The background of the slide is a photograph of an industrial facility, likely an oil or gas processing plant, silhouetted against a sunset sky. A large, bright orange and yellow flare is visible on the right side of the image. The text is overlaid in white, bold, sans-serif font.

**GRADUATE STUDENT DISSERTATION  
RESEARCH IN THE TEXAS FEDERAL  
STATISTICAL RESEARCH DATA  
CENTER (TXRDC): THE EFFECTS OF  
ORGANIZATIONAL CHARACTERISTICS  
AND THE CHARACTERISTICS OF  
ORGANIZATIONAL INSTITUTIONAL  
ENVIRONMENTS ON TEXAS OIL AND  
GAS VENTING AND FLARING  
PRACTICES**

**A PRESENTATION BY:**

**KATE WILLYARD, PHD CANDIDATE  
TEXAS A&M, DEPARTMENT OF SOCIOLOGY  
DECEMBER 9, 2016**

# **DISCLAIMER**

**Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau.**

# PRESENTATION OBJECTIVES

- **Theoretical Background of Project**
  - What has been done
  - What needs to be done
- **Project Description**
  - Overview
  - Data
  - Benefits
  - Why the RDC is needed
- **The RDC Process**
  - Benefits of dissertation using an RDC
  - Timeline
  - How to get started
  - What to consider when writing a proposal

# AN ORGANIZATIONAL THEORY OF THE ENVIRONMENT

- **Organizational Political Economy of the Environment (Prechel 2015)**
  - Resource Dependence Theory (Pfeffer and Salancik 1978)
    - Organizations are open systems
    - Organizations are constrained due to resource dependence
    - Organizations interact with their social environment and have the capacity to change structural relations within and between organizations
    - Organizational and political characteristics affect the opportunities, incentives and dependencies that influence strategic decisions regarding the eco-efficiency (i.e., the minimization of environmental costs) of industrial production practices (Prechel and Zheng 2012)
  - Modern Political Economy (Polanyi 1944)
    - Corporate organizations are embedded within their political institutional environment
    - The degree of political embeddedness varies across time and between organizations
    - Variations in the degree of embeddedness affect the different incentives and opportunities for corporations to pollute at a higher or lower rate (Prechel and Zheng 2012)

# PRIOR RESEARCH

- **Plant-Level Findings**

- Size (+)
- Race (+)
- Class (+)
- Subsidiaries (+)
- Absentee Management x Community Organization (-)

- **Ultimate Parent Company-Level Findings**

- Size (-)
- Subsidiaries (+)
- Capital Dependence (+)
- Complexity (+)
- Age (+)

# MISSING LINKS

- **Research has not examined the effects of branch plants and operating companies**
- **Research has yet to examine the oil and gas industry**
- **Why? Data Access!**
  - Oil and gas industry is exempt from EPA's TRI
  - Oil and gas industry is required to report to the EPA GHGRP at the shale-level only if oil and gas production operations are estimated to emit more than 25,000 metric tons of greenhouse gasses within the year
  - Dunn & Bradstreet fails to identify oil and gas wells as branch plants
  - Compustat identifies oil and gas production activity at the level of the parent company

# PROJECT OVERVIEW

## FSRDC Research Project Main Research Question:

- How do organizational characteristics and the characteristics of the organization's institutional environment at the gas well/oil lease, operating company, and ultimate parent company-level relate to Texas onshore oil and gas extraction facility (NAICS 211) venting and flaring practices?

## Dissertation Research Question:

- How do organizational characteristics and the characteristics of the organization's political environment at the gas well, and operating company-level relate to Texas onshore oil and gas extraction facility (NAICS211) venting and flaring rates?

# MAJOR ISSUES

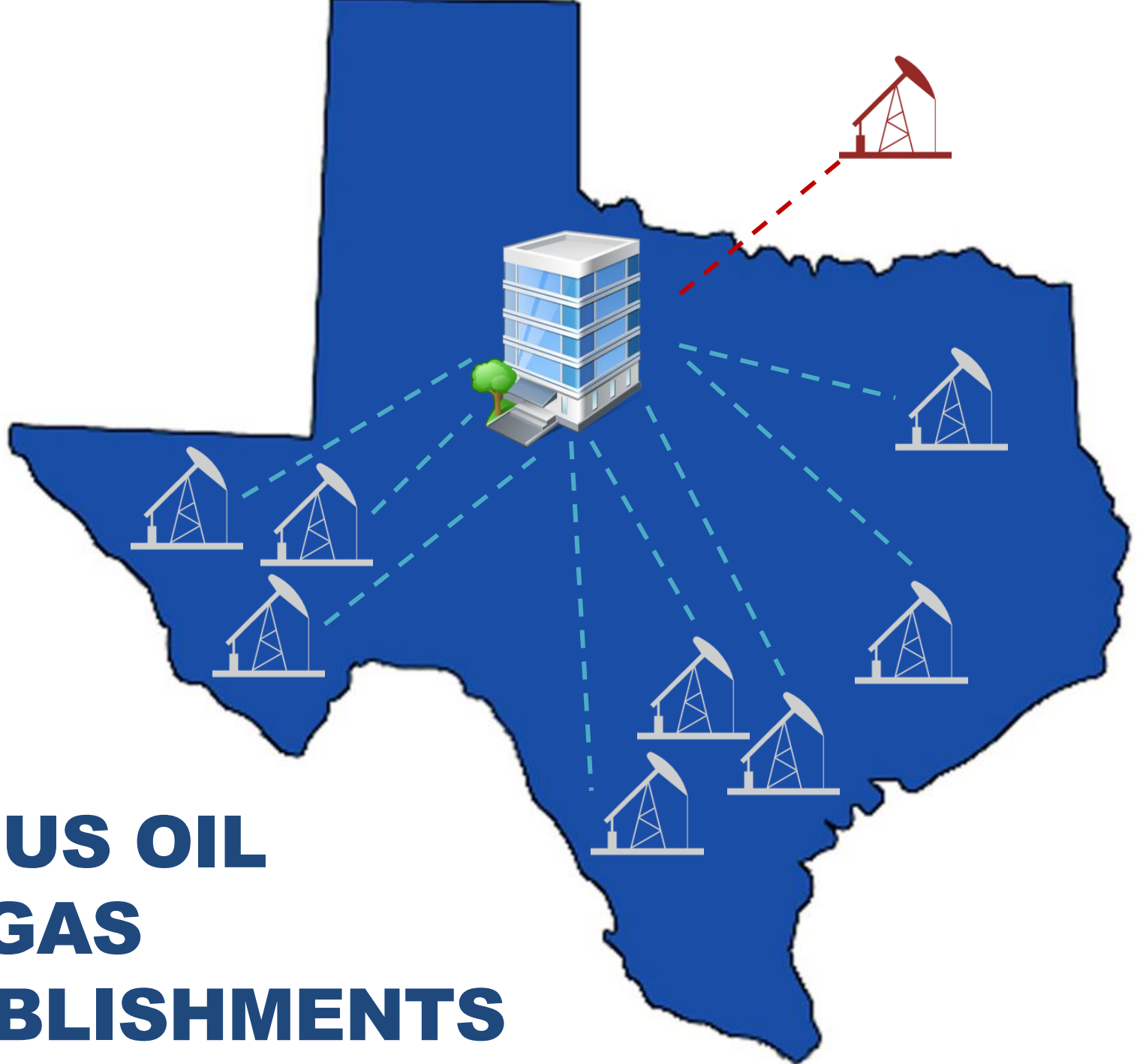
## **Oil and Gas Extraction Industry Is Organized Differently Than Most Sectors**

- **Operating company controls one or more branch plants (i.e., gas wells/oil leases) in one or more states**
- **Operating companies can be ultimate parent companies, subsidiaries, affiliates, or independent companies**

## **Census Terms for an Oil and Gas Extraction Establishment Is Different Than Most Sectors**

- **Census defines an establishment as a single physical location where business is conducted or where services or industrial operations are performed (in business terms, it is a plant)**
- **Census defines an oil and gas extraction establishments (NAICS 211) as oil and gas field activities in one state or offshore area operated by the reporting company (in business terms, it is central headquarters controlling operations at one or more branch plants in a state)**

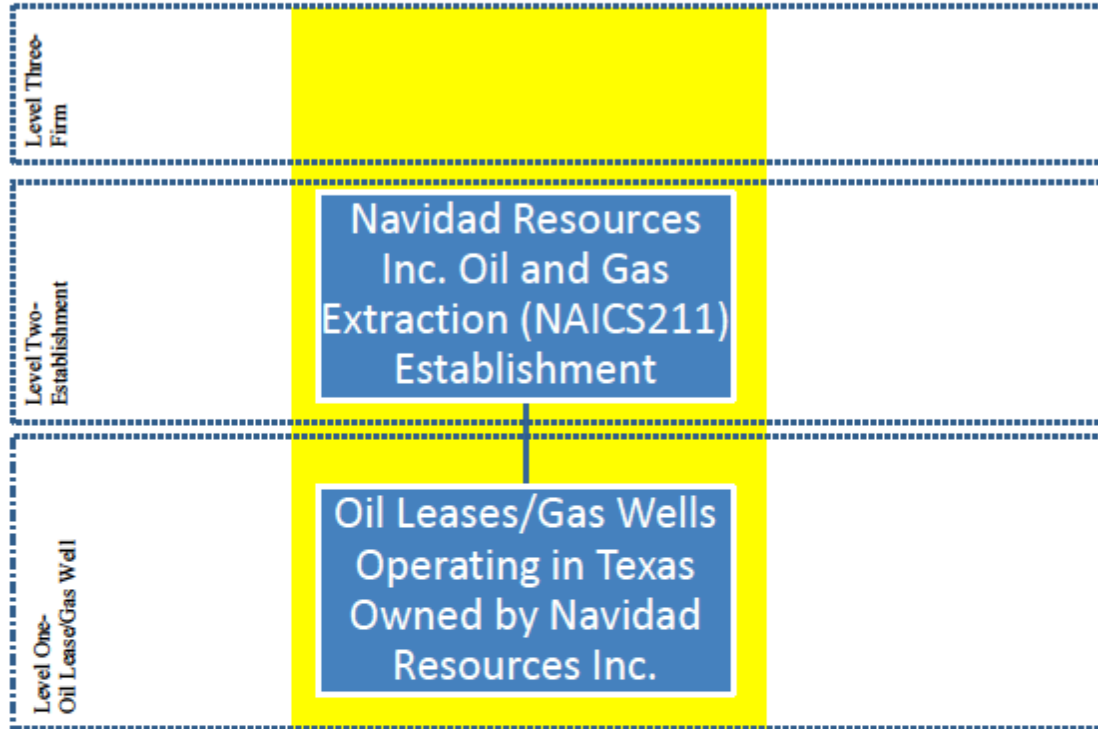




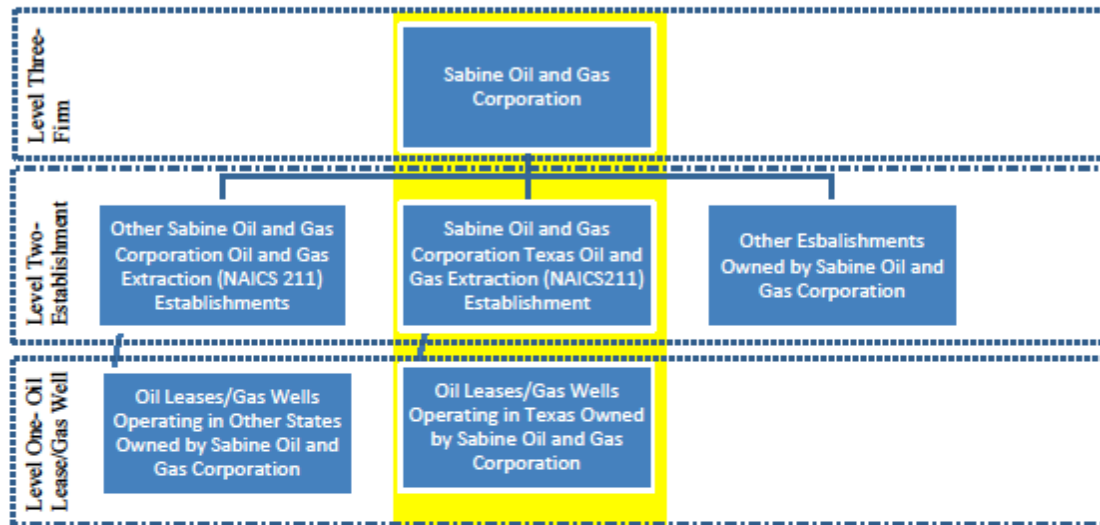
# CENSUS OIL AND GAS ESTABLISHMENTS

Example of Structure of Oil Leases/Gas Wells Owned by Single-Unit Enterprise

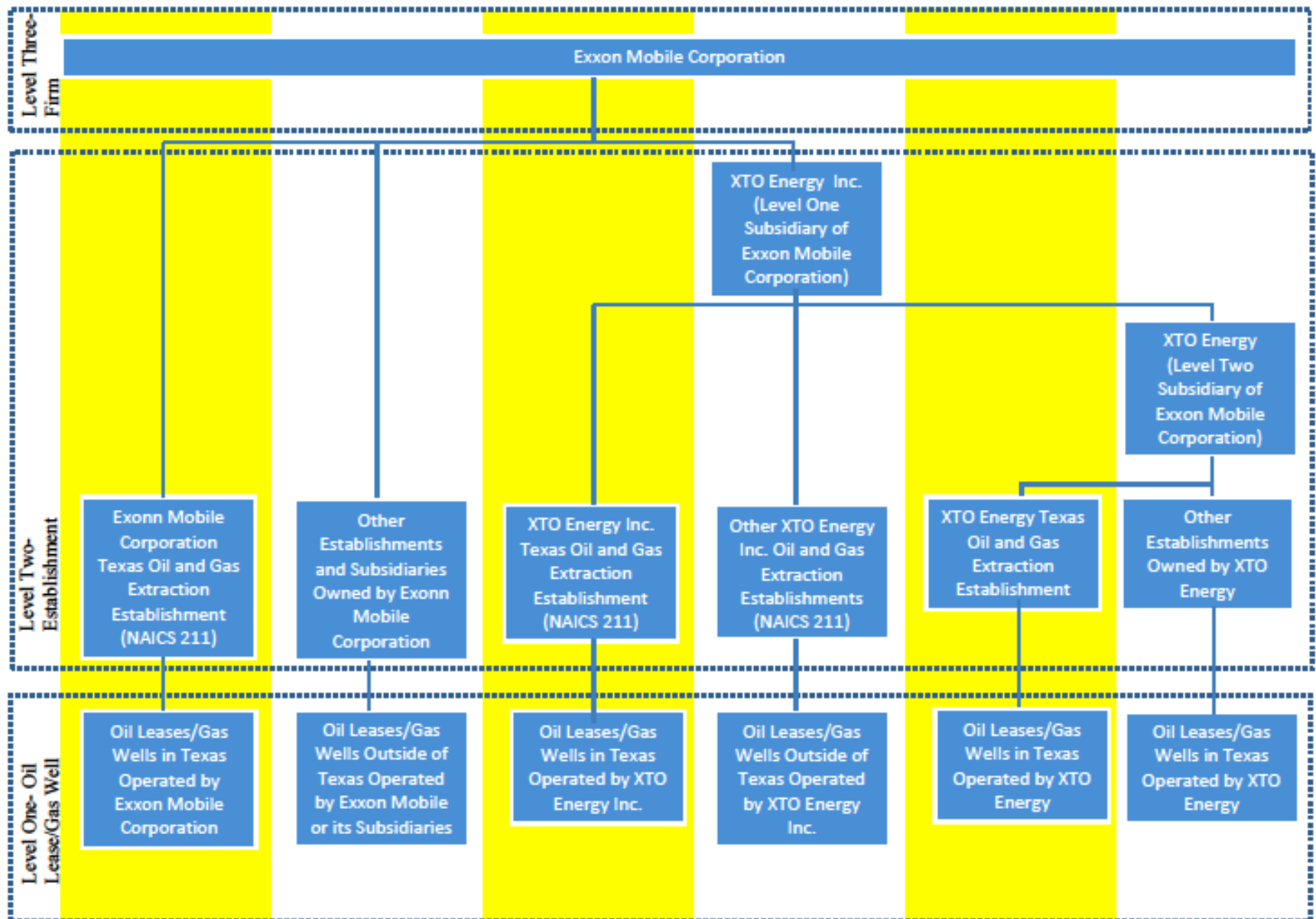
(i.e., establishment not embedded within a firm)



**Example of Structure of Oil Leases/Gas Wells Owned by Multi-Unit Enterprise (i.e., firm),  
Organized as an Independent Company, Not a Multilayer Subsidiary Organization**



### Example of Structure of Oil Leases/ Gas Wells Owned by Multi-Unit Enterprise (i.e., firm), Organized as a Multilayer Subsidiary Organization



# DATA

- **Researcher Provided Data**

- Texas Railroad Commission Databases
- Environmental Protection Agency Greenhouse Gas Report
- United States Energy Information Administration Natural Gas Interstate and Intrastate Pipelines Shapefile
- Census TIGER/Line shapefiles
- Guidestar
- National Center for Charitable Statistics
- Lexis Nexis Corporate Affiliations

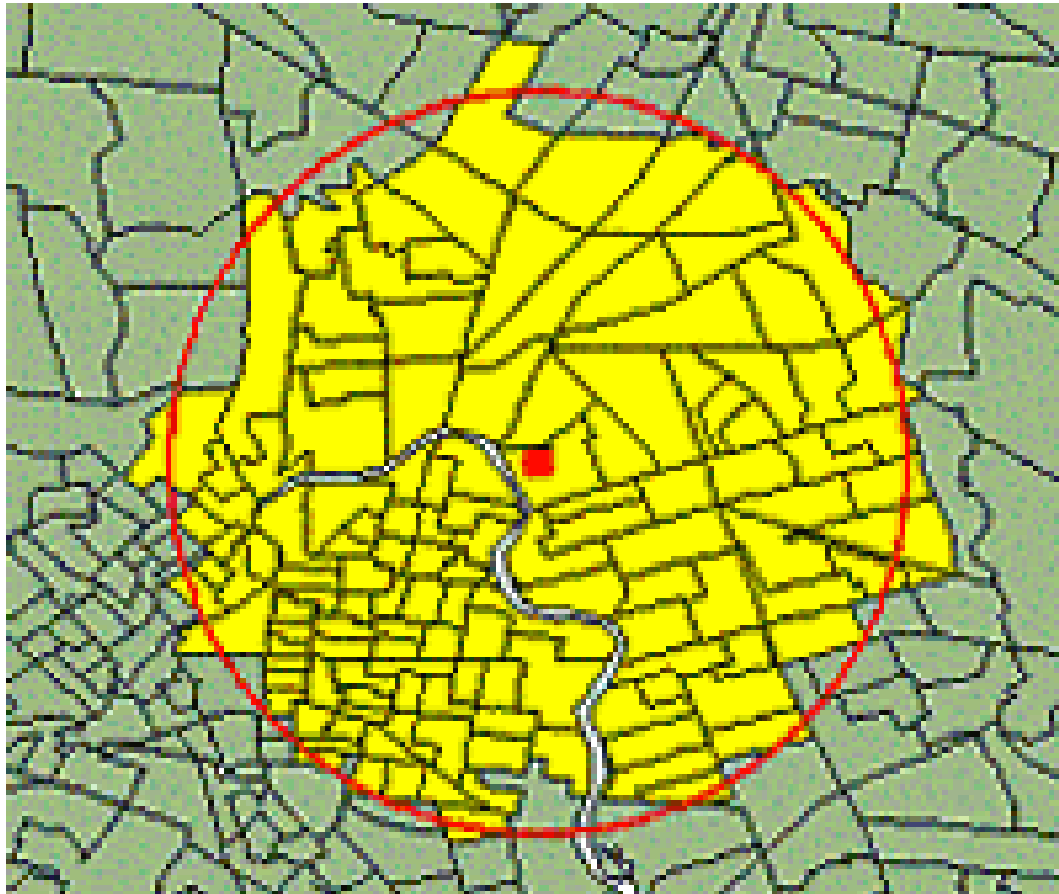
- **Census Provided Data**

- Longitudinal Business Database, 2002-2013
- Report of Organization, 2002-2013
- Census of Mining, 2002-2012
- Business Register/SSEL, 2002-2013
- Decennial Census, 2010
- American Community Survey, 2002-2014

# CONNECTING DATASETS

- **Public data matched using:**
  - Texas Railroad Commission unique identifiers
  - Geographic Information System
- **Public data matched to restricted RDC data using:**
  - Census block IDs
  - Operating company name and address
  - Firm EIN
  - Census establishment unique identifiers
  - Census firm unique identifiers

# CENTROID POINT IN BLOCK POLYGON SPATIAL OVERLAY APPROACH TO CALCULATE GAS WELL COMMUNITY CHARACTERISTICS



# FIRST CENSUS BENEFIT

**Criterion 5- Understanding and/or improving the quality of data produced through a Title 13, Chapter 5 survey, census or estimate:**

- **Finding the difference between estimates of establishment oil and gas production totals produced through the Census of Mining to estimates produced through Texas Railroad Commission records.**
- **Producing summary statistics of the types of establishments that reported oil and gas production to the Texas Railroad Commission that did not report to the Census of Mining**



# **SECOND CENSUS BENEFIT**

**Criterion 8- Identifying the limitations of, or improving, the underlying Business Register, Master Address File, and industrial and geographic classification schemes used to collect the data:**

- **Identifying the establishments that reported oil and gas extraction activities to the Texas Railroad Commission**
- **Produce summary statistics of the establishment industrial classification schemes and the production totals of establishments within different industrial classification schemes**

# **THIRD CENSUS BENEFIT**

**Criterion 11- Preparing estimates of population and characteristics of population as authorized under Title 13, Chapter 5:**

- **Producing coefficient estimates of the characteristics of oil leases/gas wells, extraction establishments and/or ultimate parent companies relationships to gas well/oil lease venting and flaring practices.**

# NECESSITY OF RDC DATA

- **Community Data**
  - Need lower-level geography to better measure the characteristics of communities within a short distance of the gas well/oil lease
- **Business Data**
  - Establishment-level data is not publically available
  - Firm-level data is not publically available
- **Federal Tax Information (FTI) Need**
  - Some data needed to match/link datasets
  - Other fields required to identify establishments of the same industry
  - Most needed as predictor variables or statistical controls

# **PROS AND CONS OF DISSERTATION RESEARCH IN AN RDC**

## **Pros**

- **Larger community data sample sizes**
- **Access to business data not publically available**
- **Eligible for the Center for Economic Studies Dissertation Mentorship Program**
- **Part of post-graduation research plan**
- **Increase competitiveness for post-docs and other positions at top research universities with access to a RDC**

## **Cons**

- **Time**
- **Complexity**

# **TIMELINE PART 1- STEPS COMPLETED**

- **December 8, 2014: Started to get Special Sworn Status (SSS) for affiliation with Dr. Harland Prechel's approved project**
- **May 8, 2015: Obtained SSS**
- **September 25, 2015: Attended RDC Proposal Development Workshop**
- **September 30, 2015: Meet with TXRDC Administrator**
- **April 8, 2016: First Draft of Dissertation Proposal Submitted**
- **May 10, 2016: First Draft of RDC Proposal Submitted**
- **October 10, 2016: Defended Dissertation Proposal**
- **December 1, 2016: Final Draft of RDC Proposal Submitted for Census and Internal Revenue Service Internal Review**

# **TIMELINE PART 2- IDEAL TIMELINE**

## **2017**

- **Connect public datasets**
- **Build Geographic Information System using public data**
- **Conduct preliminary analysis using public data in non-restricted computing environment**
- **Obtain access to RDC data**
- **Start merge of public data with restricted data**

## **2018**

- **Conduct analysis for 2-Level HLM w/limited restricted variables (i.e., dissertation analysis)**
- **Assess goodness of fit, quality of estimates for 2-level HLM**
- **Submit first request for disclosure**
- **Complete dissertation**

# **TIMELINE PART 3- IDEAL TIMELINE CONTINUED**

## **2019**

- **Finish connecting all datasets**
- **Conduct analysis for complete 2-Level HLM, and assess goodness of fit, and quality of estimates**
- **Draft paper with dissertation findings**

## **2020**

- **Draft paper with findings for complete 2-Level HLM**
- **Conduct analysis for 3-Level HLM, and assess goodness of fit, quality of estimates for 2-level HLM**
- **Draft paper with findings for 3-Level HLM**

# **TIMELINE PART 3- IDEAL TIMELINE CONTINUED**

## **2021**

- **Finish addressing referee reports from journals where project papers were submitted for publication**
- **Complete Census Benefits**
- **Describe how the project met Title 13, Chapter 5 requirements**
- **Complete CES working paper**
- **Submit Post-Project Certification to CES**
- **Tie any other loose strings**



# HOW TO GET STARTED

- **Meet with your advisor FIRST**
- **Meet with the TXRDC Administrator AFTER you have talked with your advisor, and have a good idea of what your dissertation question will be**
- **Do research on RDC research**
  - Attend the proposal development workshop
  - Find codebooks of public Census community data
  - Find forms filled out by companies for the Economic Census
  - Don't expect the TXRDC Administrator will do the work for you

# WHAT TO CONSIDER WHEN WRITING A PROPOSAL

- **Talk to and listen to the RDC Administrator**
- **Develop a 5 year timeline for projects that work with economic data**
- **Do research on:**
  - The questionnaires for the data in which you are interested
  - The sample sizes for the data in which you are interested
  - Completed RDC projects that use similar datasets
  - Methods past RDC projects used to match datasets and their reported match rates
- **Wait until after your dissertation proposal has been defended until you submit your proposal**

# ADVICE ON PROJECT MANAGEMENT

- Before you even enter the RDC, read the Thin Client User Guide at <http://txrdc.tamu.edu/computing-software/>
- Plan at least a whole working week to learn the system
- Keep good records/organization practices; document everything as you go
- Stay organized! I recommend using the practices described by Dr. Scott Long in his book “The Workflow of Data Analysis Using Stata”
- Don’t expect to take out results emerging from point and click methods
- Best practice is to write programs where the resulting log file is automatically saved to an assigned folder

# DISCLOSURE ADVICE

- **Document your work, what programs were run, and where the resulting log files are located**
- **I recommend starting your program with a two digit number regarding the step number and ending with the date it was run and give the resulting log file the same name**
  - 01-importRRC1\_20161209.do/01-importRRC1\_20161209.log
- **Save the files/ results with the date they were created**
  - merged2012RRCssel\_20161129.dta
- **Follow the Thin Client User Guide instructions precisely**
- **Be sure you are disclosing what you want to disclose, because it may be difficult to make small changes and try to release similar output at a later time**

# MORE ADVICE FOR GRAD STUDENTS

- **Keep your advisor involved in every step of the process; your advisor must be fully on-board**
- **It is a big commitment to carry out a project, so plan for your project to go well beyond your dissertation**
  - Projects that use business data are typically 5 years
  - Think of ways your project can help you with your early career
- **Expect for it to take way more time to finish your dissertation**
- **Institute a back-up plan**
  - Although they try to be considerate, Census doesn't care about your graduation timeline and how you may need data to be released by a certain time in order to graduate
  - Be sure to develop a plan with your advisor so that you can graduate without using RDC data (and then finishing your project as part of your early career research)

**KATE WILLYARD**  
**DEPARTMENT OF SOCIOLOGY**  
**TEXAS A&M UNIVERSITY**  
**[KATE.WILLYARD@TAMU.EDU](mailto:KATE.WILLYARD@TAMU.EDU)**

**QUESTIONS?**  
**CONTACT ME!**