Data Visualization

Seeing the forest and the trees.

What We'll Do

- Basic tenants of Data Viz
- Methods
  - Temporal, quantitative, relational
- Types of Data
- Why visualization works
  - Laws – similarity and proximity

What We'll Do

- Basic tenants of Data Viz
- Methods
  - Temporal, quantitative, relational
- Types of Data
- Why visualization works
  - Laws – similarity and proximity

Play

Discuss

This is a hands on, interactive class.

And you thought you were safe.

What do we mean by Data Visualization?

In the simplest terms, data viz is a process/discipline of turning raw data into visual depictions of that data by means of visual "cues" such as color, size, and location.

This will become clear in a few slides.

"Our perception of patterns in the objects that we see is fundamental to the sense-making process. ***Pattern perception supports abstract thinking.****"*

Stephen Few of Perceptual Edge from "Visual Pattern Recognition"

Edward Tufte, in his book *"Visual Explanations"*, asks…

> *"What are the standards for evaluating visual evidence, especially for making decisions and reaching conclusions?"*

> *"Visualization provides a powerful means of making sense of data. By mapping data attributes to visual properties such as position, size, shape, and color, visualization designers leverage perceptual skills to help users discern and interpret patterns within data."*

Jeffrey Heer – Stanford
Ben Schneiderman – Univ of Maryland
"Interactive Dynamics for Visual Analysis"

"…many claim that data visualization often answers questions they didn't know they had." Accenture

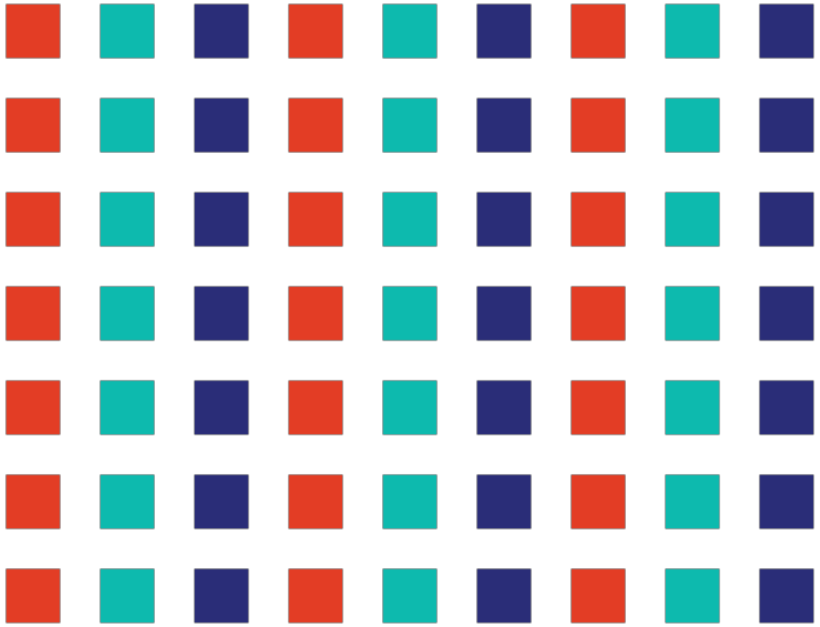
Raise your hand if your eyes went straight to

the red text

on the bottom of the previous page.

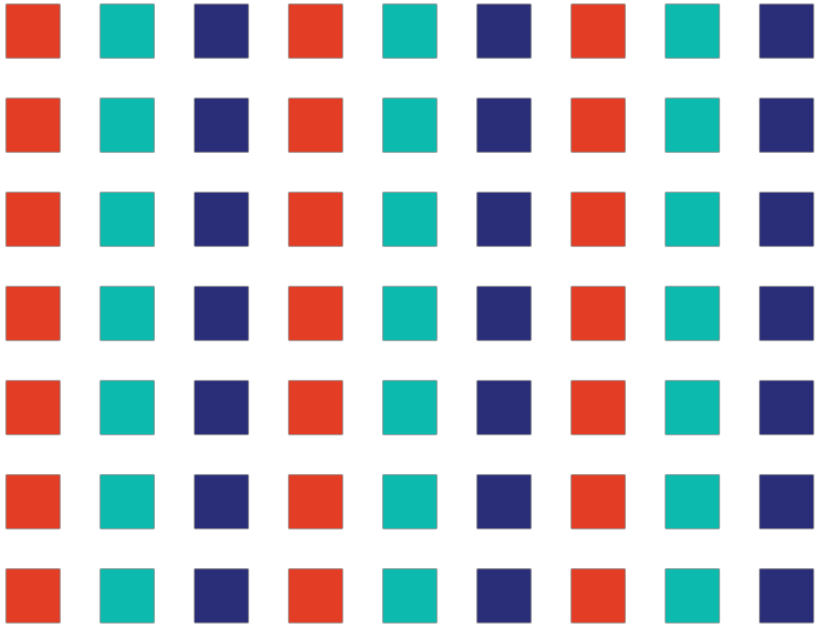
Let’s Dig In

Pattern Recognition

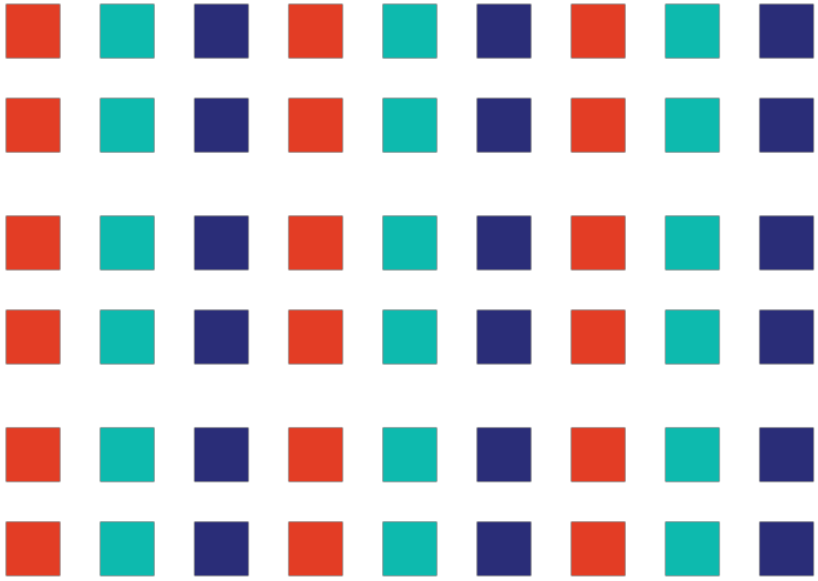


Our brains are hardwired for this.
We recognize through visuals.

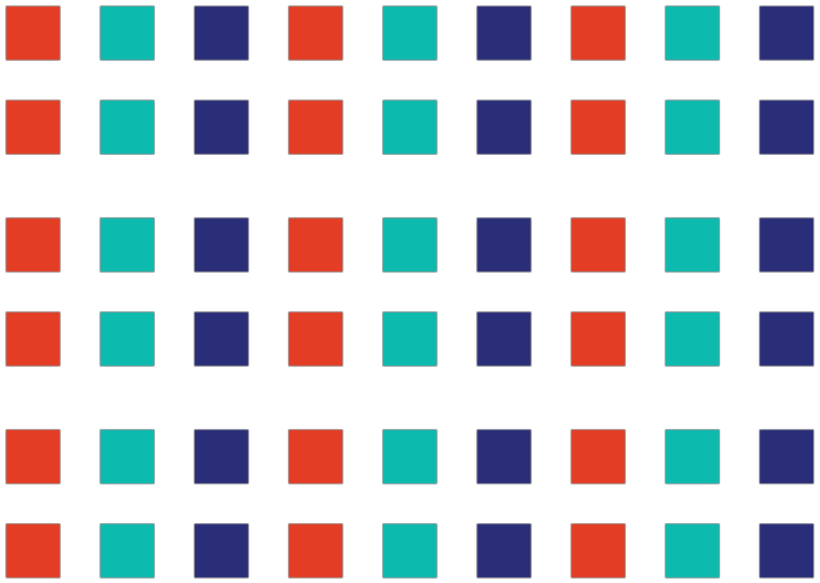
How many patterns can we find?



And now, how does our perception of the patterns change?



And now, how does our perception of the patterns change?

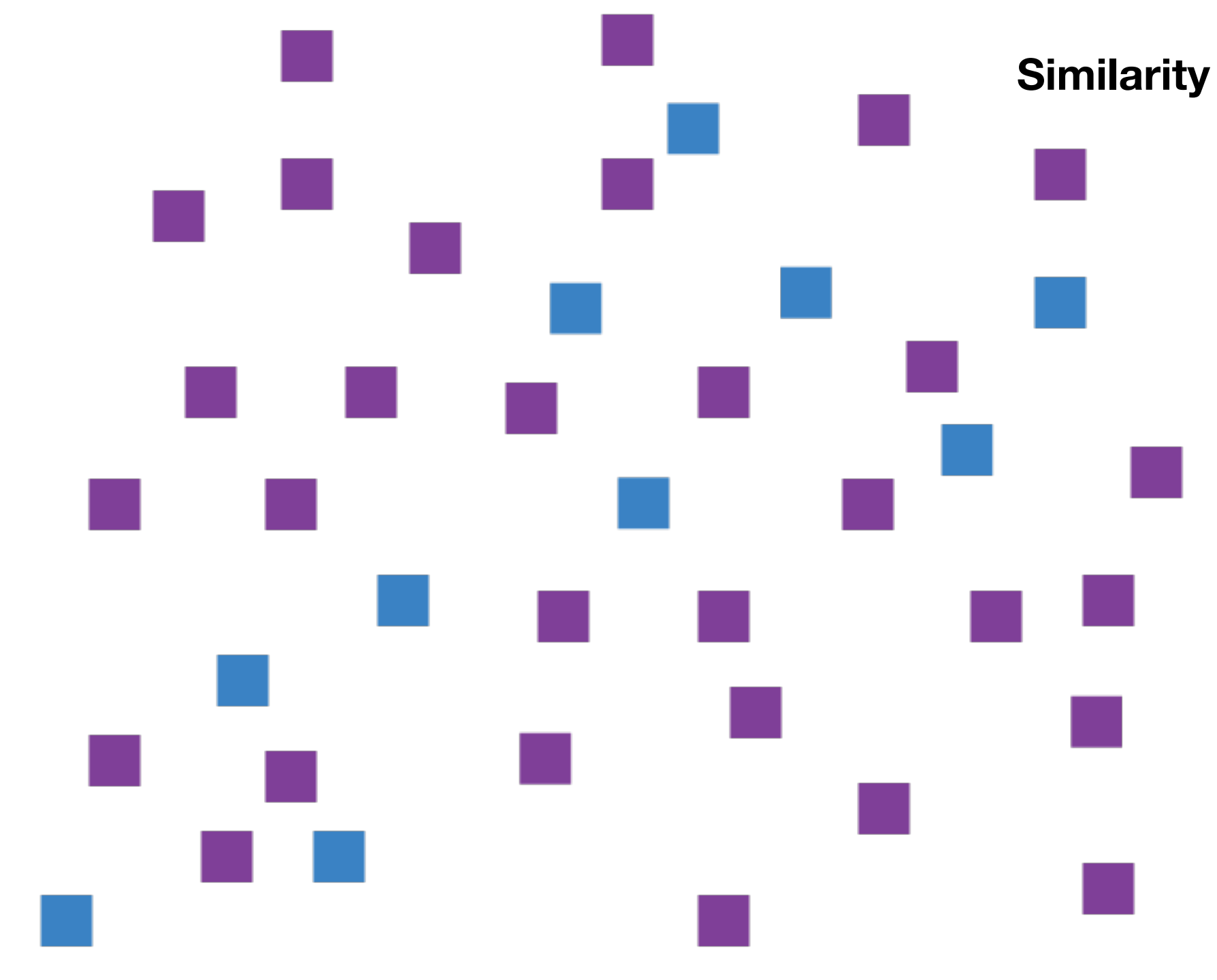


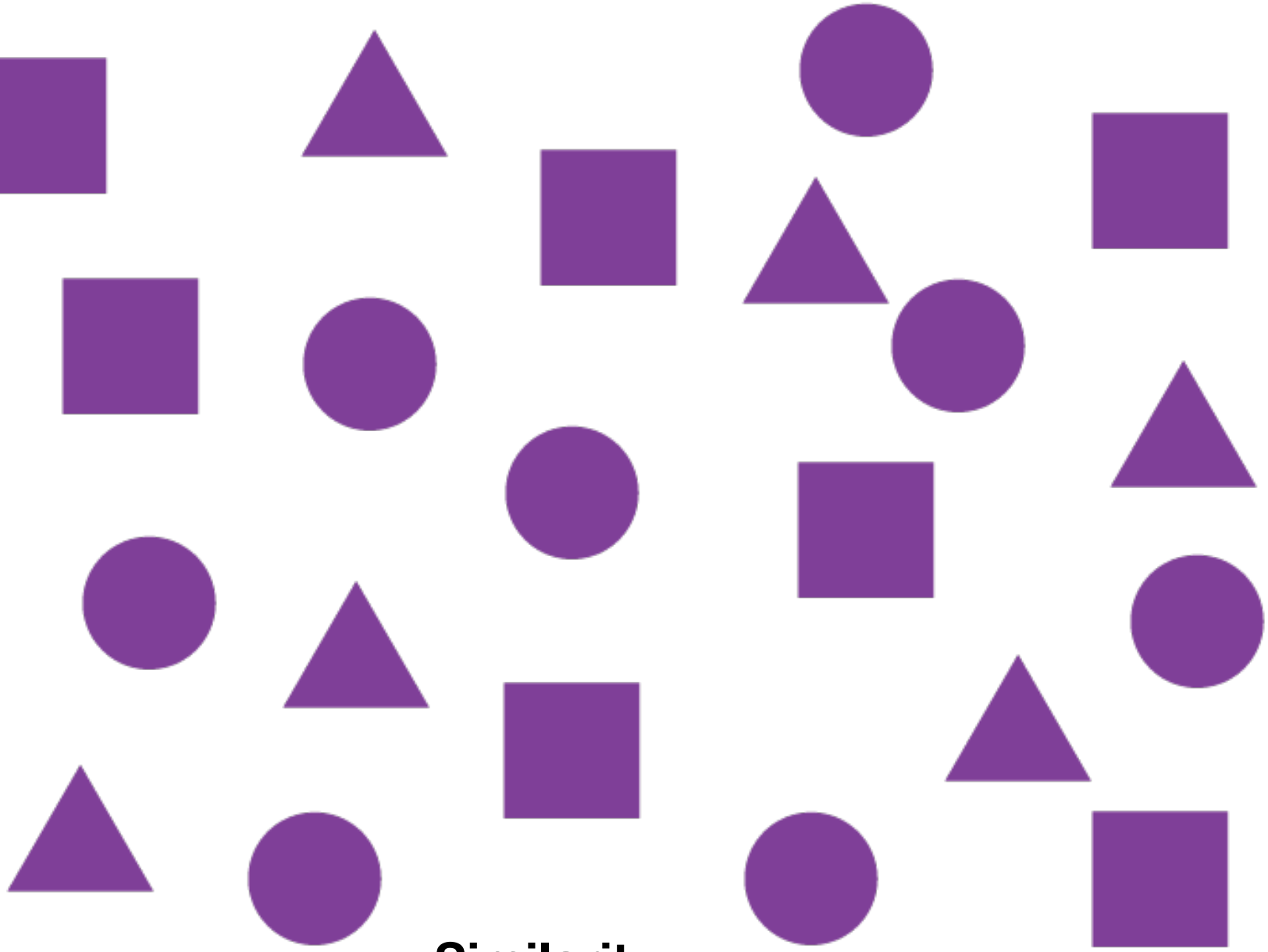
Slight changes can give us different ideas.

Proximity

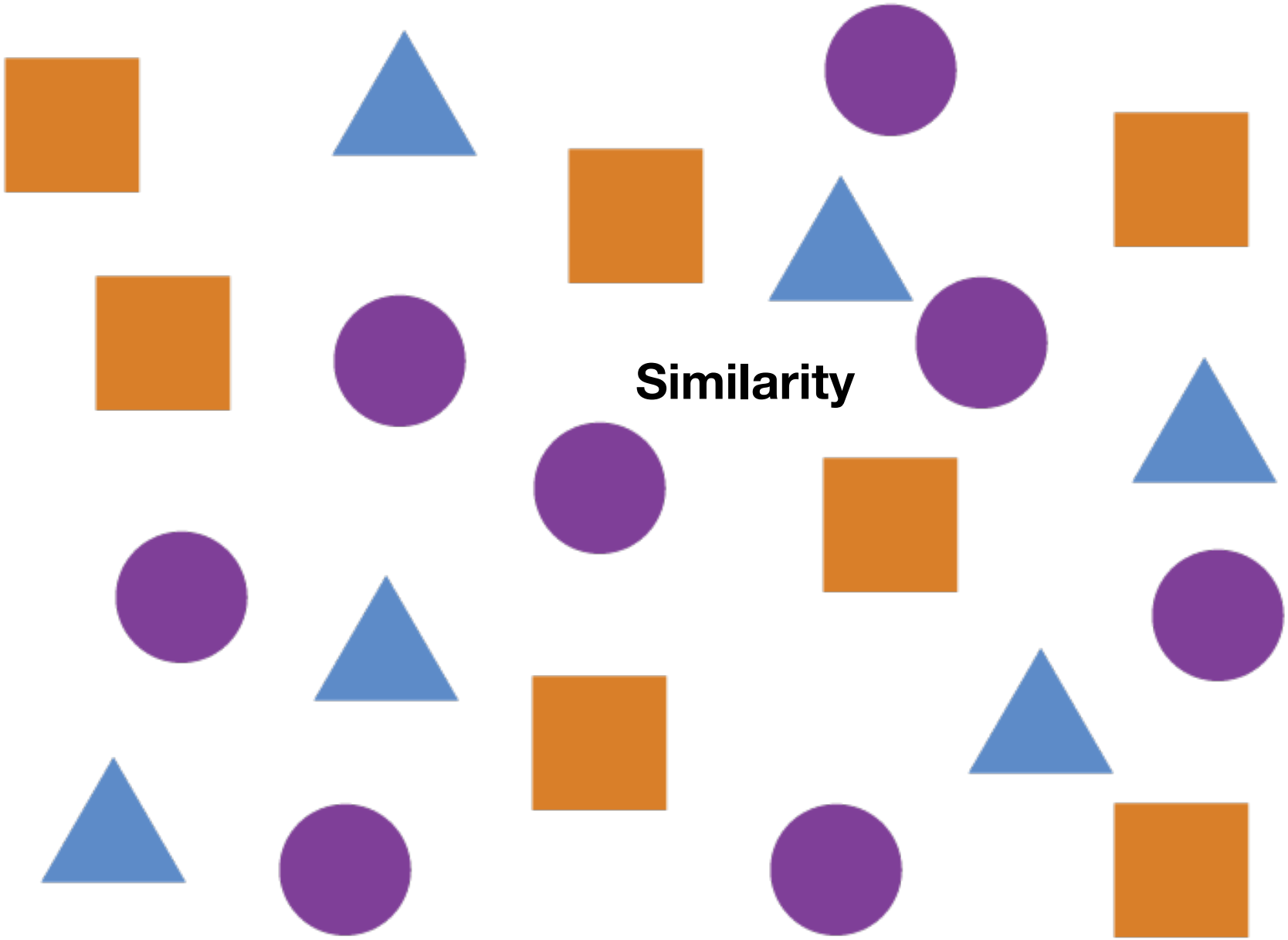
Proximity

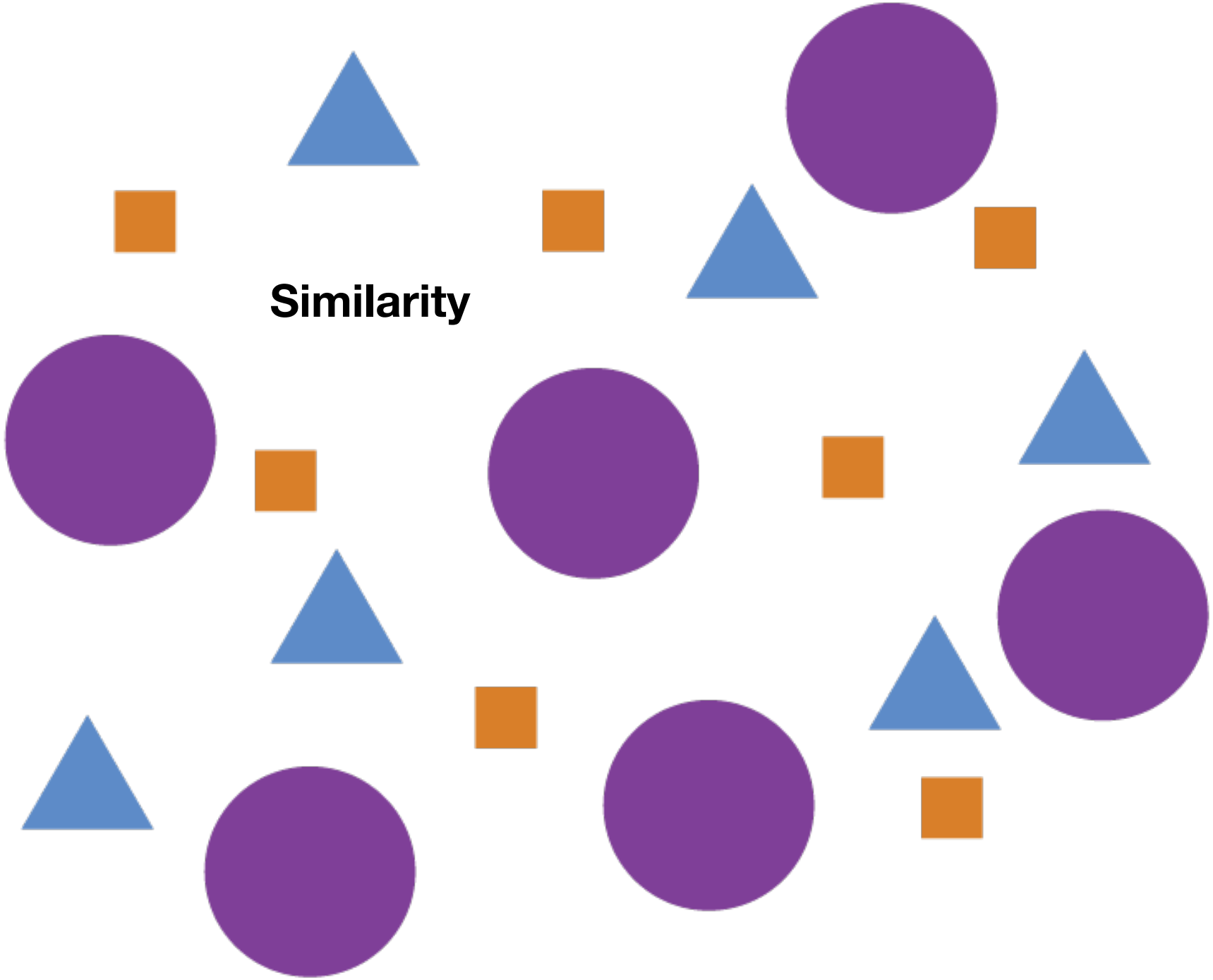
Proximity


Similarity



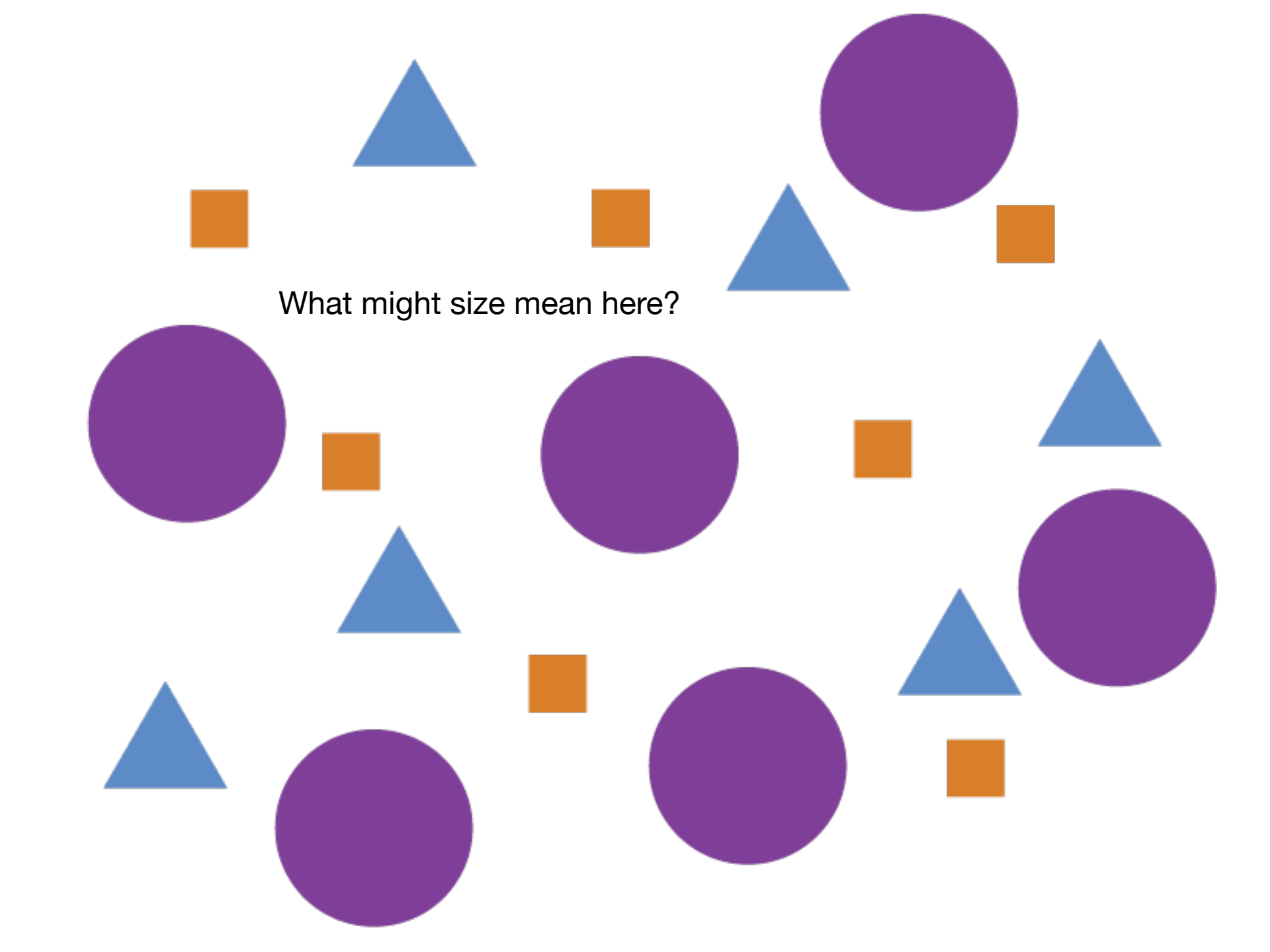
Similarity


Similarity

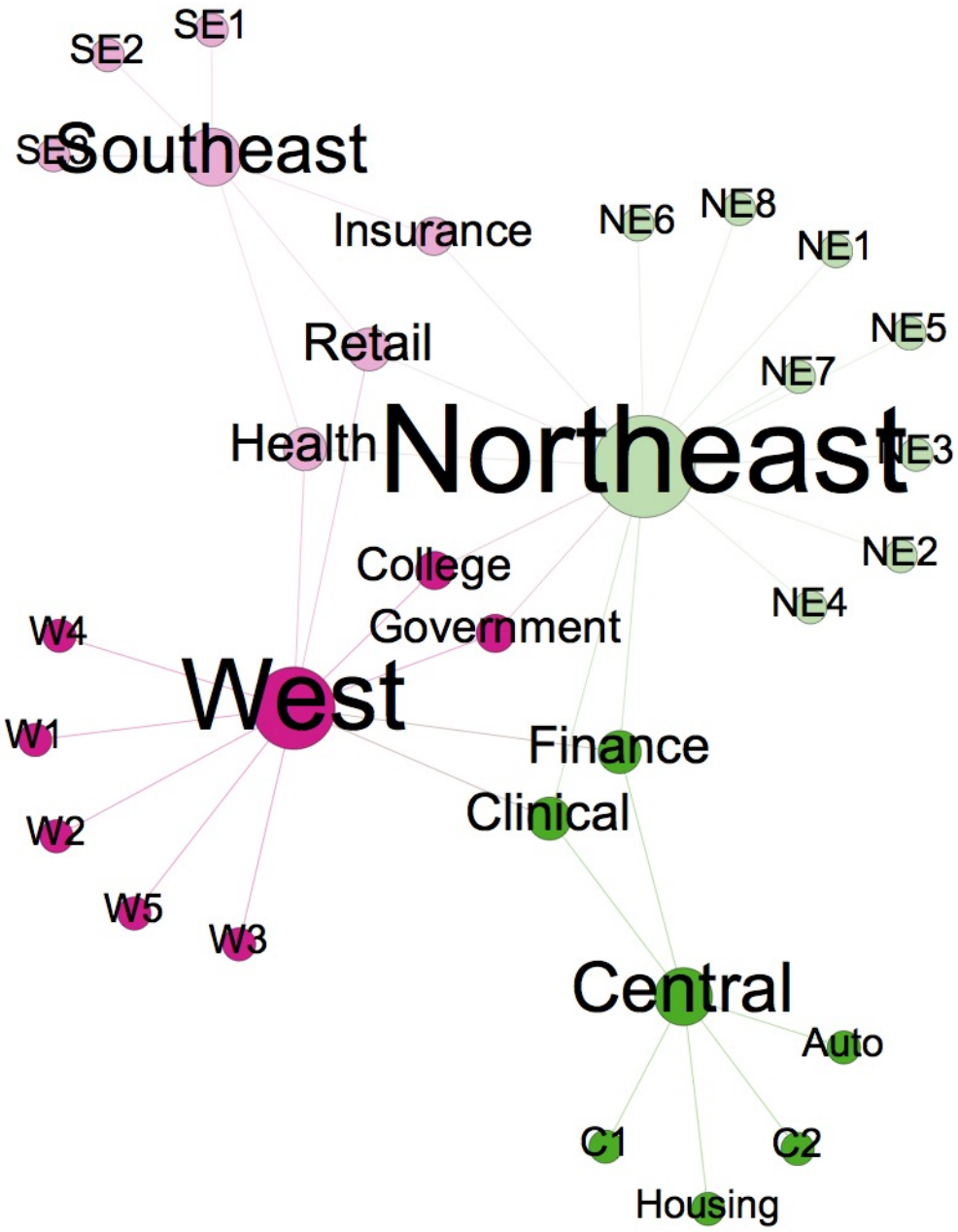

Similarity

What does size mean?

What does size mean?


What might size mean here?

Network Diagram



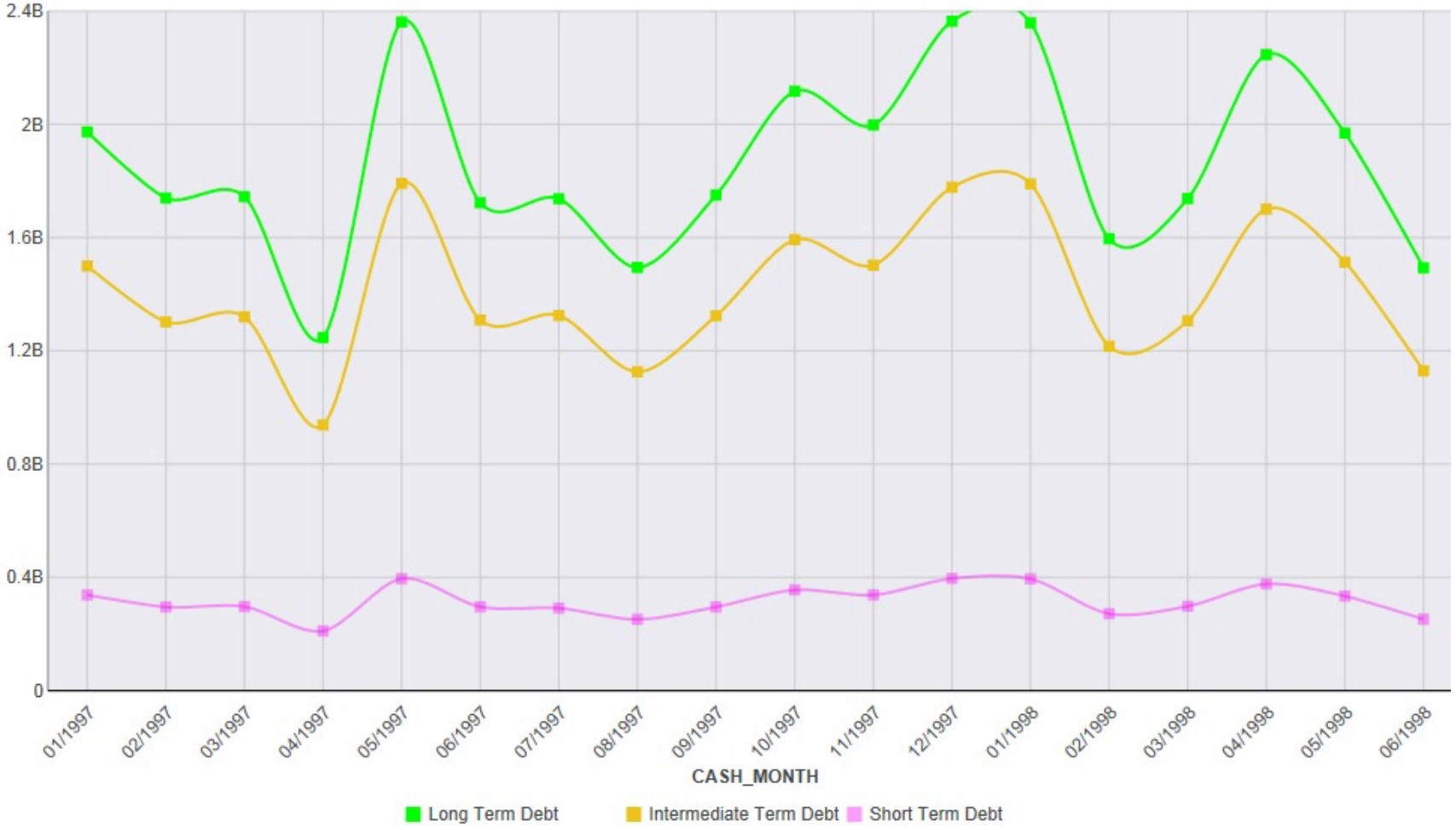
If we can understand how perception works, our knowledge can be translated into rules for displaying information.

Colin Ware, *Information Visualization: Perception for Design, Second Edition* (San Francisco, Morgan Kauffman, 2004), xxi

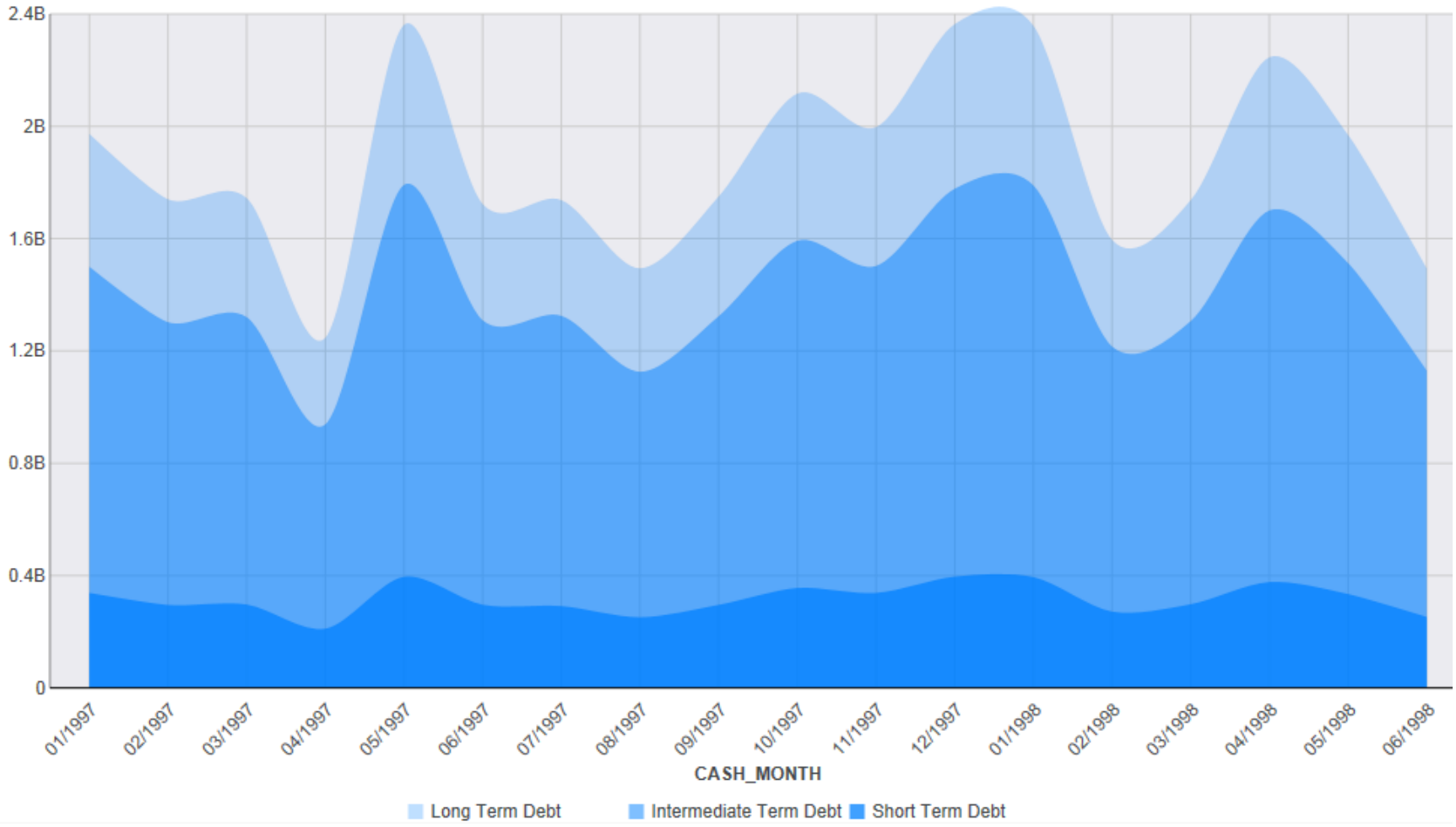
Methods

Matching Data Types to Visualization

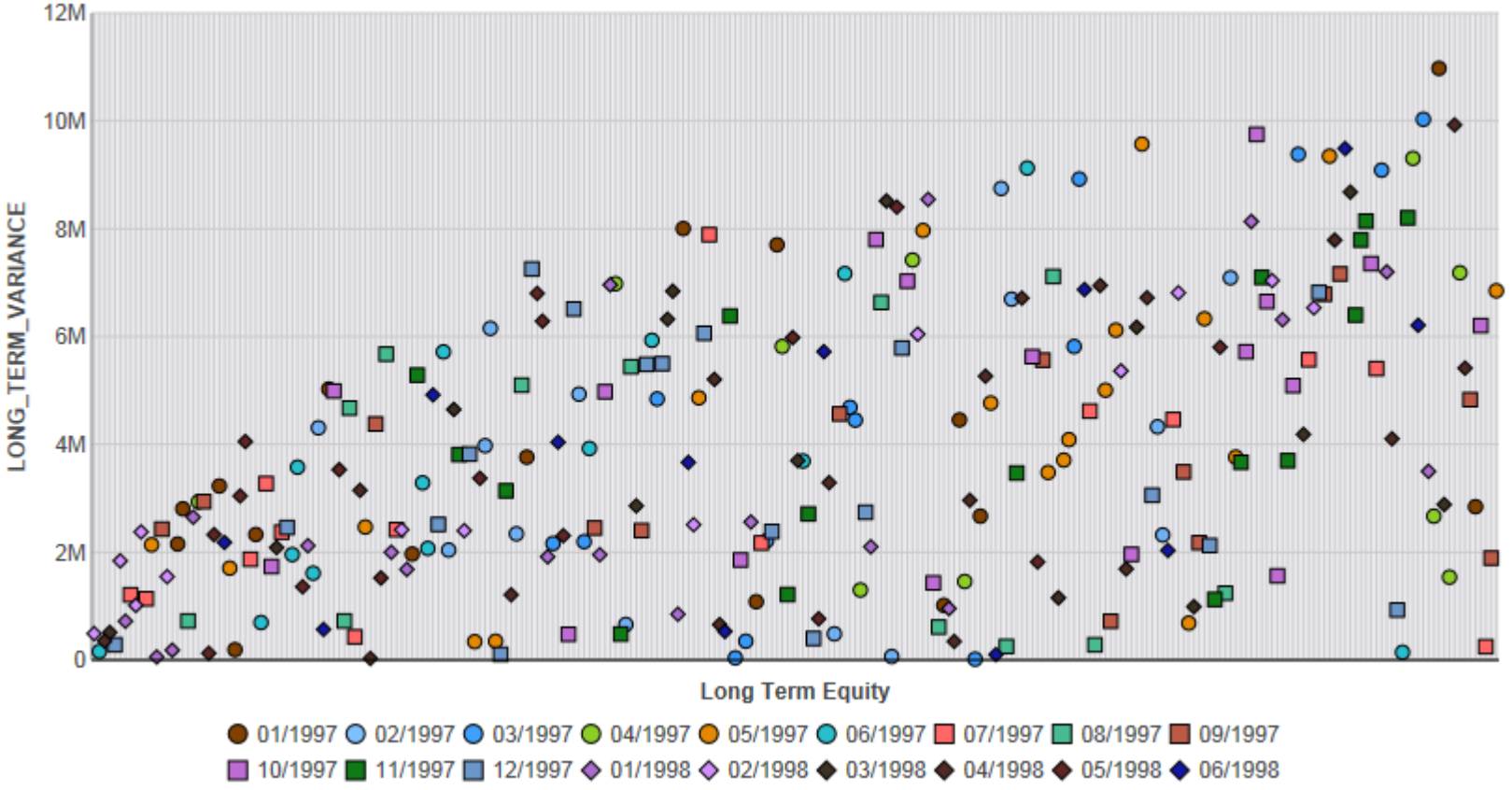
Temporal – time based data
Line chart



Temporal – time based data
Filled area chart

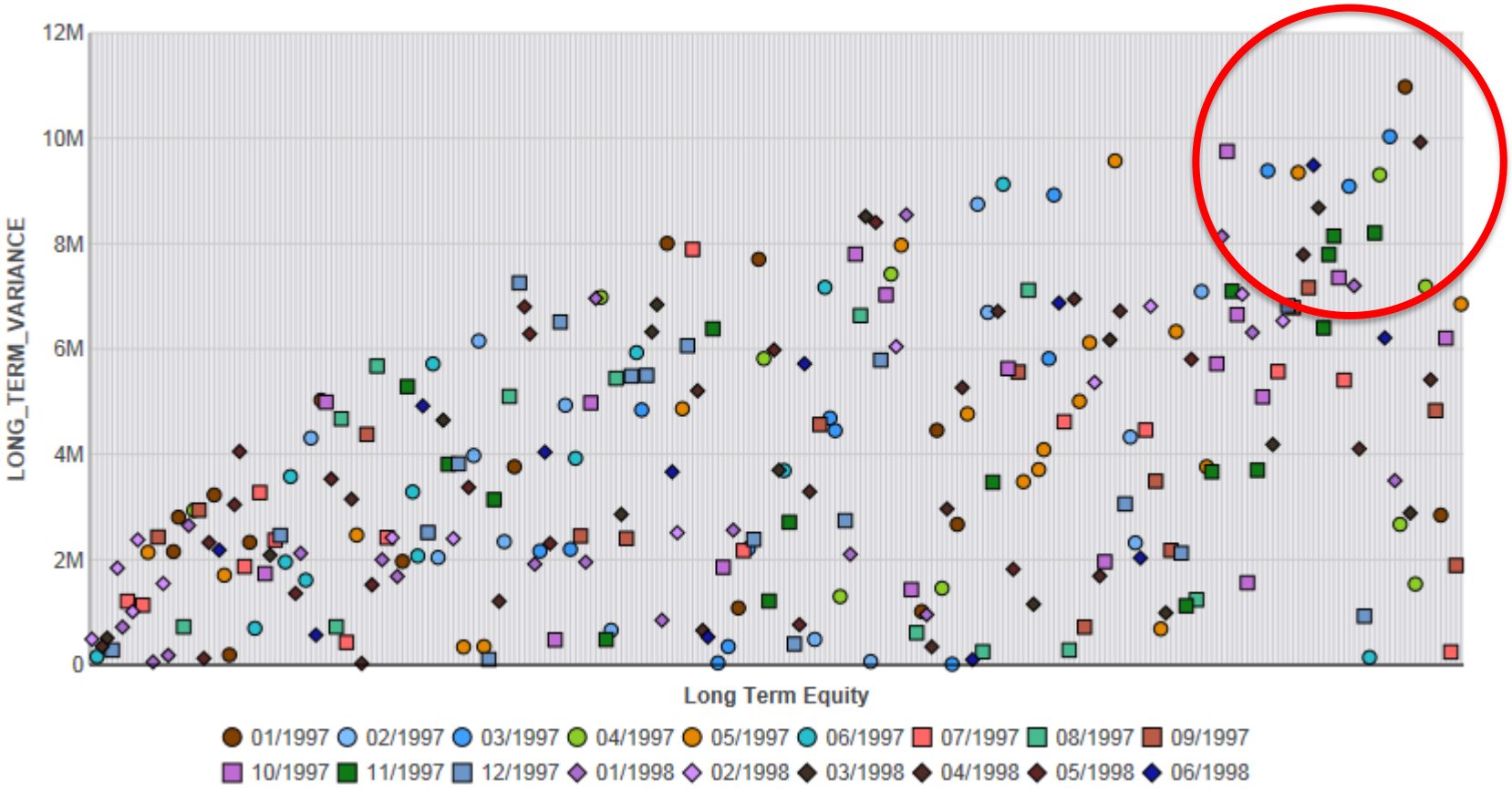


Quantitative data – 2 variables
Scatterplot

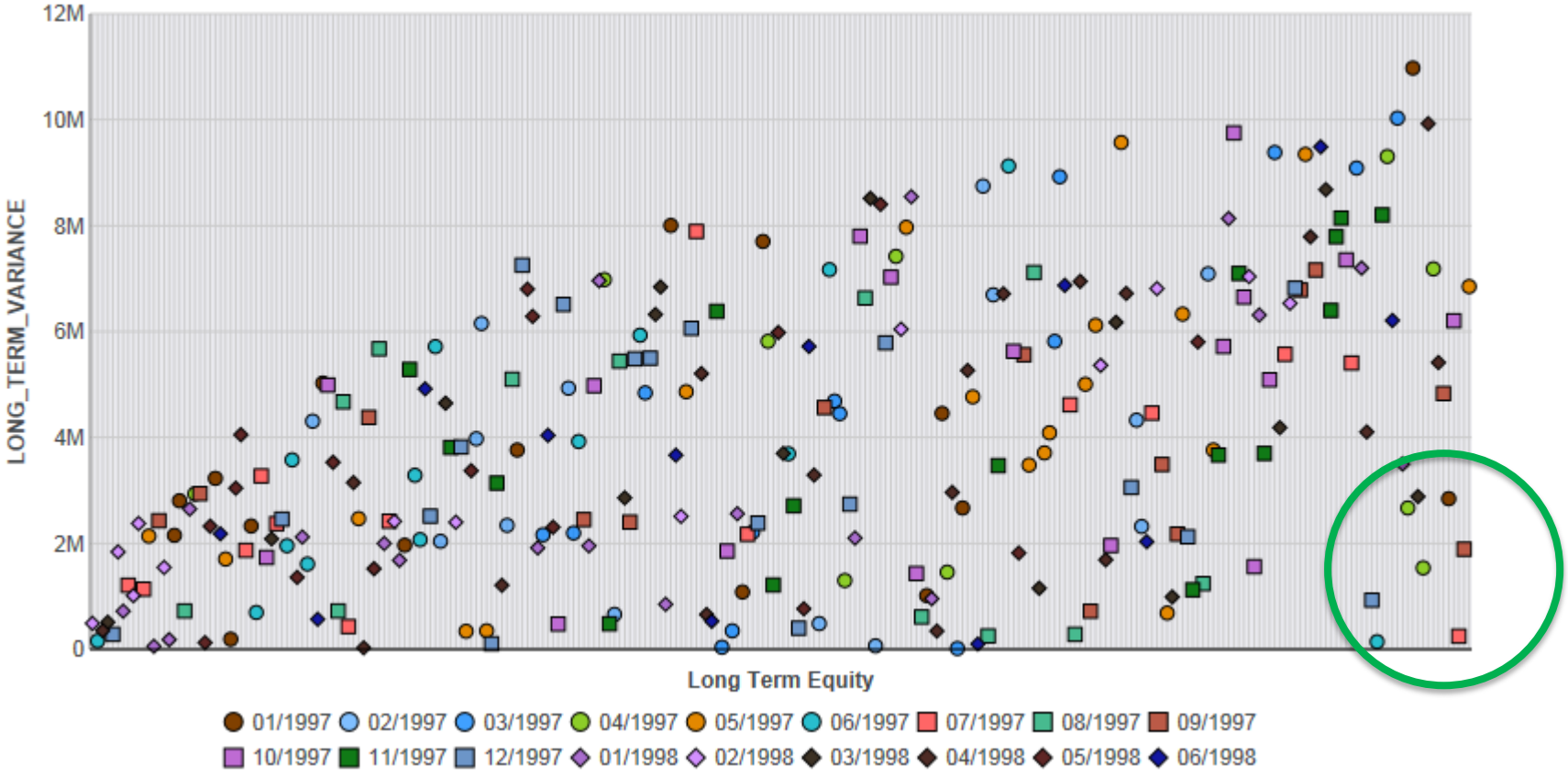


Quantitative data – 2 variables
Scatterplot

We don’t want this!

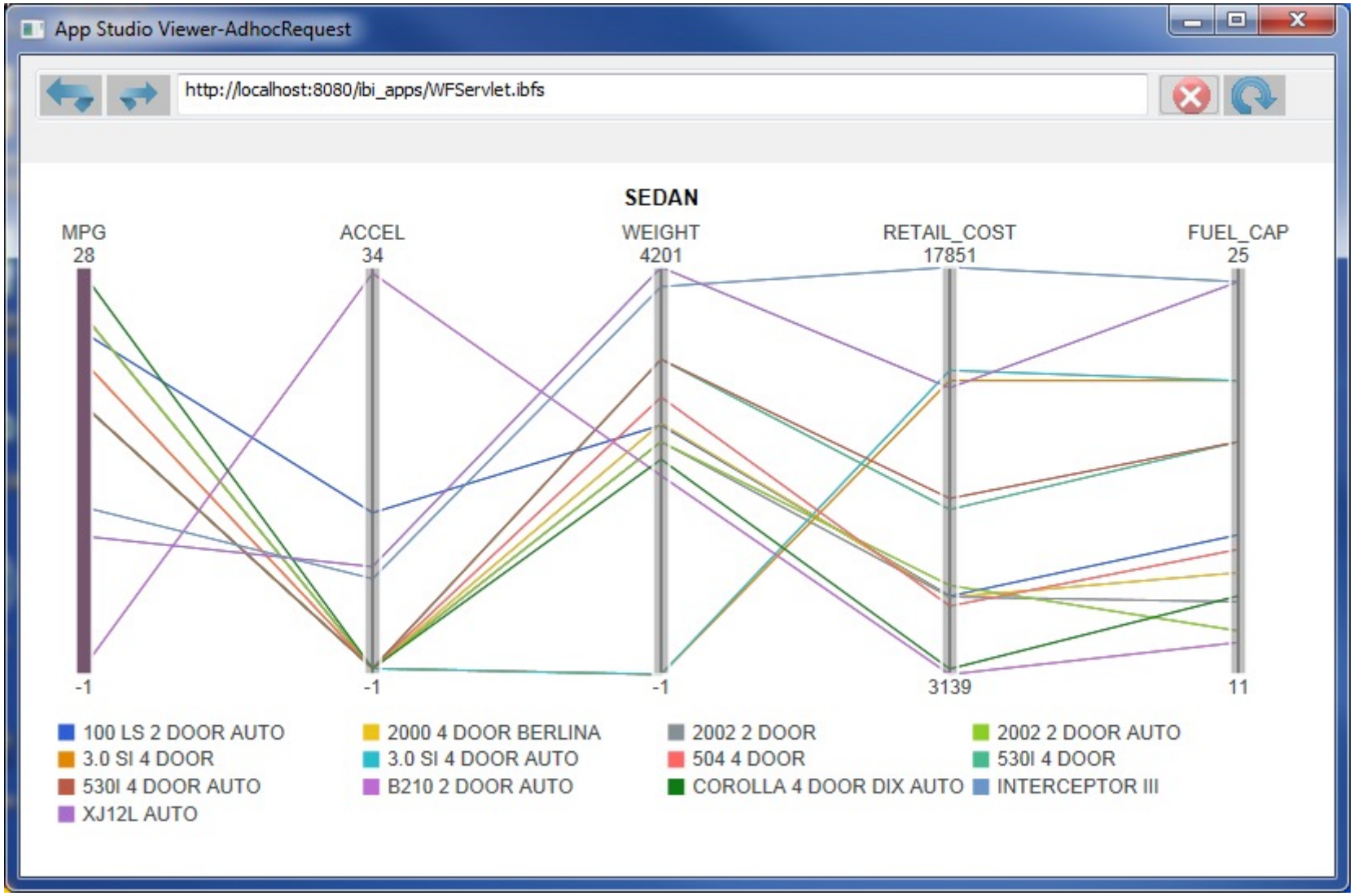


Quantitative data – 2 variables
Scatterplot



We want this!

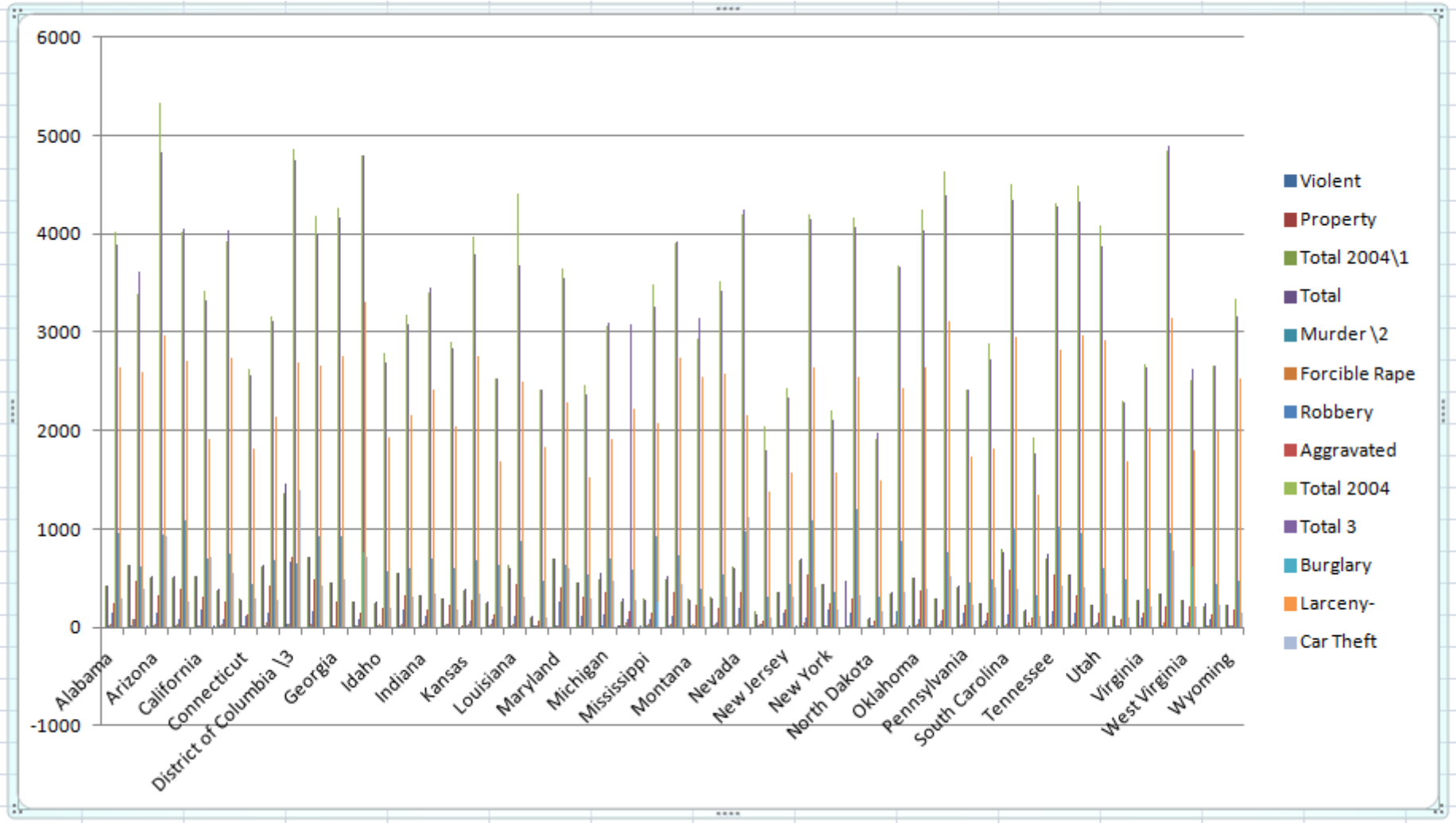
Product data (relationships) – multivariate Parallel Coordinates Chart

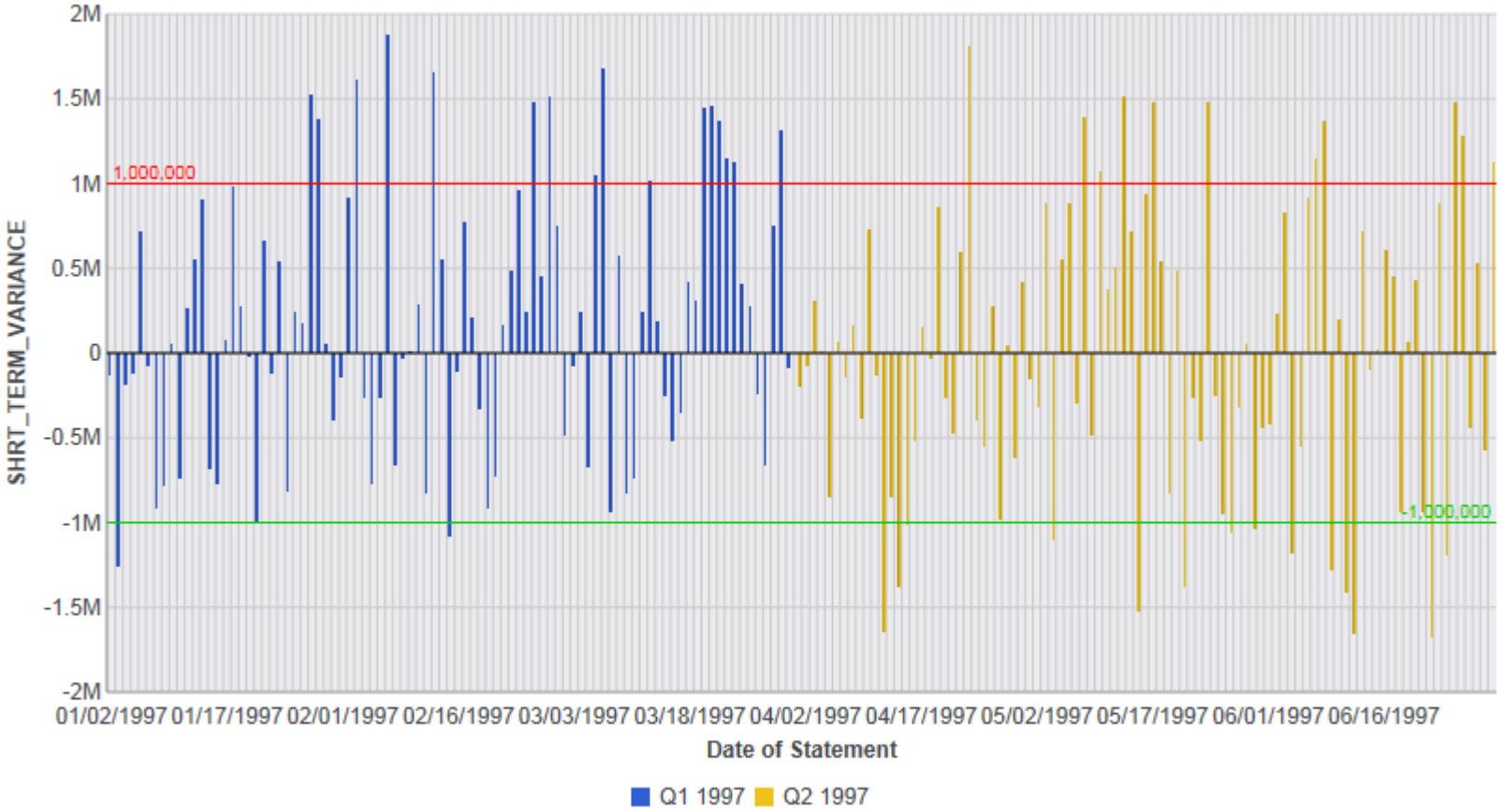


Data on Crime Rates

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	State	Violent	Property	Total 2004	Total	Murder \2	Forcible R	Robbery	Aggravate	Total 2004	Total 3	Burglary	Larceny-	Car Theft
2	Alabama	0	0	427	431.7	8.2	34.3	141.4	247.8	4025	3892.1	953.8	2650	288.3
3	Alaska	0	-0.1	632.3	631.9	4.8	81.1	80.9	465.1	3382.8	3612.5	622.5	2599.1	391
4	Arizona	0.1	0	504.4	513.2	7.5	33.8	144.4	327.4	5340.5	4838	948.4	2965.2	924.4
5	Arkansas	0	0	502.3	527.5	6.7	42.9	91.1	386.8	4013	4057.9	1084.6	2711.2	262.1
6	California	0	0	527.8	526.3	6.9	26	176.1	317.3	3419	3322.6	693.3	1916.5	712.8
7	Colorado	0.1	0	372	396.5	3.7	43.4	84.6	264.7	3919.3	4039.5	744.8	2735.2	559.5
8	Connectic	0	0	289	274.5	2.9	20	113	138.6	2627.2	2558	437.1	1824.1	296.8
9	Delaware	0	0	615	632.1	4.4	44.7	154.8	428.2	3163.9	3111.4	688.9	2144	278.5
10	District of	0	0.1	1369.4	1459	35.4	30.2	672.1	721.3	4859.1	4747	649.7	2694.9	1402.3
11	Florida	-0.1	0	711.8	708	5	37.1	169.4	496.6	4179.7	4007.9	926.3	2658.3	423.3
12	Georgia	0	0	451	448.9	6.2	23.6	154.8	264.3	4265.9	4172.3	931	2751.1	490.2
13	Hawaii	0	-0.1	254.6	255.1	1.9	26.9	78.5	147.8	4792.8	4792.6	767.9	3308.4	716.4
14	Idaho	0	0	247.4	256.8	2.4	40.4	18.6	195.4	2794.4	2697.9	564.4	1931.7	201.8
15	Illinois \4	-0.1	0	545.7	551.5	6	33.7	181.7	330.2	3186.1	3080.3	606.9	2164.8	308.6
16	Indiana	-0.1	0	325.9	323.7	5.7	29.6	108.6	179.9	3397.6	3456.3	697.6	2412	346.7
17	Iowa	-0.1	0	287.8	291.3	1.3	27.9	38.9	223.3	2905.3	2833.7	606.4	2042.7	184.6
18	Kansas	0	0.1	377.9	387.4	3.7	38.4	65.3	280	3973.5	3787	689.2	2758.1	339.6
19	Kentucky	0	-0.1	245.1	266.8	4.6	34	88.4	139.8	2537.7	2530.5	634	1685.8	210.8
20	Louisiana	0	-0.1	640	594.4	9.9	31.4	118	435.1	4410.2	3683.1	870.6	2494.5	318.1
21	Maine	0	0	103.7	112.2	1.4	24.7	24.4	61.7	2409.6	2413.1	478.5	1832.6	102
22	Maryland	0	0	700.6	703	9.9	22.6	256.7	413.8	3640.2	3544.1	641.4	2294.3	608.4
23	Massachu	0	0	460.2	456.9	2.7	27.1	119	308.1	2459.7	2363.6	541.1	1527.4	295.1
24	Michigan	0	0	492.2	552.1	6.1	51.3	131.8	362.9	3057.6	3091.1	696.8	1917.8	476.5
25	Minnesot	0.1	0.1	269.8	297	2.2	44	92	158.7	3039	3084.1	578.9	2226.9	278.2
26	Mississipp	0.1	0	295.4	278.4	7.3	39.3	82.3	149.4	3478.5	3260.1	919.7	2083.9	256.5
27	Missouri	0	-0.1	490.1	525.4	6.9	28	124.1	366.4	3903.5	3927.5	738.3	2746.2	443.1
28	Montana	0	0	293.8	281.5	1.9	32.2	18.9	228.5	2936.2	3142.9	389.2	2543	210.7
29	Nebraska	0	0	308.6	287	2.5	32.9	59.1	192.5	3520.6	3423.2	532.4	2574.3	316.5
30	Nevada	0	0	616.4	606.8	8.5	42.1	194.7	361.5	4206.6	4241.5	972.4	2153.9	1115.2
31	New Ham	0	0	169.5	132	1.4	30.9	27.4	72.3	2040.1	1796.4	317	1377.3	102.1
32	New Jerse	0	0	356.3	354.7	4.8	13.9	151.6	184.4	2429.2	2333	447.1	1568.4	317.5
33	New Mexi	0.1	0	687.4	702.2	7.4	54.1	98.7	541.9	4197.7	4148.3	1093.9	2639.9	414.5
34	New York	0	0	440.4	445.8	4.5	18.9	182.7	239.7	2198.6	2108.5	353.3	1569.6	185.6
35	North Car	0	0	447.8	468.1	6.7	26.5	145.5	289.4	4160.2	4075.1	1201.1	2546.2	327.8
36	North Dak	0	0	87.7	98.2	1.1	24.2	7.4	65.5	1916.6	1978.2	311.9	1500.3	166
37	Ohio	-0.1	0	338.7	351.3	5.1	39.8	163.1	143.4	3673.2	3662.7	872.8	2429	360.9
38	Oklahoma	0.1	0	500.5	508.6	5.3	41.7	91	370.5	4242.1	4042	1006	2644.2	391.8

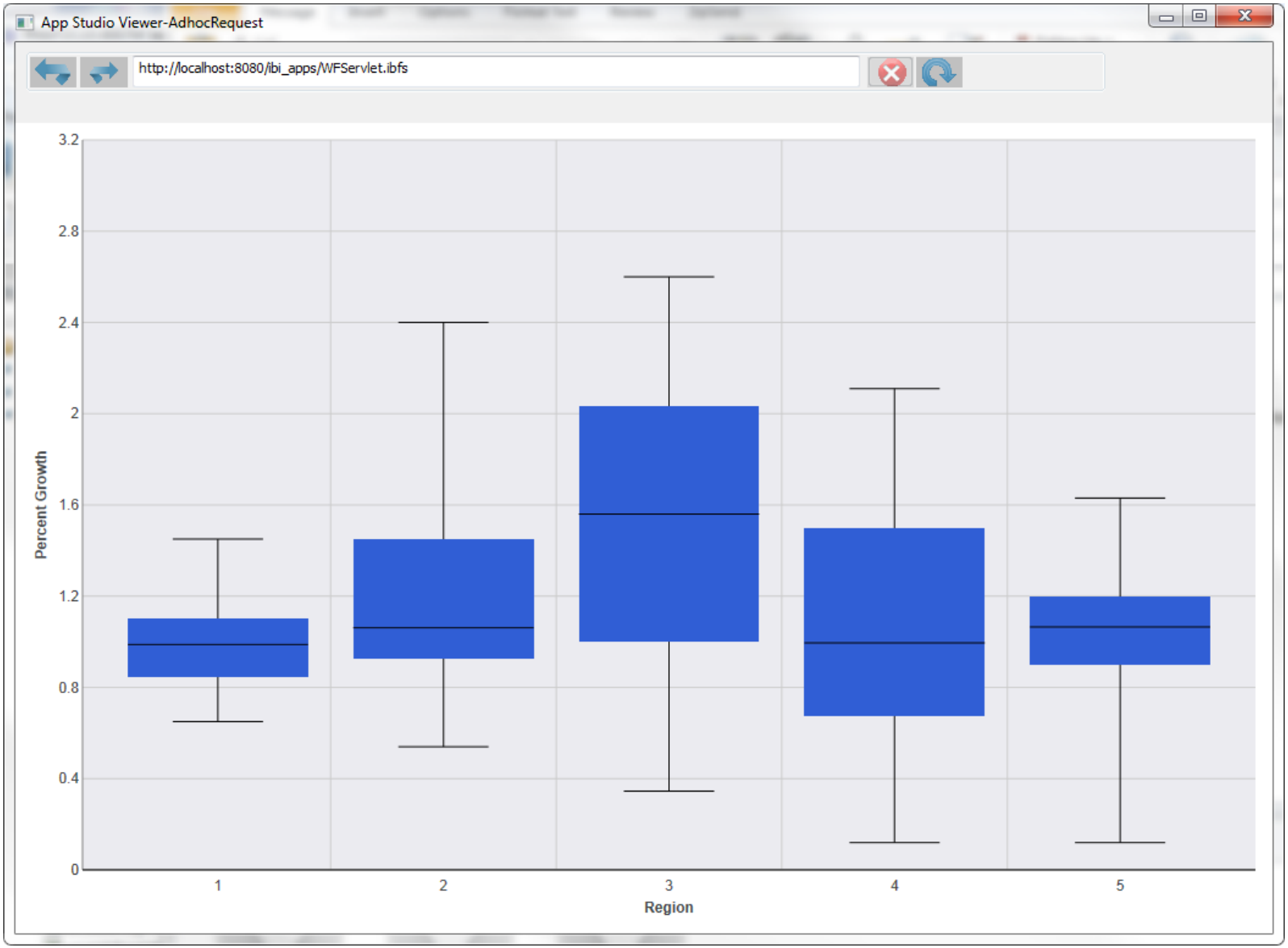
Standard Visualization

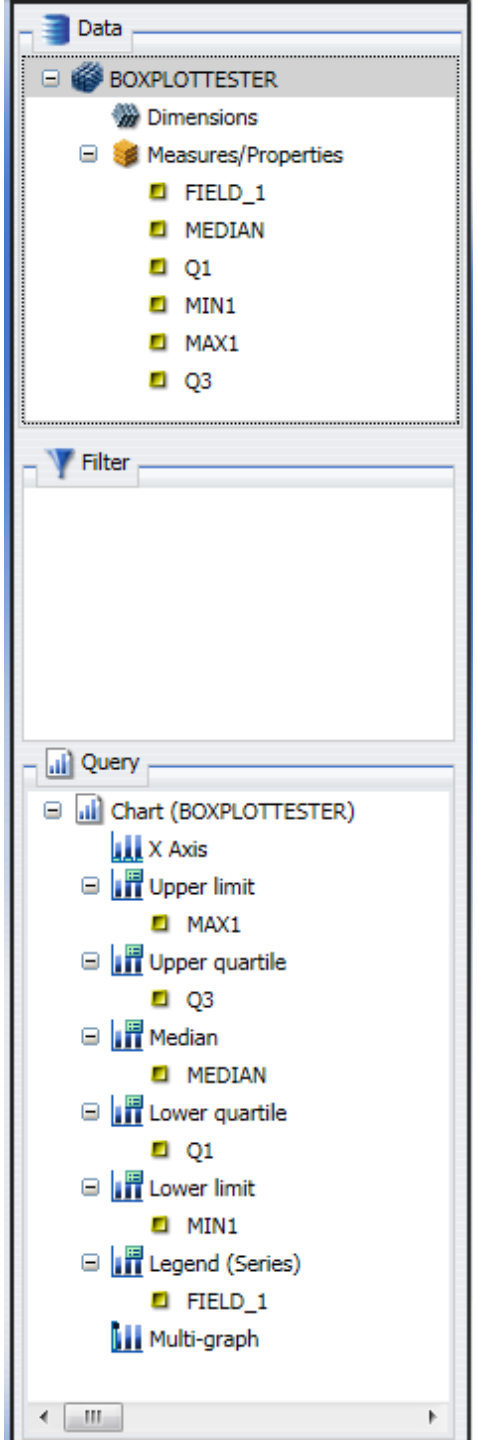



2M
1.5M
1M
0.5M
0
-0.5M
-1M
-1.5M
-2M
SHRT_TERM_VARIANCE
1,000,000
-1,000,000
01/02/1997
01/17/1997
02/01/1997
02/16/1997
03/03/1997
03/18/1997
04/02/1997
04/17/1997
05/02/1997
05/17/1997
06/01/1997
06/16/1997
Date of Statement
Q1 1997
Q2 1997

Let’s take a look at some files.

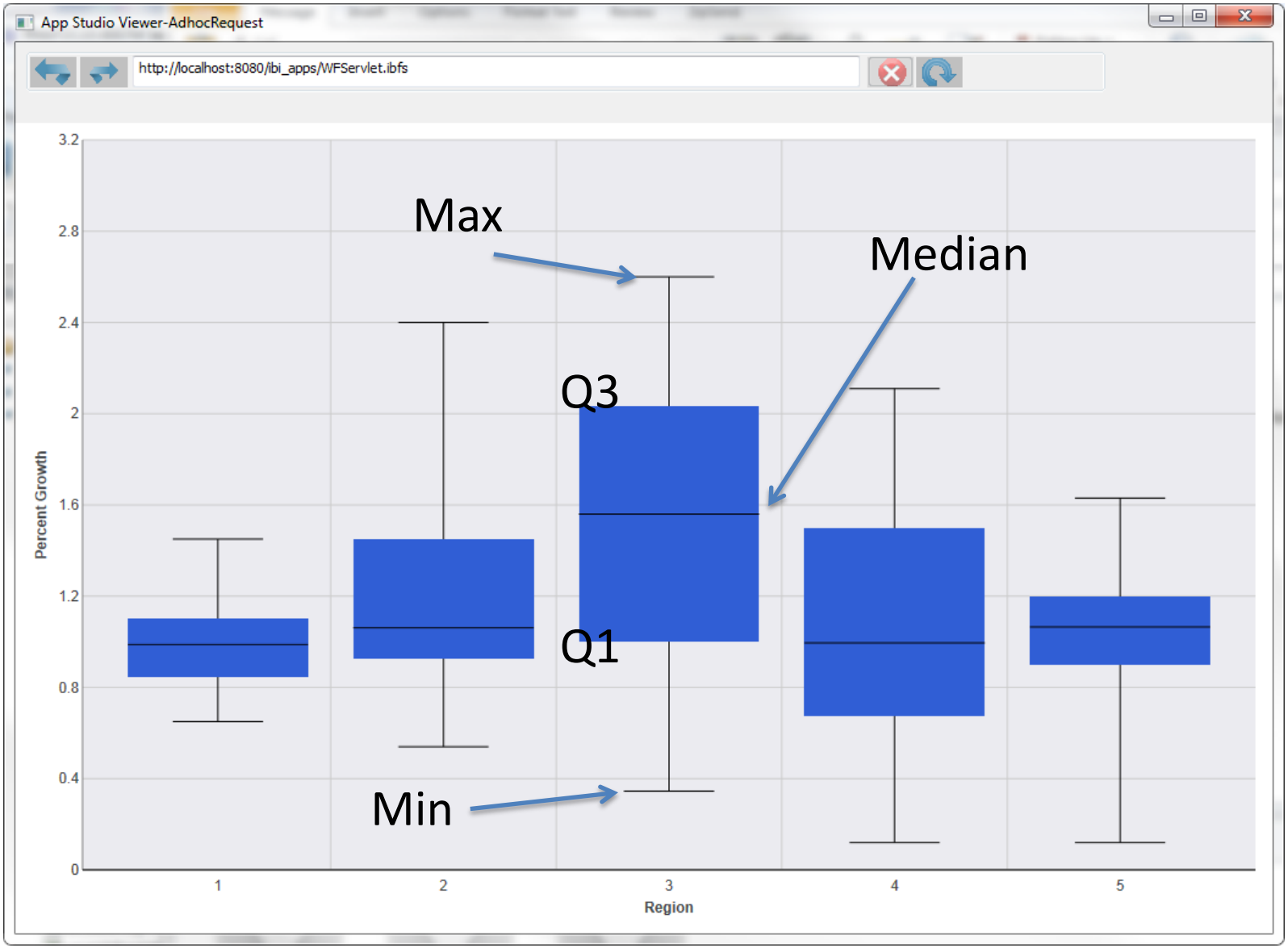
Open boxplotfromexceldata.fex





Components:

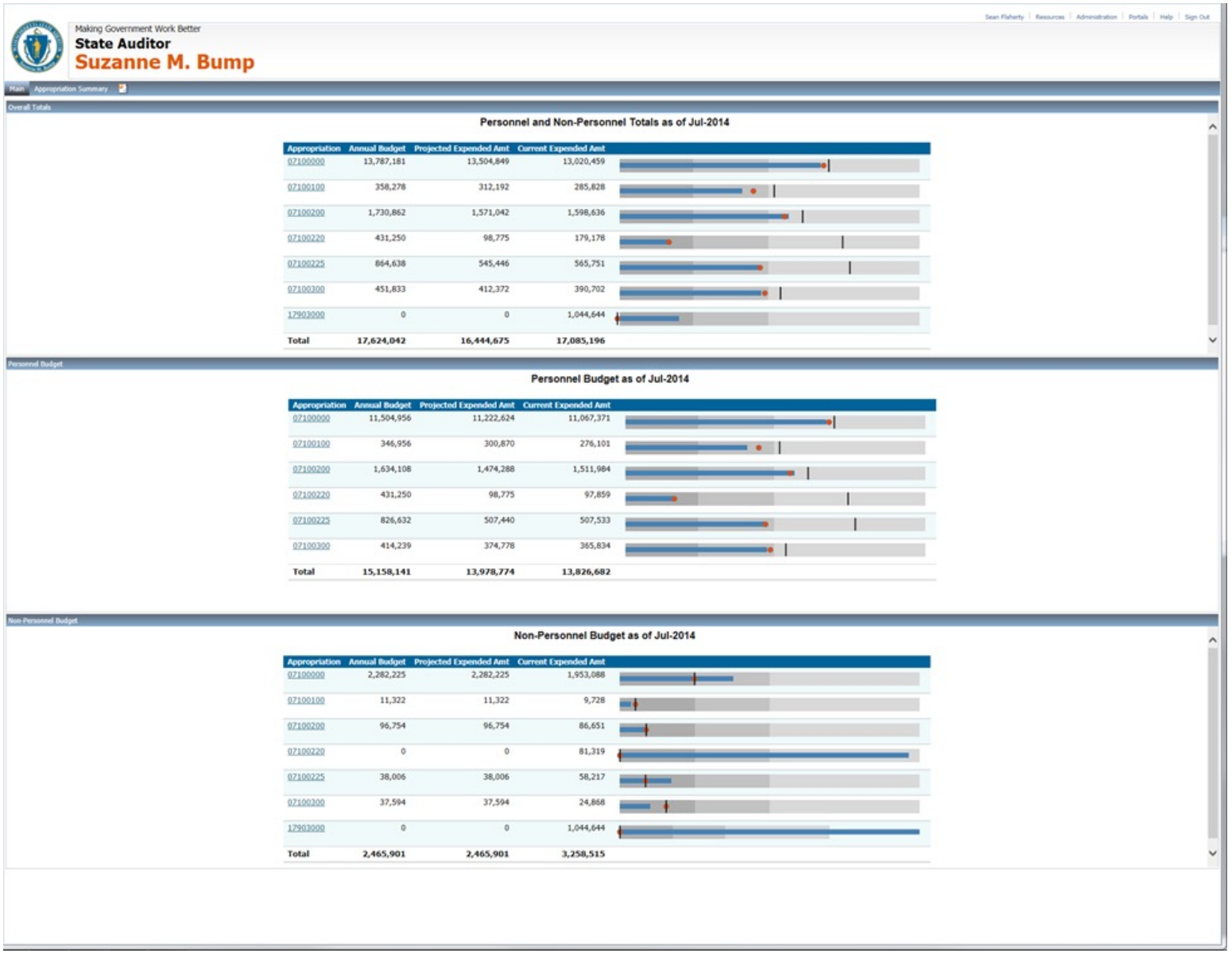
- **Median** – the number in the middle of the set when sorted in ascending order.
- **Min** – The smallest number in the set.
- **Max** – The largest number in the set.
- **Lower Quartile(Q1)** – the median of the numbers smaller than the set Median.
- **Upper Quartile(Q3)** – the median of the numbers larger than the set Median.


App Studio Viewer-AdhocRequest
http://localhost:8080/ibi_apps/WFServlet.ibfs
Max
Median
Q3
Q1
Min
Percent Growth
3.2
2.8
2.4
2
1.6
1.2
0.8
0.4
0
1
2
3
4
5
Region

What happens if we look at the same data in a different way?

Open worldmarketsscatter.fex

I have the code for this – email jeff_hendrickson@ibi.com to get it.



Data Types

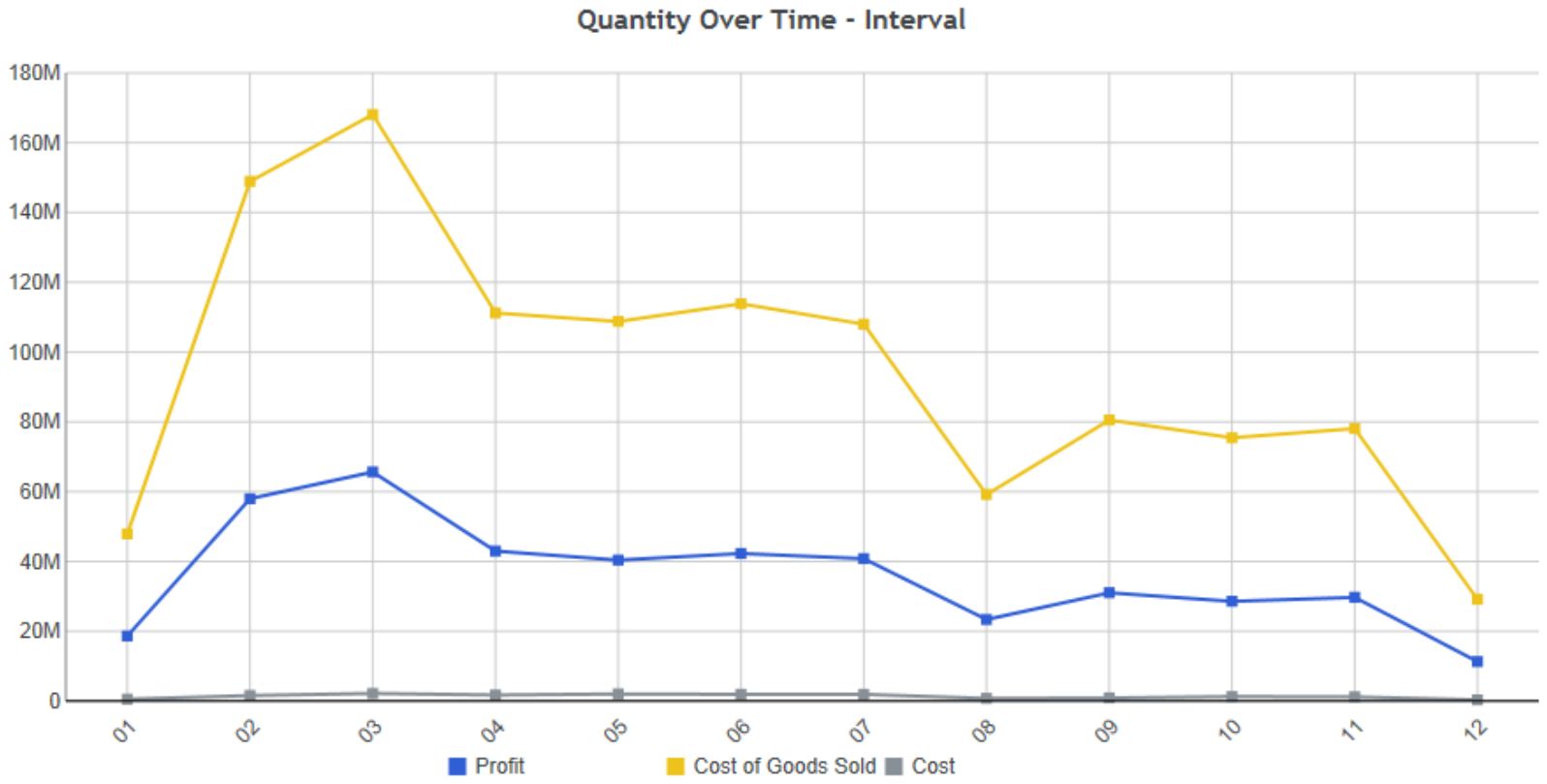
Quantitative and Continuous vs Categorical

Quantitative information forms the core of what organizations must know to operate effectively. The current emphasis on metrics, Key Performance Indicators(KPIs), Balanced Scorecards, and performance dashboards demonstrates the importance of numbers to organizations today.

Stephen Few “Show Me the Numbers”

First a few distinctions or rules.

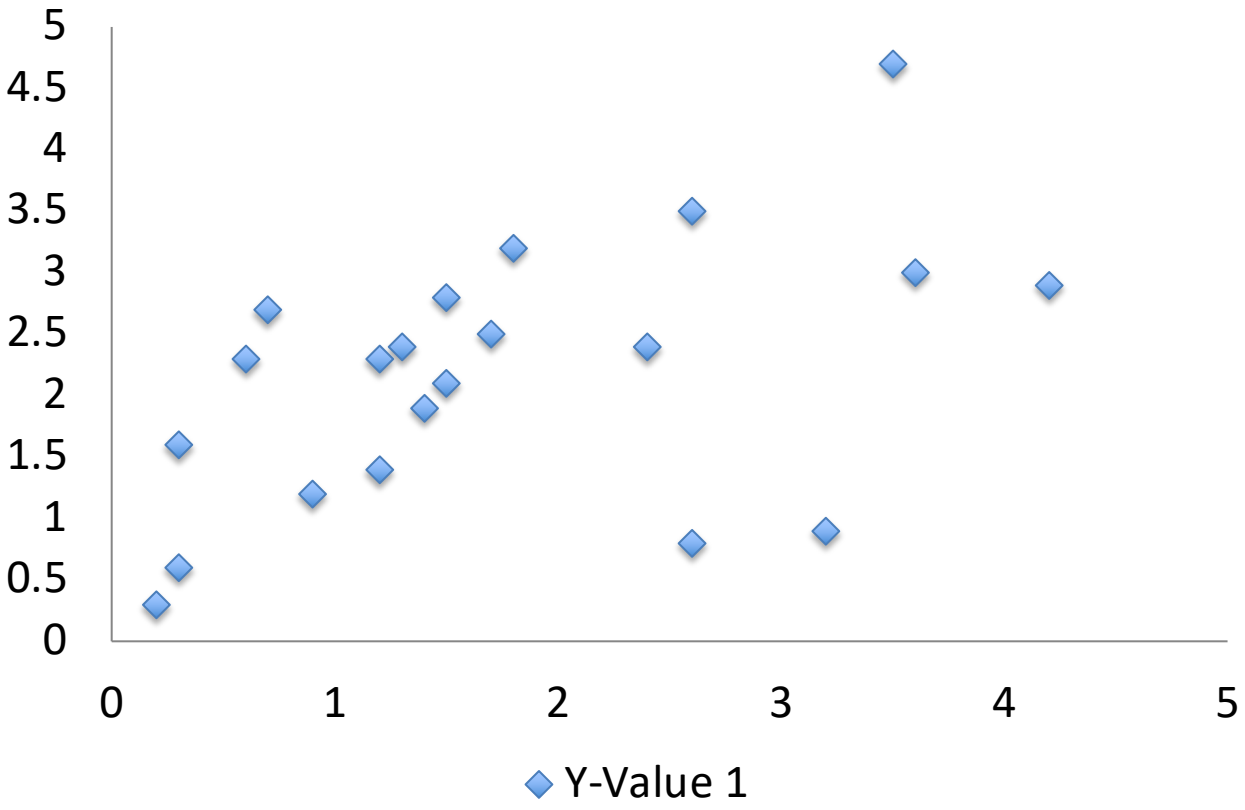
Two dimensional graphs usually have one quantitative scale and one categorical scale.



Categorical - Interval

Exception - Scatterplots

Two quantitative scales.

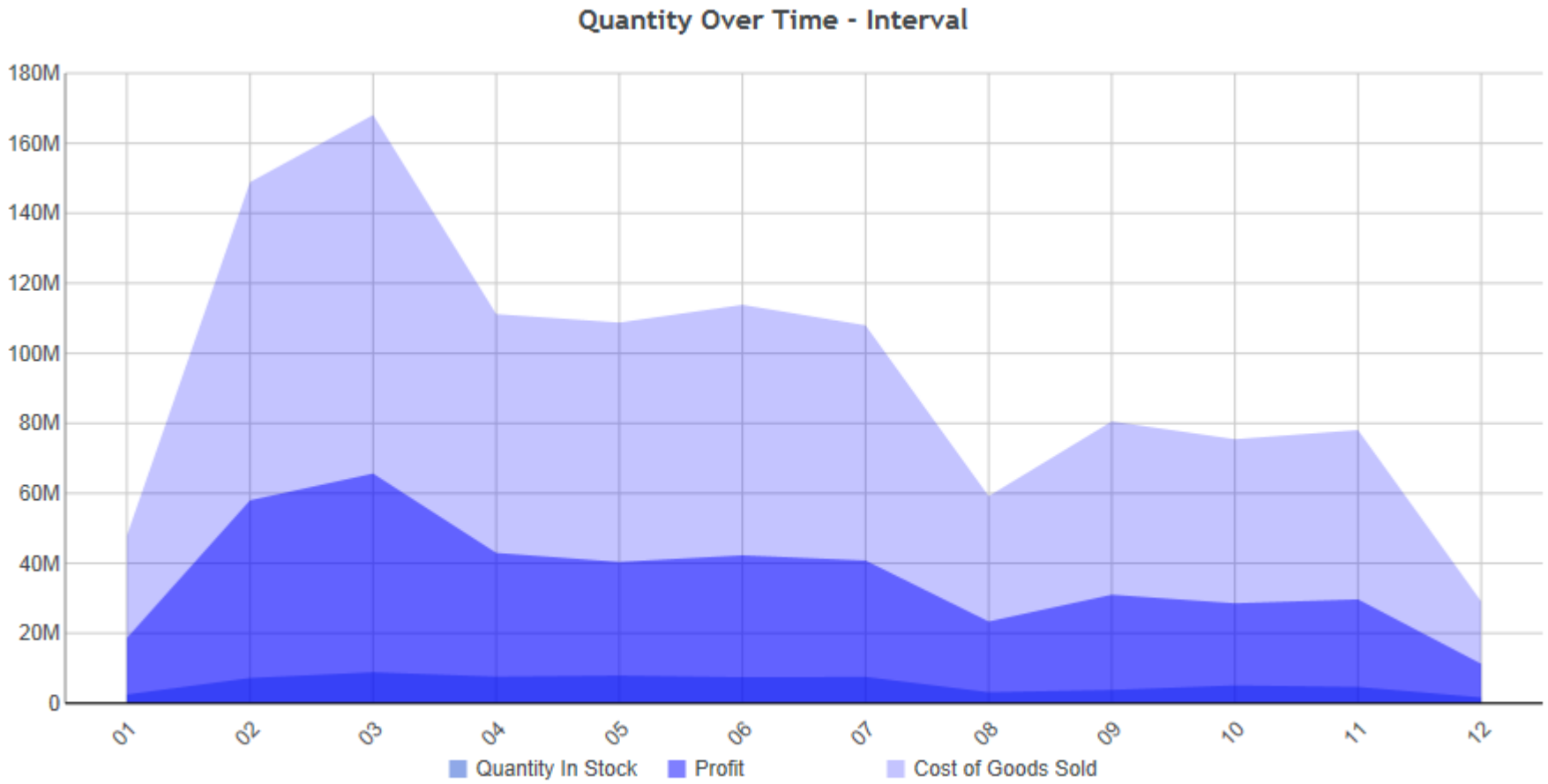


Explanatory vs Exploratory

Explanatory visualizations show charts/methods against known metrics or queries. *We know what we want to look at* and the vizzie helps us see it clearly.

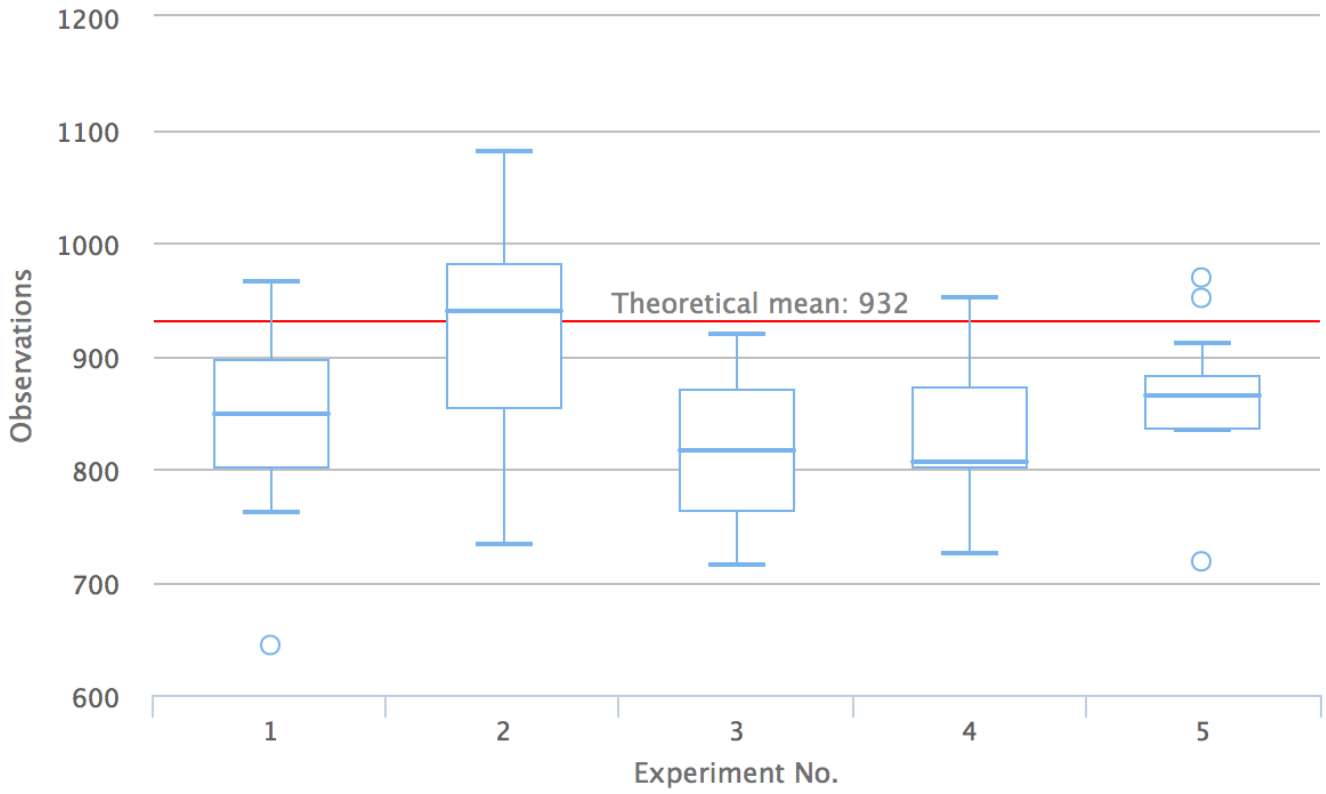
Exploratory visualizations help us find the questions that we may be missing. *We don't know what we're looking at or looking for.* A good vizzie in this instance will help us discover new ways to look at our data and will allow us to find anomalies or deeper patterns.

Explanatory – Focused Investigation

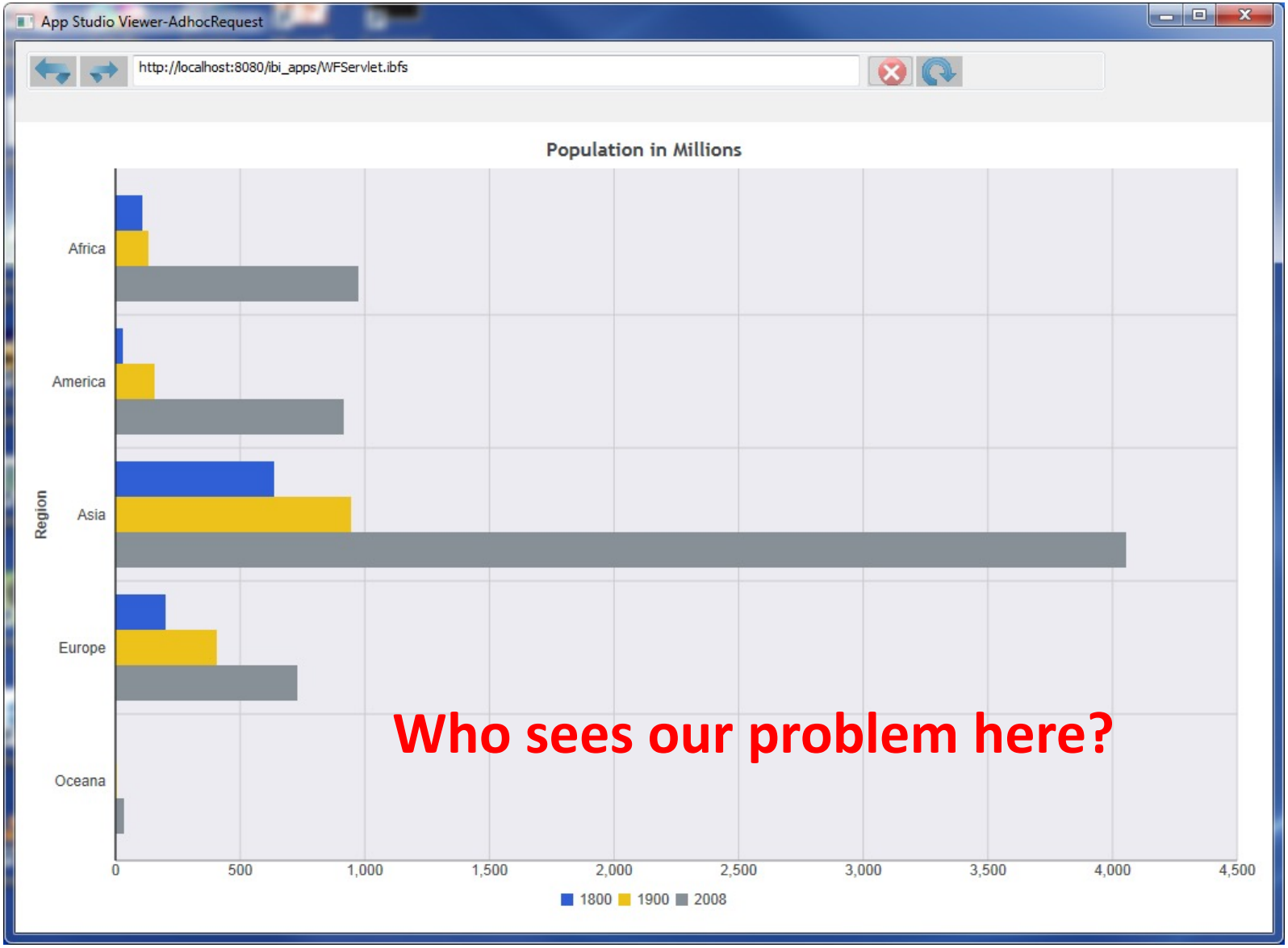


This is a chart that will run daily, weekly --- at regular intervals for reporting.

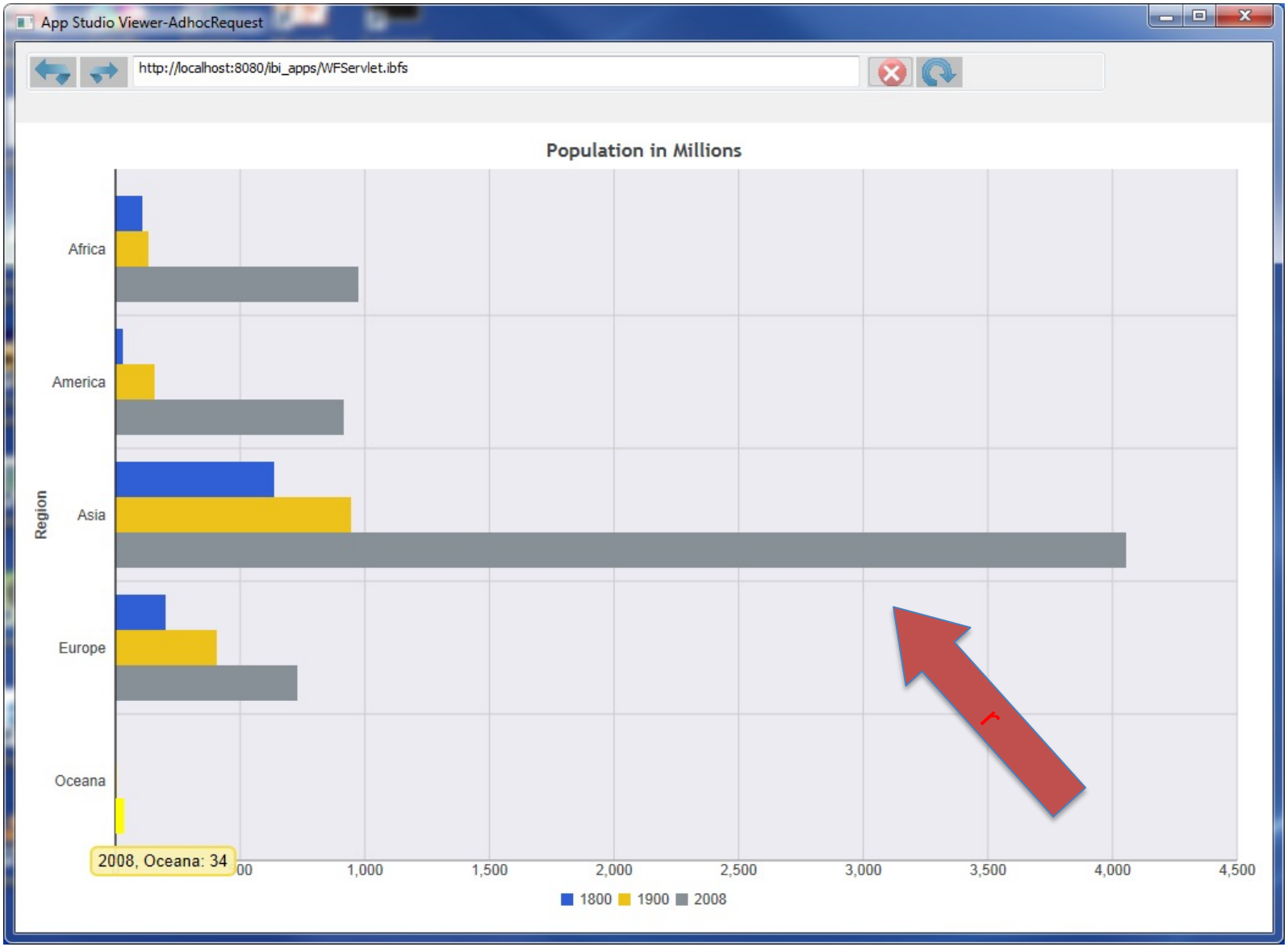
Explanatory



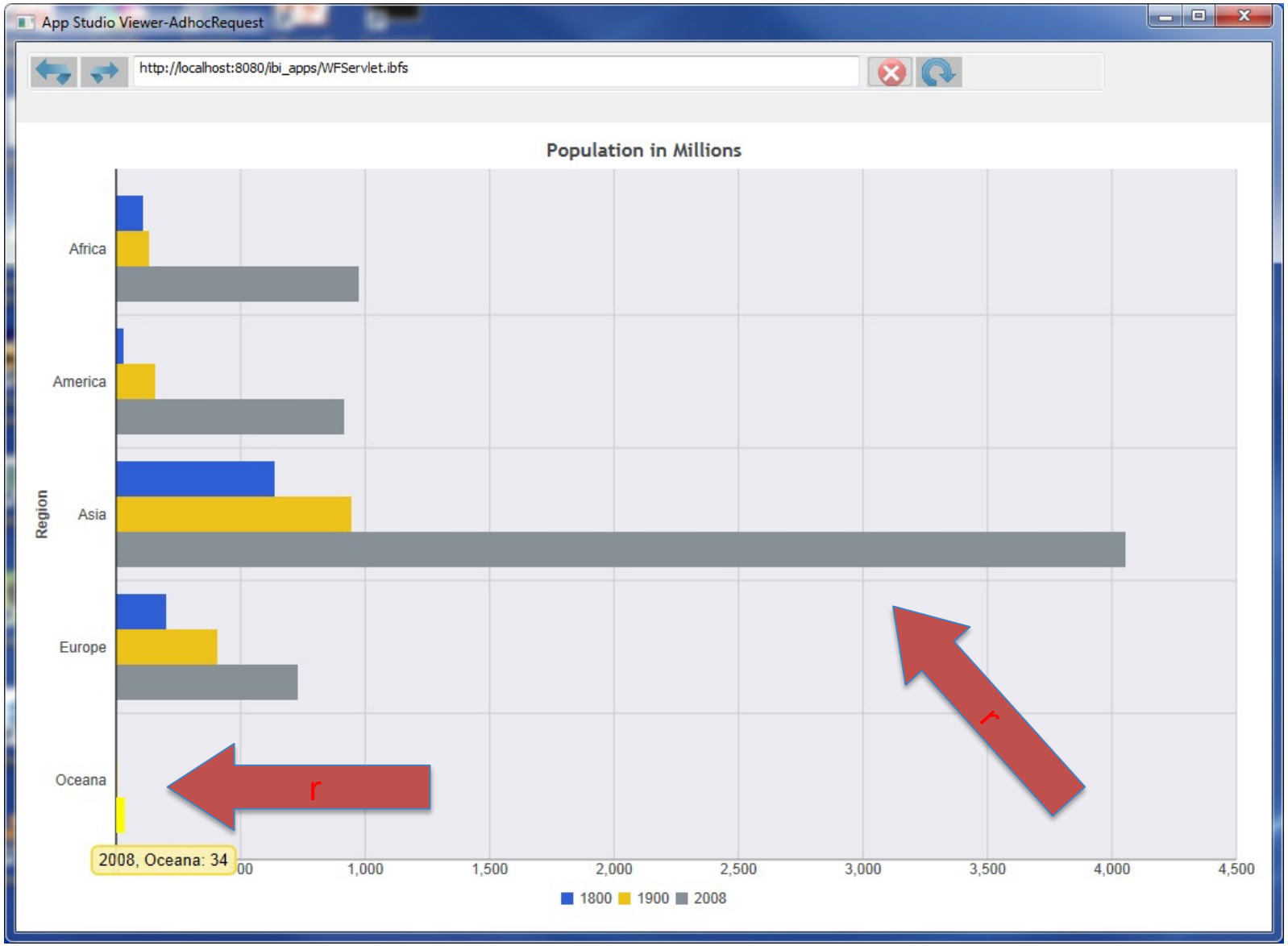
Explanatory



Parameterization

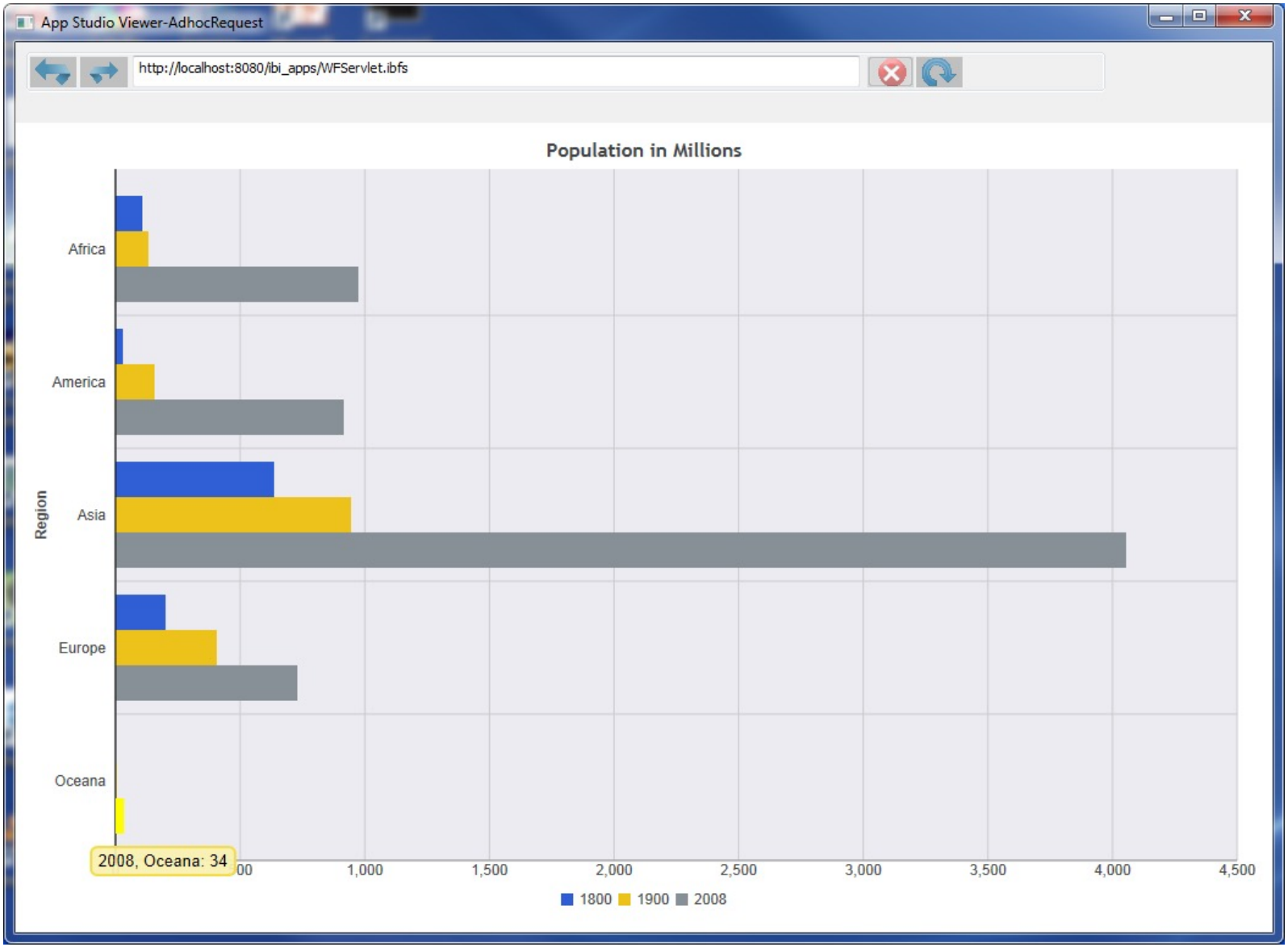


Parameterization



The numbers are so vastly different that the small amounts are almost invisible.

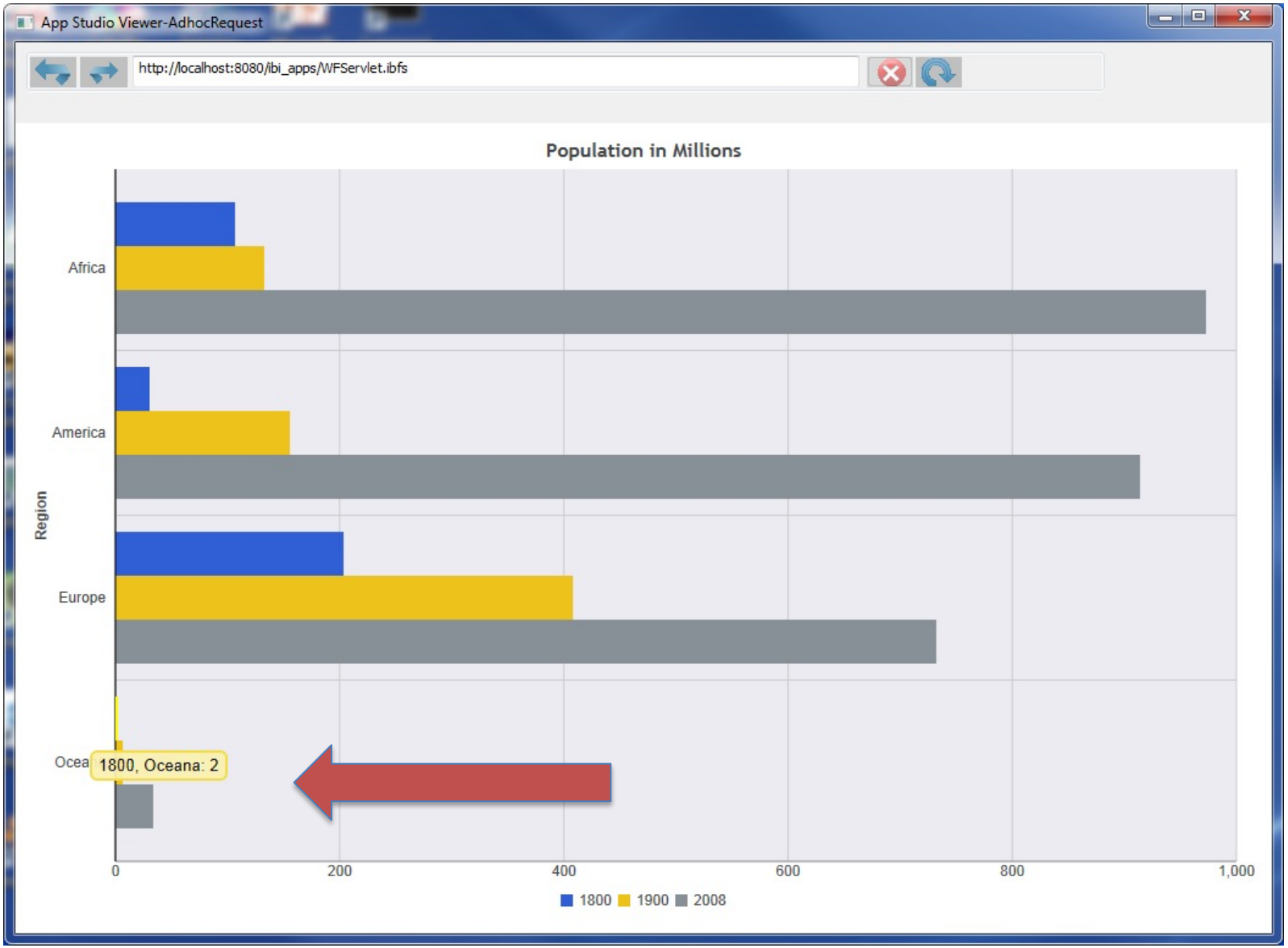
Parameterization



Back to start – we can use paramaterization.

Parameterization

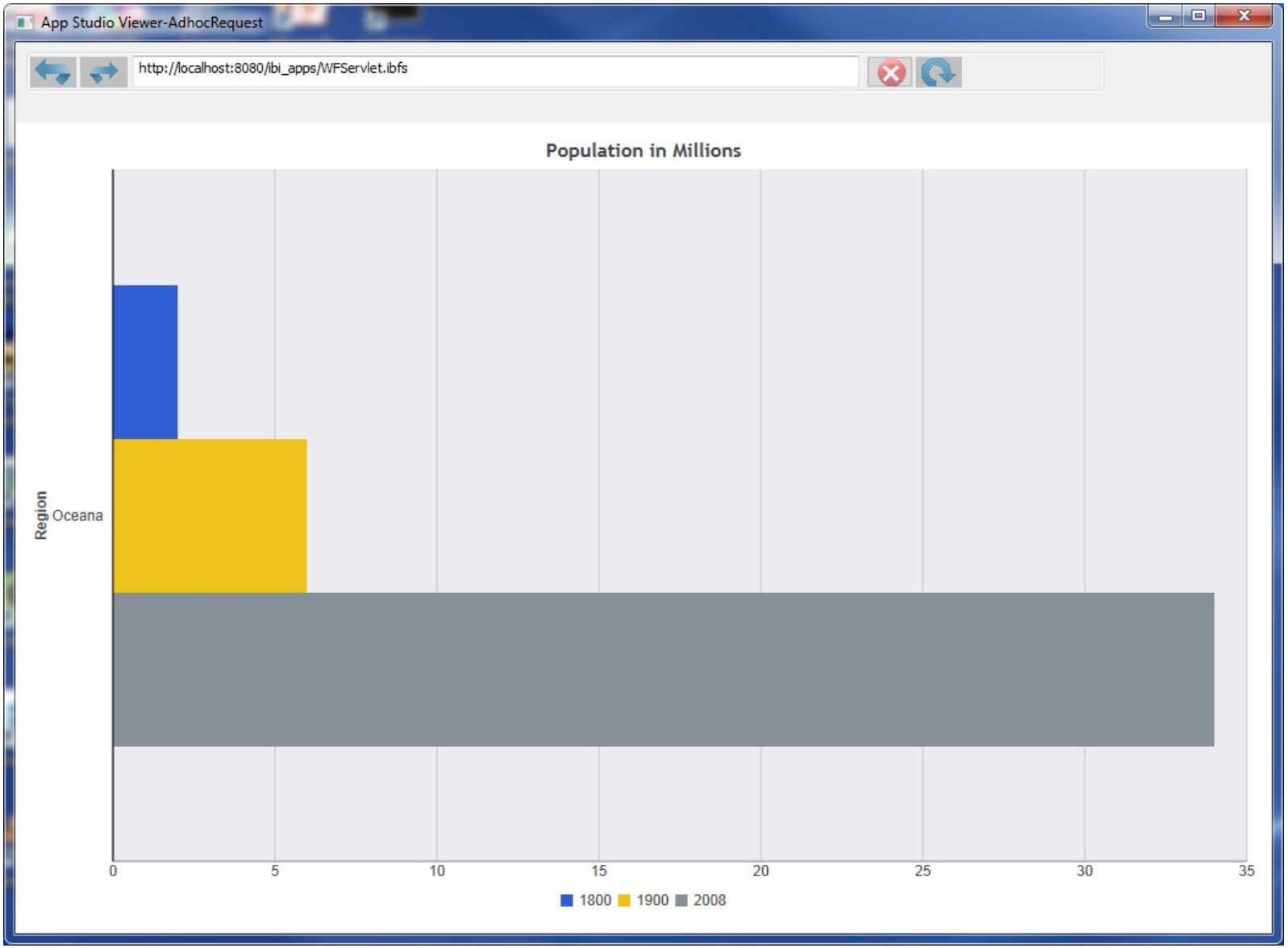
To get to here - optionally.



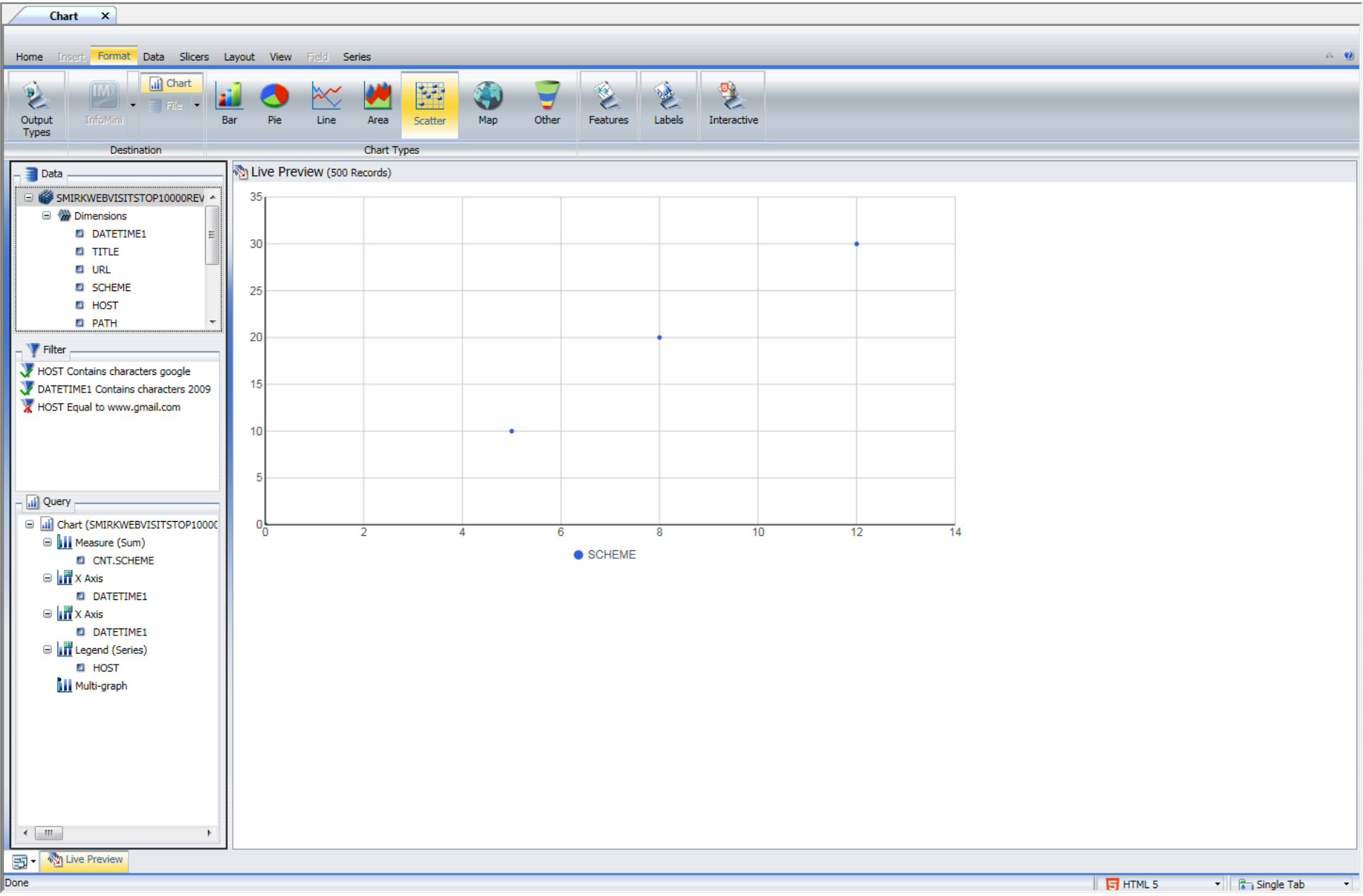
Parameterization

And then to here – Oceana only.

Open popinmillionsregiongrowth.fex

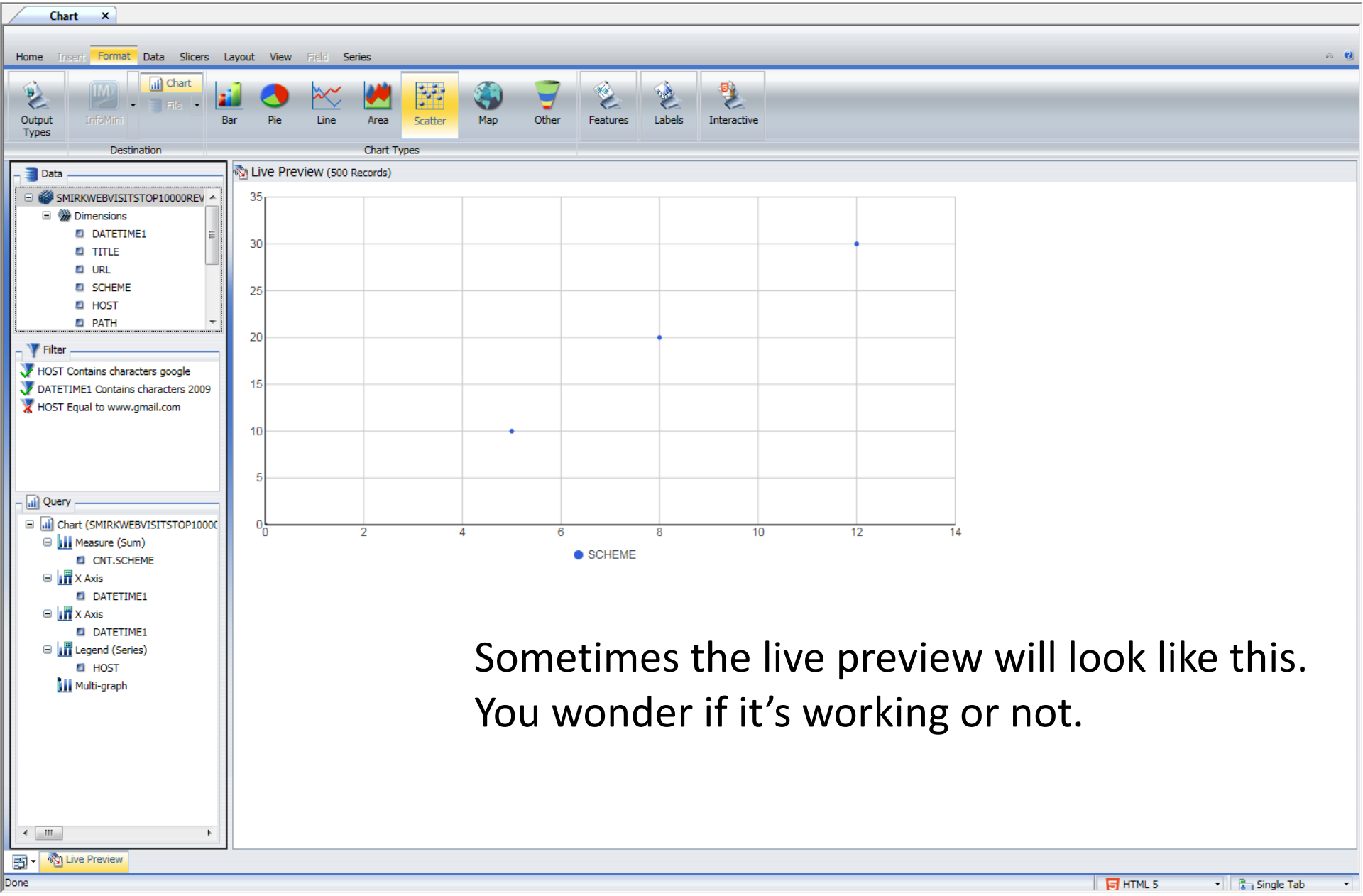


vbulscatternomarkers – not in your image



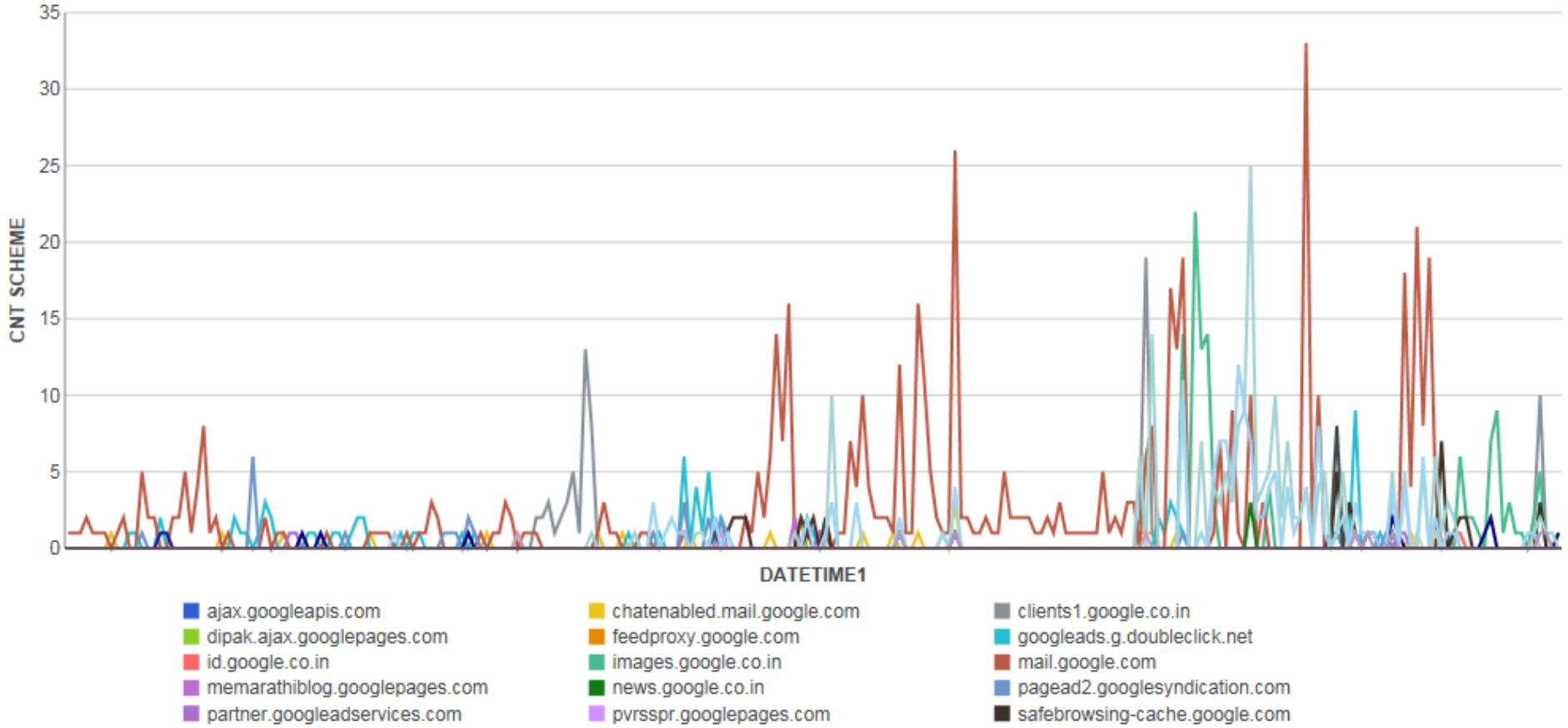
Let’s explore a bit.

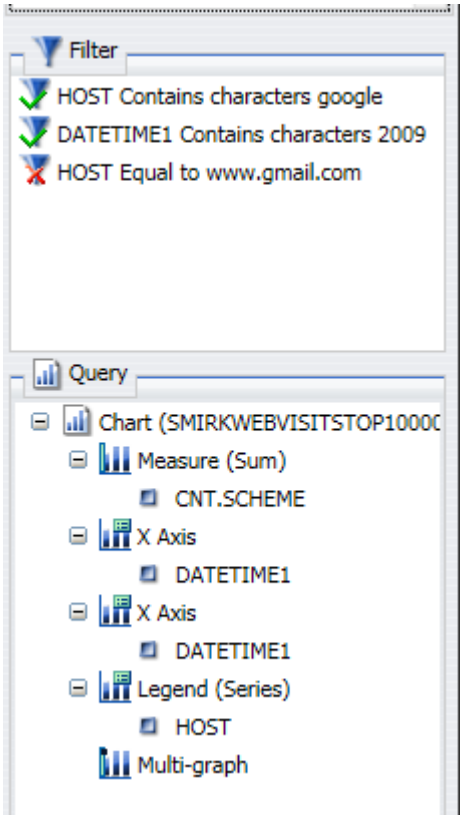
vbulgooglefilterscatternomarkers – not in your image

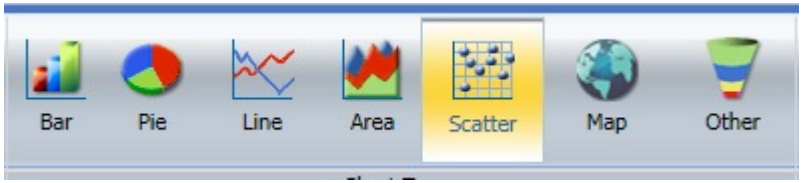


Sometimes the live preview will look like this.
You wonder if it's working or not.

It’s working.


35
30
25
20
15
10
5
0
CNT SCHEME
DATETIME1
ajax.googleapis.com
chatenabled.mail.google.com
clients1.google.co.in
dipak.ajax.googlepages.com
feedproxy.google.com
googleads.g.doubleclick.net
id.google.co.in
images.google.co.in
mail.google.com
memarathiblog.googlepages.com
news.google.co.in
pagead2.googlesyndication.com
partner.googleadservices.com
pvrsspr.googlepages.com
safebrowsing-cache.google.com

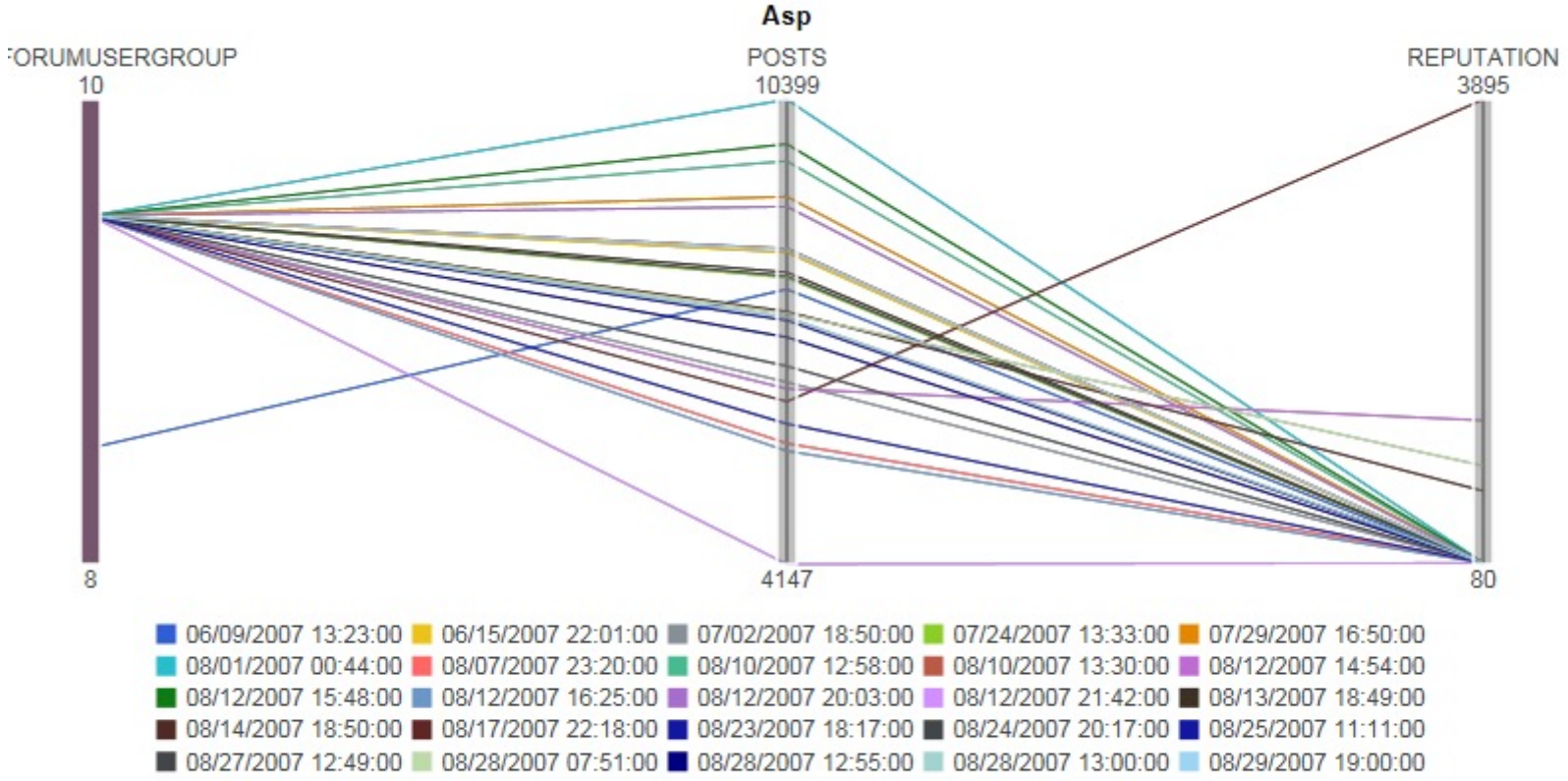

Filter
HOST Contains characters google
DATETIME1 Contains characters 2009
HOST Equal to www.gmail.com
Query
Chart (SMIRKWEBVISITSTOP1000
Measure (Sum)
CNT.SCHEME
X Axis
DATETIME1
X Axis
DATETIME1
Legend (Series)
HOST
Multi-graph


Bar
Pie
Line
Area
Scatter
Map
Other

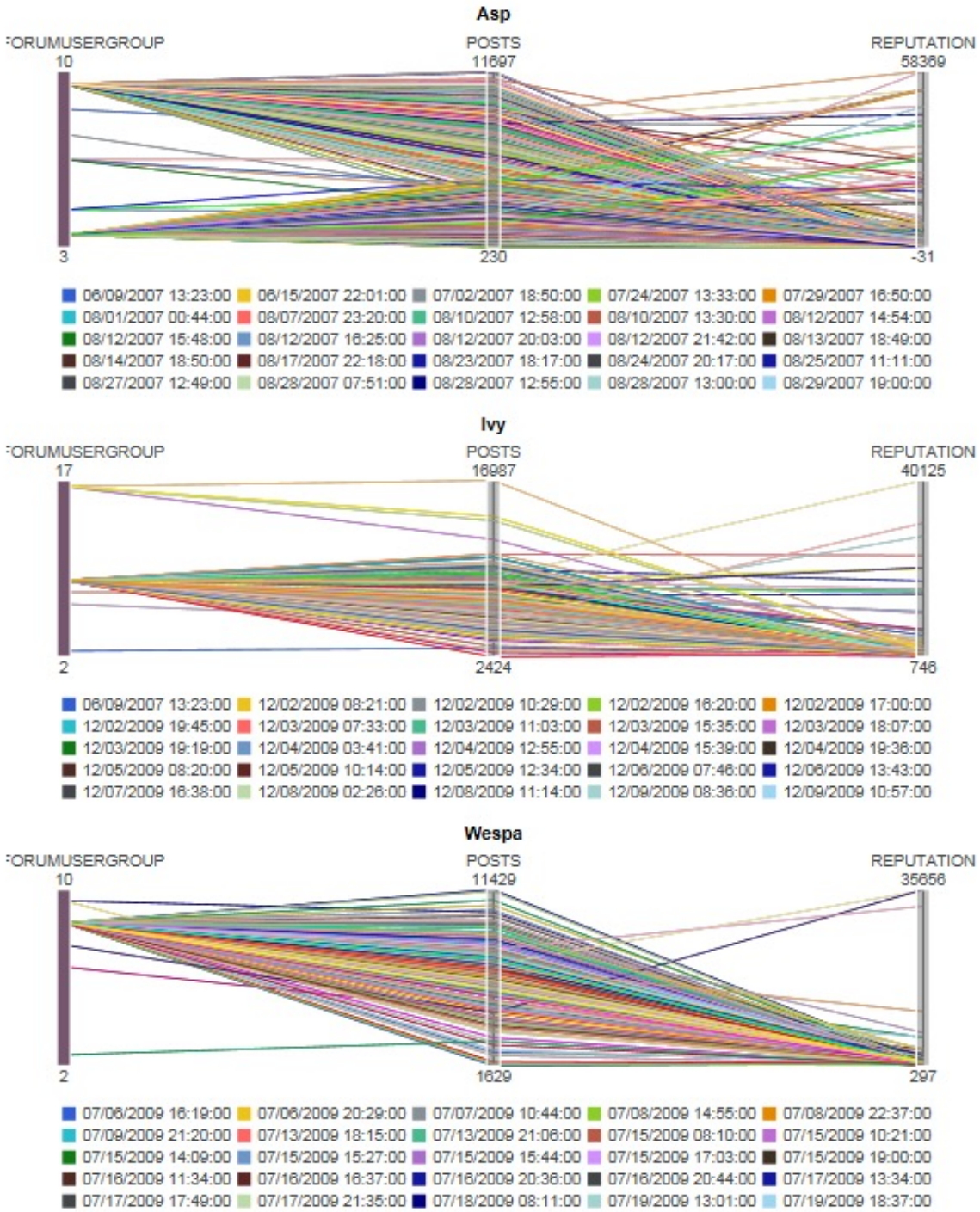
Parabox

Known in the industry mostly as Parallel Coordinates.

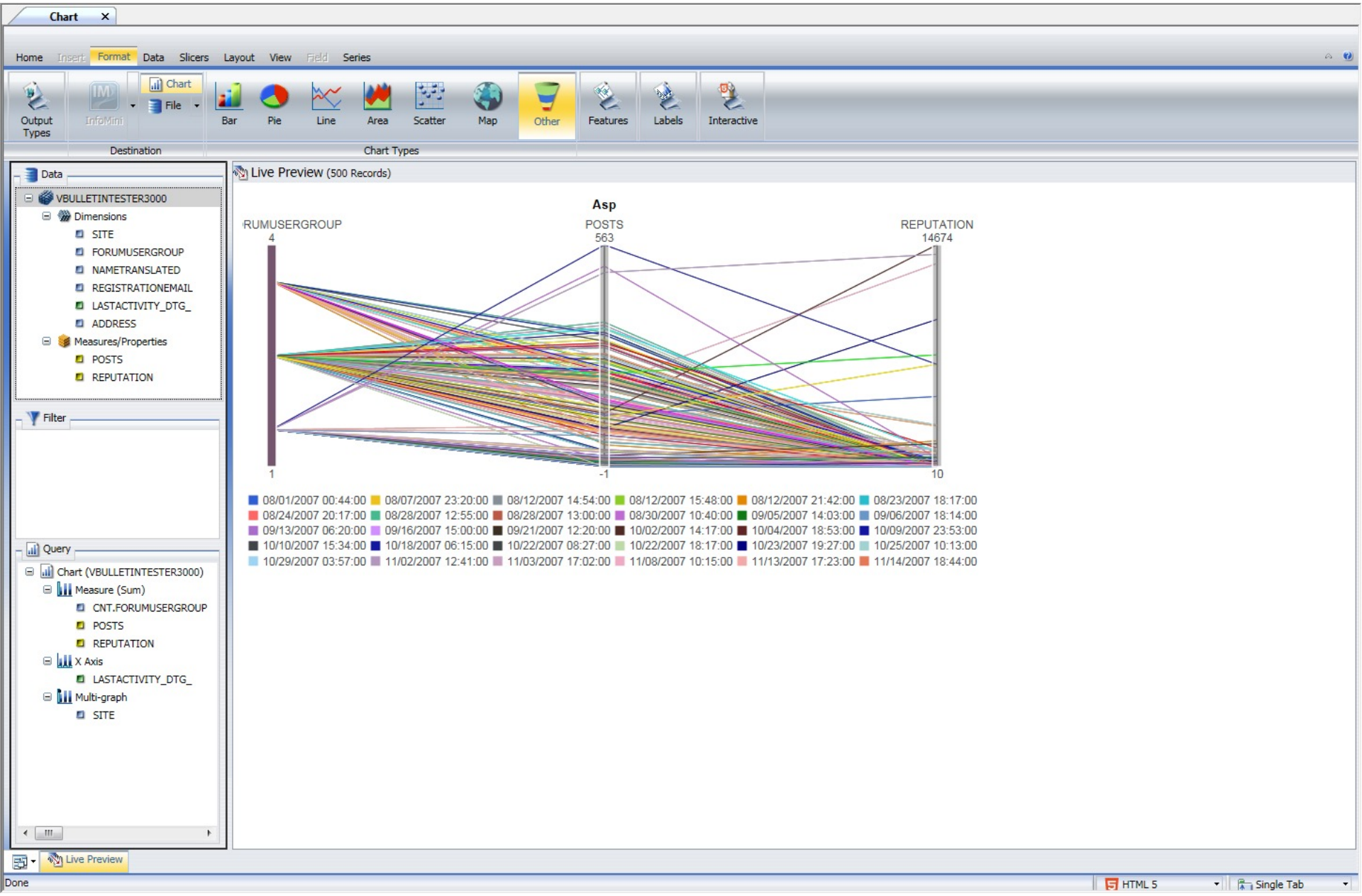
Open parcoordsvbul3000aspfilter.fex



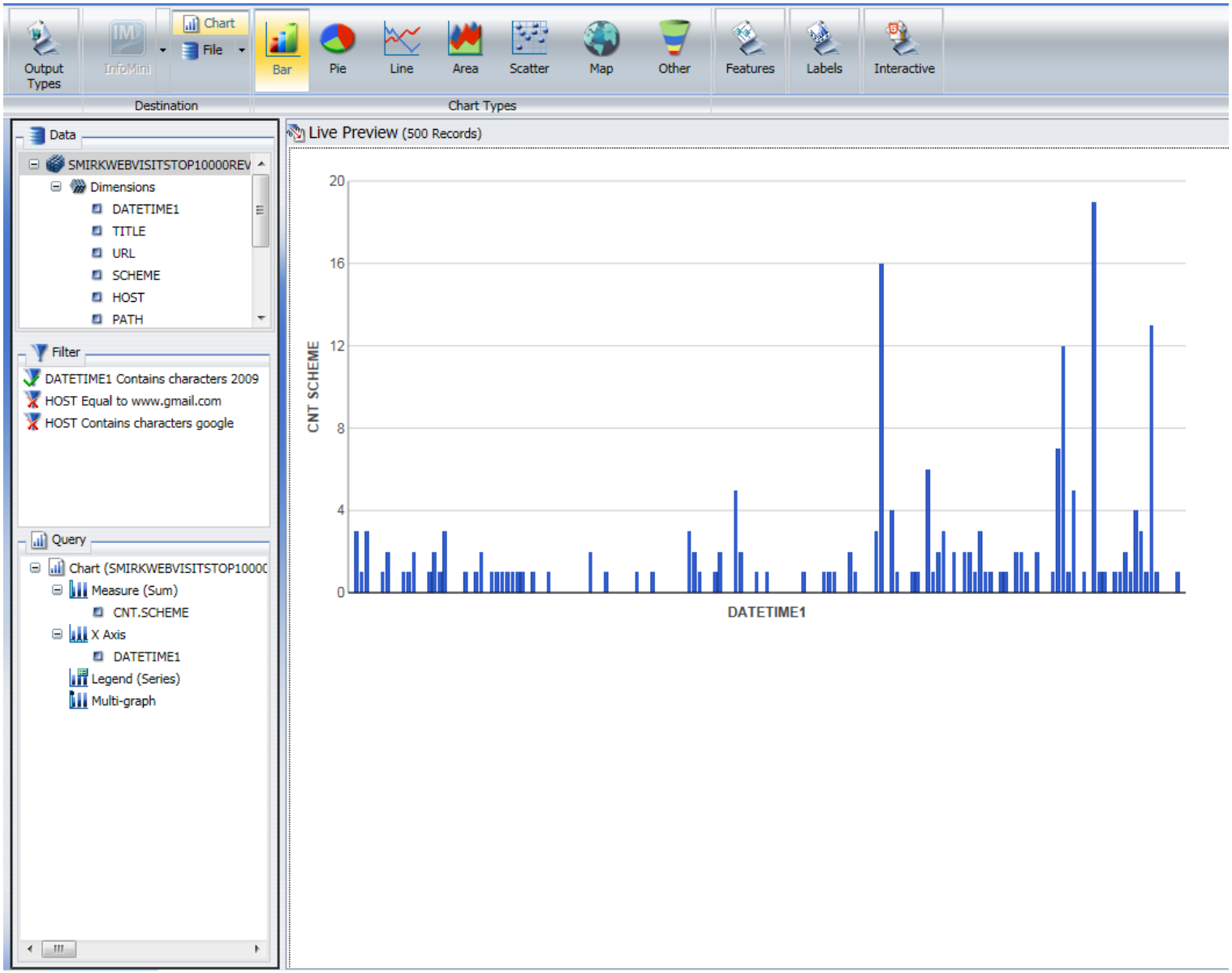
We're going to fix it to make this.



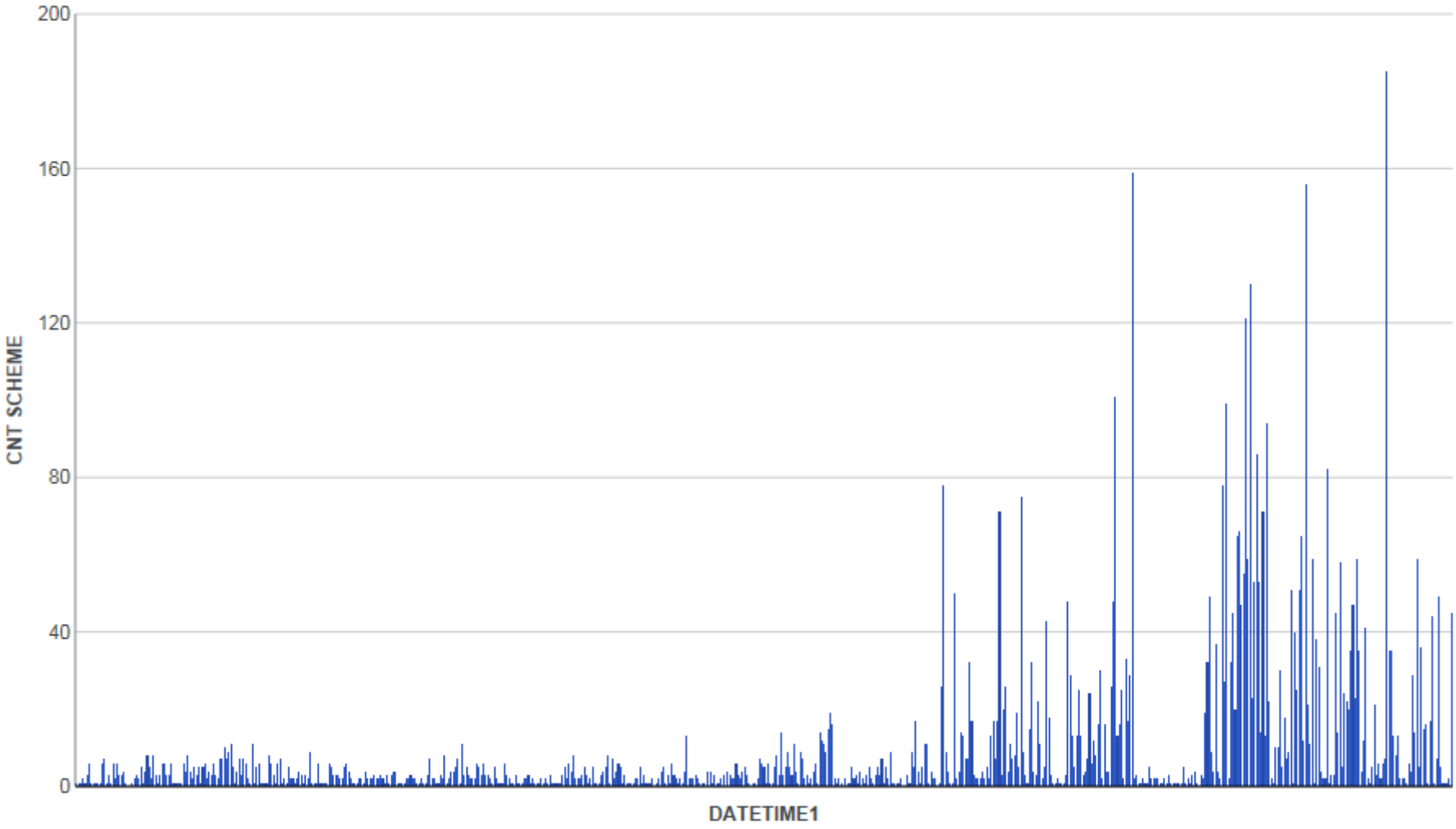
Here's what it will look like in Live Preview



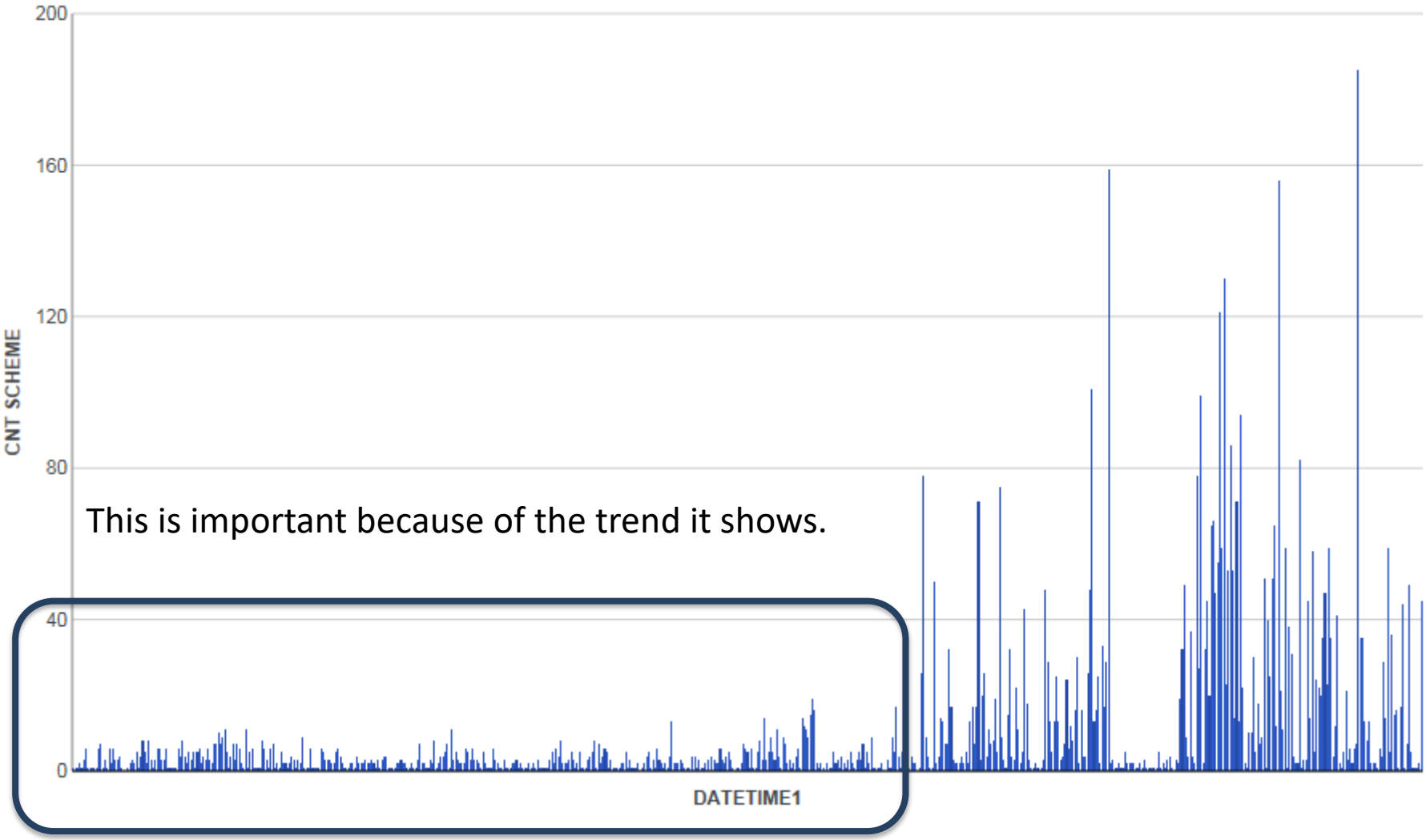
Open vbulsmirk1bar.fex



Good example of how a bar chart can be used for exploratory analysis.

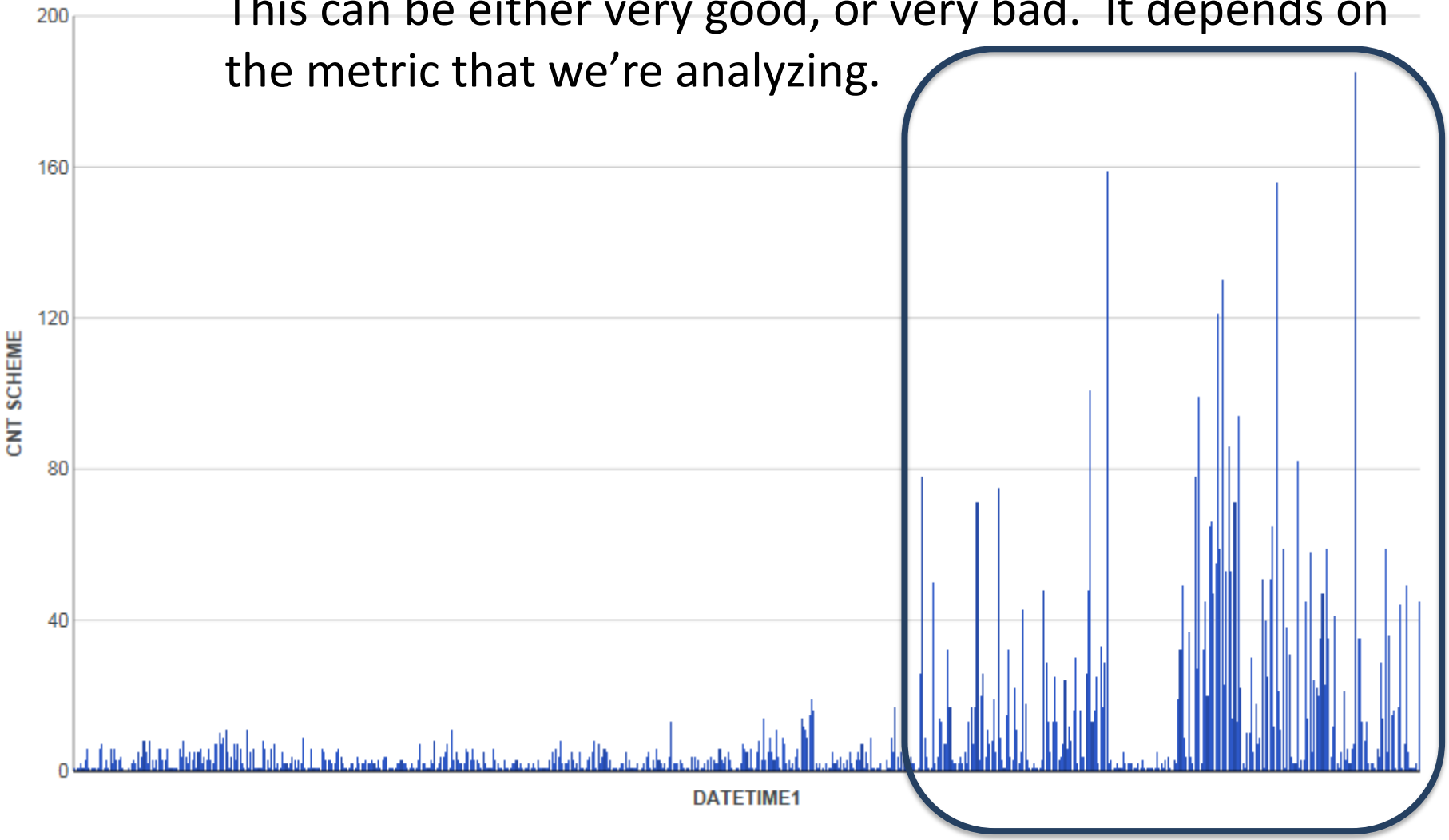


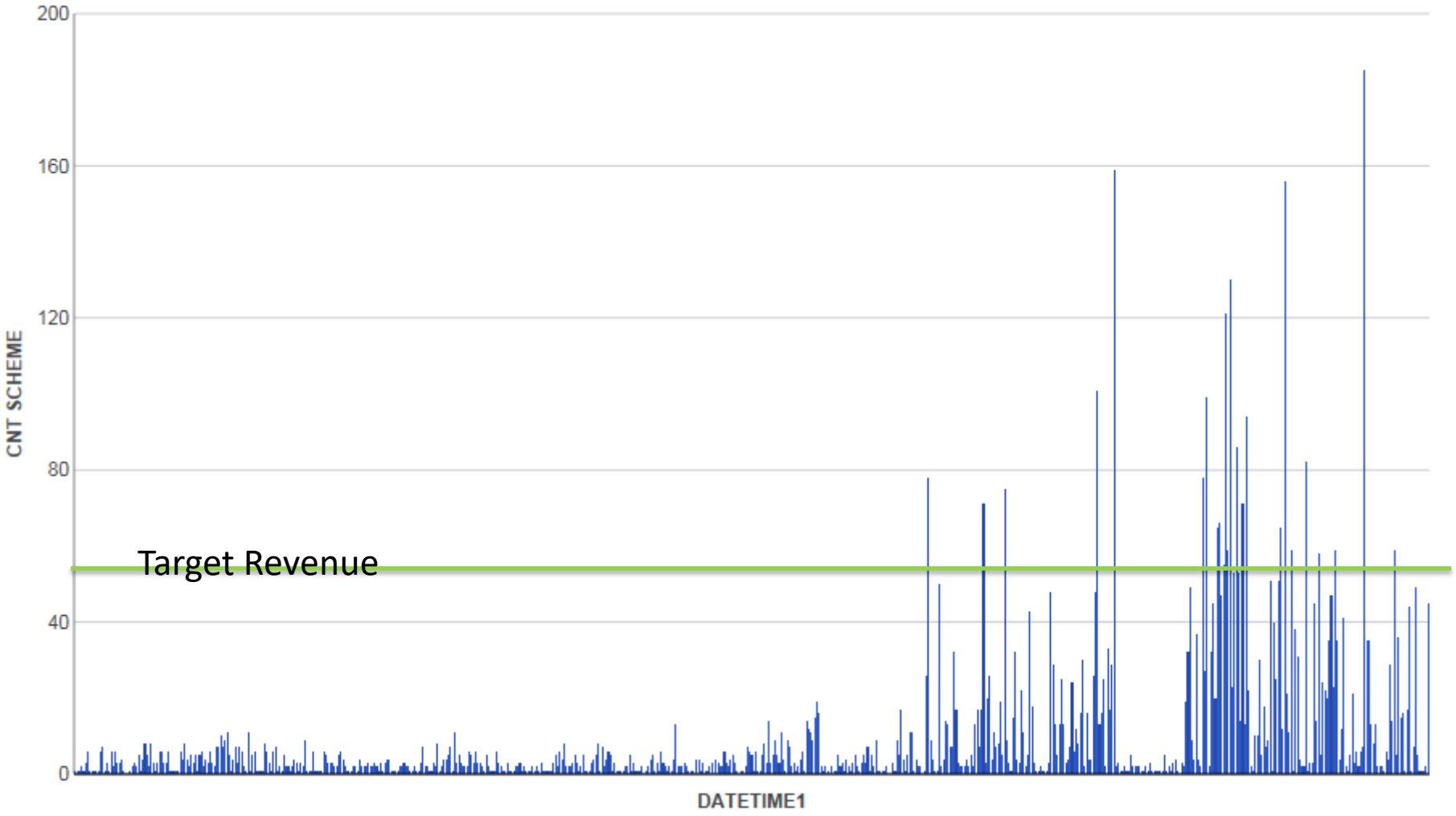
It’s not important for us to know the specific dates yet.

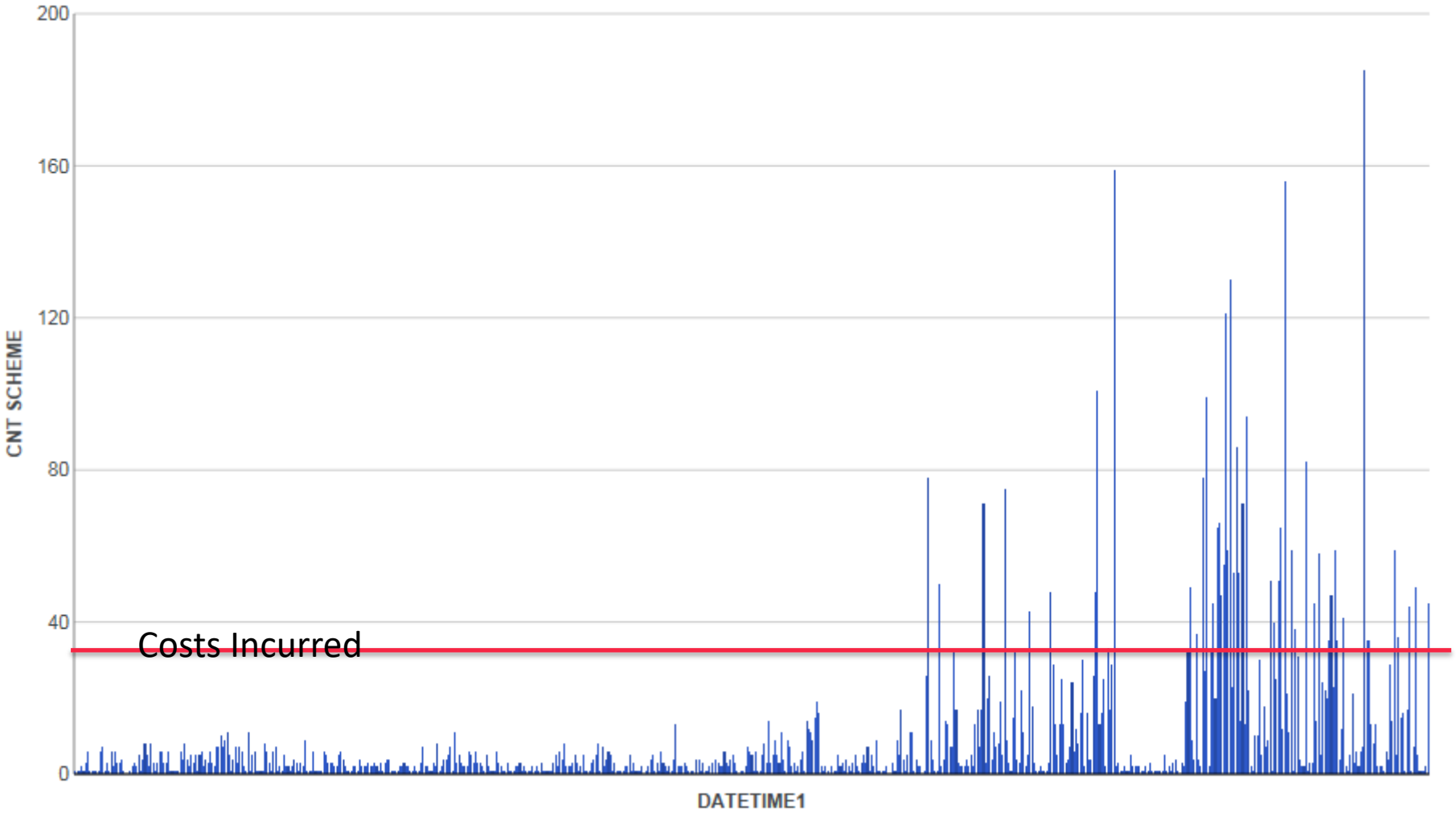


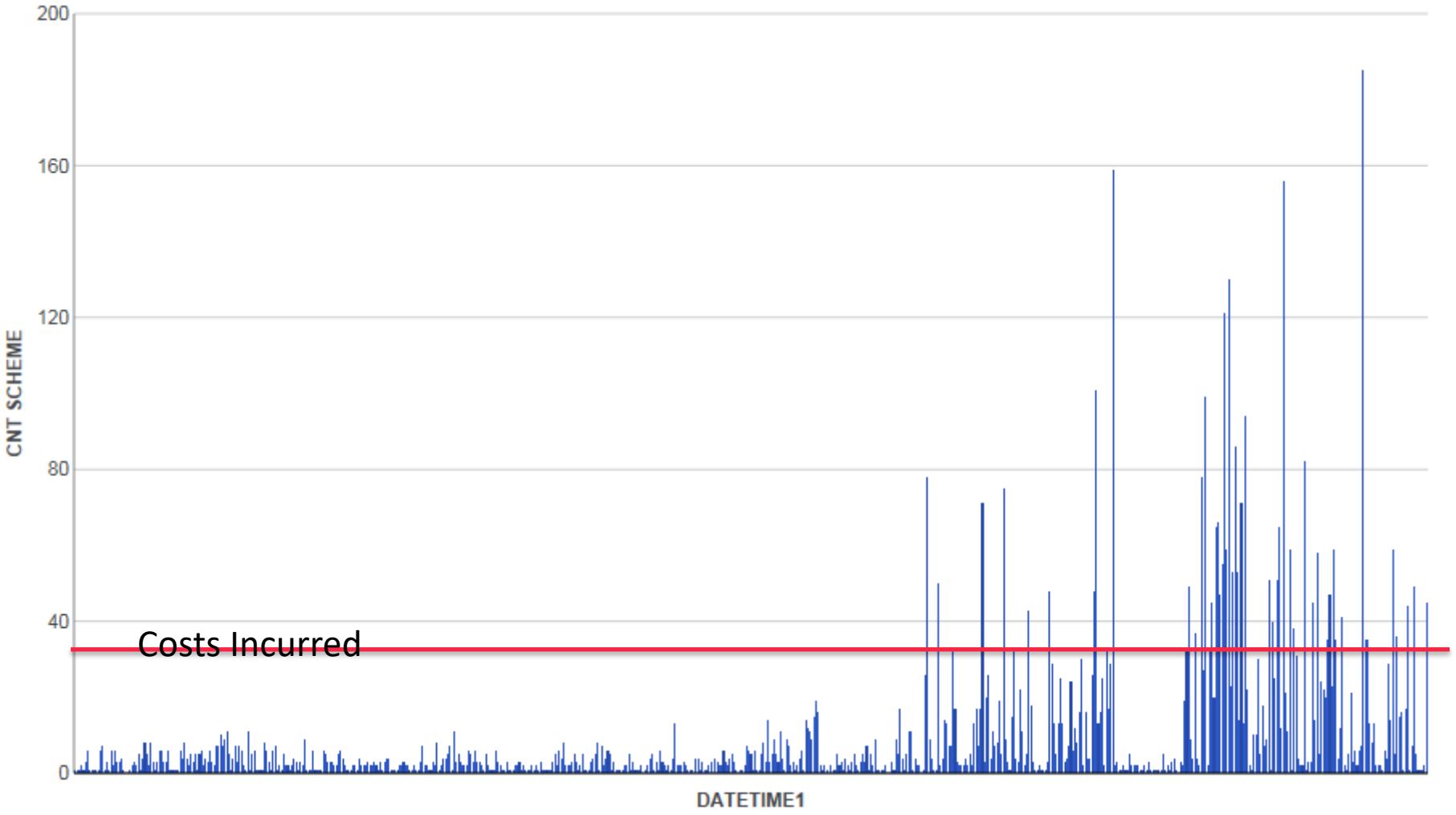
This is important because of the trend it shows.

This can be either very good, or very bad. It depends on the metric that we're analyzing.




200
160
120
80
40
0
CNT SCHEME
Target Revenue
DATETIME1


200
160
120
80
40
0
CNT SCHEME
Costs Incurred
DATETIME1



Maybe this is a yearly occurrence.

This is why “CONTEXT” and “INTENT” are so important to us.

Your
turn.
Deeper
Dive.

You are finished!