Security in knowledge

Data Analysis & Visualization for Security Professionals

Jay Jacobs

Verizon

Bob Rudis

Liberty Mutual Insurance

RSA CONFERENCE 2013

Session ID: GRC-T18

Session Classification: Intermediate

Key Learning Points

Key Learning Points

- data helps our understanding of our environment

Key Learning Points

- data helps our understanding of our environment
- solutions are more from thinking than buying

Key Learning Points

- data helps our understanding of our environment
- solutions are more from thinking than buying
- visualizations help communicate complexity quickly

Key Learning Points

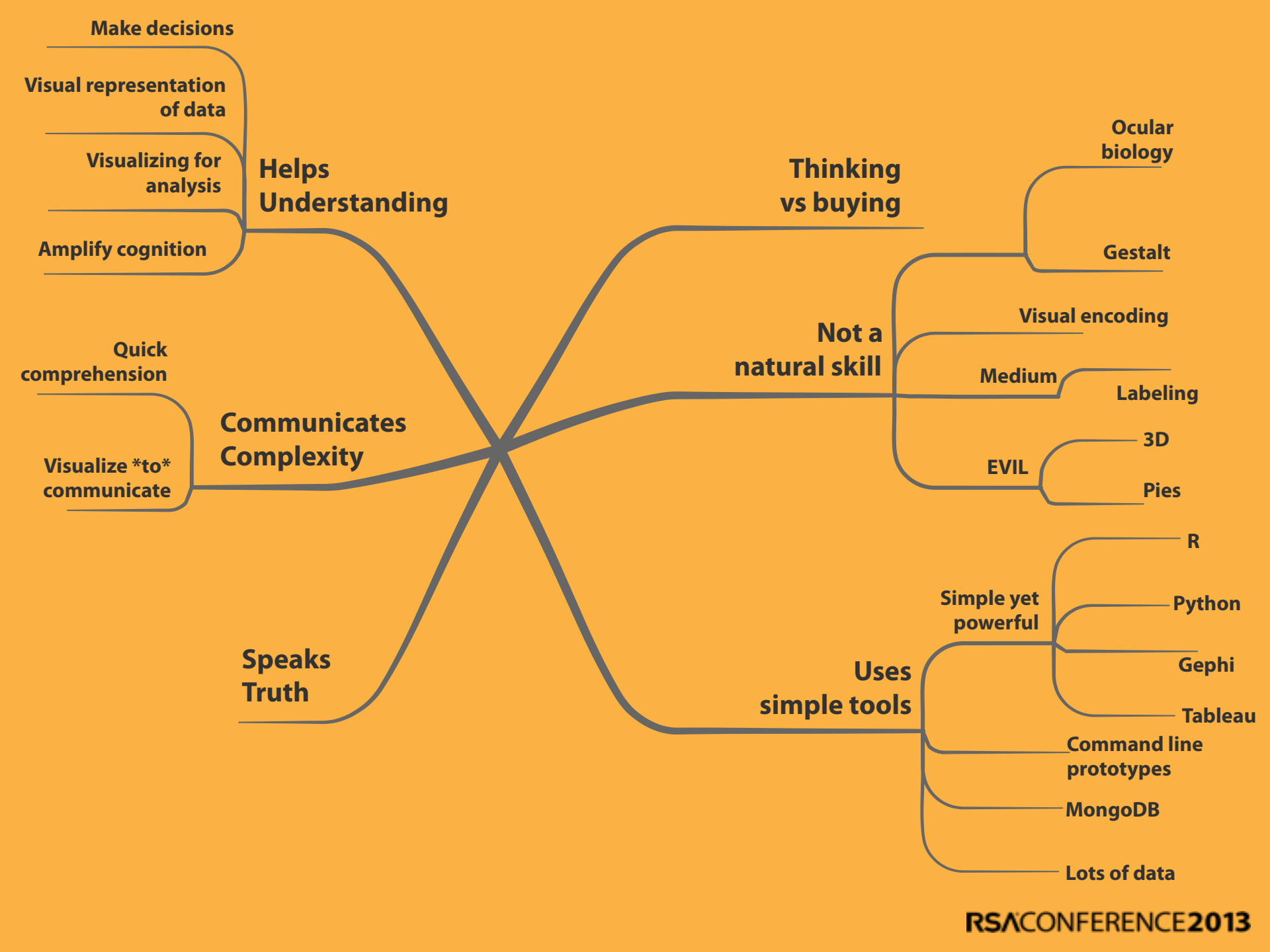
- data helps our understanding of our environment
- solutions are more from thinking than buying
- visualizations help communicate complexity quickly
- data visualization is not a natural skill, it must be learned

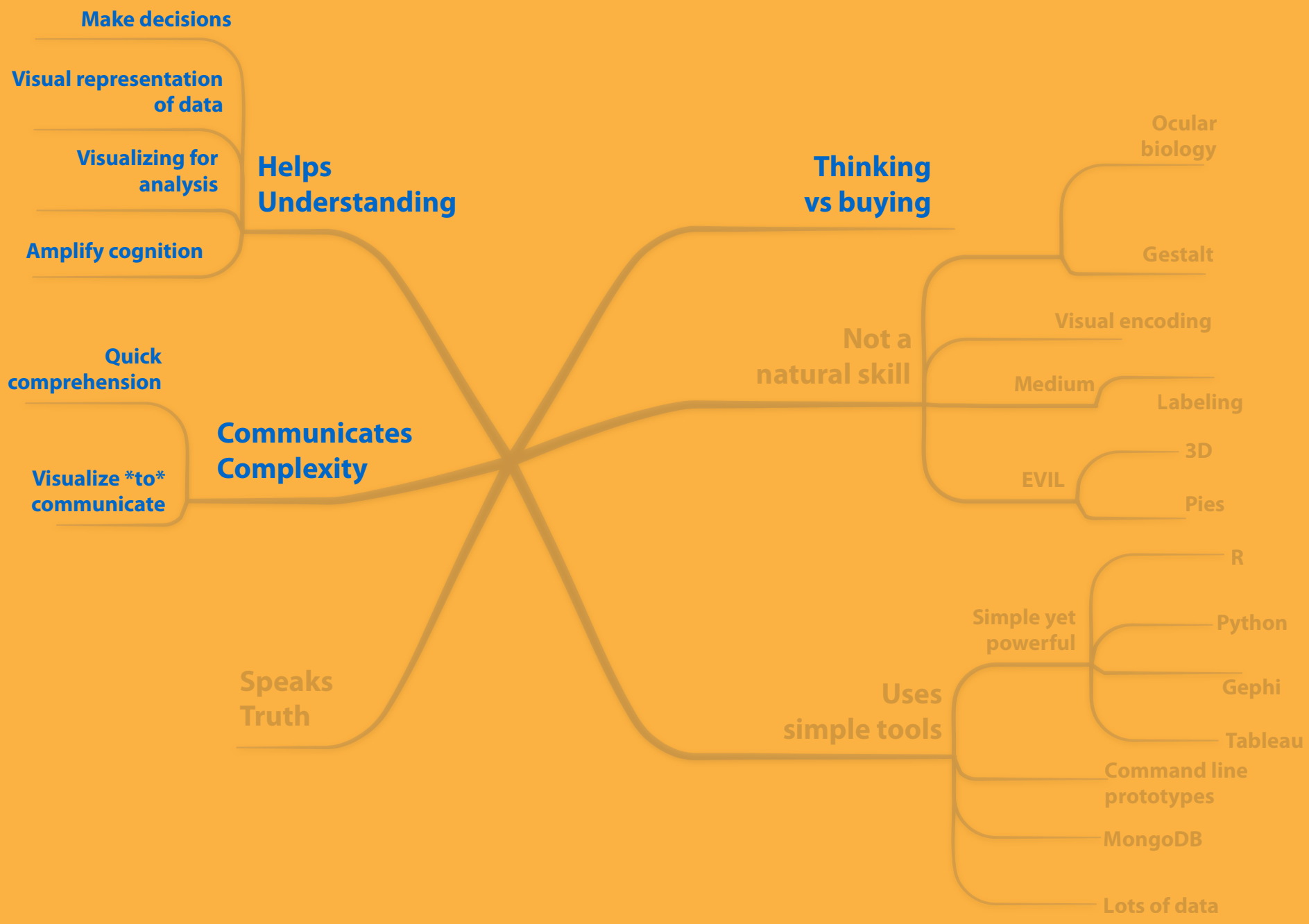
Key Learning Points

- data helps our understanding of our environment
- solutions are more from thinking than buying
- visualizations help communicate complexity quickly
- data visualization is not a natural skill, it must be learned
- be truthful: message should match the data

Key Learning Points

- data helps our understanding of our environment
- solutions are more from thinking than buying
- visualizations help communicate complexity quickly
- data visualization is not a natural skill, it must be learned
- be truthful: message should match the data
- simple tools can be, data scientist you need not be


Make decisions
Visual representation of data
Visualizing for analysis
Amplify cognition
Helps Understanding
Quick comprehension
Visualize *to* communicate
Communicates Complexity
Speaks Truth
Thinking vs buying
Ocular biology
Gestalt
Not a natural skill
Visual encoding
Medium
Labeling
EVIL
3D
Pies
Uses simple tools
Simple yet powerful
R
Python
Gephi
Tableau
Command line prototypes
MongoDB
Lots of data


Make decisions
Visual representation of data
Visualizing for analysis
Amplify cognition
Helps Understanding
Quick comprehension
Visualize *to* communicate
Communicates Complexity
Speaks Truth
Thinking vs buying
Not a natural skill
Ocular biology
Gestalt
Visual encoding
Medium
Labeling
EVIL
3D
Pies
Uses simple tools
Simple yet powerful
R
Python
Gephi
Tableau
Command line prototypes
MongoDB
Lots of data

“…use information to better understand our world and make more informed decisions …”

Stephen Few

“…use inform
better unders
world and make more
informed decisions …”

Data helps our understanding of our environment

Stephen Few

Our Goal:

To **amplify cognition of data** through **visual representation** and **presentation**.

Our Goal:

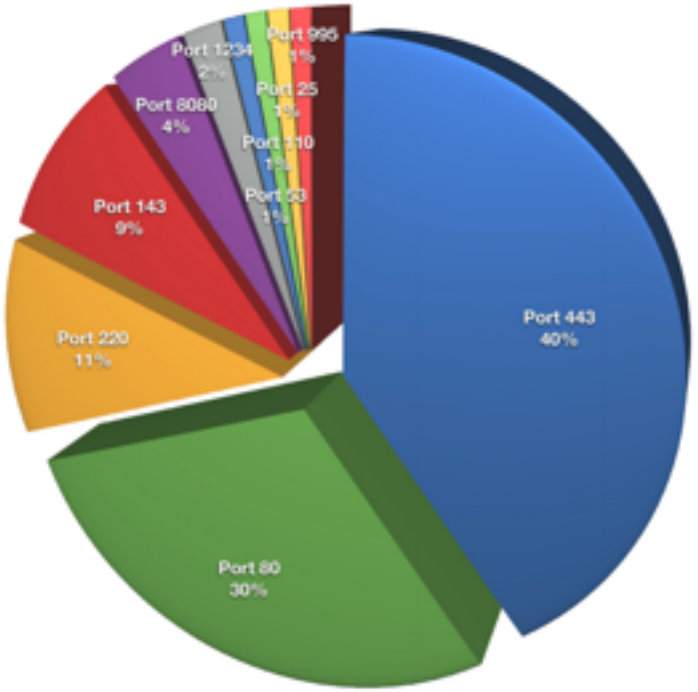
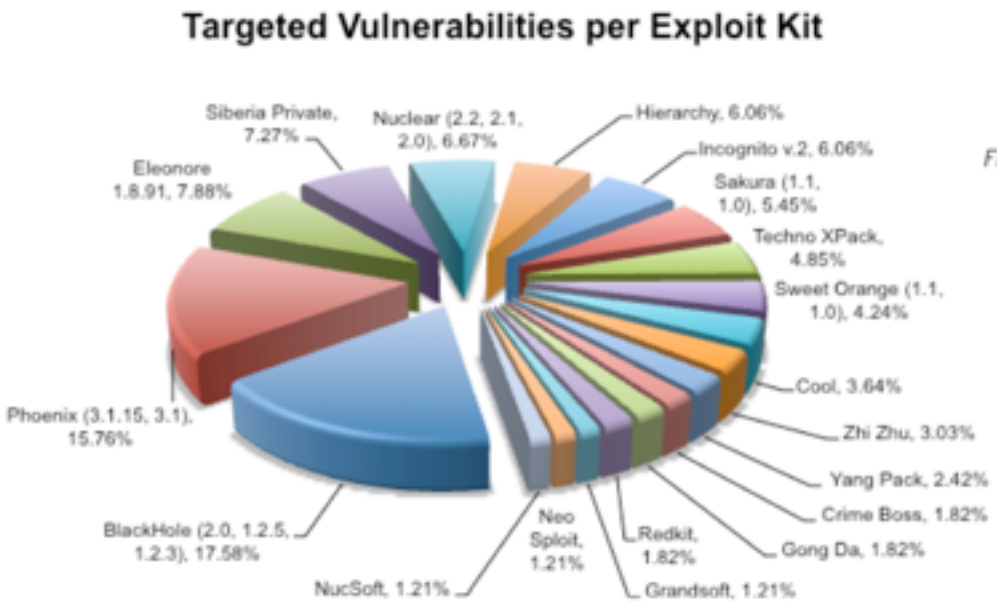


Figure 5: Most commonly used ports by PoisonIvy samples found in Japan from 2008 to 2012



Our Goal:

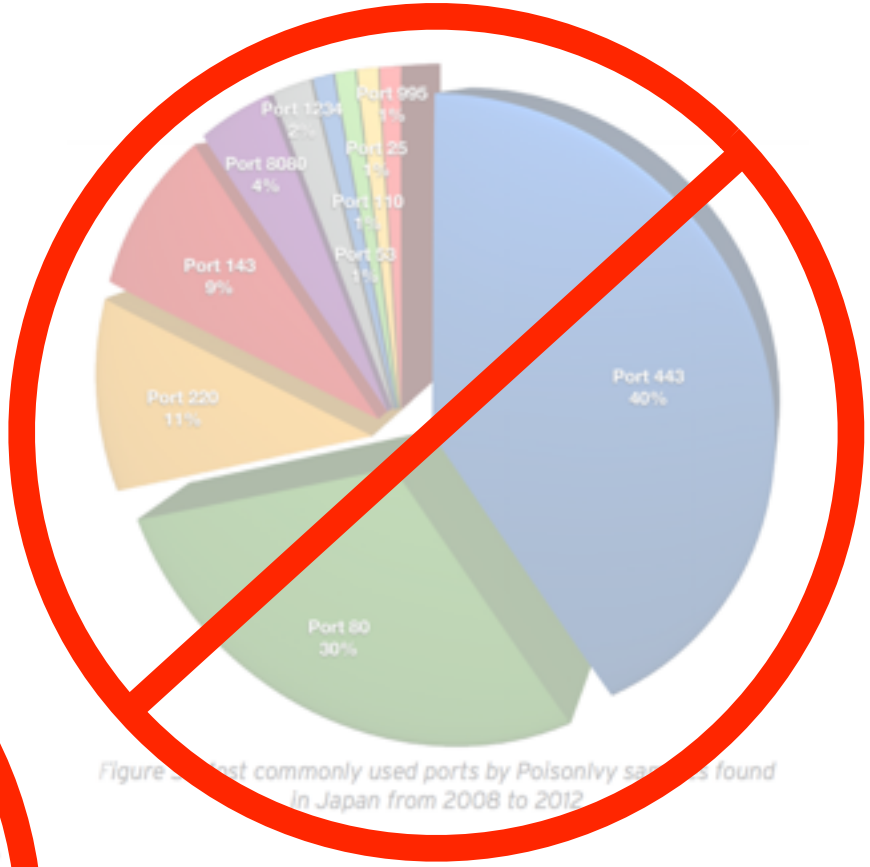


Figure 3: Most commonly used ports by Poisonivy samples found in Japan from 2008 to 2012

Visualizing for **Analysis**

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Visualizing for **Analysis**

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

All four data sets:

Mean of x: 9.0

Variance of x: 11.0

Mean of y: 7.5

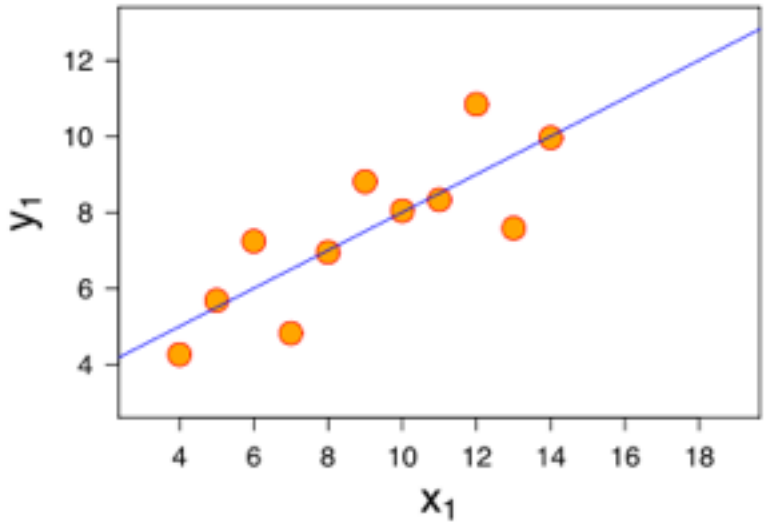
Variance of y: 4.1

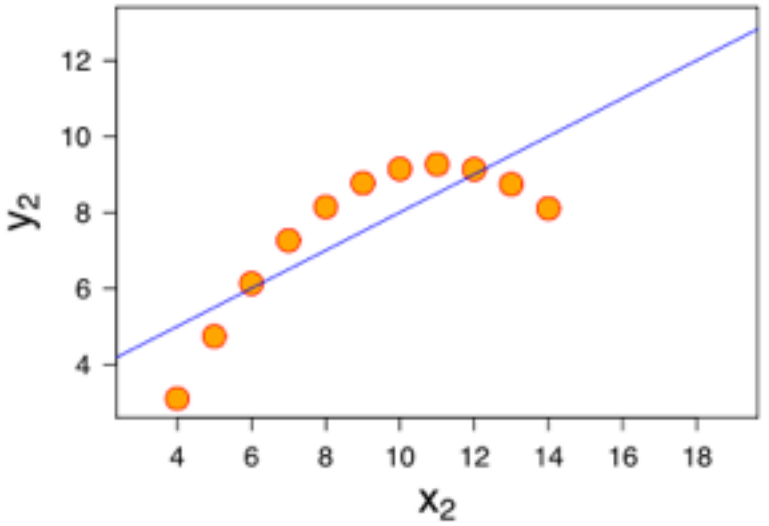
Correlation x,y: 0.816

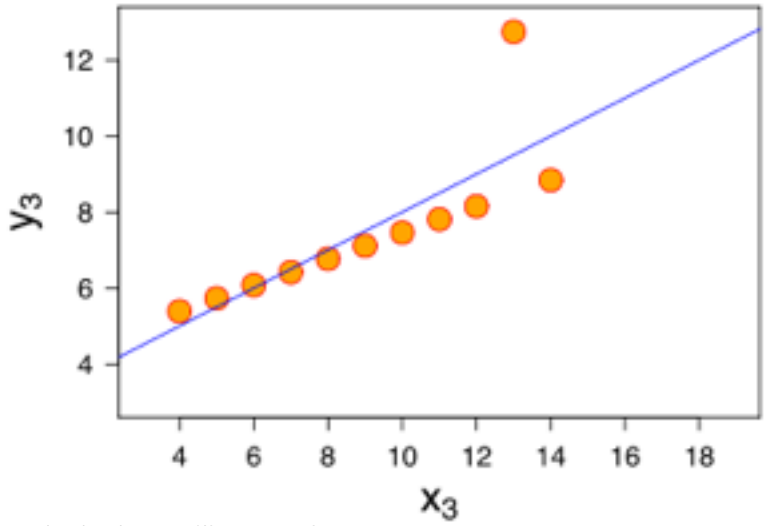
Linear Regression:

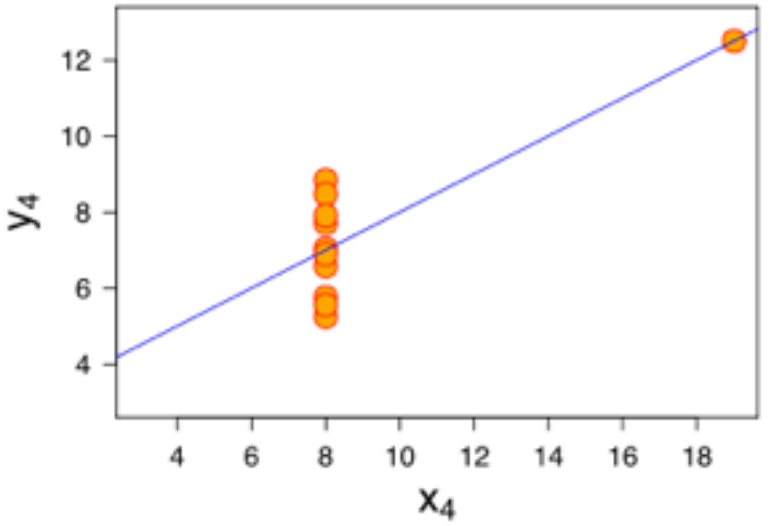
$$y = 3 + 5x$$

Visualized...

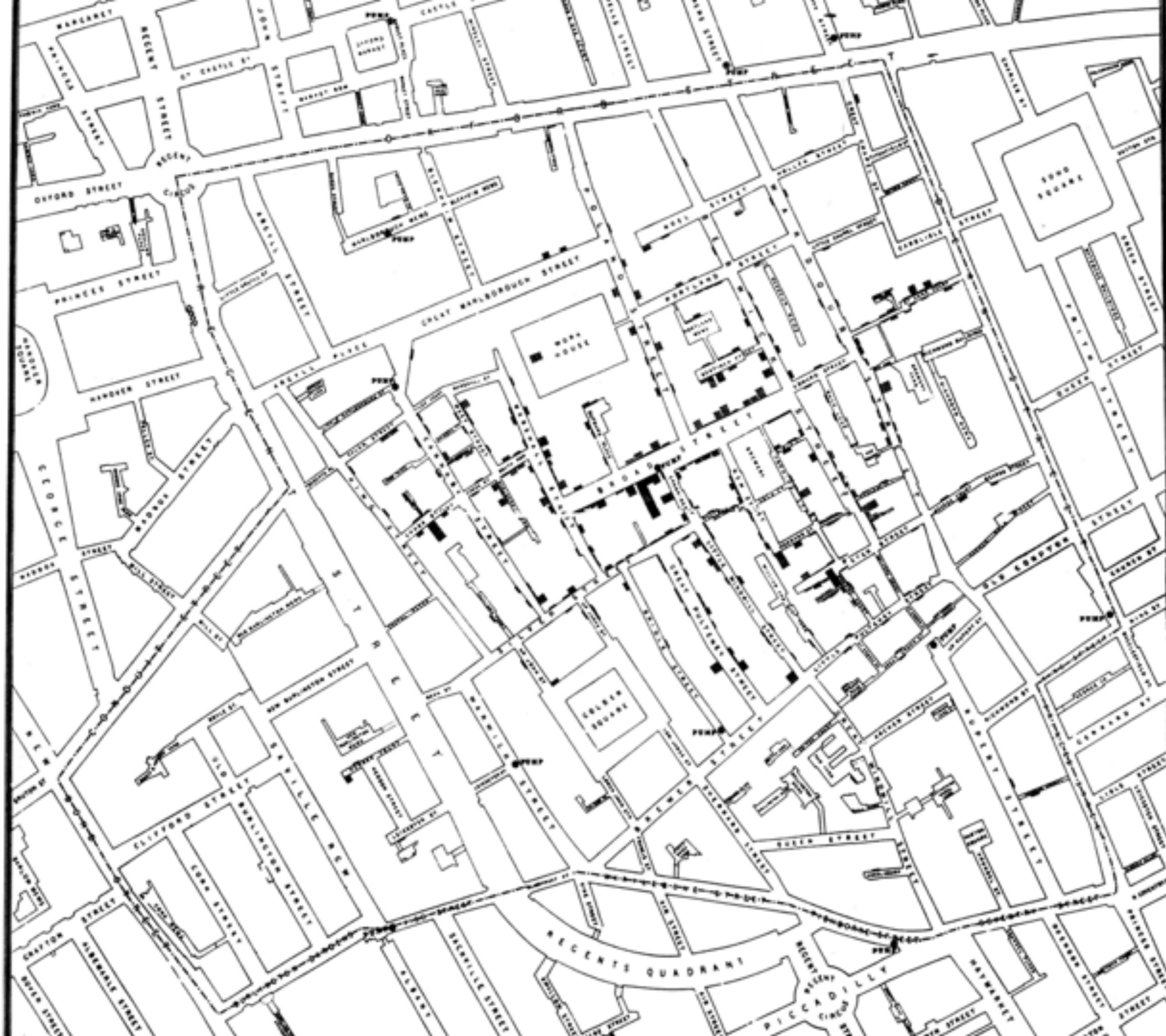








http://en.wikipedia.org/wiki/File:Anscombe%27s_quartet_3.svg

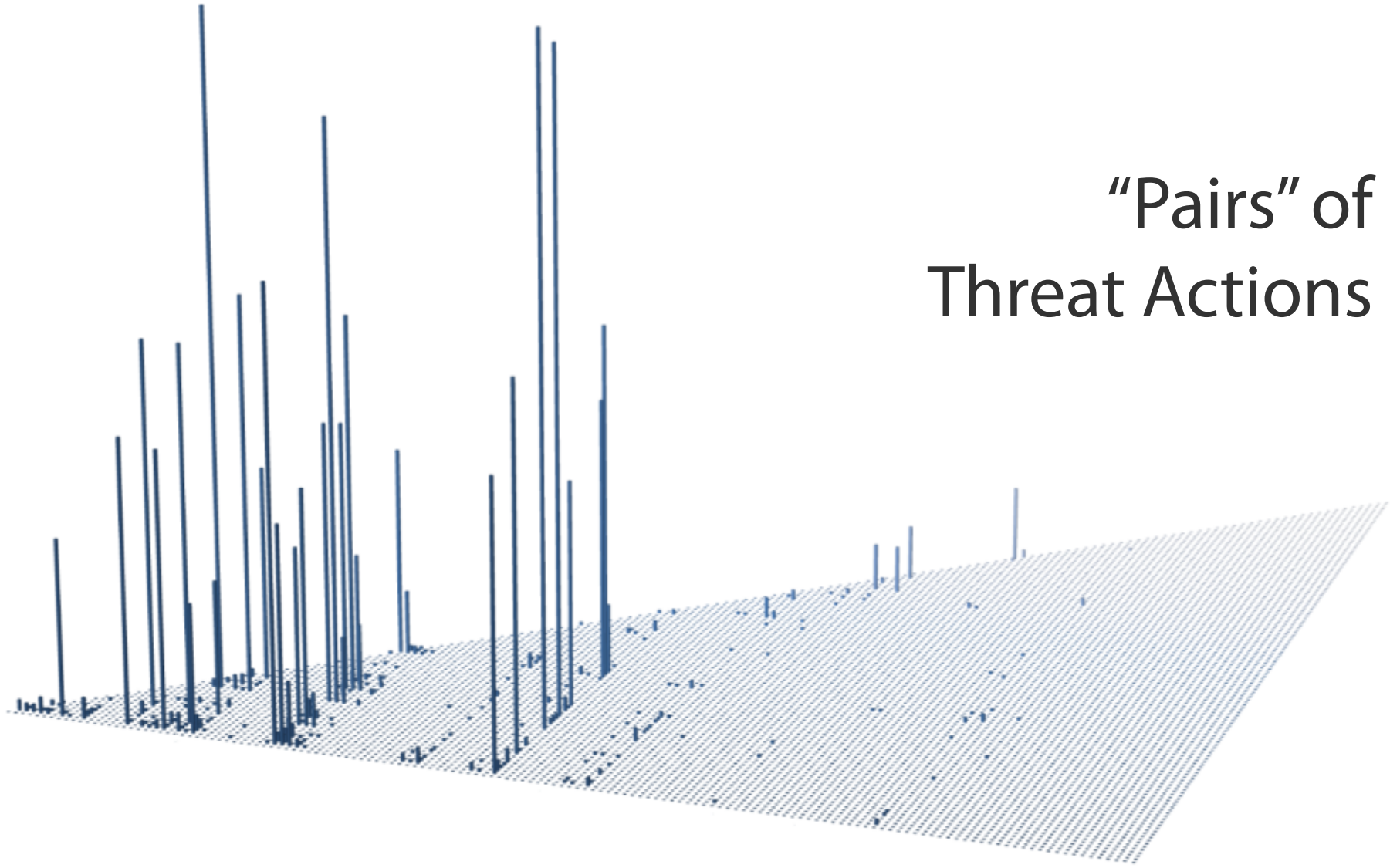


OXFORD STREET
PRINCES STREET
HANOVER STREET
HANOVER SQUARE
GREAT MARLBOROUGH STREET
ARGYLL PLACE
WORK HOUSE
BROAD STREET
PORTLAND STREET
SOHO SQUARE
FRITH STREET
OLD COMPTON
GOLDEN SQUARE
GEORGE STREET
CLIFFORD STREET
SAVILLE ROW
REGENTS QUADRANT
PICCADILLY
REGENT CIRCUS

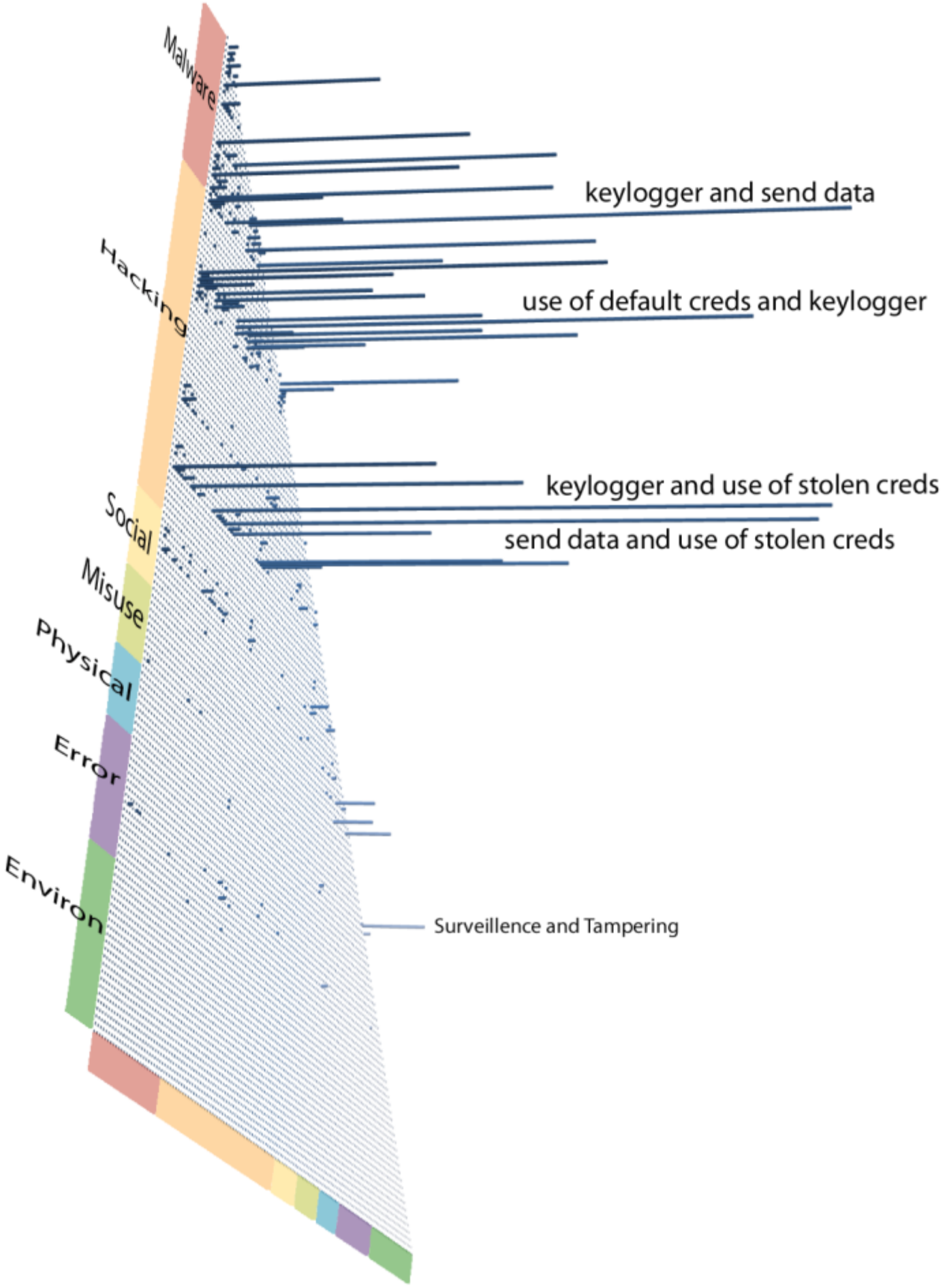
http://www.r-bloggers.com/john-snow%E2%80%99s-famous-cholera-analysis-data-in-modern-gis-formats/


Solutions are more from thinking than buying

Visualizing for **Analysis**:



"Pairs" of
Threat Actions


Malware
Hacking
Social
Misuse
Physical
Error
Environ
keylogger and send data
use of default creds and keylogger
keylogger and use of stolen creds
send data and use of stolen creds
Surveillence and Tampering

Visualizing to **Communicate**:

The night before Hurricane Sandy...

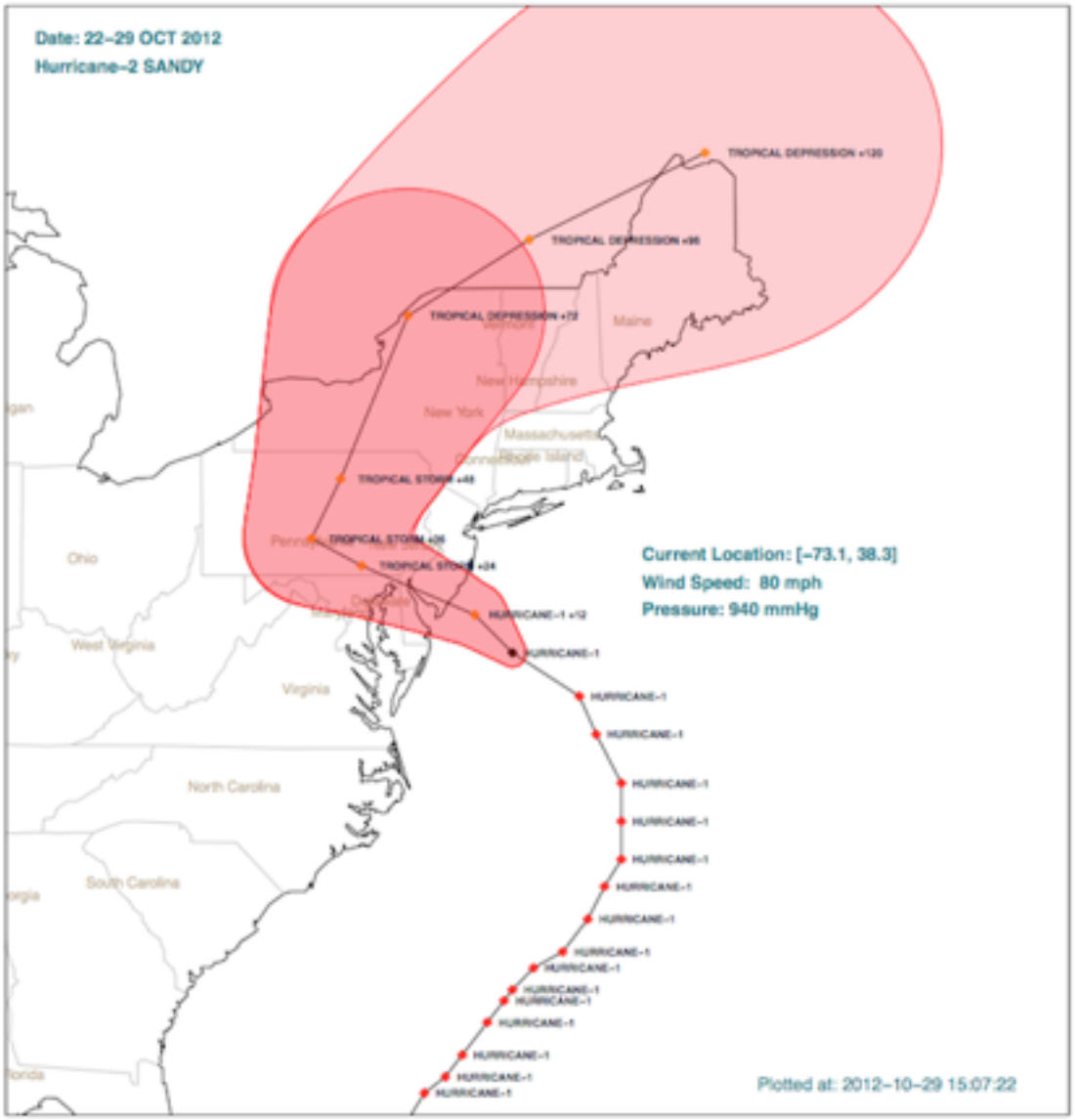
Date: 22-29 OCT 2012
Hurricane-2 SANDY

ADV	LAT	LON	TIME	WIND	PR	STAT
1	13.50	-78.00	10/22/15Z	25	1003	TROPICAL DEPRESSION
1A	13.50	-78.50	10/22/18Z	25	1003	TROPICAL DEPRESSION
2	12.50	-78.50	10/22/21Z	35	999	TROPICAL STORM
2A	12.70	-78.70	10/23/00Z	40	998	TROPICAL STORM
3	12.70	-78.60	10/23/03Z	40	998	TROPICAL STORM
3A	12.90	-78.70	10/23/06Z	40	998	TROPICAL STORM
4	13.30	-78.60	10/23/09Z	40	998	TROPICAL STORM
4A	13.40	-77.90	10/23/12Z	40	997	TROPICAL STORM
5	13.80	-77.80	10/23/15Z	45	993	TROPICAL STORM
5A	14.10	-77.60	10/23/18Z	45	993	TROPICAL STORM
6	14.30	-77.60	10/23/21Z	45	993	TROPICAL STORM
6A	14.80	-77.50	10/24/00Z	45	993	TROPICAL STORM
7	15.20	-77.20	10/24/03Z	50	989	TROPICAL STORM
7A	15.70	-77.10	10/24/06Z	55	988	TROPICAL STORM
8	16.30	-77.00	10/24/09Z	60	986	TROPICAL STORM
8A	16.60	-76.90	10/24/12Z	60	983	TROPICAL STORM
9	17.10	-76.70	10/24/15Z	70	973	HURRICANE-1
9A	17.60	-76.80	10/24/18Z	70	973	HURRICANE-1
10	18.30	-76.60	10/24/21Z	70	970	HURRICANE-1
10A	18.70	-76.40	10/25/00Z	75	968	HURRICANE-1
11	19.40	-76.30	10/25/03Z	80	954	HURRICANE-1
11A	20.10	-75.90	10/25/06Z	95	957	HURRICANE-2

. . .
. . .

Visualizing to **Communicate**:

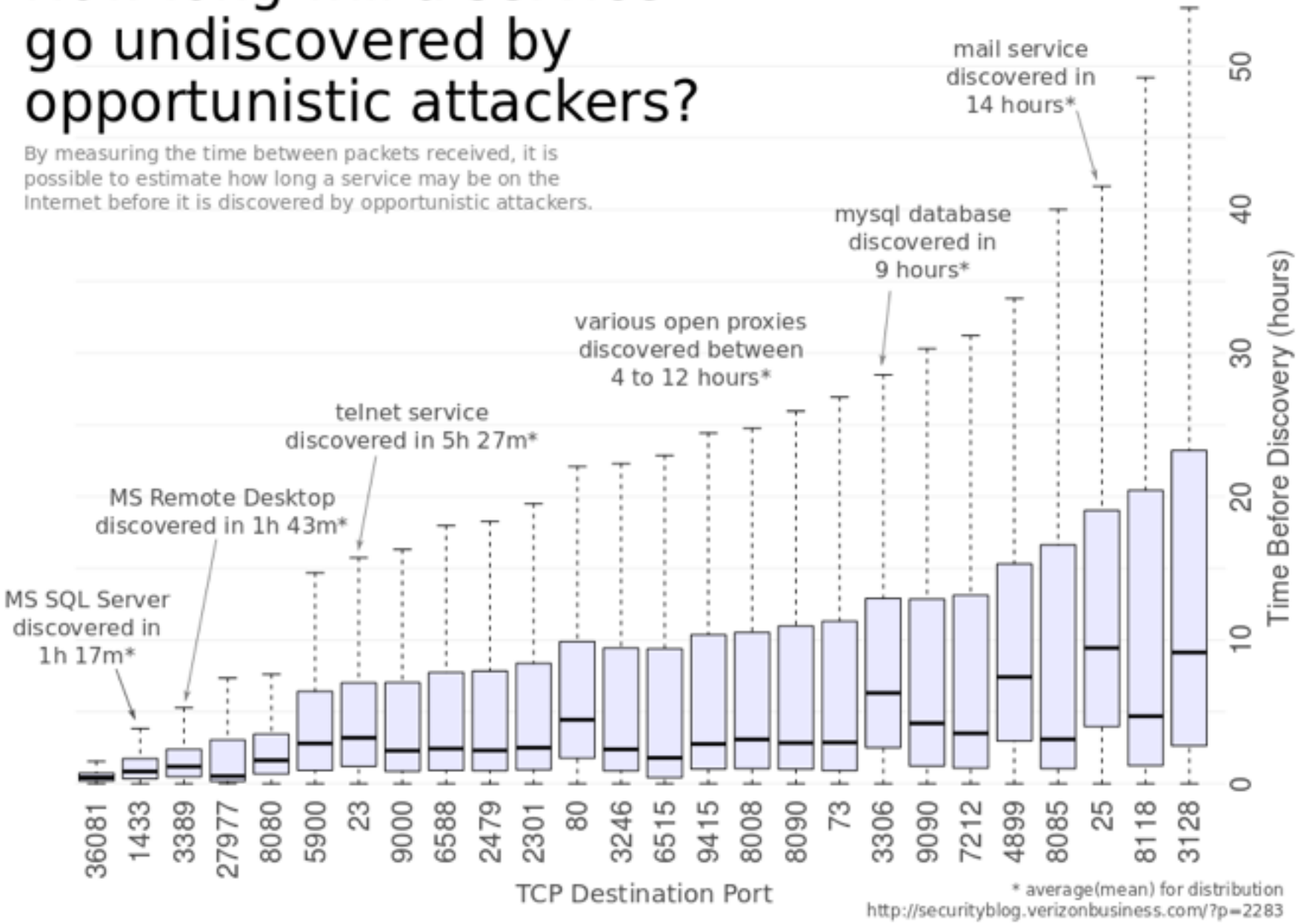
The night before Hurricane Sandy...

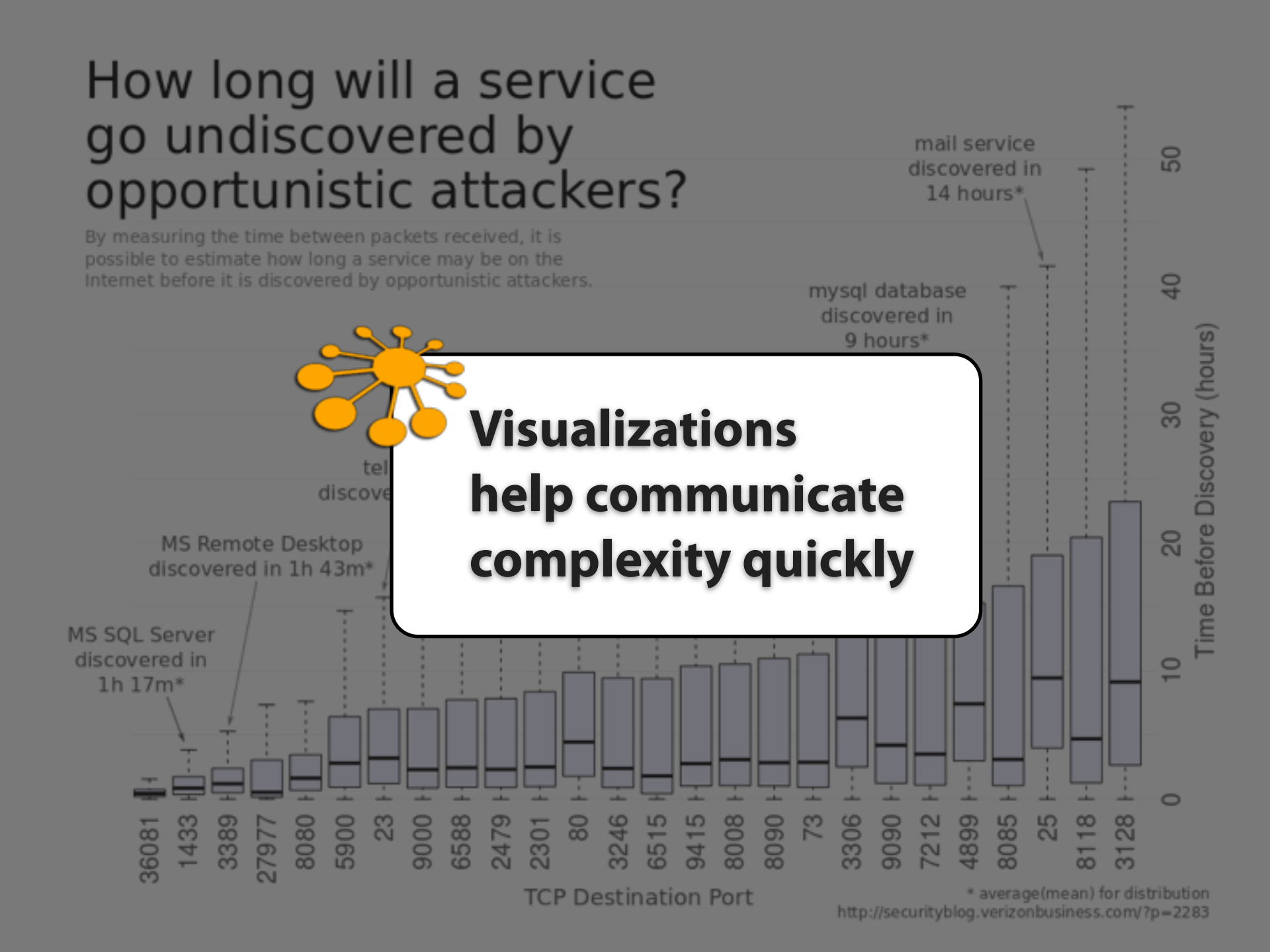



...and three
months later
Mapping a Storm
Mapping a Storm

How long will a service go undiscovered by opportunistic attackers?

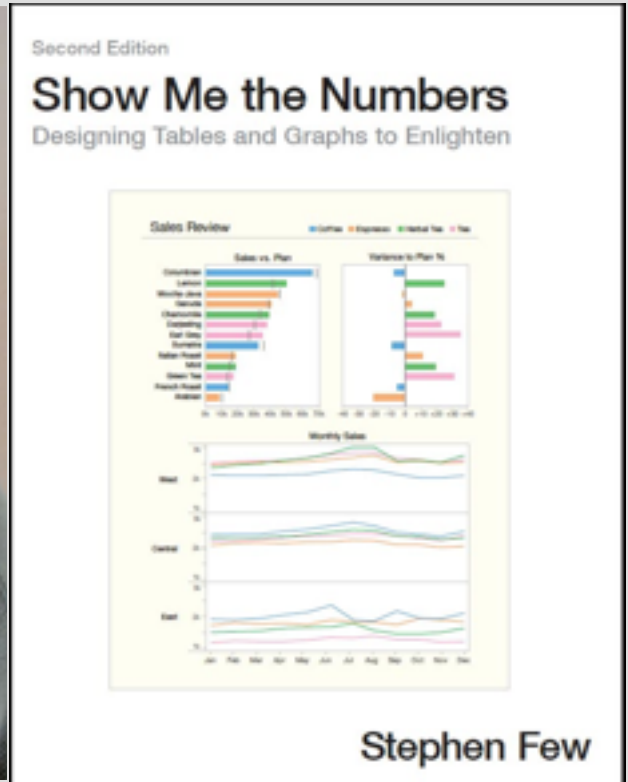
By measuring the time between packets received, it is possible to estimate how long a service may be on the Internet before it is discovered by opportunistic attackers.



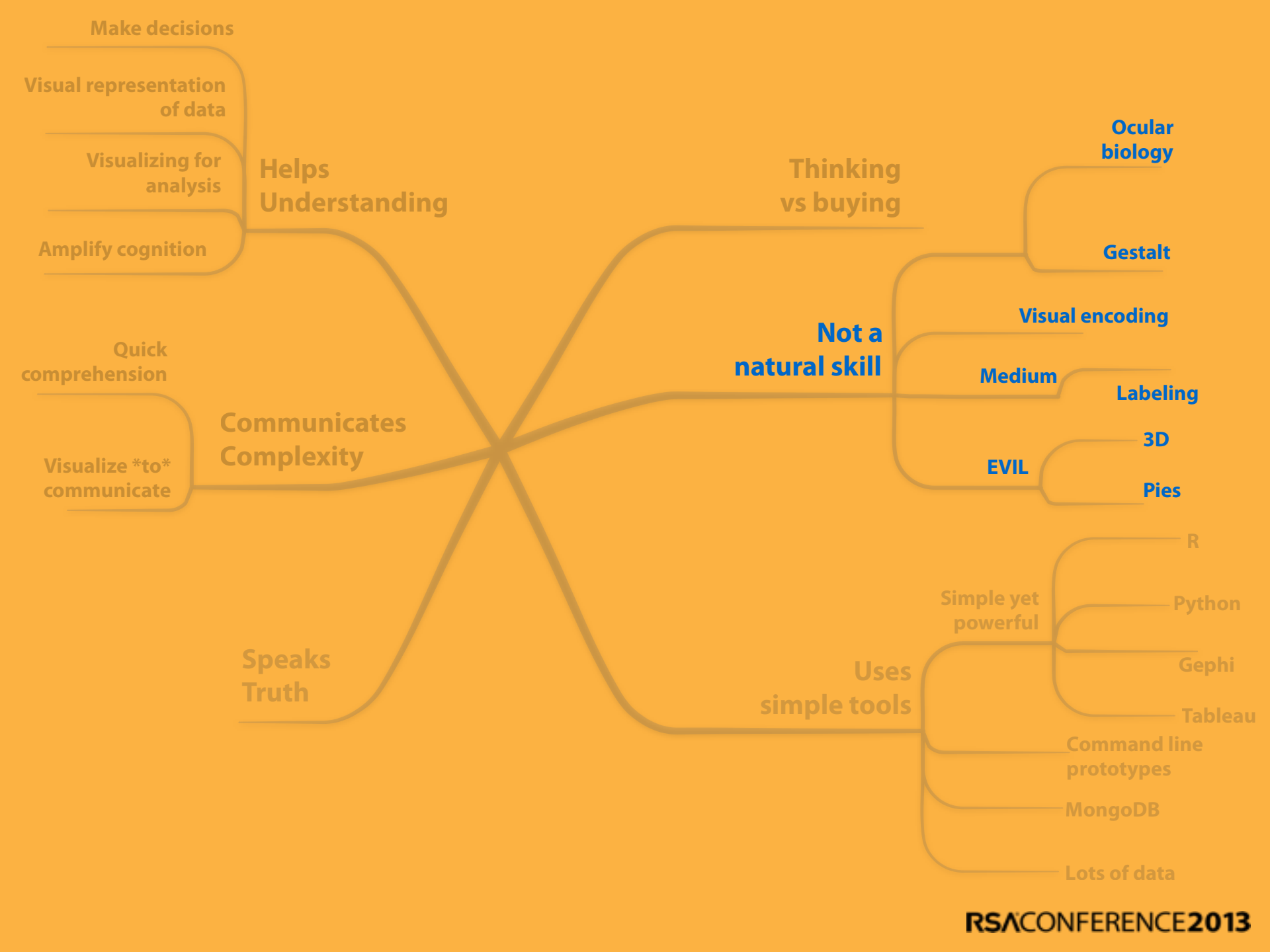

How long will a service go undiscovered by opportunistic attackers?
By measuring the time between packets received, it is possible to estimate how long a service may be on the Internet before it is discovered by opportunistic attackers.
mail service discovered in 14 hours*
mysql database discovered in 9 hours*
MS Remote Desktop discovered in 1h 43m*
MS SQL Server discovered in 1h 17m*
Visualizations help communicate complexity quickly
Time Before Discovery (hours)
0
10
20
30
40
50
36081
1433
3389
27977
8080
5900
23
9000
6588
2479
2301
80
3246
6515
9415
8008
8090
73
3306
9090
7212
4899
8085
25
8118
3128
TCP Destination Port
* average(mean) for distribution
http://securityblog.verizonbusiness.com/?p=2283

"[*Tables and graphs*] are so common many of us assume that knowledge of their effective use is common as well.
I assure you, it is not."

Stephen Few
Show Me the Numbers: Designing Tables and Graphs to Enlighten




2005
2006
2007
2008
2009
2010
2011
2012
2013
iformia Attorney Gen
Databreaches.net
Dataloss DB
S via Databreaches.
HS via PHIPrivacy.n
Media
NAID
PHIPrivacy.net
ecurity Breach Lette
count
Month
Type
CARD
DISC
HACK
INSD
PHYS
PORT
STAT
UNKN


Make decisions
Visual representation of data
Visualizing for analysis
Amplify cognition
Helps Understanding
Quick comprehension
Visualize *to* communicate
Communicates Complexity
Speaks Truth
Thinking vs buying
Ocular biology
Gestalt
Not a natural skill
Visual encoding
Medium
Labeling
EVIL
3D
Pies
Uses simple tools
Simple yet powerful
R
Python
Gephi
Tableau
Command line prototypes
MongoDB
Lots of data

Visualizing

Encoding

with shape, size, color and position

using categorical or quantitative variables

possibly over space or time

Accuracy of Decoding

More

Position along a common scale

Position on identical but nonaligned scales

Length

Angle / Slope

Area

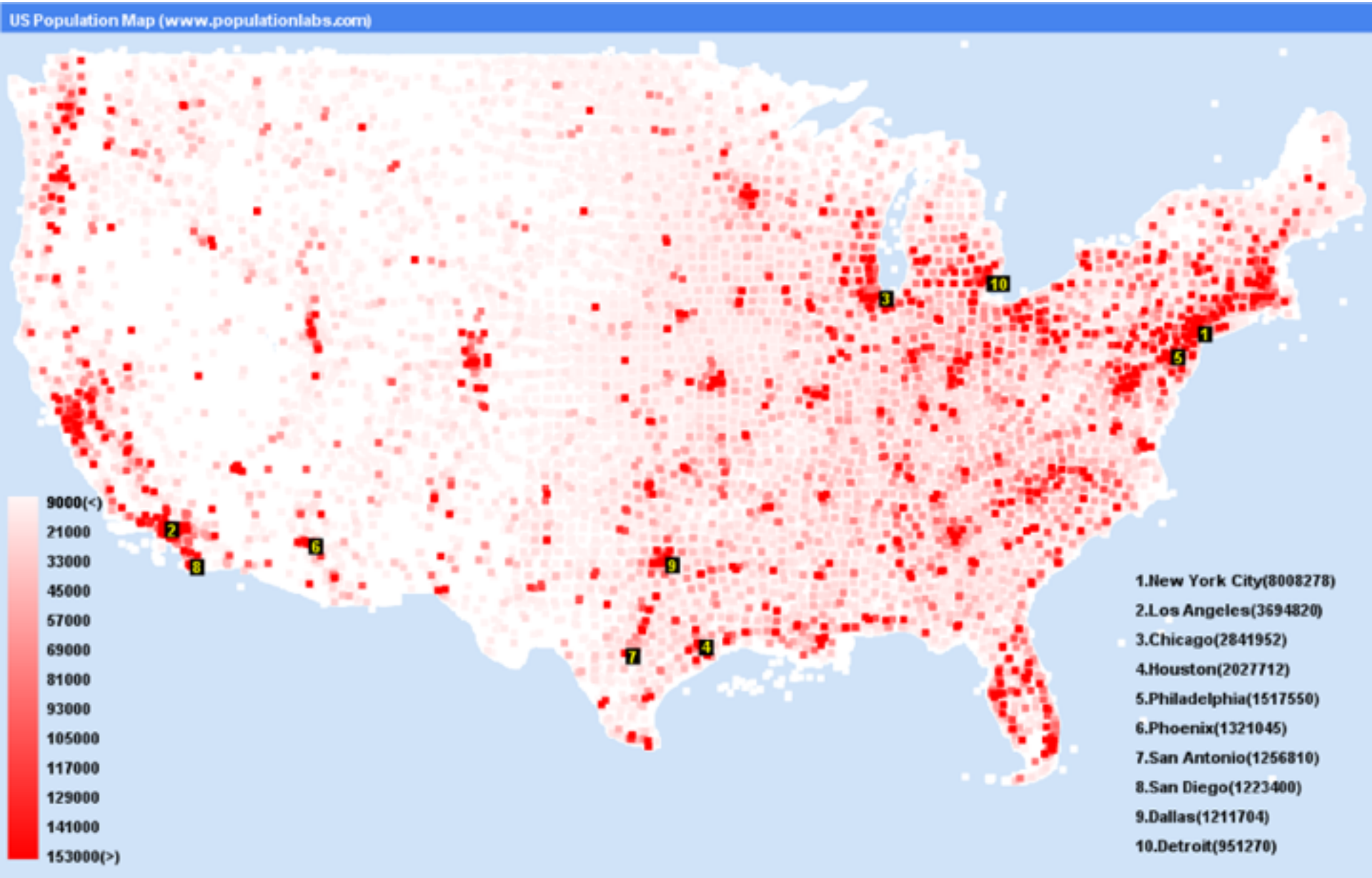
Volume / Density / Saturation

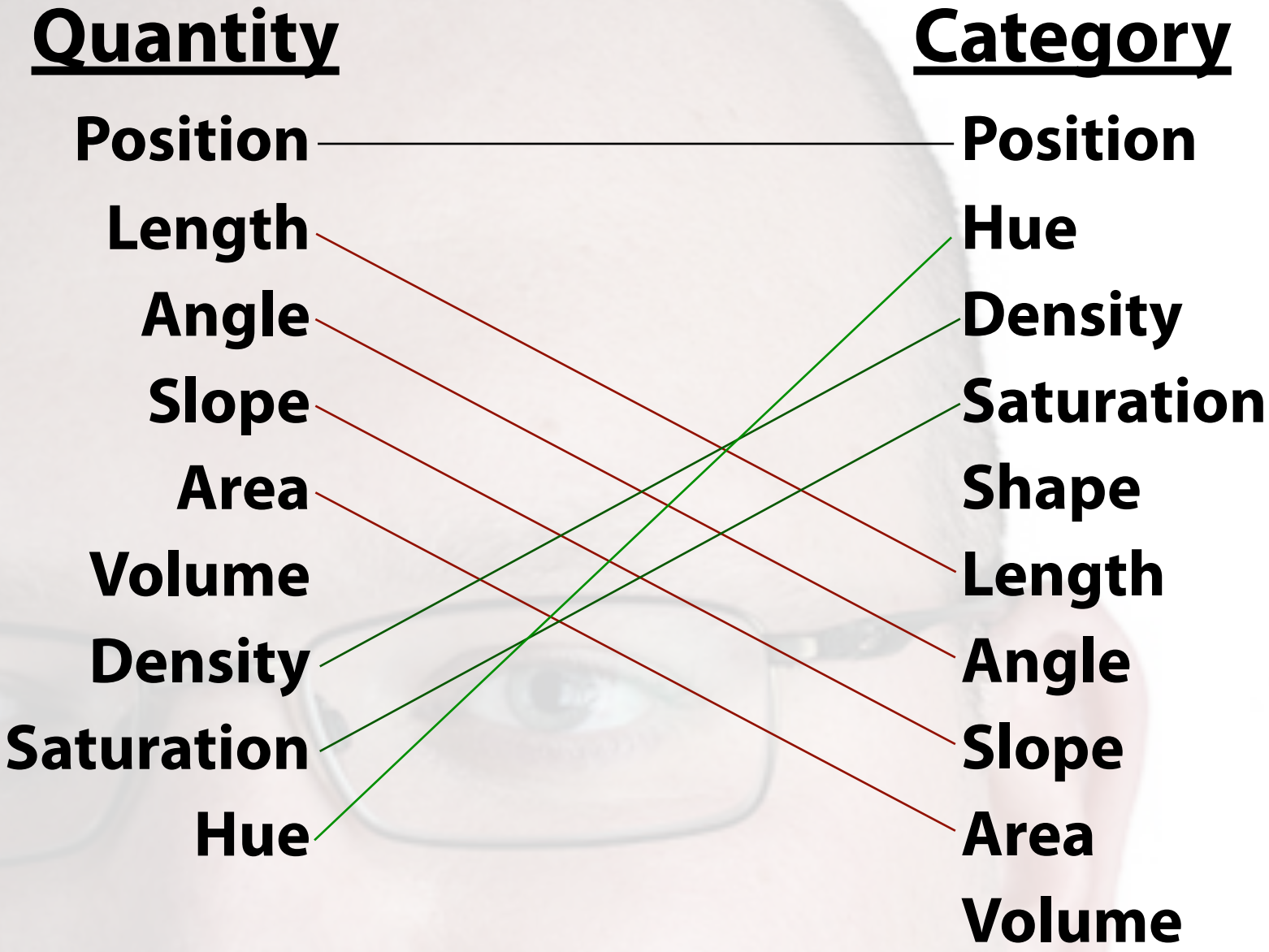
Hue

Less

"Graphical perceptions and Graphical Methods for Analyzing Scientific Data", Cleveland and McGill, Science,
Photo by Kevin Riggins, https://plus.google.com/u/0/photos/115846783938665223975/albums/5695093594342187601/5697948193623743506

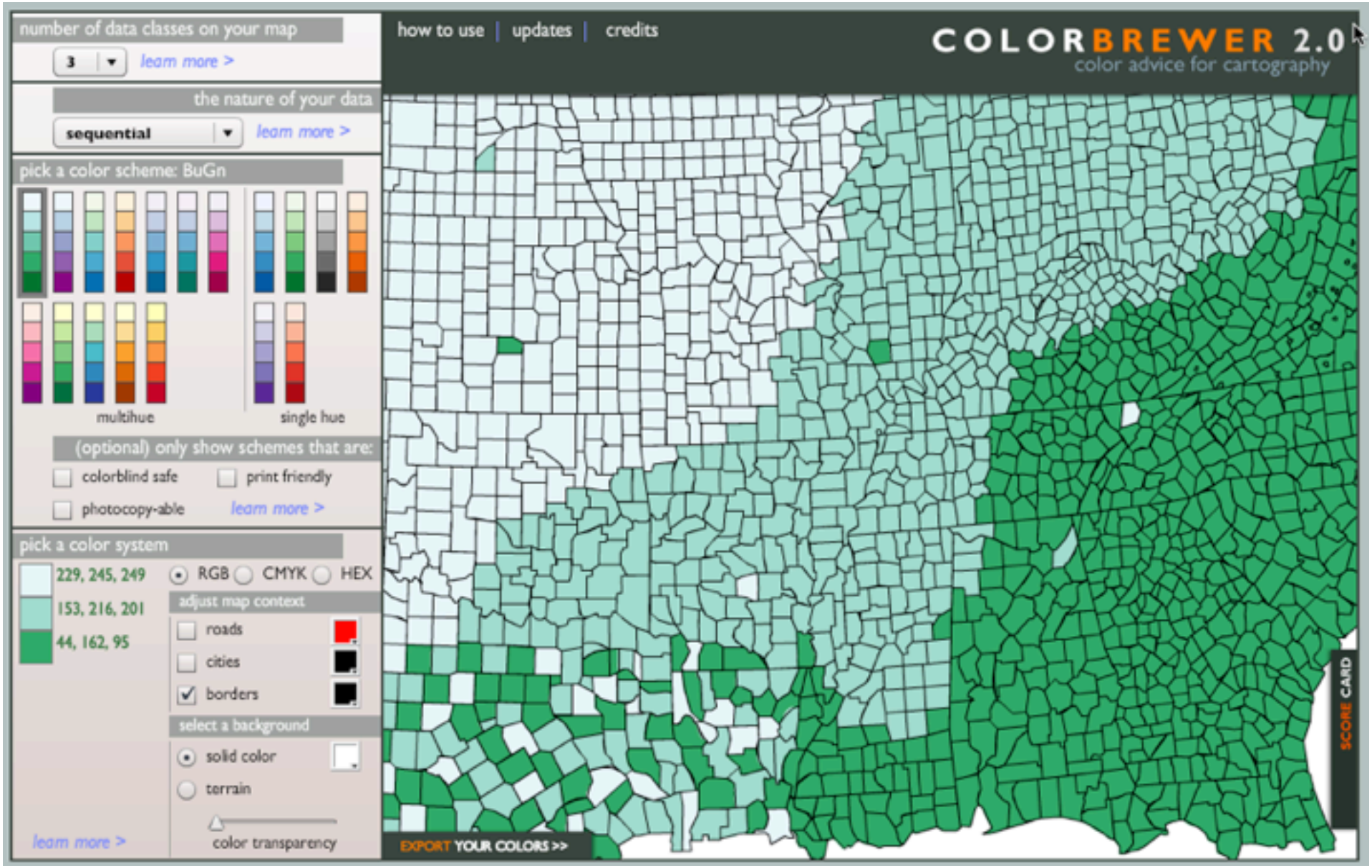
Communicate Quantity with Saturation?

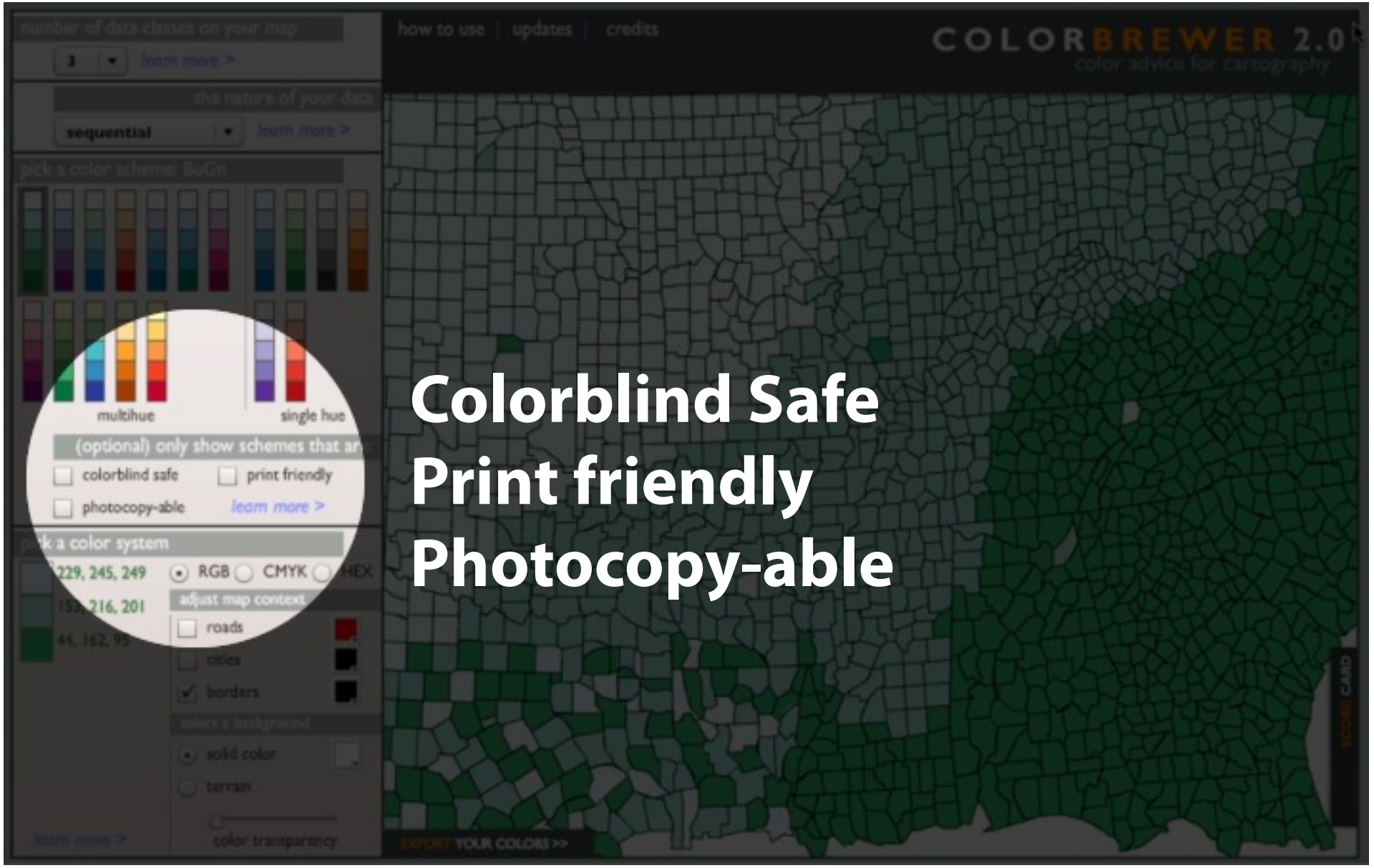




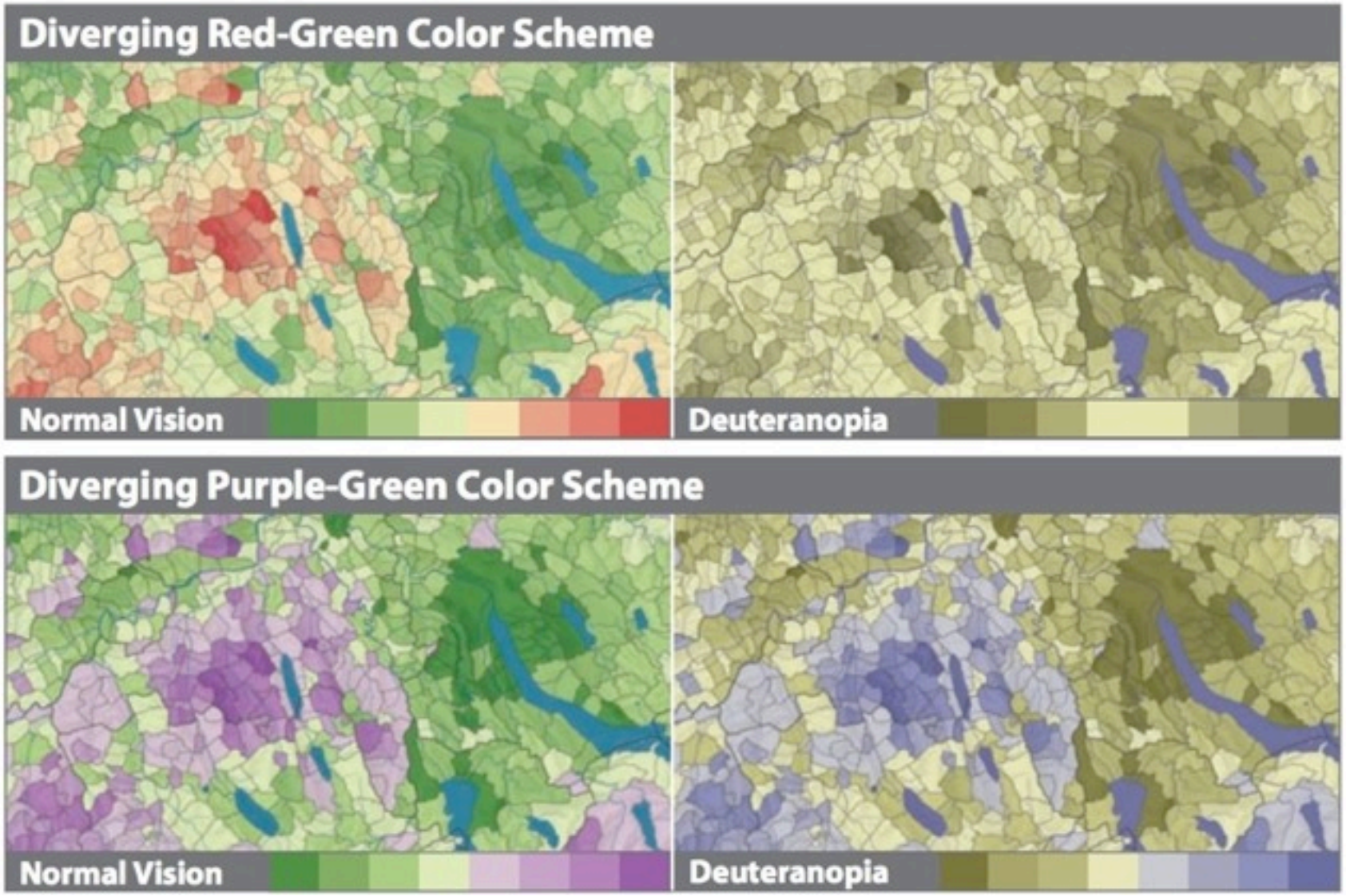
From: http://assassin.cs.rpi.edu/~cutler/classes/visualization/F10/papers/p110-mackinlay.pdf

Photo by Kevin Riggins, http://www.flickr.com/photos/krandj/7688630288/in/set-72157630847690018/


number of data classes on your map
3
learn more >
the nature of your data
sequential
learn more >
pick a color scheme: BuGn
multihue
single hue
(optional) only show schemes that are:
colorblind safe
print friendly
photocopy-able
learn more >
pick a color system
229, 245, 249
153, 216, 201
44, 162, 95
RGB
CMYK
HEX
adjust map context
roads
cities
borders
select a background
solid color
terrain
color transparency
learn more >
how to use | updates | credits
COLORBREWER 2.0
color advice for cartography
EXPORT YOUR COLORS >>
SCORE CARD


how to use
updates
credits
COLORBREWER 2.0
color advice for cartography
number of data classes on your map
learn more >
the nature of your data
sequential
learn more >
pick a color scheme: BuGn
multihue
single hue
(optional) only show schemes that ar
colorblind safe
print friendly
photocopy-able
learn more >
k a color system
229, 245, 249
RGB
CMYK
HEX
216, 201
44, 162, 95
adjust map context
roads
cities
borders
solid color
terrain
color transparency
learn more >
EXPORT YOUR COLORS >>
Colorblind Safe
Print friendly
Photocopy-able

Color blindness is common



Print Friendly…?

Pop Quiz:
What do you see?

Figure 10: Threat agents over time by percent of breaches

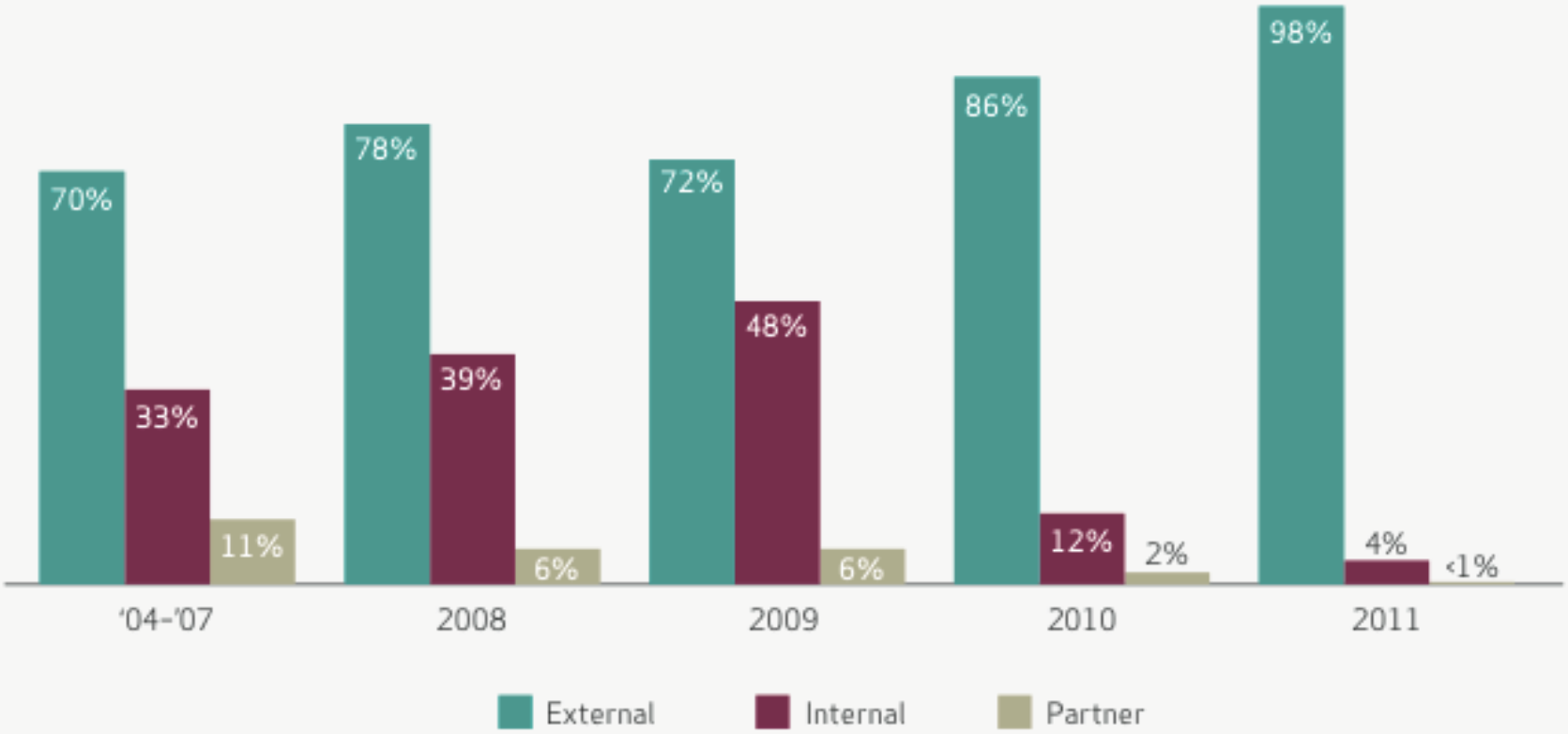
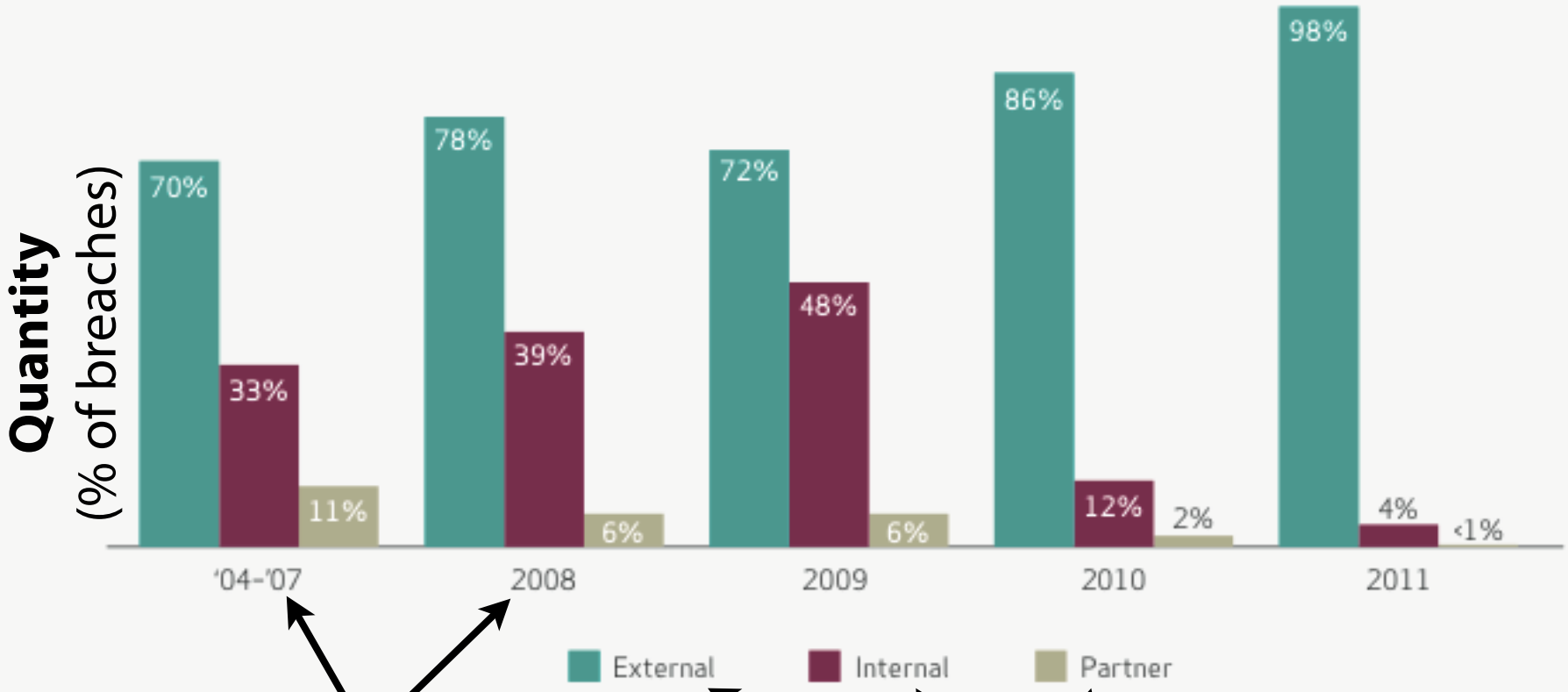


Figure 10: Threat agents over time by percent of breaches



Category (year)

Category (actor)

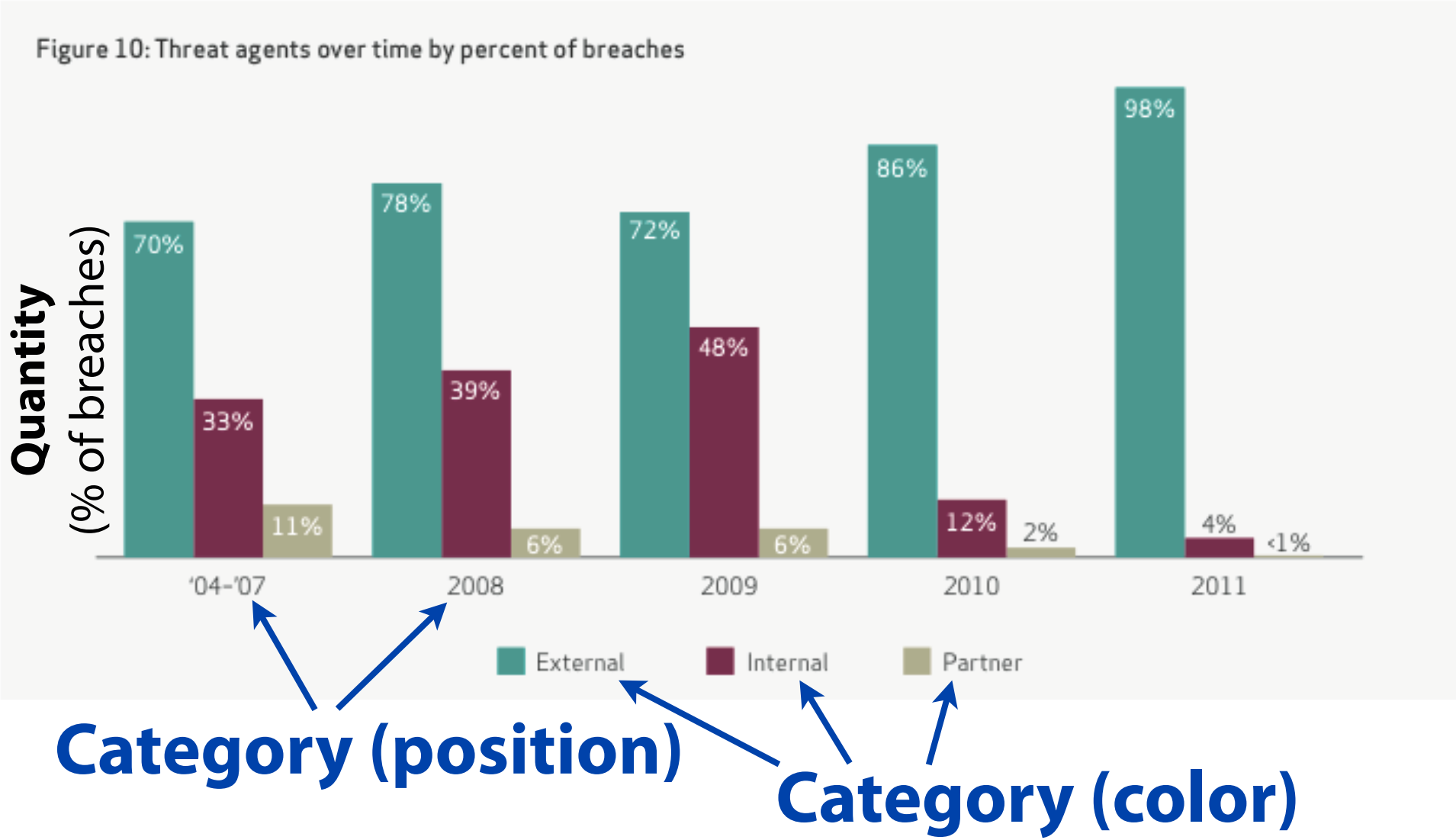


Figure 10: Threat agents over time by percent of breaches
Quantity
(% of breaches)
70%
33%
11%
78%
39%
6%
72%
48%
6%
86%
12%
2%
98%
4%
<1%
'04-'07
2008
2009
2010
2011
External
Internal
Partner
Category (position)
Category (color)

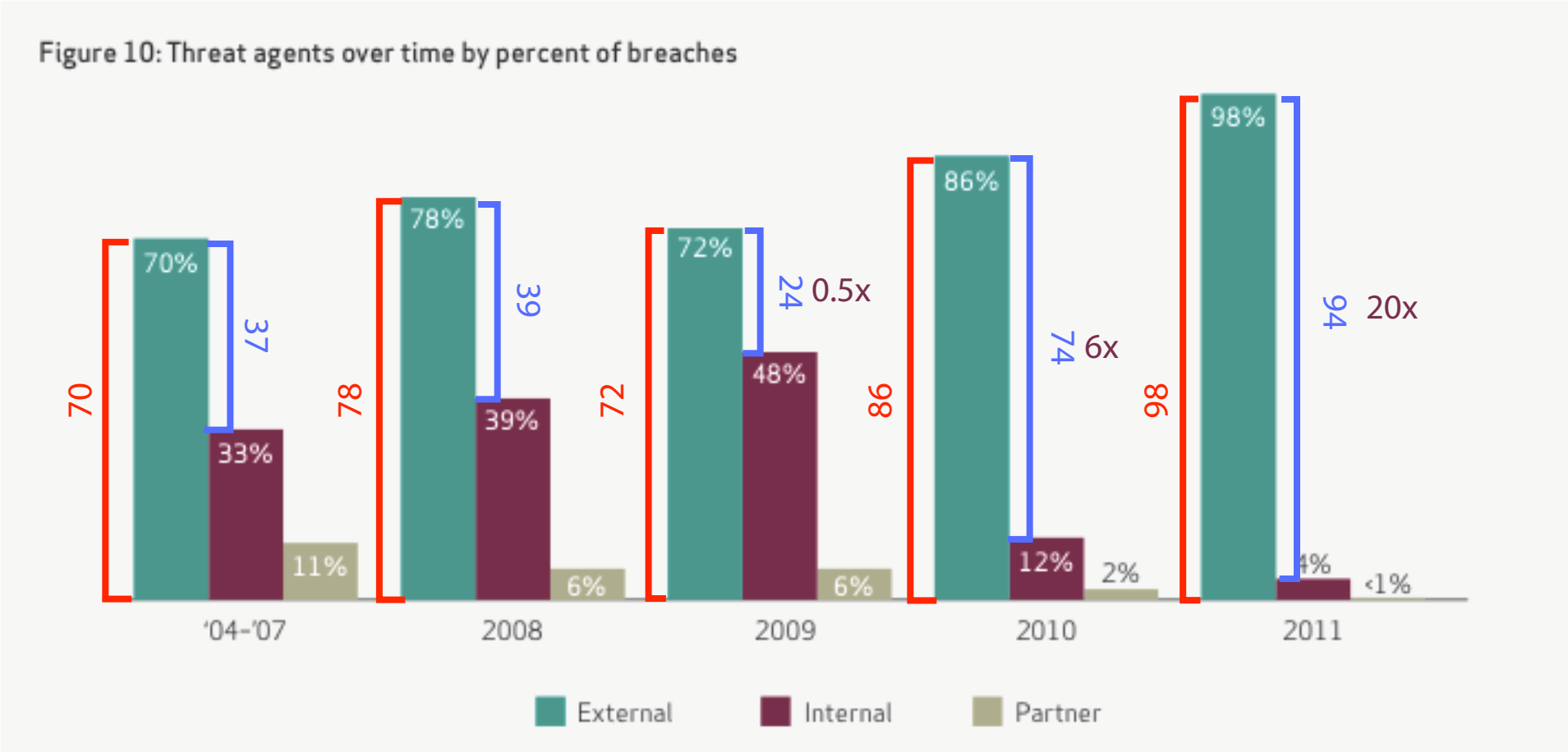
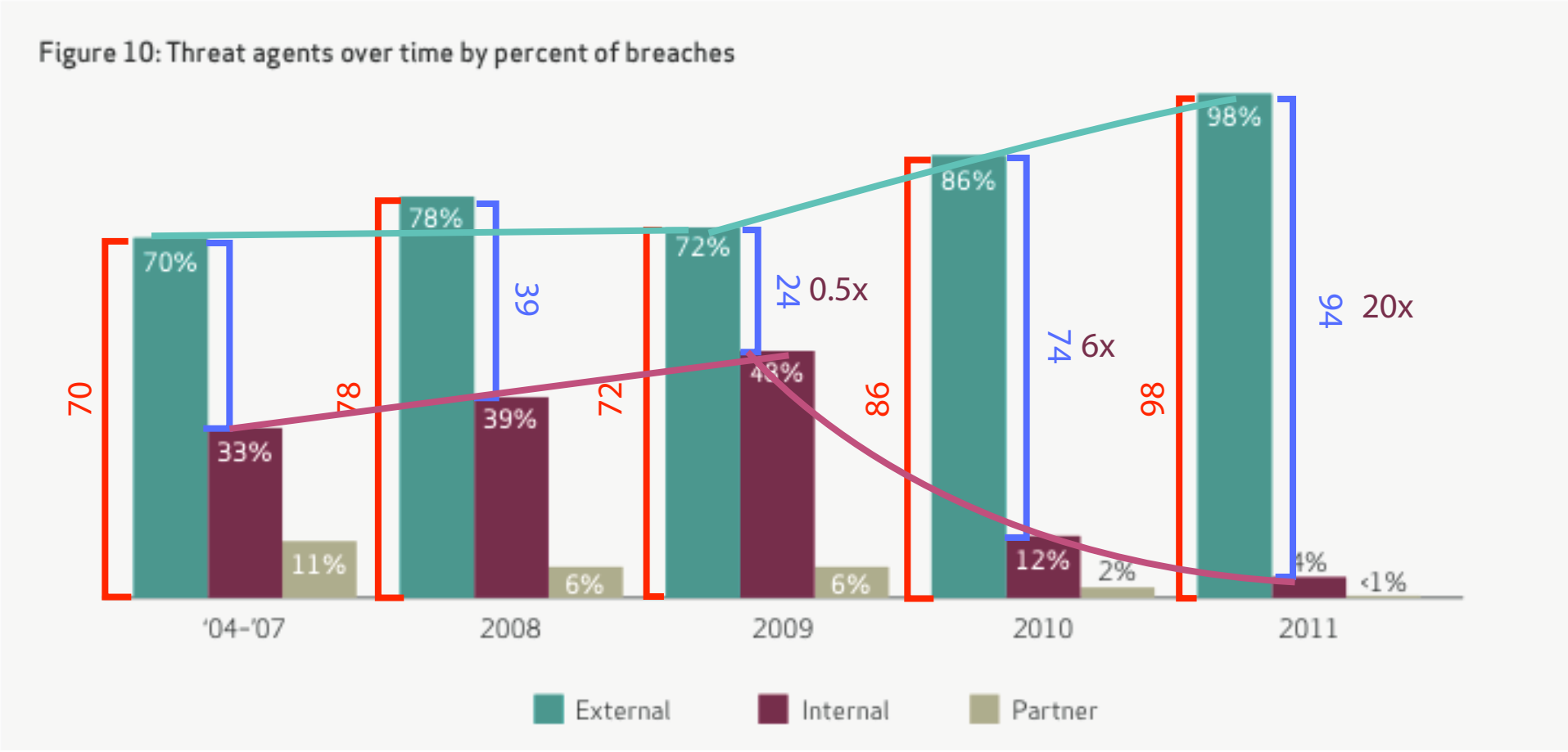


Figure 10: Threat agents over time by percent of breaches
70
70%
37
33%
11%
78
78%
39
39%
6%
72
72%
24
0.5x
48%
6%
98
86%
74
6x
12%
2%
86
98%
94
20x
4%
<1%
'04–'07
2008
2009
2010
2011
External
Internal
Partner

...length



…Position on common scale

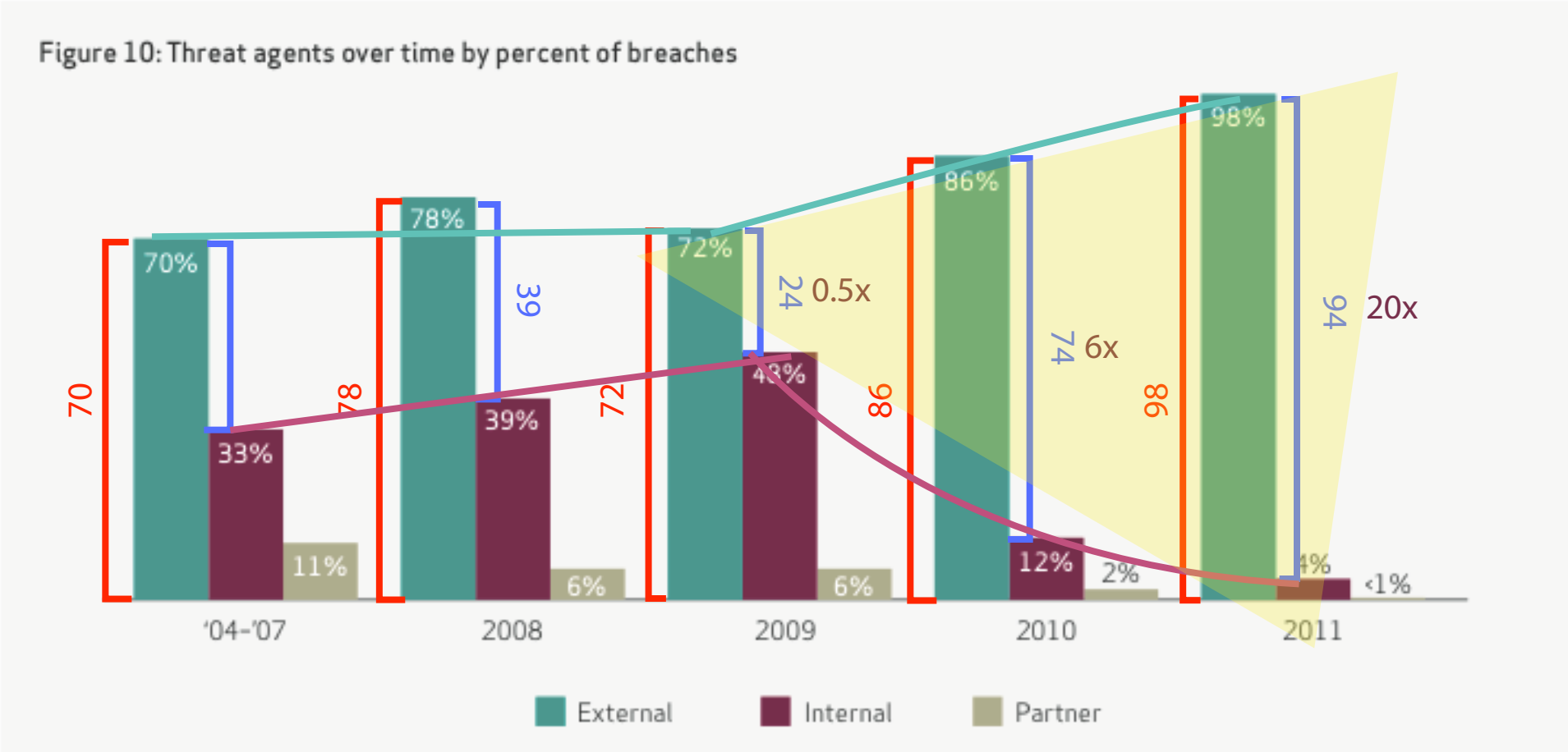
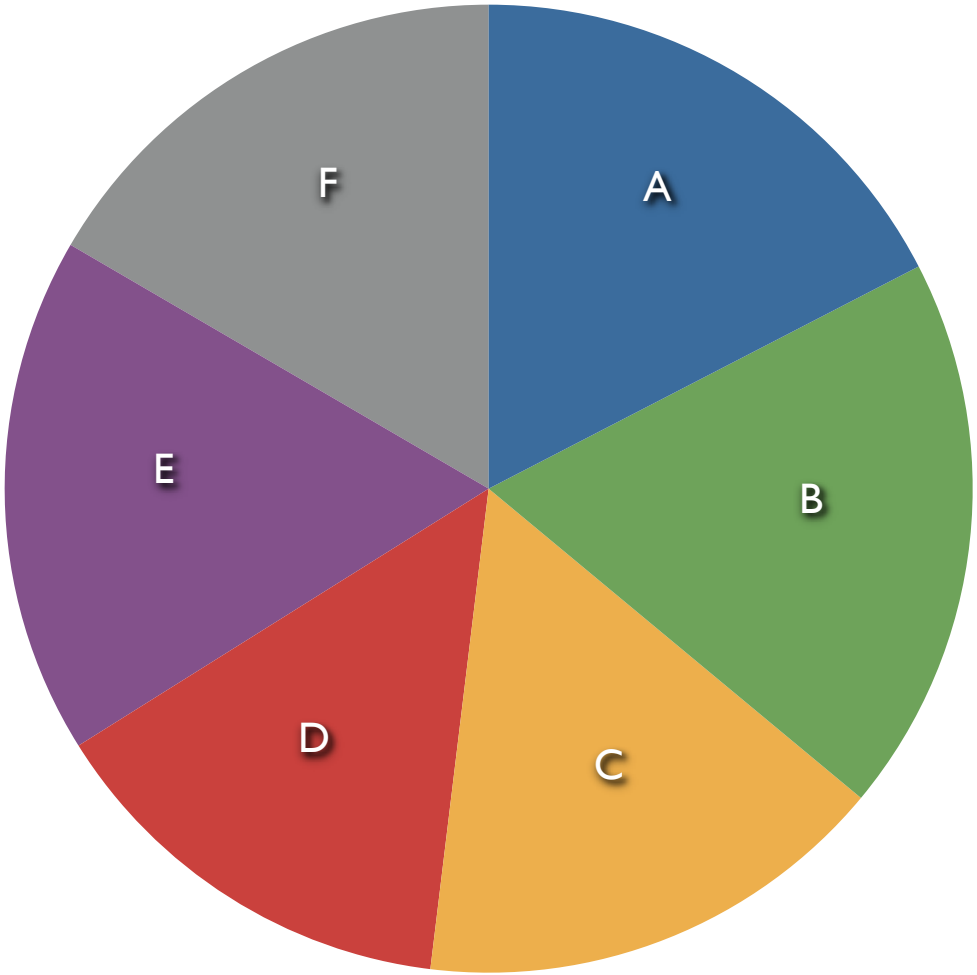


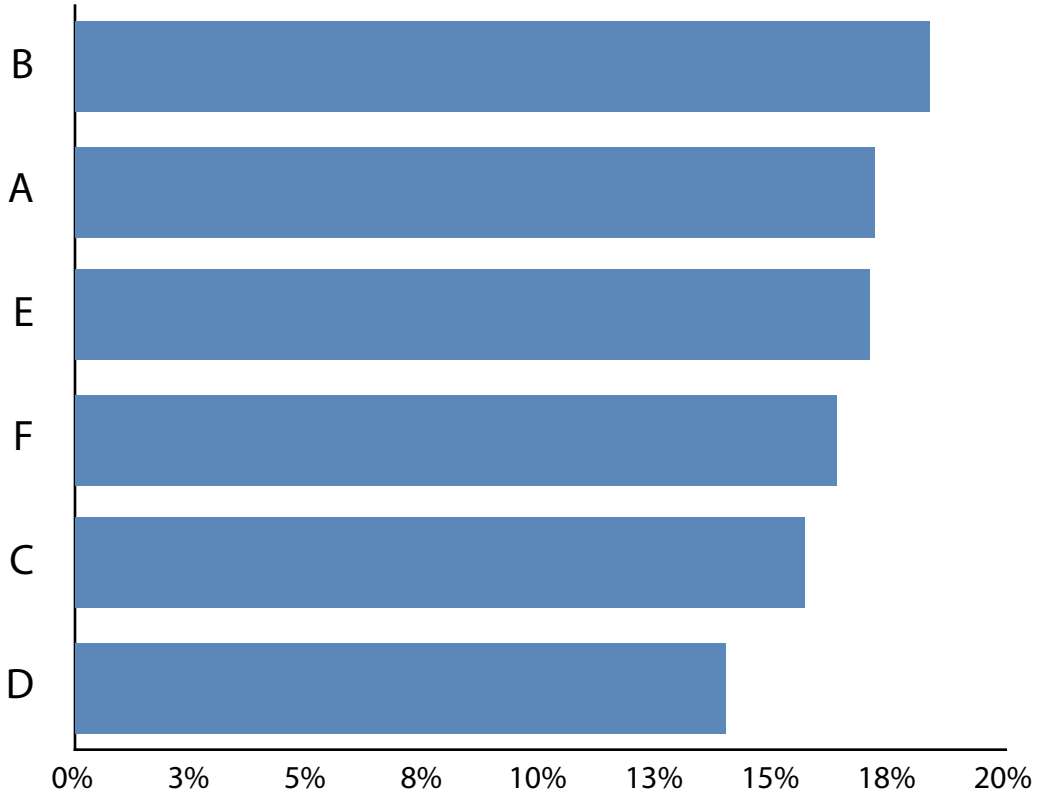
Figure 10: Threat agents over time by percent of breaches
70%
33%
11%
78%
39%
6%
72%
43%
6%
86%
12%
2%
98%
4%
<1%
70
78
39
72
24
0.5x
98
74
6x
86
94
20x
'04–'07
2008
2009
2010
2011
External
Internal
Partner

…patterns!

Pop Quiz #2:
Which is larger?


A
B
C
D
E
F

Which is Larger?



Position and Length makes the same values easier to compare


Pie I have eaten
Pie I have not yet eaten
Hasnamat.com

Caution!

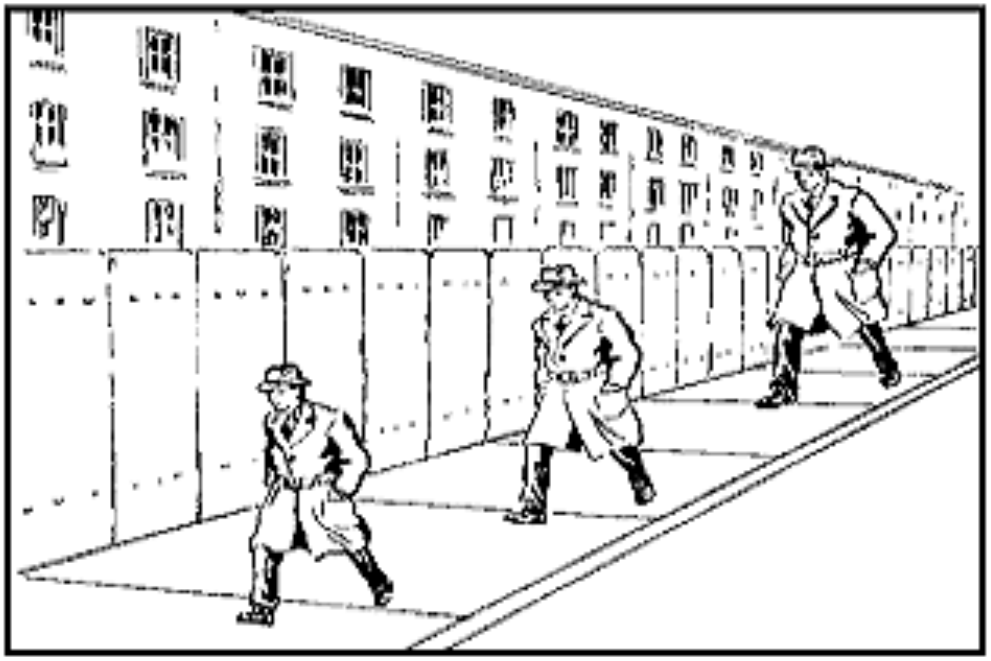
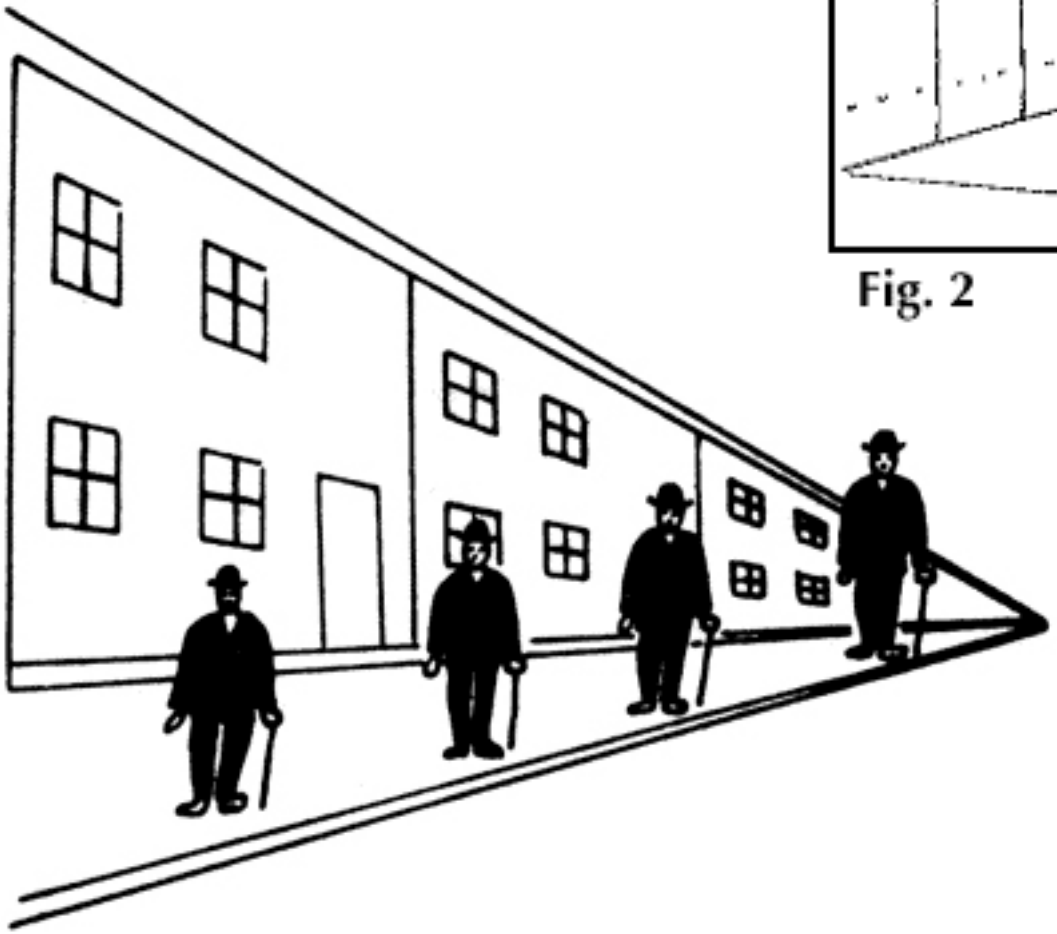
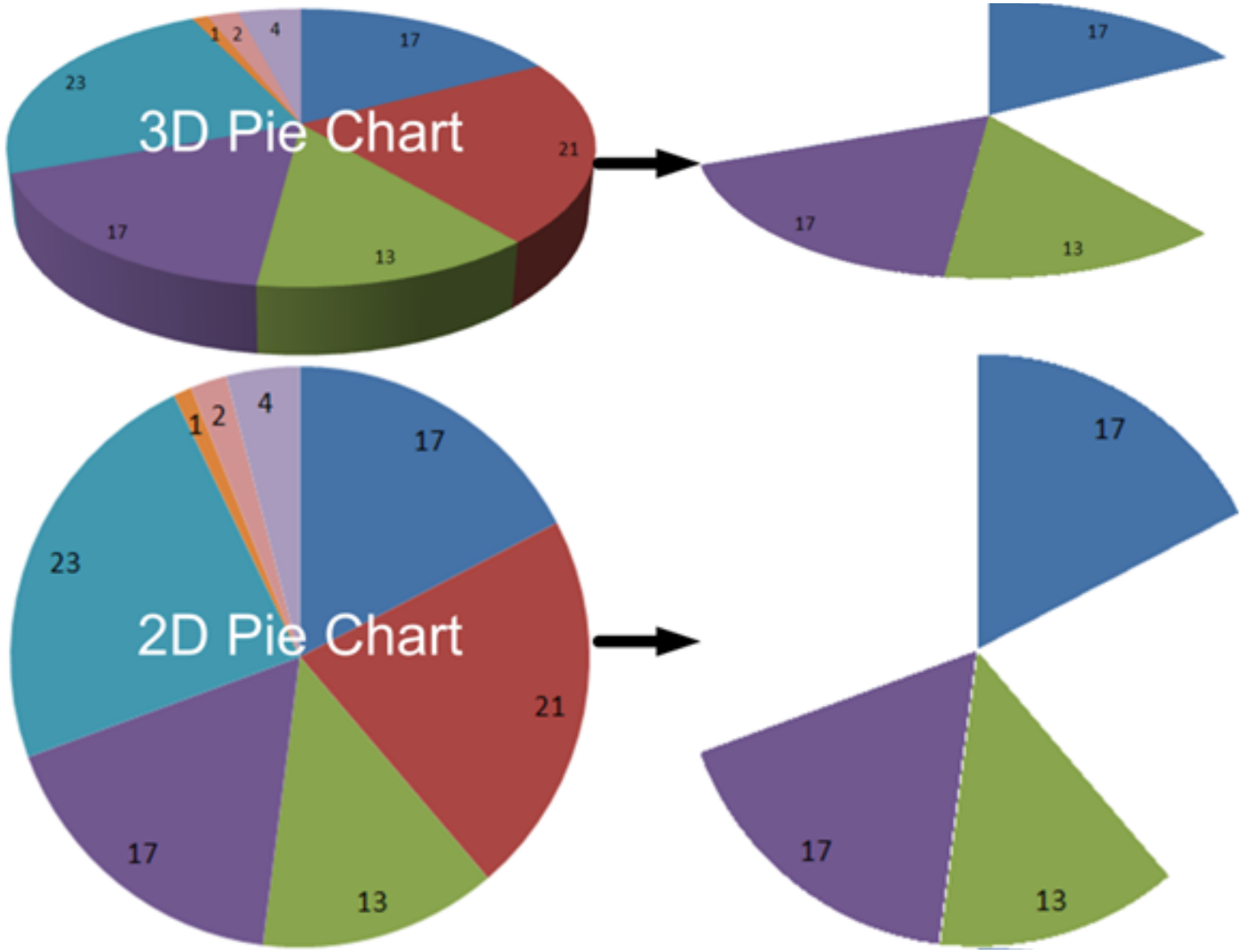


Fig. 2

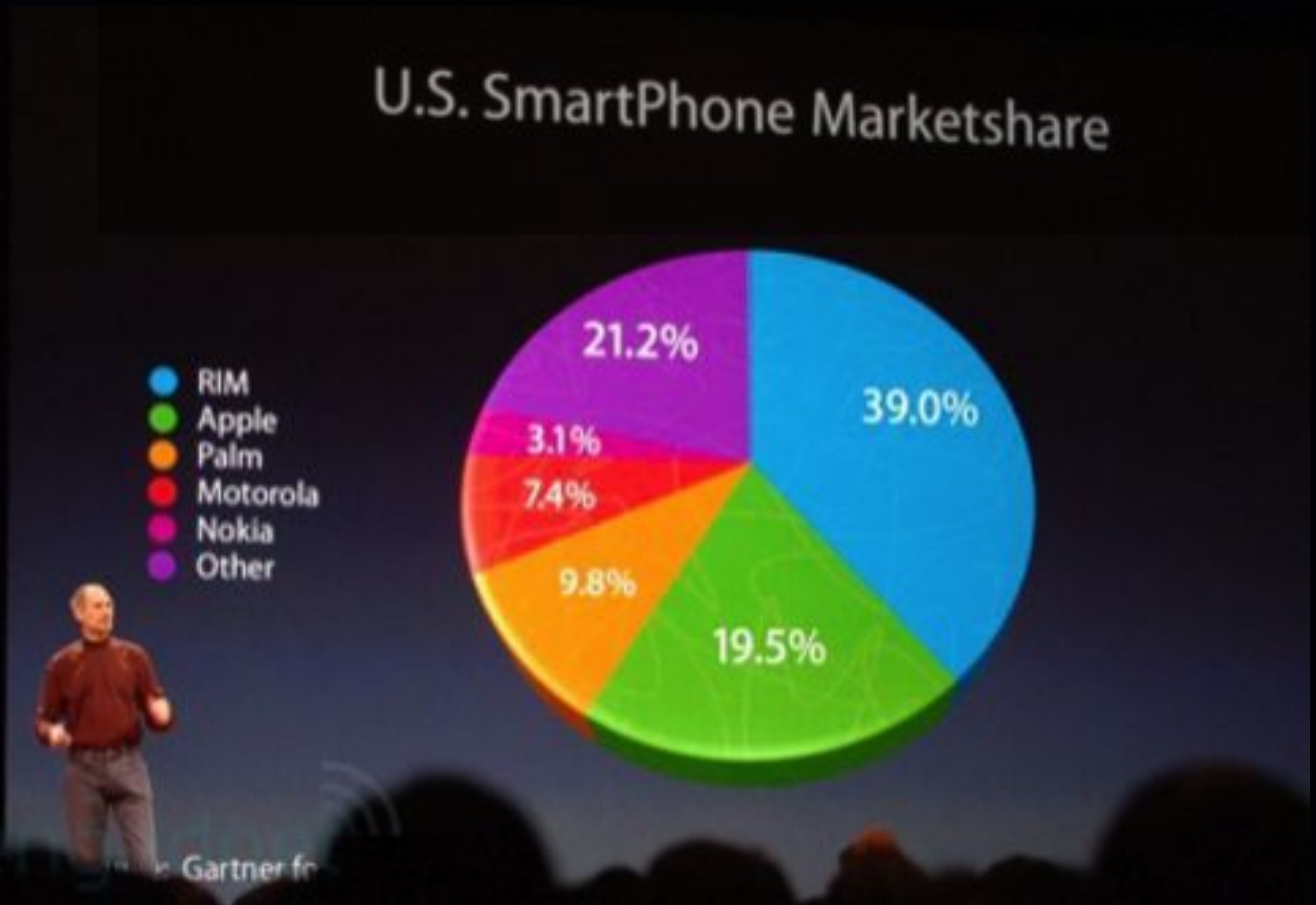


Adding a third dimension on two-dimensional medium creates perspective...

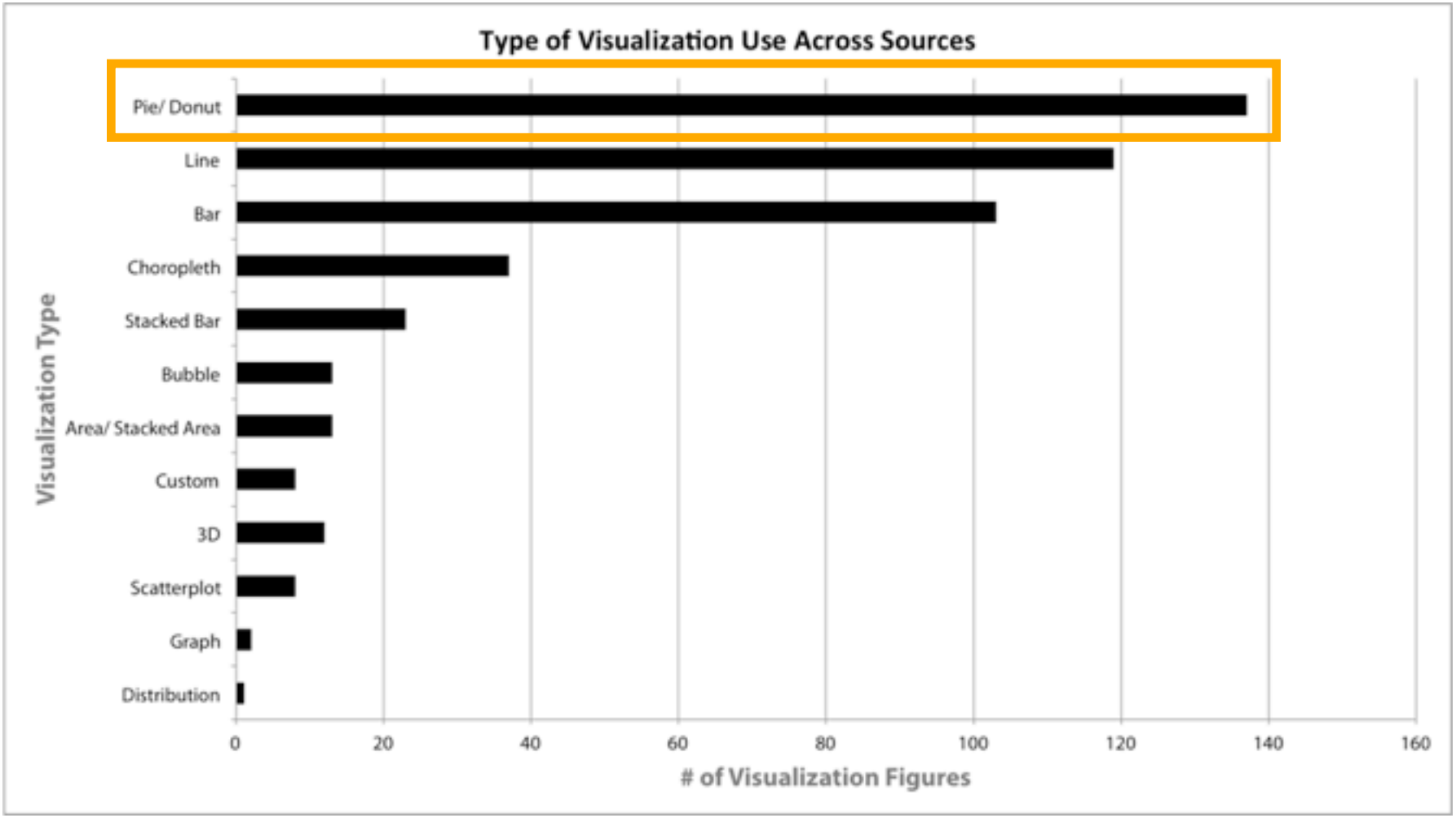


http://www.dashboardinsight.com/articles/digital-dashboards/building-dashboards/the-case-against-3d-charts-in-dashboards.aspx

Pie Charts


U.S. SmartPhone Marketshare
RIM
Apple
Palm
Motorola
Nokia
Other
39.0%
19.5%
9.8%
7.4%
3.1%
21.2%
Gartner

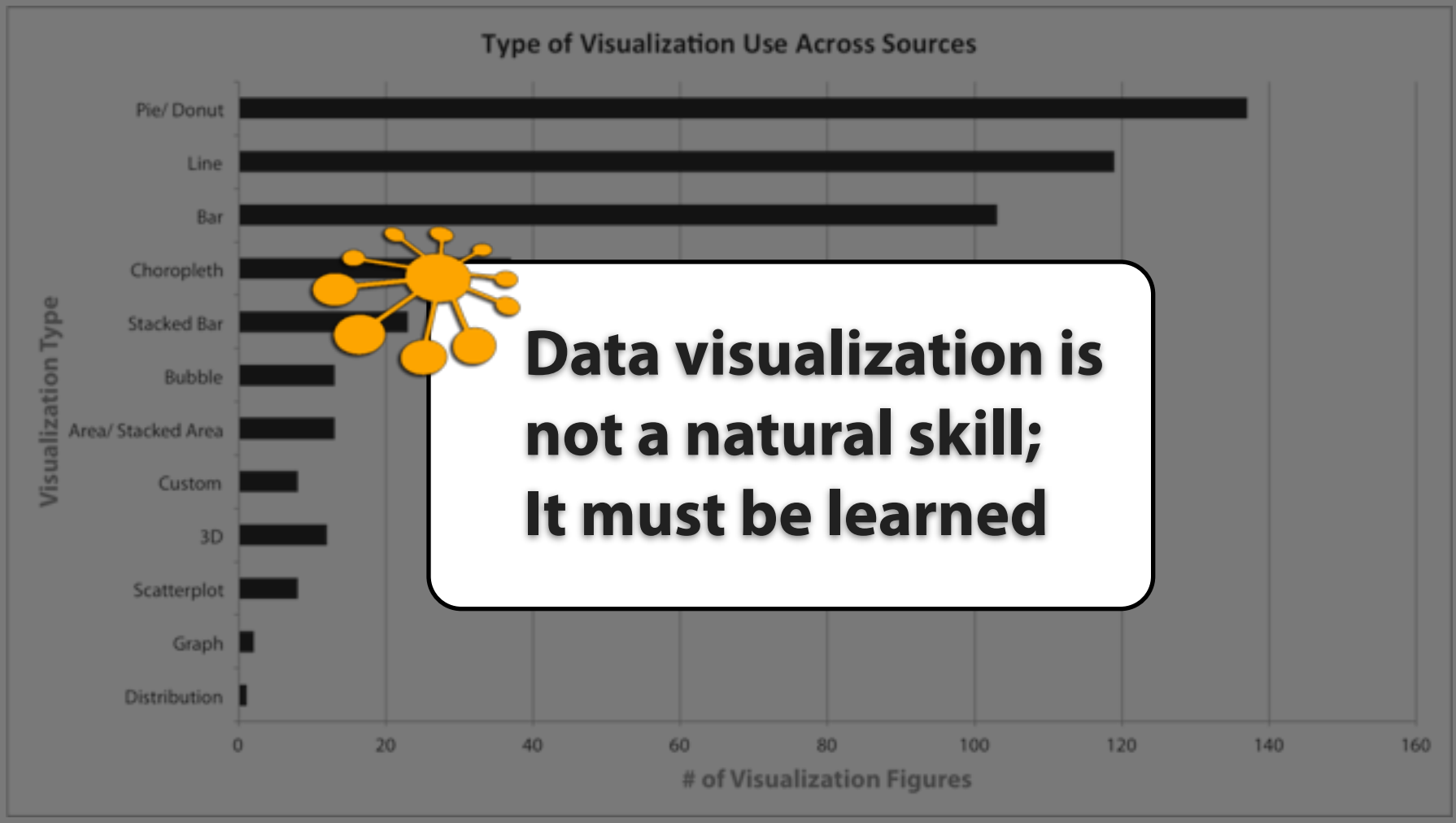
How are we as an Industry?



It seems y'all need to go on a diet (too much pie)

over 20 industry reports pulled from @jcran - http://bit.ly/QGqJdV

How are we as an Industry?



It seems y'all need to go on a diet (too much pie)

over 20 industry reports pulled from @jcran - http://bit.ly/QGqJdV

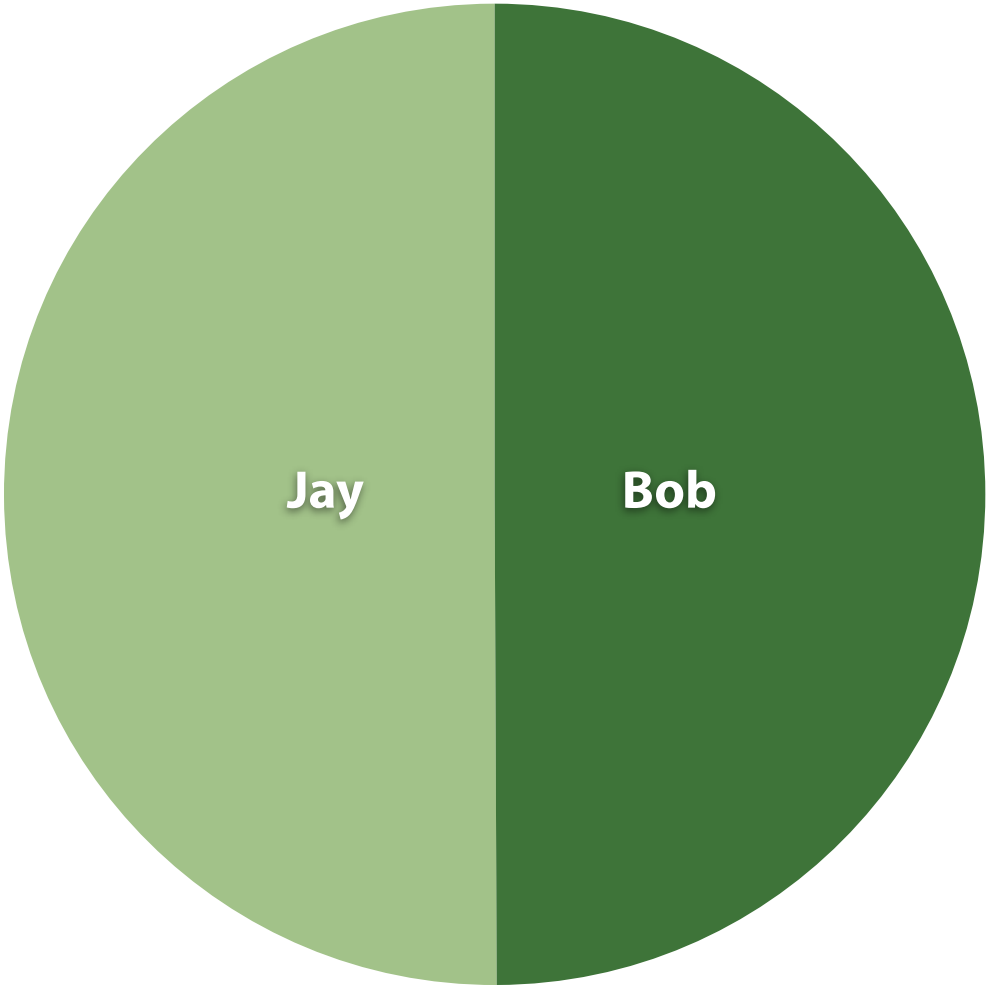
Avoid them, people don't decode well

Use them, people learn how to decode

If you must use Pie Charts...

- Never in 3D
- Limit categories, 3 to 6
- Start at 12, clockwise decreasing in quantity
- Avoid if angles are small or values are close

Slide Workload Distribution



Tufte Takeaways

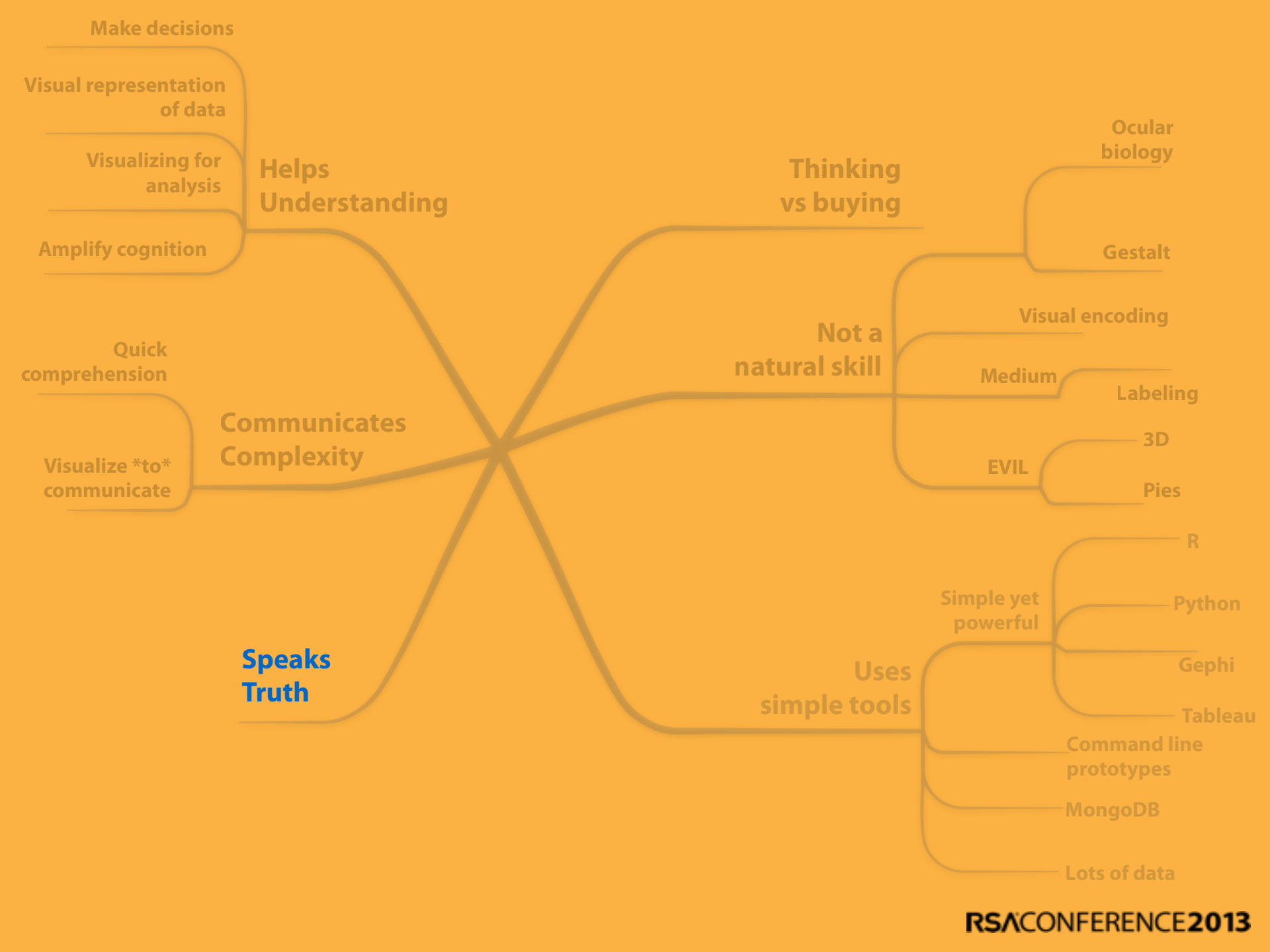
Chart Junk: the stuff that doesn't change when the data changes

Data Ink Ratio: what percentage of your ink shows data

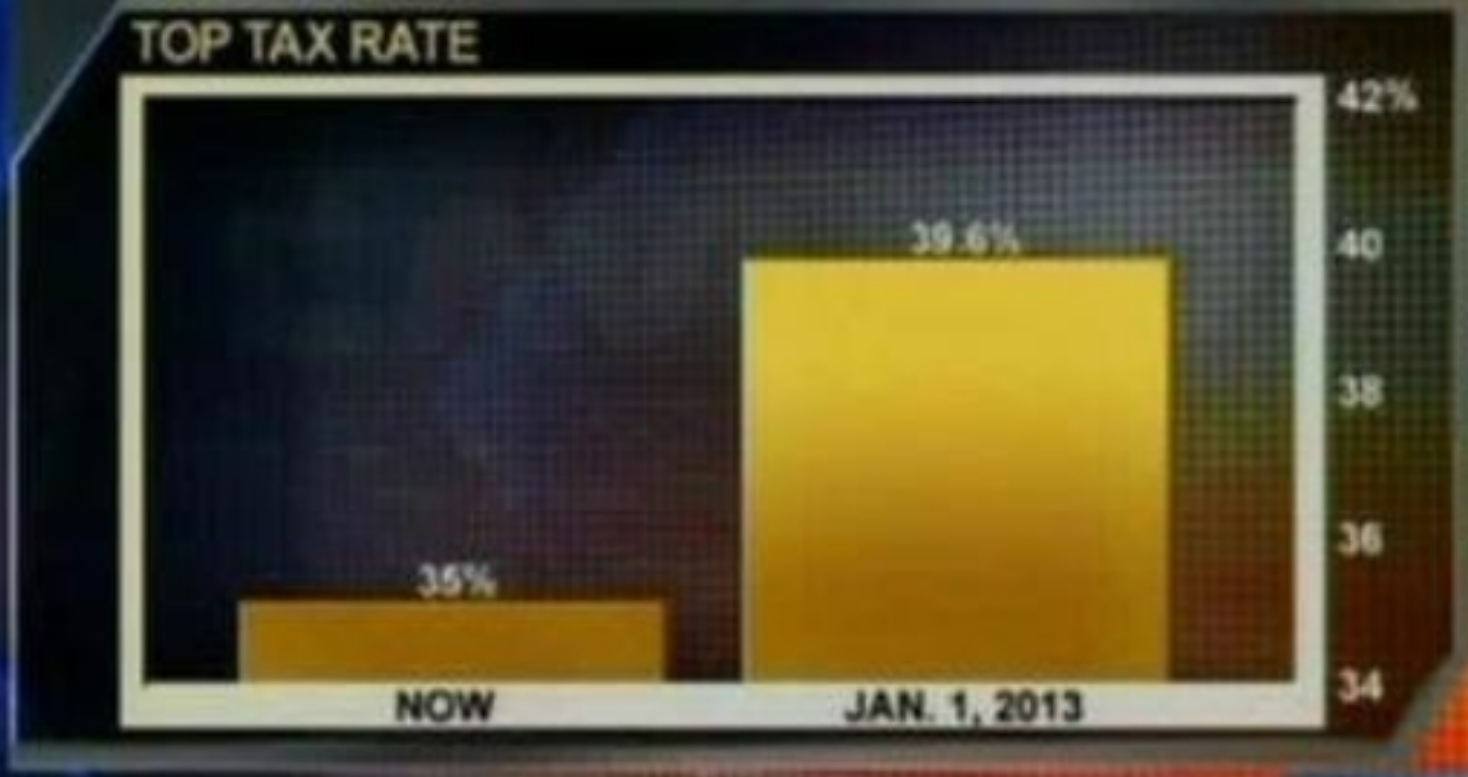
Smallest Effective Difference: the least you can do to highlight

http://www.edwardtufte.com/tufte/

http://anddesignmagazine.blogspot.com/2010/10/edward-tufte-creates-art-and.html


Make decisions
Visual representation of data
Visualizing for analysis
Amplify cognition
Helps Understanding
Quick comprehension
Visualize *to* communicate
Communicates Complexity
Speaks Truth
Thinking vs buying
Ocular biology
Gestalt
Not a natural skill
Visual encoding
Medium
Labeling
EVIL
3D
Pies
Uses simple tools
Simple yet powerful
R
Python
Gephi
Tableau
Command line prototypes
MongoDB
Lots of data

IF BUSH TAX CUTS EXPIRE



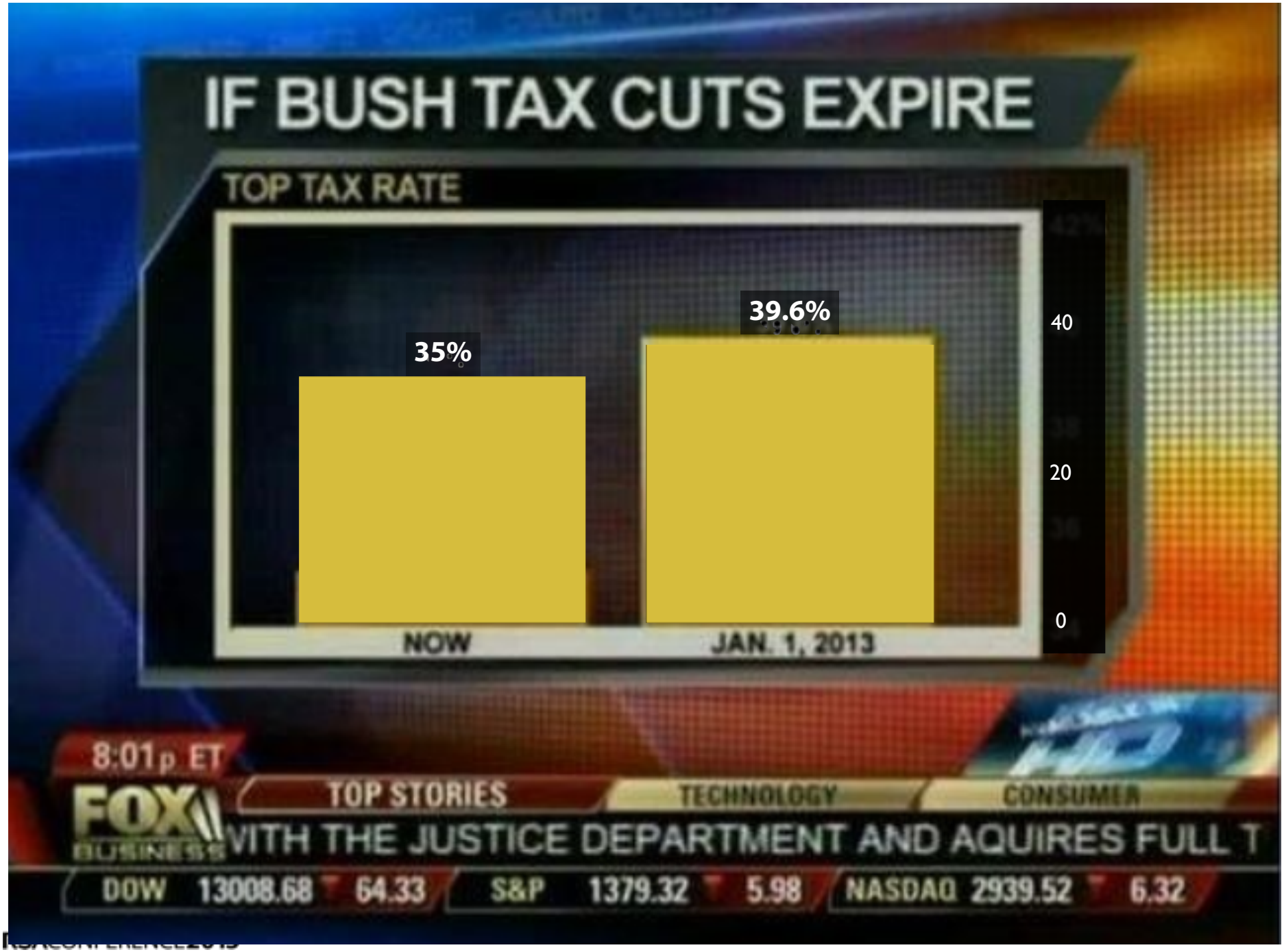
8:01p ET

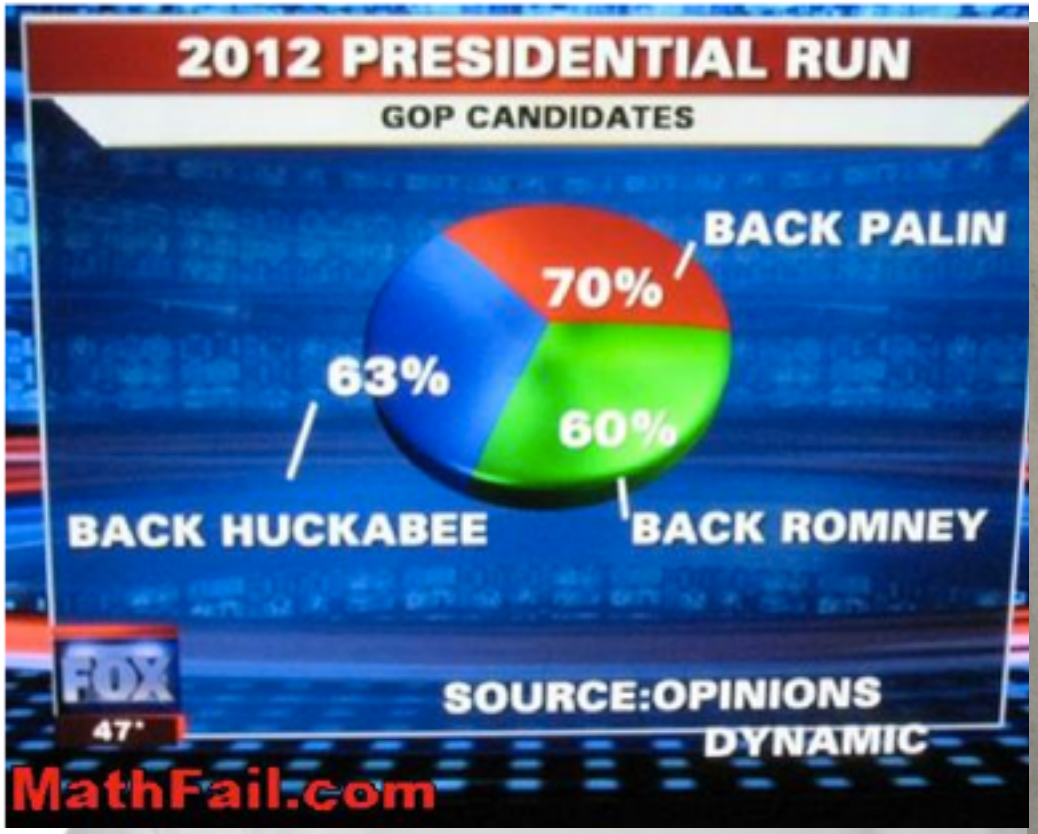
FOX BUSINESS

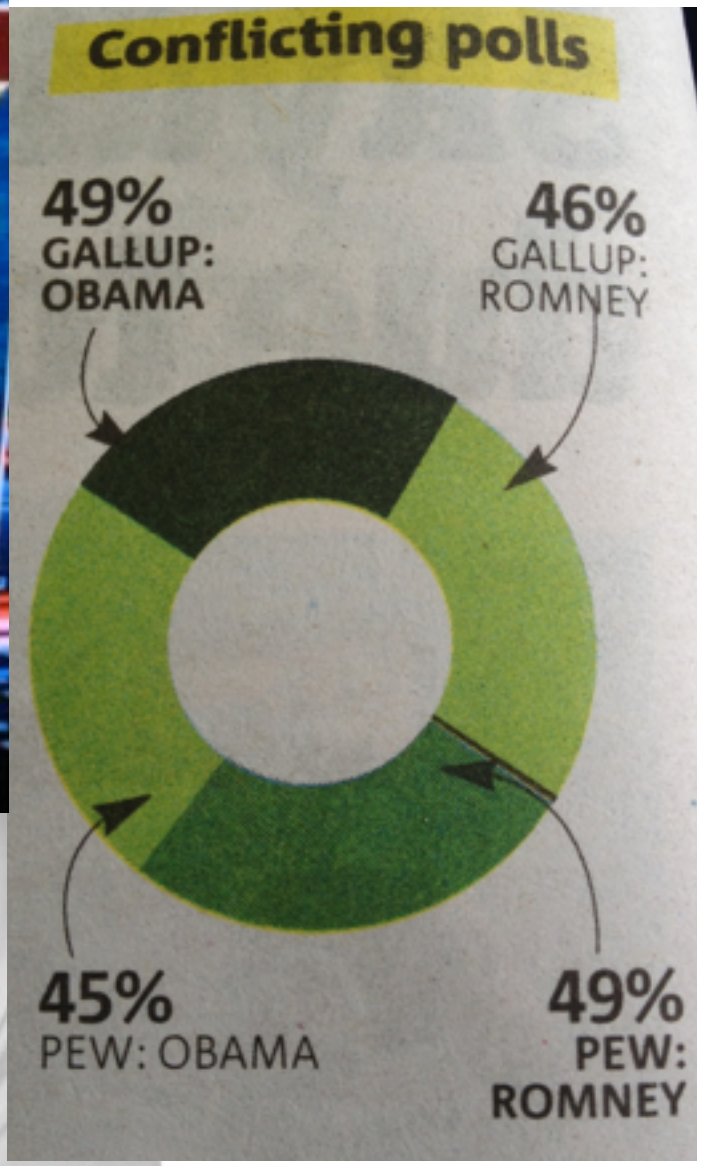
TOP STORIES | TECHNOLOGY | CONSUMER

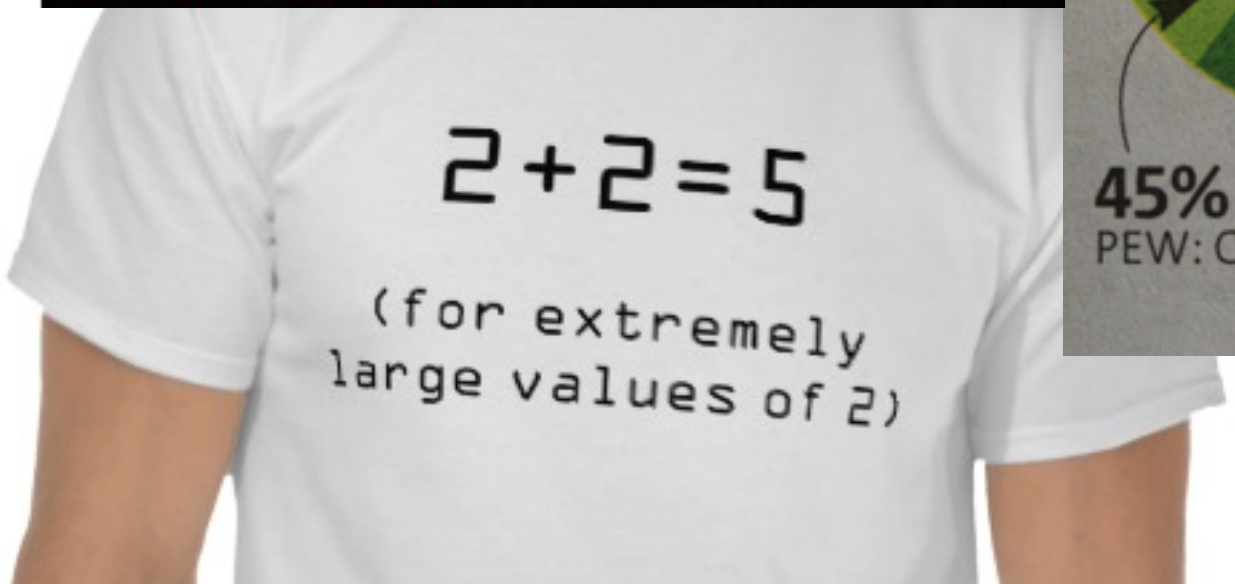
VITH THE JUSTICE DEPARTMENT AND AQUIRES FULL T

DOW 13008.68 ▼ 64.33 | S&P 1379.32 ▼ 5.98 | NASDAQ 2939.52 ▼ 6.32


IF BUSH TAX CUTS EXPIRE
TOP TAX RATE
35%
39.6%
40
20
0
NOW
JAN. 1, 2013
8:01p ET
FOX BUSINESS
TOP STORIES
TECHNOLOGY
CONSUMER
VITH THE JUSTICE DEPARTMENT AND AQUIRES FULL T
DOW 13008.68 64.33
S&P 1379.32 5.98
NASDAQ 2939.52 6.32

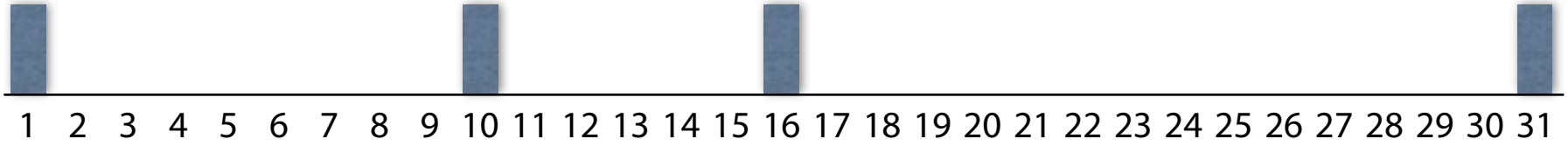

2012 PRESIDENTIAL RUN
GOP CANDIDATES
BACK PALIN
70%
63%
60%
BACK HUCKABEE
BACK ROMNEY
FOX
47'
SOURCE:OPINIONS
DYNAMIC
MathFail.com


Conflicting polls
49%
GALLUP:
OBAMA
46%
GALLUP:
ROMNEY
45%
PEW: OBAMA
49%
PEW:
ROMNEY

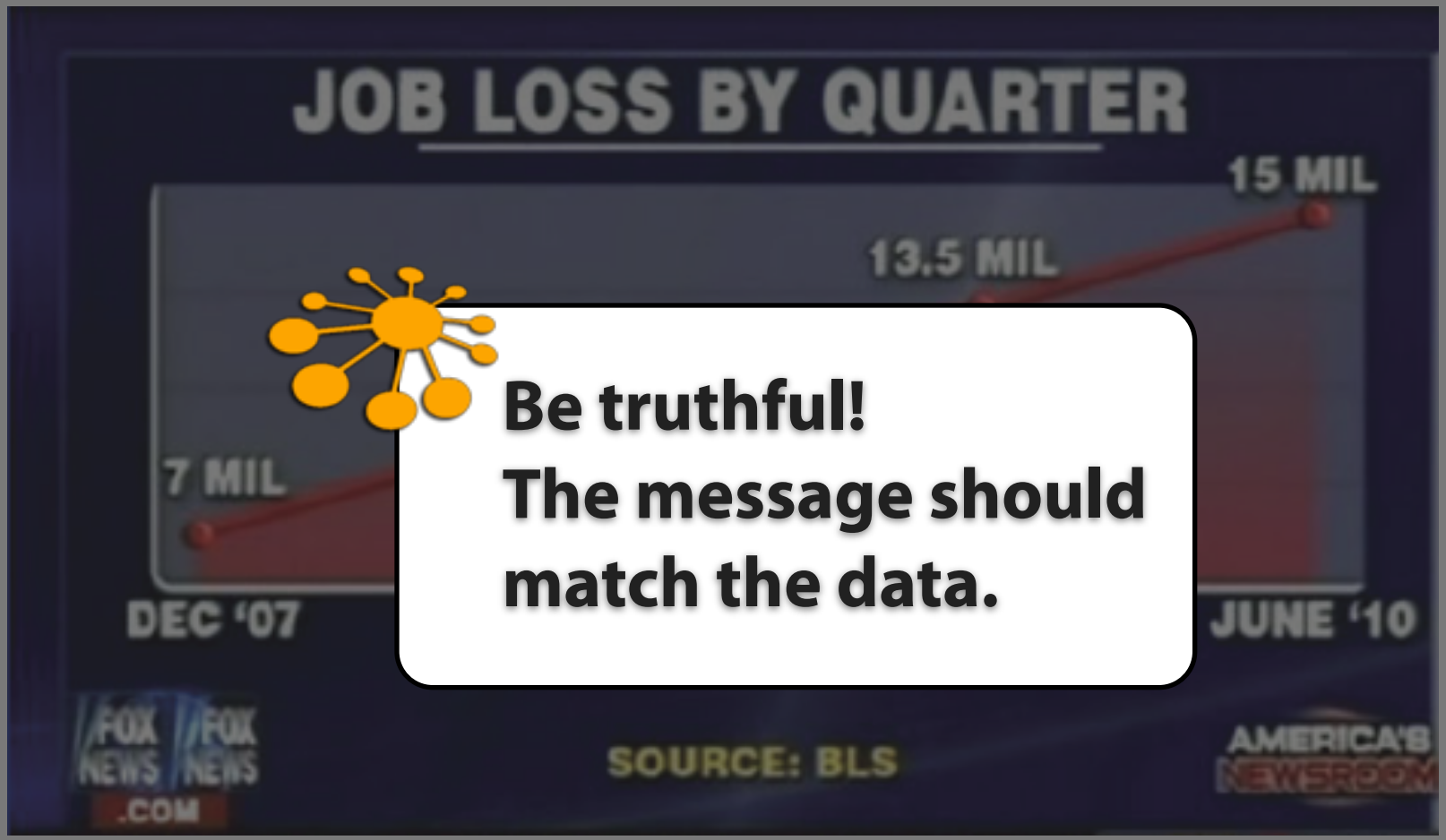

2+2=5
(for extremely
large values of 2)

Selection Bias?

[1st, 10th, 16th, and 31st month]

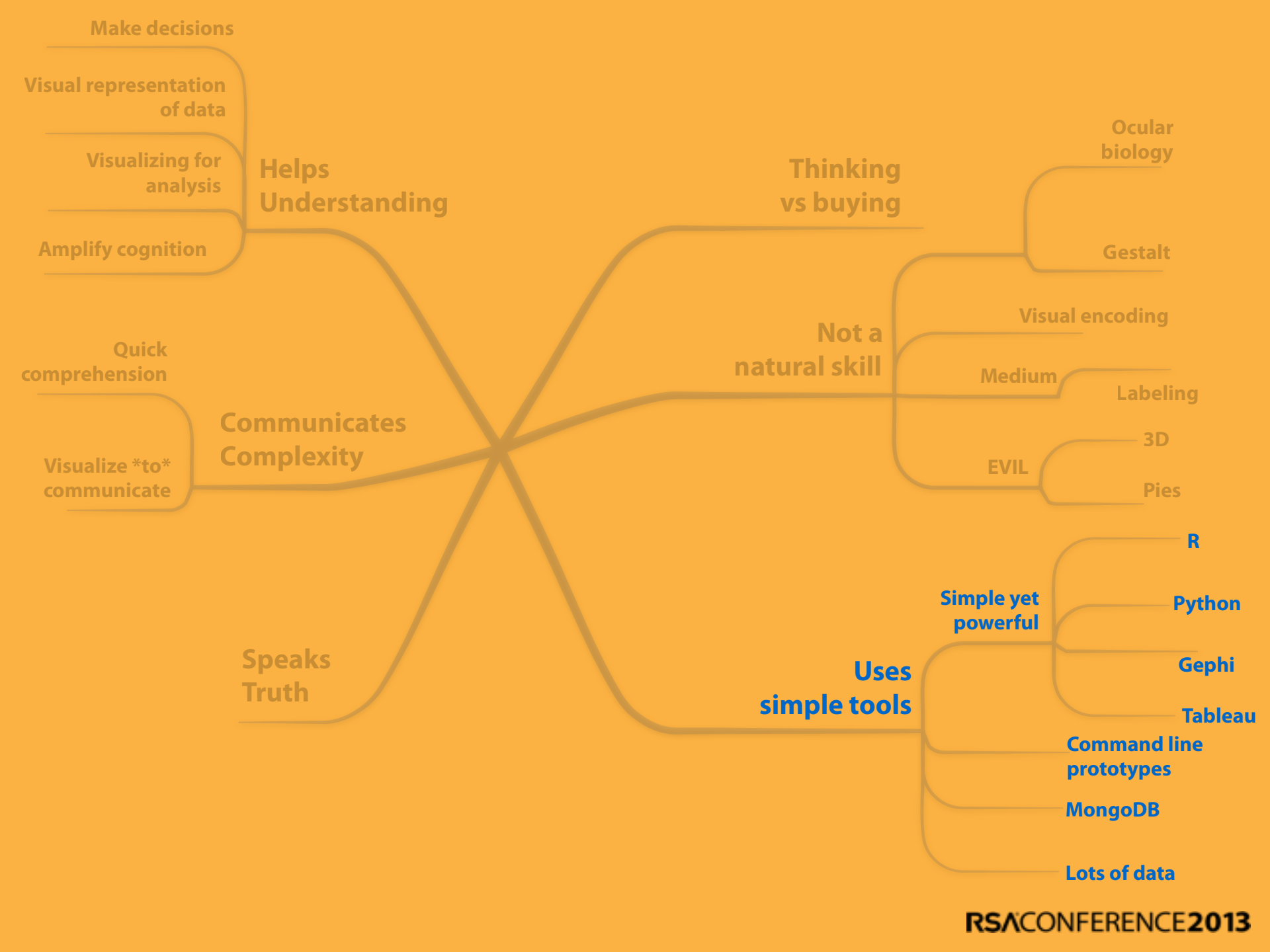


Selection Bias?



Be truthful!
The message should match the data.

[1st, 10th, 16th, and 31st month]


Make decisions
Visual representation of data
Visualizing for analysis
Amplify cognition
Helps Understanding
Quick comprehension
Visualize *to* communicate
Communicates Complexity
Speaks Truth
Thinking vs buying
Ocular biology
Gestalt
Not a natural skill
Visual encoding
Medium
Labeling
EVIL
3D
Pies
Uses simple tools
Simple yet powerful
R
Python
Gephi
Tableau
Command line prototypes
MongoDB
Lots of data

Firewall Logs Are A Good Example (Use case #1)

Source:

```
2011-04-13 08:52:52     Local4.Info     192.168.1.1     :Apr
13 08:52:52 PDT: %ASA-session-6-302013: Built inbound TCP
connection 41997795 for Workstations:192.168.2.133/4873
(192.168.2.133/4873) to Servers:192.168.1.6/135
(192.168.1.6/135)
```

Normalized:

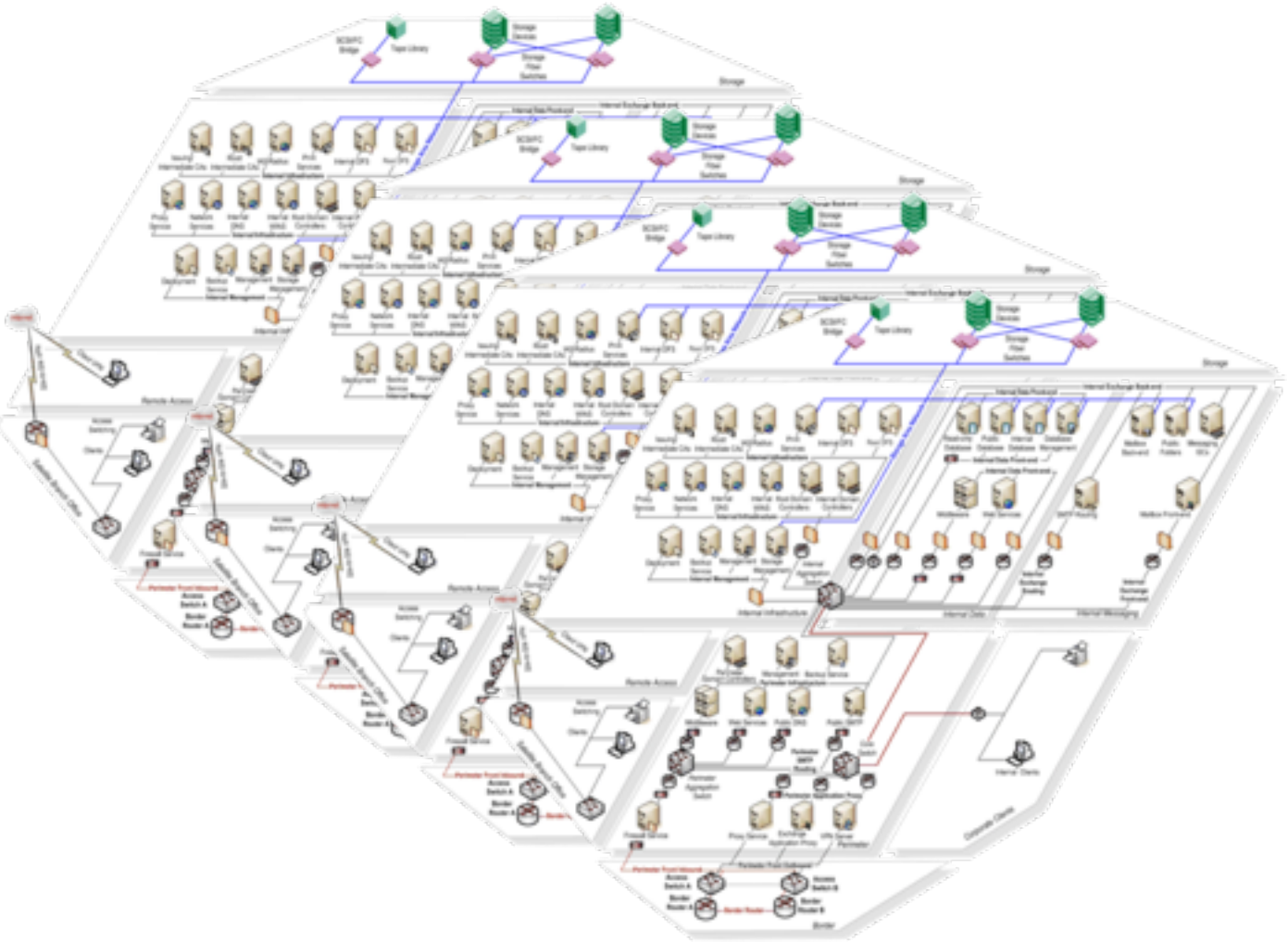
Date/time,Syslog priority,Operation,Message code,Protocol,Source IP,Destination IP,Source hostname,Destination hostname,Source port,Destination port,Destination service,Direction,Connections built,Connections torn down

```
13/Apr/2011 08:52:52,Info,Built,ASA-session-6-302013,TCP,
192.168.2.133,192.168.1.6,(empty),(empty),
4873,135,epmap,inbound,1,0
```

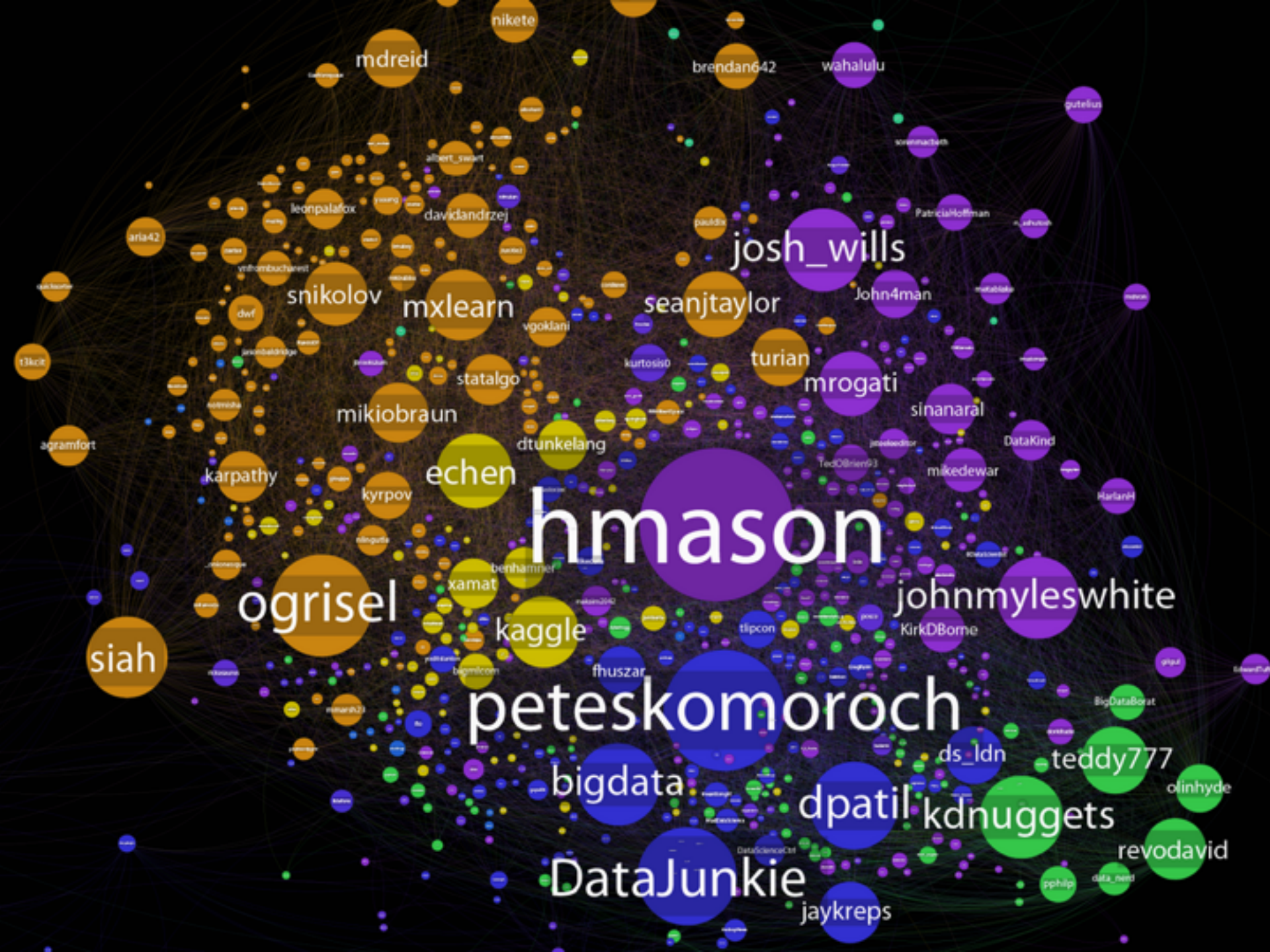
```
$ grepfield -p  Built 20110413_20110414_fw_log* | aggregate -p -k
6 -c 6 -d \, | sort -n -t, -k2 | tail
192.168.2.98,1558
192.168.2.11,2752
192.168.2.46,3457
192.168.1.6,10753
(empty),39931
10.200.150.208,920071
10.200.150.207,1017366
10.200.150.206,1145203
10.200.150.209,1145360
10.200.150.201,1165935
```

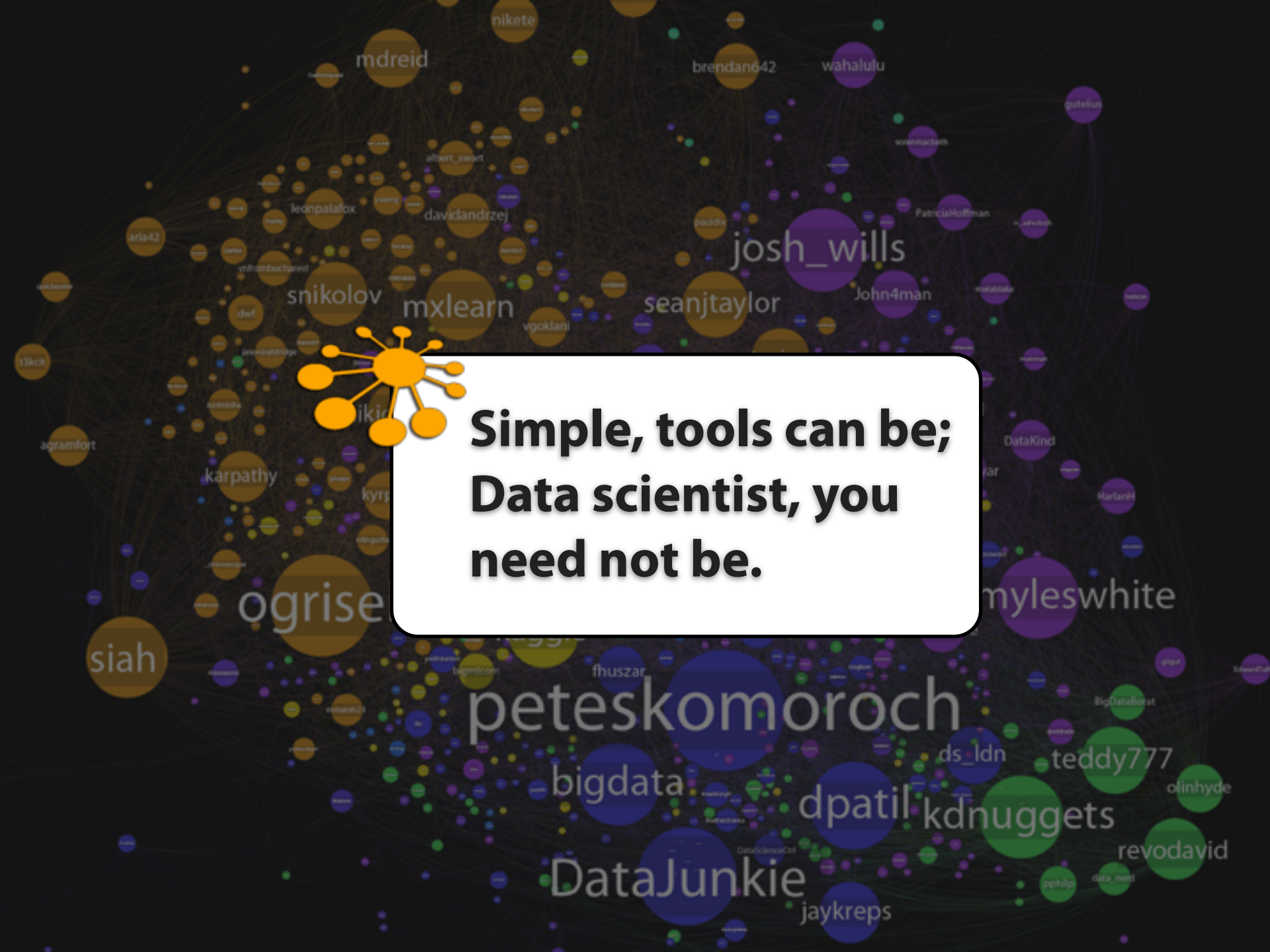

http://code.google.com/p/crush-tools/

- 10,452,115 events
- 1.3GB of data
- 4.5 hours (not even one day)
- **1 firewall**



Command-line tools aren't enough


nikete
mdreid
brendan642
wahalulu
albert_swart
leonpalafox
davidandrzej
aria42
josh_wills
PatriciaHoffman
snikolov
mxlearn
seanjtaylor
John4man
vgokhani
dwf
turian
statalgo
mrogati
mikiobraun
sinanaral
agramfort
dtunkelang
DataKind
karpathy
echen
mikedewar
kyrpov
hmason
xamat
ogrisel
johnmyleswhite
kaggle
KirkDBorne
siah
fhuszar
peteskomoroch
BigDataBorat
ds_ldn
teddy777
bigdata
dpatil
kdnuggets
olinhyde
revodavid
DataJunkie
jaykreps



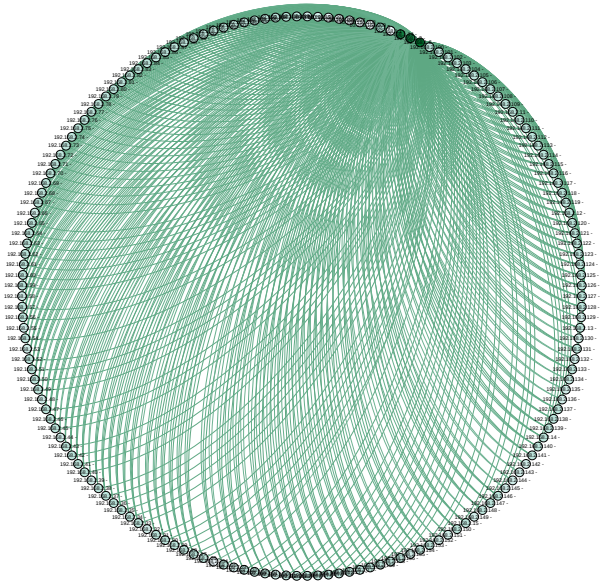
Simple, tools can be; Data scientist, you need not be.

```
[   {
      "Syslog priority": "Info",
      "Protocol": "TCP",
      "Destination IP": "192.168.1.6",
      "Destination port": "135",
      "Source IP": "192.168.2.133",
      "Connections torn down": "0",
      "Direction": "inbound",
      "Connections built": "1",
      "Message code": "ASA-session-6-302013",
      "Date/time": "13/Apr/2011 08:52:52",
      "Destination service": "epmap",
      "Source port": "4873",
      "Destination hostname": "(empty)",
      "Source hostname": "(empty)",
      "Operation": "Built"
   },
   {
      "Syslog priority": "Info",
      "Protocol": "TCP",
      "Destination IP": "192.168.1.6",
      "Destination port": "43025",
      "Source IP": "192.168.2.133",
      "Connections torn down": "0",
      "Direction": "inbound",
      "Connections built": "1",
      "Message code": "ASA-session-6-302013",
      "Date/time": "13/Apr/2011 08:52:52",
      "Destination service": "43025_tcp",
      "Source port": "4874",
      "Destination hostname": "(empty)",
      "Source hostname": "(empty)",
      "Operation": "Built"
    }   ]
```

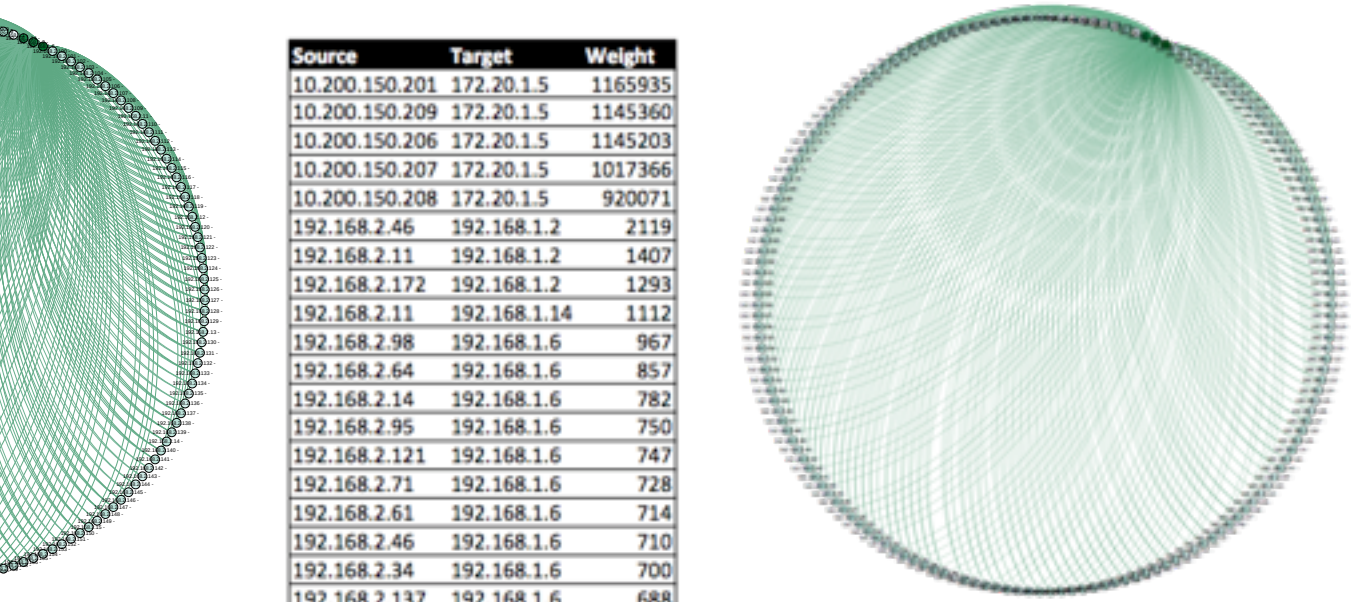
```
[   {
      "Syslog priority": "Info",
      "Protocol": "TCP",
      "Destination IP": "192.168.1.6",
      "Destination port": "135",
      "Source IP": "192.168.2.133",
      "Connections torn down": "0",
      "Direction": "inbound",
      "Connections built": "1",
      "Message code": "ASA-session-6-302013",
      "Date/time": "13/Apr/2011 08:52:52",
      "Destination service": "epmap",
      "Source port": "4873",
      "Destination hostname": "(empty)",
      "Source hostname": "(empty)",
      "Operation": "Built"
   },
   {
      "Syslog priority": "Info",
      "Protocol": "TCP",
      "Destination IP": "192.168.1.6",
      "Destination port": "43025",
      "Source IP": "192.168.2.133",
      "Connections torn down": "0",
      "Direction": "inbound",
      "Connections built": "1",
      "Message code": "ASA-session-6-302013",
      "Date/time": "13/Apr/2011 08:52:52",
      "Destination service": "43025_tcp",
      "Source port": "4874",
      "Destination hostname": "(empty)",
      "Source hostname": "(empty)",
      "Operation": "Built"
    }   ]
```

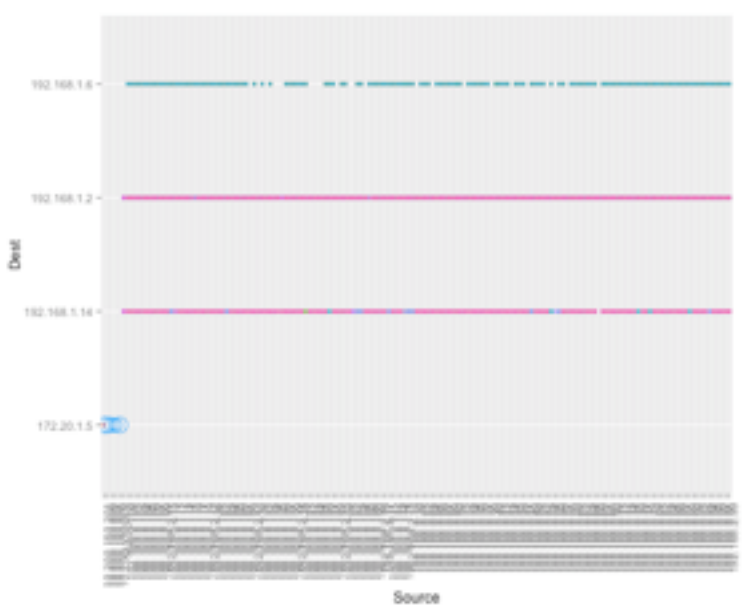

```
#!/usr/bin/python

import csv
import json
import sys

csv_file = open(sys.argv[1],"r")
reader = csv.reader(csv_file)
header = reader.next()
for row in reader:
    jsondict = {}
    for i in range(len(header)):
        jsondict[header[i]] = row[i]
    print json.dumps(jsondict)
```

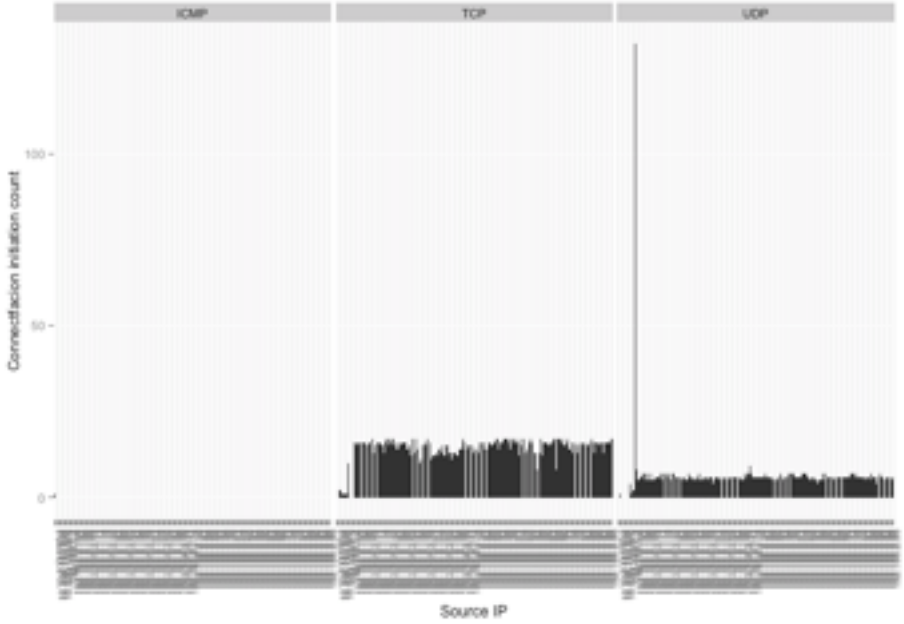
```
mongoimport
db.fw.aggregate( [  { $match : { day : "13" } },  # match the first day
                    { $group : { _id : "$src", count : { $sum : 1 } } },  # group and count source
                    { $project : { _id : 0, src : "$_id", count: "$count" } },  # project into structure
                    { $sort : { count : -1, _id : -1 } },   # sort by counts
                    { $limit : 10 } ] )  # show just top 10

        {       "count" : 1165935,
                "src" : "10.200.150.201" },
        {       "count" : 1145360,
                "src" : "10.200.150.209" },
        {       "count" : 1145203,
                "src" : "10.200.150.206" },
        {       "count" : 1017366,
                "src" : "10.200.150.207" },
        {       "count" : 920071,
                "src" : "10.200.150.208" },
        {       "count" : 39931,
                "src" : "(empty)" },
        {       "count" : 10753,
                "src" : "192.168.1.6" },
        {       "count" : 3457,
                "src" : "192.168.2.46" },
        {       "count" : 2752,
                "src" : "192.168.2.11" },
        {       "count" : 1558,
                "src" : "192.168.2.98" }
"ok" : 1
```

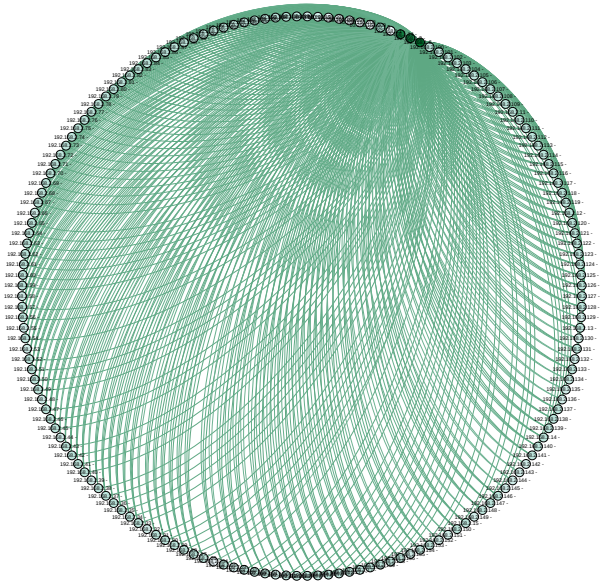


Source	Target	Weight
10.200.150.201	172.20.1.5	1165935
10.200.150.209	172.20.1.5	1145360
10.200.150.206	172.20.1.5	1145203
10.200.150.207	172.20.1.5	1017366
10.200.150.208	172.20.1.5	920071
192.168.2.46	192.168.1.2	2119
192.168.2.11	192.168.1.2	1407
192.168.2.172	192.168.1.2	1293
192.168.2.11	192.168.1.14	1112
192.168.2.98	192.168.1.6	967
192.168.2.64	192.168.1.6	857
192.168.2.14	192.168.1.6	782
192.168.2.95	192.168.1.6	750
192.168.2.121	192.168.1.6	747
192.168.2.71	192.168.1.6	728
192.168.2.61	192.168.1.6	714
192.168.2.46	192.168.1.6	710
192.168.2.34	192.168.1.6	700
192.168.2.137	192.168.1.6	688
192.168.2.97	192.168.1.6	654

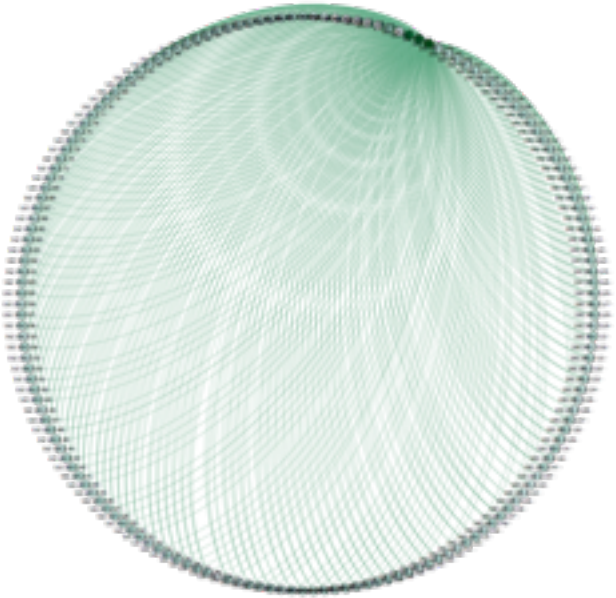


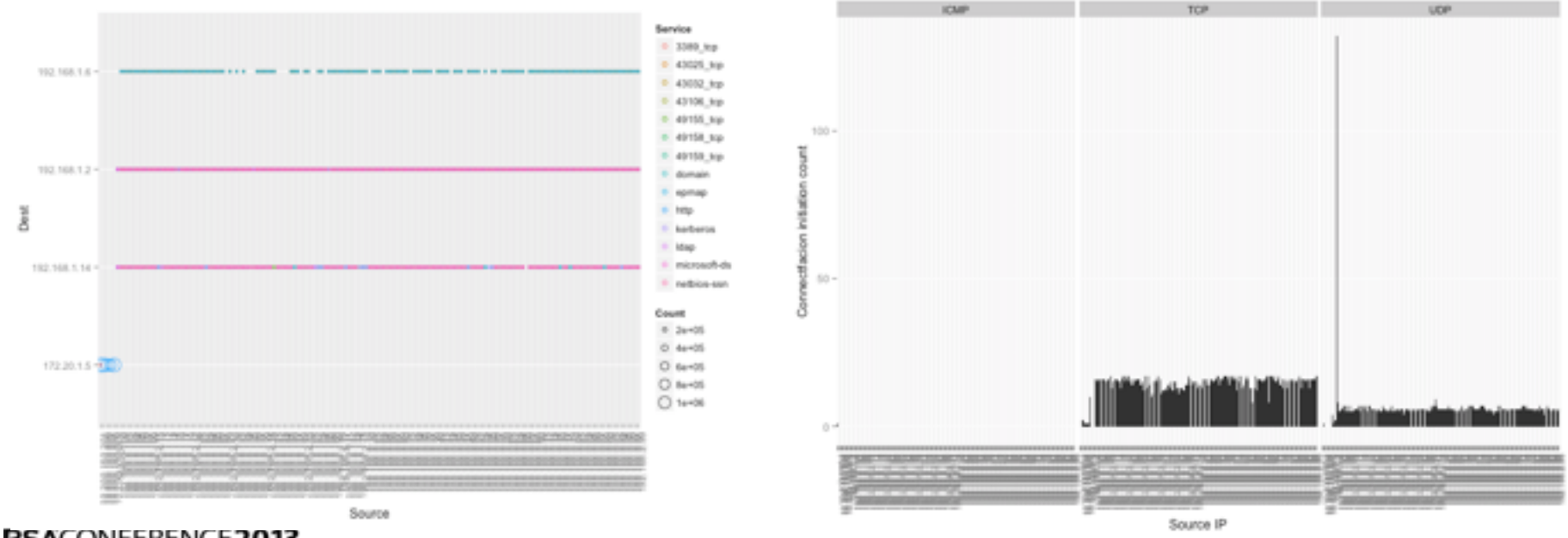


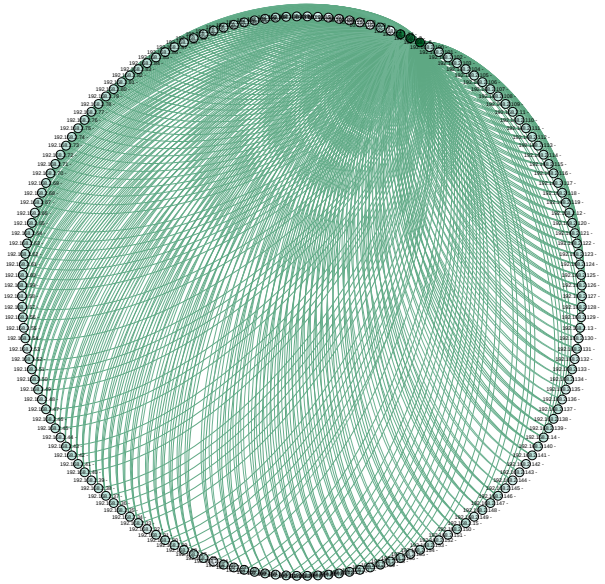




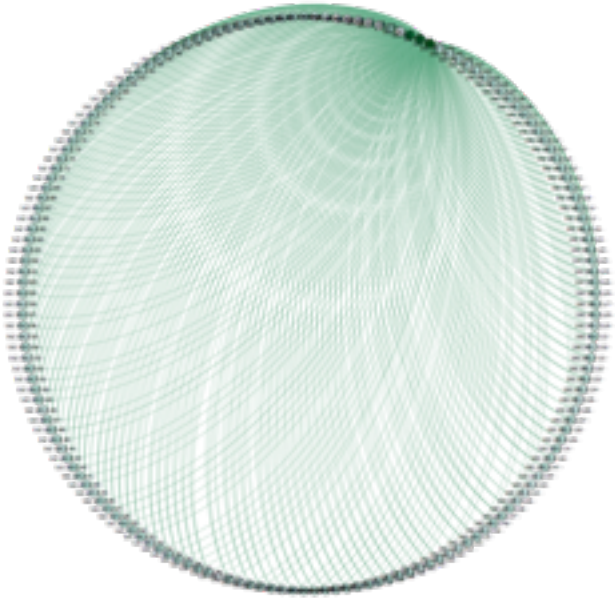
Source	Target	Weight
10.200.150.201	172.20.1.5	1165935
10.200.150.209	172.20.1.5	1145360
10.200.150.206	172.20.1.5	1145203
10.200.150.207	172.20.1.5	1017366
10.200.150.208	172.20.1.5	920071
192.168.2.11	192.168.1.2	1407
192.168.2.172	192.168.1.2	1293
192.168.2.11	192.168.1.14	1112
192.168.2.98	192.168.1.6	967
192.168.2.64	192.168.1.6	857
192.168.2.14	192.168.1.6	782
192.168.2.95	192.168.1.6	750
192.168.2.121	192.168.1.6	747
192.168.2.71	192.168.1.6	728
192.168.2.61	192.168.1.6	714
192.168.2.46	192.168.1.6	710
192.168.2.34	192.168.1.6	700
192.168.2.137	192.168.1.6	688
192.168.2.97	192.168.1.6	654

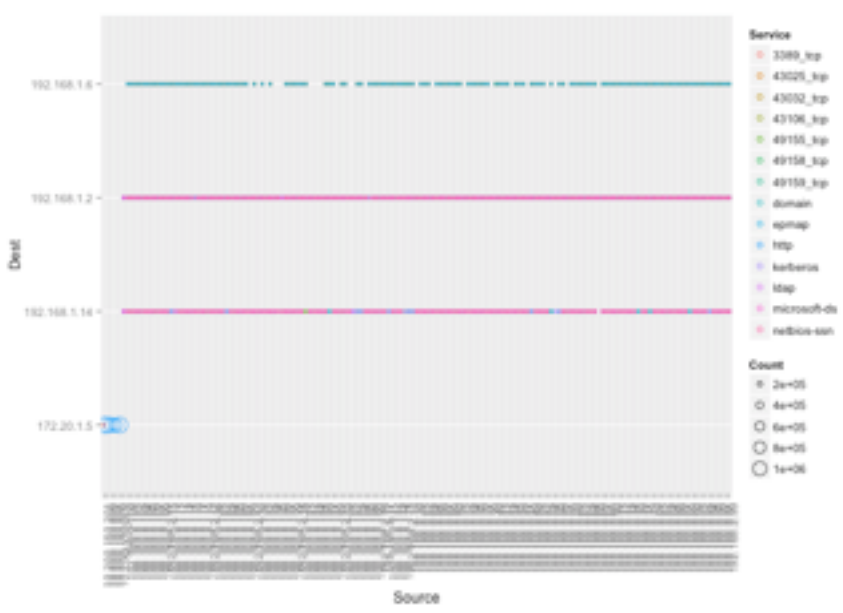


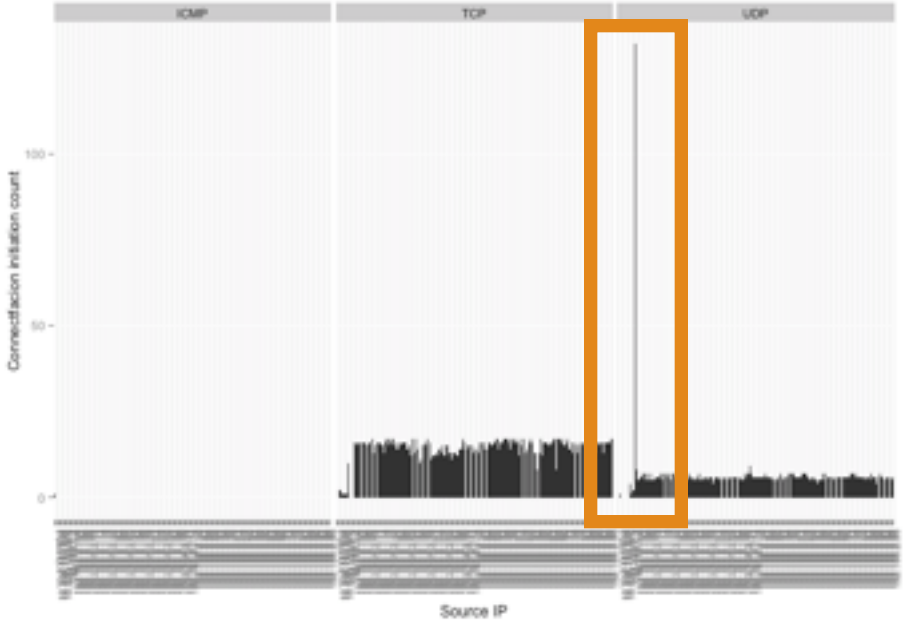


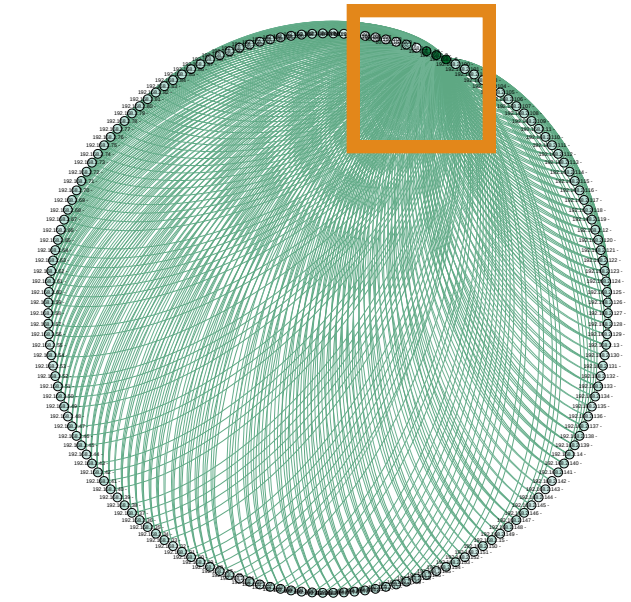


Source	Target	Weight
10.200.150.201	172.20.1.5	1165935
10.200.150.209	172.20.1.5	1145360
10.200.150.206	172.20.1.5	1145203
10.200.150.207	172.20.1.5	1017366
10.200.150.208	172.20.1.5	920071
192.168.2.11	192.168.1.2	1407
192.168.2.172	192.168.1.2	1293
192.168.2.11	192.168.1.14	1112
192.168.2.98	192.168.1.6	967
192.168.2.64	192.168.1.6	857
192.168.2.14	192.168.1.6	782
192.168.2.95	192.168.1.6	750
192.168.2.121	192.168.1.6	747
192.168.2.71	192.168.1.6	728
192.168.2.61	192.168.1.6	714
192.168.2.46	192.168.1.6	710
192.168.2.34	192.168.1.6	700
192.168.2.137	192.168.1.6	688
192.168.2.97	192.168.1.6	654

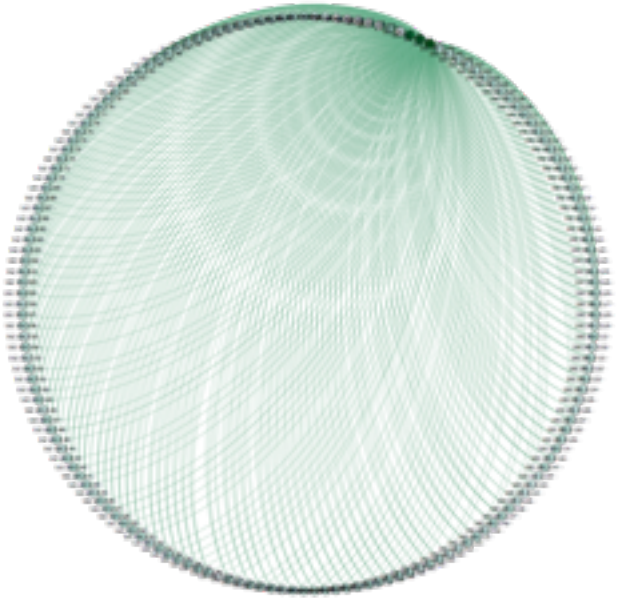


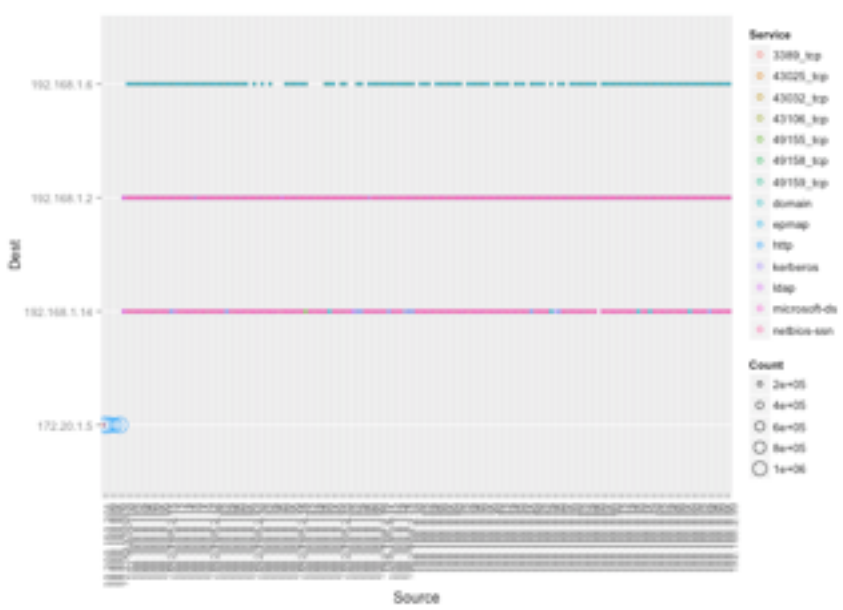


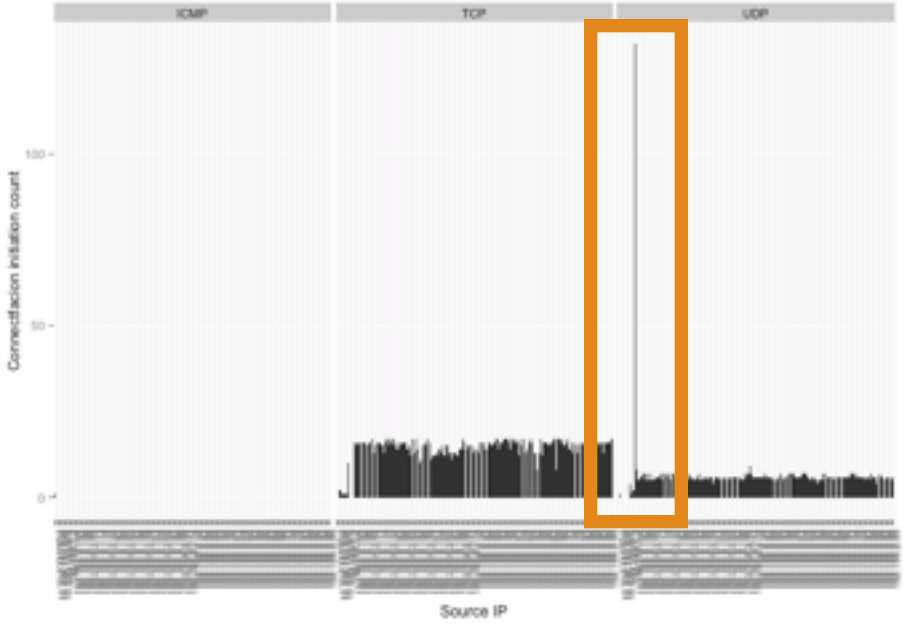


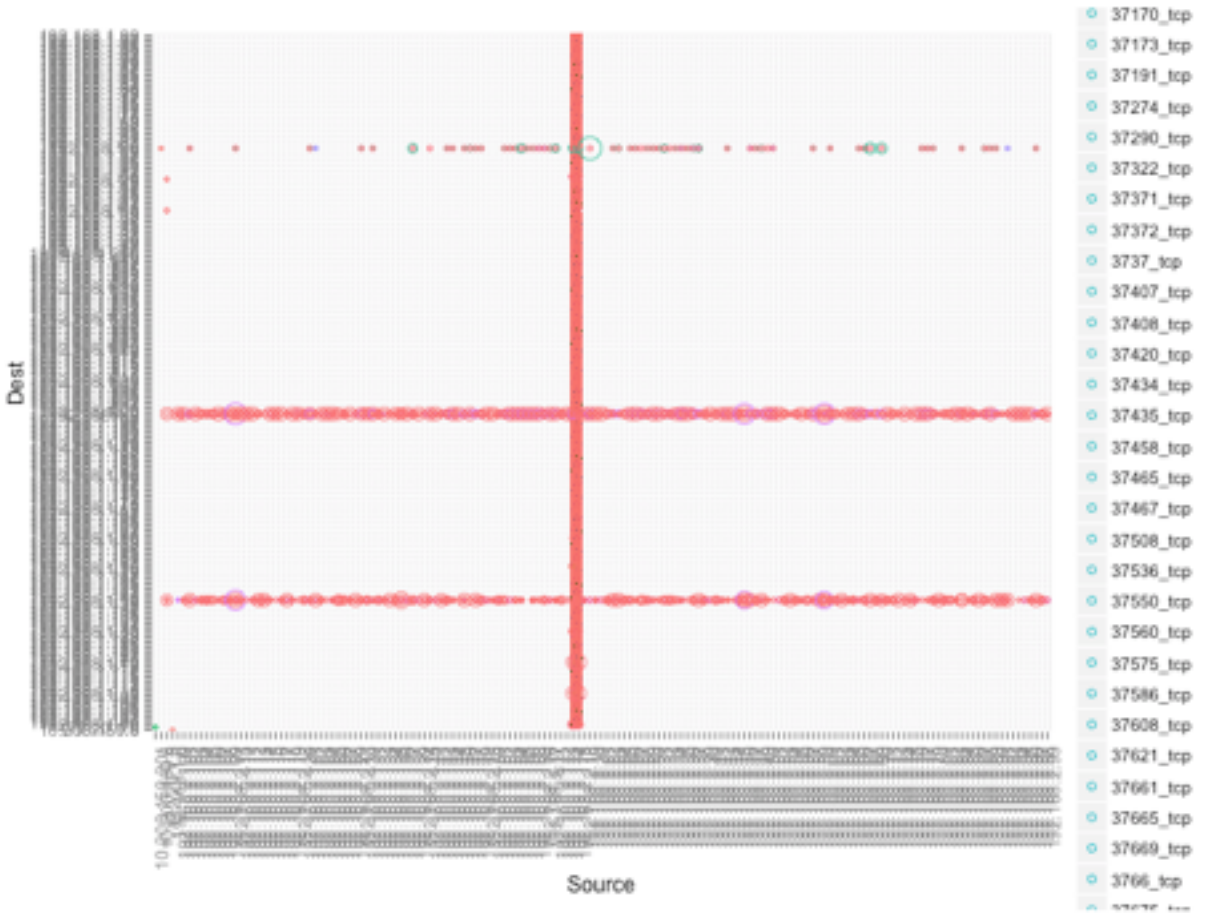


Source	Target	Weight
10.200.150.201	172.20.1.5	1165935
10.200.150.209	172.20.1.5	1145360
10.200.150.206	172.20.1.5	1145203
10.200.150.207	172.20.1.5	1017366
10.200.150.208	172.20.1.5	920071
192.168.2.11	192.168.1.2	1407
192.168.2.172	192.168.1.2	1293
192.168.2.11	192.168.1.14	1112
192.168.2.98	192.168.1.6	967
192.168.2.64	192.168.1.6	857
192.168.2.14	192.168.1.6	782
192.168.2.95	192.168.1.6	750
192.168.2.121	192.168.1.6	747
192.168.2.71	192.168.1.6	728
192.168.2.61	192.168.1.6	714
192.168.2.46	192.168.1.6	710
192.168.2.34	192.168.1.6	700
192.168.2.137	192.168.1.6	688
192.168.2.97	192.168.1.6	654









Source	Count
192.168.2.175	442,351
192.168.2.174	435,431
192.168.1.6	2,880
1921.68.1.2	2,599
192.168.2.11	2,557
192.168.2.46	2,518

Analysis → Action: **Investigate 174/175**

Packets over Time...

http://www.youtube.com/watch?v=mGr1GpV-YcE

Packets over Time...

http://www.youtube.com/watch?v=mGr1GpV-YcE

Use Case #2: Geo-location of IP addresses

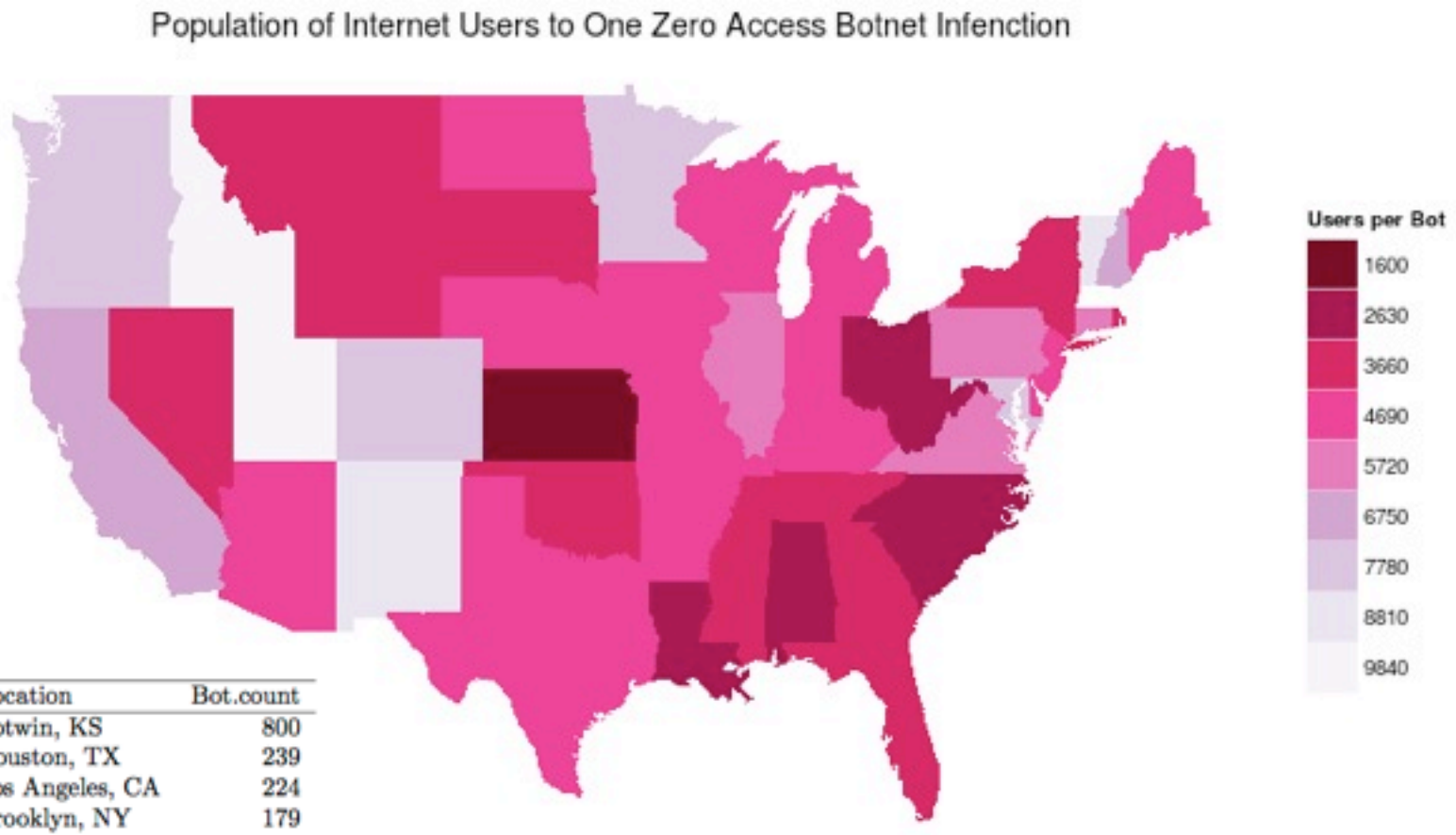
“Some botnets are so big… you can see them from space (or at least, Google Earth).”

http://www.f-secure.com/weblog/archives/00002428.html
http://www.f-secure.com/weblog/archives/00002430.html

F-Secure releases 140,000 ZeroAccess geolocations

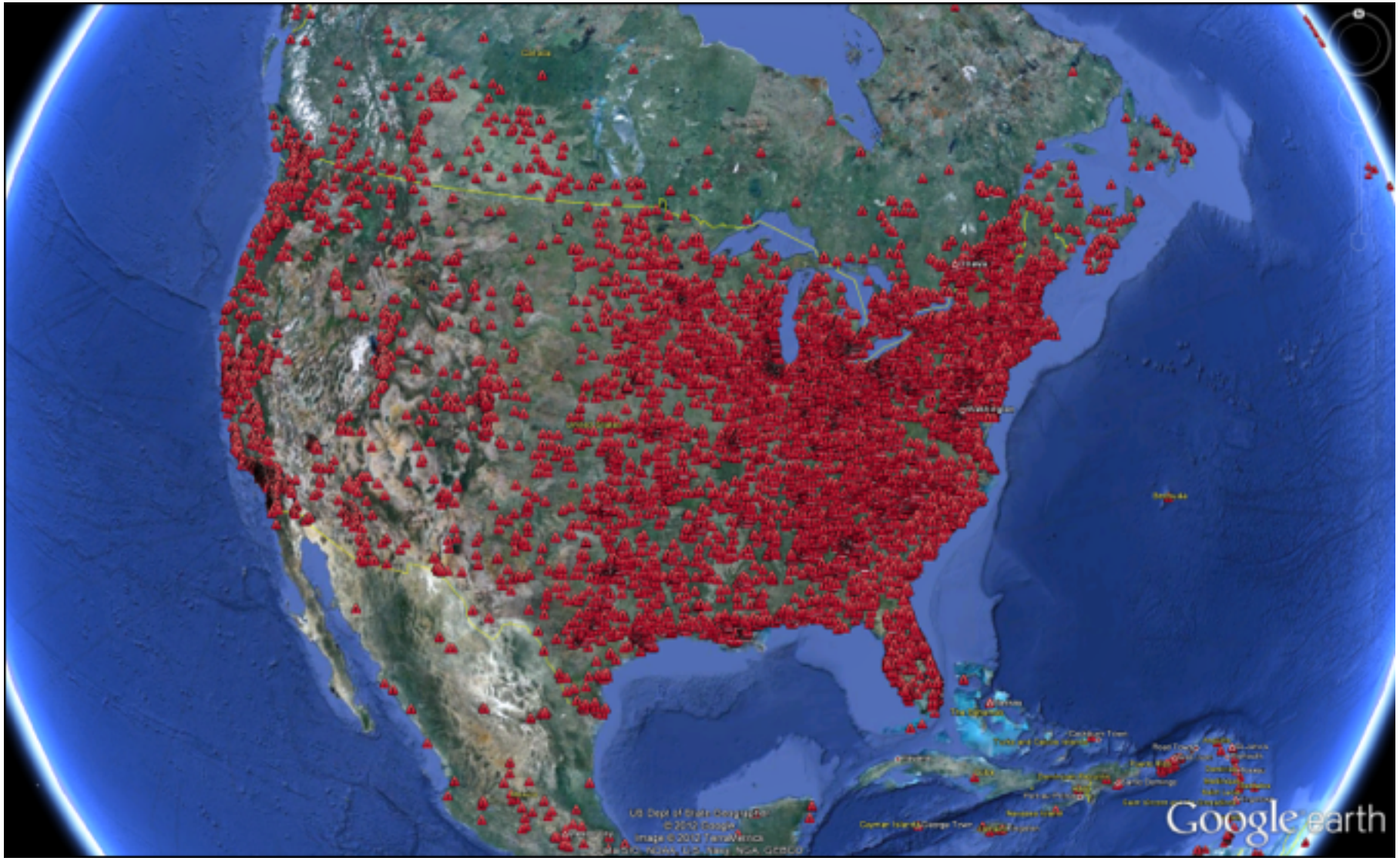
IN,"18.975,72.8258"
TR,"41.0186,28.9647"
US,"41.0399,-81.4802"
TR,"37.9158,40.2189"
VE,"10.5,-66.9167"
US,"39.1111,-94.6904"
RO,"44.1167,24.35"
RO,"44.1167,24.35"
RO,"46.5667,26.9"
JP,"35.685,139.7514"
BR,"-21.1333,-48.9667"
IN,"15.15,76.9333"
CA,"45.6333,-72.9333"
IT,"44.8,10.3333"
US,"43.2166,-78.0584"
CO,"4.6492,-74.0628"
RO,"46.35,25.8"
US,"40.1083,-83.144"
PL,"50.0783,19.2253"
CA,"45.3833,-72.7333"
SE,"59.5167,17.25"
US,"33.6813,-116.9593"
US,"33.9129,-118.3439"
CA,"45.1833,-73.4"

Careful How Data is Parsed

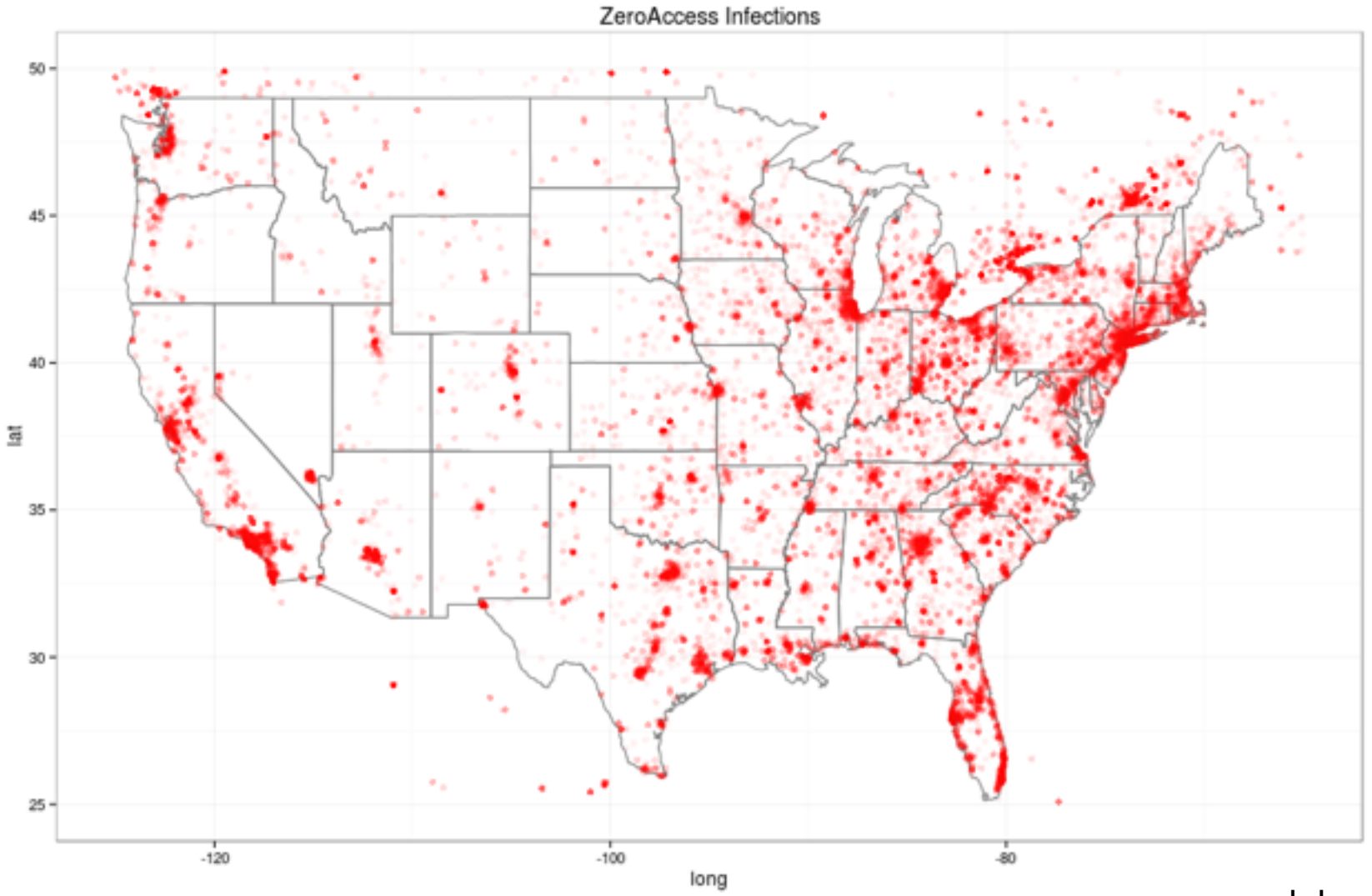


Rank	Location	Bot.count
1	Potwin, KS	800
2	Houston, TX	239
3	Los Angeles, CA	224
4	Brooklyn, NY	179
5	Miami, FL	169
6	Las Vegas, NV	168
7	Phoenix, AZ	141
8	Chicago, IL	116
9	Bronx, NY	115
10	New York, NY	110

Google Maps

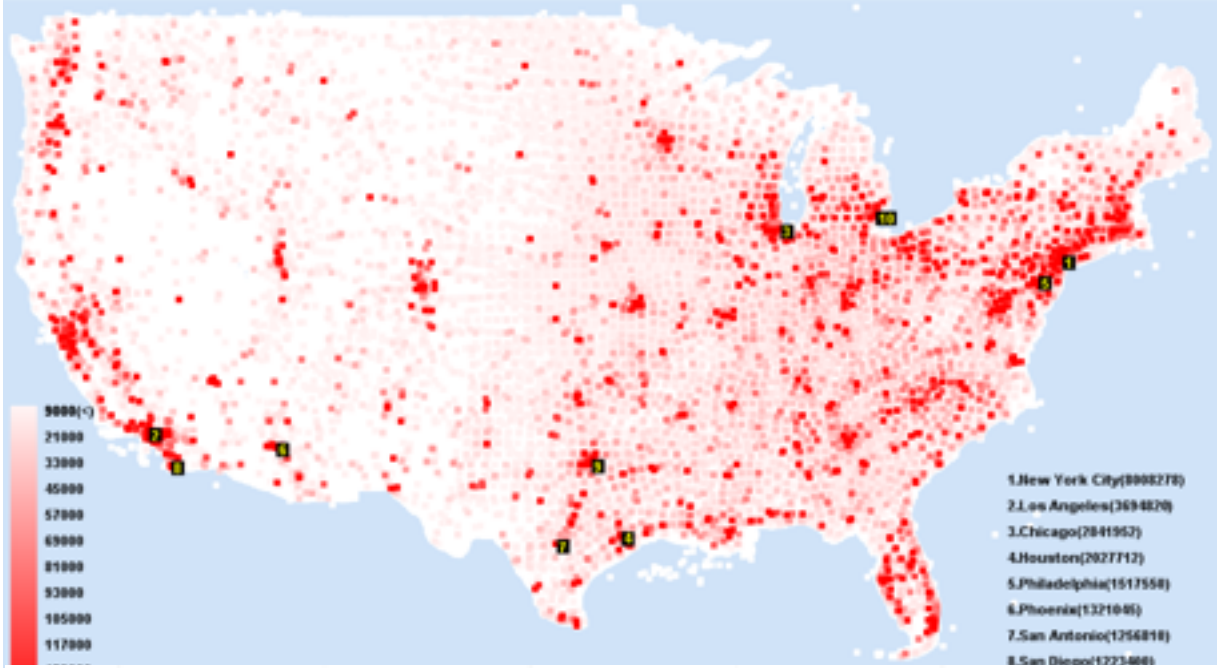


Not Google Maps

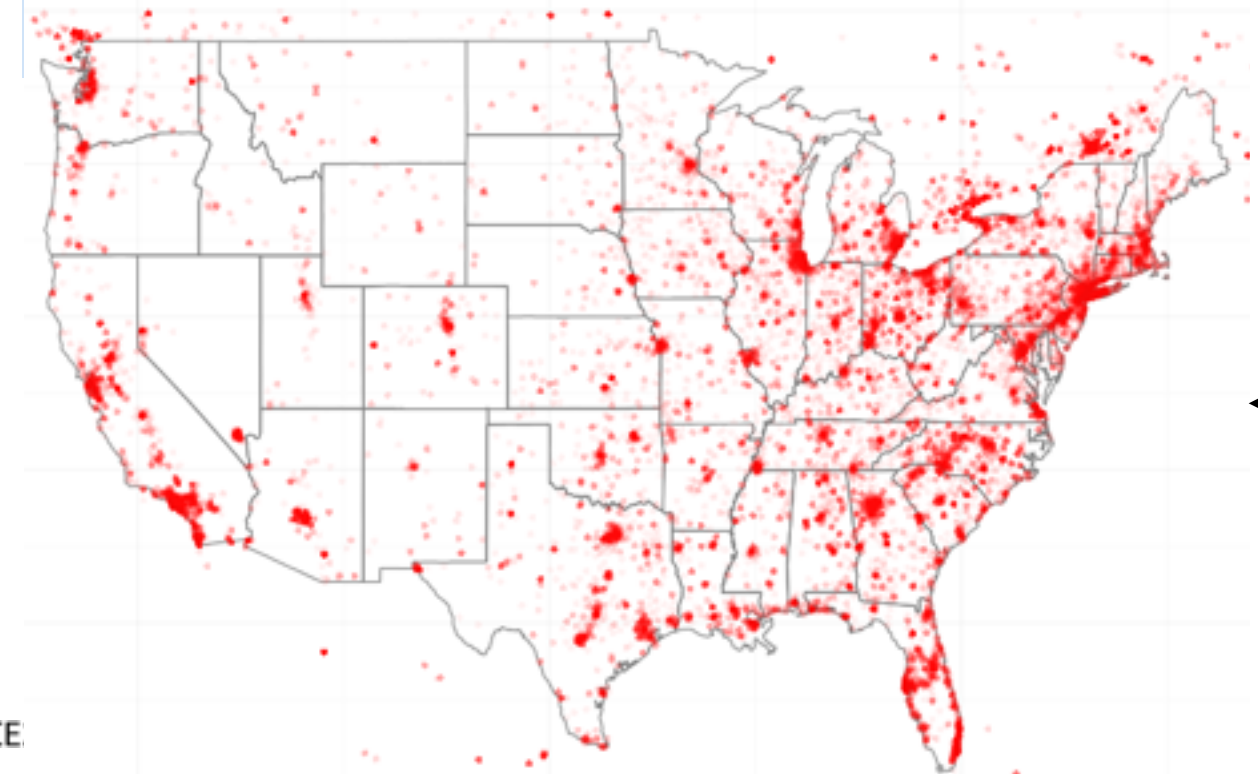


alpha = 1/33

The Story

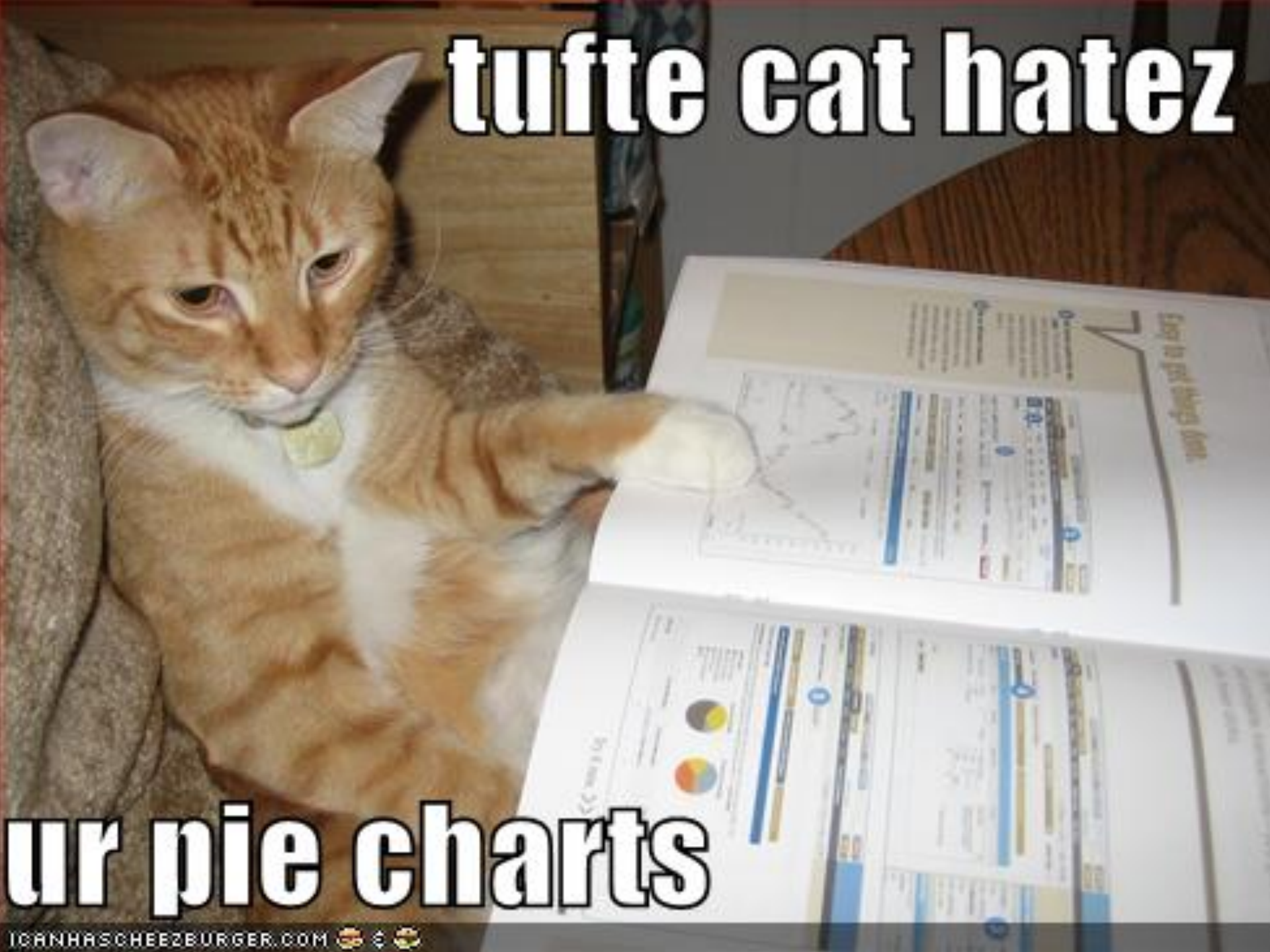


Population



Our Data

and to wrap things up....


tufte cat hatez
ur pie charts
ICANHASCHEEZBURGER.COM

Key Learning Points

- data helps our **understanding of our environment**
- solutions are more from **thinking than buying**
- visualizations help **communicate complexity quickly**
- data visualization is **not a natural skill**, it must be learned
- be truthful: **message should match the data**
- simple tools can be, **data scientist you need not be**

Bob Rudis
@hrbrmstr

Jay Jacobs
@jayjacobs

Security in knowledge