Open Source, Scale and Reproducibility Using GIS: Discovering the World Beyond Point-and-Click and ArcGIS

Ömer Özak

Dept. Economics, SMU

Here's How I Do GIS

April 28, 2021

The Voyage of Homo-œconomicus into GIS

Ömer Özak

Dept. Economics, SMU

Here's How I Do GIS

April 28, 2021

Homo-œconomicus...or Economic Man

" the concept in many economic theories portraying humans as consistently rational and narrowly self-interested agents who usually pursue their subjectively-defined ends optimally."

Plan for today

1. The Big Bang
2. The Dark Ages
3. The Age of Discovery
4. The Modern Era
5. The Future
6. Q&A

Why GIS?

The question that started it all

Why GIS?

The question that started it all

- What is the effect of geographical isolation on economic development?

Why GIS?

The question that started it all

- What is the effect of geographical isolation on economic development?
    - How to measure?

Why GIS?

The question that started it all

- What is the effect of geographical isolation on economic development?
    - How to measure?
    - Measure for Pre-industrial era

Why GIS?

The question that started it all

- What is the effect of geographical isolation on economic development?
  - How to measure?
  - Measure for Pre-industrial era
  - Changes due to technology

Why GIS?

The question that started it all

- What is the effect of geographical isolation on economic development?
  - How to measure?
  - Measure for Pre-industrial era
  - Changes due to technology
- Need geographically organized data

  $\Longrightarrow$ Geographic Information Systems (GIS)

Why a New Measure?

Why a New Measure?

- Common approach: Geodesic distances

Why a New Measure?

- Common approach: Geodesic distances
    - As the crow flies

Why a New Measure?

- Common approach: Geodesic distances
  - As the crow flies
  - Assumes flying technology or flat world

Why a New Measure?

- Common approach: Geodesic distances
  - As the crow flies
  - Assumes flying technology or flat world
  - Measurement error

Why a New Measure?

- Common approach: Geodesic distances
  - As the crow flies
  - Assumes flying technology or flat world
  - Measurement error

Country 1	Country 2	Distance	Country 1	Country 2	Distance
Costa Rica	Panama	514.3561	Germany	Poland	515.774
Phillipines	Brunei	1262.339	Yemen	Sudan	1254.947
Irak	Romania	2002.218	Ghana	Gambia	2002.745

Proposed Solution

Construct a measure that

Proposed Solution

Construct a measure that

- Controls for

Proposed Solution

Construct a measure that

- Controls for
    - Human biological constraints

Proposed Solution

Construct a measure that

- Controls for
  - Human biological constraints
  - Geographical conditions

Proposed Solution

Construct a measure that

- Controls for
    - Human biological constraints
    - Geographical conditions
    - Technological conditions

Proposed Solution

Construct a measure that

- Controls for
    - Human biological constraints
    - Geographical conditions
    - Technological conditions
- Has meaning

Solution: The Voyage of Homo-œconomicus

Combine data on

Solution: The Voyage of Homo-œconomicus

Combine data on

- Infantry movement

Solution: The Voyage of Homo-œconomicus

Combine data on

- Infantry movement
- Geographical conditions

Solution: The Voyage of Homo-œconomicus

Combine data on

- Infantry movement
- Geographical conditions
- Ship speeds in different eras

Solution: The Voyage of Homo-œconomicus

Combine data on

- Infantry movement
- Geographical conditions
- Ship speeds in different eras

$\Longrightarrow$

Solution: The Voyage of Homo-œconomicus

Combine data on

- Infantry movement
- Geographical conditions
- Ship speeds in different eras

$\Longrightarrow$

- Human Mobility Index (HMI)

Solution: The Voyage of Homo-œconomicus

Combine data on

- Infantry movement
- Geographical conditions
- Ship speeds in different eras

$\Longrightarrow$

- Human Mobility Index (HMI)
- HMI with Seafaring pre-1500CE (HMISea)

Solution: The Voyage of Homo-œconomicus

Combine data on

- Infantry movement
- Geographical conditions
- Ship speeds in different eras

$\Longrightarrow$

- Human Mobility Index (HMI)
- HMI with Seafaring pre-1500CE (HMISea)
- HMI with Seafaring pre-steam engine (HMIOcean)

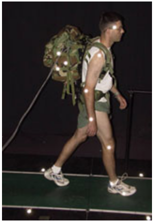
How?

Construct cost of movement using data from

How?

Construct cost of movement using data from

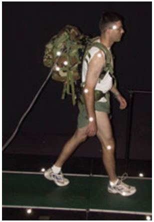
- U.S. Army data on human mobility (Hayes, 1994)



How?

Construct cost of movement using data from

- U.S. Army data on human mobility (Hayes, 1994)

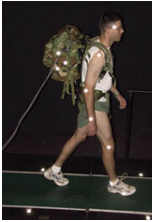


- Travel Time on Land $= f(slope, temp, rel.\ hum., terrain, sky)$

How?

Construct cost of movement using data from

- U.S. Army data on human mobility (Hayes, 1994)

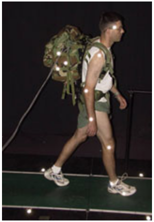


  - Travel Time on Land $= f(slope, temp, rel.\ hum., terrain, sky)$
- Historical data on seafaring in Old World (pre-1500CE) (Casson, 1951, 1989)

How?

Construct cost of movement using data from

- U.S. Army data on human mobility (Hayes, 1994)



  - Travel Time on Land $= f(slope, temp, rel.\ hum., terrain, sky)$
- Historical data on seafaring in Old World (pre-1500CE) (Casson, 1951, 1989)
- Historical data on seafaring (pre-steam engine) (García-Herrera, Können, Wheeler, Prieto, Jones, and Koek, 2005)

Optimal Paths

- With cost surface find minimum travel time between locations

Optimal Paths

- With cost surface find minimum travel time between locations
- Concerns

Optimal Paths

- With cost surface find minimum travel time between locations
- Concerns
  - Raster (matrix of locations) size (12,837; 43,345), i.e. 556,419,765 cells!

Optimal Paths

- With cost surface find minimum travel time between locations
- Concerns
  - Raster (matrix of locations) size (12,837; 43,345), i.e. 556,419,765 cells!

$\Longrightarrow$ Potentially huge graph

Optimal Paths

- With cost surface find minimum travel time between locations
- Concerns
  - Raster (matrix of locations) size (12,837; 43,345), i.e. 556,419,765 cells!

  ⟹ Potentially huge graph
  - 200+ countries

Homo-œconomicus meets GIS

Start with traditional approach...

Homo-œconomicus meets GIS

Start with traditional approach...

- Go to GIS course/workshop (1 or 2 weeks)

Homo-œconomicus meets GIS

Start with traditional approach...

- Go to GIS course/workshop (1 or 2 weeks)
    - Intro to ArcGIS (point-and-click)

Homo-œconomicus meets GIS

Start with traditional approach...

- Go to GIS course/workshop (1 or 2 weeks)
  - Intro to ArcGIS (point-and-click)
  - Get data from TIGER or ArcGIS

Homo-œconomicus meets GIS

Start with traditional approach...

- Go to GIS course/workshop (1 or 2 weeks)
    - Intro to ArcGIS (point-and-click)
    - Get data from TIGER or ArcGIS
    - Make maps using shapefiles in ArcGIS

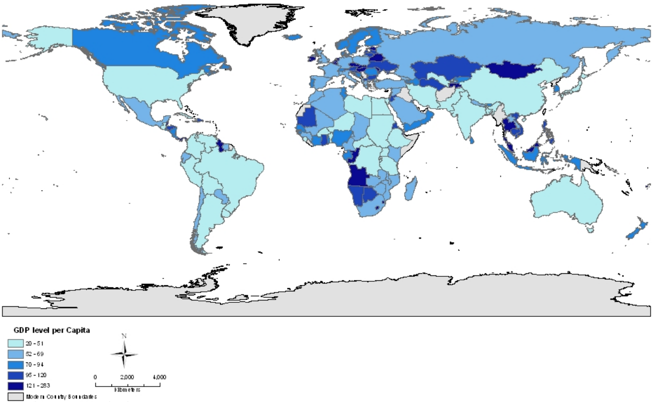
Homo-œconomicus meets GIS

Start with traditional approach...

- Go to GIS course/workshop (1 or 2 weeks)
    - Intro to ArcGIS (point-and-click)
    - Get data from TIGER or ArcGIS
    - Make maps using shapefiles in ArcGIS
    - Some spatial stats (compute Moran I & II)

My First Map (True!) – Trade as share of GDP

Trade as Percentage of GDP in the World in year 2000



Source : Data Source : World Development Indicators 2005.
The World Bank.

copyright Ömer Özak 2006

Initiative for Spatial Literacy - GIS@SMU

- Resources for learning GIS at SMU
  - Data
  - Courses
  - Workshops & Events
  - Email list gis_support@smu.edu
  - Super friendly & dedicated library staff and faculty
    - Sylvia Jones, Science Research Librarian
    - Jessie Zarazaga, Project Director, Initiative for Spatial Literac
  - https://www.smu.edu/libraries/fondren/services/gis

Homo-œconomicus meets GIS

Seems easy & straightforward...Thesis's gonna be ready in 2 weeks!

What do you mean the data is not in a shapefile?

- Search for data to construct cost surface

What do you mean the data is not in a shapefile?

- Search for data to construct cost surface
    - Where do I search for data?

What do you mean the data is not in a shapefile?

- Search for data to construct cost surface
    - Where do I search for data?
    - No shapefiles...what?!

What do you mean the data is not in a shapefile?

- Search for data to construct cost surface
    - Where do I search for data?
    - No shapefiles...what?!
    - What is a raster?!

What do you mean the data is not in a shapefile?

- Search for data to construct cost surface
    - Where do I search for data?
    - No shapefiles...what?!
    - What is a raster?!
- Raster = Matrix

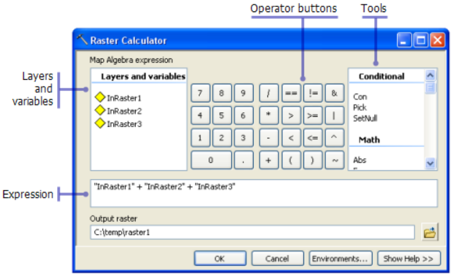
What do you mean the data is not in a shapefile?

- Search for data to construct cost surface
    - Where do I search for data?
    - No shapefiles...what?!
    - What is a raster?!
- Raster = Matrix
- Shapefile $\Longrightarrow$ points, lines, polygons, etc.

Point-and-click and more in ArcGIS

To solve my problem I need more tools...

- Raster Calculator



Map Algebra expression

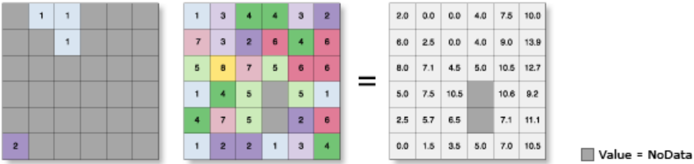
"InRaster1" + "InRaster2" + "InRaster3"

Output raster

C:\temp\raster1

Point-and-click and more in ArcGIS

- Cost Distance Function



	1	1			
		1			
2					

Source_Ras

1	3	4	4	3	2
7	3	2	6	4	6
5	8	7	5	6	6
1	4	5		5	1
4	7	5		2	6
1	2	2	1	3	4

Cost_Ras

=

2.0	0.0	0.0	4.0	7.5	10.0
6.0	2.5	0.0	4.0	9.0	13.9
8.0	7.1	4.5	5.0	10.5	12.7
5.0	7.5	10.5		10.6	9.2
2.5	5.7	6.5		7.1	11.1
0.0	1.5	3.5	5.0	7.0	10.5

Cost_Dist

▨ Value = NoData

ArcGIS headache I

- Read manual and figured out how to do it...at least in theory...

ArcGIS headache I

- Read manual and figured out how to do it...at least in theory...
- Problem...does not work on my laptop...why?!

ArcGIS headache I

- Read manual and figured out how to do it...at least in theory...
- Problem...does not work on my laptop...why?!
    - I have a Mac...need virtual machine

ArcGIS headache I

- Read manual and figured out how to do it...at least in theory...
- Problem...does not work on my laptop...why?!
    - I have a Mac...need virtual machine
    - I have student edition

ArcGIS headache I

- Read manual and figured out how to do it...at least in theory...
- Problem...does not work on my laptop...why?!
  - I have a Mac...need virtual machine
  - I have student edition
  - **Need Spatial Analyst License = $2,500!!!**

ArcGIS headache I

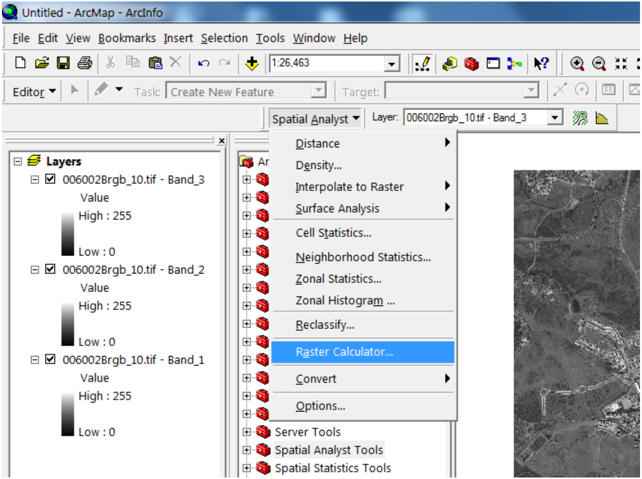
- Read manual and figured out how to do it...at least in theory...
- Problem...does not work on my laptop...why?!
  - I have a Mac...need virtual machine
  - I have student edition
  - **Need Spatial Analyst License = $2,500!!!**
- Ok...work in university computers...GIS lab

ArcGIS headache I

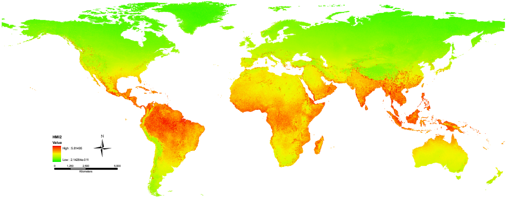
- Read manual and figured out how to do it...at least in theory...
- Problem...does not work on my laptop...why?!
  - I have a Mac...need virtual machine
  - I have student edition
  - **Need Spatial Analyst License = $2,500!!!**
- Ok...work in university computers...GIS lab
  - Get access to 1 computer...

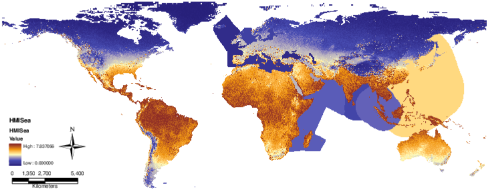
Produce Raster in ArcGIS

- Construct HMI data



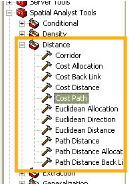
HMI & HMISea

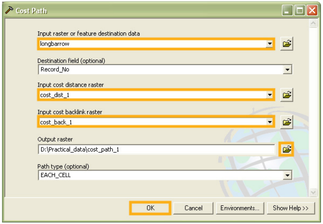

HMI2
Value
High : 5.81495
Low : 2.14284e-011
N
0
1,250
2,500
5,000
Kilometers


HMISea
HMISea
Value
High : 7.837056
Low : 0.000000
N
0
1,350
2,700
5,400
Kilometers

ArcGIS headache II

- Construct Optimal Routes and Times





Cost Path

Input raster or feature destination data: longbarrow

Destination field (optional): Record_No

Input cost distance raster: cost_dist_1

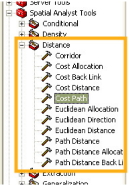
Input cost backlink raster: cost_back_1

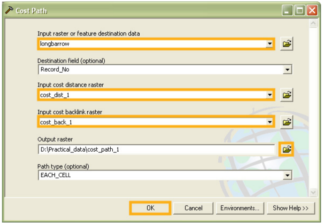
Output raster: D:\Practical_data\cost_path_1

Path type (optional): EACH_CELL

ArcGIS headache II

- Construct Optimal Routes and Times




Cost Path

Input raster or feature destination data: longbarrow

Destination field (optional): Record_No

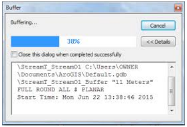
Input cost distance raster: cost_dist_1

Input cost backlink raster: cost_back_1

Output raster: D:\Practical_data\cost_path_1

Path type (optional): EACH_CELL

- Problem...it takes more than 1 day per source!


Buffer

Buffering...

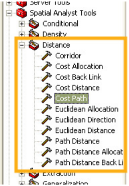
38%

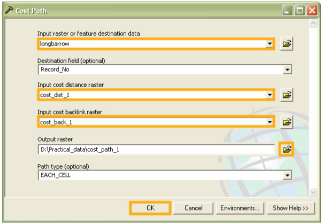
□ Close this dialog when completed successfully

```
\StreamT_Stream01 C:\Users\OWNER
\Documents\ArcGIS\Default.gdb
\StreamT_Stream01_Buffer "11 Meters"
FULL ROUND ALL # PLANAR
Start Time: Mon Jun 22 13:38:46 2015
```

ArcGIS headache II

- Construct Optimal Routes and Times




Cost Path

Input raster or feature destination data: longbarrow

Destination field (optional): Record_No

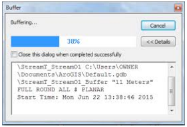
Input cost distance raster: cost_dist_1

Input cost backlink raster: cost_back_1

Output raster: D:\Practical_data\cost_path_1

Path type (optional): EACH_CELL

- Problem...it takes more than 1 day per source!


Buffer

Buffering...

38%

□ Close this dialog when completed successfully

```
\StreamT_Stream01 C:\Users\OWNER
\Documents\ArcGIS\Default.gdb
\StreamT_Stream01_Buffer "11 Meters"
FULL ROUND ALL # PLANAR
Start Time: Mon Jun 22 13:38:46 2015
```

$\Longrightarrow$ $>$ 1 year to compute data!!!

Solution…Parallelize!

- Find multiple computers and **repeat exact same** process for different sources

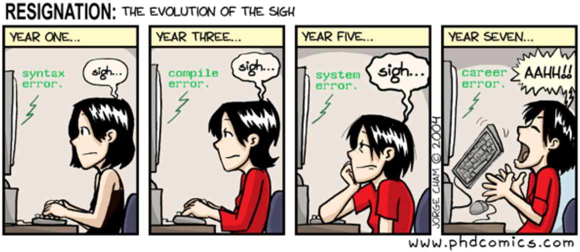
Solution...Parallelize!

- Find multiple computers and **repeat exact same** process for different sources

- Need scripting...still slow in ArcGIS!

OMG...Now what???!!!

HOW'S THE DEBUGGING GOING?
JORGE CHAM © 2005
www.phdcomics.com


RESIGNATION: THE EVOLUTION OF THE SIGH
YEAR ONE...
syntax error.
sigh...
YEAR THREE...
compile error.
sigh...
YEAR FIVE...
system error.
sigh...
YEAR SEVEN...
career error.
AAHH!!
JORGE CHAM © 2004
www.phdcomics.com

Main take aways

ArcGIS and point-and-click

- Advantages

Main take aways

ArcGIS and point-and-click

- Advantages
    - Easy to install and start using for basic things

Main take aways

ArcGIS and point-and-click

- Advantages
  - Easy to install and start using for basic things
  - Good for mapping

Main take aways

ArcGIS and point-and-click

- Advantages
  - Easy to install and start using for basic things
  - Good for mapping
- Disadvantages

Main take aways

ArcGIS and point-and-click

- Advantages
    - Easy to install and start using for basic things
    - Good for mapping
- Disadvantages
    - Expensive

Main take aways

ArcGIS and point-and-click

- Advantages
  - Easy to install and start using for basic things
  - Good for mapping
- Disadvantages
  - Expensive
  - Computations are slow...very slow

Main take aways

ArcGIS and point-and-click

- Advantages
  - Easy to install and start using for basic things
  - Good for mapping
- Disadvantages
  - Expensive
  - Computations are slow...very slow
  - Support is slow

Main take aways

ArcGIS and point-and-click

- Advantages
    - Easy to install and start using for basic things
    - Good for mapping
- Disadvantages
    - Expensive
    - Computations are slow...very slow
    - Support is slow
    - Difficult replication

      (very important! AER requires replication code for maps!)

Main take aways

ArcGIS and point-and-click

- Advantages
  - Easy to install and start using for basic things
  - Good for mapping
- Disadvantages
  - Expensive
  - Computations are slow...very slow
  - Support is slow
  - Difficult replication

    (very important! AER requires replication code for maps!)
  - Not scripting friendly

Main take aways

ArcGIS and point-and-click

- Advantages
  - Easy to install and start using for basic things
  - Good for mapping
- Disadvantages
  - Expensive
  - Computations are slow...very slow
  - Support is slow
  - Difficult replication

    (very important! AER requires replication code for maps!)
  - Not scripting friendly
  - Only Windows compatible

Arrival at SMU

- Although many disadvantages to ArcGIS, it is costly to change

Arrival at SMU

- Although many disadvantages to ArcGIS, it is costly to change
- Luckily for me when I moved to SMU I got a push

Arrival at SMU

- Although many disadvantages to ArcGIS, it is costly to change
- Luckily for me when I moved to SMU I got a push
    - Couldn't get ArcGIS installed

Arrival at SMU

- Although many disadvantages to ArcGIS, it is costly to change
- Luckily for me when I moved to SMU I got a push
  - Couldn't get ArcGIS installed
  - My computer is a Mac...and you know,
    once you go Mac, you don't go back!

Arrival at SMU

- Although many disadvantages to ArcGIS, it is costly to change
- Luckily for me when I moved to SMU I got a push
  - Couldn't get ArcGIS installed
  - My computer is a Mac...and you know,
    once you go Mac, you don't go back!
  - Virtual Machine had problems with ArcGIS

Arrival at SMU

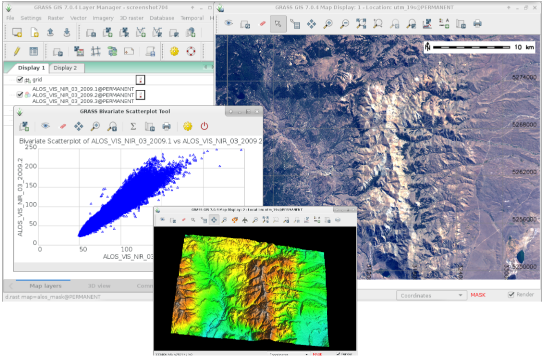
- Although many disadvantages to ArcGIS, it is costly to change
- Luckily for me when I moved to SMU I got a push
  - Couldn't get ArcGIS installed
  - My computer is a Mac...and you know,
    once you go Mac, you don't go back!
  - Virtual Machine had problems with ArcGIS

⟹ Time to try something different

Can I overcome disadvantages?

Free Point-and-click solutions

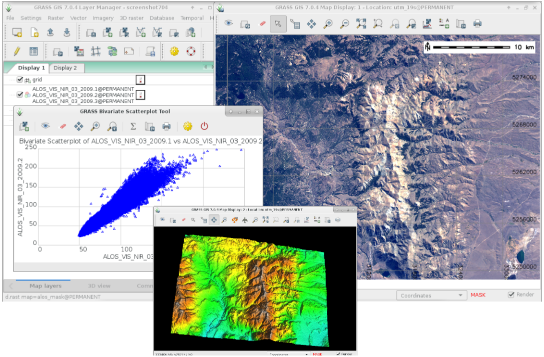
- GRASS (Geographic Resources Analysis Support System):


GRASS GIS 7.0.4 Layer Manager - screenshot704
File Settings Raster Vector Imagery 3D raster Database Temporal
Display 1
Display 2
grid
ALOS_VIS_NIR_03_2009.1@PERMANENT
ALOS_VIS_NIR_03_2009.2@PERMANENT
ALOS_VIS_NIR_03_2009.3@PERMANENT
GRASS Bivariate Scatterplot Tool
Bivariate Scatterplot of ALOS_VIS_NIR_03_2009.1 vs ALOS_VIS_NIR_03_2009.2
ALOS_VIS_NIR_03_2009.2
ALOS_VIS_NIR_03
250
200
150
100
50
0
0
50
100
Map layers
3D view
d.rast map=alos_mask@PERMANENT
GRASS GIS 7.0.4 Map Display: 1 - Location: utm_19s@PERMANENT
10 km
5274000
5268000
5262000
5256000
5250000
Coordinates
MASK
Render
GRASS GIS 7.0.4 Map Display: 2 - Location: utm_19s@PERMANENT

Can I overcome disadvantages?

Free Point-and-click solutions

- GRASS (Geographic Resources Analysis Support System):

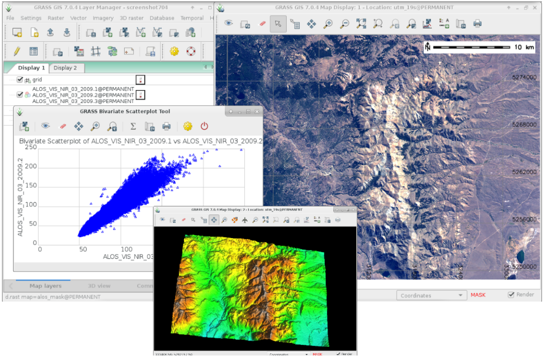

GRASS GIS 7.0.4 Layer Manager - screenshot704
File Settings Raster Vector Imagery 3D raster Database Temporal
Display 1
Display 2
☑ grid
☑ ALOS_VIS_NIR_03_2009.1@PERMANENT
ALOS_VIS_NIR_03_2009.2@PERMANENT
ALOS_VIS_NIR_03_2009.3@PERMANENT
GRASS Bivariate Scatterplot Tool
Bivariate Scatterplot of ALOS_VIS_NIR_03_2009.1 vs ALOS_VIS_NIR_03_2009.2
250
200
150
100
50
0
ALOS_VIS_NIR_03_2009.2
0
50
100
ALOS_VIS_NIR_03
Map layers
3D view
d.rast map=alos_mask@PERMANENT
GRASS GIS 7.0.4 Map Display: 1 - Location: utm_19s@PERMANENT
10 km
5274000
5268000
5262000
5256000
5250000
Coordinates
MASK
☑ Render
GRASS GIS 7.0.4 Map Display: 2 - Location: utm_19s@PERMANENT

- Originally developed by the U.S. Army Construction Engineering Research Laboratories

Can I overcome disadvantages?

Free Point-and-click solutions

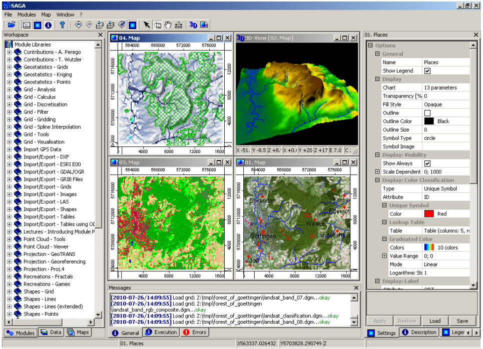
- GRASS (Geographic Resources Analysis Support System):


GRASS GIS 7.0.4 Layer Manager - screenshot704
File Settings Raster Vector Imagery 3D raster Database Temporal H
Display 1
Display 2
grid
ALOS_VIS_NIR_03_2009.1@PERMANENT
ALOS_VIS_NIR_03_2009.2@PERMANENT
ALOS_VIS_NIR_03_2009.3@PERMANENT
GRASS Bivariate Scatterplot Tool
Bivariate Scatterplot of ALOS_VIS_NIR_03_2009.1 vs ALOS_VIS_NIR_03_2009.2
250
200
150
100
50
0
ALOS_VIS_NIR_03_2009.2
0
50
100
ALOS_VIS_NIR_03
Map layers
3D view
Comm
d.rast map=alos_mask@PERMANENT
GRASS GIS 7.0.4 Map Display: 1 - Location: utm_19s@PERMANENT
10 km
5274000
5268000
5262000
5256000
5250000
3120
3180
3240
3300
Coordinates
MASK
☑ Render
GRASS GIS 7.0.4 Map Display: 2 - Location: utm_19s@PERMANENT
Coordinates
MASK
☑ Render

  - Originally developed by the U.S. Army Construction Engineering Research Laboratories
  - Now part of OSGEO (Open Source Geospatial Foundation)

Free Software/Open Source

- SAGA (System for Automated Geoscientific Analyses):



```
[2010-07-26/14:09:55] Load grid: Z:\tmp\forest_of_goettingen\landsat_band_07.dgm...okay
[2010-07-26/14:09:55] Load grid: Z:\tmp\forest_of_goettingen
\landsat_band_rgb_composite.dgm...okay
[2010-07-26/14:09:55] Load grid: Z:\tmp\forest_of_goettingen\landsat_classification.dgm...okay
[2010-07-26/14:09:55] Load grid: Z:\tmp\forest_of_goettingen\landsat_band_08.dgm...okay
```

01. Places X563337.026432 Y5703828.290749 Z

Options

General

Name	Places
Show Legend	☑

Display

Chart	13 parameters
Transparency [%	0
Fill Style	Opaque
Outline	□
Outline Color	■ Black
Outline Size	0
Symbol Type	circle
Symbol Image	

Display: Visibility

Show Always	☑
Scale Dependent	0; 1000

Display: Color Classification

Type	Unique Symbol
Attribute	ID

Unique Symbol

Color	🟥 Red

Lookup Table

Table	Table (columns: 5, r

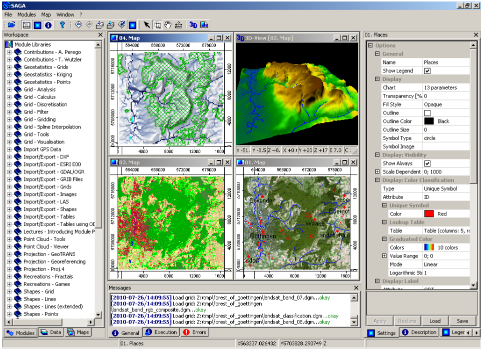
Graduated Color

Colors	10 colors
Value Range	0; 0
Mode	Linear
Logarithmic St	1

Display: Label

Free Software/Open Source

- SAGA (System for Automated Geoscientific Analyses):



```
[2010-07-26/14:09:55] Load grid: Z:\tmp\forest_of_goettingen\landsat_band_07.dgm...okay
[2010-07-26/14:09:55] Load grid: Z:\tmp\forest_of_goettingen
\landsat_band_rgb_composite.dgm...okay
[2010-07-26/14:09:55] Load grid: Z:\tmp\forest_of_goettingen\landsat_classification.dgm...okay
[2010-07-26/14:09:55] Load grid: Z:\tmp\forest_of_goettingen\landsat_band_08.dgm...okay
```

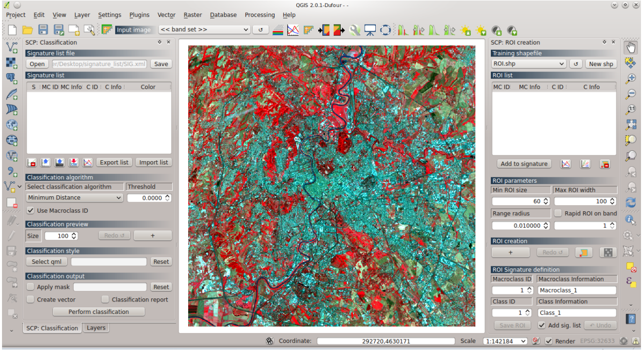
01. Places X563337.026432 Y5703828.290749 Z

Options	
General	
Name	Places
Show Legend	☑
Display	
Chart	13 parameters
Transparency [%	0
Fill Style	Opaque
Outline	□
Outline Color	Black
Outline Size	0
Symbol Type	circle
Symbol Image	
Display: Visibility	
Show Always	☑
Scale Dependent	0; 1000
Display: Color Classification	
Type	Unique Symbol
Attribute	ID
Unique Symbol	
Color	Red
Lookup Table	
Table	Table (columns: 5, rc
Graduated Color	
Colors	10 colors
Value Range	0; 0
Mode	Linear
Logarithmic Stı	1
Display: Label	

- Originally developed by Dept. of Physical Geography, Göttingen and Hamburg

What I use...QGIS!

- QGIS (Quantum GIS):



<< band set >>

er/Desktop/signature_list/SIG.xml

Select classification algorithm: Minimum Distance

Threshold: 0.0000

☑ Use Macroclass ID

Size: 100

☐ Apply mask

☐ Create vector

☐ Classification report

Training shapefile: ROI.shp

Min ROI size: 60

Max ROI width: 100

Range radius: 0.010000

☐ Rapid ROI on band: 1

Macroclass ID: 1

Macroclass Information: Macroclass_1

Class ID: 1

Class Information: Class_1

☑ Add sig. list

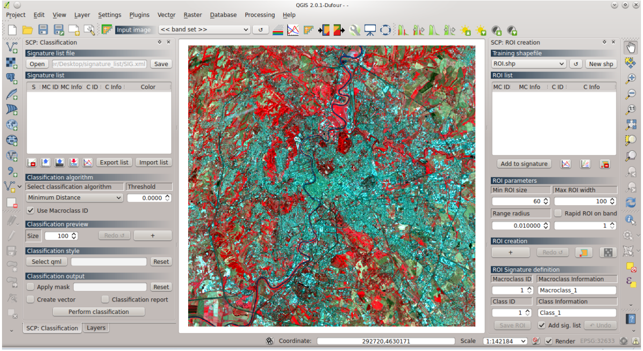
Coordinate: 292720,4630171

Scale: 1:142184

☑ Render

What I use...QGIS!

- QGIS (Quantum GIS):



er/Desktop/signature_list/SIG.xml

Select classification algorithm: Minimum Distance

Threshold: 0.0000

☑ Use Macroclass ID

Size: 100

□ Apply mask

□ Create vector

□ Classification report

Training shapefile: ROI.shp

Min ROI size: 60

Max ROI width: 100

Range radius: 0.010000

□ Rapid ROI on band: 1

Macroclass ID: 1

Macroclass Information: Macroclass_1

Class ID: 1

Class Information: Class_1

☑ Add sig. list

Coordinate: 292720,4630171

Scale: 1:142184

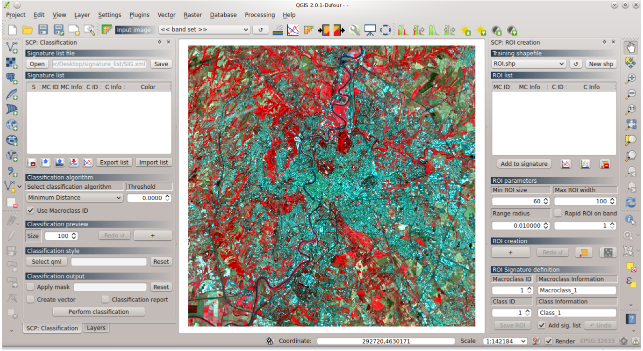
☑ Render

EPSG:32633

- Part of OSGEO (Open Source Geospatial Foundation)

What I use...QGIS!

- QGIS (Quantum GIS):



Signature list file: er/Desktop/signature_list/SIG.xml

Signature list

Classification algorithm

Select classification algorithm: Minimum Distance

Threshold: 0.0000

☑ Use Macroclass ID

Classification preview

Size: 100

Classification style

Classification output

□ Apply mask

□ Create vector

□ Classification report

Training shapefile: ROI.shp

ROI list

ROI parameters

Min ROI size: 60

Max ROI width: 100

Range radius: 0.010000

□ Rapid ROI on band: 1

ROI creation

ROI Signature definition

Macroclass ID: 1

Macroclass Information: Macroclass_1

Class ID: 1

Class Information: Class_1

☑ Add sig. list

Coordinate: 292720,4630171

Scale: 1:142184

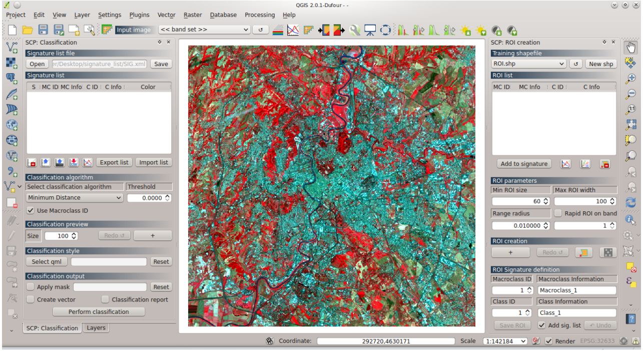
☑ Render

EPSG:32633

- Part of OSGEO (Open Source Geospatial Foundation)
- Can integrate GRASS & SAGA!

What I use...QGIS!

- QGIS (Quantum GIS):



Signature list file: er/Desktop/signature_list/SIG.xml

Select classification algorithm: Minimum Distance
Threshold: 0.0000
☑ Use Macroclass ID

Classification preview: Size 100

☐ Apply mask
☐ Create vector
☐ Classification report

Training shapefile: ROI.shp

Min ROI size: 60
Max ROI width: 100
Range radius: 0.010000
☐ Rapid ROI on band: 1

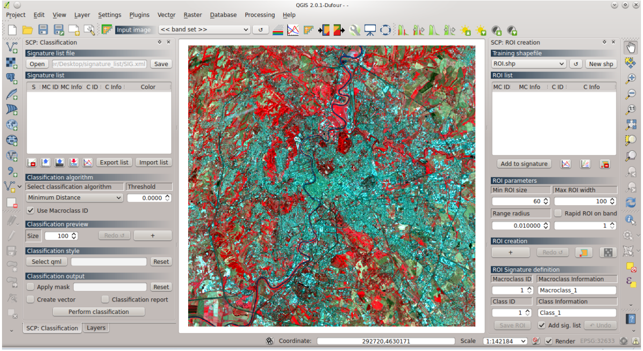
Macroclass ID: 1
Macroclass Information: Macroclass_1
Class ID: 1
Class Information: Class_1
☑ Add sig. list

Coordinate: 292720,4630171
Scale 1:142184
☑ Render EPSG:32633

- Part of OSGEO (Open Source Geospatial Foundation)
- Can integrate GRASS & SAGA!
- Large user group, plug-ins, programmable

What I use...QGIS!

- QGIS (Quantum GIS):



Signature list file: er/Desktop/signature_list/SIG.xml

Select classification algorithm: Minimum Distance

Threshold: 0.0000

☑ Use Macroclass ID

Size: 100

☐ Apply mask

☐ Create vector

☐ Classification report

Training shapefile: ROI.shp

Min ROI size: 60

Max ROI width: 100

Range radius: 0.010000

☐ Rapid ROI on band: 1

Macroclass ID: 1

Macroclass Information: Macroclass_1

Class ID: 1

Class Information: Class_1

☑ Add sig. list

Coordinate: 292720,4630171

Scale: 1:142184

☑ Render

- Part of OSGEO (Open Source Geospatial Foundation)
- Can integrate GRASS & SAGA!
- Large user group, plug-ins, programmable
- Ömer's Basic QGIS Tutorial

What this solves

- Free as in

What this solves

- Free as in
  - beer (no cost)

What this solves

- Free as in
  - beer (no cost)
  - freedom (users have the freedom to run, copy, distribute, study, change and improve the software)

What this solves

- Free as in
  - beer (no cost)
  - freedom (users have the freedom to run, copy, distribute, study, change and improve the software)
- Works many OS: Android, Linux, OSX, Windows

What this solves

- Free as in
  - beer (no cost)
  - freedom (users have the freedom to run, copy, distribute, study, change and improve the software)
- Works many OS: Android, Linux, OSX, Windows
- Open source $\Longrightarrow$

What this solves

- Free as in
  - beer (no cost)
  - freedom (users have the freedom to run, copy, distribute, study, change and improve the software)
- Works many OS: Android, Linux, OSX, Windows
- Open source $\Longrightarrow$
  - Large community/support

What this solves

- Free as in
  - beer (no cost)
  - freedom (users have the freedom to run, copy, distribute, study, change and improve the software)
- Works many OS: Android, Linux, OSX, Windows
- Open source $\Longrightarrow$
  - Large community/support
  - Change code

What this solves

- Free as in
  - beer (no cost)
  - freedom (users have the freedom to run, copy, distribute, study, change and improve the software)
- Works many OS: Android, Linux, OSX, Windows
- Open source $\Longrightarrow$
  - Large community/support
  - Change code
  - Propose plug-ins, features, etc.

What this solves

- Free as in
    - beer (no cost)
    - freedom (users have the freedom to run, copy, distribute, study, change and improve the software)
- Works many OS: Android, Linux, OSX, Windows
- Open source $\Longrightarrow$
    - Large community/support
    - Change code
    - Propose plug-ins, features, etc.
- But, not parallelizable...

How I Learned to Stop Clicking and Love the Code

Finally started using

python™

IP[y]: IPython
Interactive Computing

jupyter

python™

- General Purpose Programming Language

python™

- General Purpose Programming Language
    - Open source

python™

- General Purpose Programming Language
  - Open source
  - Easy to learn and code

python™

- General Purpose Programming Language
  - Open source
  - Easy to learn and code
  - Clean code

python™

- General Purpose Programming Language
  - Open source
  - Easy to learn and code
  - Clean code
  - Powerful

python™

- General Purpose Programming Language
  - Open source
  - Easy to learn and code
  - Clean code
  - Powerful
  - Versatile

python™

- General Purpose Programming Language
  - Open source
  - Easy to learn and code
  - Clean code
  - Powerful
  - Versatile
  - Lots of packages to get things done

- General Purpose Programming Language
  - Open source
  - Easy to learn and code
  - Clean code
  - Powerful
  - Versatile
  - Lots of packages to get things done
  - Large community (Stackoverflow, Github, Bitbucket)

python™

- General Purpose Programming Language
  - Open source
  - Easy to learn and code
  - Clean code
  - Powerful
  - Versatile
  - Lots of packages to get things done
  - Large community (Stackoverflow, Github, Bitbucket)
  - Used in ArcGIS, QGIS, Google, Yahoo!, LANL, Netflix, National Weather Service, NASA, etc.

IP[y]: IPython Interactive Computing

- Interactive Python

IP[y]: IPython Interactive Computing

- Interactive Python
  - GUI/Kernel for Python/Jupyter

IP[y]: IPython Interactive Computing

- Interactive Python
  - GUI/Kernel for Python/Jupyter
  - Easy to use

IP[y]: IPython Interactive Computing

- Interactive Python
  - GUI/Kernel for Python/Jupyter
  - Easy to use
  - Visually appealing

IP[y]: IPython Interactive Computing

- Interactive Python
  - GUI/Kernel for Python/Jupyter
  - Easy to use
  - Visually appealing
  - Interactive data visualization

IP[y]: IPython
Interactive Computing

- Interactive Python
  - GUI/Kernel for Python/Jupyter
  - Easy to use
  - Visually appealing
  - Interactive data visualization
  - Debugging

IP[y]: IPython Interactive Computing

- Interactive Python
  - GUI/Kernel for Python/Jupyter
  - Easy to use
  - Visually appealing
  - Interactive data visualization
  - Debugging
  - Tab completion

IP[y]: IPython

Interactive Computing

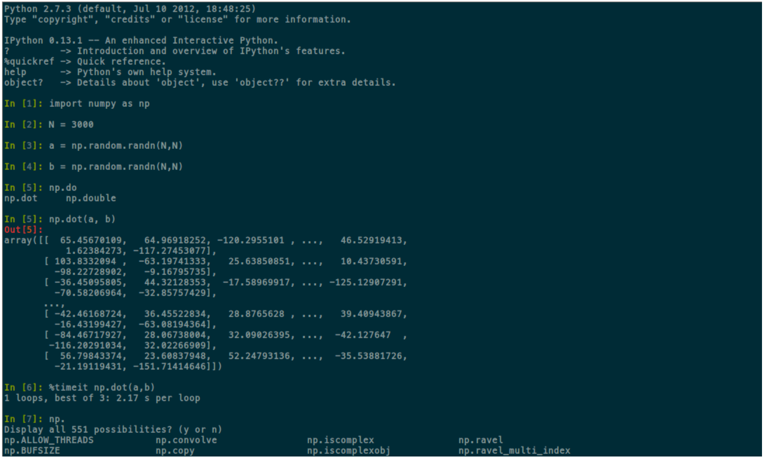
- Interactive Python
  - GUI/Kernel for Python/Jupyter
  - Easy to use
  - Visually appealing
  - Interactive data visualization
  - Debugging
  - Tab completion
  - High performance tools for parallel computing

IP[y]: IPython
Interactive Computing

- Interactive Python
  - GUI/Kernel for Python/Jupyter
  - Easy to use
  - Visually appealing
  - Interactive data visualization
  - Debugging
  - Tab completion
  - High performance tools for parallel computing
  - Open source

IP[y]: IPython
Interactive Computing

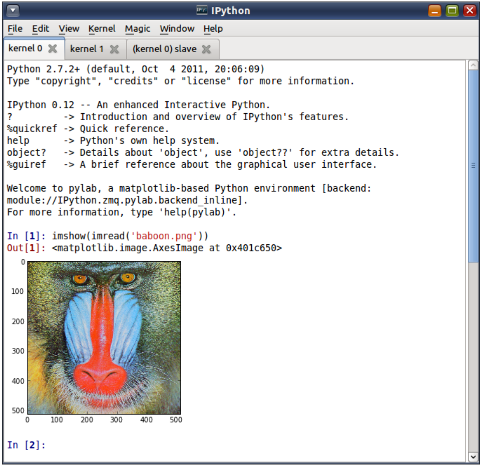
- Terminal/Command Line



```
Python 2.7.3 (default, Jul 10 2012, 18:48:25)
Type "copyright", "credits" or "license" for more information.

IPython 0.13.1 -- An enhanced Interactive Python.
?         -> Introduction and overview of IPython's features.
%quickref -> Quick reference.
help      -> Python's own help system.
object?   -> Details about 'object', use 'object??' for extra details.

In [1]: import numpy as np

In [2]: N = 3000

In [3]: a = np.random.randn(N,N)

In [4]: b = np.random.randn(N,N)

In [5]: np.do
np.dot     np.double

In [5]: np.dot(a, b)
Out[5]:
array([[  65.45670109,   64.96918252, -120.2955101 , ...,   46.52919413,
           1.62384273, -117.27453077],
       [ 103.8332094 ,  -63.19741333,   25.63850851, ...,   10.43730591,
         -98.22728902,   -9.16795735],
       [ -36.45095805,   44.32128353,  -17.58969917, ..., -125.12907291,
         -70.58206964,  -32.85757429],
       ...,
       [ -42.46168724,   36.45522834,   28.8765628 , ...,   39.40943867,
         -16.43199427,  -63.08194364],
       [ -84.46717927,   28.06738004,   32.09026395, ...,  -42.127647  ,
        -116.20291034,   32.02266909],
       [  56.79843374,   23.60837948,   52.24793136, ...,  -35.53881726,
         -21.19119431, -151.71414646]])

In [6]: %timeit np.dot(a,b)
1 loops, best of 3: 2.17 s per loop

In [7]: np.
Display all 551 possibilities? (y or n)
np.ALLOW_THREADS           np.convolve                np.iscomplex               np.ravel
np.BUFSIZE                 np.copy                    np.iscomplexobj            np.ravel_multi_index
```

IP[y]: IPython
Interactive Computing

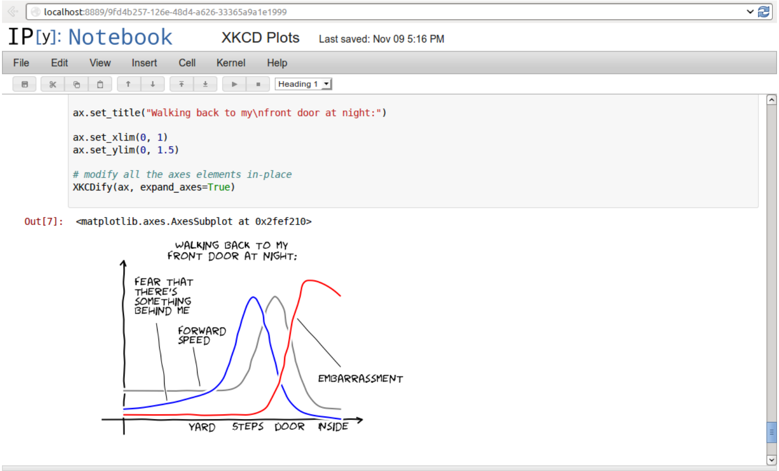
- QtConsole



```
Python 2.7.2+ (default, Oct  4 2011, 20:06:09)
Type "copyright", "credits" or "license" for more information.

IPython 0.12 -- An enhanced Interactive Python.
?         -> Introduction and overview of IPython's features.
%quickref -> Quick reference.
help      -> Python's own help system.
object?   -> Details about 'object', use 'object??' for extra details.
%guiref   -> A brief reference about the graphical user interface.

Welcome to pylab, a matplotlib-based Python environment [backend:
module://IPython.zmq.pylab.backend_inline].
For more information, type 'help(pylab)'.

In [1]: imshow(imread('baboon.png'))
Out[1]: <matplotlib.image.AxesImage at 0x401c650>

In [2]:
```

IP[y]: IPython Interactive Computing

- Notebook (Web Application)



localhost:8889/9fd4b257-126e-48d4-a626-33365a9a1e1999
IP[y]: Notebook
XKCD Plots
Last saved: Nov 09 5:16 PM
File
Edit
View
Insert
Cell
Kernel
Help
Heading 1

```
ax.set_title("Walking back to my\nfront door at night:")

ax.set_xlim(0, 1)
ax.set_ylim(0, 1.5)

# modify all the axes elements in-place
XKCDify(ax, expand_axes=True)
```

```
Out[7]: <matplotlib.axes.AxesSubplot at 0x2fef210>
```

- Open source, interactive data science and scientific computing across over 40 programming languages!

- Open source, interactive data science and scientific computing across over 40 programming languages!
    - Spin off from IPython

- Open source, interactive data science and scientific computing across over 40 programming languages!
  - Spin off from IPython
  - Based on Notebook

- Open source, interactive data science and scientific computing across over 40 programming languages!
  - Spin off from IPython
  - Based on Notebook
  - Kernels for IPython, R, Stata, Julia, Scala, etc.

- Open source, interactive data science and scientific computing across over 40 programming languages!
    - Spin off from IPython
    - Based on Notebook
    - Kernels for IPython, R, Stata, Julia, Scala, etc.
    - Big data ready...Spark, Dask

- Open source, interactive data science and scientific computing across over 40 programming languages!
  - Spin off from IPython
  - Based on Notebook
  - Kernels for IPython, R, Stata, Julia, Scala, etc.
  - Big data ready...Spark, Dask
- Share notebooks (Web, LaTeX)

- Open source, interactive data science and scientific computing across over 40 programming languages!
    - Spin off from IPython
    - Based on Notebook
    - Kernels for IPython, R, Stata, Julia, Scala, etc.
    - Big data ready...Spark, Dask
- Share notebooks (Web, LaTeX)
- Use multiple language simultaneously (e.g. Python & R & Stata)

If they're so good, why aren't we all using it?

- Lack of knowledge

If they're so good, why aren't we all using it?

- Lack of knowledge
- Change is costly

If they're so good, why aren't we all using it?

- Lack of knowledge
- Change is costly
- Disadvantages of General & Open Source Software

If they're so good, why aren't we all using it?

- Lack of knowledge
- Change is costly
- Disadvantages of General & Open Source Software
    - Installation may not be straightforward

If they're so good, why aren't we all using it?

- Lack of knowledge
- Change is costly
- Disadvantages of General & Open Source Software
    - Installation may not be straightforward
    - Learning curve

If they're so good, why aren't we all using it?

- Lack of knowledge
- Change is costly
- Disadvantages of General & Open Source Software
    - Installation may not be straightforward
    - Learning curve
    - Not always GIS ready

If they're so good, why aren't we all using it?

- Lack of knowledge
- Change is costly
- Disadvantages of General & Open Source Software
  - Installation may not be straightforward
  - Learning curve
  - Not always GIS ready

$\Longrightarrow$ Scary!

But...

- Becoming easier & more generalized

But...

- Becoming easier & more generalized
    - OSGeo distributes installers

But...

- Becoming easier & more generalized
    - OSGeo distributes installers
    - Continuum distributes Ananconda & Conda

But...

- Becoming easier & more generalized
    - OSGeo distributes installers
    - Continuum distributes Ananconda & Conda
    - Enthought distributes Canopy

But...

- Becoming easier & more generalized
  - OSGeo distributes installers
  - Continuum distributes Ananconda & Conda
  - Enthought distributes Canopy
  - Homebrew on OSX

But...

- Becoming easier & more generalized
  - OSGeo distributes installers
  - Continuum distributes Ananconda & Conda
  - Enthought distributes Canopy
  - Homebrew on OSX
  - Scientific Linux

But...

- Becoming easier & more generalized
  - OSGeo distributes installers
  - Continuum distributes Ananconda & Conda
  - Enthought distributes Canopy
  - Homebrew on OSX
  - Scientific Linux
- Lots of community support

But...

- Becoming easier & more generalized
  - OSGeo distributes installers
  - Continuum distributes Ananconda & Conda
  - Enthought distributes Canopy
  - Homebrew on OSX
  - Scientific Linux
- Lots of community support
  - Fast

But...

- Becoming easier & more generalized
  - OSGeo distributes installers
  - Continuum distributes Ananconda & Conda
  - Enthought distributes Canopy
  - Homebrew on OSX
  - Scientific Linux
- Lots of community support
  - Fast
  - Solutions

But...

- Becoming easier & more generalized
    - OSGeo distributes installers
    - Continuum distributes Ananconda & Conda
    - Enthought distributes Canopy
    - Homebrew on OSX
    - Scientific Linux
- Lots of community support
    - Fast
    - Solutions
    - Suggestions

But...

- Becoming easier & more generalized
  - OSGeo distributes installers
  - Continuum distributes Ananconda & Conda
  - Enthought distributes Canopy
  - Homebrew on OSX
  - Scientific Linux
- Lots of community support
  - Fast
  - Solutions
  - Suggestions
- Ömer's installation instructions and computational guide

My Main Workhorses

Main GIS Packages I use:

My Main Workhorses

Main GIS Packages I use:

- GeoPandas (Major Project for Geometries)

My Main Workhorses

Main GIS Packages I use:

- GeoPandas (Major Project for Geometries)
- GeoRasters (My project for Rasters - Public)

My Main Workhorses

Main GIS Packages I use:

- GeoPandas (Major Project for Geometries)
- GeoRasters (My project for Rasters - Public)
- GeoStats (My project for Statistics - Private)

My Main Workhorses

Main GIS Packages I use:

- GeoPandas (Major Project for Geometries)
- GeoRasters (My project for Rasters - Public)
- GeoStats (My project for Statistics - Private)
- HMI (My project for HMI - Private)

My Main Workhorses

Main GIS Packages I use:

- GeoPandas (Major Project for Geometries)
- GeoRasters (My project for Rasters - Public)
- GeoStats (My project for Statistics - Private)
- HMI (My project for HMI - Private)
- RasterStats

My Main Workhorses

Main GIS Packages I use:

- GeoPandas (Major Project for Geometries)
- GeoRasters (My project for Rasters - Public)
- GeoStats (My project for Statistics - Private)
- HMI (My project for HMI - Private)
- RasterStats
- Fiona

My Main Workhorses

Main GIS Packages I use:

- GeoPandas (Major Project for Geometries)
- GeoRasters (My project for Rasters - Public)
- GeoStats (My project for Statistics - Private)
- HMI (My project for HMI - Private)
- RasterStats
- Fiona
- Shapely

My Main Workhorses

Main GIS Packages I use:

- GeoPandas (Major Project for Geometries)
- GeoRasters (My project for Rasters - Public)
- GeoStats (My project for Statistics - Private)
- HMI (My project for HMI - Private)
- RasterStats
- Fiona
- Shapely
- rTree

My Main Workhorses

Main GIS Packages I use:

- GeoPandas (Major Project for Geometries)
- GeoRasters (My project for Rasters - Public)
- GeoStats (My project for Statistics - Private)
- HMI (My project for HMI - Private)
- RasterStats
- Fiona
- Shapely
- rTree
- OGR/GDAL

My Workflow

- Download data

My Workflow

- Download data
- Check it in QGIS

My Workflow

- Download data
- Check it in QGIS
- Write & Test Code using IPython QtConsole or Notebook

My Workflow

- Download data
- Check it in QGIS
- Write & Test Code using IPython QtConsole or Notebook
- Deploy to Server if needed

My Workflow

- Download data
- Check it in QGIS
- Write & Test Code using IPython QtConsole or Notebook
- Deploy to Server if needed
- Get results and analyze in Stata (soon to be replaced by R or StatsModels)

My Workflow

- Download data
- Check it in QGIS
- Write & Test Code using IPython QtConsole or Notebook
- Deploy to Server if needed
- Get results and analyze in Stata (soon to be replaced by R or StatsModels)
- Write paper in LaTeX

Examples

- Ömer's intro to GIS with Python (Geometries)

Examples

- Ömer's intro to GIS with Python (Geometries)
- Ömer's intro to GIS with Python (Rasters)

Examples

- Ömer's intro to GIS with Python (Geometries)
- Ömer's intro to GIS with Python (Rasters)
- Ömer's intro to GIS with Python (Data Munging)

Examples

- Ömer's intro to GIS with Python (Geometries)
- Ömer's intro to GIS with Python (Rasters)
- Ömer's intro to GIS with Python (Data Munging)
- Caloric Suitability Index (CSI)

Examples

- Ömer's intro to GIS with Python (Geometries)
- Ömer's intro to GIS with Python (Rasters)
- Ömer's intro to GIS with Python (Data Munging)
- Caloric Suitability Index (CSI)
- Google Location History

Example: Compute Zonal Stats I

```
#!/usr/bin/env python
# coding: utf-8
'''
======================================================
Author:  Ömer Özak, 2014 (ozak at smu.edu)
Website: http://omerozak.com
GitHub:  https://github.com/ozak/
======================================================
Python Script Template
'''
from __future__ import division
import sys, os, time
# Math, data
import numpy as np
import pandas as pd
pd.set_option('display.width', 140)

# GIS packages
#from osgeo import  ogr, osr, gdal, gdalnumeric
#from gdalconst import *
#from PIL import Image, ImageDraw
```

Example: Compute Zonal Stats II

```
#from pyGDsandbox.dataIO import df2dbf, dbf2df
#from rasterstats import zonal_stats
#import shutil, glob
#import pysal as ps
#from pysal.contrib import shapely_ext
#import shapely
from shapely.geometry import Polygon, Point
#from shapely.wkt import loads, dumps
#from shapely.ops import cascaded_union
import pyproj
#import georasters as gr
import hmi
#import fiona
import geopandas as gp
#import geopandas.tools as gpt
#from geopandas.tools import sjoin
from geopy.distance import great_circle
import geostats

# Various other packages
```

Example: Compute Zonal Stats III

```
import isounidecode          # Decode and encode text

# Directory
try:
    %cd Islam
except:
    path = os.path.abspath(__file__)
    dir_path = os.path.dirname(path)
    os.chdir(dir_path)
    %bookmark Islam

# Buffer size in meters
buf = 50000

# Set paths
path='../../data/'
if not os.path.exists(path):
    os.mkdir(path)
pathout='../../data/GIS/Cities/'
if not os.path.exists(pathout):
```

Example: Compute Zonal Stats IV

```
    os.mkdir(pathout)

# Geographical characteristics of each Ethnicity in a Buffer of buffer kms
cities = pd.read_stata(path+'AllCities.dta')
cities['geometry'] = cities.apply(lambda x: Point(x.lon, x.lat), axis=1)
cities.drop('aaanameofcity',axis=1,inplace=True)
cities = gp.GeoDataFrame(cities, crs=geostats.wgs84)
cities['city']=cities.city.apply(lambda x: isounidecode.unidecode(x[:x.find(' (')])
cities.to_file(pathout+'AllCities.shp')
cities = cities.to_crs(geostats.cea)
cities.to_file(pathout+'AllCitiesCyl.shp')
cities.geometry = cities.geometry.buffer(buf)

# Create geostats object and compute statistics
Stats = geostats.geostats(cities)
Stats.geostats()

# Export data
Stats.df.to_csv(path+'/AllCities'+str(int(buf/1000))+'.csv', index=False)
```

Example: HMI Distances with MP I

```
# coding: utf-8
# #Import packages
'''
=======================================================
Author:  Ömer Özak, 2014 (ozak at smu.edu)
Website: http://omerozak.com
GitHub:  https://github.com/ozak/
=======================================================
Program to create HMIdata for Islam Project using MultiProcessing to accelerate comp
Michalopoulos and Özak (2016)
'''
from __future__ import division
# Parallel
from IPython.parallel import Client
## Setup the clients, direct views, and balanced views
c = Client()
c.ids
dview = c[:]
view = c.load_balanced_view()
dview.activate()
```

Example: HMI Distances with MP II

```
%%px --local
import sys, os, time
# Math, data
import numpy as np
import pandas as pd
pd.set_option('display.width', 140)
from scipy.interpolate import griddata
import scipy.interpolate as interpolate
from scipy.spatial import cKDTree as KDTree
# GIS packages
from osgeo import  ogr, osr, gdal, gdalnumeric
from gdalconst import *
from PIL import Image, ImageDraw
#from pyGDsandbox.dataIO import df2dbf, dbf2df
from rasterstats import zonal_stats
import shutil, glob
import pysal as ps
import shapely
from shapely.geometry import Polygon, Point
from shapely.wkt import loads, dumps
```

Example: HMI Distances with MP III

```
from pysal.contrib import shapely_ext
from shapely.ops import cascaded_union
import pyproj
import geopandas as gp
import georasters as gr
from geopy.distance import great_circle
import fiona
import hmi
import isounidecode         # Decode and encode text
import datetime
import matplotlib.pyplot as plt

# Set directories
try:
    %cd Islam
    %matplotlib inline
except:
    path = os.path.abspath(__file__)
    dir_path = os.path.dirname(path)
    os.chdir(dir_path)
```

Example: HMI Distances with MP IV

```
# Set paths
path='../../data/'
if not os.path.exists(path):
    os.mkdir(path)
pathout='../../data/HMI/'
if not os.path.exists(pathout):
    os.mkdir(pathout)
# Import Cities shapefile using GeoPandas
cities = pd.read_stata(path+'/AllCities.dta')
cities['LAT']=cities['lat']
cities['LON']=cities['lon']
cities['city']=cities.city.apply(lambda x: isounidecode.unidecode(x[:(x.find(' (')=
cities = cities[['LAT','LON','code','city','ID']]

# Define a function that calls HMI, HMISea, HMIOcean with start point only one row i
#start_points=pd.DataFrame([cities.loc[0,:]], columns=cities.columns.values)

def computeHMI(row):
    """Compute HMI for starting at row and ending in all of cities"""
    A = hmi.HMI(pd.DataFrame([row[1]], columns=cities.columns.values), cities, lat=
```

Example: HMI Distances with MP V

```
    A.HMIdistance(export_shape=True, path=pathout+str(row[1]['code']+str(row[0])))
    return A.hmidist

def computeHMISea(row):
    """Compute HMI for starting at row and ending in all of cities"""
    A = hmi.HMISea(pd.DataFrame([row[1]], columns=cities.columns.values), cities, l
    A.HMIdistance(export_shape=True, path=pathout+str(row[1]['code']+str(row[0])))
    return A.hmidist

def computeHMIOcean(row):
    """Compute HMI for starting at row and ending in all of cities"""
    A = hmi.HMIOcean(pd.DataFrame([row[1]], columns=cities.columns.values), cities,
    A.HMIdistance(export_shape=True, path=pathout+str(row[1]['code']+str(row[0])))
    return A.hmidist

# Now compute all distances in parallel, merge GeoPandas Frames
dfhmi = view.map_async(computeHMI, cities.iterrows())
dfhmisea = view.map_async(computeHMISea, cities.iterrows())
dfhmiocean = view.map_async(computeHMIOcean, cities.iterrows())
# Get results
```

Example: HMI Distances with MP VI

```
dfhmi = dfhmi.get()
dfhmisea = dfhmisea.get()
dfhmiocean = dfhmiocean.get()
# Concatenate the results
dfhmi = pd.concat(dfhmi)
dfhmisea = pd.concat(dfhmisea)
dfhmiocean = pd.concat(dfhmiocean)
# Convert to GeoPandas again
dfhmi = gp.GeoDataFrame(dfhmi, crs=hmi.cea)
dfhmisea = gp.GeoDataFrame(dfhmisea, crs=hmi.cea)
dfhmiocean = gp.GeoDataFrame(dfhmiocean, crs=hmi.cea)
# Export Shape files
dfhmi.to_file(path+'HMI10.shp')
dfhmisea.to_file(path+'HMISea10.shp')
dfhmiocean.to_file(path+'HMIOcean10.shp')

# Merge Distances and export them
cols=dfhmi.columns
cols=cols.drop('geometry')
dfout = dfhmi[cols].copy()
```

Example: HMI Distances with MP VII

```
dfout = dfout.merge(dfhmisea[['city_1','city_2','HMISea10dist','HMISea10Iso']], how
dfout = dfout.merge(dfhmiocean[['city_1','city_2','HMIOcean10dist','HMIOcean10Iso']
dfout.sort(['city_1','city_2'], inplace=True)
dfout.to_stata('../../data/HMI10dists.dta', write_index=False)
```

Example: Extension of Original Project

New Project...similar to original one but using city data

- 4669 cities

Example: Extension of Original Project

New Project...similar to original one but using city data

- 4669 cities
- 9 versions of HMI*10

Example: Extension of Original Project

New Project...similar to original one but using city data

- 4669 cities
- 9 versions of HMI*10
- Using 149 cores on server

Example: Extension of Original Project

New Project...similar to original one but using city data

- 4669 cities
- 9 versions of HMI*10
- Using 149 cores on server
- Less than 1 day for full results (data, networks, MST, etc.)

Advantages

- High speed

Advantages

- High speed
- Reproducible research

Advantages

- High speed
- Reproducible research
- Portable across computers

Advantages

- High speed
- Reproducible research
- Portable across computers
- Shareable across users

Advantages

- High speed
- Reproducible research
- Portable across computers
- Shareable across users
- Easy parallelization

Advantages

- High speed
- Reproducible research
- Portable across computers
- Shareable across users
- Easy parallelization
- Adaptability to users needs

Advantages

- High speed
- Reproducible research
- Portable across computers
- Shareable across users
- Easy parallelization
- Adaptability to users needs
- Access to large set of tools (GIS and non-GIS)

Advantages

- High speed
- Reproducible research
- Portable across computers
- Shareable across users
- Easy parallelization
- Adaptability to users needs
- Access to large set of tools (GIS and non-GIS)
- Support

When should you use?

- Point-and-click

When should you use?

- Point-and-click
    - Fast or simple tasks/analyses

When should you use?

- Point-and-click
  - Fast or simple tasks/analyses
  - Non-repeating tasks/analyses

When should you use?

- Point-and-click
  - Fast or simple tasks/analyses
  - Non-repeating tasks/analyses
- Code

When should you use?

- Point-and-click
  - Fast or simple tasks/analyses
  - Non-repeating tasks/analyses
- Code
  - Repetitive tasks/analyses

When should you use?

- Point-and-click
  - Fast or simple tasks/analyses
  - Non-repeating tasks/analyses
- Code
  - Repetitive tasks/analyses
  - Too specific or complex tasks/analyses

Standing on the Shoulders of Giants

- No need to start from scratch

Standing on the Shoulders of Giants

- No need to start from scratch
- Use other people's code

Standing on the Shoulders of Giants

- No need to start from scratch
- Use other people's code
- Reuse your own code

Standing on the Shoulders of Giants

- No need to start from scratch
- Use other people's code
- Reuse your own code
- Write your own libraries/packages

Standing on the Shoulders of Giants

- No need to start from scratch
- Use other people's code
- Reuse your own code
- Write your own libraries/packages
- Share...share...share!

Standing on the Shoulders of Giants

- No need to start from scratch
- Use other people's code
- Reuse your own code
- Write your own libraries/packages
- Share...share...share!
- Ömer's Github

Where are we going?

- Simplification

Where are we going?

- Simplification
  - Easier to install

Where are we going?

- Simplification
  - Easier to install
  - Easier to work with

Where are we going?

- Simplification
    - Easier to install
    - Easier to work with
- Power

Where are we going?

- Simplification
  - Easier to install
  - Easier to work with
- Power
  - More packages

Where are we going?

- Simplification
  - Easier to install
  - Easier to work with
- Power
  - More packages
  - More speed

Where are we going?

- Simplification
  - Easier to install
  - Easier to work with
- Power
  - More packages
  - More speed
  - More parallelization

Where are we going?

- Simplification
    - Easier to install
    - Easier to work with
- Power
    - More packages
    - More speed
    - More parallelization
- Interactivity/Interaction

Where are we going?

- Simplification
  - Easier to install
  - Easier to work with
- Power
  - More packages
  - More speed
  - More parallelization
- Interactivity/Interaction
  - Cooler graphs, widgets, dashboards

Where are we going?

- Simplification
  - Easier to install
  - Easier to work with
- Power
  - More packages
  - More speed
  - More parallelization
- Interactivity/Interaction
  - Cooler graphs, widgets, dashboards
  - with other users

Where are we going?

- Simplification
    - Easier to install
    - Easier to work with
- Power
    - More packages
    - More speed
    - More parallelization
- Interactivity/Interaction
    - Cooler graphs, widgets, dashboards
    - with other users
    - with other languages

The Voyage of Homo-œconomicus into GIS

Ömer Özak

Dept. Economics, SMU

Here's How I Do GIS

April 28, 2021