

Data Enhancement and Analysis of Federal Acquisition Databases

Ningning Wu, M. Eduard Tudoreanu, Richard Wang

Institute for Chief Data Officers
University of Arkansas at Little Rock
Massachusetts Institute of Technology

Problem Statement

- A wealth of information is collected and available
 - Governmental sources
 - Social media
 - Websites
- Valuable information could be used for acquisition
 - Risk assessment, time assessment, logistics, employment bottlenecks
- Federal acquisition and purchasing databases have rich information about past contracts
 - › More knowledge could be extracted

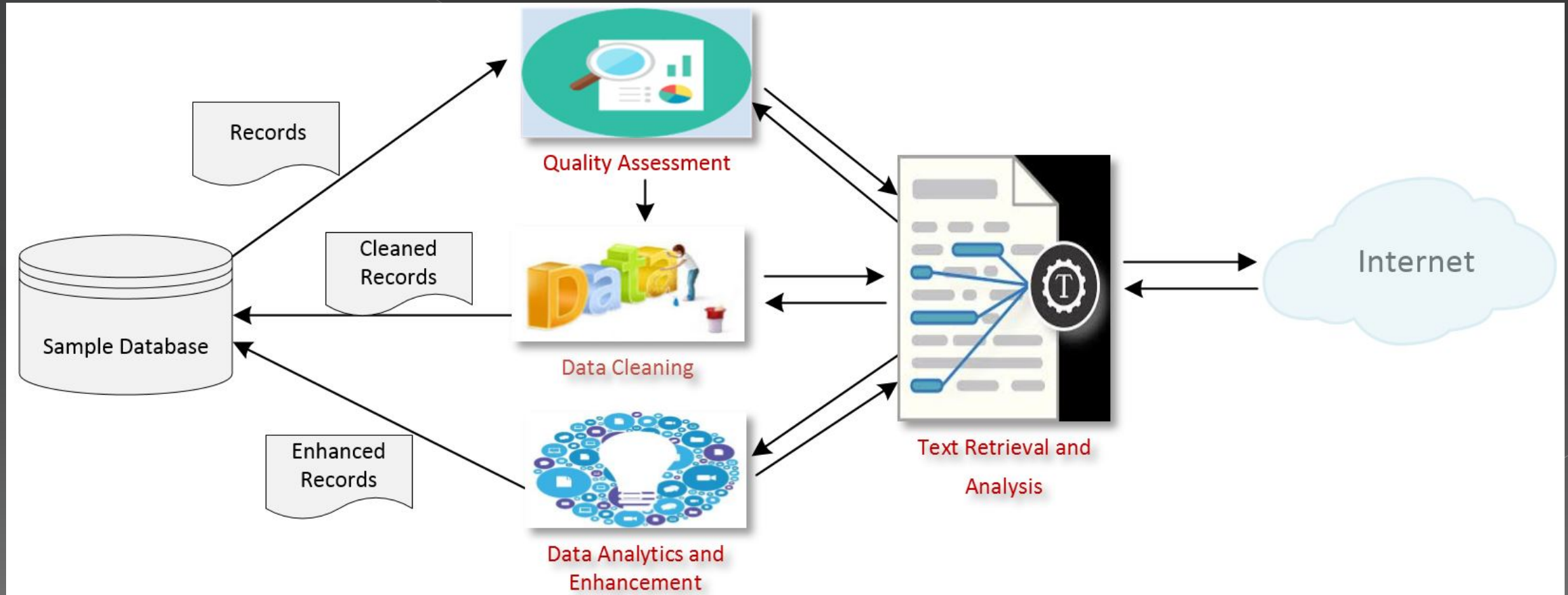
Importance to Acquisition

- Social, economic, natural changes often get captured in various data sources
 - Can provide a more global and integrated view of factors affecting contracts and projects
 - Complementary to
 - subject-matter expert knowledge
 - contractor information and expertise

Focus of Presentation - Agenda

- Data quality assessment
 - › check two purchasing databases individually and against each other
- Use cases for data-driven process improvement
 - › Unique service contractors
 - › Natural disaster high-risk contractors
 - › Undisclosed contractors

Data Enhancement and Analytics System Framework



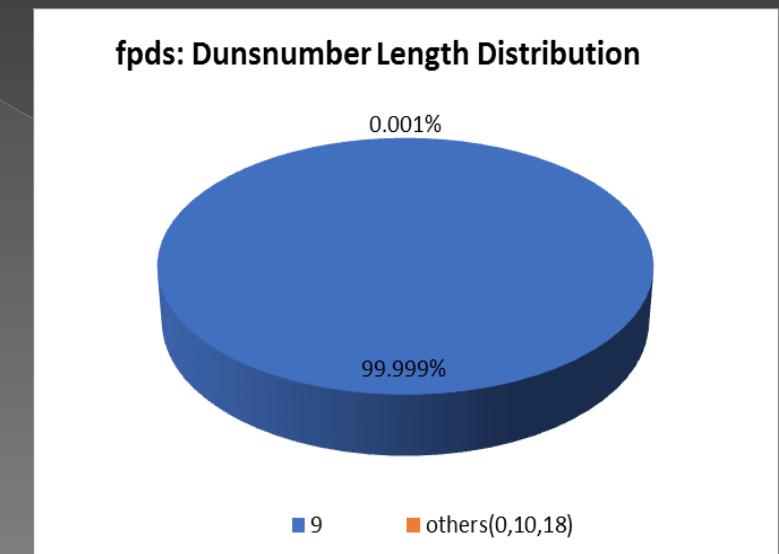
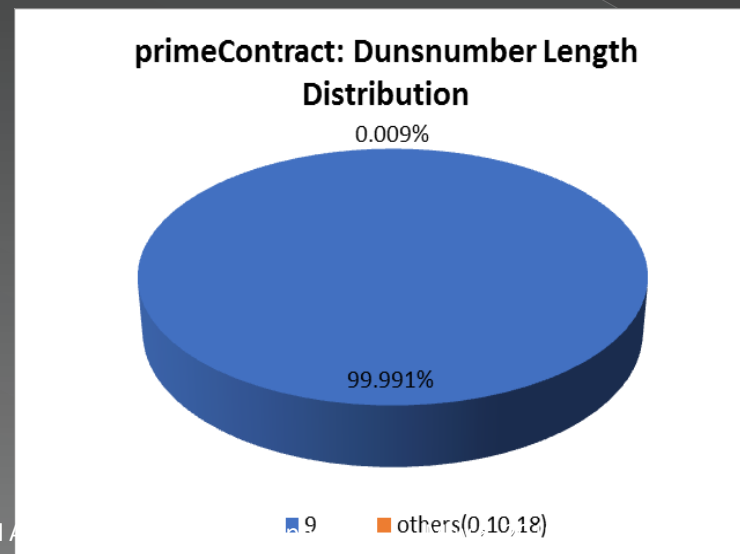
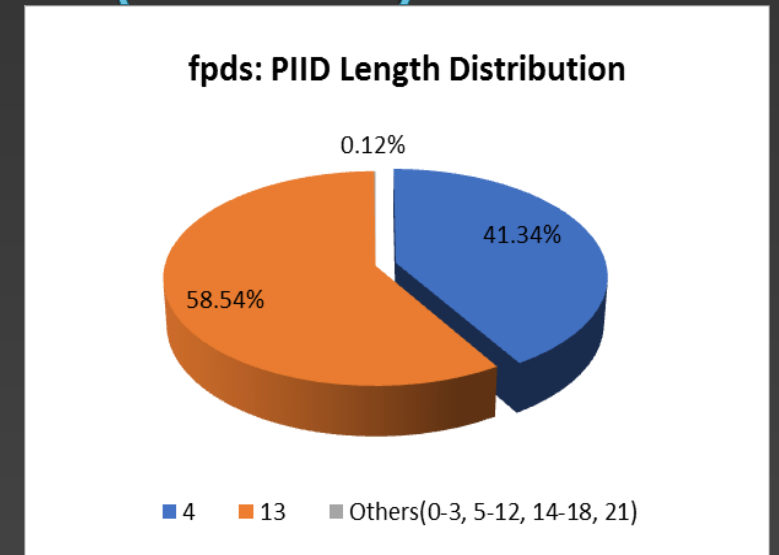
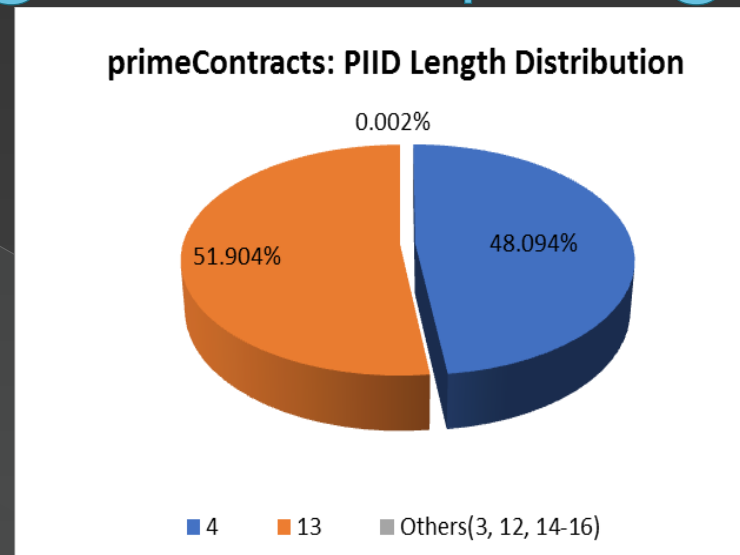
Comparison of databases of usaspending.gov and fpds.gov

Profiling of tables

Table Name	ColCnt	CompleteCols/ SingleValCols	EmptyCols	IncompleteCols
PrimeContracts	221	50/1	0	162
SubContracts	101	41/0	3	57
PrimGrants	67	32/5	2	33
SubAGrants	101	29/4	25	47
fpds	210	74/3	1	136

Comparison of databases of usaspending.gov and fpds.gov (cont.)

Quality Assessment of Field Length



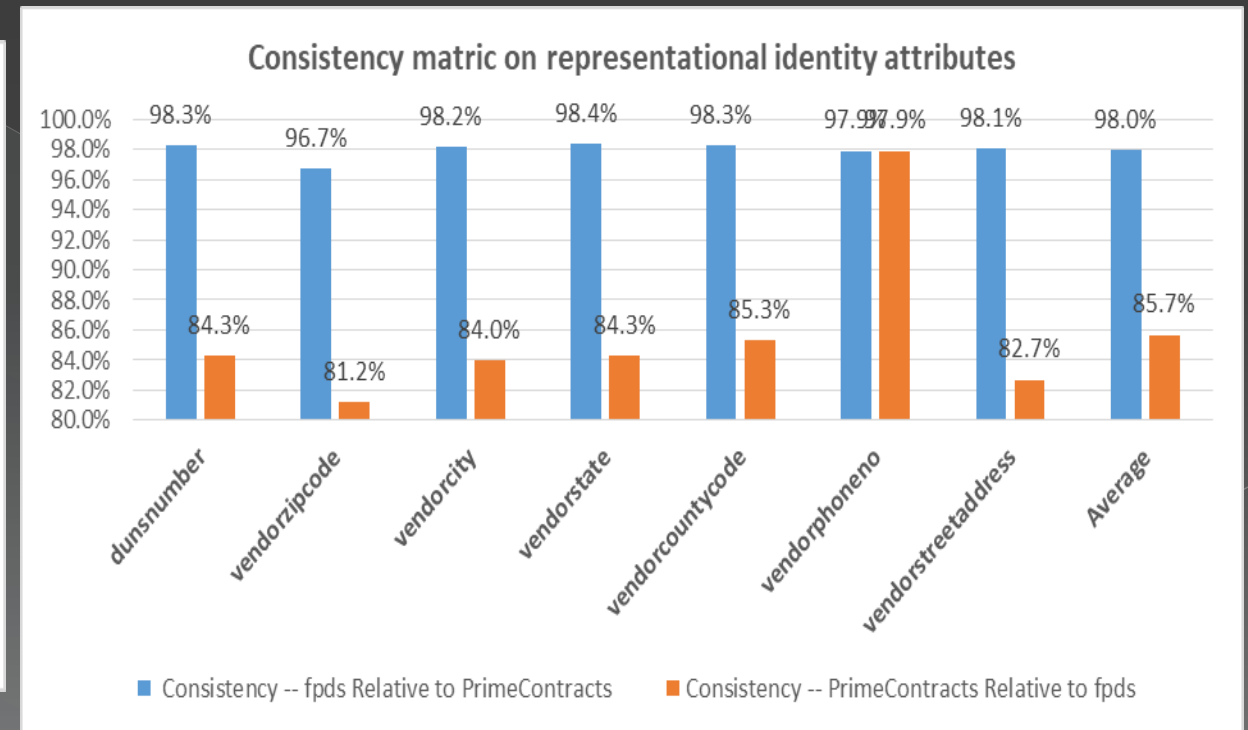
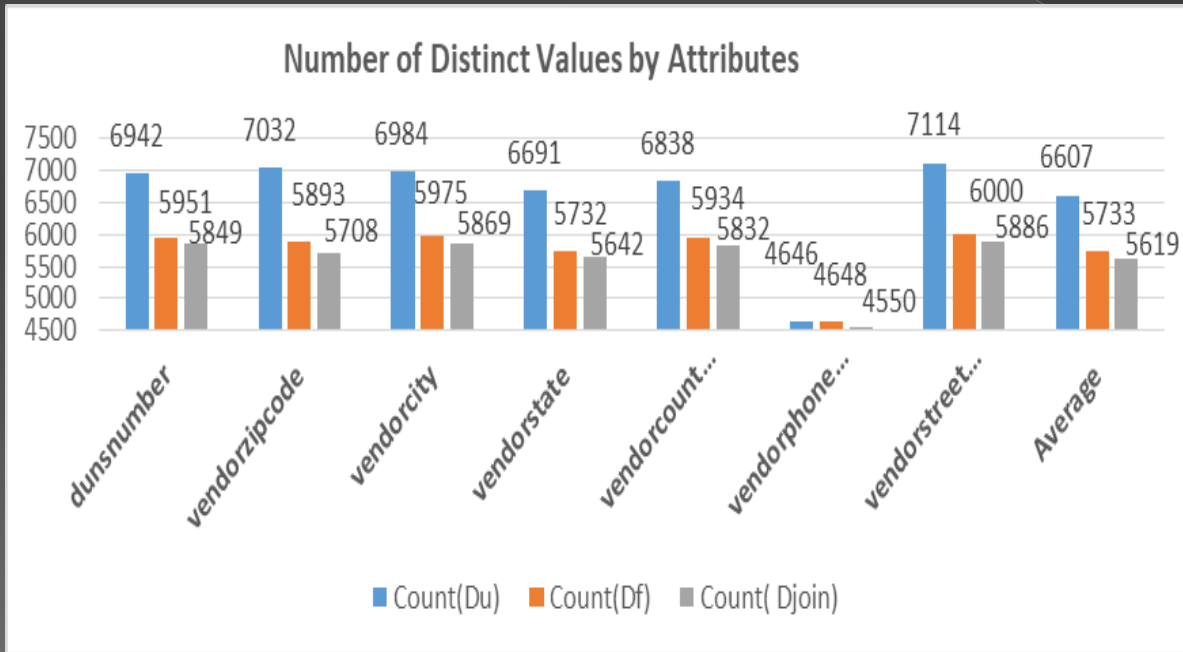
Comparison of databases of usaspending.gov and fpds.gov (cont.)

- Record mapping to measure data consistency
 - Random sample of 5000 PIID that appear in both sources
 - Records for those PIID are retrieved from primeContract (D_u) and fpds (D_f)
 - Standardize and transform data; remove duplicates
 - The equijoin is applied on two datasets; create D_{join}
 - D_{join} has common records
 - Calculate record-wise consistency measures $|D_{join}| / |D_u|$ and $|D_{join}| / |D_f|$

Comparison of databases of usaspending.gov and fpds.gov (cont.)

Record Mapping Results

- D_u consistently has more distinct values for each attribute than D_f . The number of distinct values for each attribute in table D_{join} indicates the number of common values between D_f and D_u .



Comparison of databases of usaspending.gov and fpds.gov (cont.)

Schema Mapping

Table	#Unique Attributes	#common attributes	#total attributes
Fpds	30	180	210
usaspending	41	180	221

- Example of common attributes:

Mapping Attributes		
	Attributes in fpds	Matched Attributes in PrimeContracts
1	awardID_awardContractID_PIID	piid
2	awardID_awardContractID_agencyID	agencyid
3	awardID_awardContractID_modNumber	modnumber
4	awardID_awardContractID_transactionNumber	transactionnumber
5	awardID_referencedIDVID_PIID	idvpiid

- Example of Unique attributes:

Unique Attributes		
	Unique Attributes in fpds	Unique Attributes in PrimeContracts
1	competition_idvTypeOfSetAside	congressionaldistrict
2	competition_numberOfOffersReceived	divisionnumberorofficecode
3	competition_numberOfOffersSource	emergingsmallbusinessflag
4	competition_typeOfSetAsideSource	fiscal_year
5	contractData_inherentlyGovernmentalFunction	hubzoneflag
6	contractData_listOfTreasuryAccounts_treasuryAccount_initiative	isarchitectureandengineering

Use case overview

- **Use case 1:** Unique-service contractors provide unique products and services. They could be a potential weak link in a project because it would be hard to find alternatives to fill their places if they failed
- **Use case 2:** High risk contractors are those located in areas of high risks in natural disasters such as hurricane, tornado, flood, and wild fire. They could become a risk to a project
- **Use case 3:** Some prime contractors make use of additional subcontractors, which are not necessarily known to an acquisition expert. A method of uncovering these undisclosed contractors is explored

Unique Service Contractors

- There are 379 distinct NAICS codes of all contractors in total.
- 78 NAICS codes have only one contractor associated with it. These contractors are considered as unique service contractor.
- On average, each unique service contractor is involved in 37 different projects.

Rank	No. of Distinct Projects
1	382
2	343
3	245
4	237
5	138
6	117
7	91
8	69
9	61
10	59

Top 10 contractors with most number of projects

Natural Disaster Data

- We have retrieved the natural disaster data for **each US county** between the years 1950 and 2018 from the National Centers for Environmental Information (Formerly the National Climatic Data Center NCDC).
- The data cover all types of natural disasters, including flood, tornado, hurricane, blizzard, high wind, flash flood, hail, dust storm, etc
- We also retrieved the zip codes for every US county so that we are able to produce natural disaster map by zip codes.

Natural Disaster Data (cont.)

- The project focuses on disasters that could cause severe damages and significantly affect the normal life and business operations of local communities such as Tornado, Hurricane, Flood, and Wildfire.
- The high risk areas of each disaster are defined as follows:
 - The high-risk flooding areas are identified as those that have at least 10 episodes of flood in the last twenty years;
 - The high risk hurricane areas are those that have at least one hurricane in last 20 years;
 - The high-risk wildfire areas are those that have at least one wildfire that lasted more than 1 day in last 20 years;
 - The high-risk tornado areas are those that have at least one category 3 or above tornado in the last 20 years.

Contractor Distributions in High-Risk Areas

- Number of contractor zips vulnerable to each disaster type

Disaster type	Flood	Hurricane	Tornado	Wildfire
# zip codes	5959	780	1182	1831

- Distribution of contractor principal zip and DUNS

#Disaster Types	#zip codes	%zip_population	#DUNS	% DUNS_population
1	2165	7.8%	13373	42.3%
2	3548	12.9%	10965	34.6%
3	1004	3.6%	2733	8.6%
4	69	0.25%	141	0.44%
Total:	6786	23.7%	27072	86.0%

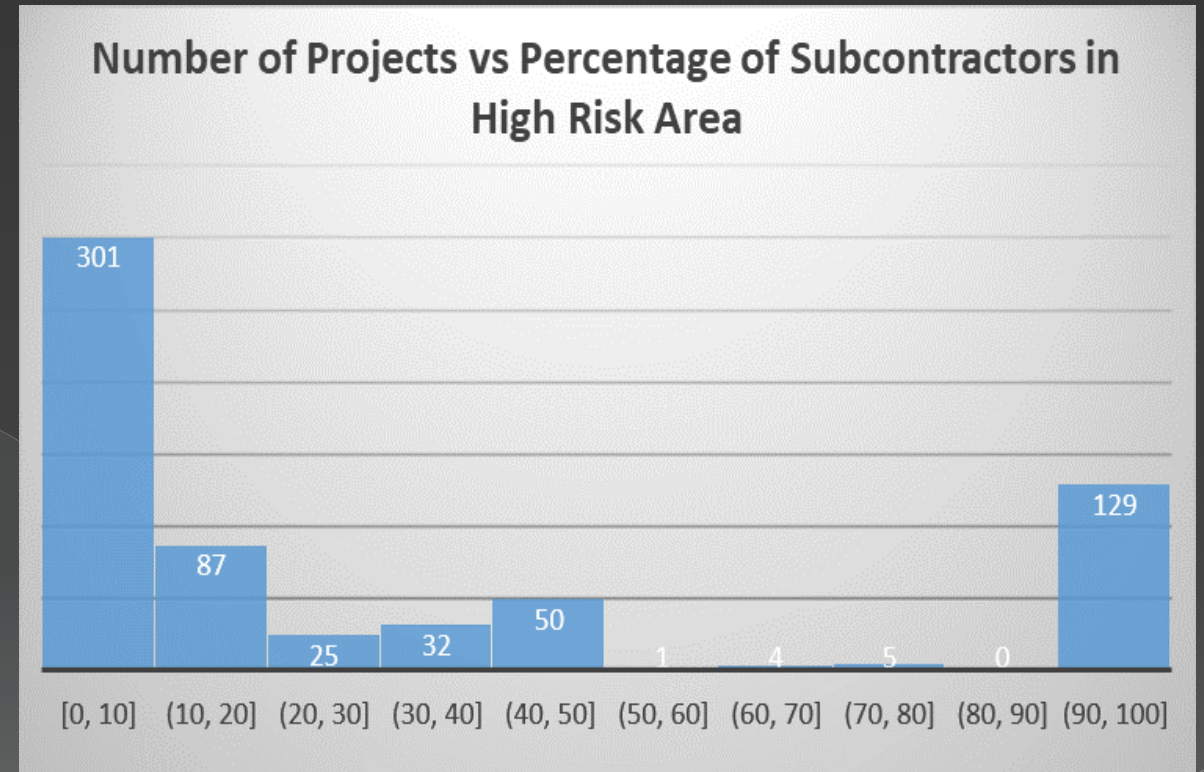
High Risk Contractors

- Contractors that located in an area vulnerable to all four disaster types are considered to have a high risk
- Top 10 projects with the most number of high-risk contractors

Rank	No. of High-risk Contractors
1	59
2	49
3	43
4	37
5	36
6	31
7	27
8	24
9	23
10	19

High Risk Contractors

- There are total 588 projects have at least one high risk subcontractors.
- Figure below shows the distribution of projects by their percentage of contractors that are vulnerable to all four types of natural disaster

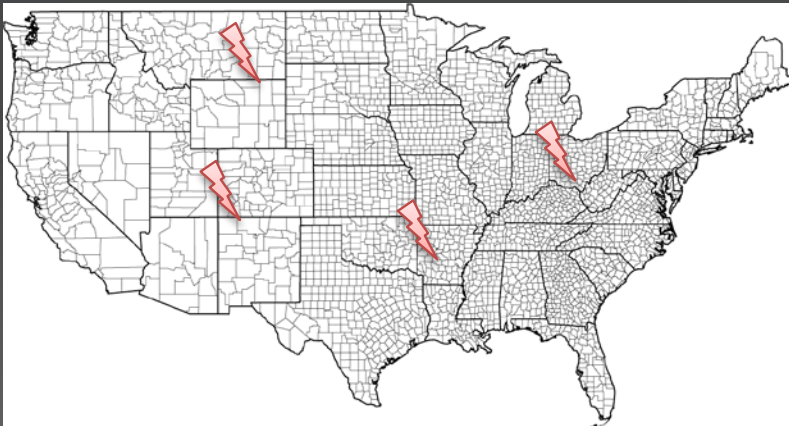


Searching for Undisclosed Contractors

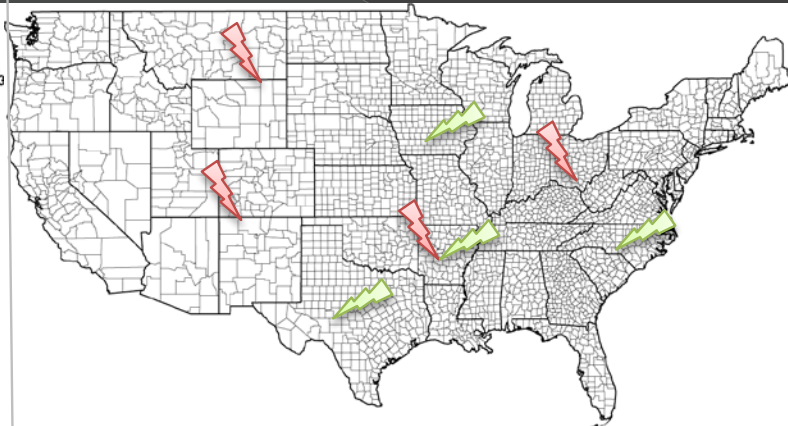
- Prime contractor may outsource tasks to undisclosed entities for
 - › competitive advantage
 - › security
 - › confidentiality
- Acquisition experts may need to take into account undisclosed subcontractors
 - › Keep them secret if desired
 - › Uncover those already participating in many projects

Look for Employment Changes

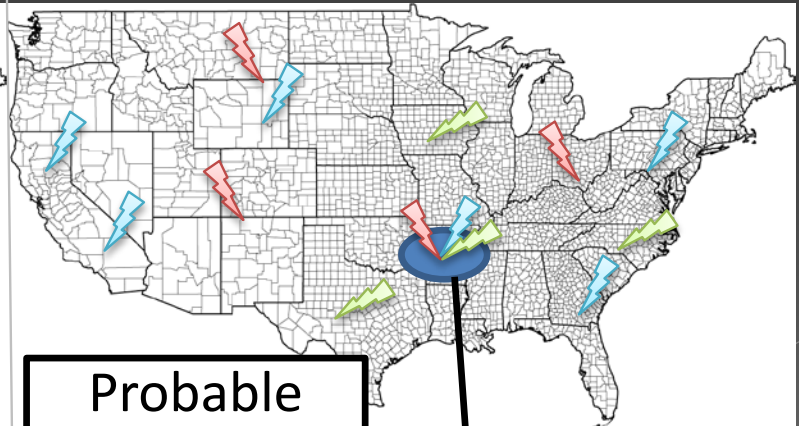
- Bureau of Labor Statistics: monthly employment information
 - > County or city level
 - > NAICS



Event 1
(contract start, end, or change)



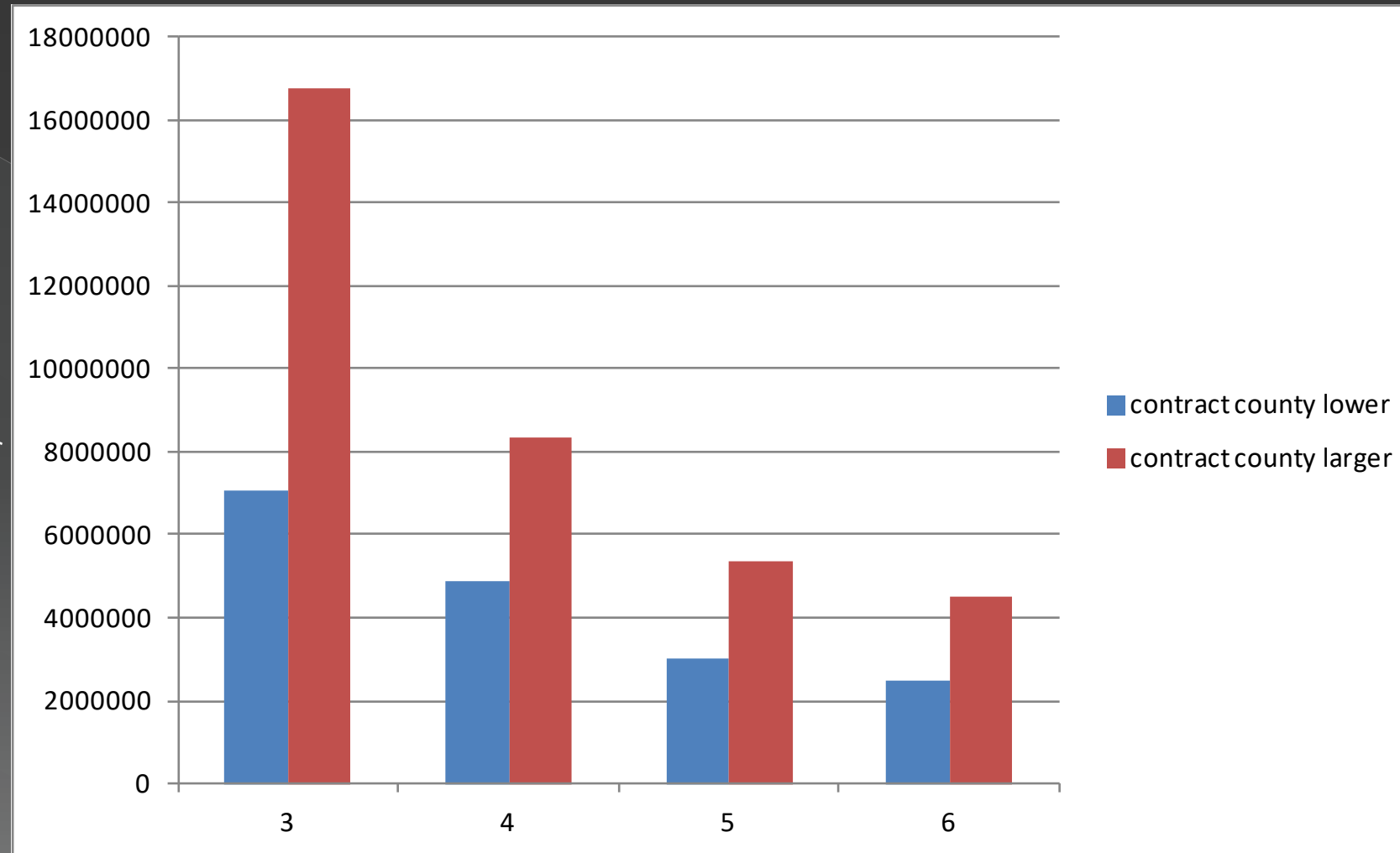
Event 2
(contract start, end, or change)



Event 3
(contract start, end, or change)

Correlate to fpds contract data

- Looked at 2016, about 23,000 contracts
- Contract start events
 - county of award above 70% of rest of counties



Concluding Thoughts

- Small unique service contractors might be potential weak link in a project
 - A couple of unique service contractors are small companies that appear to provide very unique products and services
 - Need strategic plan to handle the situations when they are unavailable.
- Information on high risk areas of natural disasters is beneficial as it would help project managers calculate the risk to project timeline and develop strategies to mitigate the risk
- Undisclosed contractors, especially those working on many contracts, may also pose a systemic risk