

GEOGRAPHIC DATA DISPLAY

1

ISE369/POL369

© Robert Kelly 2023-2026

2

Get Started

- Register for an account at Redistricting Data Hub
- <https://redistrictingdatahub.org/>

English

Redistricting Data Hub

Home Data Reports

We have all the redistricting data to draw or analyze congressional and state legislative maps, and find the best way to represent your community.

[Download Data](#)

Create a free account and download individual maps or register as an API user and download multiple maps.

The Redistricting Data Hub is your one stop shop for data on congressional, state legislative, and local Community of Interest or District maps.

[Register](#) [Download Data](#) [API](#)

2

ISE369/POL369

© Robert Kelly 2023-2026

3

Reading

- Wiki – Shapefile
en.wikipedia.org/wiki/Shapefile
- GeoPandas data structures
geopandas.org/en/stable/docs/user_guide/data_structures.html
- Pyplot tutorial
matplotlib.org/stable/tutorials/introductory/pyplot.html
- Data selection in a DataFrame
<https://www.shanelynn.ie/pandas-iloc-loc-select-rows-and-columns-dataframe/>
- Adding columns to a DataFrame
<https://www.interviewkickstart.com/learn/adding-new-column-to-existing-dataframe-in-pandas>

3

ISE369/POL369

© Robert Kelly 2023-2026

4

Reference

- Geopandas documentation
<https://geopandas.org/en/stable/docs.html>
- Shapely User Manual
<https://shapely.readthedocs.io/en/stable/manual.html>
- Installing packages
<https://www.jetbrains.com/help/pycharm/installing-uninstalling-and-upgrading-packages.html>
- Matplotlib
<https://matplotlib.org/stable/api/index.html>
- Color maps
<https://matplotlib.org/stable/tutorials/colors/colormaps.html>

4

ISE369/POL369
© Robert Kelly 2023-2026
5

Previous Lectures

- Discussed some political data
 - Contributions
 - Voter registration
 - Political regions (e.g., precincts)
 - Others

Some of the characteristics of a proposed congressional district are best shown and analyzed on a map

Search for “political maps” – what did you find in image results?

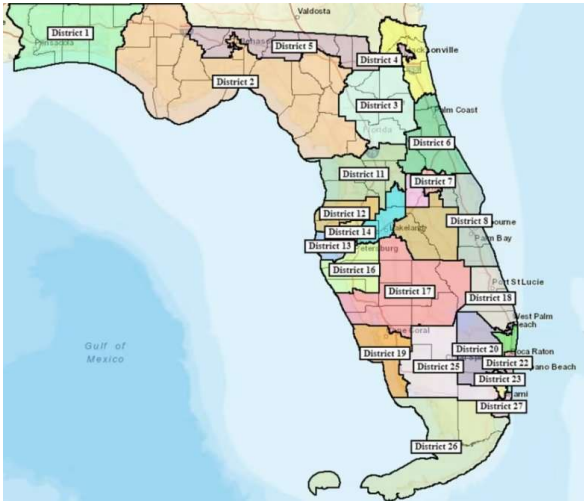


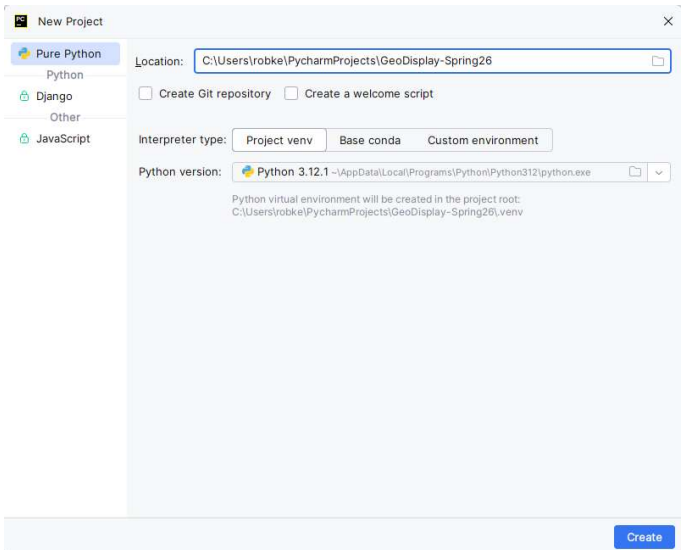
Image: news4jax.com

5

ISE369/POL369
© Robert Kelly 2023-2026
6

Install Geopandas for Windows ...

- First, set up a new PyCharm project
- Be sure to select the venv environment

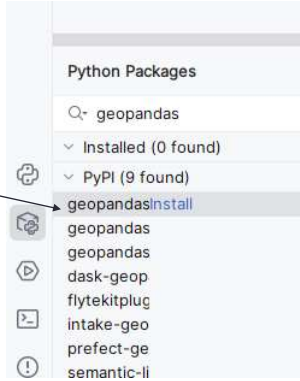


6

ISE369/POL369 © Robert Kelly 2023-2026 7

... Install Geopandas for Windows ...

- Click on the packages icon (lower left)
- Enter geopandas in search area
- Click on install (select the latest version, e.g., 1.1.2)

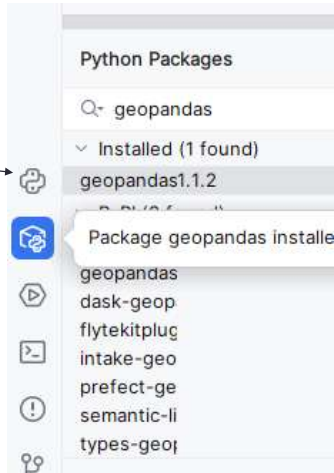


7

ISE369/POL369 © Robert Kelly 2023-2026 8

Installed Packages

- Your Python Interpreter dialog should now show many new libraries including geopandas and libraries it needs (e.g., fiona, shapely, pandas, etc.)



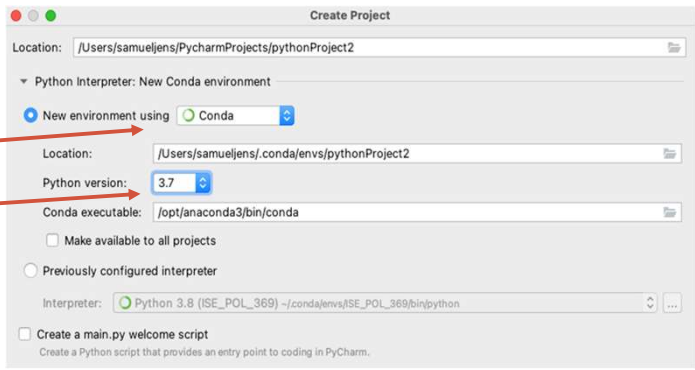
8

ISE369/POL369 © Robert Kelly 2023-2026 9

Install Geopandas for Mac...

- First, set up a new PyCharm project
- Be sure to select the Conda environment
- Select a recent version of Python

First, be sure that you have downloaded and installed Anaconda Individual Edition from <https://www.anaconda.com/products/individual>

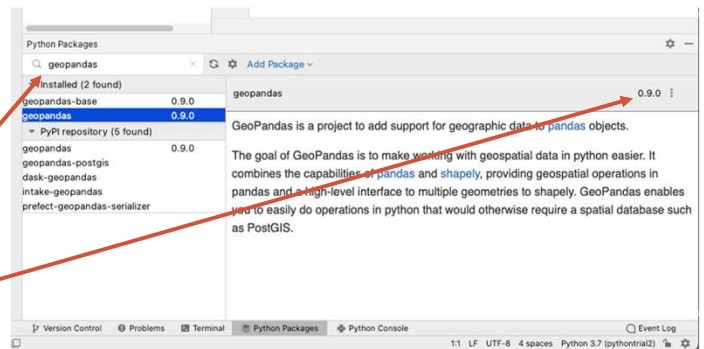


9

ISE369/POL369 © Robert Kelly 2023-2026 10

... Install geopandas on Mac ...

- Type “geopandas” in the search bar
- Click “install” on the right side [mine is installed!]



If you have any issues, please take a screenshot and post to Piazza!

10

ISE369/POL369
© Robert Kelly 2023-2026
11

Install Matplotlib

- You will also need a Python library to display boundary files you generate
- Install matplotlib in the same way you installed geopandas

Python Packages

Q- matplotlib

✓ Installed (1 found)

matplotlib 3.10.8

Package matplotlib installed

matplotlibx

matplotlib-a

matplotlib-c

matplotlibau

matplotlibhor

11

ISE369/POL369
© Robert Kelly 2023-2026
12

Download Data

- Go to the Redistricting Data Hub
- Download the District Maps for U.S. Congress 2024 plans for your favorite state
<https://redistrictingdatahub.org/data/download-data/>

For most (all?) states, you will be downloading a Shapefile

Featured Data Download

Datasets of note for this state.

Name
2020 Census Block level PL 94-171 (by table)
2020 Block, County, MCD, State, and VTD PL 94-171
State Senate Districts Plan Approved 2022 (SI)
State Assembly Districts Plan Approved 2024
Congressional Districts Plan Approved 2024
2023 CVAP Data Disaggregated to 2020 Census

State M

Browse the data individually.

12

ISE369/POL369

© Robert Kelly 2023-2026

13

Access the NY Congressional Shapefile

- File contains the boundaries of the 2024 New York Congressional districts as a Shapefile

```
import geopandas as gpd
import pandas as pd
import os
import matplotlib.pyplot as plt
os.chdir('G:\\My Drive\\Courses\\ISE369\\Data'
          + '\\GeoDataDisplayLesson'
          + '\\ny_cong_adopted_2024'
          + '\\CON24_shapefile_Feb_28_2024')
ny_cd = gpd.read_file('con24.shp')
ny_cd.plot()
plt.show()
```

There's a lot going on in this code block, so let's review it

13

ISE369/POL369

© Robert Kelly 2023-2026

14

What's in the Zip File?

Name	Date modified	Type
CON24_shapefile_Feb_28_2024	3/9/2026 4:48 PM	File folder
CON24_Feb_28_2024.dbf	2/29/2024 12:32 PM	DBF File
congress_population_2024	2/29/2024 12:31 PM	Adobe Acrobat D...
con-li	2/29/2024 12:32 PM	Adobe Acrobat D...
con-nyc	2/29/2024 12:31 PM	Adobe Acrobat D...
con-nys	2/29/2024 12:31 PM	Adobe Acrobat D...
README	2/29/2024 12:44 PM	TXT File

Name	Date modified	Type
con24.dbf	2/29/2024 12:33 PM	DBF File
con24.prj	2/29/2024 12:33 PM	PRJ File
con24.shp	2/29/2024 12:33 PM	SHP File
con24.shx	2/29/2024 12:33 PM	SHX File

14

ISE369/POL369

© Robert Kelly 2023-2026

15

Shapefiles

- Geospatial vector data format
- Developed and maintained by ESRI
- Introduced in early 1990s
- Collection of files
 - Usually stored as a zip file
 - Mandatory files (.shp, .shx, and .dbf) and other files
- Represents points, lines, polygons
- Formatted as fixed length header, followed by one or more variable length records

Dominant format for geographic data due to the market dominance of ESRI

Shapefiles are not very human readable

15

ISE369/POL369

© Robert Kelly 2023-2026

16

Display Code

```
import geopandas as gpd
import pandas as pd
import os
import matplotlib.pyplot as plt
os.chdir('G:\\My Drive\\Courses\\ISE369\\Data'
        + '\\GeoDataDisplayLesson'
        + '\\ny_cong_adopted_2024'
        + '\\CON24_shapefile_Feb_28_2024')
ny_cd = gpd.read_file('con24.shp')
ny_cd.plot()
plt.show()
```

geopandas uses
matplotlib

Change directory to the one
holding my course data

Display the
district map

Generate the
district map

Read from the directory holding the
downloaded dataset and store as a
GeoDataFrame object

16

ISE369/POL369 © Robert Kelly 2023-2026 17

Geopandas

- Open-source project that adds support for geographic data to pandas objects
- Extends the pandas DataFrame to add geometric operations
- Built on top of the Shapely geometric library for geometric operations



The diagram illustrates the structure of a GeoDataFrame. It consists of three main components: an 'index' column represented by a vertical stack of five green squares, a 'data' column represented by a 5x3 grid of yellow squares, and a 'geometry' column represented by a vertical stack of five pink squares. Below the diagram, the text 'GeoDataFrame (Source: geopandas.org)' is displayed.

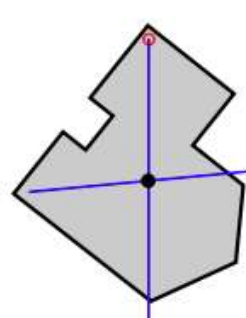
A GeoDataFrame is a DataFrame with an added geometry column

17

ISE369/POL369 © Robert Kelly 2023-2026 18

Useful GeoDataFrame Geometric Operations

- Area
- Polygon boundary
- Centroid
- Distance
- Convex hull
- Simplify
- Union
- Buffer



The diagram shows a gray-filled, irregular polygon with a black outline. A black dot represents the centroid of the polygon, with a blue line segment extending from it to a red dot on the top boundary. Another blue line segment extends horizontally from the centroid to the right edge of the polygon.

Polygon Centroids (source: Wikipedia)

18

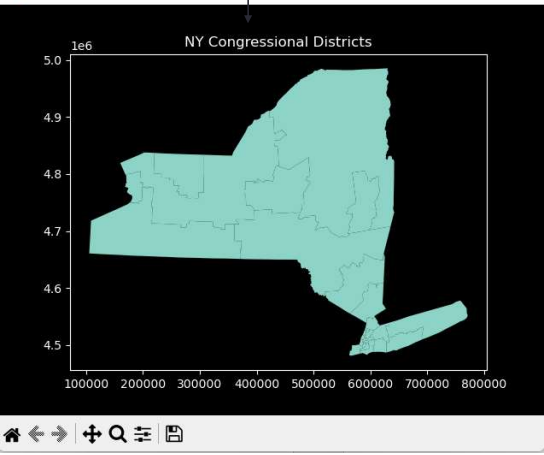
ISE369/POL369
© Robert Kelly 2023-2026
19

Matplotlib

Call the plot method on a GeoDataFrame or GeoSeries

- Python library for the creation of static, animated, and interactive visualizations
- pyplot is the main sub-library (matplotlib.pyplot)
 - Any options you can pass to pyplot in matplotlib (or line style options) can be passed to the plot() method
 - Each pyplot function makes some change to a figure (state of the plot is preserved across function calls)
 - You make changes to the display with successive calls

```
import matplotlib.pyplot as plt
plt.style.use('dark_background')
ny_cd.plot()
plt.title('NY Congressional Districts')
plt.show()
```



19

ISE369/POL369
© Robert Kelly 2023-2026
20

Python

- Notice some time we are referring to blocks of code as functions and some time as methods
- Functions
 - Block of code that has parameters
 - Used to return or pass data
 - Example


```
def factorial(n):
    return 1 if (n==1 or n==0) else n * factorial(n-1);
```
- Methods
 - Define the behavior of Python objects
 - A method operates on the data contained in the object

20

ISE369/POL369

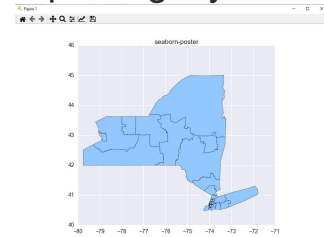
© Robert Kelly 2023-2026

21

Are We On Track?

- Download a proposed district plan for some state (not NY) from the Redistricting Data Hub
- Read the plan and assign it to a GeoDataFrame object
- Available plot styles can be accessed as a list
`styles = plt.style.available`
- Write a loop to show your state with all the available plotting styles. Inside the loop you will
 - Use the style
 - Plot the state
 - Set the plot title to be the name of your style

Check the next slide if you would like a refresh on looping over a list in Python



21

ISE369/POL369

© Robert Kelly 2023-2026

22

Python

- For loop over a list

```
# Python code to iterate over a list
list = [1, 3, 5, 7, 9]

# Using for loop
for i in list:
    print(i)
```

22

ISE369/POL369

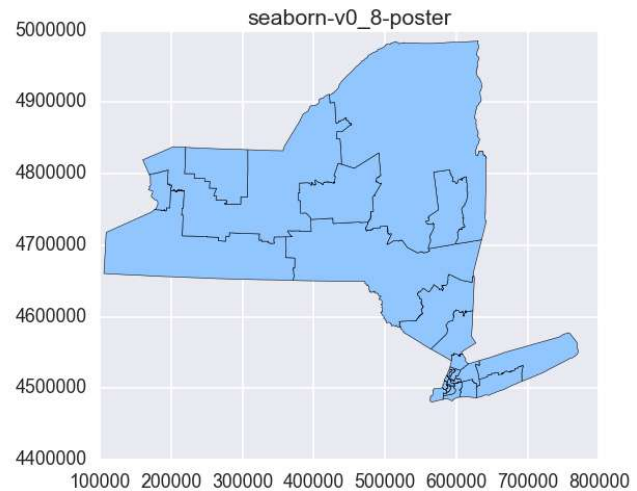
© Robert Kelly 2023-2026

23

Were We On Track?

```
print('Available styles',
plt.style.available)
styles = plt.style.available
for s in styles:
    plt.style.use(s)
    ny_cd.plot()
    plt.title(s)
    plt.show()
```

Figure 1



23

ISE369/POL369

© Robert Kelly 2023-2026

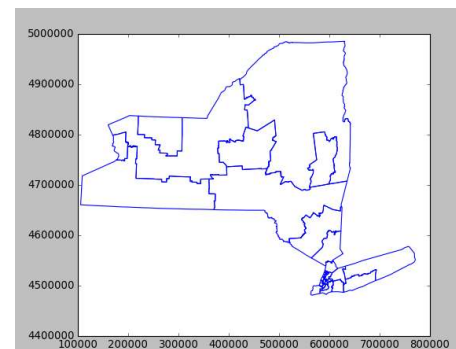
24

Show Boundaries

- But in some of the styles, the image does not clearly show the district boundaries since the boundary blends with the image background
- We can show just the boundaries with the following statements

```
...
ny_cd = geopandas.read_file('con24.shp')
plt.style.use('classic')
ny_cd.boundary.plot()
plt.show()
```

The boundary property returns a
GeoSeries, which can then be plotted



24

ISE369/POL369
© Robert Kelly 2023-2026
25

What is ny_cd?

```
...
ny_cd = geopandas.read_file('con24.shp')
print(ny_cd)
```

ny_cd is a GeoDataFrame

	OBJECTID	DISTRICT	geometry
0	1	1	POLYGON ((742169.526 4532898.369, 736334.25 45...
1	2	2	POLYGON ((679978.703 4523008.025, 679995.248 4...
2	3	3	POLYGON ((632251.823 4531012.738, 632272.906 4...
3	4	4	POLYGON ((622128.271 4510545.013, 622090.862 4...
4	5	5	POLYGON ((668338.1 4580338.582, 668338.523 4...

Row #

25

ISE369/POL369
© Robert Kelly 2023-2026
26

GeoDataFrame

- Part of the geopandas library
- A pandas.DataFrame that has a column with geometry
- A GeoDataFrame always has one GeoSeries column (geometry) that holds a special status
- Spatial methods applied to a GeoDataFrame will operate on the “geometry” column
- Geometry column accessed through `gdf.geometry`
- The name of the geometry column accessed through `gdf.geometry.name`

Geometry objects are Shapely objects

26

ISE369/POL369

© Robert Kelly 2023-2026

27

GeoDataFrame

- A GeoDataFrame extends a pandas DataFrame, so you need to import pandas to use DataFrame options

```
import geopandas
import os
import pandas as pd
os.chdir(...)
ny_cd = geopandas.read_file('con24.shp')
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
print(ny_cd)
```

Sets the DataFrame options

27

ISE369/POL369

© Robert Kelly 2023-2026

28

What are the Columns in this GeoPandas DataFrame?

```
...
for col_name in ny_cd.columns:
    print(col_name)
```

```
OBJECTID
DISTRICT
geometry
```

Let's look at this column more carefully

28

Geometry Column in a GeoDataFrame

```
...
geom_cd = ny_cd.iat[0,2]
print(geom_cd)
```

- `iat` accesses a single `DataFrame` element by integer position, so this accesses the geometry of Congressional District # 1

POLYGON ((742169.5257000001 4532898.369100001, 736334.2504000003 4529914.095799999, 735

Note that this GeoDataFrame cell contains a polygon.

Accessing Data in a DataFrame

Multiple approaches differ in what you access and what the property returns

- `iat`
 - accesses a single value for a row/column by integer position
 - Use `iat` if you only need to get or set a single value
- `at`
 - Similar to `iat`, but accesses a single value by label pair
- `iloc`
 - Integer based indexing for selection by position
 - Allowed inputs include an integer, a list or array of integers, and a slice object
- `loc`
 - Access a group of rows and columns by labels
 - Allowed inputs include a single label, a list or array of labels, and a slice object with labels

ISE369/POL369
© Robert Kelly 2023-2026
31

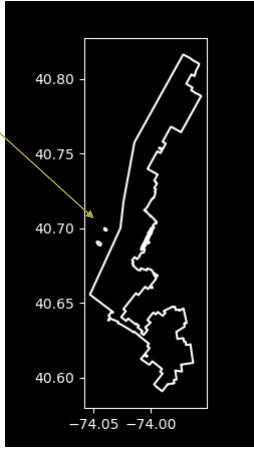
GeoPandas Data Structures

- GeoPandas implements 2 main data structures
 - GeoSeries – subclass of pandas.Series
 - GeoDataFrame – subclass of pandas.DataFrame
- GeoPandas shapes, which are Shapely objects
 - Points / Multi-Points
 - Lines / Multi-Lines
 - Polygons / Multi-Polygons

For example, a congressional district might be defined as a multi-polygon or multiple polygons if there are islands in the district

Each entry in a GeoSeries vector is a set of shapes

Shapely operations assume a cartesian coordinate system



One previous NY district (#10) contained multiple polygons

31

ISE369/POL369
© Robert Kelly 2023-2026
32

Shapely

- A Python package for computational geometry
- Fundamental geometric objects
 - Point – e.g., Point class
 - Curve – e.g., LineString class
 - Surface – e.g., Polygon class

Shapely assumes all features exist in the same Cartesian plane

- Geometric relationships
 - Contains
 - Intersects
 - Overlaps
 - Touches
- Geometric operations
 - Buffer
 - Convex hull
 - intersection

32

ISE369/POL369

© Robert Kelly 2023-2026

33

Shapely Attributes

- Area
- Bounds
- Length
- Minimum_clearance

33

ISE369/POL369

© Robert Kelly 2023-2026

34

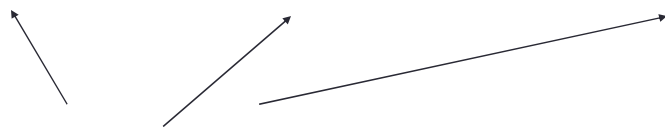
District Geometry

```
...  
geom_cd = ny_cd.iat[0,2]  
print(geom_cd)
```

```
POLYGON ((742169.5257000001 4532898.369100001, 736334.2504000003 4529914.095799999, 736334.2504000003 4532898.369100001, 742169.5257000001 4532898.369100001))
```

A polygon is
specified as a
series of points

What are these coordinates?



34

ISE369/POL369

© Robert Kelly 2023-2026

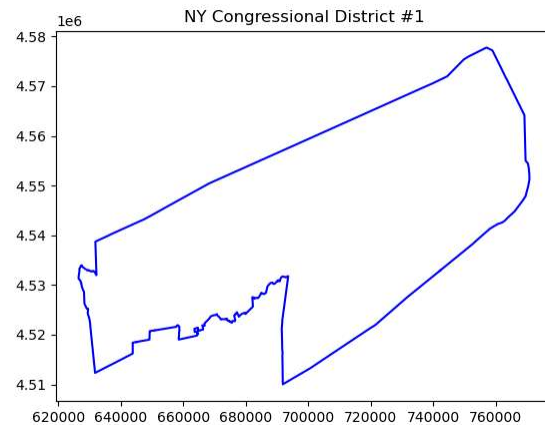
35

Are We On Track?

- For the state you downloaded, analyze the geometry of an “interesting” Congressional District
 - What is the geometric area?
 - What is the perimeter?

```
type of geom_cd1 is Polygon
Area of CD1 is 0.5396835201245012
Length of CD1 is 4.117221284683758
POLYGON ((-73.296459 40.934796999999996, -73.296554 40.934796999999996, -73.296554 40.934796999999996, -73.296459 40.934796999999996, -73.296459 40.934796999999996))
```

Use `iat` or `at` to get the geometry of a district



35

ISE369/POL369

© Robert Kelly 2023-2026

36

Were We on Track?

```
...
geom_cd1 = ny_cd.at[0,'geometry']
print('type of geom_cd1 is ', geom_cd1.geom_type)
print('Area of CD1 is ', geom_cd1.area)
print('Length of CD1 is ', geom_cd1.length)
print(geom_cd1)
```

```
type of geom_cd1 is Polygon
Area of CD1 is 4877024416.905213
Length of CD1 is 376948.36952022655
POLYGON ((742169.5257000001 4532898.369100001, 736334.2500000001 4532898.369100001, 736334.2500000001 4532898.369100001, 742169.5257000001 4532898.369100001, 742169.5257000001 4532898.369100001))
```

But what are these units?

36

ISE369/POL369

© Robert Kelly 2023-2026

37

How Do We Add a Column to a DataFrame?

- Frequently, data you need is contained in different data sources, for example
 - Geometry
 - Election results
 - Campaign contributions
 - Demographic data
- Approach options
 - Declaring a new list as a column – insert data as the last column
 - Using `DataFrame.insert()` – insert a column anywhere

37

ISE369/POL369

© Robert Kelly 2023-2026

38

Example ...

```
...
ny_cd = gpd.read_file('con24.shp')
print("\n" + 'DataFrame column names')
print(ny_cd.columns.tolist())
ny_cd['values'] = range(26)
print("\n" + 'DataFrame column names after insert')
print(ny_cd.columns.tolist())
```

columns returns the column labels

Generates 26 integers, from 0 to 25

The values column was added

```
DataFrame column names
['OBJECTID', 'DISTRICT', 'geometry']

DataFrame column names after insert
['OBJECTID', 'DISTRICT', 'geometry', 'values']
```

38

ISE369/POL369

© Robert Kelly 2023-2026

39

Python

- `range()` – built-in function that returns a range object
- Range object is a sequence of integers
- Range object is generated with a specified (or default) start, stop, and increment
- Sequence functions – `len()`, `index()`, `min()`, `max()`, `count()`
- Useful in loops

```
for i in range(5):
    print(i)
```



```
0
1
2
3
4
```

39

ISE369/POL369

© Robert Kelly 2023-2026

40

... Example

columns returns the column labels

```
...
ny_cd = gpd.read_file('con24.shp')
print("\n" + 'DataFrame column names')
print(ny_cd.columns.tolist())
ny_cd['values'] = range(26)
print("\n" + 'DataFrame column names after insert')
print(ny_cd.columns.tolist())
print(ny_cd)
```

	OBJECTID	...	values
0	1	...	0
1	2	...	1
2	3	...	2
3	4	...	3
4	5	...	4

40

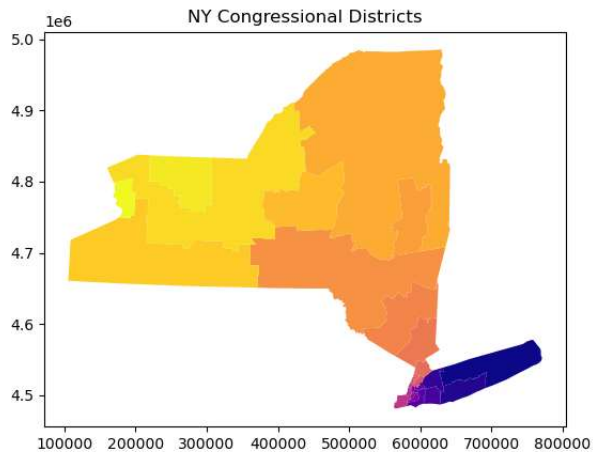
ISE369/POL369

© Robert Kelly 2023-2026

41

How Do We Color Regions in a Map?

- Use columns in the GeoDataFrame that specify
 - Geometry of each region – column labelled 'geometry'
 - Data value for each region – you specify label (e.g., 'values')
- For each region, the value is mapped into a color using a color map
 - Standard color maps are provided with matplotlib
 - You can specify your own color map



Color map for above plot is plasma

41

ISE369/POL369

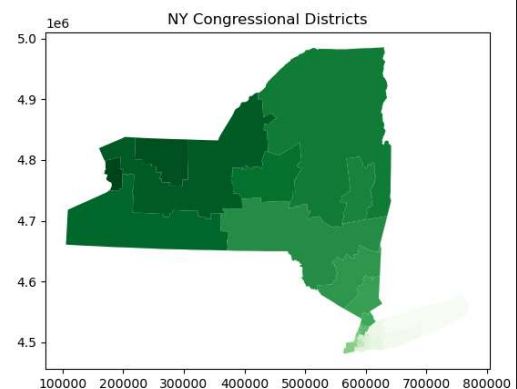
© Robert Kelly 2023-2026

42

Example

A single hue color map is best for ordered data

```
...
ny_cd = gpd.read_file('con24.shp')
ny_cd['values'] = range(26)
ny_cd.plot(column='values', cmap='Greens')
plt.title('NY Congressional Districts')
plt.show()
```



We could instead code

```
ny_cd['values'] = range(0, 26, 1)
```

The districts in the file are in numerical order, so the shades of green give an indication of district order

42

ISE369/POL369
© Robert Kelly 2023-2026
43

Color Maps

- Matplotlib provides many predefined color maps
- You can define your own color maps (e.g., Red/Blue vote share)

Track: Insert a values column into your state GeoDataFrame, and try different color maps

Color map gallery at
https://matplotlib.org/stable/gallery/color/colormap_reference.html

Sequential colormaps

Diverging colormaps

43

ISE369/POL369
© Robert Kelly 2023-2026
44

Example – Combining Two Data Sources

- In a previous example we generated shaded regions of a map with a sequence of colors
- Now we generate shaded regions of a map based on a separate data source
- We will shade the map based on the percentage of Republican registrations in a Congressional district

NY Republican Registration (%)

44

ISE369/POL369

© Robert Kelly 2023-2026

45

Example-Setup

- We get district registration data from <https://elections.ny.gov/enrollme nt-congressional-district>
- We store in congress_feb26.xlsx
- File is an Excel spreadsheet with titles (not pandas labels)
- We eliminate the titles directly in Excel, producing the spreadsheet fragment shown on the right

	A	B	C	D	E	F
1	DISTRICT	COUNTY	STATUS	DEM	REP	CO
2						
3	1	Suffolk	Active	180,008	191,206	11,53
4	1	Suffolk	InActive	9,648	10,006	61
5	1	Suffolk	Total	189,656	201,212	12,14
6						
7	1	District Total	Active	180,008	191,206	11,53
8	1	District Total	InActive	9,648	10,006	61
9	1	District Total	Total	189,656	201,212	12,14
10						
11	2	Nassau	Active	14,716	29,055	93
12	2	Nassau	InActive	1,032	1,425	5
13	2	Nassau	Total	15,748	30,480	98
14						
15	2	Suffolk	Active	160,670	137,282	8 56

45

ISE369/POL369

© Robert Kelly 2023-2026

46

Example – Code ...

```

os.chdir('G:/My Drive/Courses/ISE369/Data'
         + '/GeoDataDisplayLesson')
print(os.getcwd())
nyreg=pd.read_excel('congress_feb26.xlsx')
nyreg=nyreg[['DISTRICT', 'COUNTY', 'STATUS', 'DEM', 'REP', 'TOTAL']]
nyreg=nyreg[nyreg['COUNTY']=='District Total']
nyreg=nyreg[nyreg['STATUS']=='Total']
nyreg['Rep%'] = nyreg['REP']/nyreg['TOTAL']
rep_pct = nyreg['Rep%'].tolist()

```

Select only the rows that
contain totals for each district

Generate a list containing the
Republican percentages

Create a new column that
contains the Republican
registration percentage

46

ISE369/POL369 © Robert Kelly 2023-2026 47

... Example - Code

Change current directory to one containing the district geometry

```
os.chdir('G:/My Drive/Courses/ISE369/Data'
         + '/GeoDataDisplayLesson/ny_cong_adopted_2024'
         + '/CON24_shapefile_Feb_28_2024')
ny_cd = gpd.read_file('con24.shp')
ny_cd['values'] = rep_pct
ny_cd.plot(column='values', cmap='Reds')
plt.title('NY Congressional Districts')
plt.show()
```

47

ISE369/POL369 © Robert Kelly 2023-2026 48

Are We On Track?

- Calculate the Democratic percentage of registered voters in NY
- Use the calculated values to generate a choropleth map of NY using shades of blue

48