

Predictors of Permanent Housing for Homeless Families

Data Science for Social Good

University of Washington, eScience Institute

August 20, 2015

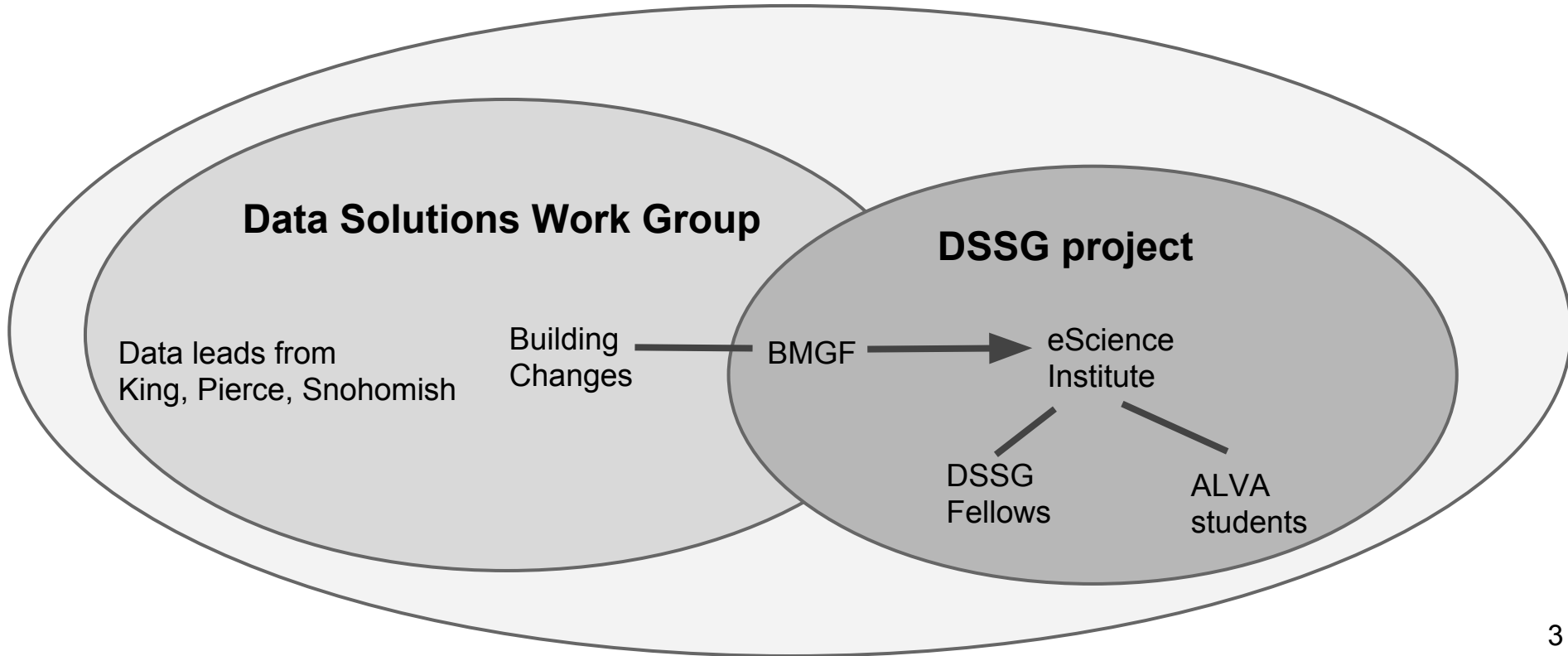


Background

- Over **4,000** homeless families in the Tri-county (Snohomish, King, Pierce) area every year.
- BMGF and Building Changes: cut family homelessness **by half** by 2020.
- Make family homelessness **rare, brief, and one-time**.



Multi-stakeholder Collaboration



ALVA Project:



Xitlalit Sanchez



Cameron Holt

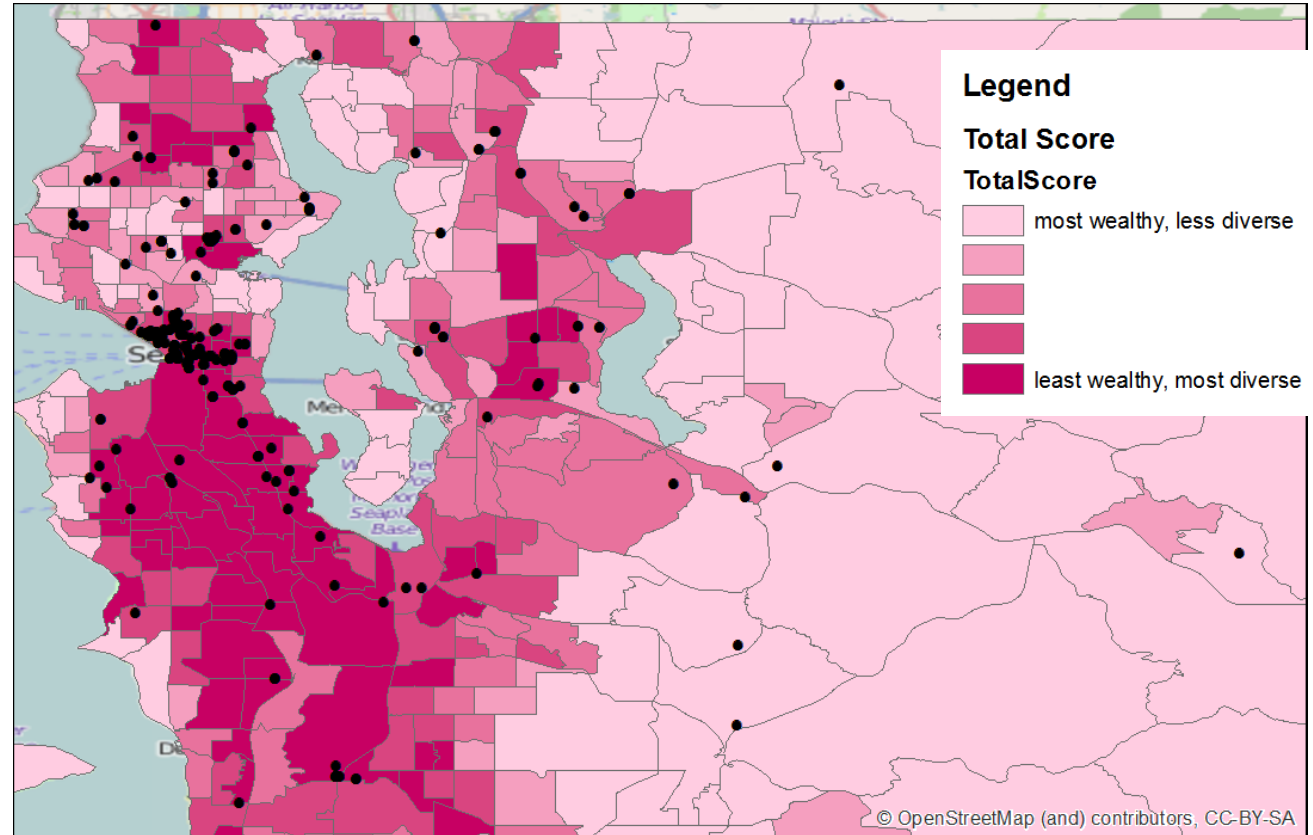
- **DATA:**
 - Census data at the census-tract level (King County GIS Data Portal)
 - Geographic data about homeless shelters and other services (King County)
- **Question:**

How are homelessness services geographically distributed?

 - Relative to population density
 - Relative to wealth, racial diversity, English proficiency
 - Relative health services

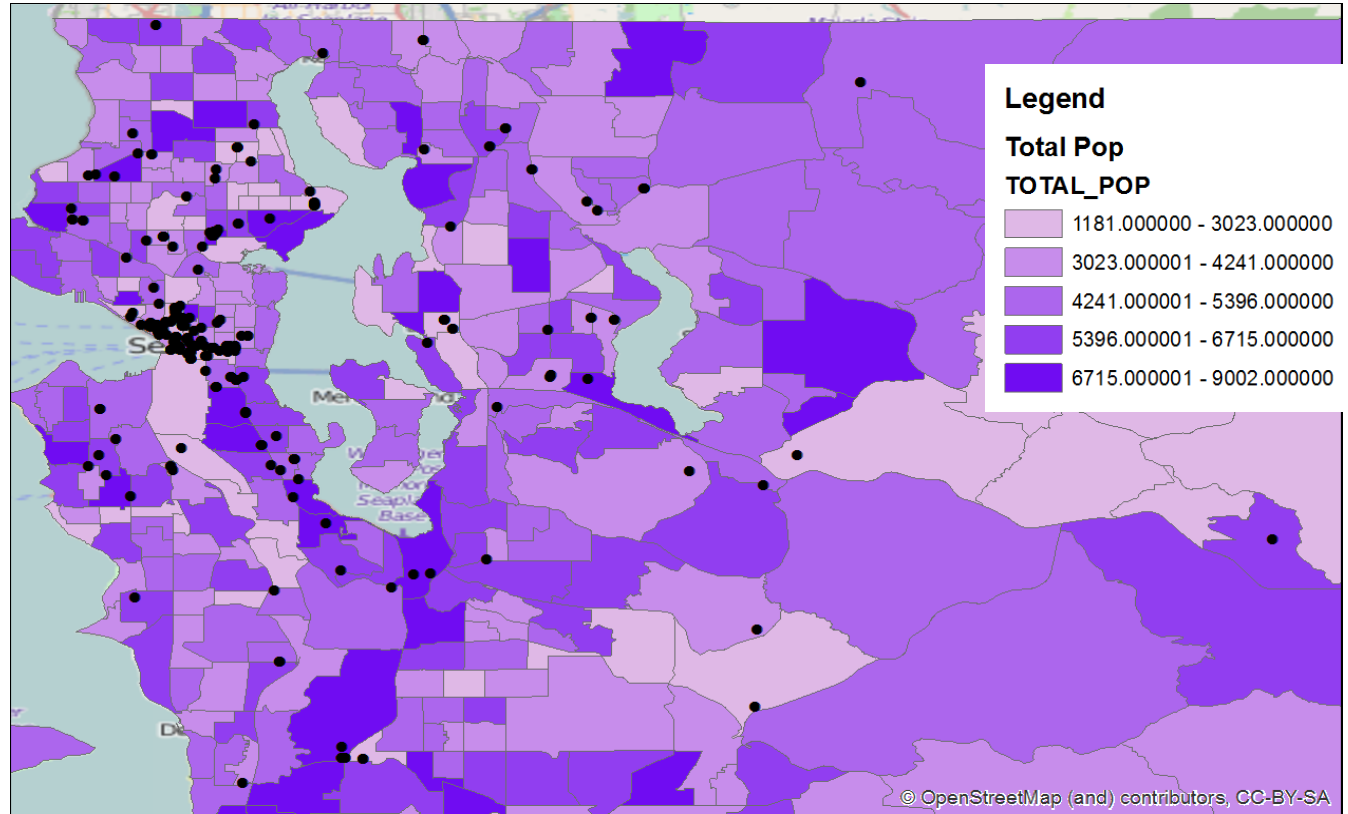
ALVA: demographics in King County

The majority of homeless shelters are located in the least wealthy and most diverse tracts.



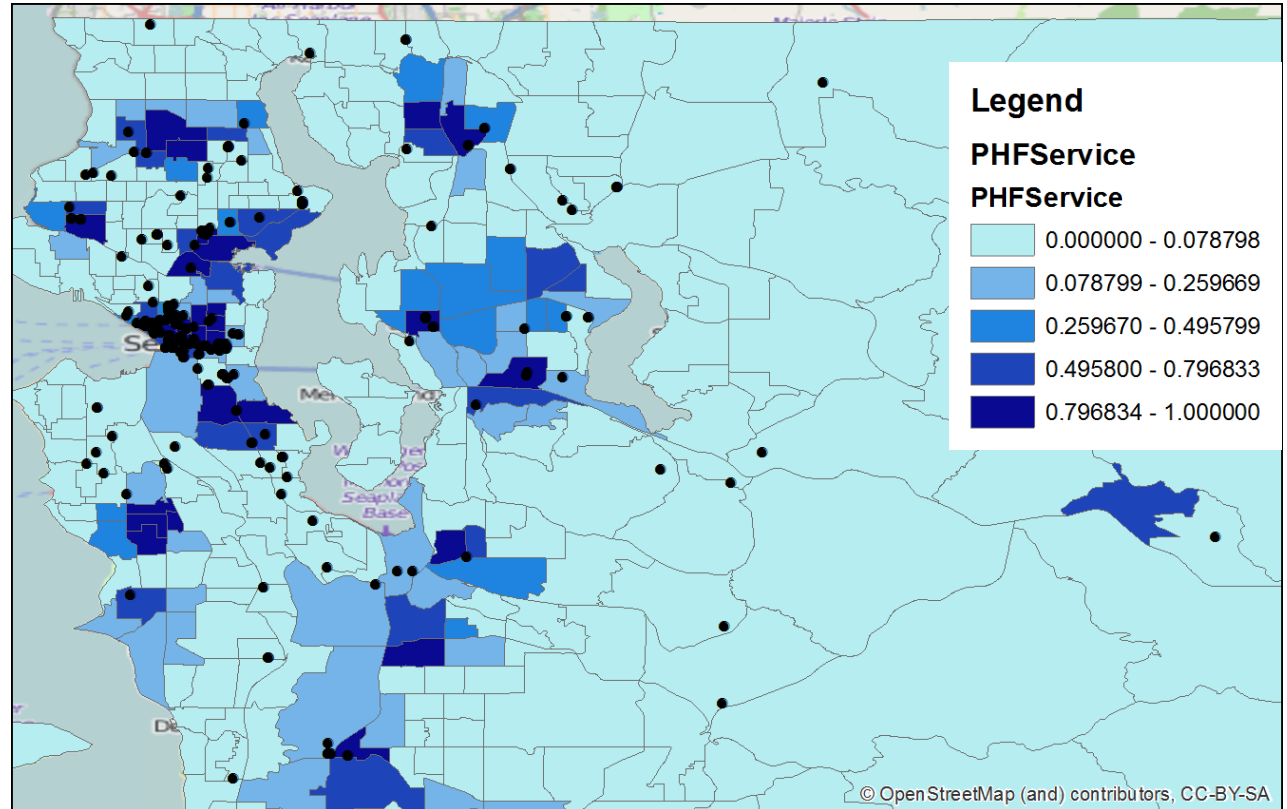
ALVA: shelters in King County

Shelters are **well dispersed** throughout the area, across both populous and less populous tracts



ALVA: Public Health Facilities in King County

A large number of homeless shelters are in close proximity to public health facilities



DSSG Project

Data description:

- Homelessness Management Information System (HMIS)
- Individuals' **enrollments**
 - **Programs:** Emergency Shelter, Transitional Housing, Rapid Re-Housing, etc.
 - **Destination:** Permanent Housing, Temporary Housing, etc.
 - **Demographics:** age, income, disabilities, etc.

Time frame: (enrolled during these times)

- King (2011 - 2014)
- Pierce (2010 - mid-2015)
- Snohomish (2010 - mid-2015)

Research Questions

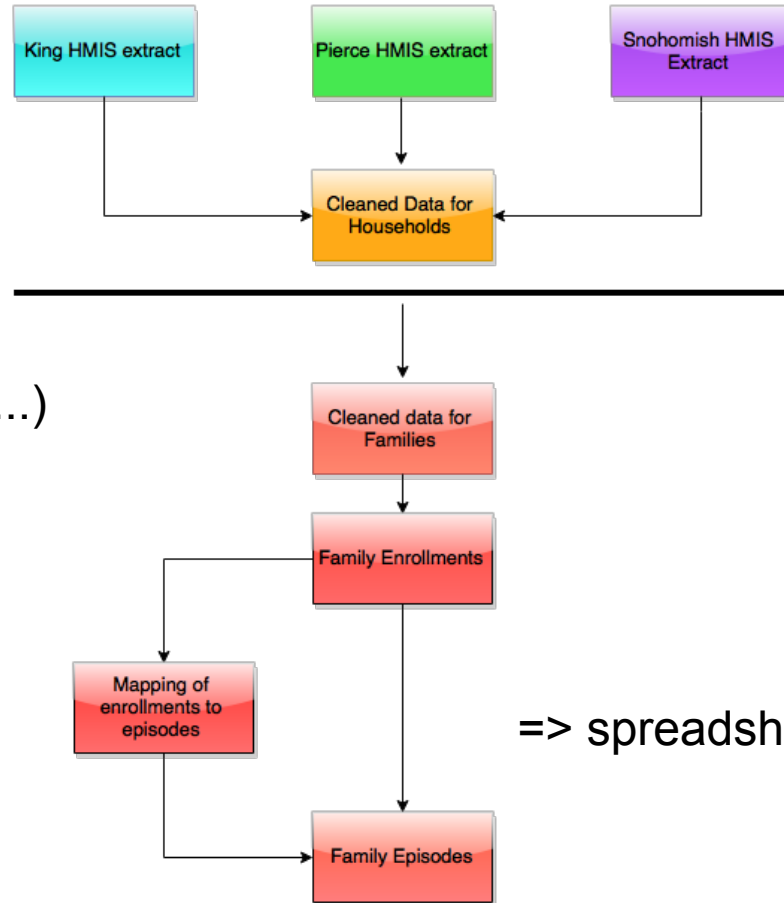
- How do families move through programs before exiting?
- What are the barriers and facilitators for families finding permanent housing?
- What factors increase or decrease a family's length of stay in a homeless shelter or program?

Data Caveats

- **No data** on people who were not able to enroll
- **No data** on people after they exit the HMIS system
- **Data excluded** if people have not exited
- **Data excluded** if there were ambiguous coding errors

Most importantly... data from each county were entered, represented, and extracted differently

Data Processing Pipeline



Python (Pandas, Numpy,...)

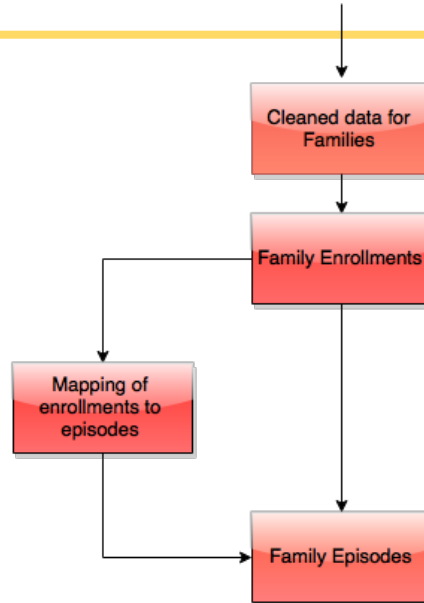
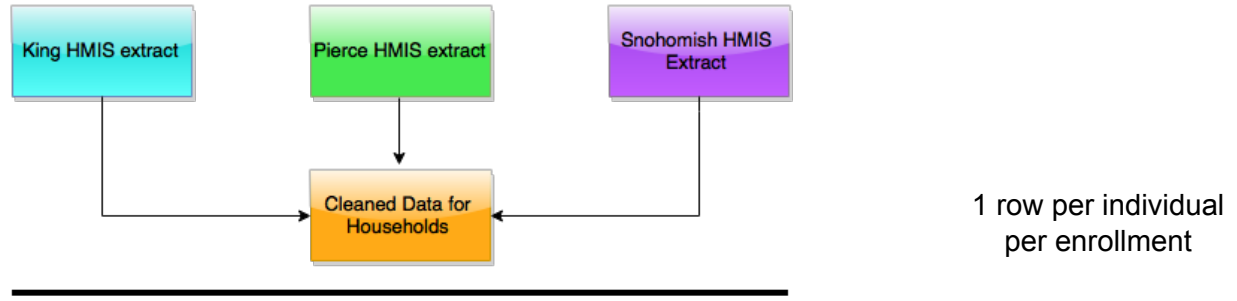
=> spreadsheets (CSV, Tableau, etc.)

Data Processing Pipeline

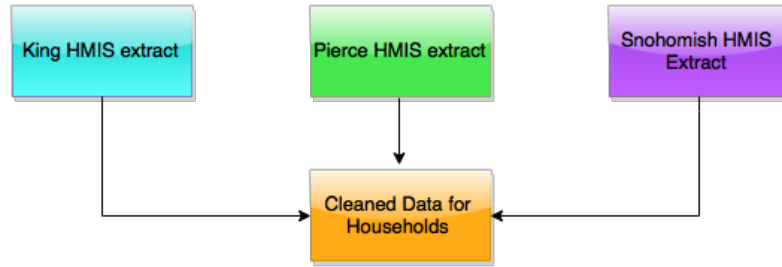
(1) Cleaning

Aligning data across all counties

Identify households over time
(e.g. clustering)



Data Processing Pipeline

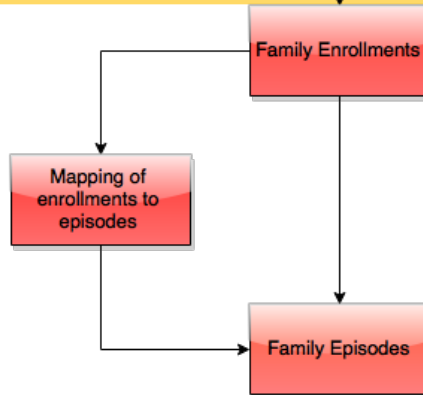
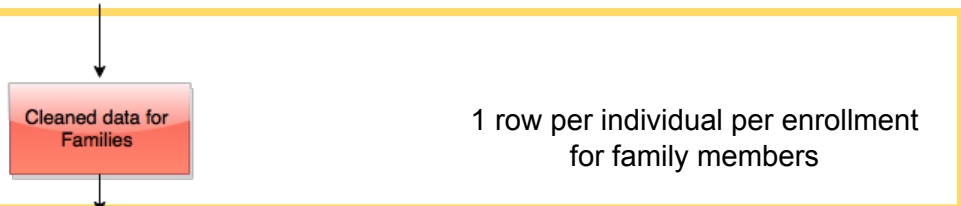


(2) Extracting Families

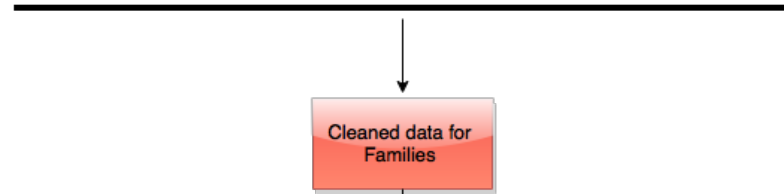
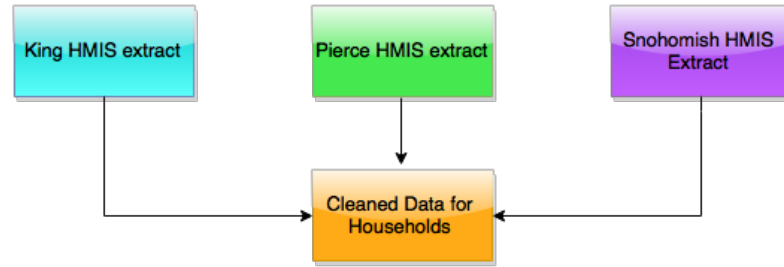
Definition of "family":

- **at least 1 adult and at least 1 child (cutoff 18 yrs)**

Drop non-families

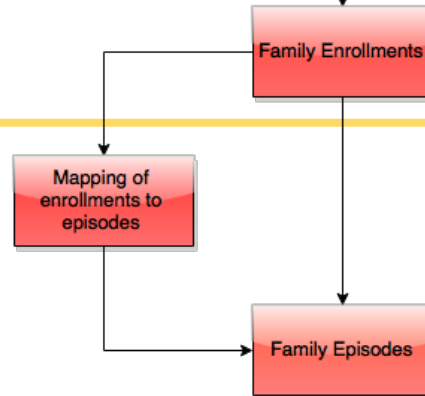


Data Processing Pipeline



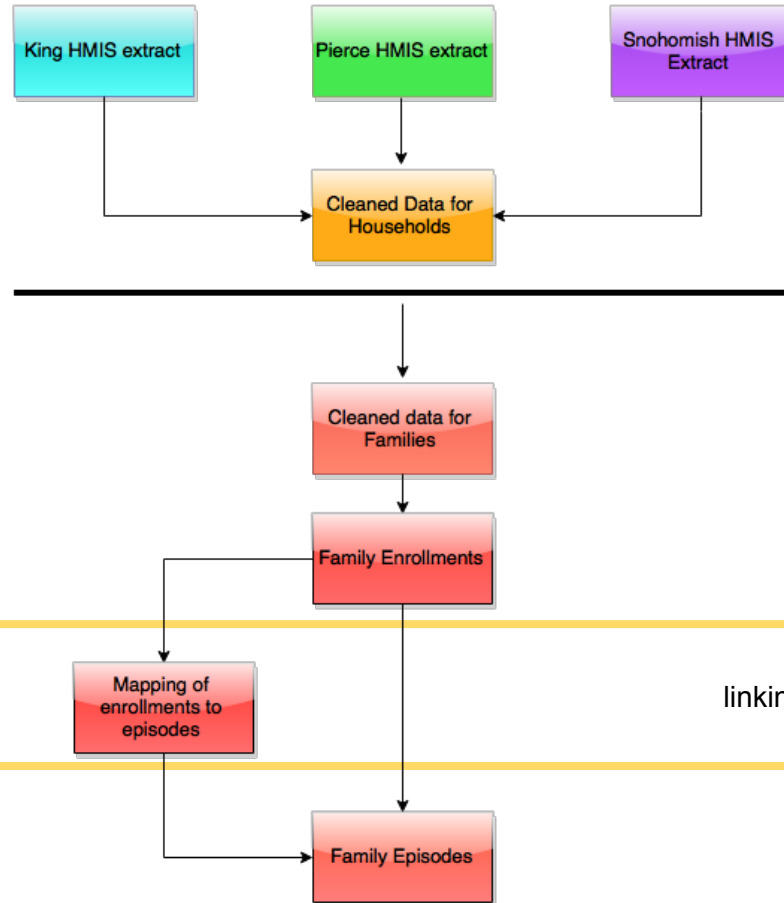
(3) Aggregating into families

Create family aggregate variables
(income, # adults, # children, etc)



1 row per family per enrollment

Data Processing Pipeline

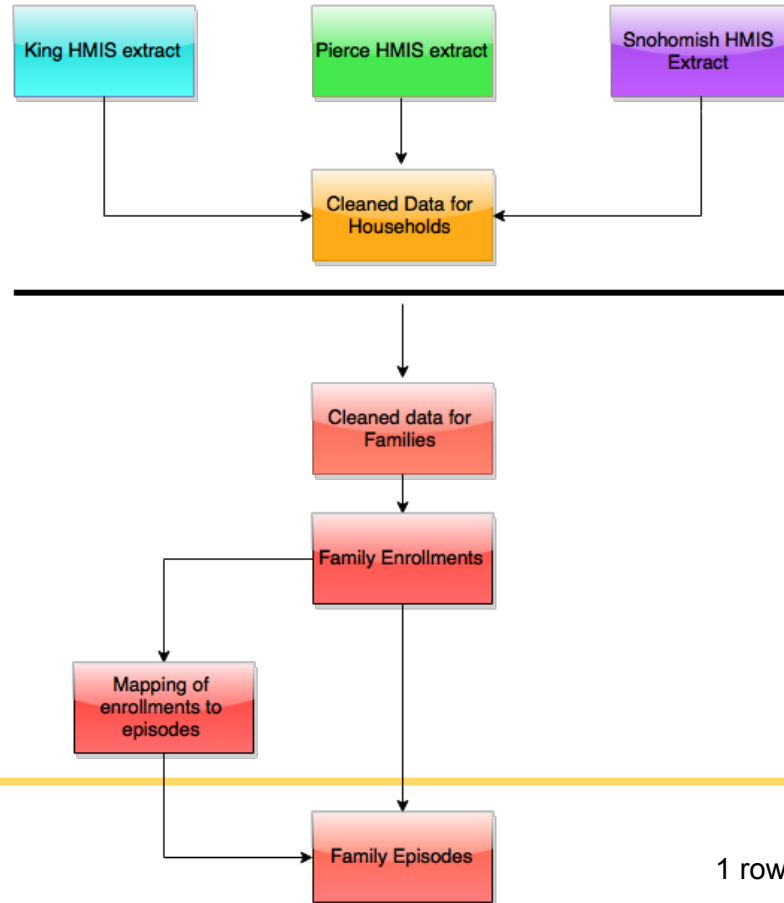


(4) Identifying episodes

Enrollments with exit & entry separated by ≤ 30 days

linking table between enrollments and episodes

Data Processing Pipeline



(5) Aggregating into episodes
Create family/episode aggregate variables

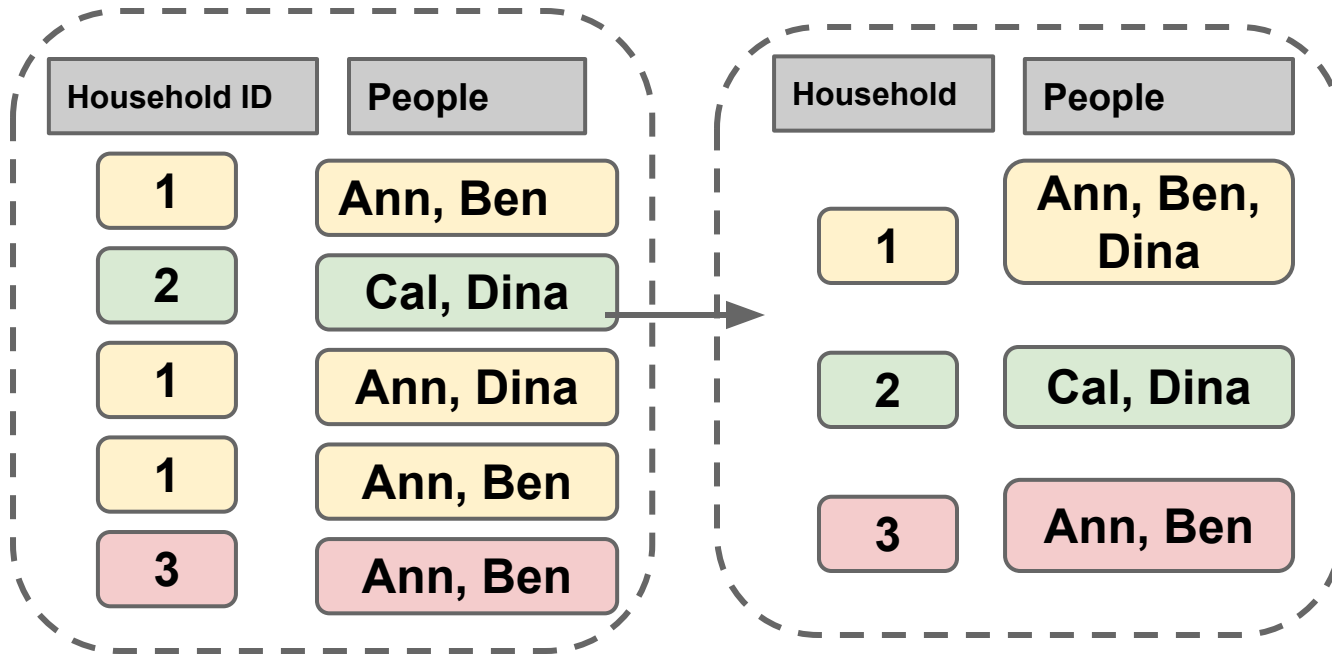
1 row per family per episode

Data cleaning : defining households

Data in HMIS provided as individuals and their enrollments

Problem: how do we know which individuals belong to the same household?

Clustering : group information



| | A | B | C | D |
|---|----------|----------|----------|---|
| A | - | 0.5 | ∞ | 1 |
| B | 0.5 | - | ∞ | 1 |
| C | ∞ | ∞ | - | 1 |
| D | 1 | 1 | 1 | - |

**Group-based
Distance Table**

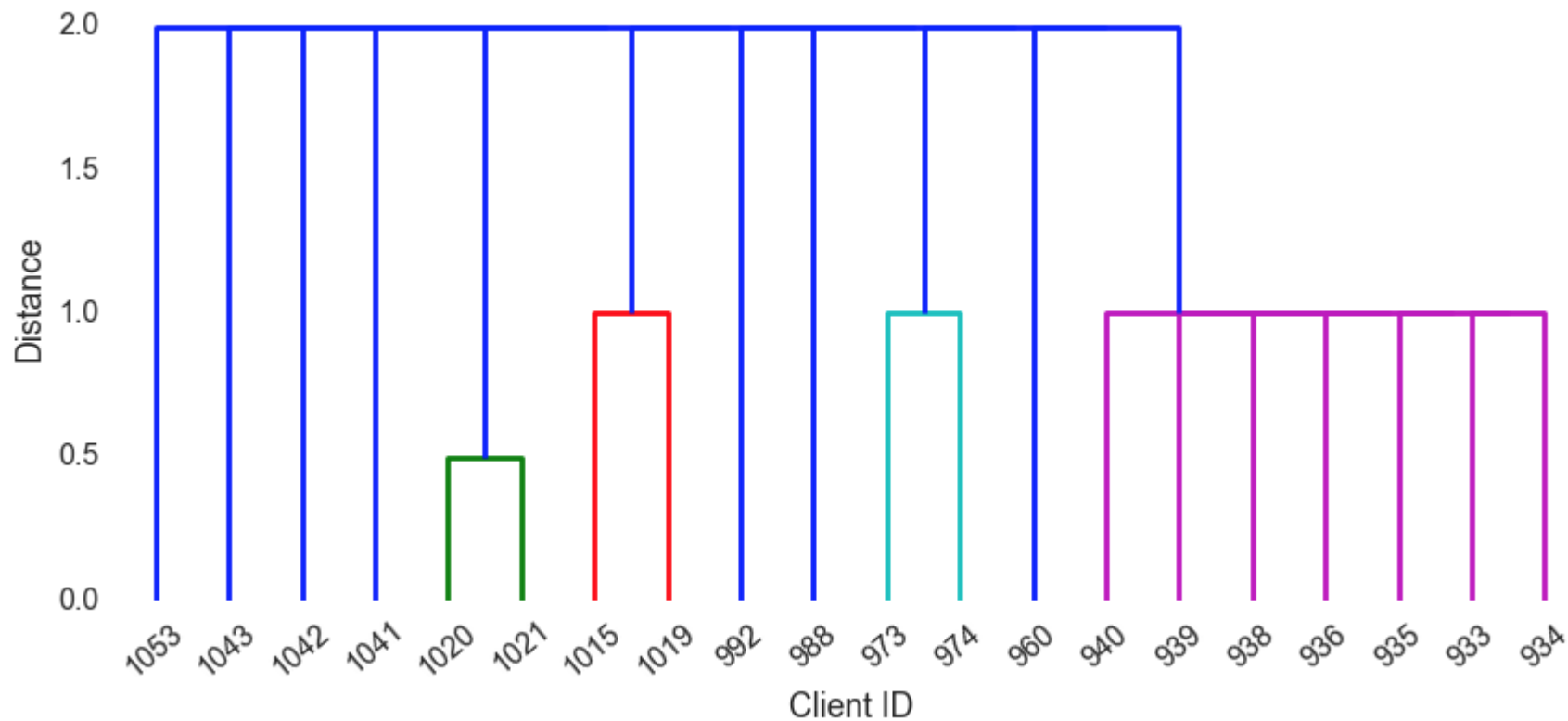
Clustering : temporal information

| PERSON | ENTRY TIME | EXIT TIME |
|--------|------------|-----------|
| Ann | 1 | 5 |
| Ann | 7 | 12 |
| <hr/> | | |
| Ben | 7 | 15 |
| Ben | 1 | 5 |
| Cal | 1 | 3 |
| Cal | 1 | 6 |
| Cal | 7 | 12 |
| Dina | 1 | 6 |

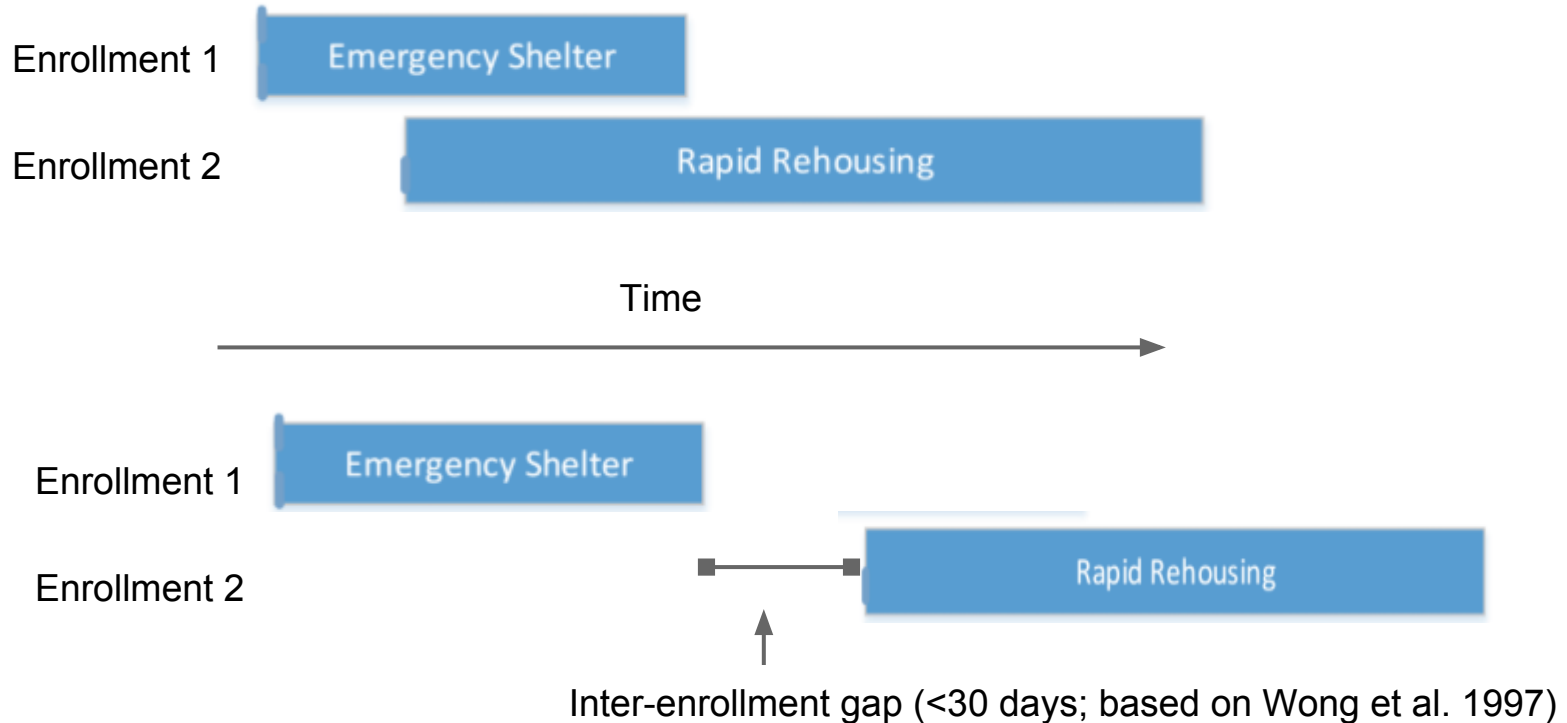
| | A | B | C | D |
|---|----------|----------|----------|----------|
| A | - | 1 | 1 | ∞ |
| B | 1 | - | ∞ | ∞ |
| C | 1 | ∞ | - | 1 |
| D | ∞ | ∞ | 1 | - |

Time-based
Distance Table

Define households



Defining episodes

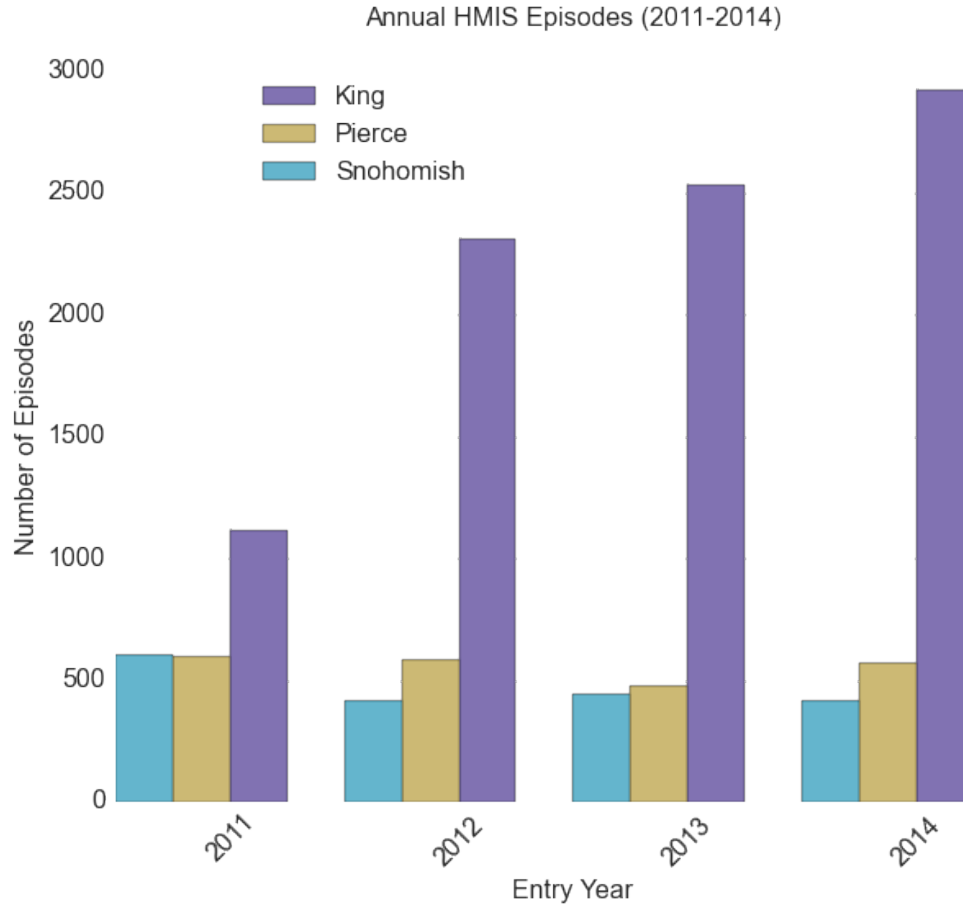


Summary Statistics of Families

| | Total Population in 2013 | Number of Families in Dataset | Average Number of Enrollments/Episode | Average Number of Children | Average Number of Parents |
|------------------|---------------------------------|--------------------------------------|--|-----------------------------------|----------------------------------|
| King | 2,044,000 | 8,681 | 1.21 | 2.02 | 1.20 |
| Pierce | 819,743 | 3,038 | 1.23 | 2.31 | 1.75* |
| Snohomish | 745,913 | 2,707 | 1.25 | 1.89 | 1.25 |

*Represents number of adults (rather than parents)

Annual HMIS Episodes

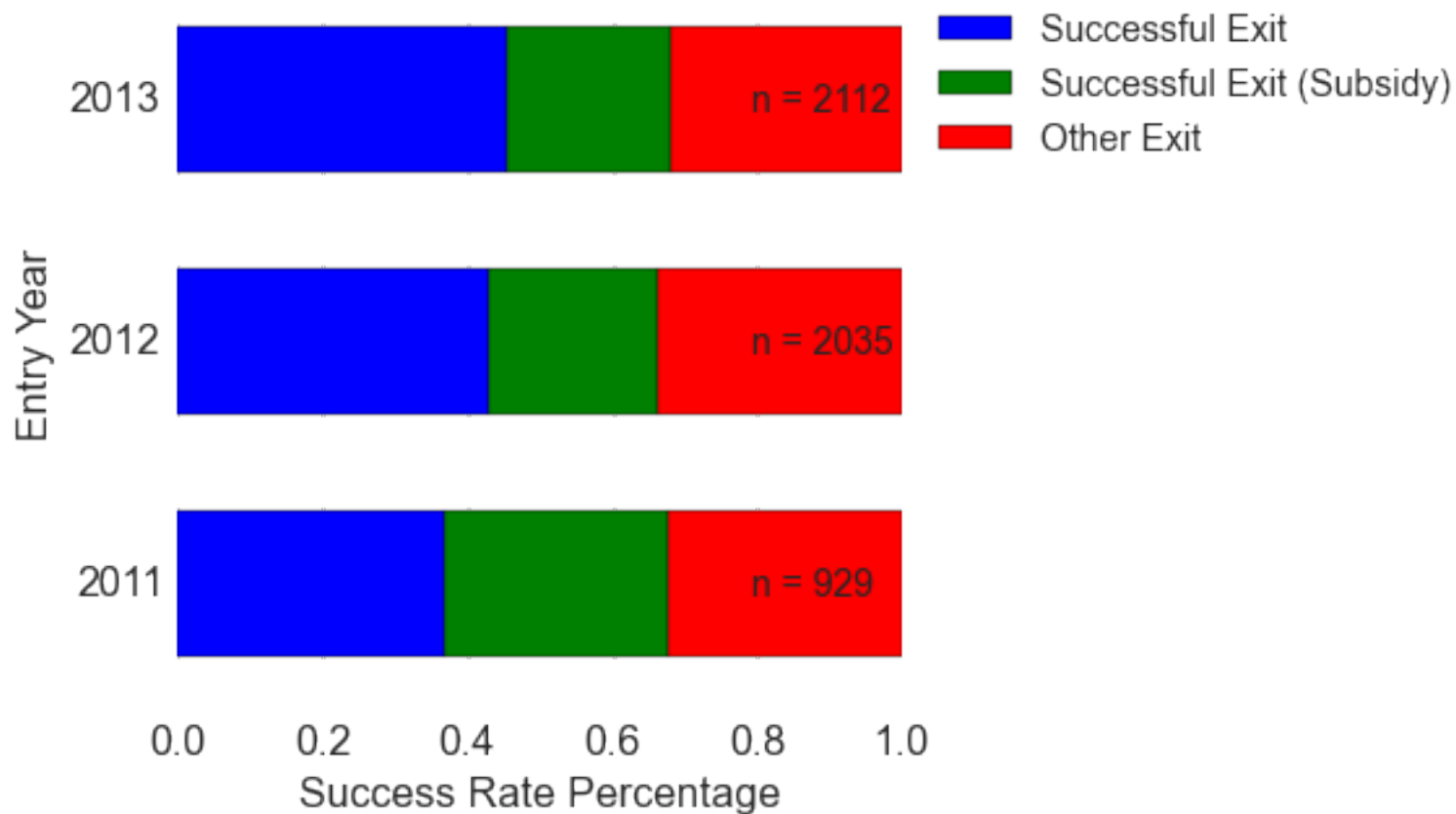


Variations in episode volume across time & programs are based on **supply**, not demand.

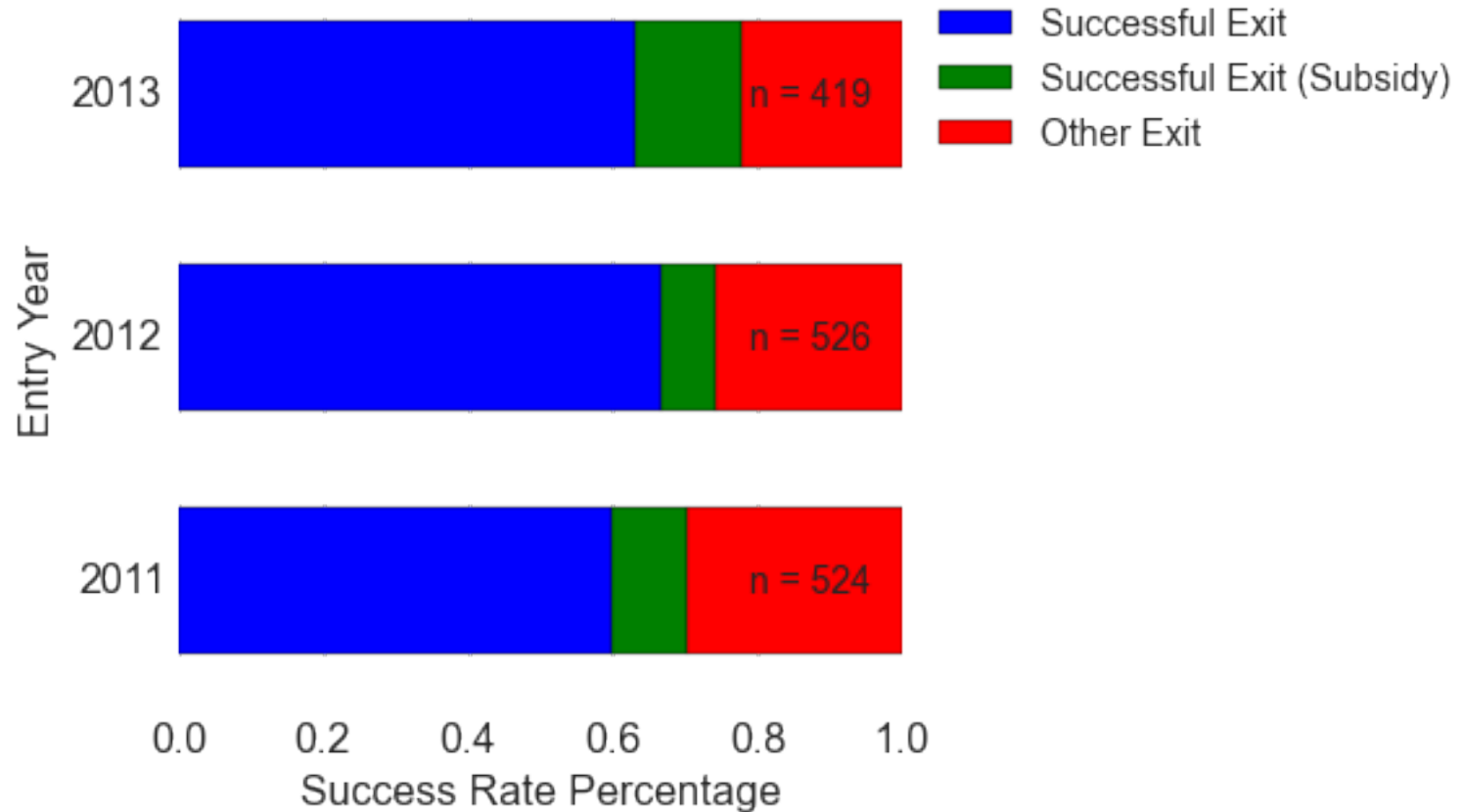
Defining Successful Exit

- **Successful Exit (no subsidy)**
 - Finding permanent housing with no ongoing support
- **Successful Exit (with subsidy)**
 - Finding permanent housing with ongoing support
- **Other Exit**
 - Exited HMIS system without permanent housing
 - Temporarily living with family or friends, living in an unsafe environment, jail, unknown

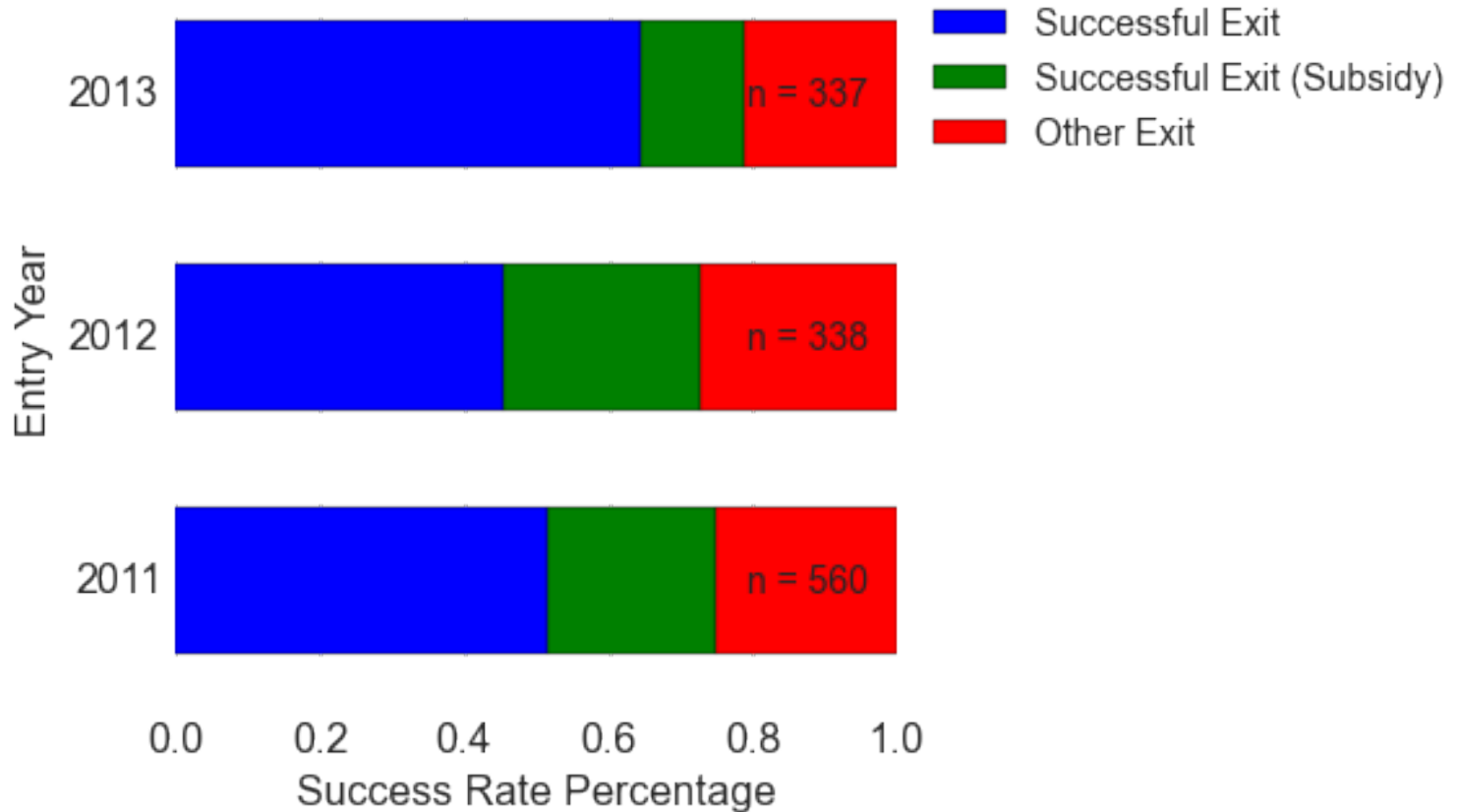
King



Pierce



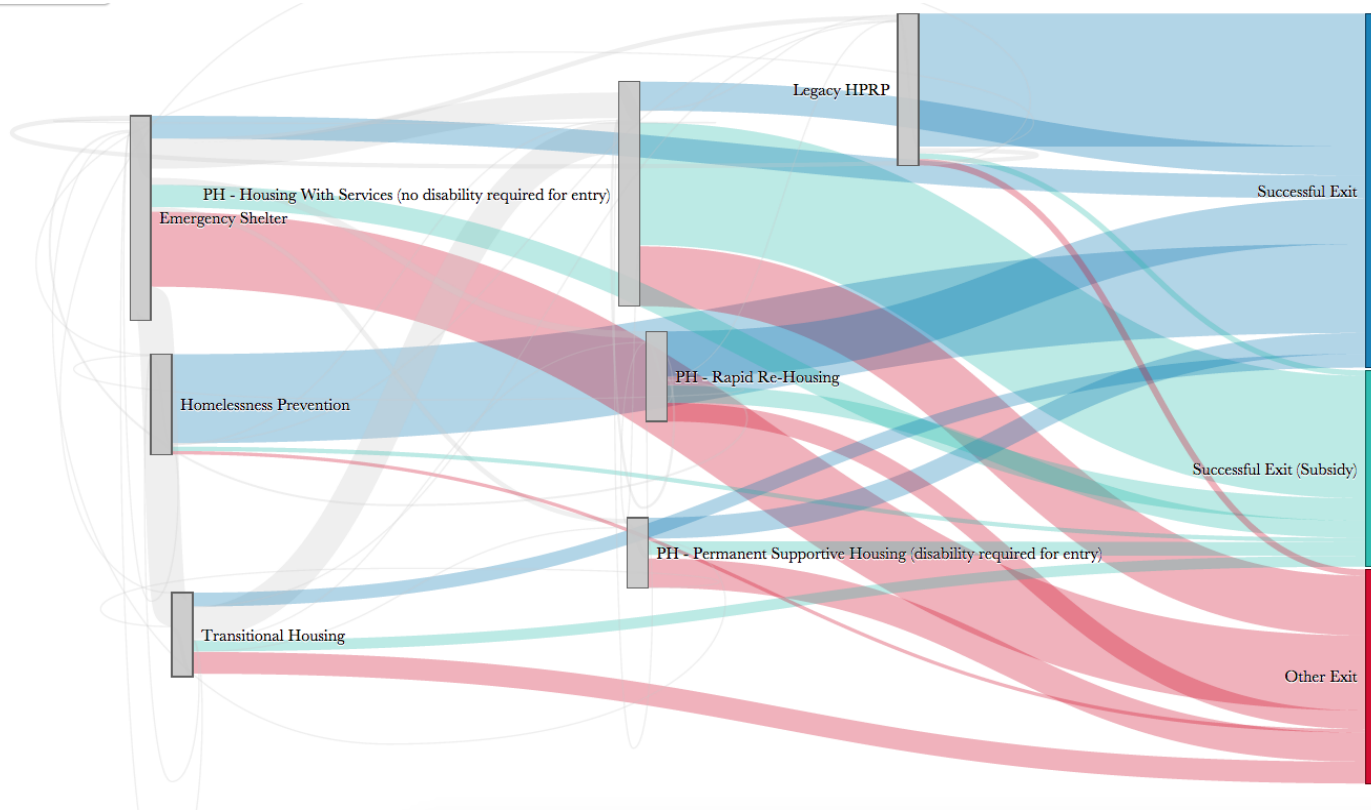
Snohomish



How do families move through programs?

Sankey Diagrams

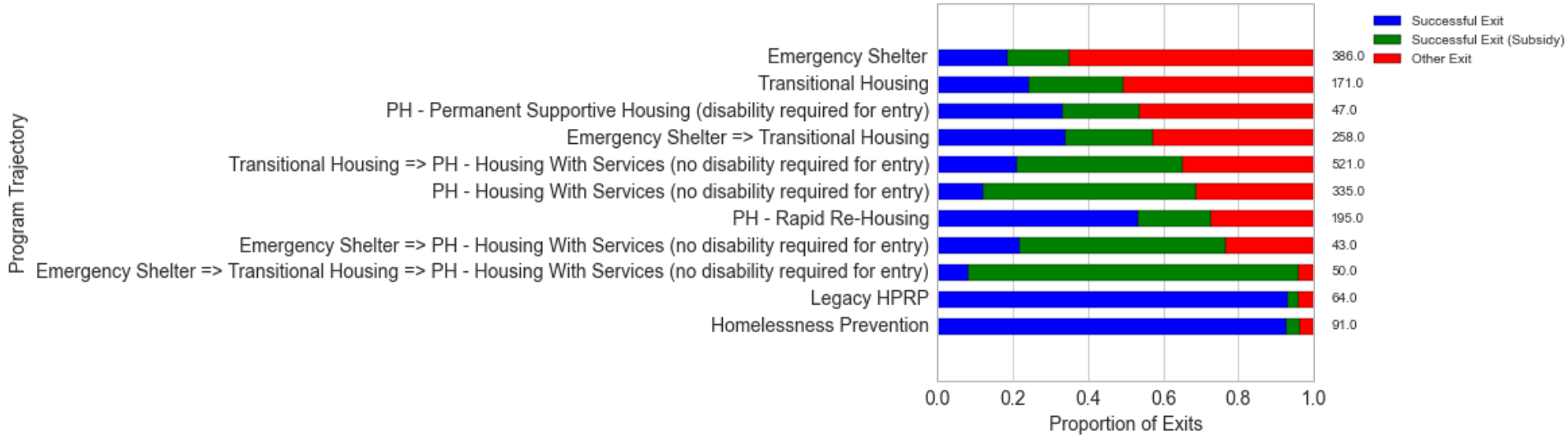
<http://tinyurl.com/dssg-homeless>



Trajectories

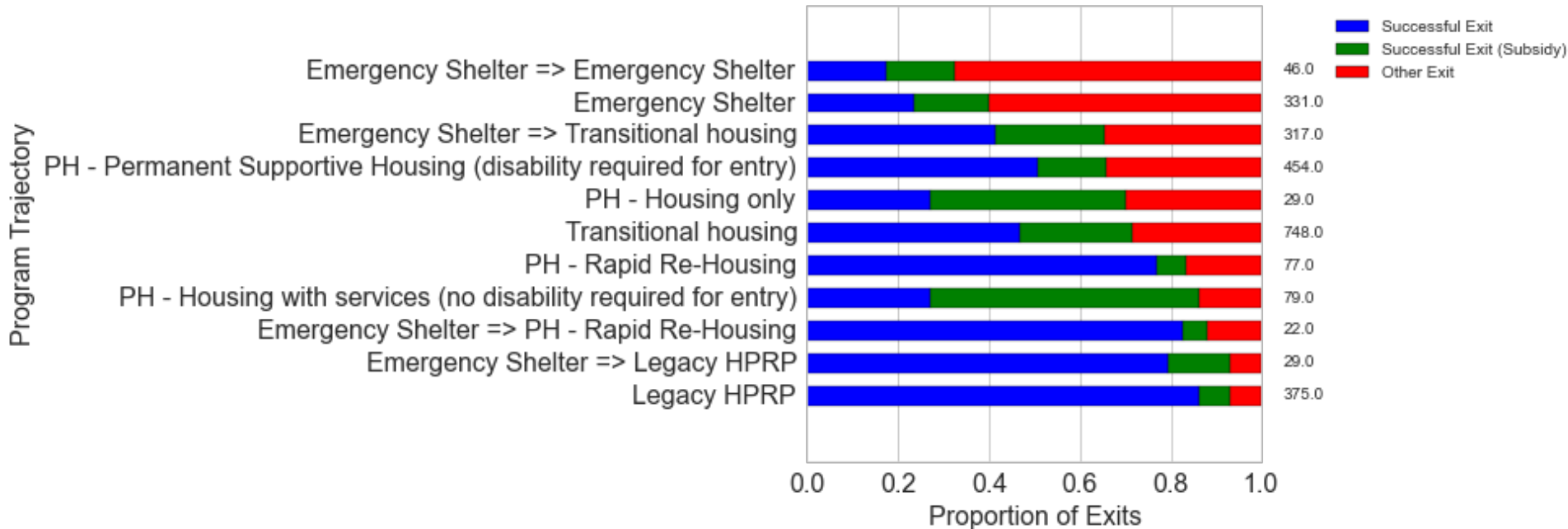
- How do families transition through the programs?
- How do the full paths affect their outcomes?

Snohomish



Note: some programs are targeted at different populations: comparing 'success' rates is problematic!

Pierce



Note: some programs are targeted at different populations: comparing 'success' rates is problematic!

Summary and conclusions

- Our analysis can help domain experts identify programs/trajectories that help homeless families find permanent housing.
- Can help address data quality issues
- Important caveats:
 - This analysis did not (so far) take into account demographic characteristics
 - Programs/demographics interact!

Future Directions

- Further statistical analysis:
 - Logistic regression (Following Rocha et al.)
 - Decision trees
- Other types of data:
 - Coordinated entry system
 - Location and time specific data (e.g. about unemployment, housing stock, etc.)

Thank you!

- ALVA: Xitlalit Sanchez, Cameron Holt (thanks also to Anthony Arendt, eScience, and Kia Guarino and Io Blair-Freese, BMGF).
- BMGF: Neil Roche, Anjana Sundaram, Juan Sanchez, Anne Martens.
- Building Changes: Mei Ling Ellis, Christena Coutsubos
- County Data Leads:
 - Snohomish: Jess Jorstad, Alex Vallandry
 - King: Amanda Thompkins, Ann Ku
 - Pierce: Valerie Pettit, Geoffrey Champion