

O'REILLY®

Security

BUILD BETTER DEFENSES

oreillysecuritycon.com

#oreillysecurity

Foundations of Security Data Science Exploratory Data Analysis

Jay Jacobs
Charles Givre
Bob Rudis

Communication

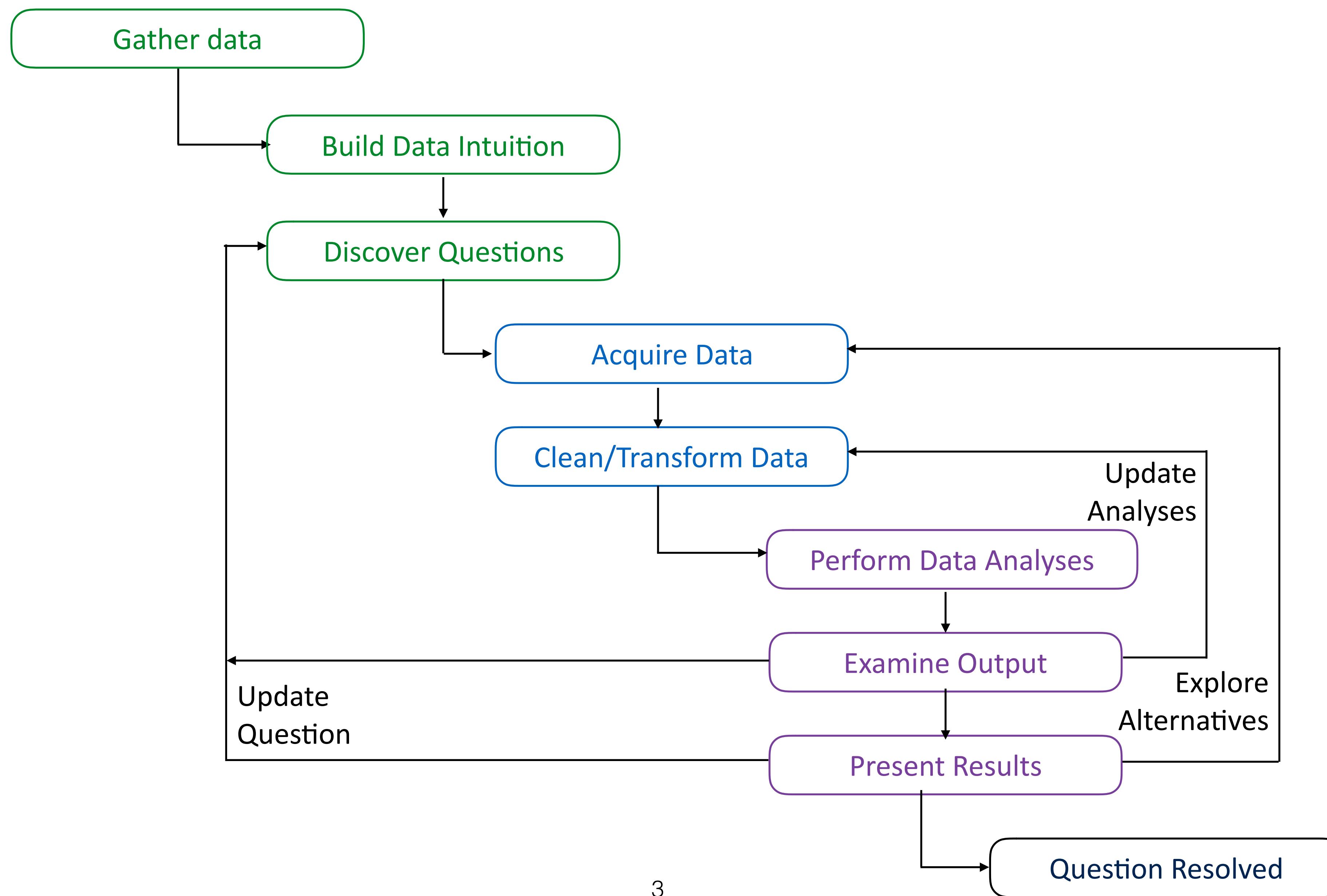
“The true promise of the information age isn’t tons of data but decisions and actions that are better because they’re based on an understanding of what’s really going on in the world.

Any knowledge you gain that could be used to make better decisions will amount to nothing if you can’t communicate it to others in a way that makes sense to them.

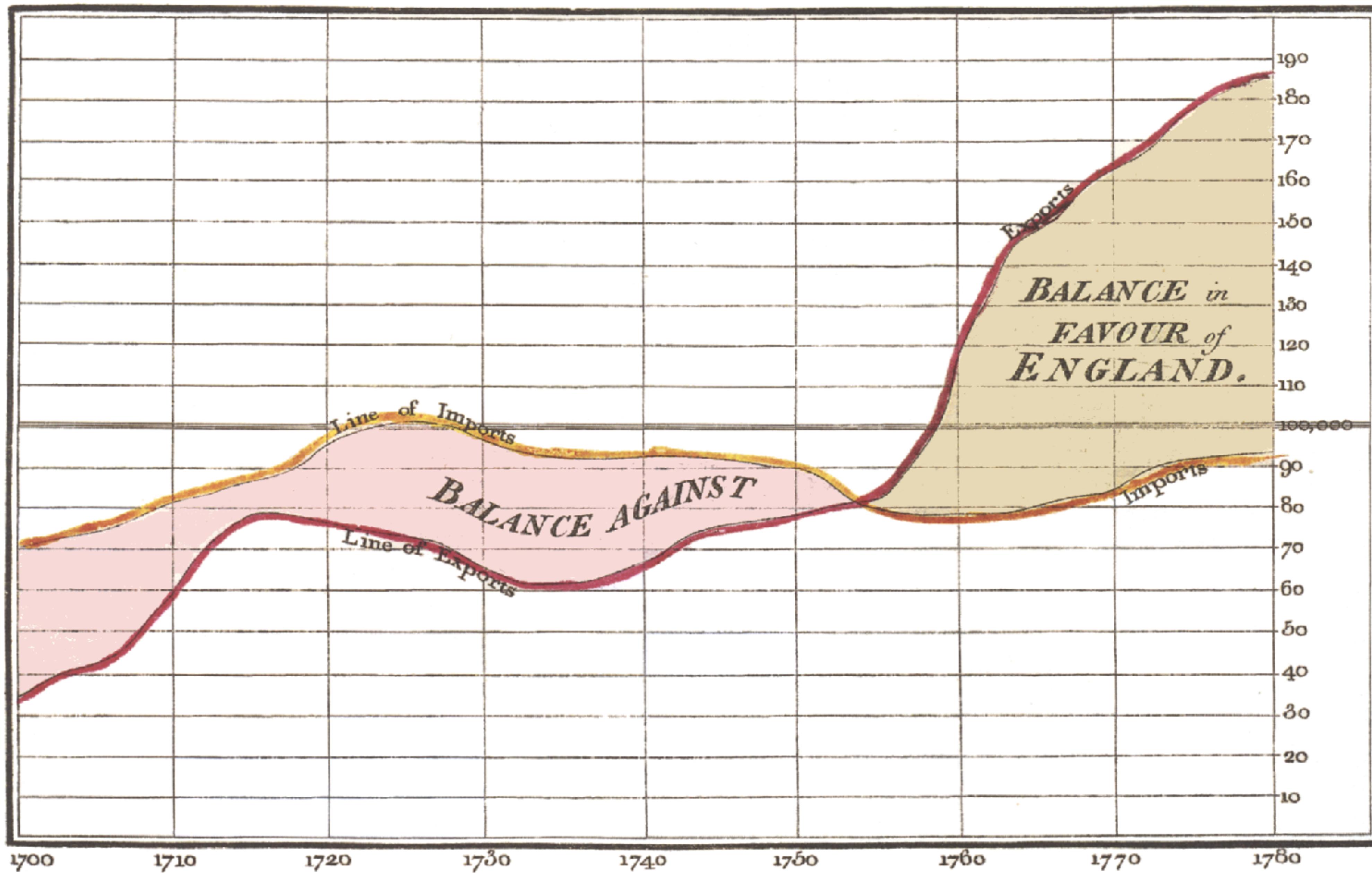
*Stephen Few
“Show me the Numbers”*



Reality of Research Workflow



Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.



The Bottom line is divided into Years, the Right hand line into £10,000 each.

Napoleon's March to Moscow, The War of 1812

Carte Figurative des pertes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite
Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. Chiers, de Séguir, de Fezensac, de Chambray et le journal médical de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk en Malibow et qui rejoignirent Orscha en Wilésk, avaient toujours marché avec l'armée.

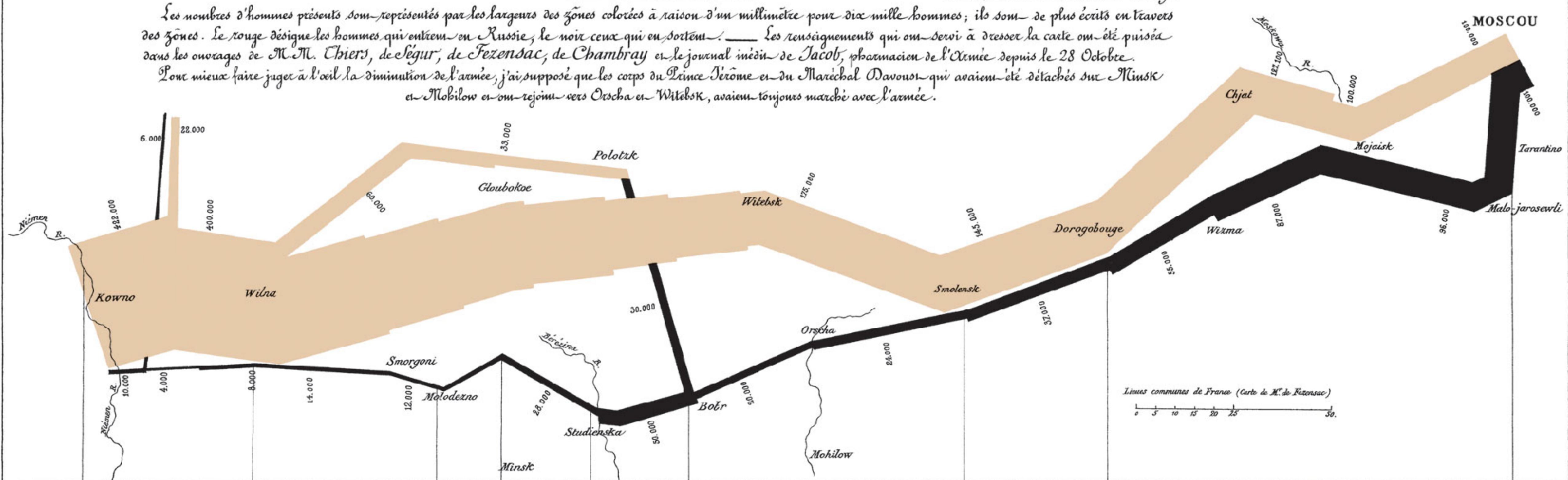


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les Cosaques passent au galop
le Niemen gelé.

-26° le 2 X^{bre}

-30° le 6 X^{bre}
-24° le 1^{er} X^{bre}
-20° le 28 9^{bre}
-11°

-21° le 14 9^{bre}
-9° le 9 9^{bre}

Pluie 24 8^{bre}

Zéro le 18 8^{bre}
5
10
15
20
25
30 degrés

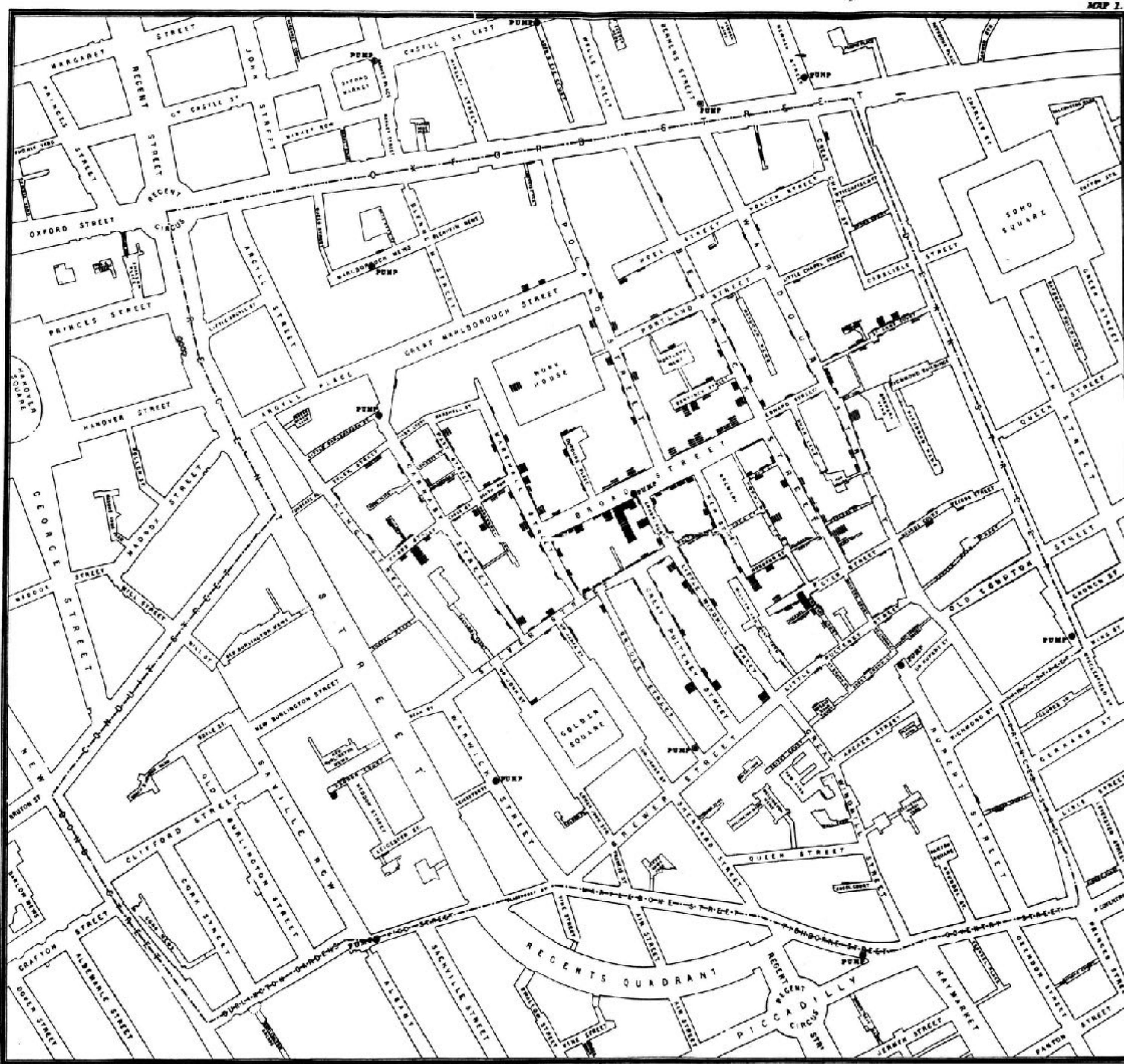
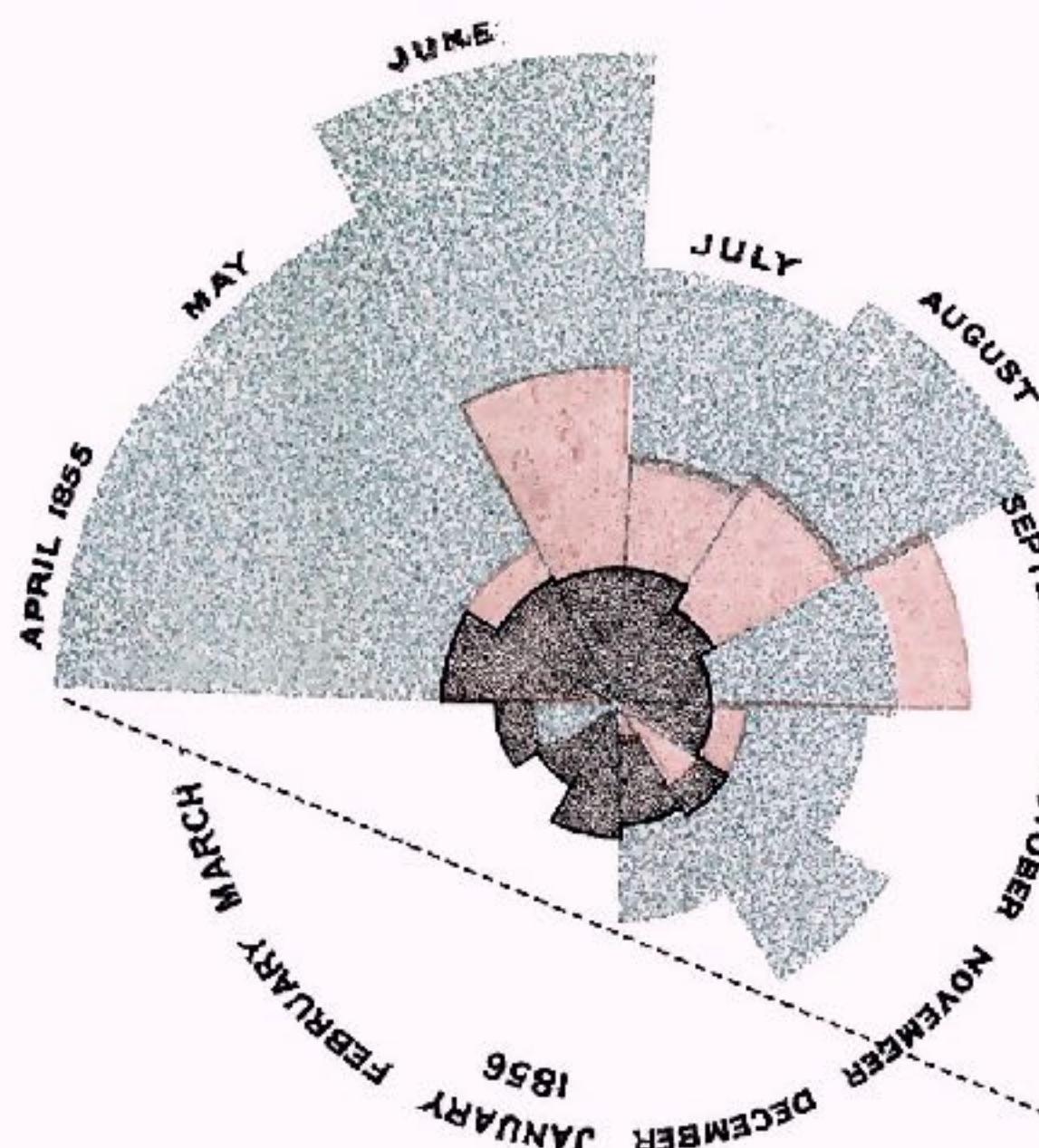


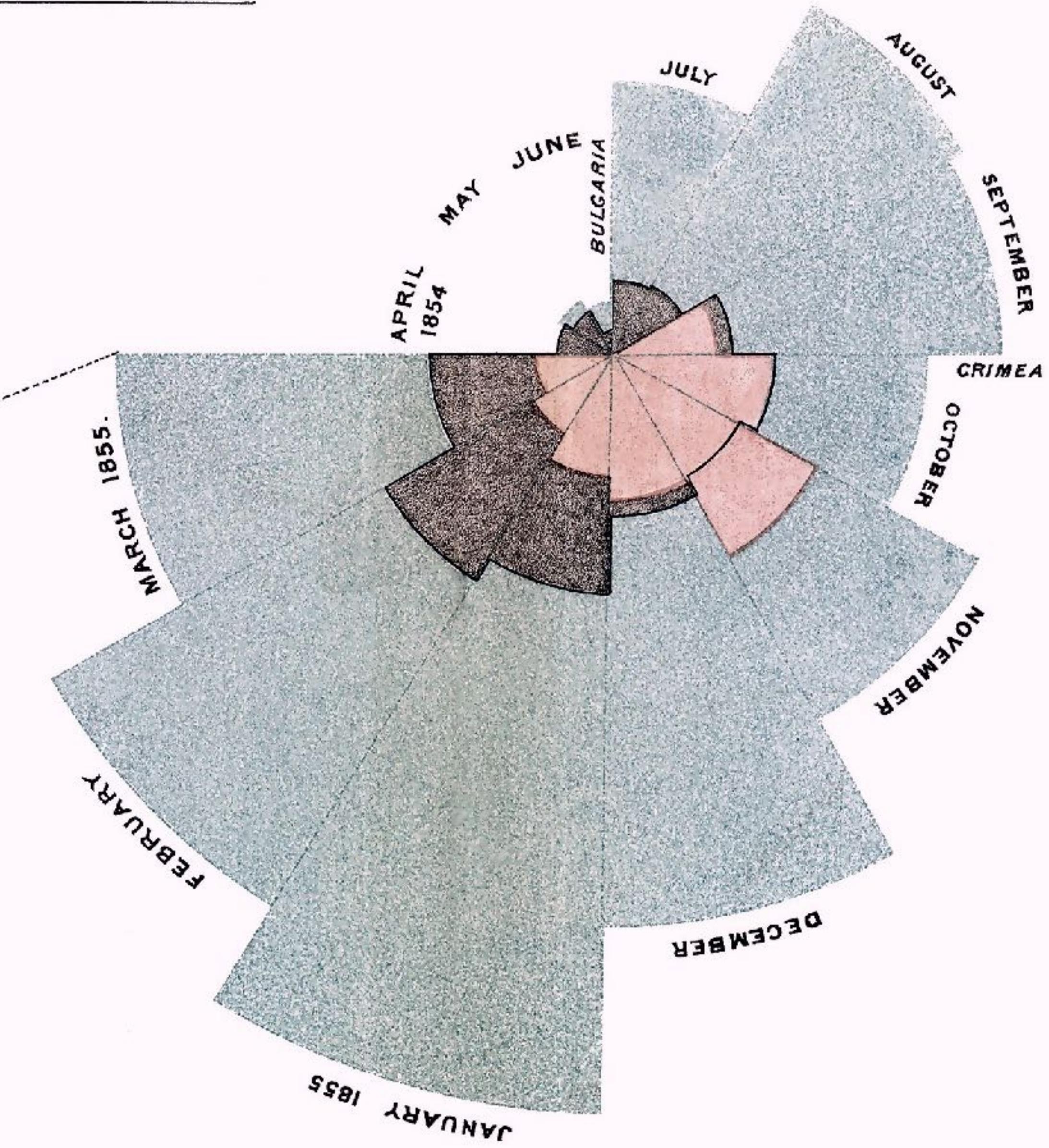


DIAGRAM OF THE CAUSES OF MORTALITY
IN THE ARMY IN THE EAST.

2.
APRIL 1855 TO MARCH 1856.



1.
APRIL 1854 TO MARCH 1855.



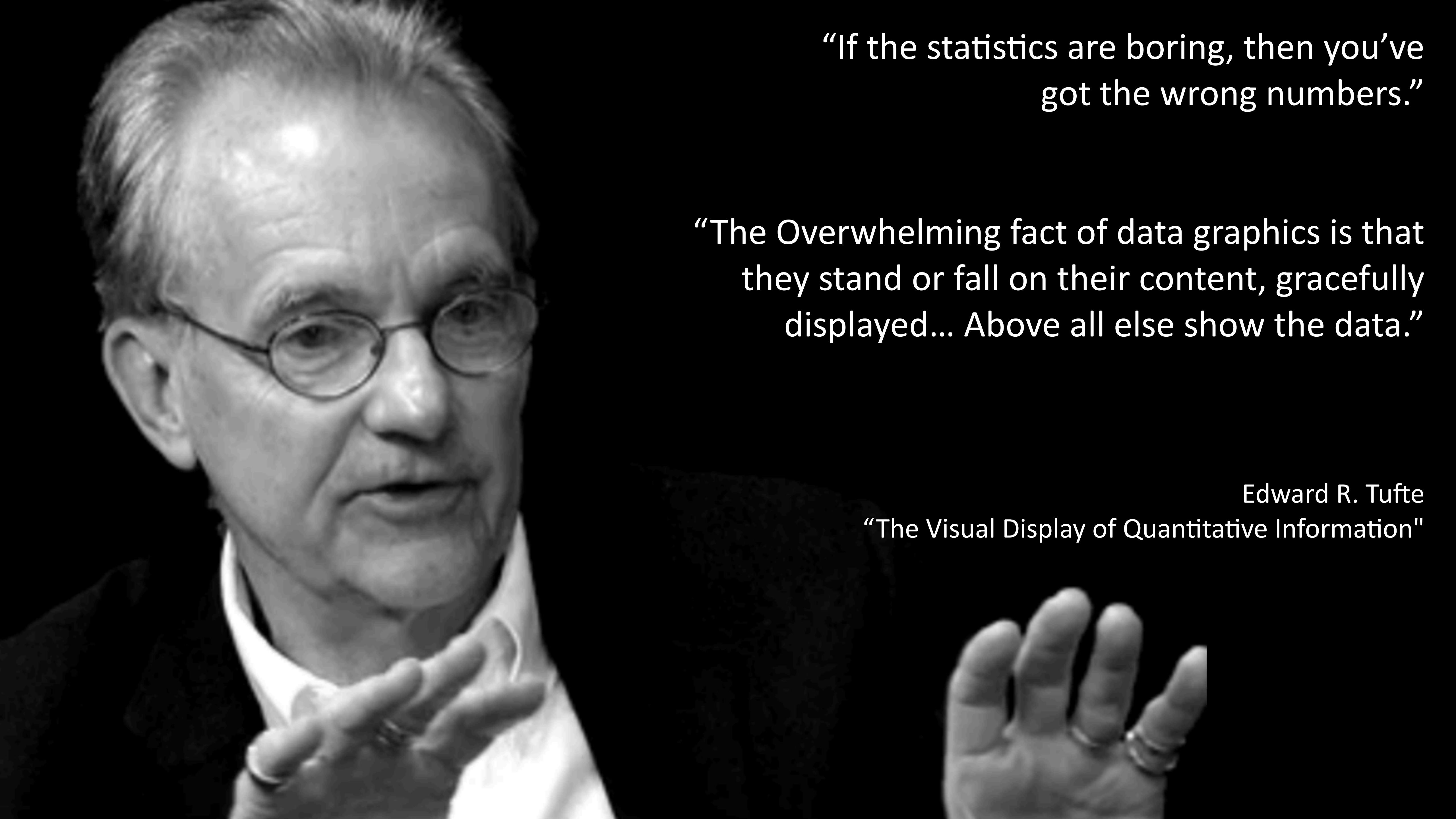
The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases; the red wedges measured from the centre the deaths from wounds; & the black wedges measured from the centre the deaths from all other causes.

The black line across the red triangle in Nov^r 1854 marks the boundary of the deaths from all other causes during the month.

In October 1854, & April 1855, the black area coincides with the red; in January & February 1856, the blue coincides with the black.

The entire areas may be compared by following the blue, the red & the black lines enclosing them.



“If the statistics are boring, then you’ve got the wrong numbers.”

“The Overwhelming fact of data graphics is that they stand or fall on their content, gracefully displayed... Above all else show the data.”

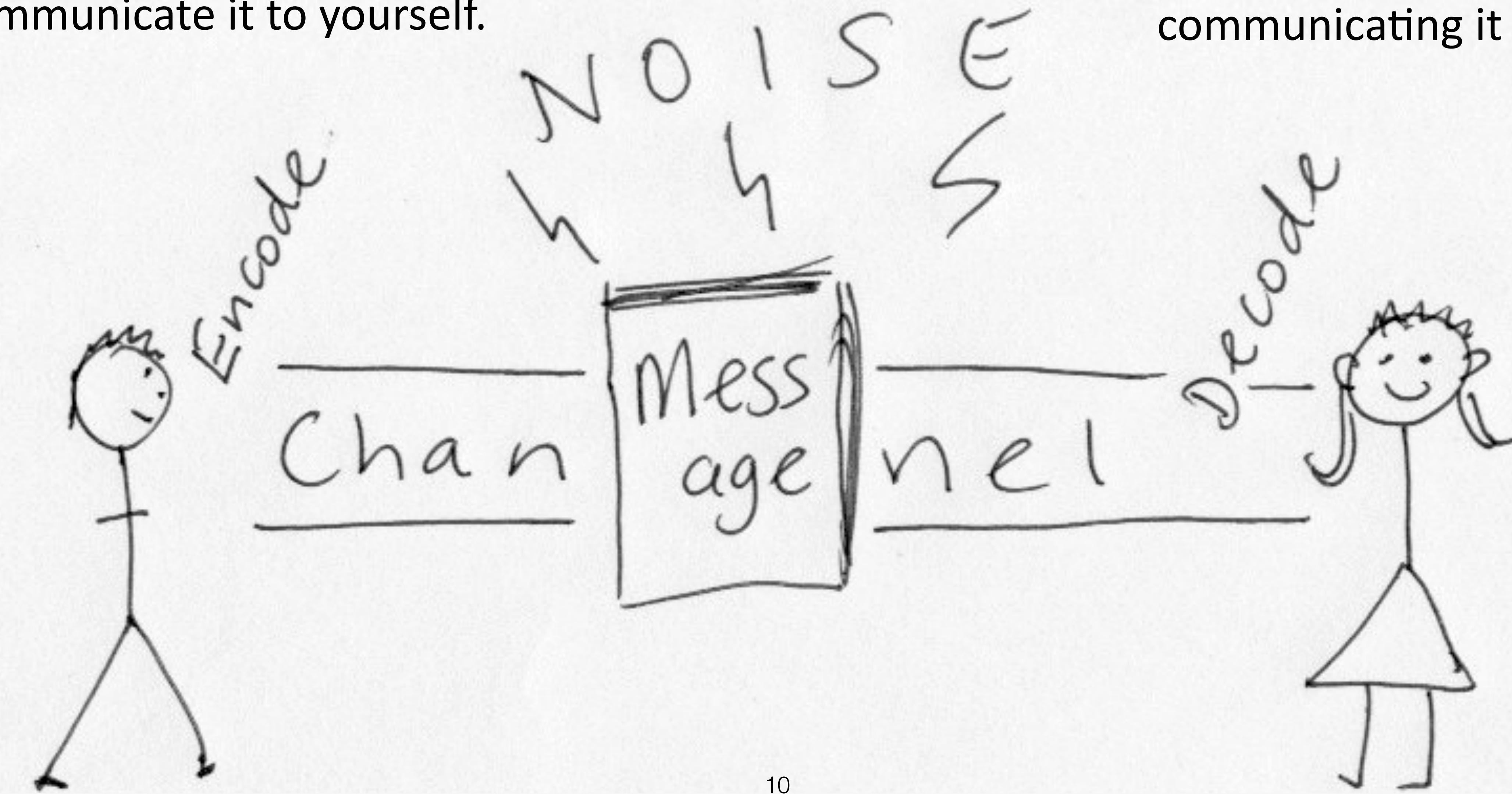
Edward R. Tufte

“The Visual Display of Quantitative Information”

Visualization is Communication

For yourself: To see into the data
and communicate it to yourself.

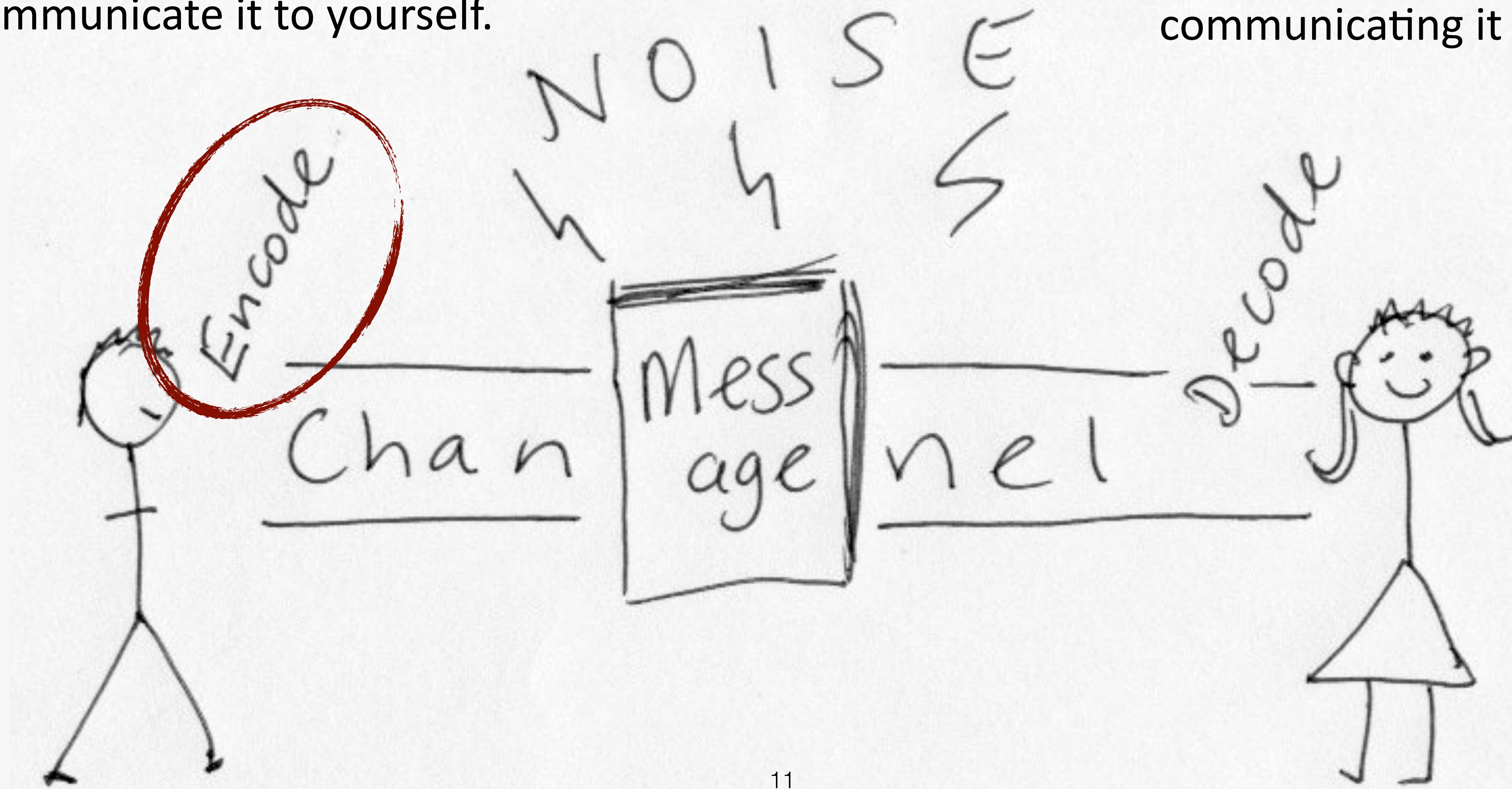
For Others: After finding a story,
communicating it to others.



Visualization is Communication

For yourself: To see into the data
and communicate it to yourself.

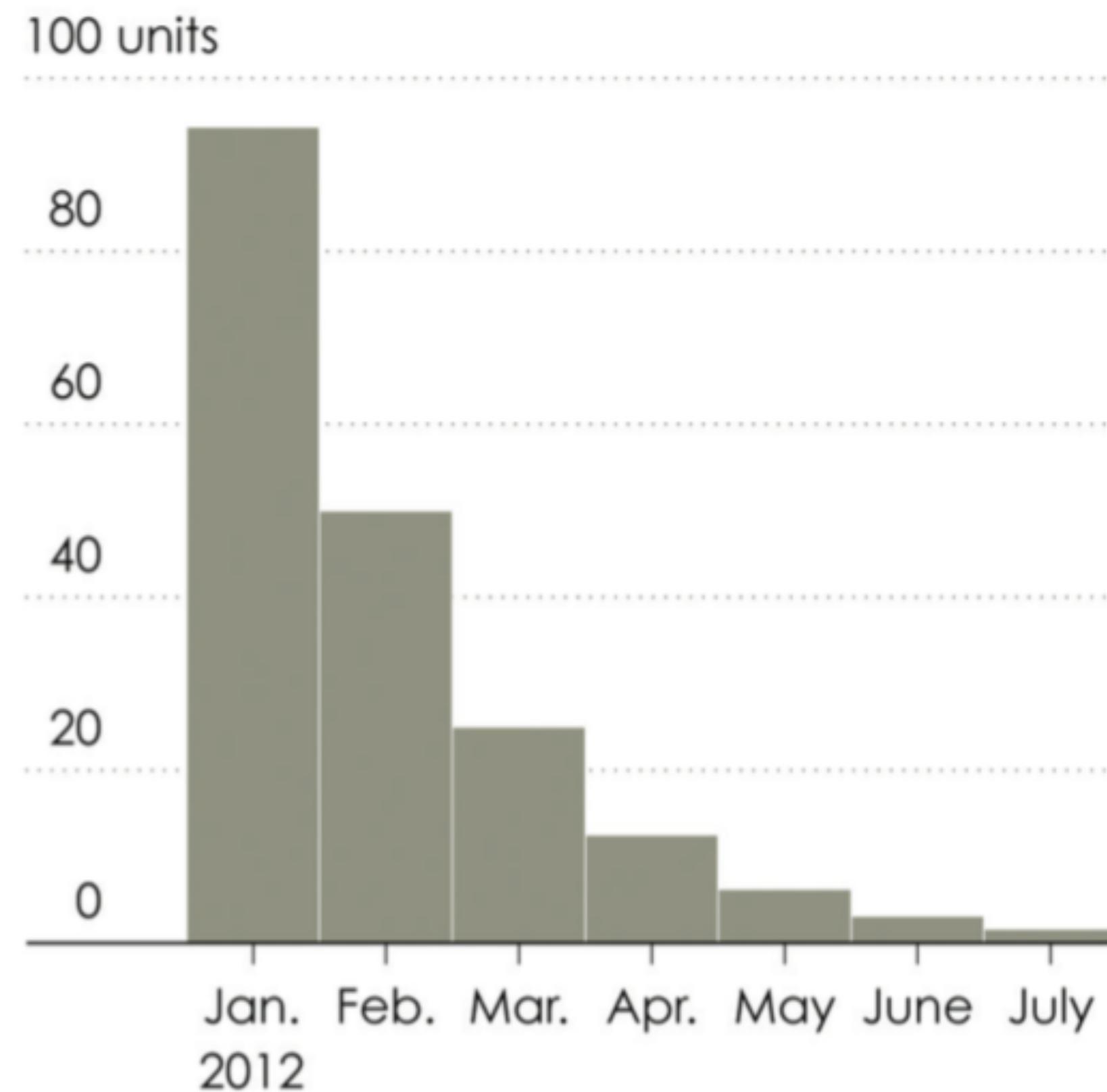
For Others: After finding a story,
communicating it to others.



Building Blocks

Title of this Graph

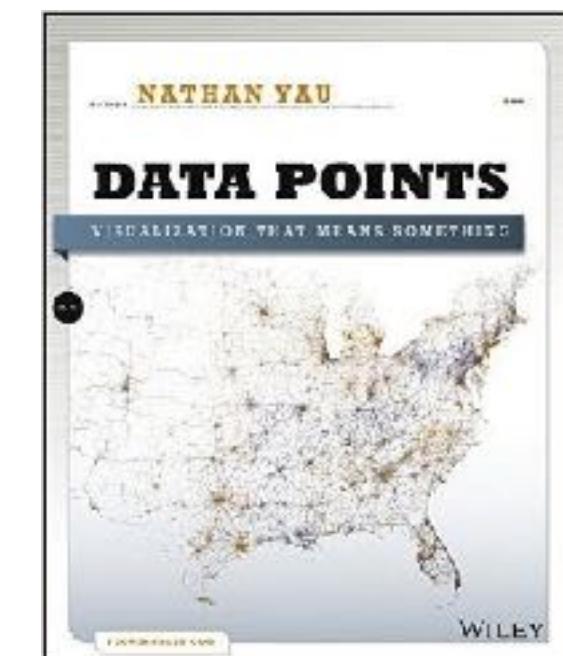
A description of the data or something worth highlighting to set the stage.



Source: Somewhere reputable

Visualizations are made up of multiple pieces working together.

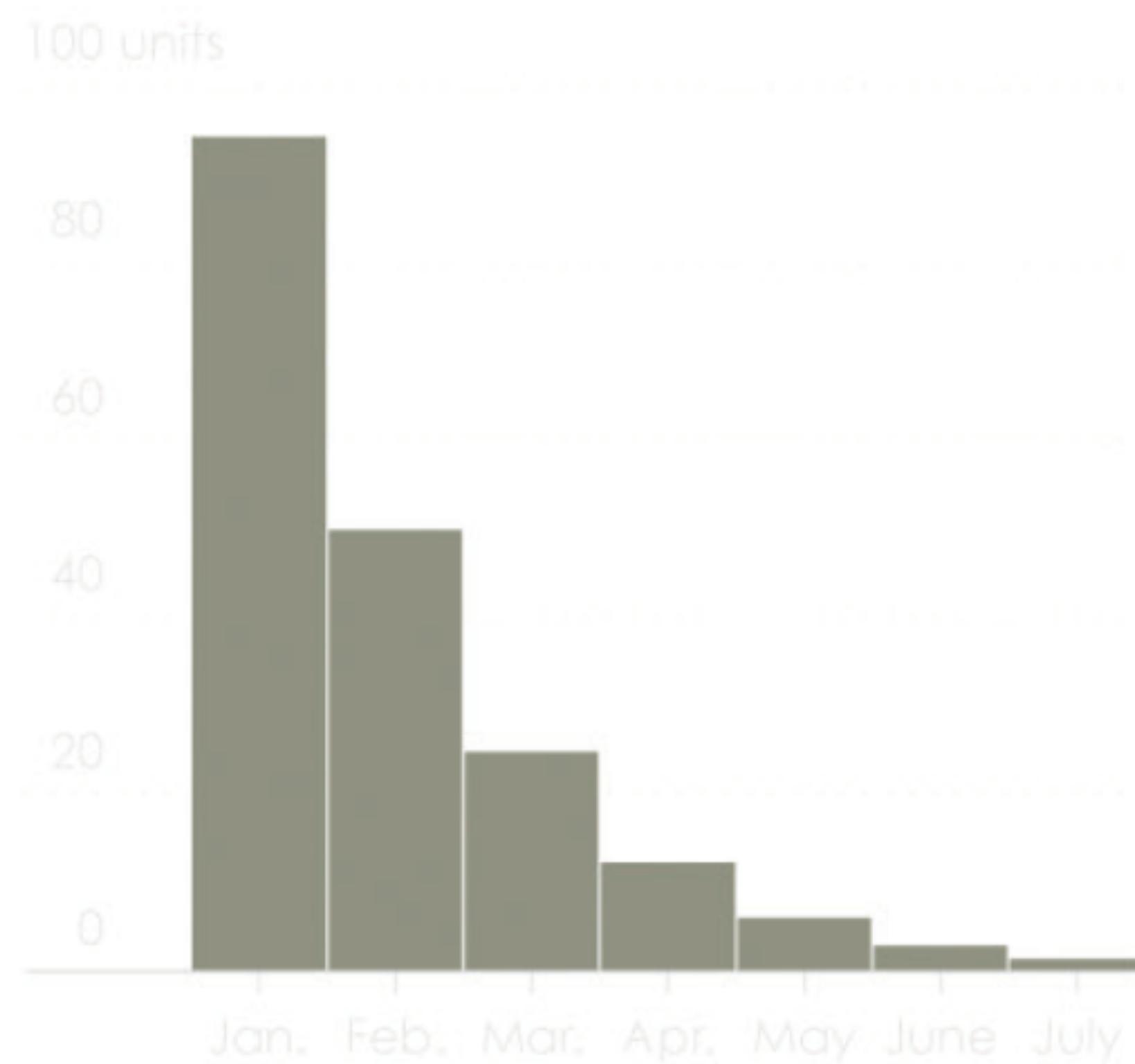
From Nathan Yau



Visual Cues

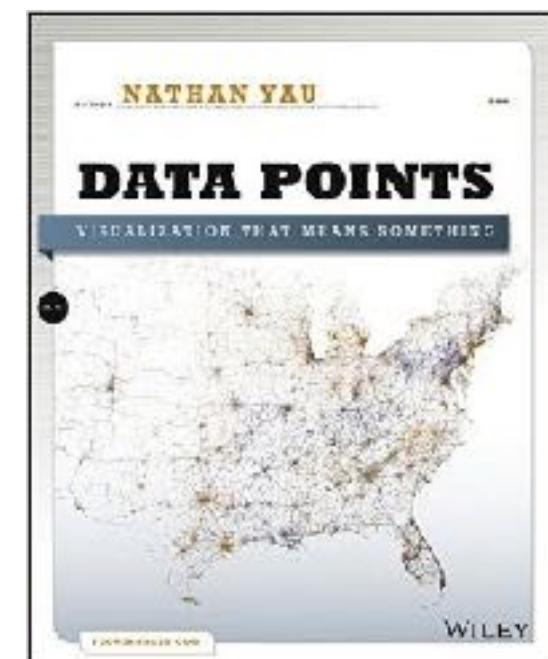
Title of this Graph

A description of the data or something worth highlighting to set the stage.

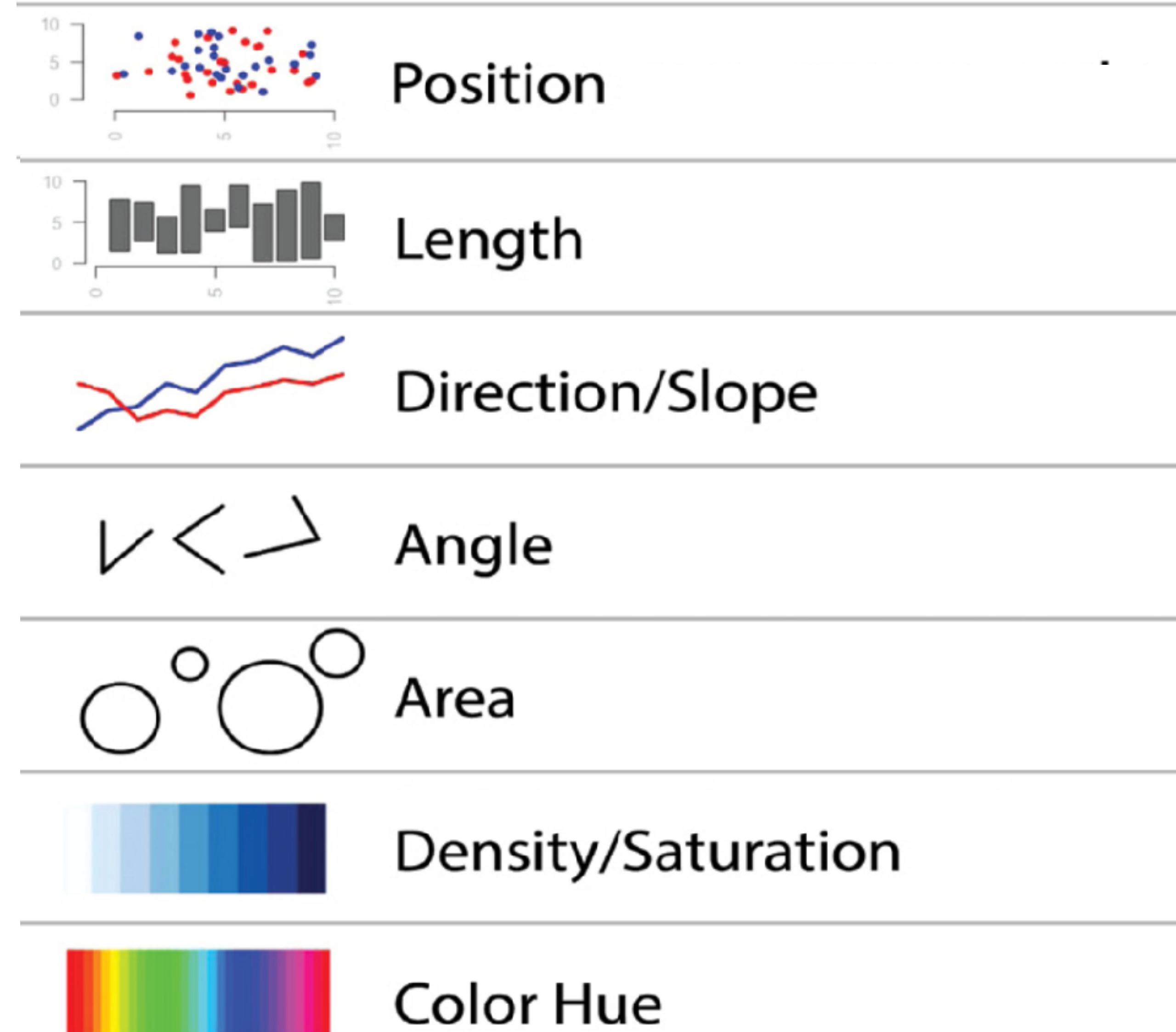


“Visualization involves encoding data with shapes, colors, and sizes. Which cues you choose depends on your data and your goals.”

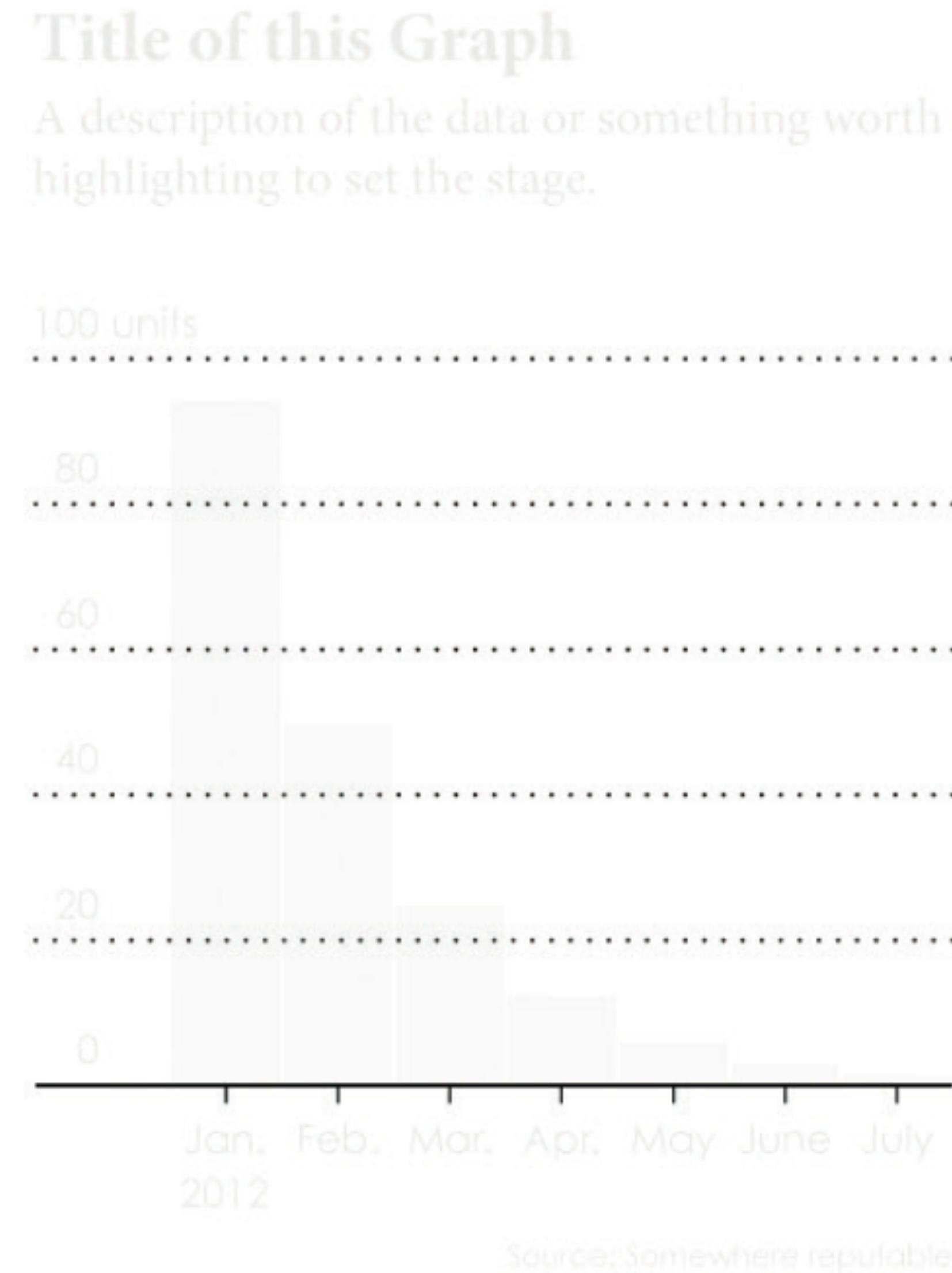
From Nathan Yau



Visual Cues

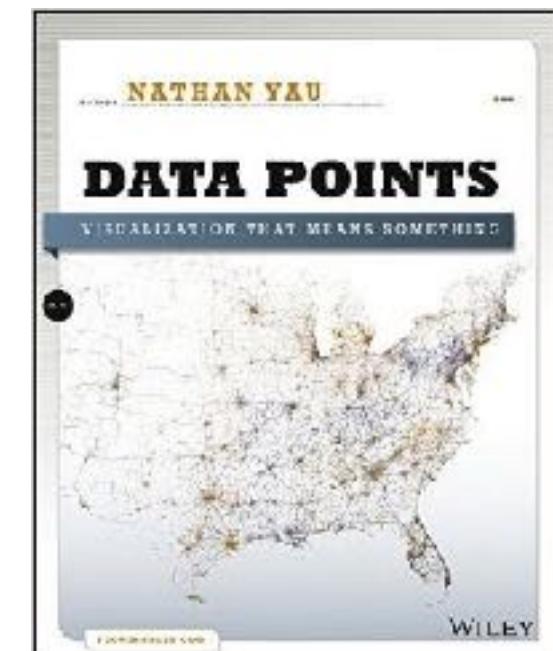


Coordinate System



Scatter plot versus pie chart, map projections use a variety of different coordinate systems.

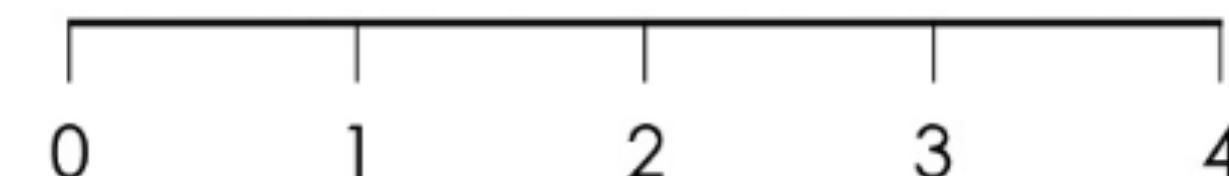
From Nathan Yau



Scales

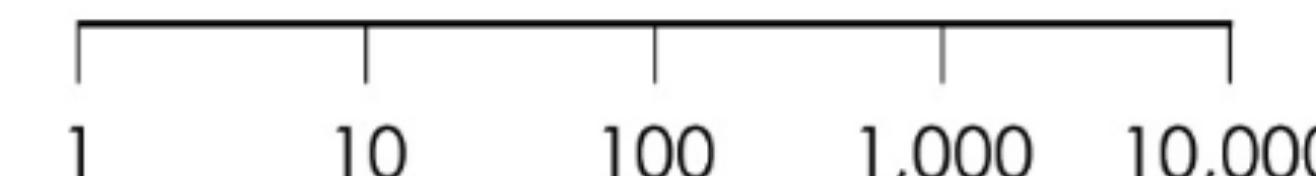
Linear

Values are evenly spaced



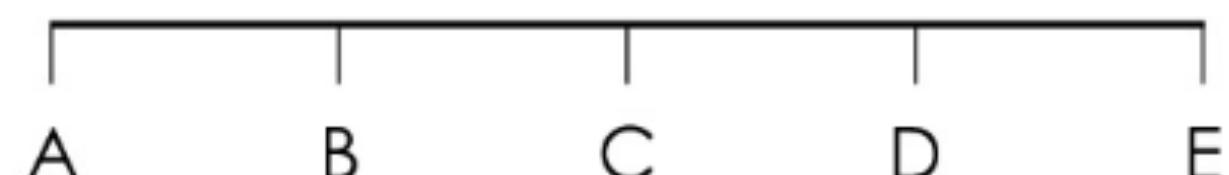
Logarithmic

Focus on percent change



Categorical

Discrete placement in bins



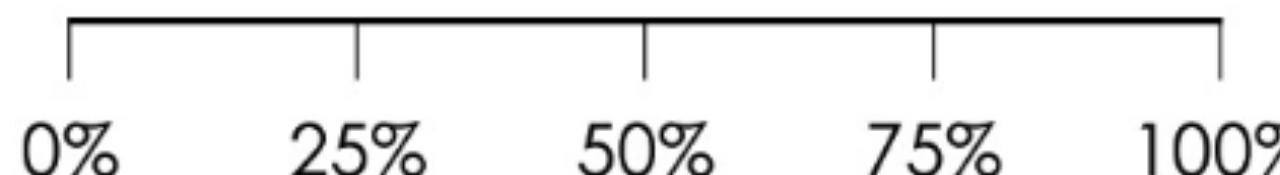
Ordinal

Categories where order matters



Percent

Representing parts of a whole



Time

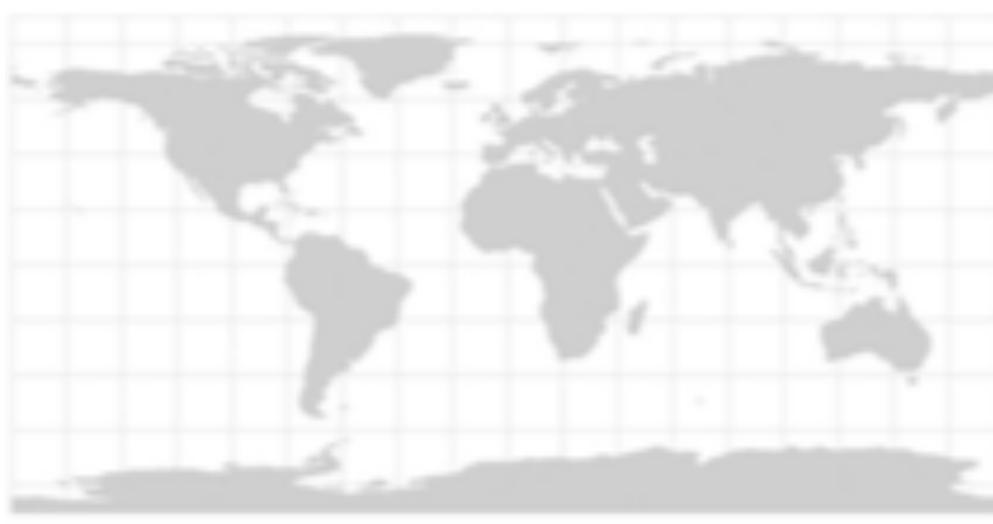
Units of months, days, or hours



Map Projections

Equirectangular

Typically used for thematic mapping, but doesn't preserve area or angle



Albers

Scale and shape not preserved; angle distortion is minimal



Mercator

Preserves angles and shapes in small areas, making it good for directions



Lambert conformal conic

Better for showing smaller areas and often used for aeronautical maps.



Sinusoidal

Preserves area; useful for areas near the prime meridian

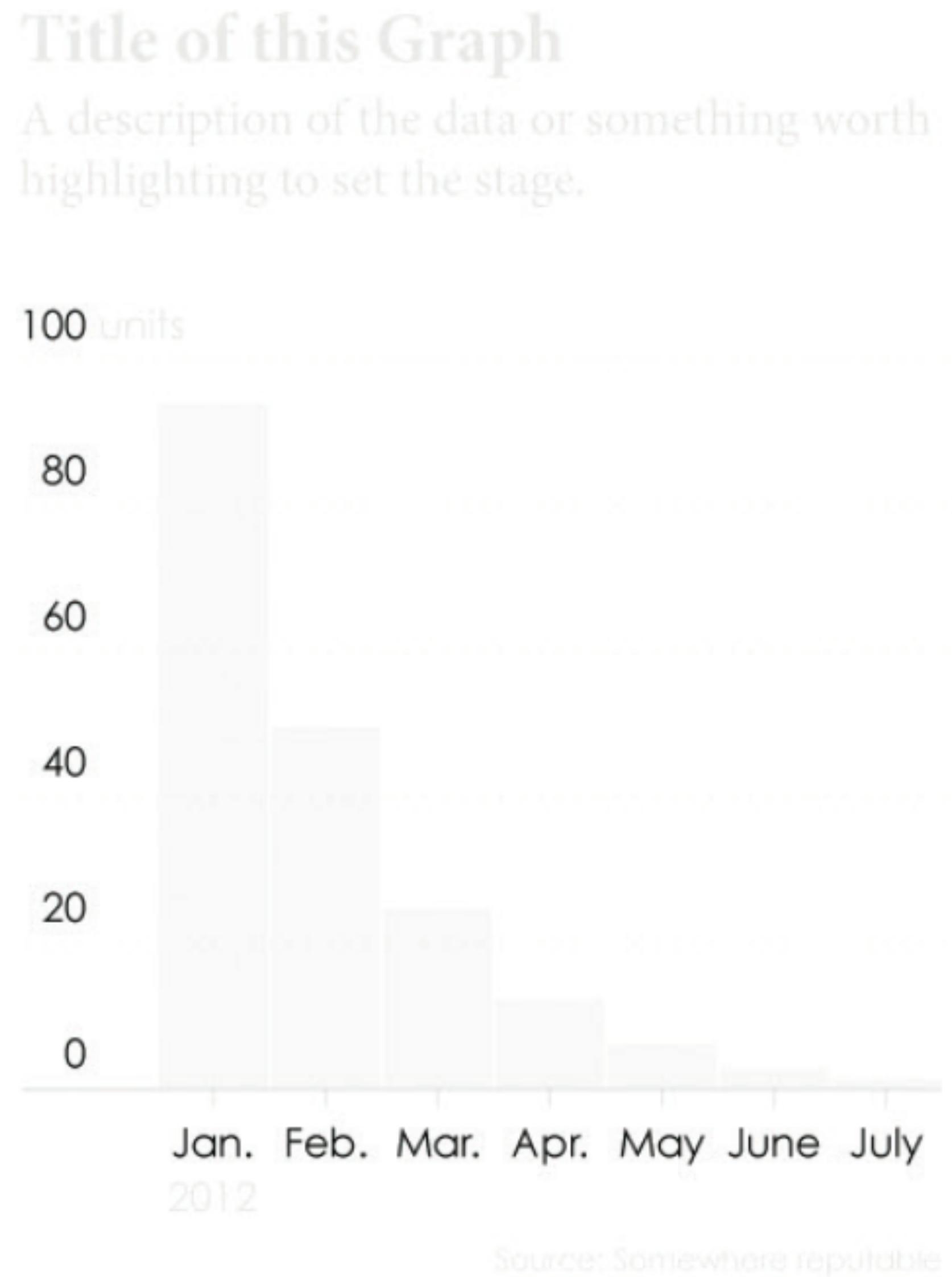


Polyconic

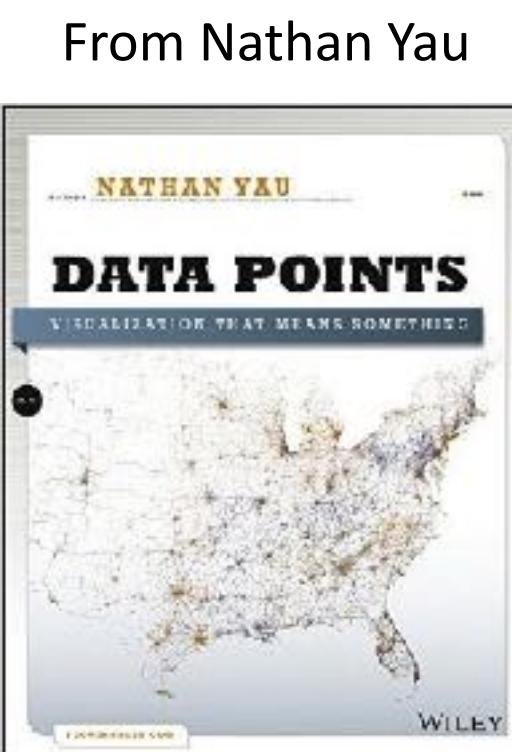
Was often used to show US in the mid-1900s; little distortions in small areas near meridian



Scale



Linear, logarithmic,
categorical, time, spatial,
and so on.



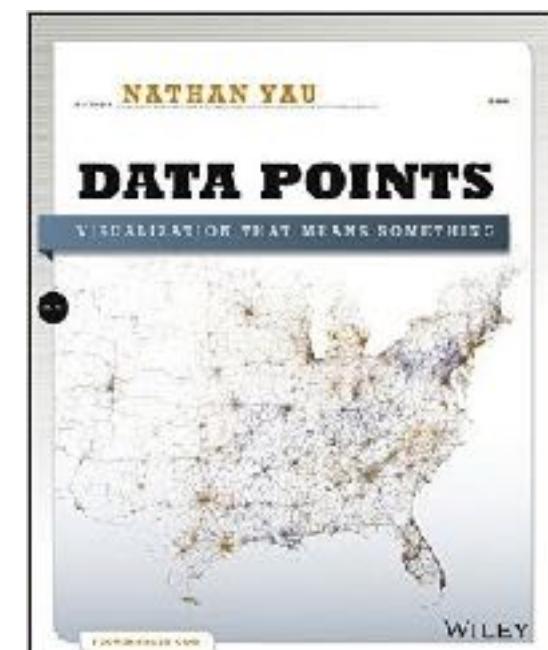
Title of this Graph

A description of the data or something worth highlighting to set the stage.



Title, subtitle, caption, clarifications of what the data represents all help the reader understand the visualization.

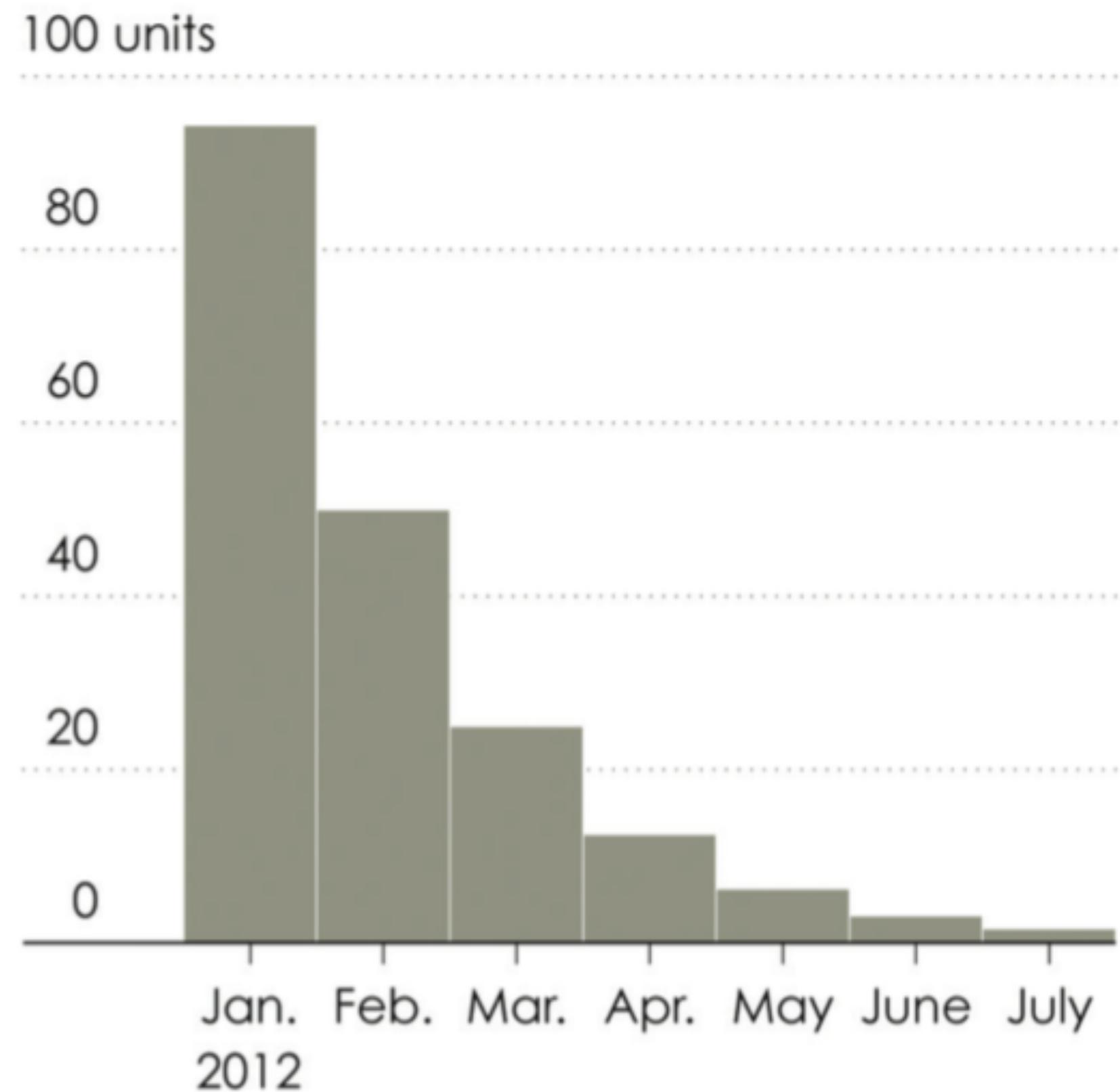
From Nathan Yau



Sum of the parts

Title of this Graph

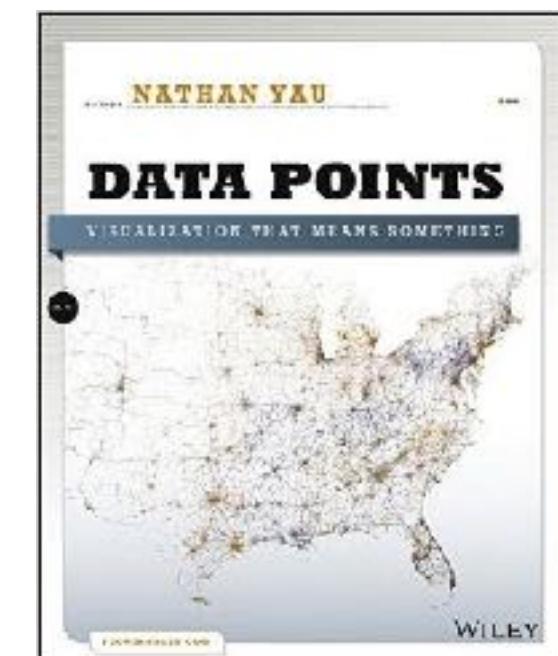
A description of the data or something worth highlighting to set the stage.



Source: Somewhere reputable

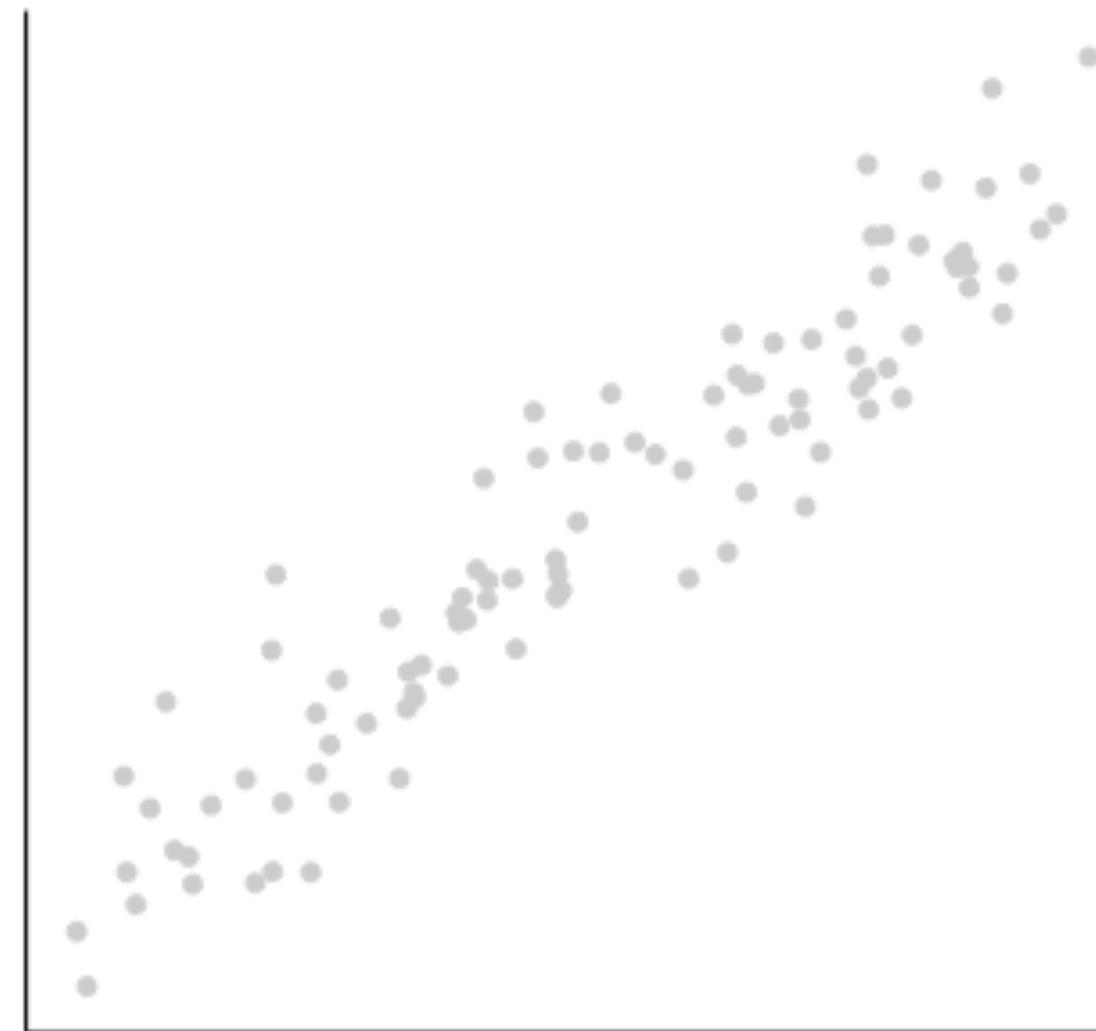
Everything comes together to (hopefully) create a coherent message for the reader.

From Nathan Yau

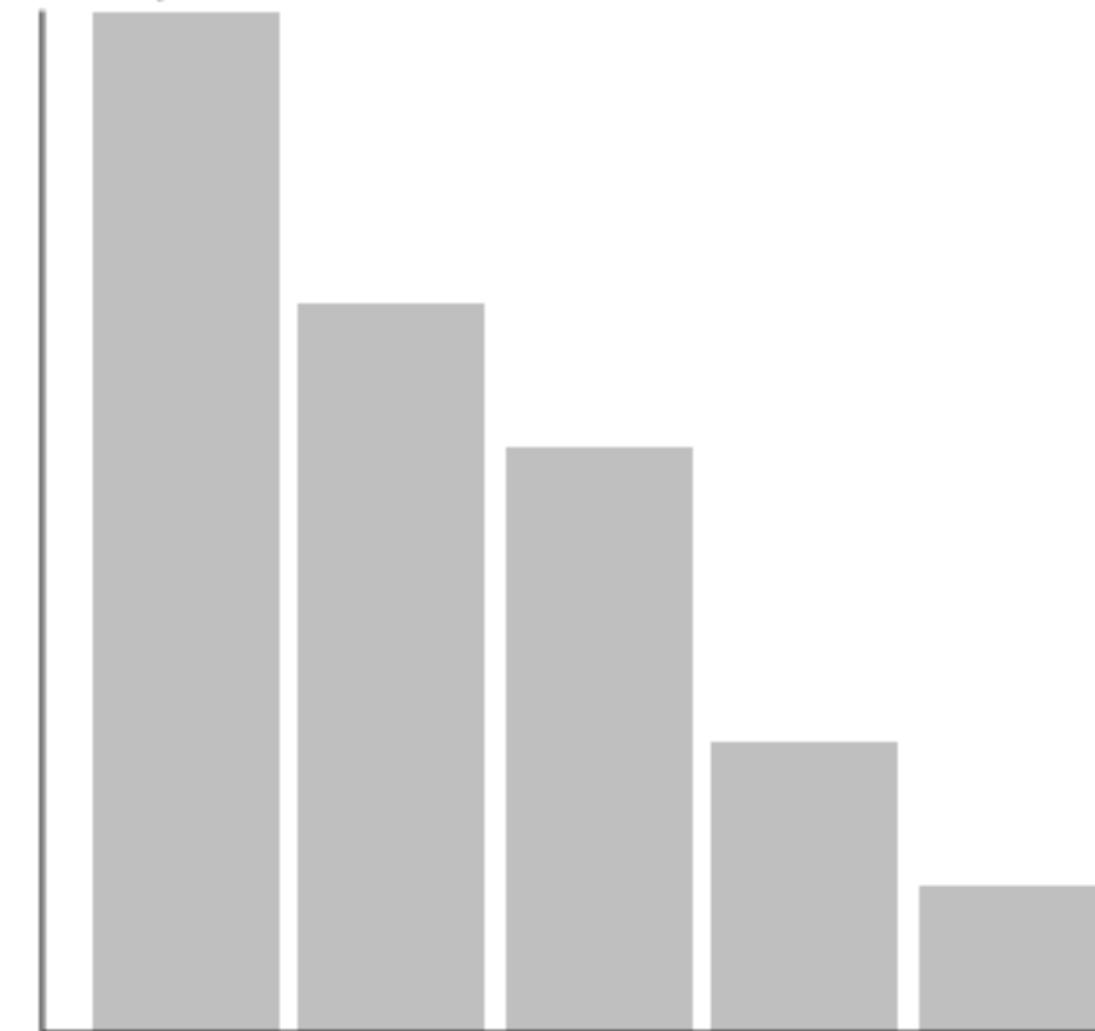


Common Visualizations

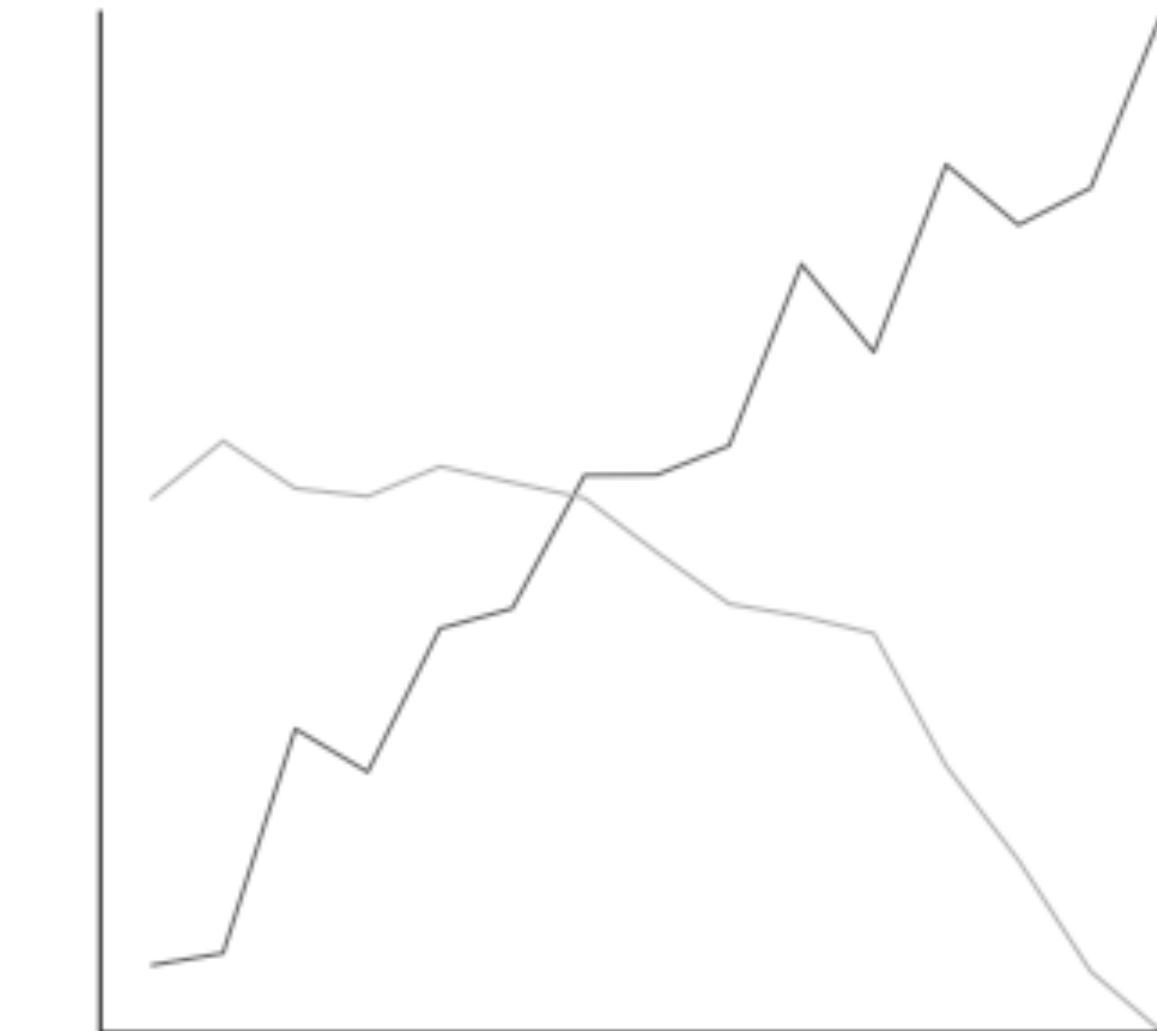
Scatter Plot



Bar plot



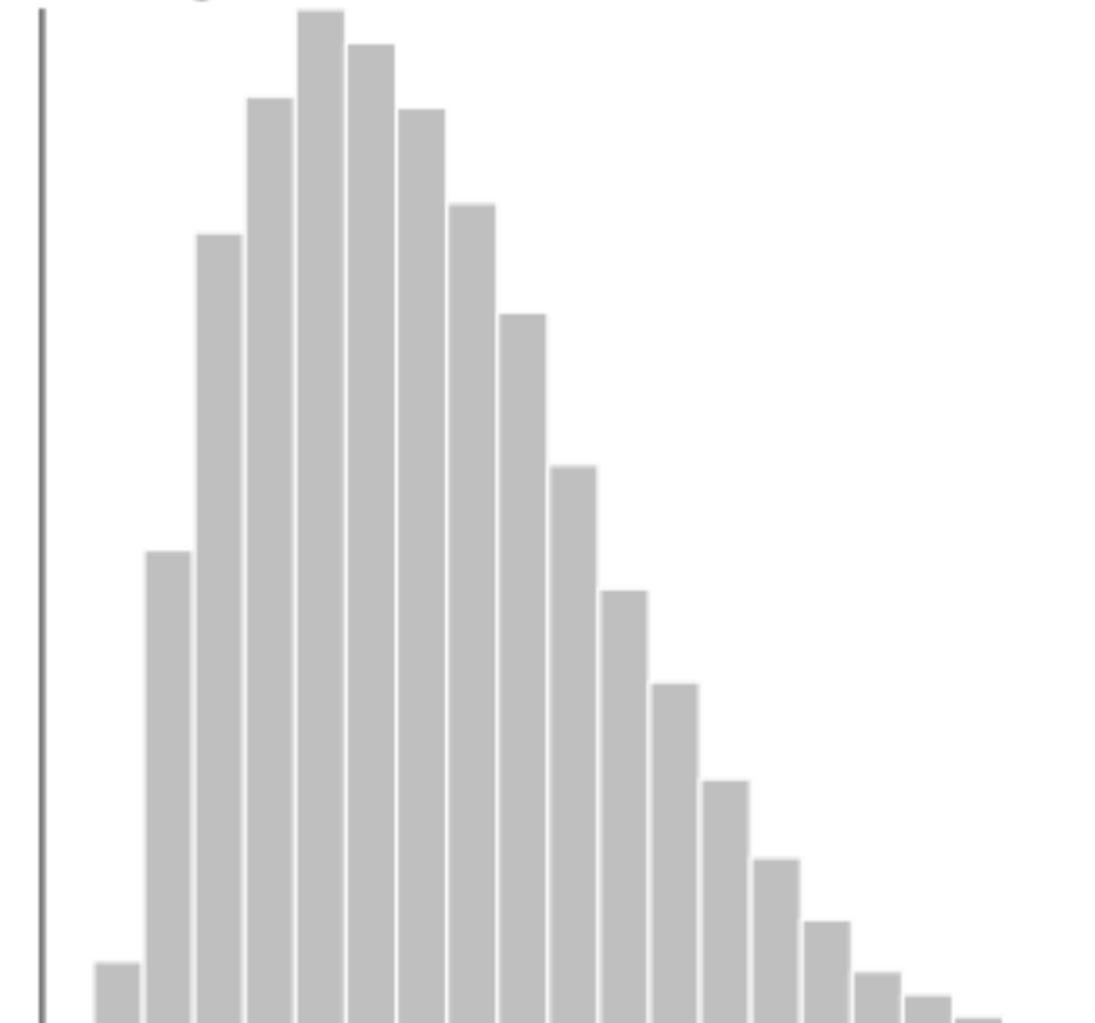
Line Plot



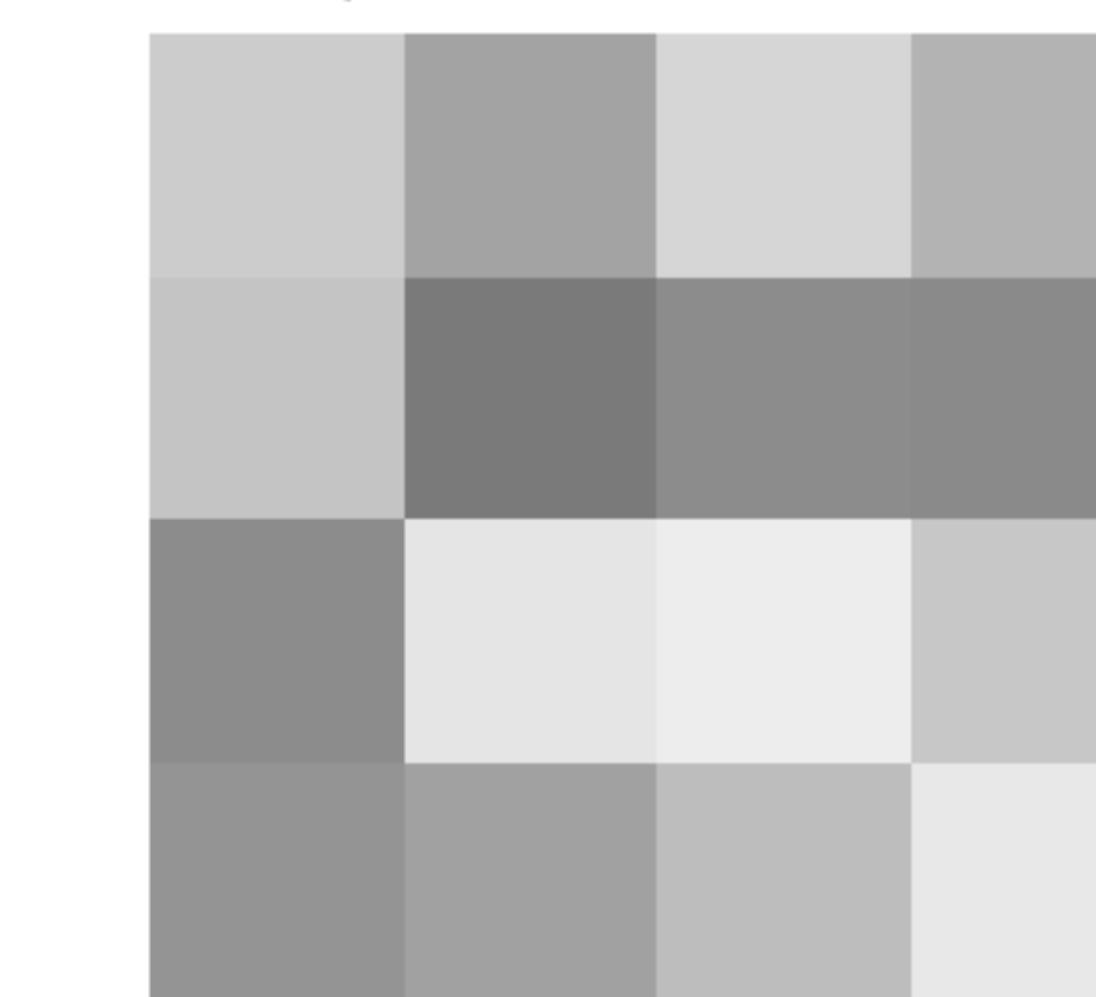
Pie Chart



Histogram



Heat Map



Visualization Building Blocks

Visual Cues

- Position, Length, Size, Shape, Color, Angle, Area

Coordinate System

- Cartesian, Polar, other projection

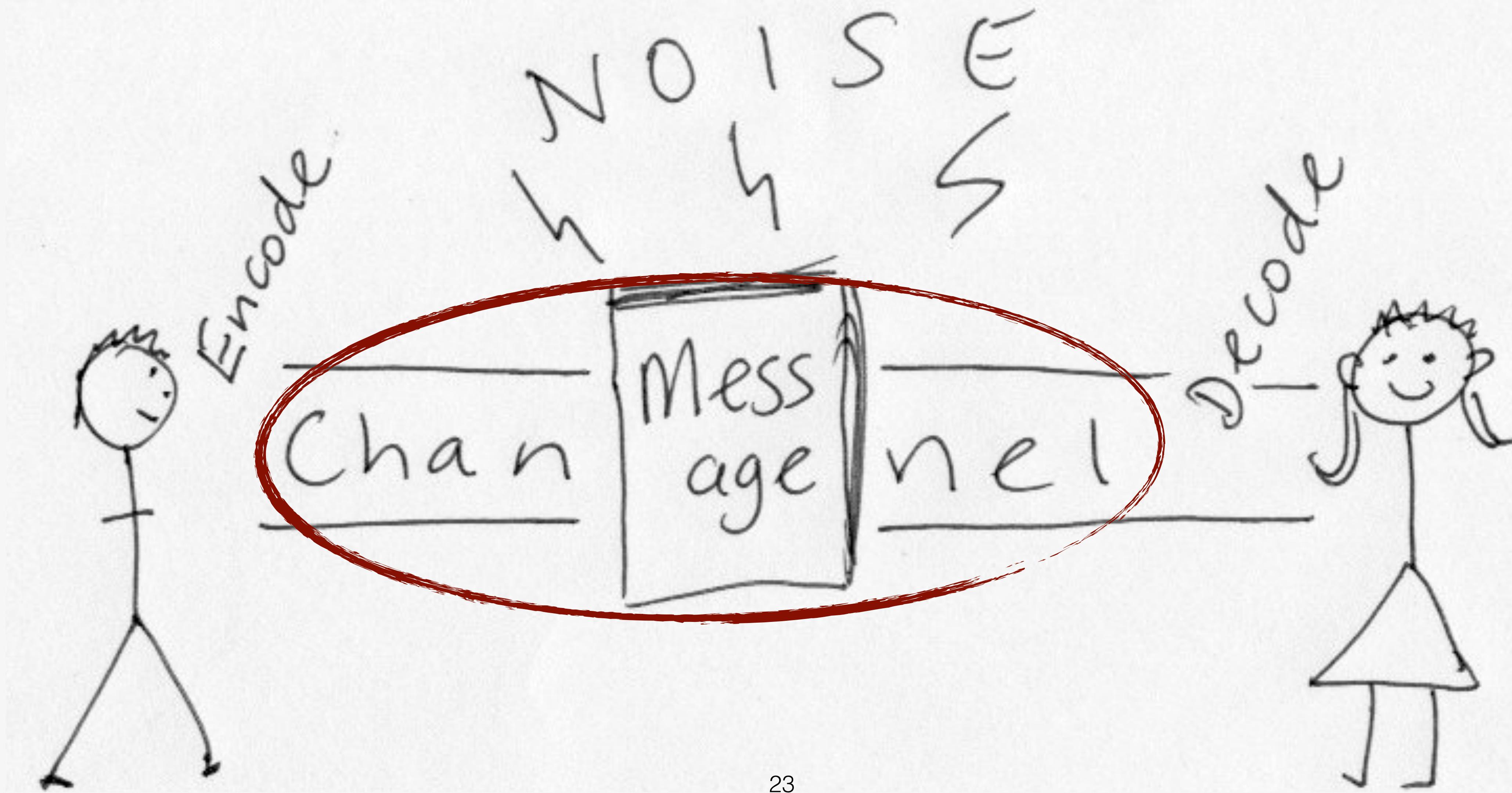
Scale

- Linear, logarithmic, categorical, time

Context

- Title, subtitle, labels, source, annotations

Be aware of the Channel



Printer Friendly?

Obama's Divided Nation

is over divided in 50 riven gath- ct its presi- viding f econ- n re- evi- Mr. close e has been and cked true nney and drawn attention to what hap-



Source: AP

Obama spoke to tum replied:

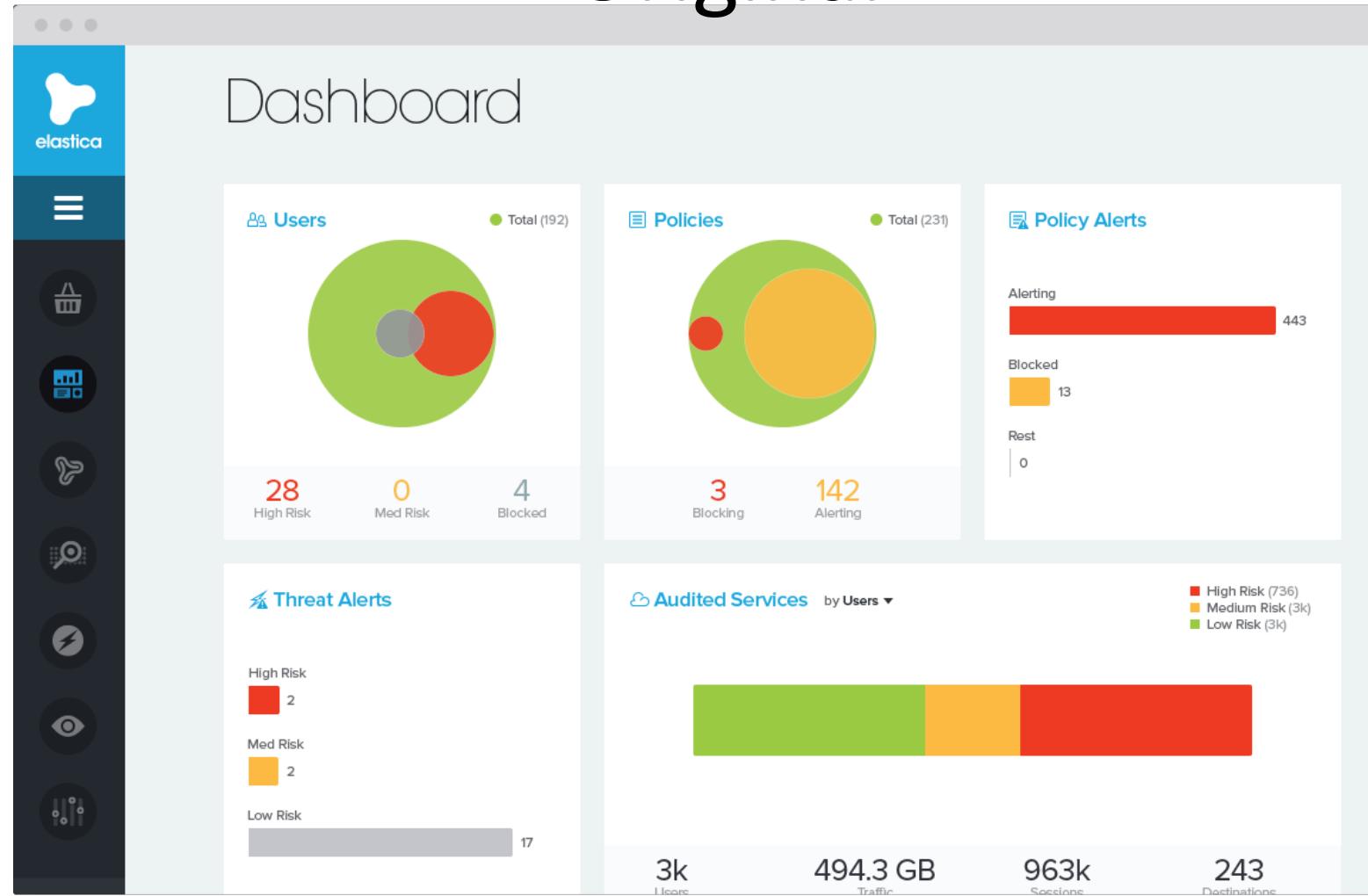
problem verbally that in revert that 9 who Tues 42% Sand tie p fact vot Mr. pol ac th he is B i

Color Blindness



Color Blindness

Original



Green-Blind / Deuteranopia



Red-Blind / Protanopia



Blue-Blind / Tritanopia



Color Brewer

<http://colorbrewer2.org/>

Number of data classes: 3

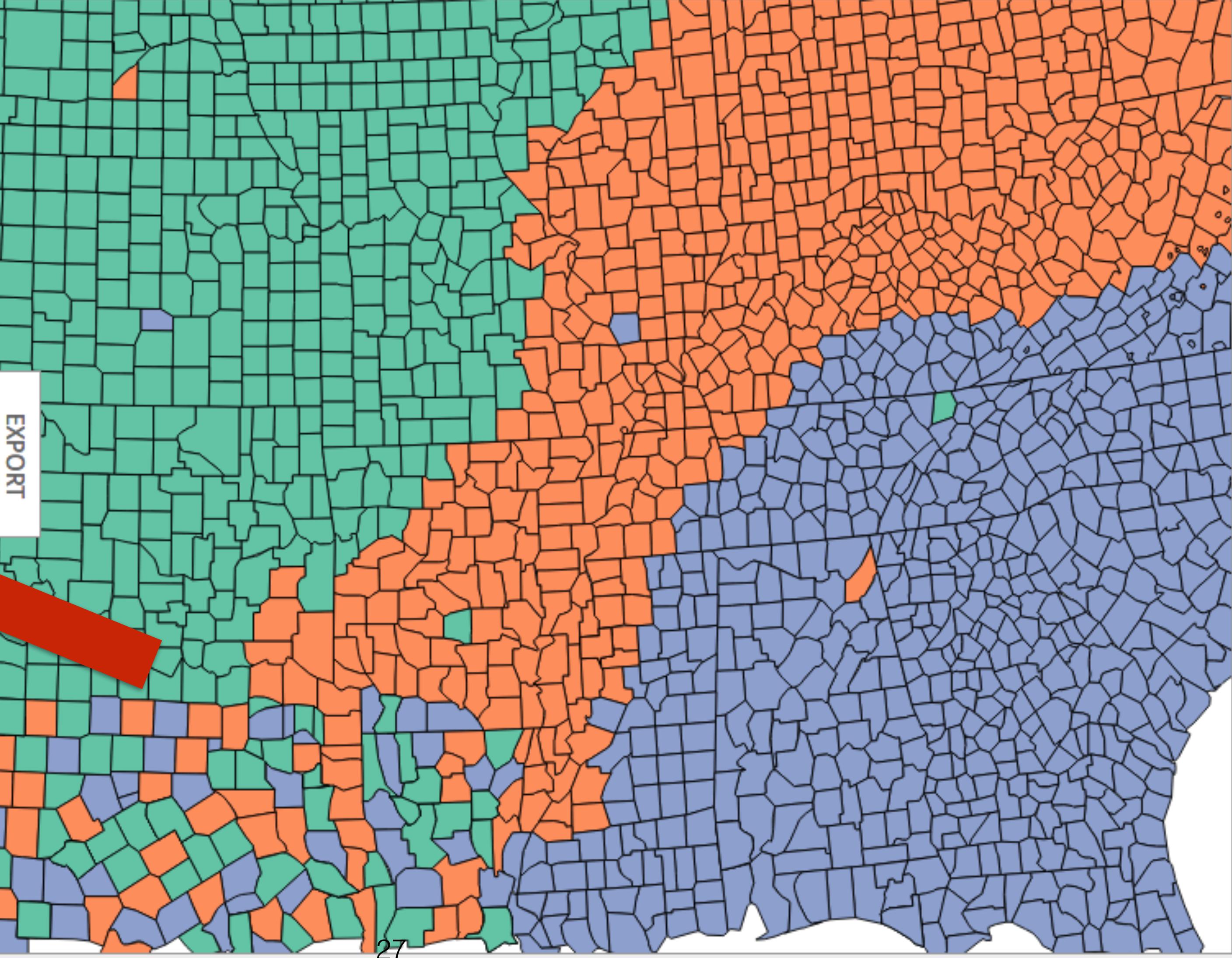
Nature of your data:
 sequential diverging qualitative

Pick a color scheme:


Only show:
 colorblind safe print friendly photocopy safe

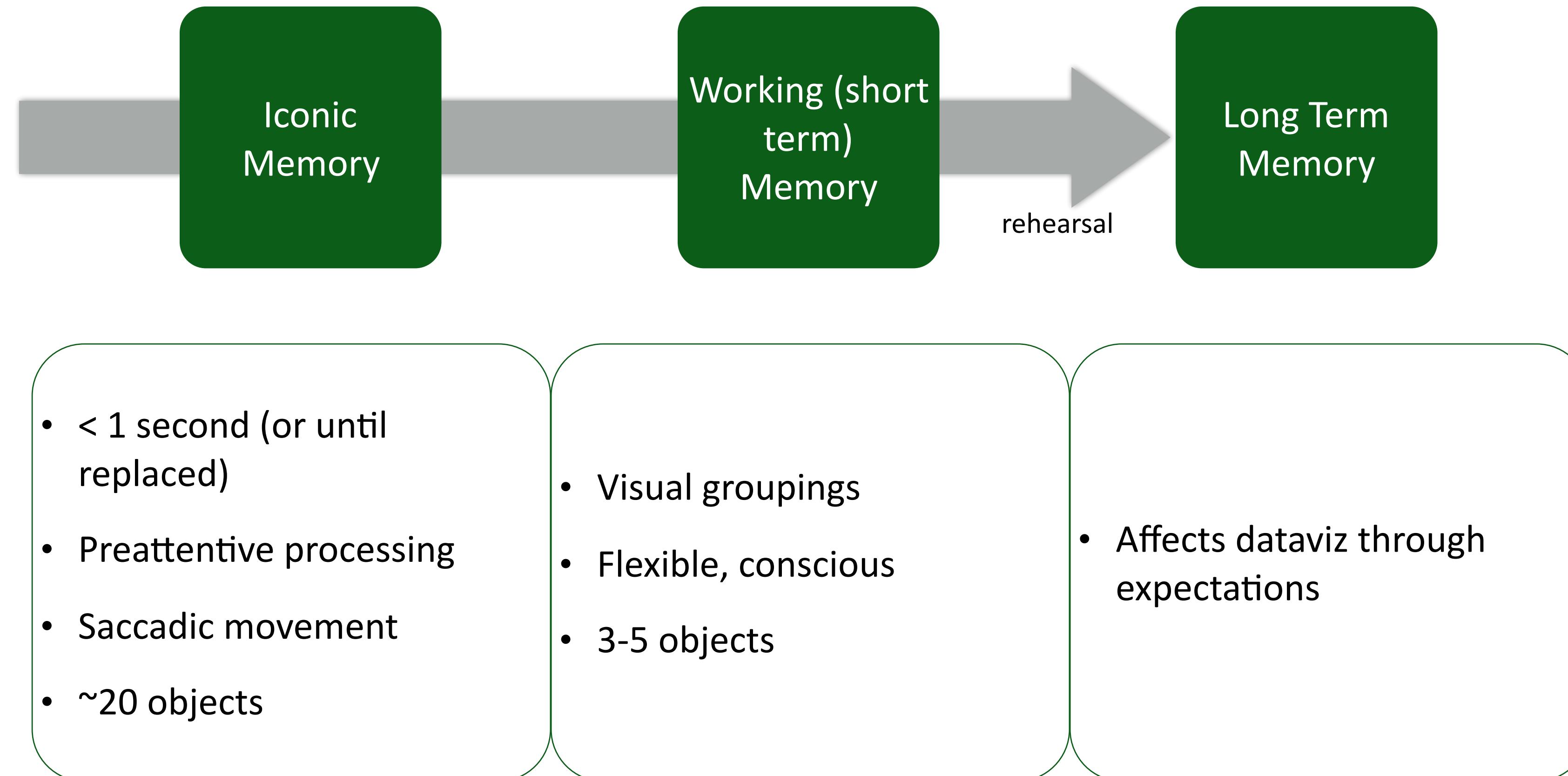
Context:
 roads cities borders


Background:
 solid color terrain
color transparency

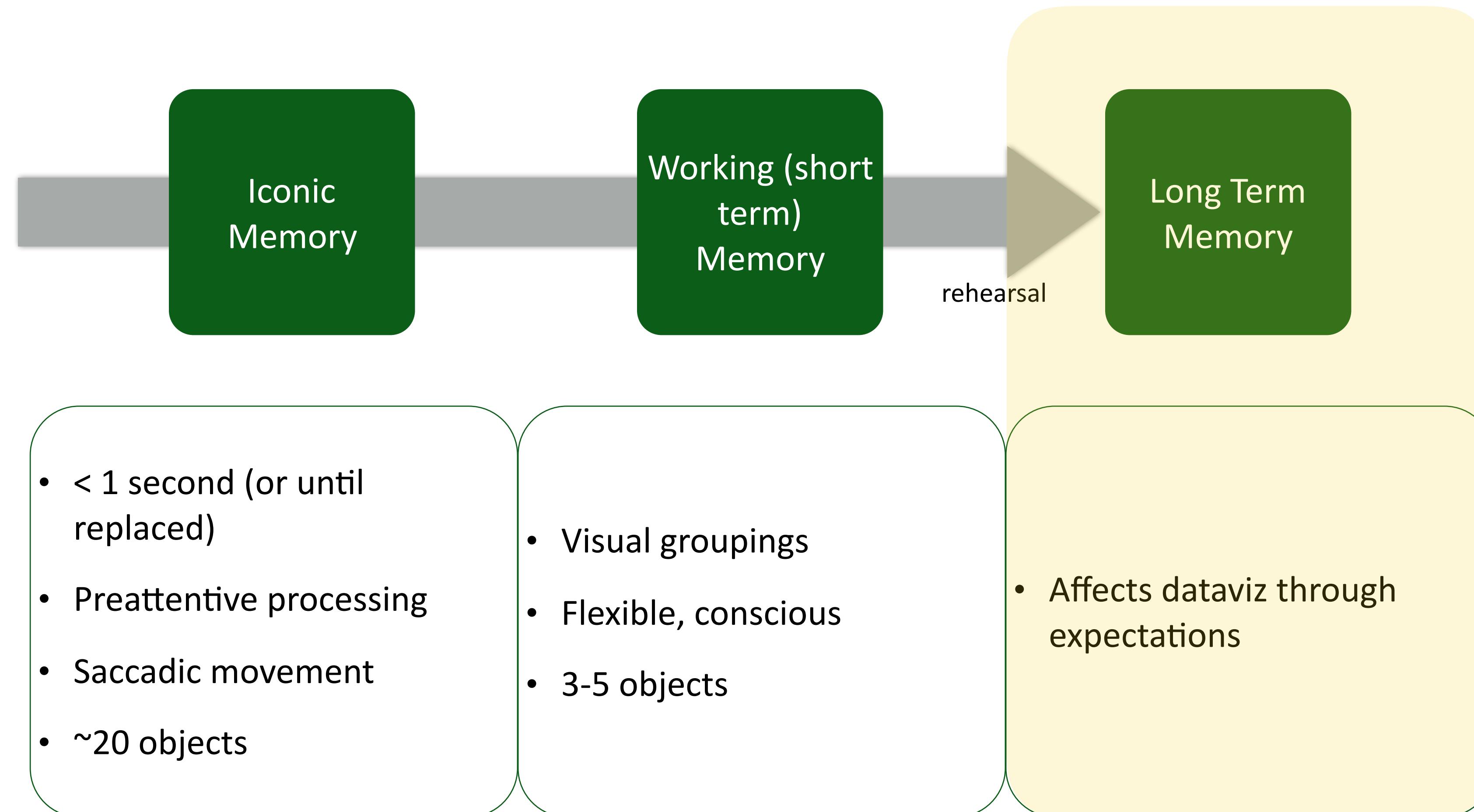
3-class Set2

EXPORT
HEX
#2a52be #8da6a5 #f4a460

The Visual Processing System

Human Visual Processing



Human Visual Processing



February 23, 2008

 SIGN IN TO E-MAIL OR SAVE THIS

FEEDBACK

The Ebb and Flow of Movies: Box Office Receipts 1986 – 2008

Summer blockbusters and holiday hits make up the bulk of box office revenue each year, while contenders for the Oscars tend to attract smaller audiences that build over time. Here's a look at how movies have fared at the box office, after adjusting for inflation.

Find Movie Go

June July Aug. Sept. Oct. Nov. Dec. 2007 Jan. 2008 Feb.

Each shape shows how one film did at the box office.

↑ Height shows weekly box office revenue ↓

← Width → shows longevity

The area of the shape (and its color) corresponds to the film's total domestic gross, through Feb. 21

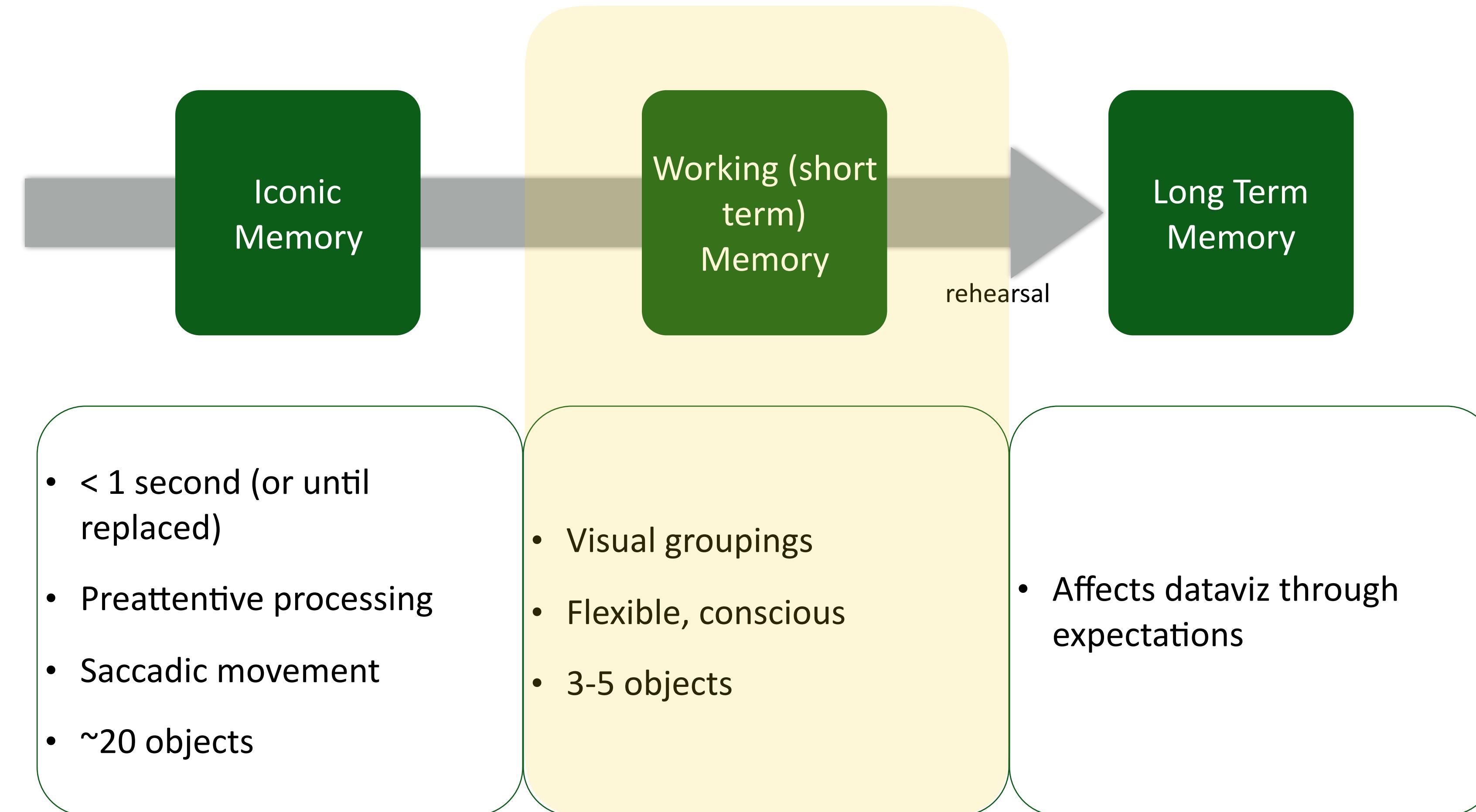
\$862 million
250
100
25
1

• Transformers
• Ocean's Thirteen
• Pirates of the Caribbean: At World's End
• Evan Almighty
• Live Free or Die Hard
• Fantastic Four: Rise of the Silver Surfer
• Ratatouille
• The Simpsons Movie
• I Now Pronounce You Chuck and Larry
• Hairspray (2007)
• Harry Potter and the Order of the Phoenix
• Superbad
• Rush Hour 3
• The Bourne Ultimatum
• Bee Movie
• Alvin and the Chipmunks
• National Treasure: Book of Secrets
• Enchanted
• American Gangster
• Juno
• I Am Legend

Sources: Baseline StudioSystems; Box Office Mojo

Mathew Bloch, Lee Byron, Shan Carter and Amanda Cox

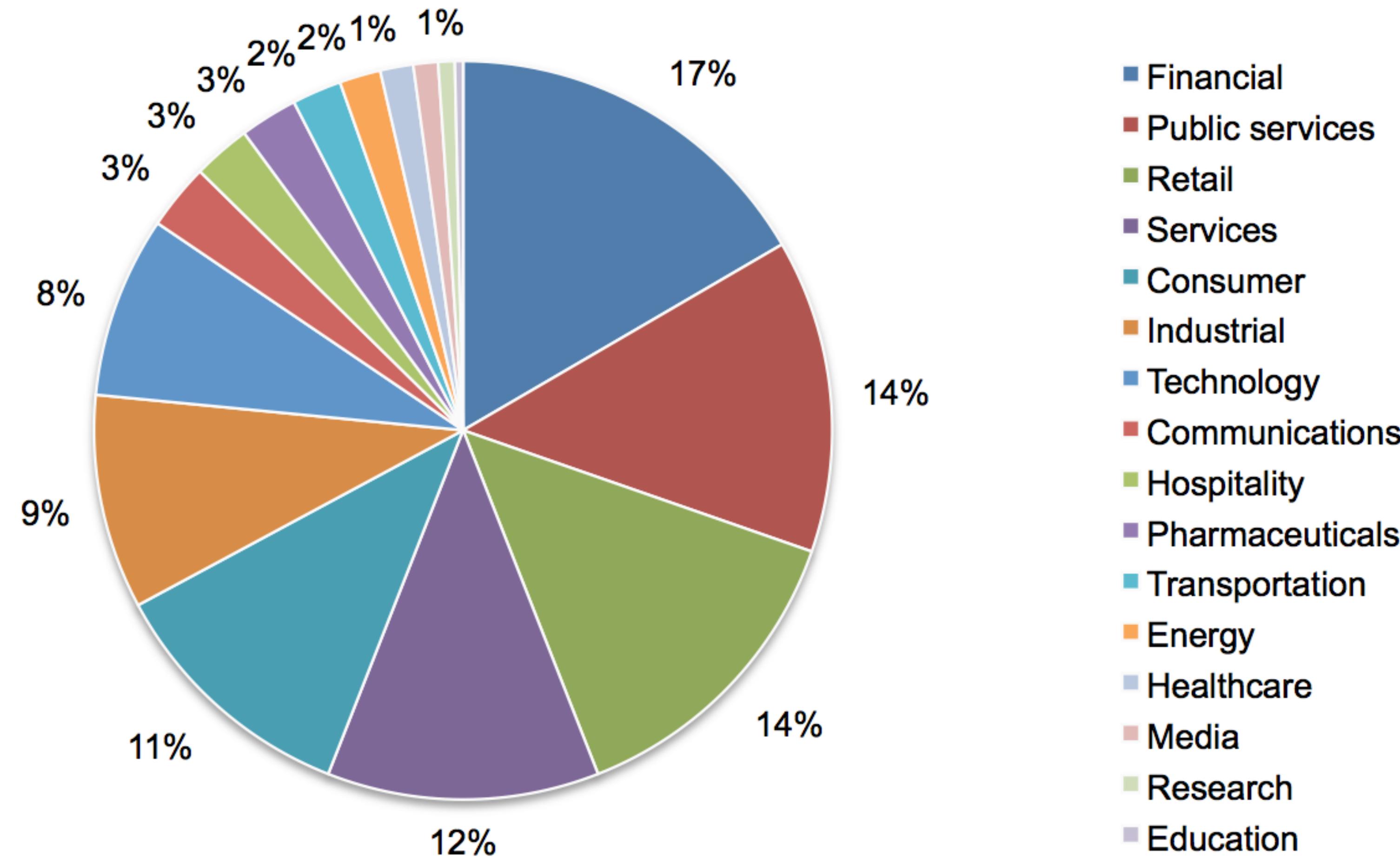
Human Visual Processing



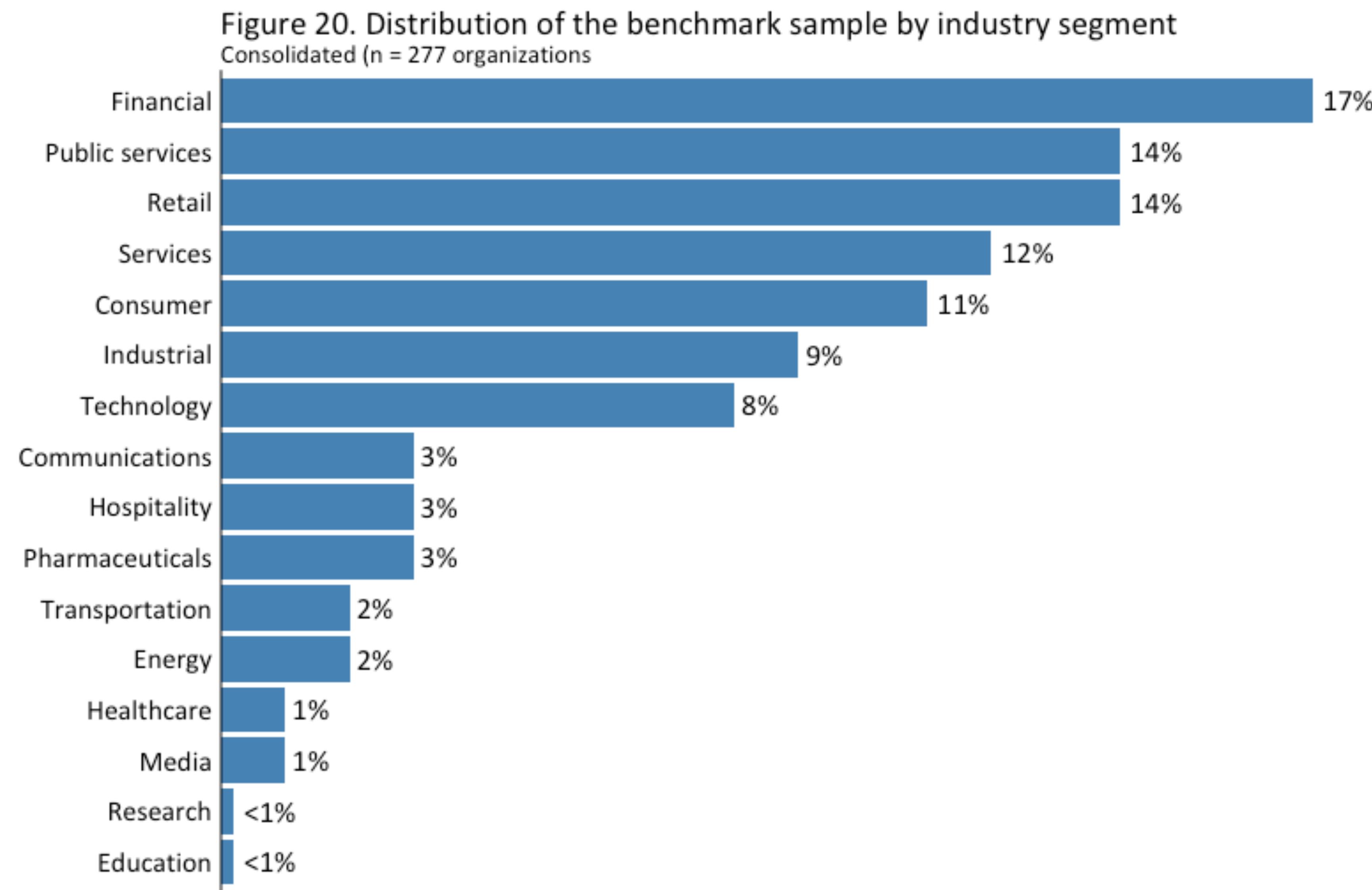
Overworking Working Memory

Figure 20. Distribution of the benchmark sample by industry segment

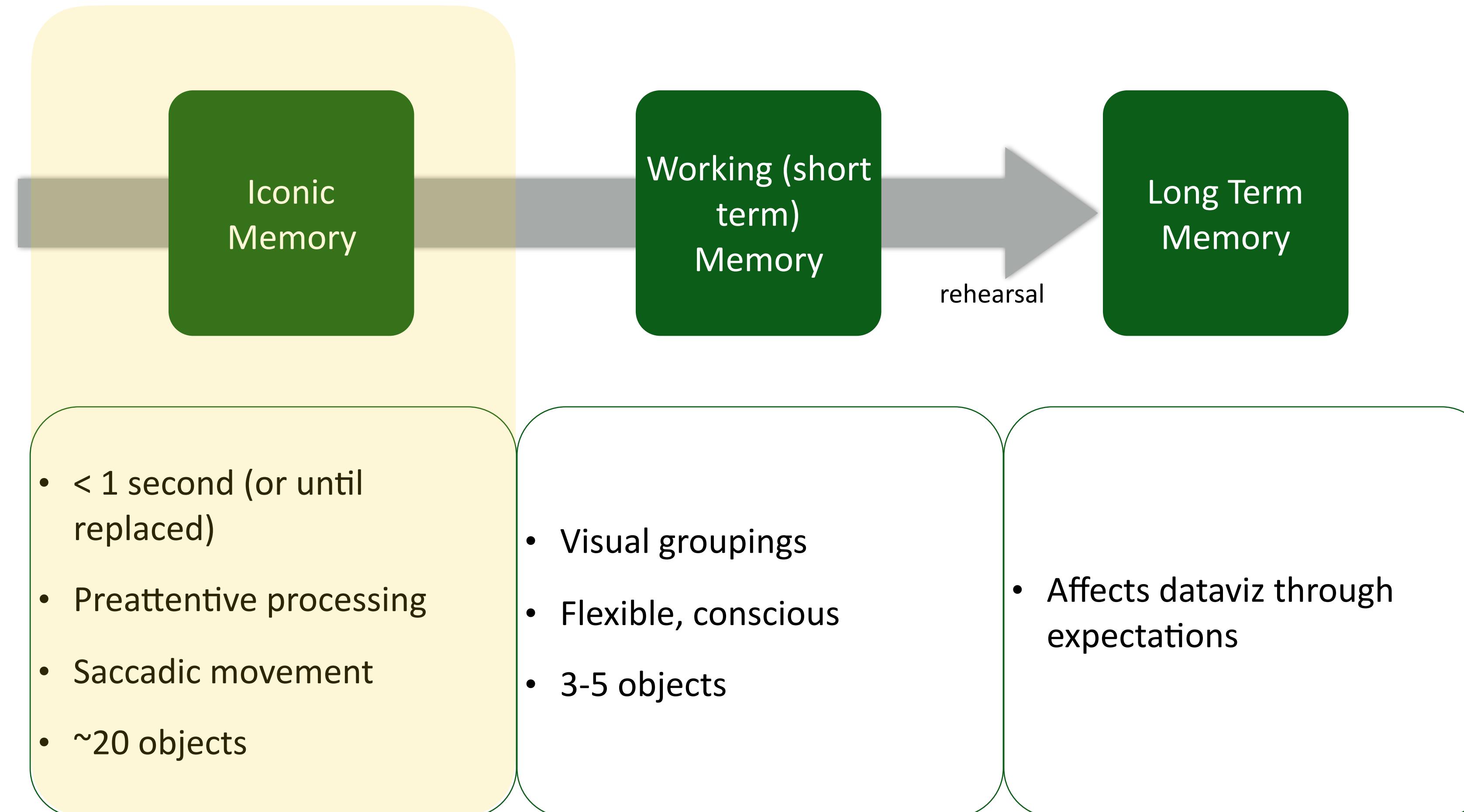
Consolidated (n = 277 organizations)



Working Memory: Direct Labels



Human Visual Processing



Pre-attentive Processing

Count the number of times “X” appears:

V 3 Jp d G I u Zy B h I G Jv b 2 s g a X M g a G F y Z C B 3
b 3 J r L C B i d X Q g b 25 I I H N p Z G U g c G V y a y B p
c y B 3 Z S B n Z X Q g d G 8 g a W 5 q Z W N 0 I G V h c 3 R I
c i B I Z 2 d z I G x p a 2 U g d G h p c y 4 g I E I m I H I v
d S d 2 Z S B m b 3 V u Z C B 0 a G I z L C B z Z W 5 k I H V z
I G E g b W V z c 2 F n Z S B v b I B 0 d 2 I 0 d G V y I Ch A
a H J i c m 1 z d H I g Y W 5 k I E B q Y X I q Y W N v Y n M p
I H N h e W I u Z v A i S G F w c H k a R W F z d G V y I i E =

Pre-attentive Processing

Count the number of times “X” appears:

V 3 Jp d G I u Z y B h I G Jv b 2 s g a X M g a G F y Z C B 3
b 3 J r L C B i d X Q g b 25 I I H N p Z G U g c G V y a y B p
c y B 3 Z S B n Z X Q g d G 8 g a W 5 q Z W N 0 I G V h c 3 R I
c i B I Z 2 d z I G x p a 2 U g d G h p c y 4 g I E I m I H I v
d S d 2 Z S B m b 3 V u Z C B 0 a G I z L C B z Z W 5 k I H V z
I G E g b W V z c 2 F n Z S B v b I B 0 d 2 I 0 d G V y I Ch A
a H J i c m 1 z d H I g Y W 5 k I E B q Y X I q Y W N v Y n M p
I H N h e W l u Z v A i S G F w c H k a R W F z d G V y I i E =

Saccadic Eye Movement



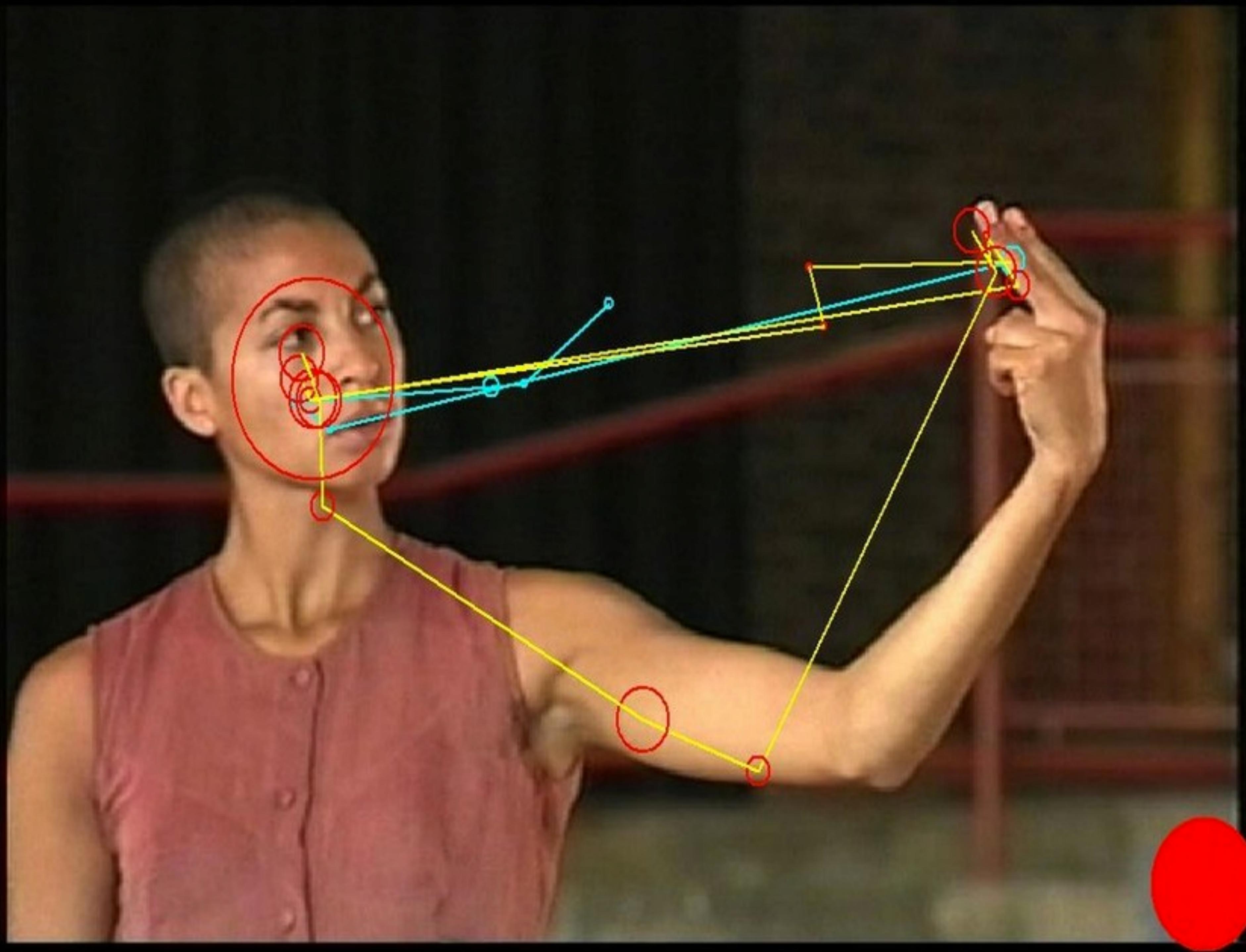
Saccadic Eye Movement

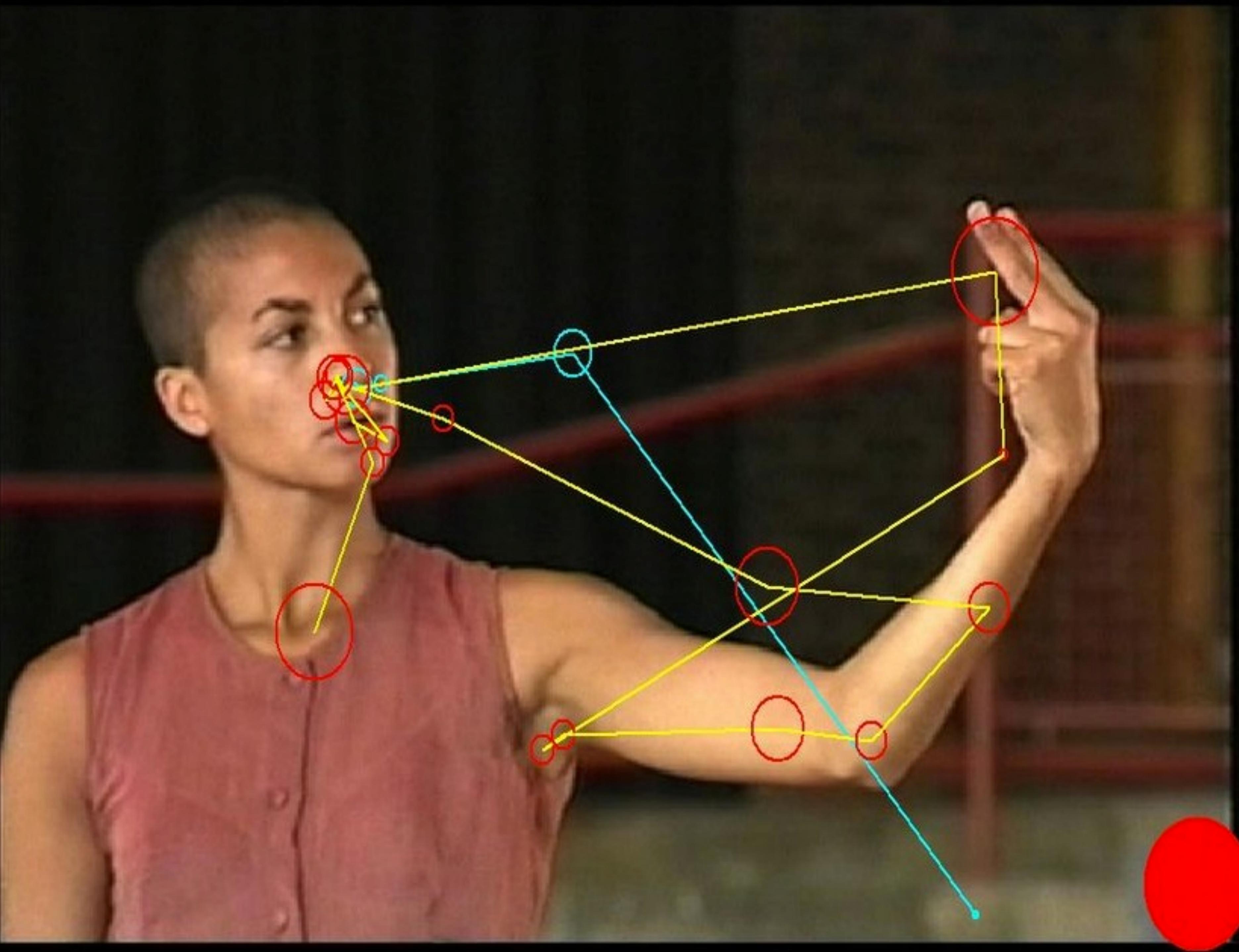
- Pre-attentively Driven
- Ballistic
- Visual Suppression
- Cannot Change after start

Saccadic Eye Movement



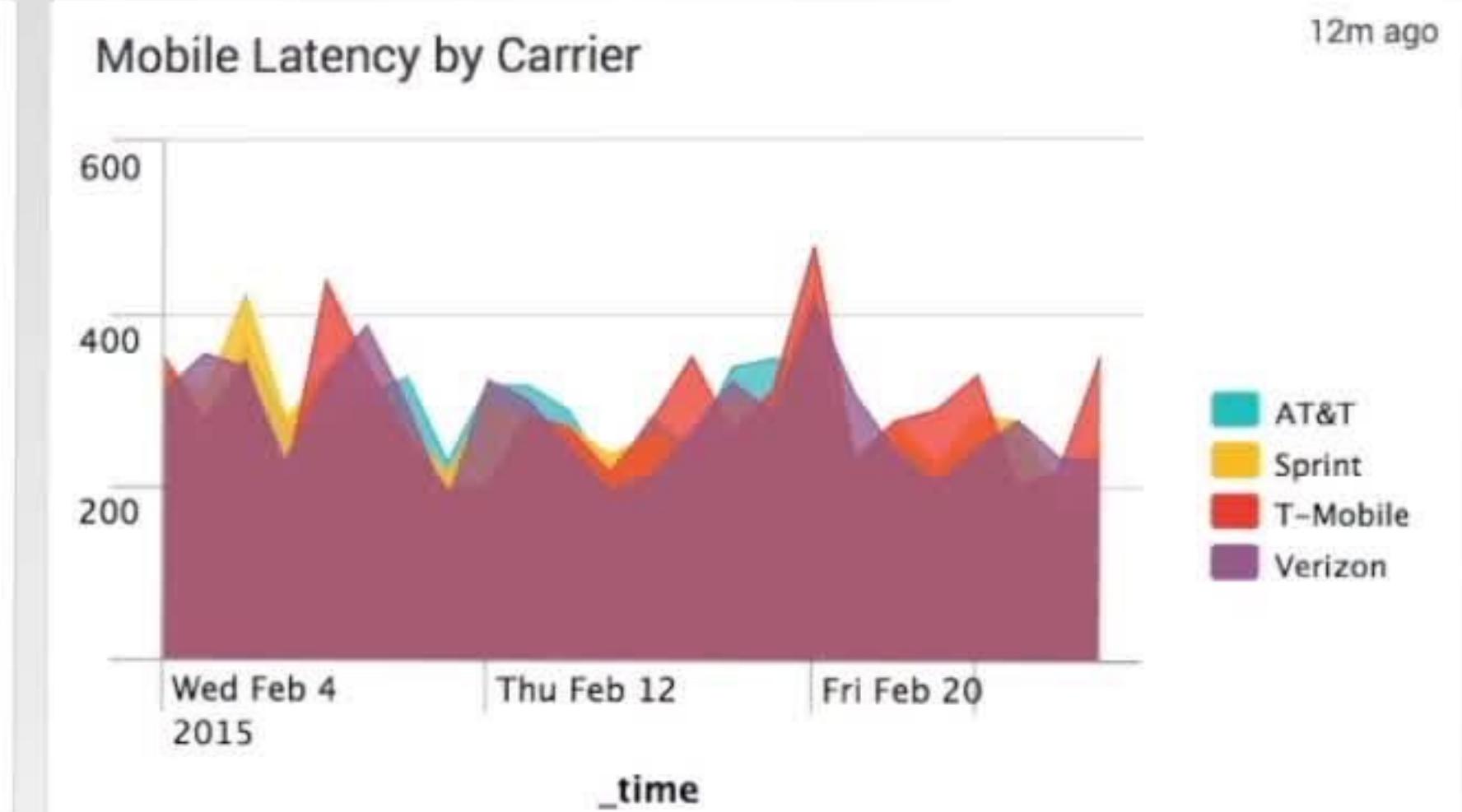
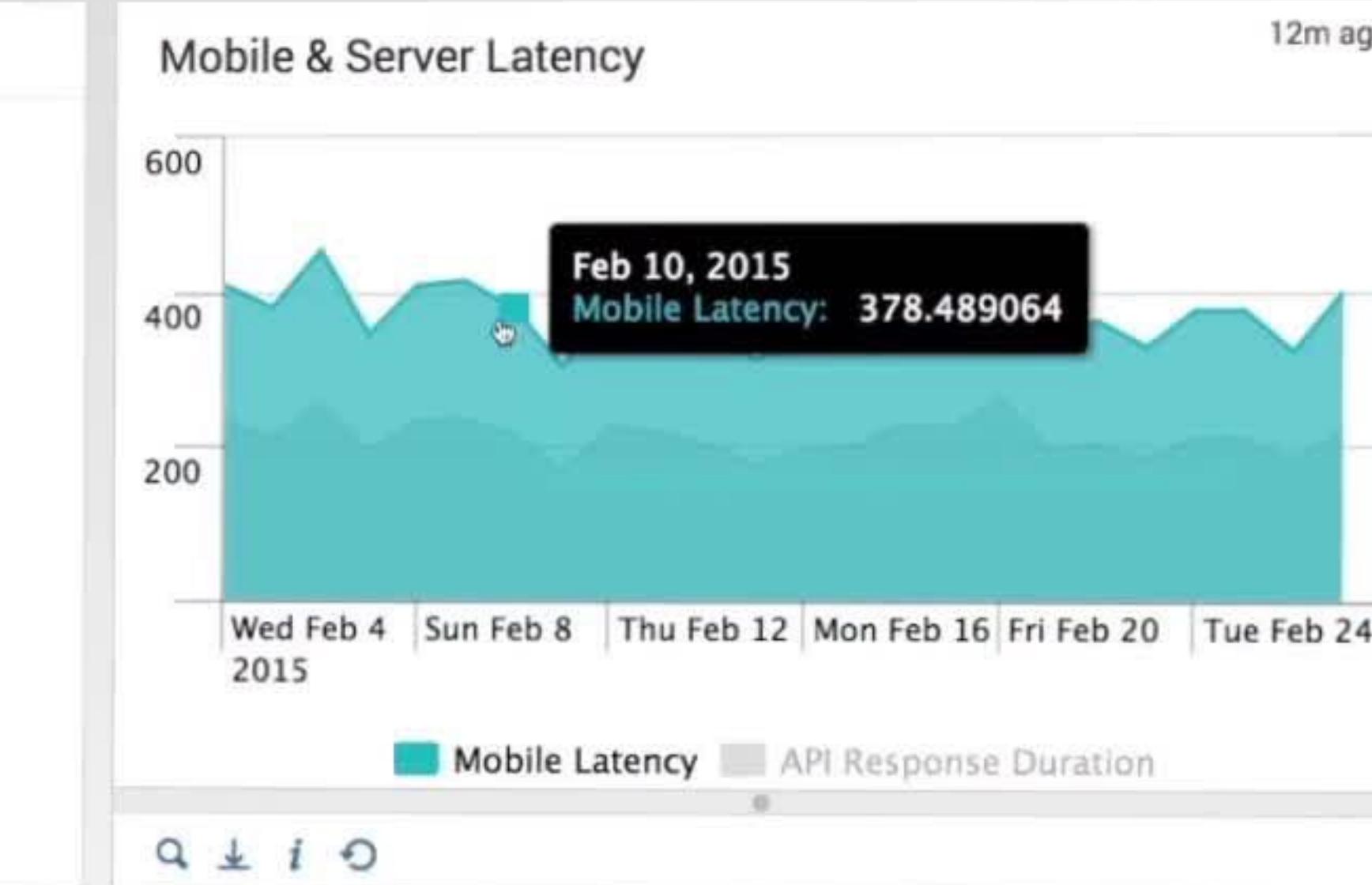
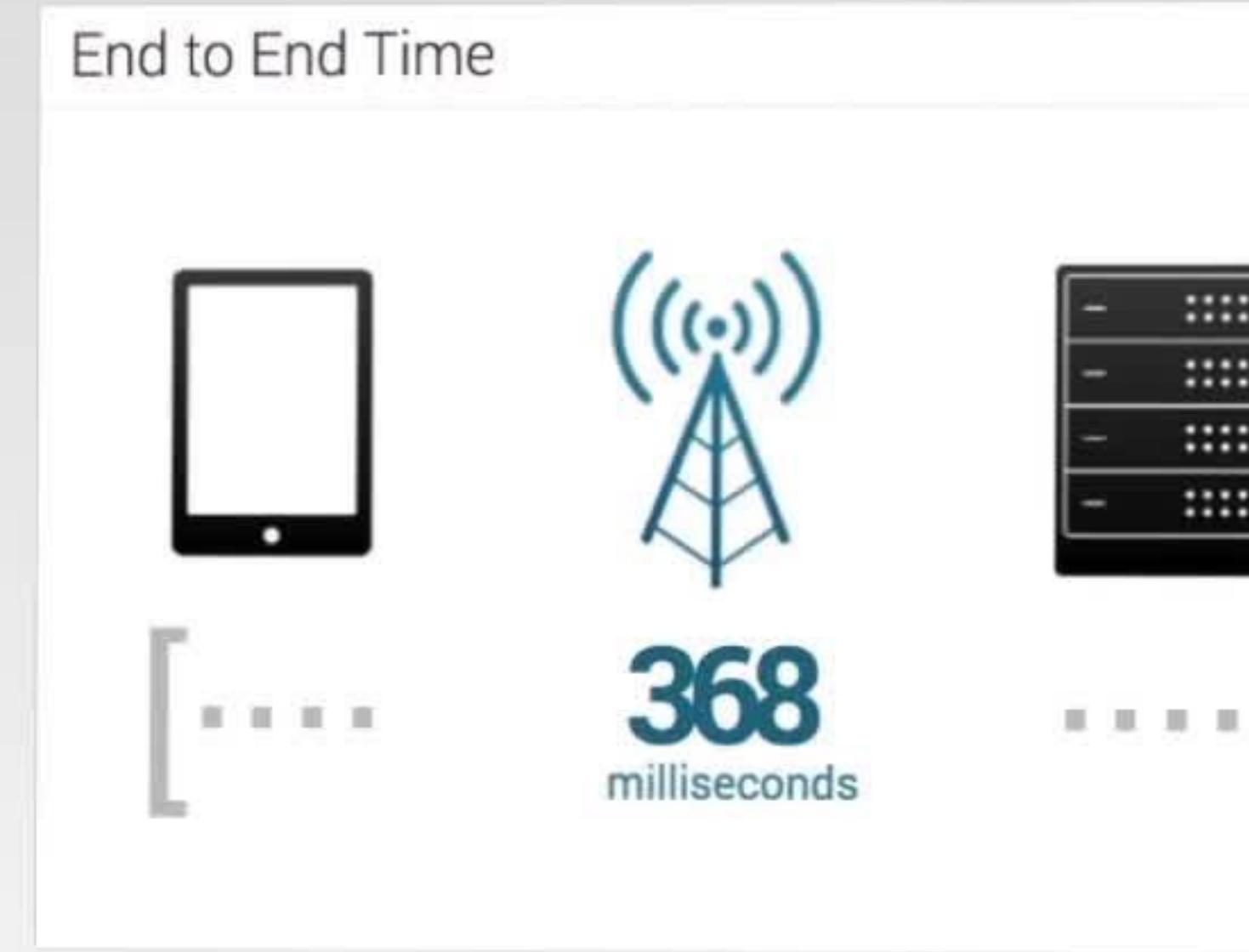






Mobile Ops Dashboard

Edit More Info

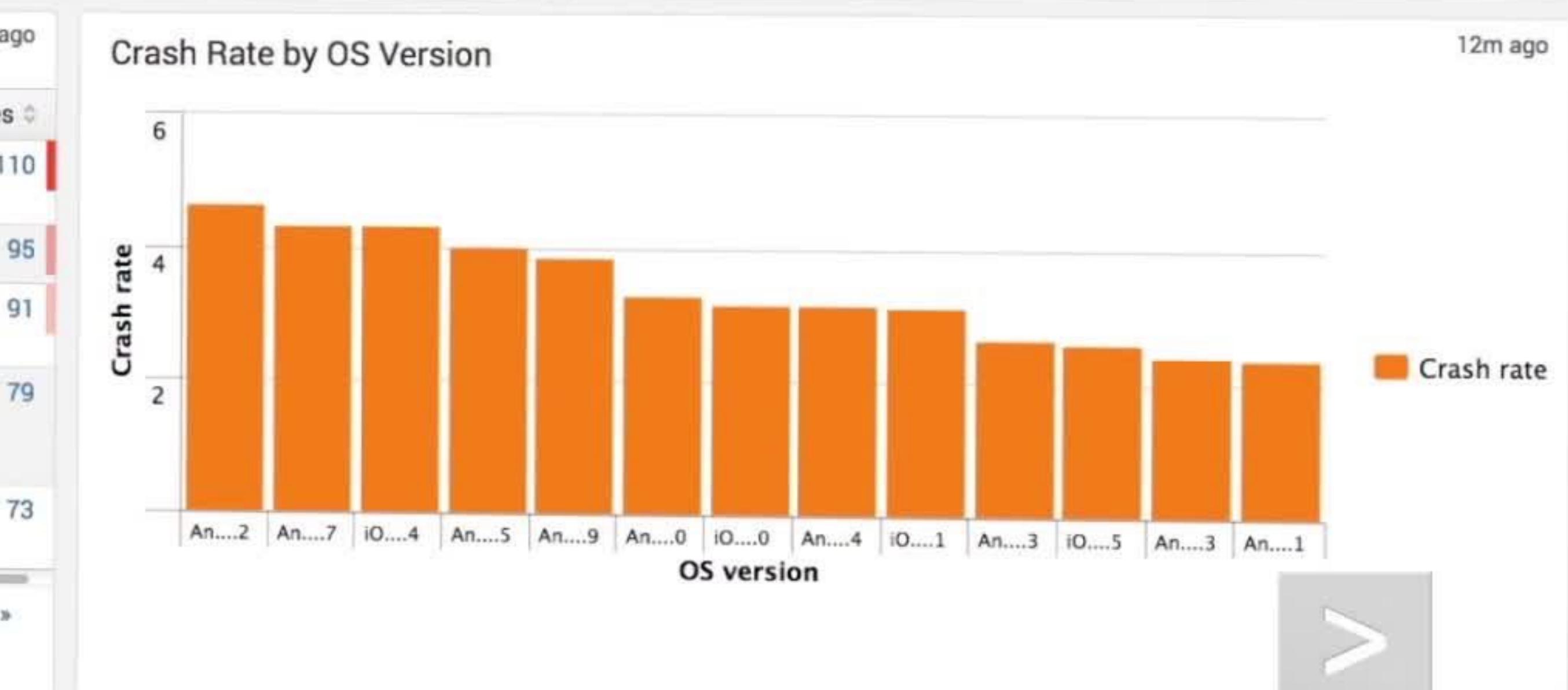


Mobile App Crashes

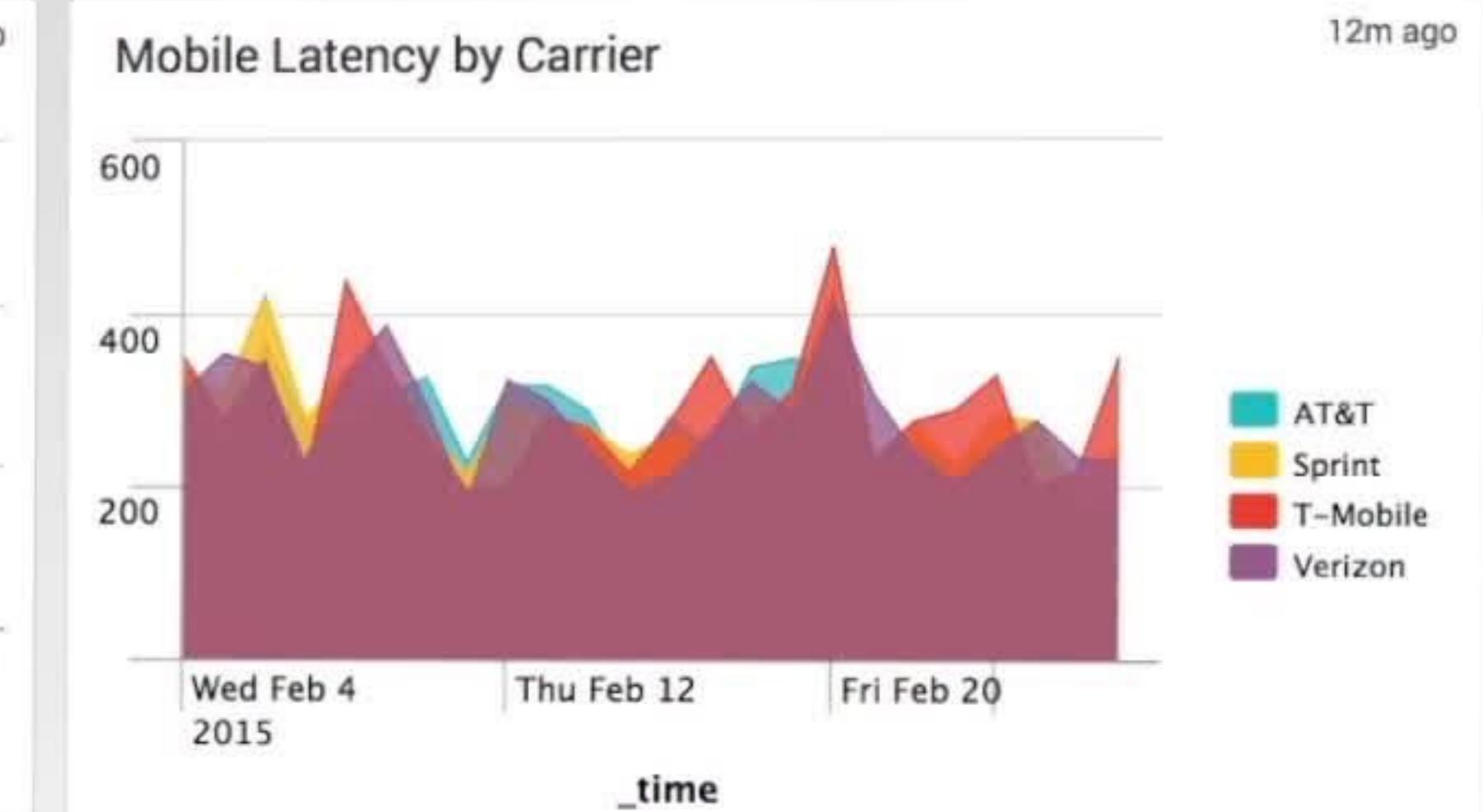
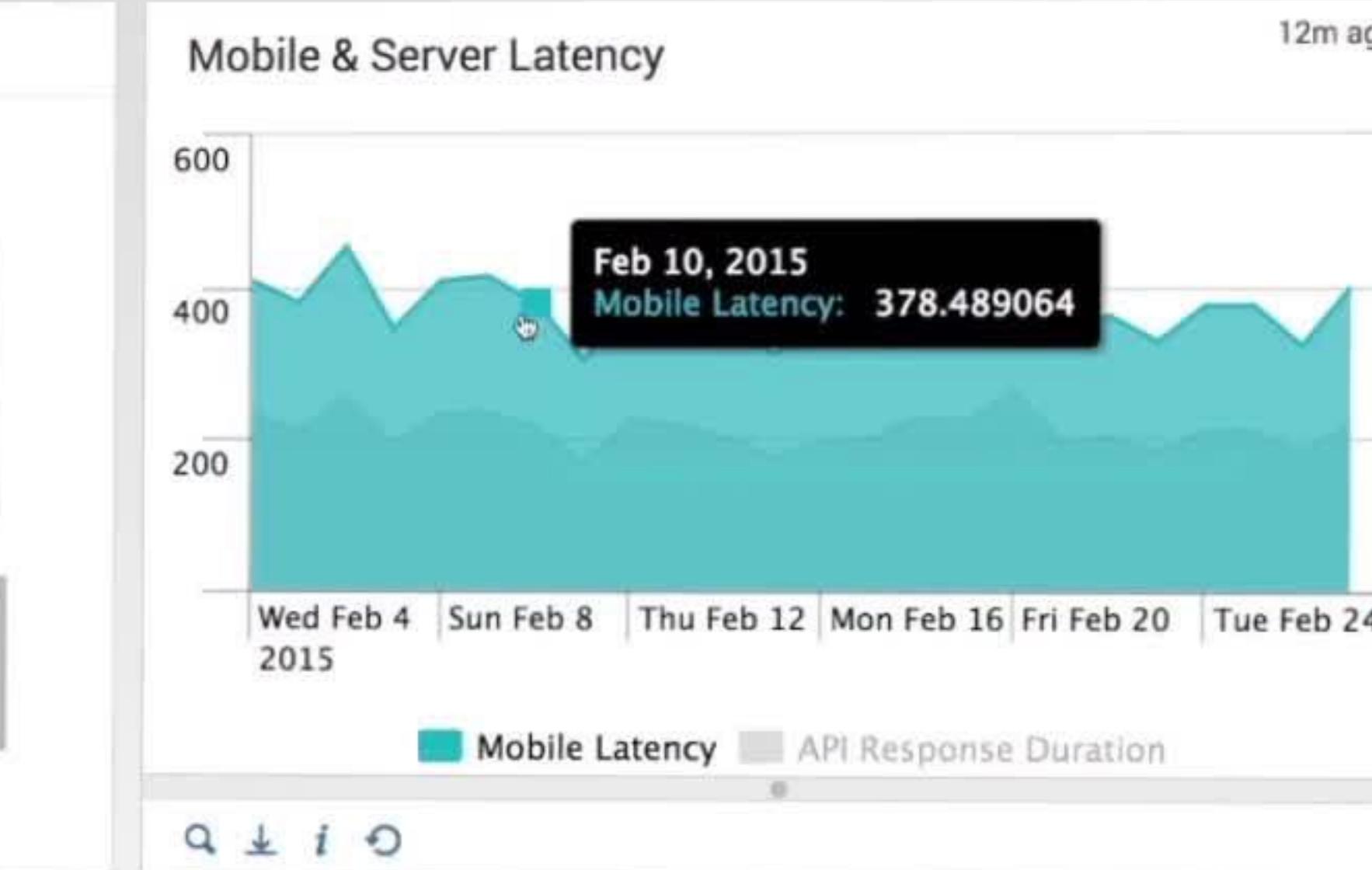
12m ago

	platform	message	where	Trend	Occurrences
1	iOS	SIGSEV	[CrashController Handled1]		110
2	Android	java.lang.NullPointerException	AsyncThread.java:55		95
3	iOS	SIGILL	[CrashController Handled1]		91
4	Android	java.lang.IllegalStateException: Could not execute method of the activity	AsyncThread.java:55		79
5	iOS	SIGILL	[UIKITController Handled5]		73

< prev 1 2 3 next >

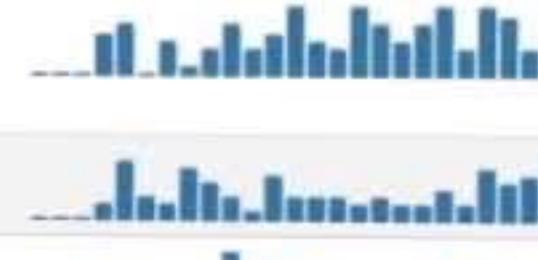
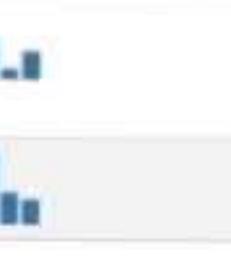
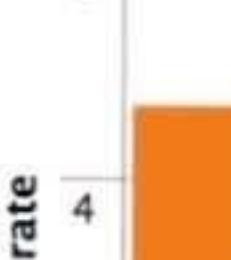


Mobile Ops Dashboard

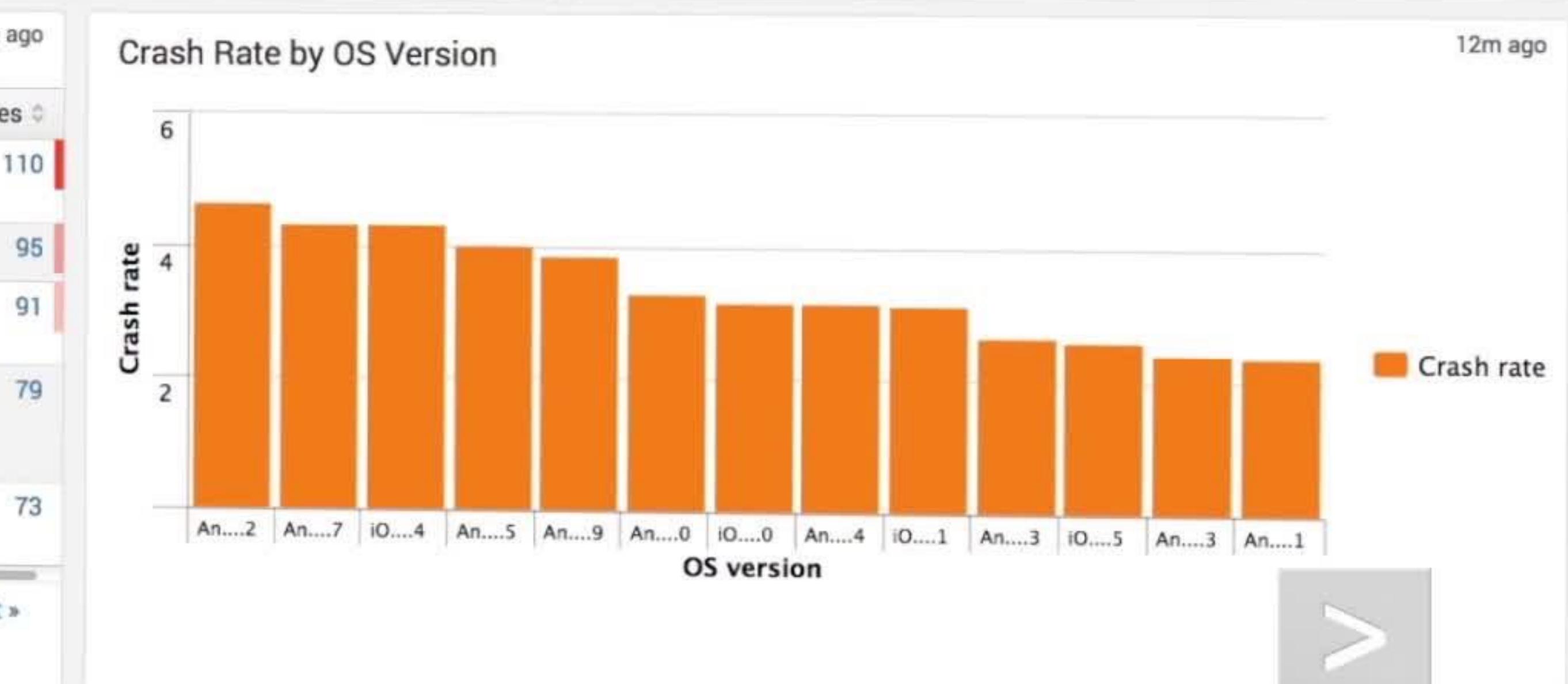
Edit More Info  

Mobile App Crashes

12m ago

	platform	message	where	Trend	Occurrences
1	iOS	SIGSEV	[CrashController Handled1]		110
2	Android	java.lang.NullPointerException	AsyncThread.java:55		95
3	iOS	SIGILL	[CrashController Handled1]		91
4	Android	java.lang.IllegalStateException: Could not execute method of the activity	AsyncThread.java:55		79
5	iOS	SIGILL	[UIKITController Handled5]		73

« prev 1 2 3 next »



[Available reports for dashboard](#)

Dashboard > Device Dashboard > Event Status

Incidents: Top Incidents Categories Ranked By Count (last 30 min...)

Category	Count
Overall Health Issue	~100
Reconn	~100
Traffic Anomaly	~100
Ungrouped	~100
Hardware Issue	~100
Logon Failure	~100
Config Change	~100
Service Non-responsive	~100
Multiple Exploit	~100
Inst Software Change	~100

Total Event Rate (Per Sec) (last 30 mins,4,8 hours updated@19:33)

Event Type	Rate (Per Sec)
Blue	~150
Yellow	~150
Red	~150

Events: Top Dest IP By Count (last 8 hours updated@19:33)

IP Address	Count
172.16.10.36	312,206
172.16.10.16	299,195
172.16.10.15	296,025
172.16.10.18	246,624

Top Ports, Events By Severity, Count (last 1 hour updated@19:43)

Legend:

- TCP(6),HTTPS(443),WEB-MISC PCT Client_Hello ove
- UDP(17),DOMAIN(53),Denied IP traffic by policy,Seve
- UDP(17),59866,Windows Filtering blocked a connectio
- UDP(17),57968,Windows Filtering blocked a connectio
- UDP(17),49359,Windows Filtering blocked a connectio

Events: Top Network Connections, Events By Severity, Count (las...

Legend:

- 172.16.20.121
- 192.168.24.11
- 224.0.0.252,1
- 224.0.0.252,1
- 224.0.0.252,1
- 224.0.0.252,1
- 224.0.0.252,1
- 224.0.0.252,1
- 224.0.0.252,1

Value: 2,090

Settings

Title: Top Rpt IP By Event Rate

Description:

Condition: Please see Report Template section for details.

Display: Combo View

Time: Last 1 hour

of results: 10

OK Cancel

Top Events By Severity, Count (last 1 hour updated@19:43)

Event Name	COUNT	Trend
A user account was locked	2	
Scanning detected and suc	5	
Host Vulnerability score ex	1	
Memory hardware health:	20	
Power supply hardware he	19	
WEB-MISC PCT Client_Hell	2	
Windows Filtering blocked	10,674	
Denied IP traffic by policy	2,105	
Hardware health status	19	
Power supply hardware he	19	

Incidents: Top Incidents Ranked By Severity, Count (last 1 hour u...

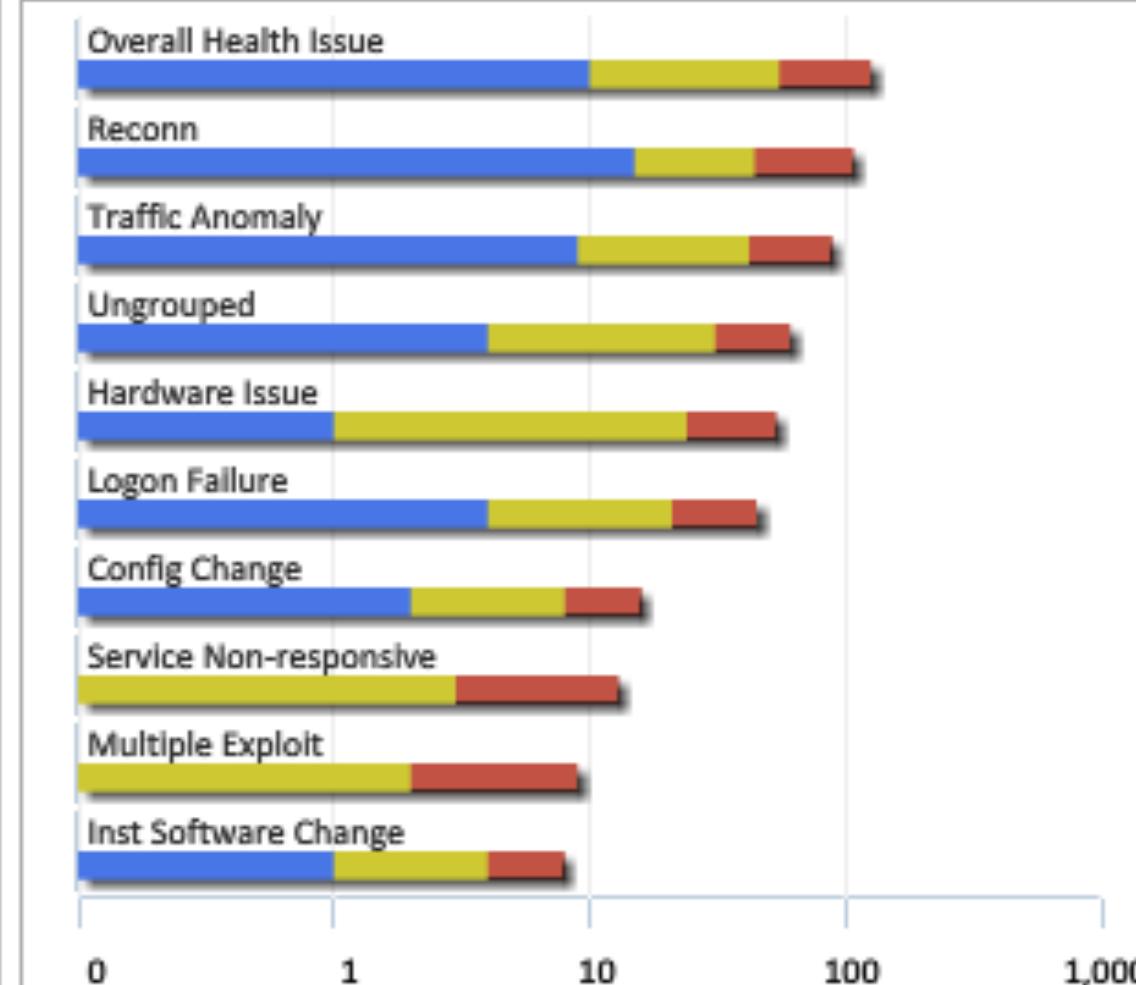
Event Name	COUNT	Trend
Account Lockout: Server	2	
Important application stop	1	
User_Added_to_Domain_	1	
Biz Srv Performance Heat	3	
Difference in Running and	2	

[Available reports for dashboard](#)

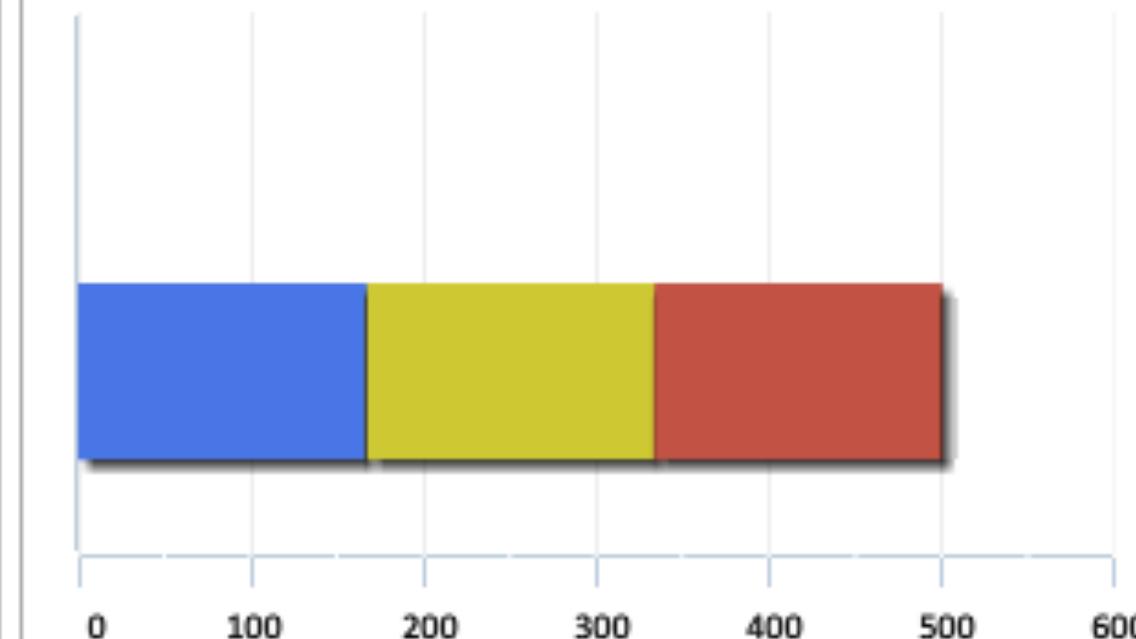
- [Summary Dashboards](#)
 - Biz Service Summary
- [Device Summary](#)
 - All Devices
 - Servers
 - Network Devices
 - My Devices
- [VMWare Summary](#)
- [Hardware Summary](#)
- [Storage Summary](#)
- [Application Summary](#)
- [Incident Dashboard](#)
- [Availability Dashboard](#)
- [Performance Dashboard](#)
 - BizSvc
 - Device
- [Security Dashboard](#)
 - BizSvc
 - Device
- [Biz Svc Dashboard](#)
- [Device Dashboard](#)
 - Network
 - Server
 - Application
- [Event Status](#)
- [Storage](#)
- [My Dashboard](#)
- [Ungrouped](#)

Dashboard > Device Dashboard > Event Status

Incidents: Top Incidents Categories Ranked By Count (last 30 min...)



Total Event Rate (Per Sec) (last 30 mins,4,8 hours updated@19:33)

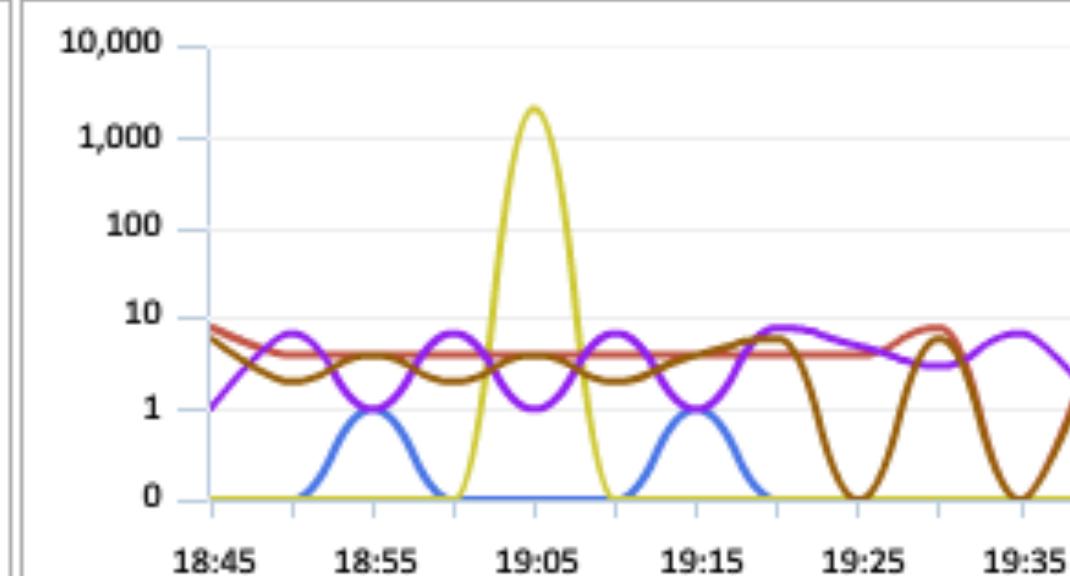


Events: Top Dest IP By Count (last 8 hours updated@19:33)



Available reports for dashboard

Top Ports, Events By Severity, Count (last 1 hour updated@19:43)



Event Name COUNT Trend

A user account was locked	2	
Scanning detected and suc	5	
Host Vulnerability score ex	1	
Memory hardware health:	20	
Power supply hardware he	19	
WEB-MISC PCT Client_Hell	2	
Windows Filtering blocked	10,674	
Denied IP traffic by policy	2,105	
Hardware health status	19	
Power supply hardware he	19	

Settings

Title: Top Rpt IP By Event Rate

Description:

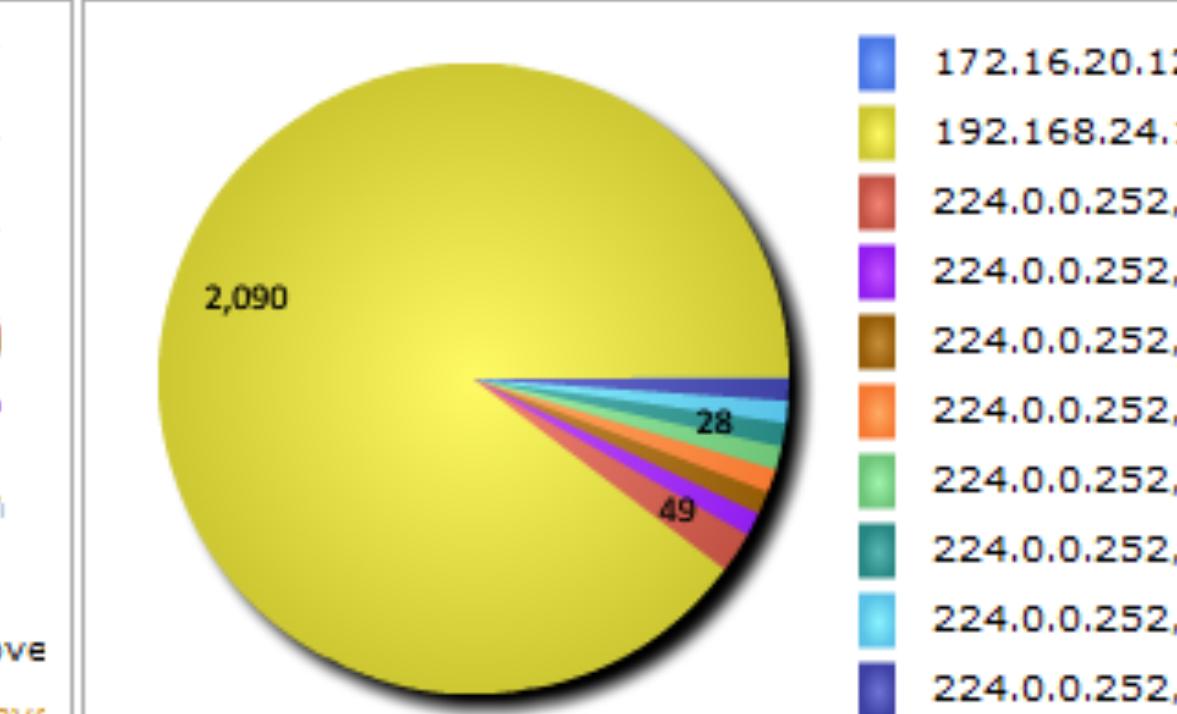
Condition: Please see Report Template section for details.

Display: Combo View

Time: Last 1 hour

of results: 10

Events: Top Network Connections, Events By Severity, Count (las...



Top Events By Severity, Count (last 1 hour updated@19:43)

Event Name	COUNT	Trend
A user account was locked	2	
Scanning detected and suc	5	
Host Vulnerability score ex	1	
Memory hardware health:	20	
Power supply hardware he	19	
WEB-MISC PCT Client_Hell	2	
Windows Filtering blocked	10,674	
Denied IP traffic by policy	2,105	
Hardware health status	19	
Power supply hardware he	19	

Incidents: Top Incidents Ranked By Severity, Count (last 1 hour u...

Event Name	COUNT	Trend
Account Lockout: Server	2	
Important application stop	1	
User_Added_to_Domain_	1	
Biz Svc Performance Heal	3	
Difference in Running and	2	

Navigation Menu

Incident Response

- Security Alerts
- Security Incidents
- Incident Investigations
- Forensic Analysis
- Incident Response Tasks
- Incident Journal

Data Breach Response

- Data Breaches
- Breach Tasks
- Breach Risk Assessment
- Notifications & Call Trees
- Notification History

SOC Program Management

- Shift Handover
- SOC Procedure Library
- Security Controls
- SOC Policies
- Contacts
- Teams
- Team Members
- Competencies
- Training & CPE Tracking

Issue Management

- Findings
- Remediation Plans
- Exception Requests
- Policy Change Requests

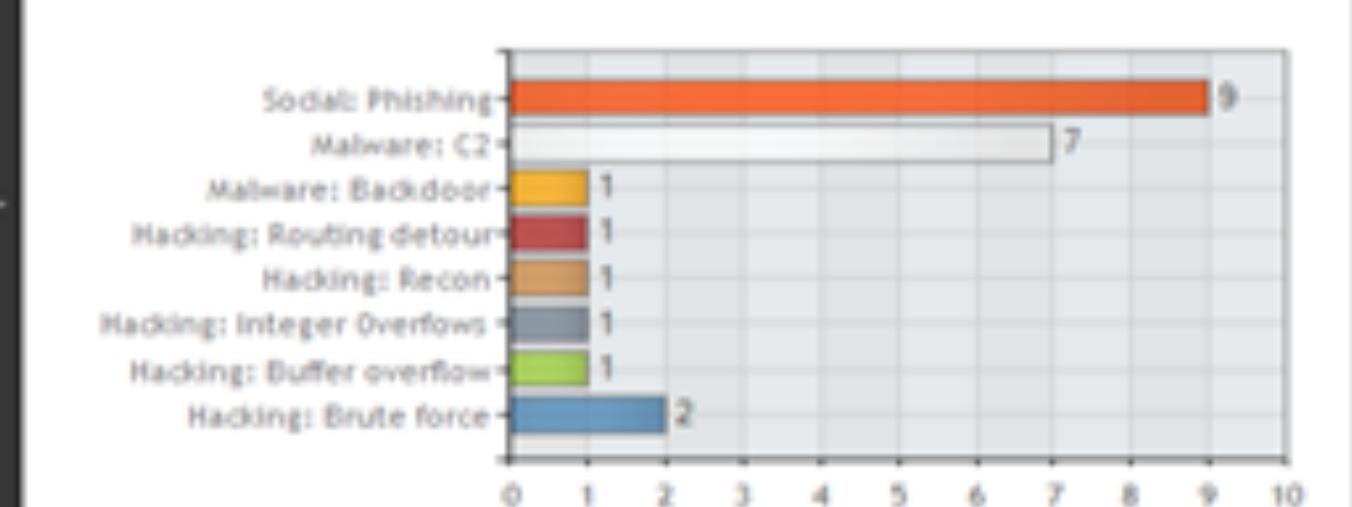
Dashboard: SOC Manager

Security Alerts by External Destination

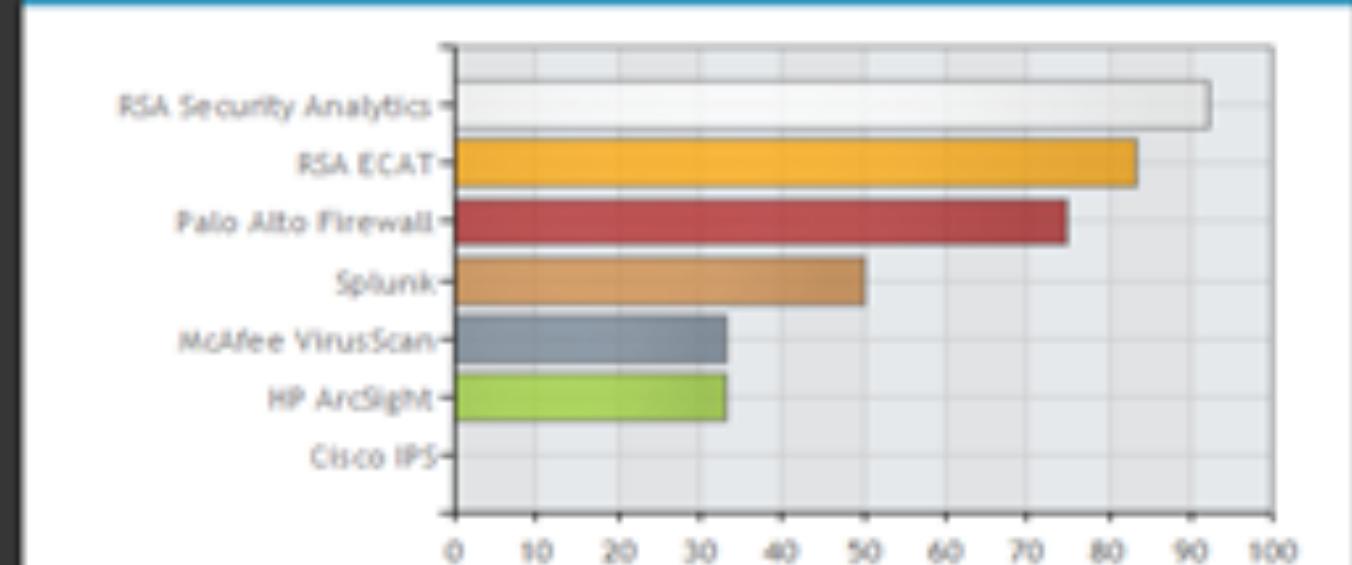


Threat Metrics

Incidents by Threat Categories

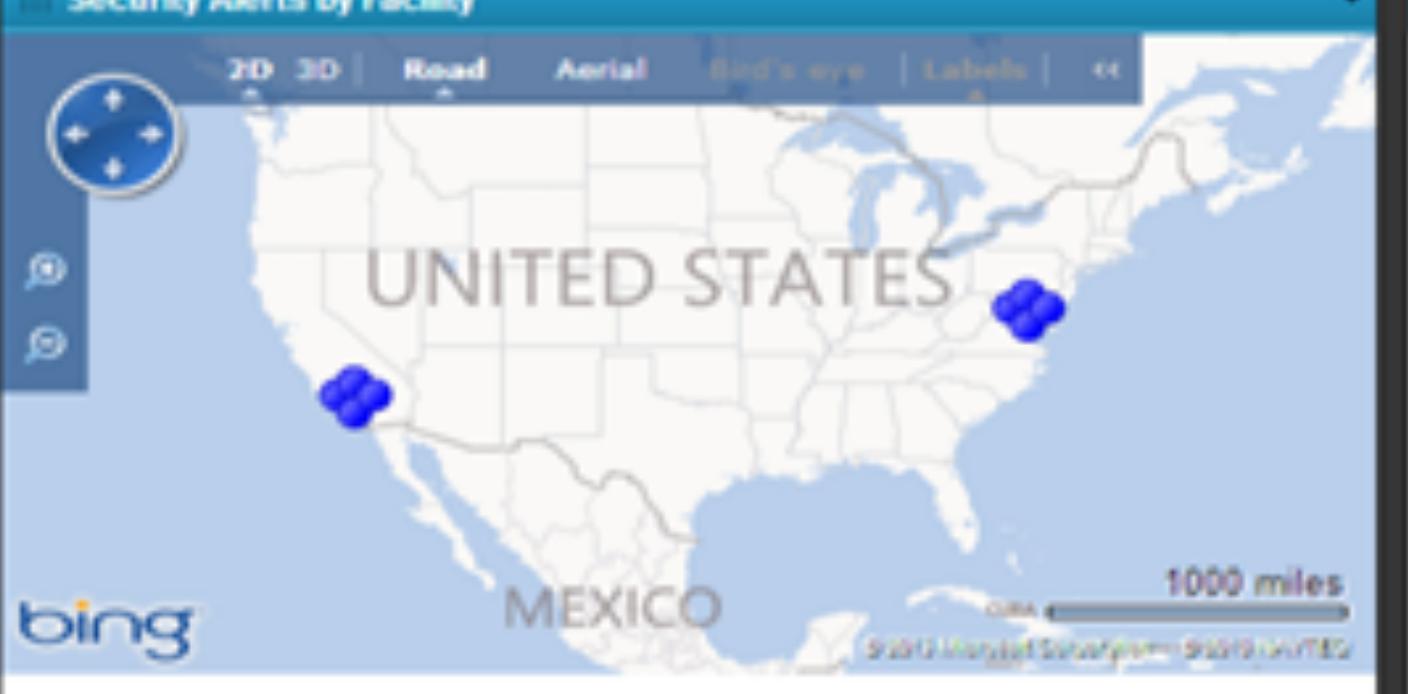


Security Tools Efficacy



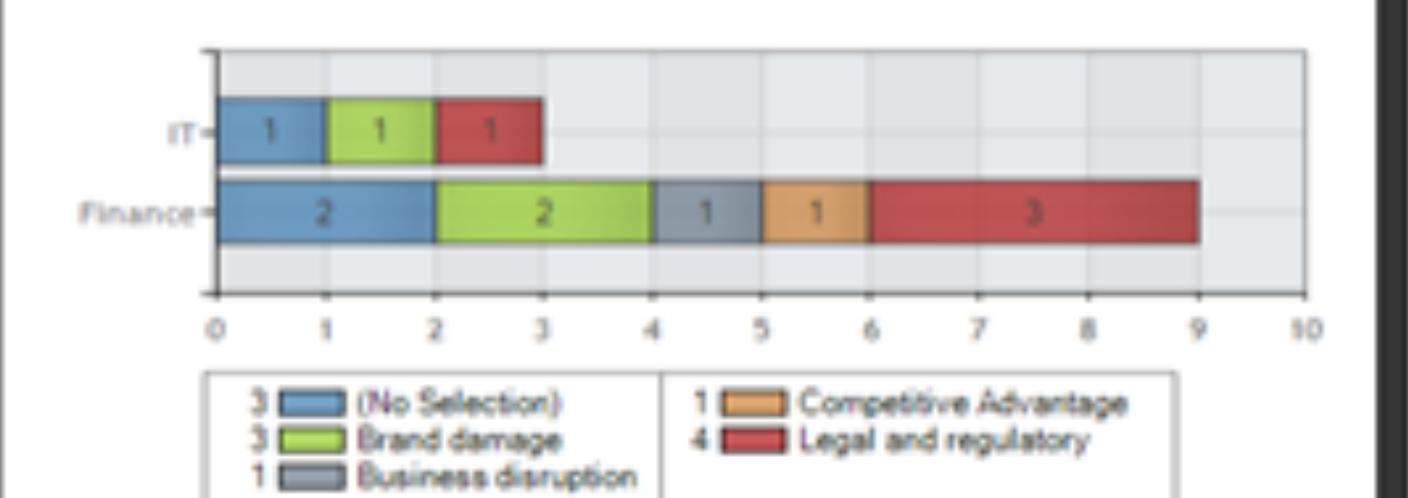
Welcome, Jim SOC Manager

Security Alerts by Facility



Impact Metrics

Incidents by Business Impact



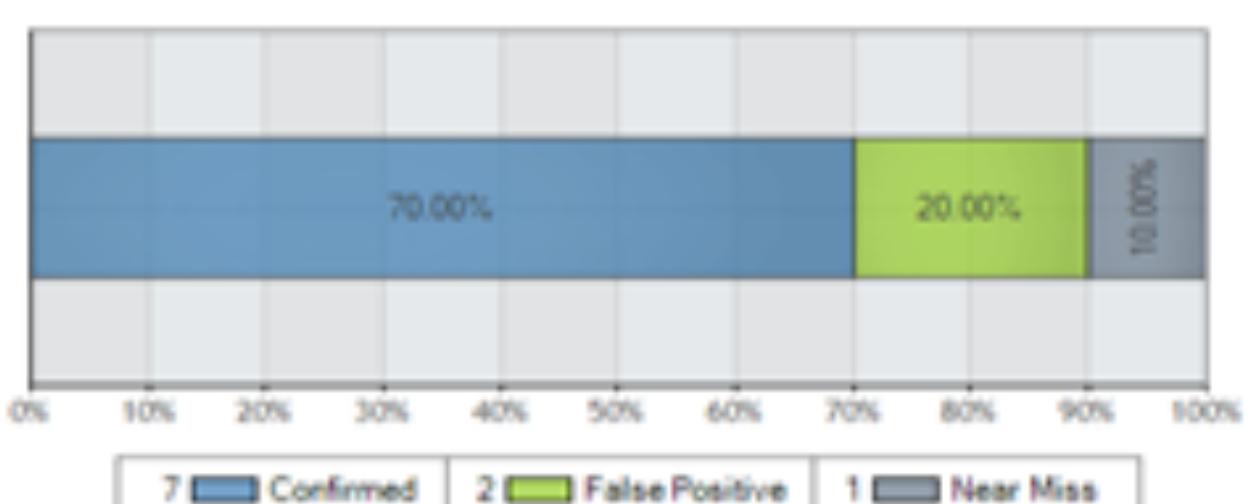
Data Breaches

ID	Date of Breach	Breach Name	Breach Type	Breach Status
DB-205533	10/3/2013 10:30 AM	PII Disclosed	Unauthorized Disclosure	Active

Page 1 of 1 (1 records)

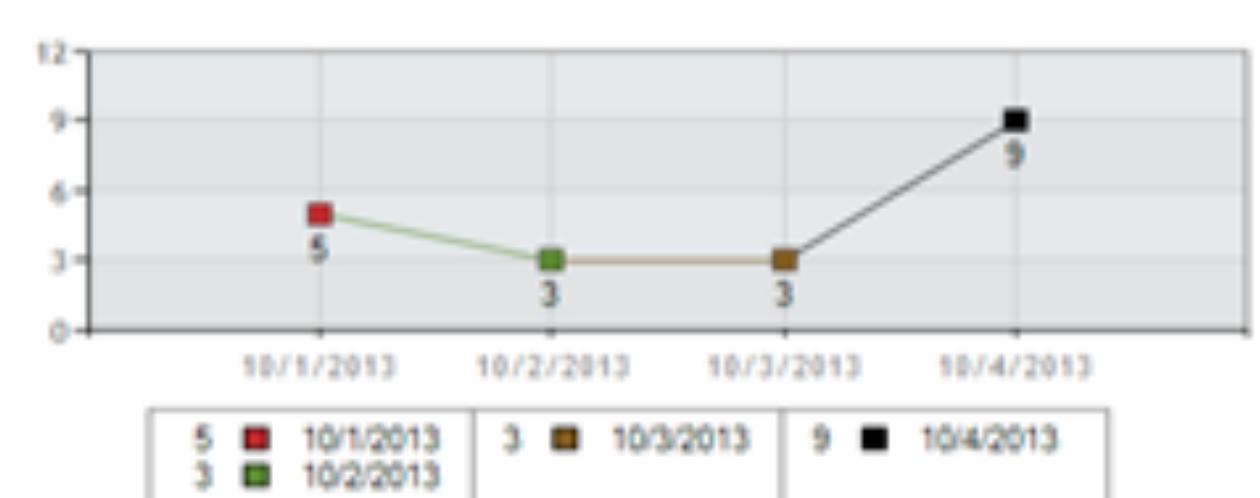
Incident Metrics

Incidents by Confirmation



Trend for All Queues

Incident Trends - Daily



Declared Incidents

ID	Incident Summary	Priority	Threat Category	Threat Valid	Incident Status
INC-6	Beaconing activity grouped by destination IP: 202.111.175.31	P-1	Malware C2	Suspected	Escalated

Page 1 of 1 (1 records)

Navigation Menu

Incident Response

- Security Alerts
- Security Incidents
- Incident Investigations
- Forensic Analysis
- Incident Response Tasks
- Incident Journal

Data Breach Response

- Data Breaches
- Breach Tasks
- Breach Risk Assessment
- Notifications & Call Trees
- Notification History

SOC Program Management

- Shift Handover
- SOC Procedure Library
- Security Controls
- SOC Policies
- Contacts
- Teams
- Team Members
- Competencies
- Training & CPE Tracking

Issue Management

- Findings
- Remediation Plans
- Exception Requests
- Policy Change Requests

Dashboard: SOC Manager

Welcome, Jim SOC Manager

Security Alerts by External Destination

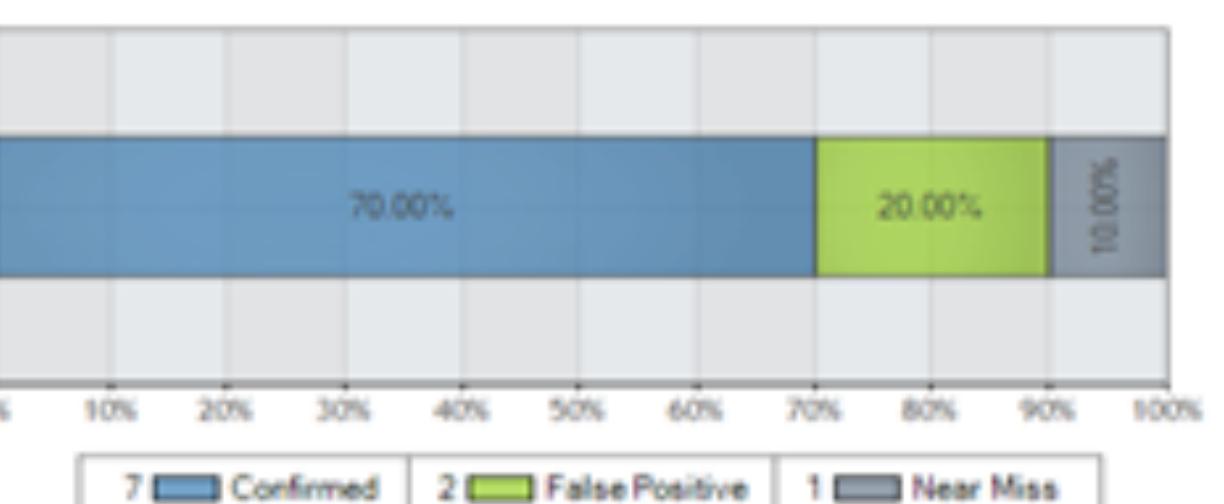


Security Alerts by Facility



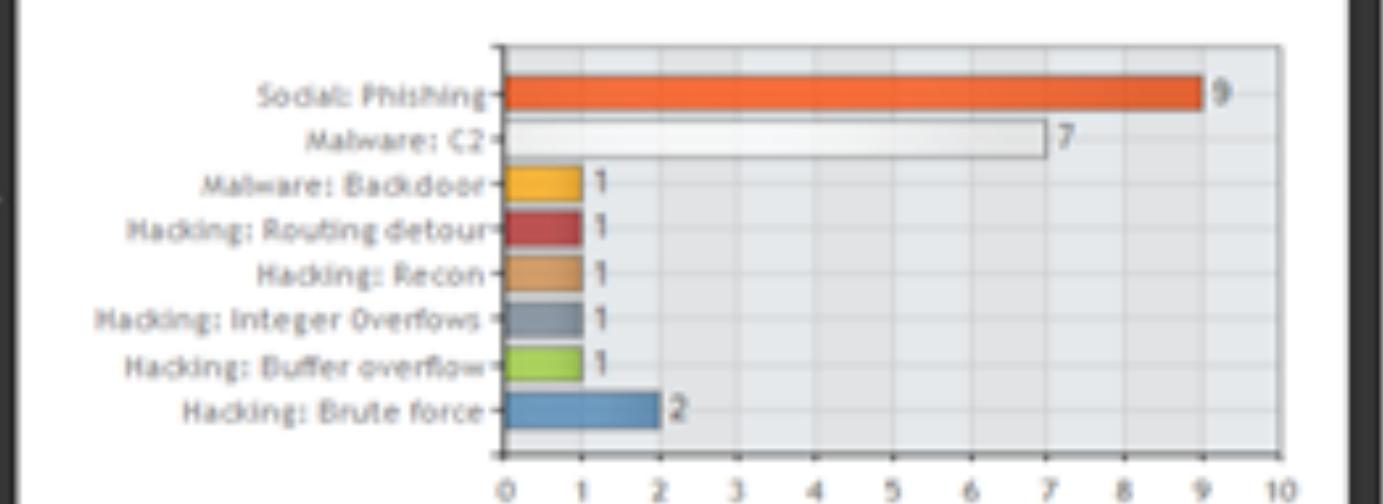
Incident Metrics

Incidents by Confirmation



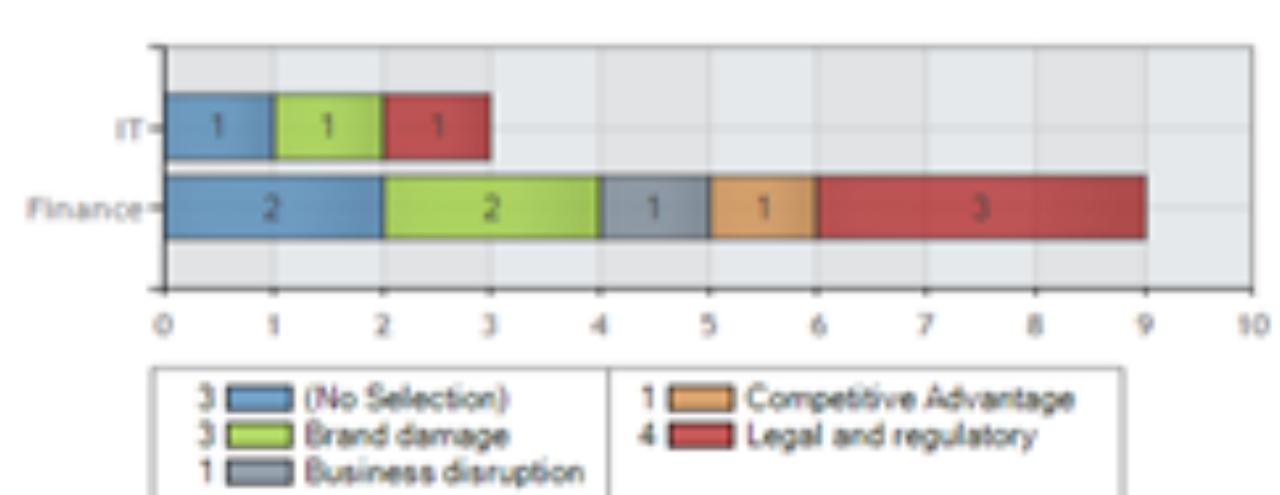
Threat Metrics

Incidents by Threat Categories



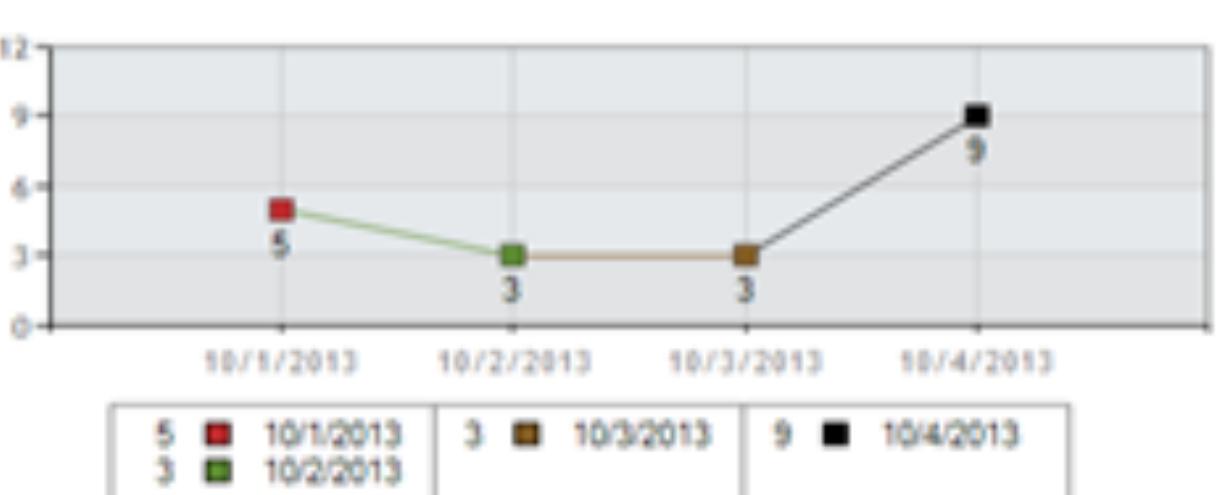
Impact Metrics

Incidents by Business Impact

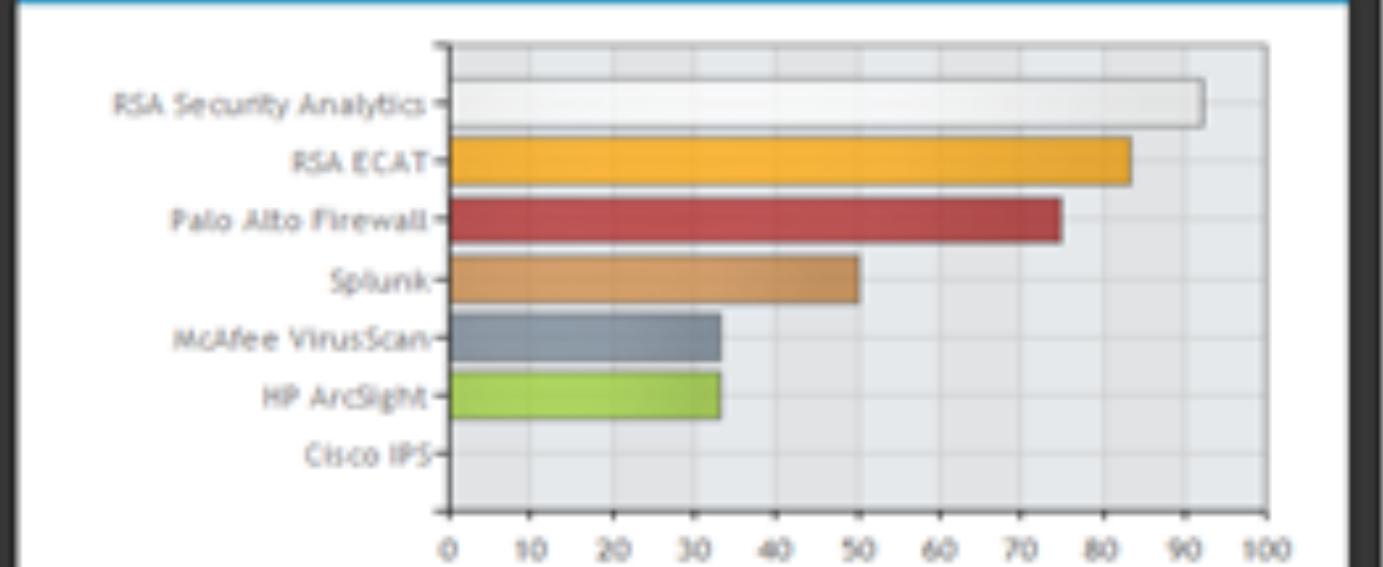


Trend for All Queues

Incident Trends - Daily



Security Tools Efficacy



Data Breaches

ID	Date of Breach	Breach Name	Breach Type	Breach Status
DB-205533	10/3/2013 10:30 AM	PII Disclosed	Unauthorized Disclosure	Active

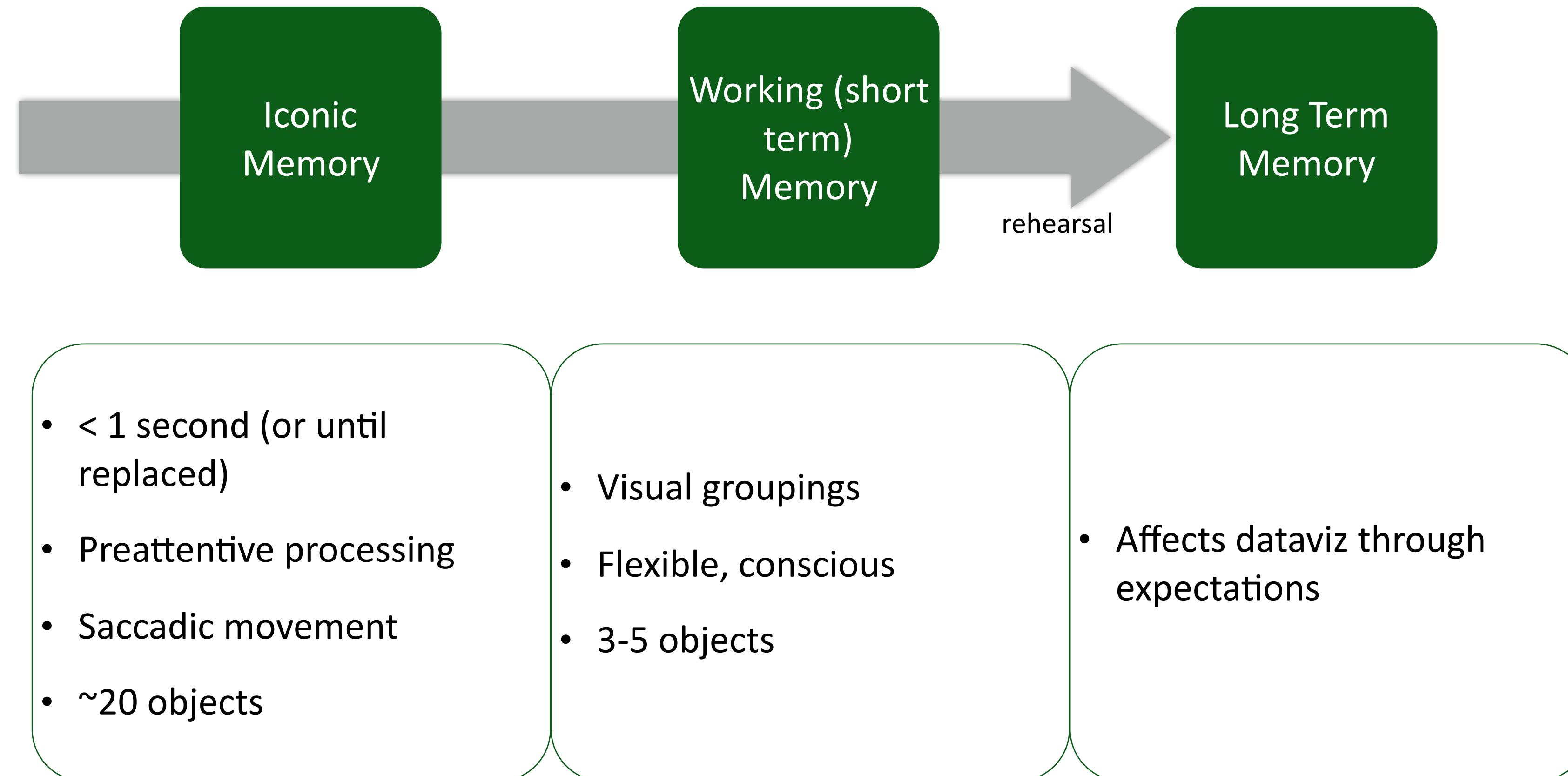
Page 1 of 1 (1 records)

Declared Incidents

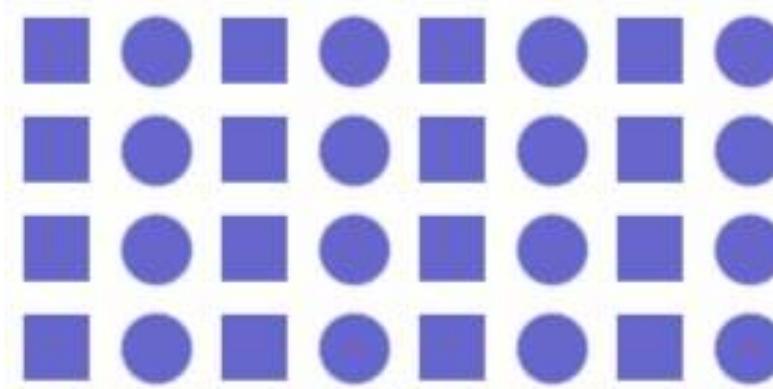
ID	Incident Summary	Priority	Threat Category	Threat Valid	Incident Status
INC-6	Beaconing activity grouped by destination IP: 202.111.175.31	P-1	Malware C2	Suspected	Escalated

Page 1 of 1 (1 records)

Human Visual Processing



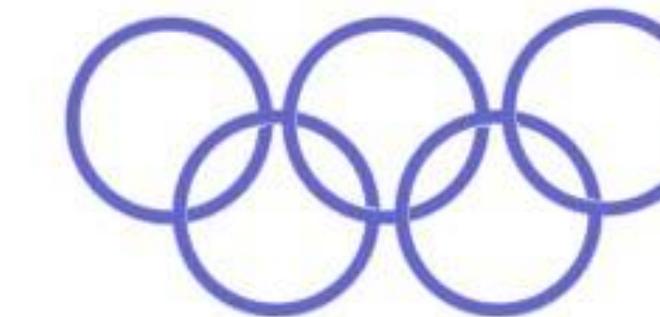
Gestalt Laws



Law of Similarity:

Items that are similar tend to be grouped together.

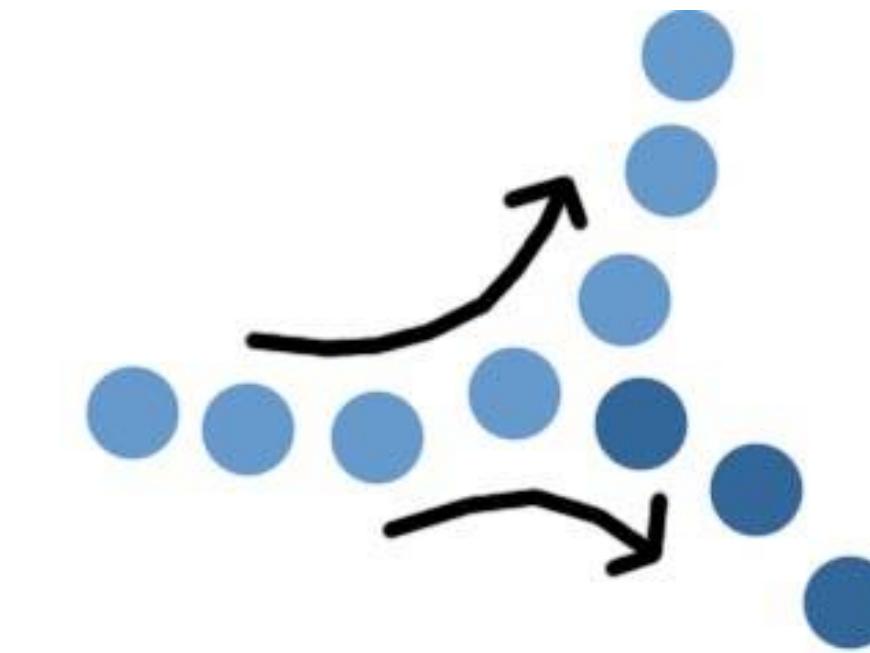
In the image above, most people see vertical columns of circles and squares.



Law of Pragnanz:

Reality is organized or reduced to the simplest form possible.

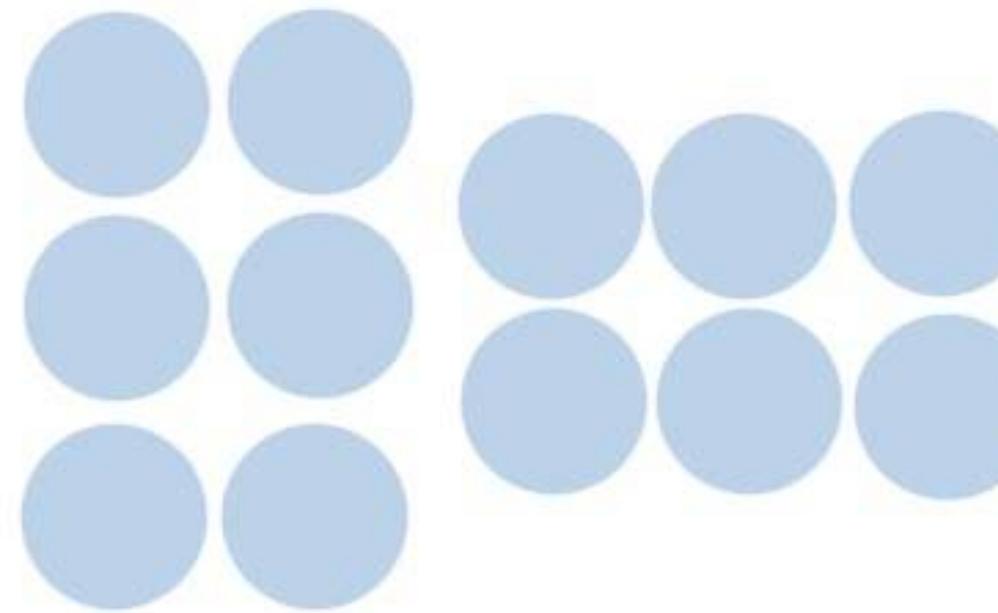
For example, we see the image above as a series of circles rather than as many much more complicated shapes.



Law of Continuity:

Lines are seen as following the smoothest path.

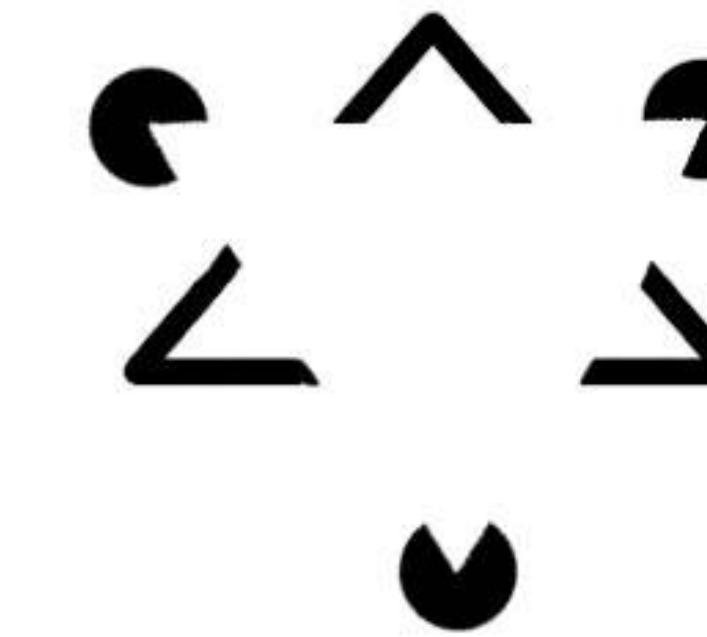
In the image above, the top branch is seen as continuing the first segment of the line. This allows us to see things as flowing smoothly without breaking lines up into multiple parts.



Law of Proximity:

Objects near each other tend to be grouped together.

The circles on the left appear to be grouped in vertical columns, while those on the right appear to be grouped in horizontal rows.



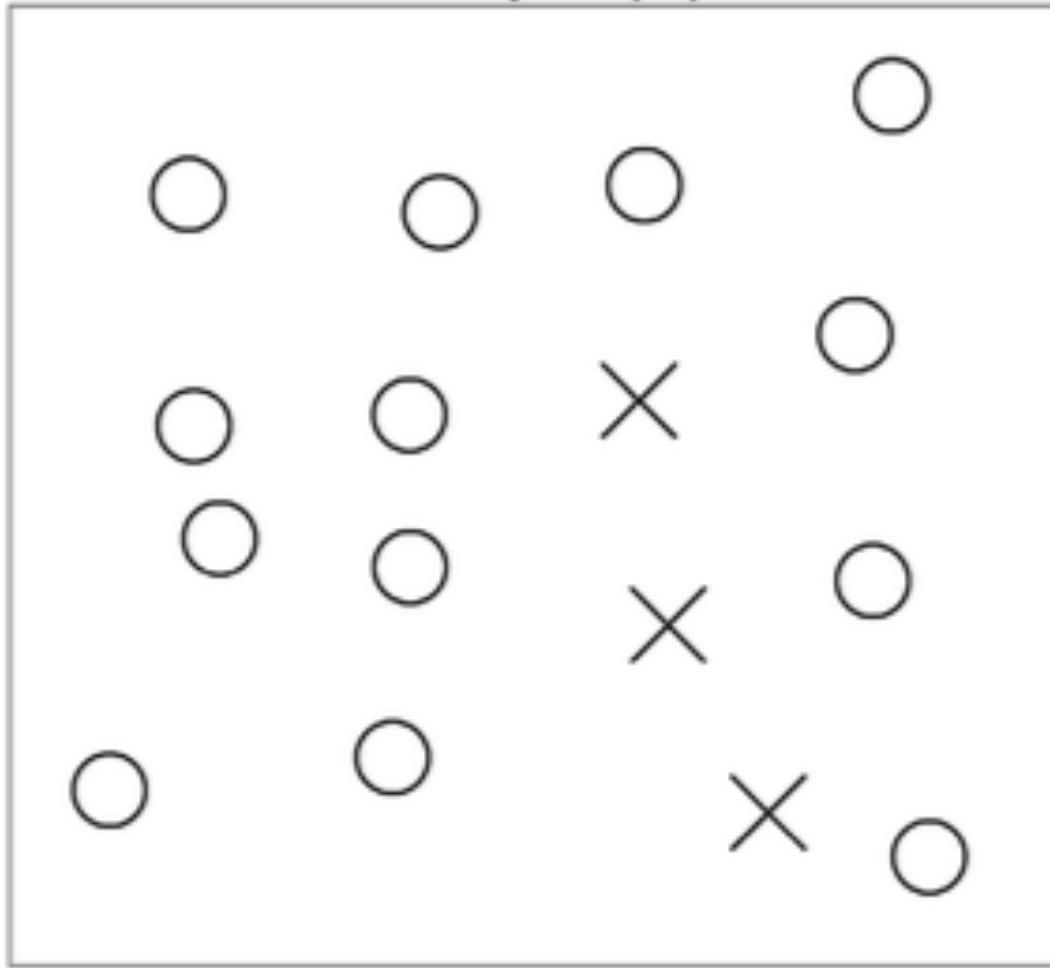
Law of Closure:

Objects grouped together are seen as a whole.

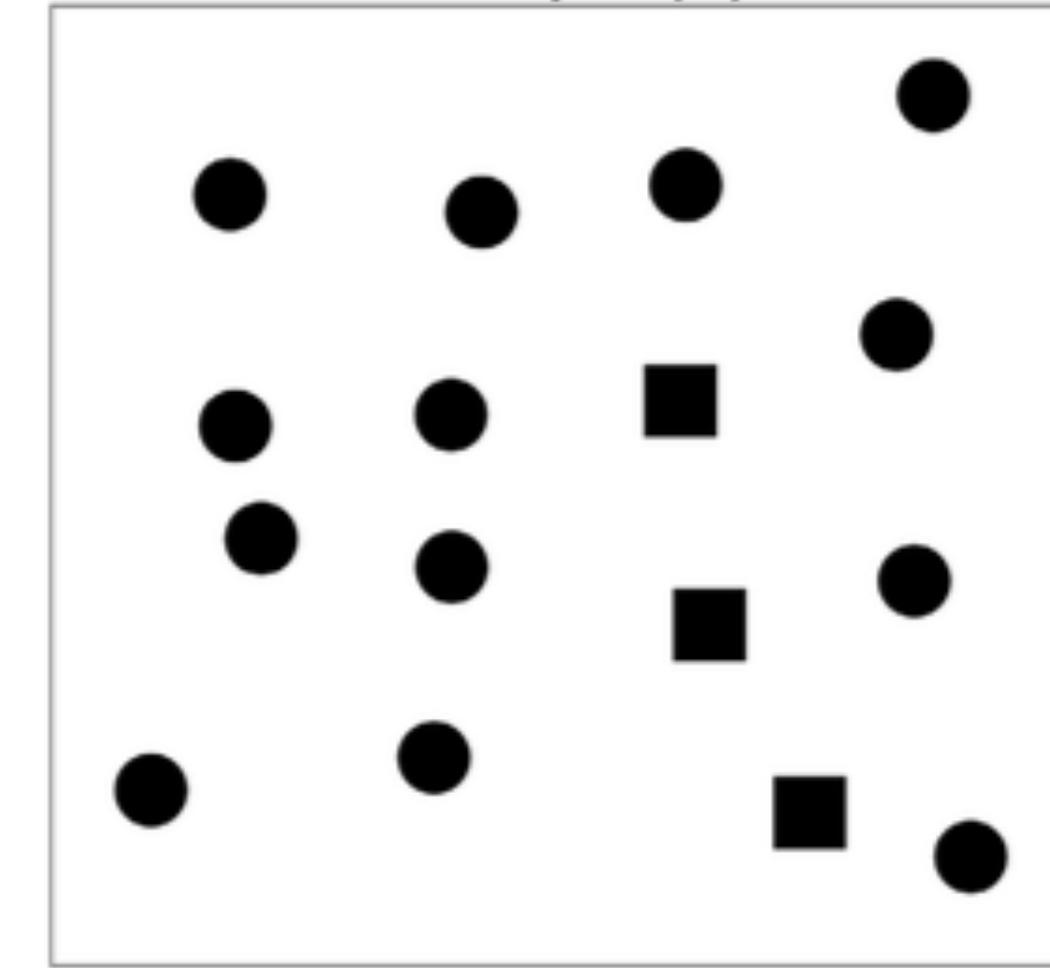
We tend to ignore gaps and complete contour lines. In the image above, there are no triangles or circles, but our minds fill in the missing information to create familiar shapes and images.

Groupings

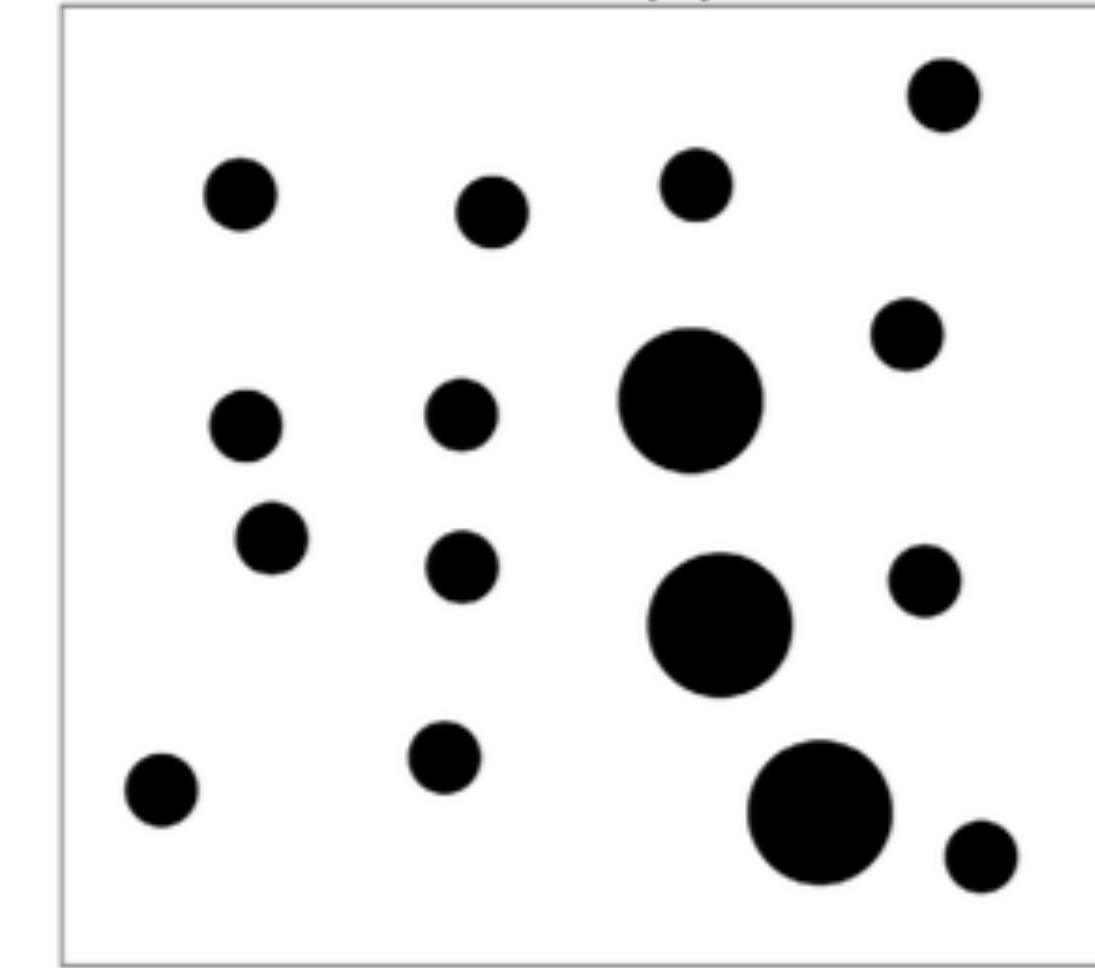
Shape (a)



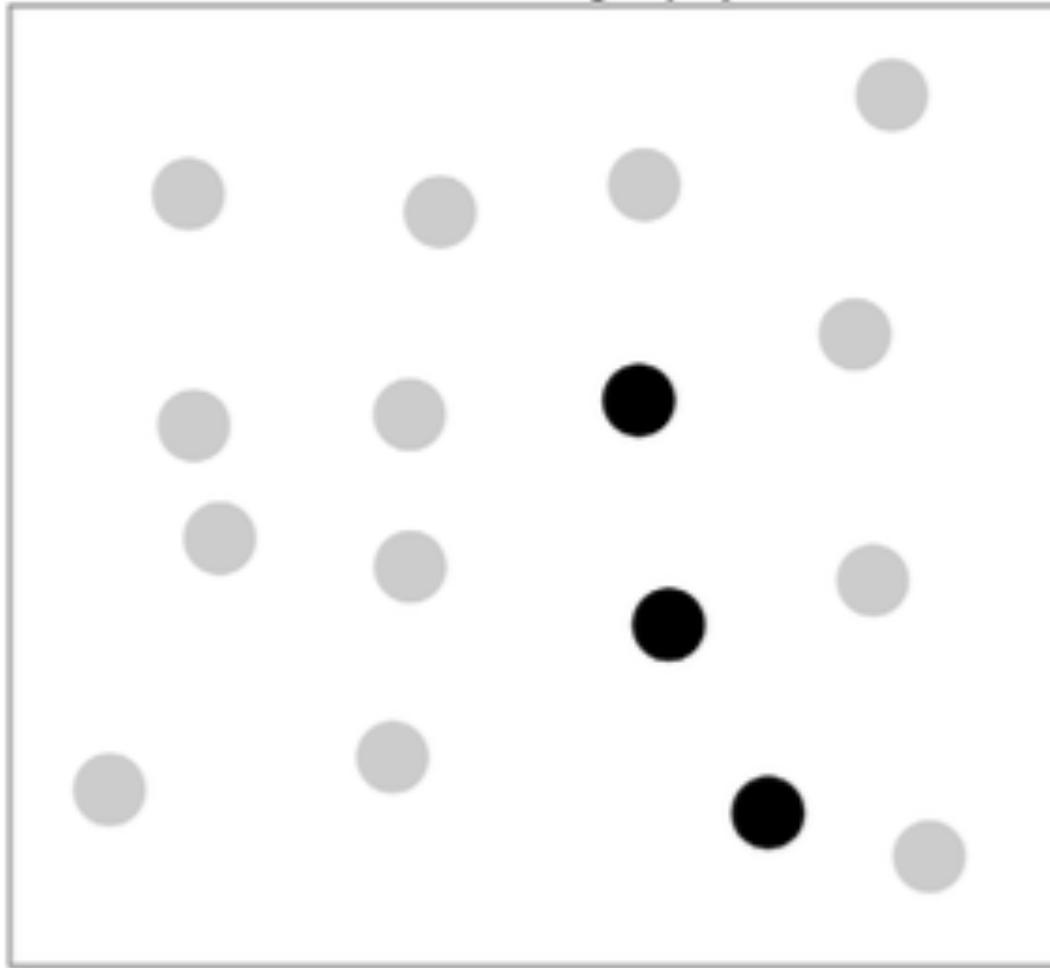
Shape (b)



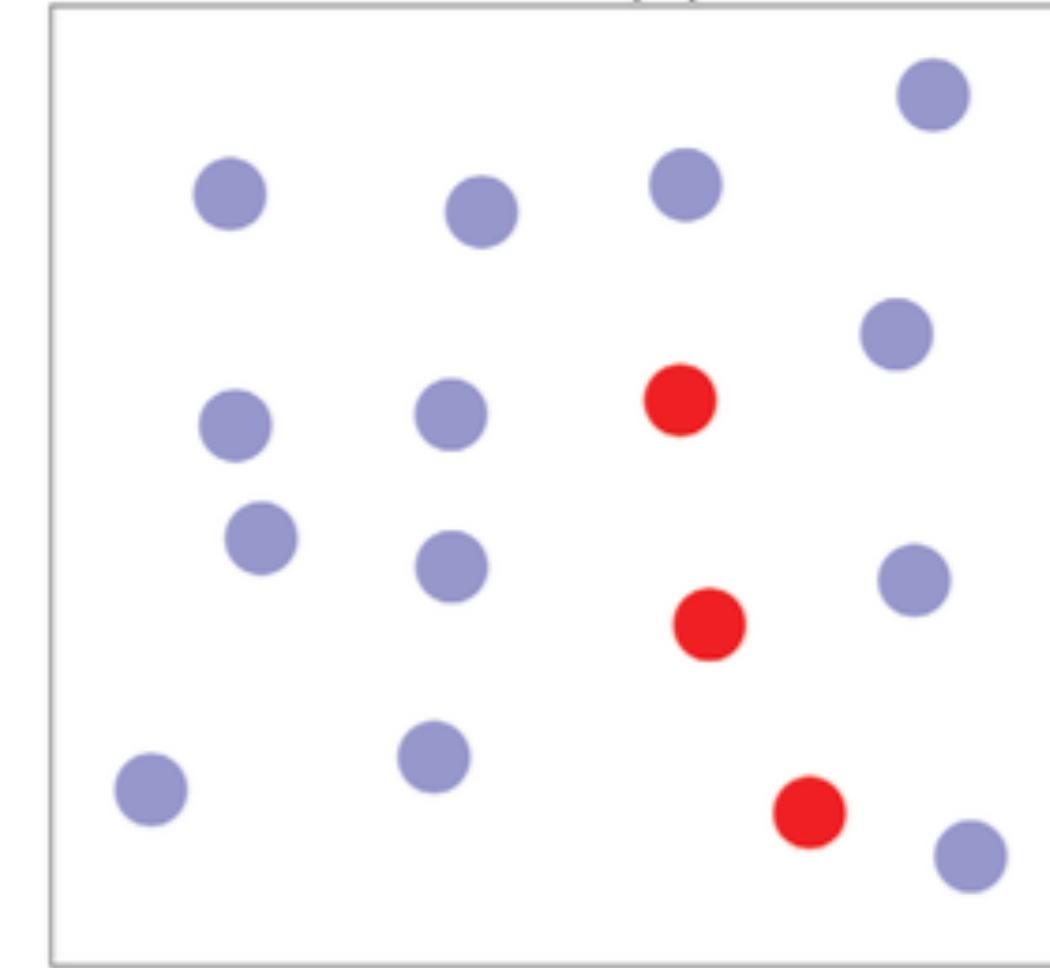
Size (c)



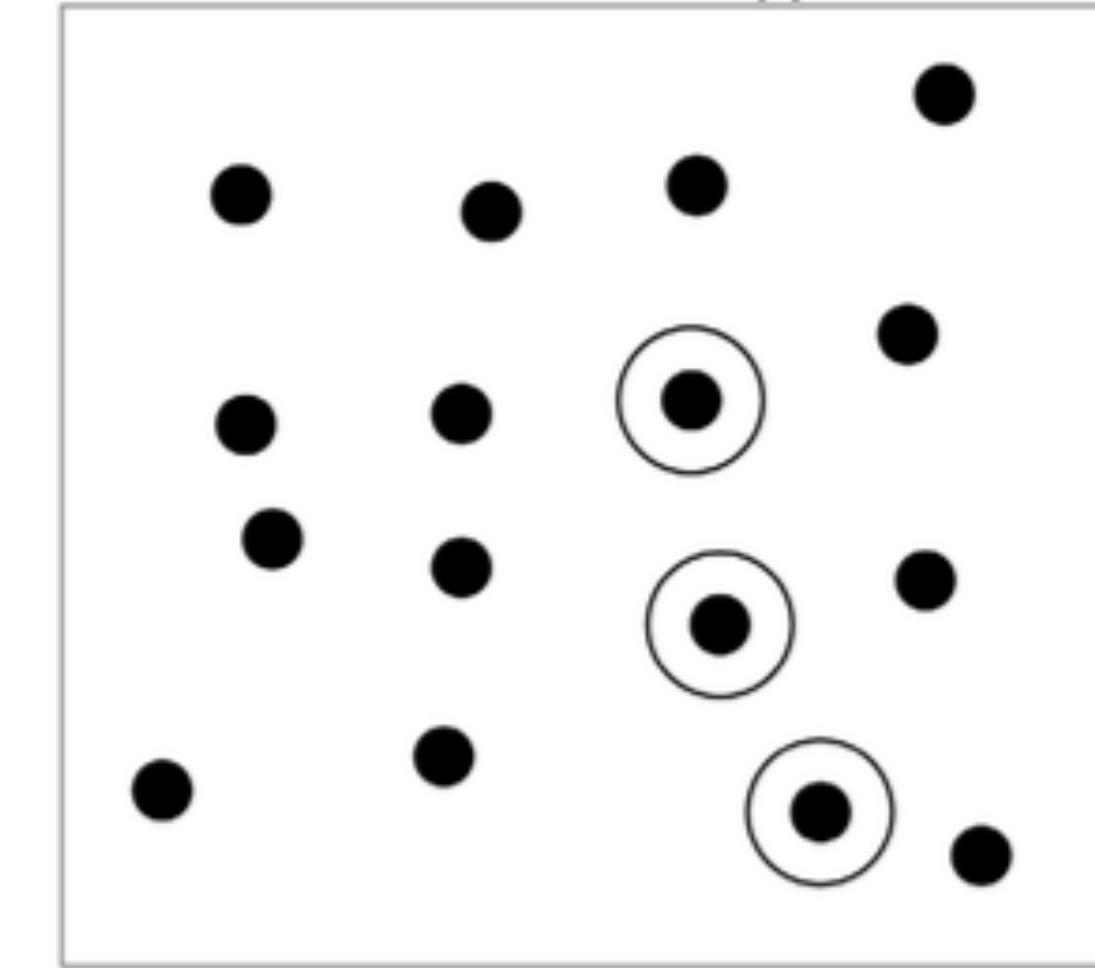
Intensity (d)



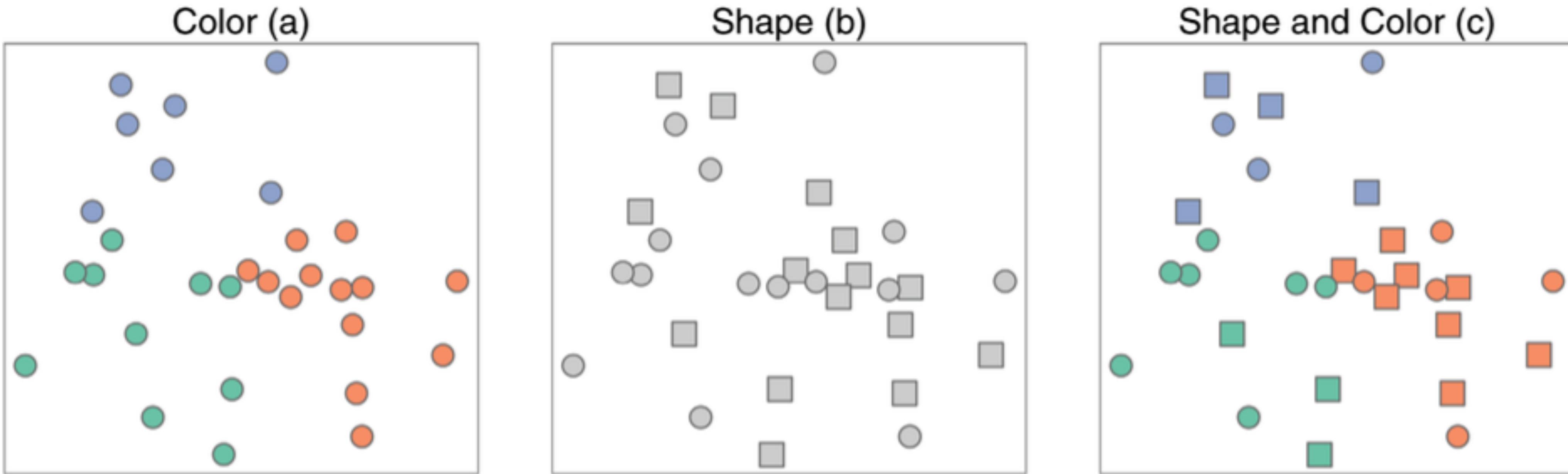
Hue (e)



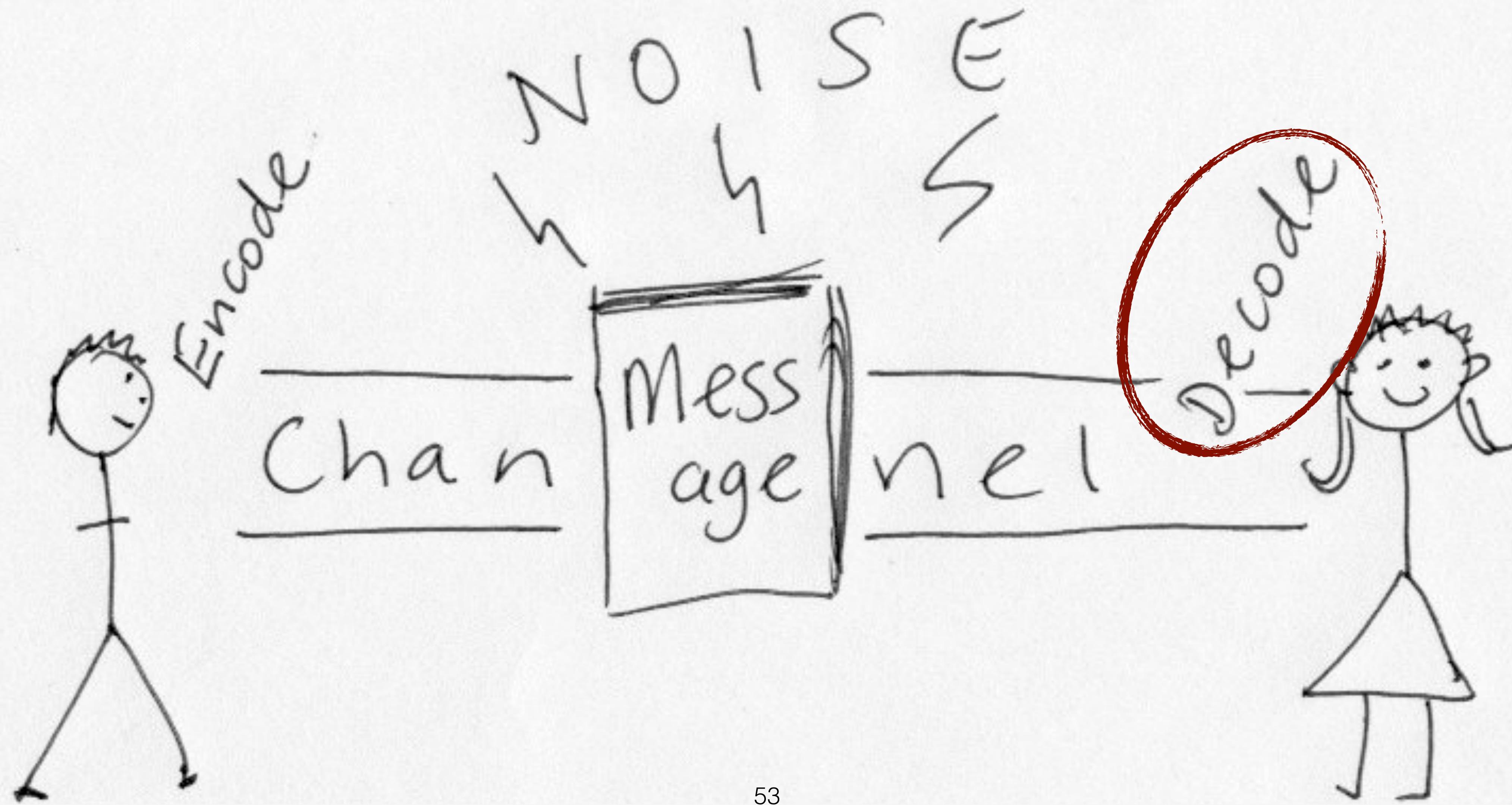
Enclosure (f)



Too many encodings...

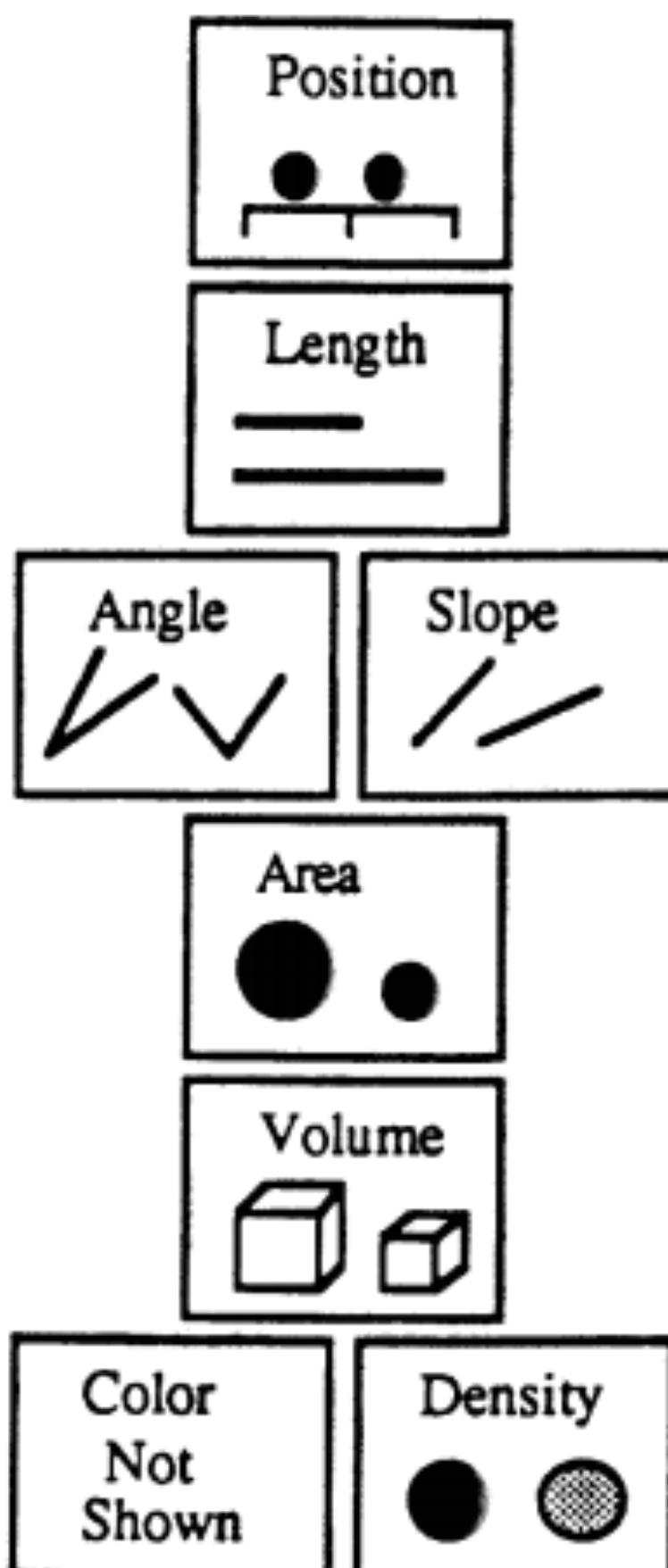


Targeting the Audience



Ease of Decoding

More accurate

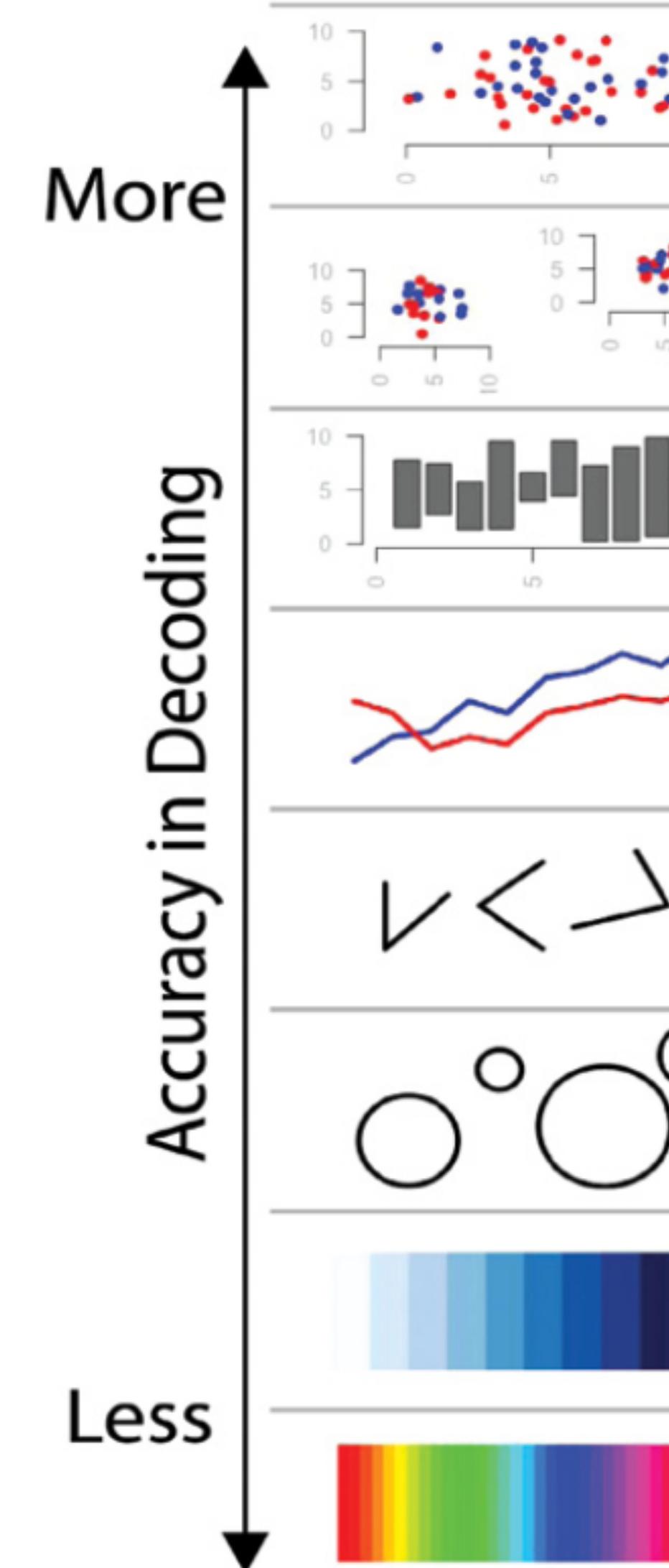


Less accurate

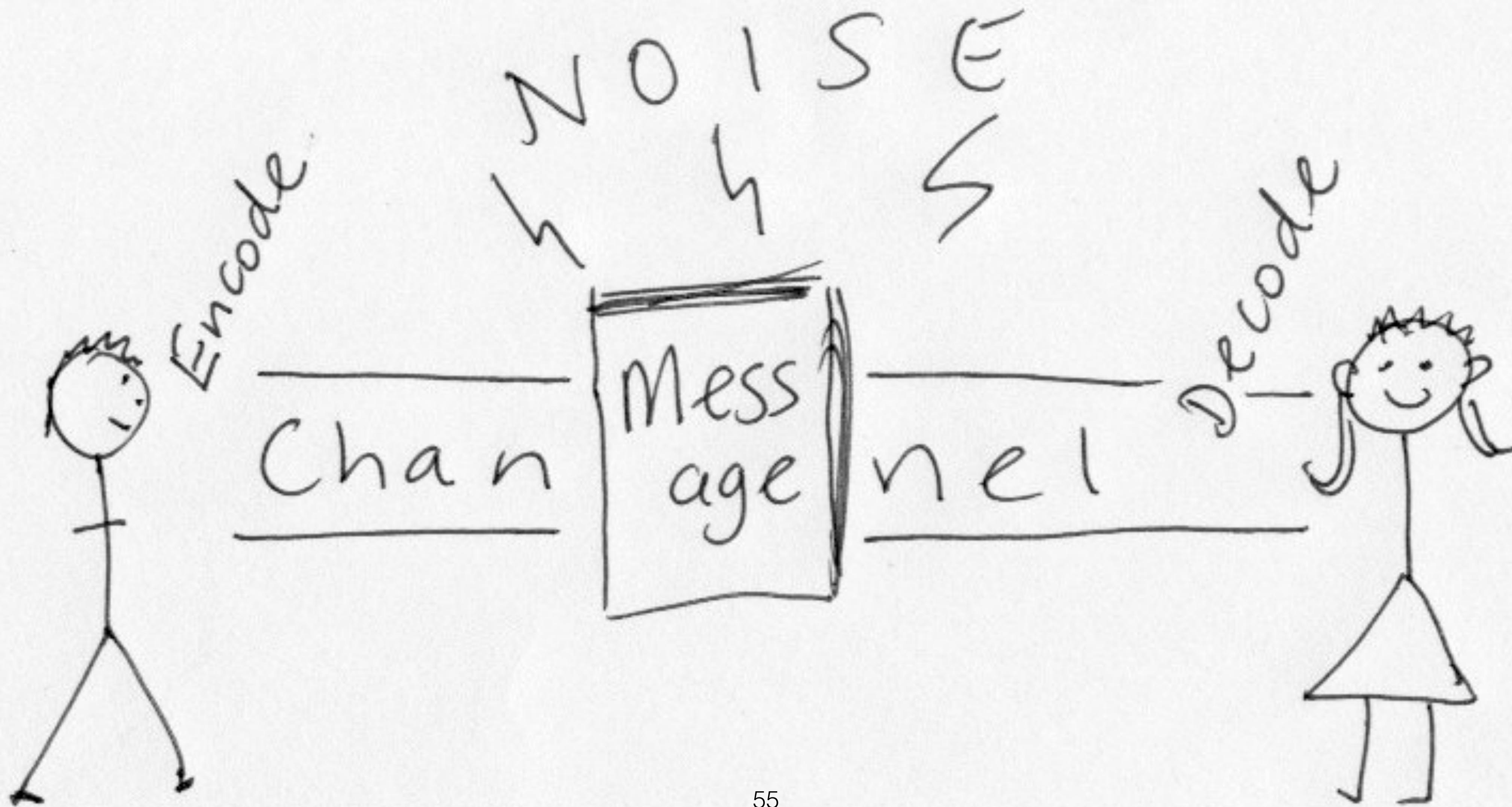
"Automating the Design of Graphical Presentations of Relational Information", J. D. Mackinlay

More

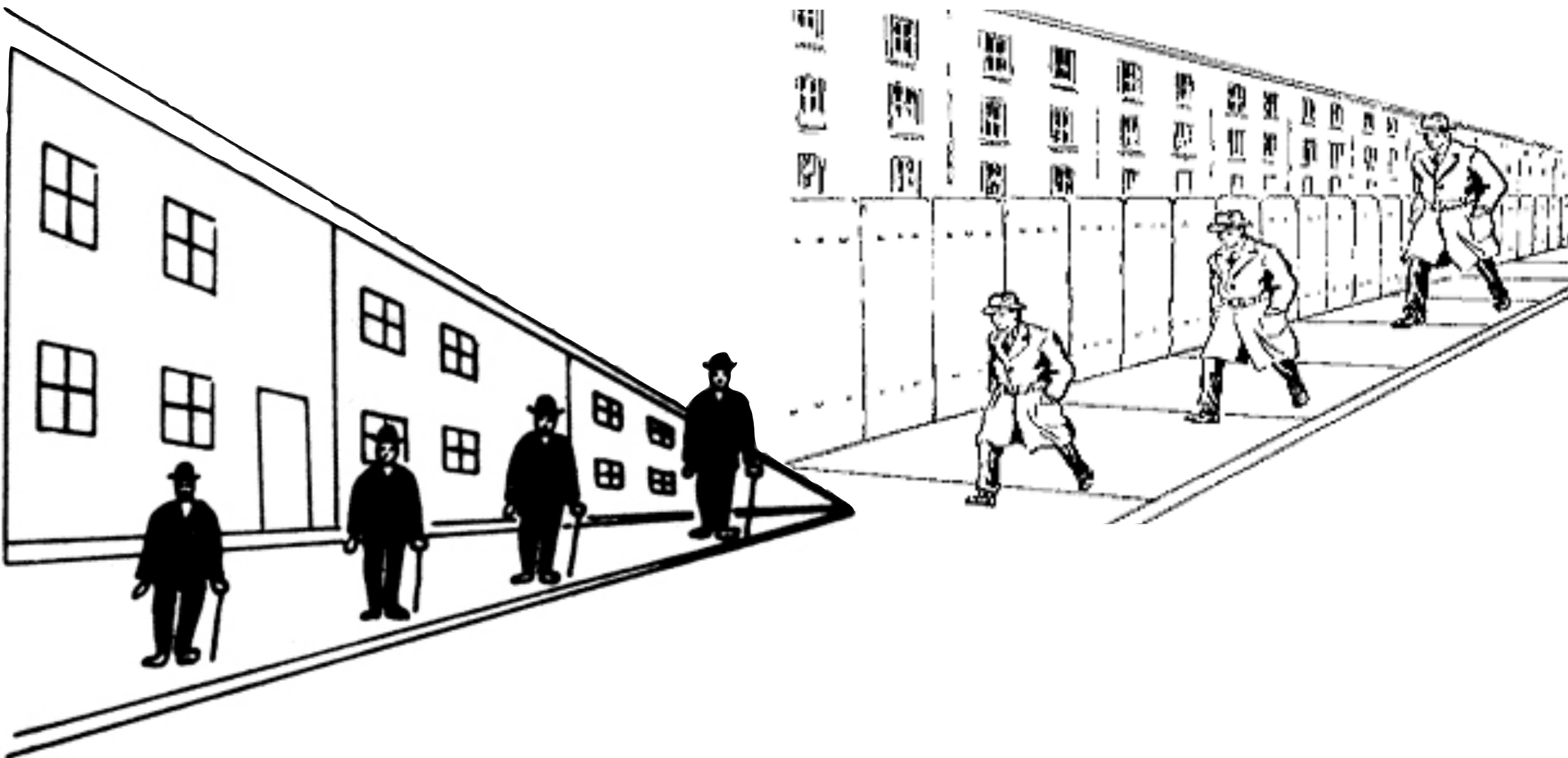
Less



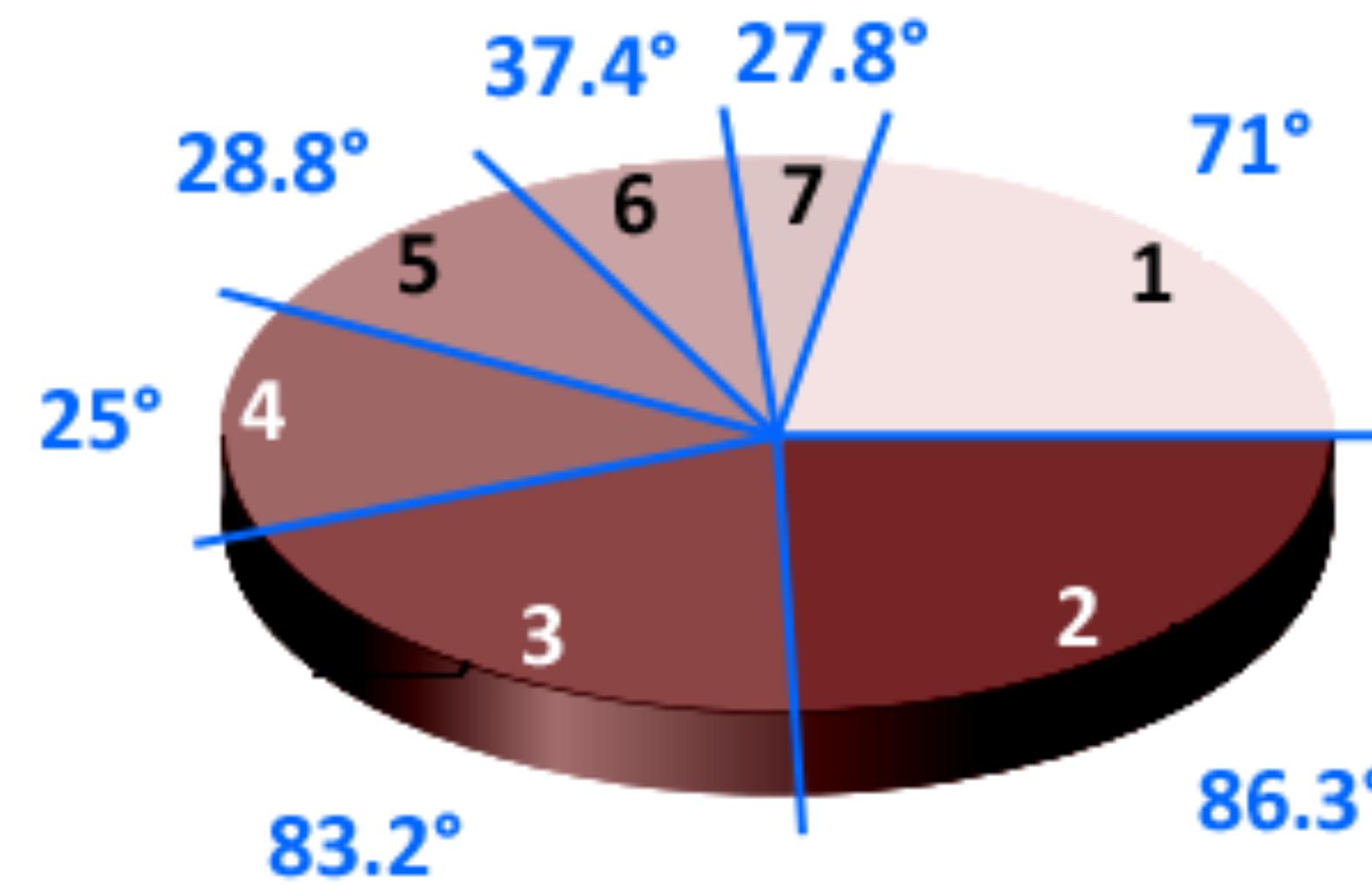
Everything Else



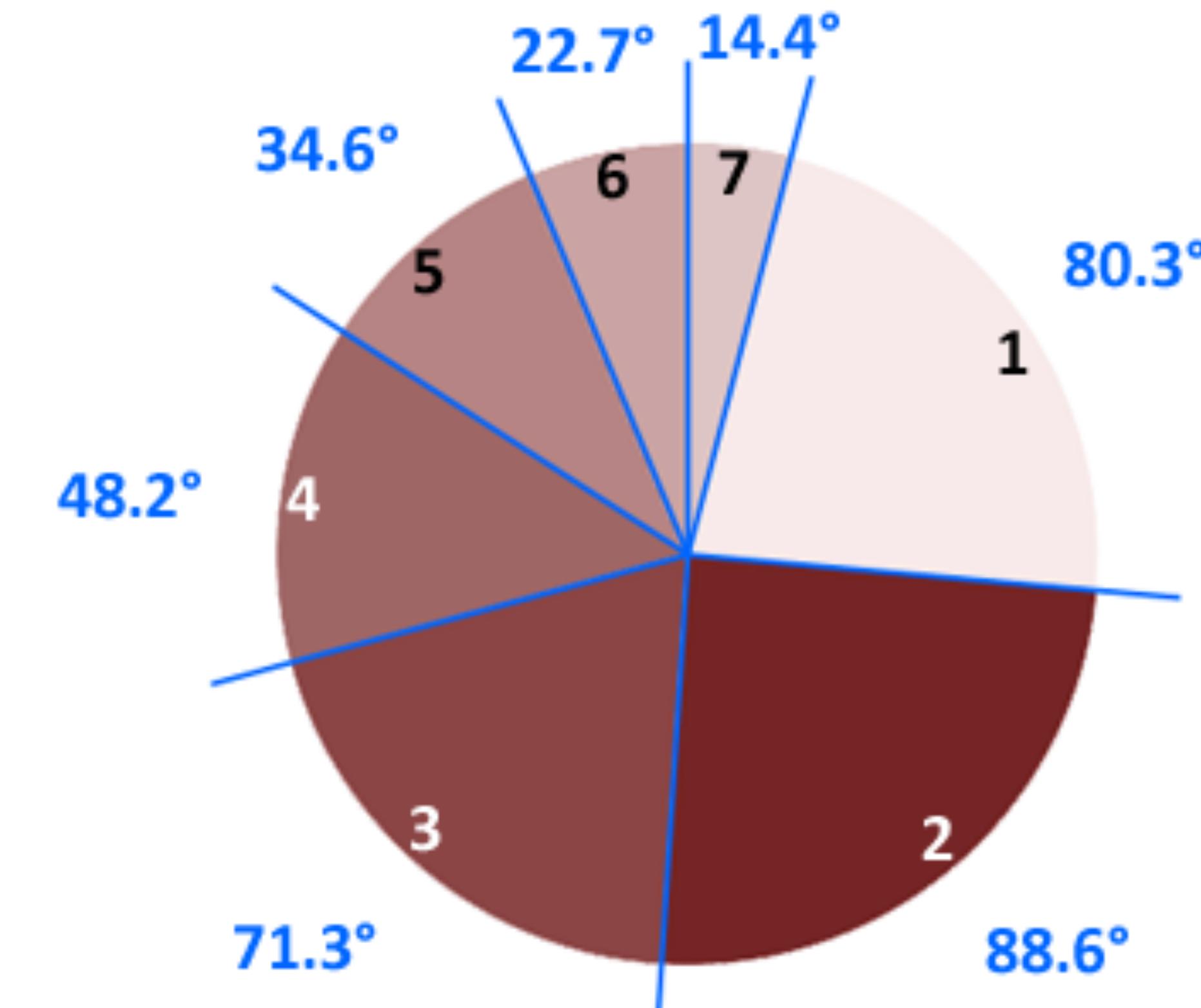
Avoid 3D Perspectives



3D Pie Charts Distort Angles

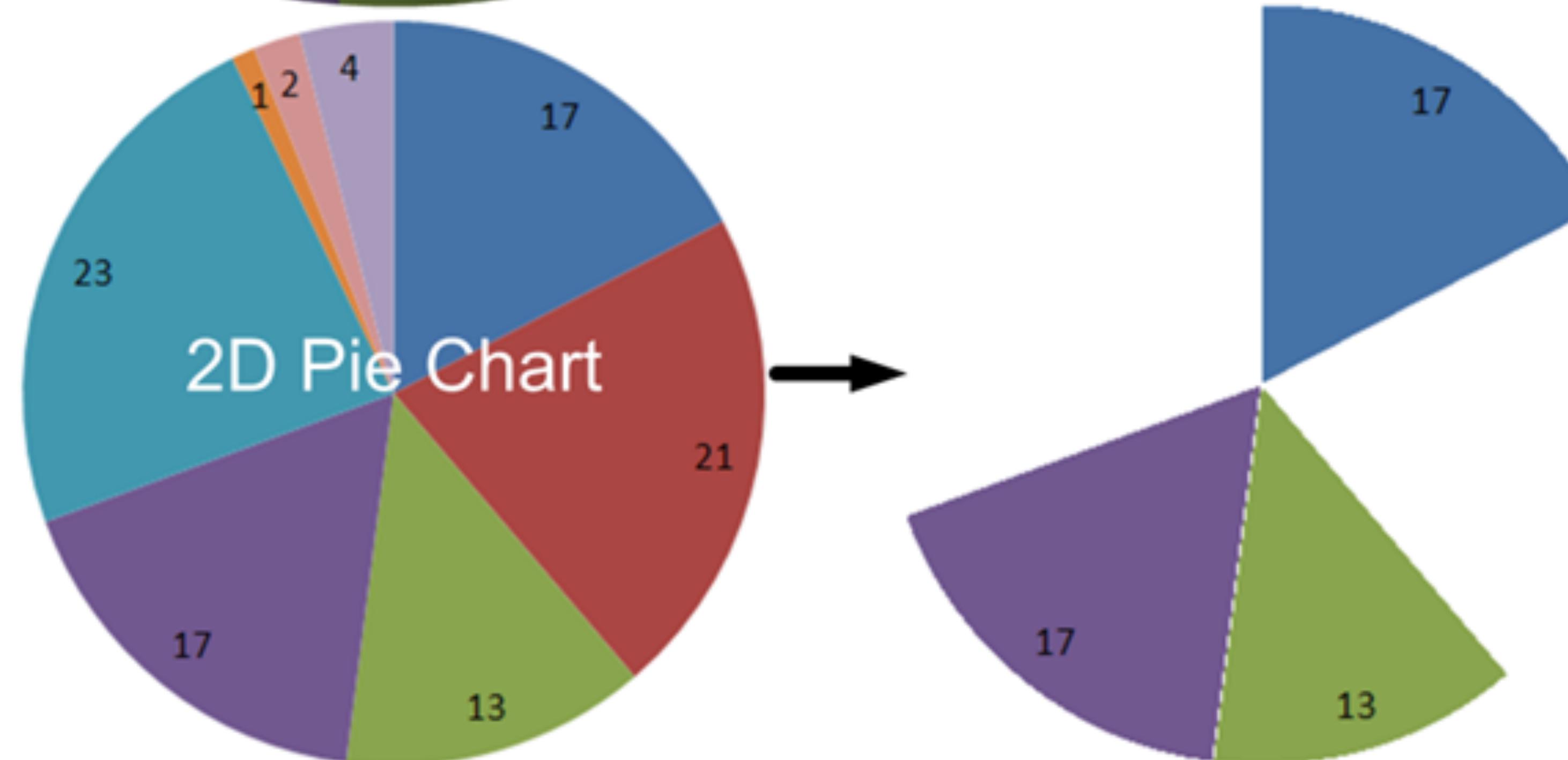
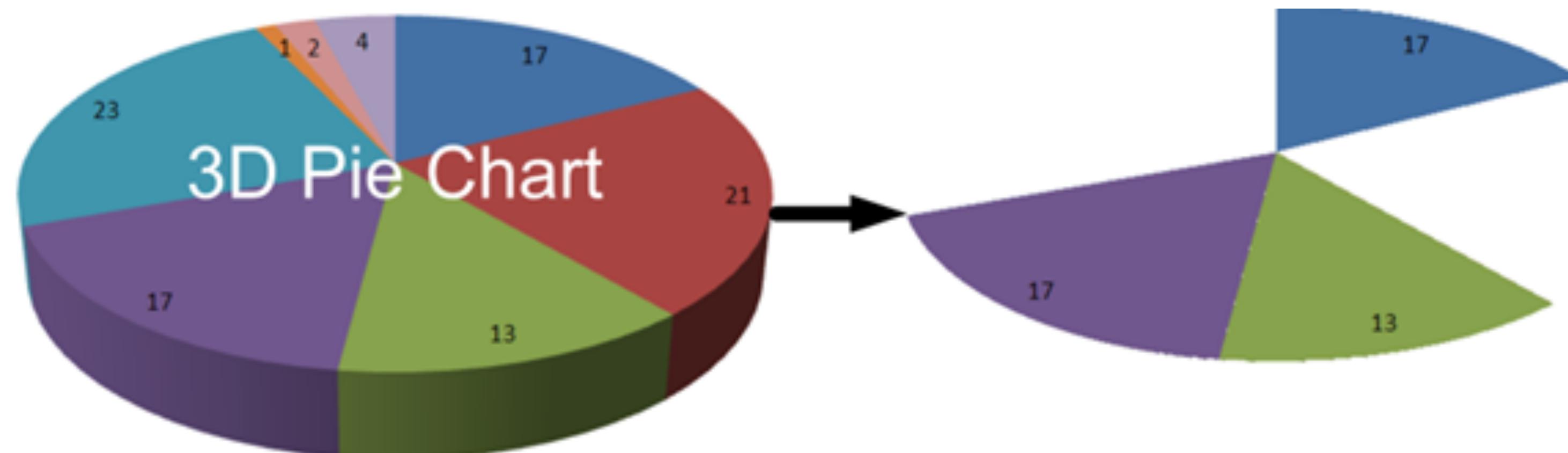


Angles on the original pie chart



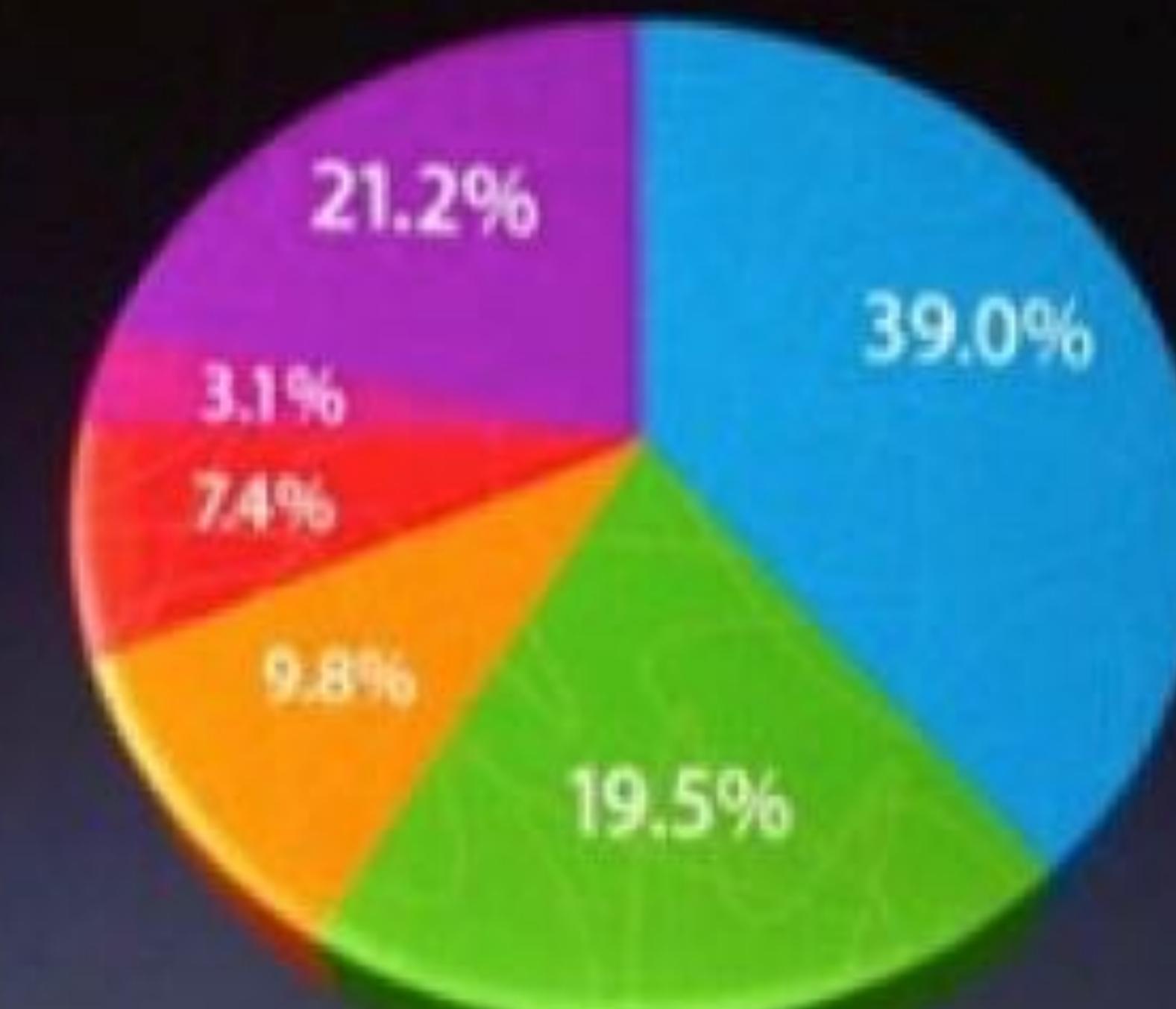
Angles on a non-3D pie chart

3D Pie Charts Distort Area



U.S. SmartPhone Marketshare

- RIM
- Apple
- Palm
- Motorola
- Nokia
- Other



Gartner

3D Perspectives

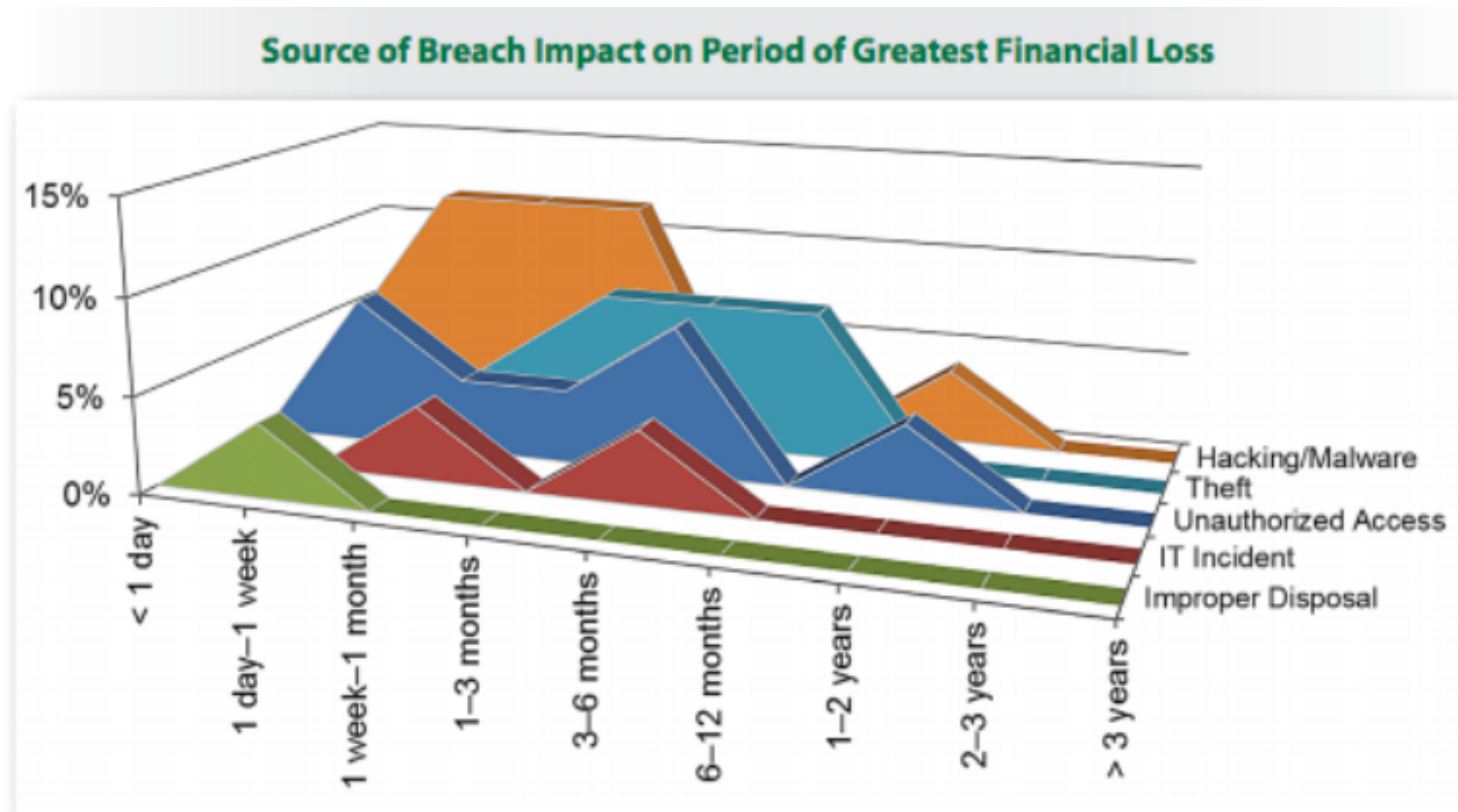


Figure 8. Source of Breach Effect on Period of Greatest Financial Loss

Setting up for visualization

```
import pandas as pd  
import matplotlib.pyplot as plt  
import numpy as np  
%matplotlib inline  
pd.options.display.mpl_style = 'default'
```

```
library(tidyverse)
```

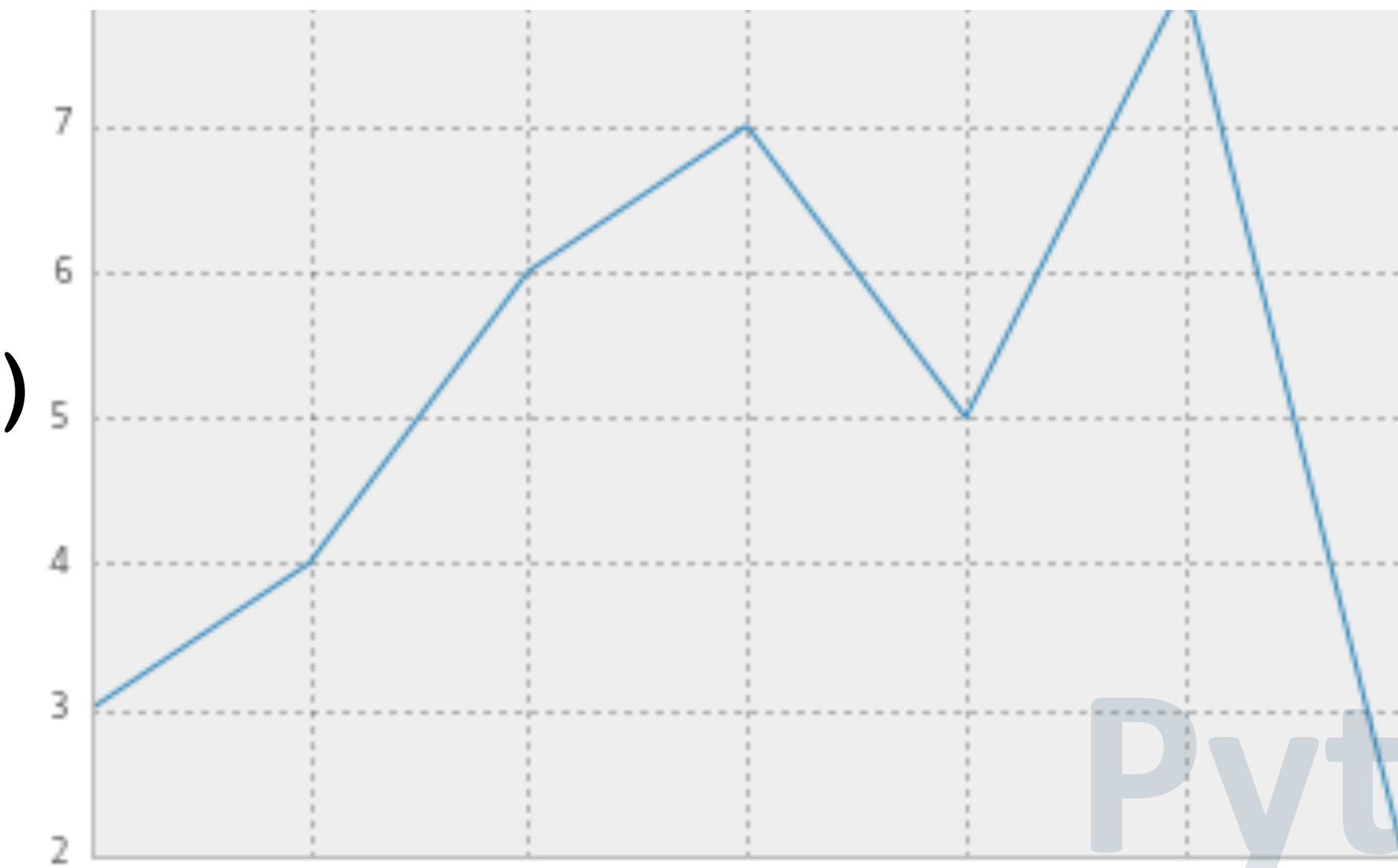
ggplot2, for data visualisation.
dplyr, for data manipulation.
tidyR, for data tidying.
readr, for data import.
purrr, for functional programming.
tibble, a modern re-imagining of data frames.

Python

R

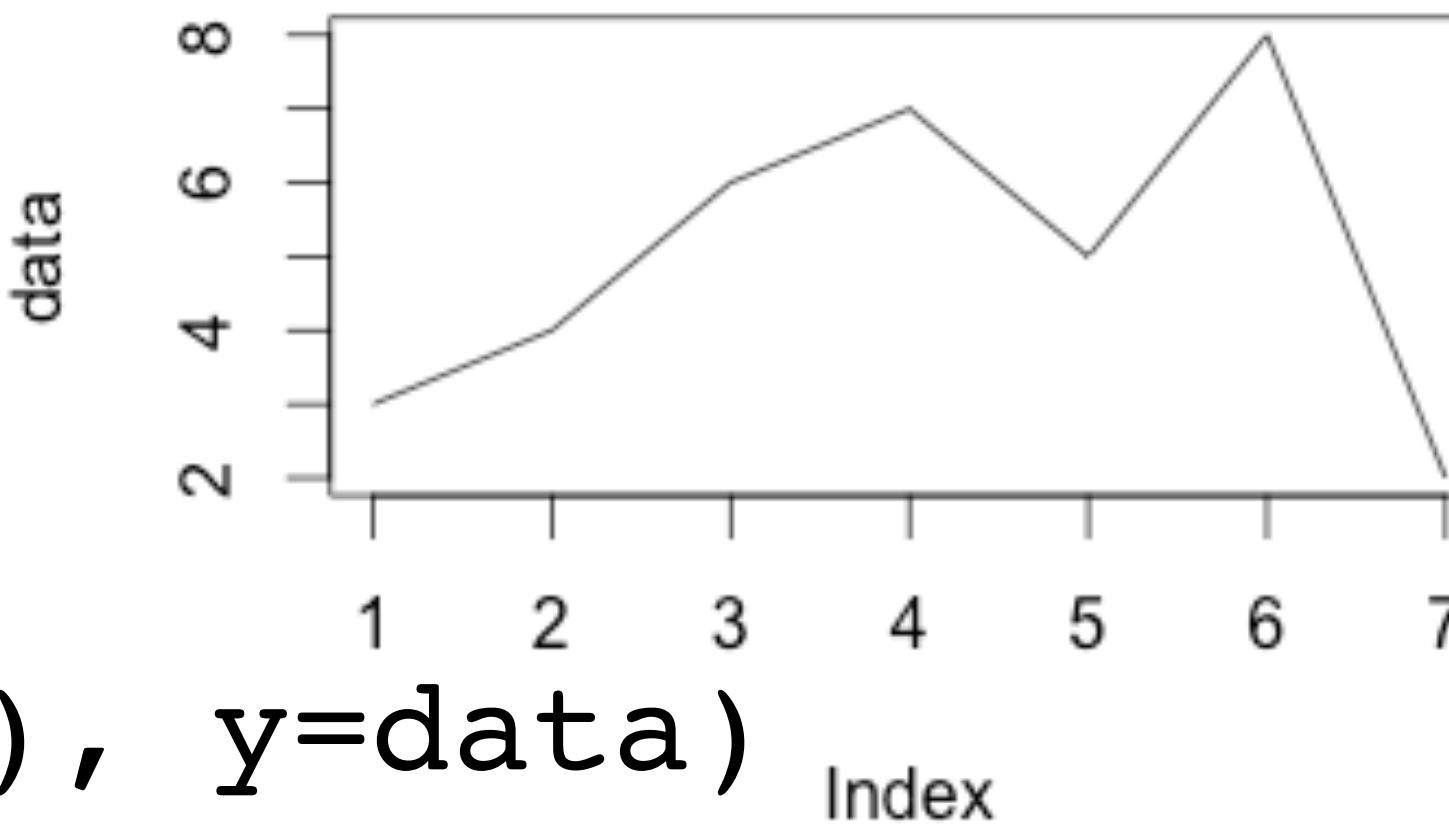
Line Plots

```
data = pd.Series([3,4,6,7,5,8,2])  
graph = data.plot()
```



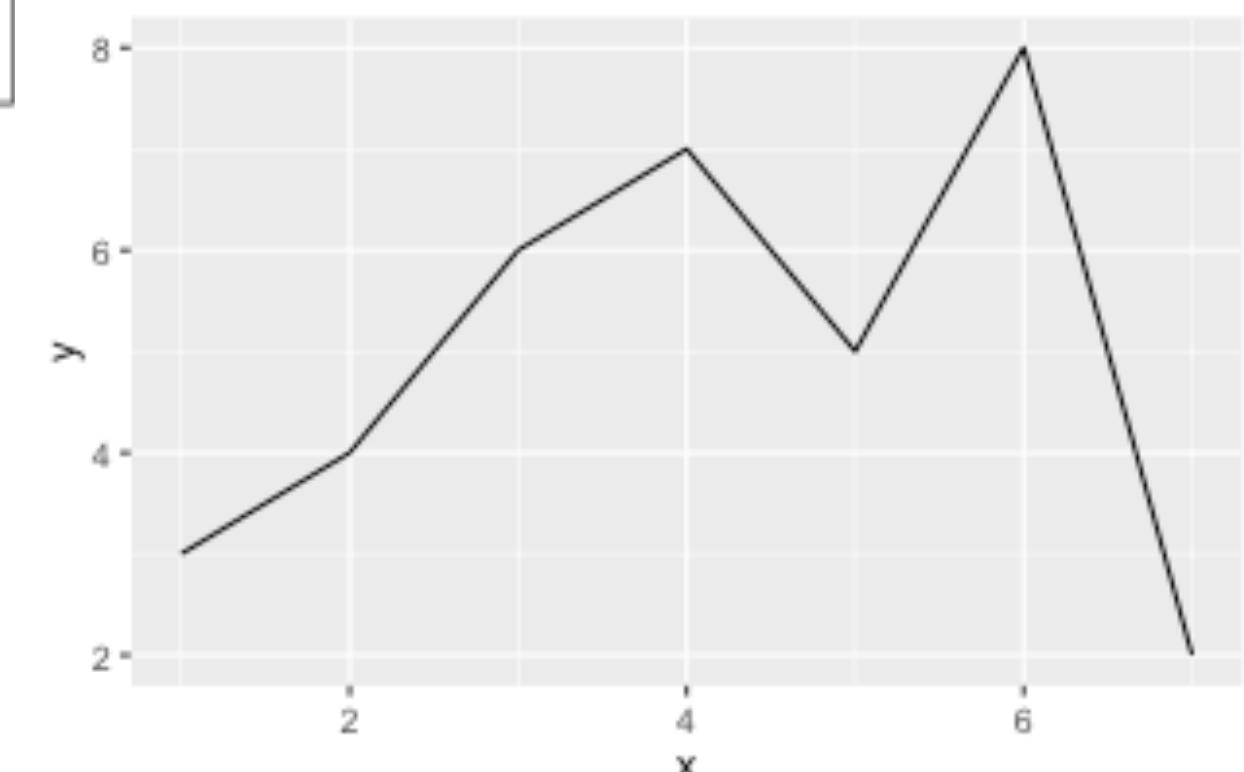
Python

```
data <- c(3,4,6,7,5,8,2)  
plot(data, type="l")
```



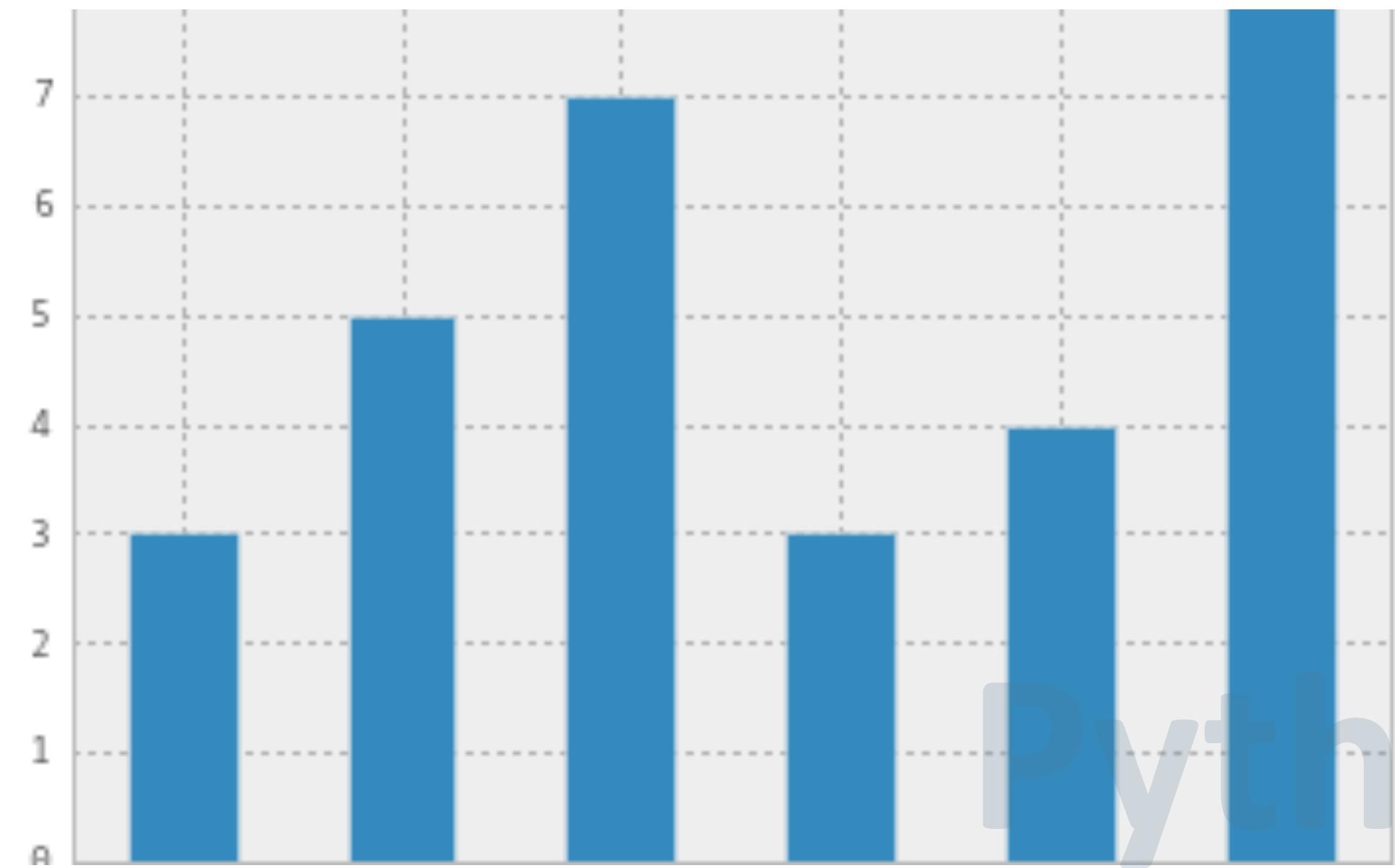
R

```
dataf <- data.frame(x=seq(7), y=data)  
ggplot(dataf, aes(x, y)) +  
geom_line()
```

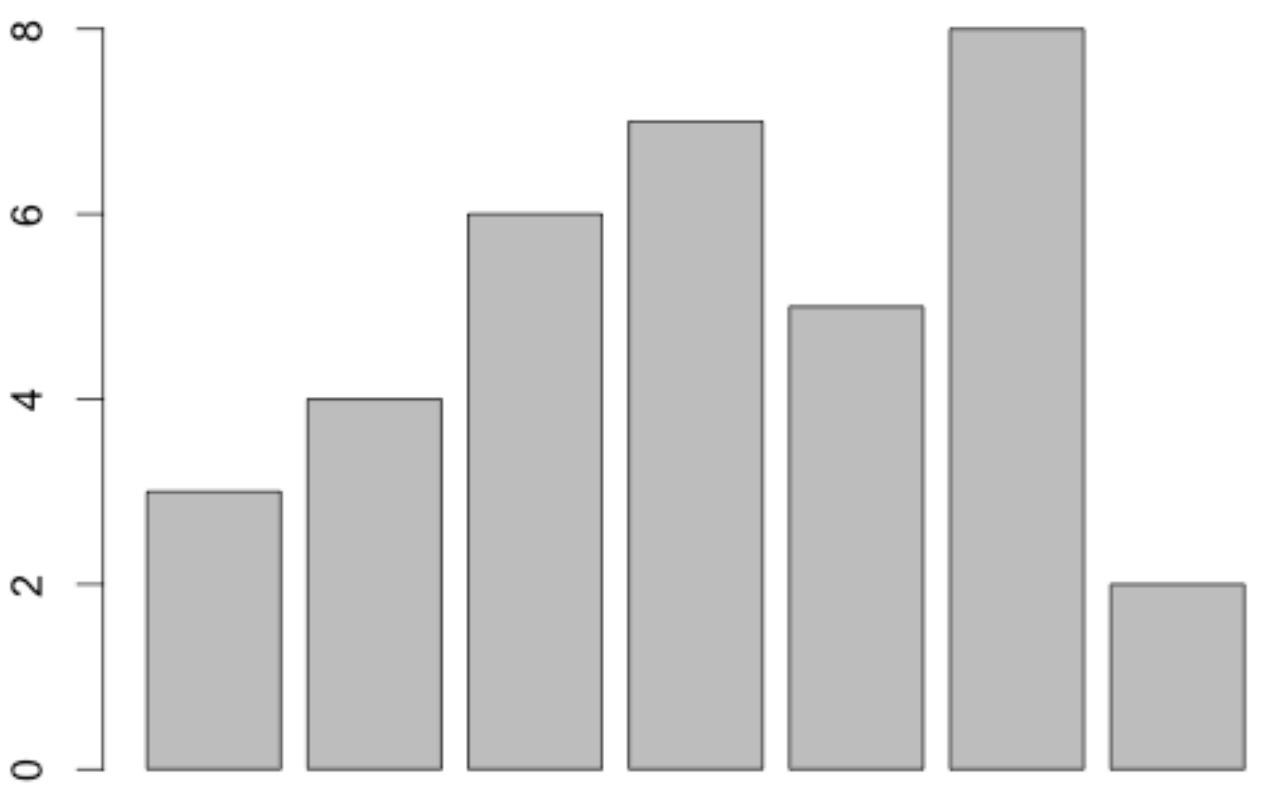


Bar Plots

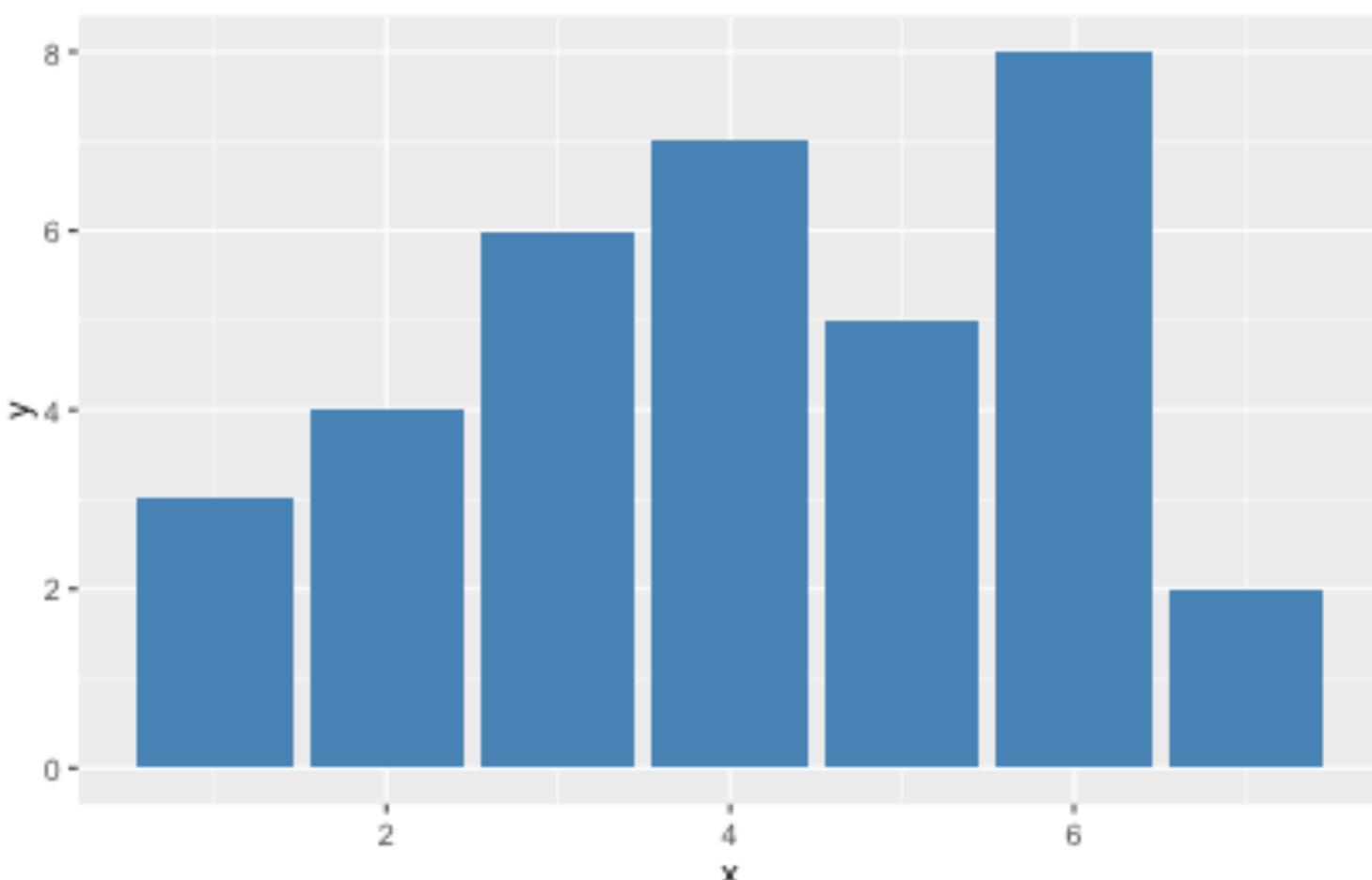
```
barchart = data.plot( kind="bar" )
```



```
barplot(data)
```

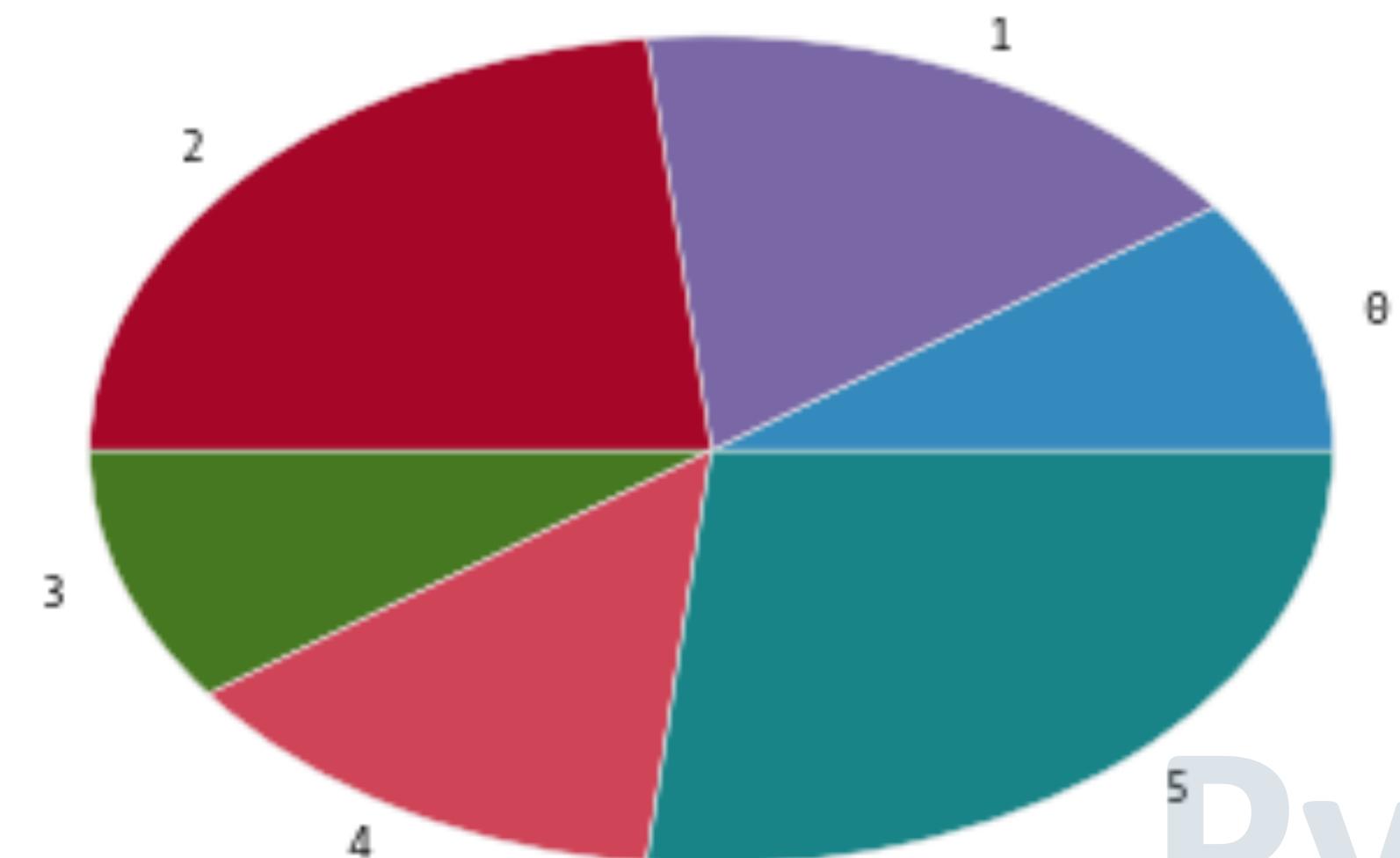


```
ggplot(dataf, aes(x, y)) +  
geom_bar(stat="identity", fill="steelblue")
```



Pie Charts

```
piechart = data.plot( kind="pie" )
```



Python

• • •

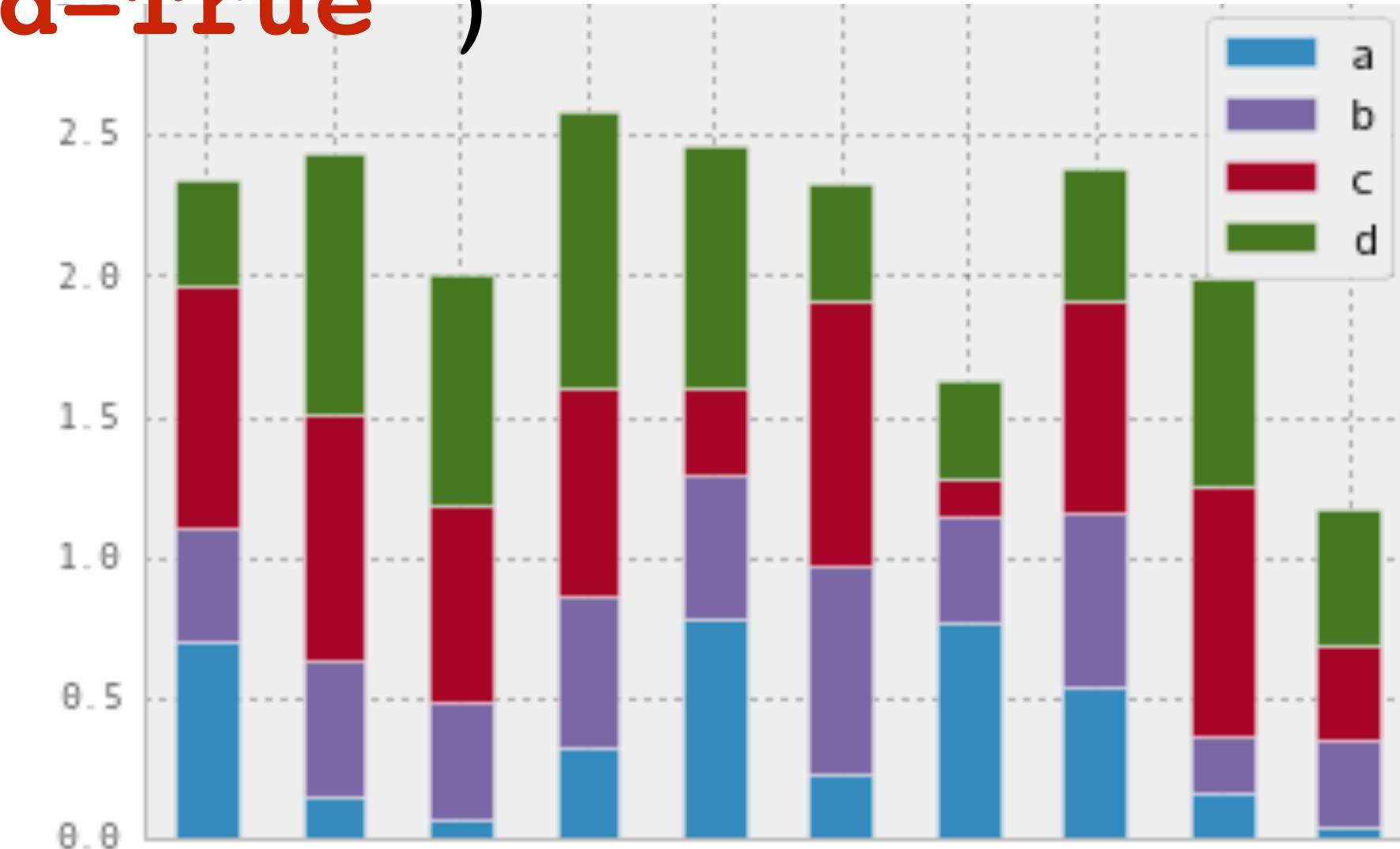
HERE'S A NICKEL,
KID. GET YOUR-
SELF A BETTER
COMPUTER.



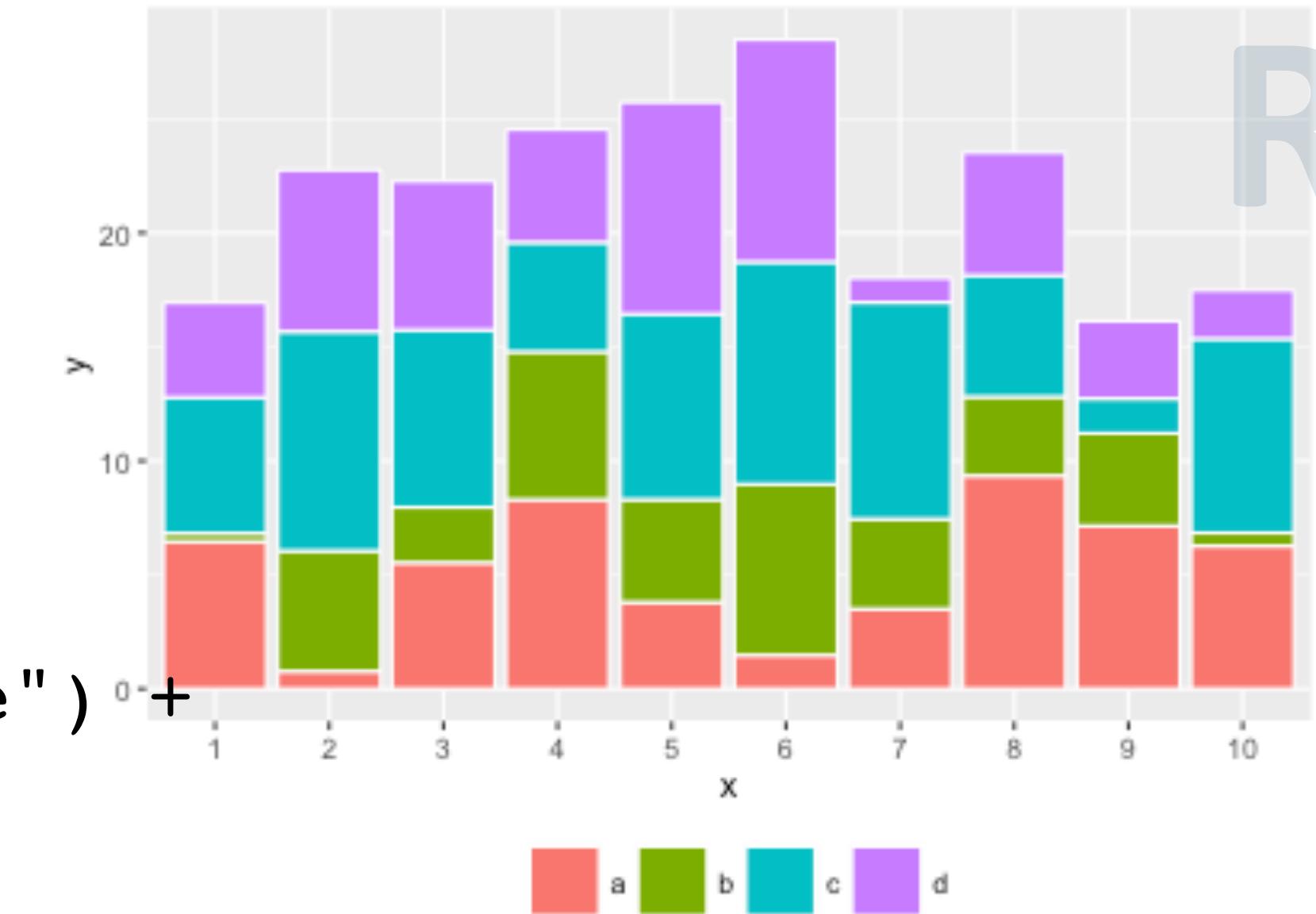
R

Stacked Bar Charts

```
df2.plot( kind='bar' , stacked=True )
```



Python

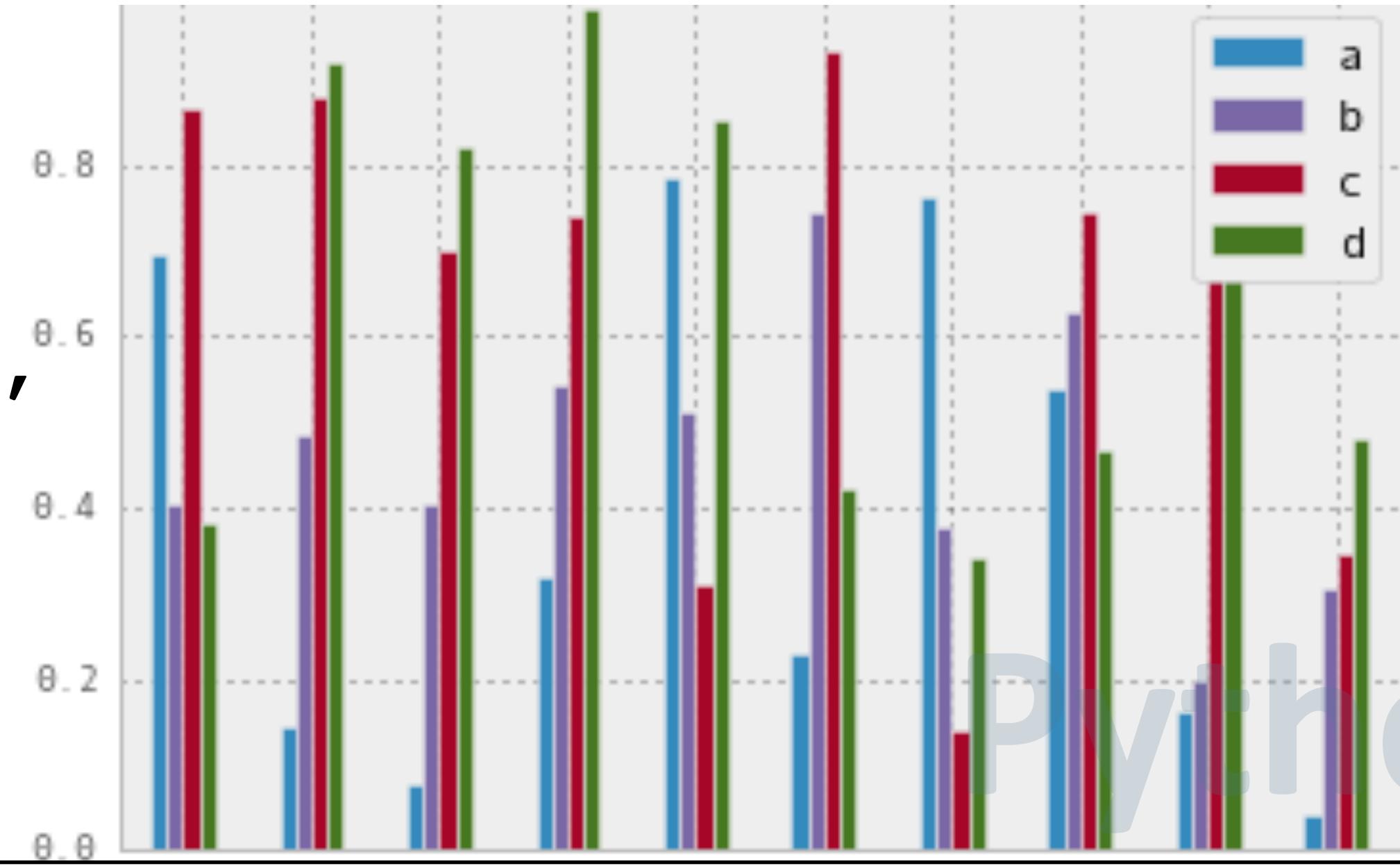


R

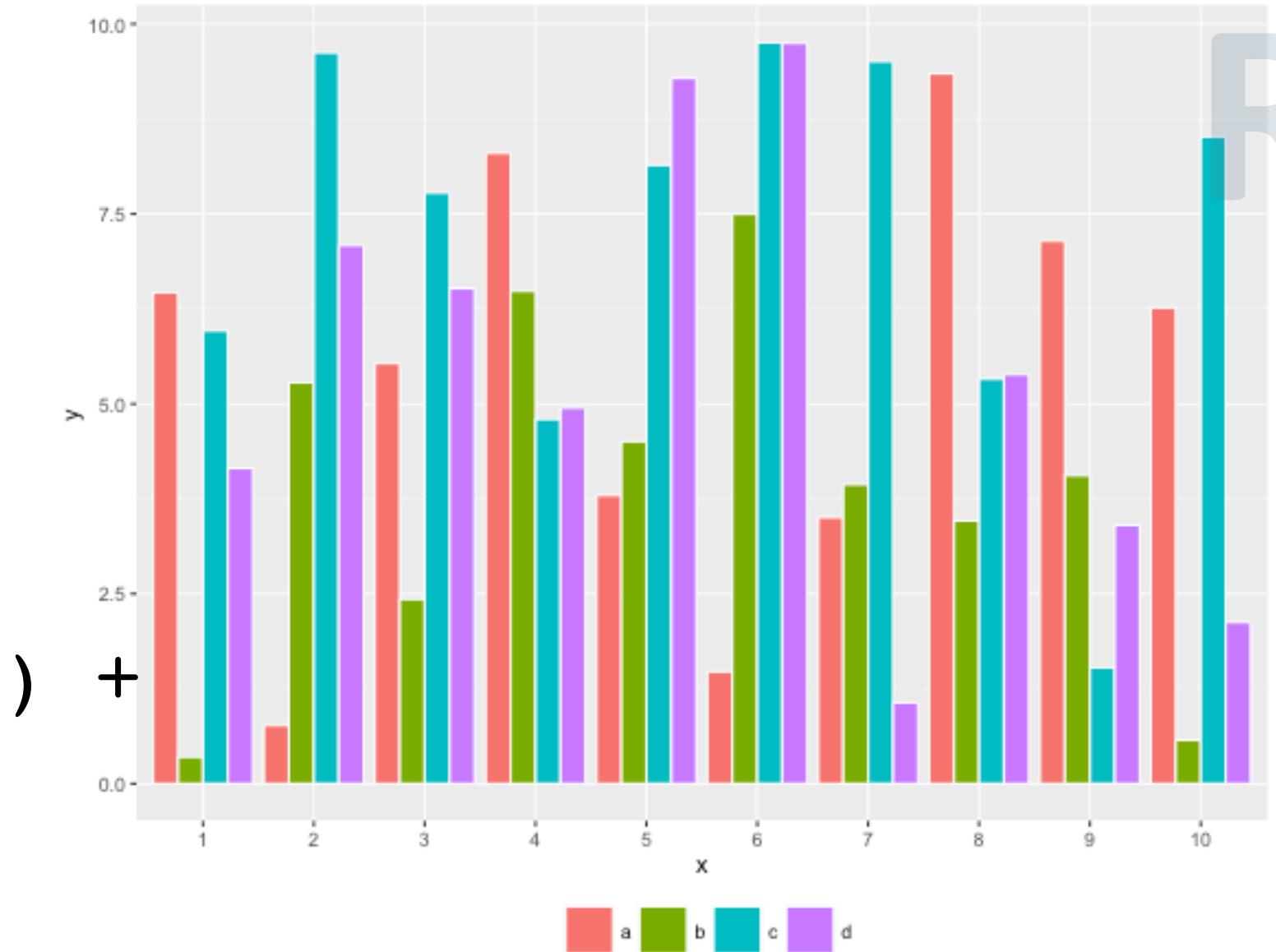
```
ggplot(df2, aes(x,y, fill=cat)) +  
  geom_bar(stat="identity", position="stack", color="white") +  
  theme(legend.position="bottom",  
        legend.title = element_blank())
```

Grouped Bar Charts

```
df2 = pd.DataFrame(np.random.rand(10, 4),  
columns=[ 'a' , 'b' , 'c' , 'd' ] )  
df2.plot( kind='bar' )
```

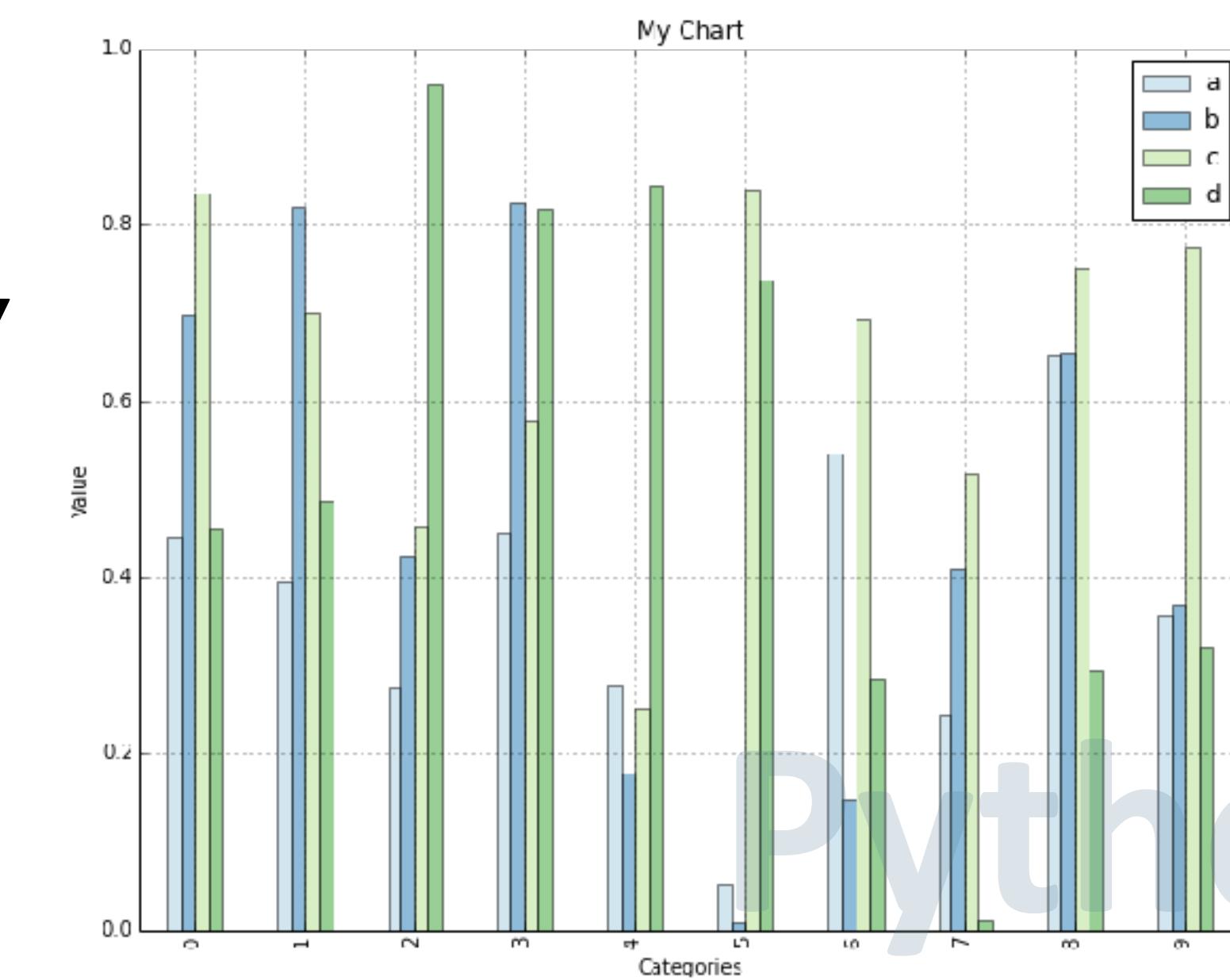


```
df2 <- tibble(x=factor(rep(seq(10), each=4)),  
               y=runif(40,0,10),  
               cat=rep(letters[1:4], 10))  
  
ggplot(df2, aes(x,y, fill=cat)) +  
  geom_bar(stat="identity", position="dodge", color="white") +  
  theme(legend.position="bottom",  
        legend.title = element_blank())
```

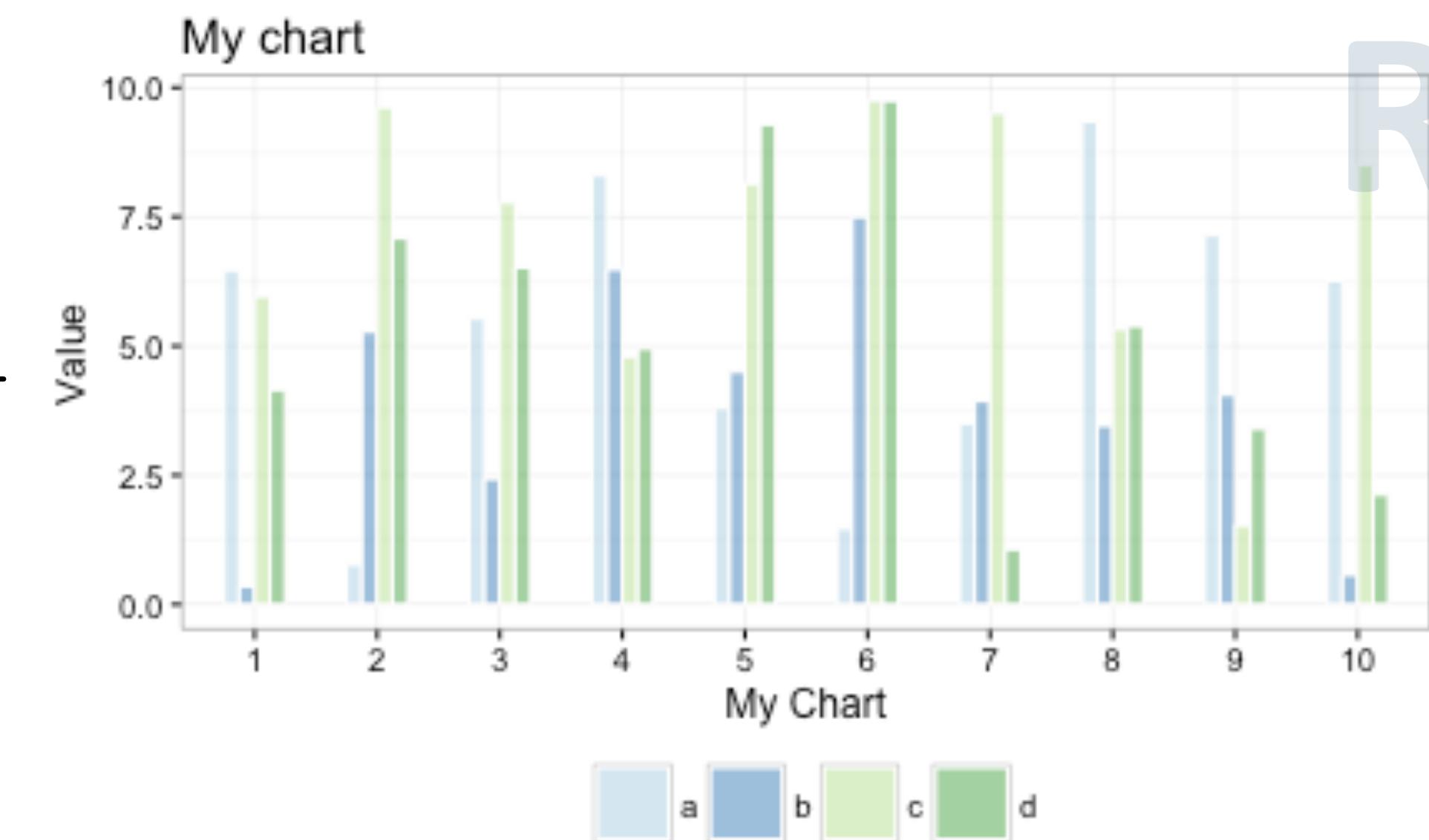


Grouped Bar Chart (custom fill)

```
df2.plot( kind='bar',
          color=('#a6cee3', '#1f78b4', '#b2df8a', '#33a02c' ),
          alpha=0.5,
          width=0.5,
          figsize=(10, 8))
plt.title( "My Chart" )
plt.xlabel( "Categories" )
plt.ylabel( "Value" )
```

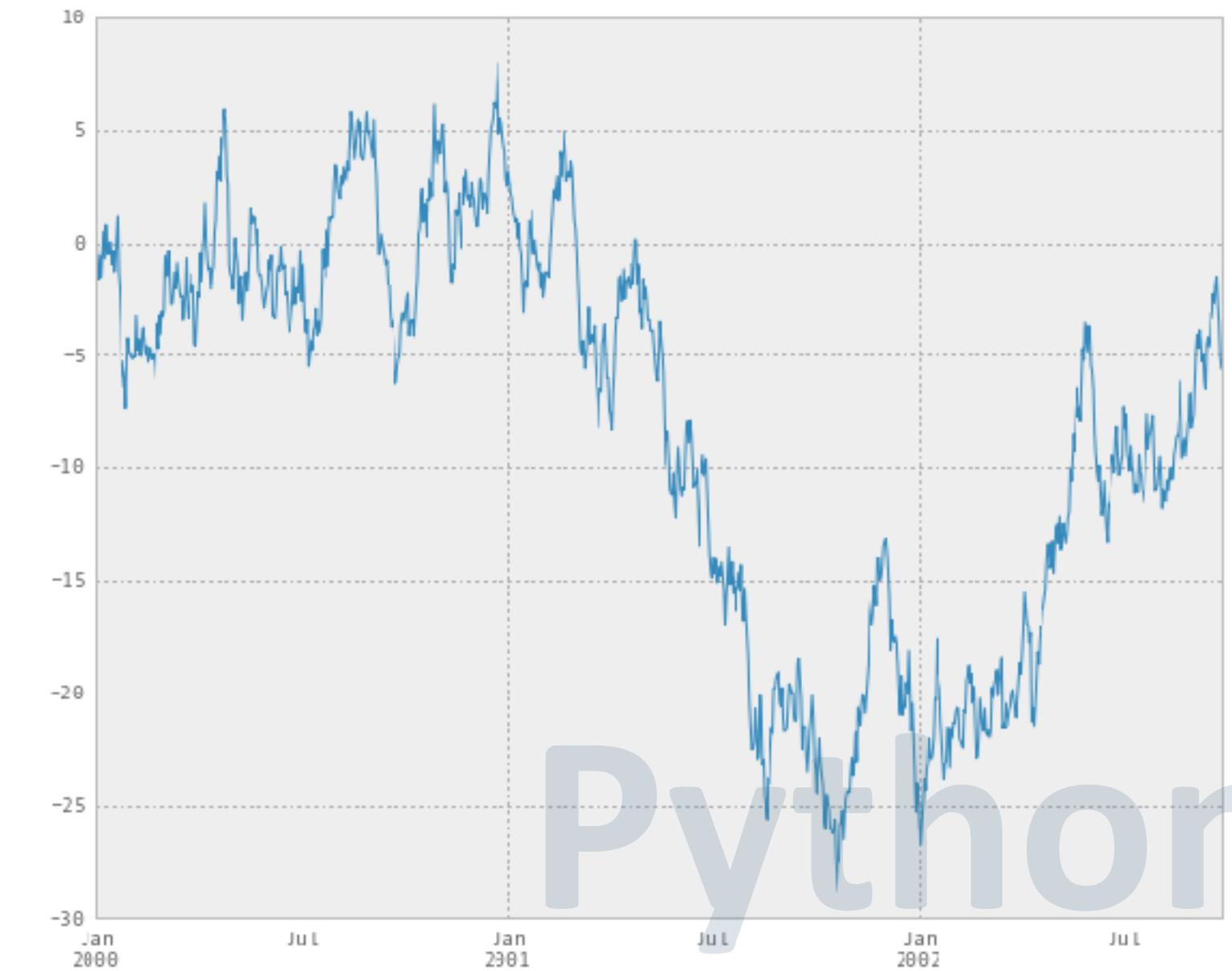


```
ggplot(df2, aes(x, y, fill=cat)) +
  geom_bar(stat="identity", position="dodge",
           color="white", alpha=1/2, width=0.5) +
  scale_color_manual(values=c('#a6cee3', '#1f78b4',
                             '#b2df8a', '#33a02c')) +
  ggtitle("My chart") +
  labs(x="My Chart", y="Value") +
  theme_bw() +
  theme(legend.position="bottom",
        legend.title = element_blank())
```

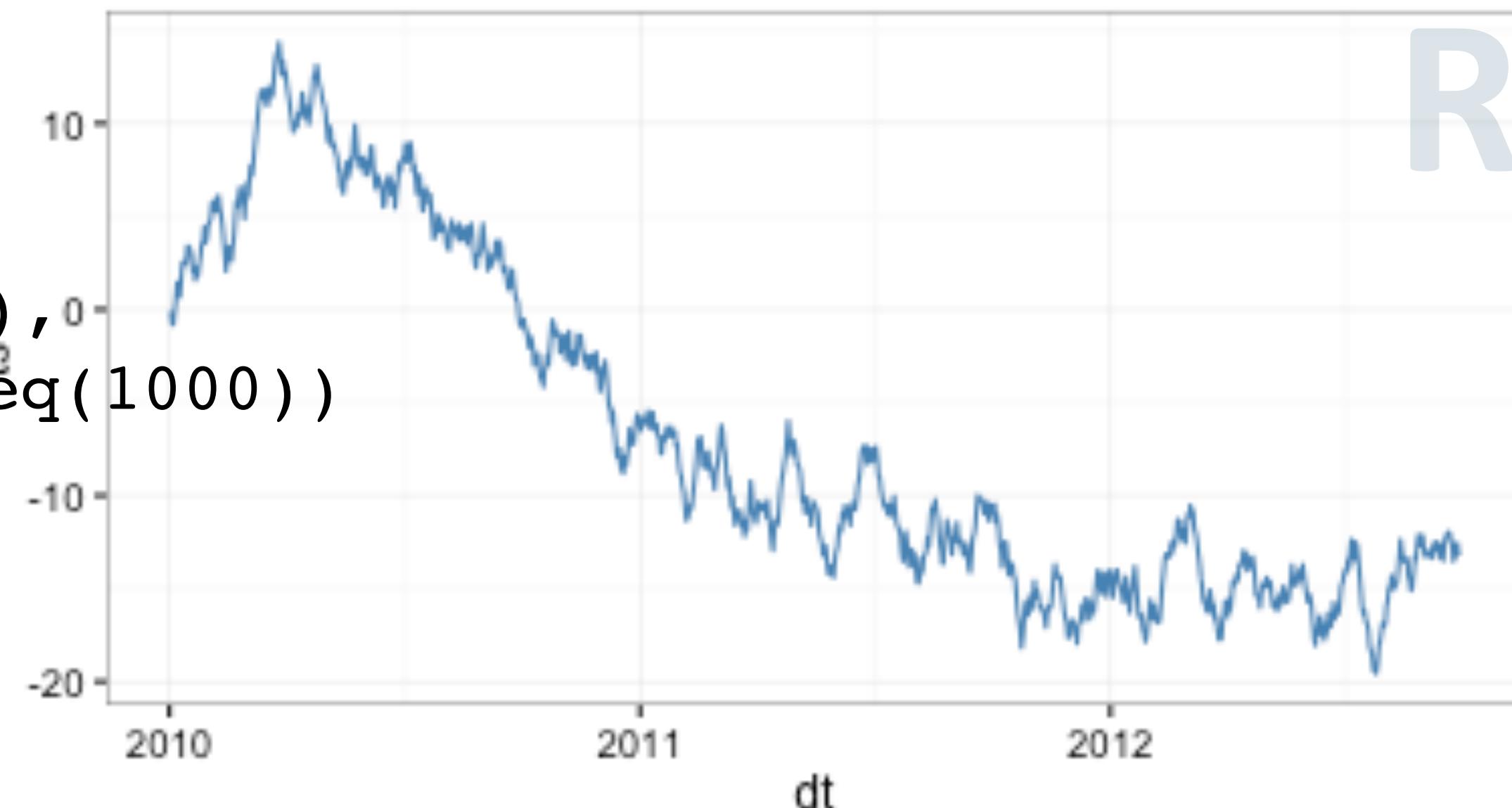


Time series (line plot)

```
ts = pd.Series(np.random.randn( 1000 ),  
    index=pd.date_range('1/1/2000',  
    periods=1000))  
ts = ts.cumsum()  
timeseriesChart = ts.plot( figsize=(10, 8) )
```

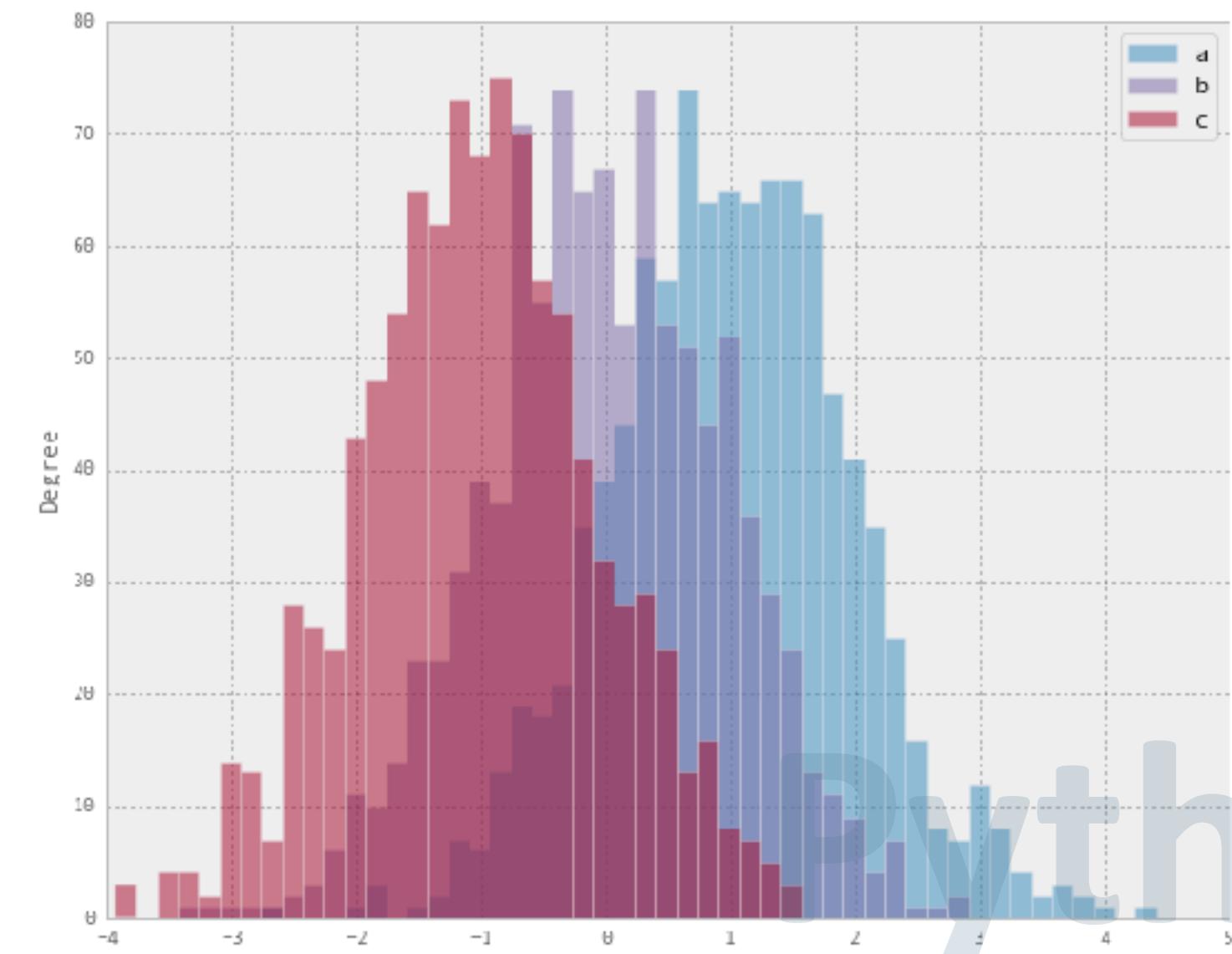


```
toplot <- data.frame(ts=cumsum(runif(1000, -1, 1)),  
    dt=as.Date("2010-01-01") + seq(1000))  
ggplot(toplot, aes(x=dt, y=ts)) +  
    geom_line(color="steelblue") +  
    theme_bw()
```

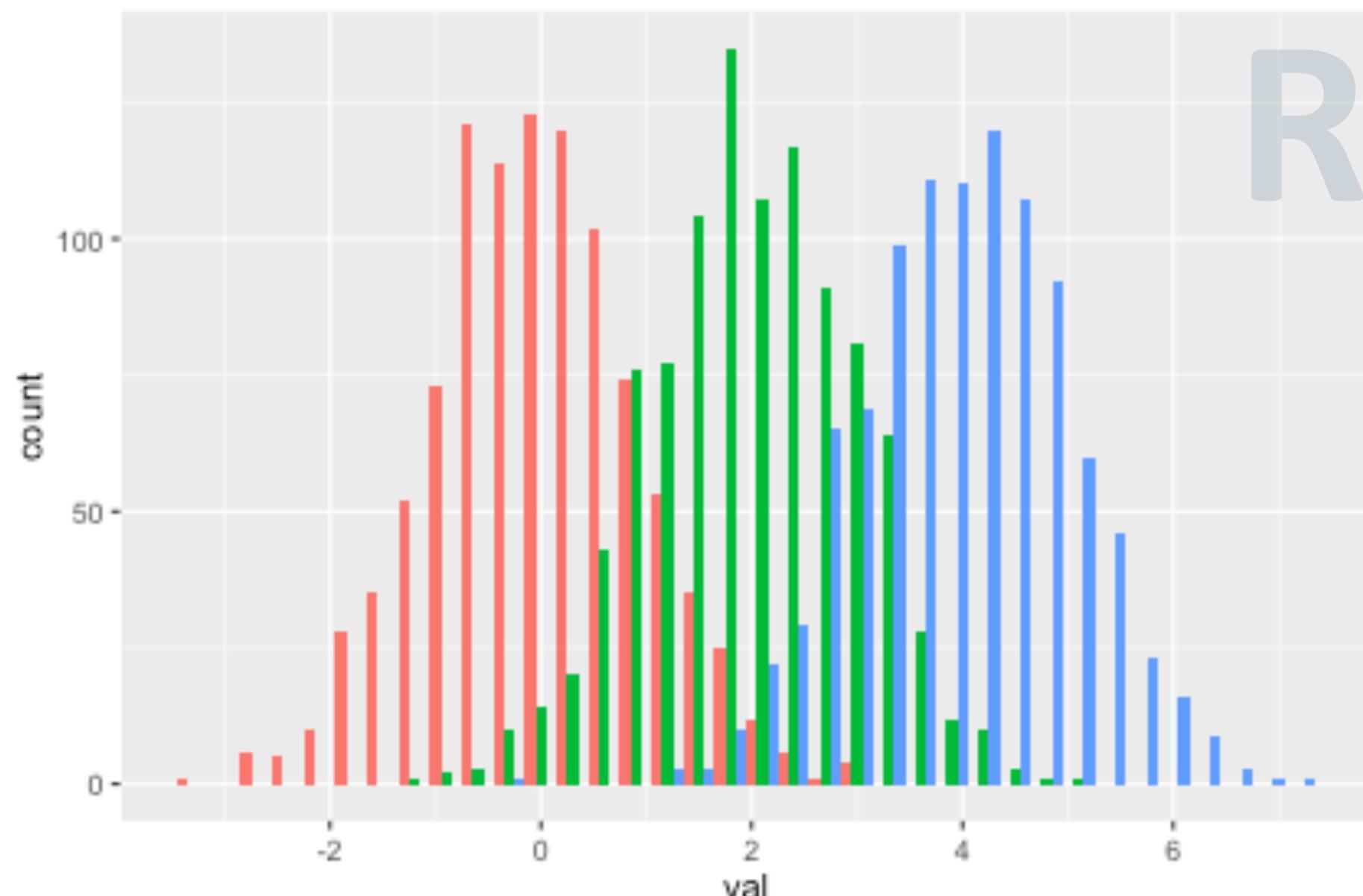


Histograms

```
df4.plot(kind='hist',  
         alpha=0.5,  
         bins=50 )
```

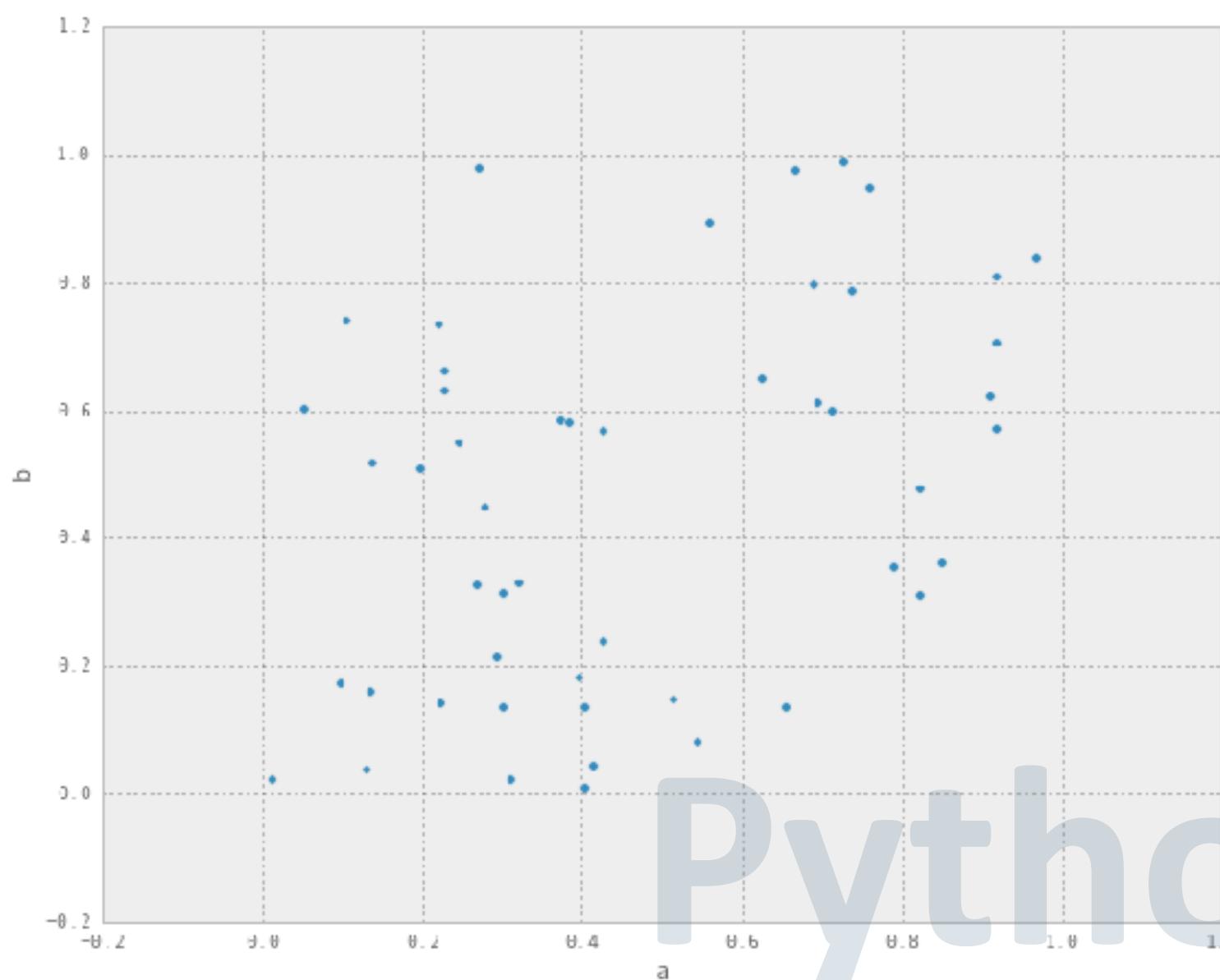


```
ggplot(df4, aes(val, fill=cat)) +  
  geom_histogram(position="dodge",  
                 binwidth = 0.3)
```



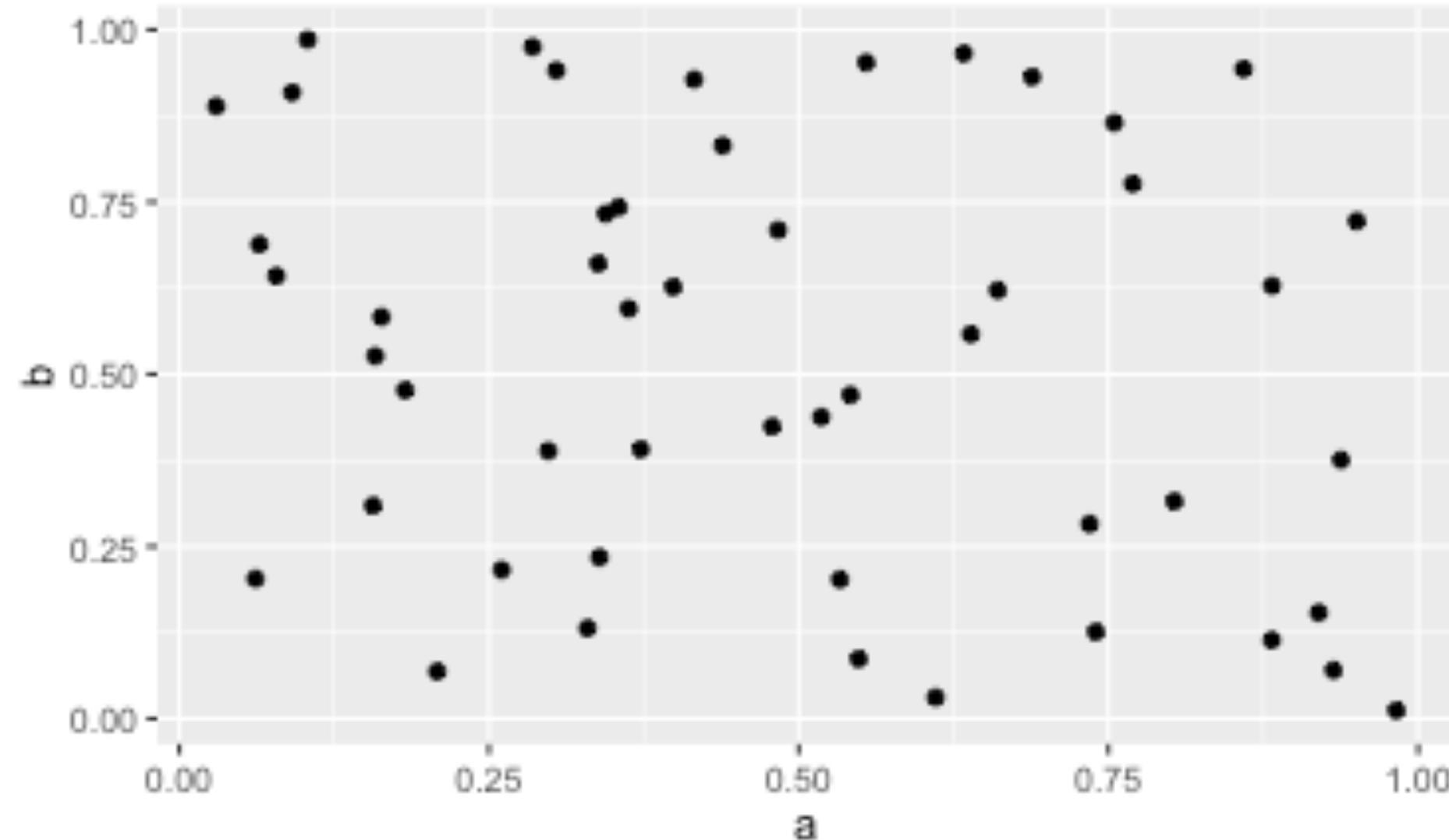
Scatter Plot

```
df5 = pd.DataFrame(np.random.rand(50, 4),  
columns= [ 'a' , 'b' , 'c' , 'd' ] )  
df5.plot(kind='scatter', x='a' , y='b' )
```



Python

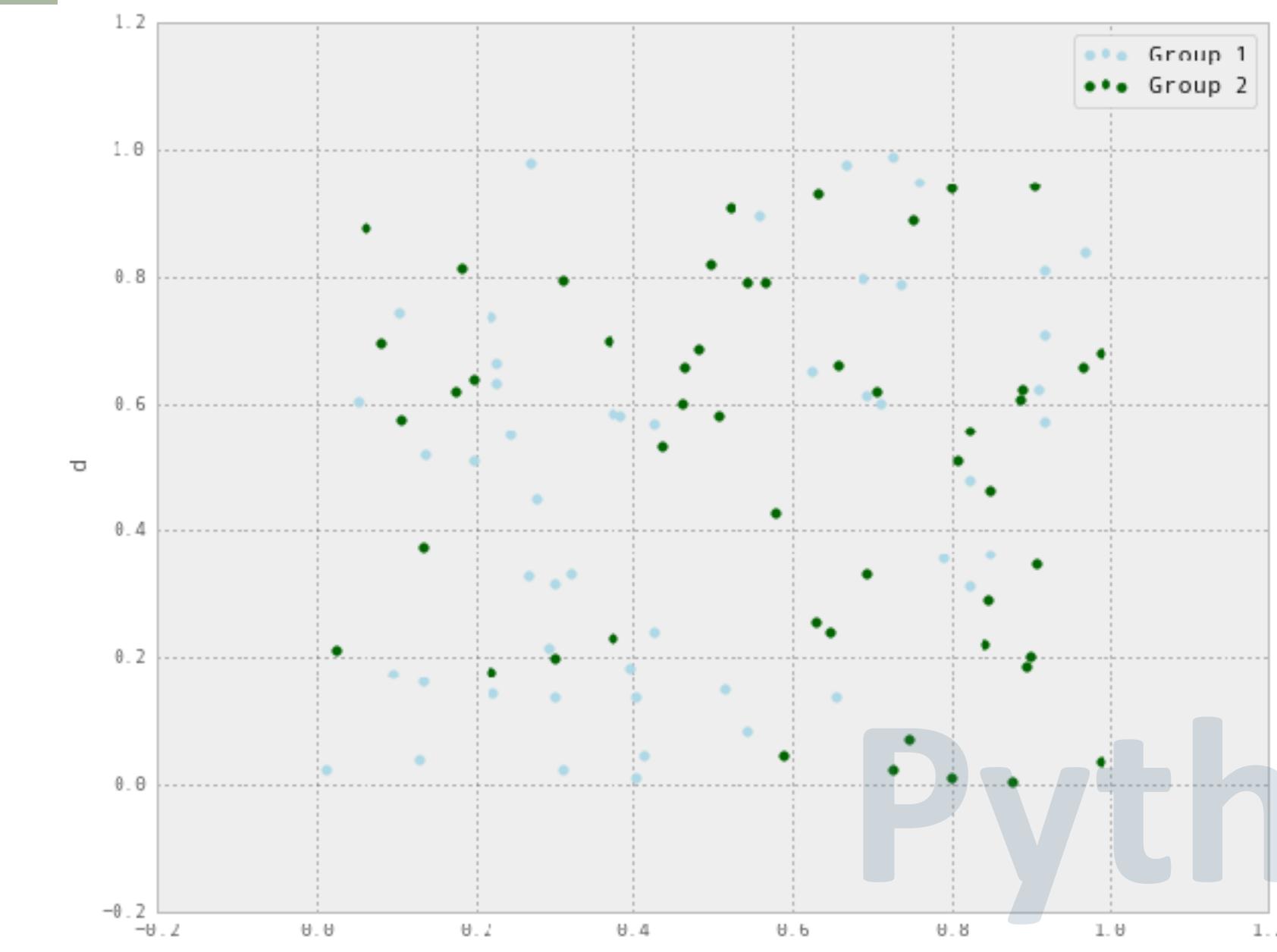
```
df5 <- data.frame(a=runif(50) , b=runif(50) ,  
c=runif(50) , d=runif(50))  
ggplot(df5, aes(x=a, y=b)) + geom_point()
```



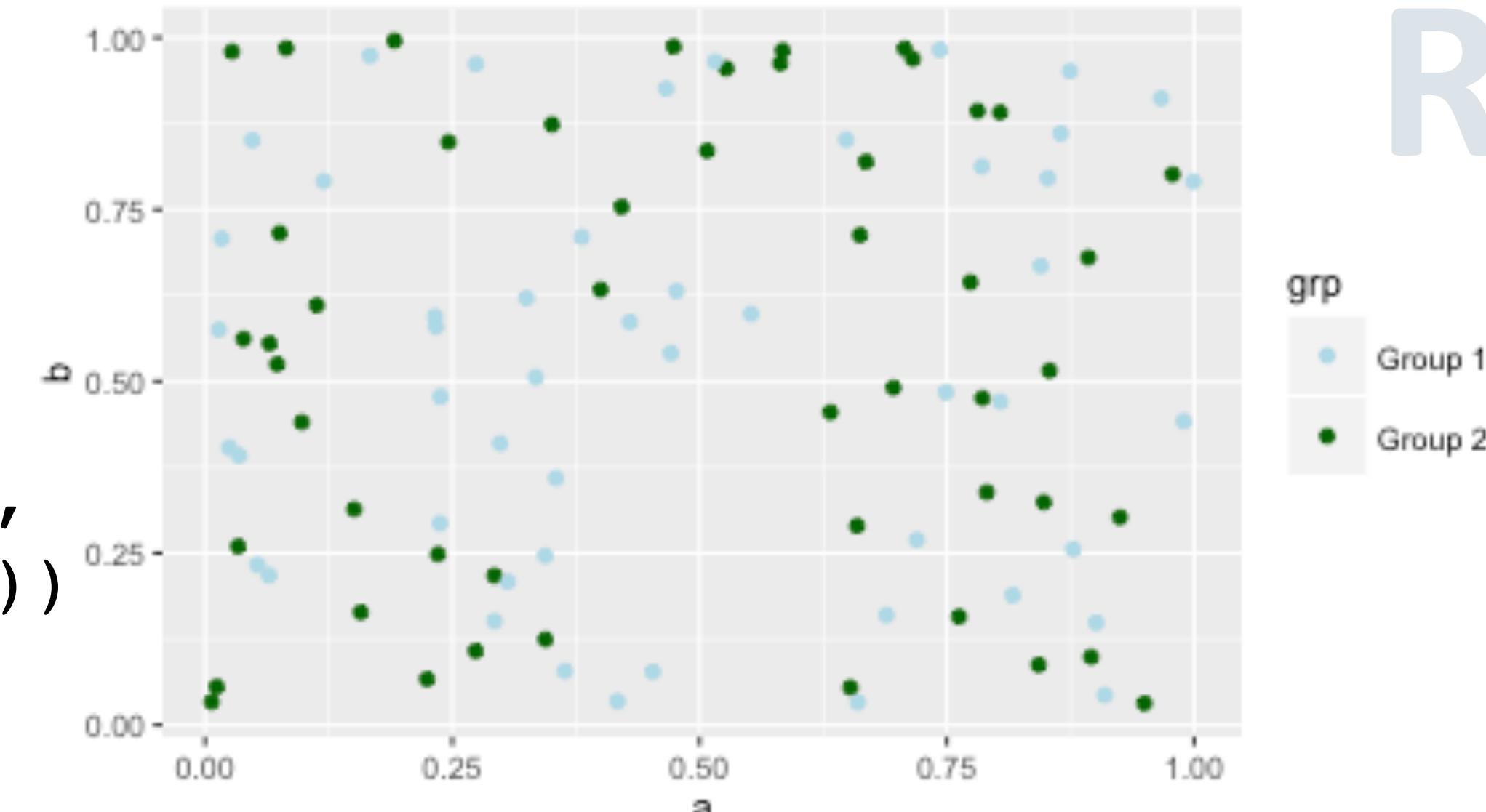
Scatter plot with groups

```
ax = df5.plot(kind='scatter', x='a', y='b',  
               color='LightBlue',  
               label='Group 1',  
               figsize=(10, 8))
```

```
df5.plot(kind='scatter', x='c', y='d',  
          color='DarkGreen',  
          label='Group 2',  
          ax=ax)
```

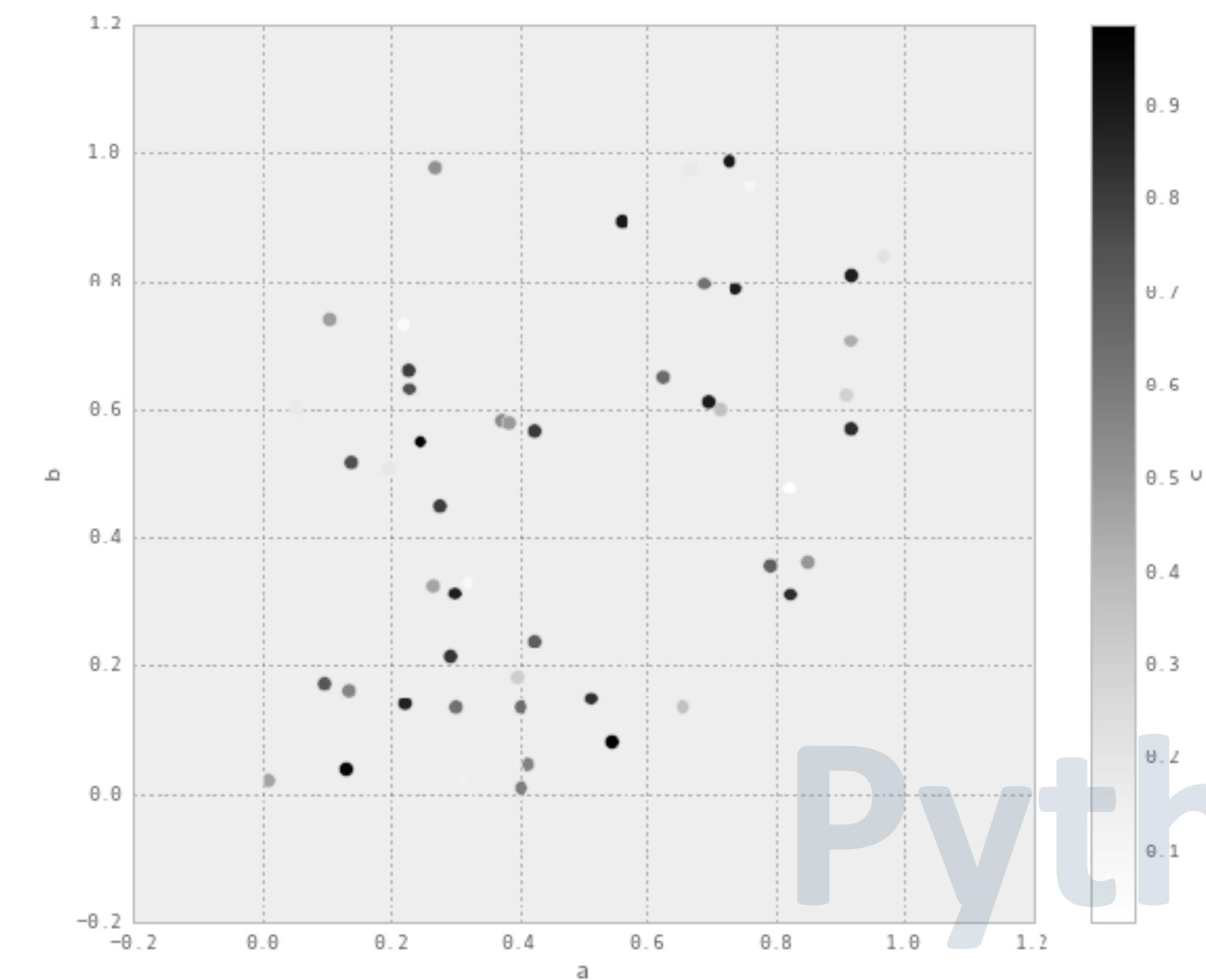


```
df5 <- data.frame(grp=paste("Group", c(1,2)),  
                   a=runif(100), b=runif(100))  
ggplot(data=df5, aes(x=a, y=b, color=grp)) +  
  geom_point() +  
  scale_color_manual(values=c("Group 1"="lightblue",  
                             "Group 2"="darkgreen"))
```

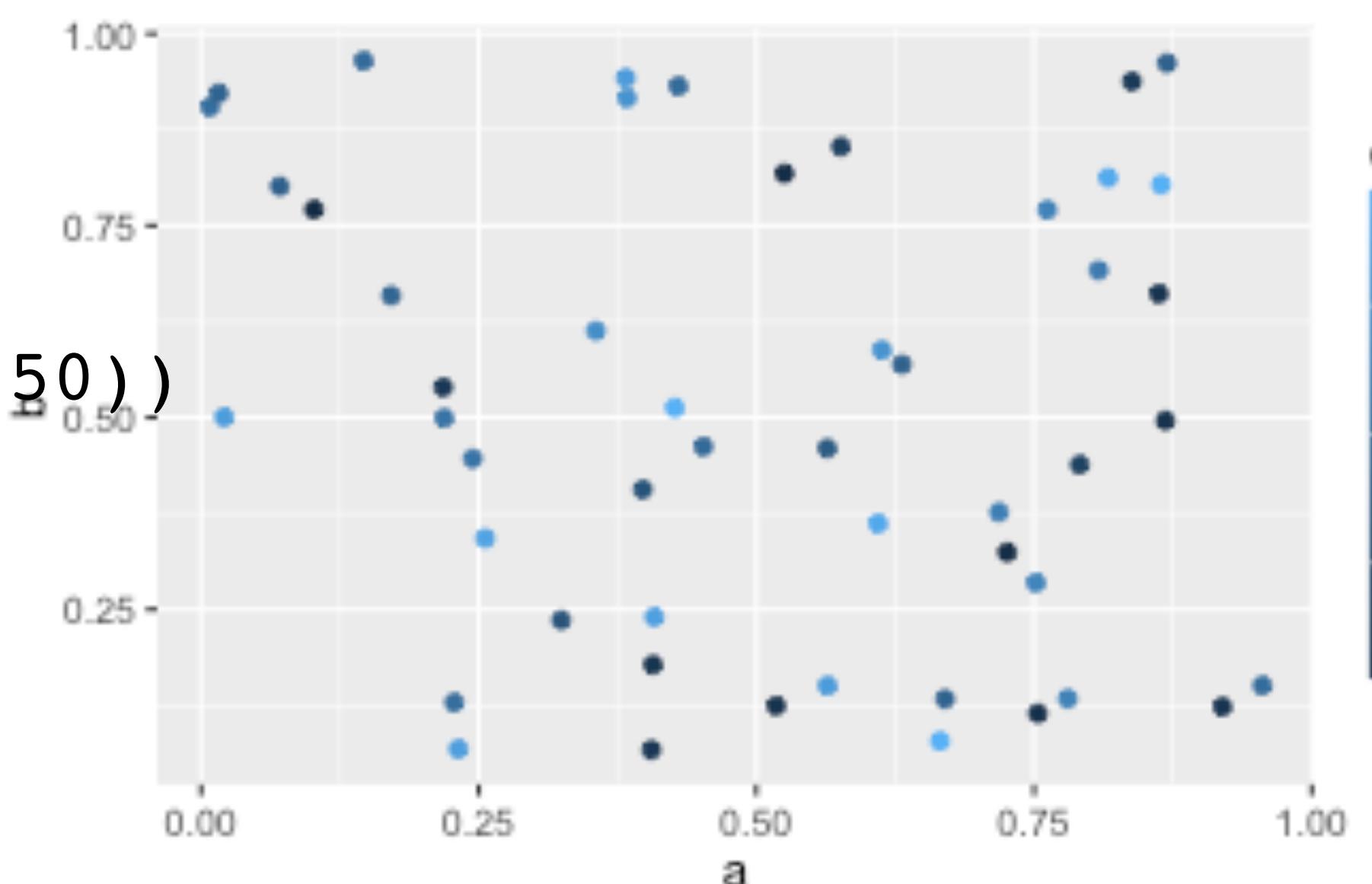


Scatter plots with color density representing data

```
df5.plot(kind='scatter', x='a', y='b',  
c='c', s=50 )
```

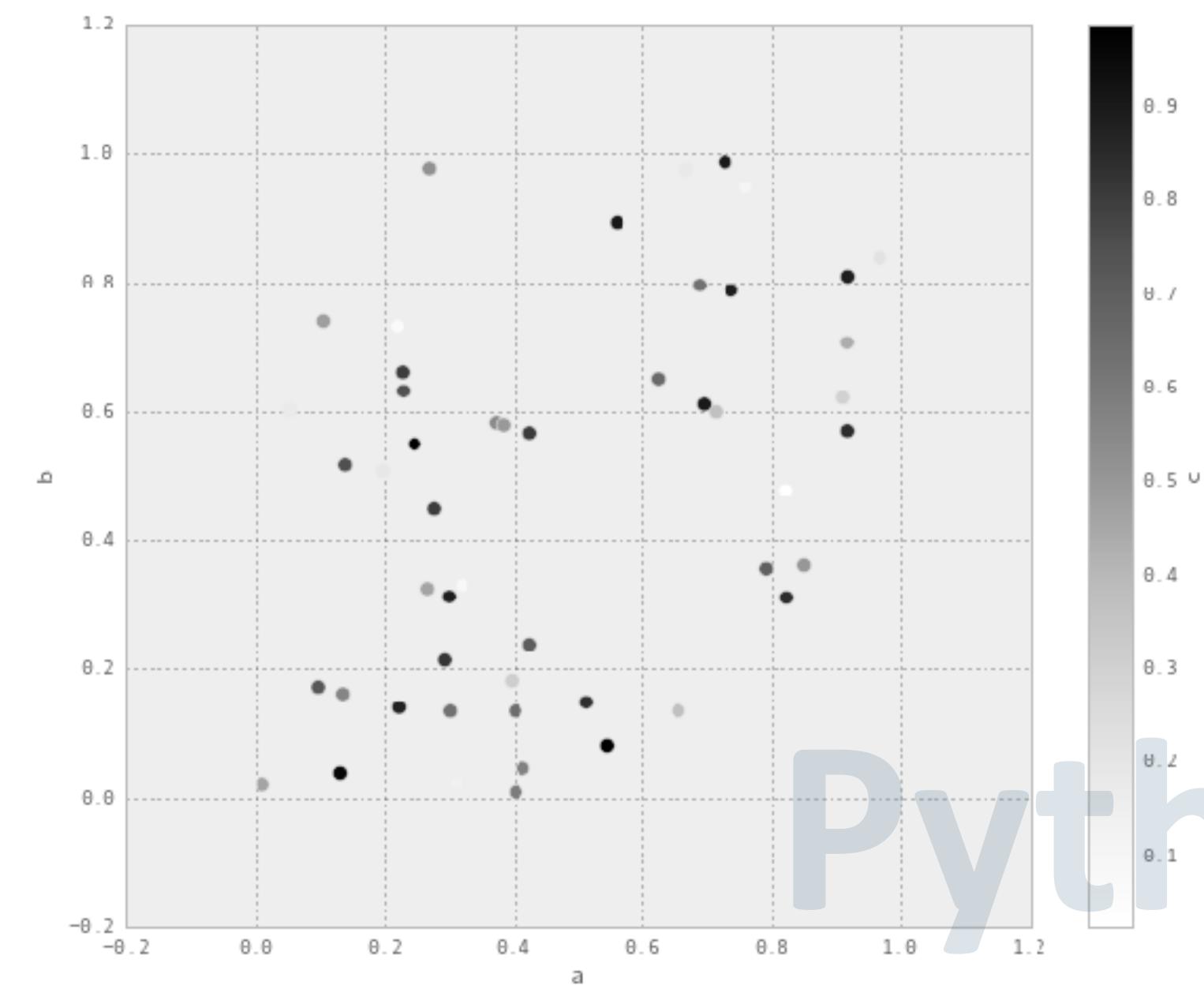


```
df5 <- data.frame(a=runif(50), b=runif(50), c=runif(50))  
ggplot(df5, aes(a,b,color=c)) + geom_point()
```



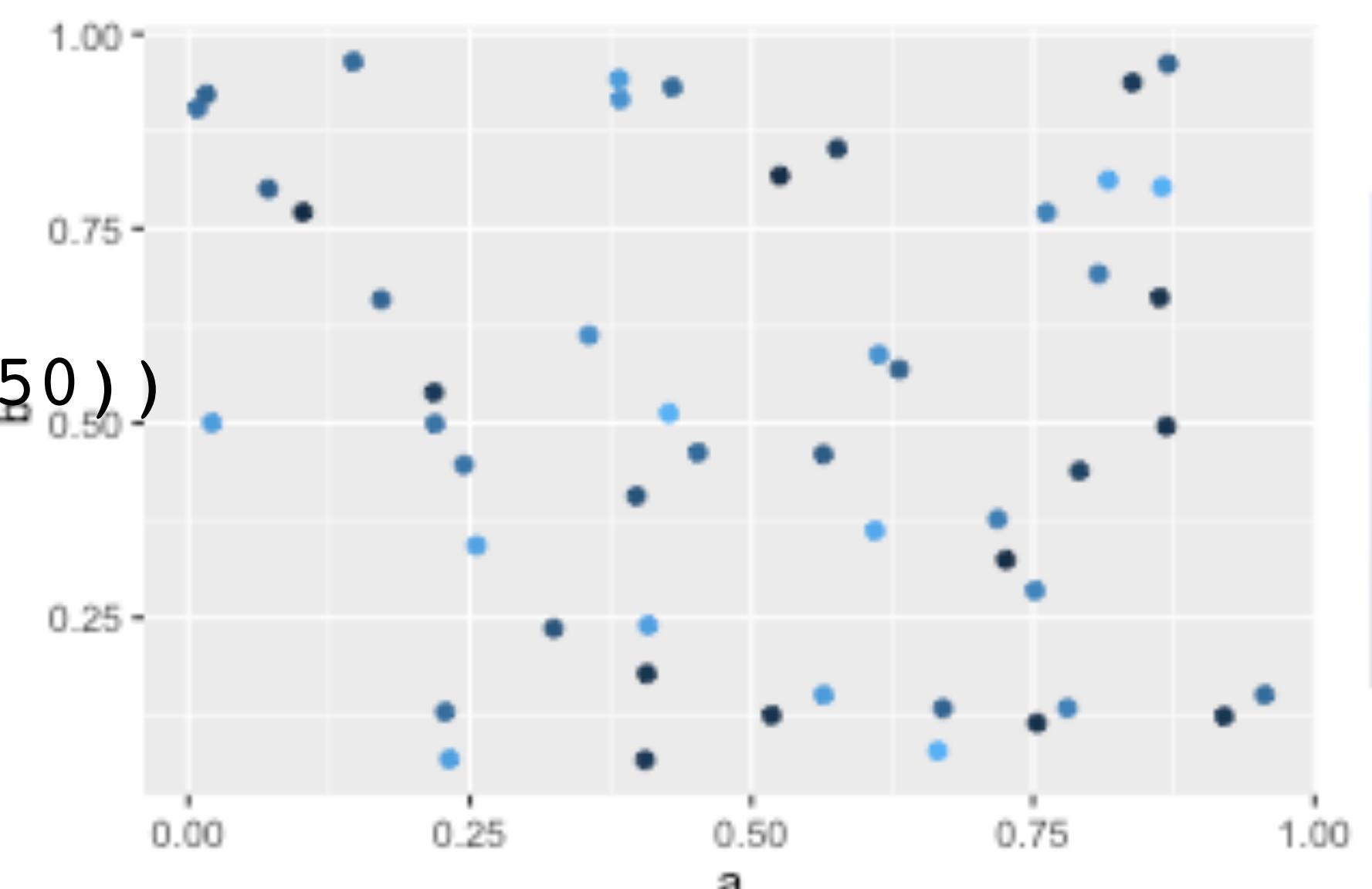
Setting up for visualization

```
df5.plot(kind='scatter', x='a', y='b',  
c='c', s=50 )
```



Python

```
df5 <- data.frame(a=runif(50), b=runif(50), c=runif(50))  
ggplot(df5, aes(a,b,color=c)) + geom_point()
```



R

O'REILLY®

Security

BUILD BETTER DEFENSES

Interactive Visualizations

oreillysecuritycon.com

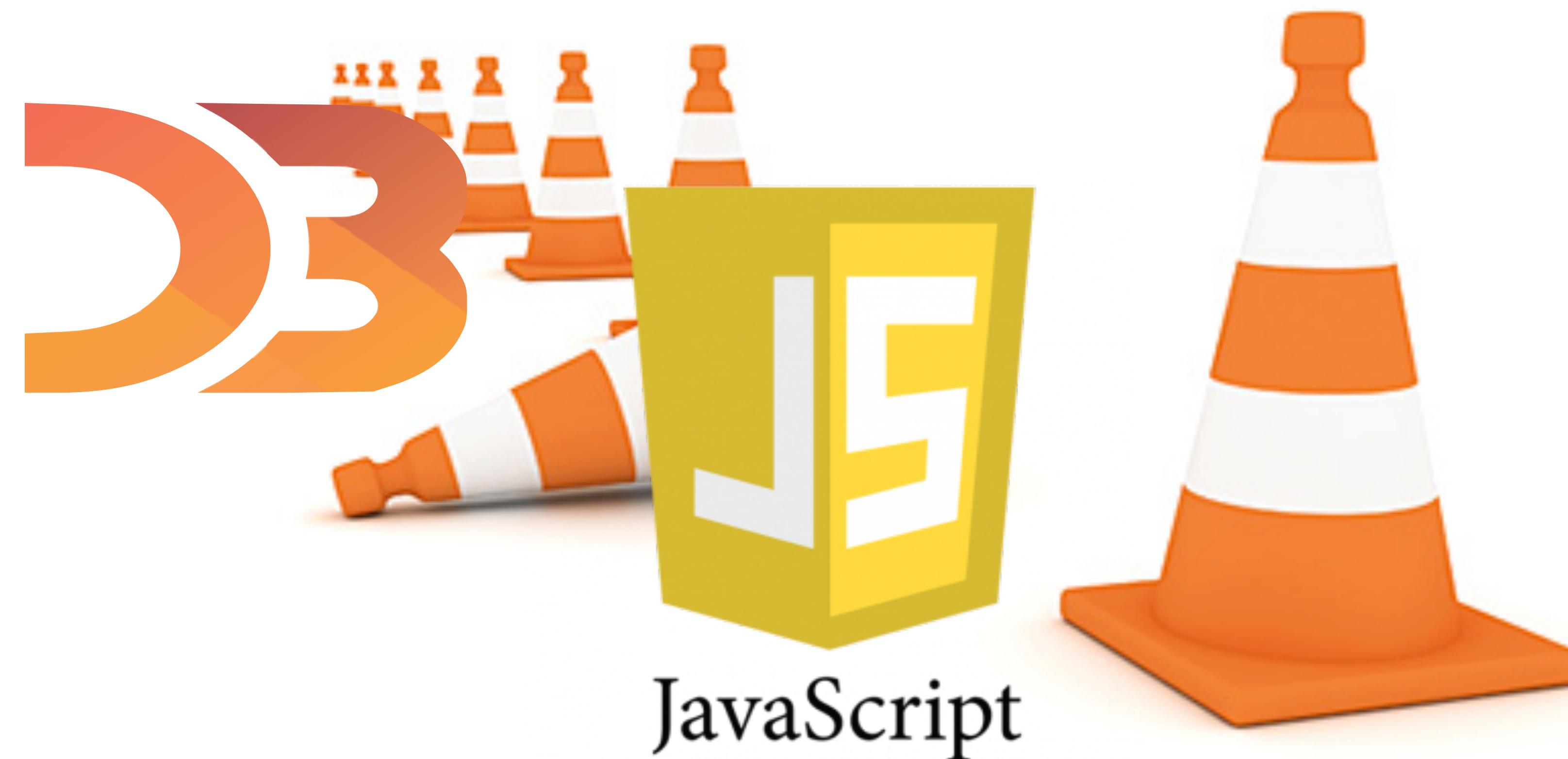
#oreillysecurity



QlikView



Interactive Visualizations





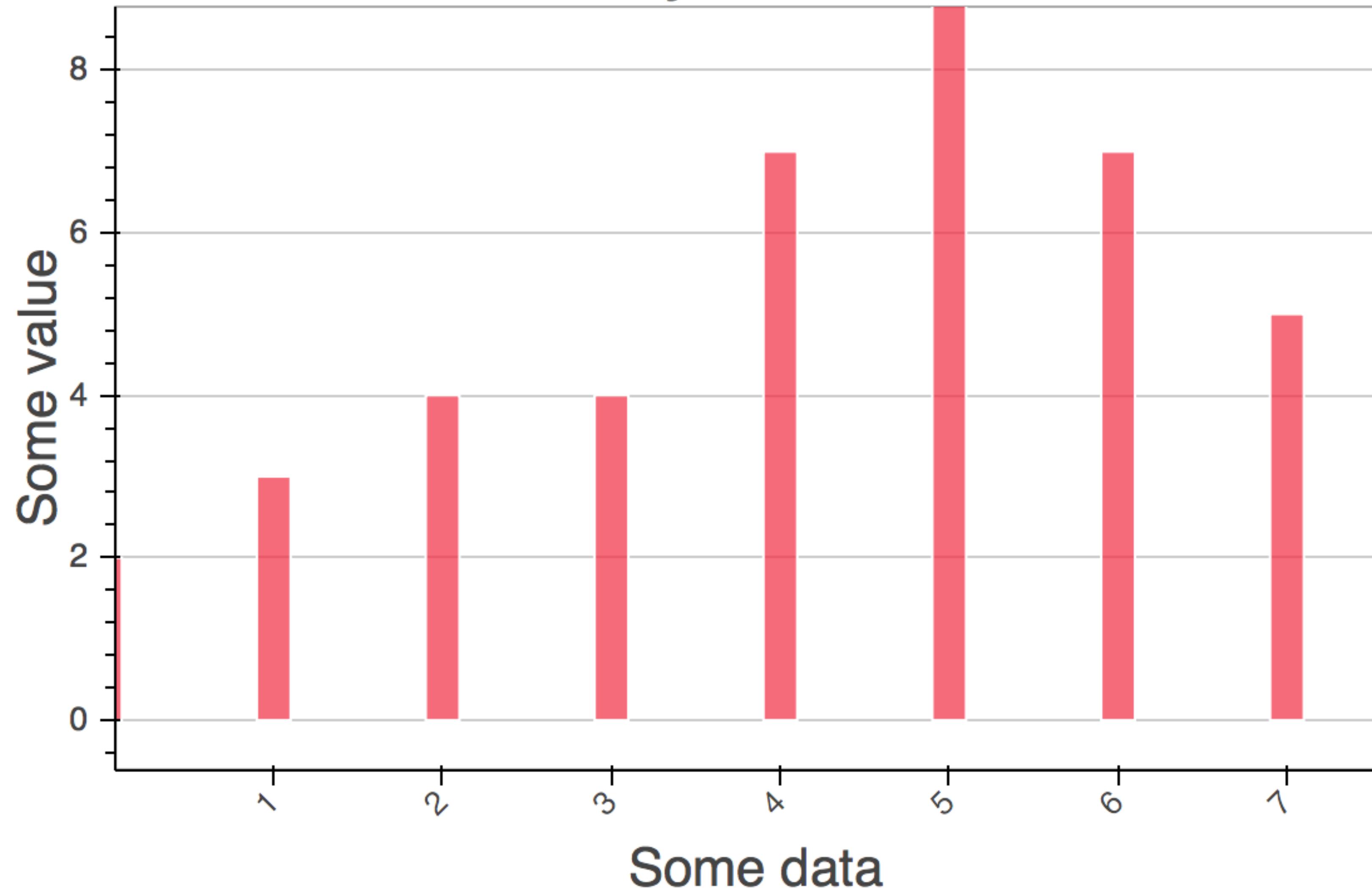
Requires Shiny Server (open source)
<http://shiny.rstudio.com/gallery/>



Introducing Bokeh



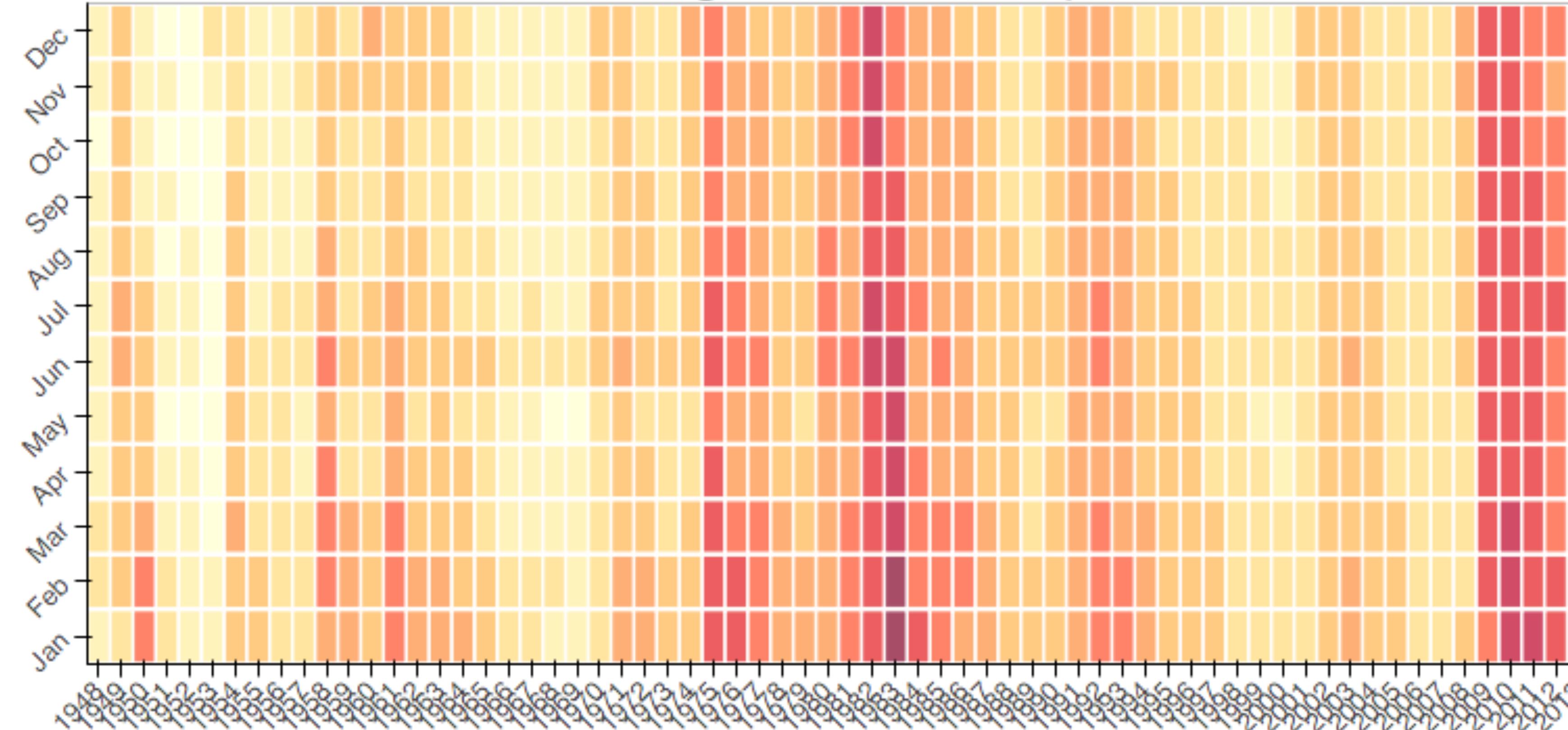
My Chart





⊕ 🔎 ⌂ ⏷ ⏵ ⏶ ⏴ ?

categorical heatmap



```
from bokeh.charts import HeatMap, output_file, show
from bokeh.palettes import YlOrRd9 as palette
from bokeh.sampledata.unemployment1948 import data

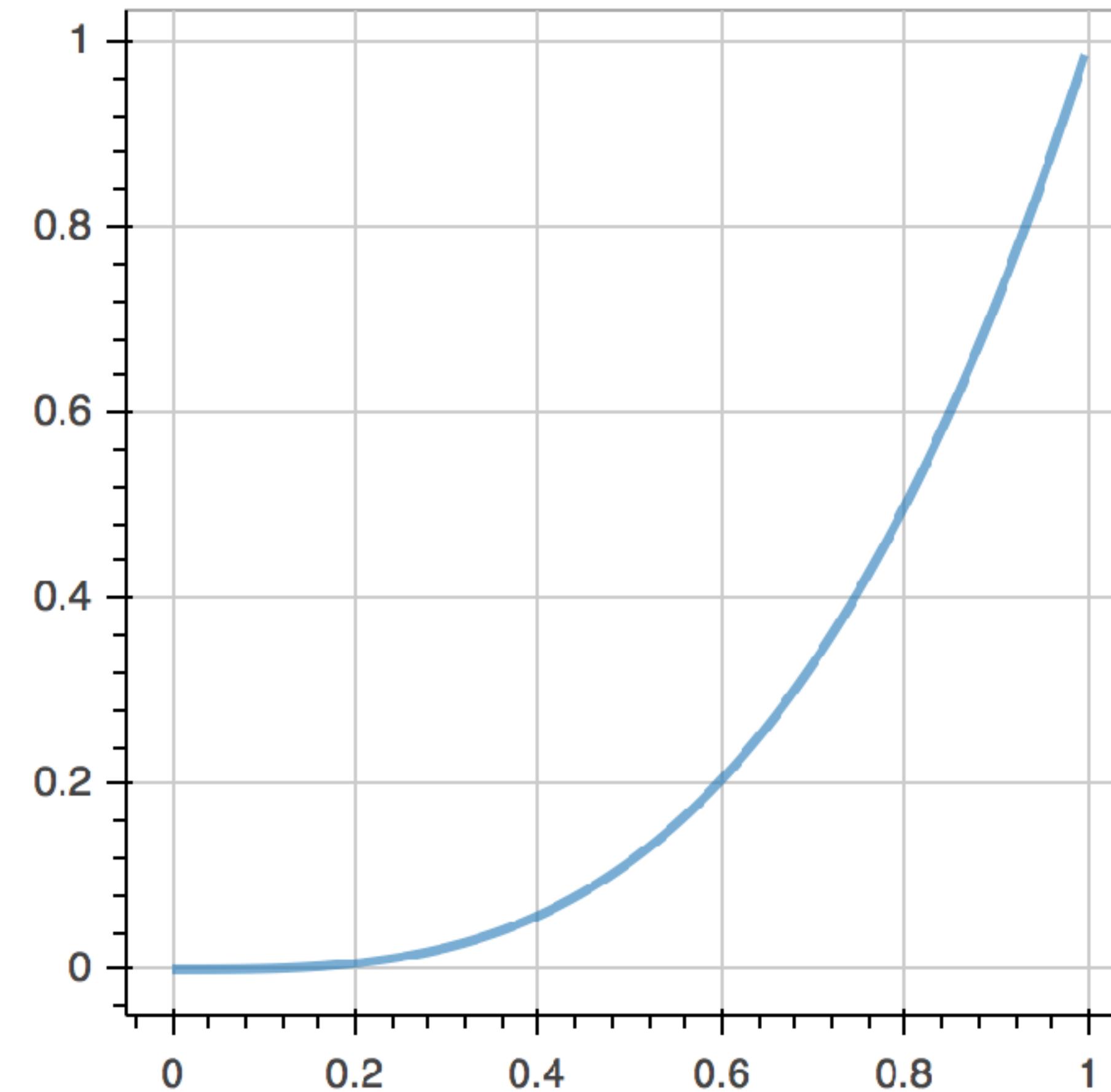
# pandas magic
df = data[data.columns[:-1]]
df2 = df.set_index(df[df.columns[0]].astype(str))
df2.drop(df.columns[0], axis=1, inplace=True)
df3 = df2.transpose()

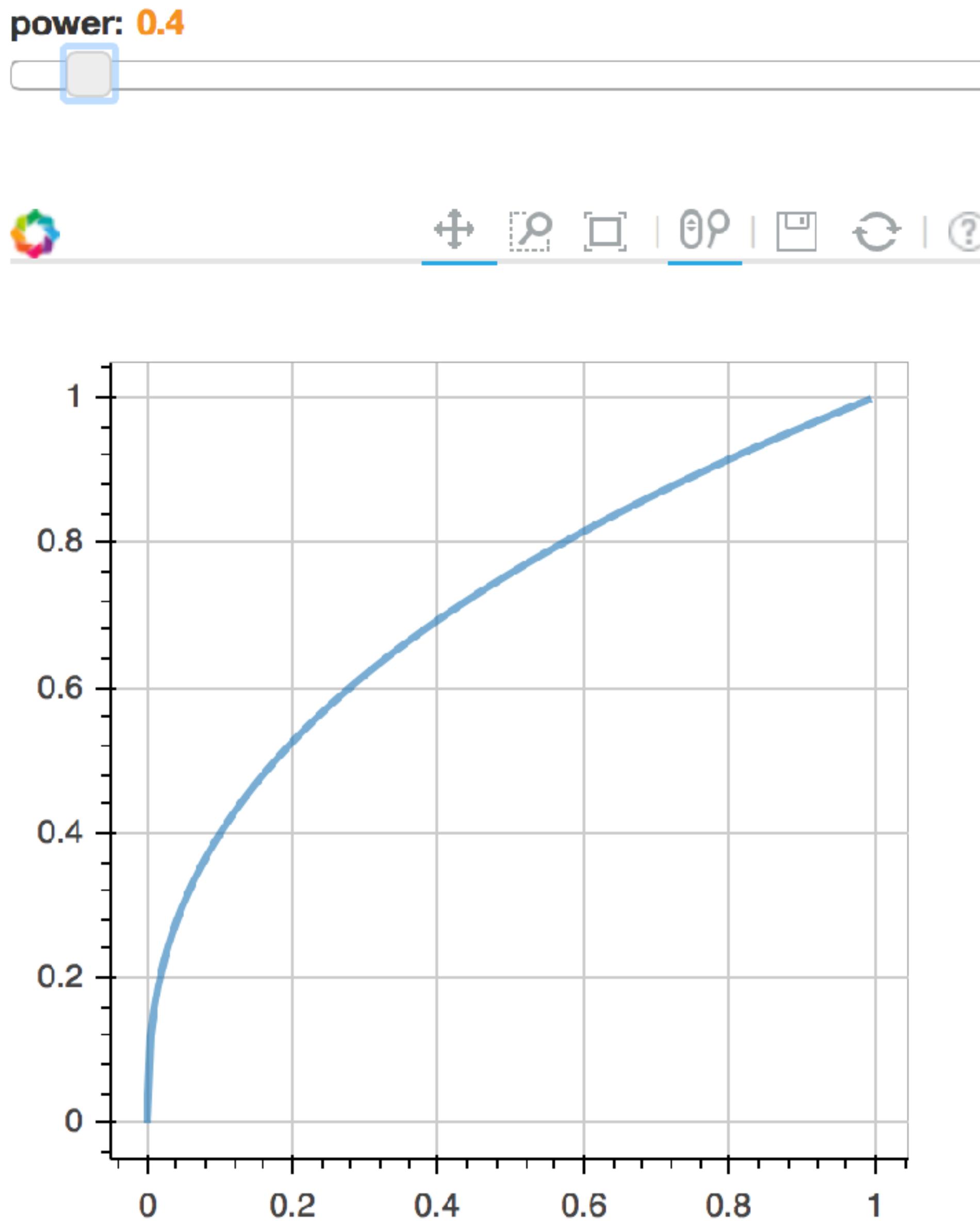
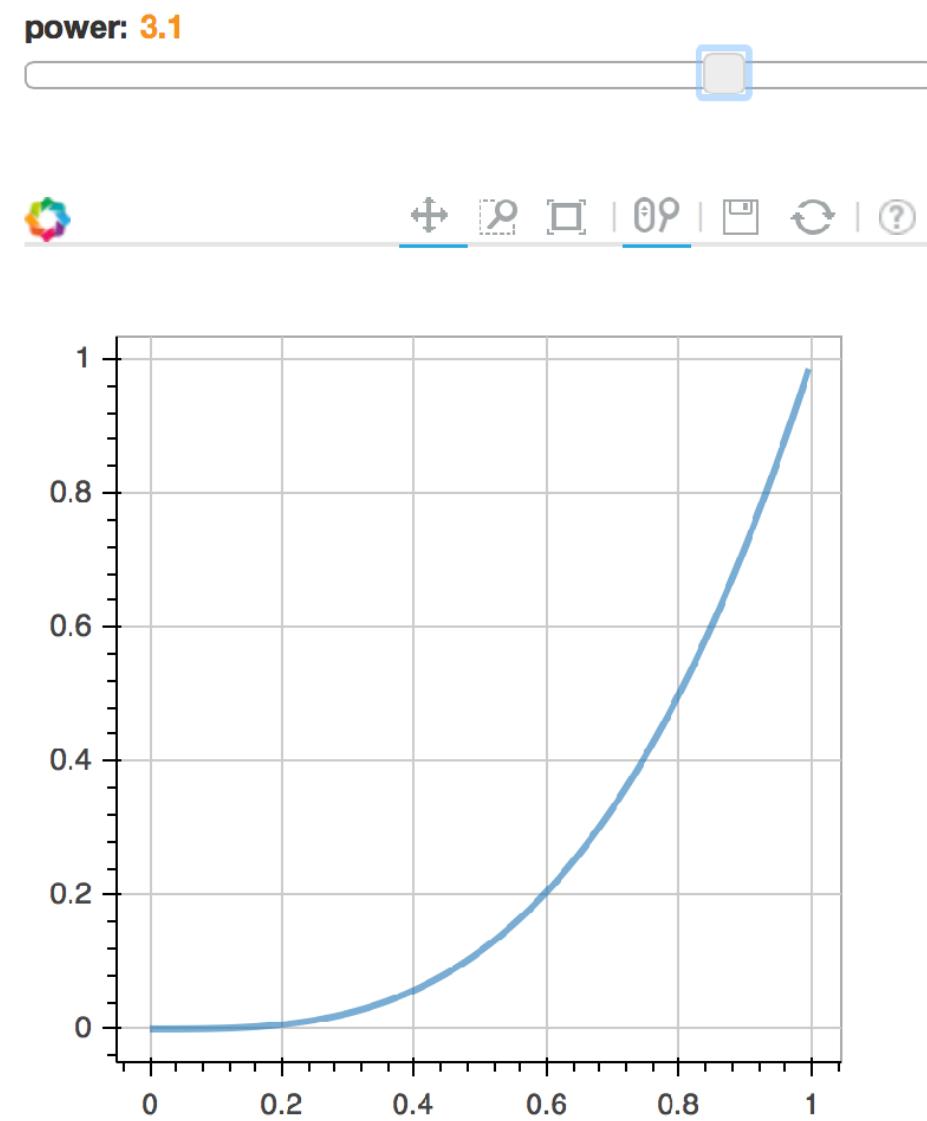
output_file("cat_heatmap.html")

palette = palette[::-1] # Reverse the color order so dark red is highest un
hm = HeatMap(df3, title="categorical heatmap", width=800, palette=palette)

show(hm)
```

power: 3.1





```
from bokeh.io import vform
from bokeh.models import Callback, ColumnDataSource, Slider
from bokeh.plotting import figure, output_file, show

x = [x*0.005 for x in range(0, 200)]
y = x

source = ColumnDataSource(data=dict(x=x, y=y))

plot = figure(plot_width=400, plot_height=400)
plot.line('x', 'y', source=source, line_width=3, line_alpha=0.6)

callback = Callback(args=dict(source=source), code="""
    var data = source.get('data');
    var f = cb_obj.get('value')
    x = data['x']
    y = data['y']
    for (i = 0; i < x.length; i++) {
        y[i] = Math.pow(x[i], f)
    }
    source.trigger('change');
""")
slider = Slider(start=0.1, end=4, value=1, step=.1, title="power", callback=callback)

layout = vform(slider, plot)

show(layout)
```

Using Bokeh

```
from bokeh.plotting import output_notebook  
output_notebook()
```

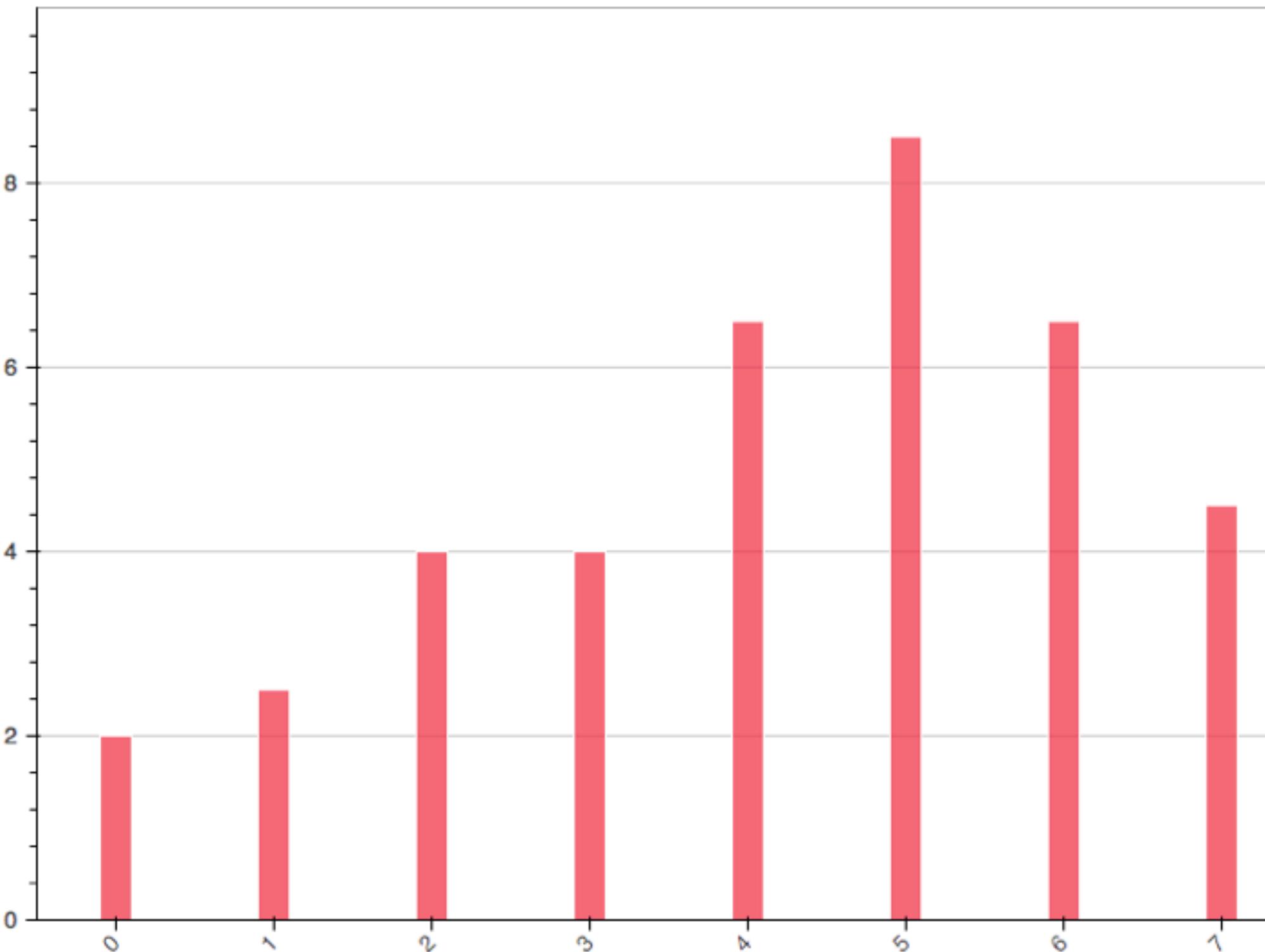
- Area (Overlapped & Stacked)
- Bar (Grouped & Stacked)
- BoxPlot
- Donut (Ick!)
- HeatMap
- Line
- Scatter
- Step
- Timeseries

```
from bokeh.charts import <chartname>
from bokeh.io import show
```

```
from bokeh.charts import Bar, show  
data = [2,3,4,4,7,9,7,5]  
  
barchart = Bar( data, notebook=True )  
show()
```

```
from bokeh.charts import Bar  
from bokeh.io import show  
data = [2,3,4,4,7,9,7,5]
```

```
barchart = Bar( data, notebook=True )  
show( bar chart )
```



Method Chaining



```
barchart =  
    Bar( data, title="My Chart", xlabel="Categories", ylabel="Value",  
notebook=True )  
  
show( barchart )
```

```
barchart =  
    Bar( data, title="My Chart", xlabel="Categories", ylabel="Value",  
notebook=True )  
  
barchart.show()
```

```
barchart3 = Bar(data)  
    .title( "My Chained Chart" )  
    .notebook( True )  
    .legend( True )  
    .xlabel( "Categories" )  
    .ylabel( "Value" )  
  
show( barchart3 )
```

Bokeh Config Options

- **title** (str): the title of your plot.
- **xlabel** (str): the x-axis label of your plot.
- **ylabel** (str): the y-axis label of your plot.
- **legend** (str, bool): the legend of your plot.
- **xscale** (str): the x-axis type scale of your plot.
- **yscale** (str): the y-axis type scale of your plot.
- **width** (int): the width of your plot in pixels.
- **height** (int): the height of you plot in pixels.
- **tools** (bool): to enable or disable the tools in your plot.
- **filename** (str or bool): the name of the file where your plot will be written.
- **server** (str or bool): the name of your plot in the server.
- **notebook** (bool):if you want to output (or not) your plot into the IPython notebook.

O'REILLY®

Security

BUILD BETTER DEFENSES

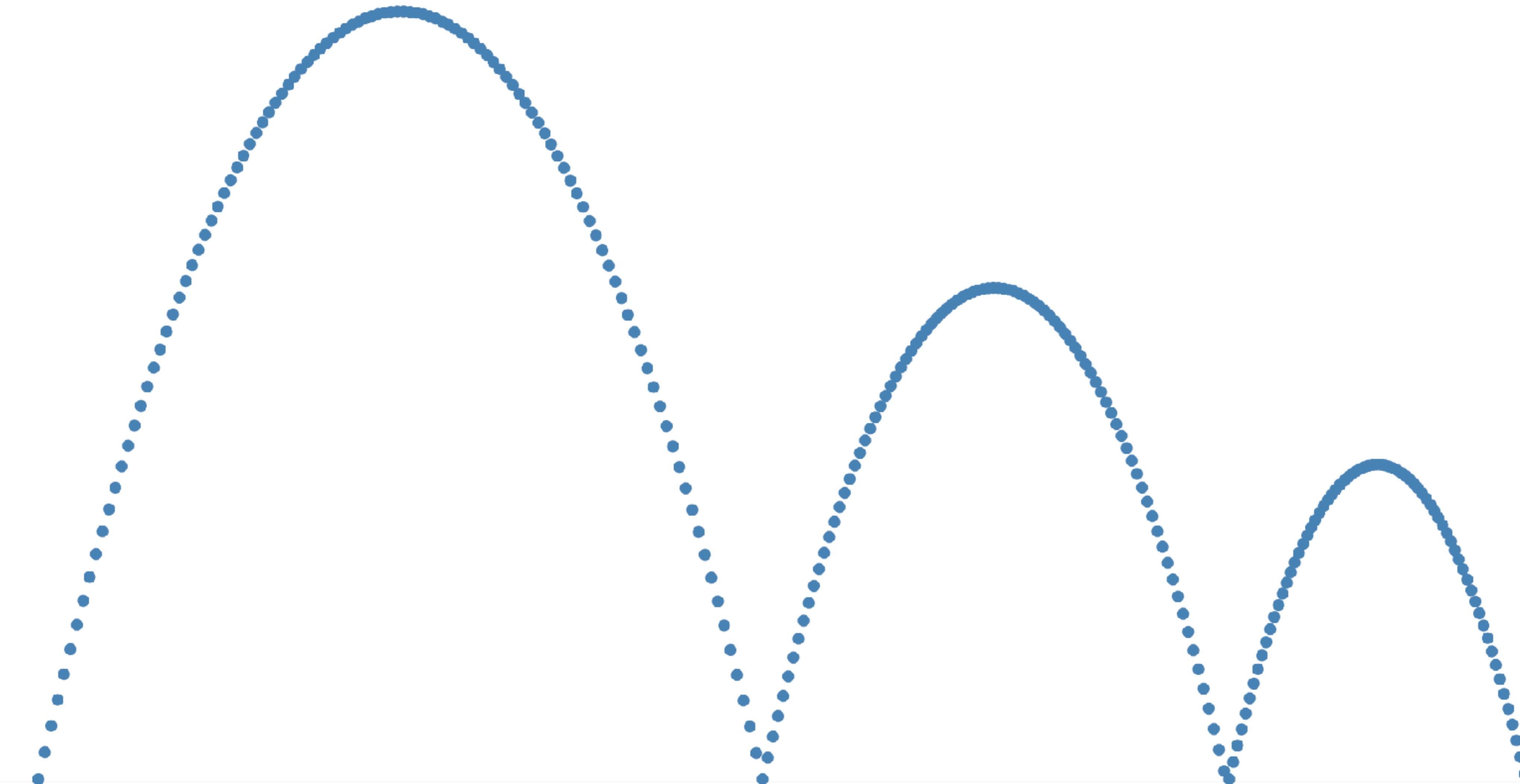
oreillysecuritycon.com
#oreillysecurity

Visualizations in Motion

Wrap up: Fun with Visualizations

$$x = v_0 t \cos(\theta)$$

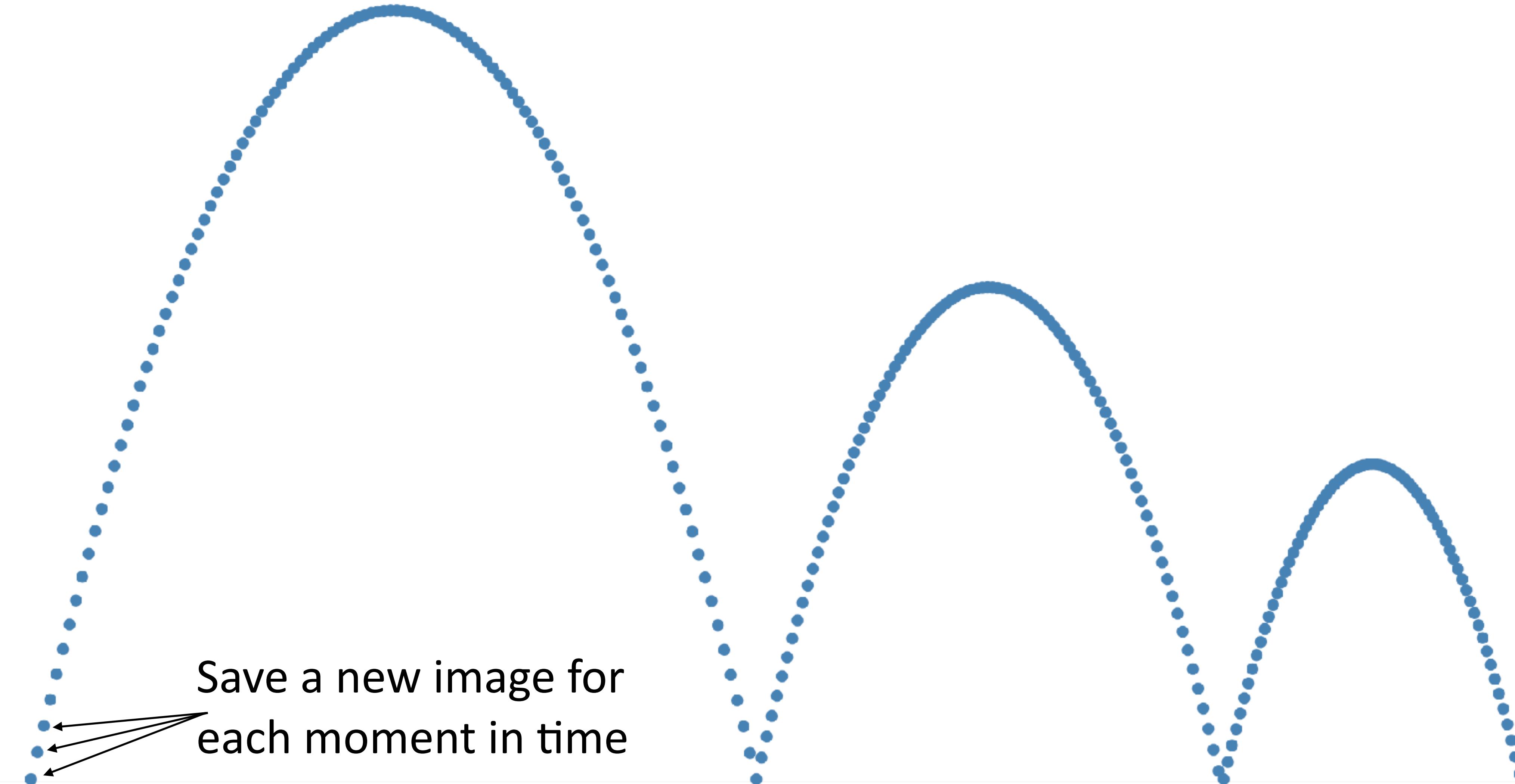
$$y = v_0 t \sin(\theta) - \frac{1}{2} g t^2$$



Wrap up: Fun with Visualizations

$$x = v_0 t \cos(\theta)$$

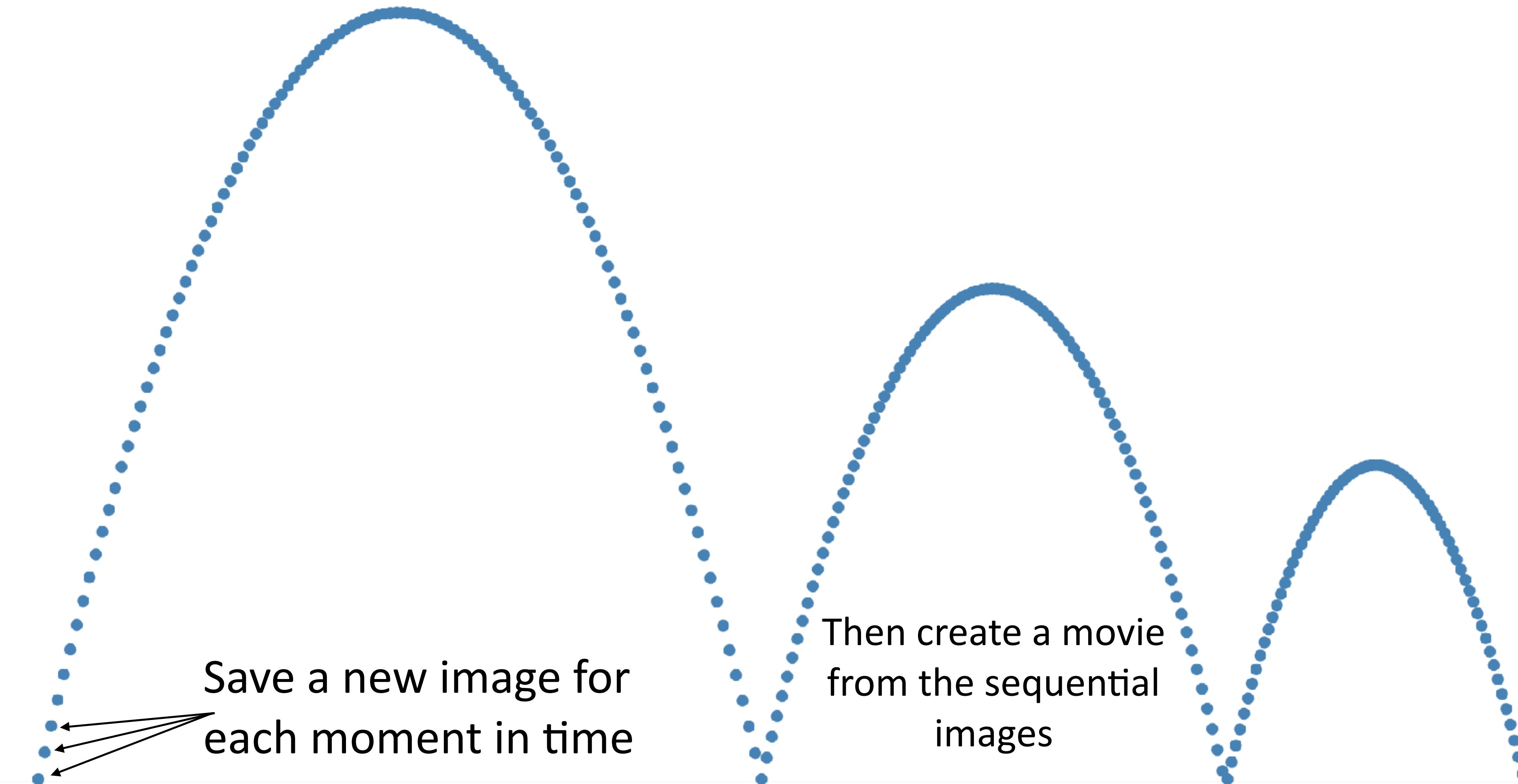
$$y = v_0 t \sin(\theta) - \frac{1}{2} g t^2$$



Wrap up: Fun with Visualizations

$$x = v_0 t \cos(\theta)$$

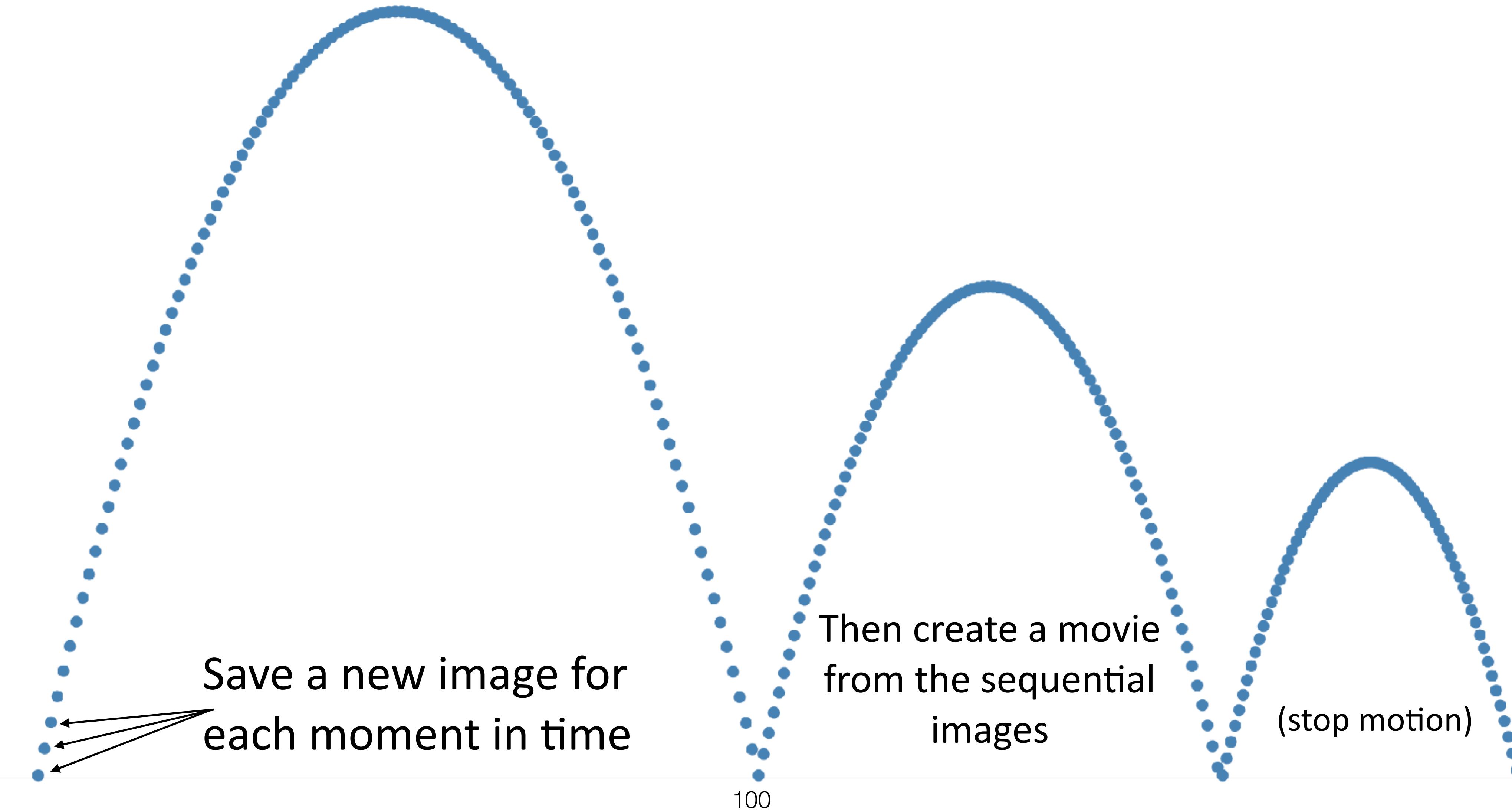
$$y = v_0 t \sin(\theta) - \frac{1}{2} g t^2$$



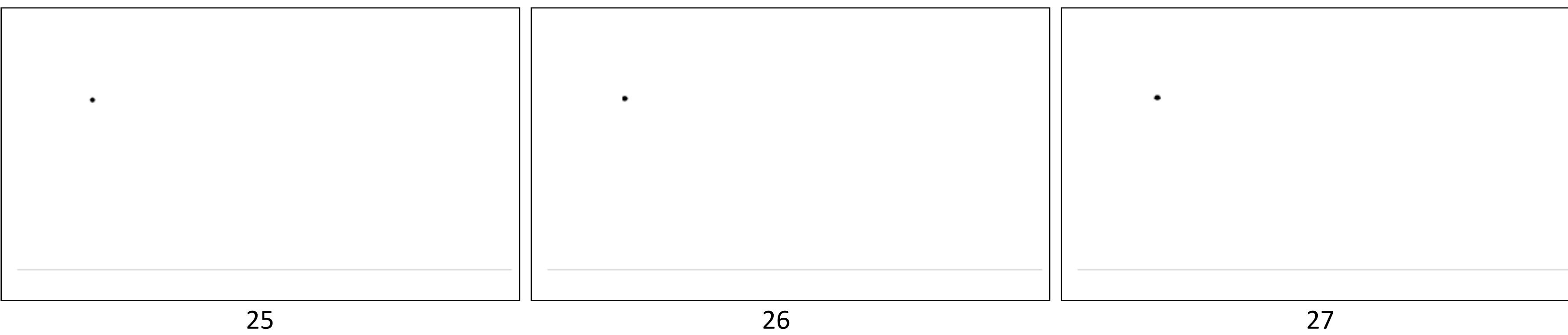
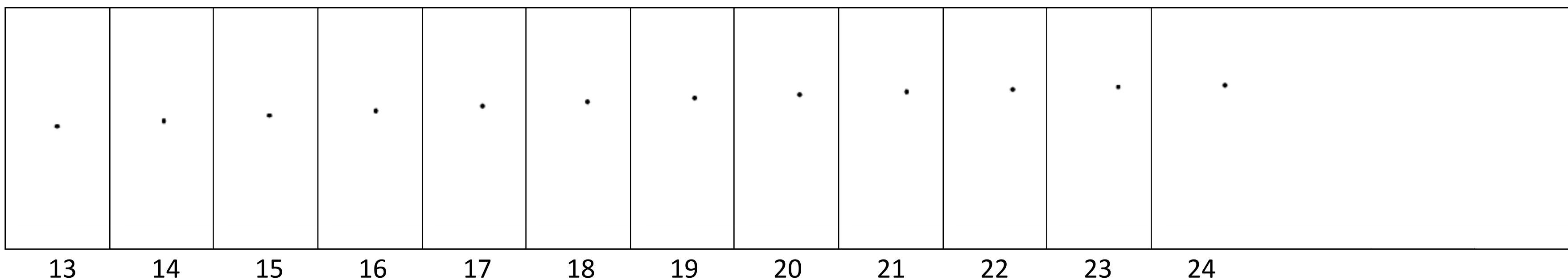
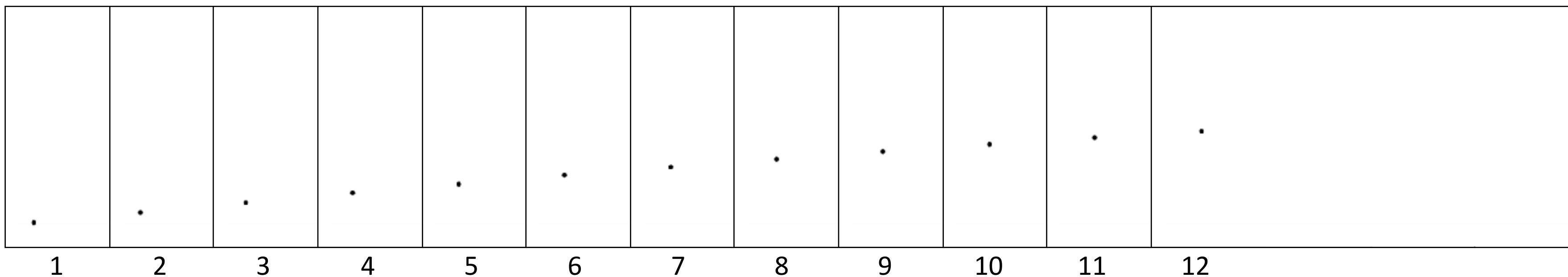
Wrap up: Fun with Visualizations

$$x = v_0 t \cos(\theta)$$

$$y = v_0 t \sin(\theta) - \frac{1}{2} g t^2$$



Creating Frame by Frame



Showtime!

But wait there's more

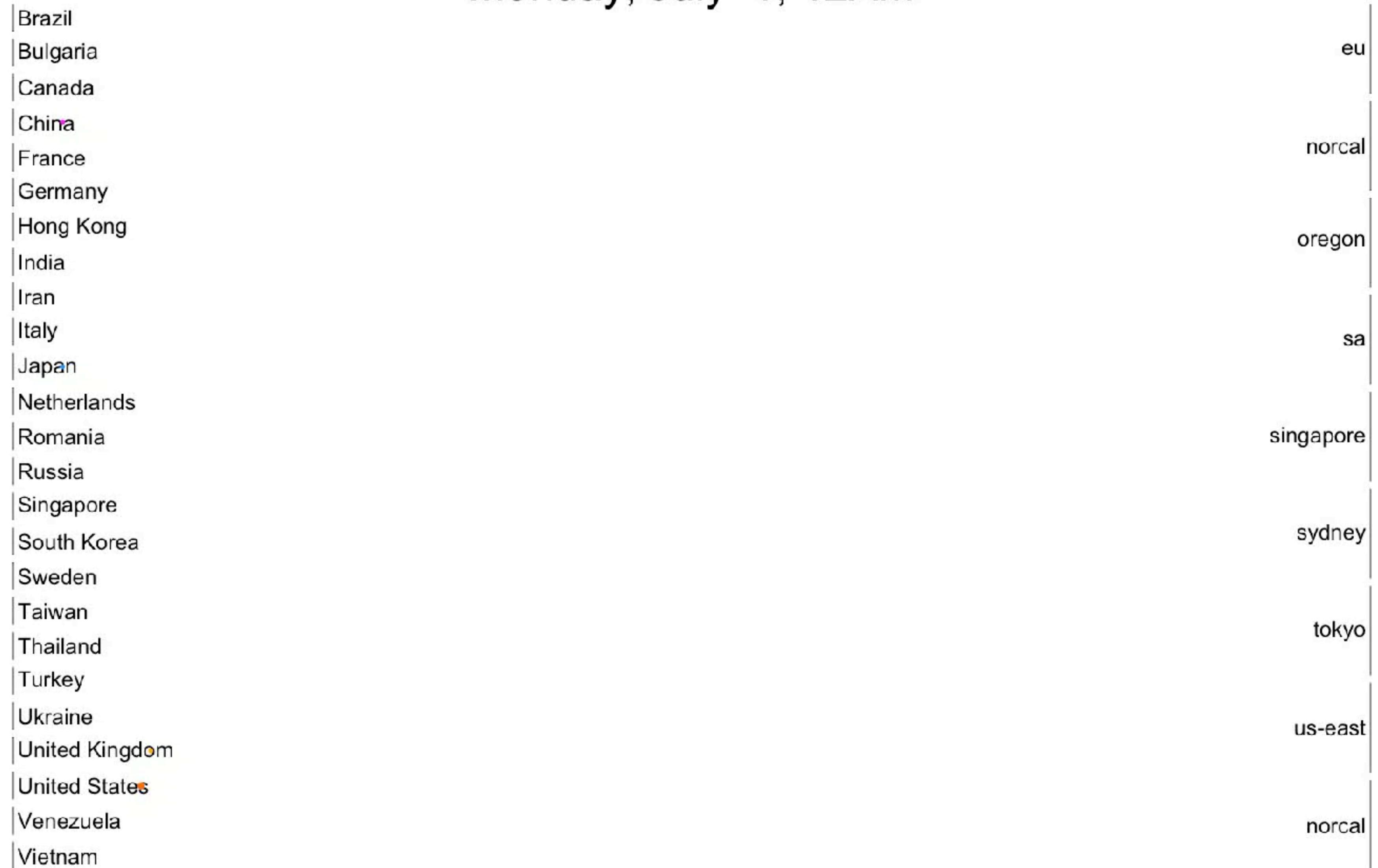


...and more!



Serious Fun

Monday, July 1, 12AM



UDP

Monday, July 1, 12AM

TCP

Argentina

Brazil

garia

nada

china

ance

Germany

Hong Kong

India

can

valy

upan

Orlando

Conclusion

69

1wan

10

Page 1

Kingdom

United States

Venezuela

tham

<http://datadrivensecurity.info>

Using avconv for stop motion

```
avconv -f image2 -y -i <basedir> -r 25 -b 256  
-s 1920x1080 -an <output.mp4>
```

{ LAB } time

Please go forth and visualize...

