

A 21st Century Census Curated Data Enterprise:

Demonstration Use Case on Skilled Nursing Facilities

Presentation to FCSM
October 24, 2023

Team: Vicki Lancaster, Stephanie
Shipp, Aaron Schroeder, Dawen
Xie, Henning Mortveit, Samarth
Swarup, Sallie Keller



BIOCOMPLEXITY
INSTITUTE



21st Century Census Curated Data Enterprise

The 21st Century Census Curated Data Enterprise (CDE) is a proposed shift of focus away from relying solely on individual surveys and towards creating statistical products defined by stakeholders' purposes and uses –
Statistical Products First.

The statistical product first process starts with research questions and ends with products that provide data-driven insights into these questions.

Demonstration Use Cases are being used to test this idea out and shed light on what capabilities will be needed to develop and deploy the CDE.

Each Use Case follows the CDE Framework.

Curated Data Environment Framework



The **outer rectangle** identifies the guiding principles for ethical transparent and reproducible product development and dissemination.

The **inner rectangle** identifies the statistical product development steps.

Curated Data Environment Framework

One can think of this as the scientific method for research using multiple data sources and types that are discovered and wrangled rather than collected based on a statistical model like an experimental design.



Purpose & Use

Nursing home became death trap

Four nursing home residents have died at an evacuation facility, Louisiana officials say.

Eight Dead From Sweltering Nursing Home as Florida Struggles After Irma

Hurricane Ian Hits Nursing Homes Already Beset By Staffing Crisis

US nursing homes aren't safe amid climate crisis, report finds. And it's getting worse.

THE WHITE HOUSE 

FACT SHEET: Protecting Seniors by Improving Safety and Quality of Care in the Nation's Nursing Homes

FEBRUARY 28, 2022 • STATEMENTS AND RELEASES

PURPOSE: Conduct a realistic appraisal of the skilled nursing facility that includes the physical facility, community in which the SNF resides, and the ability of the nursing staff to get to work and function.

USE: Provide a baseline assessment of the emergency preparedness of SNFs and the communities in which they reside that can be used by SNFs, emergency responders, and policy makers.

1. Are Skilled Nursing Facilities (SNF) prepared for extreme climate events?
2. Will the nursing staff be able to get to the facility in the event of an extreme climate event?
3. Will communities be able to provide support to SNFs?

Subject Matter Input

We met with SNF administrators, resident advocates, & researchers through discussions & listening sessions and conducted a library search.

Subject Matter Input

Data Discovery

Data Ingest & Governance

PURPOSE: Conduct a realistic appraisal of the skilled nursing facility that includes the physical facility, community in which the SNF resides, the ability of the nursing staff to get to work and function.

USE: Provide a baseline assessment of the emergency preparedness of SNFs and the communities in which they reside that can be used by SNFs, emergency responders, and policy makers.

Statistics Development

Fitness-for-Purpose

Data Wrangling

Data Discovery

Based on subject matter input & literature we made a wish list of variables for SNFs, climate threat, & community preparedness.

Subject Matter Input Data Discovery Data Ingest & Governance

PURPOSE: Conduct a realistic appraisal of the skilled nursing facility that includes the physical facility, community in which the SNF resides, the ability of the nursing staff to get to work and function.

USE: Provide a baseline assessment of the emergency preparedness of SNFs and the communities in which they reside that can be used by SNFs, emergency responders, and policy makers.

Statistics Development Fitness-for-Purpose Data Wrangling

Data List

Baseline Skilled Nursing Facility

Skilled Nursing Facilities

- Type (profit, not for profit)
- Chain status
- Last sold
- Emergency deficiencies
- Fire Safety deficiencies
- Inspections dates

Residents

- Total # of Residents
- % > 65
- % Minority
- % on Medicaid
- Cognitive function
- Mobility
- Physical Health

Nursing Staff (NS)

- Total # of Nursing Staff
- Where NS live
- How NS get to work
- Composition of NS family

Climate Change Threat

Riverine Flooding

- # of events
- Annualized frequency
- Risk score

Coastal Flooding

- # of events
- Annualized frequency
- Risk score

Hurricanes

- # of events
- Annualized frequency
- Risk score

Preparedness

Skilled Nursing Facilities

- # Emergency deficiencies and Inspections over 4yrs
- # Fire safety deficiencies and Inspections over 4yrs

Community Built Environment

- # Hospitals
- # Fire Stations
- # Red Cross Chapters
- # National Shelter System Facilities
- # Urgent Care Facilities
- # Emergency Medical Service Stations
- # Primary care providers
- Transportation infrastructure

Community Population Resilience

- % < High School
- % Disabilities
- % Unemployed
- % No vehicle
- GINI Index

Data Ingest & Governance

Set up a GitHub repository to store public data, metadata, code, statistical products, data processes, & relevant literature on SNF resilience & regulations.

Subject Matter Input Data Discovery Data Ingest & Governance

PURPOSE: Conduct a realistic appraisal of the skilled nursing facility that includes the physical facility, community in which the SNF resides, the ability of the nursing staff to get to work and function.

USE: Provide a baseline assessment of the emergency preparedness of SNFs and the communities in which they reside that can be used by SNFs, emergency responders, and policy makers.

Statistics Development Fitness-for-Purpose Data Wrangling

Appendix C. Census Skilled Nursing Facility Use Case GitHub Repository

There are three main folders in the GitHub Repository,

1. **Data** – Owners, Facility, Nursing Staff, Residents, Community, Resilience, Demographic + Meta Data for each folder
2. **Source Code**
 - Discovery (EDA-Exploratory Data Analysis (choropleths, box plots, isochrones)
 - Analysis (County Assets- Infrastructure and Workers, Deficiency Indexes, population resiliency, Probability of getting to SNF, and City of Richmond, VA example)
3. **Documents**
 - Literature and other Products (data tables, visualizations, presentations, and reports)
 - Processes (commute vulnerability) and derived variables (e.g., county resilience, shelter and emergency facilities, deficiency indices, average daily nursing staff, and probability of work not getting to work).

Table of Contents for the GitHub Repository

https://github.com/uva-bi-sdad/census_cde_demo_2

1. Data Folder

- Subfolder: **Virginia Skilled Nursing Facility**
 - Subfolder: **Owners**
 - File: [va cms ownership 2022-08.csv](#)
 - ✓ Subfolder: **Meta Data**
 - File: [cms five star users guide 2022-01.pdf](#)
 - File: [cms primary data dictionary.xlsx](#)
 - Subfolder: **Facility**
 - File: [us cms inspection dates 2022-06.csv](#)
 - File: [va cms fire safety deficiencies 2022-12.csv](#)
 - File: [va cms provider final 2022-07.csv](#)
 - ✓ Subfolder: **Meta Data**
 - File: [cms deficiency citation descriptions 2022-07.xlsx](#)

Data Wrangling

Subject Matter Input Data Discovery Data Ingest & Governance

PURPOSE: Conduct a realistic appraisal of the skilled nursing facility that includes the physical facility, community in which the SNF resides, the ability of the nursing staff to get to work and function.
USE: Provide a baseline assessment of the emergency preparedness of SNFs and the communities in which they reside that can be used by SNFs, emergency responders, and policy makers.

Statistics Development Fitness-for-Purpose Data Wrangling

288 SNFs in Virginia, 283 had sufficient staffing data to be included in this Use Case.

Data Wrangling Process

Profiling	Preparation	Linkage	Exploration
Quantity, Structure, Metadata, Provenance	Cleaning, Transforming, Restructuring,	Ontology, Selection & Alignment, Harmonization	Visualization, Exploratory Data Analyses, Characterization

Data wrangling issues:

- Nonstandard names for hospitals and classification system
- Absence of data dictionaries that would allow users to compare variables across data sets
- Inconsistent geographic information

Fitness-for-Purpose

Subject Matter Input Data Discovery Data Ingest & Governance

PURPOSE: Conduct a realistic appraisal of the skilled nursing facility that includes the physical facility, community in which the SNF resides, the ability of the nursing staff to get to work and function.

USE: Provide a baseline assessment of the emergency preparedness of SNFs and the communities in which they reside that can be used by SNFs, emergency responders, and policy makers.

Statistics Development Fitness-for-Purpose Data Wrangling

Exploratory data analyses were done to ensure the SNF Deficiency Indices had sufficient variability to discriminate between high and low-performing SNFs.

Emergency Preparedness Deficiency Indicator

Descriptive Statistics

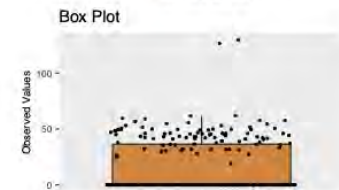
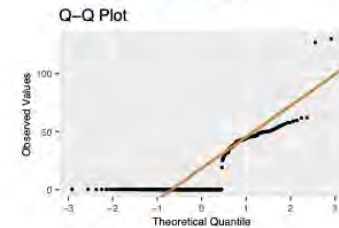
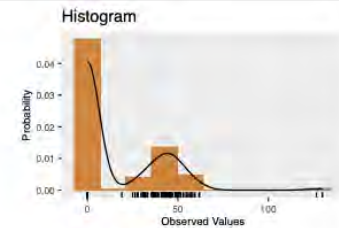
N = 283
Min. = 0
Q(.25) = 0
Median = 0
Q(.75) = 36.416667
Max. = 129.833333
Mean = 14.6223
SD = 22.8097
CS = 1.4981
CK = 5.7315
CV = 1.5599

Inferential Statistics

D'Agostino & Pearson
Test for Normality
K2 = 86.3642
p(K2) = 0

Test for Skewness
Z(CS) = 7.9659
p(Z(CS)) = 0

Test for Kurtosis
Z(CK) = 4.7862
p(Z(CK)) = 0



Fire Life Safety Code Deficiency Indicator

Descriptive Statistics

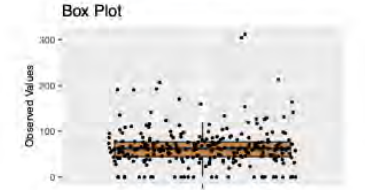
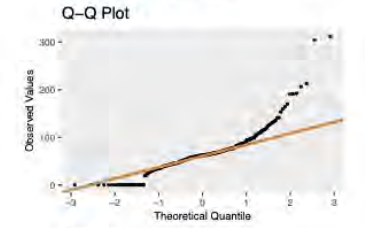
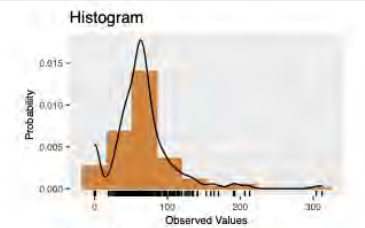
N = 283
Min. = 0
Q(.25) = 44.650794
Median = 62.025641
Q(.75) = 76.636364
Max. = 312
Mean = 65.2057
SD = 42.04
CS = 1.9222
CK = 11.1016
CV = 0.6447

Inferential Statistics

D'Agostino & Pearson
Test for Normality
K2 = 140.9591
p(K2) = 0

Test for Skewness
Z(CS) = 9.297
p(Z(CS)) = 0

Test for Kurtosis
Z(CK) = 7.3841
p(Z(CK)) = 0



Statistical Development

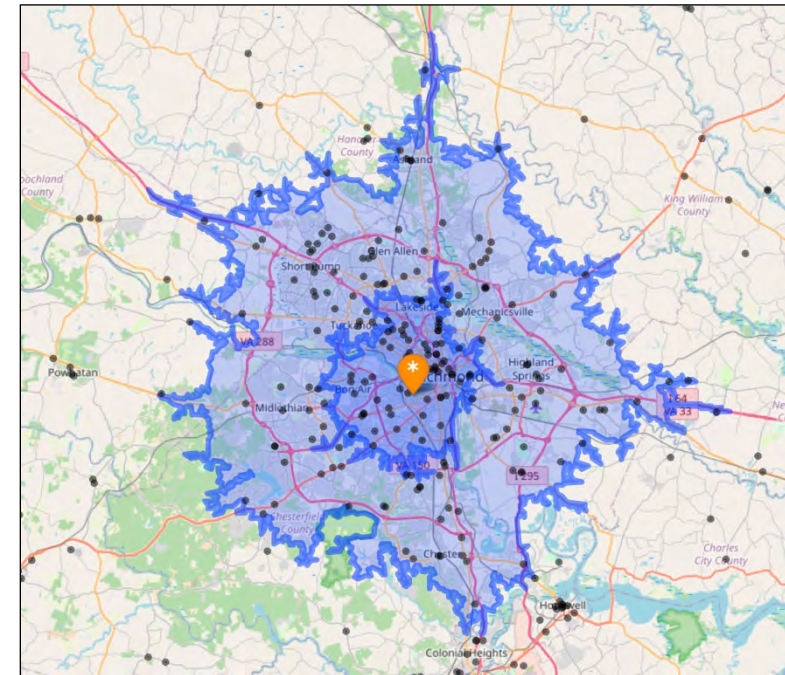
- Integrated Use Case Database
- Composite Indices
 - SNF Deficiency Indices
 - Community Resiliencies Indices
- Choropleth & Isochrone Maps
- Modeling Results
- Simulation Estimates
- R code use to
 - Download, clean & integrate multiple data sources
 - Produce the various statistical products

Subject Matter Input Data Discovery Data Ingest & Governance

PURPOSE: Conduct a realistic appraisal of the skilled nursing facility that includes the physical facility, community in which the SNF resides, the ability of the nursing staff to get to work and function.

USE: Provide a baseline assessment of the emergency preparedness of SNFs and the communities in which they reside that can be used by SNFs, emergency responders, and policy makers.

Statistics Development Fitness-for-Purpose Data Wrangling



Are SNF prepared for emergencies?

CMS Skilled Nursing Home Deficiencies as a proxy for preparedness

Emergency Preparedness (EP)

- Emergency plans
- Communication plan
- Training and testing
- Policies and procedures

Life Safety Code (LSC) Deficiencies

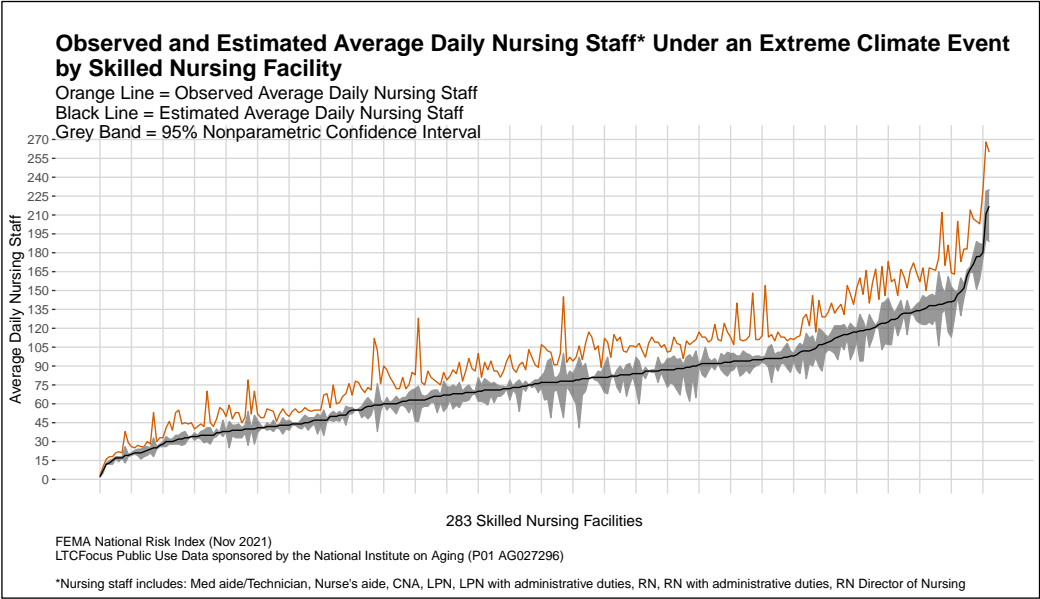
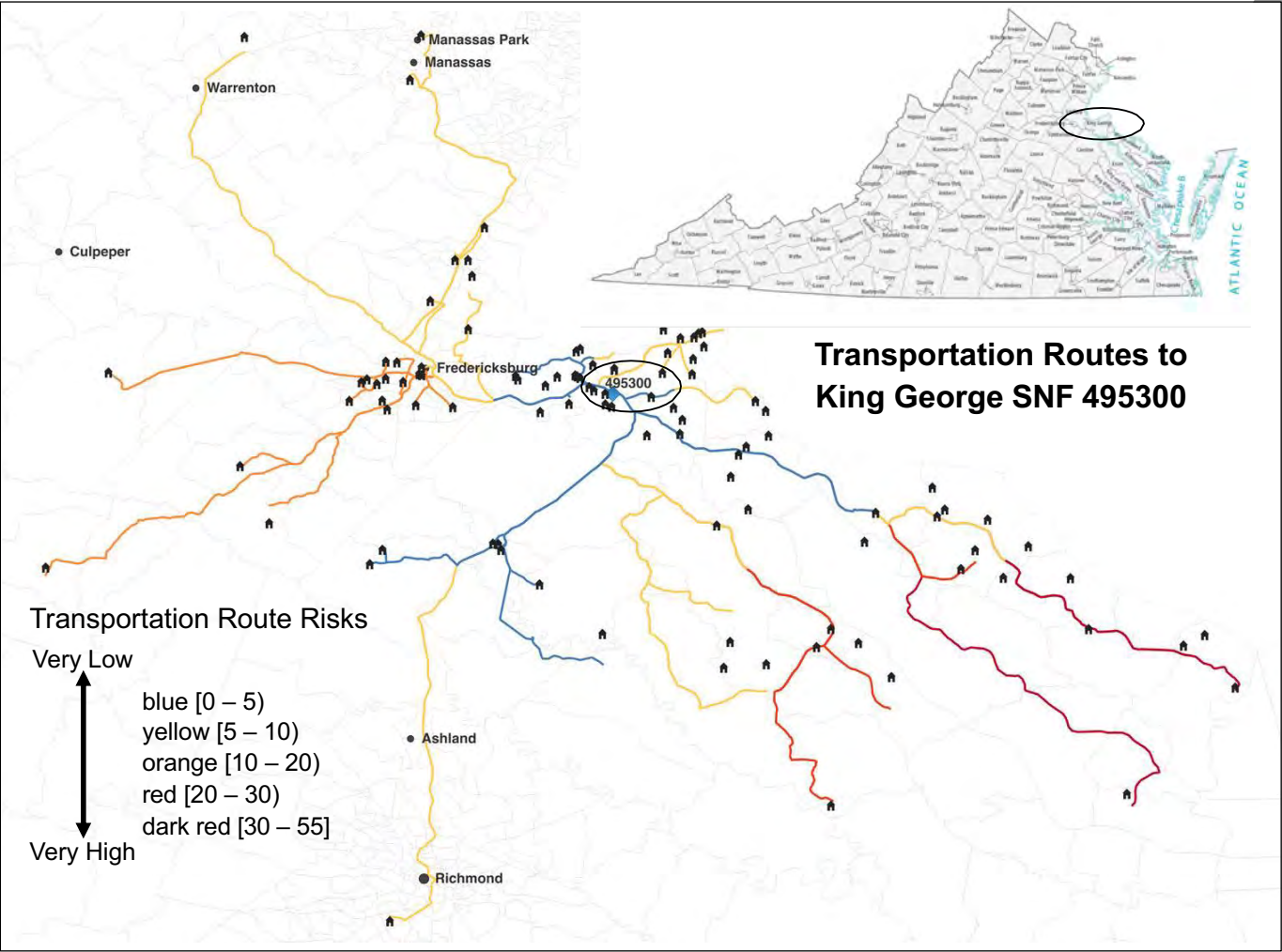
Fire protection requirements

- Designed to provide a reasonable degree of safety from fire, smoke, and panic
- Construction
- Protection
- Operational features

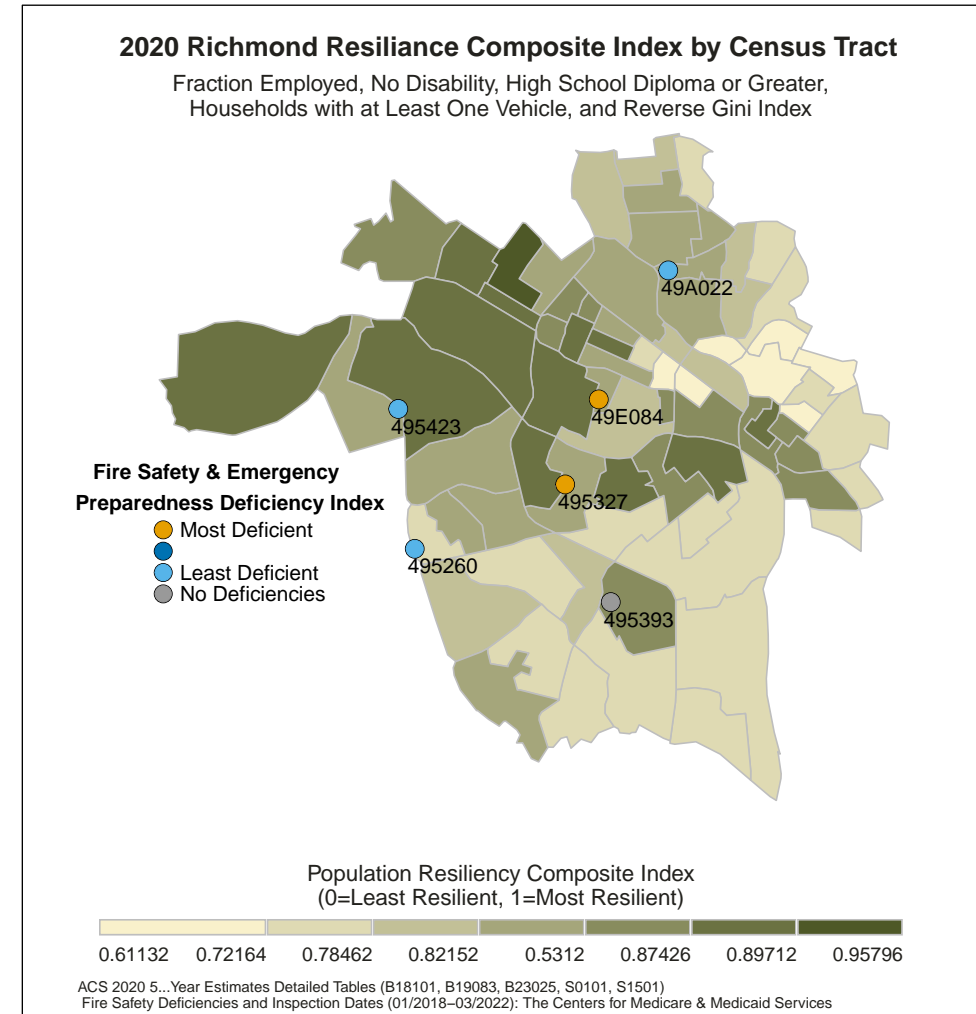
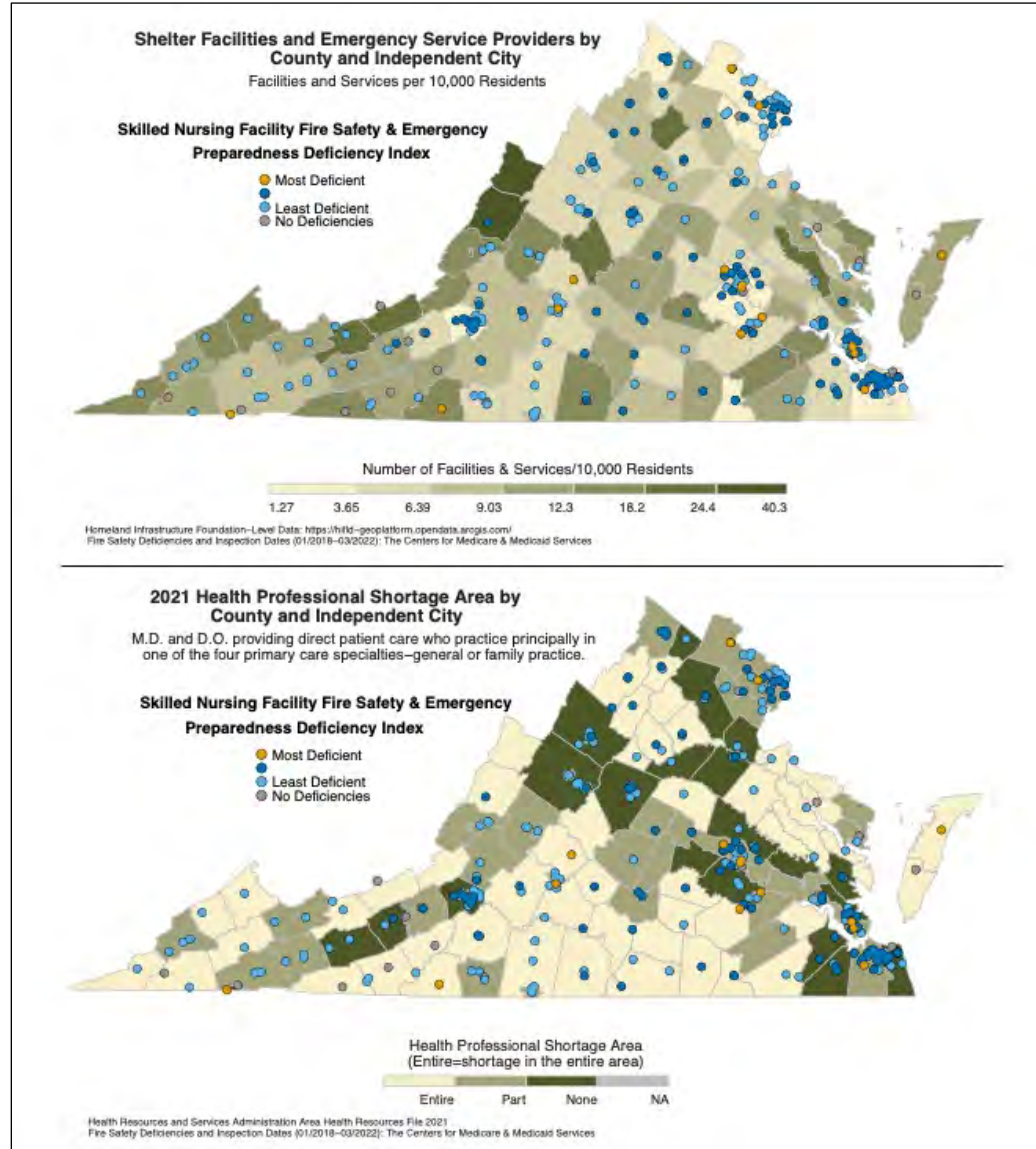
Fire Safety & Emergency Preparedness Deficiency Index

- Most Deficient
- Least Deficient
- No Deficiencies

Will nursing staff reach the facility?



Can communities provide support?



What CDE capabilities identified in this Use Case?

CDE Capabilities developed:

- Identify the home location of SNF workers.
- SNF workers' transportation routes + risk of flooding by route
 - Monte Carlo simulation to estimate the average probability of getting to work and the percentage loss in nursing staff based on the average daily skilled nursing staff numbers from CMS.
- SNF combined deficiency index
 - SNF Emergency Preparedness deficiencies metric
 - SNF Life Safety Code (fire) deficiencies metric
- Community resilience indicators
 - Number of shelter facilities and emergency service providers per 10,000 people.
 - Mapping health professional shortage areas.

What are the gaps to be filled in future use cases?

- Provide crosswalks that link location variables such as lats & long, FIPS codes, ZIP codes, addresses, etc.
- Resolve inconsistencies and omissions observed across data sources on the same topic. In this Use Case, three data sets on hospitals had different names and federal IDs such as Medical Provider Number.
- Provide private-equity ownership data, there is no definitive source that is publicly available (GAO 2023).

Technical Report: Lancaster V, Shipp S, Keller S, Schroeder A, Mortveit H, Swarup S, Xie D, (2023), Census Curated Data Enterprise Use Case Demonstration: Climate Resiliency of Skilled Nursing Facilities. Proceedings of the Biocomplexity Institute, Technical Report. TR# 2023-53. University of Virginia. <https://doi.org/10.18130/ce97-sp05>