

# California Military Economic Impact Study Process Guide

Britnee Pannell & Sumeet Bedi

2022-01-21



# Contents

<b>1</b>	<b>Introduction</b>	<b>5</b>
<b>2</b>	<b>Software Requirements</b>	<b>7</b>
2.1	Download Necessary R Packages . . . . .	7
<b>3</b>	<b>Data Requirements</b>	<b>9</b>
3.1	Employment Data . . . . .	9
3.2	USASpending Data . . . . .	10
3.3	Additional Data . . . . .	10
<b>4</b>	<b>Setting Up the Repository</b>	<b>13</b>
4.1	Obtain Repository . . . . .	13
<b>5</b>	<b>Data Dictionary for Repository</b>	<b>15</b>
<b>6</b>	<b>Methods</b>	<b>17</b>
6.1	Parameters File Instructions . . . . .	17
6.2	Make User Specific Data Files . . . . .	17
6.3	The Code . . . . .	17
<b>7</b>	<b>Using IMPLAN</b>	<b>19</b>
<b>8</b>	<b>Conclusion/ Discussion</b>	<b>21</b>
<b>9</b>	<b>What's Next?</b>	<b>23</b>
<b>10</b>	<b>License your GitBook</b>	<b>25</b>



# Chapter 1

## Introduction

The California Military Economic Impact Study report series began in 2018, when the California Governor’s Office of Planning & Research and the Governor’s Military Council requested the California Research Bureau (CRB, a unit of the California State Library) to conduct this study. This report provides detailed statewide and localized economic impacts of federal national security activity in the state of California. The federal agencies identified as relevant to national security includes the Departments of Defense (DOD), Homeland Security (DHS), Veterans Affairs (VA), and specified sub-agencies of The Department of Energy (DOE). The type of economic activity detailed includes spending (contracts, grants, veterans’ benefits, and SmartPay charge card) and employment data (civilian and military). The first report, published in August 2018, utilized federal fiscal year 2016 data, while the second report, published in December 2019, used fiscal year 2018 data.

Following these 2 reports, the Office of Planning & Research and the CRB secured grant funding from the DOD to support two full-time equivalent (FTE) positions to develop supplemental reports that help localize the economic impacts detailed in the statewide report. In December 2020, the CRB produced the 2020 California Military Economic Impact Study, and followed this third edition of the statewide report with 2 first-time supplements that discuss the economic impacts in every county and congressional district in California. In December 2021, the Research Bureau completed the fourth version of the statewide report as well as the second edition of the county and congressional district supplements. After completion of the second round of supplements, the CRB was tasked with producing this process guide document.

This process guide and supporting documentation were developed in order to allow other states to replicate the methodology of this study for their respective geography of interest. Additionally, this documentation serves to provide the rationale behind how the data was gathered, wrangled, and analyzed in order to justify the conclusions in our main reports.



## Chapter 2

# Software Requirements

To recreate California's study or to perform studies in additional areas, specialized software is needed to obtain and process Federal data. Fortunately all software is free and available online. (Where possible, make sure to use the most recent and fully updated versions of the software).

1. The R coding language from [cloud.r-project.org](https://cloud.r-project.org). This language is used to obtain and process data.
2. RStudio Desktop from [rstudio.com](https://rstudio.com). RStudio is the integrated development environment (IDE) used to run R scripts and develop code.
3. Git from [git-scm.com](https://git-scm.com). A comprehensive guide to installing Git is available at [happygitwithr.com](https://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.

### 2.1 Download Necessary R Packages

`dplyr`, `httr`, `jsonlite`, `openxlsx`, `readxl`, `tidyverse` (double check that our code actually uses all of these packages)

The next section details data necessary to complete this project as well as how to obtain it.





## Chapter 3

# Data Requirements

This section details how to obtain each data type, whose categories are defined in the “requirements” section. The data is broken out into 3 types: employment, USAspending, and all other additional data. There is no guarantee that the data sources described below will exist in this form indefinitely, so care will be taken to keep this document as up to date as possible.

### 3.1 Employment Data

This report focused on 2 types of employment data: civilian and military. These 2 employment data types were found on 2 separate websites:

**Civilian employment** was obtained from the Office of Personnel Management’s (OPM) FedScope website, which provides federal workforce data. From this link, one can click on “Employment” under the “Status Data” bullet point to access quarterly employment data cubes. Given that the report focuses on federal fiscal years, the employment data of interest is of Quarter 3 of a given year (i.e. September). When one clicks on a September cube of their desired year, one can filter the data by agency type (“Cabinet Level”) and location (“United States” and a state of one’s choice) in order to get the civilian employment numbers for the Departments of Defense (which is an aggregate of the Air Force, Army, Defense, and Navy cabinet agencies), Homeland Security, Veterans Affairs, and Energy. This data can be exported out as a PDF, CSV, or Excel file.

**Military employment** was obtained from the Defense Manpower Data Center (DMDC) website, which serves under the Office of the Secretary of Defense. From this link, one can hover over to DoD Data/Reports and select “Statistics & Reports”. On this page, scroll to the “DoD Personnel, Workforce Reports & Publications” text, and click on its hyperlink. This opens up a new window

to access the military employment data. Similar to civilian employment, the data of interest is from September of a given year. One can scroll to the section “Military and Civilian Personnel by Service/Agency by State/Country (Updated Quarterly)”, and download the Excel file for September of their desired year.

## 3.2 USASpending Data

This report utilized USASpending.gov for a large majority of the direct spending done by federal national security agencies in California. The spending types obtained from this site include contracts and grants for DOD, DHS, VA and DOE, as well as direct payments from VA (i.e. veterans’ benefits). From this link, one can hover over to “Download” and click on “Custom Award Data”. On this webpage, one can select the appropriate spending award types (contracts, grants, and/or direct payments), awarding agency (DOD, DHS, VA, DOE), recipient location (United States for “Country”, and a specific state should you want to filter the data further), action date range (a fiscal year “FY 20XX”), and file format. This download process must be done once per agency.

For obtaining USASpending data, we have developed code that automates this entire process and easily captures all USASpending data in one instance. Please refer to our methods section of this process guide for more details.

## 3.3 Additional Data

### 3.3.1 SmartPay Data via FOIA

The remaining portion of direct spending detailed in these reports was SmartPay, a charge card program for federal employees. In order to obtain this data, Freedom of Information Act requests (commonly known as FOIAs) were sent out to the federal national security agencies.

For more information about SmartPay, please visit the General Services Administration’s (GSA) SmartPay website.

### 3.3.2 Data Obtained Online

A variety of data sources needed to be obtained online in order to assist with the data processing for this report. This includes:

- Federal Real Public Property data from the GSA. At this link, the GSA provides an Excel file that breaks down how much federal property is

spread out across the United States. This data is used to help localize the number of statewide DHS and VA civilian employees across counties and districts.

- American Community Survey data from the Census Bureau. At this link, type into the search bar at the center of the webpage “DP03”. This returns a table of “Selected economic characteristics”. From this table, one can filter to the appropriate geography (state, county, congressional district, etc.), year, and such to download into an Excel file. This data is used to help localize the number of statewide military personnel across counties and districts.
- NAICS to IMPLAN crosswalk from IMPLAN. At this link, IMPLAN provides a variety of crosswalks that one can use to relate data from other source to IMPLAN’s 546 industries. On this page, one can go to the second heading titled “2017 NAICS to IMPLAN 546 Industries” and download the Excel file. This crosswalk is utilized to help relate the USAspending contracts data to IMPLAN sectors in order to run that spending data through the IMPLAN software.

### **3.3.3 Data Provided Raw/Self-Made**

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

# Setting Up the Repository

### 4.1 Obtain Repository

Instructions on how to get repo containing code and raw data provided to end users.

Include link to repo and specifically Readme.md file (won't be able to link this until that repo finds its forever home)

Do we want to have a chapter on just setting up the repo and running through each section and file it contains? AKA a data dictionary for the Repo? To insert before the methods section?? So maybe 2 additional sections



## Chapter 5

# Data Dictionary for Repository

This section details each file contained in the code repository and a summary of what it is used for in the project. For a more detailed explanation, please see the “Methods” section. [add link later]





# Chapter 6

## Methods

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

### 6.1 Parameters File Instructions

Include detailed info on Parameters file and how to set it up for a specific project. Default entries are for this study of California. Readme will eventually link to this section

### 6.2 Make User Specific Data Files

Edit framework files in raw data of repo and move into appropriate section of repo.

Here is where you obtain your employment data and make the region specific file to read in those numbers (if we don't just put them in the parameters file?)

### 6.3 The Code

Some words here, maybe a rough over view of the process before getting into the specific details?

Explanation of code - Clear Environment - Load Libraries - Load Parameters File

Load Functions - obtain USAspending - filter USAspending - split USAspending

- 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.

## Chapter 7

# Using IMPLAN

Once the IMPLAN activity sheets for your desired geographies have been produced, it is time to upload them into IMPLAN and allow the tool to run its analysis. When you log into IMPLAN, the homepage should display 3 boxes to navigate through: Regions, Impacts, and Projects.

- **Regions:** Click on Regions to begin your analysis. This should open up a map of the United States. Select the region of interest for the analysis (state, county, congressional district, etc.) and give this project a name. After saving the project, move on to the Impacts tab.
- **Impacts:** In the Impacts tab, look for the button to the right of the “Save” button. Click on that, and select “upload event template”. From there, find the activity sheet for the geography of interest and select that for upload. Ensure that “Industry Change” and “Household Spending” can be selected, and choose those for uploading into IMPLAN. The data points for the respective IMPLAN sectors should then auto-populate on IMPLAN. If your analysis is including SmartPay spending data, be sure to manually add that in as type “Institutional Spending”, specification 11001, and the value amount. Choose to select all events (the button to the left of the “Save” button) and drag them over to the geography’s model on the right side of the IMPLAN screen. After that, select to run the model at the bottom right. Depending on how big the model, IMPLAN will take time to generate the results. Once the model is done with its analysis, click on the “View Results” button.
- **Results:** Once on this page, one can sort through the results of the IMPLAN activity sheet that was ran for a geography and download CSV files of a variety of data points. For this report, the data points of interest included economic indicators like output and full-time employment (FTE) opportunities, both generally and across industries, as well as government and tax revenue generated.

As part of our final step before developing and writing the reports, our team created code to combine all the input and output data from IMPLAN into one big spreadsheet which was used to create the graphs and charts seen throughout the reports.

## Chapter 8

# Conclusion/ Discussion

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.

Utilizing IMPLAN to manage government spending data doesn't come with a lot of instructions

You have to trust that the data entry is accurate, many instances of older NAICS codes exist and this has the potential to introduce errors into the code.

Limited time to process the data due to a new method of processing the data



## Chapter 9

# 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 10

# 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](#).