Patterns of Social Vulnerability
An Interactive Dashboard to Explore Risks to Public Health on the US County Level

Darius Coelho[1], Nikita Gupta[2], Eric Papenhausen[1], Klaus Mueller[1,2]

[1] Akai Kaeru LLC
[2] Stony Brook University Computer Science

Preliminaries

Our software looks for **patterns** in high-dimensional data

A **pattern** is a subgroup of data points that share similar characteristics or features

For e.g. a set of similar US counties associated with a COVID-19 death rate that is higher on average than the US county average

Patterns are automatically mined via subgroup discovery

Let's look at an **example**

First: The Data

Social Vulnerability Index (SVI) factors for 3,006 US counties
- compiled by the US Center for Disease Control (CDC)

Grouped into 4 related themes of 15 factors:
- **Socioeconomic status:** below poverty, unemployed, income, no high school diploma
- **Household composition & disability:** age 65 or older, age 17 or younger, older than age 5 with a disability, single-parent households
- **Minority & language:** minority, speak English "less than well"
- **Housing type & transportation:** multi-unit structures, mobile homes, crowding, no vehicle, group quarters

The Data – Extended

A dataset from Kaggle refines these measures to

- **Demographic features:** race, gender, and more
- **Further health & social risks:** smoking & drug use habits, teenage pregnancies, sleep deprivedness, housing debt, vaccination rate, and many more

Data was collected from 200 publicly available COVID-19 related datasets, using sources like Johns Hopkins, the WHO, World Bank, the New York Times, and many others

We found several redundancies in this dataset and carefully trimmed it to the set of **241 factors**

A Pattern with 3 Factors (Factor 1 Only)

Target variable **Death rate** (# of COVID-19 deaths per # of county residents, prob. log scale)

two US counties with a very high **death rate**
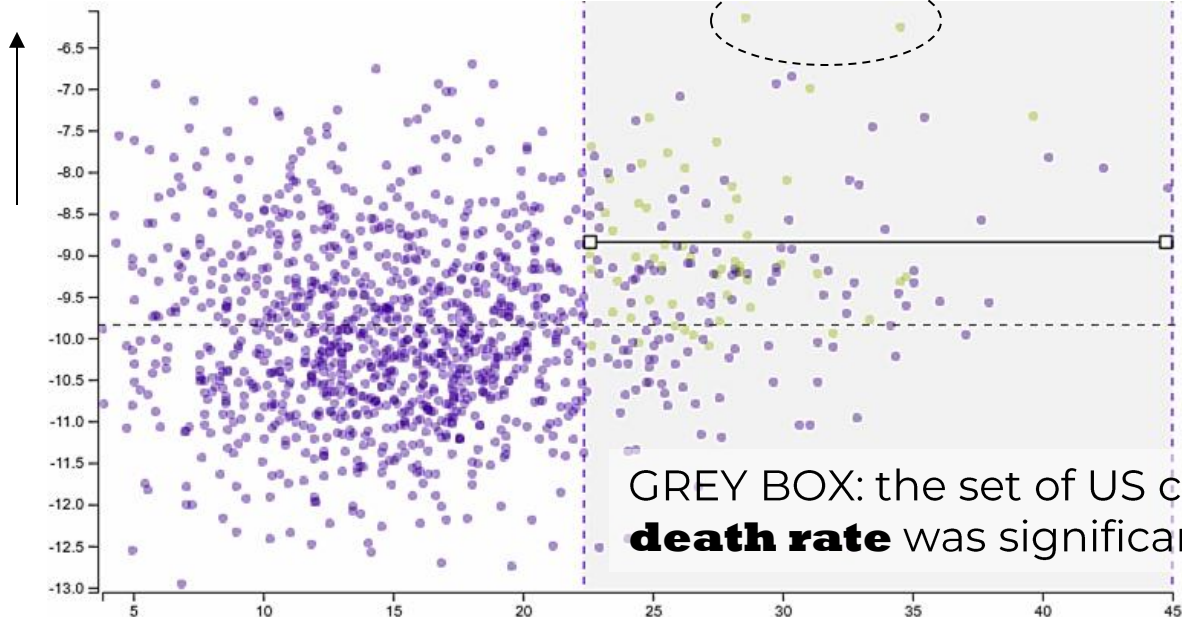
death rate for the grey box's set of US counties

death rate averaged over all US counties
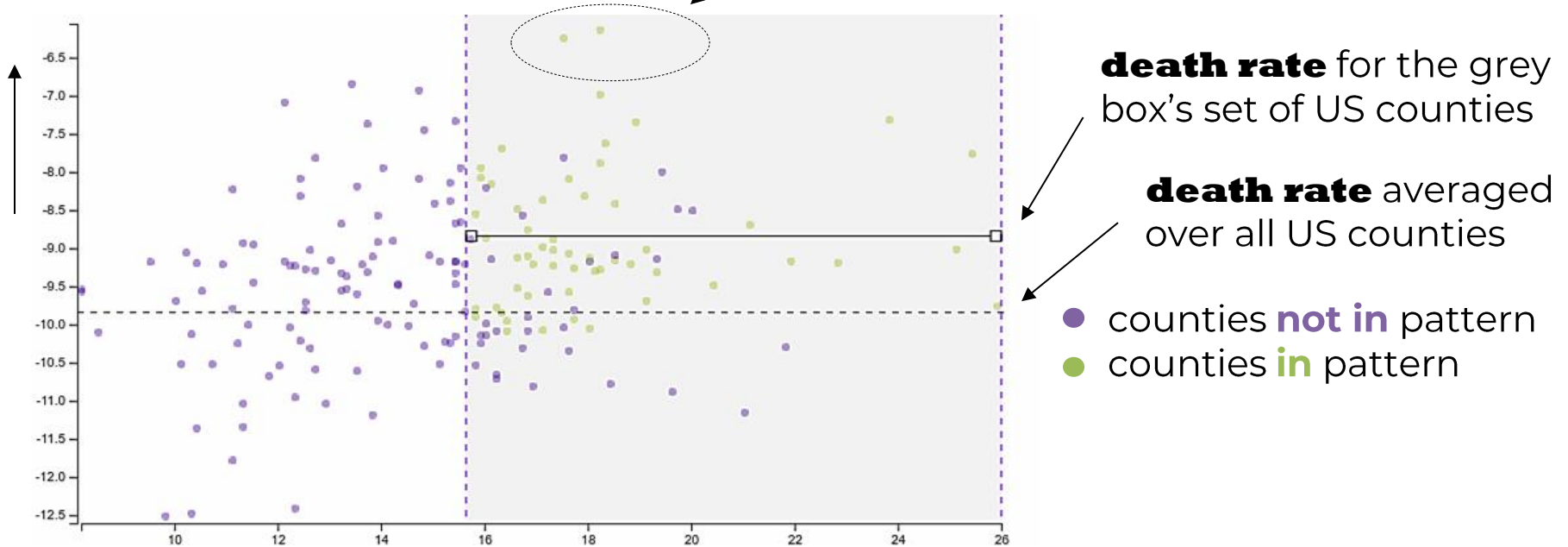
● counties **not in** pattern
● counties **in** pattern

GREY BOX: the set of US counties for which the COVID-19 **death rate** was significantly higher than average

Pattern factor #1: **County poverty rate**

A Pattern with 3 Factors (Add Factor 2)
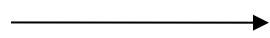
Target variable **Death rate**

two US counties with a
very high **death rate**

death rate for the grey
box's set of US counties

death rate averaged
over all US counties

● counties **not in** pattern
● counties **in** pattern

Adding pattern factor #2: **% Population aged over 65**
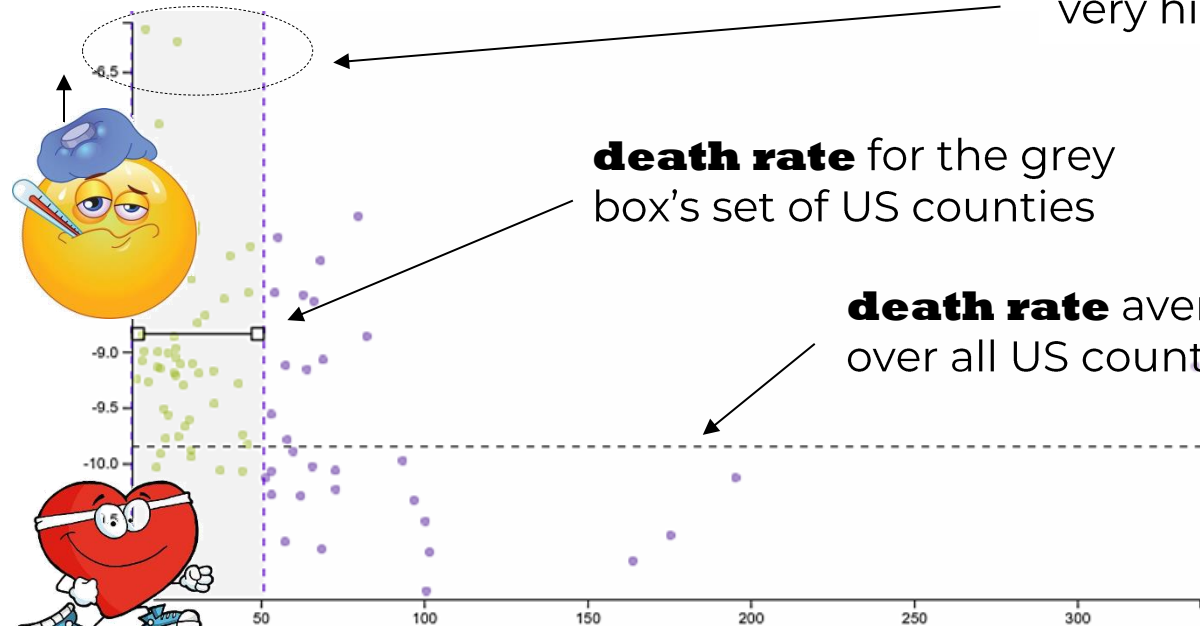
A Pattern with 3 Factors (Add Factor 3)

Target variable **Death rate**

two US counties with a
very high **death rate**

- ● counties **not in** pattern
- ● counties **in** pattern

death rate for the grey
box's set of US counties

death rate averaged
over all US counties

Adding pattern factor #3: **Population density**

Summary

Adding the third factor fully separates **in-pattern counties** from **out-pattern counties**

- Mann-Whitney and $\chi 2$ tests are used for statistical significance

This particular pattern describes counties with

- high poverty rate **AND**
- high percentage of older people **AND**
- low population density

These are **joint conditions** making a county more vulnerable

- in this case we tested for higher-than-usual COVID-19 death rate
- other risks could also be tested for, given data

Where are These Counties?
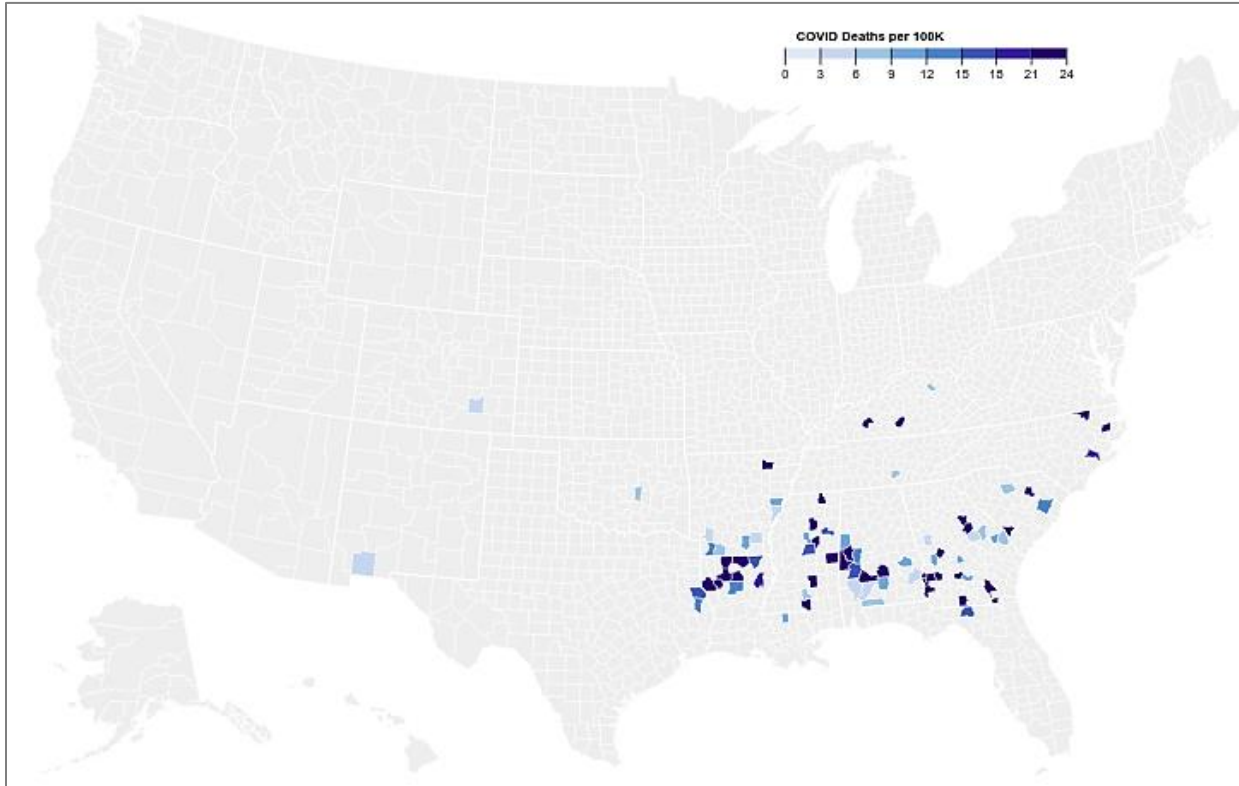
A Few Other Patterns

In May 2020 we found **297 patterns** of US counties
- each consisting of 1 or 2 or 3 factors
- easy to describe
- all tell an intuitive story that explains the vulnerability

Some more examples – counties with a…
- sleep-deprived, uninsured AND uneducated population
- low % Asian, minority AND poor black youth population
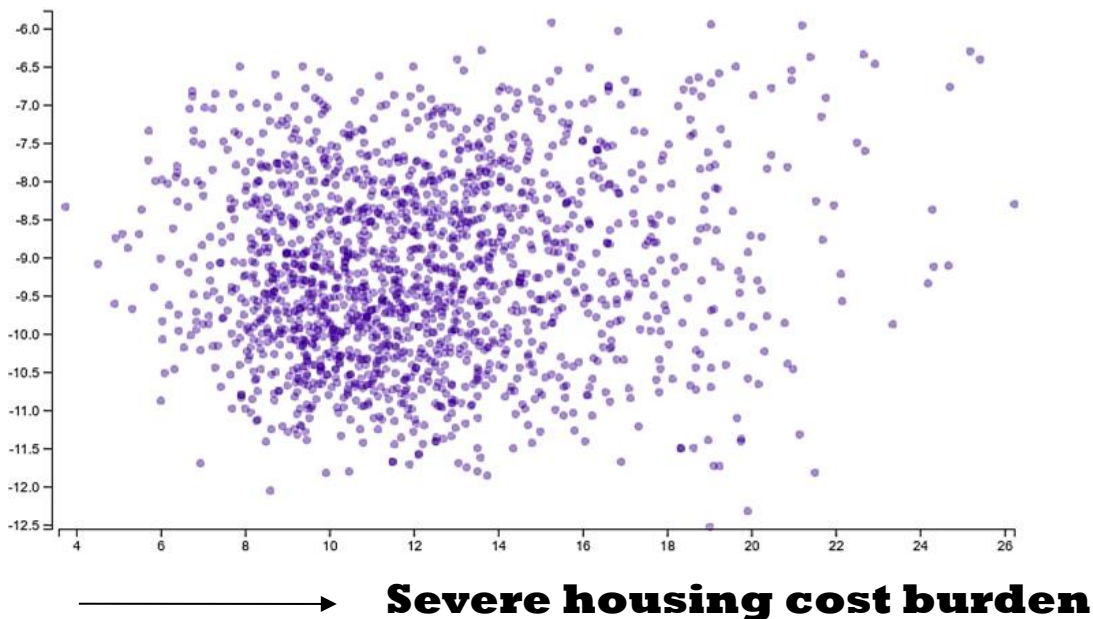
are at elevated risk

Correlation Patterns (1)

We also found interesting correlations within some patterns

Target variable **Death rate**
(# of COVID-19 deaths per # of county residents, prob. log scale)

Pearson correlation = **0.1**
- practically non-existent
- so there is **no correlation** between housing debt and COVID-19 death rate?
- standard methods would imply this
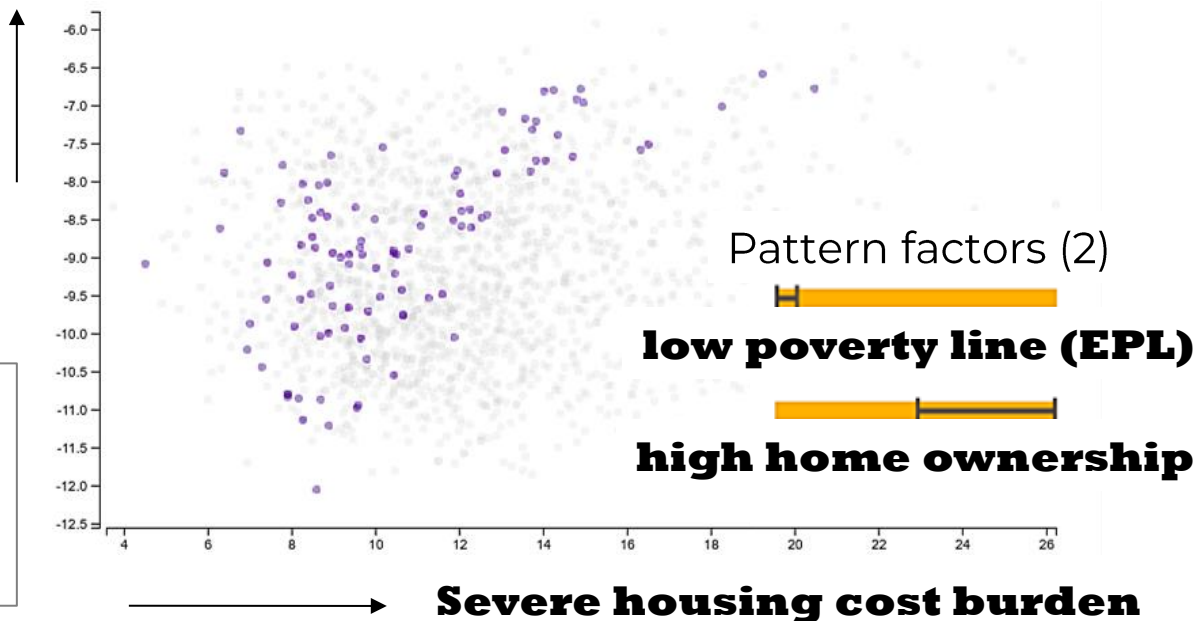

Severe housing cost burden

Correlation Patterns (2)

We also found interesting correlations within some patterns

Target variable **Death rate**
(# of COVID-19 deaths per # of county residents, prob. log scale)

Pattern factors (2)

low poverty line (EPL)

high home ownership

Severe housing cost burden

Pears. correlation = **0.62**

- essentially, folks who live beyond their means are more vulnerable

Where are These Counties?
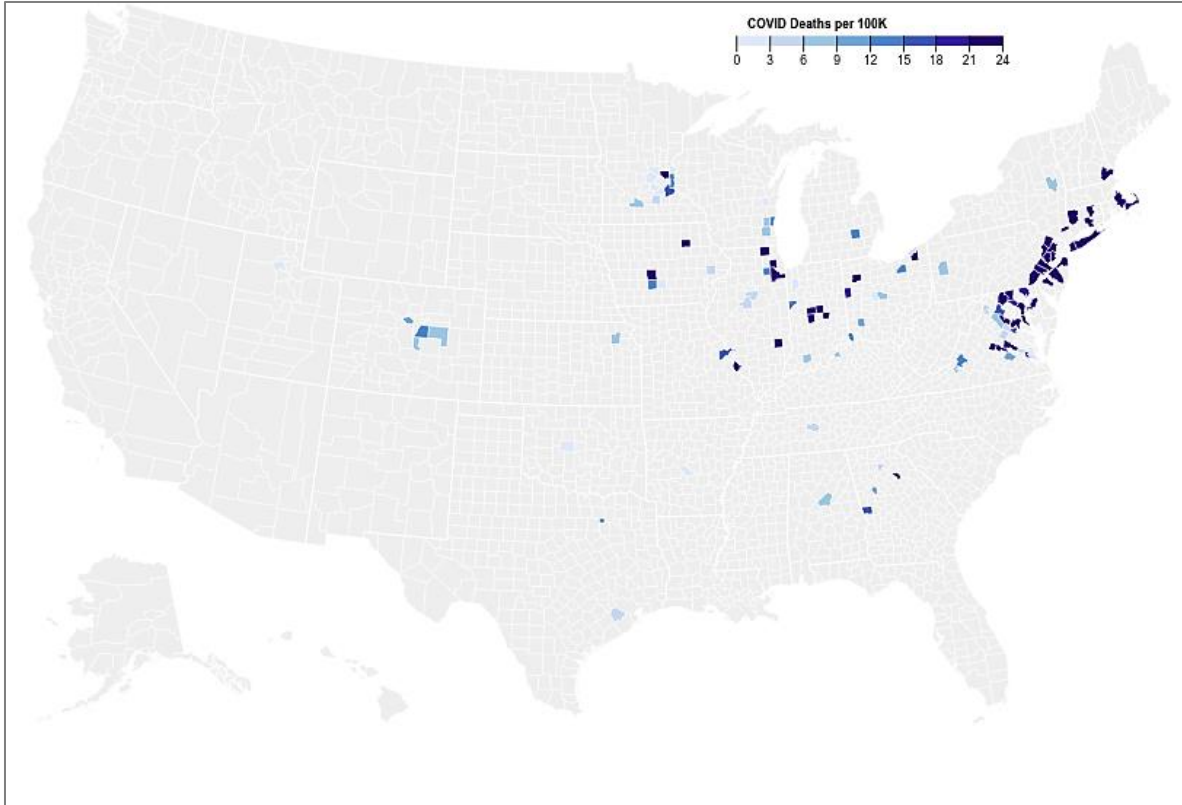
Strong Predictive Capabilities (1)

We observed for our May 2020 patterns

- for **98% of them** the death rate growth in June 2020 was **2-3 times higher** than the US average
- the other 2% grew at the average pace
- none slowed in growth below the US-average
- these trends continued in July and even August

See Them on a Map (May 2020)

COVID Deaths per 100K

0   3   6   9   12 15 18 21 24

May
10
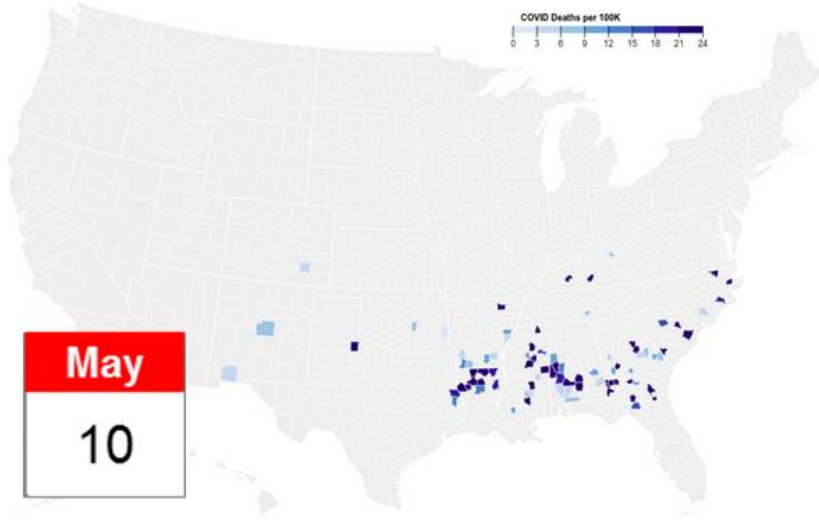
Sparsely populated counties with a poor and aging population

See Them on a Map (June 2020)
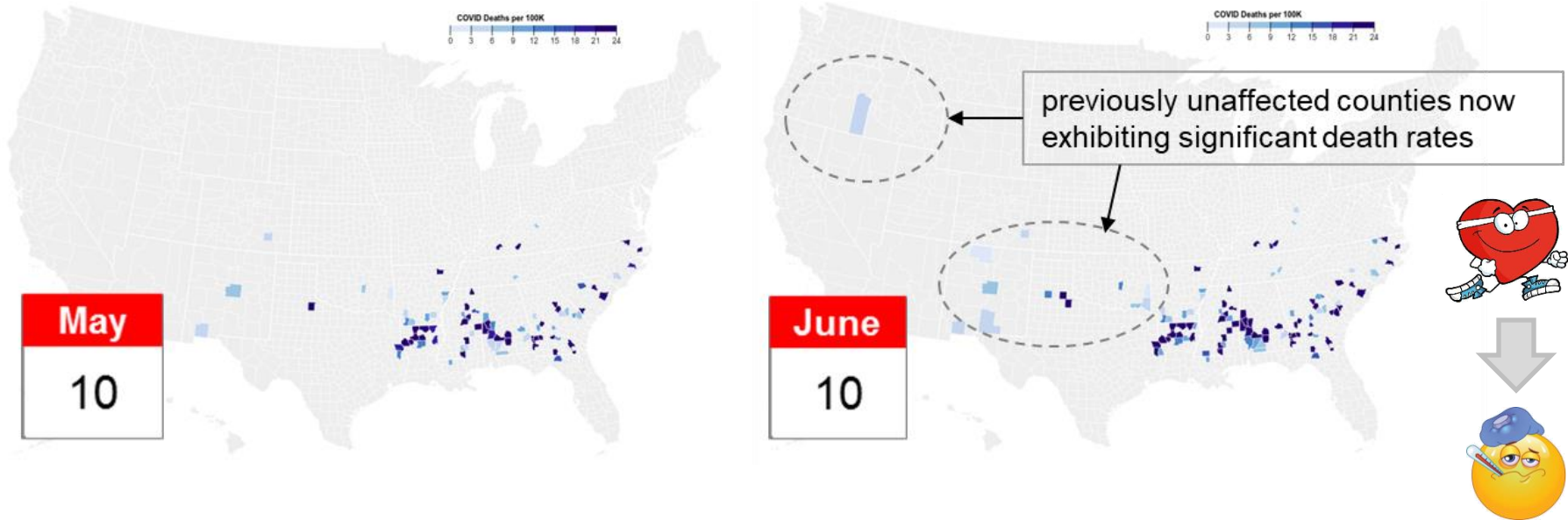
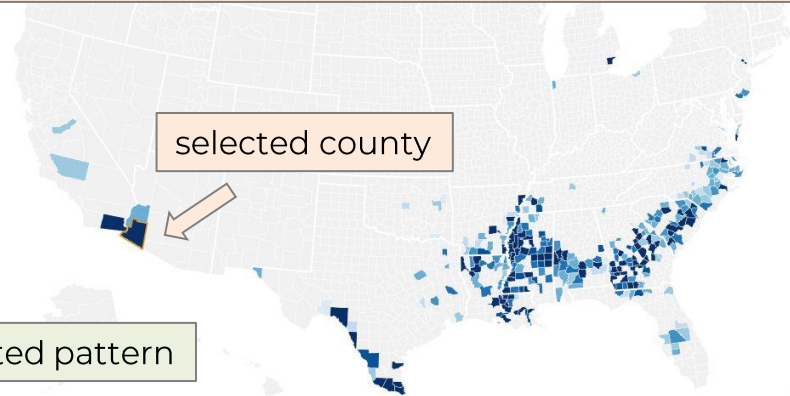Sparsely populated counties with a poor and aging population

Our Covid-19 Risk Explorer Dashboard

County Information Panel shows COVID-19 death rate over time and the most prominent risk factors for the selected county

Interactive Geomap shows the US counties that have the selected pattern shaded by COVID-19 death rate

selected county

Pattern Information Panel visualizes the risk factors of the currently selected pattern

selected pattern

Risk Pattern Browser, each tile is a pattern shaded and ordered by COVID-19 death rate

Our Covid-19 Risk Explorer Dashboard

[dashboard](#)   [video](#)

What Does It Mean?

Imagine you are a health official in **Cameron County, TX**

- you're **worried** about your county's residents and their vulnerability to COVID-19

- but you find it difficult to **assess the risk**

- this makes it difficult to do **effective messaging and policymaking**

What You Can Do!

You can use the two US counties we just found, **Passiac, NJ** and **McKinley, NM** to help you in that mission

- these counties have the **same risk patterns and top risk factors** than your county, **Cameron, TX**
- they showed up on COVID-19's death map **several months earlier**
- they either already **successfully mitigated the virus** or are on a good track for it

The similarities & virus mitigation successes should help in deriving an **effective prevention policy and persuasive message**

Take-Aways

- **Outlined a methodology** that can group socio-economic indicators of public health into 1-3 factor patterns learnt from observational data

- **Devised an interactive visual dashboard** by which the patterns can be explored in the context of the communities' geographical locations

- **Studied how policymakers and health officials** could use it to predict and manage the risk community has with respect to some natural hazard or public health emergency

Questions?

COVID-19 RISK EXPLORER

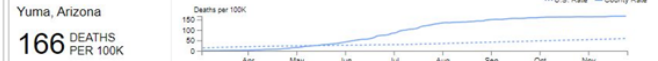**Interactive Geomap** shows the US counties that have the selected pattern shaded by COVID-19 death rate

selected county

selected pattern

County Information Panel shows COVID-19 death rate over time and the most prominent risk factors for the selected county

Yuma, Arizona
166 DEATHS PER 100K

Top 3 Risk Features:

% population uninsured
avg. GPA
% population w/o high schoo...

Pattern Information Panel visualizes the risk factors of the currently selected pattern

This pattern appears in 301 counties which have an avg. death rate of 119.4 per 100K

PM25 particle matter pollution
% adults with frequent phys...
% minority population

RISK PATTERN BROWSER

Risk Pattern Browser, each tile is a pattern shaded and ordered by COVID-19 death rate