
U.S. House of Representatives: 6 Years of Spending Data

M Daniel A Turse
GASP Conference
2018-10-24



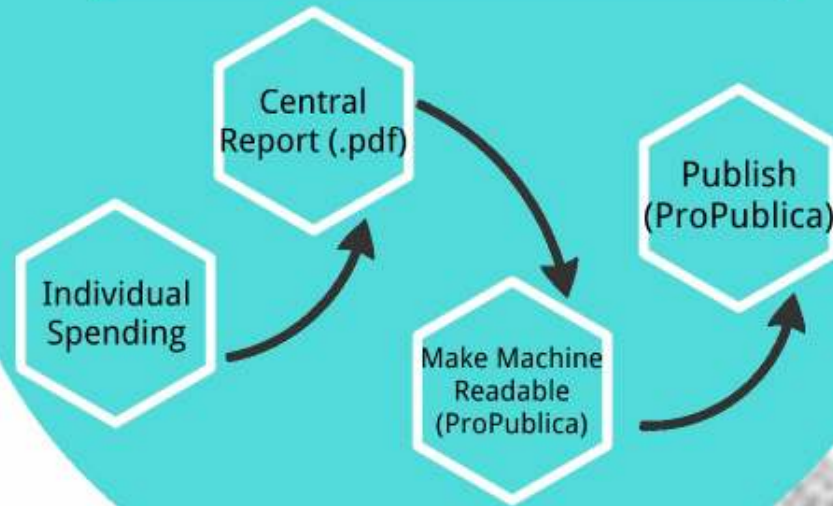
Introduction

Data Pipeline

STATEMENT OF DISBURSEMENTS						
DATE	VEHICLE NO.	PLATE	SERVICE DATES	DESCRIPTION	YES	AMOUNT QUARTERLY AMOUNT
MEMBERS REPRESENTATIONAL ALLOW - Con.						
SOLUTION 3071 CAS - Con.						
03-22	AP	E8C26215	04/01/18	04/15/18	TOWPARKINGTOLLS	5.00
03-25	AP	E8C29000	04/01/18	05/01/18	COMMERCIAL TRANSPORTATION	250.28
03-25	AP	E8C29000	05/01/18	05/15/18	COMMERCIAL TRANSPORTATION	278.28
03-25	AP	E8C29000	05/15/18	05/31/18	COMMERCIAL TRANSPORTATION	333.28
03-26	AP	E8C24126	04/01/18	04/15/18	TOWPARKINGTOLLS	14.55
03-26	AP	E8C29152	04/01/18	05/01/18	COMMERCIAL TRANSPORTATION	684.68
03-25	AP	E8C26159	04/01/18	04/15/18	COMMERCIAL TRANSPORTATION	1,341.54
03-25	AP	E8C26158	04/01/18	05/01/18	PRIVATE AUTO RELEASE	126.88
03-25	AP	E8C26158	05/01/18	05/15/18	TOWPARKINGTOLLS	17.00
03-25	AP	E8C26157	04/01/18	04/15/18	PRIVATE AUTO RELEASE	343.17
03-25	AP	E8C26157	04/15/18	04/30/18	PRIVATE AUTO RELEASE	300.62
03-25	AP	E8C26157	05/01/18	05/15/18	TOWPARKINGTOLLS	12.00
03-25	AP	E8C26158	04/01/18	05/01/18	TOWPARKINGTOLLS	46.88
03-27	AP	E8C26258	04/23/18	05/04/18	TOWPARKINGTOLLS	52.00
OTHER COMMUNICATION FEES:					TRAVEL TOLLS	9,842.04

Objectives

Data Pipeline





Objectives

General Exploration

Spending Patterns

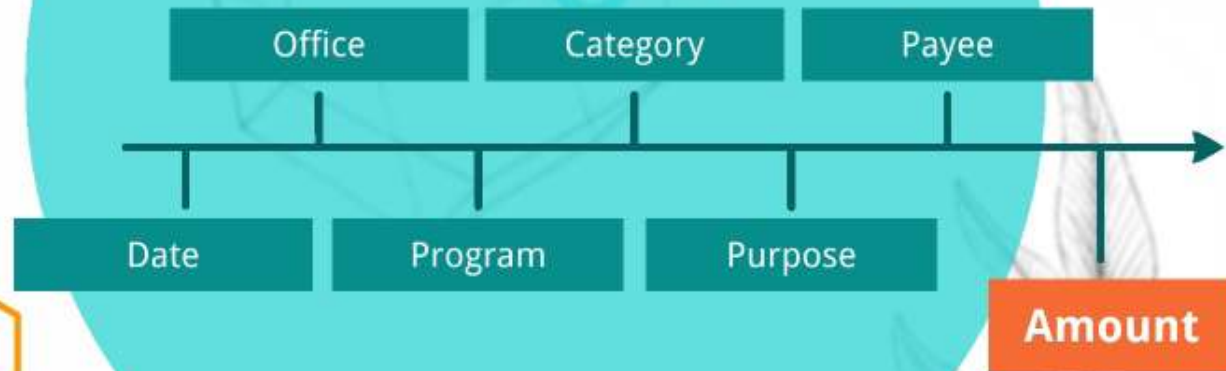
Spending Anomalies

U.S. House of Representatives: 6 Years of Spending Data

M Daniel A Turse
GASP Conference
2018-10-24



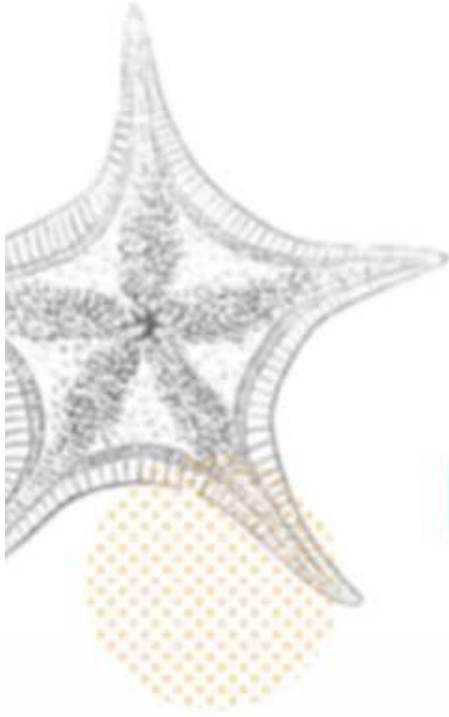
Tabular Data



U.S. House of Representatives: 6 Years of Spending Data

M Daniel A Turse
GASP Conference
2018-10-24





Text Munging

New York City
 sudden revelation of
 the memory that haunts
 her to this day
 North American
 chocolate and
 mint icecream
 on the pavement..
 Millsbury Avenue
 and from her
 office, the same routine everyday,
 coffee with a newspaper, work,
 at the...
 he stopped
 north and
 thought to
 everyone the
 summer-long vacatio
 New York City
 the
 great
 Empre
 of
 dinner at Le Bla

stringr

stringdist

tidytext & hunspell

stringr

purpose
OFFICE SUPPLIES (OUTSIDE)
OFFICE SUPPLIES (OUTSIDE) QTY - 60
OFFICE SUPPLIES (OUTSIDE) QTY - 2
OFFICE SUPPLY (TRANSFER)
COMPUTER HARDW PURCH LESS THAN \$25 000
WARRANTIES
MAINTENANCE / REPAIRS
OFFICE EQUIP PURCH LESS THAN \$25 000
MAINTENANCE / REPAIRS
MAINTENANCE / REPAIRS
WARRANTIES
COMPUTER SOFTW PURCH LESS THAN \$10 000 QTY - 10
MAINTENANCE / REPAIRS QTY - 45
OFFICE EQUIP PURCH LESS THAN \$25 000
MAINTENANCE / REPAIRS
OFFICE SUPPLIES (OUTSIDE)
OFFICE SUPPLIES (OUTSIDE) QTY - 7
OFFICE EQUIPMENT PURCH GREATER THAN OR = \$25K
INTERN

stringdist

var_og	level1	level2	jw_score
office	new members orientation	new member orientation	0.014492754
office	hon tom price	hon tom rice	0.025641026
office	hon patrick j murphy	hon patrick murphy	0.033333333
office	hon vern buchanan	hon vern g buchanan	0.035087719
office	hon doris o matsui	hon doris matsui	0.037037037
office	hon mike rogers	hon mike j rogers	0.039215686
office	hon nita m lowey	hon nita n lowey	0.041666667
office	hon paul tonko	hon paul d tonko	0.041666667
office	hon rush d holt	hon rush holt	0.044444444
office	hon paul ryan	hon paul d ryan	0.044444444
office	hon theodore e deutch	hon theodore deutch	0.049289891
office	hon xavier becerra	hon xavier becarra	0.05664488
office	cloakroom personnel maj	cloakroom personnel min	0.057971014
office	hon richard e neal	hon richard r neal	0.066448802
office	afs immediate office	cao immediate office	0.066666667

tidytext & hunspell

word	WordCnt_ Num	WordCnt_ NumCum	WordCnt_ Pct	WordCnt_ PctCum	RowNum_ Min	purpose_mod
advisor	287	287	8.88	8.88	30	press advisor
legis	229	516	7.08	15.96	170	policy advisor legis couns
dep	176	692	5.44	21.40	13	dep chief of staff commun ops
coord	98	790	3.03	24.44	169	member srvs and outreach coord
svcs	88	878	2.72	27.16	110	deputy director of member svcs
serv	72	950	2.23	29.38	141	web dev hst email rlted serv
mem	63	1,013	1.95	31.33	46	dep dir mem srv outreach
const	61	1,074	1.89	33.22	1104	caseworker dir const svcs rep
coor	60	1,134	1.86	35.08	1109	assistant dist dir grants coor
corres	59	1,193	1.82	36.90	37	staff assistant corres
ofc	57	1,250	1.76	38.66	603	liaison mem ofc prog corr
mbr	54	1,304	1.67	40.33	196	dir of mbr outreach vice chair
constit	50	1,354	1.55	41.88	870	constit svc representative
subc	38	1,392	1.18	43.06	2368	shared subc staff dir
stf	37	1,429	1.14	44.20	1049	professional stf mbr
sys	34	1,463	1.05	45.25	179	sr staff asst sys admin
coun	33	1,496	1.02	46.27	1284	professional staff member coun
svc	31	1,527	0.96	47.23	228	sr adv dir of member svc
dpty	28	1,555	0.87	48.10	987	dpty chf of staff for admin

U.S. House of Representatives: 6 Years of Spending Data

M Daniel A Turse
GASP Conference
2018-10-24

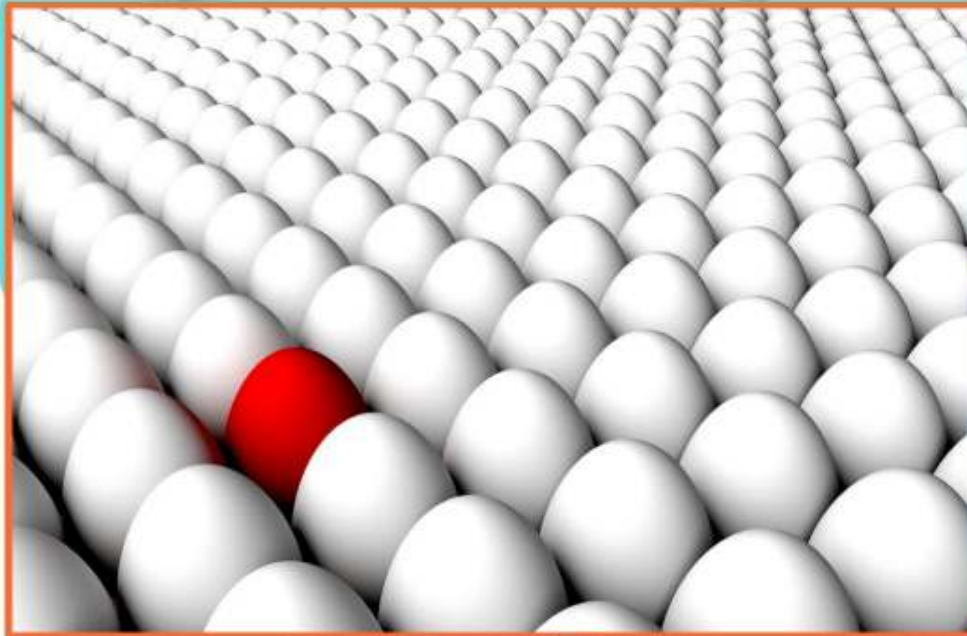


Anomaly Detection Methods

z score

boxplot.stats

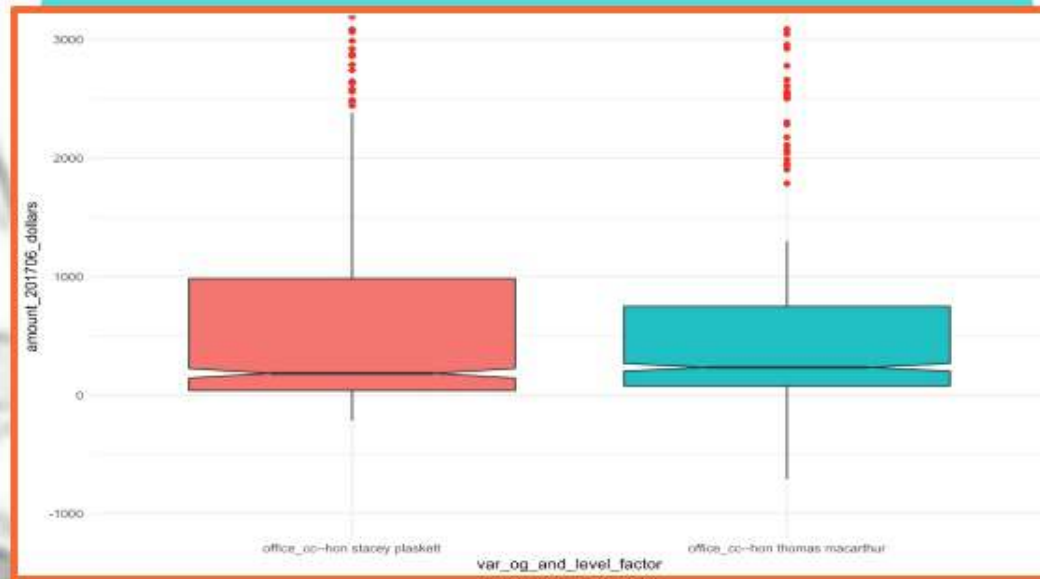
AnomalyDetection



Z SCORE

calculate the z-score by various groups (e.g., purpose) and compare

boxplot.stats



Anomaly Detection

Algorithm 1 S-ESD Algorithm

Input:

X = A time series
 n = number of observations in X
 k = max anomalies (iterations in ESD)

Output:

X_A = An anomaly vector wherein each element is a tuple
(*timestamp, observed value*)

Require:

$k \leq (n \times .49)$

1. Extract seasonal component S_X using STL Variant
2. Compute median \tilde{X}
/* Compute residual */
3. $R_X = X - S_X - \tilde{X}$
/* Detect anomalies vector X_A using ESD */
4. $X_A = \text{ESD}(R, k)$

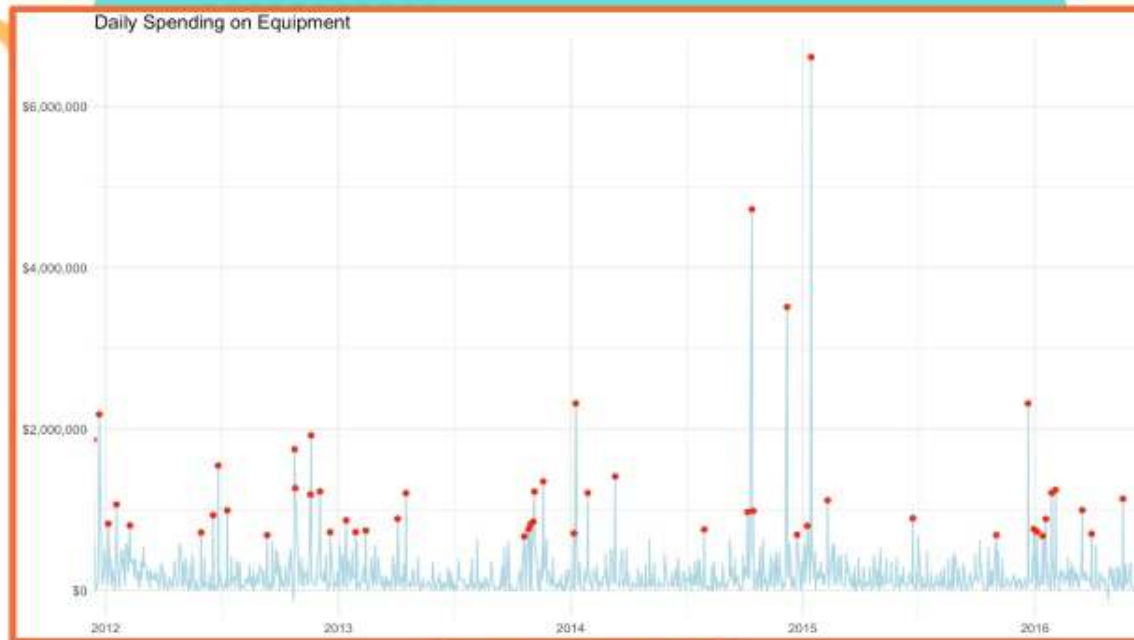
return X_A

U.S. House of Representatives: 6 Years of Spending Data

M Daniel A Turse
GASP Conference
2018-10-24



Results



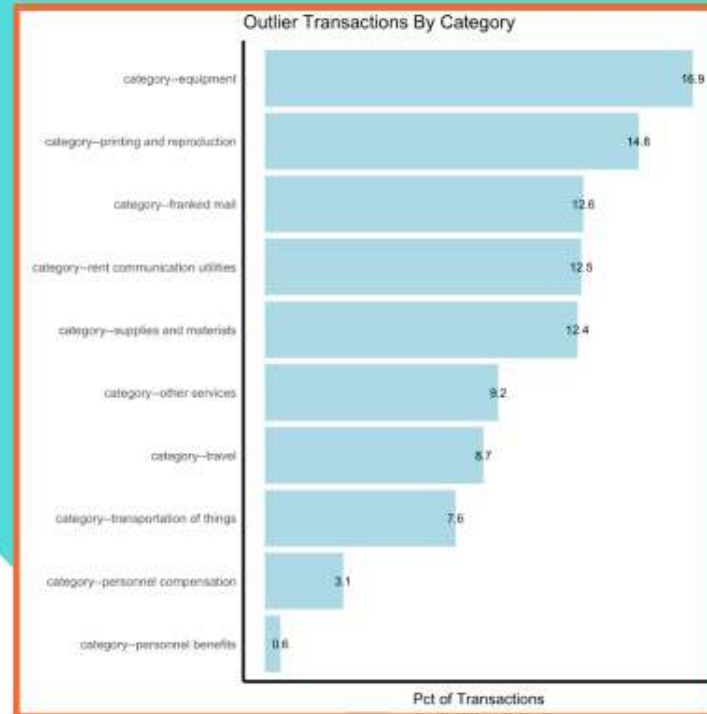
Categories

Equipment

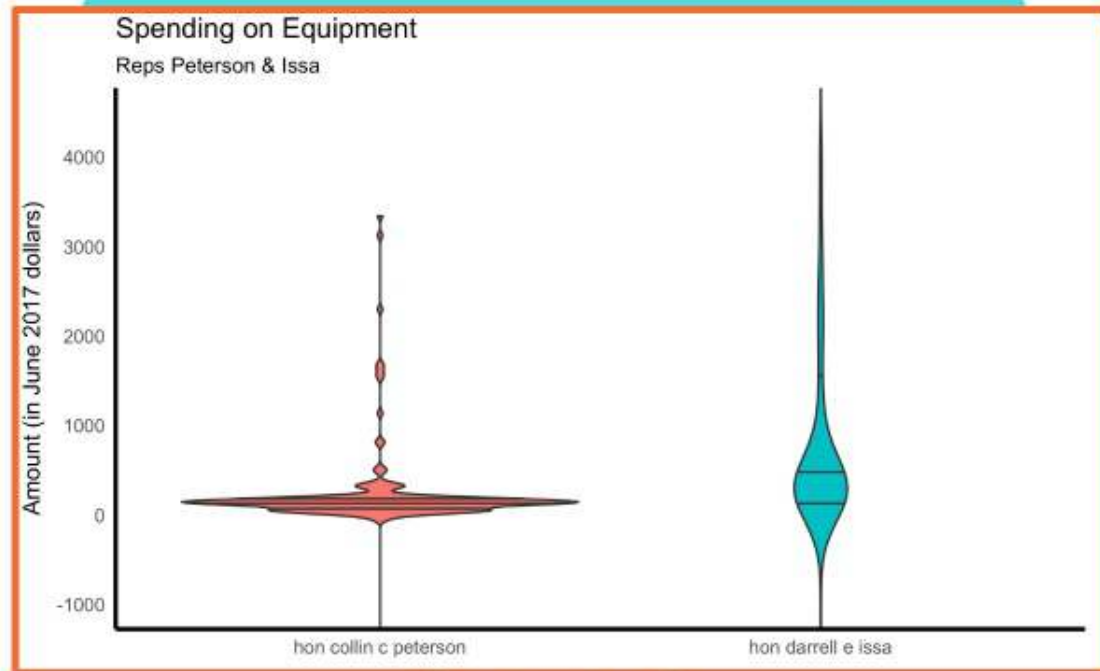
Representative
Comparison

Contact Info

Categories



Equipment



Representative Comparison

	purpose_cc_factor	office_cc_factor	cnt	amount_mean	amount_med	amount_max
1	carpet	hon darrell e issa	1	8509.810	8509.810	8509.81
2	computer hardware purchases less than	hon collin c peterson	3	2920.787	3122.320	3340.73
3	computer hardware purchases less than	hon darrell e issa	28	7994.053	5999.465	28694.69
4	computer software purchases less than	hon darrell e issa	1	2703.320	2703.320	2703.32
5	furniture and fixture less than	hon darrell e issa	1	3344.610	3344.610	3344.61
6	office equip purchases less than	hon darrell e issa	5	4808.288	2272.140	11330.73
7	warranties	hon darrell e issa	1	4120.270	4120.270	4120.27

Contact Info

turse.mdo@gmail.com

<https://www.linkedin.com/in/mdoturse/>

<http://datafordemocracy.org/>

https://github.com/Data4Democracy/house_expenditures

U.S. House of Representatives: 6 Years of Spending Data

M Daniel A Turse
GASP Conference
2018-10-24

