**D210 Task 1: Data Dashboard and Storytelling**

**Joseph Duszynski**

**ID #011237236**

**A. Interactive data dashboard**

See attached file “CDI\_2020\_and\_WGU\_data\_comparisons.twb” or hosted version at <https://public.tableau.com/app/profile/joseph.duszynski/viz/CDI_2020_and_WGU_data_comparisons/WGUand2020U_S_CDIComparisons>

**A.1. Data sets**

See “medical\_clean\_transformed.csv” (modified version of WGU’s medical dataset [1] primarily for one-hot encoding) and “U.S.\_Chronic\_Disease\_Indicators\_2020.csv”. This second file was derived from CDC data [2] that is over 350 MB in its raw form, so it was restricted to ‘YearStart’ = 2020, ‘Stratification1’ = ‘Overall’, and only retains the columns 'LocationAbbr', 'Question', 'DataValueUnit', 'DataValueType', 'DataValue', 'LowConfidenceLimit', 'HighConfidenceLimit', reducing the file size to ~1.3 MB. Alternatively, to recreate it, download the raw file from the source provided in [2], and apply the following (changing source directory as needed):

df\_ext = pd.read\_csv('data/D210-external-datasets/U.S.\_Chronic\_Disease\_Indicators\_\_CDI\_\_\_2023\_Release.csv')  
  
columns = ['LocationAbbr', 'Question', 'DataValueUnit', 'DataValueType', 'DataValue', 'LowConfidenceLimit', 'HighConfidenceLimit']  
cond = (df\_ext['YearStart'] == 2020) \* (df\_ext['Stratification1'] == 'Overall')  
  
df\_ext\_f = df\_ext[columns][cond].sort\_values(by='LocationAbbr')  
df\_ext\_f.to\_csv('data/D210-external-datasets/U.S.\_Chronic\_Disease\_Indicators\_2020.csv')

**A.2. Installation instructions**

It’s most convenient to access the dashboard at [this link](https://public.tableau.com/app/profile/joseph.duszynski/viz/CDI_2020_and_WGU_data_comparisons/WGUand2020U_S_CDIComparisons) in a browser.

Load the two data files mentioned in A.1. in the “Data Source” page and form a relationship on U.S.\_Chronic\_Disease\_Indicators\_2020[‘Location Abbr’] = medical\_clean\_transformed[‘State’].

For the “Recent limitations stacked bar chart” sheet, using the 2020 CDI data, place “Location Abbr” in “Columns”, “AVG(Data Value)” in Rows, “Question” in “Marks” with “Color” selected (using bar chart option). For filters, restrict “Question” to “Recent activity limitation among adults aged >= 18 years”, “Recent mentally unhealthy days among adults aged >= 18 years”, and “Recent physically unhealthy days among adults aged >= 18 years”. Filter “AVG(Data Value)” to non-null values in the “Special” tab. Filter “Data Value Type” by “Mean”.

For the sheet “Init-days\_and\_TC\_by\_Job\_and\_Init-admin”, use the “medical\_clean\_transformed” data and place “Job” in “Columns”. For rows, “AVG(Total Charge)” and “AVG(Initial days)”. Then place “Initial admin” in “Marks” and use the “Shape” option (for both variables in “Rows”).

The Asthma Diff, Arthritis Diff, and Overweight Diff sheets are all analogous. From the 2020 CDI data, place “Location Abbr” in “Marks” as a detail and “AVG(Data Value)” as text. Include “Question”, “Data Value Unit”, and “Data Value Type” as attributes. Set the mode to “Map”. Use the calculated field (will be presented below) “Asthma Difference” as an aggregate in “Marks” with the “Color” option selected. In “Filters”, restrict “Data Value Unit” to “%” and initially allow “Data Value Type” to take all values. Select “Show Filter” through the right click context menu which produces a “Data Value Type” widget in the upper right corner. In the widget’s context menu, select “Multiple values (list)” and “Only Relevant Values”. It’s best to set it to “Crude Prevalence” as the WGU data doesn’t adjust for age, but “Age-adjusted Prevalence” is worth including should it be desired for a user. Filter the “Question” to “Current asthma prevalence among adults aged >= 18 years”. Adjust the “Colors” in “Marks” to have a center point of 18 with Red-Blue Diverging and tick “Reversed”.

The “Asthma Difference” calculated field is:

100 \* SUM([Asthma Yes]) / COUNT([Case Order]) - AVG([Data Value])

For arthritis, copy the above sheet, change the “Question” to “Arthritis among adults aged >= 18 years”, change colors to Orange-Blue\_White Diverging with a center point of 8, and use the aggregate “Arthritis Difference” calculated field, which is:

100 \* SUM([Arthritis Yes]) / COUNT([Case Order]) - AVG([Data Value])

For overweight, again copy the asthma sheet, change the “Question” to “Overweight or obesity among adults aged >= 18 years”, change colors to Red-Black Diverging with a center point of 1, and use the aggregate “Overweight Difference” calculated field:

100 \* SUM([Overweight Yes]) / COUNT([Case Order]) - AVG([Data Value])

The sheet “CDI\_Q\_w\_ReAdmis” primarily uses the 2020 CDI data. Add “Location Abbr”, “Data Value Unit”, and “Data Value Type” to the “Marks” shelf as details. Add “AVG(Data Value)” as a “Color” (Red-Blue Diverging, Reversed) with the “Map” mode selected. Create the calculated field “Readmission percentage”:

100 \* SUM([ReAdmis Yes]) / COUNT([Case Order])

Add that as a “Label” to “Marks”. Add “Question”, “Data Value Unit”, and “Data Value Type” to “Filters”, allowing all options and using their right click context menus to “Show Filter”. Use “Multiple Values (list)” and “Only Relevant Values” for the “Data Value Unit” and “Data Value Type” widgets. For “Question”, use “Single Value (dropdown)” and “All Values in Database”. Change title to “2020 U.S. CDI Map by Survey Question (WGU readmission rate overlaid in each state)”.

For the final sheet “Condition Selector”, using the WGU data, add “State” as a “Label” to “Marks”, select “Map”, and add the calculated field “Selected Condition Percentage”:

CASE [Condition Selector]

WHEN 'Allergic Rhinitis' THEN 100 \* SUM([Hyperlipidemia Yes]) / COUNT([Case Order])

WHEN 'Anxiety' THEN 100 \* SUM([Anxiety Yes]) / COUNT([Case Order])

WHEN 'Arthritis' THEN 100 \* SUM([Arthritis Yes]) / COUNT([Case Order])

WHEN 'Asthma' THEN 100 \* SUM([Asthma Yes]) / COUNT([Case Order])

WHEN 'Back Pain' THEN 100 \* SUM([BackPain Yes]) / COUNT([Case Order])

WHEN 'Diabetes' THEN 100 \* SUM([Diabetes Yes]) / COUNT([Case Order])

WHEN 'Hyperlipidemia' THEN 100 \* SUM([Hyperlipidemia Yes]) / COUNT([Case Order])

WHEN 'Overweight' THEN 100 \* SUM([Overweight Yes]) / COUNT([Case Order])

WHEN 'Readmission' THEN 100 \* SUM([ReAdmis Yes]) / COUNT([Case Order])

WHEN 'Reflux Esophagitis' THEN 100 \* SUM([Reflux esophagitis Yes]) / COUNT([Case Order])

WHEN 'Stroke' THEN 100 \* SUM([Stroke Yes]) / COUNT([Case Order])

END

as a color (Red-Blue-White Diverging, Reversed). Create a parameter called “Condition Selector” that takes “String” as its “Data type” and has “Allowable values”:

* Allergic Rhinitis
* Anxiety
* Arthritis
* Asthma
* Back Pain
* Diabetes
* Hyperlipidemia
* Overweight
* Readmission
* Reflux Esophagitis
* Stroke

Right click on it and select “Show Parameter” in its context menu.

Under “Filters”, add “Age”, “Gender”, “Area”, “Initial admin”, “Complication risk”, “Services”, “Overweight”, and “Initial days”. For all eight of these, again select “Show Filter”. On the right edge of the screen,

* “Condition Selector”: use “Compact List”
* “Initial days”: “Range of Values” and “All Values in Database”
* “Age”: “Range of Values” and “All Values in Database”
* “Gender”: “Single Value (slider)” and “All Values in Database”
* “Overweight”: “Single Value (slider)” and “All Values in Database”
* “Area”: “Multiple Values (list)” and “All Values in Database”
* “Initial admin”: “Multiple Values (list)” and “All Values in Database”
* “Services”: “Multiple Values (list)” and “All Values in Database”
* “Complication risk”: “Multiple Values (list)” and “All Values in Database”

Additionally, create the calculated field “Number of Affected Patients”:

COUNT([Case Order])

and include it as a detail in the “Marks” shelf.

To create “Dashboard 1”, add the “Asthma Diff”, “Arthritis Diff”, and “Overweight Diff” sheets one at a time into the panel (1920 x 1080 resolution). Set the “Data Value Type” widget and color legend to floating and move them to the upper right hand corner of each map, making them roughly 640 x 1080 each. Create a text object in the bottom of the middle map (arthritis) stating: “Difference between WGU condition percentage (asthma, arthritis, overweight) and 2020 U.S. CDI. Value within each state is from the 2020 U.S. CDI data.”

Dashboard 2 is created from the stacked bar chart sheet.

Dashboard 3 uses the “Init-days\_and\_TC\_by\_Job\_and\_Init-admin” sheet with “Fit” set to “Standard” to retain the required horizontal scroll bar.

Dashboard 4 is the condition selector map with widgets rearranged to original order in the corresponding sheet.

Similarly, Dashboard 5 is the “CDI\_Q\_w\_ReAdmis” map with widget order restored.

The five dashboards are placed into a single “Story” in the following order: Dashboard 4, Dashboard 5, Dashboard 1, Dashboard 3, Dashboard 2.

**A.3. Navigation instructions**

The Tableau Public link previously mentioned is the story page “WGU and 2020 U.S. CDI Comparisons” in the “CDI\_2020\_and\_WGU\_data\_comparisons.twb” file, which consists of the five dashboards in the order described in section A.2.

Dashboard 1 has three separate maps with independent panning and zoom controls. It’s recommended to leave “Crude Prevalence” as the selected “Data Value Type”. The overlaid numerical labels are equivalent to the 2020 CDI rates for the described conditions by state. The measured difference in rates as depicted by the color gradient is the WGU dataset’s rate of the respective condition minus its rate in the CDI data. As an example, California has 28.91% occurrence of asthma in the WGU data while it’s 9.300% in the 2020 CDI data, giving a difference of 19.61% (shown in the tooltip).

Dashboard 2 is a stacked bar chart by state of recent days with activity limitations (blue on top), poor mental health (yellow in the middle), and poor physical health (green on the bottom). Mouse over the selected segment for additional details in the tooltip.

Dashboard 3 has tooltips when the cursor hovers over a point of interest. The corresponding horizontal and vertical dimensions with the “Initial admin” legend describe the data of interest.

Dashboard 4 is a map with panning and zoom controls (mouse control or upper left hand corner of map) for WGU data in the U.S. and some overseas territories. “Condition Selector” is a drop down menu for asthma, diabetes, and other conditions in the dataset that show the rates of those conditions by state according to a color gradient (see the legend underneath the selector). Additionally, the condition rates can be filtered by subsets of length of initial stay, age, gender, overweight status, area type, initial admission reason, initial service sought, and assessed complication risk. Select the desired boxes or buttons to adjust the filters as needed, initially set to allow all patients (no filtering).

Dashboard 5 is a map of the 2020 U.S. CDI data again with zoom and panning capabilities. The “Question” widget on the right hand side will render a color graded map by state for the measured data value of that question. The color legend below gives an approximate visual measure, although mousing over a state will show a tooltip with all desired information. The “Data Value Unit” and “Data Value Type” widgets will ideally be restricted to only one value each. Depending on the question, the dataset has different values recorded, such as a crude rate as well as an age-adjusted rate. If overlapping values are selected it will report an average. Asterisks for “Data Value Unit/Type” in a tooltip indicate a conflict. The overlaid numerical text on each state is the WGU readmission percentage.

**B. Panopto storytelling with data**

See the attached link: https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=ad566285-a8bb-478c-b3d5-b17301665ae0

Key Results:

1. Cumulative recent bad health days are highest in Arkansas, Kentucky, Louisiana, Nevada, and West Virginia (over 12 days across three categories).
2. Separated by job, very long (~55 days and over) and rather short (under ~15 days) initial stays tend to be for observation or done electively. Emergency admissions are quite frequently ~20-50 days.
3. WGU hospitals have unusually high asthma rates in nearly every state while western states have abnormally high arthritis in their patients. In Montana, Colorado, and Vermont their patient obesity rates are anomalously higher than the U.S. CDI data.
4. Wyoming and Montana have high arthritis rates in patients and surprisingly they become considerably higher in patients who aren’t overweight. Utah’s rates are lower overall, but similarly has the counterintuitive effect. There may be unique lifestyle factors explaining this phenomenon as it the differences by obesity are generally small in most states.
5. Readmission rates are highest in the Midwest and South. The admission type overall has no clear correlation, although some states have extreme differences by admission type. When the initial admission is for a CT scan, readmissions tend to be higher, particularly in North Dakota and Oregon, although they have a smaller case count for CT scans. Readmissions for MRIs are mostly uniform. Some states have higher readmission rates but they have small case counts as well. Readmission following blood work is most common in southern states. Overweight patients are most at risk of readmission in the South while the risk is generally much lower in western states. Readmissions by complication risk is mostly uniform across all states.
6. From the CDI data, premature mortality is high in the south but doesn’t display a clear correlation to readmission rates. Generally, high mortality and poor health indicators are concentrated in the south while binge drinking is highest in the Midwest and northern plains states. Reasons for high admission rates likely rely on multiple different factors, although there’s unexplained variance in Texas’ high readmission rate despite generally being one of the healthier states (according to CDI data surveying).

Actionable Insights:

1. Readmission occurs in patients with an initial stay of ~48 days or more. Additional data collection will be needed to understand what causes their lengthy stays and what might be done to mitigate the probability of the patient requiring readmission.
2. WGU patients have very high rates of asthma, arthritis, and obesity. Other health conditions may similarly have high rates, but that will require additional data to confirm. This needs to be thoroughly examined.
3. Readmission rates are highest in the Midwest and South, so they may need special attention from regional managers and staff.
4. Why do readmission rates vary so drastically within some states based on admission type? Are the staff in some departments better trained than others? Consult regional managers and administrators working at affected hospitals.
5. Why are readmissions higher for patients seeking CT scans? Why do obesity and patients being admitted for blood work have higher readmission rates in the South than other states? Regional managers may have a better understanding of factors within their hospitals and communities that could explain this. Perform additional research and data collection for correlations in local community health (e.g. binge drinking) and readmission rates. Ideally more granular data by county, city, or hospital.
6. Most importantly, the causes for patient readmission and length of further stays are needed.
7. Some subpopulation sizes are too small to draw confident conclusions, so comparisons with neighboring hospitals in each county or state would be beneficial.

**C.1. Dashboard alignment**

As discussed in the data dictionary [1], hospital administrators and analysts would like to understand dynamics driving readmission rates within their healthcare system (in the WGU dataset). The variables collected in the dataset are somewhat limited and are confined to 10000 patients, which could be improved by comparing to a U.S. national survey of many different health metrics and indicators. The WGU data may have abnormally high (or low) instances of certain health conditions, which would require further investigation. By comparing the WGU readmission rates by state to state-level data contained in the 2020 CDI, patterns that wouldn’t otherwise be noticed may emerge. High premature mortality and lack of access to healthcare coverage are two outside metrics that might correlate to WGU readmission rates.

The stacked bar chart gives a general measure of “unhealthy days” by state, identifying states that have underlying poor health amongst their population. The discrete scatter plot by job and admission type can be used to visually identify outliers in hospital bills and length of initial stay by job title and admission classification.

The differences in asthma, arthritis, and overweight status (Dashboard 1) display the discrepancies in rates of those conditions by state, prompting further investigation into states and conditions with profound differences.

The “Condition Selector” interactive map is a convenient way to view state-level percentages of the conditions recorded in the WGU data (e.g. asthma, hyperlipidemia) as well as the readmission rate. Filter options can confine questions and analysis to specific subgroups and show regions, demographics, and health conditions that are worthy of additional investigation or targeted changes.

The ”2020 U.S. CDI Map by Survey Question” map integrates the WGU readmission rates as a label over each state with questions of interest from the CDI dataset. For example, premature mortality may have a visually detectable relationship with readmission rates or identify anomalies worthy of further analysis.

The included dashboards allow administrators to review multiple different variables in one panel for insights into readmission rates and the overall health status of each state’s population. Additionally, Dashboard 3 provides information by job and initial admission type.

**C.2. Additional data set insights**

The WGU data has a limited set of health questions responded to in a binary (“yes”/’no”) manner. Including the U.S. 2020 CDI data allows comparison of state-level rates of asthma, arthritis, and overweight status, as well as other health measures not captured by the WGU data. Inspecting community health metrics by state may reveal correlations to readmission rates or anomalies worth investigating further. Expanding the dataset to a broader national one allows comparisons of WGU’s health system with the rest of the country outside of its health network.

**C.3. Decision-making support**

The “Condition Selector” map allows administrators and analysts to view rates for readmission and various health conditions within the WGU hospitals by state. Identifying states performing well can lead to determining best practices that would improve the states performing poorly. Additionally, states may have unique regional factors best known by regional managers that could affect their rates of these health conditions. Various filters are included to allow the rates to be viewed for subpopulations of interest, such as patients with an initial stay of 50 days or more.

Another important visualization is the “2020 U.S. CDI Map by Survey Question (WGU readmission rate overlaid in each state)” map, displaying the WGU readmission rate overlaid on each state and a color gradient determined by the selected CDI question of interest. Analysts and administrators can learn of numerous health metrics by state including tobacco and alcohol use, recently unhealthy days, mortality rates for many conditions, disease rates, vaccination rates, premature death rates, and others. States or broader regions may have their own unique health deficits driving hospital admissions and readmissions. A better understanding of community health by state allows informed decisions for preventative care, community outreach, and thoroughly addressing a patient’s health when admitted to ideally reduce their stay and lower the chance of readmission.

**C.4. Interactive controls**

The “Condition Selector” map allows the user to select readmission or a health condition of interest among the WGU dataset in the “Condition Selector” parameter field. This renders the map and associated color gradients for the given condition. Additionally, the user can filter by “Age”, “Gender”, “Area”, “Initial admin”, “Complication risk”, “Services”, “Overweight”, and “Initial days” for insights into subpopulations of interest (determined by their selection in the above factors).

The ”2020 U.S. CDI Map by Survey Question” map has a “Question” field in the upper right corner for the user to render the map according to the survey question of interest (from the 2020 U.S. CDI data). For questions with multiple recorded data values (such as a crude rate and age-adjusted rate), they can use additional filters on the right side of the panel to adjust the map data for the value of interest.

**C.5. Colorblindness**

Tableau doesn’t have native options for monochrome texturing, so in the interest of greatest accessibility, colors were selected to be of high contrast and sufficiently distinct to accommodate most forms of colorblindness (usually red-blue or blue-orange). However, for users with forms of colorblindness where these color palettes are still difficult to distinguish, mouse-over tooltips and/or selecting “Edit Colors” from the dashboard’s color legend (in the upper right corner of a visualization) can better serve their needs. Unfortunately, with a lack of texturing options, it’s impossible to fully accommodate all users.

**C.6. Data representations**

One of the most relevant conclusions from these visualizations is shown in Dashboard 1 (differences in asthma, arthritis, and being overweight between WGU hospitals and the U.S. CDI data), highlighting the extraordinary differences by state, in most cases well above state-level averages. This depiction allows administrators to quickly identify outlier states with the greatest differences (positive or negative). In particular, nearly every state has asthma rates at least 15% higher than averages from the CDI data. The maps make it very clear that rates of these conditions are abnormally high and need further investigation. This may be the case for other conditions as well, but they weren’t available in the CDI data.

The ”Condition Selector” map provides essential information by state with multiple filtering options. Administrators can focus on readmission rates or a health condition of interest, identify states that perform well or poorly, and further refine the scope with factors such as the patient’s initial stay, age, complication risk, and others. Restricting the initial stay to ~48 days or less shows there are no readmissions across the entire healthcare system. Removing that restriction clearly shows readmission rates tend to be highest in the Midwest and South overall. Adjusting the filters quickly demonstrates readmission rates for patients initially seeking blood work are considerably higher in southern states than most other states, prompting further inquiry into local community health and hospital practices in those regions.

**C.7. Audience analysis**

The senior vice president of hospital operations, as described in the data dictionary [1], oversees operations across all hospitals, creates initiatives to improve patient outcomes, and would like to learn more about patient treatments and outcomes across regions and demographic factors. They would benefit greatly by having a visual map depicting metrics for each state, which can be found in the “Condition Selector” map. The presentation spends time reviewing the capabilities of this visualization in assisting the senior vice president with their responsibilities. Additionally, the presentation reviews the 2020 U.S. CDI map displaying health metrics outside of the WGU data. The senior vice president can review national health trends by state to glean possible relationships in broader health issues by region and their regional readmission rates within the WGU healthcare system.

The vice president of research investigates patterns of patient care and looks to improve patient outcomes with strategic initiatives. This executive would similarly benefit from the “Condition Selector” map and 2020 U.S. CDI map. In addition, the aberrant rates of asthma, arthritis, and overweight patients would be of great importance to them in understanding why so many patients in the WGU system are disproportionately unhealthy to the rest of the country. Comparisons with other hospitals systems, if possible, could be a starting point. They might also consider investigating chronic health conditions leading to increased hospital admissions and other feedback effects. This is discussed in detail in the presentation to give the vice present of research subjects of interest to investigate.

The regional vice presidents are also addressed through the “Condition Selector” and 2020 U.S. CDI maps, highlighting regional successes and weaknesses. As they learn how their regions compare to other regions across the country, they can conduct further research into areas needing improvement by collaborating with senior management and other regional vice presidents.

Explain how you used audience analysis to adapt the message in your presentation. They also benefit from the WGU vs. U.S. CDI comparison maps in asthma, arthritis, and being overweight to formulate and assess reasons for their region’s performance, whether it’s better or worse than others. Lastly, the stacked bar chart allows the regional vice presidents to see how their states fare overall in terms of recently unhealthy days.

**C.8. Universal access**

The colors selected throughout the visualizations were chosen to accommodate colorblind users as much as possible. The file was made available publicly at the link mentioned in section A. and has been successfully tested on mobile devices. Legends, text, and tooltips were incorporated to improve clarity and ease of use. Lastly, the Panopto presentation in part B. may be of use to users with poor vision.

**C.9. Effective storytelling**

Hospital administrators are primarily concerned with hospital readmission rates and patient outcomes, much of which focuses on regional trends and relative differences. Constructing an easily navigable visualization with a clear depiction of those trends is most easily accomplished by a map of the country with a color gradient. The map visualizations were created with clear color distinction in mind and informative tooltips for immediate insights into regional differences. Visual clarity and effective use support administrators in their objective of understanding readmission rates, patient outcomes, and how they differ by region, demographics, and preexisting health conditions. The presentation demonstrates the utility of the maps in identifying trends among patient subpopulations of interest.

Incorporating the 2020 U.S. CDI data adds many different health metrics that weren’t previously available within the WGU dataset. Administrators can use the visualizations to understand broader health trends across the country and how they interact with rates of readmission and health conditions within patients of their healthcare system. Relevant, accurate data expanded the initial WGU dataset, allowing conclusions and a narrative that wouldn’t have been possible to see otherwise. In particular, WGU hospitals have extremely high rates of asthma, arthritis, and overweight patients compared to the rest of the country. Administrators can review state-level health metrics and determine potential areas of further research. For instance, the higher rates of binge drinking in certain northern states could play a role in readmissions, whereas higher rates of obesity and overall poor health in the south may require a different and more detailed investigation into patient readmissions. Hospital administrators likely haven’t viewed their internal data side by side with U.S. CDI data, so incorporating the two datasets in these visualizations and associated presentation should be of great interest to them.

**D. Sources**

**1.** WGU. 2024. D210 Representation and Reporting “Data Sets and Associated Data Dictionaries for Provided Data Sets”. Medical Data and Dictionary Files. Retrieved May 16, 2024, from https://access.wgu.edu/ASP3/aap/content/k8gj49f8sichedufutms.html.

**2.** CDC. 2024. U.S. Chronic Disease Indicators (CDI), 2023 Release. Retrieved May 16, 2024, from <https://data.cdc.gov/Chronic-Disease-Indicators/U-S-Chronic-Disease-Indicators-CDI-2023-Release/g4ie-h725/data_preview>.