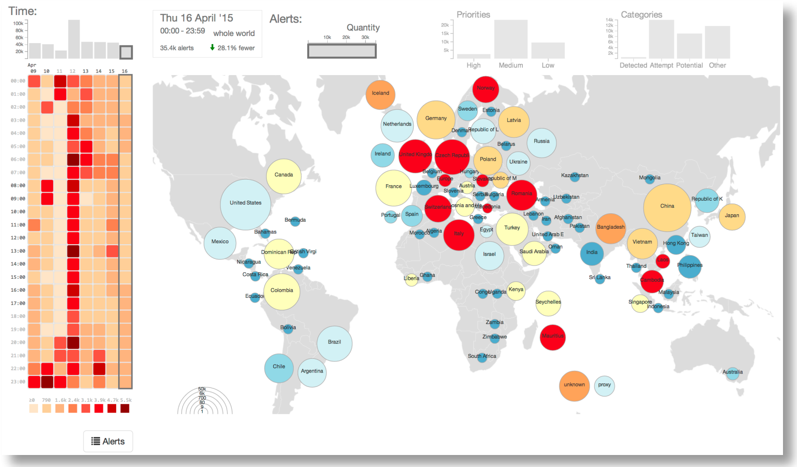
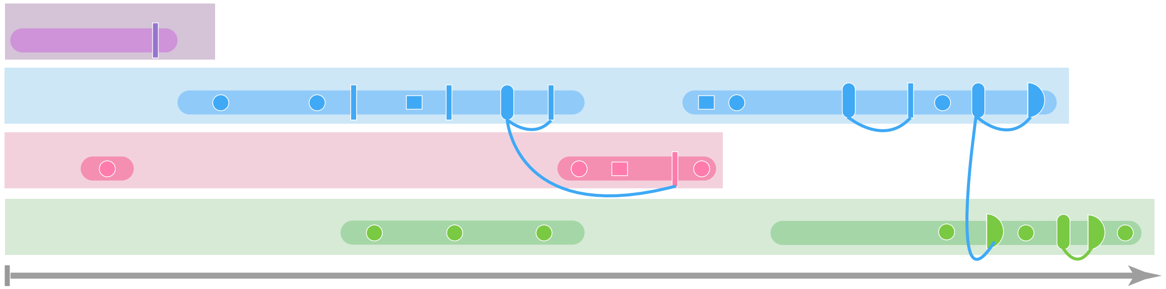
BubbleNet: A Cyber Security Dashboard for Visualizing Patterns



Sean McKenna[1,2] Diane Staheli[2] Cody Fulcher[2] Miriah Meyer[1]

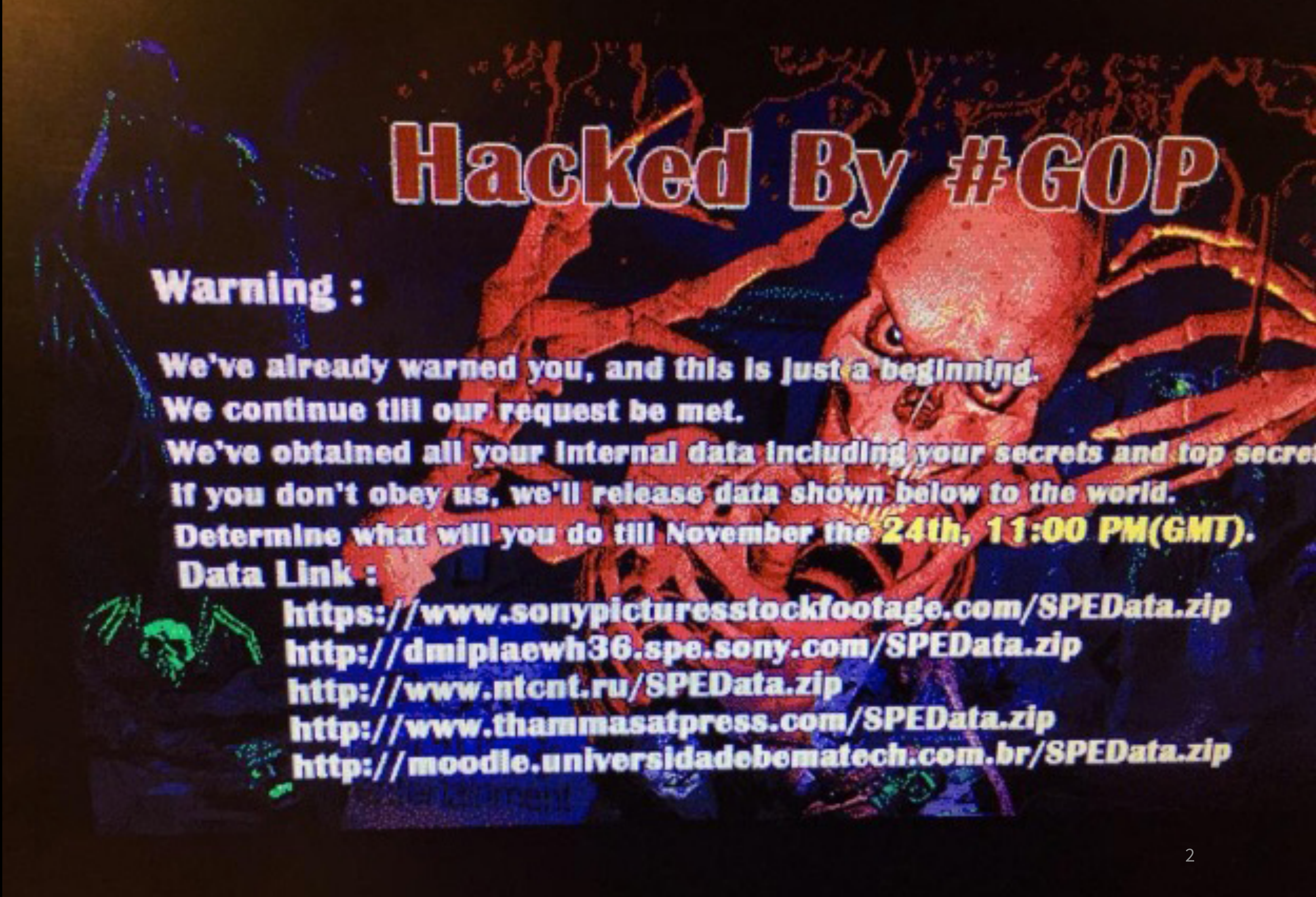
[1] *University of Utah*

[2] *MIT Lincoln Laboratory*



The Lincoln Laboratory portion of this work is sponsored by the Assistant Secretary of Defense for Research & Engineering under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

SONY
PICTURES



Hacked By #GOP

Warning :

We've already warned you, and this is just a beginning.
We continue till our request be met.
We've obtained all your internal data including your secrets and top secre
If you don't obey us, we'll release data shown below to the world.
Determine what will you do till November the 24th, 11:00 PM(GMT).
Data Link :

https://www.sonypicturesstockfootage.com/SPEData.zip
http://dmiplaewh36.spe.sony.com/SPEData.zip
http://www.ntcnt.ru/SPEData.zip
http://www.thammasatpress.com/SPEData.zip
http://moodle.universidadebematech.com.br/SPEData.zip

what was leaked?

“spoiled brat”

“minimally talented”

Challenges in Cyber Security

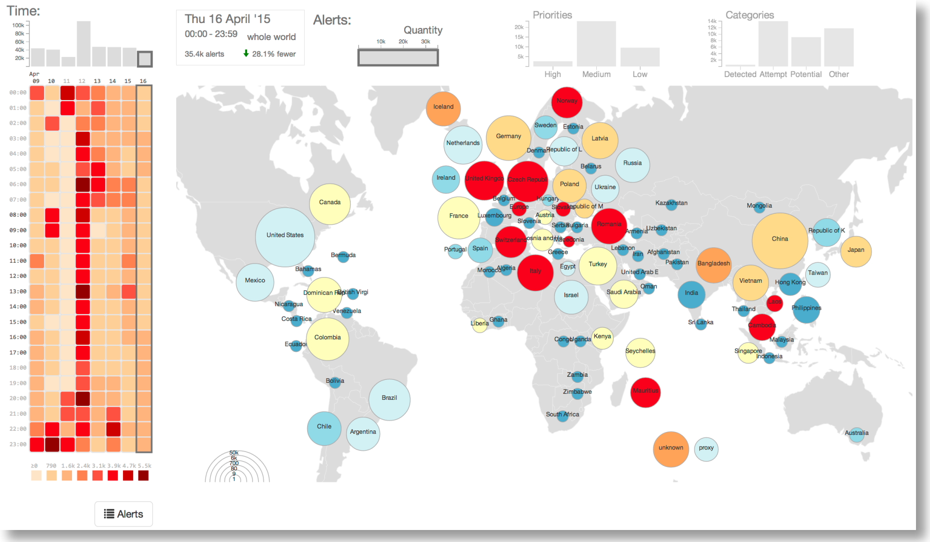
- for analysts
  - large amounts of data
  - requires human interpretation to prevent attacks
  - attacks are robust and ever-changing

- for visualization practitioners
  - analysts can distrust visualization
  - hard to compete with speed

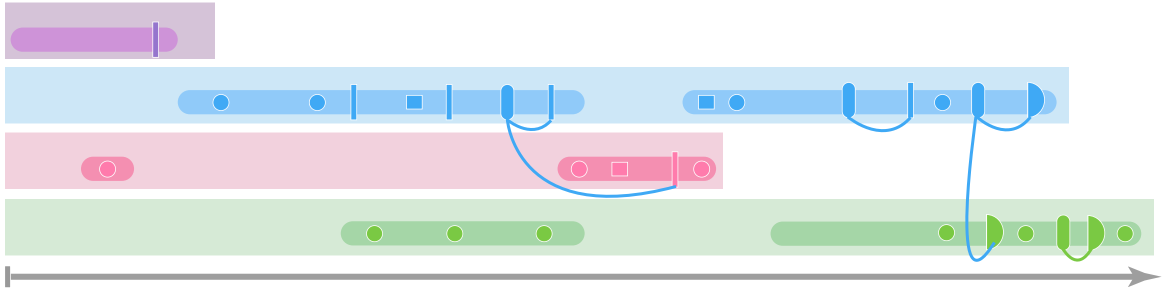
    “current main bottleneck is the ***hard drive read times****”*
  - limited access to both users and data

BubbleNet Dashboard

- conducted a **design study**
  - problem characterization
  - data and task abstraction
  - dashboard design



- focus on the **design process**
  - design methods
  - user evaluation
  - deployment



Cyber Security Visualization Tools

- most cyber security research has focused on novel representations *[Foresti ‘06, Taylor ‘09, Paul ’13, Fowler ‘14, Fischer ‘14]*

- usability and tool effectiveness have been scarcely studied

- very few discussions about tool deployment

- no end-to-end design study

Problem Characterization

- cyber security incidents can result in negative outcomes
  - information disclosure
  - theft
  - denial of service

- to prevent these, analysts find anomalies in data streams

- dashboards are a vital component of data presentation

  *"**pictures are great** when going up to management because you have **60 seconds to make your case**"*

Data and Task Abstraction

- **network record**:
  - metadata associated with the communication between two computers

- **pattern**:
  - collection of *network records* that represent some recurring or abnormal behavior

- analysts must both **discover** & **present** these *patterns*
  - identification and comparison can be supported by aggregation
  - e.g. collecting records by location on the internet

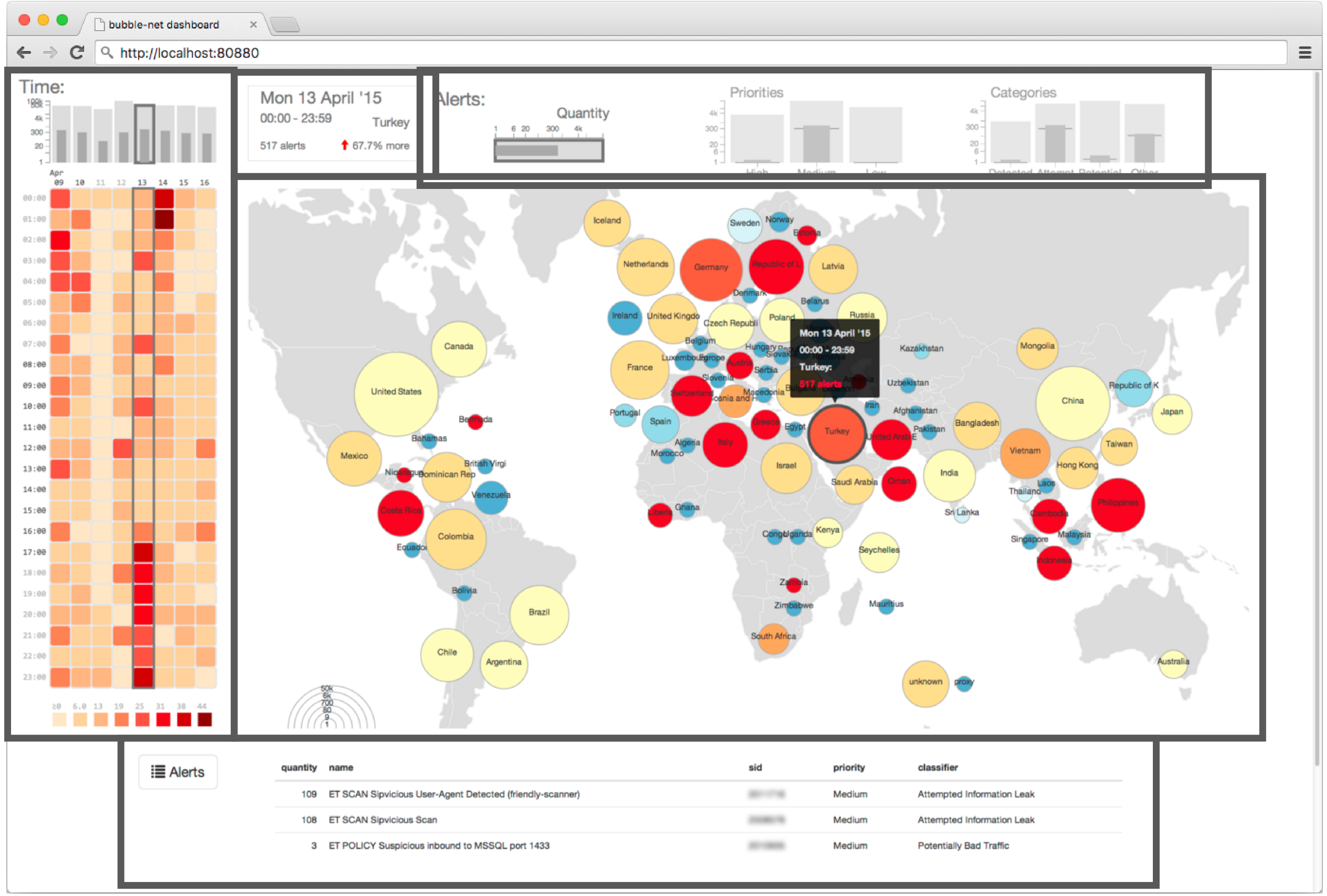
Dataset

- **intrusion detection system** (IDS) data
  - captures **alerts** – these are our *records*
  - rules triggered and may hint at potential incidents
  - requires a priori knowledge

- **aggregation** of alerts
  - by *location*: **country**
  - by *time*: **day** and **hour**
  - store amount of alerts and averages
  - keep links back to original data

BubbleNet Dashboard

- location view
- temporal views
- attribute bullet charts
- record details
- selection overview



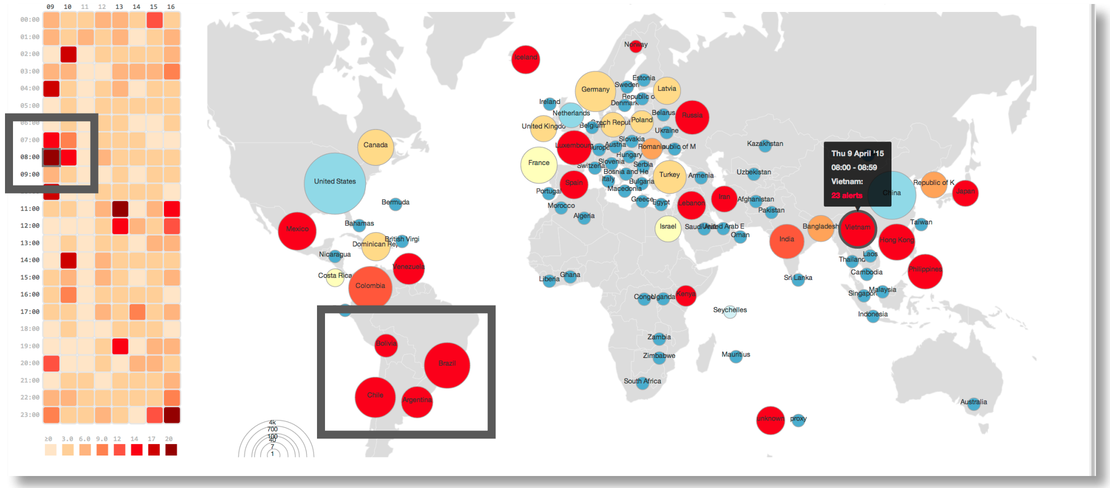
Mon 13 April '15
00:00 - 23:59
Turkey
517 alerts
↑ 67.7% more

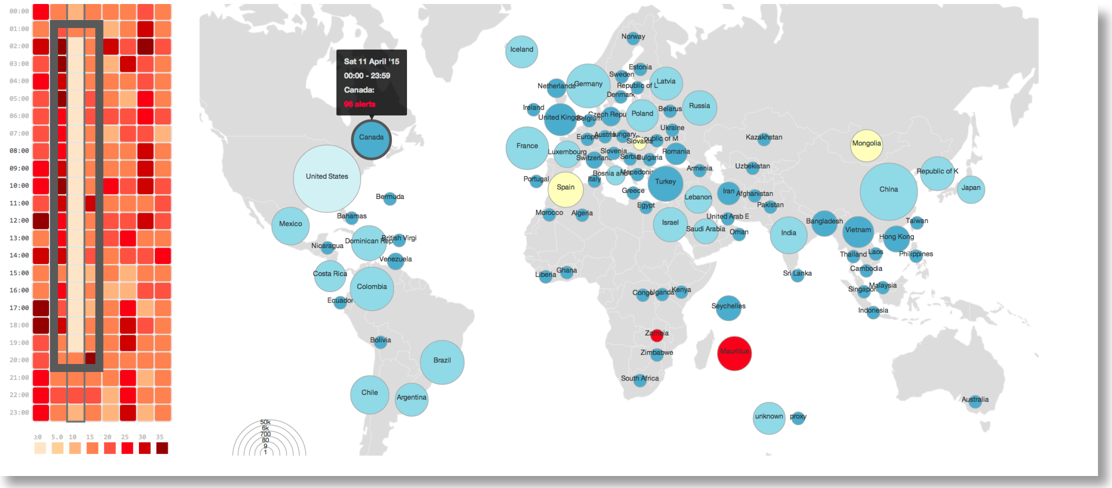
Mon 13 April '15
00:00 - 23:59
Turkey:
517 alerts

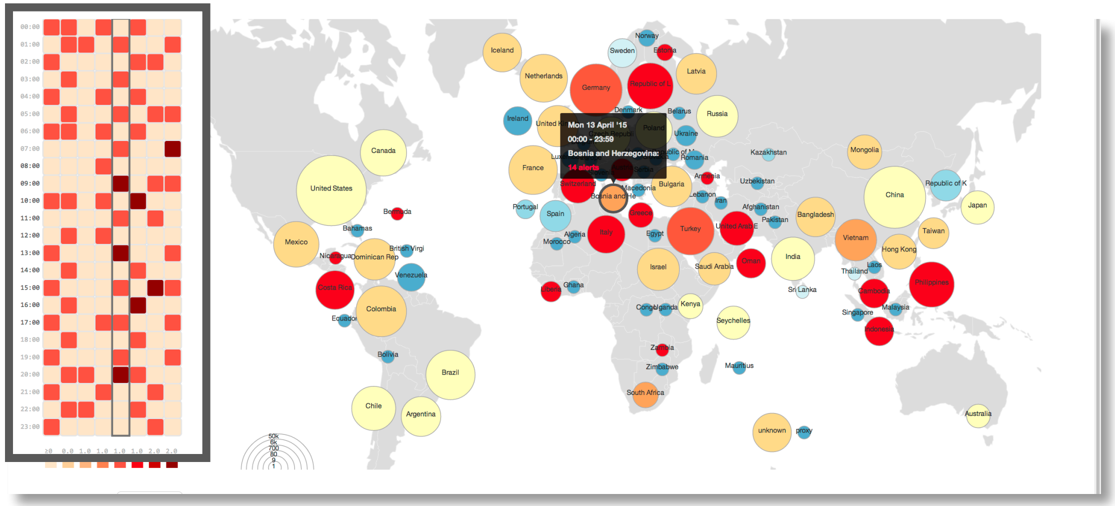
Alerts

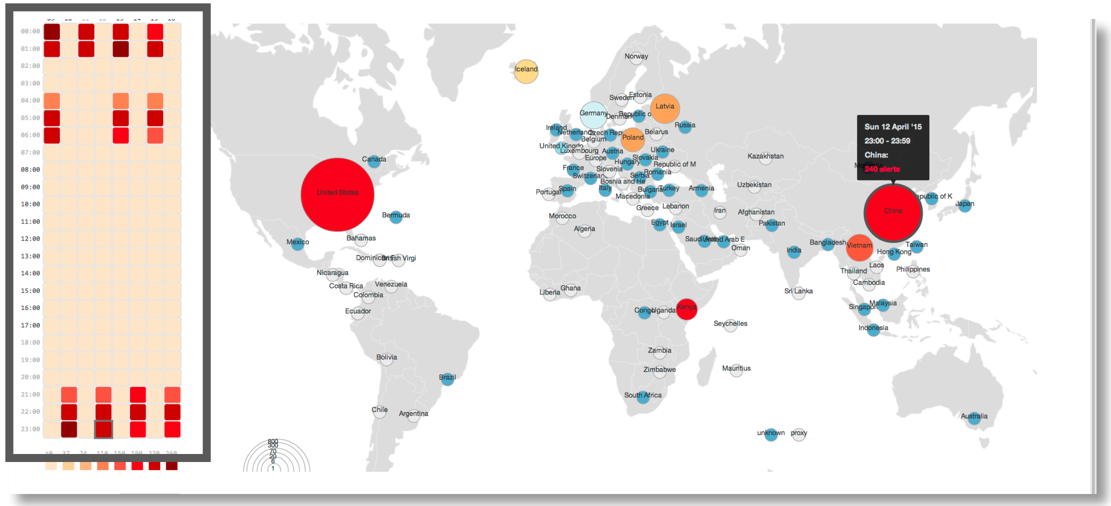
quantity	name	sid	priority	classifier
109	ET SCAN Sipvicious User-Agent Detected (friendly-scanner)	[redacted]	Medium	Attempted Information Leak
108	ET SCAN Sipvicious Scan	[redacted]	Medium	Attempted Information Leak
3	ET POLICY Suspicious inbound to MSSQL port 1433	[redacted]	Medium	Potentially Bad Traffic

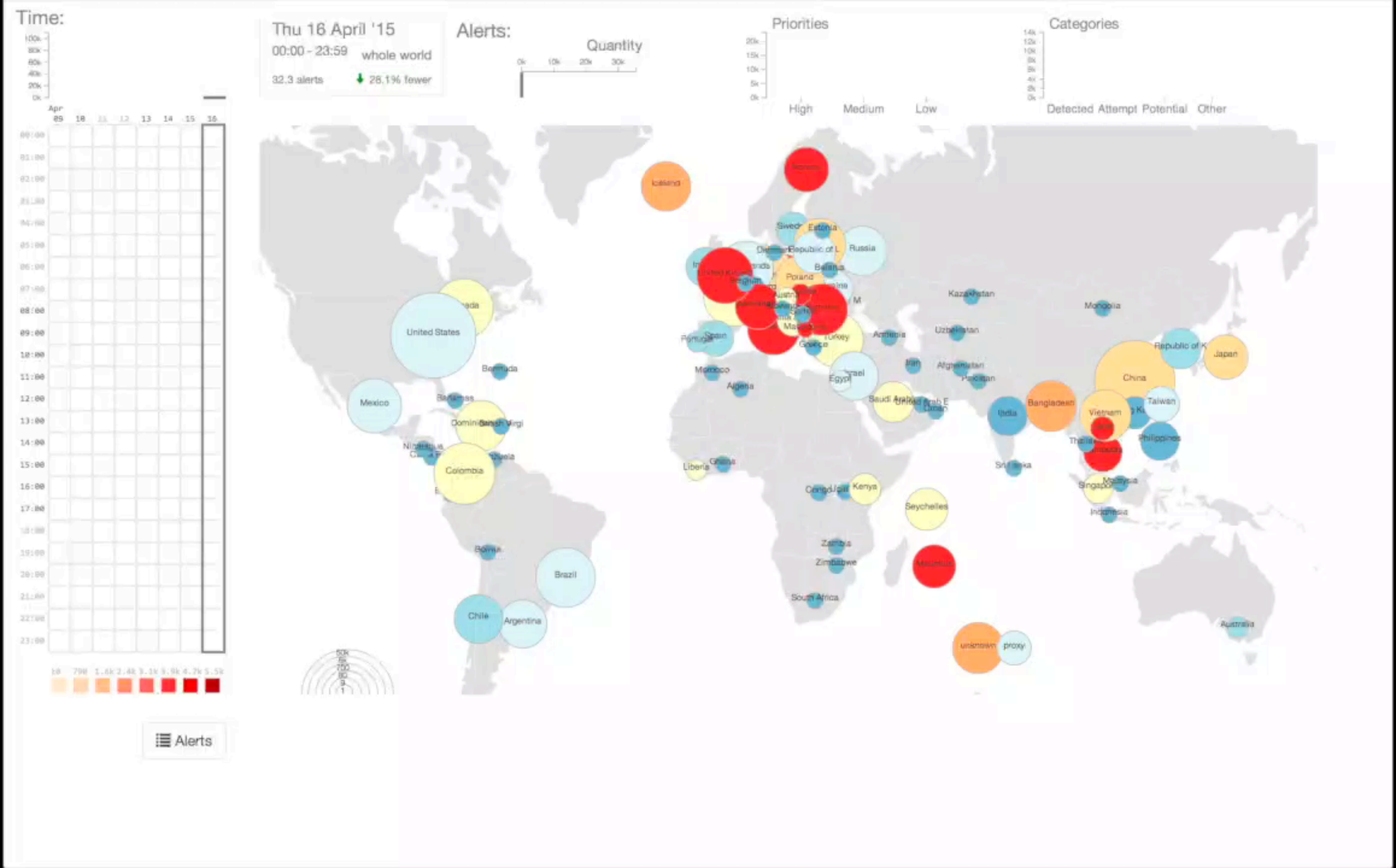
Finding Patterns in BubbleNet


Thu 9 April '15
08:00 - 08:59
Vietnam:
23 alerts


Sat 11 April '15
00:00 - 23:59
Canada:
96 alerts


Mon 13 April '15
00:00 - 23:59
Bosnia and Herzegovina:
14 alerts


Sun 12 April '15
23:00 - 23:59
China:
240 alerts

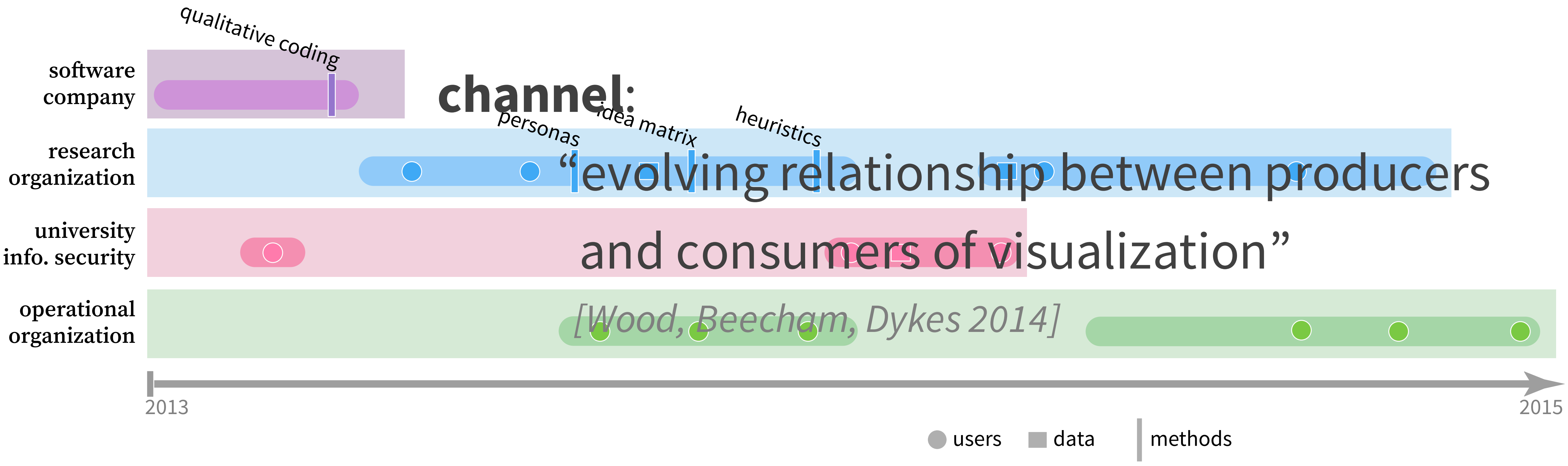


Thu 16 April '15

00:00 - 23:59 whole world

32.3 alerts ⬇ 28.1% fewer

Design Process



channel:

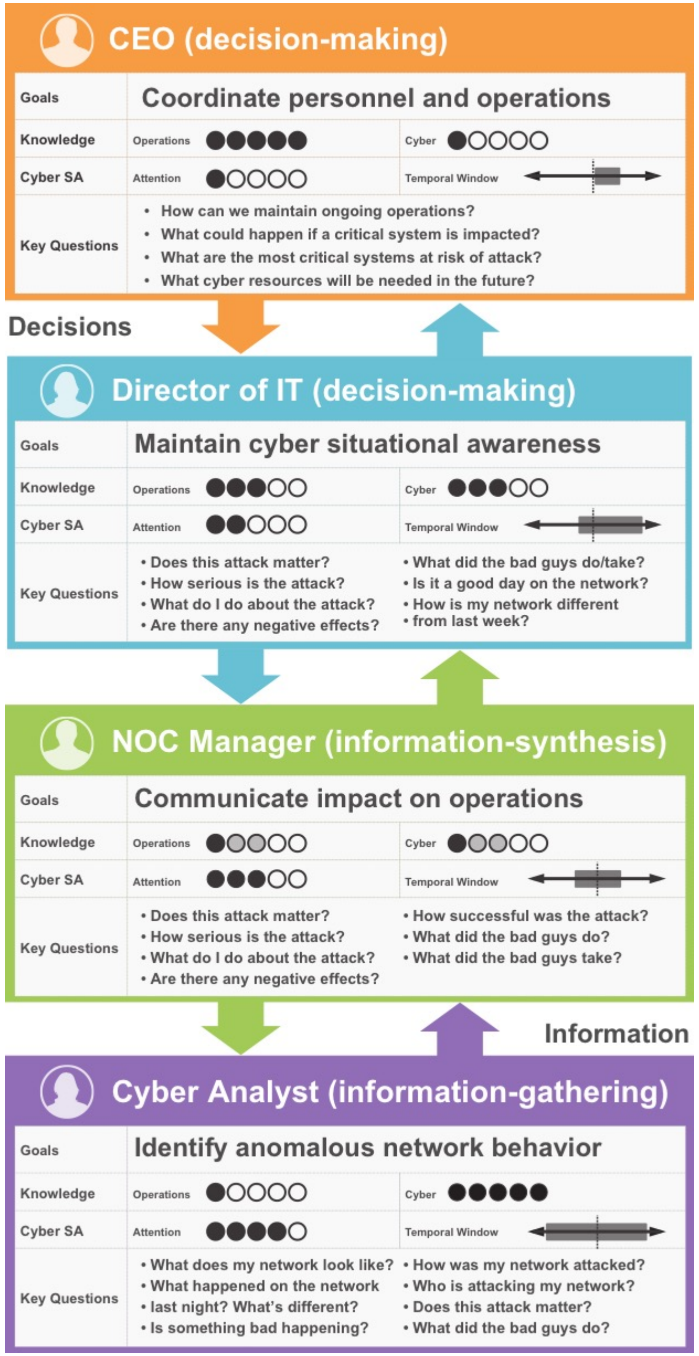
“evolving relationship between producers and consumers of visualization”

[Wood, Beecham, Dykes 2014]

Personas

- identified different potential users
- flow of information and decisions
- selected a subset to focus the design
  - analysts and managers
  - simplified requirements
  - consistent terminology

[McKenna et al. 2015]



CEO (decision-making)

Goals	Coordinate personnel and operations		
Knowledge	Operations ●●●●●	Cyber ●○○○○	
Cyber SA	Attention ●○○○○	Temporal Window	
Key Questions	• How can we maintain ongoing operations? • What could happen if a critical system is impacted? • What are the most critical systems at risk of attack? • What cyber resources will be needed in the future?		

Director of IT (decision-making)

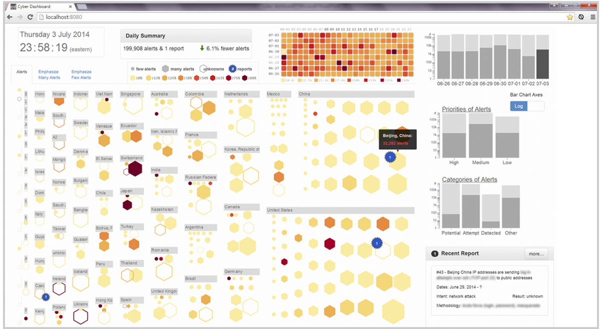
Goals	Maintain cyber situational awareness	
Knowledge	Operations ●●●○○	Cyber ●●●○○
Cyber SA	Attention ●●○○○	Temporal Window
Key Questions	• Does this attack matter? • How serious is the attack? • What do I do about the attack? • Are there any negative effects?	• What did the bad guys do/take? • Is it a good day on the network? • How is my network different • from last week?

NOC Manager (information-synthesis)

Goals	Communicate impact on operations	
Knowledge	Operations ●◐◐○○	Cyber ●◐◐○○
Cyber SA	Attention ●●●○○	Temporal Window
Key Questions	• Does this attack matter? • How serious is the attack? • What do I do about the attack? • Are there any negative effects?	• How successful was the attack? • What did the bad guys do? • What did the bad guys take?

Cyber Analyst (information-gathering)

Goals	Identify anomalous network behavior	
Knowledge	Operations ●○○○○	Cyber ●●●●●
Cyber SA	Attention ●●●●○	Temporal Window
Key Questions	• What does my network look like? • What happened on the network • last night? What's different? • Is something bad happening?	• How was my network attacked? • Who is attacking my network? • Does this attack matter? • What did the bad guys do?



#43 - Beijing China IP addresses are sending [redacted] to public addresses

Dates: June 29, 2014 - ?

Intent: network attack

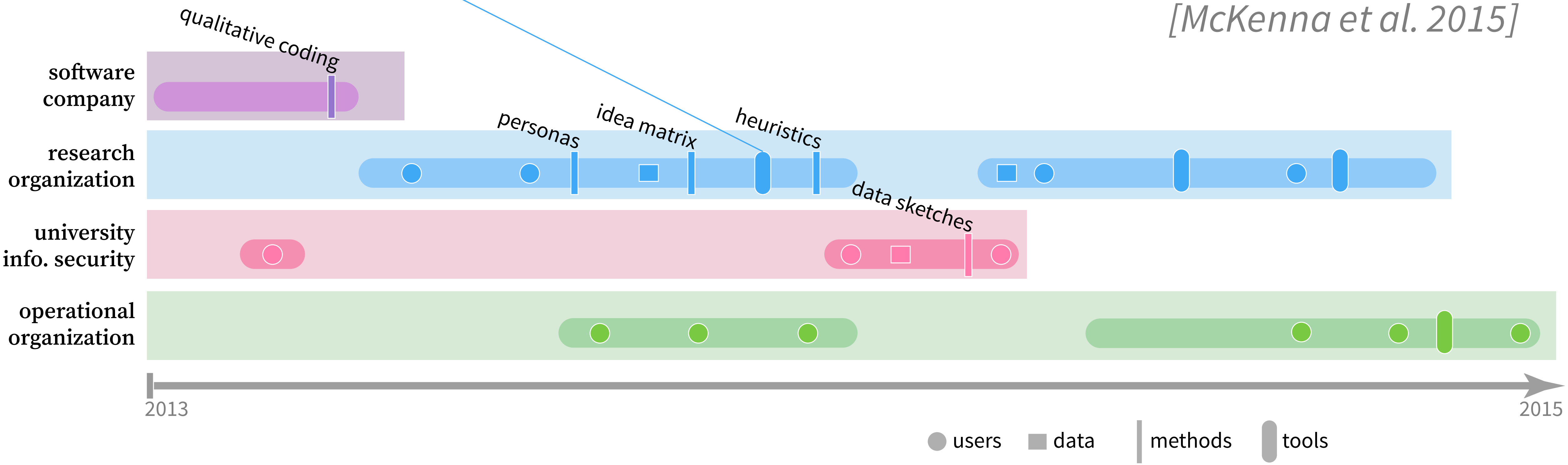
Result: unknown

Methodology: [redacted]

a) prototype I

for more on these **design methods**

[McKenna et al. 2015]



Data Sketches

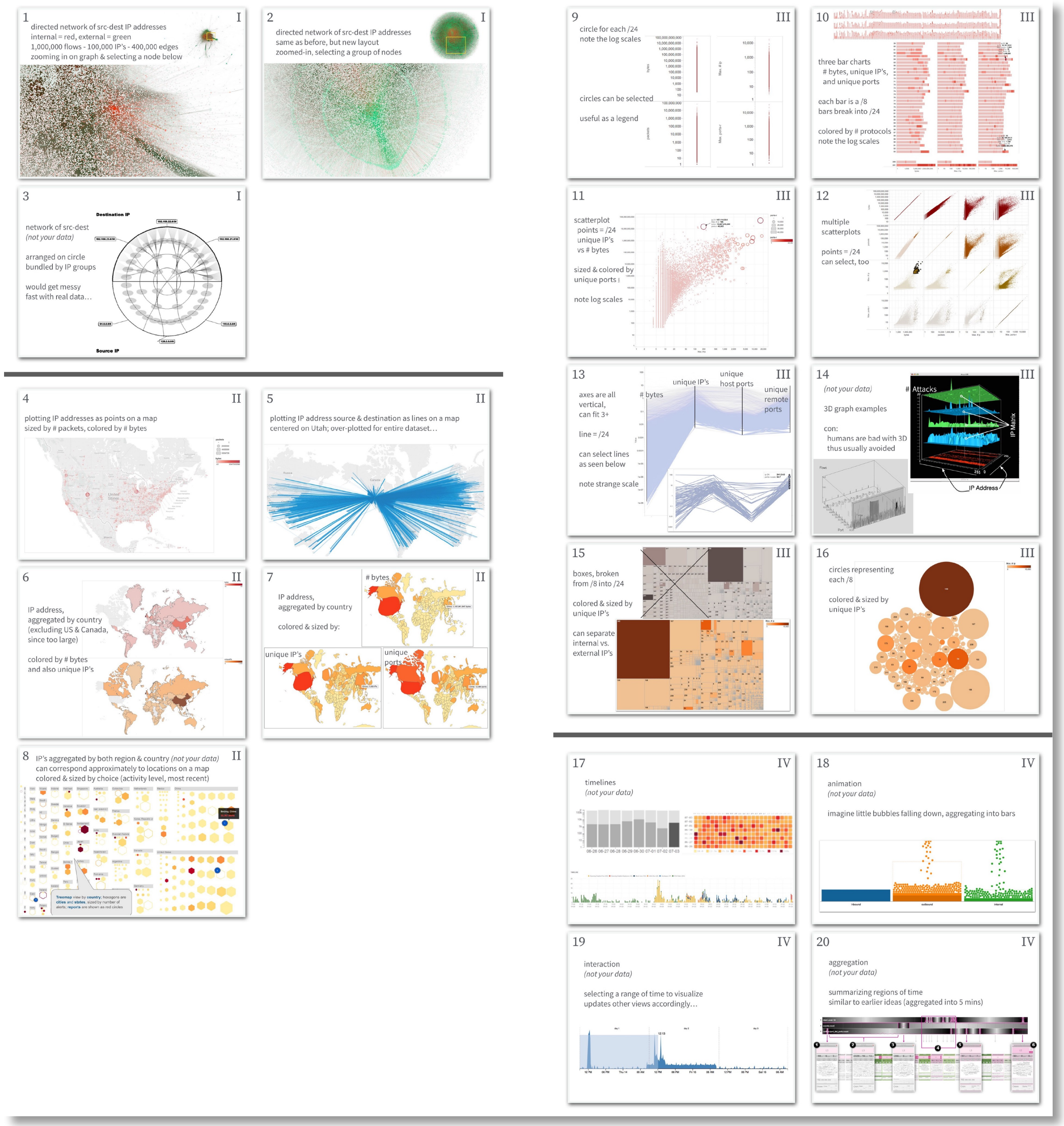
- data-driven sketches, test our abstractions

  [Lloyd & Dykes 2011]

- feedback from analyst

- provided project focus:
  - initial impressions
  - confusing encodings
  - encodings of interest

[McKenna et al. 2015]



1 I

directed network of src-dest IP addresses
internal = red, external = green
1,000,000 flows - 100,000 IP's - 400,000 edges
zooming in on graph & selecting a node below

2 I

directed network of src-dest IP addresses
same as before, but new layout
zoomed-in, selecting a group of nodes

3 I

network of src-dest
(not your data)

arranged on circle
bundled by IP groups

would get messy
fast with real data…

4 II

plotting IP addresses as points on a map
sized by # packets, colored by # bytes

5 II

plotting IP address source & destination as lines on a map
centered on Utah; over-plotted for entire dataset…

6 II

IP address,
aggregated by country
(excluding US & Canada,
since too large)

colored by # bytes
and also unique IP's

7 II

IP address,
aggregated by country

colored & sized by:

8 II

IP's aggregated by both region & country *(not your data)*
can correspond approximately to locations on a map
colored & sized by choice (activity level, most recent)

Treemap view by country; hexagons are cities and states, sized by number of alerts; reports are shown as red circles

9 III

circle for each /24
note the log scales

circles can be selected

useful as a legend

10 III

three bar charts
# bytes, unique IP's,
and unique ports

each bar is a /8
bars break into /24

colored by # protocols
note the log scales

11 III

scatterplot
points = /24
unique IP's
vs # bytes

sized & colored by
unique ports

note log scales

12 III

multiple
scatterplots

points = /24
can select, too

13 III

axes are all
vertical,
can fit 3+

line = /24

can select lines
as seen below

note strange scale

14 III

(not your data)

3D graph examples

con:
humans are bad with 3D
thus usually avoided

15 III

boxes, broken
from /8 into /24

colored & sized by
unique IP's

can separate
internal vs.
external IP's

16 III

circles representing
each /8

colored & sized by
unique IP's

17 IV

timelines
(not your data)

18 IV

animation
(not your data)

imagine little bubbles falling down, aggregating into bars

19 IV

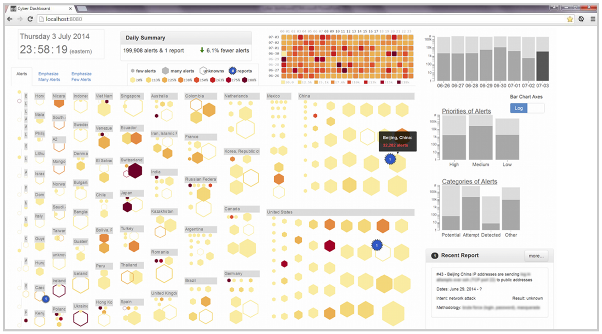
interaction
(not your data)

selecting a range of time to visualize
updates other views accordingly…

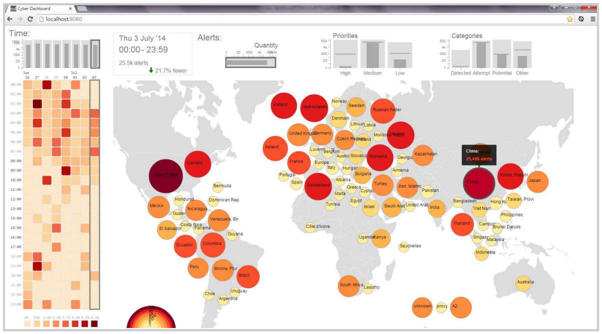
20 IV

aggregation
(not your data)

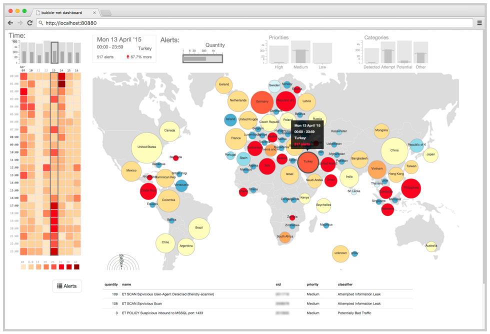
summarizing regions of time
similar to earlier ideas (aggregated into 5 mins)



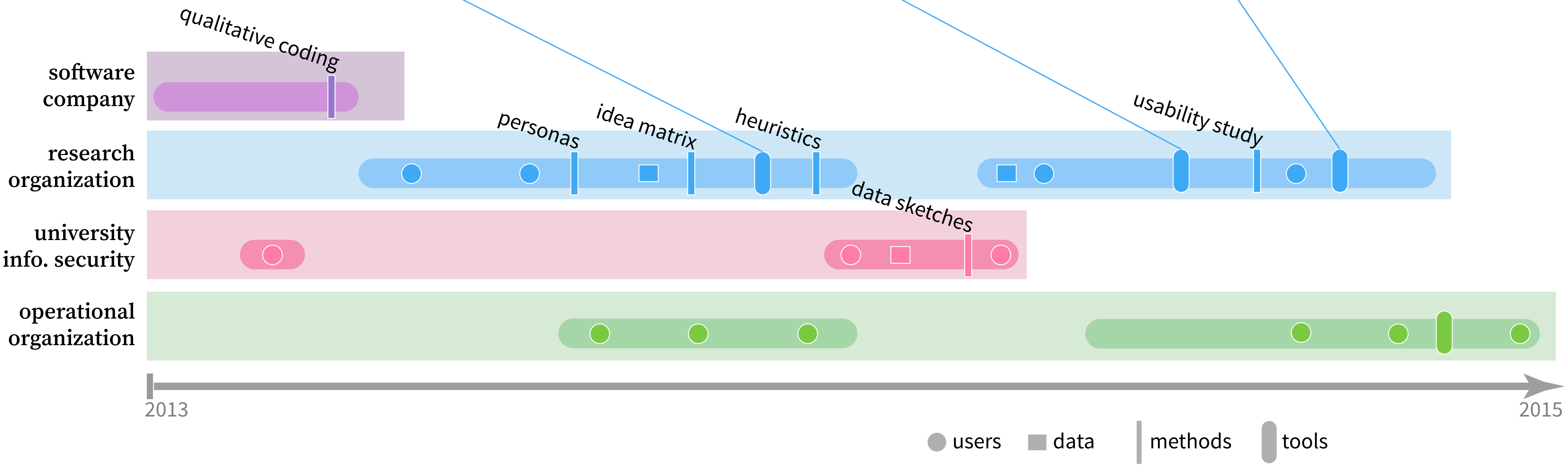
a) prototype I



b) prototype II



c) BubbleNet dashboard



Evaluation

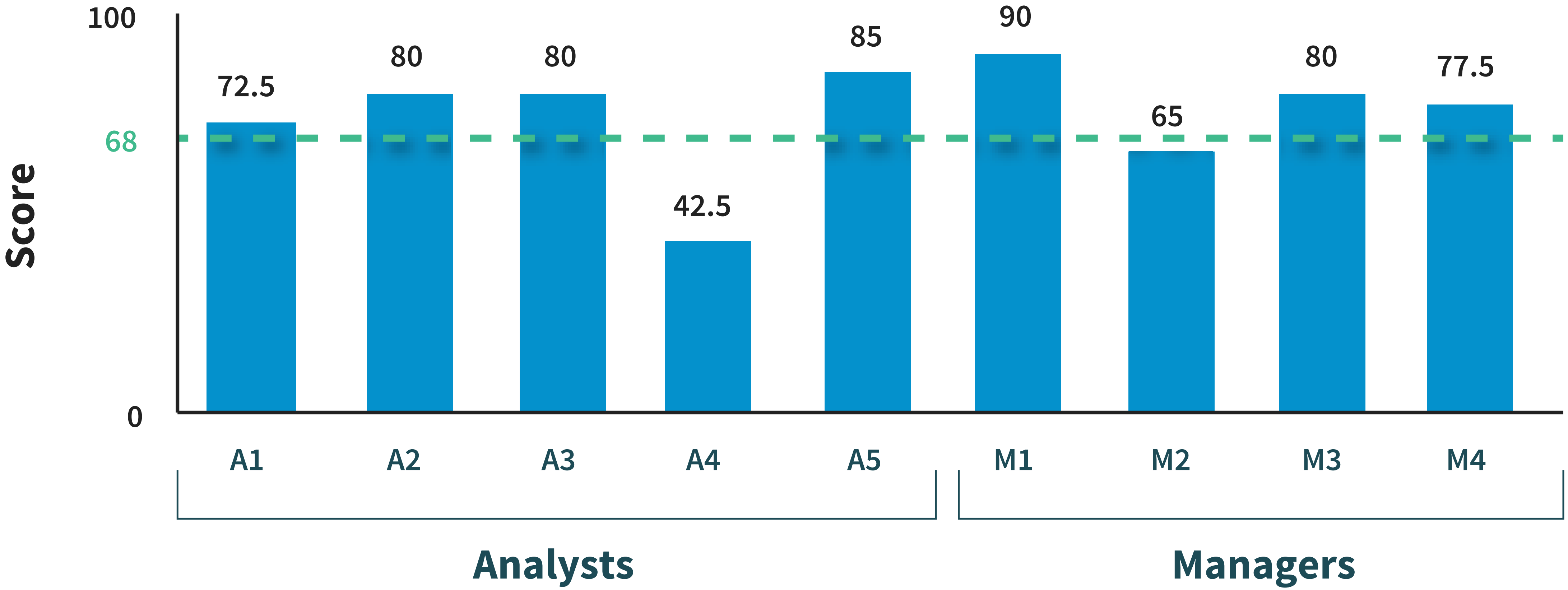
- user study
  - 5 analysts, 4 managers
  - 1-hour long, training + scenarios

- system usability scale (SUS) *[Sauro 2011]*
  - 10 questions on usability
  - yields score out of 100
  - standardized across many user interfaces

Evaluation

BubbleNet’s score: **75 / 100**

System Usability Score by User



Evaluation

- system usability scale
  - validates general principles and interaction paradigms
  - limited to usability

- think-aloud session + qualitative coding
  - pulled out key successes of the project
  - e.g. temporal pattern detection, focus on patterns, interaction feedback

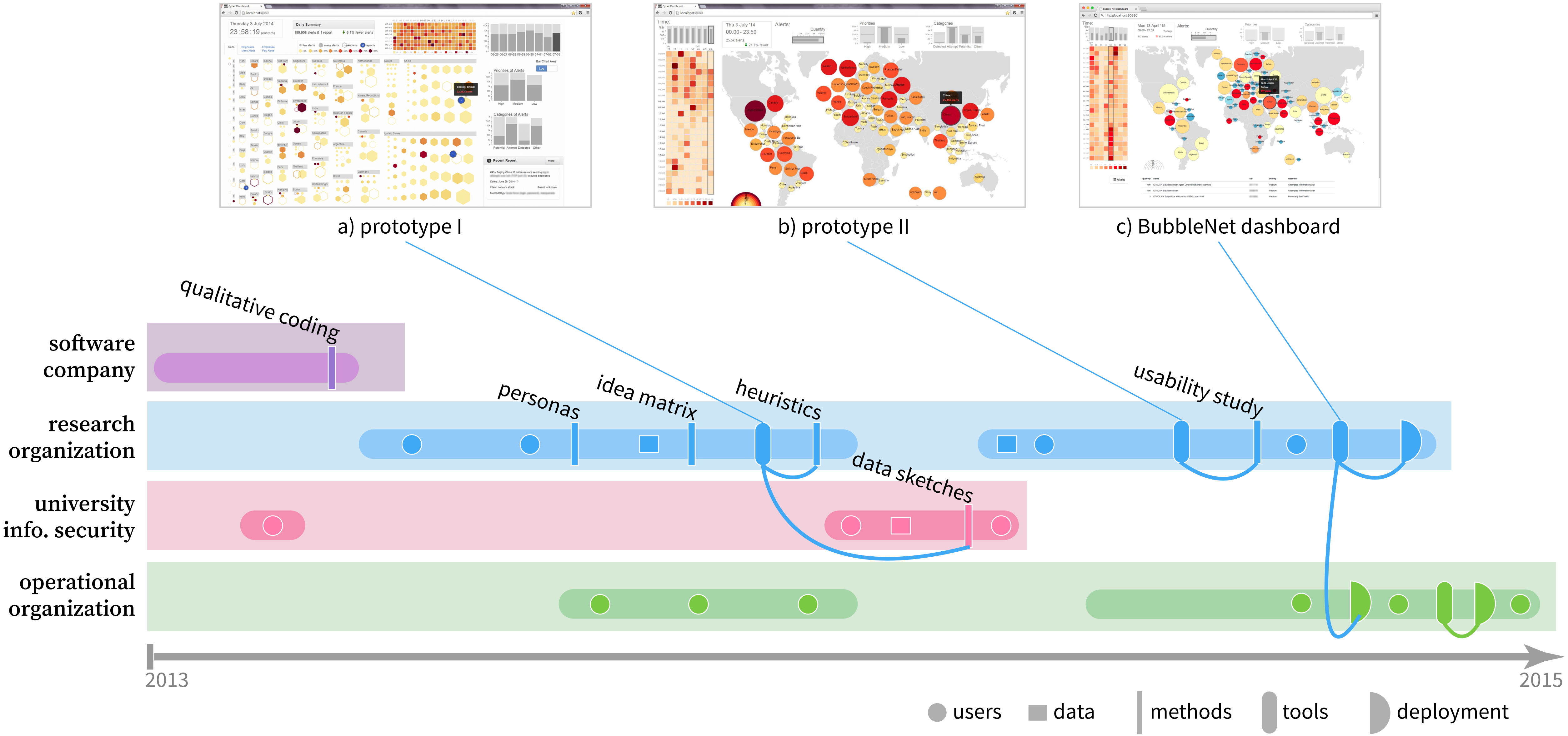
Evaluation

*“I keep getting **drawn to the heatmap** and these darker areas, because they **certainly stand out**”*

*“the majority of what we are looking for is **patterns** and this just makes patterns which is **faster**”*

*“it’s very **responsive and dynamic**; the fact that it changes as I narrow [in] is the best”*

*“I could write a splunk query to do this, but **this is easier**”*



a) prototype I

b) prototype II

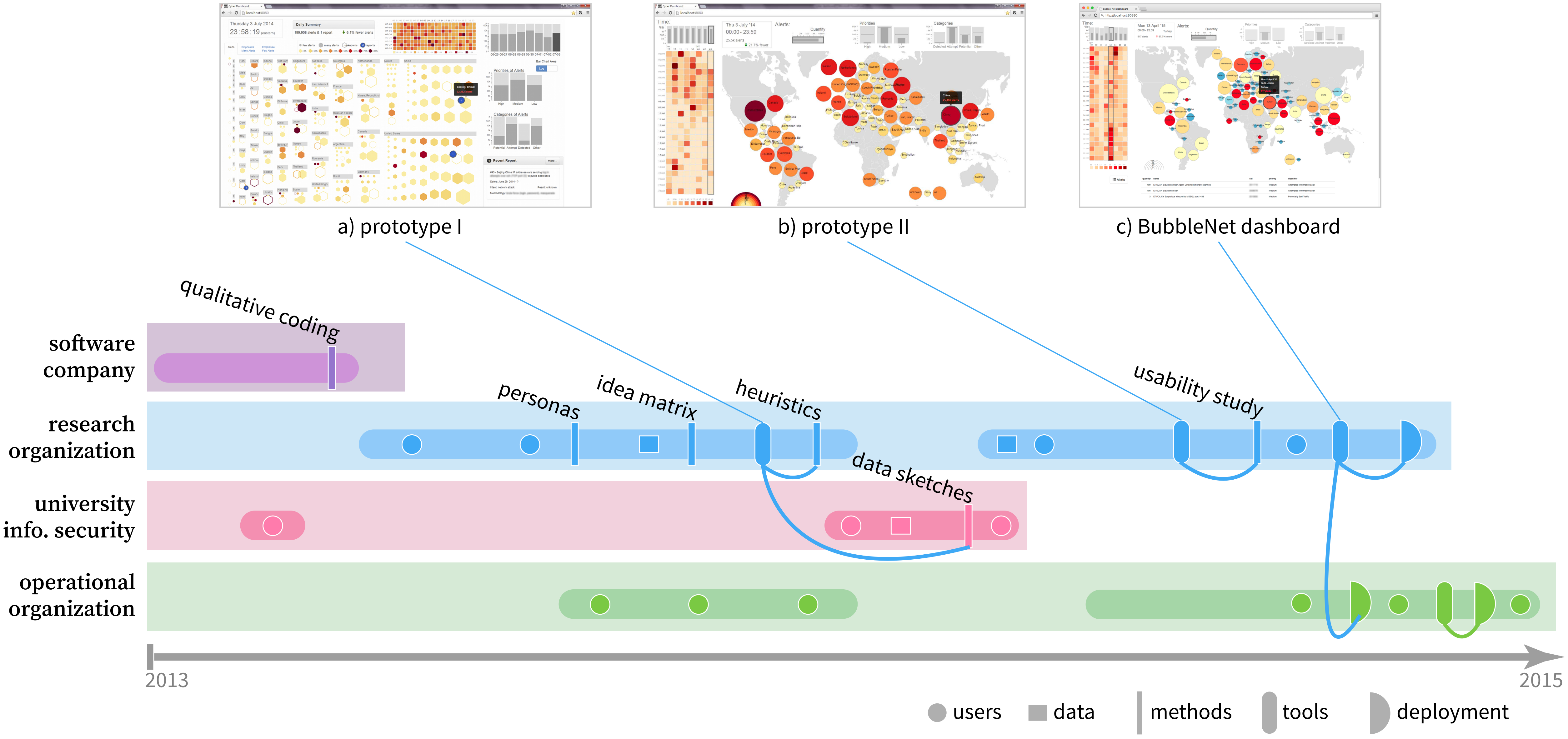
c) BubbleNet dashboard

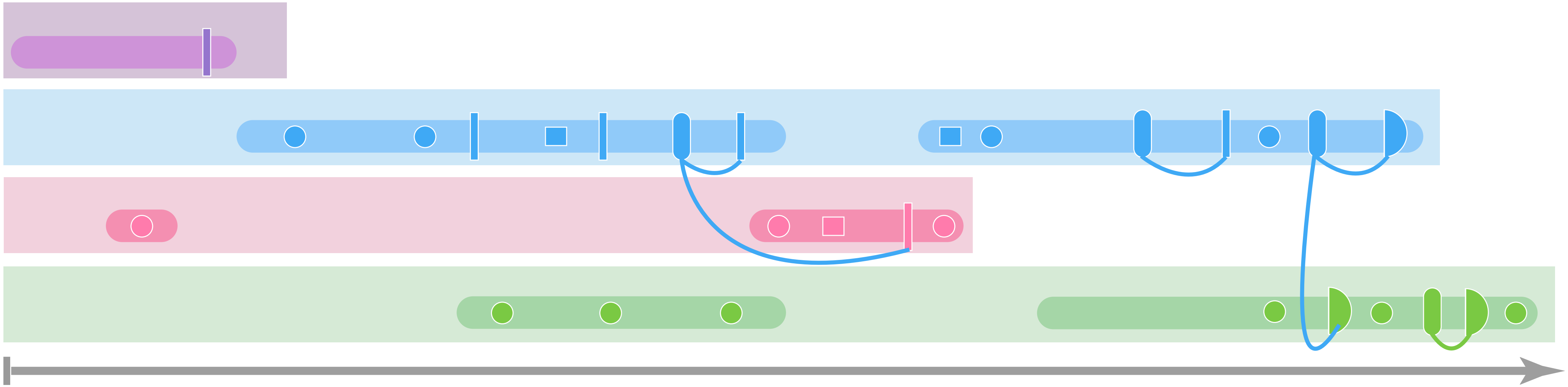
Reflections

- needs of cyber security analysts and managers are unique and challenging to accommodate simultaneously

- winnowing and casting of user roles occurred later in the design process

- task of presentation involves two or more parties, so there were users beyond just a data analyst to consider


a) prototype I
b) prototype II
c) BubbleNet dashboard
qualitative coding
software company
personas
idea matrix
heuristics
usability study
research organization
data sketches
university info. security
operational organization
2013
2015
users
data
methods
tools
deployment



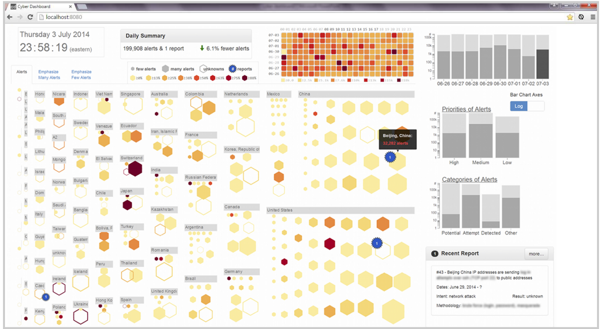
to find out more…

http://mckennapsean.com/projects/bubble-net

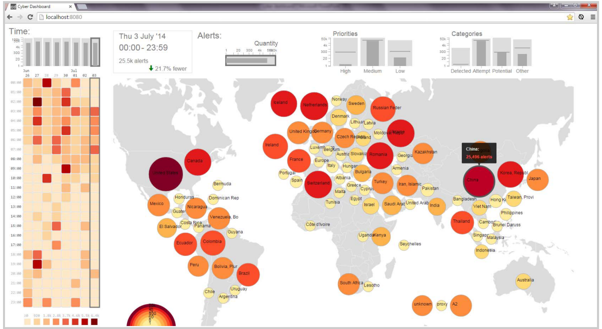
sean@cs.utah.edu

acknowledgements: Jonzy, Dan Bowden, Tamara Denning, staff members at MIT Lincoln Laboratory, and the Visualization Design Lab

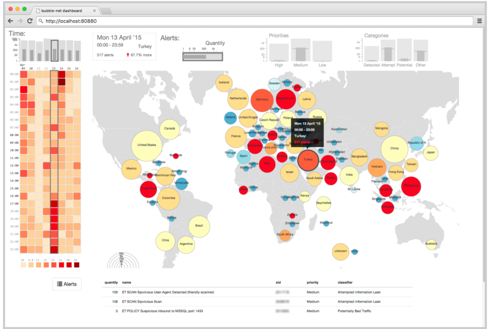
visualization design lab



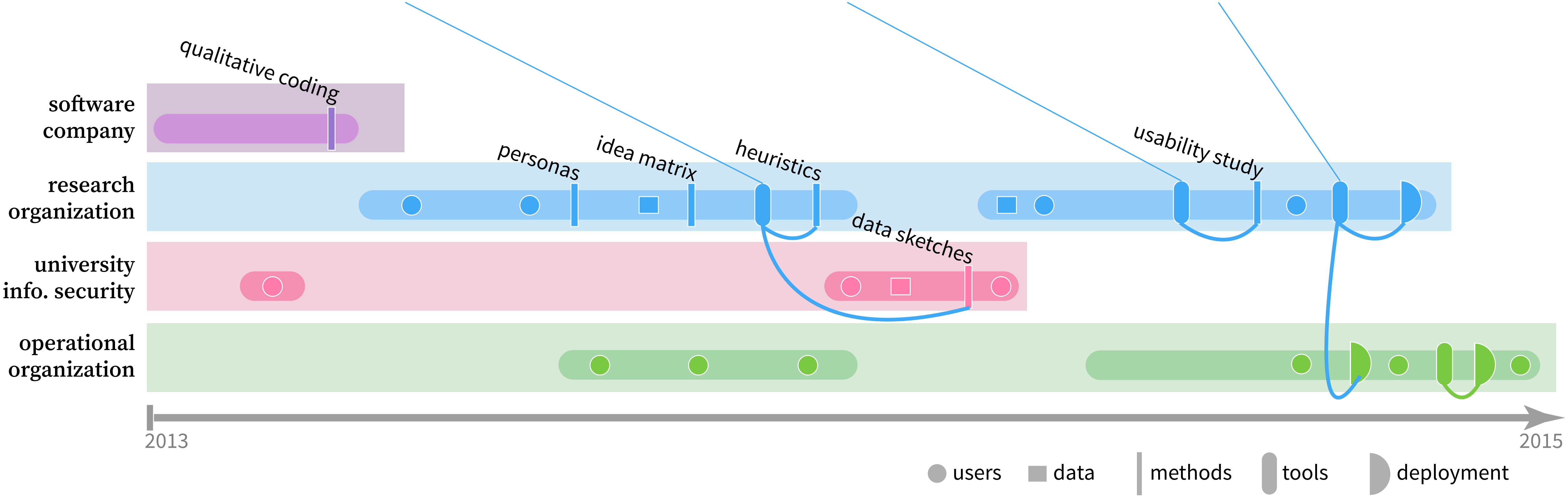
a) prototype I



b) prototype II



c) BubbleNet dashboard



qualitative coding
software company
personas
idea matrix
heuristics
usability study
research organization
data sketches
university info. security
operational organization
2013
2015
users
data
methods
tools
deployment