

# Federal Statistical Research Data Centers (FSRDC)

Samuel R. Bondurant

Dallas-Fort Worth Federal Statistical Research Data Center  
Center for Enterprise Dissemination  
US Bureau of the Census

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

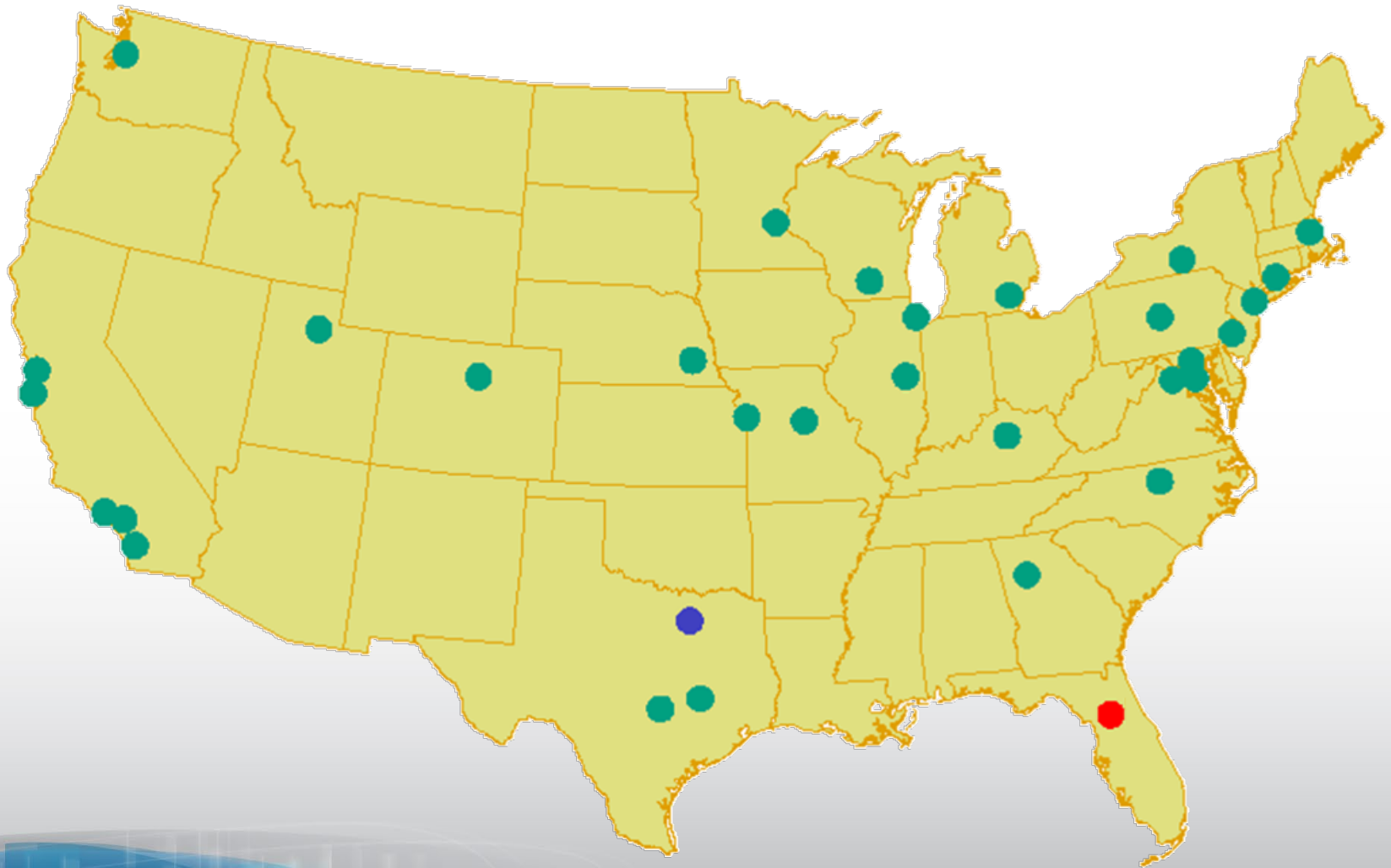
# Goals of this Presentation

- Overview of FSRDC program
  - Description of working environment
- Phases of an RDC Project
- Available data products and advantages
- How to get started

# Federal Statistical Research Data Center (FSRDC)

- Federally secure research lab where authorized researchers are granted secure access to restricted-use microdata
  - Thin client access to Census linux servers
  - Census Bureau and other statistical agencies
  - Fees to maintain facility typically paid by a local consortium of research institutions
    - Additional fees for non-Census agency projects

# Research Data Centers



# Census Research Data Centers

RDC	Open	RDC	Open	RDC	Open
Census Bureau	1982	Atlanta Fed	2011	U. Nebraska	2016
NBER	1994	U. Washington	2012	Georgetown	2017
UC Berkeley	1998	Texas A&M	2012	U. Kentucky	2017
UCLA	1998	USC	2014	Philadelphia Fed	2017
Duke/RTI	2000	UC Irvine	2014	U. Colorado Boulder	2017
Chicago Fed	2002	Penn State	2014	U. Texas Austin	2017
U. Michigan	2002	Yale	2015	Dallas Fed	2018
Cornell	2004	U. Missouri	2015	Federal Reserve	2019
CUNY Baruch	2006	U. Wisconsin	2015	U. of Utah	2020
Stanford	2009	Kansas City Fed	2016	UIUC	2021
U. Minnesota	2010	U. Maryland	2016	U. Florida	(2022)

# DFW FSRDC Consortium



TEXAS TECH  
UNIVERSITY.

UNT  
UNIVERSITY OF NORTH TEXAS

SOUTHWESTERN  
MEDICAL CENTER



TCU

UNIVERSITY OF  
TEXAS  
ARLINGTON

UT DALLAS



*The* UNIVERSITY of OKLAHOMA

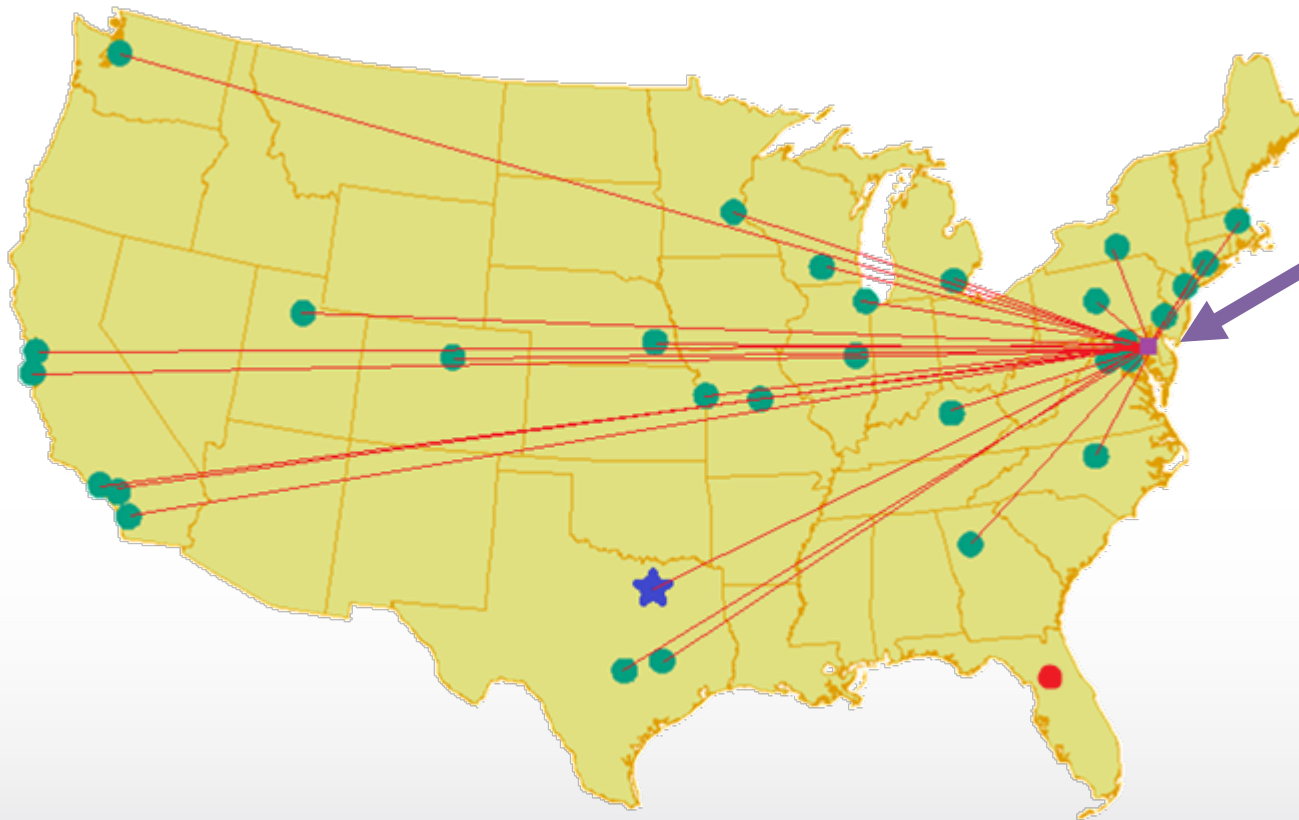


SMU

# What is an RDC?



**\*Not the actual lab**



## Census Servers

- Project space
- Statistical packages
- Data



# Why Is Census Required to Restrict Microdata Access?

- Titles 13 (Census) /26 (IRS) U.S.C. and CIPSEA protect confidentiality
  - respondent cannot be identified
  - only Census employees and temporary staff can access microdata
  - use limited to statistical purpose
  - access must potentially provide legitimate benefits to Census Bureau programs

# RDC Resources

- [www.dfwrdc.org](http://www.dfwrdc.org)
  - Data
  - Proposal Process
  - Presentations
- <https://ideas.repec.org/s/cen/wpaper.html>
  - Working paper series
  - Google “census ces working paper series”

# Quick Poll

- Type into the chat:
  - 1 – I'm very familiar with the FSRDC program
  - 2 – I'm somewhat familiar
  - 3 – I've heard about it
  - 4 – This is my first introduction to the FSRDC

# Phases of an RDC Project

- Proposal drafting
- Submit proposal for review
- Security clearance (SSS)
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census

# Phases of an RDC Project

- **Proposal drafting**
- Submit proposal for review
- Security clearance (SSS)
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census

# \*Census\* Proposal Process

- Pre-Preliminary proposal
  - Send Wenhua and I an email
- Preliminary proposal submission portal
  - ResearchDataGov
  - <https://www.icpsr.umich.edu/web/pages/appfed/index.html>
- Proposal development
  - Involve RDC Admin (potentially other Census staff)
  - Research\_Proposal\_Guidelines.doc located in dfwrdc.org
  
- Note: on the dfwrdc.org website
  - Census proposal guidelines (pdf)
  - Links to NCHS, AHRQ, BLS, BEA proposal processes

# \*Census\* Proposal

## 1. Proposal

- Overview
- Requested Datasets and years
  - Why you need internal data
- Methodology
  - Linking process
  - Estimating equations
  - Measures constructed
- Expected Output
- Duration and Funding

## 2. Benefits to the Census Bureau

- Minimum of two (one being the working paper)

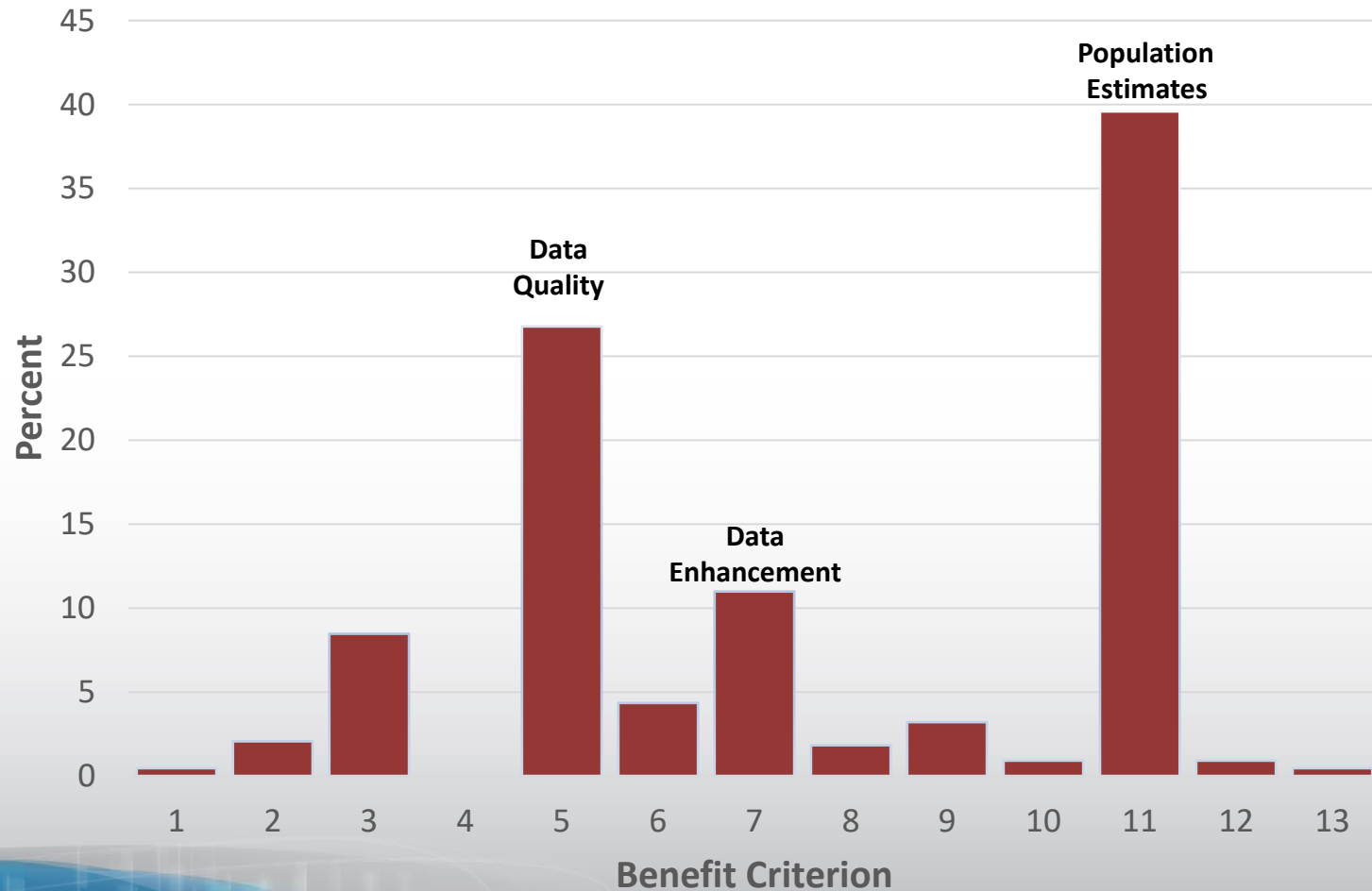
## 3. Abstract

# 13 Criteria for Benefits

1. Evaluating practices, continued relevance
2. Demog., econ. processes that affect Census programs
3. Increasing utility of data for analyzing public programs, public policy, and/or demographic, economic, or social conditions
4. Conducting data collection, processing or dissemination
- 5. Understanding/improving the quality of data**
6. Improved methodology to collect, measure, or tabulate
- 7. Enhancing the data collected (e.g. imputation, links)**
8. Limitations/improving the Business Register
9. New data collection needs
10. Constructing, verifying, improving sampling frames
- 11. Estimates/characteristics of population**
12. Methodology for estimating non-response
13. Developing statistical weights for a survey



# Distribution of Benefit Criterion



# Phases of an RDC Project

- Proposal drafting
- Submit proposal for review
- Security clearance (SSS)
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census

# Phases of an RDC Project

- Proposal drafting
- **Submit proposal for review**
- Security clearance (SSS)
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census

# Proposal Review

- Census Review
  - Scientific merit
  - Need for non-public data
  - Risk of disclosure
  - Feasibility
  - Benefits to Census strength
- Additional Agency Review, if needed
  - Economic data (IRS), Numident (SSA), LEHD (states), CPS (BLS)

# Phases of an RDC Project

- Proposal drafting
- Submit proposal for review
- Security clearance (SSS)
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census

# Phases of an RDC Project

- Proposal drafting
- Submit proposal for review
- **Security clearance (SSS)**
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census

# Special Sworn Status (SSS)

- All persons working in the RDC require SSS
  - Census, NCHS, AHRQ, BLS, BEA, etc.
  - Census absorbs SSS cost, other agencies might not
- Couple months to process (longer for non-citizens)
  - Application process can start after Census project approval
  - Non-citizens must reside in the U.S. for 3 of last 5 years to qualify

# Phases of an RDC Project

- Proposal drafting
- Submit proposal for review
- Security clearance (SSS)
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census



# Phases of an RDC Project

- Proposal drafting
- Submit proposal for review
- Security clearance (SSS)
- **RDC analysis**
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census

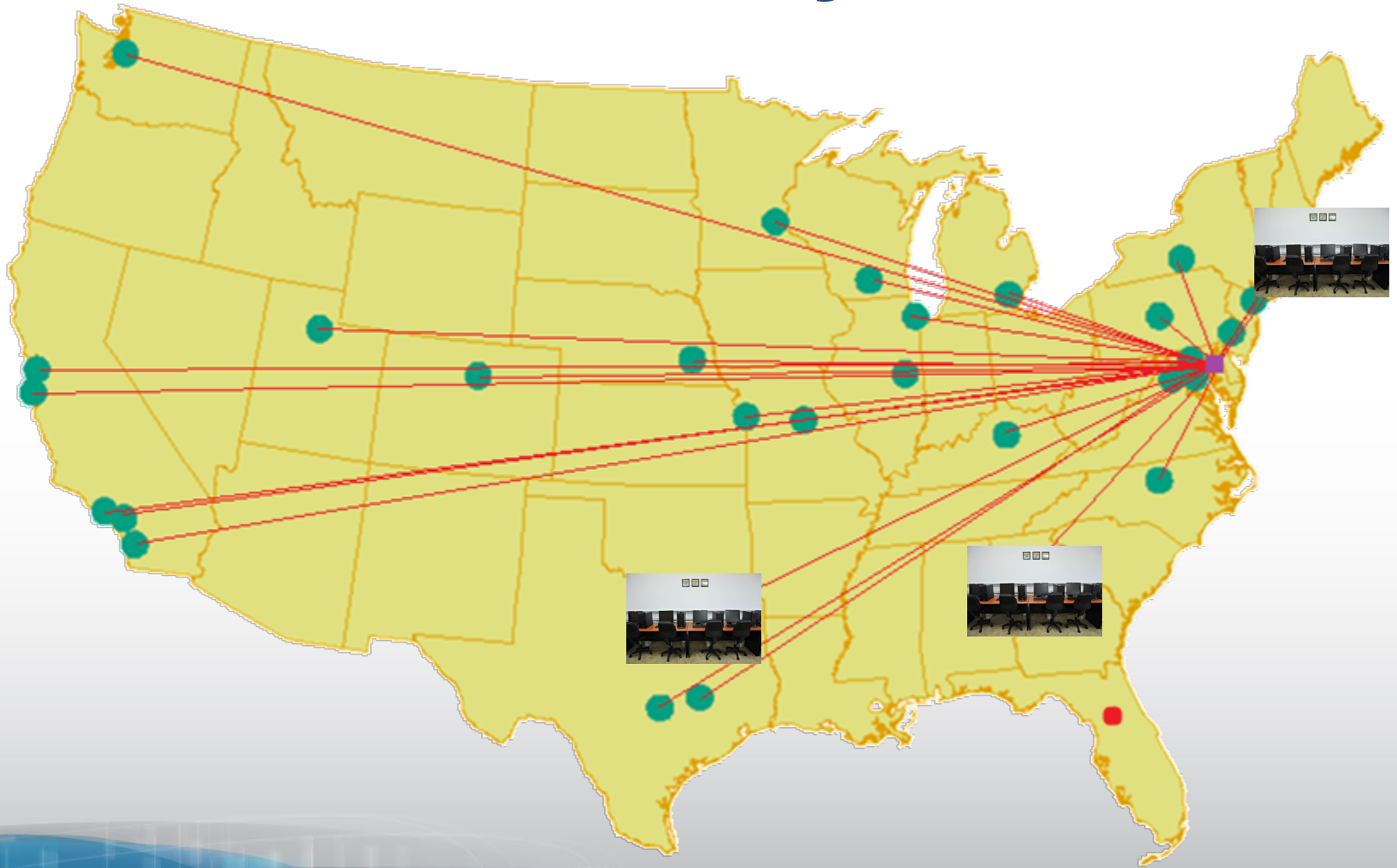
# RDC Analysis

- All work involving restricted-use data will occur in the RDC
- SAS, Stata, R, python, GeoDA, etc. available on server
- Data Staff at HQ uploads user provided data/code to project space
- Large data files, takes time to organize
- Projects lasts from 3-5 years



**\*Not the actual lab**

# RDC Analysis



# Virtual RDC Analysis

- Title 13 data only (non-FTI data)
- Access the RDC environment from home
  - Start project in the physical RDC
  - Additional VRDC orientation
  - Work the remainder of project from home
- Simulate the security requirements of the RDC

# Phases of an RDC Project

- Proposal drafting
- Submit proposal for review
- Security clearance (SSS)
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census

# Phases of an RDC Project

- Proposal drafting
- Submit proposal for review
- Security clearance (SSS)
- RDC analysis
- **Statistical output for Disclosure Avoidance Review**
- Provide working paper and benefit to Census

# Disclosure Avoidance Review

- All output from an RDC project must go through disclosure avoidance review
  - Involved process assisted by RDC Admin
- Disclosure requirements different across data
- HQ Disclosure Office will email you approved output

# Phases of an RDC Project

- Proposal drafting
- Submit proposal for review
- Security clearance (SSS)
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census



# Phases of an RDC Project

- Proposal drafting
- Submit proposal for review
- Security clearance (SSS)
- RDC analysis
- Statistical output for Disclosure Avoidance Review
- Provide working paper and benefit to Census

# 13 Criteria for Benefits

1. Evaluating practices, continued relevance
2. Demog., econ. processes that affect Census programs
3. Increasing utility of data for analyzing public programs, public policy, and/or demographic, economic, or social conditions
4. Conducting data collection, processing or dissemination
- 5. Understanding/improving the quality of data**
6. Improved methodology to collect, measure, or tabulate
- 7. Enhancing the data collected (e.g. imputation, links)**
8. Limitations/improving the Business Register
9. New data collection needs
10. Constructing, verifying, improving sampling frames
- 11. Estimates/characteristics of population**
12. Methodology for estimating non-response
13. Developing statistical weights for a survey

Working paper



# Timeline

- Census review, (1-3 months)
  - Assign to reviewers
  - Reviewing committee
  - (if needed) IRS review (3-7 additional months)
- After approval, SSS Packet
- Security clearance (2-3 months)
  - Additional time for non-citizens
  
- Project work from 3-5 years

# Why Submit an RDC Proposal?

- Data availability
- Data advantages
- High level research

# Why Submit an RDC Proposal?

- Data availability
- Data advantages
- High level research

# Data Availability

- Census Bureau Data
  - Economic Data
    - establishment or firm level
  - Demographic Data
    - household or individual level
  - Combined Employer/Employee Data
    - Longitudinal Employer-Household Dynamics (LEHD)
  - UMETRICs Data
    - Criminal Justice Administrative Records System (CJARS)

---

- Agency for Healthcare Research and Quality (AHRQ)
- National Center for Health Statistics (NCHS)
- Bureau of Labor Statistics
  - NLS, NCS, SOII, CFOI
- Bureau of Economic Analysis

# Economic Data

Economic Censuses	Years
Auxiliaries (AUX)	1977-2012
Construction Industries (CCN)	1972-2017
Fin., Insurance, Real Estate (CFI)	1992-2017
Manufacturers (CMF)	1963-2017
Mining (CMI)	1987-2017
Retail Trade (CRT)	1977-2017
Services (CSR)	1977-2017
Trans., Comms., Utilities (CUT)	1987-2017
Wholesale Trade (CWH)	1977-2017

Establishment Surveys	Years
Annual Survey of Manufacturers (ASM)	1973-2019
Current Industrial Reports (CIR)	1993-2011
Manufacturing Energy Consumption Survey (MECS)	1985-2018
Medical Expenditure Panel Survey – Insurance Component (MEPS-IC)	1996-2020
National Employer Survey (NES)	1994-1998
Plant Capacity Utilization (PCU)	1974-2018
Survey of Manufacturing Technology (SMT)	1988-1993
Survey of Pollution Abatement Costs and Expenditures (PACE)	1979-2005

Business Register Data	Years
Compustat-SSEL Bridge (CSB)	1976-2016
Form 5500 Bridge File	1992-2018
Integrated Longitudinal Business Database (ILBD)	1977-2018
Longitudinal Business Database (LBD)	1976-2019
Ownership Change Database (OCD)	1963-2002
Standard Statistical Establishment List/ Business Register (SSEL/BR)	1974-2019

Firm Surveys	Years
Annual Business Survey	2017
Annual Capital Expenditures Survey (ACES)	1993-2019
Annual Retail Trade Survey (ARTS)	1999-2018
Business Expenditures Survey (BES)	1992-2007
Business R&D and Innovation Survey (BRDIS)	2008-2018
Enterprise Summary Report (ESR)	1977-1992
Exporter Database (EDB)	1992-2010
Quarterly Financial Report (QFR)	1977-2019
Service Annual Survey (SAS)	1991-2019
Survey of Business Owners (SBO)	1982-2012
Survey of Industrial R&D (SIRD)	1972-2007

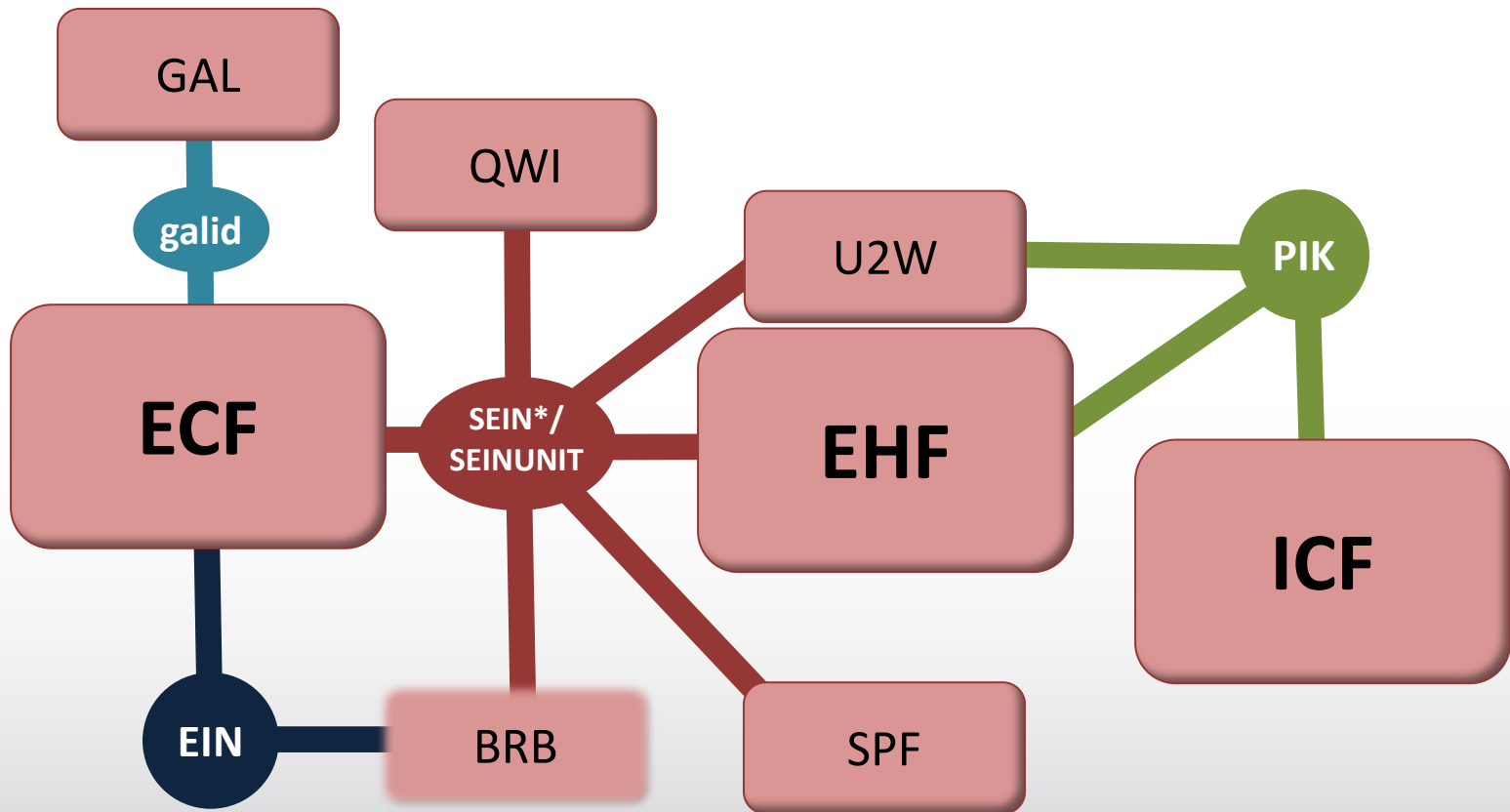
Transactions Data	Years
Commodity Flow Survey (CFS)	1993-2017
Foreign Trade Data - Export (EXP)	1992-2019
Foreign Trade Data - Import (IMP)	1992-2019
Longitudinal Foreign Trade Transactions Data (LFTTD)	1992-2019

# Demographic Data

Data Set	RDC Years
American Community Survey (ACS)	1996-2020
American Housing Survey (AHS)	1984-2019
Current Population Survey (CPS) – March Supplement	1967-2020
CPS Supplements	(1997-2017)
Decennial Census	1950-2010
National Crime Victimization Survey (NCVS)	2006-2017
NCVS School Crime, Police Public Contact Supplements	(2009-2017)
National Longitudinal Survey (NLS) – Young/Mature Men/Women	1966-1999
National Longitudinal Mortality Study (NLMS)	1973-2011
Survey of Income and Program Participation (SIPP)	1984-2020
Census Numident	Present



# LEHD System



\*Majority are matched using just the SEIN and not the SEIN+SEINUNIT pairing

# Data from Partnering Agencies Accessible in the RDC

- Agency for Healthcare Research and Quality
- National Center for Health Statistics
- Bureau of Labor Statistics
- Bureau of Economic Analysis
- Criminal Justice Administrative Records System
  
- Proposal process is with individual agencies and not Census
  - Security background check still with Census

# AHRQ Data Sets

- Medical Expenditure Panel Survey – Household Component (MEPS-HC)
- Household Component-Insurance Component linked file (1996-1999, 2001)
- Nursing Home Component (1996)
- Medical Provider Component (except directly identifiable data)
- Two-Year, Two-Panel Files
- Area Resource File (county-level data that can be linked to MEPS-HC)

# NCHS Data Sets

- **National Health and Nutrition Examination Survey (NHANES)**
- **National Health Care Surveys**
  - National Ambulatory Medical Care Survey (NAMCS) and National Hospital Ambulatory Medical Care Survey (NHAMCS)
  - National Hospital Discharge Survey (NHDS)
  - National Nursing Home Survey (NNHS) and National Nursing Assistant Survey (NNAS)
  - National Home and Hospice Care Survey (NHHCS) and National Home Health Aide Survey (NHHAS)
- **National Health Interview Survey (NHIS)**
- **National Survey of Family Growth (NSFG)**
- **State and Local Area Integrated Telephone Survey (SLAITS)**
- **NCHS Data Linkage Activities**
  - Linked Mortality Data Products
  - Linked Medicare Enrollment and Claims Files Data
  - Linked Social Security Benefit History Data
  - Linked Air Quality Data
- **National Vital Statistics System (NVSS)**

# Why Submit an RDC Proposal?

- Data availability
- Data advantages
- High level research

# Why Submit an RDC Proposal?

- Data availability
- Data advantages
- High level research

# RDC Economic Data Advantages

- No publicly-available microdata
  - Internal data at establishment and firm level
  - Universal scope
  - Detailed industry and geography
  - Privately collected business data is costly to access
- Linking Data
  - Consistent identifiers
  - Business register
    - External data

# Research Example

- Greenstone, Hornbeck, Moretti (2010), “Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings,” Journal of Political Economy
  - RDC microdata
    - Census of Manufactures (CMF)
    - Annual Survey of Manufactures (ASM)
    - Business Register (SSEL/BR)
  - Winning and losing counties have similar trends in incumbents’ TFP prior to a large new plant opening.
  - Five years after the opening, incumbent plants’ TFP is 12 percent higher in winning counties.



# RDC Demographic Data Advantages

- Individual and household data
- Restricted use RDC data
  - more detailed geography
  - less top-coding/censoring
  - additional variables
    - Date of birth
  - larger sample
  - link across other demographic data
    - Longitudinal linking

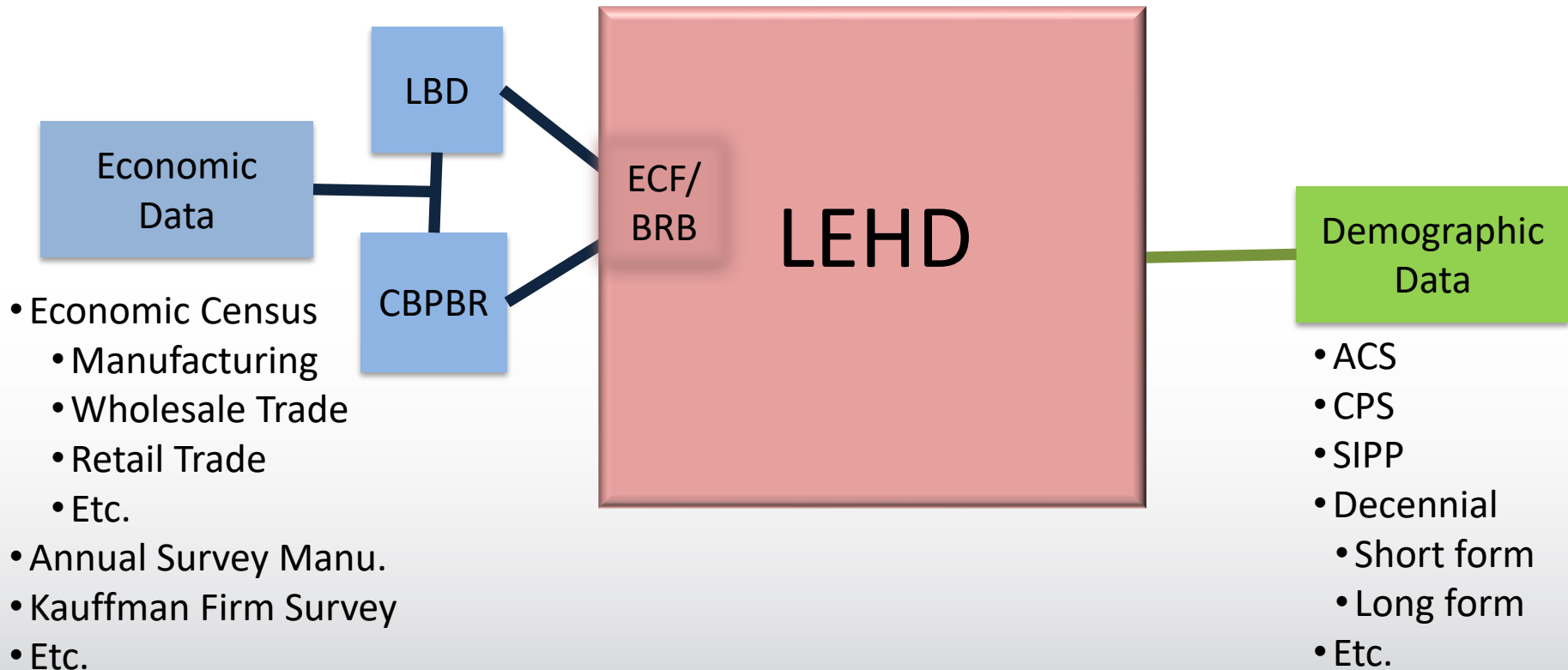
# Research Example

- Angrist and Chen (2011) “Schooling and Vietnam-Era GI Bill: Evidence from the Draft Lottery,” AEJ: Applied Economics
  - RDC microdata
    - Decennial Censuses
    - BLS – National Longitudinal Survey
  - Use census birth date to link to Vietnam draft lottery dates to use as an instrument for military service.
  - Long run consequence of Vietnam service were found to be schooling gains and no earnings losses.

# More Detailed Geography

Data Set	Geography
American Housing Survey (AHS)	Tract
American Community Survey (ACS)	Block
Current Population Survey (CPS) – March Supplement	Tract
Decennial Census	Block
National Crime Victimization Survey (NCVS)	Block
National Longitudinal Survey (NLS) – Young/Mature Men/Women	Tract
National Longitudinal Mortality Study (NLMS)	County
Survey of Income and Program Participation (SIPP)	Tract/County
Master Address File Extract/Auxiliary Reference File	Address

# Longitudinal Employer-Household Dynamics (LEHD)



# Agency for Healthcare Research and Quality (AHRQ)

- Fully specified ICD-9 medical condition codes
- Fully specified industry and occupation codes
- Lower levels of Geography
  - State and county FIPS codes
  - Census tract and block-group codes
- Non-public use data elements
  - Asset information
  - Imputed NDC codes
- Federal and state marginal tax rates

# National Center for Health Statistics (NCHS)

## Types of Restricted Variables

- Geographic Variables
- Linked Data Products
- Genetic Variables (NHANES phenotype data)
- Temporal Variables
  - e.g. dates of birth, death, exams
- Detailed Race/Ethnicity Variables
- Sensitive Variables
  - e.g. youth sexual behavior and mental health
- Long Term Health Care Survey Merging Variables

# Why Submit an RDC Proposal?

- Data availability
- Data advantages
- High level research

# Why Submit an RDC Proposal?

- Data availability
- Data advantages
- High level research



# Publications by RDC Researchers and CES Staff: 2018 and Forthcoming

Econ Journals (by rank)	RDC		Total
	Researchers	CES staff	
AAA (1-5)	12	2	14
AA (6-20)	14	7	21
A (21-102)	22	15	37
B (103-258)	12	11	23
C (259-562)	5	6	11
D (562-1,202)	0	0	0
<hr/>			
<b>Journals outside of Econ</b>	14	11	25
<b>Book Chapters</b>	3	11	14
<b>TOTAL</b>	82	63	145

# How to get started?

[samuel.r.bondurant@census.gov](mailto:samuel.r.bondurant@census.gov)

[wenhua.di@dal.frb.gov](mailto:wenhua.di@dal.frb.gov)

- Send us a quick proposal pitch
  - Send me simple data questions
- Request a meeting in person
  - Include an itinerary if possible
- [www.census.gov/fsrdc/about](http://www.census.gov/fsrdc/about)
- [www.dfwrdc.org](http://www.dfwrdc.org)