The University of Texas at Austin
Hildebrand Department of Petroleum and Geosystems Engineering
Cockrell School of Engineering

MAY 8, 2020

CS 327E ELEMENTS OF DATABASES: PROJECT FINAL PRESENTATION

VAXONOMICS

VEE LEE KOH AND NOAH PLACKE
The University of Texas at Austin

PROJECT OVERVIEW

Primary Dataset :

- Vaccine Adverse Event Reporting System (VAERS) 2018 by CDC
    - Adverse Events, Vaccines, and Symptoms

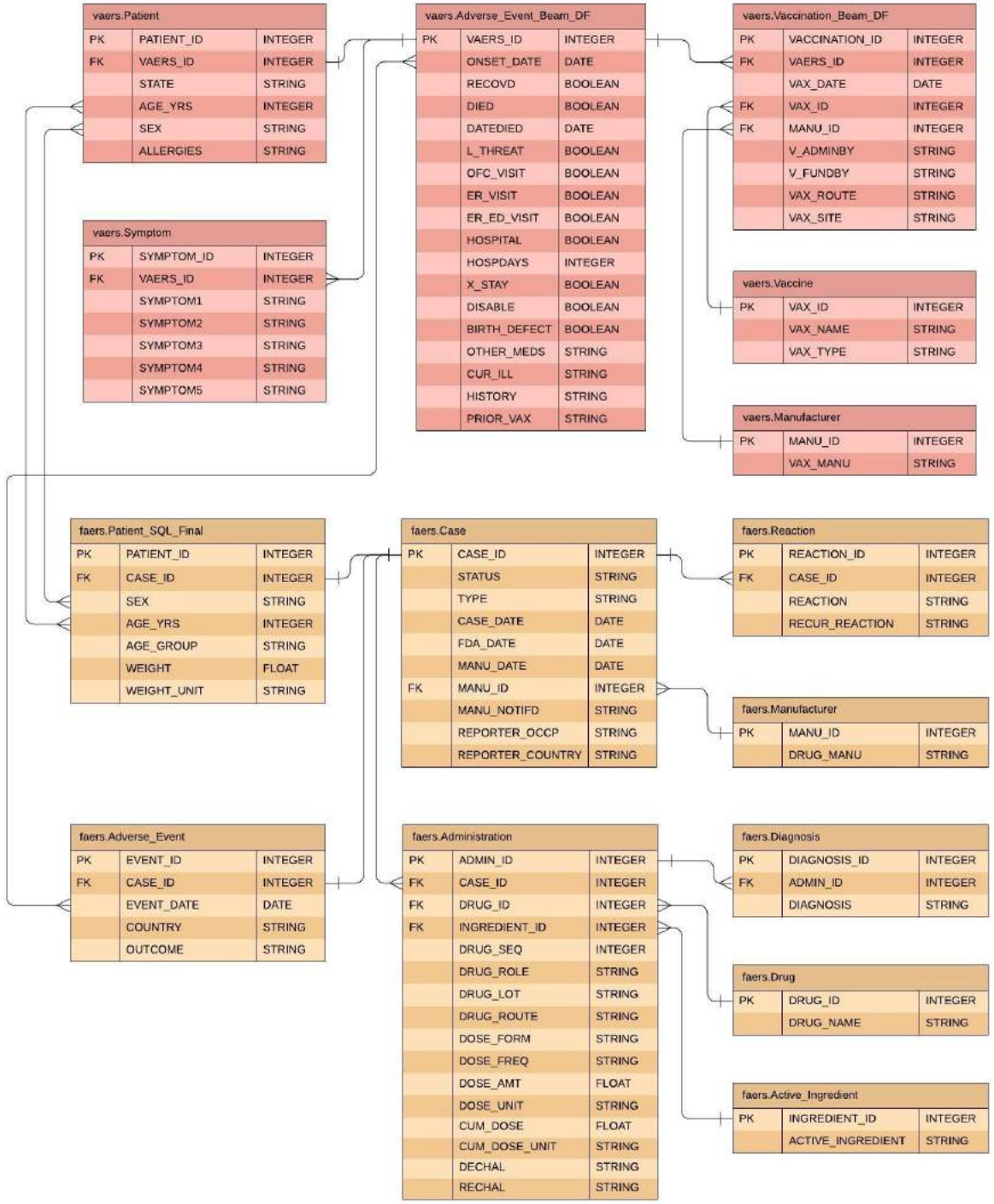
Secondary Dataset :

- FDA Adverse Event Reporting System (FAERS) Q4 2018 by FDA
    - Demographics, Drugs, Reactions, Outcomes, and Diagnosis

***Goal of Project** :*

- Investigate if there is correlation between adverse events and attributes of patients, vaccines, drugs, or external factors

MODELED TABLES ERD



The University of Texas at Austin
Hildebrand Department of Petroleum and Geosystems Engineering
Cockrell School of Engineering

Beam Transforms on VAERS Dataset

***Adverse Event Table** :*

- Convert String attributes into Booleans
    - Recovered and Birth defect
- Standardize values of Boolean attributes
    - Died, Life threatening, Hospitalized, ER visit, Disabled, etc.

***Vaccination Table** :*

- Standardize unknown values of String attributes

```python
class FormatRECOVDFn(beam.DoFn):
  def process(self, element):
    # Process to convert RECOVD attribute into boolean
    # Attribute type will be changed to boolean in schema
    event_record = element

    # get RECOVD attribute
    recovd = event_record.get('RECOVD')

    # print current RECOVD values
    print('Current RECOVD: ', recovd)

    # Convert RECOVD value into None if Unknown (U)
    if recovd == 'U':
        recovd = None
    # Convert RECOVD value into True if Yes (Y)
    elif recovd == 'Y':
        recovd = True
    # Convert RECOVD value into False if No (N)
    elif recovd == 'N':
        recovd = False

    # print new RECOVD values
    print('New RECOVD: ', recovd)

    # update advserse event records
    event_record['RECOVD'] = recovd

    # return advserse event records
    return [event_record]
```

ParDo Transform for ‘Recovered’ attribute

SQL Transforms on FAERS Dataset

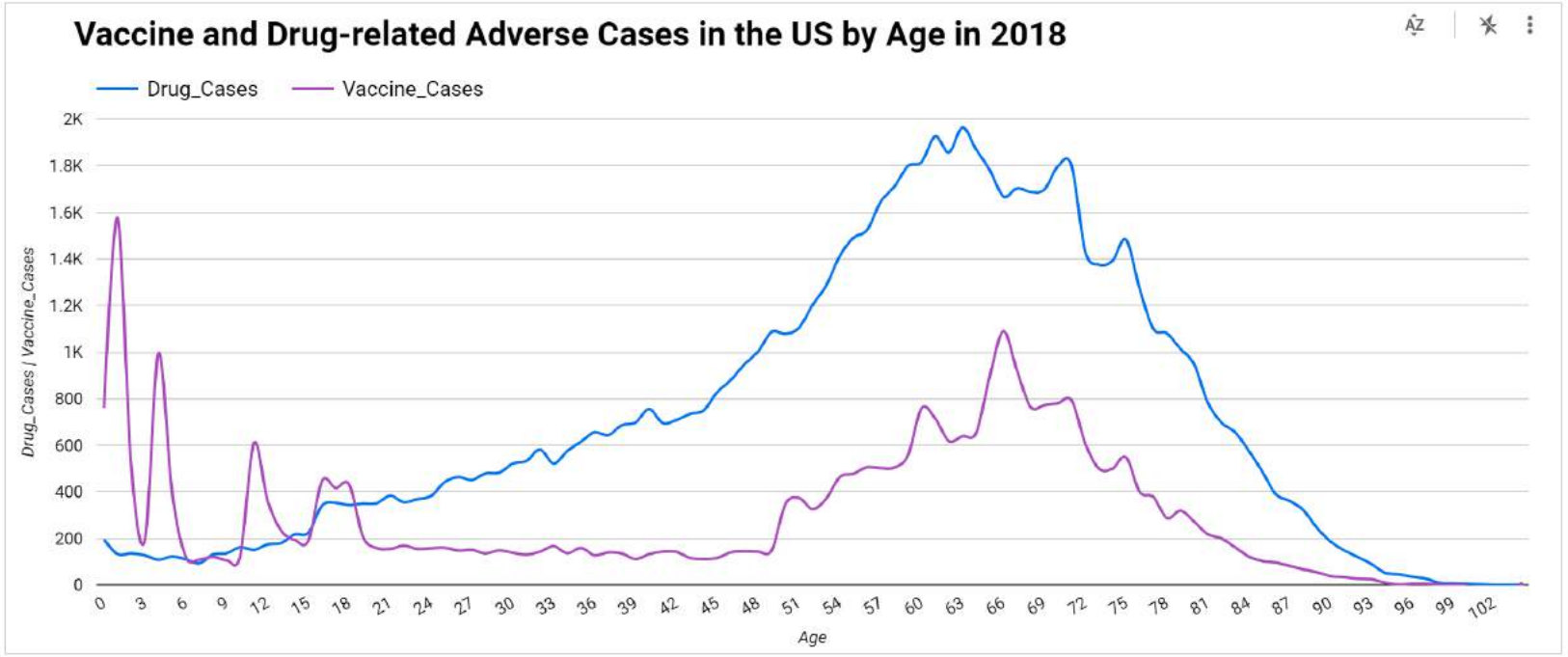
***Patient Table** :*

- Standardize **age** values into a single unit
    - Intermediate tables : Convert decades, months, weeks, days, and hours into years
    - Final table : Union all tables
    - Check if all records retained

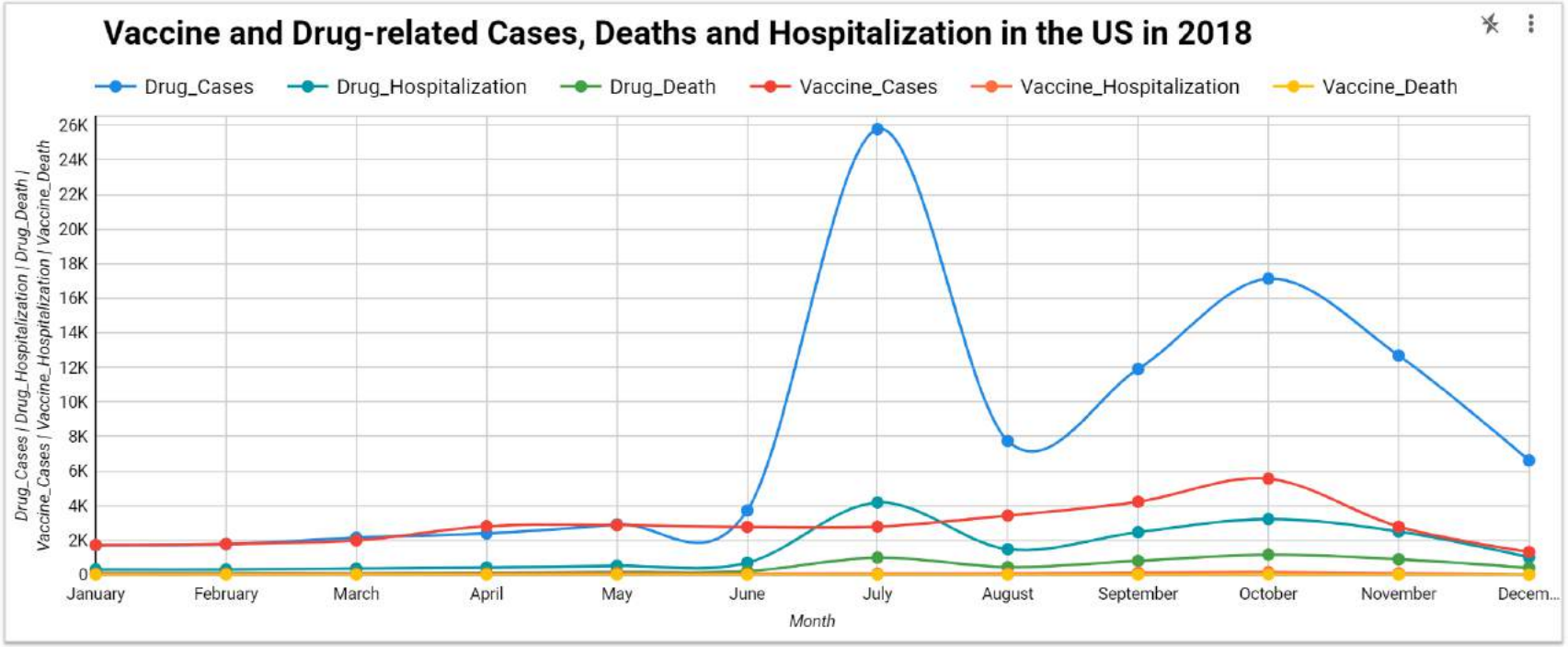
Cross - Dataset Queries

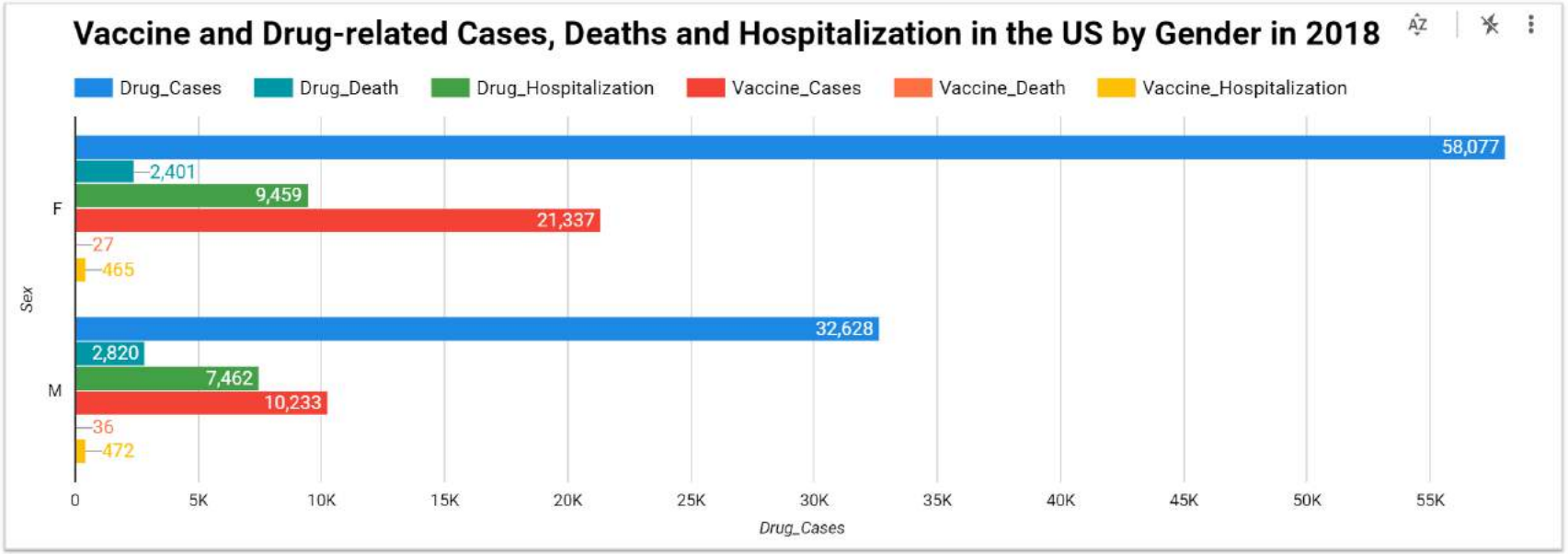
Joining Datasets *:*

- Full Outer Join : Vaccine and drug-related adverse events by **age**
- Inner Join : Adverse events, deaths and hospitalizations by **month**
- Inner Join : Adverse events, deaths and hospitalizations by **gender**



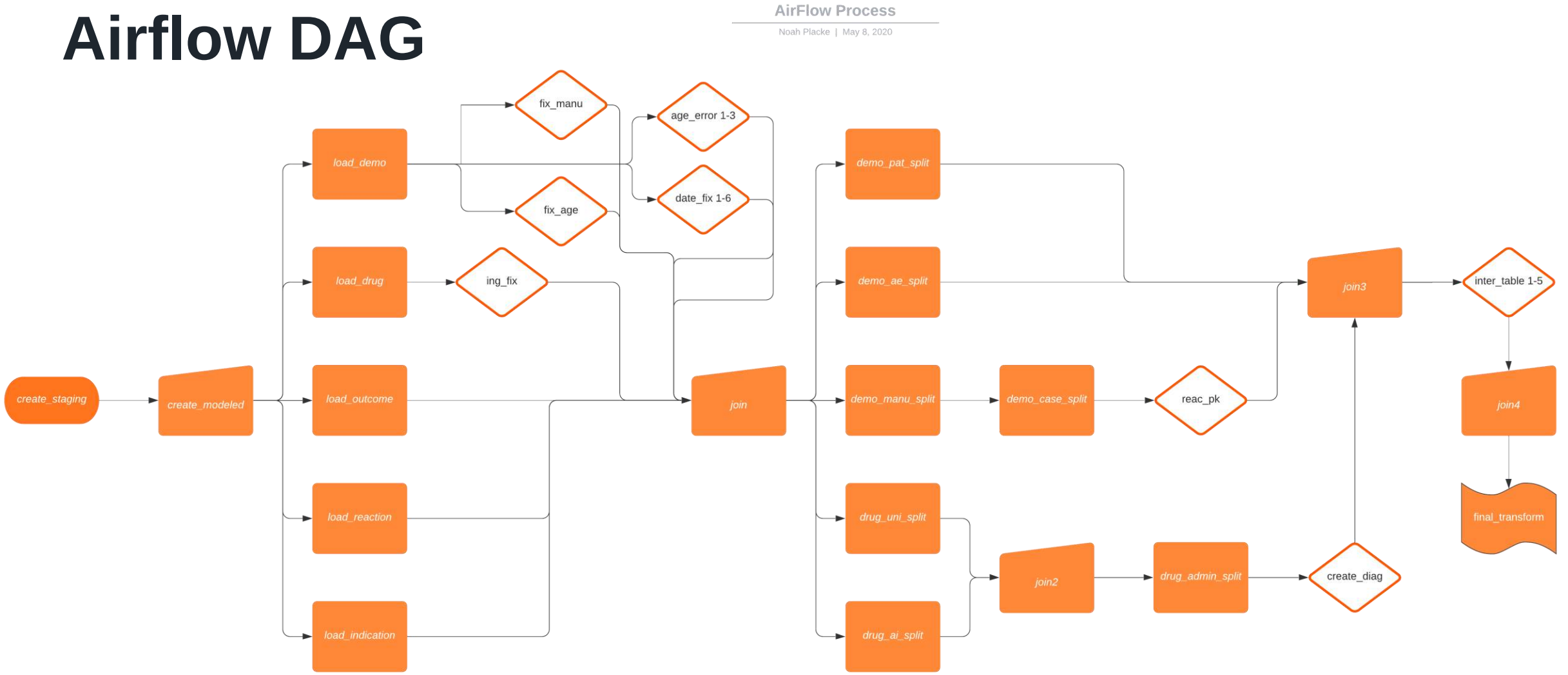
CROSS - DATASET VISUALS





The University of Texas at Austin
Hildebrand Department of Petroleum and Geosystems Engineering
Cockrell School of Engineering

Airflow DAG


AirFlow Process
Noah Placke | May 8, 2020
fix_manu
age_error 1-3
load_demo
date_fix 1-6
fix_age
load_drug
ing_fix
create_staging
create_modeled
load_outcome
join
load_reaction
load_indication
demo_pat_split
demo_ae_split
demo_manu_split
demo_case_split
reac_pk
join3
inter_table 1-5
join4
drug_uni_split
join2
drug_admin_split
create_diag
drug_ai_split
final_transform

Future Improvements to Solution

***Data Required for Better Solution** :*

- Residential states of patients → Assess insurance coverage & healthcare quality by state
- Disease data → Join datasets by disease and count adverse events
- Standardize manufacturer names → Join datasets by manufacturer

THANK YOU!