

Hack Oregon Data Notes

- Data cleaning
 - Data oddities
 - Generic names; finding the grass root
- Data type descriptions and definitions
 - Mysterious values and columns
 - Which rows should actually be counted as contributions?

Data oddity: max transactions by 1 committee: 4999?

Why do the PACs with the 15 largest numbers of transactions all have exactly 4999 transactions?

```
SELECT "Filer", count("Tran_Id")  
FROM fins  
GROUP BY "Filer"  
ORDER BY count("Tran_Id") DESC
```

Committee name (from filer column)	Transactions per committee
Dentists of Oregon PAC	4999
Oregon Realtors Political Action Committee	4999
Oregon Trial Lawyers Association PAC	4999
Promote Oregon Leadership PAC	4999
Farmers Employee and Agent PAC of Oregon	4999
Oregon Right to Life PAC	4999
Local 48 Electricians PAC	4999
Oregon Family Council PAC	4999
Lane County Republican Central Committee	4999
Friends of Chris Dudley	4999
Oregon Medical PAC	4999
Oregon Health Care Association PAC	4999
Kate Brown Committee	4999
Citizen Action for Political Education	4999
Future PAC, House Builders	4999
Oregon Council of National Electrical Contractors Association Chapter PAC	4649

Generic contributor names

Contributor_Payee	Count(Tran_Id)
Miscellaneous Cash Contributions \$100 and under	73382
Miscellaneous Cash Expenditures \$100 and under	33704
Miscellaneous Personal Expenditures \$100 and under	13939
	5173
Miscellaneous In-Kind Contributions \$100 and under	2774
C&E Systems	2444

```
select "Contributor_Payee", count("Tran_Id")
from fins
group by "Contributor_Payee"
order by count("Tran_Id") desc
```

100 and under =
grass root, but it's all
under one name:

"Miscellaneous Cash
Contributions \$100
and under"

Perhaps we
could append the
address to the
generic names to
try to get a
better idea of
the number of
unique
contributors?

Sub_Type column has 30 unique values:
which should we include when aggregating total donated? received?

Sub_Type	count	total \$ Amount	Sub_Type	count	total \$ Amount
	2		Loan Forgiven (Non-Exempt)	101	689404.46
Cash Expenditure	172199	203447150	Return or Refund of Contribution	746	605245.58
Cash Contribution	257784	197815190.3	Account Payable Rescinded	98	190669.62
In-Kind Contribution	18489	21741208.1	Interest/Investment Income	5435	182568.45
Expenditure Made by an Agent	1937	16031746.71	Miscellaneous Account Receivable	201	164940.23
Account Payable	6843	15074182.57	Uncollectible Pledge of Cash	121	161729.33
Loan Received (Non-Exempt)	797	5499504	Unexpended Agent Balance	21	155111.64
Personal Expenditure for Reimbursement	29490	4241767.91	In-Kind/Forgiven Personal Expenditures	229	149276.18
Miscellaneous Other Receipt	8595	3485937.93	In-Kind/Forgiven Account Payable	59	93124.41
Items Sold at Fair Market Value	15173	3434928.16	Loan Received (Exempt)	11	35446.33
Refunds and Rebates	1474	2201682.59	Nonpartisan Activity	40	24505.72
Loan Payment (Non-Exempt)	411	1895111.52	Loan Payment (Exempt)	16	14354
Pledge of Cash	2350	1778273.1	Uncollectible Pledge of In-Kind	2	12646.22
Lost or Returned Check	1089	1442554.51	Pledge of Loan	4	820
Pledge of In-Kind	154	1156413	Cash Balance Adjustment	330	-43303.76
Miscellaneous Other Disbursement	1180	885922.09			

Original (amended)

Tran_Id	Original_Id	Date	Original	Filer	Contributer	Sub_Type	Amount
812090	811405	7/1/10	Amended	Oregon Business Association PAC	Promote Oregon Leadership PAC (682)	Cash Expenditure	500

I've checked three of the "original_id"s for rows marked "Amended" and haven't found the records for any.

Amount and Aggregate_Amount

Tran_I d	Original_I d	Date	Original	Filer	Contributer	Sub_Type	Amount	Aggregate_A mount
200720	200720	1/2/08	Original	Basic Rights Oregon Equality PAC	Bank of America	Cash Expenditure	282.33	667.76
200721	200721	1/2/08	Original	Basic Rights Oregon Equality PAC	Miscellaneous Cash Contributions \$100 and under	Cash Contribution	25	0
180680	180680	12/31/07	Original	Basic Rights Oregon Equality PAC	Bank of America	Cash Expenditure	3	385.43
180685	180685	12/31/07	Original	Basic Rights Oregon Equality PAC	Leila Wice	Cash Contribution	360	360

What is aggregate amount referring to?

We've decided we should ignore this column – it can only serve to confuse and waste time

Missing committee ids

- 6821 rows of fins data have committee ids that are not found in the comms table
 - (56538 have committee ids that are found in the comms table)

```
select *  
from fins  
where "Contributor_Payee_Committee_ID"  
      not in      (select "Committee_Id"  
                    from comms)  
and "Contributor_Payee_Committee_ID" is not null  
/*6821 rows*/
```

```
select *  
from fins  
where "Contributor_Payee_Committee_ID"  
      in      (select "Committee_Id"  
                from comms)  
and "Contributor_Payee_Committee_ID" is not null  
/*56538 rows*/
```


Data idiosyncrasies

- Candidates switching parties (see candidate filings table)
 - Many switch parties at some point
 - Some switch more than others
 - ex: Arthur Robinson

Election.Txt	Candidate.Office	Filetype.Descr	Party.Descr	Filed.Date
2014 Primary Election	Representative in Congress, 4th District	Fee	Republican	11/5/13
2012 General Election	Representative in Congress, 4th District	Minor Party	Constitution	7/31/12
2012 General Election	Representative in Congress, 4th District	Nominated	Republican	6/14/12
2012 Primary Election	Representative in Congress, 4th District	Fee	Republican	9/8/11
2010 General Election	Representative in Congress, 4th District	Minor Party	Constitution	8/19/10
2010 General Election	Representative in Congress, 4th District	Minor Party	Independent	8/18/10
2010 General Election	Representative in Congress, 4th District	Nominated	Republican	6/17/10
2010 Primary Election	Representative in Congress, 4th District	Fee	Republican	3/9/10

Making an simple interaction format (sif) table

- SIF: Used to define networks
 - columns: from, to and amount
 - Added column “date” to assure uniqueness
- Not all entities/nodes have numeric ids, but, where they exists they should be used as unique identifiers
 - Those without ids are assigned the value “Committee_Payee” field
 - Those NULLs for id and Committee_Payee are given the name of the Sub_Type

sub_type	Number of sub_type occurrences	Total \$ amount associated with sub_types having blank “contributor_payee” field
Nonpartisan Activity	1	25
Interest/Investment Income	655	15908.72
Cash Balance Adjustment	269	-90170.91
Items Sold at Fair Market Value	4248	1928354.16