California Military Economic Impact Study Process Guide

Britnee Pannell & Sumeet Bedi

2022-01-19

Contents

Chapter 1

Introduction

Introduce the project Add blurb about the history of the project and how it has evolved since then.

- Who is funding project
- Where is the focus of the project?
- When does the project take place?
- How does the project get accomplished? (rough overview)

Why are we doing this project? What does it hope to answer? Specifically mention this documentation is to allow other areas/states to have a path to follow if they wish to duplicate this work.

This also serves to document and justify our conclusions in the main reports if anyone wants to 'check our work.'

Chapter 2

Requirements for Project

The primary requirements to recreate California's study or perform studies in additional areas are federal spending data and the software used to obtain and process that data. Fortunately all software and most Federal data is free and available online.

2.1 Software Requirements

1. The R coding language from cloud.r-project.org. This language is used to obtain and process data.

2. RStudio Desktop from rstudio.com. RStudio is the integrated development environment (IDE) used to run R scripts and develop code.

3. Git from git-scm.com. A comprehensive guide to installing Git is available at happygitwithr.com. Git allows version control of edits across a multi-person team of researchers.

4. If a Github account is needed, one can sign up and register at https://github.com/. Individual free plans are available, as well as free upgrades for qualifying academic purposes.Github is used to develop and host this project.

2.2 Data Requirements

Data used to complete this study was of three varieties: employment and spending data for The Department of Defense (DOD), The Department of Homeland Security (DHS), The Department of Veteran's Affairs (VA), and the Department

of Energy (DOE), additional data obtained by submitting FOIA requests, and data sheets required for processing employment and spending data for upload into IMPLAN.

- Employment Data
- Obtainable With Code Unfortunately, there is no simple way to obtain employment data with code.
- Manual Data Retrieval Employment data can be obtained from several sites. There is no guarantee that these websites will exist in this form indefinitely. Care will be taken to keep this document as up to date as possible.
- Department of Defense Employment(DOD): Civilian employment from FedScope, and military employment from DMDC.
- Department of Homeland Security Employment(DHS): Civilian employment from FedScope.
- Department of Veterans Affairs(VA): Civilian employment from FedScope.
- Department of Energy(DOE): Civilian employment from FedScope.

The next section details how to obtain this necessary data.

Chapter 3

How to Obtain Necessary Data

Basically an additional chapter break to specify how to get each data type after defining categories in the "requirements" section

3.0.1 Obtainable With Code

Unfortunately, there is no simple way to obtain employment data with code.

3.0.2 Manual Data Retrieval

Employment data can be obtained from several sites. There is no guarantee that these websites will exist in this form indefinitely. Care will be taken to keep this document as up to date as possible.

- Department of Defense Employment(DOD): Civilian employment from FedScope, and military employment from DMDC.
- Department of Homeland Security Employment(DHS): Civilian employment from FedScope.
- Department of Veterans Affairs(VA): Civilian employment from FedScope.
- Department of Energy(DOE): Civilian employment from FedScope.

3.1 Spending Data

3.1.1 Code to Obtain Spending Data

3.1.2 Spending Data Obtained via FOIA

3.2 Raw Data Provided by

Include details on Data provided in the "data/raw" folder in the code repo- and the justifications on why it was included and not others

Here is where things get a little annoying - Each file used, with an explanation of where to get it and how to navigate the sites used for obtaining data - What information each file provides - Detailed information on how to make a file custom to specific data needs (where applicable) - Any notes on how files may differ according to region and individual project goals

Chapter 4

Methods

The following section details how to use the data and R code provided as well as an explanation of how the code works.

4.1 Process Outline

The over all process for this project is as follows:

- Data was Obtained
- Data was Filtered for relevance.
- Errors in data were for checked and repaired where found.
- Data was formatted for use in IMPLAN.
- Data was run through IMPLAN.
- IMPLAN outputs were graphically displayed and distributed via report.

4.2 Process in Detail

- Obtain data
  - Spending data
    - Grants
    - Contracts
    - SmartPay (FOIA Required)
  - Employment Data
    - Military Personnel
    - Civilian Employment

DMDC- download and parse csv FedScope- Initially have users go to website and save values of interest to a separate csv file to use in IMPLAN later Eventually set up a code to generate these values acording to state and national level based on parsing out the download from the site.

  - NAICS to IMPLAN crosswalks
  - Spreadsheets provided to aid processing and format outputs

- Process data

  - Clean contracts and grant data-
  - Clean spending data
  - Error check contract spending data

  Will need to go into detail about changes in the code between this year (2021) and subsequent years

  More detailed mention of how the error checking of the USASpending.gov contract data is needed A detailed walk through of how to manually check data and use the multiple NAICS to IMPLAN crosswalks to catch data Mention how IMPLAN automatically removes any codes having to do with construction so those have to be manually coded

  Some errors occur due to the transaction not being given a NAICS code, those need to be manually fixed

  Issues occur with NAICS codes that apply to multiple IMPLAN codes- give detailed explanation of how this was worked around and data was processed and added back to the main cleaned data.

- Run Data Through IMPLAN At this point we stop giving details to users about subsequent processes- we are not responsible for teaching users how to use IMPLAN. We should go over the general steps in what we did next to process data from IMPLAN, and how it was displayed graphically to educate the customers and summarize results for easier understanding

Chapter 5

Using IMPLAN

Place holder for section on how to enter the output files into implan and what IMPLAN analysis was run so that our study can be repeated.

Should not go into too much detail, as full instructions on how IMPLAN works is outside of the scope of this process guide.

Chapter 6

Conclusion/ Discussion

- Importance of modern techniques to get more efficient data analysis in a timely fashion
- Other closing remarks

Pitfalls, how re factoring code caught some of them. How this process will result in a more robust study over subsequent years

Government spending data is very difficult to obtain and there is not a lot of good documentation to help lay people use this data Hopefully we provide some guidelines and aid in discovering and processing this data so that quality studies can come about and Government spending can become more transparent.

Feel good hopeful stuff next.

Chapter 7

What's Next?

A section on where we hope to add and develop this process. Potentially the section to outline changes to the code we have already made for the upcoming 2021 report.

Chapter 8

License your GitBook

In the spirit of Open Science, it is good to think about making your course materials Open Source. That means that other people can use them. In principle, if you publish materials online without license information, you hold the copyright to those materials. If you want them to be Open Source, you must include a license. It is not always obvious what license to choose.

The Creative Commons licenses are typically suitable for course materials. This GitBook, for example, is licensed under CC-BY 4.0. That means you can use and remix it as you like, but you must credit the original source.

If your project is more focused on software or source code, consider using the GNU GPL v3 license instead.

You can find more information about the Creative Commons Licenses here. Specific licenses that might be useful are:

- CC0 ("No Rights Reserved"), everybody can do what they want with your work.
- CC-BY 4.0 ("Attribution"), everybody can do what they want with your work, but they must credit you. Note that this license may not be suitable for software or source code!

For compatibility between CC and GNU licenses, see this FAQ.