DATA ANALYTICS AND IMPROPER PAYMENTS

Nick Purse, JD, CIA

Utah Office of the State Auditor (OSA)

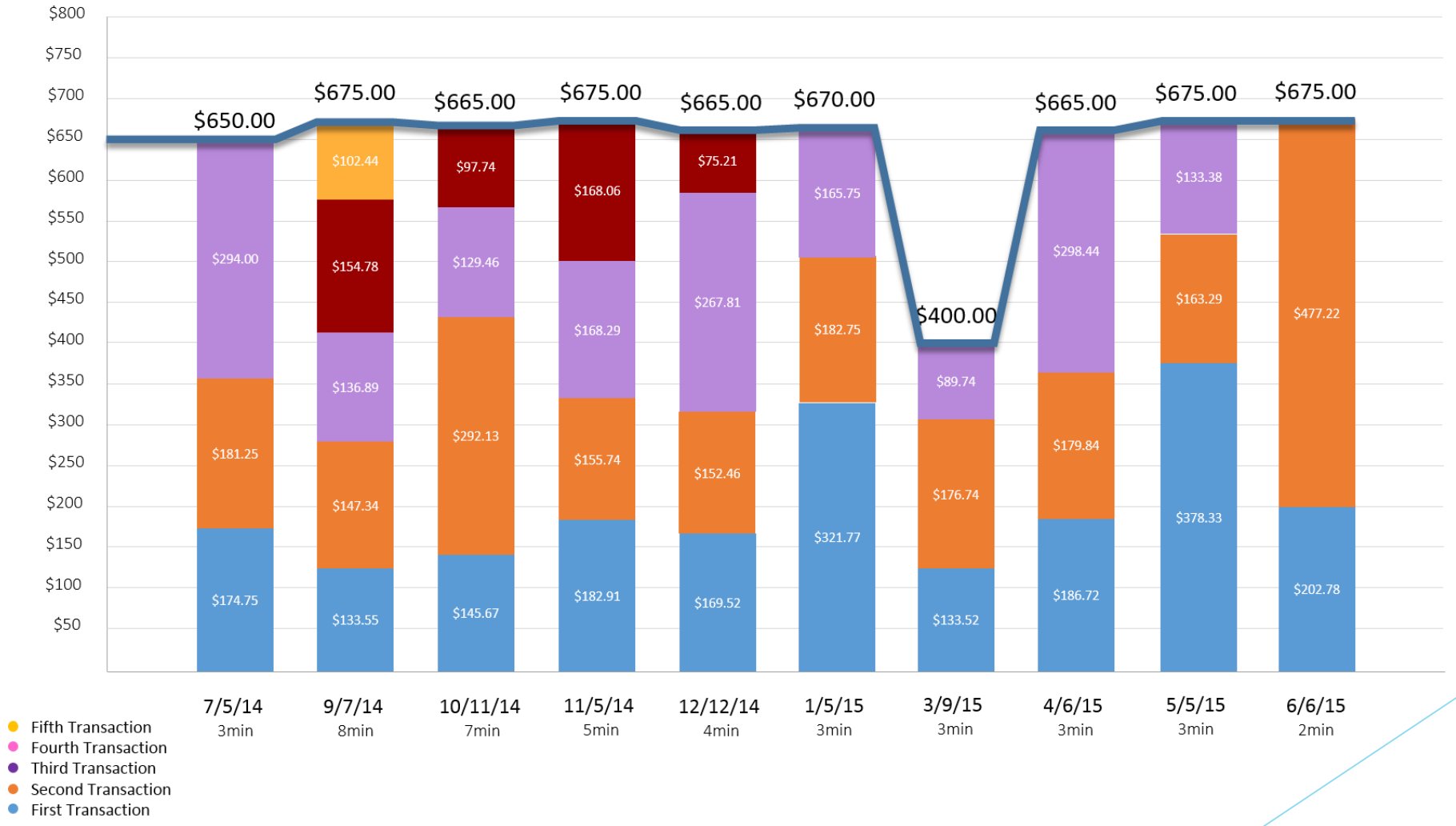
Mountain & Plains Intergovernmental Forum

August 30-31, 2017

INTRODUCTION

- Question: How would you identify a specific transaction pattern for one particular SNAP recipient from among hundreds of millions of SNAP EBT card transactions with the following parameters?
  - Minimum of two transactions in a rolling, five-minute period
  - Same vendor
  - Transactions must total an even dollar amount
  - Cumulative transaction amount must not equal total EBT balance

INTRODUCTION



Source: PA No. 15-04 *A Performance Audit of Data Analytics Techniques to Detect SNAP Abuse*, OSA 45-46 (January 5, 2016).

INTRODUCTION

- Overview of data analytics and improper payments
- Introduction to available data analytics tools
- Examination and examples of core data analytics techniques

Data Analytics

- Data analytics is the process by which large amounts of data are reviewed to discover anomalies, patterns, commonalities, or other user-defined metrics to facilitate increased compliance, greater efficiency, or additional understanding of a particular government program.

Improper Payments: Definition

- What makes a payment “improper”?
  - Criteria: What does applicable law, rule, or policy say?

Improper Payments: Indicators

- Where do you begin?

Source: Photo by Ethan Sykes on Unsplash.

Improper Payments: FREQUENCY

Source: Photo by Álvaro Bernal on Unsplash.

- Totals across rolling periods of time
- Example: Rapid successive SNAP EBT transactions

Transaction	Time	Amount
1	12:00	$10.00
2	12:01	$5.00
3	12:07	$3.00
4	12:09	$10.00
5	12:10	$1.00
6	12:14	$6.00
7	12:35	$5.00

Improper Payments: FREQUENCY

Source: Photo by Álvaro Bernal on Unsplash.

- Totals across rolling periods of time
- Example: Rapid successive SNAP EBT transactions

Transaction	Time	Amount	Amount	Amount	Amount
1 →	12:00	$10.00	$10.00	$10.00	$10.00
2 ✖	12:01	$5.00	$5.00	$5.00	$5.00
3 →	12:07	$3.00	$3.00	$3.00	$3.00
4 →	12:09	$10.00	$10.00	$10.00	$10.00
5 →	12:10	$1.00	$1.00	$1.00	$1.00
6 ✖	12:14	$6.00	$6.00	$6.00	$6.00
7 ✖	12:35	$5.00	$5.00	$5.00	$5.00

Improper Payments: FREQUENCY

Source: Photo by Álvaro Bernal on Unsplash.

- Totals across rolling periods of time
- Example: Rapid successive SNAP EBT transactions

Transaction	Time	Amount
1	12:00	$10.00
2	12:01	$5.00
3	12:07	$3.00
4	12:09	$10.00
5	12:10	$1.00
6	12:14	$6.00
7	12:35	$5.00

$15.00

$20.00

$17.00

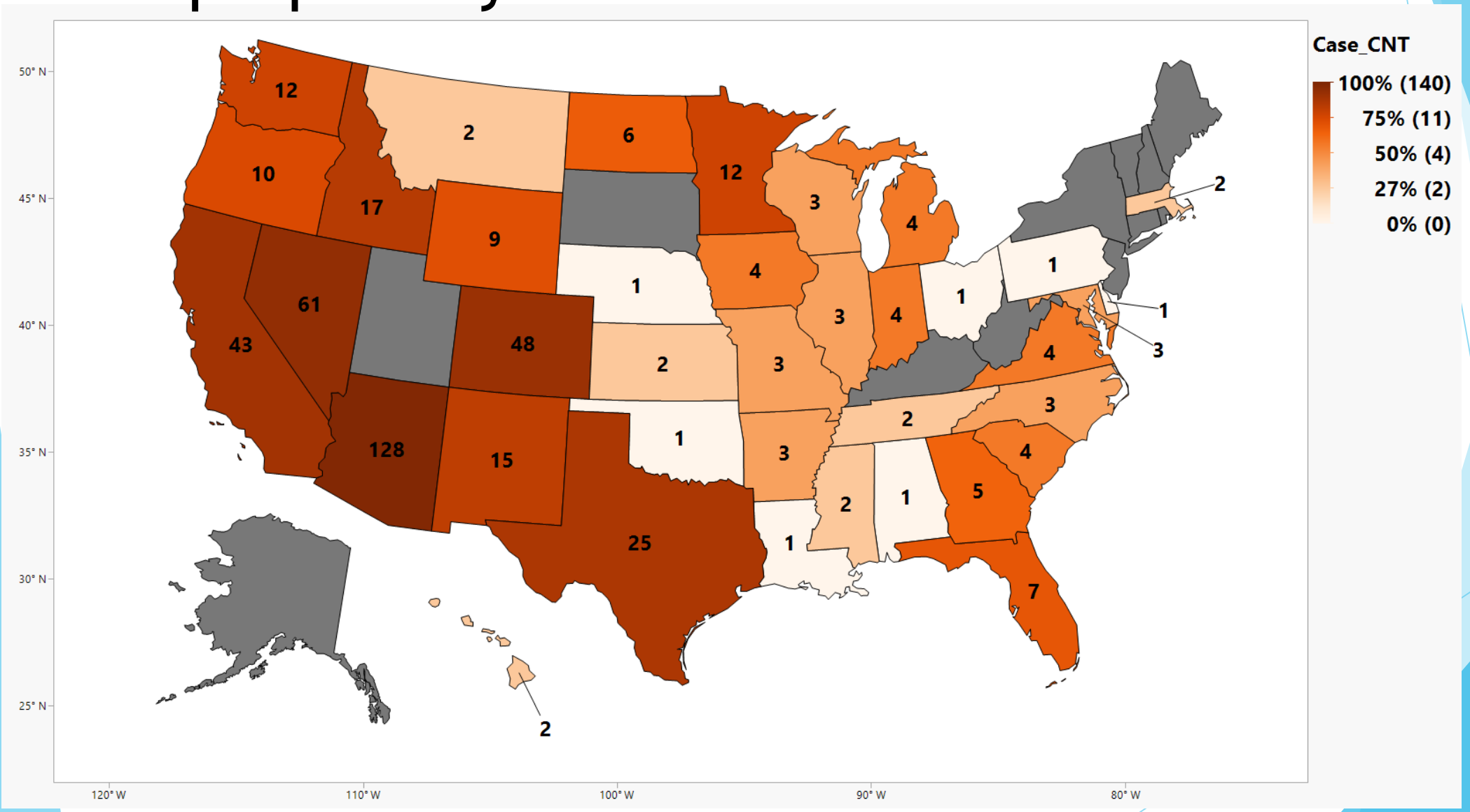
$7.00

Improper Payments: LOCATION

- Example: Out-of-state SNAP EBT transactions
  - **636** SNAP recipients spent almost **$1.4 million** while using their EBT card *exclusively* outside of Utah for **6** consecutive months or longer

Source: PA No. 15-04 *A Performance Audit of Data Analytics Techniques to Detect SNAP Abuse*, OSA 21-22 (January 5, 2016).

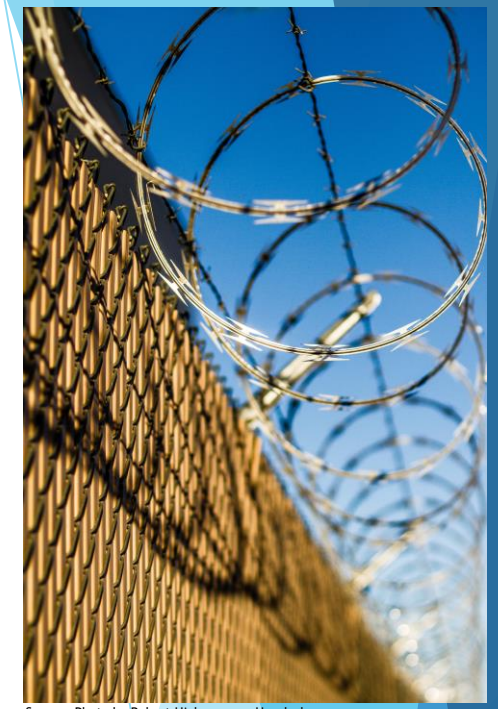
Improper Payments: LOCATION



Source: PA No. 15-04 *A Performance Audit of Data Analytics Techniques to Detect SNAP Abuse*, OSA 21-22 (January 5, 2016).

Improper Payments: LOCATION

- Result:
  - Little Rock, Arkansas recipient charged with 17 counts of fraud



Source: Photo by Robert Hickerson on Unsplash.

Source: *Workforce Services Helps to Bring Down Multi-State Fraudster*, Utah Department of Workforce Services (June 30, 2016).

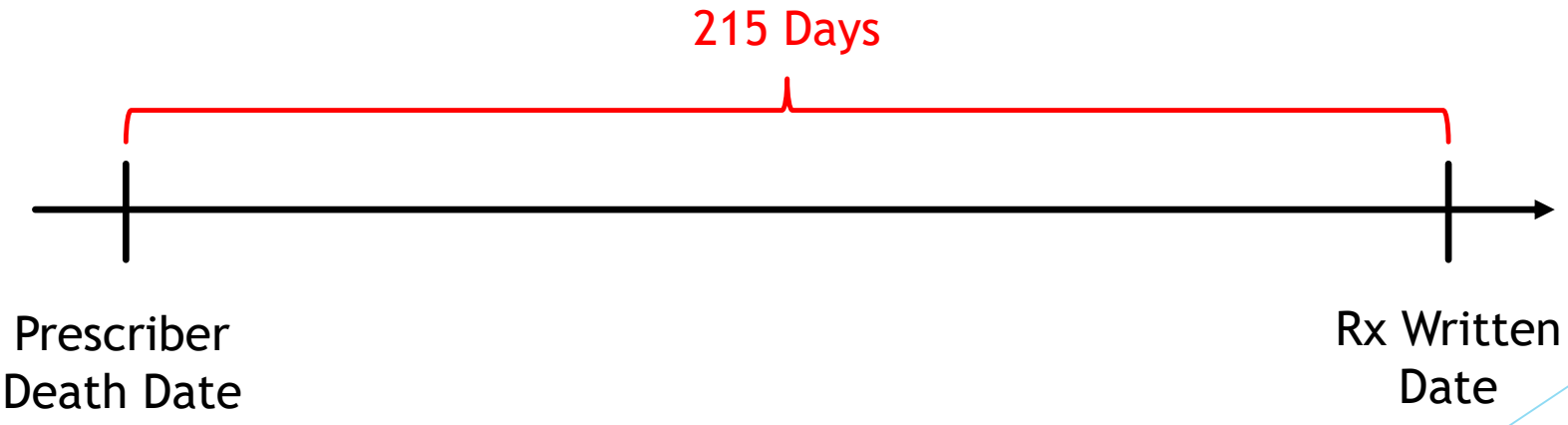
Improper Payments: AMOUNT

- Example: Even dollar SNAP EBT transactions
  - Investigations unit inadvertently limited algorithm to detect only even dollar transactions that were multiples of 10 (e.g., $50, $60, etc.)
  - What about transactions made in rapid succession at the same vendor that total to an even dollar amount?

Source: PA No. 15-04 *A Performance Audit of Data Analytics Techniques to Detect SNAP Abuse*, OSA 53 (January 5, 2016).

Improper Payments: DATES

- Too early or too late?
- Example:
  - Utah Medicaid paid for a prescription *written* 215 days *after* a prescriber's death



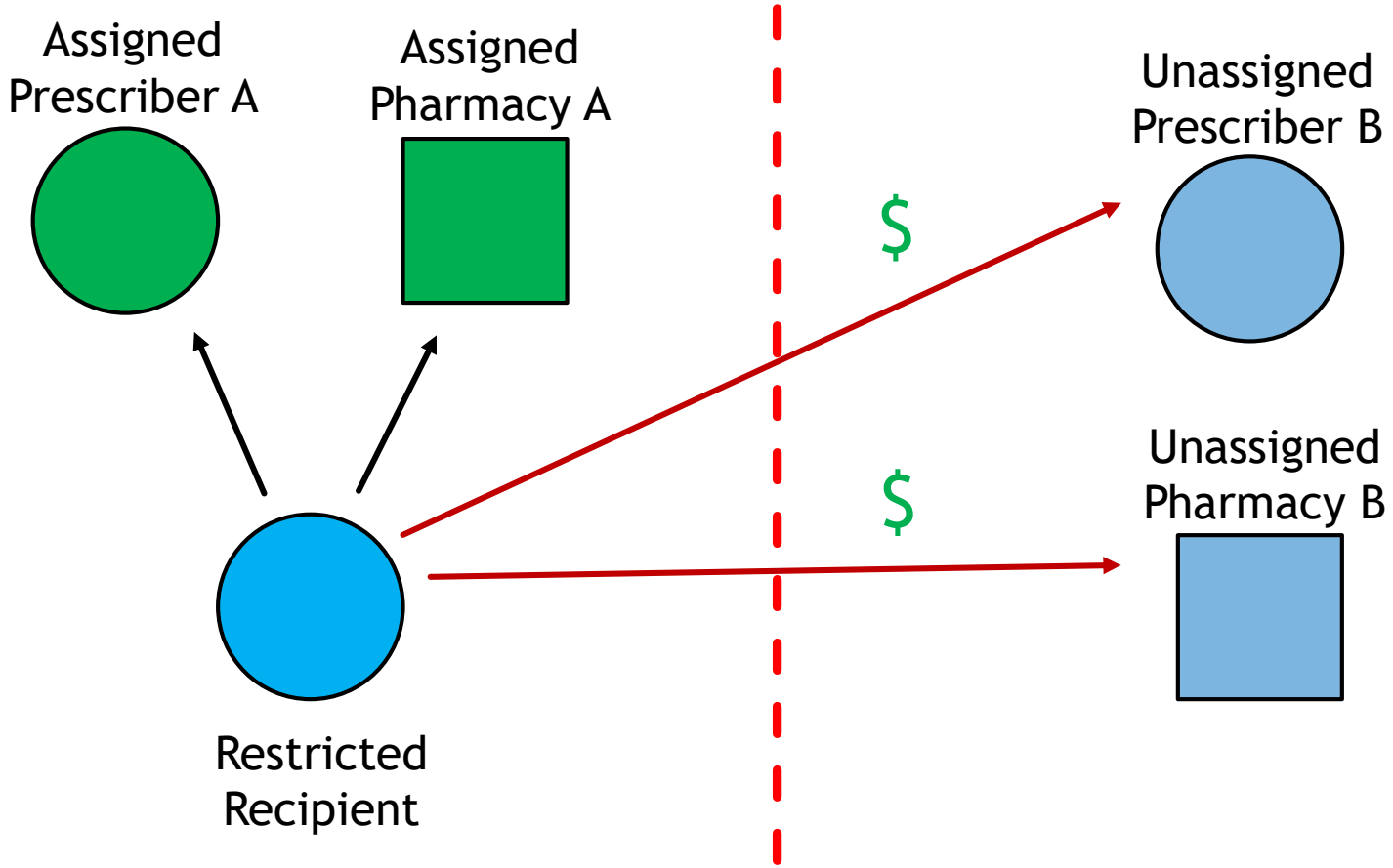
Source: PA No. 16-02 *A Performance Audit of Medicaid Prescription Drug Controls*, OSA 9-10 (August 17, 2016).

Improper Payments: ELIGIBILITY

- Example: Unauthorized vendors
  - Utah Medicaid paid for (1) prescriptions written by *unassigned* prescribers and (2) prescriptions dispensed at *unassigned* pharmacies

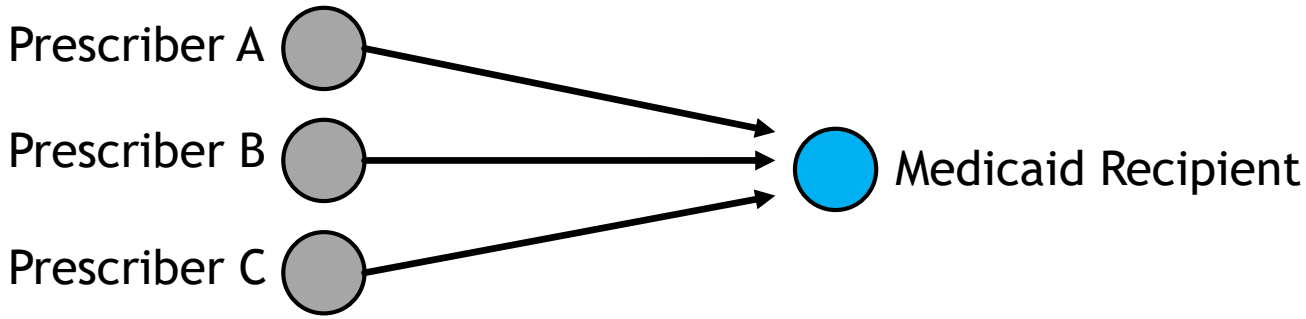
Source: PA No. 16-02 *A Performance Audit of Medicaid Prescription Drug Controls*, OSA 35 (August 17, 2016).

Improper Payments: ELIGIBILITY



Improper Payments: ELIGIBILITY

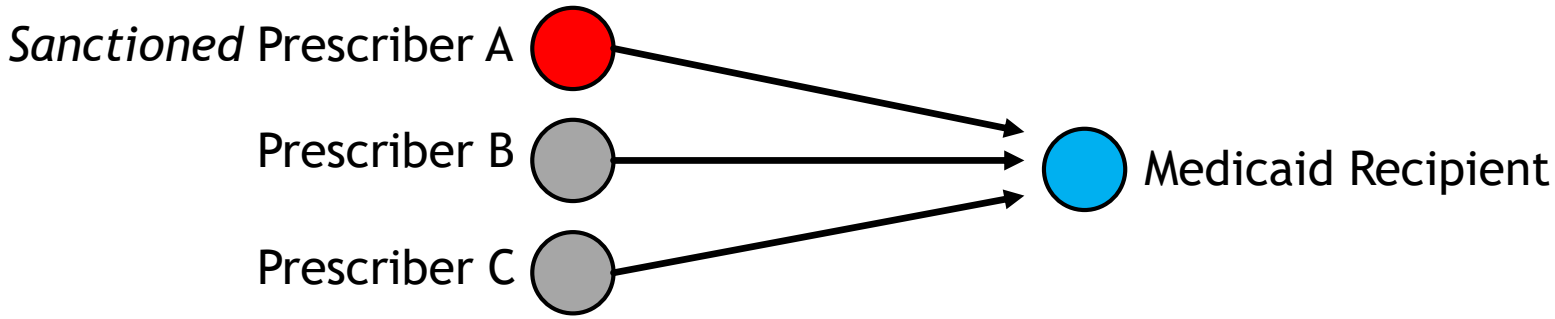
- Unauthorized vendors
- Example:
  - $ for Rx written by (1) *sanctioned* & (2) *unenrolled* prescribers



Source: PA No. 16-02 *A Performance Audit of Medicaid Prescription Drug Controls*, OSA 23 (August 17, 2016).

Improper Payments: ELIGIBILITY

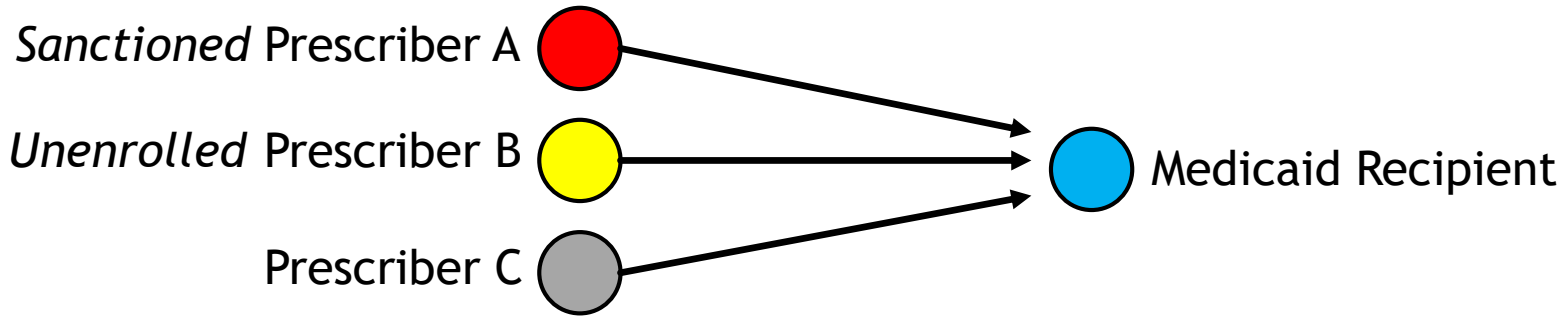
- Unauthorized vendors
- Example:
  - $ for Rx written by (1) *sanctioned* & (2) *unenrolled* prescribers



Source: PA No. 16-02 *A Performance Audit of Medicaid Prescription Drug Controls*, OSA 23 (August 17, 2016).

Improper Payments: ELIGIBILITY

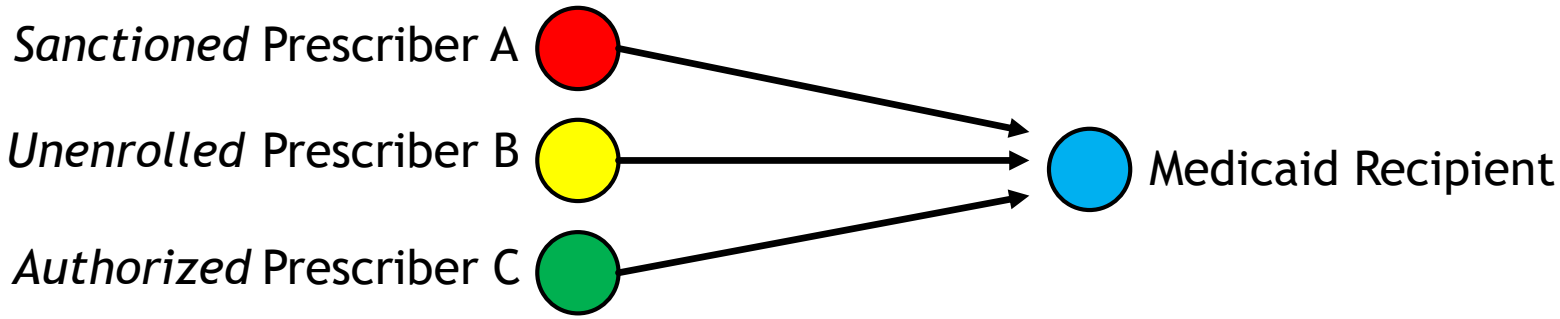
- Unauthorized vendors
- Example:
  - $ for Rx written by (1) *sanctioned* & (2) *unenrolled* prescribers



Source: PA No. 16-02 *A Performance Audit of Medicaid Prescription Drug Controls*, OSA 23 (August 17, 2016).

Improper Payments: ELIGIBILITY

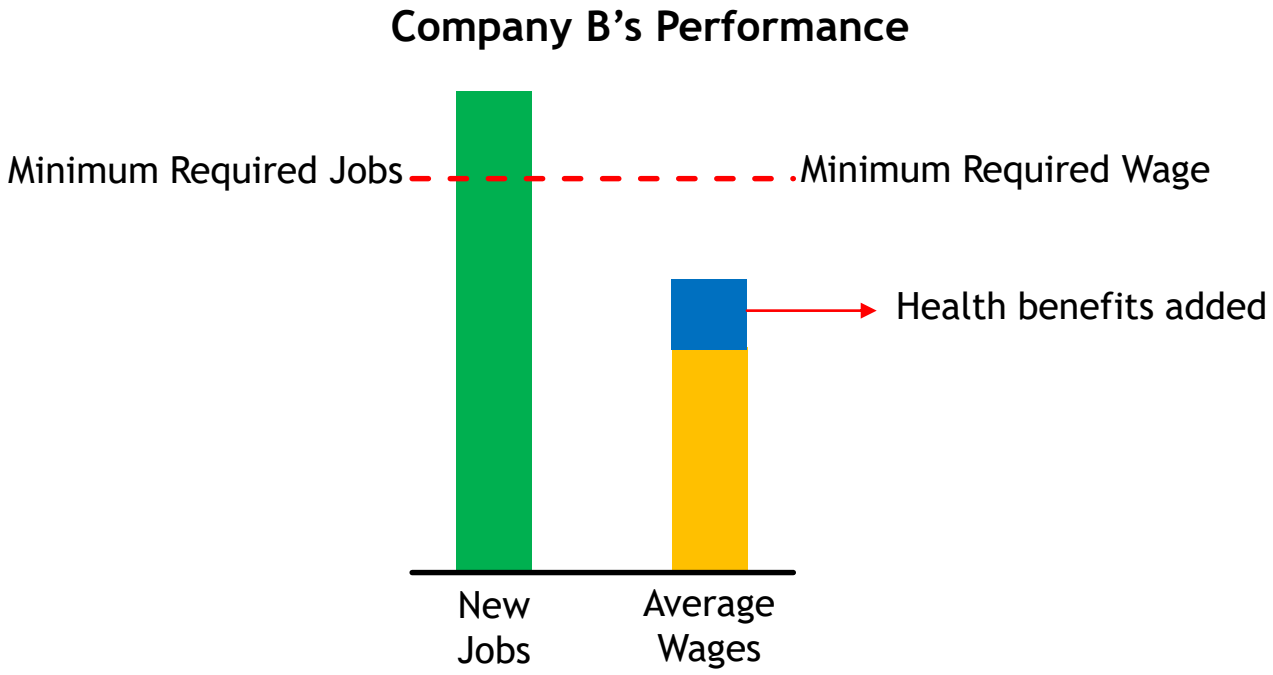
- Unauthorized vendors
- Example:
  - $ for Rx written by (1) *sanctioned* & (2) *unenrolled* prescribers



Source: PA No. 16-02 *A Performance Audit of Medicaid Prescription Drug Controls*, OSA 23 (August 17, 2016).

Improper Payments: CONTRACTS

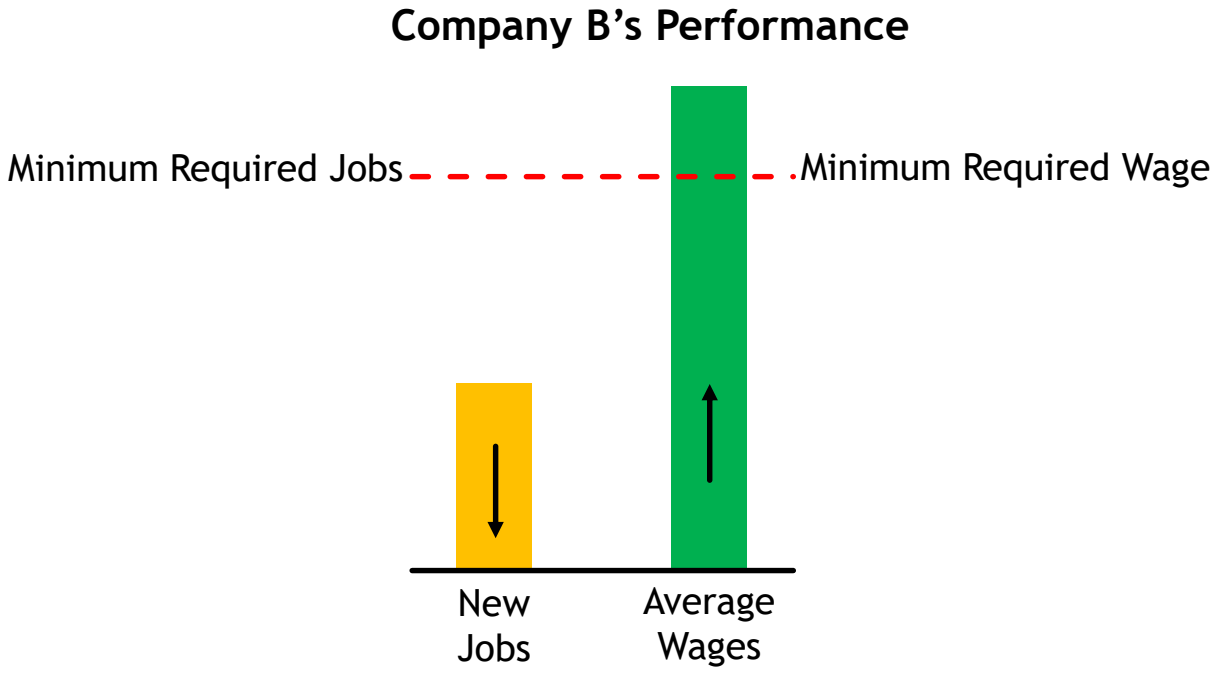
- Contractual terms and conditions
- Example: Corporate Tax Incentive Awards



Source: PA No. 14-03 *A Performance Audit of GOED's Corporate Incentives Program,* OSA 18-19 (October 14, 2014).

Improper Payments: CONTRACTS

- Contractual terms and conditions
- Example: Corporate Tax Incentive Awards



Source: PA No. 14-03 *A Performance Audit of GOED's Corporate Incentives Program,* OSA 18-19 (October 14, 2014).

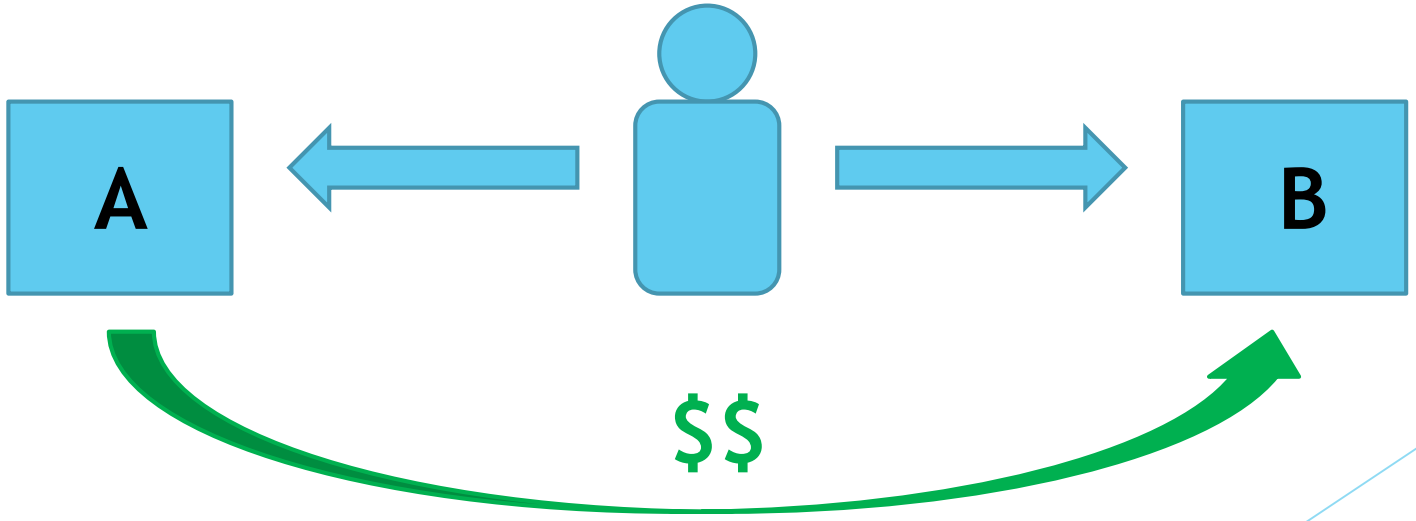
Improper Payments: CONTRACTS

- Contractual terms and conditions
- Example: Corporate Tax Incentive Awards
  - Inappropriate tax incentive award payment despite failure to meet required contractual conditions
  - Result: H.B. 129 passed 96-0

Source: PA No. 14-03 *A Performance Audit of GOED's Corporate Incentives Program,* OSA 18-19 (October 14, 2014).

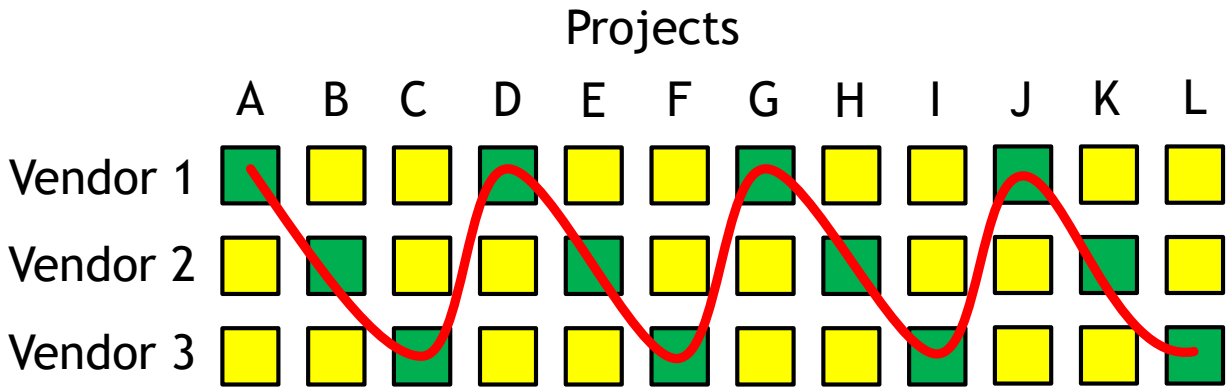
Improper Payments: PROCUREMENT

- Purchasing laws and rules
- Example: Related-party transactions
  - Does a common individual fill official roles with both the contracting entity and the vendor?



Improper Payments: PROCUREMENT

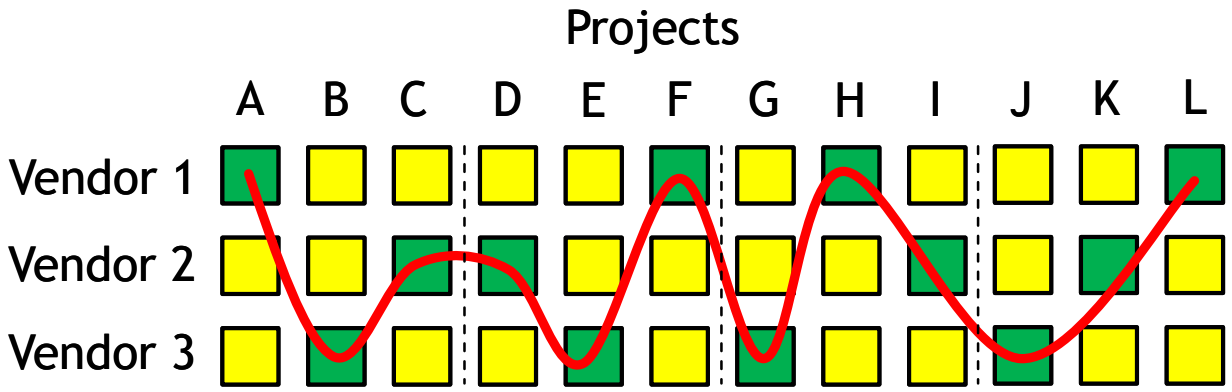
- Purchasing laws and rules
- Example: Split purchases
  - Data summaries *by vendor* (and date?)
- Example: Bid collusion



Source: *Visualizing Fraud Patterns: Exposing the Hidden Threats*, NASACT Webinar (March 26, 2015).

Improper Payments: PROCUREMENT

- Purchasing laws and rules
- Example: Split purchases
  - Data summaries *by vendor* (and date?)
- Example: Bid collusion



Source: *Visualizing Fraud Patterns: Exposing the Hidden Threats*, NASACT Webinar (March 26, 2015).

Improper Payments: DOCUMENTATION

- Discrepancies/errors/missing information?
- Example: Doctored credit card statements

Transaction	Receipt Date	Statement Date	Transaction	Amount
A	04/13	04/31	The Home Depot	$144.67
B	04/13	04/13	America's Best Vaule	$111.58
C	06/25	06825	Crystal Inn Cedar	$83.48

Source: PA No. 16-01 *A Performance Audit of Utah Communications Authority Financial Management and Transparency,* OSA 16-17 (May 18, 2016).

Improper Payments: DOCUMENTATION

- Discrepancies/errors/missing information?
- Example: Doctored credit card statements

Transaction	Receipt Date	Statement Date	Transaction	Amount
A	04/13	04/31	The Home Depot	$144.67
B	04/13	04/13	America's Best Vaule	$111.58
C	06/25	06825	Crystal Inn Cedar	$83.48

- $1.2 million embezzlement = 27-month prison sentence

Source: PA No. 16-01 *A Performance Audit of Utah Communications Authority Financial Management and Transparency,* OSA 16-17 (May 18, 2016).

Improper Payments: RECAP

- Don’t forget to account for multiple indicators!
  - Frequency + amount
  - Location + date
  - Eligibility + documentation

Source: PA No. 16-01 *A Performance Audit of Utah Communications Authority Financial Management and Transparency,* OSA 16-17 (May 18, 2016).

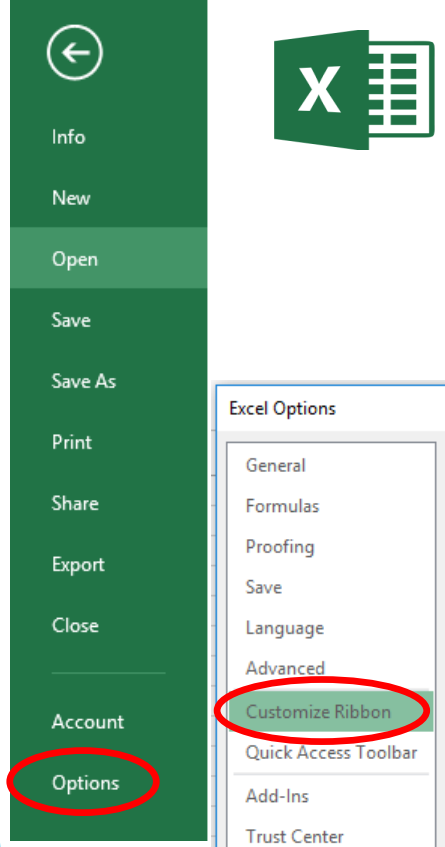
Available Tools

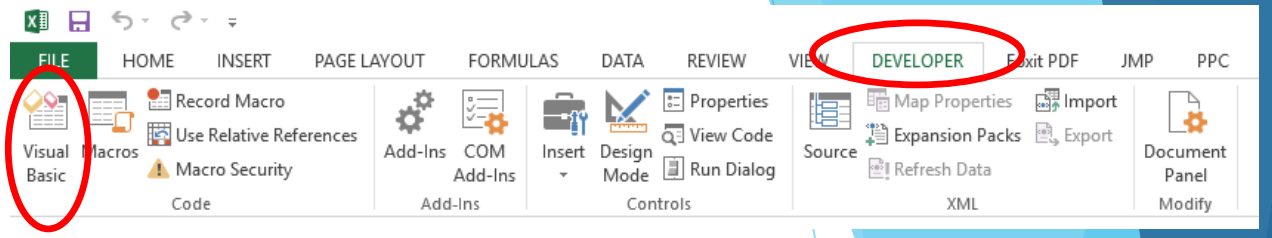
Source: Photo by Todd Quackenbush on Unsplash.

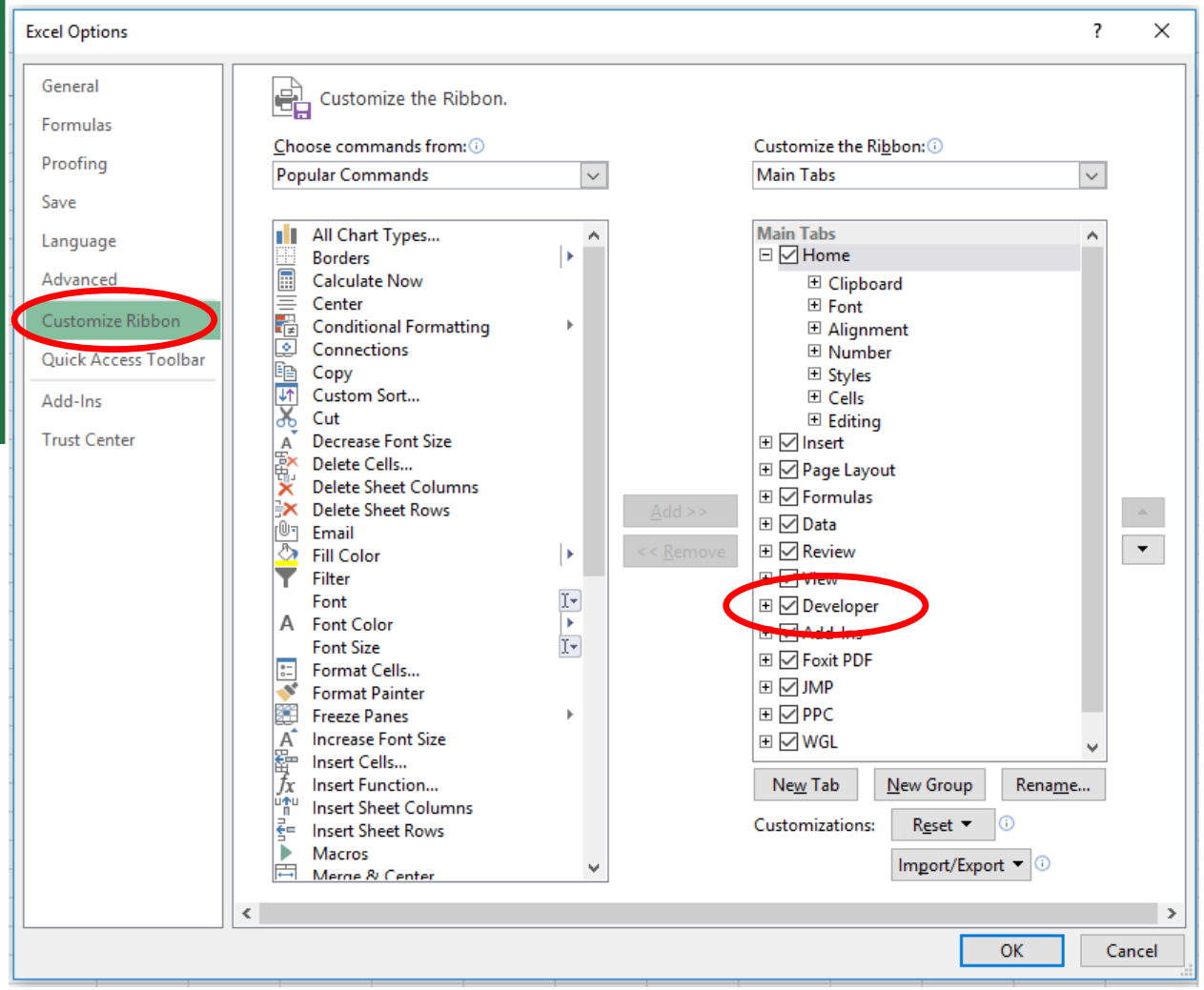
Available Tools: Open Source

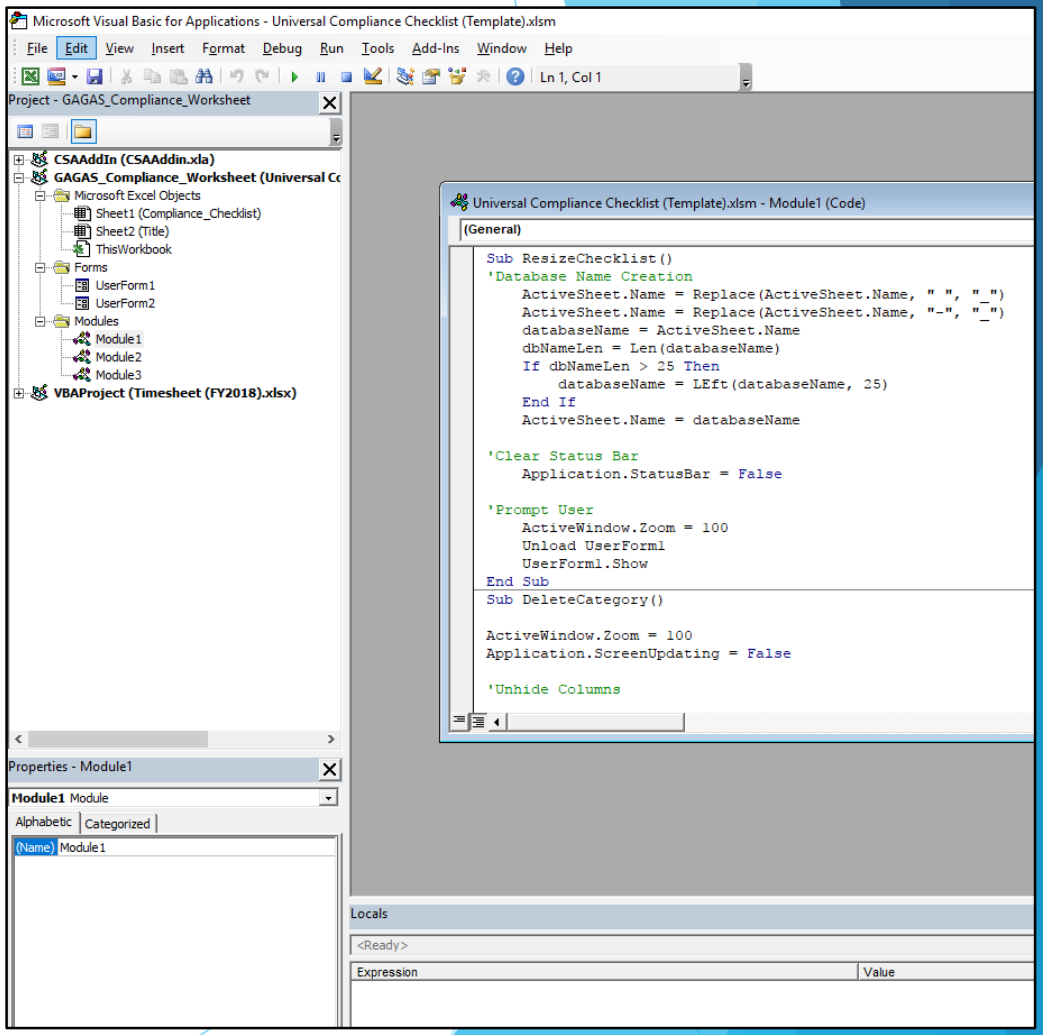
Available Tools: Licensed Software

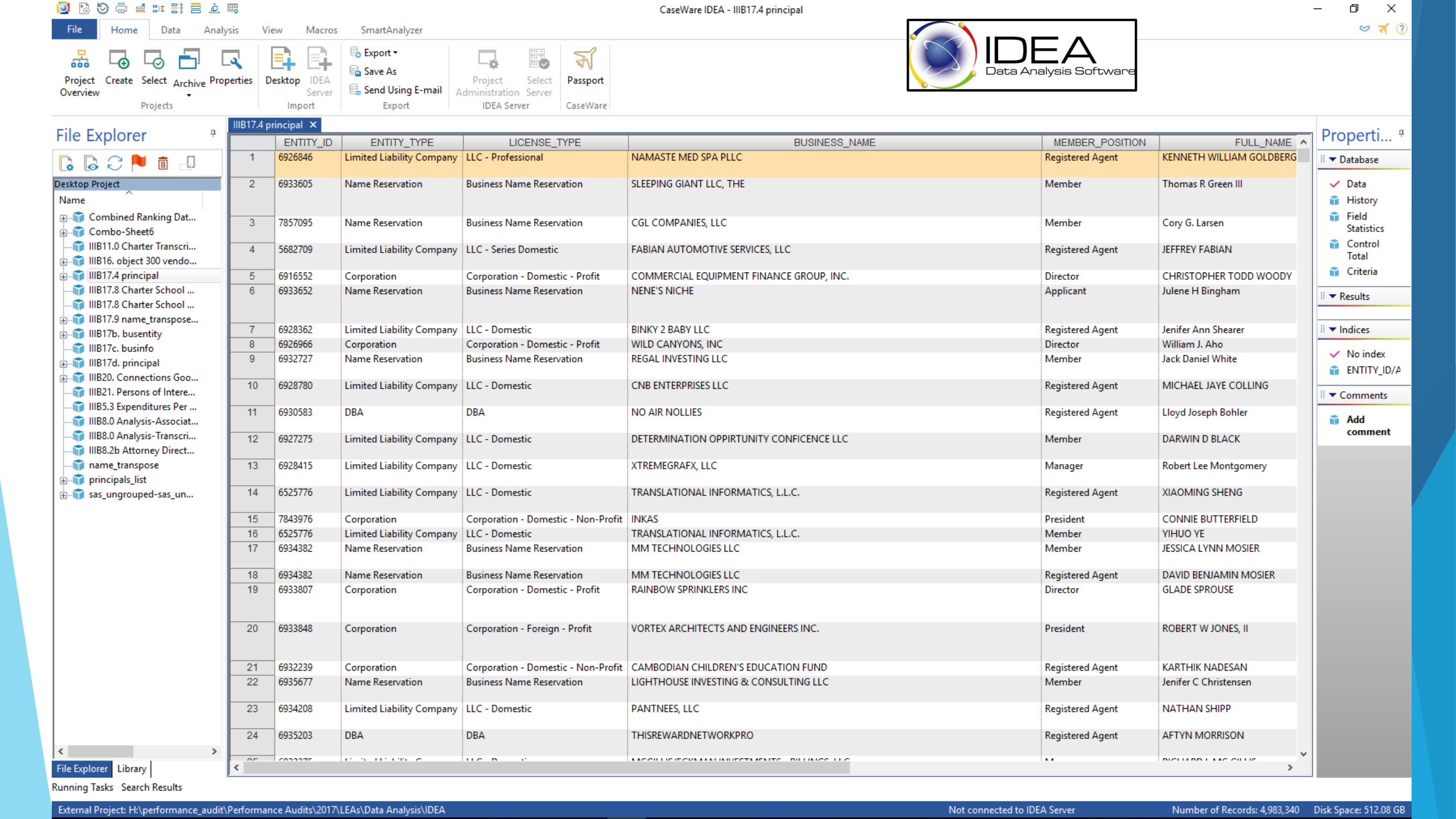
Visual Basic for Applications (VBA)



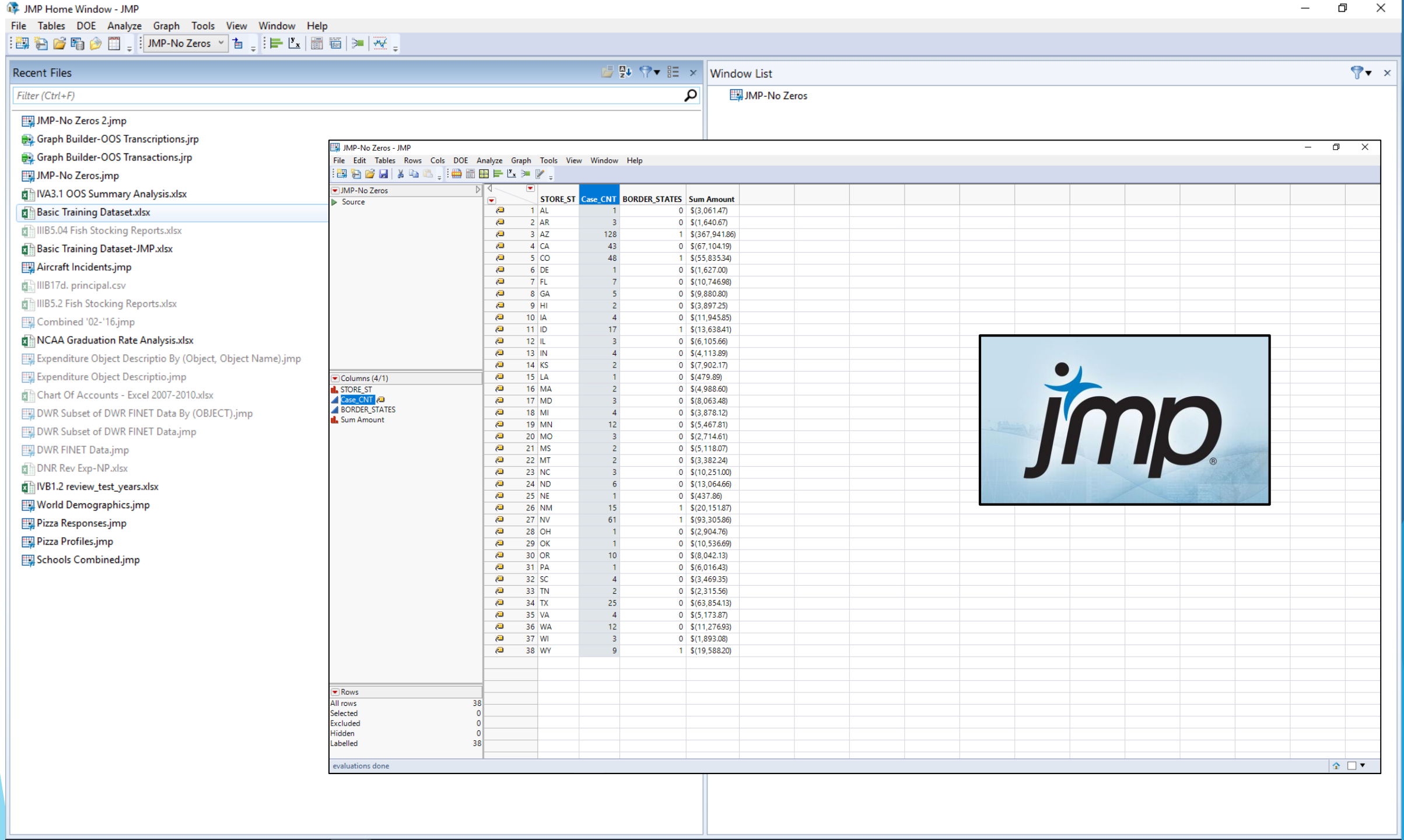








	ENTITY_ID	ENTITY_TYPE	LICENSE_TYPE	BUSINESS_NAME	MEMBER_POSITION	FULL_NAME
1	6926846	Limited Liability Company	LLC - Professional	NAMASTE MED SPA PLLC	Registered Agent	KENNETH WILLIAM GOLDBERG
2	6933605	Name Reservation	Business Name Reservation	SLEEPING GIANT LLC, THE	Member	Thomas R Green III
3	7857095	Name Reservation	Business Name Reservation	CGL COMPANIES, LLC	Member	Cory G. Larsen
4	5682709	Limited Liability Company	LLC - Series Domestic	FABIAN AUTOMOTIVE SERVICES, LLC	Registered Agent	JEFFREY FABIAN
5	6916552	Corporation	Corporation - Domestic - Profit	COMMERCIAL EQUIPMENT FINANCE GROUP, INC.	Director	CHRISTOPHER TODD WOODY
6	6933652	Name Reservation	Business Name Reservation	NENE'S NICHE	Applicant	Julene H Bingham
7	6928362	Limited Liability Company	LLC - Domestic	BINKY 2 BABY LLC	Registered Agent	Jenifer Ann Shearer
8	6926966	Corporation	Corporation - Domestic - Profit	WILD CANYONS, INC	Director	William J. Aho
9	6932727	Name Reservation	Business Name Reservation	REGAL INVESTING LLC	Member	Jack Daniel White
10	6928780	Limited Liability Company	LLC - Domestic	CNB ENTERPRISES LLC	Registered Agent	MICHAEL JAYE COLLING
11	6930583	DBA	DBA	NO AIR NOLLIES	Registered Agent	Lloyd Joseph Bohler
12	6927275	Limited Liability Company	LLC - Domestic	DETERMINATION OPPIRTUNITY CONFICENCE LLC	Member	DARWIN D BLACK
13	6928415	Limited Liability Company	LLC - Domestic	XTREMEGRAFX, LLC	Manager	Robert Lee Montgomery
14	6525776	Limited Liability Company	LLC - Domestic	TRANSLATIONAL INFORMATICS, L.L.C.	Registered Agent	XIAOMING SHENG
15	7843976	Corporation	Corporation - Domestic - Non-Profit	INKAS	President	CONNIE BUTTERFIELD
16	6525776	Limited Liability Company	LLC - Domestic	TRANSLATIONAL INFORMATICS, L.L.C.	Member	YIHUO YE
17	6934382	Name Reservation	Business Name Reservation	MM TECHNOLOGIES LLC	Member	JESSICA LYNN MOSIER
18	6934382	Name Reservation	Business Name Reservation	MM TECHNOLOGIES LLC	Registered Agent	DAVID BENJAMIN MOSIER
19	6933807	Corporation	Corporation - Domestic - Profit	RAINBOW SPRINKLERS INC	Director	GLADE SPROUSE
20	6933848	Corporation	Corporation - Foreign - Profit	VORTEX ARCHITECTS AND ENGINEERS INC.	President	ROBERT W JONES, II
21	6932239	Corporation	Corporation - Domestic - Non-Profit	CAMBODIAN CHILDREN'S EDUCATION FUND	Registered Agent	KARTHIK NADESAN
22	6935677	Name Reservation	Business Name Reservation	LIGHTHOUSE INVESTING & CONSULTING LLC	Member	Jenifer C Christensen
23	6934208	Limited Liability Company	LLC - Domestic	PANTNEES, LLC	Registered Agent	NATHAN SHIPP
24	6935203	DBA	DBA	THISREWARDNETWORKPRO	Registered Agent	AFTYN MORRISON


JMP Home Window - JMP
File Tables DOE Analyze Graph Tools View Window Help
JMP-No Zeros
Recent Files
Filter (Ctrl+F)
JMP-No Zeros 2.jmp
Graph Builder-OOS Transcriptions.jrp
Graph Builder-OOS Transactions.jrp
JMP-No Zeros.jmp
IVA3.1 OOS Summary Analysis.xlsx
Basic Training Dataset.xlsx
IIIB5.04 Fish Stocking Reports.xlsx
Basic Training Dataset-JMP.xlsx
Aircraft Incidents.jmp
IIIB17d. principal.csv
IIIB5.2 Fish Stocking Reports.xlsx
Combined '02-'16.jmp
NCAA Graduation Rate Analysis.xlsx
Expenditure Object Descriptio By (Object, Object Name).jmp
Expenditure Object Descriptio.jmp
Chart Of Accounts - Excel 2007-2010.xlsx
DWR Subset of DWR FINET Data By (OBJECT).jmp
DWR Subset of DWR FINET Data.jmp
DWR FINET Data.jmp
DNR Rev Exp-NP.xlsx
IVB1.2 review_test_years.xlsx
World Demographics.jmp
Pizza Responses.jmp
Pizza Profiles.jmp
Schools Combined.jmp
Window List
JMP-No Zeros
JMP-No Zeros - JMP
File Edit Tables Rows Cols DOE Analyze Graph Tools View Window Help
JMP-No Zeros
Source
Columns (4/1)
STORE_ST
Case_CNT
BORDER_STATES
Sum Amount
Rows
All rows 38
Selected 0
Excluded 0
Hidden 0
Labelled 38
evaluations done
STORE_ST Case_CNT BORDER_STATES Sum Amount
1 AL 1 0 $(3,061.47)
2 AR 3 0 $(1,640.67)
3 AZ 128 1 $(367,941.86)
4 CA 43 0 $(67,104.19)
5 CO 48 1 $(55,835.34)
6 DE 1 0 $(1,627.00)
7 FL 7 0 $(10,746.98)
8 GA 5 0 $(9,880.80)
9 HI 2 0 $(3,897.25)
10 IA 4 0 $(11,945.85)
11 ID 17 1 $(13,638.41)
12 IL 3 0 $(6,105.66)
13 IN 4 0 $(4,113.89)
14 KS 2 0 $(7,902.17)
15 LA 1 0 $(479.89)
16 MA 2 0 $(4,988.60)
17 MD 3 0 $(8,063.48)
18 MI 4 0 $(3,878.12)
19 MN 12 0 $(5,467.81)
20 MO 3 0 $(2,714.61)
21 MS 2 0 $(5,118.07)
22 MT 2 0 $(3,382.24)
23 NC 3 0 $(10,251.00)
24 ND 6 0 $(13,064.66)
25 NE 1 0 $(437.86)
26 NM 15 1 $(20,151.87)
27 NV 61 1 $(93,305.86)
28 OH 1 0 $(2,904.76)
29 OK 1 0 $(10,536.69)
30 OR 10 0 $(8,042.13)
31 PA 1 0 $(6,016.43)
32 SC 4 0 $(3,469.35)
33 TN 2 0 $(2,315.56)
34 TX 25 0 $(63,854.13)
35 VA 4 0 $(5,173.87)
36 WA 12 0 $(11,276.93)
37 WI 3 0 $(1,893.08)
38 WY 9 1 $(19,588.20)
jmp

SAS

File Edit View Tools Data Solutions Window Help

Explorer

Contents of 'Work'

Check Check2 Names Names_t Name_fin Utah_chart...

Log - (Untitled)

```
NOTE: There were 1517 observations read from the data set WORK.NAME_FIN.
NOTE: The data set WORK.CHECK has 601 observations and 4 variables.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.03 seconds
      cpu time            0.01 seconds


23
24   proc freq data=check noprint;
25    tables full_name/ missing out=check2;
26   run;

NOTE: There were 601 observations read from the data set WORK.CHECK.
NOTE: The data set WORK.CHECK2 has 569 observations and 3 variables.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds
```

VIEWTABLE: Work.Check (Frequency Counts and Percentages)

	FULL_NAME	ENTITY_ID	Frequency Count	Percent of Total Frequency
1	AARON MORRISON	6777552	3	0.1977587343
2	ABIGAIL WRIGHT-GRISSOM	7617998	3	0.1977587343
3	ABRAHAM TENG	5698676	1	0.0659195781
4	ADAM BURRIS	4898401	4	0.2636783125
5	ADAM JOHNSTON	6140457	3	0.1977587343
6	ADRIENNE LANSING	8278873	1	0.0659195781
7	ALAN DANIELS	5903476	1	0.0659195781
8	ALAN SHINO	8002503	1	0.0659195781
9	ALAN STOKES	8356448	1	0.0659195781
10	ALISHA JOHNSON	5927257	1	0.0659195781
11	ALLISON CLINGER	4992418	2	0.1318391562
12	ALYSSA LARSON	6104643	1	0.0659195781
13	AMIE CAMPBELL	5347893	2	0.1318391562
14	AMY SMITH	9108496	1	0.0659195781
15	AMY HUGHES	5821318	5	0.3295978906
16	AMY WADSWORTH	5081630	2	0.1318391562
17	AMY WYLIE	8237628	1	0.0659195781
18	ANDREA JOHNSON	7013445	3	0.1977587343
19	ANDREA OLDING	9565078	1	0.0659195781
20	ANDREW BURT	6449928	1	0.0659195781
21	ANDY HALE	7954876	2	0.1318391562
22	ANGELA D ROWLAND	8237628	1	0.0659195781
23	ANGELA FANJUL	6247831	1	0.0659195781
24	ANGELA RASMUSSEN	5898486	2	0.1318391562
25	ANN GUBLER	6728440	2	0.1318391562
26	ANNA KRAMER	7939585	1	0.0659195781
27	ANNA STANTON	7954876	1	0.0659195781
28	ANNA TREVINO	5698676	2	0.1318391562
29	ANNABEL SHEINBERG	5081630	1	0.0659195781

name transpose

```
proc import datafile="H:\performance_audit\Performance Audits\2017\LEAs\Data Analysis\IDI
  out=names replace;
run;

proc sort data=names; by full_name entity_id; run;

proc transpose data=names out=names_t;
  by full_name entity_id;
  var name1-name69;
run;

data name_fin;
  set names_t;
  if col1="" then delete;
  drop _NAME_ _LABEL_;
run;

proc freq data=name_fin noprint;
  tables full_name*entity_id/ missing out=check;
run;

proc freq data=check noprint;
  tables full_name/ missing out=check2;
run;
```

Results Explorer

Output - (Untitled) Log - (Untitled) name transpose VIEWTABLE: Work.Che...

NOTE: Table has been opened in browse mode.

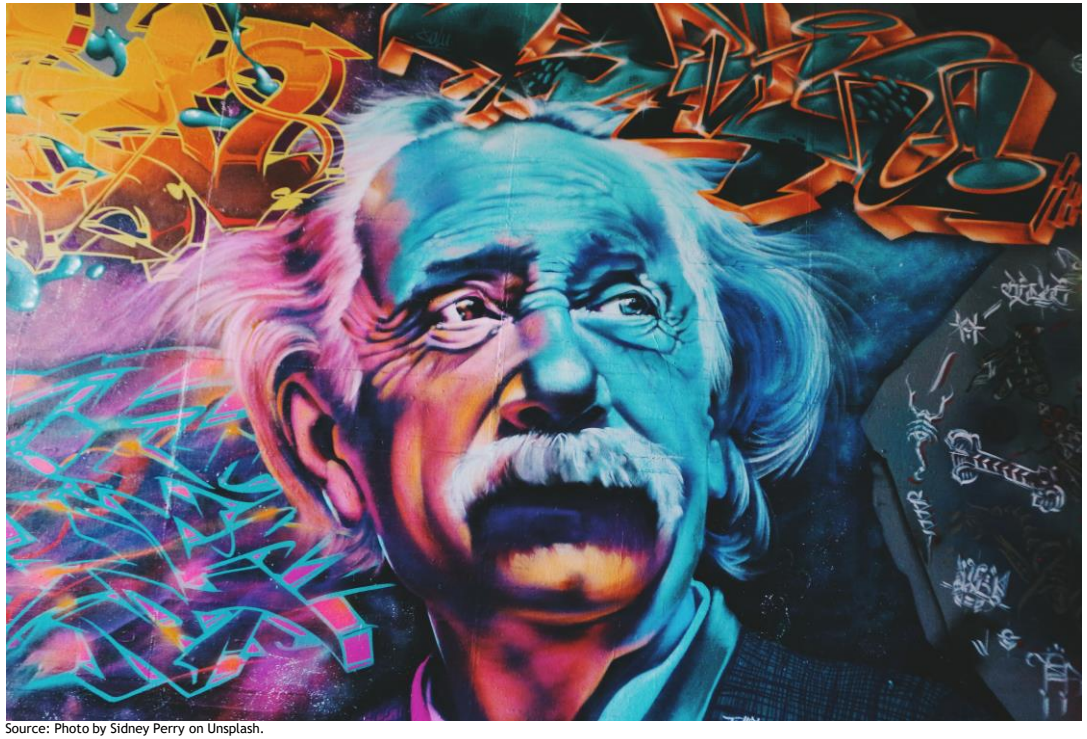
C:\Users\npurse

Available Tools: Considerations

- Price (e.g., open source vs. licensed software)
- Programming ability (e.g., point-and-click vs. running scripts)
- Visualization (e.g., geographic plotting)
- Data Type (e.g., PDF, CSV, Excel)
  - Text fields
- Volume
- Speed
- Automation
- Test Type (e.g., statistical sampling vs. entire universe)

Techniques: Preliminary Considerations

- Data analytics is both a science and an art. The key is thinking about data in new ways.



Source: Photo by Sidney Perry on Unsplash.

Techniques: Preliminary Considerations

- Understanding the data
  - Type codes (e.g., final payment, reversed, etc.)
  - Data dictionaries

Techniques: Preliminary Considerations

- Data cleanliness
  - Is related data standardized across related tables?

	Number/Numeric	Date	Text/String
Date Field	42370.00	1/1/2016	January 1, 2016

Techniques: Preliminary Considerations

- Data cleanliness
  - Is related data standardized across related tables?

	Number/Numeric	Date	Text/String
Date Field	42370.00	1/1/2016	January 1, 2016
User ID	42114	4/20/2015	042114

Techniques: Preliminary Considerations

- Data cleanliness
  - Leading or trailing spaces?

Name	Date	Amount
Robert Jones	1/1/2016	$5.00
Robert Jones	1/1/2016	$5.00
Robert Jones	1/1/2016	$5.00

Techniques: Preliminary Considerations

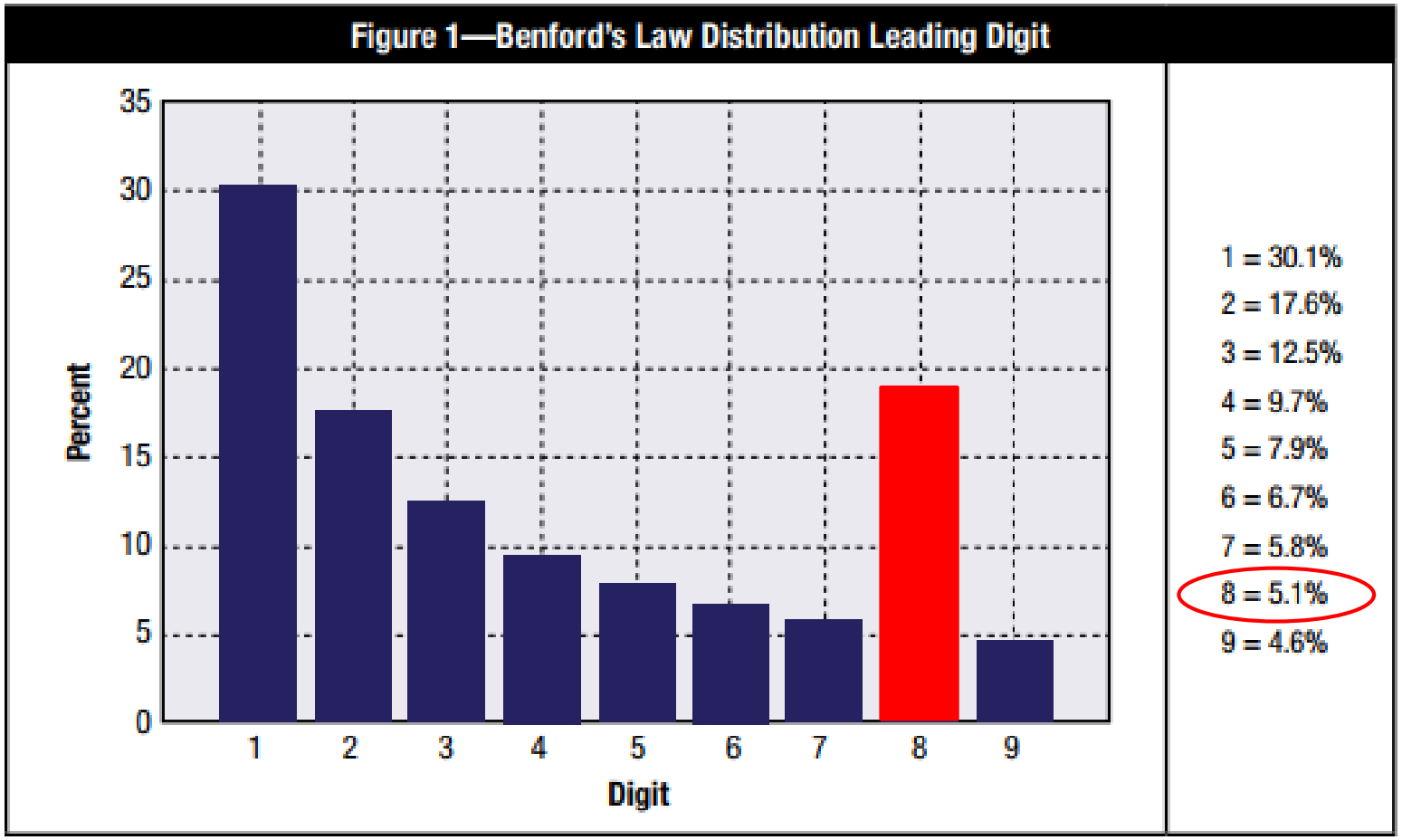
- Applicable criteria
  - Federal & state laws/regulations
  - Internal policies
- Audit objectives

Source: Photo by Claire Anderson on Unsplash.

Techniques: Benford's Law

- "Leveraging Data Analytics in Audits and Investigations" (August 2016)
  - Jamie Ralls, CFE, ACDA (Oregon Audits Division)
  - https://www.regonline.com/custImages/290000/299144/2016/Aug_11_Data_Analytics.pdf

Techniques: Benford's Law



Source: *Understanding and Applying Benford's Law*, IT Audit Basics, ISACA Journal, Vol. 3, 2011 (available at: https://www.isaca.org/Journal/archives/2011/Volume-3/Pages/Understanding-and-Applying-Benfords-Law.aspx)

Techniques: Sort Data

- Purpose:
  - Prepare data for some other data manipulation

Techniques: Sort Data

Name	Date	Amount
A	3/2/16	$10.00
B	3/4/16	$20.00
C	3/6/16	$5.00
A	3/8/16	$30.00
A	3/8/16	$25.00
B	3/21/16	$10.00

Techniques: Sort Data

School	Vendor	Amount
A	Not provided	$XXXXXXXX
B	Not applicable	$XXXXXXXX
C	Not provided	$XXXXXXXX
D	Not applicable	$XXXXXXXX
E	Not applicable	$XXXXXXXX
F	Not provided	$XXXXXXXX
G	ABC, Inc.	$XXXXXXXX
H	XYZ, LLC.	$XXXXXXXX

Techniques: Filter/Extract Data

- Purposes:
  - Identify specific data points
  - Create subsets

Techniques: Filter/Extract Data

Name	Date	Amount
A	3/2/16	$10.00
B	3/4/16	$20.00
C	3/6/16	$5.00
A	3/8/16	$30.00
A	3/8/16	$25.00
B	3/21/16	$10.00

Criteria: Amount > 10

Name	Date	Amount

Techniques: Filter/Extract Data

Name	Date	Amount
A	3/2/16	$10.00
B	3/4/16	$20.00
C	3/6/16	$5.00
A	3/8/16	$30.00
A	3/8/16	$25.00
B	3/21/16	$10.00

Criteria: Amount > 10

Name	Date	Amount

Techniques: Summarize Data

- Purposes:
  - Identify specific data
  - Combine databases

Techniques: Summarize Data

Name	Date	Amount
A	3/2/16	$5.00
B	3/4/16	$20.00
C	3/6/16	$5.00
A	3/8/16	$30.00
A	3/8/16	$25.00
B	3/21/16	$10.00

Sum

Name	Amount
A	$60.00
B	$30.00
C	$5.00

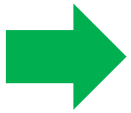
Mean

Name	Amount
A	$30.00
B	$15.00
C	$5.00

Techniques: Summarize Data

Sum

Name	Date	Amount
A	3/2/16	$5.00
B	3/4/16	$20.00
C	3/6/16	$5.00
A	3/8/16	$30.00
A	3/8/16	$25.00
B	3/21/16	$10.00

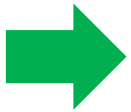


Name	Date	Amount
A	3/2/16	$5.00
A	3/8/16	$55.00
B	3/4/16	$20.00
B	3/21/16	$10.00
C	3/6/16	$5.00

Techniques: Summarize Data

Sum

Name	Date	Amount
A	3/2/16	$5.00
B	3/4/16	$20.00
C	3/6/16	$5.00
A	3/8/16	$30.00
A	3/8/16	$25.00
B	3/21/16	$10.00



Name	Date	Amount
A	3/2/16	$5.00
A	3/8/16	$55.00
B	3/4/16	$20.00
B	3/21/16	$10.00
C	3/6/16	$5.00

Techniques: Summarize Data

Name	SSN	Address
A	1	X
B	2	Y
C	3	Z
D	4	X
E	5	X
F	6	X
G	7	X
H	8	X
I	9	X

Address	Count
X	7
Y	1
Z	1

Louisiana Department of Labor

- Hurricane Katrina
- Address 450 miles away from NOLA
- 51 total debit cards requested
- 27-month jail sentence

Source: *Visualizing Fraud Patterns: Exposing the Hidden Threats*, NASACT Webinar (March 26, 2015).

Techniques: Merge/Join Data

- Purpose: Consolidate data
  - Important Consideration: One to many? One to one?

Techniques: Merge/Join Data (One to One)

Table 1: Items Sold

Name ←	Date ←	Item
A	3/2/16	Chair
A	3/8/16	Desk ✖
B	3/4/16	Sofa
B	3/21/16	Printer ✖
C	3/6/16	Lamp

Table 2: Quantity Sold

Name	Date	Quantity
A	3/2/16	3
A	3/3/16	15
A	3/9/16	6
B	3/4/16	8
B	3/22/16	5
C	3/6/16	4

Techniques: Merge/Join Data (One to One)

Name	Date	Item	Quantity
A	3/2/16	Chair	3
A	3/8/16	Desk	- ✖
B	3/4/16	Sofa	8
B	3/21/16	Printer	- ✖
C	3/6/16	Lamp	4

Techniques: Merge/Join Data (One to Many)

Table 1: Sales Team

Name	Title
A	Staff
B	Supervisor
C	Manager

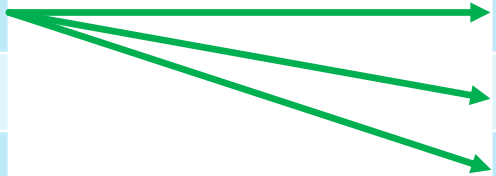


Table 2: Items Sold

Name	Date	Item
A	3/2/16	Chair
A	3/3/16	Desk
A	3/8/16	Sofa
B	3/4/16	Printer
B	3/21/16	Lamp
C	3/6/16	Chair

Techniques: Merge/Join Data (One to Many)

Table 1: Sales Team

Name	Title
A	Staff
B	Supervisor
C	Manager

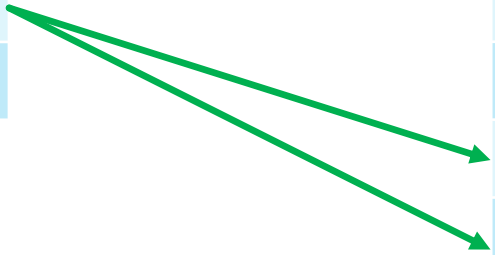


Table 2: Items Sold

Name	Date	Item
A	3/2/16	Chair
A	3/3/16	Desk
A	3/8/16	Sofa
B	3/4/16	Printer
B	3/21/16	Lamp
C	3/6/16	Chair

Techniques: Merge/Join Data (One to One)

Name	Title	Date	Item
A	Staff	3/2/16	Chair
A	Staff	3/3/16	Desk
A	Staff	3/8/16	Sofa
B	Supervisor	3/4/16	Printer
B	Supervisor	3/21/16	Lamp
C	Manager	3/6/16	Chair

Techniques: Append Data

- Purposes:
  - Consolidate data
  - Combine databases

Techniques: Append Data

Table 1: 2016

Name	Date	Amount
A	3/2/16	$10.00
A	3/8/16	$30.00
A	3/8/16	$25.00
B	3/4/16	$20.00
B	3/21/16	$10.00
C	3/6/16	$5.00

Table 2: 2017

Name	Date	Quantity
A	5/2/17	$15.00
A	5/8/17	$3.00
A	6/8/17	$5.00
B	6/4/17	$34.00
B	7/21/17	$18.00
C	8/6/17	$53.00

Table 1: 2016

Name	Date	Amount
A	3/2/16	$10.00
A	3/8/16	$30.00
A	3/8/16	$25.00
B	3/4/16	$20.00
B	3/21/16	$10.00
C	3/6/16	$5.00

Table 2: 2017

Name	Date	Quantity
A	5/2/17	$15.00
A	5/8/17	$3.00
A	6/8/17	$5.00
B	6/4/17	$34.00
B	7/21/17	$18.00
C	8/6/17	$53.00

2016 & 2017

Name	Date	Amount
A	3/2/16	$10.00
A	3/8/16	$30.00
A	3/8/16	$25.00
B	3/4/16	$20.00
B	3/21/16	$10.00
C	3/6/16	$5.00

Name	Date	Quantity
A	5/2/17	$15.00
A	5/8/17	$3.00
A	6/8/17	$5.00
B	6/4/17	$34.00
B	7/21/17	$18.00
C	8/6/17	$53.00

Techniques: Text Field Analysis

	Accident Number	Event Date	Investigation Type	Location	Country	Latitude	Longitude	Airport Code	Airport Name	Injury Severity	Fatal	Aircraft Damage	Registration Number	Make
1	CHI01LA057	01/01/2001	Accident	Lincoln, IL	United States	•	•	3LC	Lincoln Airport	Non-Fatal	Non-Fatal	Substantial	N737WQ	Cessna
2	CHI01LA061	01/01/2001	Accident	Port Huron, MI	United States	•	•	PHN	St. Clair County Int'l	Non-Fatal	Non-Fatal	Substantial	N2184N	Mooney
3	LAX01LA068	01/01/2001	Accident	Placerville, CA	United States	•	•		Private	Non-Fatal	Non-Fatal	Destroyed	N94LW	Wallace
4	MIA01LA054	01/02/2001	Accident	Naples, FL	United States	26.15° N	81.78° W	APF	Naples Municipal	Non-Fatal	Non-Fatal	Substantial	N45CF	Beech
5	ANC01LA028	01/03/2001	Accident	Atmautluak, AK	United States	60.87° N	162.28° W	4A2	Atmautluak	Non-Fatal	Non-Fatal	Substantial	N19771	Cessna
6	NYC01IA072	01/03/2001	Incident	Covington, KY	United States	•	•			Incident	Non-Fatal	Minor	N933CA	Bombardier
7	CHI01LA060	01/04/2001	Accident	Sioux Falls, SD	United States	•	•	FSD	JOE FOSS FIELD	Non-Fatal	Non-Fatal	Substantial	N727SP	Cessna
8	DEN01LA038	01/04/2001	Accident	Eden, UT	United States	41.30° N	111.82° W	NONE	Farm Field	Non-Fatal	Non-Fatal	Substantial	N26HV	Aviat
9	FTW01LA045	01/04/2001	Accident	BRISTOW, OK	United States	35.14° N	96.43° W	3F7	Jones Memorial	Non-Fatal	Non-Fatal	Substantial	N68472	Cessna
10	IAD01LA022	01/04/2001	Accident	Schenectady, NY	United States	42.85° N	73.93° W	SCH	SCHENECTADY CO...	Non-Fatal	Non-Fatal	Substantial	N435JL	Learjet
11	FTW01LA046	01/05/2001	Accident	DEL RIO, TX	United States	30.00° N	100.00° W	NONE	Wardlaw White Ran...	Non-Fatal	Non-Fatal	Substantial	N45373	Cessna
12	LAX01IA073	01/05/2001	Incident	Honolulu, HI	United States	•	•			Incident	Non-Fatal		N470EV	Boeing
13	LAX01LA069	01/05/2001	Accident	Livermore, CA	United States	37.68° N	121.82° W	LVK	Livermore Municipal	Non-Fatal	Non-Fatal	Substantial	N6165M	Cessna
14	MIA01LA055	01/05/2001	Accident	Cleveland, TN	United States	•	•			Non-Fatal	Non-Fatal	Destroyed	N112EC	Hiller
15	DEN01LA039	01/06/2001	Accident	Spanish Fork, UT	United States	40.18° N	111.70° W			Non-Fatal	Non-Fatal	Destroyed	N108ES	Cessna
16	FTW01LA064	01/06/2001	Accident	Pine Bluff, AR	United States	•	•			Non-Fatal	Non-Fatal	Substantial	N8984B	Cessna
17	LAX01FA070	01/06/2001	Accident	San Luis Obispo,...	United States	•	•	SBP	San Luis Obispo	Fatal(2)	Fatal	Destroyed	N383CA	Cessna
18	SEA01LA035	01/06/2001	Accident	EASTSOUND, WA	United States	•	•			Non-Fatal	Non-Fatal	Substantial	N1703D	Cessna
19	IAD01LA023	01/07/2001	Accident	LINCOLN PARK, ...	United States	•	•	N07	LINCOLN PARK	Non-Fatal	Non-Fatal	Substantial	N699LP	Cessna
20	IAD01LA024	01/07/2001	Accident	Concord, NH	United States	43.20° N	71.50° W	CON	Concord Muni Airp...	Non-Fatal	Non-Fatal	Substantial	N54931	Cessna

Source: Photo by Dominik Kollau on Unsplash.

Narrative Cause
The pilot's failure to maintain directional control on the runway. Factors relating to this accident were the aborted takeoff attempt, the pilot's improper in-flight planni...
The pilot failed to maintain directional control of the airplane and the runway selected resulted in a tailwind condition. Factors associated with the accident were the g...
The failure of the student pilot to maintain adequate ground clearance while hovering.
The failure of the pilot to obtain assistance from the FBO in the form of a marshaller and failure of the pilot to maintain clearance resulting in the on-ground collision w...
The pilot's failure to maintain a proper glidepath during final approach. A factor associated with the accident was soft terrain.
Missing exhaust nozzle bolts for undetermined reasons. A factor was inadequate maintenance inspection of the affected area.
Aircraft directional control not being maintained by the student pilot during the takeoff roll. Factors to the accident were the snow bank and the student pilot's lack of...
the pilot's failure to maintain aircraft control during a landing attempt. A contributing factor was his failure to check the snow conditions in the field before the flight.
The pilot's inadequate compensation for the crosswind conditions, which resulted in the airplane striking a snow bank during the landing flare/touchdown. A factor wa...
The pilot's improper trim setting, which resulted in a runway overrun and impact with a fence.
The pilot's inadequate compensation for the winds. A factor was the windshear.
the failure of the lighting dimmer switch or circuit components for undetermined reasons, which resulted in smoke in the cockpit.
The student pilot's inadequate compensation for a tailwind during final approach and her improper recovery from a bounced landing.
The PIC's failure to follow safe operating procedures for the maintenance of the rotorcraft's external aerial application equipment, resulting in inadvertent activation of...
Improper weather evaluation by both the pilot and pilot/passenger, and the pilot's inadvertent VFR flight into IMC resulting in his spatial disorientation. Factors were t...
the pilot's failure to maintain directional control during the forced landing on a gravel road. Contributing factors were the pilot's delay in using carburetor heat and th...
The pilot's failure to maintain a proper climb rate to VFR conditions on-top.
The pilot's failure to use carburetor heat prior to reducing engine power to idle while operating in serious carburetor icing conditions. Factors include serious carbureto...
the pilot's failure to maintain proper runway alignment during landing. Factors in the accident were the night lighting and the snow bank.
The flight instructor's improper decision to land downwind on the snow and ice covered runway with a tailwind, and his failure to perform a go-around. Factors in the ...
The loss of control on landing due to the student's improper recovery from a bounced landing, and the resulting nose over on the grass runway.
The failure of maintenance personnel to properly reconnect the aircraft's elevator control system, resulting in an in-flight disconnection of the elevator control and jam...
The loss of engine power during a normal descent due to fuel starvation for undetermined reasons.
The pilot's failure to adequately compensate for wind conditions while taxiing the airplane.
The vehicle driver's inadvertent failure to place the column shifter into the parking gear. A factor to the accident was the vehicle.
Fuel exhaustion due to the failure of the instructor to ensure there was sufficient fuel for the flight. Factors were the instructor's inadequate supervision of the student ...
the loss of engine power for undetermined reasons. A contributing factor was the lack of suitable terrain for the forced landing.
fuel exhaustion during approach due to the pilot's failure to refuel.
The pilot's failure to execute the published missed approach procedure. A factor was heavy snowfall during the approach.
the loss of engine power during takeoff resulting from the inadequate engagement of the throttle torque tube rod end into the support bushing by unknown mainten...
One or both of the rear ballonet air relief valves remained in an open position for undetermined reasons, which caused an out-of-balance trim condition; also causal w...
The failure of the pilot to conduct proper preflight planning, resulting in loss of engine power while in cruise flight due to fuel exhaustion, and the subsequent emerge...
Pilot's failure to maintain aircraft control while landing. Factors include a snow covered runway.
the pilot's failure to maintain directional control. Factors were the crosswind, the pilots lack of total experience, and the snow covered terrain adjacent to the runway.
The unsuitable terrain for landing encountered by the pilot. Factors included the loss of power for undetermined reasons and the rough terrain.
the ingestion of ice/slush into both engines, which resulted in dual engine power fluctuations.
The pilot's improper fuel management in cruise flight that resulted in a loss of engine power, because of fuel starvation, and a collision with a fence during the subsequ...
The flight instructor's failure to ensure (supervision) the student had an adequate supply of fuel available, and the student's failure to refuel the aircraft sufficently resul...
The pilot's inflight decision to continued visual flight into instrument meteorological conditions resulted in the inflight collision with trees. Low ceilings and trees were f...
the loss of power to both engines for undetermined reasons during approach. Contributing factors were the pilot's failure to properly position the fuel selector in acco...

Aircraft Incidents - Text Explorer of Narrative Cause - JMP

Text Explorer for Narrative Cause

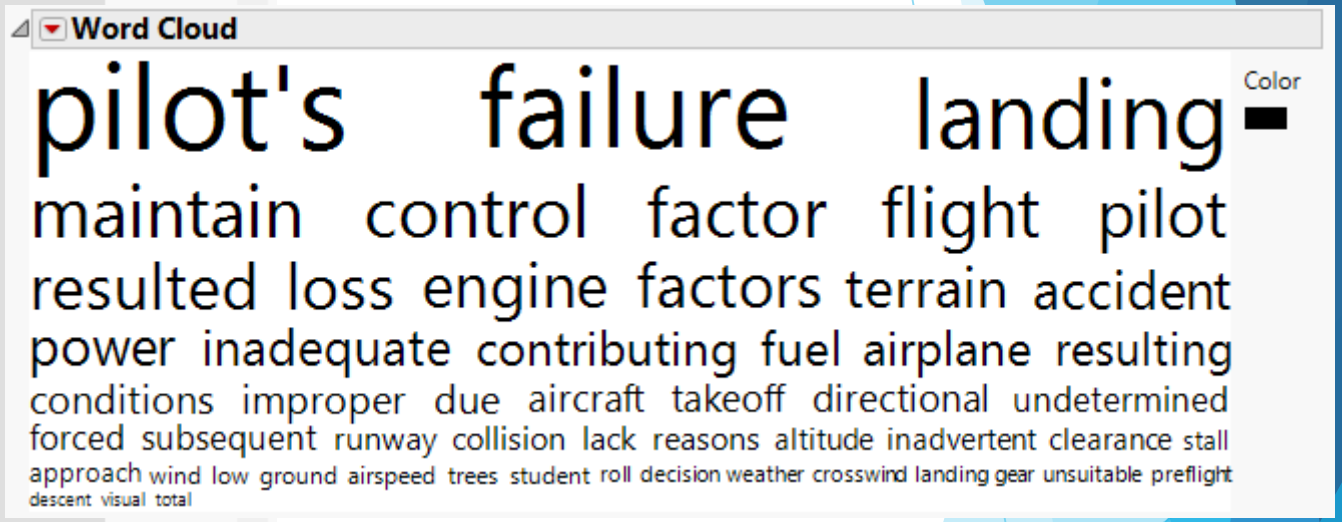
Number of Terms	Number of Cases	Total Tokens	Tokens per Case	Number of Non-empty Cases	Portion Non-empty per Case
2574	1906	51330	26.9307	1902	0.9979

Term and Phrase Lists

Term	Count
pilot's	1157
failure	1045
landing	846
maintain	575
control	563
factor	542
flight	537
pilot	509
resulted	489
loss	488
engine	458
factors	457
terrain	391
accident	370
power	364
inadequate	335
contributing	331
fuel	323
airplane	319
resulting	290
conditions	259
improper	246
due	244
aircraft	227
takeoff	225
directional	222
undetermined	214
forced	207
subsequent	205
runway	190
collision	186
lack	175
reasons	172
altitude	166
inadvertent	158
clearance	156
stall	136
approach	133

Phrase	Count	N
pilot's failure	484	2
failure to maintain	458	3
pilot's failure to maintain		
engine power		
loss of engine		
loss of engine power		
directional control		
forced landing		
contributing factor		
pilot's inadequate		
maintain directional control		
maintain directional		
undetermined reasons		
failure to maintain directional		
pilot's improper		
contributing factors		
aircraft control		
landing gear		
inadvertent stall		
fuel exhaustion		
failure of the pilot		
associated with the accident	76	4
control of the airplane	76	4
landing roll	75	2
loss of control	72	3
unsuitable terrain	72	2
power due	71	2
hard landing	70	2
engine power due	69	3
suitable terrain	69	2
power for undetermined	67	3
inadequate preflight	67	2
lack of suitable	66	3
student pilot's	66	2
lack of suitable terrain	65	4
power for undetermined reasons	65	4
wind conditions	65	2
engine power for undetermined	63	4

Word Cloud



Example 1: SNAP

- Objective: Determine six months of consecutive out-of-state transactions

	A	B	C
1	Name	Month	State
2	A	1	UT
3	B	1	UT
4	C	1	UT
5	A	2	UT
6	B	2	FL
7	C	2	UT
8	A	3	UT
9	B	3	GA
10	C	3	CO
11	A	4	UT
12	B	4	FL
13	C	4	UT
14	A	5	NV
15	B	5	FL
16	C	5	UT
17	A	6	NV
18	B	6	FL
19	C	6	UT
20	A	7	UT
21	B	7	FL
22	C	7	ID
23	A	8	UT
24	B	8	FL
25	C	8	UT
26	A	9	UT
27	B	9	FL
28	C	9	UT
29	A	10	WY
30	B	10	GA
31	C	10	UT

Sort

	A	B	C
1	Name	Month	State
2	A	1	UT
3	A	2	UT
4	A	3	UT
5	A	4	UT
6	A	5	NV
7	A	6	NV
8	A	7	UT
9	A	8	UT
10	A	9	UT
11	A	10	WY
12	B	1	UT
13	B	2	FL
14	B	3	GA
15	B	4	FL
16	B	5	FL
17	B	6	FL
18	B	7	FL
19	B	8	FL
20	B	9	FL
21	B	10	GA
22	C	1	UT
23	C	2	UT
24	C	3	CO
25	C	4	UT
26	C	5	UT
27	C	6	UT
28	C	7	ID
29	C	8	UT
30	C	9	UT
31	C	10	UT

=IF(C2<>"UT",1,0)

	A	B	C	D
1	Name	Month	State	OOS
2	A	1	UT	0
3	A	2	UT	0
4	A	3	UT	0
5	A	4	UT	0
6	A	5	NV	1
7	A	6	NV	1
8	A	7	UT	0
9	A	8	UT	0
10	A	9	UT	0
11	A	10	WY	1
12	B	1	UT	0
13	B	2	FL	1
14	B	3	GA	1
15	B	4	FL	1
16	B	5	FL	1
17	B	6	FL	1
18	B	7	FL	1
19	B	8	FL	1
20	B	9	FL	1
21	B	10	GA	1
22	C	1	UT	0
23	C	2	UT	0
24	C	3	CO	1
25	C	4	UT	0
26	C	5	UT	0
27	C	6	UT	0
28	C	7	ID	1
29	C	8	UT	0
30	C	9	UT	0
31	C	10	UT	0

=IF(A2=A1,IF(D2=1,E1+1,0),0)

	A	B	C	D	E
1	Name	Month	State	OOS	Count
2	A	1	UT	0	0
3	A	2	UT	0	0
4	A	3	UT	0	0
5	A	4	UT	0	0
6	A	5	NV	1	0
7	A	6	NV	1	1
8	A	7	UT	0	0
9	A	8	UT	0	0
10	A	9	UT	0	0
11	A	10	WY	1	0
12	B	1	UT	0	0
13	B	2	FL	1	0
14	B	3	GA	1	1
15	B	4	FL	1	2
16	B	5	FL	1	3
17	B	6	FL	1	4
18	B	7	FL	1	5
19	B	8	FL	1	6
20	B	9	FL	1	7
21	B	10	GA	1	8
22	C	1	UT	0	0
23	C	2	UT	0	0
24	C	3	CO	1	0
25	C	4	UT	0	0
26	C	5	UT	0	0
27	C	6	UT	0	0
28	C	7	ID	1	0
29	C	8	UT	0	0
30	C	9	UT	0	0
31	C	10	UT	0	0

=IF(A2=A1,IF(D2=1,E1+1,0),0)

	A	B	C	D	E
1	Name	Month	State	OOS	Count
2	A	1	UT	0	0
3	A	2	UT	0	0
4	A	3	UT	0	0
5	A	4	UT	0	0
6	A	5	NV	1	0
7	A	6	NV	1	1
8	A	7	UT	0	0
9	A	8	UT	0	0
10	A	9	UT	0	0
11	A	10	WY	1	0
12	B	1	UT	0	0
13	B	2	FL	1	0
14	B	3	GA	1	1
15	B	4	FL	1	2
16	B	5	FL	1	3
17	B	6	FL	1	4
18	B	7	FL	1	5
19	B	8	FL	1	6
20	B	9	FL	1	7
21	B	10	GA	1	8
22	C	1	UT	0	0
23	C	2	UT	0	0
24	C	3	CO	1	0
25	C	4	UT	0	0
26	C	5	UT	0	0
27	C	6	UT	0	0
28	C	7	ID	1	0
29	C	8	UT	0	0
30	C	9	UT	0	0
31	C	10	UT	0	0

	A	B
1	Name	Count (Max)
2	A	1
3	B	8
4	C	0

Example 2: Medicaid

- Objective: Determine whether any prescriptions were written after death

Example 2: Medicaid

Table 1: Eligible Prescribers

Name	SSN	DOB
1	111-11-1111	*
2	222-22-2222	2/2/70
3	*	3/3/70
4	444-44-4444	4/4/70

Table 2: Death Data

Name	SSN	DOB	DOD
4	444-44-4444	4/4/70	8/25/15
9	999-99-9999	9/9/70	9/14/15
3	333-33-3333	3/3/70	11/24/15
1	555-55-5555	5/5/70	1/5/16
2	222-22-2222	2/2/70	2/27/16
1	111-11-1111	1/1/70	12/1/16

Table 3: Pharmacy Claims

Prescriber	Rx Date
1	3/2/16
2	3/3/16
3	3/8/16
4	3/4/16
4	3/21/16
1	3/6/16

Example 2: Medicaid

Table 1: Eligible Prescribers

Name	SSN	DOB
1	111-11-1111	*
2 →	222-22-2222	2/2/70
3	*	3/3/70
4 →	444-44-4444	4/4/70

Table 2: Death Data

Name	SSN	DOB	DOD
4	444-44-4444	4/4/70	8/25/15
9	999-99-9999	9/9/70	9/14/15
3	333-33-3333	3/3/70	11/24/15
1	555-55-5555	5/5/70	1/5/16
2	222-22-2222	2/2/70	2/27/16
1	111-11-1111	1/1/70	12/1/16

Example 2: Medicaid

Table 1: Eligible Prescribers

Name	SSN	DOB
1	111-11-1111	*
2	222-22-2222	2/2/70
3	*	3/3/70
4	444-44-4444	4/4/70

Table 2: Death Data

Name	SSN	DOB	DOD
4	444-44-4444	4/4/70	8/25/15
9	999-99-9999	9/9/70	9/14/15
3	333-33-3333	3/3/70	11/24/15
1	555-55-5555	5/5/70	1/5/16
2	222-22-2222	2/2/70	2/27/16
1	111-11-1111	1/1/70	12/1/16

Example 2: Medicaid

Table 1: Eligible Prescribers

Name	SSN	DOB
1	111-11-1111	*
2	222-22-2222	2/2/70
3	*	3/3/70
4	444-44-4444	4/4/70

Table 2: Death Data

Name	SSN	DOB	DOD
4	444-44-4444	4/4/70	8/25/15
9	999-99-9999	9/9/70	9/14/15
3	333-33-3333	3/3/70	11/24/15
1	555-55-5555	5/5/70	1/5/16
2	222-22-2222	2/2/70	2/27/16
1	111-11-1111	1/1/70	12/1/16

Example 2: Medicaid

Tables 1 + 2: Eligible Prescribers + Death Data

Name	SSN	DOB	DOD
1	111-11-1111	*	12/1/16
2	222-22-2222	2/2/70	2/27/16
3	*	3/3/70	11/24/15
4	444-44-4444	4/4/70	8/25/15

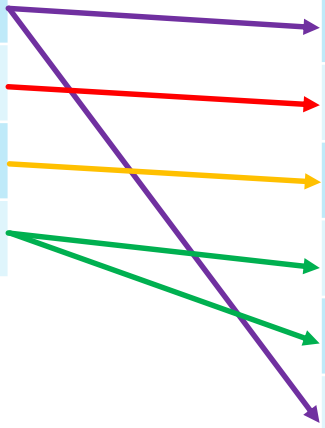


Table 3: Pharmacy Claims

Prescriber	Rx Date
1	3/2/16
2	3/3/16
3	3/8/16
4	3/4/16
4	3/21/16
1	3/6/16

Example 2: Medicaid

Final Merged/Joined Table

Name	SSN	DOB	DOD	Rx Date
1	111-11-1111	*	12/1/16	3/2/16
1	111-11-1111	*	12/1/16	3/6/16
2	222-22-2222	2/2/70	2/27/16	3/3/16
3	*	3/3/70	11/24/15	3/8/16
4	444-44-4444	4/4/70	8/25/15	3/4/16
4	444-44-4444	4/4/70	8/25/15	3/21/16

Example 2: Medicaid

Final Merged/Joined Table

Name	SSN	DOB	DOD	Rx Date	After Death?
1	111-11-1111	*	12/1/16	3/2/16	0
1	111-11-1111	*	12/1/16	3/6/16	0
2	222-22-2222	2/2/70	2/27/16	3/3/16	1
3	*	3/3/70	11/24/15	3/8/16	1
4	444-44-4444	4/4/70	8/25/15	3/4/16	1
4	444-44-4444	4/4/70	8/25/15	3/21/16	1

Example 2: Medicaid

Final Merged/Joined Table

Name	SSN	DOB	DOD	Rx Date	After Death?
1	111-11-1111	*	12/1/16	3/2/16	0
1	111-11-1111	*	12/1/16	3/6/16	0
2	222-22-2222	2/2/70	2/27/16	3/3/16	1
3	*	3/3/70	11/24/15	3/8/16	1
4	444-44-4444	4/4/70	8/25/15	3/4/16	1
4	444-44-4444	4/4/70	8/25/15	3/21/16	1

Example 2: Medicaid

Final Summary

Name	After Death (Sum)
1	0
2	1
3	1
4	2

SUMMARY

- Data analytics can help identify improper payments
- Various data analytics tools are available
- Data analytics techniques are as varied as you are creative!

Questions?

Nick Purse

npurse@utah.gov

801.538.1338