Socio-Economic Aspects of Urban Development

Unit 6: Analyzing Geo-Spatial Data in QGIS

Prof. Melanie Krause, Ph.D.

Winter Semester 26/27

Overview of Unit 6

- How can we use geo-spatial data in urban analysis?
  - What is geo-spatial data?
  - Two types of geo-spatial data
    - ★ Vector data: Points, lines, polygons
    - ★ Raster data
  - Working with geo-spatial data in QGIS
  - Simple Operations in QGIS: Exercises
    - ★ Producing maps of city districts: Example of Leipzig
    - ★ Calculating mean building height, mean nighttime lights and the number of art galleries for each district of Leipzig

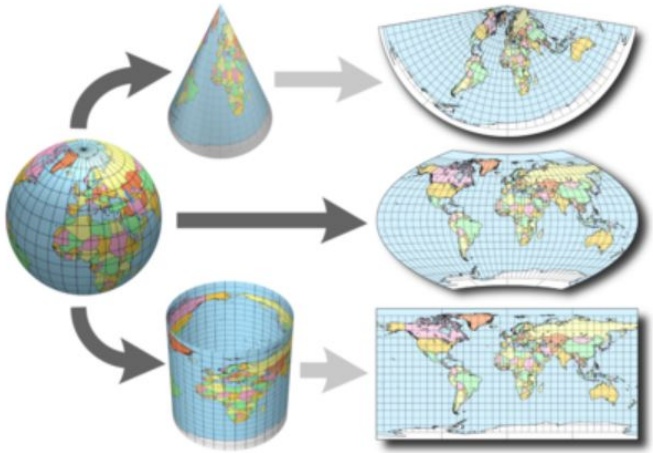
What is Geo-Spatial Data?

- **Geo-spatial data**, also called GIS data (Geographical Information System), are observations that are attached with **geographical identifiers**.
- Because of its place-based identifiers, for example latitude and longitude of a point, geo-spatial data can be visualized on a map.
- GIS data allows wide-ranging data analysis, measurements and geographical calculations
  - Visualize a map of the geographical distribution of some variables
  - Analyze the spatial distribution of a variable: Which district of the city has the highest building height?
  - Compute distances from neighborhoods to points of interest nearby: How far is it from your house to the nearest restaurant?
  - Construct whole infrastructure networks

The Coordinate Reference System

- The geographical identifiers are referenced into a **coordinate reference system (CRS)**.
- **Geographical CRS**: Spherical coordinates, decimal degrees:
  - Example: Leipzig is at
    51° 20' 24.7164" N (degrees, minutes, seconds Northern latitude)
    12° 21' 36.3708" E (degrees, minutes, seconds Eastern longitude)
  - Problem: Degrees do not reflect the same distance around the globe. They are much closer to each other at the poles.
  - For calculating areas and distances, we need a **Projected CRS**.
  - A projected CRS represents the earth's sphere (3D) in a map (2D). There are different projected CRSs depending on how you do the projection. The choice of the CRS can influence the distance and area you measure. But none is perfect: All projected CRS are distorted at same way or on some place on earth.

The Coordinate Reference System



- The Web Mercator projection is well-known. On a map, it inflates areas with distance from the equator so that the polar regions are grossly exaggerated.
- The Universal Transverse Mercator (UTM) projection is useful when calculating distances and areas on small scales. It tries to minimize distortions by dividing the earth into 60 zones, projecting each individually. Leipzig is in Zone UTM 33 North.
- An equal area projection (such as Mollweide) is useful for calculating large areas or distances.
- *When working on small scales, such as within cities, the choice of the CRS is not that important. Just be aware that you have to choose one and be consistent!*

Vector and Raster Data

Geo-spatial data comes in two main types

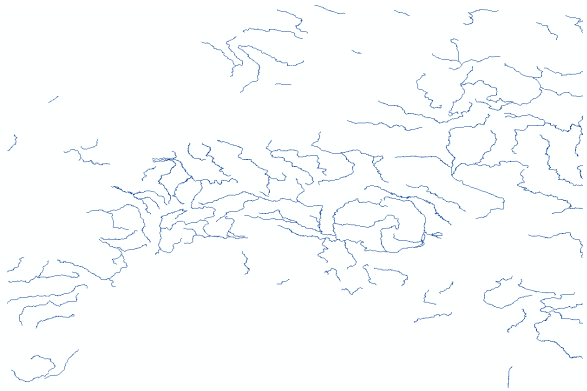
- **Vector Data** is often packed in shapefiles (*.shp)
  - Represent points, lines or polygons
  - Useful to store discrete features like roads and city blocks
  - Also important for drawing boundaries between neighborhoods and cities
  - If we want to compute the number of cafés within a district, we need to have the district as a polygon because of the boundaries
- **Raster Data** is often packed in GeoTIFF (*.tif)
  - Represent gridcells and their interiors
  - Useful to store elevation, temperature, nighttime luminosity
  - Often derived from satellite imagery
- As a file format, GeoPackages (*.gpkg) are becoming more popular. They are file containers like (*.zip) and can contain vector or raster data.

Vector and Raster Data

Vector data: Countries of Europe as polygons



Vector data: Rivers of Europe as lines

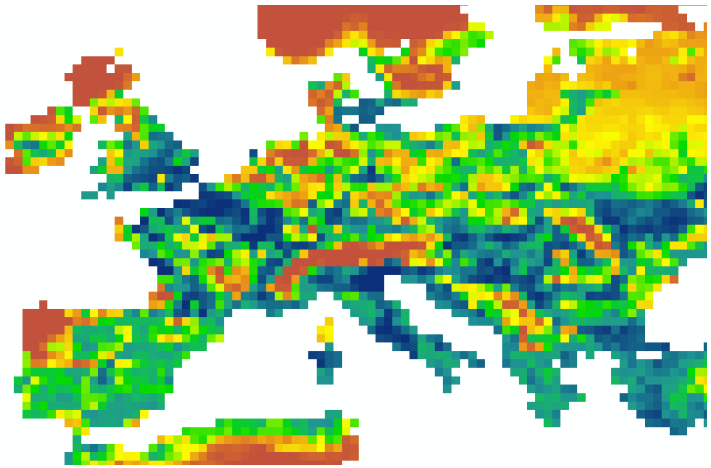


Vector and Raster Data

Vector data: Cities of Europe as points



Agricultural suitability of Europe as raster data: Each cell in the raster has a value

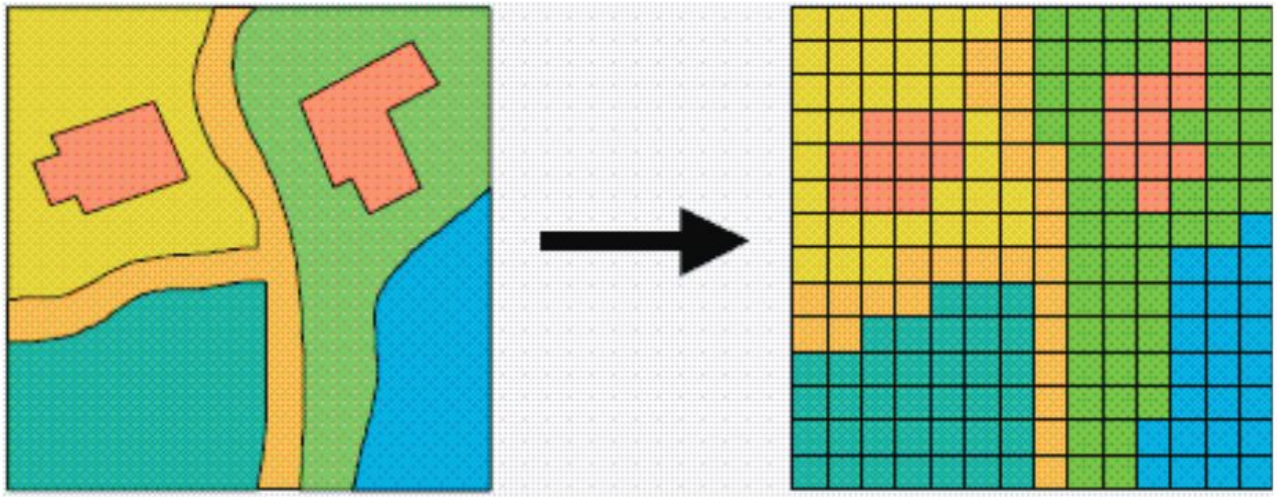


1	1	0	0
	1	2	2
4	0	0	2
4	0	1	1

Figure: Agricultural suitability of Europe in a raster

Vector and Raster Data

Comparison of vector and raster data



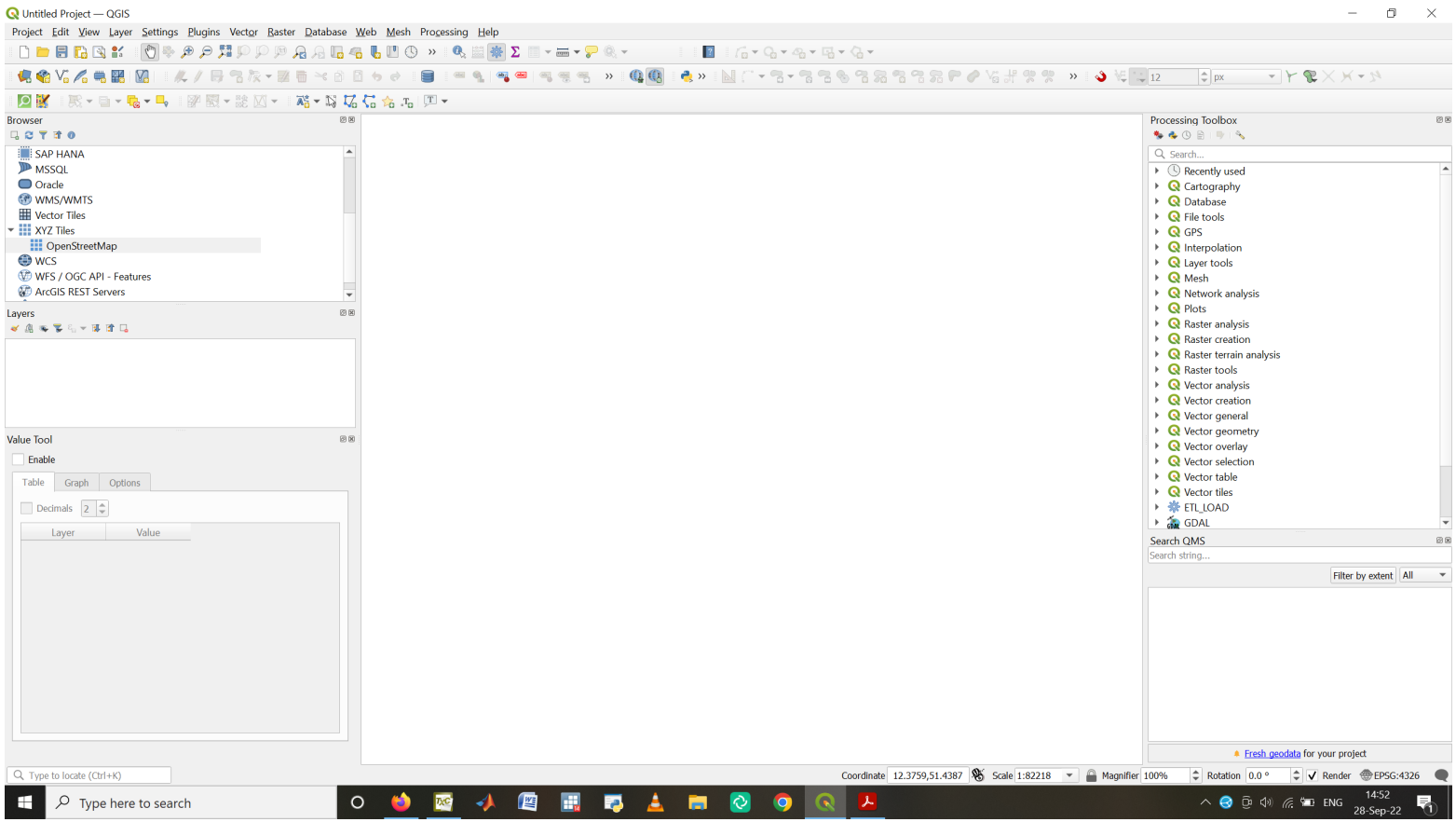
"Raster is faster but vector is corrector."

What is QGIS?

- **QGIS** is a free, widely-used, specialized software for geo-spatial analysis
- Easy to learn with extensive documentation and tutorials
  - Download it from https://www.qgis.org/download
  - Tutorials: https://docs.qgis.org/3.40/pdf/en/, http://www.qgistutorials.com/en/
- It has a neat graphical user interface (GUI) so that you see what you are doing - very helpful for geo-spatial analysis, in particular for beginners.
- It has a vast community continuously updating and extending the software.
- It has many useful packages and plug-ins that are already installed, while others can be added as needed.
- Whether creating a simple map or doing more advanced operations, everyone can get something out of it rather quickly.

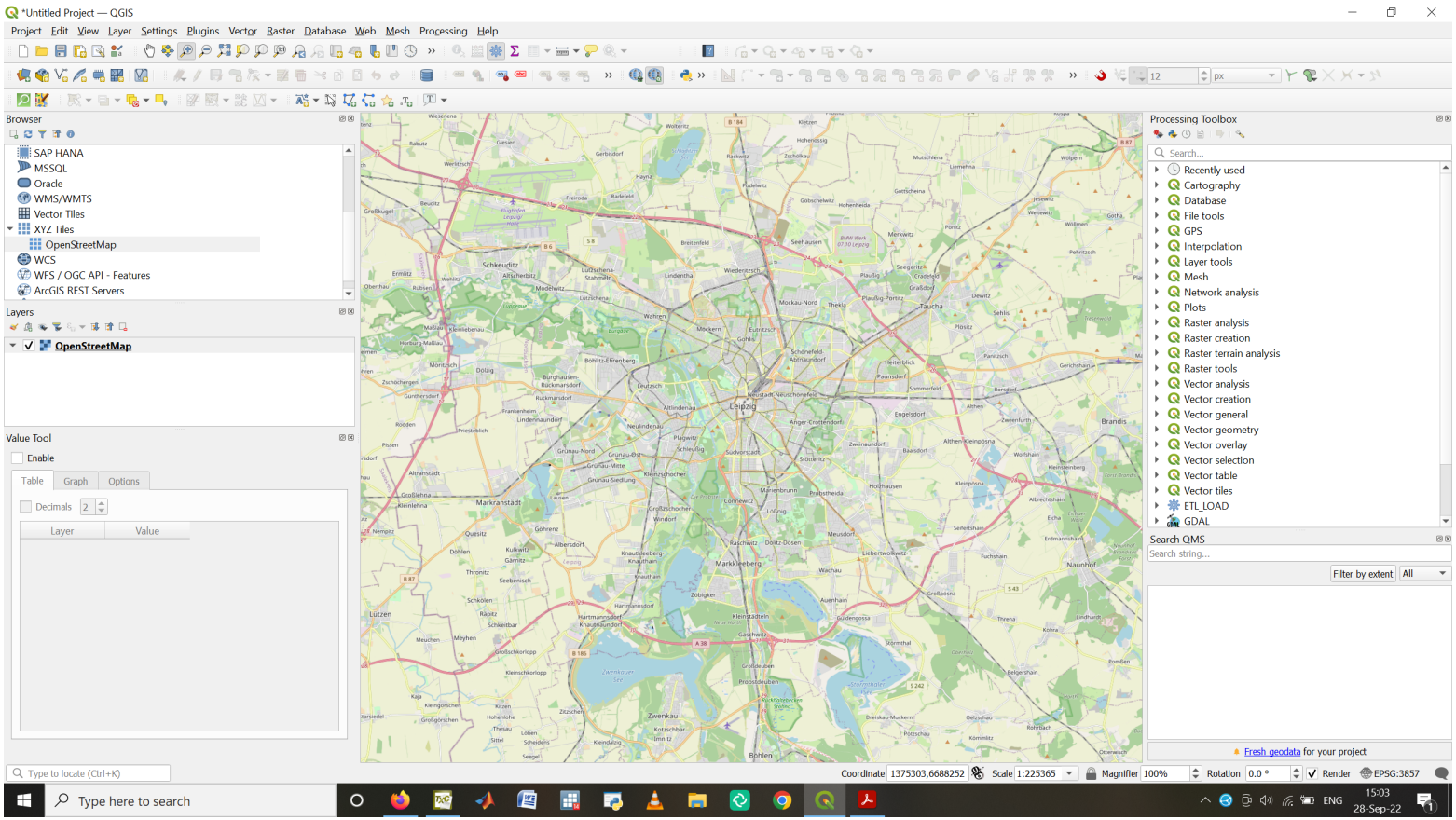
QGIS Application: Screen Layout

- Get started with QGIS! Start a new project, inspect the screen layout and the drop-down menus.
  - You can customize your screen by adding panels: Go to the drop-down menu *View -> Panels* and tick e.g. the *Processing Toolbox*.



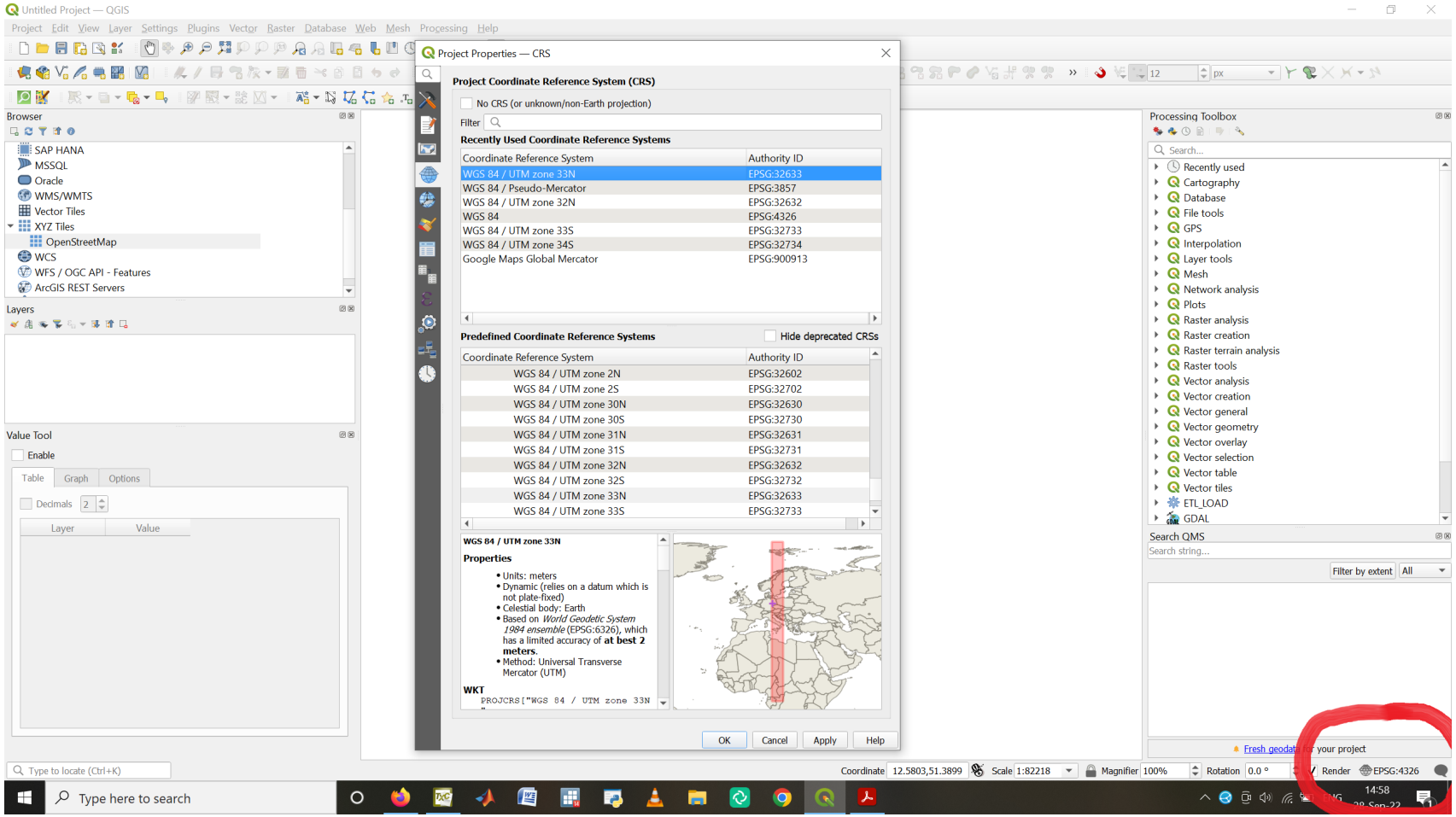
QGIS Application: Background Map

- QGIS comes with a map from Open Street Map. You find it under *XYZ Tiles* of your *Browser* panel (*View -> Panels -> Browser*). Zoom (the glasses with the + sign) to your place of interest, here Leipzig.



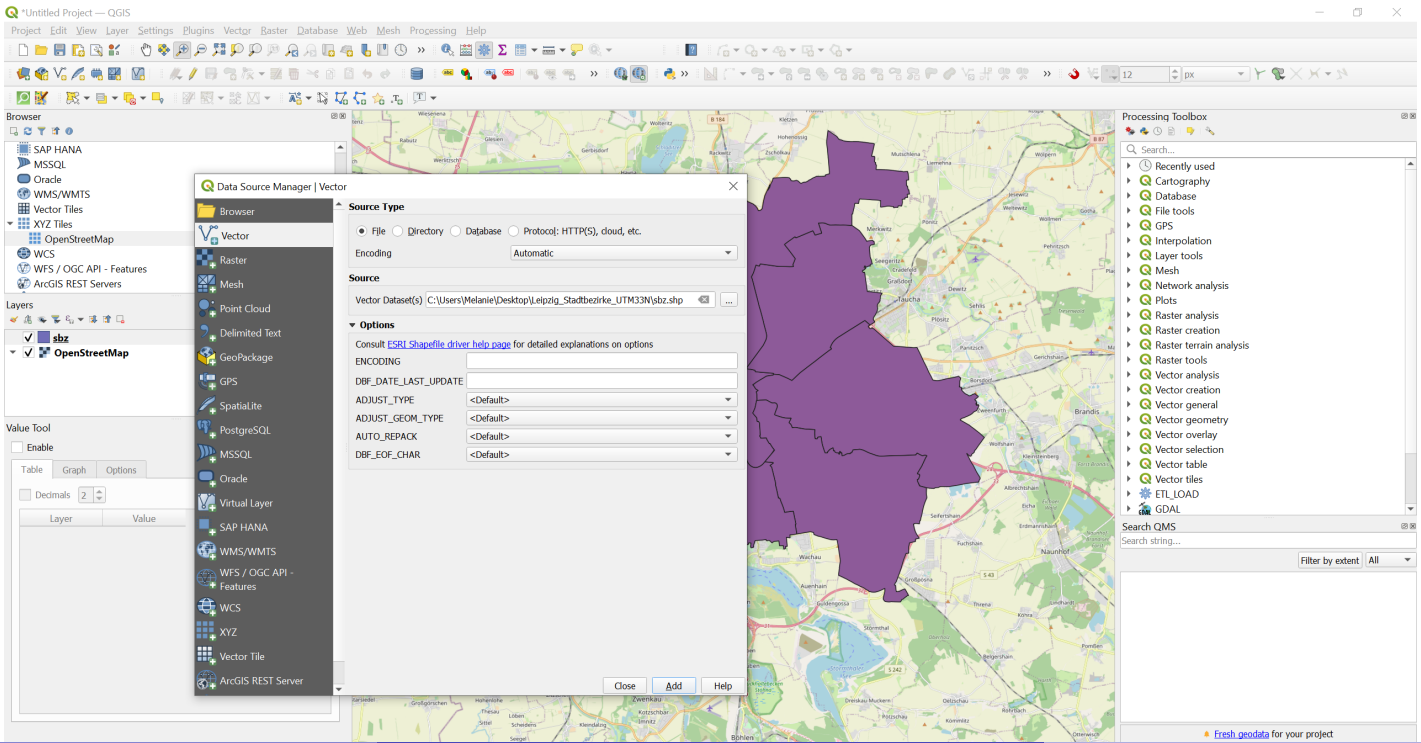
QGIS Application: Coordinate Reference System

- You can adjust the coordinate reference system at the bottom right of the screen, choosing WGS 84 / UTM Zone 33 N (EPSG 32633) (for our section of the Northern hemisphere). Click 'Apply'.



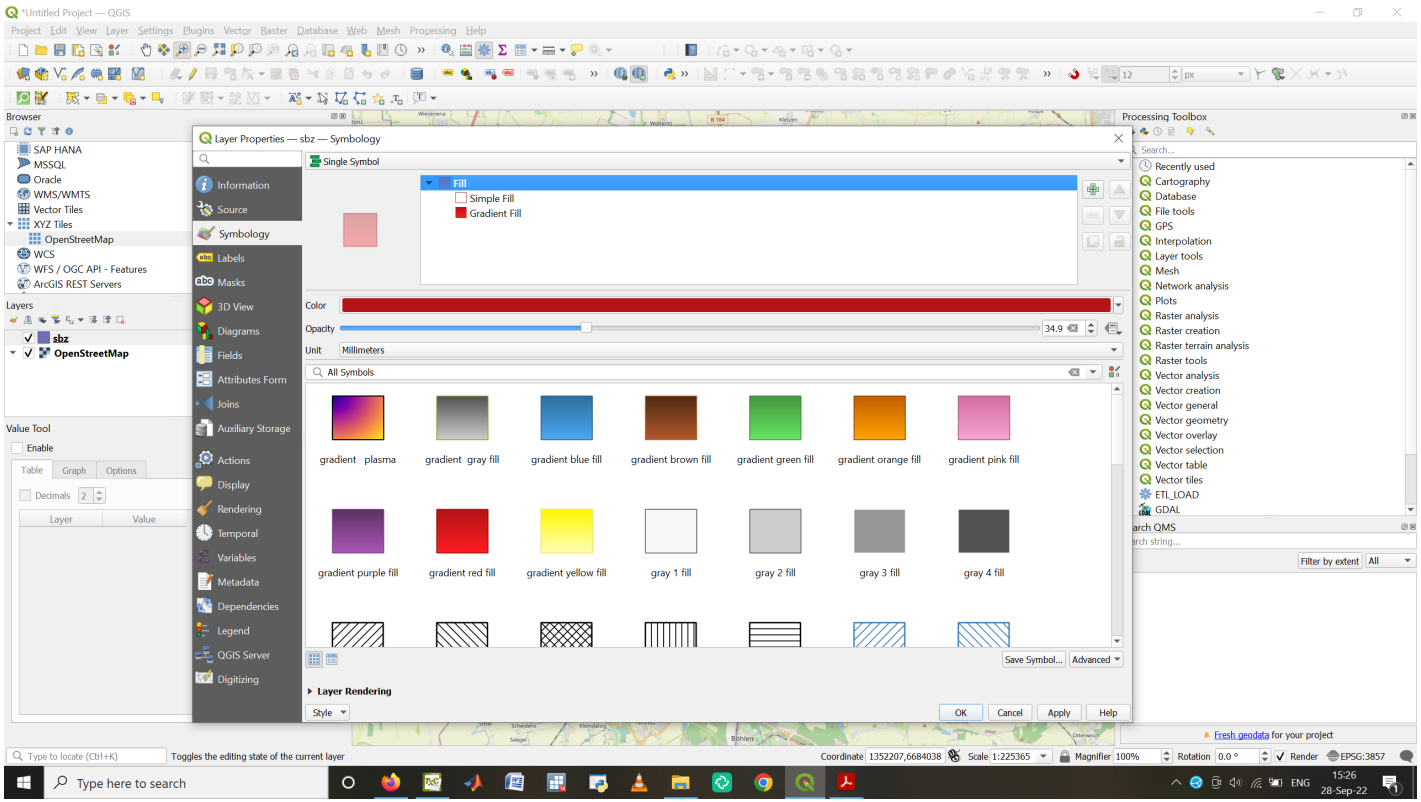
QGIS Application: Add a Vector Layer (City Districts)

- To analyze data district by district, we need the boundaries of the individual city districts as polygons. Download the *sbz* files from the QGIS folder on Moodle to your computer (original source: City of Leipzig, https://www.leipzig.de/buergerservice-und-verwaltung/unsere-stadt/gebietsgliederung-und-strassennamen/kommunale-gebietsgliederung).
- Add this new layer: *Layer -> Add Layer -> Add Vector Layer* and choose the location of the *sbz.shp* shapefile on your computer.



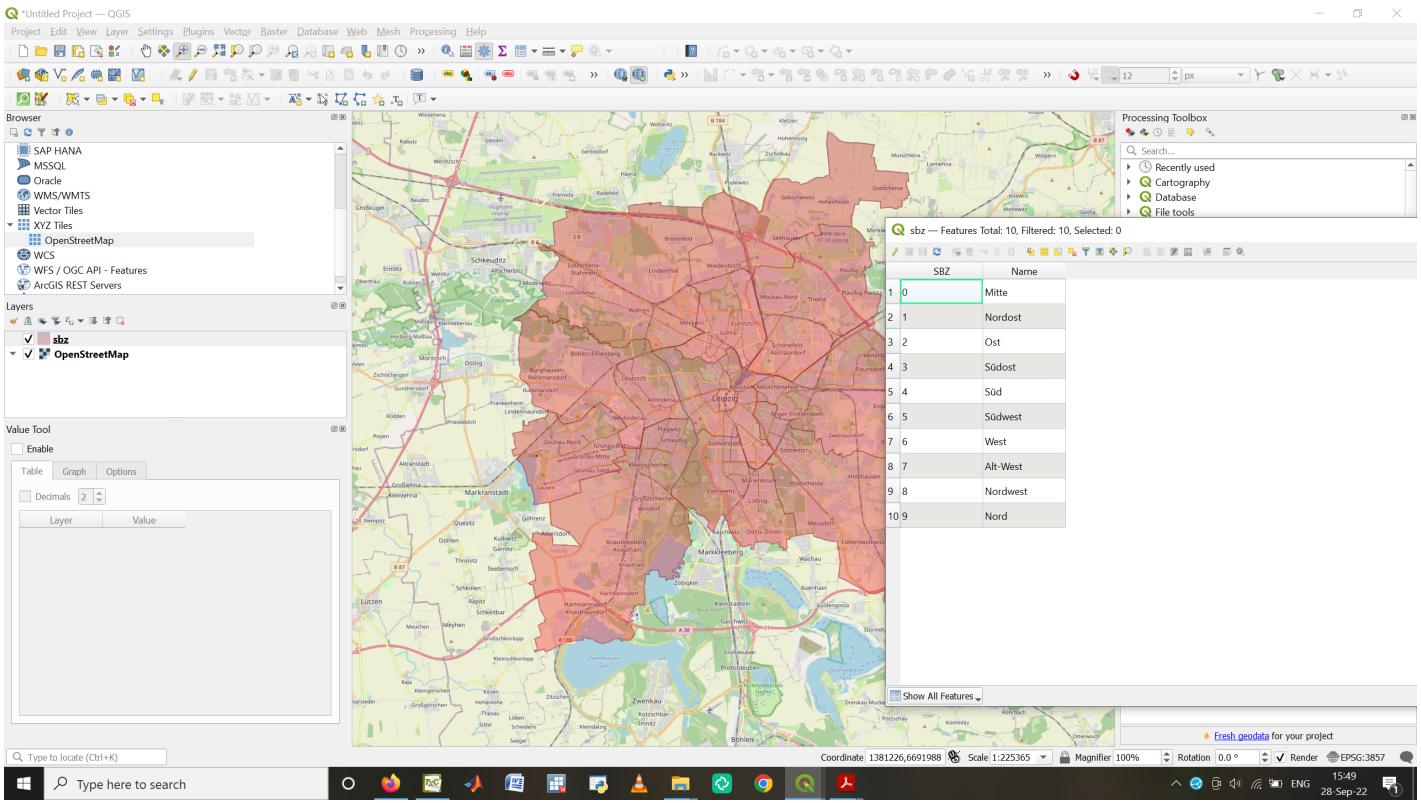
QGIS Application: Modifying a Vector Layer

- New layers appear on top of existing layers and may hide the ones beneath.
- Let us change the color and transparency of the polygons by right-clicking on the layer and selecting *Properties -> Symbology*. Adjust the color and the opacity (how transparent the layer is). Click *Apply*.



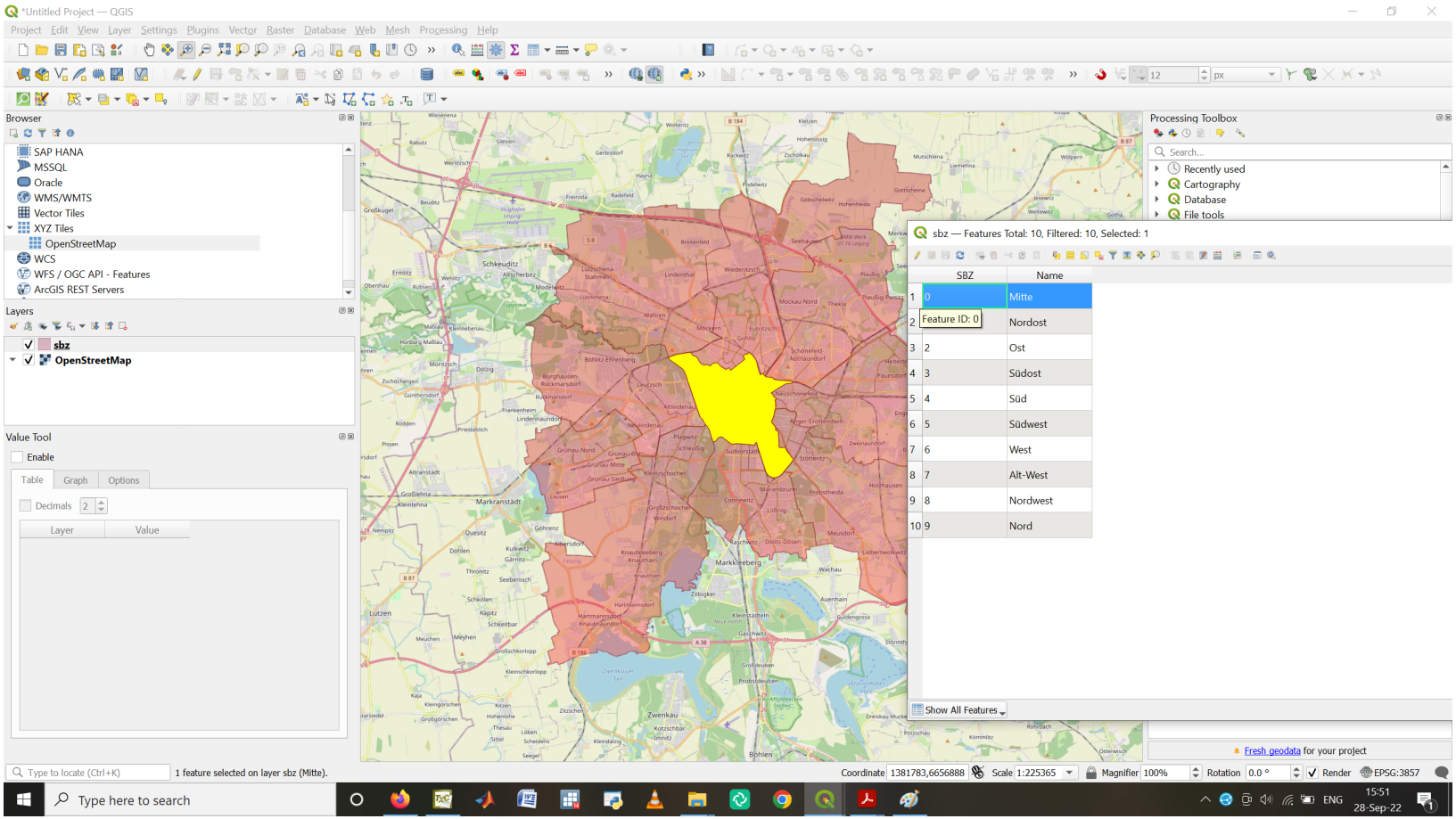
QGIS Application: The Attribute Table

- Check out the attribute table: Each vector layer has an attribute table that contains information on all the elements (called features) of the layer, here the city districts and their names. We can later add more information and export the attribute table as CSV file.
- You get to the attribute table by right-clicking on the vector layer and selecting *Open Attribute Table*.



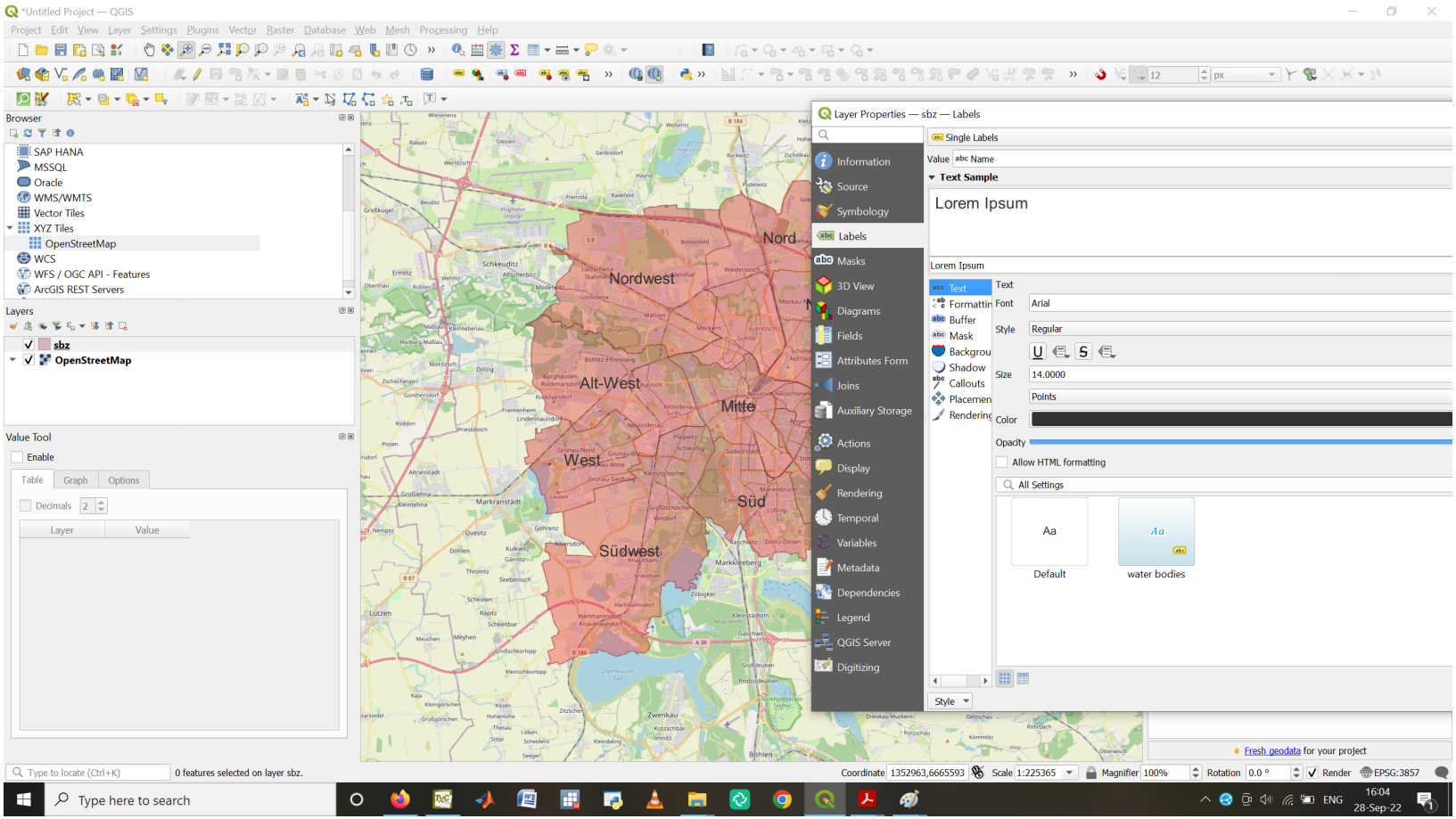
QGIS Application: The Attribute Table

- Click on one feature (city district) in the attribute table and see how the corresponding polygon gets highlighted in yellow in the map. With the *Deselect* icon on top of the attribute table, the selection of the city district vanishes again.



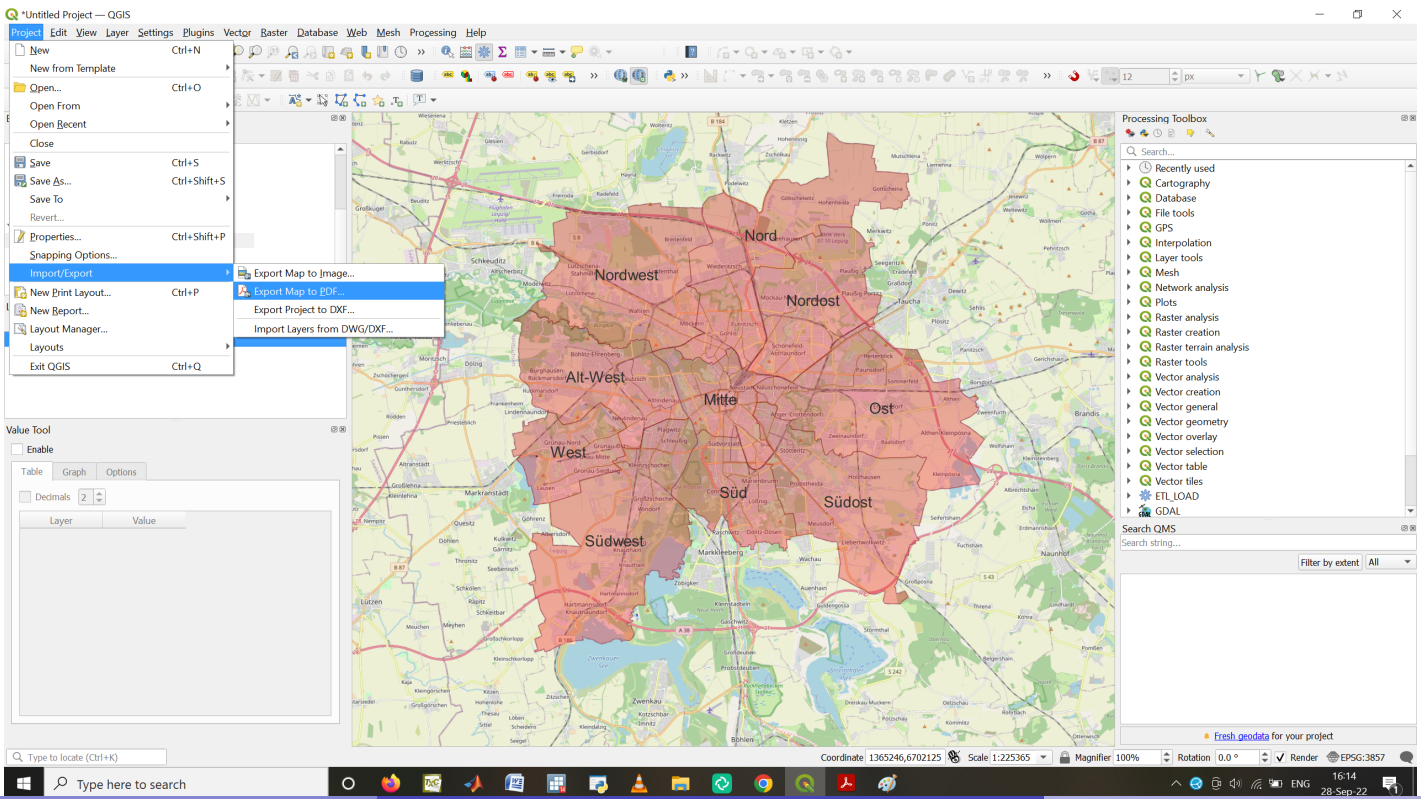
QGIS Application: Adding Labels

- How can we add the names of the district to the polygons on the map? Right-click on the layer and select *Properties -> Labels -> Single Labels*. Make sure that the *Name* column is chosen as *Value* and click *Apply*. You may adjust the text formatting of the labels under *Text*.



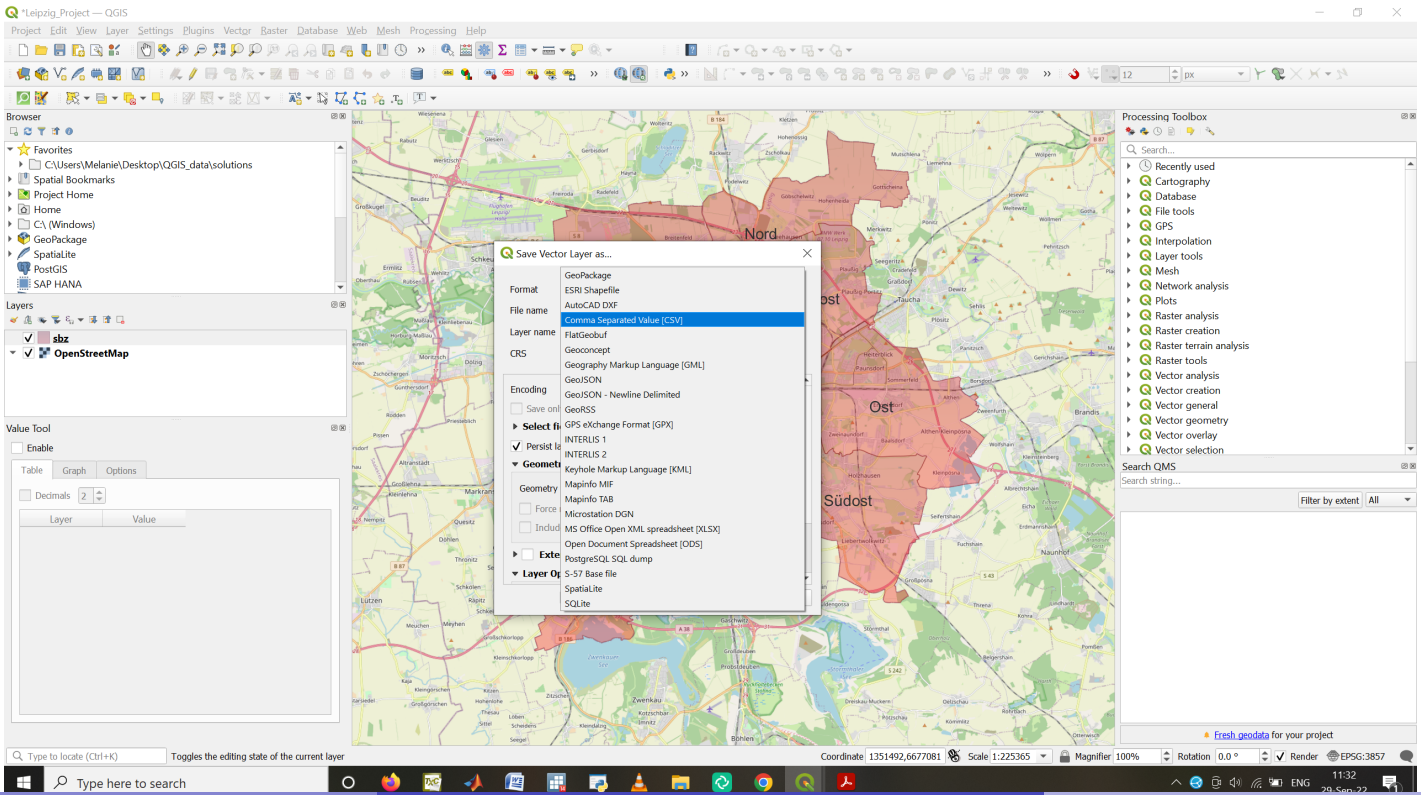
QGIS Application: Export Your Map

- If your goal was to create a map and you are happy with it, you can export it as an image of a PDF via *Project -> Import/Export -> Export Map to PDF*.
- There is A LOT more you can do in QGIS. But remember to save your intermediate project when you close QGIS so that you can return to it later *Project -> Save As*. Note: You have to keep the layers in your original location on your computer, because QGIS will refer to these locations when you open the project again.



QGIS Application: Export Your Attribute Table

- The attribute table does not only contain the initial object information, but also the results of calculations that we are going to do now (which will add further columns for each city district).
- You can export your attribute table as a CSV file for further use outside of QGIS (statistical analysis): Right-click on the *sbz* layer: *Export -> Save Features as* and choose *Comma Separated Values (CSV)* from the drop-down menu.

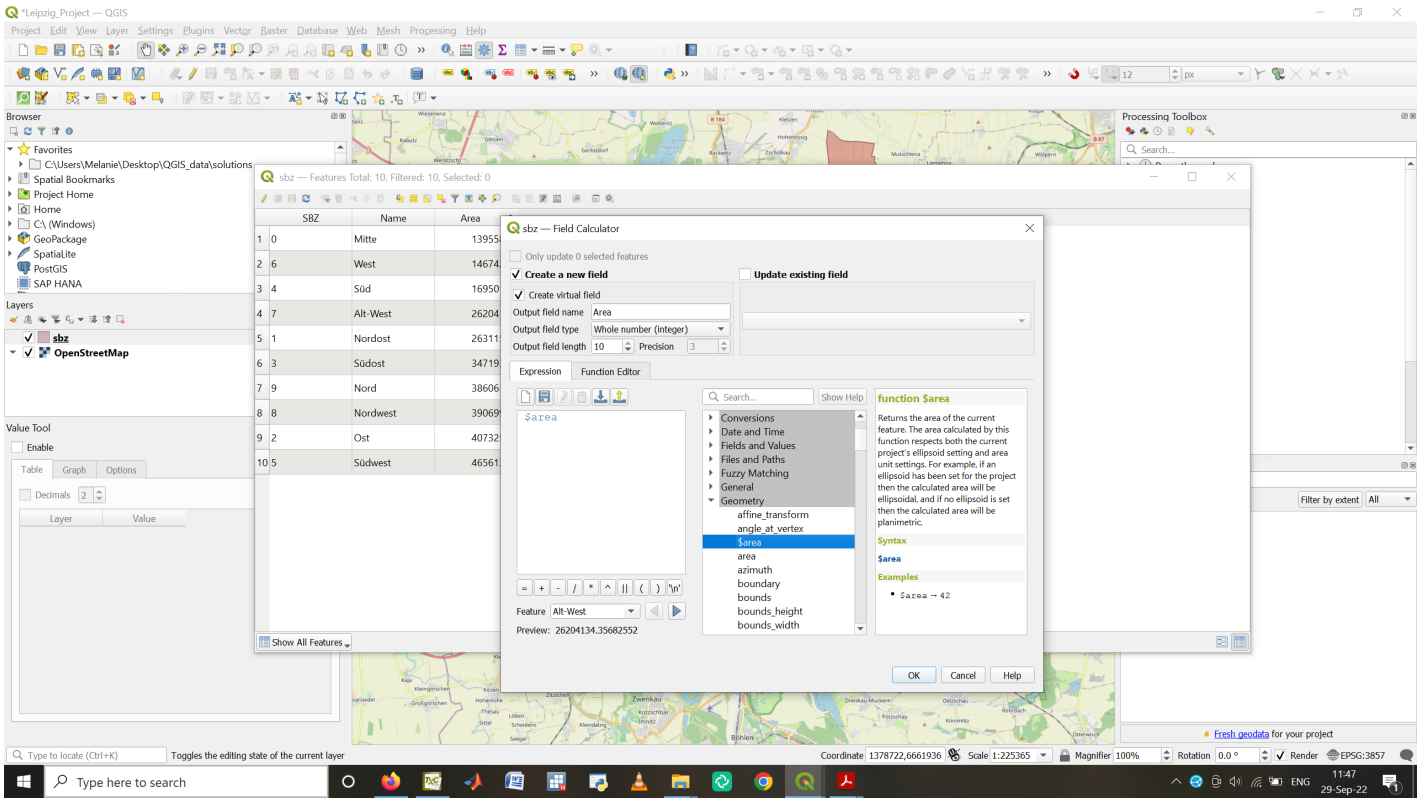


So what else to do in QGIS?

- Apart from a nice graphical representation of data in form of a **map**, QGIS allows us to **measure things** by using their geographical properties
  - The attributes of the object are stored in the attribute table. The **Field Calculator** allows you to add further attributes as columns.
    - ★ Common calculations are the length of a line (e.g. length of a river) or area of a polygon (e.g. area of a building)
  - The **Geoprocessing Tools** contain tools to manipulate vector data, creating a new vector layer that fulfills certain conditions
    - ★ Common vector data manipulations are calculating the intersection of polygons (e.g. the part of city district that was built after year x) or a buffer around a object (e.g. the number of offices within a mile around a coffee shop).
  - The **Zonal Statistics** plugin allows you to extract various attributes of the raster data within polygons, which are then added to the attribute table of the polygon
    - ★ Common calculations are the mean of pixel values (e.g. average nighttime light intensity)

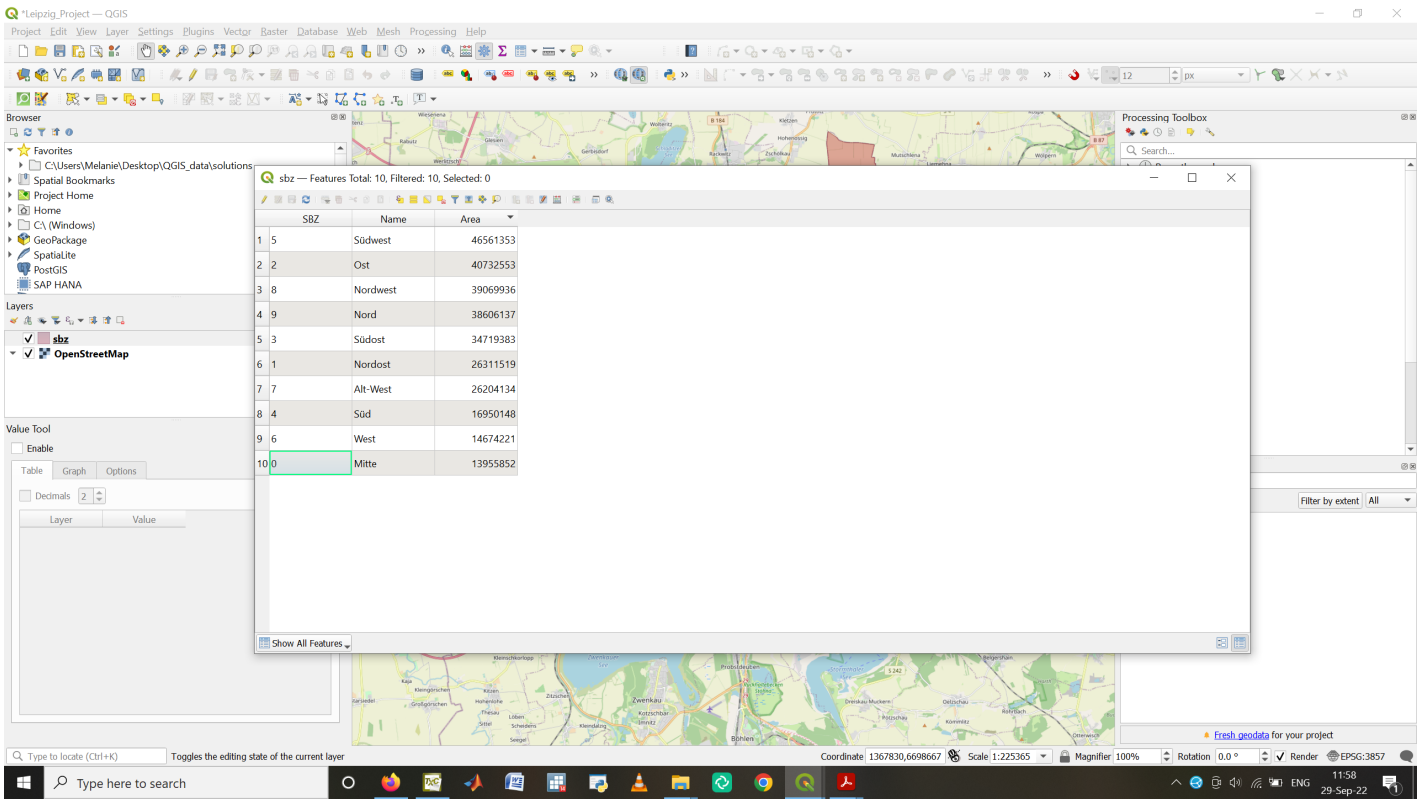
QGIS Application: Computing the Area of a Polygon

- Which district of Leipzig is the largest? Calculate the area!
- Right-click on the layer to open the attribute table. Click on the little abacus icon (counting frame) that reads **Open Field Calculator**. Create a new virtual field with the title 'Area'. For the computation, select *Geometry -> $area*, click on it and press *OK*.



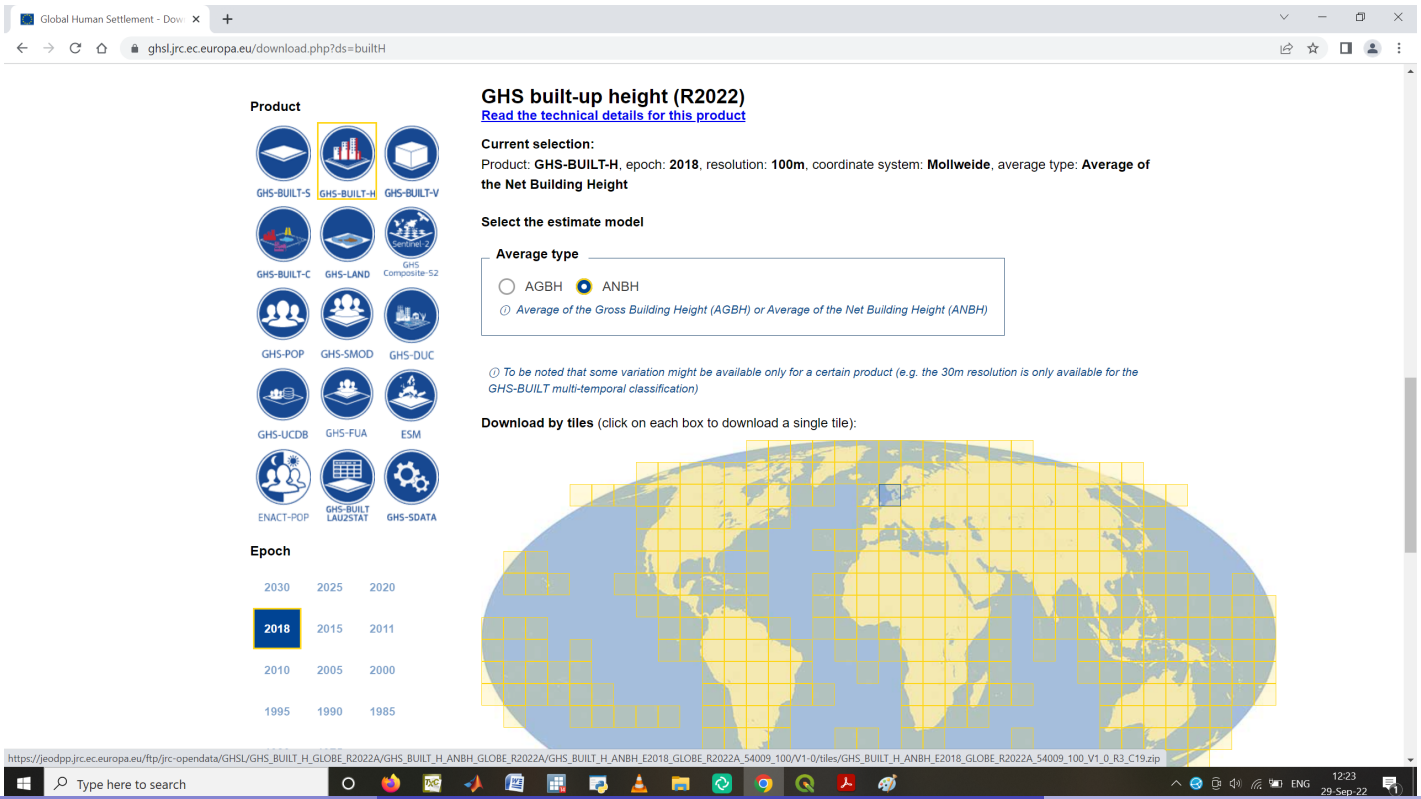
QGIS Application: Computing the Area of a Polygon

- A new column has appeared in the attribute table, the area measured in square meters. (How do you know the units? Check them under *Project -> Properties -> General settings - > Units for area measurement*)
- You can sort the area column to see that South West Leipzig of has the largest area (46.56 square kilometers), followed by East.



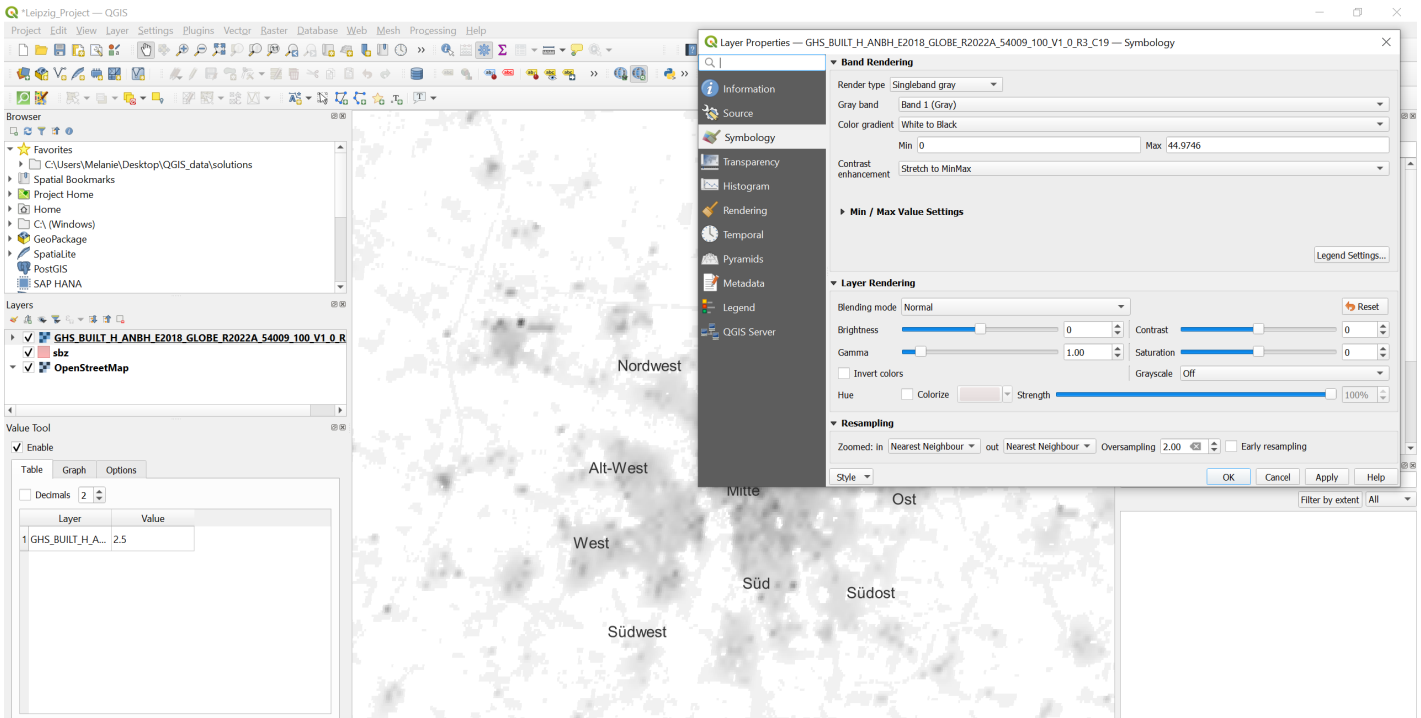
QGIS Application: Computing Average Building Height

- Which district has the highest average building height? Can we detect the pattern of decreasing building height predicted by the Alonso-Muth-Mills model?
- We will use data on building height of the Global Human Settlement Layers from the European Commission, year 2018, resolution 100m. It can be downloaded as a raster file: https://ghsl.jrc.ec.europa.eu/download.php?ds=builtH Select the tile of the earth you are interested in and check if you want net or gross building height



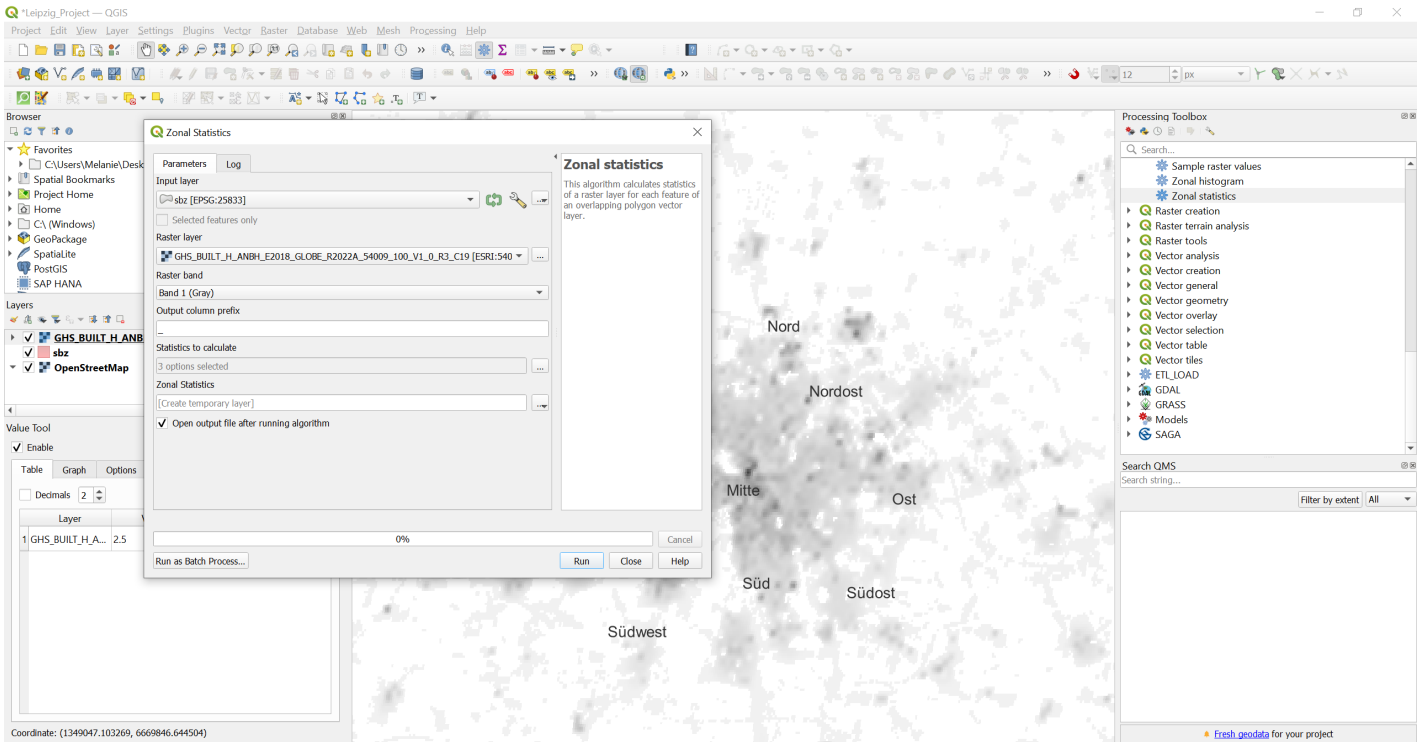
QGIS Application: Computing Average Building Height

- Here, we use net building height, which is building volume divided over the building footprint, hence we get an average height of all buildings.
- Gross building height is the building volume divided by total area, hence it will be lower if there are few buildings and a lot of free space.
- Add the layer: *Layer -> Add layer -> Add raster layer*. The layer may look very dark. But you may change the colors from 'black to white' to 'white to black' by right-clicking on the layer *Properties -> Symbology -> Color Gradient*



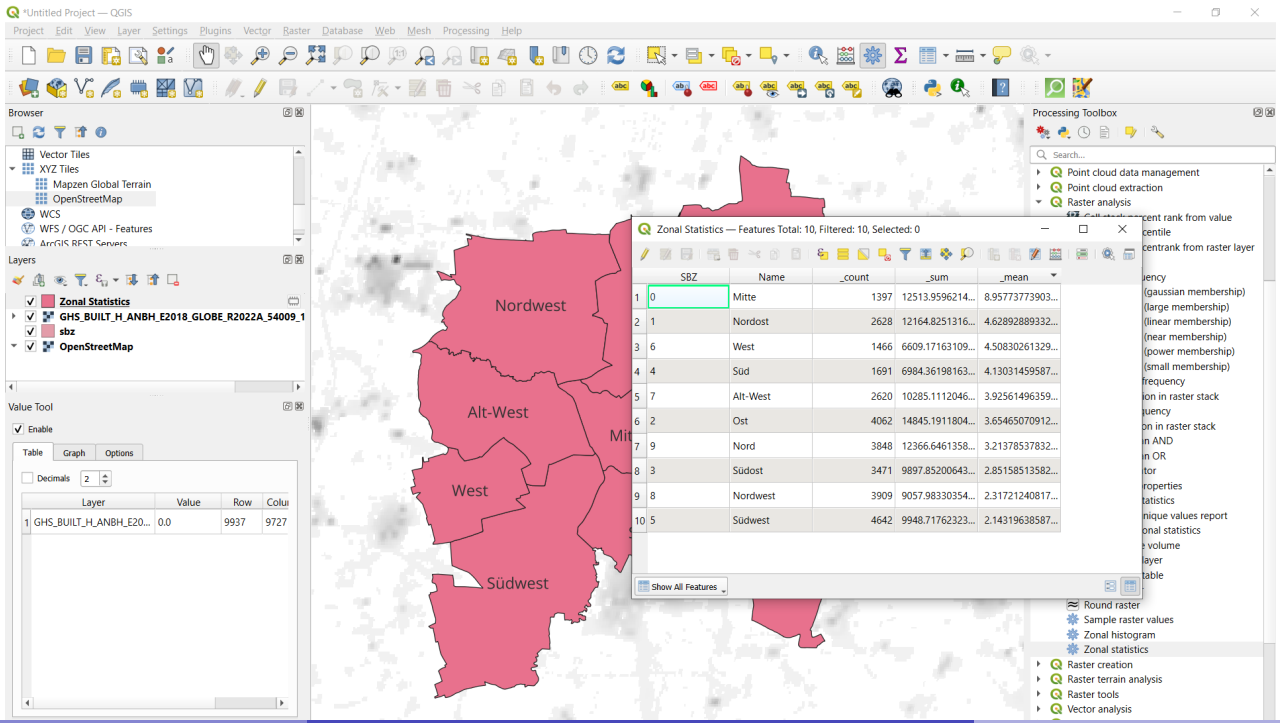
QGIS Application: Computing Average Building Height

- Inspect the 100m building height raster. With *Plugins -> Value Tool*, you can enable the value tool in the bottom left corner, so that you can draw your mouse across the raster and see the building height in that cell.
- Which district has the highest average building height? We have to calculate mean values of the cells that fall within a district polygon! Go to *Processing Toolbox* → *Raster Analysis* → *Zonal Statistics*. As the input vector layer, select *sbz*. As raster layer, choose the building height layer.



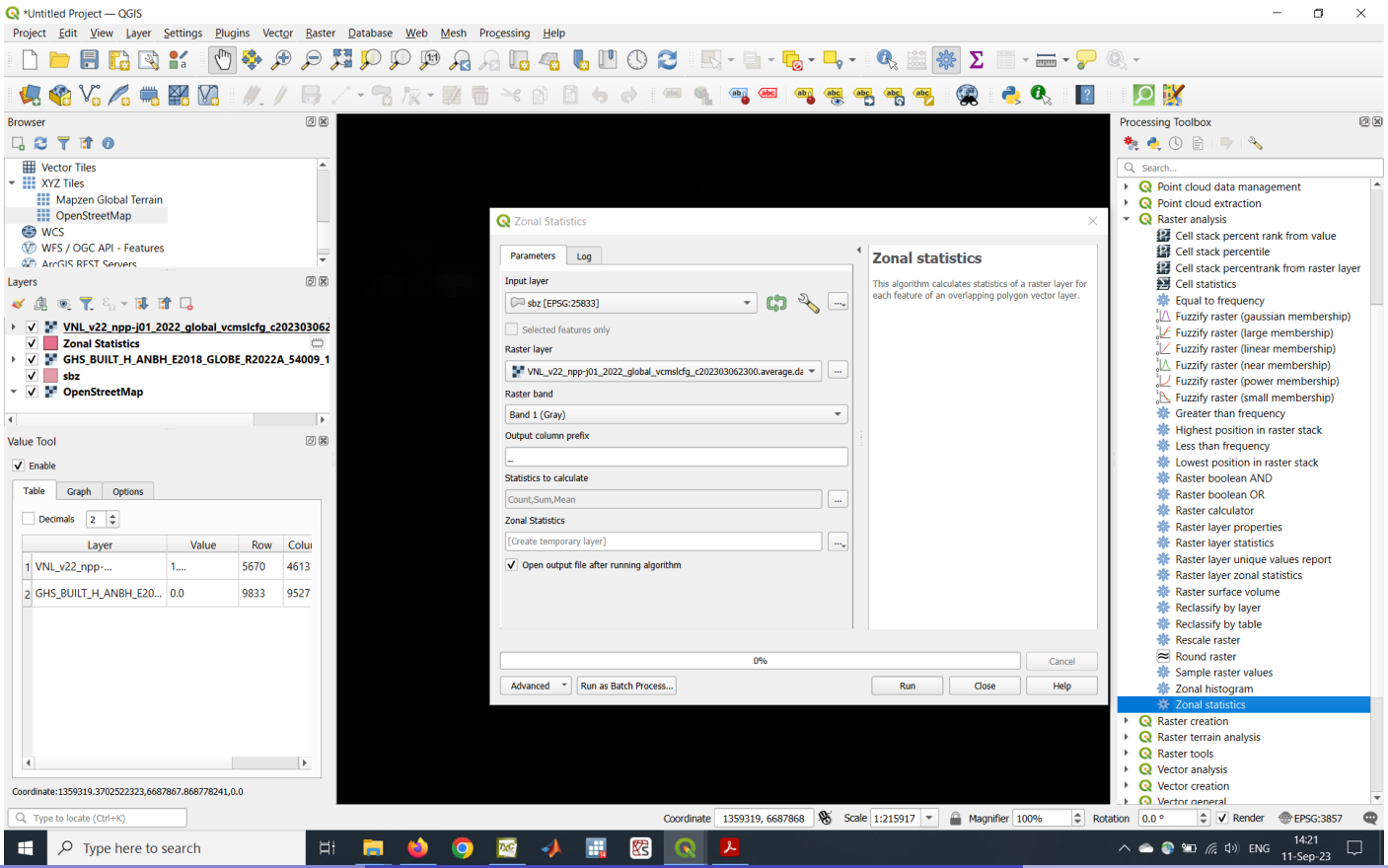
QGIS Application: Computing Average Building Height

- A new layer called *Zonal Statistics* has been created. Right-click on it to see the attribute table: It shows the number of cells (*_count*) per district, the total summed-up building height (*_sum*) per district as well as the mean building height (*_mean*) per district.
- The center of Leipzig (Mitte) has a mean building height of 8.96m, while it is only 2.14m in the South West.
- Note that you can rename a layer by right-clicking on it. Make sure to save newly created layers before closing QGIS!



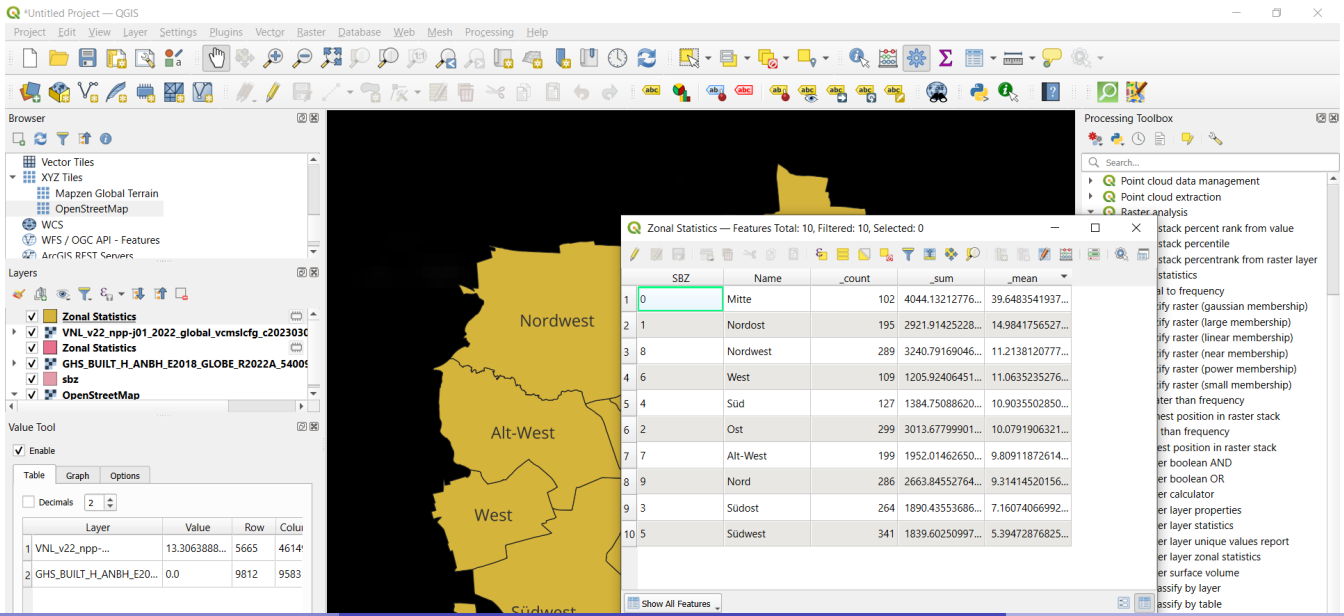
QGIS Application: Computing Average Nighttime Light

- Exercise: Will you obtain similar pattern for nighttime lights as for building height in the districts? Go to https://eogdata.mines.edu/products/vnl/ and download the *Annual VNL V2.2* from 2022 (raster file, but has to be un-zipped first). Add the layer and follow the same steps as before to calculate zonal statistics for average nighttime lights in the districts!



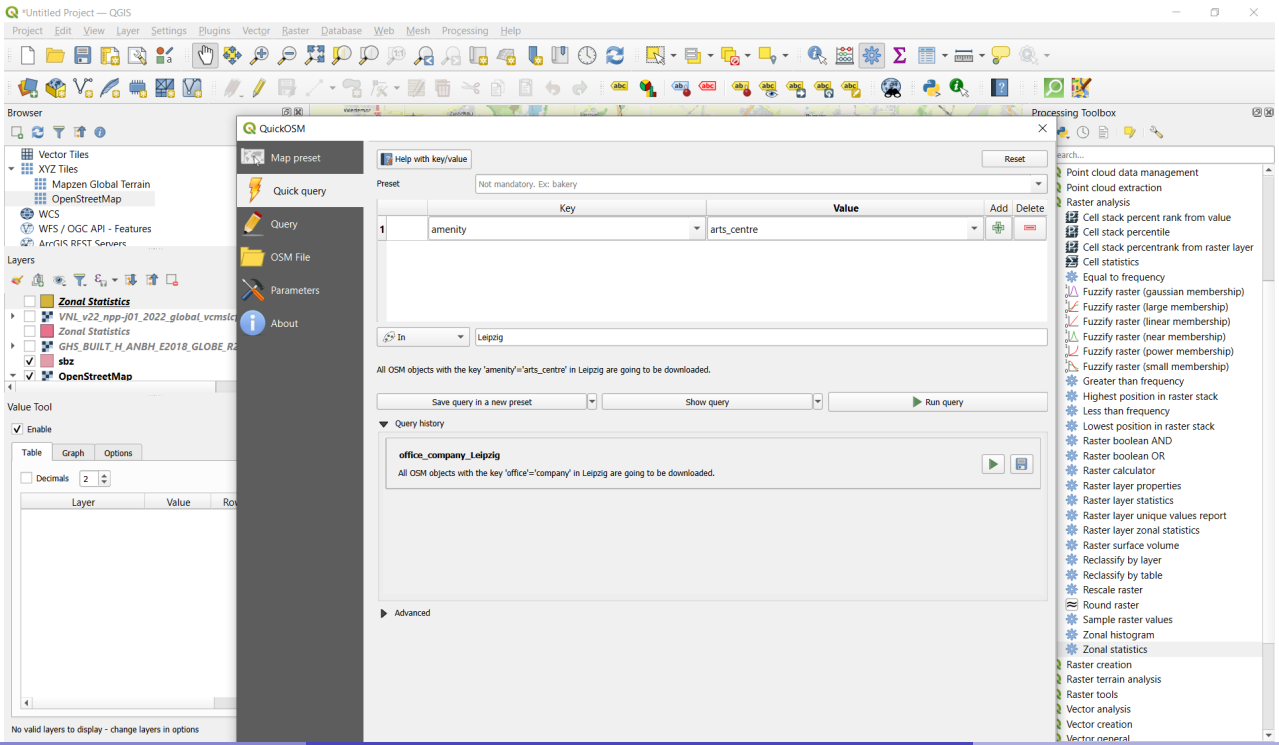
QGIS Application: Computing Average Nighttime Light

- The attribute table of the newly created *Zonal Statistics* layer shows: Mean lights are also highest in the city center (39.65), where they are nearly eight times as high as in the South West (5.39). The lights are measured in nano watt per steradian per square centimeter.
- Similar to building heights, nighttime lights show a pattern of decreasing intensity from the center to the outskirts.
- More detailed analysis - and more QGIS practise: Instead of the 10 city districts (*Stadtbezirke*), use the 95 sub-districts (*Ortsteile*) as units of analysis! Get the vector shapefile from https://www.leipzig.de/buergerservice-und-verwaltung/unsere-stadt/gebietsgliederung-und-strassennamen/kommunale-gebietsgliederung



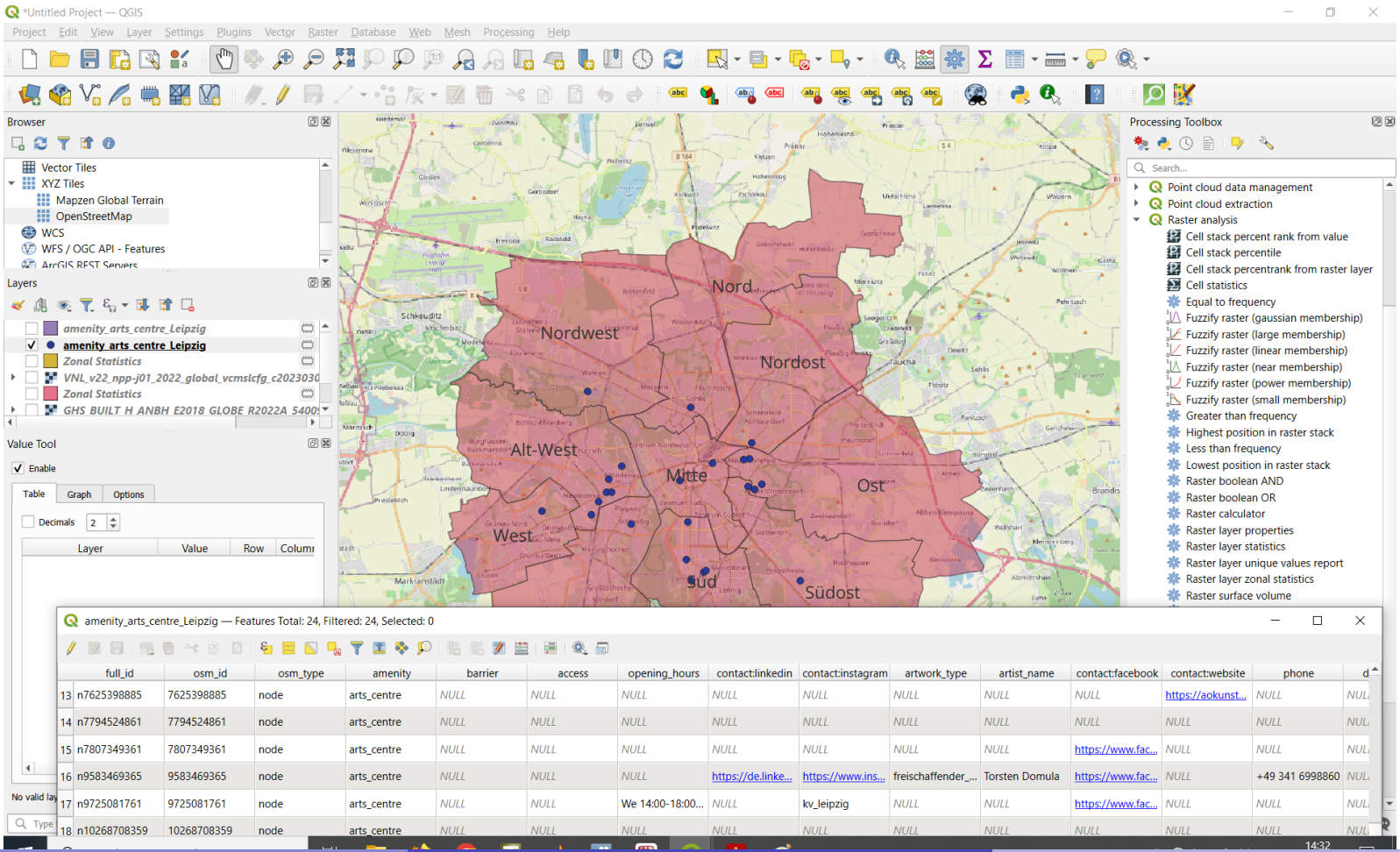
QGIS Application: Art Galleries from OSM

- Finally, *The QuickOSM* (Open Street Map) plugin is very useful when working with amenities and streetmaps within cities. It uses an overpass API acting as a database on the web. You can install the *The QuickOSM* plugin in QGIS by clicking on *Plugins* and selecting it.
- Selecting *Key = amenity, Value = arts_centre, IN: Leipzig, RUN QUERY* yields two layers with arts galleries, one with the galleries as points (more useful), one with some galleries as polygons



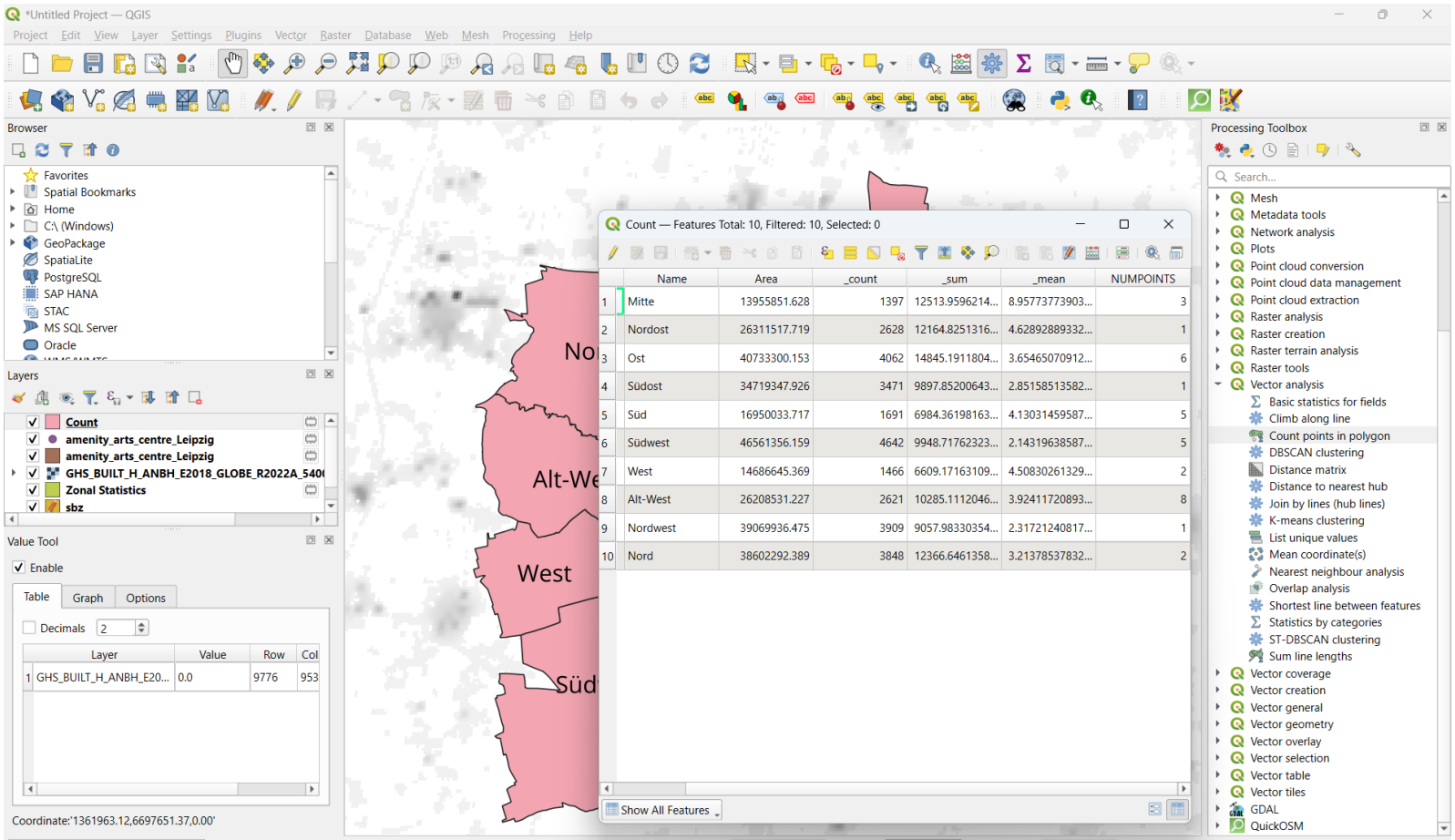
QGIS Application: Art Galleries from OSM

- You can uncheck all the layers you don't want to be displayed. Look at the locations of the galleries (as points) and find more information in the attribute table (right-click on the points layer)



QGIS Application: Art Galleries from OSM

- Count the art galleries in each district: *Vector -> Analysis Tools -> Count Points in Polygon*. As the polygon layer, select *sbz*. As the points layer, select the *OSM Query* points layer. Look at the attribute table of the newly created layer: 6 art galleries in the East and 8 in Alt-West - is this already a sign of future gentrification?



Wrap-Up of Unit 6

- **Geo-spatial data** is place-based data that comes with geographical identifiers, such as coordinates.
- It can be visualized, analyzed and manipulated with geographical information system software such as **QGIS**.
- Geo-spatial data comes in two main forms: **Vector data** (points, lines, polygons) and **raster data** that stores values of gridcells.
- The polygons of a vector layer can represent city districts, cities or countries. The information stored in a raster file (nighttime lights, building height, temperature) can be interacted with the borders of a polygon layer (such as city districts) to compute the average lights or mean building height in that polygon. The result of the computation is added to the **attribute table**.
- The Quick OSM plug-in is useful for adding Open Street Map data on amenities and analyzing them at the city district level.
- The attribute table can be **exported as a CSV file** for further statistical analysis.