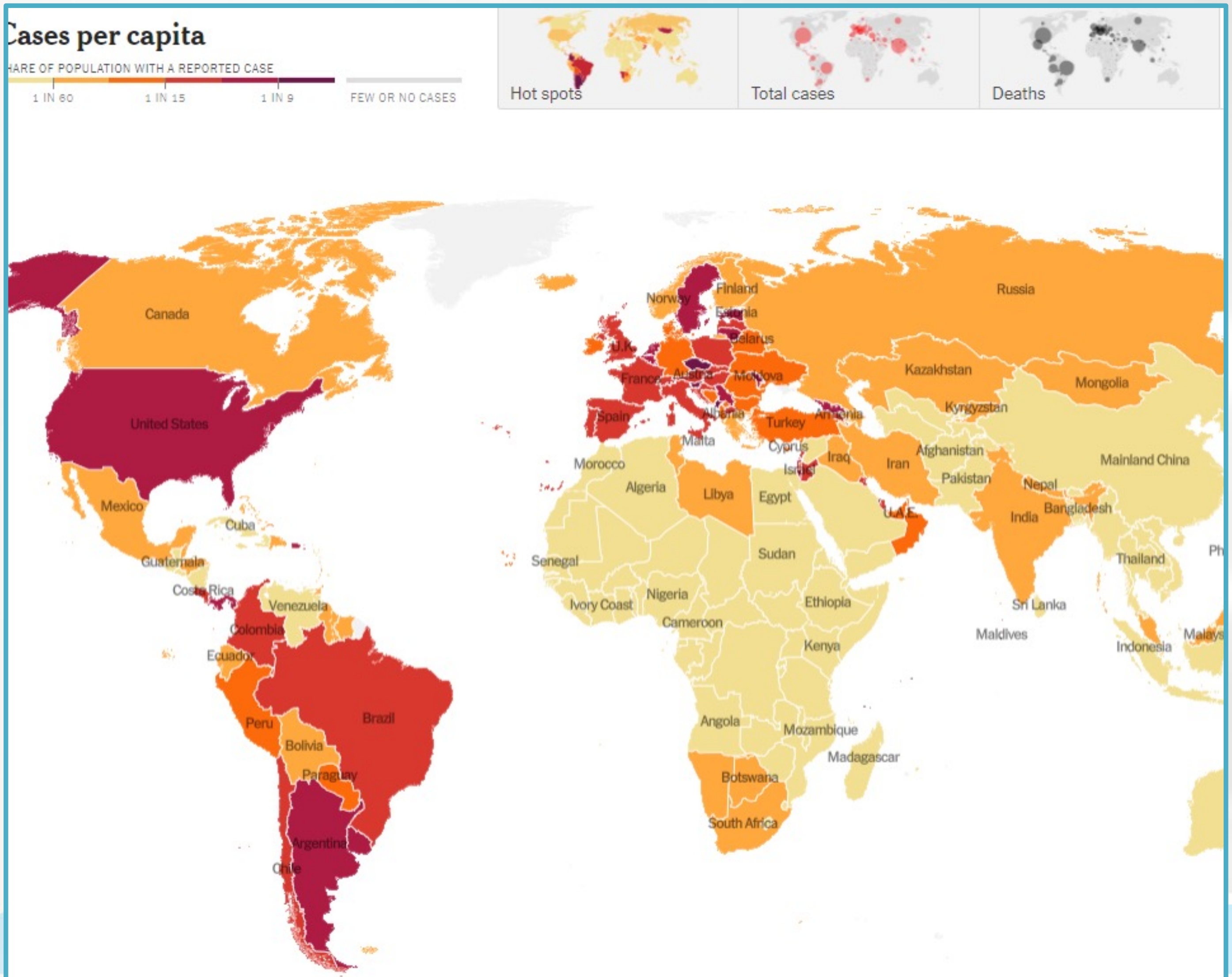
Hospital Analytics

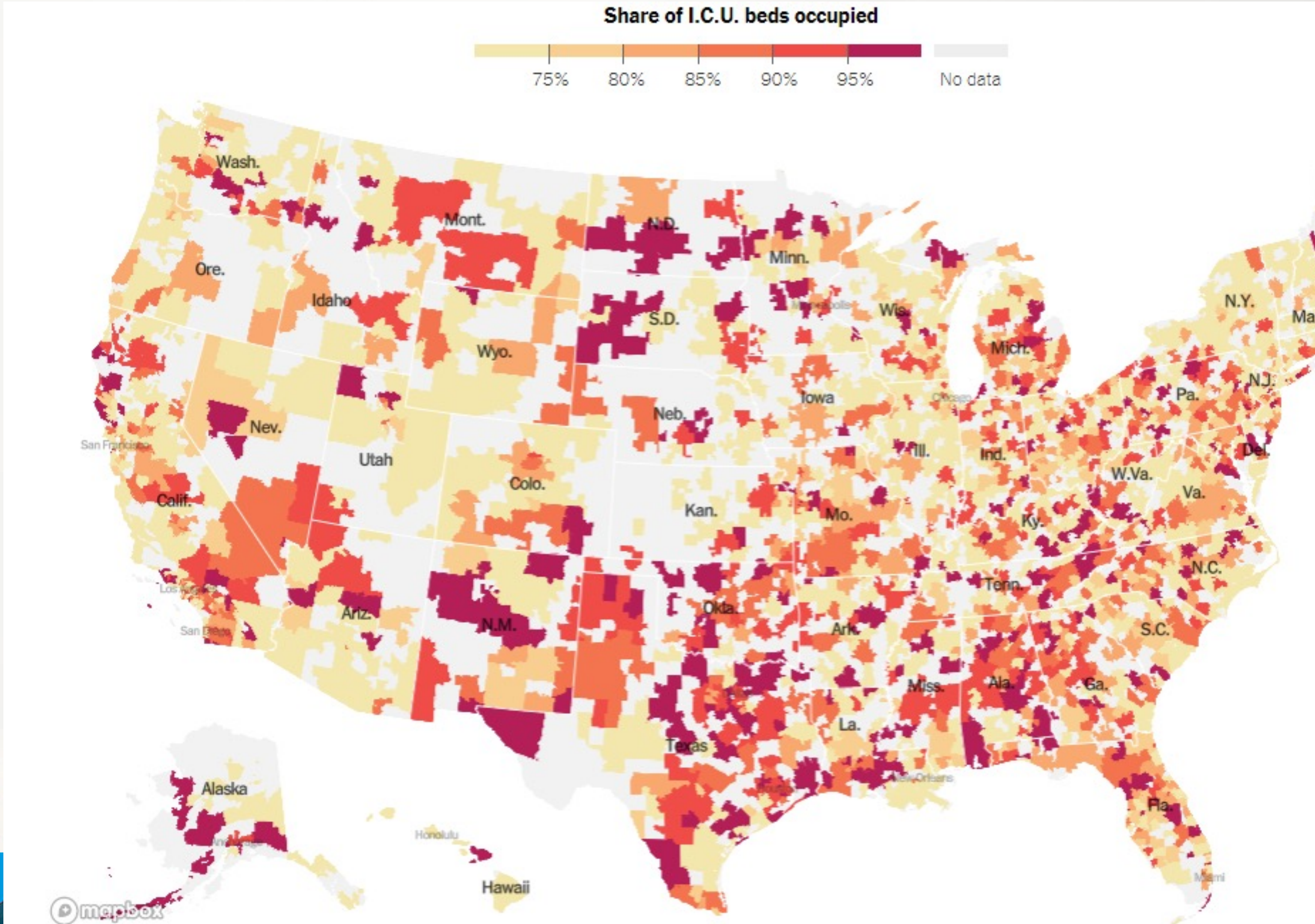
Determining Patient Length of Stay

- Elissa Carroll, Dylan Fajardo, Tim Hulak, Jason Tompkins

Introduction

- 31Dec2019: 1st case in Wuhan China
- 9Jan2020: Identified as novel coronavirus
- Jan 2020: Cases reported in US
- Feb 2020: Some states with community spread
- March 2020: All States with community spread




Share of I.C.U. beds occupied
75%
80%
85%
90%
95%
No data
Wash.
Mont.
N.D.
Minn.
Ore.
Idaho
S.D.
Wis.
N.Y.
Wyo.
Mich.
Iowa
Pa.
N.J.
Nev.
Neb.
Ill.
Ind.
Del.
Utah
W.Va.
Colo.
Va.
Calif.
Kan.
Mo.
Ky.
N.C.
Tenn.
Okla.
Ariz.
N.M.
Ark.
S.C.
Miss.
Ala.
Ga.
La.
Texas
Fla.
Alaska
Honolulu
Hawaii
mapbox
fppt.com

- *Health Care Management* is the overall management of all aspects of a hospital or clinic. Are there areas that could have been better managed?
  - Patient Length of Stay (LOS)
    - Optimizing hospital efficiency and flow
    - Increased likelihood to develop hospital-related infection

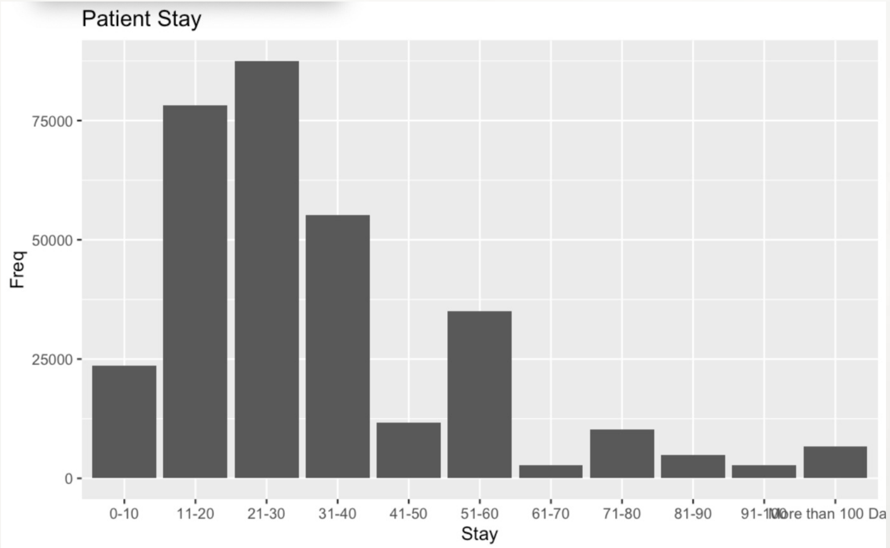
About the Data

Attribute	Description/Values
case_id	CaseID given by Hosp
Hospital_code	Code for Hosp (1-38)
Hospital_type_code	Code for type of Hosp (a-g)
City_Code_Hospital	City code for Hosp (1-13)
Hospital_region_code	Hosp Region Code (X,Y,Z)
Available.Extra.Rooms.in.Hospital	**Num available rooms**
Department	**Radiotherapy, Anesthesia, Gynecology, TB & Chest disease, Surgery**
Ward_Type	**Type of Ward (P,Q,R,S,T,U)**
Ward_Facility_Code	Facility of Ward (A,B,C,D,E,F)
Bed.Grade	**Condition of Bed in Ward (1,2,3,4,NA)**
Patientid	Unique Patient identifier
City_Code_Patient	City code for Patient (1-38)
Type.of.Admission	**Emergency, Trauma, Urgent**
Severity.of.Illness	**Recorded @ admission (Extreme, Moderate, Minor)**
Visitors.with.Patient	**Num visitors with patient (0-32)**
Age	**10 bins, no specific age**
Admission_Deposit	**$1,800-$11,008**
Stay (Length of Stay, LOS)	**11 bins; no exact value.**

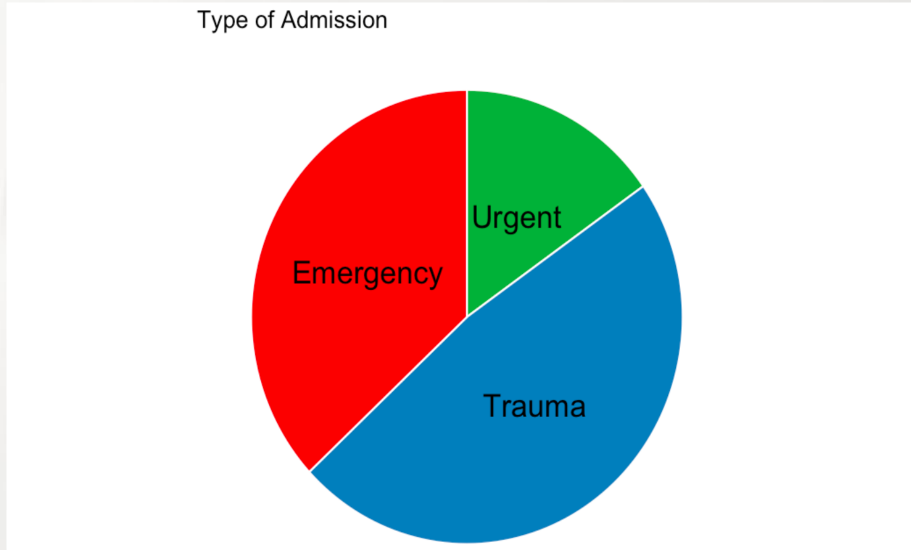
	case_id <int>	Hospital_code <int>	Hospital_type_code <chr>	City_Code_Hospital <int>	Hospital_region_code <chr>	Available.Extra.Rooms.in.Hospital <int>	Department <chr>	Ward_Type <chr>
1	1	8	c	3	Z	3	radiotherapy	R
2	2	2	c	5	Z	2	radiotherapy	S
3	3	10	e	1	X	2	anesthesia	S
4	4	26	b	2	Y	2	radiotherapy	R
5	5	26	b	2	Y	2	radiotherapy	S
6	6	23	a	6	X	2	anesthesia	S

Ward_Facility_Code <chr>	Bed.Grade <dbl>	patientid <int>	City_Code_Patient <dbl>	Type.of.Admission <chr>	Severity.of.Illness <chr>	Visitors.with.Patient <int>	Age <chr>	Admission_Deposit <dbl>	Stay <chr>
F	2	31397	7	Emergency	Extreme	2	51-60	4911	0-10
F	2	31397	7	Trauma	Extreme	2	51-60	5954	41-50
E	2	31397	7	Trauma	Extreme	2	51-60	4745	31-40
D	2	31397	7	Trauma	Extreme	2	51-60	7272	41-50
D	2	31397	7	Trauma	Extreme	2	51-60	5558	41-50
F	2	31397	7	Trauma	Extreme	2	51-60	4449	11-20

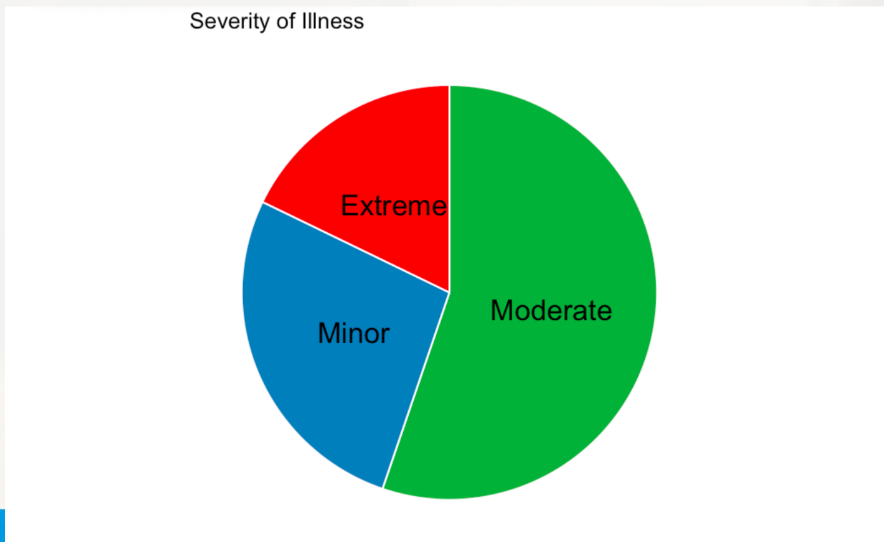
EXPLORATORY DATA ANALYSIS



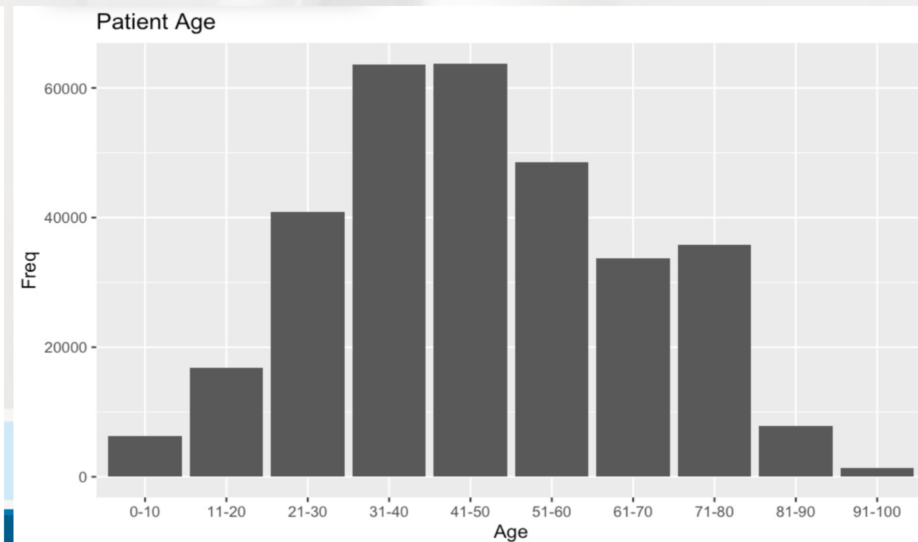
Patient Stay
75000
50000
25000
0
Freq
0-10
11-20
21-30
31-40
41-50
51-60
61-70
71-80
81-90
More than 100 Da
Stay



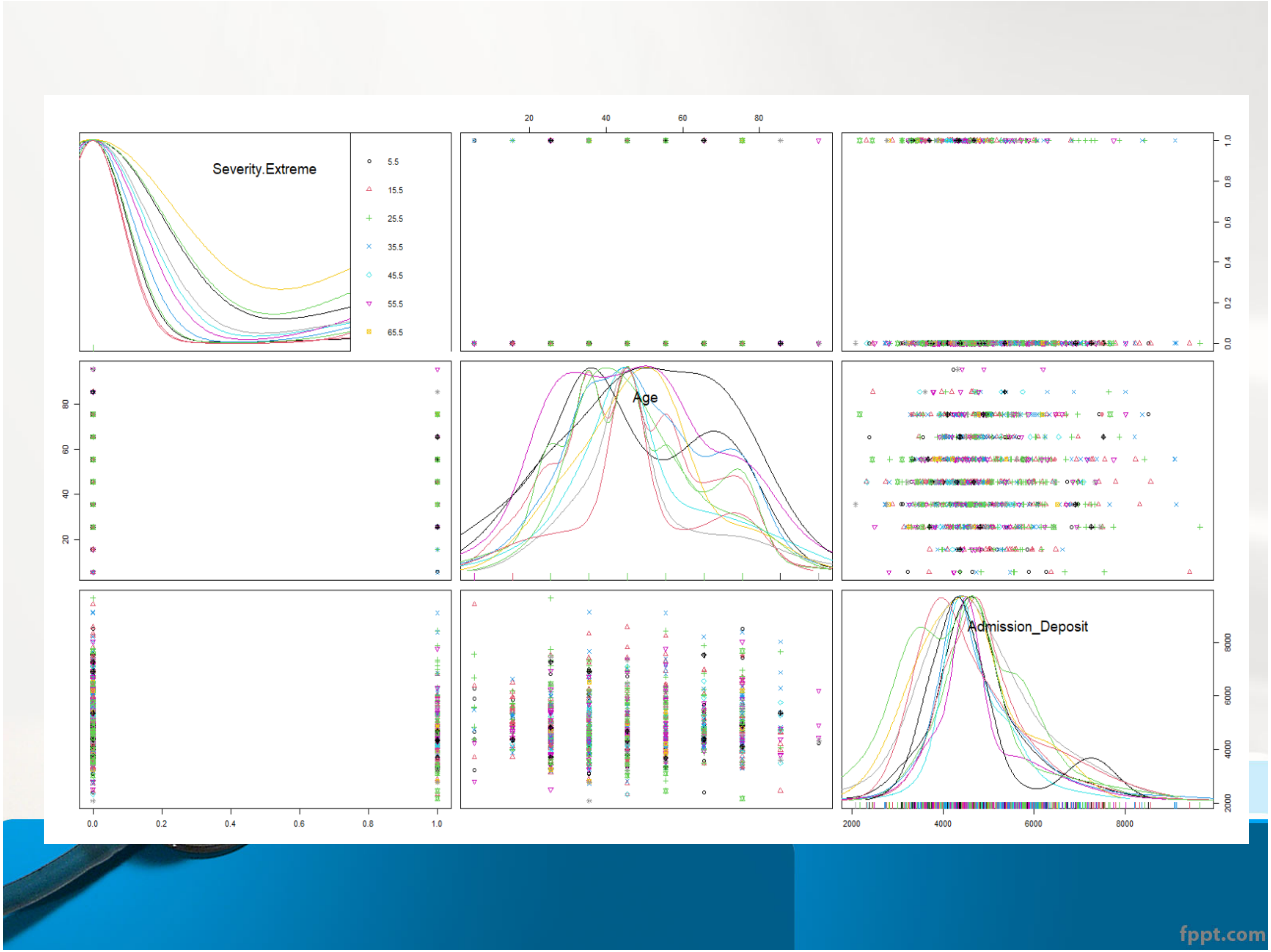
Type of Admission
Urgent
Emergency
Trauma



Severity of Illness
Extreme
Moderate
Minor



Patient Age
60000
40000
20000
0
Freq
0-10
11-20
21-30
31-40
41-50
51-60
61-70
71-80
81-90
91-100
Age



Severity.Extreme
5.5
15.5
25.5
35.5
45.5
55.5
65.5
Age
Admission_Deposit
fppt.com

	Length of Stay						
	0-20 (N=101743)	**21-40 (N=142650)**	**41-60 (N=46761)**	**61-80 (N=12998)**	**81-100 (N=7603)**	**More than 100 Days (N=6683)**	**Total (N=318438)**
Hospital_type_code							
a	50366 (49.5%)	64260 (45.0%)	18728 (40.1%)	4646 (35.7%)	2852 (37.5%)	2573 (38.5%)	143425 (45.0%)
b	19078 (18.8%)	31926 (22.4%)	10802 (23.1%)	3525 (27.1%)	1809 (23.8%)	1806 (27.0%)	68946 (21.7%)
c	13906 (13.7%)	20748 (14.5%)	7048 (15.1%)	2041 (15.7%)	1164 (15.3%)	1021 (15.3%)	45928 (14.4%)
d	5737 (5.6%)	9233 (6.5%)	3473 (7.4%)	947 (7.3%)	601 (7.9%)	398 (6.0%)	20389 (6.4%)
e	8449 (8.3%)	10584 (7.4%)	3596 (7.7%)	1021 (7.9%)	613 (8.1%)	507 (7.6%)	24770 (7.8%)
f	3355 (3.3%)	3990 (2.8%)	2164 (4.6%)	526 (4.0%)	392 (5.2%)	276 (4.1%)	10703 (3.4%)
g	852 (0.8%)	1909 (1.3%)	950 (2.0%)	292 (2.2%)	172 (2.3%)	102 (1.5%)	4277 (1.3%)

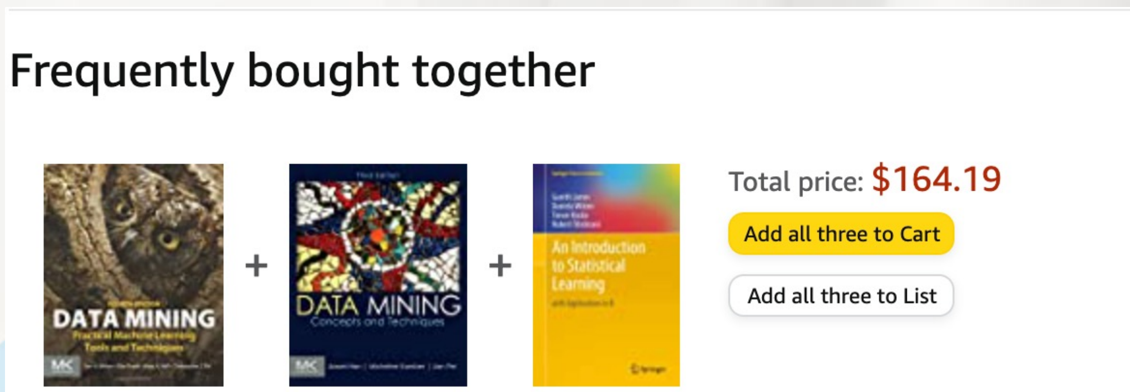
ANALYSIS AND MODELS

Analysis and Models

- ARM
- Decision Tree
- KNN
- Naïve Bayes

Analysis and Models: ARM

- Association Rule Mining is typically used to identify “rules” between items in an itemset.
- For example, Amazon recommends items that are frequently bought together in order to increase the likelihood that a customer will purchase more items.



Analysis and Models: ARM

- For this analysis, Association Rule mining was used to determine if there were any rules that may help identify how long a patient will stay in the hospital based in their Age, Severity of Illness, Type of Admission, and which Department they were admitted for.
- The Apriori algorithm was used to make these connections. This algorithm identifies the frequent individual items in the dataset and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the dataset.
- Apriori pruning principle: If there is any itemset that is infrequent, its superset should not be generated or tested!
- Apriori Primarily 3 Metrics to evaluate the rules:
  - **Lift:** Measure of dependent or correlated events. Association rules should have >1 lift to be meaningful.
  - **Support:** Fraction of transactions that contain both X and Y
  - **Confidence:** How frequently items in Y appear in transactions that contain X

Analysis and Models: ARM
Stay = 0 – 10 Days

- Severity of Illness: Minor to Moderate
- Admission Type: Emergency
- Department : Gynecology
  - **Support:** 0.015 – 0.037
  - **Confidence:** 0.072 – 0.137
  - **Lift:** 0.97 – 1.85

	lhs <chr>	<chr>	rhs <chr>	support <dbl>	confidence <dbl>	coverage <dbl>	lift <dbl>	count <int>
[1]	{Type.of.Admission=Emergency,Severity.of.Illness=Minor}	=>	{Stay=0-10}	0.015	0.137	0.11	1.85	4842
[2]	{Department=gynecology,Type.of.Admission=Emergency}	=>	{Stay=0-10}	0.037	0.122	0.30	1.64	11663
[3]	{Department=gynecology,Type.of.Admission=Emergency,Severity.of.Illness=Moderate}	=>	{Stay=0-10}	0.019	0.119	0.16	1.60	5991
[4]	{Type.of.Admission=Emergency,Severity.of.Illness=Moderate}	=>	{Stay=0-10}	0.023	0.117	0.20	1.58	7320
[5]	{Department=gynecology,Severity.of.Illness=Minor}	=>	{Stay=0-10}	0.020	0.093	0.22	1.25	6455
[6]	{Department=gynecology,Severity.of.Illness=Moderate}	=>	{Stay=0-10}	0.031	0.072	0.43	0.97	9785

Analysis and Models: ARM
Stay = 11 – 20 Days

- Severity of Illness: Minor
- Admission Type: Emergency / Trauma
- Department : Gynecology
- Age: Young to Middle Aged (21-30 and 31-40)
  - **<u>Support:</u>** 0.016 – 0.067
  - **<u>Confidence:</u>** 0.30 - 0.32
  - **<u>Lift:</u>** 1.2 – 1.3

	lhs <chr>	<chr>	rhs <chr>	support <dbl>	confidence <dbl>	coverage <dbl>	lift <dbl>	count <int>
[1]	{Severity.of.Illness=Minor,Age=21-30}	=>	{Stay=11-20}	0.016	0.32	0.048	1.3	4992
[2]	{Type.of.Admission=Trauma,Severity.of.Illness=Minor}	=>	{Stay=11-20}	0.036	0.31	0.116	1.3	11545
[3]	{Severity.of.Illness=Minor,Age=31-40}	=>	{Stay=11-20}	0.016	0.31	0.052	1.3	5141
[4]	{Department=gynecology,Severity.of.Illness=Minor}	=>	{Stay=11-20}	0.067	0.31	0.218	1.3	21476
[5]	{Type.of.Admission=Emergency,Severity.of.Illness=Minor}	=>	{Stay=11-20}	0.034	0.31	0.111	1.3	10846
[6]	{Department=gynecology,Type.of.Admission=Emergency,Severity.of.Illness=Minor}	=>	{Stay=11-20}	0.029	0.31	0.093	1.2	9091
[7]	{Department=gynecology,Type.of.Admission=Trauma,Severity.of.Illness=Minor}	=>	{Stay=11-20}	0.027	0.30	0.090	1.2	8634

Analysis and Models: ARM
Stay = 21 – 30 Days

- Severity of Illness: Moderate
- Admission Type: Trauma
- Department : Gynecology
- Age: Middle-Aged to Older (31-40 and 41-50)
  - **Support:** 0.016 – 0.087
  - **Confidence:** 0.32 – 0.35
  - **Lift:** 1.2 – 1.3

	lhs <chr>	<chr>	rhs <chr>	support <dbl>	confidence <dbl>	coverage <dbl>	lift <dbl>	count <int>
[1]	{Department=gynecology,Type.of.Admission=Trauma,Severity.of.Illness=Moderate,Age=31-40}	=>	{Stay=21-30}	0.016	0.35	0.045	1.3	4989
[2]	{Type.of.Admission=Trauma,Severity.of.Illness=Moderate,Age=31-40}	=>	{Stay=21-30}	0.019	0.35	0.055	1.3	6112
[3]	{Department=gynecology,Type.of.Admission=Trauma,Severity.of.Illness=Moderate,Age=41-50}	=>	{Stay=21-30}	0.016	0.34	0.047	1.2	5002
[4]	{Type.of.Admission=Trauma,Severity.of.Illness=Moderate,Age=41-50}	=>	{Stay=21-30}	0.020	0.33	0.059	1.2	6267
[5]	{Department=gynecology,Type.of.Admission=Trauma,Age=31-40}	=>	{Stay=21-30}	0.025	0.33	0.075	1.2	7876
[6]	{Type.of.Admission=Trauma,Age=31-40}	=>	{Stay=21-30}	0.030	0.33	0.093	1.2	9671
[7]	{Department=gynecology,Type.of.Admission=Trauma,Severity.of.Illness=Moderate}	=>	{Stay=21-30}	0.066	0.32	0.204	1.2	21057
[8]	{Type.of.Admission=Trauma,Severity.of.Illness=Moderate}	=>	{Stay=21-30}	0.087	0.32	0.272	1.2	27775

Analysis and Models: ARM
Stay = 31 – 40 Days

- Severity of Illness: Mixed
- Admission Type: Trauma
- Department : Gynecology
- Age: Middle-Aged to Older (31-40 or 41-50)
  - **Support:** 0.017 - 0.068
  - **Confidence:** 0.18 – 0.19
  - **Lift:** 1.0 – 1.1

	lhs <chr>	<chr>	rhs <chr>	support <dbl>	confidence <dbl>	coverage <dbl>	lift <dbl>	count <int>
[1]	{Department=gynecology,Type.of.Admission=Trauma,Severity.of.Illness=Minor}	=>	{Stay=31-40}	0.017	0.19	0.090	1.1	5386
[2]	{Department=gynecology,Type.of.Admission=Trauma}	=>	{Stay=31-40}	0.068	0.19	0.362	1.1	21569
[3]	{Department=gynecology,Type.of.Admission=Trauma,Severity.of.Illness=Moderate}	=>	{Stay=31-40}	0.038	0.19	0.204	1.1	12129
[4]	{Type.of.Admission=Trauma,Age=31-40}	=>	{Stay=31-40}	0.017	0.19	0.093	1.1	5503
[5]	{Type.of.Admission=Trauma,Severity.of.Illness=Minor}	=>	{Stay=31-40}	0.021	0.19	0.116	1.1	6823
[6]	{Type.of.Admission=Trauma,Severity.of.Illness=Extreme}	=>	{Stay=31-40}	0.017	0.18	0.091	1.1	5326
[7]	{Type.of.Admission=Trauma,Severity.of.Illness=Moderate}	=>	{Stay=31-40}	0.050	0.18	0.272	1.1	15958
[8]	{Type.of.Admission=Trauma,Age=41-50}	=>	{Stay=31-40}	0.018	0.18	0.099	1.1	5771
[9]	{Department=gynecology,Severity.of.Illness=Extreme}	=>	{Stay=31-40}	0.025	0.18	0.137	1.0	7835

Analysis and Models: ARM
Stay = 71 – 80 Days

- Severity of Illness: Extreme
- Admission Type: Trauma
- Department : Gynecology
- Age: Older to Senior (41 – 50 and 71 – 80)
  - **Support:** 0.0032 – 0.0064
  - **Confidence:** 0.041 – 0.051
  - **Lift:** 1.3 – 1.6

	lhs <chr>	<chr>	rhs <chr>	support <dbl>	confidence <dbl>	coverage <dbl>	lift <dbl>	count <int>
[1]	{Department=gynecology,Type.of.Admission=Trauma,Severity.of.Illness=Extreme}	=>	{Stay=71-80}	0.0035	0.051	0.068	1.6	1106
[2]	{Type.of.Admission=Trauma,Severity.of.Illness=Extreme}	=>	{Stay=71-80}	0.0044	0.049	0.091	1.5	1406
[3]	{Department=gynecology,Severity.of.Illness=Extreme}	=>	{Stay=71-80}	0.0064	0.047	0.137	1.5	2040
[4]	{Department=gynecology,Age=71-80}	=>	{Stay=71-80}	0.0032	0.041	0.078	1.3	1034
[5]	{Department=gynecology,Type.of.Admission=Trauma,Age=41-50}	=>	{Stay=71-80}	0.0032	0.041	0.078	1.3	1012

Analysis and Models: ARM
Stay = More than 100 Days

- Severity of Illness: Extreme / Moderate
- Admission Type: Trauma / Emergency
- Department : Gynecology / Radiotherapy
  - **<u>Support:</u>** 0.0016 – 0.0043
  - **<u>Confidence:</u>** 0.032 – 0.036
  - **<u>Lift:</u>** 1.5 - 1.7

	lhs <chr>	<chr>	rhs <chr>	support <dbl>	confidence <dbl>	coverage <dbl>	lift <dbl>	count <int>
[1]	{Type.of.Admission=Emergency,Severity.of.Illness=Extreme}	=>	{Stay=More than 100 Days}	0.0023	0.036	0.062	1.7	723
[2]	{Type.of.Admission=Trauma,Severity.of.Illness=Extreme}	=>	{Stay=More than 100 Days}	0.0033	0.036	0.091	1.7	1049
[3]	{Department=radiotherapy,Severity.of.Illness=Moderate}	=>	{Stay=More than 100 Days}	0.0017	0.034	0.051	1.6	546
[4]	{Department=gynecology,Type.of.Admission=Trauma,Severity.of.Illness=Extreme}	=>	{Stay=More than 100 Days}	0.0023	0.034	0.068	1.6	731
[5]	{Department=gynecology,Type.of.Admission=Emergency,Severity.of.Illness=Extreme}	=>	{Stay=More than 100 Days}	0.0016	0.032	0.049	1.5	502
[6]	{Department=gynecology,Severity.of.Illness=Extreme}	=>	{Stay=More than 100 Days}	0.0043	0.032	0.137	1.5	1376

Length of Stay Results: ARM

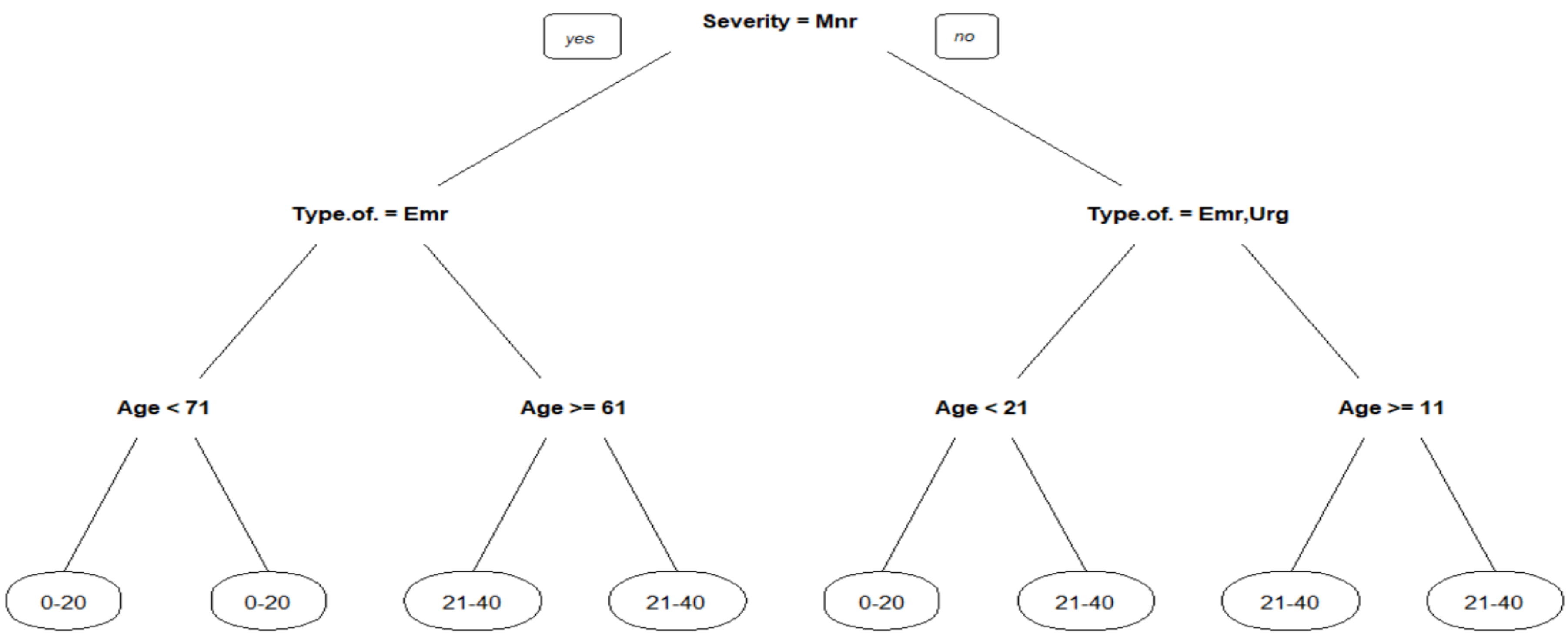
- **0-10 Days:** Minor to Moderate Emergency Injuries
- **11-20 Days:** Middle-aged patients with minor/moderate emergency injuries
- **21-30 Days:** Young to middle aged patients with Moderate trauma injuries
- **31-40 Days:** Older Trauma patients with Mixed illnesses
- **41-50 Days:** Moderate illness with Emergency/Trauma
- **51-60 Days:** Moderate illness with Emergency/Trauma
- **61 -70 Days:** Older to Senior patients with Extreme illness and trauma
- **71-80 Days:** Older to Senior patients with Moderate trauma
- **81-90 Days:** Moderate trauma patients
- **91-100 Days:** Moderate to Extreme Trauma patients
- **100+ Days:** Moderate to Extreme illness with Emergency / Trauma patients who are admitted to Radiotherapy

Summary: *Older, more severely injured patients can be expected to have an extended stay while younger patients with minor/moderate injuries tend to have a relatively abbreviated stay.*

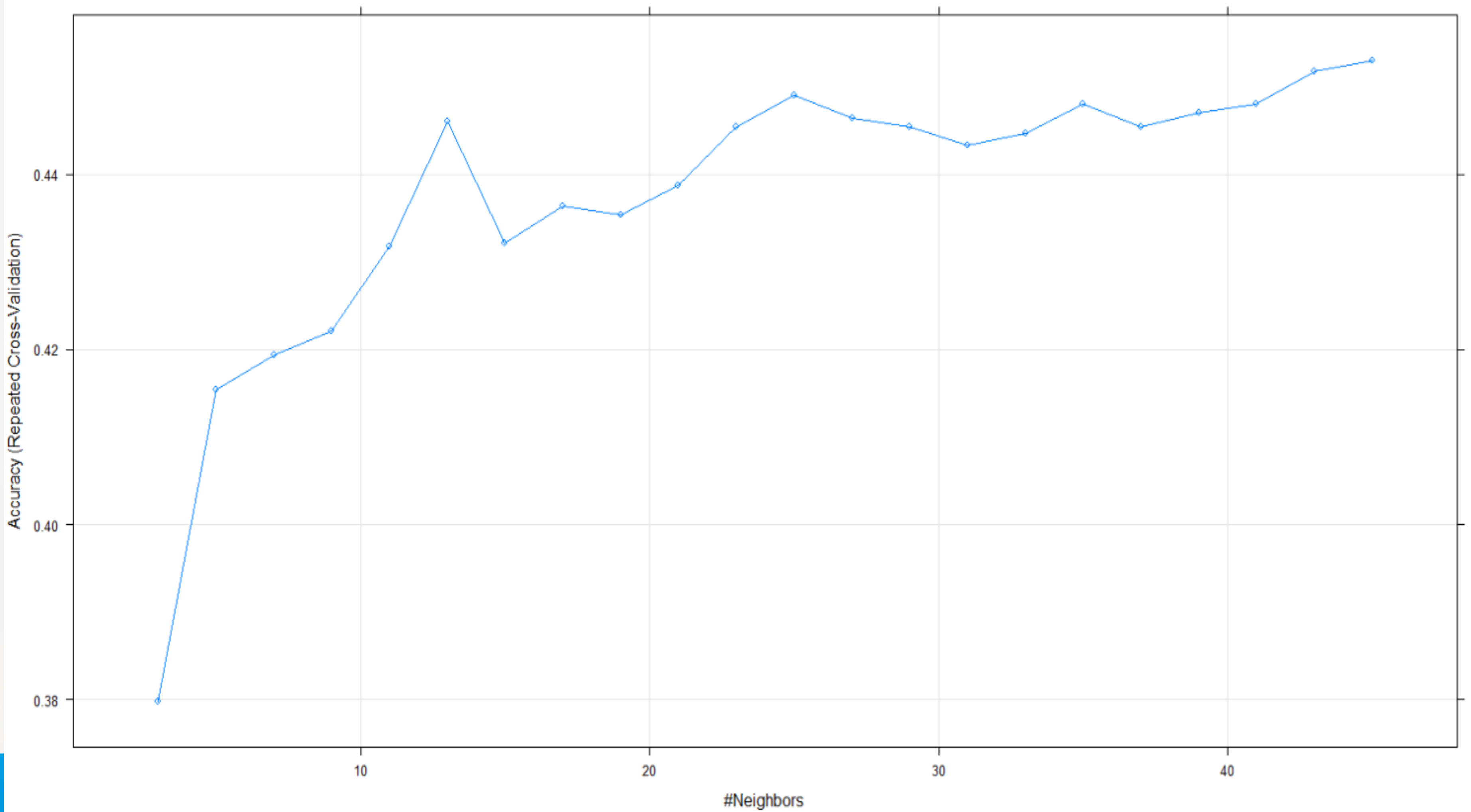
Analysis and Models: Decision Trees

- Tried 3 Different Trees
- Third decision tree
- Constraints
  - Minsplit 20, minbucket = 6, maxdepth, cp = -1
- 3 independent variables
  - Type of Admission, Severity of Illness, Age

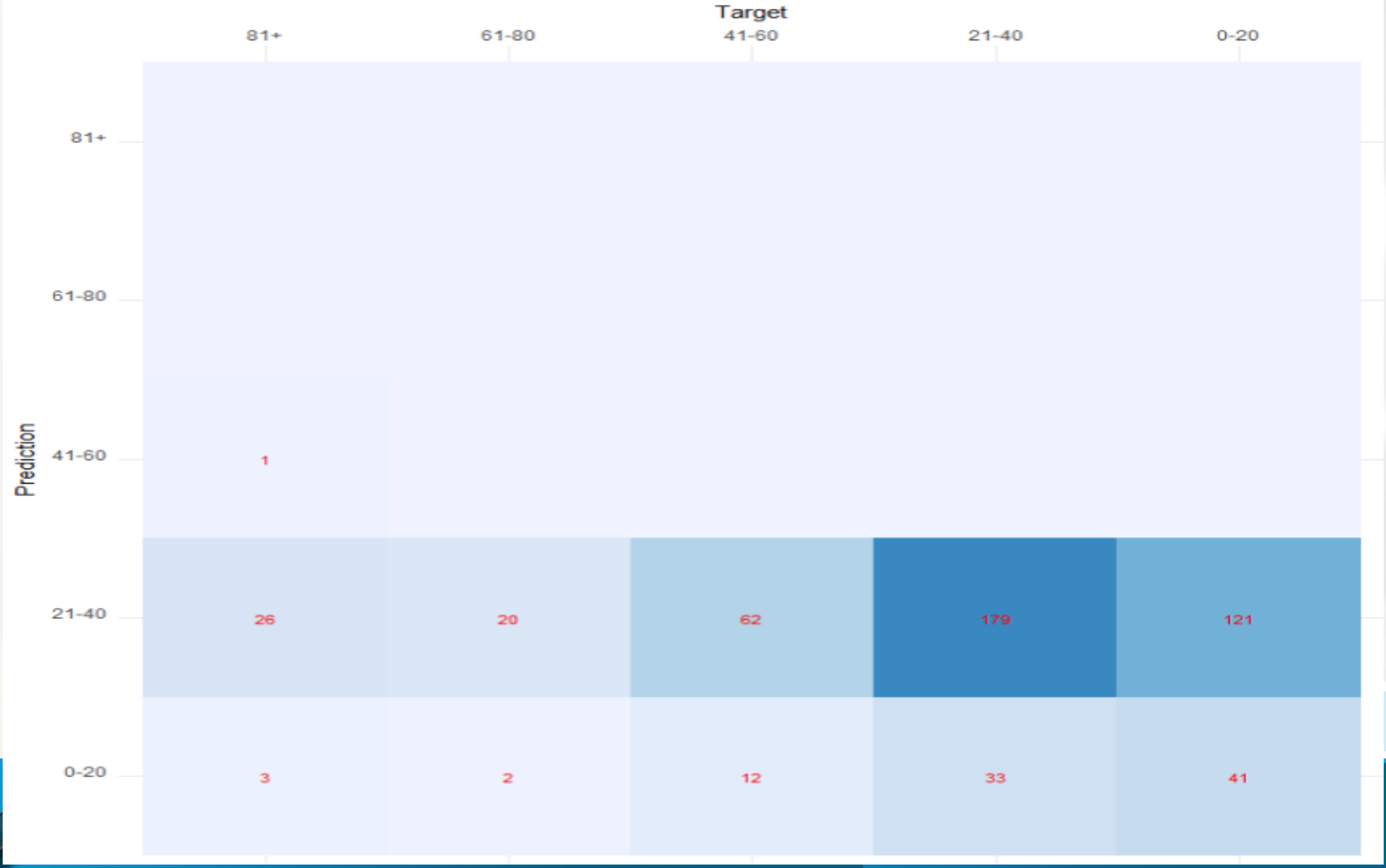
Analysis and Models: Decision Trees



Analysis and Models: KNN



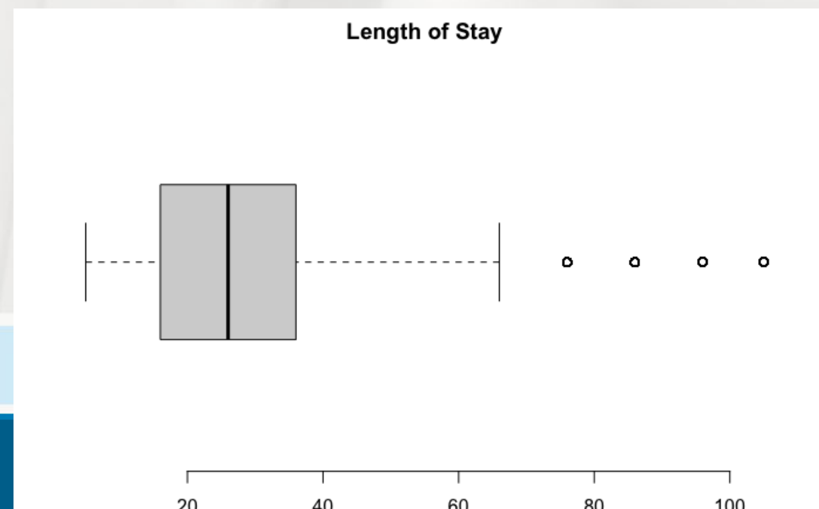
Analysis and Models: KNN (Cross-Validation)



Analysis and Models: Naïve Bayes

- Naïve Bayes is probabilistic algorithm based on Bayes theorem.
- How it works:
  - Step 1: Convert the data set into a frequency table
  - Step 2: Create a likelihood table by finding the probabilities
  - Step 3: Use Naïve Bayesian equation to calculate the posterior probability for each class
  - **The class with the highest posterior probability is the outcome of prediction**
- The model will be evaluated by its prediction accuracy

- For this analysis, Naïve Bayes was used to determine a patient's Length of Stay based on a variety of variables.
- Additionally, we binarized our Length of Stay variable to 'plus61' then used Naïve Bayes to determine whether a patient would stay longer than 60 days.
- All numerical variables were converted to factors.



Analysis and Models: Naïve Bayes - 1

- Variables: Hospital_type, Hospital_region, Extra.Rooms, Department, Ward_Type, Bed.Grade, Type.of.Admission, Severity.of.Illness, Age, Admission_Deposit
- Prediction Class: Stay
- Accuracy: 46.03%

	Reference					
Prediction	0	1	2	3	4	5
0	31820	27474	8873	2287	1478	1313
1	66807	110734	33893	9394	5020	4178
2	2932	4250	3855	1233	1023	1015
3	6	1	1	6	4	2
4	10	9	12	5	11	10
5	168	182	127	73	67	165

Analysis and Models: Naïve Bayes - 2

- Variables: Hospital_type, Hospital_region, Extra.Rooms, Department, Ward_Type, Bed.Grade, Type.of.Admission, Severity.of.Illness, Age, Admission_Deposit
- Prediction Class: plus61
- Accuracy: 91.30%

	Reference	
Prediction	0	1
0	290300	26864
1	854	420

Analysis and Models: Naïve Bayes – 3

- Variables: Hospital_type, Hospital_region, Extra.Rooms, Department, Ward_Type, Bed.Grade, Type.of.Admission, Severity.of.Illness, Age, Admission_Deposit, Hospital_code, Visitors
- Prediction Class: plus61
- Accuracy: 92.15%

Model 2

	Reference	
Prediction	0	1
0	290300	26864
1	854	420

Model 3

	Reference	
Prediction	0	1
0	284000	17840
1	7154	9444

Length of Stay Results: Naïve Bayes

Predicting ‘Stay’

- **Model 1:**
  - Accuracy: **46.03%**
- **Model 4: Include ’Hospital_code’ & ‘Visitors.with.Patient’**
  - Accuracy: **50.37%**

Predicting ‘plus61’

- **Model 2:**
  - Accuracy: **91.37%**
  - Low true positive rate (1.8%)
- **Model 3: Include ’Hospital_code’ & ‘Visitors.with.Patient’**
  - Accuracy: **92.15%**
  - True positive rate improved to 34.61%

Summary: A Naïve Bayes model may have been limited because ‘Stay’ was heavily concentrated between 0 and 40 days. Converting ‘Stay’ to a binary variable, ‘plus61’ yielded high accuracy, but our models were still limited in terms of true positive rate.

Conclusions

- Older, more severely injured patients can be expected to have an extended stay while younger patients with minor/moderate injuries tend to have a relatively abbreviated stay.
- Hospital code and patient visitors increased True Positive Rate, which supports the accuracy in predicting length of stay.
- Predicting 0-20 or 21-40 seem to be much more reliable in terms of predictors than any other length of stay.

Looking Forward

- Given the lower accuracy of some of the models, more nuanced data with respect to the length of stay, would greatly assist making predictions.

QUESTIONS