



HIV/AIDS:

A Comprehensive National and State-Wide Analysis

W209 - Final Project Progress Report
Ramya Girish, Chris Beecroft, Tim Alt



Background and Use Cases

- ▶ First case discovered in 1981
- ▶ Affects marginalized demographics in larger numbers
- ▶ **Richest data is collected at the local, grassroots level**
- ▶ **No comprehensive national, top-down approach exists**

Our Users

Who benefits from this information?

1. Insurance agencies
2. Pharmaceutical Companies
3. Local HIV/LGBT Agencies with less resources
 - Is coverage provided for the most impacted?
 - Where is coverage lacking?
 - What are the trends of new infections?
 - What are other states doing to combat new infections, death rates and reduce prevalence?

Our Data and Technology

- Original data is from AIDSvu
- Distilled from 29 Excel files to 5 data sets across 3 areas

New Infections

(State & County data)

Description, years,
included
info/demographics

AIDS Related Deaths

(State data)

Description, years, included
info/demographics

Prevalence

(State & County data)

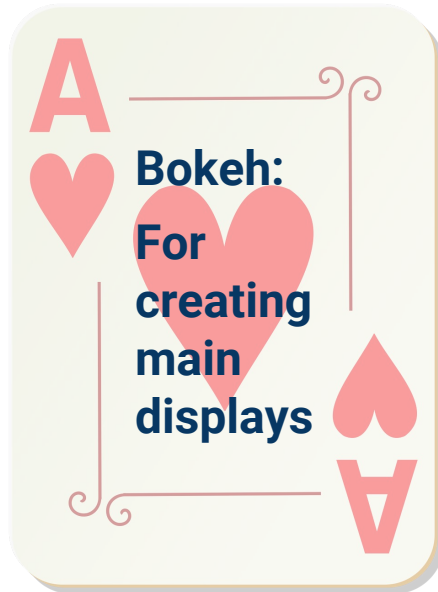
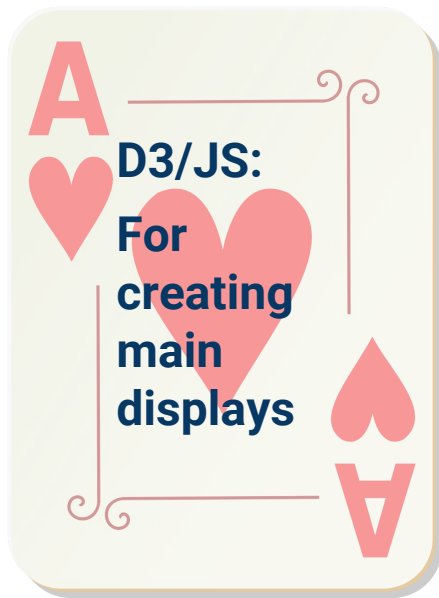
Description, years,
included
info/demographics

Our Data and Technology

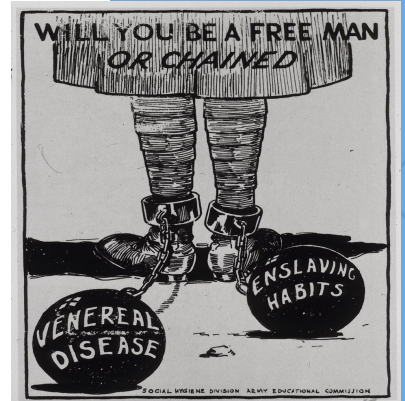
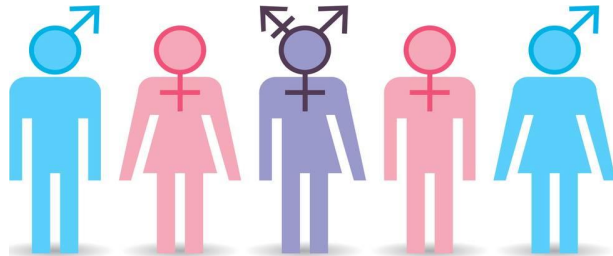
Data Cleanup

- Numbers inconsistently encoded
 - Numbers
 - Quoted strings with commas
- Different character encodings were used
 - (Latin1, MacRoman/Windows)
 - Data reduced to 7-bit ASCII
- Formats changed between years
 - Needed to Line up columns
 - Produce files with only common columns

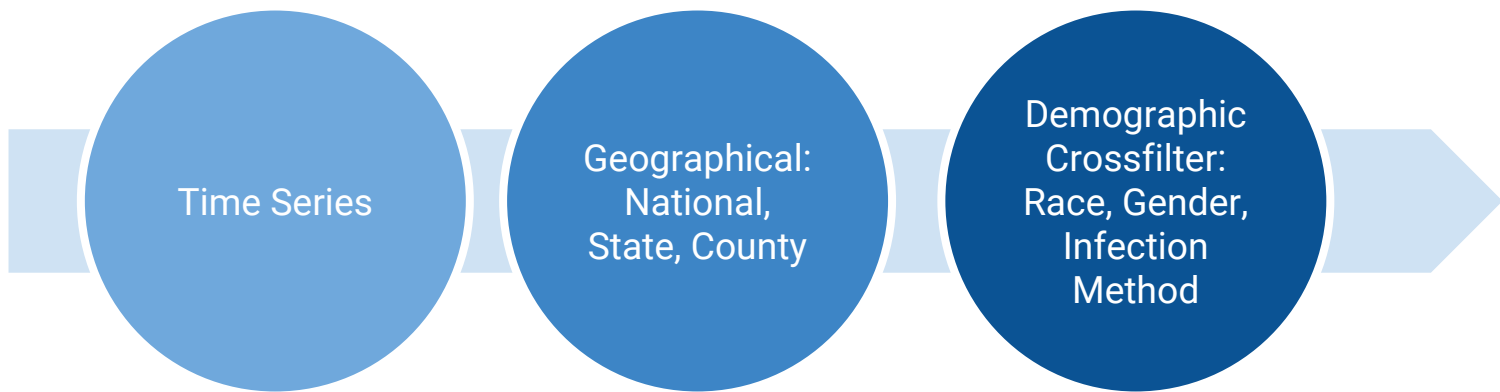
Our Data and Technology



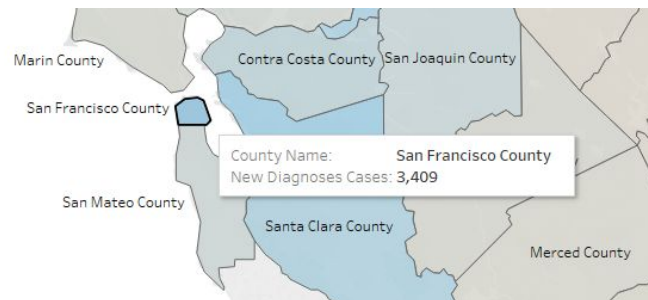
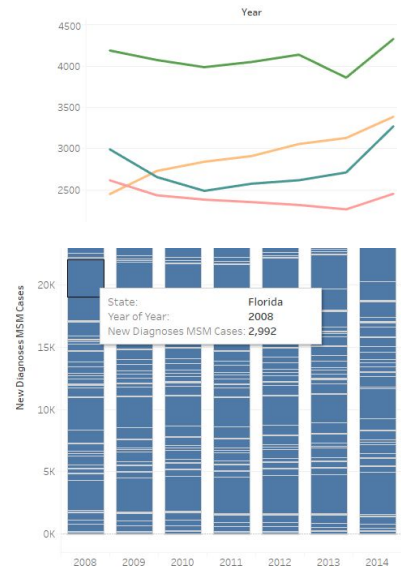
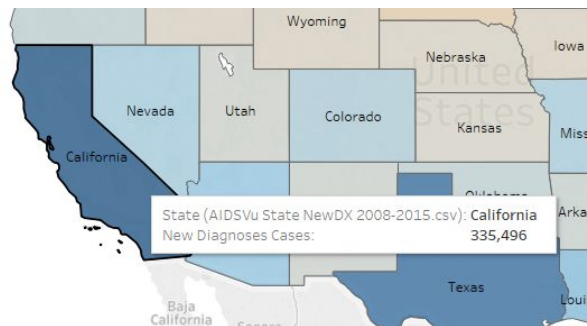
Our Story



Our Tasks



Implementation and Mock-ups



Questions and Feedback

For feedback, please forward to:

Tim

- timothymalt@berkeley.edu