A1: Data Sets

The dashboard containing my data analysis is attached to this submission. I would've published it to Tableau Public, but that appears to not be an option for this class. This data analysis is identical to the one I performed in D210, and thus it makes use of materials generated for that project, though it accomplishes this through a different means to satisfy the requirements of this particular project for D211. Again, two datasets are used for this analysis:

- The WGU medical dataset, provided by WGU for several of the classes in the MSDA program
- The 2013-14 National Health and Nutrition Examination Survey (NHANES), provided by the United States Centers of Disease Control & Prevention

A2: Dashboard Installation

The dashboard is provided in a .twbx format, which can be opened with Tableau Reader. Because it is not clear to me what evaluators are or are not able to do with the virtual machines, I have included instructions using either the D211.zip file submitted alongside this report (if the evaluator can easily transfer this file to the virtual machine) as well as a link to an identical version of D211.zip that can be downloaded via Google Drive to the virtual machine. Which process is used is up to the evaluator.

In order to do this on the Labs on Demand virtual machine, the evaluator must do the following:

- **IF POSSIBLE:** Take the D211.zip file included alongside this submission, and open File Explorer in the virtual machine, navigating to 'C:\Users\Public\Downloads' and pasting the provided D211.zip file to this folder.
- **ALTERNATIVELY:** CTRL+click on this link to download D211.zip from my Google Drive. Click the download icon in the top right of the screen to download the file. Open File Explorer and navigate to the Downloads folder where this file was saved. Right Click on D211.zip to cut the file, and then navigate to 'C:\Users\Public\Downloads', where you'll paste the file.
- Right click on D211.zip and select "Extract All…". Extract the files to 'C:\Users\Public\Downloads'.
- On the desktop, double click on pgAdmin to open it. In the left side navigation pane, navigate down to the "medical_data" database (one level below "Databases").
- Right click on the "medical_data" database and select Query Tool.
- When the Query Tool panel opens on the right side of the pgAdmin window, select the "Open File" button (second button from left).
- In the Select File dialog box, navigate to 'C:\Users\Public\Downloads' and select the SQL Commands.txt file.
    - It may be necessary to change the bottom-right "Format" drop-down menu to reflect "All Files", rather than just SQL files.
- Once the Query Tool has imported the contents of the SQL Commands.txt file, hit the Play button in the top right corner to execute the query.
- Once the query is complete (this may take several seconds), close pgAdmin and return to the desktop.
- Double click on Tableau 2021.4 to open Tableau Desktop.

- At the top left, click File > Open and navigate to 'C:\Users\Public\Downloads'. Select the 20230103 Dashboard.twbx file to open it.
- When prompted to sign in to complete the connection to a server, use the username "postgres" and the password "Passw0rd!"

At this point, the dashboard is "installed" and may be interacted with.

A3: Dashboard Usage

After opening the provided dashboard, I suggest navigating to the "PRESENTATION" story tab at the bottom of the screen. This can be made into a full screen presentation by using the projector screen button at the end of the ribbon at the top of the screen. The two boxes at the top of the story can be thought of as "tabs" like you may find in your web browser.

The "Dashboard 1" tab is an interactable dashboard that provides visualizations about the WGU Hospital System's patient demographics, such as patient age, number of children, or gender. This data can be filtered in multiple ways. In the bottom right of this dashboard are filters for both age and gender. By adjusting the age slider, the user can change the range of ages reported. For example, setting the slider to 40 – 50 will show the distribution of patients aged 40 – 50, the number of children for patients aged 40 – 50, and the genders of patients aged 40 – 50. Similarly, clicking on any gender in the gender filter (which also functions as a lend) will update all of the visualizations to reflect only the selected gender. Alternatively, clicking on a portion of the visualization will filter in the same way. For example, clicking on the "male" portion of the pie charts visualizing patient gender will update all data in the dashboard to reflect only male patients, similar to using the filter in the bottom right. Clicking on the "male" portion of the pie chart (or on "Male" in the filter) will "release" this filter and return the visualizations to the original view.

The "Dashboard 2" tab is another interactable dashboard, providing visualizations and data about selected serious medical conditions observed in both the CDC NHANES dataset and the WGU dataset. The rate of diagnosis for each medical condition is provided as a key performance indicator (KPI), demonstrating how many patients amongst both the WGU and CDC datasets have this condition. This is further broken down in the heatmap to the right of each KPI, demonstrating the raw count of patients with that diagnosis by age group, for both the CDC and WGU datasets. For reference, the WGU dataset has approximately 150% more observations than the CDC's NHANES data (10,000 rows vs ~6,100). The same filters for age and gender are provided in the top right of this dashboard, and they are used in the same fashion as they were in the previous dashboard.

A4: SQL Code

```
/* Before dealing with external data, I need to clean up the medical data provided by WGU to be
in the format that I want it to be. That's going to primarily include editing certain Yes/No string
fields to instead be True/False booleans. */

-- Modify highblood pressure to boolean, fix 'hignblood' typo in column name
```

```
ALTER TABLE public.patient ALTER hignblood TYPE bool USING CASE WHEN hignblood
='No' THEN FALSE ELSE TRUE END;
ALTER TABLE public.patient ALTER COLUMN hignblood SET DEFAULT FALSE;
ALTER TABLE public.patient RENAME COLUMN hignblood TO high_bp;

-- Modify prior stroke to boolean
ALTER TABLE public.patient ALTER stroke TYPE bool USING CASE WHEN stroke ='No'
THEN FALSE ELSE TRUE END;
ALTER TABLE public.patient ALTER COLUMN stroke SET DEFAULT FALSE;

-- Modify prior arthritis to boolean
ALTER TABLE public.servicesaddon ALTER arthritis TYPE bool USING CASE WHEN
arthritis ='No' THEN FALSE ELSE TRUE END;
ALTER TABLE public.servicesaddon ALTER COLUMN arthritis SET DEFAULT FALSE;

-- Modify prior diabetes to boolean
ALTER TABLE public.servicesaddon ALTER diabetes TYPE bool USING CASE WHEN
diabetes ='No' THEN FALSE ELSE TRUE END;
ALTER TABLE public.servicesaddon ALTER COLUMN diabetes SET DEFAULT FALSE;

-- Modify prior hyperlipidemia to boolean
ALTER TABLE public.servicesaddon ALTER hyperlipidemia TYPE bool USING CASE WHEN
hyperlipidemia ='No' THEN FALSE ELSE TRUE END;
ALTER TABLE public.servicesaddon ALTER COLUMN hyperlipidemia SET DEFAULT
FALSE;

-- Modify prior asthma to boolean
ALTER TABLE public.servicesaddon ALTER asthma TYPE bool USING CASE WHEN asthma
='No' THEN FALSE ELSE TRUE END;
ALTER TABLE public.servicesaddon ALTER COLUMN asthma SET DEFAULT FALSE;

-- Modify prior overweight to boolean
ALTER TABLE public.servicesaddon ALTER overweight TYPE bool USING CASE WHEN
overweight ='No' THEN FALSE ELSE TRUE END;
ALTER TABLE public.servicesaddon ALTER COLUMN overweight SET DEFAULT FALSE;

/* Because my research involves comparing the WGU population to another, I need to label the
WGU data as having come from the WGU dataset */
ALTER TABLE public.patient ADD COLUMN source text;
UPDATE public.patient SET source = 'WGU';

/* All prior usage of this dataset indicates that Gender 'Prefer not to answer' should be construed
as non-binary. Updating this accordingly. */
UPDATE public.patient SET gender = REPLACE(gender, 'Prefer not to answer', 'Non-Binary');

/* Comparison dataset top-codes age at 80, while the WGU dataset does not. To make a clean
comparison, I am top-coding the WGU Age column at 80. */
UPDATE public.patient SET age = 80 WHERE age > 80;

/* The NHANES data consists of 6 different CSV files and hundreds of columns, of which I'm
only interested in a handful. I'm going to ignore the 4 CSV files that I'm not using, as this would
```

be hundreds of lines of code for no particular purpose. Suffice it to say, though, I could use this same process to import those CSVs, if needed. I do, unfortunately, have to import the entirety of these CSVs, and then isolate only the columns that I actually want for my visualization. I'm leaving the original CDC column names for the columns that I don't care about, and I'm using human-readable names for the columns that I'm interested in. */

```
CREATE TABLE public.nhanes_demographic
(
	seqn text, SDDSRVYR text, RIDSTATR text, gender text, age numeric, RIDAGEMN
text, RIDRETH1 text, RIDRETH3 text, RIDEXMON text, RIDEXAGM text, DMQMILIZ text,
DMQADFC text, DMDBORN4 text, DMDCITZN text, DMDYRSUS text, DMDEDUC3 text,
DMDEDUC2 text, DMDMARTL text, RIDEXPRG text, SIALANG text, SIAPROXY text,
SIAINTRP text, FIALANG text, FIAPROXY text, FIAINTRP text, MIALANG text,
MIAPROXY text, MIAINTRP text, AIALANGA text, DMDHHSIZ text, DMDFMSIZ text,
small_children numeric, older_children numeric, DMDHHSZE text, DMDHRGND text,
DMDHRAGE text, DMDHRBR4 text, DMDHREDU text, DMDHRMAR text, DMDHSEDU
text, WTINT2YR text, WTMEC2YR text, SDMVPSU text, SDMVSTRA text, INDHHIN2 text,
INDFMIN2 text, INDFMPIR text,
	CONSTRAINT participant_id PRIMARY KEY (seqn)
);

ALTER TABLE public.nhanes_demographic
	OWNER TO postgres;

-- Import the demographic CSV into the demographic table
COPY nhanes_demographic
FROM 'C:\Users\Public\Downloads\demographic.csv'
DELIMITER ','
CSV HEADER;

/* Now do the same process for the questionnaire data... all 963 columns of it. Seems like maybe
this project could specify some allowance for this sort of thing! */

CREATE TABLE public.nhanes_questionnaire
(
	seqn text,	ACD011A text,	ACD011B text, ACD011C text, ACD040 text,
	ACD110 text,	ALQ101 text,	ALQ110 text,	ALQ120Q text, ALQ120U text,
	ALQ130 text,	ALQ141Q text, ALQ141U text, ALQ151 text,	ALQ160 text,
	high_bp text,	BPQ030 text,	BPD035 text,	BPQ040A text, BPQ050A text,
	BPQ056 text,	BPD058 text,	BPQ059 text,	hyperlipidemia text,	BPQ060 text,
	BPQ070 text,	BPQ090D text, BPQ100D text, CBD070 text,	CBD090 text,
	CBD110 text,	CBD120 text,	CBD130 text,	HSD010 text,	HSQ500 text,
	HSQ510 text,	HSQ520 text,	HSQ571 text,	HSQ580 text,	HSQ590 text,
	HSAQUEX text,	CSQ010 text,	CSQ020 text,	CSQ030 text,	CSQ040 text,
	CSQ060 text,	CSQ070 text,	CSQ080 text,	CSQ090A text, CSQ090B text,
	CSQ090C text, CSQ090D text, CSQ100 text,	CSQ110 text,	CSQ120A text,
	CSQ120B text, CSQ120C text, CSQ120D text, CSQ120E text, CSQ120F text,
	CSQ120G text, CSQ120H text, CSQ140 text,	CSQ160 text,	CSQ170 text,
	CSQ180 text,	CSQ190 text,	CSQ200 text,	CSQ202 text,	CSQ204 text,
	CSQ210 text,	CSQ220 text,	CSQ240 text,	CSQ250 text,	CSQ260 text,
	AUQ136 text,	AUQ138 text,	CDQ001 text,	CDQ002 text,	CDQ003 text,
```

CDQ004 text, CDQ005 text, CDQ006 text, CDQ009A text, CDQ009B text, CDQ009C text, CDQ009D text, CDQ009E text, CDQ009F text, CDQ009G text, CDQ009H text, CDQ008 text, CDQ010 text, diabetes text, DID040 text, DIQ160 text, DIQ170 text, DIQ172 text, DIQ175A text, DIQ175B text, DIQ175C text, DIQ175D text, DIQ175E text, DIQ175F text, DIQ175G text, DIQ175H text, DIQ175I text, DIQ175J text, DIQ175K text, DIQ175L text, DIQ175M text, DIQ175N text, DIQ175O text, DIQ175P text, DIQ175Q text, DIQ175R text, DIQ175S text, DIQ175T text, DIQ175U text, DIQ175V text, DIQ175W text, DIQ175X text, DIQ180 text, DIQ050 text, DID060 text, DIQ060U text, DIQ070 text, DIQ230 text, DIQ240 text, DID250 text, DID260 text, DIQ260U text, DIQ275 text, DIQ280 text, DIQ291 text, DIQ300S text, DIQ300D text, DID310S text, DID310D text, DID320 text, DID330 text, DID341 text, DID350 text, DIQ350U text, DIQ360 text, DIQ080 text, DBQ010 text, DBD030 text, DBD041 text, DBD050 text, DBD055 text, DBD061 text, DBQ073A text, DBQ073B text, DBQ073C text, DBQ073D text, DBQ073E text, DBQ073U text, DBQ700 text, DBQ197 text, DBQ223A text, DBQ223B text, DBQ223C text, DBQ223D text, DBQ223E text, DBQ223U text, DBQ229 text, DBQ235A text, DBQ235B text, DBQ235C text, DBQ301 text, DBQ330 text, DBQ360 text, DBQ370 text, DBD381 text, DBQ390 text, DBQ400 text, DBD411 text, DBQ421 text, DBQ424 text, DBD895 text, DBD900 text, DBD905 text, DBD910 text, CBQ596 text, CBQ606 text, CBQ611 text, CBQ505 text, CBQ535 text, CBQ540 text, CBQ545 text, CBQ550 text, CBQ552 text, CBQ580 text, CBQ585 text, CBQ590 text, DED031 text, DEQ034A text, DEQ034C text, DEQ034D text, DEQ038G text, DEQ038Q text, DED120 text, DED125 text, DLQ010 text, DLQ020 text, DLQ040 text, DLQ050 text, DLQ060 text, DLQ080 text, DPQ010 text, DPQ020 text, DPQ030 text, DPQ040 text, DPQ050 text, DPQ060 text, DPQ070 text, DPQ080 text, DPQ090 text, DPQ100 text, DUQ200 text, DUQ210 text, DUQ211 text, DUQ213 text, DUQ215Q text, DUQ215U text, DUQ217 text, DUQ219 text, DUQ220Q text, DUQ220U text, DUQ230 text, DUQ240 text, DUQ250 text, DUQ260 text, DUQ270Q text, DUQ270U text, DUQ272 text, DUQ280 text, DUQ290 text, DUQ300 text, DUQ310Q text, DUQ310U text, DUQ320 text, DUQ330 text, DUQ340 text, DUQ350Q text, DUQ350U text, DUQ352 text, DUQ360 text, DUQ370 text, DUQ380A text, DUQ380B text, DUQ380C text, DUQ380D text, DUQ380E text, DUQ390 text, DUQ400Q text, DUQ400U text, DUQ410 text, DUQ420 text, DUQ430 text, ECD010 text, ECQ020 text, ECD070A text, ECD070B text, ECQ080 text, ECQ090 text, WHQ030E text, MCQ080E text, ECQ150 text, FSD032A text, FSD032B text, FSD032C text, FSD041 text, FSD052 text, FSD061 text, FSD071 text, FSD081 text, FSD092 text, FSD102 text, FSD032D text, FSD032E text, FSD032F text, FSD111 text, FSD122 text, FSD132 text, FSD141 text, FSD146 text, FSDHH text, FSDAD text, FSDCH text, FSD151 text, FSQ165 text, FSQ012 text, FSD012N text, FSD230 text, FSD225 text, FSQ235 text, FSQ162 text, FSD650ZC text, FSD660ZC text, FSD675 text, FSD680 text, FSD670ZC text, FSQ690 text, FSQ695 text, FSD650ZW text, FSD660ZW text, FSD670ZW text, HEQ010 text, HEQ020 text, HEQ030 text, HEQ040 text, HIQ011 text, HIQ031A text, HIQ031B text, HIQ031C text, HIQ031D text, HIQ031E text, HIQ031F text, HIQ031G text, HIQ031H text, HIQ031I text, HIQ031J text, HIQ031AA text, HIQ260 text,

```
    HIQ105 text,  HIQ270 text,  HIQ210 text,  HOD050 text,  HOQ065 text,
    HUQ010 text,  HUQ020 text,  HUQ030 text,  HUQ041 text,  HUQ051 text,
    HUQ061 text,  HUQ071 text,  HUD080 text,  HUQ090 text,  IMQ011 text,
    IMQ020 text,  IMQ040 text,  IMQ070 text,  IMQ080 text,  IMQ090 text,
    IMQ045 text,  INQ020 text,  INQ012 text,  INQ030 text,  INQ060 text,
    INQ080 text,  INQ090 text,  INQ132 text,  INQ140 text,  INQ150 text,
    IND235 text,  INDFMMPI text,  INDFMMPC text,  INQ244 text,
    IND247 text,  MCQ010 text,  MCQ025 text,  asthma text,  MCQ040 text,
    MCQ050 text,  AGQ030 text,  MCQ053 text,  MCQ070 text,  MCQ075 text,
    overweight text,  MCQ082 text,  MCQ084 text,  MCQ086 text,  MCQ092 text,
    MCD093 text,  MCQ149 text,  MCQ151 text,  arthritis text,  MCQ180A text,
    MCQ195 text,  MCQ160N text,  MCQ180N text,  MCQ160B text,
    MCQ180B text,  MCQ160C text,  MCQ180C text,  MCQ160D text,
    MCQ180D text,  MCQ160E text,  MCQ180E text,  stroke text,
    MCQ180F text,  MCQ160G text,  MCQ180G text,  MCQ160M
text,  MCQ170M text,  MCQ180M text,  MCQ160K text,  MCQ170K text,
    MCQ180K text,  MCQ160L text,  MCQ170L text,  MCQ180L text,
    MCQ160O text,  MCQ203 text,  MCQ206 text,  MCQ220 text,  MCQ230A text,
    MCQ230B text,  MCQ230C text,  MCQ230D text,  MCQ240A text,
    MCQ240AA text,  MCQ240B text,  MCQ240BB text,  MCQ240C text,
    MCQ240CC text,  MCQ240D text,  MCQ240DD text,  MCQ240DK
text,  MCQ240E text,  MCQ240F text,  MCQ240G text,  MCQ240H text,
    MCQ240I text,  MCQ240J text,  MCQ240K text,  MCQ240L text,
    MCQ240M text,  MCQ240N text,  MCQ240O text,  MCQ240P text,
    MCQ240Q text,  MCQ240R text,  MCQ240S text,  MCQ240T text,
    MCQ240U text,  MCQ240V text,  MCQ240W text,  MCQ240X text,
    MCQ240Y text,  MCQ240Z text,  MCQ300A text,  MCQ300B text,
    MCQ300C text,  MCQ365A text,  MCQ365B text,  MCQ365C text,
    MCQ365D text,  MCQ370A text,  MCQ370B text,  MCQ370C text,
    MCQ370D text,  MCQ380 text,  OCD150 text,  OCQ180 text,  OCQ210 text,
    OCQ260 text,  OCD270 text,  OCQ380 text,  OCD390G text,  OCD395 text,
    OHQ030 text,  OHQ033 text,  OHQ770 text,  OHQ780A text,  OHQ780B text,
    OHQ780C text,  OHQ780D text,  OHQ780E text,  OHQ780F text,
    OHQ780G text,  OHQ780H text,  OHQ780I text,  OHQ780J text,
    OHQ780K text,  OHQ555G text,  OHQ555Q text,  OHQ555U text,
    OHQ560G text,  OHQ560Q text,  OHQ560U text,  OHQ565 text,
    OHQ570Q text,  OHQ570U text,  OHQ575G text,  OHQ575Q text,
    OHQ575U text,  OHQ580 text,  OHQ585Q text,  OHQ585U text,
    OHQ590G text,  OHQ590Q text,  OHQ590U text,  OHQ610 text,
    OHQ612 text,  OHQ614 text,  OHQ620 text,  OHQ640 text,  OHQ680 text,
    OHQ835 text,  OHQ845 text,  OHQ848G text,  OHQ848Q text,
    OHQ849 text,  OHQ850 text,  OHQ855 text,  OHQ860 text,  OHQ865 text,
    OHQ870 text,  OHQ875 text,  OHQ880 text,  OHQ885 text,  OHQ895 text,
    OHQ900 text,  OSQ010A text,  OSQ010B text,  OSQ010C text,  OSQ020A text,
    OSQ020B text,  OSQ020C text,  OSD030AA text,  OSQ040AA text,
    OSD050AA text,  OSD030AB text,  OSQ040AB text,  OSD050AB
text,  OSD030AC text,  OSQ040AC text,  OSD050AC text,  OSD030BA
text,  OSQ040BA text,  OSD050BA text,  OSD030BB text,  OSQ040BB
text,  OSD050BB text,  OSD030BC text,  OSQ040BC text,  OSD050BC
text,  OSD030BD text,  OSQ040BD text,  OSD050BD text,  OSD030BE
text,  OSQ040BE text,  OSD030BF text,  OSQ040BF text,  OSD030BG
```

text, OSQ040BG text, OSD030BH text, OSQ040BH text, OSD030BI text, OSQ040BI text, OSD030BJ text, OSQ040BJ text, OSD030CA text, OSQ040CA text, OSD050CA text, OSD030CB text, OSQ040CB text, OSD050CB text, OSD030CC text, OSQ040CC text, OSQ080 text, OSQ090A text, OSQ100A text, OSD110A text, OSQ120A text, OSQ090B text, OSQ100B text, OSD110B text, OSQ120B text, OSQ090C text, OSQ100C text, OSD110C text, OSQ120C text, OSQ090D text, OSQ100D text, OSD110D text, OSQ120D text, OSQ090E text, OSQ100E text, OSD110E text, OSQ120E text, OSQ090F text, OSQ120F text, OSQ090G text, OSQ100G text, OSD110G text, OSQ120G text, OSQ090H text, OSQ120H text, OSQ060 text, OSQ072 text, OSQ130 text, OSQ140Q text, OSQ140U text, OSQ150 text, OSQ160A text, OSQ160B text, OSQ170 text, OSQ180 text, OSQ190 text, OSQ200 text, OSQ210 text, OSQ220 text, PFQ020 text, PFQ030 text, PFQ033 text, PFQ041 text, PFQ049 text, PFQ051 text, PFQ054 text, PFQ057 text, PFQ059 text, PFQ061A text, PFQ061B text, PFQ061C text, PFQ061D text, PFQ061E text, PFQ061F text, PFQ061G text, PFQ061H text, PFQ061I text, PFQ061J text, PFQ061K text, PFQ061L text, PFQ061M text, PFQ061N text, PFQ061O text, PFQ061P text, PFQ061Q text, PFQ061R text, PFQ061S text, PFQ061T text, PFQ063A text, PFQ063B text, PFQ063C text, PFQ063D text, PFQ063E text, PFQ090 text, PAQ605 text, PAQ610 text, PAD615 text, PAQ620 text, PAQ625 text, PAD630 text, PAQ635 text, PAQ640 text, PAD645 text, PAQ650 text, PAQ655 text, PAD660 text, PAQ665 text, PAQ670 text, PAD675 text, PAD680 text, PAQ706 text, PAQ710 text, PAQ715 text, PAQ722 text, PAQ724A text, PAQ724B text, PAQ724C text, PAQ724D text, PAQ724E text, PAQ724F text, PAQ724G text, PAQ724H text, PAQ724I text, PAQ724J text, PAQ724K text, PAQ724L text, PAQ724M text, PAQ724N text, PAQ724O text, PAQ724P text, PAQ724Q text, PAQ724R text, PAQ724S text, PAQ724T text, PAQ724U text, PAQ724V text, PAQ724W text, PAQ724X text, PAQ724Y text, PAQ724Z text, PAQ724AA text, PAQ724AB text, PAQ724AC text, PAQ724AD text, PAQ724AE text, PAQ724AF text, PAQ724CM text, PAQ731 text, PAD733 text, PAQ677 text, PAQ678 text, PAQ740 text, PAQ742 text, PAQ744 text, PAQ746 text, PAQ748 text, PAQ755 text, PAQ759A text, PAQ759B text, PAQ759C text, PAQ759D text, PAQ759E text, PAQ759F text, PAQ759G text, PAQ759H text, PAQ759I text, PAQ759J text, PAQ759K text, PAQ759L text, PAQ759M text, PAQ759N text, PAQ759O text, PAQ759P text, PAQ759Q text, PAQ759R text, PAQ759S text, PAQ759T text, PAQ759U text, PAQ759V text, PAQ762 text, PAQ764 text, PAQ766 text, PAQ679 text, PAQ750 text, PAQ770 text, PAQ772A text, PAQ772B text, PAQ772C text, PAAQUEX text, PUQ100 text, PUQ110 text, RHQ010 text, RHQ020 text, RHQ031 text, RHD043 text, RHQ060 text, RHQ070 text, RHQ074 text, RHQ076 text, RHQ078 text, RHQ131 text, RHD143 text, RHQ160 text, RHQ162 text, RHQ163 text, RHQ166 text, RHQ169 text, RHQ172 text, RHD173 text, RHQ171 text, RHD180 text, RHD190 text, RHQ197 text, RHQ200 text, RHD280 text, RHQ291 text, RHQ305 text, RHQ332 text, RHQ420 text, RHQ540 text, RHQ542A text, RHQ542B text, RHQ542C text, RHQ542D text, RHQ554 text, RHQ560Q text, RHQ560U text, RHQ570 text, RHQ576Q text, RHQ576U text, RHQ580 text, RHQ586Q text, RHQ586U text, RHQ596 text, RHQ602Q text, RHQ602U text, RXQ510 text, RXQ515 text, RXQ520 text, RXQ525G text, RXQ525Q text, RXQ525U text, RXD530 text, SLD010H text, SLQ050 text, SLQ060 text, SMQ020 text,

```
        SMD030 text,   SMQ040 text,   SMQ050Q text,          SMQ050U text,
        SMD055 text,   SMD057 text,   SMQ078 text,   SMD641 text,   SMD650 text,
        SMD093 text,   SMDUPCA text,         SMD100BR text,         SMD100FL text,
        SMD100MN text,         SMD100LN text,         SMD100TR text,         SMD100NI
text,   SMD100CO text,         SMQ621 text,   SMD630 text,   SMQ661 text,   SMQ665A text,
        SMQ665B text,          SMQ665C text,          SMQ665D text,          SMQ670 text,
        SMQ848 text,   SMQ852Q text,         SMQ852U text,          SMAQUEX2 text,
        SMD460 text,   SMD470 text,   SMD480 text,   SMQ856 text,   SMQ858 text,
        SMQ860 text,   SMQ862 text,   SMQ866 text,   SMQ868 text,   SMQ870 text,
        SMQ872 text,   SMQ874 text,   SMQ876 text,   SMQ878 text,   SMQ880 text,
        SMAQUEXx text,         SMQ681 text,   SMQ690A text,          SMQ710 text,
        SMQ720 text,   SMQ725 text,   SMQ690B text,          SMQ740 text,   SMQ690C text,
        SMQ770 text,   SMQ690G text,         SMQ845 text,   SMQ690H text,
        SMQ849 text,   SMQ851 text,   SMQ690D text,          SMQ800 text,   SMQ690E text,
        SMQ817 text,   SMQ690I text,  SMQ857 text,   SMQ690J text,  SMQ861 text,
        SMQ863 text,   SMQ690F text, SMQ830 text,   SMQ840 text,   SMDANY text,
        SMAQUEXy text,         SXD021 text,   SXQ800 text,   SXQ803 text,   SXQ806 text,
        SXQ809 text,   SXQ700 text,   SXQ703 text,   SXQ706 text,   SXQ709 text,
        SXD031 text,   SXD171 text,   SXD510 text,   SXQ824 text,   SXQ827 text,
        SXD633 text,   SXQ636 text,   SXQ639 text,   SXD642 text,   SXQ410 text,
        SXQ550 text,   SXQ836 text,   SXQ841 text,   SXQ853 text,   SXD621 text,
        SXQ624 text,   SXQ627 text,   SXD630 text,   SXQ645 text,   SXQ648 text,
        SXQ610 text,   SXQ251 text,   SXQ590 text,   SXQ600 text,   SXD101 text,
        SXD450 text,   SXQ724 text,   SXQ727 text,   SXQ130 text,   SXQ490 text,
        SXQ741 text,   SXQ753 text,   SXQ260 text,   SXQ265 text,   SXQ267 text,
        SXQ270 text,   SXQ272 text,   SXQ280 text,   SXQ292 text,   SXQ294 text,
        WHD010 text,   WHD020 text,   WHQ030 text,   WHQ040 text,   WHD050 text,
        WHQ060 text,   WHQ070 text,   WHD080A text,          WHD080B text,
        WHD080C text,          WHD080D text,          WHD080E text,          WHD080F text,
        WHD080G text,          WHD080H text,          WHD080I text, WHD080J text,
        WHD080K text,          WHD080M text,          WHD080N text,          WHD080O
text,   WHD080P text,          WHD080Q text,          WHD080R text,          WHD080S text,
        WHD080T text,          WHD080U text,          WHD080L text,          WHD110 text,
        WHD120 text,   WHD130 text,   WHD140 text,   WHQ150 text,   WHQ030M text,
        WHQ500 text,   WHQ520 text,
        CONSTRAINT seqn_id PRIMARY KEY (seqn)
);

ALTER TABLE public.nhanes_questionnaire
        OWNER TO postgres;

-- Import the questionnaire CSV into the questionnaire table
COPY nhanes_questionnaire
FROM 'C:\Users\Public\Downloads\questionnaire.csv'
DELIMITER ','
CSV HEADER;

/* Now I have to clean this data up, turning coded values into human readable ones, handling
nulls, etc. */

-- Modify gender to strings from coded values
```

```
UPDATE public.nhanes_demographic SET gender = CASE WHEN gender = '1' THEN 'Male'
WHEN gender = '2' THEN 'Female' END;

-- Modify arthritis to strings from coded values, make it boolean too
ALTER TABLE public.nhanes_questionnaire ALTER arthritis TYPE bool USING CASE
WHEN arthritis = '1' THEN TRUE ELSE FALSE END;

-- Modify asthma to strings from coded values, make it boolean too
ALTER TABLE public.nhanes_questionnaire ALTER asthma TYPE bool USING CASE WHEN
asthma = '1' THEN TRUE ELSE FALSE END;

-- Modify diabetes to strings from coded values, make it boolean too
ALTER TABLE public.nhanes_questionnaire ALTER diabetes TYPE bool USING CASE
WHEN diabetes = '1' THEN TRUE ELSE FALSE END;

-- Modify high_bp to strings from coded values, make it boolean too
ALTER TABLE public.nhanes_questionnaire ALTER high_bp TYPE bool USING CASE
WHEN high_bp = '1' THEN TRUE ELSE FALSE END;

-- Modify hyperlipidemia to strings from coded values, make it boolean too
ALTER TABLE public.nhanes_questionnaire ALTER hyperlipidemia TYPE bool USING CASE
WHEN hyperlipidemia = '1' THEN TRUE ELSE FALSE END;

-- Modify overweight to strings from coded values, make it boolean too
ALTER TABLE public.nhanes_questionnaire ALTER overweight TYPE bool USING CASE
WHEN overweight = '1' THEN TRUE ELSE FALSE END;

-- Modify stroke to strings from coded values, make it boolean too
ALTER TABLE public.nhanes_questionnaire ALTER stroke TYPE bool USING CASE WHEN
stroke = '1' THEN TRUE ELSE FALSE END;

-- Add my source column to the NHANES data so it can be differentiated in Tableau from the
WGU data
ALTER TABLE public.nhanes_demographic ADD COLUMN source text;
UPDATE public.nhanes_demographic SET source = 'CDC';

-- WGU dataset omits minors, apply same rule to CDC dataset
DELETE FROM nhanes_questionnaire WHERE seqn IN (SELECT seqn AS id FROM
nhanes_demographic WHERE age < 18);
DELETE FROM nhanes_demographic WHERE seqn IN (SELECT seqn AS id FROM
nhanes_demographic WHERE age < 18);

/* Add some foreign keys to the WGU dataset to maintain referential integrity (why are these
missing?) and to the NHANES dataset as well, to clearly connect them */
ALTER TABLE public.servicesaddon ADD FOREIGN KEY (patient_id) REFERENCES
public.patient(patient_id);
ALTER TABLE public.survey_responses_addon ADD FOREIGN KEY (patient_id)
REFERENCES public.patient(patient_id);
ALTER TABLE public.nhanes_questionnaire ADD FOREIGN KEY (seqn) REFERENCES
public.nhanes_demographic(seqn);
```

```
/* Finally, create the UNION table with all of the data for the 12 columns I'm actually interested
in by creating two temp tables, UNION ALL together into a third table, then drop the temp tables
*/

CREATE TABLE wgu_temp AS (
        SELECT pat.age, pat.gender, pat.children, cond.arthritis, cond.asthma, cond.diabetes,
pat.high_bp, cond.hyperlipidemia, cond.overweight, pat.stroke, pat.source
        FROM patient AS pat
        INNER JOIN servicesaddon AS cond
        ON pat.patient_id = cond.patient_id);
CREATE TABLE nhanes_temp AS (
        SELECT demo.age, demo.gender, demo.small_children + demo.older_children AS
children, ques.arthritis, ques.asthma, ques.diabetes, ques.high_bp, ques.hyperlipidemia,
ques.overweight, ques.stroke, demo.source
        FROM nhanes_demographic AS demo
        INNER JOIN nhanes_questionnaire AS ques
        ON demo.seqn = ques.seqn);
CREATE TABLE study_data AS (
        SELECT *
        FROM wgu_temp
        UNION ALL
        SELECT *
        FROM nhanes_temp);
DROP TABLE IF EXISTS wgu_temp;
DROP TABLE IF EXISTS nhanes_temp;
```

B: Panopto Presentation

My Panopto presentation covering sections B1 – B7 of the project rubric can be found here.

C1: Alignment of Dashboard with Organizational Needs

As mentioned in my prior works with this dataset, the WGU medical data dictionary describes a particular interest in predicting readmission of previously hospitalized patients, but it doesn't preclude examination of other trends or alternative insights, including using other datasets to provide context for this data. As a result, I took a broad approach to look at trends in demographics and medical conditions for the patients in the WGU hospital system. With the CDC's NHANES data being representative of the United States population as a whole, I compared this with the WGU dataset to examine both the demographics (age, gender, number of children) of WGU Hospital System's patients vs the NHANES data, as well as the occurrence of serious medical conditions such as diabetes, high blood pressure, and hyperlipidemia. Insights in this regard can inform decisions such as whether limited resources should be allocated towards expanding pediatric care or geriatric care, or determining if WGU's patients have a higher (or lower) rate of a particular condition relative to what would be expected. These sorts of insights are immensely useful to any organization because they help enable both planning for the future and assessment of the present.

C2: Justify Business Intelligence Tool

The business intelligence tool used for this analysis is Tableau Desktop. Tableau is a program that allows for the easy creation of complex and interactable data visualization. This helps with communicating conclusions of data analysis, as well as making it easier to sift through complicated data to find new relationships or interesting observations.

C3: Data Preparation

The data preparation process involved the following steps:

1) Recast relevant WGU medical data from Yes/No string to True/False Booleans
2) Labelling the WGU medical data with a 'source' column so it could be easily distinguished in Tableau from CDC's NHANES data
3) Replace gender values of "Prefer not to answer" with "Non-Binary", consistent with the data dictionary in previous classes that have used this dataset
4) Top-code WGU age data at age 80, to provide a consistent comparison against the NHANES data, which also top codes any participant over the age of 80 as being 80 years old
5) Create tables for both the NHANES demographic and services addon data, using human-readable column names for the columns of interest and leaving all other columns (over 1000 columns) with their coded names provided by the CDC
6) Import the demographic.csv file to the nhanes_demographic table, and import the questionnaire.csv file to the nhanes_questionnaire table
7) Recast coded condition values (1 = has the condition, 2 = doesn't have the condition, etc.) into human-readable True/False Booleans
8) Label the CDC NHANES data with a 'source column so it could be easily distinguished in Tableau from the WGU medical data
9) Drop NHANES demographic and condition data for all participants under the age of 18, to provide a consistent comparison against the WGU data, which does not include any minors
10) Generate temporary tables of the WGU data that I'm interested in and the matching CDC data, and UNION ALL them together into a 'study_data' table for import into Tableau. Drop the temporary tables afterwards.

C4: Dashboard Creation

Following are the directions for creating the story that I presented in my data analysis, including the dashboards that were provided in that story and the worksheets that were imported into the dashboards.

AGES:

- Create Ages (bin 2) bins for histograms (bin size = 2)
- Drag Source & Age (bin 2) fields to columns
- Drag Age to rows, change aggregation to Count
- Change Title to "Number of Patients by Age (2 year groups)", Tableau Bold, Center Align
- Change Tooltip to: <CNT(Age)> of <Source> Patients are age <Age (bin 2)>

- Set to Fit Width
- Right click Age > Show Filter, modify under filters to Dimension, show filter, Apply to all using this data source

GENDER:

- Drag Source field to columns, Drag Gender field to Color
- Set Marks to Pie
- Drag Gender field to Angle, change measure type to Count
- Click colors, and edit color palette to something more intuitive (blue = male, pink = female, yellow = nonbinary)
- Click Label, check Show Mark Labels
- Hide field labels for columns
- Set to Fit Width
- Change Title to “Patient Genders”, Tableau Bold, Center Align
- Change Tooltip to: <CNT(Gender)> of <Source> Patients identify as <Gender>
- Right click Gender > Show as Filter, Modify by removing All button, Add Apply Button, Apply to all using this Data Source

CHILDREN:

- Create Children (bin 1) bins for histogram (bin size = 1)
- Drag Source and Children (bin 1) fields to columns
- Drag Children field to rows, change aggregation to Count
- Apply Quick Table Calculation to CNT(Children) using Pct of Total, compute using Pane (Across)
- Set to Show Mark Labels, Fit to Width
- Hide Field Labels for columns
- Change Title to “Number of Children by Patient”, Tableau Bold, Center Aligned
- Change Axis to “Percentage of Patients”
- Change Tooltip: <% of Total CNT(Children)> of <Source> Patients have <Children (bin 1)> children

DIABETES KPI (repeat for High BP KPI, Hyperlipidemia KPI):

- Drag Source field to columns
- Drag Diabetes field to Rows
- Drag count of full dataset to Text, Apply Quick Table Calculation as Pct of Total, Compute using Table (Down)
- Right click False and choose to Hide
- Right click True and uncheck Show Header
- Hide Field Labels for columns
- Modify Title to “Diabetes Rate”

- Right click Column Headers, Format to Tableau Medium 12 pt, Center Align, Light gray shading
- Right click cells, Format Pane to Tableau Book 10 pt, click to Format Alignment, Center Align
- Modify tooltip to read: <% of Total CNT(study_data)> <Source> Patients diagnosed as diabetic
- Drag column borders to resize as appropriate
- *Repeat process identically for High BP and Hyperlipidemia, substituting those fields for diabetes and updating title/tooltip info as appropriate*

AGE VS DIABETES (repeat for Age vs High BP, Age vs Hyperlipidemia)

- Create Age (bin 5) bins for heatmap (bin size = 5)
- Drag Source field to Rows, Drag Age (bin 5) to columns
- Drag Diabetes field to Color, change aggregation to Count
- Drag Diabetes field to Filter, select Arthritis = True
- Set to Show Mark Labels, Fit to Width
- Edit Title to "Number of Patients with Diabetes by Age (5 year groups)", Tableau Bold, Center Align
- Hide Field Labels for columns and rows
- *Repeat process identically for High BP and Hyperlipidemia, substituting those fields for diabetes and updated title/tooltip info as appropriate*

GENDER LEGEND:

- Drag Gender field to Rows
- Drag Gender field to Color
- Eliminate all titles
- Import to Dashboard as its own sheet
- Activate "Use As Filter"

DASHBOARD 1:

- Import Age worksheet onto top portion of page, Children worksheet to bottom left, Gender to bottom right
- Make Age Filter floating, and place in empty space below Gender
- Import Gender Legend Worksheet, make floating, and place in empty space below Gender
- Activate "Use As Filter" for all components

DASHBOARD 2:

- Import Age vs Diabetes to top of page, Age vs High BP worksheet to middle of page, Age vs Hyperlipidemia worksheet to bottom of page.
- Import Horizontal container to top of page. Adjust edges of all three figures and container as necessary.

- Set legends for Age vs Diabetes, Age vs High BP, and Age vs Hyperlipidemia to floating, and place each under the associated plot
- Insert Text to Horizontal Container and provide title “Age Breakdown and KPI’s by Condition” and directions “Filter by Age or Gender using the filters to the right”
- Drag Age Filter and Gender Filter into Horizontal container, to the right of title/directions text
- Import Vertical Container to left side of page
- Import Diabetes KPIs, High BP KPIs, and Hyperlipidemia KPIs to Vertical Container, with each positioned to the left of the associated age plot for that condition. Adjust spacing and edges as necessary, using “Distribute Evenly” and “Fit to Width”.
- Activate “Use As Filter” for all components

C5: Results of Data Analysis

There were two primary insights that I took from this analysis. First, the demographics of the WGU Hospital System patients are significantly different from the NHANES data, which is representative of the US as a whole. Specifically, the WGU patients are dramatically older than the sample collected by the CDC, with over 13% of patients being 80 years old or older compared to ~5.7% of participants in the NHANES survey. At the same time, WGU Hospital System patients have many more children than the NHANES survey participants. The second insight that I took from this analysis was that while WGU Hospital System patients have similar rates of hyperlipidemia as participants in the NHANES study do, WGU patients have much higher rates of high blood pressure and severely increased rates of being diabetic.

C6: Limitation of Data Analysis

There are definitely some limitations to this data analysis that are worth addressing. First of all, the analysis is predicated on the idea that the NHANES data is representative of the entire United States population. This is difficult to define as being correct or not, as the NHANES data would need to be either compared against extensive data for every single American (which does not exist) or compared against other representative samples (and each representative sample would have to address the same conditions, questions, etc.)

Another very significant consideration is that while the NHANES data is supposed to be representative of the US population at large as a point of comparison for the WGU data, the WGU patient population may not necessarily be consistent with the patterns of the overall US population, despite being made up of people from the US population. This is because the WGU dataset is made up of people who have already been hospitalized, meaning that it is not just 10,000 members of the US population but more specifically, 10,000 members of the US population of people who have been hospitalized in whatever timeframe the WGU data is taken from. This is a much smaller population, and because people who are hospitalized are by definition dealing with a serious health issue, this subset of the US population might have different characteristics than the entire US population does.

Other limitations may also exist with this analysis. For example, the CDC’s data does not seem to account for non-binary gendered patients, while WGU does treat non-binary patients, though they are a small portion of the whole. This leads to a lack of comparison for non-binary patients, and any

conclusions that may try to be drawn about this group are subject to issues based on the small sample size. Other issues exist with the age of patients in this study. While the CDC top-codes participants ages at 80 years old, WGU does not, and patterns may change as patients age past 80 years old. WGU's dataset also omits minor patients, despite the fact that minors are hospitalized. Minors made up nearly 40% of the NHANES data, but they were omitted in this study to create consistent comparisons. WGU's choice to omit minors from their data skews their demographics and can create misleading conclusions if this analysis is too broadly implemented.

D: Sources

William Townsend D210 Task 1 Project was used throughout this project, as it consisted of a similar analysis, substituting SQL for data preparation rather than Python.

Centers for Disease Control & Prevention: 2013-14 National Health and Nutrition Examination Survey (NHANES), hosted on Kaggle, is the dataset that was used for this project. Dictionaries provided by the CDC were used to understand the variable names for column in the demographic and questionnaire portion of the dataset. Translations of values (1 = True, 2 = False, etc.) were provided by the CDC here.