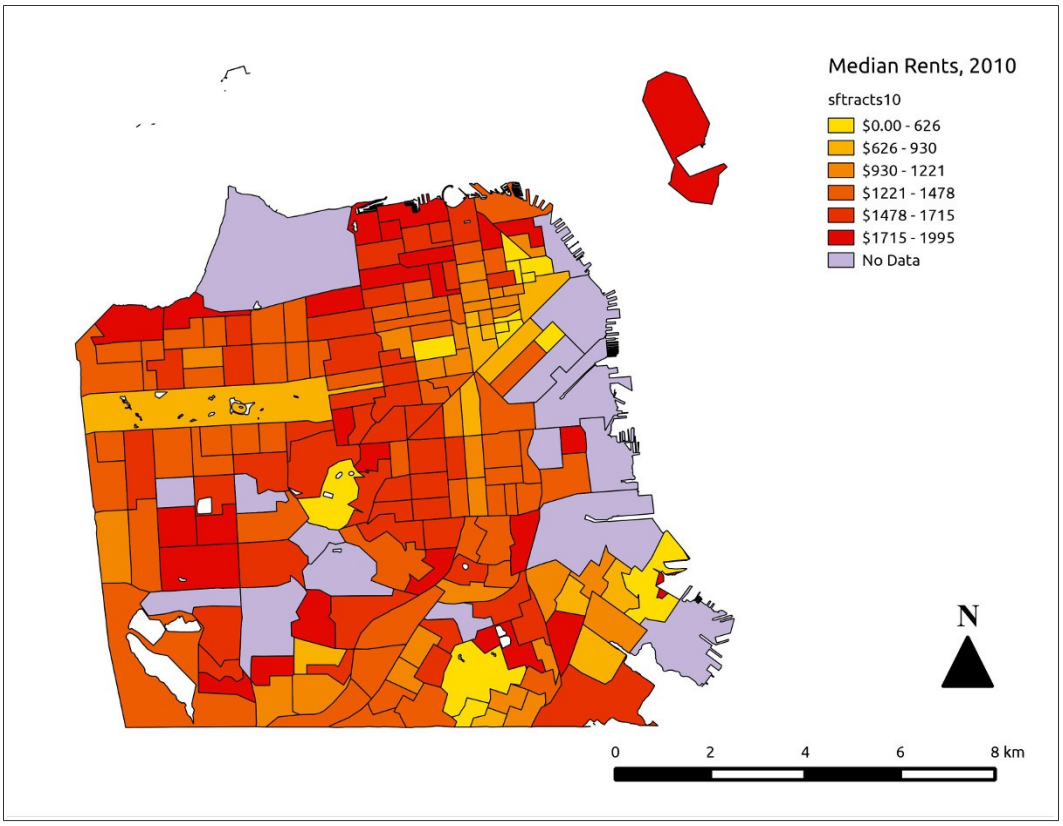
USP/PLSI 492: Research Methods
Instructor: Pietro Calogero

Making shapefiles pretty in QGIS

In this third part of this process, we are going to make a map like the one below:

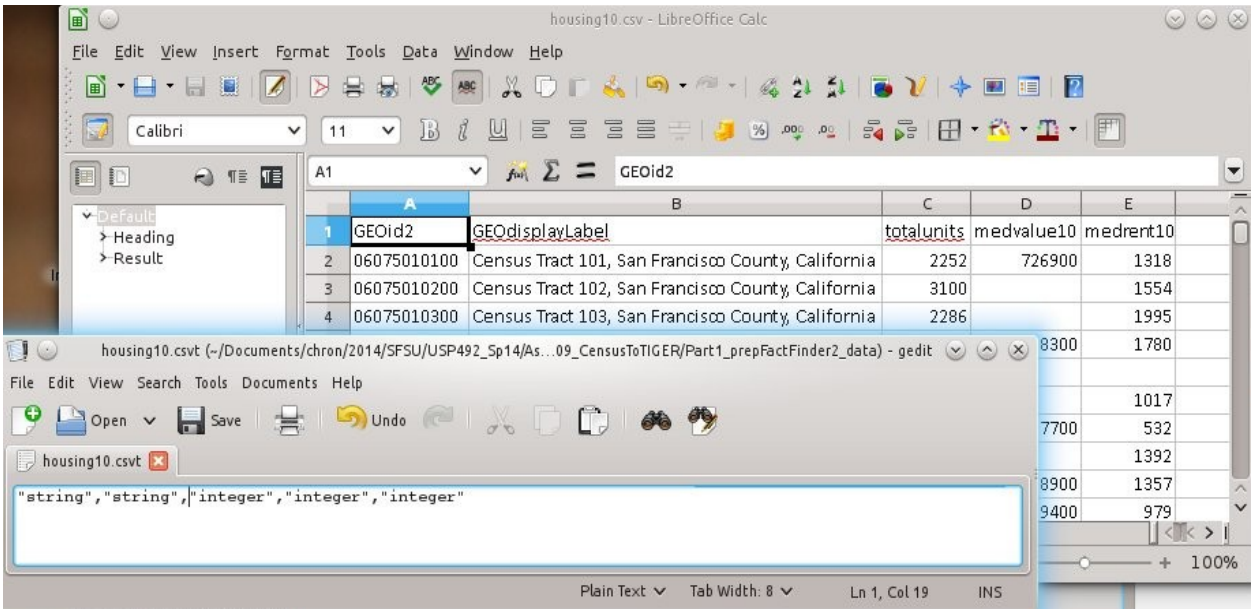


GIS can be used for very sophisticated data analysis of many types. But the most fundamental purpose of GIS is data-visualization, which we use for policy analysis and development. In this case I have taken both tabular data and map data from the US Census website to show median rents by census-tract in San Francisco in 2010. The data may need some checking, and missing data needs to be filled in or explained. But even as it is, such a map can already be used as the basis for studies of gentrification and relative affordability in different parts of the city.

In previous exercises we have learned how to get and format data and TIGER files from the U.S. Census. Here we will bring these data together.

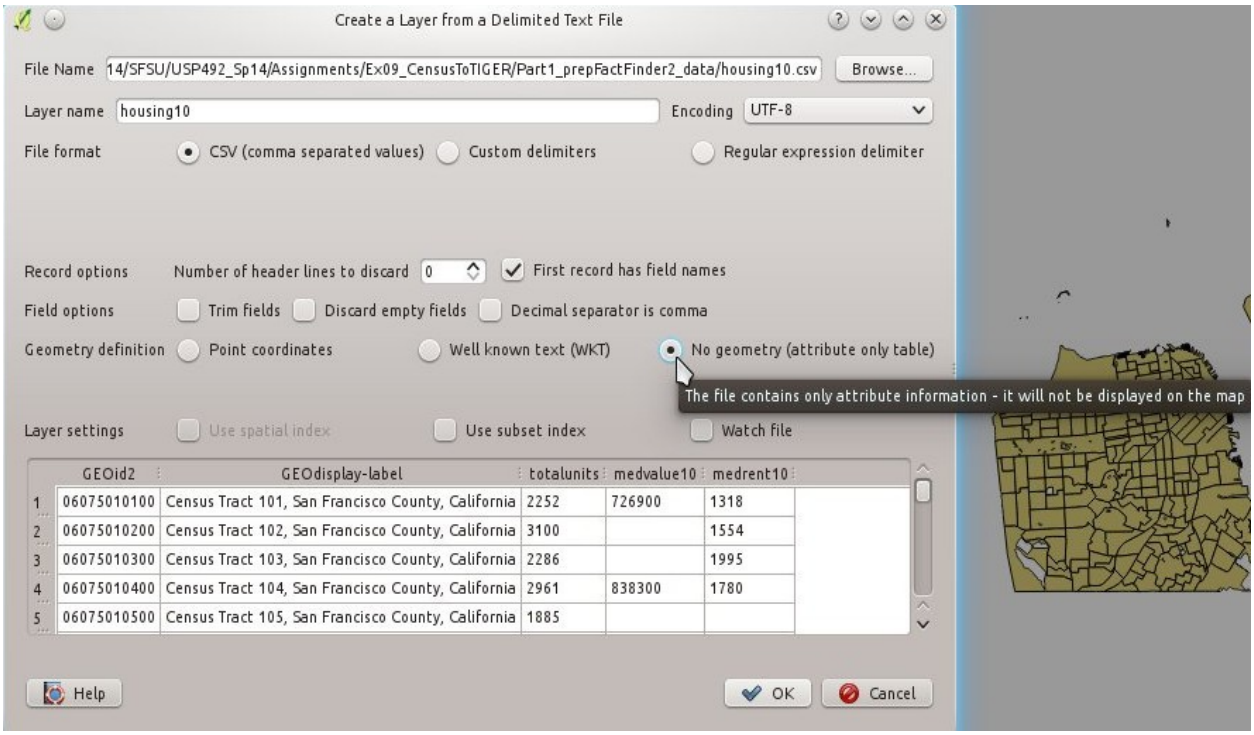
1. Join the American FactFinder (AFF) data to your shapefile.

In QGIS, this requires some additional steps not needed in ArcMAP. You will need to look at your data structure, then create a text file which describes your data structure, called a CSVT file. You will need to save this in the same folder as the CSV file, with the same filename before the dot. QGIS will read the CSVT file and use it as a guide for how to import the CSV data.

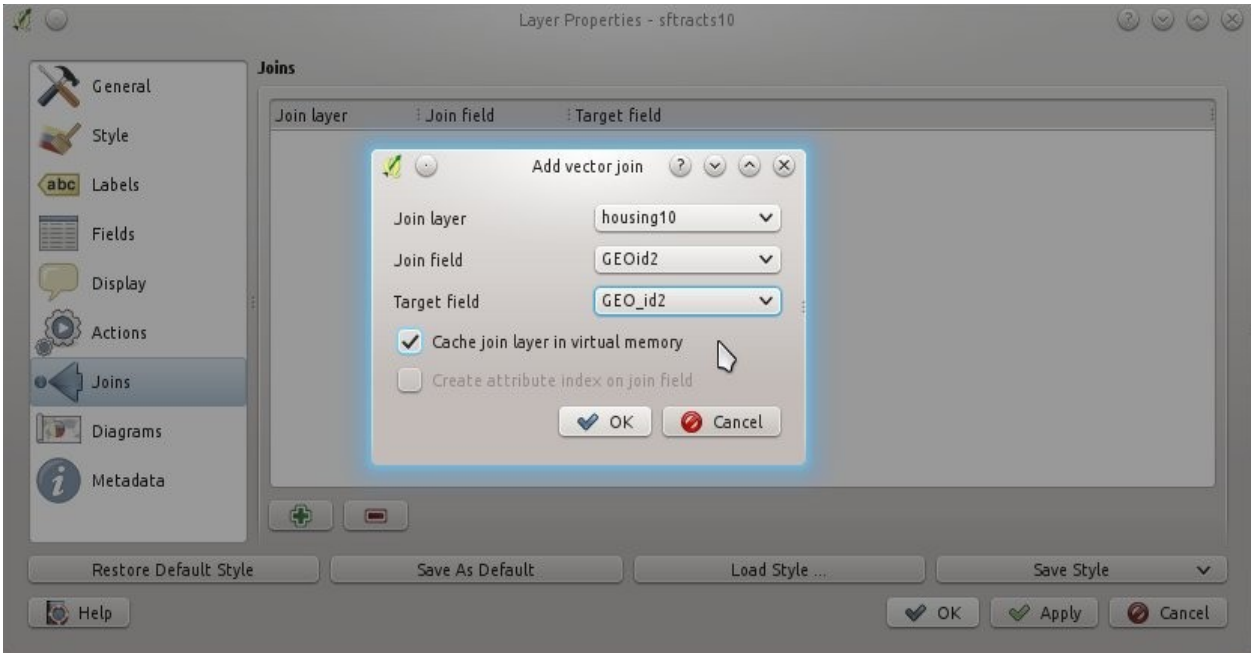


Open a very basic text editor, such as Notepad on Windows (here I am using Gedit because I am on Linux). Type in the data-type for each column. In this example, my first two columns should be read as text-strings, not as calculable numbers. So I call them “string” and “string”, separated only by a comma (no spaces). In the last three columns there is numeric data, and I want QGIS to read it as whole integers. So I add “integer”,“integer”,“integer” to specify that these three columns should be read as whole-number integers. I save this file with the same main filename, but with a different suffix from the original CSV file: I save it as **Housing10.csvt**.

Now I can add the CSV file to the QGIS workspace. QGIS will also read the .CSVT file use it as the guide for how to interpret the data-type in the CSV. In the screenshot below, note that I specify the **Geometry definition** as “No geometry (attribute only table)”. Then it will show up in the **Layers** pane on the left, but of course nothing new will show up in the main window, because there are no spatial entities (no points, no lines, no polygons) associated with this CSV file.



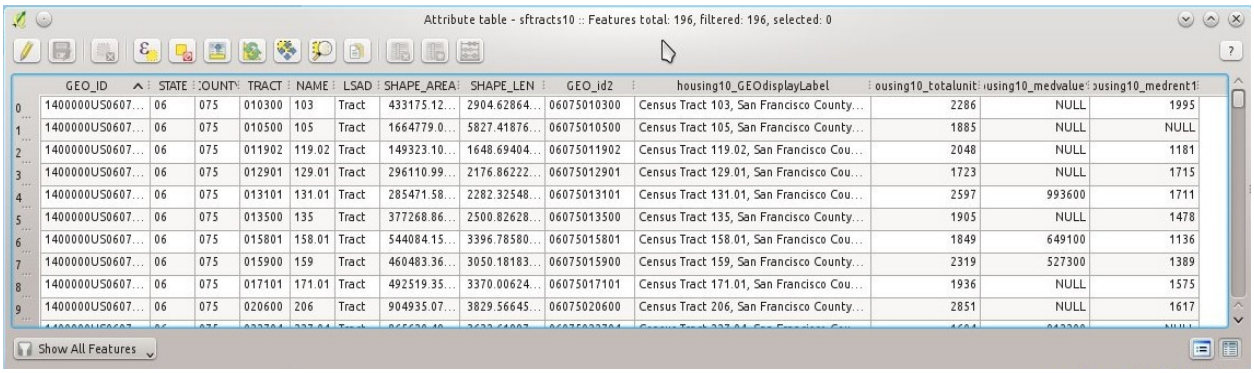
Next, in the **Layers** pane, select the shapefile of census-tracts. Then either double click it or right-click and choose “Properties...” In the Properties window, in the left pane, one of the options is **Joins**. Click that. Initially the right-hand pane will be blank. Click the green plus-sign at the bottom, and it will invoke the **Add vector join** sub-window:



The **Join field** in the CSV file is the first column, which is what appears in the drop-down options by default. In the **Target field,** the first column of the shapefile’s attribute table also appears by default. But

this is not the one we will use. Instead, we will choose “GEO_id2”, which was the field we created in the last exercise by concatenating the State, County, and Tract codes into a single, 11-character geocode. Hit OK in both the sub-window and the **Layer Properties** window.

Now open the Attribute table of the census-tracts shapefile again. The additional columns from the CSV file should be added. If all the values are null, that means either the text format in the CSV headers or the data-format in the corresponding CSVT file are somehow off; try reviewing those carefully.

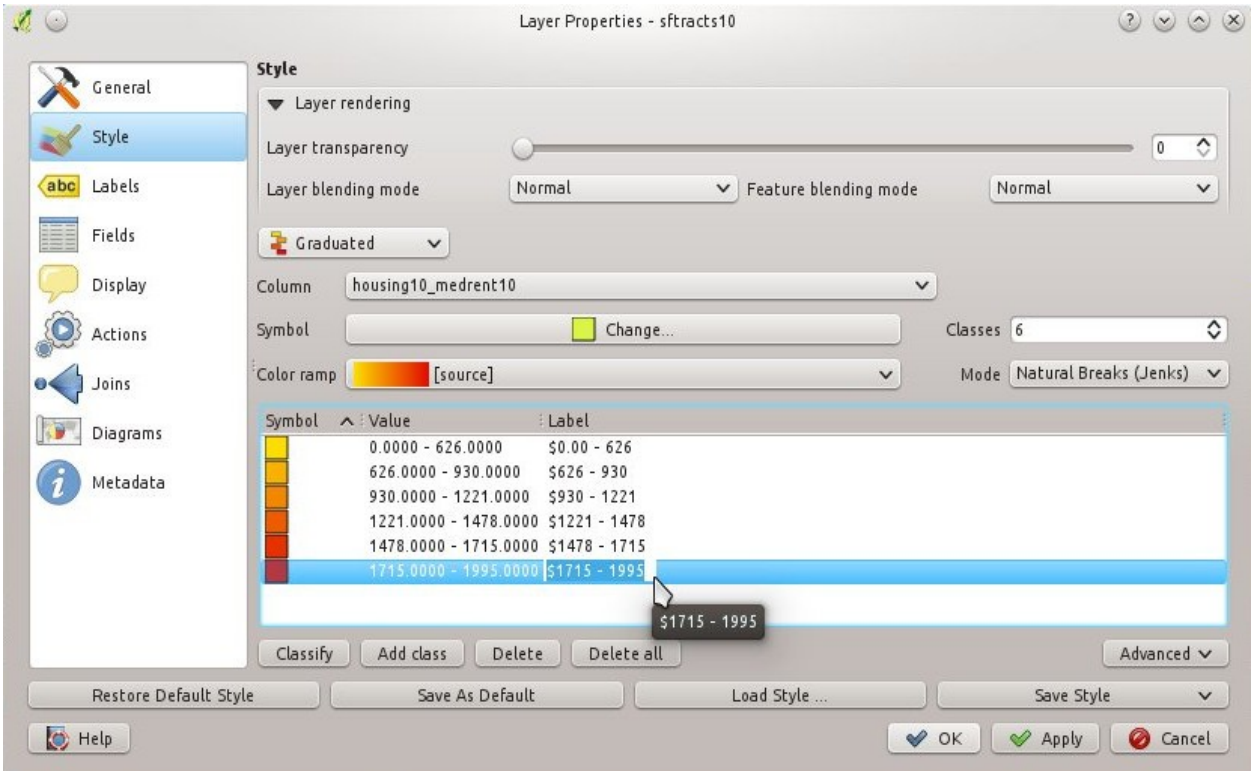


	GEO_ID	STATE	COUNTY	TRACT	NAME	LSAD	SHAPE_AREA	SHAPE_LEN	GEO_id2	housing10_GEOdisplayLabel	ousing10_totalunit	using10_medvalue	ousing10_medrent1
0	1400000US0607...	06	075	010300	103	Tract	433175.12...	2904.62864...	06075010300	Census Tract 103, San Francisco County...	2286	NULL	1995
1	1400000US0607...	06	075	010500	105	Tract	1664779.0...	5827.41876...	06075010500	Census Tract 105, San Francisco County...	1885	NULL	NULL
2	1400000US0607...	06	075	011902	119.02	Tract	149323.10...	1648.69404...	06075011902	Census Tract 119.02, San Francisco Cou...	2048	NULL	1181
3	1400000US0607...	06	075	012901	129.01	Tract	296110.99...	2176.86222...	06075012901	Census Tract 129.01, San Francisco Cou...	1723	NULL	1715
4	1400000US0607...	06	075	013101	131.01	Tract	285471.58...	2282.32548...	06075013101	Census Tract 131.01, San Francisco Cou...	2597	993600	1711
5	1400000US0607...	06	075	013500	135	Tract	377268.86...	2500.82628...	06075013500	Census Tract 135, San Francisco County...	1905	NULL	1478
6	1400000US0607...	06	075	015801	158.01	Tract	544084.15...	3396.78580...	06075015801	Census Tract 158.01, San Francisco Cou...	1849	649100	1136
7	1400000US0607...	06	075	015900	159	Tract	460483.36...	3050.18183...	06075015900	Census Tract 159, San Francisco County...	2319	527300	1389
8	1400000US0607...	06	075	017101	171.01	Tract	492519.35...	3370.00624...	06075017101	Census Tract 171.01, San Francisco Cou...	1936	NULL	1575
9	1400000US0607...	06	075	020600	206	Tract	904935.07...	3829.56645...	06075020600	Census Tract 206, San Francisco County...	2851	NULL	1617

Note that in this example, there are a few NULL data cells because the original census data had some NULL cells. In this real-world example, our data-set is not perfect, not complete. In the initial graphic on page 1, note that I explicitly showed tracts with no data.

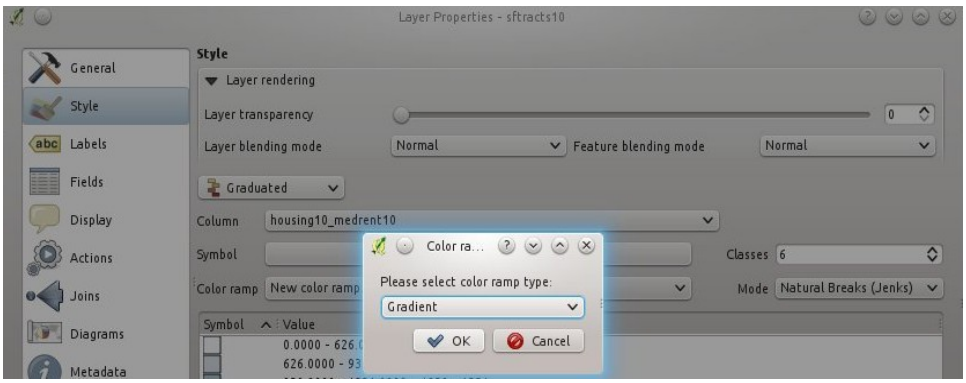
[I continue with the exercise on the next page]

2. Create a Choropleth Map

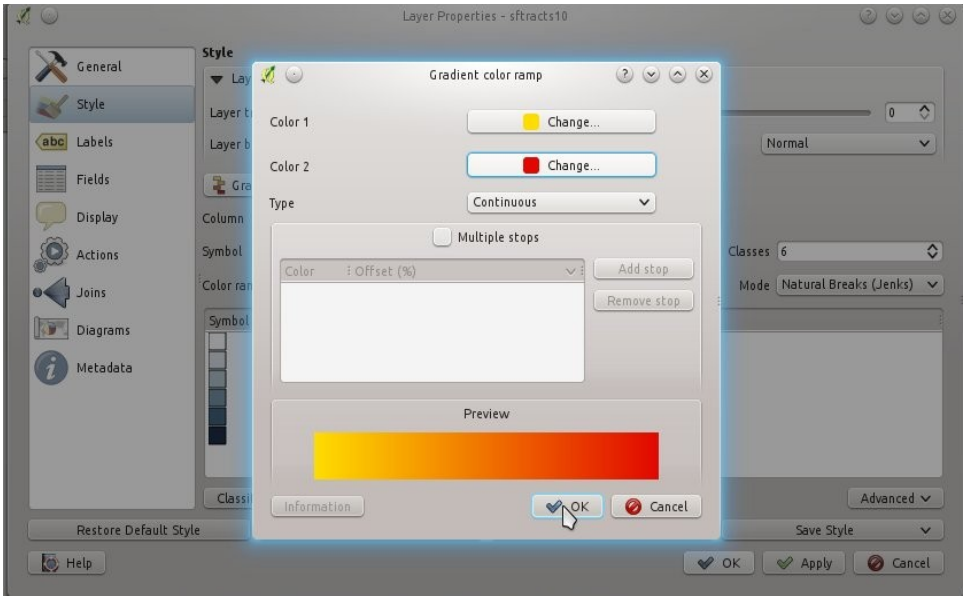


In this map I am going to show increasing rents with increasingly dark blue.

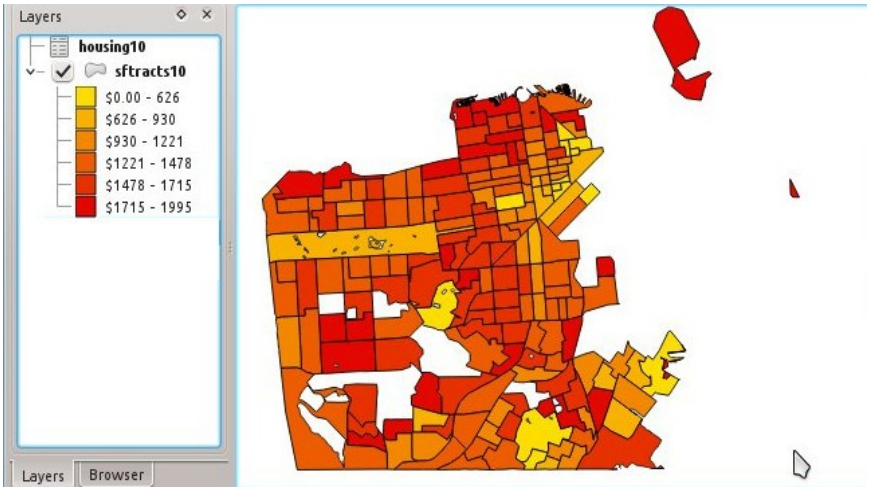
1) I go to the shapefiles Properties by double-clicking on it.
2) I select the **Style** tab in the left pane.
3) I change the area-fill of the polygons from “Single symbol” to “Graduated”.
4) I choose the data-column that will be the basis for colors: “housing10_medrent10”
5) For the **Color ramp,** I created a custom yellow-red gradient (explained below)
6) For **Classes** of data I select “6”
7) For **Mode** of data-classification I select “Natural Breaks (Jenks)”
8) Finally, in the **Label** column, I manually retype the values into more human-readable format. This is how the classes will be displayed in any graphics I produce, so it is worthwhile doing this now.



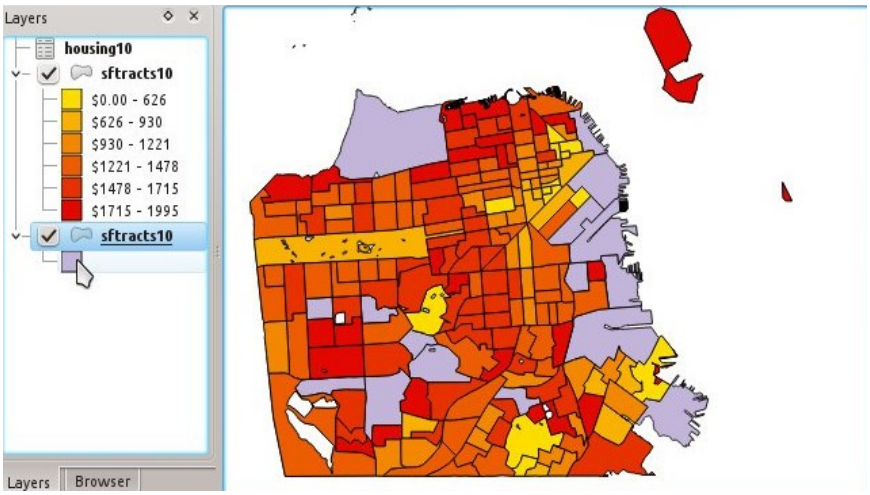
In the **Color ramp** drop-down field I scroll to the bottom and select “New color ramp...” I choose “Gradient”. In the next window, I select Yellow as the first, and Red as the final color in the gradient:



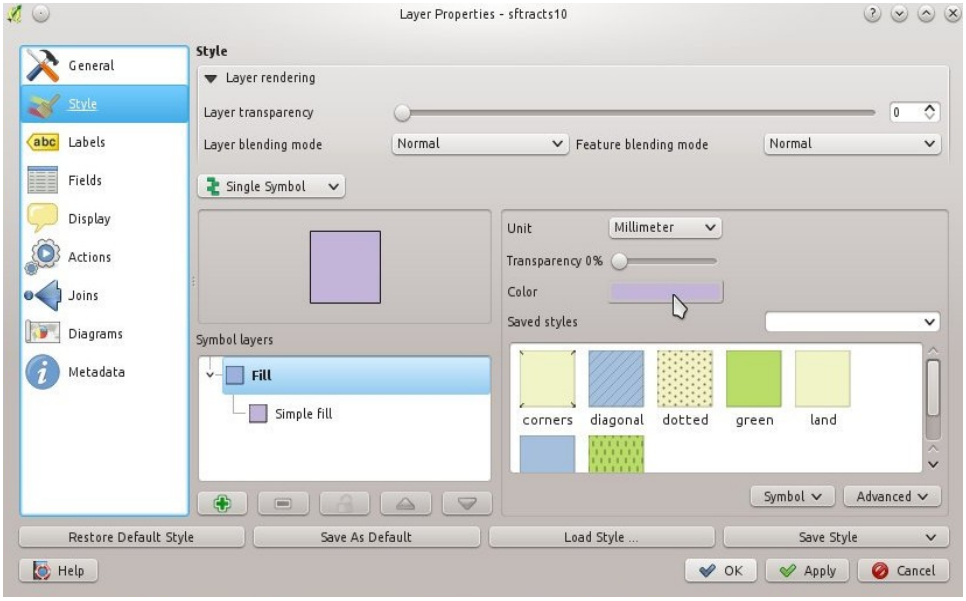
The resulting map (below) looks a little weird. Census tracts with no data in the median-rents column don’t show up at all. This also happens with ArcMAP, so the way to clarify your data-display is the same in either program.



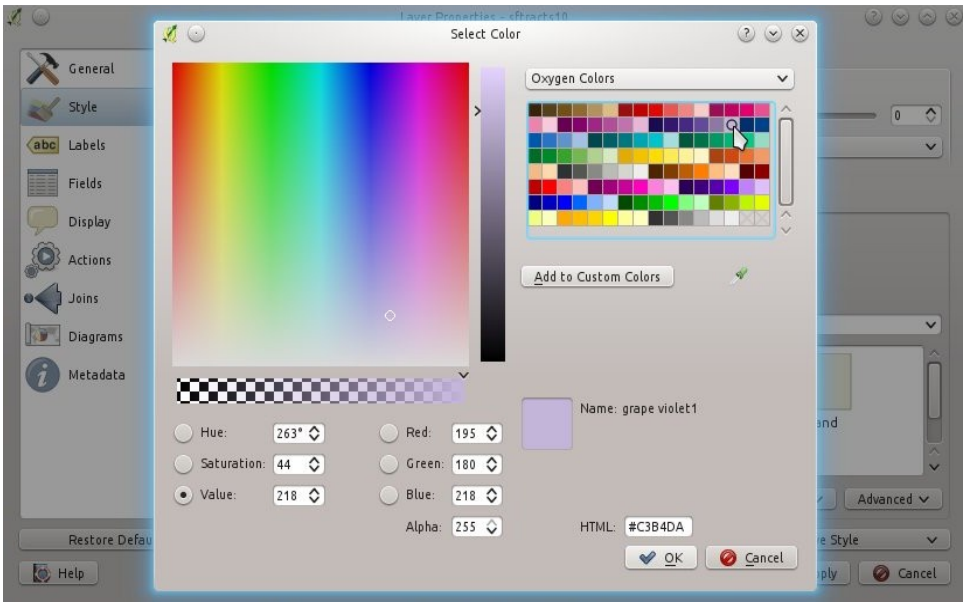
Census-tracts with NULL data in the median-rent field are not displayed at all. This makes the map hard to interpret. I would rather show that they have no data in them, by showing them very differently. What I do is place a whole second copy of exactly the same shapefile ***underneath*** the yellow-red shapefile. All the polygons in the underlay shapefile are colored lavender. Notice how the Presidio and Hunter’s Point are now visible:



I set the NULL color to be very distinct from the yellow-red gradient of the tracts which do have data. To do this, I open up the **Properties** of the underlying shapefile. In the **Style** pane I leave the fill-color as "Single Symbol", and in the **color** field (in the right-hand sub-pane) is set the fill color.



and then:



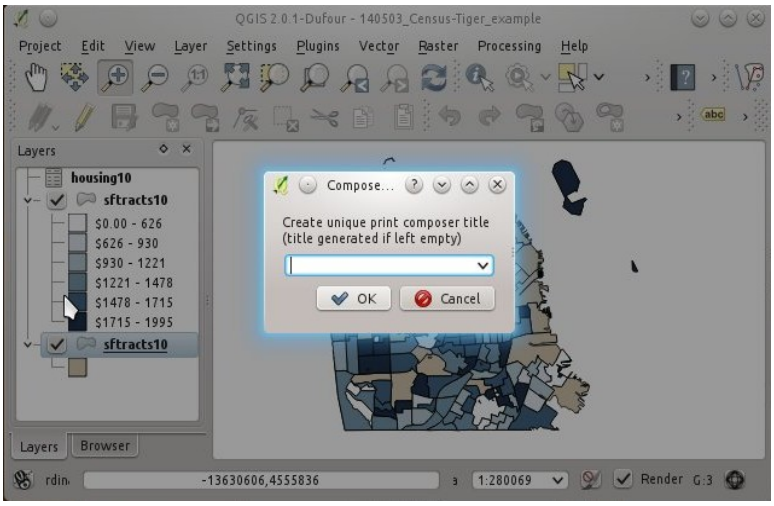
I now have a choropleth map of median rents in each census tract of San Francisco.

3. Formatting as a presentation-map

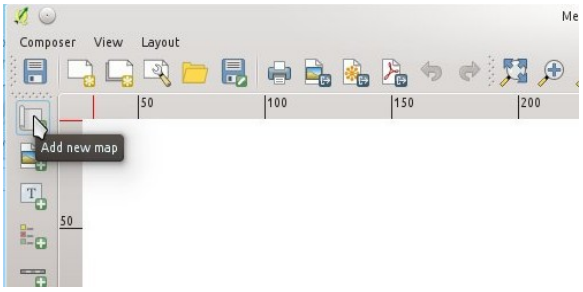
In ArcMAP, you toggle from Data View into Layout View in order to compose a graphic output. That is similar to AutoCAD's "layerspace" versus "paperspace". In QGIS, you use a "Print Composer" instead. The only advantage to the Print Composer is that you can set up multiple compositions of the same data, so that you can manage two different graphic outputs such as "before and after" maps of the same area.

Go to **Project > New print composer...** and name your new output. I named mine "MedianRents10" because this particular choropleth map only shows that data.

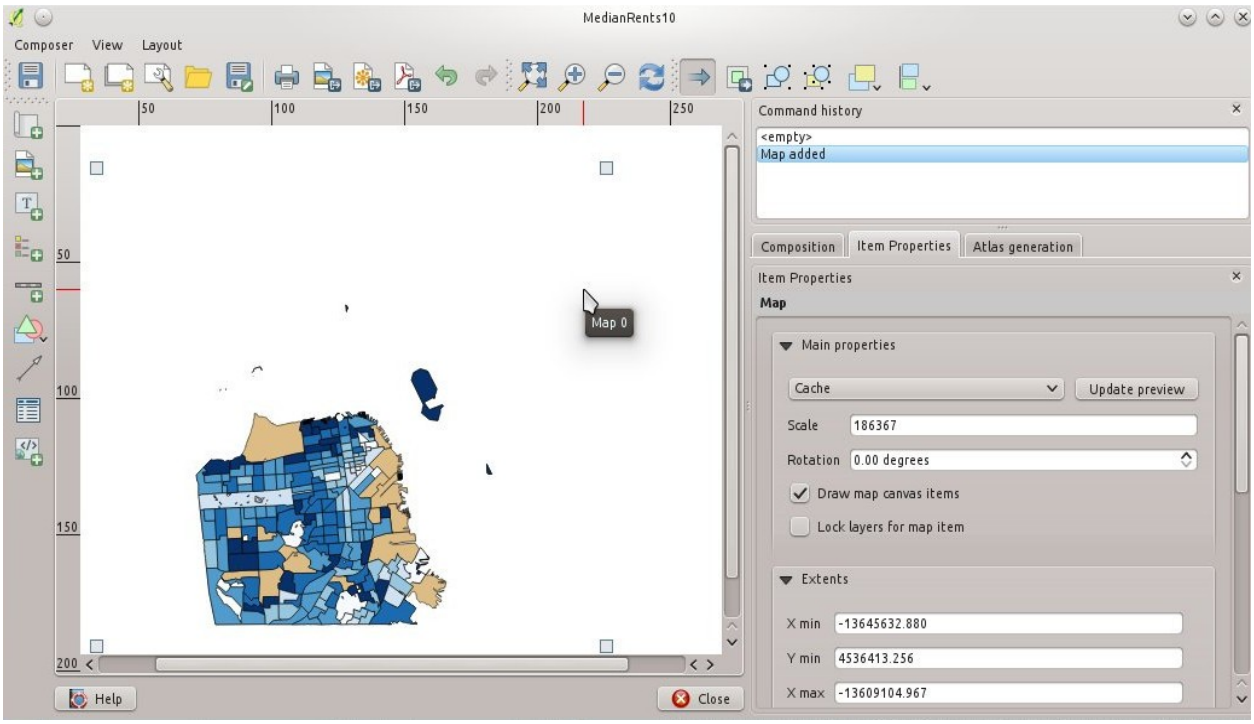
NOTE: in the first part of this Step 3 I am using an earlier color-scheme for my map. Further on, I will show how to update the Print Composer to reflect changed colors.



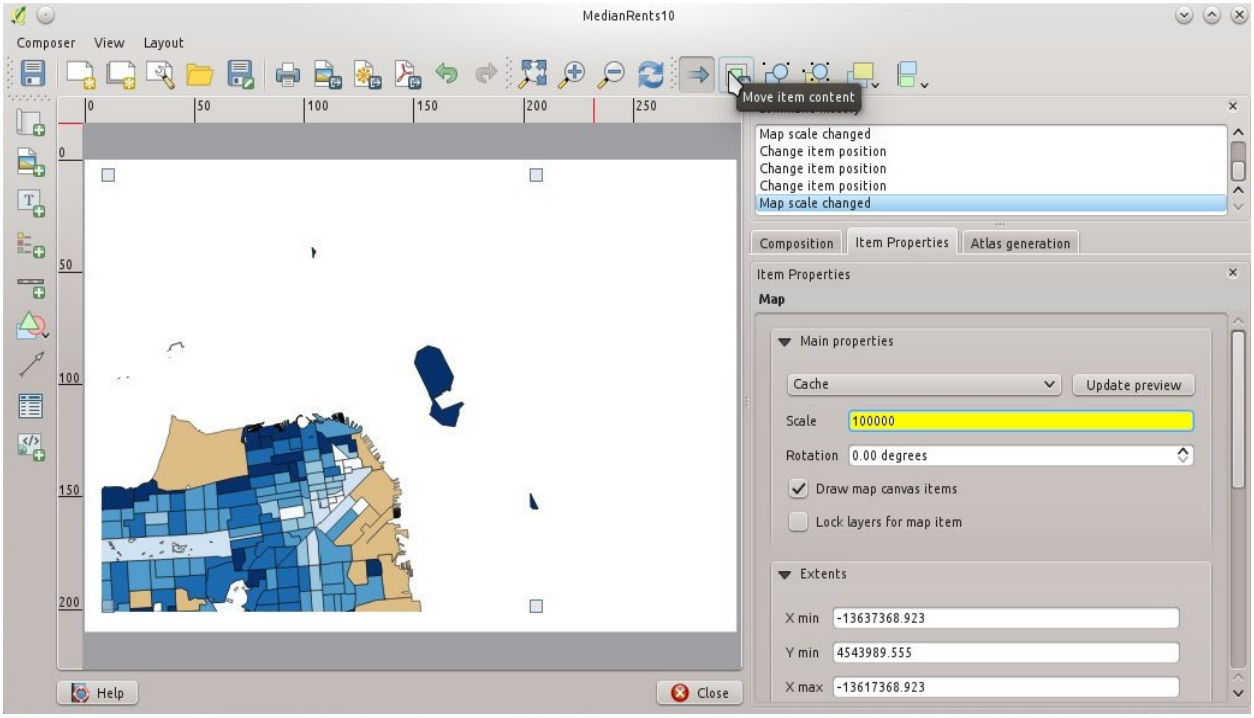
The Print composer opens as a blank at first, by default. Click **Add new map** and then click-and-drag a new window across the blank canvas.



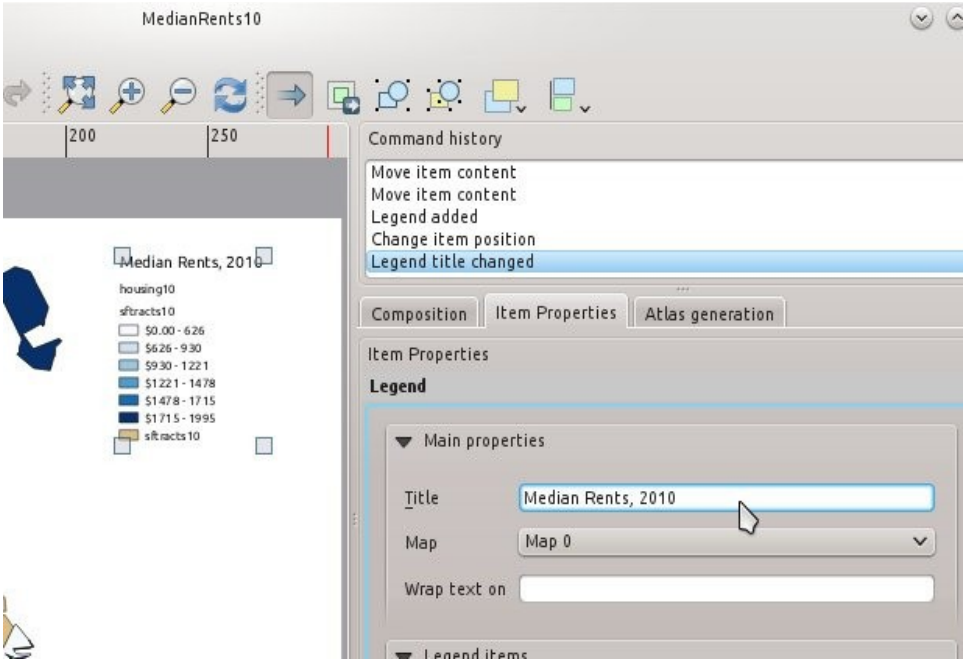
This will add a new map, which closely corresponds to what you have visible in your data window at the moment that you invoke the **Print composer**.



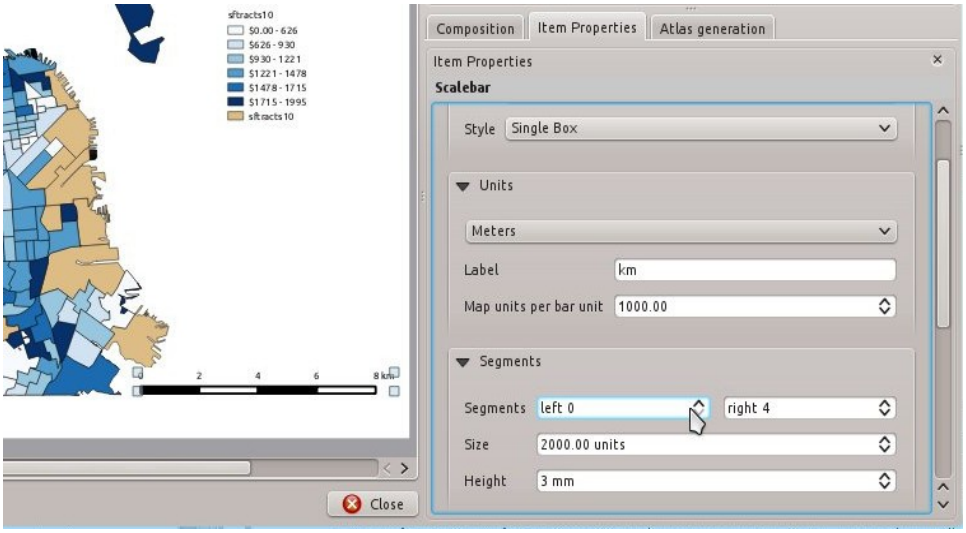
I rescaled the map in the Main properties pane to 1:100,000 (see below, highlighted in yellow)



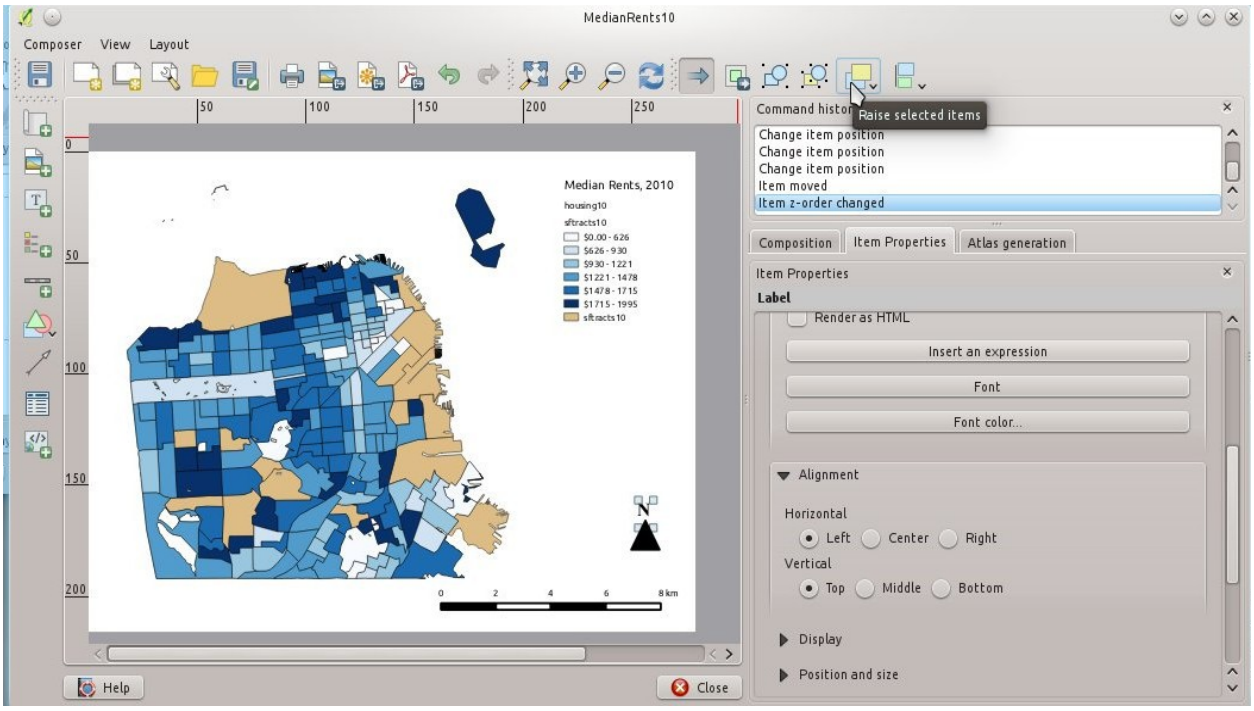
Then I selected the **Move item content** tool to adjust the position of the map-graphic.



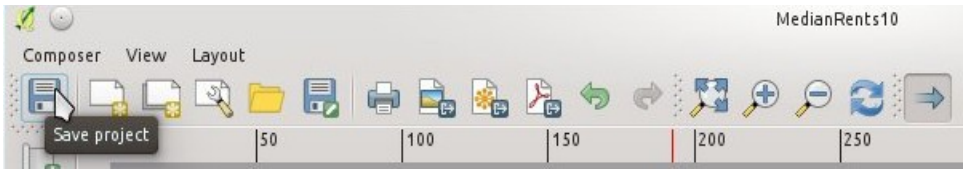
Next, I added a Legend and renamed the title (the **Legend** button is on the left).



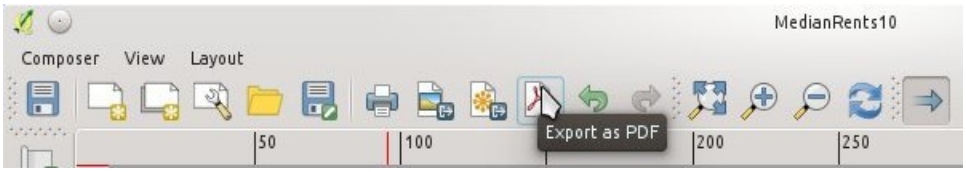
Then I added a scale bar. In my system this is all set up as metric by default. You may want to re-set your default units to feet and miles, but I will not cover that process in this tutorial. In the **Item properties** tab on the right, I am able to control the design (single box), the units (meters), the label (km), the number of Map units (m) per scale bar unit (1000), the number of Segments on the left (0) and right (4) of the scale bar. This makes the scale bar more legible.



Then I added a north arrow, and then an “N” above the arrow. Unfortunately the white backspace of the “N” overlapped the top of the arrow, so I used the **Raise selected items** to raise the arrow symbol on top of the “N”.



Once built, I saved this print-composition.

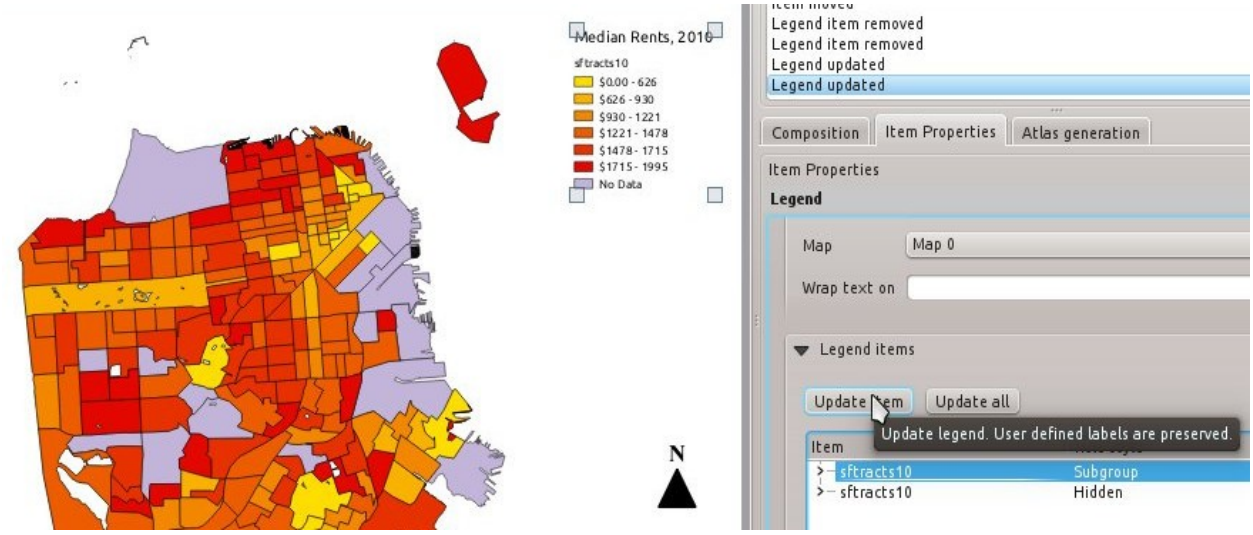


Finally, I exported the print-composition as a PDF. On my system it complains a bit about file-compatibility, but it worked. If you prefer you could export as an image file.

Revising your Print-Composition:

In the process of building this tutorial I changed my mind about the color-scheme I wanted to use for my chropleth map. The initial white-to-blue scheme was confusing, because tracts with the lowest median rents were colored white, whereas tracts with NULL data did not appear; leaving the white background visible. Thus I changed to the yellow-red color gradient for legibility.

Now I need to update my Print Composer. When I open it again, the main map has automatically updated to the new colors, but the Legend has not.



When I select the Legend in the main layout window, I can change its settings in the **Item properties** tab in the right-hand pane. First, I removed the display of the Housing10 layer, because it has no direct graphic representation (and that data is in fact displayed through the sftracts10 layer as joined attribute data). Second, I changed the display-name of the lavender under-layer to “No data”. Third, I clicked “Update item” so that the Print composer would re-read the display-settings from the working file and re-set the colors to the yellow-red sequence. That is how I produced the graphic at the beginning of this document.