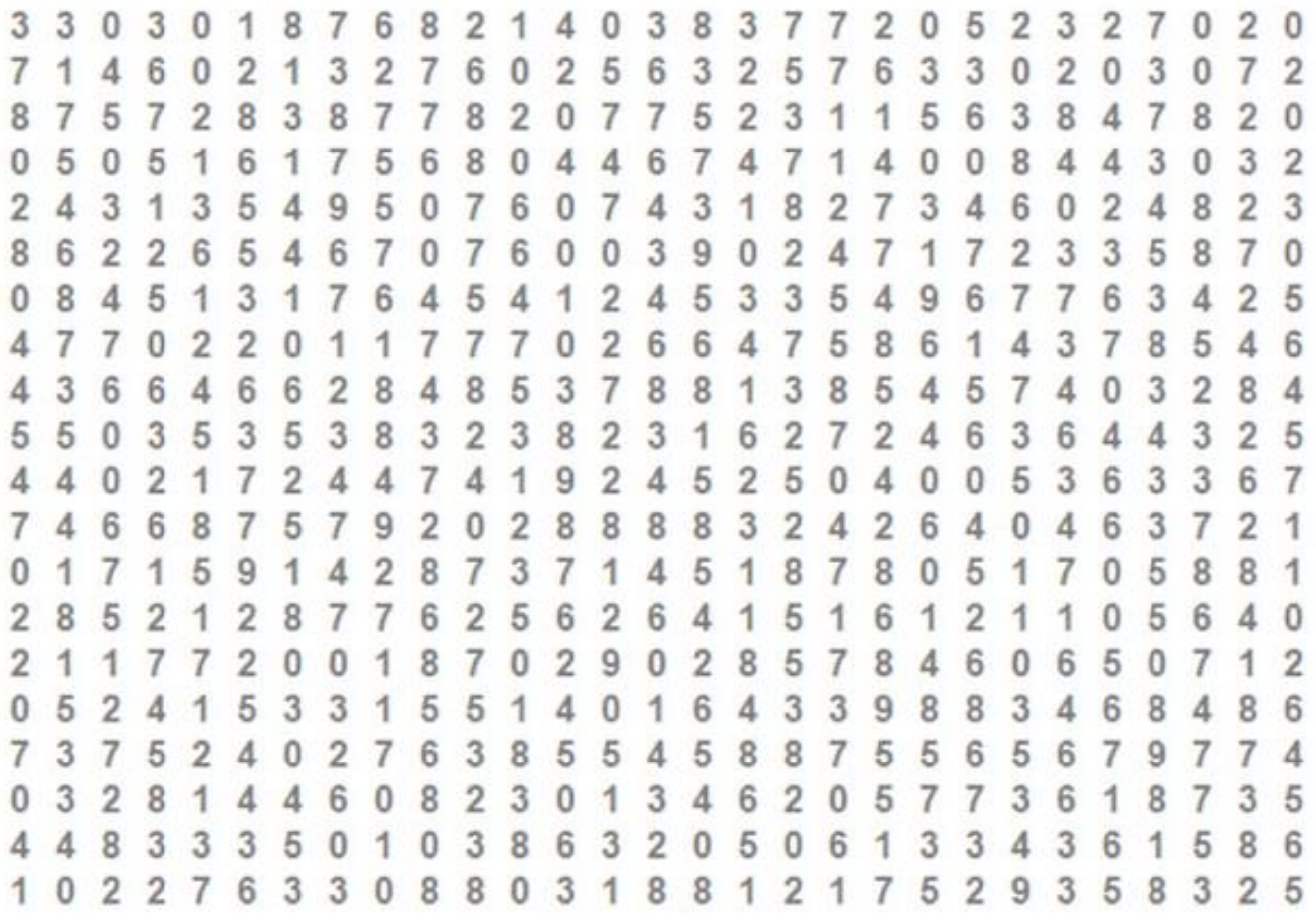
Visual Analytics
http://www.viscenter.uncc.edu/courses/visanalytics.html

Nominal Sense-Making Process

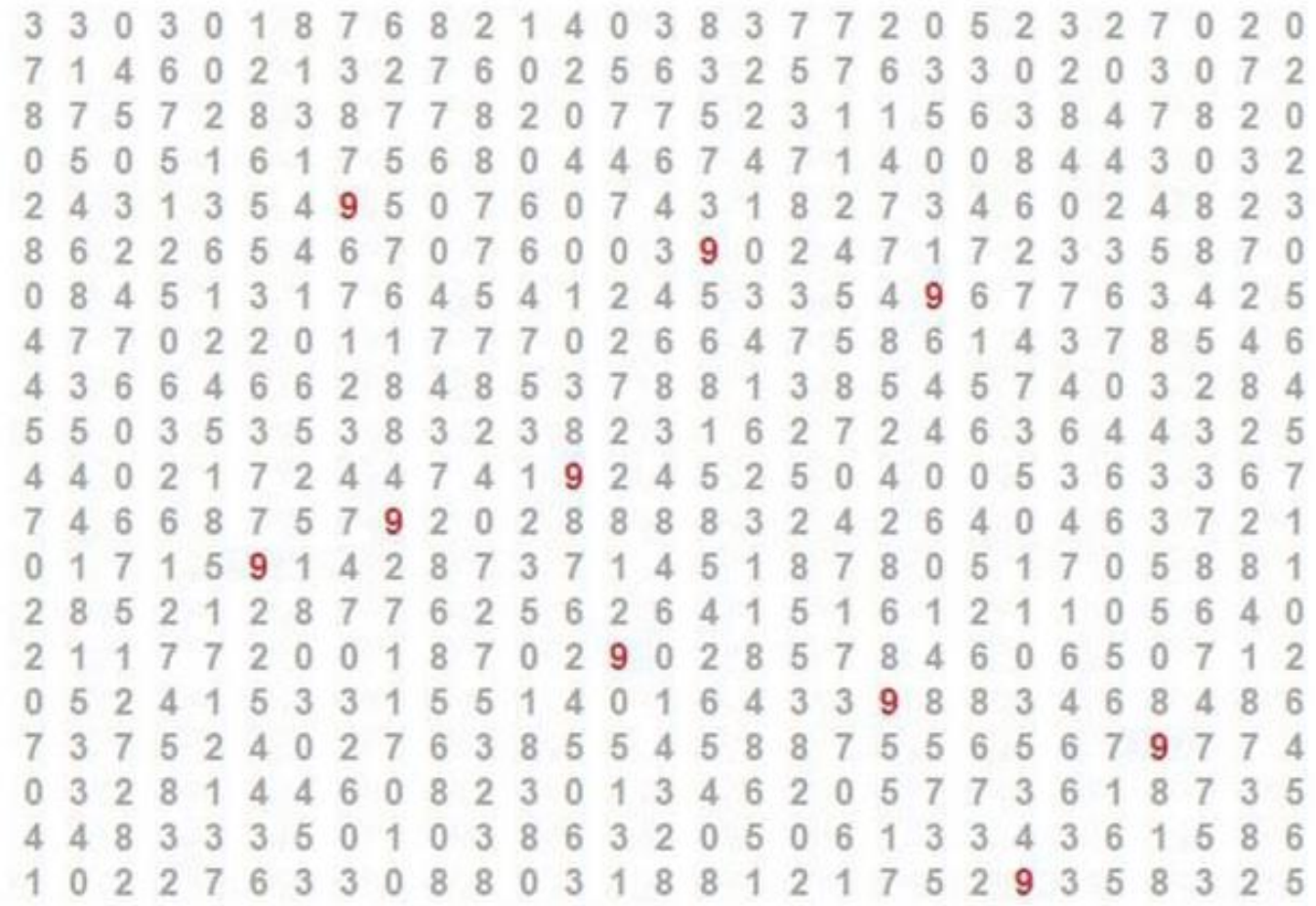


Use Visualizations to Evoke The Right Thoughts

How Many 9s Do You See?



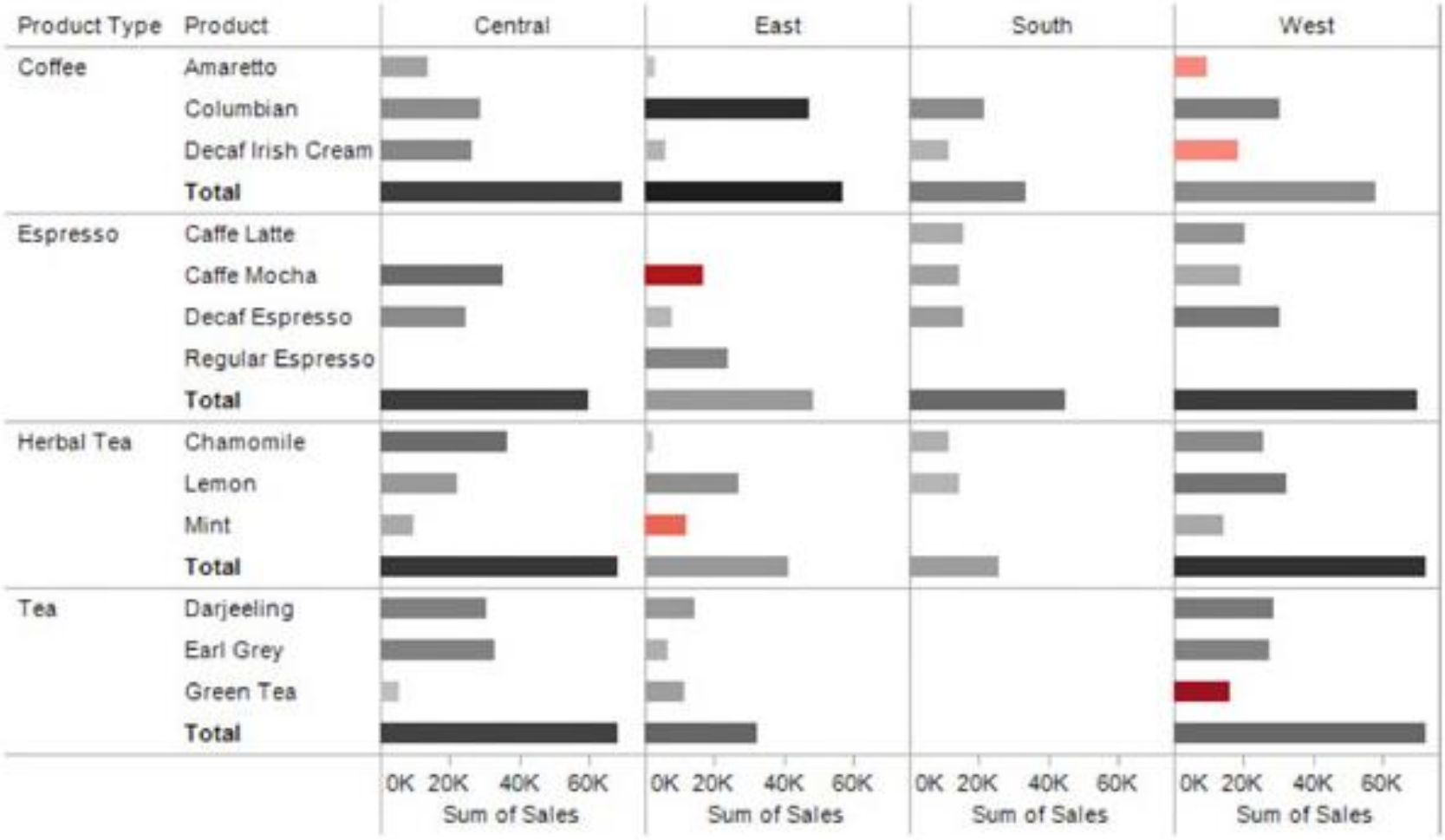
How Many 9s Do You See?



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

		Central		East		South		West	
Product Type	Product	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

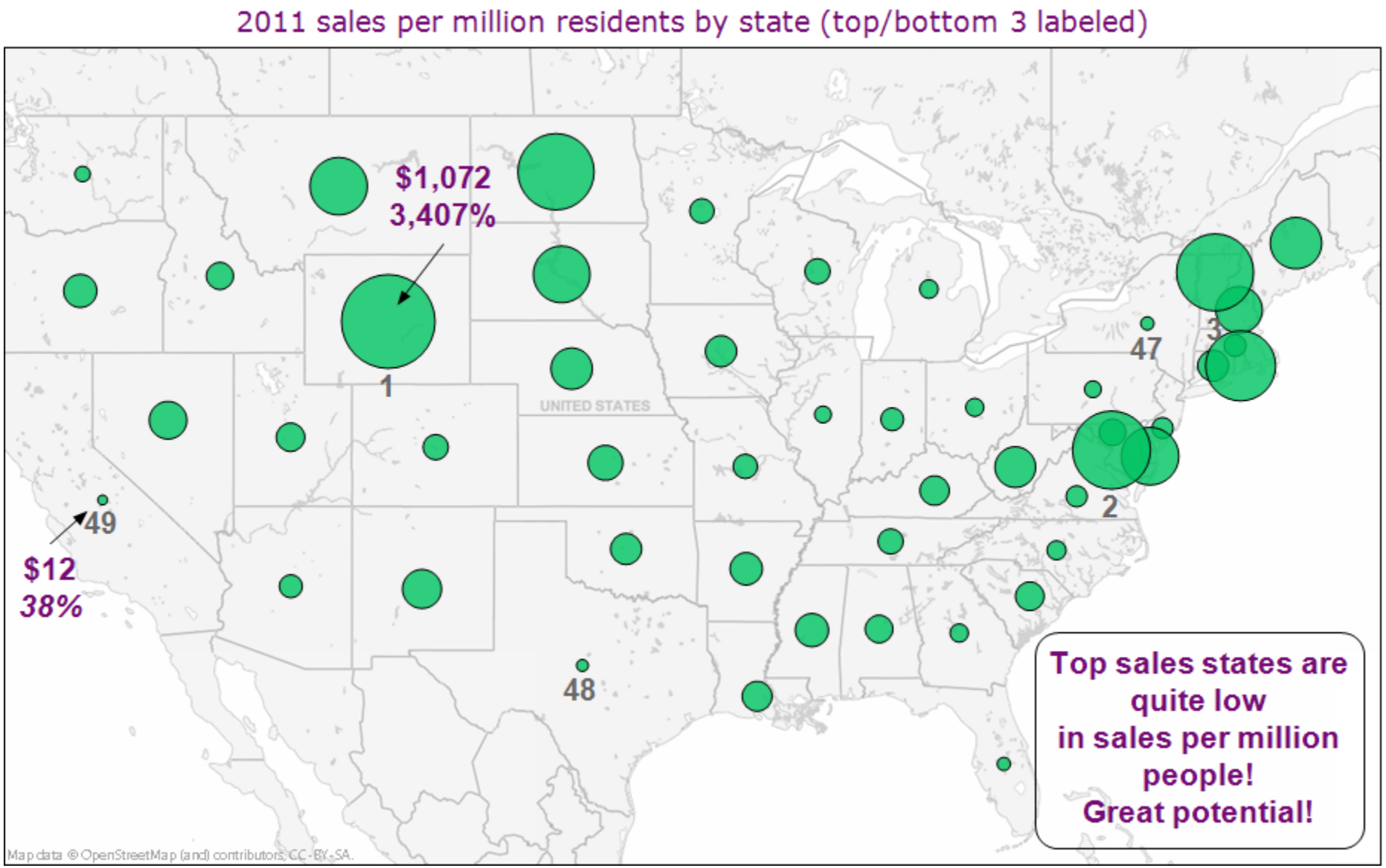
Maria
Senior Sales Analyst
March 15th, 2012

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



Potential sales by state???

+Is there a better metric?

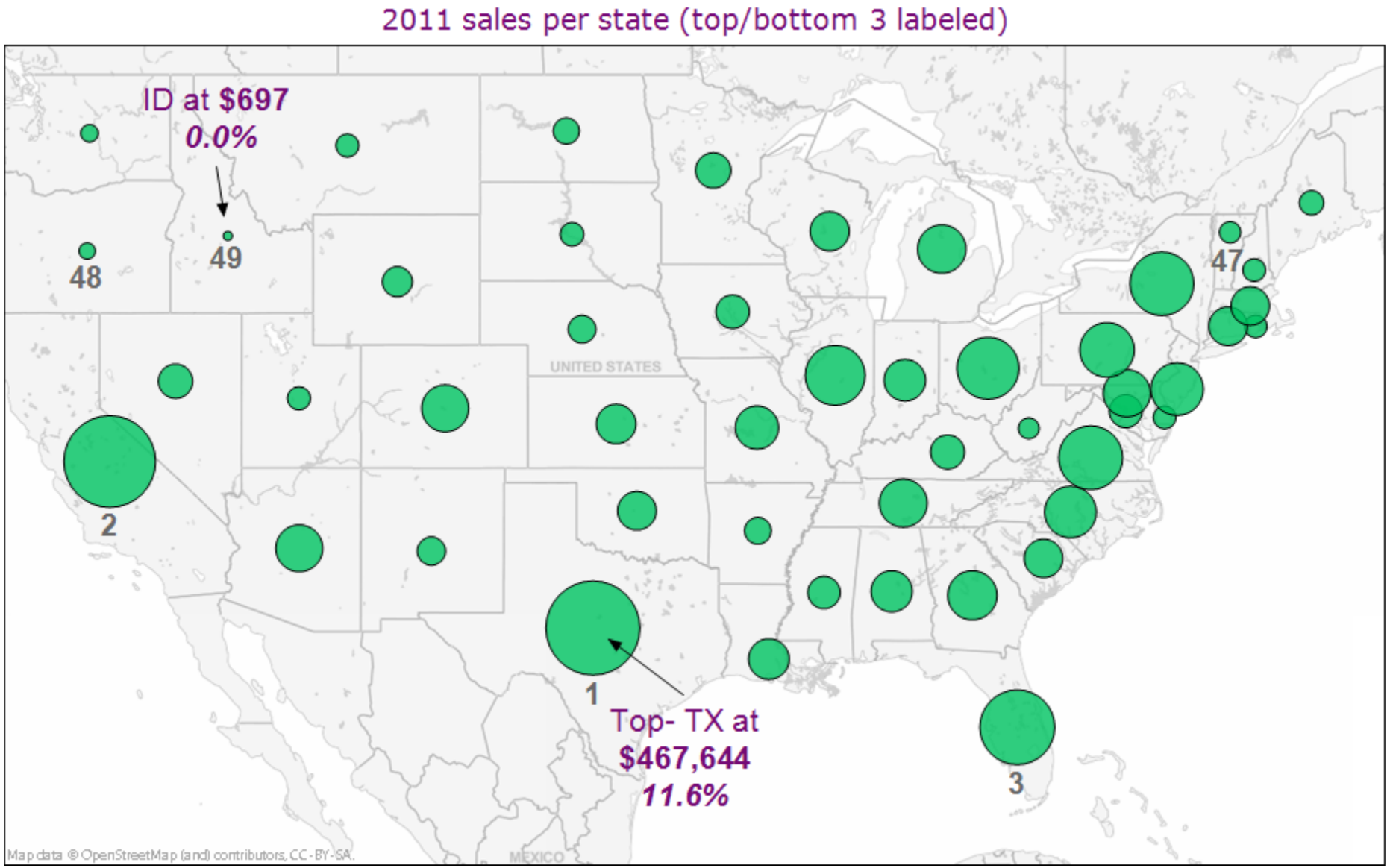
+The emphasis is on **potential**

Average sale per capita for top states

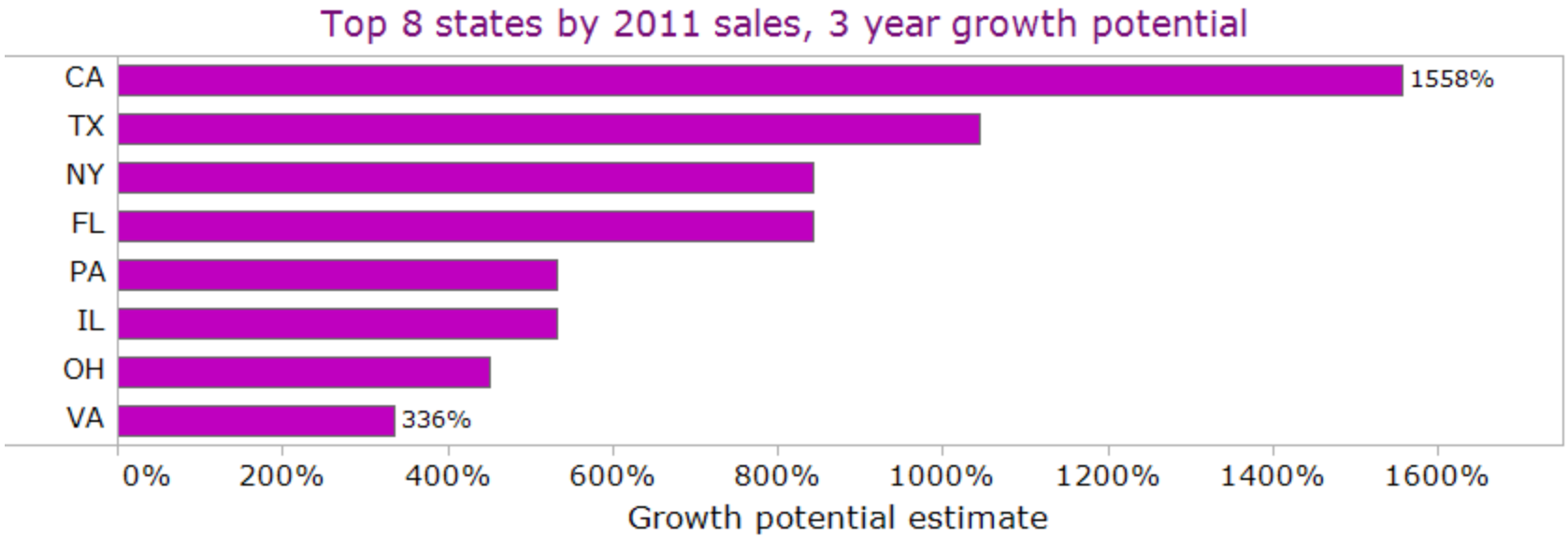
multiplied by

Current population of top sales states

Sales per State/Capita x Capita



Highest growth potential in top 8

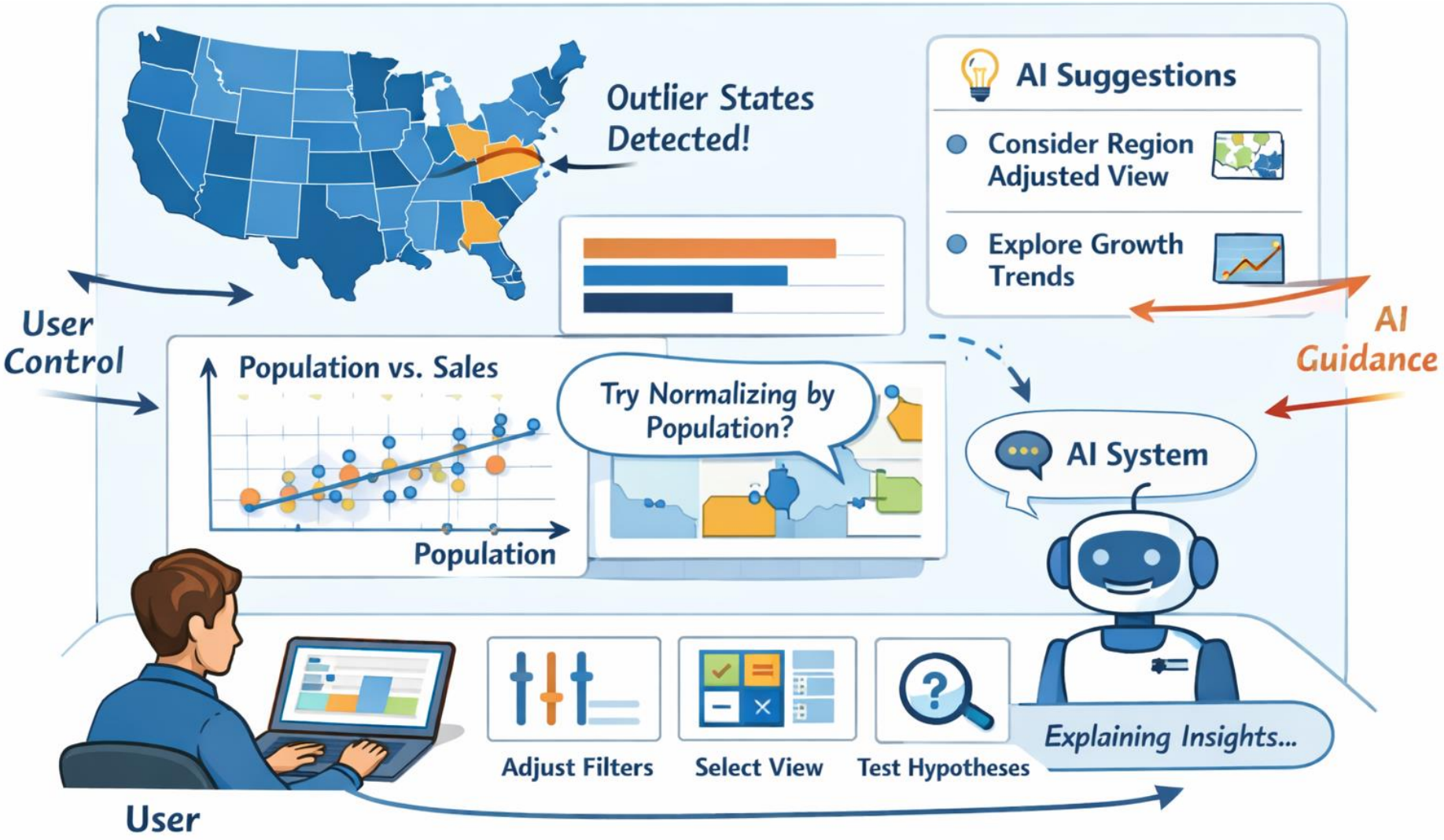


+ If we were to pick just one state, California has the greatest potential

+ 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

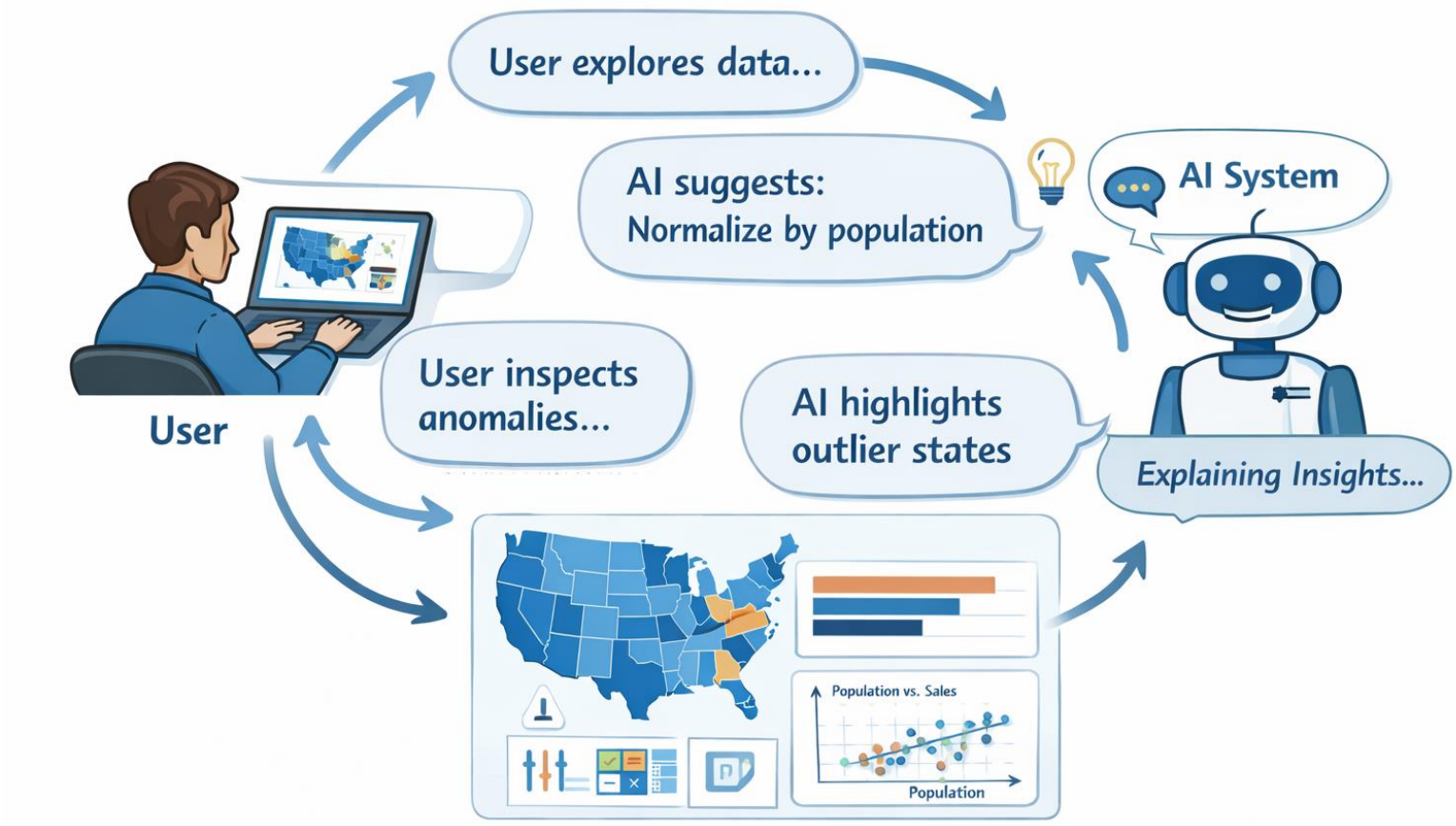
Encourages multiple hypotheses (ACH!)

- system suggests different perspectives
- user evaluates them

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:

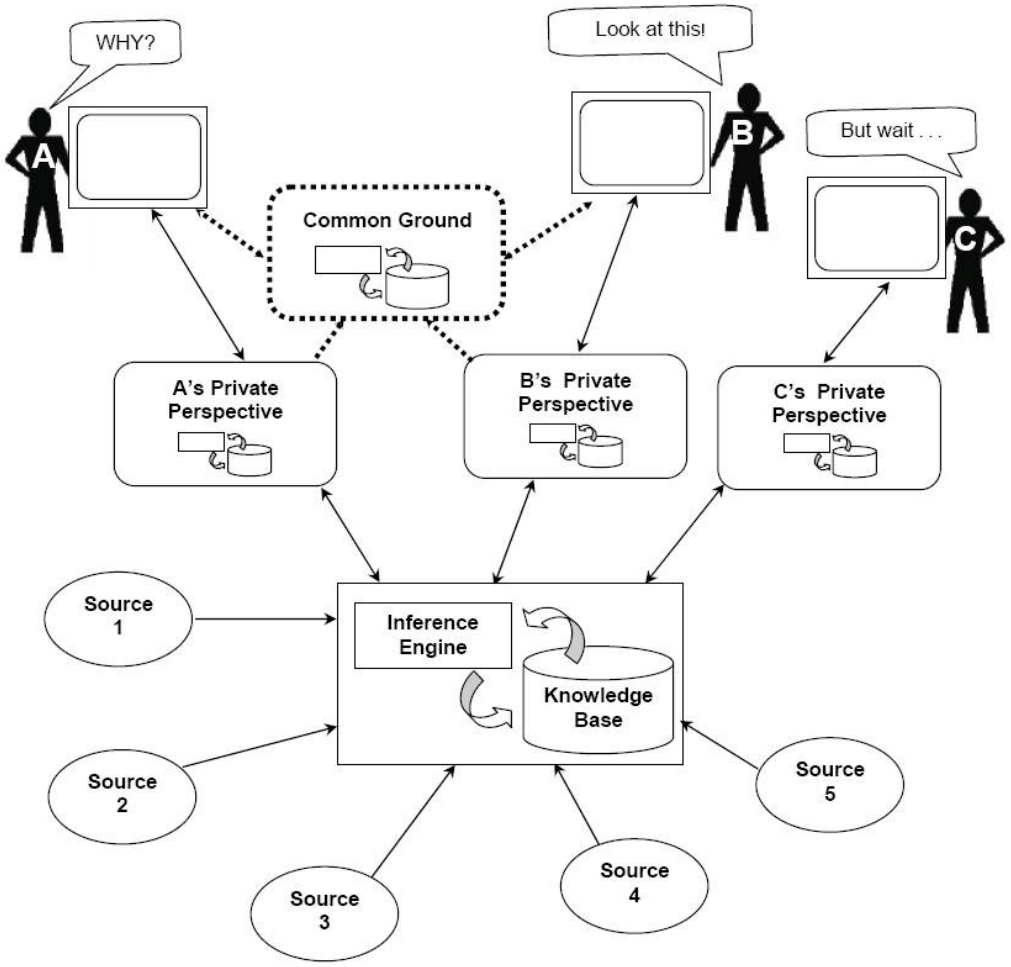
Individual sensemaking and hypothesis formation

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

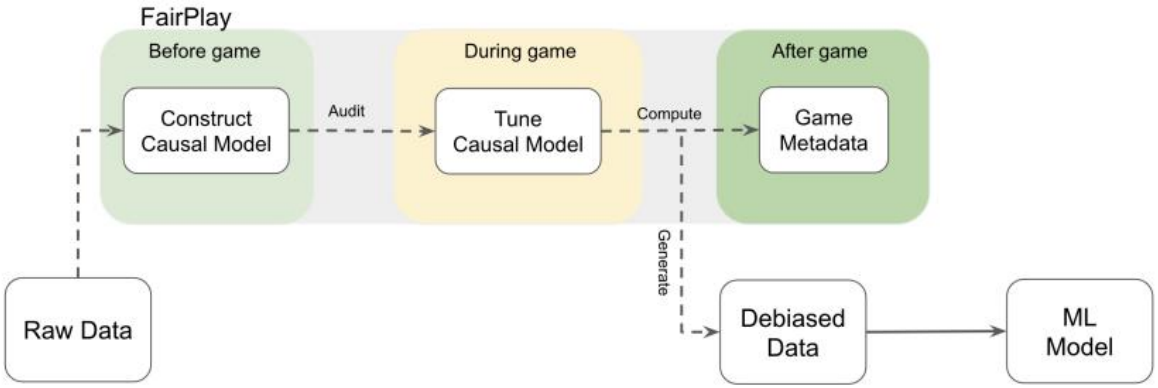
- Web-based system for collaborative dataset debiasing

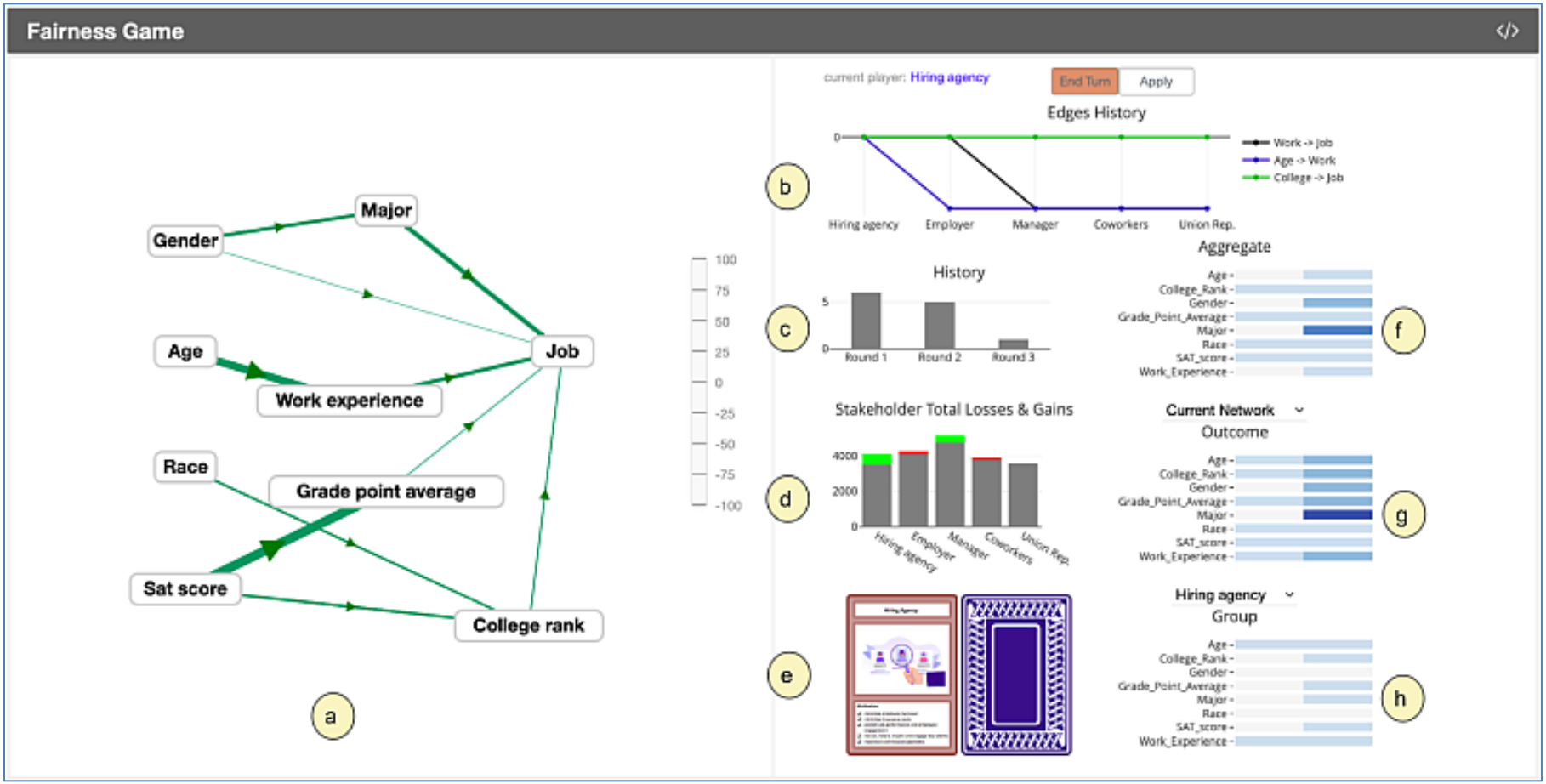
Stakeholders interact to

- 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.

T. Behzad, M. Singh, A. Ripa, K. Mueller, FairPlay: A Collaborative Approach to Mitigate Bias in Datasets for Improved AI Fairness, *Proc. of the ACM on HCI (CSCW)*, 9(2):1-30, 2025 (link)

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