North Central Washington Epi Consortium Community Health Indicator Construction: Data Download and Cleaning Methods Document

June 2024

**This documents the construction of Community Health Assessment Metrics for the NCW Epi Consortium Data Dashboards.** Each indicator is listed under its corresponding data source. There is also a corresponding Excel spreadsheet “*NCW Epi Dashboards\_indicator information*” which includes the relevant information that should be included in the dashboards, such as a blurb on “why this indicator is important” written by NCW Epi Consortium Members, unit of indicator, a source for the data that should be listed, etc. Please refer to this document and that spreadsheet when updating the NCW Dashboards.

Additionally, after all of the data has been updated and cleaned according to the instructions below, the actual Power BI dashboards will need to be updated. Instructions to do that are included in the document “*NCW Epi Dashboards\_refresh PowerBI instructions*”.

This document is organized by data source. The instructions for downloading each indicator are underneath their corresponding data source.

**Selection of Indicators**

Community Health Assessments were queried from LHJs within the NCW Epi Consortium. Similar indicators were aggregated together to come up with a list of over 100 different health indicators. From this list, indicators used in 2 or more previous Community Health Assessments were included within the final list. LHJs then met and decided on other indicators that were important to LHJs. In September 2023, the NCW Epi consortium decided on 25 indicators to prioritize for the initial construction of the dashboards. Those 25 (*with the exception of Overdose Deaths because of data access issues as of 5/16/2024*) are included in this document, but we anticipate that more indicators will be added over time.

# Community Health Assessment Tool (CHAT)

**Description of data source**: The Community Health Assessment tool, or CHAT, is an online query system for population health-based data sets maintained by the Washington State Department of Health. The tool allows users to conduct analyses for community health assessment. Lots of information is collected via CHAT, such as on birth risks, pregnancy outcomes, communicable diseases, life expectancy, and deaths and hospitalizations.

More information can be found on the [DOH website here](https://doh.wa.gov/public-health-healthcare-providers/public-health-system-resources-and-services/community-health-assessment-and-improvement/chat#:~:text=CHAT%20provides%20a%20secure%2C%20online,analyses%20for%20community%20health%20assessment.).

The indicators from this data source is…

* Diabetes
* Mental Health
* Flu vaccination
* Overdose Deaths
* Suicide
* Causes of Death
* Death Rate
* Life Expectancy
* Hospitalizations
* Teenage Pregnancy Rate

**Accessing data source**:This data source requires permission to access. Typically, CHAT access is reserved for direct employees of LHJs. If the person updating the dashboards is considered a 3rd party contractor to an LHJ (i.e., someone like a UW PhD student), the *responsible party of the LHJ or LHJ consortium (supervisor or point of contact) would need to expressly write an email to the DOH CHAT Epidemiologist giving permission for the contractor to have access to CHAT*. The email should look something like this:

“I, (Supervisor/ POC name), hereby give permission for (contractor name) to have access to CHAT. I agree that I am responsible for any CHAT related work conducted by (contractor name). The following individuals have been approved for access (blank, blank, blank). Access is granted from (Starting date) to (Ending Date).”

At the time of writing these instructions, Kaylee Kim was the CHAT Epidemiologist ([Kaylee.Kim@doh.wa.gov](mailto:Kaylee.Kim@doh.wa.gov)). The general CHAT DOH email is [chs.chat@doh.wa.gov](mailto:chs.chat@doh.wa.gov) which should also be emailed during this process. Email both of these addresses when you send the message above.

Once you are accepted, you will receive an email with instructions on how to access CHAT, which are very detailed, but you can reach out to the emails above if you need assistance of have further questions.

**Where to access data**: Once you have permission to access these data (as outlined above), all data for CHAT can be downloaded here: [https://secureaccess.wa.gov/](https://urldefense.com/v3/__https:/secureaccess.wa.gov/__;!!K-Hz7m0Vt54!kPH3Wvm_ulLY8UPvMf46Iq9Yo6L-pUfAME07DidUtFRafeUbbLxiP30o__O4ocpkFrOz2rQ_UzOOAy5d5OI$)

**Diabetes – County and State-Level estimates**

* **Download instructions** 
  1. Go to [https://secureaccess.wa.gov/](https://urldefense.com/v3/__https:/secureaccess.wa.gov/__;!!K-Hz7m0Vt54!kPH3Wvm_ulLY8UPvMf46Iq9Yo6L-pUfAME07DidUtFRafeUbbLxiP30o__O4ocpkFrOz2rQ_UzOOAy5d5OI$) and click “Access Now” next to “Community Health Assessment Tool provided by Department of Health”
  2. Select “BRFSS” from the “Topics” tab
  3. You will be sent to a “Request” tab where you can select your query preferences
     1. Statistic or Measure: “Age-Adjusted Proportion”
     2. BRFSS Variable: “Diabetes dx-Yes”
     3. Geography: select State, Chelan, Douglas, Grant, Kittitas, and Okanogan (you can select all at once by holding down “CTRL” and selecting these counties)
     4. Year Rollup: 1 Year
     5. Year: select your *SINGLE* year of interest
        + **Previous years should already be downloaded and available in the dashboard data repository. The newest year of data, or the data that are not already available in the repository, are the only datasets that should be downloaded, and they should be downloaded as single years. This makes the data cleaning and compiling process much easier moving forward**
     6. Gender: All (Combined)
     7. Race: All (Combined)
  4. Hit “Submit”, which should then take you to a Results page
  5. In the results page, select all available fields (Year, Geography, BRFSS Variable, Gender, Race, Proportion, Lower CI, Upper CI, SE, RSE, Count, Population). *Be sure to hit “Refresh Table”* when you do this, otherwise all of the selected fields will not populate
  6. Under Export, download data as a CSV and then hit “Download”
  7. You should be able to hit “Request” across the top or “Back” and redo steps 3-6 with other years of data if needed
* **Storing Data**
  1. The data should download as a generic “BRFSS\_output”, but rename as follows:
     1. DataSource\_Indicator\_Level\_Years
     2. Example: CHAT\_Diabetes\_All\_2016
  2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/CHAT/Diabetes** folder. This will allow for the data cleaning process below to be automated
* **Clean Data** 
  1. You will use the “clean\_diabetes.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Diabetes data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Diabetes data into the **ncw\_data/cleaned\_data** folder as **diabetes\_data.csv**. This data file will be the one that is pulled into PowerBI

**Flu Vaccination – County and State-Level estimates**

* **Download instructions** 
  1. For downloading data for this indicator, you will use the same steps as you did for the Diabetes indicator above. *However, for 3b, select “Flu Shot or Vaccine-Yes”*. All other instructions remain the same
* **Storing Data**
  1. The data should download as a generic “BRFSS\_output”, but rename as follows:
     + DataSource\_Indicator\_Level\_Years
     + Example: CHAT\_FluVaccination\_All\_2016
  2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/CHAT/Flu Vaccination** folder. This will allow for the data cleaning process below to be automated
* **Clean Data** 
  1. You will use the “clean\_fluvax.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Flu vaccination data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Flu vaccination data into the **ncw\_data/cleaned\_data** folder as **fluvax\_data.csv**. This data file will be the one that is pulled into PowerBI

**Mental Health – County and State-Level estimates**

* **Download instructions** 
  1. For downloading data for this indicator, you will use the same steps as you did for the Diabetes indicator above. *However, for 3b, select “Mental Health-Not Good”*. All other instructions remain the same
* **Storing Data**
  1. The data should download as a generic “BRFSS\_output”, but rename as follows:
     + DataSource\_Indicator\_Level\_Years
     + Example: CHAT\_MentalHealth\_All\_2016
  2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/CHAT/Mental Health** folder. This will allow for the data cleaning process below to be automated
* **Clean Data** 
  1. You will use the “clean\_mh.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Mental Health data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Mental Health data into the **ncw\_data/cleaned\_data** folder as **mh\_data.csv**. This data file will be the one that is pulled into PowerBI

**Teenage Pregnancy Rate – County and State-Level estimates**

* **Download instructions** 
  1. Go to [https://secureaccess.wa.gov/](https://urldefense.com/v3/__https:/secureaccess.wa.gov/__;!!K-Hz7m0Vt54!kPH3Wvm_ulLY8UPvMf46Iq9Yo6L-pUfAME07DidUtFRafeUbbLxiP30o__O4ocpkFrOz2rQ_UzOOAy5d5OI$) and click “Access Now” next to “Community Health Assessment Tool provided by Department of Health”
  2. Select “Pregnancy and Abortion” from the “Topics” tab
  3. You will be sent to a “Create Query” tab where you can select your query preferences
     1. Primary Statistic or Measure: “Pregnancy Rate”
     2. Year: select year of interest
        + **Previous years should already be downloaded and available in the dashboard data repository. The newest year of data, or the data that are not already available in the repository, are the only datasets that should be downloaded, and they should be downloaded as single years. This makes the data cleaning and compiling process much easier moving forward**
     3. Geography: select Chelan, Douglas, Grant, Kittitas, and Okanogan (you can select all at once by holding down “CTRL” and selecting these counties) and also click the “Include State Total” box at the bottom
     4. Age: for Age, go to the “Create Group” tab, and move the 15-17 and 18-19 groups to the Selected Items column. Name this group “Teen Group” in the Group Name box, and then click “Create Group”. This should preserve this age group of interest
     5. Gestational Age: gestational age should be “All Gestational Weeks (Combined)” automatically, but if it is not, make sure all gestational ages are selected
  4. Hit “Submit” which should take you to a Query page
  5. Scroll to the bottom and make sure data is to be exported as a CSV, and then export
  6. If you need to do this more than one, you should be able to “Revise Query” towards the top
* **Storing Data**
  1. The data should download as a generic “temp”, but rename as follows:
     1. DataSource\_Indicator\_Level\_Years
     2. Example: CHAT\_TeenagePregnancy\_All\_2016
  2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/CHAT/Teenage Pregnancy** folder. This will allow for the data cleaning process below to be automated
* **Clean Data** 
  1. You will use the “clean\_teenpreg.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Teenage Pregnancy data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Teenage Pregnancy data into the **ncw\_data/cleaned\_data** folder as **teenpreg\_data.csv**. This data file will be the one that is pulled into PowerBI. This is the same process as below, as the Rmarkdown file combines the State, individual county, and NCW Epi estimates

**Teenage Pregnancy Rate – Combined NCW Epi LHJs**

* **Download instructions** 
  1. For downloading data for this indicator, you will use the same steps as you did for the “Teenage Pregnancy Rate – County and State-Level estimates” indicator above. *However, for 3c, go to the “Create Group” tab, and select Chelan, Douglas, Grant, Kittitas, and Okanogan all at once, and create a group called “NCW”. Be sure to uncheck the “Include State Total” button. You should then use this group to create a single estimate for all 5 counties combined.* All other instructions remain the same, including the instructions for 3d (i.e., restricting to mothers 15-19 years of age)
* **Storing Data**
  1. The data should download as a generic “temp”, but rename as follows:
     + DataSource\_Indicator\_Level\_Years
     + Example: CHAT\_ TeenagePregnancy\_NCW\_2016
  2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/CHAT/Teenage Pregnancy** folder. This will allow for the data cleaning process below to be automated
* **Clean Data** 
  1. You will use the “clean\_teenpreg.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Teenage Pregnancy data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Teenage Pregnancy data into the **ncw\_data/cleaned\_data** folder as **teenpreg\_data.csv**. This data file will be the one that is pulled into PowerBI. This is the same process as above, as the Rmarkdown file combines the State, individual county, and NCW Epi estimates

**Suicide – County and State-Level estimates**

* **Download instructions** 
  1. Go to [https://secureaccess.wa.gov/](https://urldefense.com/v3/__https:/secureaccess.wa.gov/__;!!K-Hz7m0Vt54!kPH3Wvm_ulLY8UPvMf46Iq9Yo6L-pUfAME07DidUtFRafeUbbLxiP30o__O4ocpkFrOz2rQ_UzOOAy5d5OI$) and click “Access Now” next to “Community Health Assessment Tool provided by Department of Health”
  2. Select “Mortality” from the “Topics” tab
  3. You will be sent to a “Create Query” tab where you can select your query preferences
     1. Primary Statistic or Measure: “Age-Adjusted Rate”
     2. Diagnosis: Under NCHS113 Groupings, scroll down and select “Intentional self-harm (suicide)”
     3. Year: select year of interest
        + **Previous years should already be downloaded and available in the dashboard data repository. The newest year of data, or the data that are not already available in the repository, are the only datasets that should be downloaded, and they should be downloaded as single years. This makes the data cleaning and compiling process much easier moving forward**
     4. Geography: select Chelan, Douglas, Grant, Kittitas, and Okanogan (you can select all at once by holding down “CTRL” and selecting these counties) and also click the “Include State Total” box at the bottom
     5. Age: make sure “All Ages (Combined)” is selected
     6. Gender: make sure “All Genders (Combined)” is selected
     7. Race: make sure “All Races (Combined)” is selected
     8. Ethnicity: make sure “All (Combined)” is selected
  4. Hit “Submit” which should take you to a Query page
  5. Scroll to the bottom and make sure data is to be exported as a CSV, and then export
  6. If you need to do this more than one, you should be able to “Revise Query” towards the top
* **Storing Data**
  1. The data should download as a generic “temp”, but rename as follows:
     1. DataSource\_Indicator\_Level\_Years
     2. Example: CHAT\_Suicide\_All\_2016
  2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/CHAT/Suicide** folder. This will allow for the data cleaning process below to be automated
* **Clean Data** 
  1. You will use the “clean\_suicide.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Suicide data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Suicide data into the **ncw\_data/cleaned\_data** folder as **suicide\_data.csv**. This data file will be the one that is pulled into PowerBI. This is the same process as below, as the Rmarkdown file combines the State, individual county, and NCW Epi estimates

**Suicide – Combined NCW Epi LHJs**

* **Download instructions** 
  1. For downloading data for this indicator, you will use the same steps as you did for the “Suicide – County and State-Level estimates” indicator above. *However, for 3d, go to the “Create Group” tab, and select Chelan, Douglas, Grant, Kittitas, and Okanogan all at once, and create a group called “NCW”. Be sure to uncheck the “Include State Total” button. You should then use this group to create a single estimate for all 5 counties combined.* All other instructions remain the same
* **Storing Data**
  1. The data should download as a generic “temp”, but rename as follows:
     + DataSource\_Indicator\_Level\_Years
     + Example: CHAT\_Suicide\_NCW\_2016
  2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/CHAT/Suicide** folder. This will allow for the data cleaning process below to be automated
* **Clean Data** 
  1. You will use the “clean\_suicide.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Suicide data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Suicide data into the **ncw\_data/cleaned\_data** folder as **suicide\_data.csv**. This data file will be the one that is pulled into PowerBI. This is the same process as above, as the Rmarkdown file combines the State, individual county, and NCW Epi estimates

**Life Expectancy – County and State-Level estimates**

* **Download instructions** 
  1. Go to [https://secureaccess.wa.gov/](https://urldefense.com/v3/__https:/secureaccess.wa.gov/__;!!K-Hz7m0Vt54!kPH3Wvm_ulLY8UPvMf46Iq9Yo6L-pUfAME07DidUtFRafeUbbLxiP30o__O4ocpkFrOz2rQ_UzOOAy5d5OI$) and click “Access Now” next to “Community Health Assessment Tool provided by Department of Health”
  2. Select “Life Expectancy” from the “Topics” tab
  3. You will be sent to a “Create Query” tab where you can select your query preferences
     1. Primary Statistic or Measure: “Life Expectancy In Years”
     2. Year: select year of interest
        + **Previous years should already be downloaded and available in the dashboard data repository. The newest year of data, or the data that are not already available in the repository, are the only datasets that should be downloaded, and they should be downloaded as single years. This makes the data cleaning and compiling process much easier moving forward**
     3. Geography: select Chelan, Douglas, Grant, Kittitas, and Okanogan (you can select all at once by holding down “CTRL” and selecting these counties) and also click the “Include State Total” box at the bottom
     4. Age: <1
     5. Gender: make sure “All Genders (Combined)” is selected
     6. Race: make sure “All Races (Combined)” is selected
     7. Ethnicity: make sure “All (Combined)” is selected
  4. Hit “Submit” which should take you to a Query page
  5. Scroll to the bottom and make sure data is to be exported as a CSV, and then export
  6. If you need to do this more than one, you should be able to “Revise Query” towards the top
* **Storing Data**
  1. The data should download as a generic “temp”, but rename as follows:
     1. DataSource\_Indicator\_Level\_Years
     2. Example: CHAT\_LifeExpectancy\_All\_2016
  2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/CHAT/Life Expectancy** folder. This will allow for the data cleaning process below to be automated
* **Clean Data** 
  1. You will use the “clean\_lifeexp.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Life Expectancy data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Life Expectancy data into the **ncw\_data/cleaned\_data** folder as **lifeexp\_data.csv**. This data file will be the one that is pulled into PowerBI. This is the same process as below, as the Rmarkdown file combines the State, individual county, and NCW Epi estimates

**Life Expectancy – Combined NCW Epi LHJs**

* **Download instructions** 
  1. For downloading data for this indicator, you will use the same steps as you did for the “Life Expectancy – County and State-Level estimates” indicator above. *However, for 3c, go to the “Create Group” tab, and select Chelan, Douglas, Grant, Kittitas, and Okanogan all at once, and create a group called “NCW”. Be sure to uncheck the “Include State Total” button. You should then use this group to create a single estimate for all 5 counties combined.* All other instructions remain the same
* **Storing Data**
  1. The data should download as a generic “temp”, but rename as follows:
     + DataSource\_Indicator\_Level\_Years
     + Example: CHAT\_LifeExpectancy\_NCW\_2016
  2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/CHAT/Life Expectancy** folder. This will allow for the data cleaning process below to be automated
* **Clean Data** 
  1. You will use the “clean\_lifeexp.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Life Expectancy data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Life Expectancy data into the ncw\_data/cleaned\_data folder as lifeexp\_data.csv. This data file will be the one that is pulled into PowerBI. This is the same process as above, as the Rmarkdown file combines the State, individual county, and NCW Epi estimates

**Death Rate – County and State-Level estimates**

* **Download instructions** 
  1. Go to [https://secureaccess.wa.gov/](https://urldefense.com/v3/__https:/secureaccess.wa.gov/__;!!K-Hz7m0Vt54!kPH3Wvm_ulLY8UPvMf46Iq9Yo6L-pUfAME07DidUtFRafeUbbLxiP30o__O4ocpkFrOz2rQ_UzOOAy5d5OI$) and click “Access Now” next to “Community Health Assessment Tool provided by Department of Health”
  2. Select “Mortality” from the “Topics” tab
  3. You will be sent to a “Create Query” tab where you can select your query preferences
     1. Primary Statistic or Measure: “Age-Adjusted Rate”
     2. Diagnosis: from the NCHS113 Groupings in the Code Sets menu, make sure the Selected Items on the right says “All Combined”. This should be the default
     3. Year: select year of interest
        + **Previous years should already be downloaded and available in the dashboard data repository. The newest year of data, or the data that are not already available in the repository, are the only datasets that should be downloaded, and they should be downloaded as single years. This makes the data cleaning and compiling process much easier moving forward**
     4. Geography: select Chelan, Douglas, Grant, Kittitas, and Okanogan (you can select all at once by holding down “CTRL” and selecting these counties) and also click the “Include State Total” box at the bottom
     5. Age: make sure “All Ages (Combined)” is selected
     6. Gender: make sure “All Genders (Combined)” is selected
     7. Race: make sure “All Races (Combined)” is selected
     8. Ethnicity: make sure “All (Combined)” is selected
  4. Hit “Submit” which should take you to a Query page
  5. Scroll to the bottom and make sure data is to be exported as a CSV, and then export
  6. If you need to do this more than one, you should be able to “Revise Query” towards the top
* **Storing Data**
  1. The data should download as a generic “temp”, but rename as follows:
     1. DataSource\_Indicator\_Level\_Years
     2. Example: CHAT\_DeathRate\_All\_2016
  2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/CHAT/Death Rate** folder. This will allow for the data cleaning process below to be automated
* **Clean Data** 
  1. You will use the “clean\_deathrate.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Death Rate data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Death Rate data into the **ncw\_data/cleaned\_data** folder as **deathrate\_data.csv**. This data file will be the one that is pulled into PowerBI. This is the same process as below, as the Rmarkdown file combines the State, individual county, and NCW Epi estimates

**Death Rate – Combined NCW Epi LHJs**

* **Download instructions** 
  1. For downloading data for this indicator, you will use the same steps as you did for the “Death Rate – County and State-Level estimates” indicator above. *However, for 3d, go to the “Create Group” tab, and select Chelan, Douglas, Grant, Kittitas, and Okanogan all at once, and create a group called “NCW”. Be sure to uncheck the “Include State Total” button. You should then use this group to create a single estimate for all 5 counties combined.* All other instructions remain the same
* **Storing Data**
  1. The data should download as a generic “temp”, but rename as follows:
     + DataSource\_Indicator\_Level\_Years
     + Example: CHAT\_DeathRate\_NCW\_2016
  2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/CHAT/Death Rate** folder. This will allow for the data cleaning process below to be automated
* **Clean Data** 
  1. You will use the “clean\_deathrate.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Death Rate data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Death Rate data into the **ncw\_data/cleaned\_data** folder as **deathrate\_data.csv**. This data file will be the one that is pulled into PowerBI. This is the same process as above, as the Rmarkdown file combines the State, individual county, and NCW Epi estimates

**Causes of Death – County and State-Level estimates**

* **Download instructions** 
  1. Go to [https://secureaccess.wa.gov/](https://urldefense.com/v3/__https:/secureaccess.wa.gov/__;!!K-Hz7m0Vt54!kPH3Wvm_ulLY8UPvMf46Iq9Yo6L-pUfAME07DidUtFRafeUbbLxiP30o__O4ocpkFrOz2rQ_UzOOAy5d5OI$) and click “Access Now” next to “Community Health Assessment Tool provided by Department of Health”
  2. Select “Mortality” from the “Topics” tab
  3. You will be sent to a “Create Query” tab where you can select your query preferences
     1. Primary Statistic or Measure: “Age-Adjusted Rate”
     2. A screenshot of a computer

        Description automatically generatedDiagnosis: In the “Code Sets” drop down menu, select “Leading Causes”. Click “Add Top Level” which should pull all of the causes on the death on the left to the right menu under “Selected Items”. *Of note: make sure that the subitems underneath the causes of death are not selected. For example, for “Major cardiovascular disease”, only move the main item to the “Selected Items”, i.e., not “Diseases of heart” or “Atherosclerosis”*
     3. Year: select year of interest
        + **Previous years should already be downloaded and available in the dashboard data repository. The newest year of data, or the data that are not already available in the repository, are the only datasets that should be downloaded, and they should be downloaded as single years. This makes the data cleaning and compiling process much easier moving forward**
     4. Geography: select Chelan, Douglas, Grant, Kittitas, and Okanogan (you can select all at once by holding down “CTRL” and selecting these counties) and also click the “Include State Total” box at the bottom
     5. Age: make sure “All Ages (Combined)” is selected
     6. Gender: make sure “All Genders (Combined)” is selected
     7. Race: make sure “All Races (Combined)” is selected
     8. Ethnicity: make sure “All (Combined)” is selected
  4. Hit “Submit” which should take you to a Query page
  5. Scroll to the bottom and make sure data is to be exported as a CSV, and then export
  6. If you need to do this more than one, you should be able to “Revise Query” towards the top
* **Storing Data**
  1. The data should download as a generic “temp”, but rename as follows:
     1. DataSource\_Indicator\_Level\_Years
     2. Example: CHAT\_CausesofDeath\_All\_2016
  2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/CHAT/Causes of Death** folder. This will allow for the data cleaning process below to be automated
* **Clean Data** 
  1. You will use the “clean\_cod.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Cause of Death data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Cause of Death data into the **ncw\_data/cleaned\_data** folder as **cod\_data.csv**. This data file will be the one that is pulled into PowerBI. This is the same process as below, as the Rmarkdown file combines the State, individual county, and NCW Epi estimates

**Causes of Death – Combined NCW Epi LHJs**

* **Download instructions** 
  1. For downloading data for this indicator, you will use the same steps as you did for the “Causes of Death – County and State-Level estimates” indicator above. *However, for 3d, go to the “Create Group” tab, and select Chelan, Douglas, Grant, Kittitas, and Okanogan all at once, and create a group called “NCW”. Be sure to uncheck the “Include State Total” button. You should then use this group to create a single estimate for all 5 counties combined.* All other instructions remain the same
* **Storing Data**
  1. The data should download as a generic “temp”, but rename as follows:
     + DataSource\_Indicator\_Level\_Years
     + Example: CHAT\_CausesofDeath\_NCW\_2016
  2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/CHAT/Causes of Death** folder. This will allow for the data cleaning process below to be automated
* **Clean Data** 
  1. You will use the “clean\_cod.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Cause of Death data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Cause of Death data into the **ncw\_data/cleaned\_data** folder as **cod\_data.csv**. This data file will be the one that is pulled into PowerBI. This is the same process as above, as the Rmarkdown file combines the State, individual county, and NCW Epi estimates

**Hospitalization – County and State-Level estimates**

* **Download instructions** 
  1. Go to [https://secureaccess.wa.gov/](https://urldefense.com/v3/__https:/secureaccess.wa.gov/__;!!K-Hz7m0Vt54!kPH3Wvm_ulLY8UPvMf46Iq9Yo6L-pUfAME07DidUtFRafeUbbLxiP30o__O4ocpkFrOz2rQ_UzOOAy5d5OI$) and click “Access Now” next to “Community Health Assessment Tool provided by Department of Health”
  2. Select “Hospitalization” from the “Topics” tab
  3. You will be sent to a “Create Query” tab where you can select your query preferences
     1. Primary Statistic or Measure: “Age-Adjusted Rate”
     2. Topic Filters: Do not select either option, leave default setting
     3. Diagnosis: In the “Code Sets” drop down menu, select “ICD10cm Raw Codes”. Click “Add Top Level” which should pull all of the causes on the death on the left to the right menu under “Selected Items”. *Of note: make sure that the subitems underneath the causes of hospitalization are not selected*
     4. Year: select year of interest
        + **Previous years should already be downloaded and available in the dashboard data repository. The newest year of data, or the data that are not already available in the repository, are the only datasets that should be downloaded, and they should be downloaded as single years. This makes the data cleaning and compiling process much easier moving forward**
     5. Geography: select Chelan, Douglas, Grant, Kittitas, and Okanogan (you can select all at once by holding down “CTRL” and selecting these counties) and also click the “Include State Total” box at the bottom
     6. Age: make sure “All Ages (Combined)” is selected
     7. Gender: make sure “All Genders (Combined)” is selected
  4. Hit “Submit” which should take you to a Query page
  5. Scroll to the bottom and make sure data is to be exported as a CSV, and then export
  6. If you need to do this more than one, you should be able to “Revise Query” towards the top
* **Storing Data**
  1. The data should download as a generic “temp”, but rename as follows:
     1. DataSource\_Indicator\_Level\_Years
     2. Example: CHAT\_Hospitalization\_All\_2016
  2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/CHAT/Hospitalizations** folder. This will allow for the data cleaning process below to be automated
* **Clean Data** 
  1. You will use the “clean\_hosp.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Hospitalizations data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Hospitalizations data into the **ncw\_data/cleaned\_data** folder as **hosp\_data.csv**. This data file will be the one that is pulled into PowerBI. This is the same process as below, as the Rmarkdown file combines the State, individual county, and NCW Epi estimates

**Hospitalization – Combined NCW Epi LHJs**

* **Download instructions** 
  1. For downloading data for this indicator, you will use the same steps as you did for the “Hospitalization – County and State-Level estimates” indicator above. *However, for 3e, go to the “Create Group” tab, and select Chelan, Douglas, Grant, Kittitas, and Okanogan all at once, and create a group called “NCW”. Be sure to uncheck the “Include State Total” button. You should then use this group to create a single estimate for all 5 counties combined.* All other instructions remain the same
* **Storing Data**
  1. The data should download as a generic “temp”, but rename as follows:
     + DataSource\_Indicator\_Level\_Years
     + Example: CHAT\_Hospitalization\_NCW\_2016
  2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/CHAT/Hospitalizations** folder. This will allow for the data cleaning process below to be automated
* **Clean Data** 
  1. You will use the “clean\_hosp.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Hospitalizations data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Hospitalizations data into the **ncw\_data/cleaned\_data** folder as **hosp\_data.csv**. This data file will be the one that is pulled into PowerBI. This is the same process as above, as the Rmarkdown file combines the State, individual county, and NCW Epi estimates

**Overdose Deaths – County and State-Level estimates**

* **Data Access Instructions:** As of early 2024, the “Overdose Death Data” is no longer stored directly in the CHAT data system. Instead, there is a dashboard *within CHAT* that you need to request additional access to. You will need to reach out to [chs.chat@doh.wa.gov](mailto:chs.chat@doh.wa.gov) and [cds-dashboards@doh.wa.gov](mailto:cds-dashboards@doh.wa.gov) asking for permission to access the access to the "Internal Opioid and Drug Overdose Data Dashboard". Someone from DOH (as of April 2024, that was Bert Plants, PhD; [bert.plants@doh.wa.gov](mailto:bert.plants@doh.wa.gov)) will send you a confidentiality agreement to use the Opioid Dashboard. Sign the confidentiality agreement and then send it back to them. Follow-up on access as needed
* **Download Instructions 🡪 do not exist yet as DOH has not given access permission**

# American Community Survey (ACS)

The American Community Survey Data can be access in two ways. First is by downloading Excel files from the American Community Survey Website. The second, and preferred way, is to pull ACS data straight into R using the *tidycensus* package. To run *tidycensus*, you will need an API key issued by the census department. Here is a link with information on how to get a key and install it: <https://search.r-project.org/CRAN/refmans/tidycensus/html/census_api_key.html>. The website where the API key can be requested is here: <https://api.census.gov/data/key_signup.html>. Once you obtain the key you store the key in R using the `census\_api\_key` R function. This should only need to be done once.

A diagram of a data flow

Description automatically generated

As we are using *tidycensus* to pull in data, there isn’t much additional work that is needed to download the data, since it will be pulled directly into R and the data management and restructuring for PowerBI will be automated in the using the “clean\_american\_community\_survey.Rmd” file in the “cleaning\_scripts” folder. Additionally, ACS data files can be extremely large if downloaded locally onto your computer, so pulling data in through R is preferred. When using *tidycensus*, you will pull in the data you’re interested in through variable codes. Note that codes sometimes change year to year, and therefore you’ll need to verify the codes for subsequent years. Codes can be checked via loading in all the variables. This code is already in the R scripts. The Rmarkdown file should function, for all of the indicators below to pull in, manipulate and visualize all of the indicators.

When updating dashboards for new years, edit the line of code “end\_year <- 2022” and replace “2022” with the end year that you want.

ACS in small population areas is only put out in 5-year estimates, therefore the data you’ll be pulling in is just the 5-year estimates.

## Health Insurance

## SNAP Program

## Poverty

## Disability

# CDC National Environmental Public Health Tracking Network

**Description of data source**: The National Environmental Public Health Tracking Network (Tracking Network) brings together health data and environmental data from national, state, and city sources and provides supporting information to make the data easier to understand. The Tracking Network has data and information on environments and hazards, health effects, and population health.

**Description of measure**: The indicator from this data source is **Air Quality (PM2.5: Percent of Days over Air Quality Standard (Monitor + Modeled Data)).** Air quality can be measured and defined in a number of ways, but we are using “PM2.5: Percent of Days over Air Quality Standard (Monitor + Modeled Data).” This indicator is defined as the annual percent of days over the 24-hr National Ambient Air Quality Standards (NAAQS) for PM2.5, calculated by summing the number of days each year in which the maximum PM2.5 concentration within each county exceeded the [24-hr PM2.5 NAAQS of 35 µg/m3](https://www.epa.gov/criteria-air-pollutants/naaqs-table). Because many US counties do not have sufficient air quality data to derive PM2.5 data from monitor data alone (i.e., measured data), CDC National Environmental Public Health Tracking Network fills in these gaps by supplementing with modeled estimates of PM2.5 concentrations. These modeled estimates are derived from the Downscaler (DS) model. This indicator includes solely monitor (measured) data in counties and years where there is sufficient monitor data and DS modeled estimates for counties that lack monitor data. Additional information on this specific indicator can be found here: <https://ephtracking.cdc.gov/indicatorPages>

**Accessing data source**: To access this data, visit this site: <https://ephtracking.cdc.gov/>

**Instructions for downloading, storing, and cleaning data**:

## Air Quality (Percent of Days over PM2.5 Air Quality Standard)

* County level estimates (only county level information available, none for all of WA state)
  + **Download Data**
    1. Go to <https://ephtracking.cdc.gov/> and then click on the button that says “Data Explorer Tool”. This should take you to the CDC Query Panel
    2. Under “STEP 1: CONTENT”, specify the following:
       - Select Content Area: Air Quality
       - Select Indicator: National Ambient Air Quality Standards (NAAQS)
       - Select Measure: PM2.5: Percent of Days over Air Quality Standard (Monitor + Modeled Data)
    3. Under “STEP 2: GEOGRAPHY TYPE”, specify State By County
    4. Under “STEP 3: GEOGRAPHY”, select Washington
    5. Under “STEP 4: TIME”, select the year you would like to download
       1. Even if you can download multiple years of data, download one year at a time for ease of data cleaning
    6. As of October 2023, there should be no “Advanced Options” for you to select under “STEP 5: ADVANCED OPTIONS”
    7. Click GO 🡪 which should result in a map of Washington counties appearing
    8. In the top right corner above the map, click on the downward arrow that will allow you the “Export Data”
    9. Repeat this process for as many years as needed
       - There should be an orange “SELECT DATA” button in the top left corner that will allow you to keep all the same settings but change just the year
  + **Store Data** 
    1. A zip file should be downloaded to your computer with a CSV file, an HTML file, and a PDF file. Transfer just the CSV file to your CDC National Environmental Public Health Tracking Network data folder. If needed, transfer the HTML and PDF file (i.e., if documentation has changed)
    2. The CSV will download with the general name “data\_XXXXXX”
       - Rename CSV in the following format:
         * DataSource\_Indicator\_Level\_Year
         * Example: CDCNEPHTN\_AirQual24hr\_All\_2015
    3. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/CDC National Environmental Public Health Tracking Network/Air Quality 24 hr** folder. This will allow for the data cleaning process below to be automated
  + **Clean Data** 
    1. You will use the “clean\_airqual.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Air Quality data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Air Quality data into the **ncw\_data/cleaned\_data** folder as **airqual\_data.csv**. This data file will be the one that is pulled into PowerBI

# Department of Children, Youth and Family (DCYF)

**Description of data source**: Data on child abuse and neglect comes from the Washington State Department of Children, Youth & Families (DCYF), but the actual data extracted for this indicator comes from the [Child Well-Being Data Portal](https://portal.cssat.org/about), joint project of the Center for Social Sector Analytics & Technology (CSSAT) and Partners for Our Children (POC). The Child Well-Being Data Portal has access to quarterly deidentified data extracts from FamLink, the case management information system used by DCYF. They also use data from the US Census Bureau, WA OFM, the National Data Archive of Child Abuse and Neglect (NDACAN), as needed.

**Description of measure** The indicators from this data source is **Child Abuse or Neglect Referrals**. This indicator shows “investigations and assessments of reported child abuse and/or neglect. It shows how many households, per 1,000, had at least one screened-in report.” More information on this indicator can be found [here](https://portal.cssat.org/visualizations/ia-rates). This information can also be found in the “INFO” tab of Child Well-Being Data Portal dashboard (explained below).

**Accessing data source**: To access this data, visit this site: <https://portal.cssat.org/>

**Instructions for downloading, storing, and cleaning data**:

## Child Abuse or Neglect Referrals

* County and State level estimates (one data download for both)
  + **Download Data**
    1. Go do <https://portal.cssat.org/>
    2. Scroll down to the Explore visualizations section, find the box that says Investigations & Assessments (Rate) and click “VIEW”. This should send you to a graph of Investigations & Assessments (Rate)
    3. On the left side of the screen, you will see a list of filters. For Display, select “All Investigations & Assessments” (this will likely be the default/already selected)
    4. For Date & Time, make sure “Year” is selected
       - Unfortunately, these data cannot be downloaded year by year at this time. Even if you select just a range of time you are interested at this step, when you go to download a CSV, all years of data will pull into the data file
    5. For Demographics
       - Ensure that only “All Households” is selected under “Age of Children in Household”
       - Ensure that only “All” is selected under “Race/Ethnicity”
    6. Under Location, select the **Chelan, Douglas, Grant, Kittitas, and Okanogan counties, as well as “All” (very end of list)**. In total, you should have 6 boxes checked (5 counties and 1 “All” which reports statewide data)
    7. Scroll back up and click “Update”. You should now see 6 lines of data (one for each county and one for the entire state)
    8. From the ribbon across the top, select “DOWNLOAD”, and download as “DATA (.CSV OR .TXT)
  + **Store Data** 
    1. The file will download with the general name “Investigations\_&\_Assessments\_(Rate)” regardless of the years selected. Rename in the following format:
       - DataSource\_Indicator\_Level\_Years
       - Example: DCYF\_AbuseNeglect\_All\_2006\_2019
    2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/DCYF/Abuse and Neglect** folder. This will allow for the data cleaning process below to be fairly automated
  + **Clean Data** 
    1. You will use the “clean\_abuseneg.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Abuse and Neglect data. Before you run this Rmarkdown file, swap out the new name of the data in the “file” object path with the new data. You will likely need to change it from “DCYF\_AbuseNeglect\_All\_2006\_***2019***” to “DCYF\_AbuseNeglect\_All\_2006\_***2020***” or whatever the newest year is
    2. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Abuse and Neglect data into the **ncw\_data/cleaned\_data** folder as **abuseneg\_data.csv**. This data file will be the one that is pulled into PowerBI

# Healthy Youth Survey (HYS)

**Description of data source**: The Healthy Youth Survey (HYS) is a biennial survey administered to 6th-12th grade students across Washington State. It is a collaborative effort across many different agencies and contractors. Its purpose is to measure health risk behaviors, such as substance use, mental health, and violence, that contribute to illness, death, and social problems among youth across out state. More can be learned about this data source and find some FAQs [here](https://www.askhys.net/About), and HYS Survey Questionnaires can be found [here](https://www.askhys.net/Resources/SurveyQuestionnaires).

These data are updated every 2 years, as the HYS is only conducted every 2 years. However, there *was no survey in 2020 due to the COVID-19 pandemic*, so the data goes from 2016, 2018 to 2021. The HYS has county and state level data available, but are accessed differently as of the writing of this report. Chelan and Douglas are separated out.

Demographic information is not available in the spreadsheets sent by Dr. Simckes (described below), but some demographic information is available on the [HYS Data Dashboard](https://www.askhys.net/SurveyResults/DataDashboard) and could be extracted in the future.

The indicators from this data source are…

1. **Substance Use among 10th graders** 
   * The indicator we are interested in is “percent of 10th graders who report using a substance (cigarettes, alcohol, marijuana, e-cigarettes and all other illegal substances) in the last 30 days”. There are 6 variables (for 2016, 2018, and 2021 HYS):
     1. D20: alcohol
     2. D14: cigarettes
     3. D90\_16: e-cigarettes
     4. D21\_16: marijuana
     5. D15: chewing tobacco, snuff, or dip (other tobacco)
     6. D63: another illegal drug (not counting alcohol, tobacco, or marijuana)

In 2021, the wording of some of the 30-day substance use questions were altered, which means that the *ReportVariableLabel* variable in the HYS 2021 data is slightly different. The differences are outlined in the table below. However, the *Variable* variable remained the same throughout.

|  |  |  |
| --- | --- | --- |
| **Variable** | **ReportVariableLabel** | **Years** |
| D20 | During the past 30 days, on how many days did you drink a glass, can or bottle of alcohol (beer, wine, wine coolers, hard liquor)? | 2016, 2018 |
| Drink a glass, can or bottle of alcohol? | 2021 |
| D14 | During the past 30 days, on how many days did you smoke cigarettes? | 2016, 2018, 2021 |
| D90\_16 | During the past 30 days, on how many days did you use an electronic cigarette, also called e-cigs, or vape pens? | 2016, 2018 |
| During the past 30 days, on how many days did you use an electronic cigarette, also called e-cigs, JUUL, or vape pens? | 2021 |
| D21\_16 | During the past 30 days, on how many days did you use marijuana or hashish (weed, hash, pot)? | 2016, 2018 |
| Use marijuana or hashish? | 2021 |
| D15 | During the past 30 days, on how many days did you use chewing tobacco, snuff, or dip? | 2016, 2018, 2021 |
| D63 | During the past 30 days, on how many days did you not counting alcohol, tobacco, or marijuana, use another illegal drug? | 2016, 2018 |
| Not counting alcohol, tobacco, or marijuana, use another illegal drug? | 2021 |

1. **Suicidal Ideation among 10th graders**
   * The indicator we are interested in is “percent of 10th graders who reported seriously considering attempting suicide within the last year.” The variable (for 2016, 2018, and 2021 HYS) was H54
   * The question in the HYS is asked as “During the past 12 months, did you ever seriously consider attempting suicide?” with answer options of “Yes” and “No”
2. **Youth Dental Care among 10th graders**
   * The indicator we are interested in is “percent of 10th graders who report seeing a dentist for a check-up, exam, teeth cleaning or other dental work in the last year.” The variable (for 2016, 2018, and 2021 HYS) was H25
   * The question in the HYS is asked as “When was the last time you saw a dentist for a check-up, exam, teeth cleaning, or other dental work?” with answer options of “During the past 12 months”, “Between 12 and 24 months ago”, “More than 24 months ago”, “Never”, and “Not Sure”

**Accessing data source**: To access this county-level data, you will need to reach out to the Epi that runs the HYS. As of late 2023, that was Maayan Simckes, PhD ([maayan.simckes@doh.wa.gov](mailto:maayan.simckes@doh.wa.gov)).

**Instructions for downloading and storing**:

Most recently, Dr. Simckes solely sent the county level data files for each year. Within these excel documents, each grade of student will have its own tab within the excel document. You will need to extract the 10th grade tab for data cleaning and visualization. The process for county level and state level are described below.

## Substance Use

* County level estimates
  + **Download data** 
    1. Email [maayan.simckes@doh.wa.gov](mailto:maayan.simckes@doh.wa.gov) to request county-level estimate for the HYS year of data your need. It is helpful to explain the purpose of your request (NCW Epi dashboards), and what variables you are using for context (see above). If Dr. Simckes is no longer the point of contact, email [healthy.youth@doh.wa.gov](mailto:healthy.youth@doh.wa.gov) or browse the website for a new contact
  + **Store Data** 
    1. Store the data as “County-level HYS (YEAR) Analysis File” (likely will not have to change anything). Make sure it is stored in the **ncw\_data/raw\_data/HYS** folder
* State level estimates
  + **Download data** 
    - Alcohol
      1. Go to <https://www.askhys.net/SurveyResults/DataDashboard>, the HYS dashboard feature
      2. On the righthand side, select “10th Grade”
      3. For location, select any location (the statewide information is available on all dashboards)
      4. For topic, select “Alcohol” and for measure, select “Current Alcohol Drinking”
      5. In the top right corner, select “Data View” for easier access to statewide values
      6. Follow the Store Data instructions below, and be sure that you put the values into the correct tab on the CSV
    - Cigarettes
      1. Follow the instructions above laid out for alcohol, but for Step 4: topic should be “Tobacco” and measure should be “Current Cigarette Smoking”. Store data per instructions in Step 6
    - E-cigarettes
      1. Follow the instructions above laid out for alcohol, but for Step 4: topic should be “Tobacco” and measure should be “Current E-Cigarette/Vape Use”. Store data per instructions in Step 6
    - Marijuana
      1. Follow the instructions above laid out for alcohol, but for Step 4: topic should be “Marijuana” and measure should be “Current Marijuana Use”. Store data per instructions in Step 6
    - Other tobacco
      1. Follow the instructions above laid out for alcohol, but for Step 4: topic should be “Tobacco” and measure should be “Current Smokeless Tobacco Use”. Store data per instructions in Step 6
    - Another illegal drug
      1. Follow the instructions above laid out for alcohol, but for Step 4: topic should be “Other Drugs” and measure should be “Current Other Drug Use (Not A, T, M)”. Store data per instructions in Step 6
  + **Store Data** 
    1. For the newest year of data, pull in the “State %” value from the dashboard into the “Percent” column in the “State-level HYS” Excel file in the data folder. Also extract the “State +/-” for the year of interest and put it into the CI (confidence interval) column in the Excel file. Note that there is a different tab along the bottom of the Excel sheet for each indicator we’re interested in. Save the Excel sheet. Make sure it is stored in the **ncw\_data/raw\_data/HYS** folder

## Suicidal Ideation

* County level estimates
  + These instructions are the same as above for Substance Use. If you have completed those steps above, nothing further is needed, as all county level estimates come in the same Excel file. Make sure it is stored in the **ncw\_data/raw\_data/HYS** folder
* State level estimates
  + **Download data** 
    1. Go to <https://www.askhys.net/SurveyResults/DataDashboard>, the HYS dashboard feature
    2. On the righthand side, select “10th Grade”
    3. For location, select any location (the statewide information is available on all dashboards)
    4. For topic, select “Mental Health” and for measure, select “Contemplation of Suicide”
    5. In the top right corner, select “Data View” for easier access to statewide values
  + **Store Data** 
    1. For the newest year of data, pull in the “State %” value from the dashboard into the “Percent” column in the “State-level HYS” Excel file in the data folder. Also extract the “State +/-” for the year of interest and put it into the CI (confidence interval) column in the Excel file. Note that there is a different tab along the bottom of the Excel sheet for each indicator we’re interested in. Save the Excel sheet. Make sure it is stored in the **ncw\_data/raw\_data/HYS** folder

## Youth Dental Care

* County level estimates
  + These instructions are the same as above for Substance Use. If you have completed those steps above, nothing further is needed, as all county level estimates come in the same Excel file. Make sure it is stored in the **ncw\_data/raw\_data/HYS** folder
* State level estimates
  + **Download data** 
    1. There is no dashboard available for dental care estimates, so we will need to go into the report to extract these 10th grade dental care estimates.
    2. Go to <https://www.askhys.net/SurveyResults> and under “Survey Results” select “QxQ Analysis”
    3. In the “Analysis Parameters” tab, select the year you’re interested in, both sexes, *Grade 10*, and make sure “State” is selected. Click “Continue to Query Builder”
    4. Under “Select a data category…” select “Access to Care” and find the variable H25 “Access to Dental Care”. Drag this H25 variable into the “Row Variables” selection. In the “Description” box, select “Surveyed”. Hit “Submit”
    5. The variable we are interested in here is “During the past 12 months” since we would like to know what percentage of 10th graders have seen a dentists recently
  + **Store Data** 
    1. For the newest year of data, pull in the “State %” value from the dashboard into the “Percent” column in the “State-level HYS” Excel file in the data folder. Also extract the “State +/-” for the year of interest and put it into the CI (confidence interval) column in the Excel file. Note that there is a different tab along the bottom of the Excel sheet for each indicator we’re interested in. Save the Excel sheet. Make sure it is stored in the **ncw\_data/raw\_data/HYS** folder

**Instructions for cleaning data**: These are instructions for using the Rmd file for cleaning all HYS data. You will use the “clean\_hys.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw HYS data. Again, make sure all the raw HYS data is in the **ncw\_data/raw\_data/HYS** folder. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted HYS data into the **ncw\_data/cleaned\_data** folder as:

* youthalc\_data.csv
* youthcig\_data.csv
* youthecig\_data.csv
* youthmarijuana\_data.csv
* youthtobacco\_data.csv
* youthothdrugs\_data.csv
* youthsi\_data.csv
* youthdent\_data.csv

These data files will be the one that is pulled into PowerBI.

# CDC PLACES

**Description of data source**: CDC PLACES is a collaboration between CDC, the Robert Wood Johnson Foundation, and the CDC Foundation. It is an expansion of the [500 Cities Project](https://www.cdc.gov/places/about/500-cities-2016-2019/index.html), which was the original CDC initiative collecting the same health variables but for the 500 largest cities in the US (ended in 2019). PLACES reports county, place, census tract, and ZCTA data and uses small area estimation methods to obtain measures for the entirety of the US. [More information on this data source can be found here](https://www.cdc.gov/places/about/index.html). The main data source for the majority of health variables of interest in CDC PLACES come from the [Behavioral Risk Factor Surveillance System (BRFSS)](https://www.cdc.gov/brfss/index.html).

The indicator from this data source is **Stroke among adults aged ≥18 years**. This indicator falls under the CDC PLACES umbrella of “Health Outcomes.” The variable is defined as the *age-adjusted prevalence of stroke* in the county, or the age-adjusted lifetime self-reported prevalence of stroke in the county, for a given year. More information can be found [at this site](https://www.cdc.gov/places/measure-definitions/health-outcomes/index.html#stroke)

As of end of 2023, only county-level estimates are available for this stroke prevalence indicator, i.e., state-level estimates of stroke prevalence will not be included on dashboards as they are not made available by this data source.

**Accessing data source**: To access this data, visit this site: <https://www.cdc.gov/places/index.html>

**Instructions for downloading, storing, and cleaning data**:

## Stroke

* County-level estimates (no state-level available in CDC PLACES)
  + **Download Data**
    1. Go to <https://www.cdc.gov/places/index.html> and click on “Data Portal”
    2. In the bar above the header that says “ Welcome to the 500 Cities & PLACES Data Portal”, search for the following: “PLACES: Local Data for Better Health, County Data”. Many results will come up, but we want the following datasets:
       - PLACES: Local Data for Better Health, County Data **YEAR** release
    3. Click on the dataset you are interested in downloading
    4. **A screenshot of a computer

       Description automatically generated**In the top right corner, click “Export” and download as a “CSV” F
       - CDC PLACES does not have a user interface that allows users to download subsets of data, meaning you’ll have to download the entire dataset for the United States, and then we will restrict the data we use in the data cleaning process (described below)
  + **Store Data**
    1. The file will download with the name “PLACES\_\_Local\_Data\_for\_Better\_Health\_\_County\_Data\_**YEAR**\_release\_**DATE OF DOWNLOAD**” regardless of the year. Rename in the following format for consistency:
       - DataSource\_Indicator\_Level\_Years
       - Example: PLACES\_Stroke\_Counties\_2023
    2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/PLACES** folder. This will allow for the data cleaning process below to be automated
  + **Clean Data** 
    1. You will use the “clean\_stroke.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Stroke data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Stroke data into the **ncw\_data/cleaned\_data** folder as **stroke\_data.csv**. This data file will be the one that is pulled into PowerBI

# Washington State Department of Commerce Point in Time Count

**Description of data source**: The Annual Point in Time Count is conducted yearly in Washington. It is a statewide count of all the persons staying in temporary housing programs (sheltered count) and places not meant for human habitation (unsheltered count). You can learn more about the Annual Point in Time Count in Washington state [here](https://www.commerce.wa.gov/serving-communities/homelessness/annual-point-time-count/).

Generally speaking, much of these data are suppressed especially in the counties of the NCW Epi consortium. Per Brian Fullerton: **any numbers less than 11 were suppressed for confidentiality of people put into the Homeless Management Information System (HMIS), as well as any total values where you can easily figure out that number through subtraction.**

The indicator from this data source is **Persons experiencing homelessness per 10,000 population (Sheltered and Unsheltered).** We are interested in the rate per 10,000 persons experiencing homelessness estimated from the PIT Count, both sheltered and sheltered. The PIT data only provides raw value for each geography, but the NCW Epi team opted to standardize these values using American Community Survey estimates (described below).

**Levels of data availability for download**:

* Homelessness: both county-level and state-level (total) estimates of people experiencing homelessness from the Point In Time Count are included in this data source. As noted above: any numbers less than 11 were suppressed for confidentiality of people put into the Homeless Management Information System (HMIS), as well as any total values where you can easily figure out that number through subtraction.

**Instructions for downloading, storing, and cleaning data**:

## Homelessness

Due to the variable nature of how the data is reported year to year, a full data cleaning script was not created. When new annual data is available, it should be entered into the cleaned data file by hand, and then combined with state and county level population estimates to standardize the homelessness count estimates, as the PIT count data solely reports raw numbers.

First, you will need to access the data. There are a few ways that PIT count data can be accessed. 

* The first option is to access PDFs here: <https://deptofcommerce.app.box.com/s/ek9pu2w07oz8d77gq6c1rlpxuwcw0515>
* The second option is to reach out to Brian Fullerton at the WA State Department of Commerce requesting these data in Excel format: [brian.fullerton@commerce.wa.gov](mailto:brian.fullerton@commerce.wa.gov)
  + This is a preferred format/methods for storage and documentation purposes

From these spreadsheets or PDFs, identify the value for ***total persons experiencing homelessness for WA state and each county***. Enter the new data into the cleaned data sheet following the format below. You will make new rows for each county and the state total for the year you are reporting on (5 new rows per year)

* **Location**: State or County (Chelan-Douglas, Grant, Kittitas, Okanogan, WA)
  + NOTE: Chelan-Douglas should always be combined into a single estimate for consistency across years
* **Year**: Year which the data refers to
* **Value**: The total number of persons experiencing homelessness for that location and year
* **Plus\_Minus**: This cell will be blank unless a total was suppressed. If suppression occurred (usually denoted by a \* where the total value should be), please refer to the instructions for a Suppressed Total below
* **Indicator**: homelessness

**Example:**

Here is an example, using the data sheet for 2022. Relevant county values are highlighted in purple. First the county totals:

**A screenshot of a spreadsheet

Description automatically generated**

Note: Chelan and Douglas are sometimes reported together, sometimes separately. In cases where they are reported separately, *sum their totals for a Chelan-Douglas value and only enter this row* into the cleaned data sheet.

State value (also highlighted in purple).

A screenshot of a computer

Description automatically generated

**Clean Data Sheet Format**

(2022 date highlighted in purple from the example above)

A table with numbers and names

Description automatically generated

Double Check: Because of the possibility of human error in hand entering data, it is advised to have a second person double check the work.

Suppressed Total: Some county totals are suppressed. In all cases through the writing of these instructions (April 2024), we are able to estimate a total within ±10 of the actual value. After consulting with the NCW Epi Consortium, it was determined that the midpoint of the estimate would be used for dashboards and the ± would be denoted. Because the data is reported in different formats over the years, various examples with instructions are included to serve as guidelines for future data entry.

**Example 1: Kittitas 2023**

The totals for 2023 are a sum of sheltered and unsheltered persons. We want to report the value of total for our dashboards but it was suppressed for Kittitas County. We know the value is between 35 and 45. Following the guidelines we established, we take the midpoint, 40, as our value for our cleaned data sheet and denote that the ± is 5.

A screenshot of a computer

Description automatically generated

**Example 2: Chelan-Douglas 2019**

There may be cases when two values that contribute to the total are unknown. In 2019, Chelan and Douglas counties were reported separately. Therefore, we need to sum their totals. 

A screenshot of a table

Description automatically generated

Both totals are suppressed. We estimate both following our midpoint protocol and report the ±.

Chelan: Chelan’s total is the sum of Persons from Households w/out minors (285), Persons from Households with minors (104), and Persons from Households with only minors (<11)

* Chelan County Total = 285 + 104 + 5(±5) = 394(±5)

Douglas: Douglas’s total is the sum of Persons from Households w/out minors (11), Persons from Households with minors (<11), and Persons from Households with only minors (0)

* Douglas County Total = 11 + 5 (±5) + 0 = 16(±5)

Chelan-Douglas: Summing these totals for Chelan-Douglas’s total

* 394(±5) + 16(±5) = 410(±10)

Combining with ACS estimates to obtain rates: Now that you have your updated “homelessness\_data” spreadsheet, you will need to combine these counts with estimates of the overall population from Chelan-Douglas, Grant, Kittitas, Okanogan, and Washington State so that you can calculate rates of homelessness per 10,000 individuals, which will allow for comparisons across years and locations. An Rmarkdown file, called "clean\_homelessness.Rmd" in the “cleaning\_scripts” folder, is provided to help you with this process. This file will pull 5-year American Community Survey (ACS) estimates from the US Census Bureau using the 'tidycensus' R package, and then combine these estimates with your PIT Count data in a systematic way. **The script was written so that the process is fairly automated, but please be sure to read the file and update things as necessary.** The final dataset that is output will be called "homelessness\_data\_popest.csv", which is what you will use in the actual dashboard, as it contains the population standardized rate of interest. The instructions for how to use this Rmarkdown file are included in the file itself.

# Washington State Immunization Information System (WA IIS)

**Description of data source**: The Washington State Immunization Information System (WA IIS) is a web-based, lifetime registry that keeps track of immunization records for people of all ages in our state. The vast majority of data in the WA IIS is entered electronically through EHRs. More information on the WA IIS can be found [here](https://doh.wa.gov/public-health-healthcare-providers/healthcare-professions-and-facilities/data-exchange/immunization-information-system).

The two indicators from this data source are **Infant Vaccination** (more information on the Infant Vaccination indicator, which falls under the umbrella of “Immunization Data” can be found [here](https://doh.wa.gov/data-and-statistical-reports/washington-tracking-network-wtn/immunization-data)) and **Kindergarten Vaccination** (more information on the Kindergarten Vaccination indicator, which falls under the umbrella of “School Immunization Data” can be found [here](https://doh.wa.gov/data-and-statistical-reports/washington-tracking-network-wtn/school-immunization)).

**Accessing data source**:

* For Infant Vaccination data, go here:   
  <https://doh.wa.gov/data-statistical-reports/washington-tracking-network-wtn/immunization-data/county-hedis-measures-dashboard>
* For Kindergarten Vaccination, go here: <https://doh.wa.gov/data-and-statistical-reports/washington-tracking-network-wtn/school-immunization/dashboard>

**Instructions for downloading, storing, and cleaning data**:

## Infant Vaccination

* County and State level estimates (one data download for both)
  + **Download Data**
    1. Go to <https://doh.wa.gov/data-statistical-reports/washington-tracking-network-wtn/immunization-data/county-hedis-measures-dashboard> and select the Data Table Download tab from the ribbon
    2. From the Data Year drop down, select the *year* of data you’re interested in
       - Even if you are looking to download multiple years of data, download one year at a time for ease of data cleaning
    3. From the Vaccine Series options, select “HEDIS Combo 10”
    4. For Vaccine Component, make sure “HEDIS Combo 10” is selected
    5. For Sex, make sure “All” is selected
    6. For Geography, de-select “All”, and then select all of the following:
       - Chelan, Douglas, Grant, Kittitas, Okanogan, and Statewide
       - A close-up of a computer screen

         Description automatically generatedOnce you do this, you should see 6 rows of data in total
    7. At the bottom of the screen, click on the download symbol that looks like this 🡪
    8. Select “Crosstab”, and then under Select Format click the “Excel” option
       - We normally use CSV file format, but for this specific indicator, we have found that the CSV file format was downloading with some issues
    9. Repeat this process for all of the years of data you need
  + **Store Data**
    1. The file will download with the general name “HEDIS Export Worksheet” regardless of the year. Rename in the following format:
       - DataSource\_Indicator\_Level\_Years
       - Example: WAIIS\_InfantVax\_All\_2018
    2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/WA IIS/Infant Vaccination** folder. This will allow for the data cleaning process below to be automated
  + **Clean Data** 
    1. You will use the “clean\_infvax.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Infant Vaccination data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Infant Vaccination data into the **ncw\_data/cleaned\_data** folder as **infvax\_data.csv**. This data file will be the one that is pulled into PowerBI

## Kindergarten Vaccination

* County level estimates
  + **Download data** 
    1. Go to <https://doh.wa.gov/data-and-statistical-reports/washington-tracking-network-wtn/school-immunization/dashboard> and select the County Data Download tab from the ribbon
    2. For School Year, select the years you are interested in
       1. Even if you are looking to download multiple years of data, download one year at a time for ease of data cleaning
    3. For Disease, select “Overall”
    4. For Grade, select “Kindergarten”
    5. Select **Chelan, Douglas, Grant, Kittitas, and Okanogan counties**. You can select all at once by holding down ‘Ctrl’ as you select the counties
    6. A screenshot of a data

       Description automatically generatedHover over one of the counties you selected and click “Keep Only”. This will restrict the data that are shown to our 5 counties of interest
    7. At the bottom of the screen, click on the download symbol
    8. Select “Crosstab”, and then under Select Format click the “Excel” option
       1. We normally use CSV file format, but for this specific indicator, we have found that the CSV file format was downloading with some issues
    9. Repeat this process for all of the years of data you need
  + **Store Data** 
    1. The file will download with the general name “County Data Download” regardless of the year. Rename in the following format:
       1. DataSource\_Indicator\_Level\_Years
       2. Example: WAIIS\_KindergartenVax\_County\_2016\_2017
    2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/WA IIS/** **Kindergarten Vaccination** folder. This will allow for the data cleaning process below to be automated
  + **Clean Data** 
    1. You will use the “clean\_kindvax.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Kindergarten Vaccination data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Kindergarten Vaccination data into the **ncw\_data/cleaned\_data** folder as **kindvax\_data.csv**. This data file will be the one that is pulled into PowerBI. These instructions are the same as below because the Rmarkdown file cleans both the state and county level estimates
* State level estimates
  + **Download Data**
    1. Go to <https://doh.wa.gov/data-and-statistical-reports/washington-tracking-network-wtn/school-immunization/dashboard> and select the Yearly Trends tab from the ribbon
    2. In the drop-down menu under Grade, select “Kindergarten”
       1. We are interested in the “Kindergarten Immunization Status by School Year” figure. There are a few ways to pull data from this dashboard, but the way described here allows for the easiest formatting for data visualization
    3. Select the year of data for “Immunization Status: Complete” you are interested in. At the time of writing these instructions, that was the darkest blue line in the “Kindergarten Immunization Status by School Year” figure. Once selected, you should see your dot outlined in black
       1. Even if you are looking to download multiple years of data, download one year at a time for ease of data cleaning
    4. A screenshot of a computer

       Description automatically generatedHover over one of the dots, and select the “View Data…” tab (highlighted with red box below)

* + 1. This will open up a new window. Click “Download” in the top right corner, which will download a CSV file
    2. Repeat this process for all of the years of data you need
  + **Store Data** 
    1. The file will download with the general name “Yearly Complete\_data”. Rename in the following format:
       1. DataSource\_Indicator\_Level\_Years
       2. Example: WAIIS\_KindergartenVax\_State\_2019\_2020
    2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/WA IIS/** **Kindergarten Vaccination** folder. This will allow for the data cleaning process below to be automated
  + **Clean Data**
    1. You will use the “clean\_kindvax.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Kindergarten Vaccination data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Kindergarten Vaccination data into the **ncw\_data/cleaned\_data** folder as **kindvax\_data.csv**. This data file will be the one that is pulled into PowerBI. These instructions are the same as above because the Rmarkdown file cleans both the state and county level estimates

# Health Resources & Services Administration (HRSA)

**Description of data source**: The data here is from the HRSA Area Health Resources Files (AHRF). This data source includes information on healthcare professionals, facilities, economics, hospital utilization, and similar topics at the state and county level for the entire US. These data started being collected in 2019.

The two indicators from this data source are **Access to Dentists** and **Access to Physicians.**

As a note, when these data are downloaded at both the county and state level, many rows of data will not be needed (i.e., Washington counties and states will be included). More information on definitions and methods can be found here: <https://data.hrsa.gov/Content/Documents/topics/AHRF%20Definition.pdf>

**Accessing data source**: To access this data, visit this site:   
<https://data.hrsa.gov/topics/health-workforce/ahrf>

**Demographics availability caveat for this data source**: There does not appear to be demographic specific information available. While there is some demographic information seemingly available, it is not meaningful. For example, the “Population, Subcategory” just changes the denominator, which does not represent differential access by race, ethnicity, poverty level, etc.

**Instructions for downloading, storing, and cleaning data**:

## Access to Dentists

* County level estimates
  + **Download Data** 
    1. Go to <https://data.hrsa.gov/topics/health-workforce/ahrf> and scroll down to “Explore Workforce Data”
    2. Ensure you have “Clinicians” selected at the top
    3. For Data Source, select the time period of interest
       1. Even if you can download multiple years of data, download one year at a time for ease of data cleaning
    4. For Health Profession, select “Dentist (County Level File)”
    5. For Health Profession Subcategory, select “Total Active Dentist”
    6. For Population Category, select “All (County Level File)”
    7. For Population Subcategory*,* select “Population, All”
    8. For State, select “Washington”
    9. Hit submit
    10. You’ll see an Export Data option at the bottom of the dashboard. Export data as an “XLSX” file
    11. Repeat this process for all of the years of data you need
  + **Store Data** 
    1. The file will download with the general name “AHRFDashboard” regardless of the statistic of interest. Rename in the following format:
       1. DataSource\_Indicator\_Level\_Years
       2. Example: HRSA\_Dentists\_County\_2021\_2022
    2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/HRSA/Dentists** folder. This will allow for the data cleaning process below to be automated
  + **Clean Data** 
    1. You will use the “clean\_dentists.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Access to Dentists data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Access to Dentists data into the **ncw\_data/cleaned\_data** folder as **dentists\_data.csv**. This data file will be the one that is pulled into PowerBI. These instructions are the same as below because the Rmarkdown file cleans both the state and county level estimates
* State level estimates
  + **Download Data** 
    1. Repeat “Access to Dentists, County level estimates, Download Data” Steps 1-6 from above
    2. ***However***, for Step 8, select “All States” for State
    3. Hit submit
    4. Export data as an “XLSX” file
    5. Repeat this process for all of the years of data you need
  + **Store Data** 
    1. Again, the file will download with the general name “AHRFDashboard” regardless of the statistic of interest. Rename in the following format:
       1. DataSource\_Indicator\_Level\_Years
       2. Example: HRSA\_Dentists\_State\_2021\_2022
    2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/HRSA/Dentists** folder. This will allow for the data cleaning process below to be automated
  + **Clean Data** 
    1. You will use the “clean\_dentists.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Access to Dentists data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Access to Dentists data into the **ncw\_data/cleaned\_data** folder as **dentists\_data.csv**. This data file will be the one that is pulled into PowerBI. These instructions are the same as above because the Rmarkdown file cleans both the state and county level estimates

## Access to Physicians

* County level estimates
  + **Download Data** 
    1. Go to <https://data.hrsa.gov/topics/health-workforce/ahrf> and scroll down to “Explore Workforce Data”
    2. Ensure you have “Clinicians” selected at the top
    3. For Data Source, select the time period of interest
       1. Even if you can download multiple years of data, download one year at a time for ease of data cleaning
    4. For Health Profession, select “Physicians (County Level File)”
    5. For Health Profession Subcategory, select “Active Physicians”
    6. For Population Category, select “All (County Level File)”
    7. For Population Subcategory*,* select “Population, All”
    8. For State, select “Washington”
    9. Hit submit
    10. You’ll see an Export Data option at the bottom of the dashboard. Export data as an “XLSX” file
    11. Repeat this process for all of the years of data you need
  + **Store Data** 
    1. The file will download with the general name “AHRFDashboard” regardless of the statistic of interest. Rename in the following format:
       1. DataSource\_Indicator\_Level\_Years
       2. Example: HRSA\_Physicians\_County\_2021\_2022
    2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/HRSA/Physicians** folder. This will allow for the data cleaning process below to be automated
  + **Clean Data** 
    1. You will use the “clean\_physicians.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Access to Physicians data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Access to Physicians data into the **ncw\_data/cleaned\_data** folder as **physicians\_data.csv**. This data file will be the one that is pulled into PowerBI. These instructions are the same as below because the Rmarkdown file cleans both the state and county level estimates
* State level estimates
  + **Download Data** 
    1. Go to <https://data.hrsa.gov/topics/health-workforce/ahrf>
    2. Repeat “Access to Physicians, County level estimates, Download Data” Steps 1-7 from above
    3. ***However***, for Step 8, select “All States” for State
    4. Hit submit
    5. Export data as an “XLSX” file
    6. Repeat this process for all of the years of data you need
  + **Store Data** 
    1. Again, the file will download with the general name “AHRFDashboard” regardless of the statistic of interest. Rename in the following format:
       1. DataSource\_Indicator\_Level\_Years
       2. Example: HRSA\_Physicians\_State\_2021\_2022
    2. In your NCW GitHub repository, put this data in the **ncw\_data/raw\_data/HRSA/Physicians** folder. This will allow for the data cleaning process below to be automated
  + **Clean Data** 
    1. You will use the “clean\_physicians.Rmd” file in the NCW GitHub repository in the “cleaning\_scripts” folder to clean the raw Access to Physicians data. Run the Rmarkdown file in its entirety, which will output the cleaned and formatted Access to Physicians data into the **ncw\_data/cleaned\_data** folder as **physicians\_data.csv**. This data file will be the one that is pulled into PowerBI. These instructions are the same as above because the Rmarkdown file cleans both the state and county level estimates