acumen

insight

ideas

attention

reach

expertise

depth

agility

talent

In The Trenches: Computer Forensics and Data Mining

John Mallery
Managing Consultant
BKD, LLP
816.221.6300

experience BKD LLP
CPAs & Advisors

Agenda

- Describe my perspective
- Talk about cell phones
- New stuff I'm seeing
- Data Mining
- Lot's of lively discussion

acumen
insight
ideas
attention
reach
expertise
depth
agility
talent

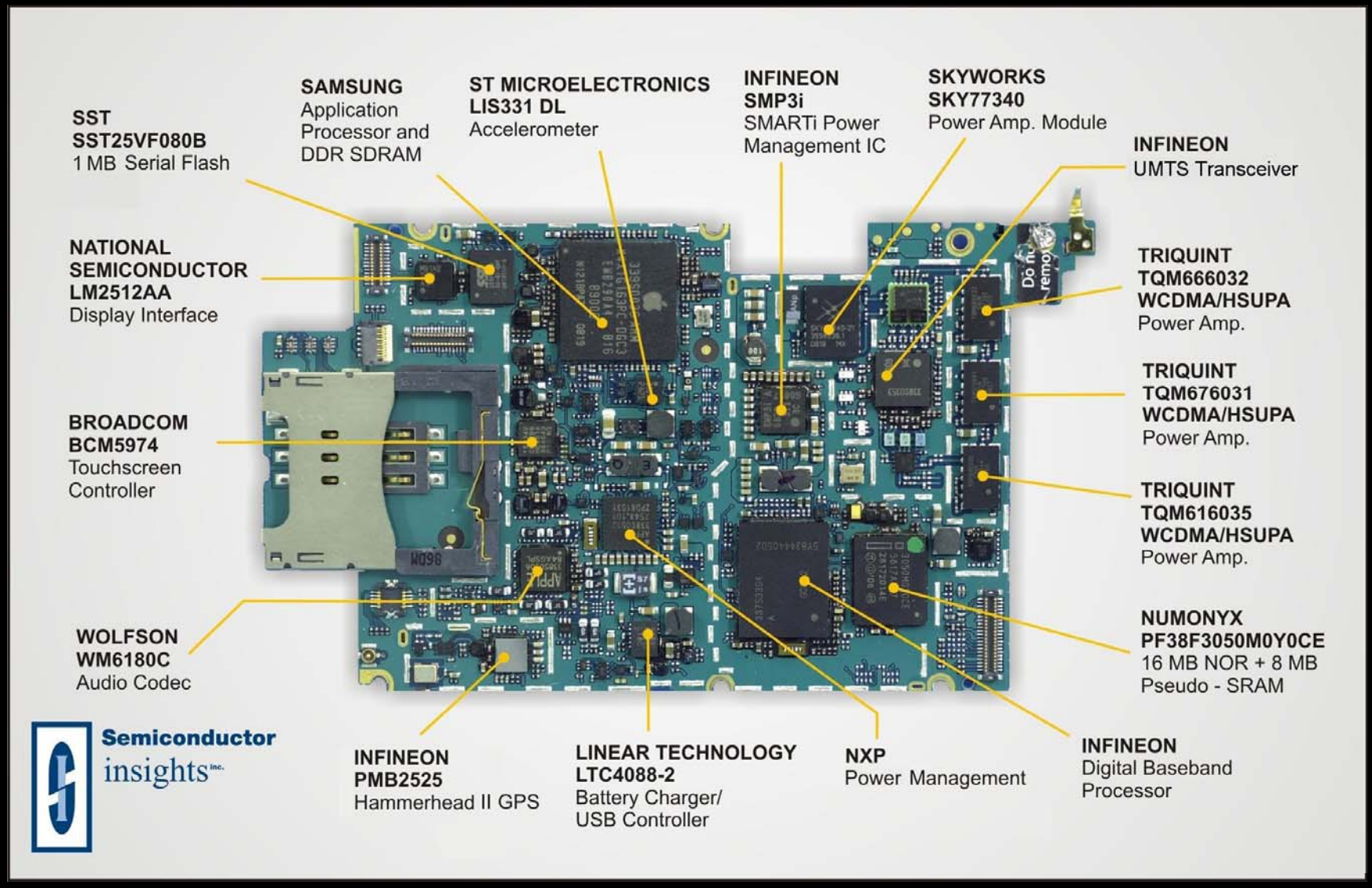
Cell Phone Forensics

- We are seeing more and more requests for cell phone analysis.
- Problem – no standardization, so it is nearly impossible to keep up with cables and tools
- No one tool does it all.

Cell Phone Forensics

- But, backups can be recovered from the computers they sync to.

acumen
insight
ideas
attention
reach
expertise
depth
agility
talent



http://www.rapidrepair.com/guides/iphone3g/iphone3grepairguide.html

However…

- iPhone Backups are created every time the phone is synced
- Windows – C:\Documents & Settings\USER\Application Data\Apple Computer\MobileSync\ Backup
- Mac ~/Library/Application Support/MobileSync/Backup/ “hex folder name”

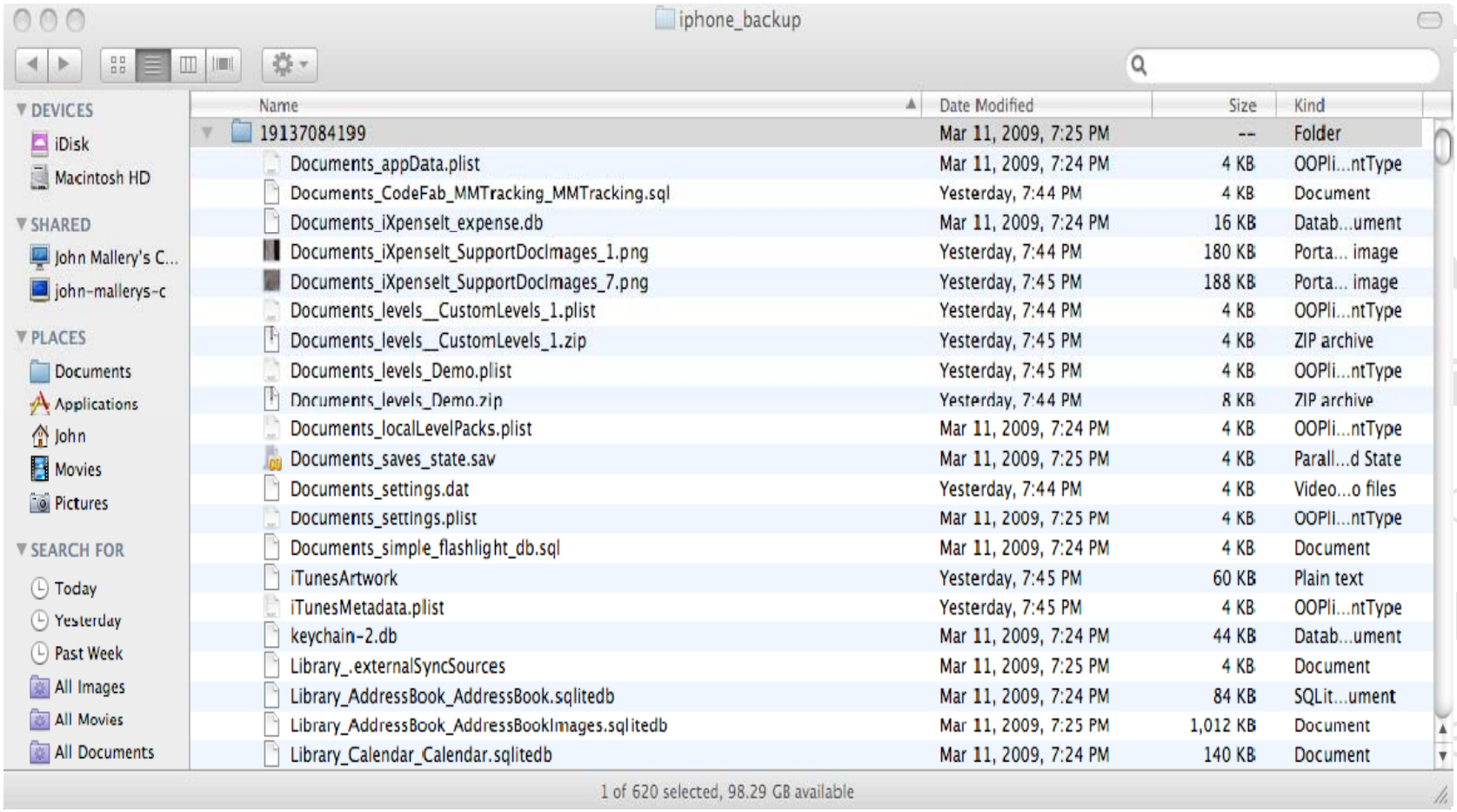
acumen
insight
ideas
attention
reach
expertise
depth
agility
talent

Tools

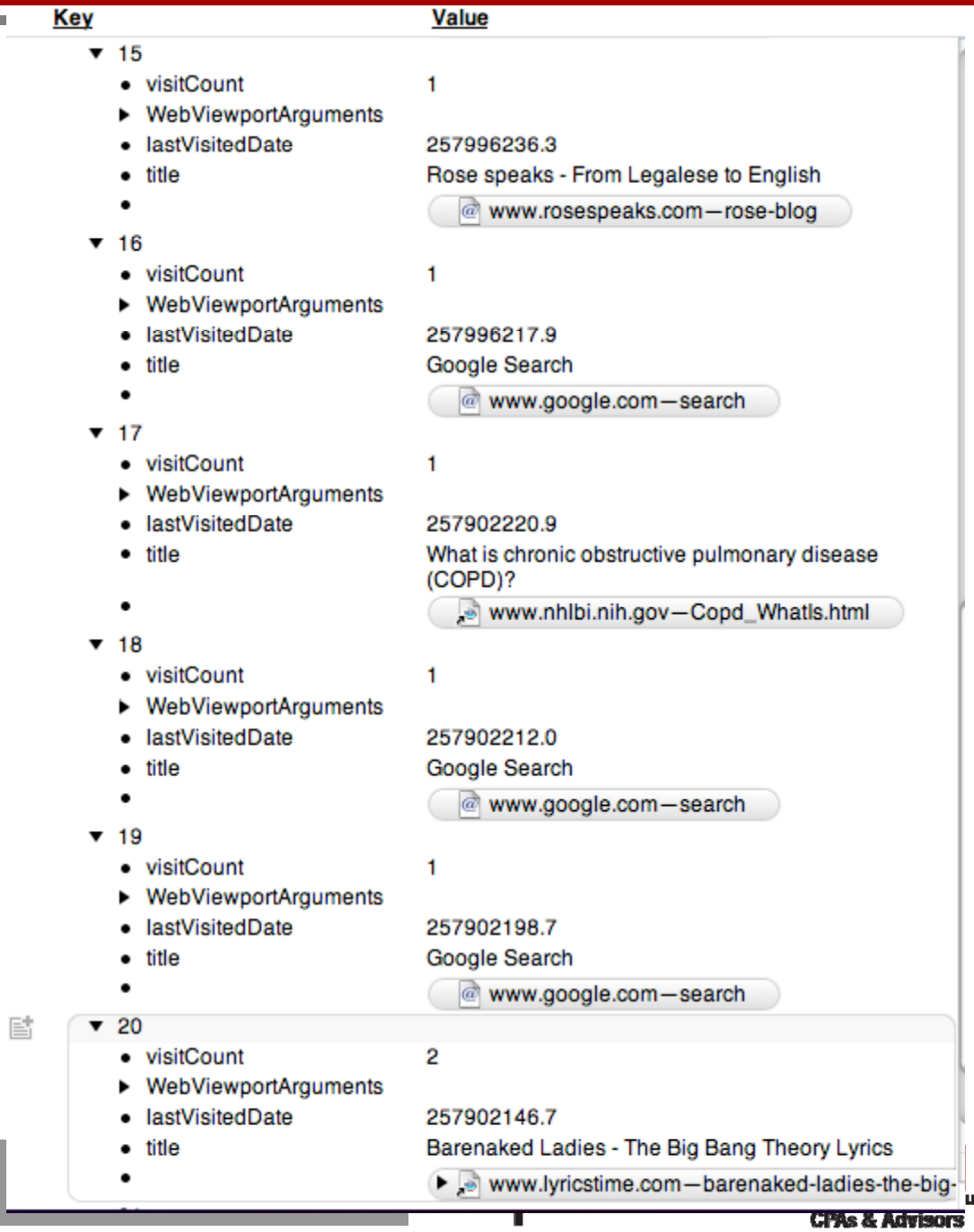
- Black Bag Tech – http://www.blackbagtech.com
- MobileSync Browser http://homepage.mac.com/vaughn/msync/
- iPhoneParser http://www.macosxforensics.com/Downloads/files/iPhoneParser.app.zip

iPhoneParser

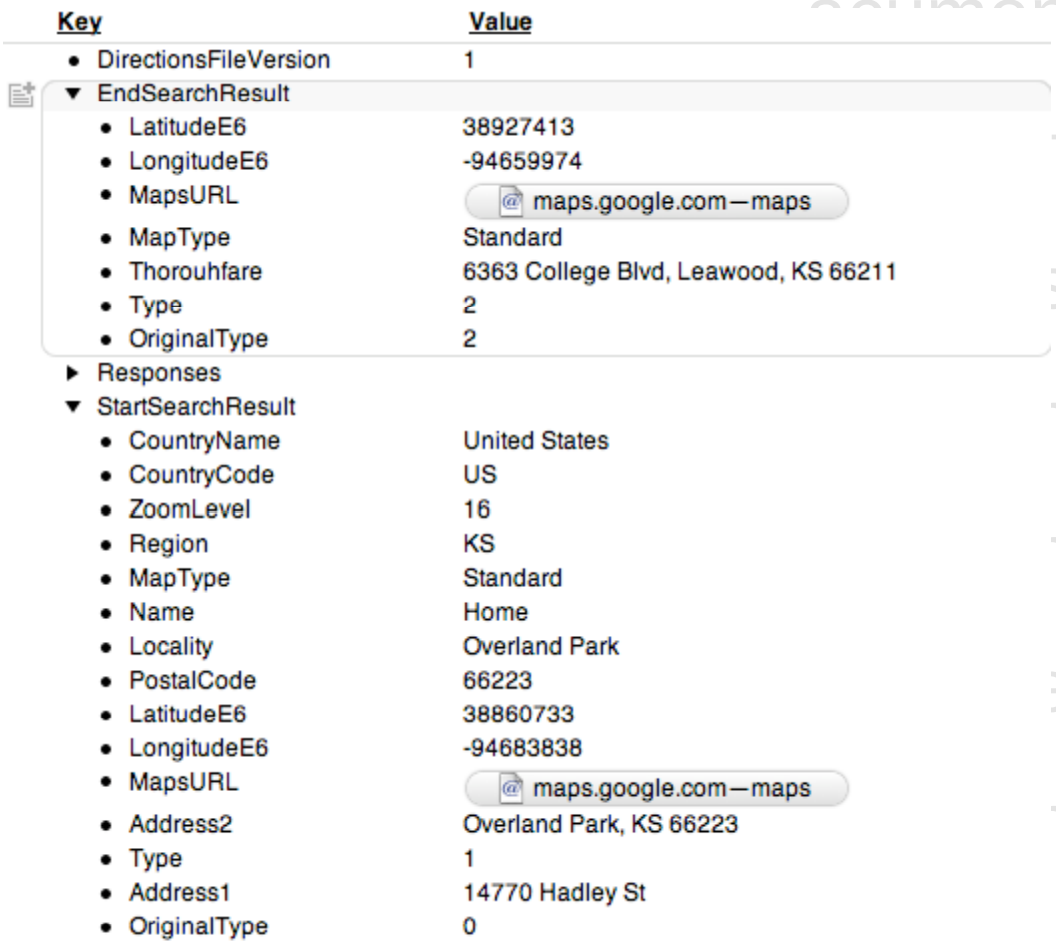
Creates iphone_backup folder on Desktop



Library_Safari_History.plist

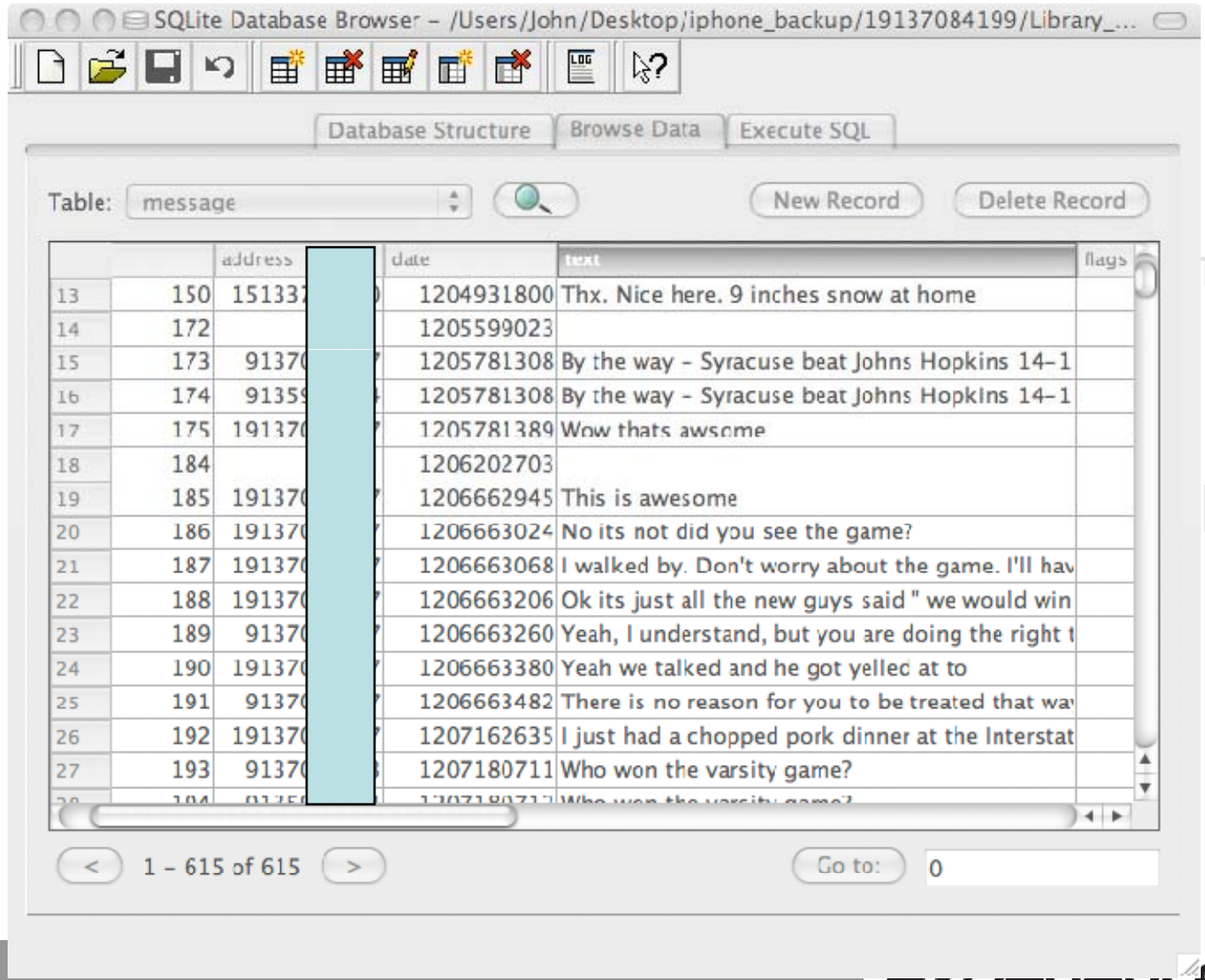


Library_Maps_Directions.plist

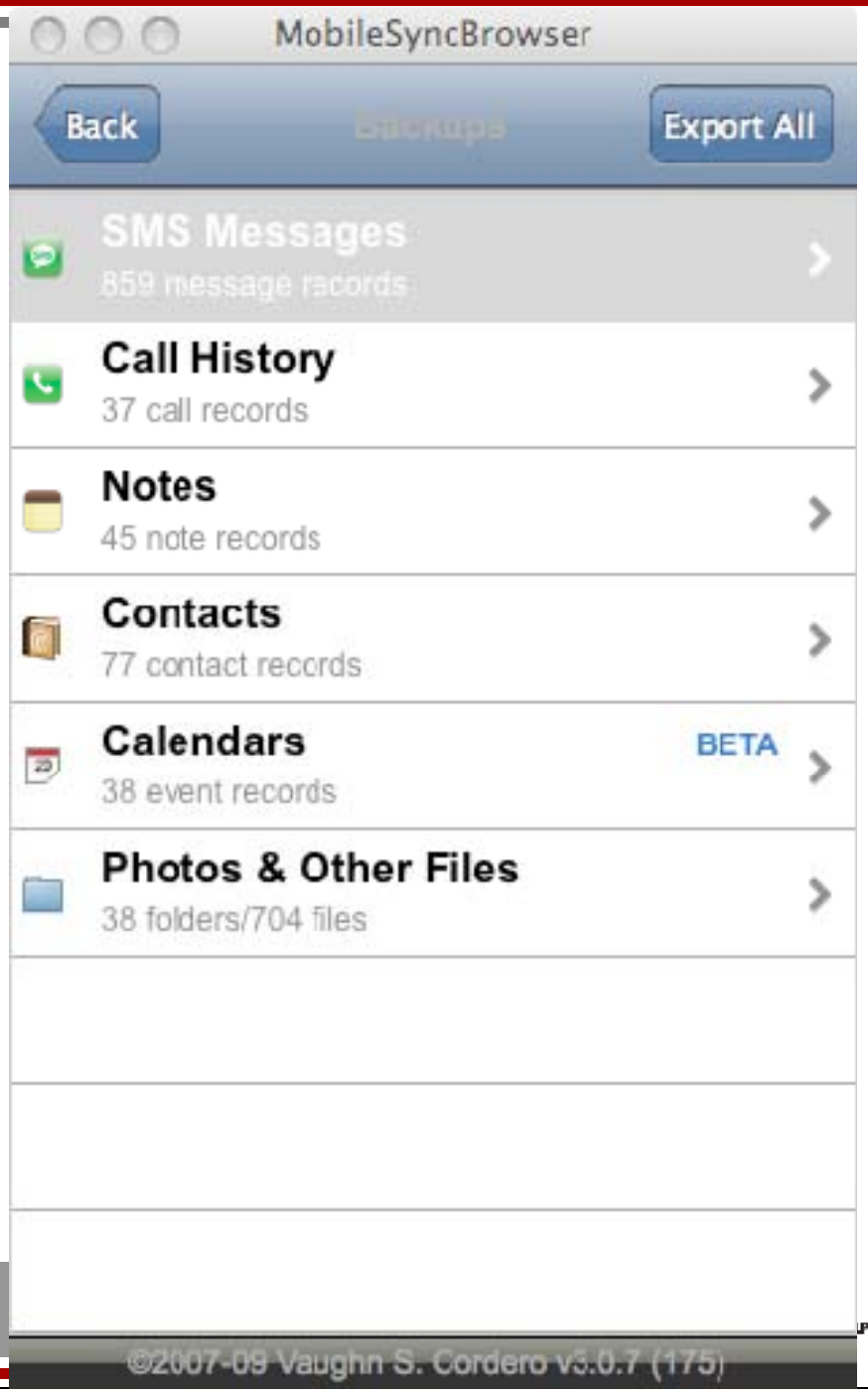


Key	Value
DirectionsFileVersion	1
EndSearchResult	
LatitudeE6	38927413
LongitudeE6	-94659974
MapsURL	maps.google.com—maps
MapType	Standard
Thorouhfare	6363 College Blvd, Leawood, KS 66211
Type	2
OriginalType	2
Responses	
StartSearchResult	
CountryName	United States
CountryCode	US
ZoomLevel	16
Region	KS
MapType	Standard
Name	Home
Locality	Overland Park
PostalCode	66223
LatitudeE6	38860733
LongitudeE6	-94683838
MapsURL	maps.google.com—maps
Address2	Overland Park, KS 66223
Type	1
Address1	14770 Hadley St
OriginalType	0

Library_SMS_sms.db http://sourceforge.net/projects/sqlitebrowser/



http://homepage.mac.com/vaughn/msync/



But…

- With iTunes 9, you now have the ability to encrypt your iPhone backup

acumen
insight
ideas
attention
reach
expertise
depth
agility
talent

iPhone – Voice Memo App

- Creates voice memos as **m4a** files.
- Can be emailed as attachments
- Attachments named “Memo.m4a”
- Not keyword searchable

iPod Stuff

Diagnostic and Disk Modes

acumen

insight

ideas

attention

reach

expertise

depth

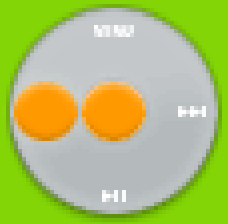
agility

talent

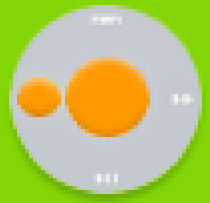
First Generation
"Scroll Wheel"

Second Generation
"Touch Wheel"

Third Generation
"Touch Wheel"



Fourth Generation
"Click Wheel"

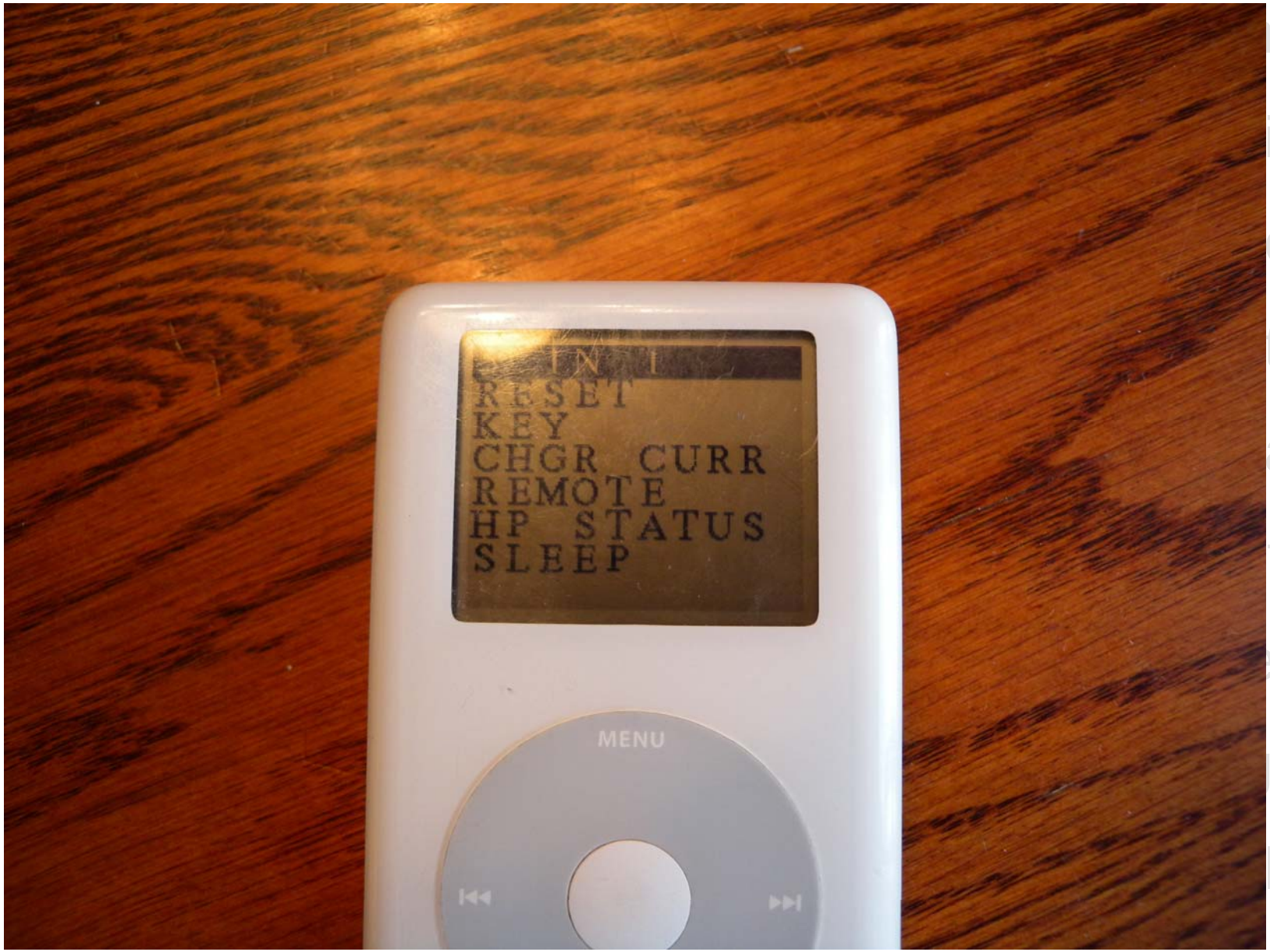


Fifth Generation
"Click Wheel"

Hard Reset

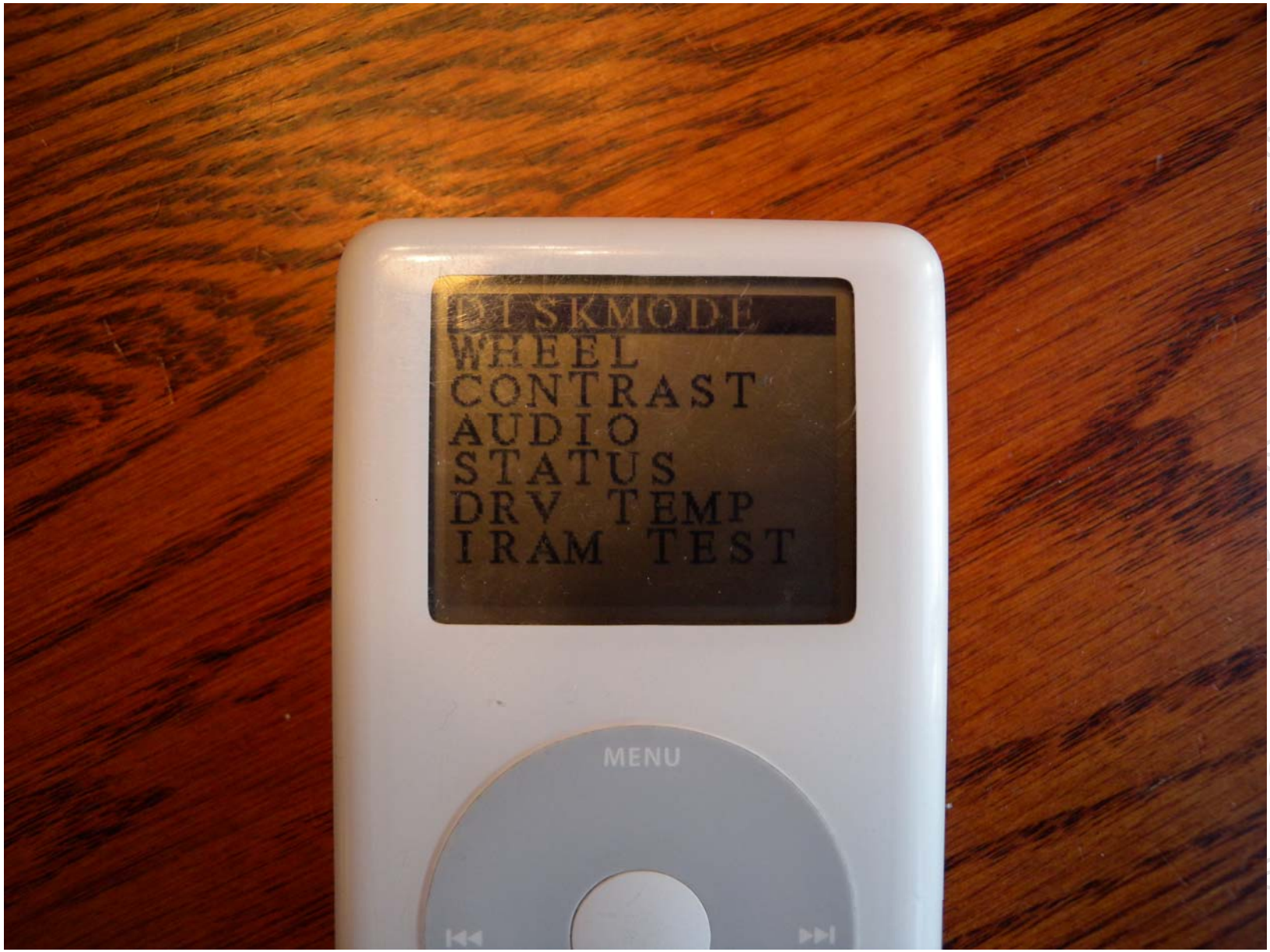
Diagnostic Mode

Disk Mode


RESET
KEY
CHGR CURR
REMOTE
HP STATUS
SLEEP
MENU

men
sight
leas
tion
each
rtise
epth
gility
lent

experience BKD LLP
CPAs & Advisors



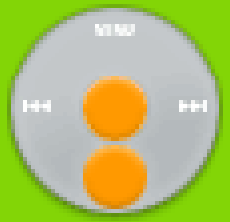
men
sight
deas
ntion
each
rtise
epth
gility
alent

experience BKD LLP
CPAs & Advisors

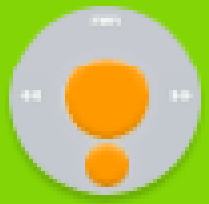
First Generation
"Scroll Wheel"

Second Generation
"Touch Wheel"

Third Generation
"Touch Wheel"



Fourth Generation
"Click Wheel"



Fifth Generation
"Click Wheel"

Hard Reset | Diagnostic Mode | Disk Mode

Stranger Devices

- Crane black box
- Computer from a surgical robot
  - Automatically records procedure as default
  - Patient dies
  - Relevant video has been deleted
  - Oops

Still seeing

- Technology implemented without any consideration to:
  - ✓ Legal requirements
  - ✓ Document retention
  - ✓ Document/File management
  - ✓ Internal controls
  - ✓ Security or Privacy

acumen
insight
ideas
attention
reach
expertise
depth
agility
talent

Example

- Dentist's office has a backup of their "system" on a hard drive in a safe
- Safe gets stolen
- Dentist's office want's to know if PII is accessible
- Developer says "no" our database is in a proprietary and closed format.
- However…

Example

- Name, address, phone number, SSN, patient notes, and patient id number all accessible by opening the backup file in a hex editor.
- Many hex editors are free!!

Another example

- Nurses decide they don't want to change in the nurses dressing room
- Change in an area monitored by a CCTV camera
- Sue for sexual harassment
- Unable to view video files except on server they were originally created upon
- Can't be viewed by the court, lawers, etc.

Forensic Data Mining

acumen

insight

ideas

attention

reach

expertise

depth

agility

talent

experience BKD LLP
CPAs & Advisors

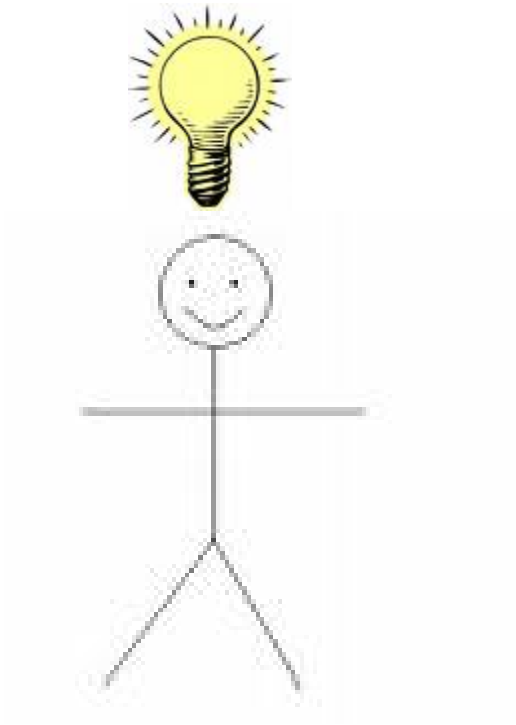
Forensic Data Mining

"Advanced data analysis used to identify activity patterns in financial and customer data not discernible through a manual review process."

"The process of discovering meaningful new relationships, patterns and trends by sifting through data using pattern recognition technologies as well as statistical and mathematical techniques."

The Data Mining Continuum

Hypothesis Testing
(Symptom-Based)



Knowledge Discovery
("Symptomless")

Why it is Effective

- While 70% of all frauds are found by tips, accidental discovery and disclosure…

  30% of all frauds are found by analysis
  (David Coderre, "Fraud Detection")

- Majority of data is in electronic format

- Data sets are massive in size and often proprietary in format

- "100% analysis is the most effective way to analyze for fraud" (Dr. Conan Albrecht, BYU)

Common Areas

- ✓ Fictitious (ghost) employees
- ✓ Shell companies and “phoenix operators”
- ✓ Loan fraud and other banking schemes
- ✓ Merger and acquisition due diligence
- ✓ Foreign Corrupt Practices Act investigations
- ✓ Money laundering
- ✓ Insurance claims fraud
- ✓ Subprime lending
- ✓ Embezzlement and financial statement fraud

Forensic Data Mining

Fraud Symptoms

acumen
insight
ideas
attention
reach
expertise
depth
agility
talent

experience BKD LLP
CPAs & Advisors

Fraud Symptoms

Payroll

Employees with no deductions
Activity subsequent to termination or before hire
Employee with no sick/vacation/timeoff
High pay vs department baselines
Duplicate phone number(s)
Duplicate addresses
Duplicate direct deposit accounts
Short duration of hire/termination
Same employee assigned to multiple departments
Timecard anomalies (threshold punchouts)
In payroll but not on phone list or active employee files

Fraud Symptoms

Vendors or Customers (Companies, Banks, etc.)

- Name similarity (phonetics, etc.)
- Acceleration (systematic spending increases)
- Employee address matches customer/vendor address
- Customer Tax ID matches another customer Tax ID
- Customer/vendor phone number matches employee phone
- Duplicate invoices or slightly altered attributes
- Sudden spike in invoice volume or activity
- Missing contact information (address, phone, names)
- High volume of transactions ending in 0 or 5
- Unusual activity compared to similar vendors or customers
- Weekend or holiday transaction dates
- Transactions processed at unusual hours
- Address is PO Box, maildrop, prison or high-risk ZIP code
- "Dormant" account suddenly active

Loan Master File

CUSTOMER NAME	ADDRESS	city	PHONE	TIN_SSN	AMT
TERRY CEO	PO BOX 145	SOMEWHERE	CEO Cell	500-17-8762	352,800.00
TERRY CEO	PO BOX 145	SOMEWHERE	CEO Cell	500-17-8762	100,000.00
P & Q BUILDERS	P O BOX 145	SOMEWHERE	CEO Cell	24-3784029	600,000.00
P & Q BUILDERS	P O BOX 145	SOMEWHERE	CEO Cell	24-3784029	269,172.81
P & Q BUILDERS	P O BOX 145	SOMEWHERE	CEO Cell	24-3784029	200,000.00
P & Q DISTRIBUTING INC	PO BOX 247	CITY	555-555-1234	24-3784029	100,000.00
P & Q DISTRIBUTING INC	PO BOX 247	CITY	555-555-1234	24-3784029	10,000.00

(1) Name similarity
(2) Customer address matches CEO address
(3) Customer phone matches CEO cell phone
(4) Customer TIN matches other customer TIN

DEBIT

GENERAL LEDGER [redacted] BANK

Date 1-25-01

ACCOUNT NUMBER [redacted]

Account LOAN SUSPENSE

DESCRIPTION	AMOUNT	
P & Q Builders	600,000	00
Note # [redacted]400		
Approved by SA TOTAL	600,000	00

RP-59

[redacted] 76000" 95 "0060000000"

CEO's Personal
Checking Account

Forensic Data Mining

acumen
insight
ideas
ttention
reach
e
n
agility
talent

Less Obvious Relationships:
Addresses and Geocoding

NAME	ADDRESS	CITY	STATE	ZIP
Employee	123 Maple Street	Our Town	MO	64678
Vendor	123 Maple Street	Our Town	MO	64678

Fictitious Company

NAME	ADDRESS	CITY	STATE	AP STAFF
Syntec Corporation	1221 East Kearney	Springfield	MO	Devon

Cross Reference Against:

- ✓ Maildrops (Mailbox Services)
- ✓ Correctional Facilities
- ✓ High-Risk ZIP Codes

→ The UPS Store
1221 East Kearney
Springfield, MO

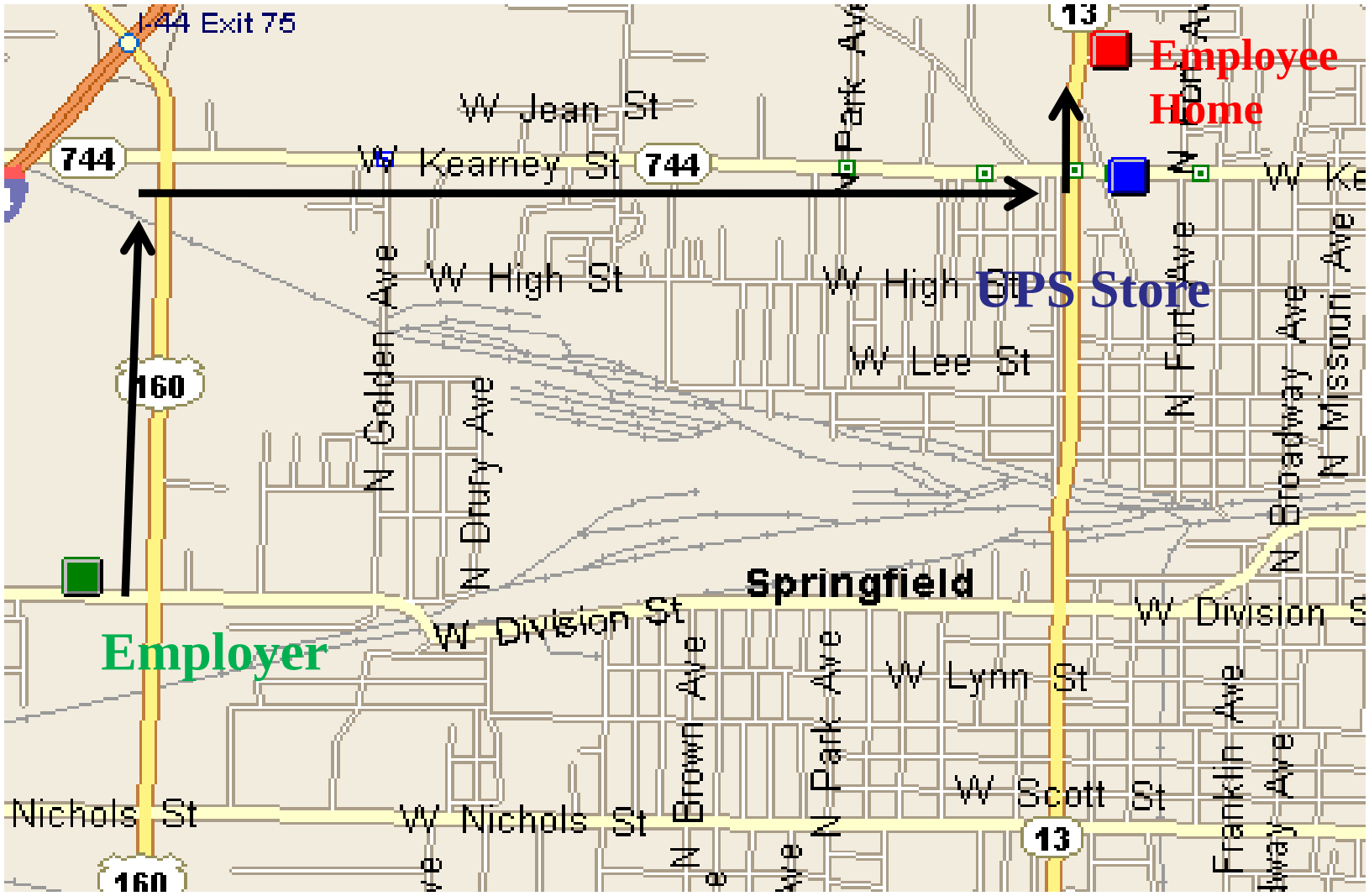
Fictitious Company

NAME	ADDRESS	CITY	STATE	Latitude	Longitude
Syntec Corporation	1221 East Kearney	Springfield	MO	37.320552	-98.536550
Devon (AP Staff)	312 East Warwick	Springfield	MO	37.320289	-98.538360

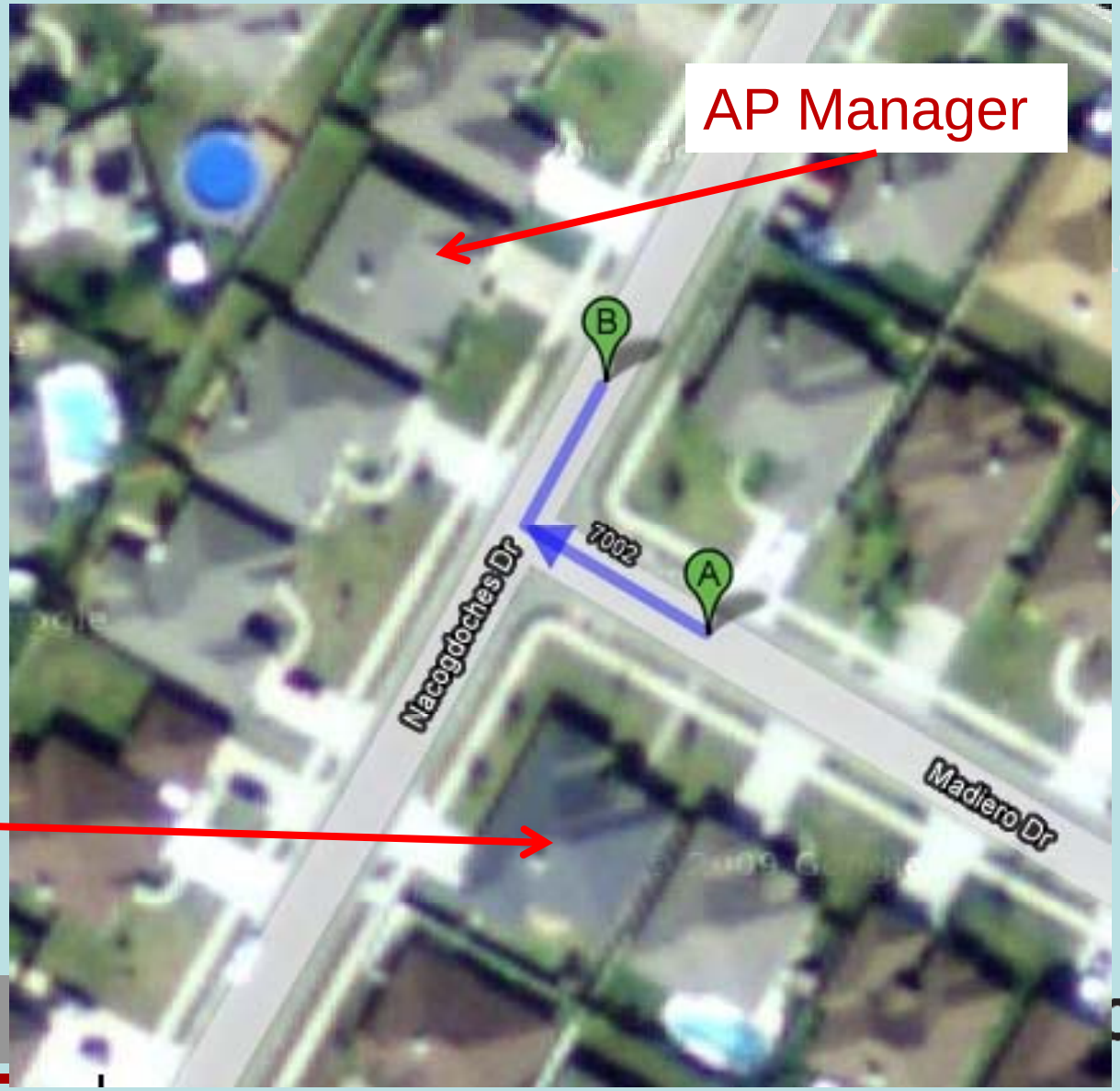
$$\Delta\hat{\sigma} = \arctan\left(\frac{\sqrt{(\cos\phi_f \sin\Delta\lambda)^2 + (\cos\phi_s \sin\phi_f - \sin\phi_s \cos\phi_f \cos\Delta\lambda)^2}}{\sin\phi_s \sin\phi_f + \cos\phi_s \cos\phi_f \cos\Delta\lambda}\right)$$

965 Feet

Mapping Employee-Vendor Relationship



Geocoding

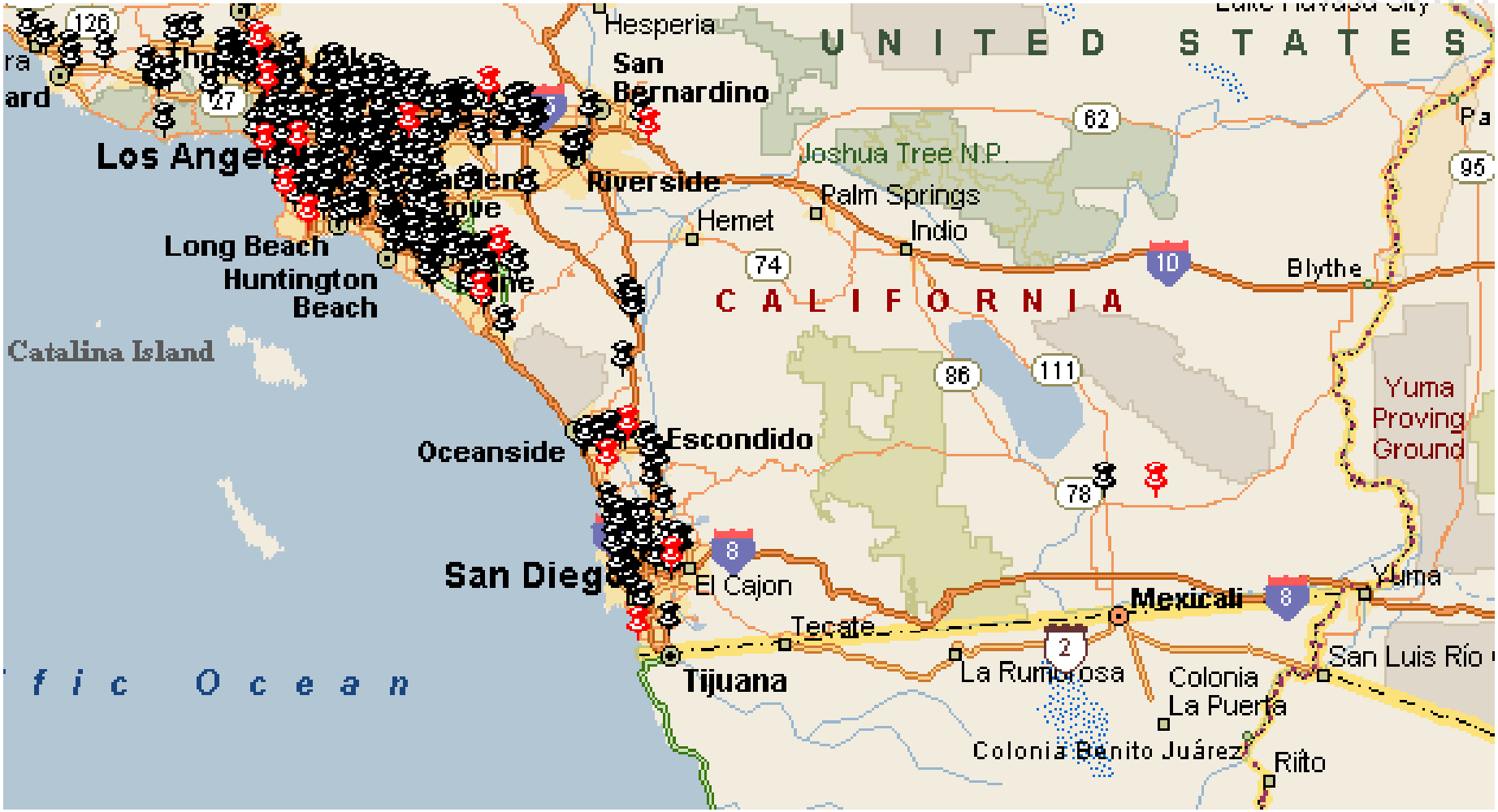


Vinny's Salvage Yard

acumen
insight
ideas
ttention
reach
xpertise
depth
agility
talent

ce BKD LLP
CPAs & Advisors

Visual Mapping



Data Mining

acumen

insight

ideas

attention

reach

expertise

depth

agility

talent

Benford's Law

(aka Digital Frequency Analysis)

Benford's Law

Benford's Law
Expected Digit Frequencies

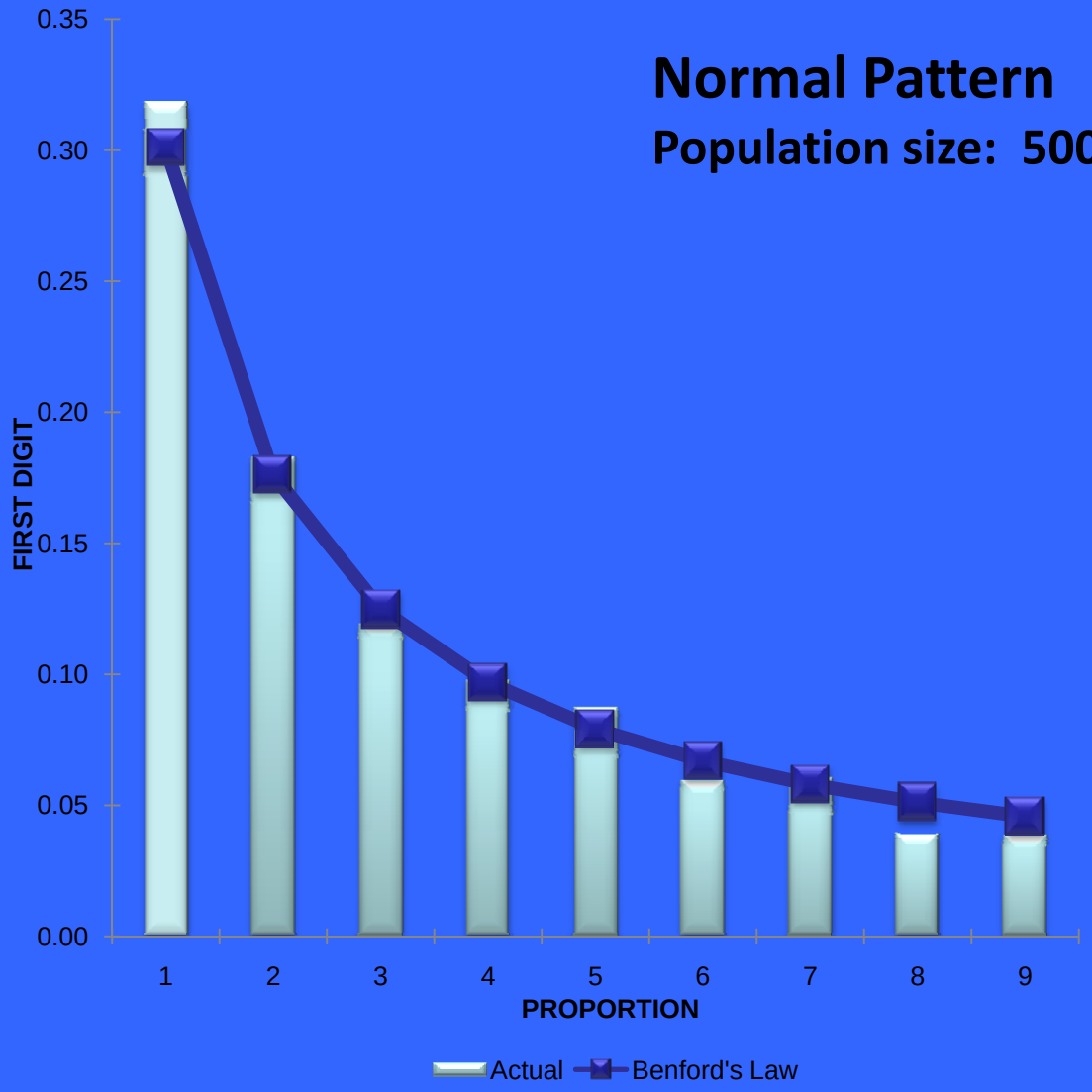
First Digit	Expected Frequency	Second Digit	Expected Frequency
1	30.10%	0	11.97%
2	17.61%	1	11.39%
3	12.49%	2	10.88%
4	9.69%	3	10.43%
5	7.92%	4	10.03%
6	6.69%	5	9.67%
7	5.80%	6	9.34%
8	5.12%	7	9.04%
9	4.58%	8	8.76%
		9	8.50%

1. Not random as one would expect
2. Also works on 1st 2 digits, 3 digits and decimals

acumen
insight
ideas
attention
reach
expertise
depth
agility
talent

Benford's Law

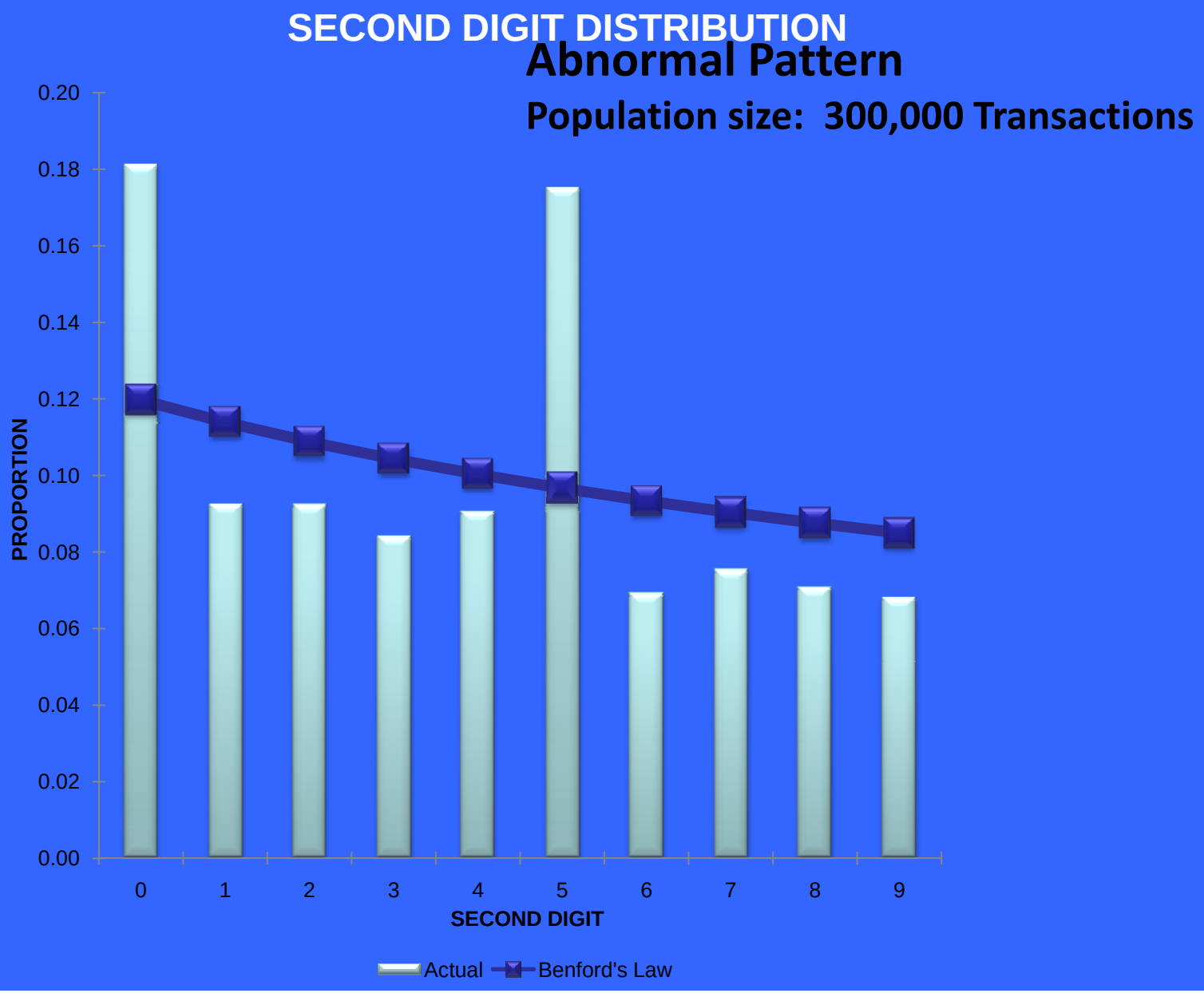
FIRST DIGIT DISTRIBUTION



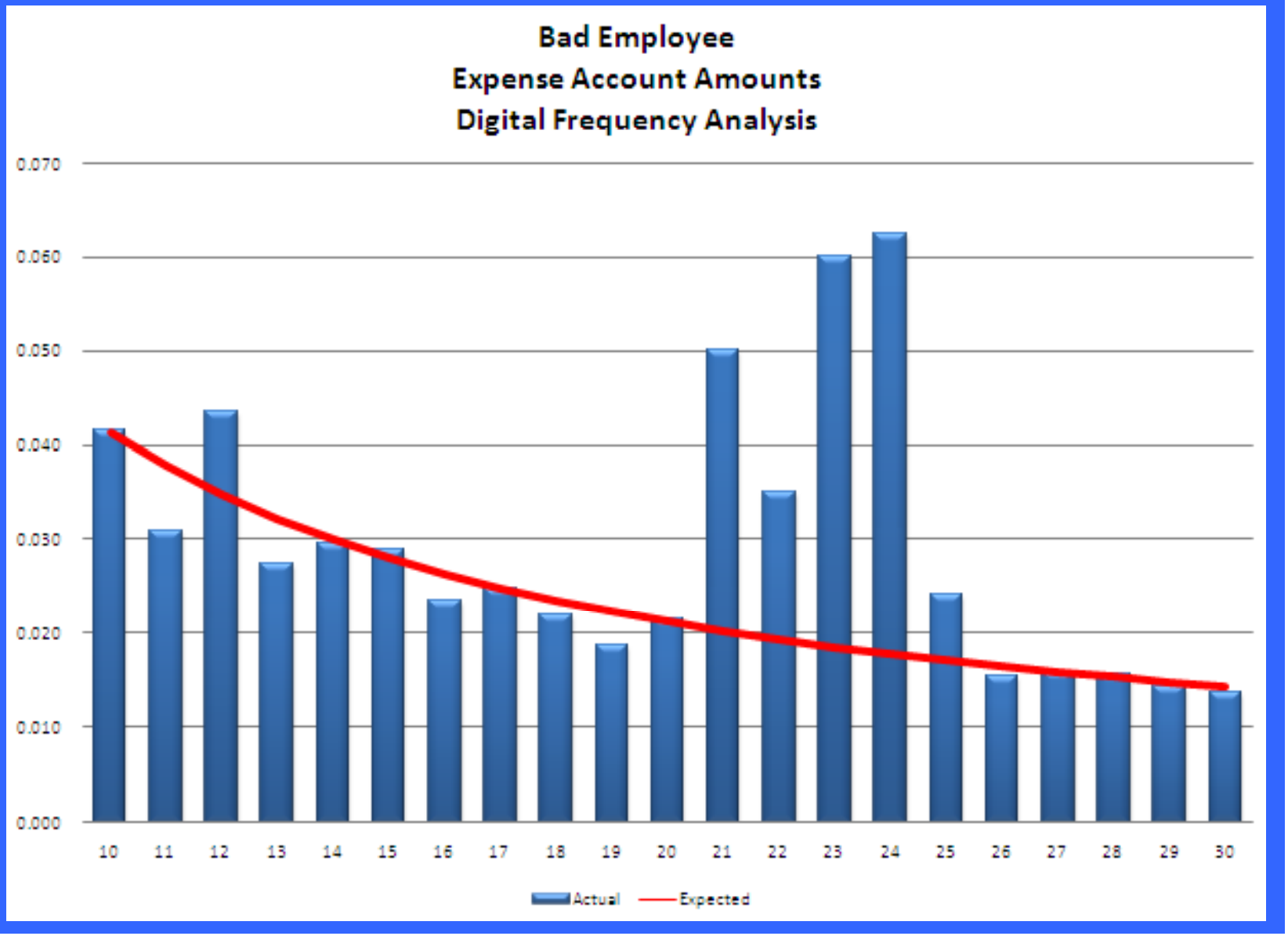
Normal Pattern

Population size: 500,000 Transactions

Benford's Law



Expense Account Padding



Expense Account Padding

Data from One-week Business Trip

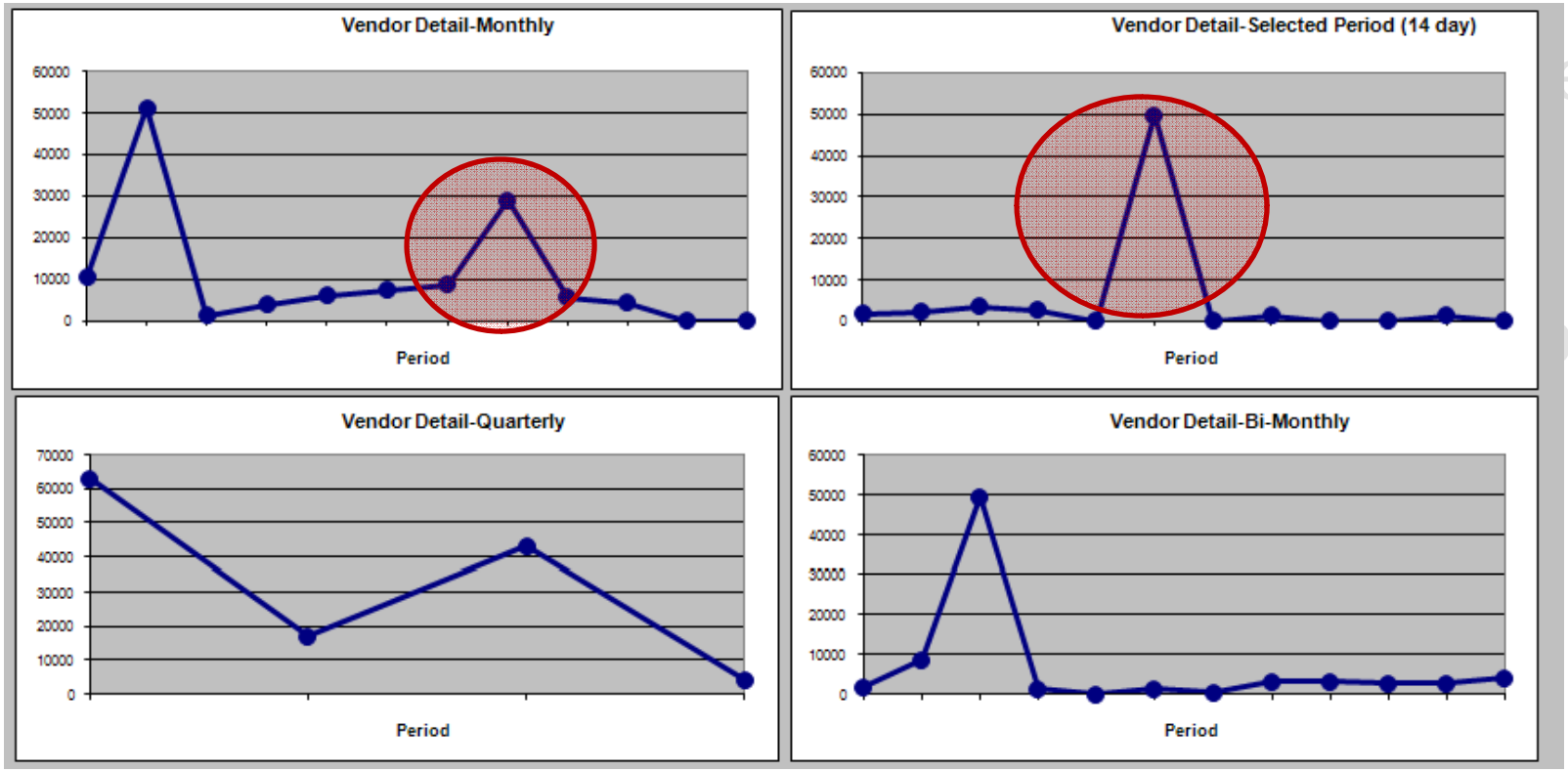
Dinner Expenses

Employee 1	**Employee 2**	Employee 3	Employee 4
$ 12.84	**$ 21.95**	$ 16.89	$ 14.11
15.55	**23.45**	14.12	14.91
13.67	**24.15**	15.03	15.84
14.98	**23.95**	12.68	16.78
$ 57.04	**$ 93.50**	$ 58.72	$ 61.64

Spending limit per meal without receipt is $25

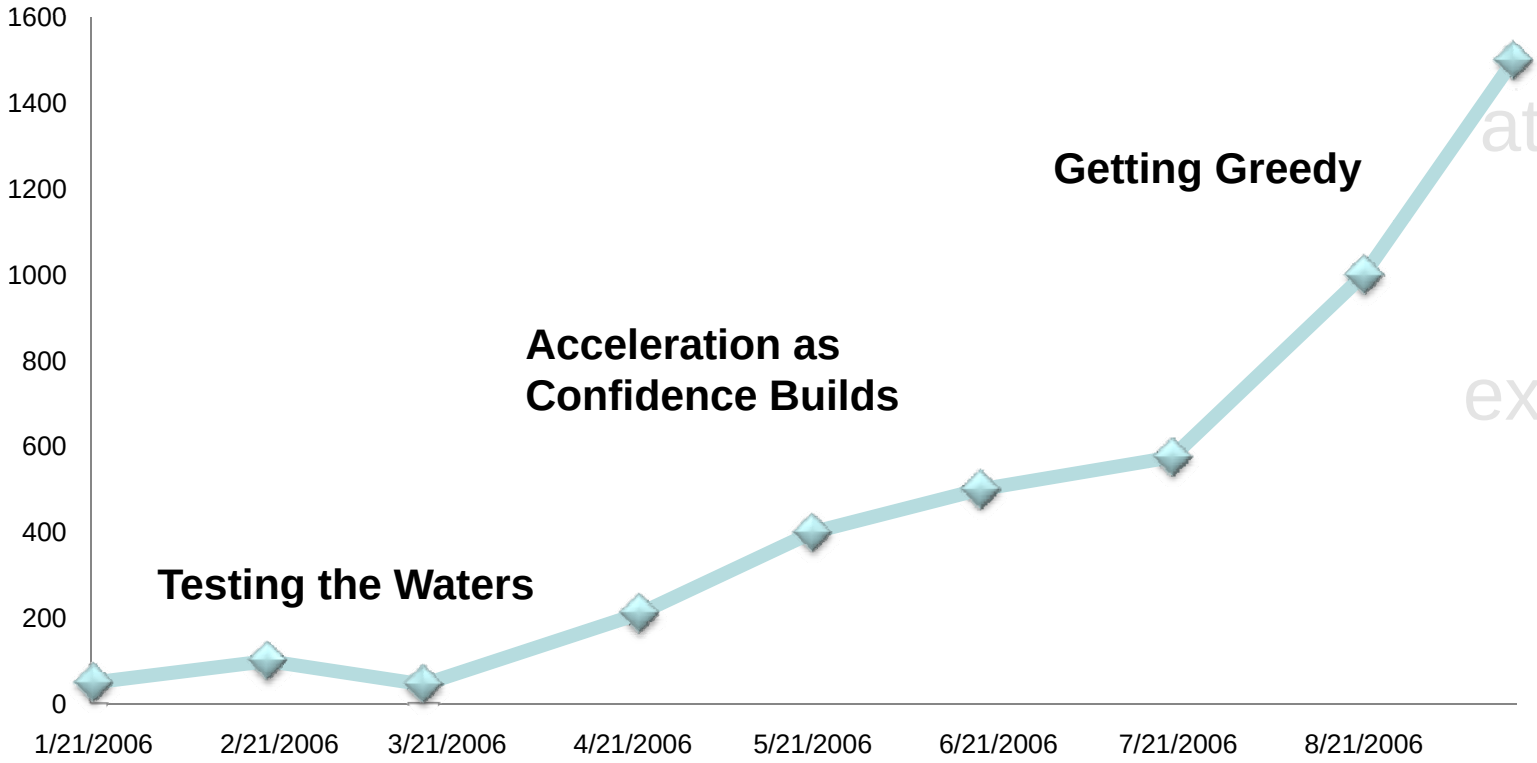
Data Mining

Time Series



Time Series

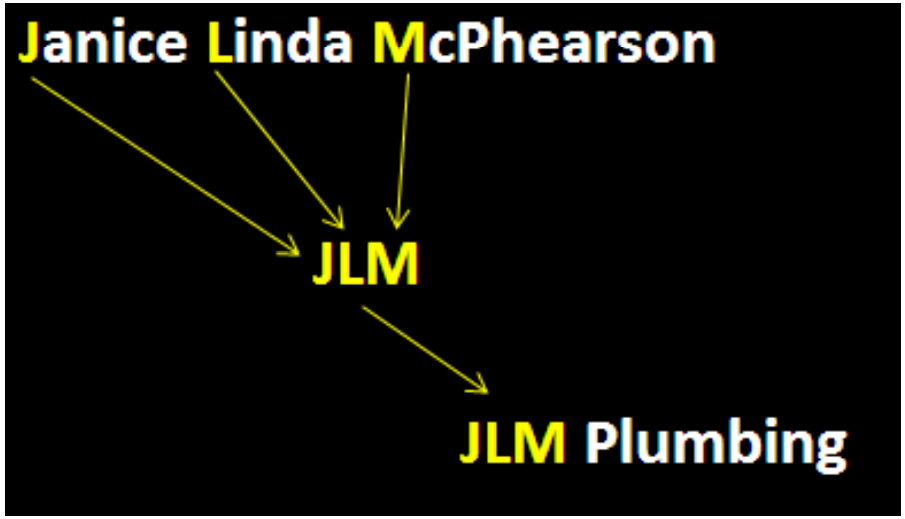
Vendor: JLM Plumbing AP Clerk: Janice McPhearson



acumen
insight
ideas
attention
reach
expertise
depth
agility
talent

Name Manipulation

1. **Acronym / Initials**



2. **Anagrams**

Name
CASHDAVID
DAVISCHAD

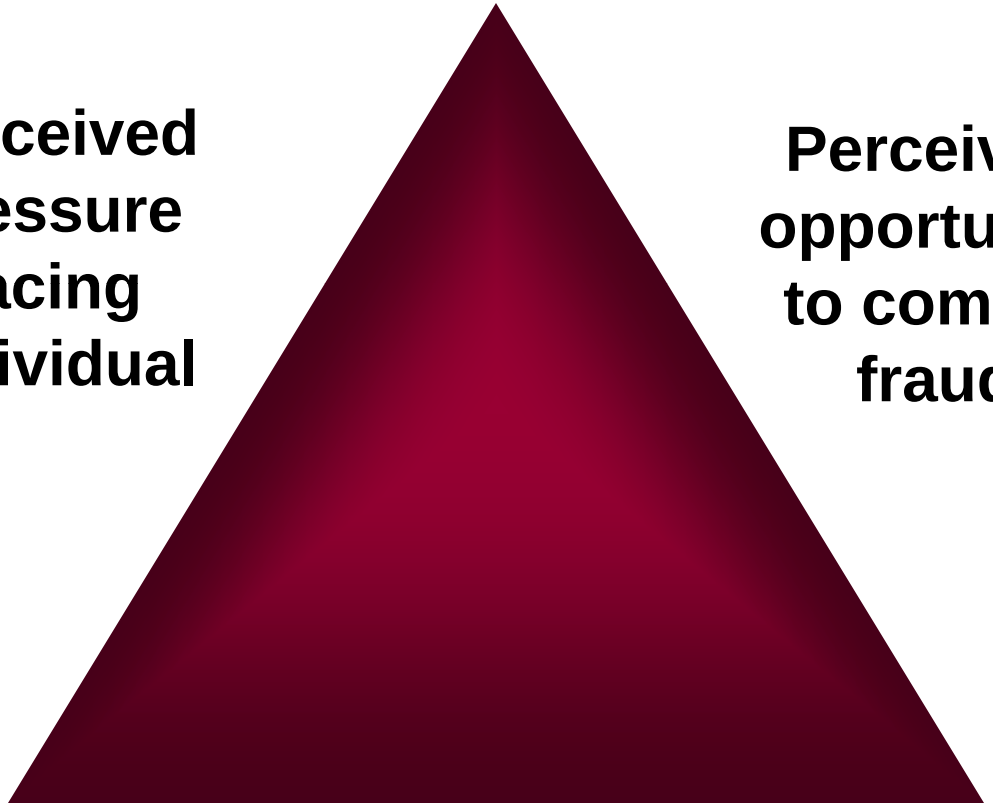
3. **Fictitious Names**

- Mick E. Mowse
- Princess Ariel
- George Ruth
- John Dough

4. **Others**

- Substitution
- Insertion or Omission
- Transposition
- Numb3r Subst1tut10n

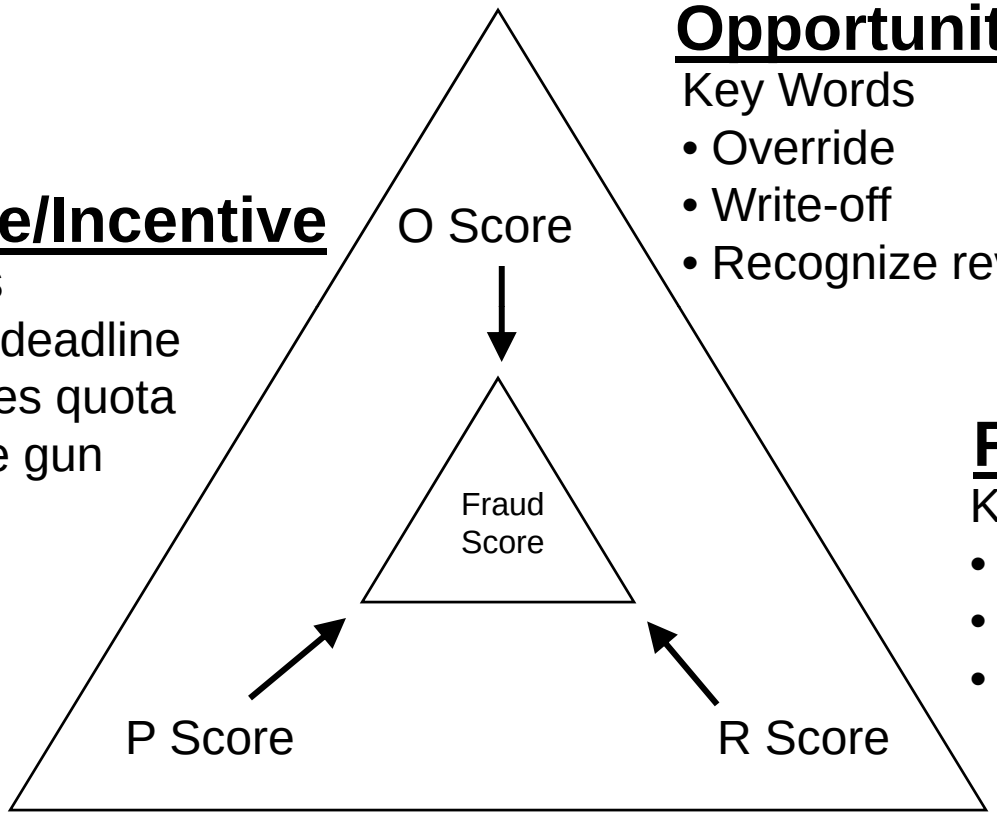
The Fraud Triangle



acumen
insight
ideas
attention
reach
expertise
depth
agility
talent

experience BKD LLP
CPAs & Advisors

Fraud Triangle Analytics



Opportunity

Key Words
- Override
- Write-off
- Recognize revenue

Pressure/Incentive

Key Words
- Meet the deadline
- Make sales quota
- Under the gun

Rationalization

Key Words
- I think it's OK
- Sounds reasonable
- I deserve

Source: "Detecting Fraud by Integrating E-mail Analytics with the Fraud Triangle," Fraud Magazine, May/June 2009

Secondary List - General Terms

Terms related to discussions in email, chat or memos.

Keyword Terms - accounting	Keyword Terms - general	Keyword Terms - general	Keyword Terms - general
won't realize	bigger fish	keep quiet	stealth
won't notice	I am leaving	keep it quiet	underhand
won't catch	take a position	under wraps	underhanded
help to reconcile	accept a position	under the table	under handed
missing checks	pursue a career	is confidential	under-hand
does not reconcile	decided to leave	keep it confidential	sneak
not reconciling	made a decision to	it is a secret	sneaky
doesn't reconcile	grass is green	a secret	sub-rosa
won't reconcile	convince me to stay	proceed w/12 cauti	surreptitious
will not reconcile	monster.com	be careful	undercover
match up	outta	carefully	under cover
matchup	out of here	hush	has no idea
get audited	new opportunit	hush hush	in the dark
audit	new challenge	hush-hush	cloak
	other opportunit	hushed	cloak and dagger
	exciting opportunit	this is a private	cloaked
	new company	private matter	concealed
	franchise	this is private	concealment
	new business	top secret	compete with
	business plan	don't tell	compete against
	compete	dont tell	betray
	non w/10 compete	call me at	disguise
	resign	call me at home	hide
	step dow	call me on my cell	hidden
	pursue other	my cell	suspect
	no longer be	clandestine	suspicious
	canned	covert	bribe
	birds of a feather	furtive	bribery
	flock together	conceal	extortion
	black mail	raise capital	push money
	blackmail	raise money	new enterprise
	black-mail	raise financing	established companies
	shady	get financing	companies in industry
	shiesty	get capital	industry leader
	sheisty	get money	top dog
	sketchy	my own money	market share
	we have to be careful	my investor	market penetration
	new venture	our investor	market leader
			gain market

The Cutting Edge

"**Symptomless Detection**" – Finding answers to questions that haven't even been asked.

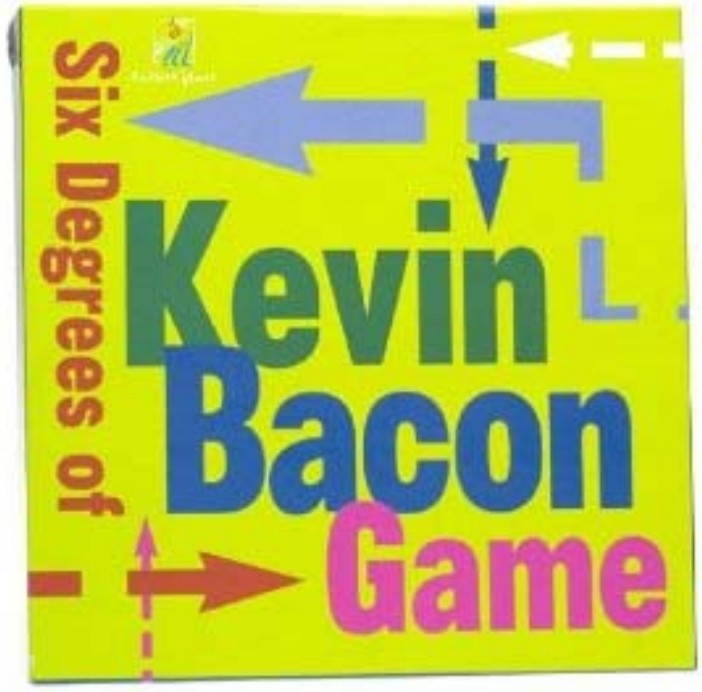
- Concept Searching – Detection based on tone, recurring themes and communication nuances
- Non-Obvious Relationship Association (Colleen McCue)
- Neural Networks and Artificial Intelligence
- Statistical-based prediction of events (Web Bot Project)

acumen
insight
ideas
attention
reach
expertise
depth
agility
talent

The Cutting Edge

Non-Obvious Relationship Association (NORA)

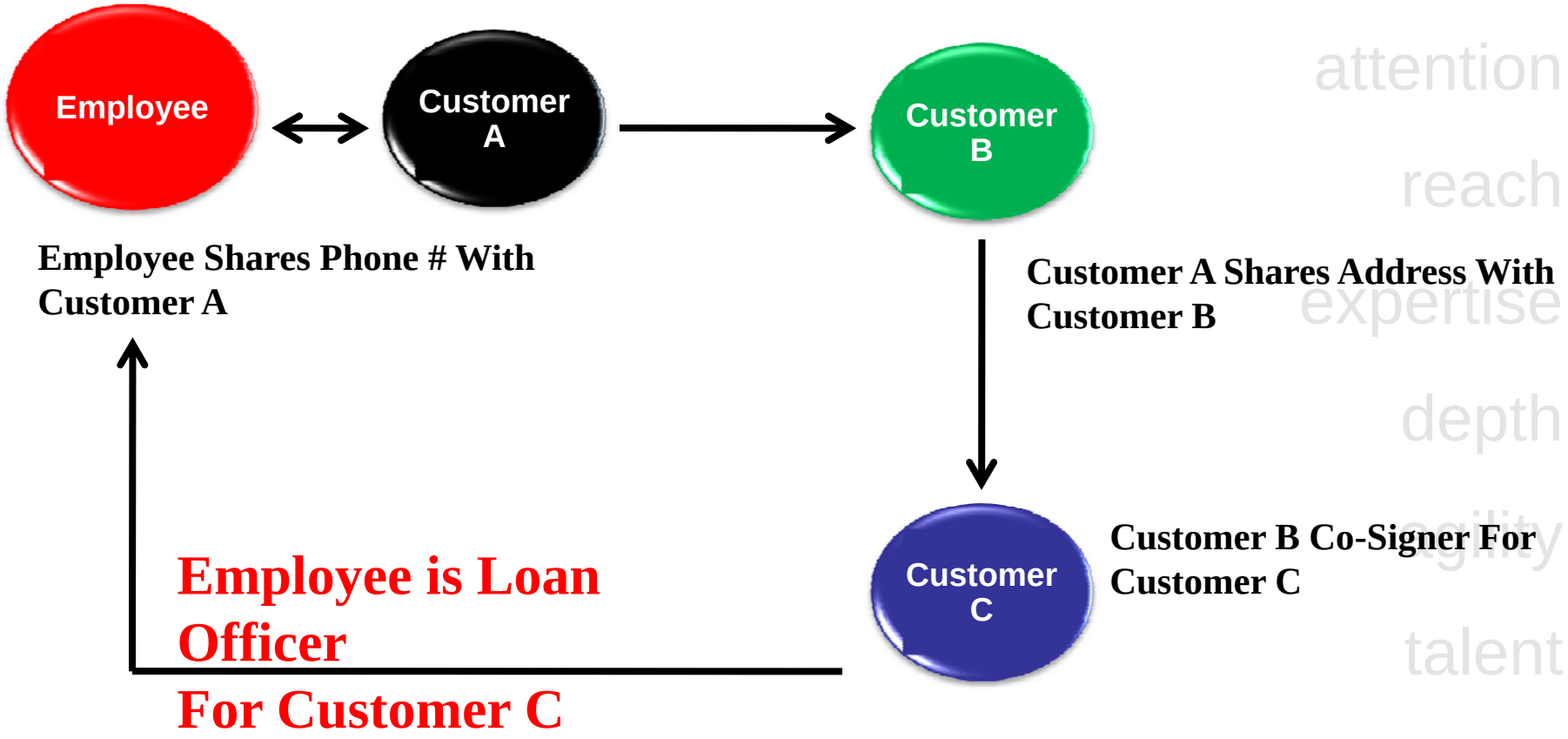
Items related by degrees of separation



Carrie Fischer was in *Star Wars*
with
Harrison Ford who was in *The Fugitive*
with
Tommy Lee Jones who was in *Batman Forever*
with
Val Kilmer who was in *Heat*
with
Robert Dinero who was in *Sleepers*
with
KEVIN BACON!

The Cutting Edge

NORA Example



The Cutting Edge

acumen
insight
ideas
attention
reach
expertise
depth
agility
talent

Neural Networks, Statistics and Concept Searching

- Uses mathematical algorithms to mimic the human neural network, and “learns” the conceptual meaning of words and phrases from a test set of documents (“digital bloodhound”).
- The more documents the engine “sees”, the more accurate its grasp of human language.
- Adept at detecting current conditions and predicting likelihood of future events based on language and patterns in corporate documents and email.

Read More About It…

"Fraud Examination" – Steve Albrecht and Conan Albrecht

"Fraud Detection" – David Coderre

"Digital Analysis Using Benford's Law – Mark Nigrini

"Data Mining and Predictive Analysis"
Intelligence Gathering and Crime Analysis - Colleen McCue

"Forensic Data Mining: Finding Needles in the Haystack" –
Archived Webcast at http://www.bkd.com/service/Forensics/Webcast/

Questions?

John Mallery
BKD, LLP
Twelve Wyandotte Plaza
120 W. 12th Street, Suite 1200
Kansas City, MO 64105
816.701.0267
lmorrow@bkd.com

acumen

insight

ideas

attention

reach

expertise

depth

agility

talent