

USA **FACTS**

Open Source Dataset Research & Analysis Government Finances User Guide

Brandon Cantey, Yuming Huang, Dillon Morley, Dennis Poludnev, Amari Smith

May 5th, 2021

Class: INST490

Instructor: Mary Ann Francis

Client: Mary Ann Francis

Director: Bill Kules

Table of Contents

ABSTRACT	2
INTRODUCTION	3
DATASETS	3
ACCESSING THE DATA	3
VIEWABLE	4
TOOLS & FEATURES	5
DOWNLOADABLE	7
TOOLS & FEATURES	7
TECHNICAL SPECIFICATIONS	8
DATA DEMONSTRATION	9
DOWNLOADING THE DATA	9
CLEANING THE DATA	10
VISUALIZING THE DATA	14
LIMITATIONS OF THE DATA	18

ABSTRACT

The Open Dataset Research & Analysis (ODRA) Project is a semester-long initiative focused on creating documentation on how to use open-source data for research. ODRA's user guide can be utilized by open-source researchers seeking a foundation of how to work with open data effectively and efficiently. This user guide focuses on using nonpartisan organization USAFacts.org for research. However, the methods and processes outlined can be applied when making use of other data sources.

INTRODUCTION

USAFacts.org is a non-profit organization that provides nonpartisan data on the nation's finances, safety & security, economy, and people & society. The organization's top goal is to make sure that facts are provided. According to the organization, they believe that "Democracy is only successful when it's grounded in the truth". That is why USAFacts.org provides metrics on more than 90,000 entities within all US governments, from Federal records to a growing collection of records of State and County data. The organization also created a 10-K model, which is a comprehensive report filed annually by a publicly traded company about its financial performance, which these companies must submit annually to the SEC for transparency and accountability to their investors.

USAFacts.org relies solely on data from over about 70 government agencies. While the data is reliable as it is coming from government agencies, it is possible for some contradiction due to revisions from original sources. Being considered as the largest source for standardized US Government data, USAFacts.org presents data in various forms for efficient and effective analysis.

DATASETS

This guide focuses on data available in the Government Finances segment of USAFacts.org. Government Finances represent what the Government does with US tax dollars. These finances cover how much the government spends and brings in, along with the amount of debt our nation is in. As the guide moves forward, you will learn more about each section, what it holds, how to access the data, and what the data highlights.

ACCESSING THE DATA

Government Finances are divided into many sections. To access the data, enter usafacts.org into your web browser. When the page loads, hover your mouse over the "Data" tab at the top. A drop down will appear, and you will click on the "Government Finances" listing, which is the first list available.

The sections of Government Finances are *Debt*, *Government-Run Business*, *Revenue*, *Spending*, and *Government Employment*. Each section can vary in how it can be accessed. Some of the data can be downloaded, while some can only be viewed.

Nonetheless, the data can be accessed and analyzed by users. This chapter will focus on the accessibility of the data and the tools that can be used when analyzing.

VIEWABLE

Three of the subsections of Government Finances are available in a web-based format. This means that the data cannot be downloaded in a format such as a CSV, but it can be viewed and analyzed on the web page provided by USAFacts.org. The following sections are available in *Viewable* format:

- **Government Employment:** The data given in this section of Government Finances provides quantitative information on various sub-topics that make up Government Employment. In this dataset, you will find data dating from 1980 to 2018 on information such as “Government Employment”, “Compensation of Government Employees (Aggregate)”, “Compensation of Government Employees (Per Person)”, and “Wages of State & Local Employees (Per Person Per Hour)”. This section highlights the number of employees and the amount of money used for their compensation.
- **Revenue:** This section of Government Finances is designed to present data and information on the Government’s Revenue from 1980 to 2018. The main table summarizes overall Revenue between Tax Revenue and Non-Tax Revenue. These two rows can be expanded to break down the numbers even further.

Revenue	1980	1981	1982	1983	1984
+ Tax Revenue	\$726.85b	\$829.13b	\$867.46b	\$869.67b	\$970.68b
+ Non-Tax Revenue	\$42.66b	\$56.37b	\$60.57b	\$73.19b	\$78.35b
Total Revenue	\$769.71b	\$885.49b	\$928.03b	\$942.86b	\$1.05t

- **Spending:** This section of Government Finances displays the spending data of the US Government from 1980 to 2018. The data captures spending towards programs such as Medicaid, Social Security, and unemployment insurance. It also includes money spent on programs that does not affect citizens directly such

Spending By Mission	1991	1992	1993	1994	1995
+ Establish Justice and Ensure Domestic Tranquility	\$136.56b	\$148.01b	\$154.15b	\$164.74b	\$178.82b
+ Provide for the Common Defense	\$321.98b	\$350.11b	\$345.78b	\$338.03b	\$327.9b
+ Promote the General Welfare	\$421.54b	\$415.03b	\$408.44b	\$444.98b	\$448.86b
+ Secure the Blessings of Liberty to Ourselves and...	\$980.62b	\$1.05t	\$1.1t	\$1.14t	\$1.23t

as military spending, foreign aid, and infrastructure spending.

TOOLS & FEATURES

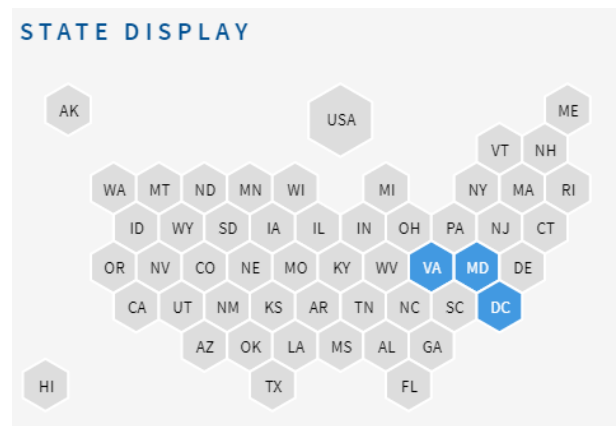
USAfacts.org has provided tools & features that can be used to manipulate the data. They can help with filtering, sorting, expanding, and decreasing the data. Below are of list of the tools available for the Viewable data:

- **Compare By Government Type:** Users can decide at what levels of government do they want to the data. The four options are *All*, *Combined*, *State & Local*, and *Federal*. *All* presents the data at each level explicitly, showing numbers by level. *Combined* presents the data from *Federal and State & Local* as one. Then, the specific levels of *Federal and State & Local* show data based on the selected level.
 - Notice that the image above is an example of when a user selects *All*. If the user was to choose one of the other options, data would be categorized specifically by that compare type.

Federal	2	263,840
State & Local	7	2,605,545
Combined	9	2,869,385

- **State Display :** Users can select which states they want to view data for. They can choose one state, multiple states, or the entire USA.

Total Spending			
District of Columbia	\$5.29b	\$5.86b	\$6.62b
Maryland	\$24.99b	\$26.44b	\$28.34b
Virginia	\$29.34b	\$31.41b	\$34.22b



- **Mission vs Function:** For some of the data, USAfacts.org has a structure that gives the user different viewing options. Users can choose to view the data by “Mission” or by “Function.”

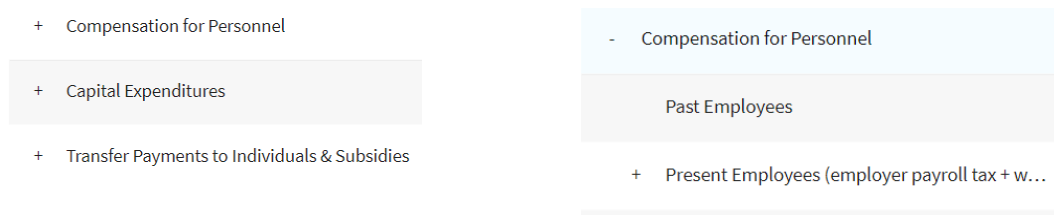
Spending by Mission

Spending by Function

- **Mission:** This option displays data in certain programs that are aligned with four constitutional missions based on the preamble to the Constitution: Establish Justice and Ensure Domestic Tranquility, Provide for the Common Defense,

Promote the General Welfare, and Secure the Blessings of Liberty to Ourselves and Our Prosperity.

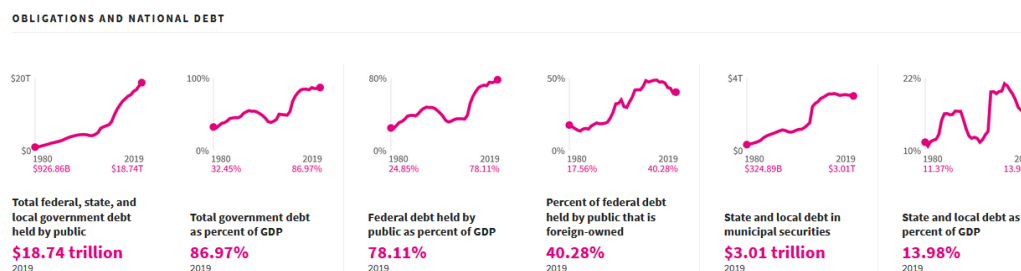
- Function:** This option shows data by functional categories such as compensation for current and past employees, capital expenditures, transfer payments to individuals, interest on the debt, and payments for goods and services.
- Expand (+):** Another feature to help breakdown the data is the expand (+) function. Essentially, if a category of data type has a “+” beside it, that means the category has subcategories, in which the data can further be broken down for analysis.
 - Along with the “+” function, if the user wants to wrap up or decrease the data breakdown, they can then click “-” which will summarize the data again.



DOWNLOADABLE

Two of the sections of Government Finances are downloadable in CSV format. Those sections are Debt and Government-Run Business. These differ from the viewable format because not only can the data be downloaded, but USAFacts.org also provides user-interactive visualizations, along with summary tables for the data.

- Debt:** This section of Government Finances focuses on the nation’s debt between 1980 to 2019. Debt is divided into different subcategories such as Government Debt held by public, Government debt as % of GDP, Federal debt held by public as % of GDP, and more. Each subcategory has a visual displaying the data over time, along with a CSV of the data available to download.



- Government-Run Business:** This section of Government Finances provides data on Government-Run Businesses at the Federal and State & Local levels of Government. At the Federal level, there is business data such as profit from FDIC, US Postal Service, and Export-Import Bank. At the State & Local level, there is data regarding the net profit

from public hospitals, toll highways, liquor stores, water utilities, and lotteries. Federal business data dates as far back as 1926 to as late as 2019. State & Local data is dated as far back as 1980 to as late as 2016.

TOOLS & FEATURES

- **Data Adjustments:** Data can be adjusted before being downloaded as a CSV file. The two adjustment options are Adjust for Inflation and Per Capita Adjustment. Depending on the data, only one option may be available. However, if both are available, you can only choose one at a time.

DATA ADJUSTMENTS

☒ Adjust for Inflation
 ☐ Per Capita Adjustment

Download Data

- **Table:** A table is provided that lists the data by category by year. There is a scrolling bar at the bottom of the table that allows the user to see data across the different years.
 - Notice that the expanding (+) function discussed in the Viewable data section is available here as well.

	1980	1981	1982	1983	1984	1985
- US Postal Service net profit	-\$1.25b	-\$1.43b	-\$154.15m	-\$1.11b	-\$1.24b	-\$1.35b
- Expenditures						
Gross expenditure	\$20.7b	\$22.6b	\$24.1b	\$26.1b	\$2	
+ Revenues						

- **Explorer:** The Explorer tab allows the user to interact with the data and create different types of visuals. Instead of having to download the data and use a visualization software, USAFacts.org provides this tool for users to make various charts directly on the web page.
 - Users can “**Select Chart**” which indicates what kind of chart they would like to view such as bar and trend charts.

Interact with the data

SELECT CHART

SELECT DATA

Public hospitals net profit

☐ Revenues
 ☐ Gross revenue
 ☒ Expenditures

toggle all

DATA ADJUSTMENTS

☐ Adjust for Inflation
 ☐ Per Capita Adjustment

- o “**Select Data**” is available as well, which allows for data to be toggled by the user to analyze different trends instead of all at once.
 - o The “**Data Adjustments**” option is available here also, to view data by the current value of the dollar.
-

TECHNICAL SPECIFICATIONS

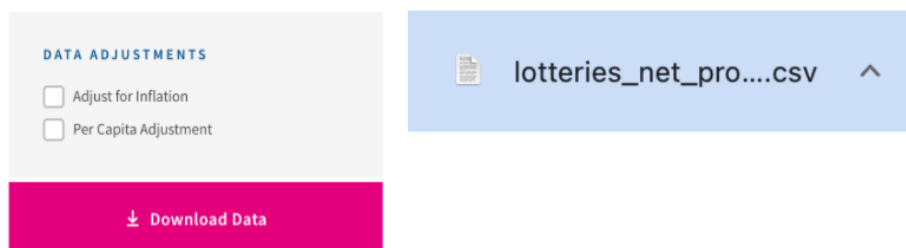
To work with the data, it is best to have a web browser, spreadsheet software, and visualization software available. Through the demonstration below, the following applications were used: Google Chrome, Google Sheets, and Tableau.

DATA DEMONSTRATION

For the purposes of this demonstration, I will be focusing on the State and Local Government-Run Businesses section of Government Finances. For the Downloading and Cleaning the Data section of this user guide, I will be using the “Lotteries Net Profit” dataset as my example. For Visualizing the Data section of this user guide, I will be using a cleaned dataset that is a combination of all State/Local Government-Run Businesses. This dataset along with the other cleaned datasets from USAFacts.org can be downloaded through our group [Kaggle](#).

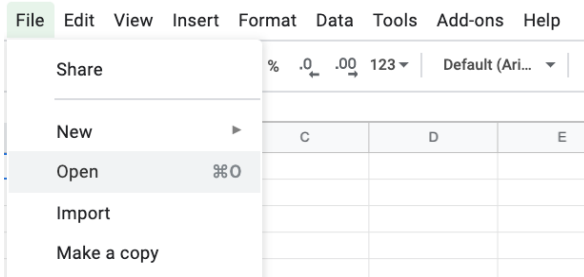
DOWNLOADING THE DATA

To download the data from the Government-Run Business being viewed, click the “**Download Data**” button on the right side of the screen as shown below. The button will turn pink when you hover over it. After clicking the “**Download Data**” button, a CSV file labeled with the Business you are viewing will begin to download.



Importing the Data into Google Sheets (or Microsoft Excel)

Now that the data is downloaded, open Google Sheets and click the “File” button in the top left corner. Then, from the drop down, click “Open” and upload the file from your computer.



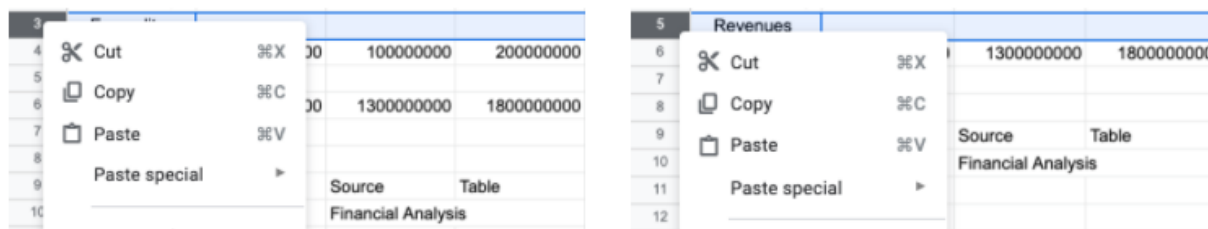
The CSV file has the Years (1980 - 2016) set as the columns in your data frame. The “Lotteries Net Profit” row corresponds to the net profit in USD. For both expenditures and revenues, there is a blank row and then a row that shows their “Gross” value.

	A	B	C	D	E	F	G	H
1	Years	1980	1981	1982	1983	1984	1985	1986
2	Lotteries net profit	1029796000	1216762000	1630330000	2027139000	2684522000	3523411000	4690624000
3	Expenditures							
4	Gross Expenditures	1000000000	1000000000	2000000000	3000000000	3000000000	4000000000	6000000000
5	Revenues							
6	Gross Revenues	1200000000	1300000000	1800000000	2300000000	3000000000	3900000000	5200000000
7								
8	Sources:							
9		Agency	Source	Table	Uri	Direct Link	Updated	Update Schedule
10		USAFacts	Financial Analysis		https://usafacts.org/usafacts-financial-analysis-methodology/			
11	Exported on:	04/19/2021						
12								
13								

CLEANING THE DATA

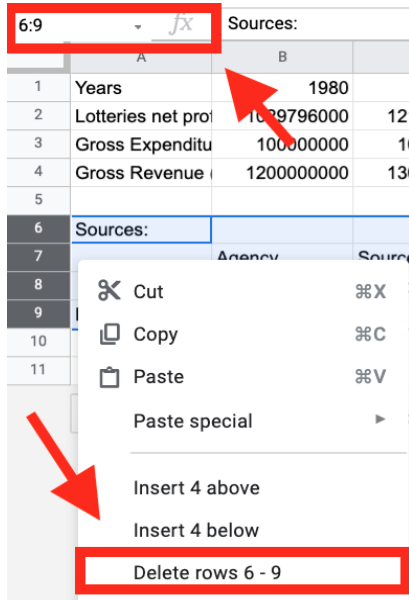
As shown in the previous section, when you upload a CSV file to Google Sheets, there are a few issues with formatting that need to be addressed. To make your data more formal and cleaner, you need to remove the blank rows for Expenditures and Revenues, then transpose your columns and rows to set “Years” as our row values.

Removal of Blank Rows for Expenditures & Revenues: To Remove the blank rows for Expenditures and Revenues, you want to right click on the row index that corresponds to the row you are trying to delete. In this example, you will right click on indices **3** and **5**. After right clicking on the index, you will scroll down to the “Delete Row” button and delete the blank row for Expenditures and Revenues.



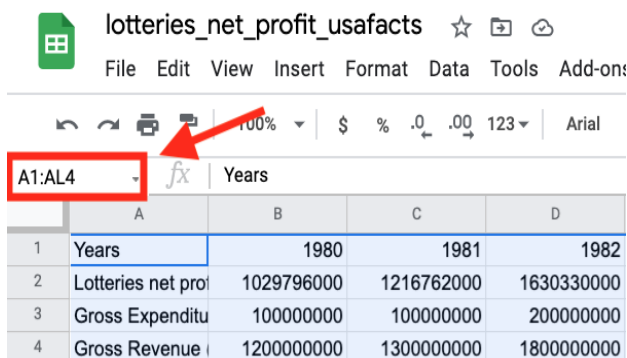
Removal of Sources:

The sources section at the bottom of your dataset will be in the way when you transpose your data, so you want to delete these rows prior to transposing your data. To do this, you must enter your desired cells in the “**Name Box**,” which can be found in the top left of your screen under the “**Undo**” button. In this box, you will enter “**6:9**,” which corresponds to rows six through nine. Next, you want to right click on the highlighted rows and select the ‘**Delete Rows 6-9**’ button.

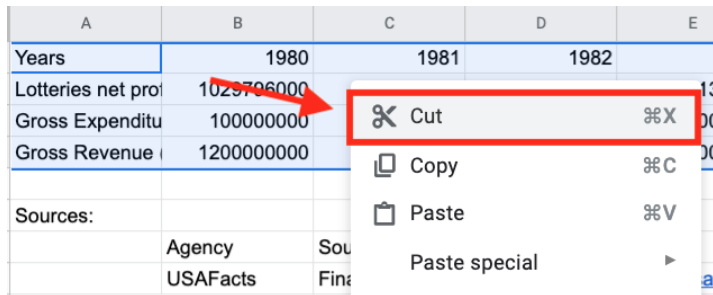


Transpose Column and Row Values:

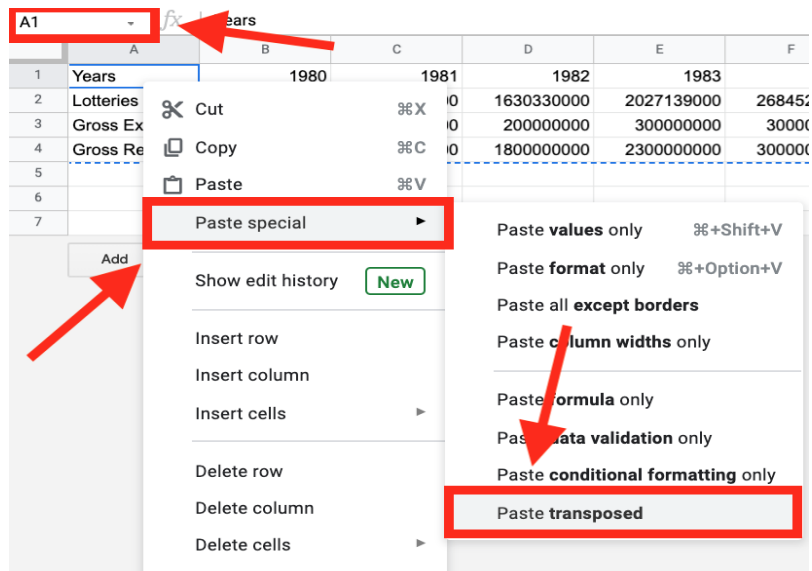
The last step in cleaning your data will be to transpose your rows and columns. This will allow you to have the years as your row values and your net profit, expenditures, and revenues as the column values. The first step is to select all the data in your Google Sheet. To do this, you will enter “**A1:AL4**” into your “**Name Box**” at the top of your sheet, which will select all your desired cells.



Once all your cells are highlighted blue, you will then right click anywhere on your cells and select the **“Cut”** button.

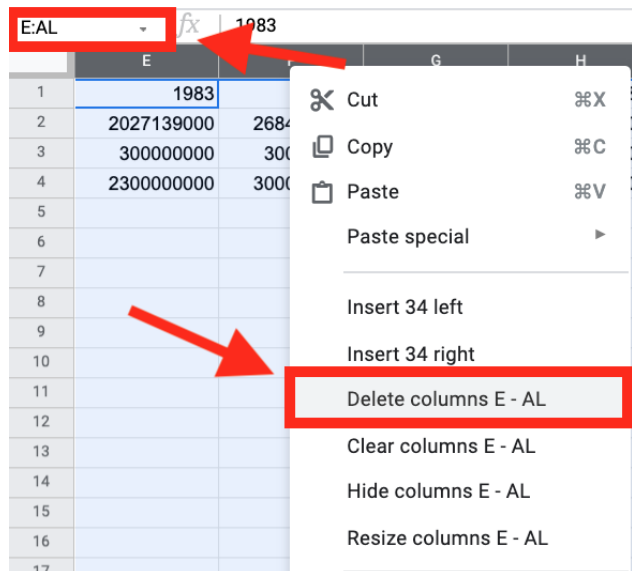


Next, you want to change your **“Name Box”** cell value to **“A1”**, and then right click on the cell labeled **“Years”** and click the **“Paste Special”** button. This button will open up a list of paste options and you will select the **“Paste Transposed”** button at the bottom of the list.

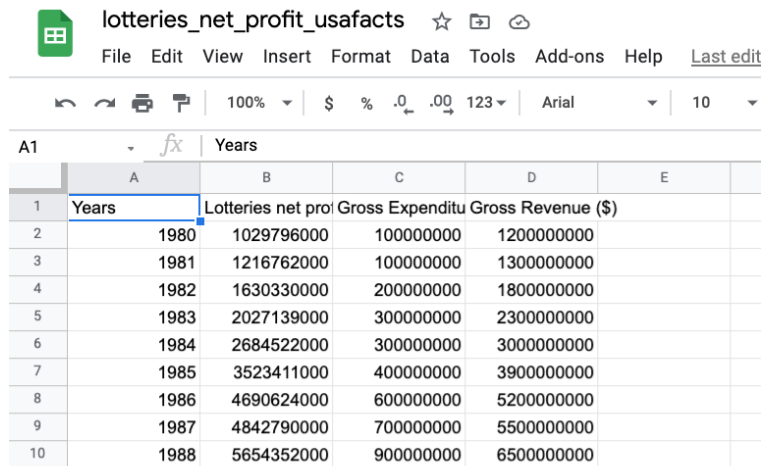


Removal of Extra Rows:

Once you paste your data transposed, it will fill columns A through D. The last step is to remove the rows that were not pasted over. To do this, you need to change the value in your **“Name Box”** to **“E:AL”** to select your desired columns. Next, you will right click on any of the blue highlighted area and select the **“Delete Columns E - AL”** button.



After removing these final columns, you will have your cleaned dataset which should look like the image below. You have “**Years**” as your rows, and “**Net Profit**”, “**Gross Expenditures**”, “**Gross Revenues**” as your columns.

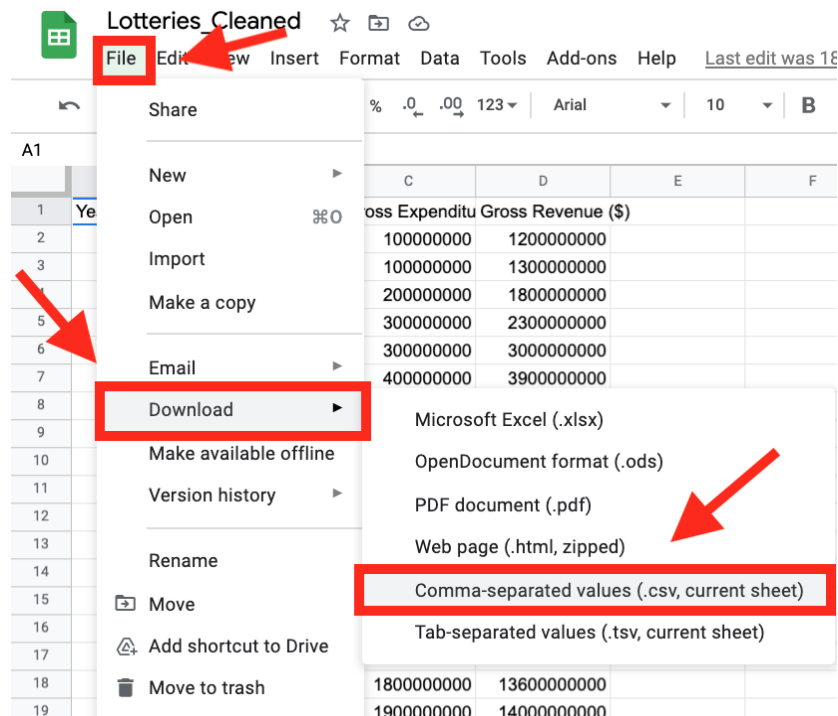


The screenshot shows a Google Sheets interface with the title 'lotteries_net_profit_usafacts'. The menu bar includes File, Edit, View, Insert, Format, Data, Tools, Add-ons, Help, and a 'Last edit' link. The toolbar shows a zoom level of 100%, currency symbols, and font settings (Arial, 10). The spreadsheet has a column header 'Years' in cell A1 and a row header 'Years' in cell A2. The data is organized into columns: 'Years' (A), 'Lotteries net profit' (B), 'Gross Expenditures' (C), and 'Gross Revenue (\$)' (D). The data spans from 1980 to 1988.

	A	B	C	D	E
1	Years	Lotteries net profit	Gross Expenditures	Gross Revenue (\$)	
2	1980	1029796000	100000000	1200000000	
3	1981	1216762000	100000000	1300000000	
4	1982	1630330000	200000000	1800000000	
5	1983	2027139000	300000000	2300000000	
6	1984	2684522000	300000000	3000000000	
7	1985	3523411000	400000000	3900000000	
8	1986	4690624000	600000000	5200000000	
9	1987	4842790000	700000000	5500000000	
10	1988	5654352000	900000000	6500000000	

Saving Cleaned Dataset as CSV File:

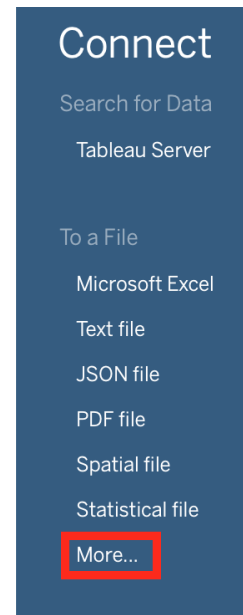
Using the cleaned dataset in Google Sheets or Excel, you first want to save the CSV file to your computer. To do this, you want to click the “**File**” tab at the top of your screen, then select the “**Download**” button, which will give the filetype option for your download. Select the “**Comma-separated values (.csv, current sheet)**” button and your cleaned dataset will download onto your computer.



VISUALIZING THE DATA

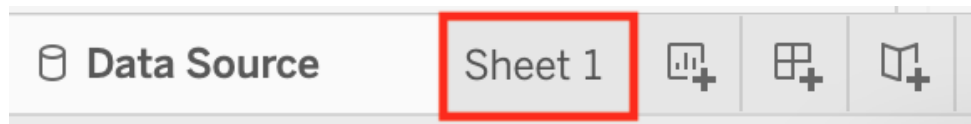
Importing Data:

Using the cleaned dataset from Google Sheets or Microsoft Excel, it is extremely easy for users to create powerful visualizations in Tableau. The first step in creating visualizations is to import our dataset into Tableau. When opening Tableau, you want to connect to a file under the “**Connect**” section on the left side of the screen. Under the “**To a file**” sub-section, you want to click the “**more...**” button and then select your cleaned CSV file from your files.



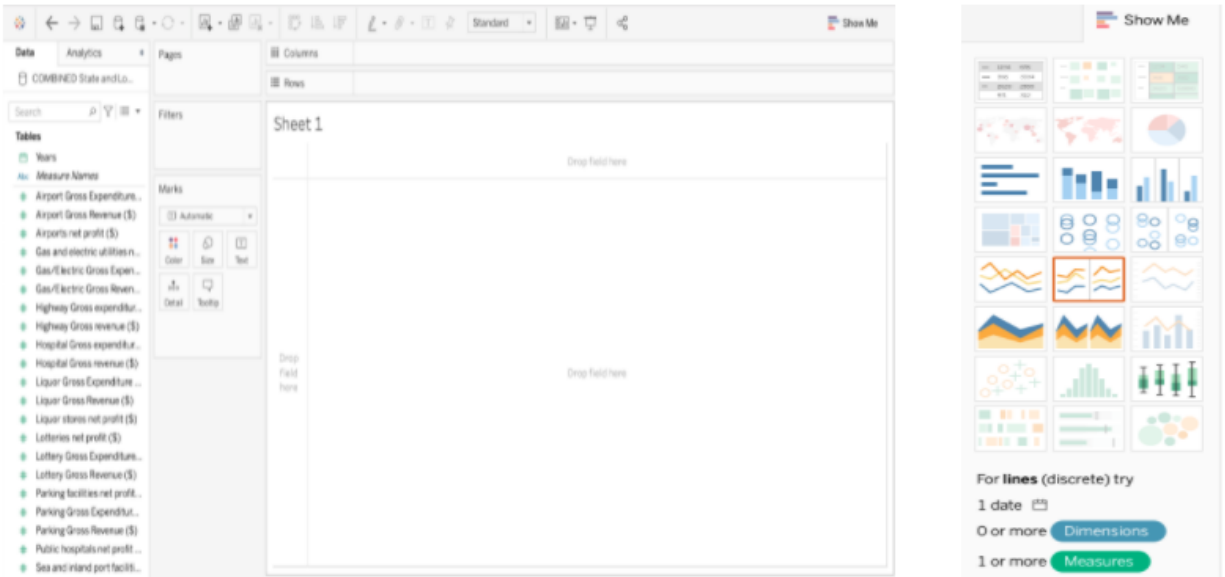
Creating a Sheet:

Upon connecting your cleaned CSV file, you want to click the “**Sheet 1**” tab at the bottom of the screen. This is where you can create your visualizations using your newly imported cleaned CSV file.



Creating a Visualization:

Once you are on your “**Sheet 1**,” you can select the different features you want to include in your visualizations by dragging elements into either the “**Rows**” section or the “**Columns**” section. After adding in your different features into both the “**Rows**” and “**Columns**” section, you can click on the “**Show Me**” button and see which graphs you can create using your data. If a graph is highlighted, you can click on it and see your data visualized as shown below.



The following are examples of visualizations that have been created in Tableau. You can create the same type of visualizations using the various datasets provided in our [Kaggle](#).

Figure 1: Example of a simple table using “Years” as our row feature, and “Measure Names” for our column value.

Columns

Measure Names

Rows

YEAR(Years)

State/Local Government-Run Businesses Net Profit Totals

Year of Yea..	Airports net profit..	Gas and electric ut..	Liquor stores net ..	Lotteries net profit..	Parking facilities n..	Public hospitals n..
1980	-675,230,000	-1,983,966,000	609,147,000	1,029,796,000	-7,897,000	-5,351,753,000
1981	-645,060,000	-2,216,364,000	552,265,000	1,216,762,000	-8,482,000	-4,607,346,000
1982	-567,782,000	-1,708,481,000	501,232,000	1,630,330,000	-12,102,000	-5,574,218,000
1983	-478,150,000	-1,122,151,000	494,696,000	2,027,139,000	-1,164,000	-6,761,092,000
1984	-839,413,000	175,766,000	498,802,000	2,684,522,000	41,939,000	-6,563,867,000
1985	-626,183,000	603,633,000	412,536,000	3,523,411,000	-17,784,000	-7,403,874,000
1986	-802,773,000	-629,495,000	445,981,000	4,690,624,000	-50,588,000	-7,508,709,000
1987	-1,088,876,000	1,226,990,000	422,666,000	4,842,790,000	-76,197,000	-7,616,632,000
1988	-1,241,955,000	2,575,341,000	452,898,000	5,654,352,000	-48,030,000	-9,170,117,000
1989	-1,057,274,000	3,538,754,000	443,669,000	7,059,265,000	41,028,000	-9,304,547,000
1990	-1,310,150,000	3,832,537,000	514,759,000	7,485,308,000	105,195,000	-9,348,662,000
1991	-1,514,959,000	4,764,614,000	566,462,000	7,664,880,000	110,378,000	-9,317,039,000
1992	-2,317,365,000	4,262,831,000	541,830,000	7,844,161,000	108,287,000	-5,197,186,000
1993	-2,641,959,000	5,046,003,000	579,776,000	8,826,523,000	304,198,000	-5,146,026,000
1994	-2,340,211,000	5,189,453,000	621,920,000	9,794,293,000	348,536,000	-2,876,132,000
1995	-734,358,000	5,754,958,000	616,976,000	10,809,971,000	292,648,000	-2,537,463,000
1996	-623,541,000	6,652,537,000	633,520,000	11,811,550,000	232,472,000	-1,829,253,000
1997	-1,072,254,000	6,401,882,000	659,134,000	12,100,219,000	383,541,000	-189,439,000
1998	-936,196,000	6,715,744,000	761,322,000	12,095,577,000	386,689,000	-1,773,144,000
1999	-2,415,190,000	7,625,171,000	698,142,000	12,324,788,000	341,499,000	-2,218,035,000
2000	-2,076,441,000	7,634,397,000	790,982,000	12,156,586,000	468,076,000	-3,952,878,000

Figure 2: Example of a simple bar chart to show which state/local business have an overall negative or positive net profit from 1980 - 2016 using “**Measure Values**” as our column value and “**Measure Names**” as our row value.

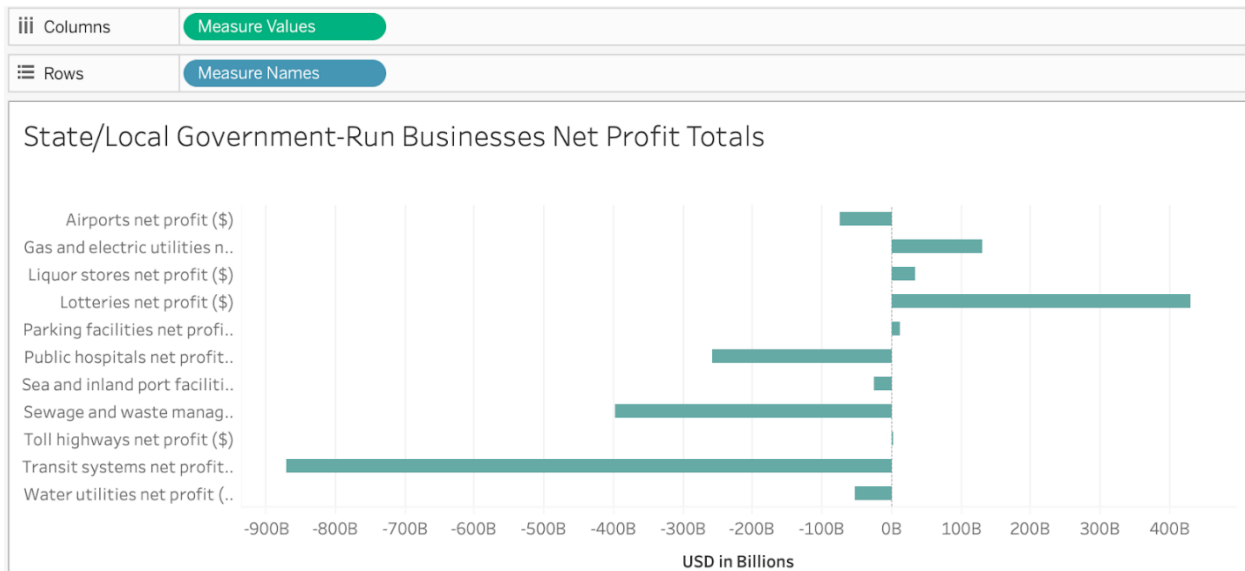


Figure 3: Creating an Interactive Dashboard showing graphs for Net Profit, Expenditures, and Revenue with a filtering option for each graph.

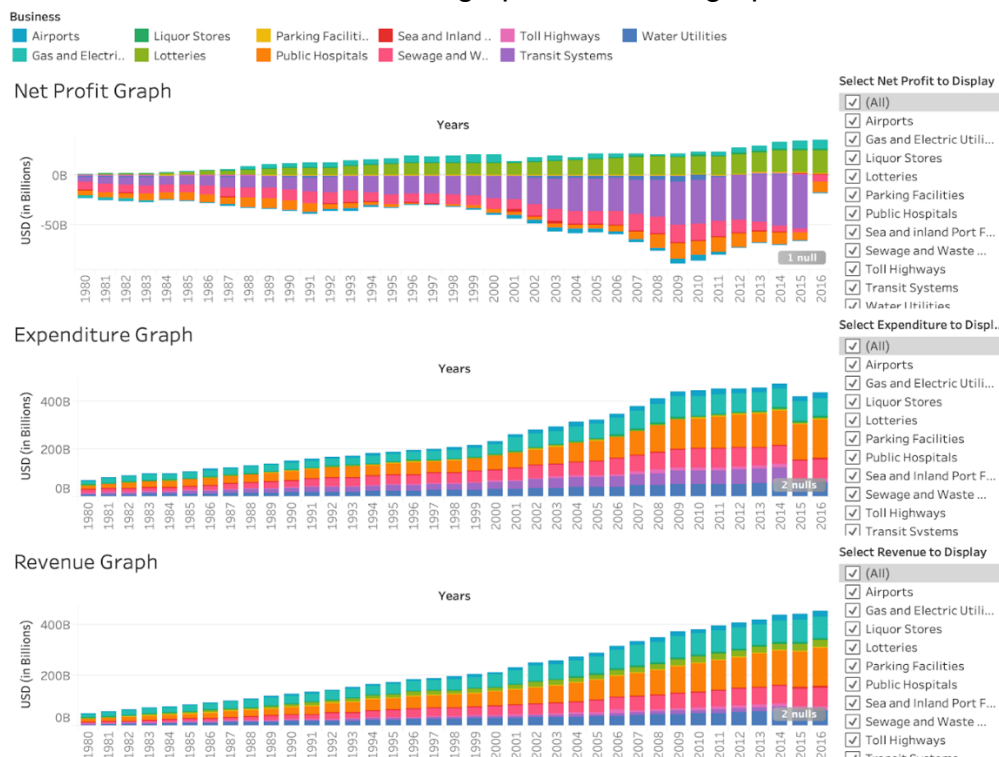


Figure 4: Filtering to view Liquor Stores vs. Lotteries:



LIMITATIONS OF THE DATA

The data provided by USAFacts.org is quite substantial. However, with the current magnitude of data, there are still some limitations that exist amongst the Government Finances data. The major limitation that exists is the inconsistency between accessibility of the data. For the Debt and Government-Run Business section of Government Finances you can access and download the data as a CSV file. However, for Government Employees, Revenue, and Spending sections you can only view the data on their website. The main limitations that exist are within these three sub-sections of Government Finance. Also, when using these three datasets, there was missing data or data that would not load at all. Another issue that occurred when using the data from USAFacts.org was that there were inconsistencies with the dates for each dataset. For example, different government businesses had different years which made it harder to analyze. Due to these issues, the capabilities for this project were limited.