An Overview:

# Analytics in [Procurement] Fraud Investigations

SA Erik Halvorson, DOE OIG

# Learning Objectives:

1. Understand how the use of analytics can augment previously used investigative and oversight techniques.

2. Understand how a fraud analytics framework can be implemented based on crawl, walk, run method.

3. Know how to use initial organization and descriptive statistics techniques to help identify outliers for follow up.

4. Observe possibilities for more advanced analytics strategies.

# Introduction:

Anti-Fraud Program Manager with DOE OIG's Office of Investigations for New Funding Endeavors with 13 years as a SA and 18 years as a LEO. Experience ranges from nuclear security to military police patrol/operations to Investigations into child exploitation and large white-collar fraud.

Experienced in domestic and international law enforcement. Assigned to DOJ's ICCTF in Saudi Arabia and joint constabulary interface in UK. Assigned to 6 countries and based in 4 states. Degrees in Criminal Justice, Accounting, Applied Analytics, and finishing a PhD in Research Methods and Statistics with a focus in Analytics and Program Evaluation.

(720) 474-8553

erik.Halvorson@doe.gov

Golden Rule/Disclosure:

There is no silver bullet in the world of Analytics. This is especially true in the much smaller world of fraud analytics. Each situation involves its own unique solution and use of data.

# What is/are Analytics:

Analytics = Tactics

Where tactics incorporate the *intentional* use of items on scene to gain an advantage towards successfully completing an endeavor.

Fraud analytics, therefor, is the use data in an *intentional* way to gain an investigative advantage towards successfully mitigating fraud.

Just like tactics, we can learn through crawl, walk, run model:

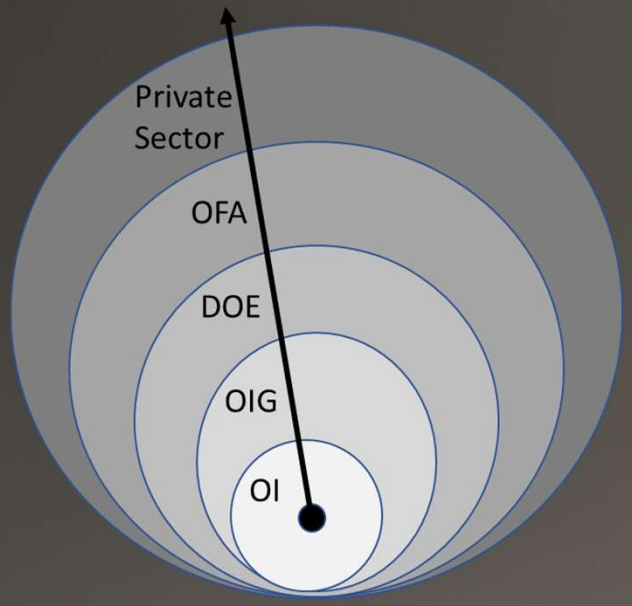
Continuing to beat this metaphor, Analytics = Tactics 

Crawl: Foundational questions - What data do you have? What problems are you trying to solve? How do you organize?

Walk: Employing data in one-off type situations to combat a specific fraud risk.

Run: Employing at-scale solutions to detect, mitigate, and combat fraud.

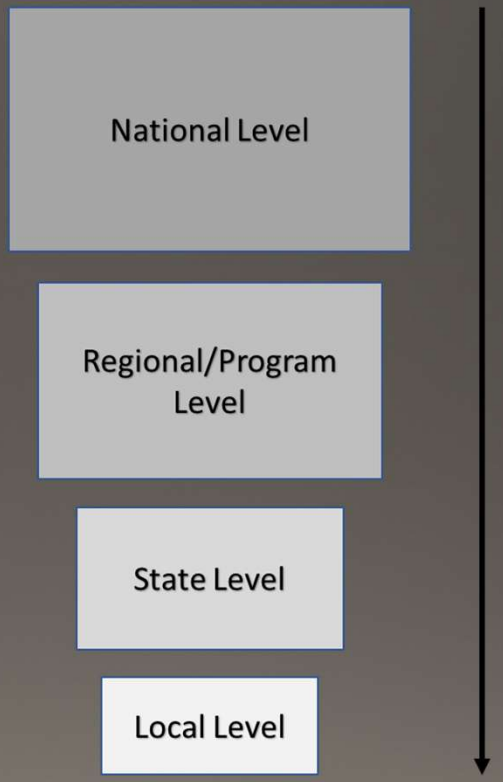
# Crawl:
# My Philosophy




Working Inside Out

And

Top Down



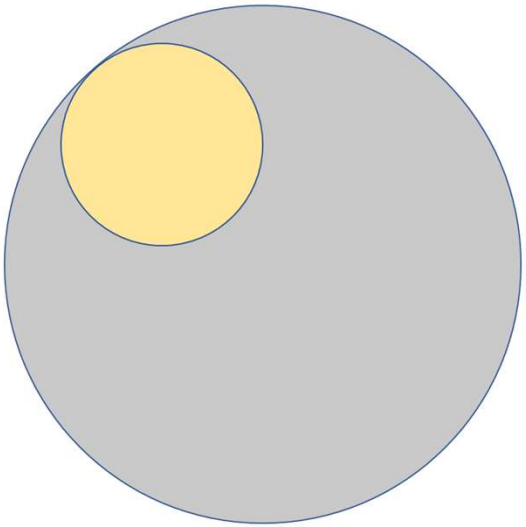
Crawl:

Why it Matters – in Law Enforcement we see an increase in responsibility and a decrease in resources.

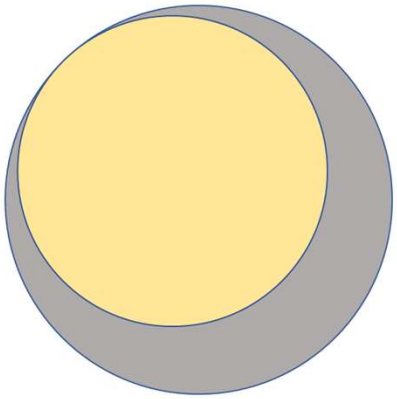
Additionally, the change in perception of LEOs and Oversight professionals makes it harder to get information in more traditional ways.

Analytics can fill gaps and bulwark your cases/proactive efforts.

Proportional Increase in *FUNDING* due to IIJA, IRA,& Chips Acts

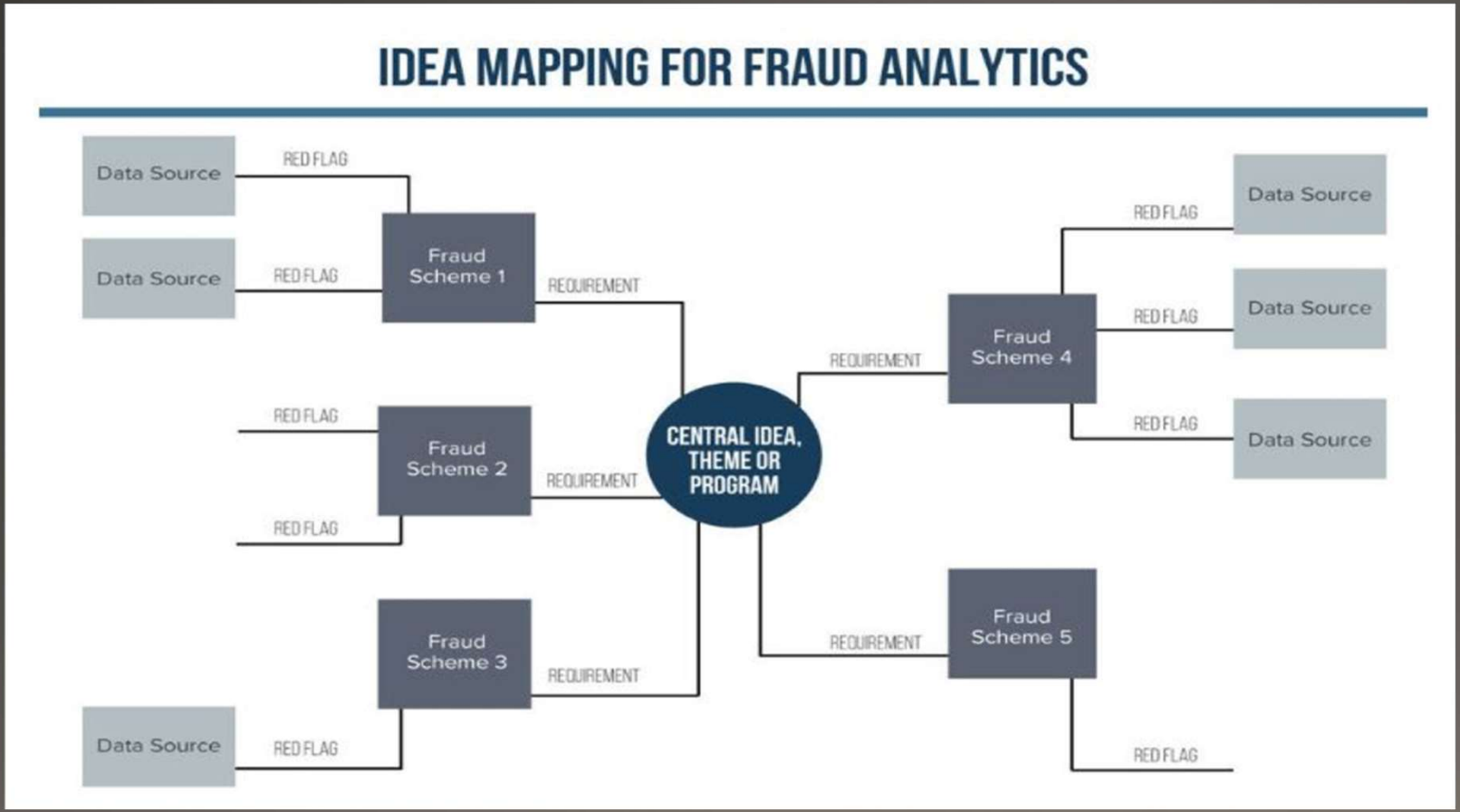


VS

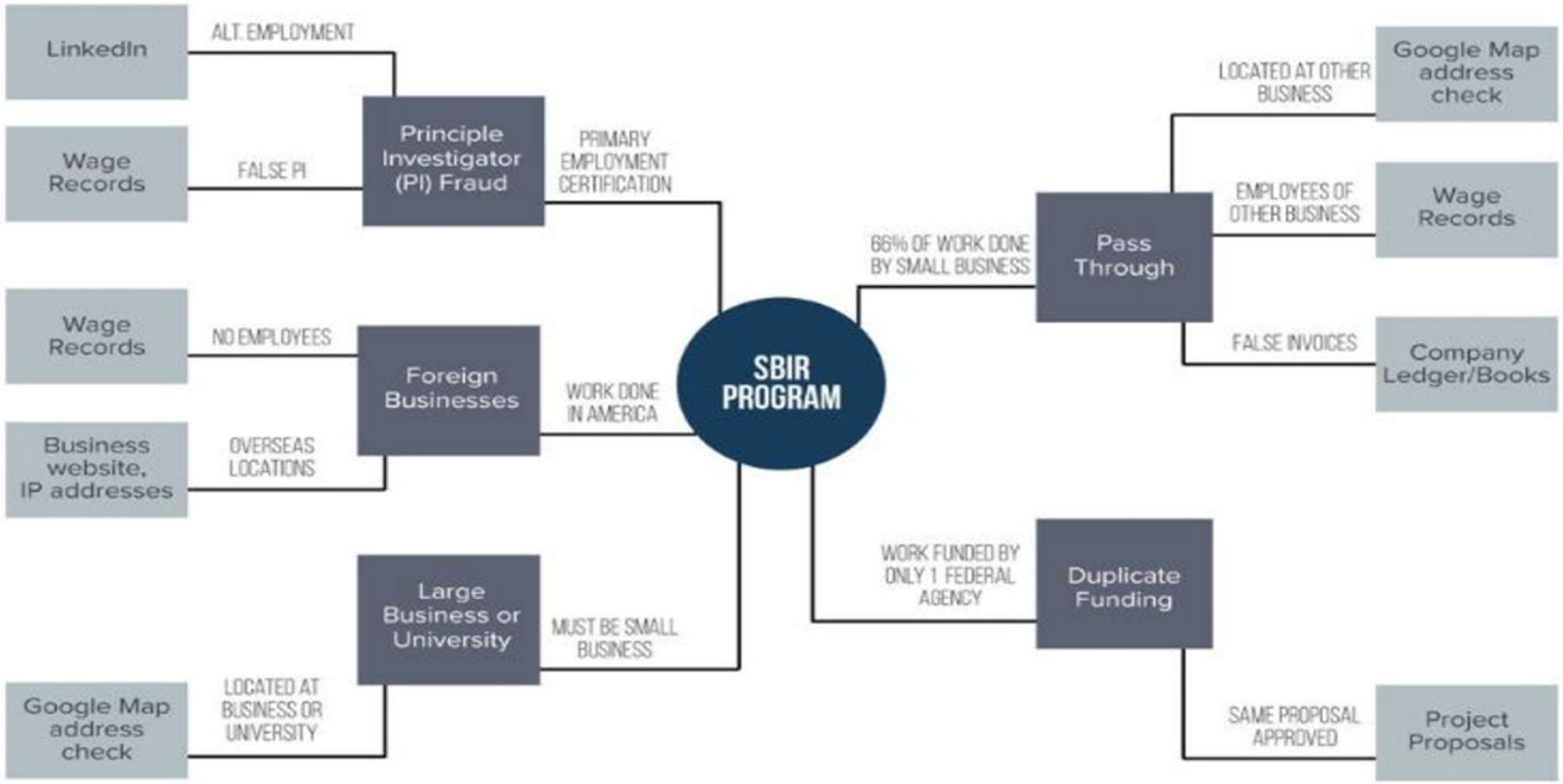


Proportional Increase in OI *MANPOWER* due to the New Funding Acts

# Crawl:
# Foundational Questions & Organization -



# EXAMPLE OF IDEA MAPPING FOR FRAUD ANALYTICS



# Crawl:

Descriptives – Use data to move beyond our [gut] instincts:

- Central Tendency
- Variability
- Frequency

Questions these basic statistics can answer:

1. How much do we spend on X.
2. How often are purchases made?
3. What does an average company look like?

All of these allow us to remove our assumptions and work to identify outliers.

# Crawl:

All of these basic things can be done in excel.

For central tendency use the =average and =median calculations

For variability use the =stdev.s calculation

For frequency use the =min, =max, and =frequency calculations

For example, using SBIR data for Department of Energy over 10 years:

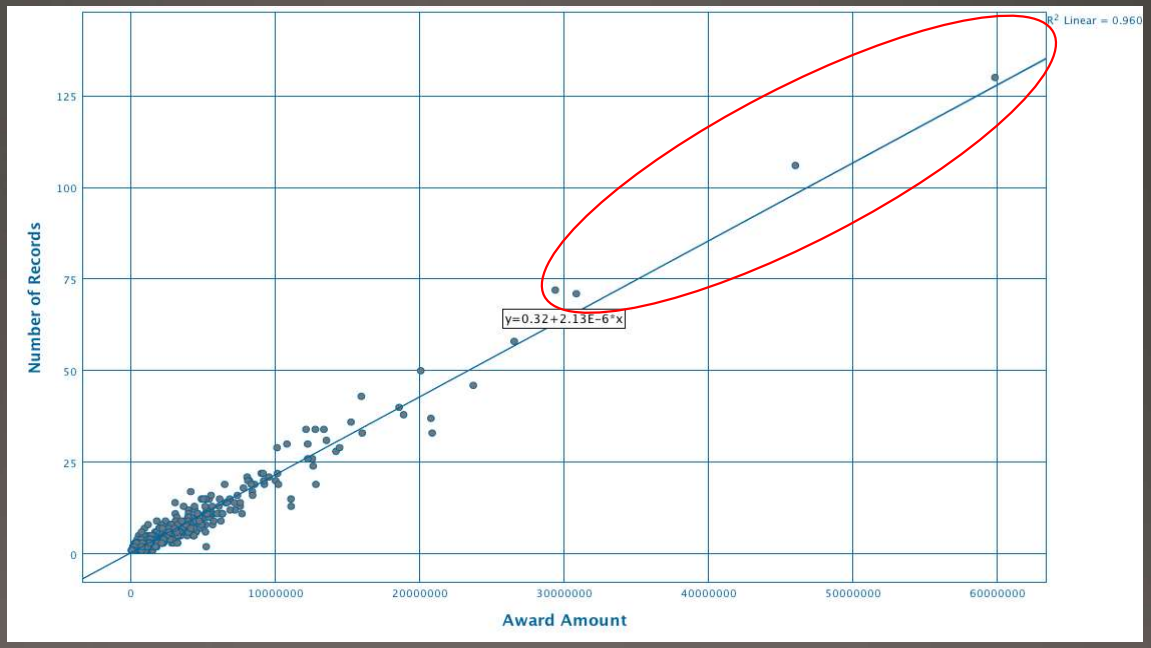
*We awarded 4,795 grants to 3,214 companies, totaling $2.06B. The average award amount was $430,508. The smallest award was $49k and the largest was $2.9m. The standard deviation of the awards was 422k.*

## Walk:

Employing (Meta)Data- When we know what's normal, we can start to see what's not.

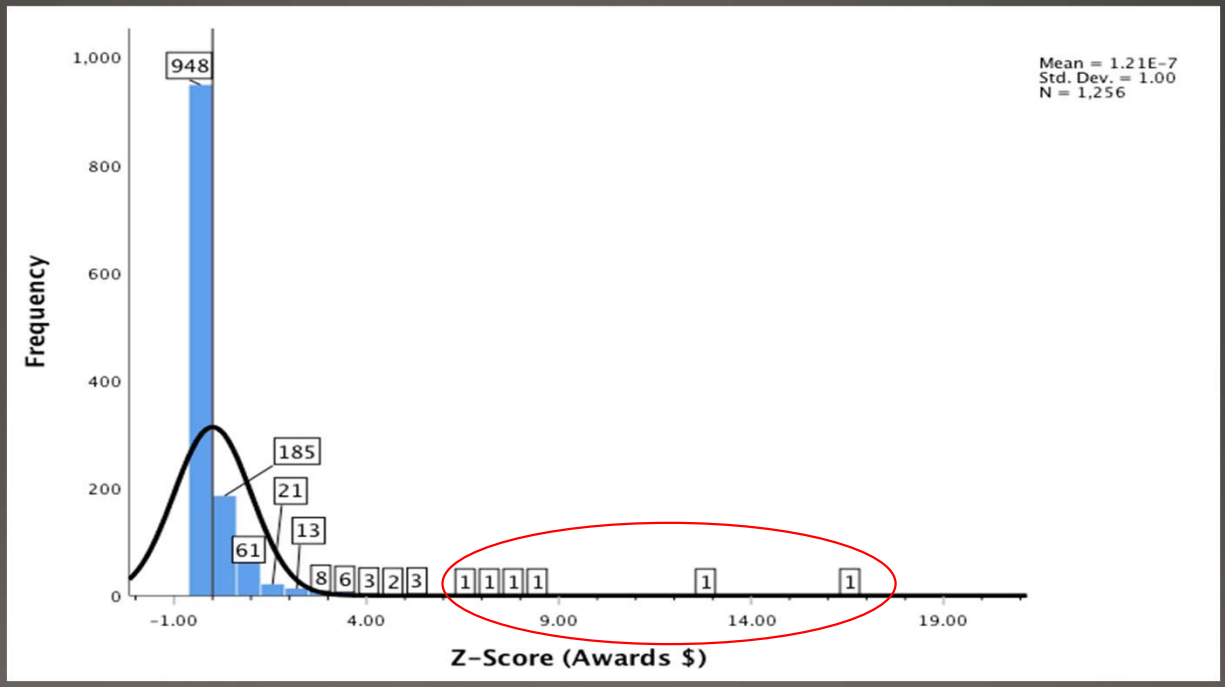
### Regression Analysis:

- Plots your two variables into standard scatter plot.
- Adds a line of best fit that equally splits the data points and allows an estimate.
- Visually shows potential outliers.

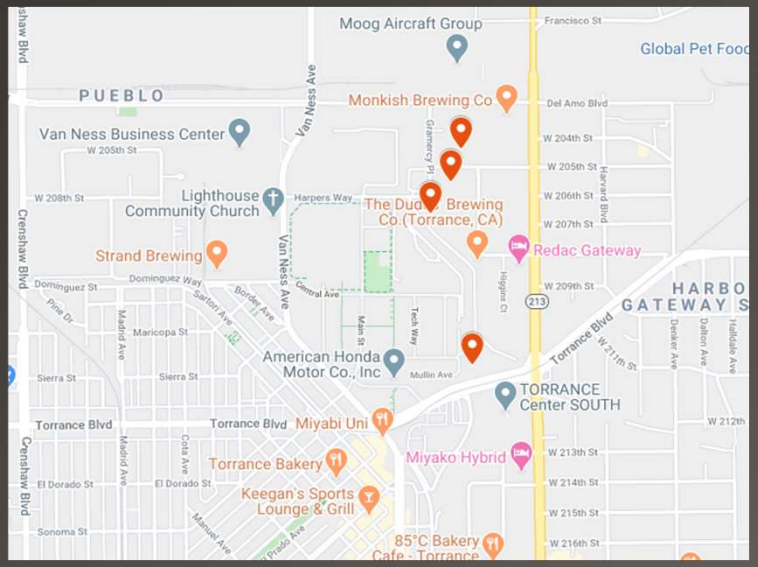


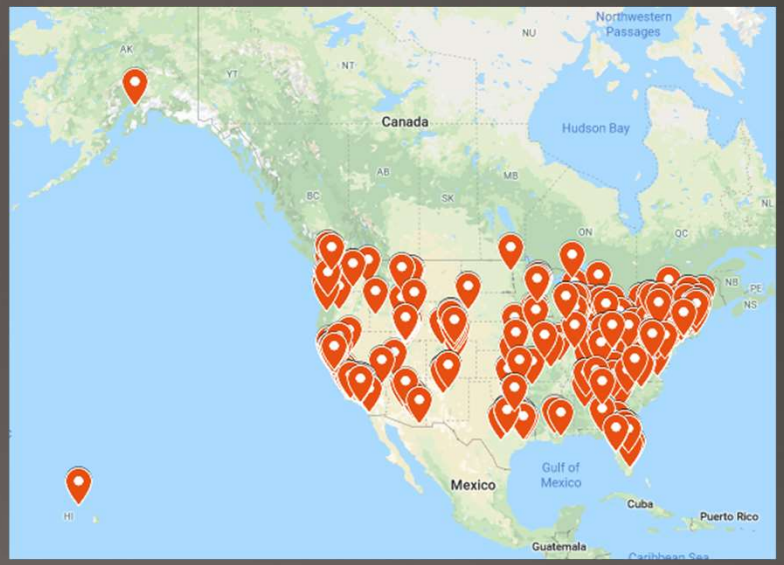
# Walk:
# Employing (Meta)Data- Z-Score Analysis:

- Converts variables to a scaled value. If >3.0 mathematically is an outlier.
- Can be done in excel using individual value, average value (mean), and standard deviation.
- To calculate in excel use the following formula =(value-mean)/standard deviation.

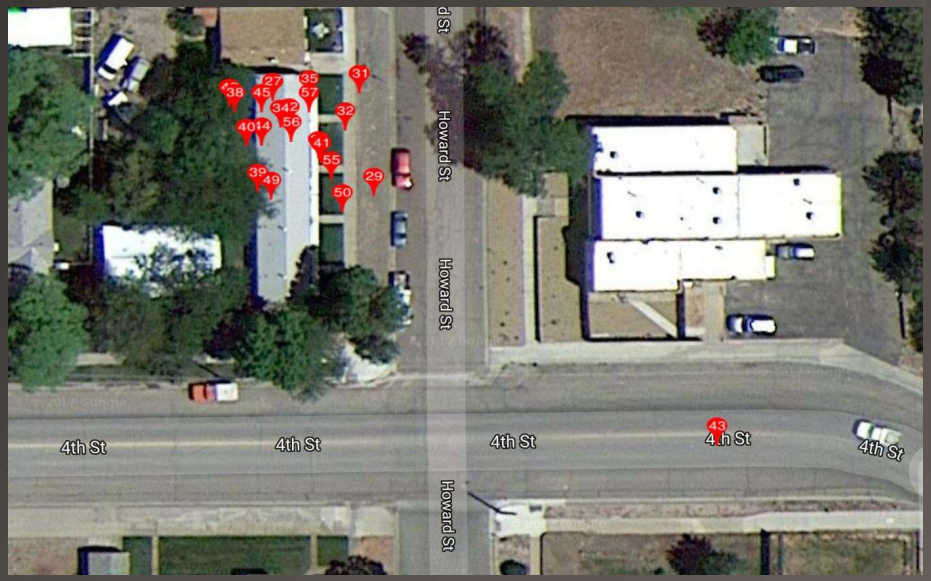


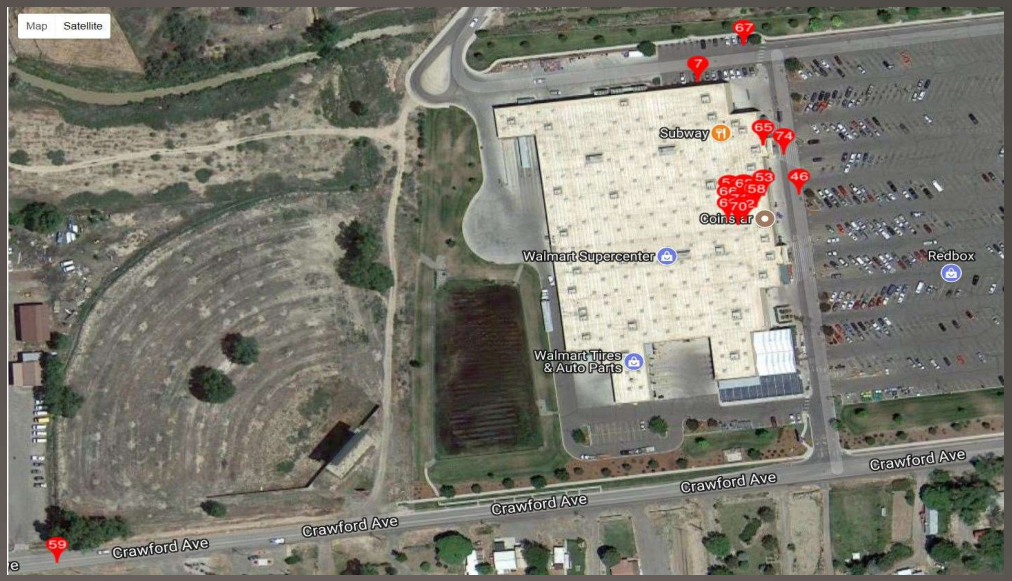
# Walk:
# Employee (Meta)Data- Mapping locally (vs. Nationally)



# Walk:
# Employing (Meta)Data- Mapping Employee Information



# Walk:
# Employing (Meta)Data- Ripping IP Addresses from Emails

Received: from mailgate.doe.gov (unknown [**205.254.128.11**]) by vsmtpx-e05-02.localdomain (Postfix) with ESMTP id 875AF20097 for <XXXXXX@science.doe.gov>; Wed, 5 Feb 2020 08:52:45 +0000 (GMT)

Received: from mail-pg1-x532.google.com (mail-pg1-x532.google.com [IPv6:2607:f8b0:4864:20::532]) (using TLSv1.2 with cipher ECDHE-RSA-AES128-GCM-SHA256 (128/128 bits)) (No client certificate

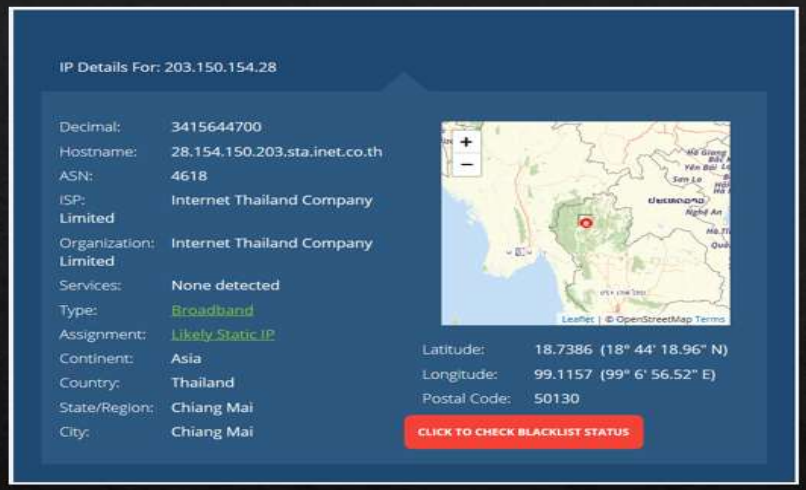
requested) by mailgate.doe.gov (Postfix) with ESMTPS id 2D8C0F1B062 for

<xxxxxxx@science.doe.gov>; Wed, 5 Feb 2020 03:52:37 -0500 (EST)

Received: by mail-pg1-x532.google.com with SMTP id z124so610869pgb.13 for

<xxxxxxx@science.doe.gov>; Wed, 05 Feb 2020 00:52:37 -0800 (PST)

Received: from [TARGET COMPUTER NAME] (28.154.150.203.sta.inet.co.th. [203.150.154.28]) by smtp.gmail.com with ESMTPSA id k3sm27080502pgc.3.2020.02.05.00.52.31 for <xxxxx@science.doe.gov> (version=TLS1_2cipher=ECDHE-RSA-AES128-SHA bits=128/128); Wed, 05 Feb 2020 00:52:33 -0800 (PST)




Target's Computer Name/Profile

The ISP

The IP Address

# Jogging? Corruption Ratios:

Used to detect potential bid rigging.
Created in Swiss economists David Imhoff and Yavuz Karagok.
Has worked in Switzerland, Japan, and now being used in America.
Pair with mapping to look for groups of collaborators in specific regions.

Variance Screen Test = CVj = σj/μj

Standard Deviation: 34764.1021

Mean (Average): 613,201

CVj = 0.056692823

| Company Name: | Bid Amount: | Difference: |
|---|---|---|
| A | $ 528,173.00 | 71,826 |
| B | $ 599,999.00 | 1,001 |
| C | $ 601,000.00 | 6,128 |
| D | $ 607,128.00 | 15,207 |
| E | $ 622,335.00 | 5,665 |
| F | $ 628,000.00 | 7,000 |
| G | $ 635,000.00 | 12,000 |
| H | $ 647,000.00 | 3,175 |
| I | $ 650,175.00 | -650,175 |


| Difference Between Two Lowest Bids: | 71,826 |
|---|---|
| Standard Deviation of Losing Bids: | 18518.17693 |
| RDj = | 3.878675546 |

| Company Name: | Bid Amount: | Difference: |
|---|---|---|
| A | $ 528,173.00 | 71,826 |
| B | $ 599,999.00 | 1,001 |
| C | $ 601,000.00 | 6,128 |
| D | $ 607,128.00 | 15,207 |
| E | $ 622,335.00 | 5,665 |
| F | $ 628,000.00 | 7,000 |
| G | $ 635,000.00 | 12,000 |
| H | $ 647,000.00 | 3,175 |
| I | $ 650,175.00 | -650,175 |

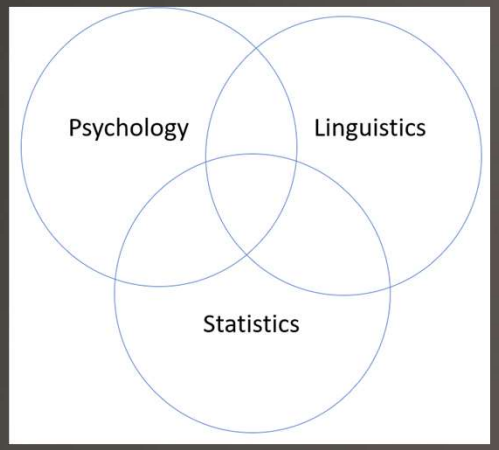
# Run – Modeling and Advanced Ideas

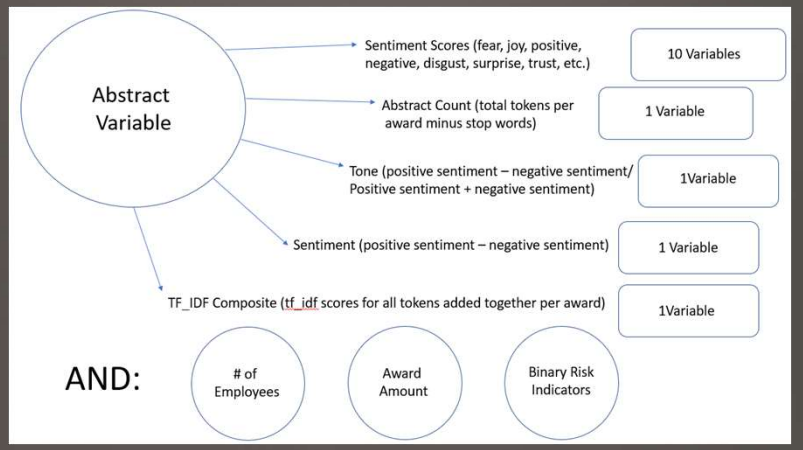
As we move beyond local or one-off type of analysis we need to start looking for data that can be used at scale.

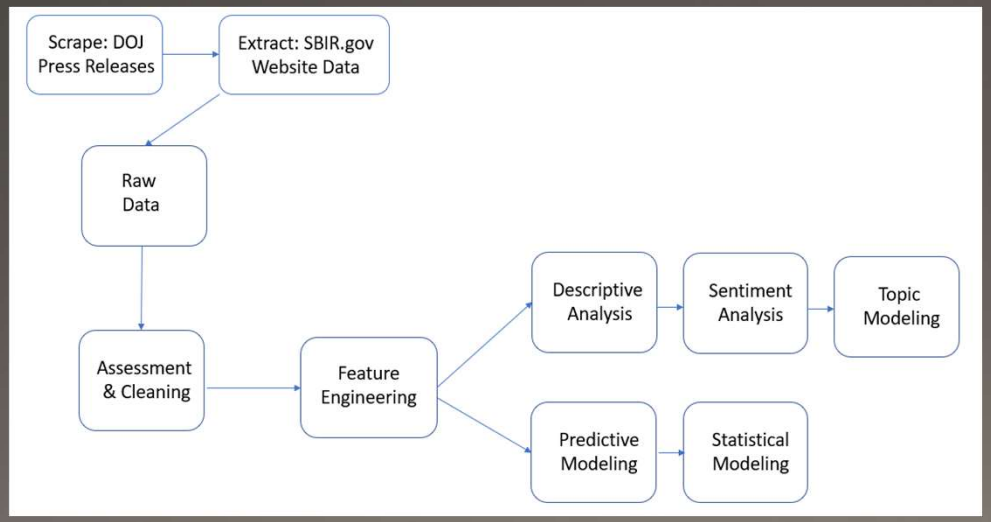
Using data creatively can help solve a lot of problems.

But, these techniques also need specialized knowledge and experience to build out.

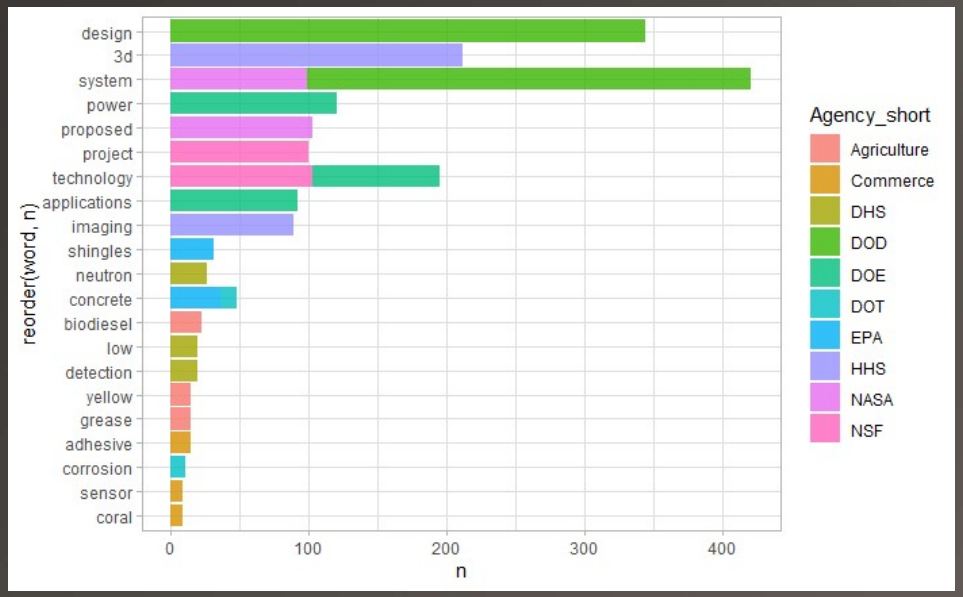
Lets look at an example.

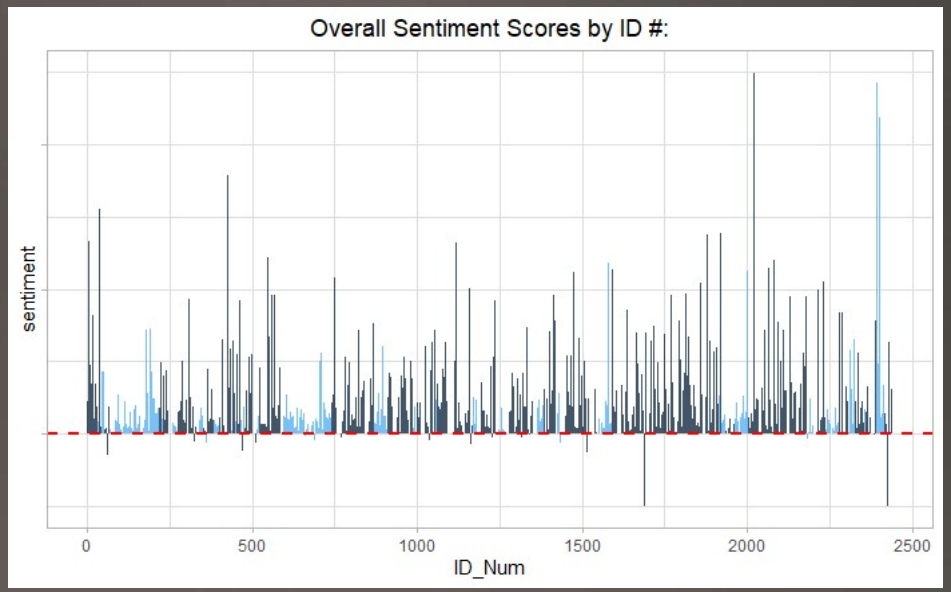


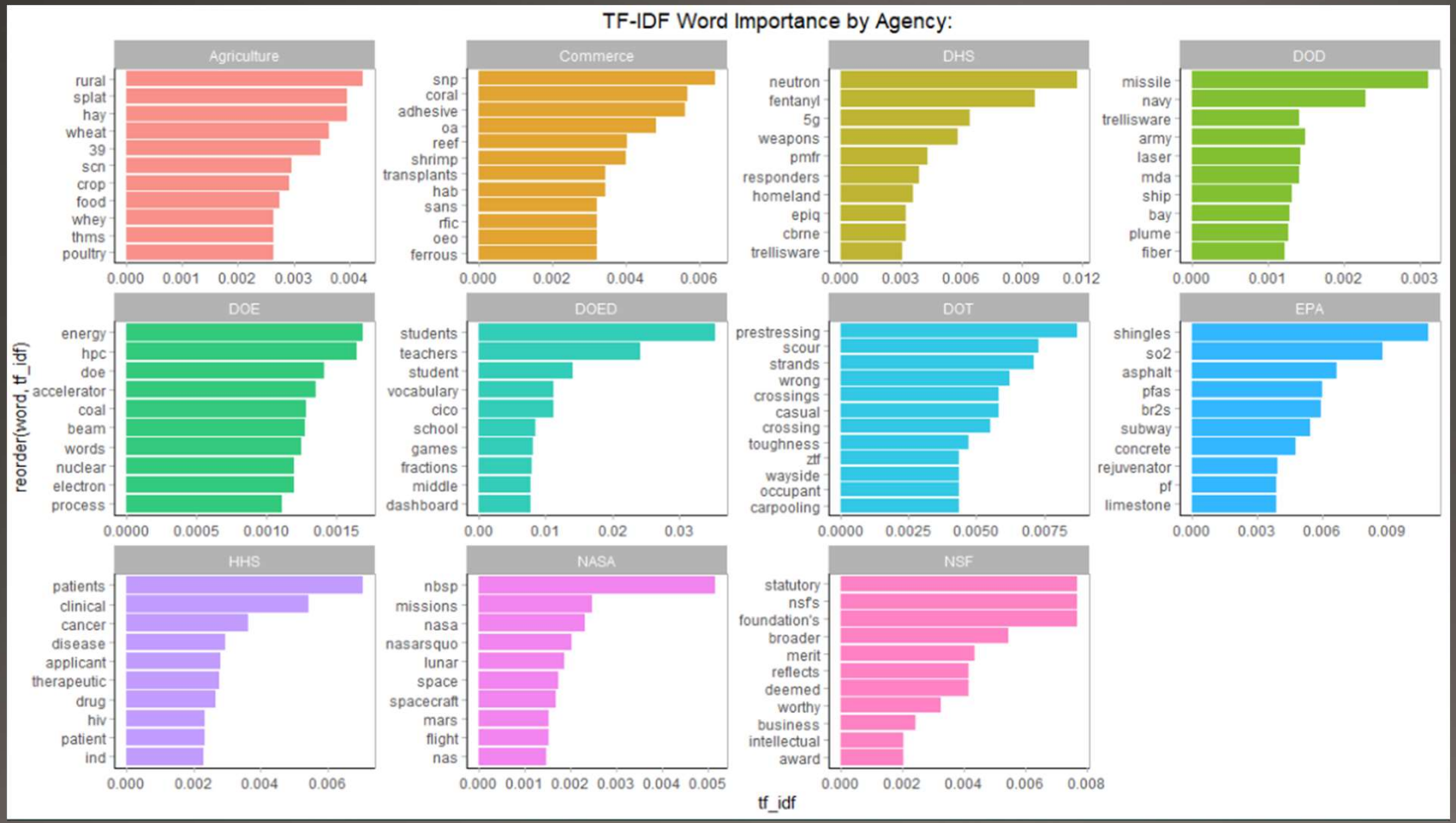



# Run – Modeling and Advanced Ideas

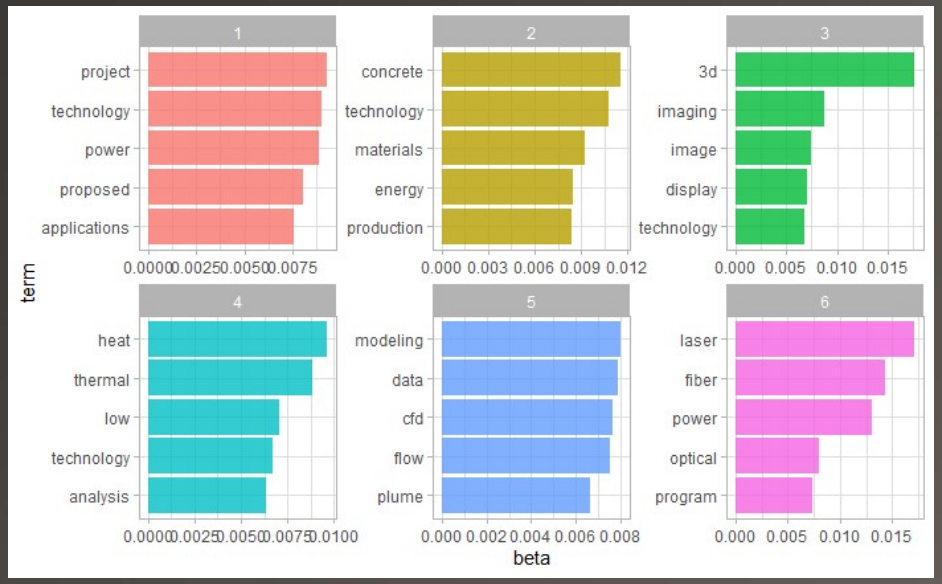


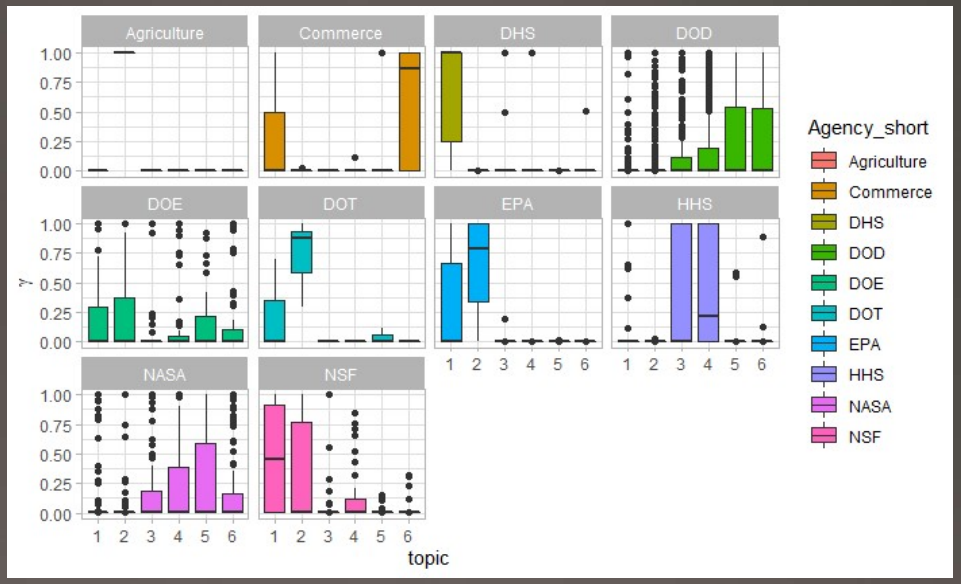



# Run – Modeling and Advanced Ideas

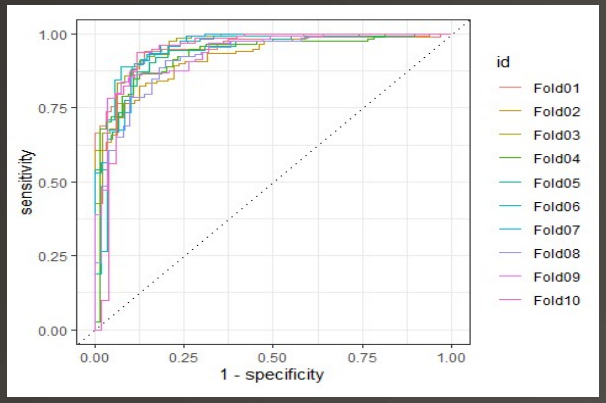


# Run – Modeling and Advanced Ideas



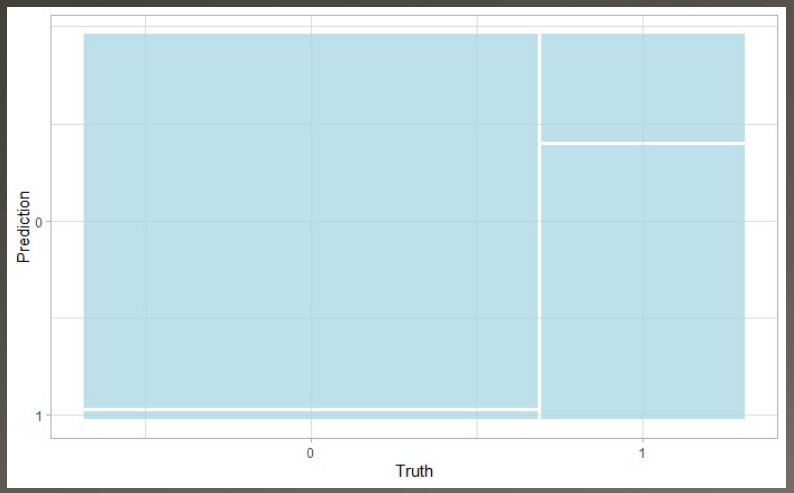
# Run – Modeling and Advanced Ideas

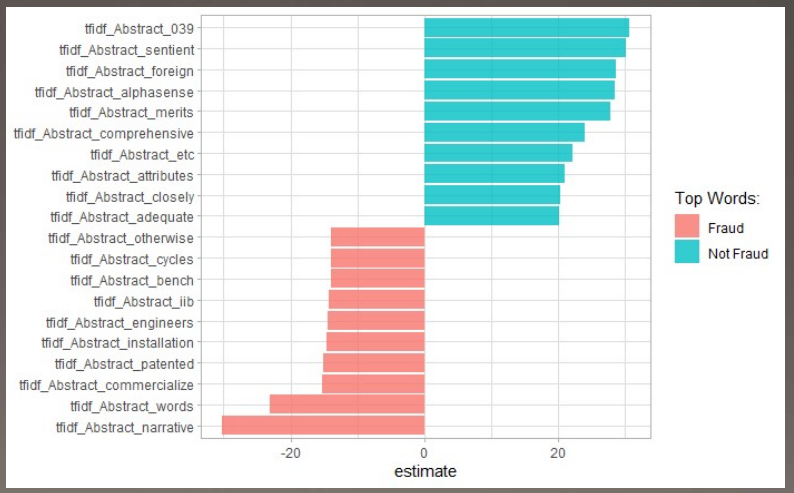



```
## # A tibble: 2 × 4
##   .metric  .estimator .estimate .config
##   <chr>    <chr>          <dbl> <chr>
## 1 accuracy binary         0.900 Preprocessor1_Model1
## 2 roc_auc  binary         0.961 Preprocessor1_Model1
```

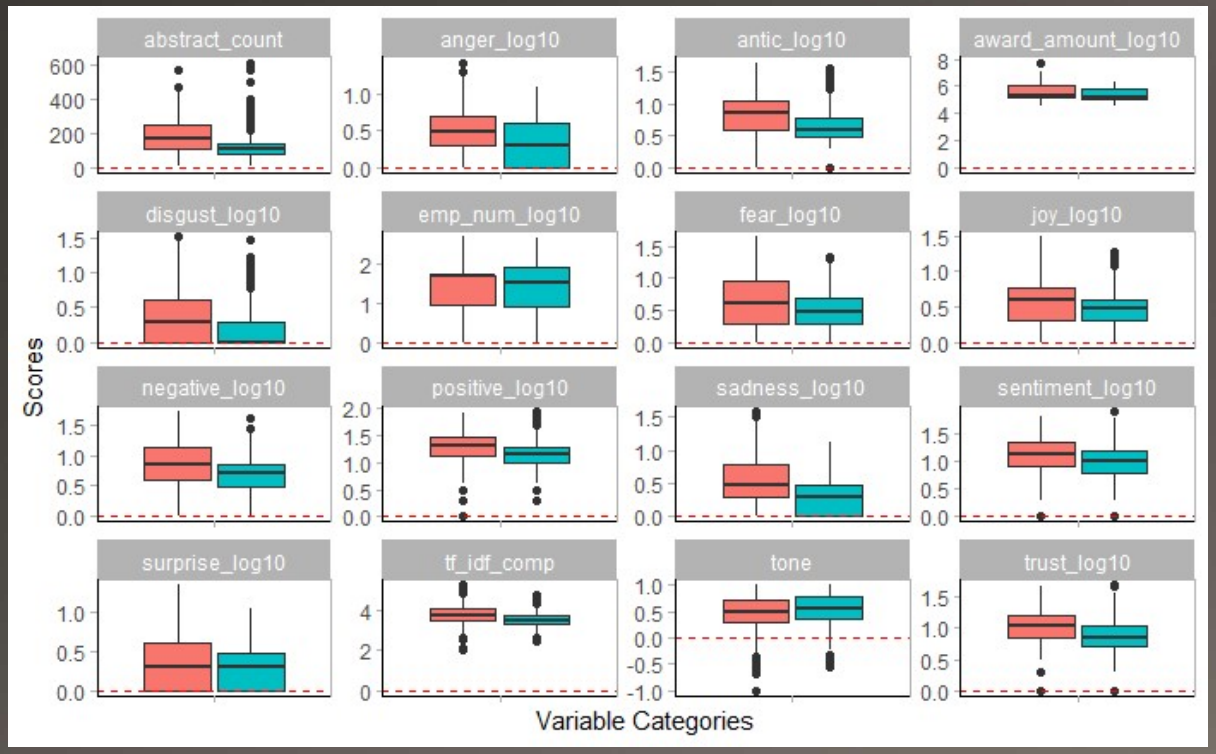
```
final_wf_elasticnet <- extract_workflow(final_fitted_elasticnet)
```

```
conf_mat_resampled(final_fitted_elasticnet, tidy = FALSE) %>%
  autoplot(type = "heatmap")
```



# Run – Modeling and Advanced Ideas




| | ci.upper | p.value |
|---|---|---|
| Abstract Count | 68.49022201 | 0.000000000 |
| Sentiment - Anger | 0.16487757 | 0.000000000 |
| Sentiment - Anticipation | 0.22930318 | 0.000000000 |
| Award Amount | 0.26533851 | 0.000000000 |
| Sentiment - Disgust | 0.20143444 | 0.000000000 |
| Employee Number | 0.05532811 | 0.844462972 |
| Sentiment - Fear | 0.20545817 | 0.000000000 |
| Sentiment - Joy | 0.21333964 | 0.000000000 |
| Sentiment - Negative | 0.23784917 | 0.000000000 |
| Sentiment - Positive | 0.18616591 | 0.000000000 |
| Sentiment - sadness | 0.24316141 | 0.000000000 |
| Sentiment - Sentiment | 0.14242356 | 0.000000000 |
| Sentiment - surprise | 0.12187229 | 0.000000000 |
| TF-IDF Composite | 0.30420608 | 0.000000000 |
| Tone | -0.03578976 | 0.000002614 |
| Sentiment - Trust | 0.18472990 | 0.000000000 |

# Conclusion:

Parting Shots:

1. Analytics require creative problem solving, just like investigations and oversight in general.
2. Analytics can help you both proactively and reactively.
3. Data is everywhere, leverage it to help you instead of letting it drown you.
4. Analytics are not the end all be all of investigations, we need to maintain proficiency in traditional oversight methods too. ***But…*** analytics can impact such a wide range of your work, day to day, that it may be one of the most helpful skills you develop.

# Questions?



(720) 474-8553


erik.Halvorson@doe.gov