Carnegie Mellon University

CMU MSP Program ✖ LivaNova DRE Patients Estimation

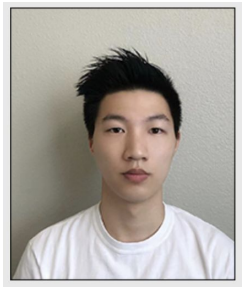
27 April 2022

Zhuoheng Han
Sifeng Li
Ziyan (Cecilia) Xia
Yuqing Xu

Agenda

- Team Introduction
- Executive Overview
- Methodology
- Results
- Recommendations

Team Introduction

			
Zhuoheng Han	Sifeng Li	Ziyan (Cecilia) Xia	Yuqing Xu
Role: Technical lead	Role: Documentation, Presenter	Role: Statistician	Role: Statistician
Interests: Watching NBA	Interests: TV shows	Interests: Oyster	Interests: Reading novels
Project Sponsors - LivaNova: Kayla Frisoli, Cheyenne Ehman *Faculty Advisor - Carnegie Mellon University: Nicole Goodman*			

Executive Overview

Goal: Use public data sources to estimate the number of drug resistant epilepsy (DRE) patients by both hospital and zip code

So what? Save cost of paying external data providers; Can be used to help optimize selling LivaNova VNS devices to targeted hospitals treating DRE patients

Now what? LivaNova has a reproducible dataset and models that can be used to better understand the market of their VNS devices for DRE patients

publicly available sources

6

Collected 6 data sets from 4 major sources

ZIP codes

33k

Estimate number of DRE patients in 33k ZIP codes

hospitals

5k

Estimate number of DRE patients in 5k hospitals

Methodology Overview

- Collect and clean data
- DRE patients by ZIP code
  - Model DRE patients by ZIP code
  - Visualize results
- DRE patients by hospital
  - Prepare data
    - Merge disparate datasets
    - Impute missing values
  - Model DRE patients by hospital
  - Visualize results

DRE patients by ZIP code

DRE patients by hospital

Data Collection

Demographics:	**Hospitals:**	**Geographic:**
• Number of epilepsy patients by state	• Hospitals with ZIP code, • Number of neuro specialists, • Number of discharges from seizures, • Total number of beds, • CEC status	• Population, • Position and distance information by ZIP code
https://www.cdc.gov/mmwr/volumes/66/wr/mm6631a1.htm	https://healthdata.gov/Hospital/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/anag-cw7u/data https://data.cms.gov/provider-summary-by-type-of-service/medicare-inpatient-hospitals/medicare-inpatient-hospitals-by-provider-and-service https://www.naec-epilepsy.org/about-epilepsy-centers/find-an-epilepsy-center https://www.cms.gov/Regulations-and-Guidance/Administrative-Simplification/NationalProvIdentStand/DataDissemination	https://simplemaps.com/data/us-zips

DRE Patients by ZIP Code

Estimating DRE Patients by ZIP Code

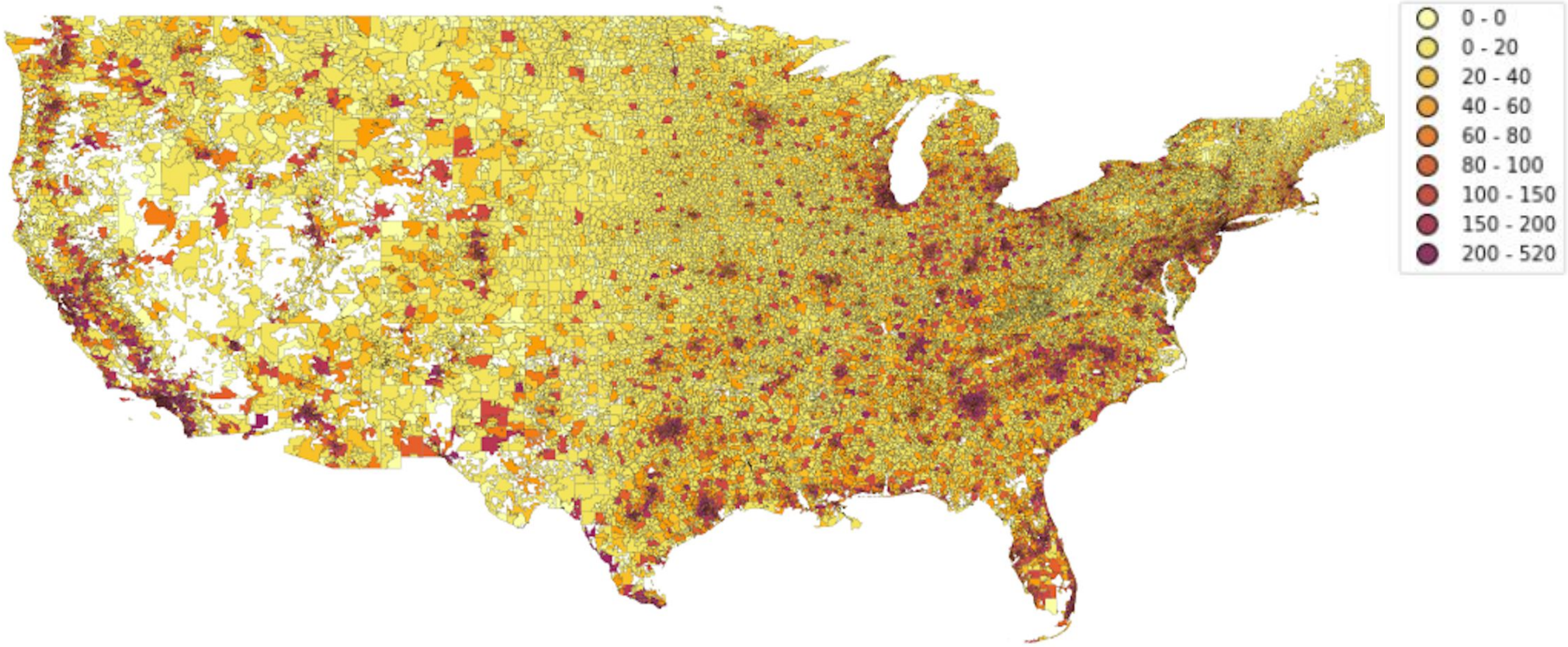
PA Example

$$\text{DRE by ZIP Code} = \frac{\text{ZIP Population}}{\text{State Population}} \times \text{State Epilepsy Population} \times \text{DRE Multiplier } (⅓)$$

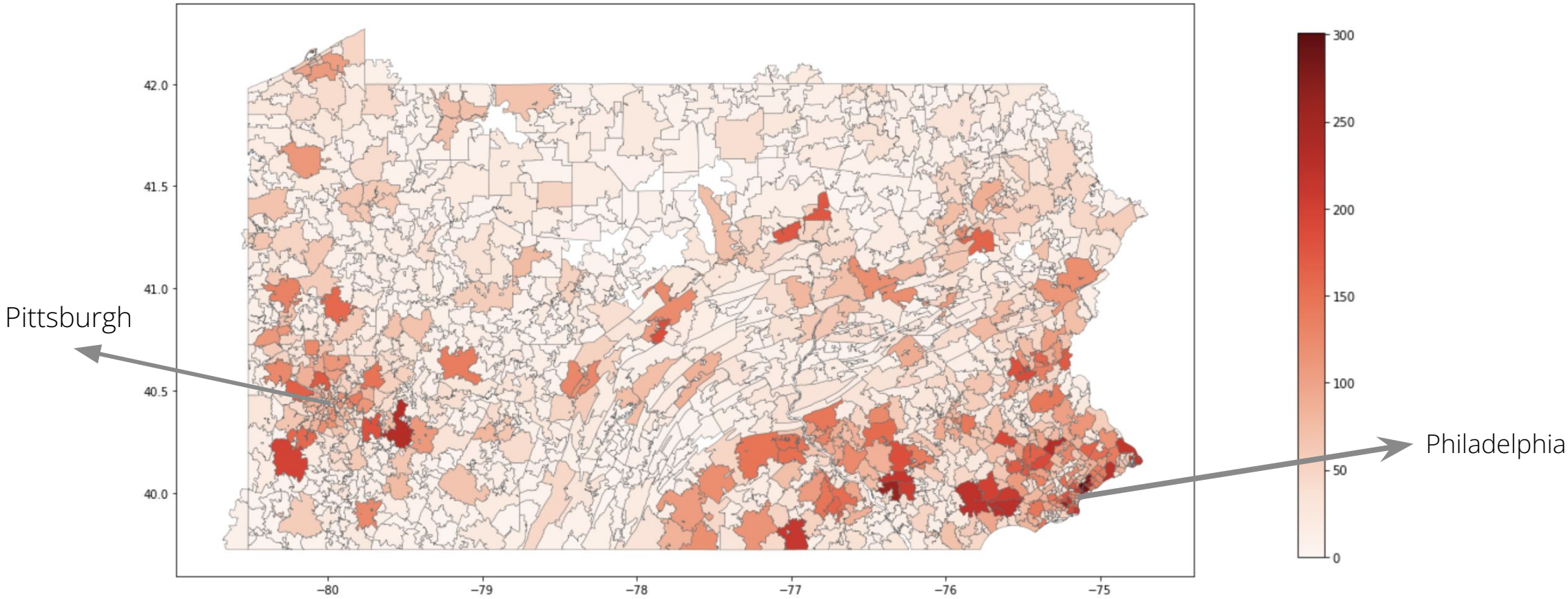
$$\text{DRE for 15237 in Pittsburgh} = \frac{42{,}561}{12{,}791{,}171} \times 156{,}200 \times ⅓ = 173$$

Note: We assume that epilepsy populations are proportional to zip code populations

Visualizing DRE Patients by ZIP Code (US)

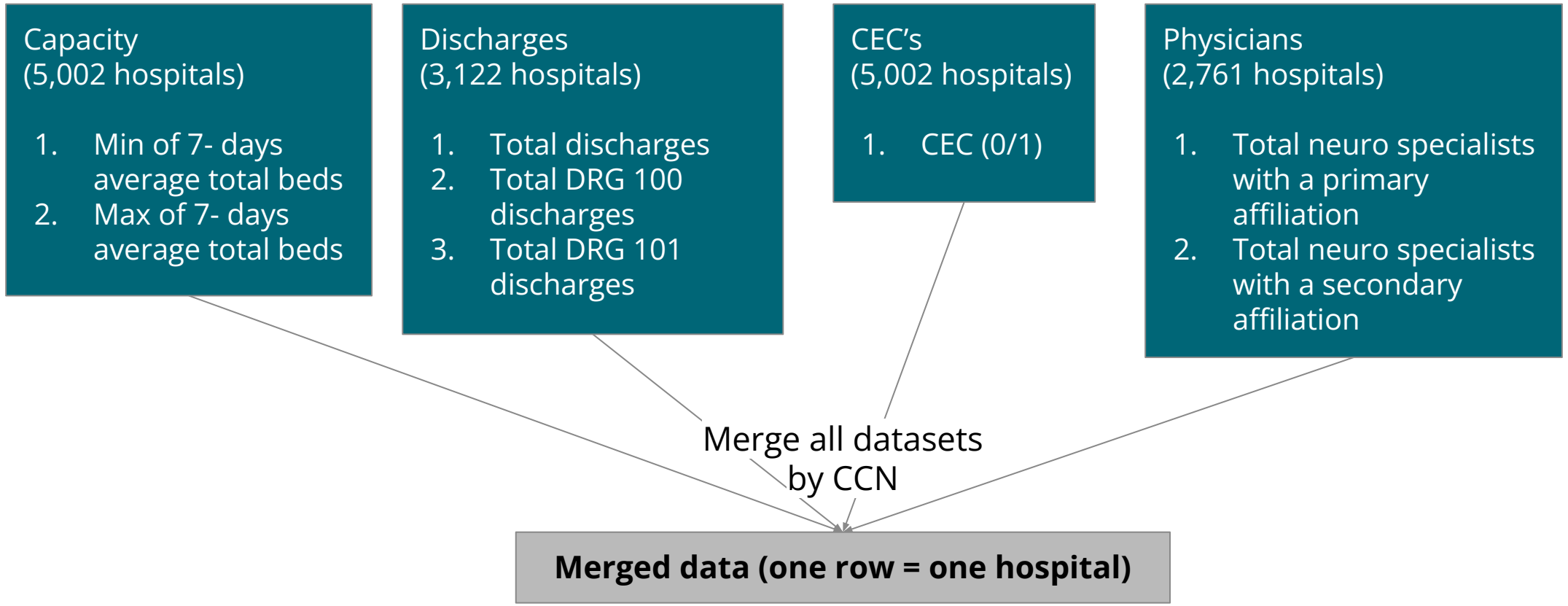


Visualizing DRE Patients by ZIP Code (PA)



DRE Patients by Hospital

Merging Disparate Datasets

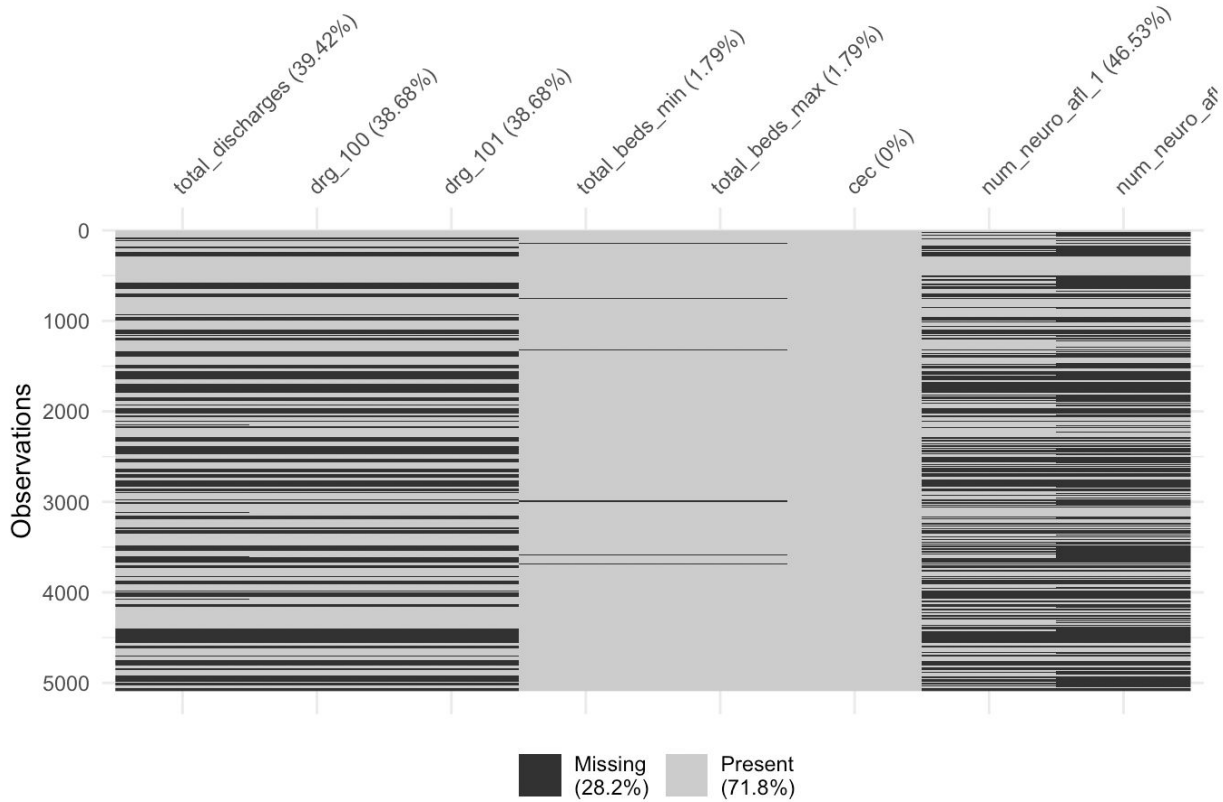


https://simplemaps.com/data/us-zips
https://www.cdc.gov/mmwr/volumes/66/wr/mm6631a1.html

Imputing Missing Values

Used MICE library with PMM (predictive mean matching) method to replace all NA's in numeric columns

See appendix for more details



Hospital Model Summary

- Allocate DRE patients to CEC and Non-CEC's
  - # of primary affiliated neuro specialists created the best split based on comparisons to internal validation

- Distribute DRE patients across hospitals
  - Discharges, neuro specialists, and hospital beds are all positively correlated with DRE

Hospital Model: Step 1

Distribute DRE Patients across CEC's

$$\text{CEC DRE \%} = \frac{\text{\# of CEC Neuro Specialists}}{\text{\# of All Neuro Specialists}} = \frac{10{,}311}{10{,}311 + 20{,}150} = 33.8\%$$

$$\text{Non-CEC DRE \%} = 66.2\%$$

Based on 1,523,075 DRE patients in the U.S.,

CEC DRE = 515,558 and Non-CEC DRE = 1,007,517

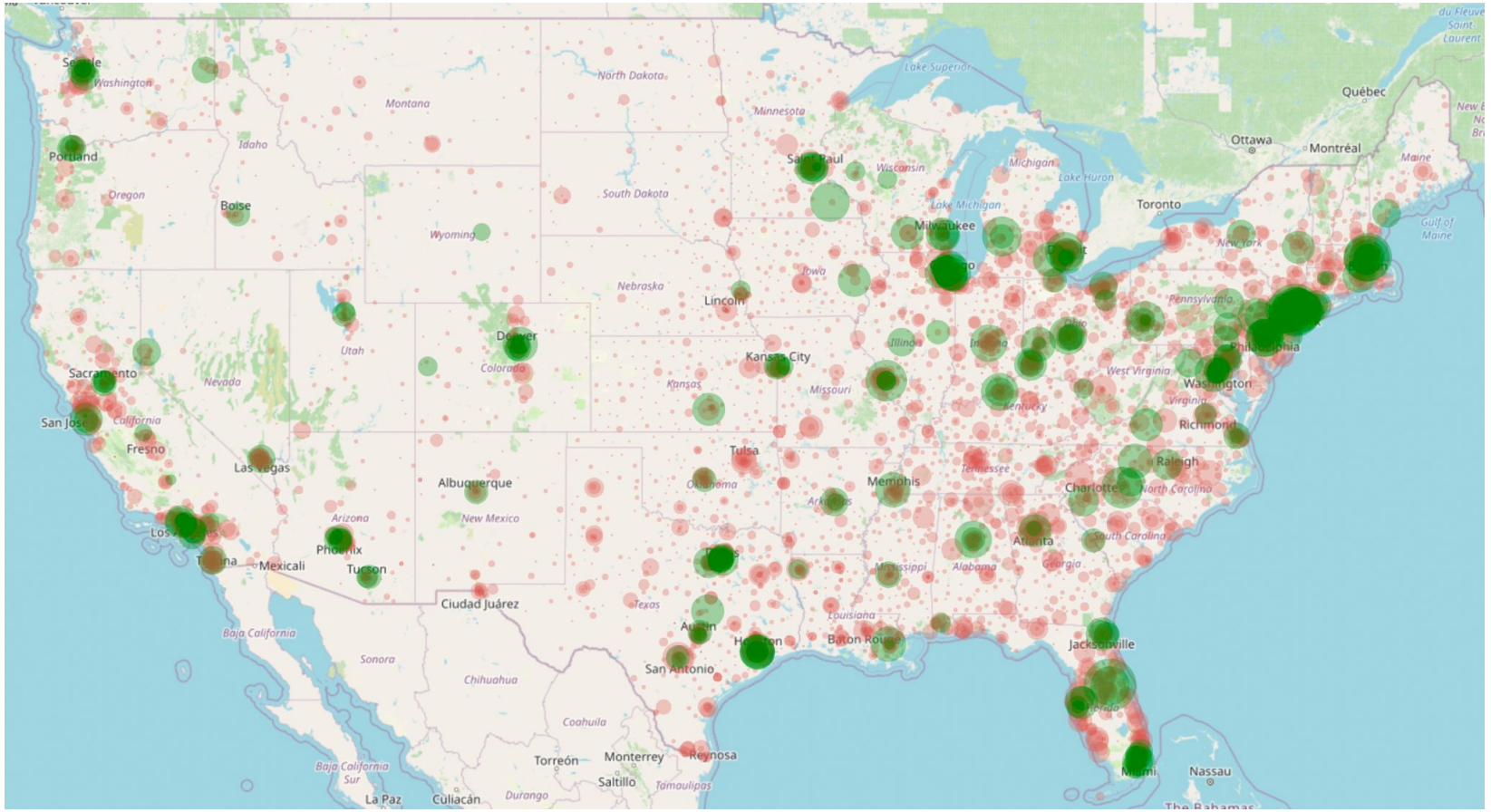
Hospital Model: Step 2

- Within a CEC or Non-CEC, weight (hospital multiplier) each hospital as follows:
  - Normalize all variables* using min-max normalization
  - Average all variables together
  - Scale metric so that the weights sum to one

UAB Hospital	=	CEC DRE	×	Hospital Multiplier	=	Hospital DRE
UAB Hospital	=	515,558	×	0.7%	=	3,718

*total discharges, seizure discharges, neuro specialists, hospital beds

Visualizing DRE Patients by Hospital



Green: CEC

Red: Non-CEC

Comparing Estimates to Internal Data

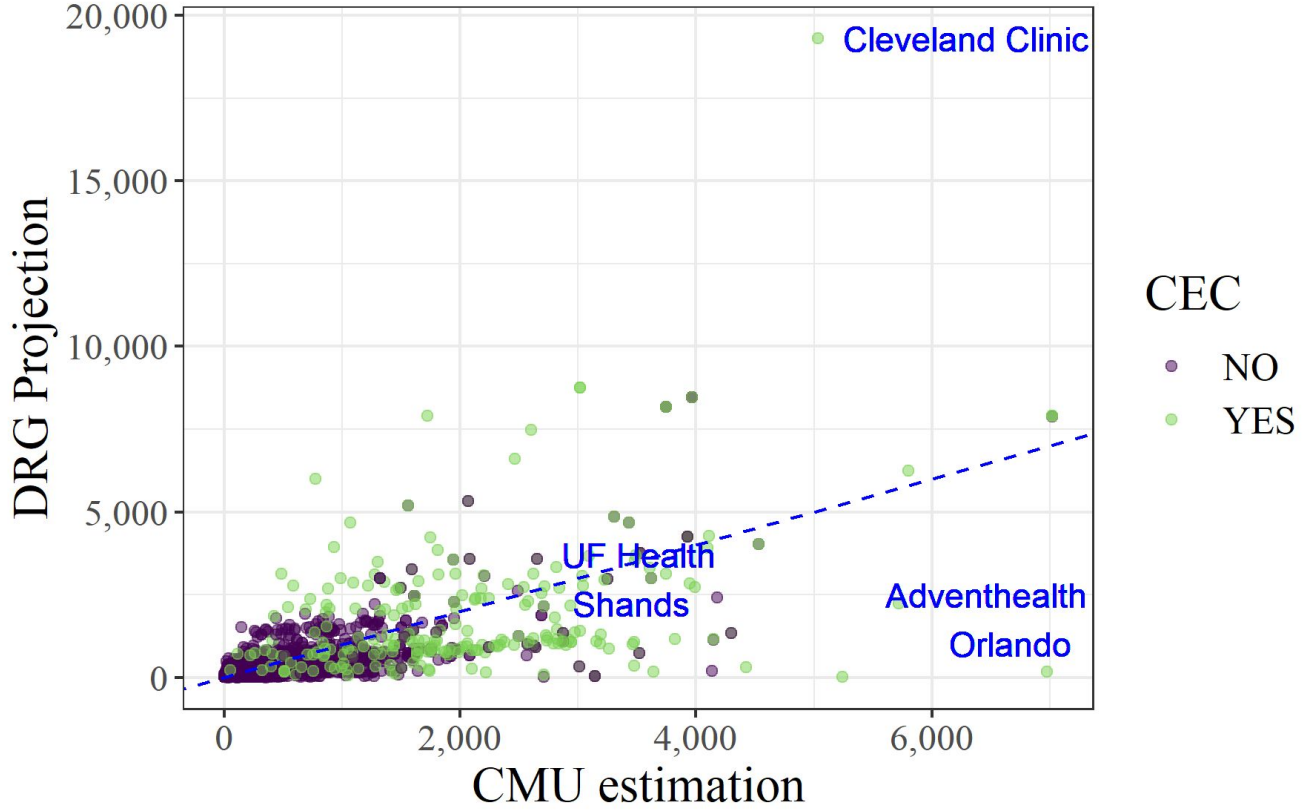
We compare our estimates to internal projections based on claims data

- Internal: 1.5M DRE patients projected across 6,062 hospitals
- CMU: 1.5M DRE patients projected across 5,091 hospitals

There are some hospitals we cannot capture with all public data. For example, Kessler Institute for Rehabilitation - West Orange and West Oaks Hospital are captured by DRG and not CMU, but all such hospitals have less than 600 projected DRE patients.

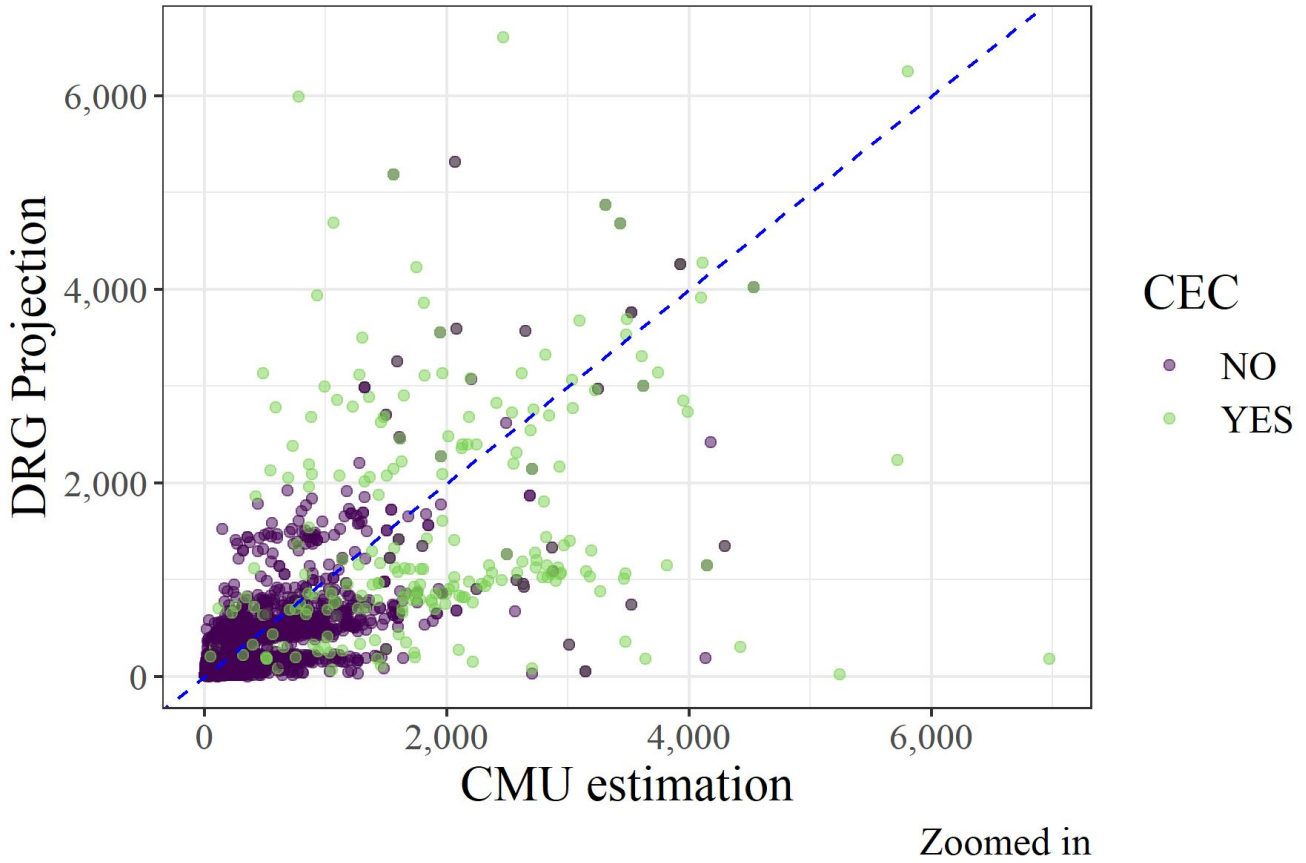
Comparing Estimates to Internal Data

- 4,708 hospitals have estimates from both DRG and CMU
- Outliers
  - CMU "underestimates" DRE patients at the Cleveland Clinic
  - CMU "overestimates" at Adventhealth Orlando



Comparing Estimates to Internal Data

- Zoomed to avoid outliers
- Overall correlation 73%
  - CEC's: 40% correlation
  - Non-CEC's: 75% correlation
- Model better estimates Non-CEC's



Discussion

- Maintenance: When the data is updated publicly, we can pull the new data and routinely update the analysis for comparison to internal data

- Drawbacks:
  - Public resources are not exhaustive and contain missing data
  - Public data sources vary across different years
  - Methods required assumptions that are difficult to validate

Main Takeaways & Recommendations

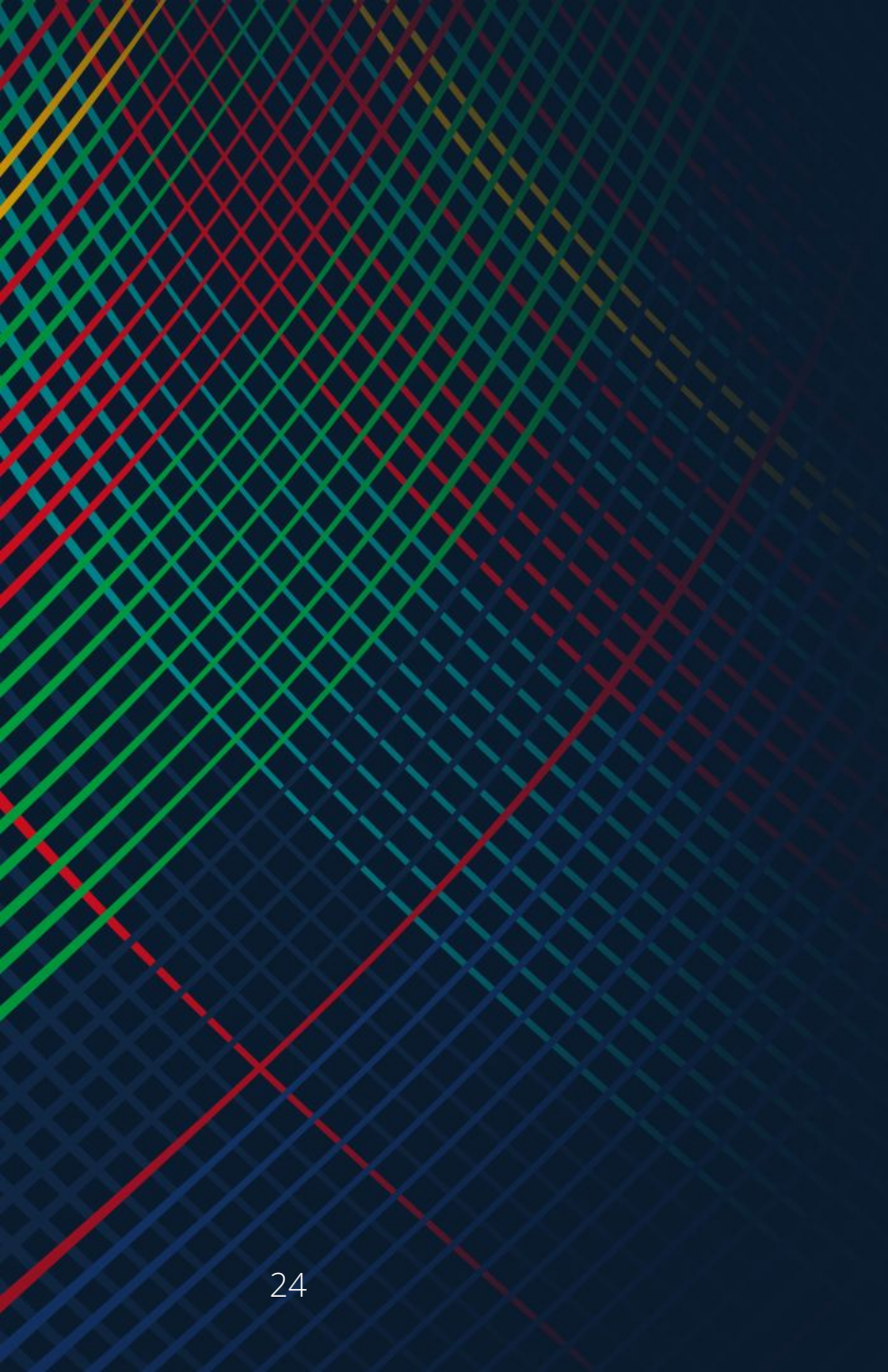
- CMU's ZIP code model provides LivaNova with a new method for understanding where DRE patients physically are located
- This project identified an alternate approach to distribute DRE by hospital
  - Allows for internal validation
  - The Northeast and California have the most DRE patients and CEC hospitals and should be treated accordingly
- Houston ZIP code 77030 contains few DRE patient residents, but contains many hospitals that treat high volumes of DRE patients

Recommendations:

- Complete a deep, hospital-by-hospital validation of the existing approach
- Consider an approach that starts with the patient in their ZIP code and then distribute them to relevant hospitals using distance from their home

Contacts

Zhuoheng Han	zhuohenh@andrew.cmu.edu
Sifeng Li	sifengl@andrew.cmu.edu
Ziyan Xia	zxia2@andrew.cmu.edu
Yuqing Xu	yuqingxu@andrew.cmu.edu



Thank You

Appendix

Background

LivaNova

- Medical technology company focused on head and heart therapies

This Project's Focus

- Epilepsy patients
  - A disorder in which nerve cell activity in the brain is disturbed, causing seizures
- Drug resistant epilepsy (DRE) patients
  - Epilepsy patients who fail two antiepileptic drug therapies

Project Timeline

Phase	Step	Date	Key Milestone
Phase 1: Kickoff	Negotiate with client to align our understanding of the scope and the clients'	2/28/2022	Kickoff Meeting
Phase 2: Research	Find all datasets that could be useful	3/1/2022-3/28/2022	Ready to do the data cleaning
Phase 3: Data Cleaning	Deal with missing values and unreasonable values Estimation by zip codes	3/22/2022-4/1/2022	Ready to do the data exploratory
Phase 4: Data exploratory	Perform EDA and imputation on missing values	4/1/2022-4/11/2022	Find interesting patterns of our data
Phase 5: Model building	Determine influential factors on number of patients, estimate the number of patients by hospitals	4/11/2022-4/22/2022	The accuracy or other evaluation criterion meet our needs
Phase 6: Deliverables	Prepare for the final presentation	4/23/2022-4/26/2022	Final presentation

Merged Data Description

ccn: CMS Certification Number (CCN) of the given facility

total_discharges: The number of discharges billed by all providers for inpatient hospital services

drg_100: The number of discharges billed by DRG 100 for inpatient hospital services

drg_101: The number of discharges billed by DRG 101 for inpatient hospital services

total_beds_min: The minimum number for an average total number of all staffed inpatient and outpatient beds in the hospital reported during the 7-day period (different collection weeks)

total_beds_max: The maximum number for an average total number of all staffed inpatient and outpatient beds in the hospital reported during the 7-day period (different collection weeks)

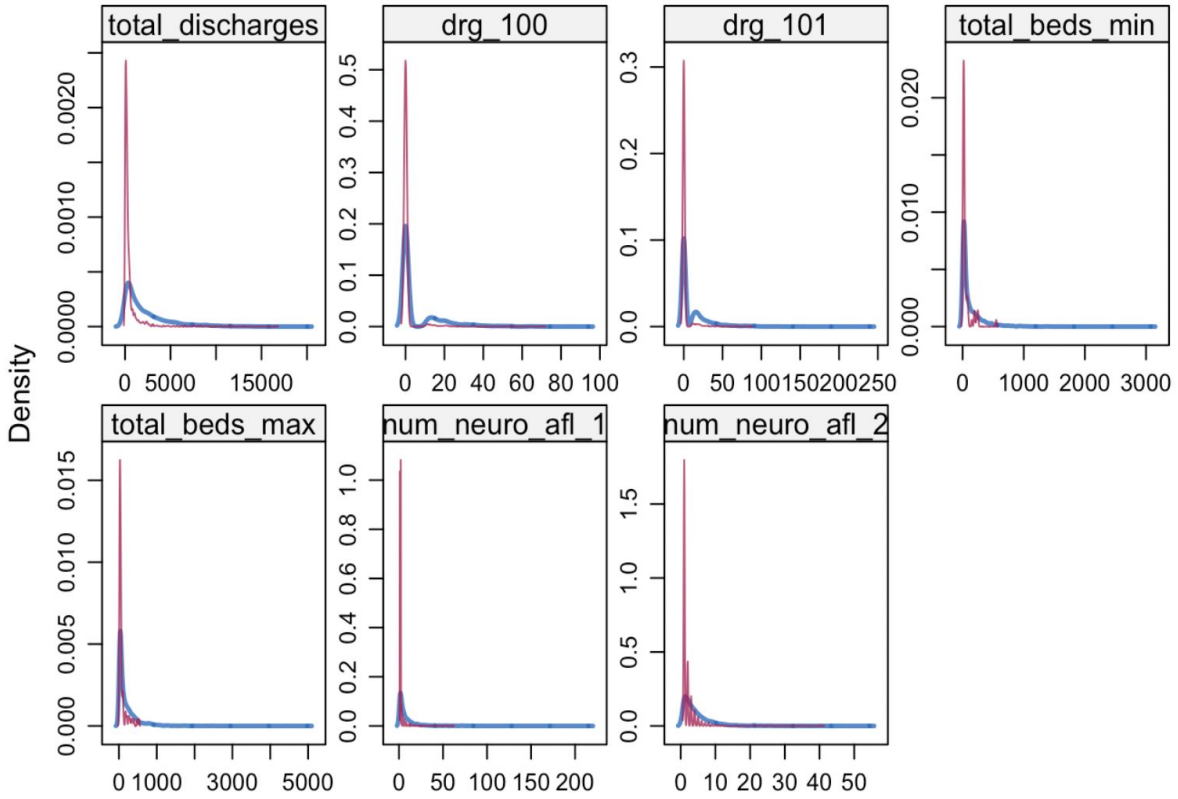
cec: Whether the hospital is accredited as CEC centers

num_neuro_afl_1: The total number of neuro specialists if this hospital is the first affiliation

num_neuro_afl_2: The total number of neuro specialists and neurosurgeons if this hospital is the second affiliation

Impute Missing Values

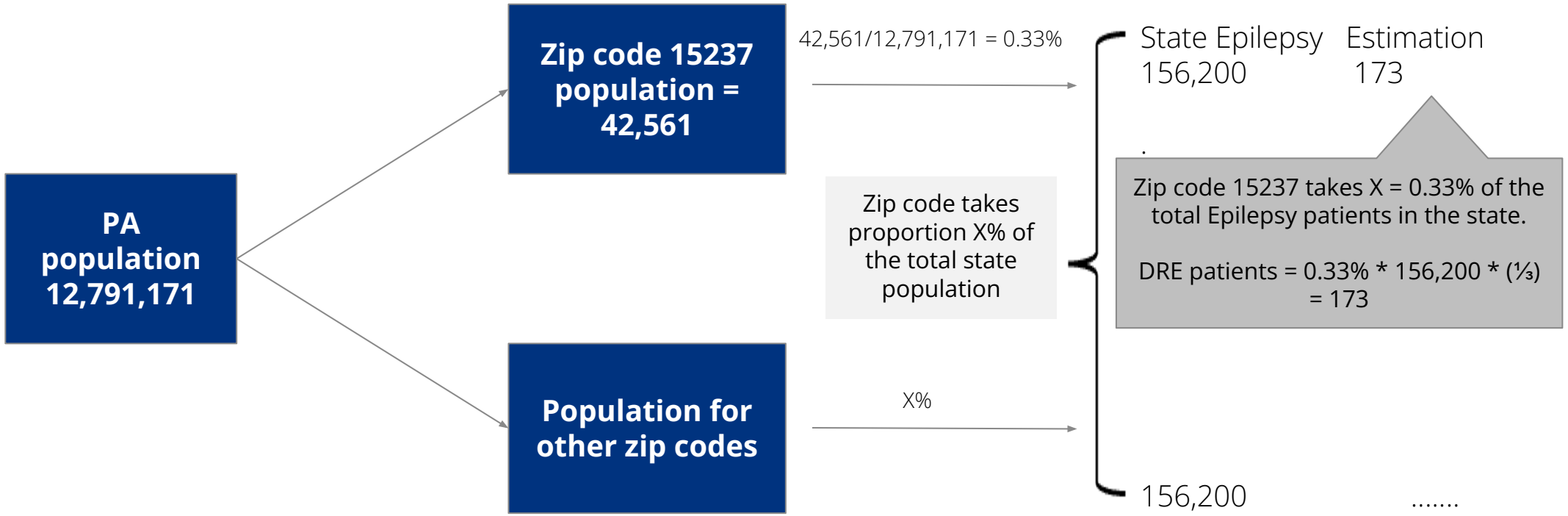
- Used MICE library with PMM (predictive mean matching) method to replace all NAs in numeric columns
  - PMM produces imputed values that are much more like real values. To be more specific, variables in our dataset are skewed, the imputed values calculated by this method would show skewness



Correlation Plot

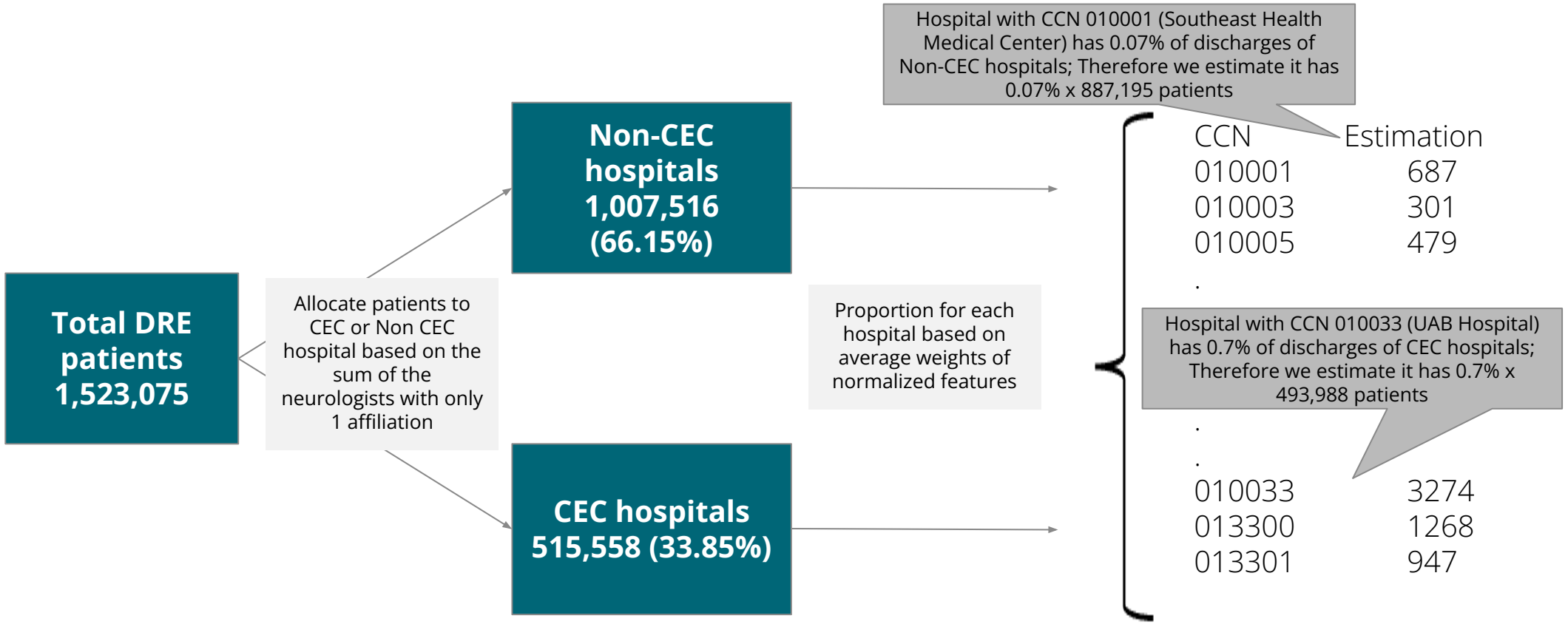
Modeling DRE Patients by Zip Code

PA Example



DRE by Hospital Model

Overview: Building a Model



Assumptions

- We assume one third of epilepsy patients have DRE
- We assume DRE is allocated to CEC's and Non-CEC's differently
- We assume the number of neuro specialists affiliated with a hospital gives information about how DRE is split by CEC's and Non-CEC's
- We assume discharges, number of neurologists, and hospital beds are positively correlated with DRE at the hospital level
- We assume discharges, number of neurologists, and hospital beds are all equally influential when allocating DRE

 https://www.ilae.org/patient-care/drug-resistant-epilepsy

Checking assumptions

Discharges, hospital beds, and the # of neuro specialists are positively correlated with projected DRE (DRG projections)

- Variables have a higher correlation for Non-CEC's

CEC	Total Discharges	DRG 100	DRG 101	Total Beds Min	Total Beds Max	Primary Affiliation Neurologists	Secondary Affiliation Neurologists
No	.59	.58	.64	.67	.62	.68	.31
Yes	.34	.30	.36	.37	.39	.39	.14

DRE By Hospital - Model Building

To estimate DRE at the hospital level, we must make 2 decisions:

- How many DRE patients should be allocated to CEC's and Non-CEC's?
- How should we allocate DRE patients to individual hospitals?

How many DRE patients should be allocated to CEC's and Non-CEC's?

We use the number of neurologists who listed each hospital as their first affiliation (N) to determine the split of DRE between CEC's and Non-CEC's as follows:

$$\text{CEC \% allocation} = \frac{\sum_{cec=1} N_{CEC=1}}{\sum_{cec=1} N_{CEC=1} + \sum_{cec=0} N_{CEC=0}} * 100\% = \frac{10,311}{10,311 + 20,150} * 100\% = 33.8\%$$

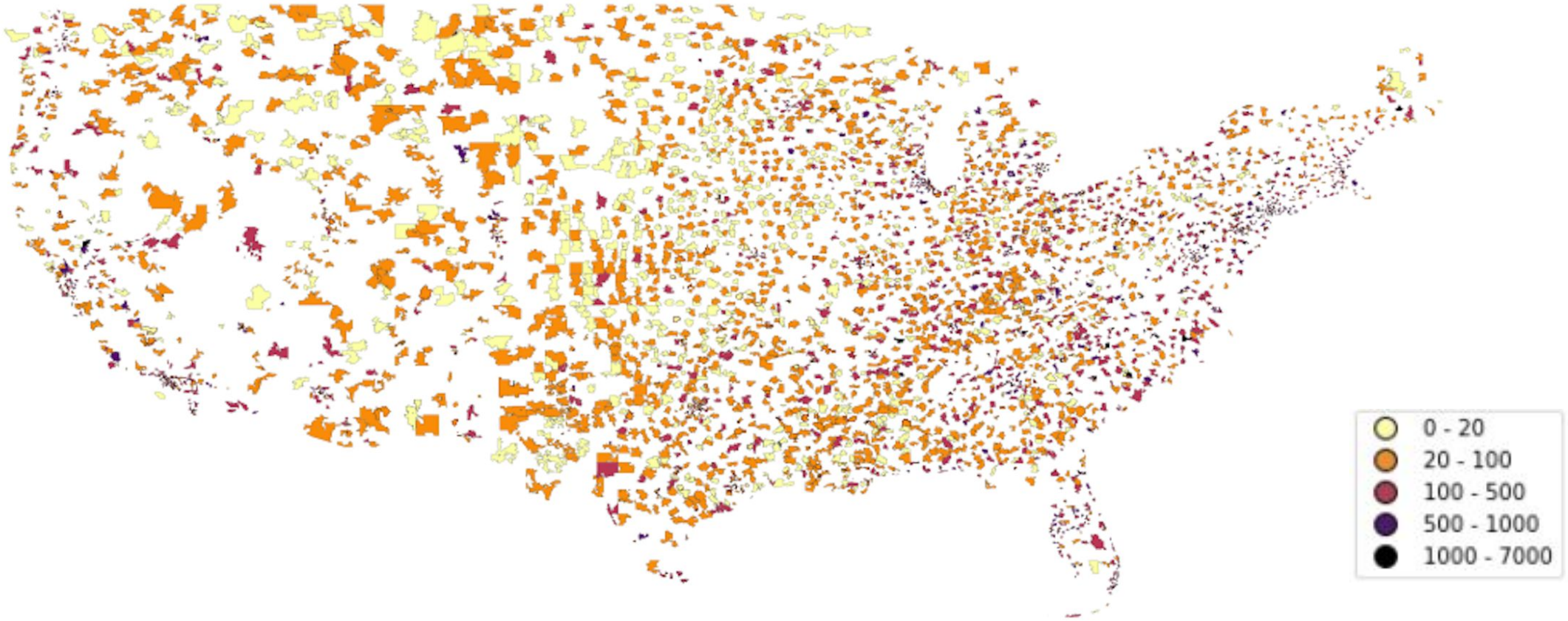
$$\text{Non-CEC \% allocation} = \frac{\sum_{cec=1} N_{CEC=1}}{\sum_{cec=1} N_{CEC=1} + \sum_{cec=0} N_{CEC=0}} * 100\% = \frac{20,150}{10,311 + 20,150} * 100\% = 66.2\%$$

If we assume that there are 1,523,075 DRE patients in the U.S, 515,558 DRE patients will be allocated to CEC hospitals and 1,007,516 DRE patients will be allocated to non-CEC hospitals.

How should we allocate DRE to individual hospitals?

- For both CEC's and Non-CEC's, we create a weighting system using total discharges, seizure discharges, affiliated neurologists, and min/max hospital bed information at the hospital level
- We get hospital allocation weights as follows:
  - Normalize all variables (Min-Max Normalization)
  - Average all variables together
  - Scale metric so that the weights sum to one

Visualizing DRE Patients by Hospital



Main Takeaways & Recommendations

The number of DRE patients by zip code is highly correlated with the population in zip code.

Recommendations:

- Invest more in areas having large population
  - potential action: increase ad campaigns
- Increase the public awareness in areas having small population
  - potential action: educate, give public lectures on specific professional knowledge

Main Takeaways & Recommendations

- **Northeast** has the most significant number of DRE patients and CEC hospitals.
  - Recommendation 1: Allocate more sales people to areas covering **Boston, New York City, Philadelphia, Washington DC**.
  - Recommendation 2: Sell more VNS devices to CEC hospitals in Northeast, and the top four are :
    - **Weill Cornell Medical Center (7020) Average: 2122**
    - **NYU Langone Tisch Hospital (5802)**
    - **Massachusetts General Hospital (5721)**
    - **Long Island Jewish Medical Center (5243)**

- For **West Coast**, the most significant number of DRE patients and number of CEC hospitals appear in California.
  - Recommendation 3: Research on major regions San Francisco Bay Area and Greater Los Angeles Area, sell more VNS devices to the following hospitals:
    - **UCSF Medical Center (3523)**
    - **Cedars-Sinai Medical Center (2939)**

Comparison on Actual Patients and Expected Patients

- Actual state epilepsy population vs. Expected state epilepsy population
- expected state epilepsy population calculated by: (state population/total population) * total Epilepsy population
- Visualize difference: Actual state epilepsy population-Expected state epilepsy population
- The difference ranges from 11589.6 (CA) to -6502.91 (TX)
- Darker colors means lpositively larger difference