Value in the DETAILS

Understanding detailed data through VISUAL EXPLORATION

Richard Brath
Rob Harper
I’ll be visually exploring all kinds of interesting patterns in tweets about Trump.

But wait – first what do we mean by visual exploration?
Visual Exploration

Natural Selection?
Visual Exploration Process

1. Collect a lot of data
2. Observe some interesting patterns
3. Hypothesize about why they exist
4. Refine and build models
Understanding detail data

Can you see the patterns?
How about these algorithms abilities to find the patterns?

No Clustering  Linkage Clustering  K-Means Clustering  Distribution-based Clustering  Density-based Clustering
We see patterns all the time, and quite easily

We tend to group things based on visual cues, such as proximity, alignment and containment.
Seeing detailed patterns

Perception can be whole. Once you see it, you won’t un-see it.

First publication of the picture probably in Life Magazine:58;7 1965-02-19, p 120.
Also, check out the movie by Wim van de Grind (http://www.michaelbach.de/ot/cog_dalmatian/index.html)
So what?

Powerful human perception system

- Detects **patterns** in complex data
- Can find **patterns** based on different criteria
- Can find **patterns** at different scales
So what?

Exploratory Data Analysis stems from John Tukey’s work in the early 1960s. EDA can be characterized by
a. understanding "what is going on here?"
b. graphic representations of data
c. tentative model building and hypothesis generation
d. robust measures, re-expression, and subset analysis
e. skepticism, flexibility.

The goal of Exploratory Data Analysis is to discover **patterns** in data.

paraphrased from John Behrens, Arizona State University, Principles and Procedures of Exploratory Data Analysis, American Psychological Association, 1997.
So, why are we summarizing big data into bar charts?
So, why we rolling-up big data into visualizations of 1000 points?

A Tour through the Visualization Zoo
Heer, Bostock, Ogievetsky
http://homes.cs.washington.edu/~jheer/files/zoo/
So, how do we visualize 100m data points?
Plot all the detail by connection
Plot all the detail by order
Plot all the detail by time
Visual Patterns in Time Plots

- time
- some other variable
- event
- boundary
- periodic
- level (threshold)
- anomaly
- trend
Bitcoin Transactions

Load & Verify

Bitcoin Address
14sScGvGttxNbfqFuStoXN7EYxdaN1JML

Bitcoin Amount:

Strength in Numbers

Private Key
5J9nozYs1GnpIap6DeUHarmsTHC68fYtwes2pHGePcxSySpInL

Spend

© 2015 Uncharted Software Inc.
Every single bitcoin transaction
Financial Markets
Stock Market: One Day Trades on 6800 stocks

ARCA stocks ranked by percent change on day

© 2015 Uncharted Software Inc.
Bank of America, first ½ hour, every trade, every order event
Bank of America, first 1/2 hour, every order event

Pre-market

Sell Orders (red)
All the preset limit orders come online at 9:30

After open the price holds tight at $12.06.

Then drops two cents.

When the price does hit $12, there is a lot of order activity.

There are a lot of orders activity at at $12.00. This indicates strong support.

The next level forms around $11.95

And then $11.92

This pattern is fairly consistent order activity that looks like it's always 25 cents away from the trades.
Vertical axis normalized by last trade price
Bank of America, first $\frac{1}{2}$ hour, last price line
Technical Approach
Exploratory Big Data Analysis

MAP

CONNECT

ORDER

TIME
Exploratory Big Data Analysis – Rich Interactions

LAYERS

ZOOM

MINE

FILTER
Spark as a Service

Expressive API
Technical Approach

Data Pipeline
- Loading
- Filtering
- Transformation
- Sentiment
- Serialization

Results Generation
- Input mapping
- Projection
- “Pixel”-level analytics
- Area-level analytics
- Dataset-level analytics

Spark Cluster as a Service
Technical Approach

- configuration
- density
- distributions
- transformations
- term frequencies
- summary statistics
- filters
- data drilldown
- sentiment
- density
- distributions
- transformations
- term frequencies
- summary statistics
- filters
- data drilldown

Spark Cluster as a Service

API
- Executor
- Executor
- Executor
- Executor
- Executor

Spark

Executor

RabbitMQ

Shared Storage

Redis

tiling.jar
Technical Approach

github.com/unchartedsoftware
Trump Tweets

Donald J. Trump @realDonaldTrump

#AskTrump Getting ready to answer your questions.
1:07 PM - 21 Sep 2015

1,037  2,429
Tweets per day: #makeamericagreatagain, #trump2016, and #donaldtrump
August 28th — September 27th
Trump Mountain Plot – Cohort Patterns

4000 coordinated accounts
Trump Top Hashtags

Hashtags associated with US Presidential Candidate Donald Trump, showing their popularity over time.
Trump Top Hashtags

cristianoronaldoid
american
headlineswewouldlove
historyinthemaking
rare honored tshirt
design tcot pijn

megynkelly
sheskindahotvma
npr jorgeramos
msnbc worthitvma
cnn almaliki
foxnews
dumptrump
thehomeofhiphop
worthitvma
JORKER AMOS
tcot iowa univision
tfm video urmarketeam
cnn jorgeramos
kkk tcot iowa
foxnews rare

gopdebate
megynkelly
sheskindahotvma
npr jorgeramos
msnbc worthitvma
cnn almaliki
foxnews
dumptrump
thehomeofhiphop
worthitvma
JORKER AMOS
tcot iowa univision
tfm video urmarketeam
cnn jorgeramos
kkk tcot iowa
foxnews rare
Trump Mountain Plot with Sentiment
Trump Mountain Plot with Sentiment – Megyn / Jorge
Trump Retweets
@Ashton5SOS: Not sure about this whole Donald trump thing

@ruinedpicnic: Donald trump looks like the villain in a movie where the hero is a dog

@jasonmustian: I think Trump just did all the emoji faces in 7 seconds

@DooleyFunnyAF: Can't believe Donald Trump winning the election
@DooleyFunnyAF
Can't believe Donald Trump winning the election
Trump Retweets and Favorites

@DooleyFunnyAF
Can’t believe Donald Trump winning the election
Trump Retweets and Favorites

@ruinedpicnic
Donald trump looks like the villain in a movie where the hero is a dog

@DooleyFunnyAF
Can't believe Donald Trump winning the election
KEY TAKE AWAYS

1. Plot all the data

MAP

CONNECT

ORDER

TIME

2. Explore it

LAYERS

ZOOM

MINE

FILTER
More Info

Richard Brath
rbrath@uncharted.software
416-203-3003 x 242

Robert Harper
rharper@uncharted.software
@rdhharper