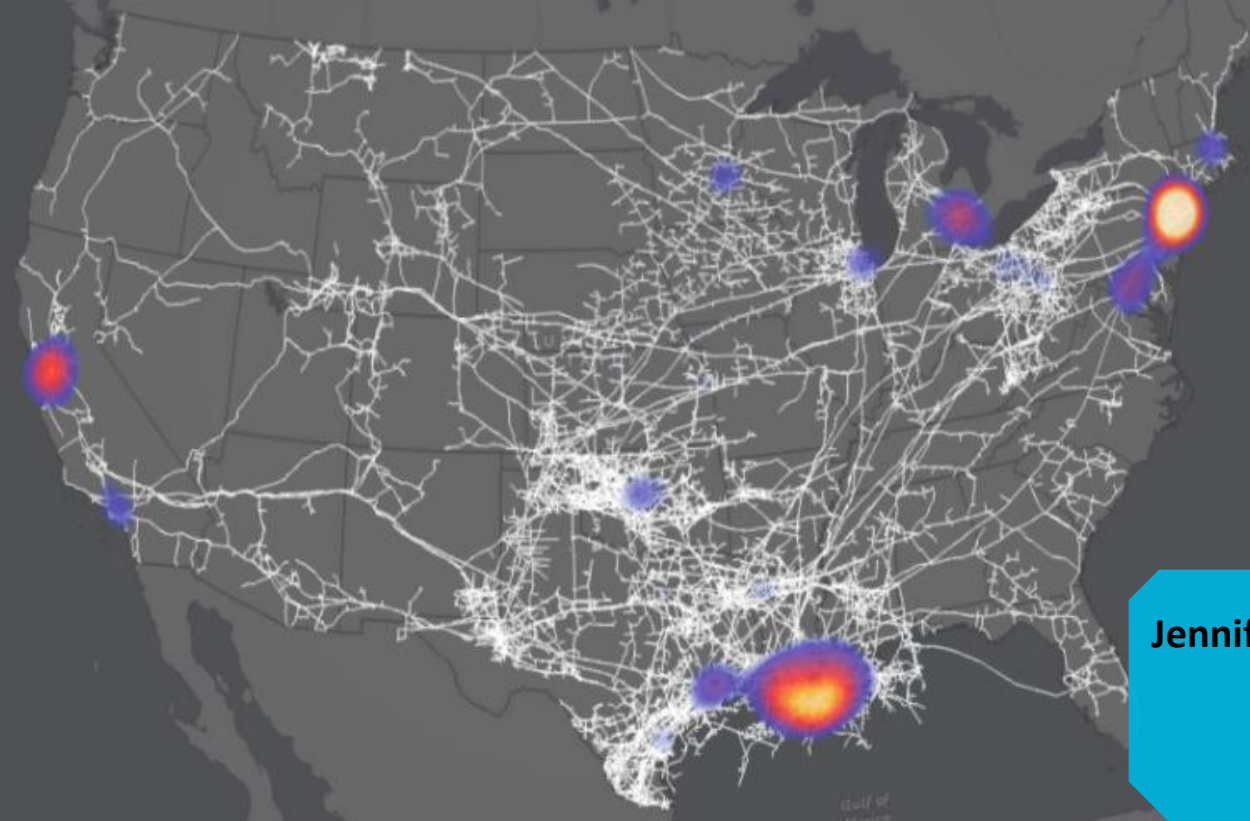
Geo-Machine Learning for Evaluating Natural Gas Pipeline Risk

NATIONAL ENERGY TECHNOLOGY LABORATORY



Jennifer Bauer, Devin Justman, Kelly Rose

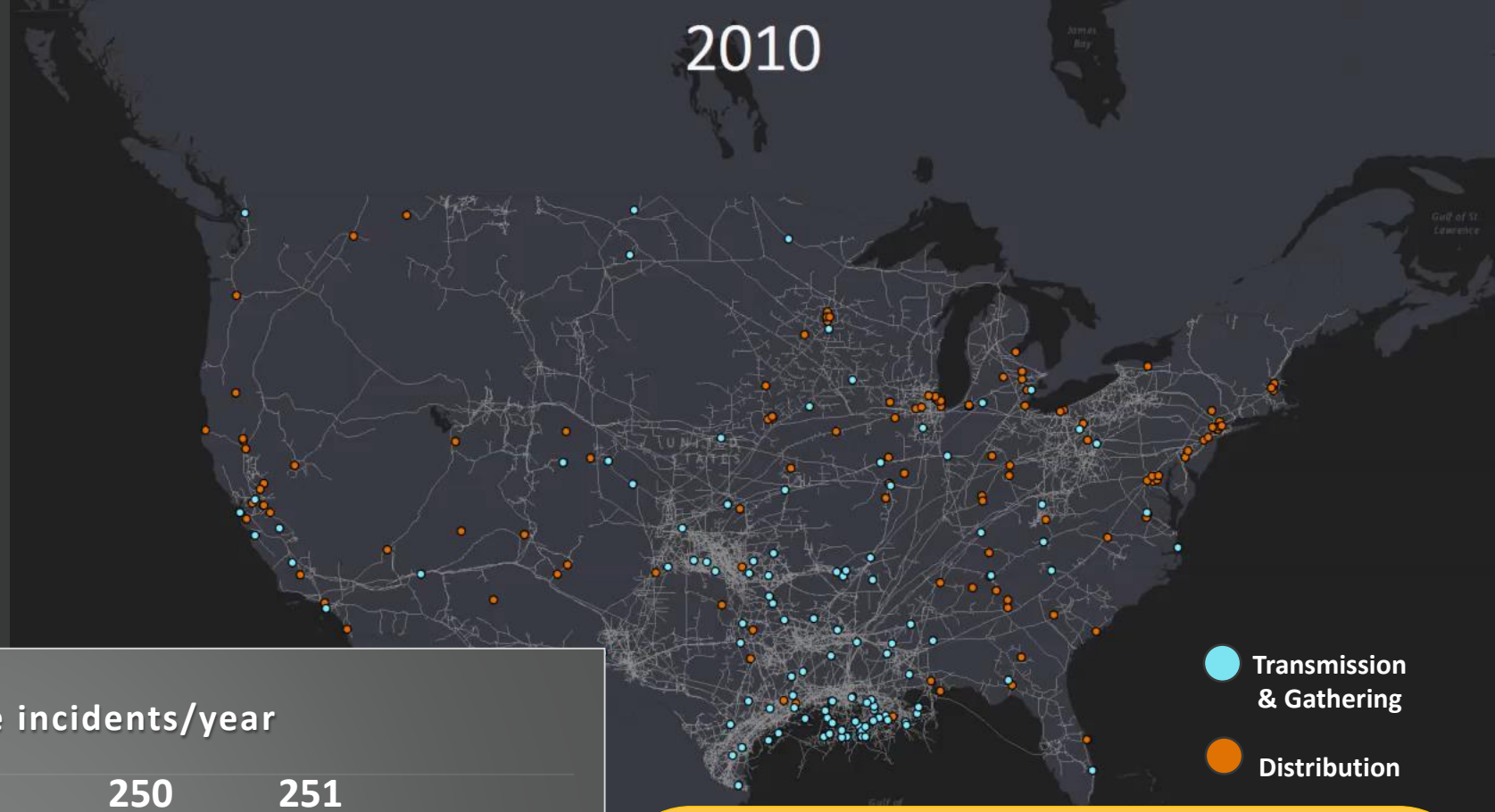
2019 Esri User Conference

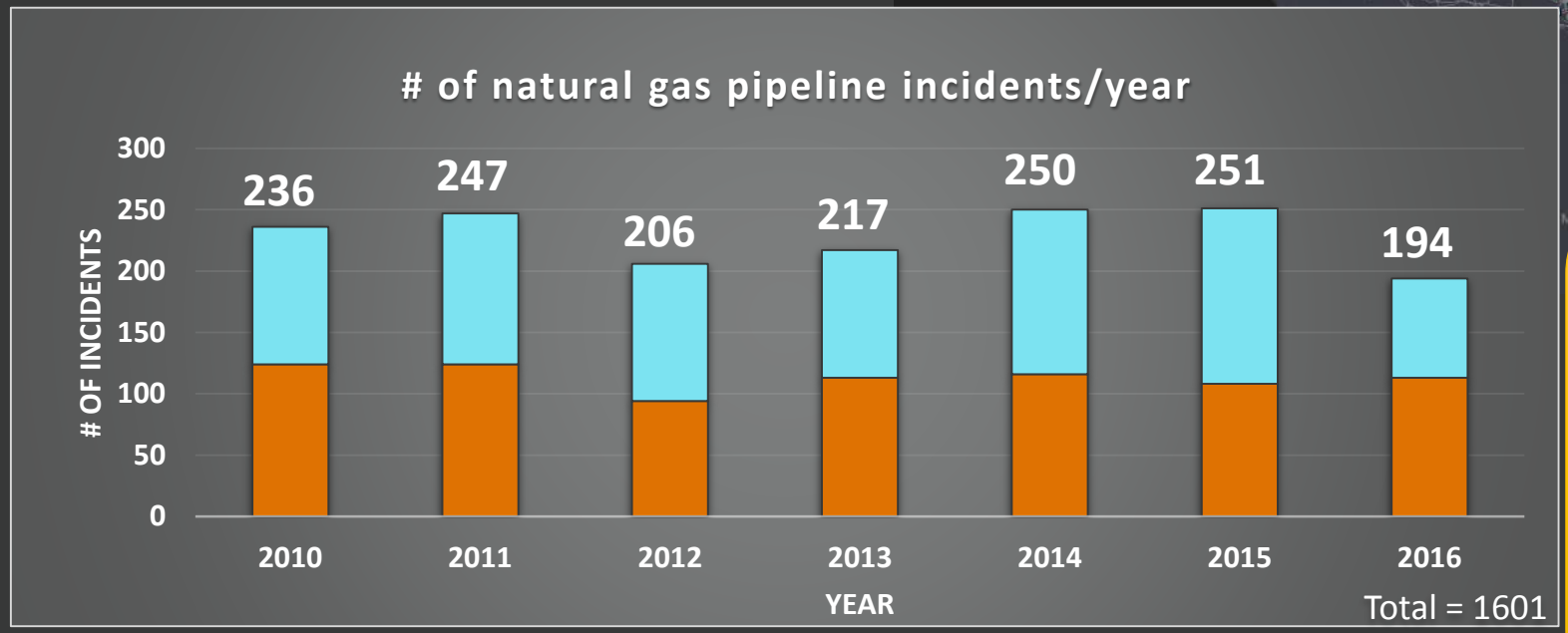
July 10, 2019

Solutions for Today | Options for Tomorrow

U.S. DEPARTMENT OF ENERGY

Across more than **2.5 million miles** of natural gas pipelines[1], an average of **240 incidents** occur each year[2]



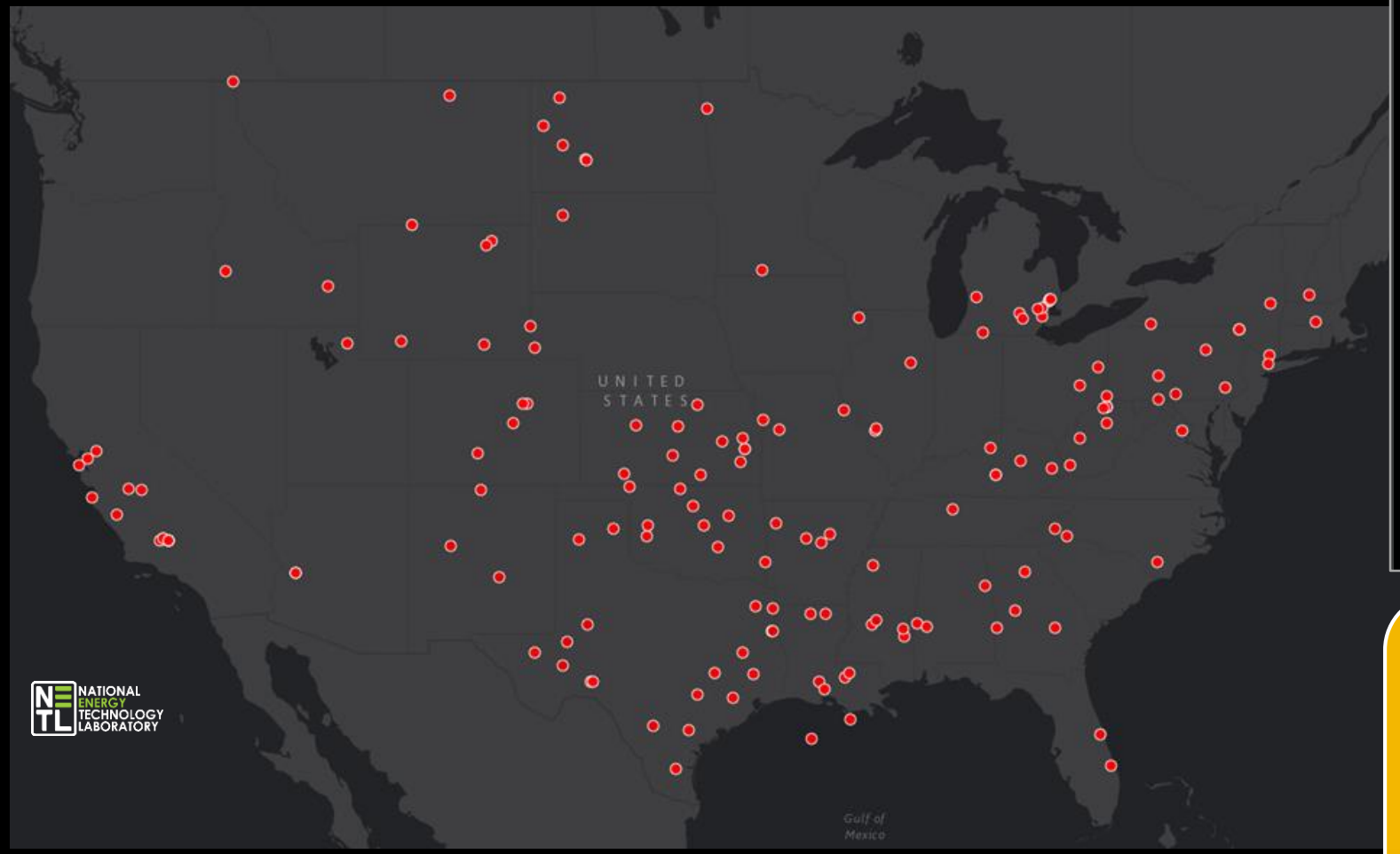


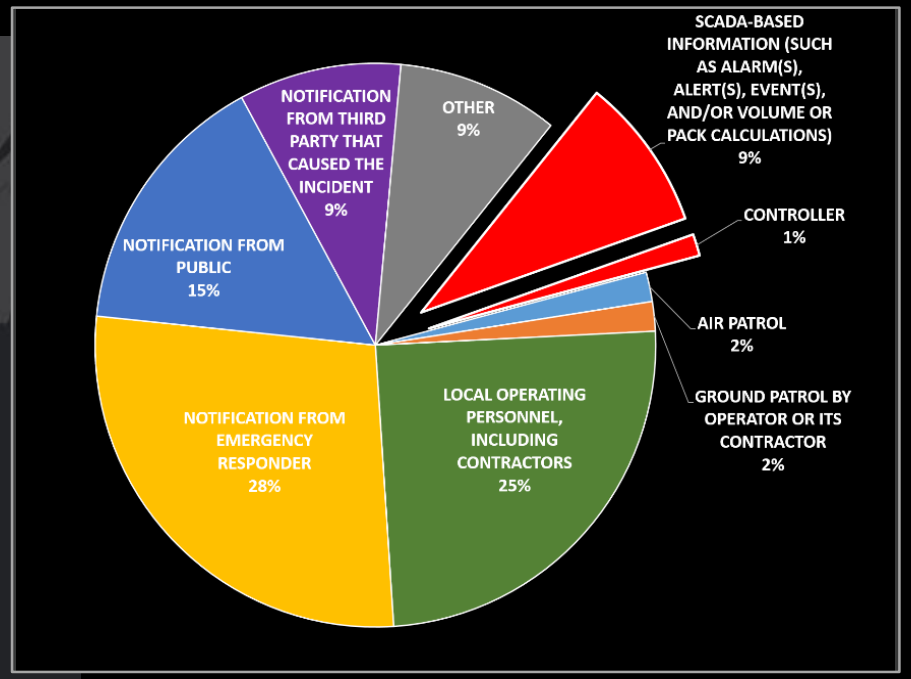
Operators have a **1 in 10,500 chance** of picking the right mile of pipeline to invest personnel and resources on inspections, monitoring, and other activities to prevent these failures.

U.S. DEPARTMENT OF ENERGY | NATIONAL ENERGY TECHNOLOGY LABORATORY

[1] EIA, 2018; [2] PHMSA, 2019; original data from Pipeline and Hazardous Materials Safety Administration (PHMSA) & US Energy Information Administration (EIA)

Need for technologies, methods, and analytics to inform sensor deployment and utilization, as well as reduce potential risks and hazards.

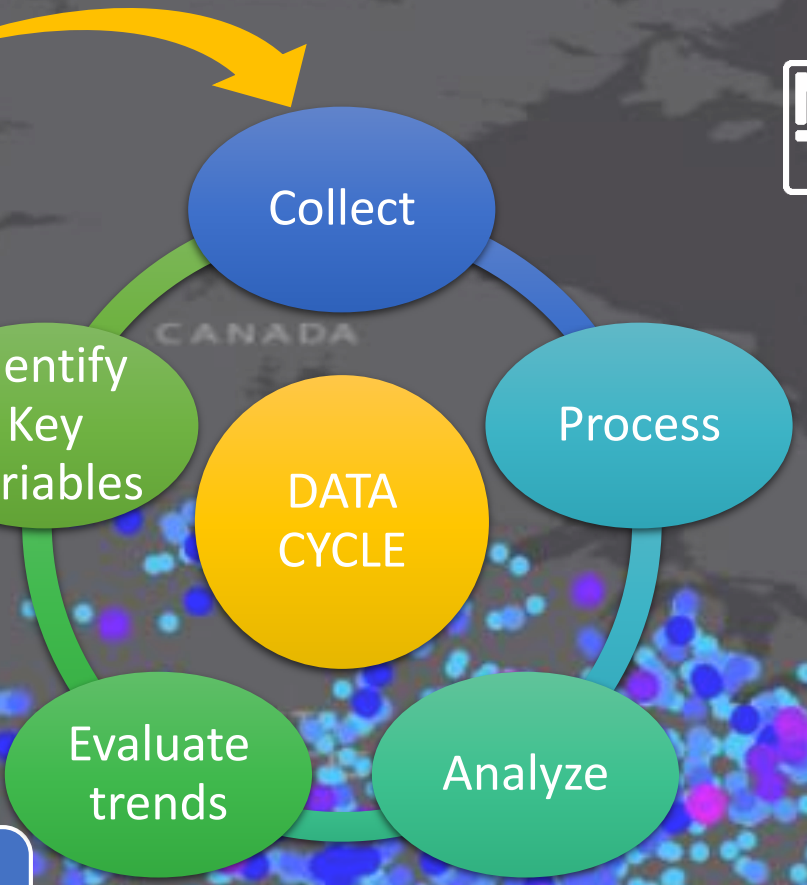




Integrating machine learning with spatio-temporal analytics offers the means to identify new insights

Original data sources: Pipeline and Hazardous Materials Safety Administration (PHMSA) & US Energy Information Administration (EIA)

Perform Geo-Data Analytics

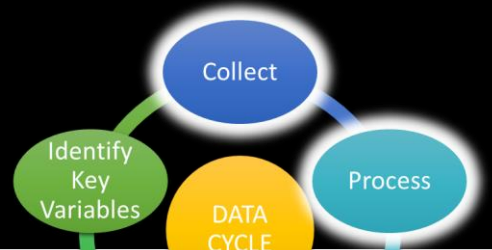


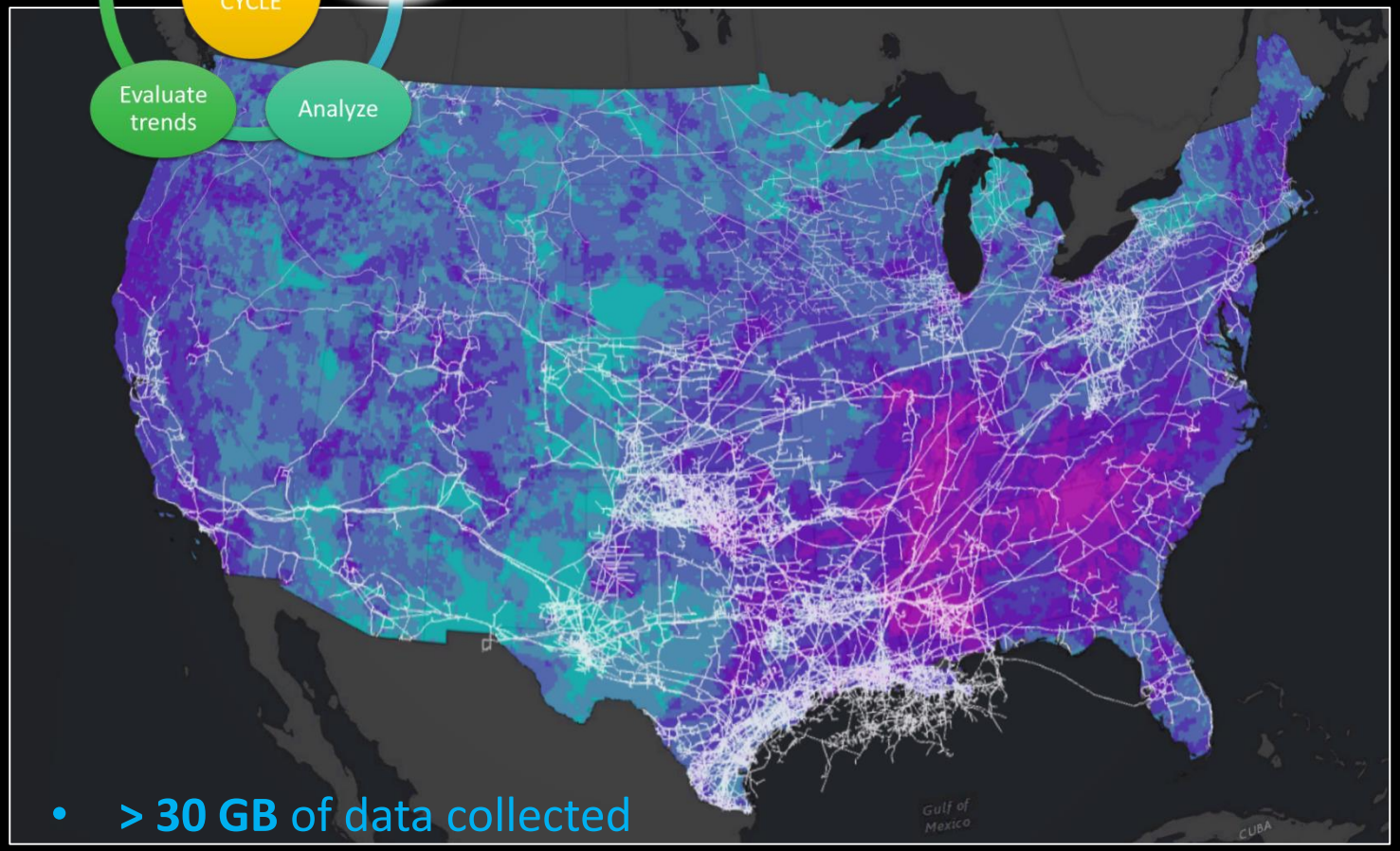
Inform sensor deployment and utilization

Reduce Risk & Impacts

- Quantify potential benefit of new technologies
- Optimize development & deployment
- Predict pipeline integrity risk
- Tools to support methane quantification

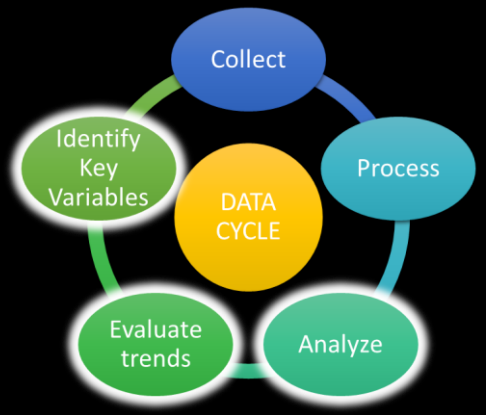
Data Collection & Processing





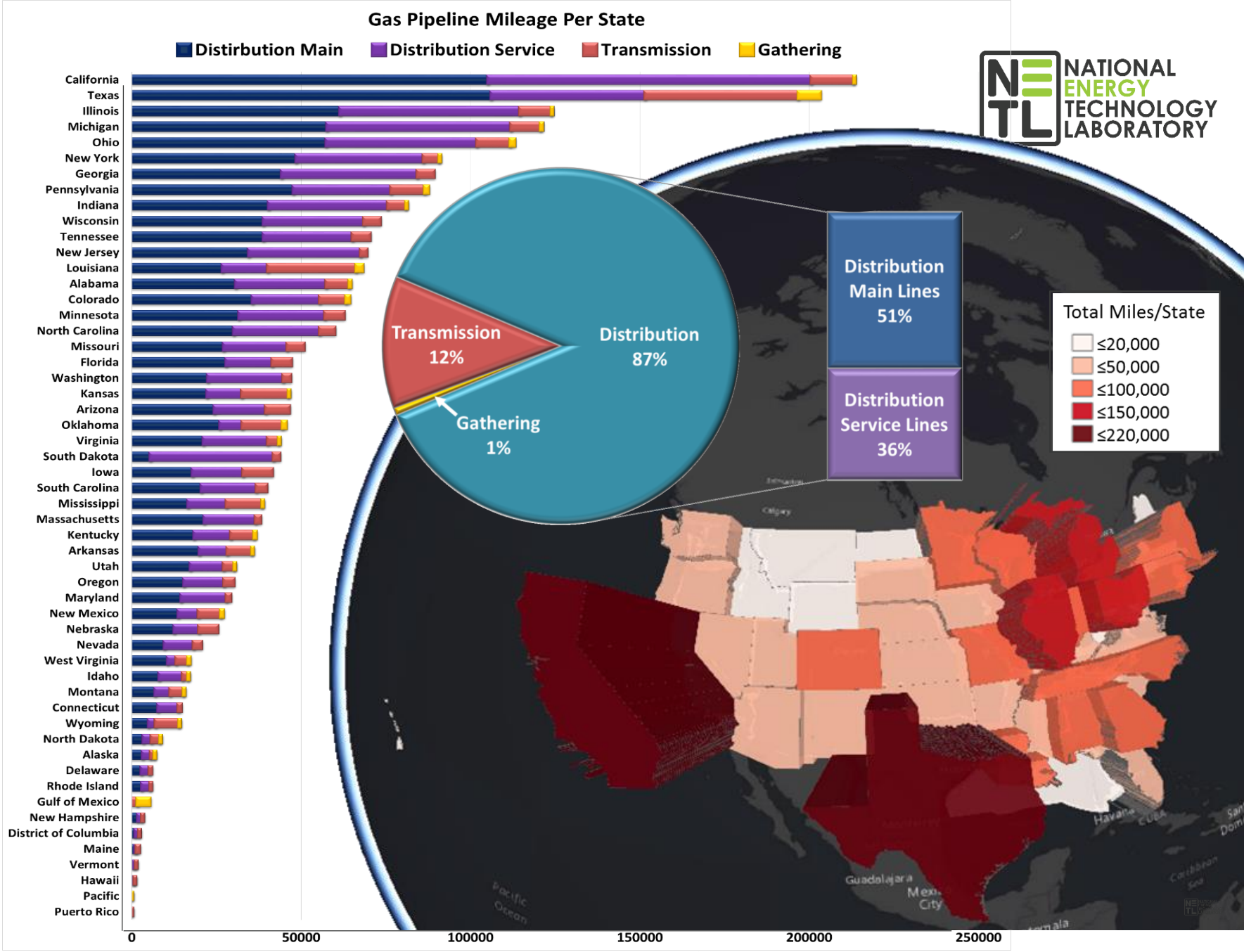
- **> 30 GB** of data collected
- Covers more than **45 years** of incidents
- Database consists of more than **200 different variables**

- Continue **adding to NETL databases** with the **most recent information** on the existing U.S. natural gas pipeline infrastructure, historic incidents, and proxy datasets
- Began **incorporating scripts** to **rectify data discrepancies, help process data**, and **automate database updating**
  - Evaluating range of tools that use natural language processing (NLP), fuzzy logic, etc. to help label and structure data from various sources and formats



Analyze, Evaluate & Identify

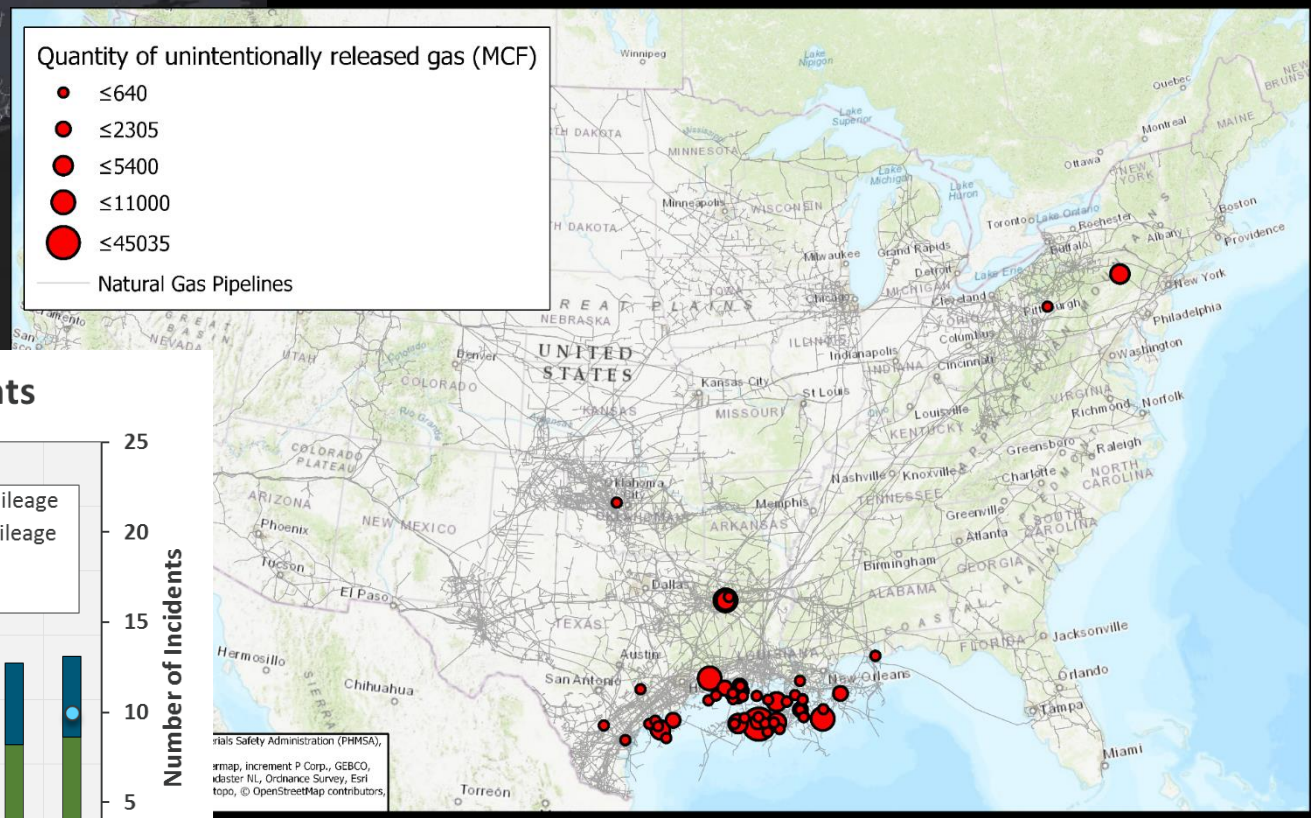
- Comb through data using exploratory analytics, as well as spatio-temporal analytics to understand the data and begin to assess relationships between data variables

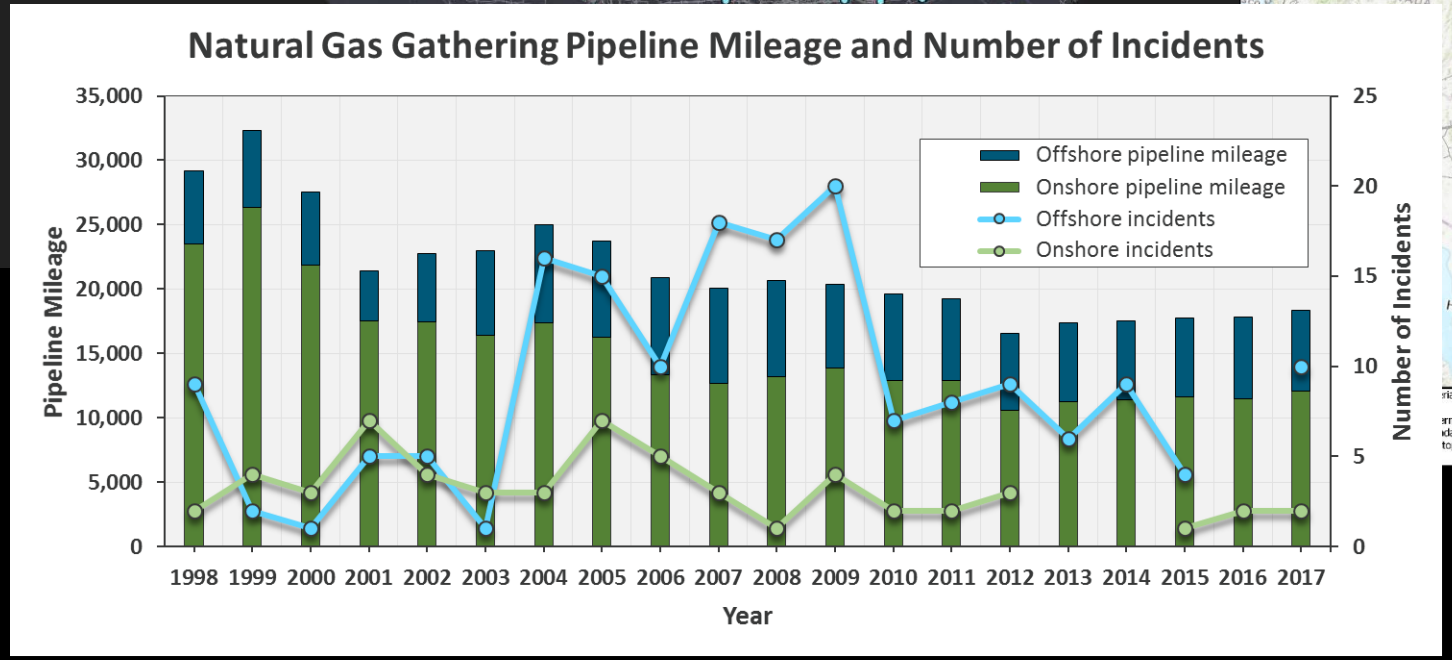


NATIONAL ENERGY TECHNOLOGY LABORATORY

[1] EIA, 2018; Source: Pipeline and Hazardous Materials Safety Administration (PHMSA)

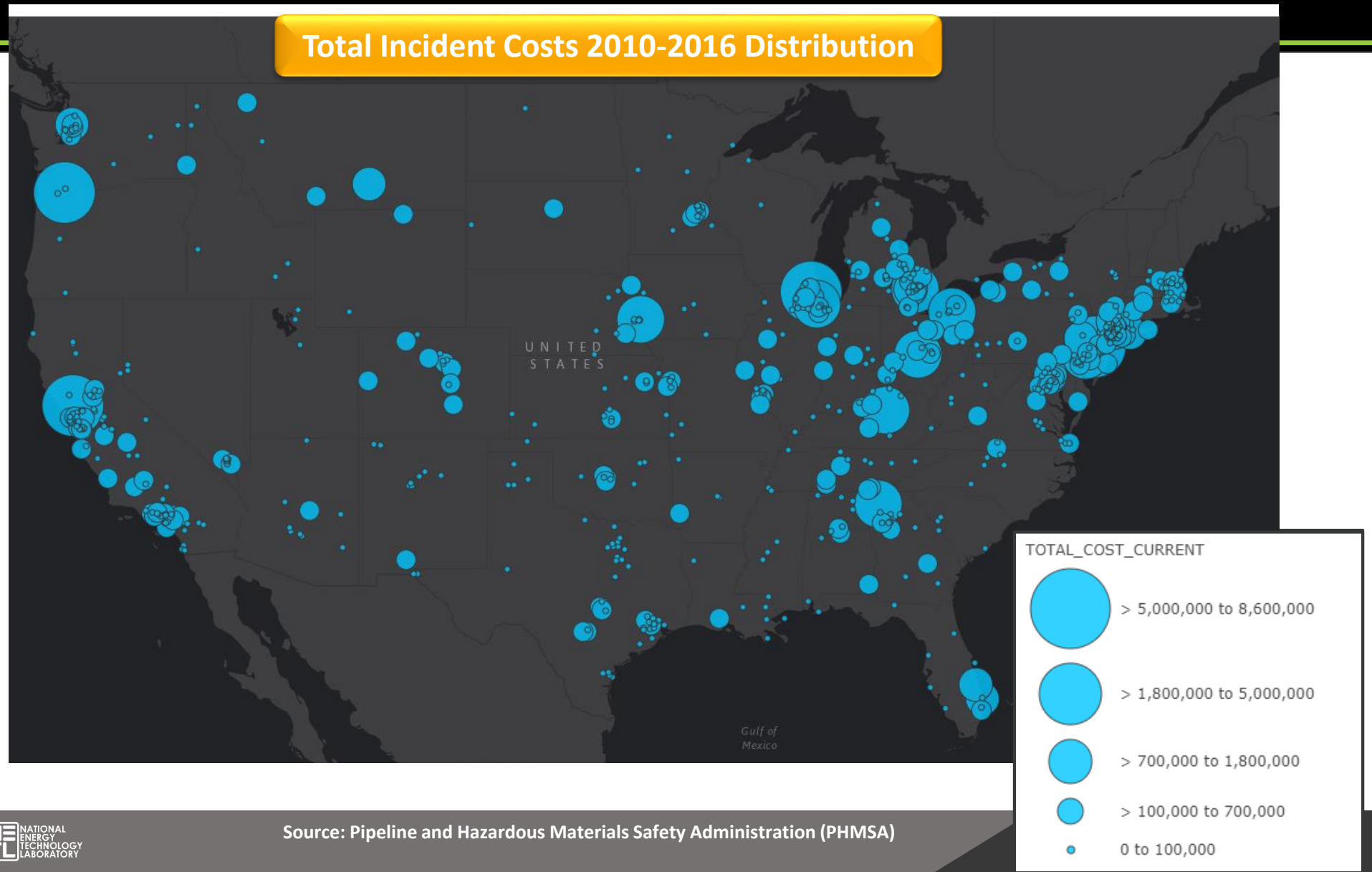
Pipeline Incidents: 2010





Data source: Pipeline and Hazardous Materials Safety Administration (PHMSA)

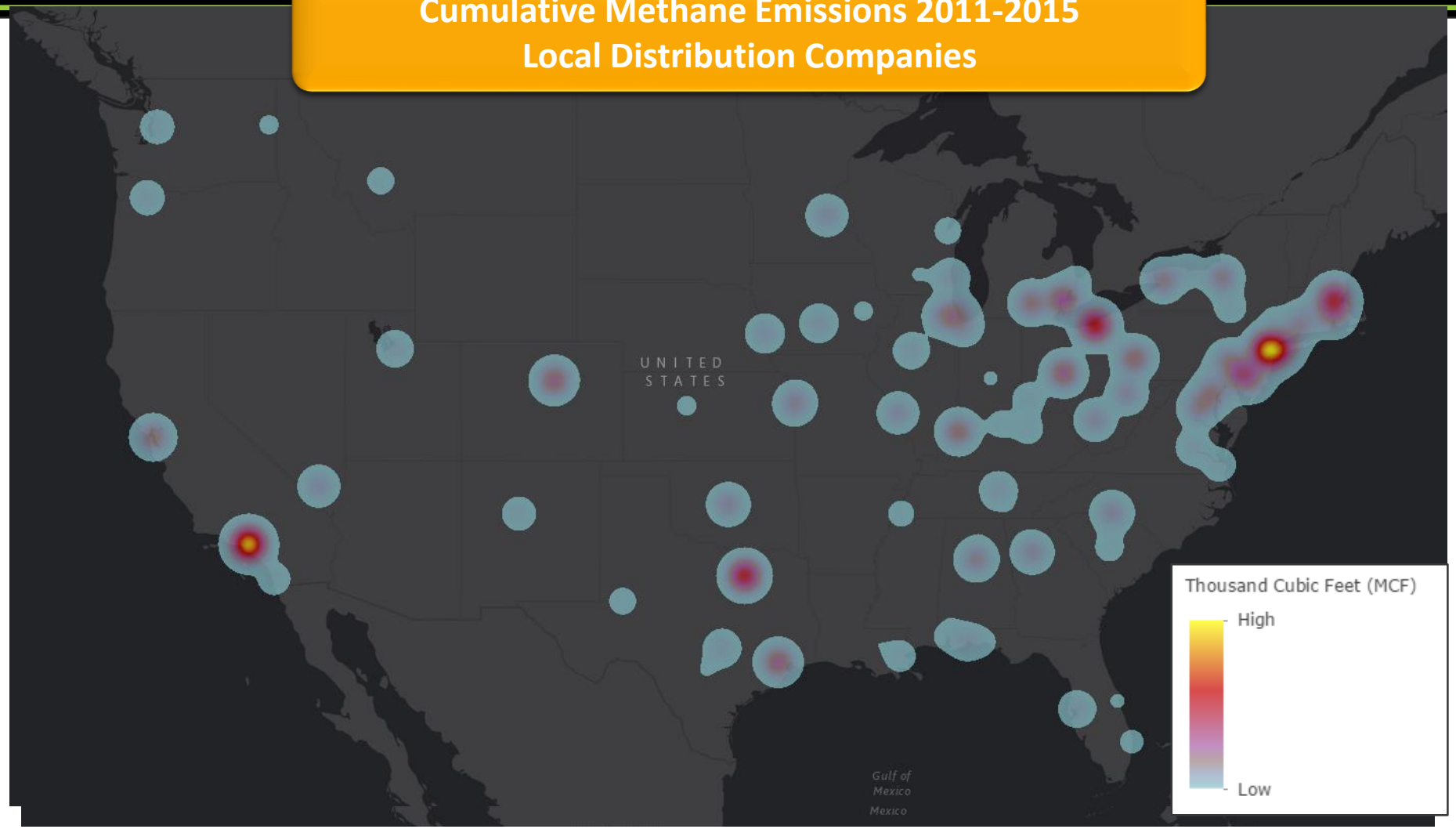
Identifying key spatio-temporal trends



Source: Pipeline and Hazardous Materials Safety Administration (PHMSA)

Identifying key spatio-temporal trends

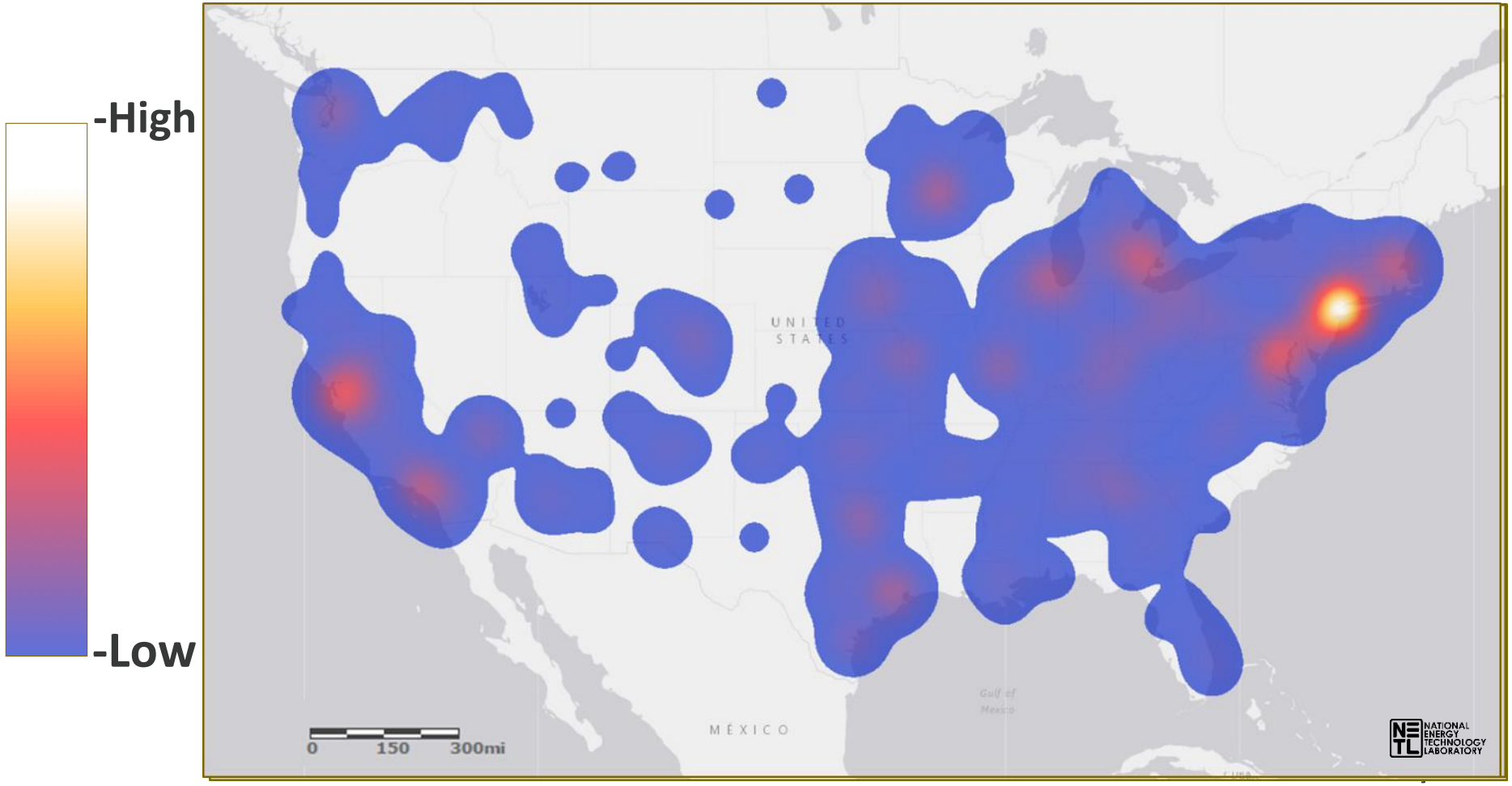
Cumulative Methane Emissions 2011-2015
Local Distribution Companies



Source: Environmental Protection Agency (EPA)

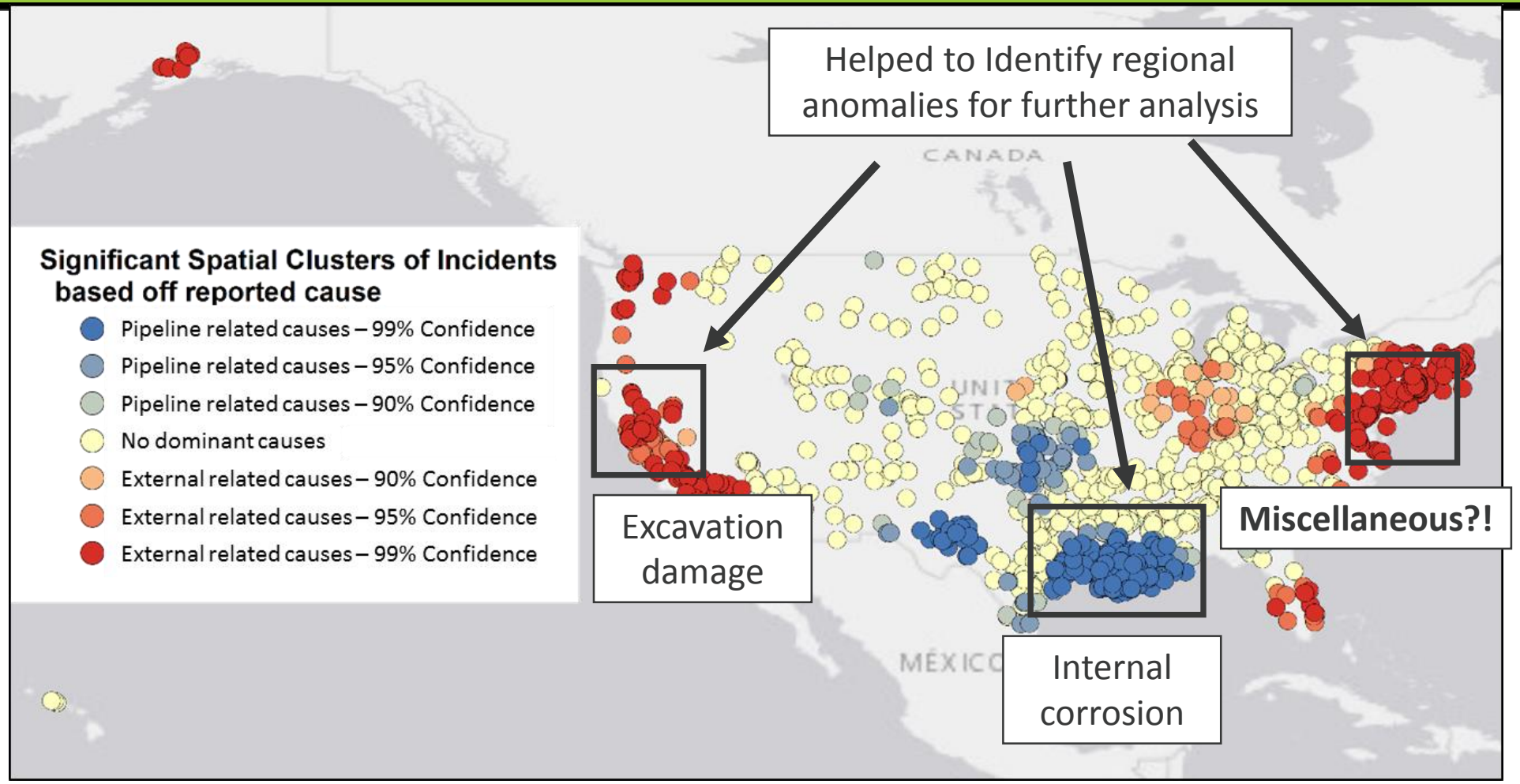
Identifying key spatio-temporal trends

Heat map for Gas Distribution incidents



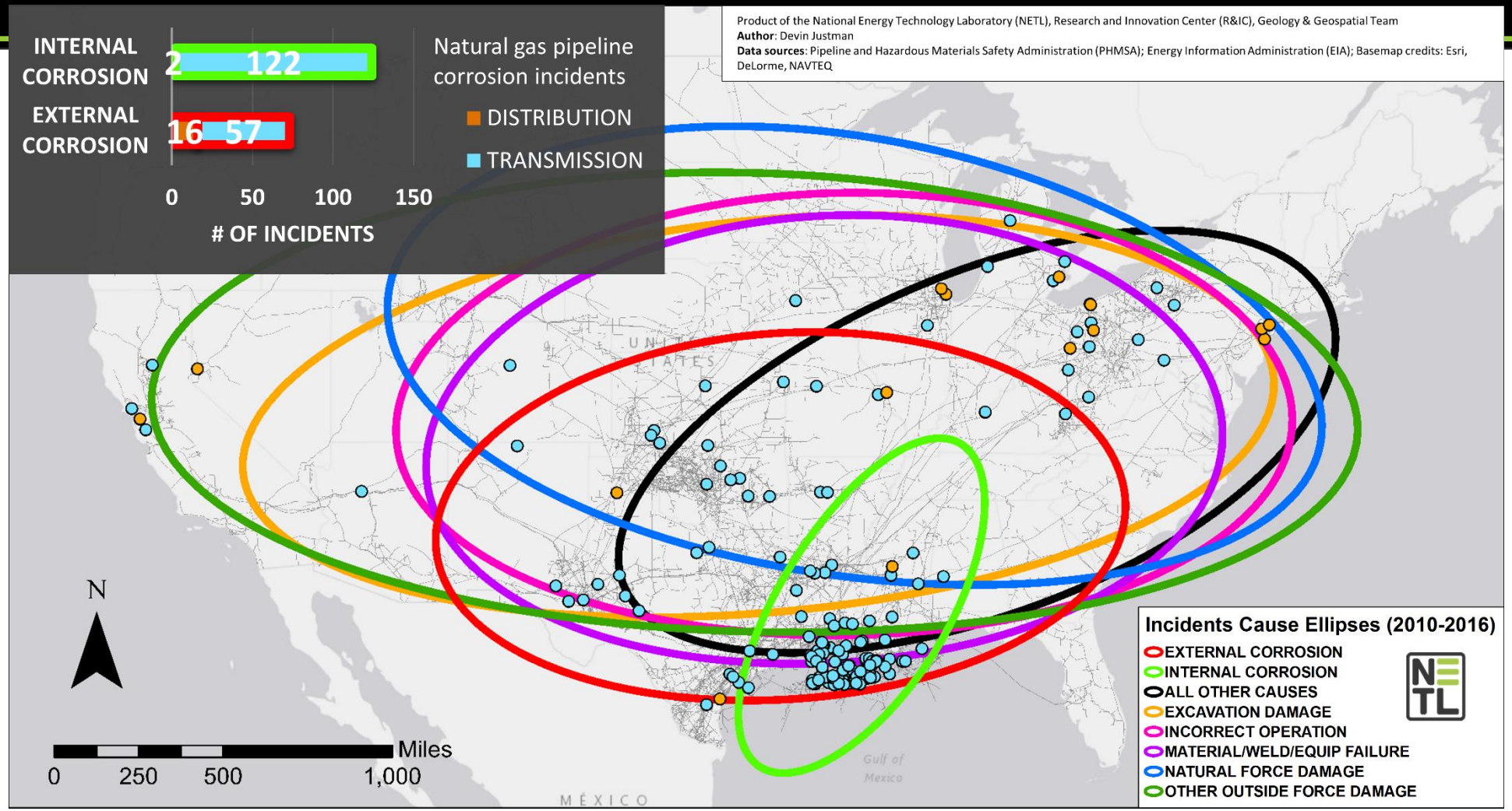
Source: Pipeline and Hazardous Materials Safety Administration (PHMSA)

Identifying key spatio-temporal trends



Source: Pipeline and Hazardous Materials Safety Administration (PHMSA)

Identifying key spatio-temporal trends



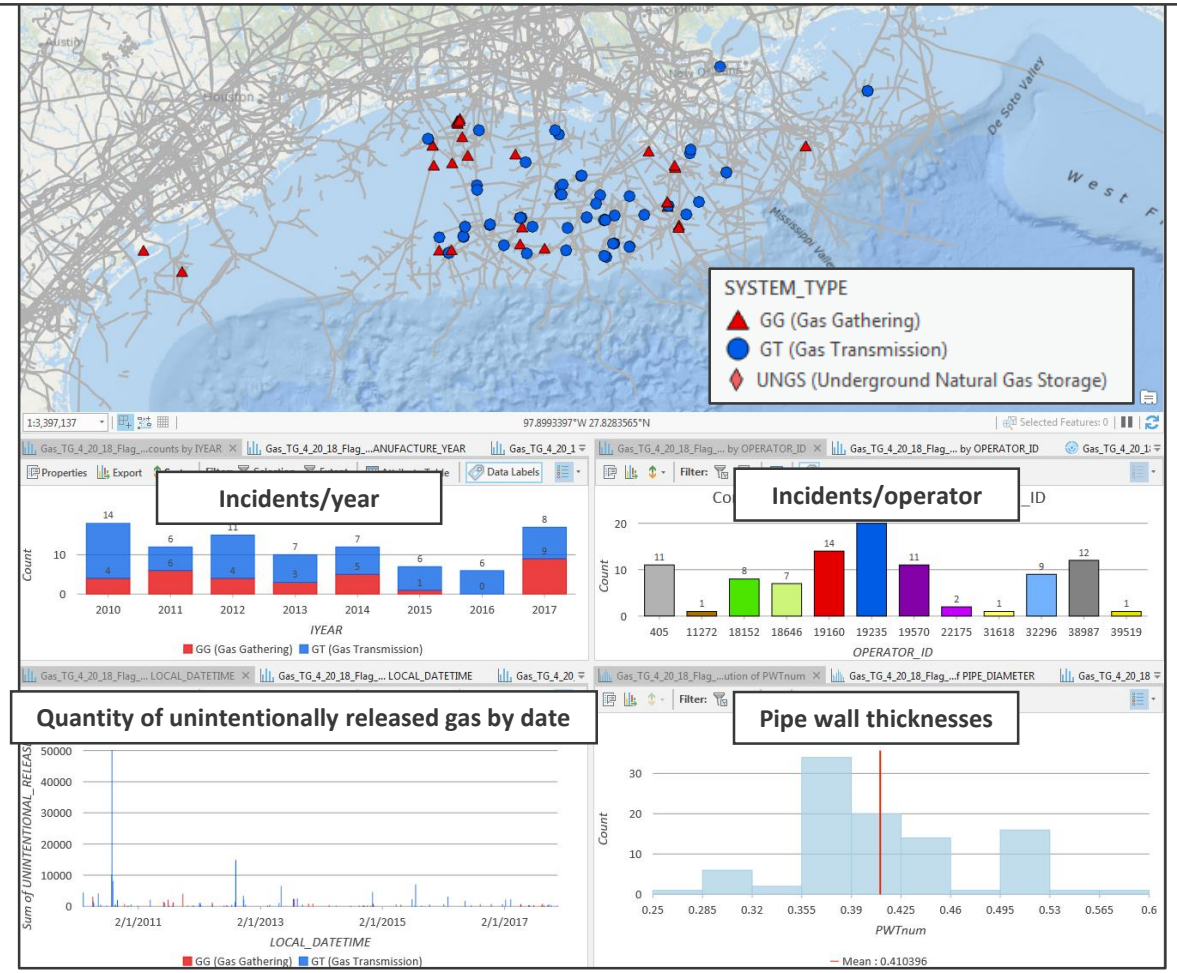
Source: Pipeline and Hazardous Materials Safety Administration (PHMSA)

Identifying key spatio-temporal trends

Of the 97 incidents (65 transmission/32 gathering) :

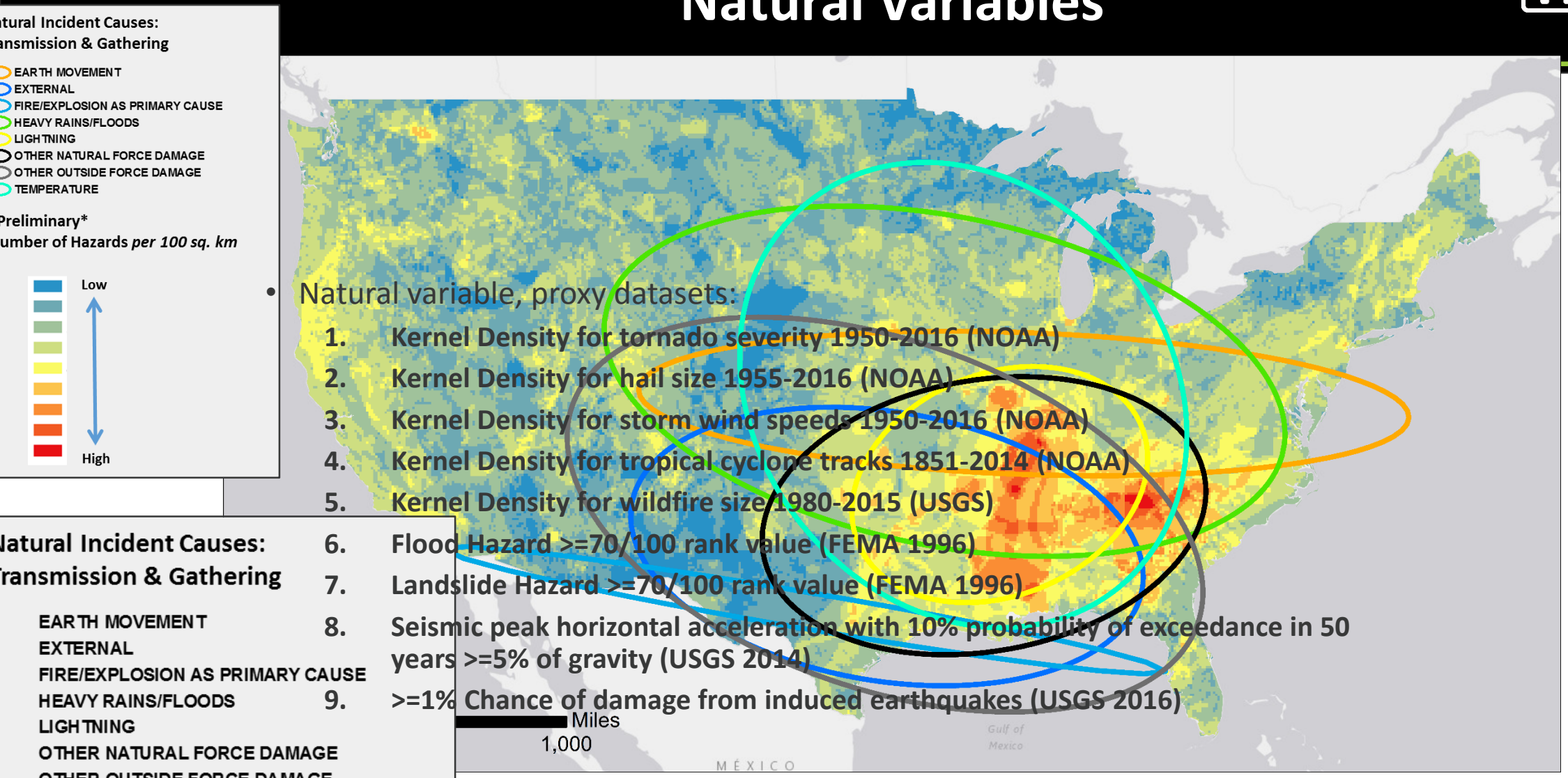
- All classified as pinhole leaks with most occurring (71%) at a low point in pipe
  - 48% microbiological cause of corrosion
  - 20% water/acid presence
- Only 7% of associated pipelines were configured to accommodate inspection tools
- 63.9% of pipeline associated with incidents have operational factors that significantly complicate the execution of an internal inspection tool run
  - 46% low/absence of flow
- Only 9% of associated pipelines had a protective interior coating
- Only 10% of associated pipelines were cleaned/de-watered routinely
- 92% of associated pipelines have not had a hydro or other pressure test conducted since original construction
- 20% of associated pipelines are still shutdown since incident occurred
- 16% of associated pipelines were abandoned or planning to be abandoned

Internal Corrosion incidents in natural gas transmission/gathering pipelines in the Gulf of Mexico (2010 – 2018)

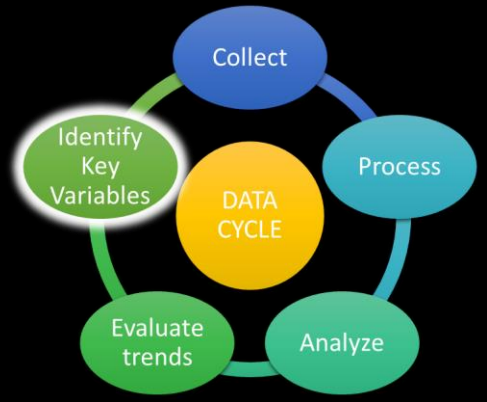


Source: Pipeline and Hazardous Materials Safety Administration (PHMSA)

Failure Causes related to Natural Variables



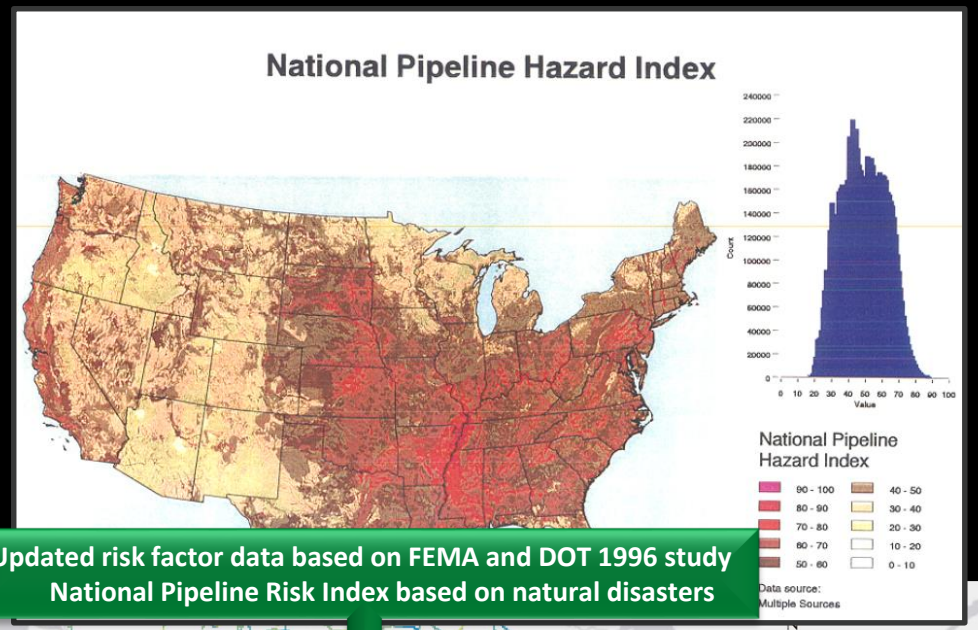
- Natural variable, proxy datasets:
  1. Kernel Density for tornado severity 1950-2016 (NOAA)
  2. Kernel Density for hail size 1955-2016 (NOAA)
  3. Kernel Density for storm wind speeds 1950-2016 (NOAA)
  4. Kernel Density for tropical cyclone tracks 1851-2014 (NOAA)
  5. Kernel Density for wildfire size 1980-2015 (USGS)
  6. Flood Hazard >=70/100 rank value (FEMA 1996)
  7. Landslide Hazard >=70/100 rank value (FEMA 1996)
  8. Seismic peak horizontal acceleration with 10% probability of exceedance in 50 years >=5% of gravity (USGS 2014)
  9. >=1% Chance of damage from induced earthquakes (USGS 2016)



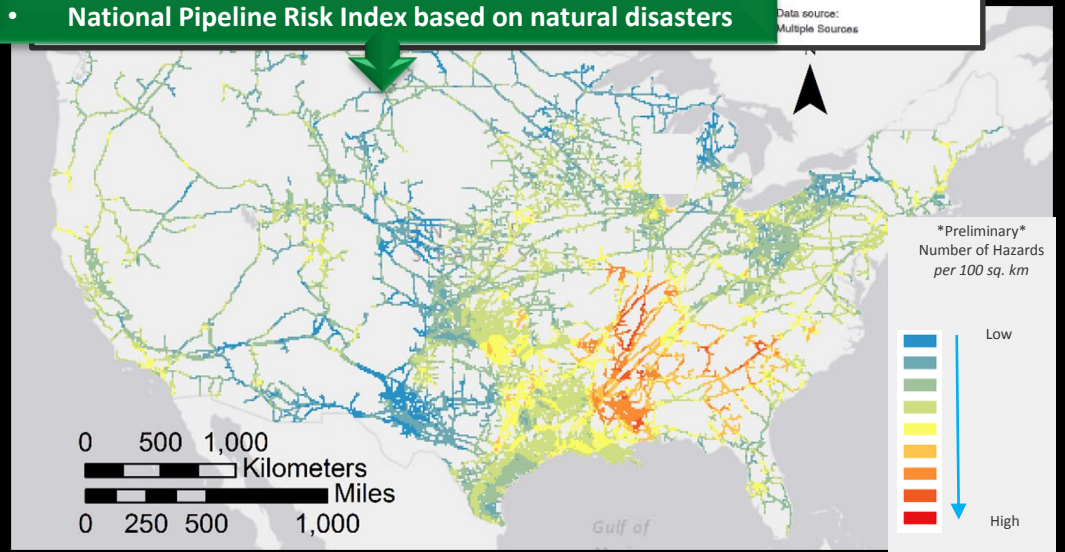
Identifying Key Variables

Used hypothesis driven methods to determine **key variables** affecting pipeline failures, such as finding from peer-reviewed literature, which represents:

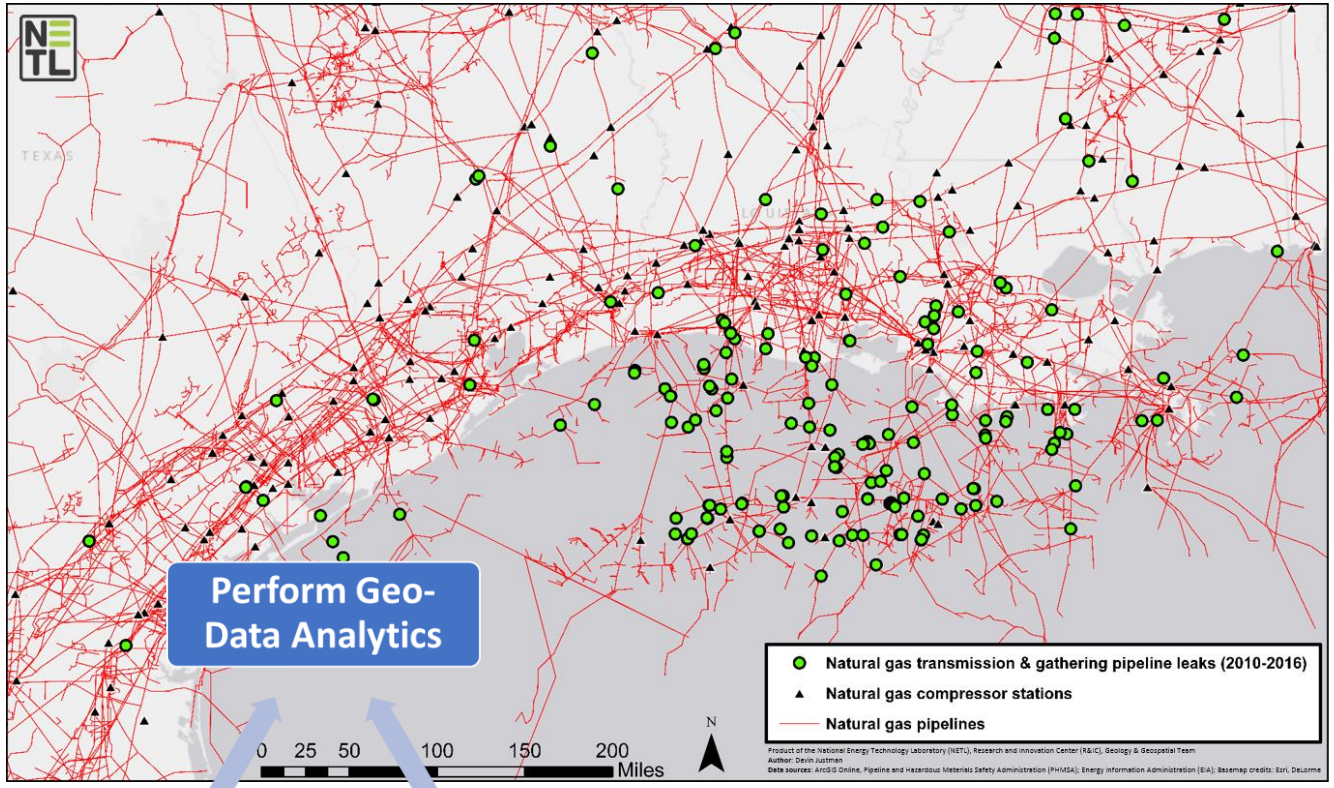
- the **existing Natural Gas Pipeline Infrastructure**,
- more than **45 years of Pipeline incidents**, and
- over **200 different internal and external factors** that affect pipeline incidents, including:
  - pipeline material, pipeline age, maintenance and construction activities, landslides, earthquakes, lightning strikes, severe rain, hail events, soil type, soil composition, land use, population growth, land development, and others



Updated risk factor data based on FEMA and DOT 1996 study
- **National Pipeline Risk Index based on natural disasters**



Building towards geo-data analytics



Perform Geo-Data Analytics

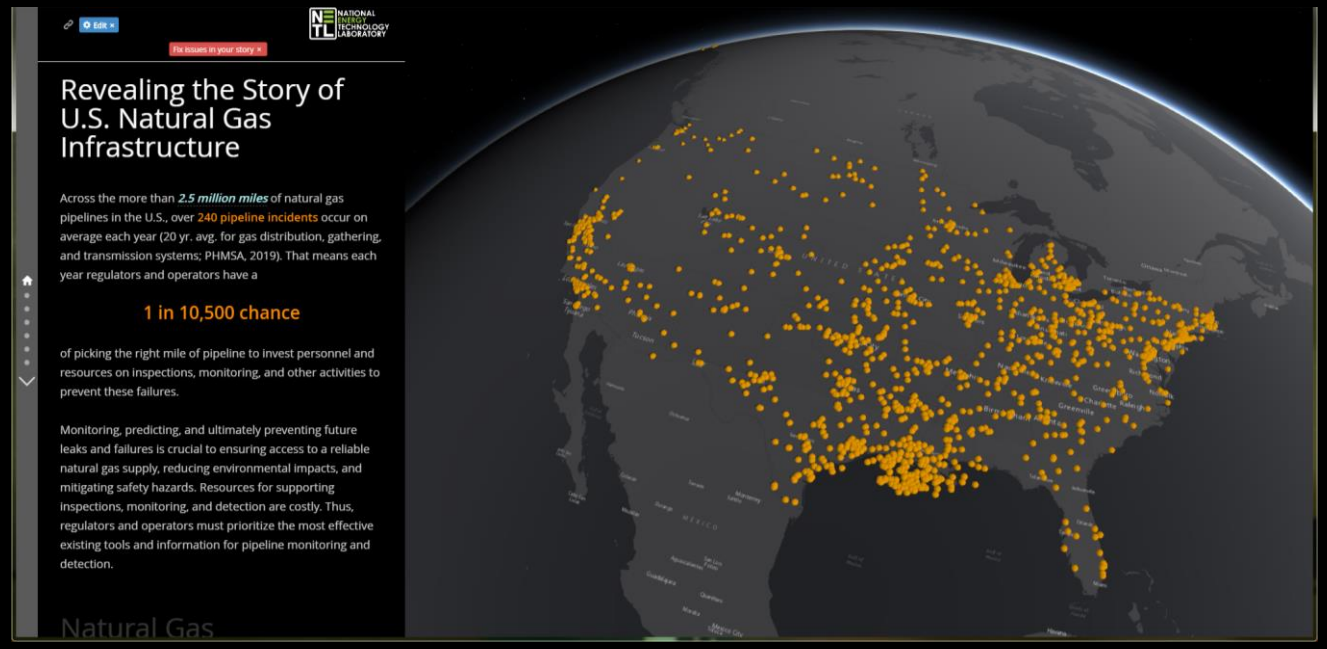
Inform sensor deployment and utilization

Reduce Risk & Impacts

- Began outlining and developing an analytical workflow to support the evaluation of benefits for deployment of new sensor technologies and other advanced pipeline materials
- Includes testing and validating key variables identified in exploratory geo-data analytics, as well as information pertinent to sensor operational characteristics from subject matter experts and foundational literature

Data Analysis & Visualization

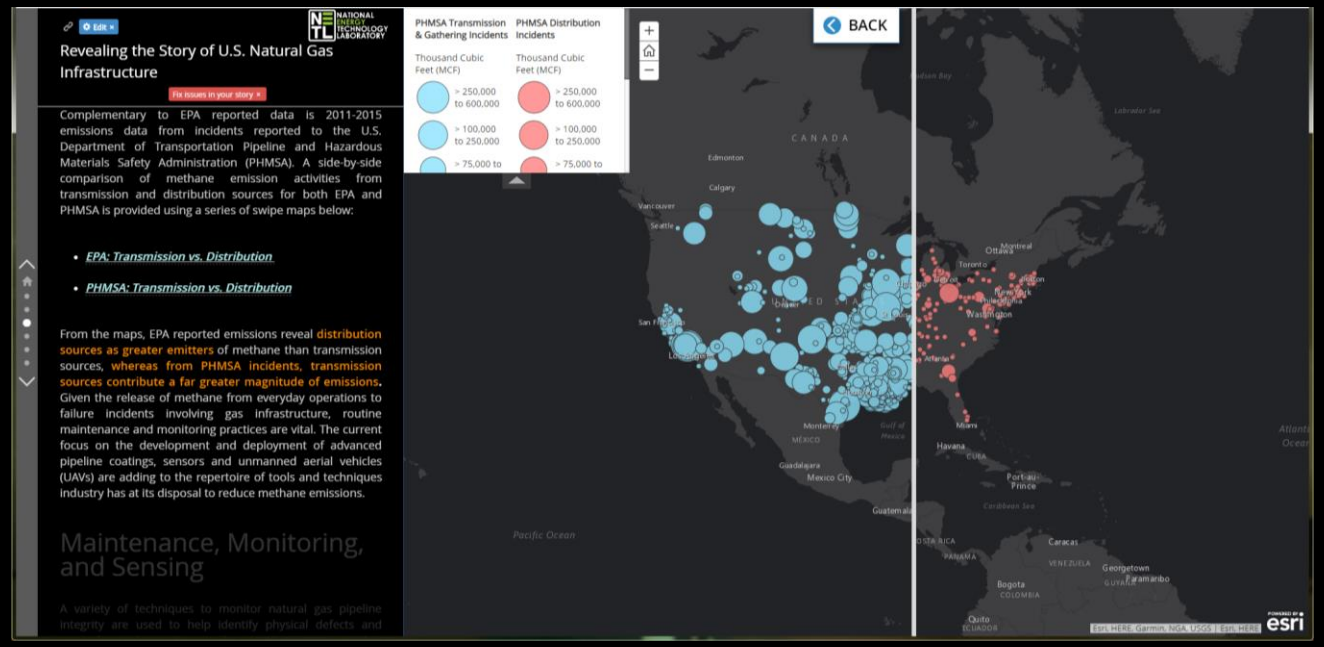
- Integrating key information, data, as well as key findings from exploratory geo-data analytics into a story map (right) and Jupyter notebook (below)
  - support spatial analytics, as well as facilitate communication of key findings for optimized sensor deployment and utilization




```python
def filter_images():
    global df
    area = geocode(location.value, out_sr=landsat.properties.spatialReference)[0]
    extent = area['extent']

    selected = landsat.filter_by(where="(Category = 1) AND (CloudCover <=0.10)",
                                 geometry=geometry.filters.intersects(extent))
    fs = selected.query(out_fields="AcquisitionDate, GroupName, Best, CloudCover, WRS_Row, WRS_Path,
Month, Name",
                        return_geometry=True,
                        return_distinct_values=False,
                        order_by_fields="AcquisitionDate")
    tdf = fs.df
    df = tdf
    tdf['Time'] = pd.to_datetime(tdf['AcquisitionDate'], unit='ms')

    if len(tdf) > 1:
        update_sliders(tdf)

    # m.draw(tdf.iloc[oldslider.value].SHAPE)

    return tdf
```

```
In [25]: df = filter_images()
```

```
In [26]: # create the action button
diffbtn = widgets.Button(description='Detect changes', disabled=False,
                         button_style='success', tooltip='Show Different Image',
                         icon='check')

def on_diffbutton_clicked(b):
    # m.clear_graphics()
    first = df.iloc[oldslider.value].OBJECTID
    last = df.iloc[newslider.value].OBJECTID
    old = landsat.filter_by('OBJECTID='+str(first))
    new = landsat.filter_by('OBJECTID='+str(last))
    diff = stretch(composite_band([ndvi(old, '5 4'),
                                   ndvi(new, '5 4'),
                                   ndvi(old, '5 4')]),
                                   stretch_type='stddev',  num_stddev=3, min=0, max=255, dra=True,
astype='u8')
```



Presentations:

Justman, D., Bauer, J., Bean, A., Rose, K., and Huerta, N., 2017, Analytical Framework for Evaluating Natural Gas Pipeline Risk. Oral presentation at the Esri User Conference, San Diego, CA, July 10-14, 2017.

Justman, D., Bauer, J., Bean, A., Rose, K., and Huerta, N., 2017, Developing an Advanced GIS Analytical Framework for Evaluating Risks to Natural Gas Pipeline Infrastructure. Poster presentation at the Esri FedGIS Conference, Washington D.C., February 13-14, 2017.

Justman, D., Rose, K., Bauer, J., 2016, Preliminary analysis integrating climate and environmental variables into a pipeline infrastructure decision support and risk evaluation framework: Leveraging NETL's novel data mining and geospatial tools, Boulder, CO July 13-15, 2016. Oral presentation at the Engineering for Climate Extremes Partnership (ECEP) Workshop at the National Center for Atmospheric Research (NCAR)

Rose, K., Baker, D.V. "Vic", Bauer, J., Dehlin, M., Jones, T.J., and Rowan, C., 2017, Working Smarter Not Harder – Developing a Virtual Subsurface Data Framework for US Energy R&D, invited talk, American Geophysical Union Annual Meeting.

References:

Bauer, J.; Rose, K. (2015). Variable Grid Method: an Intuitive Approach for Simultaneously Quantifying and Visualizing Spatial Data and Uncertainty. Transactions in GIS, 19, 377–397.

Bauer, J . R. et al. (2015). A Spatio-Temporal Approach to Analyze Broad Risks Potential Impacts Associated with Uncontrolled Hydrocarbon Release Events in the Offshore Gulf of Mexico; NETL-TRS-2-2015; EPAct Technical Report Series; U.S. Department of Energy, National Energy Technology Laboratory: Morgantown, WV, 2015; p 60.

FEMA (1996). Natural Disaster Study National Pipeline Risk Index Technical Report. Federal Emergency Management Agency Mitigation Directorate.

Petersen, M. D., Mueller, C. S., Moschetti, M. P., Hoover, S. M., Llenos, A. L., Ellsworth, W. L., ... & Rukstales, K. S. (2016). 2016 One-Year Seismic Hazard Forecast for the Central and Eastern United States from Induced and Natural Earthquakes (No. 2016-1035). US Geological Survey.

Petersen, M. D., Moschetti, M. P., Powers, P. M., Mueller, C. S., Haller, K. M., Frankel, A. D., ... & Field, N. (2014). Documentation for the 2014 update of the United States national seismic hazard maps (No. 2014-1091). US Geological Survey.