

MIMICiii Workshop

Jeon Young Seok

Saw Swee Hock School of Public Health, NUS

Email: ephjys@nus.edu.sg

Objectives

- Introduction of MIMICiii **Schema**
- Working with the MIMICiii database using **SQL**
- Simple statistical analysis using Python3 in **Colab** environment



BigQuery



MIMICiii : Introduction

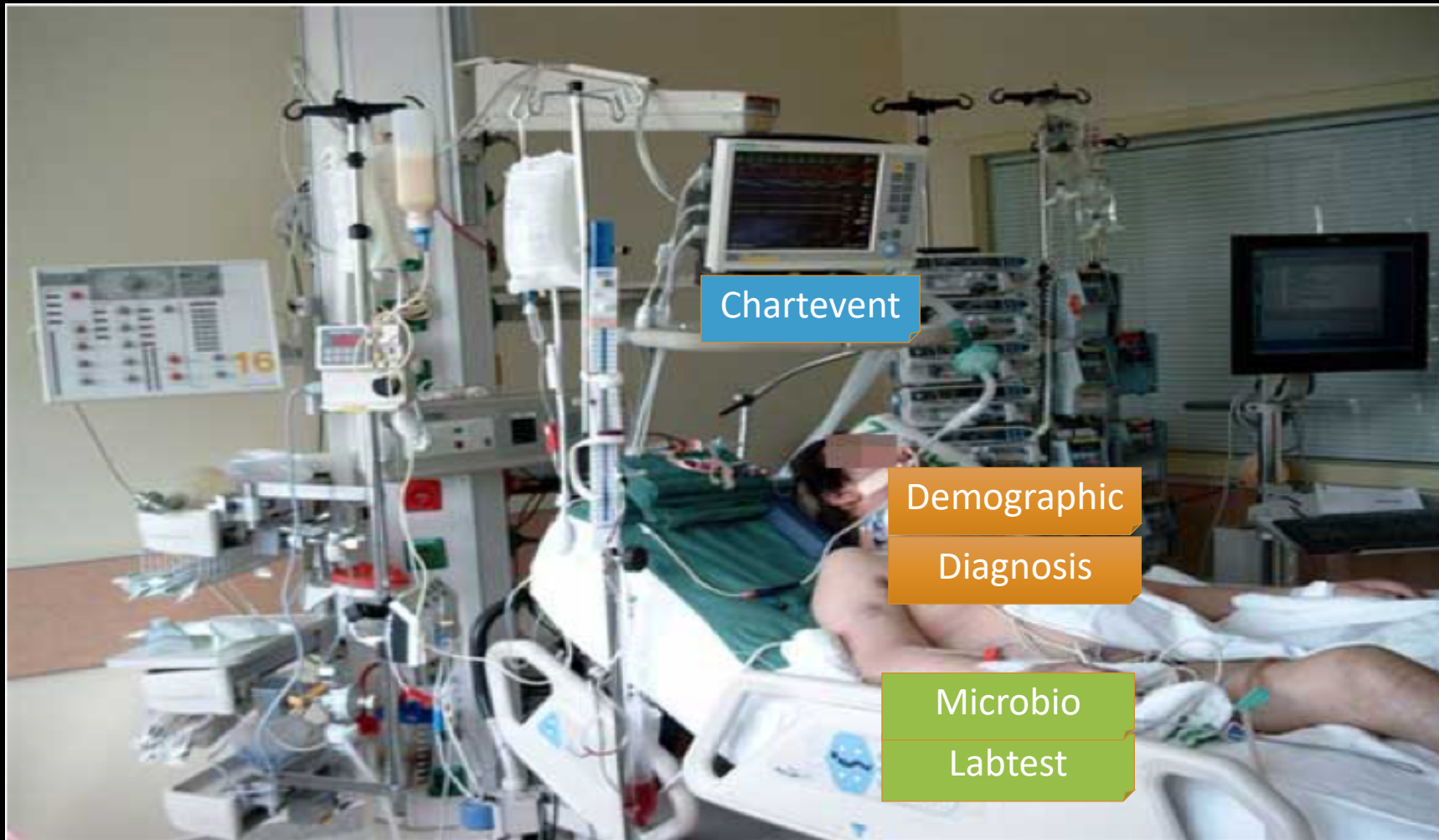
- A publicly available database released from MIT
- from Beth Israel Deaconess Medical Center between 2001 ~ 2012.
- Contains detailed information of patients' history
(Demographic, Diagnosis, Vital signs, Medication, Notevents, etc)
- Only Patients who stayed in Intense Care Unit (ICU)



MIMIC III : Introduction



MIMIC III : Introduction



MIMIC III : Numbers

27 tables

ADMISSIONS, CAREGIVERS, CHARTEVENTS,

46,000 patients

46,520 distinct patients who have gone to ICU at least once

59,000 admissions

A patient may have been admitted to the hospital more than once

60,000 icustays

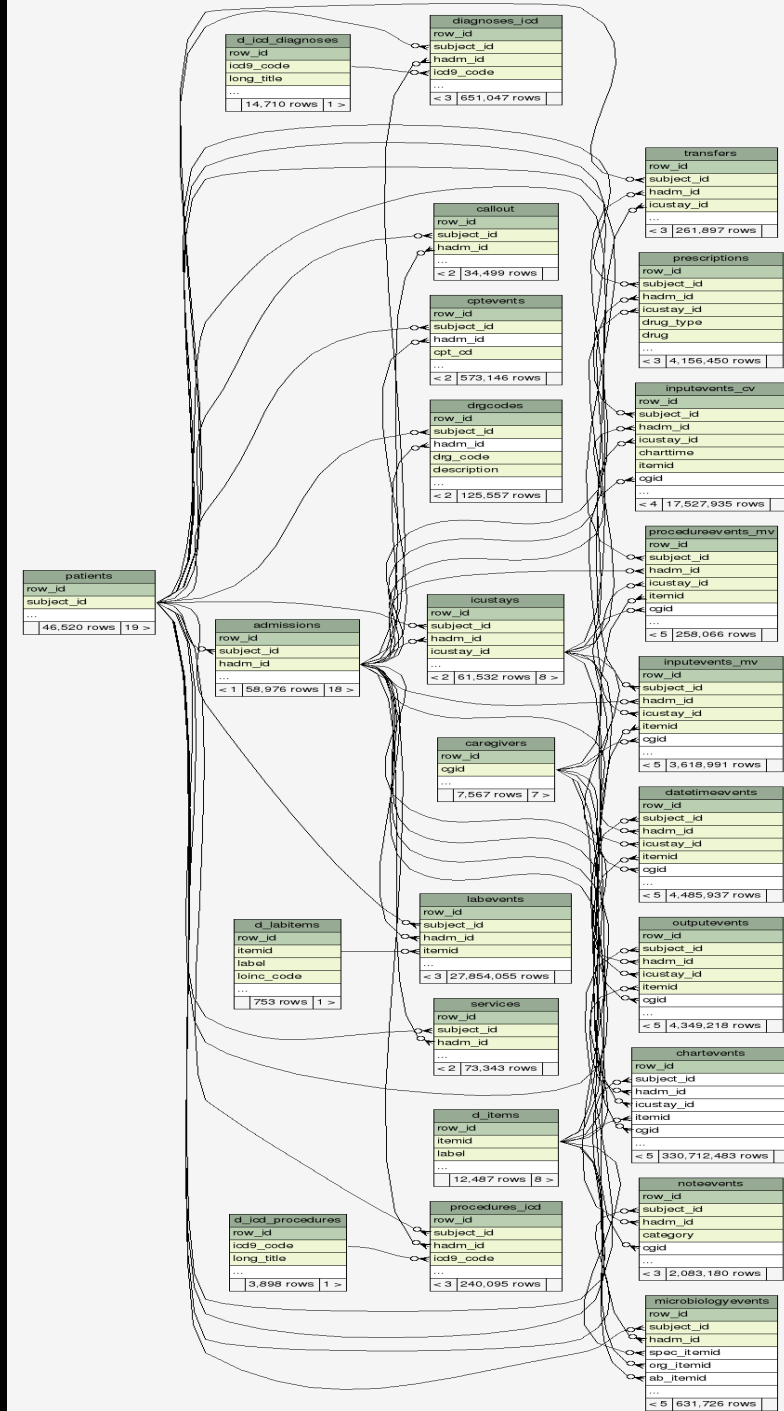
A single admission case could lead to multiple icustays

5,000 chartevents per admission case

MIMICiii : Schema

- **Patients** : Unique patients in MIMIC
- **Admissions** : Unique hospital admissions
- **Icustays** : Unique ICU stays.
- **Diagnoses_icd** : ICD diagnoses for patients
- **Chartevents** : Contains all chart data
- **D_xxx** : Dictionary Table

<https://mit-lcp.github.io/mimic-schema-spy/relationships.html>



The 3 core tables

Field name	Type
subject_id	INTEGER
gender	STRING
dob	DATETIME
dod	DATETIME
expire_flag	INTEGER

Patients

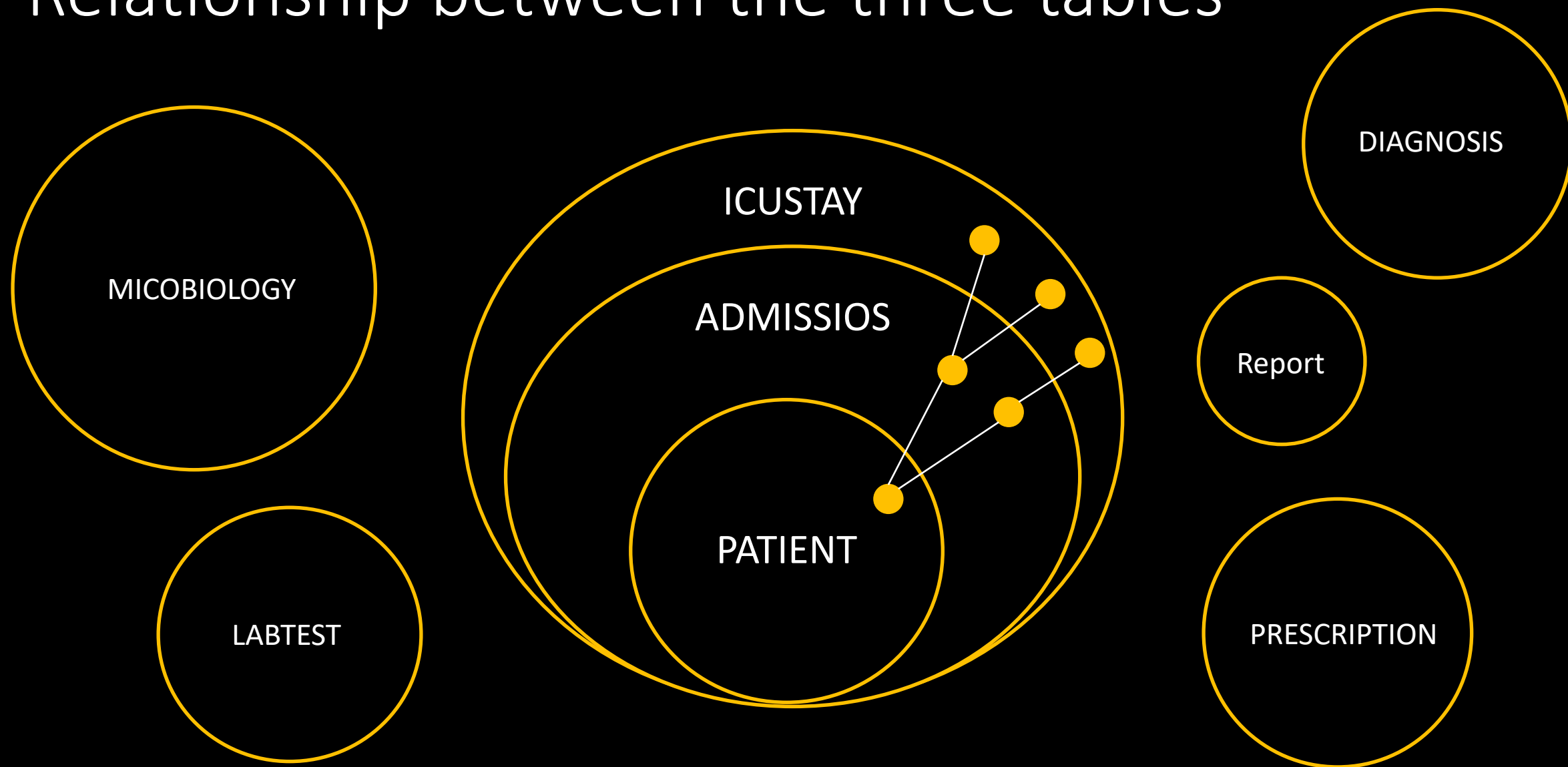
Field name	Type
subject_id	INTEGER
hadm_id	INTEGER
admittime	DATETIME
dischtime	DATETIME
deathtime	DATETIME

Admissions

Field name	Type
subject_id	INTEGER
hadm_id	INTEGER
icustay_id	INTEGER
intime	DATETIME
outtime	DATETIME

Icustays

Relationship between the three tables



BigQuery Setup

Working in BigQuery



BigQuery Project
korea-datathon-2019



Google Cloud Storage
Physionet-data:MIMIC

Working in BigQuery

- Go to **Gmail** and accept the invitation
- Go to “**My group**”
- You will now see “**Datathon Korea 2019**”

Invitation to join Datathon Korea 2018 Inbox x



Datathon Korea 2018 <datathon-korea-2018+noreply@googlegroups.com>

to me ▾

Datathon Korea 2018

Google Groups

Hi youngseokjeon74@gmail.com,

Kenneth Paik invited you to join the **Datathon Korea 2018** group.

Message from Kenneth Paik

Datathon

About this group

Datathon for Korea in 2018

Google Groups allows you to create and participate in online forums and email-based groups with a rich community experience. You can also use your Group to share documents, pictures, calendars, invitations, and other resources. [Learn more](#).

If you do not wish to be a member of this group or believe this group may contain spam, you can [report](#) the group for abuse. For additional information see our [help center](#).

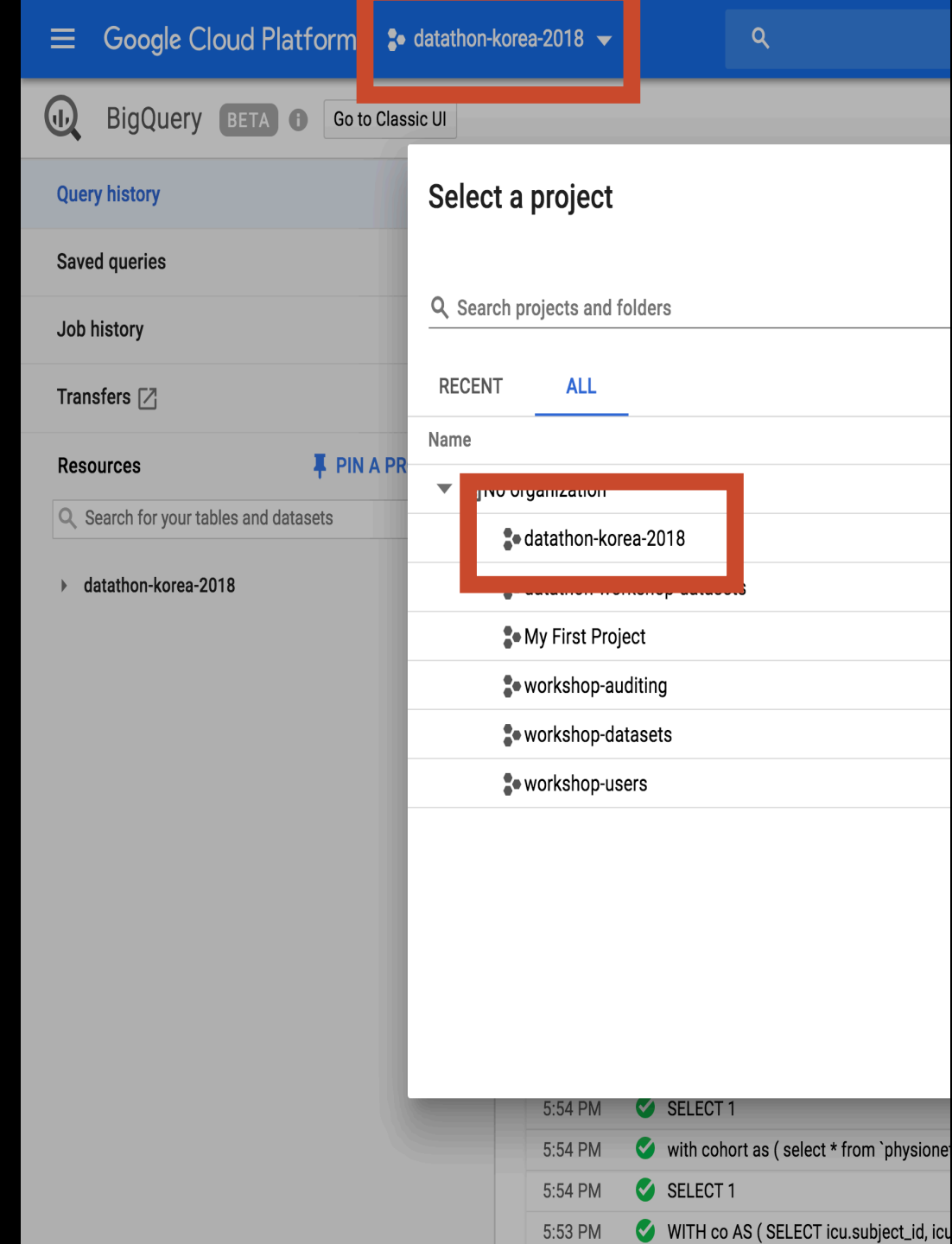
[Accept this invitation](#)

If you do not wish to be added to Google Groups in the future you can opt out [here](#).

[Start](#) a new group. [Visit](#) the help center.

Working in BigQuery

- Go to the link below
console.cloud.google.com/bigquery
- You are now a member of
“Datathon-korea-2019”



Working in BigQuery

- Click “ADD DATA” and “Pin a Project”
- Type “Physionet-data” and click OK

The screenshot displays the Google Cloud BigQuery interface. On the left, a sidebar menu lists various options: Query history, Saved queries, Job history, Transfers, Scheduled queries, BI Engine, and Resources. The 'Resources' section is expanded, showing a search bar and a list of projects: 'singapore-datathon-data', 'physionet-data' (which is selected and highlighted in blue), 'eicu_crd_demo', and 'mimicii_demo'. A red rectangular box highlights the 'ADD DATA' button with a dropdown arrow. On the right, a 'Pin a project' dialog box is open. It contains the instruction 'Select a project from the list below for easy access going forward.' and two radio button options: 'Enter a project name' (which is selected) and 'Search for a project'. Below the 'Enter a project name' option, a text input field contains the text 'physionet-data' with red dashed lines underneath it, indicating a search or validation process. Below the 'Search for a project' option, there is a button labeled 'Select a project'.

Working in BigQuery: select

```
1 SELECT
2   subject_id,
3   gender
4 FROM
5   `physionet-data.mimiciii_demo.patients`
```


Working in BigQuery: WHERE

```
1 SELECT
2     *
3 FROM
4     `physionet-data.mimiciii_demo.patients`
5 WHERE
6     gender = 'F'
7 LIMIT
8     5
```

Working in BigQuery: ADVANCED SELECT

```
1 SELECT
2   subject_id,
3   CASE WHEN
4     gender = 'M' THEN 1
5   ELSE      0   END AS gender_bi
6
7 FROM
8   `physionet-data.mimiciii_demo.patients`
```

Working in BigQuery: ADVANCED WHERE

```
1 SELECT
2     *
3 FROM
4     `physionet-data.mimiciii_demo.patients`
5 WHERE
6     subject_id > 10005 AND
7     UPPER(gender) = 'M'
```

Working in BigQuery: SUBQUERIES

```
1 SELECT
2   subject_id,
3   age
4 FROM (
5   SELECT
6     subject_id,
7     datetime_diff(CAST(dod AS datetime),
8                   CAST(dob AS datetime),
9                   year) AS age
10  FROM
11    `physionet-data.mimiciii_demo.patients` ) t
12 WHERE
13   age > 80|
```

Working in BigQuery: AGGREGATE

```
1 SELECT
2   AVG(age)
3 FROM (
4   SELECT
5     subject_id,
6     datetime_diff(CAST(dod AS datetime),
7                   CAST(dob AS datetime),
8                   year) AS age
9   FROM
10    `physionet-data.mimiciii_demo.patients` ) t
11 WHERE
12   age > 80
```

Using Colab with BigQuery

- Colab is a **Jupyter-like** service within google cloud sever
- It is possible to query database in **Bigquery**
- Your are able to develop machine learning models with pre-installed machine learning packages such as **Tensorflow, Sklearn** and **Pytorch**.
- **Only Python** is supported



Using Colab with BigQuery

- Go to the link below

<https://github.com/Youngseok0001/korea-datathon>

