FE 512 Final Report

Database Design for Hospital Service in USA

Team13: Haitao Liu, Junhan Zhou

Introduction

There are so many medical centers in the United States and different hospital have different strength and weakness

How can we extract the information we want quickly?

How can we have a better understanding of the hospital across the United State?

How can the patients choose the hospital?

In this database, we collected the basic hospital information for more than 4000 medical centers and more than 400 long term care hospital in the United States. There are different aspects to measure in each hospital including infections, efficiency, complications and deaths. In each aspect, there are different specific measure types such as head CT efficiency, serious complicatinons, giving the score of performance for a certain hospital.

We want to build a hospital information system database for patients, hospital, and government. For patients, they can get to know the hospital information(hospital type, hospital location, hosipital phone number, and emergency service) and then make choices for different situations. For hospitals, they can work more efficiency and reduce the medical waste. For government, different types of hospitals will attract different types investors. Different hospitals have different situation, some hospital are very effeiciency, but they may not doing well at other aspects. So you can search information what can help you make a better choice in this database.

Data Source

- Where it comes from?
- What it is about?

There will be six tables in our database. Most of data are from data medicare gov. (https://data.medicare.gov/ (https://data.medicare.gov/))

1. Hospital General Information(https://data.medicare.gov/Hospital-Compare/Hospital-General-Information/xubh-q36u (https://data.medicare.gov/Hospital-Compare/Hospital-General-Information/xubh-q36u))

The first one is the Hospital info table, which includes the hospital's information (Provider_ID, Hospital_Name, State, City, ZIP_Code, Phone_Number, Hospital_Type, Hospital_ownership, Emergency_service).

2. Timely and Effective Care - Hospital(https://data.medicare.gov/Hospital-Compare/Timely-and-Effective-Care-Hospital/yv7e-xc69 (https://data.medicare.gov/Hospital-Compare/Timely-and-Effective-Care-Hospital/yv7e-xc69))

   The second table is the Efficiency table, which includes Measure information (Measure_ID, Provider_ID, Department, Measure_Name, Measure_Type, Measure_Start_Date, Measure_END_Date, Score).

3. Healthcare Associated Infections - Hospital(https://data.medicare.gov/Hospital-Compare/Healthcare-Associated-Infections-Hospital/77hc-ibv8 (https://data.medicare.gov/Hospital-Compare/Healthcare-Associated-Infections-Hospital/77hc-ibv8))

   The third one, Infections table, includes infection information(Measure_ID, Provider_ID, Measure_Name, Measure_Type, Measure_Start_Date, Measure_END_Date, Score).

4. Complications and Deaths - Hospital (https://data.medicare.gov/Hospital-Compare/Complications-and-Deaths-Hospital/ynj2-r877 (https://data.medicare.gov/Hospital-Compare/Complications-and-Deaths-Hospital/ynj2-r877))

   The fourth one will be Complications_and_Deaths table which includes death information(Measure_ID, Provider_ID, Measure_Name, Measure_Type, Measure_Start_Date, Measure_END_Date, Score).

5. Long- Term Care Hospital - General Information(https://data.medicare.gov/Long-Term-Care-Hospital-Compare/Long-Term-Care-Hospital-General-Information/azum-44iv (https://data.medicare.gov/Long-Term-Care-Hospital-Compare/Long-Term-Care-Hospital-General-Information/azum-44iv))

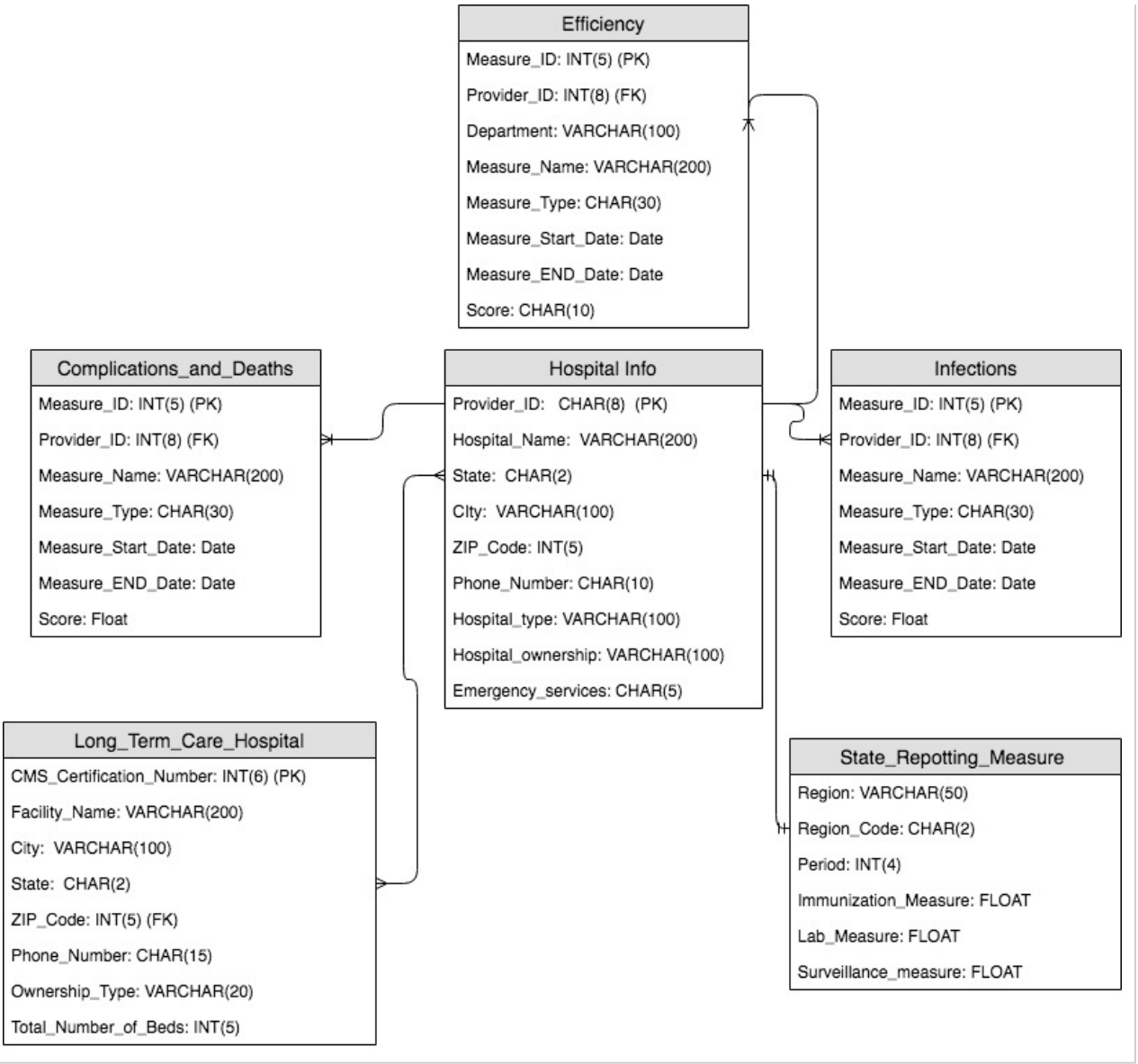
   The fifth one will be Long- Term Care Hospital table which includes long-term care information(CMS_Certification, facility_name, city, State, zipcode, phone number, ownership, and total number of beds).

6. Hospital Selection of Public Health Measures by State(https://dashboard.healthit.gov/datadashboard/documentation/hospital-public-health-reporting-documentation.php (https://dashboard.healthit.gov/datadashboard/documentation/hospital-public-health-reporting-documentation.php))

   The fifth table is State_Repotting_Measure table, which includes region information(Region, Region_Code, Period, Immunization_Measure, Lab_Measure, Surveillance_measure).

Data Model

ERD



Basically, we have one main table called Hospital Information including some basic hospital information such as phone number, location, ownership. We use provider ID as the primary key in this table bacause provider ID is a unique identifier for all the medical centers across the united state. The three measure aspect tables which are Efficiency, Infections ,Complications and Deaths use the provider ID to connect to the hospital information table. So the provider ID is both foreigh key and primary key among these 4 tables. In each aspect, we have different measure types(different measure types have different range score) to evaluate the hospital.

The long term care hospital is a kind of hospital that focuses on the patients who stay more than 25 days. This kind of hospital generally give services like respiratory,therapy,head trauma treatment and pain management. CMS Certification number as a unique identifier in this table and ZIP code as a foreigh key connect to the hospital information table.

State repotting measure is the a table includes the percentage of hospitals who reported on these measures in program years, 2013, 2014 and 2015. More specifically, Immunization_measure is the percent of Eligible Hospitals that Reported on Immunization Measure. Reportable lab results measure is the percent of Eligible Hospitals that Reported Reportable Lab Results Measure. Syndromic_surveillance_measure is the percent of Eligible Hospitals that Reported Syndromic Surveillance Measure.

Set Environment

In [2]:
```
%load_ext sql
%sql mysql+pymysql://root:@fe512_mysql/fe512db
%sql SHOW DATABASES;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
8 rows affected.
```

Out[2]:

Database
Assignments
fe512db
information_schema
mysql
performance_schema
project
sys
world

In [3]:
```
%sql USE project
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
0 rows affected.
```

Out[3]:
```
[]
```

In [4]:
```
%sql SELECT DATABASE();
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
1 rows affected.
```

Out[4]:

DATABASE()
project

Create Tables

Hospital_Information Table

In [8]:
```
%%sql
CREATE TABLE IF NOT EXISTS Hospital_Information (
    Provider_ID CHAR(8) UNIQUE NOT NULL primary key,
    Hospital_Name VARCHAR(200),
    City VARCHAR(100),
    State CHAR(3),
    Zip_Code INT(5),
    Phone_Number CHAR(10),
    Hospital_type VARCHAR(100),
    Hospital_ownership VARCHAR(100),
    Emergency_services CHAR(5)
);
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
0 rows affected.
```

Out[8]:
```
[]
```

In [11]:
```
%%sql
LOAD DATA INFILE '/home/data/Hospital__Information.csv' INTO TABLE Hospital
FIELDS TERMINATED BY ','
ENCLOSED BY ''
ESCAPED BY ''
LINES TERMINATED BY '\n'
IGNORE 1 LINES
;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
4784 rows affected.
```

Out[11]:
```
[]
```

Efficiency Table

In [15]:
```
%%sql
CREATE TABLE IF NOT EXISTS Efficiency (
    Measure_ID INT(5) UNIQUE NOT NULL primary key,
    Provider_ID INT(8),
    Department VARCHAR(100),
    Measure_Name VARCHAR(200),
    Measure_Type CHAR(30),
    Measure_Start_Date VARCHAR(20),
    Measure_END_Date VARCHAR(20),
    Score CHAR(10)
);
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
0 rows affected.
```

Out[15]:
```
[]
```

In [16]:
```
%%sql
LOAD DATA INFILE '/home/data/Efficiency.csv' INTO TABLE Efficiency
FIELDS TERMINATED BY ','
ENCLOSED BY ''
ESCAPED BY ''
LINES TERMINATED BY '\r\n'
IGNORE 1 LINES
;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
34560 rows affected.
```

Out[16]: []

Infections Table

In [19]:
```
%%sql
CREATE TABLE IF NOT EXISTS Infections (
    Measure_ID INT(5) UNIQUE NOT NULL primary key,
    Provider_ID INT(8),
    Measure_Name VARCHAR(200),
    Measure_Type CHAR(30),
    Measure_Start_Date VARCHAR(20),
    Measure_END_Date VARCHAR(20),
    Score Float
);
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
0 rows affected.
```

Out[19]: []

In [20]:
```
%%sql
LOAD DATA INFILE '/home/data/Infections.csv' INTO TABLE Infections
FIELDS TERMINATED BY ','
ENCLOSED BY ''
ESCAPED BY ''
LINES TERMINATED BY '\r\n'
IGNORE 1 LINES
;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
73674 rows affected.
```

Out[20]: []

Complications_and_Deaths Table

In [23]:
```
%%sql
CREATE TABLE IF NOT EXISTS Complications_and_Deaths (
    Measure_ID INT(5) UNIQUE NOT NULL primary key,
    Provider_ID CHAR(8),
    Measure_Name VARCHAR(200),
    Measure_Type CHAR(30),
    Score Float,
    Measure_Start_Date VARCHAR(20),
    Measure_END_Date VARCHAR(20)
);
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
0 rows affected.
```

Out[23]: []

In [24]:
```
%%sql
LOAD DATA INFILE '/home/data/Complications_and_Deaths.csv' INTO TABLE Compl
FIELDS TERMINATED BY ','
ENCLOSED BY ''
ESCAPED BY ''
LINES TERMINATED BY '\r\n'
IGNORE 1 LINES
;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
39362 rows affected.
```

Out[24]: []

Long_Term_Care_Hospital Table

In [33]:
```
%%sql
CREATE TABLE IF NOT EXISTS Long_Term_Care_Hospital (
    CMS_Certification_Number INT(6) UNIQUE NOT NULL primary key,
    Facility_Name VARCHAR(200),
    City VARCHAR(100),
    State CHAR(2),
    Zip_Code INT(5),
    Phone_Number CHAR(20),
    Ownership_type VARCHAR(20),
    Total_Number_of_Beds INT(5)
);
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
0 rows affected.
```

Out[33]: []

In [34]:
```
%%sql
LOAD DATA INFILE '/home/data/Long_Term_Care_Hospital.csv' INTO TABLE Long_T
FIELDS TERMINATED BY ','
ENCLOSED BY ''
ESCAPED BY ''
LINES TERMINATED BY '\n'
IGNORE 1 LINES
;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
388 rows affected.
```

Out[34]: []

State_Repotting_Measure Table

In [56]:
```
%%sql
CREATE TABLE IF NOT EXISTS State_Repotting_Measure (
    Region VARCHAR(50),
    Region_Code CHAR(2),
    Period INT(4),
    Immunization_Measure FLOAT,
    Lab_Measure FLOAT,
    Surveillance_measure FLOAT
);
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
0 rows affected.
```

Out[56]: []

In [59]:
```
%%sql
LOAD DATA INFILE '/home/data/public-health-measures.csv' INTO TABLE State_R
FIELDS TERMINATED BY ','
ENCLOSED BY ''
ESCAPED BY ''
LINES TERMINATED BY '\r\n'
IGNORE 1 LINES
;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
156 rows affected.
```

Out[59]: []

Data summary

In [20]:
```
%%sql
#how many hospital and Long_Term_Care_Hospital in total in this database
SELECT COUNT(distinct l.Facility_Name) AS Long_Term_Care_Hospital,COUNT(dis
FROM Long_Term_Care_Hospital AS l inner join Hospital_Information AS h
on l.State = h. state;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
1 rows affected.
```

Out[20]:

Long_Term_Care_Hospital	Hospital
388	4624

There are 388 long term care hospital and 4624 hospital in this database in total

In [18]:
```
%%sql
#see the all hospital type
SELECT DISTINCT Hospital_Information.Hospital_type
FROM Hospital_Information;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
3 rows affected.
```

Out[18]:

Hospital_type
Acute Care Hospitals
Critical Access Hospitals
Childrens

Result Explanation: There are 3 types of hospital which are Acute Care Hospitals,Critical Access Hospitals and Childrens. Acute Care Hospitals (ACH) are hospitals that provide short-term patient care. Critical Access Hospitals (CAH) are small facilities that give limited outpatient and inpatient hospital services to people in rural areas that receive cost-based reimbursement. Childrens Hospitals are hospitals that focus on the children care.

In [19]:
```
%%sql
#see all hospital ownership
SELECT DISTINCT Hospital_Information.Hospital_ownership
FROM Hospital_Information;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
11 rows affected.
```

Out[19]:

Hospital_ownership
Voluntary non-profit - Private
Government - Hospital District or Authority
Voluntary non-profit - Church
Proprietary
Government - State
Voluntary non-profit - Other
Government - Local
Government - Federal
Physician
Tribal
Government

Result Explanation: As is shown above, there are so many ownership types. In general, the ownership can be classified as Government,Voluntary non-profit,Proprietary,Physician,Tribal. You can see the some subclassification under the government and voluntary non-profit.

In [5]:
```
%%sql
#see the number of the hospitals that provide the emergency service in this
SELECT count(Provider_ID)
FROM Hospital_Information
WHERE Emergency_services="TRUE\r";
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
1 rows affected.
```

Out[5]:

count(Provider_ID)
4398

In [70]:
```
%%sql
# Ownership in Long_Term_Care_Hospital
SELECT distinct Ownership_Type
FROM Long_Term_Care_Hospital;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
5 rows affected.
```

Out[70]:

Ownership_Type
Non-profit
Government
For profit
Tribal
Physician

Data Summary

- 4784 records in Hospital__Information table
- 34560 records in Efficiency table
- 73674 records in Infections table
- 39362 Records in Complications_and_Deaths Table
- 388 Records in Long_Term_Care_Hospital Table
- 156 records in State_Repotting_Measure table

Hospital__Information table

- 388 Long_Term_Care_Hospitals and 4624 Hospitals in total
- 3 different hospital types in Hospital_Info table: Acute Care Hospitals, Critical Access Hospitals, Childrens.
- 11 different Hospital_ownership types : Voluntary non-profit - Private, Voluntary non-profit - Church, Voluntary non-profit - Other, Government, Government - Hospital District or Authority, Government - State, Government - Local, Government - Federal, Proprietary, Physician, Tribal
- 4398 Hospitals provide emergency service.

5 different Ownership Types from Long_Term_Care_Hospital table : Non-profit, Government. For profit, Tribal, Physician

In [10]:
```
%%sql
#to understand the Measure Type and Measure Name in the Efficiency table
SELECT distinct Measure_Name ,Measure_Type
FROM Efficiency;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
23 rows affected.
```

Out[10]:

Measure_Name	Measure_Type
Appropriate care for severe sepsis and septic shock	SEP_1
OP 18	OP_18b
Left before being seen	OP_22
Emergency department volume	EDV
Aspirin at Arrival	OP_4
ED1	ED_1b
Immunization for influenza	IMM_2
OP-18	OP_18c
Endoscopy/polyp surveillance: colonoscopy interval for patients with a history of adenomatous polyps - avoidance of inappropriate use	OP_30
Median Time to ECG	OP_5
ED2	ED_2b
External Beam Radiotherapy for Bone Metastases	OP_33
Door to diagnostic eval	OP_20
Head CT results	OP_23
Healthcare workers given influenza vaccination	IMM_3_OP_27_FAC_ADHPCT
Endoscopy/polyp surveillance: appropriate follow-up interval for normal colonoscopy in average risk patients	OP_29
Median Time to Transfer to Another Facility for Acute Coronary Intervention	OP_3b
Median time to pain med	OP_21
Elective Delivery	PC_01
Hospital Acquired Potentially-Preventable Venous Thromboembolism	VTE_6
Median Time to Fibrinolysis	OP_1
Fibrinolytic Therapy Received Within 30 Minutes of ED Arrival	OP_2
Improvement in Patient's Visual Function within 90 Days Following Cataract Surgery	OP_31

In [12]:
```
%%sql
#see the range of score for each measure type in Efficiency table
SELECT Measure_Type,min(Score),max(Score),avg(Score)
FROM Efficiency
WHERE Measure_Type <> "EDV"
GROUP BY Measure_Type;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
22 rows affected.
```

Out[12]:

Measure_Type	min(Score)	max(Score)	avg(Score)
SEP_1	1	98	50.55382907880133
OP_18b	100	99	141.65620857269113
OP_22	1	9	2.068510370835952
OP_4	100	99	94.67630057803468
ED_1b	100	997	275.48396624472576
IMM_2	10	99	90.78814577492992
OP_18c	100	988	253.64973102211596
OP_30	100	99	91.2838669950739
OP_5	1	9	8.467015022860876
ED_2b	1	99	101.76327054794521
OP_33	10	99	86.27892561983471
OP_20	1	92	22.28846153846154
OP_23	100	98	73.7303867403315
IMM_3_OP_27_FAC_ADHPCT	100	99	86.98057097541634
OP_29	1	99	87.62111043319096
OP_3b	102	98	61.71228070175439
OP_21	101	98	49.79444967074318
PC_01	1	9	4.116138763197586
VTE_6	1	9	7.271739130434782
OP_1	15	49	25.86
OP_2	100	96	71.08163265306122
OP_31	100	99	93.96875

In [9]:
```
%%sql
#see the range of score for each measure type in Infections table
SELECT Measure_Type,min(Score),max(Score),avg(Score)
FROM Infections
GROUP BY Measure_Type;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
36 rows affected.
```

Out[9]:

Measure_Type	min(Score)	max(Score)	avg(Score)
HAI-6-CI-LOWER	0.002	2.15	0.4506413232185509
HAI-1-DOPC-DAYS	1.0	70456.0	3164.0673773987205
HAI-2-CI-LOWER	0.003	2.64	0.41206588387794685
HAI-2-DOPC-DAYS	1.0	60007.0	3475.1052855924977
HAI-2-ELIGCASES	0.001	90.737	3.1928990873242626
HAI-2-CI-UPPER	0.118	9.581	2.0543942036548097
HAI-1-NUMERATOR	1.0	144.0	7.4688946015424165
HAI-6-DOPC-DAYS	1.0	482438.0	29406.846242111304
HAI-1-SIR	0.069	7.765	0.8907352077049614
HAI-3-ELIGCASES	0.015	44.914	1.9457409766576204
HAI-5-ELIGCASES	0.001	41.951	1.5299446422851668
HAI-5-DOPC-DAYS	1.0	500388.0	33359.30901542112
HAI-3-SIR	0.12	4.766	1.0381908907747797
HAI-6-SIR	0.04	5.305	0.8310513819068206
HAI-3-DOPC-DAYS	1.0	1520.0	95.5830639948287
HAI-4-DOPC-DAYS	1.0	1931.0	90.05406348099058
HAI-6-CI-UPPER	0.069	11.034	1.534863853737093
HAI-1-CI-UPPER	0.308	14.25	1.955648198236719
HAI-6-NUMERATOR	1.0	487.0	25.64904163912756
HAI-4-ELIGCASES	0.003	15.381	0.6950655626390781
HAI-1-ELIGCASES	0.001	77.355	2.7748133303731697
HAI-3-CI-LOWER	0.006	2.032	0.3487091870862752
HAI-1-CI-LOWER	0.003	3.787	0.34660324866028847
HAI-3-NUMERATOR	1.0	47.0	4.271631982475356
HAI-5-CI-LOWER	0.003	2.847	0.3397699237258347
HAI-4-NUMERATOR	1.0	19.0	2.185375901132853
HAI-6-ELIGCASES	0.001	366.037	14.979351640392931
HAI-2-SIR	0.055	3.972	0.9739355855754444
HAI-3-CI-UPPER	0.316	10.565	2.384044783996658
HAI-5-CI-UPPER	0.313	11.641	2.3534107620977576

HAI-2-NUMERATOR	1.0	111.0	8.16231505657093
HAI-5-SIR	0.066	6.13	1.0158893557875848
HAI-4-CI-UPPER	0.316	8.266	2.8978013455466387
HAI-4-CI-LOWER	0.008	2.525	0.31355172333843373
HAI-5-NUMERATOR	1.0	58.0	4.531759415401911
HAI-4-SIR	0.168	4.89	1.143876172979598

In [13]:
```
%%sql
#to understand the Measure Type and Measure Name in the Complications_and_D
SELECT  distinct Measure_Name ,Measure_Type
FROM Complications_and_Deaths;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
19 rows affected.
```

Out[13]:

Measure_Name	Measure_Type
Pressure sores	PSI_3_ULCER
Death rate for stroke patients	MORT_30_STK
Death rate for pneumonia patients	MORT_30_PN
Death rate for COPD patients	MORT_30_COPD
A wound that splits open after surgery on the abdomen or pelvis	PSI_14_POSTOP_DEHIS
Serious complications	PSI_90_SAFETY
Accidental cuts and tears from medical treatment	PSI_15_ACC_LAC
Serious blood clots after surgery	PSI_12_POSTOP_PULMEMB_DVT
Broken hip from a fall after surgery	PSI_8_POST_HIP
Deaths among Patients with Serious Treatable Complications after Surgery	PSI_4_SURG_COMP
Collapsed lung due to medical treatment	PSI_6_IAT_PTX
Postoperative Acute Kidney Injury Requiring Dialysis Rate	PSI_10_POST_KIDNEY
Death rate for heart failure patients	MORT_30_HF
Death rate for CABG surgery patients	MORT_30_CABG
Blood stream infection after surgery	PSI_13_POST_SEPSIS
Death rate for heart attack patients	MORT_30_AMI
Perioperative Hemorrhage or Hematoma Rate	PSI_9_POST_HEM
Rate of complications for hip/knee replacement patients	COMP_HIP_KNEE
Postoperative Respiratory Failure Rate	PSI_11_POST_RESP

In [14]:
```
%%sql
#see the range of score for each measure type in Complications_and_Deaths t
SELECT Measure_Type,min(Score),max(Score),avg(Score)
FROM Complications_and_Deaths
GROUP BY Measure_Type;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
19 rows affected.
```

Out[14]:

Measure_Type	min(Score)	max(Score)	avg(Score)
PSI_3_ULCER	0.02	12.6	0.3920983901903403
MORT_30_STK	8.9	21.4	14.238440862265966
MORT_30_PN	9.0	24.8	15.783784675787116
MORT_30_COPD	4.9	14.4	8.362775430284875
PSI_14_POSTOP_DEHIS	0.27	2.7	0.8564561380180798
PSI_90_SAFETY	0.52	4.21	1.0006571554591324
PSI_15_ACC_LAC	0.57	2.98	1.2916056334692547
PSI_12_POSTOP_PULMEMB_DVT	1.39	9.82	3.843696463041465
PSI_8_POST_HIP	0.08	0.17	0.11084930258657043
PSI_4_SURG_COMP	96.82	225.89	161.467851400752
PSI_6_IAT_PTX	0.11	0.68	0.2868233251379717
PSI_10_POST_KIDNEY	0.33	4.31	1.3086989302407956
MORT_30_HF	5.0	18.0	11.73790899350617
MORT_30_CABG	1.3	8.5	3.2063535904686753
PSI_13_POST_SEPSIS	2.09	14.45	5.248171660426669
MORT_30_AMI	8.9	18.7	13.12633390804166
PSI_9_POST_HEM	1.29	4.69	2.588323989724801
COMP_HIP_KNEE	1.2	5.4	2.5959162791204915
PSI_11_POST_RESP	1.55	34.75	8.096885491509473

Question and Answers

1.How many hospitals are not owned by the government in this database

- 3718

In [23]:
```
%%sql

SELECT count(Provider_ID)
FROM Hospital_Information
WHERE Provider_ID NOT IN (SELECT Provider_ID
                          FROM Hospital_Information
                          WHERE Hospital_ownership like "%govern%" );
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
1 rows affected.
```

Out[23]:

count(Provider_ID)
3718

2.Find the top 3 states with the most Acute Care hospitals

- TX, CA, FL

In [26]:
```
%%sql

SELECT State,count(Provider_ID)
FROM Hospital_Information
WHERE Hospital_Information.Hospital_type = 'Acute Care Hospitals'
GROUP BY State
ORDER BY count(Provider_ID) DESC
LIMIT 3;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
3 rows affected.
```

Out[26]:

State	count(Provider_ID)
TX	305
CA	294
FL	168

3.Find the number of the hospitals for the different ownerships in California

In [27]:
```
%%sql
SELECT Hospital_ownership,count(Provider_ID)
FROM Hospital_Information
WHERE State = "CA"
GROUP BY Hospital_ownership;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
9 rows affected.
```

Out[27]:

Hospital_ownership	count(Provider_ID)
Voluntary non-profit - Church	30
Voluntary non-profit - Private	105
Voluntary non-profit - Other	60
Proprietary	75
Government - Hospital District or Authority	39
Government - Local	18
Government - Federal	4
Government - State	3
Physician	5

4.Which city have most and least hospital with emergency service

- Most: Chicago (26)
- Least: Rawlins (1)

In [29]:
```
%%sql
(SELECT City,count(Provider_ID) AS count
FROM Hospital_Information
WHERE Emergency_services = 'TRUE\r'
GROUP BY City
ORDER BY count(Provider_ID) DESC
LIMIT 1)
UNION
(SELECT City,count(Provider_ID) AS count
FROM Hospital_Information
WHERE Emergency_services = 'TRUE\r'
GROUP BY City
ORDER BY count(Provider_ID)
LIMIT 1);
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
2 rows affected.
```

Out[29]:

City	count
CHICAGO	26
RAWLINS	1

5.Which 5 hospitals have the least efficiency in Aspirin at Arrival ?

In [36]:
```
%%sql

SELECT avg(Efficiency.Score) AS score,Hospital_Information.Hospital_Name
FROM Hospital_Information INNER JOIN Efficiency
ON Hospital_Information.Provider_ID = Efficiency.Provider_ID
WHERE Efficiency.Measure_Type = 'OP_4'
GROUP BY Efficiency.Provider_ID
ORDER BY avg(Efficiency.Score)
LIMIT 5;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
5 rows affected.
```

Out[36]:

score	Hospital_Name
48.0	BAKERSFIELD HEART HOSPITAL
50.0	ABRAZO SCOTTSDALE CAMPUS
50.0	MEMORIAL REGIONAL HOSPITAL
53.0	AURORA ST LUKES MEDICAL CENTER
53.0	WOMEN & INFANTS HOSPITAL OF RHODE ISLAND

6.Which 5 states have highest average score in Aspirin at Arrival among their hospital

In [35]:
```
%%sql

SELECT Region,avg(Score)
FROM Efficiency
INNER JOIN Hospital_Information
ON Efficiency.Provider_ID = Hospital_Information.Provider_ID
INNER JOIN State_Repotting_Measure
ON State_Repotting_Measure.Region_Code = Hospital_Information.State
WHERE Measure_Type = 'OP_4'
GROUP BY Region
ORDER BY avg(Score) DESC
LIMIT 5;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
5 rows affected.
```

Out[35]:

Region	avg(Score)
Hawaii	100.0
District Of Columbia	100.0
Vermont	100.0
Wyoming	99.5
Maine	98.33333333333333

7.Which state has the highest total measure score

- Indiana

In [33]:
```
%%sql
ALTER table c
ADD column Total_measure float;
UPDATE State_Repotting_Measure
SET Total_measure = Immunization_Measure + Lab_Measure + Surveillance_measu
SELECT avg(Total_measure),Region
FROM State_Repotting_Measure
GROUP BY Region
ORDER BY avg(Total_measure) desc
LIMIT 1;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
156 rows affected.
1 rows affected.
```

Out[33]:

avg(Total_measure)	Region
2.096666693687439	Indiana

8.Which states of the lab measure go down from 2013 to 2014

- Connecticut, District Of Columbia, Montana

In [32]:
```
%%sql

SELECT l.Region
FROM (SELECT Region,Lab_Measure AS lm2014
        FROM State_Repotting_Measure
        WHERE period=2014) AS l
INNER JOIN
        (SELECT Lab_Measure AS lm2013,Region
        FROM State_Repotting_Measure
        WHERE period=2013) AS j
ON l.Region=j.Region
WHERE lm2014 < lm2013;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
3 rows affected.
```

Out[32]:

Region
Connecticut
District Of Columbia
Montana

9.which 3 hospitals have the most balanced performance in efficient

In [31]:
```
%%sql
SELECT Hospital_Name,VAR_SAMP(score) AS variance,Efficiency.Provider_ID
FROM Efficiency INNER JOIN Hospital_Information ON Efficiency.Provider_ID=H
GROUP BY Efficiency.Provider_ID
ORDER BY VAR_SAMP(score) DESC
LIMIT 3;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
3 rows affected.
```

Out[31]:

Hospital_Name	variance	Provider_ID
MAYAGUEZ MEDICAL CENTER DR RAMON EMETERIO BETANCES	506745.3333333333	400103
HOSPITAL SAN FRANCISCO	502002.0	400098
HOSPITAL UPR DR FEDERICO TRILLA	302968.9166666666	400112

10.What's the score of all Surgical Site Infections for hospital with the lowest Death rate for stroke patients

In [30]:
```
%%sql
SELECT Measure_Name,Score
FROM Infections
WHERE Provider_ID = (SELECT Provider_ID
                        FROM Complications_and_Deaths
                        WHERE Measure_Type= "MORT_30_STK"
                        ORDER BY score asc
                        LIMIT 1)
        AND Measure_Name LIKE "SSI%";
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
12 rows affected.
```

Out[30]:

Measure_Name	Score
SSI - Colon Surgery: Upper Confidence Limit	3.444
SSI - Colon Surgery: Predicted Cases	6.658
SSI - Colon Surgery: Number of Procedures	260.0
SSI - Colon Surgery: Lower Confidence Limit	1.197
SSI - Abdominal Hysterectomy: Observed Cases	3.0
SSI - Abdominal Hysterectomy	1.327
SSI - Colon Surgery	2.103
SSI - Abdominal Hysterectomy: Predicted Cases	2.26
SSI - Colon Surgery: Observed Cases	14.0
SSI - Abdominal Hysterectomy: Upper Confidence Limit	3.613
SSI - Abdominal Hysterectomy: Number of Procedures	278.0
SSI - Abdominal Hysterectomy: Lower Confidence Limit	0.338

Result and Conclusion

Calculate the average head CT result efficiency by state across the USA

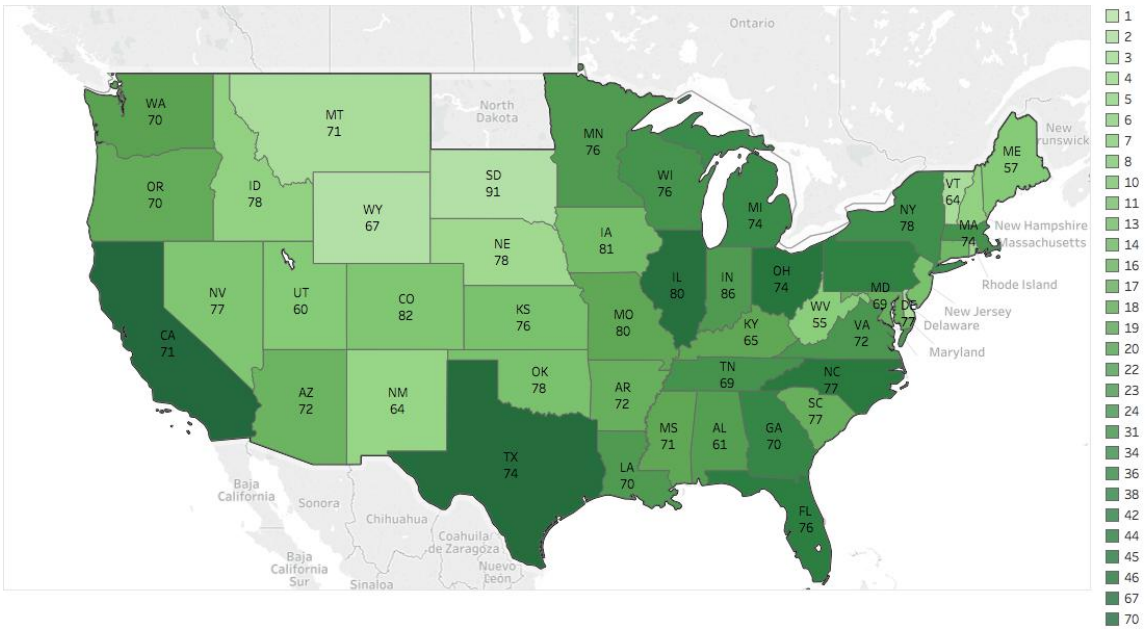
In [5]:
```
%%sql
SELECT State,avg(Score),count(Score)
FROM Efficiency AS e inner join Hospital_Information AS hi  on e.Provider_I
WHERE Measure_Type="OP_23"
GROUP BY State;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
48 rows affected.
```

Out[5]:

State	avg(Score)	count(Score)
FL	75.95238095238095	42
AL	61.25	20
GA	69.77777777777777	36
ID	78.0	5
IL	80.04347826086956	46
IN	85.9090909090909	22
IA	80.9090909090909	11
KS	75.9	10
KY	65.16666666666667	18
LA	70.36363636363636	22
ME	57.285714285714285	7
MD	68.6875	16
AK	27.0	1
MA	73.625	24
MI	73.97058823529412	34
MN	76.18181818181819	22
MS	70.83333333333333	18
MO	79.83333333333333	18
MT	71.0	2
NE	78.33333333333333	3
NV	76.75	8
NH	71.8	5
AZ	71.5	14
NJ	72.3	10
NM	64.0	4
NY	78.12903225806451	31
NC	76.81818181818181	44
OH	74.17777777777778	45
OK	78.4	10
OR	69.58823529411765	17

PA	82.52631578947368	38
AR	71.875	16
RI	54.666666666666664	3
SC	76.5	16
SD	91.0	1
TN	69.33333333333333	24
TX	74.34328358208955	67
UT	60.142857142857146	7
VT	64.0	2
VA	72.43478260869566	23
WA	69.6842105263158	19
CA	71.15714285714286	70
WV	55.166666666666664	6
WI	75.66666666666667	24
WY	67.0	1
CO	81.875	8
CT	67.76923076923077	13
DE	76.5	2



Explanation: The color shows the number of the measured hospitals in the state. The darker color,The larger number of the measured hospitals. The number shown in each state is the average score of the Head CT efficiency. We can see that the hospitals in Colorado, Missouri and Illinois have the highest head CT efficiency scores. In contrast, the efficiency scores in Utah and Maine are lowest.

Calculate the average serious comlications score by state across the USA

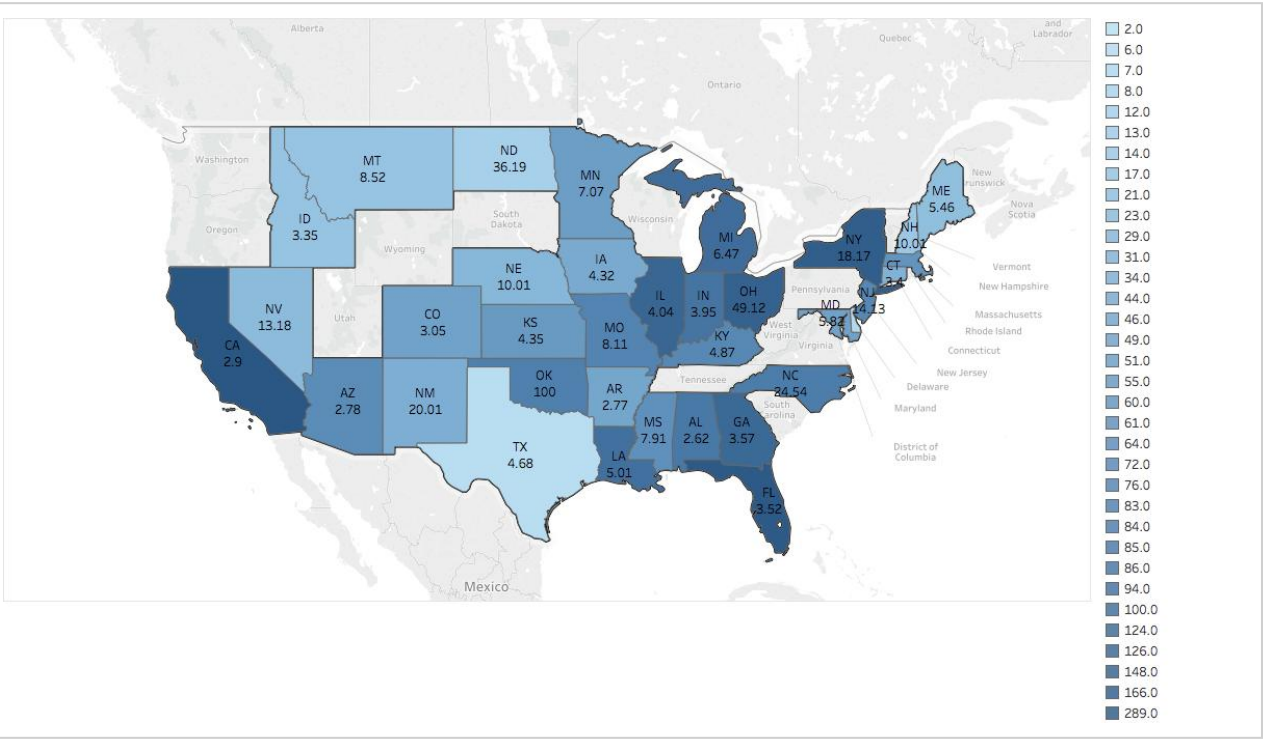
In [6]:
```
%%sql

SELECT State,avg(Score),count(Score)
FROM Complications_and_Deaths AS c inner join Hospital_Information AS hi  c
WHERE Measure_Type="PSI_90_SAFETY"
GROUP BY State;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
38 rows affected.
```

Out[6]:

State	avg(Score)	count(Score)
AL	0.9945238154558909	84
AK	0.9825000017881393	8
AZ	1.0004918037867936	61
AR	0.9715909145095132	44
CA	0.9886505174801836	289
CO	1.0089130427526392	46
CT	1.0906896570633198	29
DE	1.0433333118756611	6
DC	1.1785714541162764	7
FL	1.0101807228772037	166
GA	0.9884000009298325	100
HI	0.9141666740179062	12
ID	0.8650000052792686	14
IL	1.0081451626554612	124
IN	0.9458823505569907	85
IA	0.9938235282897949	34
KS	0.9574509859085083	51
KY	1.0246875016018748	64
LA	1.0027906970922338	86
TX	0.8900000154972076	2
ME	0.9888235330581665	17
MD	0.9969565220501112	46
MA	0.9707272681322965	55
MI	0.9796808514189213	94
MN	1.001428567633337	49
MS	1.041000000635783	60
MO	0.9837499981125196	72
MT	0.9485714307853154	14

NE	1.0200000068415767	23
NH	0.9176923082425044	13
NV	1.0871428506714957	21
NJ	1.01218749769032	64
NY	1.117297299407624	148
NM	1.0064516163641406	31
NC	0.9874698778232897	83
ND	1.098750002682209	8
OH	0.9513492030756814	126
OK	0.9853947397909666	76



Explanation: This graph shows the serious comlicatinos score in each state. The higher socre, the less chance patients get serious complicaitons. The color shows the number of the records in the state. We also can know the number of the hospitals in each states that were measured from the right side of the graph. After we standardized the value, the number shown in each state is the average score of the serious comlicatinos. It can be seen from the graph that the hospitals in Ohio and Oklahoma have high serious comlications scores.

Calculate the average pressure sores score by state across the USA

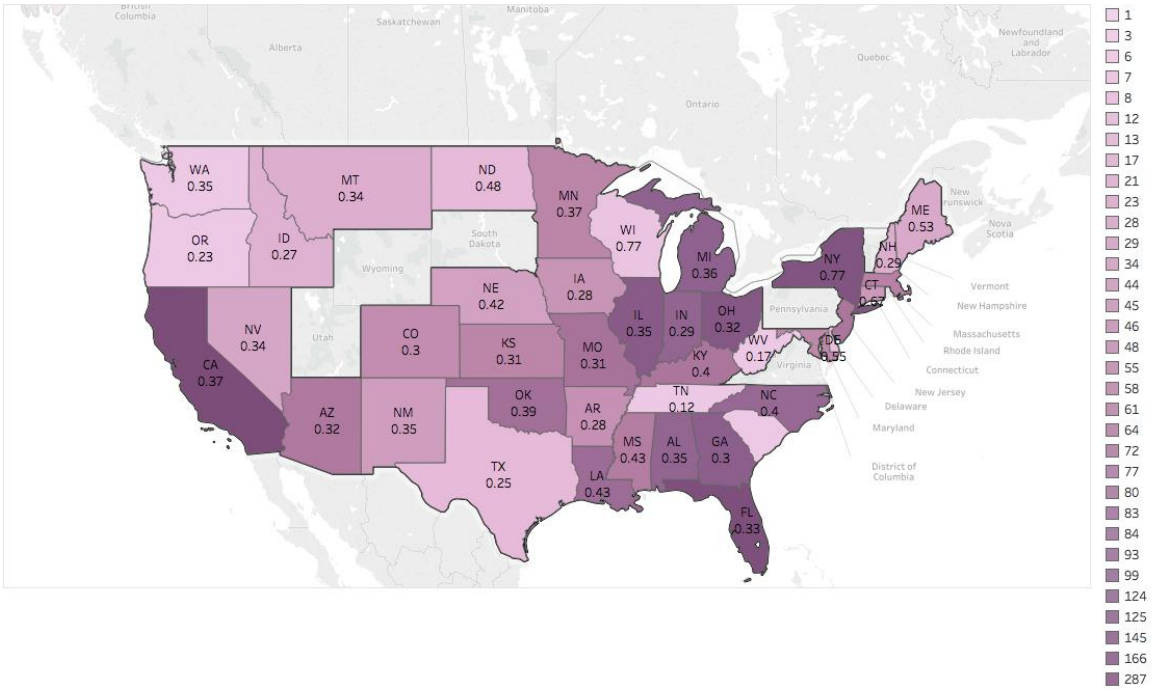
In [7]:
```
%%sql
SELECT State,avg(Score),count(Score)
FROM Complications_and_Deaths AS c inner join Hospital_Information AS hi  c
WHERE Measure_Type="PSI_3_ULCER"
GROUP BY State;
```

```
 * mysql+pymysql://root:***@fe512_mysql/fe512db
44 rows affected.
```

Out[7]:

State	avg(Score)	count(Score)
CA	0.3707317073183712	287
CO	0.30355555613835655	45
FL	0.332831327298499	166
IL	0.35387096914552874	124
IN	0.286666666689728	84
IA	0.2800000025616849	34
KY	0.3968750024214387	64
LA	0.43150000274181366	80
MI	0.36096774229920037	93
MS	0.43379310467119875	58
MO	0.30930555554934674	72
NH	0.28923076591812646	13
NY	0.7735862080650083	145
NC	0.3979518102504403	83
OH	0.32168000042438505	125
TN	0.11999999731779099	1
TX	0.2528571401323591	7
WA	0.3499999940395355	1
WV	0.17000000178813934	1
WI	0.773333340883255	3
AL	0.3460714297280425	84
AK	0.40375000424683094	8
AZ	0.3155737702475219	61
AR	0.2793181815861978	44
CT	0.6686206896757257	29
DC	0.6385714422379222	7
DE	0.550000011920929	6
GA	0.3014141401708728	99
HI	0.31083333243926364	12
ID	0.26916666453083354	12

KS	0.31437500042375177	48
ME	0.5299999998772845	17
MD	0.6215217464967914	46
MA	0.40654545676979154	55
MN	0.3689583327310781	48
SC	0.20000000298023224	1
MT	0.3438461577663055	13
NE	0.41913043966759805	23
NV	0.342380948896919	21
NJ	0.4329687522840686	64
NM	0.35499999672174454	28
ND	0.48285713366099764	7
OK	0.3859740271196737	77
OR	0.23000000417232513	1



Explanation: The color shows the number of the records in the state. If the color is darker, the number of the records are larger. The number shown in each state is the average score of the pressure sores. Pressure scores is the one of the complications after the surgery, you can see that the New York State, Connecticut and Wisconsin have the highest score while the scores in Oregon and Texas are the lowest.

Future Study

1. In this project, we only have a few aspects(Effciency, Infection, Complications_and_Deaths).

For the future work, we can add more aspects such as environment and infrastructure of the hospital. We also can add some insurance information ,so patients can make different decision choice for different situtaion.

2. The measure types for each aspect in this database are too professional and it is hard to understand from their name. We can create a reference book in the future to explain each measure tpye, so people can have a better understanding of these evaluation.
3. We collected 4000 medical centers information in this database, but we want to include as many hospitals as possible.