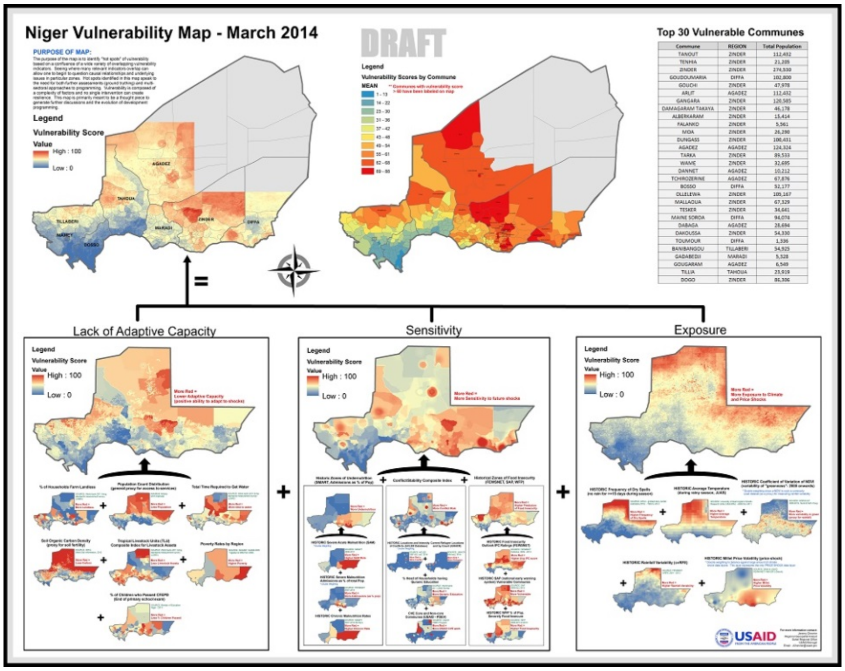
Step-by-Step Guide to Vulnerability Hotspots Mapping: Implementing the Spatial Index Approach



Top 30 Vulnerable Communes

Commune	REGION	Total Population
TANOUT	ZINDER	112,432
TENHIA	ZINDER	21,205
ZINDER	ZINDER	274,330
GOUDOUMARIA	DIFFA	102,800
GOUCHI	ZINDER	47,978
ARLIT	AGADEZ	112,432
GANGARA	ZINDER	120,585
DAMAGARAM TAKAYA	ZINDER	46,178
ALBERKARAM	ZINDER	15,414
FALANKO	ZINDER	5,561
MOA	ZINDER	26,290
DUNGASS	ZINDER	100,431
AGADEZ	AGADEZ	124,324
TARKA	ZINDER	89,533
WAME	ZINDER	32,695
DANNET	AGADEZ	10,212
TCHIROZERINE	AGADEZ	67,876
BOSSO	DIFFA	52,177
OLLELEWA	ZINDER	105,167
MALLAOUA	ZINDER	67,329
TESKER	ZINDER	34,641
MAINE SOROA	DIFFA	94,074
DABAGA	AGADEZ	28,694
DAKOUSSA	ZINDER	54,330
TOUMOUR	DIFFA	1,336
BANIBANGOU	TILLABERI	54,925
GADABEDJI	MARADI	5,328
GOUGARAM	AGADEZ	6,549
TILLIA	TAHOUA	23,919
DOGO	ZINDER	86,306

Developed by the
Center for International Earth Science Information Network (CIESIN),
The Earth Institute, Columbia University

Under contract with TetraTech/ARD for the US Agency for International Development (USAID)
Planning for Resilience in East Africa through Policy, Adaptation, Research, and Economic Development (PREPARED) Project

November 2015

Center for International Earth Science Information Network
EARTH INSTITUTE | COLUMBIA UNIVERSITY

EAC
JUMUIYA YA AFRIKA MASHARIKI

USAID | EAST AFRICA
FROM THE AMERICAN PEOPLE

Authors:

Malanding S Jaiteh, PhD
Tricia Chai-Onn
Valentina Mara
Alex de Sherbinin, PhD

Acknowledgements

This work was performed under contract with TetraTech/ARD for the US Agency for International Development (USAID) Planning for Resilience in East Africa through Policy, Adaptation, Research, and Economic Development (PREPARED) project. The work also benefited greatly from vulnerability mapping efforts under the USAID-funded African and Latin American Resilience to Climate Change (ARCC) project.

Slides on Installing and configuring R adapted from R-Cran available at:
http://cran.r-project.org/bin/windows/base/rw-FAQ.html

We wish to acknowledge the contributions of Dr. idris Bexi Wasama, University of Djibouti, who reviewed the draft manual while visiting CIESIN as a Hubert H. Humphrey Fellow through the University of California, Davis.

Cover image by Jeremy Chevrier, USAID Sahel Regional Office.

Citing this Guide

CIESIN (Center for International Earth Science Information Network), Columbia University. (2015). *A Step-by-Step Guide to Vulnerability Hotspots Mapping: Implementing the Spatial Index Approach*. Palisades NY: CIESIN. Available at http://ciesin.columbia.edu/documents/vmapping_guide_final.pdf

Table of Contents

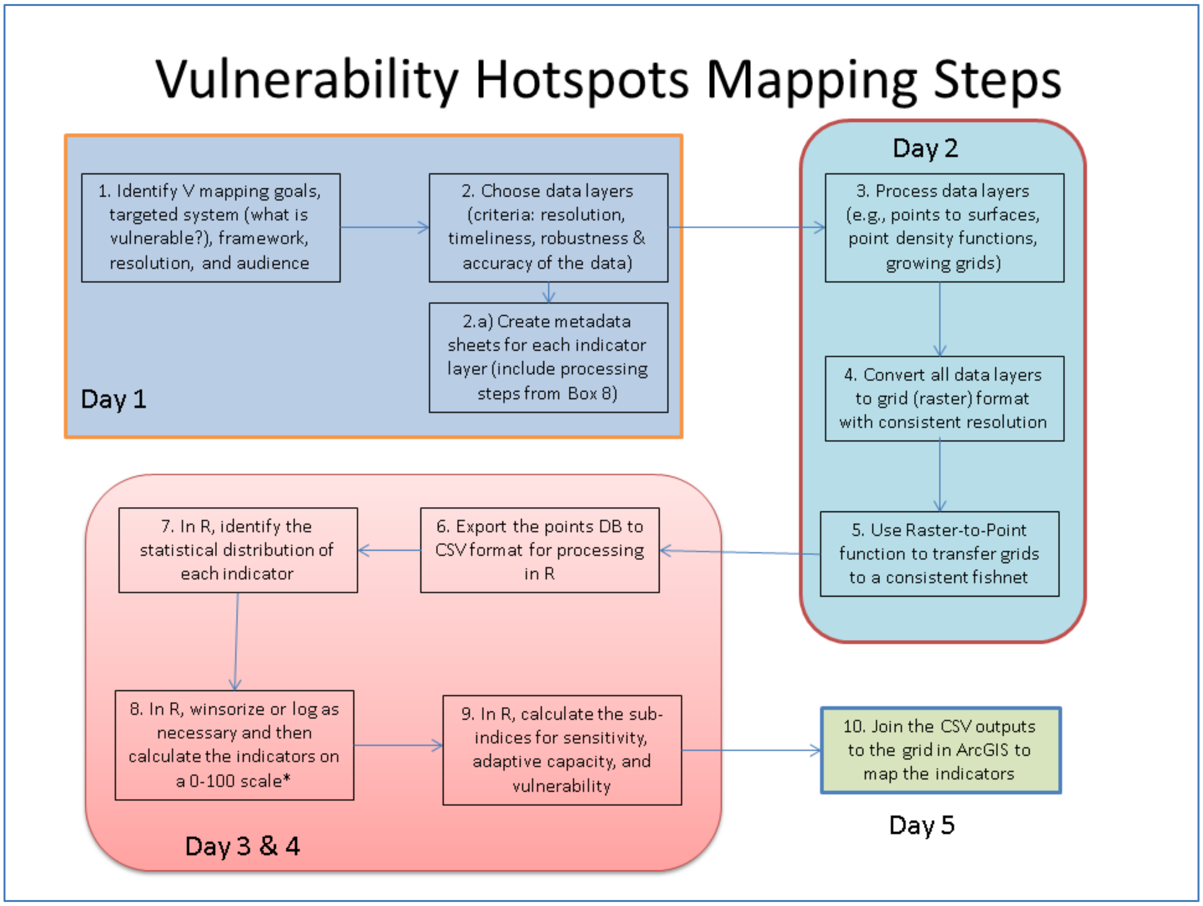
Description

This manual was designed for a five-day training program in Entebbe, Uganda, in August 2015. CIESIN provided a comprehensive training in the framework, data and methods utilized to develop a spatial vulnerability assessment for Kenya using the spatial index approach. Data sets and potential indicators to be derived from them were assembled and evaluated in advance of the workshop. The workshop's primary focus was on the methods needed to process and transform the spatial data in order to develop a spatial vulnerability index (and constituent indices for exposure, sensitivity, and adaptive capacity).

The workshop was intended for GIS analysts who regularly use ArcGIS 9.x or later versions, who are broadly familiar with a variety of geospatial data formats, and who have the ability to do advanced geospatial processing (e.g., editing data sets, creating buffers, conducting overlay analyses, running zonal statistics, etc.).

While vulnerability index maps can be produced entirely in ArcGIS environment, we use R statistical computing and graphics software for the component of part in this training. R is a free software with a wide variety of statistical and graphical techniques that have proven to be very efficient in handling large data sets as the one we encounter in spatial vulnerability index mapping.

The workshop was divided into the following modules, which are addressed in greater detail in the lectures and exercises that follow.


Vulnerability Hotspots Mapping Steps
1. Identify V mapping goals, targeted system (what is vulnerable?), framework, resolution, and audience
2. Choose data layers (criteria: resolution, timeliness, robustness & accuracy of the data)
2.a) Create metadata sheets for each indicator layer (include processing steps from Box 8)
Day 1
Day 2
3. Process data layers (e.g., points to surfaces, point density functions, growing grids)
4. Convert all data layers to grid (raster) format with consistent resolution
5. Use Raster-to-Point function to transfer grids to a consistent fishnet
6. Export the points DB to CSV format for processing in R
7. In R, identify the statistical distribution of each indicator
8. In R, winsorize or log as necessary and then calculate the indicators on a 0-100 scale*
9. In R, calculate the sub-indices for sensitivity, adaptive capacity, and vulnerability
Day 3 & 4
10. Join the CSV outputs to the grid in ArcGIS to map the indicators
Day 5

Lecture 1. Introduction to Vulnerability Hotspots Mapping

A large body of evidence going back more than two decades shows that exposure alone is not sufficient for understanding trends in disaster losses, and that social and economic vulnerability are critical ingredients (Mechler and Bouwer 2014, Cutter et al. 2003). Africa has been identified as one of the regions that is most vulnerable to climate change both in terms of exposure to climate hazards (Turco et al. 2015, Muller et al. 2014) and social vulnerability (Busby et al. 2014, Lopez-Carr et al. 2014). Tools such as spatial vulnerability assessment are useful for understanding patterns of vulnerability and risk to climate change at multiple scales, and have been applied in Africa perhaps more than any other region (e.g. Busby et al. 2014, Lopez-Carr et al. 2014, and Midgely et al. 2011). The demand for vulnerability maps among development agencies and governments is increasing as greater emphasis is placed on scientifically sound methods for targeting adaptation assistance (de Sherbinin 2014a).

Mapping is useful because climate variability and extremes, the sensitivity of populations and systems to climatic stressors, and adaptive capacities are all spatially differentiated. The interplay of these factors produces different patterns of vulnerability. Typically spatial vulnerability assessment involves data integration in which geo-referenced socio-economic and biophysical data, including those derived from remote sensing, are combined with climate data to understand patterns of vulnerability and, in turn, inform where adaptation may be required. Maps have proven to be useful boundary objects in multi-stakeholder discussions, providing a common basis for discussion and for deliberations over adaptation planning (Preston et al. 2011). Maps can help to ground discussions on a solid evidence base, especially in developing country contexts where geographic information may not be easily accessible for all stakeholders.

Spatial data integration and spatial analysis have become standard tools in the toolkit of climate change vulnerability assessments. The United Nations Environment Programme (UNEP) Programme of Research on Climate Change Vulnerability, Impacts and Adaptation (PROVIA) Research Priorities on Vulnerability, Impacts and Adaptation (PROVIA 2013) highlights “measuring and mapping vulnerability” as a first priority for supporting adaptation decision-making. In many cases vulnerability assessment (VA) is synonymous with spatial vulnerability assessment, owing in part to an understanding that vulnerability and its constituent components exhibit high degrees of spatial and temporal heterogeneity

(Preston et al., 2011). The purposes vary according to the specific study, but spatial VAs are generally intended to identify areas at potentially high risk of climate impacts — so-called climate change "hotspots" (de Sherbinin 2013) — and to better understand the determinants of vulnerability in order to identify planning and capacity building needs.

Any vulnerability mapping effort needs to be guided by a theoretical framework. de Sherbinin (2014b) provides a review of a number of different frameworks. In addition, the selection of indicators should be guided by theoretical linkages to the concept of interest – whether vulnerability, or its constituent elements such as exposure, sensitivity, and adaptive capacity. In a mapping effort for Mali (de Sherbinin et al. 2015), which served as a model for the Kenya mapping described in this guide, indicator selection was guided by the literature on factors known to contribute to each component of vulnerability, as well as by data availability and quality. Each data layer was justified based on its conceptual proximity to the three vulnerability components (Hinkel 2011), and choices were consistent with the variables that have been found to be associated with harm from climate variability and change, including education levels (Lutz et al. 2014), climate variability (Hall et al. 2014), and marginal (semi-arid and arid) environments and geographically remote areas in poor developing regions (de Sherbinin et al. 2013, Gray and Moseley 2005). The guiding approach should be to identify a limited number of high-quality spatial data sets that best represent the component of interest while avoiding the temptation to add low-quality data (data of high uncertainty or coarse spatial resolution), thereby "contaminating" the results. We had reasonably high confidence in the validity and reliability of each of the data sets included; data limitations are explored in Annex IV of the overall report.

Vulnerability mapping and the quantification of vulnerability is not without shortcomings, and more critical perspectives are provided in a number of other publications (de Sherbinin 2014a, de Sherbinin 2013, Preston et al. 2011, Hinkel 2011). Users of this guide who desire a more in depth look at the challenges of vulnerability mapping, including issues around uncertainty, are advised to read these publications. Despite these caveats, the spatial vulnerability index construction methods described here are widely used in the literature and have been found to be useful to policy audiences seeking to better understand the factors contributing to vulnerability (e.g., Busby et al. 2014, Midgley et al. 2011, Preston et al. 2011, BMZ 2014). Those wishing an overview of different vulnerability mapping methods may wish to refer to de Sherbinin et al. (2014b).

The following slide show provides an overview of vulnerability mapping.

References

BMZ (German Federal Ministry for Economic Cooperation and Development). (2011). *The Vulnerability Sourcebook: Concept and guidelines for standardised vulnerability assessment;* GIZ (Deutsche Gesellschaft fur Internationale Zusammenarbeit): Berlin, Germany.

Busby, J. W.; Smith, T. G.; Krishnan, N. (2014) Climate security vulnerability in Africa mapping 3.01. Polit. Geogr., 43, 51–67

Cutter, S.L., B.J. Boruff, and W.L. Shirley. (2003) Social Vulnerability to Environmental Hazards. *Soc. Sci. Quarterly.* 84, 242–261

de Sherbinin, A., T. Chai-On, M. Jaiteh, V. Mara, L. Pistolesi, E. Schnarr, S. Trzaska. 2015. Data Integration for Climate Vulnerability Mapping in West Africa. *ISPRS International Journal of Geo-Information*. In press. http://www.mdpi.com/2220-9964/4/3.

de Sherbinin, A. (2014a). *Mapping the Unmeasurable? Spatial analysis of vulnerability to climate change and climate variability*. Dissertation in fulfillment of the PhD degree at the Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente. Enschede, Netherlands: ITC.

de Sherbinin, A. (2014b). *Spatial Climate Change Vulnerability Assessments: A Review of Data, Methods and Issues*. Technical Paper for the USAID African and Latin American Resilience to Climate Change (ARCC) project. Washington, DC: USAID. Available at http://ciesin.columbia.edu/documents/SpatialVulAsses_CLEARED_000.pdf.

de Sherbinin, A. (2013). Climate Change Hotspots Mapping: What Have We Learned? *Climatic Change*, 123(1): 23-37.

Gray, L.C., and W.G. Moseley. (2005). A geographical perspective on poverty–environment interactions. *The Geographical Journal*. DOI: 10.1111/j.1475-4959.2005.00146.x.

Hall, W., D. Grey, D. Garrick , F. Fung , C. Brown , S. J. Dadson, and C. W. Sadoff. (2014). Coping with the curse of freshwater variability. *Science*, 346(6208).

Hinkel, J. (2011). “Indicators of vulnerability and adaptive capacity”: Towards a clarification of the science–policy interface. *Global Environmental Change*, 21(1), 198-208.

Lopez-Carr, D., N. Pricope, J. Aukema, M. Jankowska, C. Funk, G. Husak, and J. Michaelsen. (2014). spatial analysis of population dynamics and climate change in Africa: potential vulnerability hot spots emerge where precipitation declines and demographic pressures coincide. *Population and Environment*, 35:323–339

Lutz, W., R. Muttarak, and E. Striessnig. (2014). Universal education is key to enhanced climate adaptation. *Science,* 346(6213): 1061-1062.

Mechler, R., and L.M. Bouwer. (2014). Understanding trends and projections of disaster losses and climate change: is vulnerability the missing link? *Climatic Change*, 133(1): 23-35.

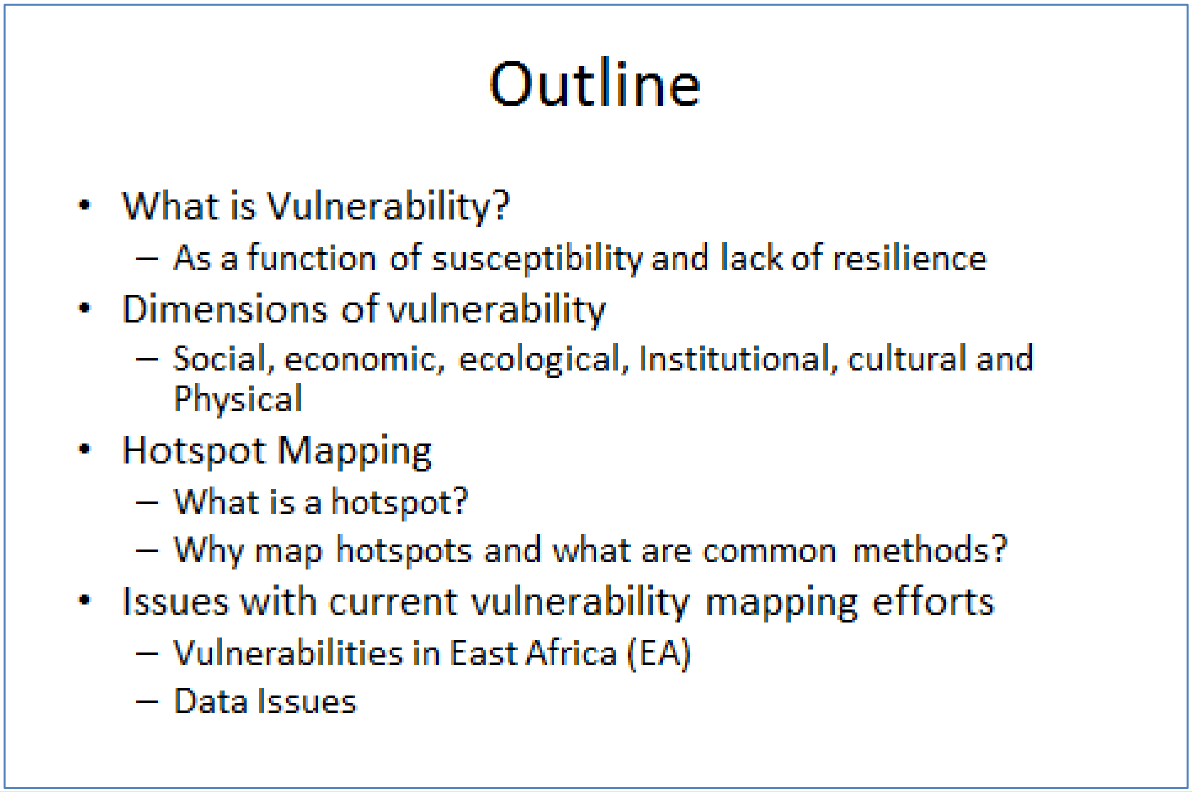
Midgley, S.J.E., Davies, R.A.G., and Chesterman, S. (2011). Climate risk and vulnerability mapping in southern Africa: status quo (2008) and future (2050). Report produced for the Regional Climate Change Programme for Southern Africa (RCCP), UK Department for International Development (DFID). Cape Town, South Africa: OneWorld Sustainable Investments.

Muller, C., K. Waha, A. Bondeau, and J. Heinke. (2014). Hotspots of climate change impacts in sub-Saharan Africa and implications for adaptation and development. *Global Change Biology*. doi: 10.1111/gcb.12586

Preston, B.L, Yuen, E.J., and Westaway, R.M. (2011). Putting vulnerability to climate change on the map: a review of approaches, benefits, and risks. *Sustainability Science, 6*(2), 177-202.

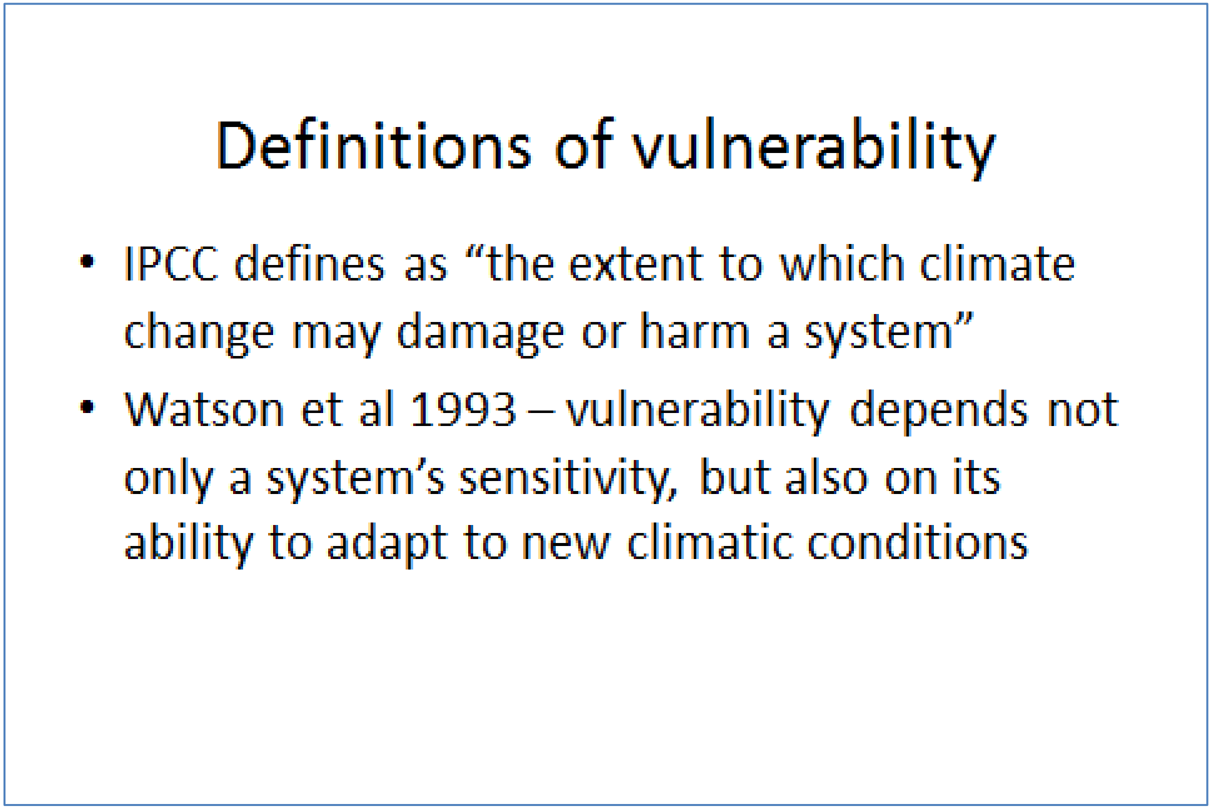
PROVIA. (2013). Research Priorities on Vulnerability, Impacts and Adaptation: Responding to the Climate Change Challenge. Nairobi, Kenya: United Nations Environment Programme.

Turco, M., E. Palazzi1, J. von Hardenberg, and A. Provenzale. 2014. Observed climate change hotspots. *Geophysical Research Letters*, doi:10.1002/2015GL063891



Outline

- What is Vulnerability?
  - As a function of susceptibility and lack of resilience
- Dimensions of vulnerability
  - Social, economic, ecological, Institutional, cultural and Physical
- Hotspot Mapping
  - What is a hotspot?
  - Why map hotspots and what are common methods?
- Issues with current vulnerability mapping efforts
  - Vulnerabilities in East Africa (EA)
  - Data Issues



Definitions of vulnerability

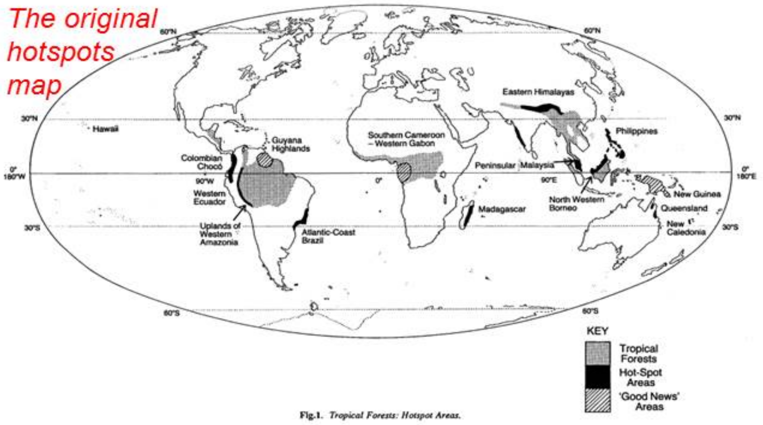
- IPCC defines as “the extent to which climate change may damage or harm a system”
- Watson et al 1993 – vulnerability depends not only a system’s sensitivity, but also on its ability to adapt to new climatic conditions

What is a vulnerability map?

- A vulnerability map gives the location of sites where people, the natural environment or property are at risk due to a potentially catastrophic event
  - E.g. a map showing the housing areas that are vulnerable to a flooding along a river.
- Vulnerability maps are most often created with the assistance of geographic information systems (GIS) but can also be created manually using background maps such as satellite imagery, property boundaries, road maps, or topographic maps.

What is a hotspot?

- Term derived from early work by Norman Myers to identify areas at risk of biodiversity loss
  - Wikipedi
    biogeog
    reservoir
    from hur



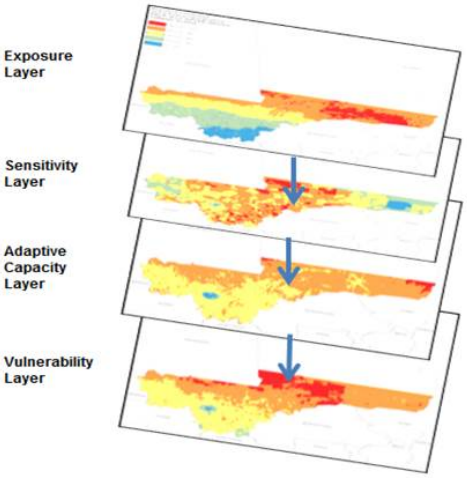
The original hotspots map

Fig.1. *Tropical Forests: Hotspot Areas.*

Source: Myers, N. 1988. Threatened biotas: “Hot spots” in tropical forests. *The Environmentalist* 8 187-208.

What is a climate change hotspot?

- Integrates *spatial variability* in:
  - Climate / biophysical changes
  - Human / system vulnerabilities

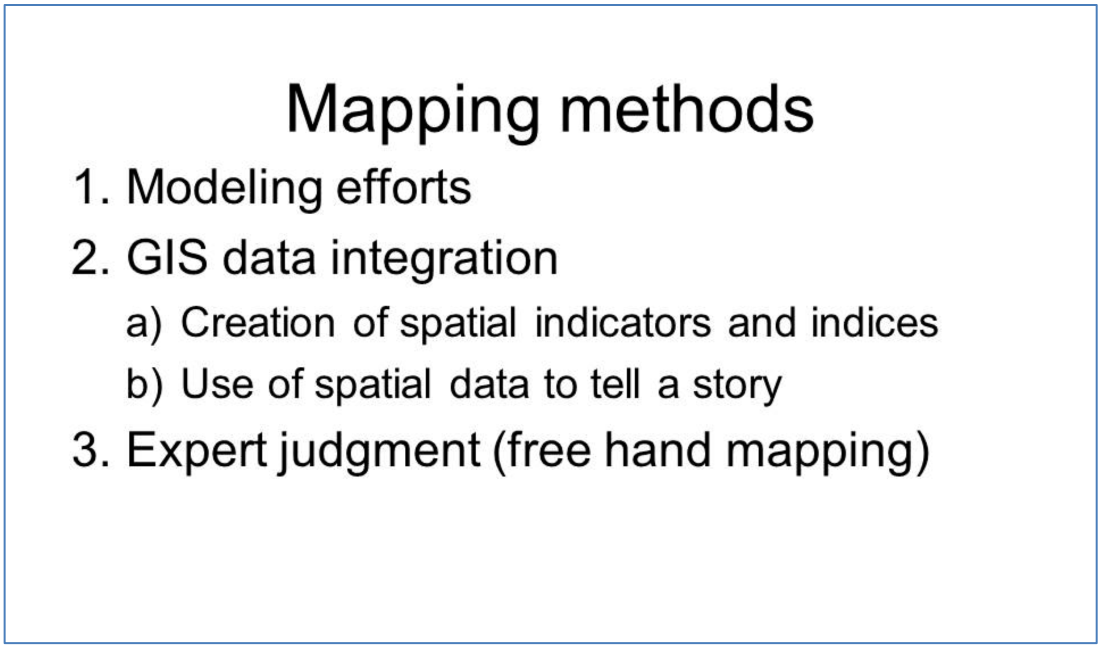


- Climate change exposure is **spatially** differentiated
- Vulnerabilities are **spatially** differentiated
- Adaptive/coping capacities are **spatially** differentiated

Mapping can illuminate key vulnerabilities in the coupled human-environment system and, in turn, inform where adaptation may be required

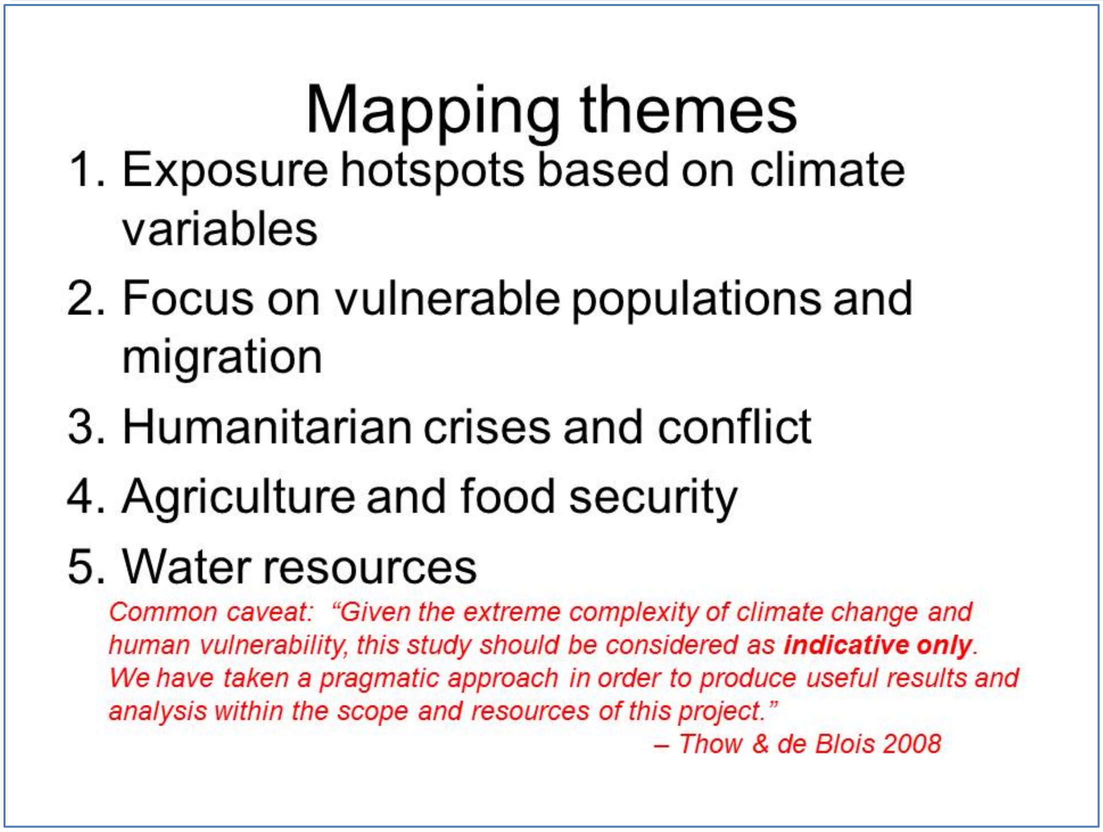
Why map vulnerability hotspots?

- Climate change vulnerability and hotspots maps convey the degree to which a system is likely to be negatively affected by climate change impacts
- By describing the nature of stressors in the most vulnerable areas, the maps help to inform the design of responses
- They are utilized by governments and a number of donors to target development, adaptation, and disaster risk reduction funding
- There is a proliferation of mapping efforts in academia and more refined methodologies are being developed



Mapping methods

1. Modeling efforts
2. GIS data integration
   a) Creation of spatial indicators and indices
   b) Use of spatial data to tell a story
3. Expert judgment (free hand mapping)



Mapping themes

1. Exposure hotspots based on climate variables
2. Focus on vulnerable populations and migration
3. Humanitarian crises and conflict
4. Agriculture and food security
5. Water resources

*Common caveat: "Given the extreme complexity of climate change and human vulnerability, this study should be considered as **indicative only**. We have taken a pragmatic approach in order to produce useful results and analysis within the scope and resources of this project."*

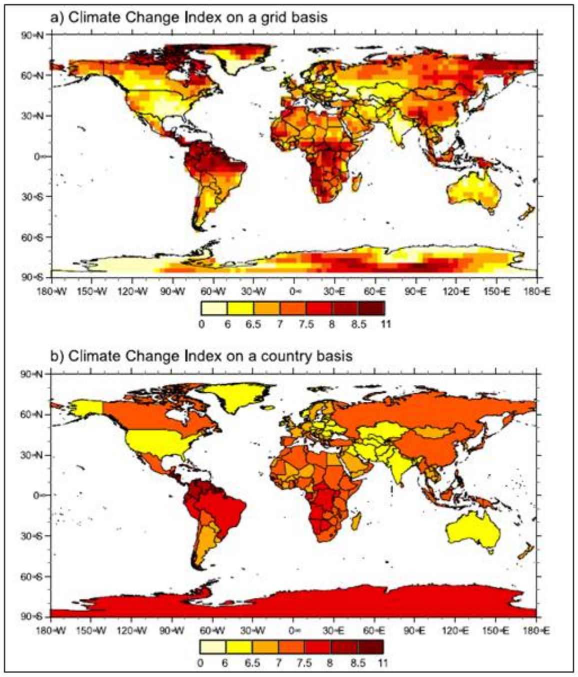
– Thow & de Blois 2008

1. Exposure Hotpots

Climate Change Index

- Tracks increases in the probabilities of extreme events relative to a 1 in 20 year return cycle for the same events under current climatology
- Based on SRES A2 and B2 scenarios
- Indicators: additional hottest years, driest years, and extremely warm/wet/dry seasons

Source: Baettig, M.B., M. Wild, and D.M. Imboden. 2007. A climate change index. *Geophysical Research Letters*, 34, L01705



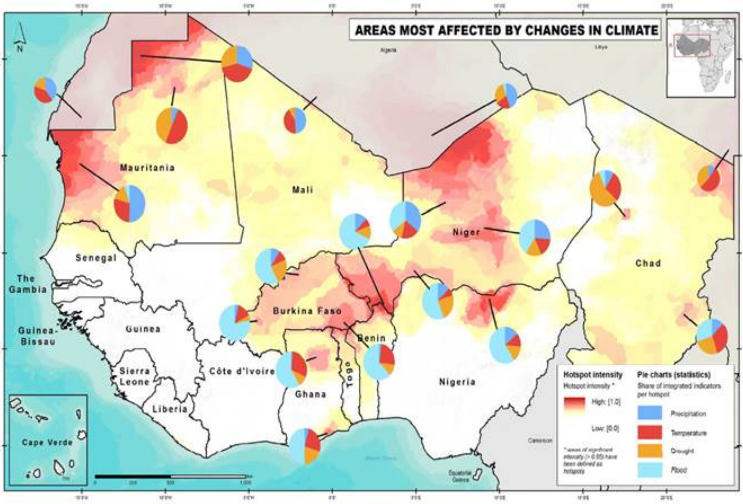
a) Climate Change Index on a grid basis

b) Climate Change Index on a country basis

90-N
60-N
30-N
0-
30-S
60-S
90-S
180-W 150-W 120-W 90-W 60-W 30-W 0- 30-E 60-E 90-E 120-E 150-E 180-E
0 6 6.5 7 7.5 8 8.5 11

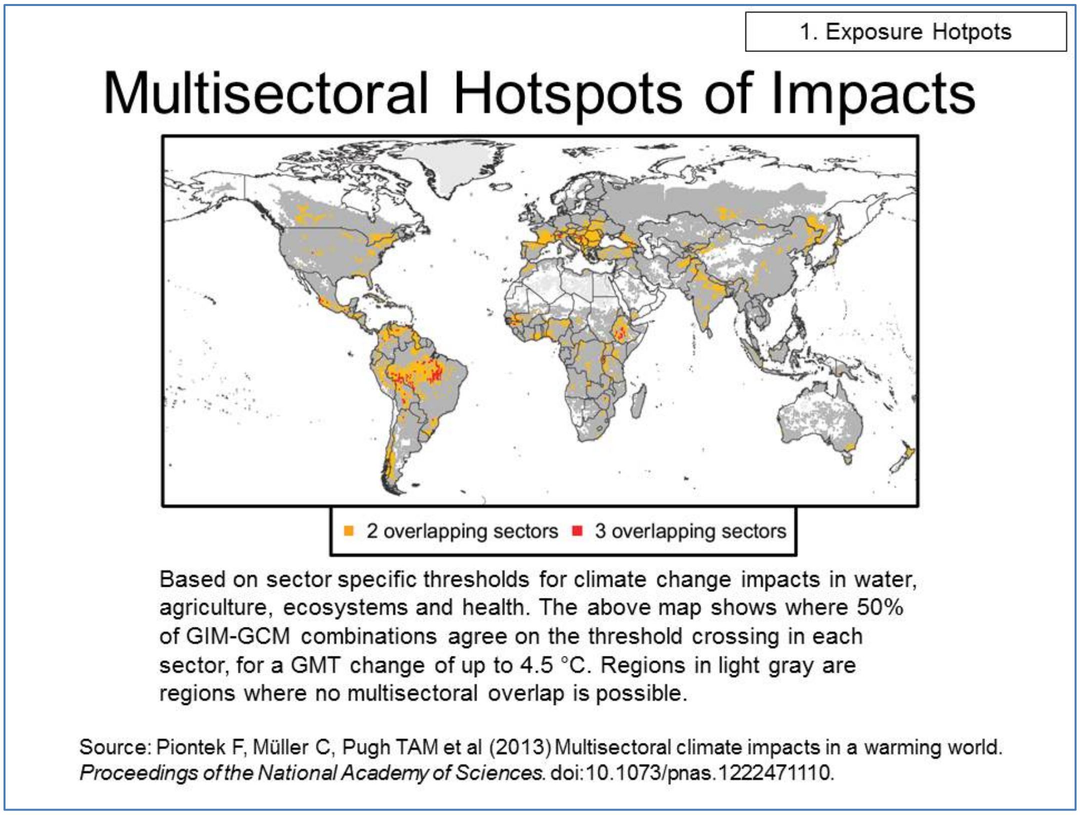
1. Exposure Hotpots

Cumulative Climate Change Index of the Sahel



Hotspots are identified as areas where temperature and precipitation trends are pronounced and drought and flood events over the past 24-36 years have been severe, with a focus on the rainy season from May to October.

Source: Hagenlocher, M., et al. 2013. Modeling Hotspots of Climate Change in the Sahel Using Object-Based Regionalization of Multidimensional Gridded Datasets. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. doi: 10.1109/JSTARS.2013.2259579



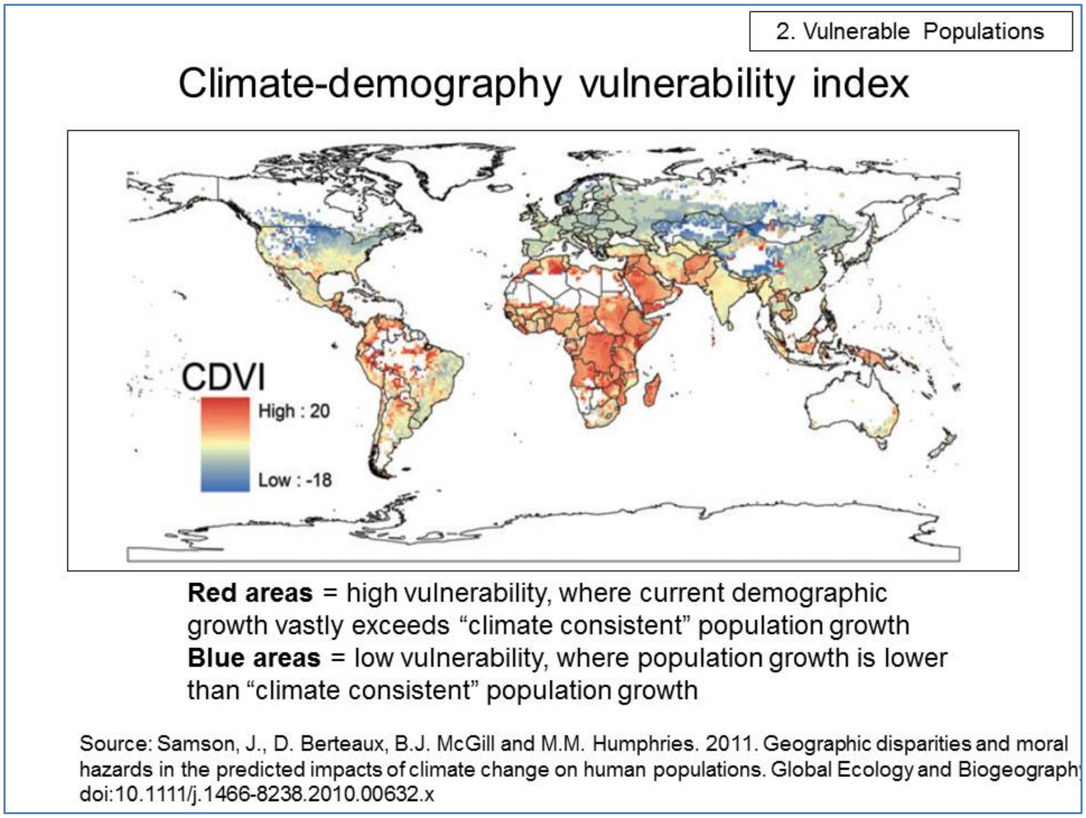
1. Exposure Hotpots

Multisectoral Hotspots of Impacts

🟧 2 overlapping sectors 🟥 3 overlapping sectors

Based on sector specific thresholds for climate change impacts in water, agriculture, ecosystems and health. The above map shows where 50% of GIM-GCM combinations agree on the threshold crossing in each sector, for a GMT change of up to 4.5 °C. Regions in light gray are regions where no multisectoral overlap is possible.

Source: Piontek F, Müller C, Pugh TAM et al (2013) Multisectoral climate impacts in a warming world. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1222471110.

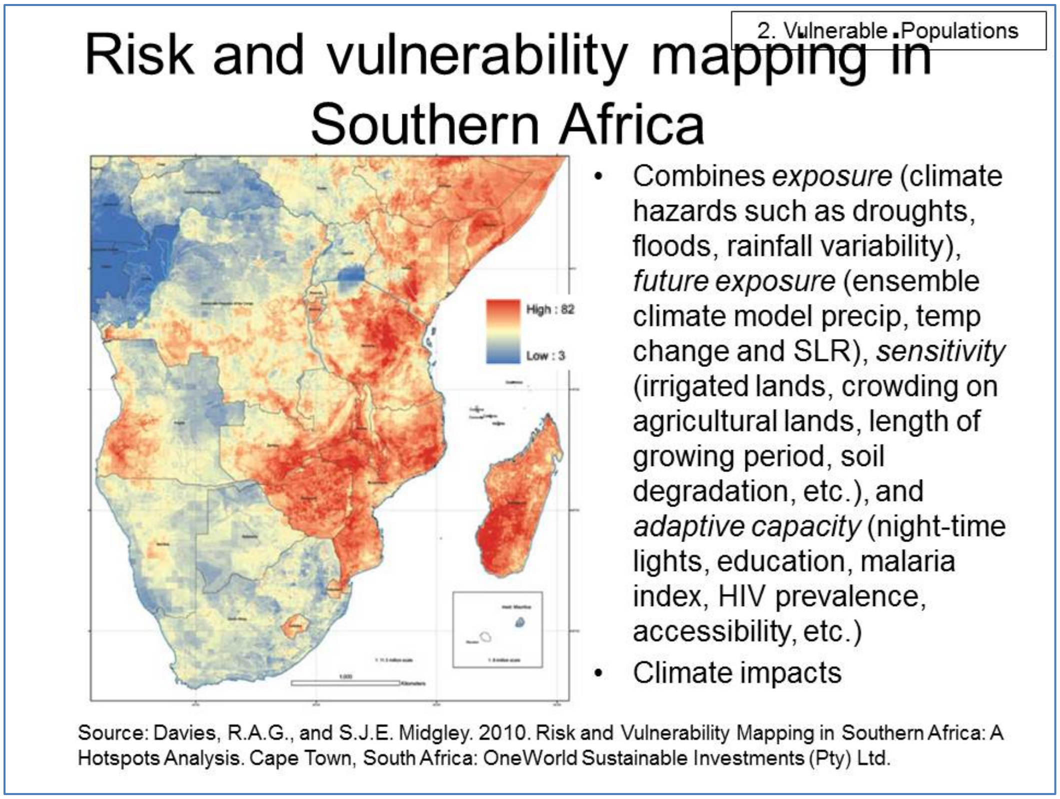


2. Vulnerable Populations

Climate-demography vulnerability index

Red areas = high vulnerability, where current demographic growth vastly exceeds "climate consistent" population growth
Blue areas = low vulnerability, where population growth is lower than "climate consistent" population growth

Source: Samson, J., D. Berteaux, B.J. McGill and M.M. Humphries. 2011. Geographic disparities and moral hazards in the predicted impacts of climate change on human populations. Global Ecology and Biogeograph
doi:10.1111/j.1466-8238.2010.00632.x

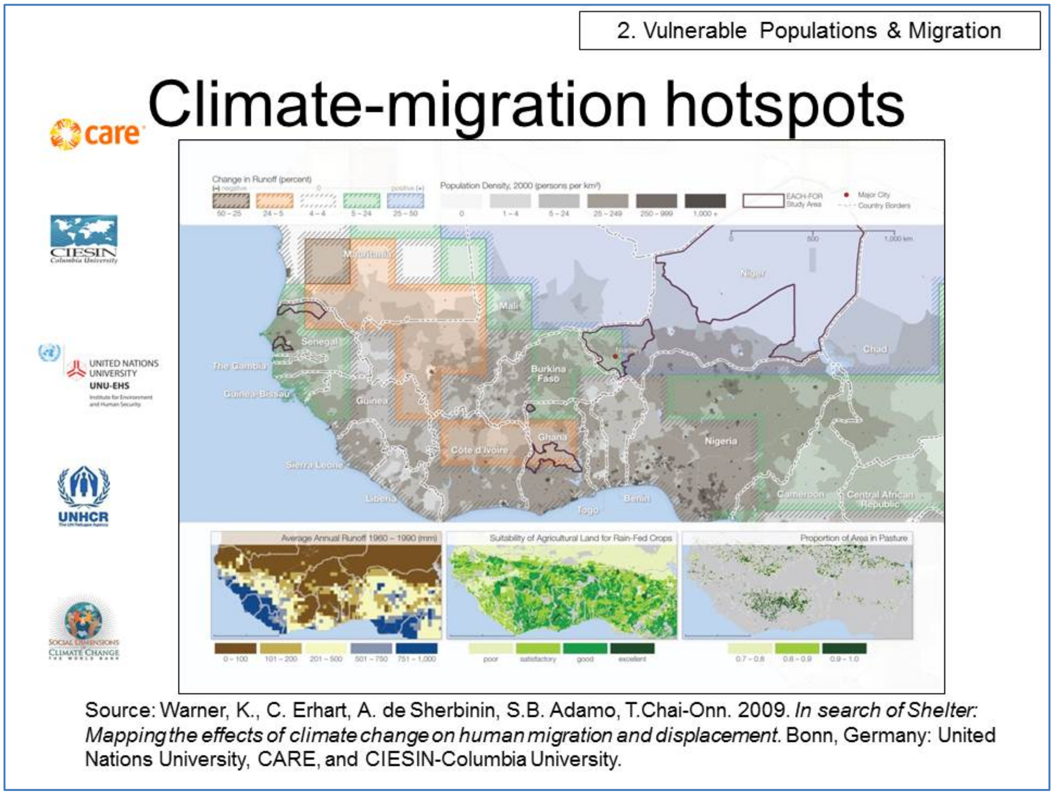


2. Vulnerable Populations

Risk and vulnerability mapping in Southern Africa

- Combines *exposure* (climate hazards such as droughts, floods, rainfall variability), *future exposure* (ensemble climate model precip, temp change and SLR), *sensitivity* (irrigated lands, crowding on agricultural lands, length of growing period, soil degradation, etc.), and *adaptive capacity* (night-time lights, education, malaria index, HIV prevalence, accessibility, etc.)
- Climate impacts

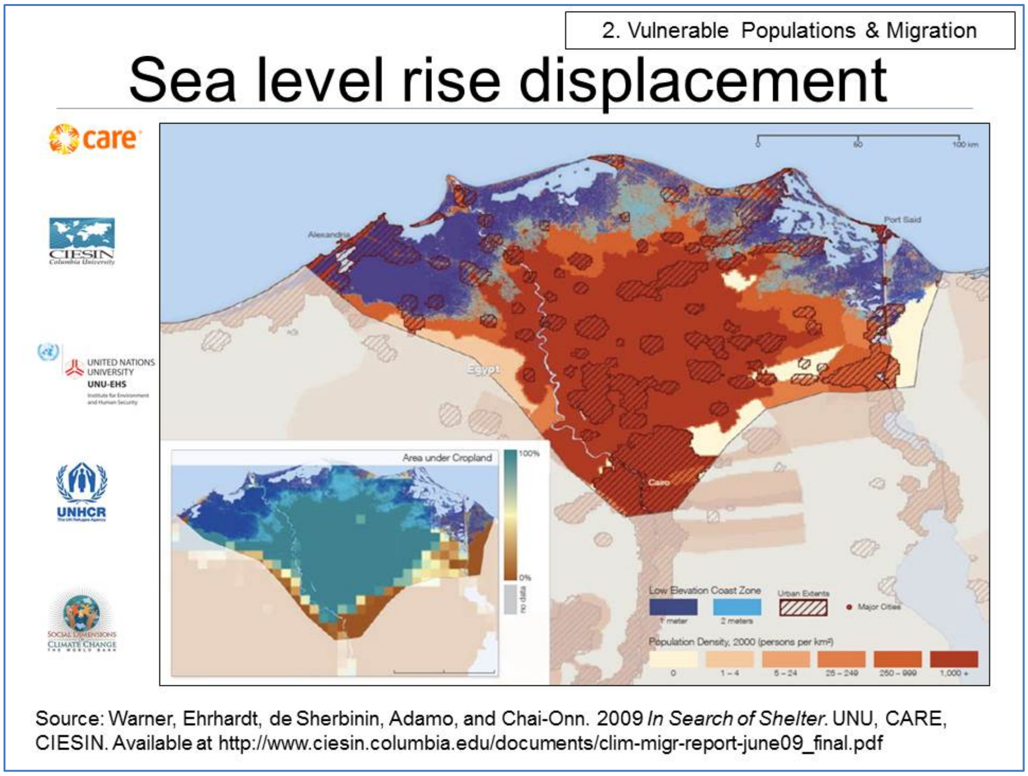
Source: Davies, R.A.G., and S.J.E. Midgley. 2010. Risk and Vulnerability Mapping in Southern Africa: A Hotspots Analysis. Cape Town, South Africa: OneWorld Sustainable Investments (Pty) Ltd.



2. Vulnerable Populations & Migration

Climate-migration hotspots

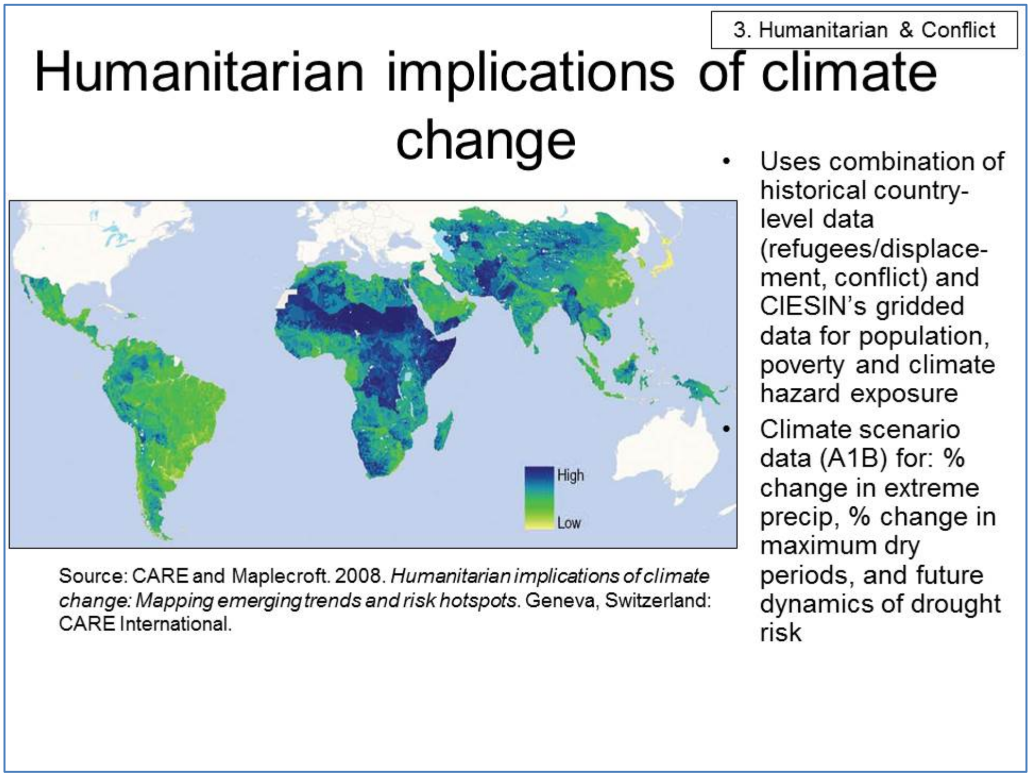
Source: Warner, K., C. Erhart, A. de Sherbinin, S.B. Adamo, T.Chai-Onn. 2009. *In search of Shelter: Mapping the effects of climate change on human migration and displacement*. Bonn, Germany: United Nations University, CARE, and CIESIN-Columbia University.



2. Vulnerable Populations & Migration

Sea level rise displacement

Source: Warner, Ehrhardt, de Sherbinin, Adamo, and Chai-Onn. 2009 *In Search of Shelter*. UNU, CARE, CIESIN. Available at http://www.ciesin.columbia.edu/documents/clim-migr-report-june09_final.pdf

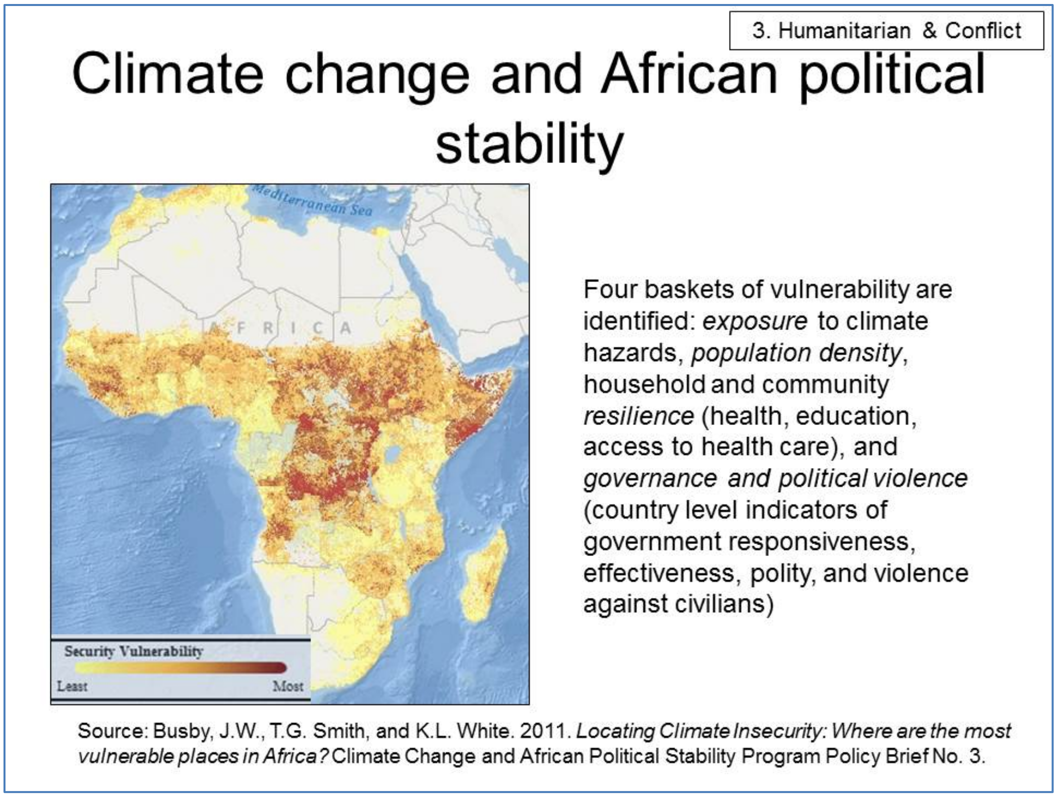


3. Humanitarian & Conflict

Humanitarian implications of climate change

- Uses combination of historical country-level data (refugees/displacement, conflict) and CIESIN's gridded data for population, poverty and climate hazard exposure
- Climate scenario data (A1B) for: % change in extreme precip, % change in maximum dry periods, and future dynamics of drought risk

Source: CARE and Maplecroft. 2008. *Humanitarian implications of climate change: Mapping emerging trends and risk hotspots*. Geneva, Switzerland: CARE International.

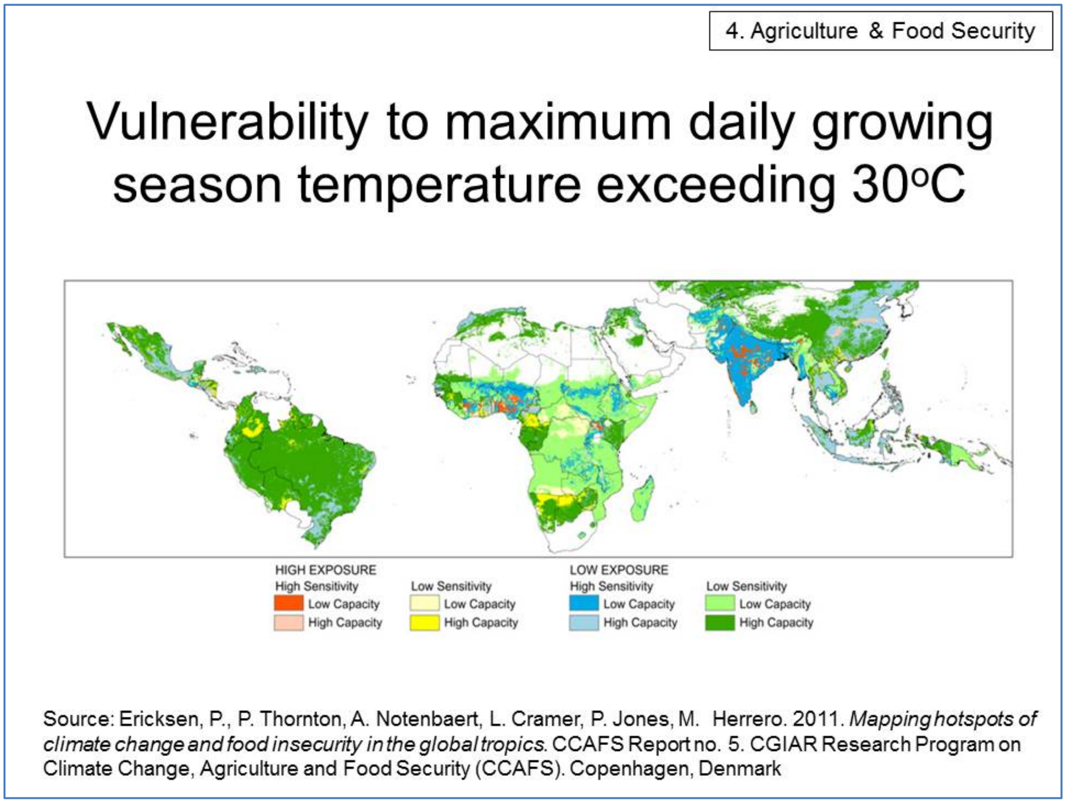


3. Humanitarian & Conflict

Climate change and African political stability

Four baskets of vulnerability are identified: *exposure* to climate hazards, *population density*, household and community *resilience* (health, education, access to health care), and *governance and political violence* (country level indicators of government responsiveness, effectiveness, polity, and violence against civilians)

Source: Busby, J.W., T.G. Smith, and K.L. White. 2011. *Locating Climate Insecurity: Where are the most vulnerable places in Africa?* Climate Change and African Political Stability Program Policy Brief No. 3.

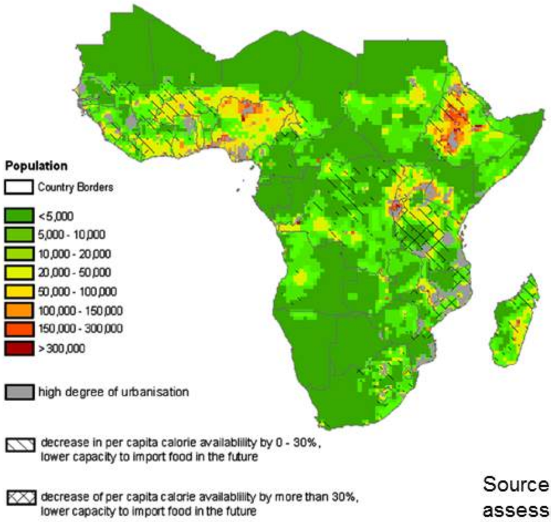


4. Agriculture & Food Security

Vulnerability to maximum daily growing season temperature exceeding 30^{o}C

Source: Ericksen, P., P. Thornton, A. Notenbaert, L. Cramer, P. Jones, M. Herrero. 2011. *Mapping hotspots of climate change and food insecurity in the global tropics.* CCAFS Report no. 5. CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS). Copenhagen, Denmark

Current and future hotspots of hunger in Sub-Saharan Africa in the context of global change


Population
□ Country Borders
🟩 <5,000
🟩 5,000 - 10,000
🟩 10,000 - 20,000
🟨 20,000 - 50,000
🟨 50,000 - 100,000
🟧 100,000 - 150,000
🟥 150,000 - 300,000
🟥 >300,000
⬜ high degree of urbanisation
▧ decrease in per capita calorie availability by 0 - 30%, lower capacity to import food in the future
▦ decrease of per capita calorie availability by more than 30%, lower capacity to import food in the future

- Areas experiencing under-nutrition were assessed
- Impact of climate change on the production of six major crops was analyzed
- Areas that are projected to experience hunger in the context of anticipated climate, social, economic, and biophysical changes were mapped

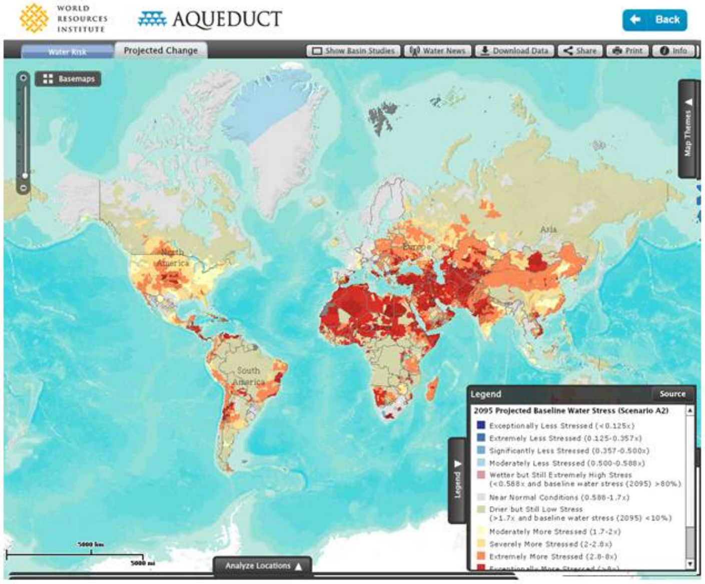
Source: Liu, J., et al. 2008. A spatially explicit assessment of current and future hotspots of hunger in Sub-Saharan Africa in the context of global change. *Global and Planetary Change*, 64 (2008): 222–235

5. Water resources

Aqueduct Water Risk Management

Creates a water risk score based on multiple climate scenarios and a hydrological model for use by "companies, investors, and governments".

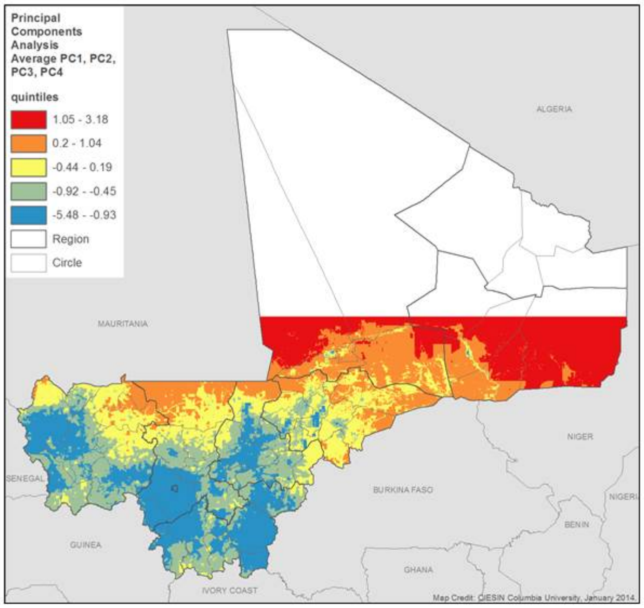
Here baseline conditions and projections to 2050 (SRES A2 scenario) are provided.


WORLD RESOURCES INSTITUTE
AQUEDUCT
Back
Water Risk
Projected Change
Show Basin Studies
Water News
Download Data
Share
Print
Info
Basemaps
Map Themes
North America
South America
Europe
Asia
5000 km
5000 mi
Analyze Locations
Legend
Legend
Source
2095 Projected Baseline Water Stress (Scenario A2)
Exceptionally Less Stressed (<0.125x)
Extremely Less Stressed (0.125-0.357x)
Significantly Less Stressed (0.357-0.500x)
Moderately Less Stressed (0.500-0.588x)
Wetter but Still Extremely High Stress (<0.588x and baseline water stress (2095) >80%)
Near Normal Conditions (0.588-1.7x)
Drier but Still Low Stress (>1.7x and baseline water stress (2095) <10%)
Moderately More Stressed (1.7-2x)
Severely More Stressed (2-2.8x)
Extremely More Stressed (2.8-8x)
Exceptionally More Stressed (>8x)

Source: http://aqueduct.wri.org/atlas.

2. Vulnerable Populations

Mali Climate Vulnerability Mapping

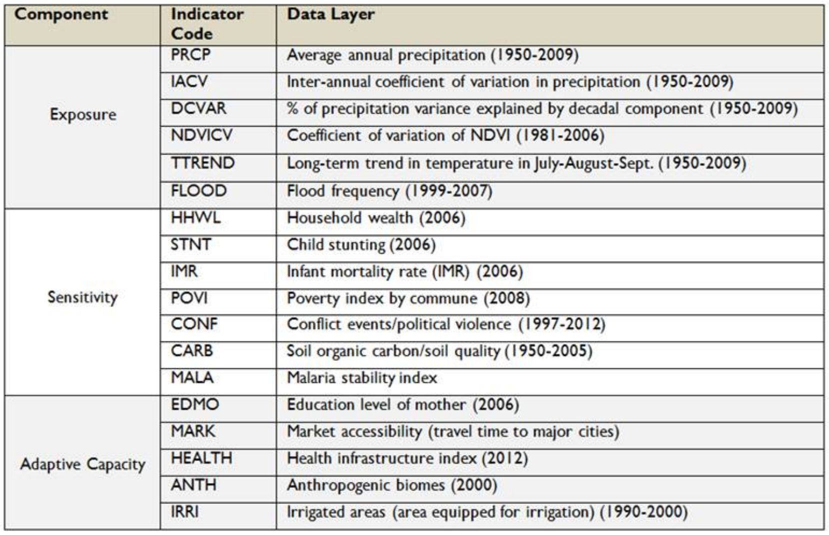


Map Credit: CIESIN Columbia University, January 2014.

- Created separate maps for:
  - Exposure
  - Sensitivity
  - Lack of Adaptive Capacity
  - Vulnerability
- Explores PCA as an alternative aggregation tool

2. Vulnerable Populations

Mali Climate Vulnerability Mapping

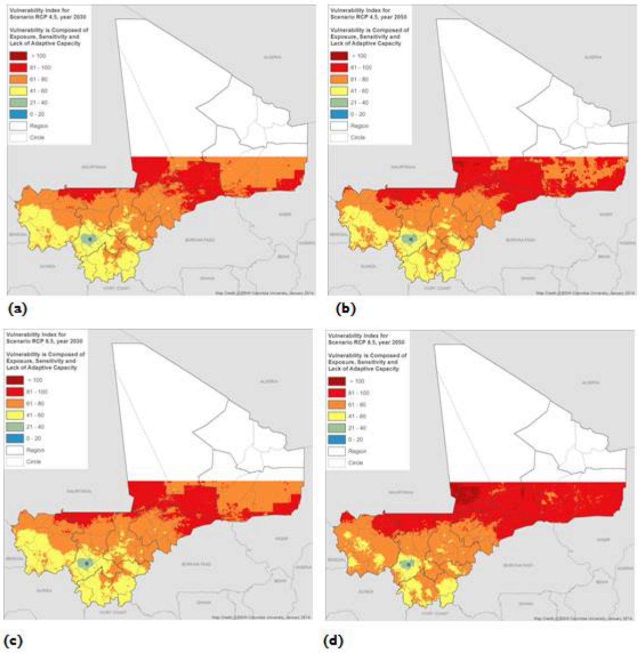


Component	Indicator Code	Data Layer
Exposure	PRCP	Average annual precipitation (1950-2009)
	IACV	Inter-annual coefficient of variation in precipitation (1950-2009)
	DCVAR	% of precipitation variance explained by decadal component (1950-2009)
	NDVICV	Coefficient of variation of NDVI (1981-2006)
	TTREND	Long-term trend in temperature in July-August-Sept. (1950-2009)
	FLOOD	Flood frequency (1999-2007)
Sensitivity	HHWL	Household wealth (2006)
	STNT	Child stunting (2006)
	IMR	Infant mortality rate (IMR) (2006)
	POVI	Poverty index by commune (2008)
	CONF	Conflict events/political violence (1997-2012)
	CARB	Soil organic carbon/soil quality (1950-2005)
	MALA	Malaria stability index
Adaptive Capacity	EDMO	Education level of mother (2006)
	MARK	Market accessibility (travel time to major cities)
	HEALTH	Health infrastructure index (2012)
	ANTH	Anthropogenic biomes (2000)
	IRRI	Irrigated areas (area equipped for irrigation) (1990-2000)

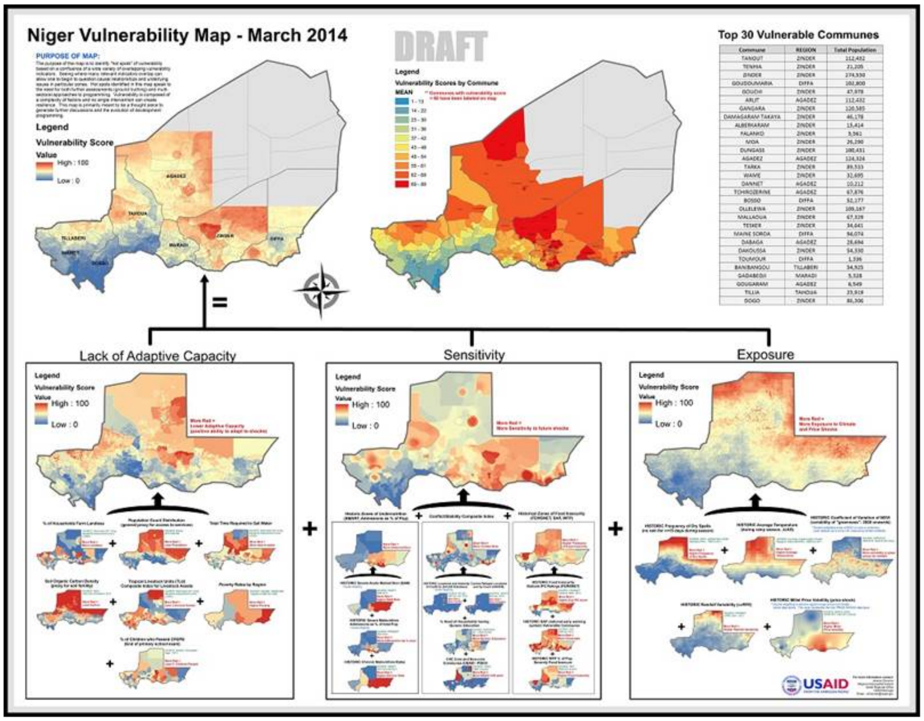
Baseline assessment using the IPCC definition of vulnerability and an additive approach

2. Vulnerable Populations

Vulnerability Index for Scenario RCP 4.5 (a) year 2030 and (b) year 2050; and RCP8.5 (c) year 2030 and (d) year 2050


Vulnerability Index for Scenario RCP 4.5, year 2030
Vulnerability Index for Scenario RCP 4.5, year 2050
Vulnerability Index for Scenario RCP 8.5, year 2030
Vulnerability Index for Scenario RCP 8.5, year 2050
Vulnerability is Composed of Exposure, Sensitivity and Lack of Adaptive Capacity
> 100
81 - 100
61 - 80
41 - 60
21 - 40
0 - 20
Region
Circle
ALGERIA
MAURITANIA
NIGER
BURKINA FASO
SENEGAL
GUINEA
IVORY COAST
(a)
(b)
(c)
(d)

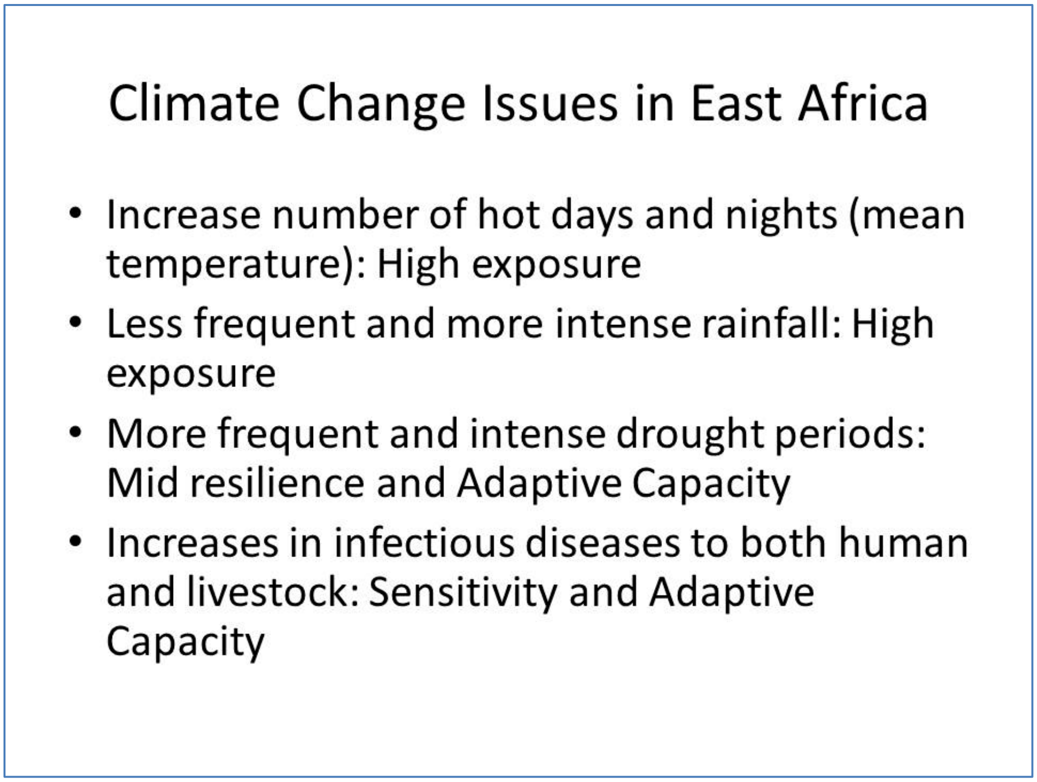
2. Vulnerable Populations



Top 30 Vulnerable Communes

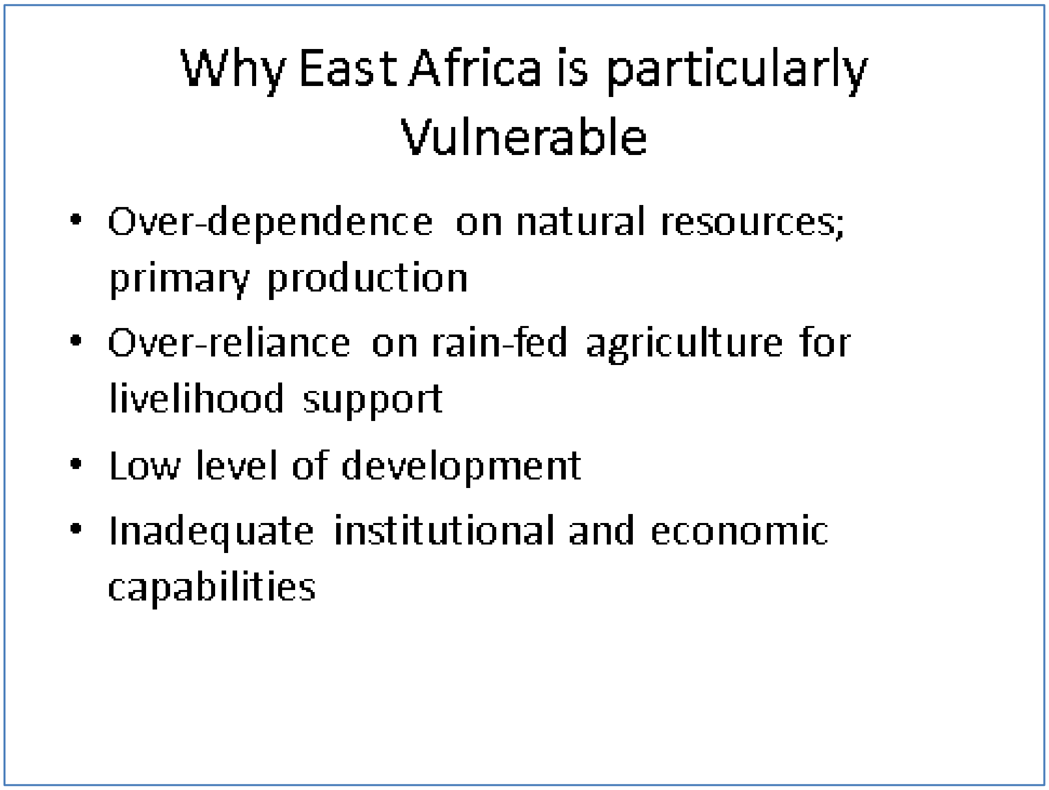
Commune	REGION	Total Population
TANOUT	ZINDER	112,452
TENHIA	ZINDER	21,205
ZINDER	ZINDER	274,550
GOUDOUMARIA	DIFFA	102,800
GOUCHI	ZINDER	47,978
ARLIT	AGADEZ	112,432
GANGARA	ZINDER	120,585
DAMAGARAM TAKAYA	ZINDER	46,178
ALBERKARAM	ZINDER	13,414
FALANKO	ZINDER	5,561
MOA	ZINDER	26,290
DUNGASS	ZINDER	100,431
AGADEZ	AGADEZ	124,324
TARKA	ZINDER	89,533
WAME	ZINDER	32,695
DANNET	AGADEZ	10,212
TCHIROZERINE	AGADEZ	67,876
BOSSO	DIFFA	52,177
OLLELEWA	ZINDER	100,167
MALLAOUA	ZINDER	67,328
TESKER	ZINDER	34,641
MAINE SOROA	DIFFA	94,074
DABAGA	AGADEZ	28,694
DAKOUSSA	ZINDER	54,530
TOUMOUR	DIFFA	1,536
BANIBANGOU	TILLABERI	34,925
GADABEDJI	MARADI	5,528
GOUGARAM	AGADEZ	6,549
TILLIA	TAHOUA	23,919
GOGO	ZINDER	80,306

Map courtesy Jeremy Chevrier, Sahel Joint Planning Cell



Climate Change Issues in East Africa

- Increase number of hot days and nights (mean temperature): High exposure
- Less frequent and more intense rainfall: High exposure
- More frequent and intense drought periods: Mid resilience and Adaptive Capacity
- Increases in infectious diseases to both human and livestock: Sensitivity and Adaptive Capacity



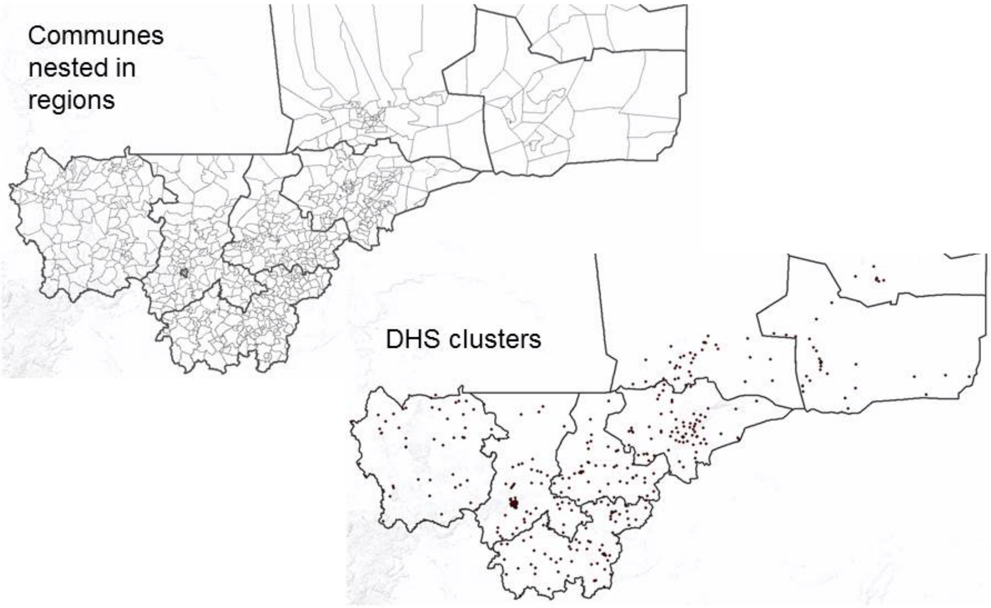
Why East Africa is particularly Vulnerable

- Over-dependence on natural resources; primary production
- Over-reliance on rain-fed agriculture for livelihood support
- Low level of development
- Inadequate institutional and economic capabilities

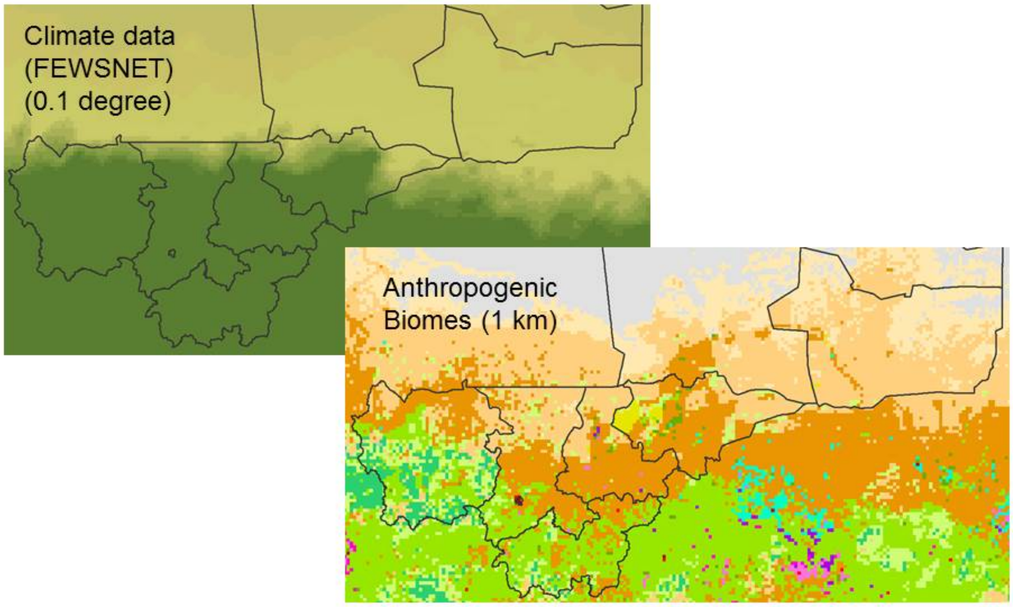
Hotspots mapping: Data issues

- Artifacts of spatial data integration:
  - Combinations of country-level, subnational and continuously varying variables such as precipitation leads to discontinuities across admin units
  - Speckling effect from higher resolution data reduces ability to interpret results

Mali VA Mapping: Sample Resolutions

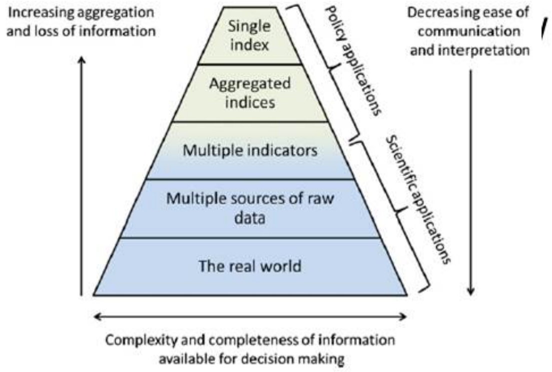


Mali VA Mapping: Sample Resolutions

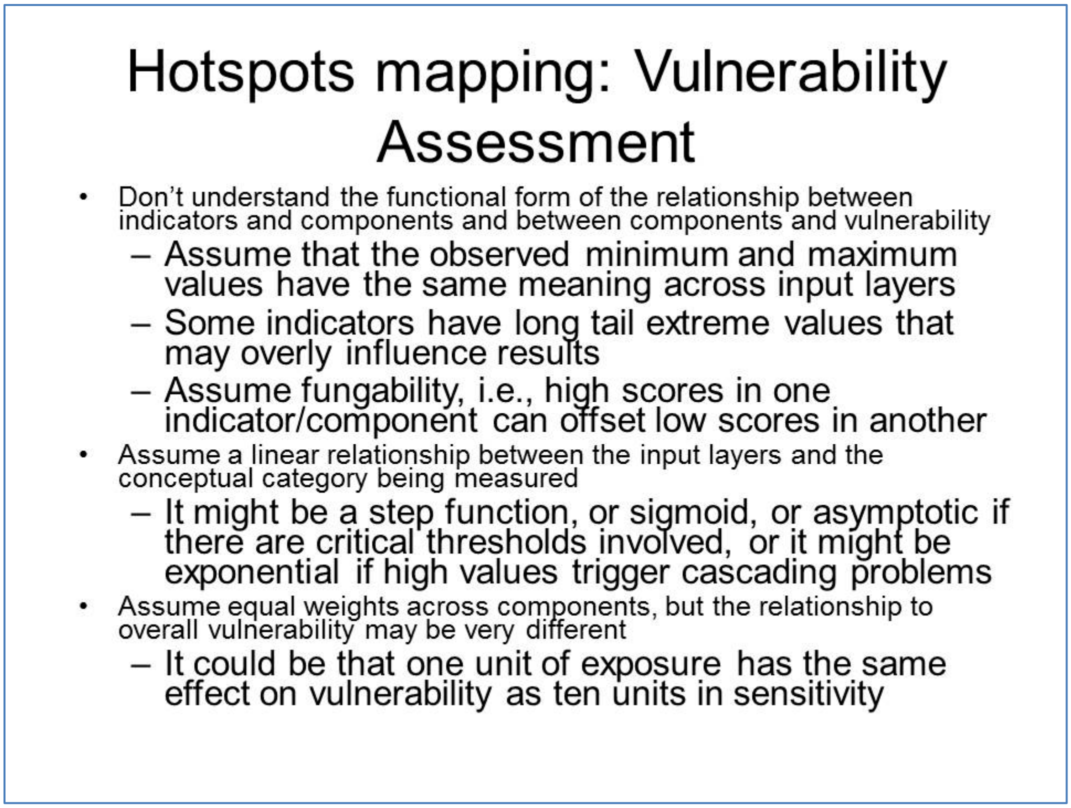


Hotspots mapping: Policy communication issues

- If results do not conform to the policy maker's own mental models, they may not be widely accepted
- Tradeoff
commun
complex



Source: Abson DJ et al. (2012). *Appl Geogr* 35:515–524



Hotspots mapping: Vulnerability Assessment

- Don’t understand the functional form of the relationship between indicators and components and between components and vulnerability
  - – Assume that the observed minimum and maximum values have the same meaning across input layers
  - – Some indicators have long tail extreme values that may overly influence results
  - – Assume fungability, i.e., high scores in one indicator/component can offset low scores in another
- Assume a linear relationship between the input layers and the conceptual category being measured
  - – It might be a step function, or sigmoid, or asymptotic if there are critical thresholds involved, or it might be exponential if high values trigger cascading problems
- Assume equal weights across components, but the relationship to overall vulnerability may be very different
  - – It could be that one unit of exposure has the same effect on vulnerability as ten units in sensitivity

Exercise 1. Developing Vulnerability Mapping Framework and Selecting Variables

1.0 Variable Listing

In this exercise we will examine the variables under each of the components, Exposure, Adaptive Capacity and Sensitivity to climate change and vulnerability

Exposure

- Seasonal temperature
- Average Annual Precipitation
- Long-term trend in temperature in July-August-Sept. (1950-2009)
- Flood frequency (1999-2007)

Adaptive Capacity

- Education level of mother (2006) point/polygon
- Market accessibility (travel time to major cities)
- Anthropogenic biomes (2000)
- Irrigated areas (area equipped for irrigation) (1990-2000)

Sensitivity

- Household wealth (2006)
- Child stunting (2006)
- Infant mortality rate (IMR) (2008)
- Soil organic carbon/soil quality (1950-2005)

1.1 Variable definition

Variable name:……….

Briefly explain three ways in which this variable may contribute to social vulnerability to climate change.

1………
……
……….
2………
……
……
3………
……
……….

List and briefly explain two alternative variables you know that can be used to substitute for this variable

1……
……….
2………
……….
3………
……

List in order of importance which component of vulnerability (exposure, adaptive capacity and sensitivity) does this belong.

1………………………………………………….

2………………………………………………..

3…………………………………………………

1.2 Final variable listing

Sensitivity:

1.
2.
3.
4.
5.

Exposure

1.
2.
3.
4.
5.

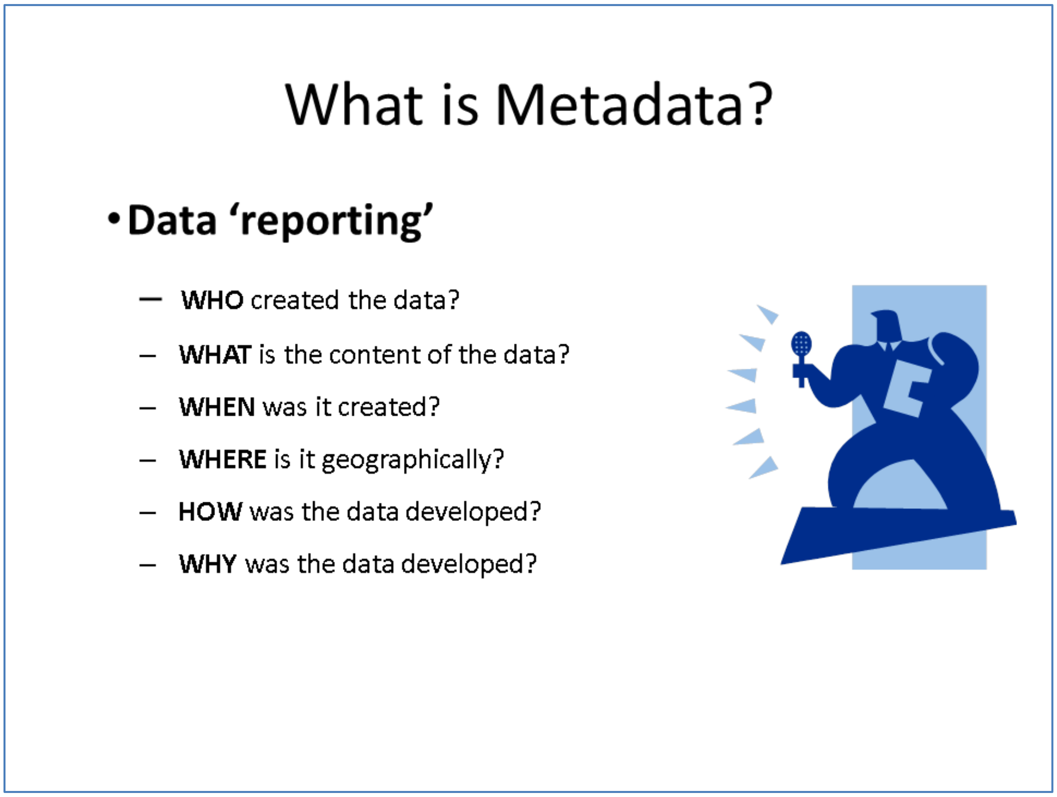
Adaptive Capacity

1.
2.
3.
4.
5.

Lecture 2. Writing metadata for spatial data

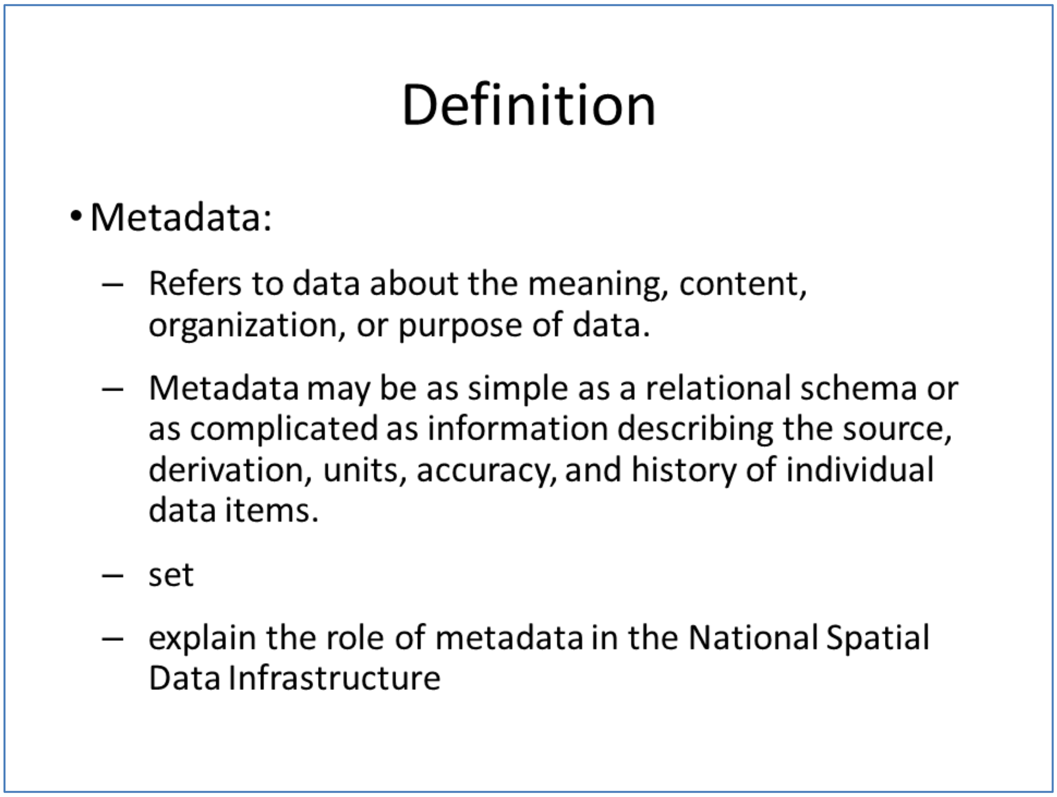
Metadata is data that describes other data. Metadata summarizes basic information about data such as the author, date created and date modified and file size. The main purpose of metadata is to facilitate in the discovery of relevant information. Metadata also helps organize electronic resources, provide digital identification, and helps support archiving and preservation of the resource. Having the ability to filter through that metadata makes it much easier for someone to locate a specific document.

Metadata in GIS describe geographic objects (such as shapefiles, satellite imagery, maps, features, or simply documents with a geospatial component). Metadata can be created manually, or by automated information processing. Manual creation tends to be more accurate, allowing the user to input any information they feel is relevant or needed to help describe the file. Automated metadata creation can be much more elementary, usually only displaying information such as file size, file extension, when the file was created and who created the file.



What is Metadata?

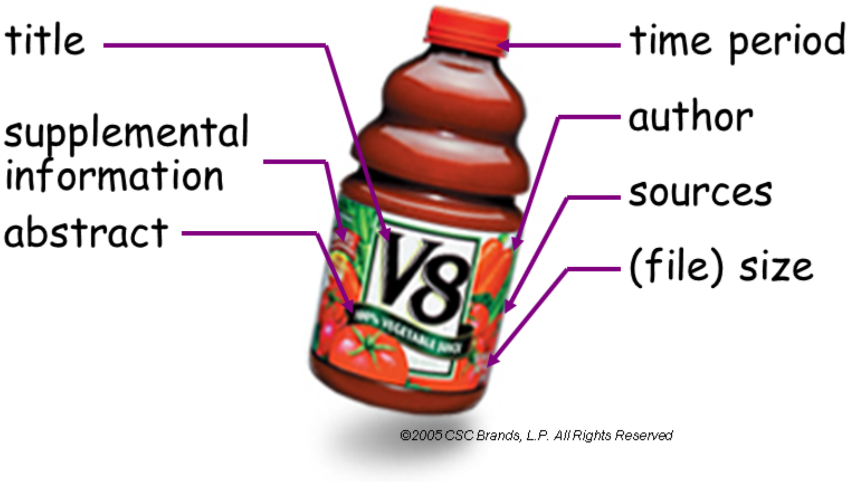
- **Data 'reporting'**
  - **WHO** created the data?
  - **WHAT** is the content of the data?
  - **WHEN** was it created?
  - **WHERE** is it geographically?
  - **HOW** was the data developed?
  - **WHY** was the data developed?



Definition

- Metadata:
  - Refers to data about the meaning, content, organization, or purpose of data.
  - Metadata may be as simple as a relational schema or as complicated as information describing the source, derivation, units, accuracy, and history of individual data items.
  - set
  - explain the role of metadata in the National Spatial Data Infrastructure

What is Metadata ?



title

supplemental information

abstract

time period

author

sources

(file) size

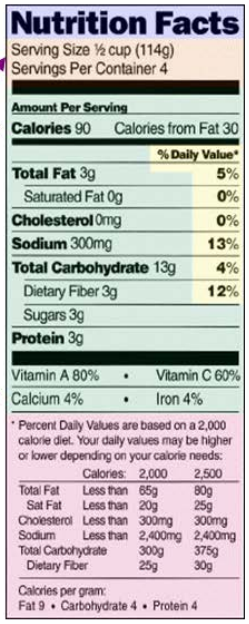
©2005 CSC Brands, L.P. All Rights Reserved

What is Metadata?

©2005 CSC Brands, L.P. All Rights Reserved

•entity

•attributes



Nutrition Facts

Serving Size ½ cup (114g)
Servings Per Container 4

Amount Per Serving

Calories 90 Calories from Fat 30	
	% Daily Value*
Total Fat 3g	**5%**
Saturated Fat 0g	**0%**
Cholesterol 0mg	**0%**
Sodium 300mg	**13%**
Total Carbohydrate 13g	**4%**
Dietary Fiber 3g	**12%**
Sugars 3g	
Protein 3g	

Vitamin A 80% • Vitamin C 60%
Calcium 4% • Iron 4%

* Percent Daily Values are based on a 2,000 calorie diet. Your daily values may be higher or lower depending on your calorie needs:

	Calories:	2,000	2,500
Total Fat	Less than	65g	80g
Sat Fat	Less than	20g	25g
Cholesterol	Less than	300mg	300mg
Sodium	Less than	2,400mg	2,400mg
Total Carbohydrate		300g	375g
Dietary Fiber		25g	30g

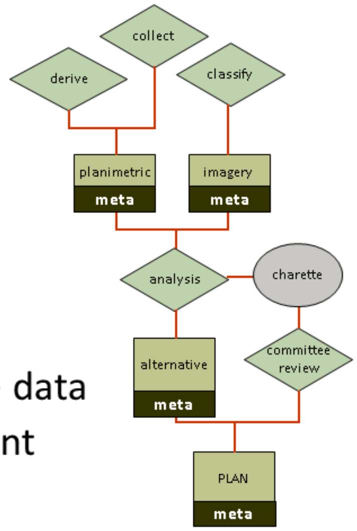
Calories per gram:
Fat 9 • Carbohydrate 4 • Protein 4

Types of Metadata

- System catalogs metadata
  - e.g. catalogs maintained by relational databases
- Relationship metadata
  - Information about relationship between data entities
- Content metadata
  - descriptions of the contents of stored data at an arbitrary granule.
- Data lineage metadata
  - it includes information about the creation of data
- Technical metadata
  - is technical information about stored data. It includes such information as the format
- Data usage metadata

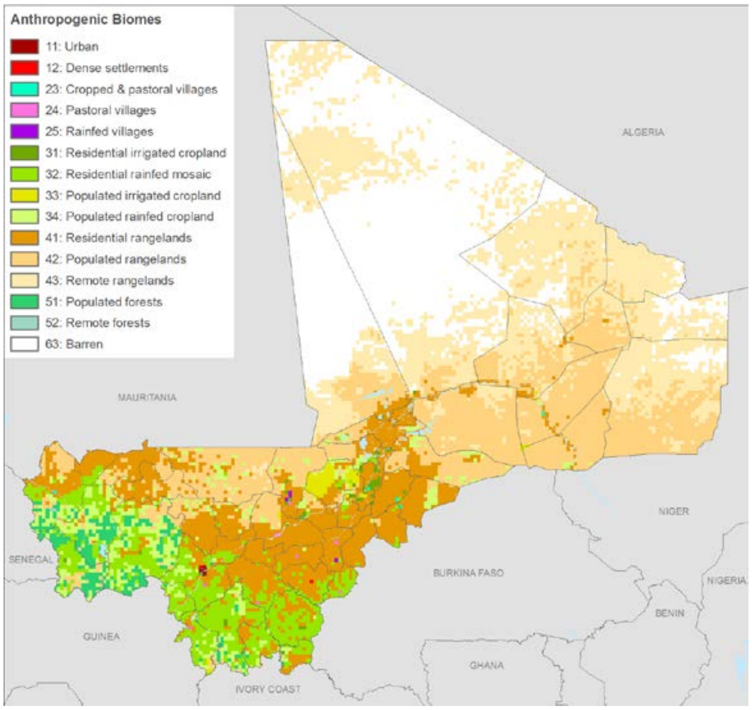
Utility of Metadata

- Metadata can support:
  - data distribution
  - data management
  - project management
- If it is:
  - considered a component of the data
  - created during data development
  - populated with rich content



Exercise 2. Writing metadata for spatial data

Anthropogenic Biomes Metadata Sheet



Title:	Anthropogenic Biomes of the World, v1 (2001-2006)
Indicator Code:	ANTH
Component:	*Adaptive Capacity*
Rationale:	*Corresponds to the field 'Rationale' in Indicator sheets in methodology. Include brief description of the indicator in the context of:* ▪ *why the issue is important globally;* ▪ *what are some of the impacts of the issue (these two may overlap);* ▪ *what the results of the indicator show; and how they can be interpreted.*
Source Data Set:	Anthropogenic Biomes of the World, v1 (2001-2006) The Anthropogenic Biomes of the World, Version 1 data set describes globally-significant ecological patterns within the terrestrial biosphere caused by sustained direct human interaction with ecosystems, including agriculture, urbanization, forestry and other land uses. Anthropogenic biomes were identified and mapped using a multi-stage procedure based on population (urban, non-urban), land use (percent area of pasture, crops, irrigation, rice and urban land) and land cover (percent area of trees and bare earth). Input datasets include population data (Landscan 2005), land use data (percent area under cropland, pasture, irrigation and rice) and land cover (vegetation, continuous fields, tree cover, and bare earth proportion layers). Dataset released in 2009. Citation: Ellis, E. C. and N. Ramankutty. 2008. Putting people in the map: anthropogenic biomes of the world. Data distributed by the Socioeconomic Data and Applications Center (SEDAC): http://sedac.ciesin.columbia.edu/es/anthropogenicbiomes.html. Accessed 9/6/2013.

Units:	Biome analysis was conducted at 5 arc minute resolution (5’ grid cells cover ~ 86 km^2 at the equator), a spatial resolution selected as the finest allowing direct use of high-quality land-use area estimates.
Computation:	The data were subset to the Mali national boundary extent using ArcGIS Extract by Mask tool and a 30 arc-second raster mask generated from a 30 arc second fishnet. Raster values were extracted using ArcGIS Extract Values to Points tool and the 30 arc-second fishnet centroids. The output was exported to a .csv table for re-coding and statistical analysis.
Statistics for raw data:	*Min = 11, Max=63, Median=43, Mean=47.15, Standard Deviation=10.62*
Scoring system:	*Raster values were re-coded as follows:* 11 (urban) = 10 12 (dense settlements) = 15 22 (irrigated villages) = 15 [not present in Mali] 23 (cropped pastoral villages) = 20 24 (pastoral villages) = 30 25 (rainfed villages) = 30 26 (rainfed mosaic villages) = 30 [not present in Mali] 31 (residential irrigated cropland) = 20 32 (residential rainfed mosaic) = 30 33 (populated irrigated cropland) = 20 34 (populated rainfed cropland) = 30 35 (remote croplands) = 50 [not present in Mali] 41 (residential rangelands) = 50 42 (populated rangelands) = 60 43 (remote rangelands) = 80 51 (populated forests) = 30 52 (remote forests) = 50 61 (wild forests) = 20 [not present in Mali] 62 (sparse trees) = 30 [not present in Mali] 63 (barren) = 100
Statistics for transformed data:	*After recode: Min=10, Max=100, Median=80, Mean=70.26, Standard Deviation=24.24*
Limitations:	The data set is a conceptual model and not intended to replace existing biome systems based on climate, terrain, and geology. Rather it is intended that wide availability of an anthropogenic biome system will encourage a richer view of human–ecosystem interactions across the terrestrial biosphere, and that this will, in turn, guide our investigation, understanding, and management of ecosystem processes and their changes at global and regional scales.
Spatial Extent:	Global
Spatial Resolution:	Raster cell sizes are 5 arc-minute or 0.08333 degree decimal (about 10 kilometers at the equator)
Year of Publication:	2009
Time Period:	2001–2006

Additional Notes:	
Date:	Obtained: 9/6/2013
Format:	Raster data available in GeoTiff and Esri Grid formats. The Africa Esri Grid was downloaded for this analysis.
File Name:	\\Dataserver0\arcc\VAs\Mali\VulnerabilityMapping\data\AnthropogenicBiomes\af_anthrome_ESRI grid\af_anthrome
Contact person:	N/A, downloaded from http://sedac.ciesin.columbia.edu/es/anthropogenicbiomes.html
Contact details:	N/A.

Exercise Task: Given the metadata template provided, add as much information for each of the variables as possible following the sample metadata provided above.

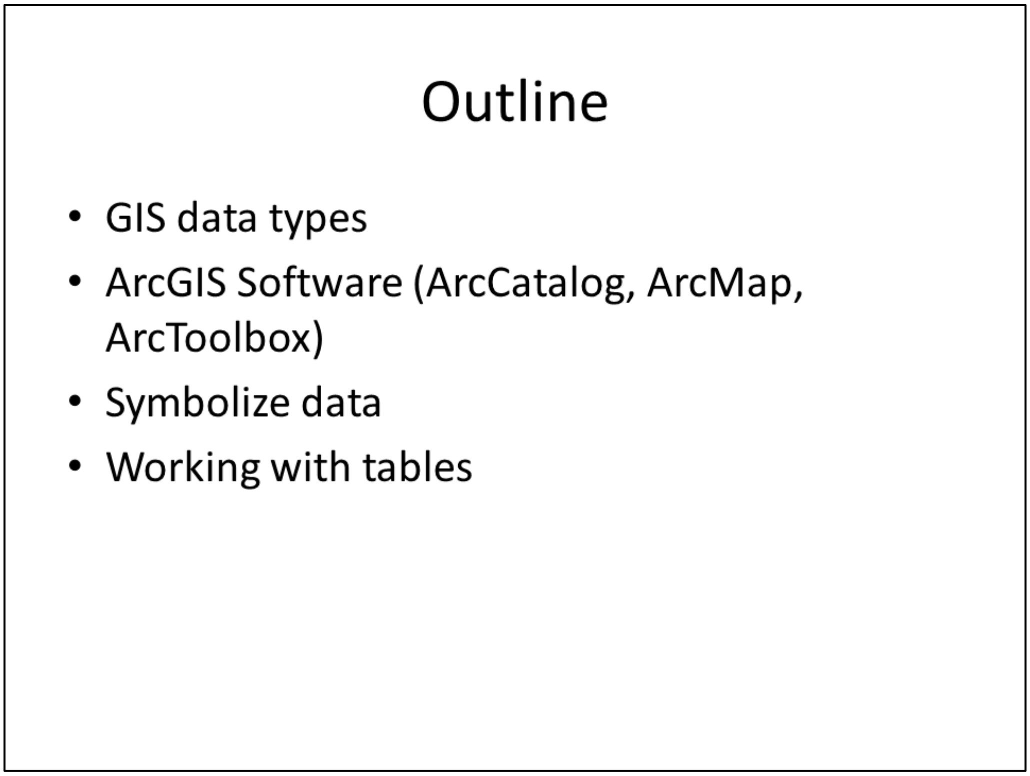
List of indicators for Vulnerability Mapping

Indicator Code	**Data Layer**
PRCP	Average annual precipitation (1950-2009)
TTREND	Long-term trend in temperature in July-August-Sept. (1950-2009)
FLOOD	Flood frequency (1999-2007)
HHWL	Household wealth (2006)
STNT	Child stunting (2006)
IMR	Infant mortality rate (IMR) (2006)
CARB	Soil organic carbon/soil quality (1950-2005)
EDMO	Education level of mother (2006)
MARK	Market accessibility (travel time to major cities)
ANTH	Anthropogenic biomes (2000)
IRRI	Irrigated areas (area equipped for irrigation) (1990-2000)

Lecture 3. Introducing ArcGIS Tools

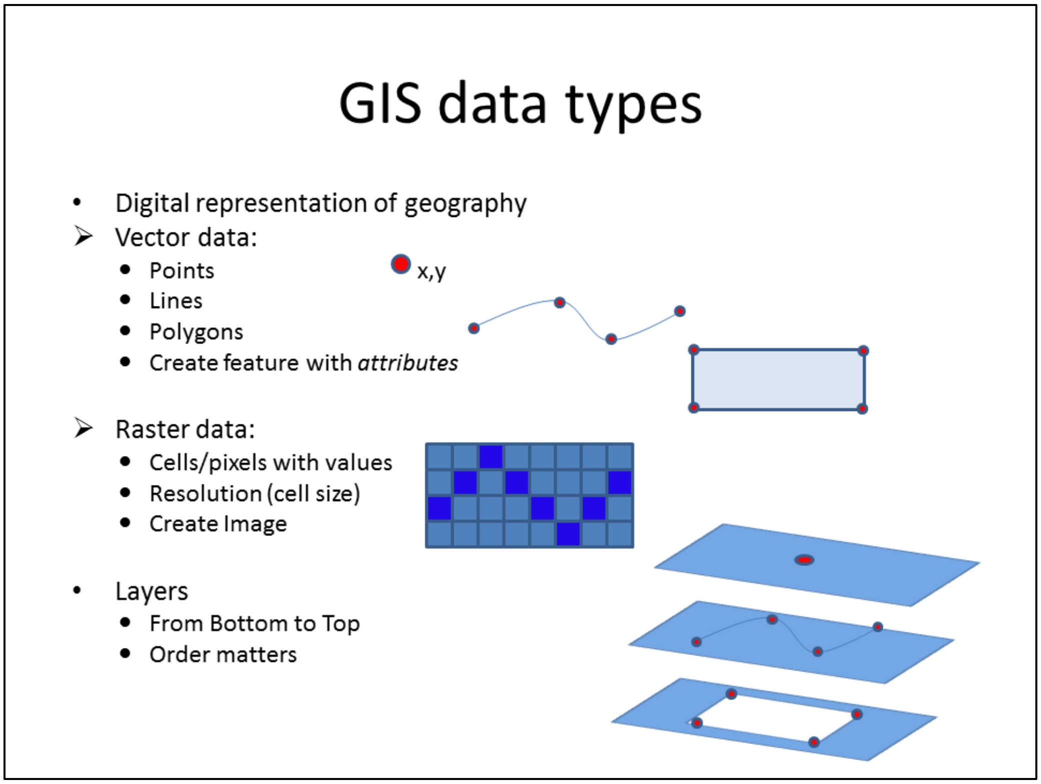
Introduction

This lecture is a refresher for those who may not have been regular users of the ArcGIS software. It will introduce the Graphic User Interface (GUI) and basic navigation to access the different tools that we may be using in the vulnerability assessment. Lecture will be followed by exercises to process the data sets in each category.



Outline

- GIS data types
- ArcGIS Software (ArcCatalog, ArcMap, ArcToolbox)
- Symbolize data
- Working with tables

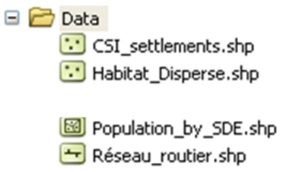


GIS data types

- Digital representation of geography
- Vector data:
  - Points
  - Lines
  - Polygons
  - Create feature with *attributes*
- Raster data:
  - Cells/pixels with values
  - Resolution (cell size)
  - Create Image
- Layers
  - From Bottom to Top
  - Order matters

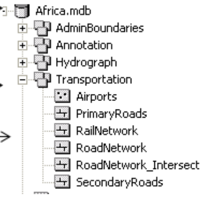
Vector Files

- Shapefile

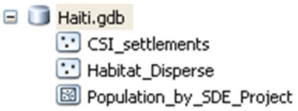


Data
CSI_settlements.shp
Habitat_Disperse.shp
Population_by_SDE.shp
Réseau_routier.shp

- Geodatabase
  - Feature dataset
  - Feature class

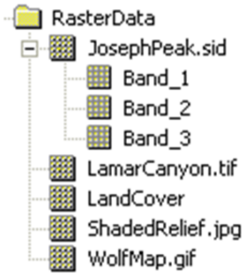


Africa.mdb
AdminBoundaries
Annotation
Hydrograph
Transportation
Airports
PrimaryRoads
RailNetwork
RoadNetwork
RoadNetwork_Intersect
SecondaryRoads



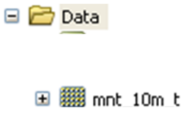
Haiti.gdb
CSI_settlements
Habitat_Disperse
Population_by_SDE_Project

Raster Files



RasterData
JosephPeak.sid
Band_1
Band_2
Band_3
LamarCanyon.tif
LandCover
ShadedRelief.jpg
WolfMap.gif

- Esri grid
- DEM (Digital Elevation Model)
- ERDAS IMAGINE
- MrSID
- TIFF
- JPEG
- GIF
- PNG



Data
mnt 10m t

Pros and Cons of Vector vs Raster

- Vector

-better for Cartography

-can have more than one attribute per Feature (columns of joined data)

Raster

-more powerful and better for spatial analysis calculations

-can only have only one value/ feature

-uses more memory

ArcGIS for Desktop

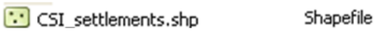
- Mostly use:
  - ArcCatalog (Manage)
  - ArcMap (Visualize)
  - ArcToolbox (Process)
- Can explore (3D):
  - ArcScene
  - ArcGlobe

ArcCatalog

- Manage your data

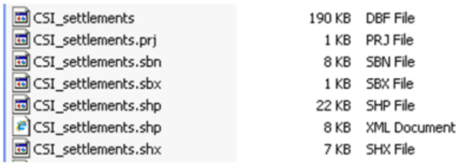
ArcCatalog

- ‘Windows’ for ArcGIS
- **Use ArcCatalog** to copy, delete, move, rename etc.
- In ArcCatalog GIS files are represented as one icon



CSI_settlements.shp	Shapefile

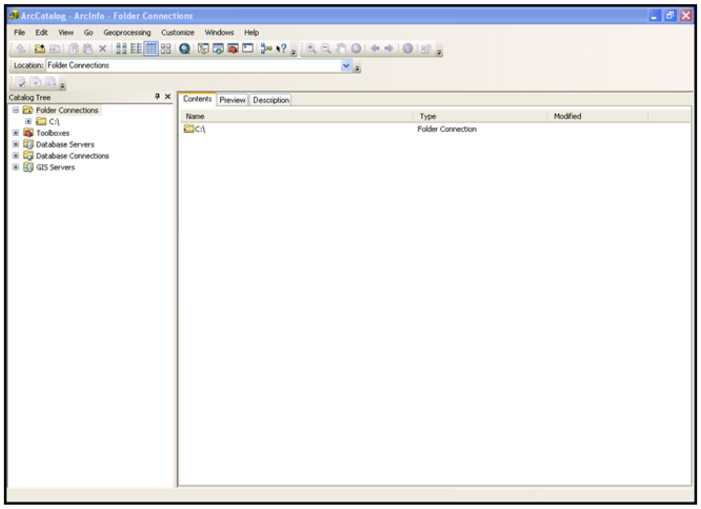
- Windows shows you the multiple files that create a single GIS file



CSI_settlements	190 KB	DBF File
CSI_settlements.prj	1 KB	PRJ File
CSI_settlements.sbn	8 KB	SBN File
CSI_settlements.sbx	1 KB	SBX File
CSI_settlements.shp	22 KB	SHP File
CSI_settlements.shp	8 KB	XML Document
CSI_settlements.shx	7 KB	SHX File

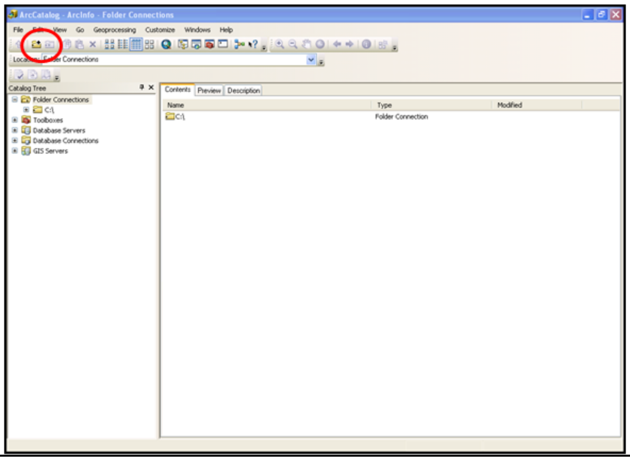
ArcCatalog

- Start>All Programs>ArcGIS>ArcCatalog 10

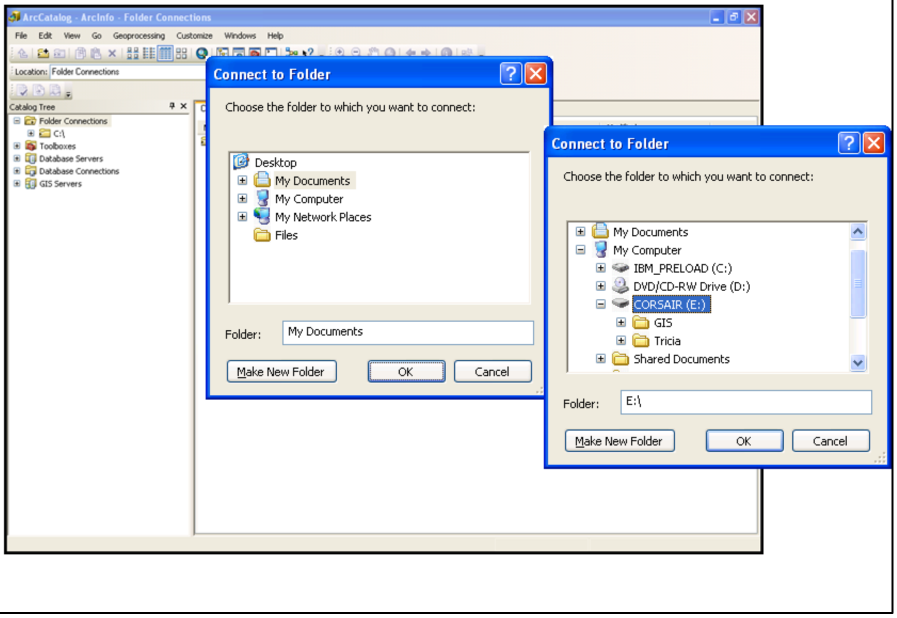

ArcCatalog - ArcInfo - Folder Connections
File Edit View Go Geoprocessing Customize Windows Help
Location: Folder Connections
Catalog Tree
Folder Connections
C:\
Toolboxes
Database Servers
Database Connections
GIS Servers
Contents Preview Description
Name Type Modified
C:\ Folder Connection

Connect To Folder

- Use Connect to folder if cannot find the drive where your data is stored (eg. E:\)


ArcCatalog - ArcInfo - Folder Connections
File Edit View Go Geoprocessing Customize Windows Help
Location: Folder Connections
Catalog Tree
Folder Connections
C:\
Toolboxes
Database Servers
Database Connections
GIS Servers
Contents Preview Description
Name Type Modified
C:\ Folder Connection

Connect To Folder



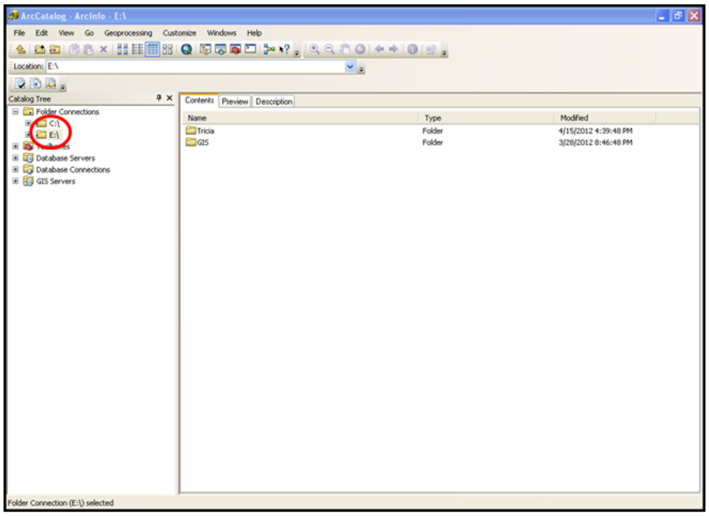
Choose the folder to which you want to connect:

Folder: My Documents

Choose the folder to which you want to connect:

Folder: E:\

Connect To Folder



Name	Type	Modified
Tricia	Folder	4/15/2012 4:39:48 PM
GIS	Folder	3/28/2012 8:46:48 PM

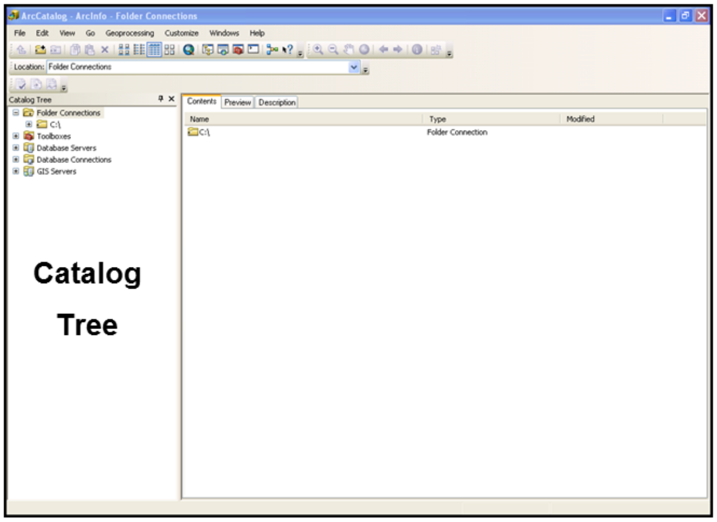
Folder Connection (E:\) selected

Connect To Folder Exercise

- Use Connect To Folder to find the **Sample** folder where you downloaded the sample data

ArcCatalog- Catalog Tree

- shows folders and layers



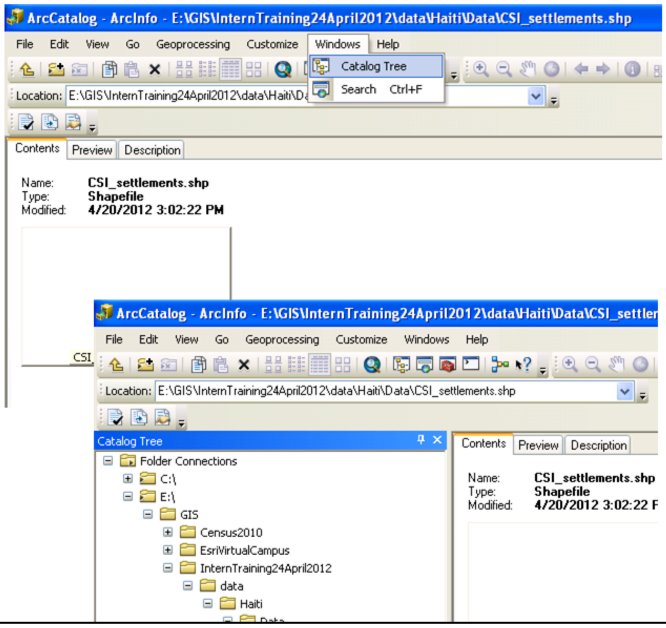
Location: Folder Connections

Name	Type	Modified
C:\	Folder Connection	

Catalog Tree

ArcCatalog- Catalog Tree

- If you close Catalog Tree by accident
- Go to Windows> Catalog Tree to bring it back



ArcCatalog - ArcInfo - E:\GIS\InternTraining24April2012\data\Haiti\Data\CSI_settlements.shp

Location: E:\GIS\InternTraining24April2012\data\Haiti\Da

Name: CSI_settlements.shp
Type: Shapefile
Modified: 4/20/2012 3:02:22 PM

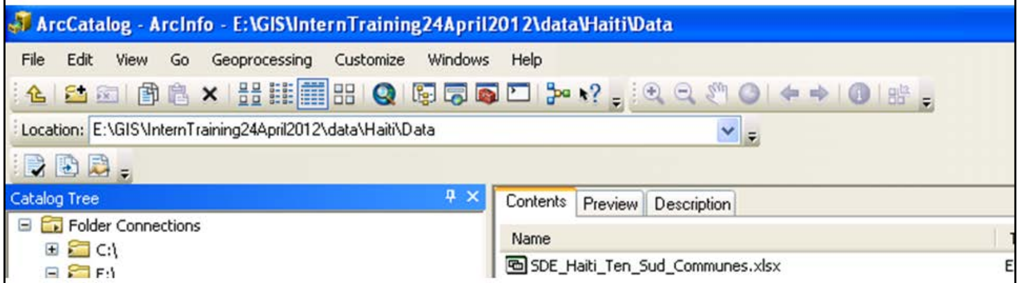
ArcCatalog - ArcInfo - E:\GIS\InternTraining24April2012\data\Haiti\Data\CSI_settler

Location: E:\GIS\InternTraining24April2012\data\Haiti\Data\CSI_settlements.shp

Name: CSI_settlements.shp
Type: Shapefile
Modified: 4/20/2012 3:02:22 F

ArcCatalog

- Quick exploration of your data
- 3 Tabs: Contents, Preview, Description



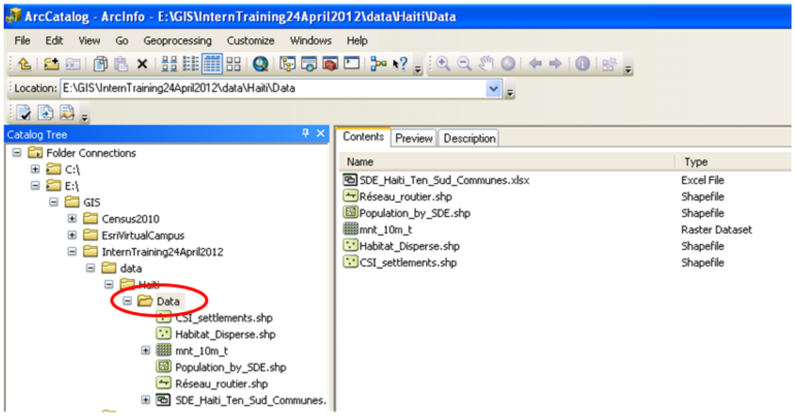
ArcCatalog - ArcInfo - E:\GIS\InternTraining24April2012\data\Haiti\Data

Location: E:\GIS\InternTraining24April2012\data\Haiti\Data

SDE_Haiti_Ten_Sud_Communes.xlsx

ArcCatalog- Contents Tab

- View Contents of folder



ArcCatalog - ArcInfo - E:\GIS\InternTraining24April2012\data\Haiti\Data

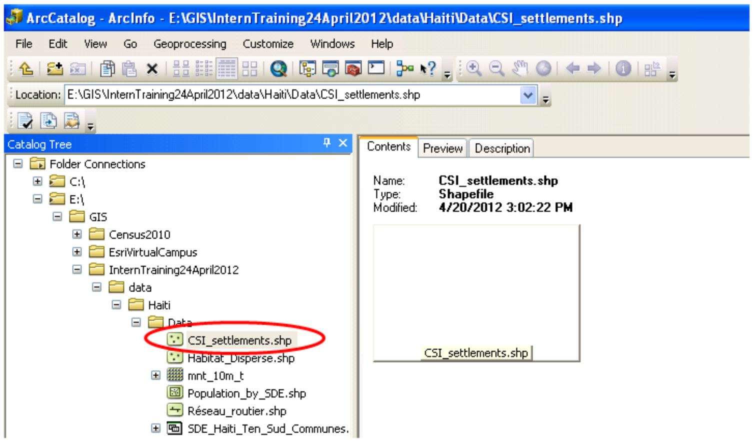
Location: E:\GIS\InternTraining24April2012\data\Haiti\Data

```
⊟ Folder Connections
  ⊞ C:\
  ⊟ E:\
    ⊟ GIS
      ⊞ Census2010
      ⊞ EsriVirtualCampus
      ⊟ InternTraining24April2012
        ⊟ data
          ⊟ Haiti
            ⊟ Data
                CSI_settlements.shp
                Habitat_Disperse.shp
              ⊞ mnt_10m_t
                Population_by_SDE.shp
                Réseau_routier.shp
              ⊞ SDE_Haiti_Ten_Sud_Communes.
```

Name	Type
SDE_Haiti_Ten_Sud_Communes.xlsx	Excel File
Réseau_routier.shp	Shapefile
Population_by_SDE.shp	Shapefile
mnt_10m_t	Raster Dataset
Habitat_Disperse.shp	Shapefile
CSI_settlements.shp	Shapefile

ArcCatalog – Contents Tab

- View data set details - Name, Type, Modified date



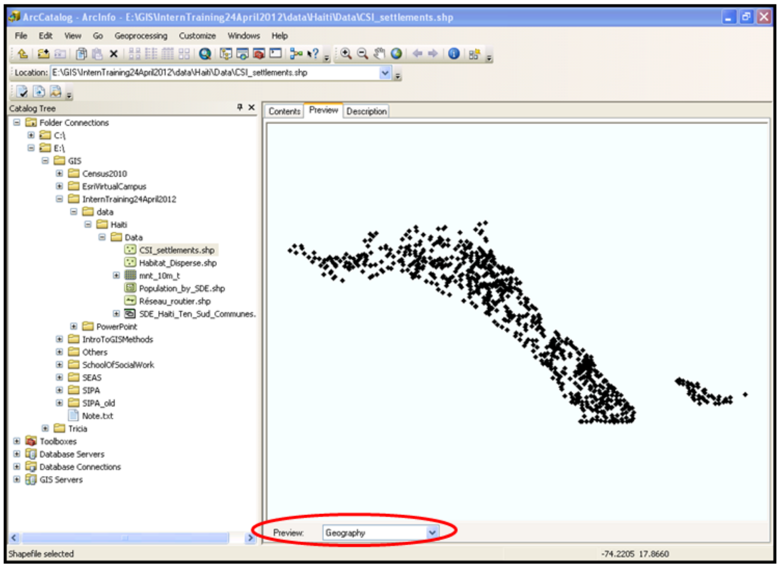
ArcCatalog - ArcInfo - E:\GIS\InternTraining24April2012\data\Haiti\Data\CSI_settlements.shp

Location: E:\GIS\InternTraining24April2012\data\Haiti\Data\CSI_settlements.shp

```
⊟ Folder Connections
  ⊞ C:\
  ⊟ E:\
    ⊟ GIS
      ⊞ Census2010
      ⊞ EsriVirtualCampus
      ⊟ InternTraining24April2012
        ⊟ data
          ⊟ Haiti
            ⊟ Data
                CSI_settlements.shp
                Habitat_Disperse.shp
              ⊞ mnt_10m_t
                Population_by_SDE.shp
                Réseau_routier.shp
              ⊞ SDE_Haiti_Ten_Sud_Communes.
```

Name: **CSI_settlements.shp**
Type: **Shapefile**
Modified: **4/20/2012 3:02:22 PM**

ArcCatalog – Preview Tab - Geography



ArcCatalog - ArcInfo - E:\GIS\InternTraining24April2012\data\Haiti\Data\CSI_settlements.shp

Location: E:\GIS\InternTraining24April2012\data\Haiti\Data\CSI_settlements.shp

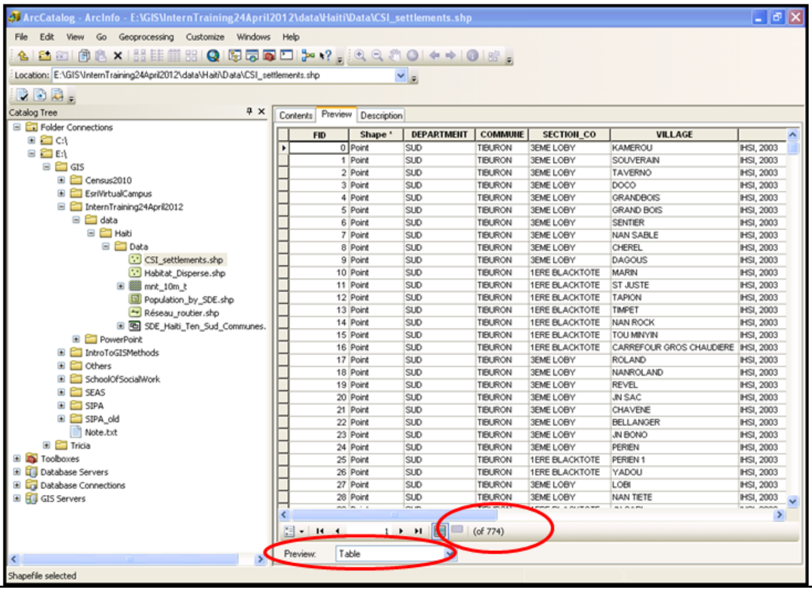
Preview: Geography

Shapefile selected

-74.2205 17.8660

ArcCatalog – Preview Tab - Table

- shows # of records (rows)



ArcCatalog - ArcInfo - E:\GIS\InternTraining24April2012\data\Haiti\Data\CSI_settlements.shp

Location: E:\GIS\InternTraining24April2012\data\Haiti\Data\CSI_settlements.shp

FID	Shape *	DEPARTMENT	COMMUNE	SECTION_CO	VILLAGE	
0	Point	SUD	TIBURON	3EME LOBY	KAMEROU	IHSI, 2003
1	Point	SUD	TIBURON	3EME LOBY	SOUVERAIN	IHSI, 2003
2	Point	SUD	TIBURON	3EME LOBY	TAVERNO	IHSI, 2003
3	Point	SUD	TIBURON	3EME LOBY	DOCO	IHSI, 2003
4	Point	SUD	TIBURON	3EME LOBY	GRANDBOIS	IHSI, 2003
5	Point	SUD	TIBURON	3EME LOBY	GRAND BOIS	IHSI, 2003
6	Point	SUD	TIBURON	3EME LOBY	SENTIER	IHSI, 2003
7	Point	SUD	TIBURON	3EME LOBY	NAN SABLE	IHSI, 2003
8	Point	SUD	TIBURON	3EME LOBY	CHEREL	IHSI, 2003
9	Point	SUD	TIBURON	3EME LOBY	DAGOUS	IHSI, 2003
10	Point	SUD	TIBURON	1ERE BLACKTOTE	MARIN	IHSI, 2003
11	Point	SUD	TIBURON	1ERE BLACKTOTE	ST JUSTE	IHSI, 2003
12	Point	SUD	TIBURON	1ERE BLACKTOTE	TAPION	IHSI, 2003
13	Point	SUD	TIBURON	1ERE BLACKTOTE	TIMPET	IHSI, 2003
14	Point	SUD	TIBURON	1ERE BLACKTOTE	NAN ROCK	IHSI, 2003
15	Point	SUD	TIBURON	1ERE BLACKTOTE	TOU MINYIN	IHSI, 2003
16	Point	SUD	TIBURON	1ERE BLACKTOTE	CARREFOUR GROS CHAUDIERE	IHSI, 2003
17	Point	SUD	TIBURON	3EME LOBY	ROLAND	IHSI, 2003
18	Point	SUD	TIBURON	3EME LOBY	NANROLAND	IHSI, 2003
19	Point	SUD	TIBURON	3EME LOBY	REVEL	IHSI, 2003
20	Point	SUD	TIBURON	3EME LOBY	JN SAC	IHSI, 2003
21	Point	SUD	TIBURON	3EME LOBY	CHAVENE	IHSI, 2003
22	Point	SUD	TIBURON	3EME LOBY	BELLANGER	IHSI, 2003
23	Point	SUD	TIBURON	3EME LOBY	JN BONO	IHSI, 2003
24	Point	SUD	TIBURON	3EME LOBY	PERIEN	IHSI, 2003
25	Point	SUD	TIBURON	1ERE BLACKTOTE	PERIEN 1	IHSI, 2003
26	Point	SUD	TIBURON	1ERE BLACKTOTE	YADOU	IHSI, 2003
27	Point	SUD	TIBURON	3EME LOBY	LOBI	IHSI, 2003
28	Point	SUD	TIBURON	3EME LOBY	NAN TIETE	IHSI, 2003

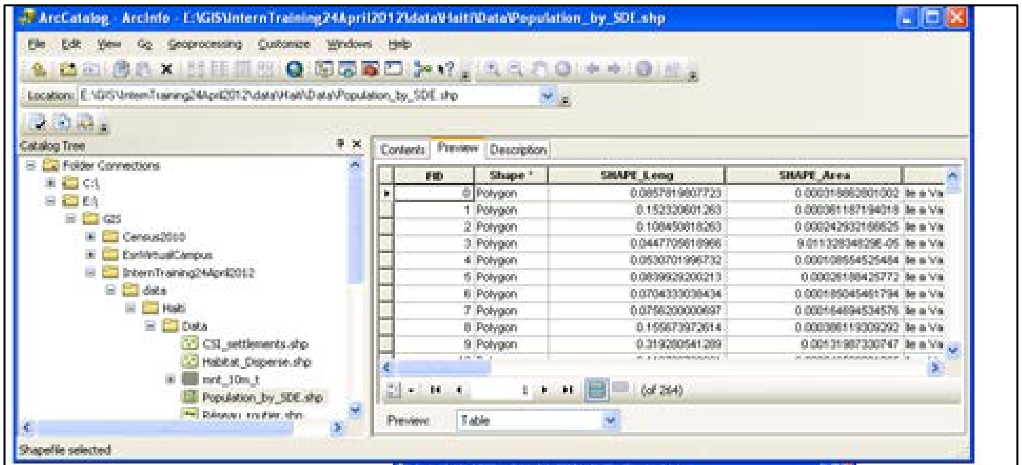
1 (of 774)

Preview: Table

Shapefile selected

ArcCatalog Questions

- How many records are in the "kenya_Admin5.shp" attribute table?
- Are there more records in the COAST Admin2 or CENTRAL?



ArcCatalog - ArcInfo - E:\GIS\InternTraining24April2012\data\Haiti\Data\Population_by_SDE.shp

Location: E:\GIS\InternTraining24April2012\data\Haiti\Data\Population_by_SDE.shp

FID	Shape *	SHAPE_Leng	SHAPE_Area	
0	Polygon	0.0857819807723	0.000318863801002	Ile a Va
1	Polygon	0.152320601263	0.000361187194018	Ile a Va
2	Polygon	0.108450818263	0.000242932168625	Ile a Va
3	Polygon	0.0447705618966	9.01132834829E-05	Ile a Va
4	Polygon	0.0530701996732	0.000108554525484	Ile a Va
5	Polygon	0.0839929200213	0.00026188425772	Ile a Va
6	Polygon	0.0704333038434	0.000185045461794	Ile a Va
7	Polygon	0.0756200000697	0.000164894534576	Ile a Va
8	Polygon	0.155673973614	0.000386119309292	Ile a Va
9	Polygon	0.319280541289	0.00131987330747	Ile a Va

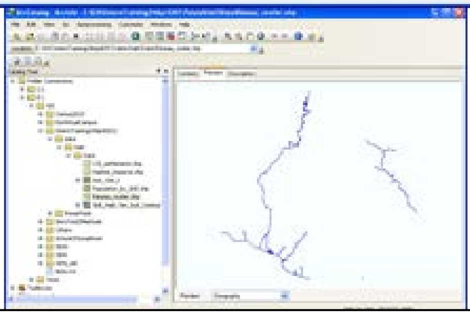
1 (of 264)

Preview: Table

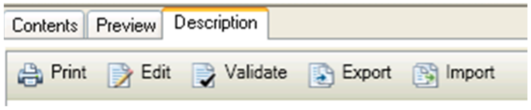
Shapefile selected

Answers:

- 1152
- CENTRAL

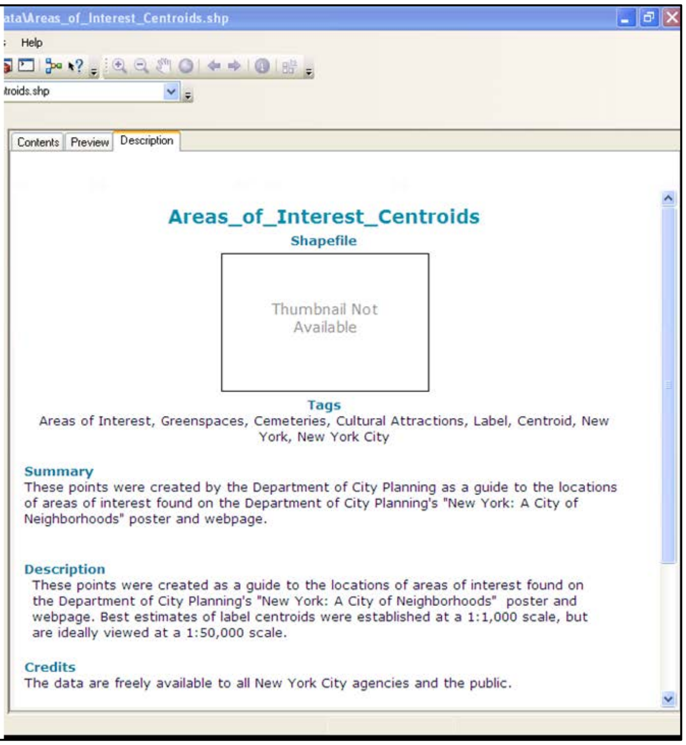


ArcCatalog – Description Tab



- Has Metadata – Data about data
- Contains: Summary, Spatial Information (Coordinate Reference), field and subtypes etc.
- It is important to document the data especially if you plan on sharing data with others
- Creates a "Data Dictionary"

- Metadata Example



Areas_of_Interest_Centroids

Shapefile

Thumbnail Not Available

Tags

Areas of Interest, Greenspaces, Cemeteries, Cultural Attractions, Label, Centroid, New York, New York City

Summary

These points were created by the Department of City Planning as a guide to the locations of areas of interest found on the Department of City Planning's "New York: A City of Neighborhoods" poster and webpage.

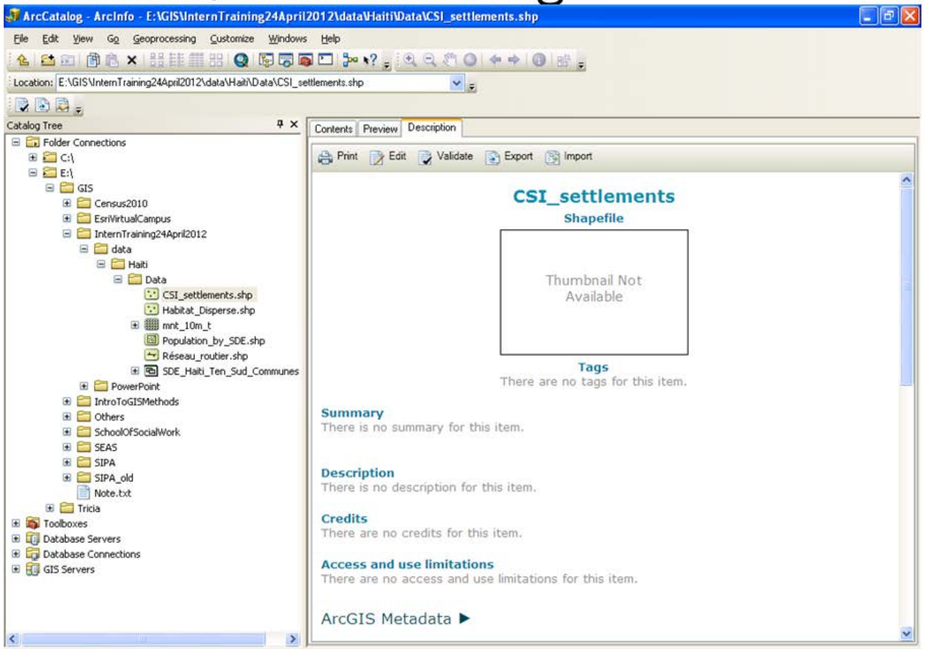
Description

These points were created as a guide to the locations of areas of interest found on the Department of City Planning's "New York: A City of Neighborhoods" poster and webpage. Best estimates of label centroids were established at a 1:1,000 scale, but are ideally viewed at a 1:50,000 scale.

Credits

The data are freely available to all New York City agencies and the public.

ArcCatalog - Metadata



ArcCatalog - ArcInfo - E:\GIS\InternTraining24April2012\data\Haiti\Data\CSI_settlements.shp

Location: E:\GIS\InternTraining24April2012\data\Haiti\Data\CSI_settlements.shp

CSI_settlements

Shapefile

Thumbnail Not Available

Tags

There are no tags for this item.

Summary

There is no summary for this item.

Description

There is no description for this item.

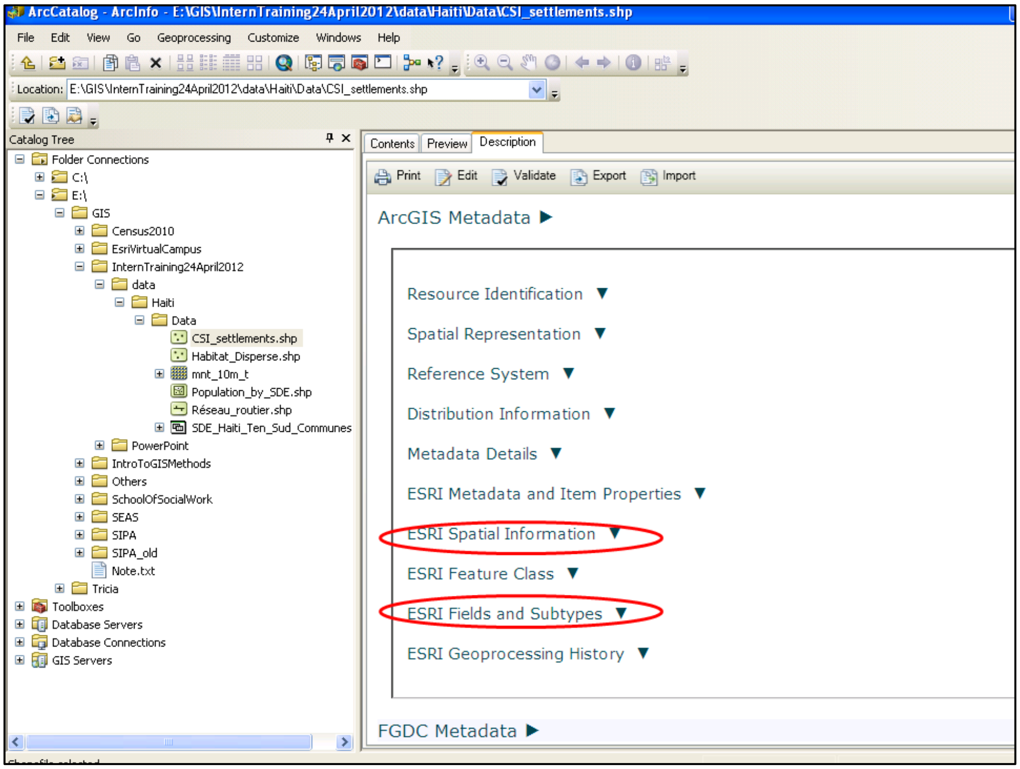
Credits

There are no credits for this item.

Access and use limitations

There are no access and use limitations for this item.

ArcGIS Metadata ▶



ArcCatalog - ArcInfo - E:\GIS\InternTraining24April2012\data\Haiti\Data\CSI_settlements.shp

Location: E:\GIS\InternTraining24April2012\data\Haiti\Data\CSI_settlements.shp

ArcGIS Metadata ▶

Resource Identification ▼

Spatial Representation ▼

Reference System ▼

Distribution Information ▼

Metadata Details ▼

ESRI Metadata and Item Properties ▼

ESRI Spatial Information ▼

ESRI Feature Class ▼

ESRI Fields and Subtypes ▼

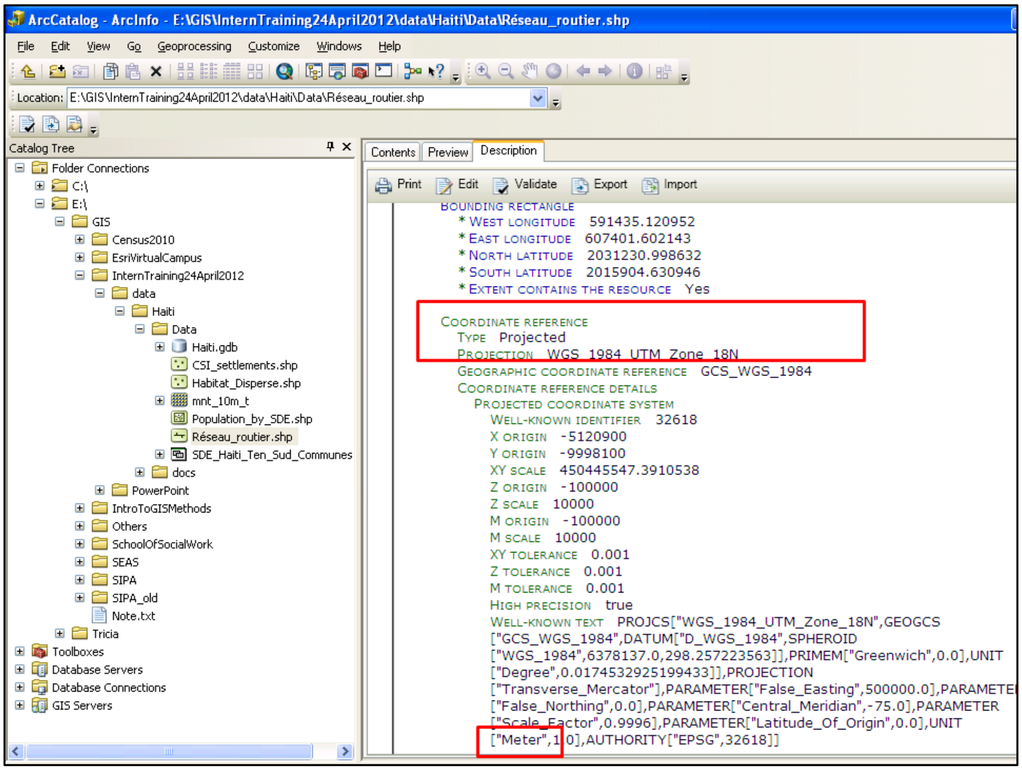
ESRI Geoprocessing History ▼

FGDC Metadata ▶

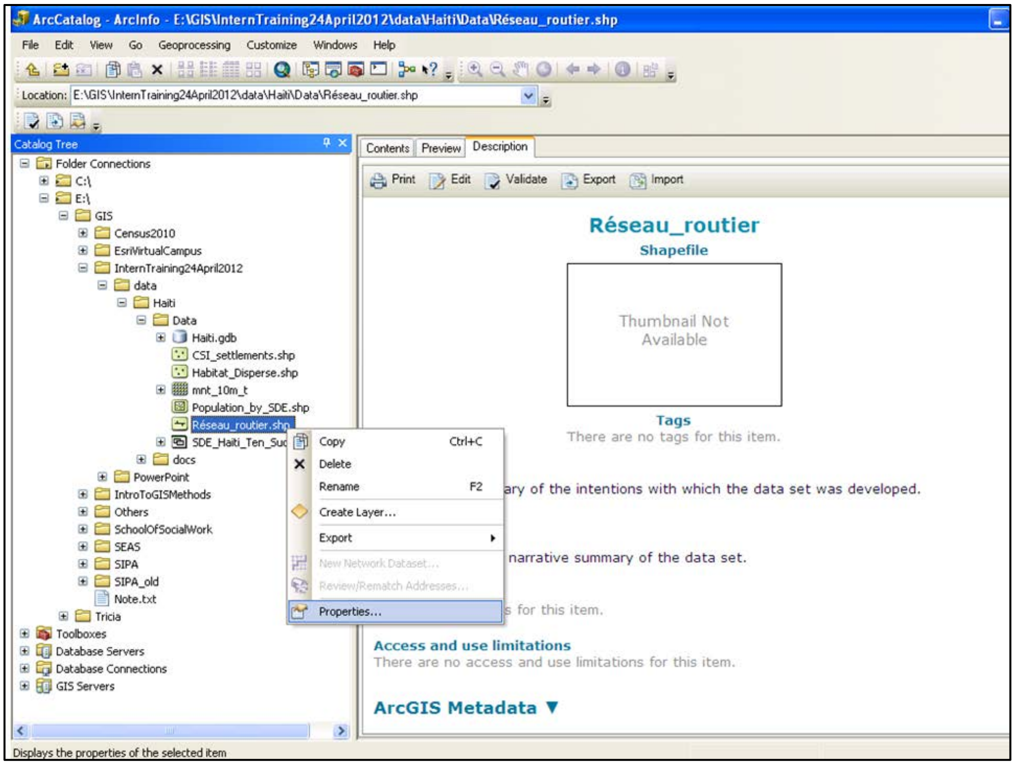
Finding Projection and units Exercise

- Look at kenya_populated_locality_projected.shp layer properties

- Question:
- What is the Coordinate Reference Type?
- What is the Projection?
- What are the units of this projection?



```
Bounding rectangle
  * West longitude   591435.120952
  * East longitude   607401.602143
  * North latitude   2031230.998632
  * South latitude   2015904.630946
  * Extent contains the resource   Yes

Coordinate reference
  Type   Projected
  Projection   WGS_1984_UTM_Zone_18N
  Geographic coordinate reference   GCS_WGS_1984
  Coordinate reference details
    Projected coordinate system
      Well-known identifier   32618
      X origin   -5120900
      Y origin   -9998100
      XY scale   450445547.3910538
      Z origin   -100000
      Z scale   10000
      M origin   -100000
      M scale   10000
      XY tolerance   0.001
      Z tolerance   0.001
      M tolerance   0.001
      High precision   true
      Well-known text   PROJCS["WGS_1984_UTM_Zone_18N",GEOGCS
      ["GCS_WGS_1984",DATUM["D_WGS_1984",SPHEROID
      ["WGS_1984",6378137.0,298.257223563]],PRIMEM["Greenwich",0.0],UNIT
      ["Degree",0.0174532925199433]],PROJECTION
      ["Transverse_Mercator"],PARAMETER["False_Easting",500000.0],PARAMETE
      ["False_Northing",0.0],PARAMETER["Central_Meridian",-75.0],PARAMETER
      ["Scale_Factor",0.9996],PARAMETER["Latitude_Of_Origin",0.0],UNIT
      ["Meter",1.0],AUTHORITY["EPSG",32618]]
```



ArcCatalog - ArcInfo - E:\GIS\InternTraining24April2012\data\Haiti\Data\Réseau_routier.shp

Location: E:\GIS\InternTraining24April2012\data\Haiti\Data\Réseau_routier.shp

Réseau_routier

Shapefile

Thumbnail Not Available

Tags

There are no tags for this item.

ary of the intentions with which the data set was developed.

narrative summary of the data set.

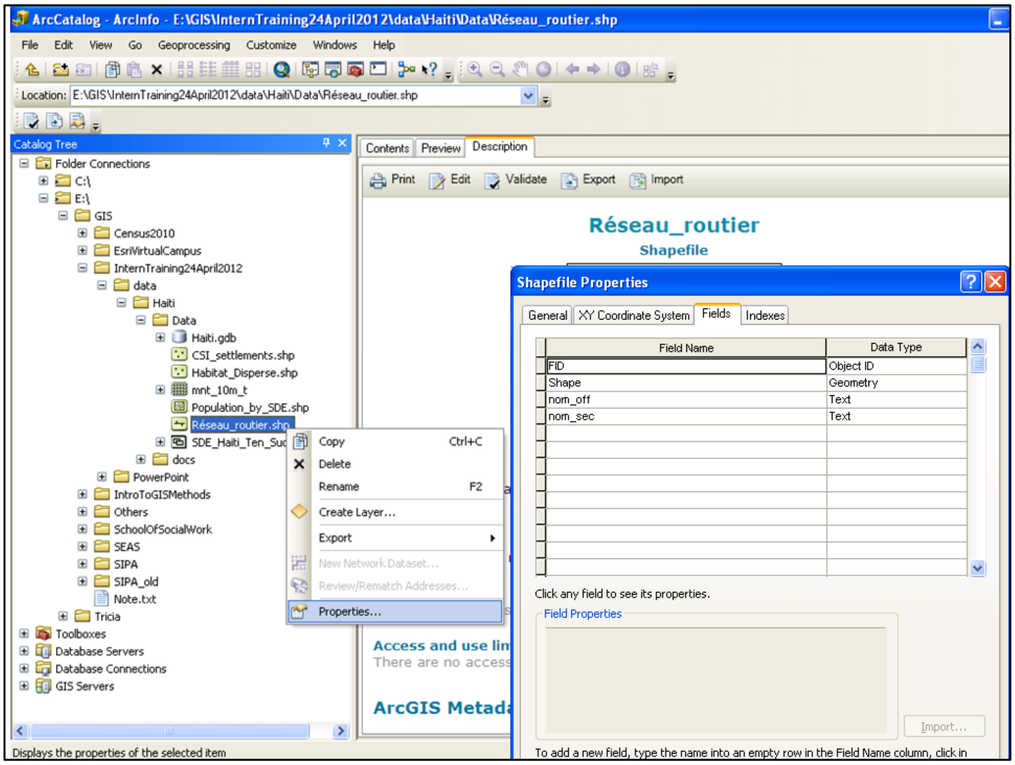
s for this item.

Access and use limitations

There are no access and use limitations for this item.

ArcGIS Metadata ▼

Displays the properties of the selected item



ArcCatalog - ArcInfo - E:\GIS\InternTraining24April2012\data\Haiti\Data\Réseau_routier.shp

Location: E:\GIS\InternTraining24April2012\data\Haiti\Data\Réseau_routier.shp

Réseau_routier

Shapefile

Access and use lim

There are no access

ArcGIS Metad

Displays the properties of the selected item

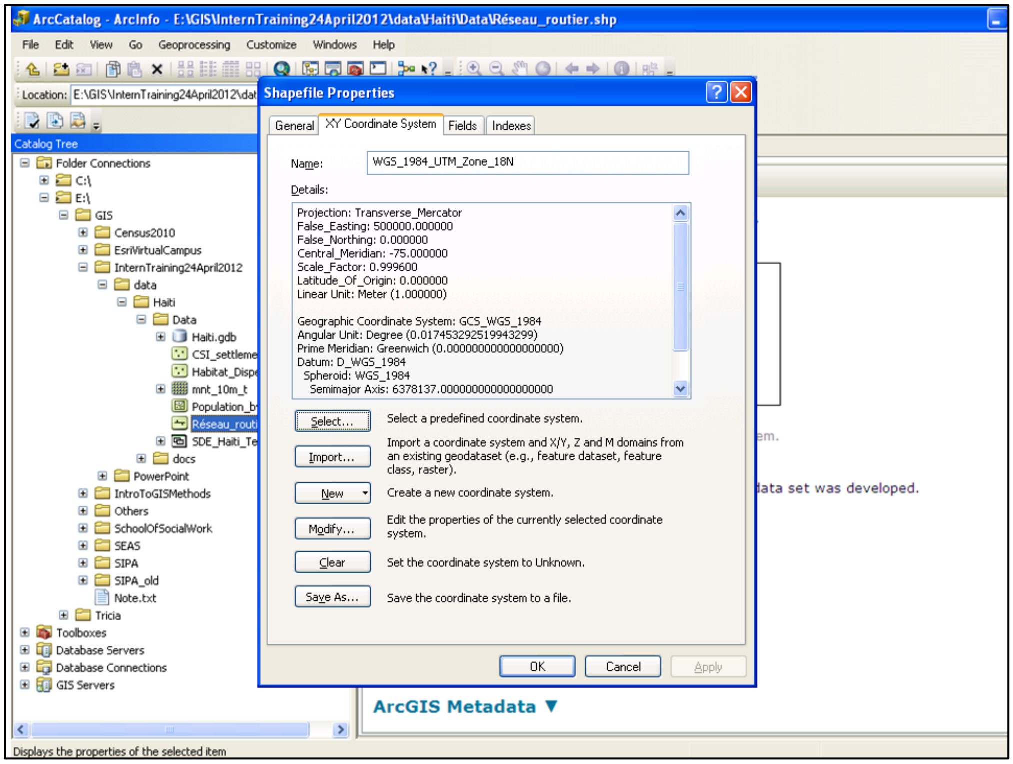
Shapefile Properties

Field Name	Data Type
FID	Object ID
Shape	Geometry
nom_off	Text
nom_sec	Text

Click any field to see its properties.

Field Properties

To add a new field, type the name into an empty row in the Field Name column, click in



ArcCatalog - ArcInfo - E:\GIS\InternTraining24April2012\data\Haiti\Data\Réseau_routier.shp

Location: E:\GIS\InternTraining24April2012\dat

Shapefile Properties

Name: WGS_1984_UTM_Zone_18N

Details:

```
Projection: Transverse_Mercator
False_Easting: 500000.000000
False_Northing: 0.000000
Central_Meridian: -75.000000
Scale_Factor: 0.999600
Latitude_Of_Origin: 0.000000
Linear Unit: Meter (1.000000)

Geographic Coordinate System: GCS_WGS_1984
Angular Unit: Degree (0.017453292519943299)
Prime Meridian: Greenwich (0.000000000000000000)
Datum: D_WGS_1984
  Spheroid: WGS_1984
   Semimajor Axis: 6378137.000000000000000000
```

Select... Select a predefined coordinate system.

Import... Import a coordinate system and X/Y, Z and M domains from an existing geodataset (e.g., feature dataset, feature class, raster).

New Create a new coordinate system.

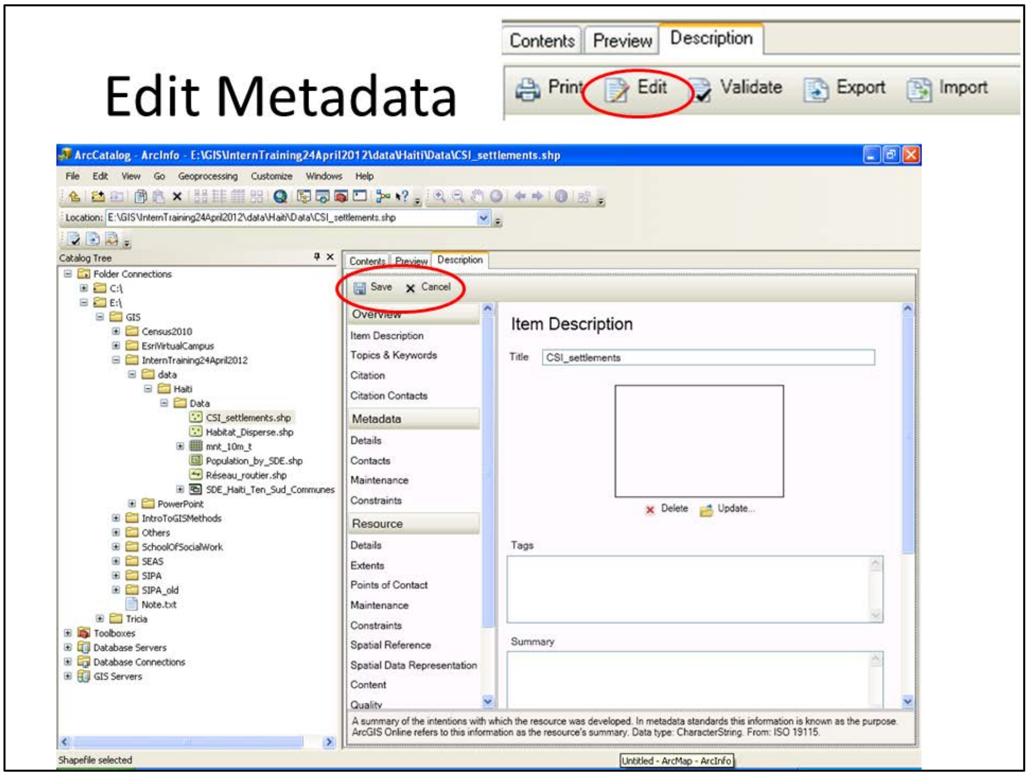
Modify... Edit the properties of the currently selected coordinate system.

Clear Set the coordinate system to Unknown.

Save As... Save the coordinate system to a file.

Displays the properties of the selected item

Edit Metadata



ArcCatalog - ArcInfo - E:\GIS\InternTraining24April2012\data\Haiti\Data\CSI_settlements.shp

Location: E:\GIS\InternTraining24April2012\data\Haiti\Data\CSI_settlements.shp

Item Description

Title: CSI_settlements

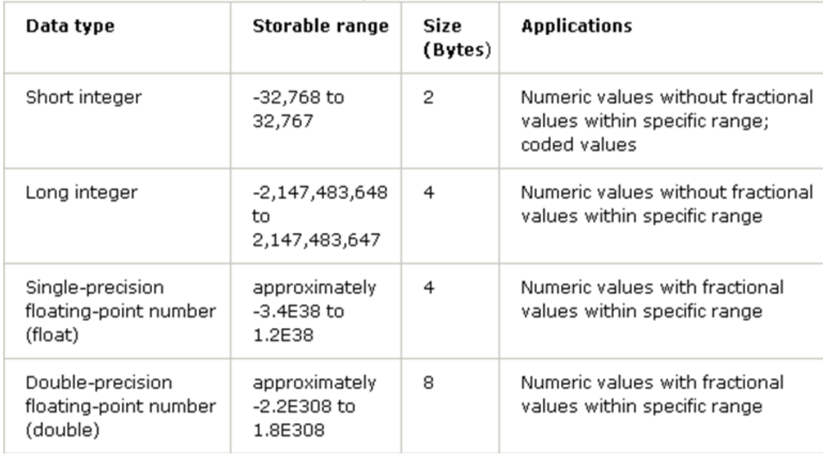
A summary of the intentions with which the resource was developed. In metadata standards this information is known as the purpose. ArcGIS Online refers to this information as the resource's summary. Data type: CharacterString. From: ISO 19115.

Shapefile selected

Untitled - ArcMap - ArcInfo

Types of Attribute Data

- String – text
- Date
- Precision – Maximum length
- Scale – Max. Decimal places



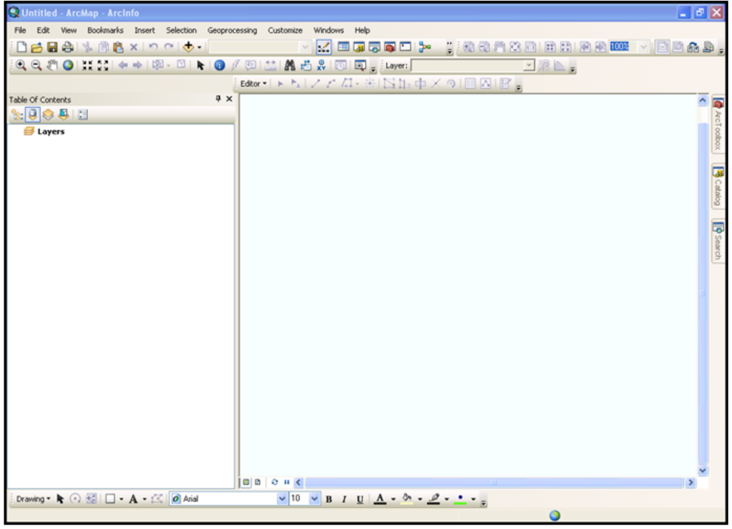
Data type	Storable range	Size (Bytes)	Applications
Short integer	-32,768 to 32,767	2	Numeric values without fractional values within specific range; coded values
Long integer	-2,147,483,648 to 2,147,483,647	4	Numeric values without fractional values within specific range
Single-precision floating-point number (float)	approximately -3.4E38 to 1.2E38	4	Numeric values with fractional values within specific range
Double-precision floating-point number (double)	approximately -2.2E308 to 1.8E308	8	Numeric values with fractional values within specific range

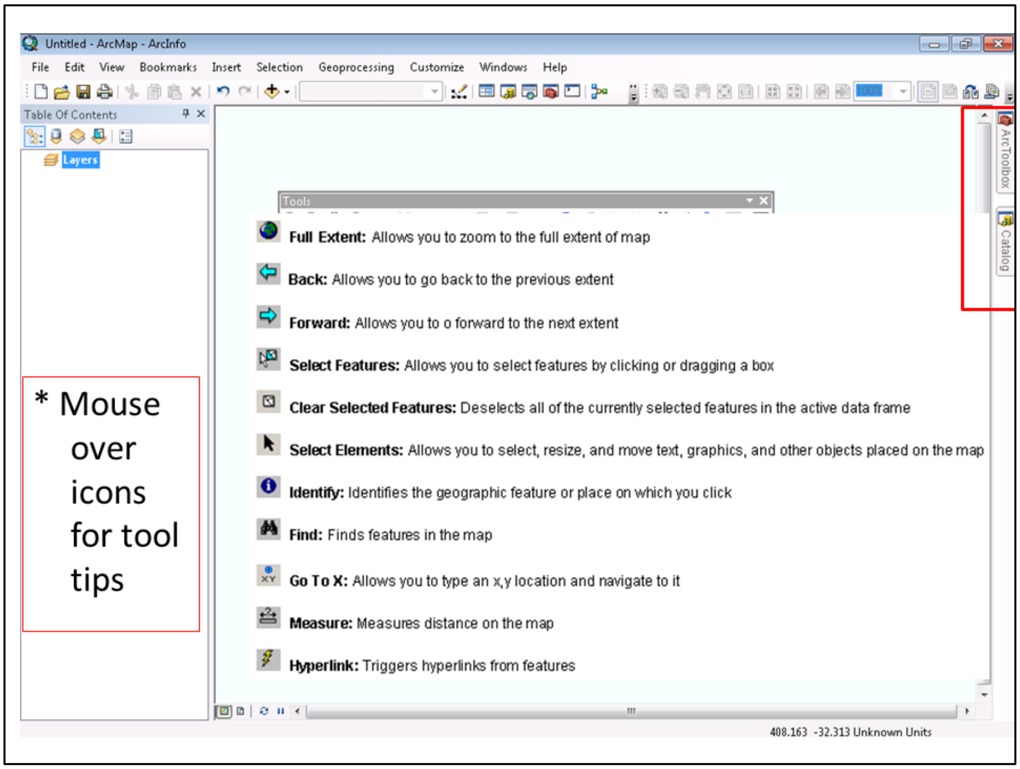
ArcMap

- Explore & Analyze data
- Create maps in a Map Document
- Save Map documents as _____.mxd

ArcMap

- Programs>ArcGIS>ArcMap


Untitled - ArcMap - ArcInfo
File Edit View Bookmarks Insert Selection Geoprocessing Customize Windows Help
Layer:
Editor
Table Of Contents
Layers
ArcToolbox
Catalog
Search
Drawing
Arial
10
B I U A


Untitled - ArcMap - ArcInfo
File Edit View Bookmarks Insert Selection Geoprocessing Customize Windows Help
Table Of Contents
Layers
Tools
ArcToolbox
Catalog
408.163 -32.313 Unknown Units

Full Extent: Allows you to zoom to the full extent of map

Back: Allows you to go back to the previous extent

Forward: Allows you to o forward to the next extent

Select Features: Allows you to select features by clicking or dragging a box

Clear Selected Features: Deselects all of the currently selected features in the active data frame

Select Elements: Allows you to select, resize, and move text, graphics, and other objects placed on the map

Identify: Identifies the geographic feature or place on which you click

Find: Finds features in the map

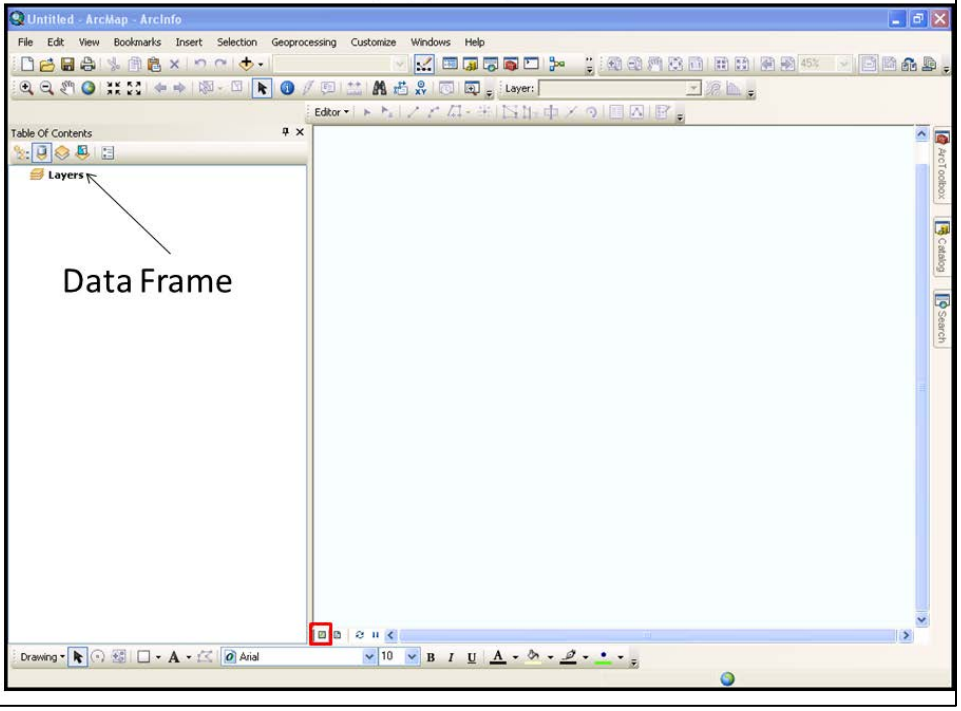
Go To X: Allows you to type an x,y location and navigate to it

Measure: Measures distance on the map

Hyperlink: Triggers hyperlinks from features

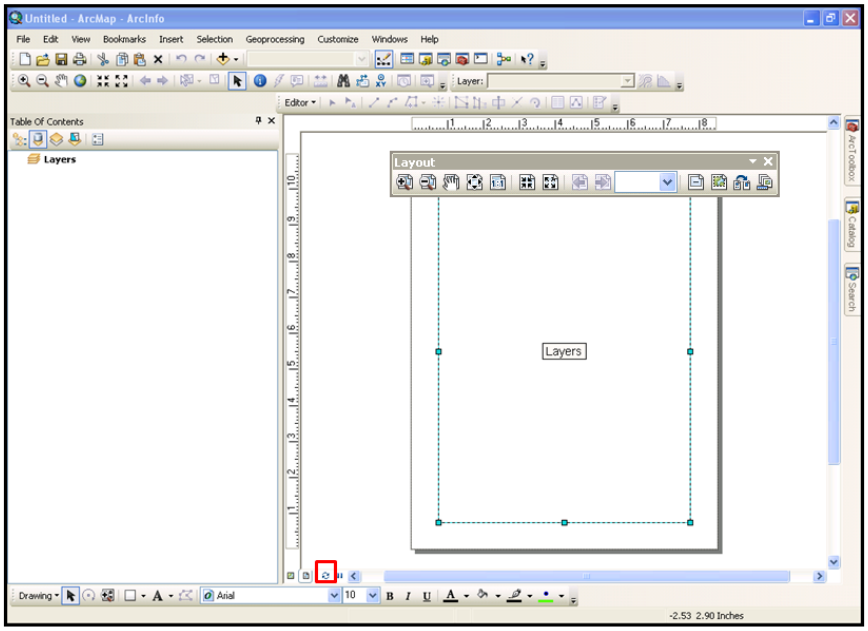
* Mouse over icons for tool tips

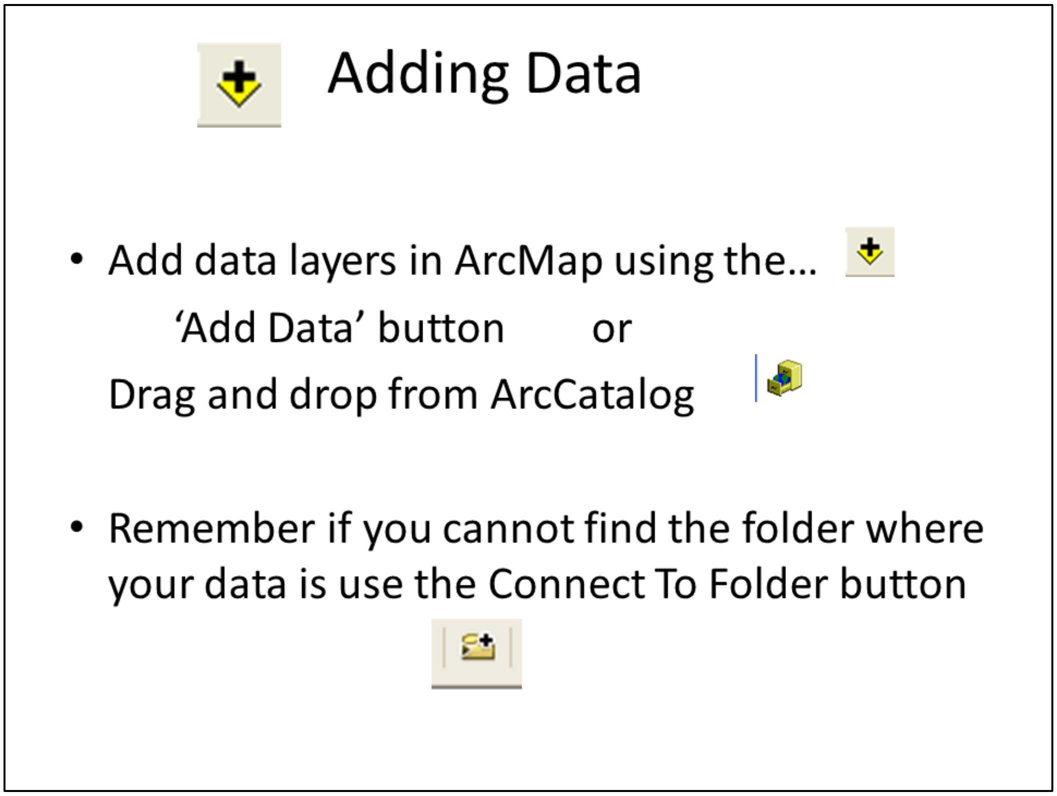
DataView


Untitled - ArcMap - ArcInfo
File Edit View Bookmarks Insert Selection Geoprocessing Customize Windows Help
Layer:
Editor
Table Of Contents
Layers
ArcToolbox
Catalog
Search
Drawing
Arial
10

Data Frame

Layout View – notice new tools


Untitled - ArcMap - ArcInfo
File Edit View Bookmarks Insert Selection Geoprocessing Customize Windows Help
Layer:
Editor
Table Of Contents
Layers
Layout
Layers
ArcToolbox
Catalog
Search
Drawing
Arial
10
-2.53 2.90 Inches



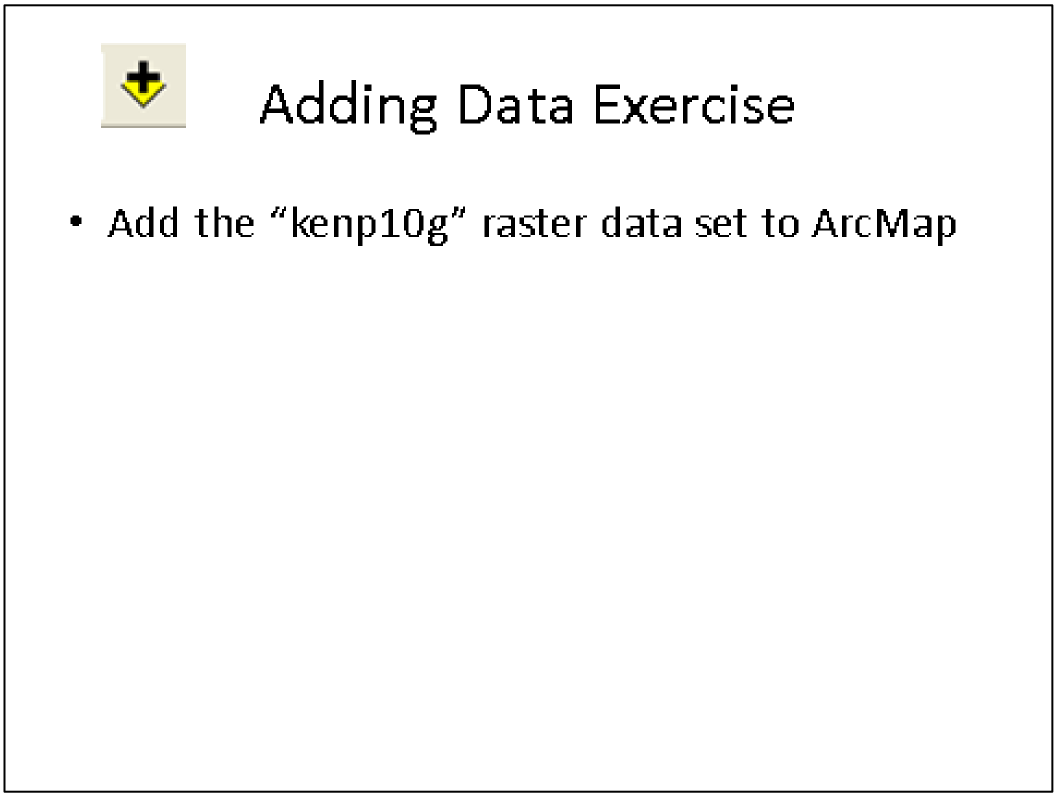
Adding Data

- Add data layers in ArcMap using the…

  ‘Add Data’ button or

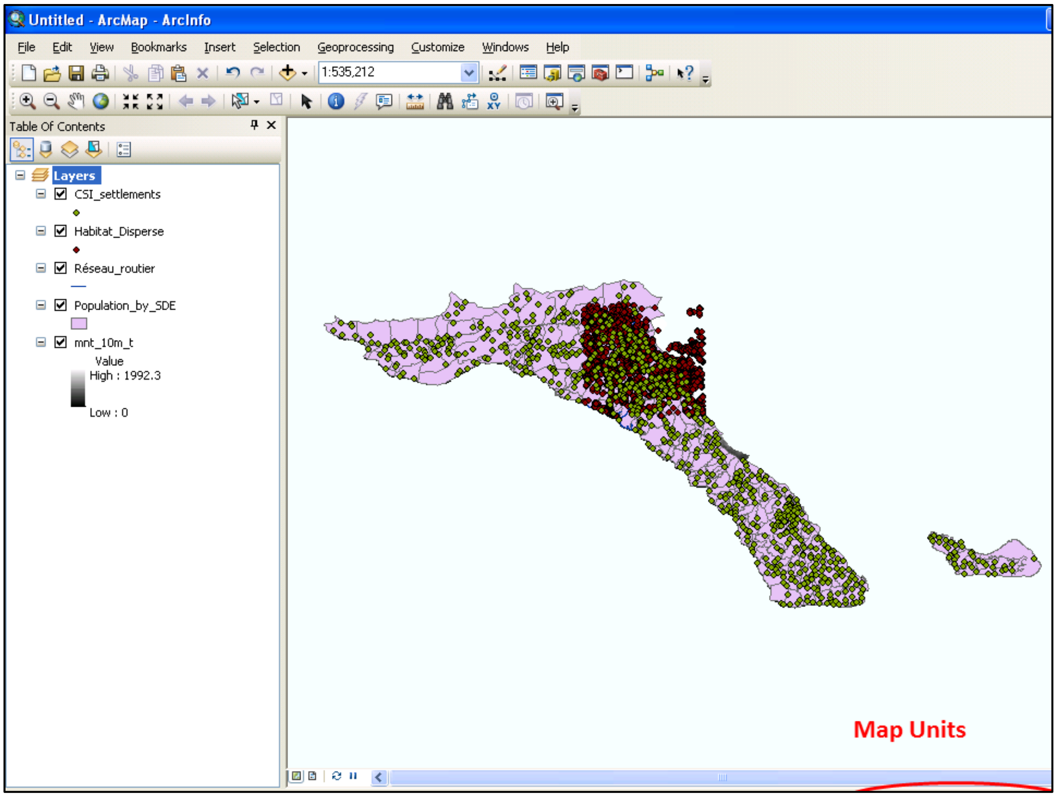
  Drag and drop from ArcCatalog

- Remember if you cannot find the folder where your data is use the Connect To Folder button



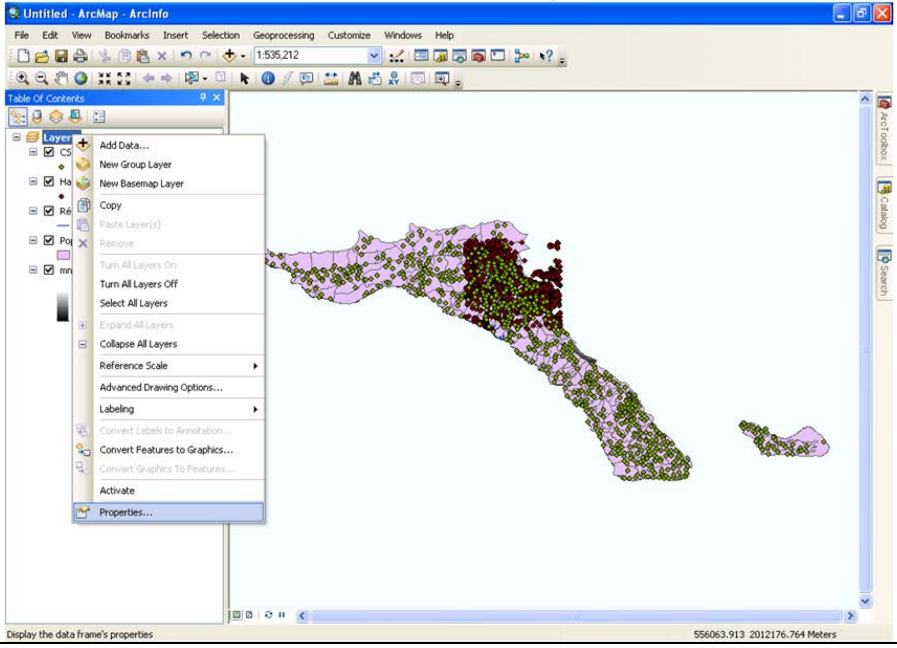
Adding Data Exercise

- Add the “kenp10g” raster data set to ArcMap



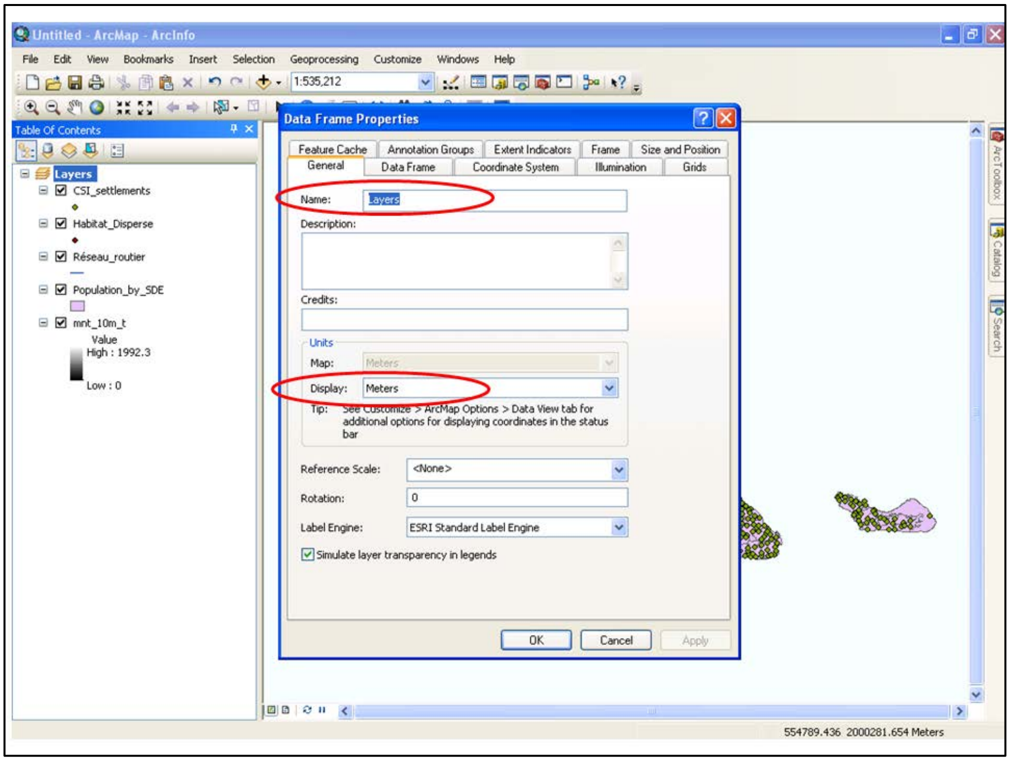
Map Units

Data Frame Properties



Display the data frame's properties

556063.913 2012176.764 Meters



Name: Layers

Description:

Credits:

Units

Map: Meters

Display: Meters

Tip: See Customize > ArcMap Options > Data View tab for additional options for displaying coordinates in the status bar

Reference Scale: <None>

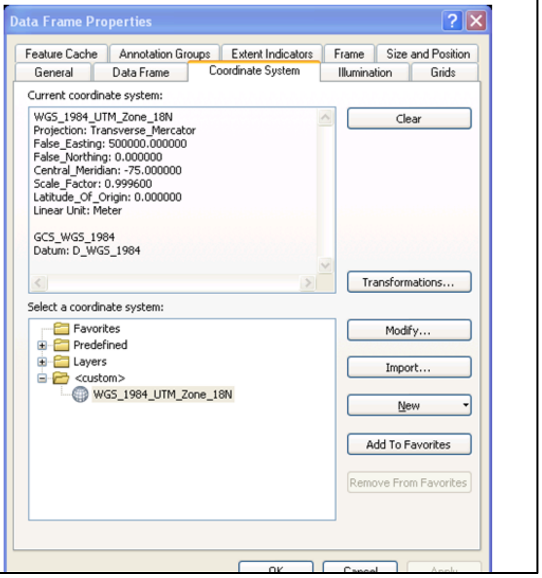
Rotation: 0

Label Engine: ESRI Standard Label Engine

☑ Simulate layer transparency in legends

Data Frame Coordinate System

- The first layer that is added to the data frame sets the projection



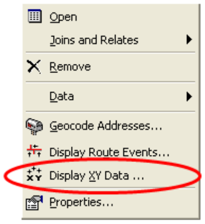
Current coordinate system:

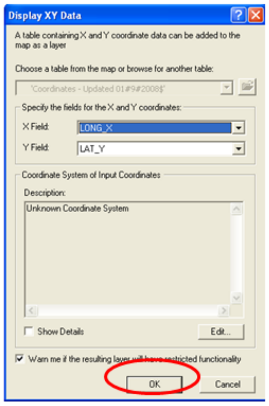
```
WGS_1984_UTM_Zone_18N
Projection: Transverse_Mercator
False_Easting: 500000.000000
False_Northing: 0.000000
Central_Meridian: -75.000000
Scale_Factor: 0.999600
Latitude_Of_Origin: 0.000000
Linear Unit: Meter

GCS_WGS_1984
Datum: D_WGS_1984
```

Select a coordinate system:

Adding XY Coordinates

- Use + and double click on the Excel Spreadsheet (select worksheet$)
- Right click on spreadsheet
- Display XY Data…>





Display XY Data

A table containing X and Y coordinate data can be added to the map as a layer

Choose a table from the map or browse for another table:

'Coordinates - Updated 01#9#2008$'

Specify the fields for the X and Y coordinates:

X Field: LONG_X

Y Field: LAT_Y

Coordinate System of Input Coordinates

Description:

Unknown Coordinate System

□ Show Details

Edit...

☑ Warn me if the resulting layer will have restricted functionality

OK

Cancel

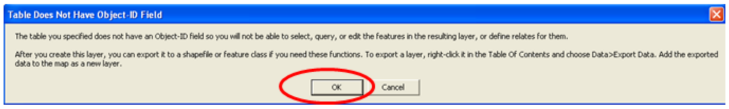


Table Does Not Have Object-ID Field

The table you specified does not have an Object-ID field so you will not be able to select, query, or edit the features in the resulting layer, or define relates for them.

After you create this layer, you can export it to a shapefile or feature class if you need these functions. To export a layer, right-click it in the Table Of Contents and choose Data>Export Data. Add the exported data to the map as a new layer.

OK

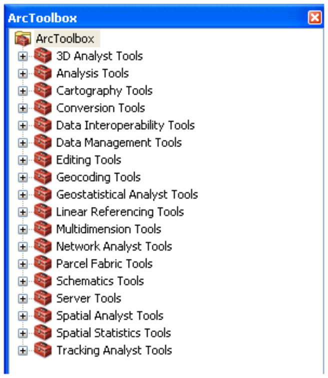
Cancel

Adding x,y data Exercise

- Add ken_admin5 centroids Excel Spreadsheet called “ken_admin5.xls” to ArcMap
- Display x,y data to see the centroids using lon for x and lat for y

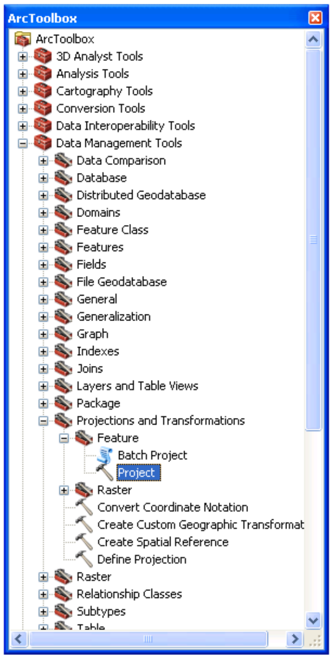
ArcToolbox

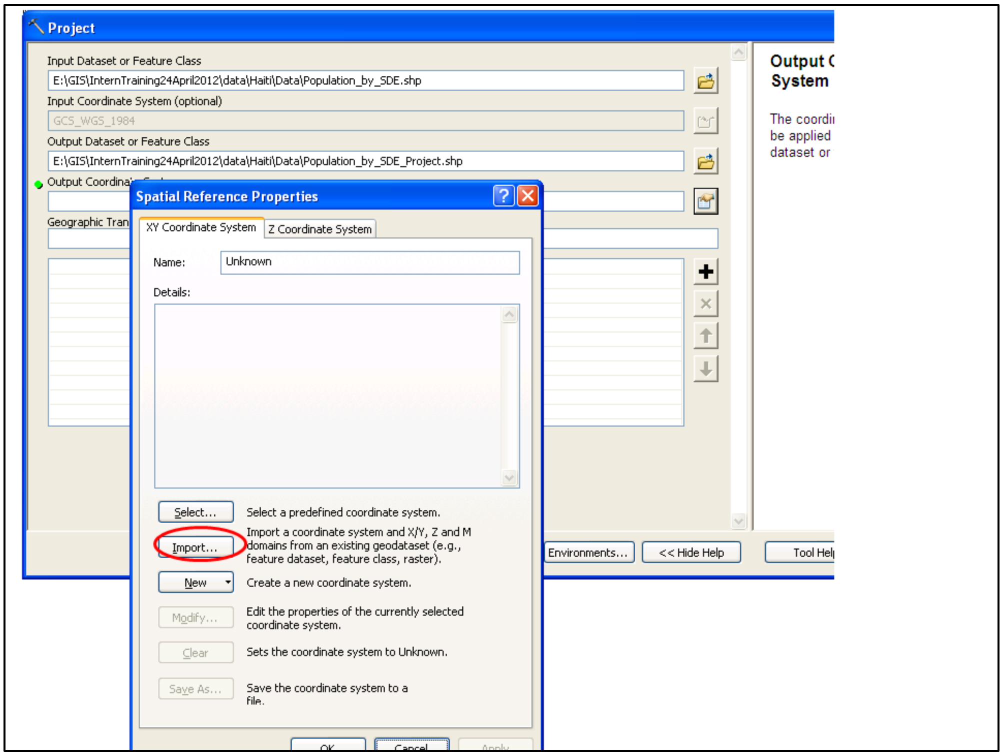
- Contains tools for processing/analyzing data
- Click on the icon to open ArcToolbox within ArcCatalog or ArcMap



Projections

- Create a new data set with an Imported projection





Project

Input Dataset or Feature Class
E:\GIS\InternTraining24April2012\data\Haiti\Data\Population_by_SDE.shp

Input Coordinate System (optional)
GCS_WGS_1984

Output Dataset or Feature Class
E:\GIS\InternTraining24April2012\data\Haiti\Data\Population_by_SDE_Project.shp

Output Coordina

Geographic Tran

Output C
System

The coordi
be applied
dataset or

Spatial Reference Properties

XY Coordinate System | Z Coordinate System

Name: Unknown

Details:

Select... Select a predefined coordinate system.

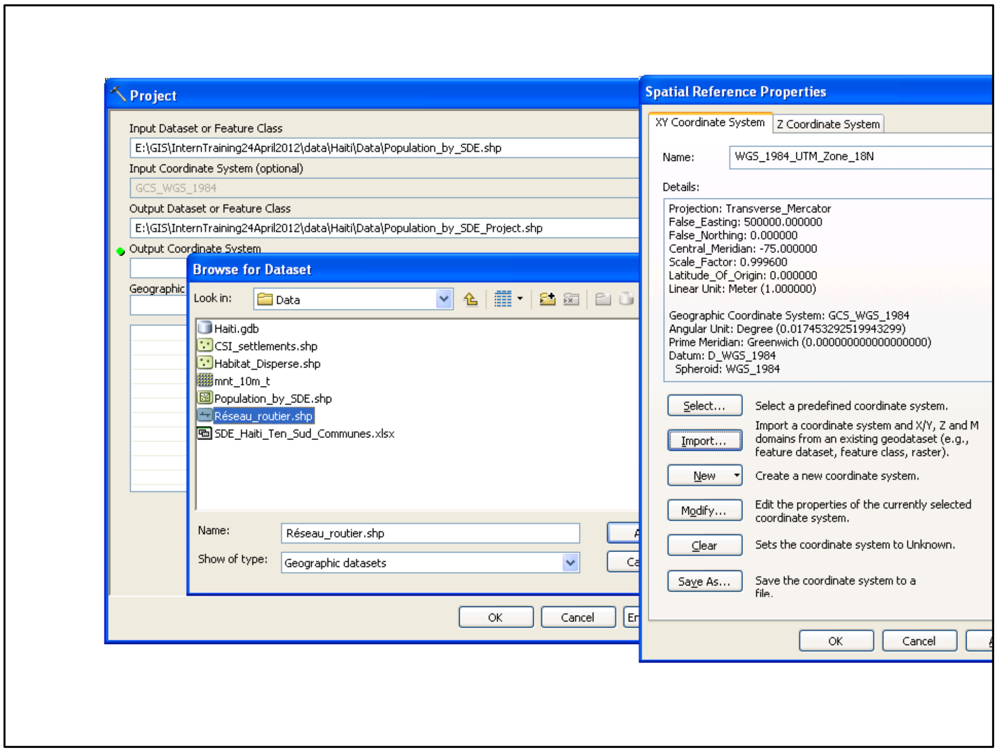
Import... Import a coordinate system and X/Y, Z and M domains from an existing geodataset (e.g., feature dataset, feature class, raster).

New Create a new coordinate system.

Modify... Edit the properties of the currently selected coordinate system.

Clear Sets the coordinate system to Unknown.

Save As... Save the coordinate system to a file.



Project

Input Dataset or Feature Class
E:\GIS\InternTraining24April2012\data\Haiti\Data\Population_by_SDE.shp

Input Coordinate System (optional)
GCS_WGS_1984

Output Dataset or Feature Class
E:\GIS\InternTraining24April2012\data\Haiti\Data\Population_by_SDE_Project.shp

Output Coordinate System

Geographic

Browse for Dataset

Look in: Data

Haiti.gdb
CSI_settlements.shp
Habitat_Disperse.shp
mnt_10m_t
Population_by_SDE.shp
Réseau_routier.shp
SDE_Haiti_Ten_Sud_Communes.xlsx

Name: Réseau_routier.shp

Show of type: Geographic datasets

Spatial Reference Properties

XY Coordinate System | Z Coordinate System

Name: WGS_1984_UTM_Zone_18N

Details:

```
Projection: Transverse_Mercator
False_Easting: 500000.000000
False_Northing: 0.000000
Central_Meridian: -75.000000
Scale_Factor: 0.999600
Latitude_Of_Origin: 0.000000
Linear Unit: Meter (1.000000)

Geographic Coordinate System: GCS_WGS_1984
Angular Unit: Degree (0.0174532925199433299)
Prime Meridian: Greenwich (0.000000000000000000)
Datum: D_WGS_1984
 Spheroid: WGS_1984
```

Select... Select a predefined coordinate system.

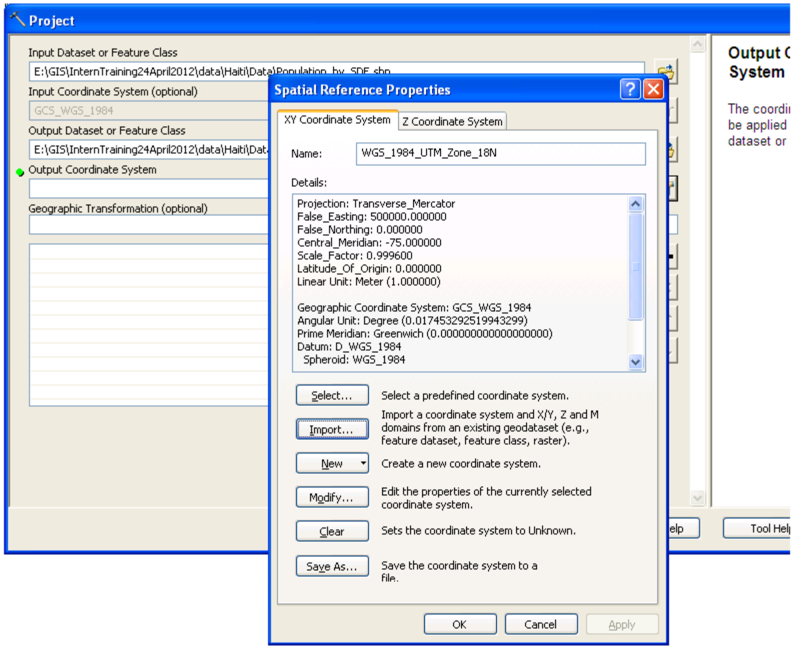
Import... Import a coordinate system and X/Y, Z and M domains from an existing geodataset (e.g., feature dataset, feature class, raster).

New Create a new coordinate system.

Modify... Edit the properties of the currently selected coordinate system.

Clear Sets the coordinate system to Unknown.

Save As... Save the coordinate system to a file.



Project

Input Dataset or Feature Class

E:\GIS\InternTraining24April2012\data\Haiti\Data\Population_by_SDE.shp

Input Coordinate System (optional)

GCS_WGS_1984

Output Dataset or Feature Class

E:\GIS\InternTraining24April2012\data\Haiti\Dat

Output Coordinate System

Geographic Transformation (optional)

Output C System

The coordi be applied dataset or

Spatial Reference Properties

Name: WGS_1984_UTM_Zone_18N

Details:

```
Projection: Transverse_Mercator
False_Easting: 500000.000000
False_Northing: 0.000000
Central_Meridian: -75.000000
Scale_Factor: 0.999600
Latitude_Of_Origin: 0.000000
Linear Unit: Meter (1.000000)

Geographic Coordinate System: GCS_WGS_1984
Angular Unit: Degree (0.0174532925199433299)
Prime Meridian: Greenwich (0.000000000000000000)
Datum: D_WGS_1984
 Spheroid: WGS_1984
```

Select... Select a predefined coordinate system.

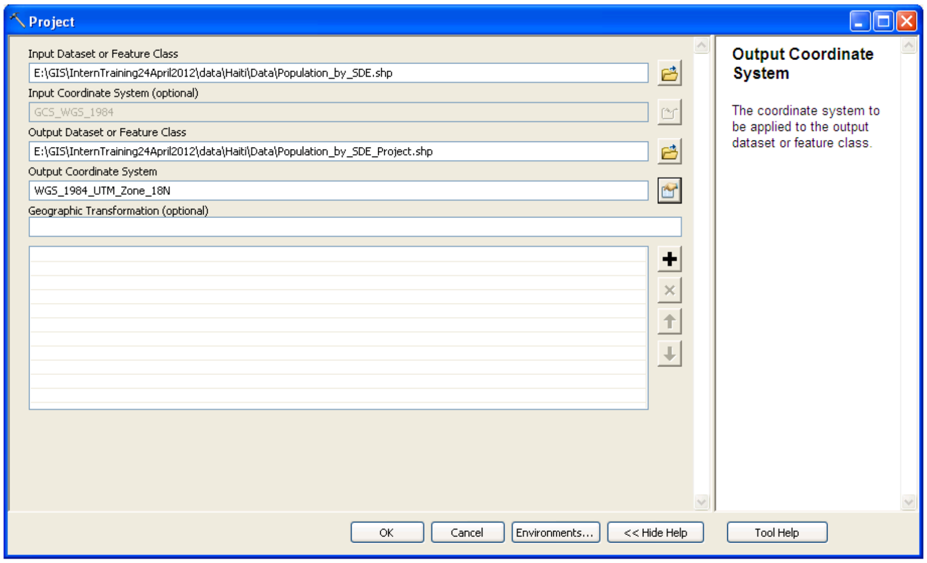
Import... Import a coordinate system and X/Y, Z and M domains from an existing geodataset (e.g., feature dataset, feature class, raster).

New Create a new coordinate system.

Modify... Edit the properties of the currently selected coordinate system.

Clear Sets the coordinate system to Unknown.

Save As... Save the coordinate system to a file.



Project

Input Dataset or Feature Class

E:\GIS\InternTraining24April2012\data\Haiti\Data\Population_by_SDE.shp

Input Coordinate System (optional)

GCS_WGS_1984

Output Dataset or Feature Class

E:\GIS\InternTraining24April2012\data\Haiti\Data\Population_by_SDE_Project.shp

Output Coordinate System

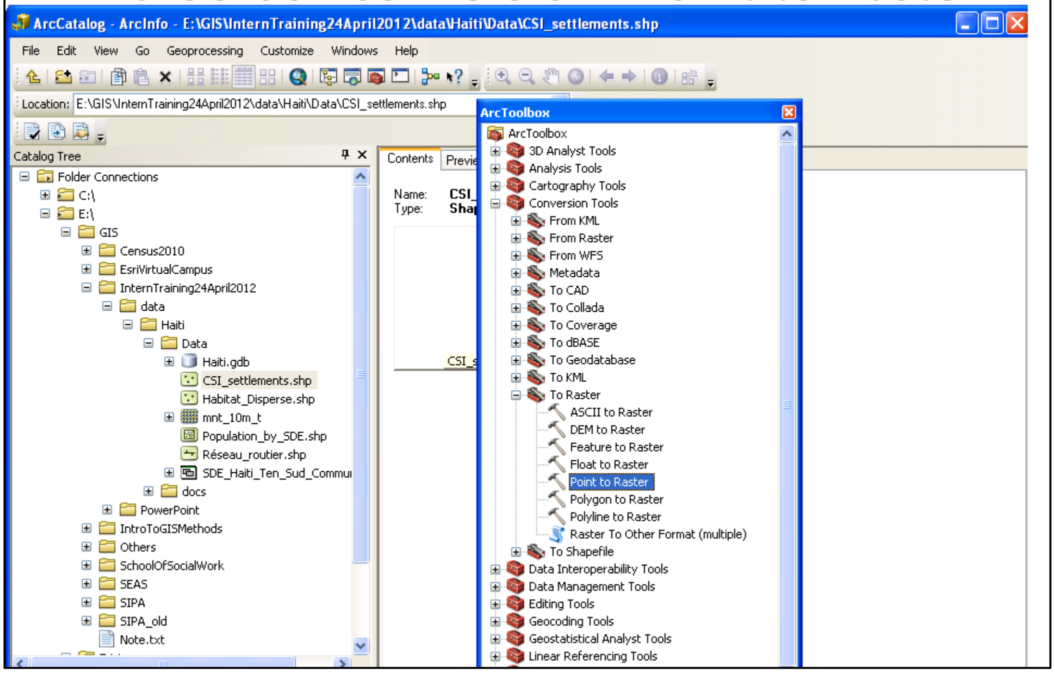
WGS_1984_UTM_Zone_18N

Geographic Transformation (optional)

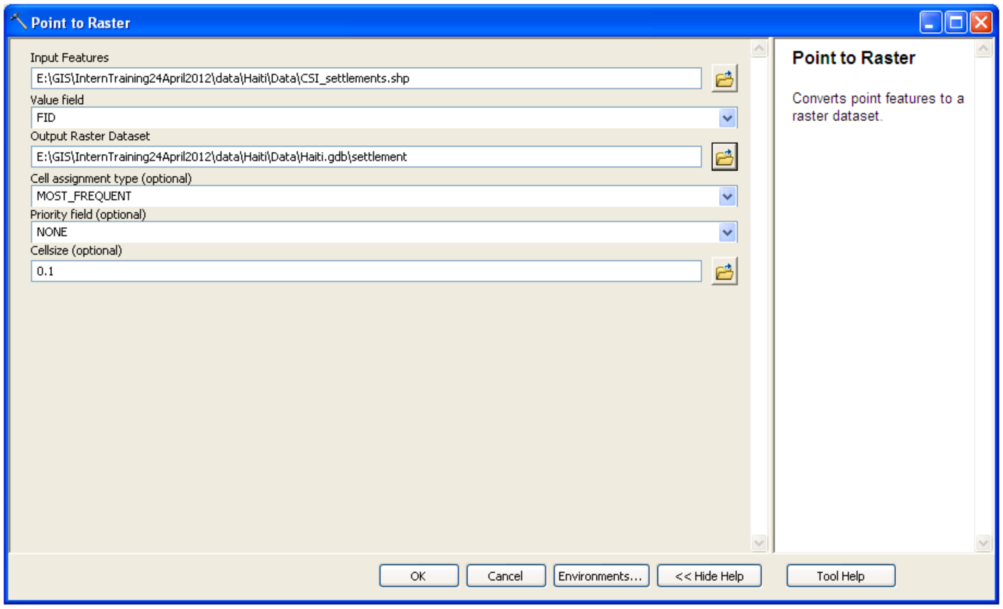
Output Coordinate System

The coordinate system to be applied to the output dataset or feature class.

ArcToolbox conversion- Point to Raster



ArcToolbox conversion- Point to Raster



Point to Raster

Input Features
E:\GIS\InternTraining24April2012\data\Haiti\Data\CSI_settlements.shp

Value field
FID

Output Raster Dataset
E:\GIS\InternTraining24April2012\data\Haiti\Data\Haiti.gdb\settlement

Cell assignment type (optional)
MOST_FREQUENT

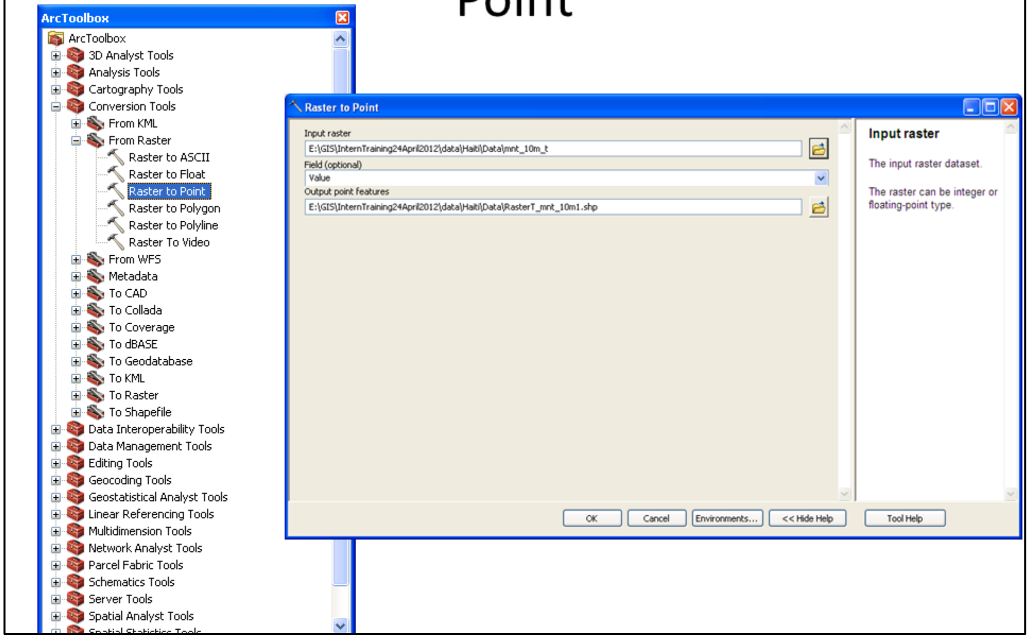
Priority field (optional)
NONE

Cellsize (optional)
0.1

Point to Raster

Converts point features to a raster dataset.

ArcToolbox Conversion – Raster to Point



Input raster
E:\GIS\InternTraining24April2012\data\Haiti\Data\mnt_10m_t

Field (optional)
Value

Output point features
E:\GIS\InternTraining24April2012\data\Haiti\Data\RasterT_mnt_10m1.shp

Input raster

The input raster dataset.

The raster can be integer or floating-point type.

Exercise 3. Data Processing in ArcGIS

In this exercise we will process the various input GIS datasets and convert export them to comma separated variable (.CSV) tables. First we will create a common reference system for our area of interest, then process vector files (points, lines and polygons) and finally process raster data.

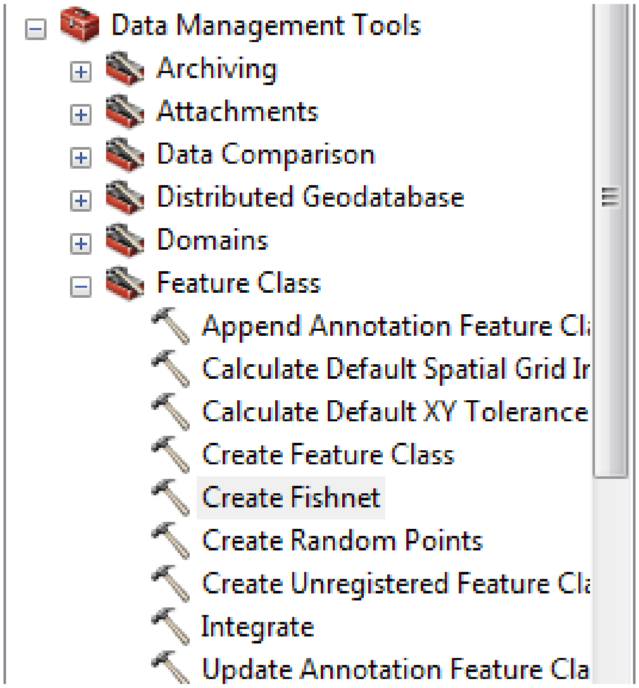
3.1 Preparing a Common Reference System

A common reference system consists of a fishnet covering the entire area of interest and a point file representing the centroids of each polygon in the fishnet. Each polygon and centroid representing a pixel location has a pixel identification number, **pixel_id**. The **pixel_id** is the attribute used to join other variables to create the tabular files.

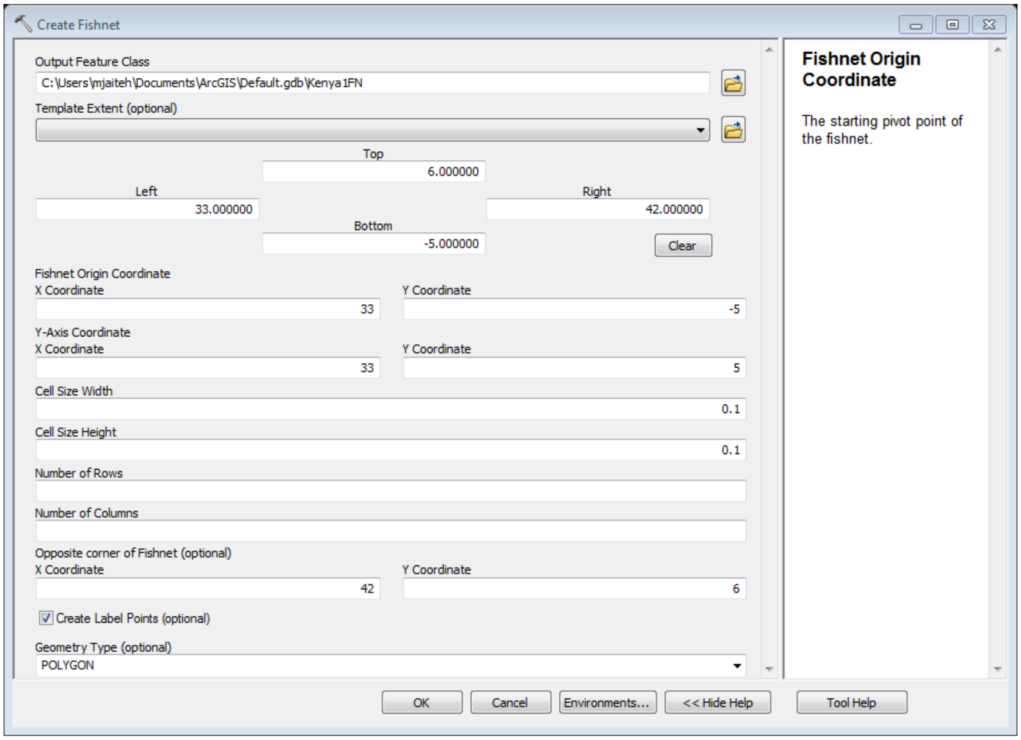
Preparing the Common Reference System involves the following steps:

3.1.1 Create Fishnet

1. Start a new ArcMap session and navigate to Create Fishnet tool under the Data management/Feature Class tool box.



2. Start the Create Fishnet tool, name *Output Feature Class* as KenyaFN, spatial extent information as shown below. Make sure to check box next to *Create Label Points* and select POLYGON as *Geometry Type.*



Create Fishnet

Output Feature Class: C:\Users\mjaiteh\Documents\ArcGIS\Default.gdb\Kenya1FN

Template Extent (optional):

Top: 6.000000

Left: 33.000000

Right: 42.000000

Bottom: -5.000000

Fishnet Origin Coordinate
X Coordinate: 33
Y Coordinate: -5

Y-Axis Coordinate
X Coordinate: 33
Y Coordinate: 5

Cell Size Width: 0.1

Cell Size Height: 0.1

Number of Rows:

Number of Columns:

Opposite corner of Fishnet (optional)
X Coordinate: 42
Y Coordinate: 6

☑ Create Label Points (optional)

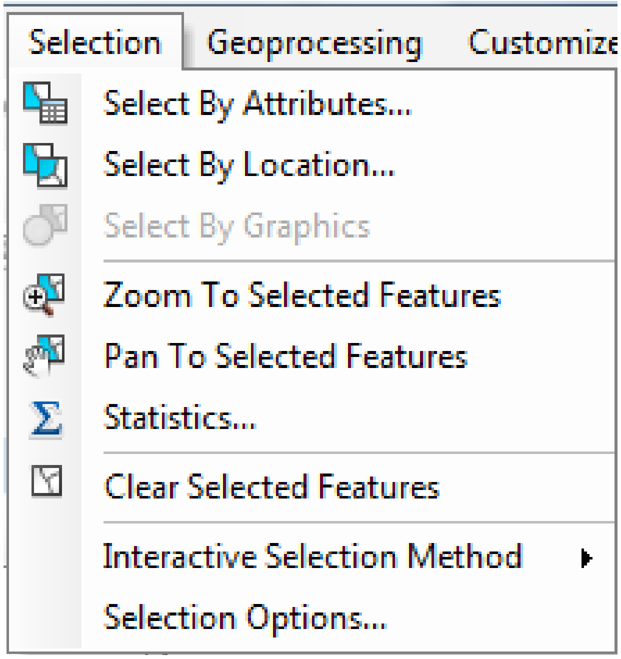
Geometry Type (optional): POLYGON

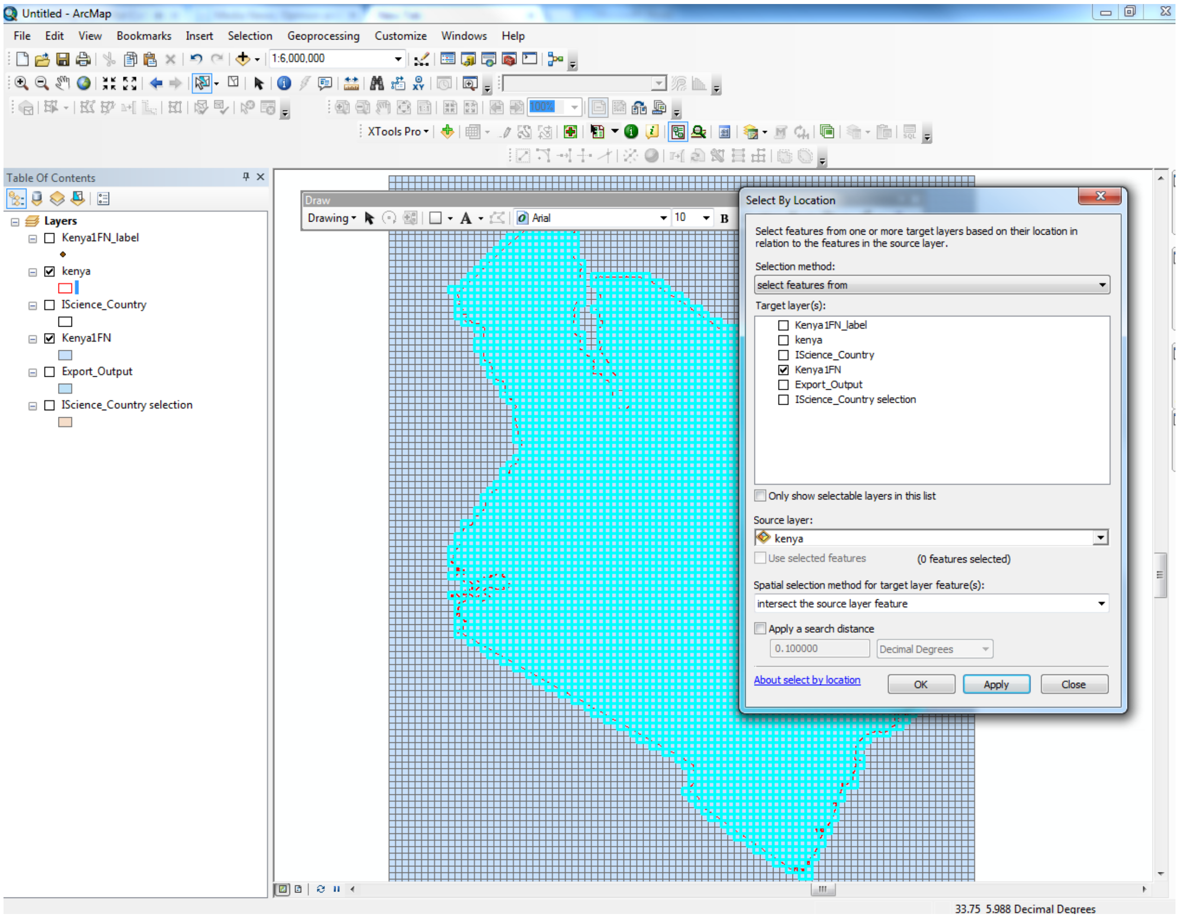
Fishnet Origin Coordinate

The starting pivot point of the fishnet.

The fishnet created extends beyond Kenya's boundary, Longitude 33E to 42E and Latitude 5S to 6N. Next we will extract only those polygons in the fishnet that intersects with the Kenya's country, using the ***Select by Location*** tool.

3. Under the selection menu, use the ***Select by Location*** option to select polygons in the fishnet that Intersect with the Kenya country boundary. (Click ***Selection*** on the menu bar and)





Select By Location

Select features from one or more target layers based on their location in relation to the features in the source layer.

Selection method: select features from

Target layer(s):

- ☐ Kenya1FN_label
- ☐ kenya
- ☐ IScience_Country
- ☑ Kenya1FN
- ☐ Export_Output
- ☐ IScience_Country selection

☐ Only show selectable layers in this list

Source layer: kenya

☐ Use selected features (0 features selected)

Spatial selection method for target layer feature(s): intersect the source layer feature

☐ Apply a search distance

0.100000 Decimal Degrees

4. Export the features selected above to a new feature class, name it **yourCountyNameID**:

   a. Use the ***Select By Location*** to :

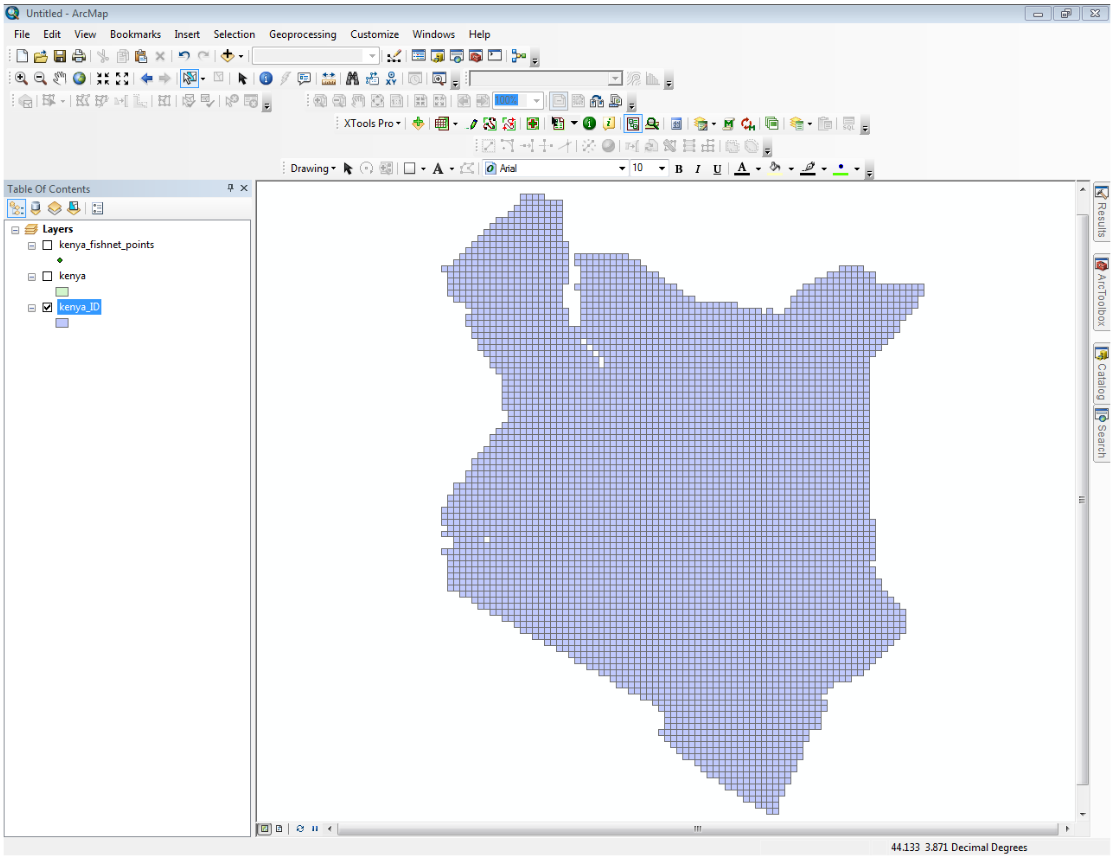
   b. Selection Method: **Select Features From**

   c. Target: “**YourCountryNameFN**”

   d. Source Layer: **yourCountry** admin. For those from Uganda this is Uganda admin layer

   e. Spatial Selection**: intersect the source layer feature**

5. Export Selected cells ”yourCountryName**ID**” fishnet polygon feature class
-Right click on YourCountryName feature >selected features> This layer is the source data

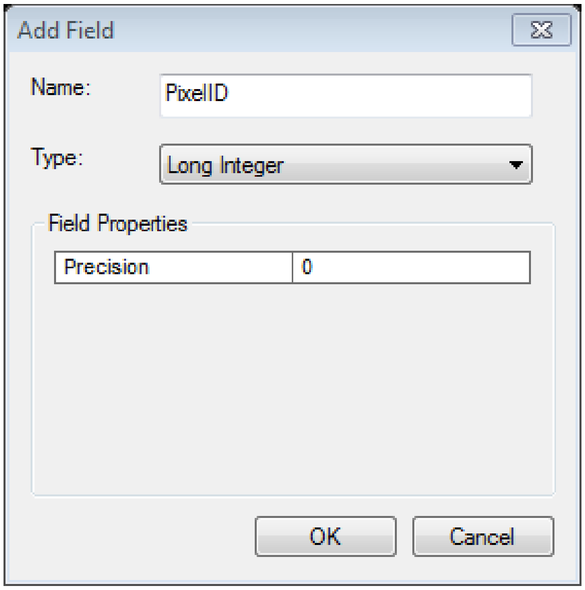


3.1.2 Create Fishnet Centroids

Select only the fishnet centroids that intersect the country boundary. Use the yourCountryNameID feature class created above to select by location the fishnet centroids that intersect with the yourCountryNameID .

1. Use the following steps to run the ***Select by Location*** tool to create fishnet centroids layer :
    - **a.** Selection Method: **Select Features From**
    - b. Target: “yourCountryNameFN_label” points
    - c. Source Layer: “Kenya boundary” polygon fishnet
    - d. Spatial Selection: **intersect the source layer feature**
    - e. Export selected points>Data>”**yourCountryName_centroids**” point feature class

2. Add a new field to the Kenya_Centroids attribute table. Call it PixelID. Make this field Long Integer.



Add Field

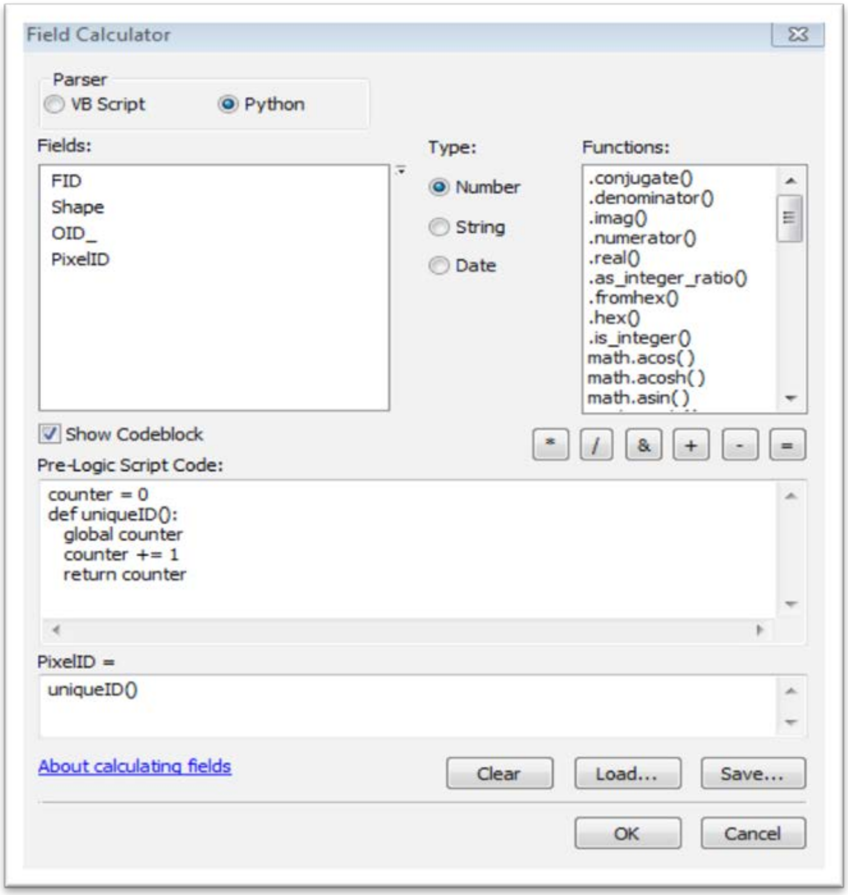
Name: PixelID

Type: Long Integer

Field Properties

Precision	0

3. Calculate PixelID values using the following code. This gives a unique identification number for each of the centroid points. Warning: Python is case sensitive. Make sure you type in the code as is with line 3-5 starting 2 spaces after line begins.



Field Calculator

Parser
○ VB Script ● Python

Fields:
FID
Shape
OID_
PixelID

Type:
● Number
○ String
○ Date

Functions:
.conjugate()
.denominator()
.imag()
.numerator()
.real()
.as_integer_ratio()
.fromhex()
.hex()
.is_integer()
math.acos( )
math.acosh( )
math.asin( )

☑ Show Codeblock

Pre-Logic Script Code:

```
counter = 0
def uniqueID():
  global counter
  counter += 1
  return counter
```

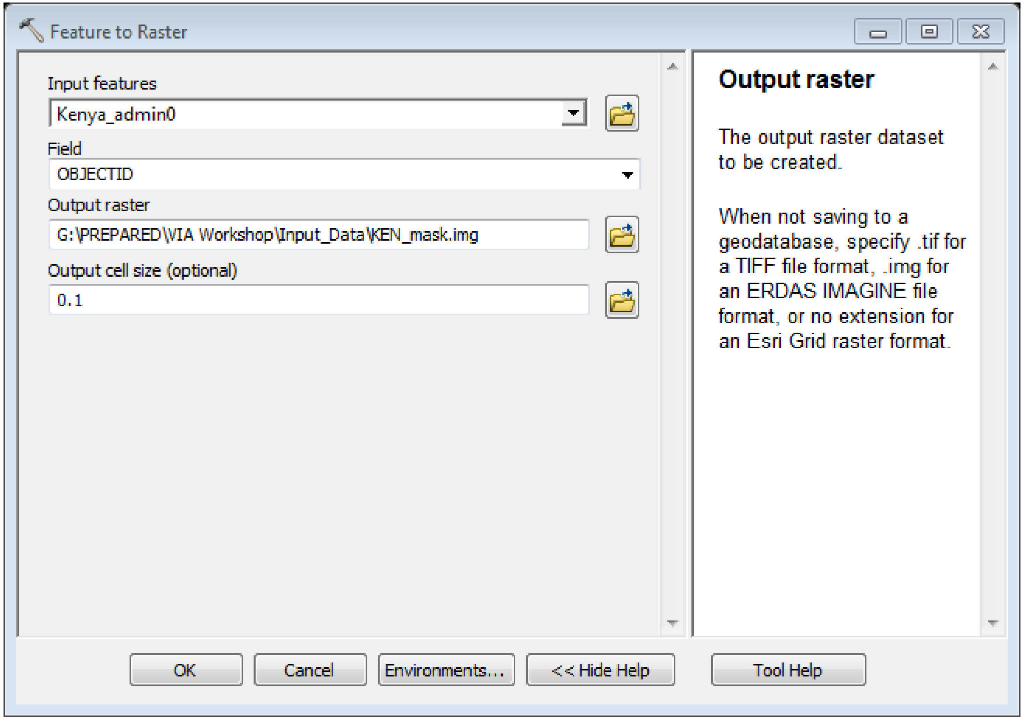
PixelID =

```
uniqueID()
```

3.1.3 Create Mask

A raster mask is dataset that defines **area of interest** in the input rasters. It is often a 1 and NoData raster, with 1 defining areas included. Cells that are NoData in a raster mask will be NoData in the output. We use a raster mask to create country raster of raster of small spatial extent from a global or regional raster.

1. To create a Kenya country raster mask, Use ***Feature To Raster*** tool to create raster of **YourCountryName** admin0 shapefile. (ArcToolbox\Conversion Tools\To Raster\Feature To Raster.)



Input features: Kenya_admin0

Field: OBJECTID

Output raster: G:\PREPARED\VIA Workshop\Input_Data\KEN_mask.img

Output cell size (optional): 0.1

Output raster

The output raster dataset to be created.

When not saving to a geodatabase, specify .tif for a TIFF file format, .img for an ERDAS IMAGINE file format, or no extension for an Esri Grid raster format.

Input features: “Kenya_admin0.shp” polygon feature class
Field: ObjectID
Output raster: **ken_mask.img**
Output Cell Size: 0.01 dd (~10km)
*Set Environment: Extent to Kenya_admin0.shp.

2. Use the Extract by Mask tool in Spatial Analyst   to extract all raster datasets to “Ken_Mask.img” extent.

3.2 Processing Vector Data

The goal in this exercise is to convert all input vector layers, tables, points, lines and polygons to text files that can be joined to the Fishnet Centroids created above. The processing steps needed are to convert all vector data to raster, then extract the values of the raster to the centroids and export the values to a file text.

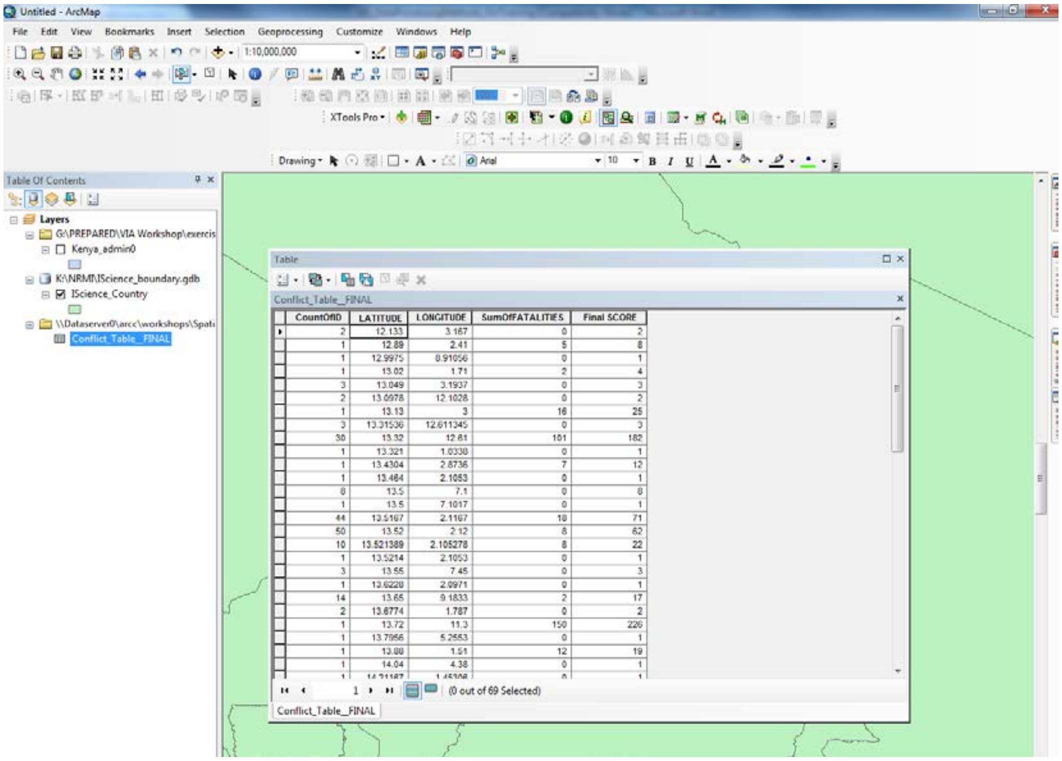
3.2.1 Point/Tabular Data

Some vector data such as survey analysis results from Microsoft Excel and other statistical packages are available as tables. These tables often contain fields representing coordinate information. The following steps are needed to process tabular vector files

1) Plot XY events layer in ArcMap
2) Export the events layer to shapefile or geodatabase Feature Class
3) Create a raster surface using one of several Interpolation methods
4) Close all ArcMap sessions before the next section

Plot XY Events layer in ArcMap

1. Start a new ArcMap session
2. Add the conflict_Table_FINAL.xlsx fine in the exercise folder to your map. This is the file containing the data we are converting.
3. Open the attribute table to confirm that there is latitude and longitude field.


Untitled - ArcMap
File Edit View Bookmarks Insert Selection Geoprocessing Customize Windows Help
1:10,000,000
XTools Pro
Drawing
Arial
10
Table Of Contents
Layers
G:\PREPARED\VIA Workshop\exercis
Kenya_admin0
K:\NRMI\IScience_boundary.gdb
IScience_Country
\\Dataserver0\arcc\workshops\Spati
Conflict_Table_FINAL
Table

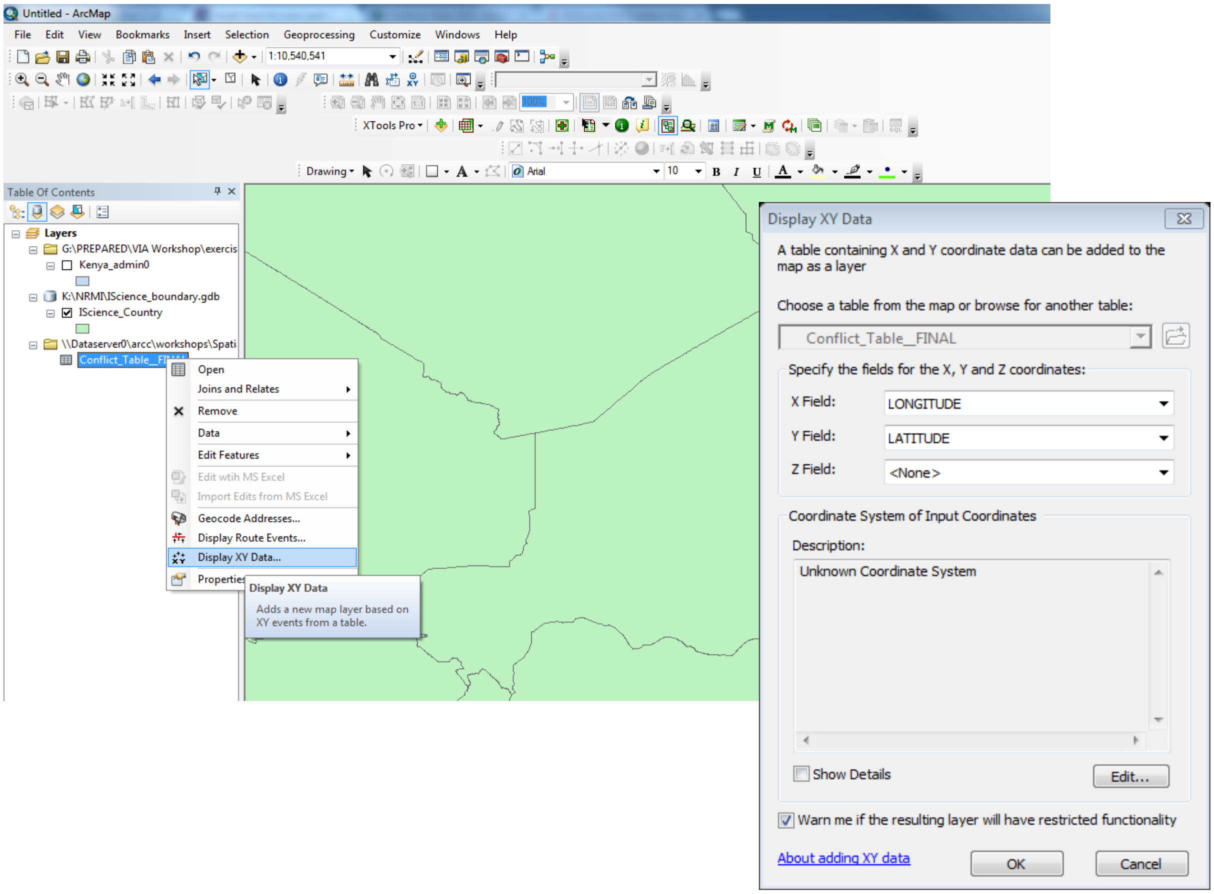
Conflict_Table_FINAL

CountOfID	LATITUDE	LONGITUDE	SumOfFATALITIES	Final SCORE
2	12.133	3.167	0	2
1	12.89	2.41	5	8
1	12.9975	8.91056	0	1
1	13.02	1.71	2	4
3	13.049	3.1937	0	3
2	13.0978	12.1028	0	2
1	13.13	3	16	25
3	13.31536	12.611345	0	3
30	13.32	12.61	101	182
1	13.321	1.0338	0	1
1	13.4304	2.8736	7	12
1	13.464	2.1053	0	1
8	13.5	7.1	0	8
1	13.5	7.1017	0	1
44	13.5167	2.1167	18	71
50	13.52	2.12	8	62
10	13.521389	2.105278	8	22
1	13.5214	2.1053	0	1
3	13.55	7.45	0	3
1	13.6228	2.0971	0	1
14	13.65	9.1833	2	17
2	13.6774	1.787	0	2
1	13.72	11.3	150	226
1	13.7966	5.2553	0	1
1	13.88	1.51	12	19
1	14.04	4.38	0	1
1	14.21187	1.45308	0	1

1 (0 out of 69 Selected)

Conflict_Table_FINAL

4. Right-click on the Table of Contents and select Display XY Data…. This opens the Display XY Data dialog box. Make sure you chose field for longitude for X Field and latitude for Y Field (see below).



Display XY Data

A table containing X and Y coordinate data can be added to the map as a layer

Choose a table from the map or browse for another table:

Conflict_Table__FINAL

Specify the fields for the X, Y and Z coordinates:

X Field: LONGITUDE

Y Field: LATITUDE

Z Field: <None>

Coordinate System of Input Coordinates

Description:

Unknown Coordinate System

☐ Show Details

☑ Warn me if the resulting layer will have restricted functionality

5. Right-click on the newly created feature class and export to a shapefile.

Create an Interpolation Surface raster

ArcGIS Spatial Analyst and Geostatistics extentions have Interpolation toolsets containing numerous tools for creating continuous surface from point data. Surface interpolation tools make predictions from point measurements for all locations in an output raster dataset, whether or not a measurement has been taken at the location.

The interpolation tools are generally divided into deterministic and geostatistical methods:

The deterministic interpolation methods assign values to locations based on the surrounding measured values and on specified mathematical formulas that determine the smoothness of the resulting surface. The deterministic methods include the Inverse Distance Weighting (IDW) Natural Neighbor, Trend, and Spline.

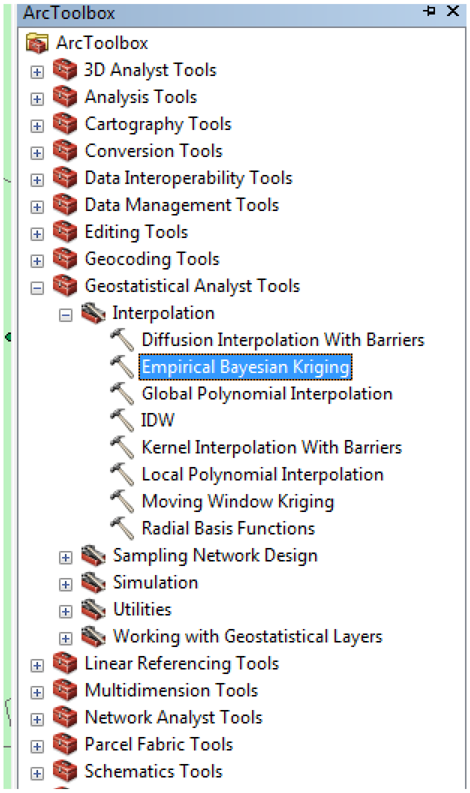
- The geostatistical methods are based on statistical models that include autocorrelation (the statistical relationship among the measured points). Because of this, geostatistical techniques not only have the capability of producing a prediction surface but also provide some measure of the certainty or accuracy of the predictions. Empirical Bayesian Kriging (EBK) is a geostatistical method of interpolation.

Interpolating Cluster Data

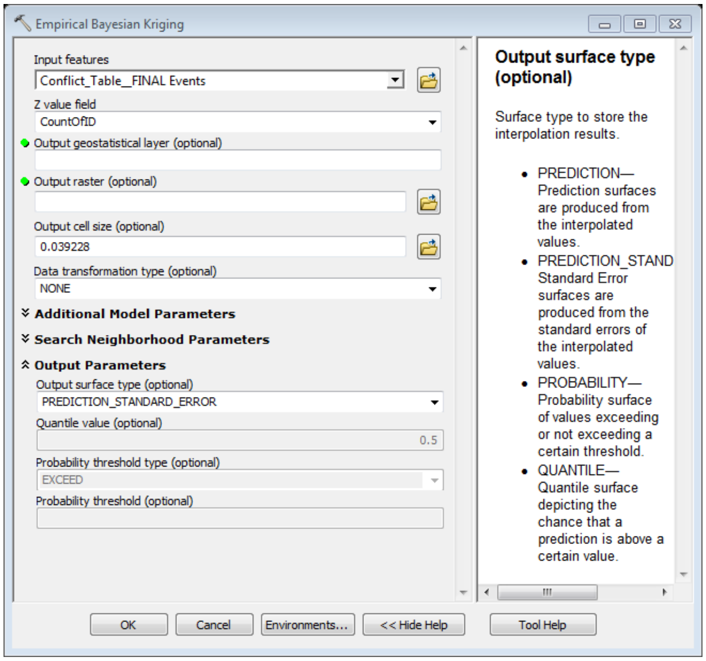
Empirical Bayesian Kriging (EBK) is a geostatistical interpolation method that automates the task of building a valid kriging model. EBK automatically calculates model parameters through a process of subsetting and simulations. In addition to creating a prediction surface, EBK also allows you to create a prediction standard error surface. For further details, please connect on ESRI website (http://www.esri.com/news/arcuser/0704/files/interpolating.pdf).

You will need to activate the Geostatistical Analyst and Spatial Analyst extensions to use interpolation tools. (Customize>Extensions).

1. Under ArcToolbox/Geostatistical Analyst Toolbox, start the Empirical Bayesian Kriging tool.



2. Set Environments: Set Mask and Raster Analysis Cell Size. This should be equal to the fishnet resolution.
3. Output Parameters: Output surface type choose PREDICTION
4. Run the EBK Tool again but this time choose for Output surface type choose PREDICTION STANDARD ERROR.



Input features
Conflict_Table__FINAL Events
Z value field
CountOfID
Output geostatistical layer (optional)
Output raster (optional)
Output cell size (optional)
0.039228
Data transformation type (optional)
NONE
Additional Model Parameters
Search Neighborhood Parameters
Output Parameters
Output surface type (optional)
PREDICTION_STANDARD_ERROR
Quantile value (optional)
0.5
Probability threshold type (optional)
EXCEED
Probability threshold (optional)

Output surface type (optional)

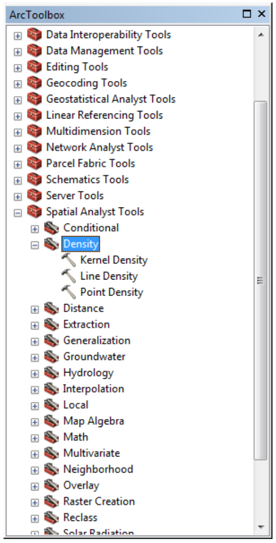
Surface type to store the interpolation results.

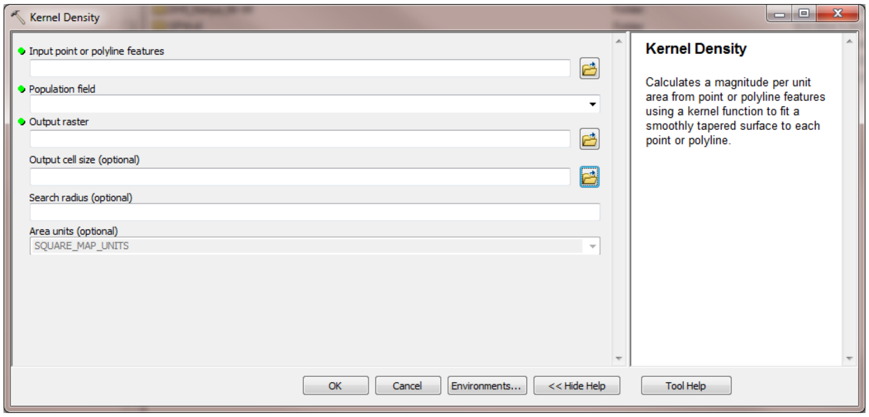
- PREDICTION—Prediction surfaces are produced from the interpolated values.
- PREDICTION_STAND Standard Error surfaces are produced from the standard errors of the interpolated values.
- PROBABILITY—Probability surface of values exceeding or not exceeding a certain threshold.
- QUANTILE—Quantile surface depicting the chance that a prediction is above a certain value.

Kernel Density/Point Density

Create a surface to calculate the density of points or number of points per unit area is another way of interpolating points. The search radius of the kernel function will vary the results. Only the points that fall within the search radius will be included in the density calculation. A large radius will create a more generalized or smoother output raster. A small radius will create a more detailed raster. The input layer should be in projected coordinate system to use the kernel density tool.

Under ArcToolbox/Spatial Analyst Tools/Density/Kernel Density





Input point or polyline features
Population field
Output raster
Output cell size (optional)
Search radius (optional)
Area units (optional)
SQUARE_MAP_UNITS

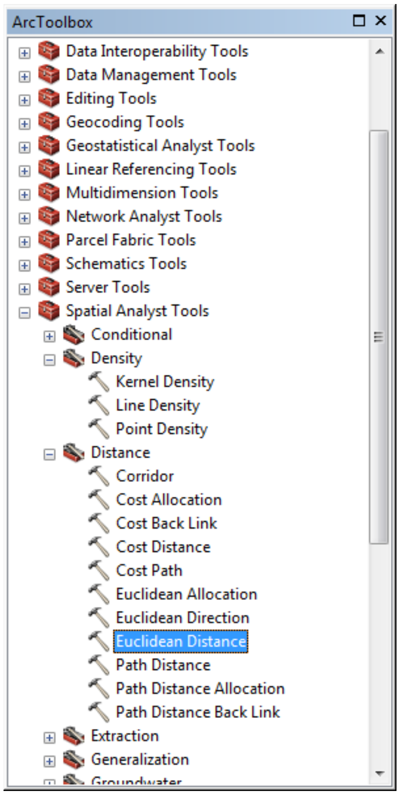
Kernel Density

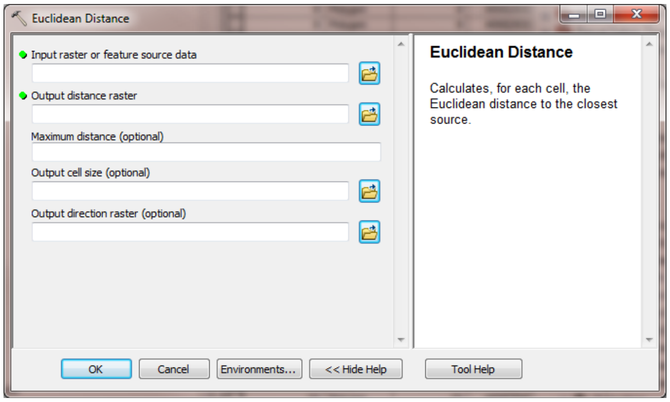
Calculates a magnitude per unit area from point or polyline features using a kernel function to fit a smoothly tapered surface to each point or polyline.

3.2.2 Line

For line data create a surface of the distance from the lines.

ArcToolbox\Spatial Analyst Tools/Distance/Euclidean Distance



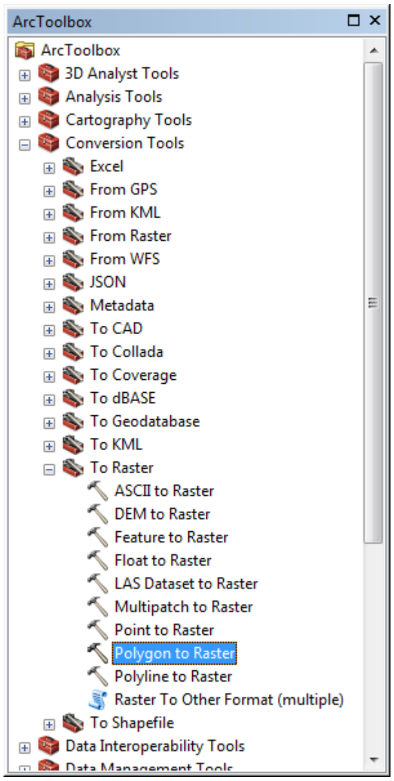

Euclidean Distance
Input raster or feature source data
Output distance raster
Maximum distance (optional)
Output cell size (optional)
Output direction raster (optional)
OK
Cancel
Environments...
<< Hide Help
Tool Help

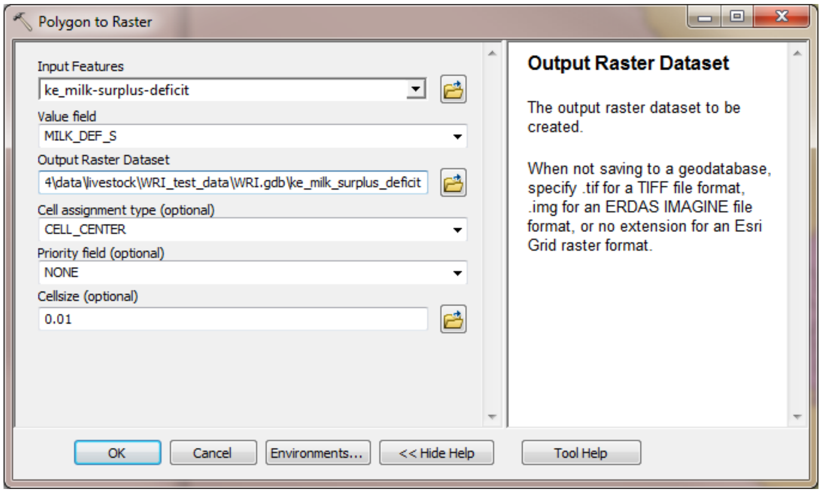
Euclidean Distance

Calculates, for each cell, the Euclidean distance to the closest source.

3.2.3 Polygons

For polygons that have integer values create a surface of the polygon data using the conversion, **Polygon to Raster** tool.
ArcToolbox\Conversion Tools\To Raster\Polygon To Raster





Input Features
ke_milk-surplus-deficit
Value field
MILK_DEF_S
Output Raster Dataset
4\data\livestock\WRI_test_data\WRI.gdb\ke_milk_surplus_deficit
Cell assignment type (optional)
CELL_CENTER
Priority field (optional)
NONE
Cellsize (optional)
0.01

Output Raster Dataset

The output raster dataset to be created.

When not saving to a geodatabase, specify .tif for a TIFF file format, .img for an ERDAS IMAGINE file format, or no extension for an Esri Grid raster format.

Example:
Input: ke_milk-surplus-deficit
Value: MILK_DEF_S
Output Raster: WRI.gdb\ke_milk_surplus_deficit
Cell assignment type: CELL CENTER
Priority field: NONE
Cellsize: Set to Kenya_Mask
Set Environments: Set Processing Extent: Kenya_Mask,
Raster Analysis: Kenya_Mask

Categorical polygon data

For polygon data that have text values, say land cover types, we will need to assign the categories a numeric equivalence before converting to raster using **Feature to Raster** tool.

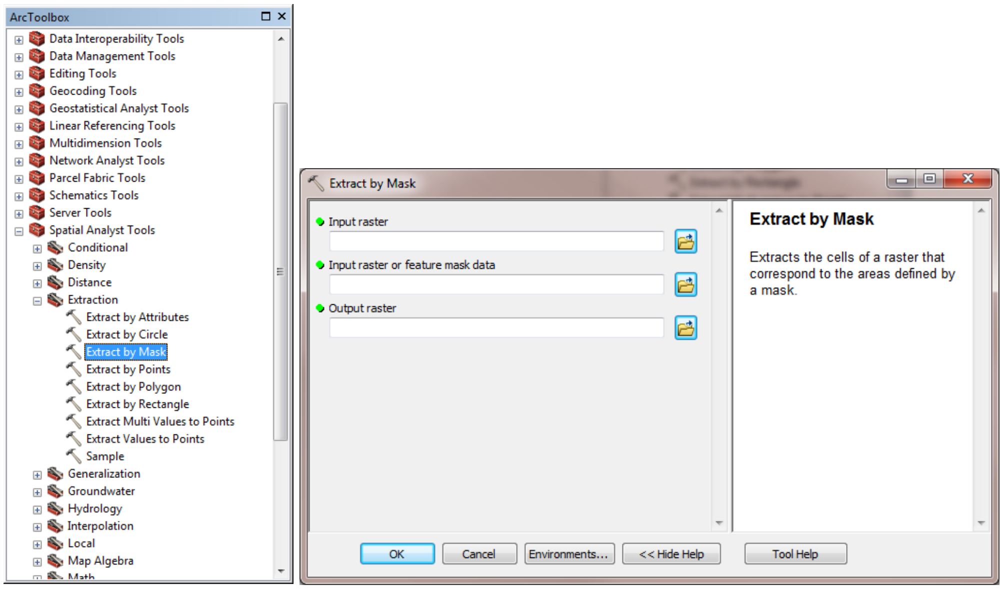
Alternatively we can execute *Spatial Join* (ArcToolbox\Analysis\Overlay\Spatial Join) in which the attributes of the polygon features are joined to the point features, i.e., the "target feature"=points and the "join feature"=polygon

3.3 Processing Raster Data

Processing Raster Data requires the use of a number of tools including, **Extract by Mask**, **Aggregate**, **Resample**, **Reclassify**, **Nibble** and **Extract Values to Points**

3.3.1 Subset Raster Data

Subset the data set to the country national boundary extent using ArcGIS *Extract by Mask* tool (ArcToolbox\SpatialAnalyst\Extraction\Extract By Mask).



Extract by Mask

Extracts the cells of a raster that correspond to the areas defined by a mask.

Input raster:Kenya _Mask

-*(Set Environment: Processing Extent to "mask"; Cell Size to the "mask".)* All cells outside the mask are assigned No Data value -9999. Check that output has same number of columns, rows and cell size for all data sets.

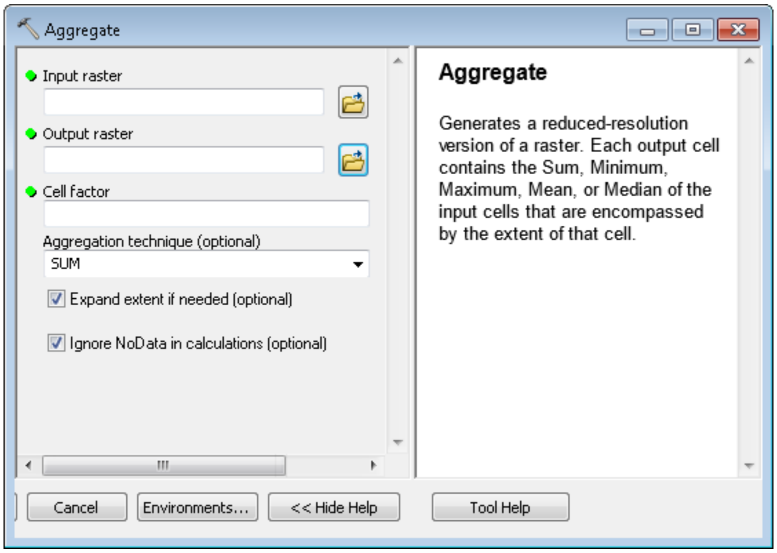
3.3.2 Matching Rasters

Ensure that all your raster files have the same resolution and extent. If not then either *aggregate* higher resolution to lower resolution, *nibble* to grow a raster to your extent or *resample* lower resolution to higher resolution.

Toolboxes\System Toolboxes\Spatial Analyst Tools.tbx\Generalization\Aggregate

Aggregate

- Aggregate tool is used to merge fine resolution raster cells into coarse resolution cell size.
- Eg. If you have a data set with a resolution of 100m then you use aggregate tool to create a raster with 1,000m cell size.
- Make sure tif's are converted to rasters before aggregating (Export>Data>Geodatabase Raster)



Aggregation technique (optional): SUM

☑ Expand extent if needed (optional)

☑ Ignore NoData in calculations (optional)

Aggregate

Generates a reduced-resolution version of a raster. Each output cell contains the Sum, Minimum, Maximum, Mean, or Median of the input cells that are encompassed by the extent of that cell.

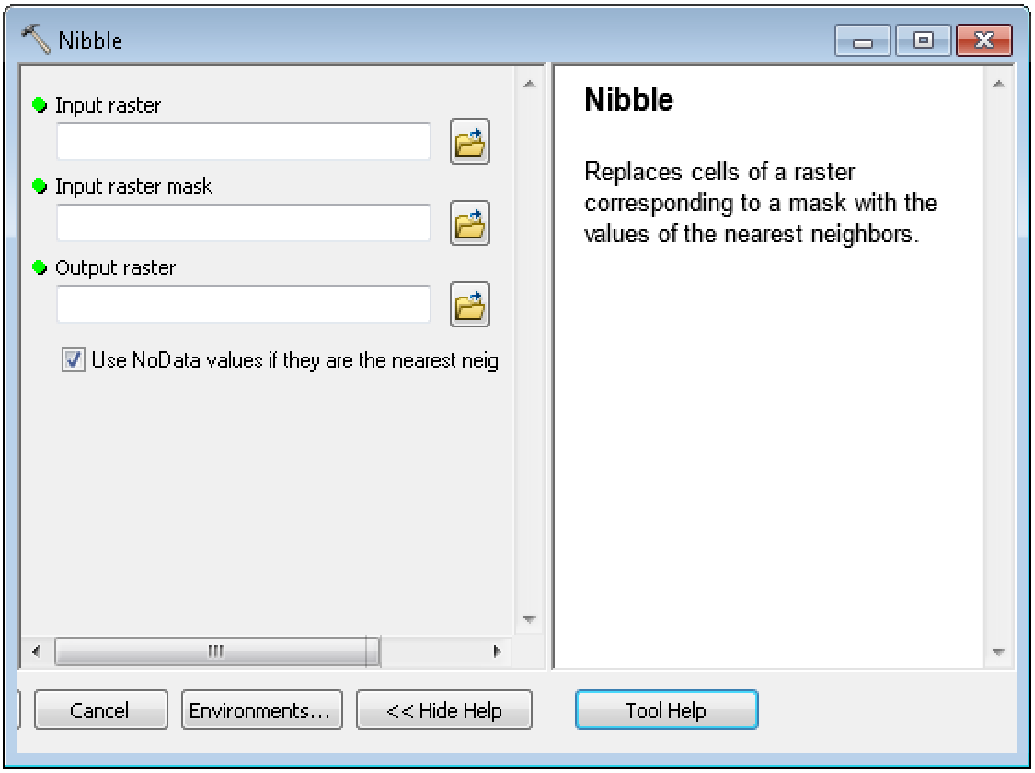
Input raster: Raster (100m)
Output raster: Aggregate_Raster
Cell Factor: **10**
Aggregate: SUM
***Keep checked: Expand extent if needed and Ignore NoData in calculations**
*Set Environments: Set Processing Extent: Kenya_Mask (1,000m), **Raster Analysis: Raster (1000m ~0.008333dd)**

*****Remember to Divide Values by Factor using Raster Calculator*****

Nibble Rasters

Sometimes we have raster files that are smaller than our Mask. In such case we use **Nibble** tool to grow raster to match the Mask raster. Nibble works only with integer raster. If you have a float or continuous raster, you will need to convert it to an integer first. Also nibbling will occur where there are no NoData values in the mask.

Toolboxes\System Toolboxes\Spatial Analyst Tools.tbx\Generalization\Nibble



Nibble

● Input raster

● Input raster mask

● Output raster

☑ Use NoData values if they are the nearest neig

Nibble

Replaces cells of a raster corresponding to a mask with the values of the nearest neighbors.

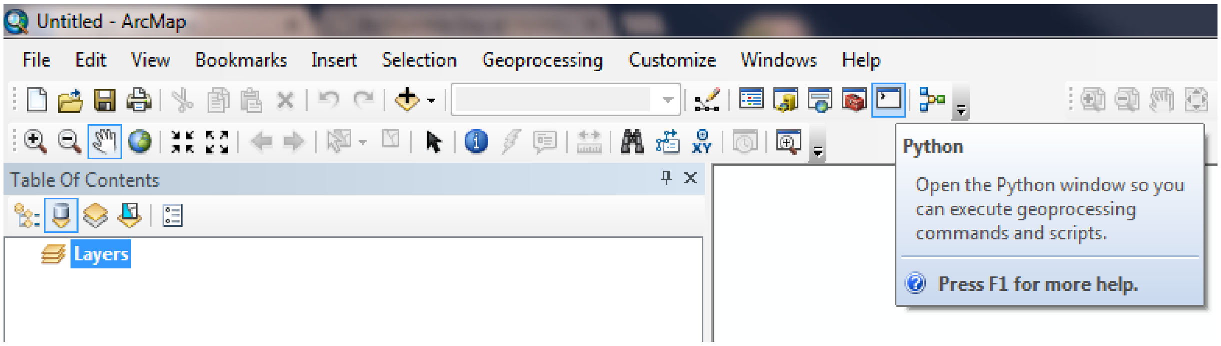
The following steps are needed to run Nibble Tools:

1) Convert input raster to integer using Raster Calculator by multiplying raster by a factor of 10
2) Create nibbling mask
3) Run Nibble
4) Export to Raster

Using Python to Nibble

In Python console:
#: indicates Python comments or notes that are not part of the script,
>: indicates what is typed in the Python window one line at a time.

1. Start Python window to execute commands. Type commands one line at a time.



Python

Open the Python window so you can execute geoprocessing commands and scripts.

Press F1 for more help.

2. Enter the following code in the console:

```
          #import arcpy library from arcpy library
          >import arcpy
           >from arcpy import env
          #import all (*) arcpy spatial analyst libraries
          >from arcpy.sa import *

      #create a new raster called test. Give ‘test’ the values of your raster file for nibbling that is an
      #integer, if not then give it a value of 1
          >test  = Con(IsNull(“<your raster file for nibbling that is an integer>”), 1, “<your raster file for
      nibbling that is an integer>”)

      #Nibble test to the original mask as integer and write values that have data. Don’t let no data
      #values influence decision.
          >nibbled = Nibble(“test”, “<your raster file for nibbling that is an integer>”, “DATA_ONLY”)
```

***Remember to divide values by the factor used to multiply**

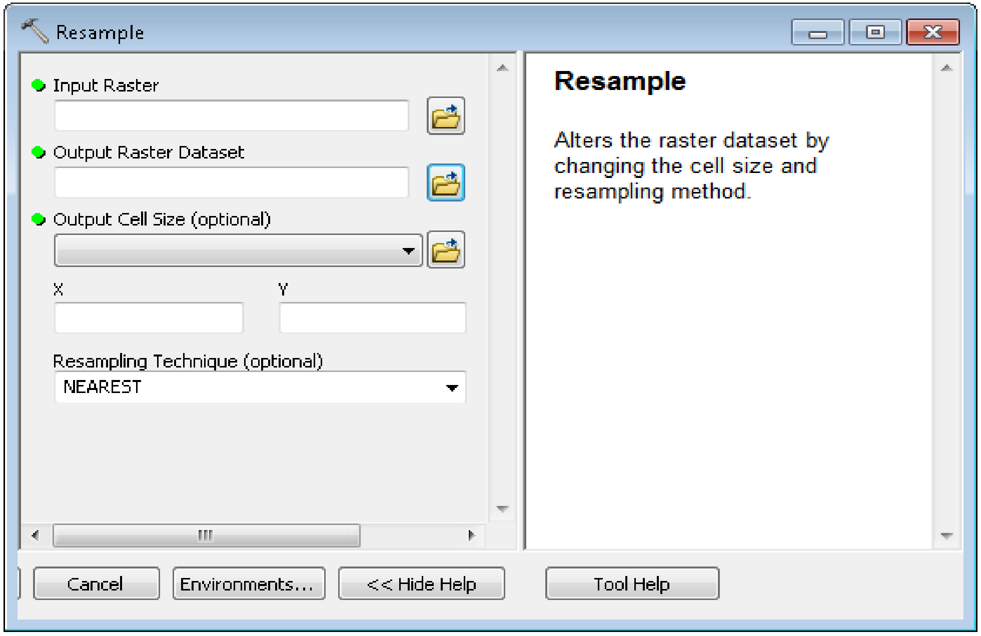
Resample Rasters

Resample (Data Management) tool, ~~a~~alters the raster dataset by changing the cell size and resampling method.

***Before you use resample make sure that your rasters line up.**
*If not use Raster Calculator and snap raster to ‘realign’ raster to Mask:
[Raster] * 1
Set Environments: Set Extent to “kenya_Mask” AND Snap Raster to “kenya_Mask”

Use *Resample* tool to get rasters to desired resolution (5 km). (ArcToolboxes\Data Management Tools.tbx\Raster\Raster Processing\Resample)

Set Environment: Extent to Niger_Mask; Raster Analysis Cell Size: 0.041667 dd (~5km)


Resample
Input Raster
Output Raster Dataset
Output Cell Size (optional)
X
Y
Resampling Technique (optional)
NEAREST
Resample
Alters the raster dataset by changing the cell size and resampling method.
Cancel
Environments...
<< Hide Help
Tool Help

Project Rasters

Make sure all rasters are in the same projection before aggregation. If not use Project Raster (Data Management) to project raster to desired projection. Project Raster tool transforms the raster dataset from one projection to another.

Tool Location: Toolboxes\System Toolboxes\Data Management Tools.tbx\Projections and Transformations\Raster\Project Raster

3.3. Converting to Tabular Data

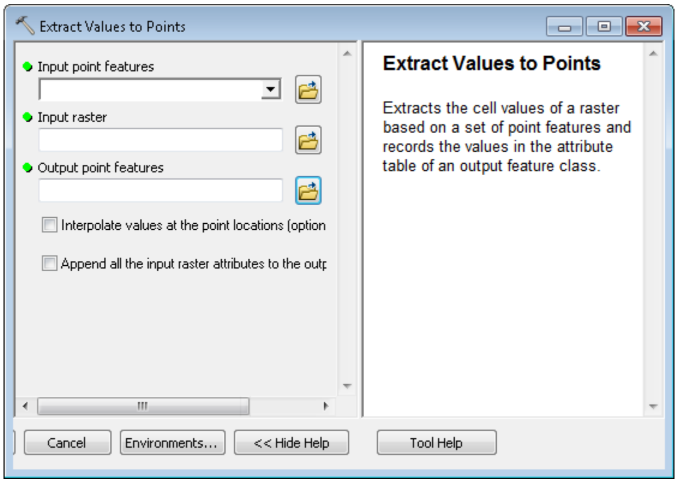
In this session we will convert all raster files into tabular comma separated variable (CSV) format files using a common referencing system identified in day 1.

3.3.1 Extract Values to Point

Once all your data is in raster format you can use your centroids to extract the raster values to the centroids.

Extract the raster cell values from to the centroids using ArcGIS *Extract Values to Points*
ArcToolbox\SpatialAnalyst\Extraction\Extract Values to Points tool
Input point features: Fishnet centroids
Input raster: [Raster] eg ACLED
Output point features: ACLED_Fishnet_Centroids

1. Add field to give field name according to CODE book
2. Choose correct data type
3. Calculate Field to RASTER_VALU
4. Remember to do any special calculations to undo factor multiplications or divisions



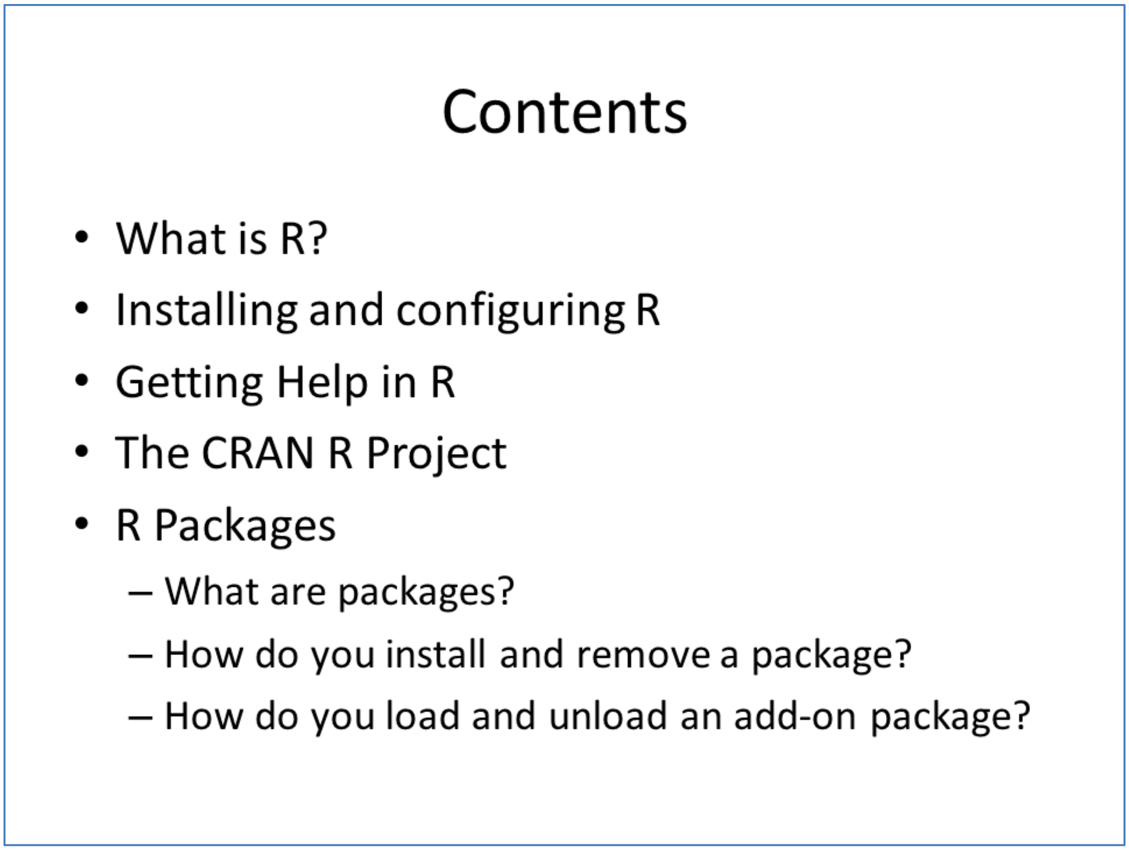
Extract Values to Points

Extracts the cell values of a raster based on a set of point features and records the values in the attribute table of an output feature class.

3.3.2 Export Values to Text Files

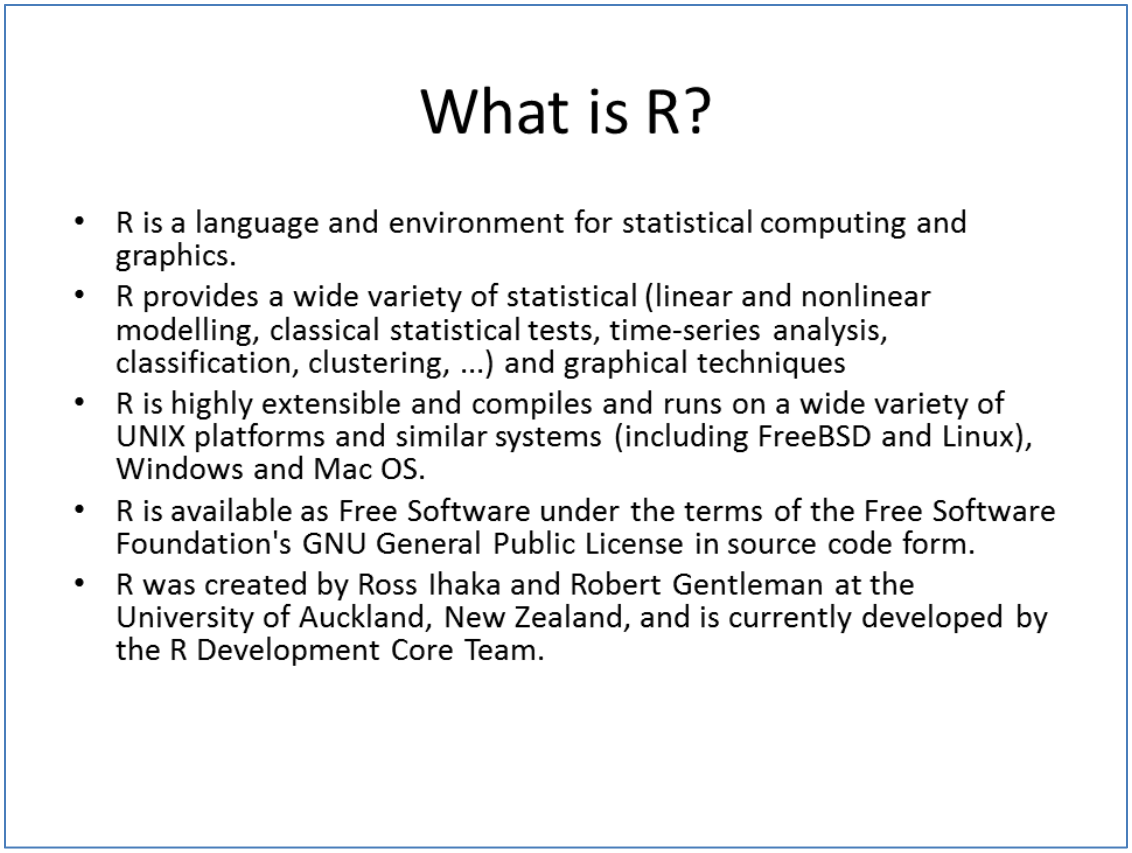
Add the point feature class to ArcMap and open the attribute table. Export table as text. In Windows you can change the file extension to .csv or write your R script to accept .txt

Lecture 4. Introductions to R Statistics Environment



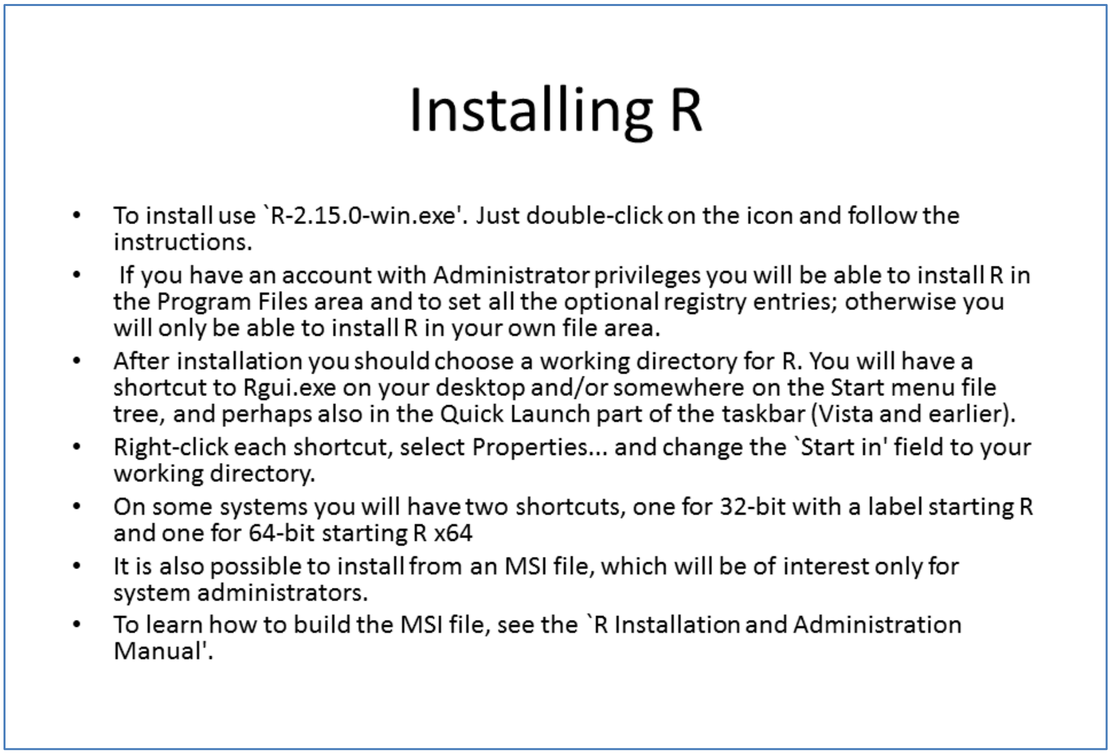
Contents

- What is R?
- Installing and configuring R
- Getting Help in R
- The CRAN R Project
- R Packages
  – What are packages?
  – How do you install and remove a package?
  – How do you load and unload an add-on package?



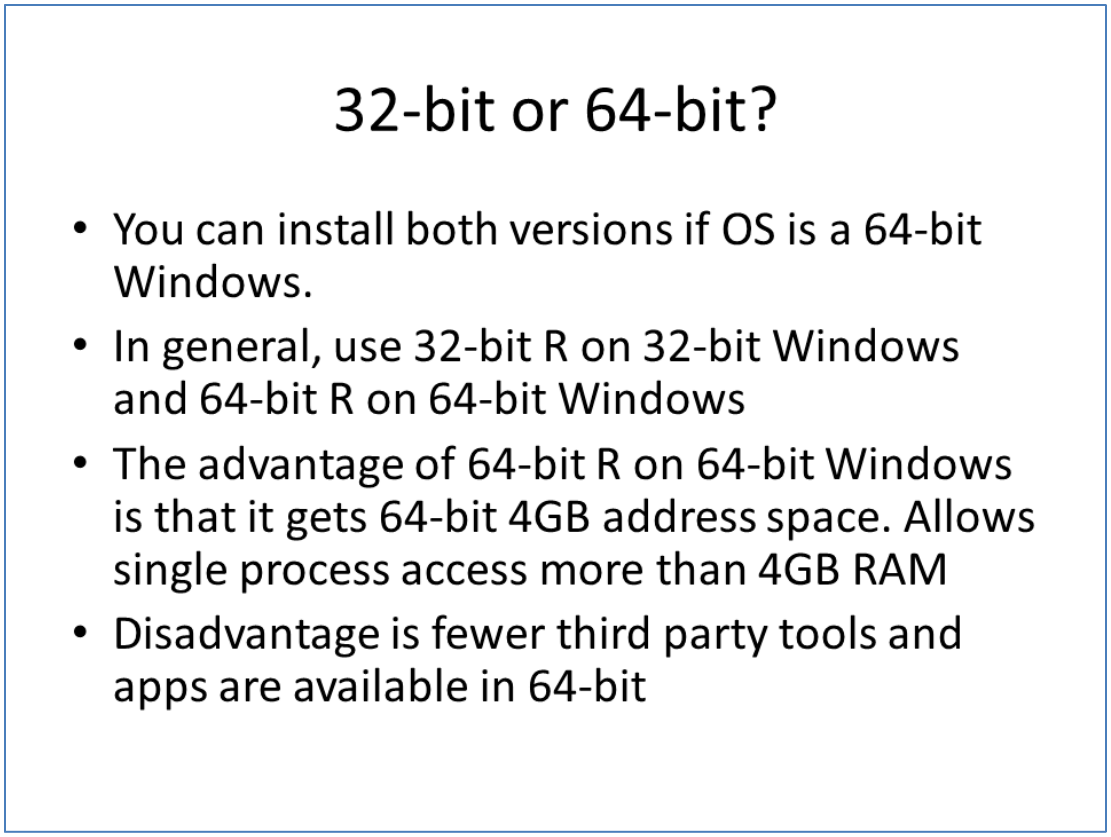
What is R?

- R is a language and environment for statistical computing and graphics.
- R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques
- R is highly extensible and compiles and runs on a wide variety of UNIX platforms and similar systems (including FreeBSD and Linux), Windows and Mac OS.
- R is available as Free Software under the terms of the Free Software Foundation's GNU General Public License in source code form.
- R was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is currently developed by the R Development Core Team.



Installing R

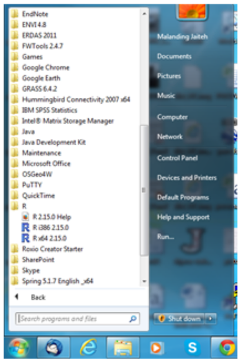
- To install use `R-2.15.0-win.exe'. Just double-click on the icon and follow the instructions.
- If you have an account with Administrator privileges you will be able to install R in the Program Files area and to set all the optional registry entries; otherwise you will only be able to install R in your own file area.
- After installation you should choose a working directory for R. You will have a shortcut to Rgui.exe on your desktop and/or somewhere on the Start menu file tree, and perhaps also in the Quick Launch part of the taskbar (Vista and earlier).
- Right-click each shortcut, select Properties... and change the `Start in' field to your working directory.
- On some systems you will have two shortcuts, one for 32-bit with a label starting R and one for 64-bit starting R x64
- It is also possible to install from an MSI file, which will be of interest only for system administrators.
- To learn how to build the MSI file, see the `R Installation and Administration Manual'.



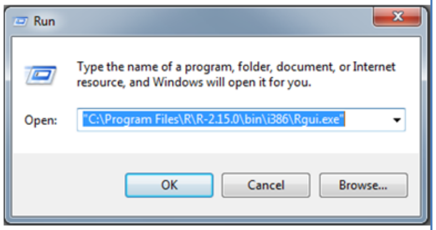
32-bit or 64-bit?

- You can install both versions if OS is a 64-bit Windows.
- In general, use 32-bit R on 32-bit Windows and 64-bit R on 64-bit Windows
- The advantage of 64-bit R on 64-bit Windows is that it gets 64-bit 4GB address space. Allows single process access more than 4GB RAM
- Disadvantage is fewer third party tools and apps are available in 64-bit

Running R



- Double-click on the shortcut generated at installation.
- If you want to set up another project, make a new shortcut or use the existing one and change the `Start in' field of the Properties.
- You may if you prefer run R from the command line by typing "C:\Program Files\R\R-2.15.0\bin\i386\Rgui.exe" in the



Type the name of a program, folder, document, or Internet resource, and Windows will open it for you.

Open: "C:\Program Files\R\R-2.15.0\bin\i386\Rgui.exe"

Update and UN-install R

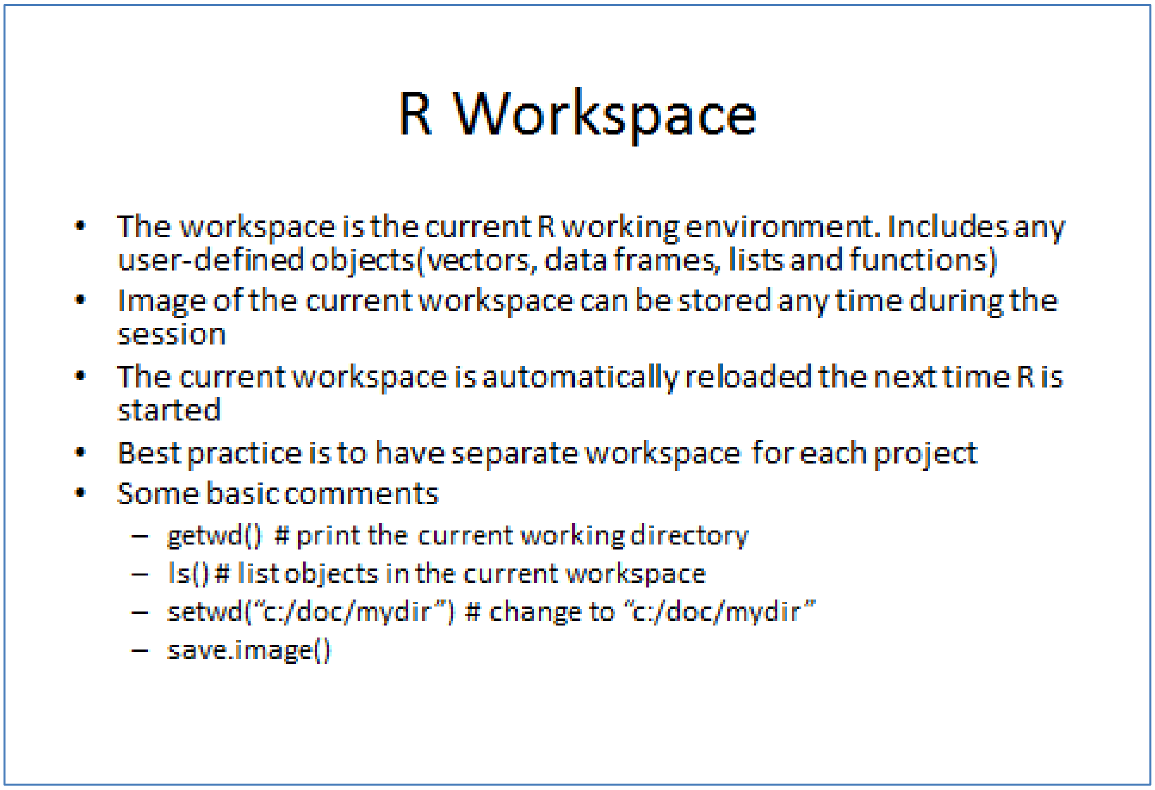
- To update to a newer version, easy thing to do is to uninstall R, install the new version, copy any installed packages to the library folder in the new installation, run update.packages(checkBuilt=TRUE, ask=FALSE) in the new R and then delete anything left of the old installation.
- Different versions of R are installed in parallel folders so you can keep old versions around if you wish.
- For those with a personal library (folder R\win-library\x.y of your home directory, R\win64-library\x.y on 64-bit builds), you will need to update that too when the minor version of R changes (e.g. from 2.14.2 to 2.15.0). A simple way to do so is to copy (say) R\win-library\2.14 to R\win-library\2.15 before running update.packages(checkBuilt=TRUE, ask=FALSE)
- To Uninstall normally you uninstall through `Add/Remove Programs' (XP) or `Programs and Features' (Vista/7) group in the Control Panel. If it does not appear there, run unins000.exe in the top-level installation directory, e.g. C:\Program Files\R\R-2.15.0
- Uninstalling R only removes files from the initial installation, not (for example) packages you have installed or updated.
- If all else fails, you can just delete the whole directory in which R was installed.

Library vs Package

- A *library* is a place (directory) where R knows to find installed packages it can use
- A *package* is a standardized collection of material extending R, e.g. providing code, data, or documentation.
- R is told to use a package (to "load" it and add it to the search path) via calls to the function library.
- library() is employed to load a package from libraries containing packages.

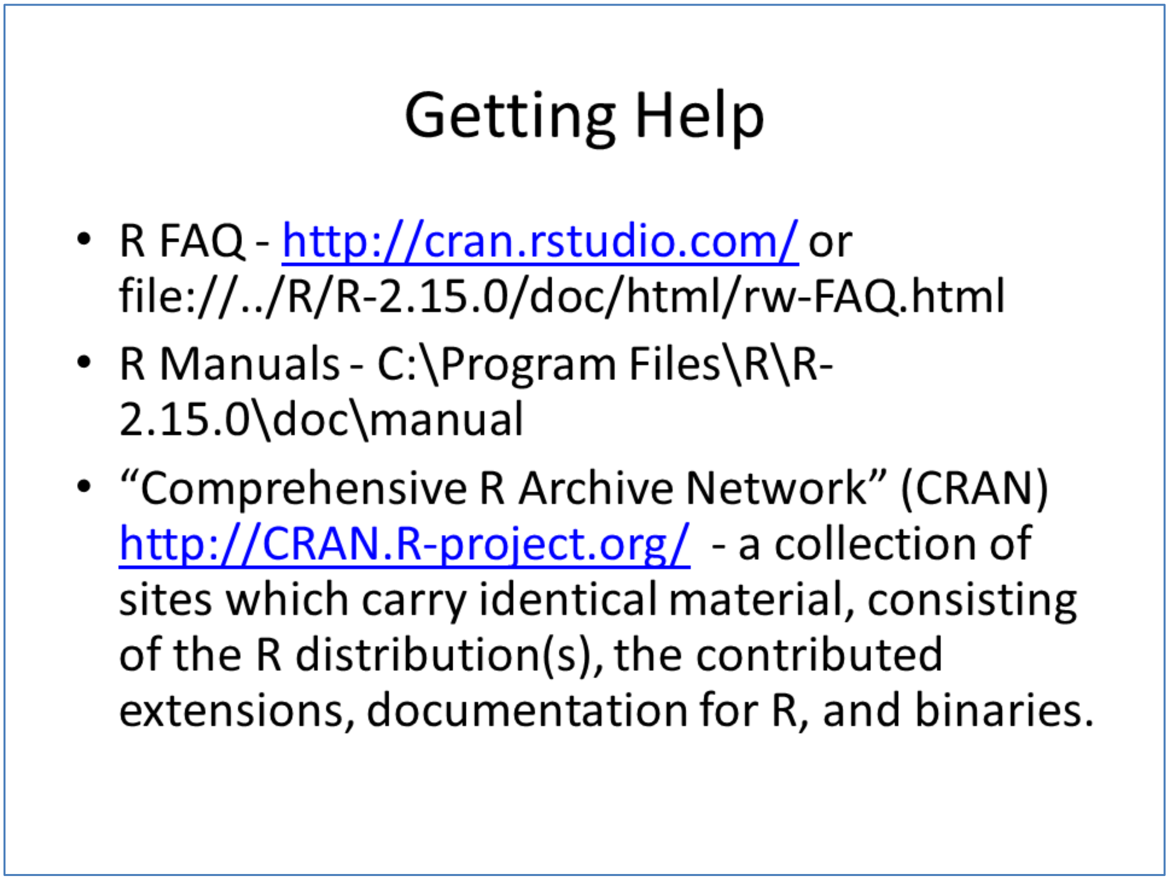
Installing and removing *packages*

- In Window you use the following:
- Install.packages("ggplot2") – Will download and install "ggplot2" package
- Update.packages("plyr") - downloads the list of available packages and their current versions, compares it with those installed and offers to fetch and install any that have later versions on the repositories.
- Use remove.packages() to remove installed packages



R Workspace

- The workspace is the current R working environment. Includes any user-defined objects(vectors, data frames, lists and functions)
- Image of the current workspace can be stored any time during the session
- The current workspace is automatically reloaded the next time R is started
- Best practice is to have separate workspace for each project
- Some basic comments
  - getwd() # print the current working directory
  - ls() # list objects in the current workspace
  - setwd("c:/doc/mydir") # change to "c:/doc/mydir"
  - save.image()



Getting Help

- R FAQ - http://cran.rstudio.com/ or file://../R/R-2.15.0/doc/html/rw-FAQ.html
- R Manuals - C:\Program Files\R\R-2.15.0\doc\manual
- "Comprehensive R Archive Network" (CRAN) http://CRAN.R-project.org/ - a collection of sites which carry identical material, consisting of the R distribution(s), the contributed extensions, documentation for R, and binaries.

Exercise 4. Installing and Configuring R Statistics

R is a language and environment for statistical computing and graphics. R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques, and is highly extensible. A major advantage of R over other statistical software is that it is freely available under the terms of the Free Software Foundation's GNU General Public License in the source code form, thus having fewer restrictions on who gets to use it for what. The Free Software status also allowed many users and developers to build a comprehensive library of functions and packages which are freely available at the Comprehensive R Archive Network (CRAN) http://cran.r-project.org/. CRAN is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R.

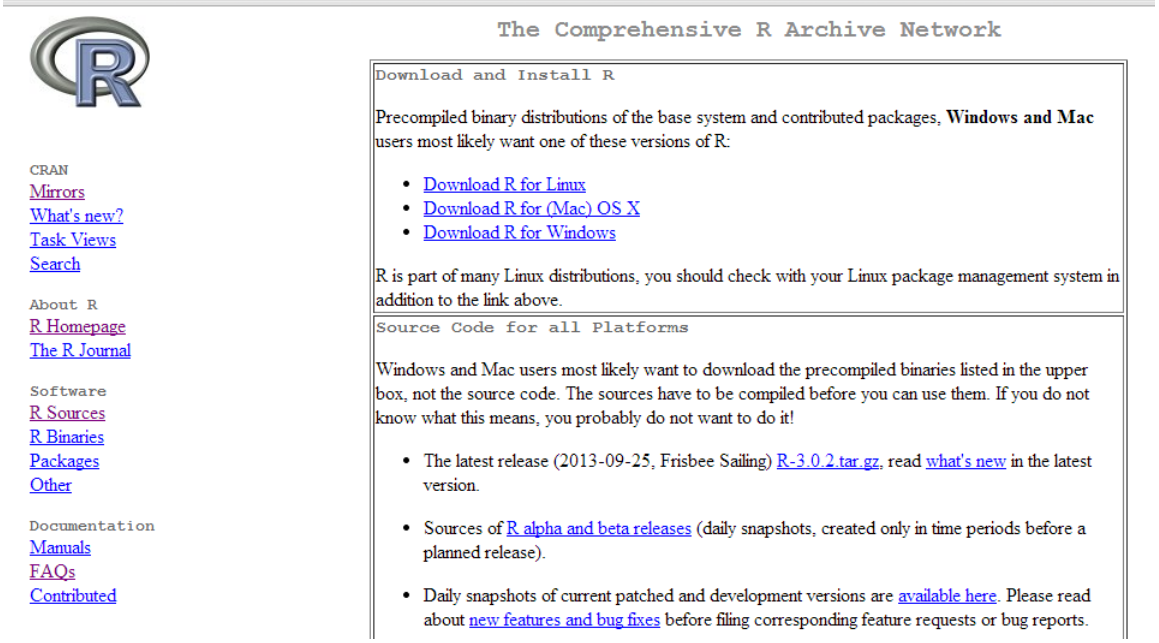
In this tutorial we will:

- Download the latest stable version of R from one of the many CRAN sites
- Install and configure the downloaded file
- Make sure the **Packages** we need for the next two days are all installed and updated

You will only need to perform the first part of this tutorial when you are installing R on your own, with administrative privileges. For most users in organizations installation, setting up environments and paths are done by the network administrator

4.1 Downloading R distribution

R distributions are available for Unix-like, Windows and Mac family of operating systems. The current release of the software is available via CRAN , http://cran.r-project.org/



The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- Download R for Linux
- Download R for (Mac) OS X
- Download R for Windows

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2013-09-25, Frisbee Sailing) R-3.0.2.tar.gz, read what's new in the latest version.
- Sources of R alpha and beta releases (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are available here. Please read about new features and bug fixes before filing corresponding feature requests or bug reports.

The rest of this tutorial will be focused on R for Windows distribution. The latest release of the Windows distribution is R 3.0.2 for Windows. For those installing R on Windows for the first time should download - http://cran.r-project.org/bin/windows/base/R-3.0.2-win.exe

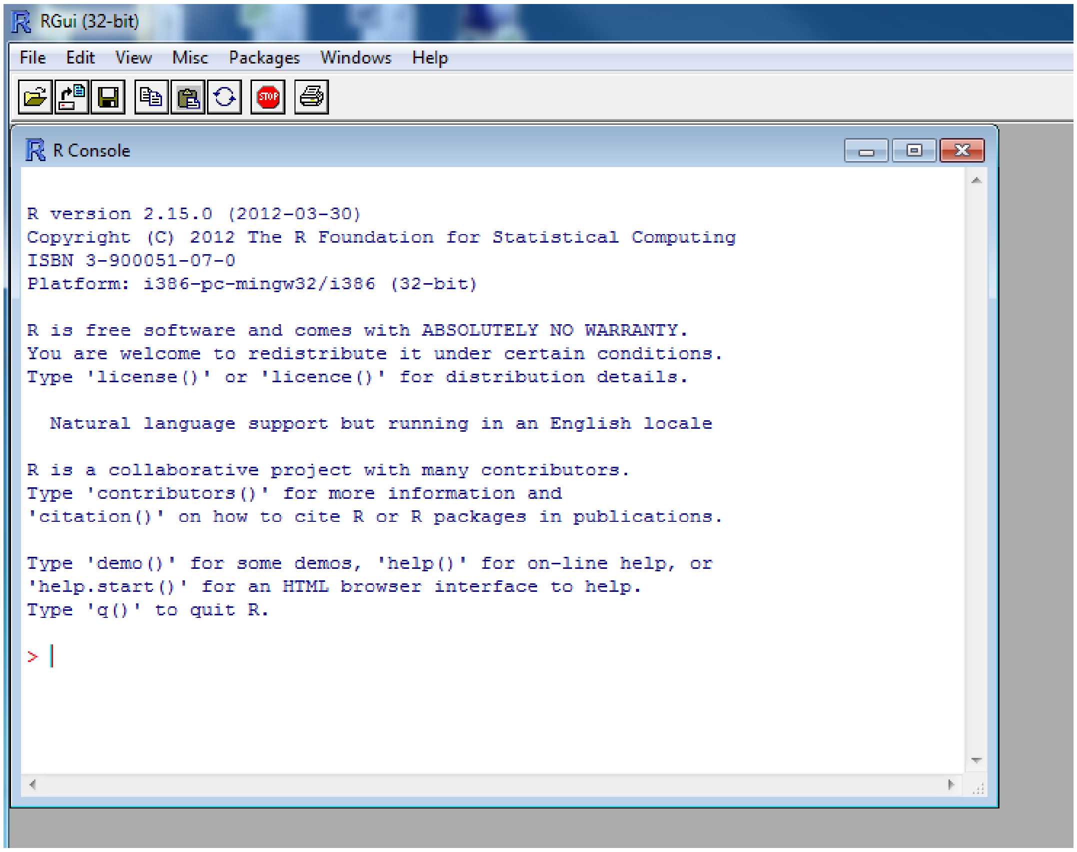
The download file, R-3.0.2-win.exe file size is 51.4MB. Depending on your internet connection, download could take only few minutes.

4.2 Installing R Installer

It is always ideal to install R with Administrator Privileges. This way the environment settings are done automatically and installed packages are available to all users.

1. To install, double-click on the downloaded file and follow the instructions to complete your installation. The installation process on Windows machines will install both the 32-bit and 64-bit R.
2. You should consider installing the open source version of RStudio. This is a recently released IDE (Integrate Development Environment) for using R. RStudio includes a variety of features including editing and managing Codes, Commands, and Files.

For this training we will need the following packages installed. Before proceeding, start R with Administrator Privileges. This will enable you to install or update any package that may be needed.



```
R version 2.15.0 (2012-03-30)
Copyright (C) 2012 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: i386-pc-mingw32/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```

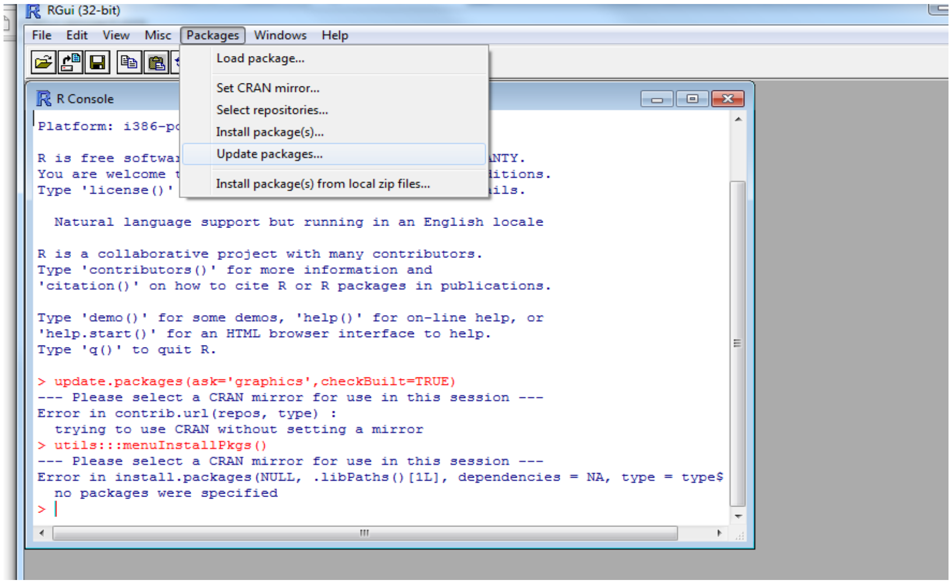
The following packages are essential for the completion of this training.

- ggplot2
- Plyr
- Foreign
- xlsReadWrite
- gridExtra
- lattice

4.2.1 Installing and updating packages

The current release includes most of the packages we need. It is good practice that we use package update to ensure that we have the most recent versions installed.

In the **Packages** menu, you have different options, select **Update packages …** to get the latest version of all installed packages (see below). The Default selection is all packages in need of update. You can narrow the selection by clearing default selection and selecting only those packages you want to update. This feature can be useful in situations where internet connectivity is poor.

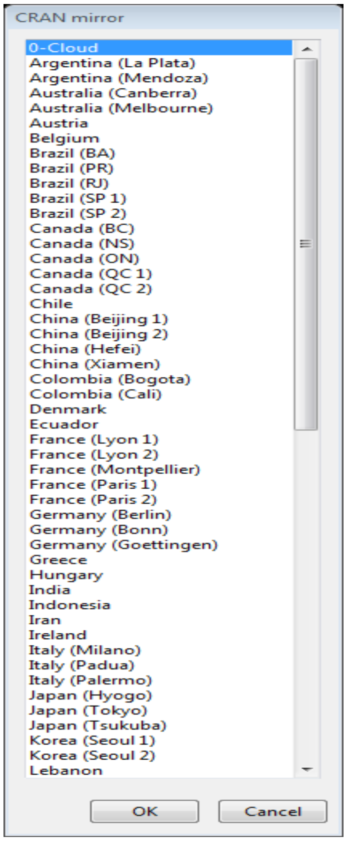


```
Platform: i386-pc

R is free softwar                              NTY.
You are welcome t                              itions.
Type 'license()'                               ils.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> update.packages(ask='graphics',checkBuilt=TRUE)
--- Please select a CRAN mirror for use in this session ---
Error in contrib.url(repos, type) :
  trying to use CRAN without setting a mirror
> utils:::menuInstallPkgs()
--- Please select a CRAN mirror for use in this session ---
Error in install.packages(NULL, .libPaths()[1L], dependencies = NA, type = type$
  no packages were specified
> |
```

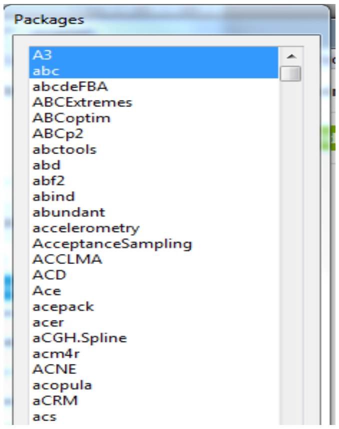
Load package - option will show all packages that are currently installed (see screenshot #). In the event that you cannot find your desired package, you can install packages using the **Install packages** option. The first time you select to install packages, you will get a list of CRAN mirror sites (screenshot #) where you can possibly download desired package; it is wise to download from closest mirror to your location. While it possible to download from any CRAN site, it is advised to download from mirror closest to you. Selecting CRAN site will open the Packages window. This has the list of all packages hosted at the CRAN site that are compatible with your OS. You can select multiple packages by holding Ctrl and select the packages.

Installing in packages or updating packages require connection to the Internet.

Windows Installation on 64-bit Windows will result in the installation of both the R 32-bit and R 64-bit versions. The installation process will place short-cuts for both the 32-bit R i386 2.15.0 and the 64-bit R x64 2.15.0 versions on the Desktop. You can start R by double-clicking on the short-cuts on the Desktop or from the Start/All Programs/R



0-Cloud
Argentina (La Plata)
Argentina (Mendoza)
Australia (Canberra)
Australia (Melbourne)
Austria
Belgium
Brazil (BA)
Brazil (PR)
Brazil (RJ)
Brazil (SP 1)
Brazil (SP 2)
Canada (BC)
Canada (NS)
Canada (ON)
Canada (QC 1)
Canada (QC 2)
Chile
China (Beijing 1)
China (Beijing 2)
China (Hefei)
China (Xiamen)
Colombia (Bogota)
Colombia (Cali)
Denmark
Ecuador
France (Lyon 1)
France (Lyon 2)
France (Montpellier)
France (Paris 1)
France (Paris 2)
Germany (Berlin)
Germany (Bonn)
Germany (Goettingen)
Greece
Hungary
India
Indonesia
Iran
Ireland
Italy (Milano)
Italy (Padua)
Italy (Palermo)
Japan (Hyogo)
Japan (Tokyo)
Japan (Tsukuba)
Korea (Seoul 1)
Korea (Seoul 2)
Lebanon



A3
abc
abcdeFBA
ABCExtremes
ABCoptim
ABCp2
abctools
abd
abf2
abind
abundant
accelerometry
AcceptanceSampling
ACCLMA
ACD
Ace
acepack
acer
aCGH.Spline
acm4r
ACNE
acopula
aCRM
acs

Exercise 5. Getting Started with R

Learning Objectives

Part II of vulnerability mapping is completed in R Statistics. The goal of this exercise is to introduce beginners to the R environment and how it works. By the end of the exercise you will learn ~~about R~~ how to:

- Start and exit the R environment
- Setup workspaces and navigate between directories
- Learn about the different data types
- Import and manipulate data frames
- Generate basic statistics and charts

Exercise data

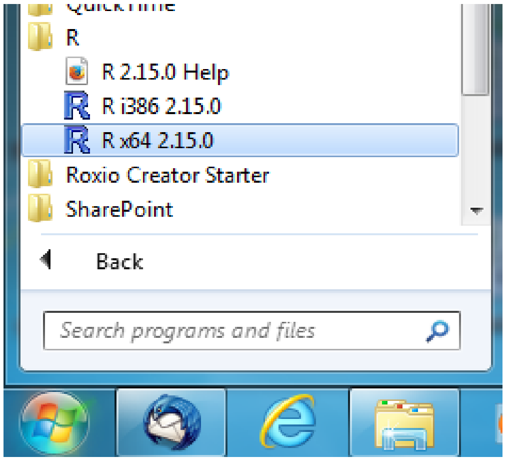
Three data files used in this exercise are contained in the data.zip. You will need to uncompressed the file into your ../VIA_R folder.

The files are:

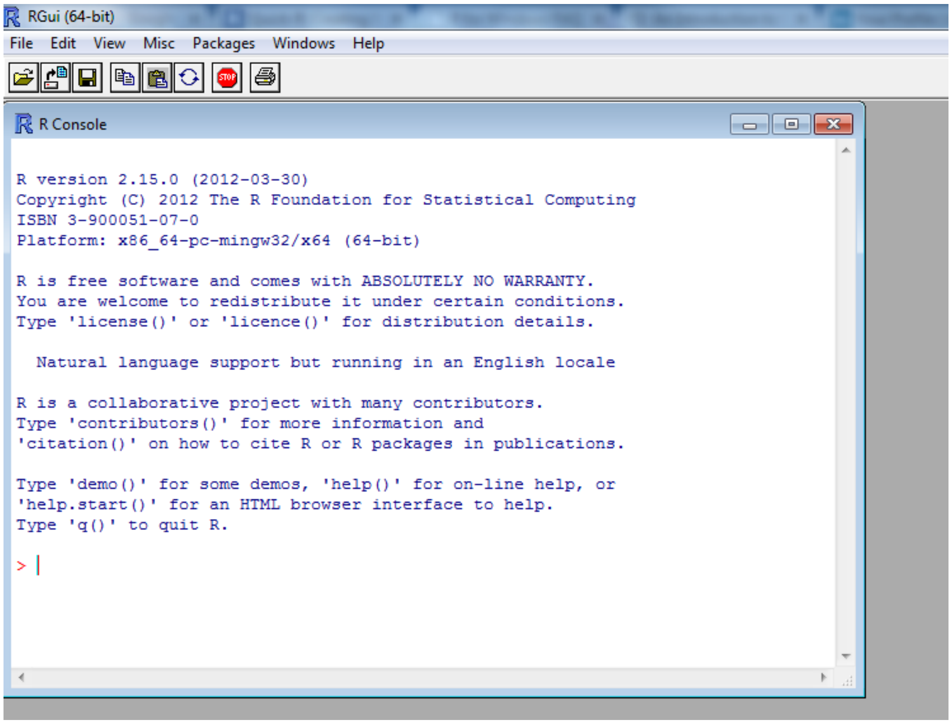
- FEWSRFE.CSV .csv file
- FEWS.TXT Text file works very much like csv file.
- POP.XLS – Excel file

5.1 Start the R environment and setting up workspaces

1. Start R by either *double-clicking* on the on your desktop or through the Start/All Programs/R/R x64 2.15.0



Starting R opens in the R Console window. The Console window is the default R Graphic User Interface (RGUI) allows interaction with the R environment. The '>' symbol by default indicates R is waiting for a command while '#" indicates the statement is a comment and not a command. R execute only a command not a comment!



```
R version 2.15.0 (2012-03-30)
Copyright (C) 2012 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: x86_64-pc-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```

2. To set your working directory:
    - Check your current directory with **getwd().** If R is newly installed on a Windows machine, the default working directory is C:/Windows/systems32.
    - Change working directory to our project folder with **setwd("D:/VIA/VIA_R")**

Note: R writes path with forward slash '/' instead of backslash '\'. It is desirable to create a separate working directory for each new project. R is *case sensitive*. getwd() is not the same as GETWD().

3. Next we will change working directory to our project folder. First, using Windows Explorer, create a new folder in the location you desire. Call it VIA_R. The path to the folder is "D:\VIA\VIA_R"

    Type **setwd("D:/VIA/VIA_R")** to set working directory to the folder you just created.
    Use get working directory **function getwd()** to confirm the change in directory

Data Types

R has a wide variety of data types are scalars, vectors (numerical, character, logical), data frames, and lists.

Vectors

The simplest data type in R, are Vectors. There are three main types of vector data namely, numeric, character and logical. A vector assignment results in an **object** in R

4. To set up a vector named x, say consisting of one number, namely 5,

```
>x <- 5
```

The symbol '**<-**' denotes an assignment. Here **x** is assigned the value 5.
Type a and <enter> to confirm assignment

```
>a <enter>
```

R screen output

```
> a
[1] 5
```

output confirms **a** is now 5

[1], indicates the display starts at the first element of **a**

A vector can consist of a list of numbers. e.g. y consisting of all prime numbers between 0 and 10

```
>y <- c(2,3,5,7)
```

The function **c()** denotes concatenating list. ***Concatenating list*** is given list argument "2,3,5,7" . The result is an object whose components are the argument list "2,3,5,7"

```
>y <enter>
> y
[1] 2 3 5 7
```

confirms **y** is a list of prime number between 0 and 10, "2,3,5,7"

Your turn.

Now try the following vector assignments:

```
a <- 5
b <- 6
```

c <- a * b # multiply a by b
d <- a / b

Q1. What is the value of a * b…………………..?

Q2. What is the value of **c** =……..?

We use the function **a[c(xxx)]** to refer to elements in a vector. For instance we can chose to display only the 1st and 4th elements in the vector **y <- c(2,3,5,7,)** by using:

- y[c(1,4)] # will display 1st and 4th elements (2 and 7) in the list
- y[c(2,3)] # will display the second and 3 elements in the list

Q3. Write the function to display only 3,5 and 7 ………………………….…………….

Table. Types of vectors

Vector	Components	type
a <-	5	Numeric vector
b<-	c(2,3,5,7)	Numeric vector
c<-	c("Columbia","Brown", "Princeton")	Character vector
d<-	c(TRUE, FALSE, TRUE, TRUE)	Logical vector

Lists

In R, a *list* is an ordered collection of objects (components). A *list* allows us to gather a variety of (possibly unrelated) objects under one name. In GIS context a *list* refers to a record in an attribute data and the objects (components) are attributes (items). You can create as many attributes as you like for a record.

A *list* is created with the function **list()**. The object names must be specified when creating a list else the **names()** function will return NULL.

5. First create 2 vectors, x and y

   x<- c(1:4) # creates a vector x<- c(1,2,3,4)
   y<- c(2:4) # creates a vector y <- c(2,3,4,)

6. Using the following vectors, x <-(1:4) and b<-(2:4) create L1 and L2

   L1 <- list(x,y)
   L2 <- list(A=x, B=y)

   Type L1 to see the content of the list

   > L1 <enter>

   - [[1]] #vector sequence number, 1
   - [1] 1 2 3 4 # values in the vector

   [[2]] # vector sequence number 2
   [1] 2 3 4 # values in the vector

   Type L2 to see content of the list

   > L2 <enter>
   $A #vector sequence number, 1

```
[1] 1 2 3 4
$B  # vector sequence number 2
[1] 2 3 4
```

Names () is functions to get or set the names of an object. For further details, please see (http://astrostatistics.psu.edu/su07/R/html/base/html/names.html)

Q. Use **names()** function to list the labels for the vectors in each list

L1 …………………………………………..
L2 ……………………………………………

Data Frames

A data frame is more like an Excel spreadsheet, SAS or SPSS datasets. A data frame contains individual columns and rows. You can create a Data Frame by importing a table form another format using the **read.table** function.

The following example demonstrate working with data frames

5.2 Importing and Exporting Data Frames

Importing Files

R has specialized functions to import from and export data frames to other formats. Depending on the format you will need to load the **foreign** packages, you may need to download the needed packages as explain in the section 4.2.1. For Excel, you need the **xlsReadWrite** package. In the following steps we will import data from .TXT, .CSV and .XLS format.

7. Start a new R session and change working directory to " **../VIA/VIA_R".** The directory contains the following POP.XLS (Excel format), FEW.TXT, Text file and

   Make sure to load the following packages:
   Foreign
   ROCBC
   xlsReadWrite

8. Import the FEWS.TXT data in R using the following:

   ```
   > FEWS <- read.table("FEWSRFE.TXT", header = TRUE, sep = ",", quote="\"")
   ```

   Note: **header = TRUE** means there is a header in the table; **sep= ","** – denotes the separator in the data is a comma (,); and quotes are denoted by " "

   This creates a data frame FEWS by importing the FEWS.TXT table
   >View(FEWS) to view the imported table

9. Importing POP.XLS file entails the following steps:

   a. call the RODBC library **Note:** the RODBC library only works within the R 32-bit installation
   >library(RODBC)

b. Assign the POP data frame equal to POP.XLS data:
> POP = odbcConnectExcel(file.choose())
file.choose() function starts a add file dialog box to allow you navigate to the "../VIA" folder and select the POP.XLS

c. Import the .CSV (comma separate variable) file using:
> FEWSRFE<- read.csv("G:/PREPARED/VIA Workshop/FEWSRFE.CSV", header = TRUE, sep = ",", quote="\"")

You now have all three data frames imported in R. You can use the following functions to explore the content:

- names(data frame) – e.g. names(FEWS) –
- View(data_frame) # note use of Upper case V
- hist(data_frame$column)
- quantile(POP$, 0.025, c(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)

Now we will join the FEWS and POP data frames using the PixelID as the join field.

First use the name() function to list the fields in each of the tables. Note the sequencing of the fields.

This requires that you to download **plyr library**
>Install.package("plyr")
>library(plyr) – loads the plyr package
Then use the following line of code to join the tables:

```
>join1<-join(FEWS[,5:6],POP[,5:6],match="first", by="PixelID")
```

The newly created larger table is called join1
The join function joins columns 5 and 6 in the POP data frame to columns 5 and 6 in the FEWS data.

PixelID is used as the join field and joining starts with first pixel in each data frame. It is important to make sure that each of the files is of the same sequencing.

Exporting data

Like import, R allows us to export data frames to a number of formats. Use the following code to export the join1 table to a comma separated variable (.CSV) file.:

```
write.table(join1, file="g:/PREPARED/VIA Workshop/joinT.csv",append = TRUE, sep= ",")
```

5.3. Learning More

Other Useful Functions

Try out the following functions and note what they do.

str(object) # structure of an object

class(object) # class or type of an object

names(object) # names

c(object,object,...) # combine objects into a vector

cbind(object, object, ...) # combine objects as columns

rbind(object, object, ...) # combine objects as rows

object # prints the object e.g typing the name of variable b will print the values of b

ls() # list current objects

rm(object) # delete an object

rbind(object, object, ...) # combine objects as rows

fix(object) # edit in place

length(object) # number of elements or components

sum(object) # sums up all values in a data frame or column. Use sum(data_frame$column) to sum values in column

data.frame(object)

Online Tutorials

An Introduction to R- Notes on R: A Programming Environment for Data Analysis and Graphics- http://cran.r-project.org/doc/manuals/R-intro.pdf

R for beginners by Emmanuel Paradis - http://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf

http://www.statmethods.net/ - Site sun by Robert I. Kabacoff providing basics on working with R

http://www.cyclismo.org/tutorial/R/genindex.html

Exercise 6. Getting Started with R

In exercise 2-3 we have processed the various input layers using various GIS methods, converted them to raster and exported each individual raster to a comma separated variable table. Each table contains a field containing a raster sequence number and a column representing variable of interest. In this exercise we will:

- Setup a project and directories
- Import all the tables into data frames in R to create one large VIA data frame~~,~~
- Use statistical data exploratory methods to determine statistical distribution of the raw data
- Variable transformations: clean, recode or winsorize
- Rescale each indicator to a 0-100 scale, where 0 equates to lower vulnerability and 100 equates to high vulnerability (invert the raw values from original data where necessary) Combine indicators to produce component scores
- Rescale component scores to 0-100
- Combine component scores to create final vulnerability map
- Principal Components Analysis and its interpretation

In exercises 4a and 4b we used command line RGui to handle vectors, list and data frames. In this exercise we will use RStudio IDE to implement the rest of the data processing.

6.1. R Studio Console

When started for the first time the RStudio GUI has three panels, namely Console, Environment/ History and Files/Plots/packages?Viewer/Help panel.

a. The Console panel is the R command panel. This is where to type in commands such as getwd() to inquire about which folder R is using as Working Directory. The Console panel divides into 2 when a data is imported. The upper half of the panel turns into a data viewer (to help visualize the data frame active in the Environment panel. The lower part of the panel remains as the Console panel.

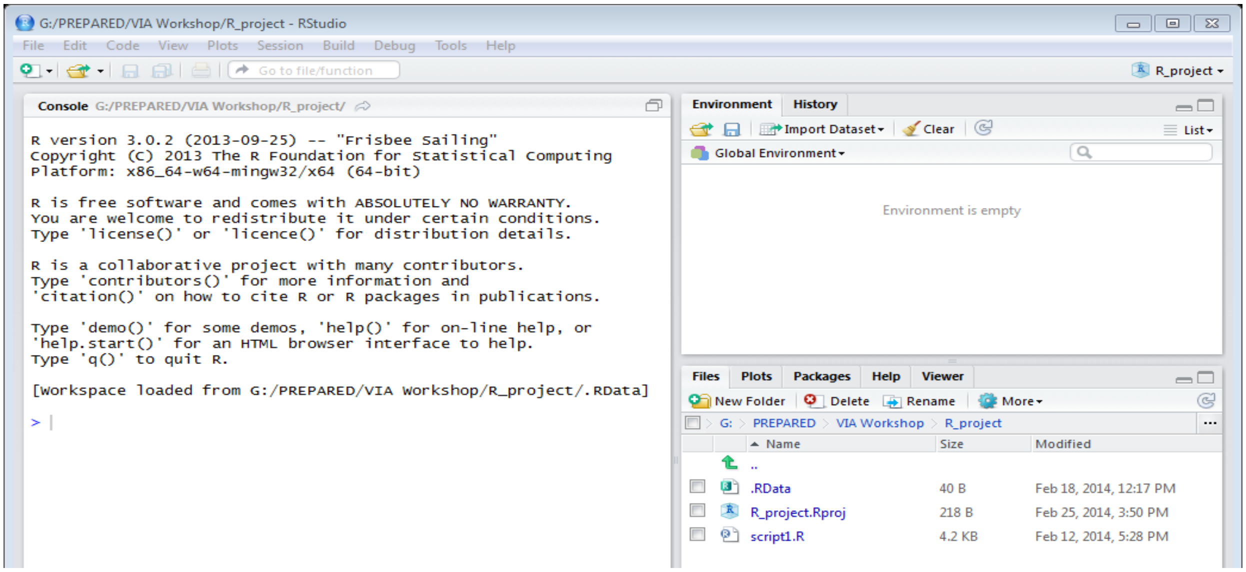
b. The Environment /History panel – The Environment panel shows the data frames in the workspace. This panel allows us to open saved environments, or import data sets. The History panel contains the list of commands and functions used in the session. You can send commands from in this panel to the Console or to Source (new scripts).

c. The Files/Plots/Packages/Help/Viewer panels – We manage files in the Files panel; Packages in Packages. Plots are drawn in the Plots panel.

6.1.1 Create a new Project

Working in RStudio IDE allows you to create separate projects for your work. Projects make it easier to divide your work into multiple contexts each with their own working directory, workspace, history and source documents.

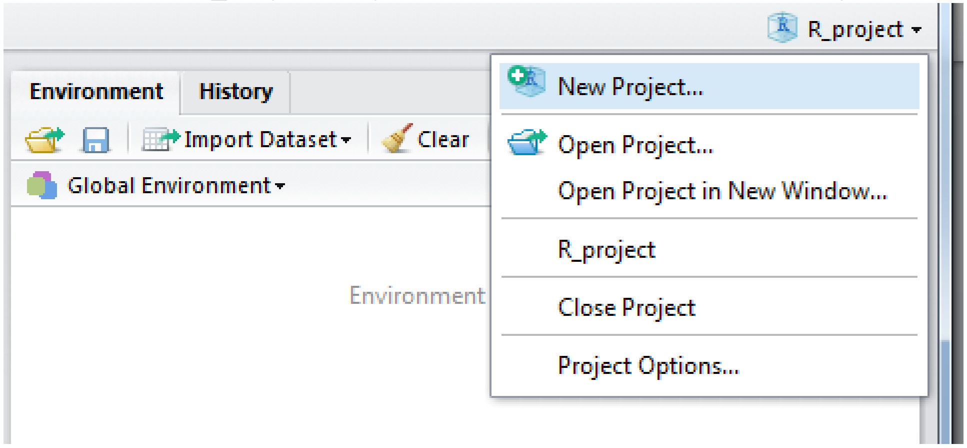
1. Start a new R Studio session from Start/All Programs/RStudio/



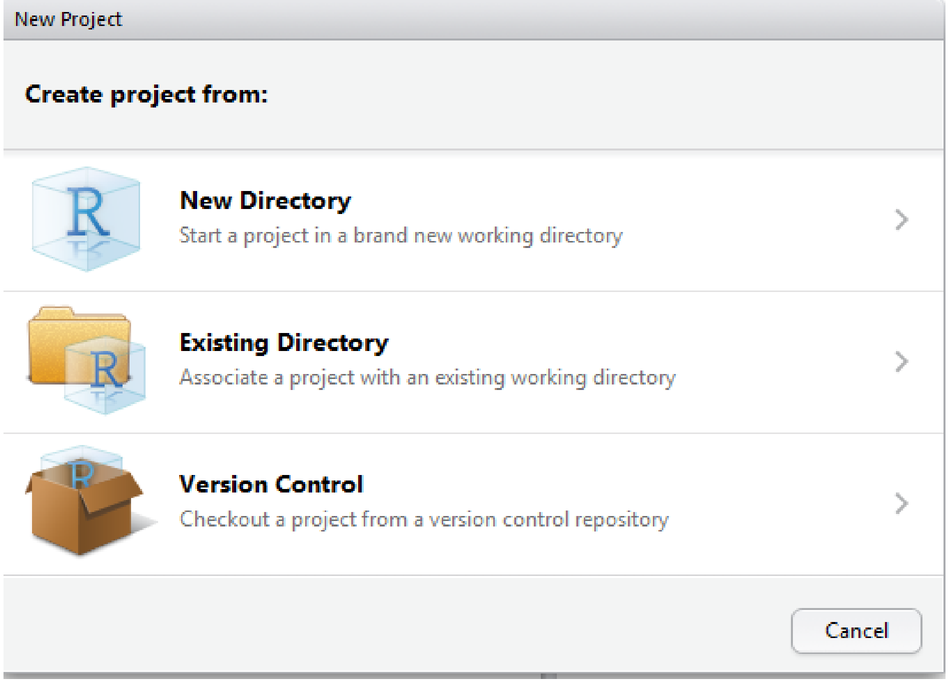
```
R version 3.0.2 (2013-09-25) -- "Frisbee Sailing"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from G:/PREPARED/VIA Workshop/R_project/.RData]

> |
```

Name	Size	Modified
..		
.RData	40 B	Feb 18, 2014, 12:17 PM
R_project.Rproj	218 B	Feb 25, 2014, 3:50 PM
script1.R	4.2 KB	Feb 12, 2014, 5:28 PM

2. Click on the R_Project drop-down window and select New Project.



Select Create project from: Existing Directory. This will allow you to navigate to the CSV folder containing the .CSV files.


New Project
Create project from:
New Directory
Start a project in a brand new working directory
Existing Directory
Associate a project with an existing working directory
Version Control
Checkout a project from a version control repository
Cancel

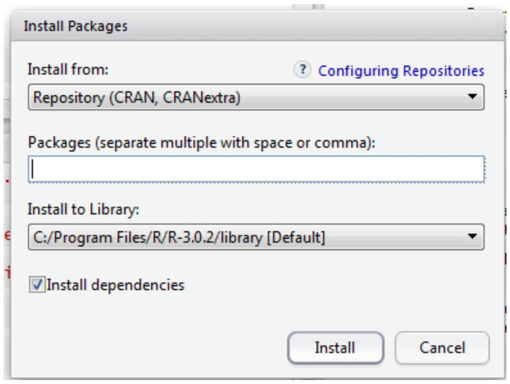
A new project in RStudio:

- Creates a project file with .Rproj extension with the directory chosen.
- Creates a hidden directory names .Rproj.user where project-specific temporary files are stored
- Loads the project into RStudio and displays its name in the Projects toolbar. In our case out new project is **CSV.Rproj**

3. In the **Packages** panel, check the box next to the following packages to load.

plyr - for join command
ggplot2 - for nice graphics
lattice - for dotplots (in case you need PCA)
stats - prcomp(in case you need PCA)
Hmisc - for calculating correlations

If in the process you get an error that package failed to load as the dependencies are not found, you can search and install the dependencies by clicking on the Install Packages button in the panel.



Install Packages

Install from: Repository (CRAN, CRANextra)

Packages (separate multiple with space or comma):

Install to Library: C:/Program Files/R/R-3.0.2/library [Default]

☑ Install dependencies

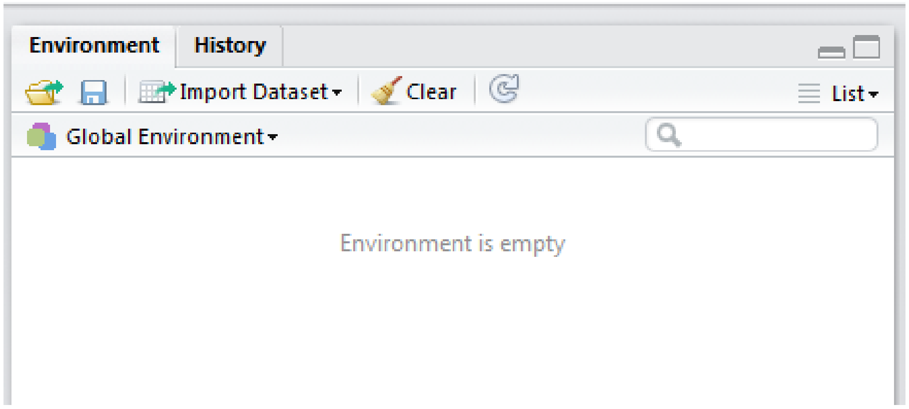
6.1.2 Reading CSV tables into R

Each variable considered for inclusion in the calculation of the vulnerability is available in CSV format. The variable tables include the point values of raster files created in exercise 2, and a raster sequence (PIXELID) field. We will use this field to merge the single variable data files into one data spreadsheet.

We are now ready to read in the .CSV files representing the input variables

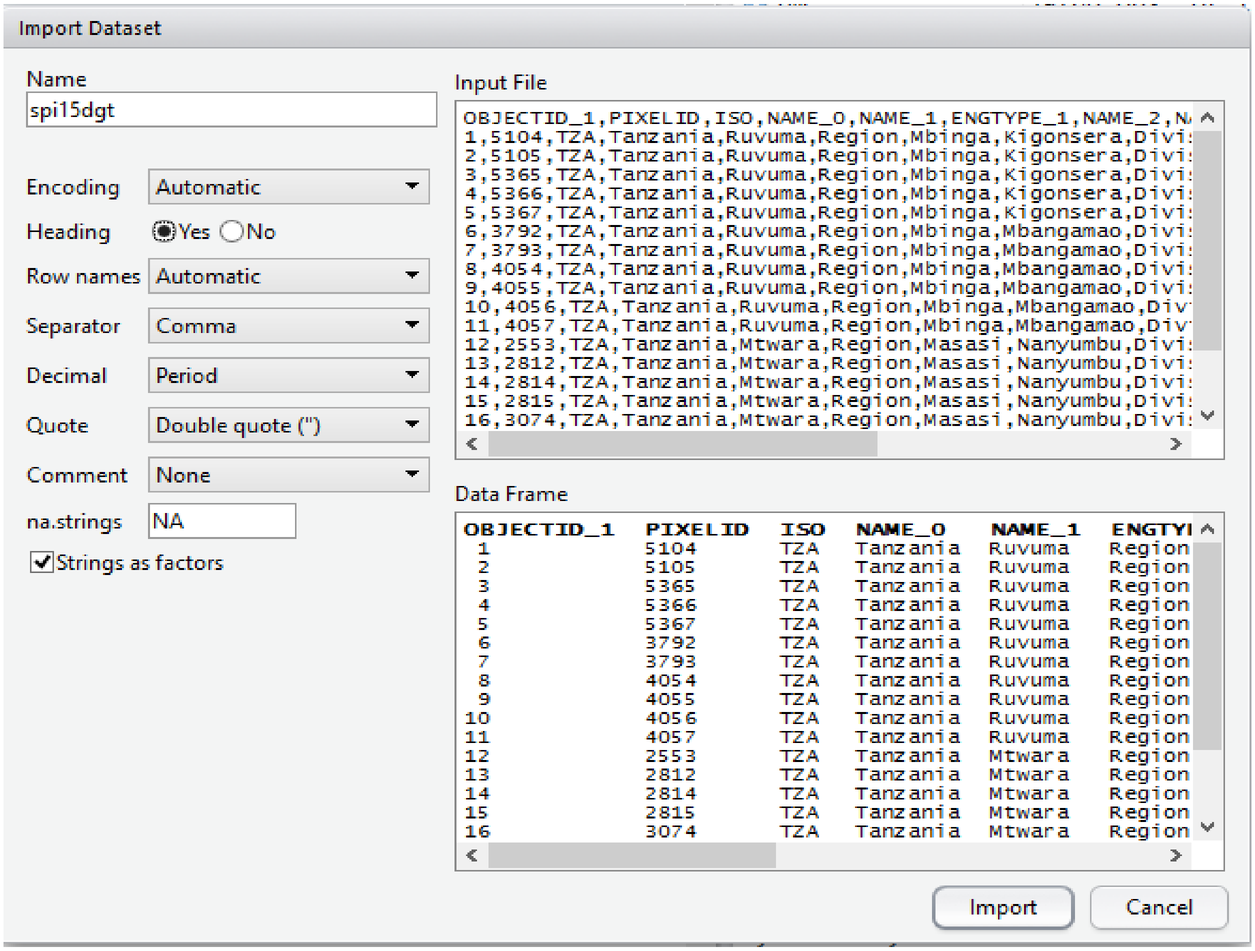
.

1. Click on the **Import Dataset** button in the Environment panel to start the RStudio import wizard~~.~~, then select from Text File



Environment is empty

This opens the *Select File to Import Dataset* dialog window.



Import Dataset

Name: spi15dgt

Encoding: Automatic

Heading: ● Yes ○ No

Row names: Automatic

Separator: Comma

Decimal: Period

Quote: Double quote (")

Comment: None

na.strings: NA

☑ Strings as factors

Input File

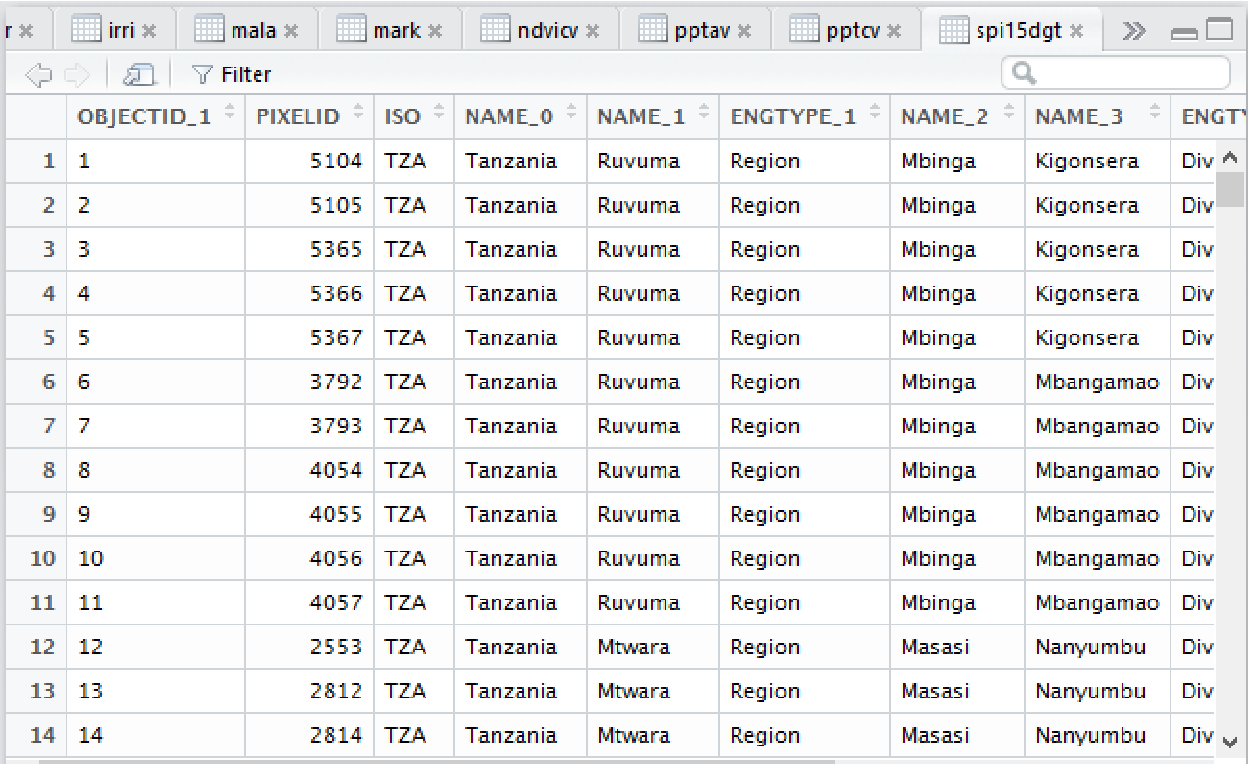
```
OBJECTID_1,PIXELID,ISO,NAME_0,NAME_1,ENGTYPE_1,NAME_2,N
1,5104,TZA,Tanzania,Ruvuma,Region,Mbinga,Kigonsera,Divi
2,5105,TZA,Tanzania,Ruvuma,Region,Mbinga,Kigonsera,Divi
3,5365,TZA,Tanzania,Ruvuma,Region,Mbinga,Kigonsera,Divi
4,5366,TZA,Tanzania,Ruvuma,Region,Mbinga,Kigonsera,Divi
5,5367,TZA,Tanzania,Ruvuma,Region,Mbinga,Kigonsera,Divi
6,3792,TZA,Tanzania,Ruvuma,Region,Mbinga,Mbangamao,Divi
7,3793,TZA,Tanzania,Ruvuma,Region,Mbinga,Mbangamao,Divi
8,4054,TZA,Tanzania,Ruvuma,Region,Mbinga,Mbangamao,Divi
9,4055,TZA,Tanzania,Ruvuma,Region,Mbinga,Mbangamao,Divi
10,4056,TZA,Tanzania,Ruvuma,Region,Mbinga,Mbangamao,Div
11,4057,TZA,Tanzania,Ruvuma,Region,Mbinga,Mbangamao,Div
12,2553,TZA,Tanzania,Mtwara,Region,Masasi,Nanyumbu,Divi
13,2812,TZA,Tanzania,Mtwara,Region,Masasi,Nanyumbu,Divi
14,2814,TZA,Tanzania,Mtwara,Region,Masasi,Nanyumbu,Divi
15,2815,TZA,Tanzania,Mtwara,Region,Masasi,Nanyumbu,Divi
16,3074,TZA,Tanzania,Mtwara,Region,Masasi,Nanyumbu,Divi
```

Data Frame

```
OBJECTID_1    PIXELID    ISO   NAME_0     NAME_1    ENGTYI
 1            5104       TZA   Tanzania   Ruvuma    Region
 2            5105       TZA   Tanzania   Ruvuma    Region
 3            5365       TZA   Tanzania   Ruvuma    Region
 4            5366       TZA   Tanzania   Ruvuma    Region
 5            5367       TZA   Tanzania   Ruvuma    Region
 6            3792       TZA   Tanzania   Ruvuma    Region
 7            3793       TZA   Tanzania   Ruvuma    Region
 8            4054       TZA   Tanzania   Ruvuma    Region
 9            4055       TZA   Tanzania   Ruvuma    Region
10            4056       TZA   Tanzania   Ruvuma    Region
11            4057       TZA   Tanzania   Ruvuma    Region
12            2553       TZA   Tanzania   Mtwara    Region
13            2812       TZA   Tanzania   Mtwara    Region
14            2814       TZA   Tanzania   Mtwara    Region
15            2815       TZA   Tanzania   Mtwara    Region
16            3074       TZA   Tanzania   Mtwara    Region
```

Make sure the dataset name in the Name field is correct, if there is Heading in the data or not and if the other variables (separator, Decimal and Quote) are correct.

The Import Dataset window also displays the Input File and the Output Data Frame. This allows you to confirm that all is well before hitting Import button to complete.



	OBJECTID_1	PIXELID	ISO	NAME_0	NAME_1	ENGTYPE_1	NAME_2	NAME_3	ENGTY
1	1	5104	TZA	Tanzania	Ruvuma	Region	Mbinga	Kigonsera	Div
2	2	5105	TZA	Tanzania	Ruvuma	Region	Mbinga	Kigonsera	Div
3	3	5365	TZA	Tanzania	Ruvuma	Region	Mbinga	Kigonsera	Div
4	4	5366	TZA	Tanzania	Ruvuma	Region	Mbinga	Kigonsera	Div
5	5	5367	TZA	Tanzania	Ruvuma	Region	Mbinga	Kigonsera	Div
6	6	3792	TZA	Tanzania	Ruvuma	Region	Mbinga	Mbangamao	Div
7	7	3793	TZA	Tanzania	Ruvuma	Region	Mbinga	Mbangamao	Div
8	8	4054	TZA	Tanzania	Ruvuma	Region	Mbinga	Mbangamao	Div
9	9	4055	TZA	Tanzania	Ruvuma	Region	Mbinga	Mbangamao	Div
10	10	4056	TZA	Tanzania	Ruvuma	Region	Mbinga	Mbangamao	Div
11	11	4057	TZA	Tanzania	Ruvuma	Region	Mbinga	Mbangamao	Div
12	12	2553	TZA	Tanzania	Mtwara	Region	Masasi	Nanyumbu	Div
13	13	2812	TZA	Tanzania	Mtwara	Region	Masasi	Nanyumbu	Div
14	14	2814	TZA	Tanzania	Mtwara	Region	Masasi	Nanyumbu	Div

The imported data will be displayed in the panel on the left. You can view the first 1000 rows of your document as well as the total number of rows in the dataset in this case 1,538,278.

Your Turn: Using the same approach to import the remaining CSV files into R.

6.1.3 Joining Imported Data Frames

We will need to join the individual data frames to create one large data frame. We will use the PIXELID field to JOIN. We can **only** join two tables at a time.

To begin, type the following in the command line console:

```
join1 <- join(variable dataset 1,variable dataset2, match="first", by="PIXELID")
```

For example, to join the Biomes (ANTH) and soil carbon (CARB) datasets, the following command was used:

```
join1 <- join(ANTH, CARB, match="first", by="PIXELID")
```

Join1 is the data frame to be created by the join
match="first" parameters tells R to start joining in the first row of the input files
by= "PIXELID" sets PIXELID as the join field

Now join the rest of the data frames, beginning with the second data set in the list to join1 to create join2, then third to join2 to create join3, and so on.

```
            join2 <- join(join1,EDUC, match="first", by="PIXELID")
```

Name the final join data frame **via_data**

We need to clean up our workspace of unnecessary data frames. This means removing all joined files except the last, **via_data**

To remove objects from a workspace, use the **rm()** function:

```
rm(join1, join2,join3,…,joinX)
```

6.2 Data Exploration

It is general practice in statistics to employ exploratory data analysis tools to check for missing data and determine central tendency of input variables.

In this step we will use the following data exploration techniques to examine each of our variables.

We begin by displaying the list of variable names in the via_data frame using the command command:

```
            names(via_data)
```

Numerical measures including:

- mean(dataframe$field)
- median(dataframe$field)
- quartile(dataframe$field)
- percentile (dataframe$field) - percentile of fields
- sd(dataframe$field) - standard deviation,
- summary(dataframe$field) – displays summary statistics (min, 1st Qu., Mean, 3rd Qu., Max) of the field in question.

Graphical methods for exploring data include,

- hist(dataframe$field) - histograms
- (dataframe$field) - Scatter plots
- (dataframe$field) - Box plots

6.2.1 Numerical measures

1. The **mean** of an observation variable is a numerical measure of the central location of the data values. The mean is the arithmetic average of a set of values ~~that is,~~; the sum of ~~its~~ data values divided by data count. For variable like with ***n*** number of pixels, the mean is given by:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} .$$

We use the function, **mean(dataframe$field)** to calculate the mean.

```
> mean(join3$ANTH)  - computes the mean of the ANTH field in the join3 data frame
```

```
> mean(join3$DCVAR,na.rm=TRUE)
[1] 25.3434?
```

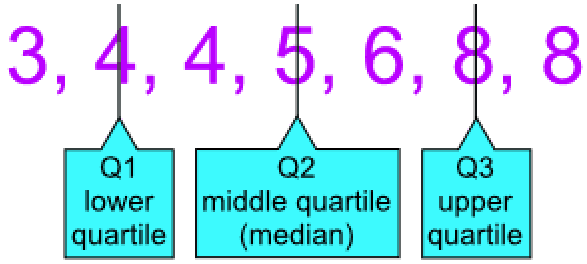
```
na.rm=TRUE  - removes missing values from the mean calculation
```

2. Median - The **median** of an observation variable is the value at the middle when the data is sorted in ascending order. It is an ordinal measure of the central location of the data values. The median of a few data points is the value of the middle observation when all the observations are lineup from lowest to highest value.

We use the function

```
> median(via_data$ANTH, na.rm=TRUE)
[1] 25.34643
```

3. Quartile – Quartiles divide data values in to 4 equal groups. The first quartile(Q1), or lower quartile is the value that cuts off 25% of the data when it is sorted in ascending order. The second quartile (Q_2) is the median of the data. The third quartile (Q_3) is the middle value between the median and the highest value of the data set, the 75%.



We apply the quantile function to compute the quartiles of eruptions.

```
> quantile(join3$DCVAR, na.rm=TRUE)
```

```
      0%       25%       50%       75%      100% 
2.378686 22.357409 25.346428 28.378949 38.683297
```

You can specify the break points using the concatenation function,

(c(0, 0.1, 0.25, 0.5, 0.75, 1.0)

```
> quantile(via_data$DCVAR, c(0, 0.1, 0.25, 0.5, 0.75, 1.0), na.rm=TRUE)

      0%       10%       25%       50%       75%      100% 
2.378686 19.316085 22.357409 25.346428 28.378949 38.683297
```

4. Percentile – The ***n**th* percentile of an observation variable is that cuts off the first ***n*** percent of the data values when it is sorted in ascending order. For instance when we sort a variable from lowest to highest observation, the 66th percentile is the that value that cuts of the first 66 percent data observation points. The **quantile()** function is used to compute the percentiles by using the concatenate, c() function.

5. The **standard deviation** (**SD**) shows how much dispersion from the average exists. It is mathematically expressed as the square root of its variance.

We apply the **sd** function to compute the standard deviation of a variable.

```
> sd(via_data$CARB, na.rm=-9999)
[1] 4807.659
```

The sd(via_data$CARB, na.rm=-9999) function above computes standard deviation of CARB variable in join3 data frame.

na.rm=-9999 Removed -9999, the no data values from the dataset in order to calculate it.If the missing datais not coded (appear as NA in the dataset), the argument is: na.rm=TRUE)

6. Summary statistics. Summary statistics are used to summarize a set of observations. The function **summary()** displays the minimum, 1st Quartile, Mean, 3rd Quartile and Max.

```
> summary(via_data$ANTH)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
11.00   41.00   43.00   47.15   63.00   63.00
```

Your turn: Now use the functions above and explore each of the variables in the via_data data frame.

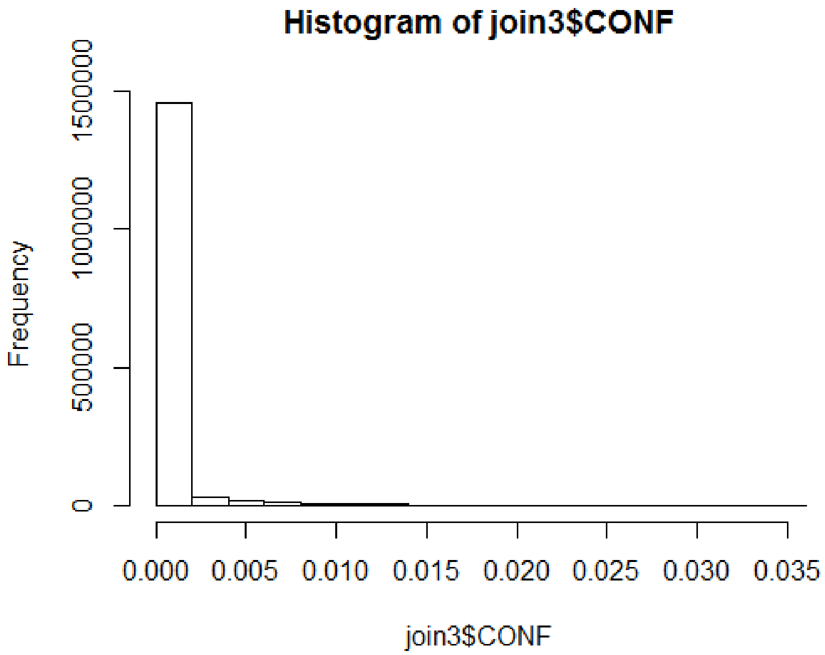
6.2.2 Graphical Methods

R statistics contain a number of tools for graphical methods for exploratory data analysis including:

- histograms
- Scatter plots
- Box plots

1. Histograms - a **histogram** consists of parallel vertical bars that graphically show the frequency distribution of a quantitative variable. A histogram is a graphical representation of the data in which the area of each bar is equal to the frequency of items found in each class.

```
> hist(via_data$CONF, main=”Conflict”, xlab=”Ratio”, ylab=”Number of grids”)
```



We can use the **ggplot()** function to make elaborate histograms with labels.

```
> plot_anth   <- ggplot( via_data, aes(x=ANTH))+geom_histogram(colour = "darkgreen", fill = "white")+scale_x_continuous("name of x axis")+scale_y_continuous("name of y axis")+ggtitle("ANTH graph")
```

plot anth – the name of your plot

ggplot() function used to declare the input data frame for a graphic and to specify the set of plot aesthetics intended to be common throughout all subsequent layers unless specifically overridden aes() aesthetics function describes how variables in the data are mapped to visual properties of geometries.

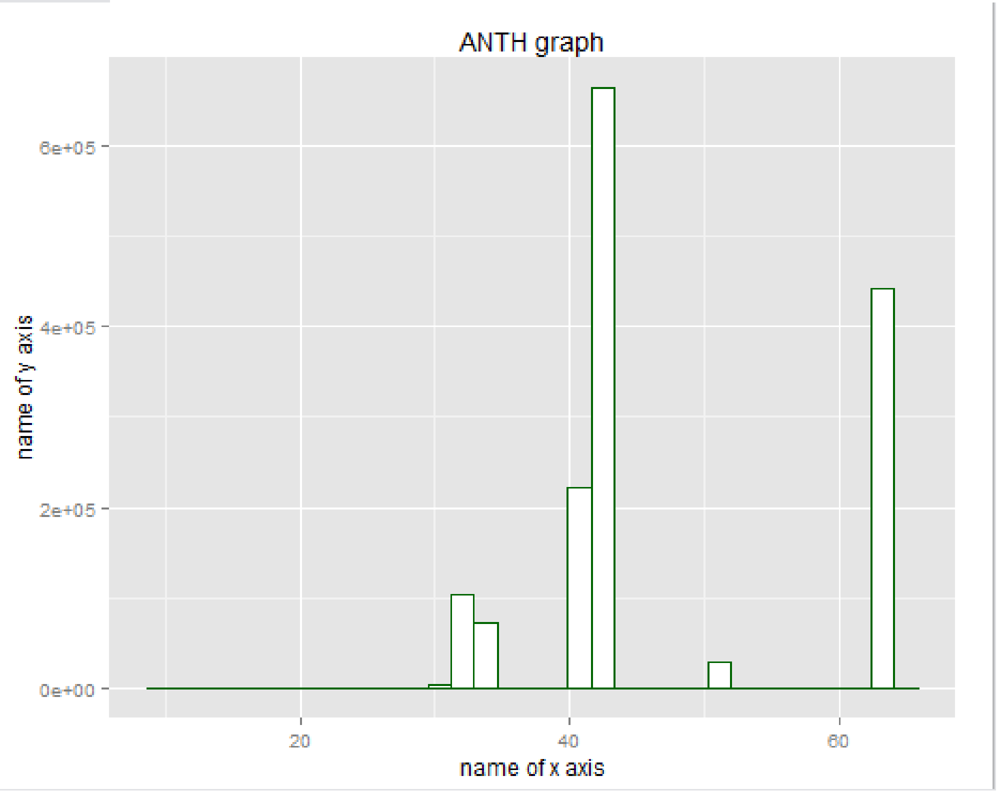
X=ANTH – states the x-axis variable

g eom_histogram() function sets line and fill color

Scale_x_continuous and scale_y_continuous sets labels for the x- and y- axis

ggtitle() sets graph title

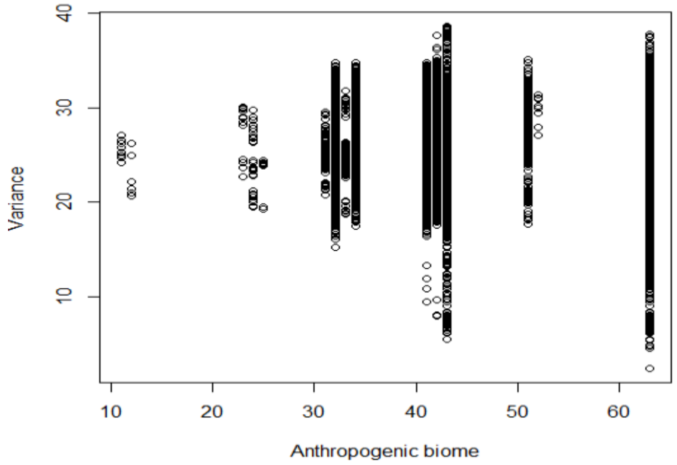
```
> plot_anth
```



2. Scatter plots . A scatter plot provides a graphical view of the relationship between two variables. The data is displayed as a collection of points, each having the value of one variable determining the position on the horizontal axis and the value of the other variable determining the position on the vertical axis.

For instance the scatter plot of ANTH and DCVAR variables is:

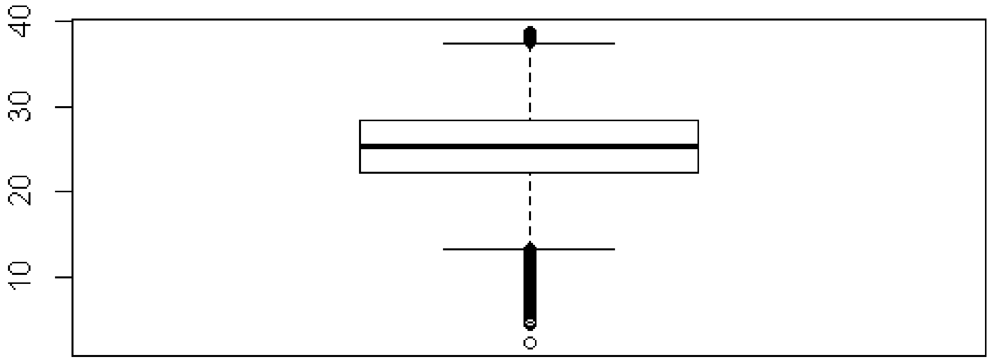
```
> plot(via_data$ANTH, via_data$DCVAR, xlab="Anthropogenic biome", ylab="Variance ")

>
```



3. Box plots - A boxplot provides a graphical view of the median, quartiles, maximum, and minimum of a data set. Box plots display differences between groups without making any assumptions of the underlying statistical distribution, and are most useful for identification of outliers.

Example below applies boxplot() function to generate the boxplot graphics

```
        > boxplot(join3$DCVAR)
```



6.3 Winsorization

Extreme values sometimes have a big effect on statistical operations. Extreme values can have positive or negative effect. Winsorization is the transformation of statistics by limiting extreme values in the statistical data to reduce the effect of possibly spurious outliers. A common method to deal with outliers is to trim or truncate the data. A 90% Winsorisation would see all data below the 5th percentile set to the 5th percentile and data above the 95th percentile set to above 95th percentile. We usually do the winsorization at one end, where we observe the presence of outliers, i.e.data is usually highly skewed to the left or to the right.

Example of Winsorization

- Order the observations by value
    - $X_{i1}, X_{i2},X_{i3}, X_{i4},X_{i5} \ldots X_{i95},X_{i96},X_{i97},X_{i98},X_{i99},X_{i100},$ where *i* denotes the i^{th} regressor
    - If Winsorizing at 5% and 95%, also called 10% Winsorization, then
        - The value for $X_{i1}, X_{i2}, X_{i3},X_{i4},$will be replaced by the value for X_{i5}
        - The value for $X_{i96}, X_{i97}, X_{i98}, X_{i99}, X_{i100},$ will be replaced by the value for X_{i95}
        - ***$X_{i5}, X_{i5},X_{i5}, X_{i5},X_{i5} \ldots X_{i95},X_{i96},X_{i96},X_{i96},X_{i96},X_{i96},$ new values***

We use the function **recode()** to set the new values below 5th and above 95th percentile to new values.

To do a 10% Winsorization we:

1. Use the > quantile(via_data$ANTH,c(0.05, 0.95)) to determine the 5th and 95th percentile values. In this case its 32 and 63

```
                    5% 95%
                   32  63
```

2. Recode all values below 5th percentile (32) in the ANTH field to 32. This creates a new field called ANTHr1

```
          >  via_data$ANTH1r <-[via_data$ANTH < 32] <-32
```

3. Recode all values above 95th percentile (63) in the ANTHr1 to 32. This creates a new field called ANTHr2

```
          > via_data$ANTH2r[a_data$ANTHr1 > 63] <-63
```

The resultant data frame contains original ANTH field and ANTHr1 field with values less than 32 (5th percentile) recoded to 32 and ANTHr2 field with values greater than 95th percentile (63) recoded to 63.

6.4 Rescale variable values to 0-100

The variables have different units of measurements, and the next step in the analysis is to bring all the data to a common 0-100 scale, where 0 equates to lower vulnerability and 100 equates to high vulnerability. For variables where *larger values represent low vulnerability* (i.e. years for education), we invert the data.

Data rescaling for variable where low values indicates low vulnerability: example= conflict

1) min_conf<-min(via_data$CONF) ~~....1~~
2) max_conf<-max(via_data$CONF) ~~......2~~
3) via_data$var_pct<-(via_data$CONF-min_conf)/(max_conf-min_conf)*100 ~~.....3~~

The command number ~~(~~1) created a new field (min_conf) and sets value to the minimum value in the CONF field.

Command ~~(~~2) creates a new field (max_conf) and computes value to maximum value in the CONF field.

Command ~~(~~3) subtract the minimum value created in ~~(~~1) from the CONF field and divide it by maximum minus minimum.

Data rescaling for variable where low values indicates high vulnerability: example= education

4) min_educ<-min(via_data$EDUC) ~~..............1~~
5) max_conf<-max(via_data$EDUC)...~~..........2~~
6) via_data$var_pct<-(via_data$EDUC-max_conf)/(min_conf-max_conf)*100~~.....3~~

6.5 Calculate composites using weighted averages, and rescale to 0-100

The input variables included in our spreadsheet are measures of three dimensions of vulnerability: adaptive capacity, exposure and sensitivity. In the following step, we will use the clean and transformed data generated in the previous steps to calculate these composites.

The calculation of the composites is based on average. If you think that some of the variables have a larger importance within the vulnerability dimension, wither because of data quality, resolution or conceptual meaning, weights are assigned to the variables to account for this.

This can be done in three steps:

1. Sum all the importance value for each of the variable. For example, if we have 5 variables, and the first 2 have double the importance of the rest of the variables, we calculate the sum of the importance as follows: 2+2+1+1+1= 7Calculate the weights:
    - for variable with 1 importance: 1/7=0.1428
    - for variables with double importance: 0.1428*2=0.2857
2. The sum of all the weights is equal to 1.
3. Calculate the weighted composite

```
via_data$adaptive<-(via_data$var1*0.2857+ via_data$var2*0.2857+
via_data$var3*0.1428+ via_data$var4*0.1428+ via_data$var5*0.1428)/5
```

The rescaling to 0-100 range is done similarly as the variable rescaling, renaming the variable adaptive_prc.

6.6 Calculate vulnerability, and rescaling to 0-100

The overall vulnerability is calculated using simple average of the three composites calculated in the previous step:

```
via_data$vulnerability <-(via_data$adaptive_prc+ via_data$exposure_prc+
via_data$sensitivity_prc)/3
```

The rescaling to 0-100 range is done similarly as the variable and composite rescaling, renaming the index vulnerability_prc.

Exercise 7. Mapping Results in ArcMap

1. In ArcCatalog import table generated from R “composites_for_mapping_table”, as a table in geodatabase.
2. In ArcMap join “composite table” by PixelID field to a copy of the “centroids” feature class.
3. If the analysis area is smaller or a subset of the “mask” then create a query by using “Select By Attributes” where PixelID IS NOT NULL.
4. Export new point feature class with the attributes from composite table “composite_table_points”
5. Rasters were created for each component and each PCA in the “composite_table_point” using ArcGIS Point to Raster tool (ArcToolbox\Conversion Tools\Raster\Point To Raster)