Census Curated Data Enterprise Use Case Demonstration: Skilled Nursing Facilities

Vicki Lancaster, https://orcid.org/0000-0002-1517-3807
Henning Mortveit, https://orcid.org/0000-0003-3363-2947
Aaron Schroeder, https://orcid.org/0000-0003-4372-2241
Samarth Swarup, https://orcid.org/0000-0003-3615-1663
Dawen Xie, https://orcid.org/0000-0001-5217-1467
Sallie Keller*, https://orcid.org/0000-0001-7303-7267
Stephanie Shipp, Project Lead, https://orcid.org/0000-0002-2142-2136
Contact: sss5sc@virginia.edu

Social and Decision Analytics Division
Biocomplexity Institute & Initiative
University of Virginia

May 5, 2020
Updated September 27, 2023

Funding: This research was sponsored by the U.S. Census Bureau Agreement No. 01-21-MOU-06 and Alfred P. Sloan Foundation grant, G-2020-14002.

Acknowledgments: We would like to thank our colleagues for their input and insights: Joseph Salvo, Edward Wu, and Joanna Schroder, Biocomplexity Institute; Zhengyuan Zhu, Iowa State University; Nicole Howell for her expert input about data sources and the nursing home industry; Deirdre Bishop, Michael Ratcliffe, and staff in the US Census Bureau's Geography Division and many others throughout the Census Bureau.

*Some work was performed under the supervision of the U.S. Census Bureau as part of the author's contractual work product. The views expressed in this perspective are those of the author and not the U.S. Census Bureau.

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

Abstract

The proposed **Curated Data Enterprise (CDE)** is a transformative approach for how the Census Bureau could accomplish its mission to develop and provide high-quality, timely, and geographically detailed statistical products by changing how it manages its data assets, incorporates information from external sources, and leverages them for the public good. The CDE explicitly focuses on curating not only the data that are part of the enterprise but also all of the processes associated with creating a statistical product, including the relevant context, and ingesting, analyzing, and producing purpose-driven statistical products on platforms that permit wide accessibility. The CDE would move the Census Bureau into a position to meet the challenges confronting statistical agencies in the 21st Century (Keller et al., 2022).

This report demonstrates the CDE concept through a Use Case on Skilled Nursing Facilities (Shipp, Salvo, & Zhang, 2022). We implemented this Use Case to begin to highlight the capabilities needed to develop and deploy the proposed CDE successfully. Lessons learned through the development will inform the design of a Use Case Research Program to test and implement the CDE concept use case by use case. It is imperative that this be done by posing relevant questions and building data insights that will inform the public and stakeholders and provide a platform for future research.

Curated Census Data Enterprise Curation Report:
A Demonstration Use Case on Skilled Nursing Facilities

Introduction

The proposed **Curated Data Enterprise (CDE)** is a transformative approach for how the US Census Bureau could accomplish its mission to develop and provide high-quality, timely, and geographically detailed statistical products by changing how it manages its data assets by incorporating information from external sources and leveraging them for the public good. To accomplish this, the CDE explicitly focuses on curating not only the data that are part of the enterprise but also the processes associated with creating a statistical product, including the relevant context, and ingesting, analyzing, and producing purpose-driven statistical products on platforms that permit wide accessibility. While our research primarily focuses on the Census Bureau implementation, the CDE is also a model for other organizations.

The CDE is part of the Census Bureau's effort to build a statistical product-focused ecosystem based on purpose and use. This requires a change in focus from managing the census and single surveys to managing an entire ecosystem of data collection, processing, and dissemination designed to use various data sources to deliver products that best address the questions posed by data users (Thieme, 2022). The CDE would move the Census Bureau into a position to meet the challenges confronting statistical agencies in the 21st Century (Keller et al., 2022). And, in the words of the Census Bureau Director, Robert Santos, "This is the essence of a curated data approach — assemble, assess and fill in the gaps to create quality statistical data products" (Santos, 2023).

The CDE is a continuously evolving infrastructure that will empower and enable researchers and their data users to develop new measures of people, places, and the economy. With its focus on purpose and use, the CDE will make it possible to leverage different data sources, including sample surveys, official administrative records, and synthetic data to provide more robust, timely, and comprehensive measures (Keller et al., 2022). The CDE framework provides a repeatable and rigorous process to ensure the development and dissemination of high-quality statistical products. Curation will be designed to document each step of the process, including decisions that involve practical tradeoffs, for example, between the provision of timelier data or more geographic detail on a particular topic, but at somewhat lower quality, which still is fit-for-purpose and can be improved over time. This report is an early example of that curation – documenting the decisions made, data sources used, methods implemented, etc. The CDE curation process will need to be partially automated and captured in an online system that others can easily retrieve.

We propose to develop the CDE use case by use case. We define a Use Case as a research method to identify, clarify, and organize requirements to create a CDE. For example, a Use Case provides concrete context that can help define a process or processes to formulate a research question, discover, acquire, curate, and store data; develop a hierarchy for data sources on the same topic; benchmark or validate synthetic data; and develop methods for creating statistics – all of which can help guide the development of the CDE to produce statistical products in new ways. The CDE Framework provides the roadmap to guide this work (see Exhibit 1).

Art-of-the-possible Use Cases are focused on current policy issues for which there is no single or definitive data source to answer policy questions. When fully implemented, these proposed Use Cases demonstrate a collection of capabilities that need to be built into the CDE, such as reusing modules and data, tracking the provenance of collected and generated data, creating and reusing synthetic data and methods to integrate many types of data, conducting statistical analyses, and reviewing data and results through an equity and ethics lens. This art-of-the-possible Use Case demonstration described herein lays the groundwork for what is needed to develop the CDE use case by use case.

The context of this demonstration project is to create a baseline picture of Skilled Nursing Facilities (SNF) in Virginia and then add data about the risk of an extreme climate event, explicitly flooding, to assess their preparedness, and to compute the probability that the nursing staff could make it to the facility during a flood event (Shipp, Salvo, & Zhang, 2022).

Developing a CDE Use Case is a data science activity and needs a well-integrated team. Our team included statisticians, economists, policy researchers, and computer scientists. The research was vetted with subject matter experts from the University of Virginia's CDE team (Keller et al., 2022) and shared with key stakeholders from the Census Bureau.

The CDE framework guided the work (Exhibit 1). This framework presents a high-level set of principles and a process for stating and documenting purpose and use, with criteria at each step.

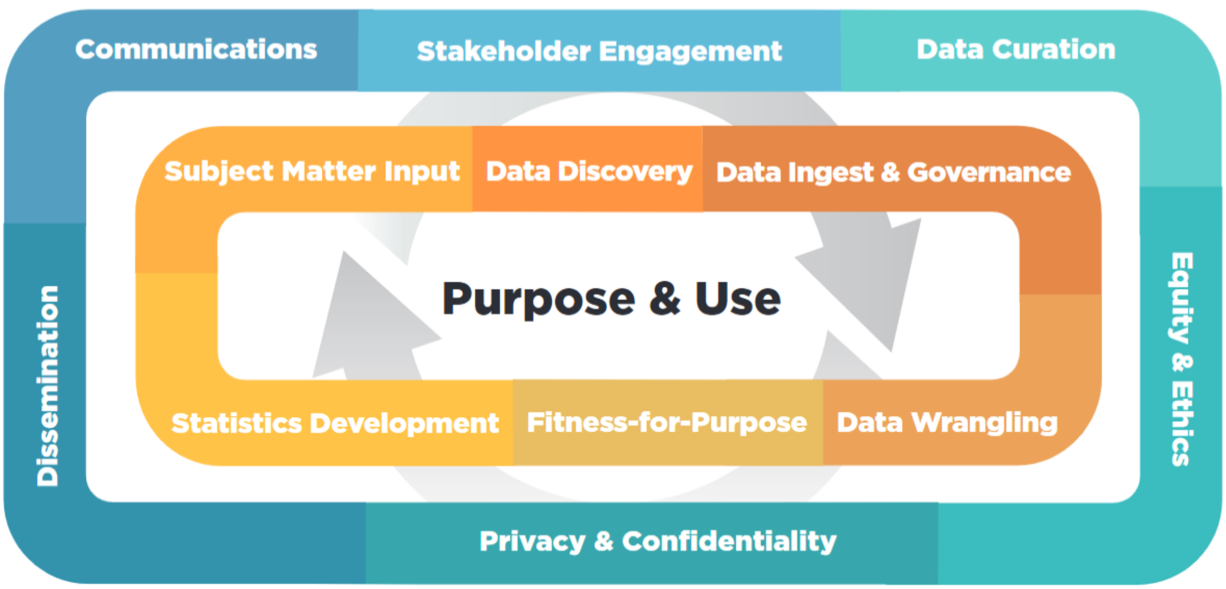

Exhibit 1. Curated Data Enterprise Framework (Keller et al., 2020)

The CDE framework starts with the research question (purpose and use) and continues through the following parallel and iterative processes. The outer processes include stakeholder engagement, communication & dissemination throughout the project, privacy & confidentiality decisions, equity & ethics reviews, and data curation. The inner processes include data discovery—inventory, screening, and acquisition; data ingestion & governance; data wrangling & profiling; fitness-for-purpose evaluation; statistics development; and subject matter input.

The plan for this Use Case (and others to follow) is to illuminate the capabilities to support the development of the CDE. We describe the framework as it unfolded for this demonstration. Curation at each step is critical to ensure transparency and repeatability of the use of the data, processes, and functions in the current and subsequent applications. Curation involves documenting, for each CDE product, the questions and context (i.e., purpose and use), inputs from which the product is derived, the data wrangling and analyses used to transform inputs into products, the CDE statistical product itself, and other processes and decisions made throughout the Use Case.

The next part of this report takes the reader through the Curated Data Enterprise Framework steps outlined in Exhibit 1, followed by four appendices. Appendix A lists the most common SNF deficiencies. Appendix B describes the CDE demonstration data for the six SNFs in the city of Richmond, Virginia. Appendix C provides the structure (data, documents, literature and products, such as visualizations, data tables, and reports and presentations) and source code (discovery and analysis) in our GitHub repository. Throughout, the reports cited are hyperlinked to their original sources (when available) and are also found in the reference list and GitHub repository. Appendix D describes alternative publicly available data sources to the proprietary ones used in this demonstration.

Walking through the CDE Framework

We describe the steps we took to develop the SNF Use Case and the decisions using the CDE framework as a guide to our work. Some topics are not covered in this Use Case, e.g., privacy and confidentiality. A Use Case does not need to cover every step to identify and develop capabilities needed for the CDE. Ultimately, a statistical product will have to include all steps.

Stakeholder Engagement

We sought input from nursing facility administrators, demographers, and researchers through discussions and listening sessions. What they told us is also conveyed in initiatives, news stories, and journal articles. As stated in The White House (2022) initiative, *Protecting Seniors by Improving Safety and Quality of Care in the Nation's Nursing Homes*, "All people deserve to be treated with dignity and respect and to have access to quality medical care." With regard to equity, there is a lack of support and response provided to Skilled Nursing Facilities during disasters. Historically, these care facilities and homes are not incorporated into existing emergency response systems. In 2004, a National Response Plan (NRP) was created by a Homeland Security Presidential Directive to align Federal capabilities and resources into a unified all-hazards approach to domestic incident management comprising eight Emergency Support Functions (ESF) (examples of ESFs include: ESF-1 Transportation; ESF-2 Communication, …, ESF-8 Public Health and Medical Services). ESF-8, Public Health and Medical Services, is responsible for public health and medical services during disasters.

Skilled nursing facilities (and other similar types of facilities) are often ignored during emergency events. During the 2004 Florida hurricane season, SNFs were given the same priority as day spas for *restoring* electricity, telephones, water, and other basic services (Hyer et al., 2006). Even worse are the deaths of nursing home residents in Louisiana following Hurricanes Katrina and Rita in 2005 (Dosa et al., 2008). This is still an issue in 2021. In Louisiana, 15 nursing home residents died when evacuated to a warehouse during Hurricane Ida (2021) and 12 died in Florida after Hurricane Irma (2017). In both instances, the deaths were attributed to extreme heat due to lack of electricity (Skarha et al., 2021).

These instances demonstrate the relevance of this Use Case and its applicability to other areas of the country. The community resilience estimates and SNF deficiency indicators developed for this Use Case could be shared with emergency support services to help identify locations needing more help during an emergency. The integrated data product, which integrates the data sources from the data discovery process with our statistical products, could be used to target populations to implement capacity-building and communication strategies to reduce vulnerability during an extreme climate event.

Purpose and Use

A Skilled Nursing Facility (SNF) is a nursing facility with the staff and equipment to provide skilled nursing care, skilled rehabilitation services, and other related health services (CMS Glossary, 2022). SNFs are federally regulated. This study focuses on SNFs, not state-regulated assisted living and other residential facilities that are alternatives to SNFs, generally for older populations.

The impact of crisis events, such as emergency preparedness response to the COVID-19 pandemic or severe climate-related events, has highlighted the need for an integrated assessment of SNFs (Skilled Nursing Facilities). The White House (2022) initiative highlights the importance of these issues. This pilot study focuses on emergency preparedness and nurse staffing, two issues raised in the White House initiative. We have reframed these issues into three purpose and use questions for this Use Case:

1. Can Skilled Nursing Facilities (SNF) workers get to work during an extreme flood event?
2. Are SNFs prepared for a flood emergency?
3. Can communities support SNFs during an emergency?

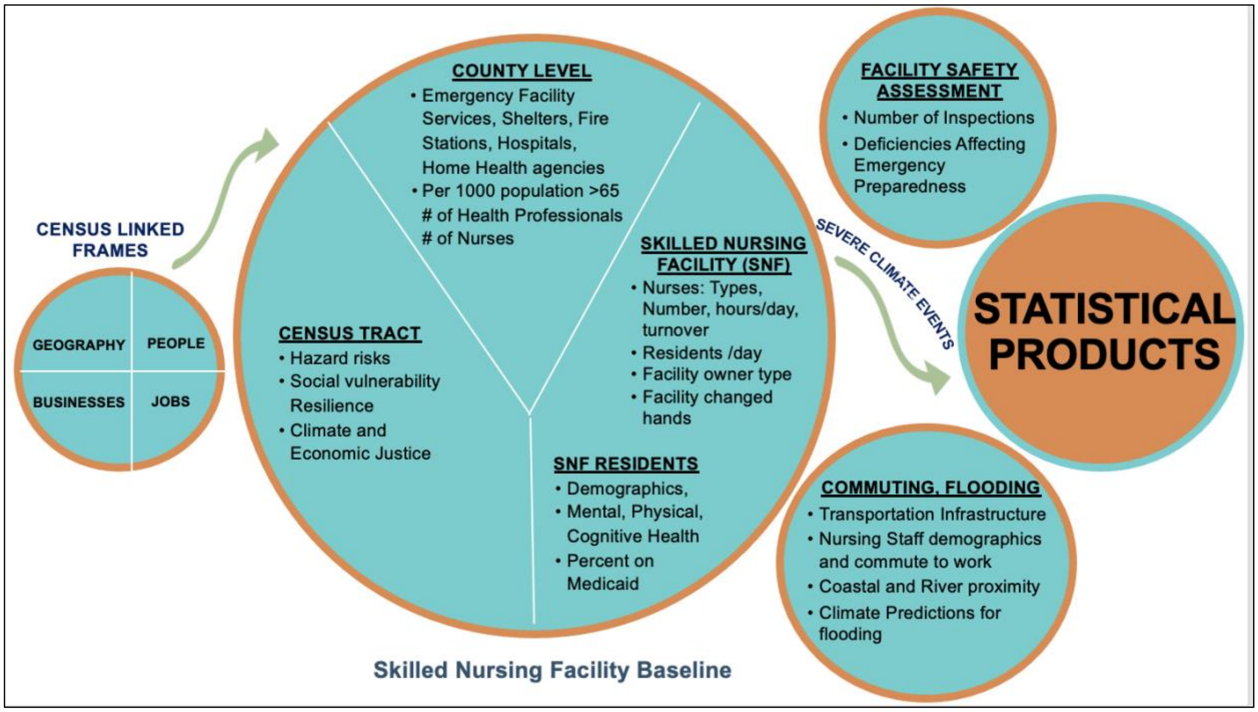

Exhibit 2. Conceptual Data Map Aligned to Purpose and Use

Creating a conceptual data map is the first step for data discovery. The team identified the kinds of data needed to implement the Use Case. It then set out to discover, screen, and acquire the needed data sources. In this demo, we undertook the project in two phases. (1) We first created a baseline data picture of SNF owners, workers, residents, and characteristics. (2) Next, we identified the factors that affect SNFs during severe climate events (e.g., flooding events). New statistical products are produced through data wrangling and statistical analysis. Although not used in this demonstration, the universal linked frames are part of the comprehensive universe of data that could be used. See Exhibit 9 for lists of actual data used and statistical products produced. These products are capabilities developed as a first step for creating the CDE.

Data Discovery

Prior to the data discovery process, we created a conceptual data map aligned with the three research questions listed above and informed by a review of the SNF literature and discussions with subject matter experts to help identify our data needs and potential data sources. The experts included nursing facility administrators, demographers, and researchers. Exhibit 2 displays the results of the process to identify what data we would want to undertake this analysis. These ideas for data align with the purpose and use questions based on our discussions with potential users.

The data sources that were discovered, evaluated for fitness-for-purpose, and selected for the Use Case are listed below. Similar to Census Bureau data assets, the data used in this pilot is a combination of publicly available data sources and restricted (proprietary) sources that will be integrated to create a public-access statistical product. In the list of data sources that follow, all are provided in the Use Case GitHub along with their metadata with the exception of three listed below under Transportation Routes which are proprietary. Two of these data sources are synthetic, in this case links to articles on how they were constructed are provided. The third data source was obtained from a private-sector vendor whose documentation is proprietary, a link is provided to their website. In the future, Census Bureau could enable for data from private-sector vendors to be included in the CDE.

- Skilled Nursing Facilities (SNF)
    - Centers for Medicare & Medicaid Services (CMS, 2022b)
        - SNF Owners, Nursing Staff
        - SNF Deficiencies for Emergency Preparedness and Fire Life Safety Code (Non-Health Deficiencies)
        - SNF Inspection Dates for Emergency Preparedness and Fire Life Safety Code
    - Long Term Care Focus (LTCFocus) Public Use Data sponsored by the National Institute on Aging (Brown University, 2022)
        - Residents

- Community Resilience, Demographics, Assets, and Risks by geography (County, City, Census tract)
    - American Community Survey
        - Population resilience indicators
    - Department of Health and Human Services, Health Resources and Services Administration, Area Resource Files (HRSA, 2022)
        - Health professional shortage areas
    - Department of Homeland Security, Homeland Infrastructure Foundation Level Data (HIFLD) (DHS, 2022)
        - Shelter facilities and emergency service providers
    - Department of Homeland Security, FEMA (Federal Emergency Management Agency)
        - FEMA Community Resilience Indicator Analysis (FEMA, 2022a)
        - FEMA National Risk Index for Natural Hazards (FEMA, 2021)

- Transportation Routes
    - The BI NSSAC synthetic population ver. 2.0 for Virginia, US (DP-US-VA-2.0). The description of the elements and data fields contained in this product are provided in *Detailed Overview and Description of Components in U.S. Pipeline 2019* (NSSAC Technical Report 2019-025) (Mortveit et al., 2023).
    - The BI NSSAC Building Database for the US (BDB-1.0). The description of the modeling and the data fields included in this product are in *NSSAC Building Knowledge Base Modeling and Implementation* (NSSAC Technical Report 2021-16). (Mortveit, Xie, & Marathe, 2023)
    - The full description of the HERE Premium Streets data (2021/Q1) is provided to customers in a proprietary document (HERE NAVSTREETS Reference Guide v15.1.pdf for their 2022 Q1 data), which, as part of their terms of use, we are unable share. Examples of fields extracted include link IDs, function class (hierarchy level), speed limits, number of lanes, bi-/uni-directionality, geometry, length, road access restrictions (e.g., bus lanes, HOV lanes), road link additional details (e.g., tunnel, bridge, ferry), and details regarding evacuation.

- FEMA (Federal Emergency Management Agency) Riverine and Coastal Flooding Maps (FEMA, 2022a)

Data Ingestion and Governance

All the public data, metadata, code, statistical products, data processes, and relevant literature on SNF resilience and regulations are stored in a GitHub repository. The structure of the repository can be viewed at CENSUS SNF Use Case GitHub Repo Structure. The GitHub Repo Table of Contents with links to the files stored in the repository is provided in Appendix C. An exception is the proprietary (commercial and non-commercial) data used to construct the transportation routes to the SNFs which cannot be shared. Proprietary data were used for this Use Case to amplify the mix of data types and governance complexities the CDE will need to address.

Data Wrangling

Of the 288 SNFs in Virginia, 283 had sufficient staffing data to be included in this Use Case. We hypothesize the missing staffing data are linked to the severe staffing shortages in Virginia during COVID based on a newspaper report that Virginia nursing homes lost more than 9,500 facility employees during the pandemic (Masters, 2021). Staffing data are submitted quarterly to CMS and are due 45 days after the end of each reporting period. Data submitted outside the deadline are not accepted.

The two predominant issues with data wrangling for this Use Case included reconciling data sources that contain data on the same topic and creating linkages between datasets. For example, we reviewed three hospital data sources:

1. Homeland Security Infrastructure Foundation-Level Data (HIFLD) (DHS, 2022)
2. HealthData.gov - COVID-19 Reported Patient Impact and Hospital Capacity by State (HHS, 2022)
3. Map of VHHA Hospital and Health System Members (Virginia Hospital & Healthcare Association, 2022)

Inconsistences and omissions observed across the three data sources included:

- non-standard names for hospitals;
- inconsistent reporting of hospital systems/conglomerates;
- non-standard classification systems for types of hospitals;
- absence of data dictionaries that would allow a user to compare variables and inconsistencies across datasets;
- inconsistent availability of hospital IDs (such as Medicare Provider Number); and
- inconsistent geographic information, including address, latitude, and longitude.

For the purposes of the demonstration, we did not attempt to reconcile these inconsistencies but decided to use a single source for shelter facility and emergency service provider data. We used HIFLD data since they provided the most current data (DHS, 2022). The use of these data reinforces the purpose of the Use Case - to illuminate the challenges in creating statistical products and what the Census Bureau would need to consider. This is an example of decisions made to expedite the creation of the statistical product and can be revisited later.

The following discussion about challenges highlights the issues to continually strive to resolve while balancing to expedite the process. This is the case when it comes to products related to popular indices, such as climate disaster risks and community resilience, each operationalized differently across the various departments and agencies within the federal government, state governments, and private and non-profit sectors. It is a large task to review the methodology and technology reports (if available) to understand their differences and decide on which versions are most relevant (fitness-for-purpose) for a particular Use Case. Again, after reviewing the options, for this Use Case, we determined that the National Risk Index for riverine and coastal floods from FEMA was the best option for climate risks estimates. The detailed technical report, *National Risk Index Technical Document* (FEMA, 2022b), provides a clear assessment of the assumptions and limitations of the data and a description of how the risk estimates were derived.

Resilience metrics are constructed from community characteristics researchers have theorized may make individuals, households, built and natural environments less resilient to natural hazards. These estimates aid emergency managers in planning and outreach by identifying vulnerable households so it is important they reflect the environment in which they are being used. Although the National Risk Index also provides estimates of social vulnerability and community resilience from the University of South Carolina's Hazards & Vulnerability Institute (Cutter et al., 2019), they are at the county level, and the latest estimates are constructed using 2015 data and 2010 geographic boundaries. We also explored using the 2019 Community Resilience Estimates (CRE) from the Census Bureau since these are the only resilience estimates constructed using both individual and household data and provide measures of variability. However, the most current CRE are from 2018 and therefore use 2010 geographic boundaries (US Census Bureau, 2021).

We, therefore, decided to estimate our own metric based on guidance from the Department of Homeland Security's technical document, *Community Resilience Indicator Analysis: County-Level Analysis of Commonly Used Indicators from Peer-Reviewed Research* (FEMA, 2022a), using the most current data from the American Community Survey (2020) and Homeland Security Infrastructure Foundation-Level Data (DHS, 2022) and 2020 geographic boundaries. Our metric for resilience focuses only on the demographic characteristics of the community since we are interested in the competition for emergency services (Summer et el., 2017). This is similar to the current CRE estimate which is a measure the risk factors, lack of resources, and weaknesses of the community. Our metric was positively correlated (0.66) with the Census Bureau's CREs. The CRE has greater discriminator power evidenced by the much larger range (maximum value minus minimum value), which makes it more suitable for separating resilient census tracts from those that will potentially require assistance, but could not be used in this demonstration, as noted above.

Similar inconsistencies made it difficult to link data sources using geographic variables. For example, we used shelter facility and emergency service provider data sources from the Homeland Security Infrastructure Foundation-Level Data that included hospitals, Red Cross Chapter Facilities, National Shelter System Facilities, emergency medical service stations, fire stations, and urgent care facilities, to calculate a metric for potential community support. The goal was to place each of these facilities in a Virginia county or independent city. Virginia is divided into 95 counties and 38 independent cities that are considered county-equivalents for census purposes, and in some cases, there is a county and a city with the same name (e.g., Richmond County and Richmond City, each in different locations in Virginia). It was necessary to canonicalize the county and city names (when available), which meant aligning upper and lower cases, removing unnecessary characters, and distinguishing between county and city.[1] (Canonicalizing is the process of converting data that involves more than one representation into a standard approved format.)

The challenge with locating shelter facilities and emergency service providers in a county or independent city was the use of different variables to identify their location (latitude and longitude, address, ZIP Code, FIPS code, and county/city name). In cases where the data source only had a ZIP or FIPS code, a HUD crosswalk was used to link the two codes; in other cases, a crosswalk that linked non-independent cities and towns to counties was used; and in others, a crosswalk that linked FIP codes to counties and independent cities. (See the resilience subfolder in Appendix C for the crosswalks.)

Statistical Analyses

Question 1: Can SNF workers make it to work during a flooding event?

Since proprietary synthetic population data and commercial sector digitized mapping data were used to construct the routes, only an outline of the computational process used to identify the routes is provided. Publicly available data from FEMA were used to estimate flooding risk along a particular route. A general description of the modeling steps and the proprietary data used to assess SNF vulnerability as a function of the nursing staff being unable to report to work due to the transportation infrastructure is provided below (Choupani & Mamdoohi, 2016).

Computational Modules[2]

Here is the basic outline of the process that uses proprietary data that starts at network construction and ends with routes. For more details, see the GitHub repository: Vulnerability of SNFs with Respect to Commuting.

1. Extract network data from HERE (2021 Q1 in this Use Case). It collects about 40 fields across several tables in this data product. These fields are described in HERE's proprietary documentation. BII has access to the documentation, but the others would have to obtain this separately.
2. Process the extracted data to form a network suitable for routing. This includes inference of speed limits for road links where such data is missing.
3. Prepare origin-destination pairs. In this case, the list of location pairs where each pair gives a worker's home and their nursing home to work locations. The person is constructed in the synthetic population pipeline, and residences and workplace are derived through the data fusion process that is used to construct the NSSAC building database.
4. Construct routes using the Quest router.

---

[1] For example, distinguishing county from city when the name is the same could be done using State/County FIPS codes. Richmond County is 51159; Richmond city is 51760.

[2] An alternative method to compute transportation routes and other information using publicly available data is presented in Appendix D.

Once the routes to an SNF were established, Question 1 was formalized as follows. Let W_i be a random variable associated with worker i at an SNF, where $W_i = 1$ indicates the worker is able to commute to work during a flood event and $W_i = 0$ indicates the opposite. Further, let p_i denote $\Pr(W_i = 1)$, i.e., the probability that $W_i = 1$. Then, $\Pr(W_i = 0) = 1 - p_i$.

The expected number of workers at the SNF is given by Equation (1),

$$\begin{aligned} E\left(\sum_i W_i\right) &= \sum_i E(W_i), \qquad \text{(by linearity of expectation)} \\ &= \sum_i p_i \times 1 + (1 - p_i) \times 0 \\ &= \sum_i p_i \end{aligned} \tag{1}$$

Thus, the expected number of workers at an SNF during a flood event is the sum of the probabilities of each of the workers being able to commute to work during a flood event (Equation 1). Depending on the information available, these probabilities can be estimated in multiple ways. For demonstration purposes, a computational model was developed using the following data:

- SNF locations in Virginia from the Centers for Medicare & Medicaid Services (CMS);
- Home locations of workers at each SNF assigned from the BI NSSAC's synthetic population and building database (Beckman et al., 1996; Bhattacharya et al., 2022)
- Virginia road network; and
- FEMA census tract-level riverine and coastal flood risks.

An alternative to the BI NSSAC synthetic population would be to use the Census Bureau's commuting data sets (e.g., Commuting Flows or the LEHD Origin-Destination Employment Statistics) to identify home census tracts for SNF workers and to use the centroid of the home census tract as the starting point for each worker (see Appendix D). Or, with proper access, these assignments and locations could be directly replaced with Census Bureau's linked data and integrated into the analyses.

From the HERE map data (Exhibit 4), the Virginia road network was used to compute each worker's likely route to their SNF using router software. Routers are commonly used within transportation and traffic simulators. The router software used for this demonstration is a highly parallelizable router previously developed in BI NSSAC and is known as the Simba router (Barrett et al., 2013).

Given the routes, the FEMA riverine and coastal flood risks were used to estimate the probabilities p_i in Equation 2. The FEMA risk data provide the riverine and coastal flood risks for each census tract in Virginia. For a route, r, let the sequence of tracts it passes through be denoted by $\tau_j,\ j = 1 \dots n$. Let f_j^r be the riverine flood risk for tract τ_j, and let f_j^c be the corresponding coastal flood risk. We use the following simple model for estimating p_i. If $f_i(r)$ is the total risk for route r taken by worker i, we assume in Equation 2 that,

$$f_i(r) = \max_j f_j^r + \max_j f_j^c, \qquad j = 1 \dots n$$
$$p_i = 1 - \frac{f_i(r)}{100}. \tag{2}$$

In the future we plan to evaluate data from the United States Geological Survey's Digital Elevation Model as an alternative to the FEMA risk data. This had the potential to provide a more accurate modeling of flood risk for individual roads rather than using risk at the census tract level which assumes all roads in the tract have the same risk.

The FEMA technical document *National Risk Index Technical Document* (FEMA, 2021) provides information on how natural hazard risks are calculated. We use these risk estimates which range from 0 to 100, as a proxy for the probability a worker can reach the SNF by dividing by 100. For example, we assume a risk is zero if there is zero probability of not being able to reach the SNF due to an extreme flood event, whereas a risk of 100 indicates the roads are underwater and the probability of not being able to reach the SNF is one. The maximum risks along transportation routes leading to a SNF range from 0 to 47.02 for riverine flooding and 0 to 39.65 for coastal flooding. We assume the combined value of the maximum riverine and coast flood risks along a worker's transportation routes, divided by 100, is the worker's probability of not getting to work during a flooding event.

Since we do not have data on the exact home locations of the nursing staff, we estimated how many could reach the facility by taking a random sample (whose size is the CMS average daily nursing staff[3] for a SNF) from the possible routes identified using the HERE Virginia road network and calculated the average with a 95% nonparametric confidence interval. See Exhibit 3 for a visualization of the estimated nursing staff for the 283 SNFs during a flood event and GitHub for the code. The 283 SNFs used in our research have a total average daily nursing staff of 12,609. Using the above approach, we estimated that 10,005 (95% CI: 9,013, 10,700) or 79% will be able to get work during an extreme flood event. The individual SNF nursing staff percentage who can make it to work ranges from 48% to 93%.

[3] Average Daily Nursing Staff is the daily number of Medical Aides and Technicians, CNAs, LPNs, LPNs with administrative duties, RNs, RNs with administrative duties, RN Director of Nursing averaged over three months.

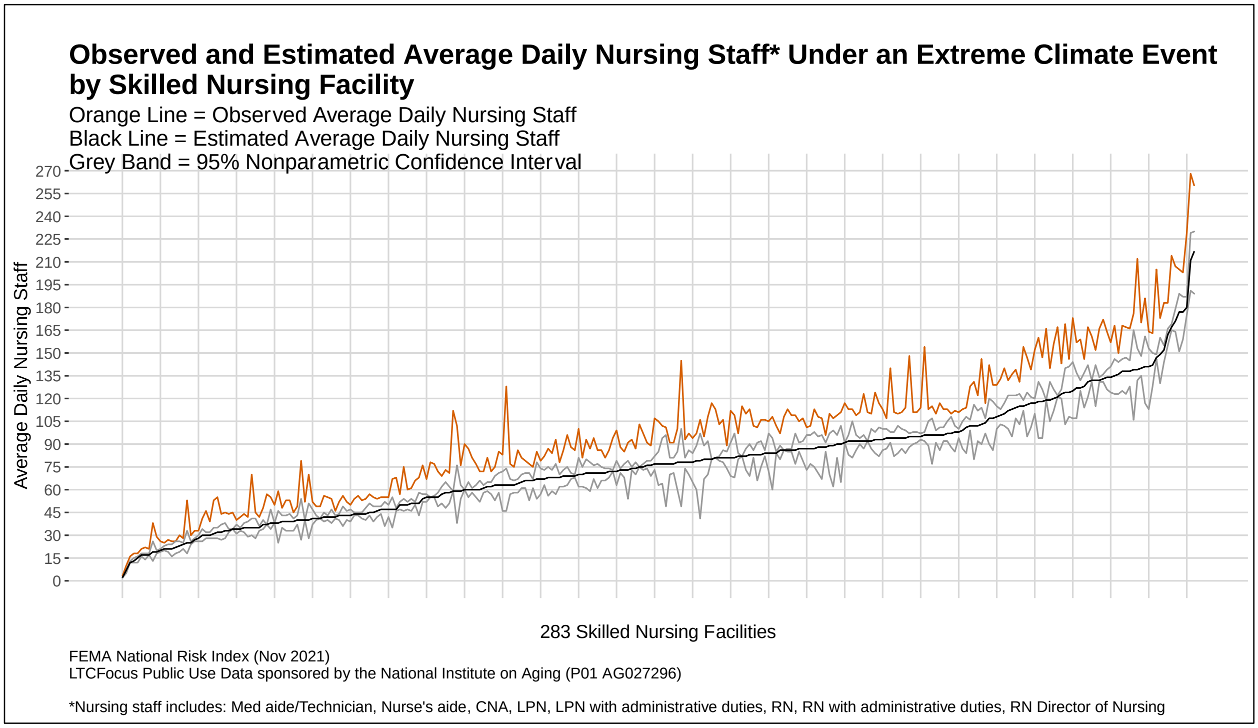

Exhibit 3. Skilled Nursing Facility Average Daily Nursing Staff

Note: The orange line displays the average daily nursing staff numbers submitted quarterly to Center for Medicare & Medicaid Services for each SNF ordered by nursing staff size. The black line displays the estimated average daily nursing staff numbers in the event of an extreme coastal and/or riverine flood event. The grey band is the 95% nonparametric confidence interval.

As an example, in King George County, the skilled nursing facility is Heritage Hall King George facility (Federal Provider Number 495300 in Exhibit 4), located near the Potomac River, which opens to the Chesapeake Bay. According to CMS, the Heritage Hall King George facility has an average daily skilled nursing staff of 41. Using the HERE Virginia road network, we identified 101 routes the staff could use to reach the facility. The combined maximum coastal and riverine flood risks along these routes ranged from 5.6 to 66.7, a random sample of 41 from the 101 routes gives an average probability of reaching the facility of 0.74 with a 95% nonparametric confidence interval of [0.65, 0.80]. These were used to estimate the average number of nursing staff at the facility, 30, during a flood event along with a 95% nonparametric confidence interval [14, 38].

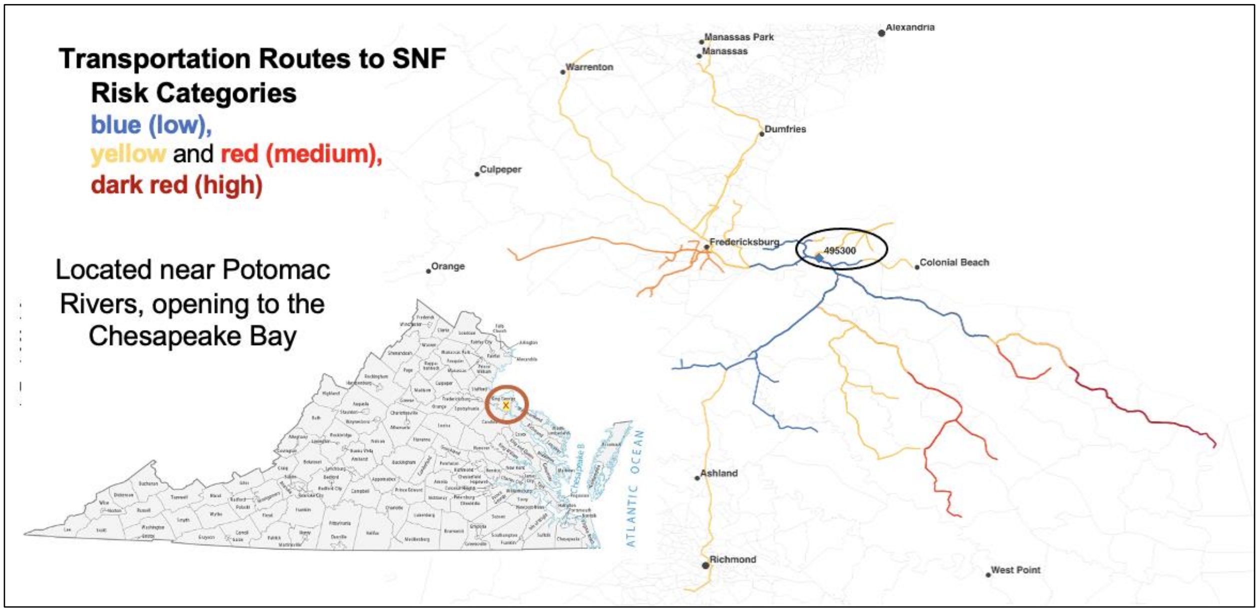

Exhibit 4. An Example of Nursing Staff Routes to Heritage Hall King George SNF

Routes that workers can take to work at Heritage Hall King George SNF FPN 495300 (identified with the black oval). The risk levels of each road are identified with colors, from low risk (blue), medium-low (yellow), orange (medium), red (medium-high), to high risk (dark red). The risk scores are used to calculate the probability of a worker getting to work during an extreme flood event.

Sufficient nursing staff is of major concern to assure resident safety and quality of care. Higher nursing staff numbers are associated with improved activities of daily living independence; less weight loss, less use of antipsychotics; and lower mortality rates (Harrington et al., 2020; White House, 2022). Because of their importance, nursing staff numbers are required to be recorded daily and reported quarterly in the Payroll Based Journal Daily Nurse Staffing Data to the Center for Medicare & Medicaid Services.

Question 2. Are SNF prepared for emergencies?

To address this question, we examined how prepared SNFs are for emergencies using inspection and deficiency data as a proxy for preparedness. SNFs are inspected annually. CMS issues deficiencies to SNFs that fail to meet federal Medicare and Medicaid preparedness standards. Every deficiency is classified into one of twelve categories based on the scope and severity of the deficiency. There are two broad types of non-health related deficiencies – Emergency Preparedness and Fire Life Safety Code deficiencies. See Appendix A for a list of the most frequent deficiencies in each group (CMS, 2022) for Virginia SNFs.

- Emergency Preparedness Deficiencies - There are four elements of emergency preparedness. They cover an emergency plan, policies and procedures, a communication plan, and training and testing.
- Fire Life Safety Code - The set of fire protection requirements are designed to provide a reasonable degree of safety from fire. They cover construction, protection, and operational features designed to provide safety from fire, smoke, and panic.

We calculated a separate Emergency Preparedness and Fire Life Safety Code deficiency index with the goal of combining them to create a single index to measure SNF relative quality to distinguish between high and low-performing SNFs. The computation of the indices has four steps.

1. *Number of Deficiencies*: For each SNF, the total number of deficiencies during the past four years, 2018-2022, was normalized by the number of SNF inspections over the same period.
2. *Time to Resolve Deficiencies*: We next computed the average number of days it took to resolve each deficiency.
3. *Scope and Severity of Deficiencies*: We then transformed the deficiency letter inspection rating for scope and severity to a numerical weight using the CMS technical guide, *Care Compare Nursing Home Five-Star Quality Rating System* (October 2022) and averaged the ratings.
4. The estimates from these three steps were summed to compute separate Emergency Preparedness and Fire Life Safety Code deficiency indices (see Exhibit 5) and are provided for reuse in a .csv file on GitHub.

Exhibit 5, displays an exploratory data analysis for each index. These analyses were done to ensure sufficient variability to discriminate between high and low-performing SNFs. For both indices, we observe two outliers on the Q-Q and box plots. The average Emergency Preparedness deficiency score (left-hand side) is 15, with half of the scores between 0 and 36. Most SNFs do not have Emergency Preparedness deficiencies (Exhibit 5, histogram and box plot). The average Fire Life Safety Code deficiency score (right-hand side) is 65, with half of the SNF scores between 45 and 77. More than a dozen SNFs with Fire Life Safety Code deficiency scores are outside the main body of the data (Exhibit 5, histogram and box plot).

We summed the Emergency Preparedness and Fire Life Safety Code indices and categorized them into high, medium, low, and no deficiencies. The combined index for each SNF is color coded on the maps to identify those with the highest combined deficiency index (orange), those with some deficiencies (dark blue for medium and light blue for low), and those with no deficiencies (grey) (Exhibits 6, 7, B1).

By carefully specifying each step in the development of a metric, this research question provides a roadmap for the key places where modeling choices have been made. By designing proper curation of this type of information, researchers can replicate and/or transparently deviate from these modeling choices to create alternative metrics.

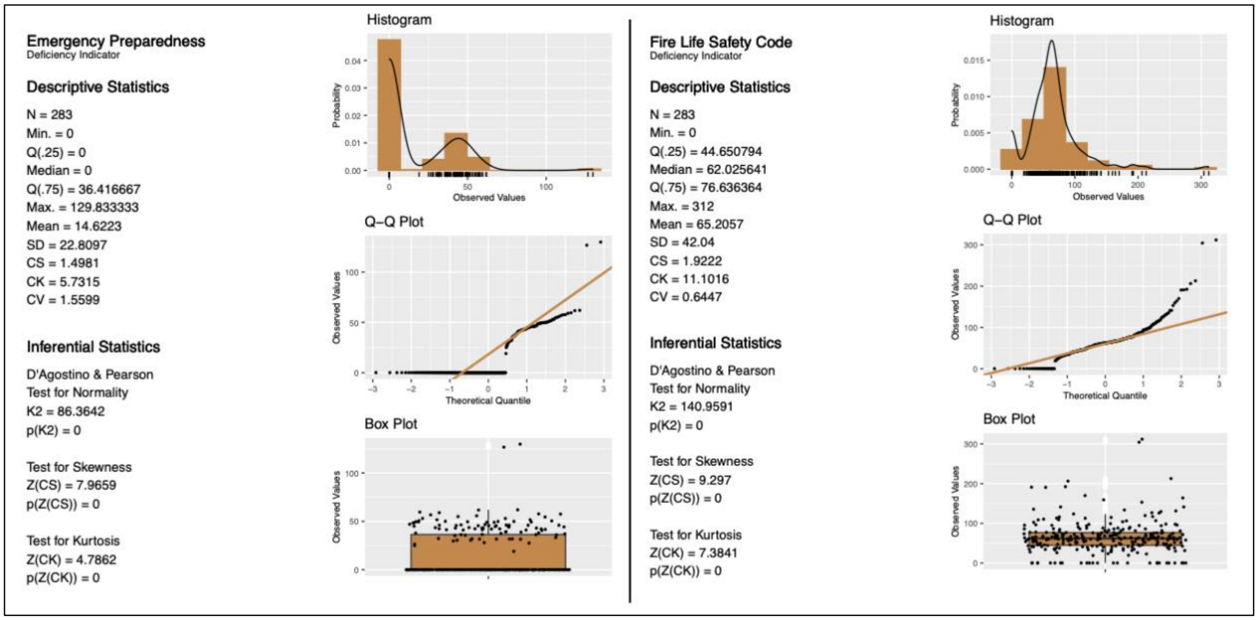

Exhibit 5. Descriptive Statistics for the Emergency Preparedness and Fire Life Safety Code Deficiencies

Legend: Min=minimum value; Q(.25)=lower quartile; Median=middle of the distribution; Q(.75)=upper quartile; Max=maximum value; Mean=average of the distribution; SD=standard deviation; CS=Coefficient of Skewness; CK=Coefficient of Kurtosis; CV=coefficient of variation; K2=Test statistic for H_0: Data=Normal; p(K2)=p-value for K2; Z(CS)=Test Statistic for CS=0 (Normal); p(Z(CS))=p-value for Z(CS); Z(CK)=Test Statistic for CK=3 (Normal); p(Z(CK))=p-value for Z(CK).

Data Sources: The Centers for Medicare & Medicaid Services (CMS) Emergency and Fire Life Safety (Non-Health) Deficiencies https://data.cms.gov/provider-data/dataset/ifjz-ge4w
Inspection Dates https://data.cms.gov/provider-data/dataset/svdt-c123

Question 3: Can the community provide support to SNF during emergency events?

To answer this question, we computed a community resiliency index using the US Census American Community Survey and the guidance provided by the *Homeland Security document Community Resilience Indicator Analysis: County-Level Analysis of Commonly Used Indicators from Peer-Reviewed Research* (FEMA, 2022a). The index was constructed by summing the county or census tract level percentages (depending on the geography being used) for the following variables:

- fraction employed,
- fraction with no disability,
- fraction with a high school diploma or greater,
- fraction of households with at least one vehicle, and
- reverse GINI Index – so all indicators are in a positive direction.

Exhibit 6 displays the combined deficiency indices, Emergency Preparedness + Fire Life Safety Code (Non-Health deficiencies), for each SNF with the choropleth map for the community resilience index at the census tract level. SNFs with the highest number of deficiencies (orange dots) tend to be in lower resilient counties (lighter color counties). We also examined the number of shelter facilities and emergency service providers along with the availability of medical staff per 10,000 residents and constructed isochrones to establish the distance from the SNF to these potential sources of support (Appendix B). Working on this component of the Use Case highlighted the need for cross-agency data, pointing to the utility of future strategic partnering between the Census Bureau, CMS, and FEMA. Even though the data products from this Use Case are publicly available, strategic partnerships would increase awareness of the available data and statistical products and could increase their use in emergency response decisions and practicing evidence-based policy.

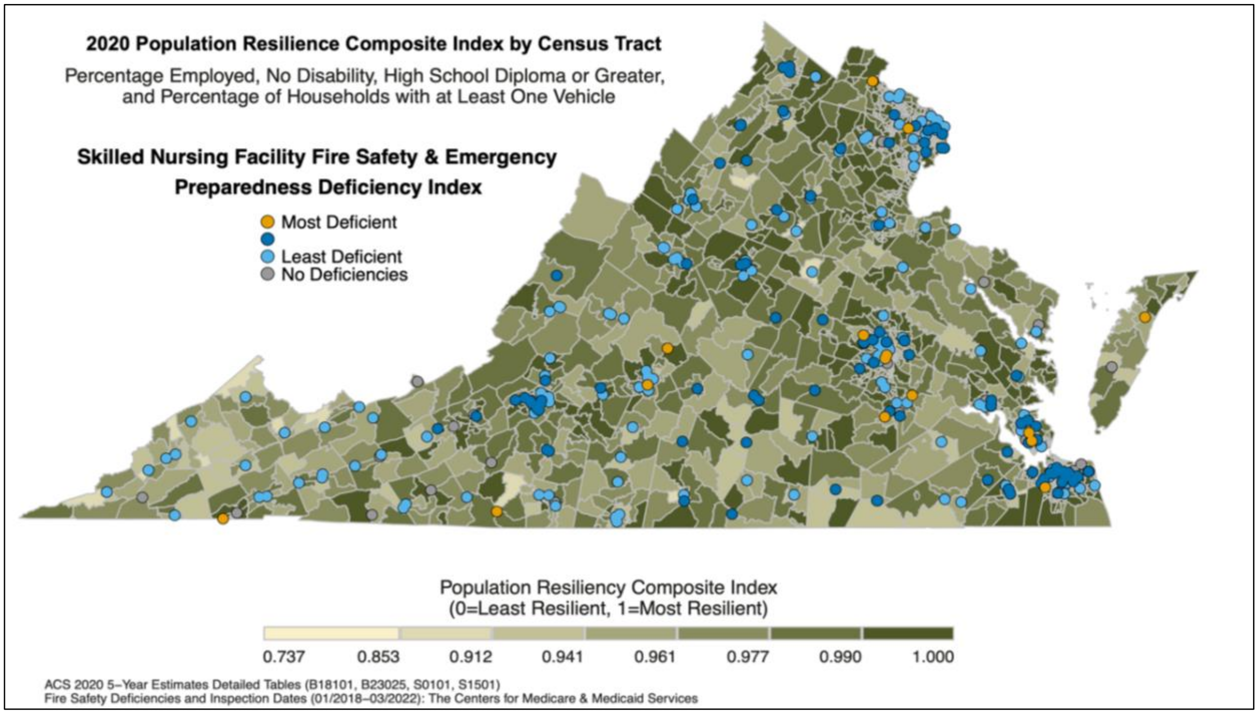

Exhibit 6. 2020 Population Resilience Composite Index for Virginia Census Tracts

The light yellow tracts are those that are least resilient, and the dark green are the most resilient. The location of the 283 SNFs are identified with filled circles, orange circles are those with the highest deficiency index and grey circles are those with no deficiencies.

In addition to describing the population using a resilience index, we also developed an index for the number of shelter facilities and emergency service providers using the Homeland Infrastructure Foundation Level Data (DHS, 2022); and the availability of medical doctors (M.D.) and Doctor of Osteopathic Medicine[4] (O.D.) who provide direct patient care using data from the Health Resources and Services Administration (HRSA), Area Resource Files. Exhibit 7 displays these indices using choropleth maps. Links to data files that contain these indices are provided in the Derived Variable subfolder in Appendix C).

The number of M.D.s and O.D.s is described as a primary care health professional shortage area. HRSA defines these contiguous areas where primary medical care professionals are overutilized,

[4] O.D.s focus on preventive medicine and comprehensive patient care.

excessively distant, or inaccessible to the population of the area under consideration. Exhibit 7 (bottom) shows that approximately one-third of the counties and independent cities have health professional shortage areas across their entire boundary and another 40 percent have shortages within parts of their boundaries.

The number of shelter facilities and emergency service providers includes Red Cross Chapter Facilities, National Shelter System Facilities, emergency medical service stations, fire stations, hospitals, and urgent care facilities per 10,000 population aggregated at the county level (Exhibit 7 top). Some health professional shortage areas occur in counties and independent cities with few shelter facilities and emergency service providers. This information could be useful for SNFs when developing evacuation plans and deciding whether to evacuate. A more detailed assessment of SNFs in the city of Richmond is provided in Appendix B.

Communication and Dissemination

The research is curated and disseminated through a GitHub repository which contains three parts – data, source code, and documents (see Appendix C for a detailed description of the contents). The GitHub Products folder has sub-folders with data tables:

- data for each SNF,
- communities (Virginia counties and independent cities);
- presentations and this report;
- visualizations (maps and statistical graphics); and
- processes which include subfolders for tables of derived variables, and
- an explanation of the synthetic data process.

The CDE approach will enable researchers to re-use these data and not have to start from scratch if they plan to study Skilled Nursing Facilities or other research questions that could use similar data. Researcher also would benefit from having a CDE that provided crosswalks for example, between various location variables such as latitude and longitude, address, ZIP Code, FIPS code, and county/city names. This would keep every researcher from having to construct the same crosswalks, save time, and decrease errors. A collection of data sets that could be used for benchmarking would help facilitate the fitness-for-use assessment. For example, in this Use Case CMS administrative data was used to benchmark synthetic data. This could be extended to data sets that could be used to benchmark statistical products such as metrics for vulnerability and resilience.

Ethics and Equity

As described in the Introduction, there are ethics and equity issues that drew us to develop this Use Case. Here we focus on equity and ethics vis-a-vis the data choices and analyses. With regard to ethical considerations with our data discovery process, fitness-for-purpose evaluation, and analyses, two questions arose:

1. What role does synthetic data have to play, and how do you benchmark it to evaluate fitness-for-purpose?
2. How do you construct and evaluate an index with the goal of identifying vulnerable populations?

Realizing the importance of nursing staff levels, we discussed and questioned whether the synthetic data had biases and were not representative of SNF residents and employees. We benchmarked the synthetic SNF nursing staff numbers against those submitted quarterly to CMS and observed they were biased low, so we decided to use the CMS data. These data were used to estimate the average number of nursing staff that could reach the facility during an extreme flood event (Exhibit 3).

In this Use Case, we were fortunate to have the "truth" to benchmark the synthetic data for the average daily nursing staff at each SNF. But this was not the case for the home locations of the nursing staff, therefore, the synthetic locations were not used since we had no way to benchmark them. Ideally, we would use the actual addresses of SNF employees. Instead, we used a simulation to estimate the average risks over routes leading to the SNF. This approach could be replaced with (or benchmarked against) the Census commuting data sets (e.g., Commuting Flows or the LEHD Origin-Destination Employment Statistics) and the home census tract used as the starting point for each worker (see Appendix D). For the number of nursing staff and their home locations, without a thorough understanding of how the synthetic data were generated, it is impossible to identify potential biases that would result in the inequitable allocation of emergency rescue resources.

How one evaluates the equity of an index is a harder task. Questions that need to be addressed include:

1. How do you select the variables used to construct an indicator to guide an equitable allocation of technical assistance?
2. What relationship between these variables is important?
3. What are the differences across the numerous publicly available resilience estimators? Do some lead to a more equitable allocation of technical assistance in the event of an extreme clime event?
4. How do you validate a resilience estimator?

The technical document, *Community Resilience Indicator Analysis: County-Level Analysis of Commonly Used Indicators from Peer-Reviewed Research* (FEMA, 2020a) identified the 20 most commonly selected variables to construct resilience estimators from peer-reviewed research. Future research will need to take this a step further and validate these indices against past extreme climate events.

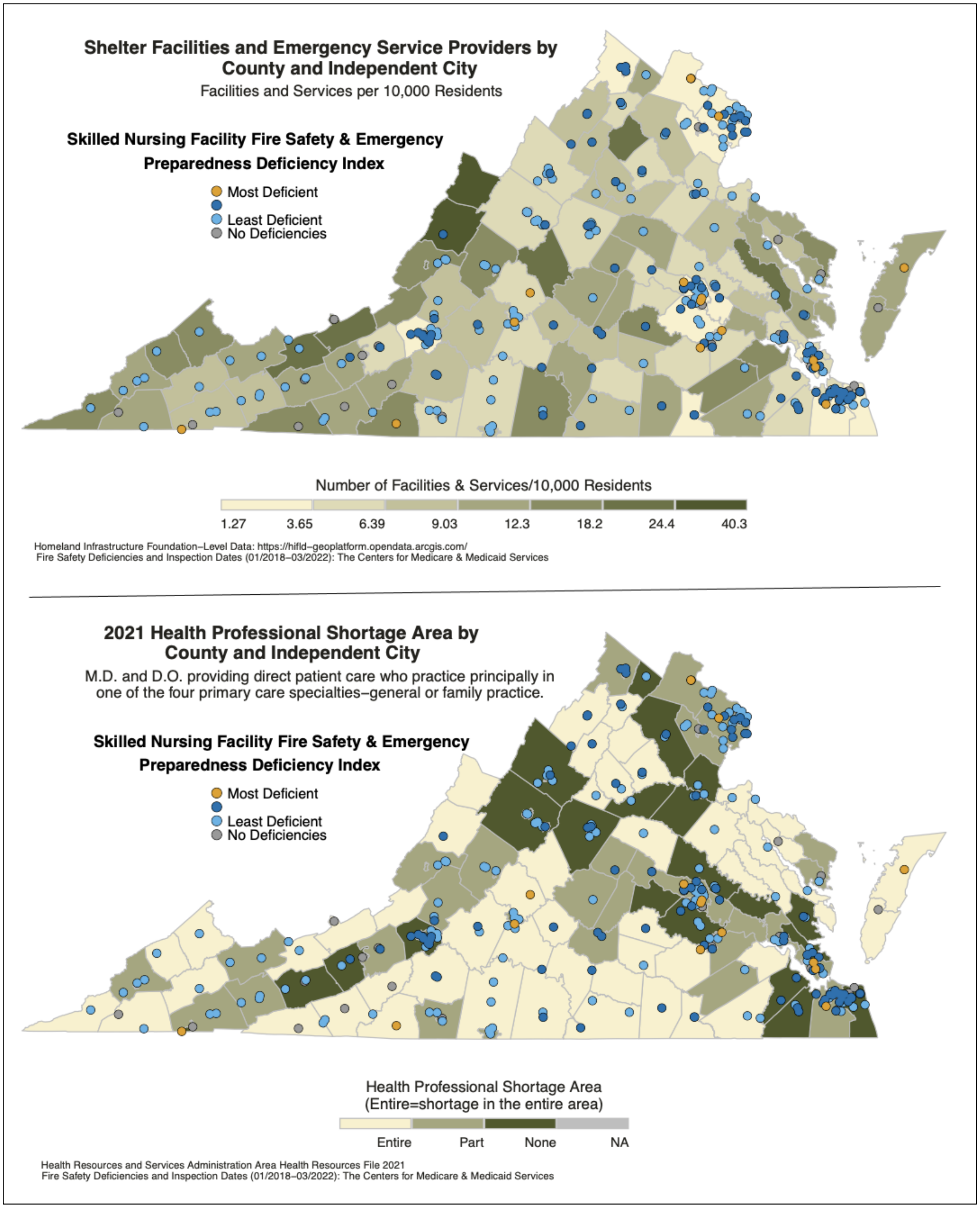

Exhibit 7. Assessment of the number of shelter facilities and emergency service providers per 10,000 population (top chart) and medically underserved areas (bottom chart).

On both maps, the lighter the color, the more in need is the population of shelter facilities and emergency services (top chart) or health professionals (bottom chart). The location of the 283 SNFs are identified with filled circles, orange circles are those with the highest deficiency index and grey circles are those with no deficiencies.

What was learned that applies to CDE development?

This Use Case highlights what we learned about needed capabilities to develop in the CDE. We developed initial criteria to build the CDE use case by use case. A valuable Use Case does not need to meet all criteria but should address many to help inform the CDE capability needs and development. These criteria were evaluated against this skilled nursing demonstration (Exhibit 8).

Exhibit 8: Criteria for selecting and implementing a Use Case and how this demonstration met or did not meet the criteria?

Initial Criteria for Use Case	How did the Use Case meet or did not meet the criteria?
Assess relevancy of the research domain	White House initiative (2022) Numerous articles on SNF issues during emergency events, e.g., Dosa et al., 2008; Hyer et al., 2006; Skarha et al., 2021; Buerhaus et al., 2022
Determine availability of data from multiple sources across multiple frames	Numerous administrative data sources are used. See *Data Discovery* section
Identify computing measurement requirements	Computing measurements developed: • 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 Emergency Preparedness deficiencies • SNF Life Safety Code (fire) deficiencies • SFN combined deficiency index • Community Resilience Indicators • Number of shelter facilities and emergency service providers per 10,000 people. • Health professional shortage areas.
Seek advice from Subject Matter Experts on data sources, the research approach, and implementation	• US Census Bureau Geography Division, Enterprise Leadership Team, and others. • University of Virginia Census Curated Data Enterprise team (Keller et al., 2022) • Nicole Howell, Executive Director of Empowered Aging
Curate and document each step in the CDE process and describe outputs produced	CENSUS SNF Use Case GitHub Repo Structure
Assess processes and data sources with an ethical and equity lens	See *Ethics and Equity* section, for example: • Bias in the synthetic worker counts model and fitness-for-purpose evaluation. • Emergency preparedness and Fire Life Safety Code index evaluation and validation as measures of vulnerability
Develop partnerships to access data from multiple types of organizations (potential is high)	Not developed in this demonstration
Assess viability of proposed platforms	Not developed in this demonstration

Exhibit 8: Criteria for selecting and implementing a Use Case and how this demonstration met or did not meet the criteria?	
Initial Criteria for Use Case	**How did the Use Case meet or did not meet the criteria?**
Identify statistical product gaps and propose new data collection	• In lieu of using proprietary synthetic data, publicly available data could be used to estimate transportation routes, e.g., US Census Bureau Commuting Flows or the LEHD Origin-Destination Employment Statistics or data from the Linked Frames. • CMS SNF ownership data do not identify SNFs with private-equity ownership and were not designed to do so. At present there is no definitive source for this information.

CDE component development and reusability

The Use Case is important for defining and designing the capabilities for the CDE to develop and maintain. For example, the CDE must have capabilities to inform CDE component development and reusability. In this Use Case, the following components are reusable – the data used in the analysis, the code, the integrated Use Case database, and the statistical products (see Exhibit 9).

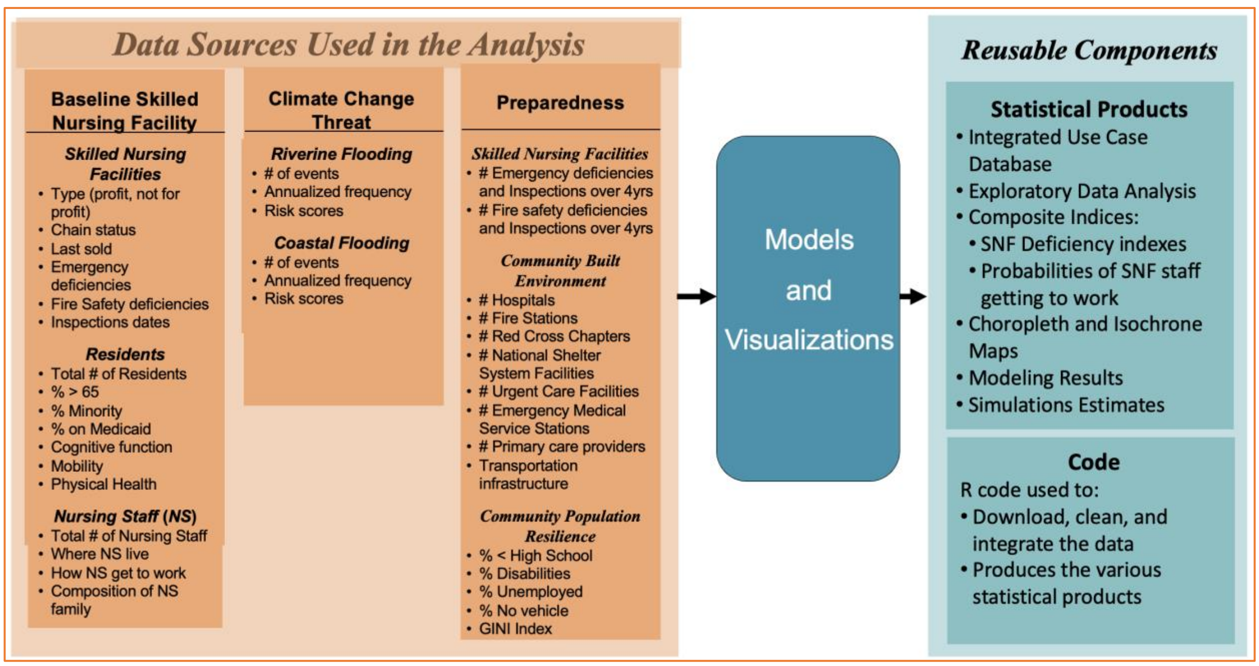

Exhibit 9. CDE component development and reusability

The team discovered the data sources through a literature review and input from experts. For this use case, the data discovery was completed in three phases: (1) create a data picture of SNF owners, workers, and residents, and the communities the SNFs reside in; (2) identify the potential risks of a severe climate event, coastal and riverine flooding; and (3) identify the potential weakness in the SNF's and community's ability to respond. Components developed, statistical products and code, and the data sources used in the analysis can be provide a starting point for the creation of the CDE.

The types of products from a Use Case that can benefit the larger community are only limited by the creativity of the researchers and stakeholders carrying out the Use Case. The products from this Use Case are re-useable code; integrated data sets across diverse topics for each SNF; maps and other visualizations; statistical products such as SNF deficiency indices and various indices that measure community and SNF resilience; the probability of a worker reaching an SNF in the event of extreme flooding; and a GitHub Repo that provides easy access to all these products plus relevant metadata, literature, and government documents and regulations. Conducting this Use Case has been an eye-opening experience as to the amount and quality of publicly available data to address our research questions. The products flowing from a diverse pool of Use Cases can only be identified as the program progresses.

As more Use Cases are prepared, the CDE capabilities will evolve, for example, to address questions about curation. Curation is done in the context of a specific purpose and use. This is an integral part of creating an end-to-end curation process for the CDE. Over time, the Census Bureau may create a curation team that builds the systems for curation by automating as many curation steps as possible. However, it is the developers of the Use Case that must complete the curation as it includes choices considered and decisions made. As learning evolves and the CDE approach becomes more accepted by other agencies, OMB, working across the federal statistical system, may promulgate government-wide standards that make data assembling and linkage easier across agencies. Finally, most of the curation documentation should be able to be made available publicly without confidentiality concerns. Other capabilities to develop are modeling and small-area estimation to improve the geographic granularity of statistical products.

This Curated Data Enterprise Skilled Nursing Facilities demonstration is the first of many Use Cases needed to define and develop CDE capabilities. Underlying each Use Case is the curation process. Curation documents each process step, including decisions that may involve tradeoffs. Curation is to support preserving and adding value to data, including organizing data to facilitate discovery and provide access, documenting data to enable the reuse of the data in scientific and programmatic research, and enhancing the value of the data enterprise through linkages between datasets and by mapping the network of interconnections between datasets, research outputs, researchers, and institutions.

In early work, a decision may be made to use data that are less timely and at greater geographic detail on a particular topic than is optimally desired for the particular purpose or use. By documenting these decisions, future research can continuously target areas to improve CDE statistical products over time. The development of the CDE builds on the Census Bureau's reputation for innovation that occurs decade over decade (Keller & Shipp, 2021).

References

Barrett, C., Bisset, K., Chandan, S., Chen, J., Chungbaek, Y., Eubank, S., Evrenosoglu, Y., Lewis, B., Lum, K., Marathe, A., Marathe, M., Mortveit, H., Parikh, N., Phadke, A. Reed, J., Rivers, C., Saha, S., Stretz, P., Swarup, S., Thorp, J., Vullikanti, A. & Xie, D. (2013, December). Planning and response in the aftermath of a large crisis: An agent-based informatics framework. In *2013 Winter, Simulations Conference (WSC)* (pp. 1515-1526). IEEE.

Beckman, R. J., Baggerly, K. A., & McKay, M. D. (1996). Creating synthetic baseline populations. *Transportation Research Part A: Policy and Practice*, *30*(6), 415-429.

Bhattacharya, P., Chen, J., Hoops, S., Machi, D., Lewis, B., Venkatramanan, S., & Marathe, M. (2022). Data-driven scalable pipeline using national agent-based models for real-time pandemic response and decision support. The International Journal of High Performance Computing Applications, 10943420221127034.

Braun, R. T., Jung, H. Y., Casalino, L. P., Myslinski, Z., & Unruh, M. A. (2021, November). Association of private equity investment in US nursing homes with the quality and cost of care for long-stay residents. In *JAMA Health Forum* (Vol. 2, No. 11, pp. e213817-e213817). American Medical Association.
https://jamanetwork.com/journals/jama-health-forum/fullarticle/2786442

Brown University. (2022). LTCFocus. https://ltcfocus.org/about

Buerhaus, P. I., Staiger, D. O., Auerbach, D. I., Yates, M. C., & Donelan, K. (2022). Nurse Employment During the First Fifteen Months of the COVID-19 Pandemic: Study examines nurse employment trends during the first fifteen months of the COVID-19 pandemic. Health Affairs, 41(1), 79-85.
https://www.healthaffairs.org/doi/abs/10.1377/hlthaff.2021.01289?journalCode=hlthaff

Choupani, A. A., & Mamdoohi, A. R. (2016). Population synthesis using iterative proportional fitting (IPF): A review and future research. *Transportation Research Procedia*, *17*, 223-233.

Centers for Medicare & Medicaid Services (CMS) Glossary (2022a). Skilled Nursing Facility. Centers for Medicare & Medicaid Services.
https://www.cms.gov/glossary?term=skilled+nursing+facility&items_per_page=10&viewmode=grid

Centers for Medicare & Medicaid Services (CMS). (2022b). Design for Care Compare Nursing Home Five-Star Quality Rating System: Technical Users' Guide. October
https://www.cms.gov/medicare/provider-enrollment-and-certification/certificationandcomplianc/downloads/usersguide.pdf

Cutter, S. L., Emrich, C. T., Gall, M., Harrison, S., McCaster, R. R., Derakhshan, S., & Pham, E. (2019). Existing Longitudinal Data and Systems for Measuring the Human Dimensions of Resilience, Health, and Well-Being in the Gulf Coast. *Washington, DC: Gulf Res Program*, 38. https://www.nationalacademies.org/_cache_80ac/content/4885770000234289.pdf

Dosa, D. M., Hyer, K., Brown, L. M., Artenstein, A. W., Polivka-West, L., & Mor, V. (2008). The controversy inherent in managing frail nursing home residents during complex hurricane emergencies. *Journal of the American Medical Directors Association*, 9(8), 599-604. https://pubmed.ncbi.nlm.nih.gov/19083295/

FEMA (2021) National Risk Index Technical Document. November. https://www.fema.gov/sites/default/files/documents/fema_national-risk-index_technical-documentation.pdf

FEMA (2022a) Community Resilience Indicator Analysis: Commonly Used Indicators from Peer-Reviewed Research: Updated for Research Published 2003-2021. US Department of Homeland Security, Federal Emergency Management Agency. September. https://www.fema.gov/sites/default/files/documents/fema_2022-community-resilience-indicator-analysis.pdf

FEMA (2022b) Riverine and Coastal Flooding Maps. US Department of Homeland Security, Federal Emergency Management Agency. https://www.fema.gov/flood-maps

GAO. (September 22, 2023). Nursing Homes: Limitations of Using CMS Data to Identify Private Equity and Other Ownership, *GAO* 23-106163, Washington, D.C. https://www.gao.gov/products/gao-23-106163

Gupta, A., Howell, S. T., Yannelis, C., & Gupta, A. (2021). Does private equity investment in healthcare benefit patients? Evidence from nursing homes (No. w28474). National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w28474/w28474.pdf

Harrington, C., Olney, B., Carrillo, H., & Kang, T. (2012). Nurse staffing and deficiencies in the largest for-profit nursing home chains and chains owned by private equity companies. *Health services research*, *47*(1pt1), 106-128. https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1475-6773.2011.01311.x?casa_token=ondm1FkgmJIAAAAA%3Abv1RyOo0iQ13nzrGTcxOAkP-YJEHafoZeGlnouLoZ_4K2b1Ki78kaFMrBxEIXksltJWfRyWbIvtr5Qc

Harrington, C., Dellefield, M. E., Halifax, E., Fleming, M. L., & Bakerjian, D. (2020). Appropriate nurse staffing levels for US nursing homes. *Health services insights*, *13*, 1178632920934785. https://journals.sagepub.com/doi/pdf/10.1177/11786329209347851

Department of Homeland Security (DHS 2022). Homeland Security Infrastructure Foundation-Level Data. https://hifld-geoplatform.opendata.arcgis.com/

Hyer, K., Brown, L. M., Berman, A., & Polivka-West, L. (2006). Establishing and Refining Hurricane Response Systems for Long-Term Care Facilities: The John A. Hartford Foundation was the lead funder of a Hurricane Summit to focus on the neglected needs of the elderly. *Health Affairs*, *25*(Suppl1), W407-W411. https://www.healthaffairs.org/doi/full/10.1377/hlthaff.25.w407?casa_token=XbJ2j-CdtssAAAAA:USJMJsZq_jlYlQlASQt4O4OYJcq_AOKjpXOx5tTMUIZxoNVXZCzj1_ejtQyLHrnTg6B1BygFuuGZ

Keller, S. A., Shipp, S. S., Schroeder, A. D., & Korkmaz, G. (2020). Doing Data Science: A Framework and Case Study. *Harvard Data Science Review*, 2(1). https://doi.org/10.1162/99608f92.2d83f7f5

Keller, S. & Shipp, S. (2021). A New Approach for the 2030 Decennial Census - A 21st Century Census for a 21st Century Society. Proceedings of the Biocomplexity Institute, TR 2022-022. https://doi.org/10.18130/3was-x496

Keller S, Prewitt K, Thompson J, Jost S, Barrett C, Nusser S, Salvo J, & Shipp S. (2022) A 21st Century Census Curated Data Enterprise. A Bold New Approach to Create Official Statistics, Technical Report BI-2022-1115. Proceedings of the Biocomplexity Institute, University of Virginia; 2022. DOI: https://doi.org/10.18130/r174-yk24.

Masters, K. (November 10, 2021). Virginia nursing homes have lost more than 9,500 employees since the start of the pandemic. Virginia Mercury. https://www.virginiamercury.com/2021/11/10/virginia-nursing-homes-have-lost-more-than-9500-employees-since-the-start-of-the-pandemic/#:~:text=Since%20February%202020%2C%20the%20state's,care%20that%20residents%20are%20receiving.

Mortveit, H., Swarup, S., Xie, D, Baek, H. Wilson, M, Venkatramanan, S., (2019) Detailed Overview and Description of Components in U.S. Pipeline 2019, NSSAC Technical Report: #2019-025, updated February 2023. BI-2022-1287 https://biocomplexity.virginia.edu/our-research/institute-publications/detailed-overview-and-description-components-us-pipeline-2019

Mortveit, H., Xie, D., Marathe, M., (2023) NSSAC Building Knowledge Base: Modeling and Implementation, NSSAC Technical Report. No. 2021-016. BI-2023-13, https://biocomplexity.virginia.edu/our-research/institute-publications/nssac-building-knowledge-base-modeling-and-implementation

Santos, R. (2023). A Look Ahead: 2023. US Census Bureau Newsroom Blogs. January. https://www.census.gov/newsroom/blogs/director/2023/01/a-look-ahead-2023.html

Skarha, J., Gordon, L., Sakib, N., June, J., Jester, D. J., Peterson, L. J., Andel, R., & Dosa, D. M. (2021, November). Association of power outage with mortality and hospitalizations among Florida nursing home residents after Hurricane Irma. In JAMA Health Forum (Vol. 2, No. 11, pp. e213900-e213900). American Medical Association. https://jamanetwork.com/journals/jama-health-forum/fullarticle/2786665

Shipp, S., Salvo, J., & Zhang, S. (2022). Understanding the State of Nursing Homes and Residential Care Facilities - A Use Case to Support the Development of the Curated Data Enterprise. Technical Report 2022-028. Proceedings of the Biocomplexity Institute, University of Virginia. https://doi.org/10.18130/5d6x-f676

Summers, J. K., Smith, L. M., Harwell, L. C., & Buck, K. D. (2017). Conceptualizing holistic community resilience to climate events: Foundation for a climate resilience screening index. GeoHealth, 1(4), 151-164.

Thieme, M. (2022). Integration Phase One Plan for the Census Bureau's Four Key Initiatives Transformation. US Census Bureau newsroom blogs. October. https://www.census.gov/newsroom/blogs/research-matters/2022/10/technology-transformation.html

US Census Bureau, (2021). 2019 Community Resilience Estimates Quick Guide. Department of Commerce. August. https://www2.census.gov/programs-surveys/demo/technical-documentation/community-resilience/cre_quickguide_2019.pdf

US Department of Health and Human Services (HHS) (2022). HealthData.gov COVID-19 Reported Patient Impact and Hospital Capacity by State. Retrieved on November 13, 2022. https://healthdata.gov/dataset/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/6xf2-c3ie

Virginia Hospital & Healthcare Association, (2022). Map of VHHA Hospital and Health System Members. Retrieved on November 13, 2022, https://www.vhha.com/about/virginia-hospitals/

White House (2022). Fact Sheet: Protecting Seniors by Improving Safety and Quality of Care in the Nation's Nursing Homes. February. https://www.whitehouse.gov/briefing-room/statements-releases/2022/02/28/fact-sheet-protecting-seniors-and-people-with-disabilities-by-improving-safety-and-quality-of-care-in-the-nations-nursing-homes/

Appendices

Appendix A. Most Frequent Virginia Skilled Nursing Facilities Non-Health Deficiencies

***Emergency Preparedness Deficiencies* (2022)**	**Freq**
Establish staff and initial training requirements	53
Establish emergency prep training and testing	47
Address subsistence needs for staff and patients	45
Address patient/client population and determine types of services needed	41
Establish roles under a Waiver declared by secretary	39
Conduct testing and exercise requirements	35
Provide family notifications of emergency plan	35
Establish policies and procedures for volunteers	34
Establish procedures for tracking staff and patients during an emergency	31
Establish policies and procedures for medical documentation	24
Conduct risk assessment and an All-Hazards approach	23
Establish methods for sharing information	23
Develop Emergency Preparedness policies and procedures	22
Establish policies and procedures for sheltering	20
List the names and contact information of those in the facility	19
Provide primary/alternate means for communication	19

***Life Safety Code Deficiencies* (2022)**	**Freq**
Inspect, test, and maintain automatic sprinkler systems	322
Install corridor and hallway doors that block smoke	209
Ensure proper usage of power strips and extension cords	181
Have properly installed electrical wiring and gas equipment	172
Have approved installation, maintenance and testing program for fire alarm systems	170
Meet other general requirements that are deficient	163
Have generator or other power source capable of supplying service within 10 seconds	159
Ensure smoke barriers are constructed to a 1-hour fire resistance rating	146
Add doors in an exit area that do not require the use of a key from the exit side unless in case of special locking arrangements	137
Keep aisles, corridors, and exits free of obstruction in case of emergency	122
Ensure that special areas are constructed so that walls can resist fire for 1- hour or have an approved fire extinguishing system	120
Ensure heating and ventilation systems that have been properly installed according to the manufacturer's instructions	117
To conduct inspection, testing and maintenance of fire doors by qualified individuals	114
Have properly located and lighted Exit signs	108
Have proper medical gas storage and administration areas	93
Provide properly protected cooking facilities	87

DATA SOURCE: The Centers for Medicare & Medicaid Services Non-Health Deficiencies
https://data.cms.gov/provider-data/dataset/ifjz-ge4w

Appendix B. Census CDE Demonstration Example - Skilled Nursing Facilities in City of Richmond, Virginia

The city of Richmond has six Skilled Nursing Facilities (SNF). These are described in this appendix to provide an example of the power of bringing data together in an integrated Curated Data Enterprise environment.

Exhibit B-1 displays the resilience index (see the section for Question 3 for the definition of the index) for the census tracts in the city of Richmond, Virginia. The resilience index across census tracts is heterogeneous, from a low of 0.61 in the south and east to a high of 0.95 in the northwest. The low population resilience indicators align with the parts of the city identified as having primary care health professional shortages. With regard to the ratio of shelter facilities and emergency service providers per 10,000 population (see county-level visual Exhibit 7), the city is in the lower quartile among Virginia counties and independent cities (the range is 1.27 to 40.32). The city has 68 shelter facilities and emergency service providers for a population of 226,610 (or 3 per 10,000 people).

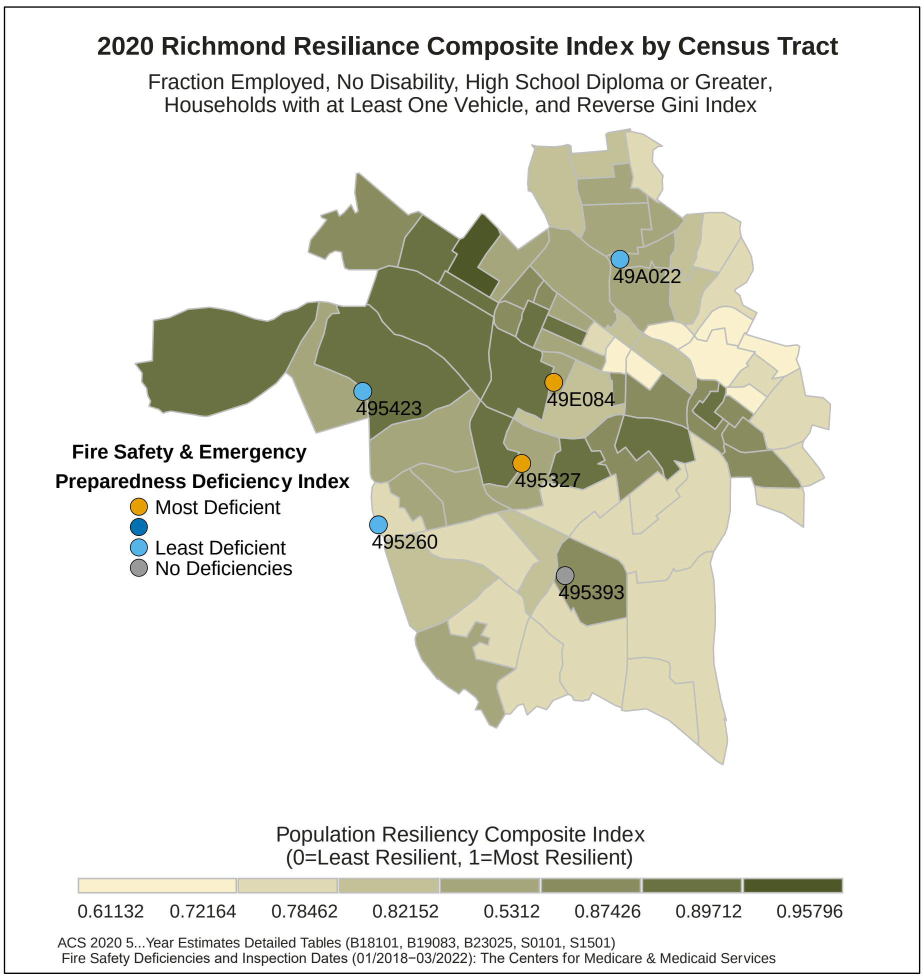

Exhibit B-1. Resilience Index for Richmond, Virginia Census Tracts, 2020. The shading indicates the least resilient census tracts (light yellow) to the most resilient (dark green). The map is overlaid with the deficiency score for each of the 6 SNFs in the city. SNFs with the most deficiencies are colored orange and those with no deficiencies are gray.

Exhibit B-2: Risk Indicators for Six SNFs in the City of Richmond						
City of Richmond & SNF Indicators	**Census Tracts**					
SNF Federal Provider Number	495260	495327	495393	495423	49A022	49E084
Maximum Riverine Flood Risk	13.80	15.51	14.57	16.39	14.41	14.48
Maximum Coastal Flood Risk	2.91	2.33	1.66	2.28	1.62	2.30
% Nursing Staff Loss if Flood Event	18%	20%	17%	21%	15%	17%
SNF Deficiency Index	Low	High	None	Low	Low	High
SNF Ownership Type	For-Profit Corp.	For-Profit Corp.	Govt. State	For-profit P'ship	Non-profit Corp.	Non-profit Corp.

Data Sources: FEMA National Risk Index (Nov 2021) and computations by the Biocomplexity team.
https://hazards.fema.gov/nri/data-resources
The first two CMS data sources are used to compute SNF deficiency indexes. The third source provides SNF information, e.g., provider number, location, ownership type, etc.

- Non-Health Deficiencies https://data.cms.gov/provider-data/dataset/ifjz-ge4w
- Inspection Dates https://data.cms.gov/provider-data/dataset/svdt-c123
- Provider Information https://data.cms.gov/provider-data/dataset/4pq5-n9py

Exhibit B-2 presents the characteristics of the six SNFs in the city of Richmond - the river and coastal flood risk scores, the percent nursing staff loss if there is a flood event, the SNF deficiency index, and the ownership type. There are relatively high riverine flood risks and low coastal flood risks. On average, 18% of the SNF workers will not make it to work if an extreme flooding event occurs. SNF 495327 (highlighted in olive green) had many emergency preparedness deficiencies but not of immediate jeopardy to residents and staff. Their deficiencies included the following:

- Emergency Preparedness: failed to conduct tests and exercises for requirements; failed to establish policies and procedures for sheltering; and
- Fire Life Safety Code (fire): failed to test and maintain the automatic sprinkler and fire alarm systems.

None of the SNFs changed ownership in 2021. There are three for-profits, two not-for-profits, and one government owned SNFs. For-profit private equity owners focus on cost efficiency and profits which researchers have found results in increases in resident mortality, hospitalizations, and emergency room visits; increases in the number of deficiencies, especially serious deficiencies. At the same time, spending per patient increases by 11% (Braun et al., 2021; Gupta et al., 2021; Harrington et al., 2012). To identify SNFs owned by private equity firms, researchers identify SNFs owned by the largest for-profit chains through Internet searches and company reports and match this information with federal secondary data for each ownership group or with private equity ownership data purchased from Pitchbook, Levin Associates: Healthcare M&A and Senior Housing M&A database, or S&P Capital IQ. This is both a time- and resource- intensive method since there is no definitive data source. In a recent research report (GAO, September 2023) the GAO concluded the CMS SNF ownership data does not provide the means to identify private equity ownership nor was it designed to. The use of these private data sources reinforces the purpose of the Use Case - to illuminate the challenges in creating statistical products and what the Census Bureau would need to consider.

The number of shelter facilities and emergency service providers can be disaggregated, and their locations plotted to provide information regarding the proximity of shelters and emergency services to SNFs. These data provided under Homeland Security Infrastructure Foundation-Level Data (DHS, 2022) are web-scraped and would therefore need to go through a more rigorous inventory process before being used for evacuation decisions. An example of how these data could be displayed is shown in Exhibit B-3, where an isochrone map is used to display the location of SNF 495327 (orange marker) and the shelter facility and emergency service provider locations (black circles) in the Richmond metropolitan area. The isochrones (red lines circling the SNF) on the map delineate drive times of 15 and 30 minutes.

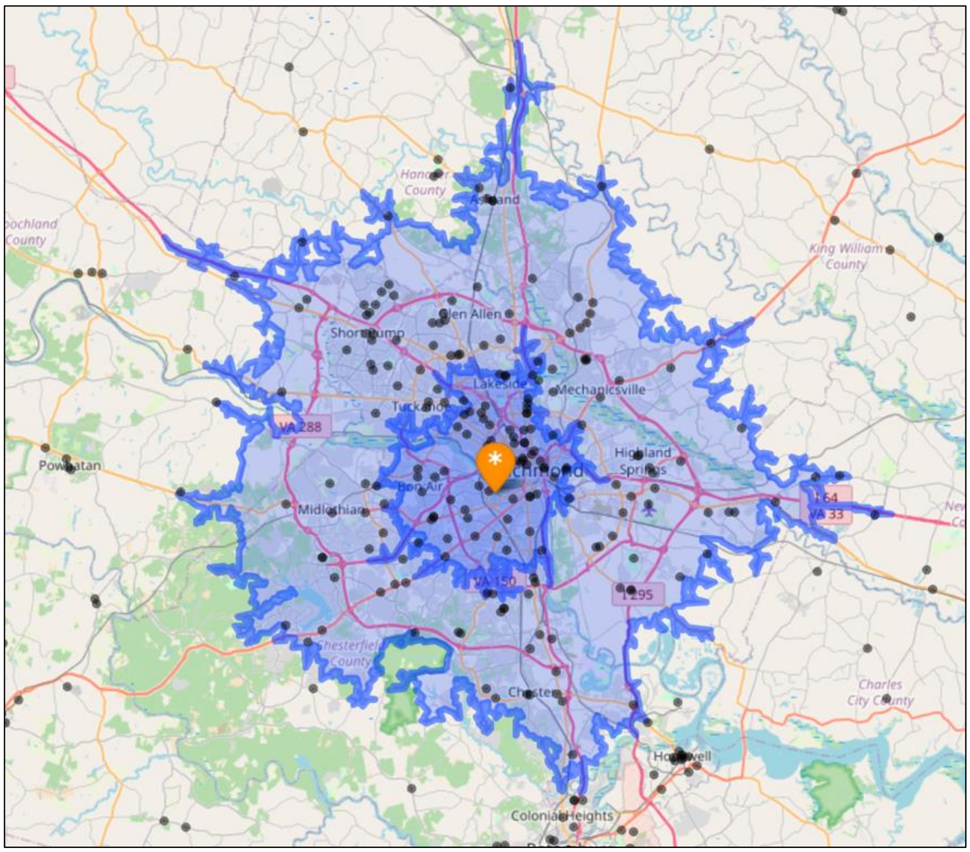

Exhibit B-3. **Isochrone Map with SNF 495327 (orange) and Emergency Resources (black circles)**

The orange marker is the location of SNF 495327, and the black circles are the location of hospitals, Red Cross Chapter Facilities, National Shelter System Facilities, emergency medical service stations, fire stations, and urgent care facilities. The isochrones outline a 15- and 30-minute drive from the SNF.

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
        - File: cms_five_star_users_guide_2022-01.pdf
        - File: cms_primary_data_dictionary.xlsx
    - Subfolder **Nursing Staff**
        - File: va_cms_pbj_puf_payroll_nursing_staff_2019-Q4.csv
    - ✓ Subfolder: **Meta Data**
        - File: cms_payroll_based_journal_daily_nurse_staffing_data_dictionary_2022-08.pdf
        - File: pbj_puf_documentation_042022.pdf
    - Subfolder **Residents**
        - File: va_ltcfocus_2020.xlsx
    - ✓ Subfolder: **Meta Data**
        - File: LTCFocus_data_dictionary_merged.xlsx
        - File: hrs_ltcfocus_data _manual_042016_v2.pdf

- Subfolder: **Community**
  - Subfolder: **Climate Change Risk**
    - File: virginia_fema_nri_county.csv
    - File: virginia_fema_nri_census_tract.csv
  - ✓ Subfolder: **Meta Data**
    - File: fema_nri_data_dictionary.xlsx
  - Subfolder: **Resilience**
    - File: va_hifld_american_red_cross_chapter_facilities_2022.csv
    - File: va_hilfd_hospitals_2022.csv
    - File: va_hilfd_emergency_medical_service_stations_2022.csv
    - File: va_hilfd_urgent_care_facilities_2018.csv
    - File: va_hilfd_national_shelter_system_facilities_2022.csv
    - File: va_hilfd_fire_stations_2020.csv
    - File: va_hrsa_ahrf_2021.csv
    - File: va_total_county_population_2020.csv
    - File: va_town_county_crosswalk.csv
    - File: va_zipcode_town_county_crosswalk.csv
  - ✓ Subfolder: **Meta Data**
    - File: hrsa_ahrf_data_dictionary_2020-2021.pdf
    - File: hifld_data_catalog-2022-Q3.xlsx
    - File: fema_resilience_analysis_and_planning_tool_2021-04.pdf
  - Subfolder: **Demographic** – files are pulled down from the CENSUS API within the Rcode; see files VA_Population_Resilience_Index_Census_Tract.R and VA_Population_Resilience_Index_County.R
    - DOC/Census/ACS/**B18101 Sex by Age by Disability**
    - DOC/Census/ACS/**B19083 Gini Index of Income Inequality**
    - DOC/Census/ACS/**B23025 Employment Status for the Populations 16 Years and Over**
    - DOC/Census/ACS/**S0101 Age and Sex**
    - DOC/Census/ACS/**S1501 Educational Attainment**

2. ***Source Code Folder***

- Subfolder: **Discovery**
    - File: EDAshape.R
      Output: eda_emergency_preparedness_deficiency_index.pdf
      Output: eda_fire_life_safe_code_deficiency_index.pdf
    - File: EDA_flatviolin.R
      Output: eda_snf_deficiency_index_by_hazard.pdf
    - File: EDA_boxplots.R
      Output: county_population_indicator_resilience_box_plots.pdf
    - File: EDA_Choropleths_Resilience_Indicator_Variables.R
      Output: census_tract_population_16_and_over_unemployed_choropleth.pdf
      Output: census_tract_population_no_hs_diploma_choropleth.p
      Output: census_tract_population_with_disability_choropleth.pdf
      Output: county_old_age_dependency_choropleth.pdf
    - File: Richmond_Isochrones_Map_All_Facilities.R

- File: Richmond_Isochrones_Map_Hospitals.R

- Subfolder: **Analyses**
  - File: SNF_Deficiency_Index.R
    Output: va_snf_deficiency_indices_k_e.csv
  - File: VA_Population_Resilience_Index_Census_Tract.R
    Output: census_tract_population_resilience_index_choropleth.pdf
    Output: va_census_tract_population_resilience.csv
  - File: VA_Population_Resilience_Index_County.R
    Output: county_population_resilience_index_with_gini_choropleth.pdf
    Output: county_population_indicator_resilience_box_plots.pdf
    Output: va_county_population_resilience.csv
  - File: VA_Probability_of_Getting_to_SNF.R
    Output:
    va_snf_estimated_average_daily_nursing_staff_during_extreme_flood_event.csv
    Output: snf_estimate_daily_nursing_staff_during_climate_event.pdf
  - File: County_Assets_Workers.R
    Output: county_health_professional_shortage_area_choropleth.pdf
  - File: County_Assets_Infrastructure.R
    Output:
    county_shelter_facilities_and_emergency_service_provider_choropleth.pdf
    Output: va_county_shelter_and_emergency_facility_resilience_index.csv
  - File: Richmond_Population_Resilience_Index.R
    Output:
    census_tract_richmond_population_resilience_index_choropleth.pdf

3. ***Documents Folder***

- Subfolder: **Products**
  - Subfolder: **Visualizations**
    - File: census_tract_population_16_and_over_unemployed_choropleth.pdf
    - File: census_tract_population_no_hs_diploma_choropleth.pdf
    - File: census_tract_population_with_disability_choropleth.pdf
    - File: census_tract_population_resilience_index_choropleth.pdf
    - File: census_tract_richmond_population_resilience_index_choropleth.pdf
    - File: county_health_professional_shortage_area_choropleth.pdf
    - File: county_old_age_dependency_choropleth.pdf
    - File: county_population_indicator_resilience_box_plots.pdf
    - File: county_population_resilience_index_with_gini_choropleth.pdf
    - File: county_shelter_facilities_and_emergency_service_provider_choropleth.pdf
    - File: eda_emergency_preparedness_deficiency_index.pdf
    - File: eda_fire_life_safe_code_deficiency_index.pdf
    - File: snf_estimate_daily_nursing_staff_during_climate_event.pdf
  - Subfolder: **Processes**
    - Subfolder: **Commute Vulnerability**
      - File: algorithm. md
    - Subfolder: **Derived Variables**

        - File: va_census_tract_population_resilience.csv
        - File: va_county_population_resilience.csv
        - File: va_county_shelter_and_emergency_facility_resilience_index.csv
        - File: va_snf_deficiency_indices_k_e.csv
        - File: va_snf_estimated_average_daily_nursing_staff_during_extreme_flood_event.csv
        - File: va_snf_prob_worker_not_getting_to_work.csv
        - File: va_snf_to_si_location_mapping.csv
        - File: va_snf_flood_risks.csv
    - Subfolder: **Data Tables**
        - Subfolder: **Skilled Nursing Facility**
            - File: va_snf_data_dictionary.csv
            - File: va_snf_table.csv
        - Subfolder: **Community**
            - File: va_county_city_data_dictionary.csv
            - File: va_county_city_data_table.csv
    - Subfolder: **Presentations and Reports**
        - File: UVA_CENSUS_Demo_2022OCT31.pdf
        - File: UVA_FCSM_2022OCT27.pdf
        - File: UVA SNF CENSUS USE CASE DRAFT 2023MAR07.pdf
- Subfolder: **Literature**
    - Subfolder: **Deficiencies**
        - File: Castle_2011.pdf
        - File: Harrington_2012.pdf
        - File: Li_2015.pdf
        - File: OIF_Trends in SNH Deficiencies_2019.pdf
        - File: REPORT-Nursing-Facilities-Staffing-Residents-and-Facility-Deficiencies-2009-2016.pdf
    - Subfolder: **Emergency Preparedness**
        - File: Laditka_2008.pdf
        - File: Benevolenza_2019.pdf
        - File: Brunkard_2013.pdf
        - File: Dosa_2006.pdf
        - File: GAO_2006.pdf
        - File: HHS_2006.pdf
        - File: HHS_2012.pdf
        - File: Willoughby_2017.pdf
        - File: cms_federal_government_emergency_preparedness_for_ltc.pdf
        - File: fcc_ndms_guide_2018-06.pdf
    - Subfolder: **Nurse Staffing Levels**
        - File: Harrington_2020.pdf
    - Subfolder: **Resilience**
        - File: Keim_2008.pdf
        - File: fema_community-resilience-indicator-analysis_2020.pdf
    - Subfolder: **Ownership**
        - File: Pradhan_2014.pdf

        - File: Bos_2017.pdf
        - File: Braun_2020.pdf
        - File: Braun_2021.pdf
        - File: Gupta_2021.pdf
        - File: americans_for_financial_reform_2020.pdf
    - Subfolder: **Quality**
        - File: Li_2020.pdf
        - File: NAS_2022.pdf
        - File: FACT SHEET_ Protecting Seniors by Improving Safety and Quality of Care in the Nation’s Nursing Homes _ The White House.pdf
    - Subfolder: **Synthetic Data**
        - File: NSSAC Technical Report 2019-025.pdf
        - File: NSSAC Technical Report 2021-016.pdf

Appendix D. Alternative Publicly Available Data Sources to Compute Transportation Routes

In the Use Case described in this report, proprietary data are used to compute transportation routes and flood risk maps. Other data, such as household composition and traffic congestion, could also be included. There are publicly available data sources that could be used.

- *FEMA data:* replace or augment FEMA flood risk maps (FEMA, 2022b) with data such as NOAA flood maps or dedicated hydrological models.
- *Household structure:* one may choose to add a worker's household composition (e.g., single parent) into the estimates of that worker's ability to report to work. One may also consider household income.
- *Route construction:* One may also incorporate damage from flooding for the routing over a transportation network. This would offer route alternatives that could factor into a worker's decision-making or ability to report to work. In addition, network loading could be modeled to estimate traffic congestion due to the impacts on the transportation infrastructure. Public transportation, if available, may also be considered.
- *A*n alternative to using the synthetic population DP-US-VA-2.0 (see *Question 1, Statistical Analysis section in paper)* to identify the worker population is as follows: for each SNF, *f*, combine the worker count for *f* obtained from federal data and LODES/ACS commute data.
    - For *f* and its count of workers, one may estimate the distribution of workers by home county (or census tract).
    - Sampling from this distribution, one may:
        1. Sample the appropriate number of residence locations within the respective home counties and use this as the origin for determining the routes to/from the SNF, and/or
        2. Identify workers within the synthetic population for Virginia (DP-US-VA-2.0) with a NAICS designation corresponding to care facilitation.
    - Either approach (1 or 2) would allow one to match worker counts accurately.