

CSE 564

VISUALIZATION & VISUAL ANALYTICS

VISUAL ANALYTICS & THE VISUAL
SENSE MAKING PROCESS

KLAUS MUELLER

COMPUTER SCIENCE DEPARTMENT
STONY BROOK UNIVERSITY

WHY VISUAL ANALYTICS?

Big Data

12+ TBs
of tweet data
every day



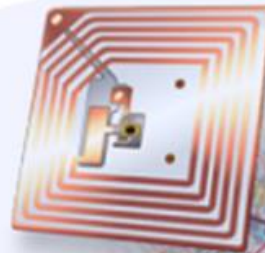
? TBs of
data every day



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
- 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™

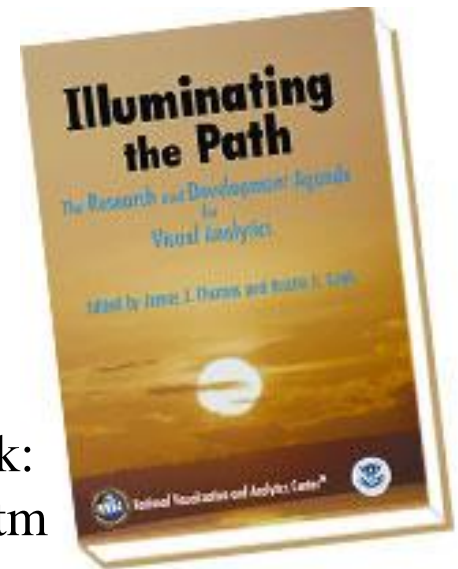
WHAT IS VISUAL ANALYTICS

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

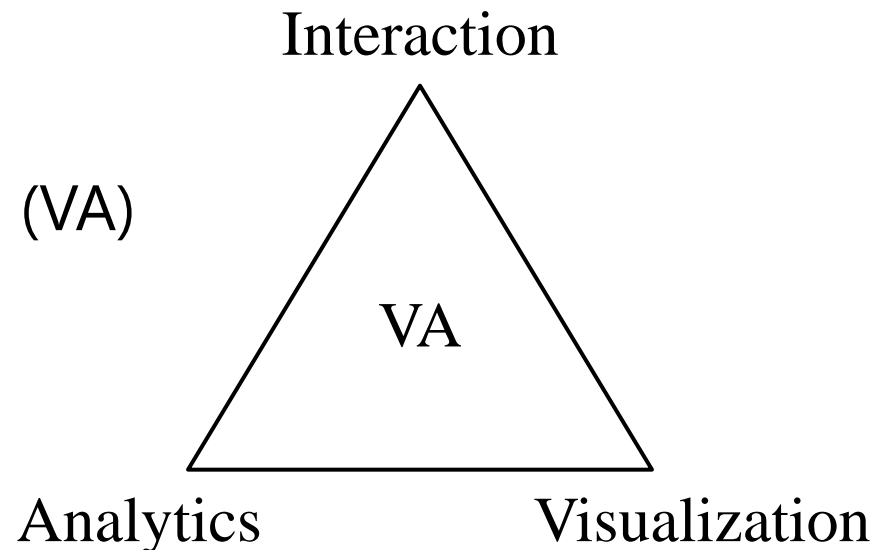
VISUAL ANALYTICS PARADIGM

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

Low signal vs. noise (SNR)

Many data types

- text, images, video, sensor data, etc.

Uncertainty

Contradictions

Omissions

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

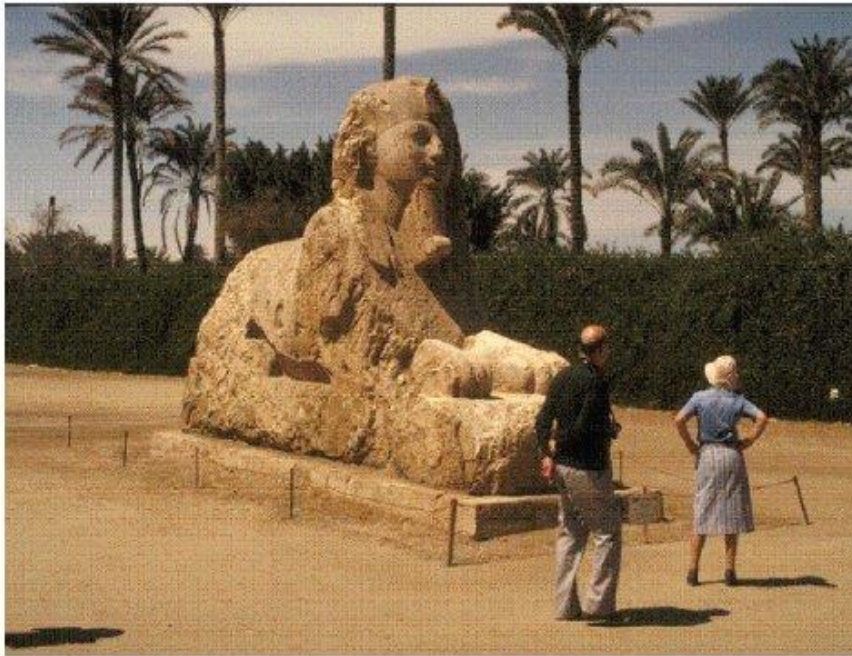
- note the Visual Analytics slogan: *Detect the Unexpected*

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?

[watch video](#) ([YouTube](#))

Video by Dan Simons (U Illinois)

EXAMPLE #3

Another distraction experiment

[watch video](#) ([YouTube](#))

Video by Dan Simons (U Illinois)

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

PERSISTENCE OF MINDSET

Another deficiency of humans

- humans tend to stick with an "opinion" for a long time
- how long does it take you to switch?



man/woman

Young/old woman



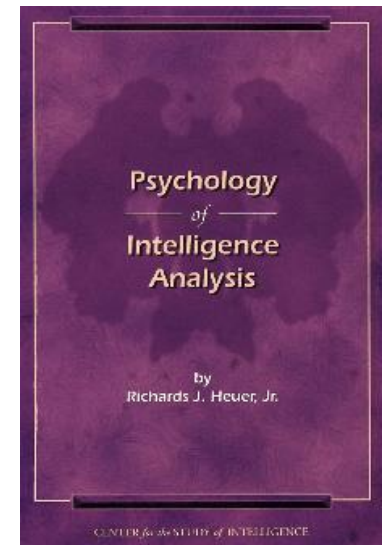
HUMAN LIMITATIONS

The Magic Number Seven

- ± 2 : the number of things most people can keep in working memory at one time
- causes problems for complicated analysis

An excellent book that has more on this topic

- discussed next



STRATEGIES FOR DEALING WITH COMPLEXITY

Decomposition

- decompose a complex problem into simpler problems
- get one's thinking straight in these simpler problems

Externalization

- get the decomposed problem out of one's head and down on paper or on a computer screen in some simplified form
- shows 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,...

THE SENSE-MAKING LOOP

Support visualization with computations for data processing

Form a loop: visualize - refine

Gather (forage) information

Re-represent

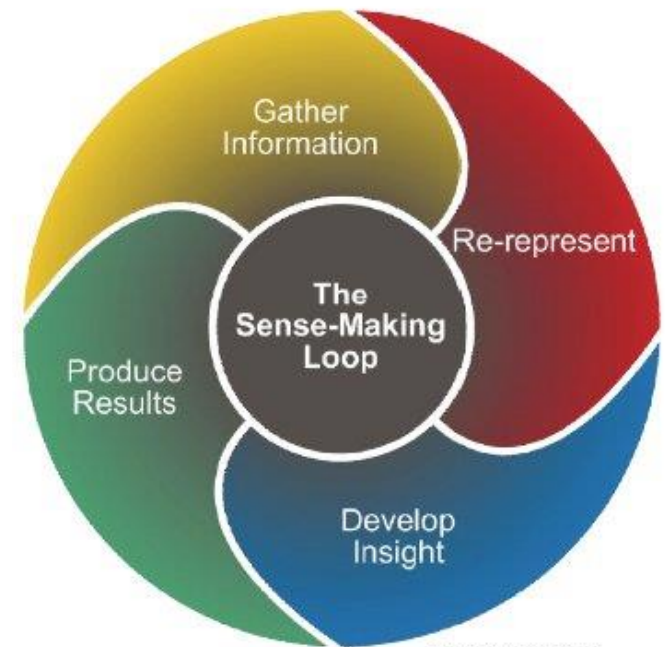
- choose form that aids analysis

Develop insight

- through manipulation of representatic

Produce results

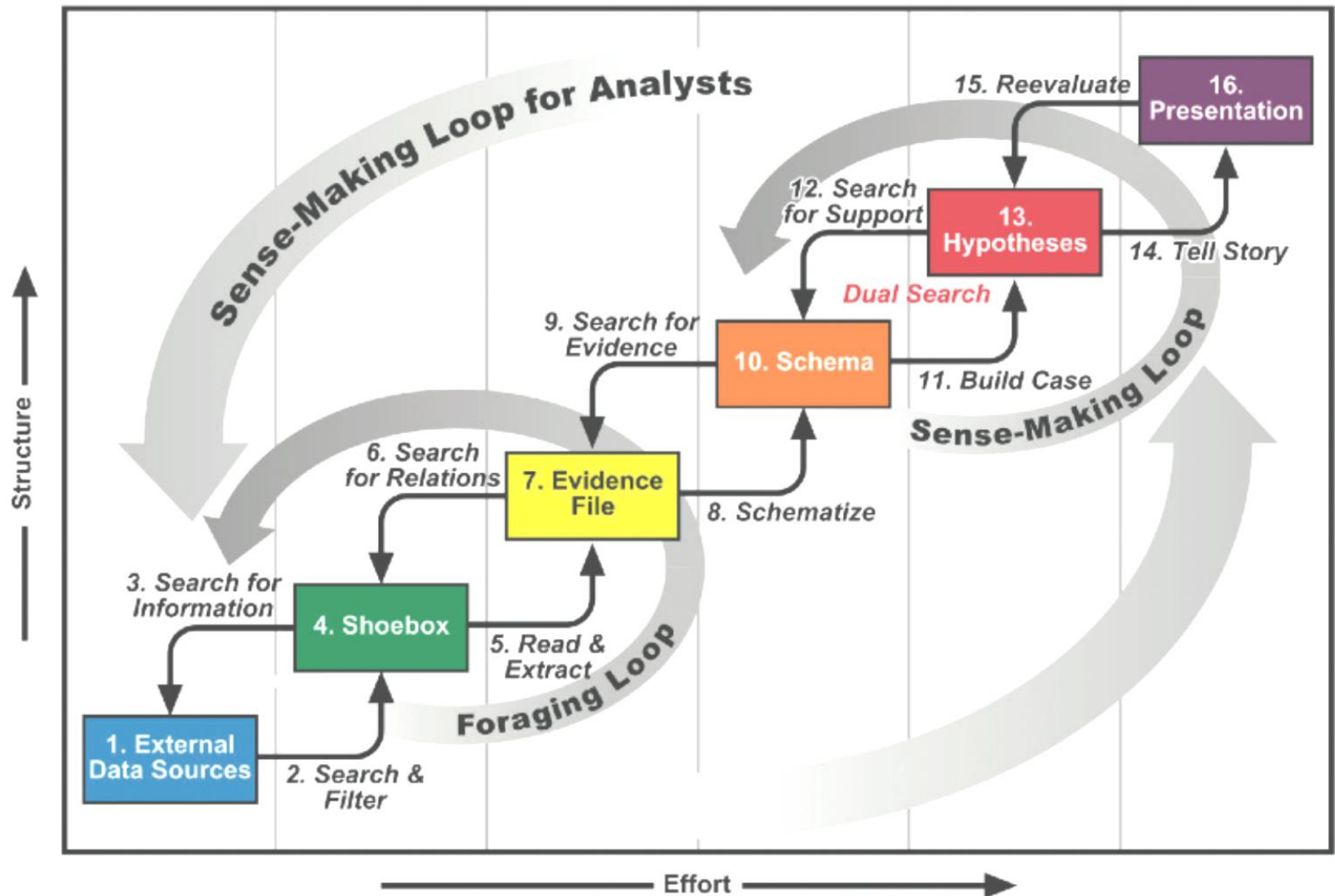
- "product"



Visual Analytics

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

NOMINAL SENSE-MAKING PROCESS



REASONING ARTIFACTS

Elemental artifacts

- source intelligence, evidence, assumptions

Pattern artifacts

- relationships, temporal and spatial structure

Higher-order knowledge constructs

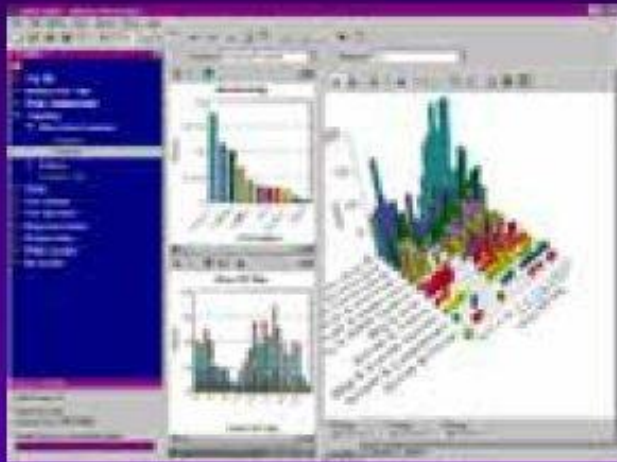
- arguments, causality, models

Complex reasoning constructs

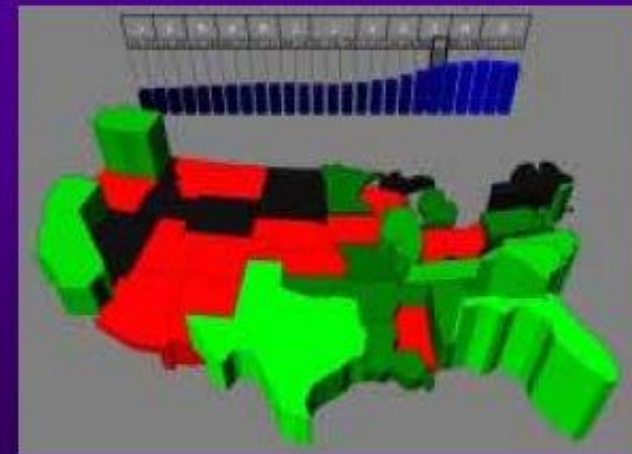
- hypotheses, scenarios

*All these become part of the Visual Analytics
sense-making (reasoning) process*

STANDARD INFORMATION DISPLAYS



Showing about 50 - 300 data values; 10-20 dimensions



Examples from the VisualInsights WebPage

CROSS-FILTERING

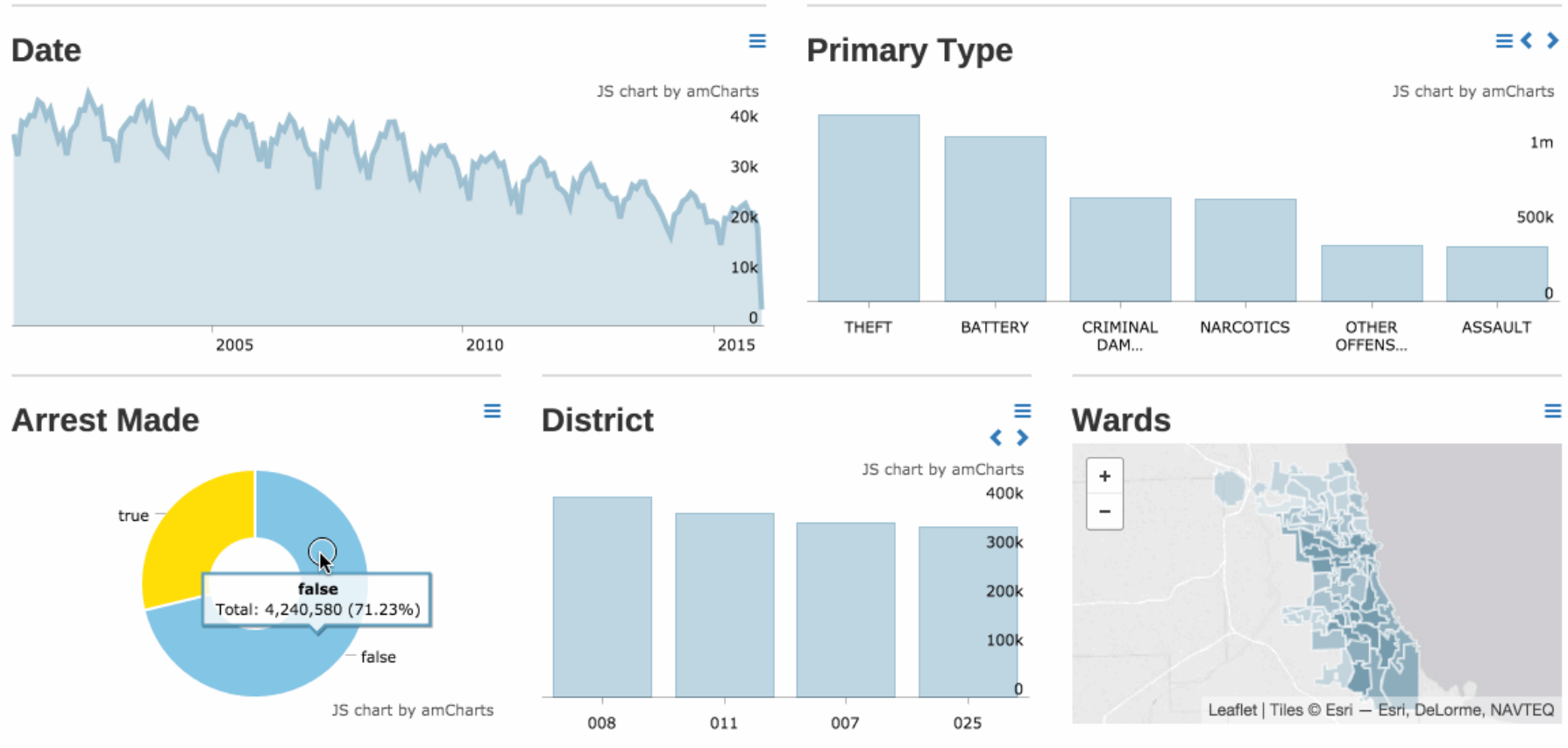
Drill down into certain aspects of the data

- essentially isolates a subset of the data
- for multivariate data with coordinated views (dashboards)

Crossfilter

- see github <https://square.github.io/crossfilter/>
- fast multidimensional filtering for coordinated views
- even with datasets containing a million or more records;

CROSS-FILTER EXAMPLE



See here for the animation: <https://discourse.metabase.com/t/cross-filtering/3466>

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

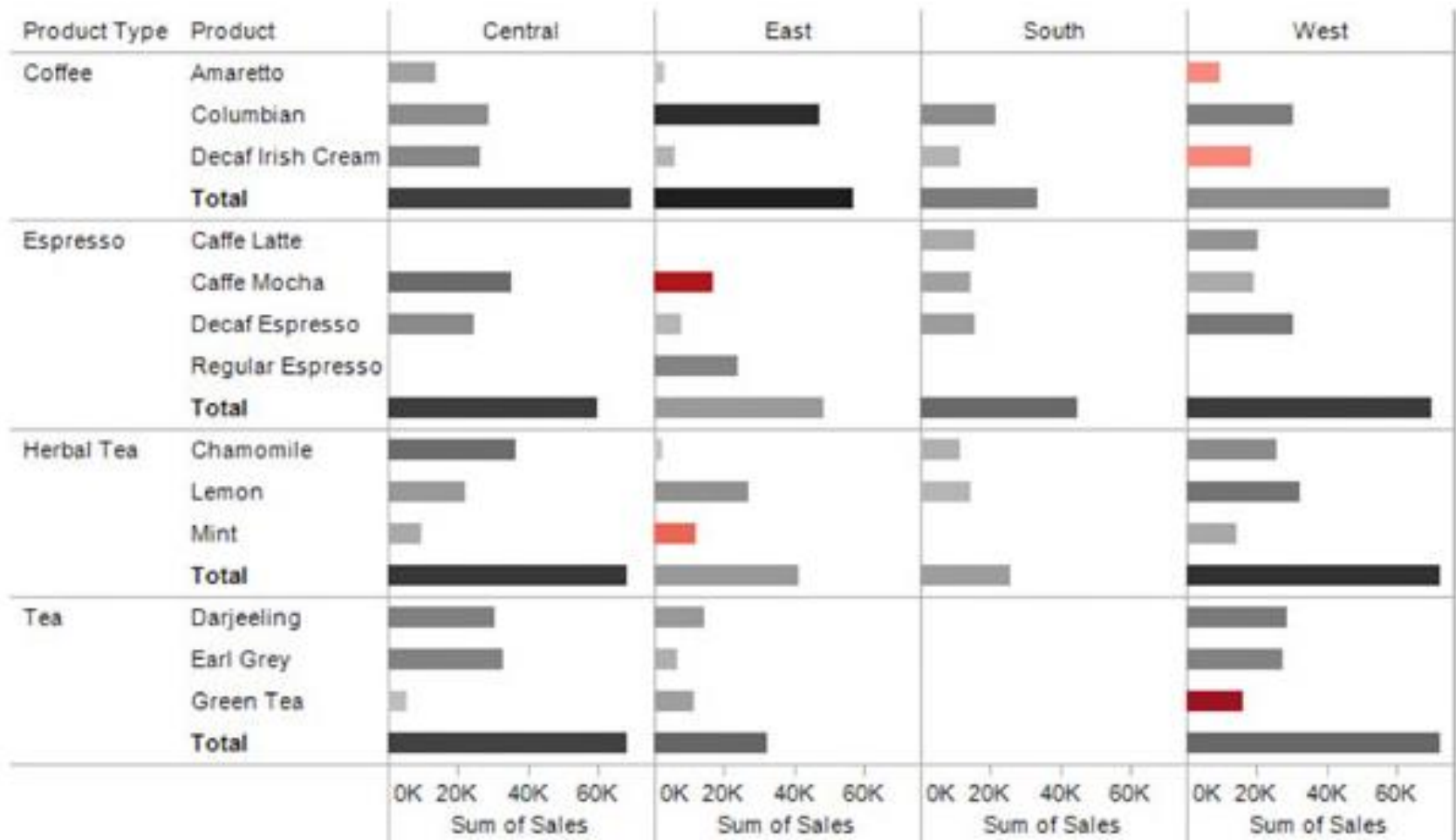
HOW MANY 9S DO YOU SEE?

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

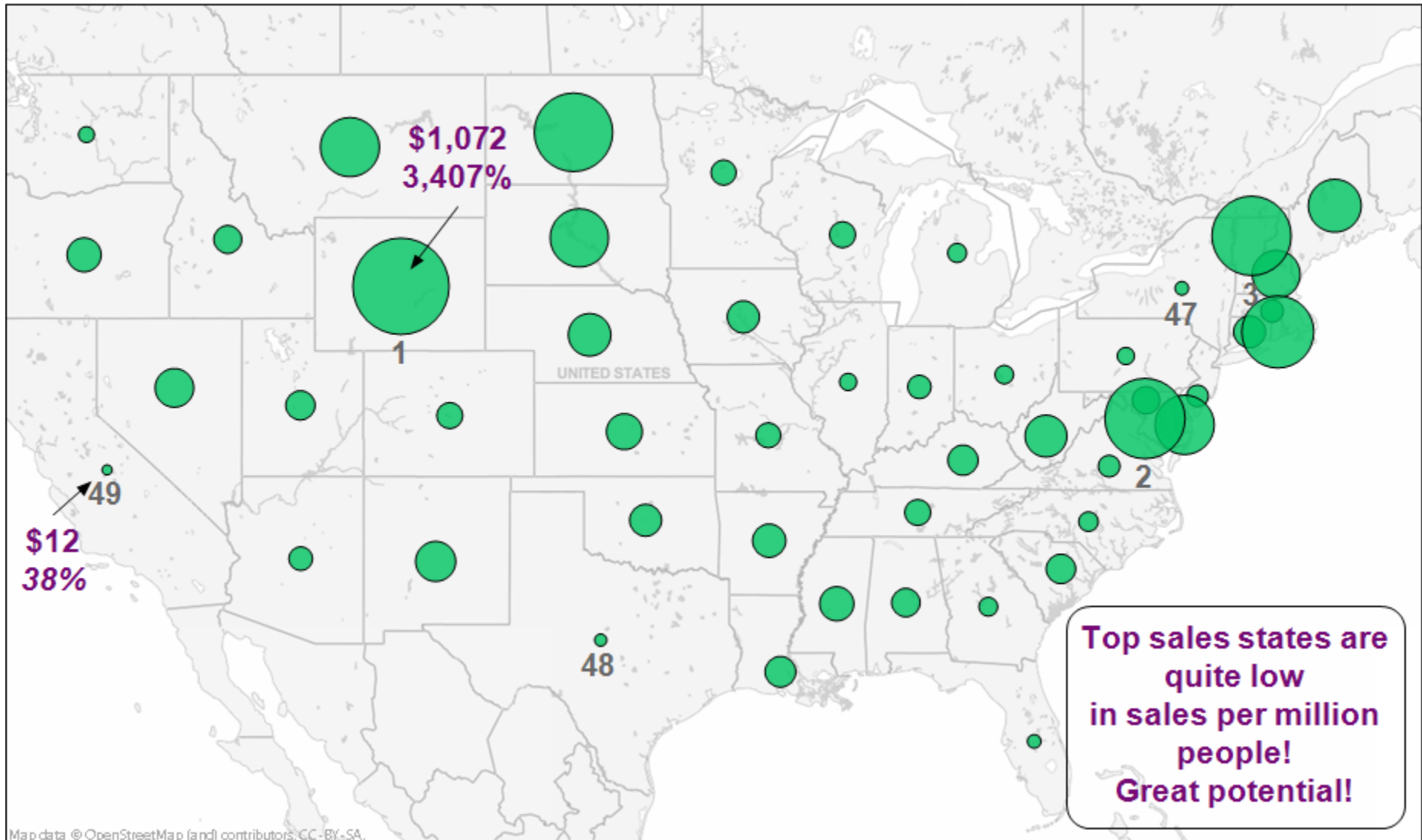
Today's question

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

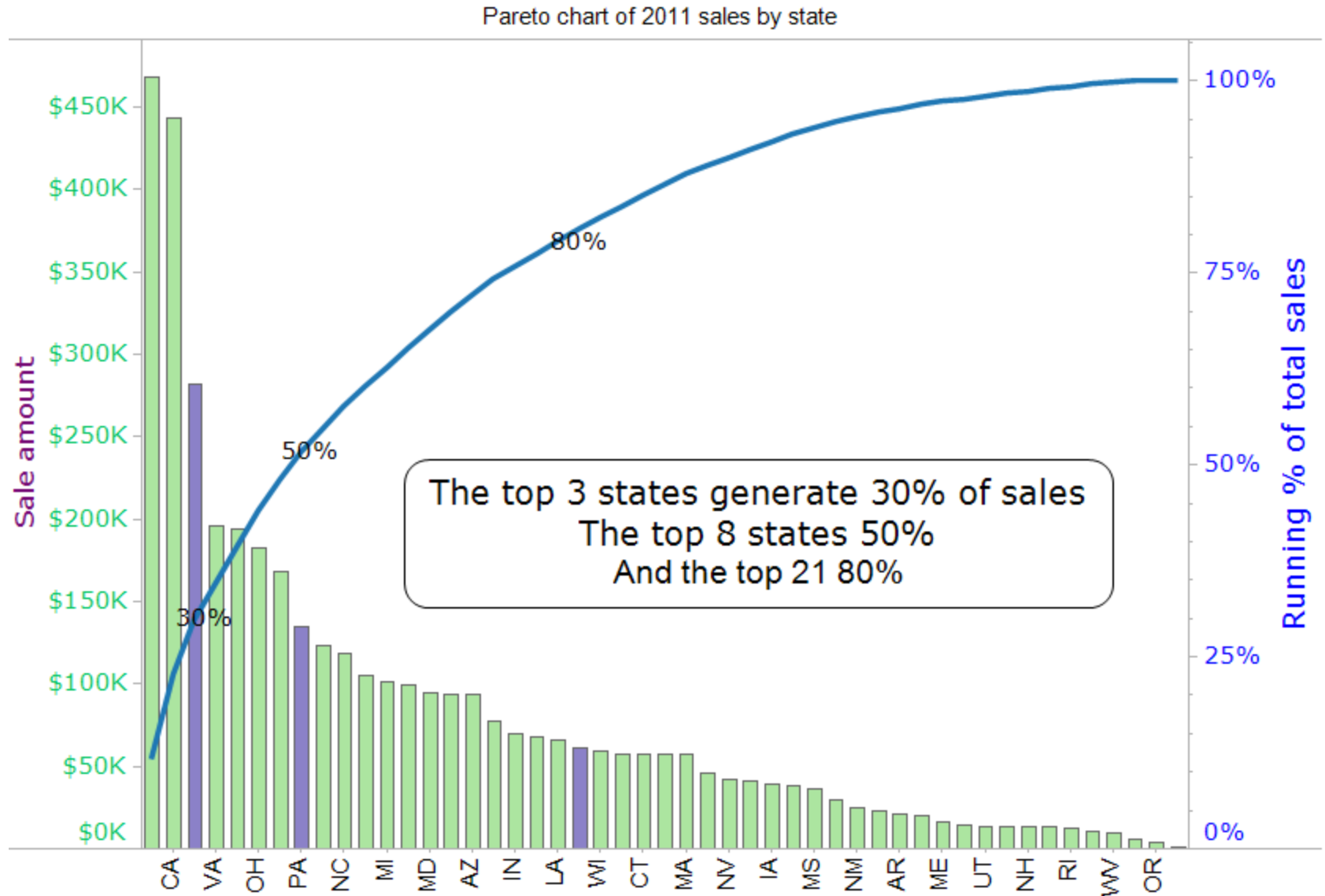
Based upon sales growth potential...

2011 sales by state

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



2011 sales by state



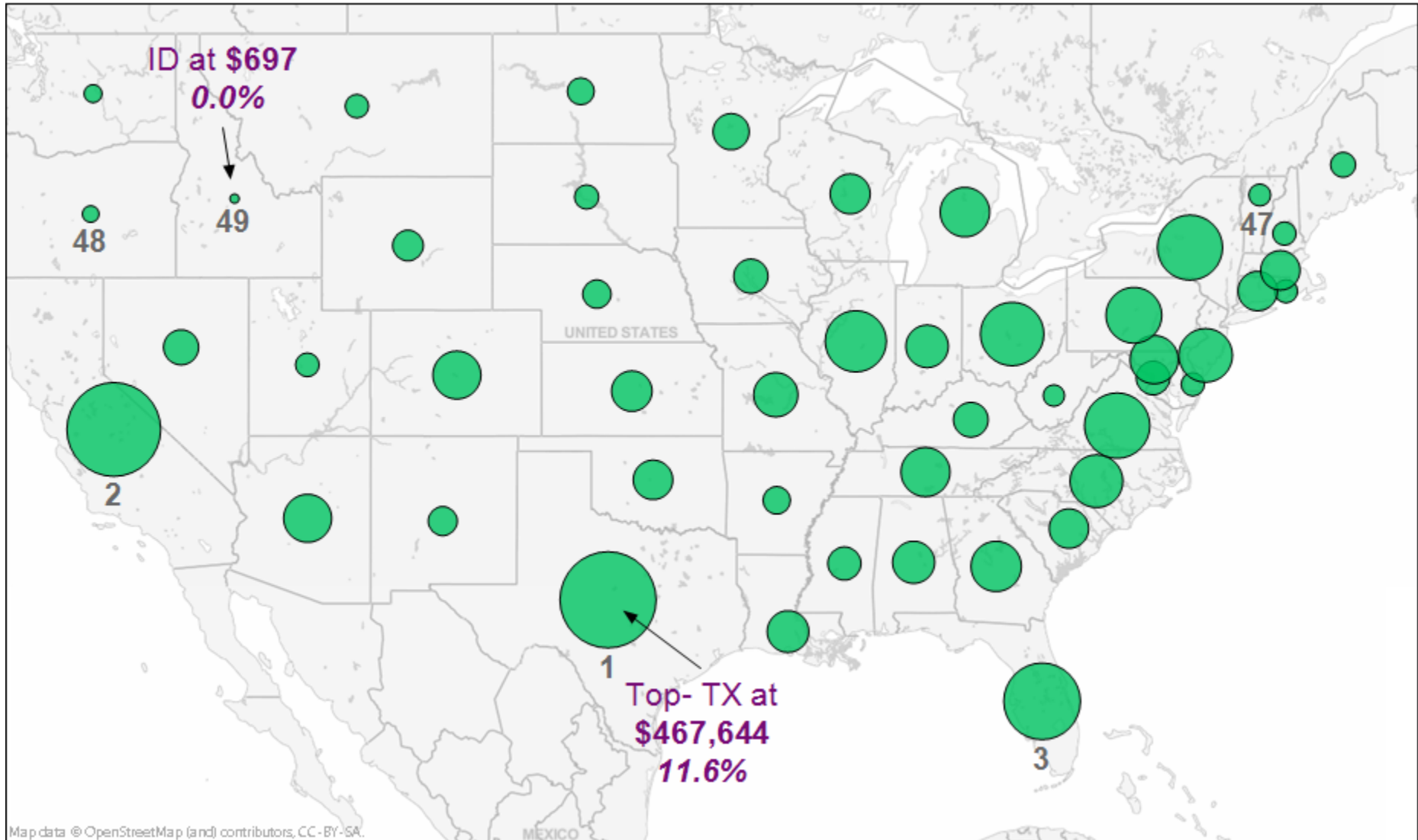
Potential sales by state???

- +Is there a better metric?
- +The emphasis is on **potential**

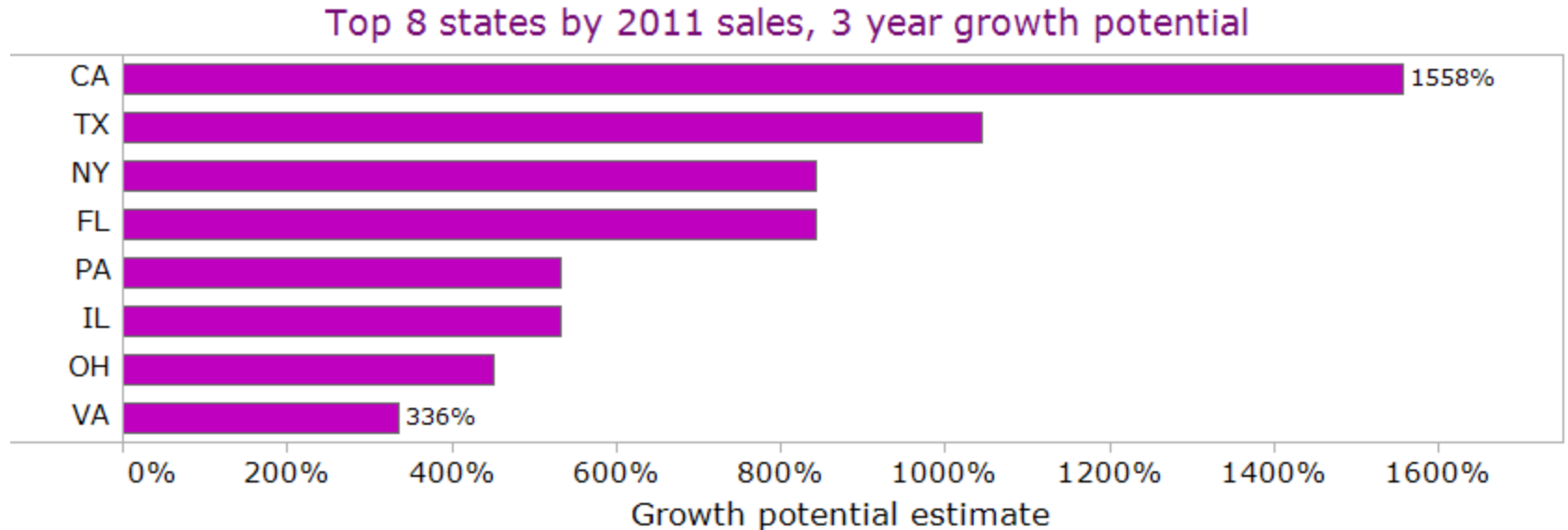
Average sale per capita for top states
multiplied by
Current population of top sales states

What are the top states based on sales per 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

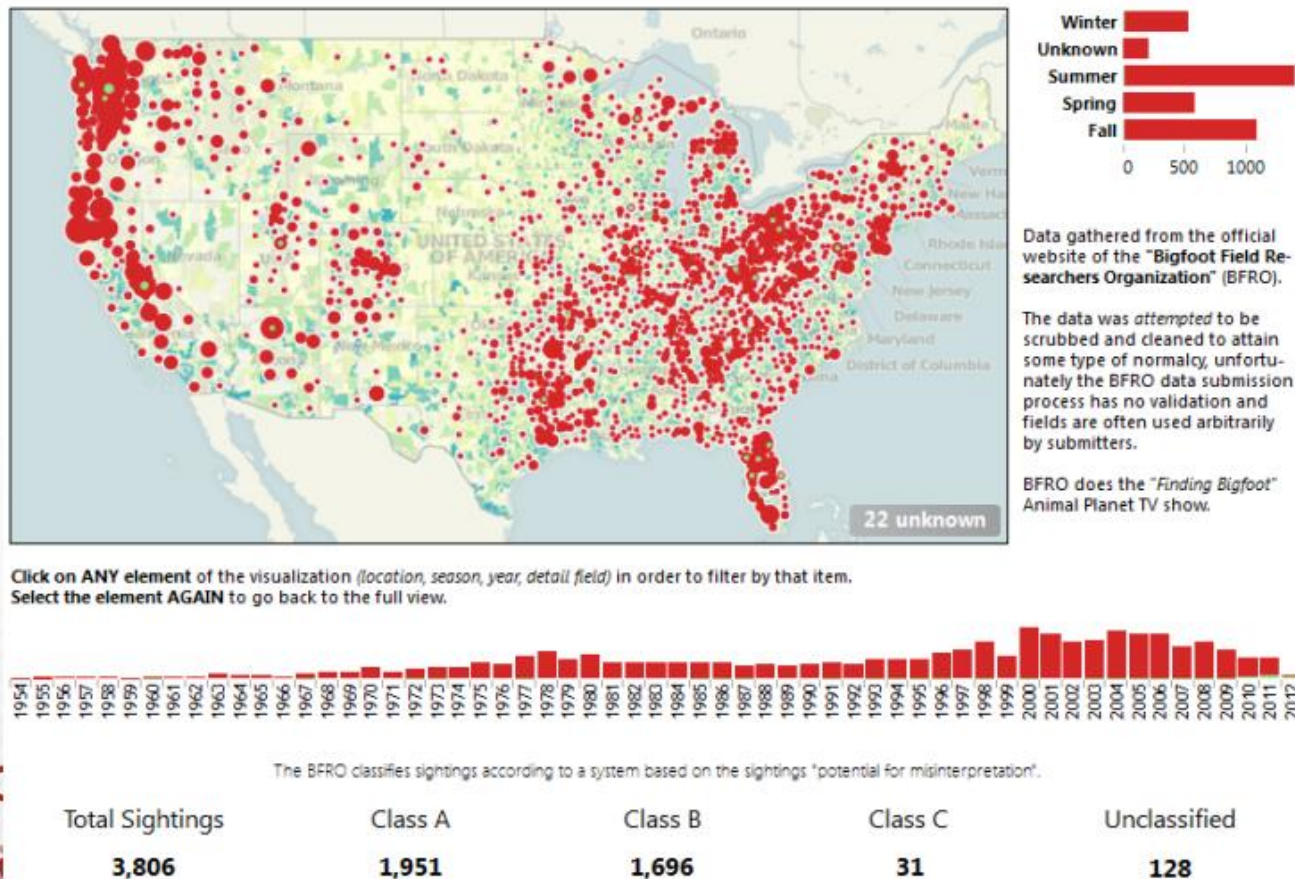
Useful metrics

1. Total sales per state was OK
2. Better: Total sales per million residents per capita is better than looking at existing customers, because we want new customers
3. Top five states to target:
90th percentile +

ACCELERATE UNDERSTANDING

Dashboards should pass the 5-second test

Finding Bigfoot



ACCELERATE UNDERSTANDING

Important rules:

- most important view goes on top or top left
- legends go near their views
- avoid using multiple color schemes
- use 5 views or fewer
- provide interactivity

