BEYOND CENSUS DATA: USING OPEN DATA TO UNDERSTAND COMMUNITIES AROUND PUBLIC LIBRARIES

Sponsor: Seattle Public Library

Intern: Karalyn Ostler

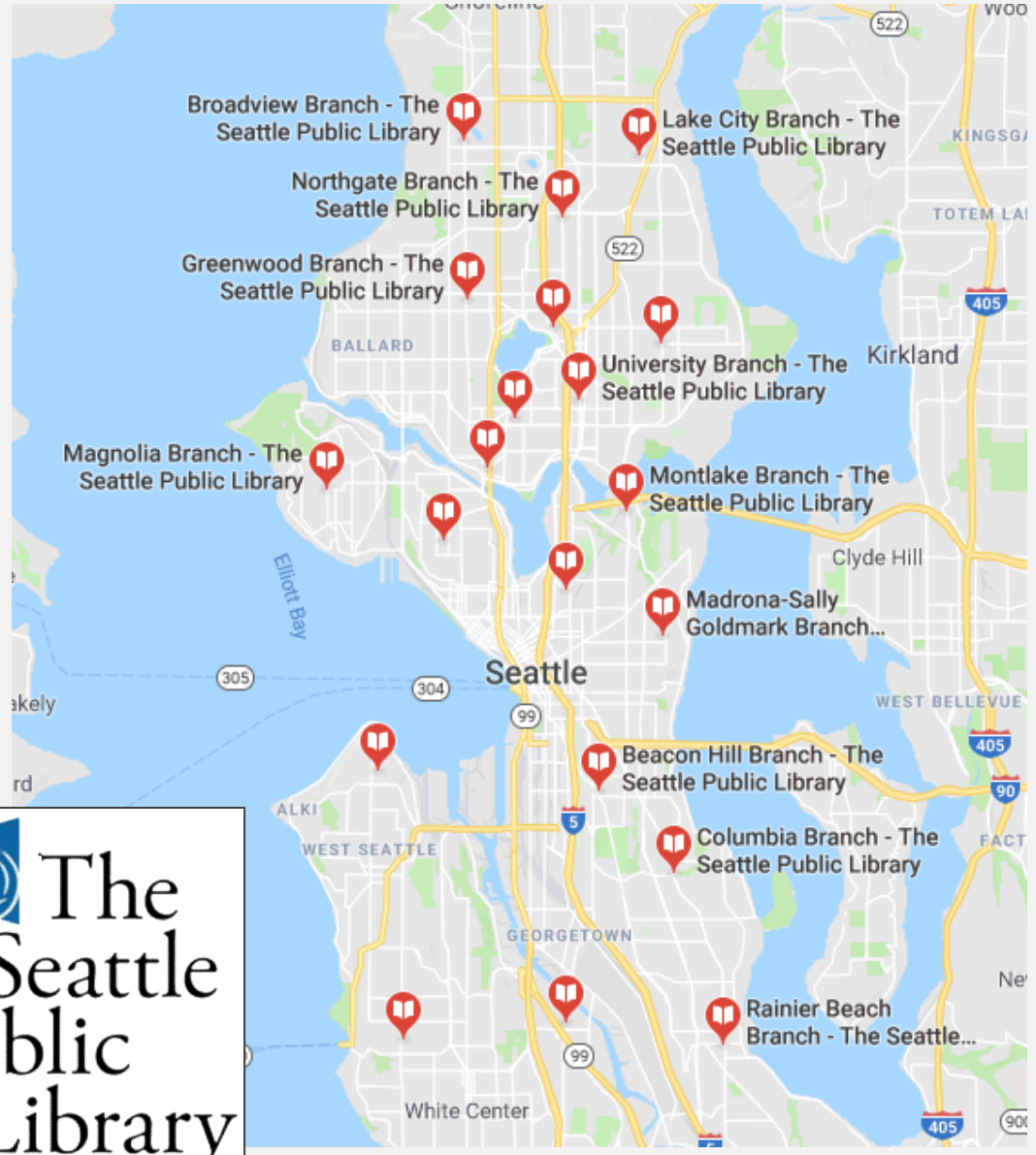
ODL
OPEN DATA LITERACY

The Seattle Public Library

SEATTLE PUBLIC LIBRARY (SPL)

- Public library system serving Seattle
- 27 Locations
- 6 regions
- Variety of services and outreach programs
- Note: Separate library system than King County Library System

- SPL Mentor: David Christensen, Data Analysis Lead



The Seattle Public Library

PROBLEM STATEMENT

- How can a large public library system like Seattle Public Libraries use external open data to help inform staff about the needs of the community to better plan services and outreach?

- How can open data be transformed into materials that are accessible and easy to understand?

PROJECT GOALS

- Identify data needs of SPL frontline staff
- Identify relevant open datasets related to Seattle and SPL
- Create materials using selected datasets for use by SPL staff

ASSESSING DATA NEEDS: INTERVIEWS

- Interviewed two SPL Regional Managers
  - Northeast Region: Francesca Wainwright
  - Southeast Region: Wei Cai
- Findings:
  - More granular data
  - Age
  - Income
  - Language Spoken at home
  - Private Schooling versus Public Schooling

SEARCH FOR SUITABLE OPEN DATA

- Local Data:
  - City of Seattle Open Data portal
  - Washington State Open data portal
- Data.gov
- Packaged Data in R
- US Census Bureau

CENSUS DATA

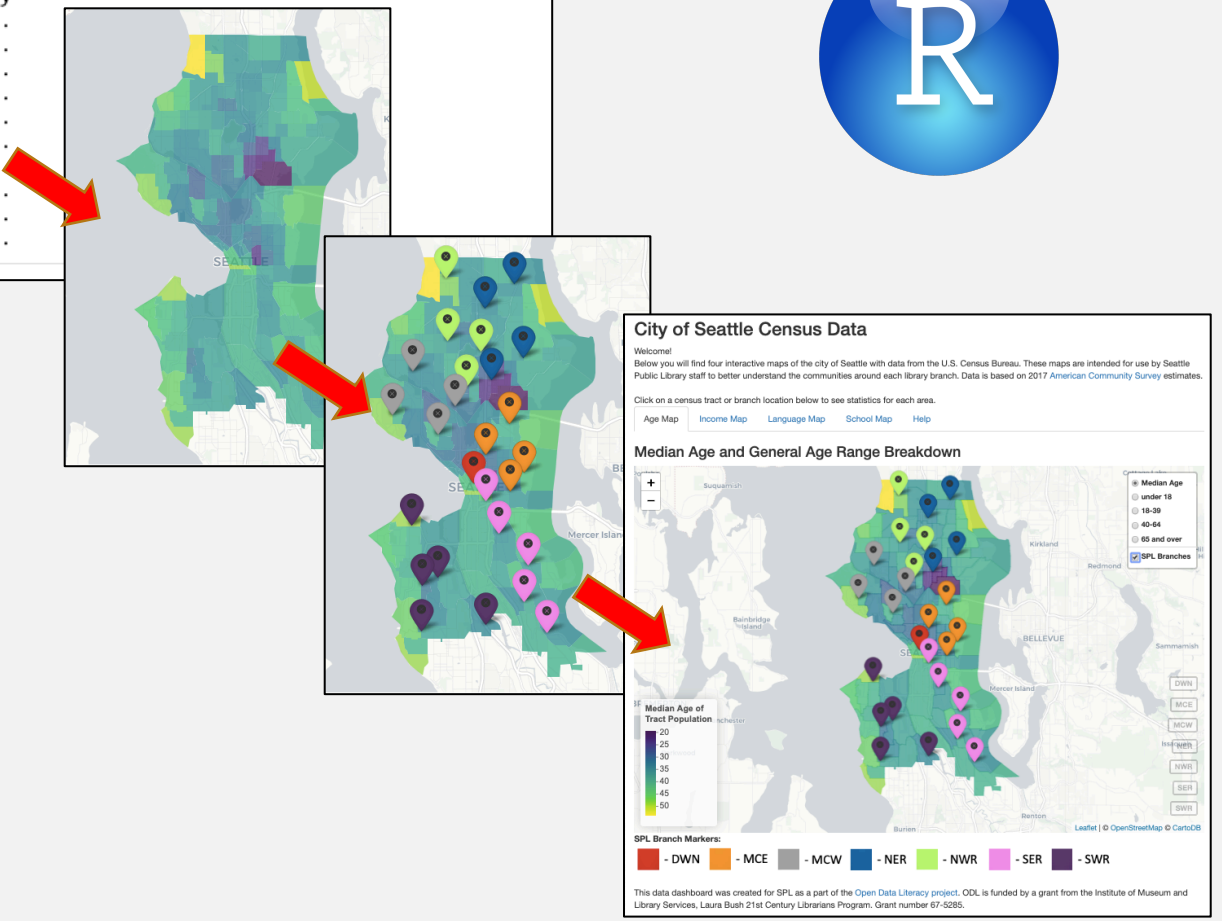
- Types of Data:
  - Censuses
    - Decennial Census
    - Economic Census
    - Census of Governments
  - Surveys
    - American Community Survey (ACS)
    - Demographic surveys
    - Economic surveys
    - Sponsored surveys

- Ways to Get Data:
  - API
  - American FactFinder

MAPS

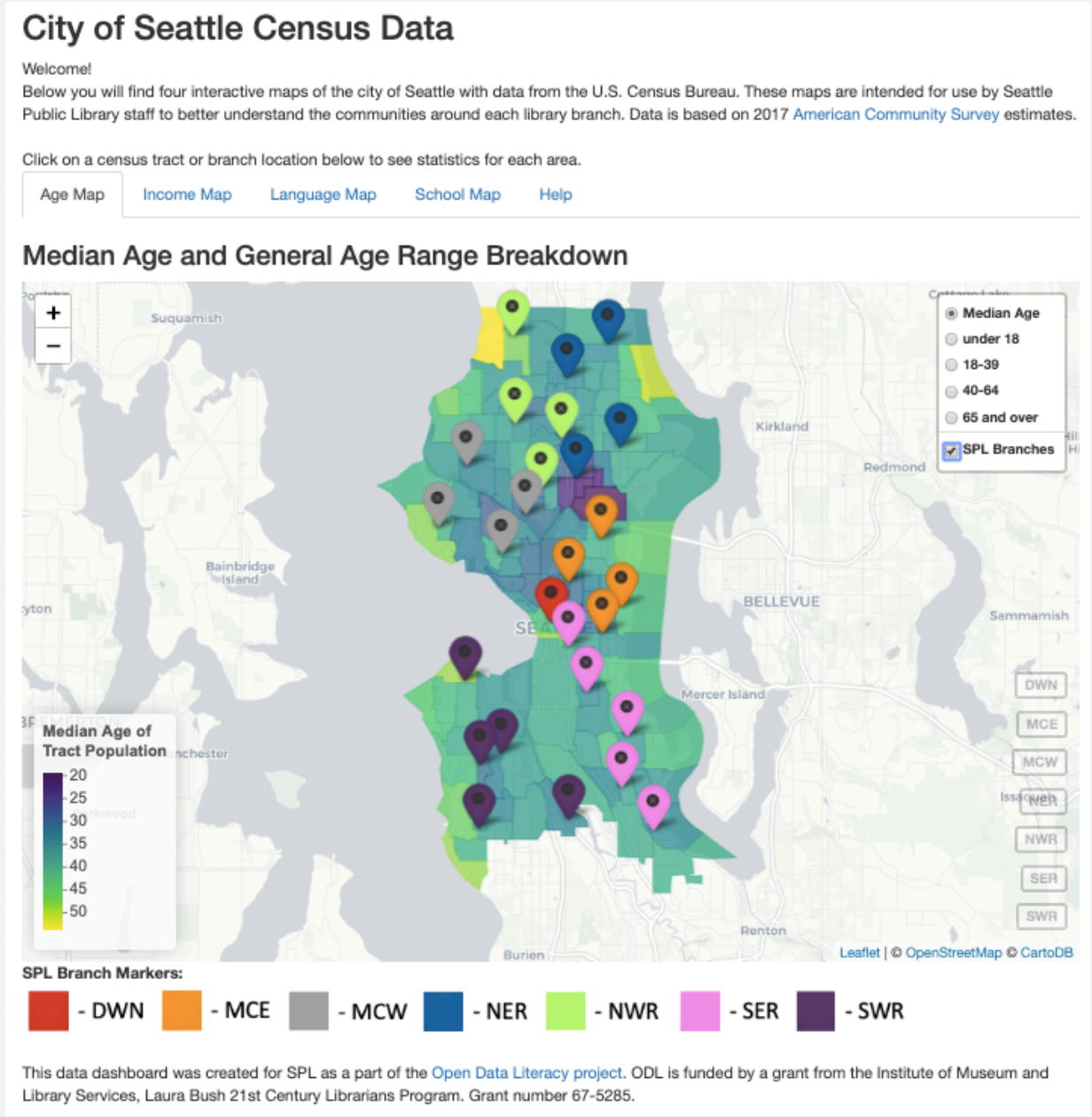
- R and Rstudio
- Download 2017 ACS data via API
- Leaflet package: Create interactive maps
- Shiny: Create application for data dashboard
- Shinyapps.io: upload dashboard to URL for sharing

```
Simple feature collection with 19502 features and 5 fields (with 49 geometries empty)
geometry type:  MULTIPOLYGON
dimension:      XY
bbox:           xmin: -122.5281 ymin: 47.0844 xmax: -121.0659 ymax: 47.78058
epsg (SRID):    4269
proj4string:    +proj=longlat +datum=NAD83 +no_defs
First 10 features:
        GEOID                                          NAME   variable estimate moe
1  53033000100 Census Tract 1, King County, Washington B01001_001     7963 529
2  53033000100 Census Tract 1, King County, Washington B01001_002     4117 397
3  53033000100 Census Tract 1, King County, Washington B01001_003      220 110
4  53033000100 Census Tract 1, King County, Washington B01001_004      384 130
5  53033000100 Census Tract 1, King County, Washington B01001_005      259 140
6  53033000100 Census Tract 1, King County, Washington B01001_006      125  86
7  53033000100 Census Tract 1, King County, Washington B01001_007       92  67
8  53033000100 Census Tract 1, King County, Washington B01001_008       50  41
9  53033000100 Census Tract 1, King County, Washington B01001_009       13  21
10 53033000100 Census Tract 1, King County, Washington B01001_010      176  70
                       geometry
1  MULTIPOLYGON (((-122.2965 4...
2  MULTIPOLYGON (((-122.2965 4...
3  MULTIPOLYGON (((-122.2965 4...
4  MULTIPOLYGON (((-122.2965 4...
5  MULTIPOLYGON (((-122.2965 4...
6  MULTIPOLYGON (((-122.2965 4...
7  MULTIPOLYGON (((-122.2965 4...
8  MULTIPOLYGON (((-122.2965 4...
9  MULTIPOLYGON (((-122.2965 4...
10 MULTIPOLYGON (((-122.2965 4...
```



MAP DASHBOARD

- URL: https://kostler.shinyapps.io/SPL-Seattle-Census-Data/
- Browser based
- Hosted by Rstudio
- Interactive
- Code easily modified:
  - Update with yearly census data releases
  - Add new data layers
  - Different city or library system



COMPARISON WITH INTERNAL DATASET

- Dataset: SPL Borrower Data
  - Divided by branches
  - # of active uses of card by age of borrower
- Age Statistics
  - Average age of borrower at Branch
  - Percentage Breakdown:
    - Under 18
    - 18-39
    - 40-64
    - Over 65

Central Library

	SPL	Census
Average Age	40.2	35.8
Under 18	4.0%	5.9%
18-39	52.1%	56.6%
40-64	34.5%	26.5%
Over 65	9.4%	11.0%

FINAL RESULTS

- Dashboard with data about the community
  - Available publicly via https://kostler.shinyapps.io/SPL-Seattle-Census-Data/
  - Four interactive maps of census data and SPL branches
- Short report for each region
  - Available datasets found
  - Stats for each branch on topics
- Documented code
  - Available via GitHub repository: https://github.com/OpenDataLiteracy/SPL-KO
  - Easily adapted by other library systems

SUSTAINABILITY

- Open file types and softwares
  - R and Rstudio
    - Preferred programming language for sponsor
    - Good visualization packages
- GitHub Repository
  - All Code
  - Publicly available

- U.S. Census Bureau Data
  - Federally maintained
  - Consistent formatting and data releases
- Dashboard on shinyapps.io
  - Can easily be modified for future years' data
  - Hosted by RStudio

CHALLENGES

- Keeping the project scale manageable, very easy to try and do too much
- First time creating maps, especially interactive maps
- Keeping things simple with visualizations, not overwhelming users

NEXT STEPS FOR SPL

- Distribute the maps and reports for use by staff
- Share maps with SPL Foundation, in annual reports about demographics served by SPL, or even in policy advocacy (e.g. Library Levee that just passed)
- Add other datasets to maps
  - School district data, low-income housing
- Update when 2018 ACS data or 2020 Census data available
- Collect feedback from frontline staff about maps and reports
  - Are they helpful? How could they be improved?

NEXT STEPS FOR ODL

- For this project:
  - Write journal article about project
  - Create Binder instance for R Notebook in GitHub Repo
  - Directed Field Work or Capstone
- In general:
  - Data Science/Coding and Librarians
    - Using Open data and open data software in an everyday workflow