The Science of Data Visualization

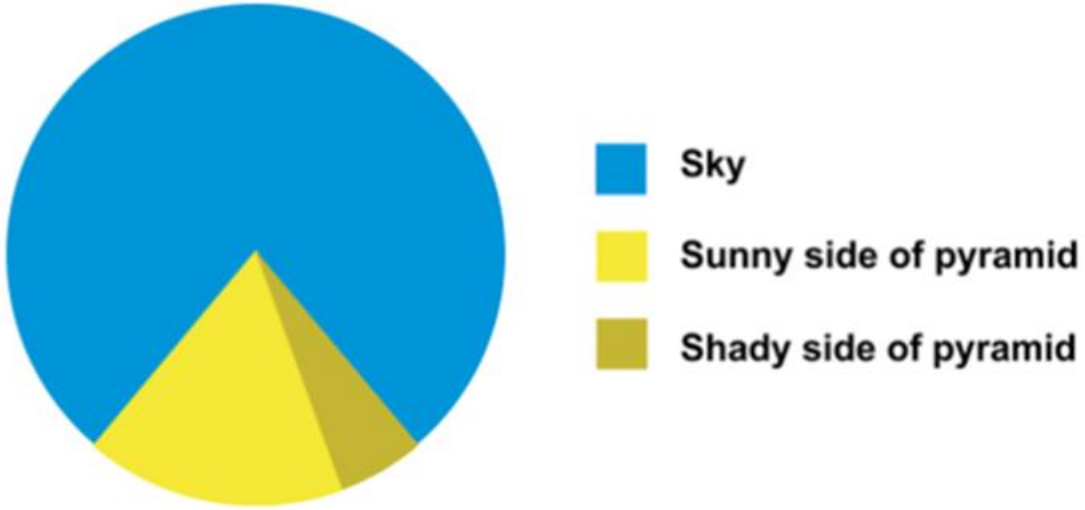
Ben Olsen
ben.olsen@matisia.com

PROFESSIONAL &
CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

Daniël Lakens
@lakens

Best pie chart ever
pic.twitter.com/B52J1g2l2y

6:35am - 30 Nov 14



4 Main Purposes of Data Visualization

- Analysis
  - Using visualizations to gain insight and assist in decision-making
- Communication
  - Passing a message to others.
- Monitoring
  - Tracking information about performance.
- Planning
  - Predicting and preparing for the future.

4 Main Purposes of Data Science?

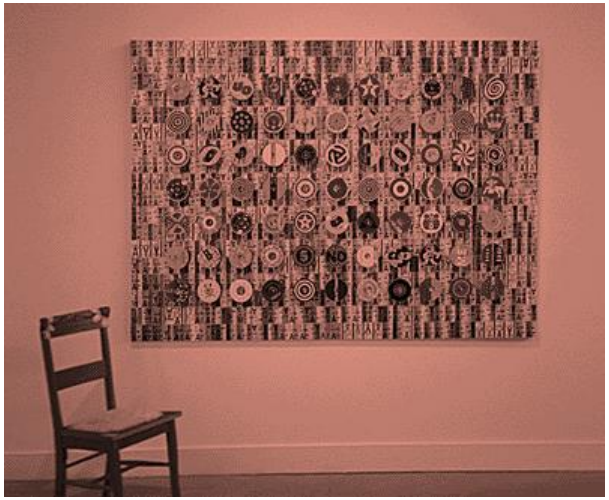
- Analysis
  - Using visualizations to gain insight and assist in decision-making
- Communication
  - Passing a message to others.
- Monitoring
  - Tracking information about performance.
- Planning
  - Predicting and preparing for the future.

Question…

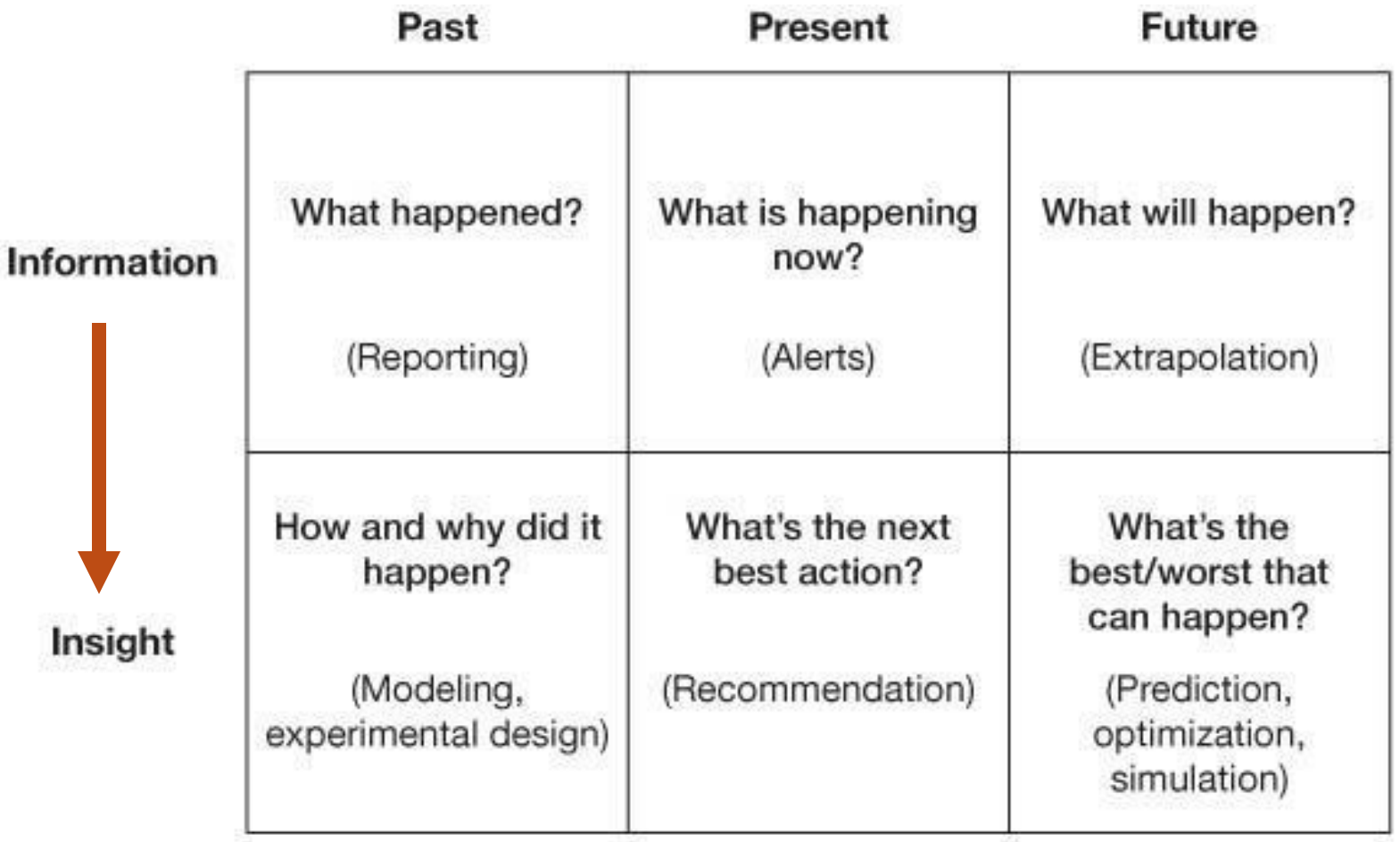
- Think about two examples of data visualization used in either your current or past workplace.
- Which of the purposes did it serve?
- Are there things you can think of that would have made it serve that purpose better?

At it’s core, data visualization is
a cognitive tool for
understanding data and
assisting decision-making as
effectively as possible

Otherwise, it’s an expensive art project



From Information to Insight

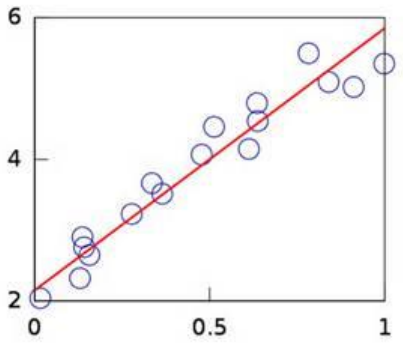


What data product would be the output here?

– Regression:

- Use statistical software to create a trend line model that gives a numeric value for a future/unknown data point.
- Example:

"Sales have been increasing by approximately 5% for the past 10 years. Thus, we can predict that sales next year are 95% likely to be between 3% and 7% higher than this year."



Here?

-Classification:

- Predict a value/class/grouping for a person/company/entity based on data for other people/companies/entities.
- Example:

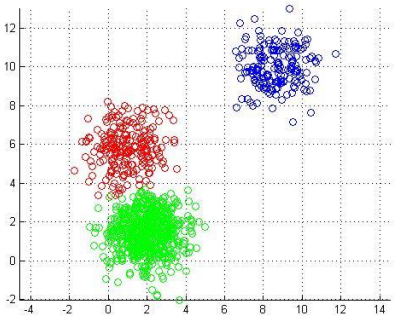
"I know that all males with red hair and size 12 shoes in our data set own a sports car. Therefore, I can predict that all future males with red hair and size 12 shoes will own a sports car. We can reliably base our decisions on this assumption."

Here?

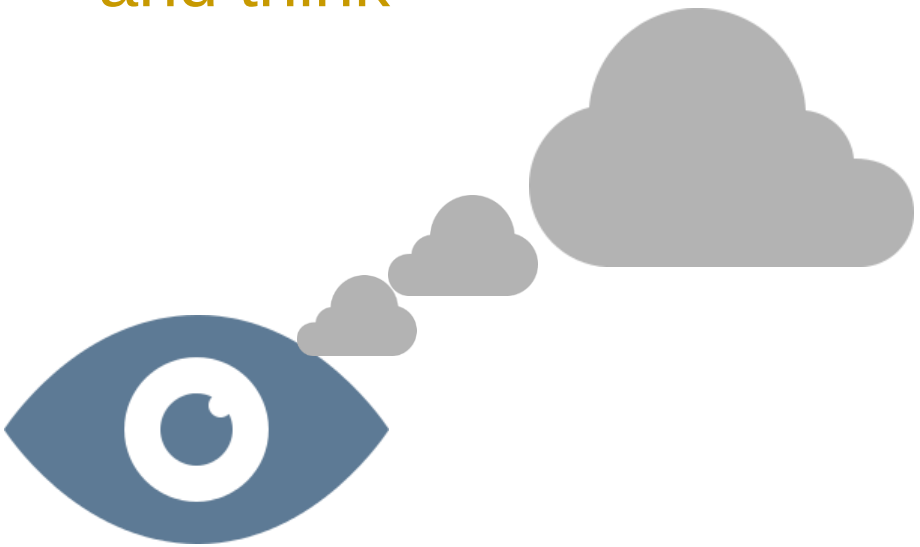
-Clustering:

- Group items together based on similarities among these items.
- Example:

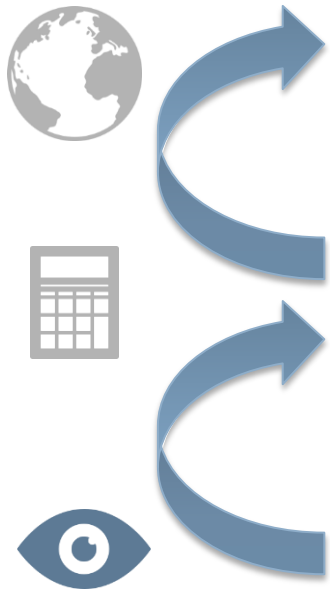
"We grouped our customers based on gender, age, and economic status. We discovered that middle class women over 70 buy our product at 10 times the rate of other customers. We should target these women in our mailings in the future."



World-class Data Visualization is done with an understanding of how we see and think



Perceptual building blocks of data visualization

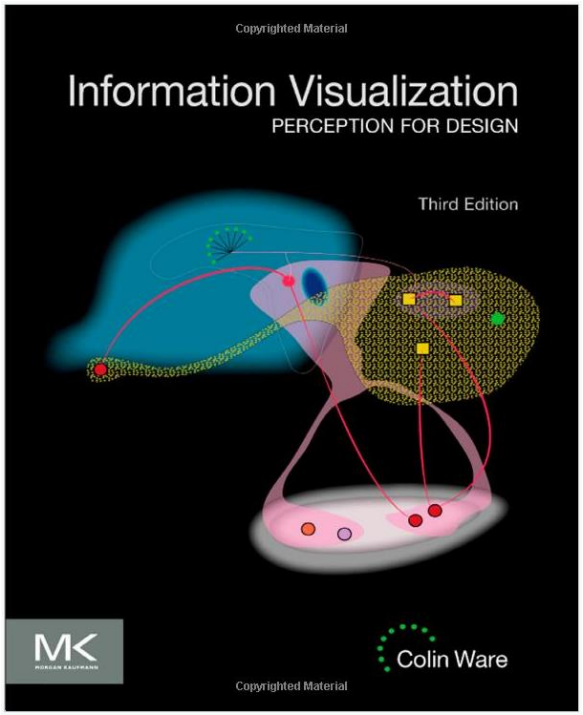


Context

Quantitative Reasoning

Visual Perception

Knowing how this works changes how you build even basic tables!!!



"Following perception-based rules, we can present our data in such a way that the important and informative patterns stand out. If we disobey these rules, our data will be incomprehensible or misleading."

-Colin Ware, from Information Visualization 3rd Edition

Foundation for ~~Guiding the Eye~~ Honest Charting

Visual Processing Stages

1) Rapid Parallel Processing

Edges, orientation, color, texture, motion

Transitory: briefly held ‘the mind’s eye’

Bottom-up, data-driven

2) Serial Goal-Directed Processing

Object recognition: visual attention & memory

Slow and serial

Uses both short and long-term memory

Top-down processing

(1) And (2) work in concert to create the way we see the world.

Preattentive Processing

- Certain basic visual properties are detected **immediately** by low-level visual system
- No need to focus attention
- Can mislead viewer
- These things “Pop” or “Pop-out”
- Tasks that can be performed in less than 200 to 250 milliseconds on a complex display

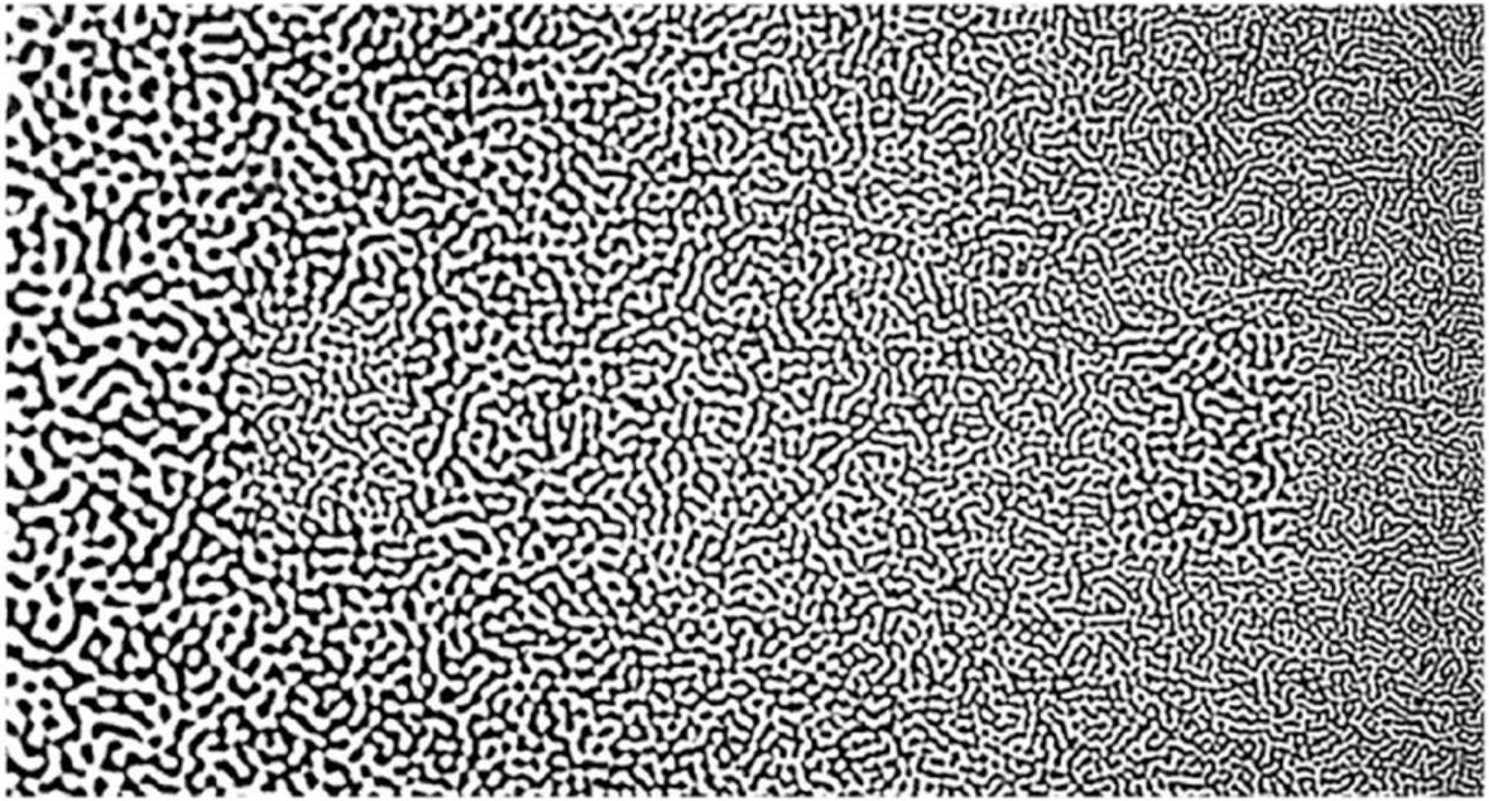
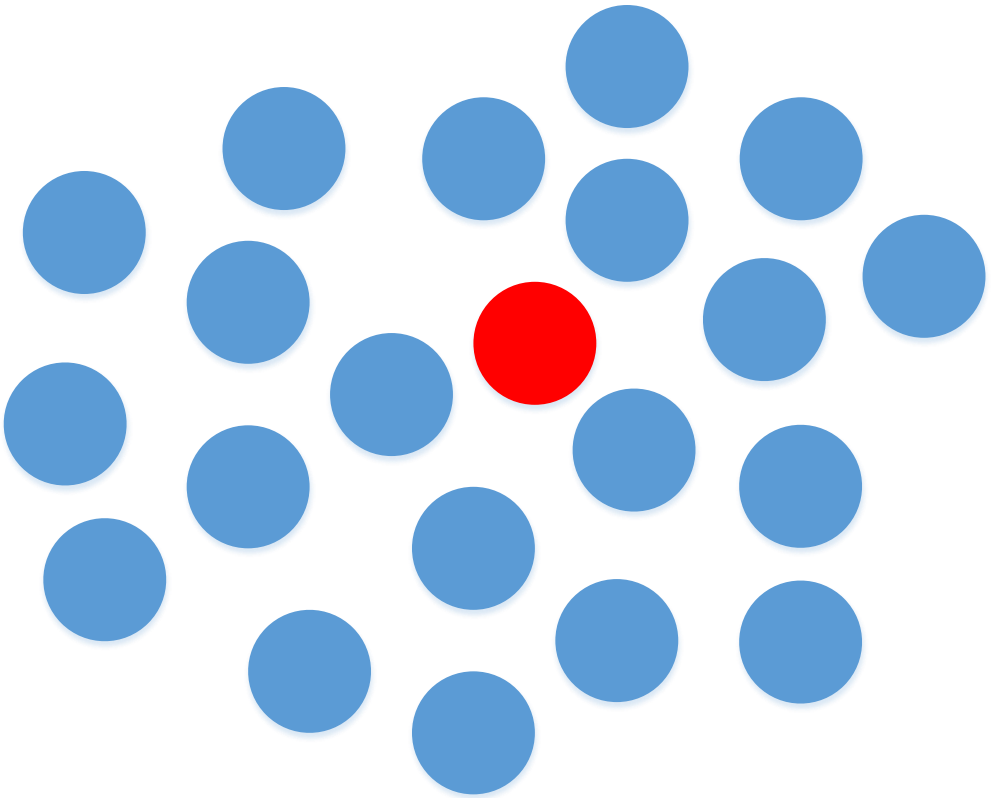


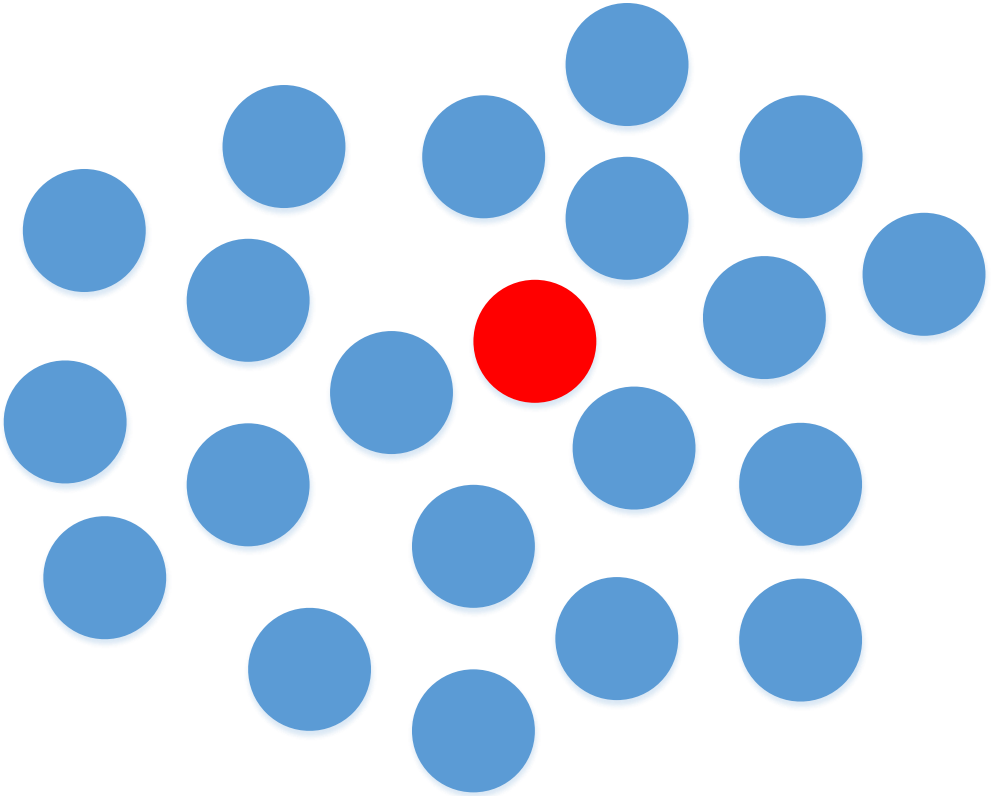
Figure 3.5. This image appears in *Information Visualization: Perception for Design*, Second Edition, Colin Ware, Morgan Kaufmann Publishers, San Francisco CA, 2004, p. 171.

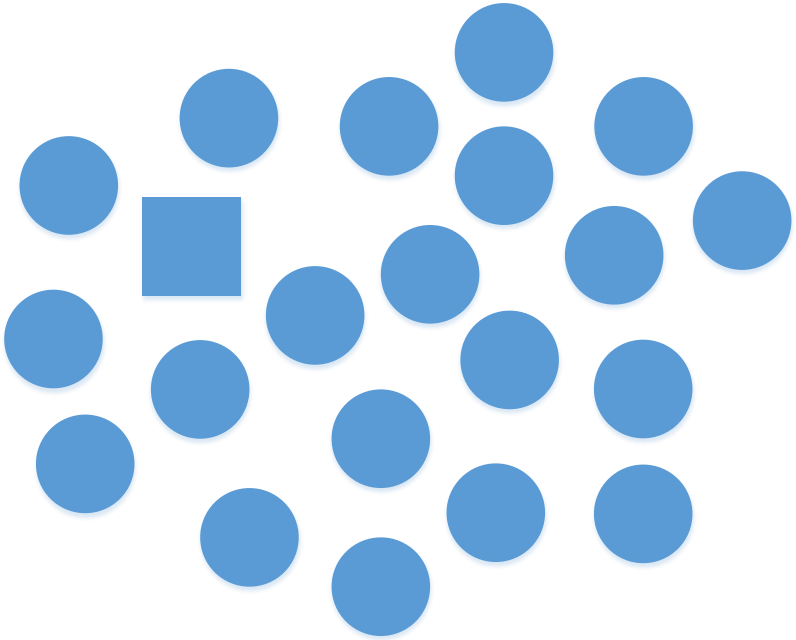
98734970275647902194720727676473490985403208029007305901270238008374082078720275008083435 4

9873497027564790219472072767647349098540320802900730590127023800837408207872027500808343454

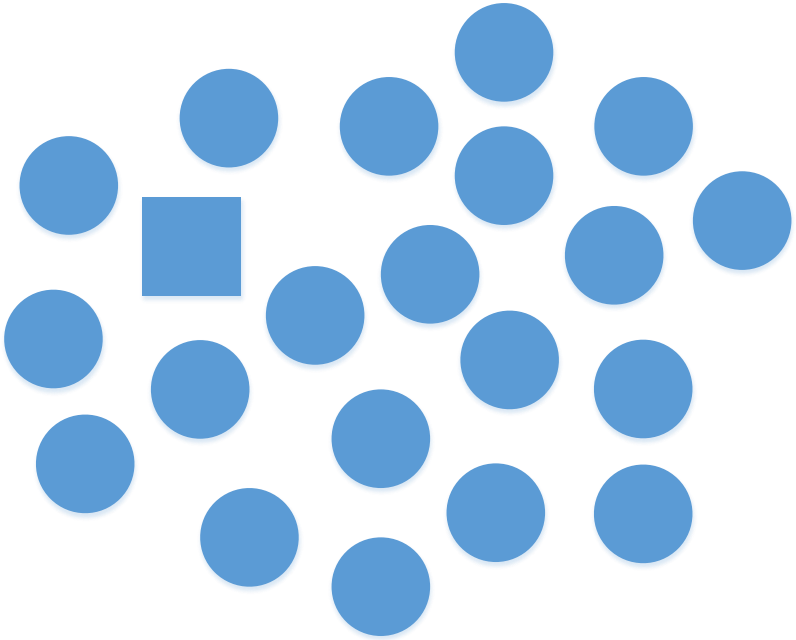


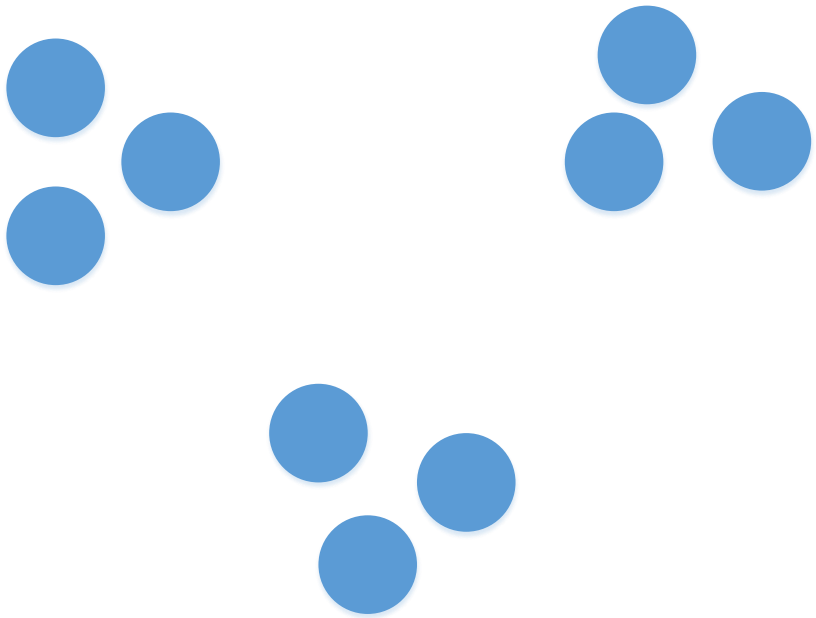
Color



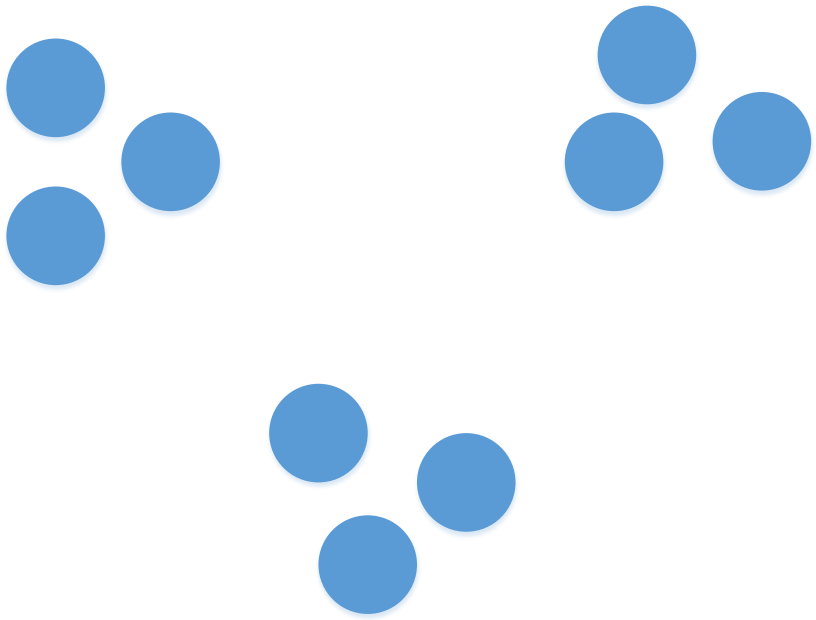


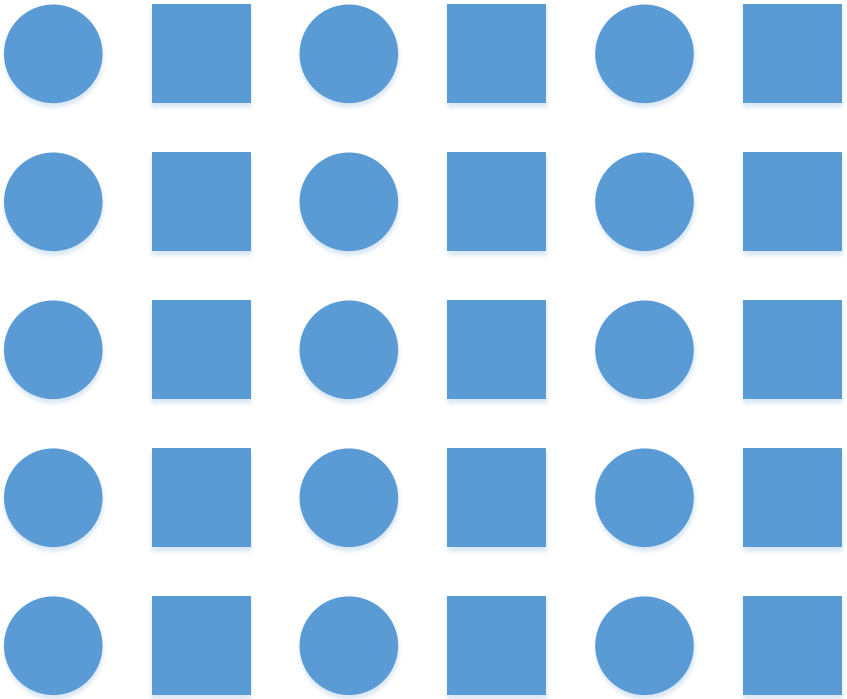
Curvature



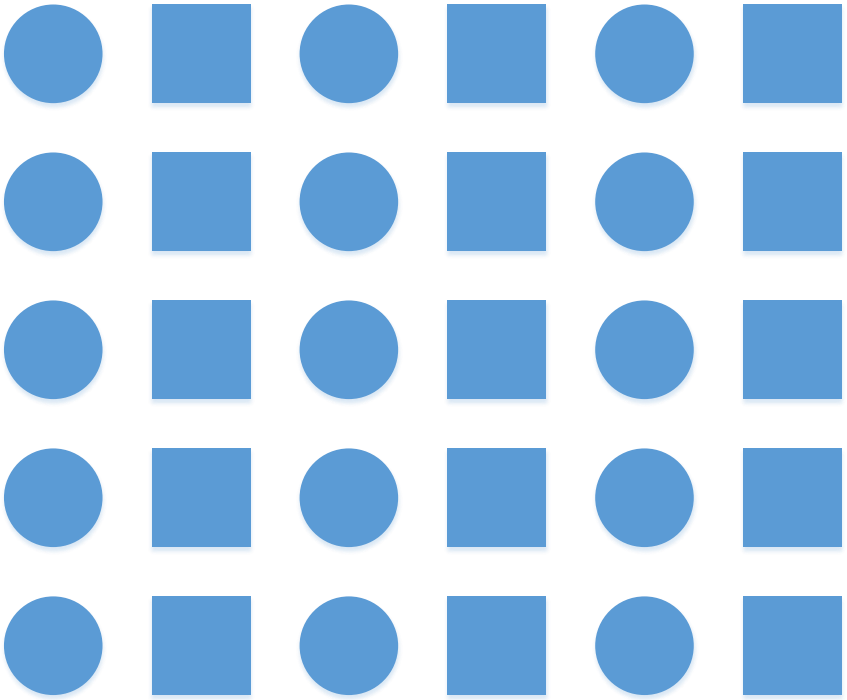


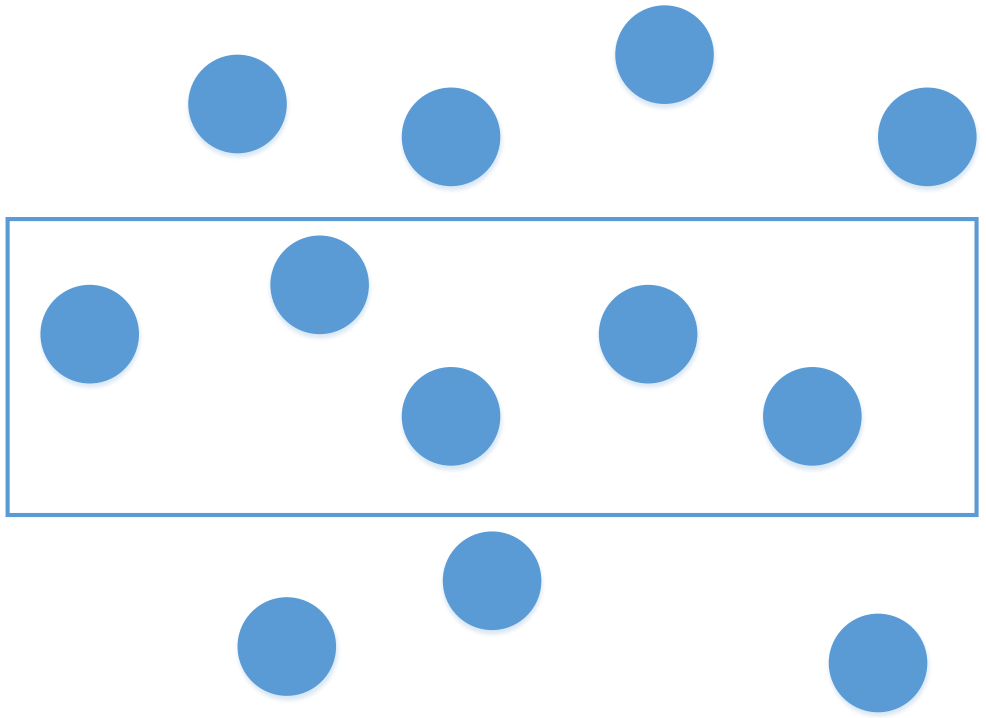
Proximity



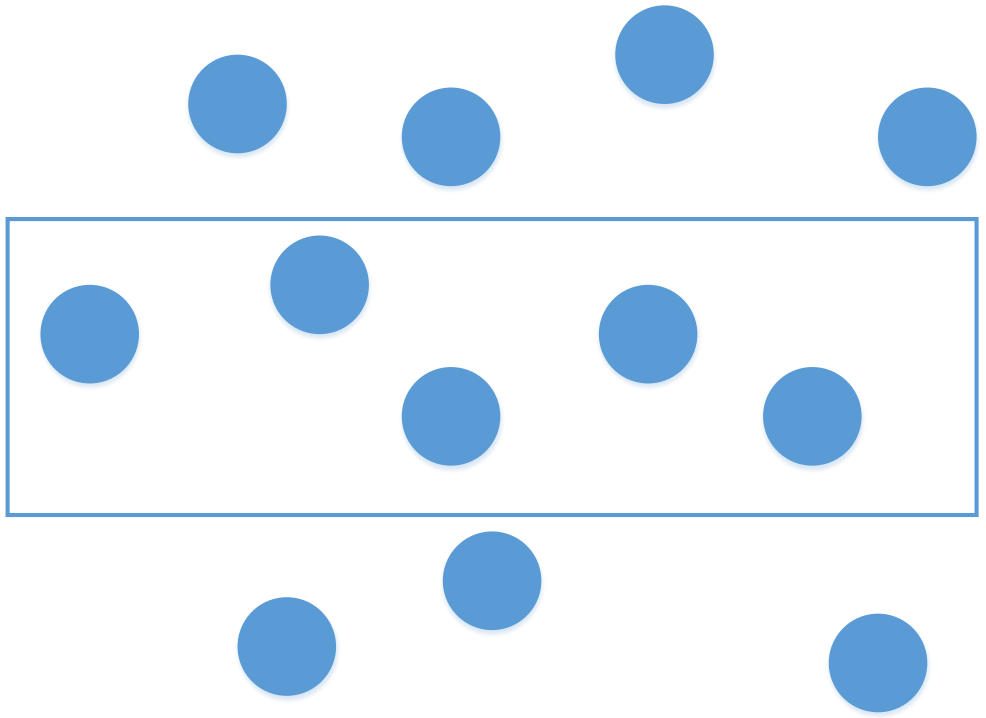


Similarity





Enclosure



Preattentive Attributes of Perception

Form
- Length
- Width
- Line orientation
- Size
- Shape
- Curvature
- Enclosure
- Blur

Color
- Hue
- Intensity

Position
- 2-D position
- Spatial Grouping
- Direction of Motion

Preattentive attributes of visual perception

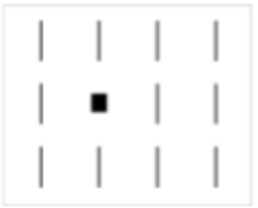
Form


Length


Width


Orientation


Size


Shape


Curvature


Enclosure


Blur

Color


Hue


Intensity

Position


2-D position


Spatial Grouping

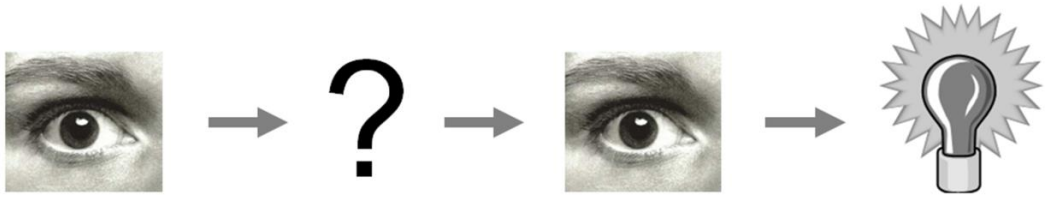
Motion

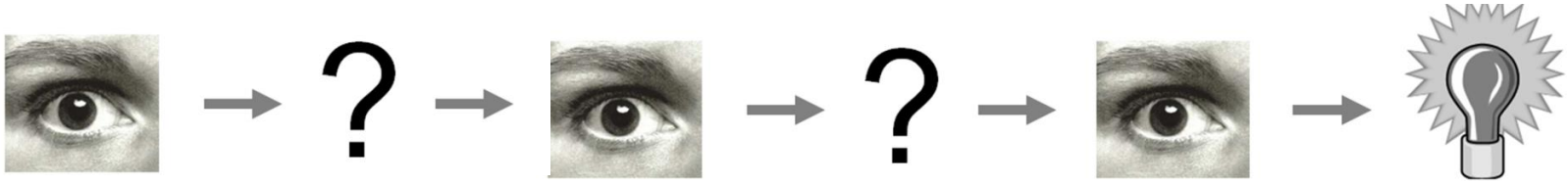

Direction of Motion

Analytical Navigation

Directed

Exploratory





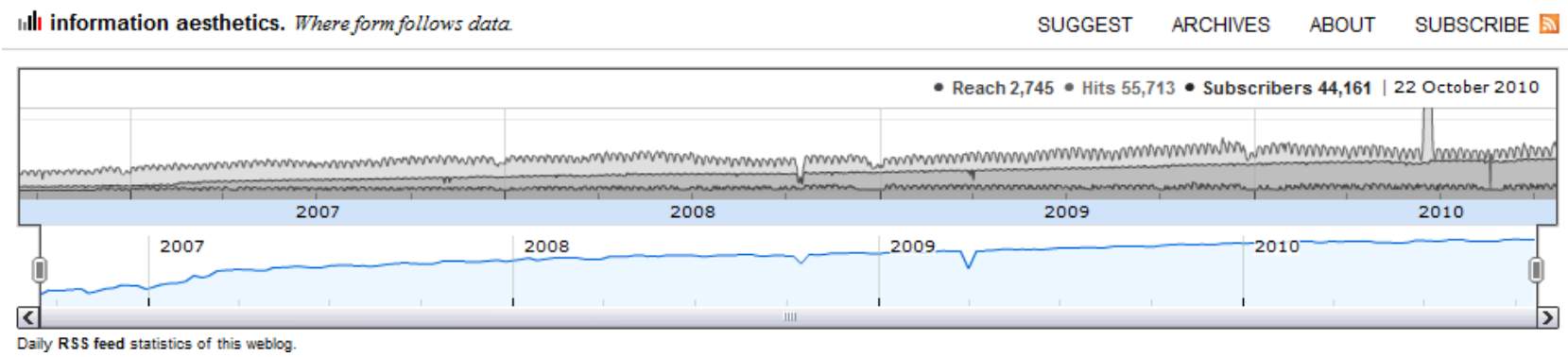
Schneiderman's Mantra

Overview First

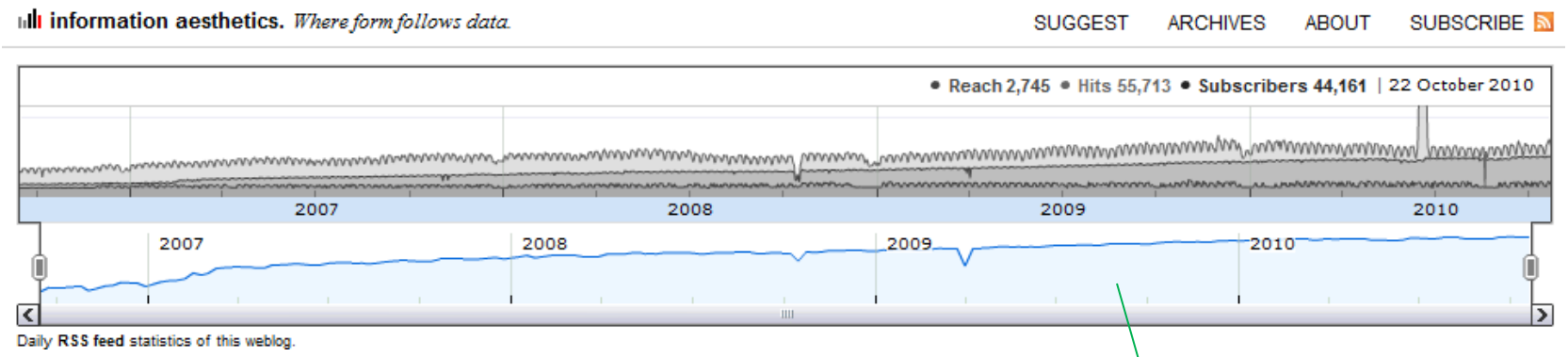
Zoom & Filter

Details on Demand

Overview First

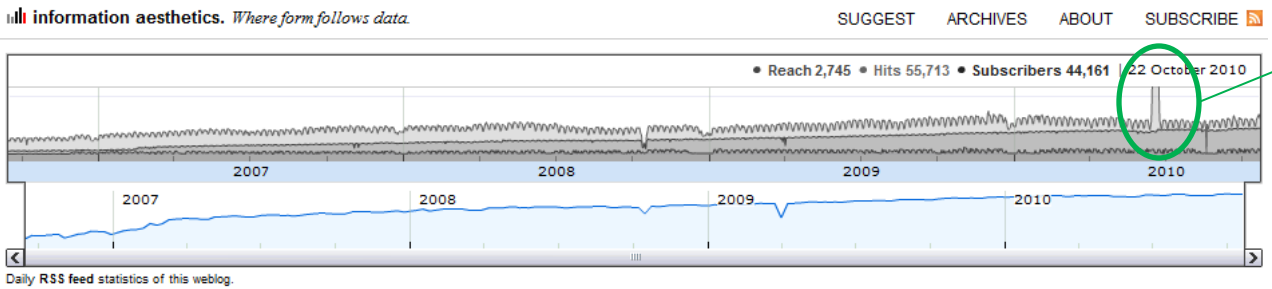


Overview First



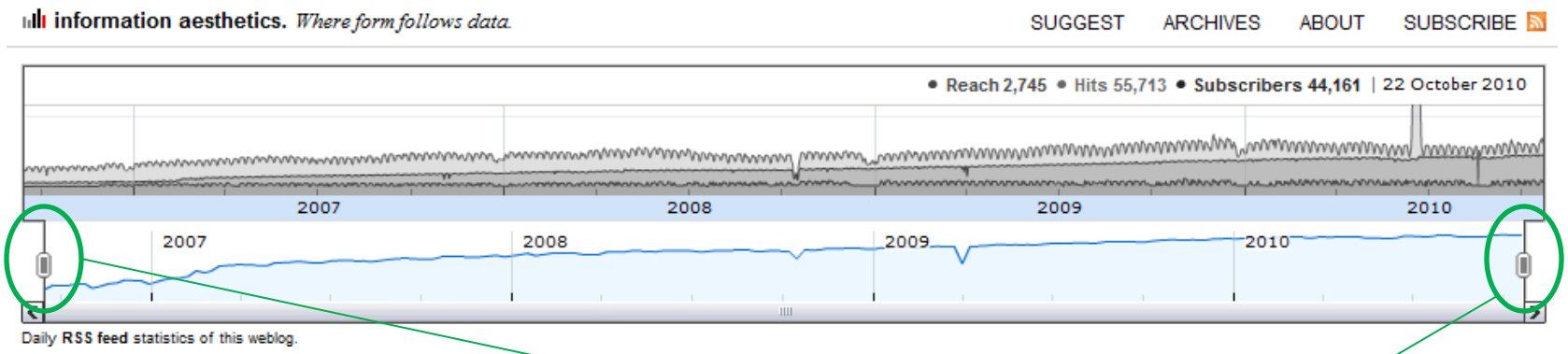
Total Subscribers

Zoom & Filter



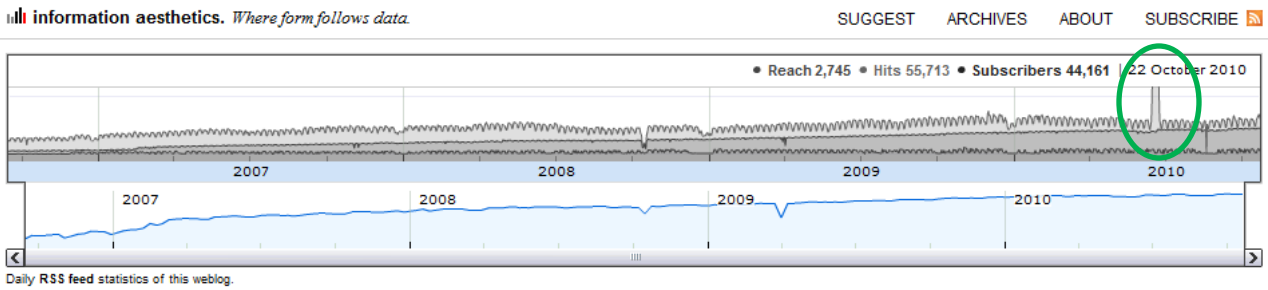
What's this?

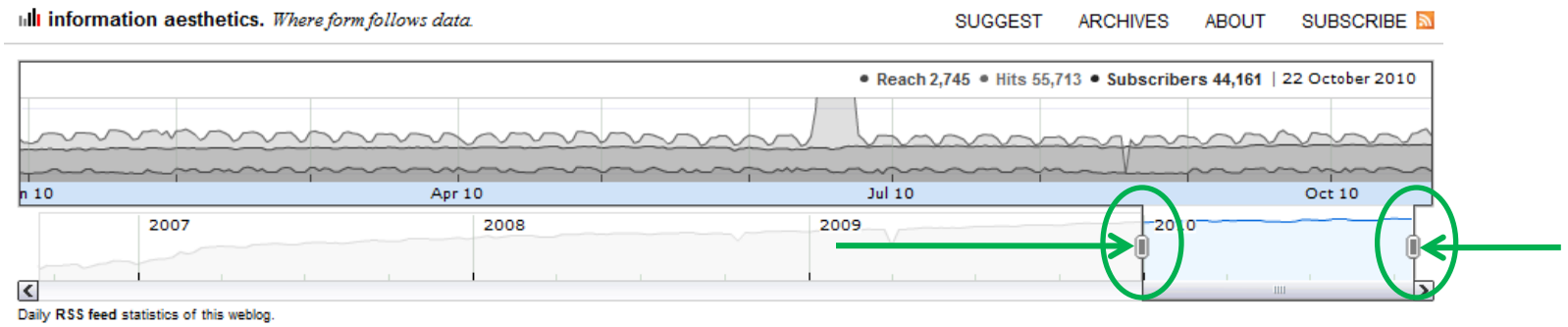
Zoom & Filter



Look, I can adjust the range…

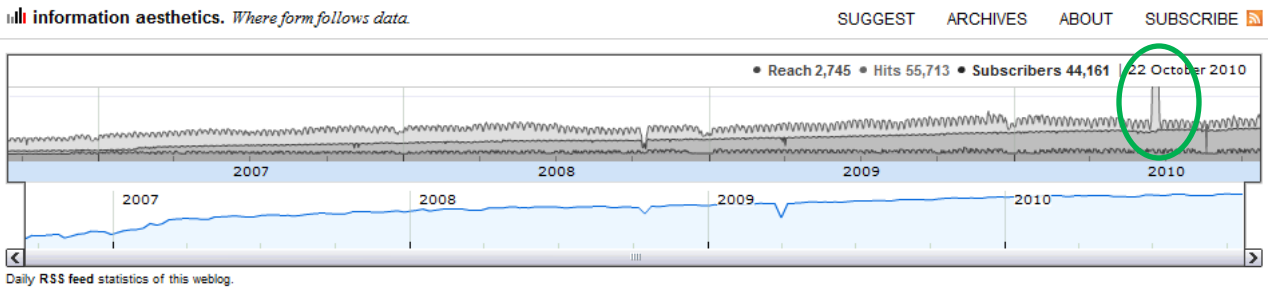
Zoom & Filter

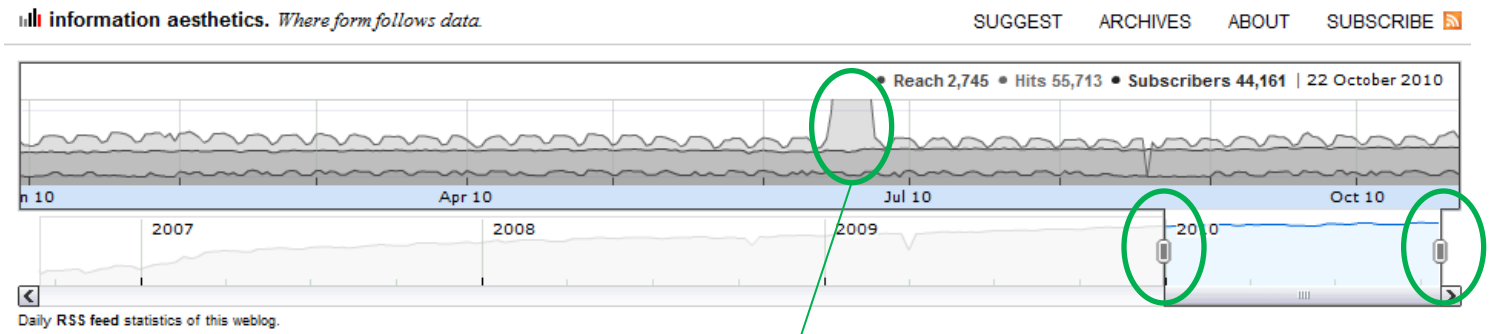




Getting closer…

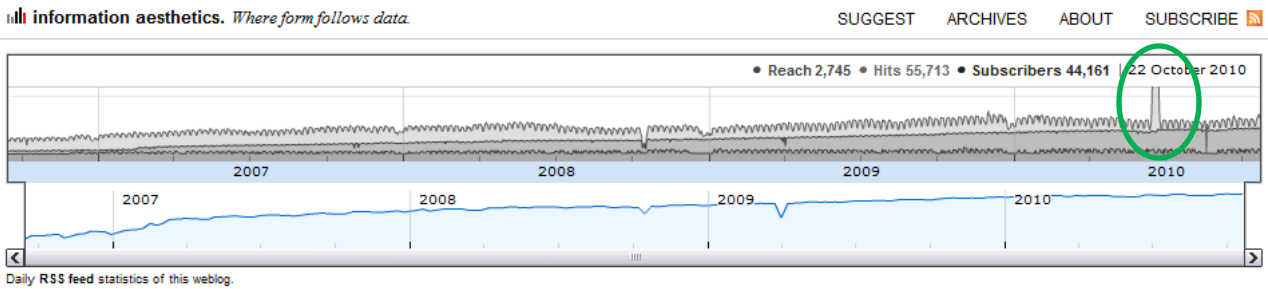
Zoom & Filter

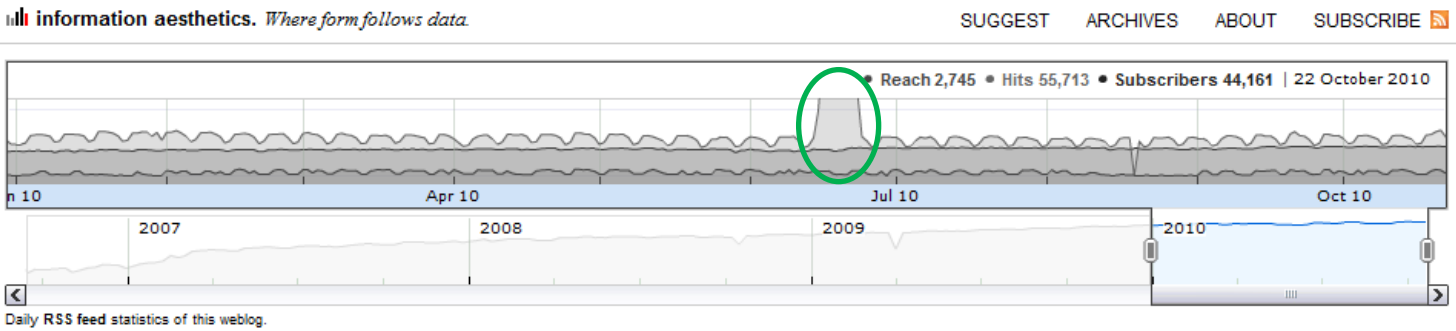


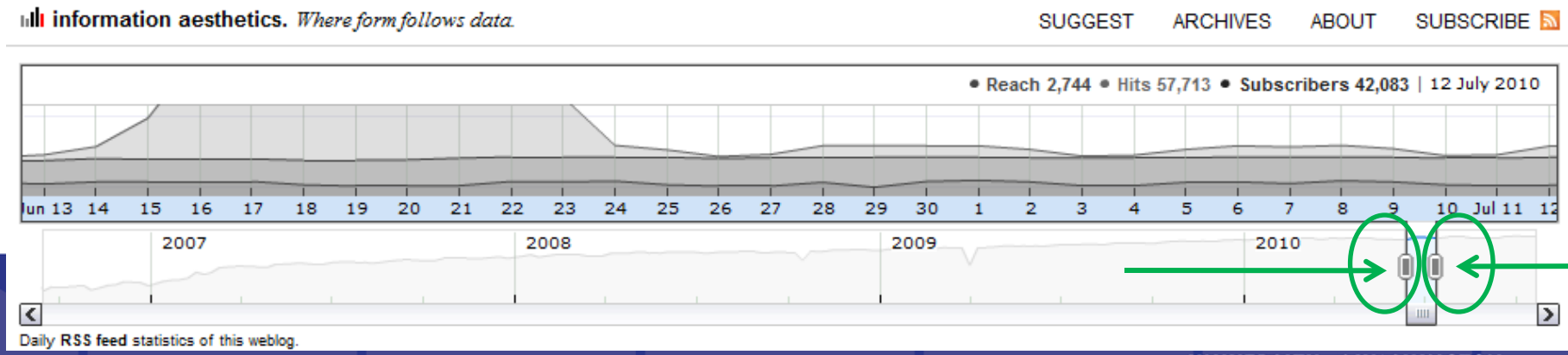


What dates are those?

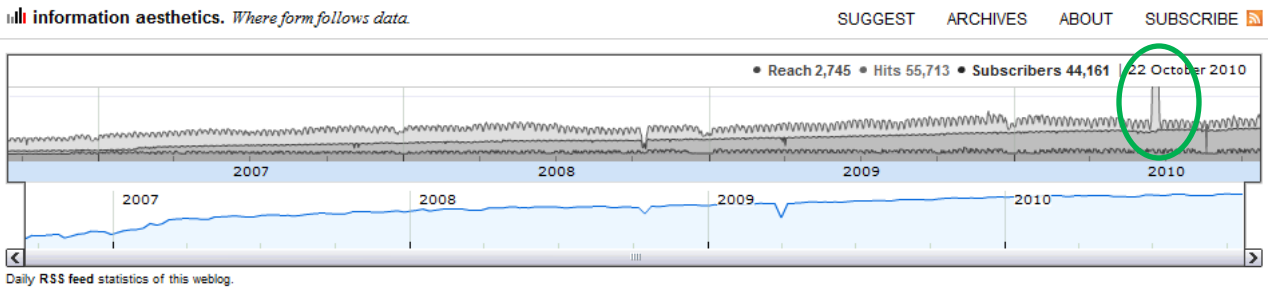
Zoom & Filter

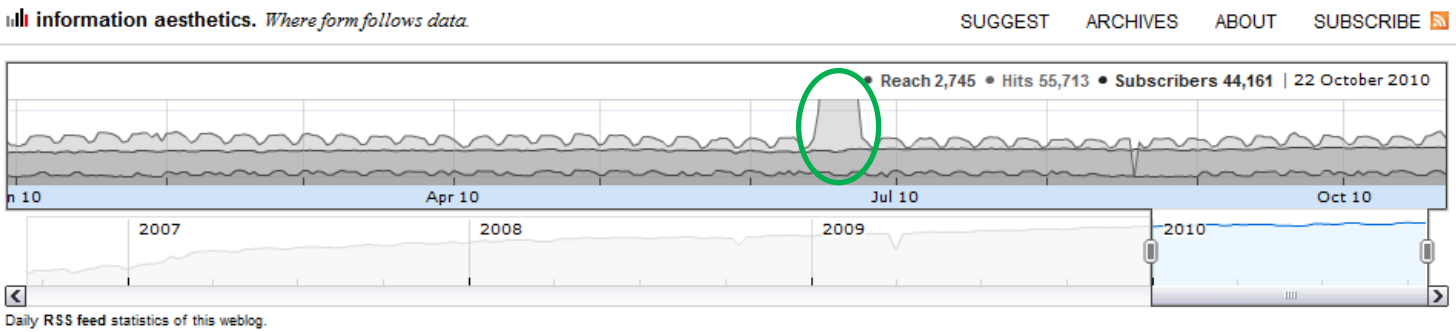

information aesthetics. Where form follows data.
SUGGEST
ARCHIVES
ABOUT
SUBSCRIBE
Reach 2,745 • Hits 55,713 • Subscribers 44,161 | 22 October 2010
2007
2008
2009
2010
Daily RSS feed statistics of this weblog.


information aesthetics. Where form follows data.
SUGGEST
ARCHIVES
ABOUT
SUBSCRIBE
Reach 2,745 • Hits 55,713 • Subscribers 44,161 | 22 October 2010
Apr 10
Jul 10
Oct 10
2007
2008
2009
2010
Daily RSS feed statistics of this weblog.

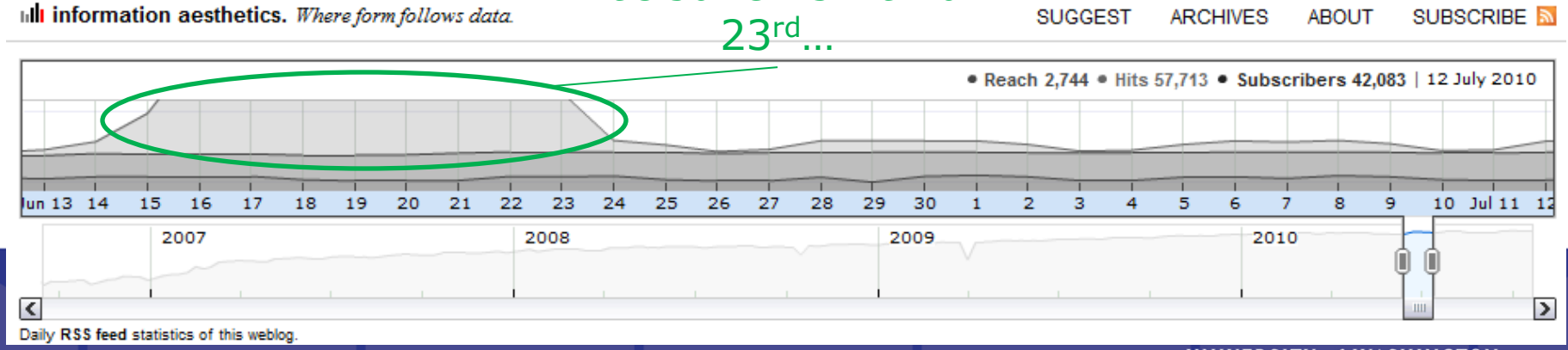

information aesthetics. Where form follows data.
SUGGEST
ARCHIVES
ABOUT
SUBSCRIBE
Reach 2,744 • Hits 57,713 • Subscribers 42,083 | 12 July 2010
Jun 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 1 2 3 4 5 6 7 8 9 10 Jul 11 12
2007
2008
2009
2010
Daily RSS feed statistics of this weblog.

Zoom & Filter

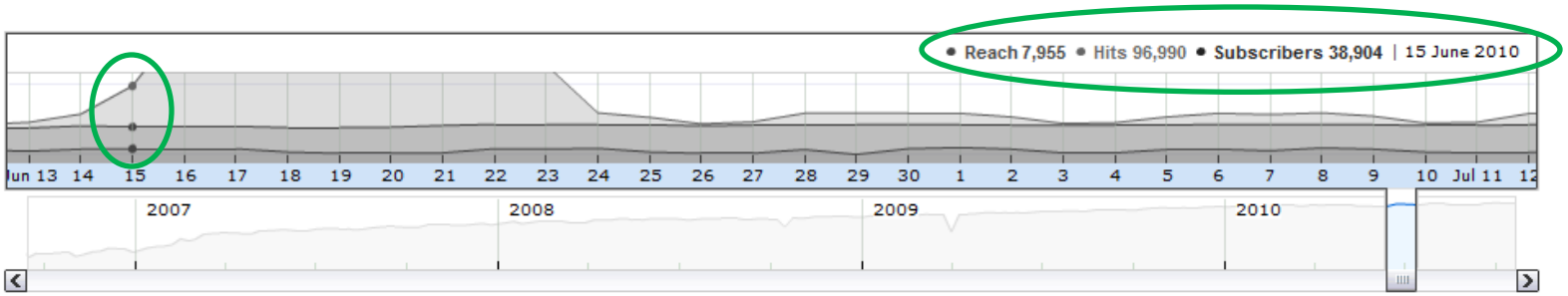




It's June 15th thru 23rd...

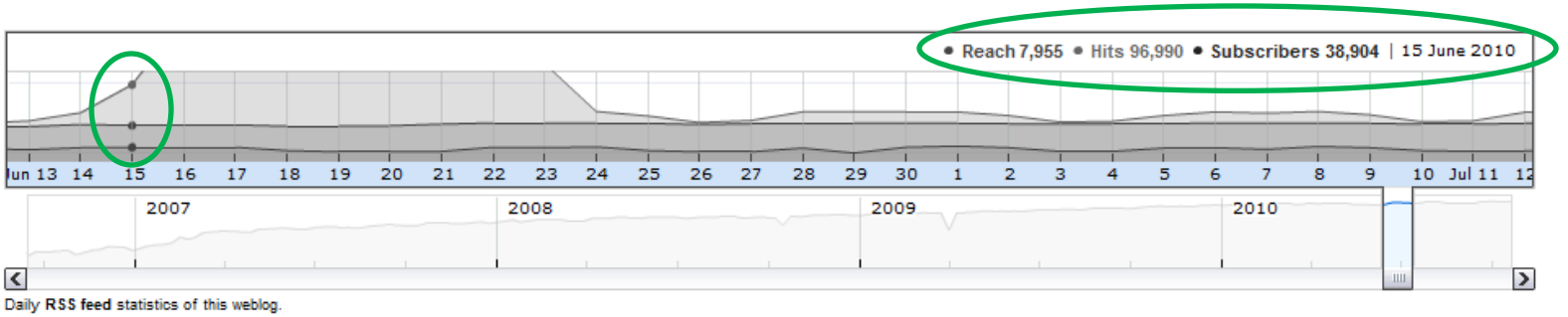


Details on Demand

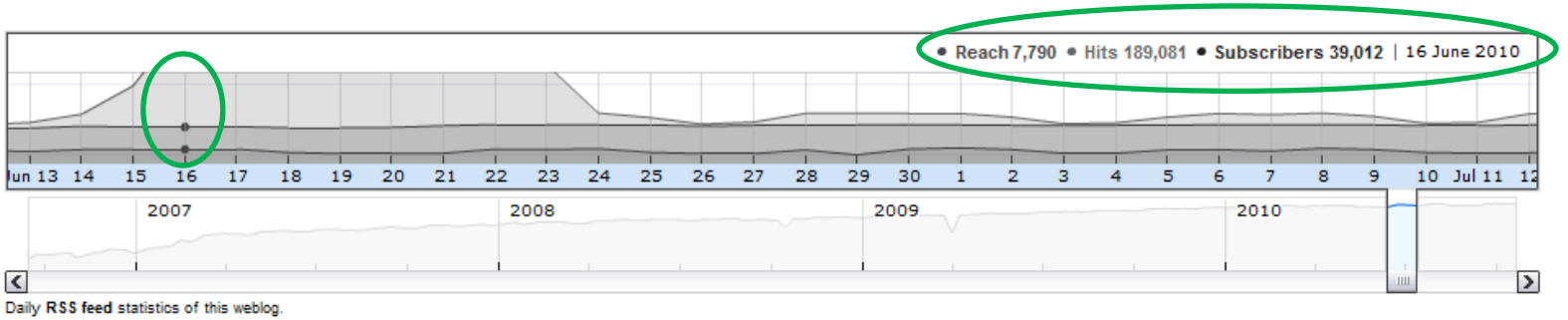


Daily **RSS feed** statistics of this weblog.

Details on Demand

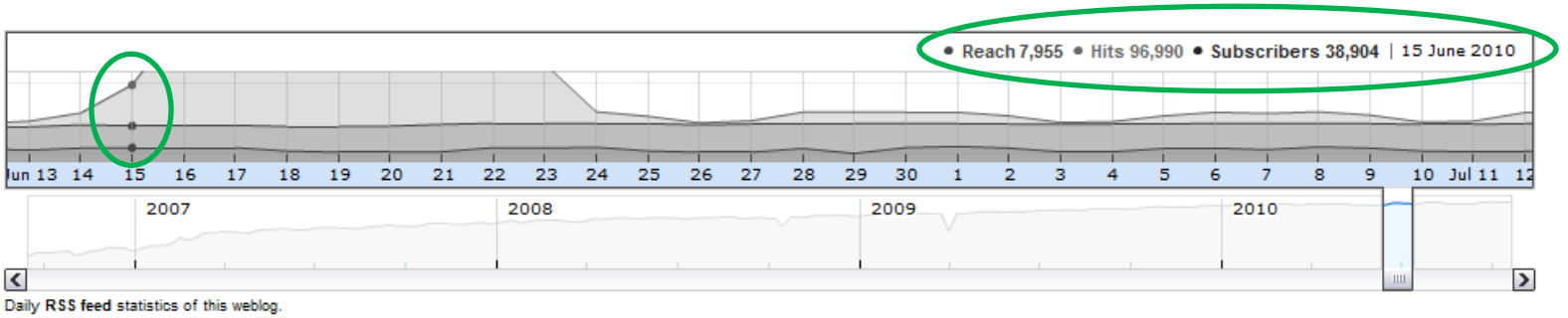


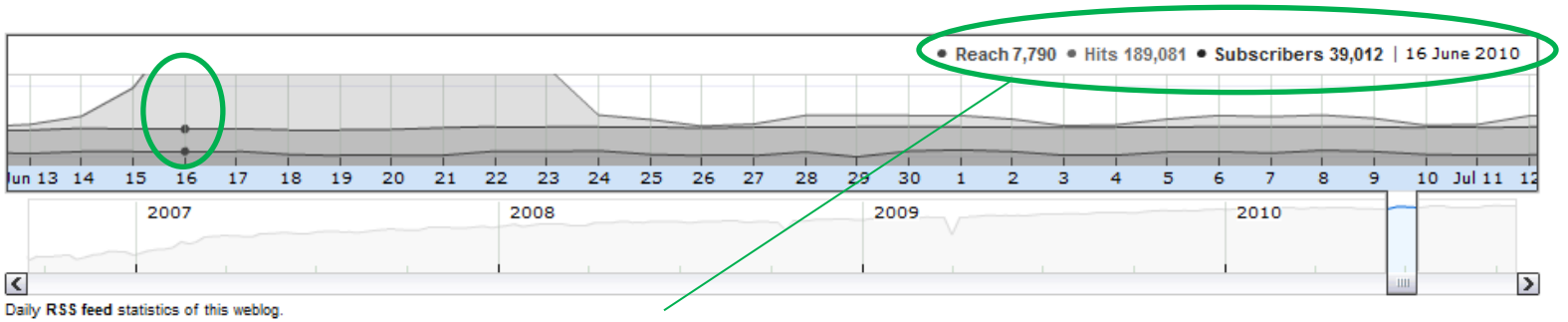
Daily **RSS feed** statistics of this weblog.



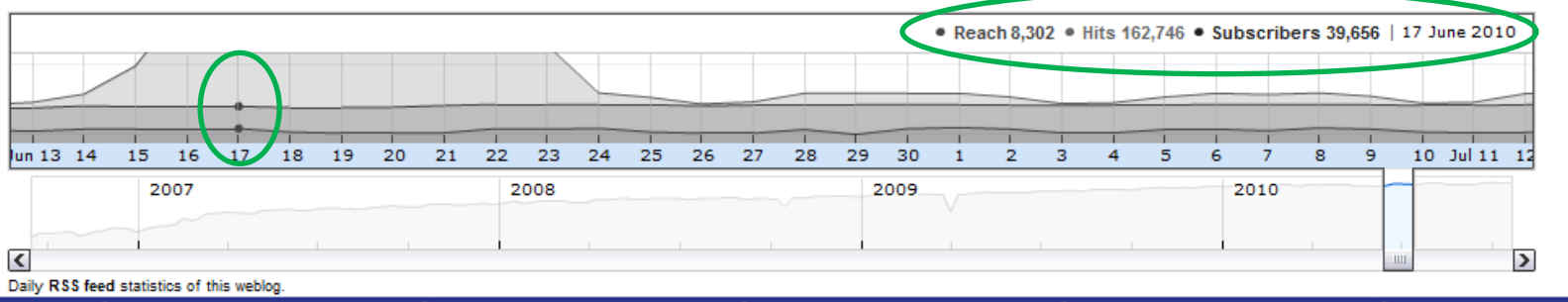
Daily **RSS feed** statistics of this weblog.

Details on Demand

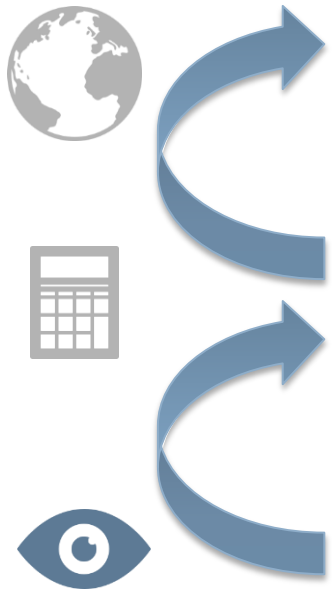




Wow, June 16th was the high point…



Perceptual building blocks of data visualization



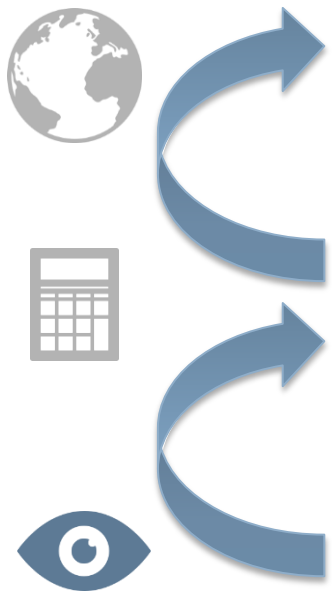
Context

Quantitative Reasoning

Visual Perception

Data Science work (modeling, sampling, etc.), BI (Calculations, KPIs, Business Rules)… the "what"

Perceptual building blocks of data visualization



Context

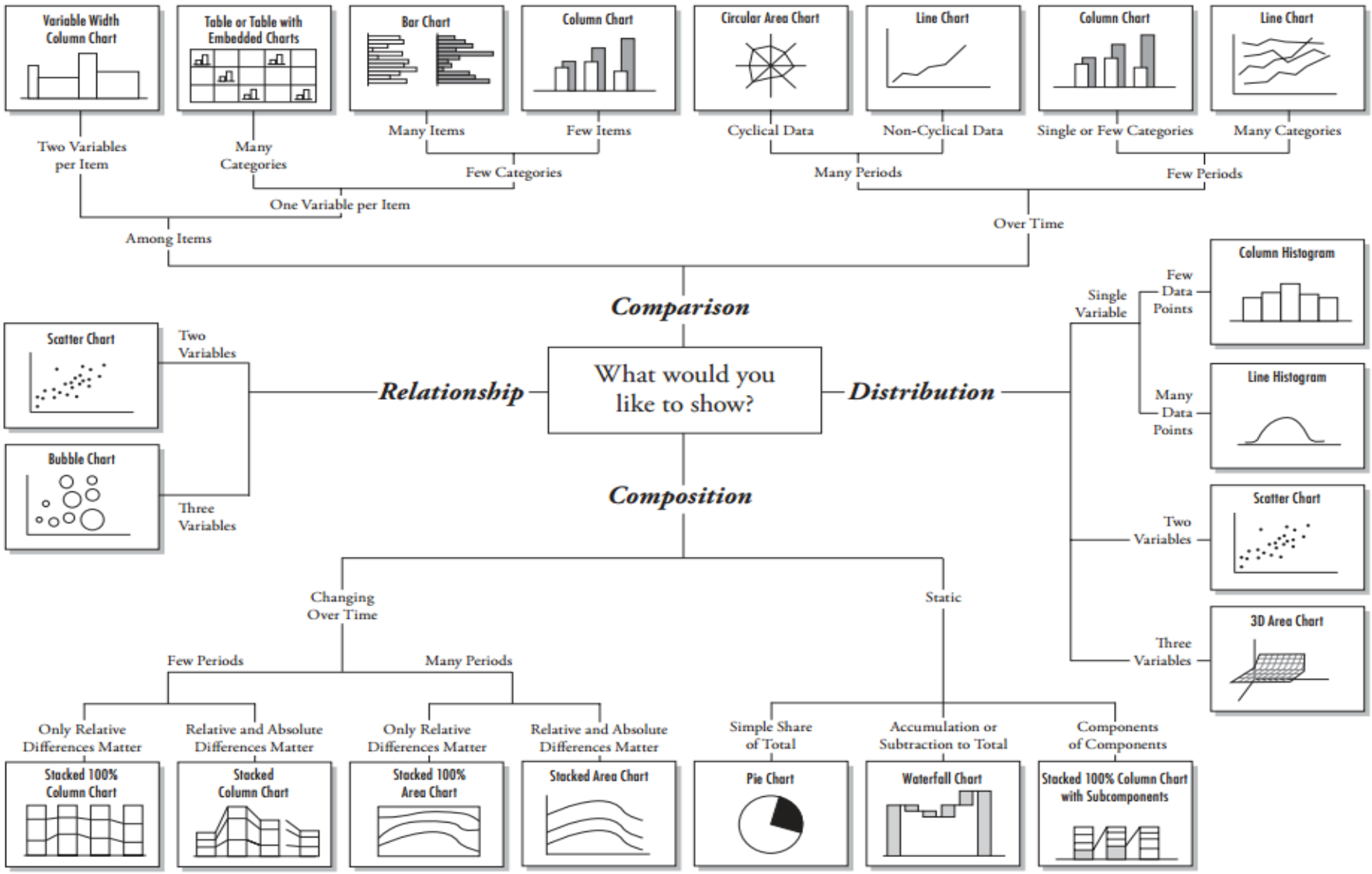
Quantitative Reasoning

Visual Perception

The science is here, albeit more hidden

World-class Data Visualization is done with an understanding of what we're building, for whom, and why

Chart Suggestions—A Thought-Starter



Creating vs Consuming Insight

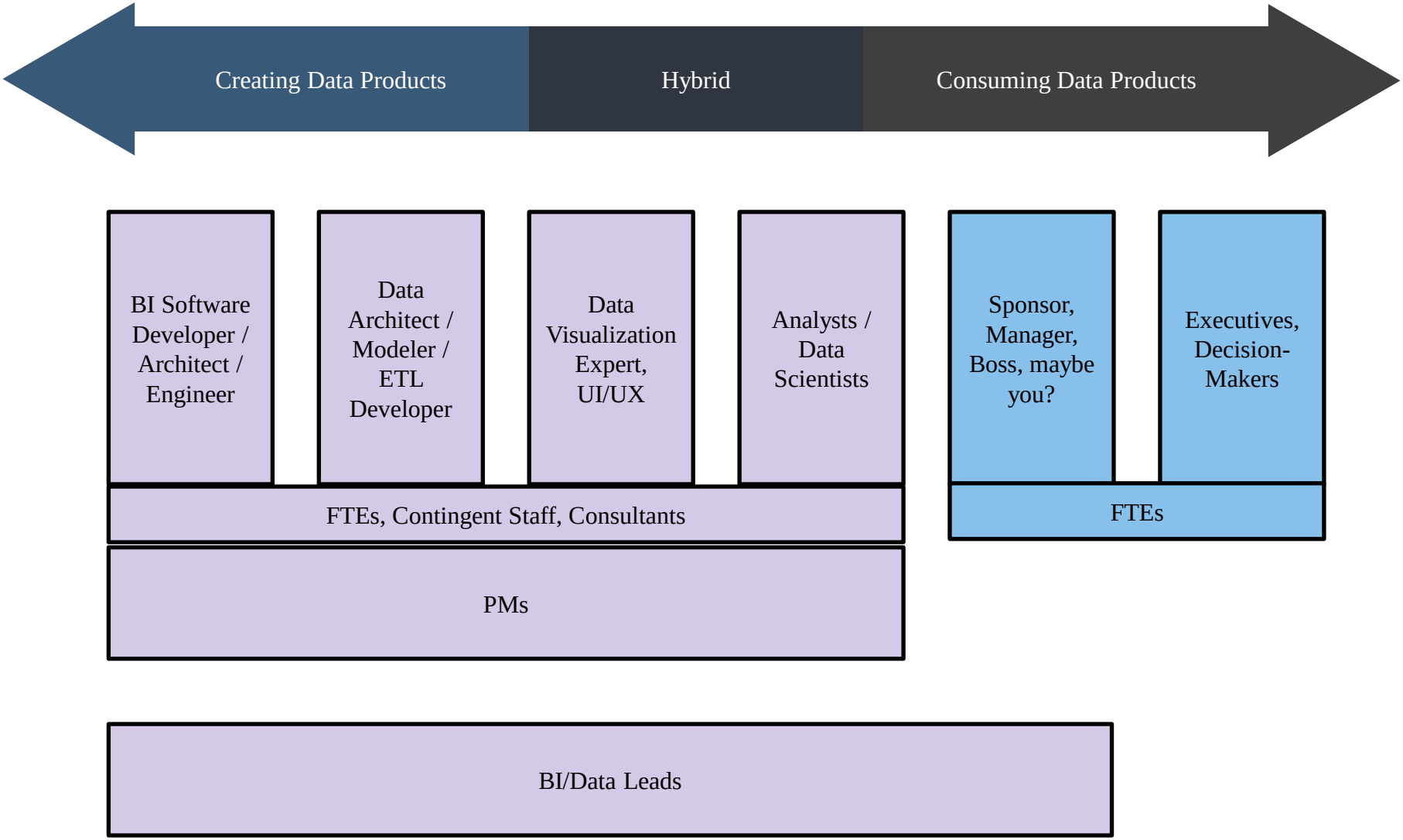
The approach to creating insight is nearly the opposite approach from consuming insight.

Creating vs Consuming Insight

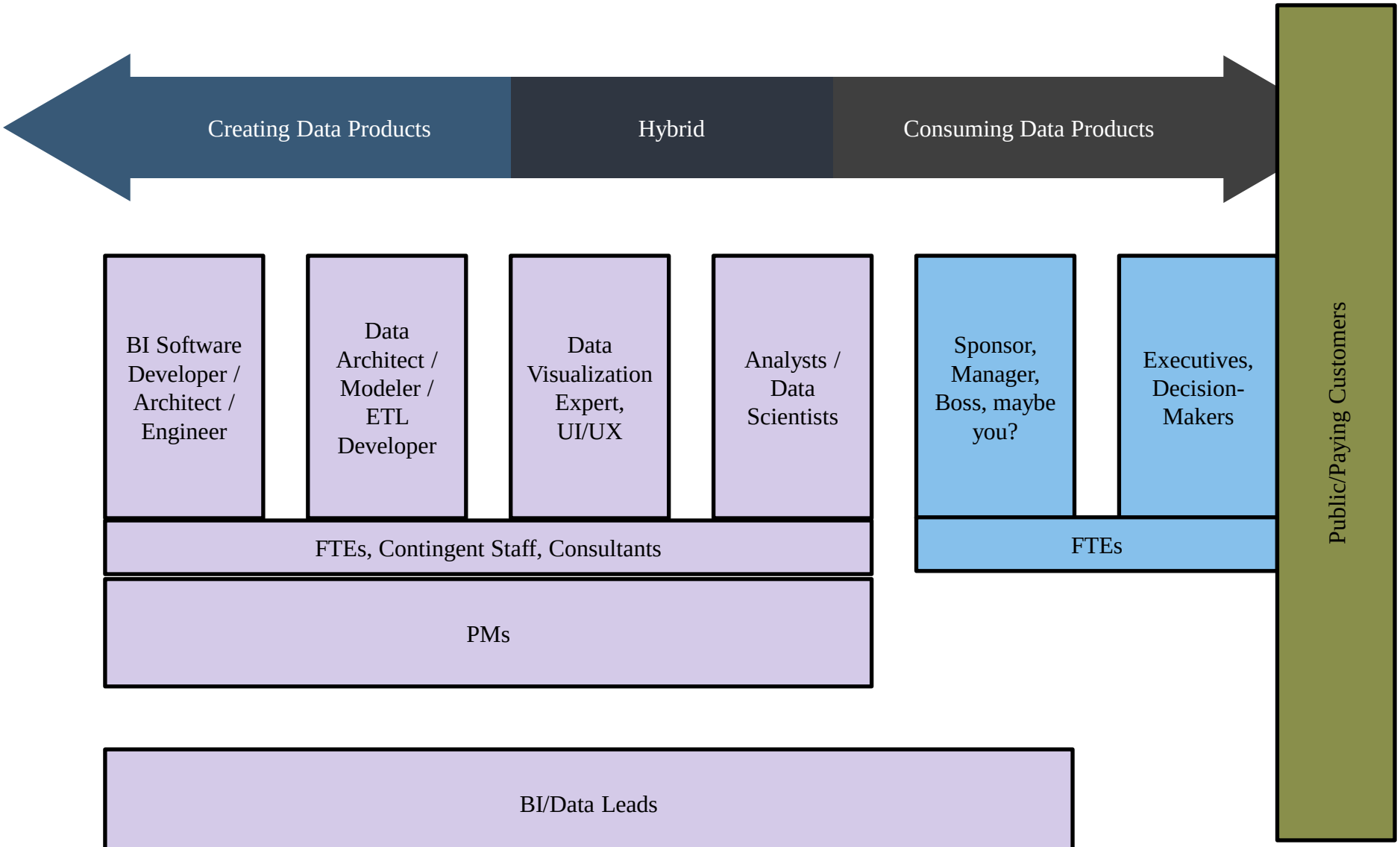
The approach to creating insight is nearly the opposite approach from consuming insight.

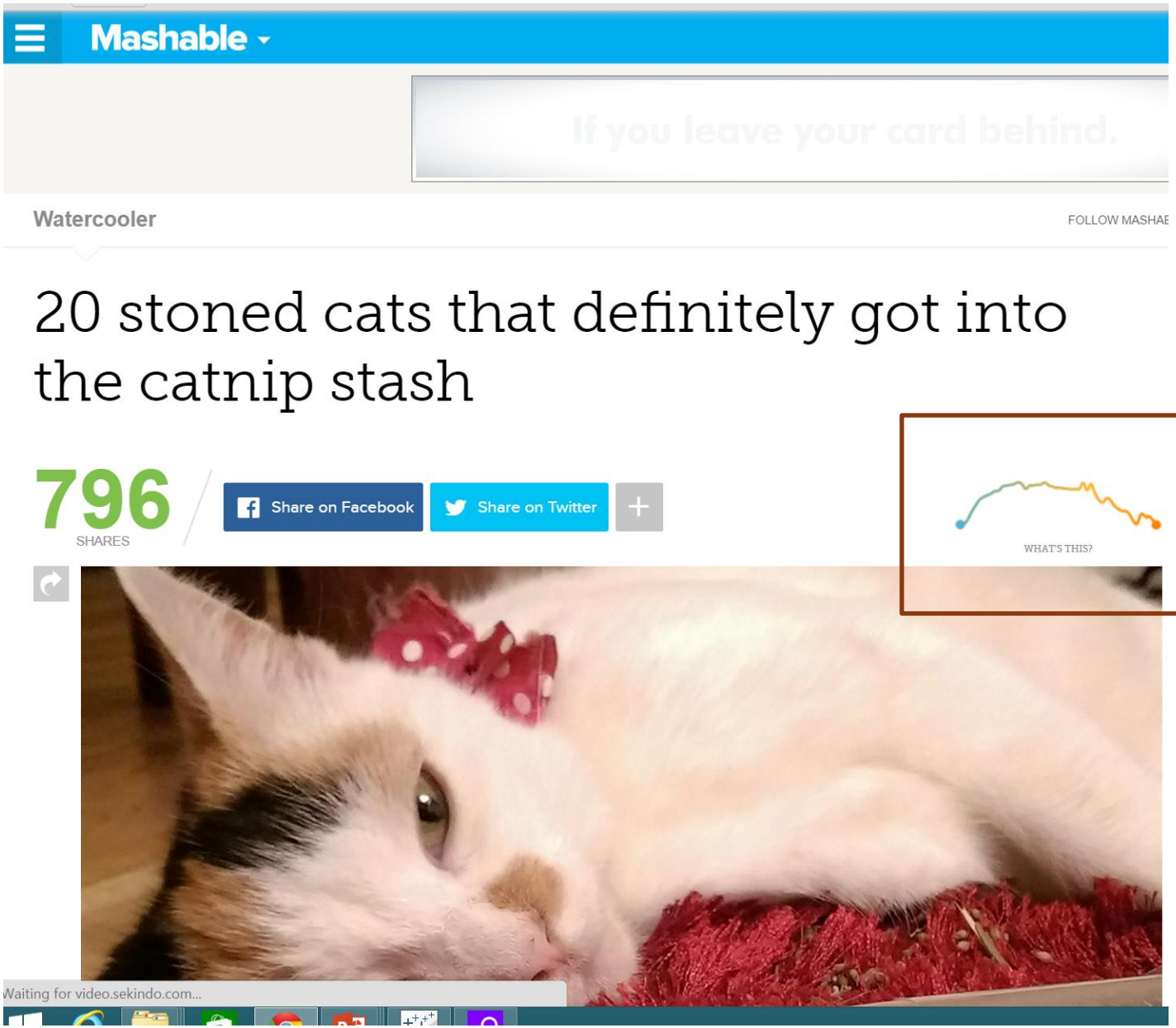
Creating Insight	Consuming Insight
1) Get data	4) Plan action
2) Build context	3) Interpret the observations
3) Find the meaning	2) Self-orient to the context
4) Make it easy to "see"	1) Internalize the framing

Full Data Professional Spectrum



Full Data Professional Spectrum



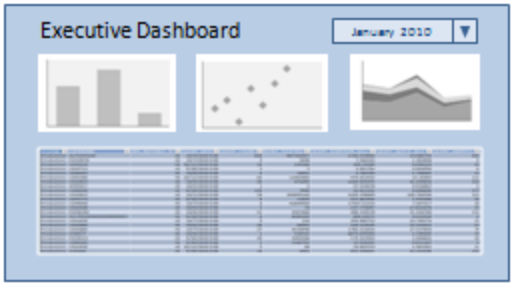


This is a data viz most definitely made by a data scientist… for customer / public consumption

Visual Artifacts

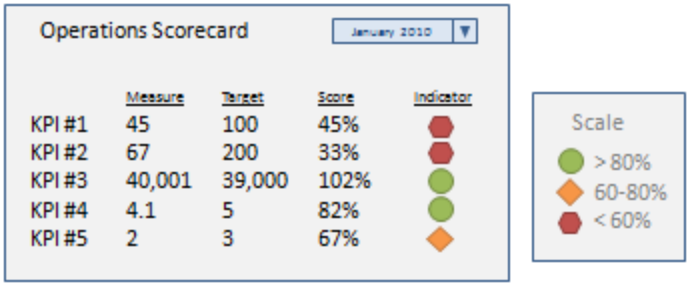
- Dashboards
- Scorecards
- Analytic Reports
- Analysis
- Transactional Reports

Visual Artifacts



- **Dashboards**
  - Role-specific, screen-based, interactive, multi-view
- Scorecards
- Analytic Reports
- Analysis
- Transactional Reports

Visual Artifacts

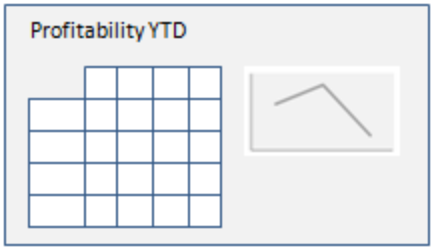


- Dashboards
  - Role-specific, screen-based, interactive, multi-view
- **Scorecards**
  - Focused performance improvement visualization
- Analytic Reports
- Analysis
- Transactional Reports

Visual Artifacts

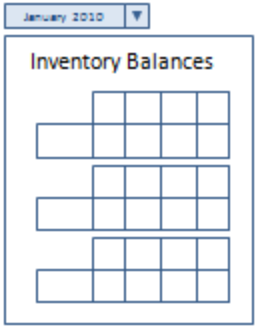
- Dashboards
  - Role-specific, screen-based, interactive, multi-view
- Scorecards
  - Focused performance improvement visualization
- **Analytic Reports**
  - Pre-designed analytic view based on analytic data
- Analysis
- Transactional Reports

Visual Artifacts



- Dashboards
  - Role-specific, screen-based, interactive, multi-view
- Scorecards
  - Focused performance improvement visualization
- Analytic Reports
  - Pre-designed analytic view based on analytic data
- **Analysis**
  - Free-form, ad-hoc, often one-time, analytic view
- Transactional Reports

Visual Artifacts



- Dashboards
  - Role-specific, screen-based, interactive, multi-view
- Scorecards
  - Focused performance improvement visualization
- Analytic Reports
  - Pre-designed analytic view based on analytic data
- Analysis
  - Free-form, ad-hoc, often one-time, analytic view
- **Transactional Reports**
  - Pre-designed informational view based on transactional data

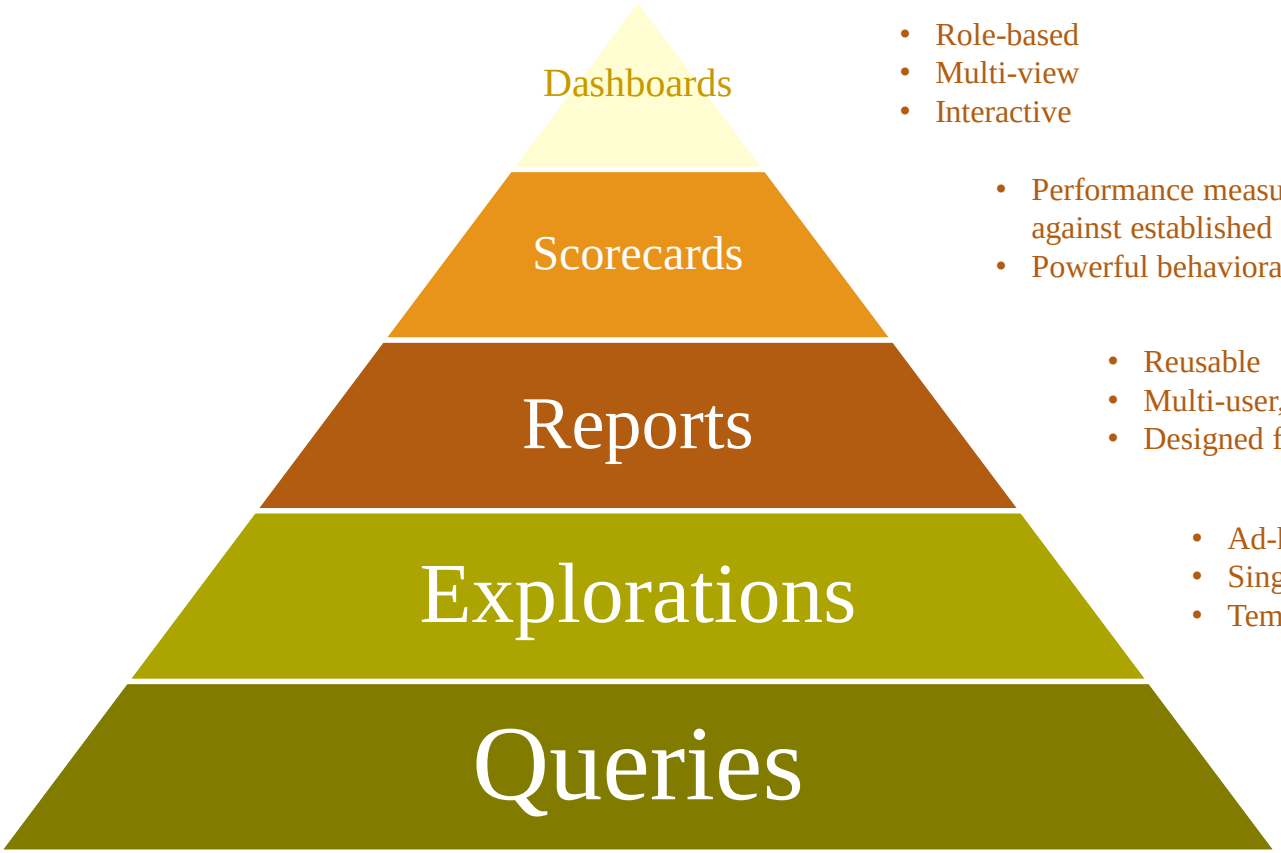
Visual Artifacts

- Dashboards
  - Role-specific, screen-based, interactive, multi-view
- Scorecards
  - Focused performance improvement visualization
- Analytic Reports
  - Pre-designed analytic view based on analytic data
- Analysis
  - Free-form, ad-hoc, often one-time, analytic view
- Transactional Reports
  - Pre-designed informational view based on transactional data

Visual Artifacts

- Dashboards
  - Role-specific, screen-based, interactive, multi-view
- Scorecards
  - Focused performance improvement visualization
- Analytic Reports
  - Pre-designed analytic view based on analytic data
- Analysis
  - Free-form, ad-hoc, often one-time, analytic view
- Transactional Reports
- Embedded data products
  - Example: cat nip velocity chart

User Interaction: Distinguishing Characteristics



- Role-based
- Multi-view
- Interactive

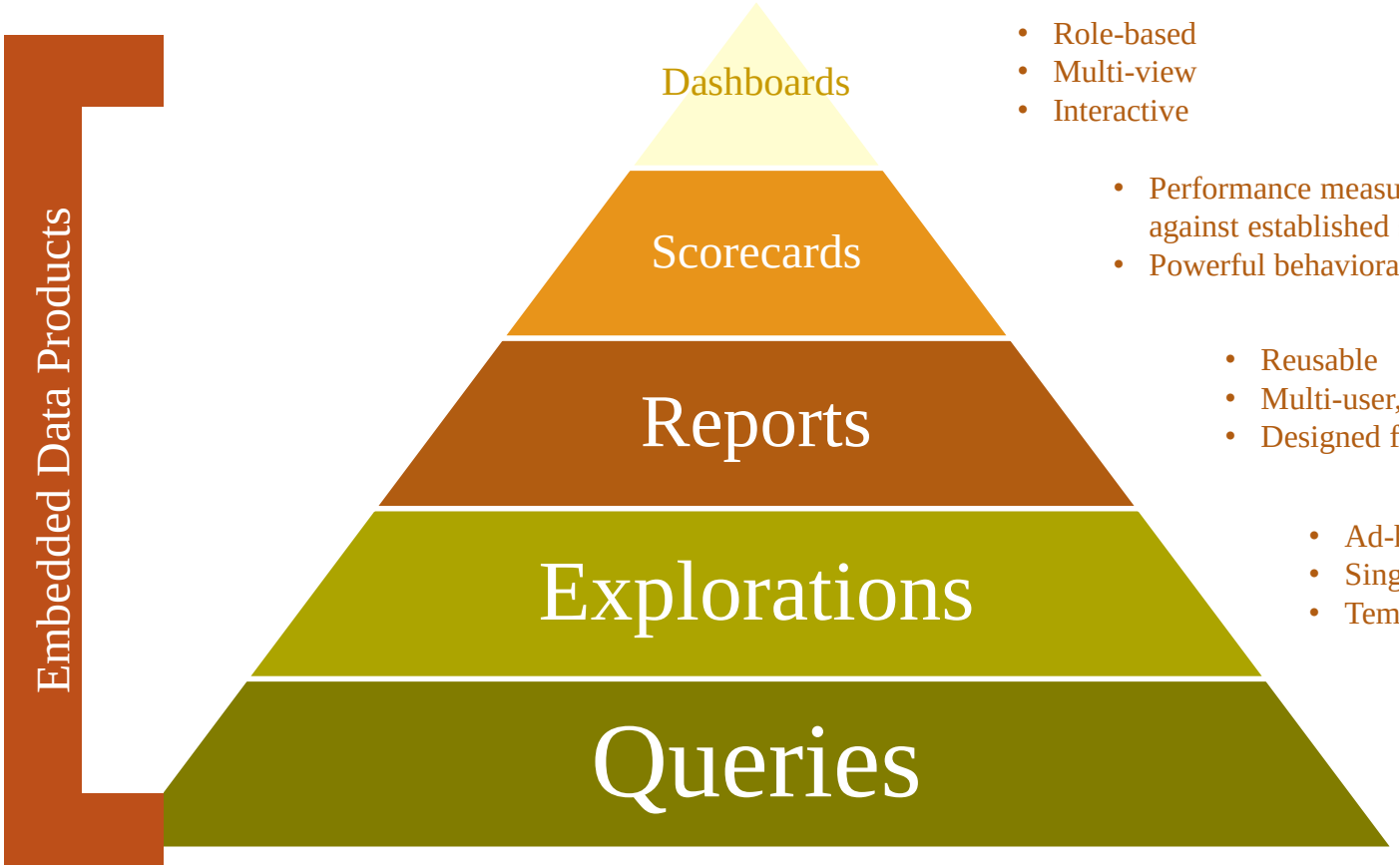
- Performance measurement against established standards
- Powerful behavioral change tool

- Reusable
- Multi-user, multi-source
- Designed format & guided discovery

- Ad-hoc , tool-dependent
- Single user, often single source
- Temporary

- Ad-hoc & form-free
- Single user, single source
- Temporary

User Interaction:
Embedded Data Products



- Role-based
- Multi-view
- Interactive

- Performance measurement against established standards
- Powerful behavioral change tool

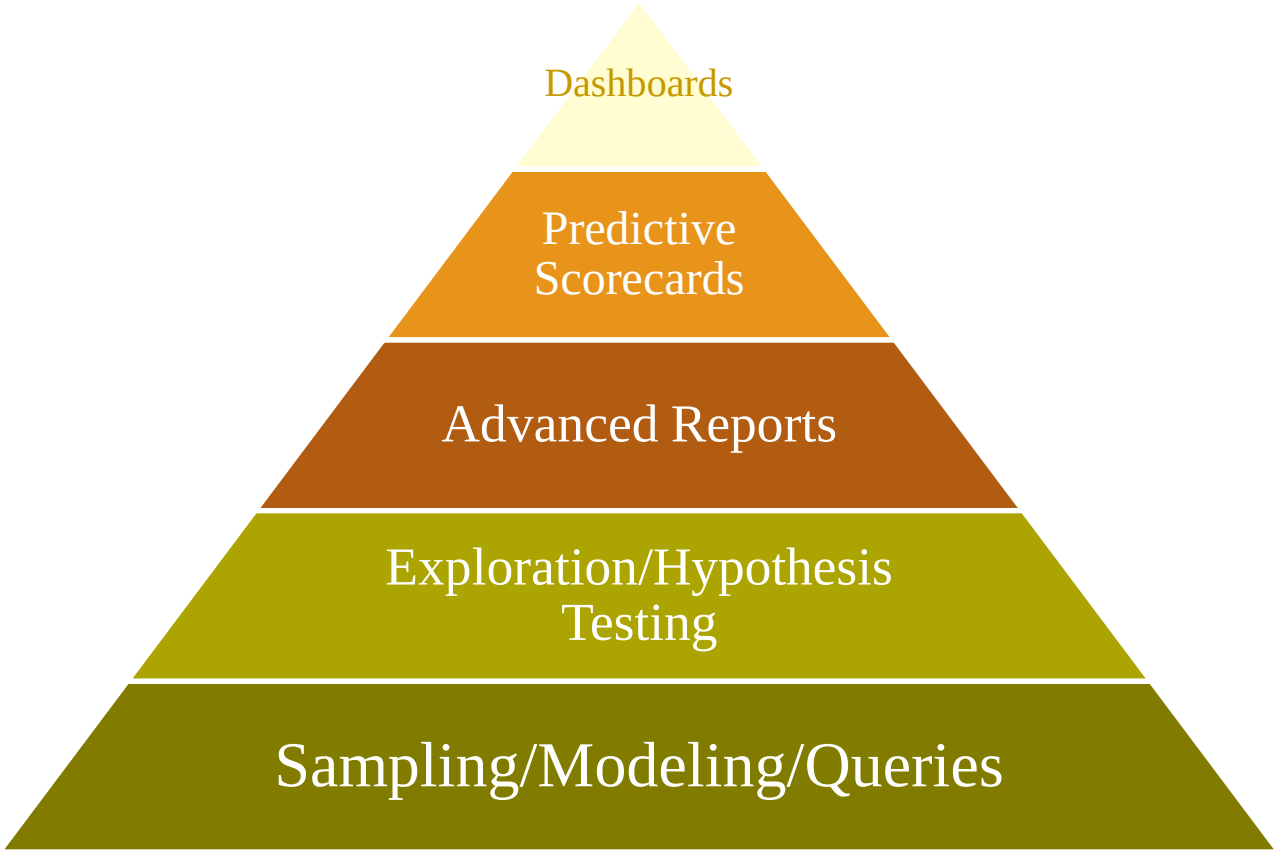
- Reusable
- Multi-user, multi-source
- Designed format & guided discovery

- Ad-hoc , tool-dependent
- Single user, often single source
- Temporary

- Ad-hoc & form-free
- Single user, single source
- Temporary

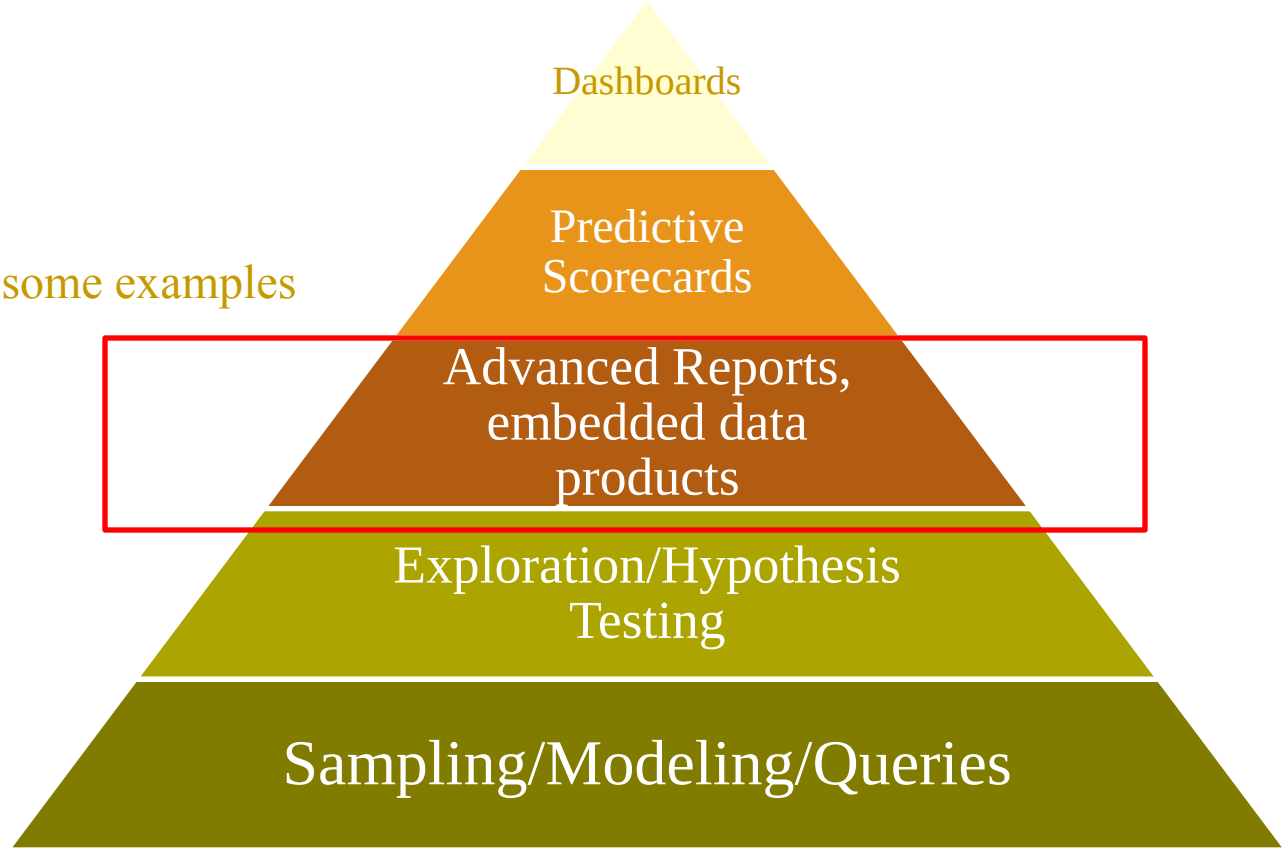
User Interaction: A Typical Portfolio for Data Professionals

User Interaction:
Visual Data Science?



User Interaction:
Visual Data Science?

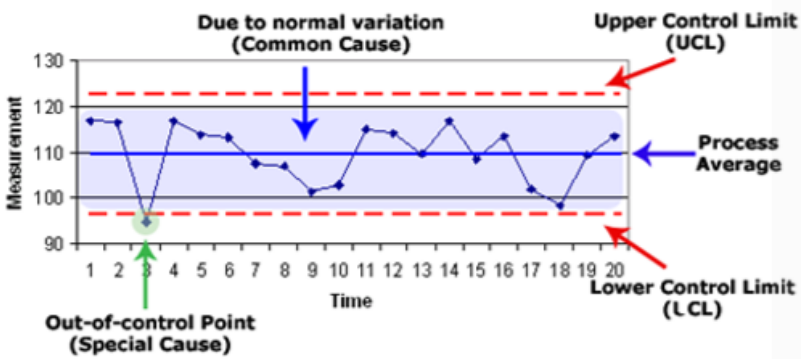
Let's see some examples

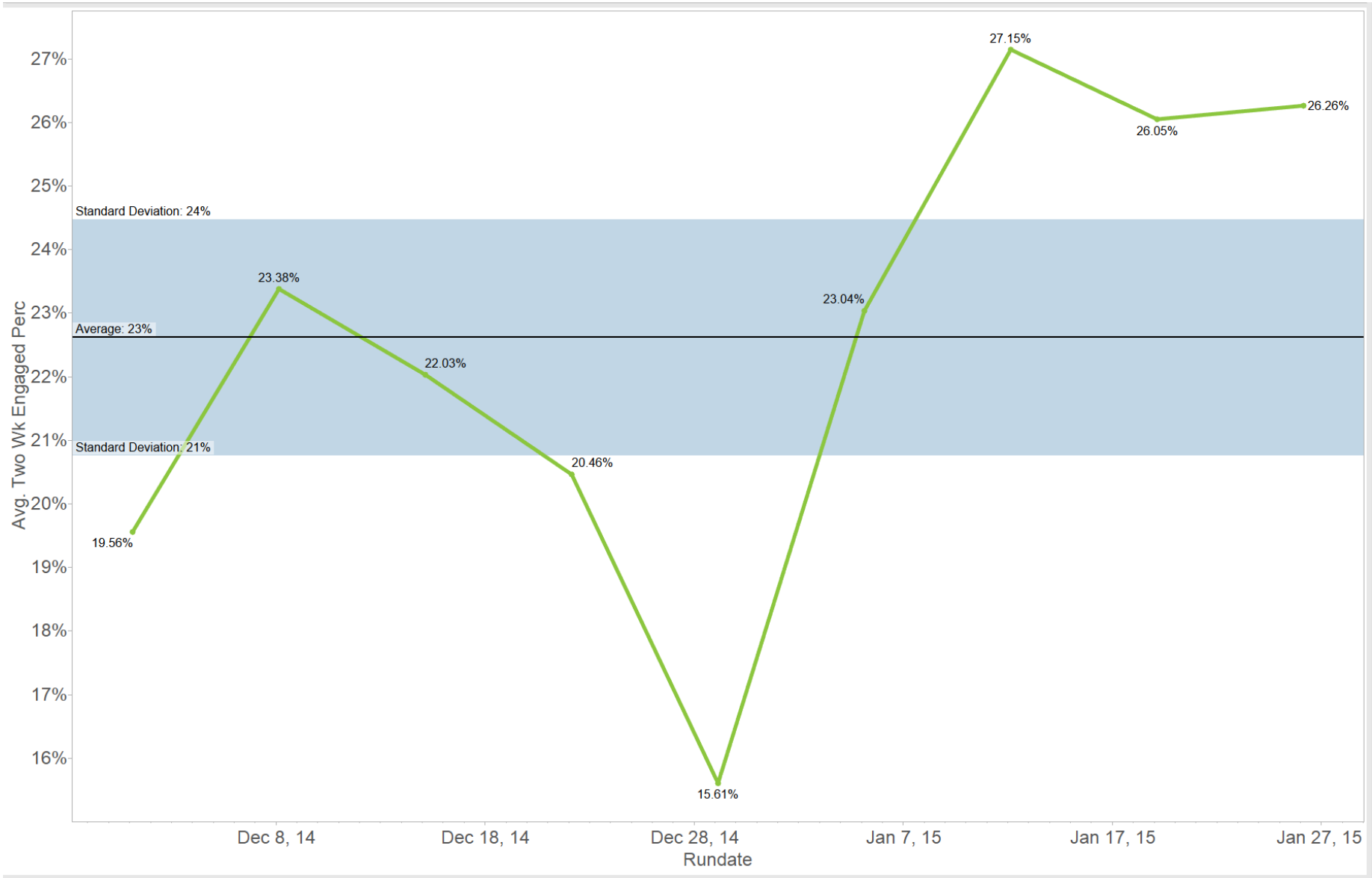


Advanced Charting often = Combination Charts

Control Chart

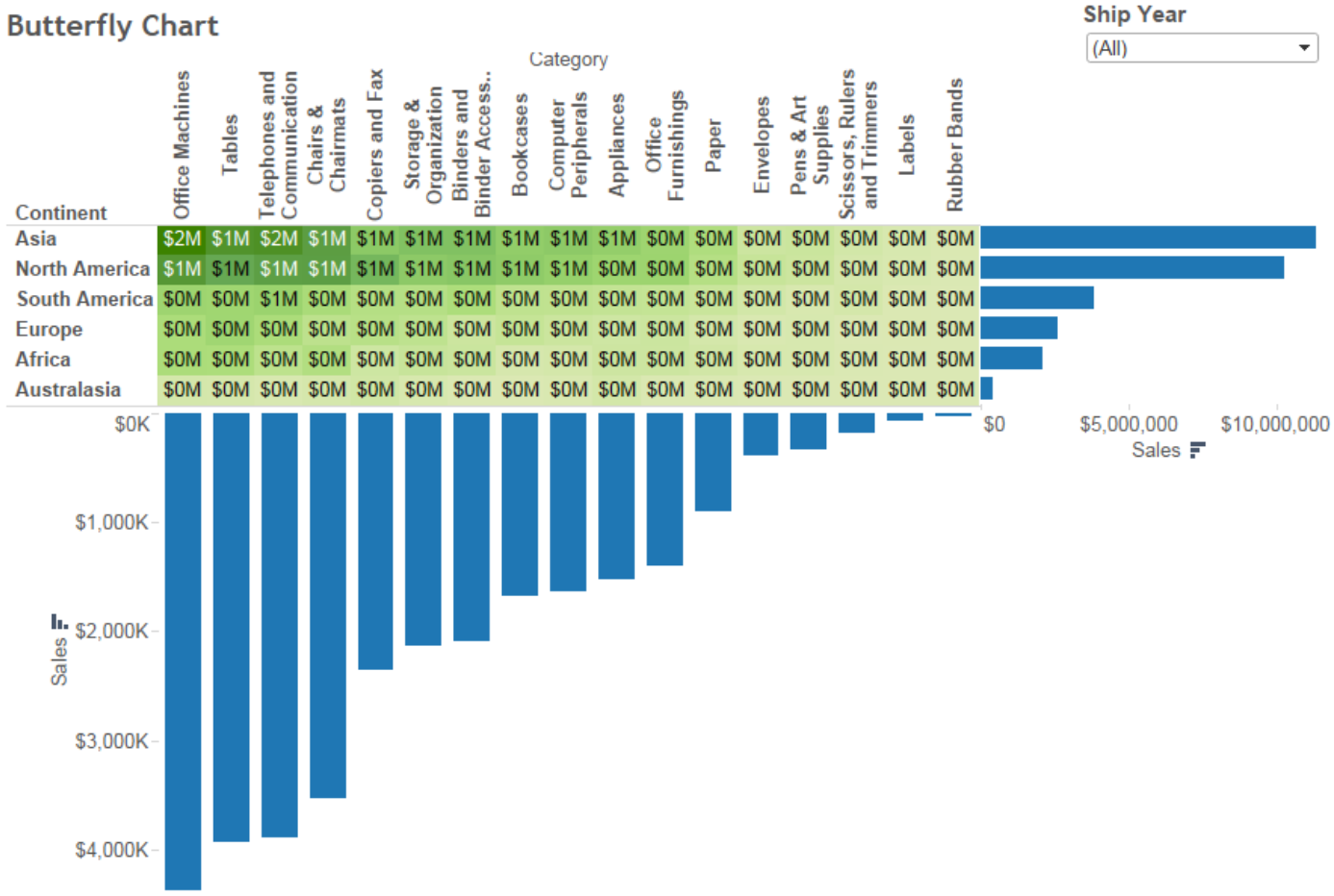
- A simple but powerful way to tell us whether variations we are seeing in a metric are due to normal statistical noise or due to some outside force
- (don't worry we won't be building the chart to the right)




Avg. Two Wk Engaged Perc
27%
26%
25%
24%
23%
22%
21%
20%
19%
18%
17%
16%
Standard Deviation: 24%
Average: 23%
Standard Deviation: 21%
19.56%
23.38%
22.03%
20.46%
15.61%
23.04%
27.15%
26.05%
26.26%
Dec 8, 14
Dec 18, 14
Dec 28, 14
Jan 7, 15
Jan 17, 15
Jan 27, 15
Rundate

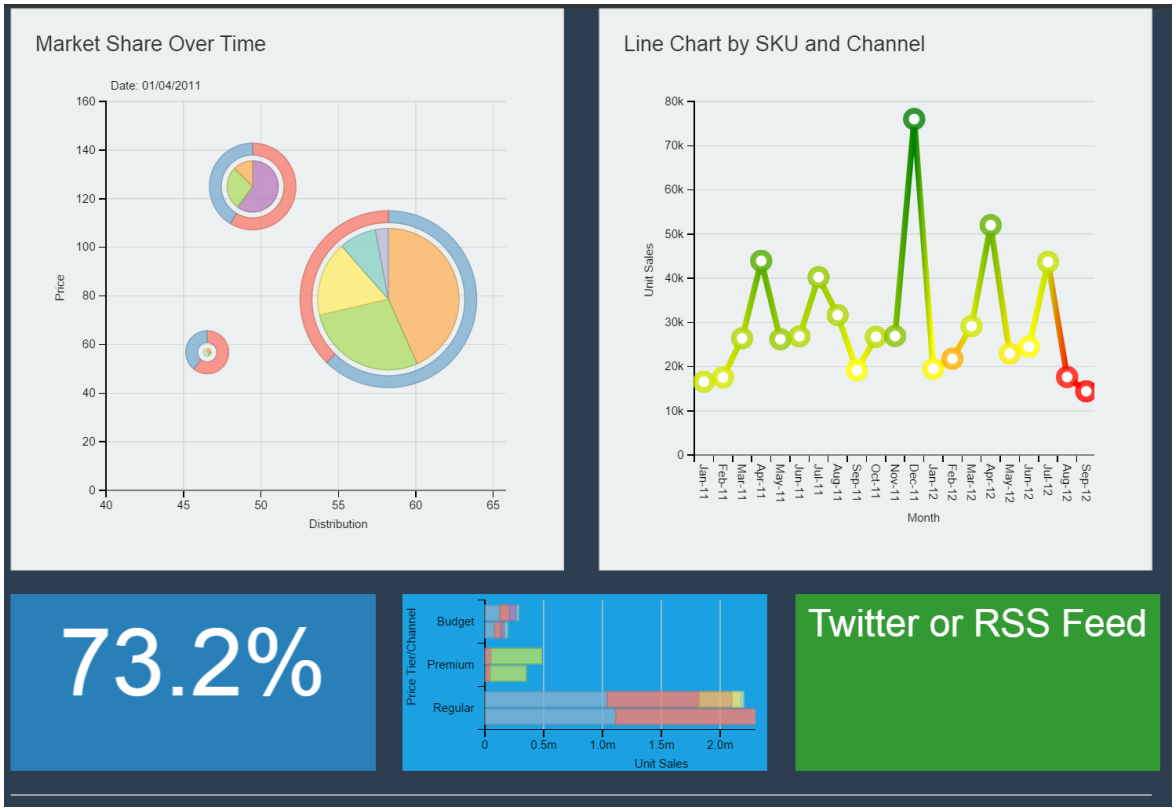
Advanced Charting often = Combination Charts

Butterfly Chart



Advanced Charting could also mean open-source (R, D3)

http://fizzyinc.co/matisia/



If you've liked what you've seen here, some follow-up resources:

1) Tableau online learning and tableau public—the hub for data viz tips and tricks and examples from one of the world's leading data viz software tools
2) Read!
    1) Stephen Few's Now You See It
    2) Edward Tufte—anything by him (The Visual Display of Quantitative Information)
    3) Colin Ware—anything by him
3) Practice: Master your basic charts—start with Excel, no one needs to know!
4) Go open-source! R, D3, etc.
5) Get involved—meetups around here abound

Q&A

Keep learning.
It's the Washington Way.

Thank You

PROFESSIONAL &
CONTINUING EDUCATION

UNIVERSITY *of* WASHINGTON