UN-REDD PROGRAMME

Tanzania Methodological Brief:

Technical report with appended training materials for the production of multiple benefits maps to inform REDD+ planning and safeguards policies using open-source GIS software

UN-REDD PROGRAMME

Technical Report: August 2013

Cambridge, United Kingdom

Produced as part of two working sessions held in Dar es Salaam and Morogoro, United Republic of Tanzania

The UN-REDD Programme is the United Nations Collaborative initiative on Reducing Emissions from Deforestation and forest Degradation (REDD) in developing countries. The Programme was launched in September 2008 to assist developing countries prepare and implement national REDD+ strategies, and builds on the convening power and expertise of the Food and Agriculture Organization of the United Nations (FAO), the United Nations Development Programme (UNDP) and the United Nations Environment Programme (UNEP).

The United Nations Environment Programme World Conservation Monitoring Centre (UNEP-WCMC) is the specialist biodiversity assessment centre of the United Nations Environment Programme (UNEP), the world's foremost intergovernmental environmental organisation. The Centre has been in operation for over 30 years, combining scientific research with practical policy advice.

Prepared by Corinna Ravilious, Matea Osti, Rebecca Mant, Lisen Runsten and Lera Miles

Copyright: FAO

Copyright release: This publication may be reproduced for educational or non-profit purposes without special permission, provided acknowledgement to the source is made. Re-use of any figures is subject to permission from the original rights holders. No use of this publication may be made for resale or any other commercial purpose without permission in writing from UNEP. Applications for permission, with a statement of purpose and extent of reproduction, should be sent to the Director, UNEP-WCMC, 219 Huntingdon Road, Cambridge, CB3 0DL, UK.

Disclaimer: The contents of this report do not necessarily reflect the views or policies of UNEP, contributory organisations or editors. The designations employed and the presentations of material in this report do not imply the expression of any opinion whatsoever on the part of UNEP or contributory organisations, editors or publishers concerning the legal status of any country, territory, city area or its authorities, or concerning the delimitation of its frontiers or boundaries or the designation of its name, frontiers or boundaries. The mention of a commercial entity or product in this publication does not imply endorsement by UNEP.

Should readers wish to comment on this document, they are encouraged to get in touch via:
ccb@unep-wcmc.org.

Citation: Ravilious, C., Osti, M., Mant, R., Runsten, L. and Miles, L. (2013) Tanzania Methodological Brief: Technical report with appended training materials for the production of multiple benefits maps to inform REDD+ planning and safeguards policies using open-source GIS software. Technical Report. Prepared on behalf of the UN-REDD Programme. UNEP World Conservation Monitoring Centre, Cambridge, UK

Acknowledgements: With thanks to the workshop participants (Richard Giliba, Veronica Hailakwahi, Ludovic RA Kashaga, Almas Kashindye, Shani Khalid, Boniface Mbilinyi, Yohane Mwampashi, Msalika Pastory) and to Anssi Pekkarinen at FAO and Chris Graves at UNEP-WCMC whose help was instrumental in enabling us to carry out effective working sessions and produce these training materials. We would also like to thank Saber Razmjooei at Lutra Consulting for providing introductory QGIS training and support to the UNEP-WCMC team. Finally, thanks to Tania Salvaterra for her work on edit and layout.

Contents

1. Introduction 5
2. Capacity Development 5
3. Open-source Software 6
   i. Why use Open-source Software? 6
   ii. Which open-source GIS software? 6
   iii. Technical training method: Xubuntu on Live USB 7
   iv. Challenges of using open-source 8
   v. Good working practices and spatial data management 8
4. Methods for development of enhanced national scale spatial datasets 9
   i. Compilation of dataset on wildlife corridors 9
   ii. Enhancing spatial data on forest reserves and Participatory Forest Management (PFM) in Tanzania 9
   iii. Compilation of dataset on population density based on 2012 census 10
   iv. Compilation of dataset on importance of forests for limiting soil erosion 10
   v. Digitization and Georeferencing of maps of bee reserves 11
5. Methods for evaluating the potential of the NAFORMA inventory to model tree species richness .. 12
6. Methods for production of maps and statistics to serve the REDD+ planning process 16

ANNEX 1 – Training materials used in working sessions 17

1. Understanding Coordinate Reference System (CRS) in QGIS 17
   1.1. Changing the default CRS (projection) settings 17
   1.2. Changing the CRS of the current project 18
   1.3. Projecting data / saving data to a new CRS 19
   1.4. Reprojection of Raster layers 20
2. Saving a QGIS project 20
3. Adding new layers 21
   3.1. Adding a vector layer 21
   3.2. Adding a Raster Layer 22
4. Changing symbology 22
   4.1. Vector symbology 22
   4.2. Raster symbology 24
5. Installing plugins 26
6. Useful plugins for spatial analysis to inform REDD+ planning and safeguards policies 27
   6.1. Qnote 27
   6.2. Group stats 27
   6.3. SEXTANTE 28

6.4. QMARXAN 29
6.5. The clipper tool 29
6.6. The Raster Calculator 30
6.7. Add delimited text 30
7. Querying data 31
8. Joining Tables 32
9. Map Layouts 35
10. Comparing area statistics generated from different projections 43
11. Calculating average above ground biomass by LULC type for Saatchi, Baccini and the NAFORMA sample of Rufiju district 45
12. Using IUCN Red List species data and generating species richness maps 48
13. Using NAFORMA data to create a map of tree species richness at the inventory locations. 61

1. Introduction

This methodological brief has been produced as part of a UN-REDD Programme project to support Tanzania in the development of multiple benefits maps to inform REDD+ safeguards policies, conducted between January and August 2013. The project had two main objectives: (1) to contribute to, and support the development of enhanced national scale spatial datasets, statistics and maps on biodiversity and ecosystem services for Tanzania for the purpose of informing REDD+ policies and measures; and (2) to build capacity within Tanzania on spatial analysis of datasets of relevance to multiple benefits and environmental safeguards for REDD+ and on approaches to developing information systems for safeguards, using open-source GIS software. Another objective of the work was to explore the data from the National Forest Inventory (NAFORMA), for the purpose of informing REDD+ planning and environmental safeguards, alongside spatial data from other sources.

Objectives 1 and 2 were achieved simultaneously through two working sessions in Tanzania of five weeks in total, covering mapping work and enhancement of national spatial datasets, using open-source GIS software (QGIS, SAGA and GRASS). The working session included eight participants from four institutions: Tanzania Forest Service, Sokoine University of Agriculture, Forestry Training Institute – Olmotonyi and FAO Tanzania. Two facilitators from UNEP-WCMC prepared and led the working sessions, with support from the NAFORMA project, the FAO-FIN Programme and UN-REDD Programme in FAO.

This methodological brief outlines the technical approaches used in the working sessions.

2. Capacity Development

A significant aim and emphasis of the project was to develop capacity within country on mapping and analyzing relevant data to inform REDD+ multiple benefits planning and safeguards policies. In order to ensure that the work was undertaken as collaboratively as possible, two working sessions were provided in Tanzania. Holding the training in Tanzania enabled the training to cater for a larger group of people, to make use of in-country expertise to a greater degree, as well as access data that would otherwise be difficult to source. Additionally, it helped ensure compatibility with in-country systems and provided the best opportunity for integrating the tools and approaches with the regular work of the working session participants.

The focus of working session 1 was predominantly GIS training. The training was centered around the exploration of a particular set of topics and questions, including exploration of the NAFORMA tree species data and the extent of occurrence data for species in the IUCN Red List. Throughout this session, the entire group of participants undertook each activity. Working session 2 built on the skills developed in working session 1 but with a stronger focus on the mapping and analysis of topics relevant to REDD+ multiple benefits to inform REDD+ planning and safeguards policies. Working session 2 relied more heavily on group discussions on how to best process and improve available data, and how to present the maps. This was followed by group work to undertake different specific tasks identified from the discussions. This problem solving approach taken in working session 2 complemented the more traditional training approach applied in working session 1.

In both working sessions, different parts of the training included, where appropriate:

1) Step-by-step training materials provided to participants for undertaking a specific task
2) Presentation of the topic followed by brief guidance notes for participants to undertake the tasks
3) On-screen, trainer-led (follow-along) exercises

4) Exploration of particular suites of tools
5) Self-planning of GIS work with trainer guidance for use of tools.

This variety of training methods allowed the participants to work both independently, in pairs and as a group to think about the task at hand and which tools they required. This approach was also intended to facilitate the transfer of the knowledge gained in the training sessions to their everyday work. The step-by-step guides were popular with participants, but if used alone they would not have provided maximum learning opportunity. Learning how to find their way around the software and exploring appropriate tools for different tasks was an important part of the learning process.

3. Open-source Software

i. Why use Open-source Software?

Open-source software is released under a license that allows software to be freely used, modified, and shared (http://opensource.org/licenses).The decision was taken to use open-source software in the training to ensure the future sustainability of the knowledge gained by the participants during the training, and for materials produced to be useful for a larger number of people, who may not be able to procure commercial software licenses. The development of open-source software is typically collaborative, meaning that developers improve the code and share the changes within the community. An example is the GNU General Public Licence v2 (http://www.gnu.org/) used for Quantum GIS (QGIS), which protects four special freedoms:

1) The freedom to run the program, for any purpose
2) The freedom to study how the program works and change it so it does your computing as you wish
3) The freedom to redistribute copies so you can help your neighbor
4) The freedom to distribute copies of your modified versions to others

This means that the software is often continuously developed and added to. By sharing in this way the whole community is given a chance to benefit from changes made by others. Access to the source code is a pre-condition for this..

One of the few disadvantages of open-source compared to proprietary software (owned by corporations) is that it might not be as fully integrated, for example it may not be as easy to link tools together into a single GIS procedure or you may need to select tools from multiple pieces of software for undertaking a particular task or set of tasks. However, the licenses for proprietary software are restrictive, and the costs of the licenses as well as additional costs for extensions to the software are often beyond the means of many individuals and organizations.

ii. Which open-source GIS software?

For this UN-REDD project Quantum GIS (QGIS) was selected as the main GIS software to use along with GRASS, SAGA and the Open Foris Geospatial Toolkit (to a lesser extent). QGIS was recommended by FAO, and the Tanzanian Forest Service and Sokoine University of Agriculture (SUA) already had some experience of using it. There are hundreds of other open-source GIS software to choose from. The website http://freegis.org provides a useful overview currently listing 356 projects.

Some of the strong advantages of QGIS are that it runs on multiple platforms (Linux, Unix, Mac OSX, Windows and Android operating systems). It supports vector and raster processing and can access and utilise tools available in other open-source GIS packages. GRASS, for example, comes as part of the QGIS installation and is fully integrated while it can be run independently. Other software can be accessed through additional freely available plug-ins. These additional plug-ins are built by the wide user community using the python scripting language. For example, the Sextante toolbox plug-in provides a geoprocessing environment similar to the ArcGIS toolbox which allows you to access tools available within QGIS, GRASS and SAGA GIS. QGIS is a commonly used alternative to the proprietary ESRI ArcGIS software with a user friendly GUI and has a similar processing power, which makes it an excellent alternative. QGIS is a lightweight installation compared to ArcGIS. It takes up less space and will run on less powerful computers. However, to ensure speed and processing ability for complex analysis, a high capacity computer is required, of similar specification recommended for ArcGIS. The greatest advantage of using Open-source GIS is that it is can be accessible to all and not limited to just 1 or 2 licensed individuals in an organization. There is a large user community and a growing wealth of online resources and training materials.

iii. Technical training method: Xubuntu on Live USB

The computer set-up used for the training throughout the working sessions was an install of the Xubuntu operating system on a live USB. This followed the approach taken by FAO in other training held under the NAFORMA project and UN-REDD Programme in Tanzania, and so complemented previous capacity building activities. FAO staff were able to provide efficient knowledge transfer to UNEP-WCMC and supplied the required ISO image containing the relevant software for installation of the live USB (see Annex 1a and 1b).

The advantages of this approach were that:

- It provided a consistent install of both operating system and software, which avoided problems related to different configurations on different computers
- Developers of open-source software often work in the Linux environment, so the Linux versions are often more stable than the Windows versions
- The software was not installed on a single machine so was portable
- Xubuntu is a lightweight Linux operating system, which work well on less powerful machines
- It is easy to customize and install other software packages (providing you have an internet connection)
- Participants gained experience with an alternative operating system to Windows
- The risk of virus infection is significantly lower in Linux operating systems than in Windows
- Other non-GIS open-source software alternatives were introduced to the participants, including LibreOffice

Booting of live USB operating systems differ slightly depending on the make and model of the computer, but this step was successful for all the machines in the training session. The main challenge of using live USBs was the introduction of a new operating system in addition to the new GIS software. However, this approach was well received by the participants. In addition to using live USBs, UNEP-WCMC made sure that all participants had a working copy of the all the GIS software (QGIS, GRASS and SAGA) installed in the Windows environment on their computers, including all the additional plug-ins for QGIS used in the working sessions. Free antivirus software were also installed where no antivirus software were present.

iv. Challenges of using open-source

The main challenge of using open-source software is the limited knowledge within Tanzania, and that ArcGIS is taught as the standard software. Although QGIS is a good alternative to ArcGIS there are still some initial challenges to be aware of when using the software, but these should not prevent users from switching to QGIS or open-source software.

- Treatment of map projections can be confusing in QGIS and can lead to mistakes
- Symbolizing raster data is less advanced in QGIS than ArcGIS
- Raster analyses is different to ArcGIS
- Development of customized tools and workflow requires python knowledge unless using the Sextante plug-in or GRASS model builder
- The model builder (Sextante plug-in) for QGIS is not stable for some tools, particularly in the Windows environment
- GRASS has its own data structure and management set up and may require further training for proficiency
- Data have to be imported into GRASS before the GRASS model builder can be used
- Tools are not so well integrated, which can make trouble shooting more difficult
- Tools are not as well tested as in ArcGIS
- Map production facilities are good but less sophisticated than ArcGIS
- Creating maps in a geographic coordinate system prevents adding a scale bar in the final layout.

v. Good working practices and spatial data management

Good spatial data management and working practice is essential for efficient management and analysis of data. Prior to the working session, UNEP-WCMC compiled a data registry consisting of global, regional and national data relevant to safeguards and multiple benefits for REDD+ in Tanzania. The data registry is a simple MS Excel table storing information about datasets, including their source, year, citation, contact information and any restrictions associated with the data. This was a useful starting point to catalogue the knowledge about potentially useful datasets available on Tanzania, including information on biodiversity, land use/land cover, management units, pressures on forest resources and ecosystem services.

In working session 1, participants were introduced to the data registry and a number of ways to plan and document their GIS work. For example, a user generated plug-in called QNote enables users to add notes directly in the map to document a description of what the QGIS map document contains, or to analyse methods applied. Another approach was to provide a technical work plan to get users to think about how to undertake a particular activity, what GIS methods tools to use and what GIS data inputs are required.

This data registry was updated during the working sessions to incorporate additional datasets identified by the participants. A metadata catalogue storing information alongside the data held within a spatial database would be ideal but in the absence of a central database for storage of spatial data this is a good starting point for collating existing knowledge and keeping track of datasets.

A spatial database is more than just a storage facility for spatial data and can be used to query and share spatial data for GIS as well as other applications. There are many advantages of centrally storing data in a spatial database and it does not necessarily mean huge costs associated with commercial software. There are open-source options available such as PostgreSQL with PostGIS (www.postgresql.org/about/news/1387) that have a long history of use with Quantum GIS, with inbuilt plug-ins to allow users to easily manage, access and retrieve data held in a PostGIS spatial database. A presentation was given on Spatial Database Management Systems in working session 2, but there was not

time to focus on this topic in depth during the working sessions or to demonstrate the setting up of a spatial database. The working session identified that future capacity development in this area would be helpful.

4. Methods for development of enhanced national scale spatial datasets

i. Compilation of dataset on wildlife corridors

Wildlife corridors are crucial habitats for the health of ecosystems and wildlife populations in Tanzania, but they are often unprotected and under pressure from human activities. For the working session, no spatial dataset was available that showed how wildlife corridors link together protected areas in Tanzania. The most recent, publicly available work on listing and mapping important large corridors in Tanzania was led by the Tanzania Wildlife Research Institute and published in 2009 (see http://www.tzwildlifecorridors.org/). In working session 2, participants used the data available on this website and in associated reports to create a dataset that illustrated the general location of corridors assessed in this work. The corridors were symbolized with dotted lines, to illustrate their general extent. The resulting map was sent to TAWIRI for comments.

ii. Enhancing spatial data on forest reserves and Participatory Forest Management (PFM) in Tanzania

Currently, there are several different spatial data versions of Tanzania's forest reserves in circulation. In working session 2, a spatial dataset with polygons of forest reserves from 2010 provided by TFS was compared to the data on forest reserves held within the World Database on Protected Areas (WDPA). A number of differences were identified and resolved both in terms of spatial boundaries and attribute information. In some instances, the data in the TFS version was deemed correct, and in other instances the WDPA was considered correct. The decisions made in the working session should be further quality checked by experts. The only attribute information that was taken into consideration in the May 2013 working session was the forest reserve name. An assumption was made that the spelling of the forest reserve name was correct in the 2010 dataset. There were additional discrepancies where a solution was not evident, and where there was not time to make further investigation.

Experts have advised that it is likely that some of the reserves in both these datasets may have been degazetted. There are also overlaps with other protected area designations that may indicate outdated information. Also, both datasets omit the Nature Reserves gazetted since 2007 – but where there are shape files available that can be added.

Furthermore, TFS also provided a non-spatial table, listing forest reserves by forest name, which had never been linked to spatial data. This data is also from 2010. A preliminary join has been made to the spatial data but further comparison of this list with the spatial data is required and along with any additional information that can be made available beyond 2010.

A related issue concerns land under Participatory Forest Management (PFM) (Community based forest management (CBFM) or Joint Forest management (JFM)). The following tables are known, but lacking spatial data linked to them:

- "Baseline for declared village land village forest reserves in Tz – 2011"
- "Baseline for signed JMAs in Tz – 2011"
- "National Census for Participatory Forest Management, Community Based Forest Management, 2008". Listed by forest name and village name. The forest name is not always a forest reserve name.

- "National Census for Participatory Forest Management, Joint Forest Management, 2008" Listed by forest name, village name and forest status (NFR or LAFR). The table also has information on protection/production status, and the current (as of 2008) stage of JFM establishment.

In the working session, a team of three participants worked to quality check these datasets. In collaboration with a UNEP-WCMC facilitator, they developed a methodology for error checking the data, and document the errors found and the changes made. Knowledge of how to systematically quality check data is a highly useful skill and this exercise was appreciated by the participants.

iii. Compilation of dataset on population density based on 2012 census

The 2012 population census report on population density per ward had just been released at the time of the working sessions and could only be obtained in a Word document. An automated extraction of the tables of population counts by ward to CSV files was made possible by FAO. These tables were joined together to form a single table, which was subsequently joined to a spatial layer of census units. This process was not straightforward as the 2012 spatial dataset for ward level boundaries was not yet available, so the 2002 ward boundaries from the previous population census had to be used. There had been changes to boundaries between 2002 and 2012. Significant effort was required to prepare the data to link the 2012 data tables to the 2002 ward polygons. Where ward boundaries had not changed the 2012 data were assigned directly to the ward. Where ward boundaries had changed significantly, efforts were required to try to modify and link the 2012 census figures to the old boundaries. The following issues were encountered and addressed:

- wards had changed name: 2012 data manually linked to old ward name
- single ward now 2 wards: 2012 population data merged and figure added to single ward
- 2 or more wards merged to form new ward.: 2003 ward boundary merged and data connected to merged ward

The population density was calculated from the population count and ward area. This map presents a significant update to representation of population data in Tanzania since 2002 and from this work it was possible to calculate change in population density. This map should be reproduced once the 2012 boundaries are release.

iv. Compilation of dataset on importance of forests for limiting soil erosion

A simple "quantitative" approach was taken for evaluating the importance of forests for soil stabilization and limiting soil erosion. Importance was estimated as a function of slope, rainfall and the presence of important features downstream that could be adversely affected by erosion such as a dam or water body. For example, forests may play an important role in soil stabilization on steep slopes in high rainfall areas with important downstream activities that benefit from clear water such as population centres, hydroelectric plants and navigation routes. This dataset can therefore be used to help answer questions such as:

- Where is forest loss likely to result in erosion and consequent sedimentation – and where is this potentially damaging?

- Where might retaining or restoring forest play an important role in retaining soil or reducing soil erosion?

The analysis was undertaken by generating single layers of low, medium and high precipitation; slope; and a binary layer of catchments of dams and water bodies. These layers were summed and clipped to the forest extent. This generated 6 classes representing low to high importance of forest for preventing soil erosion:

Slope	Precipitation	Important catchment	Sum	Slope	Precipitation	Important catchment
1	1	0	2	low slope	low precipitation	not in catchment of dam
1	1	1	3	low slope	low precipitation	in catchment of dam
1	2	0	3	low slope	medium precipitation	not in catchment of dam
2	1	0	3	slope	low precipitation	not in catchment of dam
1	3	0	4	low slope	High precipitation	not in catchment of dam
1	2	1	4	low slope	medium precipitation	in catchment of dam
2	1	1	4	slope	low precipitation	in catchment of dam
2	2	0	4	slope	medium precipitation	not in catchment of dam
3	1	0	4	high slope	low precipitation	not in catchment of dam
1	3	1	5	low slope	High precipitation	in catchment of dam
2	3	0	5	slope	medium High precipitation	not in catchment of dam
2	2	1	5	slope	medium precipitation	in catchment of dam
3	1	1	5	high slope	low precipitation	in catchment of dam
3	2	0	5	high slope	medium precipitation	not in catchment of dam
2	3	1	6	slope	High precipitation	in catchment of dam
3	3	0	6	high slope	High precipitation	not in catchment of dam
3	2	1	6	high slope	medium precipitation	in catchment of dam
3	3	1	7	high slope	High precipitation	in catchment of dam

(Colours as per legend on Map 12)

This approach could be further refined for example by adding in additional layers such as soil type or by assessing and giving weight to the different elements according to their importance before summing together.

v. Digitization and Georeferencing of maps of bee reserves

In the feedback given after working session 1, participants requested training on how to digitize maps in QGIS. An exercise in digitizing maps of proposed bee reserves was therefore conducted in working session 2. These reserves were very small and therefore not included in a map in the final report, but the exercise addressed the capacity building request.

5. Methods for evaluating the potential of the NAFORMA inventory to model tree species richness

The NAFORMA inventory provides a great amount of data on forest resources in Tanzania from over 29 000, 15m-radius plots grouped into L-shaped clusters of between 6 and 10 plots. The data collected by NAFORMA was analysed in order to assess what information it could provide about tree biodiversity. Investigations included looking at the likely proportion of species in different vegetation types recorded by the survey, and assessing whether the data could be used to create a wall-to-wall species richness map of Tanzania.

The NAFORMA survey aimed to identify the species of every tree measured within a plot. Initial investigation of the tree inventory (containing over 240 000 trees) showed that (1) the species code and the species name recorded sometimes disagreed and (2) some codes and some names recorded were not present in the NAFORMA species checklist. This was either due to the code or species being completely missing from this checklist or due to spelling errors. Due to these disagreements between species code and species name, it was not possible to just join on correct spellings from the species checklist. In order to investigate the distribution of species and the number of unique species in different locations, we needed to address the errors in the inventory, as far as possible. Without access to the field notes, errors could be more easily identified in the species names than errors in the codes. Therefore, the decision was taken to clean the species name information where spelling errors were evident, but not attempt to address possible errors in the species codes. The cleanup involved checking the species names against the NAFORMA checklist, and correcting the spelling errors found.

- Where a species name in the inventory was not listed in the checklist, further checks were made to see if the species was recorded in the African Plant Database (APD) or the Global Biodiversity Information Facility (GBIF)
- Where there was a species name that could be found in the checklist, but the code indicated in the inventory did not match, this error was not corrected
- For over 5000 tree records, the species code was “unknown” or “unlisted” but a species name was present. For these records, any spelling errors were addressed, but codes were not corrected, and these species were included in the final species count
- Where a species code was present (that matched a code in the checklist), but the species name was indicated as “unknown” or “unlisted”, the species name was added from the checklist
- The remaining trees with name recorded as “unknown” or “unlisted”, or had vernacular names only, were excluded from further analysis

A full summary of issues can be found in Annex 4. The cleaned list was required to produce the most accurate estimate possible at this stage of the total number of species recorded by NAFORMA, and the number of species within plots and within clusters of plots. When a final count of total number of species recorded in the inventory was made, this count was based on species name.

Following cleaning of the tree inventory, our calculations indicate that the NAFORMA survey recorded 1 229 species, which includes 67% of the species listed in the NAFORMA species checklist (983 species) and an additional 246 species not contained within this original expected species list. Over 29 000 trees (12 % of the total 240 000 trees measured in the survey) were also only identified to genus level. In addition, 33% of the species on the checklist were not found in the plots. As Tanzania is such a biodiverse country, the plot sampling area of just over 700m^2 is likely to have undersampled overall diversity across the country.

Additionally, the number of species identified in a plot may have been influenced by the difficulty of species identification (rare species are often mistaken for more common ones, for example[1]).

The ability of the NAFORMA survey to measure tree species richness within an area was explored by creating a graph of the number of new species found when additional plots were sampled in an area. Figure 1 shows the cumulative number of species found in selected vegetation types as more plots were sampled (within each NAFORMA plot, the vegetation type was recorded). If this curved flattened out it would have indicated that the survey had recorded most of the richness in that vegetation type. However, as the figure shows, the total number of species recorded in a vegetation type increased with the number of plots surveyed for that vegetation type; suggesting that NAFORMA did not cover enough plots to estimate the total species richness for each vegetation type. The curve of the humid montane forest is the steepest, showing without doubt that if more plots were sampled, many more species would have been found. Sampling is most complete in the open woodland habitat.

The plot level data does provide an estimate of localised species richness (which is likely to be only a portion of the species occurring within a wider area). Even so, localised species richness is highly variable; the number of species recorded within an individual plot varied from 0 to 23 (0 meaning either no trees identified to species level or no trees present).

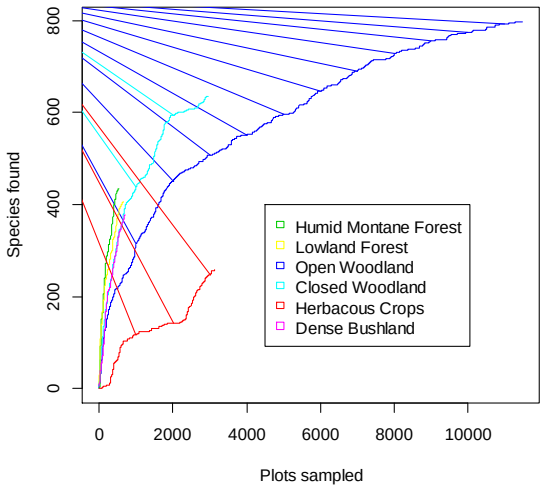


Figure 1: Cumulative number of species found in selected vegetation types.

In addition to investigating whether NAFORMA data could provide an estimate to the total species richness for different vegetation types, we also investigated whether the data could be used to estimate species richness in unsampled locations. The number of species in a location will vary depending on the characteristics of that location, including environmental attributes such as temperature and rainfall. Hence, if the relationship between a site's characteristics and the number of species is known, the potential species richness of a site can be estimated from the site's characteristics. We studied whether the NAFORMA data could allow an understanding (model) to be created on how the number of tree species varied between plots. The data were investigated using generalized linear models and generalized additive models with a quasi-poisson distribution of errors using the statistical package R. A simple model based only on the NAFORMA vegetation type of the plot explained 33% of the variability in the number of species found at a plot. A generalized linear model including additional information on the mean temperature,

[1] Ahrends, A., Rahbek, C., Bulling, M. T., Burgess, N. D., Platts, P. J., Lovett, J. C.,Marshall, A. R. (2011). Conservation and the botanist effect. *Biological Conservation*, *144*(1), 131–140.

temperature range, average rainfall, precipitation seasonality, PET, distance to road, latitude, longitude and sampling strata of a plot together only explain a further 8% of the variability in number of tree species recorded at a plot. Information on the environmental characteristics of plots was extracted from the WorldClim database, and Waterworld was used to calculate the PET. The best fitting model tested was not able to explain over 50% of the differences in tree species richness between plots and the model was particularly bad at identifying where plots are likely to have high species richness (figure 2, GLM model null deviance = 94083, residual deviance = 56944, d.f. = 28835; GAM model null deviance =94083, residual deviance =50513, d.f. = 28805). Part of the reason for the variability between plots that cannot be explained by the models tested may be due to the small size of the plots in relation to the heterogeneity of forests. Localised species richness can be very variable and the number of species recorded within an individual plot within just the closed woodlands plots of the NAFORMA survey varied from 0 to 23. As it is not possible to predict the species richness of a site with a high degree of accuracy from this model, and the richness being modeled is localised, is not possible to produce an accurate wall-to-wall map of plot scale species richness in Tanzania using this approach.

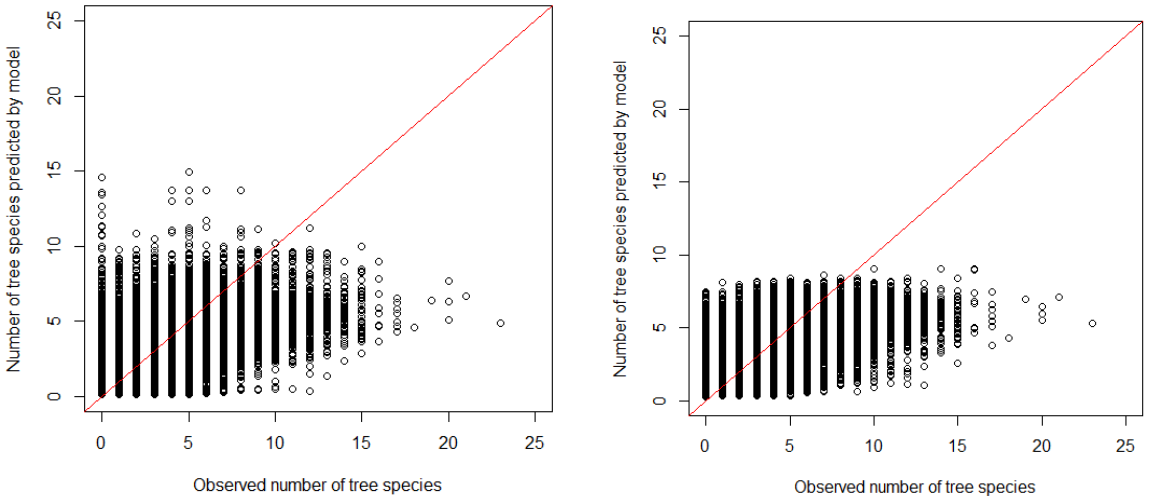


Figure 2: The observed species richness for each plot in the NAFORMA data against the tree species richness predicted by a general linear model (left) and general addititive model (right) based on the environmental characteristics of the plot (mean temperature, temperature range, average rainfall, precipitation seasonality, PET, distance to road, latitude, longitude, sampling strata and NAFORMA vegetation cover of the plot). (Plots with more than one NAFORMA vegetation cover were excluded from the analysis). If the models had explained all of the variability in the plots, all of the points would have fitted on the red line, the greater the scatter and deviance of the plots from this line the less the model explains the variability in the plots tree species richness.

The parameters which control the number of species within a plot may vary between different vegetation types. It was therefore investigated whether a greater proportion of the variability in species richness could be explained if open woodland plots were assessed alone. However, a model including just the open woodland plots only explained 15% of the variability in the number of tree species found (Generalized linear model of number of tree species in a plot by mean temperature, temperature range, average rainfall, precipitation seasonality, PET, distance to road, latitude, longitude and sampling strata using a quasi-poisson family of errors, null deviance = 24229, residual deviance = 20661, d.f.= 11484).

As the area of one plot is very small, the number of species within each cluster may provide a more representative estimate of the number of species within the wider habitat at each cluster location. The area of one cluster (0.4 to 0.7 ha), is closer to the average area predicted as needed to differentiate between the diversity of different areas (Ahrends et al. in press). Hence, it was investigated whether the number of

unique tree species recorded in a cluster could be used to produce a more specific model of tree richness in Tanzania. However, it is important to note that the number of plots in a cluster varied according to the sampling strata to which the cluster belonged. This needed to be taken into account in the model. In the cluster model across all vegetation types, 55% of the variability remained unexplained and again the model was not good at predicting the clusters with high species richness.

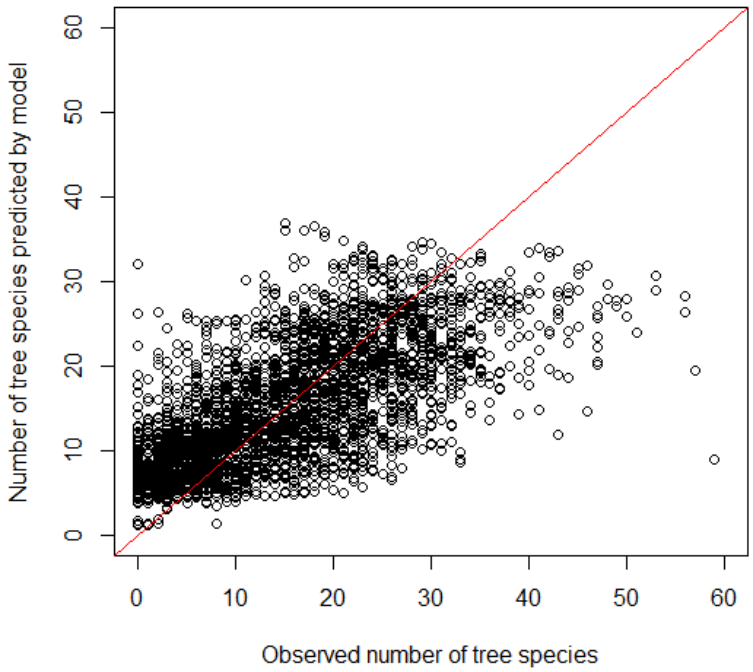


Figure 3: The observed species richness for each cluster in the NAFORMA data against the tree species richness predicted by a general linear model based on the environmental characteristics of the plot (mean temperature, temperature range, average rainfall, precipitation seasonality, PET, distance to road, latitude, longitude, sampling strata and land cover the cluster falls in). The model GAM null deviance =24383, residual deviance =13337, d.f. = 3015.

Overall, since the NAFORMA inventory does not fully register the tree species richness pattern in most vegetation types, it is not possible to simply extrapolate out from the average plot richness to the total richness of the vegetation types and the models are not good at predicting where high species richness occurs, it would be potentially misleading to create a wall to wall tree species richness map from the inventory data.

The NAFORMA data also contains information relevant to species of particular conservation importance including; species that are a) only found in (endemic to) the Eastern Arc Mountains of Tanzania or b) threatened with extinction as proposed by the Eastern Africa Plant Red List Authority. The species recorded in the NAFORMA data was compared to a list of 52 tree species listed as endemic to the Eastern Arc Mountains (Burgess et al 2007). A boundary for the Eastern Arc Mountains was obtained from Dr. Phil Platts at the University of York, which was used to extract the NAFORMA plots within the Eastern Arc area. Of the 52 species, NAFORMA field workers found 11 within the Eastern Arc Mountains (in 1639 NAFORMA inventory plots covering 116 ha across 264 clusters). Additionally, 8 species listed as Eastern Arc endemics were picked up by the NAFORMA survey outside of the Eastern Arc Mountains. Recent updated assessments using confirmed specimens suggest that three of the species are not strict endemics to the Eastern Arc; two occur in coastal areas and one in remote forest near northern part of Lake Tanganyika (Roy Gereau, pers. comm.).

The species recorded in the NAFORMA inventory were further compared to a list of 394 Tanzanian species recorded as threatened by the Eastern Africa Plant Red List Authority (EAPRLA via Roy Gereau, pers. comm.). Only 38 of these species were recorded in the NAFORMA survey which again suggests that NAFORMA sampling design, which was stratified to avoid remote and inaccessible places, made it less likely

that locations with rare species have been included, and may also reflect the fact that rare species are at higher risk of not being correctly identified in the field. The humid montane forests had a particularly high number of threatened tree species (20) given the number of plots sampled.

6. Methods for production of maps and statistics to serve the REDD+ planning process

Both working session 1 and 2 included numerous group discussions on how maps can help REDD+ planning to account for safeguards and multiple benefits, and what the best way to construct such maps might be. These discussions directed the day to day work in the working sessions and determined the appropriate maps and analyses to undertake. Working session participants and other Tanzanian colleagues often sourced additional data based on these discussions, and several solutions emerged from this process.

ANNEX 1 – Training materials used in working sessions

These materials are an improved version of the resources used during the working sessions. These are step-by-step tutorials written primarily to introduce methodologies and enable users to begin to explore the software. Although every effort has been made to provide clear guidance, it cannot be guaranteed that the materials are error-free and we welcome any feedback. They are aimed for experienced GIS users to support transition to QGIS.

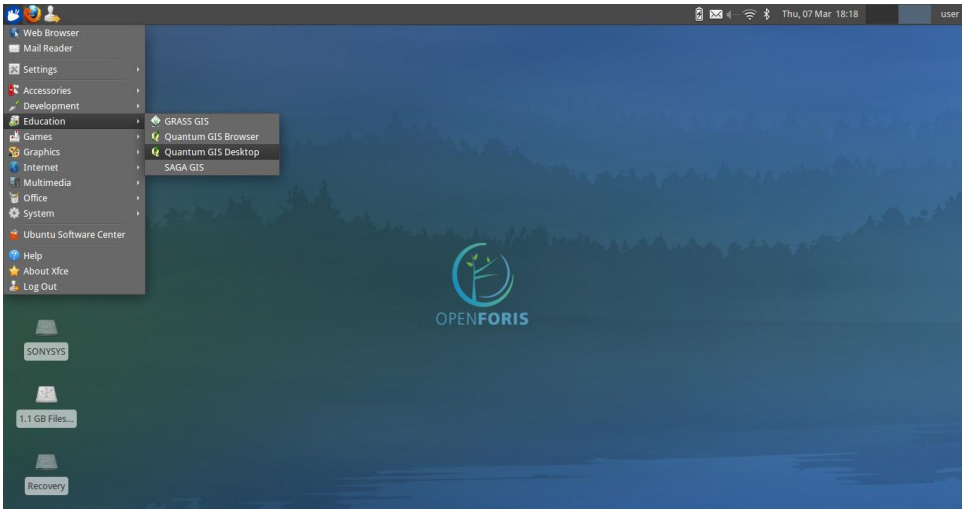
The following materials have been produced using the Xubuntu operating system running on installation on the Live USB. Instructions remain the same for the windows environment although the screen illustrations may have minor differences in their appearance. Some references are made to tutorial files shared with participants of the working sessions.

1. Understanding Coordinate Reference System (CRS) in QGIS

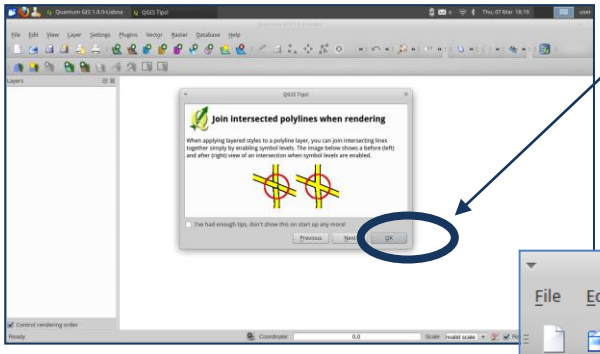
This is a VERY IMPORTANT section as there are a number of places where projections can be set or altered in QGIS and with a lack of knowledge on how projections work in QGIS, changing projections can accidently be done in the wrong way.

1.1. Changing the default CRS (projection) settings

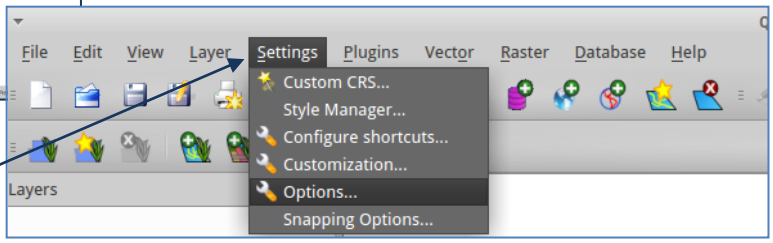
- Click on the **Applications icon** to drop down the main menus
- Click on **Education >> Quantum GIS desktop** to open QGIS



The QGIS interface will open (as below)

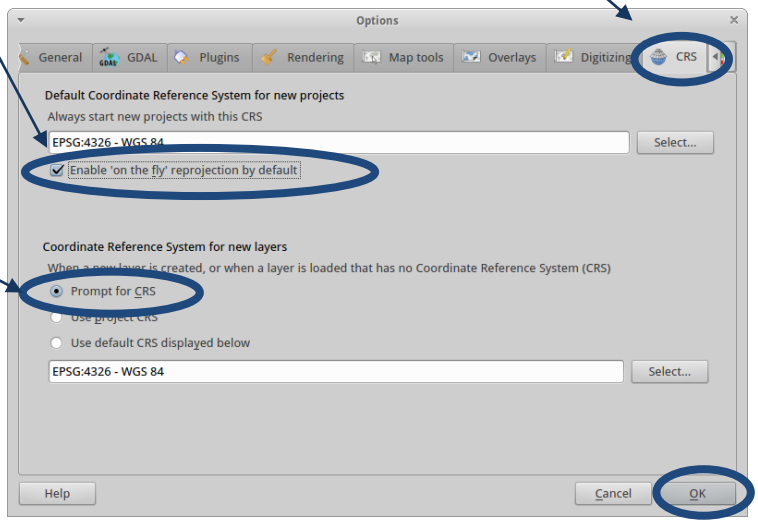


- Click **OK** to dismiss the QGIS tips



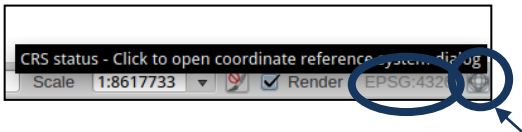
- Click on **Settings >>Options**

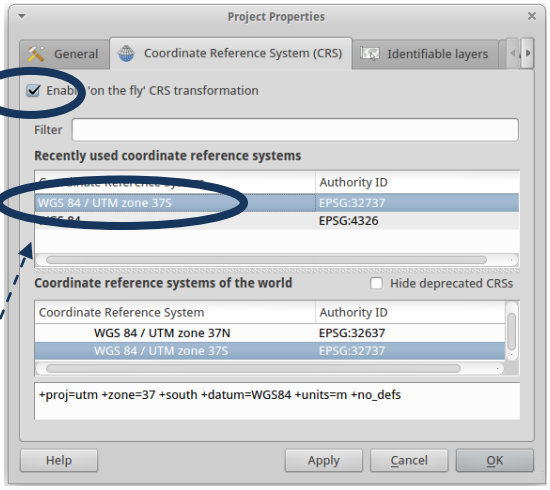
- In the options window Click on the **CRS** tab
- Click to tick **Enable 'on-the-fly' transformation by default** so that data layers with different CRSs can be displayed together automatically when added to the QGIS project
- Click to select **Prompt for CRS** so that QGIS will ask for the user for the CRS (projection) of the data when loading layers with no CRS defined
- Leave the default projection as **EPSG:4326-WGS84** (this geographic CRS will be the default for new projects)
- Click **OK** to close the Options window



*These options will be applied to all new QGIS projects but will not take effect until QGIS is closed and a new project is created. However, **IF** it is inconvenient to exit QGIS the settings for the current project can be set manually. Also there may be a requirement to set the current project to a particular CRS but not set it as the default for all other projects.*

1.2. Changing the CRS of the current project





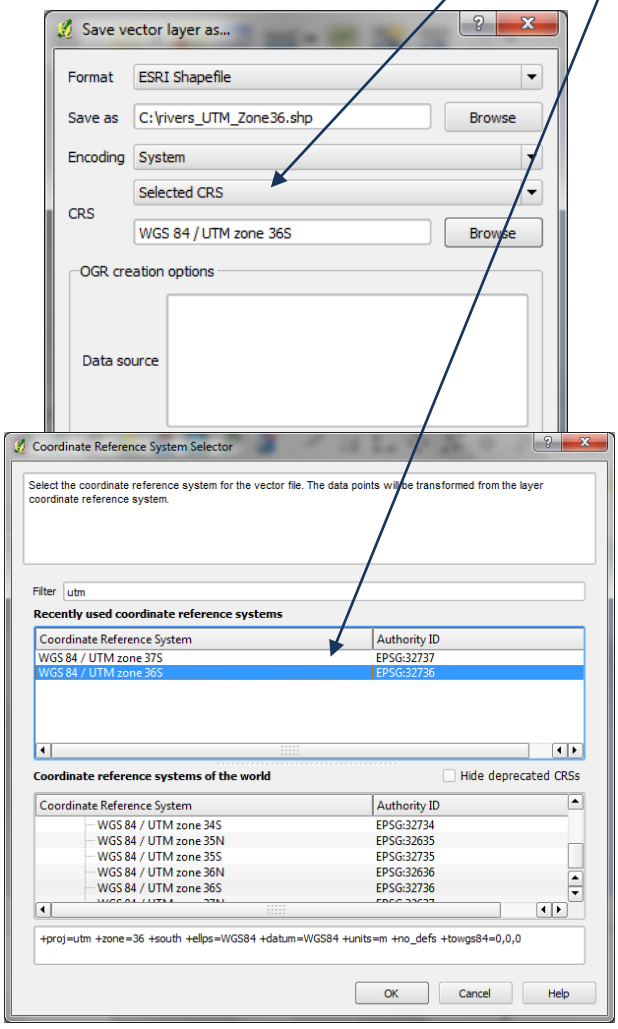
- To set the CRS for the current project Click on the **CRS icon** in the bottom right hand corner of the QGIS canvas
- Click to tick **Enable 'on-the-fly' CRS transformation** so that any data added to the QGIS project will be transformed on-the-fly from their current projection to UTM Zone 37S
- The same CRS window appears but this time it is for changing the **CRS of the <u>current</u> QGIS project ONLY.**
- The CRS was set to EPSG:4326. Change the CRS here sets the QGIS canvas to the **chosen projection**, in this case **UTM Zone 37S**.

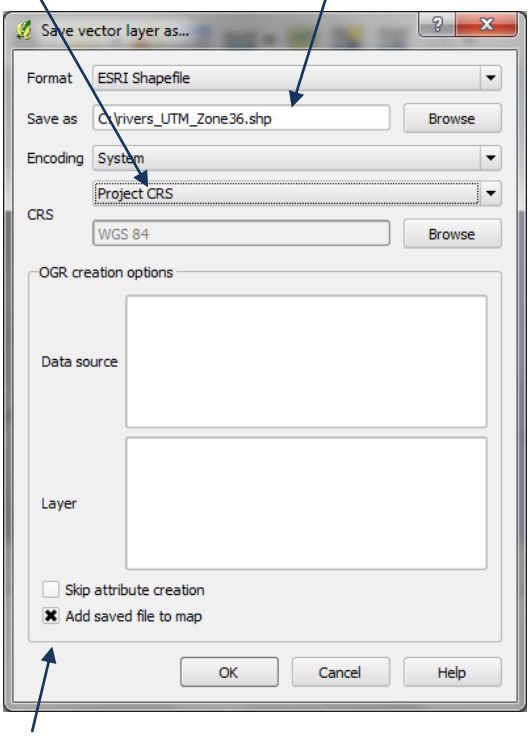
Note: 'on-the-fly transformation' does not physically change the projection of the data; it takes the data in its original projection and makes a temporary transformation as it is drawing it.

- IMPORTANT/ USEFUL Projection tips:
  - ***IMPORTANT***DO NOT **Right Click>>Set Layer CRS** on a dataset. This will NOT project the data to another projection. It will REDEFINE the projection only i.e. it will incorrectly tell the dataset it is something else. So ONLY use this to fix a dataset that has the WRONG projection defined
  - Unlike ArcGIS, **on-the-fly area calculations CANNOT be done in QGIS**, the data have to be actually saved in the projected coordinate system
  - If the required CRS cannot be found in the list, a custom CRS can be defined manually - under **setting>>custom CRS**
  - **Right click on a map layer>>Set Project CRS from Layer** to set the Project CRS (Map Canvas projection) to be the same as a particular layer
  - If the Map View is in a projected coordinate system (i.e. not EPSG: 4326) the scale of the canvas can be set at the bottom of the screen

1.3. Projecting data / saving data to a new CRS

- **Right Click** on the dataset to project. Click **Save as**
- Select the format to be saved to and give the file an appropriate name. It helps to include the projection in the filename i.e. in this case _UTM_Zone36
- Leave the encoding system as default
- Change the CRS method. Either to **Project CRS** to save the data in the same CRS as the Map Canvas or **Selected CRS** to pick the output CRS from the CRS Selector window.

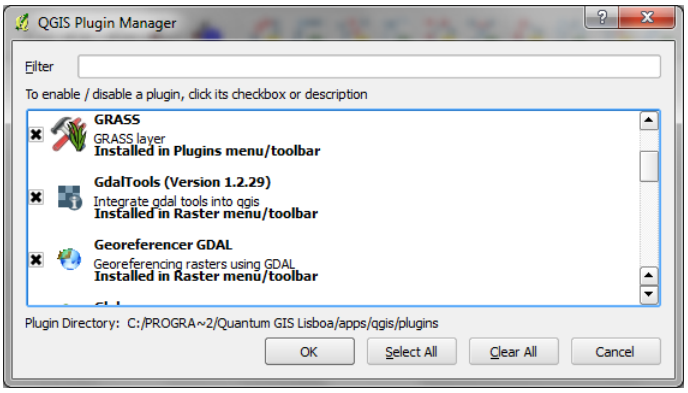




- Tick **Add** saved file to map to add the projected data to the current QGIS project
- Click **OK**

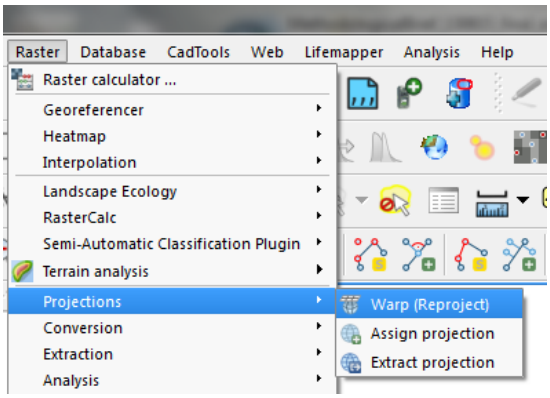
1.4. Reprojection of Raster layers

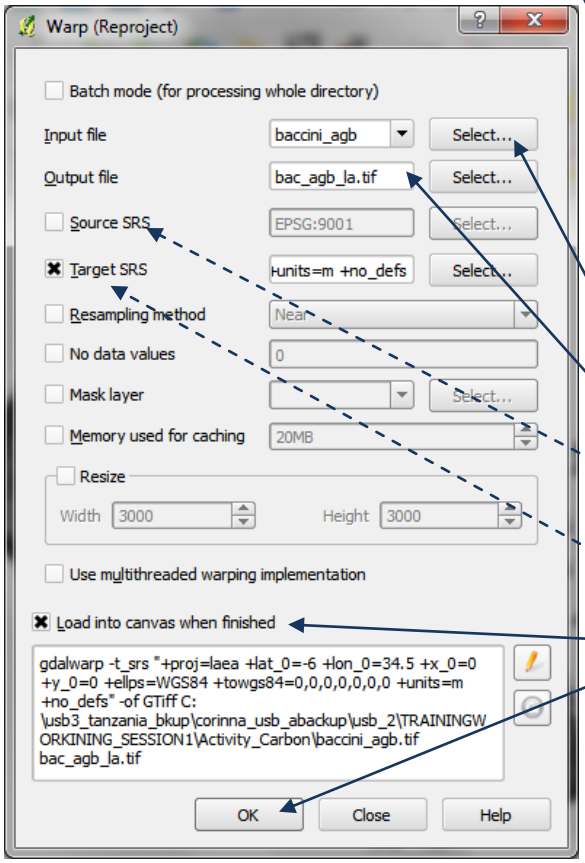
Reprojection of Raster data is done using the Warp utility of Gdaltools plugin



❖ Make sure that the **gdalTools plugin** is checked in the **QGIS Plugin Manager**

❖ From the main menu Click **Raster>>Projections>>Warp(Reproject)**





❖ Select the input raster dataset

❖ Select the location of output raster dataset and Give it a new name

❖ Click to Uncheck the **Source SRS** if the dataset knows what projection it is currently in

❖ Click to Check the **Target SRS** and select the new output projection

❖ Cick to Check **Load into canvas when finished**.

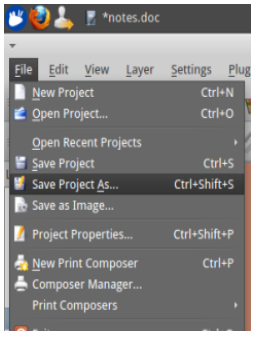
❖ Click **OK**

2. Saving a QGIS project

❖ Click **File >>Save project** as or **Save Project** from the main menu

❖ Navigate to a folder to save the project and give it a name

❖ Click **Save**

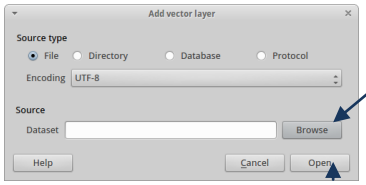


Note: The QGIS project is saved as a .qgs file. This is actually a text file containing all the information QGIS needs and is great for expert users who may wish to make edits in the text document. It also means that broken QGIS files have a good chance of being repaired as they can be opened in a text editor.

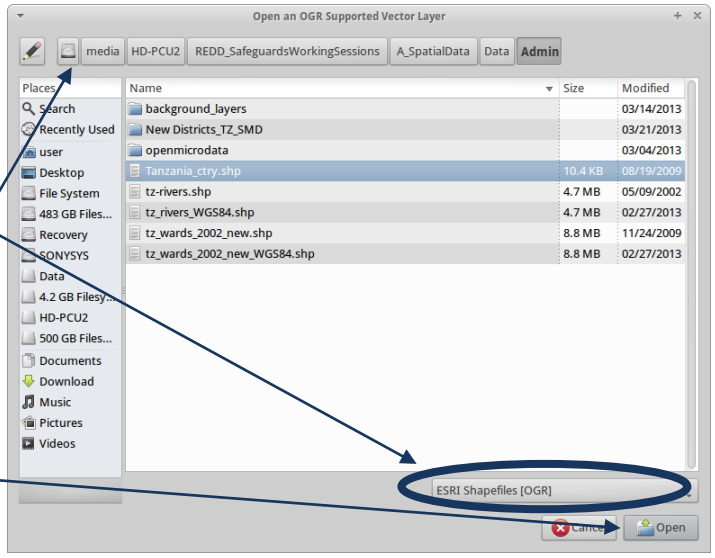
3. Adding new layers

3.1. Adding a vector layer

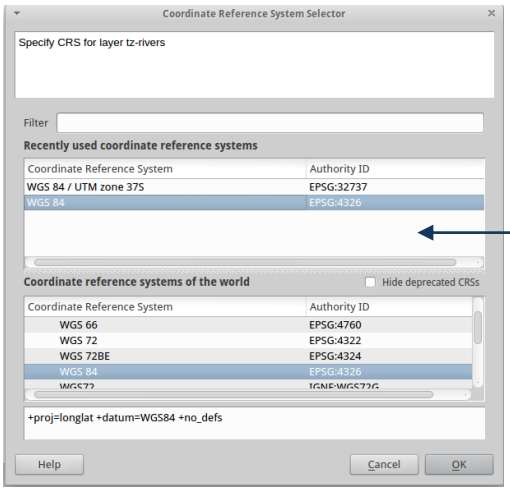
- Click either **Layer>>Add Vector Layer** or the **icon** to add vector layer to the current project



- Click **Browse**



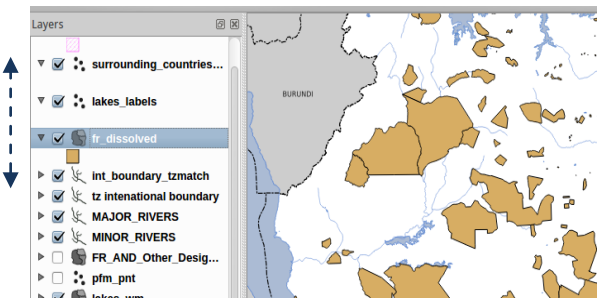
- Change the File type to **ESRI shapefile**
- Browse to the folder containing the shapfile to add i.e. in this example a country boundary file called **Tanzania_crty.shp**
- Click on the shapefile
- Click **Open**
- Click **Open**



The dataset is added to the table of contents **BUT** if the map projection information of the vector dataset is missing QGIS will ask the user to specify the CRS of the data.

- Choose the projection of the data (**NOT** what is presented here)
- If the projection is **unknown,** guess. If the data display correctly on top of other datasets for which the CRS is known then it is ok. Otherwise **right click** on the dataset and **Set Layer CRS** and try again

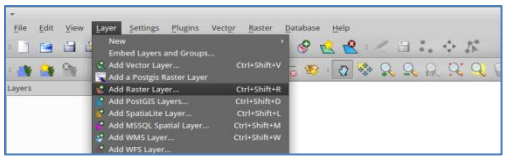
If QGIS is told by the user that the data are in CRS EPSG:4326 and the data are actually in a different projection then the data will appear in the map canvas but in the wrong place. When other datasets are added with correct CRS defined these data will not display in the same location. Remember, make sure 'the on-the-fly' projecting is ticked' or data in different CRSs will not display on top of each other.



- Layers can be **reordered** by clicking the layer in the TOC and **dragging to a new position** up or down the list

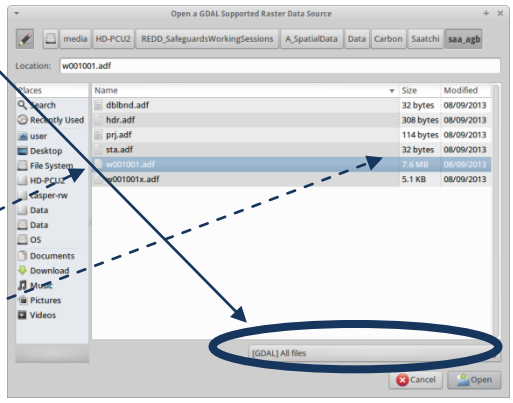
3.2. Adding a Raster Layer

- Click either **Layer>>Add Raster Layer**



or the **icon** to add a raster layer to the current project

- Leave file type as **GDAL [All files]** raster of all different types can then be added
- For data in ESRI native Grid format, this file format contains a folder with a number of standard files. The file to add is within the dataset name folder and will always be called **w001001.adf** and it will always be the largest file in the folder
- Click **Open**

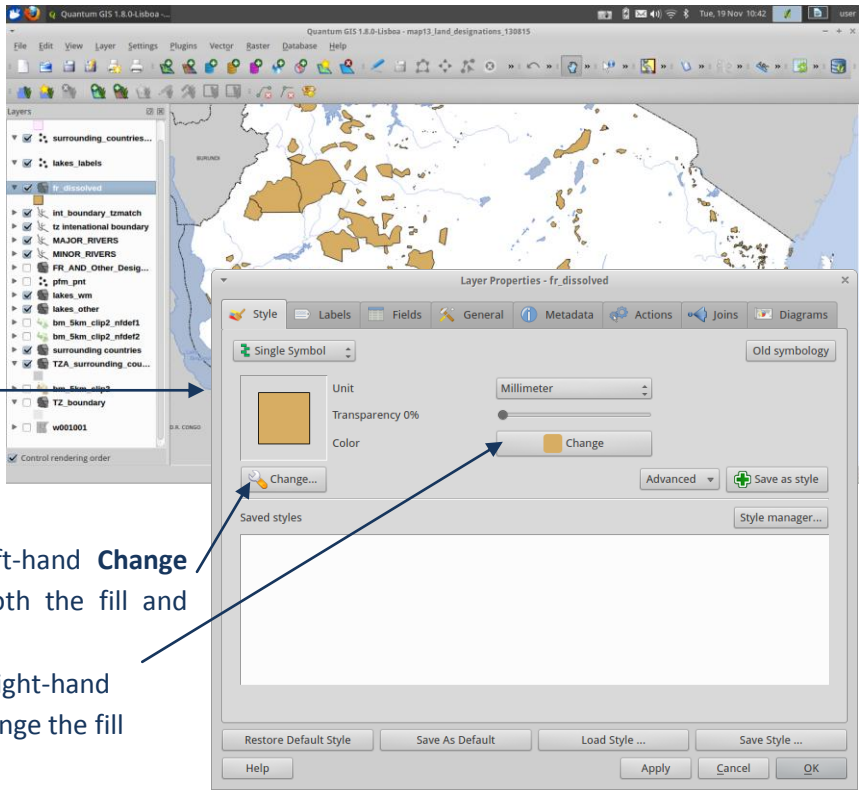


4. Changing symbology

4.1. Vector symbology

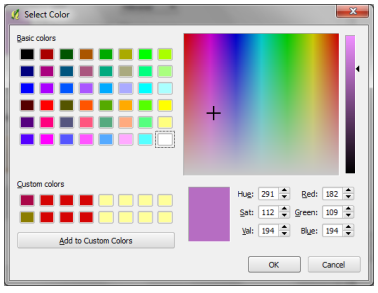
- When a dataset is added to QGIS it is added with a simple solid fill, random coloured symbol. **Double click** on the data layer in the Table of Contents to change the symbology

The layer properties window opens

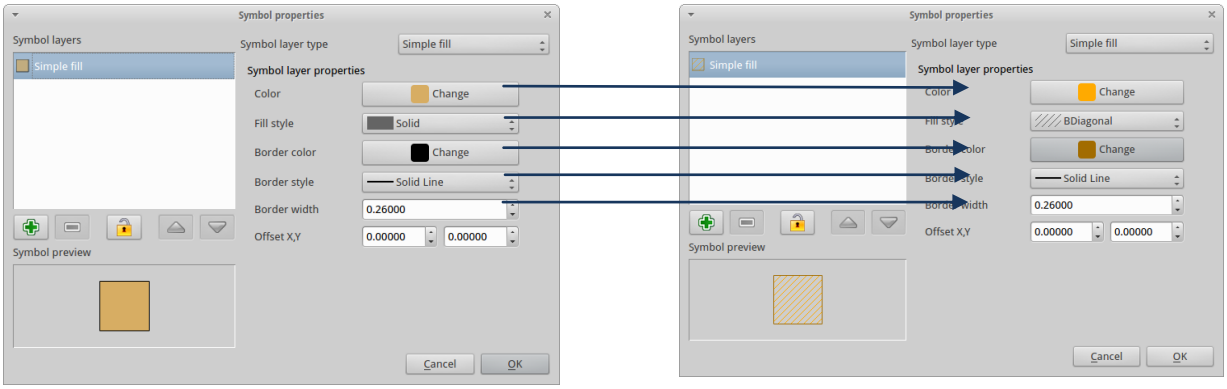


- **Double Click** the left-hand **Change** button to change both the fill and outline symbology
- Or **Double Click** the Right-hand **Change** button to change the fill color only

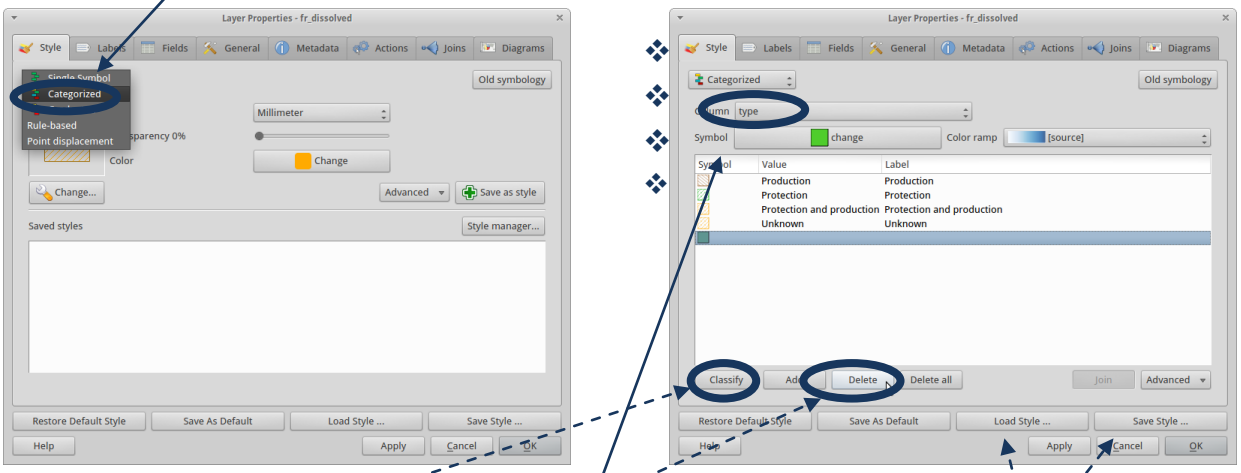
- The right-hand button opens the **Select Color** window
- The left-hand button opens the **Symbol properties** window



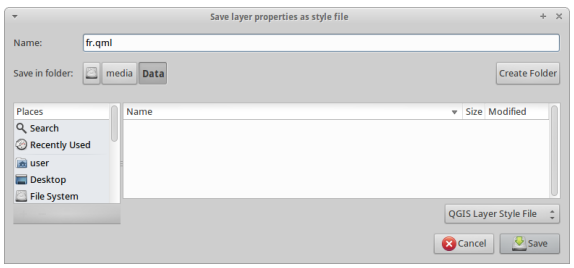
- Change the symbology as required:-



- Click **OK** to close the **Symbol properties** window
- Click **OK** to close the **Layer properties** window
- For categorical data e.g. to shade a layer based on a set of thematic values in the attribute table of a layer, Click the **Single Symbol** to drop down the options. Click **categorized**



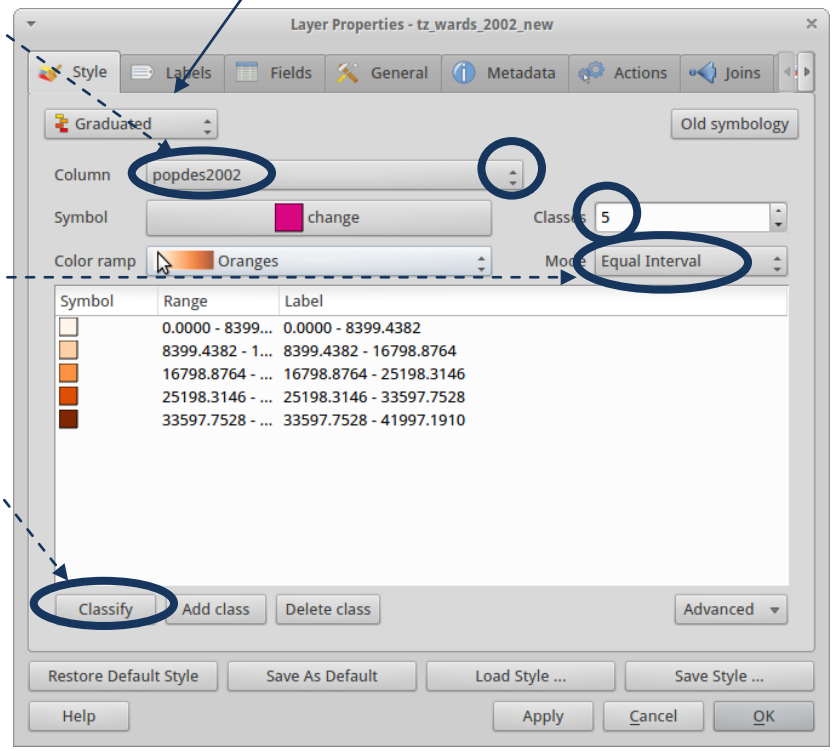
- Choose the attribute upon which to base the shading. E.g. in this example '**type**'
- Click the **Classify** button to add the unique combinations to the symbol window
- To remove any symbols (e.g. the empty one that is always added at the end), Click the symbol row, then Click **Delete**
- **Double click** on each symbol to bring up the **Symbol Selector Window** and change the symbology for each individual symbol in the same way as for the Single Symbol



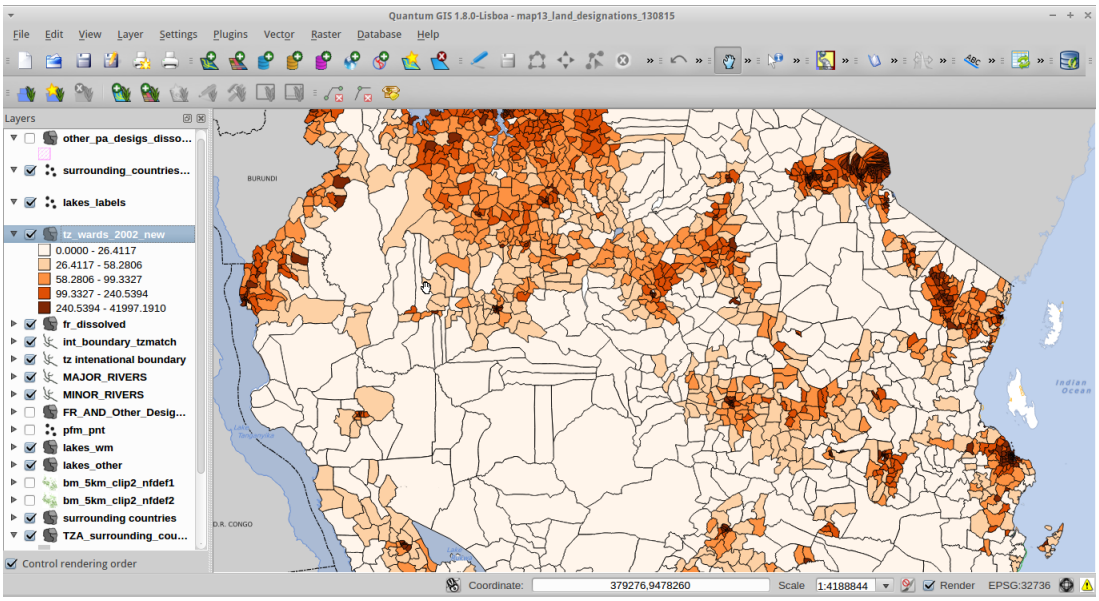
- Click **Save Style** once a set of symbols have been created. This will save the styles to a **QML** style file.

- The next time this layer is added to a QGIS project these styles can be loaded clicking the **Load Style** button

- Some data are best presented using a **graduated** symbology. E.g. population density data.
- Select the attribute from which to shade the data
- Choose the number of class breaks
- Select how the classes should be defined. E.g. **Equal Interval**
- Click **Classify**
- Chose a color ramp to shade the data
- Click **OK**

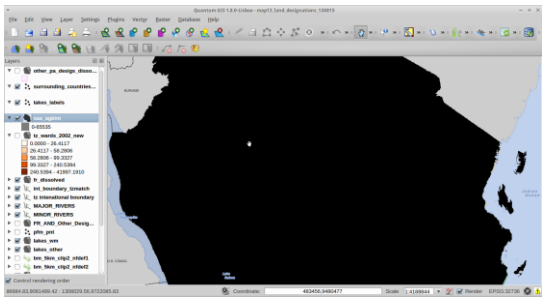


Example map below



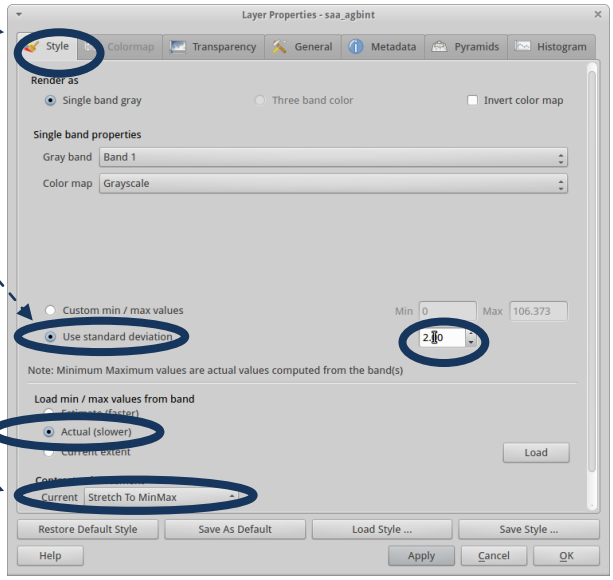
4.2. Raster symbology

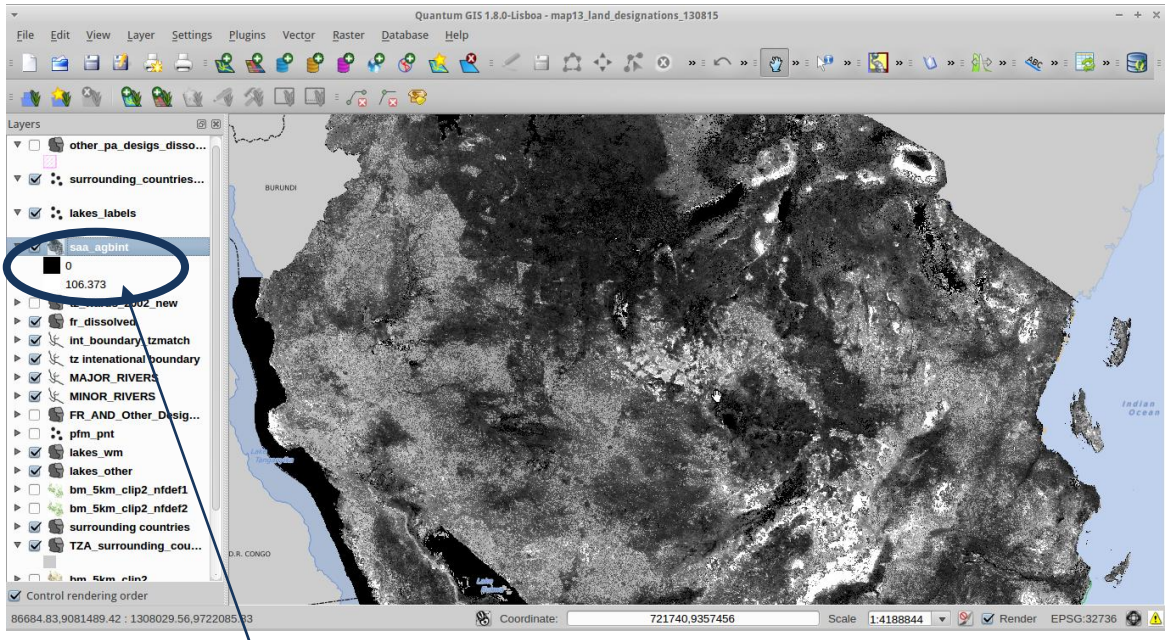
When a raster dataset is added to a QGIS project it often appears all in one color. This does not necessarily mean there is a problem with the data. It is more likely that the method of shading the data needs changing.



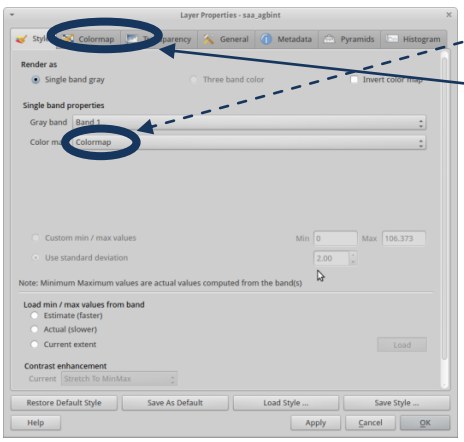
- **Double Click** on the raster dataset to bring up the **Layer Properties** window
- Click on the **Style** tab
- Click to **Use standard deviation**
- Leave the number of standard deviations as **2**
- Click on the **Actual (slower)** option in Load min/max values from band
- Click **Contrast enhancement** to **Stretch to MinMax**
- Click **OK**

The variation in the data can now be seen



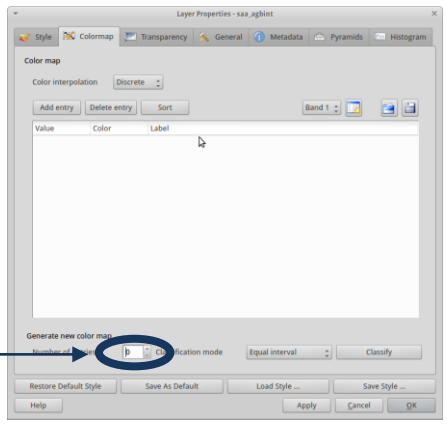


Note that in the Table of Contents the data are stretched from a low to a high value, using the above method of shading it is not possible to see class breaks.



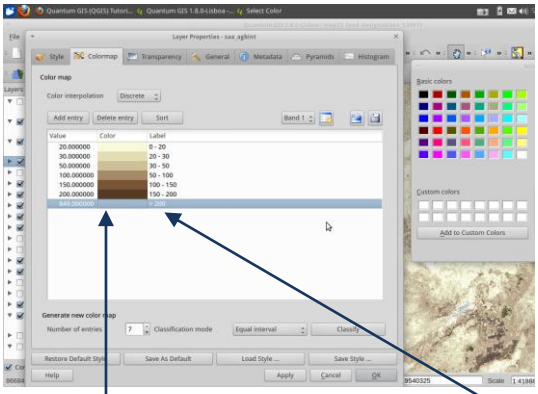
- Change the style to **Colormap**
- Click on the **colormap** tab

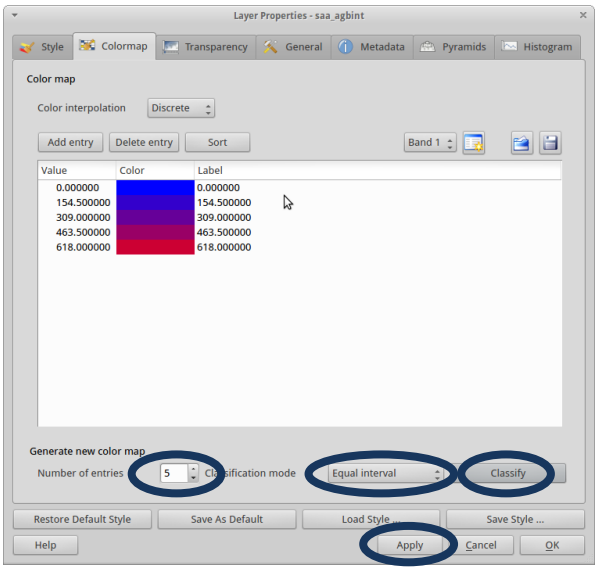
The layer properties window opens



- Click to change the number of entries (classes) to **5**.

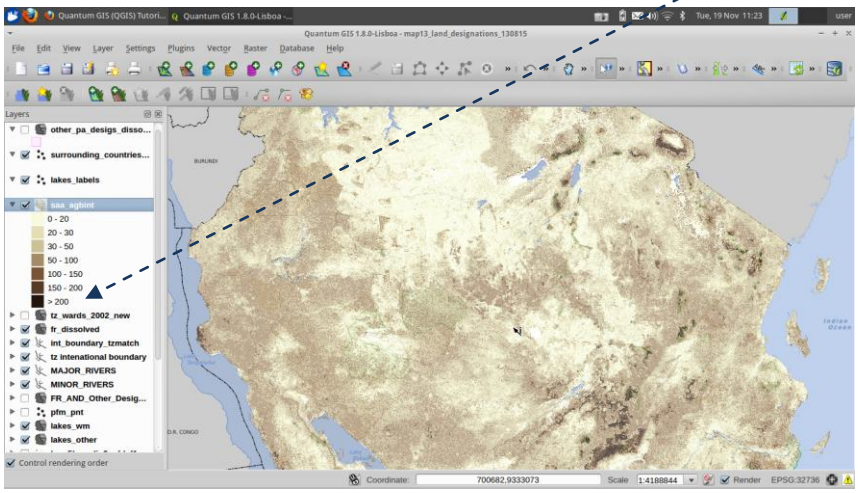
- Unfortunately there is only an option - **Equal interval** class breaks
- Click **Classify** to define the classes
- Click **Apply**





It may be necessary to define the class breaks manually if equal interval is not suitable.

- **Double Click** on each value and change it manually. **Change** the **label** for the map legend



5. Installing plugins

QGIS comes with an additional functionality in the form of 'plugins'. They are very easy to install, provided there is an internet connection to initially install them. Once installed they remain in the QGIS installation and an internet connecction is not required to run them.

- From the main menu click **Plugins>>Fetch Python Plugins**
- Click on the **Repositories tab** and tick **Show all plugins**, **even those marked as experimental** (as there are some useful plugins or updates to plugins that are still experimental)
- Check to see that the Official repository in your list of plugins by clicking on **Repositories tab**. The URL is http://plugins.QGIS.org/plugins/plugins.xml
- Click on the **Plugins tab**
- Select the desired plugin from the list and click **Install**
- The plugin's repository can also be explored from a web browser

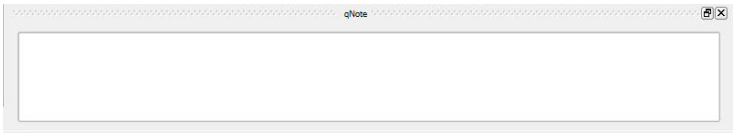
6. Useful plugins for spatial analysis to inform REDD+ planning and safeguards policies

Below is a list of a few plugins that have been particularly useful so far in carrying out spatial analysis work for REDD+ planning.

6.1. Qnote

Qnote is a great way of storing documentation about a QGIS project.

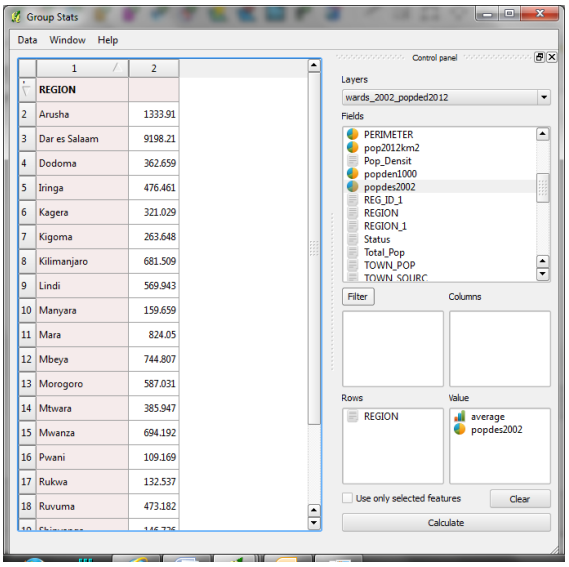
- The plugin adds an additional text window to the project in which notes can typed. It provides a method to attach metadata within the project so that it des not get lost. It is a free-text window in which information on the following might want to be stored:
  - Content of the project
  - Purpose
  - Analytical methodologies
  - Area of interest
  - Where the data came from
  - Who created the project
  - What the project was created for e.g. a report or publication
  - Version of the project / date last edited
  - Restrictions on sharing the project
- This information stored within a project is extremely valuable when sharing projects as well as providing an aid-memoire when revisiting a project at a later date
- Click on **Plugins>>manage python plugins**
- Scroll down and **tick Qnote**
- The Qnote window appears at the bottom of the QGIS project. The window can be turned on and off by Clicking **View>>Panels** menu



6.2. Group stats

Simple statistics can be calculated using the standard QGIS tools from the **Vector>>Analysis Tools>>Basic statistics** or **Vector>>Analysis Tools**>>**List Unique Values** menus. However, these are not sophisticated enough for summarizing the results of overlay analyses.

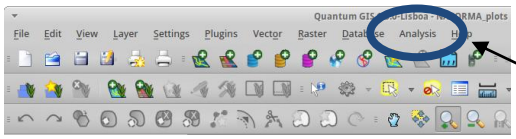
The Group Stats plugin is an advanced plugin that is useful for creating summary statistics about a dataset based on groups of features.



- Click on **Plugins>>manage python plugins**
- Scroll down and **tick Group Stats**
- Group Stats is accessed from the **Vector>>Group Stats** menu

6.3. SEXTANTE

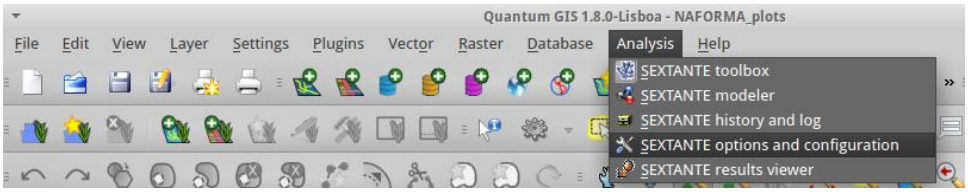
The plugin is essential for undertaking analyses with QGIS. It is a powerful plugin that integrates methods and tools from other open source software into the QGIS interface e.g. GRASS, SAGA and GDAL tools and R scripts. It provides a wealth of analysis tools that can be accessed from a single location. Tools can be run individually or through a GUI for processing work flows which is comparable to the ESRI ArcGIS Model builder environment.



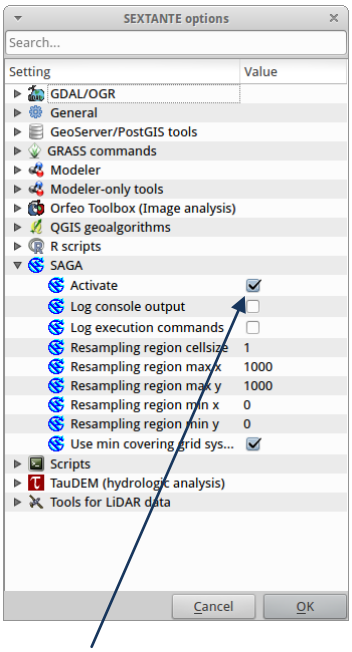
- Click on **Plugins>>manage python plugins**
- Scroll down and **tick Sextante**

There is now a new tab called **Analysis** in the QGIS menu bar

- Some of the SEXTANTE tools that are particularly useful are the SAGA GIS tools. SAGA GIS needs to be installed in order for QGIS to access the tools. If not already installed **download and install SAGA** from http://www.saga-gis.org/en/index.html
- Once installed SEXTANTE needs to be configured for the SAGA tools to become available. There are some slight differences in configuration for the Windows and Xubuntu (Linux) operating systems
- Click on **Analysis** from the main menu and go to **SEXTANTE options and configuration**

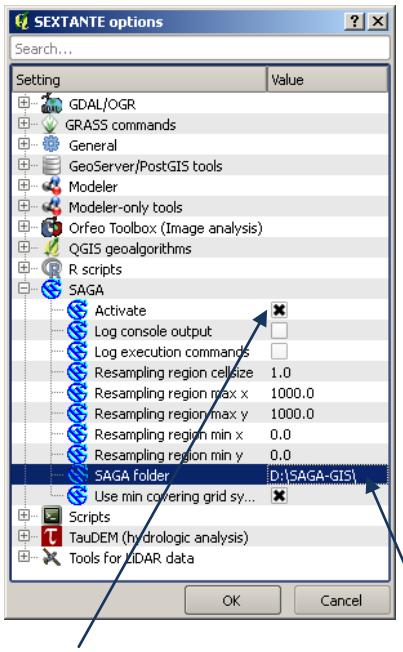


For Linux (Xubuntu) configure as below:-



Click to **Activate SAGA**

For Windows configure as below:-



Click to **Activate SAGA**

Specify the location of the SAGA installation

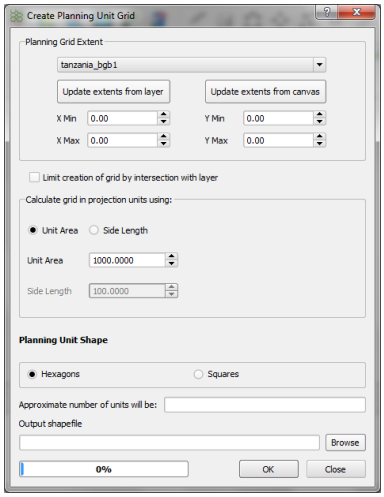
The Sextante extension forms part of the core geoprocessing in the later QGIS version 2.0.

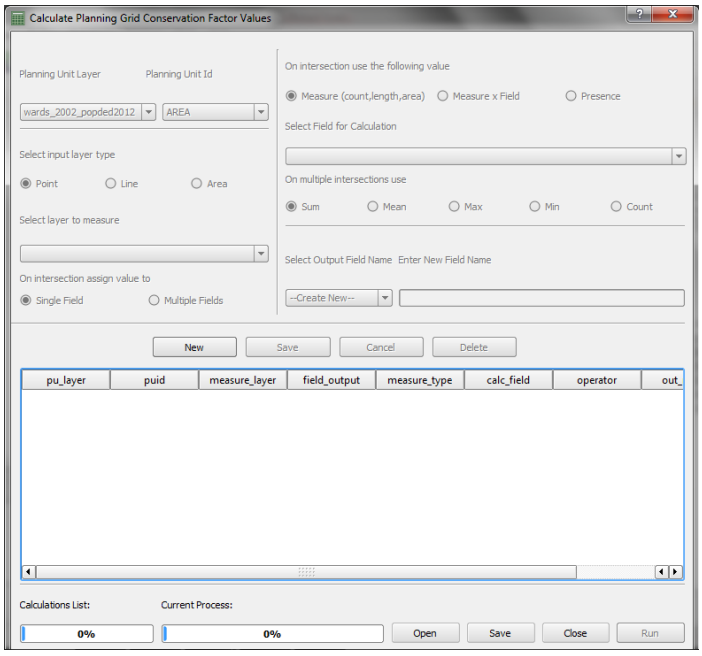
6.4. QMARXAN

Qmarxan is a set of simple to use tools for the creation of input files for the Marxan Spatial planning software from within QGIS. The tools have also proved useful for other analyses.

The two QMarxan tools that are particularly useful are:-

a) The creation of a regular grid of squares or hexagons using the **Create planning grid** tool



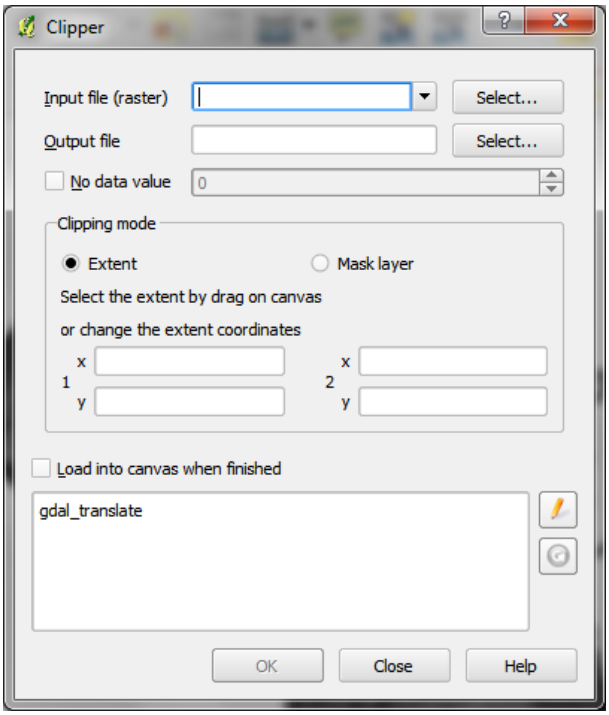


b) Using the **Calculate Conservation Values** tool to generate data that can be used to create a richness map e.g. species richness. This tool does an overlay analyses with the planning grid and a thematic dataset e.g. a dataset containing species extent of occurrence polygons.

Qmarxan currently only works with QGIS 1.8., although an updated plugin compatible for QGIS 2.0 is likely to be released in 2014.

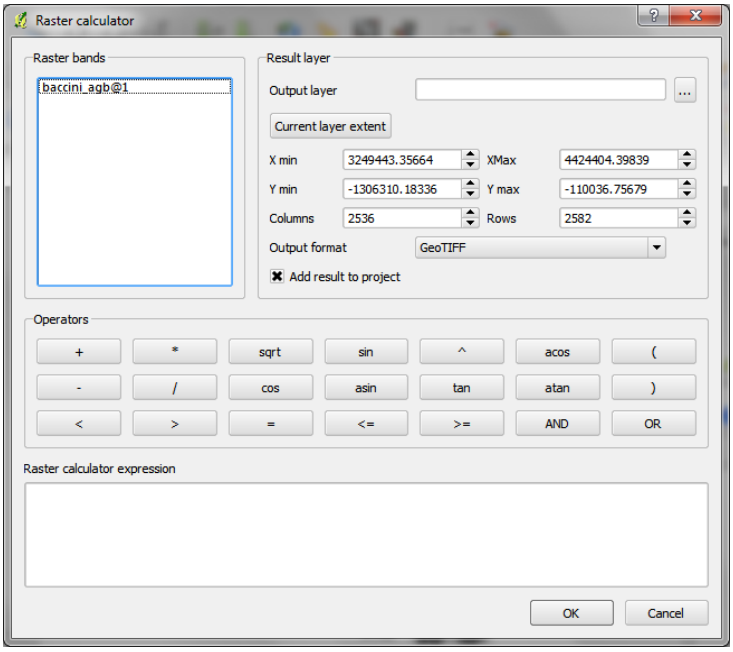
6.5. The Clipper tool

- The Clipper tool can be accessed from the **Raster>>Extraction>>Clipper** menu. It uses the gdal_translate utility to clip a raster to an area of interest based on an extent or mask layer



6.6. The Raster Calculator

The raster calculator allows you to perform mathematical operations on each cell in a raster.



E.g. multiplying values in one raster by the values in another or summing 2 raster datasets together.

- This can be useful for converting and manipulating your rasters.

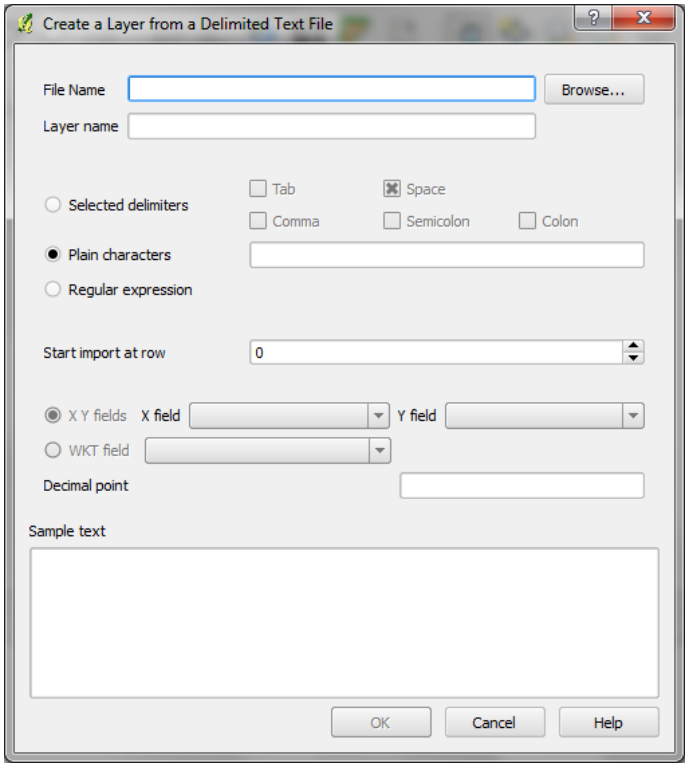
Operators include:

- Mathematical (+, -, *, /)
- Trigonometric (sin, cos, tan, asin, acos, atan)
- Comparison (<, >, =, <=, >=)
- Logical (AND, OR)

6.7. Add delimited text

The Delimited Text plugin allows a delimited text file to be uploaded as a layer in QGIS.

- The text file must be formatted with the first line of the text file a delimited header row of field names
- The data must contain an X and Y field (formatted as numeric but in any CR)
- The CRS of the XY coordinates must be known



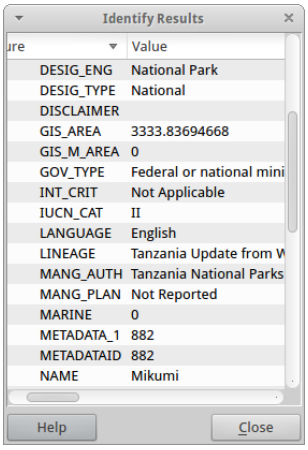
****IMPORTANT NOTE*** The 'delimited text' data in CSV format which means the data fields are often separated by commas. If Data within some of the fields also have commas this will case a problem with the import. Spreadsheet software such as LibreOffice Calc distinguish between the field delimiters and the commas within the text strings as the text strings are enclosed by " ". QGIS however does not see the " " and will read the data incorrectly. This can be resolved by opening the data in LibreOfficeCalc and saving the CSV file as tab-delimited rather than comma delimited.*

7. Querying data

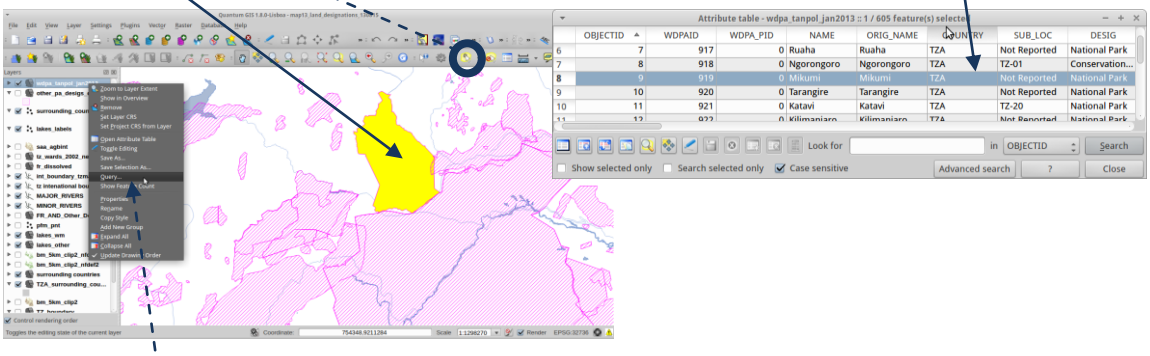
Data can be simply queried in the following 3 steps :

- First click on the **data layer** in the **TOC (table of contents)** to select the layer to query
- Click the **identify button**
- Then click on the **feature of interest** in the **map canvas** to identify the feature within the selected layer

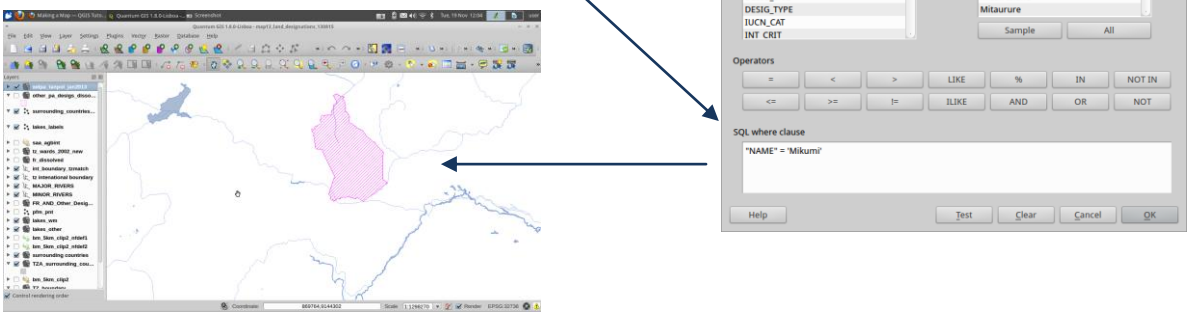
The Identify Results appears in a window showing the attributes for that feature record



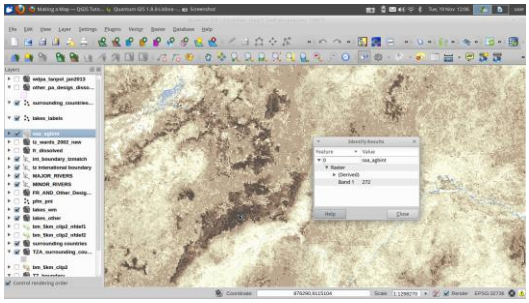
- Alternatively a feature can be selected by clicking on the **map canvas** with the **select single feature** tool and the feature is highlighted
- The attribute table can then be opened to see the **highlighted selected feature**



- If only certain polygons are required in the map display, a **Right Click>>Query** on the layer will bring up the **Query builder** window to allow the user to select features for display



- For Raster data there is **NO option to open an attribute table** as QGIS does not recognise them. The only option for rasters is to Click on the **raster data layer** in the **TOC** to select the layer to query, and then **click on the map** with the **identify button** to identify the values in the area of interest.



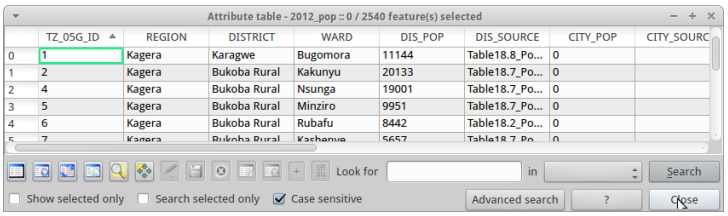
8. Joining Tables

- Joining of Tables is simple in QGIS. First **identify the datasets/tables to join**. Check there is a **common field** that can be used to join the tables together. The names of the join fields do not have to be the same – just have the same content

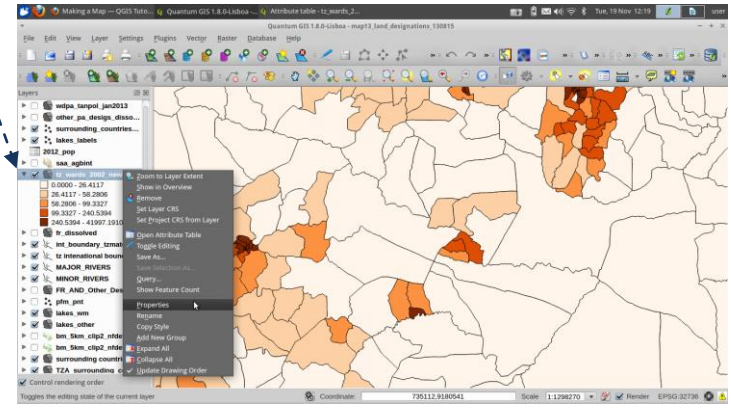
- Open the attribute tables of the dataset to be joined to explore the fields e.g. in this example a table containing population densities by ward needs to be joined to a shapefile of ward boundaries

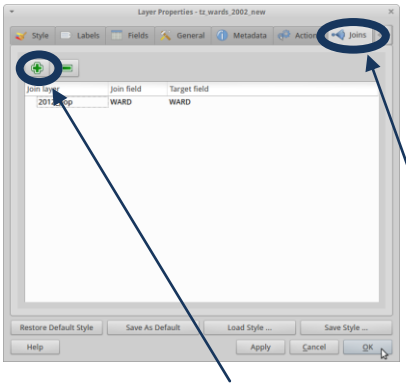
- Note the names of the field to be used in the join.

- Close the attribute tables



- **Right Click>>Properties** on the data layer to **join to** e.g in this example the ward boundaries

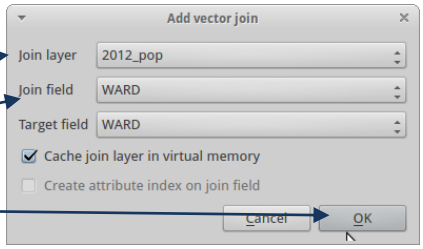


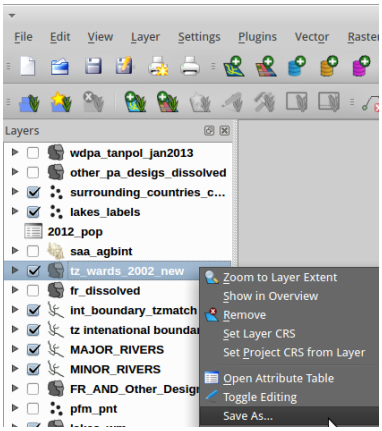


- In the **layer properties** window click on the **Joins tab**

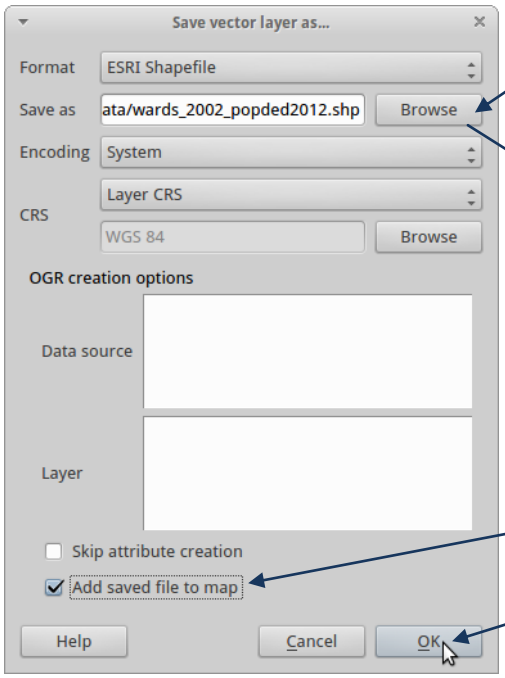
- Click on the **'+' button** to add the join

- Select the **table to join** e.g. 2012_pop table
- The **target field** and the **join fields** in this case have the same name in both datasets
- Click **OK**

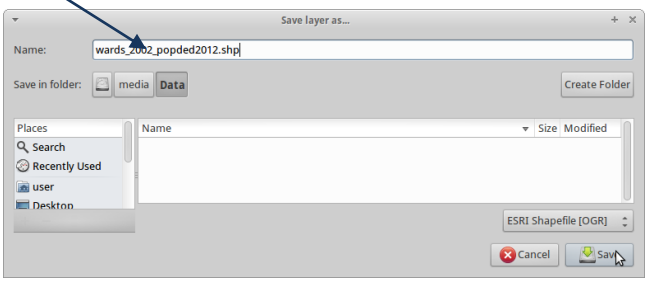




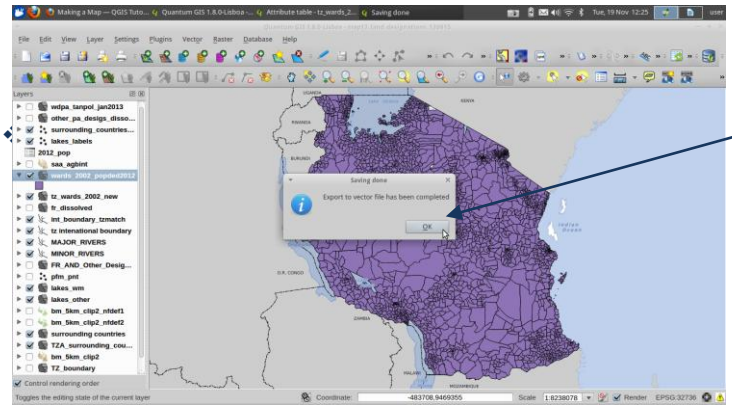
The table has been **temporarily joined** onto the shapefile. To make the join permanent, **Right Click** on the shapefile and click **Save As…**



- Click **Browse** to save the shapefile with a new name
- Keep the **CRS** the **same as the existing layer**

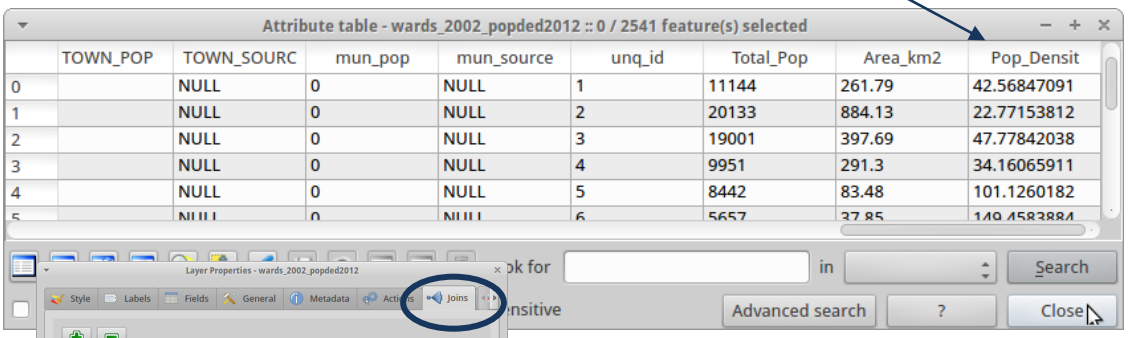


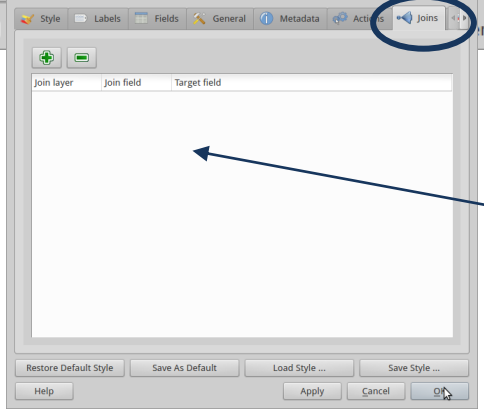
- Tick to **Add saved file to map**
- Click **OK**



- The new dataset is added to the map. Click **OK**

- Open the Attribute table
- See the fields from the population table at the end of the table

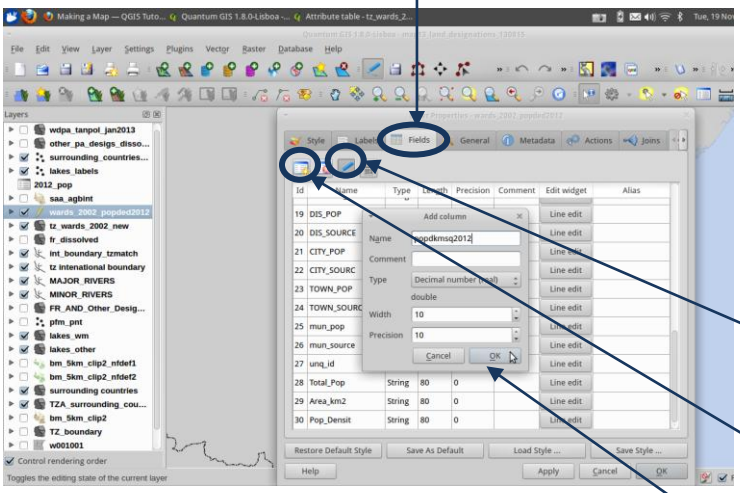




- **Double Click** on the saved dataset to open the layer properties

- Click on the **Joins tab** to see that there are no longer any joins present and therefore the population fields have been made permanent

❖ Click on the **Fields tab**



Note that all the fields joined from the CSV file are type 'string'. Some may need to be numeric but unfortunately the field types cannot be changed. It is necessary to add a new field and calculate the information across from the string field to the numeric field.

❖ Click the **toggle editing** button to start editing

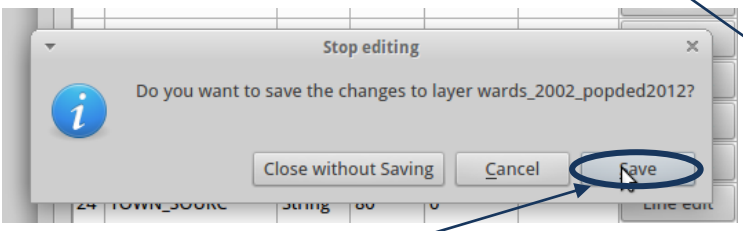
❖ Click the **Add Field** Button

❖ **Add a new field** of type **Decimal number (real)**

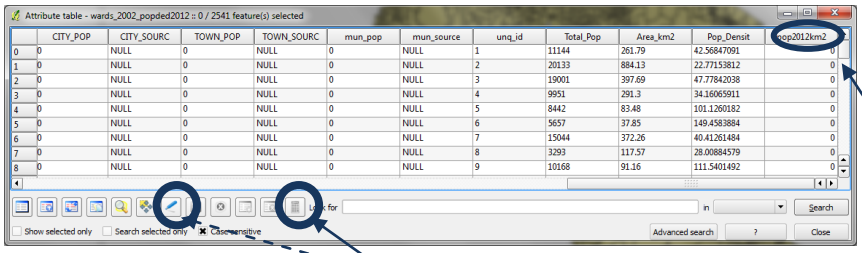
❖ Give the field a name

❖ Set Width and Precision to **10**

❖ Click the **toggle editing** button to stop editing

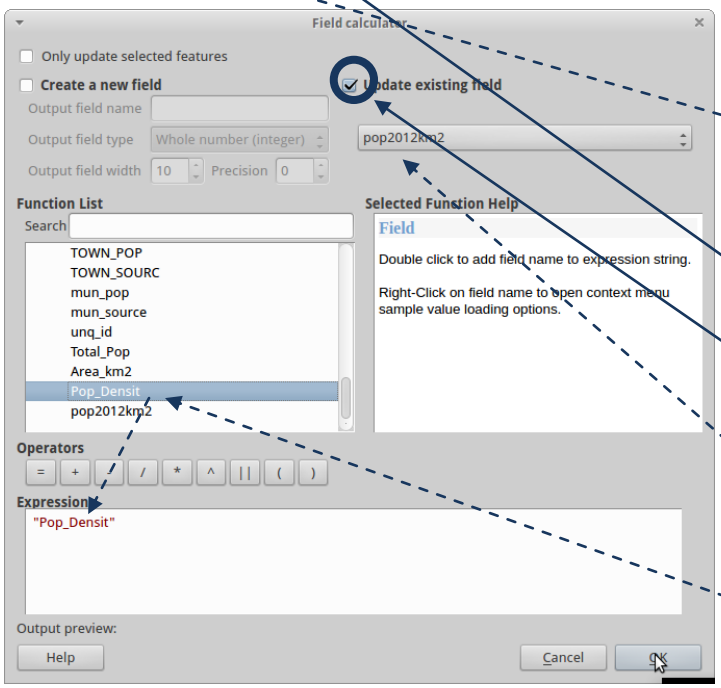


❖ Click **Save**



❖ **Right Click** on the data layer and **open the attribute table**

❖ The new 'Empty' field is located at the end of the table



❖ Click the **toggle editing** button to start editing

❖ Click the **calculator** button

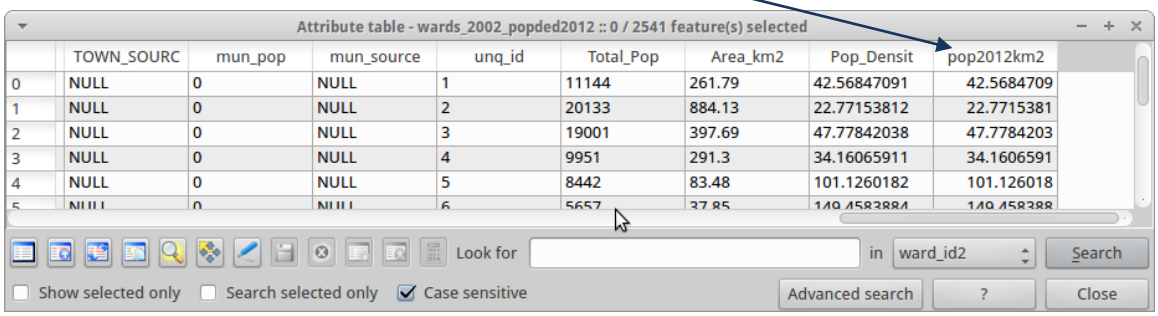
❖ In the **Field calculator** tick the **Update existing field**

❖ Choose to update the new empty field created in previous step

❖ **Double Click** on the **(string) field** which contains the **data to transfer** to the decimal field. This will bring it into the **expression** window

❖ Click **OK**

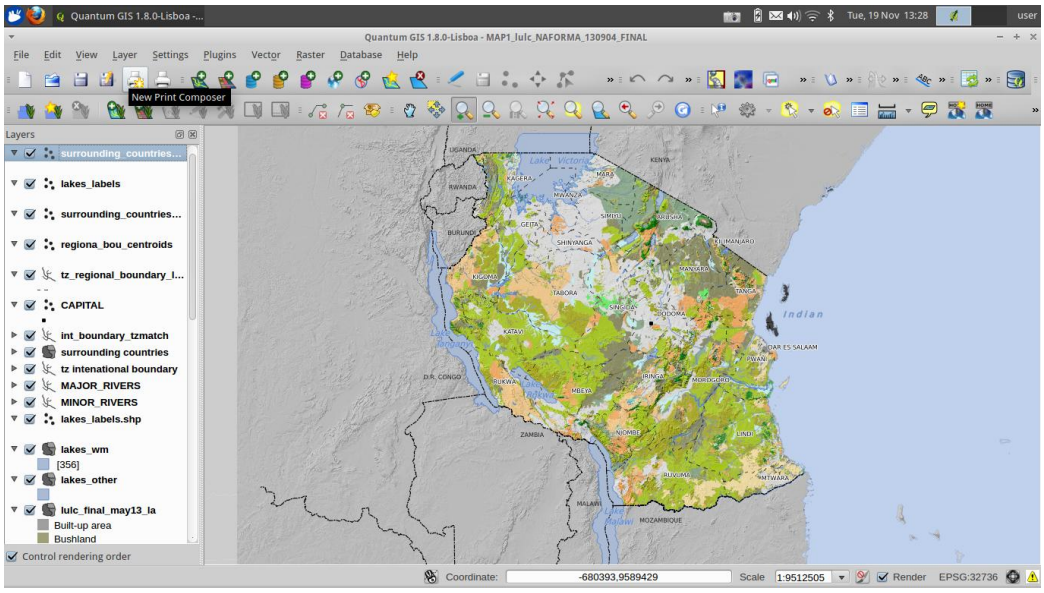
- Click the **toggle editing** button to stop editing
- Check that the attribute has been updated. It is sometimes necessary to close and reopen the attribute table

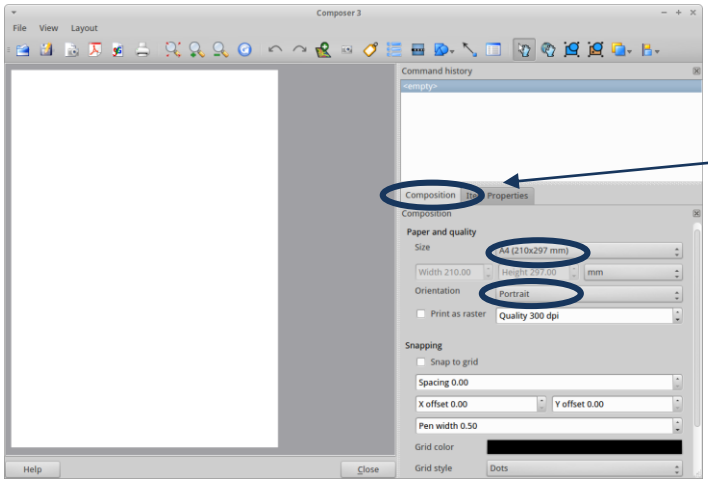


9. Map Layouts

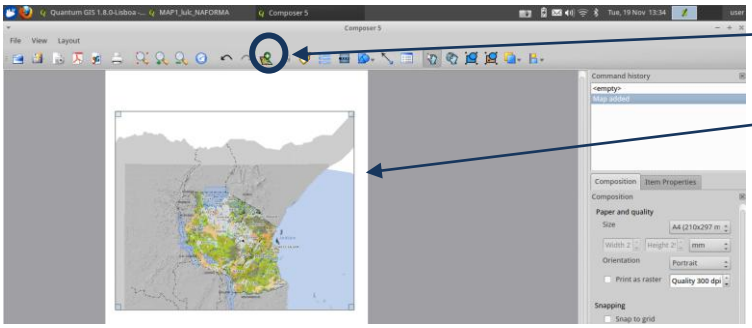
Once all the layers have been symbolised a map composition can be created. However the layers appear in the map view they will appear in the map layout so it is important to choose effective colors and symbols. The Layers can either be renamed in the Table of Contents or later when adding the legend to the map composition.

Note : In QGIS 1.8 Map layouts need to be in a projected coordinate system so the project projection needs to be set accordingly before creating the map layout. The reason for this is that the scalebar does not work correctly in a geographic CRS. This has been resolved in QGIS 2.0.

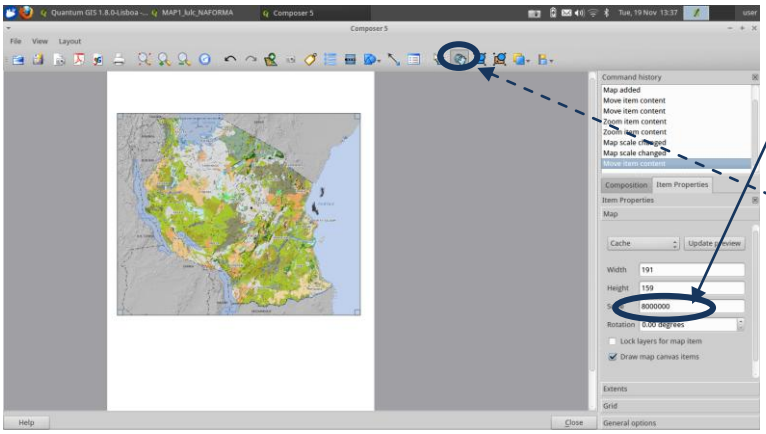




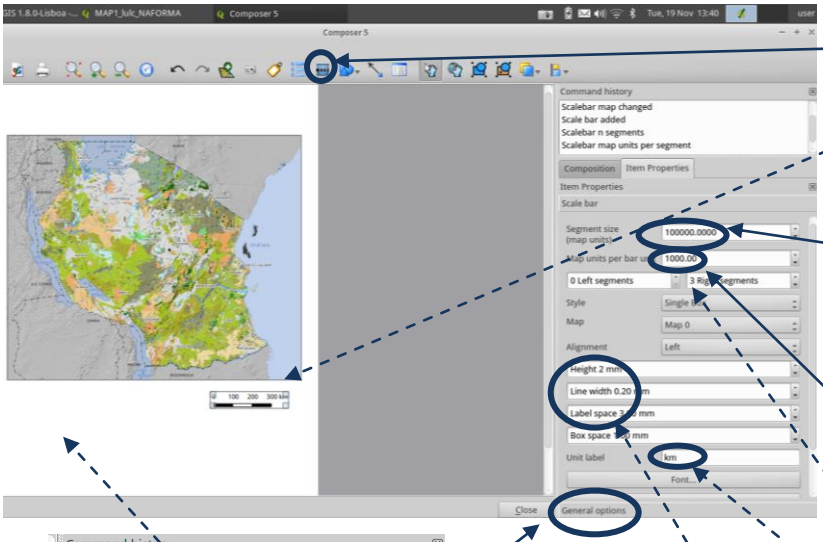
- Click on the **New Print Composer** button. An empty composer window is opened
- Click on the **Composition tab**
- Choose the **page size** and **orientation** of the map layout



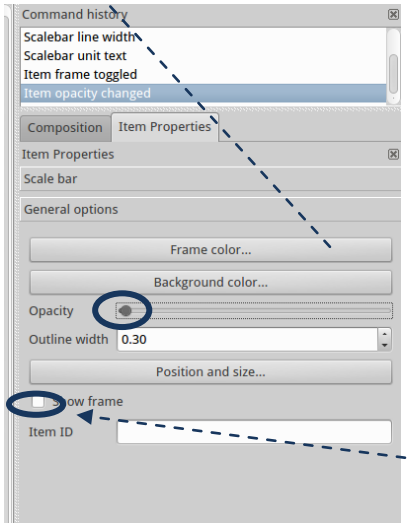
- Click the **Add new map** button
- **Drag a box** onto the map Canvas using the mouse



- Change the **Map scale** to a more appropriate scale for the composition *(a higher number will zoom out and a lower number will zoom in)*
- Click the **Move item content** button to pan the map to the desired extent



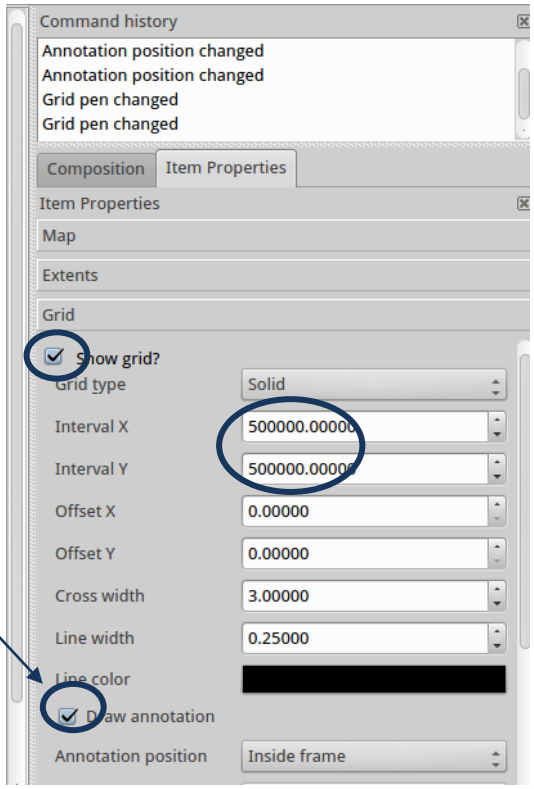
- Click the **Scalebar** button
- Then **click on the map layout** to position the scalebar
- Change the **scalebar segment size** (the data units are in meters so the size is set in **meters**)
- Change the **map units per bar** to be **1000** (to convert from meters to kilometers)



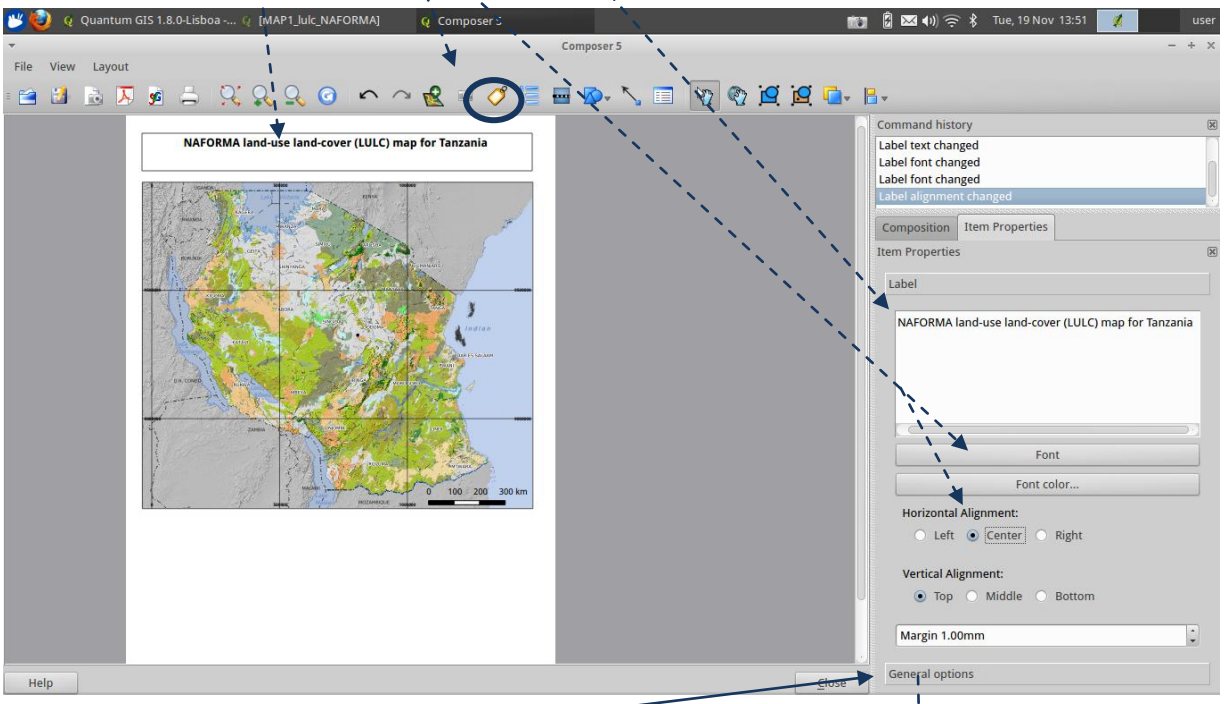
- Change segments to **3** (or 5 so the scalebar stars and ends with a black segment)
- Change Height to **2** and Line width to **0.2**
- **Add a label** to show the units of the scalebar in **km**
- Set the **font and size** of the **scalebar text**. (For an A4 map chose an appropriate font size of **6 or 7** (no more than 8)
- Then click the **General Options**
- Set **Opacity** to be at the **far left**
- Untick **Show Frame**

The scale bar should look small and neat as on the map above.

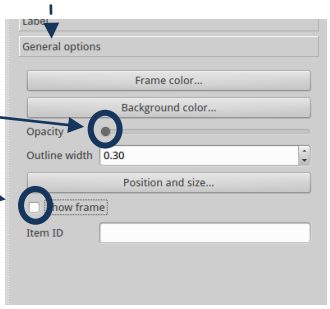
- Next **add a graticule** to the map. Click on the **Select/Move map items** button
- Click on **the map** that was dragged layout
- Click on the **Grid tab** to expand the options
- Tick **Show grid**
- Choose the **interval for the grid lines** (this will be a grid in the same units as the map projection)
- Set the **Interval X** and **Interval Y**
- Tick **Draw annotation**
- Chose **Annotation position** (inside or outside depending on preference



- Click the **Add new label** button
- Click on the map to **position the label**
- Change the **label text for a Title**
- Change the font
- Change the alignement

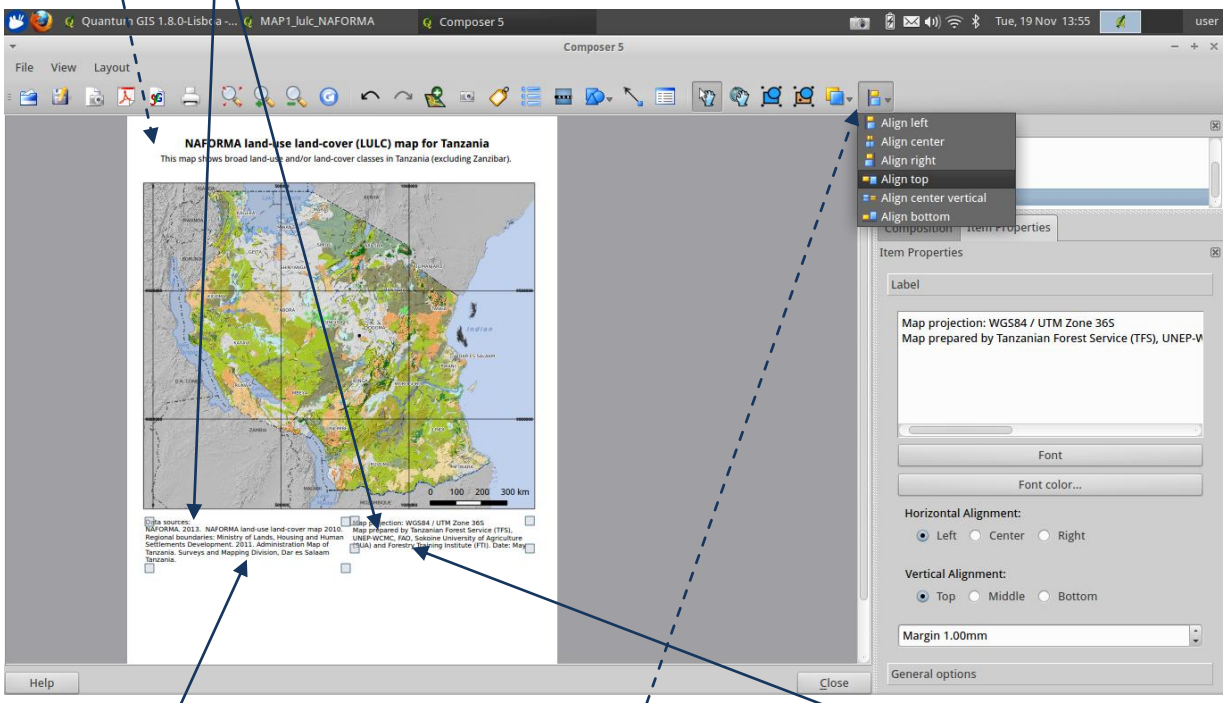


- Click on **General options**
- Set **Opacity** to **0** (i.e. far left)
- Untick **Show frame**

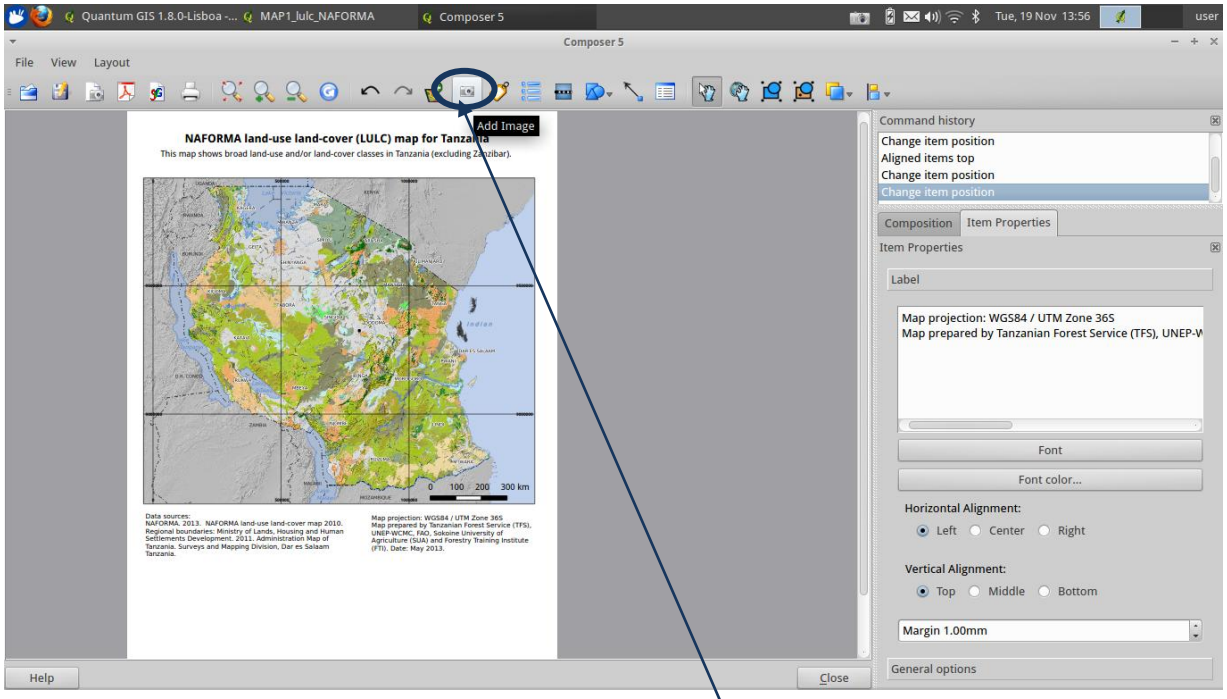


- Repeat the above to add a **subtitle / map caption** with a smaller font size
- Add smaller text for **data sources used** on the map and **map projection, creator date** etc.

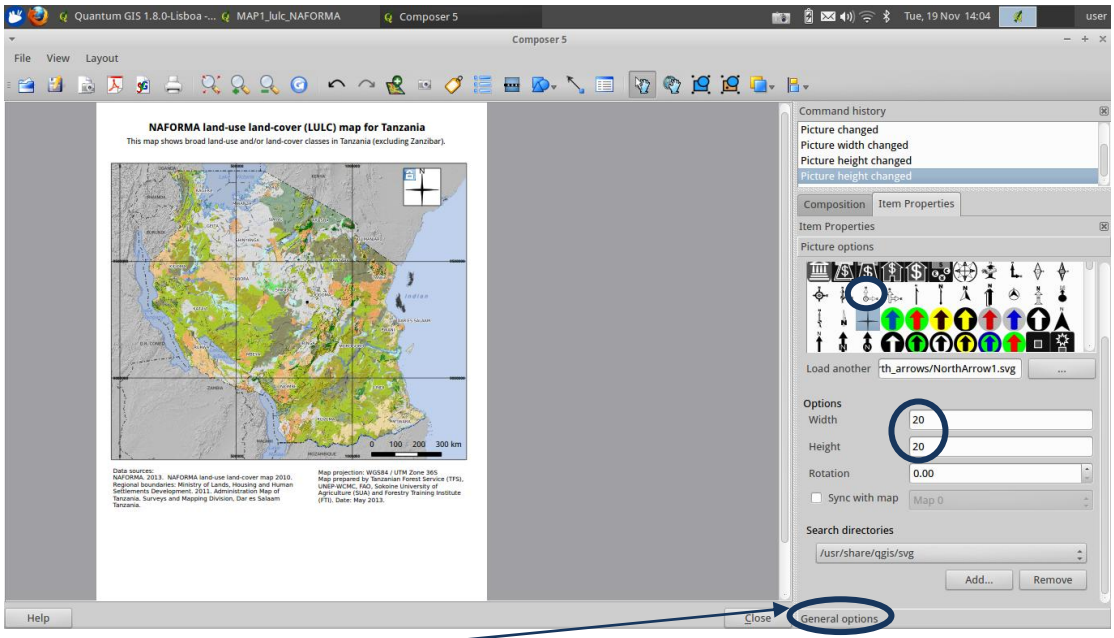
Remember to choose appropriate fonts and sizes



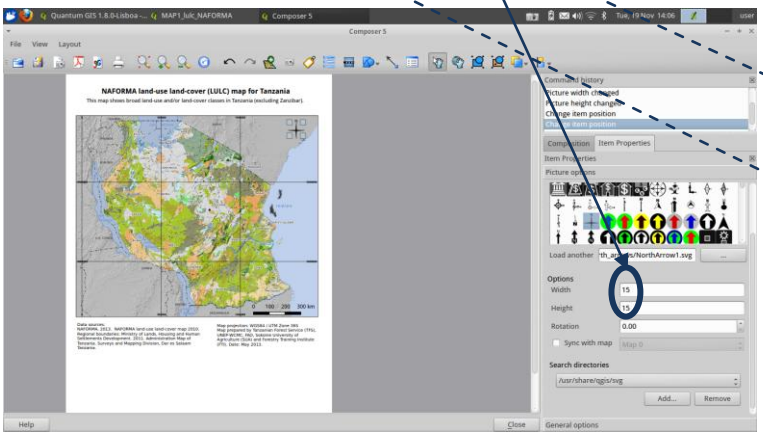
- Click on one of the text boxes and **hold down shift** and select the second one
- From the menu the text boxes can then be **aligned**

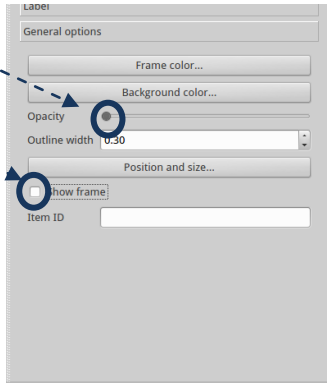


- A North Arrow is added as an image. Click on the **Add image button**
- Click on symbol to add a North Arrow of choice.
- Set the **height** and **width**

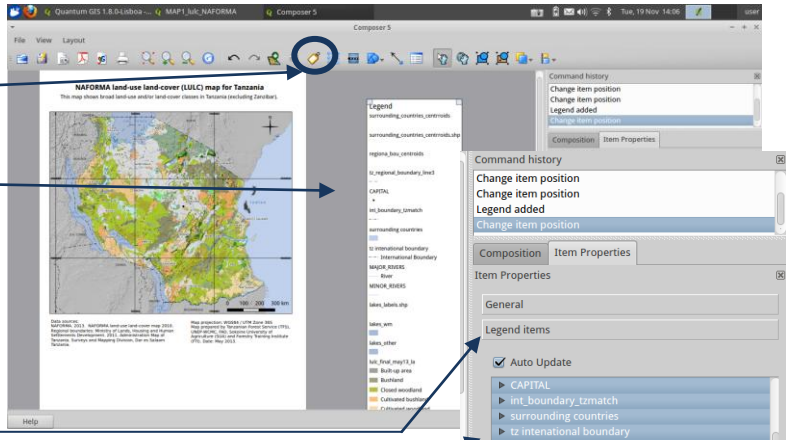


- Click on **General options**
- Set **opacity** to **0** (i.e. far left)
- Untick **Show frame**
- Adjust the size to make the north arrow smaller





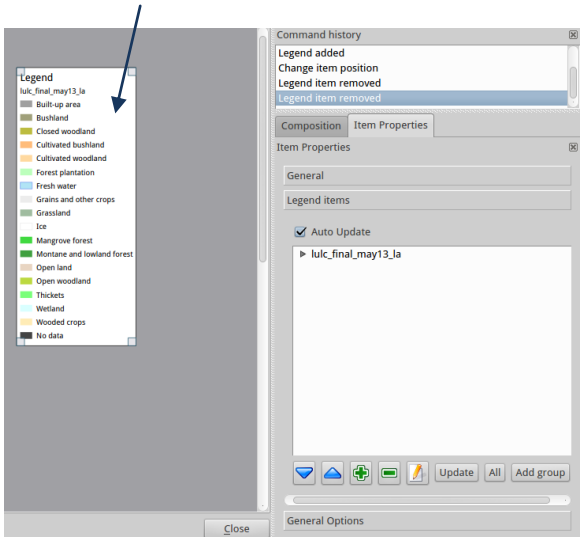
- Click on the **legend** button and **click on map** composition to add to map. Every item in the data view is added to the legend



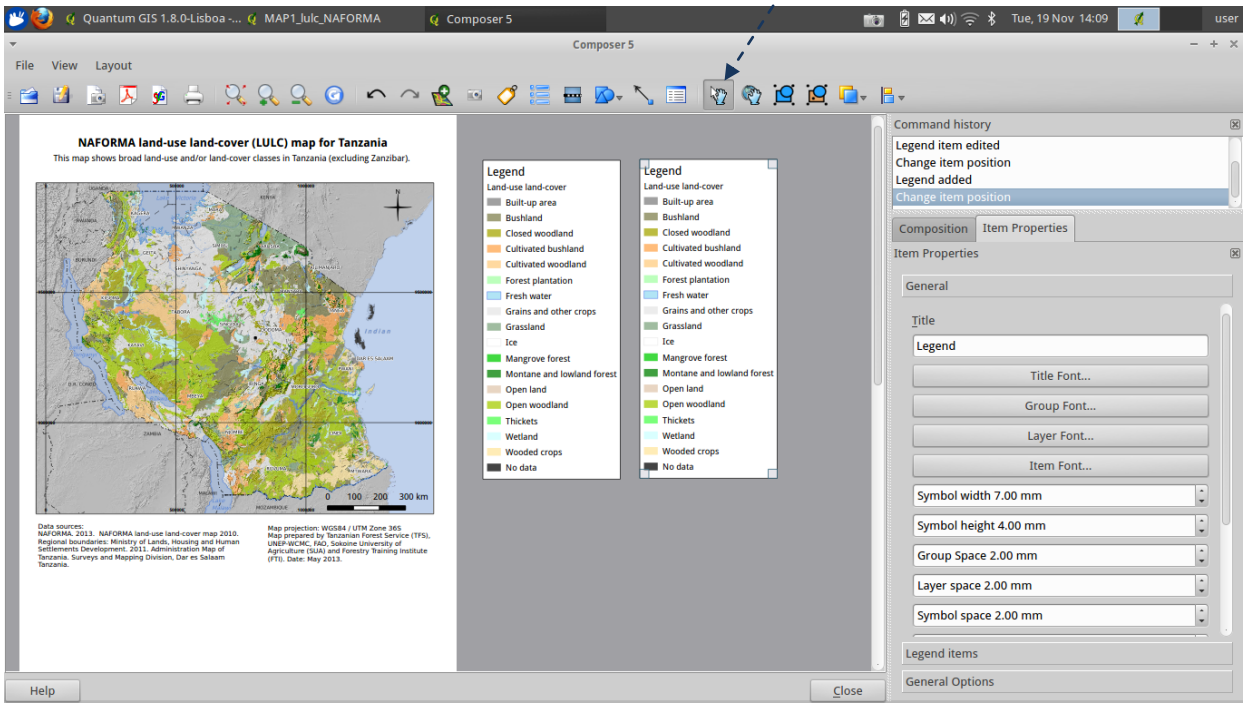
- Click on **Legend items tab**
- **Expand** the list of items
- Hold down the **shift key** and **click on layers to be removed from legend**
- Click the **'-'** button to remove them

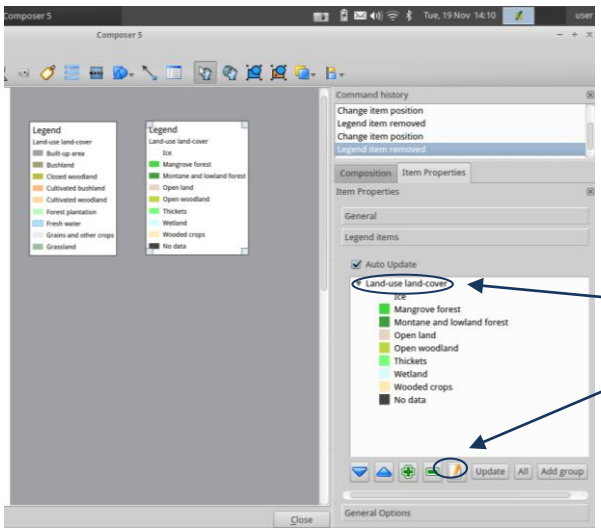
In this example all items other than the land-use land-cover have been removed.

The legend should now look smaller and with fewer layers, depending on what has been removed.

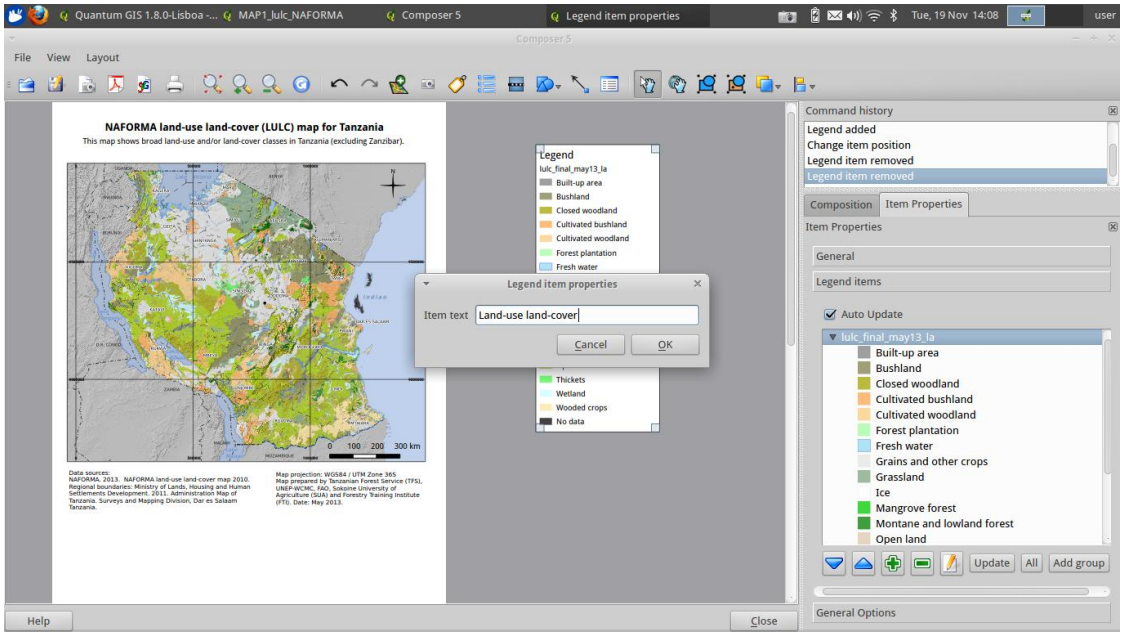


- Click on the **Select/Move** button to select the legend
- Use Ctrl-C to copy the legend and Ctrl- V to paste a duplicate

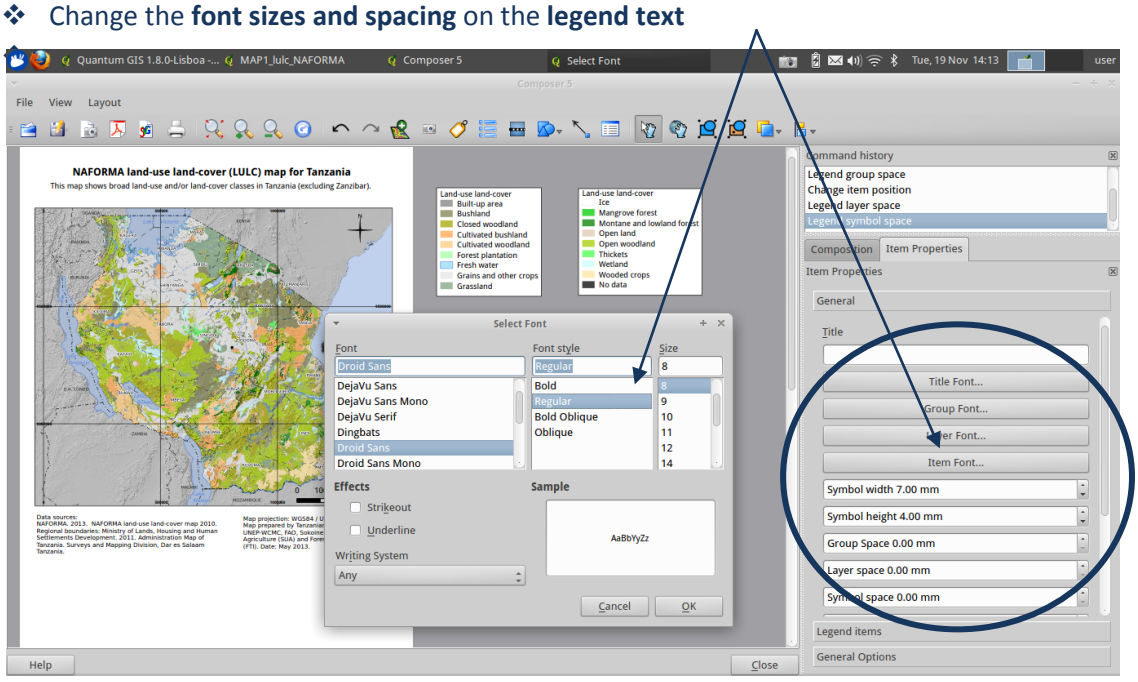


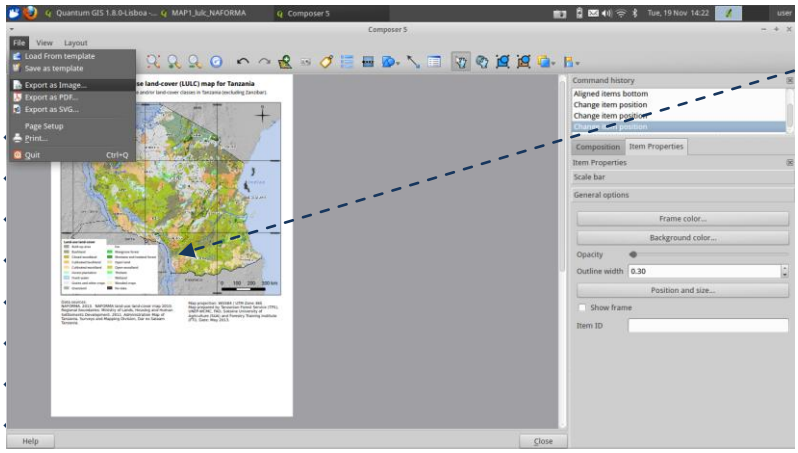


- For each legend, **highlight values to remove**. In the Left hand legend remove all classes beyond Grassland. In the right hand legend remove all classes up to and including Grassland. (Click the **'-' button to remove**)
- Click on the **layername** and click the **pencil button** to **edit or remove the layer name text**



❖ Change the **font sizes and spacing** on the **legend text**





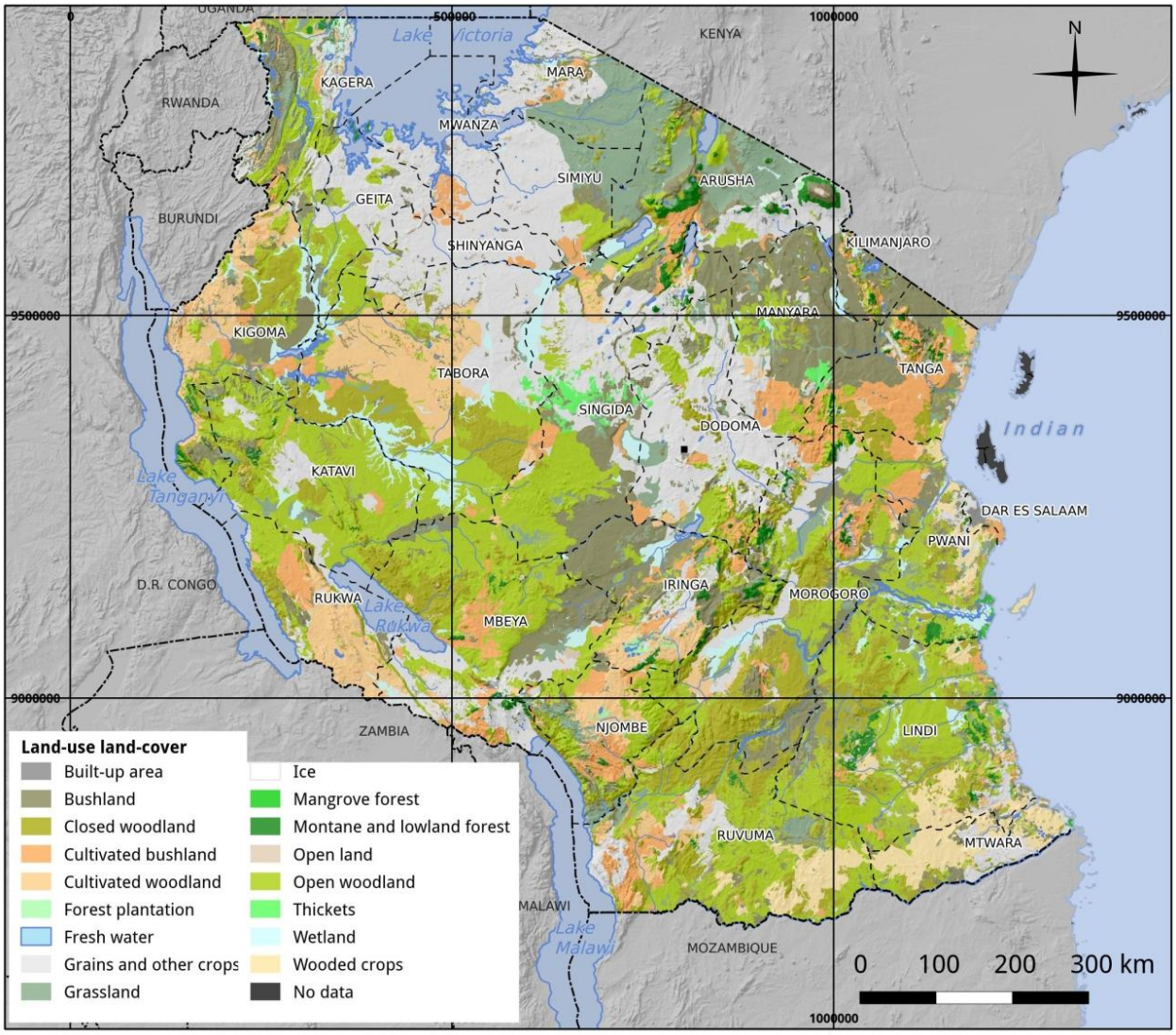
❖ **Position the legend** onto the map canvas

❖ Export the map **image for publication** to a common image format such as jpeg or tif file. Click **File>>Export as Image**

Example of the final map layout :-

NAFORMA land-use land-cover (LULC) map for Tanzania

This map shows broad land-use and/or land-cover classes in Tanzania (excluding Zanzibar).

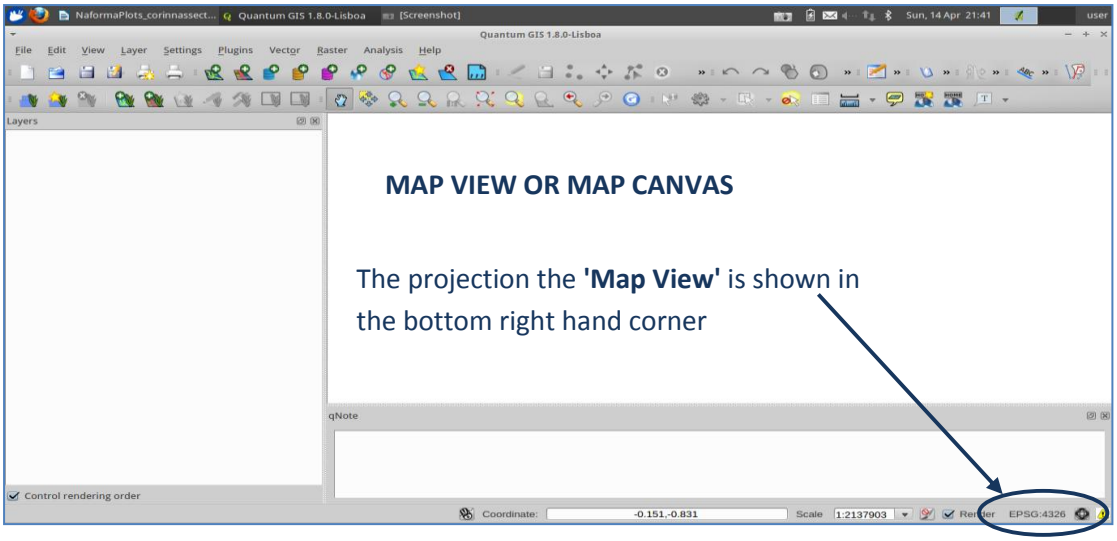


Data sources:
NAFORMA. 2013. NAFORMA land-use land-cover map 2010.
Regional boundaries: Ministry of Lands, Housing and Human Settlements Development. 2011. Administration Map of Tanzania. Surveys and Mapping Division, Dar es Salaam Tanzania.

Map projection: WGS84 / UTM Zone 36S
Map prepared by Tanzanian Forest Service (TFS), UNEP-WCMC, FAO, Sokoine University of Agriculture (SUA) and Forestry Training Institute (FTI). Date: May 2013.

10. Comparing area statistics generated from different projections

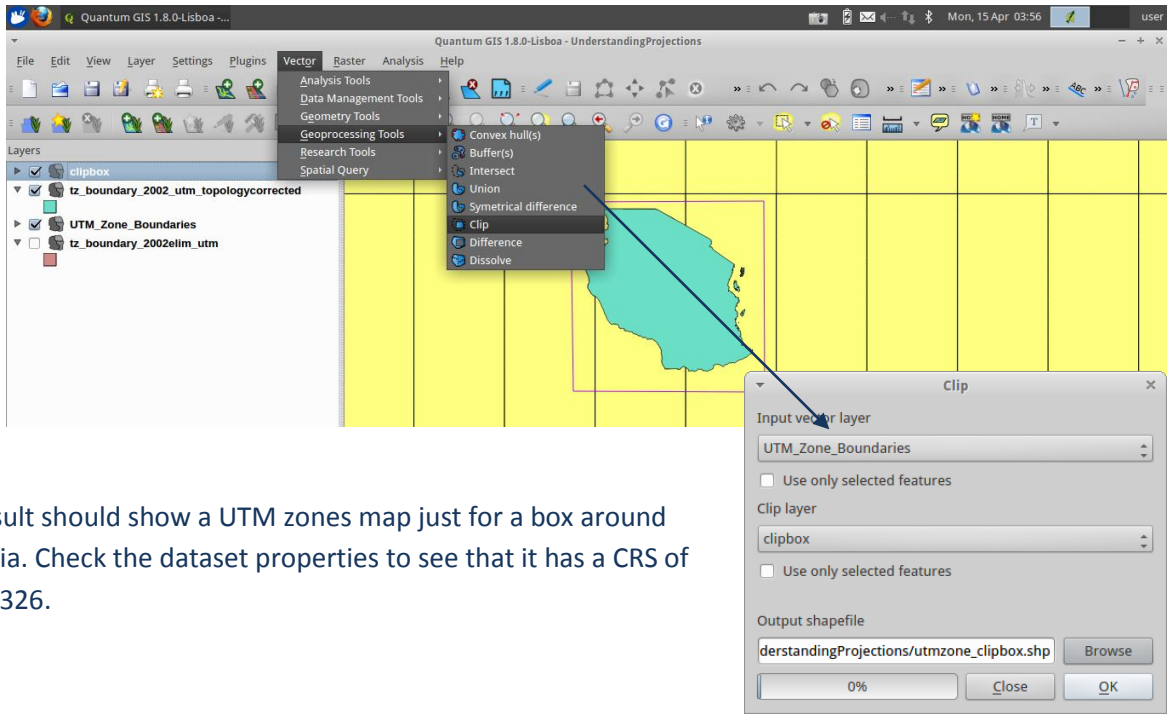
- **Change the projection** of the QGIS Map View for the **current project only** to be geographic WGS84 (EPSG:4326)



- Activate **on-the-fly projecting**
- Add the following layers to the Table of Contents (TOC) (in the working session found in the Activity_UnderstandingProjections):
    - A shapefile containing a polygon of the bounding box of Tanzania e.g. clipbox.shp
    - A shapefile of global UTM Zones e.g. utm_zone_boundaries.shp
    - A shapefile of Tanzania country boundary e.g. tz_boundary_2002_wsg84.shp

Please note the screen grabs are illustrative only and the names do match the above

- **Clip** the **global UTM boundaries** i.e. utm_zone_boundaries.shp **to** the **bounding box of Tanzania** i.e. clipbox.shp

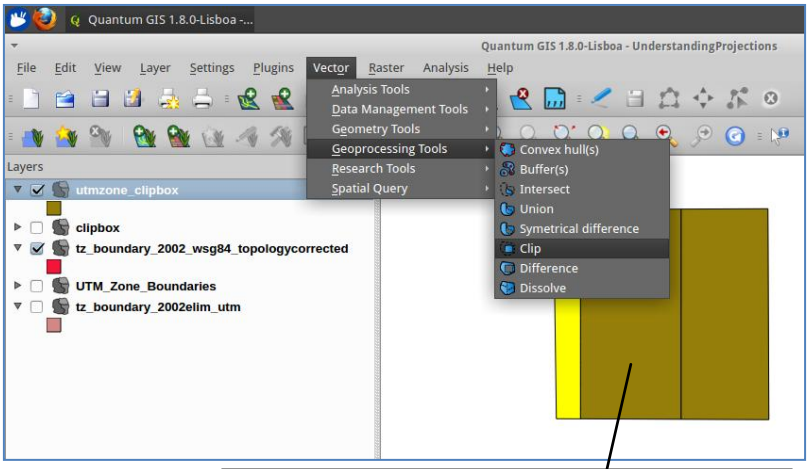


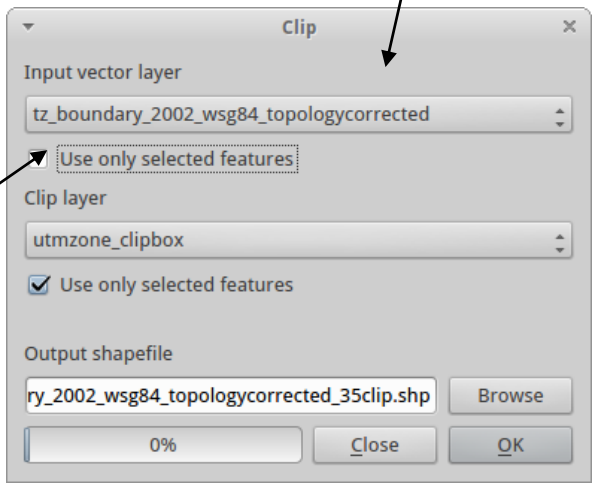
The result should show a UTM zones map just for a box around Tanzania. Check the dataset properties to see that it has a CRS of EPSG:4326.

- Select the box for **zone 35s** using the **select feature** button

  The selected zone should be yellow.

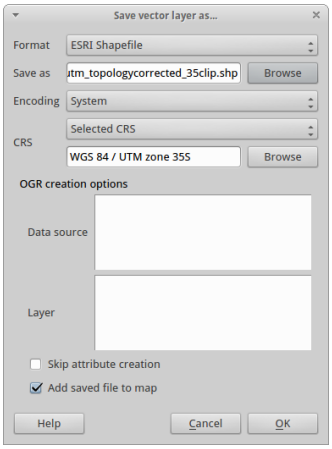
- Go to **Vector>>Geoprocessing Tools>>Clip**





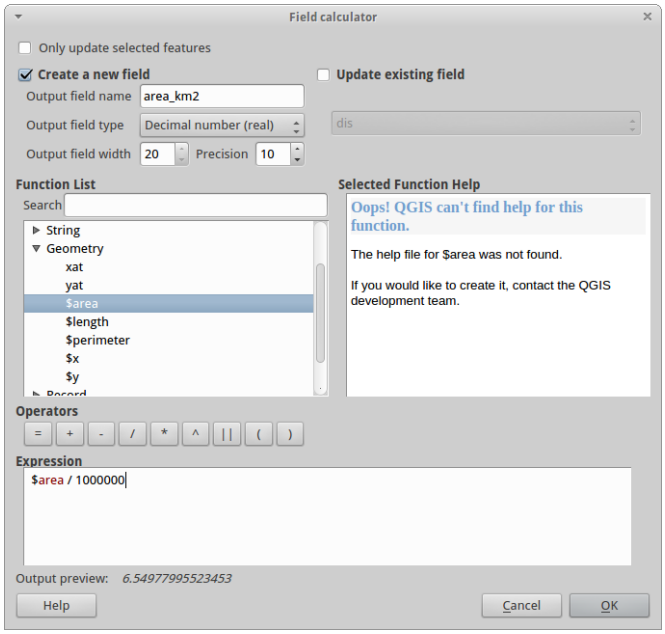
- This time do a clip using selected feature only. **Clip** the **country boundary** file to the box containing UTM zone 35
- Tick **Use only selected features**
- **Save the file** as tz_boundary_clipped35_wgs84
- **Repeat Steps 5 and 6 for zones 36 and 37**

  Save the file as tz_boundary_clipped36_wgs84 and tz_boundary_clipped37_wgs84



- **Right click** on each of the newly created files and click **save as.** Change the **CRS** to **Selected CRS** and save each dataset in their respective projections .e.g. save tz_boundary_clipped35_wgs84 from the EPSG:4326 (WGS84) to CRS UTM Zone 35 and include the projection in the output name e.g. 'name'_utmz35

- Add a **new field** to each of the 3 datasets (called area_km2) and **calculate to equal $area / 1000000**



- **Manually add the three areas values together** and make a note of the total area for Tanzania based on the 3 separate UTM zones
- **Project** the **country boundary** to the following projections
  - Full country shapefile tz_boundary_2002_wsg84.shp to **UTM Zone 36s** and name the output shapefile **tz_boundary_2002_zone36s.shp**
  - Full country shapefile tz_boundary_2002_wsg84.shp to **lambert equal area azimuthal** with **cm 34.5 and latitude of origin at -6**. Name the output shapefile **tz_boundary_2002_la.shp**
  - Add and **calculate area_km2** on these 2 files
- Compare the **area results** between the 3 examples

  a) The sum of areas calculated for the data split into the 3 UTM zones and calculated in their UTM CRS.

  b) The area calculated using UTM Zone 36s for the whole country

  3) The area calculated using Lambert Azimuthal projection for the whole country

Note the area calculated in Lambert Azimuthal Equal Area projection was closest to the areas summed from the data split into the 3 UTM zones.

11. Calculating average above ground biomass by LULC type for Saatchi, Baccini and the NAFORMA sample of Rufiju district

Prior to beginning this activity the vector and raster datasets were projected to lambert azimuthal equal area projection (see sections on projecting vector and raster data for further details of these processes).

A LULC map with only **one multipart polygon per land cover class** is needed for this analysis, this was made in the working session using the **vector>>geometry tools>>singlepart to multipart**.

- Into QGIS load the **datasets to be used in the calculations** i.e. Saatchi and Baccini above ground biomass datasets, Rufiji woody biomass NAFORMA sample, the 2010 Land-use Land-cover (LULC) shapefile and a shapefile containing a polygon bounding box of the Rufiji delta

  In the working session the following files were used:
  - Raster: saatchi_agb_la.tif
  - Raster: bacc_la.tif
  - Raster: bm_rufij_la
  - Vector: lulc_merge_la_mp2.shp
  - Vector: rufiji.shp
- Make sure that the data is projected in **Lambert Azimuthal Equal Area**, as it was agreed during the working session that this was the best projection to use for area calculations, since it proved to give a very close result to separately calculating area results for the three UTM zones covering Tanzania

Next Clip Saatchi to the country boundary

- Use the **Clip with Polygon tool in the Sextante toolbox** to clip Saatchi with tz_boundary_2002_lambert_ea_azimuth_topologycorrected.shp.

- **Save layer as** Saatchi_agb_laclip.tif

Next calculate the difference between the above ground biomass values in the Saatchi and Baccini datasets

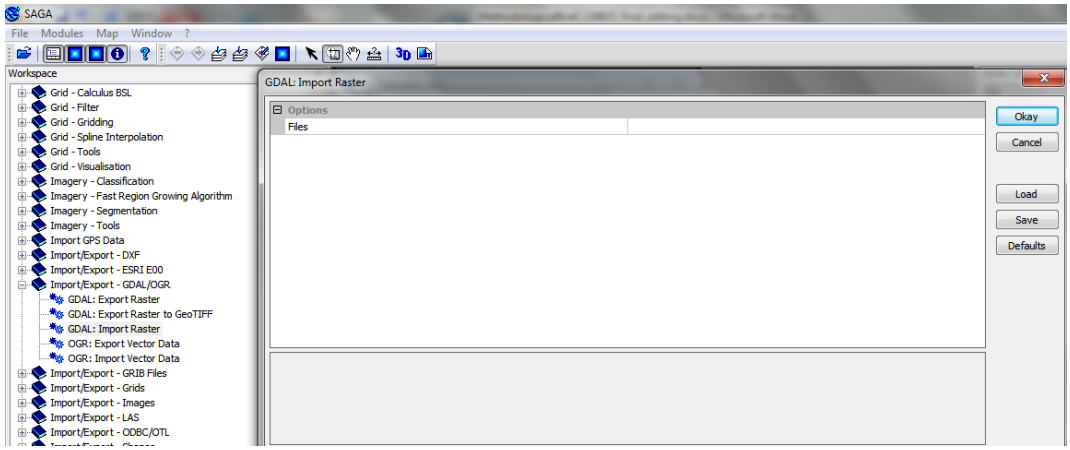
- Open the **Raster Calculator**, and use this to **subtract Baccini from Saatchi**.
  Calculate the average amount of biomass per LULC type

Note: *The biomass values are expressed in tonnes per hectare. The Rufiji district map is 30x30m resolution. Saatchi is 926m resolution and Baccini is 463m.*

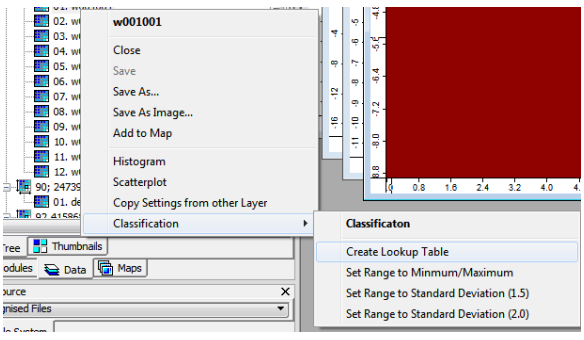
****IMPORTANT NOTE***No suitable statistics tools were found in the core suite of QGIS analysis tools. An appropriate SAGA tool was located that was accessible through the SEXTANTE toolbox but it failed to run from within QGIS in the working session so the following steps were undertaken directly within SAGA GIS.*

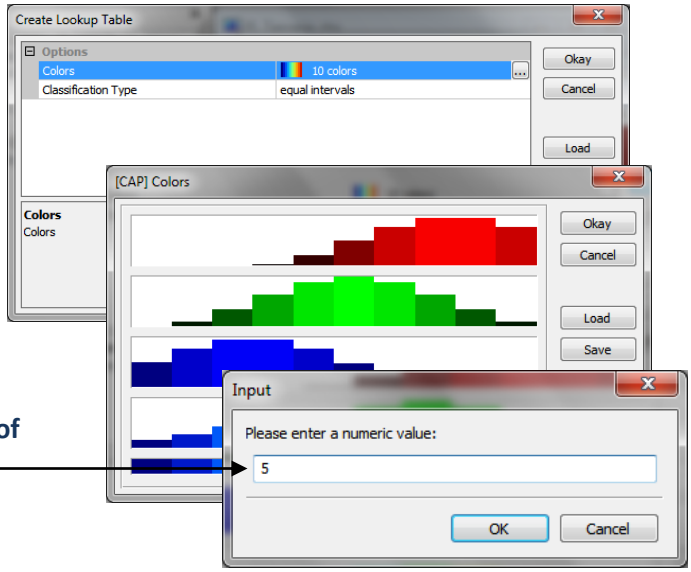
Load the clipped Saatchi, Baccini and NAFORMA Rufiji TIF files, all in Lambert Azimuthal projection into SAGA

- Open **SAGA**
- Click on the **Modules Tab**. Use the **Gdal: Import Raster tool** to load the raster datasets into SAGA



- Load the **LULC and country boundary** shapefiles using **File>>Shape>>Load**
  e.g. in the working session LULC_merge_la_mp2 and
  tz_boundary_2002_lambert_ea_azimuth_topologycorrected.shp.
- Click on the **Data tab** to see the loaded datasets
- To change the shading of the raster. **Right click** on the dataset and click on **Classification>>Create Lookup Table**





- Click on **Colors** then on the **"…"**
- Click on **Count** and change the **number of classes** to **5**

- Change the **Classification type** from Equal Intervals to **Quantiles**
- Click **OK,** then click **OK** again

Make a note of the 5 quantiles classes for each dataset

- **In SAGA**, run the tool in **Shape – Grid >> Grid statistics for polygons**, using the clipped Saatchi raster
- Click on the **data tab** and see a new copy of **lulc_merge_la_mp2**
- **Right click** on this new copy, and **save as** a new version of the LULC dataset, e.g. lulc_merge_la_mp2_saastats.shp
- **Right click on it again**, and select **Close**. Answer **yes** to deleting the selection

- Repeat this step for the Baccini and Rufiji maps as well. Remember to save each new copy of the dataset as a new shapefile

- Now, **return** to QGIS and **load the files created in SAGA into QGIS** (e.g. lulc_merge_la_mp2_saastats.shp, lulc_merge_la_mp2_baccstats.shp, lulc_merge_la_mp2_NAFstats.shp)
- Open the attribute tables and compare the results for the different datasets. A number of fields with **different statistics** have been added to the LULC map. **Note the field that shows mean value of biomass in tones per hectare per LULC type**

The next steps will clip the Saatchi and Baccini datasets to Rufiji district to compare biomass values between Baccini, Saatchi and NAFORMA for this area of interest only.

- In QGIS, **clip** the Saatchi, Baccini and NAFORMA rasters to cover only Rufiji district using the **Clip with Polygon tool in the Sextante toolbox** e.g. Clip with rufiji_la.shp
- **Save** the layers with **new names** e.g. saa_agb_la_clipruf.tif, bacc_agb_la_clipruf.tif, and NAF_agb_la_clipruf.tif, respectively

Visualize the three maps in the same color ramp.

In the working session quantiles were chosen based on the Baccini dataset to determine the class breaks to be used for all three maps. The same quantiles were necessary in order to compare the maps. The quantiles were noted in the earlier step in SAGA (**Classification>>Create Lookup Table**)

- Symbolize all five raster maps using the Baccini quantile class breaks. To symbolize one of the layers **Right Click >>Layer properties**
- Manually create the class breaks that were identified for Baccini
- **Save the style** e.g. carbon_style.qml
- To symbolize the other layers **Right Click >>Layer properties** and **Load** the saved style

12. Using IUCN Red List species data and generating species richness maps

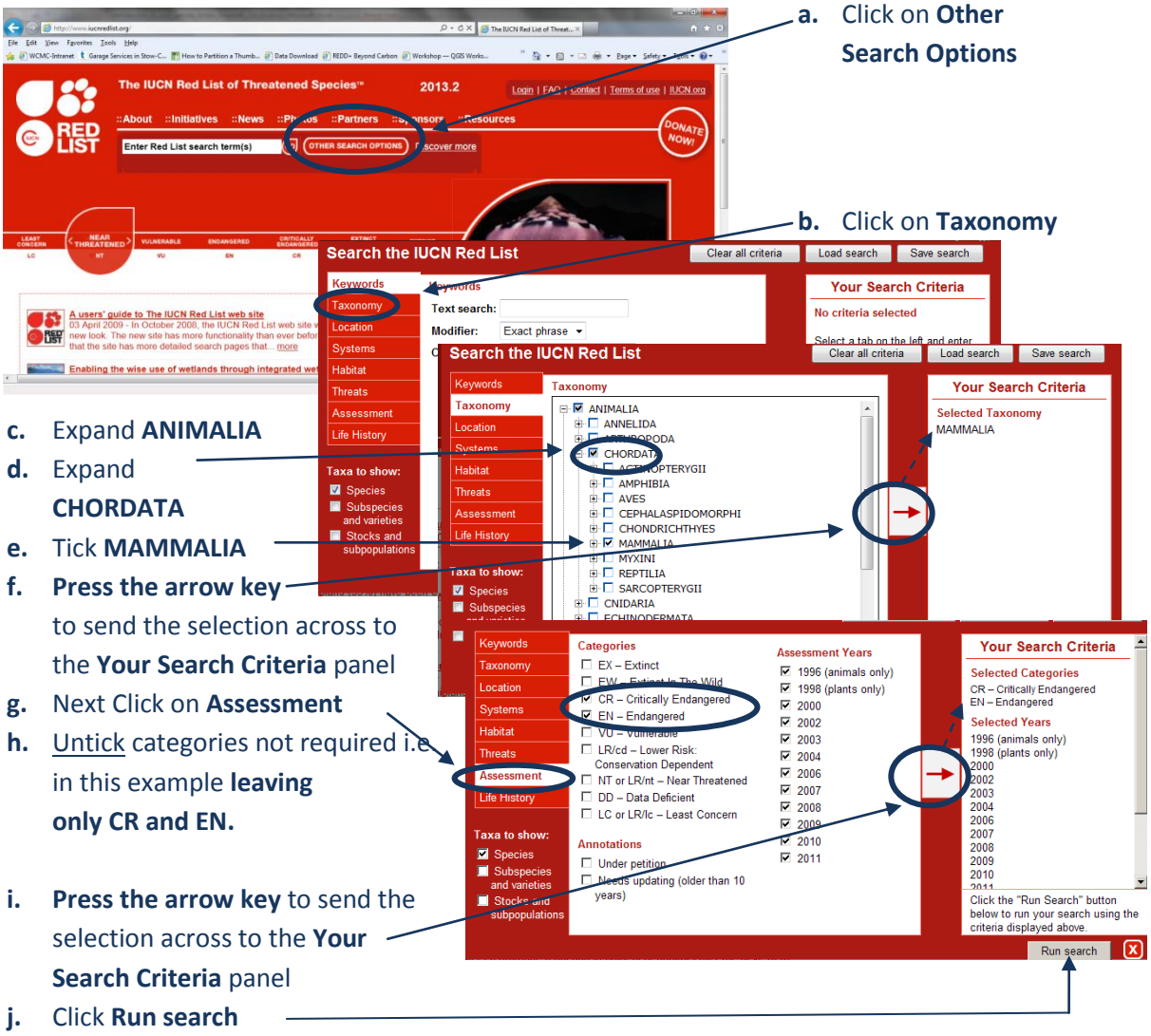
Selecting and downloading species data from the IUCN Red List website

The IUCN Red List of Threatened Species website allows users to search for and extract tabular information (in comma separated values (CSV) file format) on the status of threatened species. The website provides a user friendly interface and gives the user flexibility to customize searches based on a range of criteria. Users must register with the website to save and export customized searches.

Searching for non-spatial data

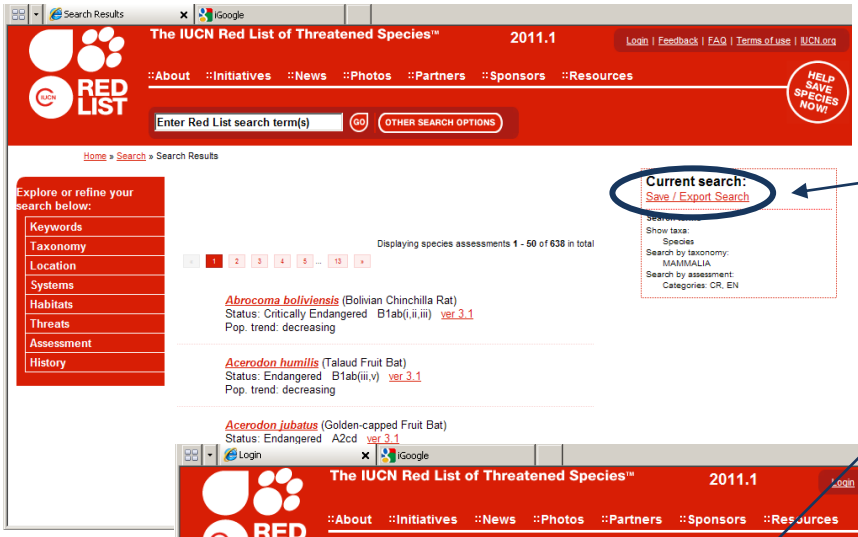
Open a web browser and go to the IUCN Red List website at http://www.iucnredlist.org/.

This search below is an EXAMPLE search for Mammals with threat status of Critically Endangered (CR) and Endangered (EN)



This search will result in a list of species within the MAMMALIA taxonomic group that have critically endangered or Endangered Red List status. The search will produce in a list of species and additional attribute data, including the threat status of each of the species.

Save the search and exporting to CSV format



a. Click **Save/Export Search**

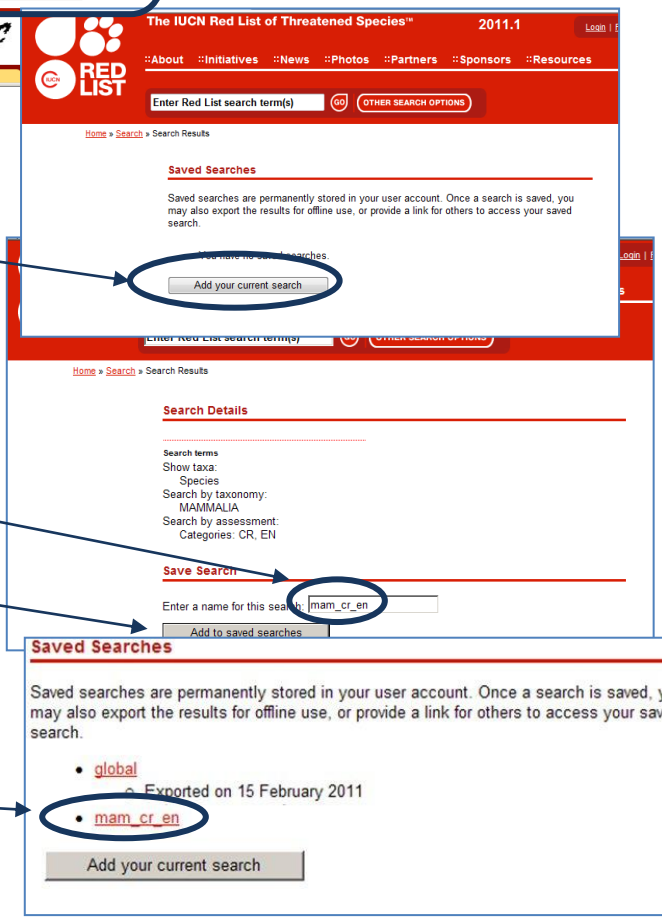
b. Click **login**

c. If already registered, **fill in your email address and password**

d. If you have not yet registered, you need to **create an account**

An account is needed in order to save and export the search results.

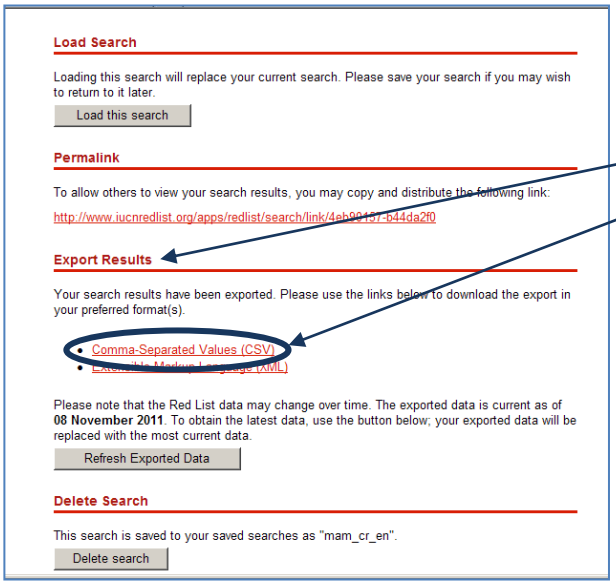
e. Once registered and logged in, Click on **Add your current Search**



f. Give the search a name e.g. mam_cr_en

g. Click **add to saved searches**

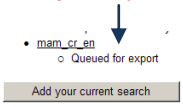
h. Click on the **saved search** e.g. mam_cr_en



i. **Scroll down to Export results**

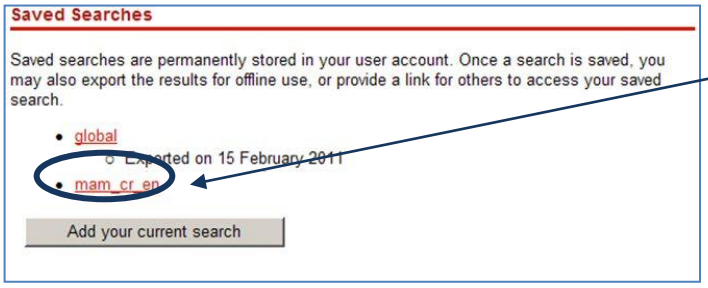
j. Click on **Comma-Separated Values (CSV)**

The dataset will then give a status of Queued for export



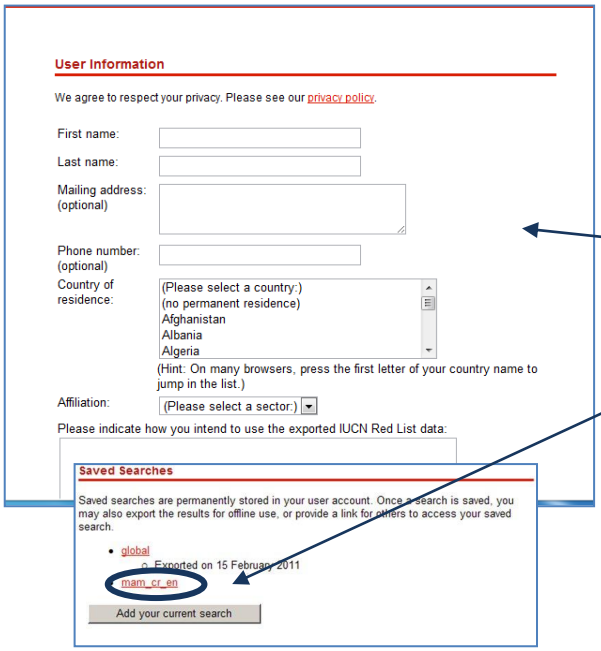
An email will be sent to you once it has been exported. This may take 24 hours.

k. Once the email has been received, login back to the Red List website.



l. Click on **mam_cr_en**

m. New users will be asked to fill out the details in the box below



The first time new users export a search, they are required to fill out some information about themselves and the intended use of the data

Click on **Supply your information** and fill in the requested details

Click on **Submit**

Click on **Mam_cr_en** again

n. **Scroll down to Export results**. Note the status now indicates that the results have been exported

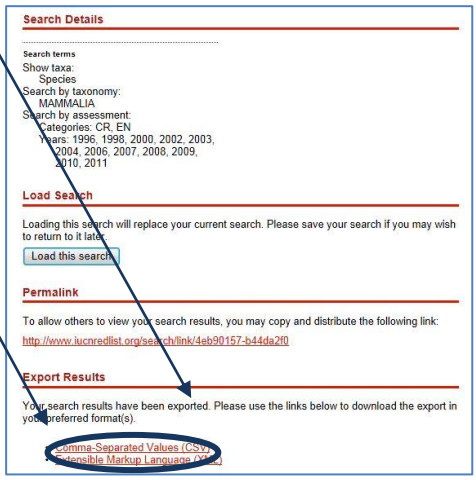
o. Click on **Comma-Separated Values (CSV)** and the zip file will download

p. If the download has placed the file in a general download folder move the zip file to a more suitable location e.g. in a project folder

q. **RenamE** the zip file to **export_mammals_date.zip** (date being the export date)

r. **Right click on zip folder**, extract the csv file

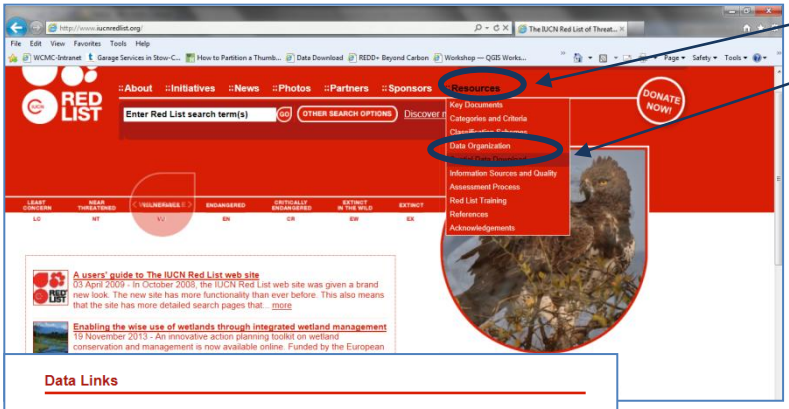
s. **Rename** the csv file to **export_mammals.csv**



Download the IUCN Red List spatial data layers

The next steps are for downloading spatial data. It is only possible to download the whole global dataset. It is not possible to filter by county prior to download. It is important to note that some of the spatial datasets are very large.

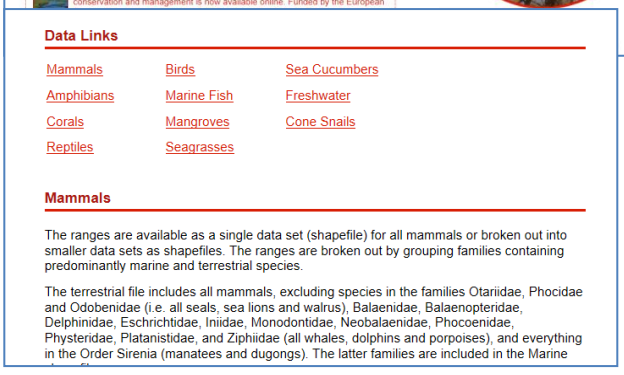
a. Open a web browser and go to the IUCN Red List website at http://www.iucnredlist.org/



b. Click on **Resources**

c. Click on **Spatial Data Download**

d. **Scroll down** on the Spatial Data page to the **Data Links** section



e. **Click the links** to navigate to each dataset

Download the Following global datasets:

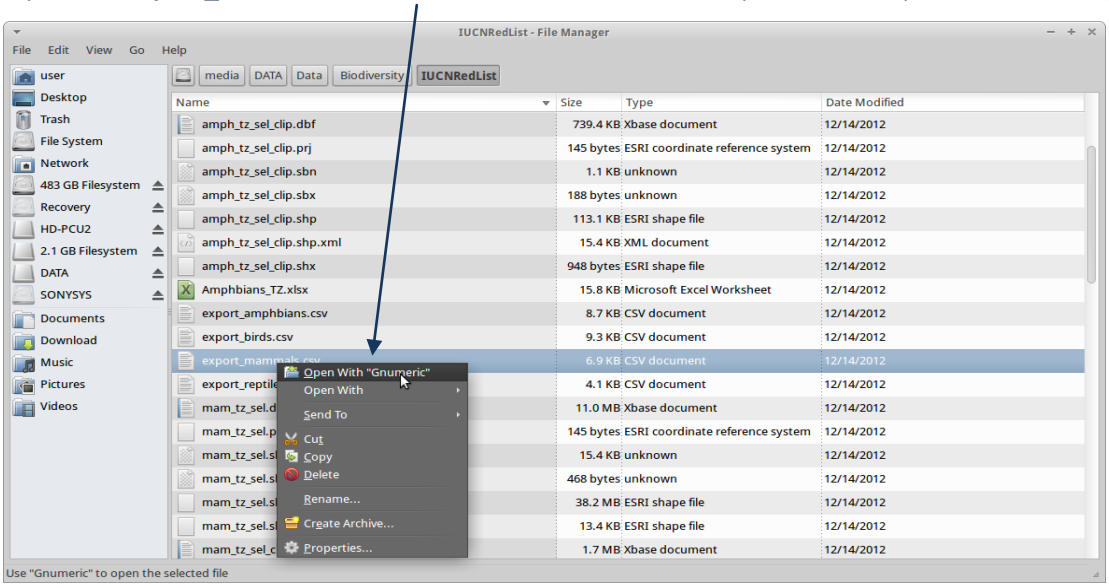
- Mammals-terrestrial
- All Amphibians
- Birds (via the link to the BirdLife Data Zone)

(Leave Reptiles for now as assessment is not yet complete for all species)

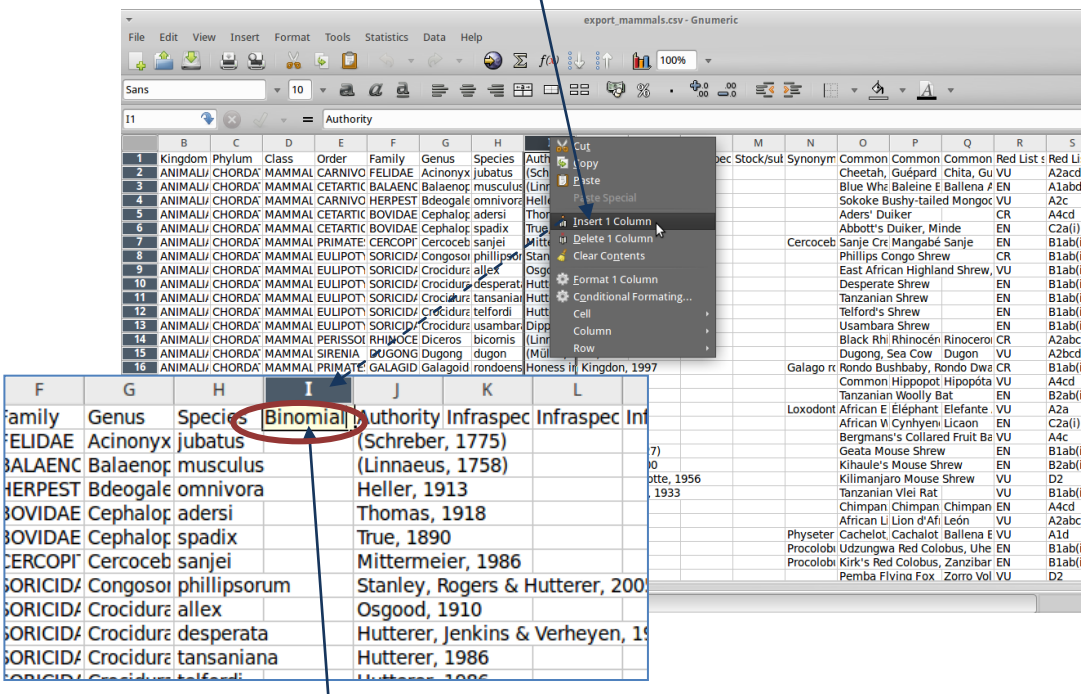
These files are all stored in geographic coordinate system (EPSG: 4326). Be aware the files are very large and will take some time to download.

Format CSV file for joining to spatial data

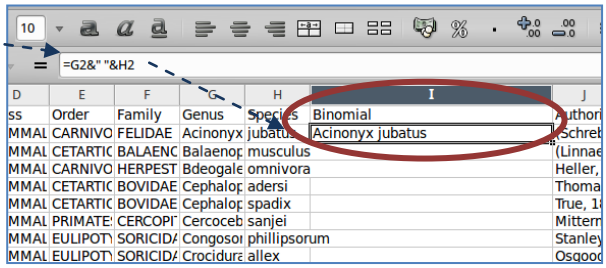
a. Open the **export_mammals.CSV file in Gnumeric** (in a basic open source equivalent to excel)



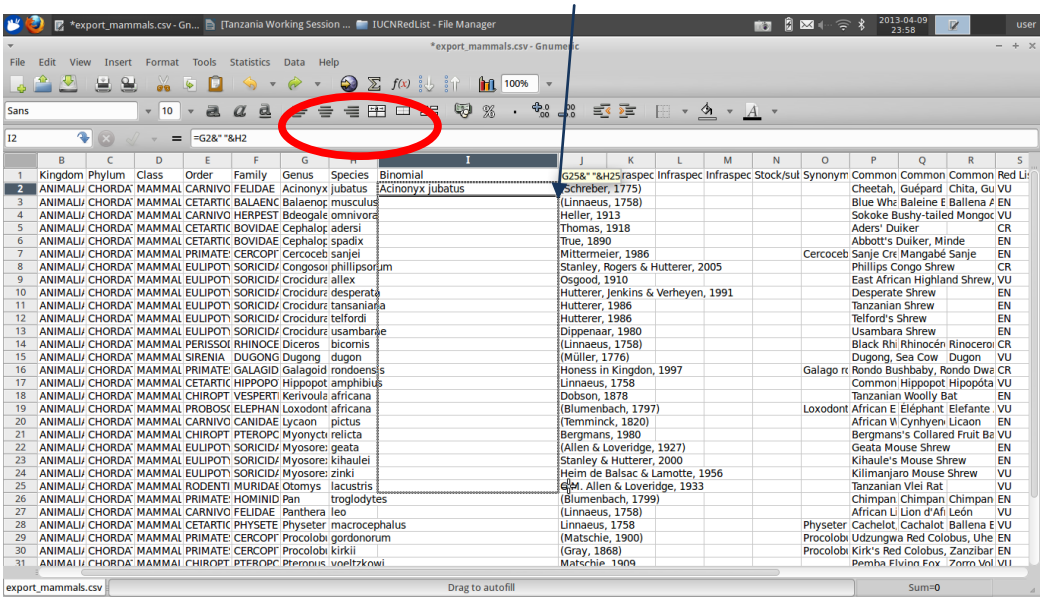
b. **Right Click** and Click **Insert 1 column** after the genus and species columns called 'binomial' (this will become a combined Genus and species field to match with a binomial field that is present in the spatial data)



c. In the 1st cell type in **=G2&" "&H2** and press **return**



d. Click into the cell and move your cursor to the bottom right hand corner until it turns into a small black cross **+**. **Drag it down to fill the rest of the column**



e. Scroll along the **column headings** of the table. Some will need to be changed as GIS software such as QGIS will not accept them. **Change the ones listed below in red**

OLD Field Name		**New Field Name**
Species ID	=	Species_ID
Kingdom	=	Kingdom
Phylum	=	Phylum
Class	=	Class
Order	=	Order
Family	=	Family
Genus	=	Genus
Species	=	Species
Binomial	=	Binomial
Authority	=	Authority
Infraspecific rank	=	Inf_rank
Infraspecific name	=	Inf_name
Infraspecific authority	=	inf_auth
Stock/subpopulation	=	stk_subpop
Synonyms	=	Synonyms
Common names (Eng)	=	com_eng
Common names (Fre)	=	com_fre
Common names (Spa)	=	com_spa
Red List status	=	rl_status
Red List criteria	=	rl_criteria
Red List criteria version	=	rl_version
Year assessed	=	year_ass
Population trend	=	poptrend
Petitioned	=	Petitioned

f. Click **File>>Save as** and save the file as e.g. **Mam_TZ.csv** and **close Gnumeric**

Clip IUCN spatial data to area of interest and check geometries for errors

a. Open **QGIS**

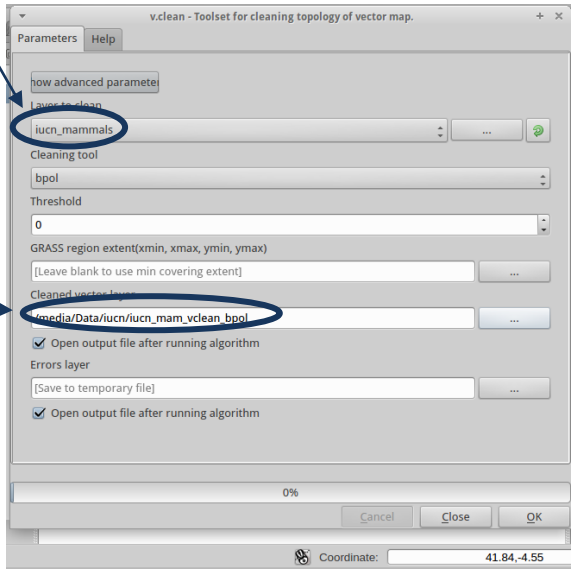
b. Load in a the **IUCN species CSV file formatted** in the previous section i.e. MAm_TZ.CSV using the **Add vector layer button** (change the **format** in the bottom right hand corner from shapefile to **CSV** to see CSV files)

c. Load in a vector boundary file covering **the area of interest** e.g. Country boundaries again using the **Add vector layer button**

d. Load in a vector boundary file of the **IUCN species** e.g. iucn_mammals_global.shp

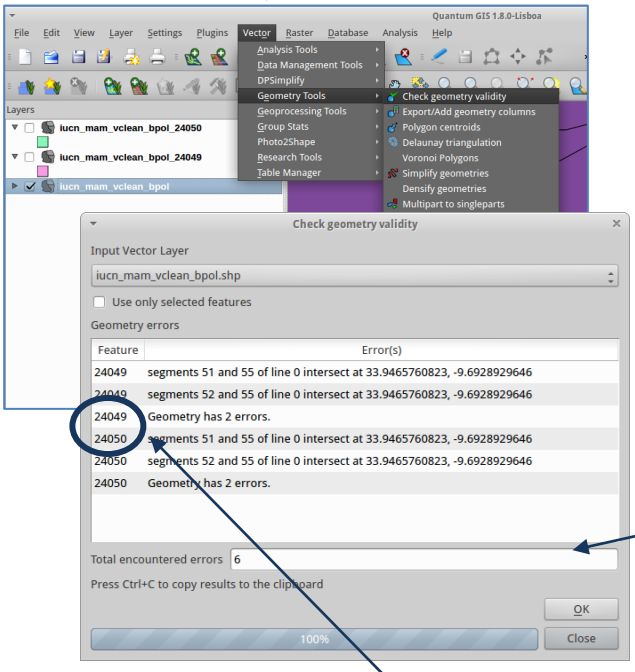
e. Run a **cleaning process** on the IUCN data as it is a complex dataset and **shapefiles are not topological datasets**, meaning that there may be errors which can cause problems in an analysis.

Give the **cleaned output** dataset a name

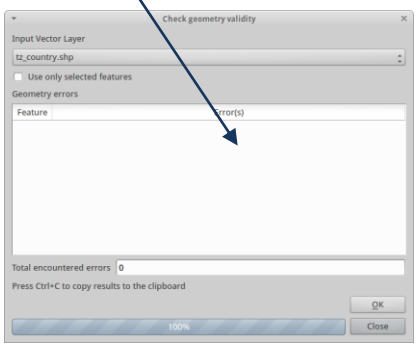


f. Check for geometry errors in cleaned IUCN dataset and the country boundaries shapefile

g. From the vector menu click **Geometry Tools>>Check geometry validity**



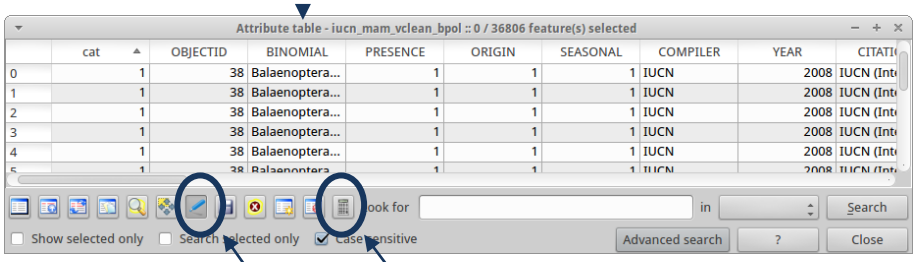
h. In this example the country boundaries file reported **0 errors**



i. <u>BUT</u> In the cleaned IUCN data there are some **errors to fix** even after the cleaning

j. Note the **Feature Ids** that contain errors. These refer to the record id's of the polygons in the spatial dataset

k. To help locate the errors **Right Click** on the **IUCN cleaned dataset** and **Open Attribute Table**



l. Click on the **Toggle Editing** button to start editing

m. Then click on the **Calculator button**

n. In the calculator window tick to **Create a new field**

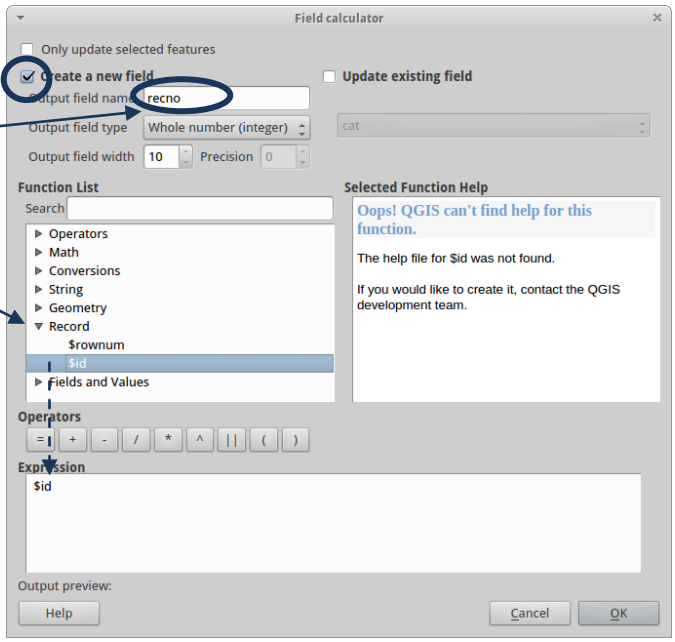
o. Call the field **recno**

p. Leave type as **Whole number (integer)**

q. Expand Record and '**Double Click** on **$id** to send it into the expression box

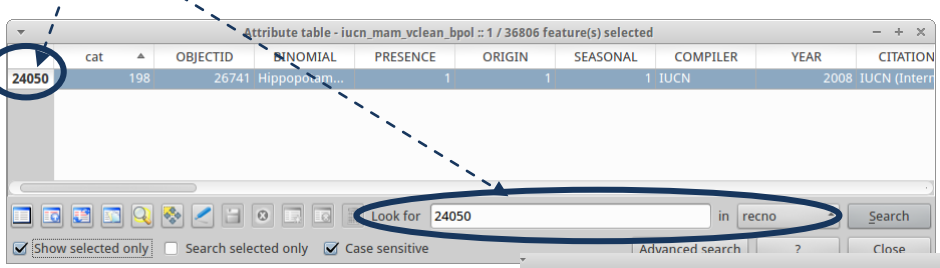
r. **Click OK**

s. Click the **Toggle Editing** button to **stop editing**

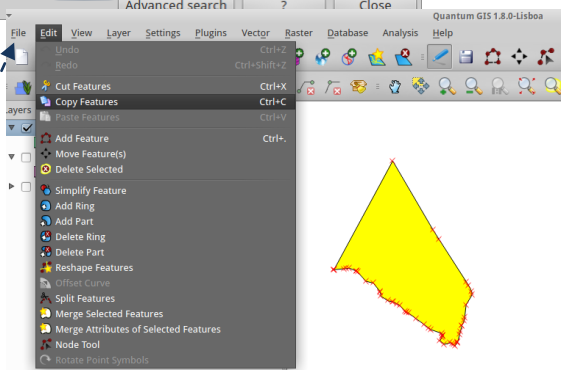


It is now possible to identify the errors that were listed in the error report.

This tutorial will not describe the editing process but a few tips are below

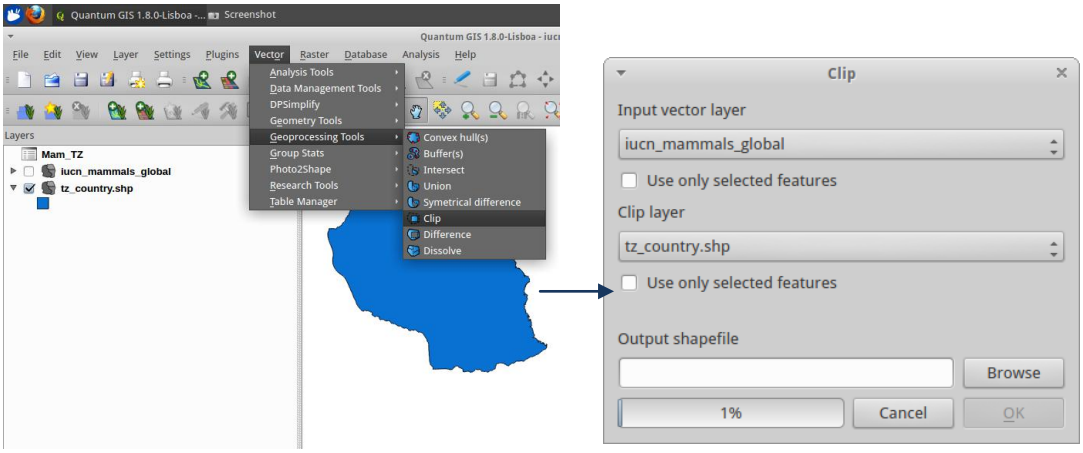


It is probably easier to select out individual error polygons, save them as a new shapefile and delete them from the main layer. Then fix the single polygon and copy and paste it back into the main layer.



t. Once editing has finished make sure **Toggle Editing** is stopped and **Check Geometry** again

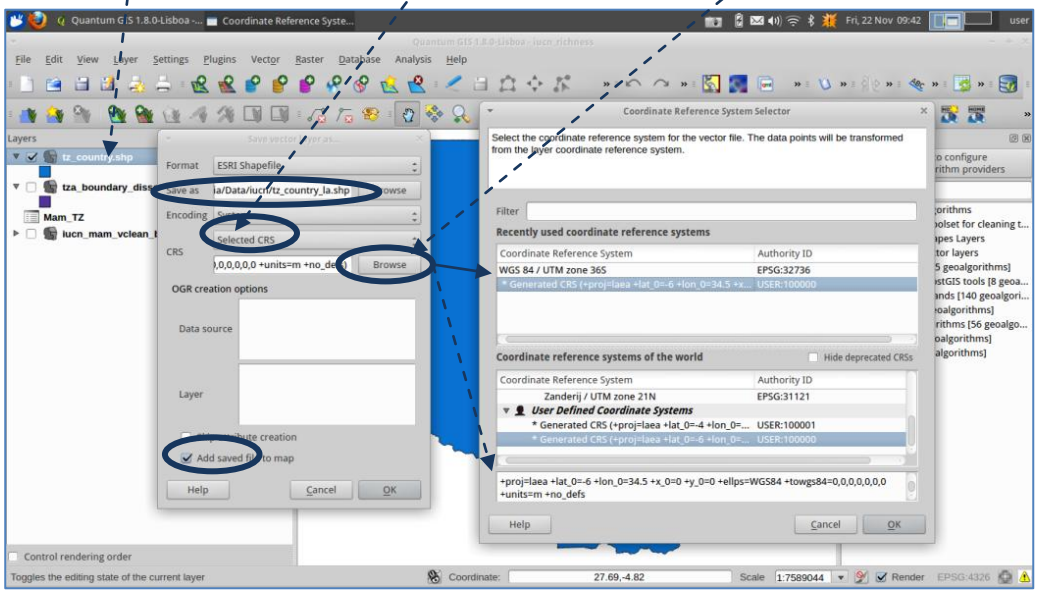
u. **Clip** the **cleaned IUCN** dataset using the **area of interest boundary** dataset



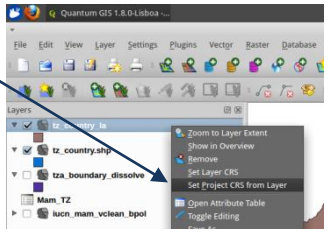
Project data to an equal area projection

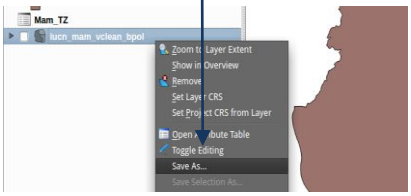
In order to generate a dataset of hexagons with a specific area, the data need to be in a CRS which has units of meters. The CRS EPSG:4326 has units of decimal degrees, therefore the data need to be projected from geographic EPSG:4326 into a different CRS. In the following example Lambert-azimuthal-equal-area projection is used.

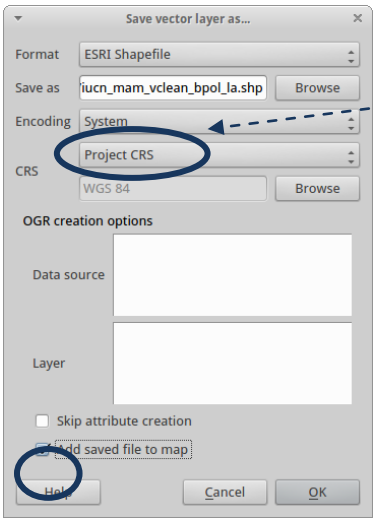
a. To project the data **Right Click** on the **country boundaries** dataset and Click **Save as...** Choose **Selected CRS** and click **Browse** to select the **output CRS** e.g. in this example Lambert -azimuthal-equal-area projection is chosen



c. **Right Click** on the **projected dataset** and **Click Set Project CRS from layer** to change the projection of the map canvas from Geographic to Lambert -Azimuthal-Equal-Area. **Right Click** on the **final cleaned IUCN species** dataset and Click **Save as...**







d. Save the layer with **Project CRS** which was set to Lambert-azimuthal-equal-area in the previous step, i.e. the output dataset will be saved in the same CRS as the Project.

e. Remove the Geographic versions of the datasets from the QGIS project to reduce confusion

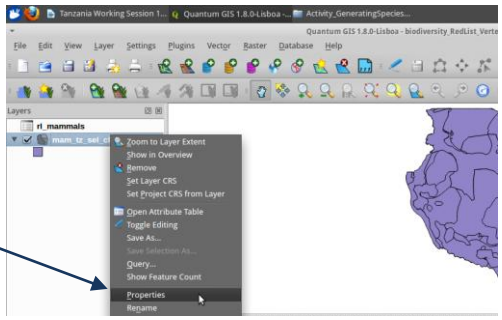
There should now be 3 files in the project :-

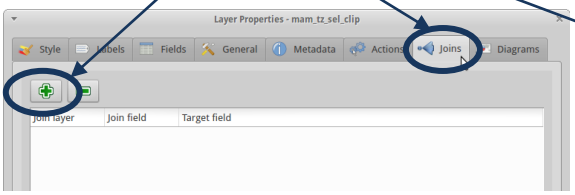
Formatted species threat categories CSV file (e.g. Mam_Tz.csv)
Area of interest shapefile in projected CRS (e.g. TZAcountry_la.shp)
Cleaned species shapfile in projected CRS (e.g. Mam_vclean_bpol_la.shp)

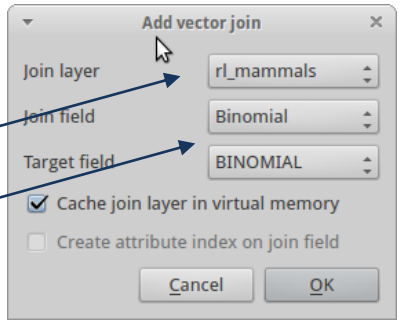
Join CSV to spatial data and create subset of the data based on Red List category

The next step is to join the threat status onto the final cleaned IUCN species layer and select only species of the selected threat categories (CR and EN)

a. **Right click** on the **species layer** (e.g. Mam_vclean_bpol_la.shp) and click **Properties**
b. Click on the **joins tab**
c. Then click **'+'**







d. Make the **join layer** the **formatted species threat categories CSV file** (e.g. Mam_Tz.csv)
e. Set the **join fields** to be **Binomial** (this field is present in both datasets and contain the species and genus information)

Because the data are managed together by IUCN then there is a perfect match between species names (binomial) in both tables.

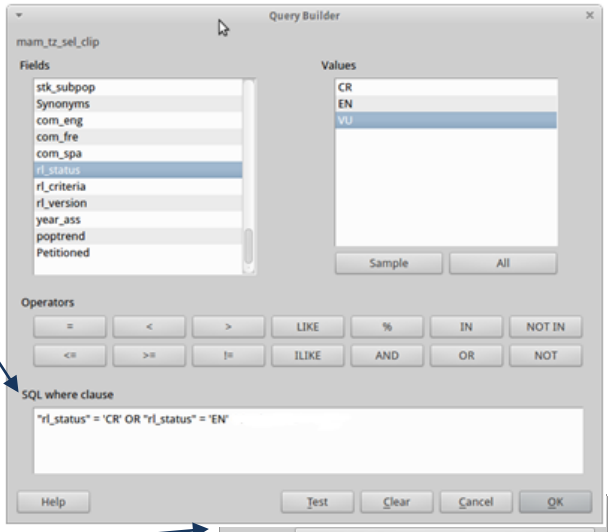
f. The join will appear in the join window. **Open the attribute table** of the **cleaned IUCN species layer** and **check the join** has worked correctly. There should be a threat category near the end of the table (this is a temporary join)

g. Close the attribute table

h. **Right Click** on IUCN species layer and click **Query**

i. Create a query to select out just those species which are of **status CR or EN**

j. Click **OK** to close the window

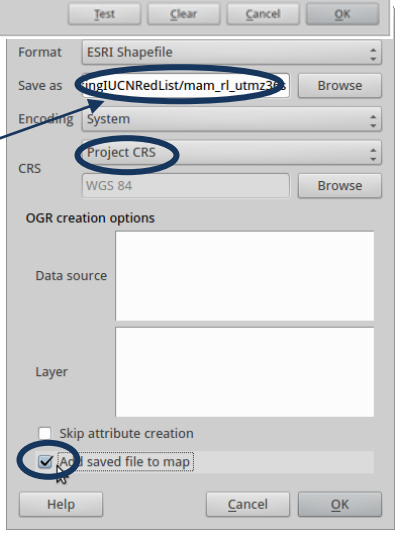


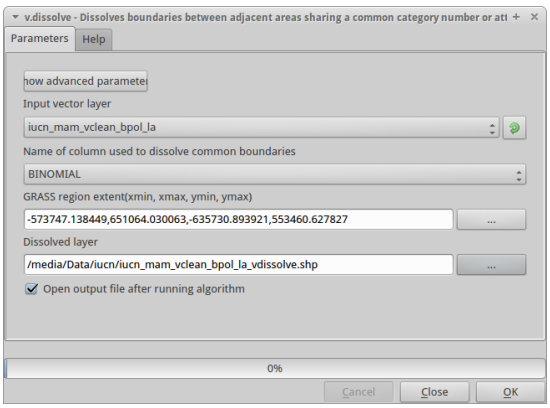
k. Now save the dataset (with its join and query set) to a new dataset

l. Save it in the **projection of the project**.

m. Save it with a **new name** e.g. Mam_RL_ENCR_la.shp

n. Tick **Add saved files to map**



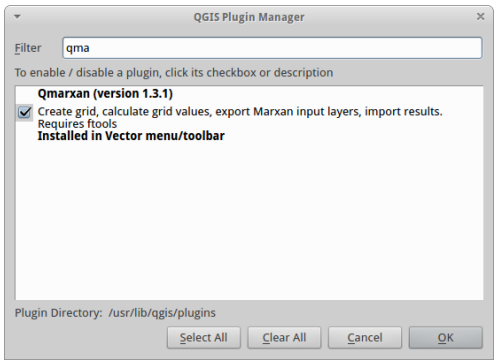


o. **Dissolve** data based on **binomial field** to remove any duplicate features

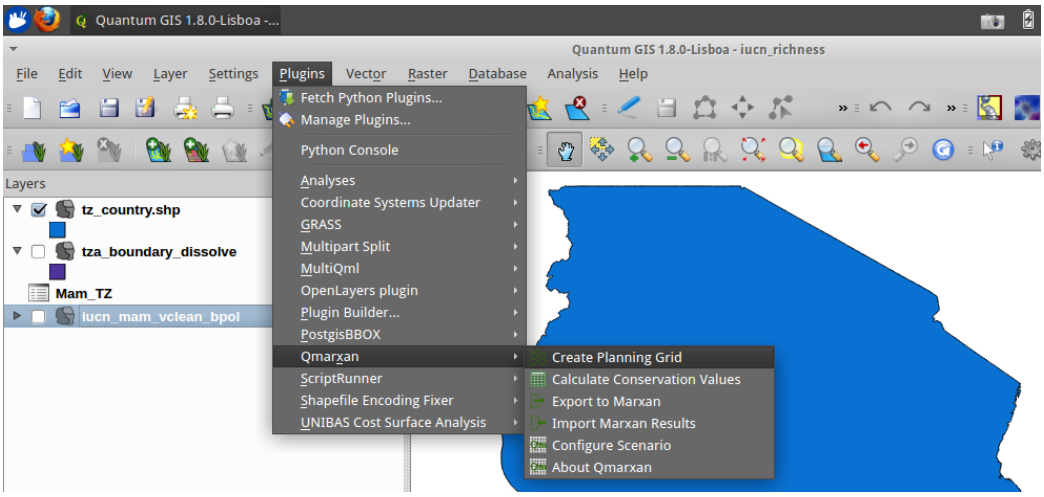
(Use the v. Dissolve GRASS tool in the Sextante toolbox NOT the one from vector>>geoprocessing tools dissolve)

Generate a dataset of hexagons or squares using the Qmarxan plugin

a. **Fetch** the **latest QMarxan plugin** from the QGIS tool repository and activate it from the **Plugins>>Manage Plugins** menu



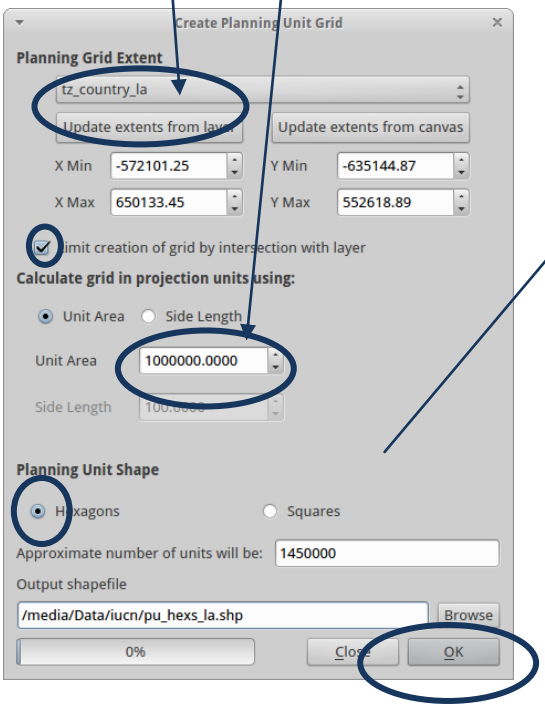
b. From the Plugins menu click **Qmarxan>>Create Planning Grid**

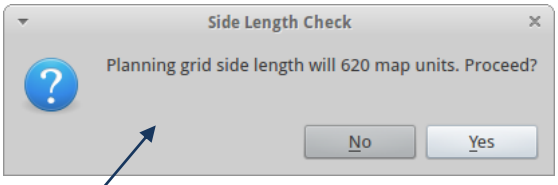


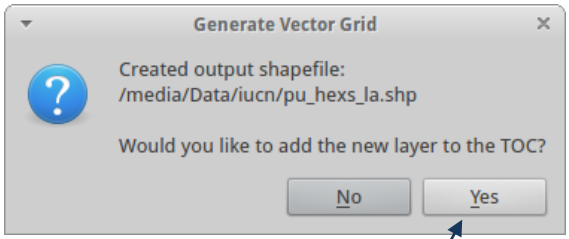
c. Select the **projected Area of interest file** e.g. tz_country_la.shp and click **Update extents from layer**

d. **Set unit area** e.g. 1000000 for 1km hexagons

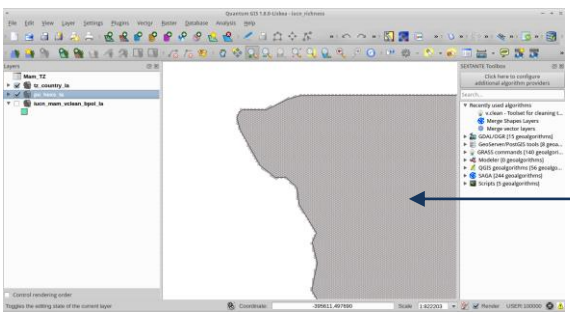
e. Tick **Limit create of grid by intersection with layer**







Click **Yes** to both windows

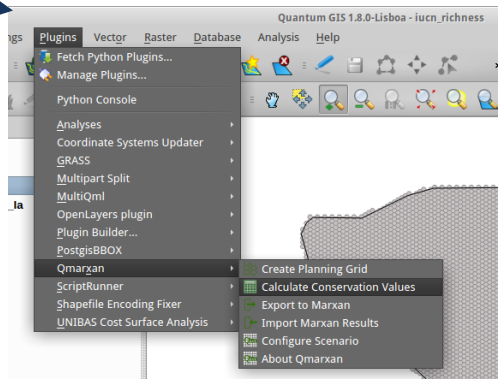


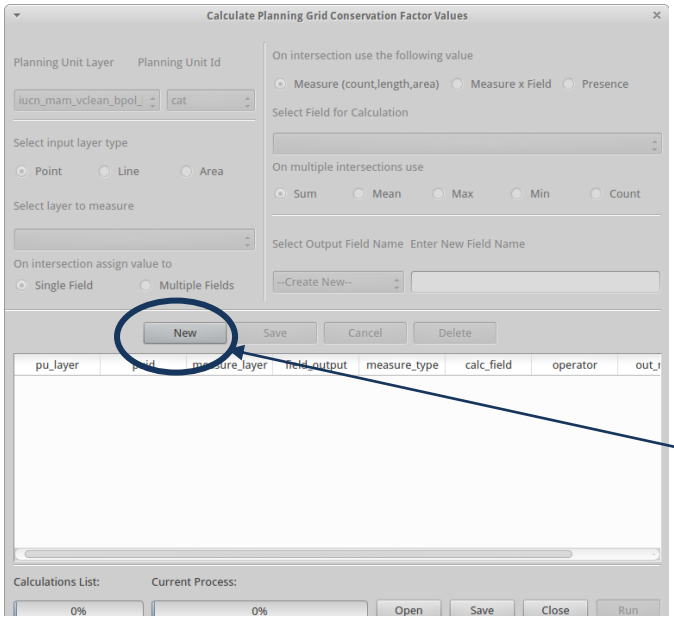
Zoom in to check the hexagons

Generate species richness using the Qmarxan plugin

a. The next step is to use the planning units file to generate a count of number of threatened species in each planning unit hexagon

b. From the **Plugins** menu click **Qmarxan>>Create Conservation Values**





c. The **Calculate Planning Grid Conservation Factor Values** window opens

d. Click **New**

a. Select the **Planning Unit Layer** generated in the previous step

b. Select the **Planning Unit Id** (this is unique ID for each hexagon in the planning units dataset upon which the summaries will be made)

c. Set **On intersection assign value** to **Single field**

d. Set the **On intersection use the following value** to **Measure (count, length area)**

e. Set the **On multiple operations use** to **Count**

f. Set the **new Output Field Name** to be added to the planning units e.g. mam

g.

h.

i.

j.

k.

l.

m.

n.

o.

p.

q.

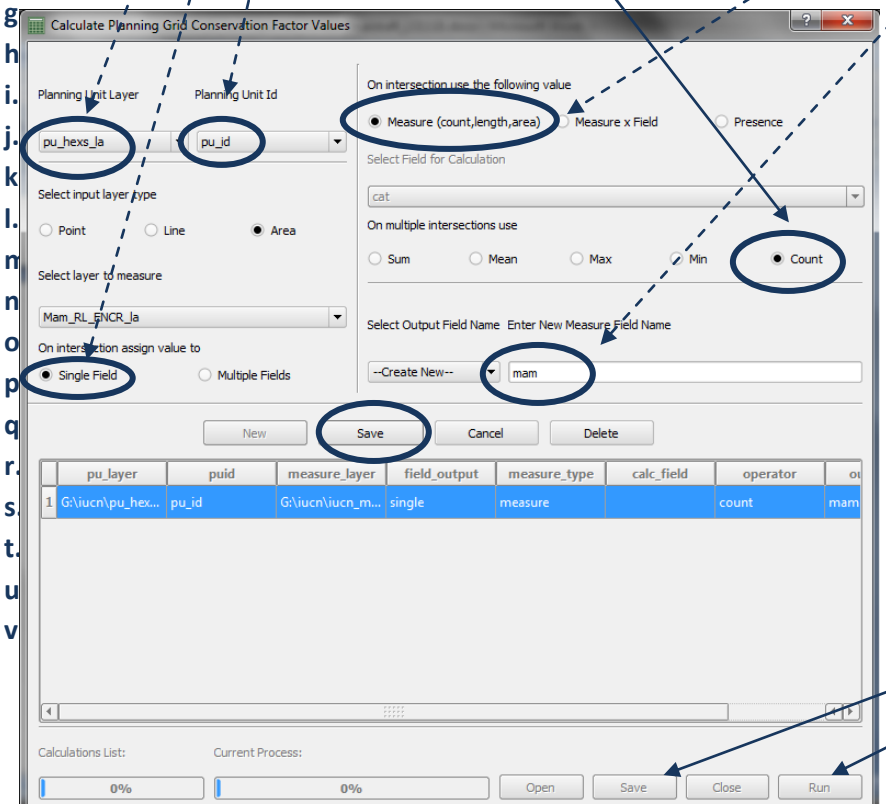
r.

s.

t.

u.

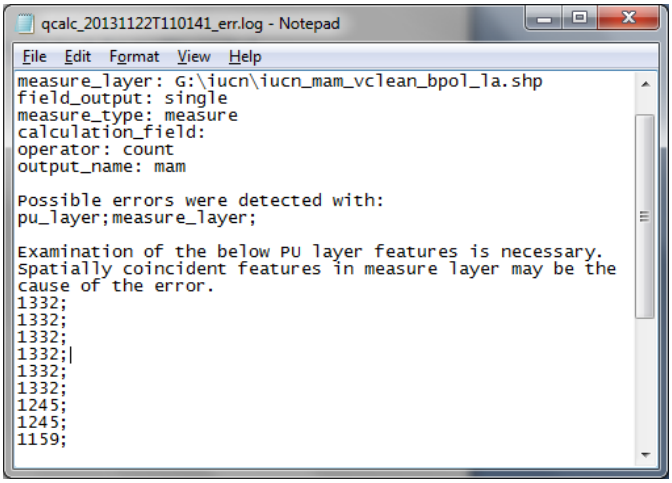
v.



Click **Save**

Click **Run**

Check any errors located in the error log and spot-check for quality control

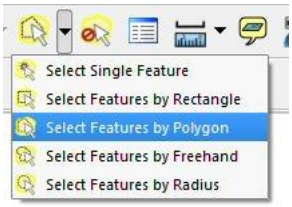


a) In the Planning Units folder **open the error log** generated by the previous step

For each of the planning units listed manually check the results of the features listed

b) **Right Click** on the **Planning Unit Layer** and **open the attribute table**
c) **Select** the **PU** and **zoom to it in the map canvas**

d) Click on the IUCN species layer in the table of contents
e) Then Click on the **Select Features by Polygon tool** and **draw a polygon** around the hexagon (PU) to select features within the IUCN dataset corresponding to that hexagon



f) The number selected should match with the count in the 'mam' field in the Planning Unit Layer. Check again by drawing the polygon around the hexagon and if necessary alter the value in the 'mam' field to match the number selected
g) Repeat for the other planning units listed in the error log

The dataset can now be shaded in graduated shading according to the count of the number of species.

13. Using NAFORMA data to create a map of tree species richness at the inventory locations

The objective of this exercise is to understand how to map the NAFORMA data to create a map of points showing tree species richness at the inventory locations. The NAFORMA database consists of a number of CSV (comma separated value) text files which contain the information collected in the National Forest inventory. The NAFORMA data are recorded by inventory plot, where each plot belongs to a cluster. There are a maximum of 10 plots per cluster, organized in an L-shape. The NAFORMA data contain latitude and longitude locations of the inventory plots expressed in UTM zones, and as the country spans three different zones (UTM zone 35s in the west, 36s in the central part of Tanzania and 37s covering the eastern part) when mapping the NAFORMA data it is important to import the data correctly in each zone.

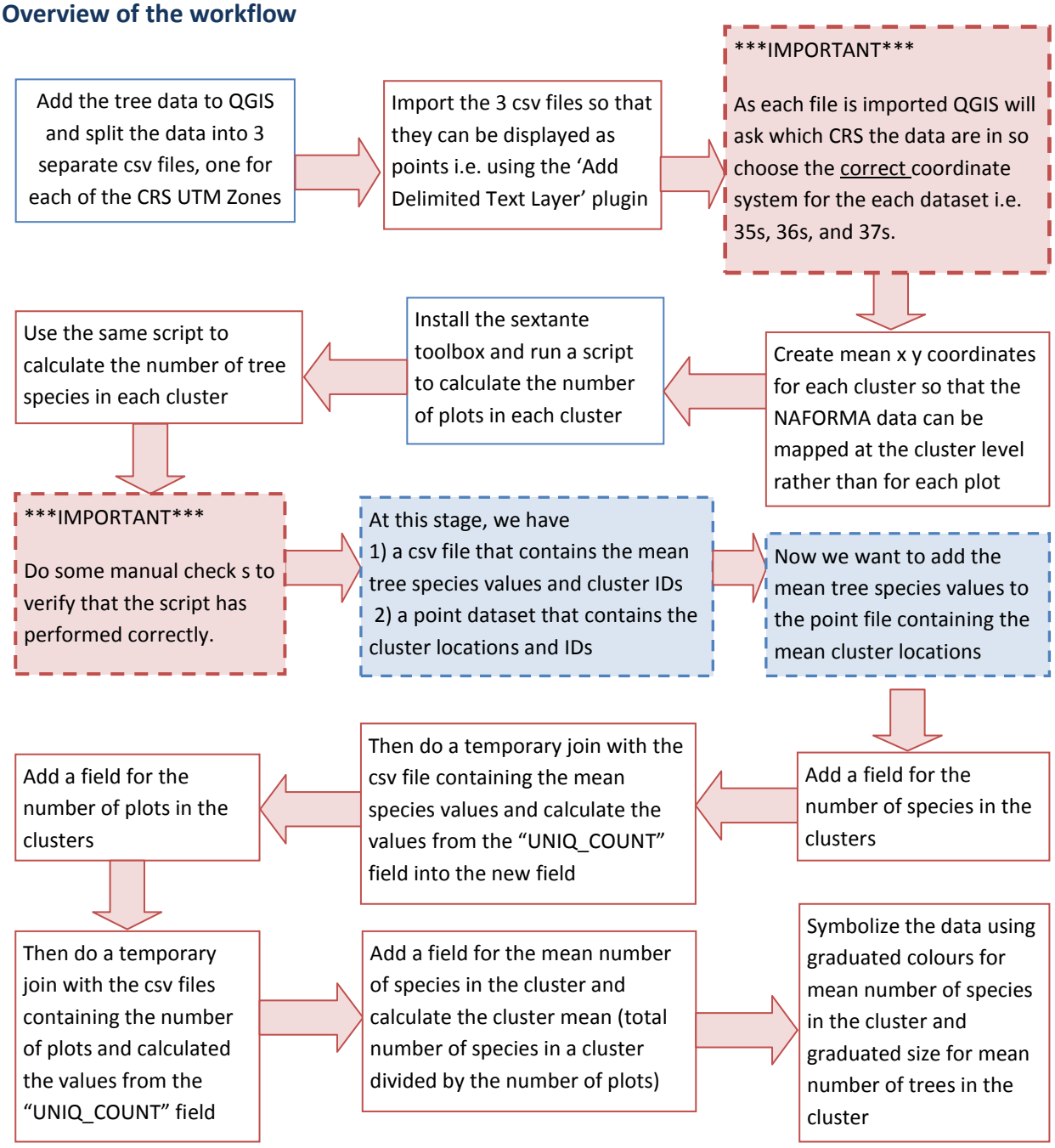
This tutorial demonstrates one method for correctly importing the tree data from the NAFORMA inventory into QGIS (an open source geographic information system). If all the data were imported in a projection of only one of the UTM zones, the clusters that are located in the other zones would get a distorted location because they would have the wrong zone projection. The tutorial also helps users gain an understanding of the structure of the NAFORMA data and covers how to present the data at the cluster level through generation of mean coordinates for the clusters and how to run an analysis to create a map by cluster showing the average tree species richness in NAFORMA plots.

TECHNICAL SUMMARY OF PROCESS

This tutorial uses two files from the NAFORMA inventory: a) the plot.csv which contains the x y location and UTM zone information for each plot and the cluster the plot belongs to and b) the tree.csv text file which contains a record of all trees measured in the NAFORMA inventory, the tree species and which plots they were measured in.

The tree.csv contains information about the trees recorded at each plot, where every row represents one tree. The x y location of the plot is also recorded in this file. Plots which do not contain any trees are excluded from the tree.csv file. The tree and plot files can be linked together by a combination of the plot id and cluster id. Other information is stored within these files but for now these are the important information to note.

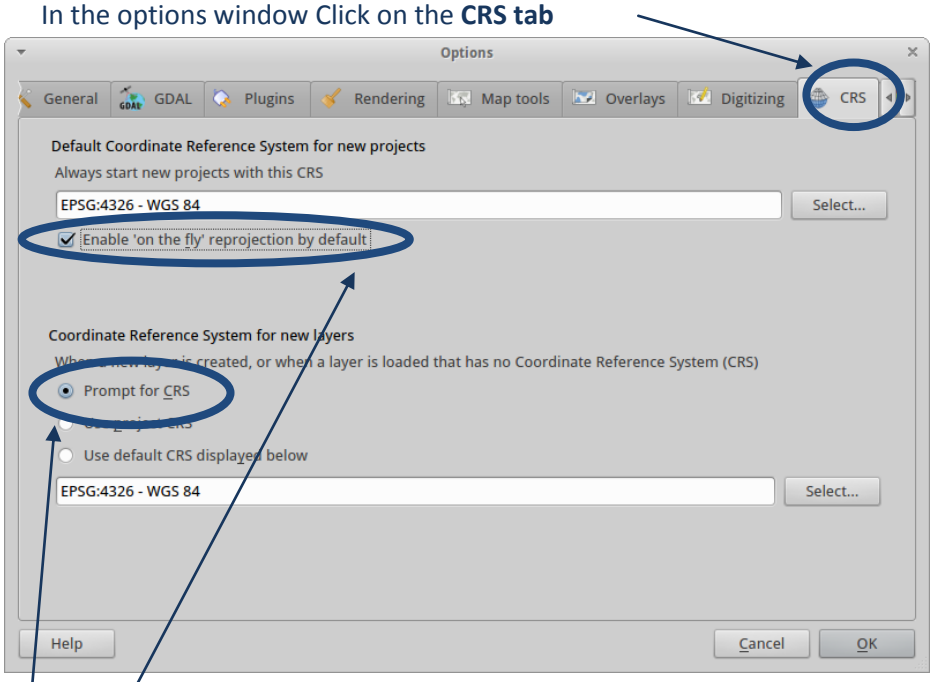
Overview of the workflow



The following document is a step by step tutorial to produce a map of points showing tree species richness at each of the NAFORMA cluster points. In the following steps the NAFORMA plot dataset will be imported into QGIS and mean coordinates will be created for each cluster.

1. First there are a few changes to be made to the setting in the QGIS interface

❖ **Open QGIS** to change some default settings. Click on **Settings>>Options**.
In the options window Click on the **CRS tab**



Click **Prompt for CRS when loading layers with no CRS defined**.

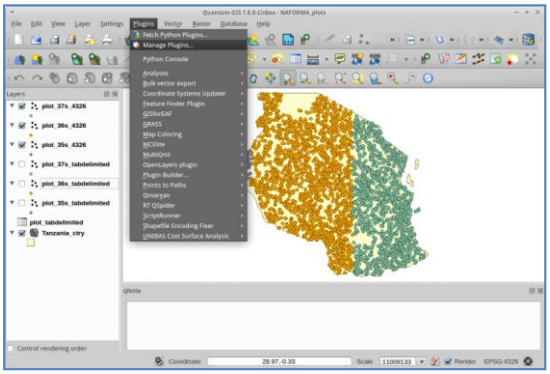
This means that QGIS will never assume it knows the CRS of data and will ask the user what projection that data are in. It stops QGIS from adding data incorrectly.

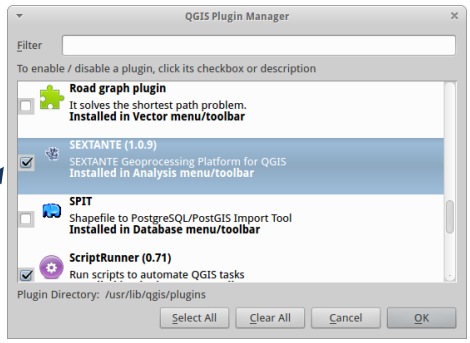
Click **Enable 'on-the-fly' reprojection by default**

Without this set, data in different projections cannot be displayed on top of each other.

Click **OK** to close the Options window.

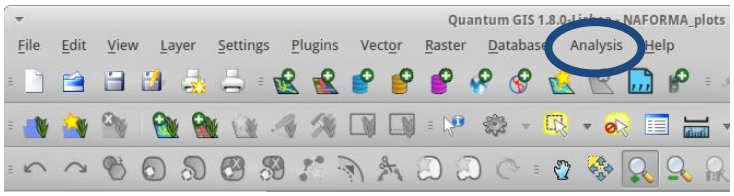
2. Activate some additional geoprocessing tools which are not automatically installed with QGIS. From the main menu click **Plugins >>Manage plugins**





The QGIS plugin manager window will appear. **Scroll down** until you see **SEXTANTE**. Put a tick next to **SEXTANTE**. Click **OK** to close.

(If the sextante tool is not visible it means that it needs to be fetched from the plugin repository. From the main menu pick "Plugins >>Fetch Python Plugins". Then re-do step 24)



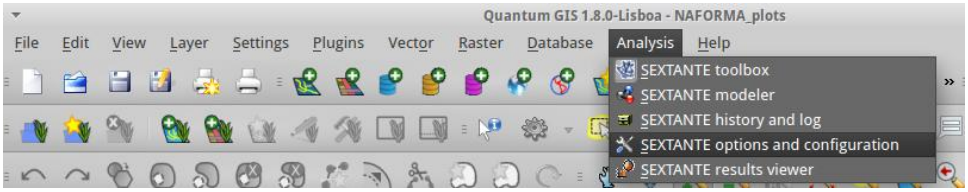
There is now a new tab called **Analysis** in the QGIS menu bar

Sextante is a geoprocessing platform running from within the QGIS interface, that enables QGIS users to access additional geoprocessing capabilities from other Open Source GIS software.

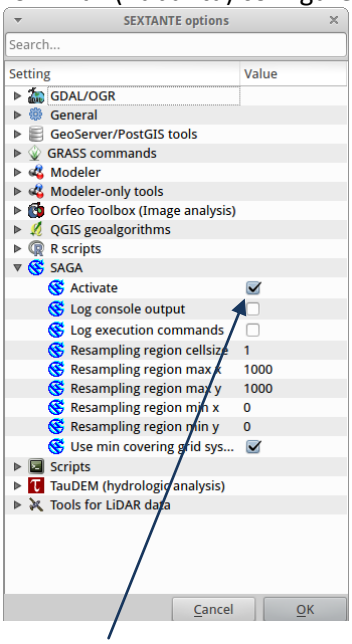
- Some of the SEXTANTE tools that are particularly useful use the SAGA GIS software. **SAGA GIS needs to be installed** in order for QGIS to access the tools. If not already installed download and install SAGA from http://www.saga-gis.org/en/index.html

- Once installed SEXTANTE needs to be configured for the SAGA tools to become available. There are some slight differences in configuration for the Windows and Xubuntu (Linux) operating systems

- Click on **Analysis** from the main menu and go to **SEXTANTE options and configuration**

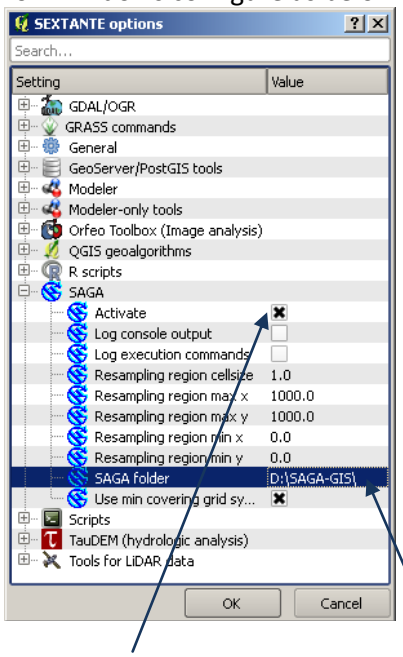


For Linux (Xubuntu) configure as below:-



Click to **Activate** SAGA

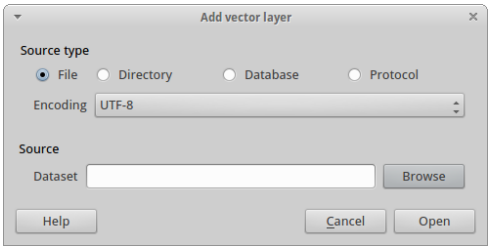
For Windows configure as below:-

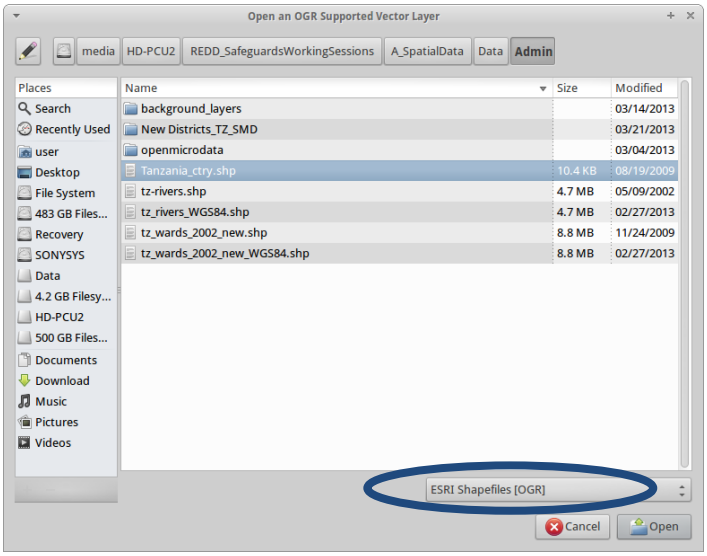


Click to **Activate** SAGA
Specify the location of the SAGA installation

3. **Now data can be added into the QGIS and be prepared for the analysis**

   - Click to **Add vector layer**. Change the file type to **ESRI shapefile** and browse to the folder containing the Tanzania country boundary file 'Tanzania_crty.shp'



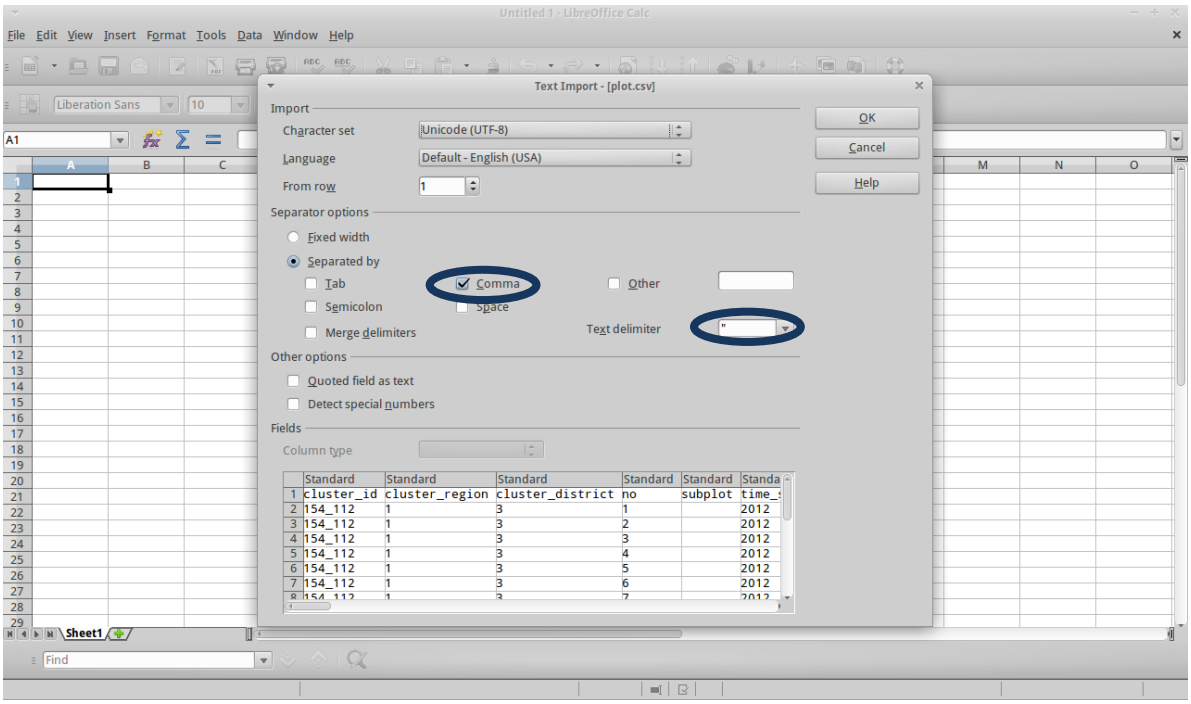


(In the working session this shapefile is located on the USB sticks in Data/Admin/Tanzania_crty.shp)

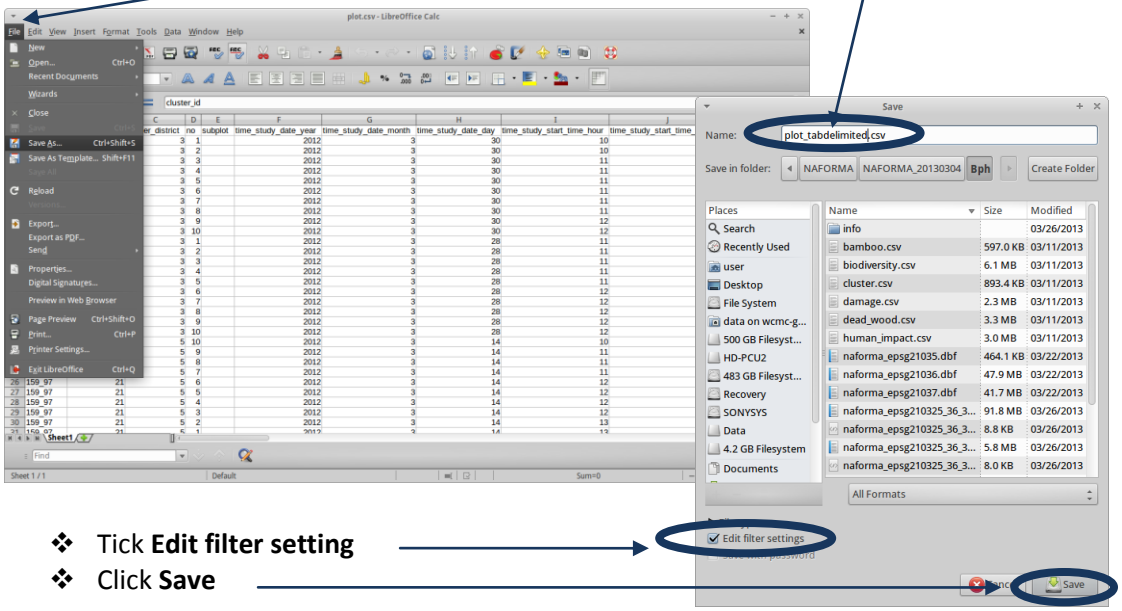
   - Click on **Tanzania_crty.shp**
   - Click **Open**

The NAFORMA 'plot' data are in CSV format which means the data fields are separated by commas. Data <u>within</u> some of the fields also have commas. Spreadsheet software such as LibreOffice Calc distinguishes between the field delimiters and the commas within the text strings as the text strings are enclosed by "". QGIS however does not see the "" and will read the data incorrectly. This can be resolved by opening the data in LibreOfficeCalc and saving the CSV file as tab-delimited rather than comma delimited.

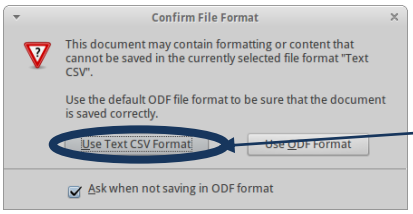
4. Open **plot.csv** in a **LibreOffice Calc** (an open source equivalent to MicroSoft Excel). Keep the **default** import settings



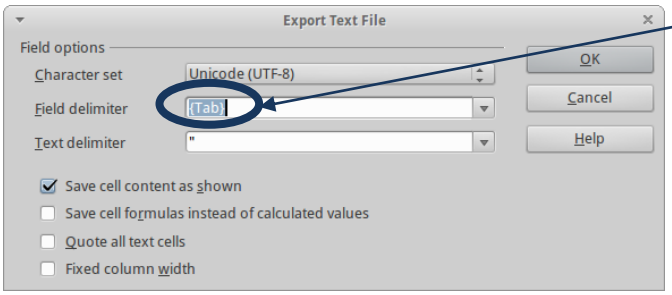
- From the main menu **Click File>>Save As** and save as **plot_tabdelimited.csv**



- Tick **Edit filter setting**
- Click **Save**



- Click **Use Text CSV Format**

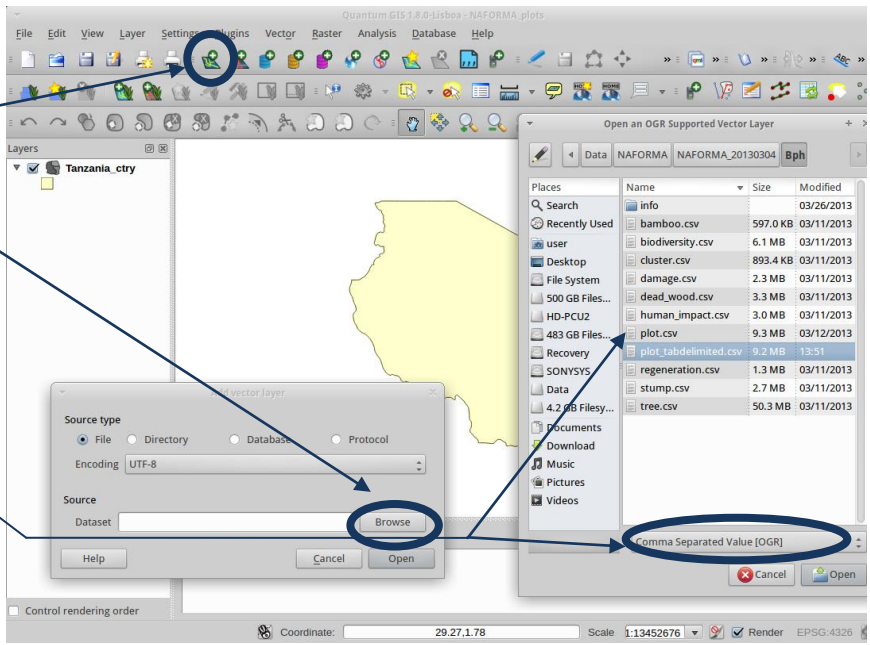


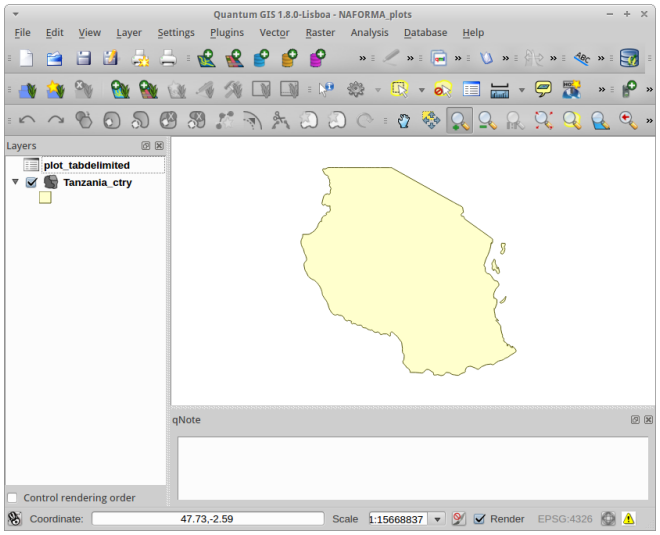
- Change **Field delimiter to {Tab}**
- Click **OK**

- Close LibreOffice Calc

5. Return to the QGIS project

- **Add vector layer**
- Browse to the folder containing the file
- Change file type to **Comma Separated Value (OGR)** and Click on **plot_tabdelimited.csv**
- Click **Open**. Then Click **Open** again

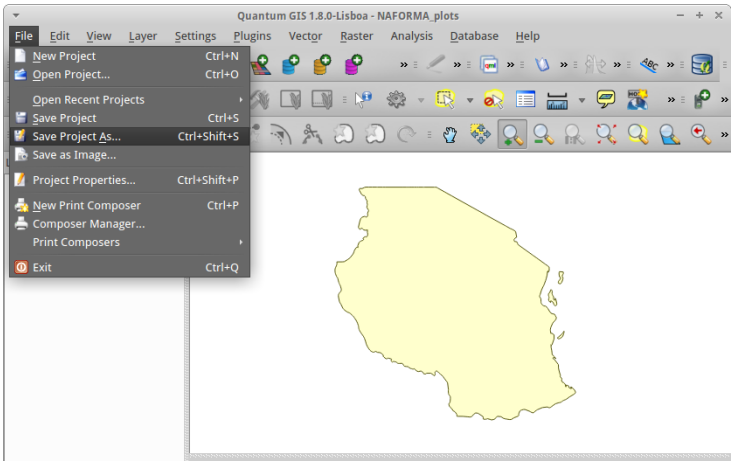




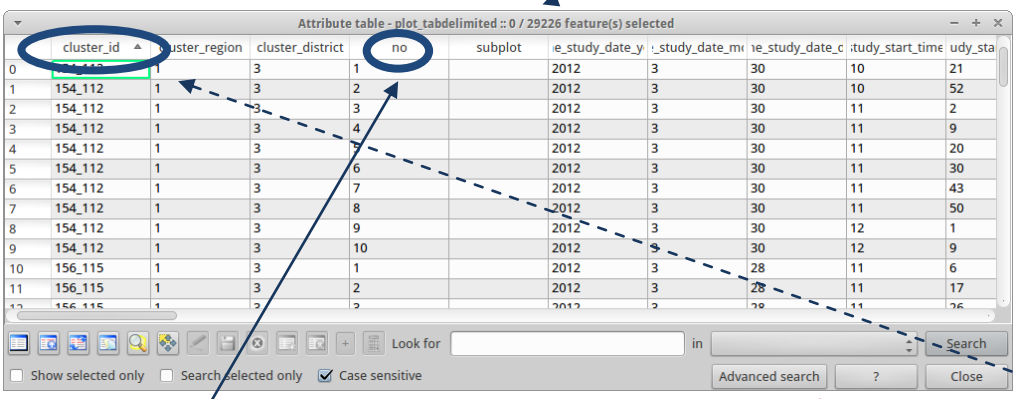
The QGIS TOC (Table of Contents) now contains 'plot_tabdelimited' and Tanzania_ctry'

Note: layers do not display the file endings e.g. .shp or .csv. Hover over layer name to see the full path, filename and file extension. The plot data do not display on the map because the data have been added to QGIS as a table.

6. Save the project. From the main menu **Click File>>Save Project as**… and navigate to the folder into which to project will be saved.
(In the working session save in the folder called Activity_ExploringNAFORMAPlotsAsPoints)

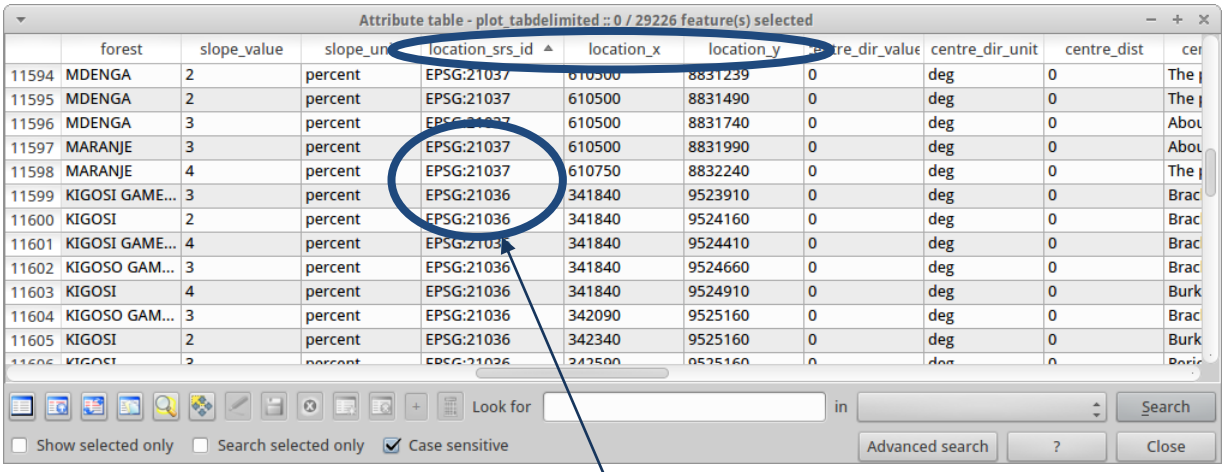


7. Become familiar with the **plot_tabdelimited** layer by opening the attribute table. **Right click >>Open attribute table**
The number of rows in the NAFORMA plots dataset (29226) is listed here. In this case there is one record (row) per inventory plot.



*Note that every row in this table represents one plot. Each plot is identified by a number '**Cluster_id**' and a plot number '**No**'.*

*Scroll to the right and look at the field names. Note that there are 3 fields containing the location information of the plots. The **location_srs_id** field contains the CRS that the X Y location coordinates were taken in and these are stored in the **location_x** and **location_y** fields.*



Note that Tanzania covers 3 UTM Zones. Scroll down the table and see that three different UTM zones are present in the data, EPSG:21035, EPSG:21036, and EPSG:21037.

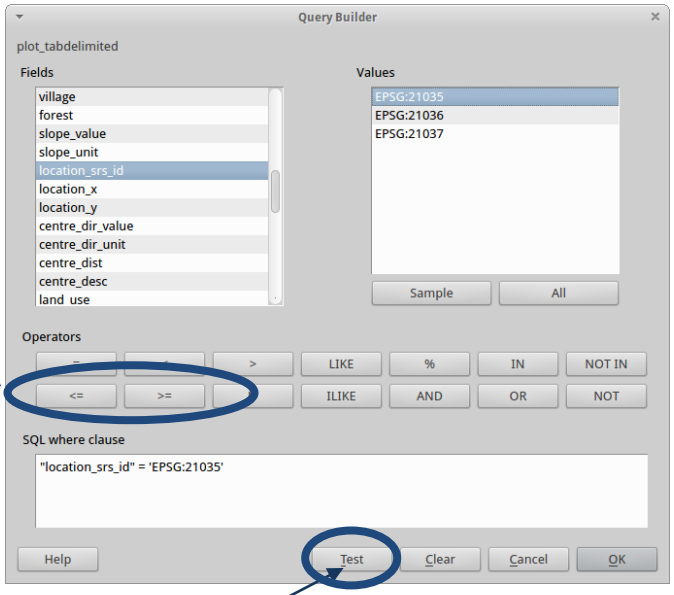
- ❖ Close the attribute table by clicking on the cross in the top right hand corner.

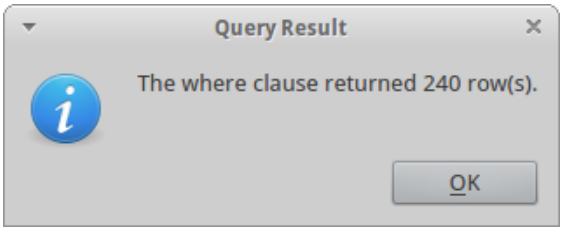
8. **The next step will be to create three separate CSV files containing the plots locations in each of the three UTM zones.**

Click on the **plot_tabdelimited** layer to select it and from the main menu bar **Click Layer >> Query**. This will bring up a new window called **Search Query Builder**

Filter the records so only rows with UTM Zone 35 are visible i.e. **location_srs_id = 'EPSG:21035'.**

*(To make the 'SQL where clause' **type** it or **double click** on the **location_srs_id** in the fields list. Single click on the "=" sign in operators list and **double click** on the value to select)*



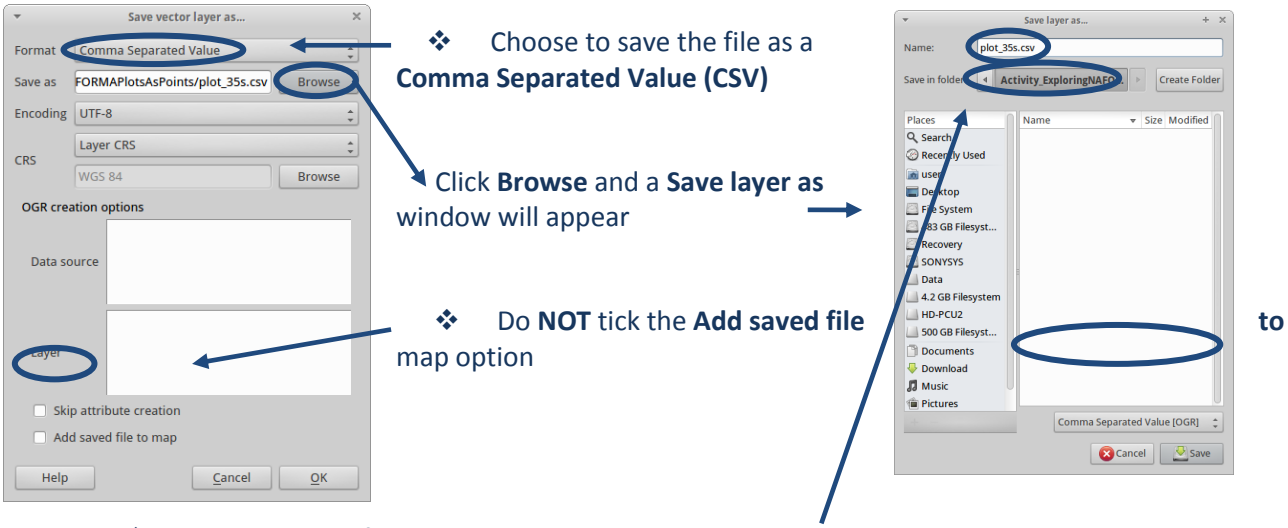


- ❖ Click **Test** to check the syntax of the query. It will return the number of rows if the syntax Is correct
- ❖ Click **OK**

Open the attribute table of the 'plot_tabdelimited' layer and see that there are only rows in UTM Zone 35 (EPSG:21035). Close the attribute table by clicking the cross in the right hand corner.

9. **Save the 'plot_tabdelimited' layer to a new CSV file for the filtered records only i.e. in this case saving only those records in Zone 35 (EPSG:21035).**

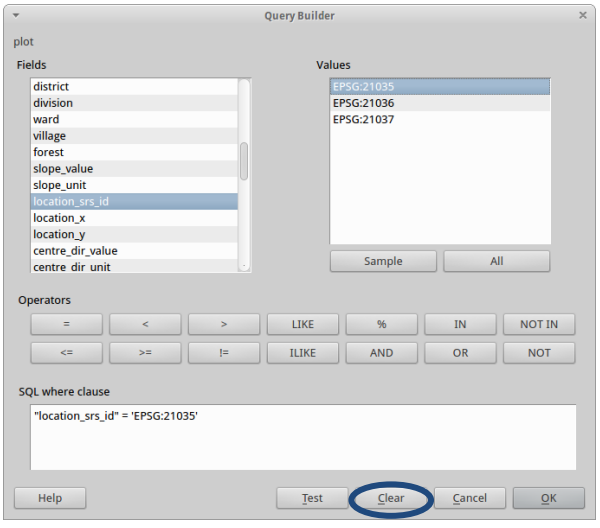
   - **Right Click** on the plot layer and **Click Save as…** . Ignore the layer CRS in this instance as we are just saving the file as a CSV (text file) which is not able to store the CRS information



   - Navigate to the folder to save the data in and save as **plot_35s.csv**
   - **Click Save** and **OK** to finish

(In the working session save in the folder called Activity_ExploringNAFORMAPlotsAsPoints)

10. **Repeat steps 9 – 10 for zones 36 and then again for 37, creating two more CSV files called plot_36s and plot_37s respectively.**



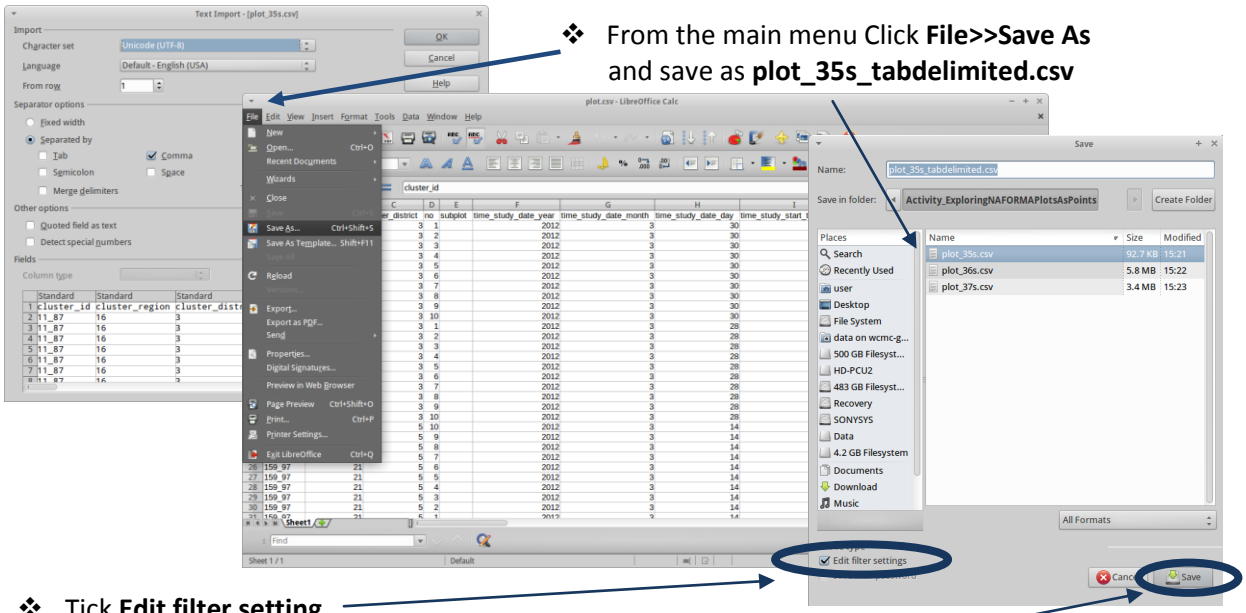
11. **Remove the query set on the 'plot_tabdelimited' layer:**

Click on **plot_tabdelimited** to select it. From the main menu bar **Click Layer >> Query** to bring up the **Search Query builder**

   - Click **Clear** and then click **OK**

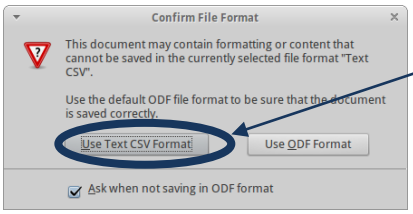
The new CSV files are unfortunately now comma delimited as QGIS does not save the data as TAB delimited so it is necessary to open the three new CSV files in LibreOffice and save them again.

12. Open **LibreOffice Calc** and open the **plot35s.csv** with the default settings



❖ From the main menu Click **File>>Save As** and save as **plot_35s_tabdelimited.csv**

❖ Tick **Edit filter setting**

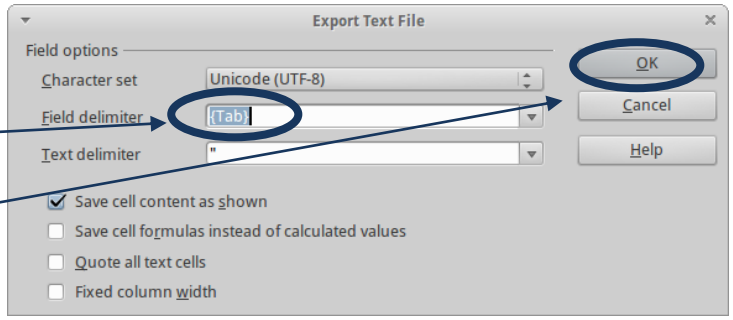


❖ Click **Save**

❖ Click **Use Text CSV Format**

❖ Change **Field delimiter to {Tab}**

❖ Click **OK**



13. Repeat Step 13 for **plot35s.csv** and **plot35s.csv**

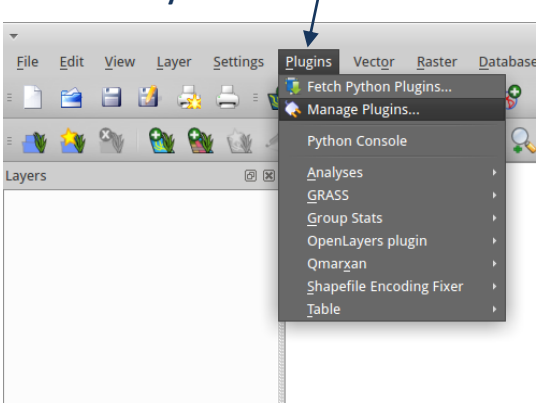
14. Once all 3 files have been saved as tab delimited **close LibreOffice Calc**

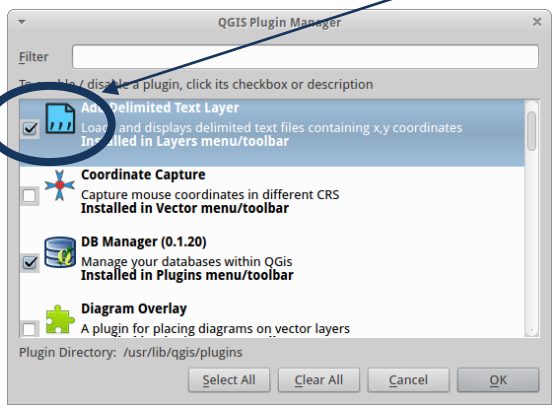
15. **Return to the QGIS** project

*****To summarize, the NAFORMA plot data have been added to QGIS and filtered by UTM Zone. The data have been split into three new CSV files so that each CSV file contains only plot records for their respective UTM zone*****

Now that the data are saved in separate files for each UTM Zone and are tab delimited they can be added to the QGIS project BUT this time the data can be added and presented spatially (not just as a table) by using a tool which tells QGIS which fields contain the x and y location information and what CRS the coordinates are stored in.

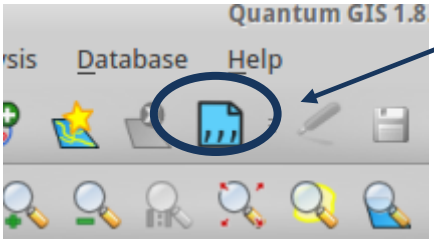
- Click **Plugins >> Manage plugins**. Check the box next to the plugin called **Add delimited text layer**. Click **OK**



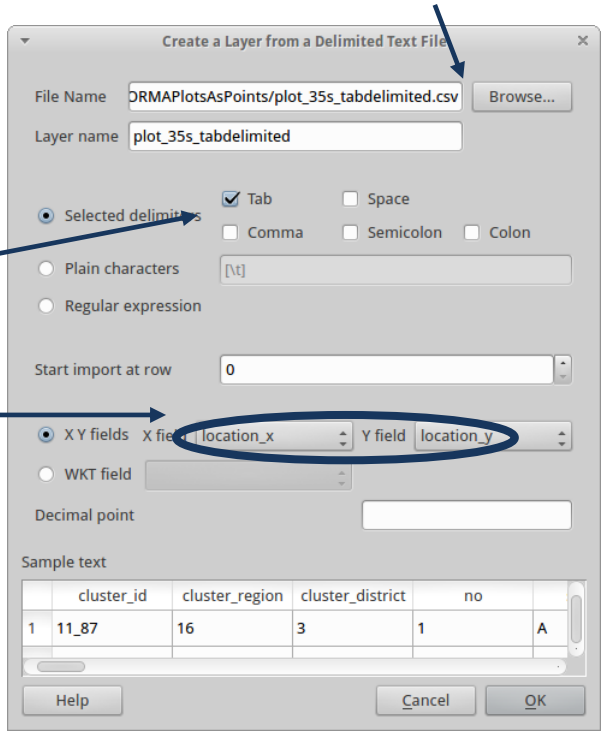


Note: if this plugin is not visible first go to 'Fetch Python plugins' and tick to install the 'Add Delimited Text Layer' plugin (an internet connection is required for this step).

- The cyan **Add delimited Text** button is added to the menu bar. This plugin allows QGIS to display the CSV file spatially



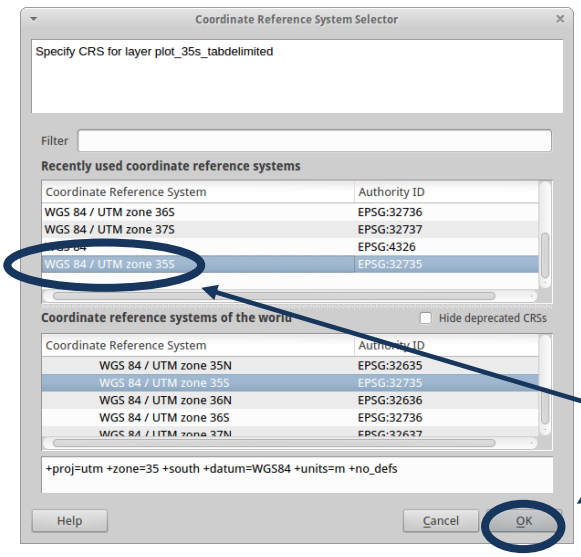
- Click on the **Add delimited Text button**
- Click **Browse** to navigate to the folder containing **plot_35s_tabdelimited.csv** . Click on the CSV file.



- Select **delimiters** and **Tab**.

- Set **x** field to **location_x** and in **y** field to **location_y**

(QGIS will try to guess the x y fields so make sure correct fields are set)



- Click **OK**

The "Coordinate Reference System Selector" window opens asking what CRS (projection) the data are stored in.

- Click on **WGS 84 / UTM zone 35s** for the plot_**35s**_tabdelimited layer
- Click **OK**

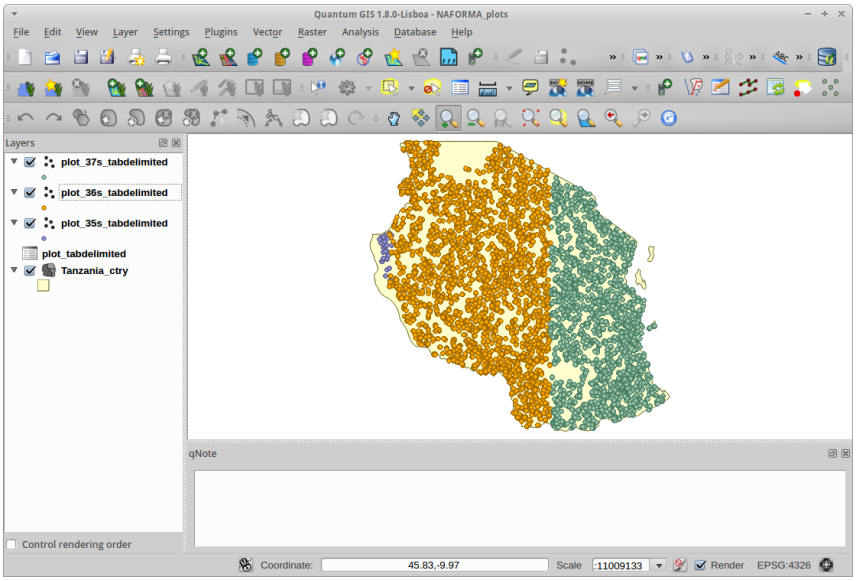
QGIS has to be told what projection the data are in because the data are coming from a CSV text file and text files do not store this information in format QGIS can read.

IMPORTANT If the "Coordinate Reference System Selector" window did not appear when adding the data, **right click** on the added layer e.g. plot_35s_layer and **Click "Set Layer CRS"**. The CRS window will then open and the CRS can be selected as in the step above. This will only happen if Step 2 was missed where the QGIS project options were set.

16. **Repeat step 18 for plot_36s_tabdelimited.csv and plot_37s_tabdelimited.csv**. Be careful to select the correct UTM zone when asked to specify the CRS.

    - **Right Click** on **Tanzania_ctry** and **Zoom to Layer Extent**

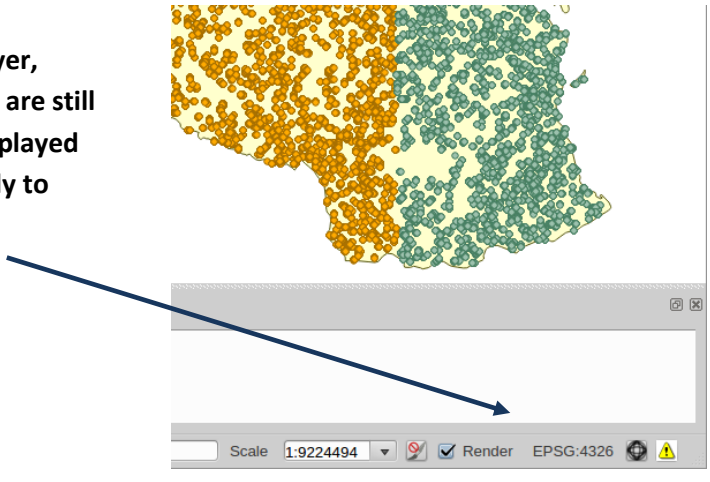
The QGIS map window should now display the Tanzania country boundary and plot_35s_tabdelimited, plot_36s_tabdelimited and plot_37s_tabdelimited as mapped points. It should look similar to the map below.



17. Check for any errors (e.g. points falling beyond the country boundary). **Right Click** on each of the plot layers and **Click Zoom to Layer Extent**. Do not try to fix the errors at this stage.

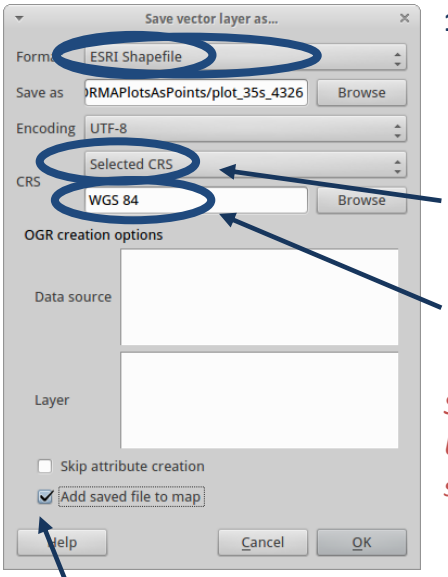
Before any errors can be fixed the CSV file needs to be converted to a global CRS such as WGS84 "EPSG:4326" as QGIS cannot select or identify the points falling outside of the UTM zone.

- **At this stage the layers plot_35s_layer, plot_36s_layer and plot_37s_layer are still CSV files which are simply being displayed spatially in QGIS projected on-the-fly to EPSG:4326 (the CRS of the project)**



The next step is to save the CSV layers as spatial data (e.g. in shapefile format). So that the data can be stored spatially in a single file the data need to be projected from their individual UTM projections into EPSG:4326

*In QGIS projecting is done from the **File>>Save Layer as** menu.*



18. **Right click** on each of the plot layers, and save them as shape files, naming them **plot_35s_4326, plot_36s_4326 and plot_37s_4326** respectively

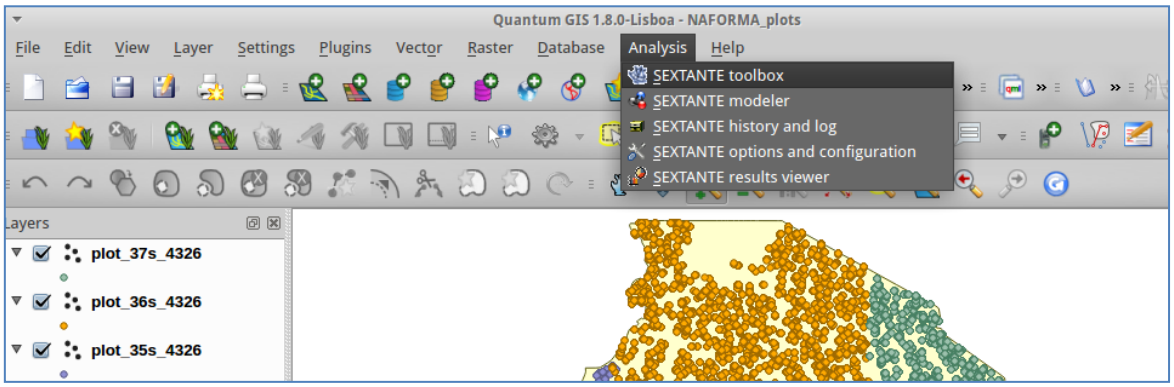
- Change to save as **Selected CRS**
- Click **Browse** and change to **WGS84 EPSG:4326**

So instead of saving as a shapefile with the current Layer CRS (i.e. UTM Zones35, 36 or 37) the shapefiles will all be projected to the same projection (EPSG:4326).

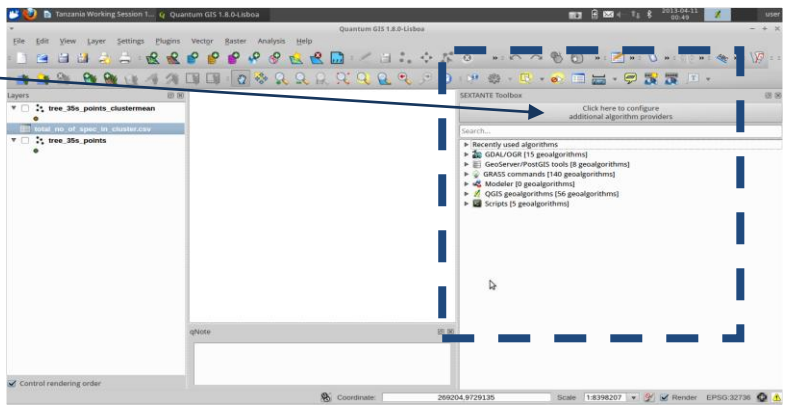
- Each time **Tick Add saved file to map**

Now that the data are in a spatial format (i.e. shapefiles) and are in the same CRS they can be merged together into a single shapefile. The 3 shapefiles have the same structure (i.e. identical fields) so the 'SAGA Merge tool in the Sextante Toolbox' can be used to merge the shapefiles into a single shapefile.

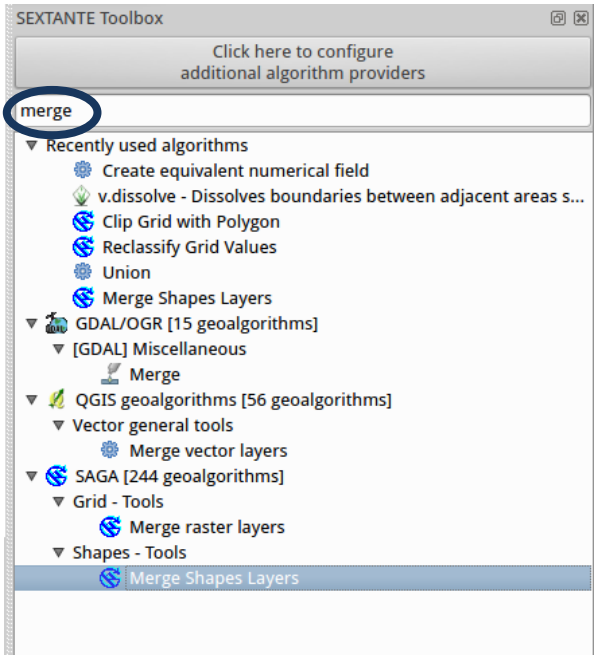
19. Click on the **Analysis>>SEXTANTE toolbox** to open the geoprocessing toolbox



The Sextante toolbox will appear. Dock it in the right hand side of the QGIS interface.



20. In the **Sextante Toolbox** type **'merge'** to help locate tools that could be used for merging the datasets

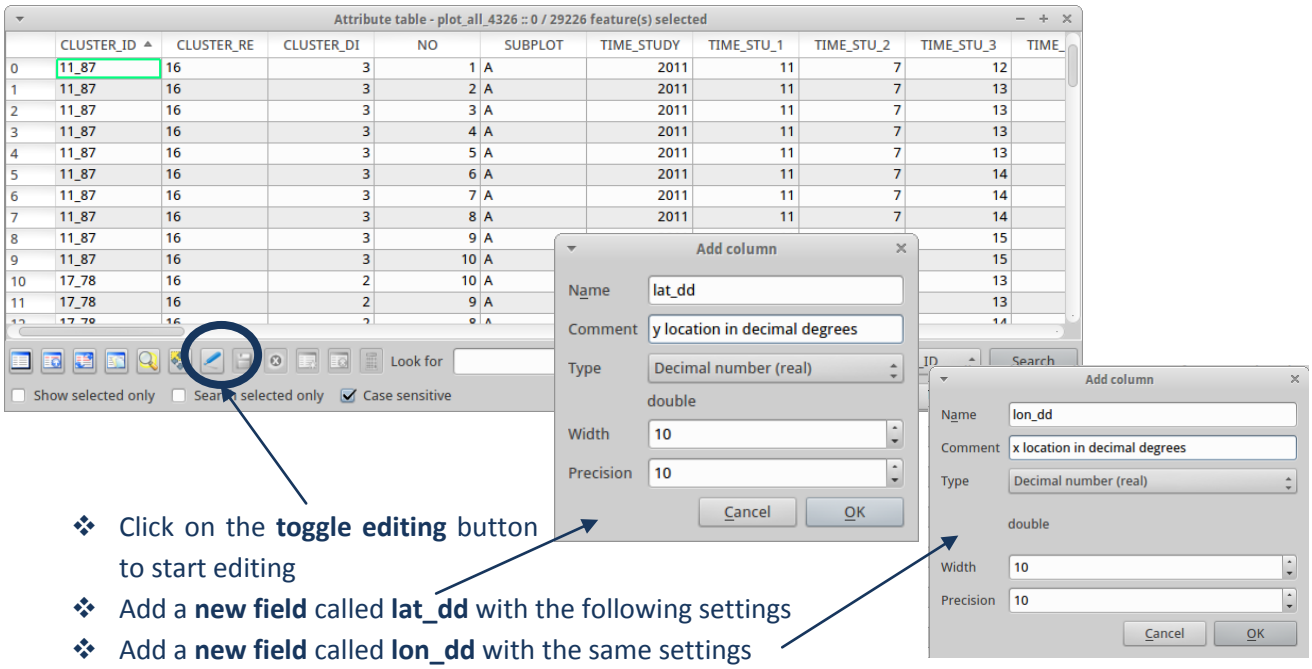


- **The errors can now be identified in the 'plots_all_4326.shp' dataset**

NOTE: If plots fall outside the Tanzania country boundary it may be possible to fix them. It is an obvious error if, for example, the x y locations are entered in the wrong way round. Otherwise the problem records should be deleted. Keep a record of any amendments or deletions.

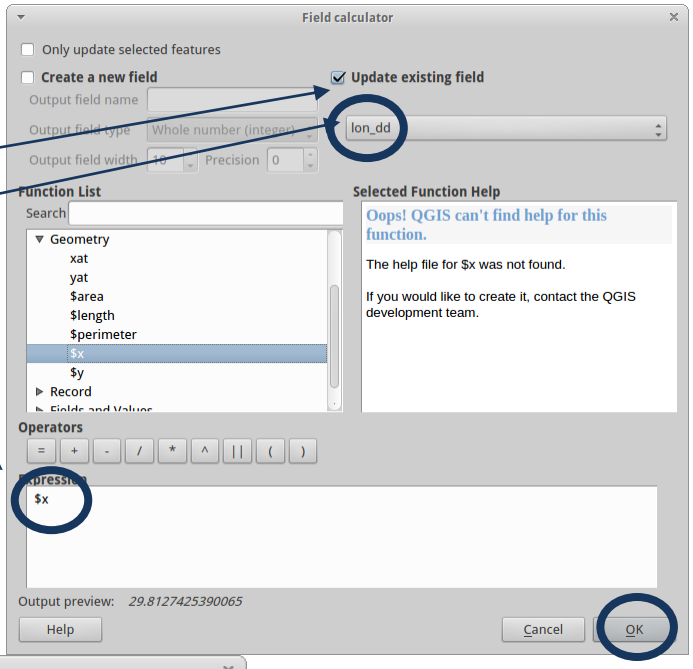
Click on the layer containing the errors in the TOC and use the identify tool to identify those points outside the country boundary. Write down the 'plot_no' and' cluster_id' of the problem records. Remove the CSV file from your QGIS project and open it in LibreOffice Calc. Go to each of records containing the location errors and either correct the locations (if known) or delete the records from the table. Document any corrections/deletions made and save the file with a new name, e.g. if you have made changes to tree_35s.csv give it a new name such as tree_35s_correct.csv. Go back to your QGIS project and repeat step 15, this time adding the corrected file tree_35s_correct.csv.

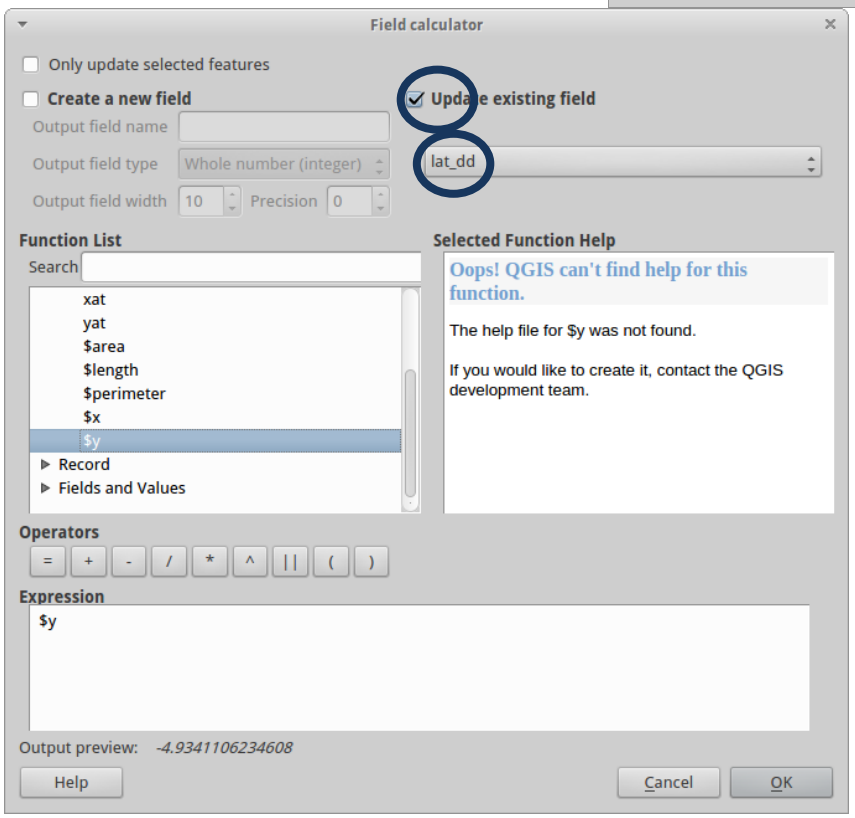
21. Right Click on **plot_all_4326** and **open the attribute table**



- Click on the **toggle editing** button to start editing
- Add a **new field** called **lat_dd** with the following settings
- Add a **new field** called **lon_dd** with the same settings

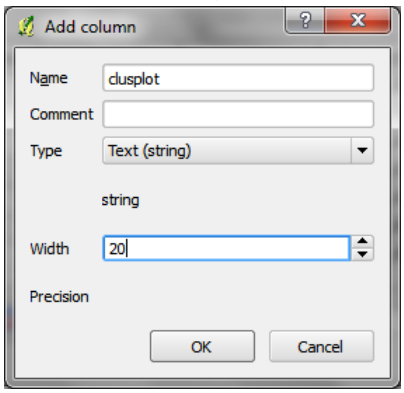
- Click the **Save** button and then click the **calculator** button
- In the **field calculator** window **Tick update existing field**
- Select **lon_dd** (as the field to calculate)
- Expand **Geometry**
- **Double click $x** to bring it down into the In the **expressions box**
- Click **OK**
- Repeat for **Lat_dd** but this time **Double Click $y**

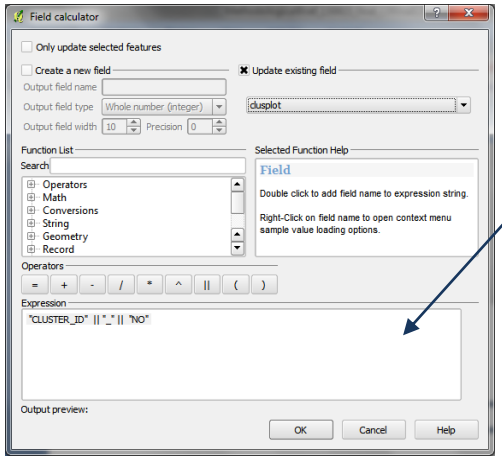




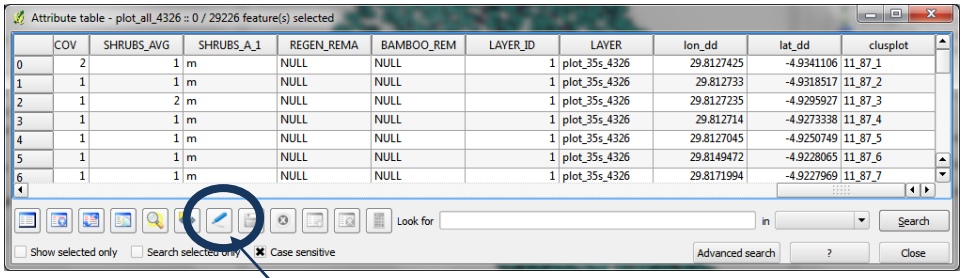
See that the attribute table now has two columns populated with longitude - latitude coordinates in decimal degrees.

- Click the **Save** button to save the edits
- **Add another field** called **clusplot** and make it type **'text'**





- Click the **Field calculator** button and calculate the field **"CLUSTER_ID" || "_" || "NO"**



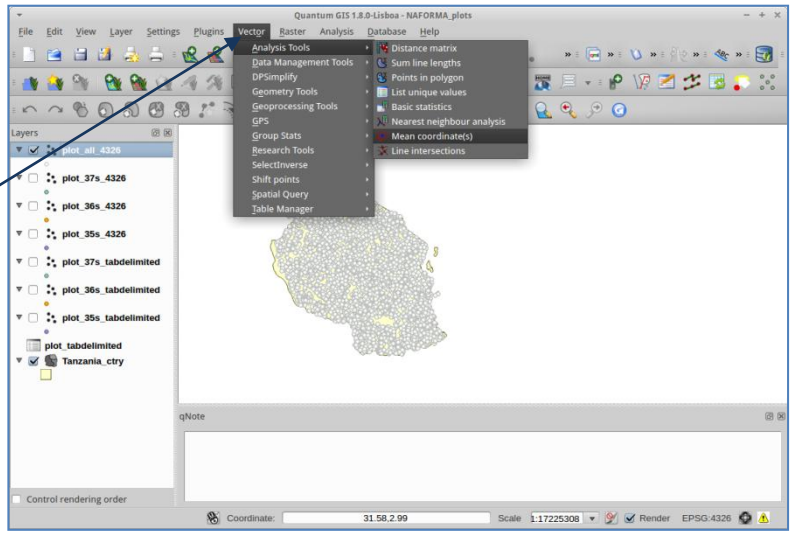
❖ Click the **toggle editing** button to **stop and save edits**

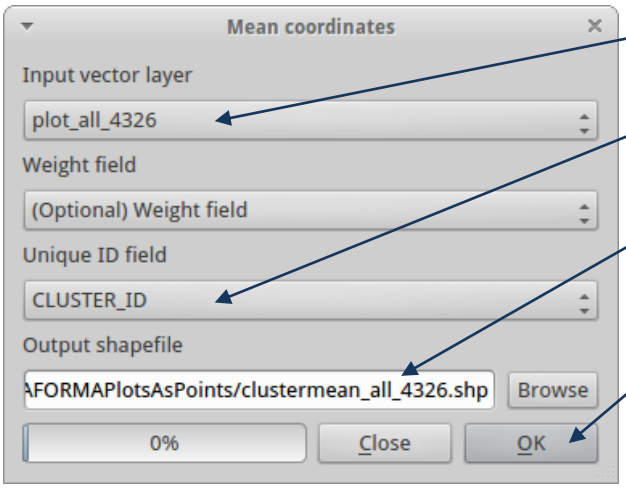
The dataset now has a unique ID (clusplot) that will be used to join the new lon_dd and lat_dd fields onto the tree data.

The next step is to create mean coordinates from the plot data for each cluster so that average values can be mapped at the cluster level. Remember that the NAFORMA inventory data have up to 10 plots in each cluster.

22. Remove all layers from the QGIS project except the 'Tanzania_ctry' and the 'plot_all_4326' layers.

❖ From the main menu Click **Vector >> Analysis tools >> Mean coordinates**





❖ Set **plot_all_4326** as **Input vector layer**
❖ Leave Weight field as it is
❖ Set **Unique ID** field as **cluster_id**

❖ Set Output shape file to **clustermean_all_4326.shp**

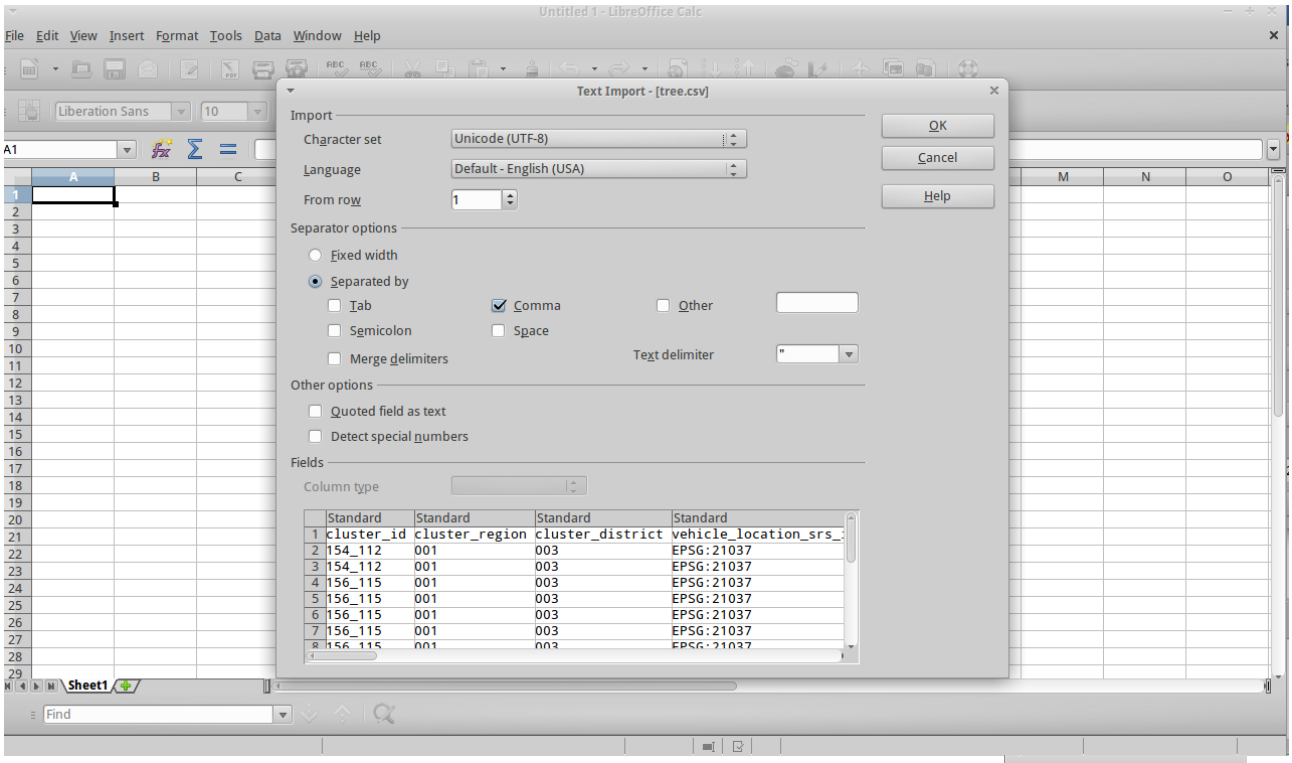
❖ Click **OK**

❖ Click **Yes** to add the **new layer to the TOC**

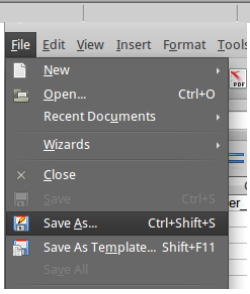
The next stage is to format and add the tree.csv data to the QGIS project. To save time processing the tree data three time for the three UTM Zones, the plot data that has just been prepared will be used to add latitude longitude coordinates in decimal degrees (geographic coordinate system EPSG: 4326) to the tree csv file so that it can added spatially to QGIS as one file.

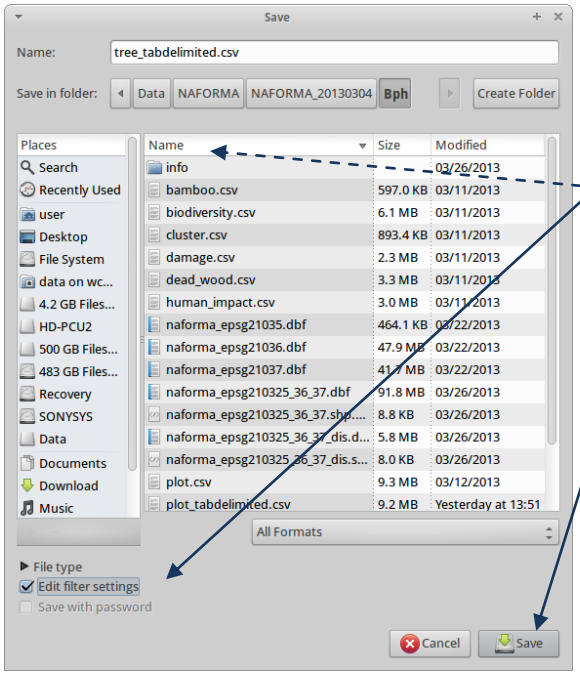
23. Open **the tree.csv** file in **Libre Office Calc** using the default Comma delimited and text delimeter options

- **Add 3 extra columns** at the end of the file. Clusplot, lon_dd and lat_dd
- Populate **lat_dd** and **lon_dd** with the number **0**



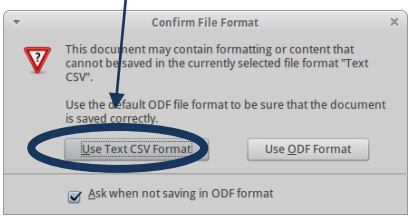
24. Save the File as



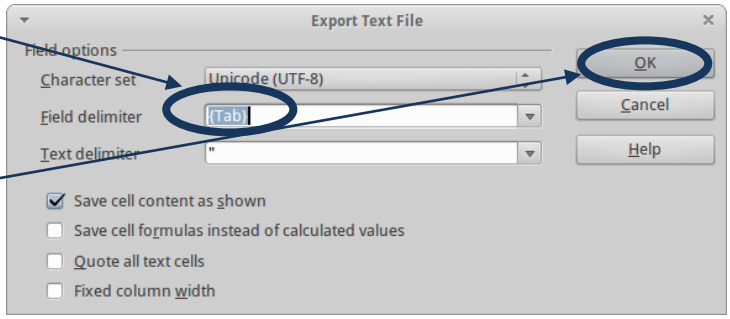


- Tick **edit filter settings**
- Name it **tree_tabdelimited.csv**

- Click **Save**
- Click Use **Text CSV Format**



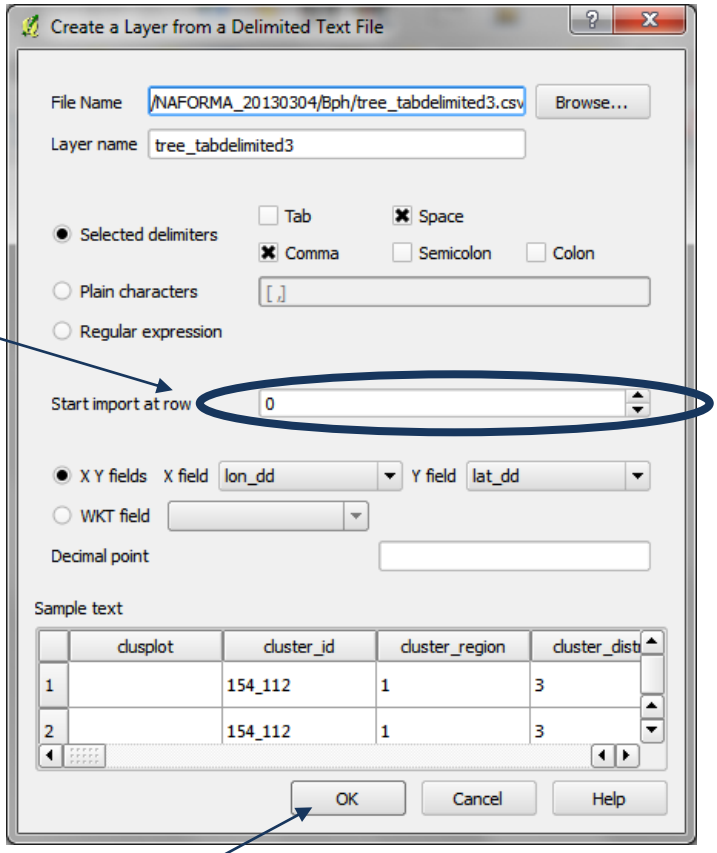
- Change **Field delimiter to {Tab}**
- Click **OK**



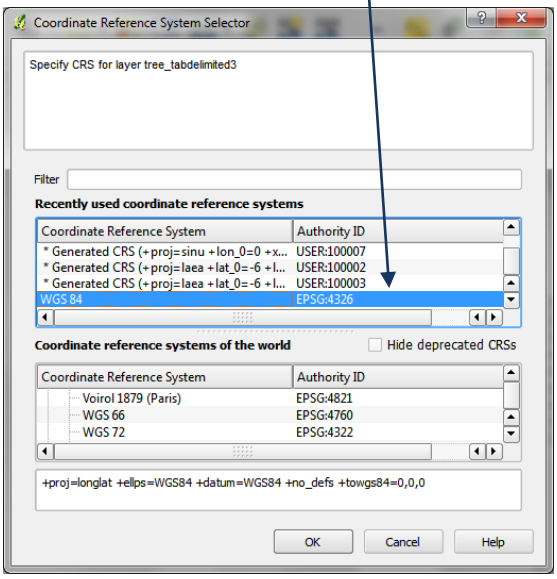
25. Return to QGIS

- Using the **Add delimited text** button add the **tree_tabdelimited.csv** file to the QGIS project
- Set the **X Y** fields at **lon_dd** and **lat_dd**

At this stage remember the values are all 0 so the points will all be mapped at 0,0



- Chose the CRS EPSG:4326

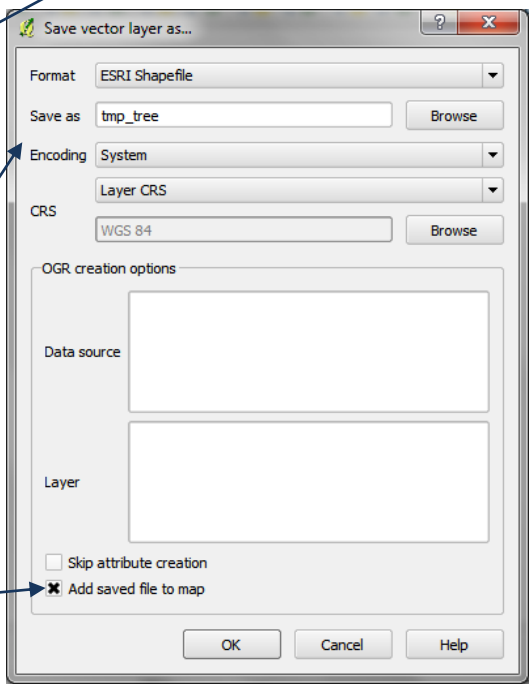


- Click **OK**

- **Right Click** on the added point visualization of tree_tabdelimited.csv and **save as tmp_tree.shp**

This step is necessary as a join is not possible to a CSV file, even if it is added to QGIS as delimited text.

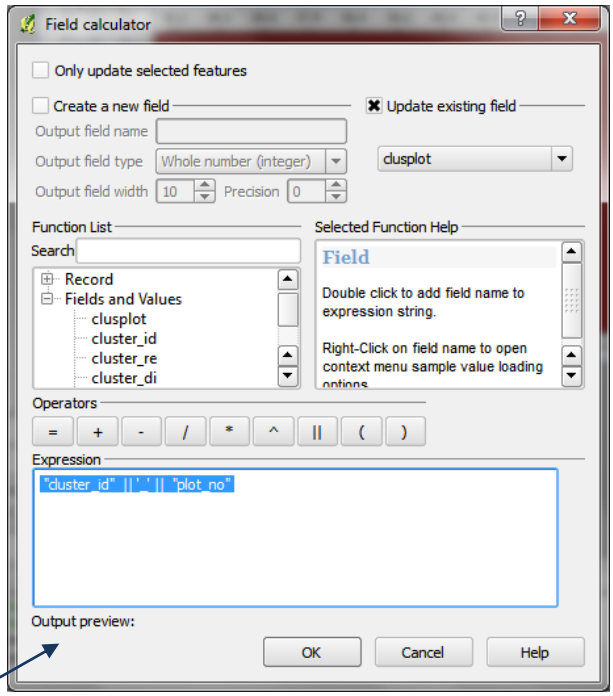
- Tick **Add saved file to map**



- Then **toggle editing on** and **calculate** the **clusplot** field:

  "cluster_id" || '_' || "plot_no"

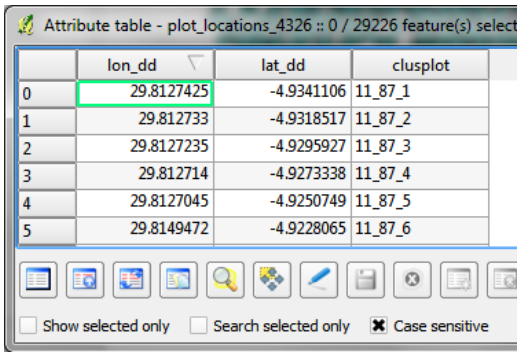
- **Toggle editing off** and **save edits**



26. Next create a small subset of the plot data which only has three fields (this will make the join with the tree data easier and less confusing)

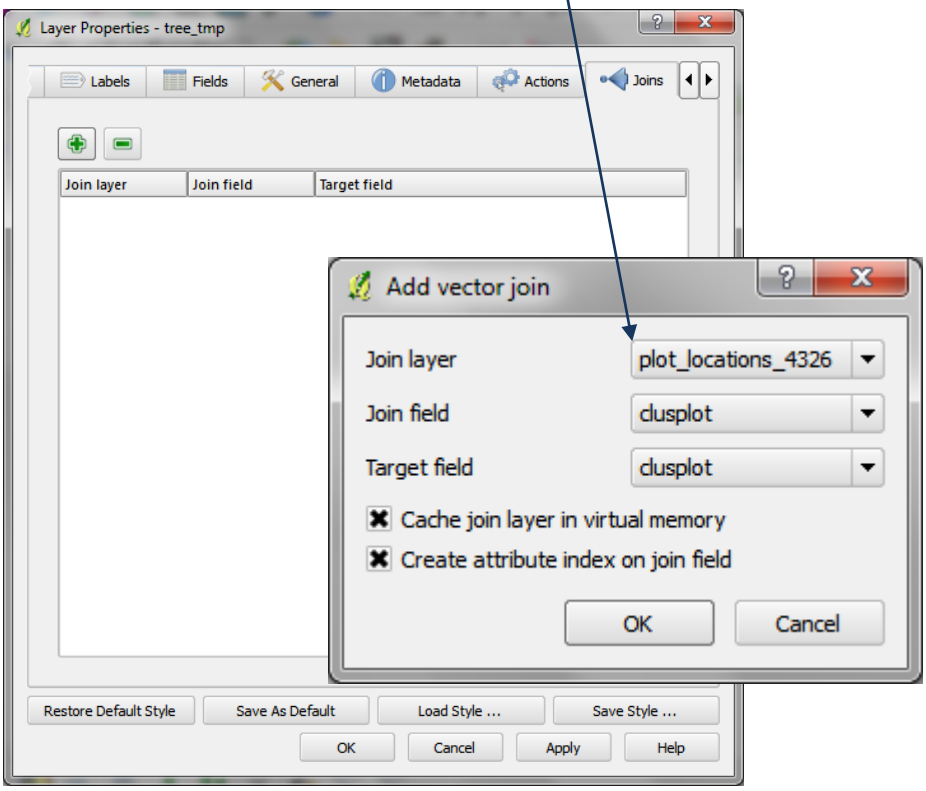
- **Right Click** on **plot_all_4326.shp** and save as a **new shapefile** plot_locations_4326.shp and **add to map**.

- Click on **plot_locations_4326.shp** and delete all fields keeping only clusplot lon_dd and lat_dd

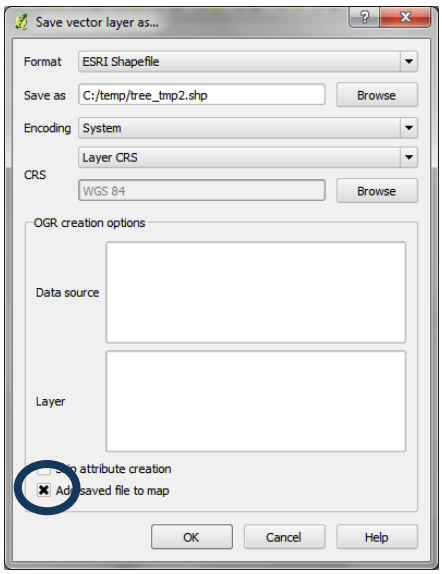


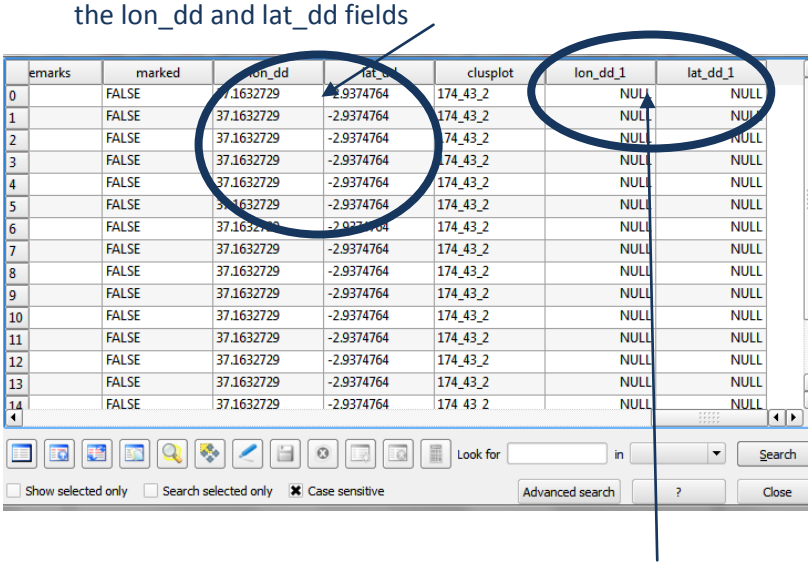
Now join the plot_locations_4326 onto tree_tmp.shp

- **Right click** on tree_tmp and click **Properties**
- Click on the **join tab** and click **'+'** to add a join



- **Right Click** on tree_tmp.shp and **Save As… tree_tmp2.shp**
- **Tick** to add new dataset to map
- Open the attribute table and notice that because the lon_dd and lat_dd fields were present in both the original datasets, the information from the join fields has moved across into the lon_dd and lat_dd fields

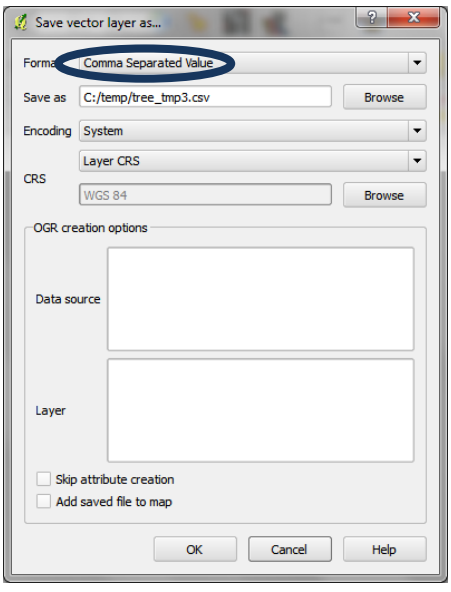




- Unfortunately it has also left 2 empty fields lon_dd_1 and lat_dd_1 field which have to be deleted manually
- Close the attribute table

27. The data in the dataset will still be displaying at 0,0 lon lat as all that has been changed are information in the attribute table. The dataset have to be saved back to CSV format in order to create a shapefile with the new lon lat information.

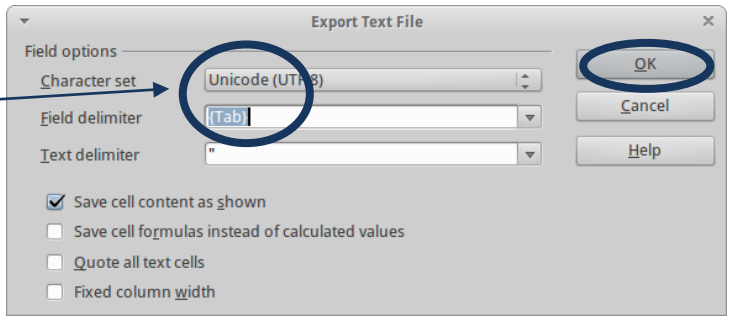
- Save the dataset as a new temporary CSV file



As QGIS only saves as comma delimited and these data have commas within fields, as well as being used to separate the fields it is necessary to open and save the data in LibreOffice Calc before using the Add delimited Text button, otherwise the data will be incorrectly imported.

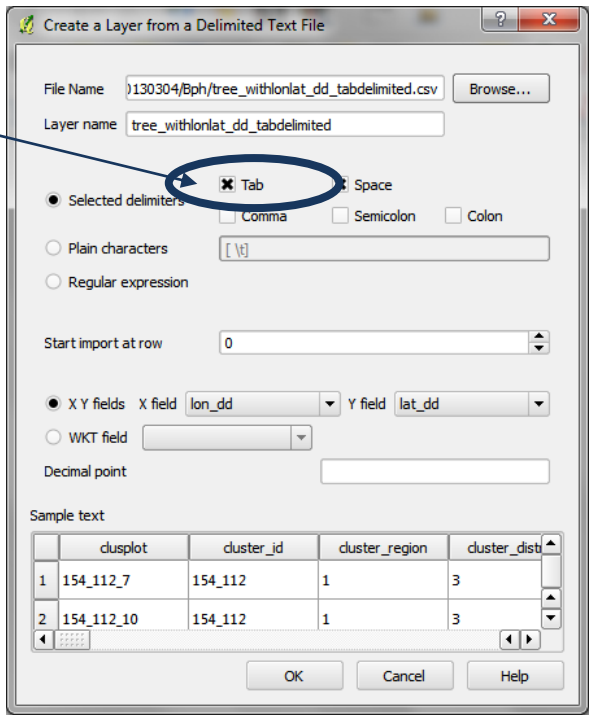
28. Open the tree_tmp3.csv file in Libre Office calc

- Open **LibreOffice Calc** and open the **tree_tmp3.csv** with the default settings
- From the main menu Click **File>>Save As** and save as '**tree_withlonlat_dd_tabdelimited.csv**
- Tick **Edit filter setting**
- Click **Save**
- Click **Use Text CSV Format**
- Change **Field delimiter to {Tab}**



- Click **OK**

29. **Return to QGIS**. Add the new CSV file with **Add Delimited Text**

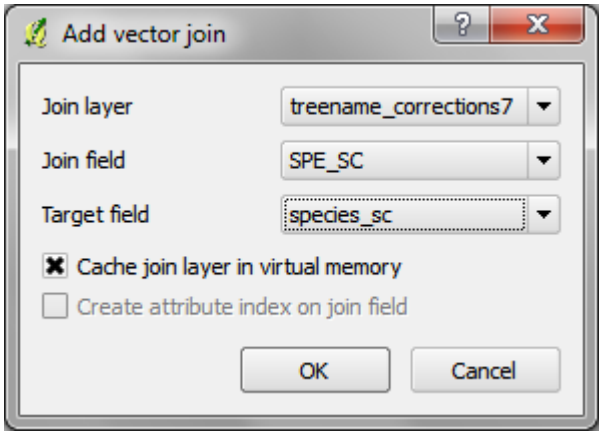


- Check the points, they should all fall correctly within Tanzania in the same locations as the plot_locations_4326.shp
- **Right Click** on the dataset and **save as tree_all_4326.shp**

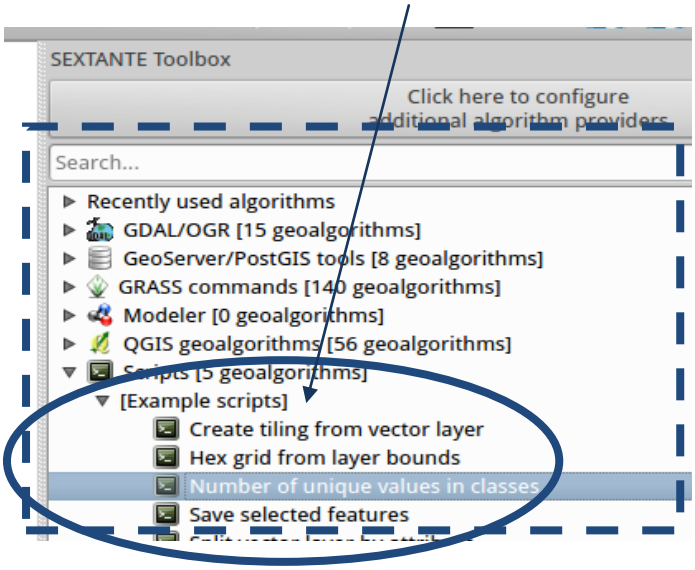
If the points and attribute table are all correct then the temporary files can be removed from the QGIS project and deleted.

30. **The next stage in this exercise is to analyze the tree data to the QGIS project and generate summary statistics of the tree data at the cluster level. The aim is to visualize mean tree density and mean tree species richness for each cluster.**

- Add the **corrected species name treename_corrections7.csv** into QGIS
- Join on the **corrected species names** to **tree_all_4326.shp**
- **Save** the dataset as **tree_all_4326j.shp** to make it a permanent join

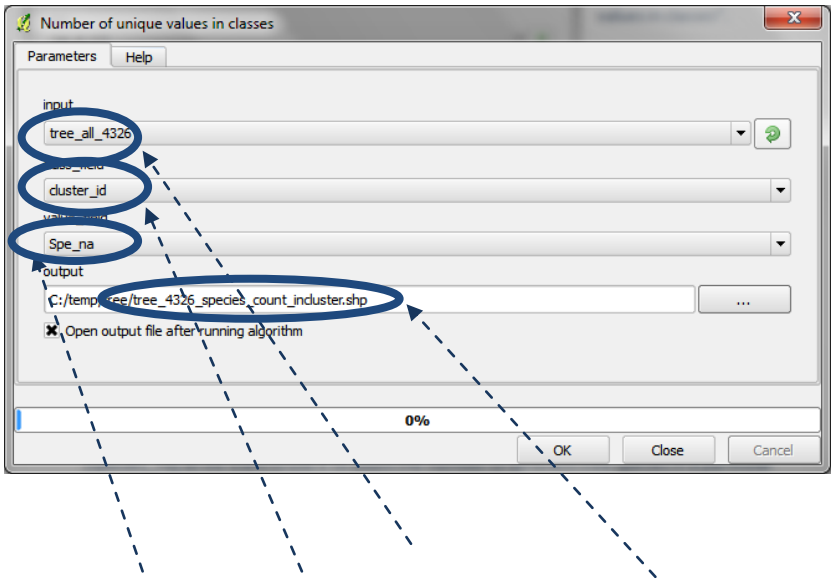


31. In the **SEXTANTE** toolbox Click on **Scripts** and expand **Example Scripts**



*You will see a script called **Number of unique values in classes**. This is the script we will use to calculate the mean tree density and mean tree species richness for the cluster*

Running a script works in the same way as other QGIS tools.



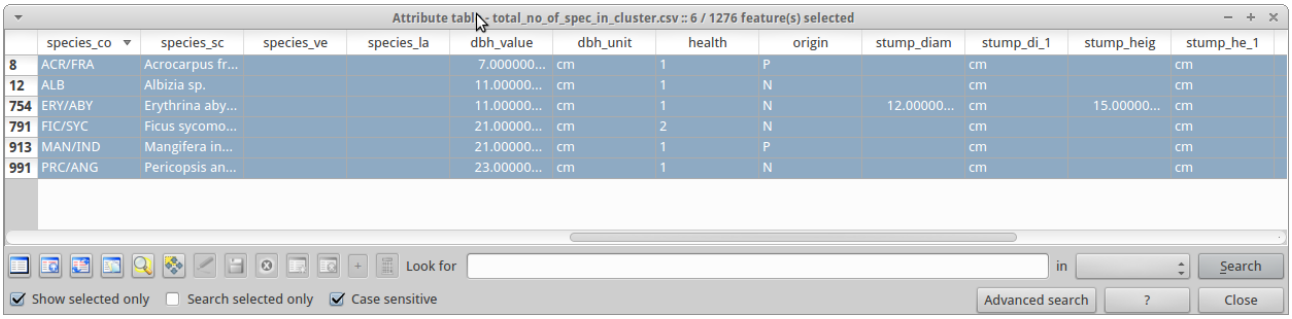
❖ **Double Click** on the **script** called **Number of unique values in classes**

We will run this for the tree species point data **tree_all_4326j**' and for each cluster_id summarise the number of species that are present in the cluster

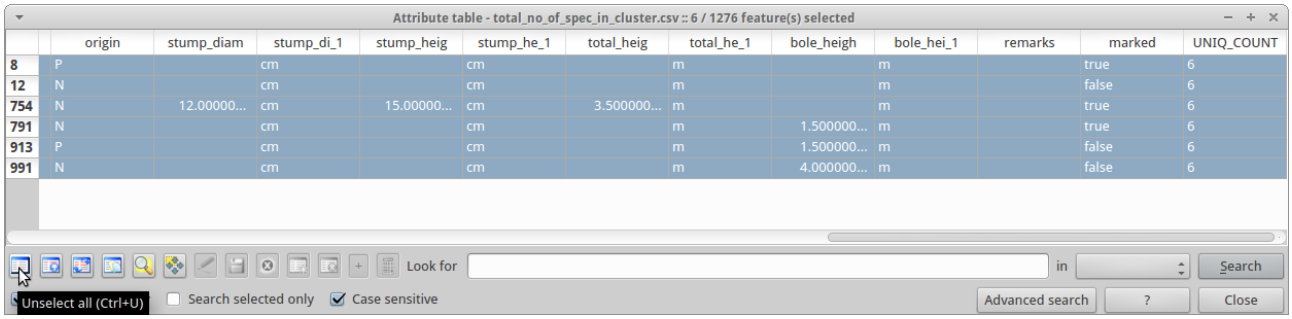
- ❖ So the **input** dataset is **tree_all_4326j.shp**
- ❖ The **class field** (or summary field) is **cluster_id**
- ❖ And the **Value field** (i.e. the field containing the data to summarise) is the **cleaned species name** (**species_na**) as we want to use it to count the number of different tree species in a particular cluster
- ❖ Give the output file a name **"tree_4326_speciescount_incluster.shp"**

This output is automatically added to your QGIS project.

32. Right click on **4326_speciescount_incluster.shp** and open the attribute table. The table show a unique list of species in each cluster as well as which species they are and how many there are in the cluster.



- ❖ Try selecting one cluster **17_178** and check to see if the number in the **UNIQ_COUNT** is the same as the number of species listed for that cluster



- ❖ Then click the **unselect all** button before closing the table

33. Calculate the total number of plots in a cluster

- Repeat the above process (steps 31 and 32) but this time summarize by cluster_id on the plot_no field

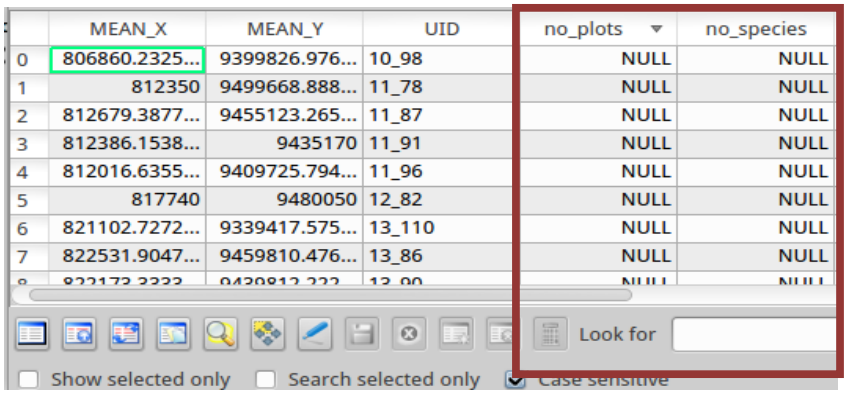
34. In the Cluster mean dataset, calculate the number of tree species per plot

Next, 2 new empty fields are added to the attribute table, into which the total number of species and total number of plots per cluster will be transferred. They will be calculated from the 2 shapefile files containing the unique counts created in the previous few steps.

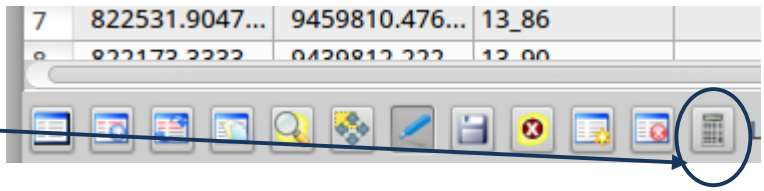
- Open the **attribute table** of **clustermean_all_4326.shp**
- Click on the **toggle editing mode button** to turn editing on
- Click on the new column button
- Add a new column as below: **no_plots** (whole number (integer) width 10)
- Add a second column **no_species** (whole number (integer) width 10)
- Then press the **toggle editing mode** button again to stop editing and save edits
- So now we have 2 new empty fields in **clustermean_all_4326.shp.**

35. To populate the no_plots field do a temporary join from the clustermean_all_4326.shp with the 4326_plotscount_incluster.shp

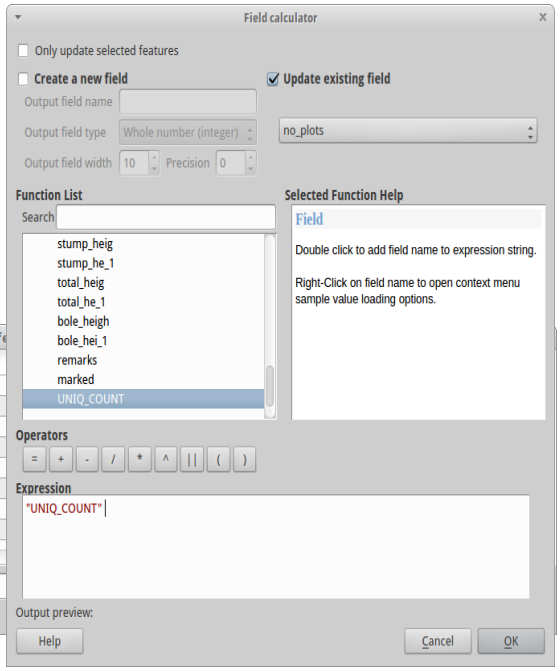
- Click on the **join tab** and then on the **'+'** button to add a join by **cluster_id**
- The **join layer** (i.e. the table you want to join to the mean cluster points) is **4326_plotcount_incluster.shp**
- The field you are going to use to do the join is **cluster_id**
- The **cluster id field** in the field in the **clustermean_all_4326.shp** dataset is **UID**
- Click OK and the **temporary join** will be made
- **Right Click** on **clustermean_all_4326.shp** and open the **attribute table**
- You will see that at the beginning of the table are the **2 new columns** we added containing **NULL (empty) values**.

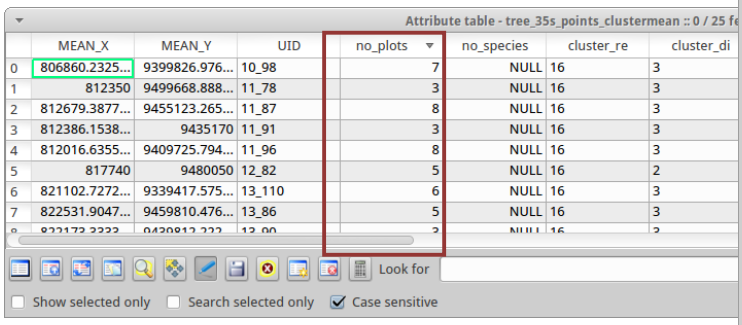


- The last field in the table shows **UNIQ_COUNT** (i.e. the no of plots in a cluster from the join file 4326_plotcount_incluster.shp)
- Click the **toggle edit mode** button
- Click on the **calculator** button which has becomes active

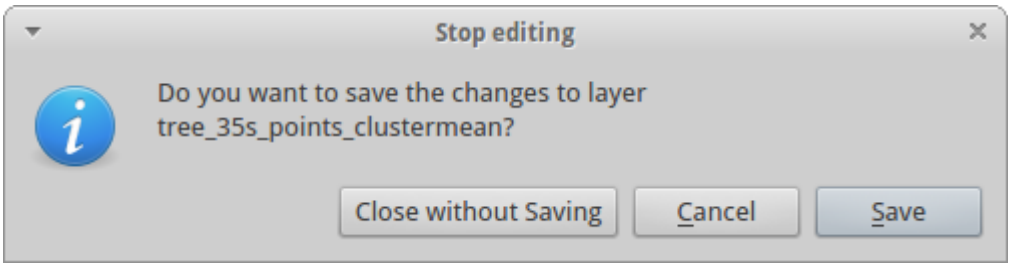


- Click on **Update existing field**
- Pick field **no_plots**
- From the function list double click on **UNIQ_COUNT**
- The expression will then be populated here
- Click **OK**
- The number of plots is now populated with the contents of the **UNIQ_COUNT** field



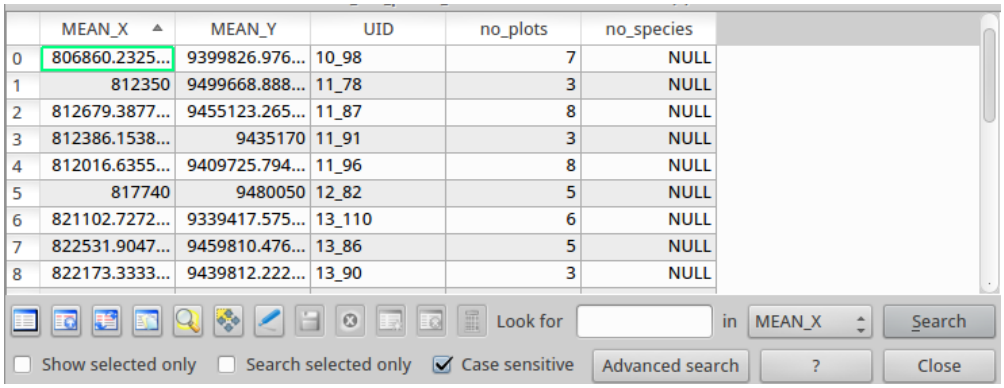


- Click the **toggle editing** button to **stop editing**



- Click **Save**
- Close the table
- Right Click on **clustermean_all_4326.shp** and click on the **join tab**

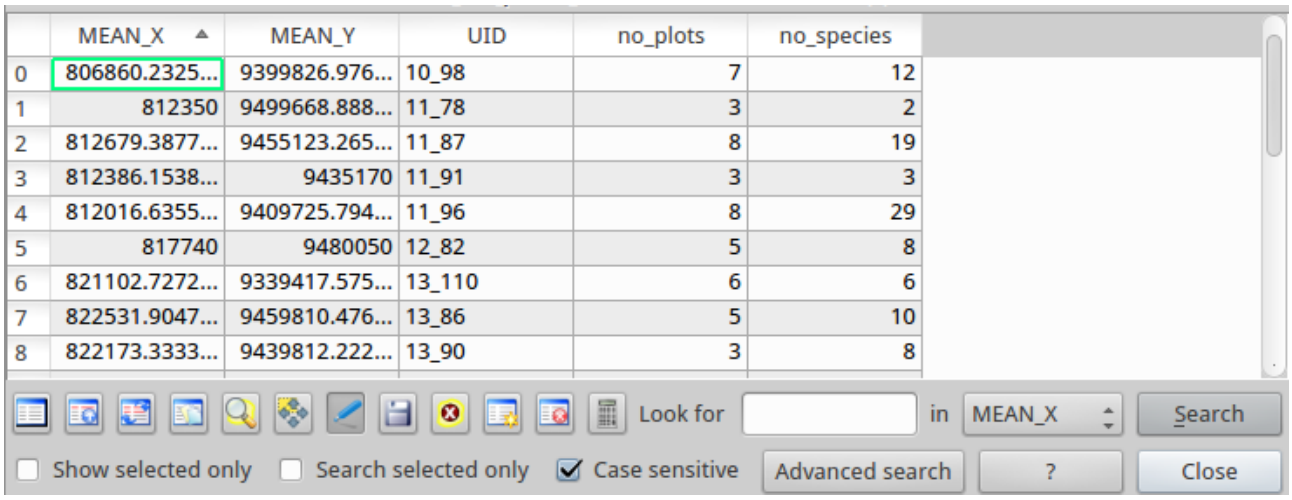
- Remove the current join by highlighting and clicking on the remove join button
- The join is now removed
- Click **Close**



- Open the attribute table **clustermean_all_4326.shp** and see that the **no_plots field** now has values

36. Repeat step 35 for the total number of species in a cluster

*At the end of this step the attribute table of **clustermean_all_4326.shp** should look similar to the below*



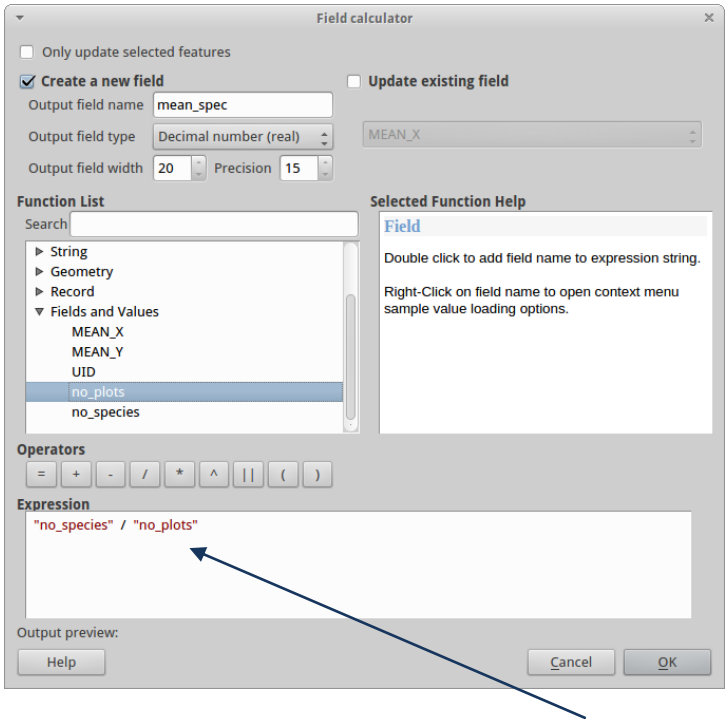
	MEAN_X	MEAN_Y	UID	no_plots	no_species
0	806860.2325...	9399826.976...	10_98	7	12
1	812350	9499668.888...	11_78	3	2
2	812679.3877...	9455123.265...	11_87	8	19
3	812386.1538...	9435170	11_91	3	3
4	812016.6355...	9409725.794...	11_96	8	29
5	817740	9480050	12_82	5	8
6	821102.7272...	9339417.575...	13_110	6	6
7	822531.9047...	9459810.476...	13_86	5	10
8	822173.3333...	9439812.222...	13_90	3	8

Look for | in MEAN_X | Search

Show selected only | Search selected only | Case sensitive | Advanced search | ? | Close

37. Now add another field to the attribute table of clustermean_all_4326.shp and calculate the mean number of species

- Click the **toggle editing mode** button so that you are in **editing mode**
- Click the **calculator** button to create and **calculate the mean values** as below:



******IMPORTANT NOTE*** There is a bug in QGIS 1.8 and instead of the expression in the screen grab, the numbers need to have the words 'toreal' wrapped around them. Otherwise decimal places will not be kept in the calculation. Type the following:-***

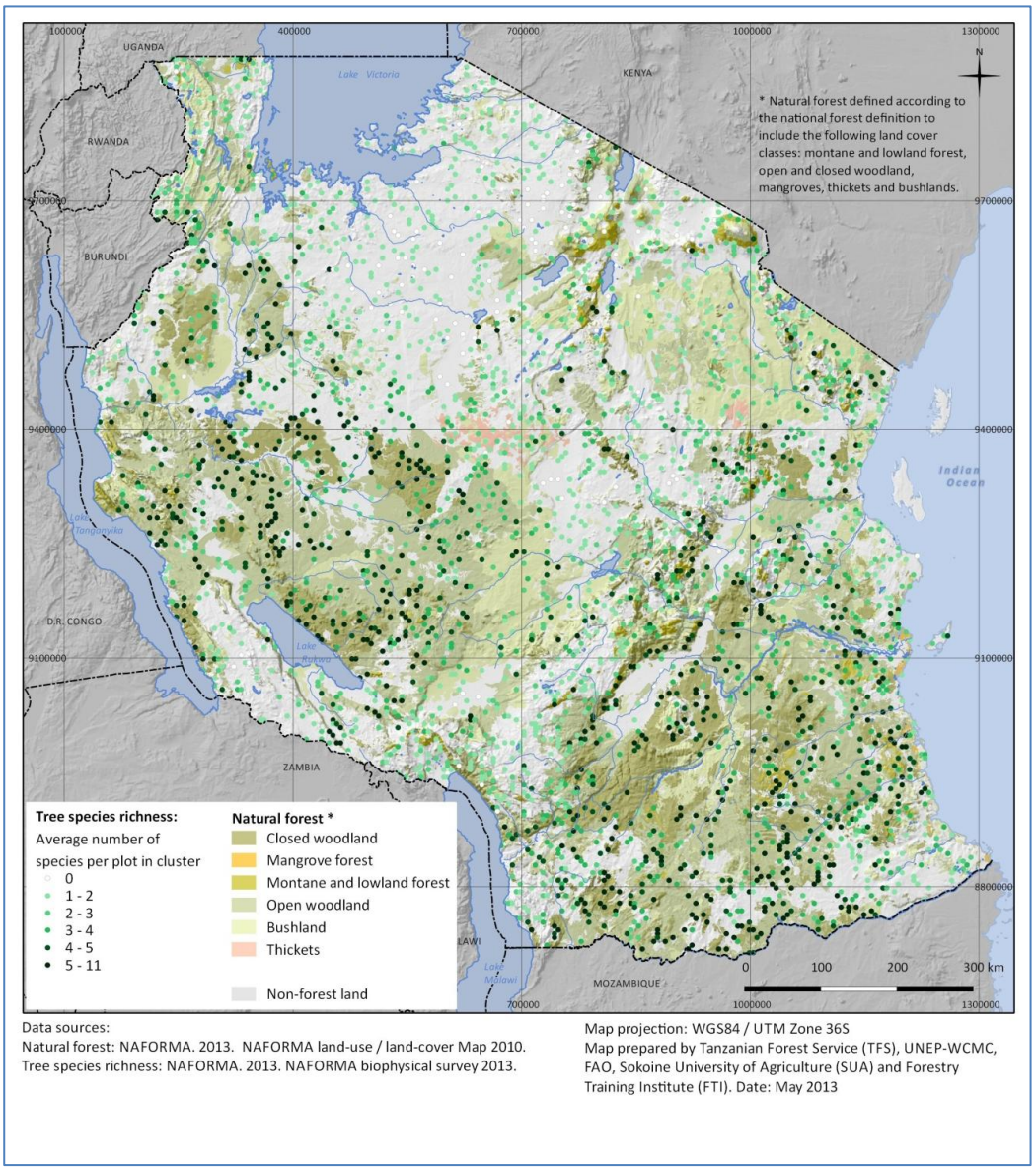
toreal ("no_species") / toreal ("no_plots")

38. Symbolize for each cluster the mean number of tree species in the cluster using a graduated color ramp

- **Right click** on using **clustermean_all_4326.shp** and click on the **style tab**
- Choose **graduated color**
- The column you want to shade on is **mean_spec** field

Example map: Average tree species richness in NAFORMA plots

The map shows the average (mean) species richness per plot for each cluster in the NAFORMA inventory.



(Note: Cartographic design and layout of the map is covered in the Map Layouts section of this report).