

Analysis of Election Results Data

1

Session Objectives

- Understand access to election results data
- Aggregation – use smaller data items to generate larger results
- Data integration - merge disparate data sources

2

Election Results

- State election offices are responsible for certifying election results
- Results are made available by the state
 - Usually the Secretary of State office
 - Responsible state offices contained in <https://www.fec.gov/introduction-campaign-finance/how-to-research-public-records/state-election-offices/>
- Results are aggregated by national organizations and in some cases made more usable

3

Redistricting Data Hub (RDH)

- Collects precinct level results from state data sources
- Merges result data with precinct boundary files
- Validates data to ensure accuracy
- Data availability usually lags the state data sources

4

Example – NY 2024 County Presidential

- We need to
 - aggregate data at the county level
 - Display counties on a map with presidential election results

```
os.chdir('G:/My Drive/Courses/ISE369/Data'
          + '/ElectionResultsData/ny_2024_gen_prec'
          + '/ny_2024_gen_all_prec')
ny2024 = gpd.read_file('ny_2024_gen_all_prec.shp')
print(ny2024.columns)
```

RDH Readme file
contains
explanation of data

```
Index(['UNIQUE_ID', 'COUNTYFP', 'GEOID24', 'CountyName', 'MuniName', 'MuniID',
      'DivType', 'DivID', 'EDID', 'EDName', 'Locality', 'LocalGEOID',
      'Congress', 'Senate', 'Assembly', 'G24COND', 'G24CONO', 'G24CONR',
      'G24PREDHAR', 'G24PREOWRI', 'G24PRERTRU', 'G24PRO1NO', 'G24PRO1YES',
      'G24SLD', 'G24SLO', 'G24SLR', 'G24SUD', 'G24SUO', 'G24SUR',
      'G24USSDGIL', 'G24USSLSAR', 'G24USSOWRI', 'G24USSRSAP', 'geometry'],
      dtype='object')
```

13,402 rows
is a typical
number of
precincts in
a large state

5

... Example ...

- Data is organized at the precinct level, but we need data at the county level
- Approach
 - Isolate needed data items
 - Aggregate precincts into counties

```
ny2024_pres = ny2024[['COUNTYFP', 'CountyName',
                      'G24PREDHAR', 'G24PRERTRU']]
ny2024_pres =
ny2024_pres.groupby('CountyName')[['G24PREDHAR',
                                     'G24PRERTRU']].sum()
print(ny2024_pres)
```

Note that ny2024_pres does not have
a geometry column, so its type is
DataFrame, not GeoDataFrame

	G24PREDHAR	G24PRERTRU
CountyName		
Alban	5802	14425
Allegany	261670	98174
Bronx	45142	44763
Broome	11424	22586
Cattaraugus	...	...
Washington	11224	17268
Wayne	17074	27291
Westchester	287434	167795
Wyoming	4929	14112
Yates	4399	6098

[62 rows x 2 columns]

6

... Example ...

- To determine the winning party in each county, we need to calculate the percentage votes
- We calculate the percentage of Harris votes in each county as a new column in the DataFrame

```
ny2024_pres['Dem %'] = ny2024_pres['G24PREDHAR'] /
    (ny2024_pres['G24PRERTRU'] + ny2024_pres['G24PREDHAR'])
```

	G24PREDHAR	G24PRERTRU	Dem %
CountyName			
Albany	92589	54560	0.629219
Allegany	5802	14425	0.286844
Bronx	261670	98174	0.727176
Broome	45142	44762	0.502102

7

... Example ...

- We now have presidential election results, but the geometry in the RDH file was only at the precinct level and we removed it to generate a DataFrame with only results
- We find a data source for NY county geometry (<https://gis.ny.gov/civil-boundaries>)

9

© Robert F. Kelly, 2026

ISE369/POL369 – Introduction to Political Informatics

10

... Example ...

- We
 - read the new geometry file,
 - create a GeoDataFrame, and
 - change the name of the county name label to be consistent with our presidential results DataFrame

```
counties = gpd.read_file('Counties.shp')
counties = counties[['NAME', 'geometry']]
counties = counties.rename(columns={'NAME': 'CountyName'})
print(counties)
```

	CountyName	geometry
0	Albany	POLYGON ((605729.038 4737866.883, 606008.186 4...
1	Allegany	POLYGON ((229573.870 4712922.870, 229713.780 4...
2	Bronx	POLYGON ((595540.679 4528518.902, 595766.658 4...

10

© Robert F. Kelly, 2026

ISE369/POL369 – Introduction to Political Informatics

11

... Example ...

- We merge the two DataFrames using CountyName as the merge column

```
ny2024_pres = pd.merge(ny2024_pres, counties, on='CountyName')
print(ny2024_pres.columns)
print(ny2024_pres)
ny2024_pres = gpd.GeoDataFrame(ny2024_pres)
```

Note: we cast ny2024_pres back to a GeoDataFrame since it now has a geometry column

```
Index(['CountyName', 'G24PREDHAR', 'G24PRERTRU', 'Dem %', 'geometry'], dtype='object')
CountyName ... geometry
0 Albany ... POLYGON ((605729.038 4737866.883, 606008.186 4...
1 Allegany ... POLYGON ((229573.870 4712922.870, 229713.780 4...
2 Bronx ... POLYGON ((595540.679 4528518.902, 595766.658 4...
3 Broome ... POLYGON ((1728800.762 4606276.376, 1728077.661 4...
```

11

© Robert F. Kelly, 2026

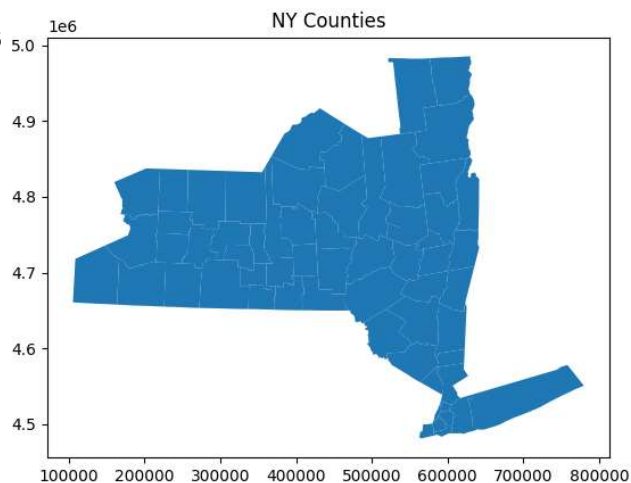
ISE369/POL369 – Introduction to Political Informatics

12

... Example ...

- We can now plot the counties

```
import matplotlib.pyplot as plt
ny2024_pres.plot()
plt.title('NY Counties')
plt.show()
```



12

© Robert F. Kelly, 2026

ISE369/POL369 – Introduction to Political Informatics

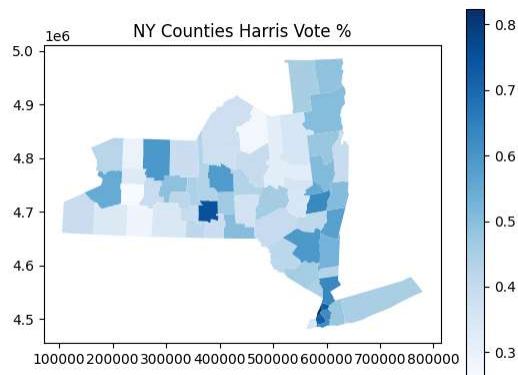
13

... Example ...

- And we can plot the Harris vote percentage as a choropleth

```
ny2024_pres.plot(column='Dem %', cmap='Blues', legend=True)
plt.title('NY Counties Harris Vote %')
plt.show()
```

Note the legend is now included in the plot

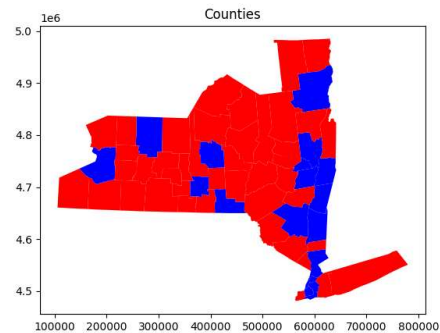


13

... Example ...

- But the map looks better when Republican districts are shown in red and Democratic districts show in blue

```
import numpy as np
ny2024_pres['Colors'] = np.where(ny2024_pres['Dem %'] >= .5, 'blue', 'red')
ny2024_pres.plot(color=ny2024_pres['Colors'])
plt.title('Counties')
plt.show()
```



15

Have You Met The Session Objectives?

- Understand access to election results data
- Aggregation – use smaller data items to generate larger results
- Data integration - merge disparate data sources

17