Characterize a Given OMOP CDM Database

Initial Set-Up

Activating Project Environment

Packages Used in Analysis

`DatabaseConnector`

Package Description:

This package will be used to create connections to the database and execute queries to obtain characterizations about the cohorts in the database.

Analysis Introduction

Background on Analysis

The intent of this report is to investigate the characterization statement:

> **Characterization Statement 1:** Characterize the individuals being seen for mental health care services (related to depression, bipolar disorder, and suicidality) at least one time – including hospitalization events.

This characterization statement is founded on the central research topic for this study:

> **Research Topic:** Based on [CLAIMS], we see X% of all persons with at least one claim indicating [DEPRESSION/BIPOLAR DISORDER/SUICIDALITY] are not seen again.

By which the phrases "...all persons..." refer to those seen by patience care provider, etc. and "...are not seen again." implies lack of adherence to care.

Data Analysis Preparation

Creating Initial Connection

Defining connection details for connecting to a given database:

Creating Initial OMOP Tables

In this case, the schema follows the OMOP CDM v5 schema.

For this analysis, we will work with the following tables from the schema:

- `PERSON`
- `LOCATION`
- `OBSERVATION_PERIOD`
- `CONDITION_OCCURRENCE`

Defining Conditions

Using the concept_ids for bipolar, depression, and suicidality, we can grab the concept set related to the each condition.

General Characterization of Data

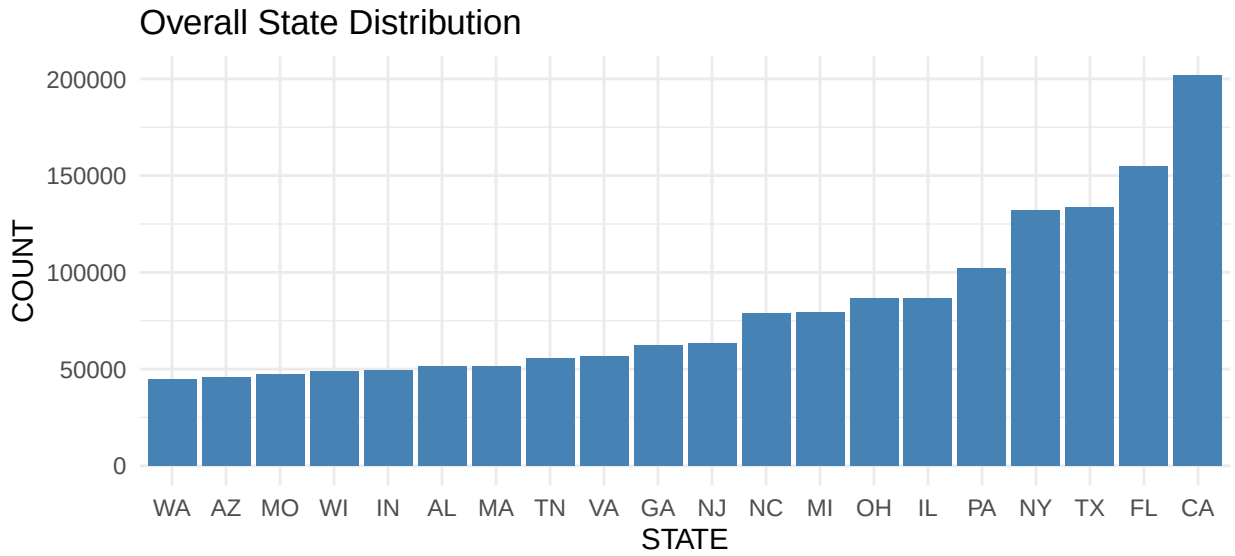
In this dataset, there are 2326856 patients. The dates in the dataset range from 2008-01-01 and 2010-12-31. To further examine this data, we can break them down across the following axes:

- State
- Age
- Race
- Gender
- Care setting

Workflow for Location/State, Race, Ethnicity, and Gender To find the the demographics of the individuals in the data set, I created a parameterized SQL containing three parameters: the table/domain, concept name (demographic is default concept name, but location is diff), and a bool var for specific demographic or entire table

1. Select the `person` table
2. Join the `person` table with the `location` table joining on `location_id`
3. Join with `concept` table using the different `concept_id` to find the names of the concept ids in the person table
4. Group by demographic (location, race, ethnicity, gender, etc)
5. Return the counts of the patients by demographics

State Breakdown

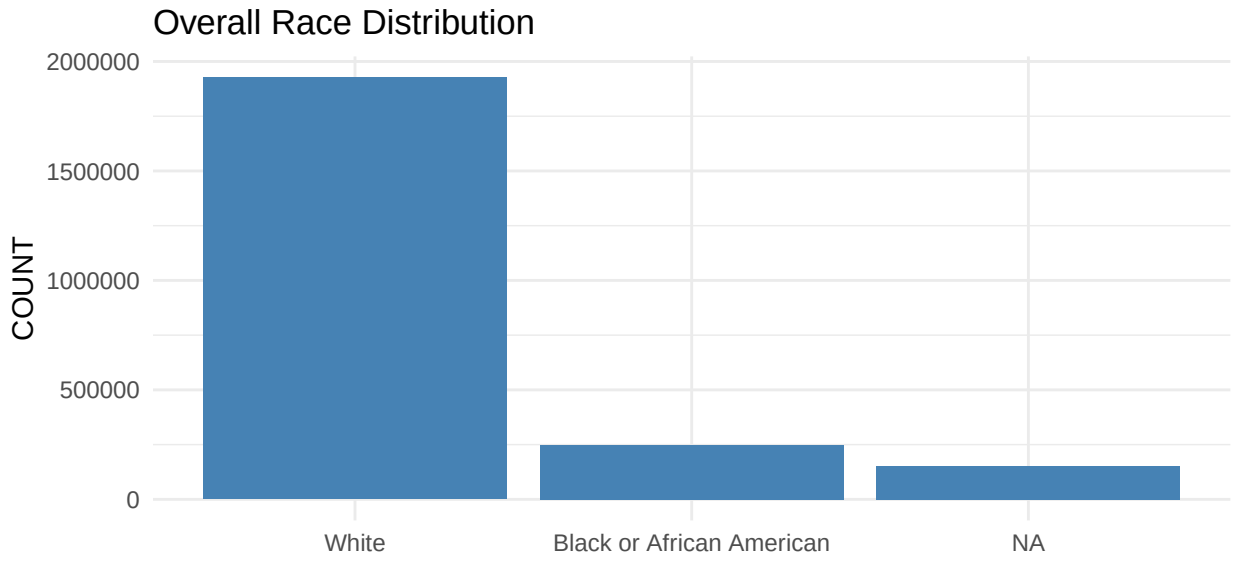


The top states are: CA, FL, TX, NY, PA, IL, OH, MI, NC, NJ

STATE	COUNT
CA	201651

STATE	COUNT
FL	155040
TX	133809
NY	132059
PA	102191
IL	86557
OH	86325
MI	79556
NC	78912
NJ	63475

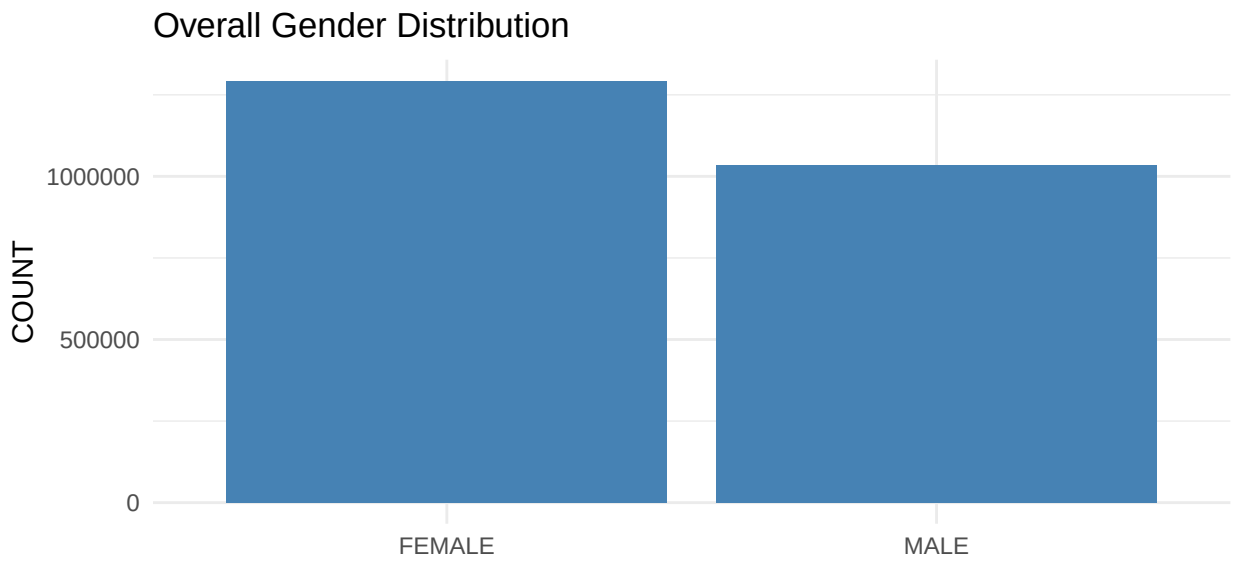
Racial Breakdown



In the dataset, there are 1926708 White persons, 247723 Black or African American persons, 152425 NA persons

RACE_CONCEPT_NAME	COUNT
White	1926708
Black or African American	247723
NA	152425

Gender Breakdown



In the dataset, there are 1292861 FEMALE persons, 1033995 MALE persons

GENDER_CONCEPT_NAME	COUNT
FEMALE	1292861
MALE	1033995

Age Breakdown To find the range of ages of the dataset:

1. Select the `visit_occurrence` table
2. Join on the `person` table using `person_id`
3. Subtract the `visit_end_date` from the `year_of_birth`
4. Report the minimum age and maximum age of the person

The age of the patient at the time of visit ranged from 25 years of age to 101 years of age.

Care Site Breakdown To understand more about the care site breakdowns of the dataset:

1. Select the `visit_occurrence` table
2. Join with the `concept` table to get the concept names using `concept_id`
3. Group by the care site settings
4. Return the counts distinct persons for the care site

In the dataset, there are 1984582 unique patients with an indicated care site. The care site setting Inpatient Visit had 111637582 visits

CONCEPT_NAME	VISIT_COUNT	PERSON_COUNT
Inpatient Visit	111637582	1984582

Inpatient Cohort

Patient Spread To find the characterization of the inpatient persons in the dataset:

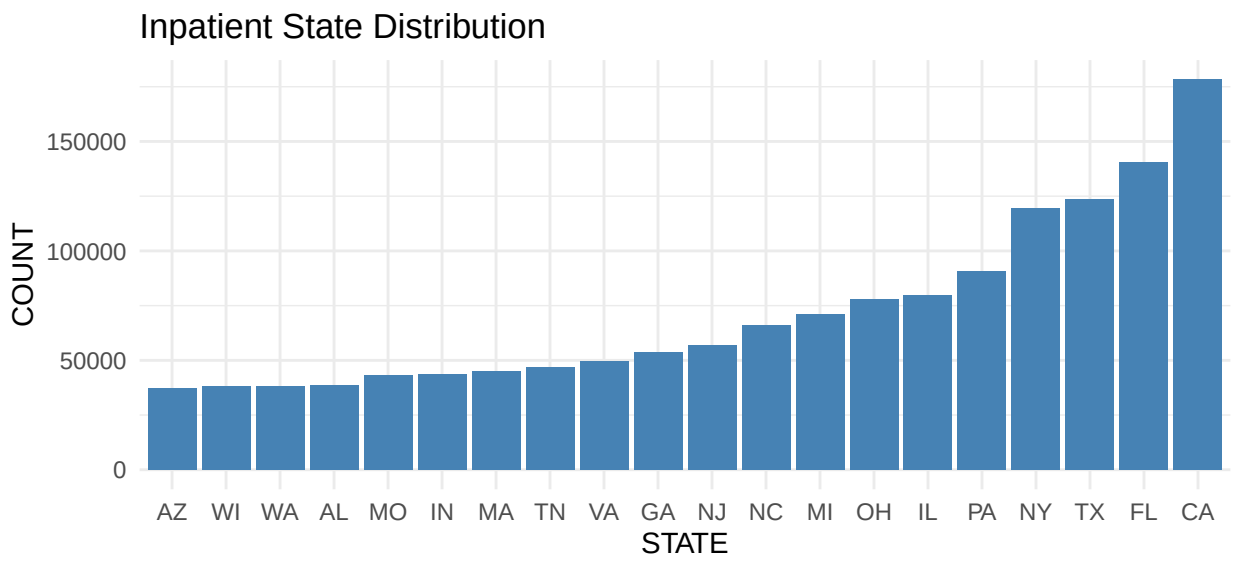
1. Select the `visit_occurrence` table
2. Join on the `person` table using `person_id`
3. Select the different demographics and join them on `concept` using `concept_id` to find the names of the race, ethnicity, and gender concept IDs.

In the inpatient cohort, there are a total of 1984582 distinct patients, and there are a total of 111637582 entries. Compared to the original unique person count in the overall dataset, which is 2326856, 342274 are not accounted for in the inpatient count.

To further examine this data, we can break them down across the following axes:

- State
- Age
- Race
- Gender
- Care setting

State Breakdown



In the inpatient cohort, the top 5 states by person count are

STATE	COUNT
CA	178254
FL	140385
TX	123530
NY	119443
PA	90893
IL	79577
OH	77830
MI	70896
NC	65822
NJ	56971

Age Breakdown For age breakdowns, inpatient counts were bucketed based on their age. The age brackets were as follows: - 0-9 - 10-19 - 20-29 - 30-39 - 40-49 - 50-59 - 60-69 - 70-79 - 80-89 - $\geq$ 90

To find the age breakdown of the dataset:

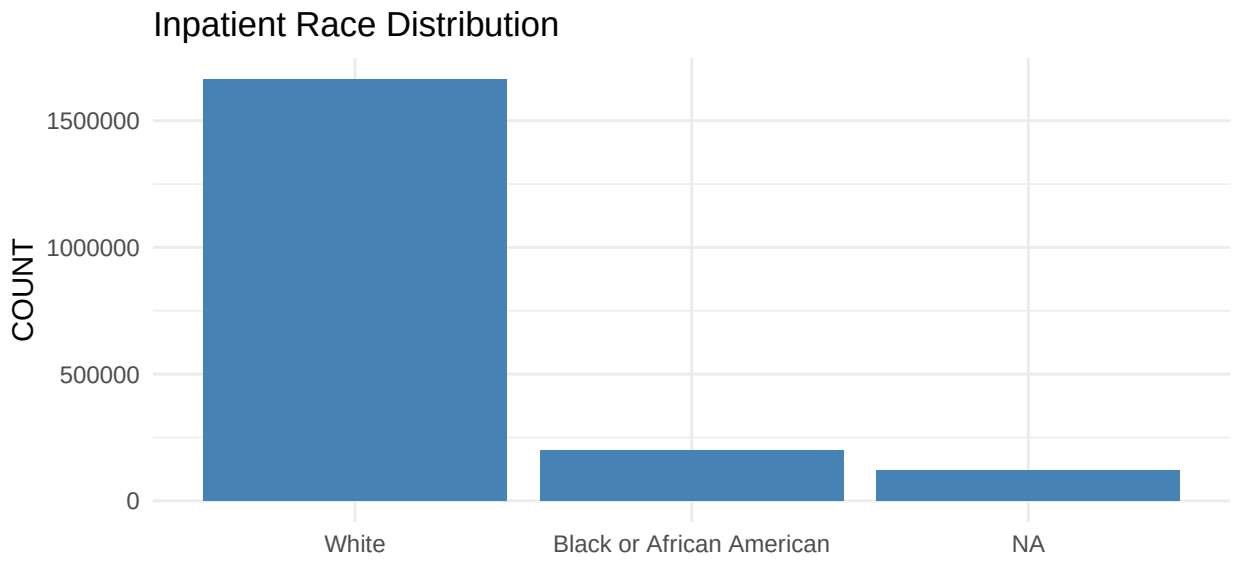
1. Select the `visit_occurrence` table
2. Join on the `person` table using `person_id`
3. Subtract the `visit_end_date` from the `year_of_birth`

In the inpatient dataset, it takes in the visit encounters, and the ages for the individuals were found according to the time of the inpatient visit.

This table shows the number of occurences of the ages in the data set.

	COUNT
AGE_0_9	NA
AGE_10_TO_19	NA
AGE_20_TO_29	587663
AGE_30_TO_39	1943789
AGE_40_TO_49	3919222
AGE_50_TO_59	6582101
AGE_60_TO_69	22398681
AGE_70_TO_79	39729991
AGE_80_TO_89	28226316
AGE_90_AND_MORE	8249819

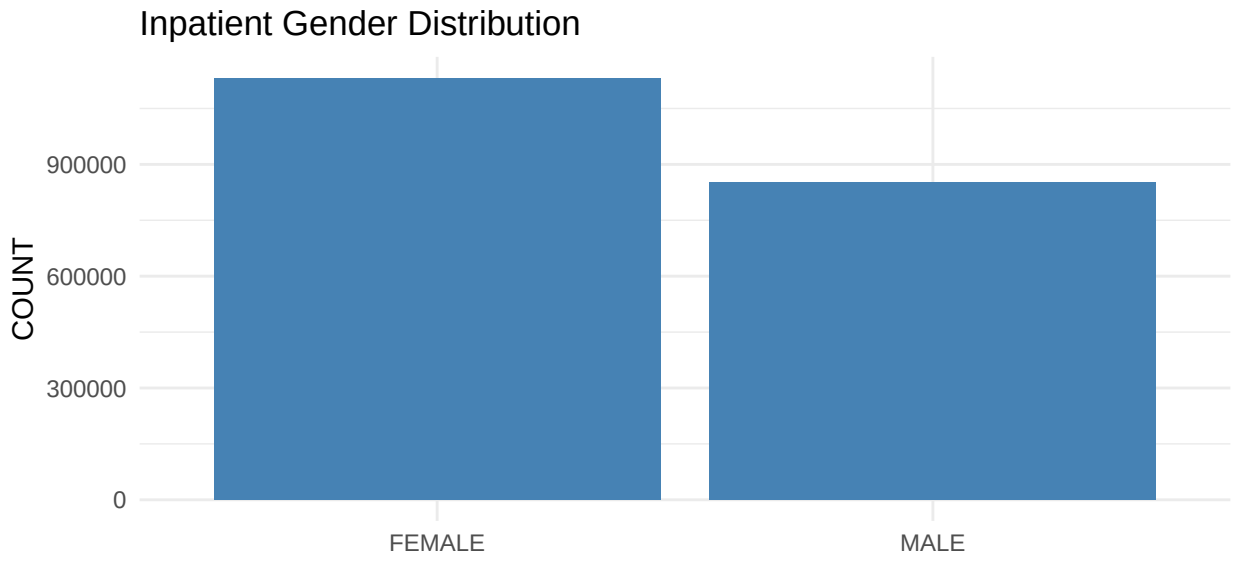
Racial Breakdown



In the inpatient cohort, there are 1664596 White persons, 198723 Black or African American persons, 121263 NA persons

RACE_CONCEPT_NAME	COUNT
White	1664596
Black or African American	198723
NA	121263

Gender Breakdown



In the inpatient cohort, there are 1131807 FEMALE persons, 852775 MALE persons

GENDER_CONCEPT_NAME	COUNT
FEMALE	1131807
MALE	852775

Condition Breakdown

To find the characterize the dataset by condition, I created a parameterized SQL that filters by the condition using the concept set, and it takes four parameters

- A list of concept ids for the condition
- X: a boolean value for an unspecified query (default True means return table)
- If X is FALSE, then query must be specified
- Y: A boolean value for a grouping the dataset
- If Y is TRUE, then a group by query must be specified

General Trends for Bipolar Disorder

Patient Spread This calculates the number of inpatient patients in the dataset:

In this dataset, there are a total of 1298679 records of bipolar disorder in the dataset, with 524063 patients

To further examine this data, we can break them down across the following axes:

- Condition
- State
- Age
- Race
- Gender
- Care setting

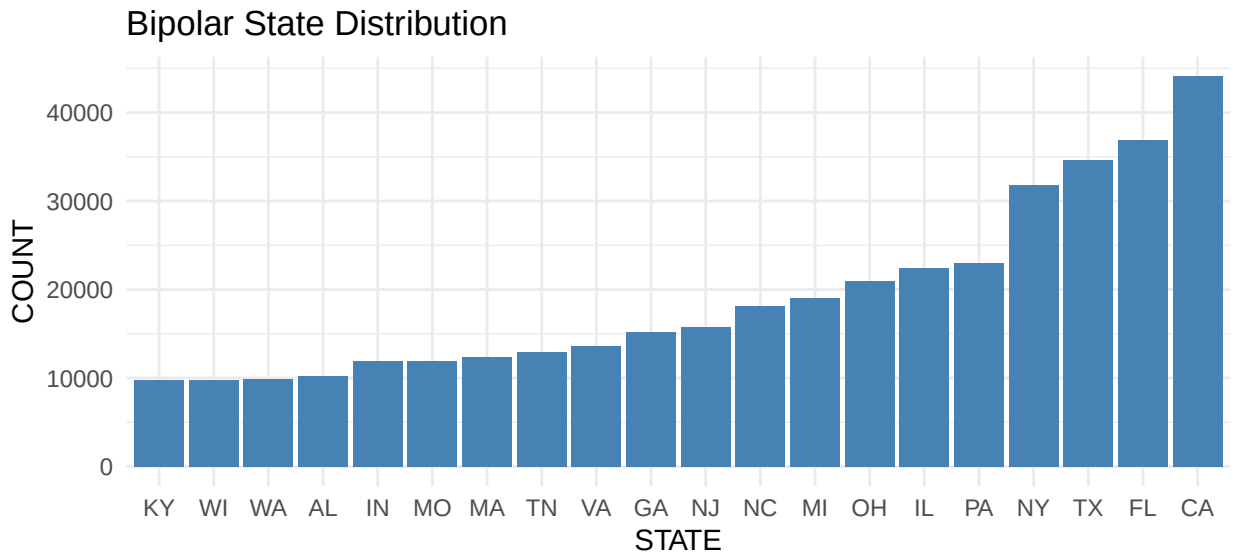
Condition Breakdown For the bipolar characterization:

1. Select the `person` table
2. Join `condition_occurrence` using `person_id`
3. Using the concept set definition for bipolar, filter the persons table for the persons with a bipolar diagnosis
4. Join the `visit_occurrence` table using `person_id`
5. Select the demographic of interest (location, race, ethnicity, or gender)
6. Group by the demographic
7. Return the person counts with bipolar for each group

Here is the breakdown of the top bipolar diagnoses:

CONDITION_CONCEPT_NAME	RECORD_COUNT	PERSON_COUNT
Bipolar disorder	274467	193040
Bipolar I disorder	88224	68356
Depressed bipolar I disorder	63863	49923
Bipolar affective disorder, current episode manic	51860	40876
Bipolar affective disorder, current episode mixed	51414	40581
Severe depressed bipolar I disorder without psychotic features	51398	38381

State Breakdown



The top ten states by person count with bipolar disorder are

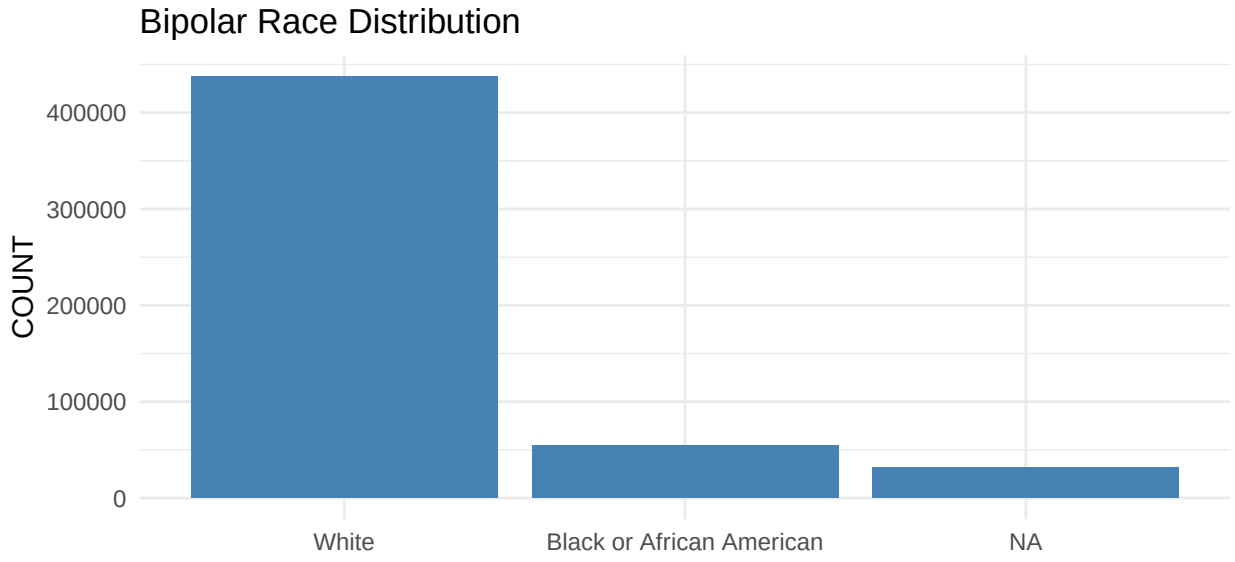
STATE	COUNT
CA	44093
FL	36922
TX	34627
NY	31864
PA	23009
IL	22406

STATE	COUNT
OH	20921
MI	19052
NC	18158
NJ	15820

Age Breakdown For age breakdowns, inpatient counts were bucketed based on their age. The age brackets were as follows: - 0-9 - 10-19 - 20-29 - 30-39 - 40-49 - 50-59 - 60-69 - 70-79 - 80-89 - $\geq$ 90

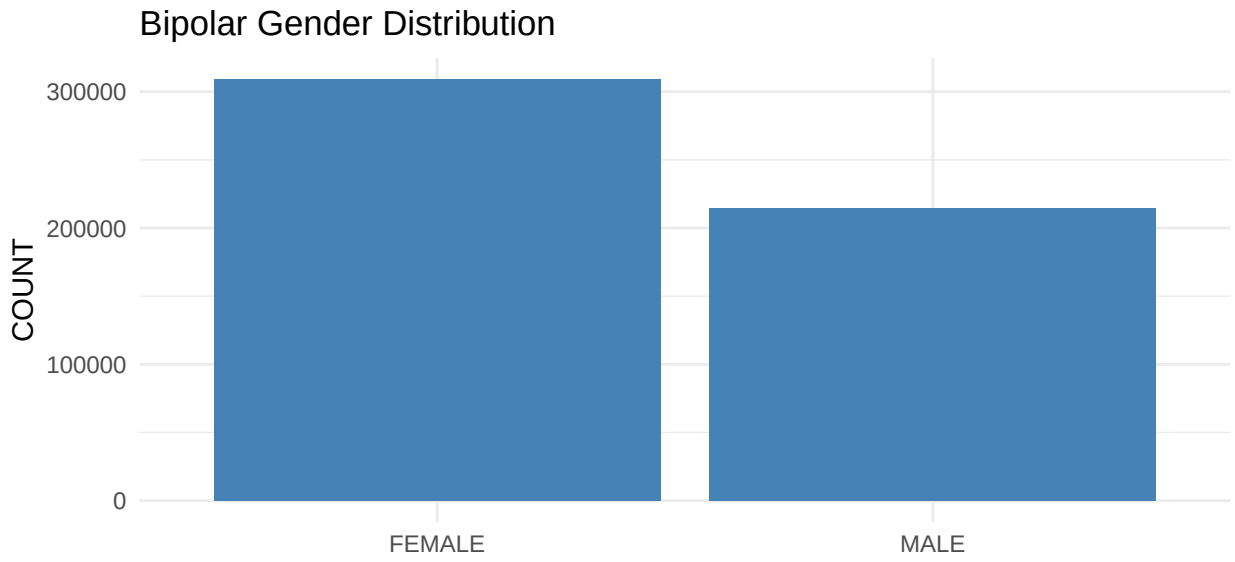
	COUNT
MIN_AGE	24
MAX_AGE	101
AGE_0_9	NA
AGE_10_TO_19	NA
AGE_20_TO_29	13646
AGE_30_TO_39	42393
AGE_40_TO_49	75261
AGE_50_TO_59	108387
AGE_60_TO_69	250005
AGE_70_TO_79	403595
AGE_80_TO_89	304294
AGE_90_AND_MORE	101098

Racial Breakdown



RACE_CONCEPT_NAME	COUNT
White	437760
Black or African American	54827
NA	31476

Gender Breakdown



GENDER_CONCEPT_NAME	COUNT
FEMALE	309470
MALE	214593

Care Site Breakdown

VISIT_CONCEPT_NAME	COUNT
Inpatient Visit	1298679

General Trends for Depression

Patient Spread This calculates the number of patients in the dataset:

In this dataset, there are a total of 1697749 records for 599108 patients. To further examine this data, we can break them down across the following axes:

- Condition
- State
- Age
- Race
- Gender
- Care setting

Condition Breakdown For the depression characterization:

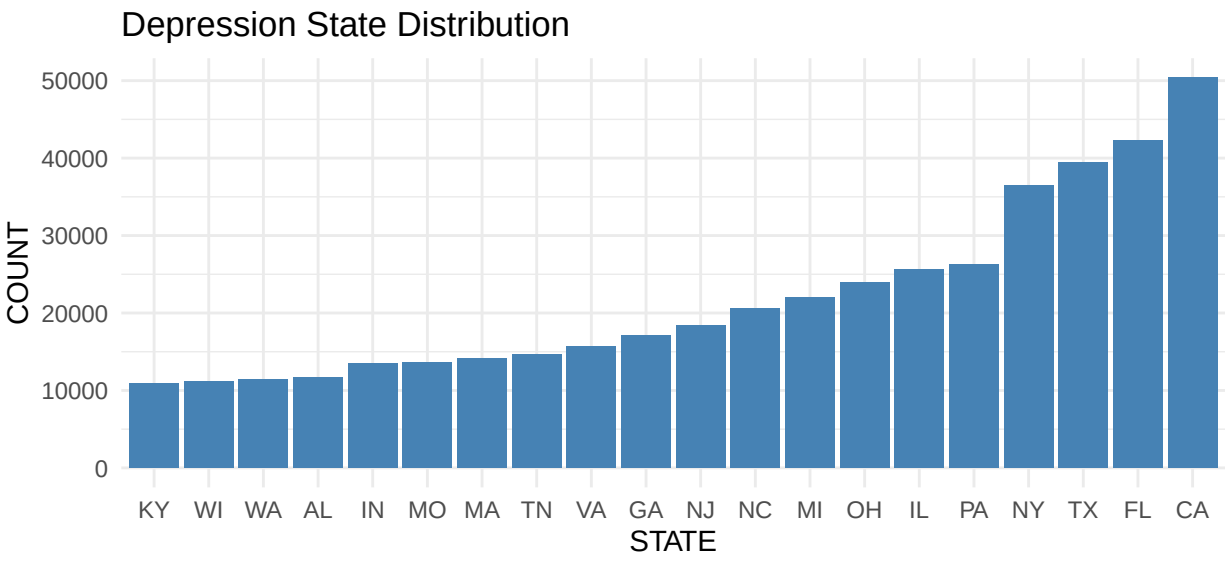
1. Select the `person` table
2. Join `condition_occurrence` using `person_id`
3. Using the concept set definition for depression, filter the persons table for the persons with a depression diagnosis
4. Join the `visit_occurrence` table using `person_id`
5. Select the demographic of interest (location, race, ethnicity, or gender)
6. Group by the demographic

7. Return the person counts with depression for each group

Here is the breakdown of the top depression diagnoses:

CONDITION_CONCEPT_NAME	RECORD_COUNT	PERSON_COUNT
Recurrent major depressive episodes, moderate	490503	243514
Dysthymia	277263	205157
Recurrent major depressive episodes	154014	104917
Single major depressive episode	142648	110259
Recurrent major depressive episodes, severe, with psychosis	104429	69300
Recurrent major depressive episodes, in full remission	75501	63448

State Breakdown



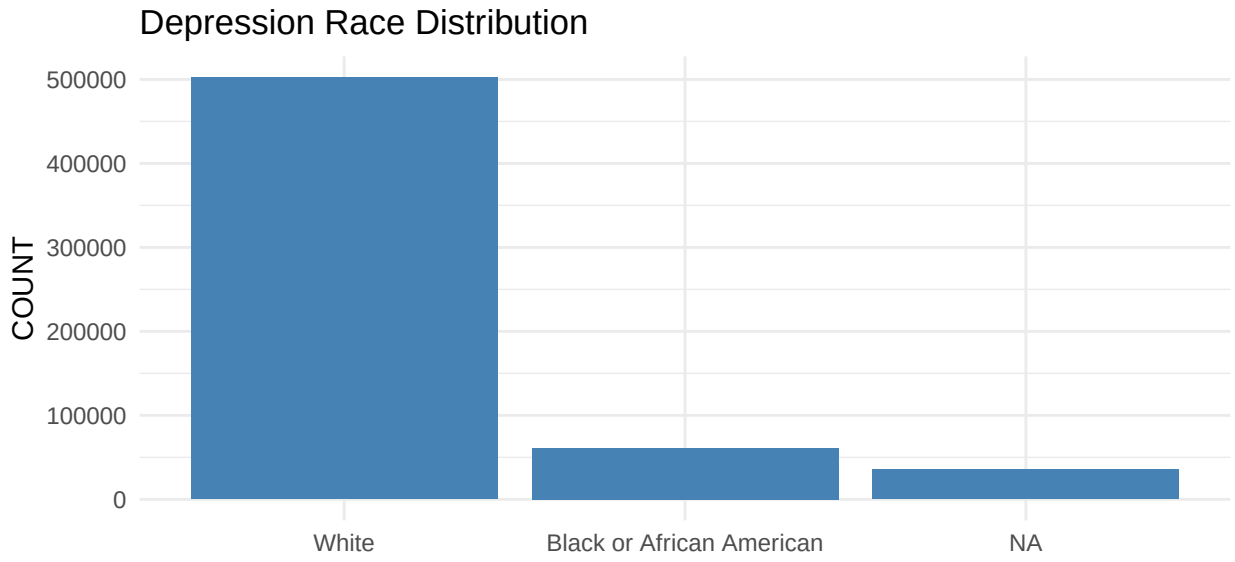
STATE	COUNT
CA	50376
FL	42229
TX	39464
NY	36449
PA	26270
IL	25572
OH	23998
MI	22051
NC	20625
NJ	18381

Age Breakdown

	COUNT
MIN_AGE	25
MAX_AGE	101
AGE_0_9	NA

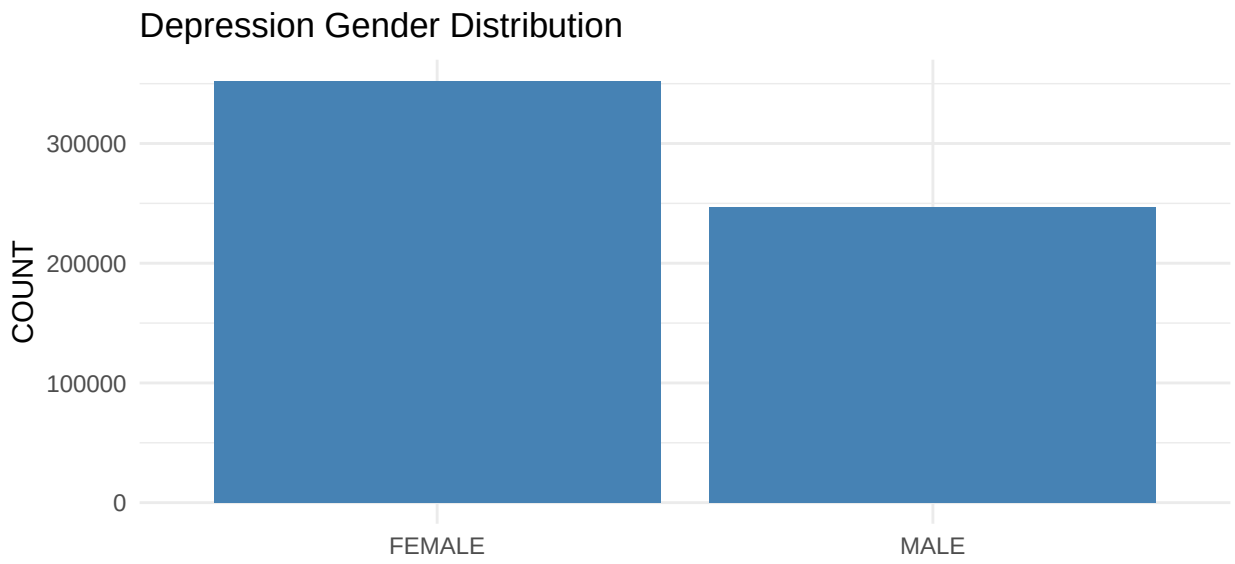
	COUNT
AGE_10_TO_19	NA
AGE_20_TO_29	14912
AGE_30_TO_39	47860
AGE_40_TO_49	88834
AGE_50_TO_59	131962
AGE_60_TO_69	328130
AGE_70_TO_79	544738
AGE_80_TO_89	408968
AGE_90_AND_MORE	132345

Racial Breakdown



RACE_CONCEPT_NAME	COUNT
White	502259
Black or African American	61460
NA	35389

Gender Breakdown



GENDER_CONCEPT_NAME	COUNT
FEMALE	352381
MALE	246727

Care Site Breakdown

VISIT_CONCEPT_NAME	COUNT
Inpatient Visit	1697749

General Trends for Suicidality

Patient Spread This calculates the number of patients in the dataset:

In this dataset, there are a total of 29342 records for 27914 patients.

To further examine this data, we can break them down across the following axes:

- Condition
- State
- Age
- Race
- Gender
- Care setting

Condition Breakdown For the suicidality characterization:

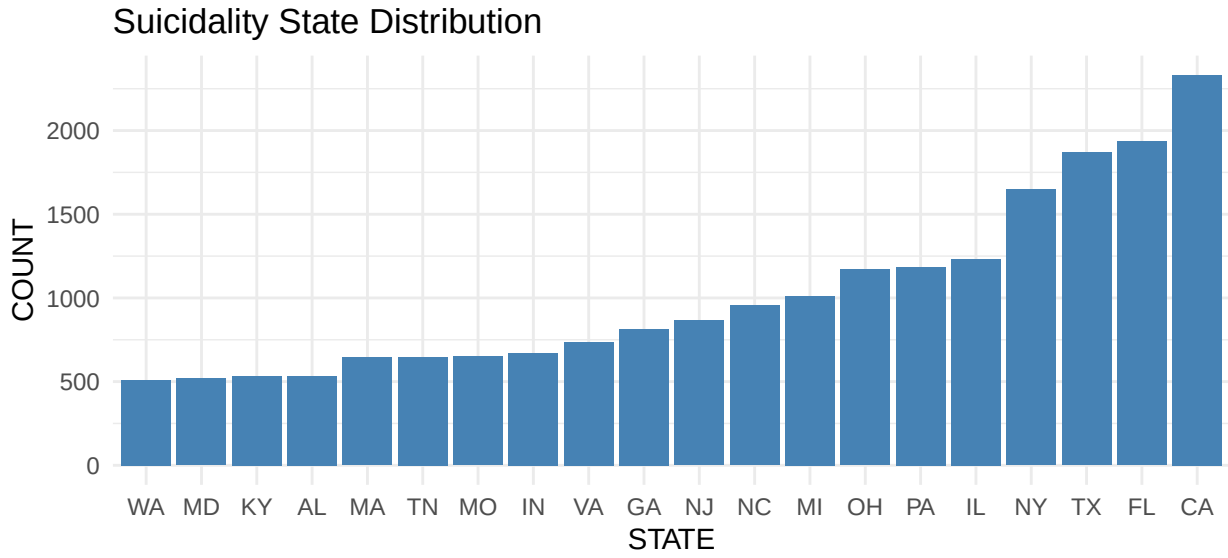
1. Select the `person` table
2. Join `condition_occurrence` using `person_id`
3. Using the concept set definition for suicidality, filter the persons table for the persons with a suicidality diagnosis
4. Join the `visit_occurrence` table using `person_id`
5. Select the demographic of interest (location, race, ethnicity, or gender)

6. Group by the demographic
7. Return the person counts with suicidality for each group

Here is the breakdown of the top suicidality diagnoses:

CONDITION_CONCEPT_NAME	RECORD_COUNT	PERSON_COUNT
Suicidal thoughts	29342	27914

State Breakdown



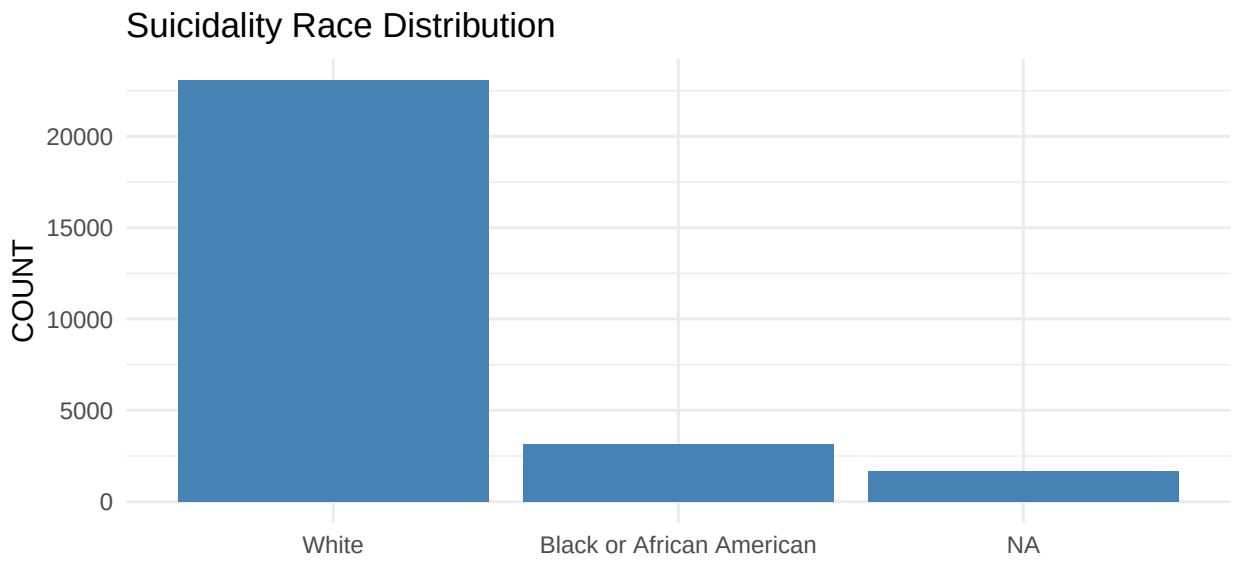
STATE	COUNT
CA	2331
FL	1934
TX	1869
NY	1649
IL	1233
PA	1185
OH	1170
MI	1011
NC	955
NJ	866

Age Breakdown

	COUNT
MIN_AGE	25
MAX_AGE	101
AGE_0_9	NA
AGE_10_TO_19	NA
AGE_20_TO_29	297
AGE_30_TO_39	940
AGE_40_TO_49	1664

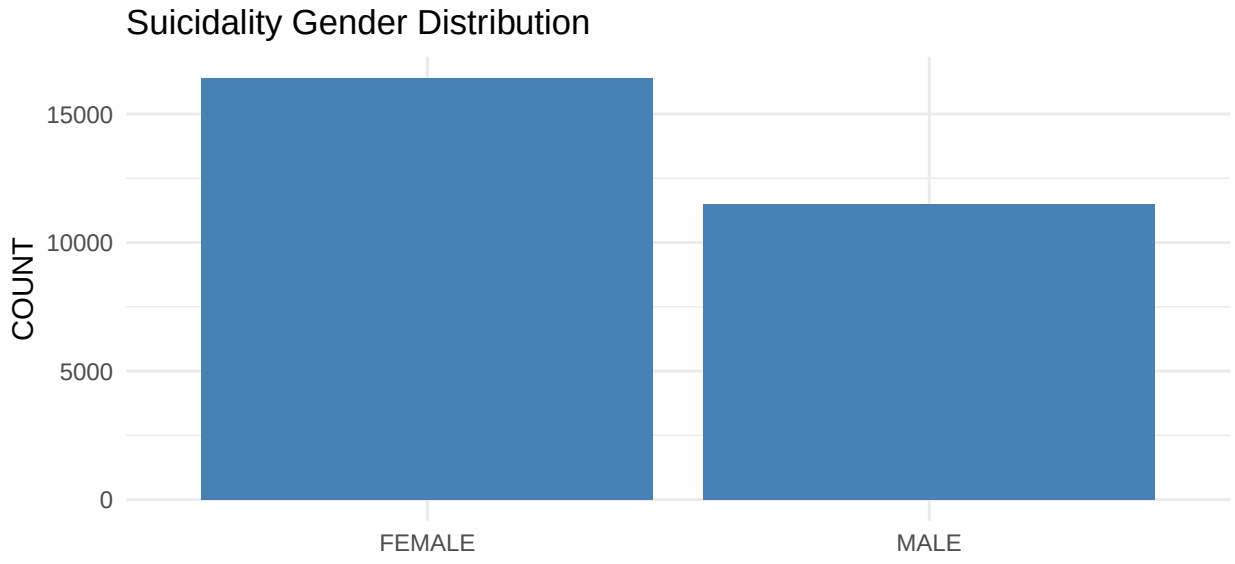
	COUNT
AGE_50_TO_59	2450
AGE_60_TO_69	5543
AGE_70_TO_79	9175
AGE_80_TO_89	6907
AGE_90_AND_MORE	2366

Racial Breakdown



RACE_CONCEPT_NAME	COUNT
White	23094
Black or African American	3129
NA	1691

Gender Breakdown



GENDER_CONCEPT_NAME	COUNT
FEMALE	16406
MALE	11508

Care Site Breakdown

VISIT_CONCEPT_NAME	COUNT
Inpatient Visit	29342