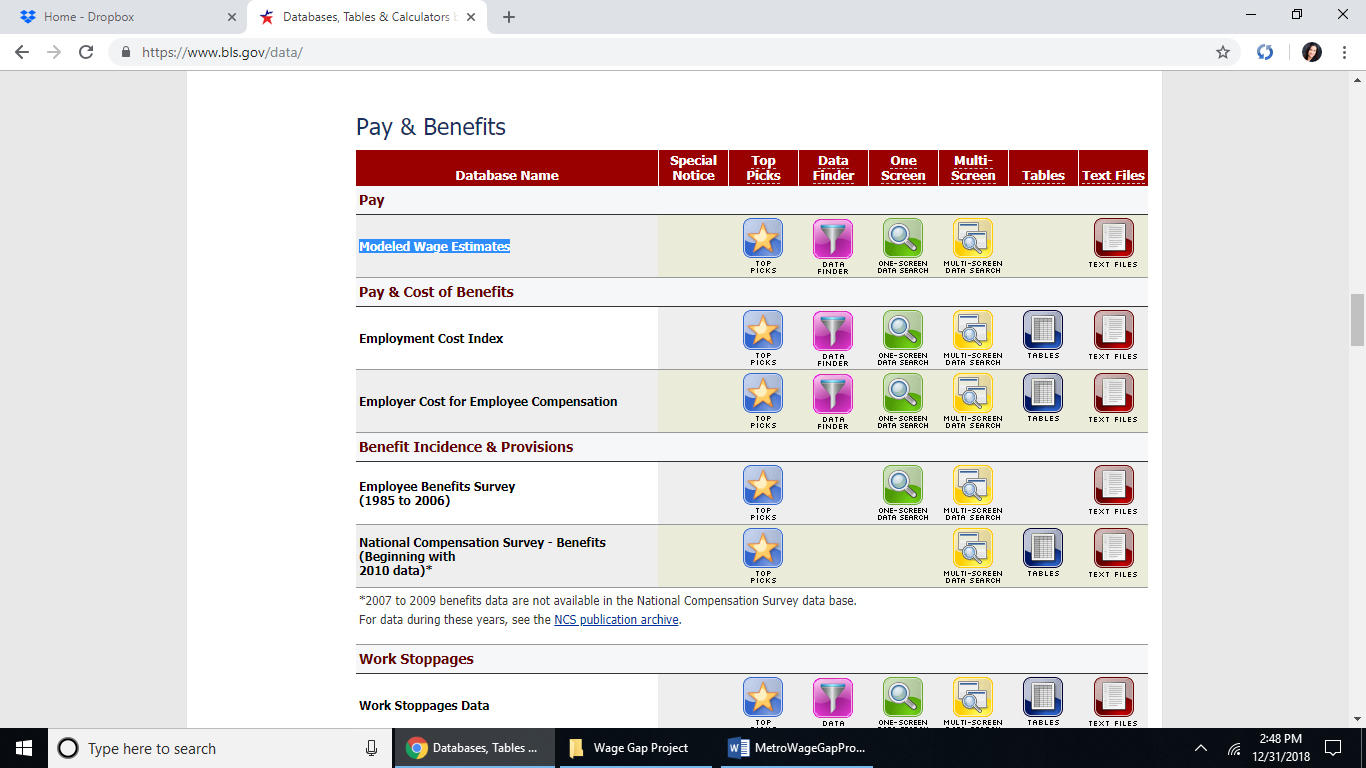
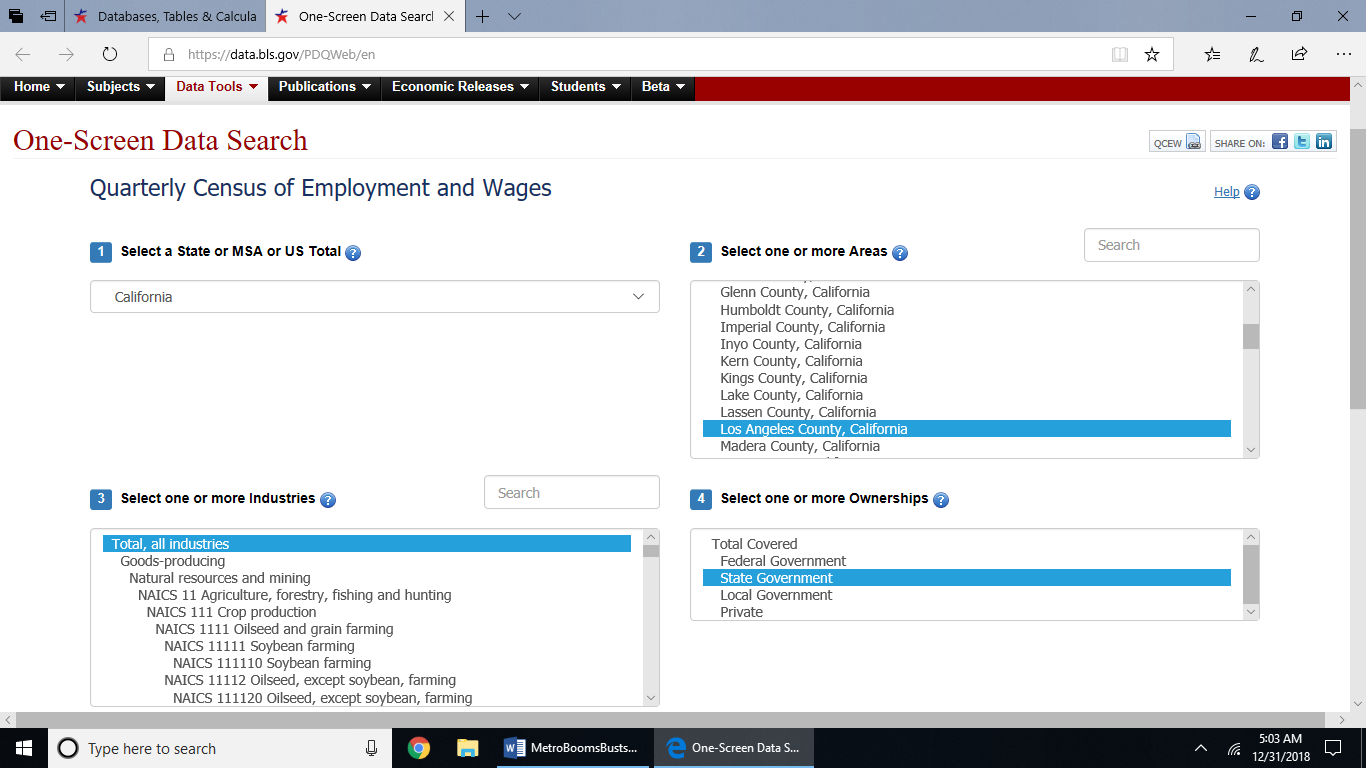
This project examines the top 15 metro regions using data from Zillow and the BLS.

The following are the links accessed 12/31/2018 for the data sets:

<https://www.bls.gov/data/> 

The top 14 metro regions from <https://www.zillow.com/Research/Data> being examined in this project:

|  |
| --- |
| New York, NY (New York County-Manhattan) |
| Los Angeles-Long Beach-Anaheim, CA |
| Chicago, IL (internet search largest in Chicago result: Cook County) |
| Dallas-Fort Worth, TX |
| Philadelphia, PA |
| Houston, TX |
| Washington, DC |
| Miami-Fort Lauderdale, FL |
| Atlanta, GA (Fulton County) |
| Boston, MA (Middlesex County) |
| San Francisco, CA |
| Detroit, MI (Wayne county the largest) |
| Riverside, CA |
| Phoenix, AZ (Maricopa County) |



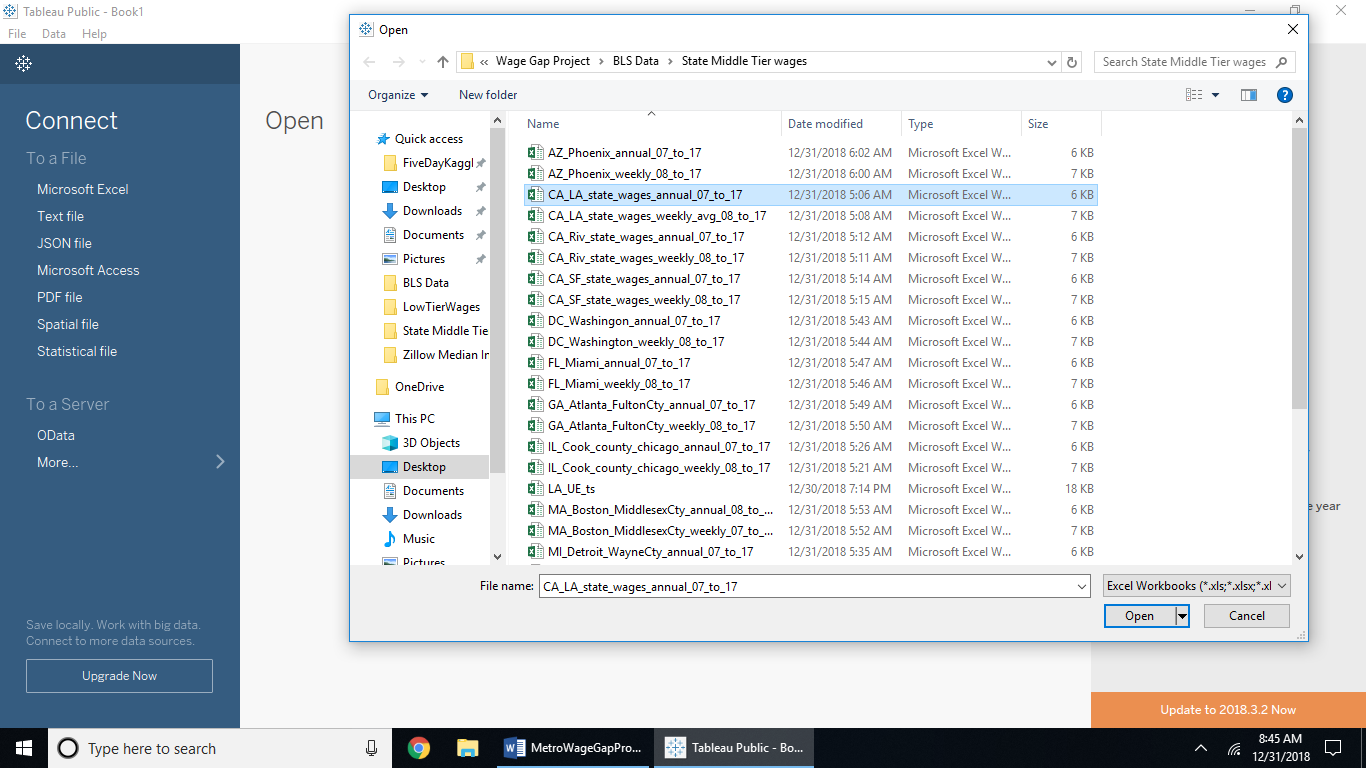
Examine the measure of wage gaps per metro area by examining state employee average annual and weekly wages of each metro location against the wage/unemployment database on all other jobs including the lowest and median jobs by metro location according to the Bureau of Labor Statistics site data. Use the main data link on the BLS website, go to Pay & Benefits section then scroll to the Quarterly heading for State and County Employment and Wages. Search by ‘One Screen Data Search’ to find regional data for top 15 metros.

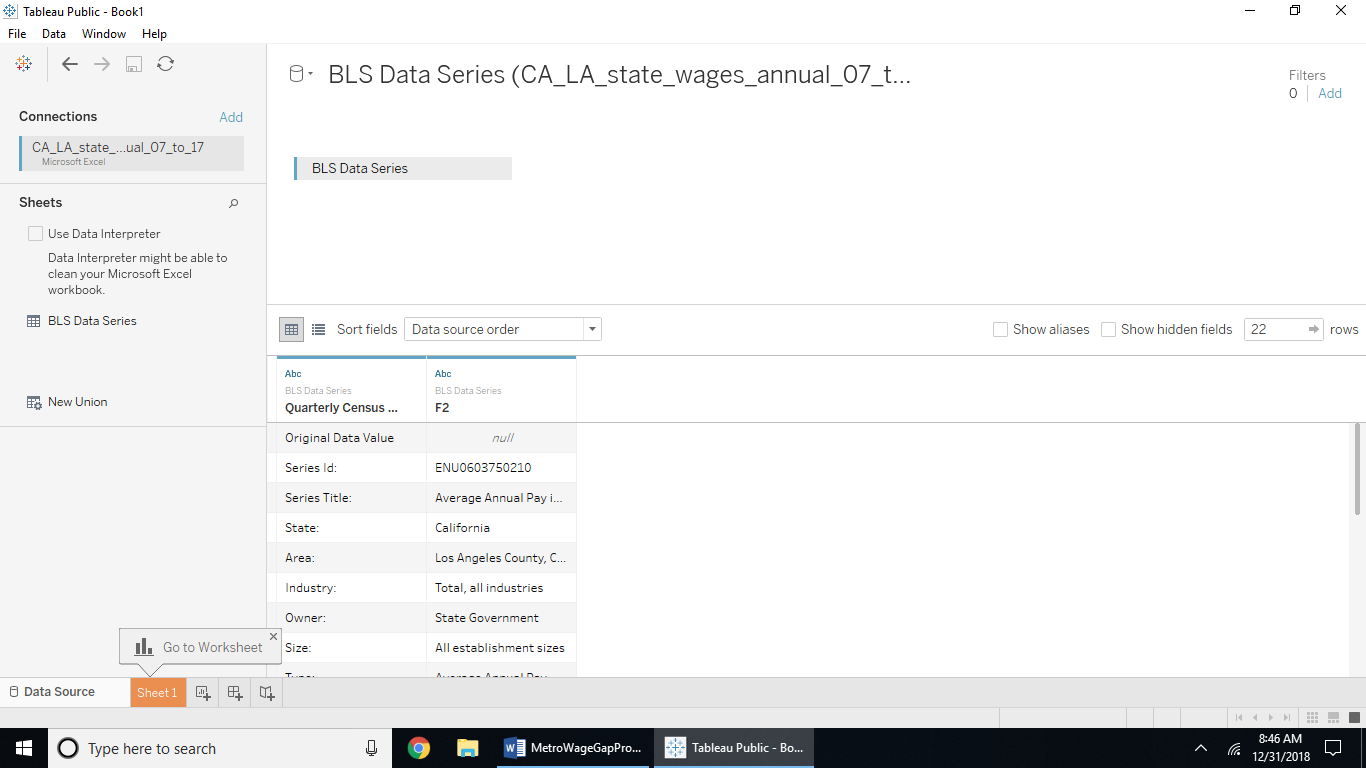
This report assumes the workers classified as non-union civilian security guards (in Dallas, TX no such occupation for security guard so customer service representative or cashier was used and also for Atlanta and Phoenix, These states also have ‘open-carry gun laws’ according to <https://www.gunstocarry.com/gun-laws-state/> accessed 12/31/2018) or non-union food preparers. The data is only given for 2017. A calculated column will be created once the data is merged that will find the average of the food preparation workers and the security guards. Because consumers need fast food or casual dining food and commercial property protection by non-public entities to park their cars and be safe from unruly patrons that may cause conflict. These job roles are among the bottom tier of workers receiving the low end of wages to calculate the wage gap between middle income state workers known to be able to afford all of the desires of any American such as multiple vehicles a house for more than four people, children, day care, vacations, health insurance, education costs to improve, and pension investments. Whereas the bottom tier of the wage gap are considered fortunate if they can afford to live in a single bedroom apartment with a pet or have an apartment with a girlfriend with either of them going to school and managing to live paycheck to paycheck and sacficices are made to either have a reliable mode of transportation or health insurance and ofcourse choose to walk away from conflict than participate in it and risk a criminal record and permanent community status demotion to ‘2nd class citizen’ with all attachments in terms of employment. This project seeks to clarify the wage gap per top 15 metro largest populations in the US. Future problems will seek to outline and clarify what the last few sentences mentioned by using the bjs and cdc for criminal and birth/death rates along these top 15 metros. To get to these data sets individually by job role and metro go to the Pay & Benefits section and the model wage estimates using the one screen data search. All files were renamed to identify what they are instead of the BLS data names of ‘SeriesReport-numericdate-ANencryption.xlsx’. These files and the original files will be in a folder to the dropbox account to access and reproduce this project results. The link to my dropbox account for the BLS Data is: https://bit.ly/2VfNluN .

The next section of this project will have to gather the 2017 data for each metro earnings for state jobs as annual median, then create a calculated field for low wage earnings per year between the low tier jobs hourly average hourly wages. Then the Zillow data will be filtered by the last available quarter of 2017 for the annual median income of home affordability, buyer-seller index, 2-bedroom rent, median home cost, house rent, number of houses bought and sold last quarter of 2017, and create a calculated field for the wage gap between middle and low-income earners on an annual basis per metro region.

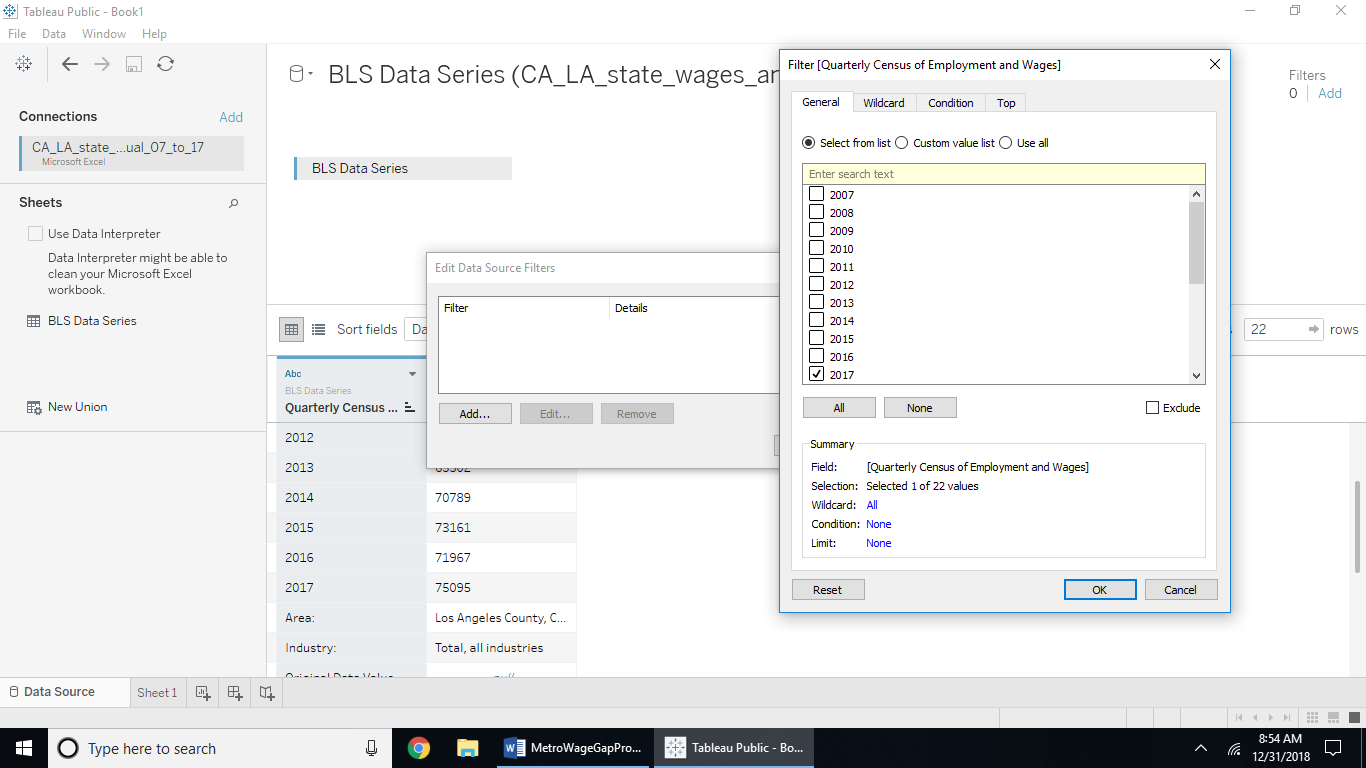
This will be a tedious process to filter and merge on regional fields along data sets from different sources and working with about 100 data sets approximately. So not all steps will be documented. But the final data set from this data warehouse of information gathered from the web will be presented from which visualizations can be created showing relationships amongst metros and the wage gap on standards of living which will later have 2017 crime and birth/death rates added.

The metro zip file downloaded Christmas day for Zillow has all the data metrics you can use by metro. The files being used from the Metro zip file of Zillow are BuyerSellerIndex\_Metro.csv, DaysOnZillow\_Public\_Metro.csv, ForeclosuresPer10Khomes\_Msa.csv, Metro\_MedianListingPrice\_3Bedroom.csv, Metro\_MedianRentalPrice\_3Bedroom.csv, Sales\_Counts\_Msa.csv . And the other Zillow files being used are: Affordability\_Income\_2018Q3.csv and Affordability\_Wide\_2018Q3.csv.

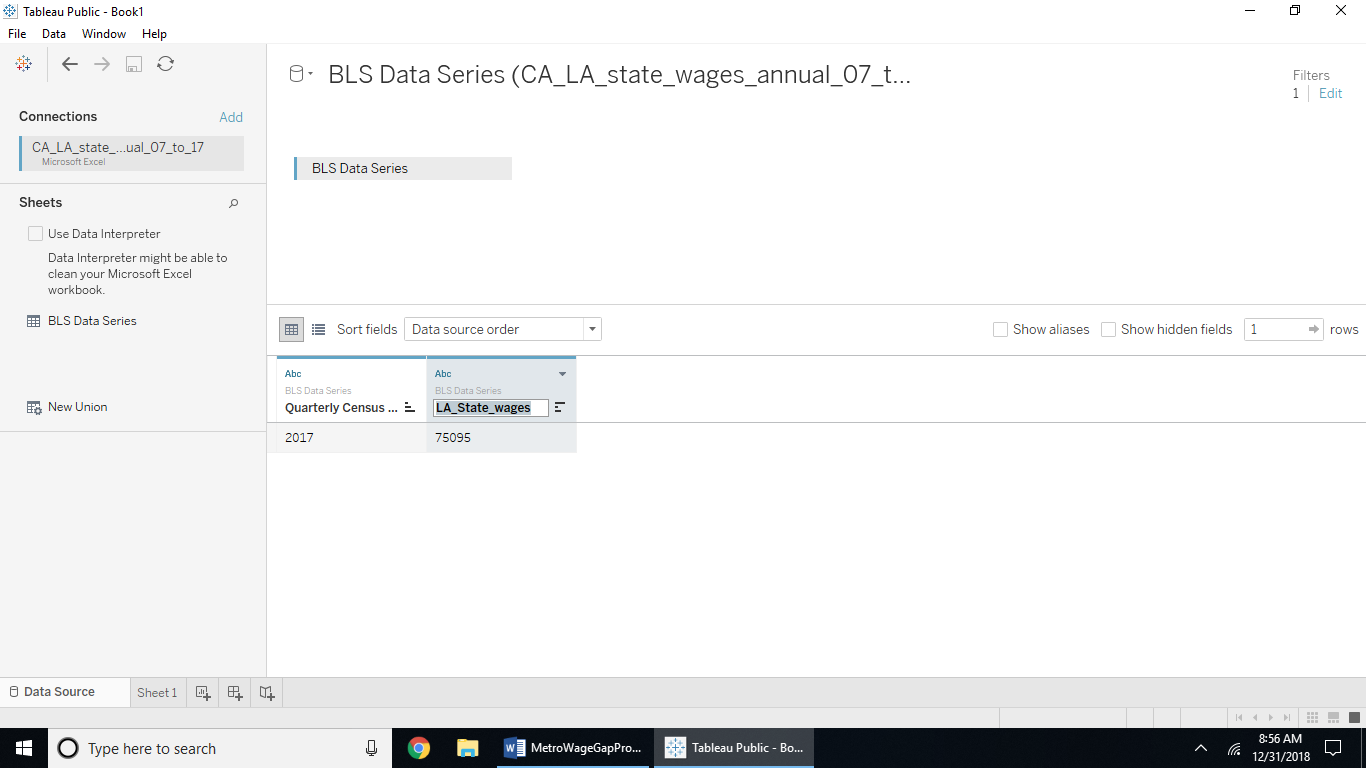




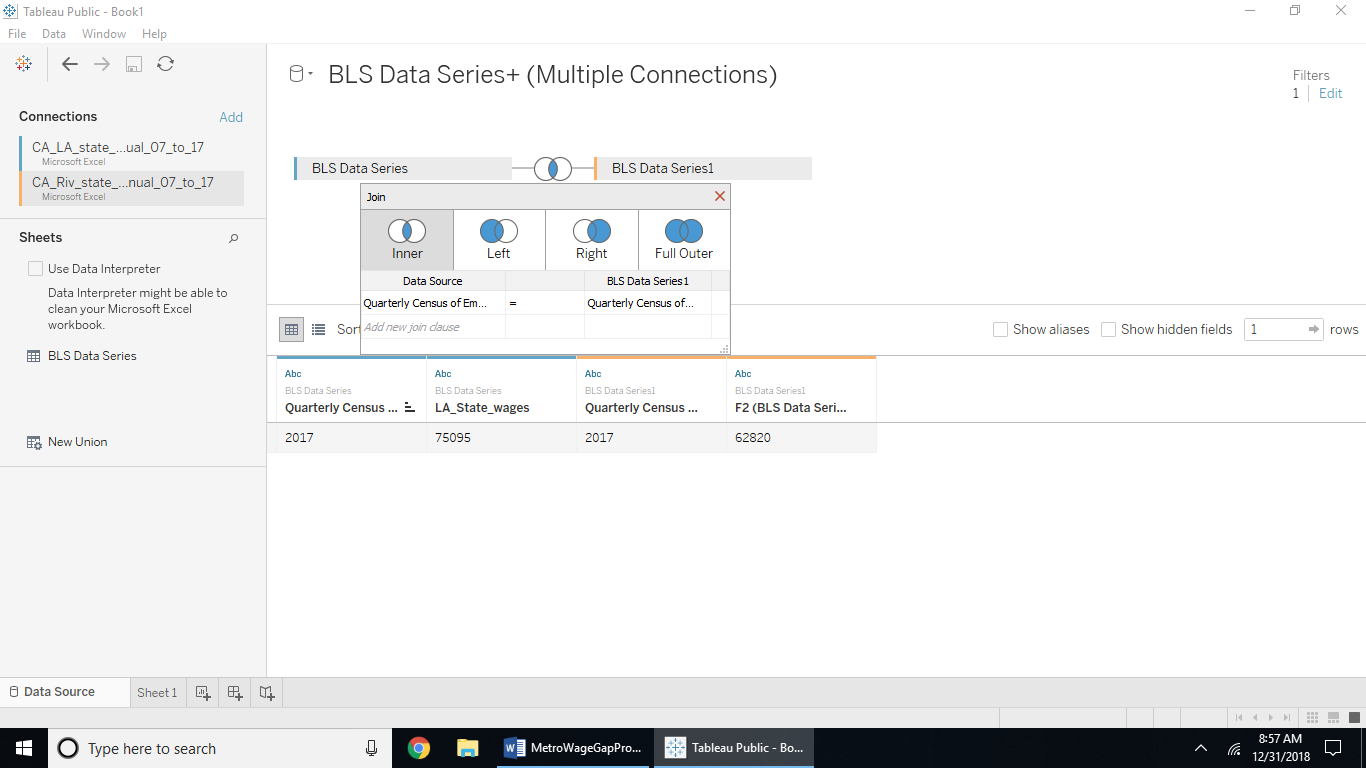
The above screenshot shows the BLS Excel file for annual data on state employees in LA imported as a Microsoft Excel file into Tableau without using the data interpreter.



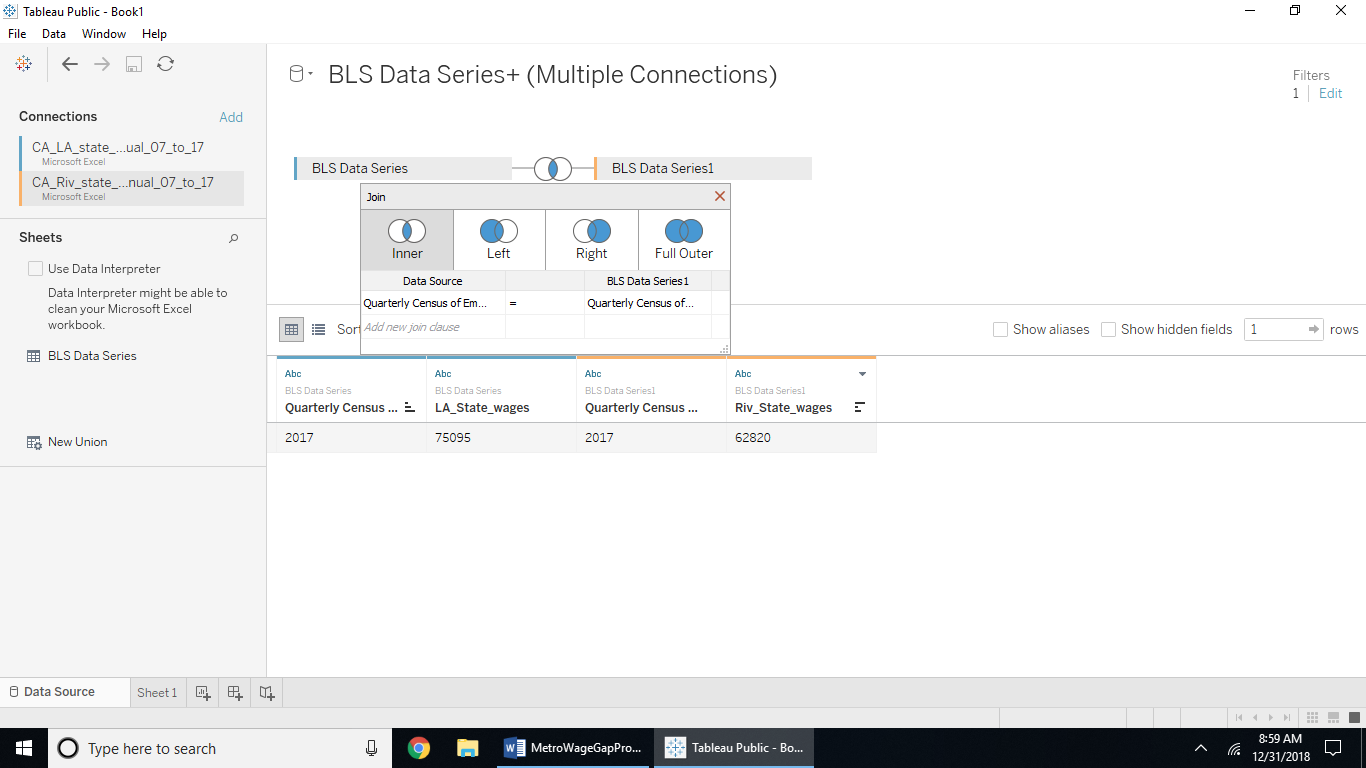
Using a filter from the Data tab to filter out year 2017 in the first column is shown in the above screenshot



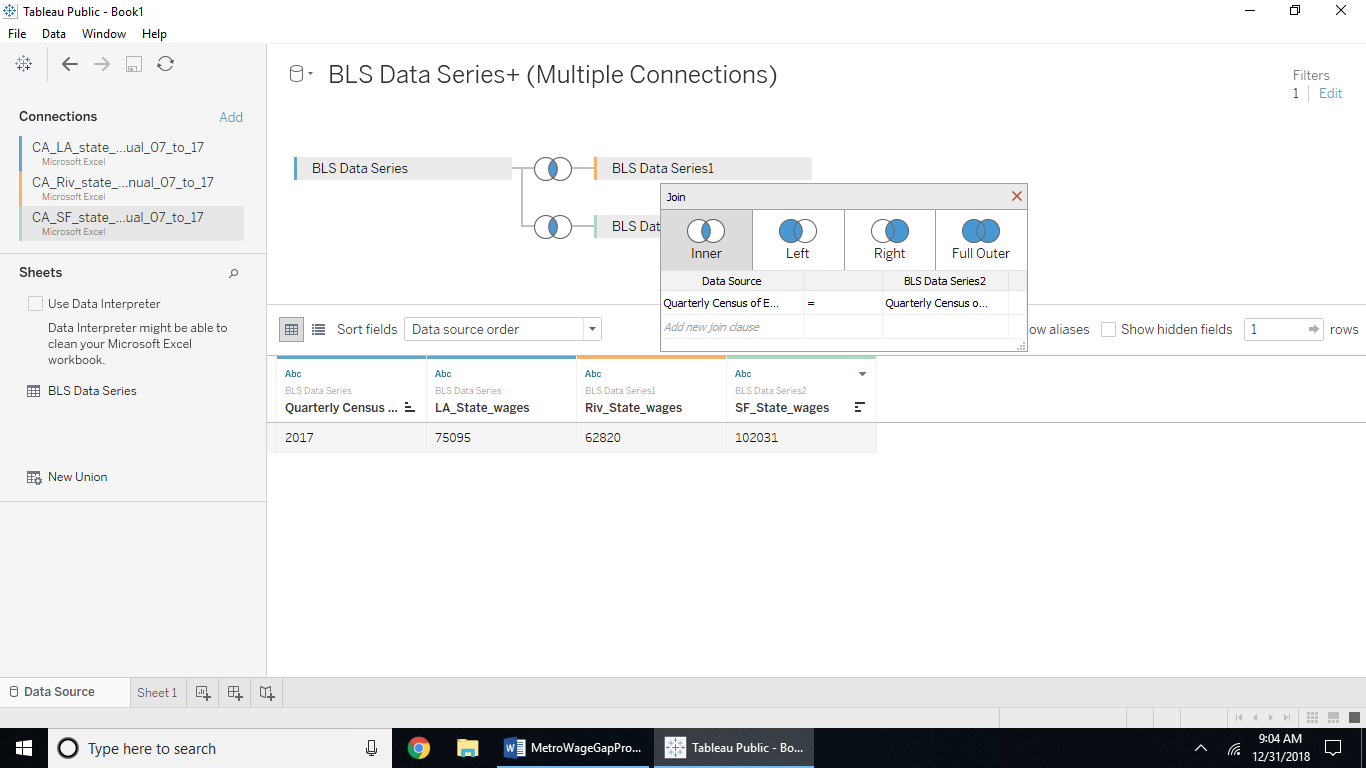
The above screenshot shows the 2017 annual wages with the ‘F2’ field name changed to LA\_State\_wages.

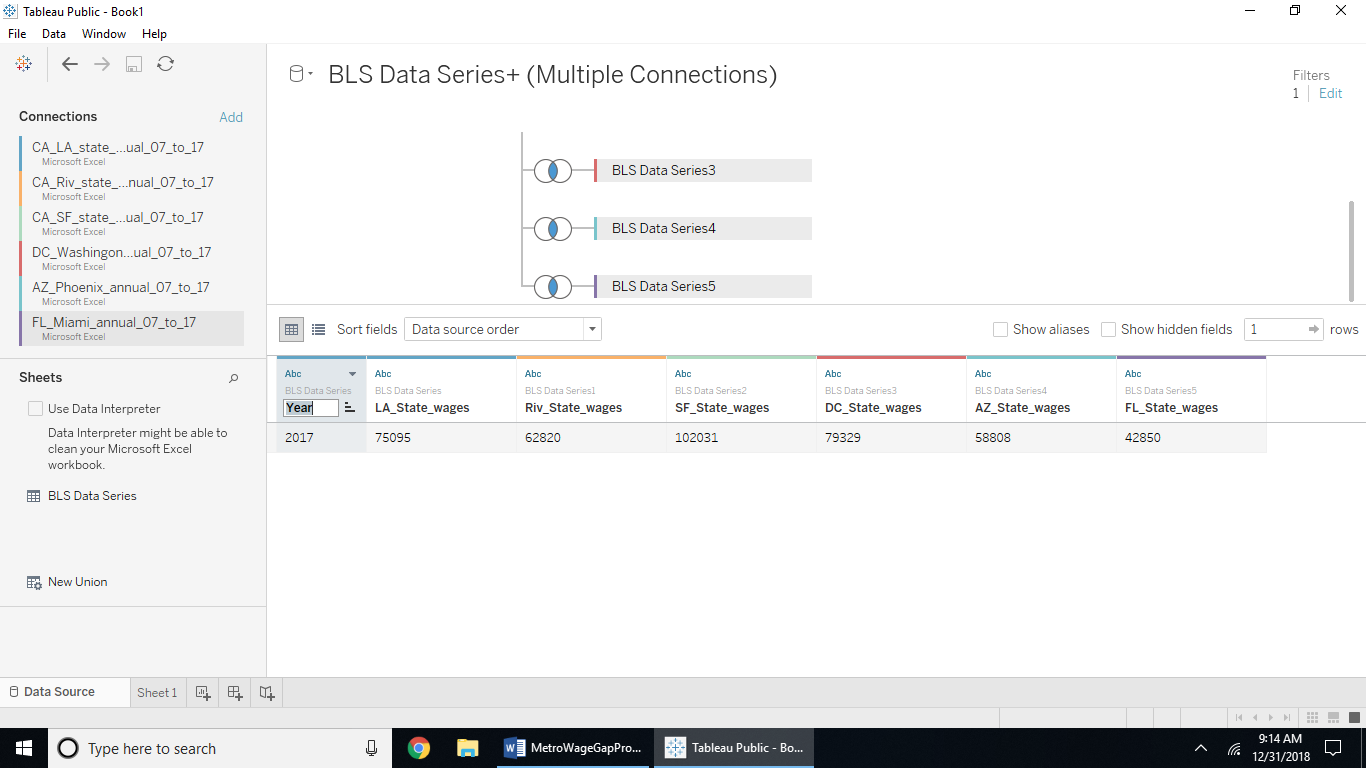


The above screenshot shows adding the Excel file for Riverside state employee annual wages and joining with an inner merge on the Quarterly field.



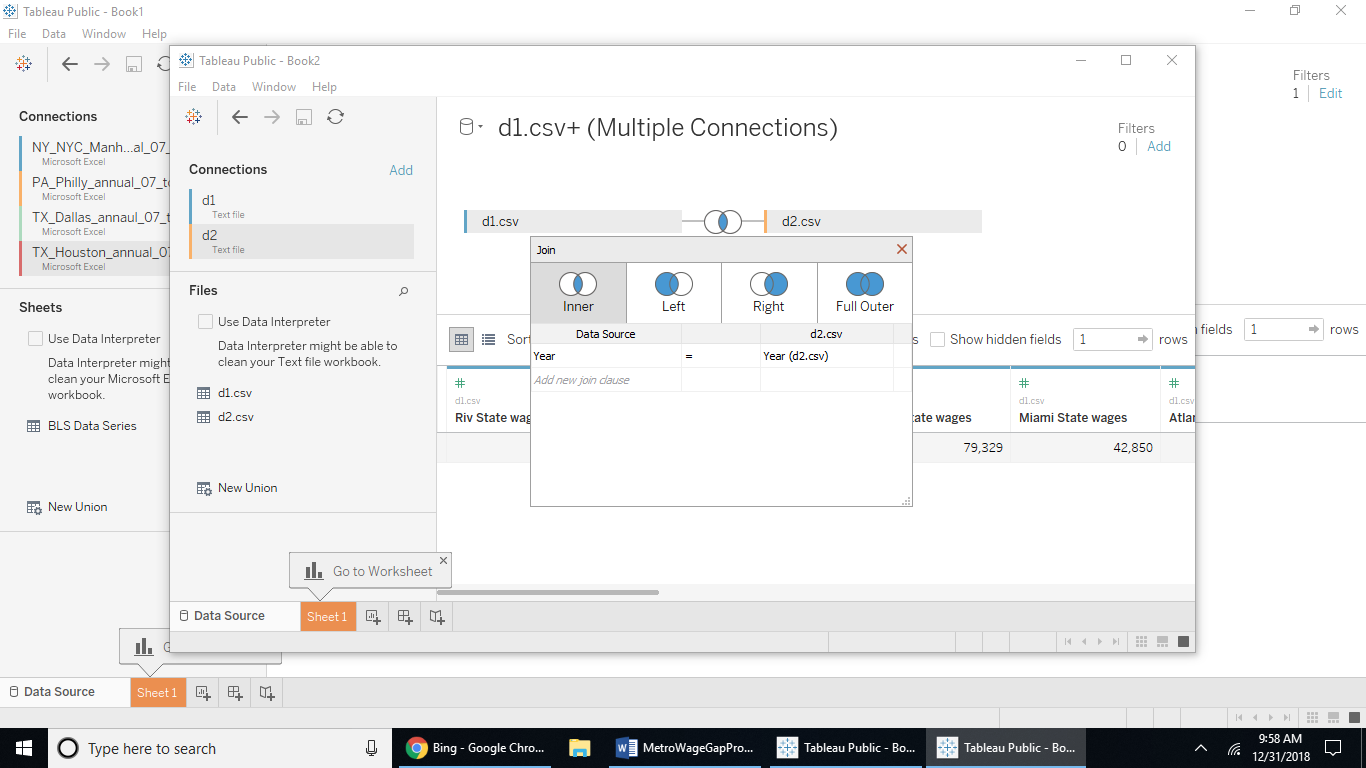
The screenshot above shows the name field for the Riverside wages in 2017 will be changed to Riv\_State\_wages. Continue doing this systematically for all top 15 metros in the BLS state wages data for each metro and naming the fields by the metro abbreviation and ‘\_State\_wages’ as shown in the next screen shot:

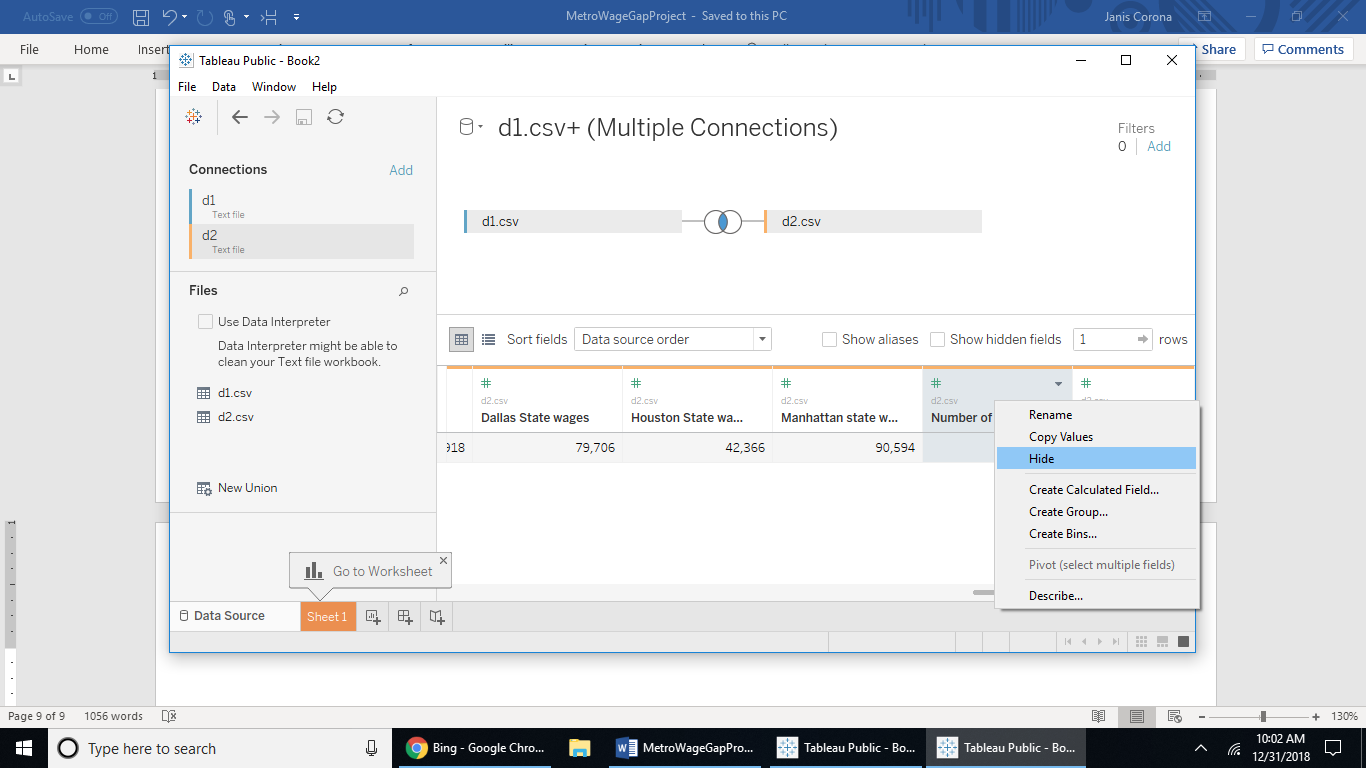




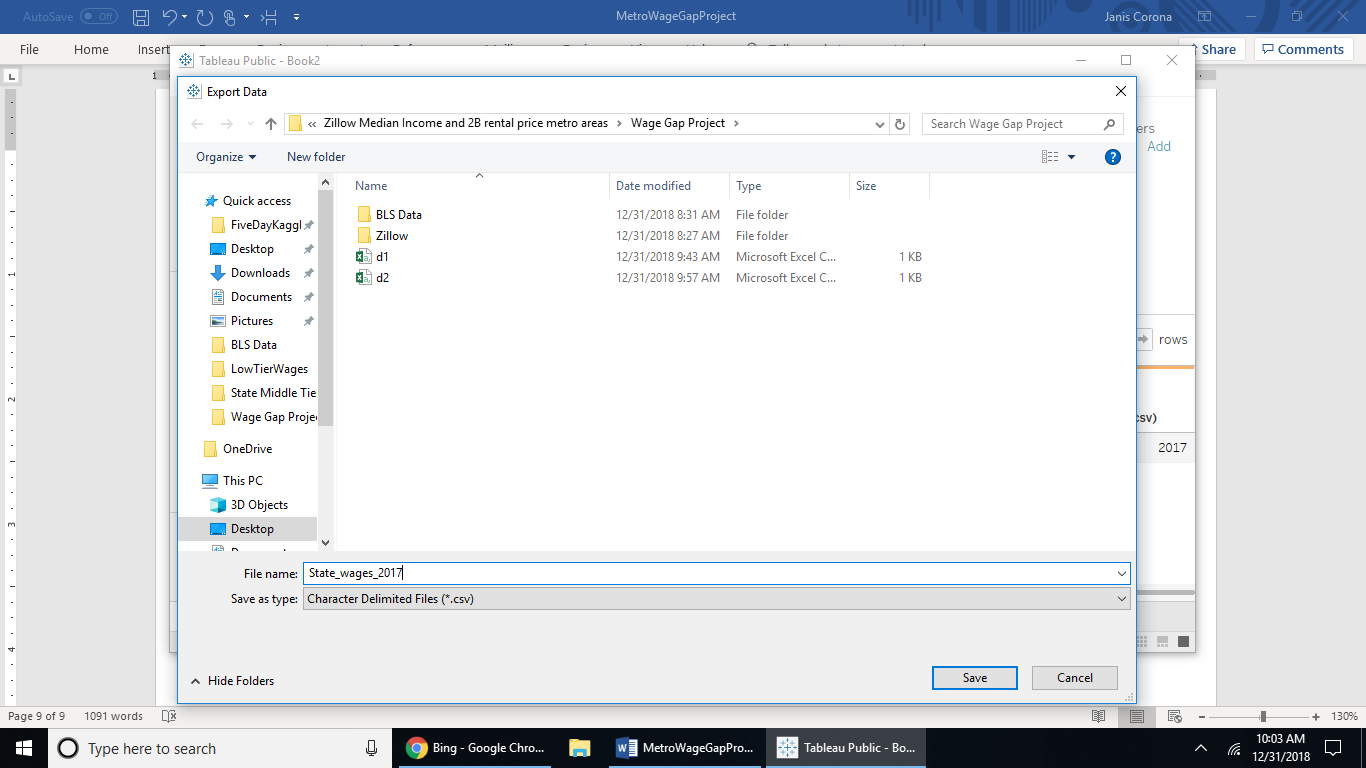
Rename the Quarterly field as ‘Year’ as you continue to add the metros to this data table.

Complete half the metros export from the Data tab as a csv file, then create a second data set that contains the other half and export as a csv file. Then open a new file in Tableau as txt for the first half metro files changing the quarterly field as ‘Year’ and hiding all other quarterly fields. My first data set is ‘d1.csv’ and then inner join the second file (‘d2.csv’) joining on the Year fields as shown below:

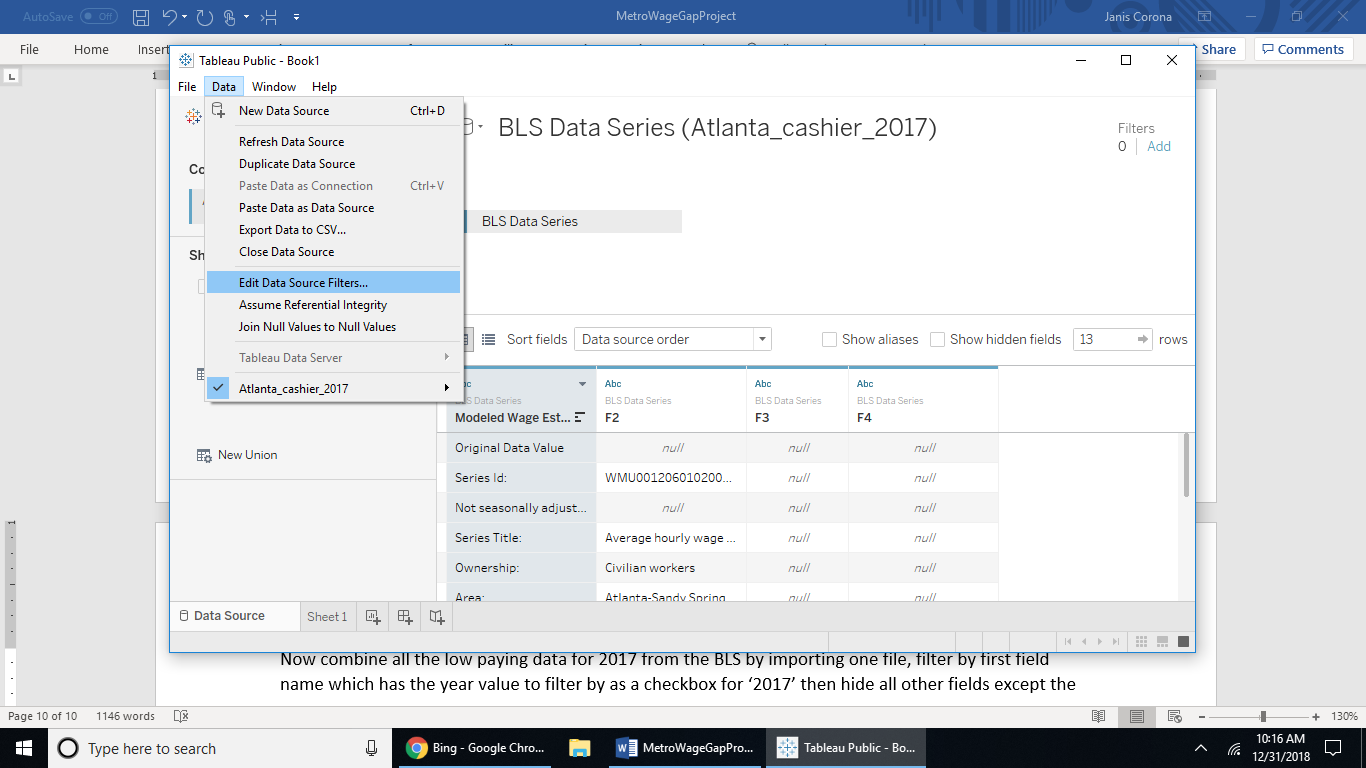
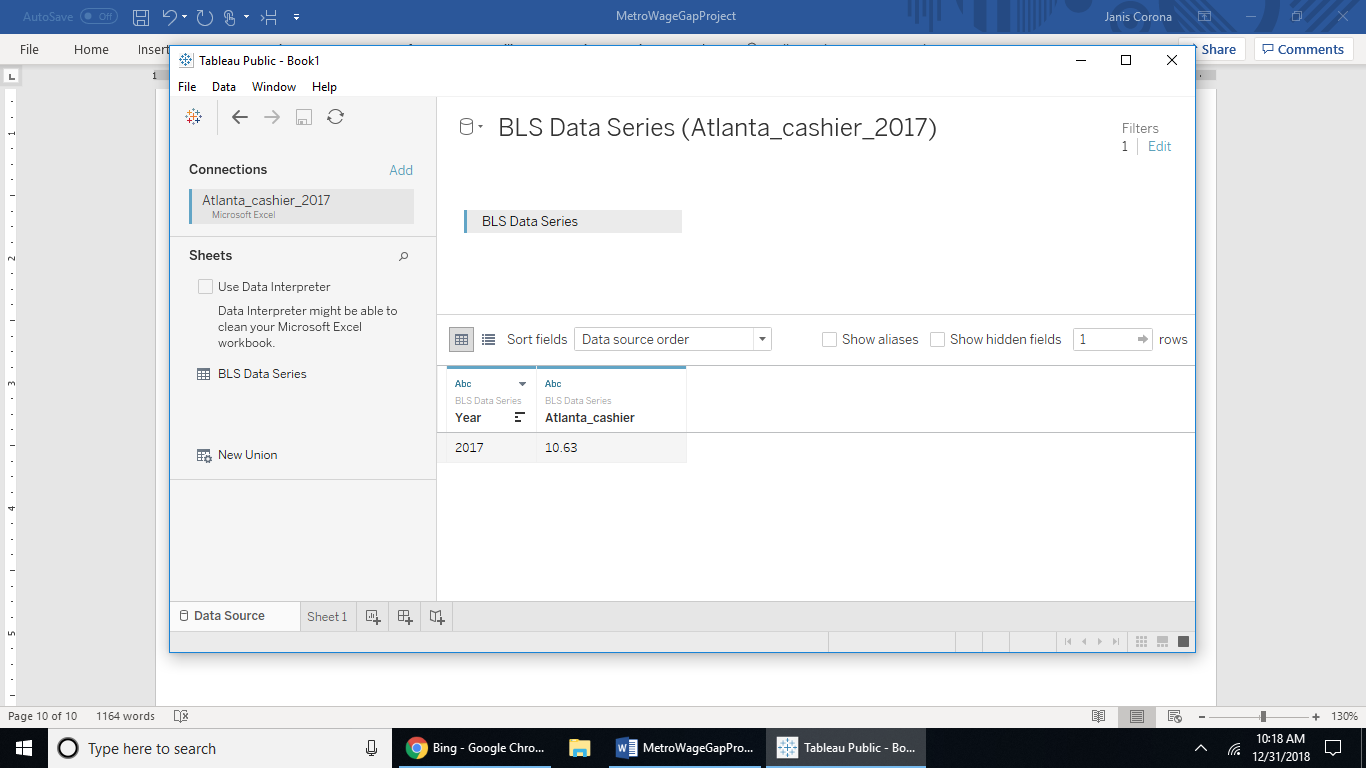




Hide the ‘Number of Records’ field and the ‘Year(d2.csv)’ field. This is the original field the d1.csv and d2.csv tables want to be joined on, but the year values are the values these two data sets have in common.

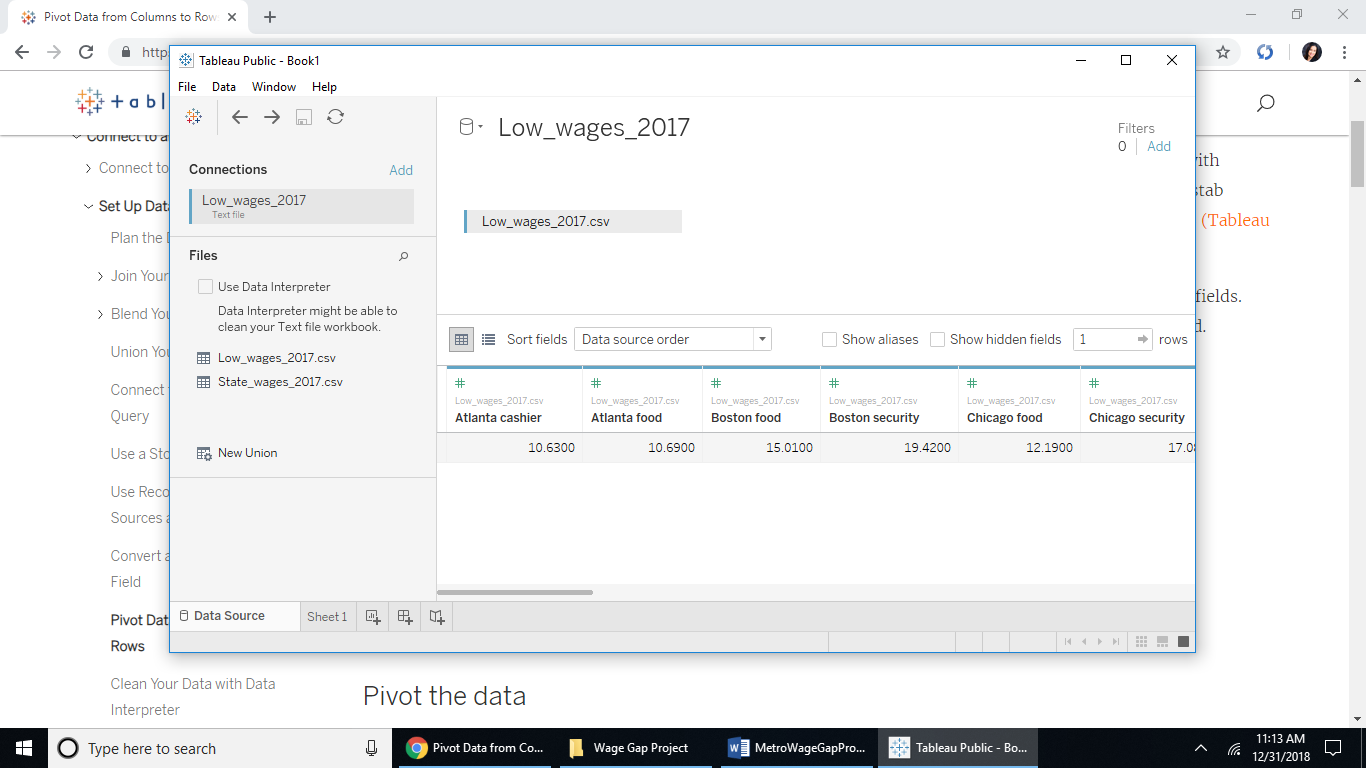


Export the merged d1 and d2 data sets as State\_wages\_2017.csv.

Now combine all the low paying data for 2017 from the BLS by importing one file, filter by first field name ‘Modeled Wage…’ which has the year value to filter by as a checkbox for ‘2017’ to keep. Then hide all other fields except the ‘F3’ field for the annual average wage for that low paying job by metro. 

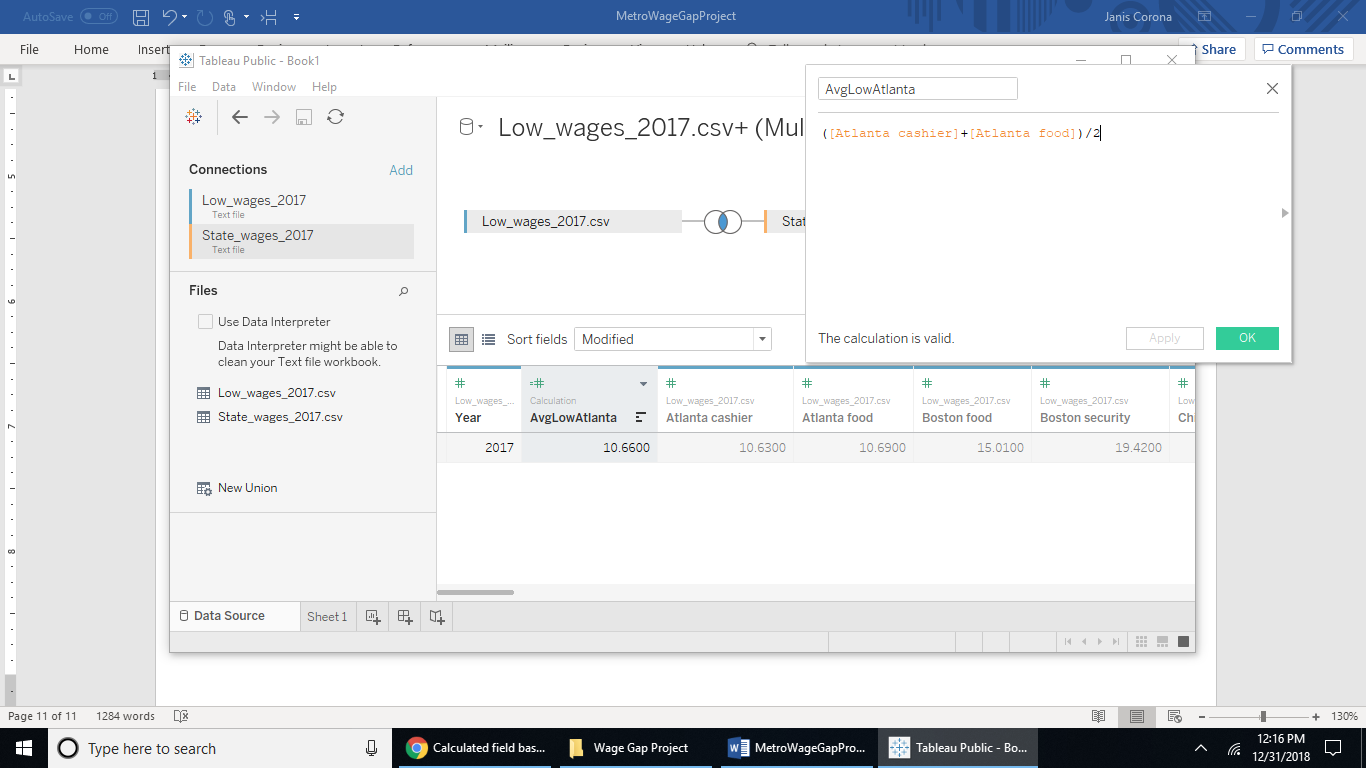
Rename the ‘Modeled Wage Est…’ field ‘Year’ and the ‘F3’ field the corresponding job and metro which in this case is ‘Atlanta\_cashier’ for this data set.

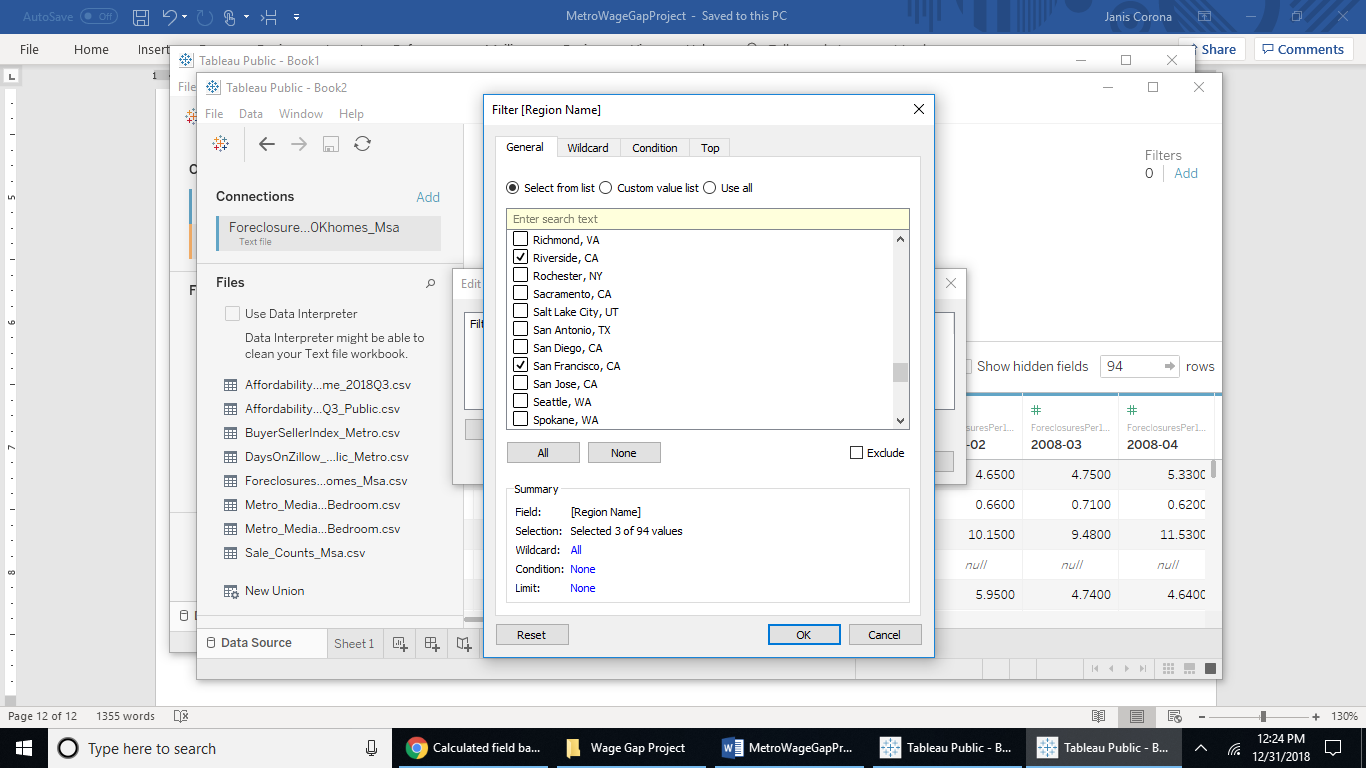
The program slows the more reports are added. So create a d3 to export with the first ½ the metro low pay values, then a d4 file to export with the remaining ½ of the metro low pay data. Then import d3 csv file and add the d4 file. Merge on the ‘Year’ field. Hide the ‘Number of Records’ fields and export to csv from the Data tab as Low\_wages\_2017.csv . You can delete the d1 through d4 intermediary data files created earlier as they won’t be needed.



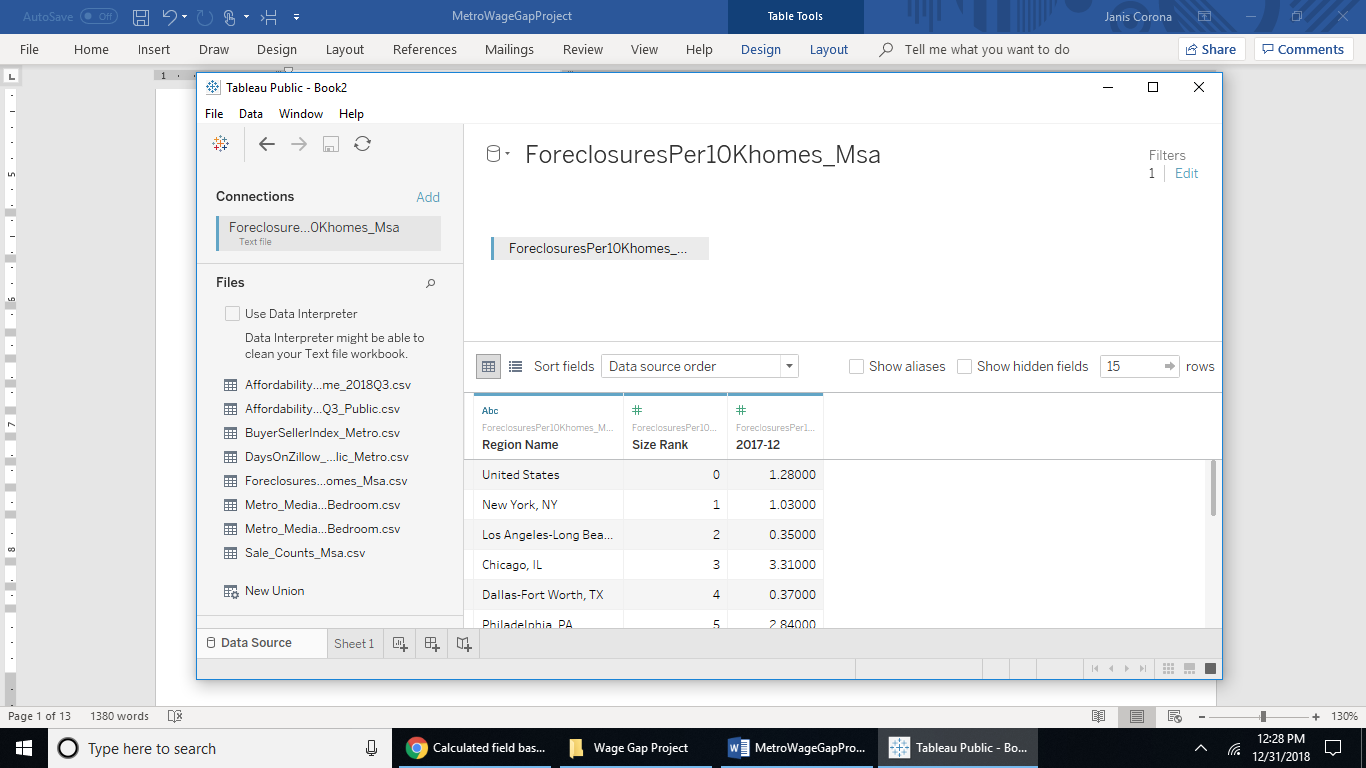
Open the Low\_wages\_2017.csv file in Tableau.

Create the calculated field for each low pay job by metro, Tableau allows grouping by filtering in a group by each low wage of food preparer or security, but I am unaware from any internet search findings of how to use an aggregate function on those groups, hence there has to be 15 calculations for the average manually as shown. Do this for each metro, then export to csv as HighLowWages.csv.

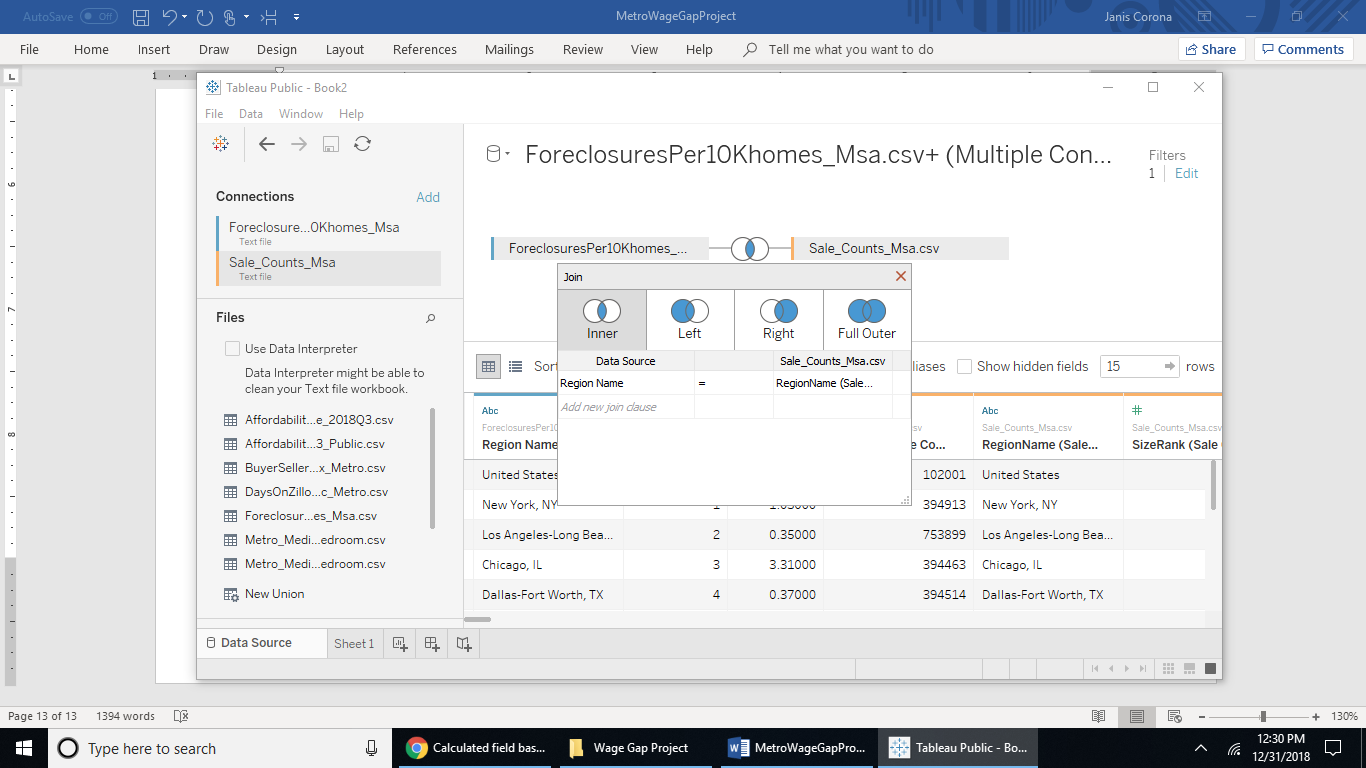


Open the Zillow foreclosures csv file as a text file in Zillow, and filter by Region name the United States and all 15 metro regions. 

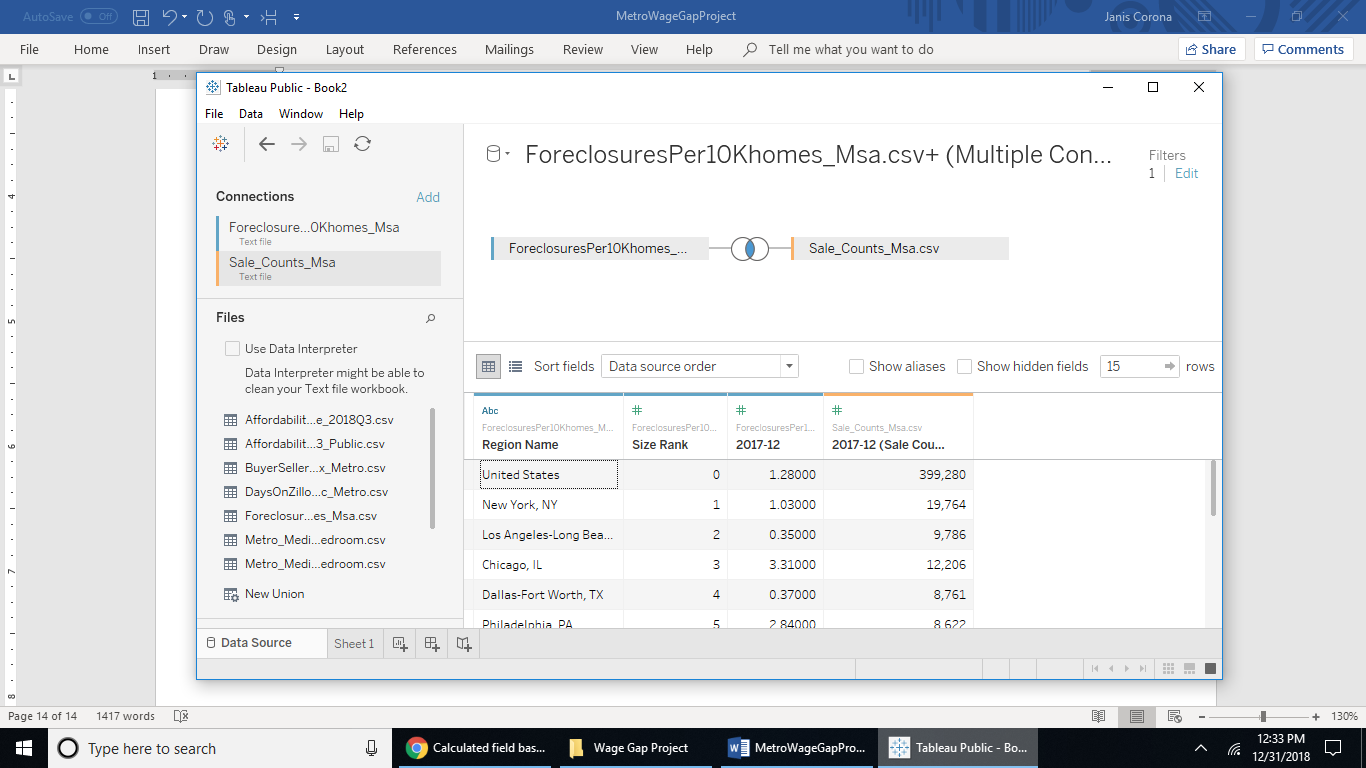
Hide all of the year fields and the ID field in this data table.



Join the Sales\_Counts\_Msa.csv file by Region Name to the Foreclosures file:

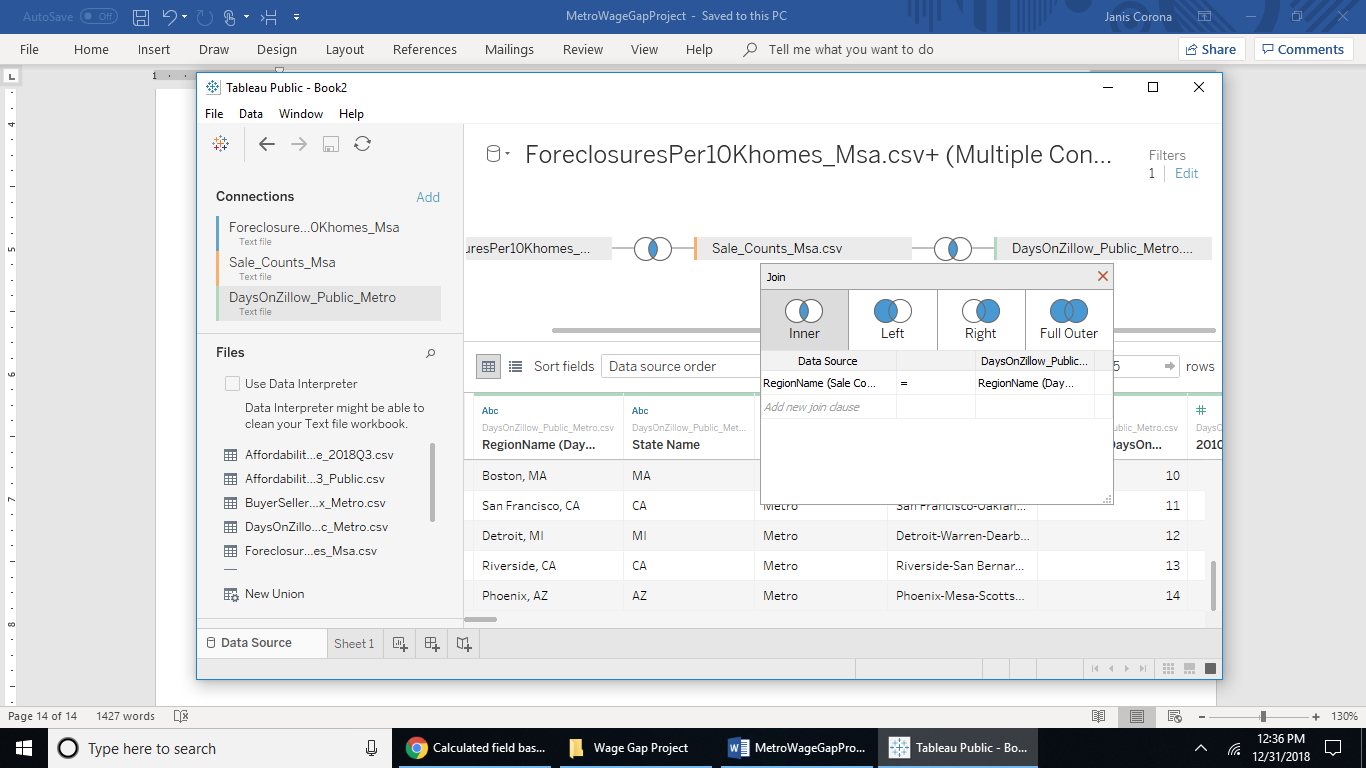


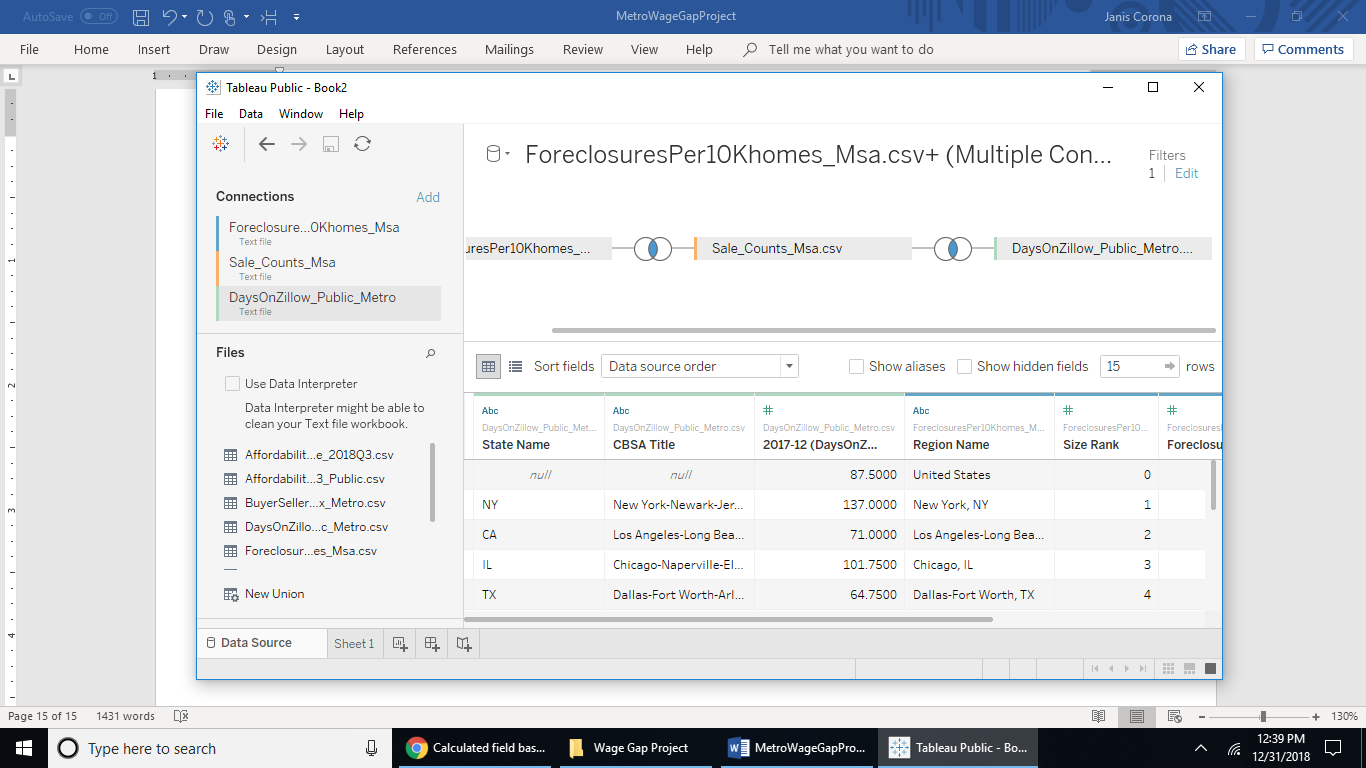
Hide all fields other than the 2017-12 field in the Sales\_Counts fields:



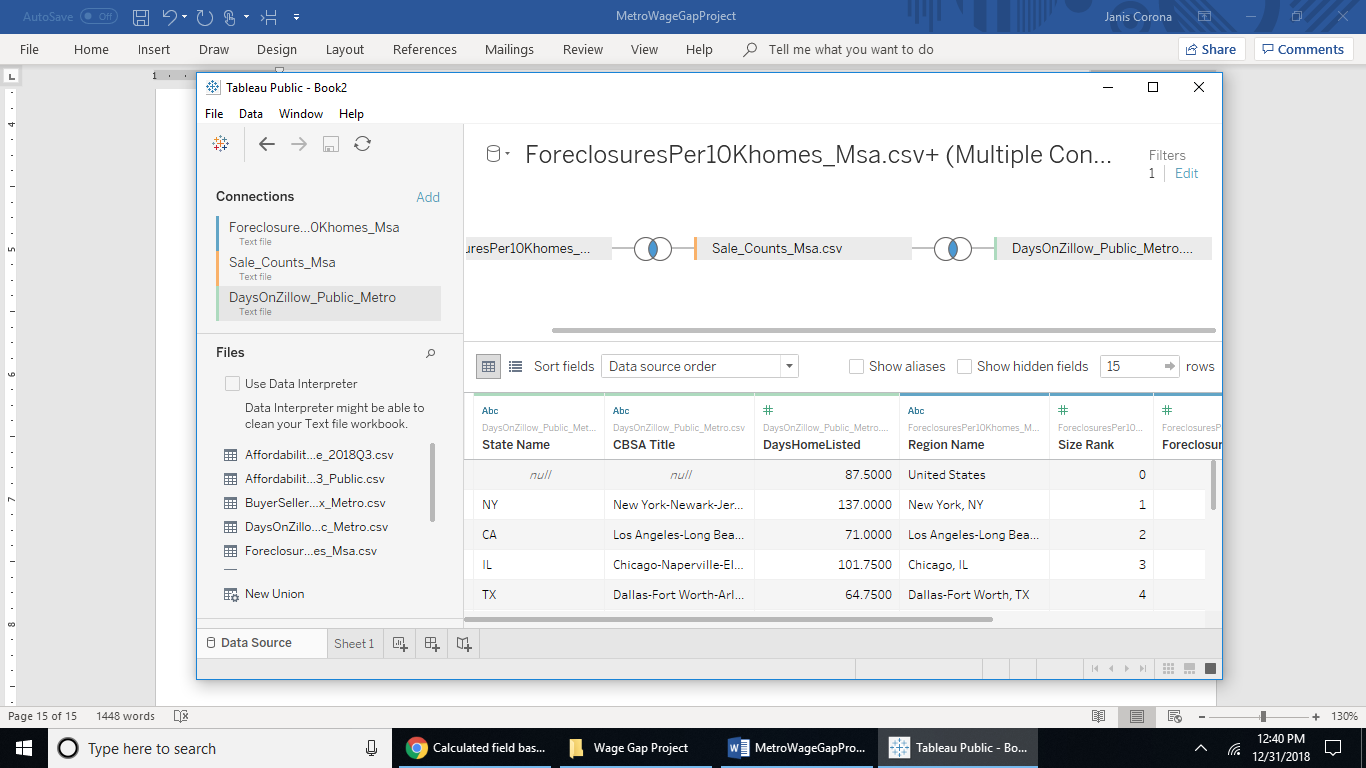
Rename the fields Foreclosures and HomeSales respective to their content.

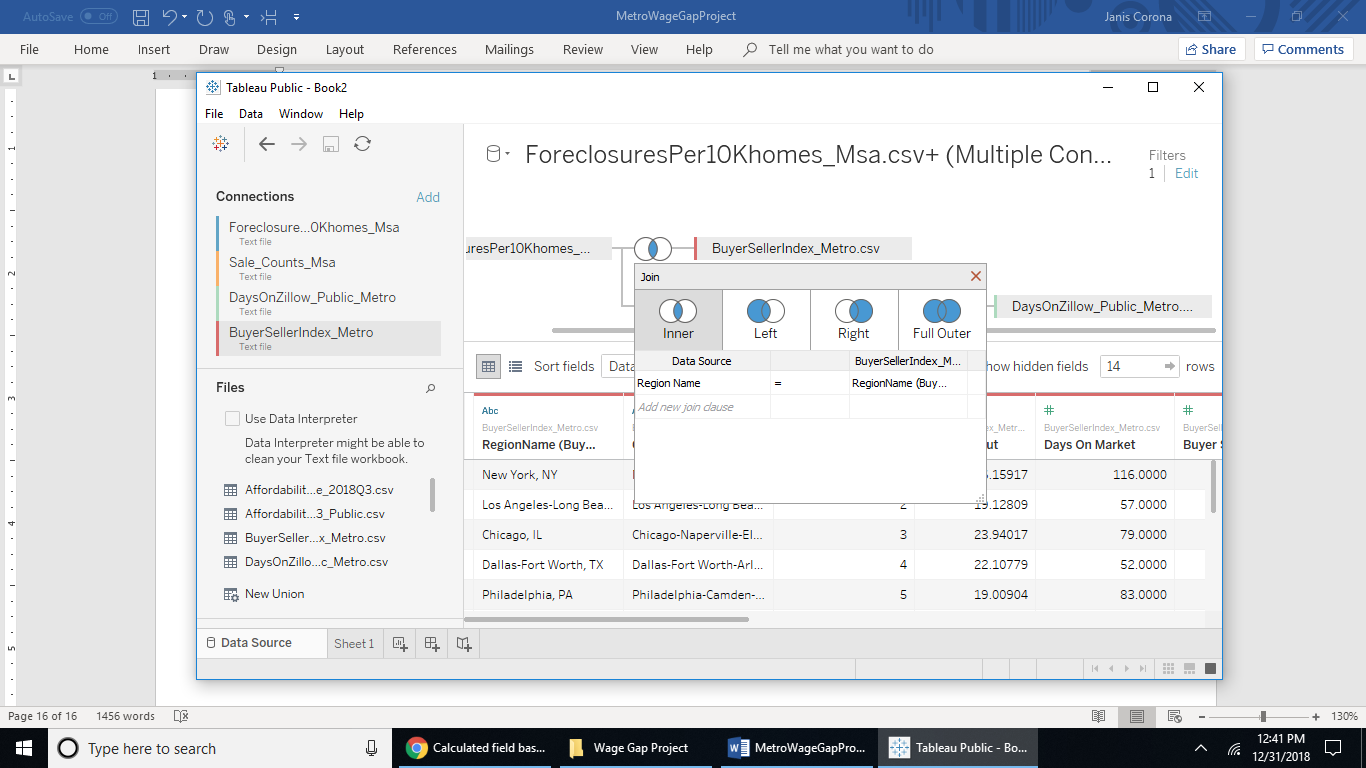
Join by Region name:



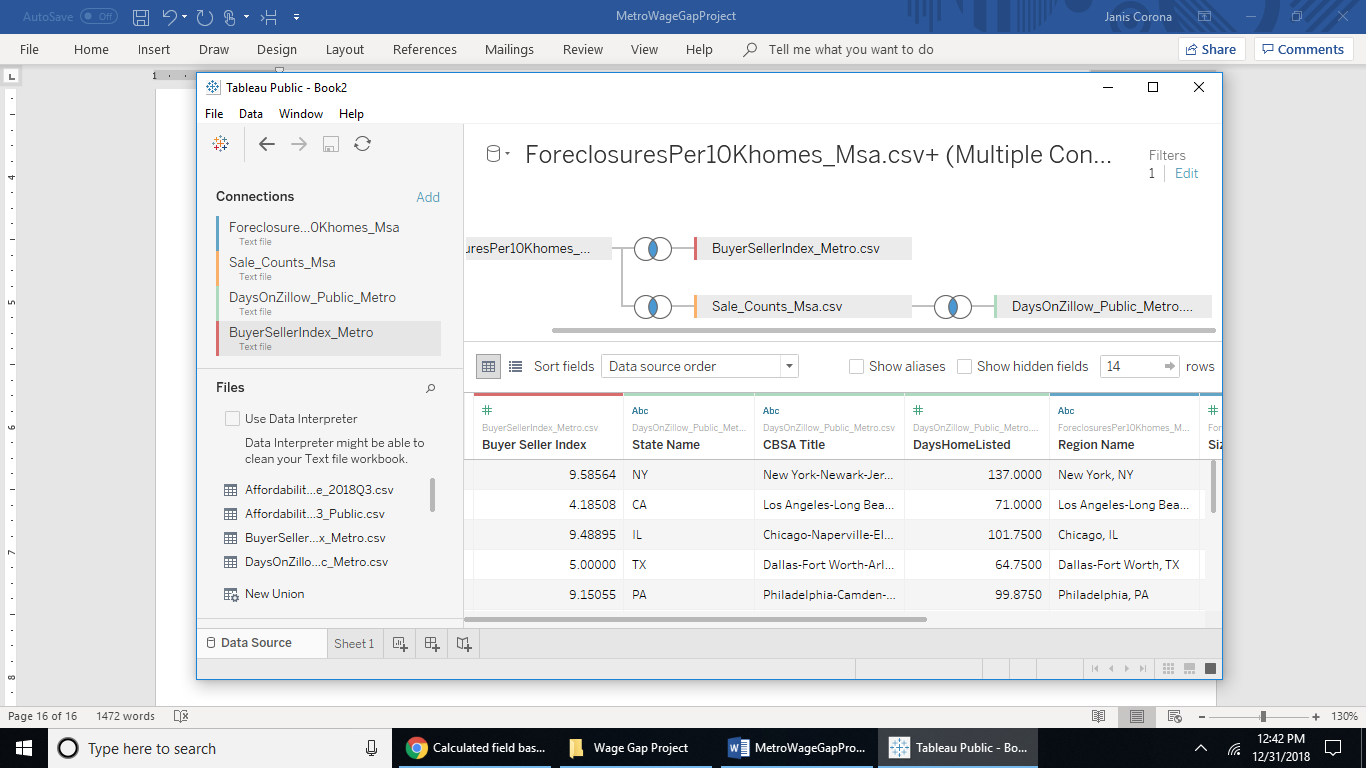


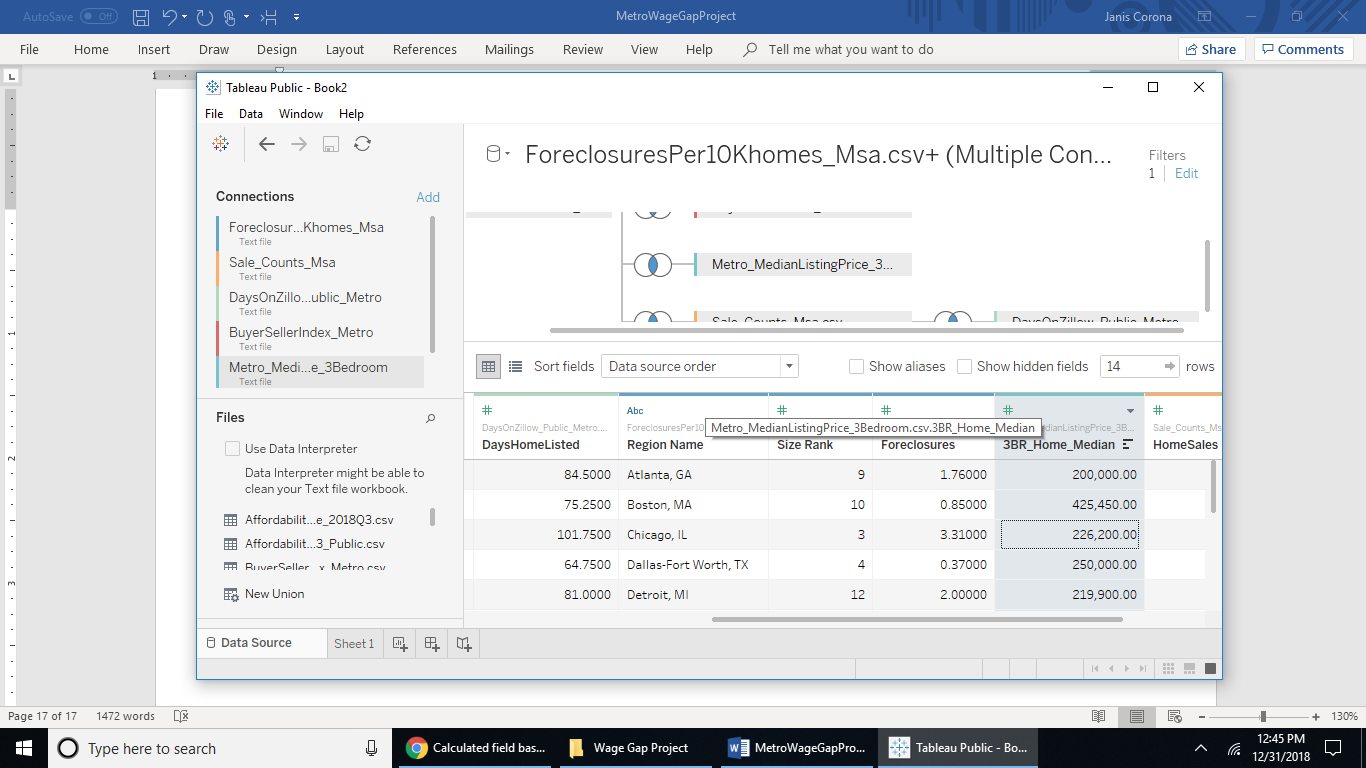
Keep the CBSA Title field, the State Name field and the 2017-12 field in the data joined to this data table, but rename it DaysHomeListed.





Inner join the Buyer-SellerIndex file by region name and keep the 2017-12 field renaming it BuyerSellerIndex.

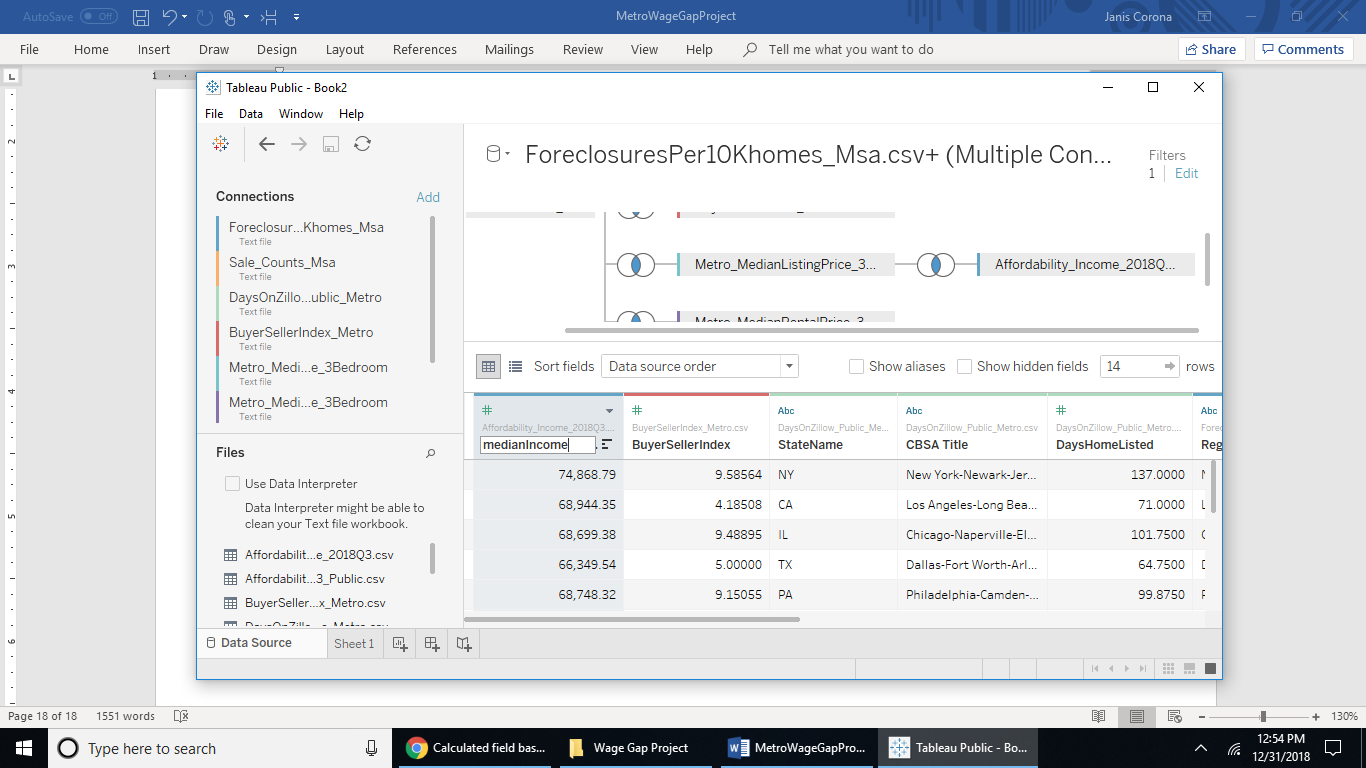


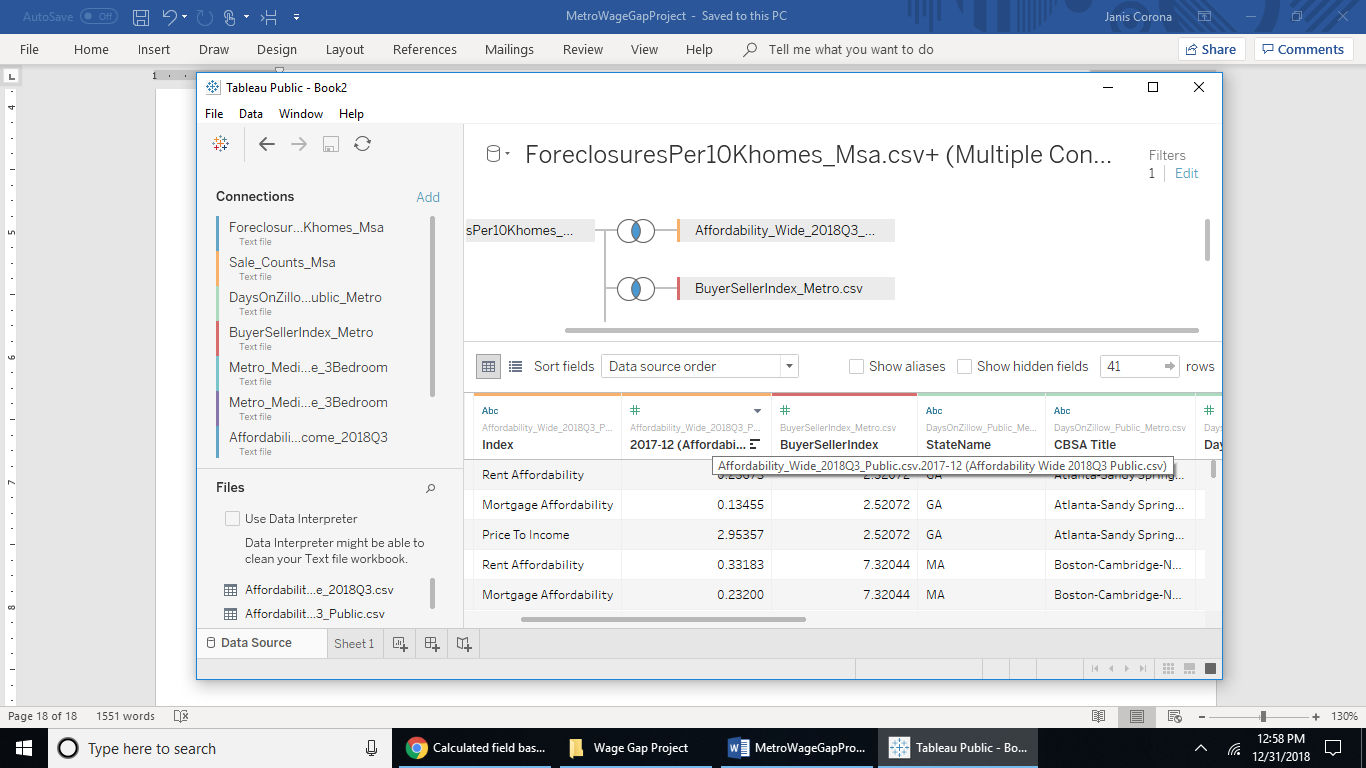
Then inner join the 3-bedroom median listing file to the data table by region name and hide all fields in the metro median home listing for 3BR except the 3bedroom median listing price field. Rename this field ‘3BR\_Home\_Median.’

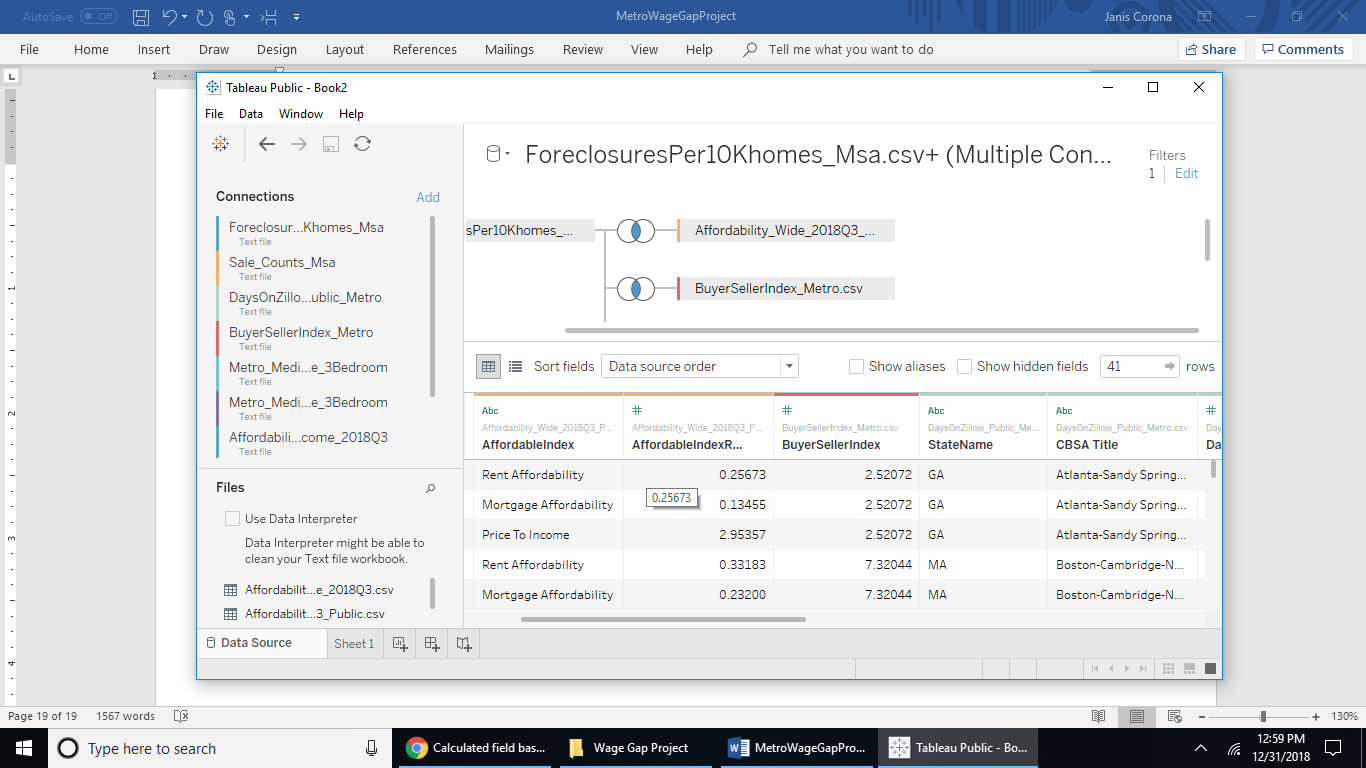
Then inner join the median rental listing price by region name and hide all fields except the 2017-12 field renaming it ‘3BR\_Rental\_Median.’



Next, add the median home affordability table by region name, hiding all fields except 2017-12 but renaming it medianIncome:

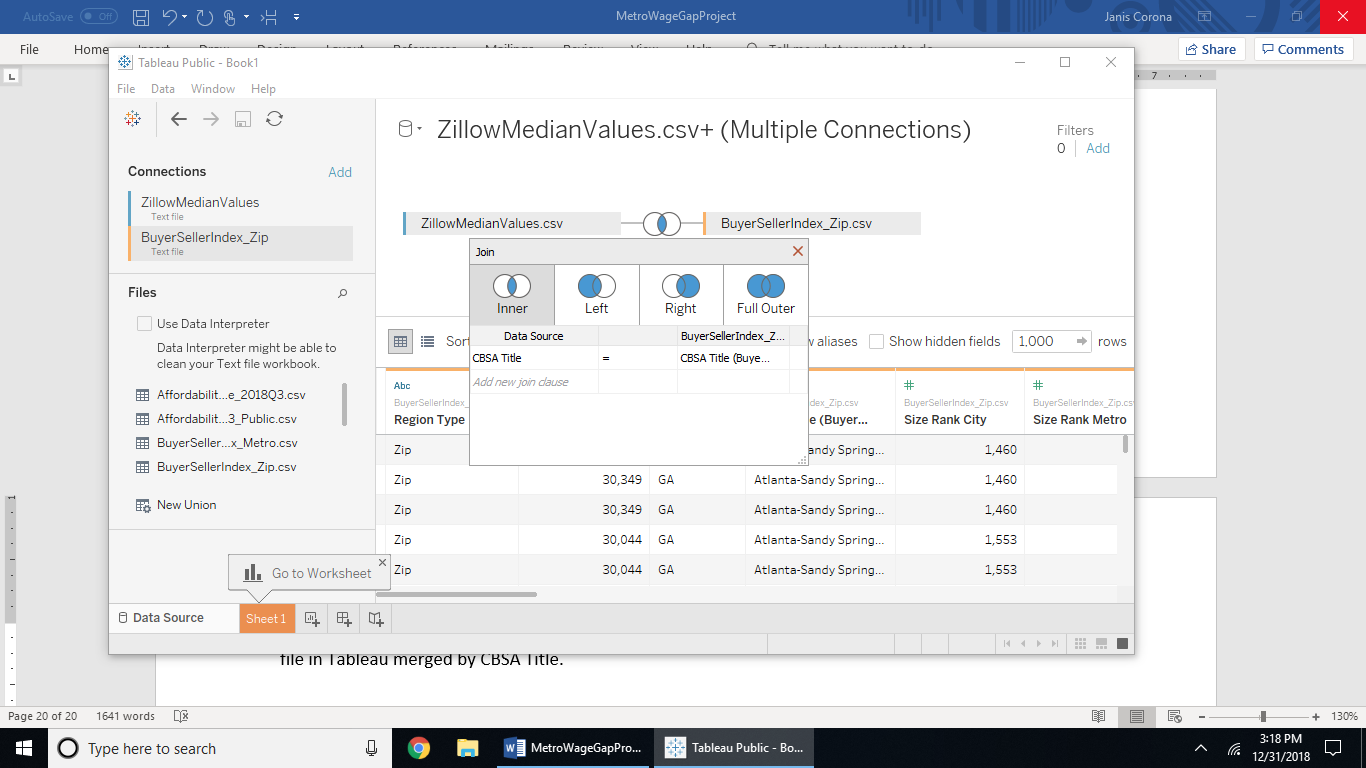


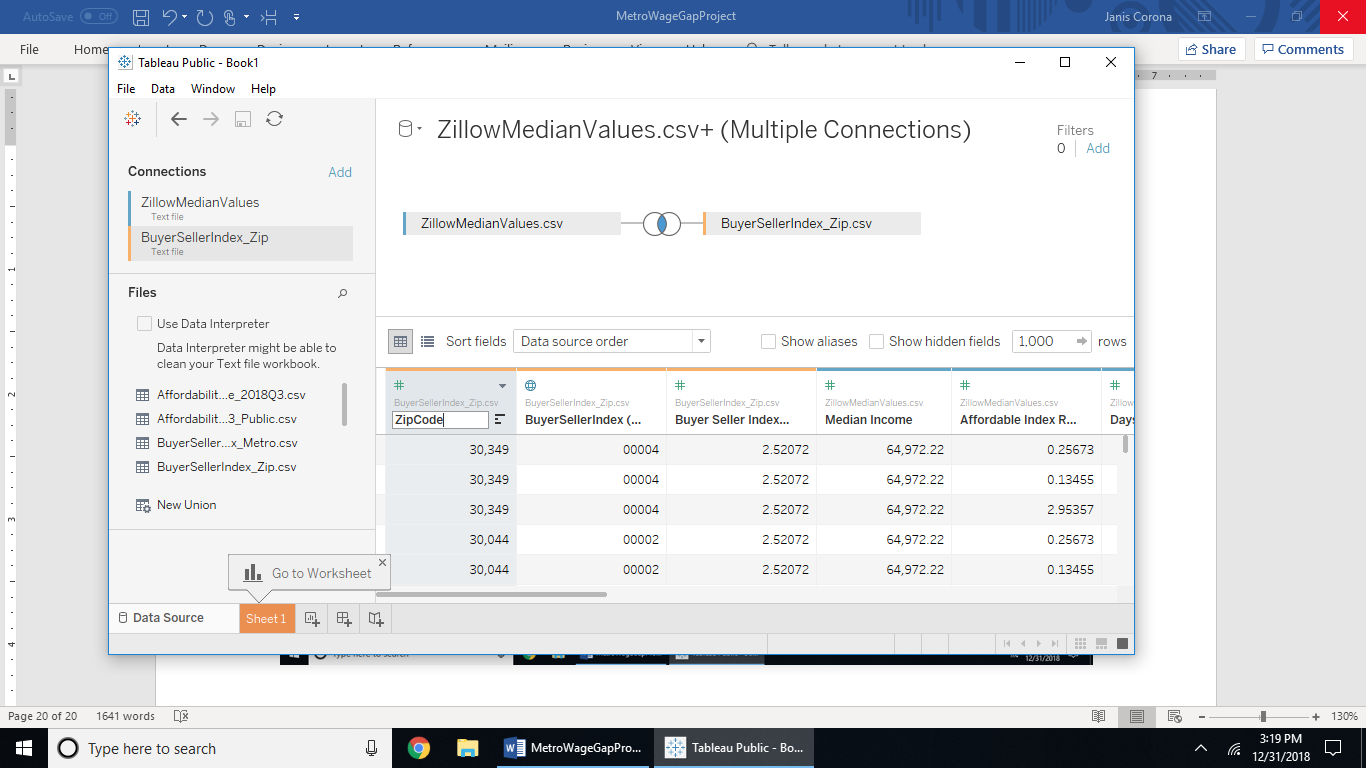
Hide all field other than the Index and 2017-12 field renaming them AffordableIndex and AffordableIndexRatio respectively.



Then export this file to csv as ZillowMedianValues.csv.

Why don’t we make the file with the zip code, so we will add the BuyerSellerIndex\_Zip.csv file from the zip code zip file because it is aligned by CBSA Title fields that we added to our data set earlier. We will remove all other fields in this zip file other than the zip codes inner joined to the ZillowMedianValues.csv file in Tableau merged by CBSA Title.

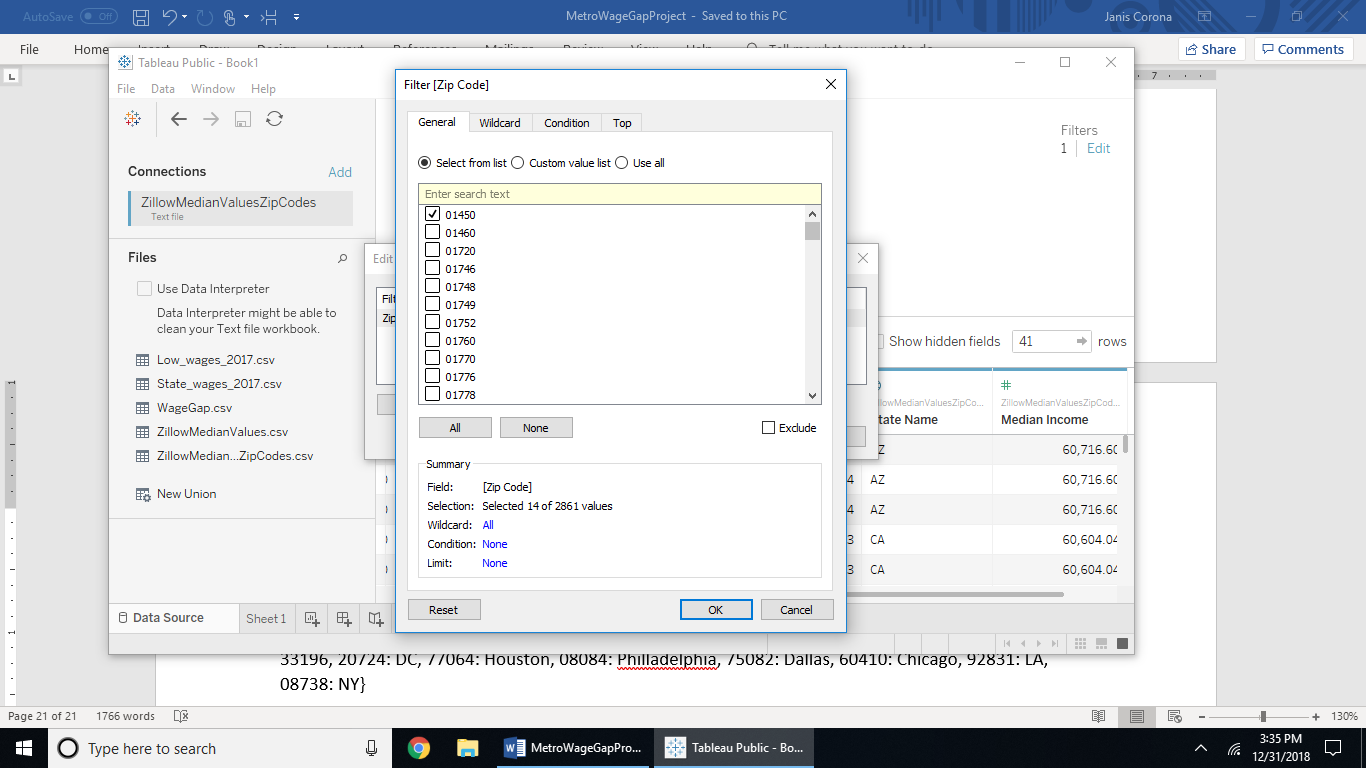




Rename the Region Name1 field ZipCode, and hide all other fields from the zip code file. Export to csv from the Data tab and save as ZillowMedianValuesZipCodes.csv. Close Tableau and open up this file just created.

This file has much more rows because of the many zip codes, go ahead and filter out only one zip code per Size Rank. These zip codes are only for mapping if desired for visualizing locational data. It doesn’t matter what zip codes you choose, but this report chose to filter out these zip codes randomly scrolling downwards by Size Rank:

{85249: Phoenix, 92210: Riverside, 48310: Detroit, 94112: San Francisco, 01450: Boston, 30087: Atlanta, 33196, 20724: DC, 77064: Houston, 08084: Philladelphia, 75082: Dallas, 60410: Chicago, 92831: LA, 08738: NY,33196:Miami}

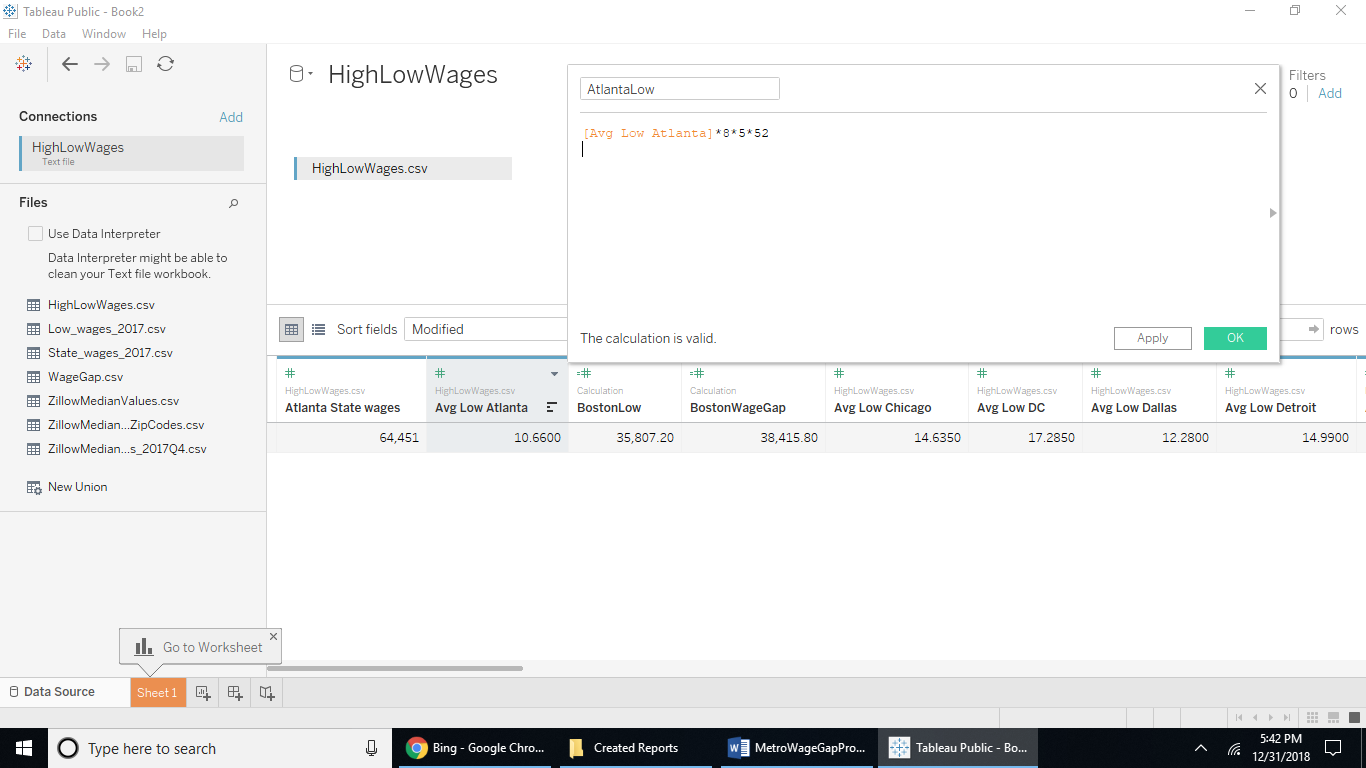


Go to the Data tab, then select Edit Data Source Filters then select the 14 zip codes by Size Rank to shrink the rows in our data. We don’t need all the added zip codes to data we collected by metro aggregates. Export the final version to csv as ZillowMedianValuesZipCodes\_2017Q4.csv, replacing the other version that had too many zip codes. Then close and reopen that last created file in Tableau ZillowMedianValuesZipCodes\_2017Q4.csv.

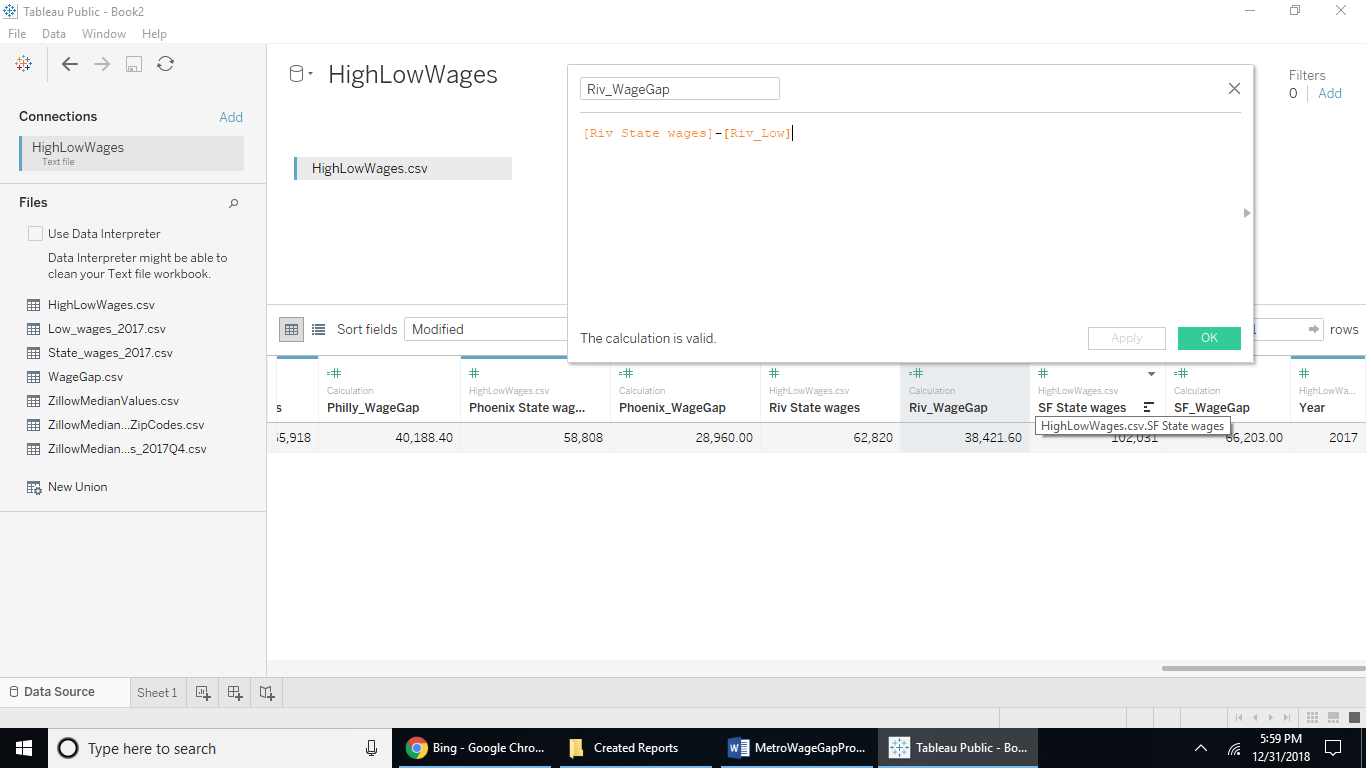
The following is a table of the fields in this data set:

|  |  |  |
| --- | --- | --- |
| Field Name | Table | Remote Field Name |
| ZipCode | BuyerSellerIndex\_Zip.csv | RegionName |
| Median Income | ZillowMedianValues.csv | medianIncome |
| Affordable Index Ratio | ZillowMedianValues.csv | AffordableIndexRatio |
| Days Home Listed | ZillowMedianValues.csv | DaysHomeListed |
| 3BR Home Median | ZillowMedianValues.csv | 3BR\_Home\_Median |
| 3BR Rental Median | ZillowMedianValues.csv | 3BR\_Rental\_Median |
| Home Sales | ZillowMedianValues.csv | HomeSales |
| Foreclosures | ZillowMedianValues.csv | Foreclosures |
| Buyer Seller Index | ZillowMedianValues.csv | BuyerSellerIndex |
| CBSA Title | ZillowMedianValues.csv | CBSA Title |
| Affordable Index | ZillowMedianValues.csv | AffordableIndex |
| Region Name | ZillowMedianValues.csv | Region Name |
| Size Rank | ZillowMedianValues.csv | Size Rank |
| State Name | ZillowMedianValues.csv | StateName |

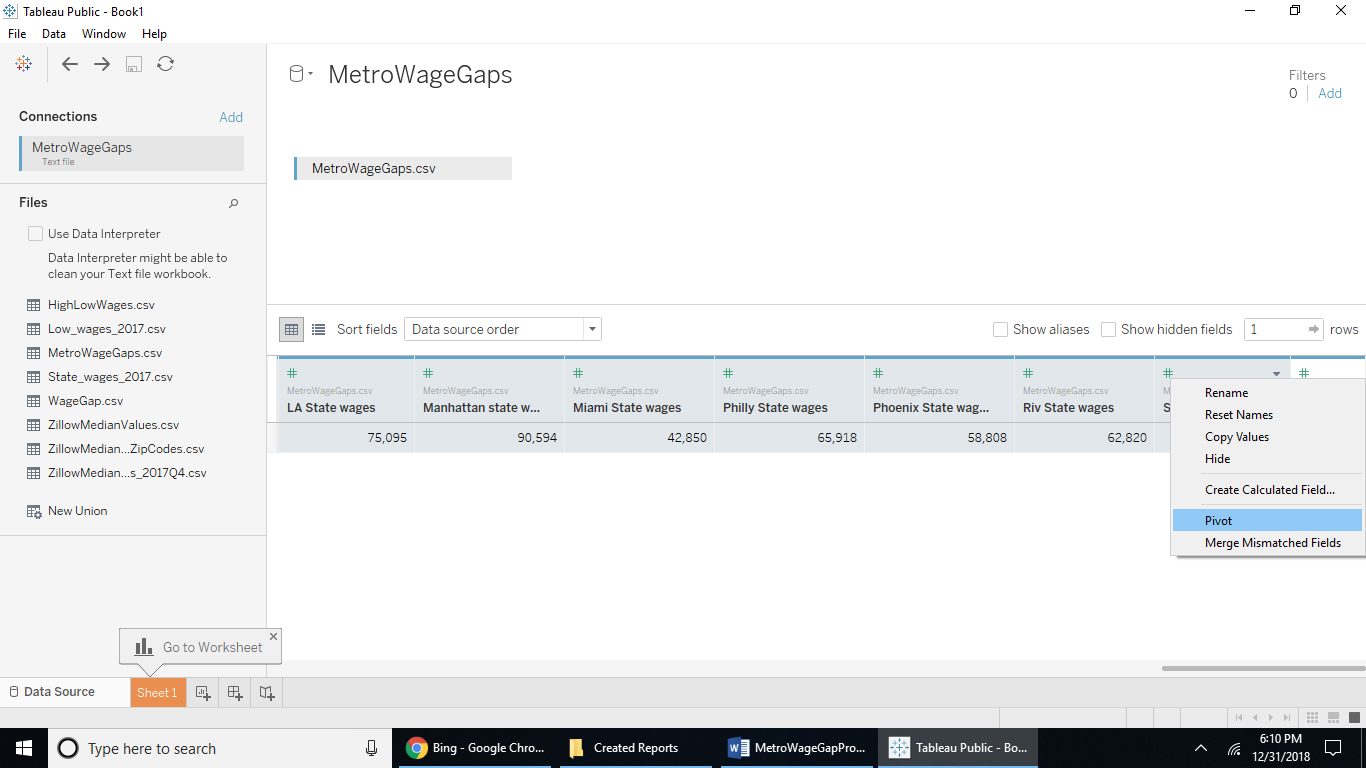
Go back to the HighLowWages.csv file created earlier from the BLS data and make more calculations for the annual of the average low wages per metro and the wage gap between state high wages and low wages by each metro.

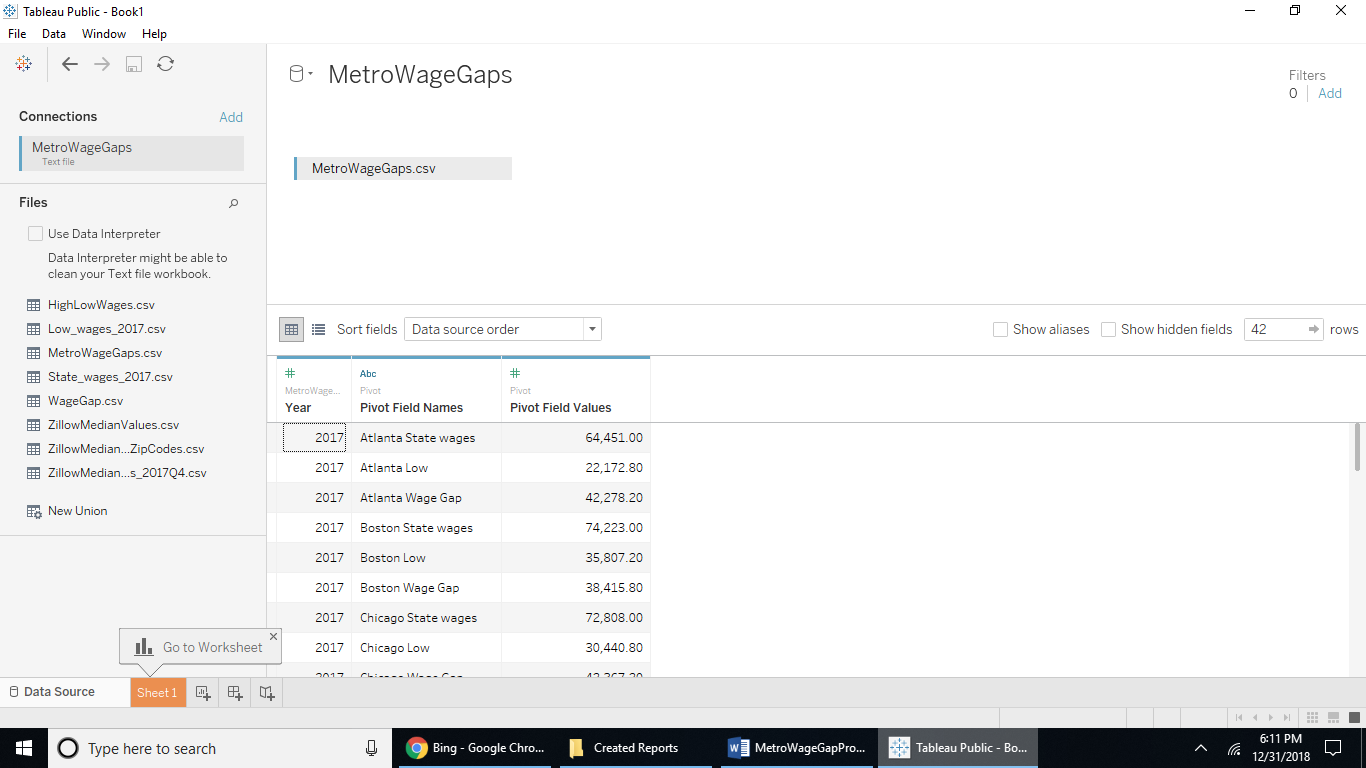
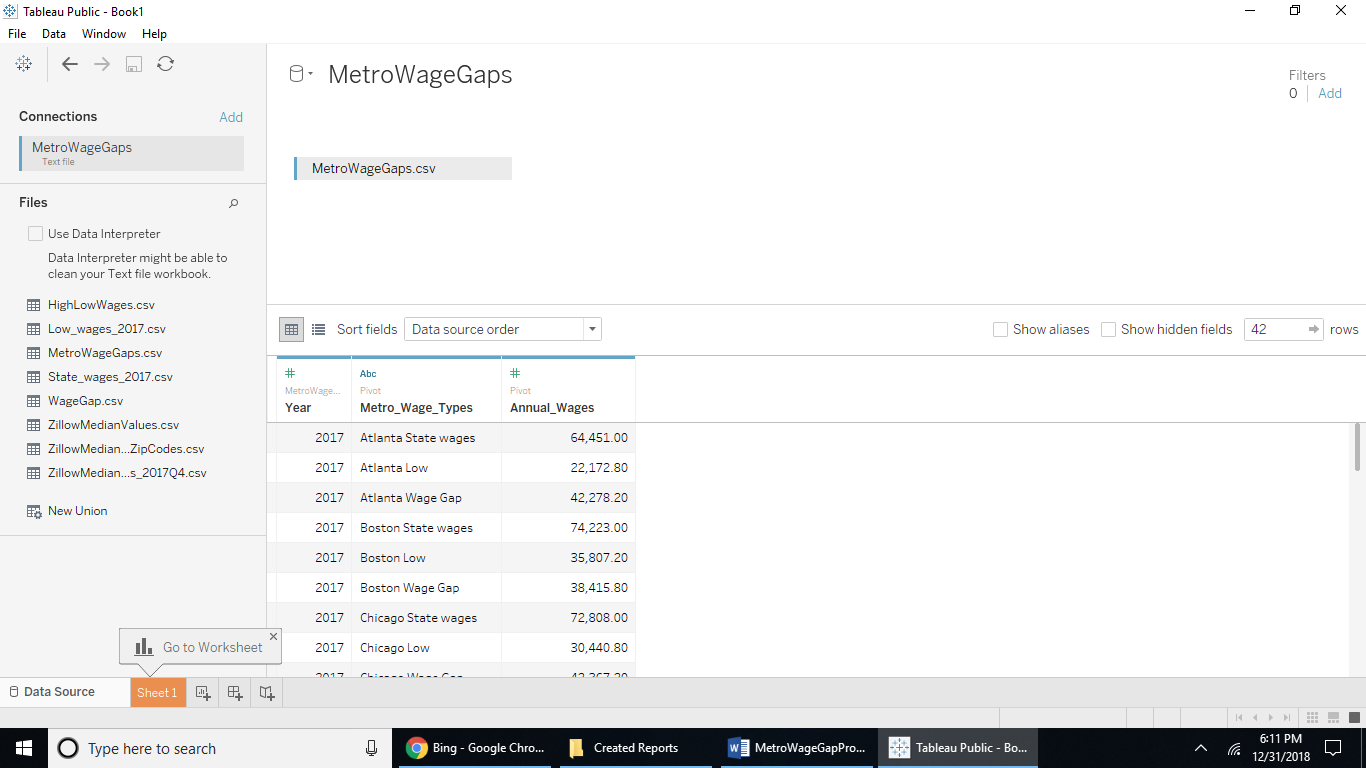


For each Avg Low <metro> field create a calculated field with <metro>Low as the name for annual low income by metro name by multiplying by 8 then by 5 then by 52 by the AvgLow<metro> fields. Then from that field create a field that calculated the wage gap with the <metro>state wage field - <metro>low field and save as <metro>WageGap. Then hide the Avg Low <metro> fields because those are in hourly values not annual like the other fields.

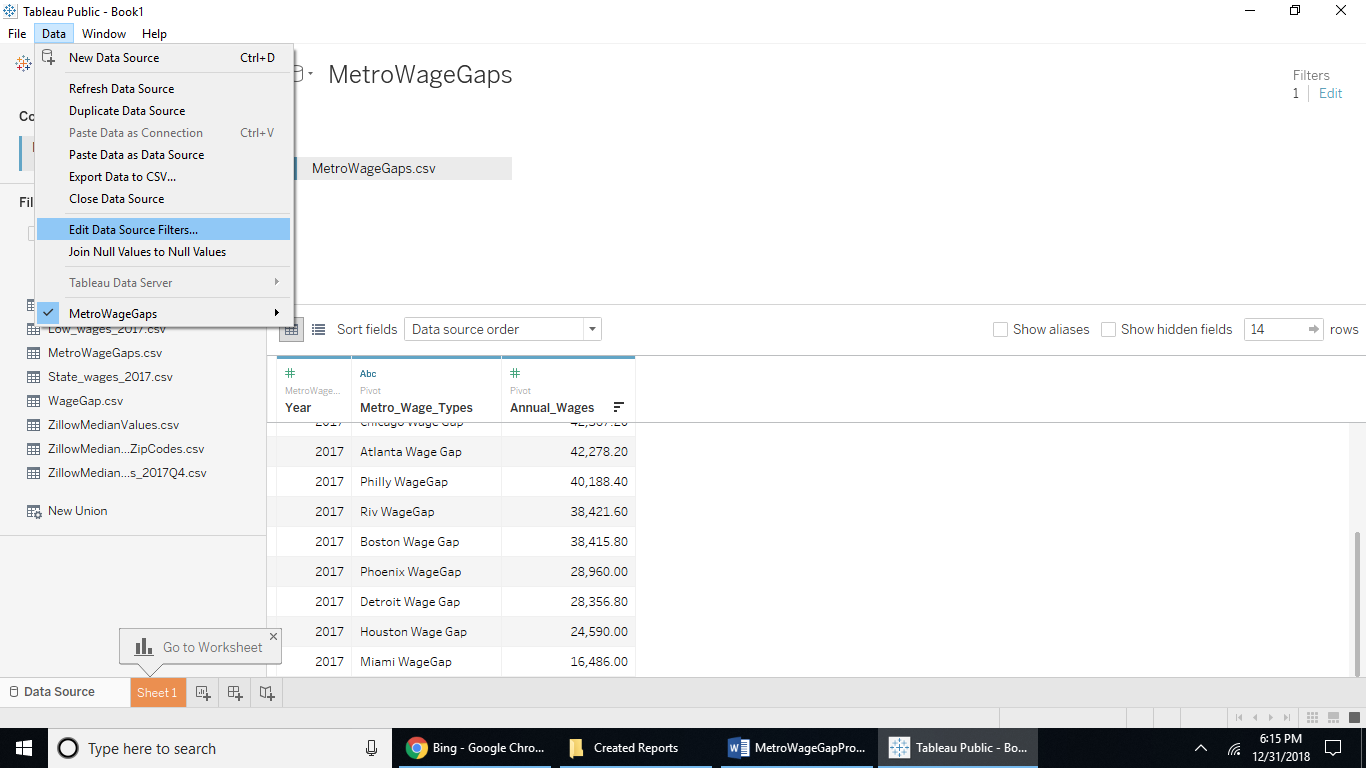


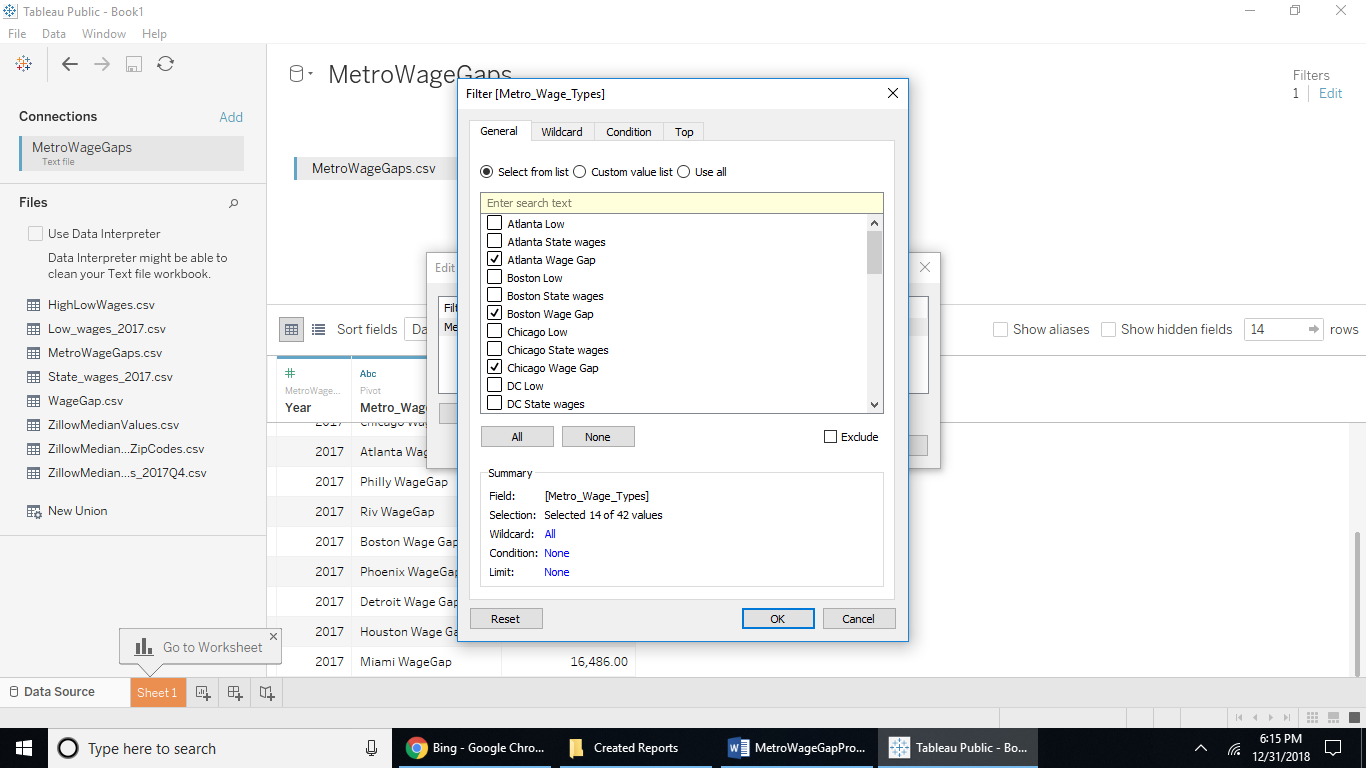
Export to csv from the data tab as MetroWageGaps.csv then reopen Tableau with that file. Pivot all the fields except the year field by shift+selecting the fields, right clicking and selecting Pivot:

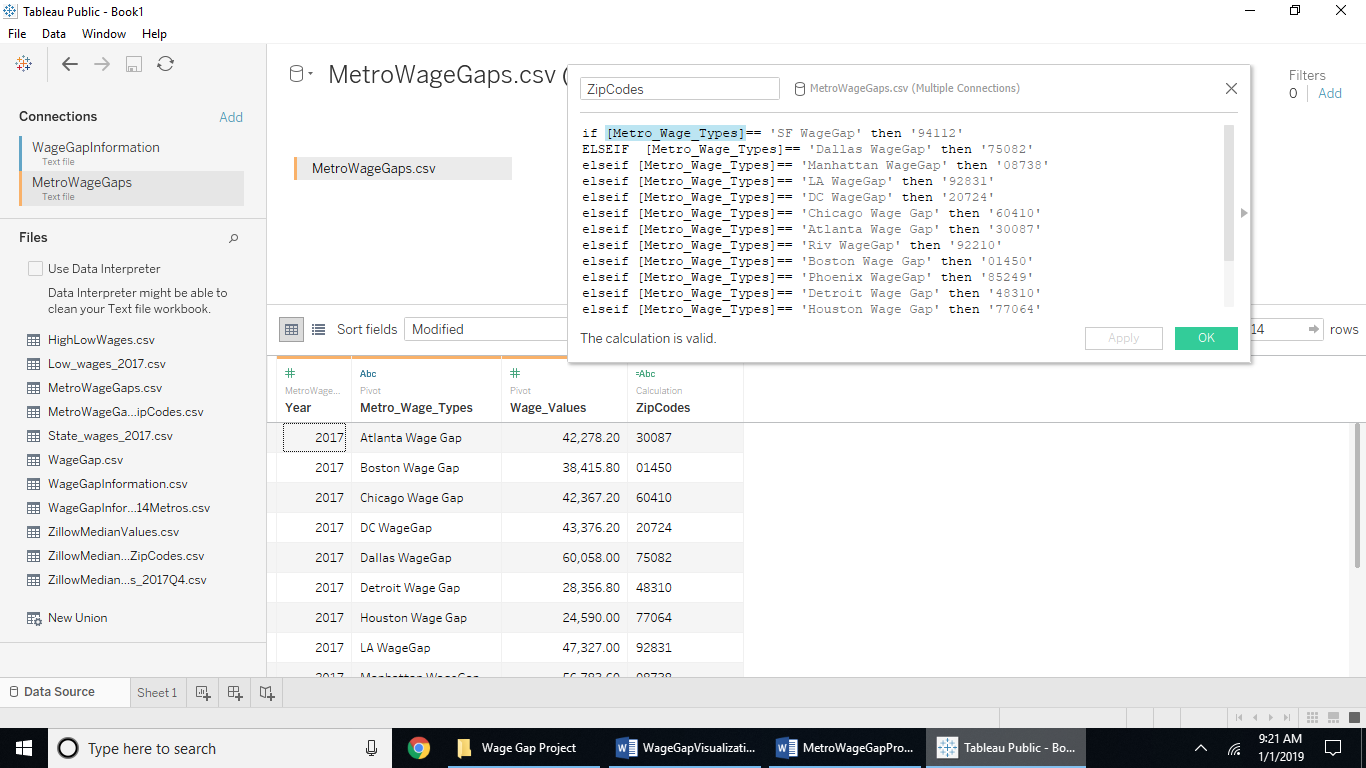


Change the field names to Metro\_Wage\_Types and Annual\_Wages for the corresponding pivot fields just created.

Now filter out only the wage gap metros by using the Data tab and selecting the Edit Data Source Filters option: 

Then scroll and rank the Annual\_Wages field from Highest to lowest. The lowest is Miami at $16,000 and the highest wage gap is San Francisco with $66,000 between state employees and low wage workers. The median wage gap of the top 14 metros by population size is Atlanta with a $42,000 wage gap.



You will need to shift select all the fields except the year field, then right click and select Pivot to get the Wage Gap values for each metro as a row, name the first pivot field ‘Metro\_Wage\_Types’ and the 2nd one ‘Wage\_Values’ then add a calculated field for Zip Codes with the following code.

Using the zip codes we entered randomly in the Zillow data earlier:

{85249: Phoenix, 92210: Riverside, 48310: Detroit, 94112: San Francisco, 01450: Boston, 30087: Atlanta, 33196, 20724: DC, 77064: Houston, 08084: Philladelphia, 75082: Dallas, 60410: Chicago, 92831: LA, 08738: NY, 33196:Miami}

we will add a zip code field to join the BLS wage gap data to the Zillow data for the top 14 metro regions, copy and paste this into a calculated field with the name ZipCode.

if [Metro\_Wage\_Types]== 'SF WageGap' then '94112'

ELSEIF [Metro\_Wage\_Types]== 'Dallas WageGap' then '75082'

elseif [Metro\_Wage\_Types]== 'Manhattan WageGap' then '08738'

elseif [Metro\_Wage\_Types]== 'LA WageGap' then '92831'

elseif [Metro\_Wage\_Types]== 'DC WageGap' then '20724'

elseif [Metro\_Wage\_Types]== 'Chicago Wage Gap' then '60410'

elseif [Metro\_Wage\_Types]== 'Atlanta Wage Gap' then '30087'

elseif [Metro\_Wage\_Types]== 'Riv WageGap' then '92210'

elseif [Metro\_Wage\_Types]== 'Boston Wage Gap' then '01450'

elseif [Metro\_Wage\_Types]== 'Phoenix WageGap' then '85249'

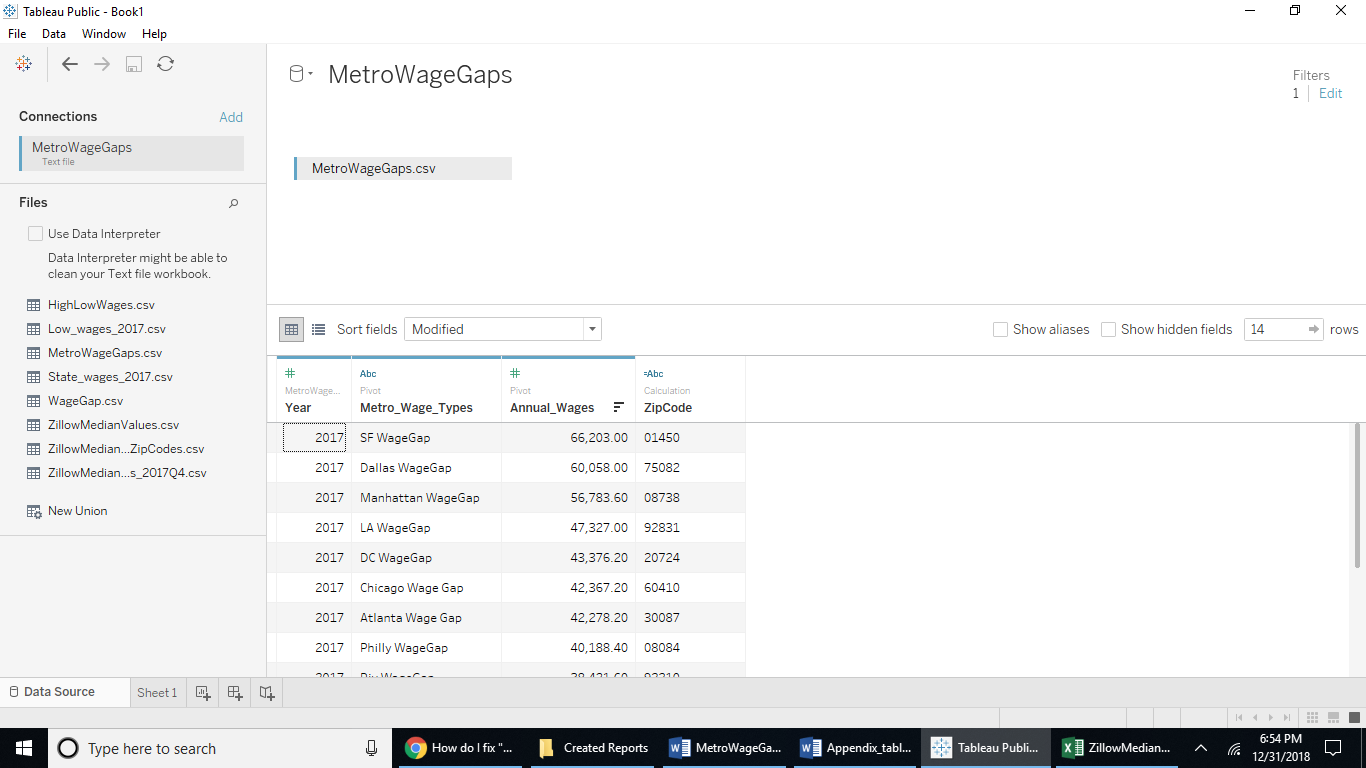
elseif [Metro\_Wage\_Types]== 'Detroit Wage Gap' then '48310'

elseif [Metro\_Wage\_Types]== 'Houston Wage Gap' then '77064'

elseif [Metro\_Wage\_Types]== 'Miami WageGap' then '33196'

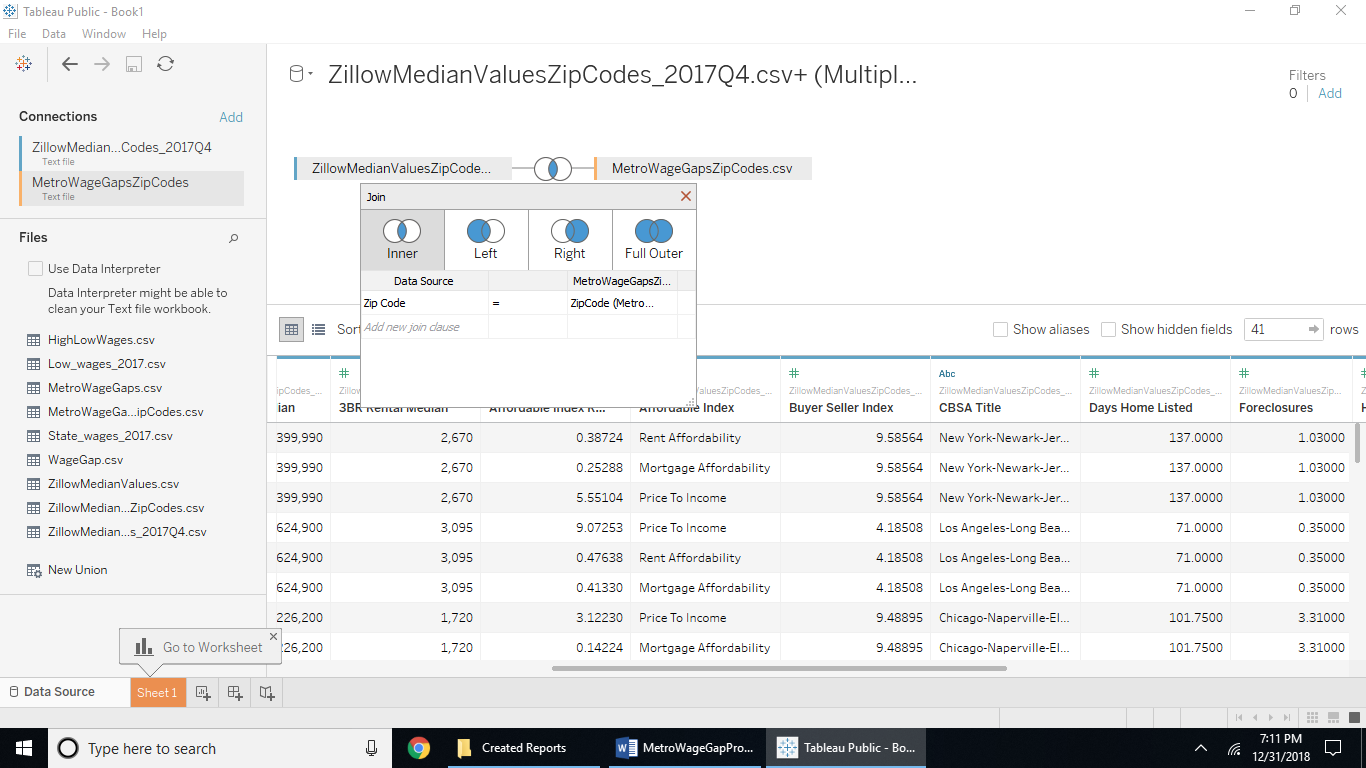
ELSEIF [Metro\_Wage\_Types]== 'Philly WageGap' then '08084'

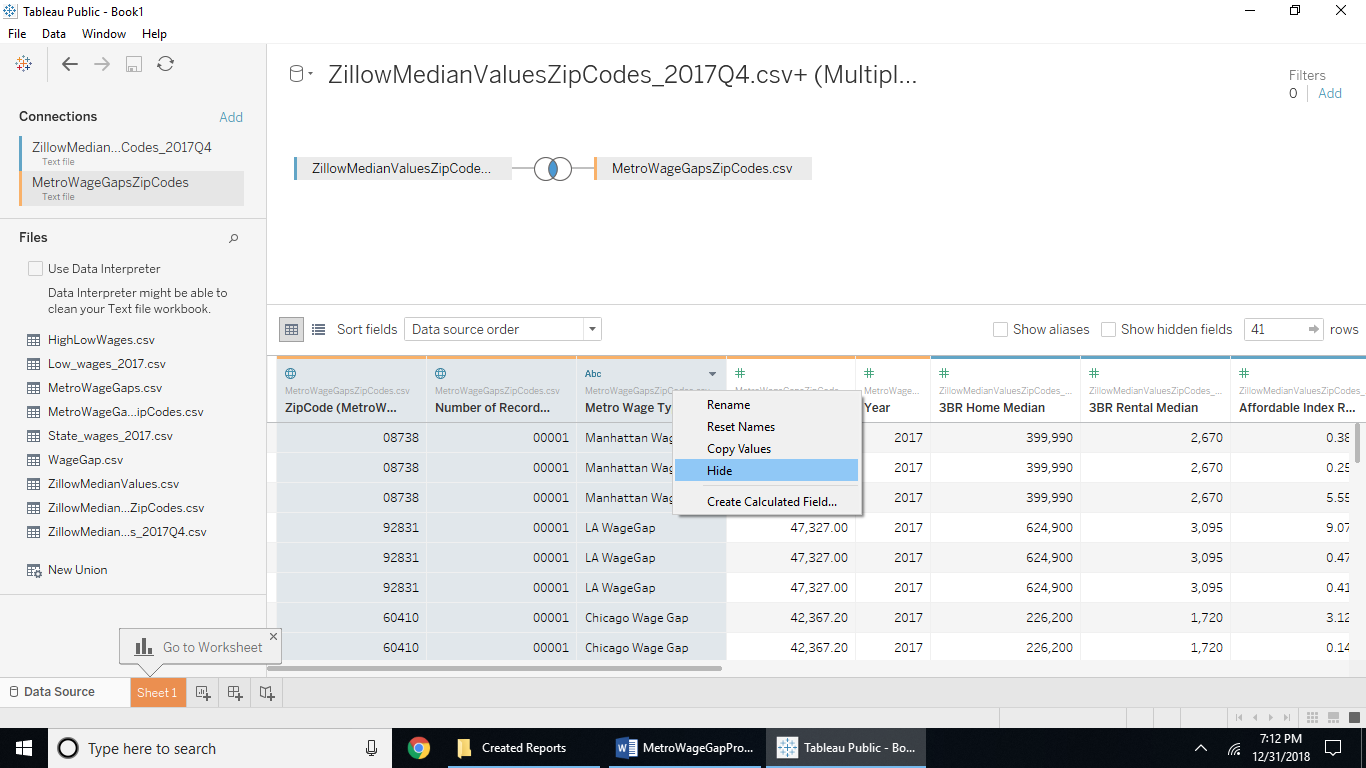
end

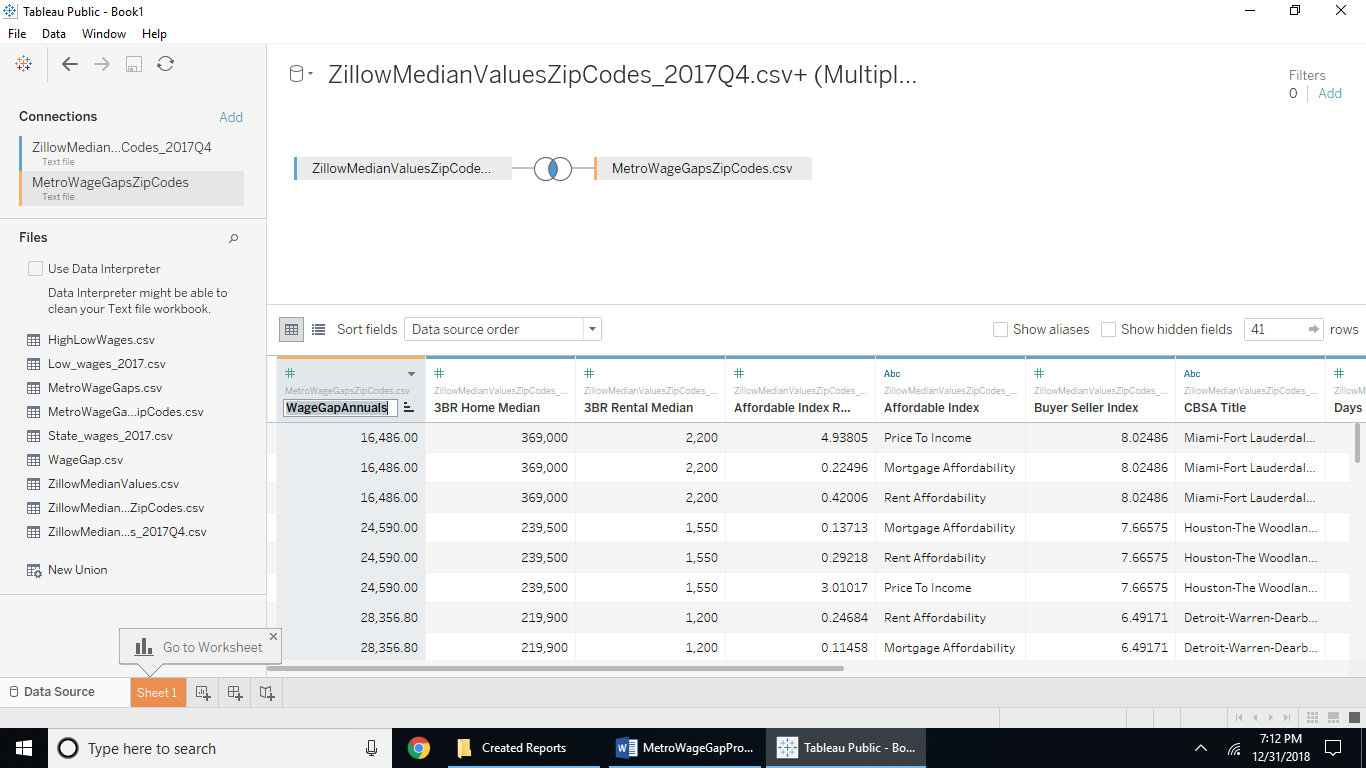


Export as csv from the Data tab with the file name MetroWageGapsZipCodes.csv.

Joing the MetroWageGapsZipCodes.csv and the ZillowMedianValuesZipCodes\_2017Q4 by ZipCode field names then hide the number of records field, one of the duplicate zipcode fields, The Metro\_Wage\_Types field and change the name of the Wage Gaps data field named WageValues to ‘Wage Gap Annuals’ to clarify this field is the wage gaps per metro. Gap represents the low pay workers of cashiers and security guards to state employee higher wages annually.







Now the data set is complete. This completes the summary on building the data table from the given link references to each data set, calculated fields, joining the subsets of data to the completed data table. Save this data set as WageGapInformationTop14Metros.csv

|  |  |  |
| --- | --- | --- |
| Field Name | Table | Remote Field Name |
| 3BR Home Median | WageGapInformationTop14Metros.csv | 3BR Home Median |
| 3BR Rental Median | WageGapInformationTop14Metros.csv | 3BR Rental Median |
| Affordable Index Ratio | WageGapInformationTop14Metros.csv | Affordable Index Ratio |
| Affordable Index | WageGapInformationTop14Metros.csv | Affordable Index |
| Wage Gap Annuals | WageGapInformationTop14Metros.csv | WageGapAnnuals |
| Buyer Seller Index | WageGapInformationTop14Metros.csv | Buyer Seller Index |
| CBSA Title | WageGapInformationTop14Metros.csv | CBSA Title |
| Days Home Listed | WageGapInformationTop14Metros.csv | Days Home Listed |
| Foreclosures | WageGapInformationTop14Metros.csv | Foreclosures |
| Home Sales | WageGapInformationTop14Metros.csv | Home Sales |
| Median Income | WageGapInformationTop14Metros.csv | Median Income |
| Region Name | WageGapInformationTop14Metros.csv | Region Name |
| Size Rank | WageGapInformationTop14Metros.csv | Size Rank |
| State Name | WageGapInformationTop14Metros.csv | State Name |
| Zip Code | WageGapInformationTop14Metros.csv | Zip Code |
| CBSA Title | WageGapInformationTop14Metros.csv | CBSA Title |
| Days Home Listed | WageGapInformationTop14Metros.csv | Days Home Listed |
| Foreclosures | WageGapInformationTop14Metros.csv | Foreclosures |