Performance Assessment D211 – Advanced Data Acquisition

Part 1

Doug Haunsperger

College of IT, Western Governors University

February 17, 2024

## A1. Data Sets

For this task, I combined the WGU medical database in the lab PostgreSQL installation with the 2017-2018 National Health and Nutrition Examination Survey (NHANES) performed by the Centers for Disease Control. I used the Demographics data from NHANES found at <https://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Demographics&Cycle=2017-2018> and combined it with the Vitamin D data set from NHANES found at <https://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Laboratory&Cycle=2017-2018> [scroll down to ‘Vitamin D’].

I prepared the data from these datasets using Excel to remove unneeded columns from the CSV file. The reduced CSV files have been uploaded with my submission as ‘NHANES\_demog.csv’ and ‘VitD.csv’.

## A2. Dashboard Installation

Step 1. Acquire the file ‘D211.twbx’ from my task submission. Alternately, download it from this link: [D211.twbx](https://westerngovernorsuniversity-my.sharepoint.com/:u:/g/personal/dhaunsp_wgu_edu/EYMF9gvF73VPqMVJMH2xXnEBVtYXQfgU5LaGikBTE_r6Lg?e=bKFMuj). Save it to your local workspace where Tableau Desktop can access it.

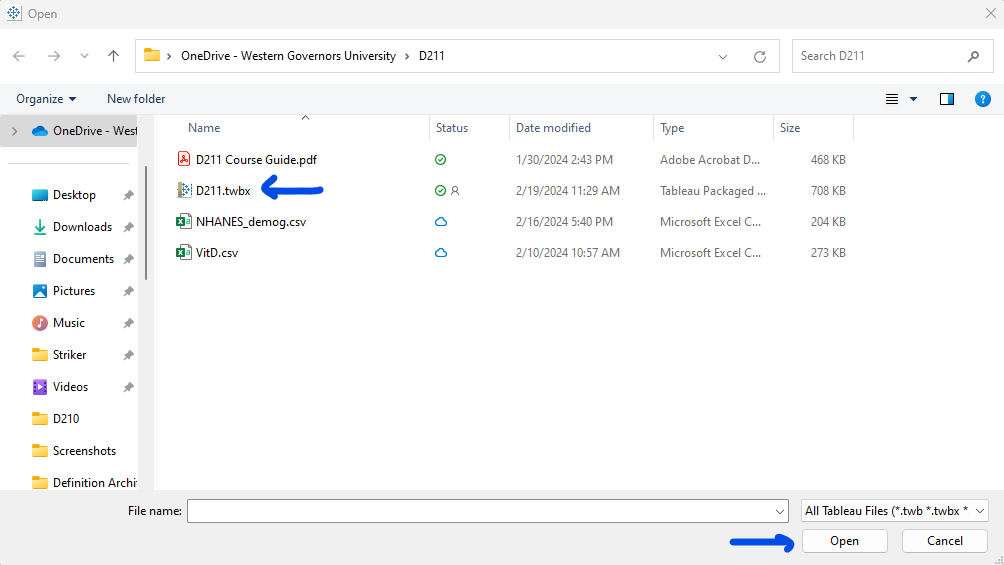
Step 2. Open Tableau Desktop

Step 3. Click ‘File’, then ‘Open’

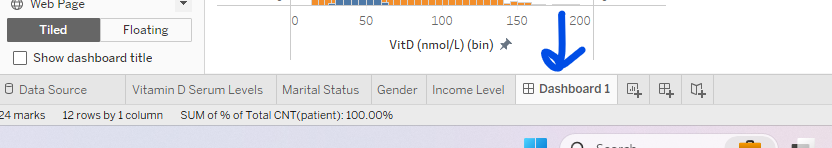
A screenshot of a computer

Description automatically generated

Step 4. From the Open File dialog, navigate to where the D211.twbx is saved and select it, then click ‘Open’.



Step 5. Select the ‘Dashboard 1’ page along the bottom.



## A3. Dashboard Navigation

On the left, this dashboard shows our patient’s Vitamin D levels (blue) compared to a national sample from the NHANES survey (orange) in histogram format. In the middle and on the right, there are demographic comparisons between our patient data and the NHANES data for marital status, gender, and income level. Each of these graphs can be used to filter the others by selecting a value. The income bins were chosen to match those in the NHANES data. By default, null Marital status is filtered out, though note that this excludes a large proportion (3685 / 9254) of the NHANES respondents, which had missing data for this variable (see CDC NHANES DEMO\_J data dictionary file).

## A4. SQL Code

To merge the NHANES data into our patient database, I first created two tables to hold the CSV import, using the following code:

CREATE TABLE nhanes\_demo

(

patient\_num varchar(10) PRIMARY KEY,

gender varchar(10),

age int,

marital varchar(30),

hh\_size int,

children\_under5 int,

children\_6to18 int,

income int

);

CREATE TABLE vitd

(

patient\_num varchar(10) PRIMARY KEY REFERENCES nhanes\_demo,

tot\_vitd numeric(5,2),

tot\_vitd\_comment int,

vitd2 numeric(5,2),

vitd2\_comment int,

vitd3 numeric(5,2),

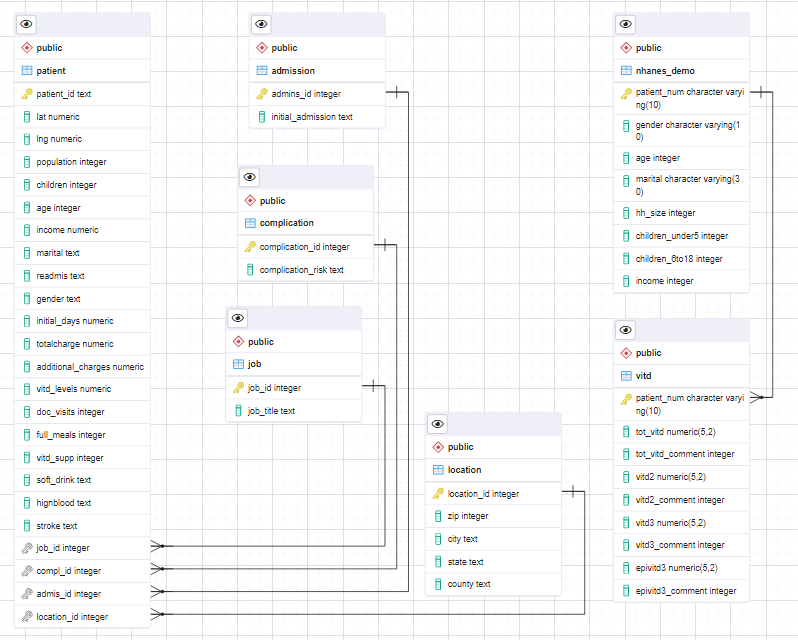
vitd3\_comment int,

epivitd3 numeric(5,2),

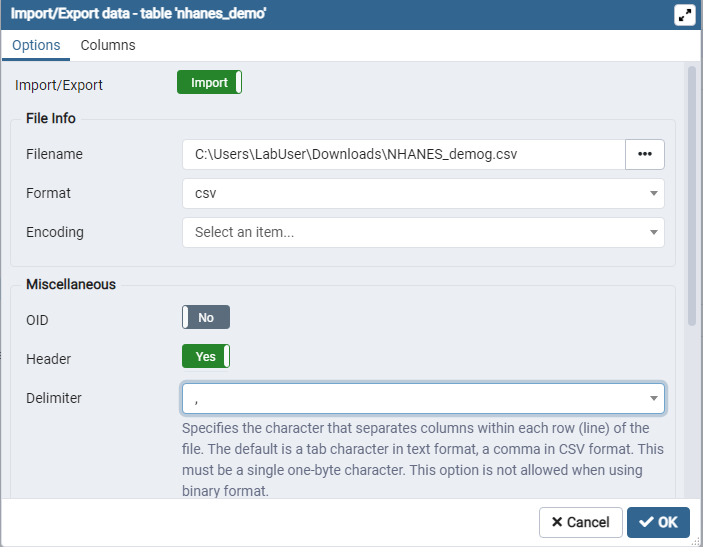
epivitd3\_comment int

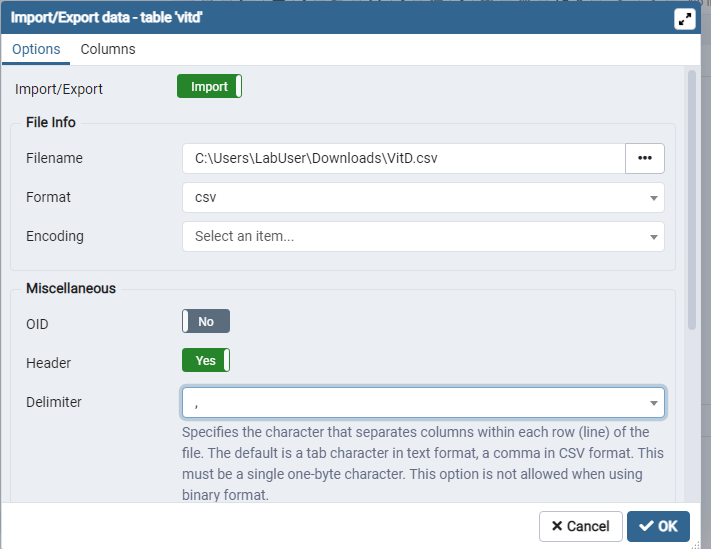
);

At this point, the ERD of the database looks like this:

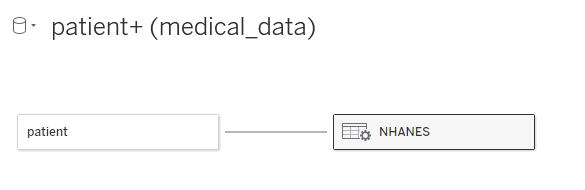


Next, I used the pgAdmin import on each new table functionality as follows:





Once the data had been imported into our medical\_data database, I used the custom SQL functionality within Tableau to join the new tables together and relate them with the patient table. The Custom SQL code also takes care of reexpressing the numerical categorical variables with meaningful text expressions.



The Custom SQL query I used was:

SELECT d.patient\_num,

v.tot\_vitd,

CASE WHEN d.gender = '1' THEN 'Male'

WHEN d.gender = '2' THEN 'Female'

ELSE 'NULL' END AS gender,

d.age,

CASE WHEN d.marital = '1' THEN 'Married'

WHEN d.marital = '2' THEN 'Widowed'

WHEN d.marital = '3' THEN 'Divorced'

WHEN d.marital = '4' THEN 'Separated'

WHEN d.marital = '5' THEN 'Never Married'

WHEN d.marital = '6' THEN 'Partnership'

ELSE 'NULL' END AS marital,

d.hh\_size,

d.children\_under5,

d.children\_6to18,

CASE WHEN d.income = '1' THEN '< $5k'

WHEN d.income = '2' THEN '$5k - $10k'

WHEN d.income = '3' THEN '$10k - $15k'

WHEN d.income = '4' THEN '$15k - $20k'

WHEN d.income = '5' THEN '$20k - $25k'

WHEN d.income = '6' THEN '$25k - $35k'

WHEN d.income = '7' THEN '$35k - $45k'

WHEN d.income = '8' THEN '$45k - $55k'

WHEN d.income = '9' THEN '$55k - $65k'

WHEN d.income = '10' THEN '$65k - $75k'

WHEN d.income = '14' THEN '$75k - $100k'

WHEN d.income = '15' THEN '> $100k'

ELSE 'NULL' END AS income

FROM public.nhanes\_demo AS d

INNER JOIN public.vitd AS v

ON d.patient\_num = v.patient\_num