Discovering the geography of data opening at Data.gov

Haoliang Yu

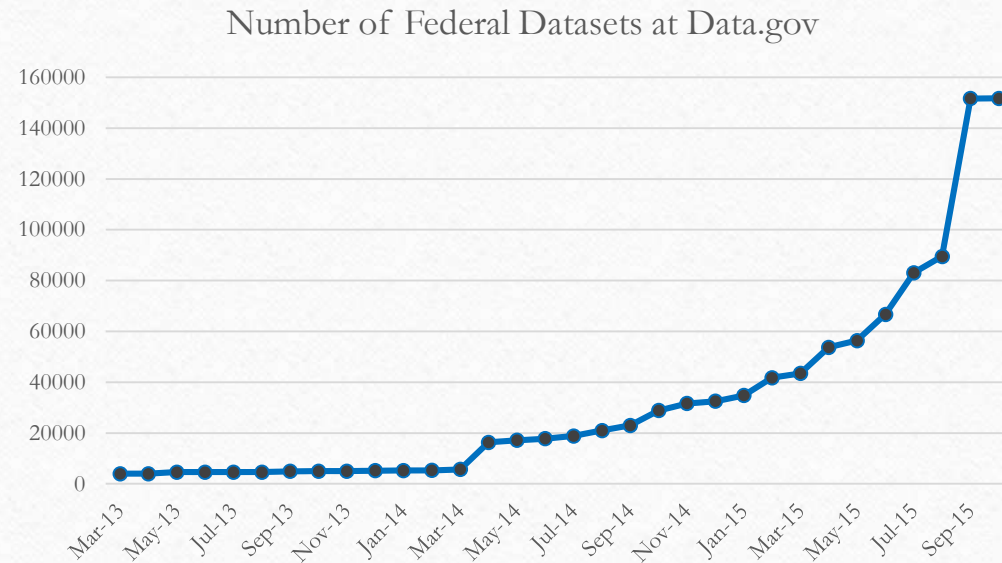NBT Solutions LLC

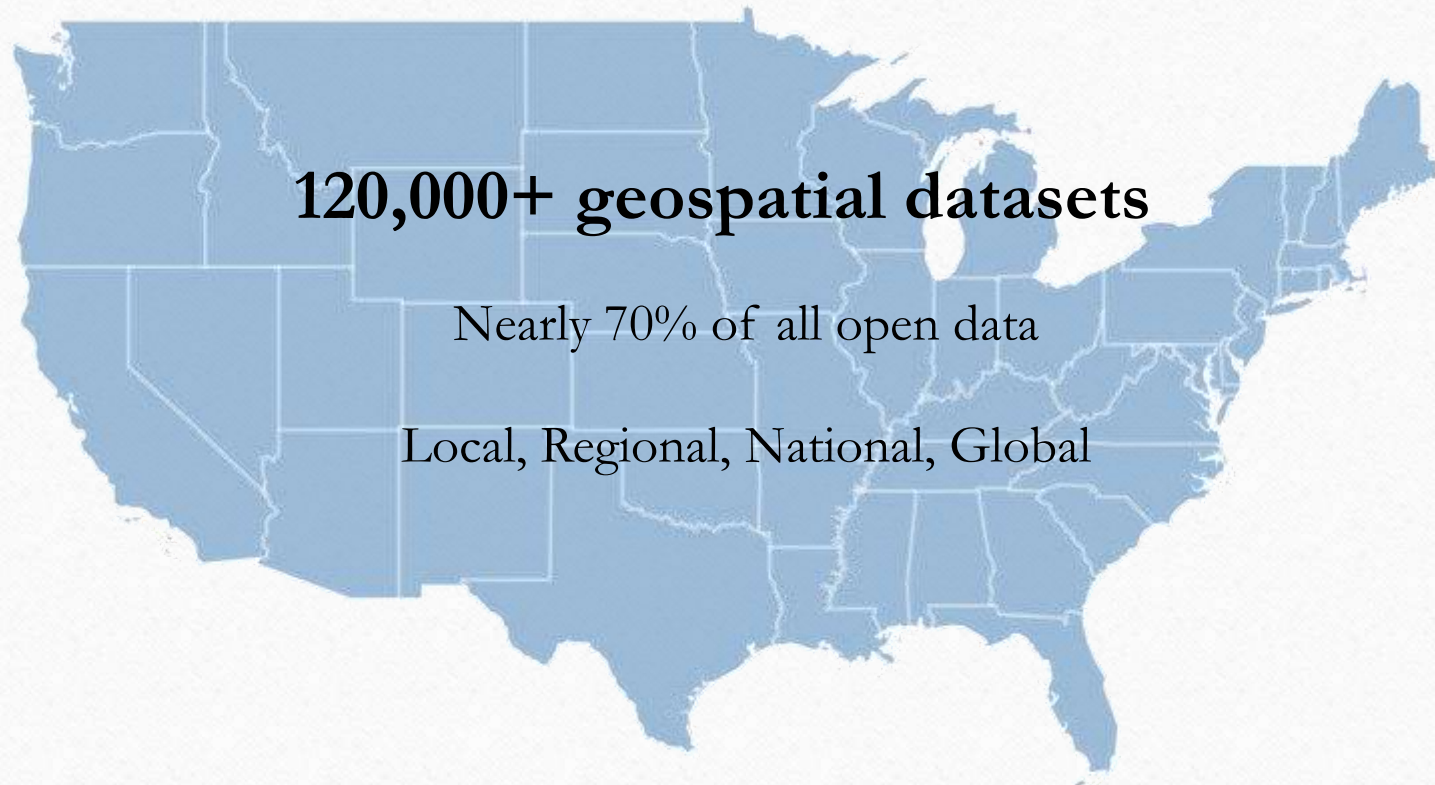haoliang.yu@nbtsolutions.com

NYGeoCon 2015

Data.gov

- A major open data portal of federal government

- Starting from 2009, tens of thousands of datasets has been published by all federal agencies

- More than 190,000 datasets are open at Data.gov.

Number of Federal Datasets at Data.gov

Geospatial Datasets at Data.gov

120,000+ geospatial datasets

Nearly 70% of all open data

Local, Regional, National, Global

Geography of Open Data

- Spatial query at data.gov - location and boundary box

- Data amount, tags, publishers, and more

- Geographic difference of data availability

- CKAN APIs

Datasets found at states
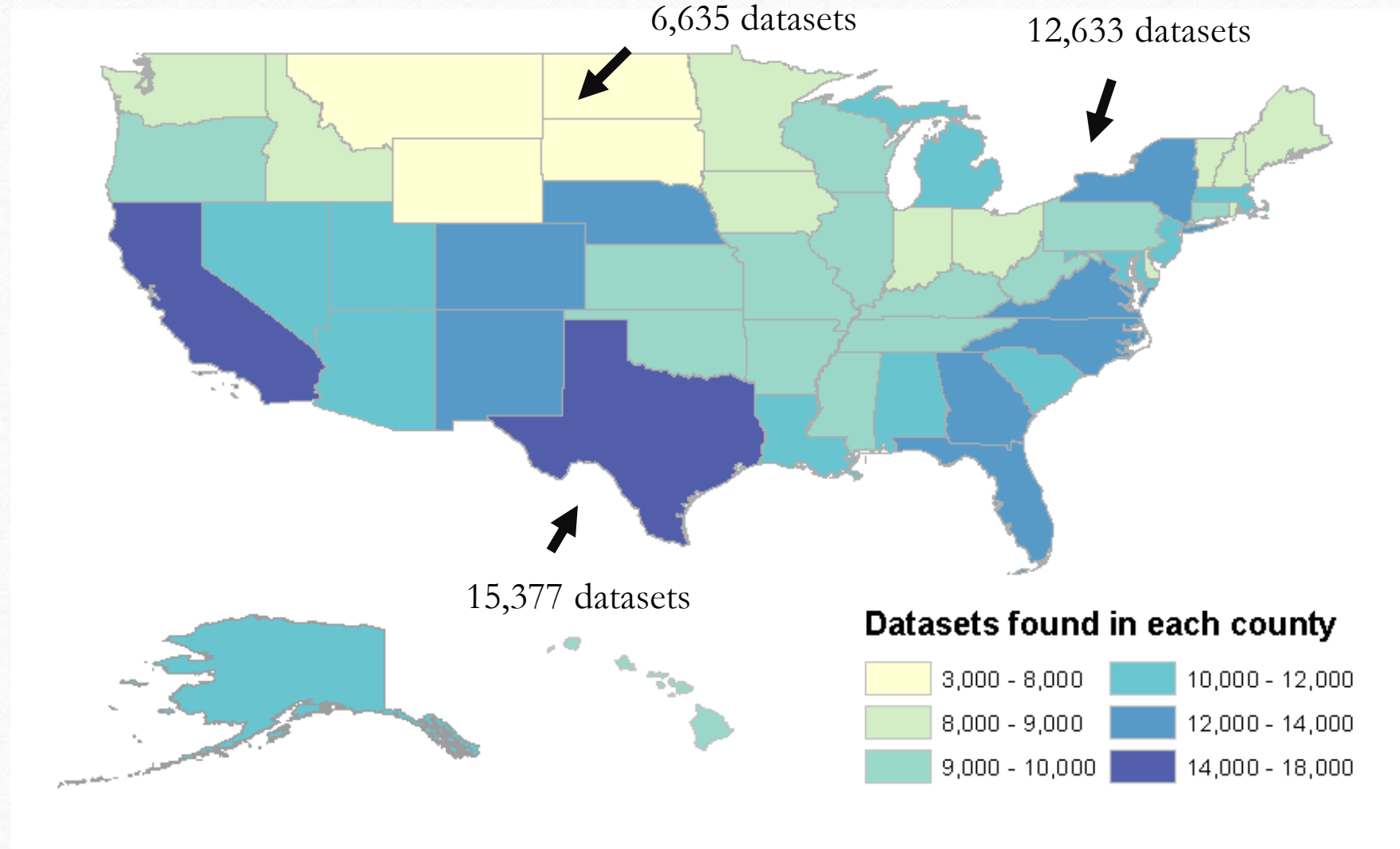
Datasets found at counties

Datasets found in each county

2,500 - 5,000	7,000 - 7,500
5,000 - 6,500	7,500 - 8,000
6,500 - 7,000	8,000+
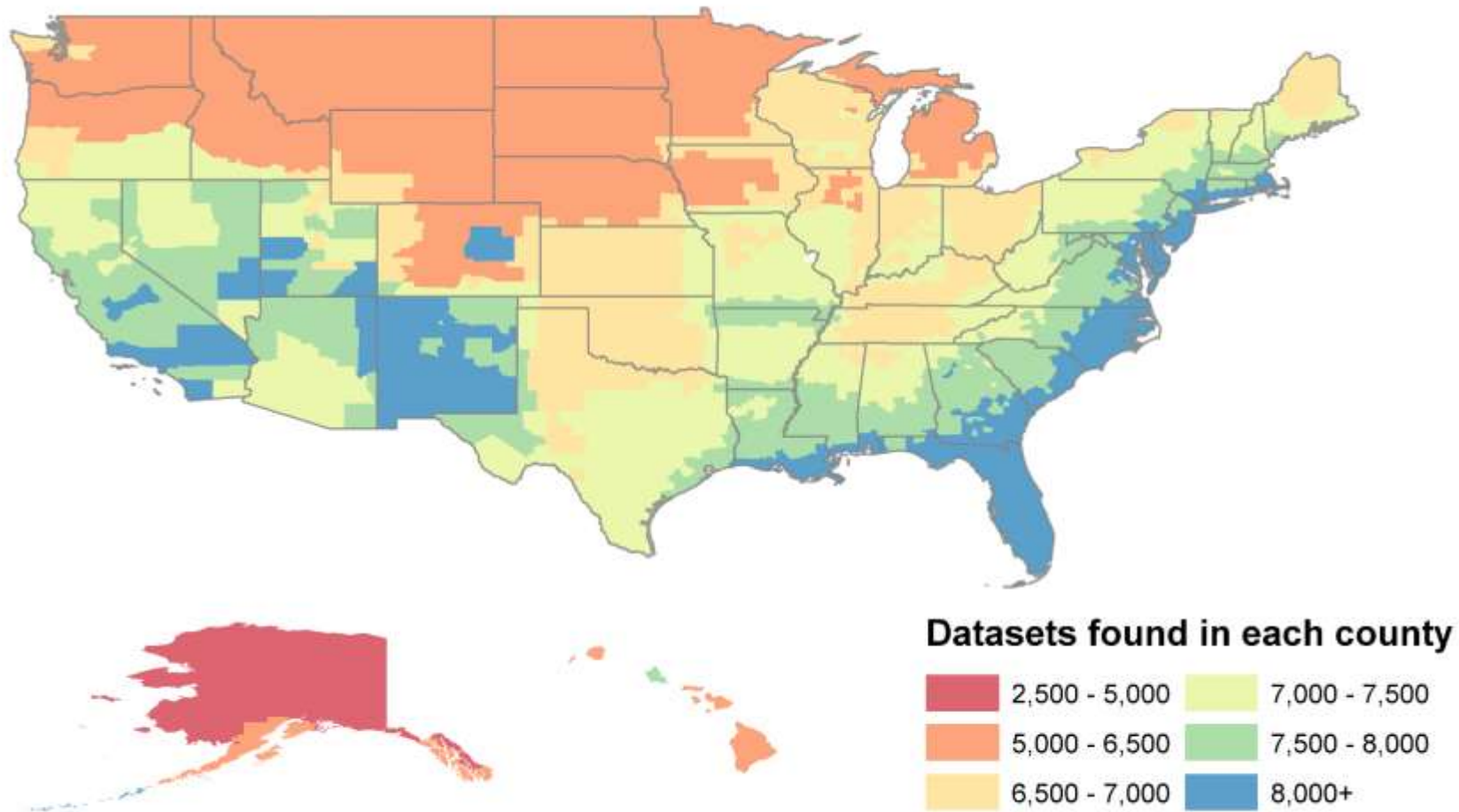
Open Data at NYS
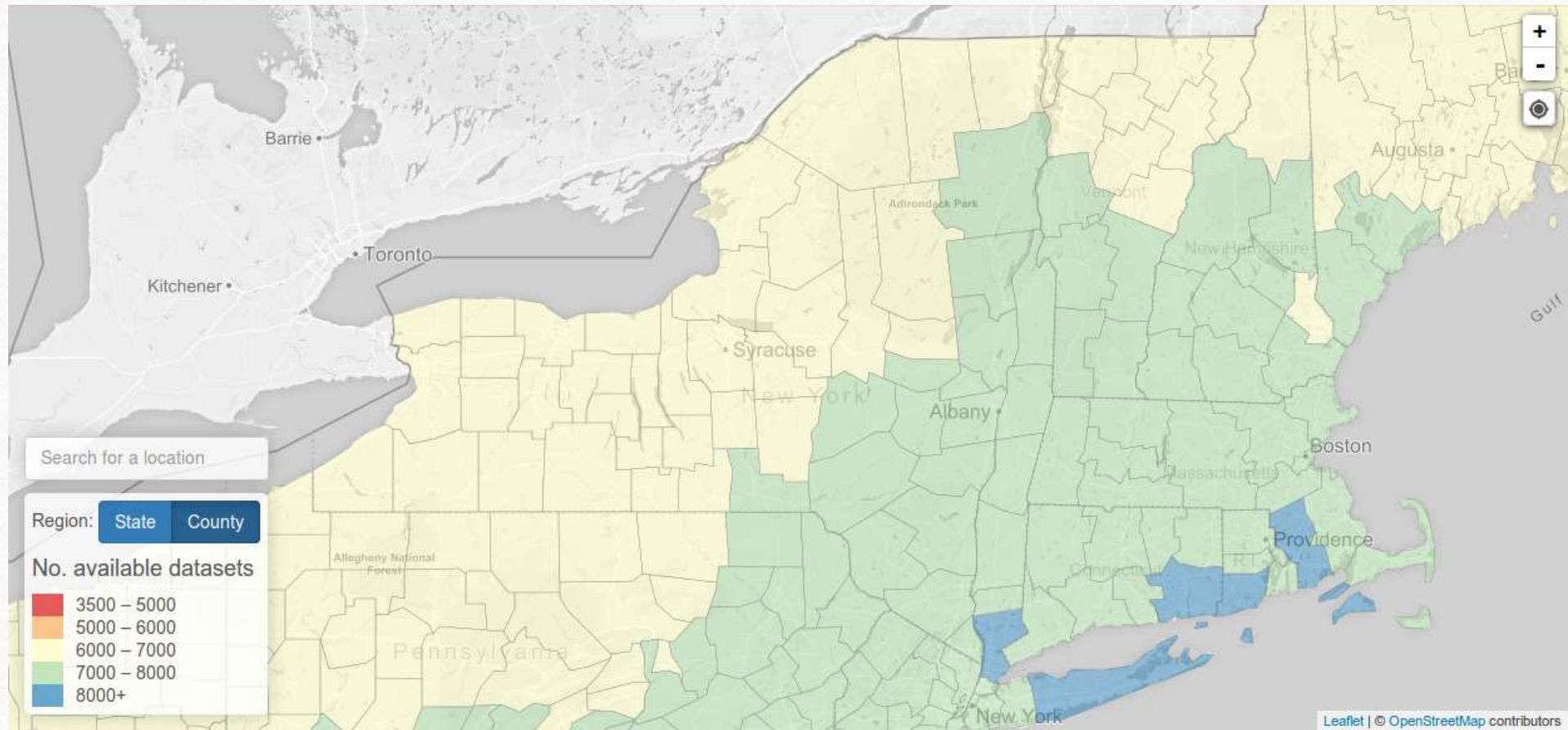
Datasets tagged with oceanography

Datasets found
[oceanography]

1000 - 2000	3000 - 3500
2000 - 2500	3500 - 4000
2500 - 3000	4000+
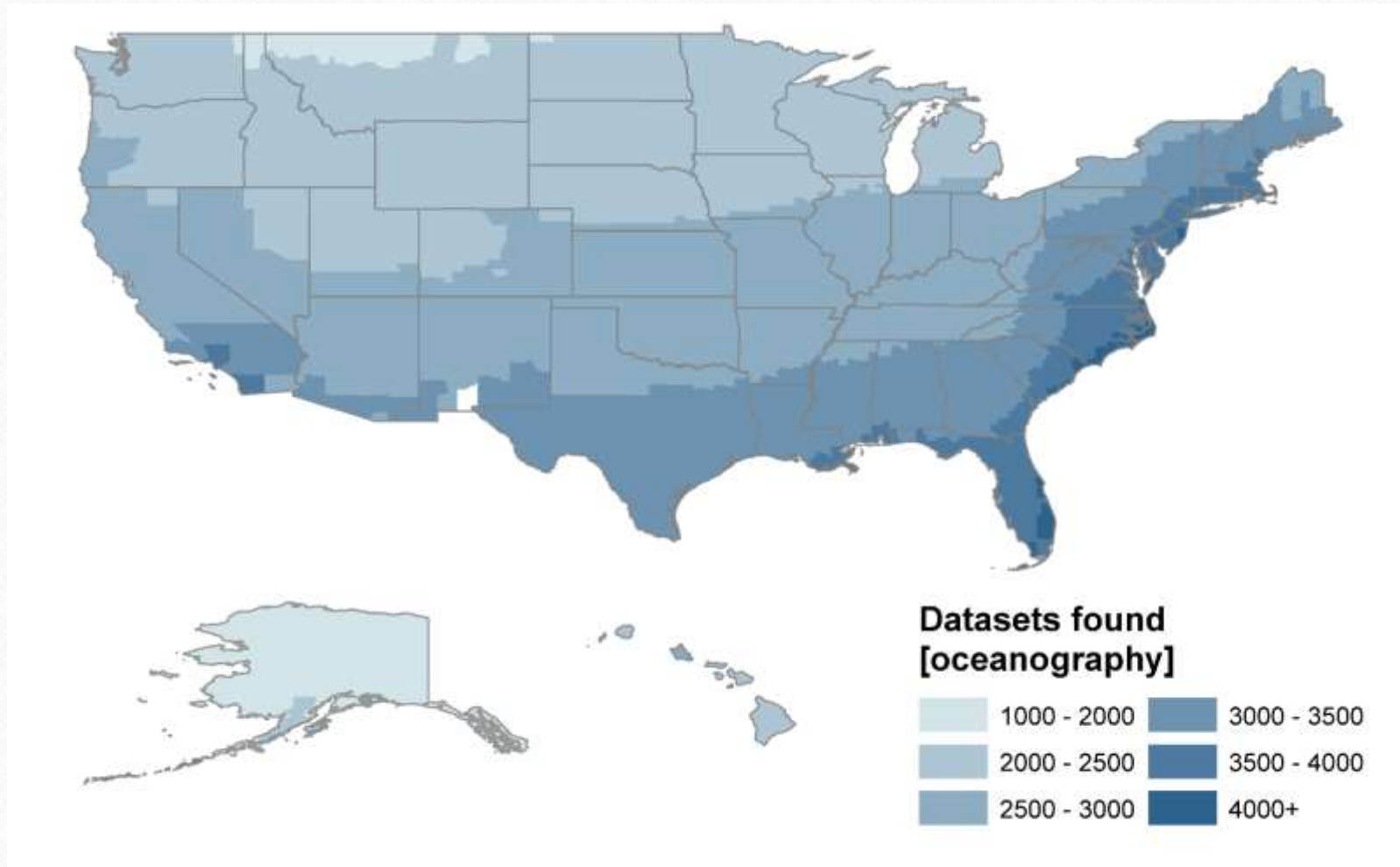
Datasets tagged with environment

Datasets found [environment]

200 - 600	750 - 800
600 - 650	800 - 850
650 - 750	850+
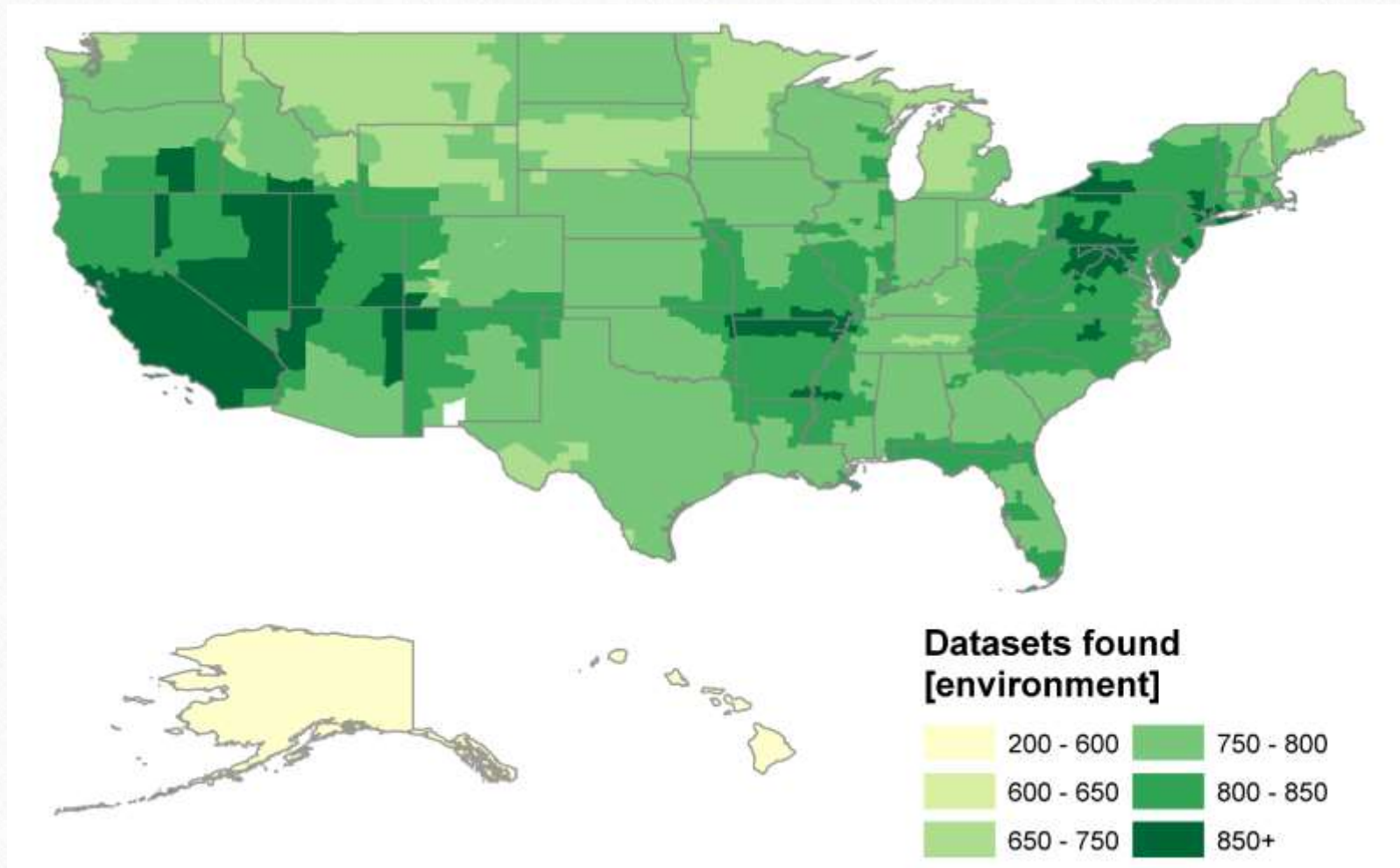
Datasets tagged with health
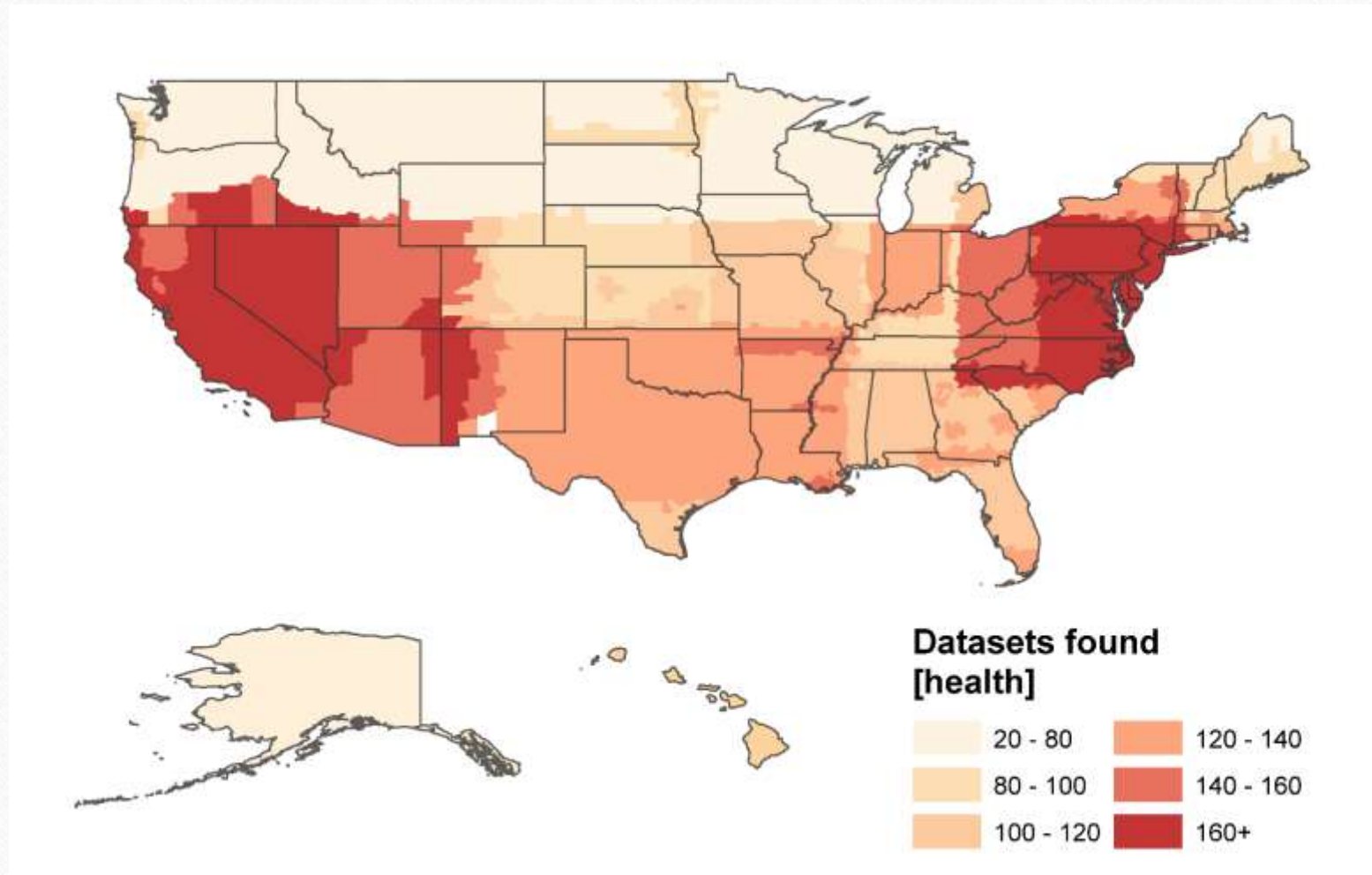
Datasets tagged with transportation

Datasets found
[transportation]

40 - 80	160 - 200
80 - 120	200 - 250
120 - 160	250+
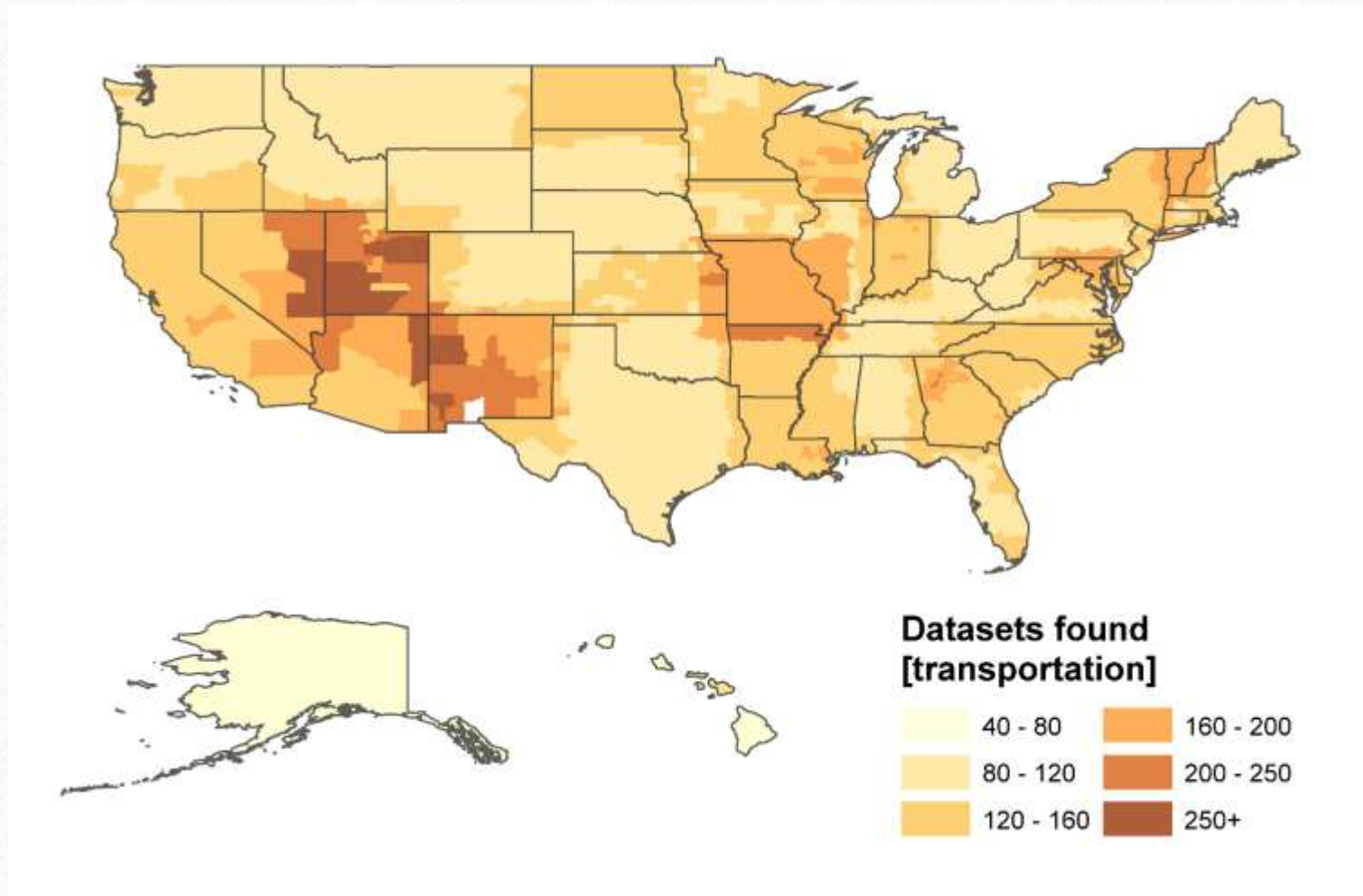
Geography of open data

- Open data is not just a collection of numbers and attributes

- Locations - geographic distribution

- Exploring the geography of open data

  - Answering the question of where

  - Another perspective to evaluate data opening

- Future: web mapping, more precise dataset number, better summary of data content

Thank you!