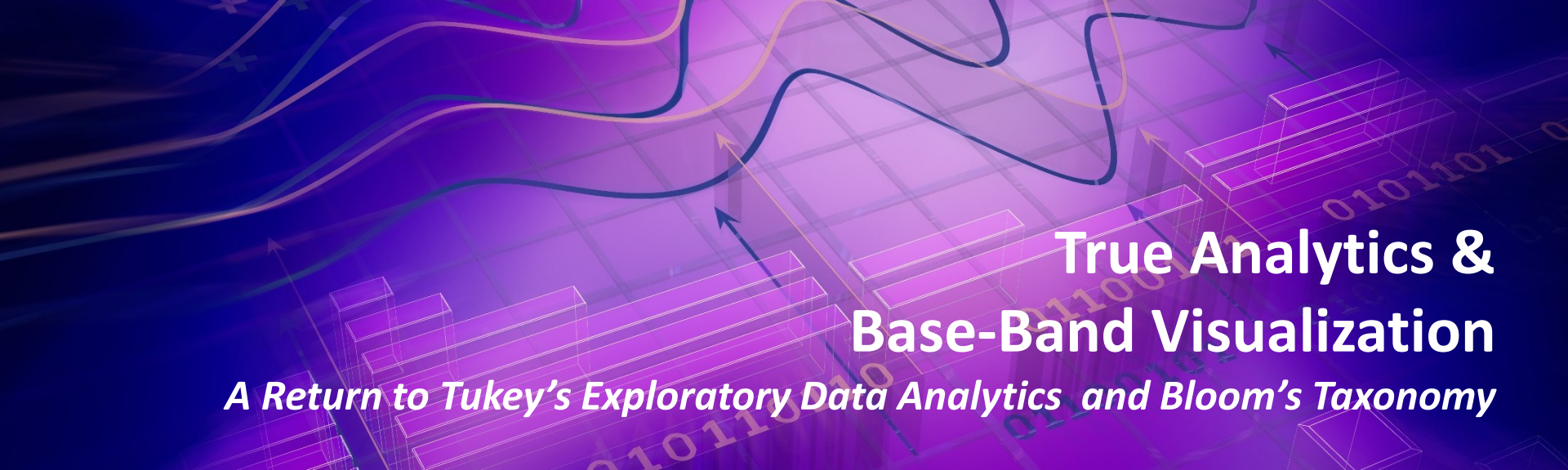




True Analytics & Base-Band Visualization

A Return to Tukey's Exploratory Data Analytics and Bloom's Taxonomy

Overwhelmed with the analytics of all that data?

**Why YOU must reset the lost art of true analytics
and lead back to leveraging data in its basic form...**

By James P. LaRue

AAS Instrument Electronics

BA Mathematics and BA in Education

MS Mathematics

PhD Applied Science and Engineering

Signal Processor and Data Scientist by Profession

Outline

Introducing YOUR Eco-System

A hierarchical sales format (with Bloom intro)

Where does Tukeys EDA enter Bloom's Taxonomy ?

It may surprise you...

A formal business and technology problem statement

A sonobuoy big data example (it is equivalent to streaming IP)

What do we mean by base-band visualization?

We're not talking pie charts, but practical and meaningful pixel arrays

Finding pattern within plasticity of 1s and 0s

Revisit the business/tech problem, plus a Model/Simulation example

The advantage to actually increasing the number of data points

A table based problem in Excel

Returning to YOUR Eco-System

Edureka: Pause for educational advertisement

The Charter Global strategic data analytics reset program

True analytics and the round table Eco-system



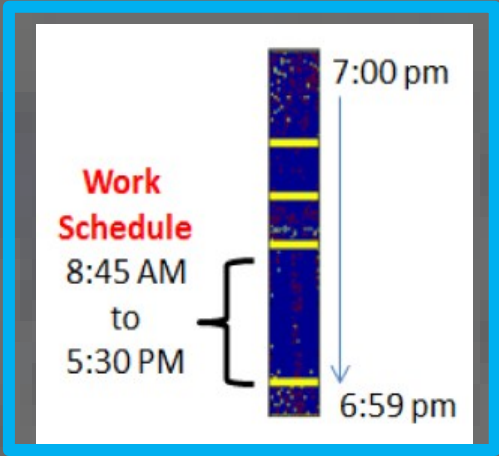
Systems Architect & Security



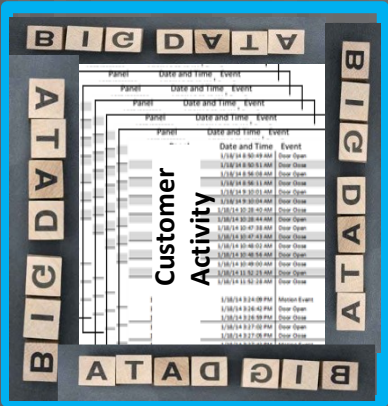
A proposed BD/BI question



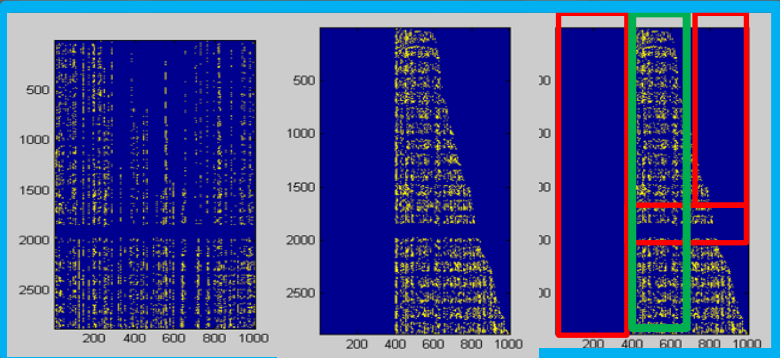
The BI/BD answer
+
ECO-derivatives



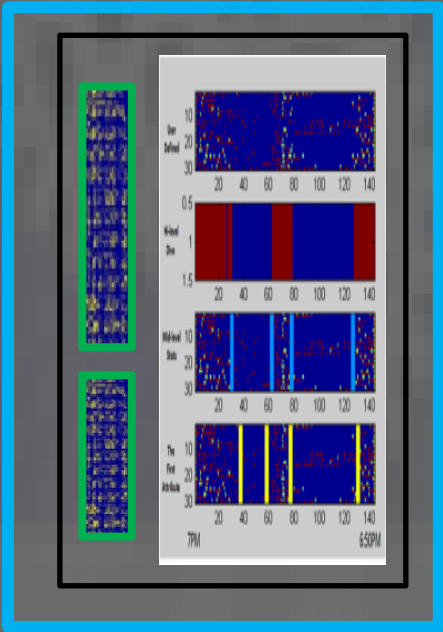
Data Source
Acquisitions and ETL



*The Eco-system of Data
requires a base-set of
thought provoking visualizations
to initiate round-table discussions
to drive cross-table observations
to empower team consensus
to draw-out winning derivatives*



Data QA-Post ETL/Pre Model



Segment Extract and Model

Legacy Data Systems & New Big Data Systems

Before:

Business Intelligence & Data Warehousing

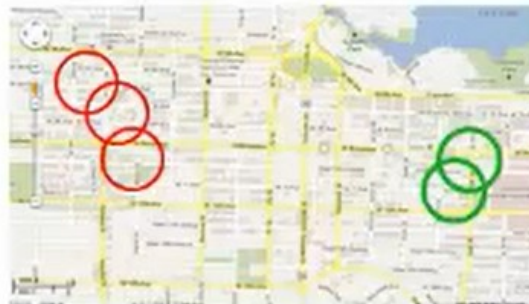
- Historical Viewpoints
- Pre-existing Reports
- Fixed-cost investments



Today:

BigData: Fast, Good & Real Time

BigData Analytics



Your Own Data:

- Unstructured
- Semi-structured
- Web store-front

Sensor Data:

- Unstructured
- Semi-structured
- High Volume
- High Variety
- High Velocity



Legacy BI and
DW Systems



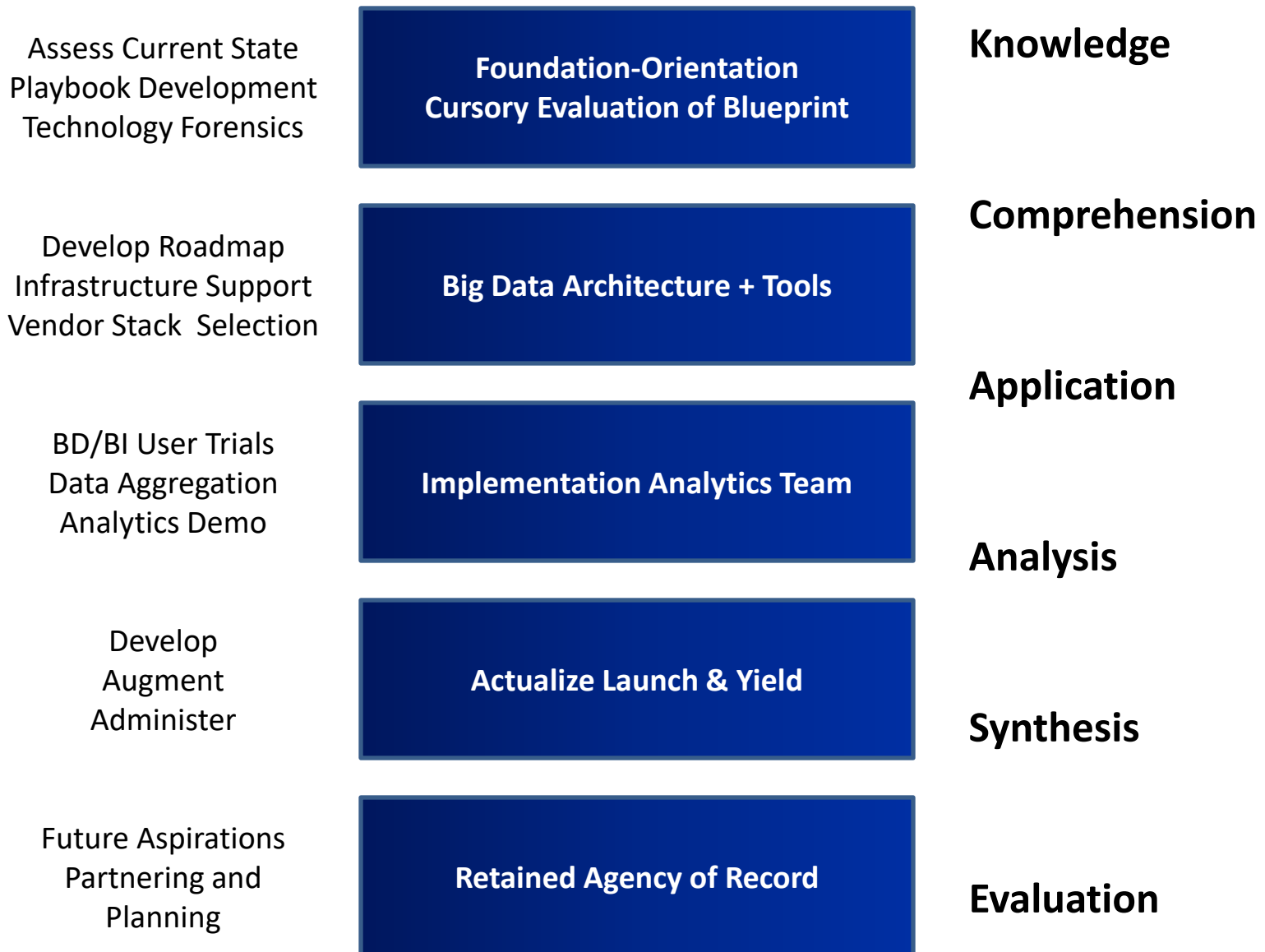
Big Data
Real-time
Analytics



Data Driven &
Rapid Response
Organization



Hierarchical Sales Format & Bloom's Taxonomy of 1956



Bloom's Taxonomy & the Cognitive Domain



Tukey's Exploratory Data Analysis (EDA)

Knowledge: assembling facts and making definitions about the data

Comprehension: translate, interpret, extrapolate, organize the data

Application: solve problems using knowledge + comprehension of the data using old models

Analysis: break data into the elements, examine the pieces, generalize the data

▶ ▶ **Fact:** John Tukey introduced the term 'bit', the contraction of Binary Digit

Synthesis: partition data elements into segments and apply old models or form new models

Evaluation: present and defend what you think you **KNOW** about the data based on model

Pie chart visualizations are for conveying knowledge, comprehension and evaluation of data

Base-band visualization is for analyzing the raw-form elements of data in pixel form

Formulas are for application and reference in evaluation

Creativity lies in synthesis and applies pressure to evaluation

http://en.wikipedia.org/wiki/Bloom%27s_taxonomy/

http://en.wikipedia.org/wiki/John_Tukey

A Formal Business & Technology Solution

Business Side

Business Outcome: Oil company to address environmentalist concerns of disturbing whale habitat and feeding, breeding, and resting. X amount of Dollars available to look for solution.

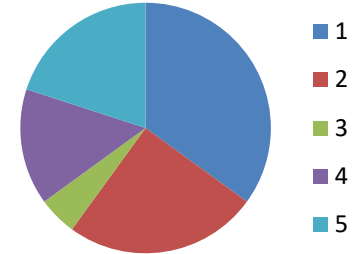
Premise 1: Underwater blasting for Seismic surveys affects habitat.

Premise 2: Whales, and other cetaceans, naturally change habitats.

Premise 3: Shipping traffic affects habitat domain.

Hypothesis to premise 1: Abrupt changes in pressure due to blasting damages the ears of the whale.

Hypothesis to premise 3: Shipping noise affects whales ability to communicate.

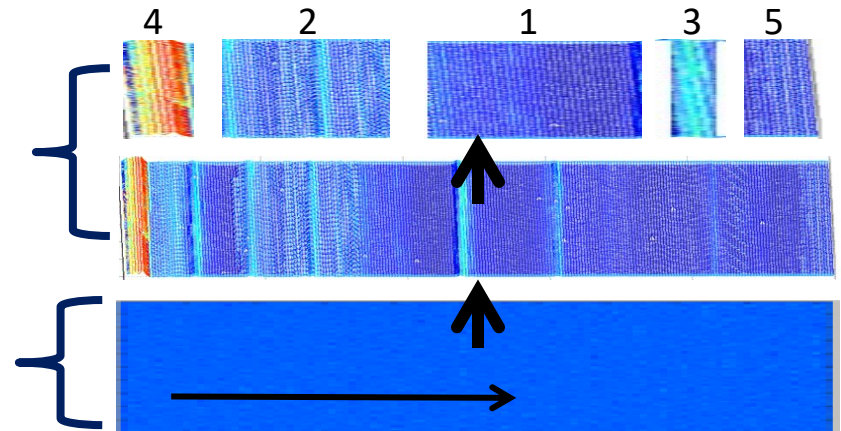


Problem Domain: How does changes in pressure link correlation between shipping traffic, seismic blasting, and whale movements?

Technology Side

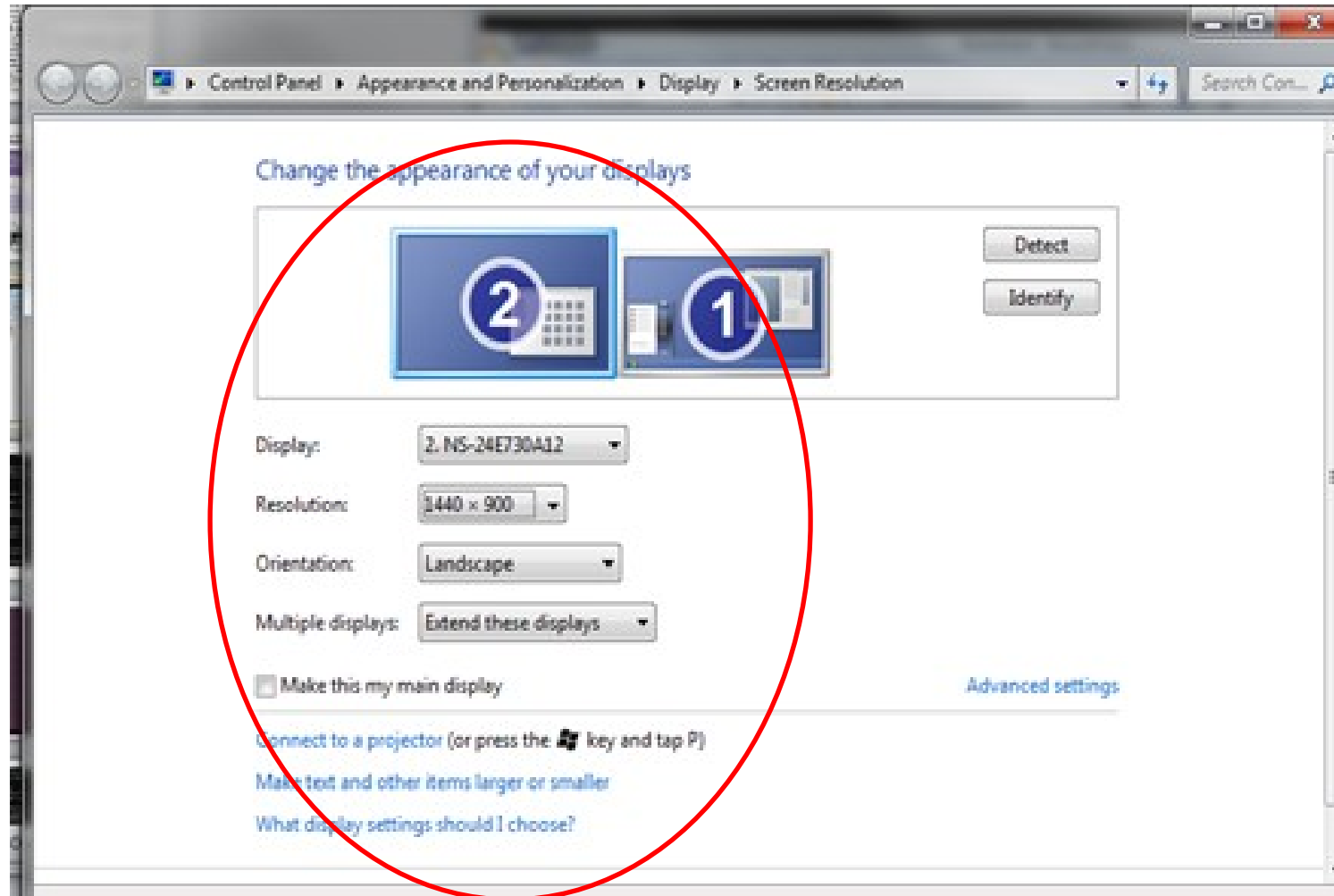
Develop Facets: Use exploitation techniques to uncover hidden attributes and then group. (K-means, higher moments, image Processing/computer vision)

Data Source: Sonobuoy recording
12000 pts/sec x 24 hrs = 1 Gpts/1 day



Base-Band Visualization Part One:

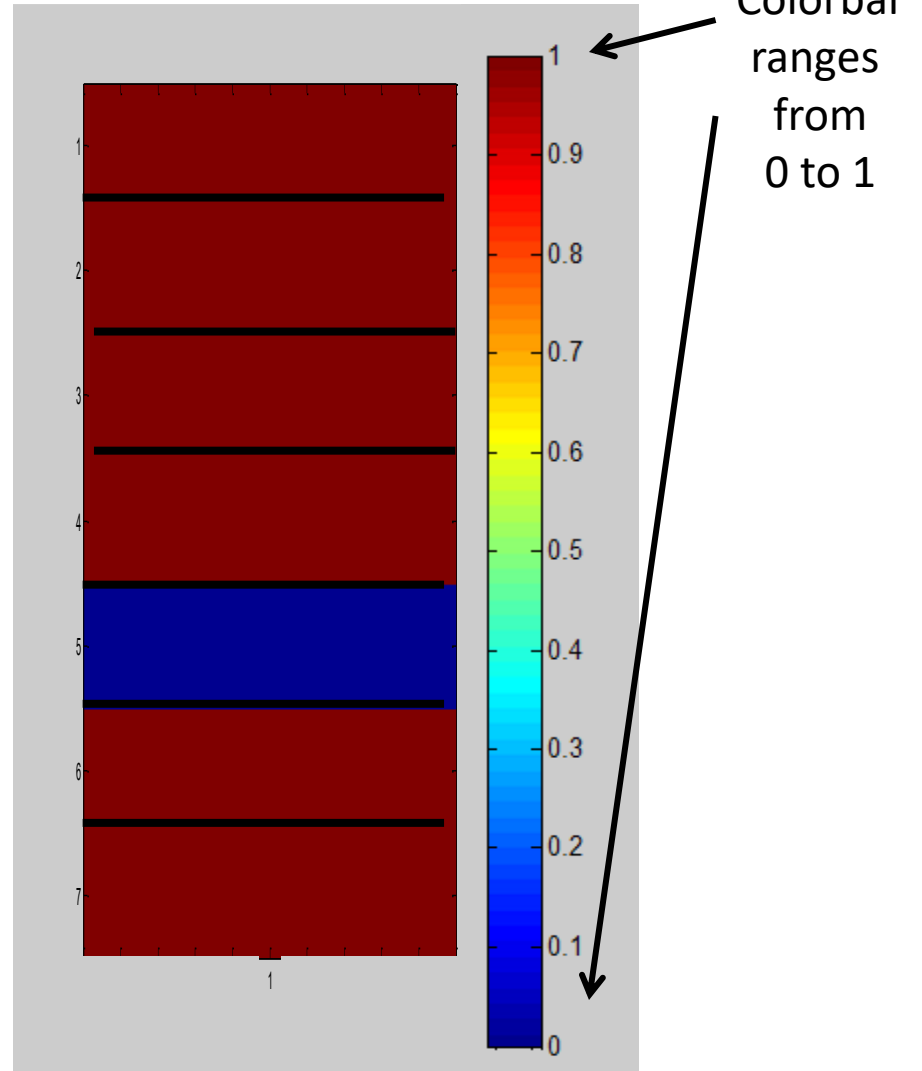
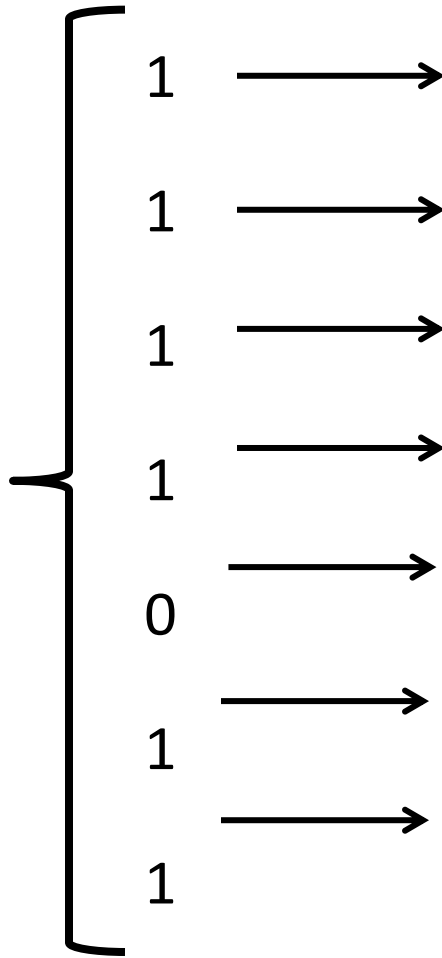
1440 x 900 pixels is a lot of pixels, so let's use them...



Base-Band Visualization Part Two:

Color the elements...

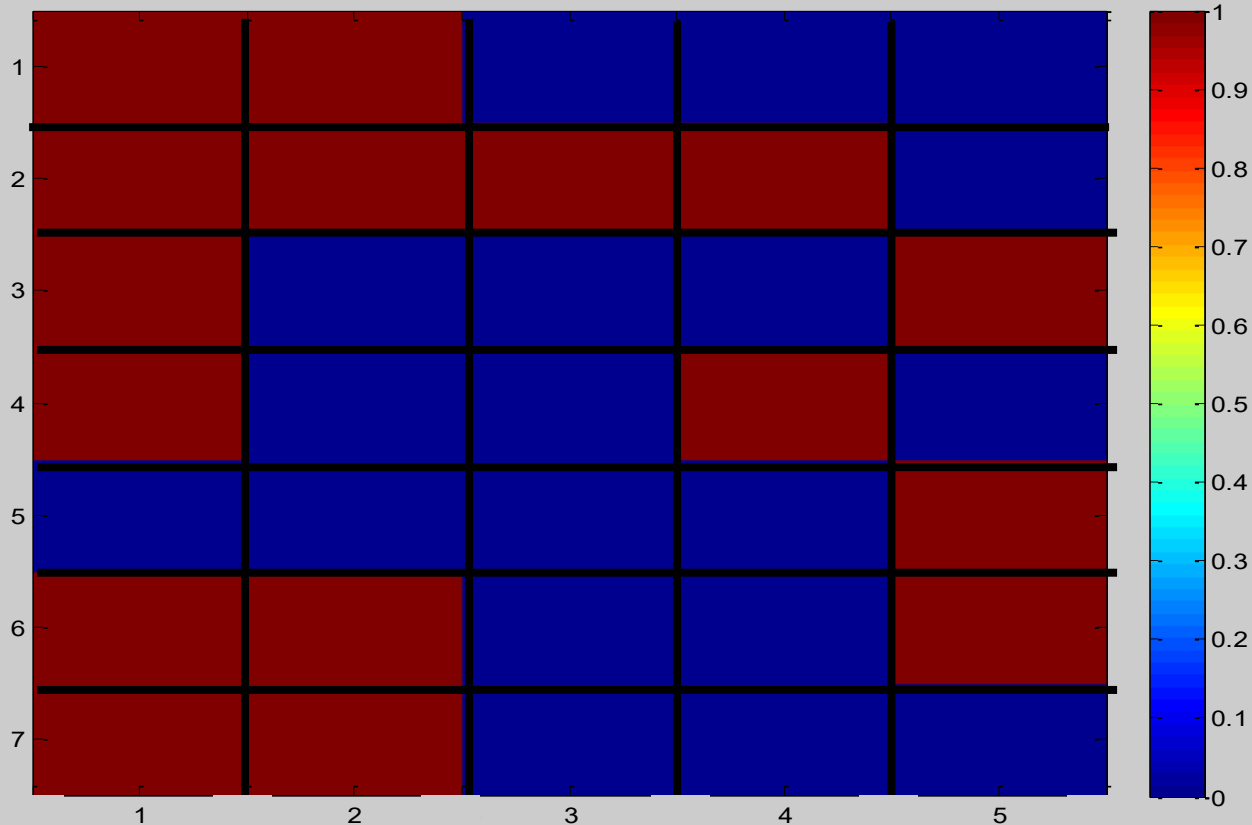
Given the
code word
elements:
1111011



A little faster now...

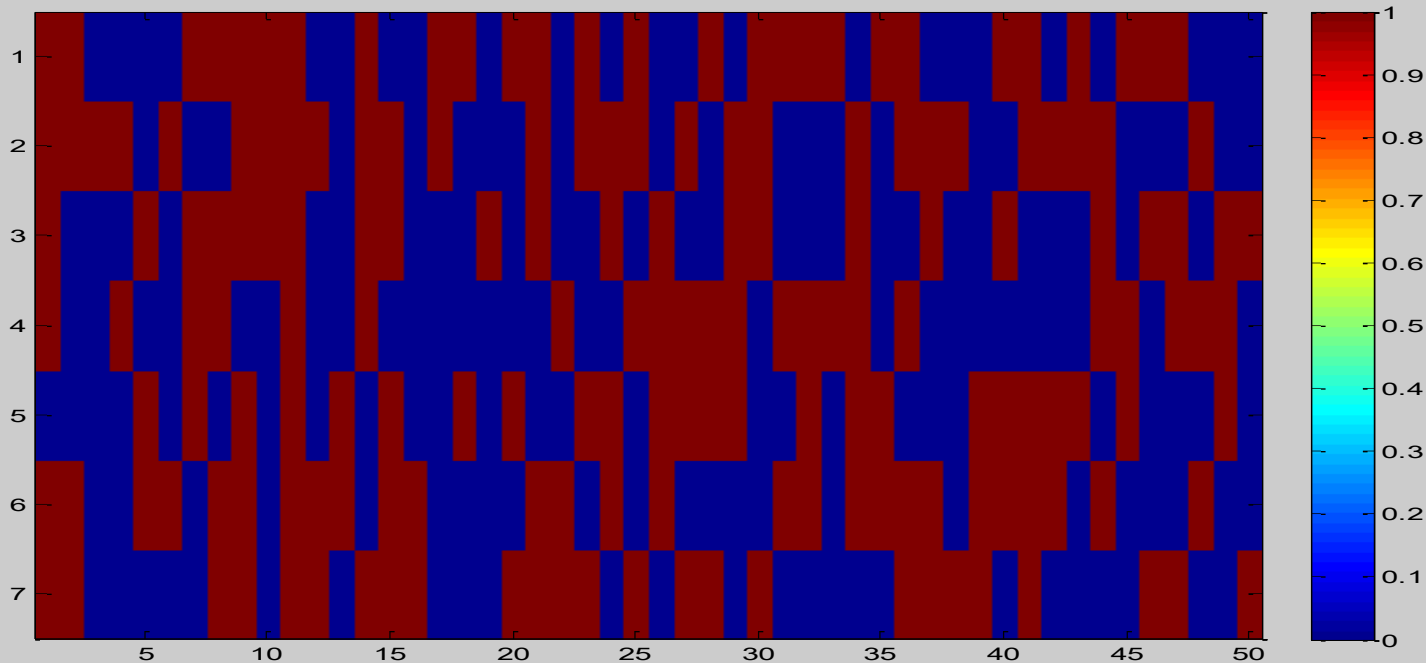
Five
Seven element
Code words
to
7x5 pixel matrix

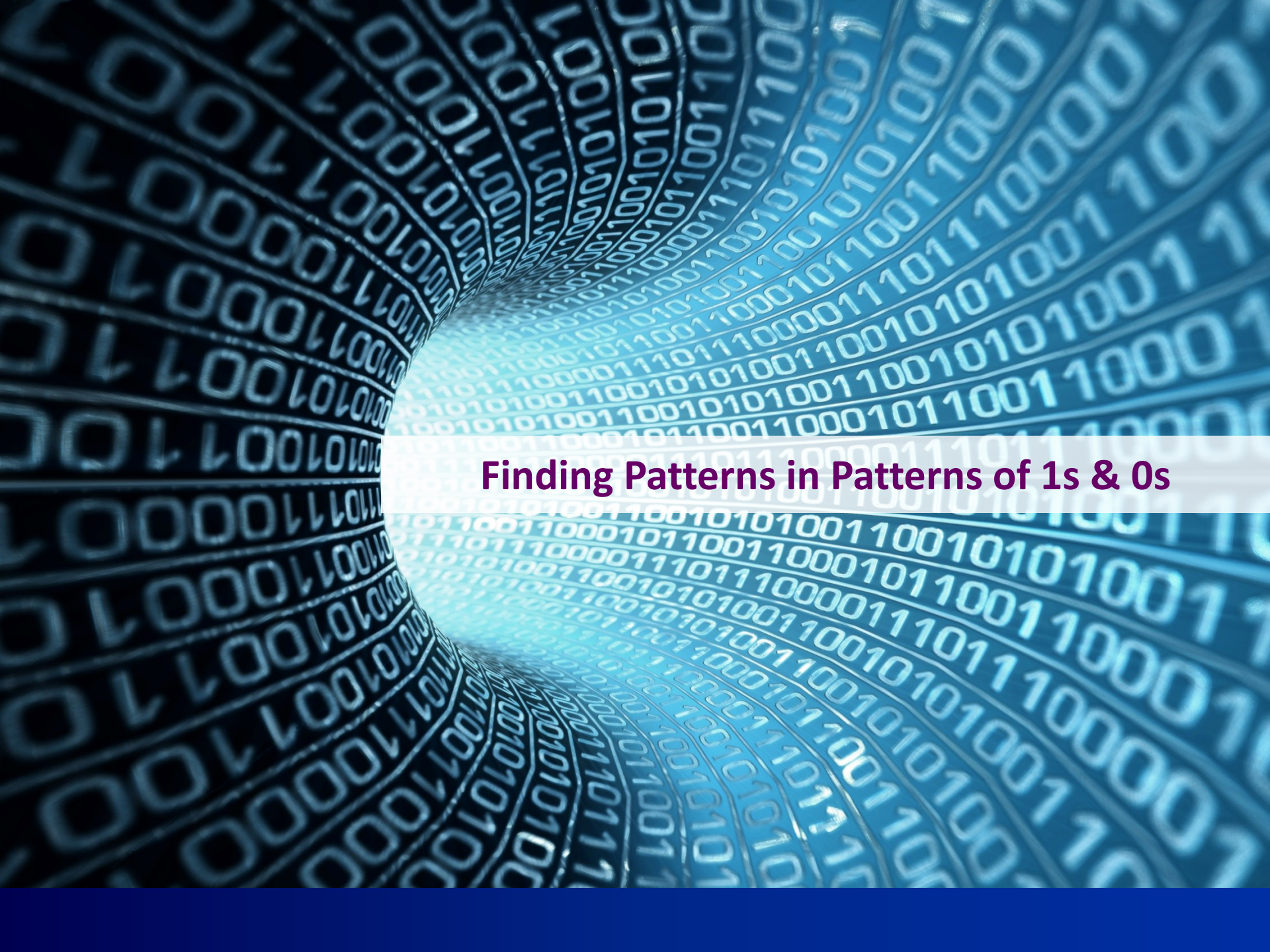
1	1	0	0	0
1	1	1	1	0
1	0	0	0	1
1	0	0	1	0
0	0	0	0	1
1	1	0	0	1
1	1	0	0	0



A 7x50 pixel matrix

```
11000011111001001101101010010111101100011010111000
11110100111101101000101110101100010111001111000100
10001011111001100010100101001100010010010001011011
10010011001001000000010011111011110100000001101110
00001010101010100101001101111001011000111110100010
11001101101110110000110101000011011110111101000100
11000001101101110001111010110100000111101000011001
```





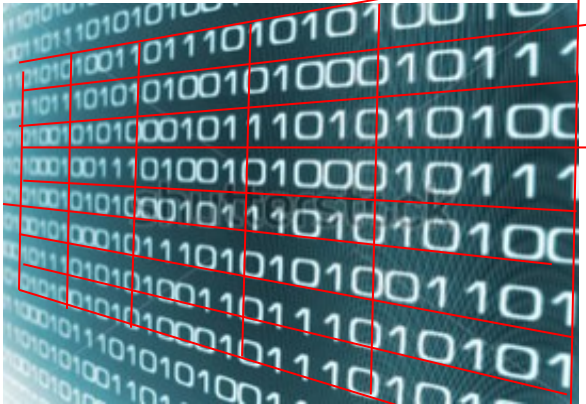
Finding Patterns in Patterns of 1s & 0s

Exercise in Pattern Digging

1



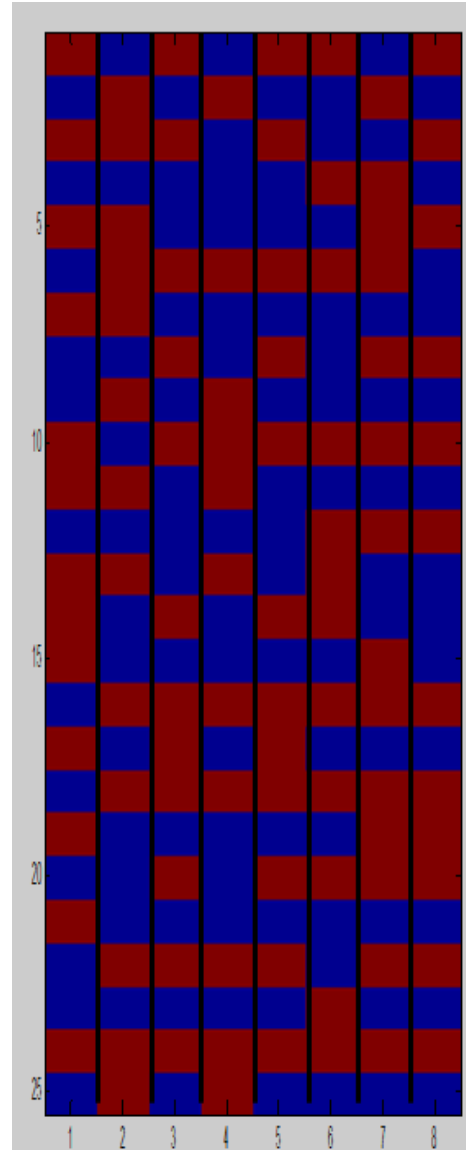
2



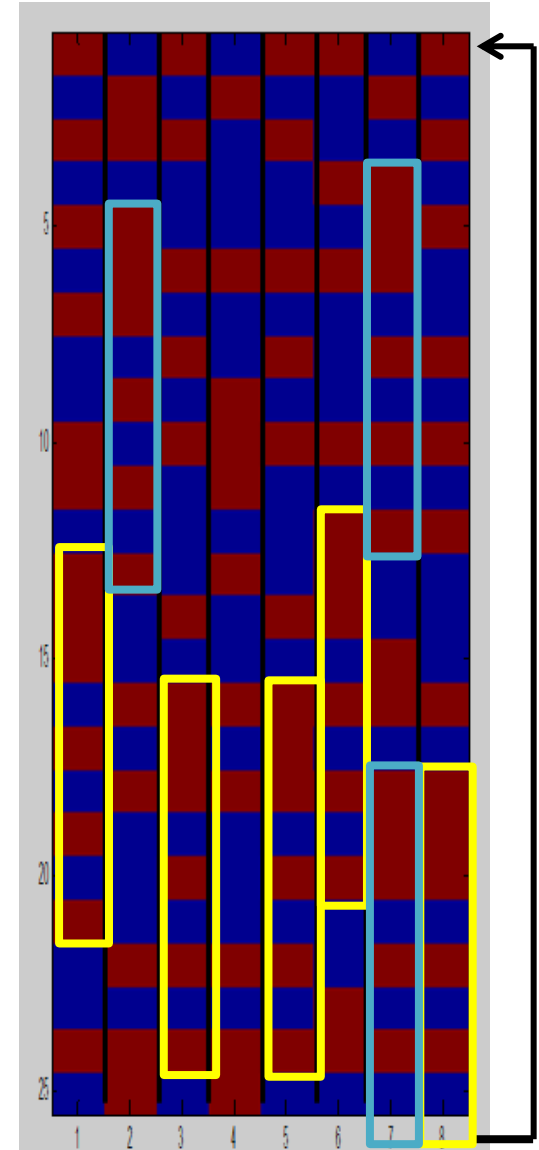
3

```
10101010011011101010100010
0110111010101001010001011
1010010101000101110101010
0100010011101001010001011
1010010101000101110101010
1001010001011101010100110
0101110101010011011101010
1010100101010001011101010
```

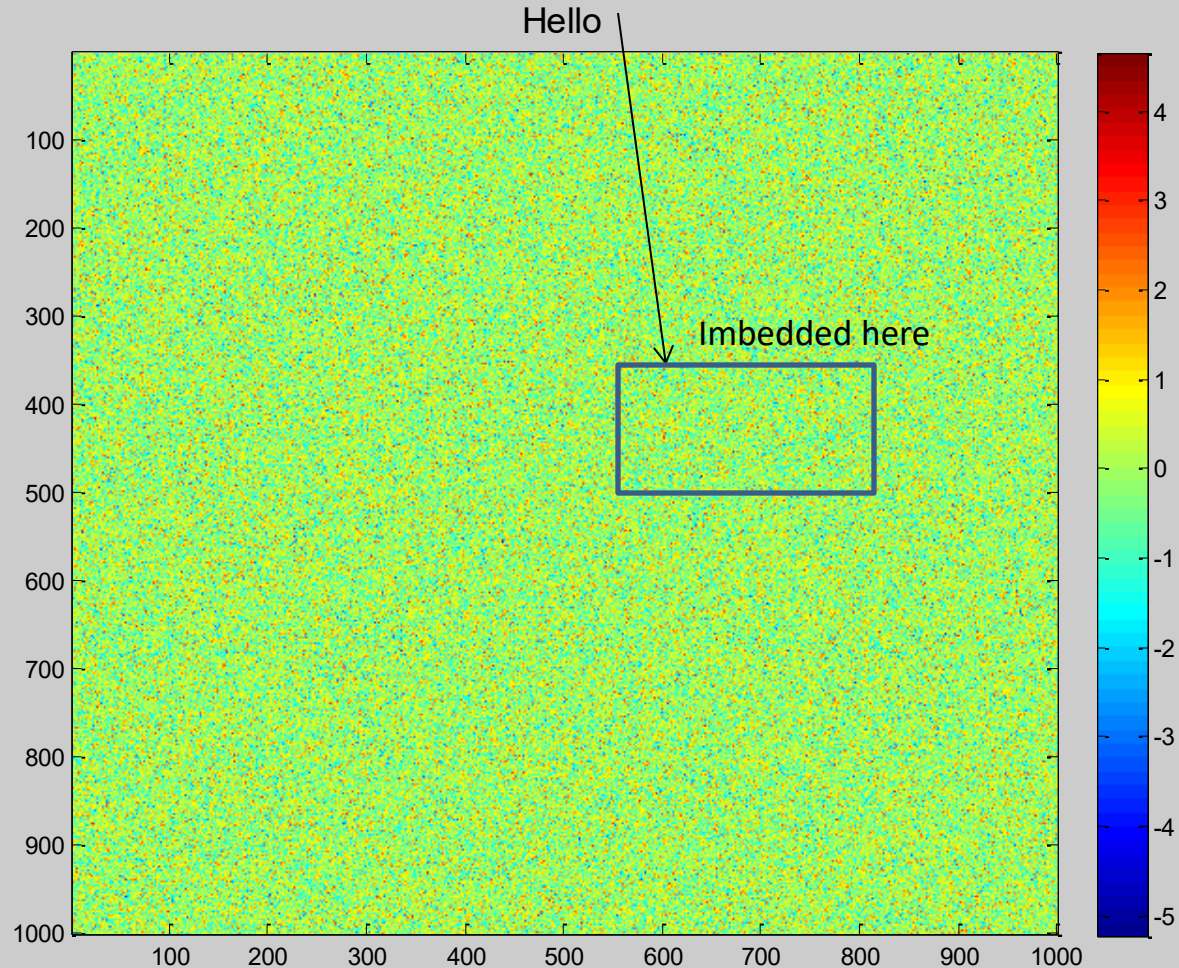
4



5



A 1000x1000 pixel matrix
1000 columns of 1000 random numbers ranging -5 to +5
1,000,000 unique colors being displayed.

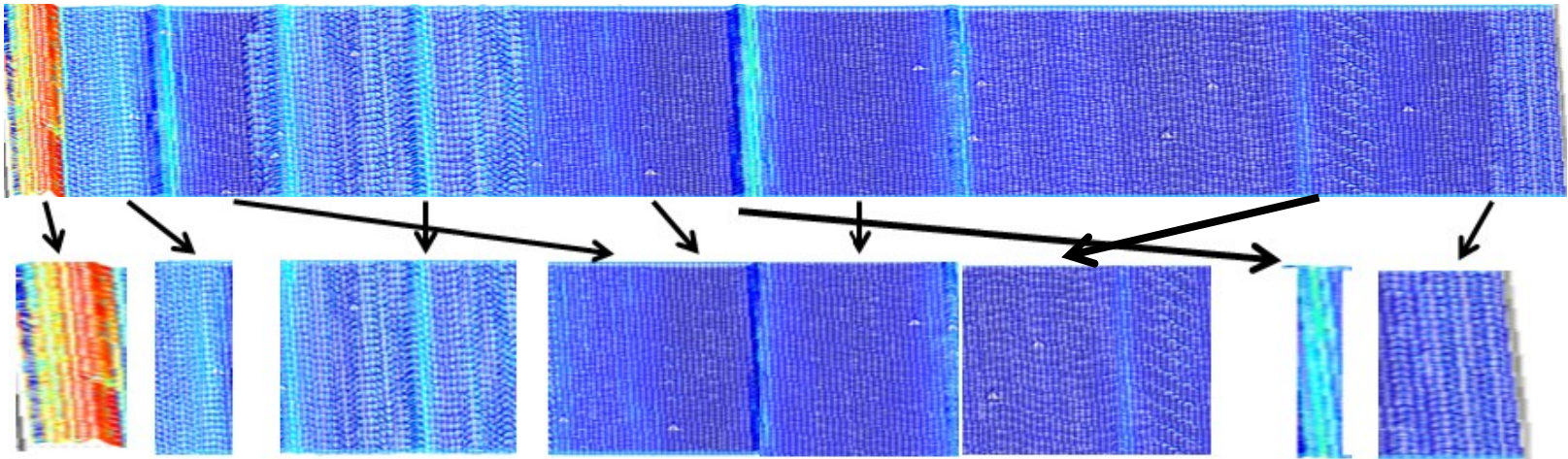


Return to the Sonobuoy Example with Tukey's EDA

We took the 1,000,000,000 acoustic sonobuoy points, transformed a little, and formed a data pool matrix of 1000 x 8000 elements. At a high level, the **information appears uniform**.

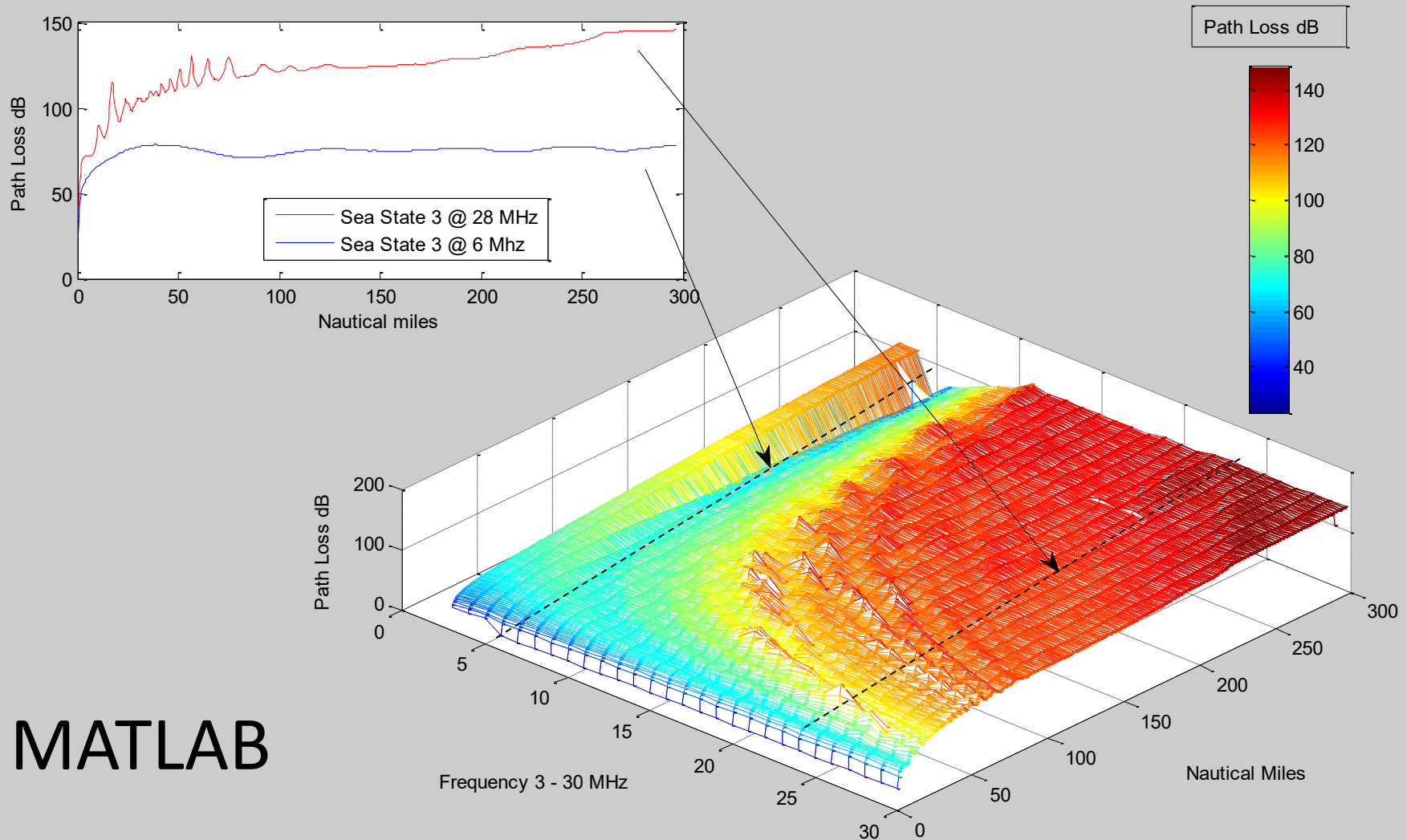


However, from the blue data pool of elements, signal processing uncovers several **underlying structures**. (buoy carrier, oil explorations, ships, storms, calm seas).



These structures form the new elements. Thus from one data source, we form several more data pools. This **segmentation** is presented to the Eco-system, to initiate round-table discussions, to drive cross-table observations, to empower team consensus.

Why look at two simple plots when you can look at 300 simultaneously ? (3-30 MHz by increments of .1)



MATLAB

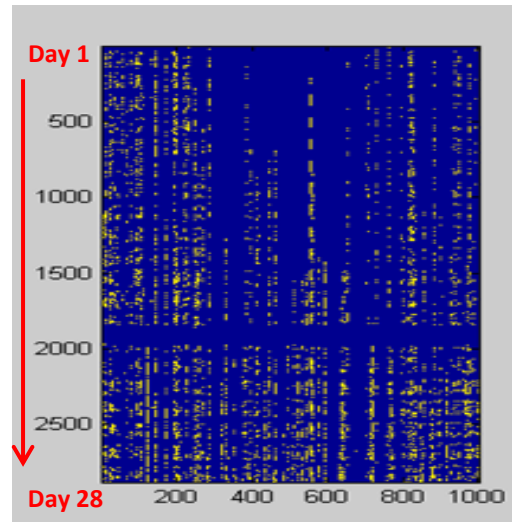
A Database Example that Moved from Row Entry to Time Domain

1000 customers were recorded for Open/Close door activity over 28 days. during the day. Activity ranged 50-750 door Open (gold)/Close (blue) total activities per customer. We expanded the table to form a uniform time scale of 100 time slots per day per home. i.e., 2800 time slots for each of the 1000 customers.

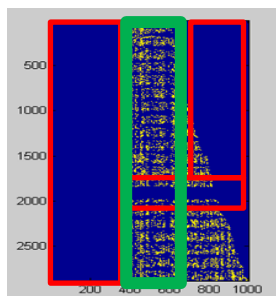
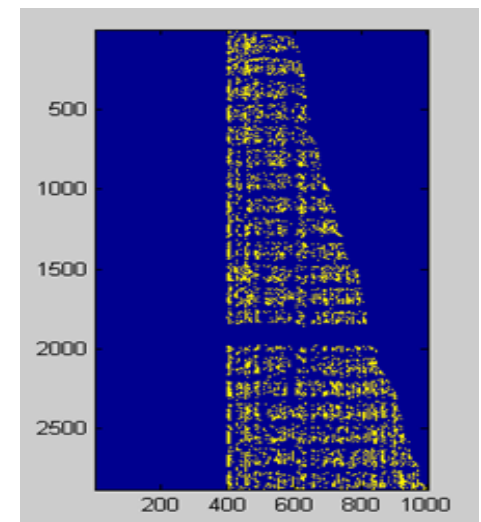
Took spreadsheet of ~78,000
lines of feature events

Customer	Date and Time	Event
102740898586	1/18/14 8:50:49 AM	Door Open
102740898586	1/18/14 8:50:51 AM	Door Close
102740898586	1/18/14 8:56:08 AM	Door Open
102740898586	1/18/14 8:56:11 AM	Door Close
102740898586	1/18/14 9:10:01 AM	Door Open
102740898586	1/18/14 9:10:04 AM	Door Close
102740898586	1/18/14 10:28:40 AM	Door Close
102740898586	1/18/14 10:28:44 AM	Door Open
102740898586	1/18/14 10:47:38 AM	Door Open
102740898586	1/18/14 10:47:43 AM	Door Close
102740898586	1/18/14 10:48:02 AM	Door Close
102740898586	1/18/14 10:48:56 AM	Door Open
102740898586	1/18/14 10:49:00 AM	Door Close
102740898586	1/18/14 11:52:25 AM	Door Open
102740898586	1/18/14 11:52:28 AM	Door Close
145677788700450	1/18/14 3:24:09 PM	Motion Event
145677788700450	1/18/14 3:26:42 PM	Door Open
145677788700450	1/18/14 3:26:59 PM	Door Close
145677788700450	1/18/14 3:27:02 PM	Door Open
145677788700450	1/18/14 3:27:05 PM	Door Close
145677788700450	1/18/14 3:27:42 PM	Motion Event
145677788700450	1/18/14 3:27:47 PM	Motion Event

Engineered time domain to
visualize as 2800x1000 matrix



Applied a cascade of
discovery transforms



Presented the 2,800,000 events in *discovery* framework to BI team

Red box: 40% of customers did not have device installed properly

Green Box: 30% had late starts

Yellow box: Data Warehouse dropped 30 hours of (paid for) recorded data

Analytics at this fundamental level is a section of QA

Base-Band Visualization of Analytics Invites a Roundtable Approach

2. ETL asks Data Warehouse
For activity on 1000 customers.
DW returns 78,000 table entries



1. BD task - work schedule

1-6: Eco-System Derivatives

Architecture/Data Storage

- DW purchase lapse

ETL

- Data Source Consistency

Modeling

- 20% valid segment

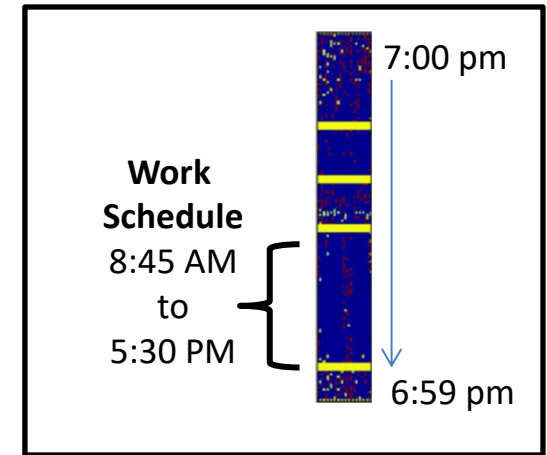
BI

- 24 Hr. Home Habits

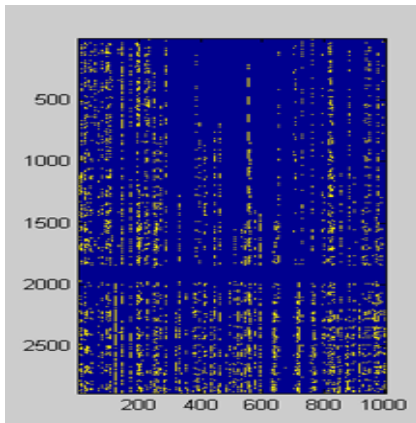
BD

- Ask Techs to check sensors

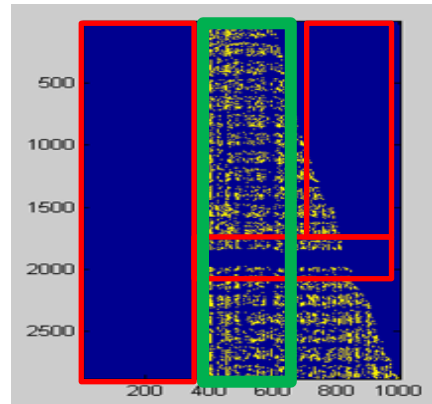
6. BD Solution



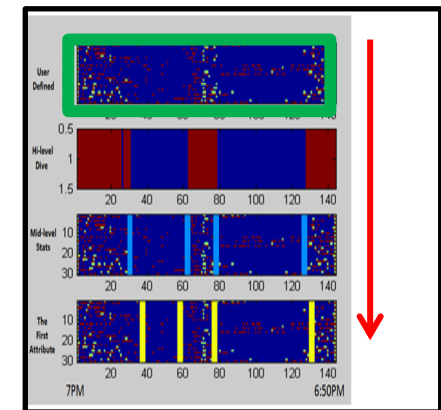
3. Engineer a structured
visualization



4. Signal Processing to see what
you have or thought you had



5. Modeling & Simulation
solution with what you have



Edureka !!

Others that are honing in on EDA and Visualization

From the Computation Institute (University of Chicago/Argonne National Labs) and AT&T Labs
<https://www.ci.uchicago.edu/blog/new-algebra-data-visualization> and
<https://www.ci.uchicago.edu/blog/new-computational-commons-cancer-genomic-data>

An **Algebraic Process for Visualization** Design by Kindlmann and Scheidegger (2014),
http://algebraicvis.net/assets/vis2014_talk_slides.pdf

Data Mining Challenges for Digital Libraries by founder of **Open Data Group, Robert Grossman**. Back in 1996 he mentions **three principle** purposes for Visual Analytics: anomaly checks, Tukey's EDA, and checking model assumptions.

[John W. Tukey](#) wrote the book "**Exploratory Data Analysis**" in 1977

From to Data Visualization Innovation Summit, April 2015, San Jose, Elijah Meeks, Senior Data Visualization Engineer at Netflix, presented, '**Beyond Line and Pie Charts: Practical Applications of Complex Data Viz**'

<https://www.codeshowse.com/> Charleston, SC May 2015, with keynote speaker Jeff Hammerbacher of Cloudera presenting his work with **Big Data and predicting the process and treatment of disease**.

Strategic Data Analytics Reset Program

True Analytics & the Roundtable Eco-System

BEFORE YOU START your investment path
(take a step back)

DEFINE THE GAME

Your Business Development Directive
(keep it purposely loose)

GET TO KNOW your BI/BD/ETL/Mod/Dev team
(collective or stove-piped)

ESTABLISH ACCESS TO your Big Data Repository
(costly and ad-hoc deck of cards)

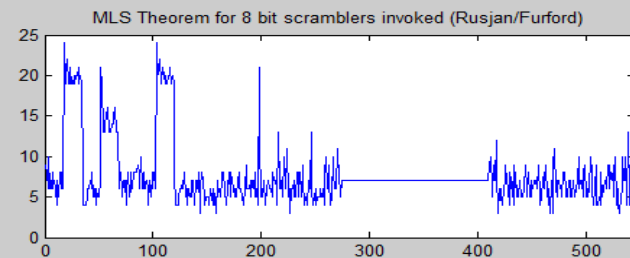
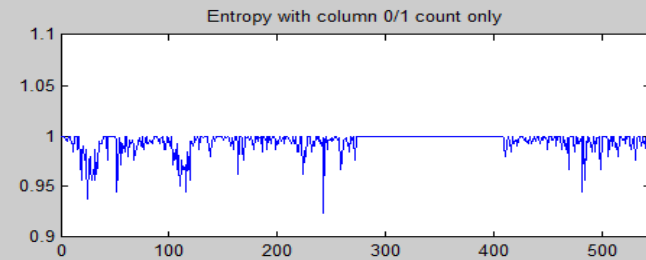
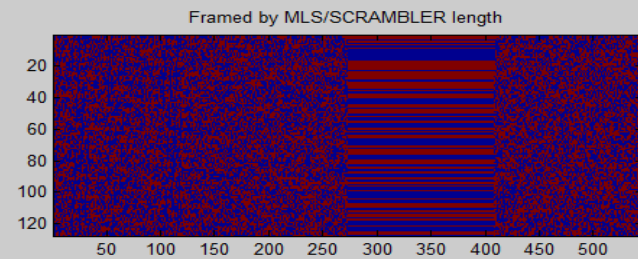
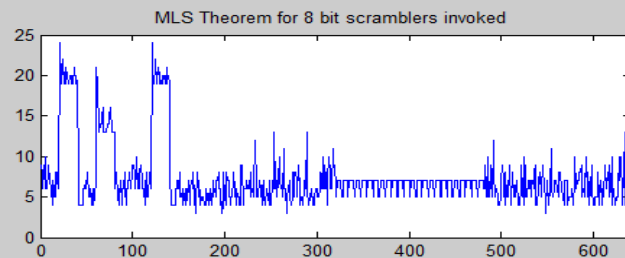
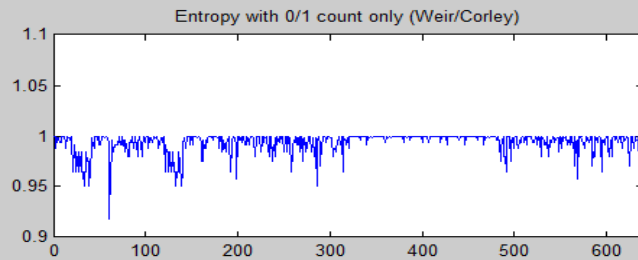
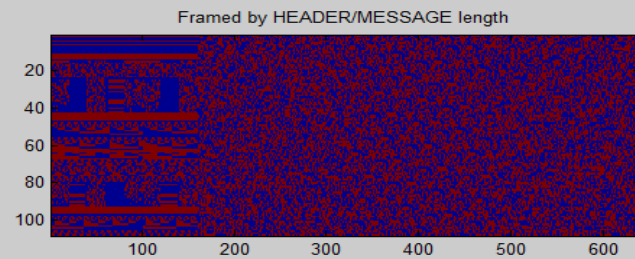
Set the odds to success

Base-band visualization (show what's in the deck)

Now, call in your players and...
STAND BACK AND LEAD



It all depends how you frame the bits,
“now you see it, now you don’t” vs “Now I don’t see it, now I do”



In bin 198 synch was lost not to noise but to the SPARSE EVENT that the scrambler collided with the message