Identifying Improper Payments and Fraud Using Data Visualization

Christopher Westphal

September 12-13, 2016

analysis365
REAL DATA · REAL RESULTS

What is an Improper Payment?

An improper payment is any payment that should not have been made or was made in an incorrect amount - including overpayments and underpayments.

- ***Duplicate payments***
- ***Payments to ineligible recipients***
- ***Incorrect amounts paid***
- ***Payments with insufficient or no documentation***

Beryl H. Davis
davisbh@gao.gov

GAO Study – Government Wide Estimates FY14

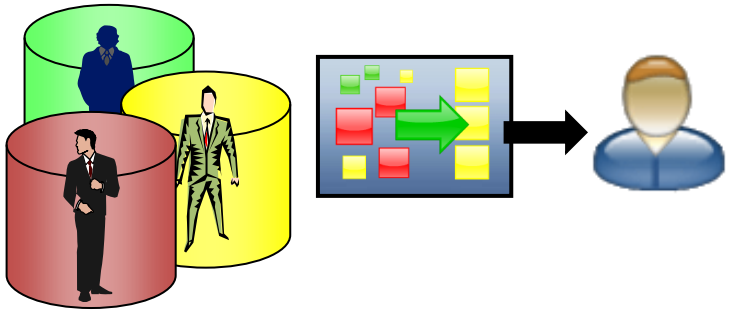
- Improper payment estimates totaled **$124.7B**
- 38 programs accounting for **$100.6B** in estimated improper payments
- Three programs accounted for **$80.9B** in improper payment, or 65% total for FY14
  - HHS – Medicaid $17.49B
  - HHS – Medicare Fee-for-Service $45.75B
  - Treasury – Earned Income Tax Credit $17.7B

Beryl H. Davis
davisbh@gao.gov

http://www.gao.gov/assets/670/669026.pdf

Similar Data, Different Content

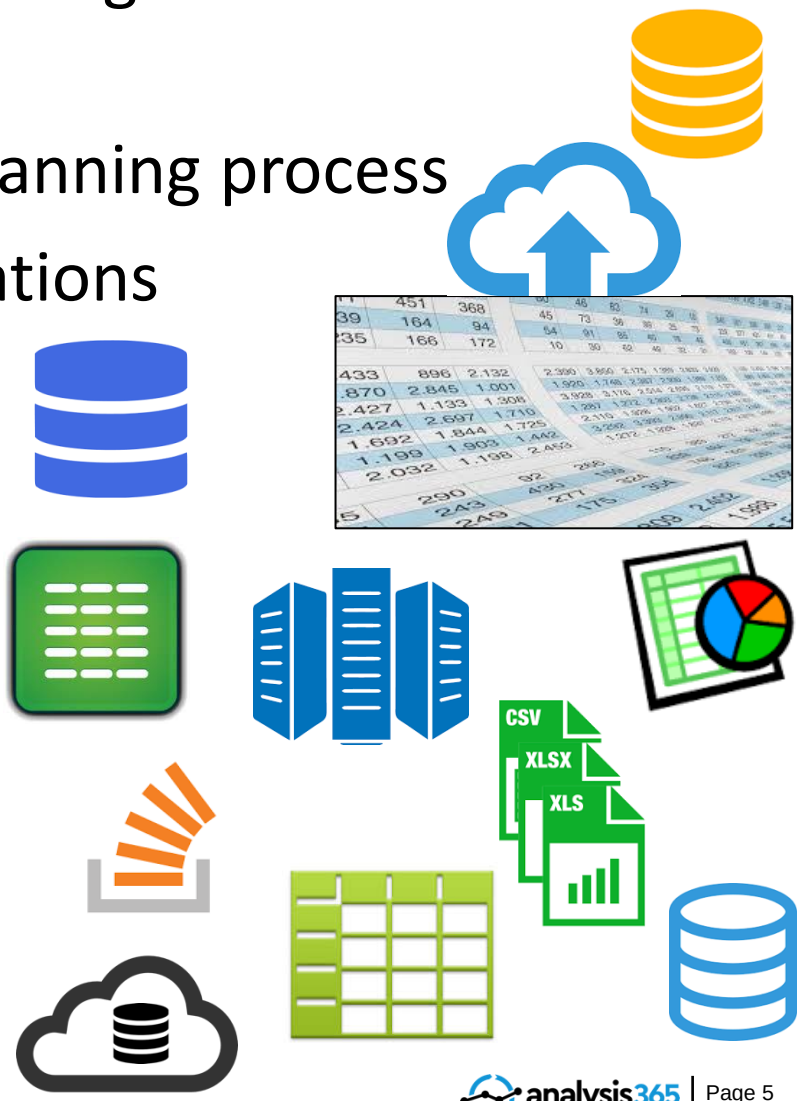
- SNAP – Supplemental Nutrition Assistance Program
- Welfare Benefits
- Energy Assistance
- Homestead Exemptions
- Unemployment Assistance
- Medicaid/Medicare
- Prescription Drugs
- Social Security Benefits
- SSA Supplemental Income
- Retirement/Pensions
- Workers Compensation
- School Lunch Program
- Direct Loan (Student)
- Public Housing / Rental Assistance
- Unregistered Businesses
- Underreported Income
- Tax Evasion
- …etc. etc. etc. etc…



How Auditors Can Use Data

Beryl H. Davis
davisbh@gao.gov

- Expand beyond sample-based testing to include analysis of entire populations
- Improved risk assessment and planning process
- Delivering better insight to operations
- Improve audit quality/accuracy
- Detect fraud (via technology)
    - Databases
    - Rule Based/Scoring
    - Statistics (w/drill down)
    - Machine Learning
    - Entity Analytics

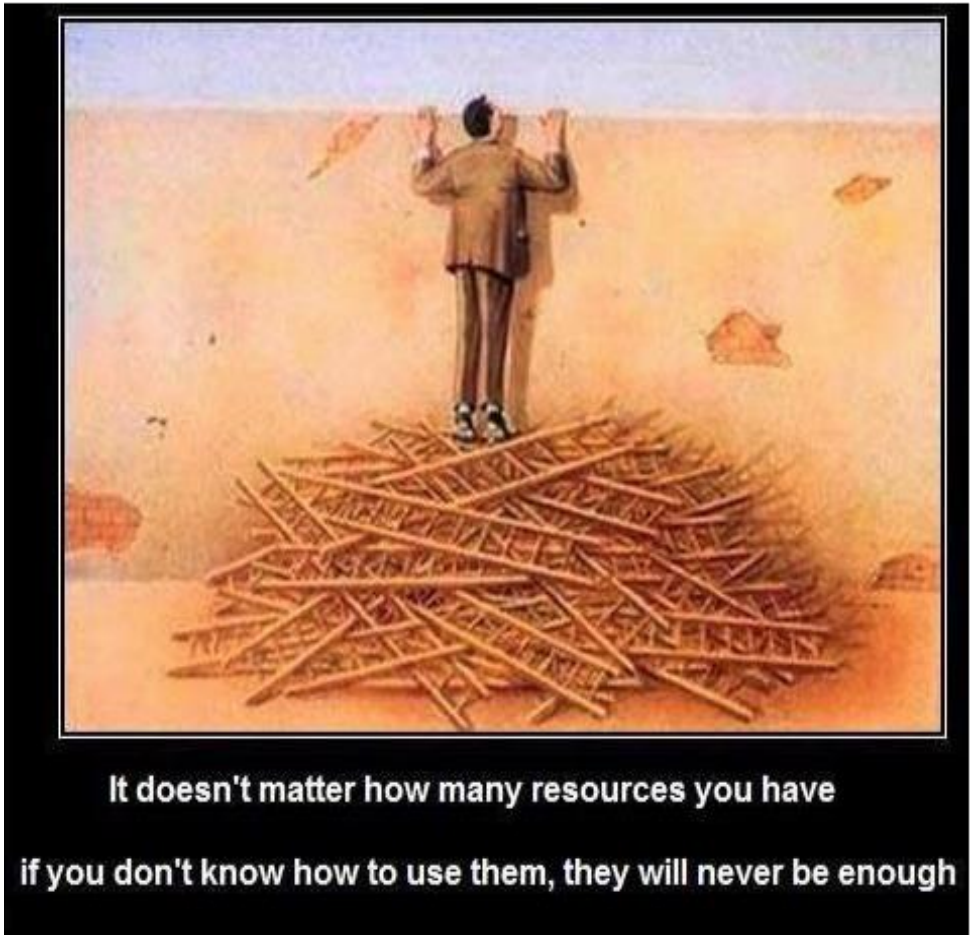


Technology Expectations for Analytics

It’s not the cost, complexity, or capability…

It’s how you use it:

- Quality vs Quantity
- The right technology to solve the right problem
- Change business processes based on discoveries
- Deliver actionable intelligence/analytics



Common Sense Patterns

- If it does not make sense…
- It is not normal…
- It seems unusual…
- Too coincidental…
- Too frequent…

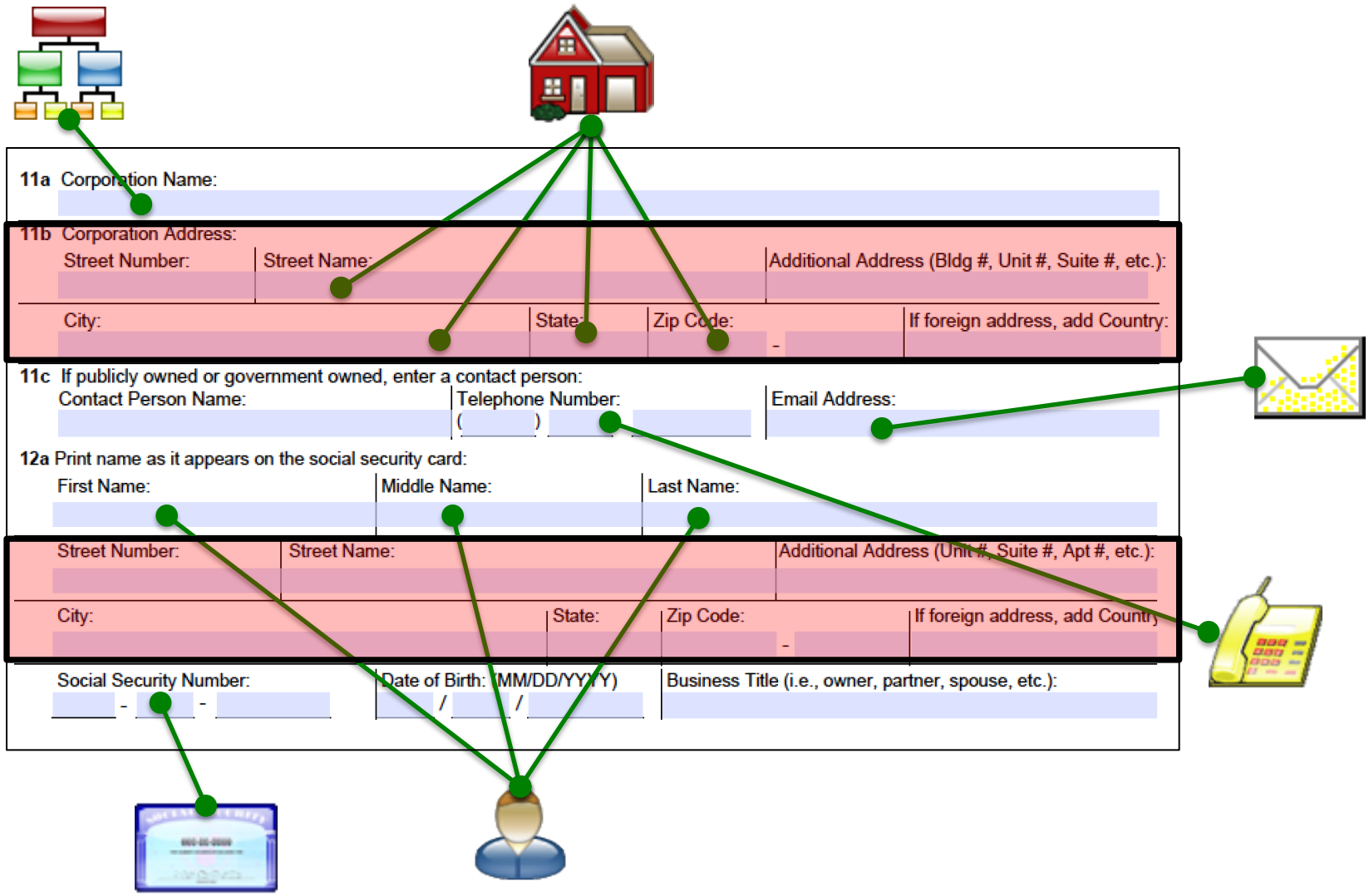
There is no **right** answer

There is no **wrong** answer

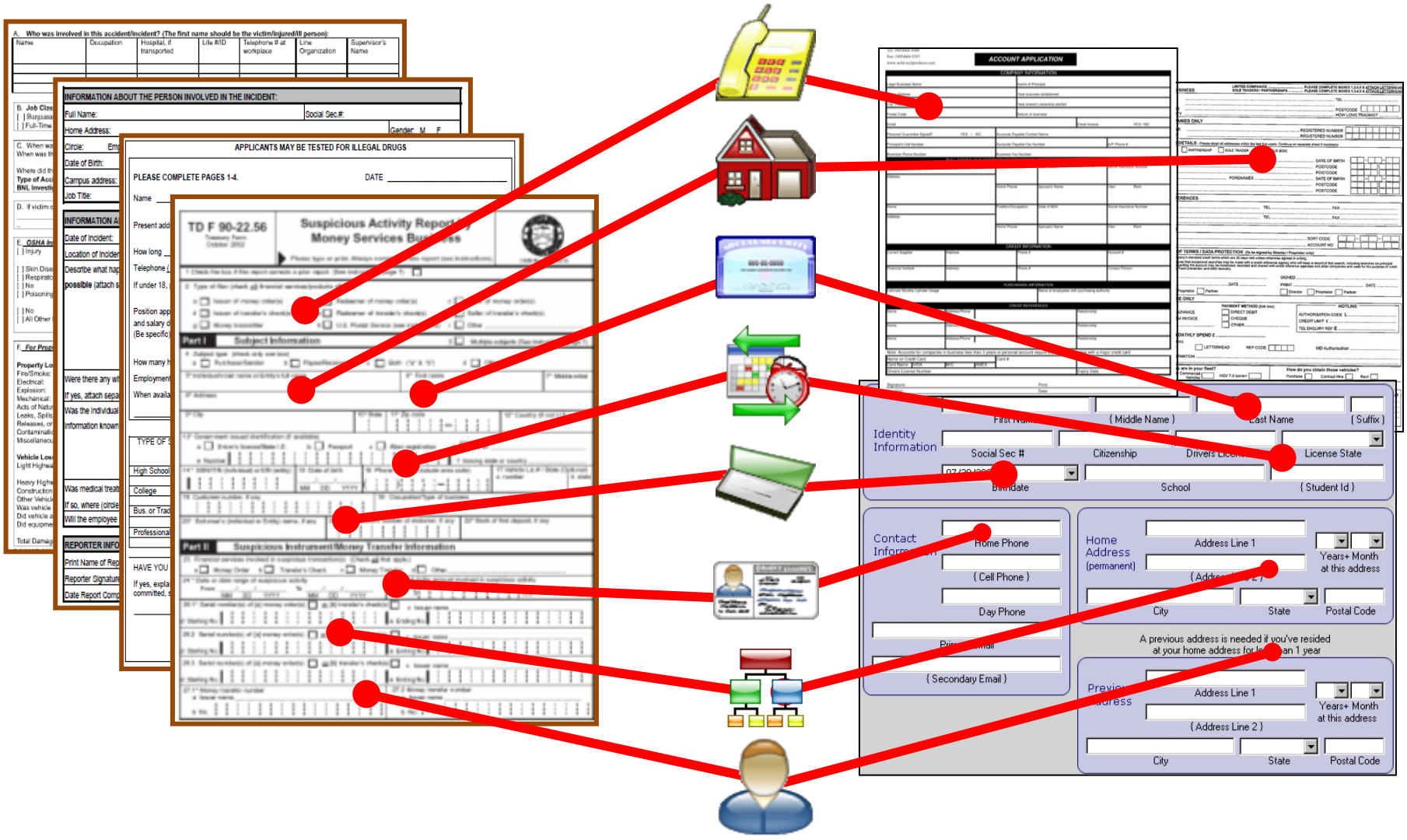
Interpretation in **context**

Context is Important

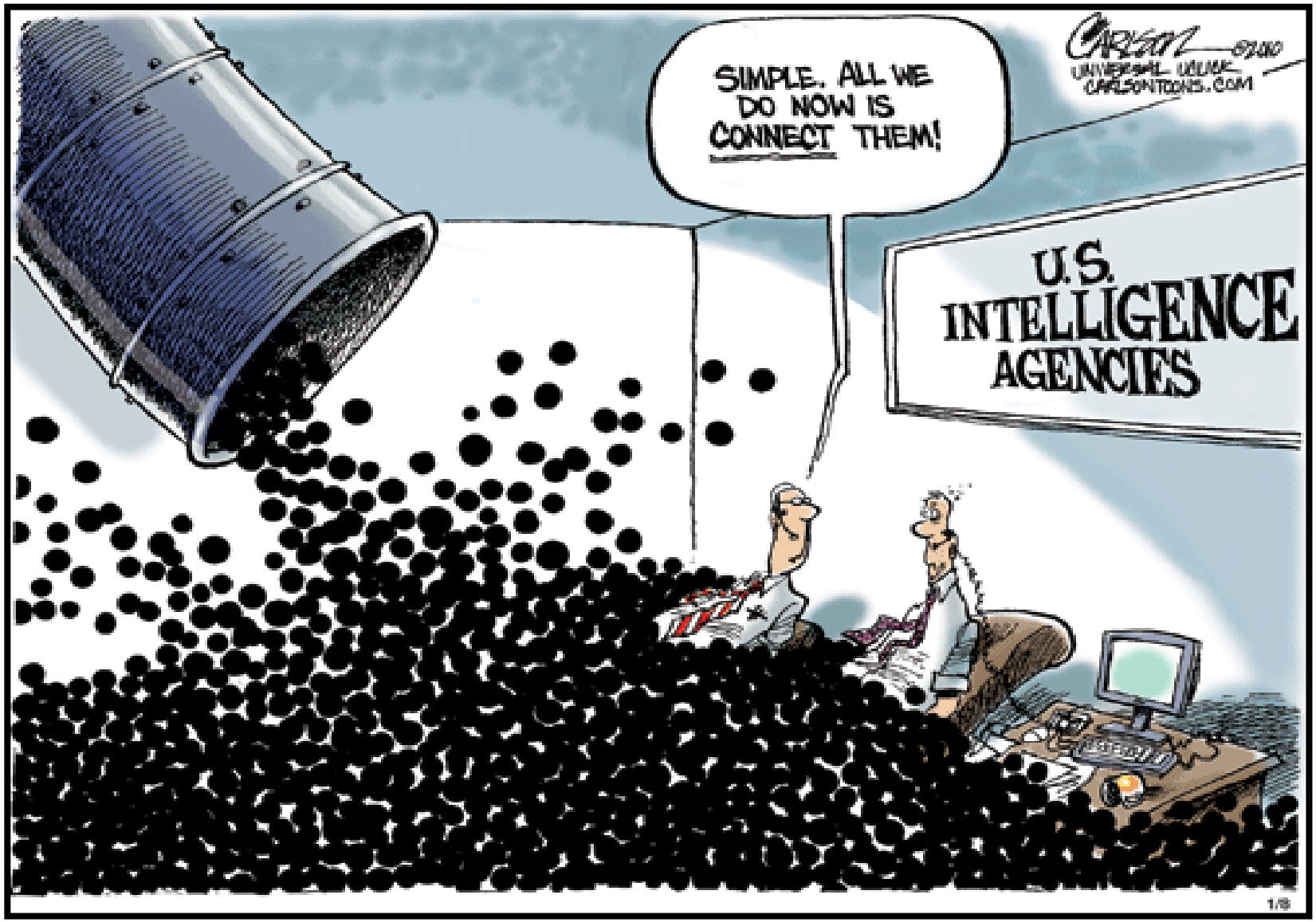
Mapping Content to Entities



Data Comes From Many Places…



Connect The Dots…

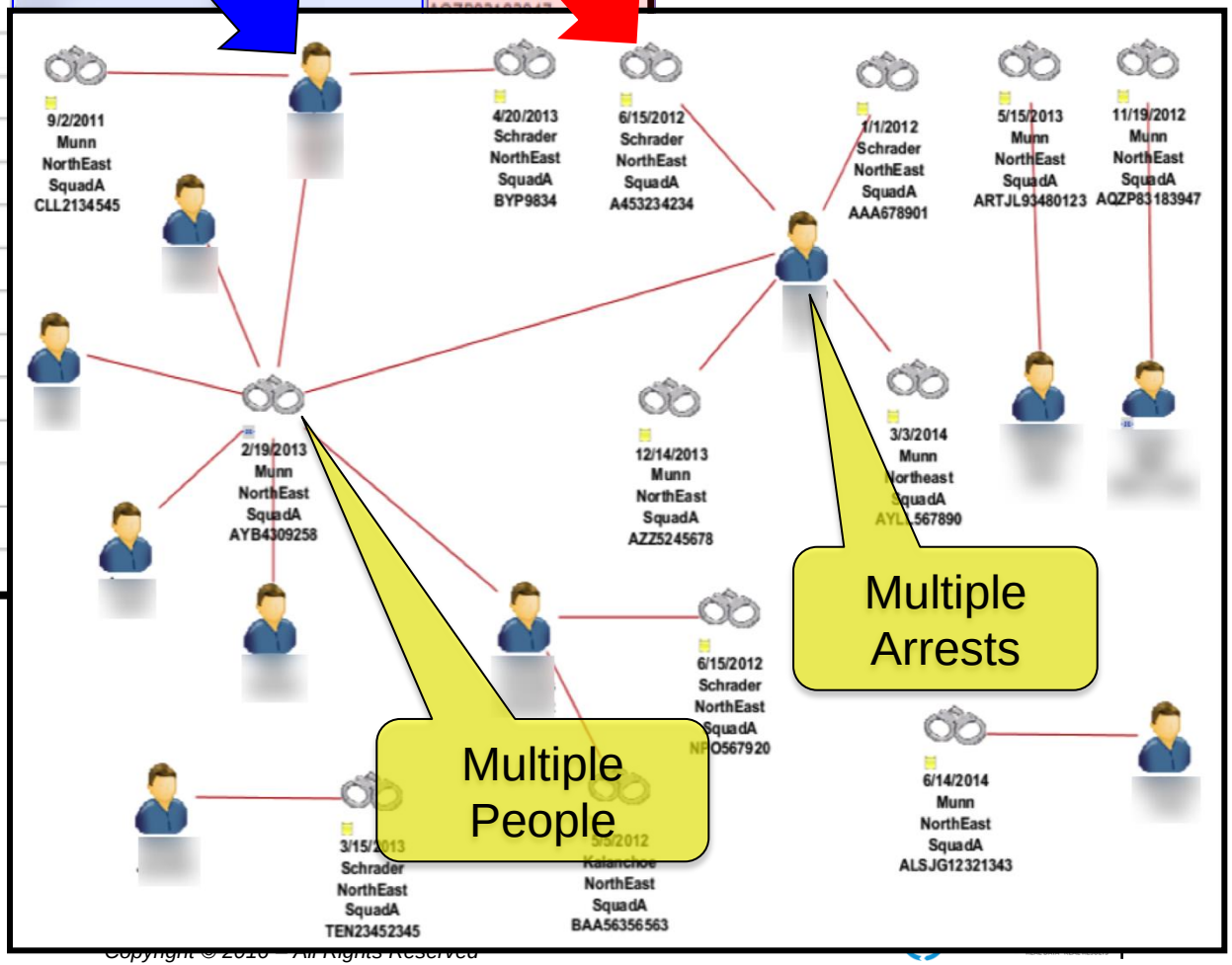


Visualizing the Connections

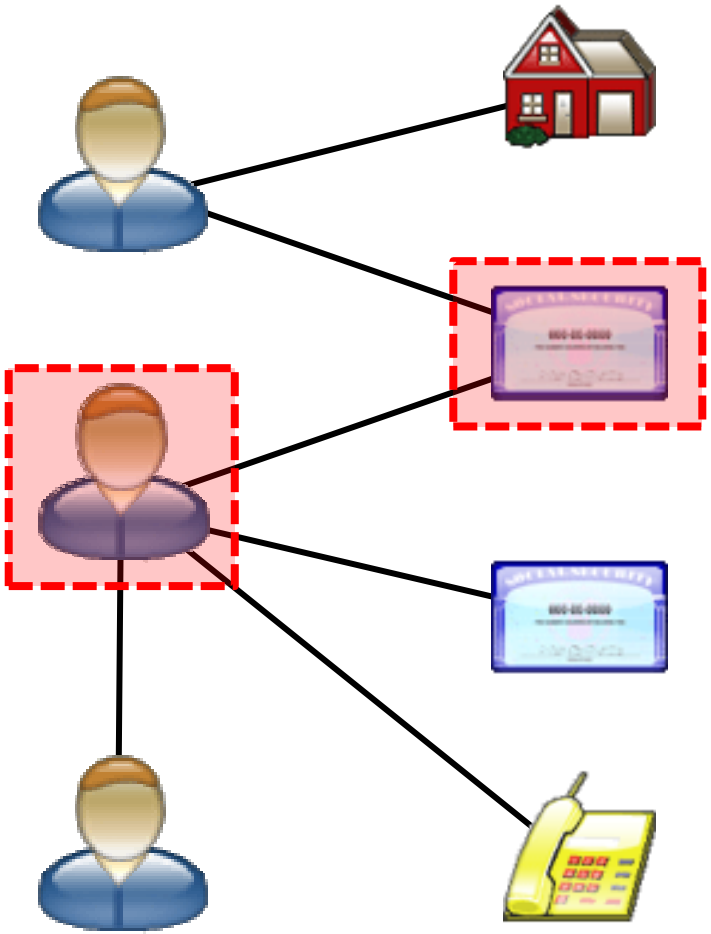
Arrest Report 2014/11/14

Result Columns: CASE_MANAGEMENT_DEMO ▾ Export: CSV

ARREST_DATE	REGION	SQUAD	DETECTIVE	ARRESTEE_LASTN	ARRESTEE_FIRSTNAME	CASE_NUMBER
15 05 2013	NorthEast	SquadA	Munn			ARTJL93480
19 11 2012	NorthEast	SquadA	Munn			
15 06 2012	NorthEast	SquadA	Schrader			
05 05 2012	NorthEast	SquadA	Kalanchoe			
01 06 2011	NorthEast	SquadA	Munn			
19 02 2013	NorthEast	SquadA	Munn			
02 09 2011	NorthEast	SquadA	Munn			
20 04 2013	NorthEast	SquadA	Schrader			
03 03 2014	Northeast	SquadA	Munn			
15 03 2013	NorthEast	SquadA	Schrader			
14 06 2014	NorthEast	SquadA	Munn			
14 12 2013	NorthEast	SquadA	Munn			
11 01 2010	NorthEast	SquadA	Kalanchoe			
15 06 2012	NorthEast	SquadA	Schrader			
01 01 2012	NorthEast	SquadA	Schrader			

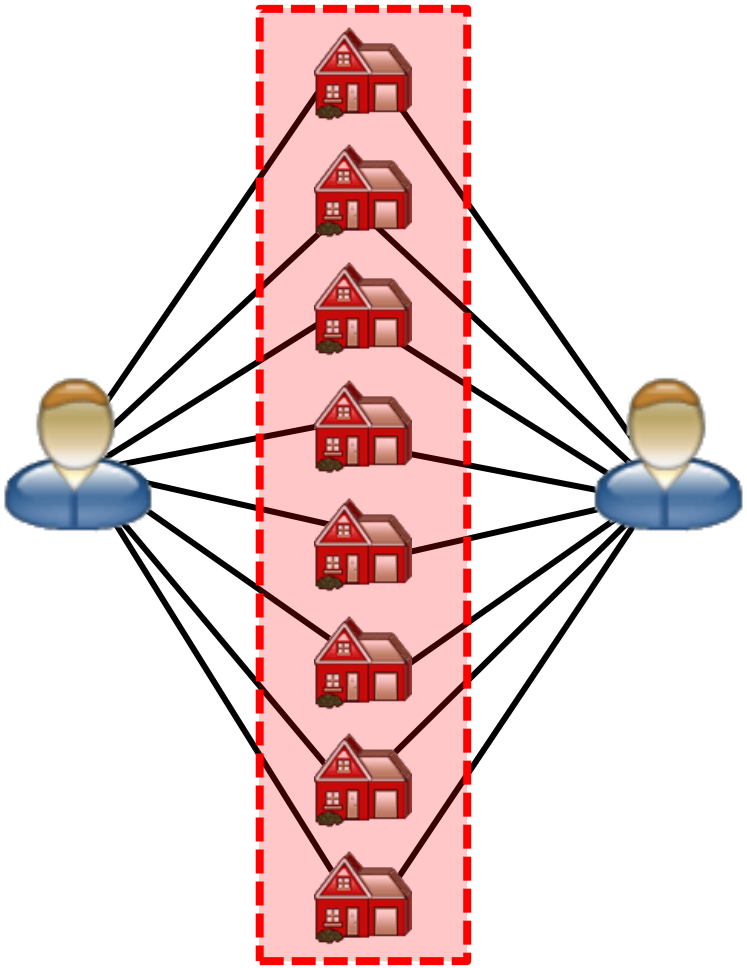


Unexpected Connections



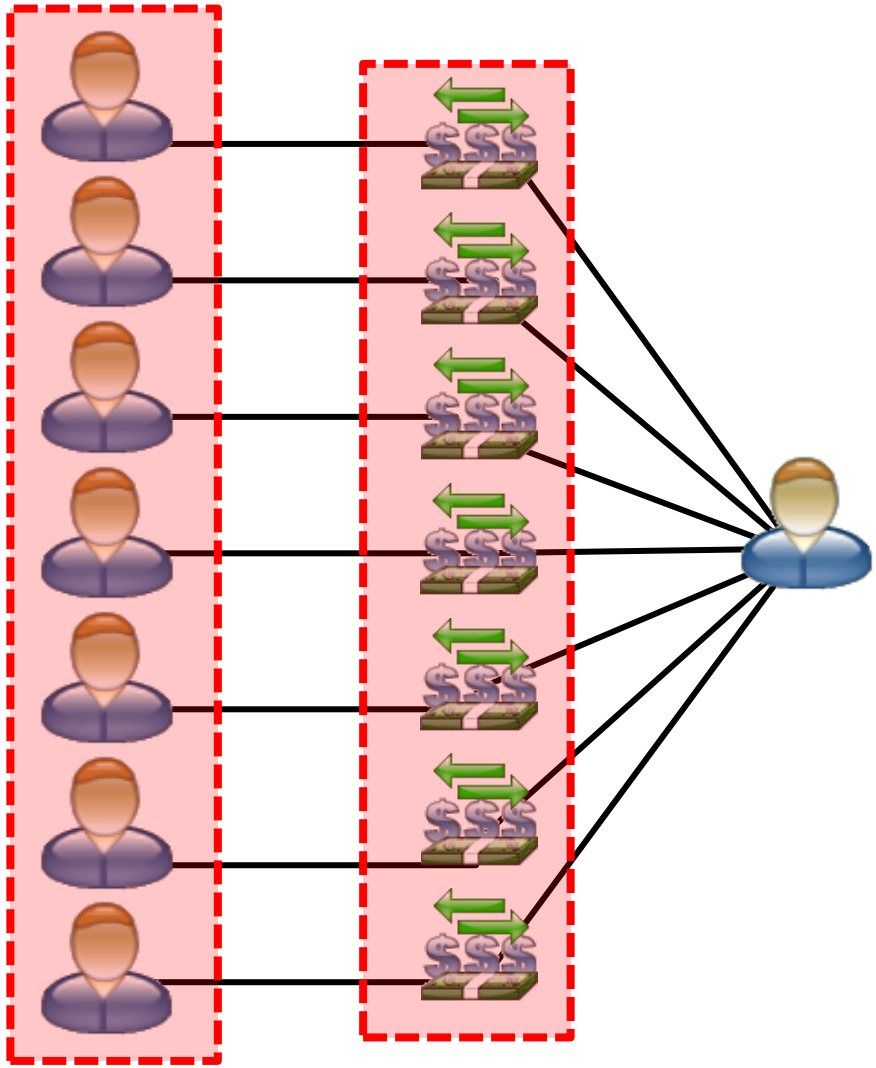
- Certain “entities” should never be shared (e.g., SSNs)
- Data prone to typos and misspellings
- Possible misrepresentation and/or falsifying data on forms
- Appearance of avoidance by varying information

Collusive Commonality



- Many patterns are exposed due to repeating behaviors
- Too many commonalities may indicate organized behaviors
- Subjects perpetrate the same crime using different information (connections)
- Only minor changes in their underlying behaviors (MO)

Accumulated Behaviors

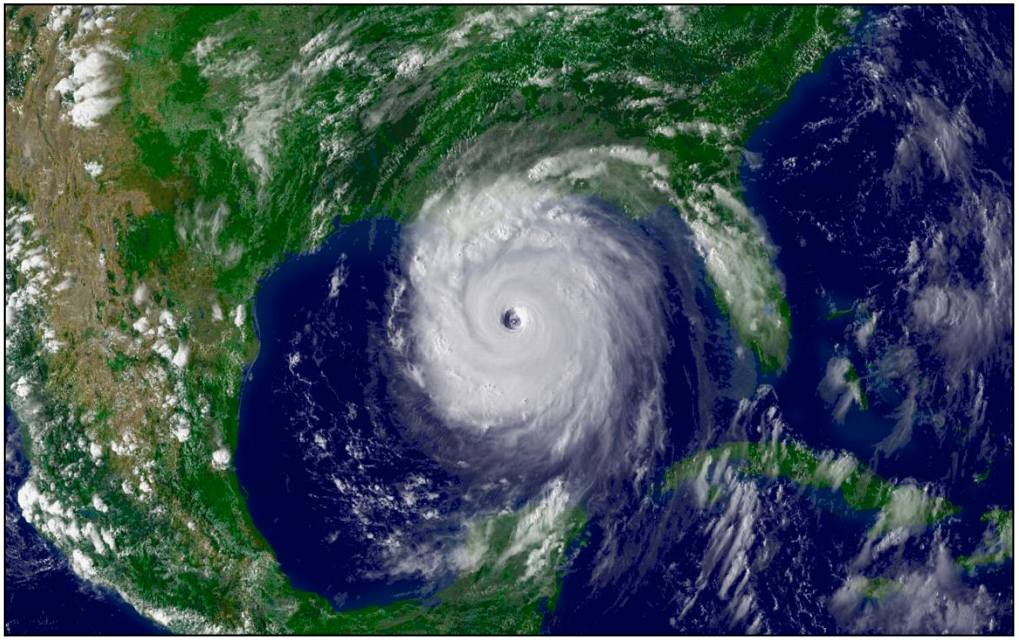


- Each unique filing looks valid – need to see it collectively
- Large numbers of discrete actions forms the bigger pattern
- Easy to avoid detection if each transaction appears legitimate
- Use mules/fronts to sustain the operations

Hurricane Katrina*

August 29, 2005
1,833 people lost their lives
1M people displaced
$120B spent post-Katrina
$75B for emergency relief
$8B distributed for…
needs assistance
housing assistance

Estimates **$1.4B** lost to fraud



- *Using FEMA hotels while receiving housing/rental allowances*
- *People making up SSNs (they did not exist / match names)*
- *Listing addresses of cemeteries / vacant lots / PO Boxes*
- *Registrants for incarcerated state/federal prisoners*

*http://www.fema.gov/louisiana-recovery-office

GAO: Questionable Debit Card Charges

http://www.gao.gov/new.items/d06844t.pdf

GAO examples of "questionable" charges for use of debit cards

Table 2: Purchases that Did Not Appear Necessary to Satisfy Immediate Emergency Needs

Vendors	Location	Nature of Transaction	Amount
Elliot's Gun Shop	Jefferson, LA	.45 caliber pistol	$1,300
D Houston	Houston, TX	Gentlemen's club	1,200
Friedman's Jewelers	Plano, TX	Diamond engagement ring	1,100
Argosy Casino	Baton Rouge, LA	7 ATM withdrawals within one day at a gambling institution	1,000
Tim Fanguy Bail Bonds	Houma, LA	Partial bail bond payment	1,000
Department of Public Safety	Baton Rouge, LA	Payment of prior traffic violations for driver's license reinstatement	700
Cat Tattoo	Addison, TX	Tattoo on arm	450
Swedish Institute	Irving, TX	Massage parlor	400
Tiger Beer and Wine	Dallas, TX	Alcohol beverages	200
Condoms To Go	Dallas, TX	Adult erotica products	150

Source: GAO analysis of debit card transactions and additional investigations.

Table 5: Examples of Questionable Use of Debit Cards

Vendor	Location	Nature of Transaction	
Jewelz	Arlington, Tex.	Diamond jewelry including watches, earrings, and a ring	$3,700
Vacation Express	Atlanta, Ga.	All inclusive 1 week Caribbean vacation resort in Punta Cana, Dominican Republic	2,200
Lesea Broadcasting	South Bend, Ind.	Donations to a faith based charity	2,000
New Orleans Saints	New Orleans, La.	5 New Orleans Saints football season tickets	2,000
Mark Lipkin	Houston, Tex.	Divorce lawyer services	1,000
Legends	Houston, Tex.	Gentleman's club	600
The Pleasure Zone	Houston, Tex.	Adult erotica products	400
Hooters	San Antonio, Tex.	Alcoholic beverages including $200 bottle of Dom Perignon champagne	300
GGW Video	Santa Monica, Calif.	Girls Gone Wild videos	300
Alamo Fireworks	San Antonio, Tex.	Fireworks	300

Louisiana Department of Labor

First Claim Submitted
- 450 miles from NOLA
- Used own name

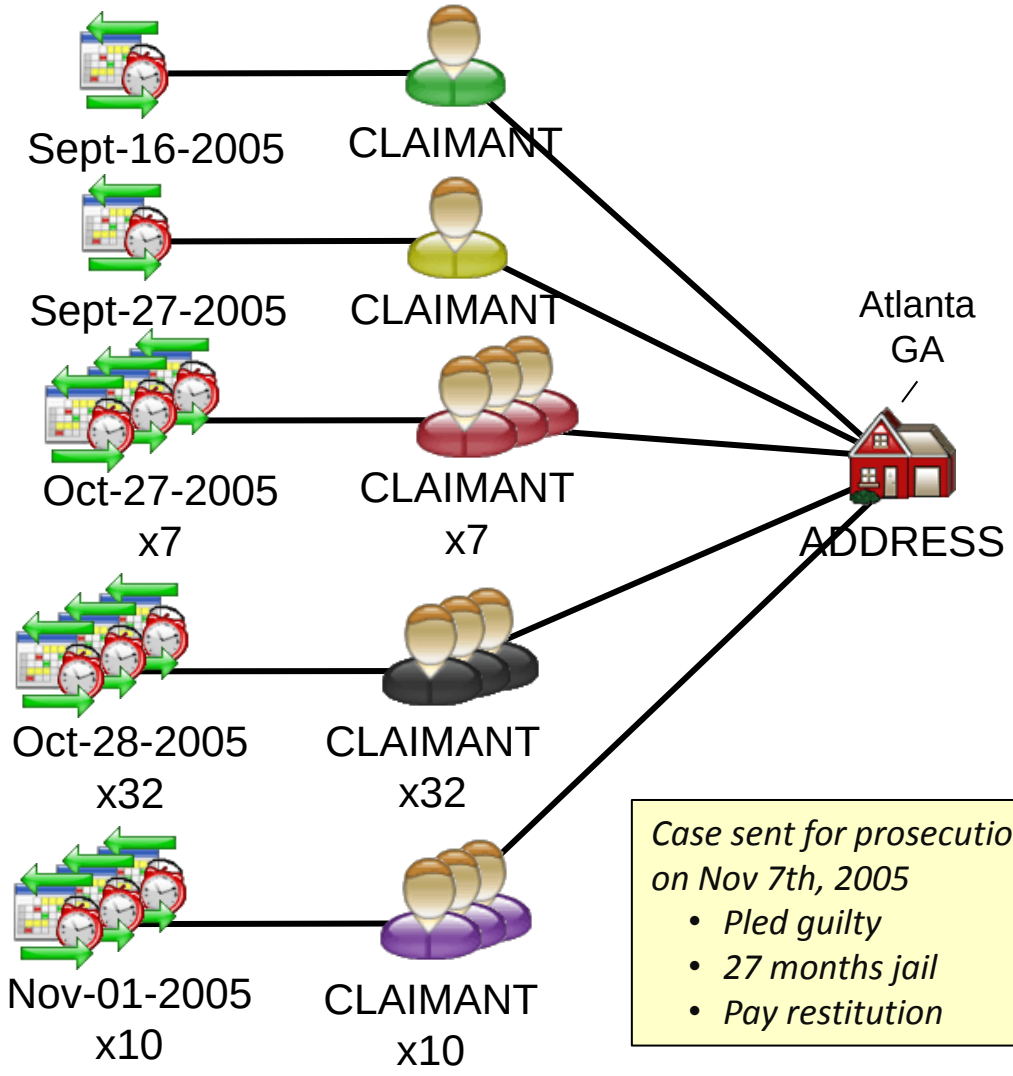
Second Claim Made
- Different name
- Same address

7 More Claims
- Same last name
- Different first name
- Same DOB
- Similar SSNs
- Same address

32 Additional Claims

10 More Claims

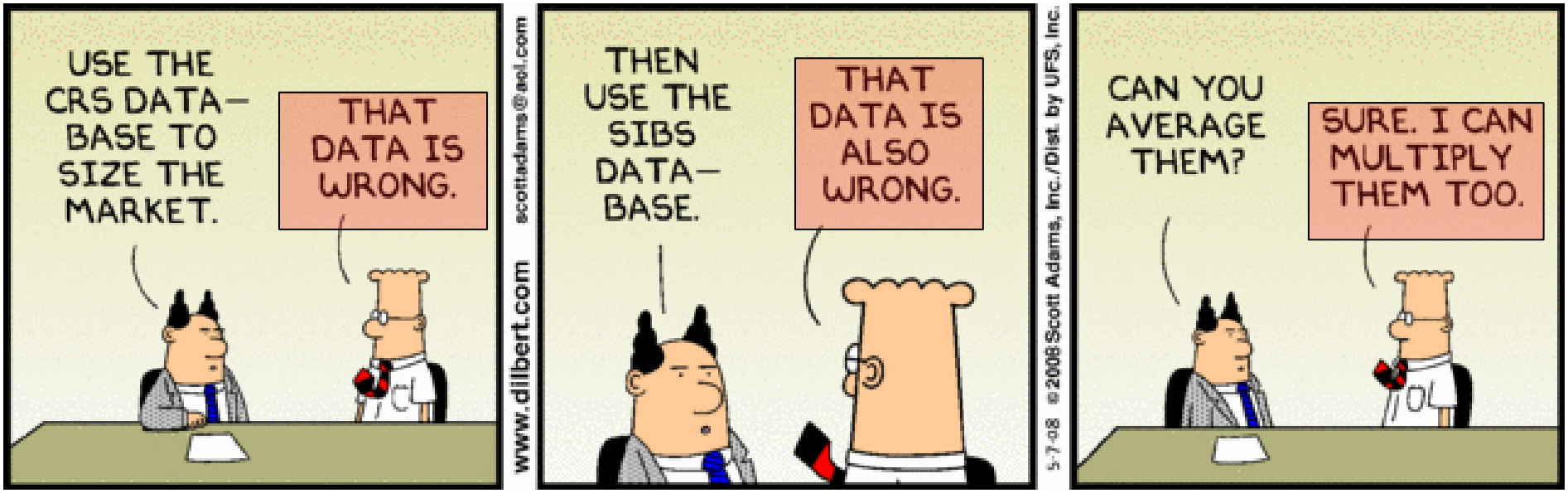
Total = 51 Debit Cards



Case sent for prosecution on Nov 7th, 2005
- *Pled guilty*
- *27 months jail*
- *Pay restitution*

Data Quality Affects Results

http://dilbert.com/strips/comic/2008-05-07/



- Poor Collection Standards
- Information Often Not Verified
- Incomplete Data
- Errors in the Data
- Misrepresentation of Data

- Inconsistencies in the Data
  - Typos/Errors
  - Misspellings
  - Transpositions
  - Transliterations/Cultural
  - Qualifiers/Prefixes/Suffixes

Medicare Fraud / Improper Payments

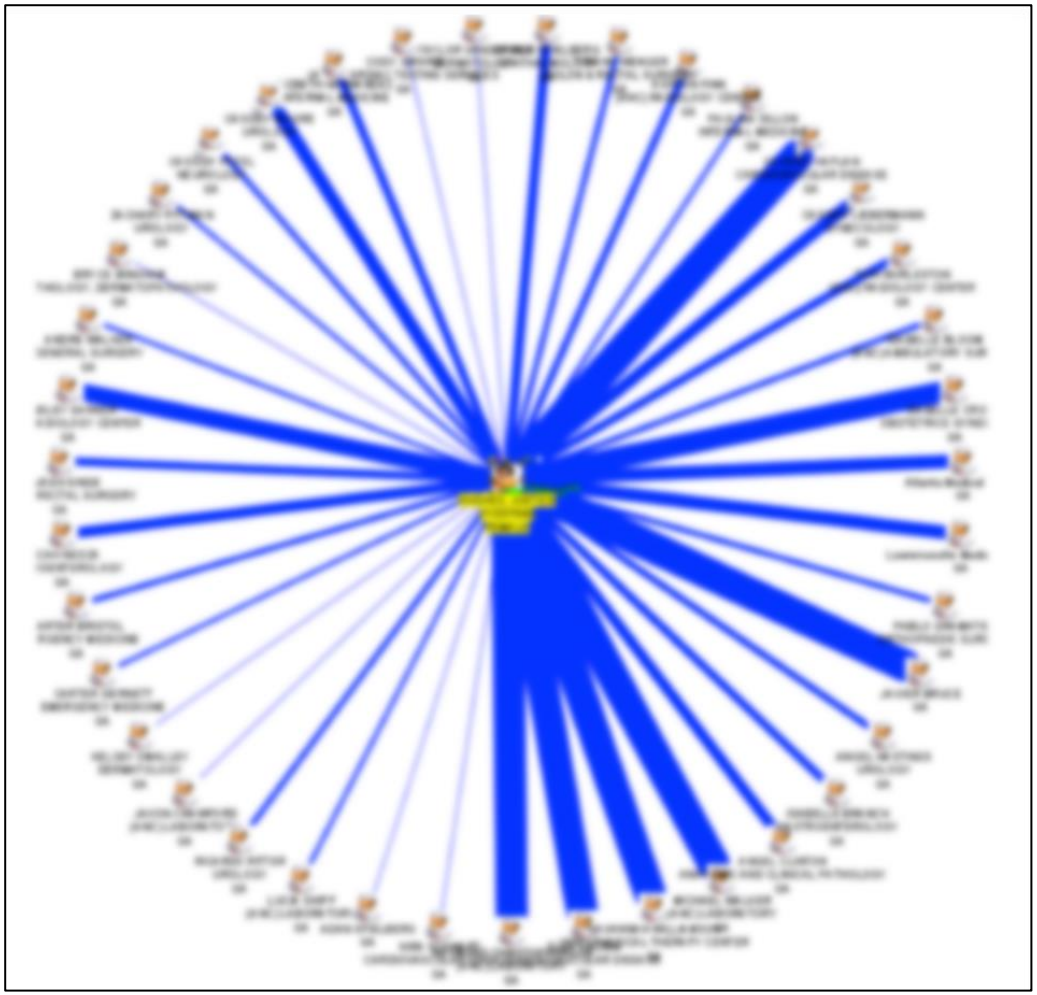
In 2014, Medicare expenditures totaled $618.7 billion

HHS – Medicare Fee-for-Service FY14 $45.75B (improper payments)

83 of the 191 countries ranked 2014 had a Gross Domestic Product (GDP) over US$45 billion

The money lost to fraud from Medicare is more than the GDP for ½ world's countries…

Excessive Medical Claims by a Single Member

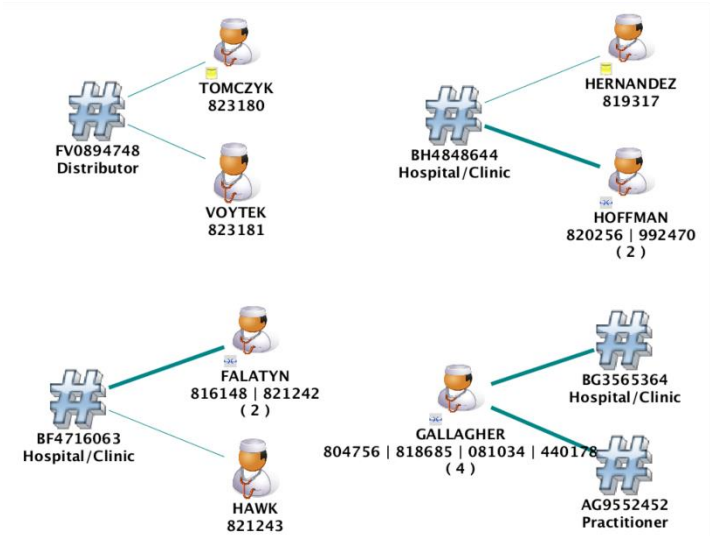


[intentionally blurred]

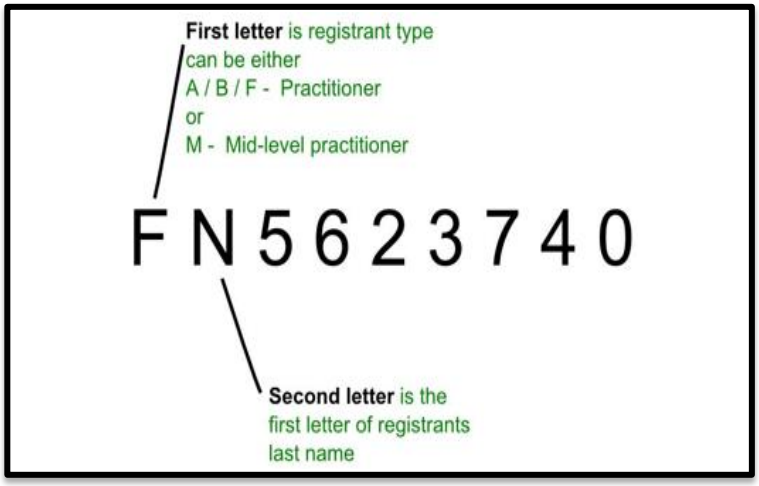
- Large utilization of provider services by a single member (~40)
- Link thickness represents repeated/frequent claims (~300)
- What is the cause?
  - Is this a very sick person?
  - Are they a hypochondriac?
  - Is this a form of identity theft?

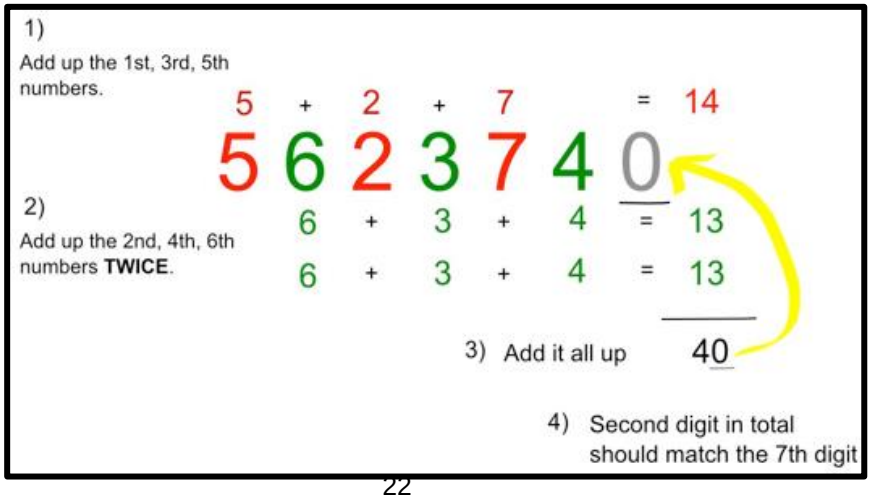
21

Questionable Connections



- Common DEA# used by multiple providers
- Several providers references duplicated in provider source





Meta Data for a Social Security Number

454 million numbers assigned
Not Valid: (000-##-####, ###-00-####, ###-##-0000)
SSN won't start with 666

Social Security Validation Check

Effective June 25, 2011, assigned randomly

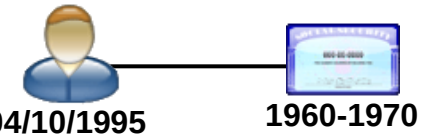
Social Security Death Master Index (Date/Name)

Appx 94 Million

Individual Taxpayer Identification Number (ITIN)

9XX-70-XXXX
9XX-99-XXXX

Social Security Calculations (Age Proximity / Location)



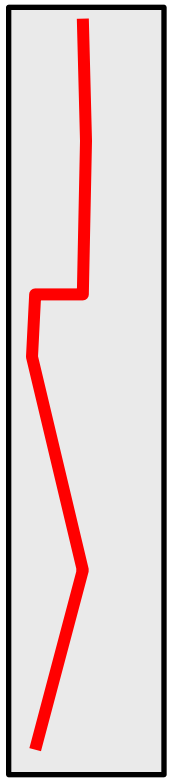
Number Verification Service

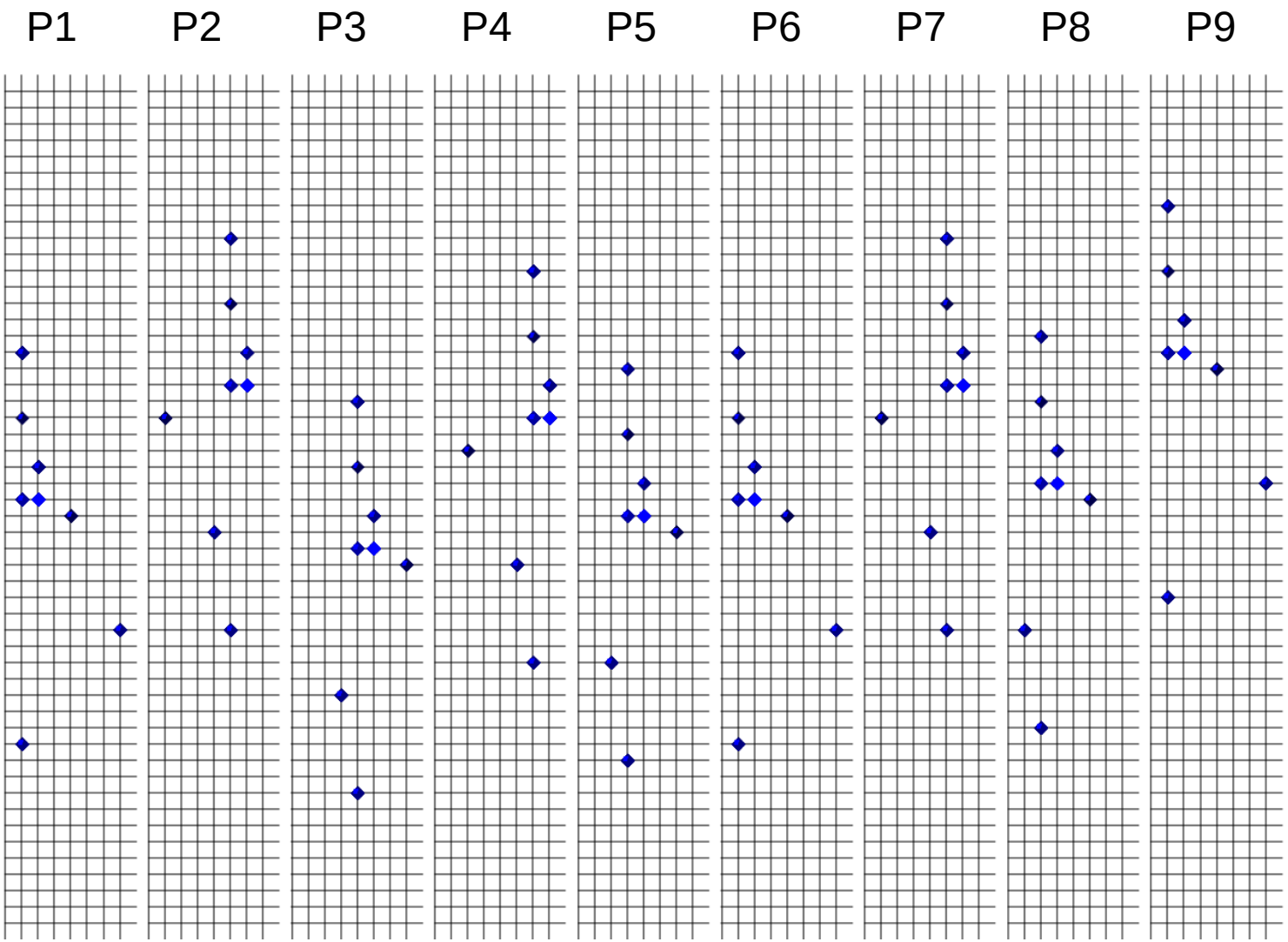
Previous Case Matches/Checks

Initial Review – Ratio Analytics

# CLAIMS	# MEMBER	ICD DESCRIPTION	PROVIDER	RATIO
825	75	LABORATORY EXAMINATION	PRV0042350	11:1
819	70	LABORATORY EXAMINATION	PRV0073294	12:1
680	62	LABORATORY EXAMINATION	PRV0067438	11:1
658	61	LABORATORY EXAMINATION	PRV0062230	11:1
624	58	LABORATORY EXAMINATION	PRV1754997	11:1
576	9	LABORATORY EXAMINATION	PRV3273933	64:1
534	71	LABORATORY EXAMINATION	PRV4120896	8:1
485	54	LABORATORY EXAMINATION	PRV0309492	9:1
459	51	LABORATORY EXAMINATION	PRV1251458	9:1
448	50	LABORATORY EXAMINATION	PRV2305050	9:1

Claim Sequencing: Reviewing 9 Entities





Weather and Environment Affects Appointments

Severe Storms

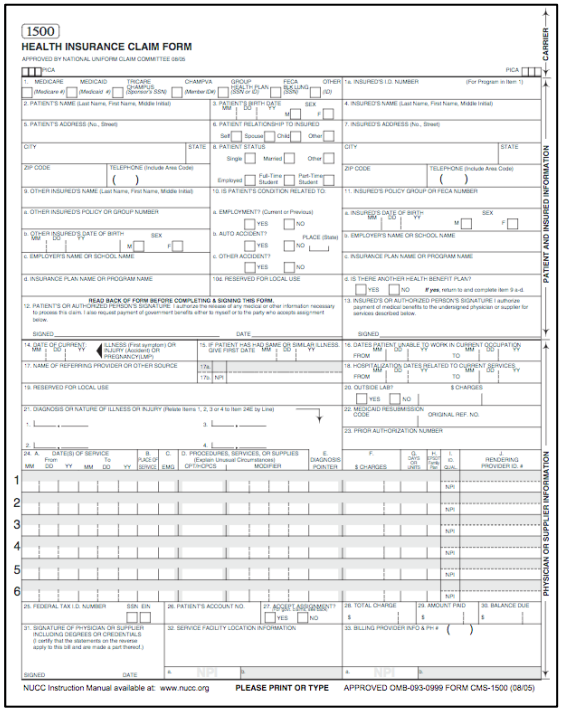
Heavy Rain

Blizzards/Snow

Hurricanes

Tornados

Extreme Temperatures



Earthquakes

Volcanoes

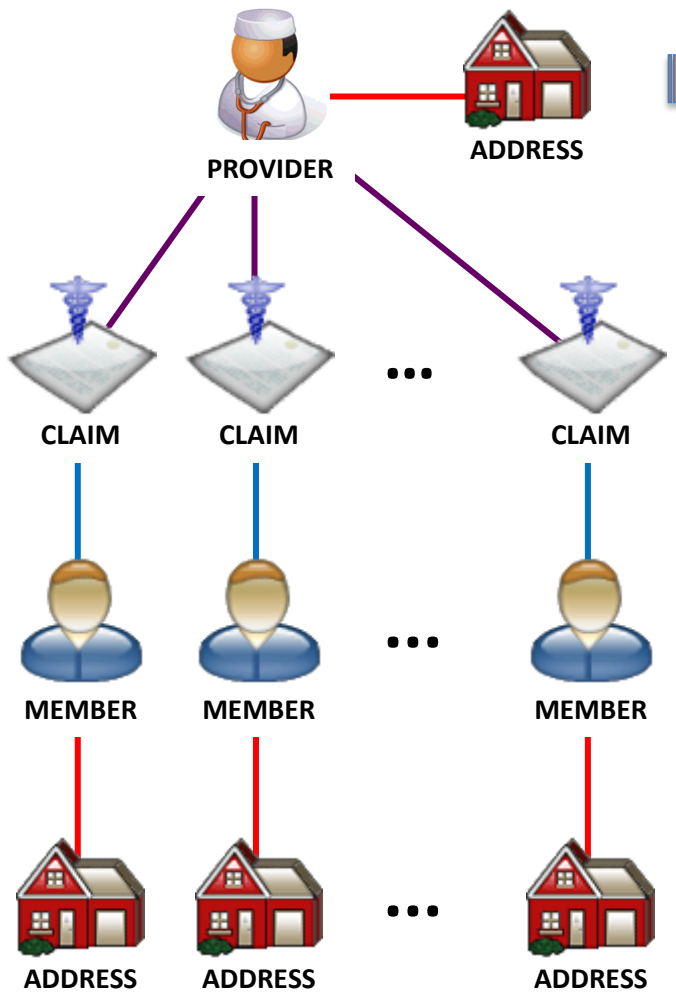
Fires / Blazes

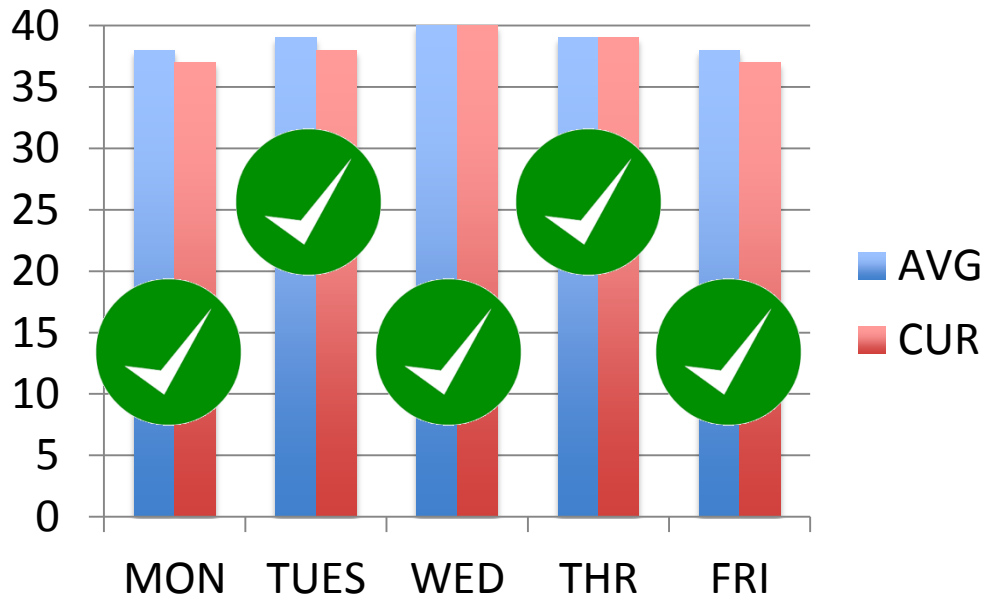
Flooding

Electrical Outages

Sickness/Vacations

Weather Related Issues - Normal

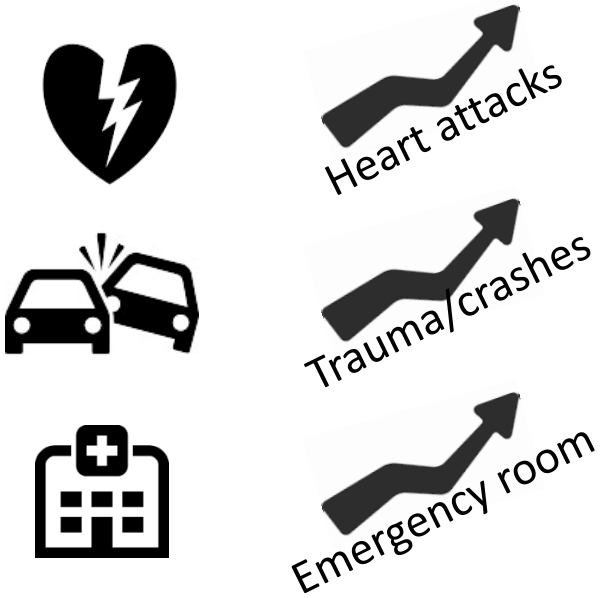


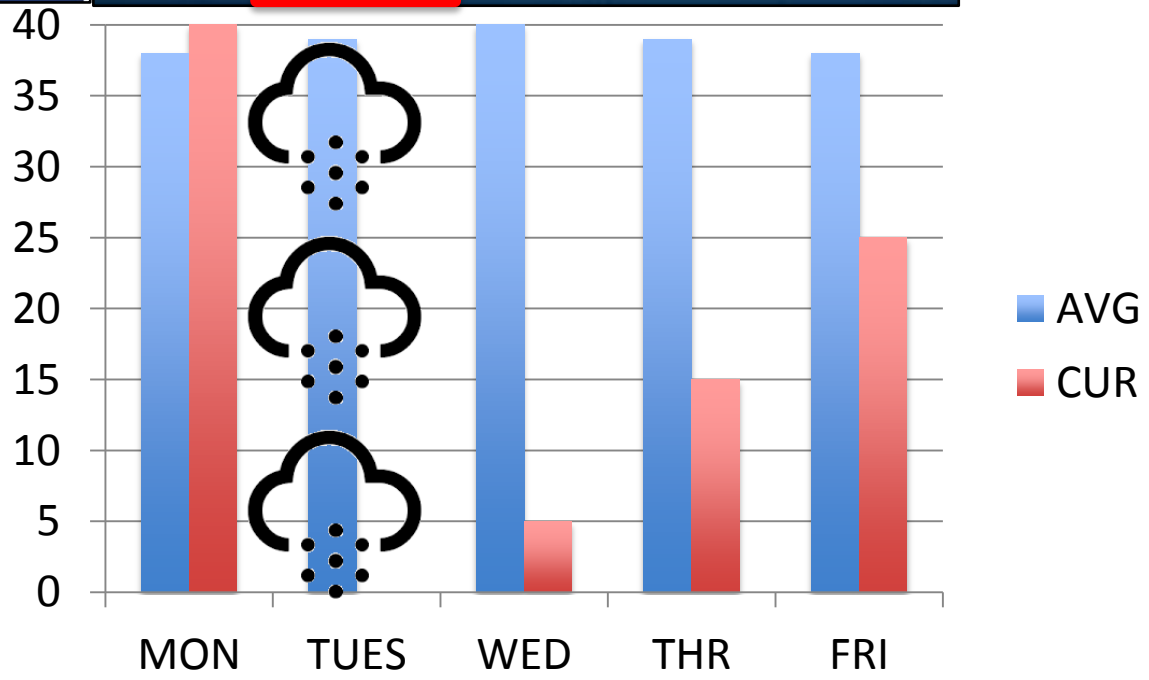


John Mininno http://www.wired.com/2016/03/john-mininno-medicare/

Weather Related Issues – Abnormal

John Mininno





http://www.wired.com/2016/03/john-mininno-medicare/

Depends on Context

What is a PDMP Program?

Note: Medicare prescription drug = $297.7 billion in 2014

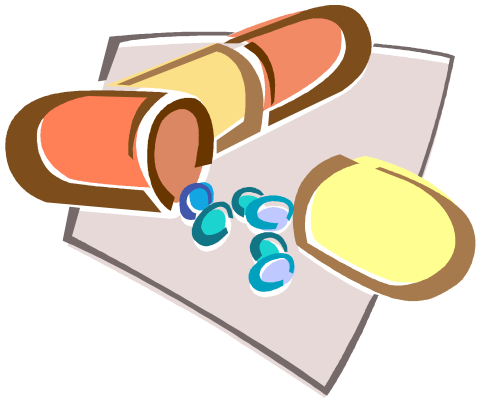
- A Prescription Drug Monitoring Program (PDMP) is operated by a state and used to collect data on Schedule II through V controlled substances from pharmacies that dispense prescriptions within the state.
  - As of Dec 2014, 49 States have PDMP
  - Missouri is only State left – pending legislation

- PDMPs are designed to support education, treatment, track trends, help with drug abuse prevention, and monitor for non-compliant behaviors of patients and practitioners.

www.namsdl.org

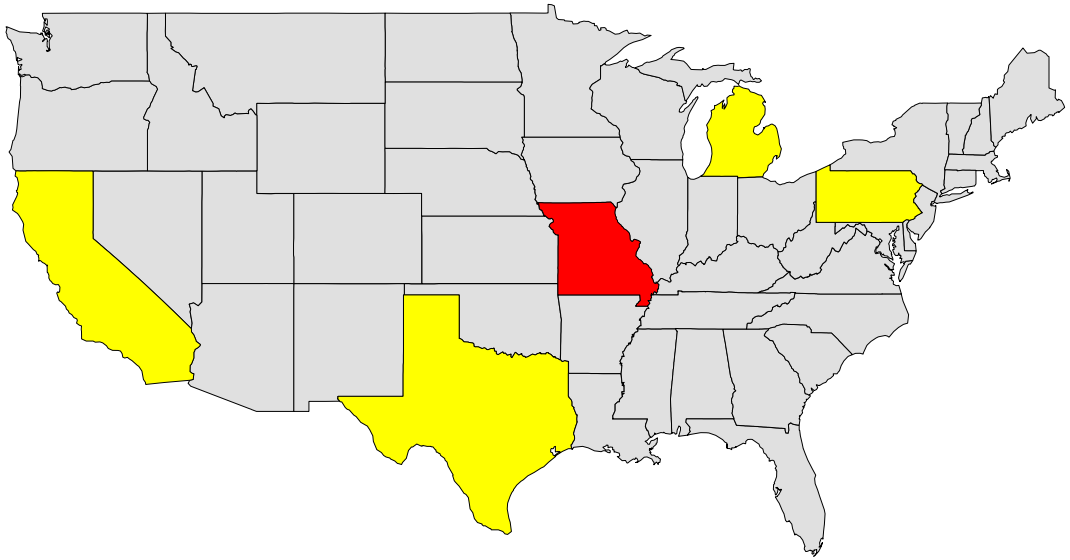
What are Schedule II Drugs?

- Schedule II drugs have a high potential for abuse and severe dependence, but have a currently accepted medical use.
- Schedule II drugs include:
  - Amphetamine (Dexedrine, Adderall, Obetrol)
  - Fentanyl (Duragesic, Oralet, Actiq)
  - Cocaine
  - Methadone (Dolophine, Methadose, Amidone)
  - Methamphetamine (Desoxyn, Crack, Speed)
  - Methylphenidate (Concerta, Ritalin)
  - Opium
  - Oxycodone (OxyContin, Percocet, Endocet, Roxicodone)
  - PCP



DEA SCHEDULE II DRUG LISTING

Sample PDMP State Profiles*



- **Michigan:**
  - Established: 1988
  - Monitors: Schedule II - V
  - Volume: 15M annual
  - Collected: Monthly
  - Dispensers: 3,000

- **California:**
  - Established: 1939
  - Monitors: Schedule II, III
  - Volume: 21M annual
  - Collected: Monthly
  - Dispensers: 155,000

- **Texas:**
  - Established: 1981
  - Monitors: Schedule II
  - Volume: 3.3M annual
  - Collected: Monthly
  - Dispensers: 4,300

- **Pennsylvania :**
  - Established: 2002
  - Monitors: Schedule II
  - Volume: 4.1M annual
  - Collected: Monthly
  - Dispensers: 3,500

* Circa 2012 from National Alliance for Model State Drug Laws (NAMSDL)

Aggregation/Summary of Data

Summary of Doctors Seen:
Identified an individual who was seen by **179** physicians within the time-frame of the Class II data.

*NODE	*COUNT OF RX_DEA_...	*GROUPED BY DATE_OF_BIRTH	*GROUPED BY FIRST_NAME
	179	11/13/1985 00:00:00 EST	ETHAN
	160	12/07/1971 00:00:00 EST	ERIC
	110		
	100	05/12/1966 00:00:00 EDT	JAMES
	87	01/18/1962 00:00:00 EST	SUSAN
	84	11/08/1974 00:00:00 EST	SHINIEH
	83	07/29/1976 00:00:00 EDT	TRACY
	77	03/13/1977 00:00:00 EST	JAMIE
	75	07/17/1962 00:00:00 EDT	SCOTT

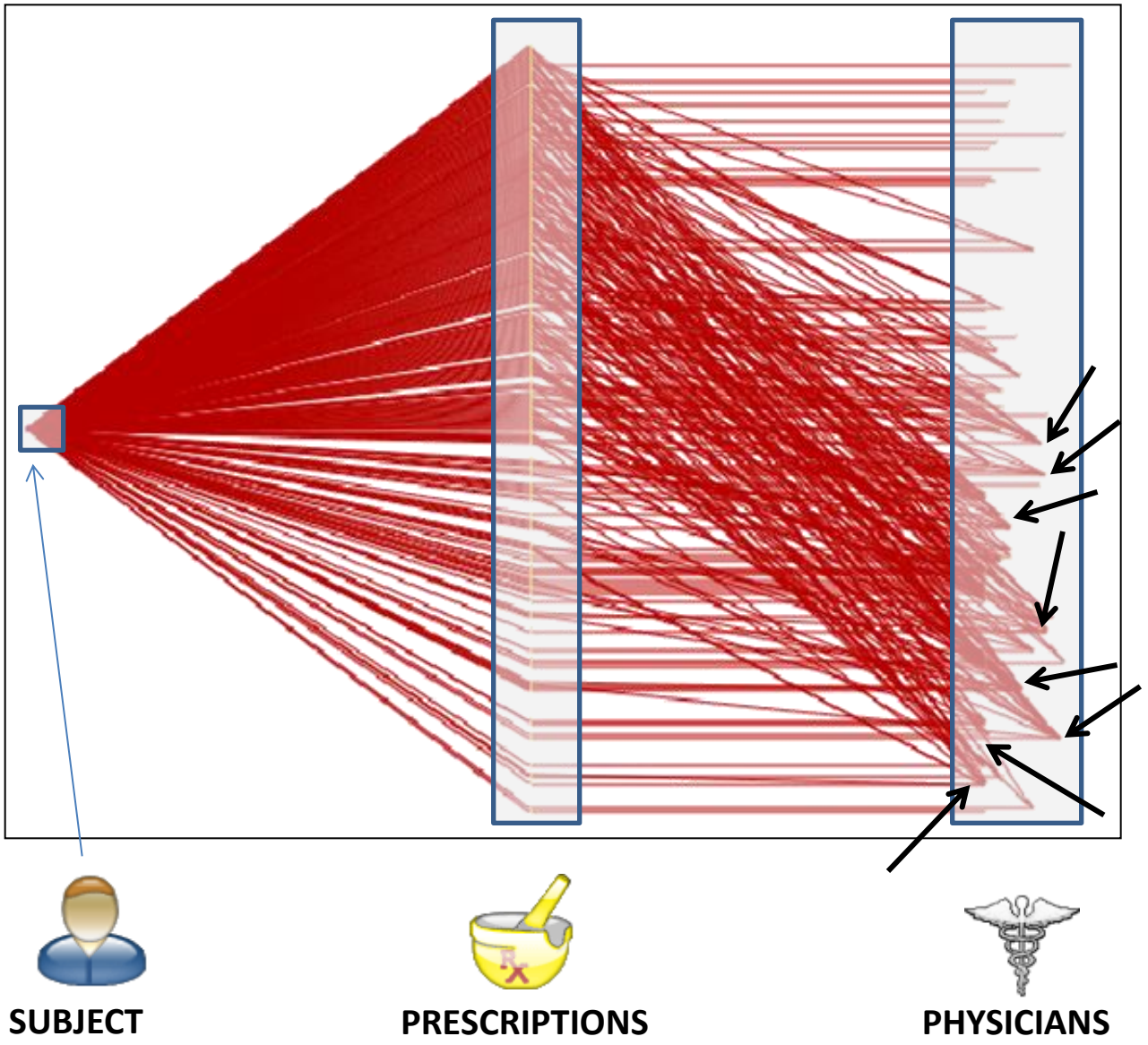
*COUNT OF RX_NU...	*GROUPED BY DATE_OF_BIRTH	*GROUPED BY FIRST_NAME
367	07/26/1954 00:00:00 EDT	GARY
338	11/13/1985 00:00:00 EST	ETHAN
259	10/31/1958 00:00:00 EST	KATHLEEN
251	12/05/1974 00:00:00 EST	DAMON
250	12/07/1971 00:00:00 EST	ERIC
248	08/24/1974 00:00:00 EDT	KATHRYN
233	10/11/1957 00:00:00 EDT	WILLIAM
211	08/28/1955 00:00:00 EDT	BERJ

Summary of Prescriptions Filled:
Identified an individual who was issued **367** prescriptions for Class II medications within the time-frame of the Class II data.

Street value of OxyContin sells for $1 per milligram and between **$40 to $80** for a single/standard tablet

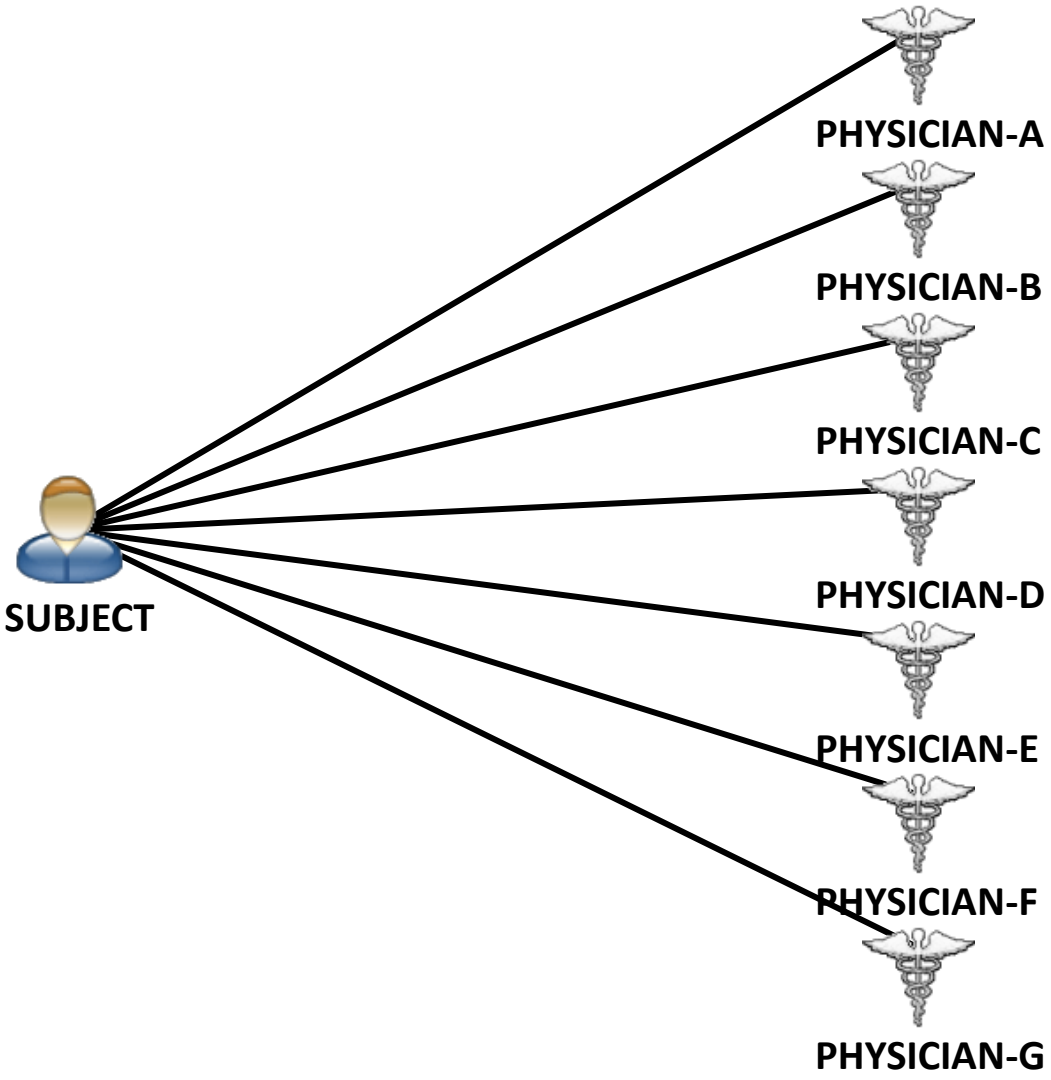
One-Many-Many

This view shows the relationship between a SUBJECT (left), to PRESCRIPTION (middle), and the related PHYSICIAN (right)



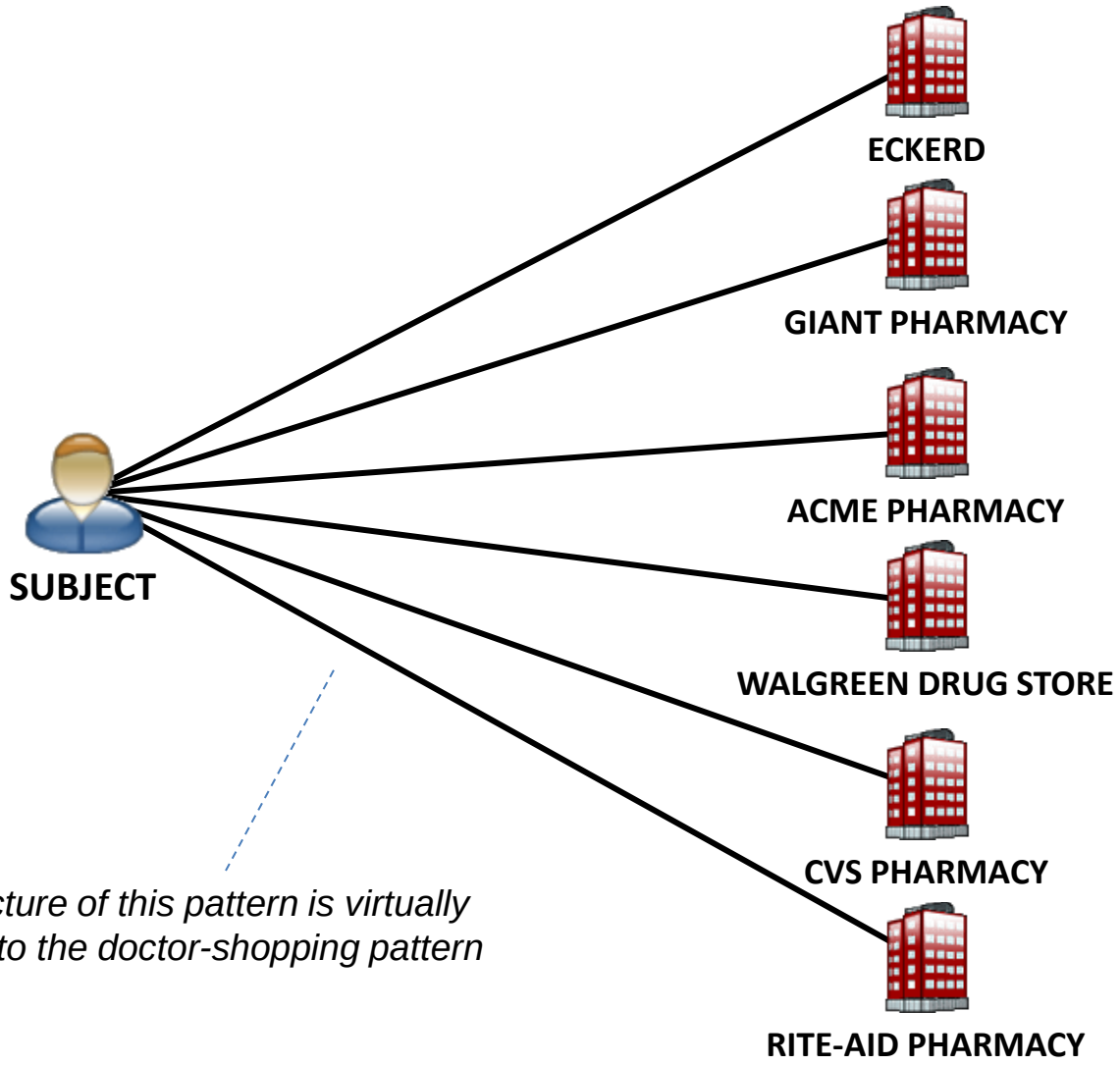
Doctor Shopping Pattern

Target suspect is related to multiple doctors for the same prescription-types



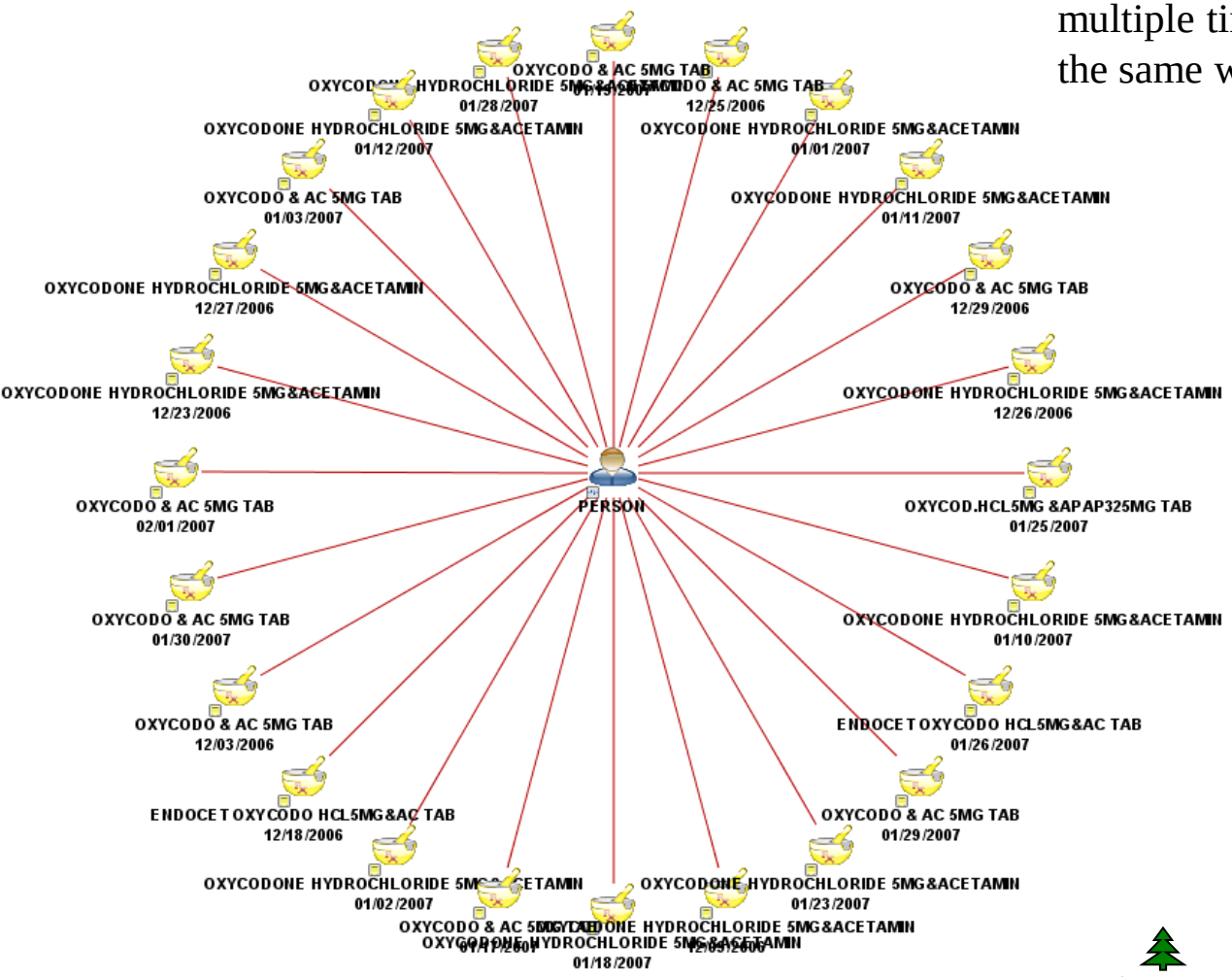
Multiple Pharmacy Usage

Target suspect uses multiple pharmacies to fill his prescriptions

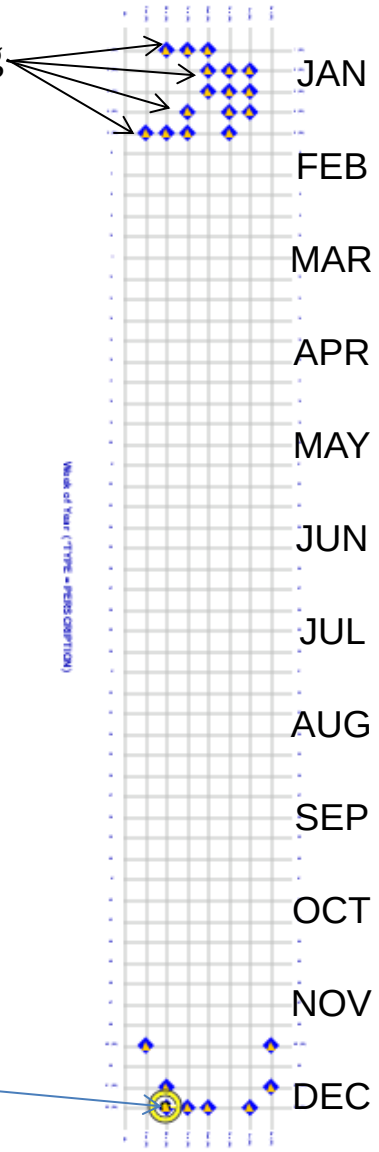


The structure of this pattern is virtually identical to the doctor-shopping pattern

Prescription Fill-Date Behavior



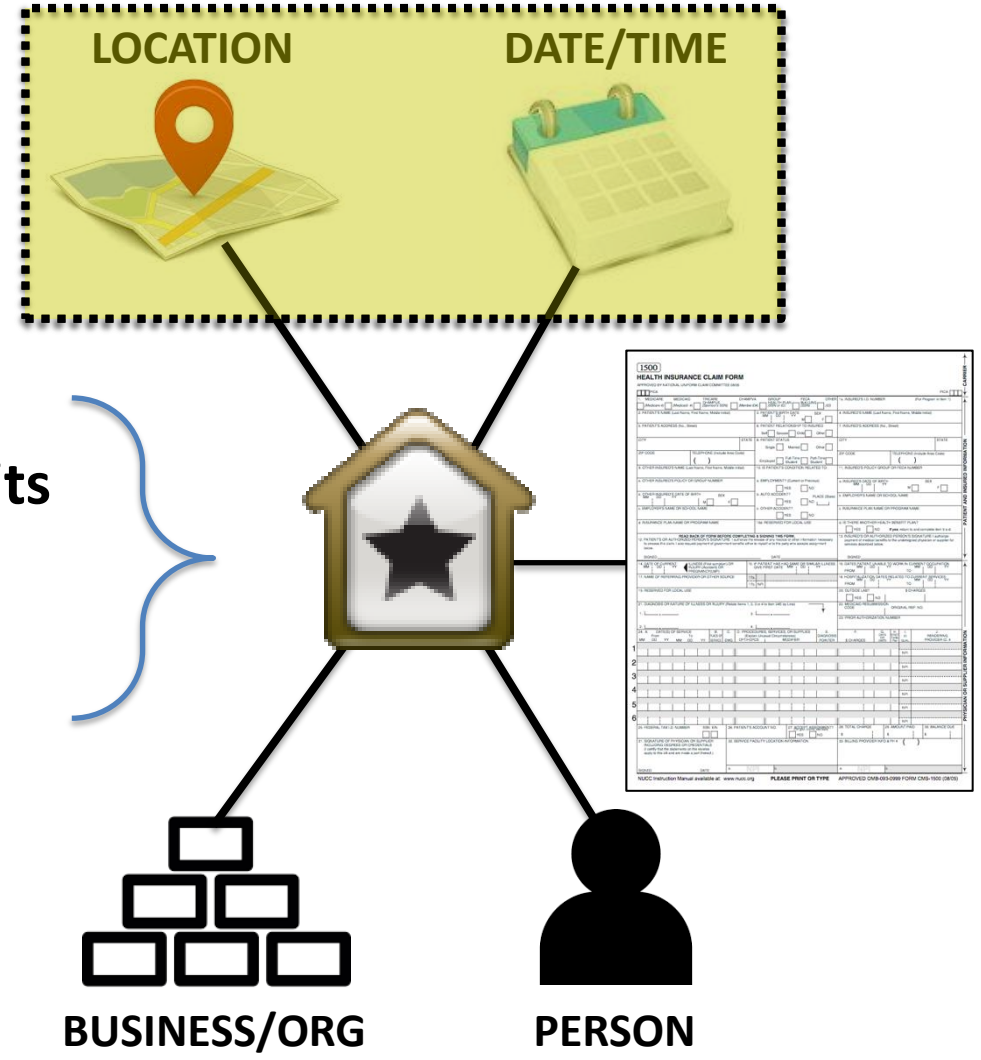
Prescription filled multiple time during the same week



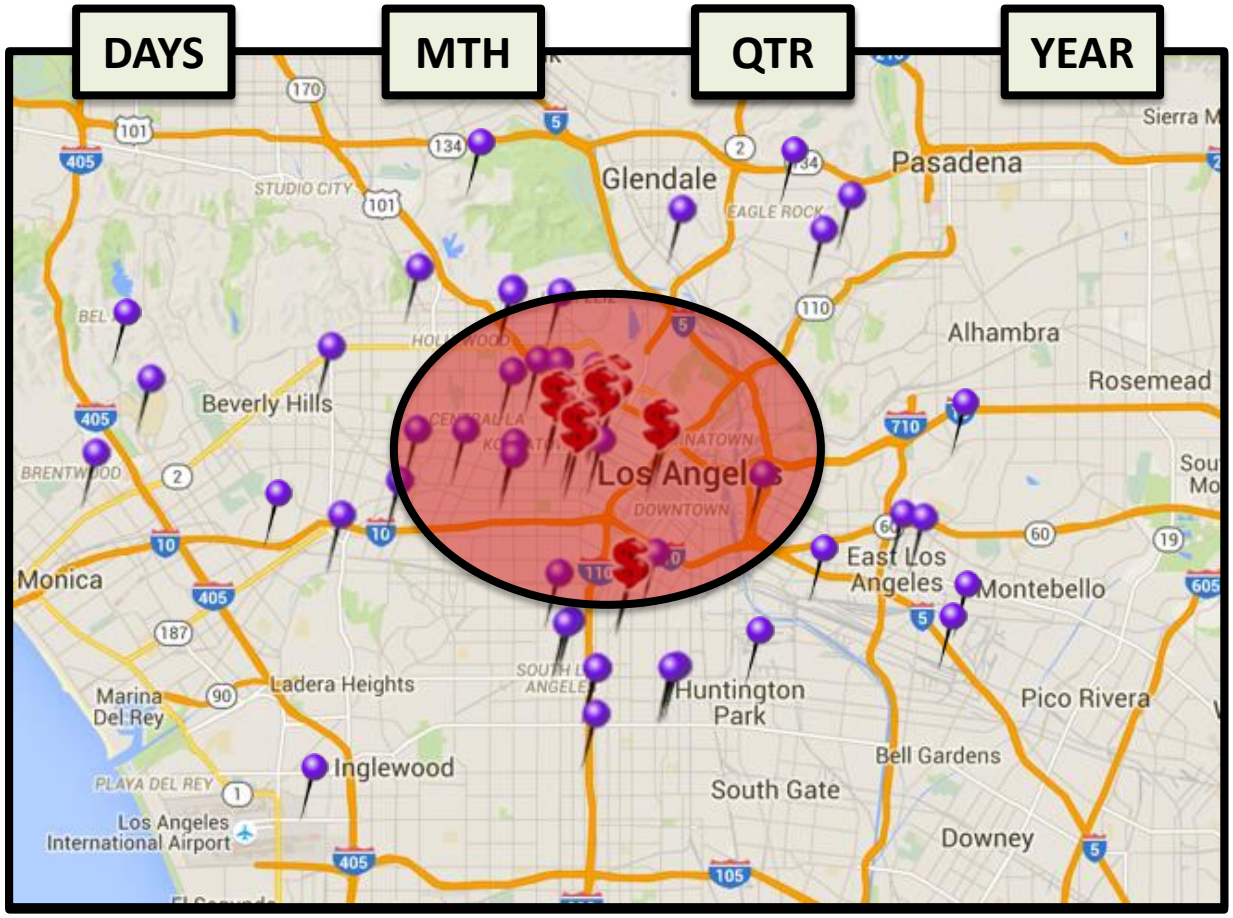
Even on Christmas Day!

Peer-to-Peer Comparisons

- Healthcare Providers
- Money Remitter Agents
- Business Tax Prep/Pay
- Supplemental Nutrition
- Household Welfare Benefits
- Property & Casualty
- Criminal Activity (LEA)
- Electricity Consumption
- Cyber Security (assets)
- Retail Stores/Shops
- Etc…



Location/Peer Comparisons



Bottom 80% of the calculated value

Top 20% (highest) of the calculated value

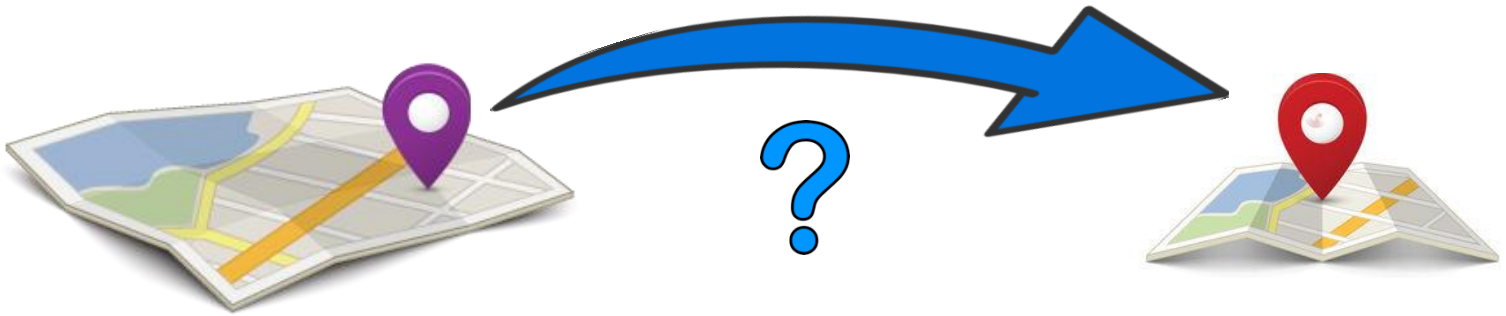
Select Date Range

Select Location

Select Type/Value

- Total # Trans
- Total $ Amount
- Number of Unique:
  - Members
  - Clients/Customers
- Other Filters
  - Occupations
  - Procedure Codes (ICD9)
- Top 10/50/100/All

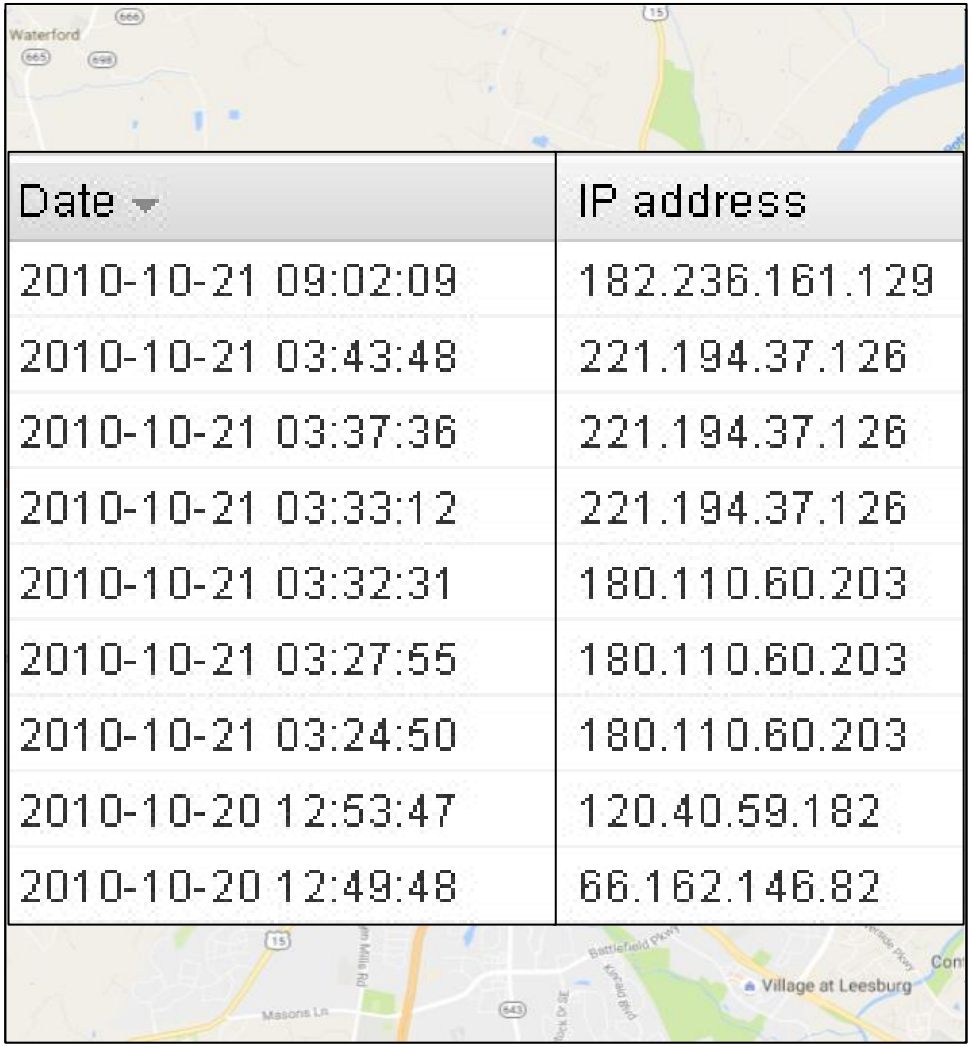
Impossible Distance Travel (IDT)



- Badge Access (Facilities, Buildings, Parking)
- Transponders (EZ-Pass, Trackers)
- Computer Logins, Accounts & VPNs
- Subscriptions (Wifi, services-Netflix)
- Social Media (Tweets/Check-Ins)
- Cell Tower – Phones/SIM cards/IoT/
- Transactions – Financial, Medical, Purchase, etc…

Impossible Distance Travel (IP Address)

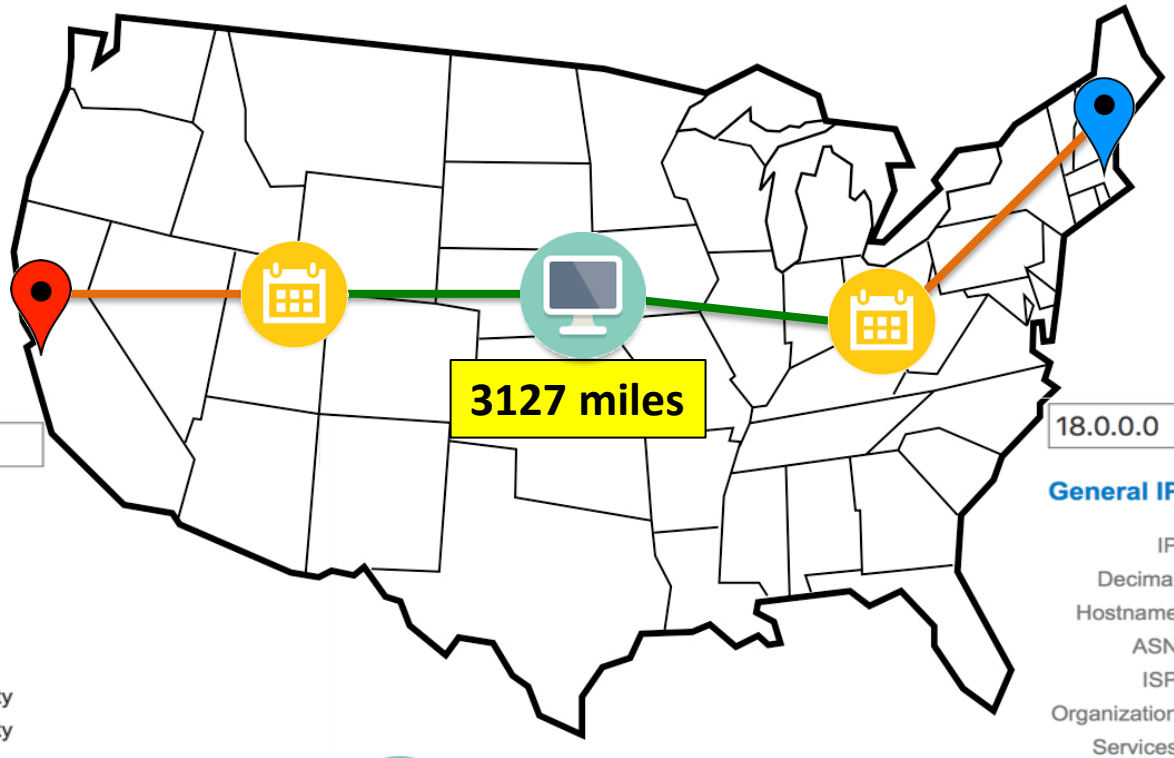
- Remote Access Logs / VPNs
  - Login IP Address
  - Time/date normalized / sync
  - UTC / NTP (network time protocol)
- Geolocation database
  - Simple: WHOIS
  - https://www.maxmind.com
  - http://www.ip2location.com/
- Exclusion lists
  - Gateways, hotels, airlines, etc.
  - Anonymizing VPN services
  - Tor Exit Nodes
  - Hosting Providers/Data Center
  - Public Proxies
- Haversine formula
  - Distance calculation
  - Lats/Longs



Date	IP address
2010-10-21 09:02:09	182.236.161.129
2010-10-21 03:43:48	221.194.37.126
2010-10-21 03:37:36	221.194.37.126
2010-10-21 03:33:12	221.194.37.126
2010-10-21 03:32:31	180.110.60.203
2010-10-21 03:27:55	180.110.60.203
2010-10-21 03:24:50	180.110.60.203
2010-10-20 12:53:47	120.40.59.182
2010-10-20 12:49:48	66.162.146.82

https://www.sans.org/reading-room/whitepapers/logging/faster-speeding-bullet-geolocation-data-account-misuse-35632

Example of Impossible Distance Travel



171.66.16.0

General IP Information

IP: 171.66.16.0
Decimal: 2873233408
Hostname: 171.66.16.0
ASN: 32
ISP: Stanford University
Organization: Stanford University
Services: None detected
Type: Broadband
Assignment: Static IP
Blacklist: Blacklist Check

Geolocation Information

Continent: North America
Country: United States
State/Region: California
City: Stanford
Latitude: 37.4178 (37° 25′ 4.08″ N)
Longitude: -122.172 (122° 10′ 19.20″ W)
Postal Code: 94305

00-15-E9-2B-99-3C

2016-08-03 10:45:22 ***171.66.16.0***

2016-08-03 10:48:27 ***18.0.0.0***

03:05

18.0.0.0

General IP Information

IP: 18.0.0.0
Decimal: 301989888
Hostname: 18.0.0.0
ASN: 3
ISP: Massachusetts Institute of Technology
Organization: Massachusetts Institute of Technology
Services: None detected
Type: Broadband
Assignment: Static IP
Blacklist: Blacklist Check

Geolocation Information

Continent: North America
Country: United States
State/Region: Massachusetts
City: Cambridge
Latitude: 42.3646 (42° 21′ 52.56″ N)
Longitude: -71.1028 (71° 6′ 10.08″ W)
Postal Code: 02139

Application for Non-Immigrant Visa

Germany

Canada

South Africa

- The date is the same [time was within 5 mins – database logs]
- The Application ID is similar/close to each other
- The IP address is captured (IRS *efile* collects this type of data)

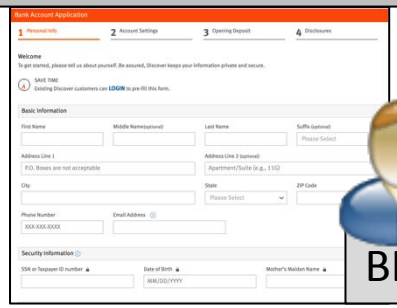
- Details of IP
  - Where is it located
  - http://dev.maxmind.com/geoip/legacy/geolite/
  - Use of proxies
  - http://dev.maxmind.com/proxy-detection/

Use of Common IP-Address via Online Applications

https://www.discovercard.com/

JOHN

https://www.discoverbank.com/

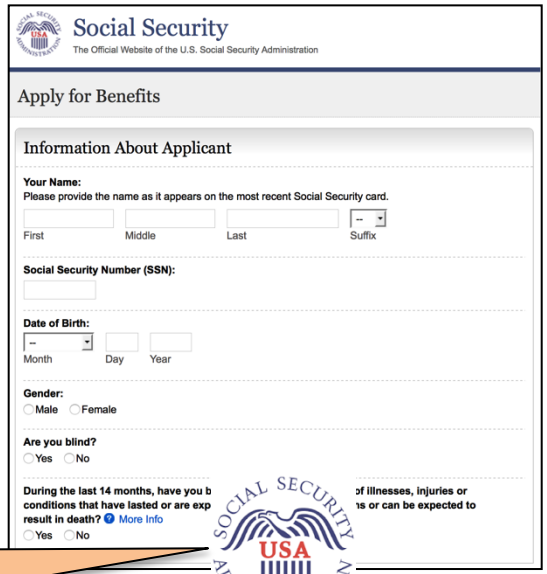


BILL

https://www.discoverpersonalloans.com

MARY

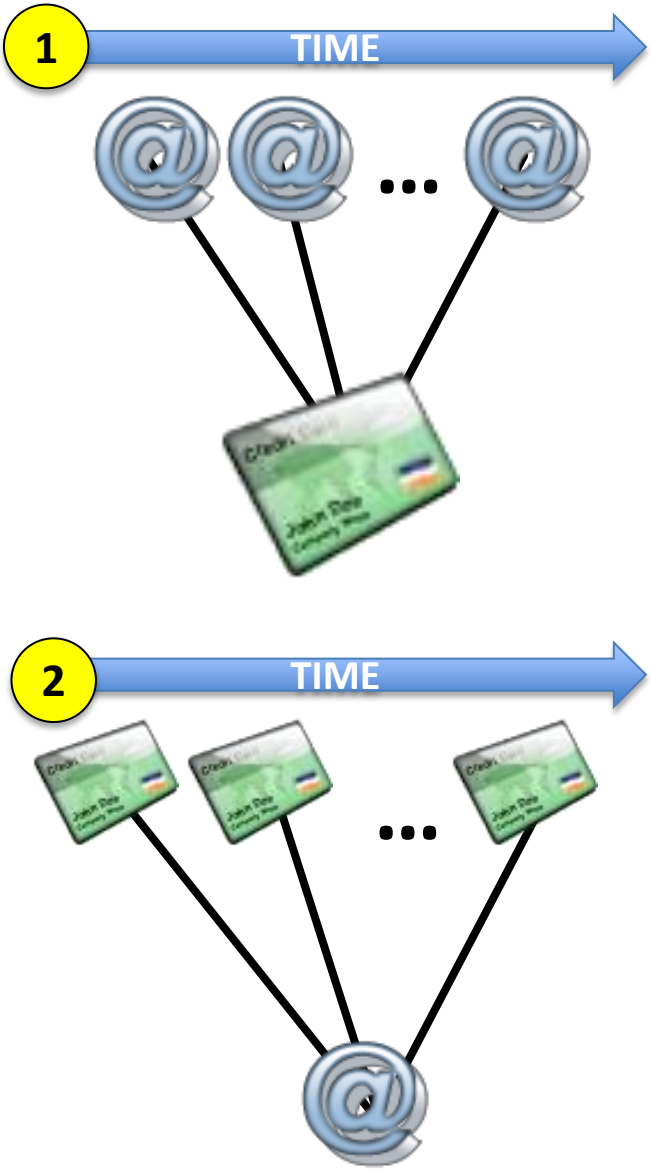
IP-ADDRESS
XXX.XXX.XXX.XXX



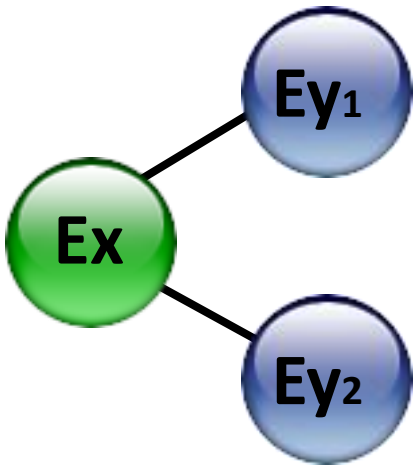
Could be any online application

- Check the date/time
  - Short duration
  - Multiple instances
- IP Domain lookup (ownership)
  - Private address (mobile/fixed)
  - Public address (Starbucks, library, etc)

Standard Representation – Commonality



Accounts applied from same IP Address
Login to multiple accounts from same IP Address
Updating multiple accounts from same IP Address



ENTITIES

- Applicants
- Accounts
- Emails
- Phones
- Transactions
- Machines
- Phones
- etc…

Using IP Info for eFile

- Internet Protocol (IP) information of the computer the Electronic Return Originator (ERO) uses to prepare the return (or originate the electronic submission of collected returns) must be included in all individual income tax returns. The required IP information includes:
  - Public/routable IP address
  - IP date / IP time
  - IP time zone

The ERO is usually the first point of contact for most taxpayers filing a return using IRS *e-file*.

The IRS requires the Internet Protocol (IP) address of the computer from which the return originated…

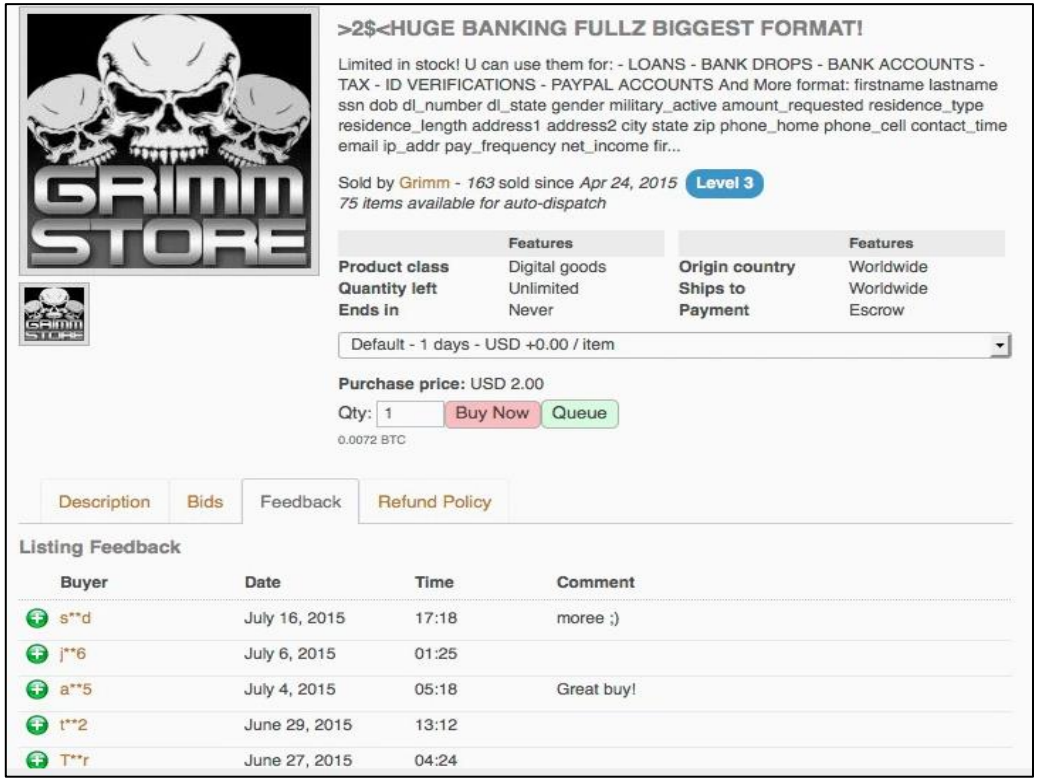
http://www.irs.gov/uac/Submitting-the-Electronic-Return-to-the-IRS

Stolen Identity Tax Refund Fraud - SIRF

1

Purchase List of Stolen Identities

- *Dentist or doctor offices ($1000 for 100 IDs)*
- *Hospitals and nursing homes*
- *Death lists (state/SSDMI)*
- Dark Web Marketplaces Median = **$21.35**



2014 Report
Steve Kroft - CBS
Corey Williams - Fraud

$21B
2016

http://www.cbsnews.com/news/tax-refund-scam-60-minutes-steve-kroft/
http://qz.com/460482/heres-what-your-stolen-identity-goes-for-on-the-internets-black-market/

Stolen Identity Tax Refund Fraud - SIRF

2

Use online Tax Preparation Site/App

- Social Security Number (SSN)
- Date of Birth (DOB)
- Fake/Bogus W2 (make up content)

H&R BLOCK

FreeTaxUSA®

TaxSlayer.com

Liberty Tax Service

TaxACT

CompleteTax® Giving You More.

efile.com Less Taxing. More Value.

e-smart tax BY LIBERTY TAX

Typically SIRF perpetrators file the false returns electronically, early in the tax filing season so that the IRS receives the false SIRF return before legitimate taxpayers have time to file their returns. By law, IRS has 6 weeks to process

Stolen Identity Tax Refund Fraud - SIRF

3 Tell the IRS where to send the money

Address **Bank Account** **Prepaid Debit Card**

From Corey Williams experience –

- Refunds were paid approximately 40% of time
- Takes 7 days for the check to arrive
- Seen 25 checks arrive in the mail to one address
- Also serving 40 months in prison…

IRS does not check for multiple refunds to same address
Does not match requested refund-returns with legitimate W2 forms
Law requires funds checks sent out within 6 weeks of request
If bogus filing is received before legitimate, the payment goes to the crooks

Green Dot Cards MoneyPak (discontinued 03/15)

- **Advance Payment Scams**
  - grants/loans/lotteries pay fees/taxes/commissions
- **Auction/Sale Scams**
  - goods never sent
- **FBI / IRS / Sheriff / LEA**
  - threaten arrest/jail for missed jury duty, taxes, etc
- **Imposter Scams**
  - friend/relative traveling in foreign country need help
- **Job Scams**
  - pay for uniforms/background/equipment
- **Inheritance Scams**
  - Nigerian prince
- **Romance Scams**
  - mail-order bride need dress / cover travel expenses
- **Utility Scams**
  - service will be cut off unless payment made

New -> Reload @ the Register

Profile of Fraud- Sift Science

- 1.3 Million online transactions/profiles
- August 2014 – August 2015 [US-only]
- Various industries (online orders and purchases)

"The Fraudiest Person in America."

www.siftscience.com

http://start.siftscience.com/hubfs/eBooks_and_White_Papers/United-States-Of-Fraud-Report.pdf

LOCATION

DEVICES

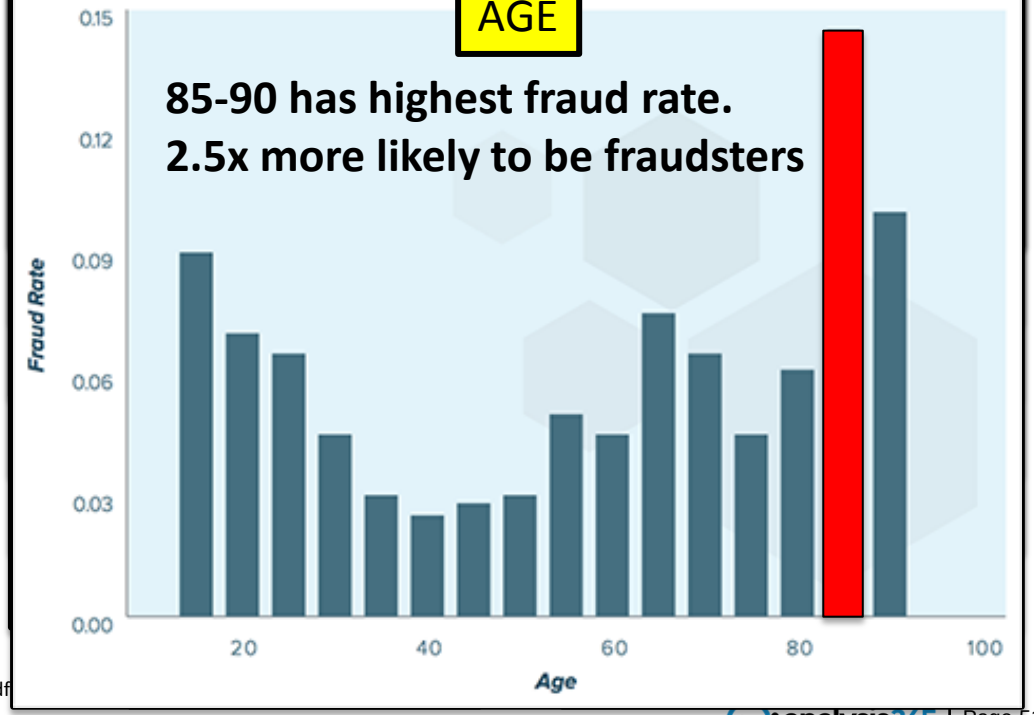
ACCOUNTS

GENDER

AMOUNT

TIME

AGE



Some Lessons Learned

- Simple solutions can yield good/decent results
- Don't necessarily need complex/expensive software
- Combine data sources to show more/better patterns
  - Simple – weather (Mininno), SSDMI,
  - Complex – case management, core systems
- Metadata is a key feature for exposing patterns
  - Dates/times (behavior)
  - Lookup tables (watch lists, locations, SSDMI)
  - Custom heuristics and rules (DR #, DEA #, etc)
- Collaboration is important
  - Share results when possible
  - Patterns will be similar across systems
  - May find same people compromising multiple systems