Group Project Presentation

DSBA 6160: Big Data Design, Storage, and Provenance

Alpha Group 1: Tony Bejos, Josh Peterson, Michael Lewis and Kevin Ovendorf

Part I

Introduction

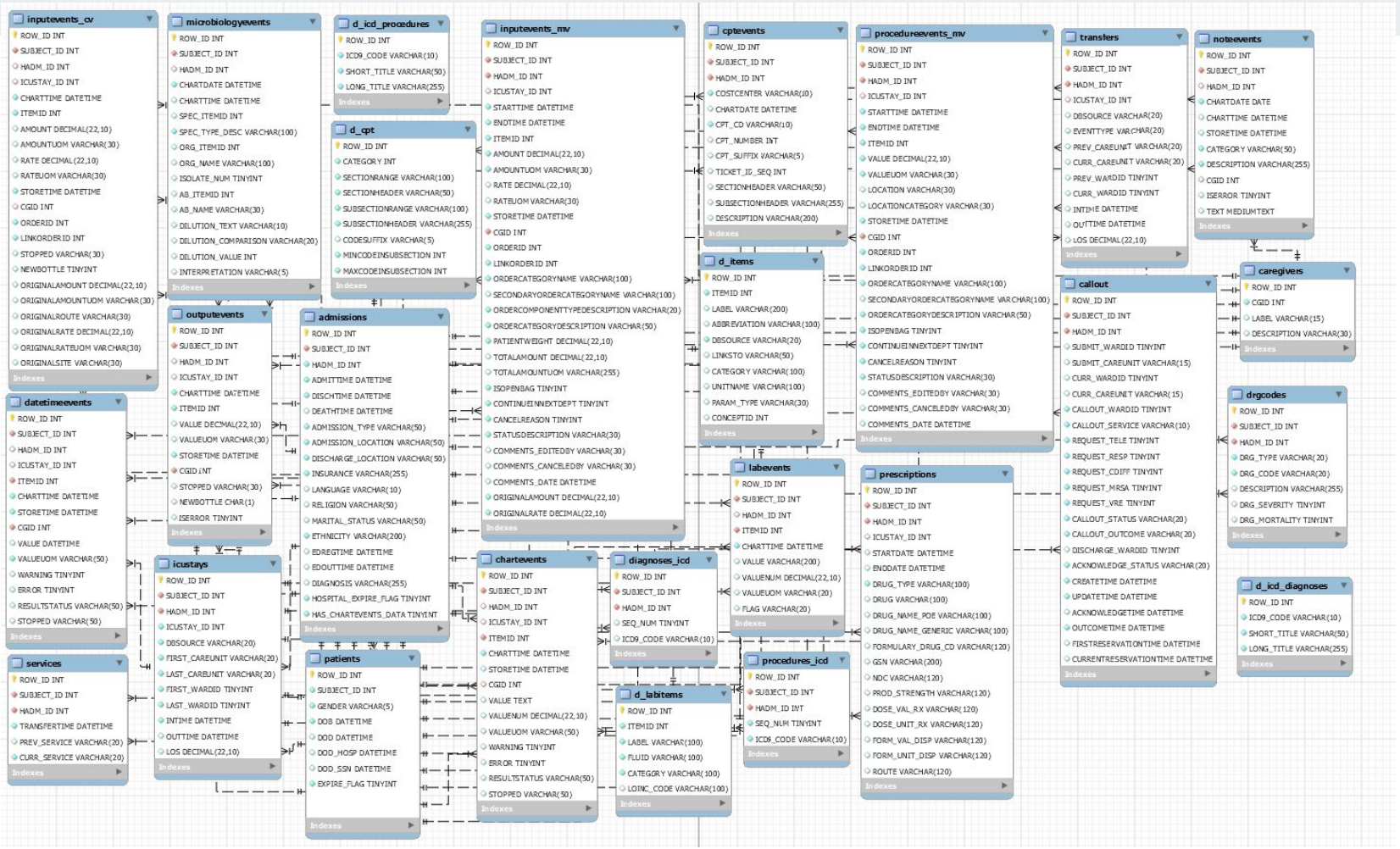
Accessing and Loading MIMIC III Database

- Attained Training Certification through CITI online training program to gain access to MIMIC III database
  - 90% Score Required
  - Learned about human research practices and regulations
  - Patient Privacy remains essential
- Prior to loading the tables into MySQL, each table reviewed and discussed by the team
  - 27 Tables
  - Millions of Rows
  - Entity vs Assosciation

MIMIC III Database Build Overview

- Created a MYSQL db named *mimic_iii_project*
- Created table schema prior to load
- Extracted all zip files and utilized LOAD DATA LOCAL IN FILE code to load
- Created indexes on all tables for ease of use

```
LOAD DATA LOCAL INFILE 'ADMISSIONS.csv' INTO TABLE ADMISSIONS
  FIELDS TERMINATED BY ',' ESCAPED BY '\\' OPTIONALLY ENCLOSED BY '"'
  LINES TERMINATED BY '\n'
  IGNORE 1 LINES
  (@ROW_ID,@SUBJECT_ID,@HADM_ID,@ADMITTIME,@DISCHTIME,@DEATHTIME,@ADMISSION_TYPE,@ADMISSION_LOCATION,@DISCHARGE_LOCATION,@INSURANCE,@LANGUAGE,@RELIGION,@MARITAL_STATUS,@ETHNICITY,@EDR
SET
  ROW_ID = @ROW_ID,
  SUBJECT_ID = @SUBJECT_ID,
  HADM_ID = @HADM_ID,
  ADMITTIME = @ADMITTIME,
  DISCHTIME = @DISCHTIME,
  DEATHTIME = IF(@DEATHTIME='', NULL, @DEATHTIME),
  ADMISSION_TYPE = @ADMISSION_TYPE,
  ADMISSION_LOCATION = @ADMISSION_LOCATION,
  DISCHARGE_LOCATION = @DISCHARGE_LOCATION,
  INSURANCE = @INSURANCE,
  LANGUAGE = IF(@LANGUAGE='', NULL, @LANGUAGE),
  RELIGION = IF(@RELIGION='', NULL, @RELIGION),
  MARITAL_STATUS = IF(@MARITAL_STATUS='', NULL, @MARITAL_STATUS),
  ETHNICITY = @ETHNICITY,
  EDREGTIME = IF(@EDREGTIME='', NULL, @EDREGTIME),
  EDOUTTIME = IF(@EDOUTTIME='', NULL, @EDOUTTIME),
  DIAGNOSIS = IF(@DIAGNOSIS='', NULL, @DIAGNOSIS),
  HOSPITAL_EXPIRE_FLAG = @HOSPITAL_EXPIRE_FLAG,
  HAS_CHARTEVENTS_DATA = @HAS_CHARTEVENTS_DATA;
```


inputevents_cv
ROW_ID INT
SUBJECT_ID INT
HADM_ID INT
ICUSTAY_ID INT
CHARTTIME DATETIME
ITEMID INT
AMOUNT DECIMAL(22,10)
AMOUNTUOM VARCHAR(30)
RATE DECIMAL(22,10)
RATEUOM VARCHAR(30)
STORETIME DATETIME
CGID INT
ORDERID INT
LINKORDERID INT
STOPPED VARCHAR(30)
NEWBOTTLE TINYINT
ORIGINALAMOUNT DECIMAL(22,10)
ORIGINALAMOUNTUOM VARCHAR(30)
ORIGINALROUTE VARCHAR(30)
ORIGINALRATE DECIMAL(22,10)
ORIGINALRATEUOM VARCHAR(30)
ORIGINALSITE VARCHAR(30)
Indexes
microbiologyevents
ROW_ID INT
SUBJECT_ID INT
HADM_ID INT
CHARTDATE DATETIME
CHARTTIME DATETIME
SPEC_ITEMID INT
SPEC_TYPE_DESC VARCHAR(100)
ORG_ITEMID INT
ORG_NAME VARCHAR(100)
ISOLATE_NUM TINYINT
AB_ITEMID INT
AB_NAME VARCHAR(30)
DILUTION_TEXT VARCHAR(10)
DILUTION_COMPARISON VARCHAR(20)
DILUTION_VALUE INT
INTERPRETATION VARCHAR(5)
Indexes
d_icd_procedures
ROW_ID INT
ICD9_CODE VARCHAR(10)
SHORT_TITLE VARCHAR(50)
LONG_TITLE VARCHAR(255)
Indexes
d_cpt
ROW_ID INT
CATEGORY INT
SECTIONRANGE VARCHAR(100)
SECTIONHEADER VARCHAR(50)
SUBSECTIONRANGE VARCHAR(100)
SUBSECTIONHEADER VARCHAR(255)
CODESUFFIX VARCHAR(5)
MINCODEINSUBSECTION INT
MAXCODEINSUBSECTION INT
Indexes
inputevents_mv
ROW_ID INT
SUBJECT_ID INT
HADM_ID INT
ICUSTAY_ID INT
STARTTIME DATETIME
ENDTIME DATETIME
ITEMID INT
AMOUNT DECIMAL(22,10)
AMOUNTUOM VARCHAR(30)
RATE DECIMAL(22,10)
RATEUOM VARCHAR(30)
STORETIME DATETIME
CGID INT
ORDERID INT
LINKORDERID INT
ORDERCATEGORYNAME VARCHAR(100)
SECONDARYORDERCATEGORYNAME VARCHAR(100)
ORDERCOMPONENTTYPEDESCRIPTION VARCHAR(20)
ORDERCATEGORYDESCRIPTION VARCHAR(50)
PATIENTWEIGHT DECIMAL(22,10)
TOTALAMOUNT DECIMAL(22,10)
TOTALAMOUNTUOM VARCHAR(255)
ISOPENBAG TINYINT
CONTINUEINNEXTDEPT TINYINT
CANCELREASON TINYINT
STATUSDESCRIPTION VARCHAR(30)
COMMENTS_EDITEDBY VARCHAR(30)
COMMENTS_CANCELEDBY VARCHAR(30)
COMMENTS_DATE DATETIME
ORIGINALAMOUNT DECIMAL(22,10)
ORIGINALRATE DECIMAL(22,10)
Indexes
cptevents
ROW_ID INT
SUBJECT_ID INT
HADM_ID INT
COSTCENTER VARCHAR(10)
CHARTDATE DATETIME
CPT_CD VARCHAR(10)
CPT_NUMBER INT
CPT_SUFFIX VARCHAR(5)
TICKET_ID_SEQ INT
SECTIONHEADER VARCHAR(50)
SUBSECTIONHEADER VARCHAR(255)
DESCRIPTION VARCHAR(200)
Indexes
d_items
ROW_ID INT
ITEMID INT
LABEL VARCHAR(200)
ABBREVIATION VARCHAR(100)
DBSOURCE VARCHAR(20)
LINKSTO VARCHAR(50)
CATEGORY VARCHAR(100)
UNITNAME VARCHAR(100)
PARAM_TYPE VARCHAR(30)
CONCEPTID INT
Indexes
procedureevents_mv
ROW_ID INT
SUBJECT_ID INT
HADM_ID INT
ICUSTAY_ID INT
STARTTIME DATETIME
ENDTIME DATETIME
ITEMID INT
VALUE DECIMAL(22,10)
VALUEUOM VARCHAR(30)
LOCATION VARCHAR(30)
LOCATIONCATEGORY VARCHAR(30)
STORETIME DATETIME
CGID INT
ORDERID INT
LINKORDERID INT
ORDERCATEGORYNAME VARCHAR(100)
SECONDARYORDERCATEGORYNAME VARCHAR(100)
ORDERCATEGORYDESCRIPTION VARCHAR(50)
ISOPENBAG TINYINT
CONTINUEINNEXTDEPT TINYINT
CANCELREASON TINYINT
STATUSDESCRIPTION VARCHAR(30)
COMMENTS_EDITEDBY VARCHAR(30)
COMMENTS_CANCELEDBY VARCHAR(30)
COMMENTS_DATE DATETIME
Indexes
transfers
ROW_ID INT
SUBJECT_ID INT
HADM_ID INT
ICUSTAY_ID INT
DBSOURCE VARCHAR(20)
EVENTTYPE VARCHAR(20)
PREV_CAREUNIT VARCHAR(20)
CURR_CAREUNIT VARCHAR(20)
PREV_WARDID TINYINT
CURR_WARDID TINYINT
INTIME DATETIME
OUTTIME DATETIME
LOS DECIMAL(22,10)
Indexes
noteevents
ROW_ID INT
SUBJECT_ID INT
HADM_ID INT
CHARTDATE DATE
CHARTTIME DATETIME
STORETIME DATETIME
CATEGORY VARCHAR(50)
DESCRIPTION VARCHAR(255)
CGID INT
ISERROR TINYINT
TEXT MEDIUMTEXT
Indexes
caregivers
ROW_ID INT
CGID INT
LABEL VARCHAR(15)
DESCRIPTION VARCHAR(30)
Indexes
callout
ROW_ID INT
SUBJECT_ID INT
HADM_ID INT
SUBMIT_WARDID TINYINT
SUBMIT_CAREUNIT VARCHAR(15)
CURR_WARDID TINYINT
CURR_CAREUNIT VARCHAR(15)
CALLOUT_WARDID TINYINT
CALLOUT_SERVICE VARCHAR(10)
REQUEST_TELE TINYINT
REQUEST_RESP TINYINT
REQUEST_CDIFF TINYINT
REQUEST_MRSA TINYINT
REQUEST_VRE TINYINT
CALLOUT_STATUS VARCHAR(20)
CALLOUT_OUTCOME VARCHAR(20)
DISCHARGE_WARDID TINYINT
ACKNOWLEDGE_STATUS VARCHAR(20)
CREATETIME DATETIME
UPDATETIME DATETIME
ACKNOWLEDGETIME DATETIME
OUTCOMETIME DATETIME
FIRSTRESERVATIONTIME DATETIME
CURRENTRESERVATIONTIME DATETIME
Indexes
drgcodes
ROW_ID INT
SUBJECT_ID INT
HADM_ID INT
DRG_TYPE VARCHAR(20)
DRG_CODE VARCHAR(20)
DESCRIPTION VARCHAR(255)
DRG_SEVERITY TINYINT
DRG_MORTALITY TINYINT
Indexes
d_icd_diagnoses
ROW_ID INT
ICD9_CODE VARCHAR(10)
SHORT_TITLE VARCHAR(50)
LONG_TITLE VARCHAR(255)
Indexes
outputevents
ROW_ID INT
SUBJECT_ID INT
HADM_ID INT
ICUSTAY_ID INT
CHARTTIME DATETIME
ITEMID INT
VALUE DECIMAL(22,10)
VALUEUOM VARCHAR(30)
STORETIME DATETIME
CGID INT
STOPPED VARCHAR(30)
NEWBOTTLE CHAR(1)
ISERROR TINYINT
Indexes
admissions
ROW_ID INT
SUBJECT_ID INT
HADM_ID INT
ADMITTIME DATETIME
DISCHTIME DATETIME
DEATHTIME DATETIME
ADMISSION_TYPE VARCHAR(50)
ADMISSION_LOCATION VARCHAR(50)
DISCHARGE_LOCATION VARCHAR(50)
INSURANCE VARCHAR(255)
LANGUAGE VARCHAR(10)
RELIGION VARCHAR(50)
MARITAL_STATUS VARCHAR(50)
ETHNICITY VARCHAR(200)
EDREGTIME DATETIME
EDOUTTIME DATETIME
DIAGNOSIS VARCHAR(255)
HOSPITAL_EXPIRE_FLAG TINYINT
HAS_CHARTEVENTS_DATA TINYINT
Indexes
labevents
ROW_ID INT
SUBJECT_ID INT
HADM_ID INT
ITEMID INT
CHARTTIME DATETIME
VALUE VARCHAR(200)
VALUENUM DECIMAL(22,10)
VALUEUOM VARCHAR(20)
FLAG VARCHAR(20)
Indexes
prescriptions
ROW_ID INT
SUBJECT_ID INT
HADM_ID INT
ICUSTAY_ID INT
STARTDATE DATETIME
ENDDATE DATETIME
DRUG_TYPE VARCHAR(100)
DRUG VARCHAR(100)
DRUG_NAME_POE VARCHAR(100)
DRUG_NAME_GENERIC VARCHAR(100)
FORMULARY_DRUG_CD VARCHAR(120)
GSN VARCHAR(200)
NDC VARCHAR(120)
PROD_STRENGTH VARCHAR(120)
DOSE_VAL_RX VARCHAR(120)
DOSE_UNIT_RX VARCHAR(120)
FORM_VAL_DISP VARCHAR(120)
FORM_UNIT_DISP VARCHAR(120)
ROUTE VARCHAR(120)
Indexes
datetimeevents
ROW_ID INT
SUBJECT_ID INT
HADM_ID INT
ICUSTAY_ID INT
ITEMID INT
CHARTTIME DATETIME
STORETIME DATETIME
CGID INT
VALUE DATETIME
VALUEUOM VARCHAR(50)
WARNING TINYINT
ERROR TINYINT
RESULTSTATUS VARCHAR(50)
STOPPED VARCHAR(50)
Indexes
icustays
ROW_ID INT
SUBJECT_ID INT
HADM_ID INT
ICUSTAY_ID INT
DBSOURCE VARCHAR(20)
FIRST_CAREUNIT VARCHAR(20)
LAST_CAREUNIT VARCHAR(20)
FIRST_WARDID TINYINT
LAST_WARDID TINYINT
INTIME DATETIME
OUTTIME DATETIME
LOS DECIMAL(22,10)
Indexes
chartevents
ROW_ID INT
SUBJECT_ID INT
HADM_ID INT
ICUSTAY_ID INT
ITEMID INT
CHARTTIME DATETIME
STORETIME DATETIME
CGID INT
VALUE TEXT
VALUENUM DECIMAL(22,10)
VALUEUOM VARCHAR(50)
WARNING TINYINT
ERROR TINYINT
RESULTSTATUS VARCHAR(50)
STOPPED VARCHAR(50)
Indexes
diagnoses_icd
ROW_ID INT
SUBJECT_ID INT
HADM_ID INT
SEQ_NUM TINYINT
ICD9_CODE VARCHAR(10)
Indexes
procedures_icd
ROW_ID INT
SUBJECT_ID INT
HADM_ID INT
SEQ_NUM TINYINT
ICD9_CODE VARCHAR(10)
Indexes
services
ROW_ID INT
SUBJECT_ID INT
HADM_ID INT
TRANSFERTIME DATETIME
PREV_SERVICE VARCHAR(20)
CURR_SERVICE VARCHAR(20)
Indexes
patients
ROW_ID INT
SUBJECT_ID INT
GENDER VARCHAR(5)
DOB DATETIME
DOD DATETIME
DOD_HOSP DATETIME
DOD_SSN DATETIME
EXPIRE_FLAG TINYINT
Indexes
d_labitems
ROW_ID INT
ITEMID INT
LABEL VARCHAR(100)
FLUID VARCHAR(100)
CATEGORY VARCHAR(100)
LOINC_CODE VARCHAR(100)
Indexes

Part II

Notable Tables and Characteristics

Table Descriptions and Characteristics

DRGCODES

- Diagnosis Related Groups are codes used by the hospital for grouping diagnoses.
- There are two sources for these codes: MS and HCFA
- Can change over time, so must also use the description field when extracting these records

ICUSTAYS

- Defines each ICU Stay, so patients can have multiple ICUStay_ID
- Provides the ICU database source relevant to that patient; they used CareVue from 2001-2008 and Metavision from 2008-2012

Table Descriptions and Characteristics

INPUTEVENTS_CV & INPUTEVENTS_MV

- These tables house the events that take place during a patient stay in either the CareVue or Metavision database.
- Each DB has similar, but different fields so must be familiar with each.

LABEVENTS

- This table contains all lab measurements for a certain patients
- Lab measurements include the results from sampling a patient specimen (ie blood, urine, saliva)
- Time for these events is logged when specimen is acquired, not when data is available.

Table Descriptions and Characteristics

MICROBIOLOGYEVENTS

- This table is similar to the LABEVENTS table except it details situations where microbiology is required
- Provides details and results from various tests involving microbiology cultures

NOTEEVENTS

- Details the patient notes taken by a doctor or nurse during the course of their ICU Stay
- Category is an important field as it can often tell us when a patient is discharged
- The TEXT field can be very long and should be extracted carefully as a result

Table Descriptions and Characteristics

ADMISSIONS

- Entity - All hospital admissions data
- Notably includes demographics of patients and admission/discharge times.
- Also include binary expiry flag that shows if patient is deceased.

CALLOUT

- Entity - Provides information about ICU discharge planning
- Shows when a caregiver suggests callout, and if patient complied
- Also has “REQUEST” fields outlining specific diagnoses

CAREGIVERS

- Entity table including caregiver information and CGID key field

Table Descriptions and Characteristics

CPTEVENTS

- Association table - outlines billing events
- One-to-many relationship with CAREGIVERS through CGID foreign key

CHARTEVENTS

- Association table w/ 330 million rows - physical measurements and “chartable” events
- 5 different foreign keys (joins PATIENTS, ADMISSIONS, ICUSTAYS, D_ITEMS, CAREGIVERS all with one-to-many relationships)

DATETIMEEVENTS

- Association table w/ 4 million rows - datetime of all measurements of a patient
- 5 different foreign keys (joins PATIENTS, ADMISSIONS, ICUSTAYS, D_ITEMS, CAREGIVERS all with one-to-many relationships)

Table Descriptions and Characteristics

PATIENTS

- Entity table - identifies the patient's sex, date of birth, date of death, and a mortality signifier
- One-to-many relationship with ADMISSIONS and ICUSTAYS

OUTPUTEVENTS

- Entity table - identifies the chart times of the output events as well as details concerning the solutions and the caregivers responsible for the admission
- This table joins with the following tables:
    - PATIENTS
    - ADMISSIONS
    - ICUSTAYS
    - D_ITEMS
    - CAREGIVERS

Table Descriptions and Characteristics

PRESCRIPTIONS

- Entity table - contains the identifiers and details about prescriptions written in the hospital
- Joins with the PATEINTS, ADMISSIONS, and ICUSTAYS tables

PROCEDUREEVENTS_MV

- Entity table - contains the identifiers and details regarding events surrounding procedures
- Joins with the following tables:
  - PATEINTS
  - ADMISSIONS
  - ICUSTAYS
  - D_ITEMS

Table Descriptions and Characteristics

PROCEDURES_ICD

- Association table - contains identifiers and the order in which procedures took place with the associated ICD-9 procedure code.
- Joins with the PATIENTS, ADMISSIONS, and D_ICD_PROCEDURES tables

SERVICES

- Association table - Contains transfer times along with the previous service performed as well as the current service being performed
- Joins with the PATIENTS and ADMISSIONS tables

TRANSFERS

- Entity table - contains the identifiers and details regarding the transfers of patients
- Joins with the PATIENTS, ADMISSIONS, and ICUSTAYS tables.

Table Descriptions and Characteristics

DICTIONARY TABLES (5 total ‘D_’ tables)

- Association tables: provide a high-level dictionary of procedural terminology used in the medical field
  - D_CPT
  - D_ICD_PROCEDURES
  - C_ICD_DIAGNOSES
  - D_ITEMS
  - D_LABITEMS
- ICD9_CODE links DIANGOSES_ICD and PROCEDURES_ICD tables separately with a one-to-one relationship
- ITEM_ID links the following tables with a one-to-one relationship:
  - CHARTEVENTS, DATETIMEEVENTS, INPUTEVENTS_MV, MICROBIOLOGYEVENTS, OUTPUTEVENTS, PROCEDUREEVENTS_MV, LABEVENTS

Table Descriptions and Characteristics

DIAGNOSES_ICD

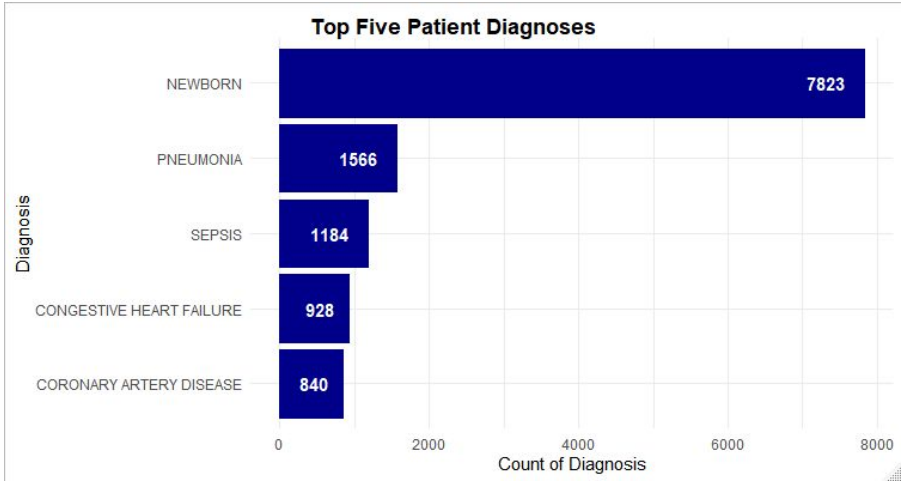
- Association table - contains diagnoses which relate to a specific hospital's admissions through the use of the ICD9 Code system.
- Links to the PATIENTS table with a one-to-many relationship
- Links to the ADMISSIONS table with a one-to-one relationship

PART III

Exploring Major Characteristics of the MIMIC III Database

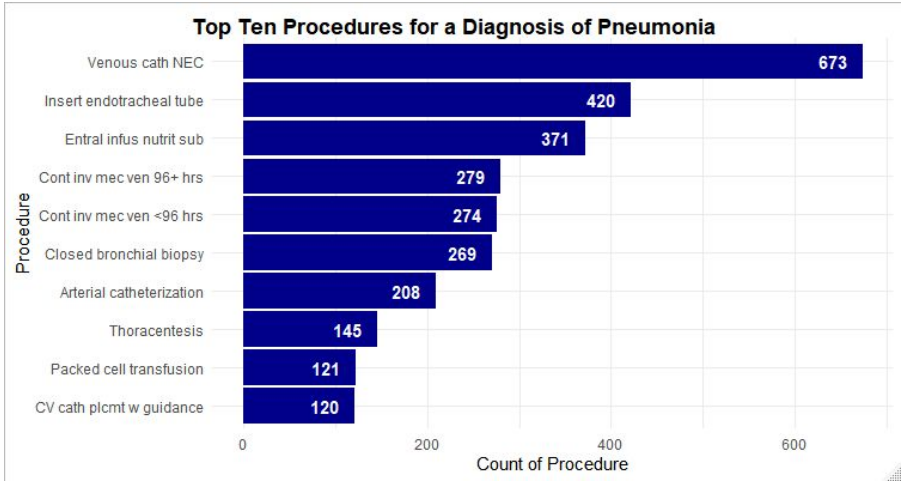
Top Five Patient Diagnoses

- Displayed is the top five patient diagnostics from the database.
- The initial exploration was to find the most common diagnoses.
- This allows further drilling down because a baseline of what patients are dealing with has been presented



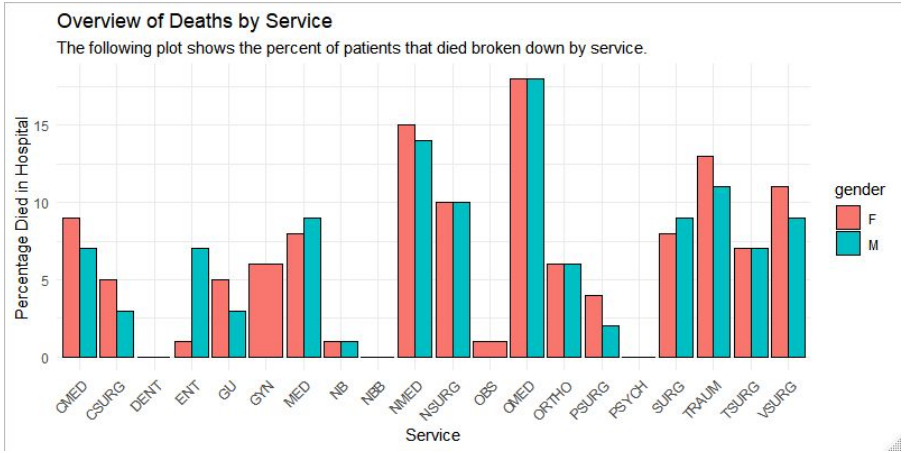
Pneumonia Procedures

- Displays the number of each procedure performed on the pneumonia diagnostic ranked from high to low, Venous Cath NEC being the most commonly performed
- This plot was derived from the top five diagnostics, pneumonia being the second most common.



Gender: Deaths By Service

- This displays the proportion of individuals, separated by gender, who died within the hospital during a particular visit.
- The largest difference between the genders was recorded at the ENT (ear, nose, and throat) where there was a difference of > 5%.
- Beyond this plot, patients were bucketed into age groups for each service to further investigate the death rate of these services.



Mortality Statistics

Goals were to find:

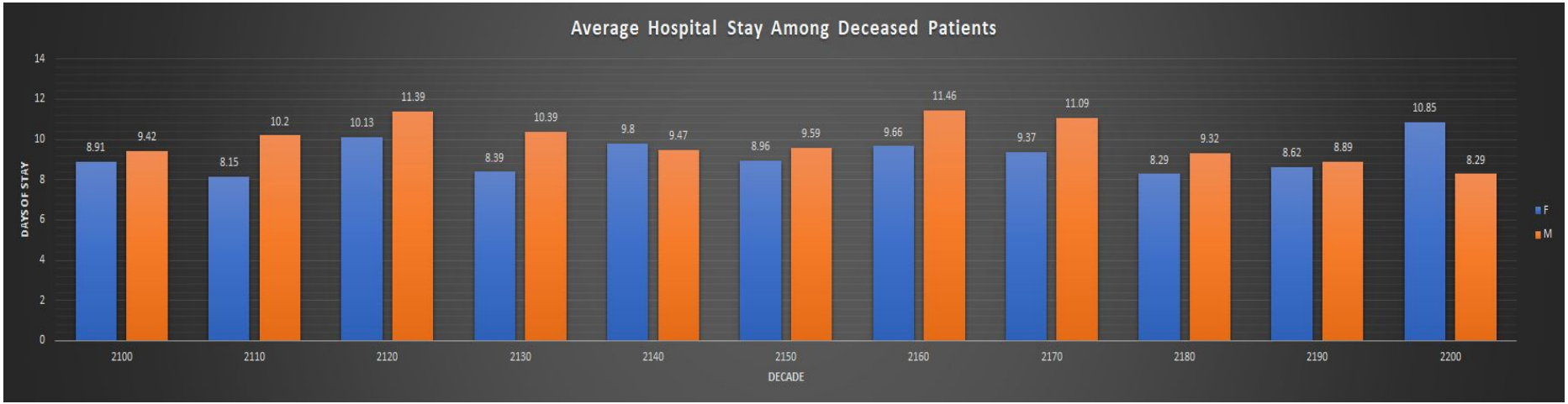
- Average hospital stay for deceased patients
- Most common diagnoses shared among deceased grouped by any time series and gender

Achieved by creating a view joining:

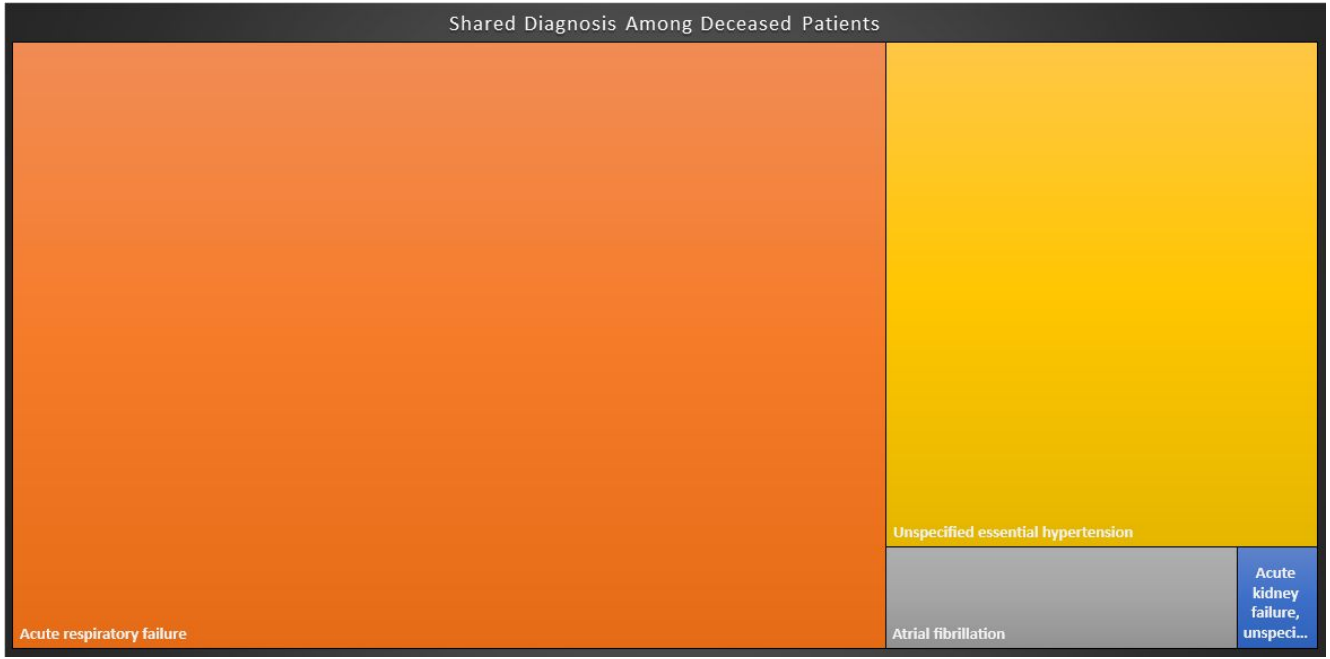
- ADMISSIONS
- PATIENTS
- DIAGNOSES_ICD
- D_ICD_DIAGNOSES

Condition - HOSPITAL_EXPIRE_FLAG = 1

Average Hospital Stay

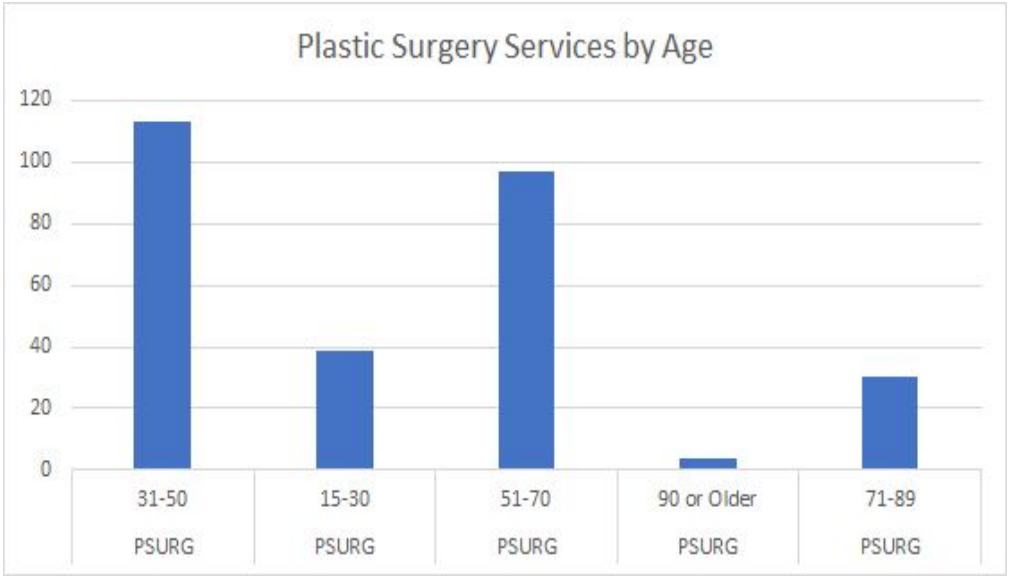


Commonly Shared Diagnoses - Deceased Patients



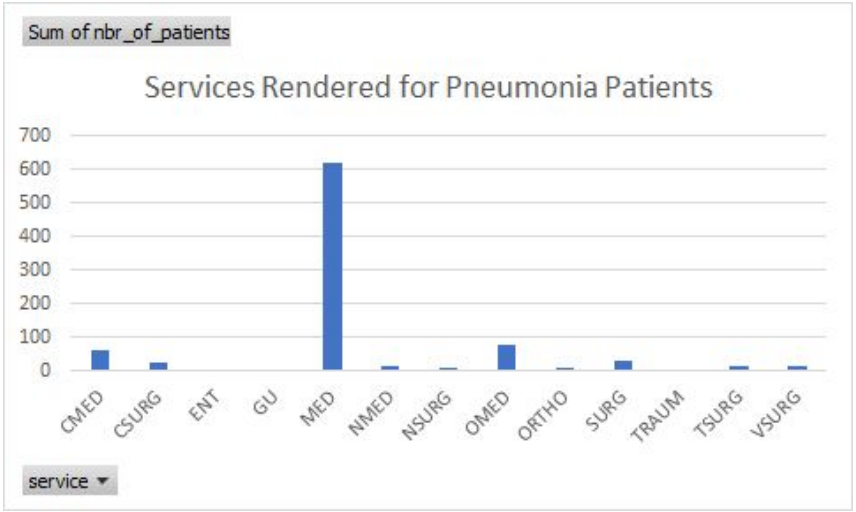
First-Visit Services Rendered by Age Group

- Displays the services rendered to first-time ICU visits by age group
- Assists hospital with predicting the resources needed and helps prioritize patients
- This graph shows the age for Plastic Surgery, shows to be a more common service for younger people in relation to other services.



First-Admit Services for Pneumonia Patients

- Search can filter by diagnosis to determine which services are commonly rendered in the case of a specific diagnosis.



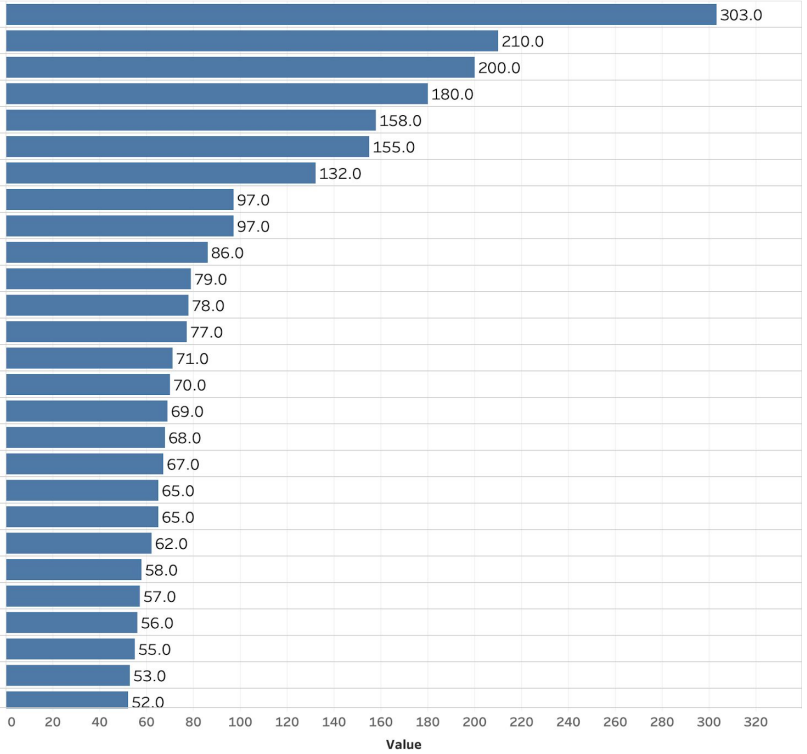
ICD9 Codes and associated diagnoses and procedure

- ICD9 codes were joined with their respective short and long descriptions of its diagnosis and procedure.
- All ICD9 codes were extracted from the DIAGNOSES_ICD and PROCEDURE_ICD tables.
- This serves as a foundation to pursue future analytics that deal with features such datetime, gender, and mortality rate.

Icd9 Code	Diagnoses Short	Diagnoses Long	Procedure Short	Procedure Long
3572	Neuropathy in diabet...	Polyneuropathy in di...	Ventr septa def rep N...	Other and unspecifie...
41400	Cor ath unsp vsl ntv/...	Coronary atheroscler...	*null*	*null*
4019	Hypertension NOS	Unspecified essentia...	Lymphatic diag proc ...	Other diagnostic pro...
V4581	Aortocoronary bypass	Aortocoronary bypas...	*null*	*null*
41071	Subendo infarct, initial	Subendocardial infar...	*null*	*null*
5849	Acute kidney failure ...	Acute kidney failure, ...	Urethral repair NEC	Other repair of urethra
4280	CHF NOS	Congestive heart fail...	*null*	*null*
25060	DMII neuro nt st unc...	Diabetes with neurol...	*null*	*null*
41401	Crnry athrscl natve v...	Coronary atheroscler...	*null*	*null*
4019	Hypertension NOS	Unspecified essentia...	Lymphatic diag proc ...	Other diagnostic pro...
2720	Pure hypercholester...	Pure hypercholester...	*null*	*null*
42731	Atrial fibrillation	Atrial fibrillation	*null*	*null*
V1582	History of tobacco use	Personal history of t...	*null*	*null*
2859	Anemia NOS	Anemia, unspecified	*null*	*null*
3572	Neuropathy in diabet...	Polyneuropathy in di...	Ventr septa def rep N...	Other and unspecifie...
78552	Septic shock	Septic shock	*null*	*null*

ICD9 Codes and associated diagnoses and procedure

Icd9 Code	Diagnoses Short	Procedure Short	Value
4019	Hypertension NOS	Lymphatic diag proc NEC	303.0
42731	Atrial fibrillation	Null	210.0
41401	Crnry athrscl natve vssl	Null	200.0
4280	CHF NOS	Null	180.0
25000	DMII wo cmp nt st uncntr	Null	158.0
5849	Acute kidney failure NOS	Urethral repair NEC	155.0
2724	Hyperlipidemia NEC/NOS	Mouth biopsy NOS	132.0
5990	Urin tract infection NOS	Null	97.0
51881	Acute respiratry failure	Null	97.0
2720	Pure hypercholesterolem	Null	86.0
2859	Anemia NOS	Null	79.0
53081	Esophageal reflux	Null	78.0
V5861	Long-term use anticoagul	Null	77.0
2449	Hypothyroidism NOS	Null	71.0
2851	Ac posthemorrhag anemia	Null	70.0
V053	Need prphyl vc vrl hepat	Null	69.0
5859	Chronic kidney dis NOS	Null	68.0
40390	Hy kid NOS w cr kid I-IV	Null	67.0
2762	Acidosis	Cleft palate correction	65.0
99592	Severe sepsis	Null	65.0
486	Pneumonia, organism NOS	Null	62.0
412	Old myocardial infarct	Splenotomy	58.0
V290	NB obsrv suspct infect	Null	57.0
496	Chr airway obstruct NEC	Excision of anus	56.0
311	Depressive disorder NEC	Temporary tracheostomy	55.0
5070	Food/vomit pneumonitis	Null	53.0
0389	Septicemia NOS	Null	52.0



Part IV

Analysis of Chapters 21 & 23

Chapter 21: Modeling Mortality

- Extracting predictor variables with the intent of building various models to predict mortality in the ICU:
- Existing ICU performance metrics used to test the validity of the model:
  - APACHE Scores
  - The simplified acute physiology score
  - The mortality probability model
  - The sequential organ failure assessment score
- New performance metrics are extracted...

Chapter 21: Extraction

- Only the first ICU admission of Adult patients is taken into account
- In order to successfully complete this task a number of adjustments were made to accommodate the requirements needed to work on the MIMIC-III database:
  - Identifiers used in the MIMIC-II database do not exist in the MIMIC-III, so a new query needed to be built to recreate the ICUSTAY_DETAIL
    - This included a ‘dense_rank’ function ‘partitioned’ over SUBJECT_ID or HADM_ID to get various ICU sequences
  - Further calculations were performed to determine the following features as well:
    - AGE
    - DAYS_TILL_DEATH_AFTER_DISCHARGE

Chapter 21: Data Preparation

- The following steps were done in R after the data was extracted, this is to ensure the data is compatible with various mortality prediction models:
  - All patients older than 89 had their aged masked in MIMIC-III
    - These records were replaced with the median age of 91.4
  - All records with incomplete data were removed

Chapter 23: Propensity Score Analysis

- Model that enables best selection for test and control groups
- Patient data must be extracted from MIMIC III using SQL to build model
  - The data to be extracted includes:
    - Demographics
    - Vital Signs
    - Basic Metabolic Panels
    - Past Medical Conditions
    - Disease Severity Scores
    - Types of Admission
    - Types of ICU
- **OUR GOAL** is to properly extract these pieces of information to feed the model for its propensity analysis and enable researchers to assign subjects to the correct group

Chapter 23: Data Extraction Process

All pulled into single view to load into an analytics platform (i.e. *R*, *Python*, *or SAS*)

- Demographics
  - joined from ADMISSIONS and PATIENTS tables
- Vital Signs
  - created “pivot” subquery as *base_events_pivot* by pulling avg values for items in CHARTEVENTS table
- Basic Metabolic Panels
  - created “pivot” subquery as *labs_pivot* by pulling max values for items in LABEVENTS table
- Past Medical Conditions
  - Separate table for reference by analyst
- Types of Admission
  - Joined admissions table
- Types of ICU
  - Joined ICUSTAYS table