Mental Health: Analyzing New York State Mental Health Patient Data

Group: Jill Book, Michaela Nolan, & Pengcheng Wu

Topic: Mental Health

Why is this important to investigate?

- According to the National Alliance of Mental Illness (NAMI), Nearly one in five U.S. adults live with a mental illness (46.6 million in 2017).
- Emotional and mental health is important because it's a vital part of people's lives and impacts their thoughts, behaviors and emotions.
- Providing preventative care and targeting high risk patients could help keep mental health patients out of the hospital and ultimately reduce health care costs.

What does the literature say?

- 2009- NYS reported $7 billion dollars in medicaid expenditures on patients with behavioral health and mental health diagnoses (Smith, Erlich et al. 2013).
- 2003- Mental Health & Substance Abuse Disorders ranked 4th most expensive condition category in Personal Health Care Spending (Dieleman et al., 2016)
- 2010 Veteran Health Administration (VHA) data showed a connection between mental health disorders and frequent ED usage (91.7% of patients who visited ED >25 times had a Mental health and/or substance abuse disorder) (Doran, Raven, & Rosenheck, 2013).

Stakeholders and Knowledge Gaps

Stakeholders:

- New York State Hospital Leadership
- New York State Office of Mental Health and its affiliated mental service providers
- New York State Insurance Companies
- All New Yorkers, especially those with problems in mental health

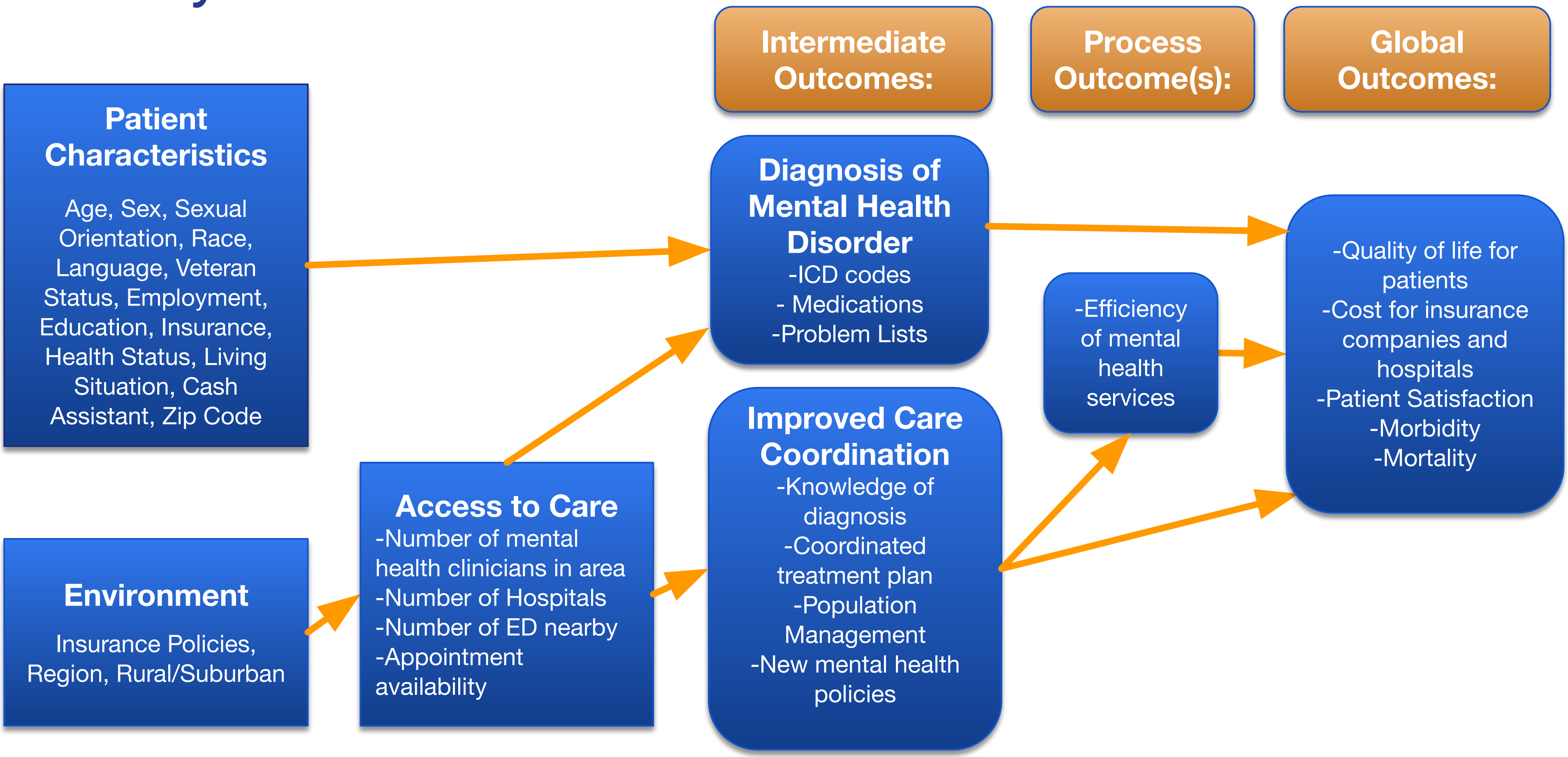
Gaps in Knowledge:

- Children/Adolescent mental health care- survey from the National Survey of American Families reported that only 21% of children who need mental health care/services actually receive treatment. (Bringewatt & Gershoff, 2010)

- Poor Mental Health Literacy- People do not have well established mental health literacy and are unable to understand signs of mental health distress. (Jorm 2000)

Analytic Framework for Mental Health Disorders

Analytic Framework



Research Question & Dataset Exploration

Research Question & Dataset

Research Question: Can analyzing demographic and diagnostic data related to mental health patients help predict/flag patients who are likely to be admitted for emergency mental health services?

Dataset: NYS Patient Characteristics Survey (PCS): 2015 from New York State Open Data and https://www.kaggle.com/new-york-state/nys-patient-characteristics-survey-pcs-2015

Data Description:
The Patient Characteristics Survey (PCS) dataset contains demographic (age, race, sex), clinical (disease history), social (work experience, education experience), and insurance characteristics for each client served by the NYS public mental health system during the week of October 19, 2015. All mental health programs licensed or funded (directly or indirectly) by the NYS Office of Mental Health were required to report.

About the Data

Types of Data:

- Almost all nominal data, such as employment status, race, ethnicity, gender identity, and presence of chronic illnesses, etc.
- 2 columns of ordinal data: number of hours worked per week (given in ranges of hours) and education status

Analysis Method:

- Due to binary predictor variable and categorical data in dataset: Logistic regression model for prediction, patient classification methodology development, and correlation analysis.

About the Data (con.)

List of Variables:

Survey Year, Program Category, Region Served, Age Group, Sex, Transgender, Sexual Orientation, Hispanic Ethnicity, Race, Living Situation, Household Composition, Preferred Language, Veteran Status, Employment Status, Number of Hours Worked Each Week, Education Status, Special Education Services, Mental Illness, Intellectual Disability, Autism Spectrum, Other Developmental Disabilities, Alcohol Related Disorder, Drug Substance Disorder, Mobility Impairment Disorder, Hearing Visual Impairment, Hyperlipidemia, High Blood Pressure, Diabetes, Obesity, Heart Attack, Stroke, Other Cardiac, Pulmonary/Asthma, Alzheimer or Dementia, Kidney Disease, Liver Disease, Endocrine Condition, Neurologic Condition, Traumatic Brain Injury, Joint DIsease, Cancer, Other Chronic Med. Condition, No Chronic Med. Condition, Unknown Chronic Med. Condition, Smokes, Received Smoking Medication, Received Smoking Counseling, Serious Mental Illness, Principal Diagnosis Class, Additional Diagnosis Class, SSI Cash Assistance, SSDI Cash Assistance, Veterans Disability Benefits, Veterans Cash Assistance, Public Assistance Cash Program, Other Cash Benefits, Medicaid and Medicare Insurance, No Insurance, Unknown Insurance Coverage, Medicaid Insurance, Medicaid Managed Insurance, Medicare Insurance, Private Insurance, Child Health Plus Insurance, Other Insurance, Criminal Justice Status, and Three Digit Residence Zip Code.

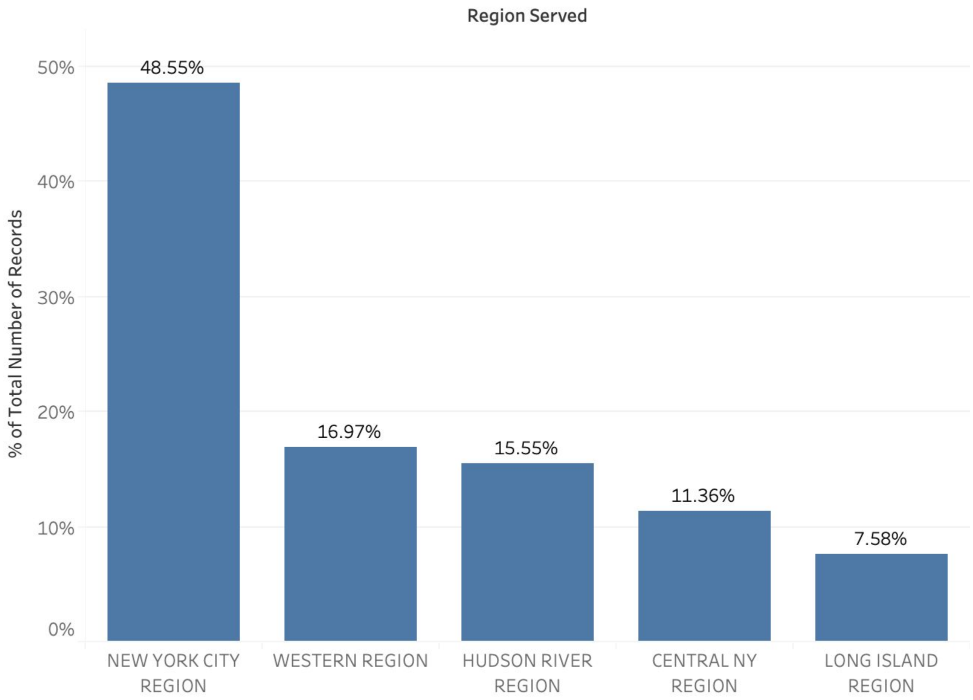
Potential Data Bias and Limitations

- **Observational Bias-** Provider reported survey data (versus patient reported survey data) and since the reported information is second hand there is a potential for bias responses
- **Sampling Bias** - the survey was conducted only over one week (October 19, 2019). The results/sample size may have been different if another time was chosen
- **Narrow Scope-** This data set only includes patients at NY state mental hospitals, so results may not be applicable to the general population
- **Unknown Responses-** There is no missing/empty data, but many responses are "unknown," which could be considered missing
- **Accuracy of Survey Responses-**
  - Some respondents may hide some private disease situations for some personal reasons
  - Some respondents may not know they have the mental disease since it's difficult to be diagnosed
  - Respondents may not be able to accurately state their health status and illness when they fill out the questionnaire because they do not have professional medical knowledge.
- **External Factors-** When data collectors organize and clean health data, it is easy to ignore the impact of uncontrollable external factors on data accuracy because there is no reliable demographic and sociological knowledge.
- **Need for medical/clinical knowledge-** Data analysts cannot find the key to the problem quickly and accurately when they do not have rich clinical and scientific experience. They need professional medical researchers to guide the research direction.

Sample Population

- 80% adults, 20% children
- 50.2% female, 49.6% male, 0.7% transgender
- 5.8% LGBTQ+, 81% heterosexual
- 23% Hispanic
- 26.1% Black, 51.5% White, 2.6% multi-racial, 16.5% other races
- 89.3% English-speaking, 7.1% Spanish-speaking, 2.9% other languages
- 2.5% veterans
- 14.9% employed, 71.2% unemployed and not looking for work, 9.8% unemployed and looking for work
- 92.9% have a diagnosed mental illness, 5.7% do not have a diagnosed mental illness
- 13.5% college or graduate degree, 13.4% some college, 54.4% middle school to high school education, 8.7% pre-kindergarten to fifth grade education, 0.4% no formal education
- 53.2% have one or more chronic medical conditions, 37.2% do not have a chronic medical condition
- 27.4% smokers, 65.6% non-smokers
- 91.2 % insured, 6% uninsured
- 11.4% Central New York region, 15.5% Hudson River region, 7.6% Long Island region, 48.5% New York City region, 17% Western region
- 2.9% experiencing homelessness

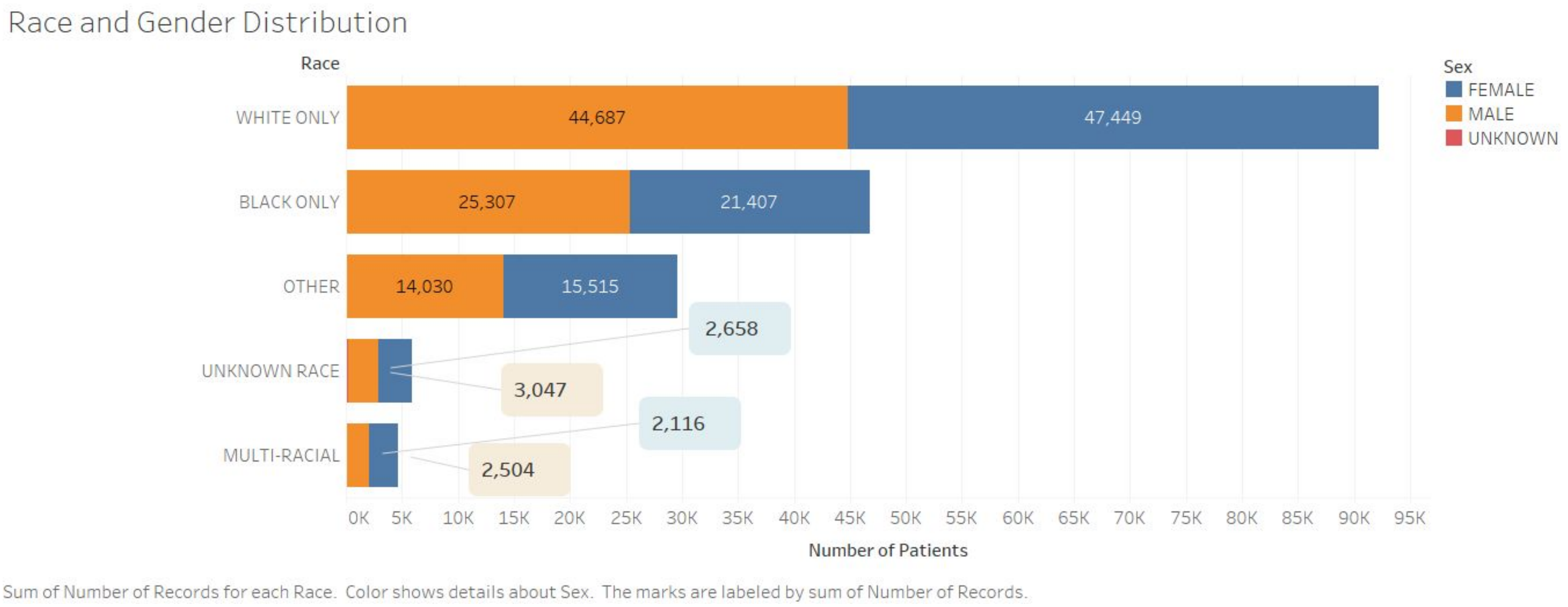
Dataset: Region of NYS, Population Age, & Education Status



Age Group	
ADULT	79.96%
CHILD	20.03%
UNKNOWN	0.01%

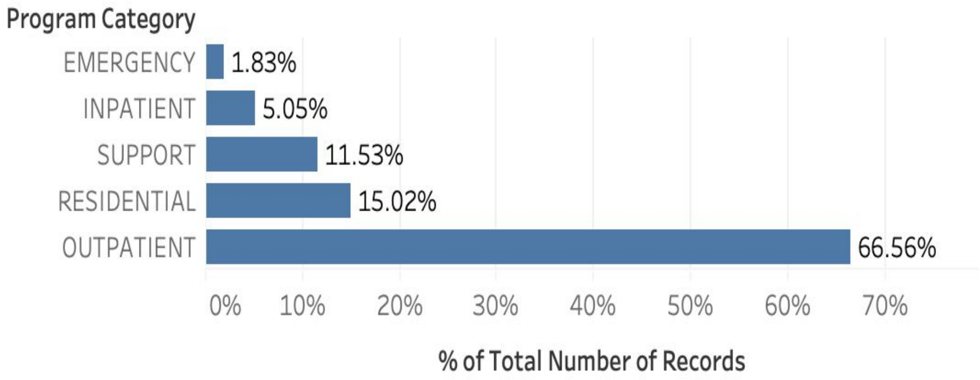
Education Status	
NO FORMAL EDUCATION	0.42%
PRE-K TO FIFTH GRADE	8.70%
MIDDLE SCHOOL TO HIGH SCHOOL	54.36%
SOME COLLEGE	13.45%
COLLEGE OR GRADUATE DEGREE	13.45%
OTHER	1.67%
UNKNOWN	7.96%

Dataset: Gender and Race Distribution

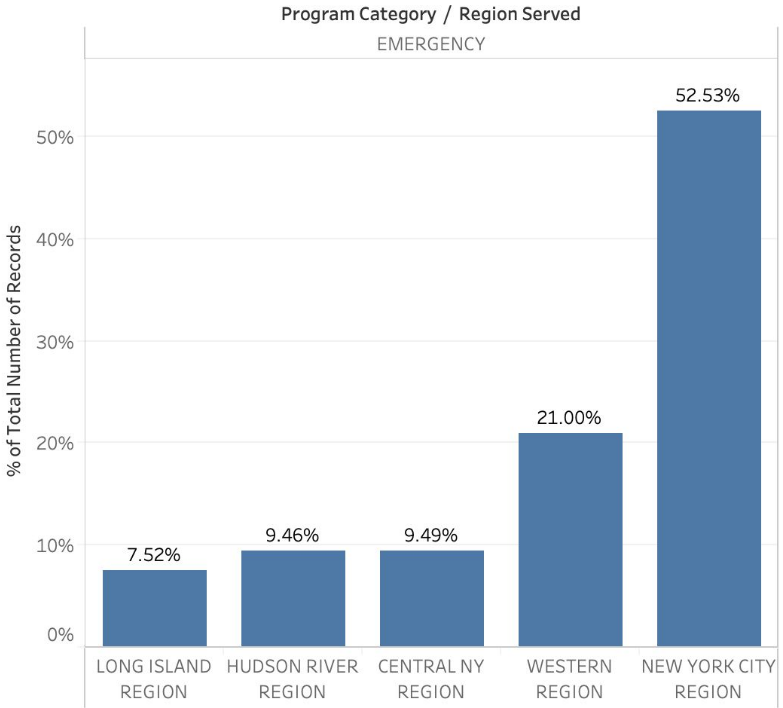


Analytics Focus: Emergency Room Visit

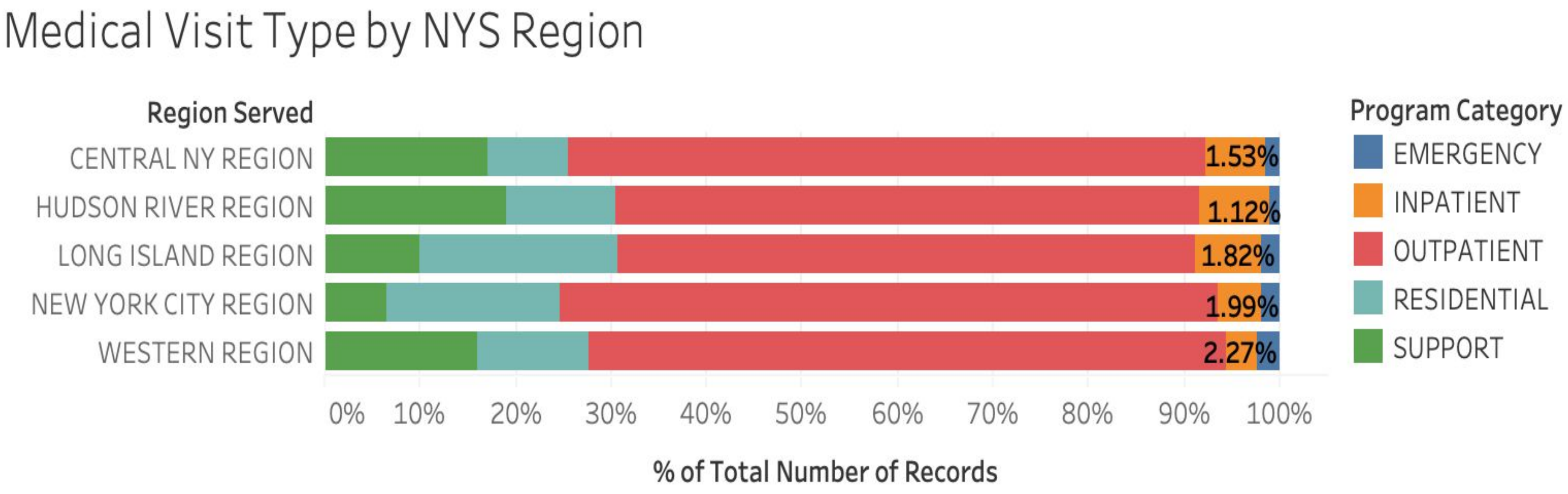
In this data set, **1.8% of patients were seen for emergency visits**, and all other visits were classified as inpatient, outpatient, residential, or support.



More than **50%** of emergency visits were from the New York City Region



Emergency Room Rate Similar Across Regions

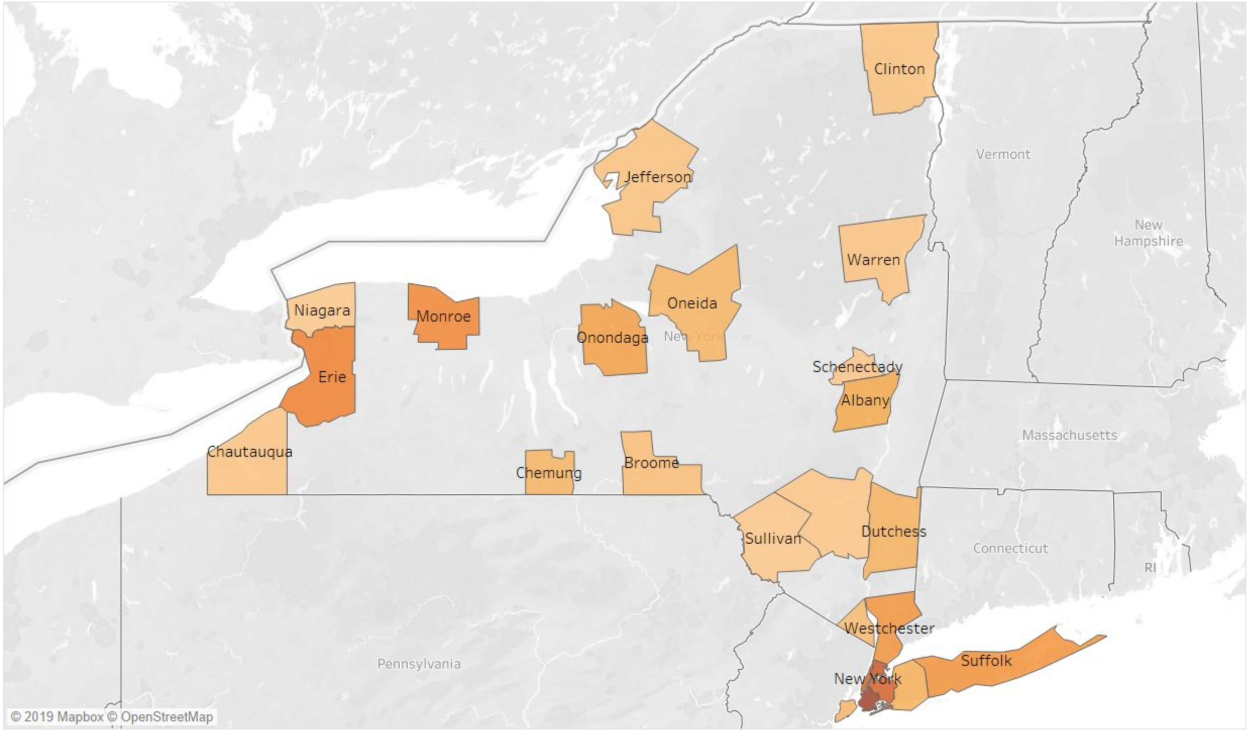


Percent of emergency visit is consistent across the different regions of New York State. Interestingly, Western New York had the highest percent of emergency visits (2.27%)

Visualization of Patients Proportion in the Map

The counties that have most proportion of emergency patients are: Erie (6.47%), Monroe (5.90%), Suffolk (5.07%), Westchester (4.81%), Queens (9.54%), Bronx (10.67%), Kings (14.02%) and New York (9.17%).

Proportion of Patients Recorded by County

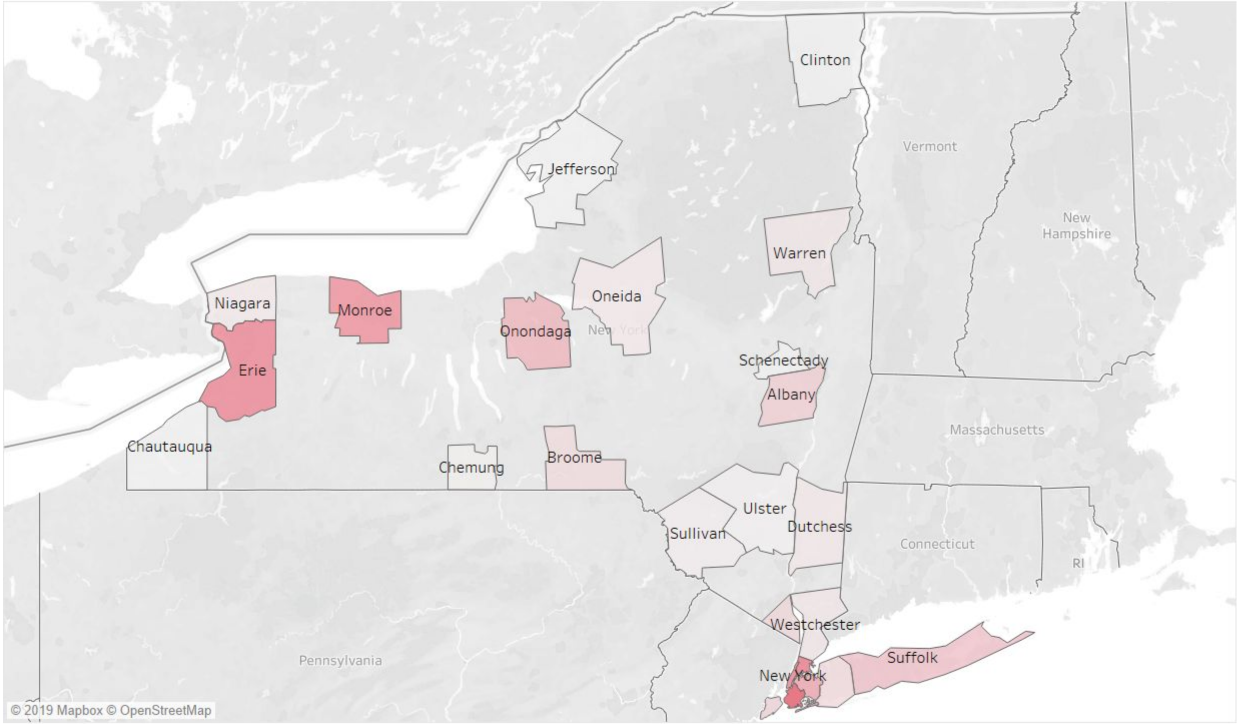


Map based on Longitude (generated) and Latitude (generated). Color shows % of Total Count of Program Category. The marks are labeled by County Name. Details are shown for County Name. Percents are based on each row of each pane of the table.

Visualization of Emergency Patients in the Map

Although the boroughs in NYC contribute close proportion of patients to the data set, emergency visits from the Queens is significantly less than the other boroughs. Westchester has lower rate of emergency visit, comparing with the total number of patients from there. Besides,Erie and Monroe in western part of the state shows a serious tendency to seek emergency care, that worths watching.

Number of Emergency Patients by County



Map based on Longitude (generated) and Latitude (generated). Color shows principal diagnal. The marks are labeled by County Name. Details are shown for County Name. The data is filtered on Program Category, which keeps EMERGENCY.

Data Analysis

Statistic Method——Chi-Squared Test

- The chi-squared test is used to determine whether there is a significant difference between the expected frequencies and the observed frequencies in one or more categories.
- Mainly considered patients' symptom characteristics and important social characteristics in different program categories.

For symptom characteristics

- **Related:** Mental Illness, Alcohol Related Disorder, Drug Substance Disorder, Mobility Impairment Disorder, Hearing Visual Impairment, Hyperlipidemia, High Blood Pressure, Diabetes, Obesity, Heart Attack, Stroke, Other Cardiac, Pulmonary Asthma, Kidney Disease, Endocrine Condition, Neurological Condition, Joint Disease, Cancer, No Chronic Med Condition, Unknown Chronic Med Condition, Smokes, Received Smoking Medication, Received Smoking Counseling, Serious Mental Illness
- **Unrelated:** Alzheimer or Dementia, Autism Spectrum, Liver Disease, Traumatic Brain Injury, Intellectual Disability, Other Chronic Med Condition

Test Results in Stata & Python

```
. tabulate agegroup programcategory, chi2

           |   Program Category
 Age Group | EMERGENCY  NONEMER.. |     Total
-----------+----------------------+----------
     ADULT |     2,662    140,551 |   143,213
     CHILD |       624     35,241 |    35,865
   UNKNOWN |         0         18 |        18
-----------+----------------------+----------
     Total |     3,286    175,810 |   179,096

          Pearson chi2(2) =   2.5882   Pr = 0.274
```

	P_value	Adjusted_p_value	Rejection
Transgender	4.387070e-01	4.768554e-01	False
Hispanic Ethnicity	9.103429e-02	1.167106e-01	False
Veteran Status	3.927511e-01	4.363901e-01	False
Mental Illness	2.035683e-53	1.696403e-52	True
Intellectual Disability	7.810170e-01	7.969561e-01	False
Autism Spectrum	2.875198e-01	3.343253e-01	False
Other Developmental Disability	1.197411e-01	1.460257e-01	False
Alcohol Related Disorder	3.492363e-94	5.820604e-93	True
Drug Substance Disorder	1.476089e-120	3.690224e-119	True
Mobility Impairment Disorder	3.972035e-06	7.092920e-06	True
Hearing Visual Impairment	1.828773e-05	3.047955e-05	True
Hyperlipidemia	7.667315e-37	3.833658e-36	True
High Blood Pressure	2.098500e-19	5.829168e-19	True
Diabetes	5.997728e-14	1.499432e-13	True
Obesity	1.941719e-34	7.468148e-34	True
Heart Attack	3.069432e-03	4.650655e-03	True
Stroke	3.854834e-03	5.668874e-03	True
Other Cardiac	2.778003e-12	6.614293e-12	True
Pulmonary Asthma	3.627416e-10	7.885688e-10	True
Alzheimer or Dementia	1.066767e-01	1.333459e-01	False
Kidney Disease	2.719320e-02	3.578052e-02	True
Liver Disease	2.841385e-01	3.343253e-01	False
Endocrine Condition	1.479935e-09	3.083199e-09	True

Top 10 vs. Last 10 Binary Variables in Python

	P_value	Adjusted_p_value	Rejection
Unknown Insurance Coverage	1.576505e-194	8.197826e-193	True
Drug Substance Disorder	1.476089e-120	3.837833e-119	True
Alcohol Related Disorder	3.492363e-94	6.053429e-93	True
Received Smoking Medication	9.495076e-86	1.234360e-84	True
Unknown Chronic Med Condition	4.847614e-79	5.041518e-78	True
Mental Illness	2.035683e-53	1.764259e-52	True
No Insurance	1.184538e-45	8.799426e-45	True
Medicaid Insurance	3.019674e-45	1.962788e-44	True
No Chronic Med Condition	1.685252e-37	9.737010e-37	True
Hyperlipidemia	7.667315e-37	3.987004e-36	True

Other Developmental Disability	1.197411e-01	1.448032e-01	False
Autism Spectrum	2.875198e-01	3.322451e-01	False
Liver Disease	2.841385e-01	3.322451e-01	False
Veterans Cash Assistance	3.288255e-01	3.717158e-01	False
Veteran Status	3.927511e-01	4.345332e-01	False
Transgender	4.387070e-01	4.752659e-01	False
Traumatic Brain Injury	4.727975e-01	5.017443e-01	False
Child Health Plus Insurance	5.540894e-01	5.762530e-01	False
Intellectual Disability	7.810170e-01	7.963310e-01	False
Other Chronic Med Condition	8.204827e-01	8.204827e-01	False

Multicollinearity Analysis Results

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	Collinearity Statistics Tolerance	VIF
1	(Constant)	.077	.004		18.374	.000		
	RegionServed	.003	.000	.026	11.033	.000	.941	1.063
	IfAdult	.004	.001	.025	8.423	.000	.593	1.687
	IfMale	.000	.000	.003	1.141	.254	.877	1.141
	Transgender	.028	.001	.052	21.056	.000	.883	1.132
	LGBTQ	-.001	.001	-.002	-.986	.324	.881	1.135
	HispanicEthnicity	-.001	.000	-.004	-1.734	.083	.817	1.224
	Race	.000	.000	-.005	-1.842	.065	.900	1.112
	LivingSituation	.003	.001	.014	4.943	.000	.692	1.445
	HouseholdComposition	.002	.000	.015	5.171	.000	.660	1.515
	PreferredLanguage	.000	.001	.001	.583	.560	.824	1.214
	VeteranStatus	.003	.001	.007	2.876	.004	.882	1.134
	EmploymentStatus	.000	.000	.002	.705	.481	.513	1.949
	NumberOfHoursWorkedEachWeek	-.001	.001	-.008	-2.319	.020	.466	2.146
	EducationStatus	-.006	.000	-.067	-25.953	.000	.810	1.234
	SpecialEducationServices	.002	.001	.005	2.209	.027	.946	1.057
	MentalIllness	-.010	.001	-.036	-14.381	.000	.856	1.168
	IntellectualDisability	.000	.001	-.001	-.553	.581	.788	1.268
	AutismSpectrum	-.004	.001	-.012	-4.274	.000	.670	1.493
	OtherDevelopmentalDisability	-.001	.001	-.002	-.778	.436	.698	1.433
	AlcoholRelatedDisorder	.006	.001	.028	9.918	.000	.680	1.471
	DrugSubstanceDisorder	.007	.001	.035	11.974	.000	.643	1.556
	MobilityImpairmentDisorder	-.001	.001	-.004	-1.430	.153	.766	1.306

Multicollinearity Analysis Results

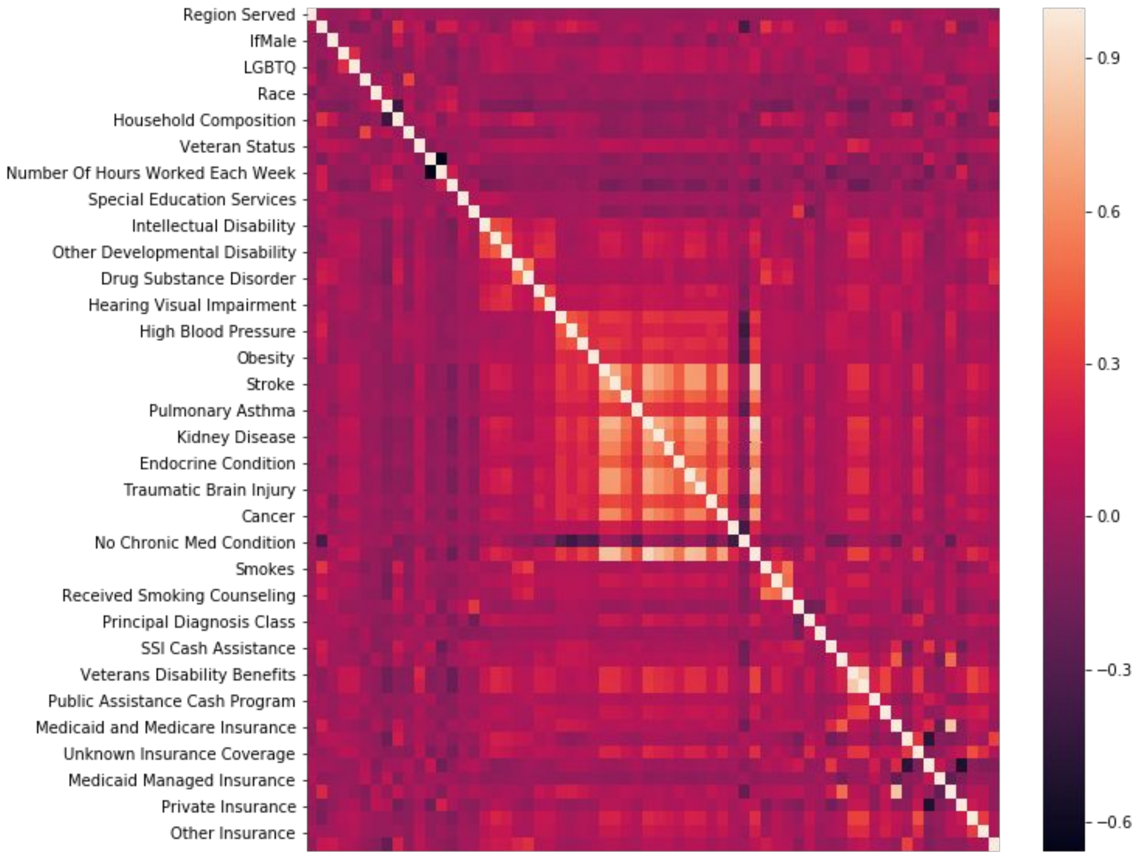
HearingVisualImpairment	-.002	.001	-.006	-2.375	.018	.810	1.234
Hyperlipidemia	-.003	.001	-.013	-4.659	.000	.706	1.416
HighBloodPressure	-.001	.000	-.005	-1.823	.068	.661	1.512
Diabetes	-.001	.001	-.003	-.925	.355	.718	1.392
Obesity	-.002	.000	-.012	-4.633	.000	.746	1.341
HeartAttack	-.001	.002	-.003	-.765	.444	.316	3.169
Stroke	-.001	.002	-.003	-.845	.398	.323	3.095
OtherCardiac	-.002	.001	-.009	-3.000	.003	.637	1.569
PulmonaryAsthma	-.001	.001	-.004	-1.515	.130	.770	1.299
AlzheimerorDementia	.003	.002	.007	1.248	.212	.166	6.017
KidneyDisease	-1.351E-6	.001	.000	-.001	.999	.381	2.625
LiverDisease	.000	.001	-.001	-.457	.648	.503	1.990
EndocrineCondition	3.034E-5	.001	.000	.040	.968	.678	1.475
NeurologicalCondition	-.001	.002	-.002	-.516	.606	.334	2.993
TraumaticBrainInjury	.000	.001	.000	-.119	.905	.371	2.698
JointDisease	-.003	.001	-.013	-4.671	.000	.724	1.380
Cancer	-.001	.001	-.003	-1.027	.304	.486	2.059
OtherChronicMedCondition	.001	.000	.008	2.791	.005	.713	1.403
NoChronicMedCondition	.002	.001	.011	3.206	.001	.435	2.297
UnknownChronicMedCondition	.002	.002	.011	1.088	.276	.058	17.330
Smokes	-5.668E-5	.000	.000	-.126	.899	.618	1.617
ReceivedSmokingMedication	.015	.001	.057	20.753	.000	.715	1.398
ReceivedSmokingCounseling	-.002	.001	-.009	-3.058	.002	.578	1.731
SeriousMentalIllness	.000	.000	.002	.715	.474	.806	1.241
PrincipalDiagnosisClass	.007	.001	.035	13.913	.000	.850	1.176

VIF>10

Multicollinearity Analysis Results

AdditionalDiagnosisClass	.002	.000	.021	8.637	.000	.930	1.075
SSICashAssistance	-.003	.000	-.023	-8.364	.000	.743	1.346
SSDICashAssistance	-.001	.000	-.008	-3.029	.002	.695	1.439
VeteransDisabilityBenefits	.012	.002	.028	5.997	.000	.252	3.962
VeteransCashAssistance	.011	.002	.023	5.059	.000	.254	3.941
PublicAssistanceCashProgram	-.002	.000	-.011	-4.207	.000	.861	1.162
OtherCashBenefits	-.001	.001	-.006	-2.309	.021	.818	1.223
MedicaidandMedicareInsurance	.001	.001	.003	.508	.612	.168	5.942
NoInsurance	.003	.001	.012	3.043	.002	.355	2.816
UnknownInsuranceCoverage	.008	.001	.020	7.012	.000	.644	1.552
MedicaidInsurance	-.004	.001	-.023	-4.684	.000	.228	4.383
MedicaidManagedInsurance	.001	.000	.003	1.216	.224	.873	1.146
MedicareInsurance	-.001	.001	-.006	-1.054	.292	.160	6.234
PrivateInsurance	.002	.001	.008	2.348	.019	.425	2.356
ChildHealthPlusInsurance	.005	.001	.011	4.079	.000	.700	1.428
OtherInsurance	-.002	.001	-.007	-2.520	.012	.787	1.270
CriminalJusticeStatus	-.003	.001	-.017	-6.118	.000	.741	1.350

Correlation Heatmap



Analysis Method 1

Logistic Regression

- Because all data is categorical, it needs to be converted to numeric data using dummy variables
- All rows containing missing values are removed
- Each response to each variable is given its own column in the data set, and a value of 0 or 1 indicates the presence of a response for each patient
- The data is split by random sampling into training data and testing data
- The logistic regression model is built using the training data
- All variables are used in the initial iteration of the model, and then each iteration removes the least significant variable until all variables are significantly contributing to the model (backwards-fitting)
- Accuracy is checked by using the model on the testing data and checking the percentages of correct and incorrect predictions

Analysis Method 2

Logistic Regression

- Similar to Method 1, but with a few important changes
- Rows with missing data are not removed
- Categorical responses are not made into new variables, all current variables stay the same
- Binary categorical responses coded: “Yes” = 1, “No” = -1, “Unknown” = 0
- Responses to all other variables are also given numeric representations, and all “Unknown” responses are marked as 0
- Data is split into testing and training data, the model is built with the training data, backwards-fitting is used to clean the model, and the accuracy is tested with the testing data

Results: Method 1

- The first iteration of the model yields many variables showing up as NA, as well as many non-significant variables
  - The NA's are partially being caused by the small sample size of emergency cases
  - Many cases where either zero patients or all patients are positive for one of the variables
  - Accuracy = 19%
- After removing NA variables and backwards-fitting the model, 40 out of 145 variables remain
  - Accuracy = 42.5%

Initial Model

	Predicted Categories		
Actual Categories		Non-emergency	Emergency
	Non-emergency	2,886	12,787
	Emergency	5	163

Final Model

	Predicted Categories		
Actual Categories		Non-emergency	Emergency
	Non-emergency	6,851	8,822
	Emergency	12	156

Variable Description	**Coefficient**	**Significance Level**
Hospital in Long Island Region	1.3685	* * *
Hospital in New York City Region	1.1300	* * *
Hospital in Western Region	1.7658	* * *
Race: White	-0.5480	* * *
Lives With Others	0.4947	* *
Veteran	-1.3078	*
Unemployed and Not Looking for Work	0.5445	* *
Unemployed and Looking for Work	0.7992	* *
Has a Mental Illness	-0.9806	* * *
Has a Developmental Disability: Other	0.7508	* *

Significance Codes: * * * = $p < 0.001$, * * = $p < 0.01$, * = $p < 0.05$

Variable Description	Coefficient	Significance Level
Has an Alcohol Related Disorder	0.5200	* *
Has a Drug or Substance Disorder	0.3571	*
Has a Mobility Impairment Disorder	-1.2322	*
Has Obesity	-0.7789	* *
Has Cardiac Problems: Other	-1.2057	*
Has Joint Disease	-0.8054	*
Received Smoking Medication	0.8460	* * *
Received Smoking Counseling	-0.3903	*
Principal Diagnosis: Mental Illness	-1.3391	* * *
Additional Diagnosis: Addiction Disorder	0.7438	* * *

Significance Codes: * * * = $p < 0.001$, * * = $p < 0.01$, * = $p < 0.05$

Variable Description	Coefficient	Significance Level
Receives SSI Benefits	-0.4712	* *
Receives SSDI Benefits	-0.5484	*
Receives Veterans Disability Benefits	1.5728	*
Public Cash Assistance Program	-0.4775	* *
Receives Other Cash Benefits	-0.8798	*
Has Medicaid Insurance	-0.3938	*
Has a Criminal Record	-0.4258	*
From Zip Code 103	0.8831	*
From Zip Code 104	0.7229	* *
From Zip Code 112	1.1074	* * *

Significance Codes: * * * = $p < 0.001$, * * = $p < 0.01$, * = $p < 0.05$

Variable Description	Coefficient	Significance Level
From Zip Code 121	2.5436	* * *
From Zip Code 127	2.2936	* *
From Zip Code 128	2.1777	* * *
From Zip Code 131	2.0432	* * *
From Zip Code 132	1.9422	* * *
From Zip Code 141	1.1905	*
From Zip Code 144	1.2257	* *
From Zip Code 145	1.4812	* * *
From Zip Code 146	1.4716	* * *
Homeless	2.6342	* * *

Significance Codes: * * * = $p < 0.001$, * * = $p < 0.01$, * = $p < 0.05$

Results: Method 2

- The first iteration of the model yields no issues with NA variables, as well as fewer non-significant variables than the first iteration in Method 1
    - Accuracy = 92.2%
- After backwards-fitting the model, 36 out of 66 variables remain
    - Accuracy = 95%

Initial Model

	Predicted Categories		
Actual Categories		Non-emergency	Emergency
	Non-emergency	72,232	5,370
	Emergency	815	679

Final Model

	Predicted Categories		
Actual Categories		Non-emergency	Emergency
	Non-emergency	74,625	2,977
	Emergency	950	544

Variable Description	Coefficient	Significance Level
Region Hospital is Located in	0.17575	* * *
Adult or Child Age Group Status	0.37401	* * *
Sex	-0.05612	*
Transgender Status	0.82415	* * *
Hispanic Ethnicity Status	-0.06765	*
Living in an Institution or Private Residence Status	0.22279	* * *
Living Alone or With Others Status	-0.27132	* * *
Employment Status	0.14010	* * *
Education Status	-0.21853	* * *
Mental Illness Status	-0.37251	* * *
Intellectual Disability Status	0.14246	* *
Autism Spectrum Status	-0.22607	* *

Significance Codes: * * * = $p < 0.001$, * * = $p < 0.01$, * = $p < 0.05$

Variable Description	Coefficient	Significance Level
Alcohol Related Disorder Status	0.20748	* * *
Drug or Substance Use Disorder Status	0.26953	* * *
Hyperlipidemia Status	-0.24711	* * *
High Blood Pressure Status	-0.09502	*
Obesity Status	-0.29746	* * *
Other Cardiac Issues Status	-0.42766	* * *
Pulmonary Asthma Status	-0.14725	* *
Liver Disease Status	-0.23336	*
Joint Disease Status	-0.34244	* * *
Unknown Chronic Medical Condition Status	0.72123	* * *
Smoking Medication Status	0.53686	* * *
Principal Diagnosis	0.41577	* * *

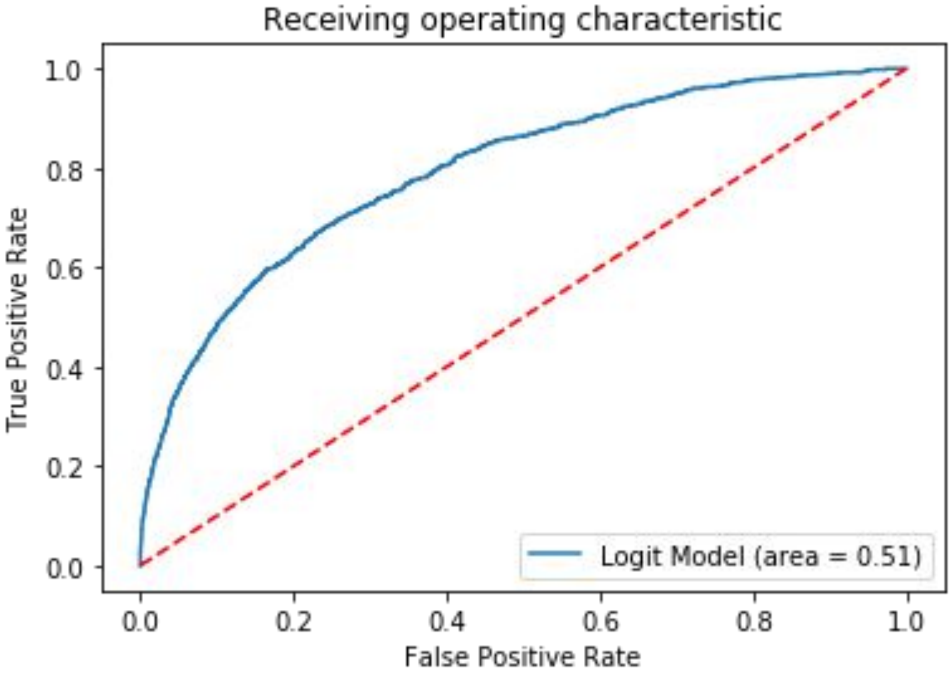
Significance Codes: * * * = $p < 0.001$, * * = $p < 0.01$, * = $p < 0.05$

Variable Description	Coefficient	Significance Level
Additional Diagnosis	-0.05497	* *
SSI Benefits Status	-0.17096	* * *
SSDI Benefits Status	-0.12541	* *
Veterans Disability Benefits Status	0.73193	* * *
Public Cash Assistance Program Status	-0.17058	* * *
Status of Having Any Insurance	0.30650	* * *
Medicaid Insurance Status	-0.11313	* *
Private Insurance Status	0.17603	* * *
Child Health Plus Insurance Status	0.37085	* * *
Status of Having Other Insurance	-0.16390	* *
Criminal Justice Record Status	-0.09696	*
Homelessness Status	1.10805	* * *

Significance Codes: * * * = $p < 0.001$, * * = $p < 0.01$, * = $p < 0.05$

Receiver Operating Characteristic Curve

Since the curve is smooth in the graph, it indicates that the logistic model obtained isn't overfitting. Meanwhile, it shows that the model can still be improved on the prediction performance of program category, especially on prediction of "EMERGENCY" category.

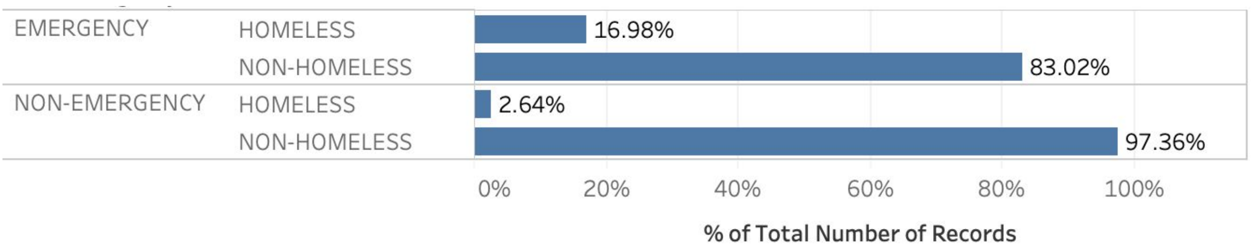


Discussion

Notable Insights

- Common theme across both models: significance of indicators of poverty
  - Not all indicators of poverty are significant, and not all significant variables are indicators of poverty
- Particularly influential variable: Homelessness
  - Had the smallest p-value in both models ($p < 2e-16$)

Descriptive Data for Homelessness in Dataset



Discussion & Next Steps for Stakeholders

The model we built can be used by hospital leadership, Emergency department administration, Quality Improvement managers, primary care providers, and others to understand which patients are at a higher risk of having an emergency room visit due to their mental health disorder & other key demographic information. These stakeholders can use this information to properly support those who are flagged as high risk and as a result try to decrease the number of emergency room visits.

Next Steps for Stakeholders:

- Apply the model to patient population of a selected region/hospital
- Share data model results with providers in an easy to read and follow dashboard
- If well received by the providers, integrate model into current EHR, scheduling system, and hospital workflow to flag patients in real time so providers can better manage their preventative care

Limitations to Keep in Mind: Stakeholders need to understand that this model was built on one week's worth of patient data in New York State. We are making the assumption that these data points and rates of emergency room visits would be reflective of an average week.

Ideal World- Room for Improvement

Room for improvement:

- If the survey was reflective of a month, quarter, or even a year’s worth of data, there would be more data available to build and test the model. This could improve the accuracy and trust by the stakeholders of the model.
- If the survey included data from a wider area, perhaps the Mid-Atlantic states (New York, New Jersey, Pennsylvania, Delaware, and Maryland) the model could impact a wider patient population. The model could also reveal different high risk red flags for patients from different states. Which would be interesting to insurance companies who have policy holders in different states.

Summary

- Improving care coordination for patients with mental health disorders should be of high importance for healthcare institutions
- Implementing a high risk patient model into the workflow of the institution's Electronic Health Record (EHR) and clinical decisions support tools is an important step in improving the quality of care given to this patient population
- While the model has a few limitations, it is a good place to start and can be improved each year the survey is completed

QUESTIONS?