# Mathematics in Election Verification

Stephanie Singer Hatfield School of Government

Portland State University

Funding from NSF Award #1936809 (CNS) GitHub.com/sfsinger19103/election\_anomaly

#### **A Fundamental Equation**



# of voters = # of ballots

#### Data Sources



# volers

Poll books?

Observation?

Voter Registration Database?

## # ballots

Count paper? Voting-counting systems?

#### Data Quality



# volers

## # ballots

Edge cases

Timing

Edge cases

Million of sheets of paper

Can we rely solely on computer-aided systems?

#### Access



# ballots

may be unavailable may be expensive (\$35K !?!) may be unavailable

### Timing



#### Timing

## What can data analysis do in time to be useful?

i.e., before certification

#### North Carolina 2018



#### Q Sections ≡ Ethe Washington Post Democracy Dies in Darkness Politics North Carolina sets new primaries, general election in scandal-plagued 9th Congressional District



#### North Carolina 2018























#### Can math help generalize?



Anomalies identified

Candidates assess, maybe push for investigation

Investigations maybe cancel certification

#### Can math help generalize?

Data analysis Anomalies identified



Automate (as much as possible) data aggregation As appropriate, use sophisticated algorithms for anomaly detection

Work in progress, funded by NSF Grant # 1936809 Main tools: Python, postgreSQL

#### Input

20 MB flat file from Board of Elections, available after voting ends

						resul	ts_	oct_20'	1811	06.tx	t						
County	Election	Date	Precinct		Contest	Group ID	)	(	Cont	est 1	Туре	Co	ontest	Name	Choice	Choice P	Party
Vote For		Electio	on Day	One Stop	)	Absentee	e by	/ Mail		I	Prov	isiona	1	Total	Votes	Real Pre	cinct
ALAMANCE		11/06/2	2018	064	1228	S	NC	COURT	0F	APPE/	ALS	JUDGE	SEAT 3	3	Michael	Monaco,	Sr.
LIB	1	59	65	2	1	127	Υ										
ALAMANCE		11/06/2	2018	03N	1228	S	NC	COURT	0F	APPE/	ALS	JUDGE	SEAT 3	3	Michael	Monaco,	Sr.
LIB	1	59	38	1	0	98	Y										
ALAMANCE		11/06/2	2018	03S	1228	S	NC	COURT	0F	APPE/	ALS	JUDGE	SEAT 3	3	Michael	Monaco,	Sr.
LIB	1	106	108	0	3	217	Y										

#### Which info do we need in addition to the info in the file??

BallotMeasureSelection.txt datafile.txt Election.txt metafile.txt name.txt Office.txt Party.txt Party.txt remark.txt ReportingUnit.txt schema\_name.txt

< 1 MB contextual info, must be created 'by hand' in advance

### Munging info (1 of 3)

• • •	l'					resul	ts_p	oct_20'	1811	06.tx	t						
County	Election	Date	Precinct		Contest	Group ID	)	(	Cont	test	Туре	e Co	ontest	Name	Choice	Choice F	Party
Vote For		Electi	on Day	One Stop	p	Absentee	e by	/ Mail			Prov	isiona	al	Total	Votes	Real Pre	ecinct
ALAMANCE		11/06/3	2018	064	1228	S	NC	COURT	0F	APPE	ALS	JUDGE	SEAT 3	3	Michael	Monaco,	Sr.
LIB	1	59	65	2	1	127	Y										
ALAMANCE		11/06/3	2018	03N	1228	S	NC	COURT	0F	APPE	ALS	JUDGE	SEAT 3	3	Michael	Monaco,	Sr.
LIB	1	59	38	1	0	98	Y										
ALAMANCE		11/06/3	2018	03S	1228	S	NC	COURT	0F	APPE	ALS	JUDGE	SEAT 3	3	Michael	Monaco,	Sr.
LIB	1	106	108	0	3	217	Y										

#### What are the columns of the file from the Board of Elections?

Column	Data Type
County	TEXT
Election Date	DATE
Precinct	TEXT
Contest Group ID	INTEGER

### Munging info (2 of 3)

						re	sults_p	oct_20	181	106.txt					
County	Election	n Date	Precinct	t	Contest	Group	ID		Con	test T	уре	Contest	Name	Choice	Choice Party
Vote For	r	Electio	n Day	0ne	Stop	Absen	tee by	Mail		P	rovisi	onal	Total	Votes	Real Precinct
ALAMANCE		11/06/2	018	064	1228	S	NC	COURT	0F	APPEA	LS JUD	GE SEAT	3	Michael	Monaco, Sr.
LIB	1	59	65	2	1	127	Y								
ALAMANCE		11/06/2	018	03N	1228	S	NC	COURT	0F	APPEA	LS JUD	GE SEAT	3	Michael	Monaco, Sr.
LIB	1	59	38	1	0	98	Y								
ALAMANCE		11/06/2	018	03S	1228	S	NC	COURT	0F	APPEA	LS JUD	GE SEAT	3	Michael	Monaco, Sr.
LIB	1	106	108	0	3	217	Y								

#### Which columns have which kind of election result?

Column Name (BoE file)	Count Type (NIST CDF)
Election Day	election-day
One Stop	stop early
Absentee by Mail	absentee-mail
Provisional	provisional
Total Votes	total

### Munging info (3 of 3)

• • •				resu	lts_pct_2	01811	06.txt				
County Election	n Date Precinc	t	Contest	Group II	D	Cont	est Type	e Contest	Name	Choice	Choice Party
Vote For	Election Day	One Sto	p	Absentee	e by Mai	1	Prov	isional	Total	Votes	Real Precinct
ALAMANCE	11/06/2018	064	1228	S	NC COUR	RT OF	APPEALS	JUDGE SEAT	3	Michael	Monaco, Sr.
LIB 1	59 65	2	1	127	Y						
ALAMANCE	11/06/2018	03N	1228	S	NC COUR	RT OF	APPEALS	JUDGE SEAT	3	Michael	Monaco, Sr.
LIB 1	59 38	1	0	98	Y						
ALAMANCE	11/06/2018	03S	1228	S	NC COUR	RT OF	APPEALS	JUDGE SEAT	3	Michael	Monaco, Sr.
LIB 1	106 108	0	3	217	Y						

#### How do we read common-data-format elements from rows of file?

<b>NIST CDF Element</b>
Office
ReportingUnit
Party

....

NIST CDF Field Name Name Name

Recipe from BoE file Contest Name County; Precinct Choice Party

#### Output

## What is the "worst" bar chart we can draw for each contest?

#### Output





#### Output



# Mathematics in Election Verification

Stephanie Singer Hatfield School of Government

Portland State University

Funding from NSF Award #1936809 (CNS)