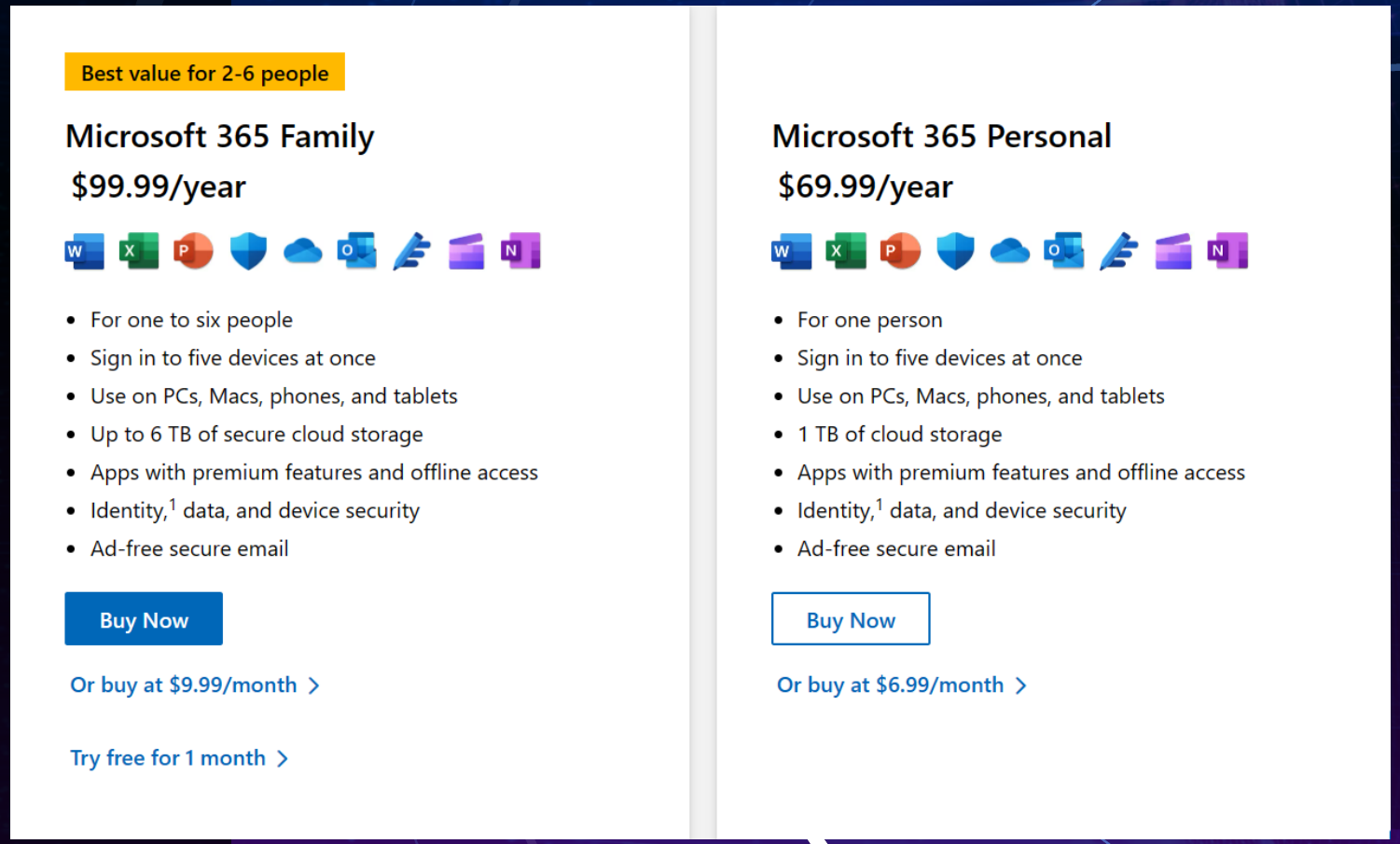
Youth Risk Behavior Surveillance System - A Dataset Guide

Including addictive, violent, and other health-related behaviors

Please Note:

- In order to utilize Microsoft Access Database (later described in this guide, page 9), users must have a Microsoft 365 subscription.
- This could incur subscription costs from $6.99 to $9.99 per month depending plan selected.
- However, another option is to email us at datajam@thedatajam.org and let us know what dataset you would like and we can download it for you and send it to you as an Excel file or Google Sheets file.



Overview

This dataset guide provides you with an overview on accessing and utilizing the Centers for Disease Control and Prevention's (CDC) Youth Risk Behavior Surveillance System (YRBSS) dataset(s). We will cover the following topics:

- How to access and download the YRBSS dataset
- Navigating the YRBSS dataset
- Utilizing the YRBSS' Analysis Tool

What exactly is the YRBSS?

- The YRBSS is a set of surveys that tracks US adolescent youth's behaviors, some of which are understood to lead to poor health outcomes (CDC)

Click here!

- There is data available spanning over 30 years!

- There are various categories of behaviors, including, but not limited to:
  - Unintentional injuries and violence
  - Alcohol, tobacco, and other drug use
  - Dietary behaviors
  - Sexual behaviors
  - Dietary behaviors
  - Physical activity
  - Other health topics

Where is the YRBSS located?

- The YRBSS is located on the CDC's webpage as an open-access database, meaning that anyone and everyone can download the data free of charge

- The following link takes you to the home page for the YRBSS
  - Youth Risk Behavior Surveillance System (YRBSS) | CDC

Click here!

- The next slide provides you with a screenshot of what to expect when you access the website

Data & Statistics Home

Adolescent Behaviors and Experiences Survey (ABES)

COVID Experiences Surveys (CovEx)

National School COVID-19 Prevention Study (NSCPS)

Youth Risk Behavior Surveillance System (YRBSS)

Overview

Why YRBS?

Results

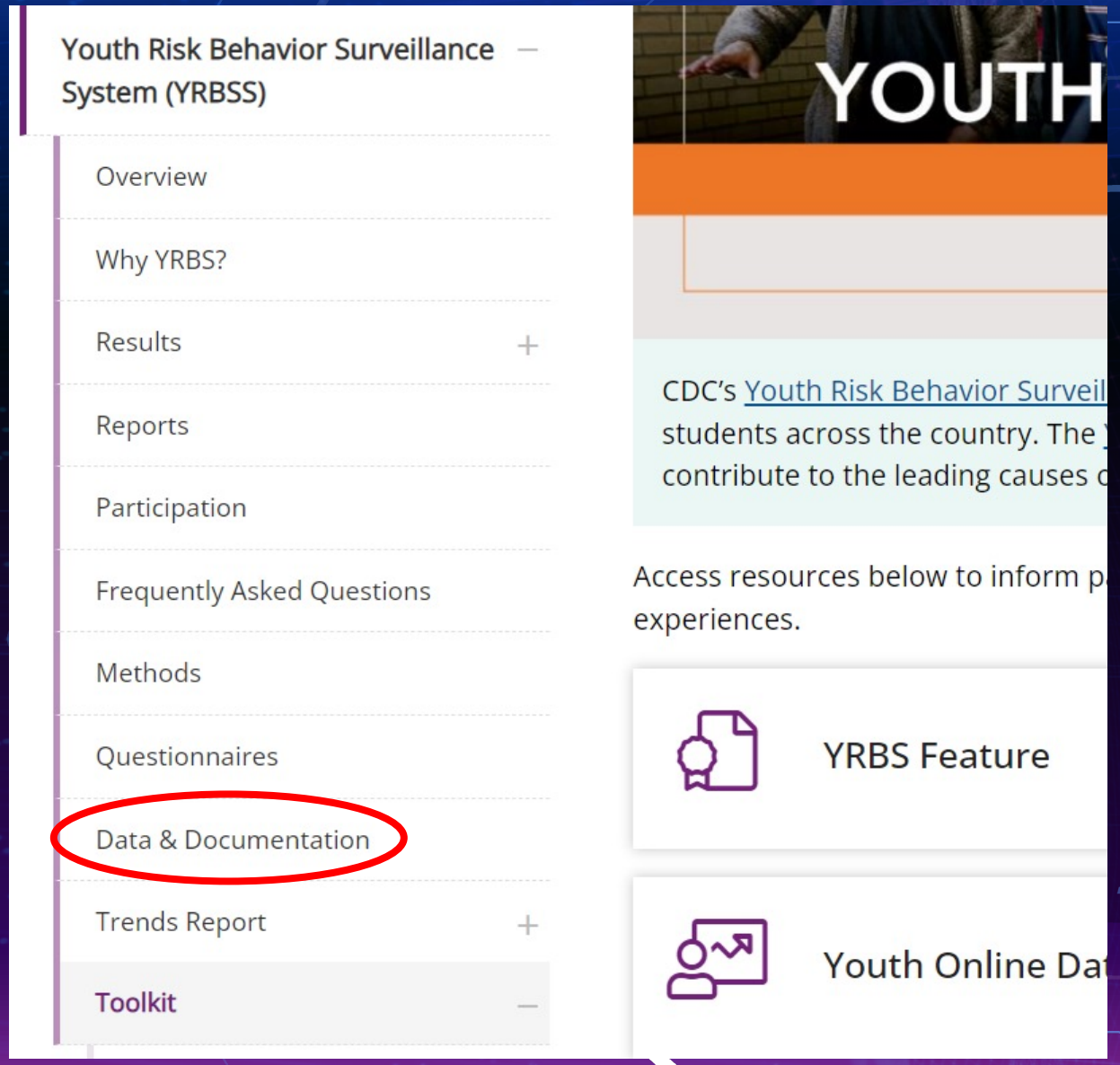
YRBS Results Toolkit

Print

CDC's Youth Risk Behavior Surveillance System (YRBSS) monitors health behaviors and experiences among high school

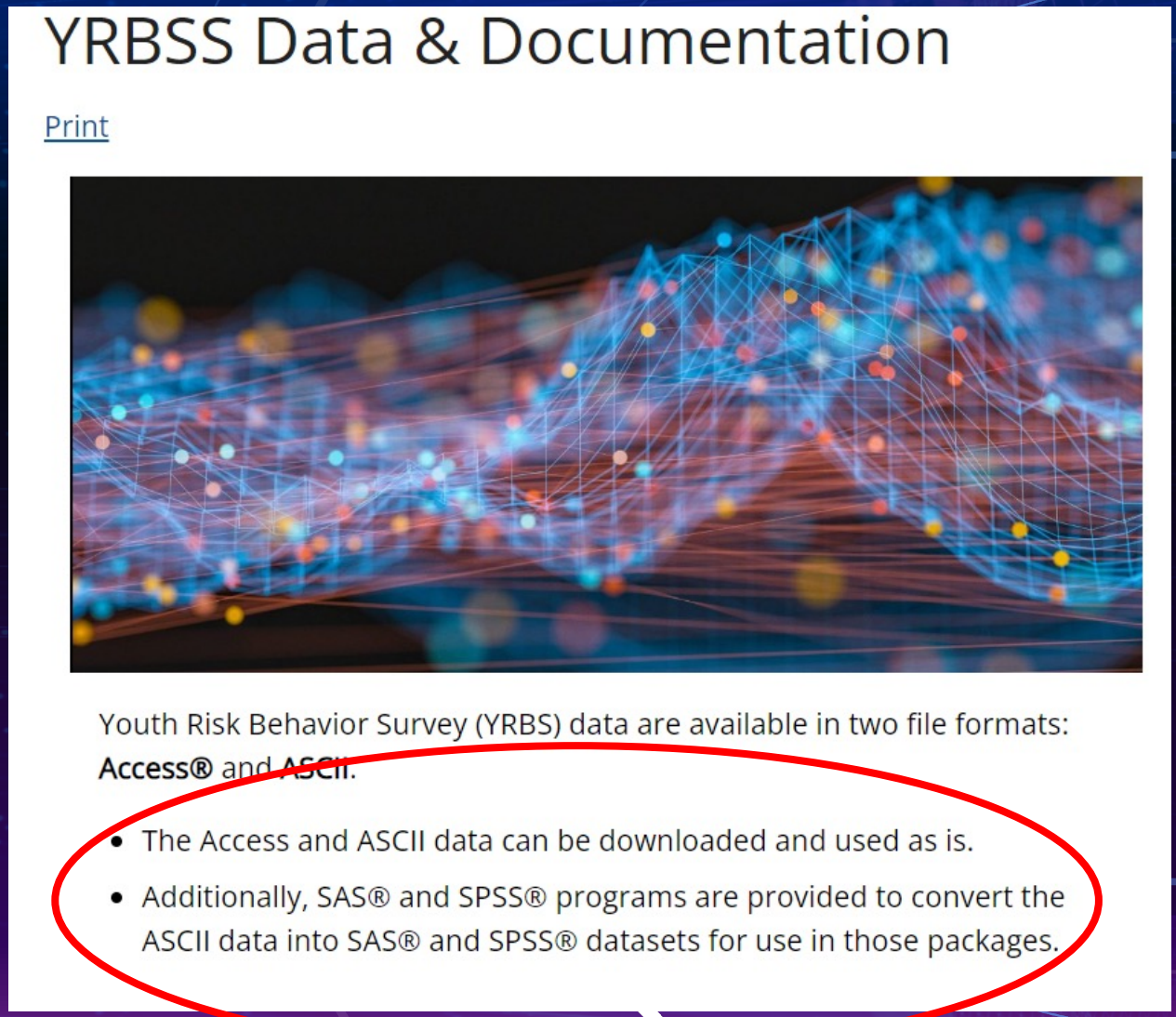
Accessing the Dataset

- On the left-hand side, scroll down to where you see "Data & Documentation" as seen in the picture here



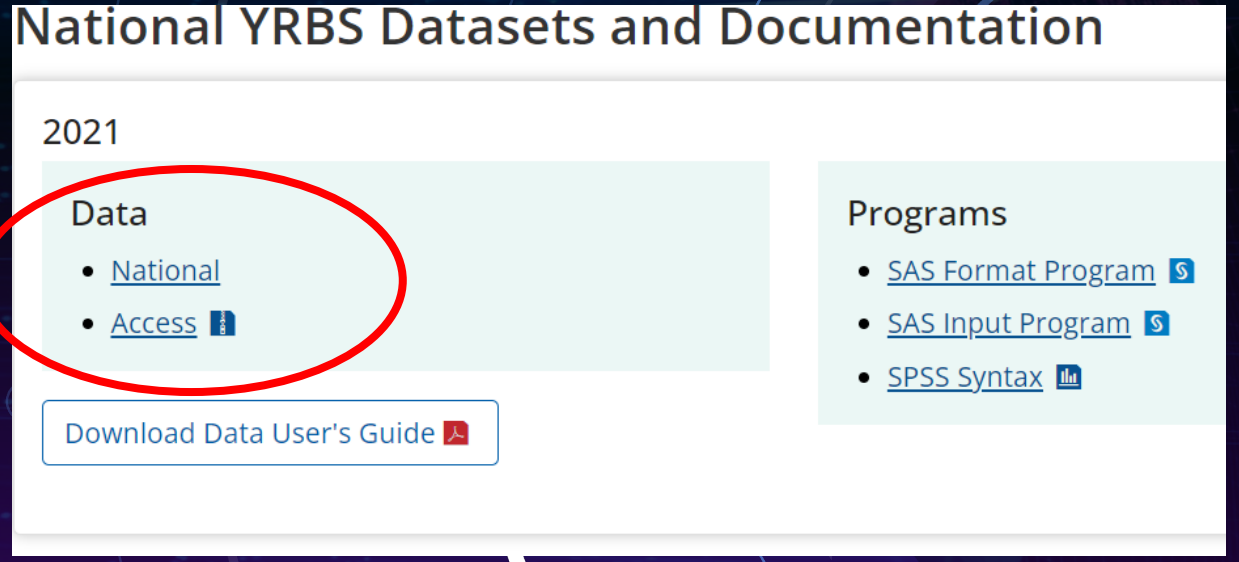
Accessing the Dataset (continued)

- Once you have opened the Data & Documentation tab, you are now able to review over 30 year's worth of datasets publicly provided by the CDC
- The default download provides you with a .DAT file



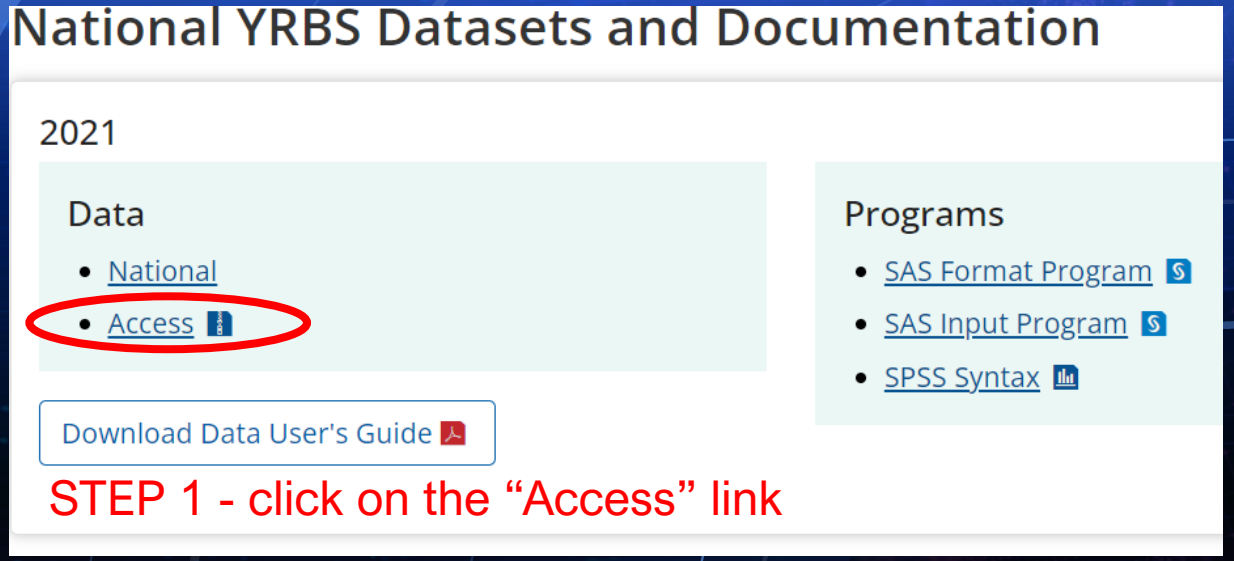
Accessing the Dataset (continued)

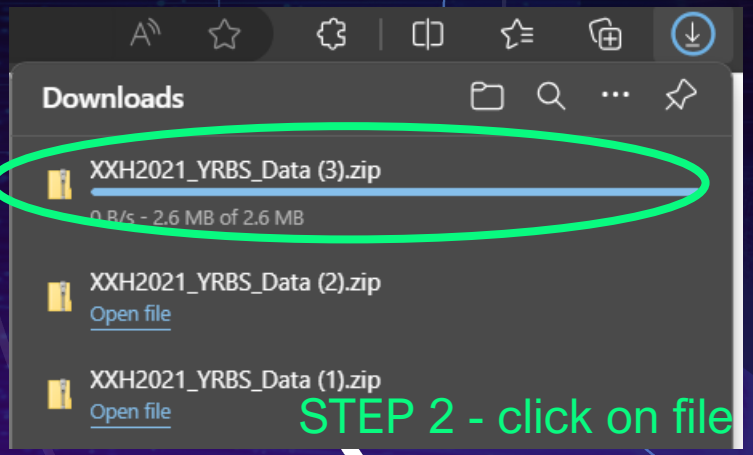
- In this guide, we will be downloading and using the data in a .DAT file
- Let's look at the 2021 Access YRBS Dataset as an example for this guide!
- **NOTE:** You can view this data on a state or district basis, too



Downloading the Dataset

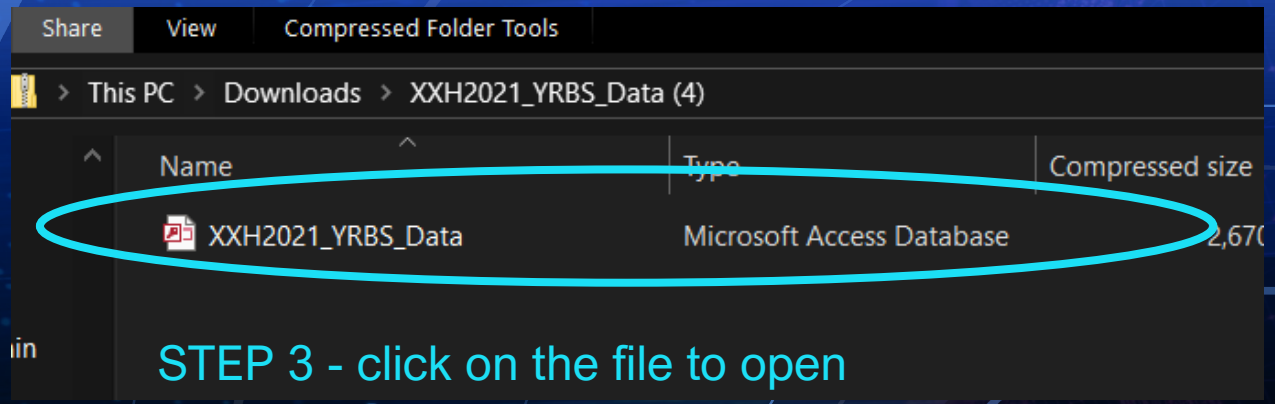
- Step 1 (red): To download the dataset, start by clicking on the 2021 “Access” link
- Step 2 (green): Click on the downloaded file in the pop-up window
- Proceed to next slide for following steps

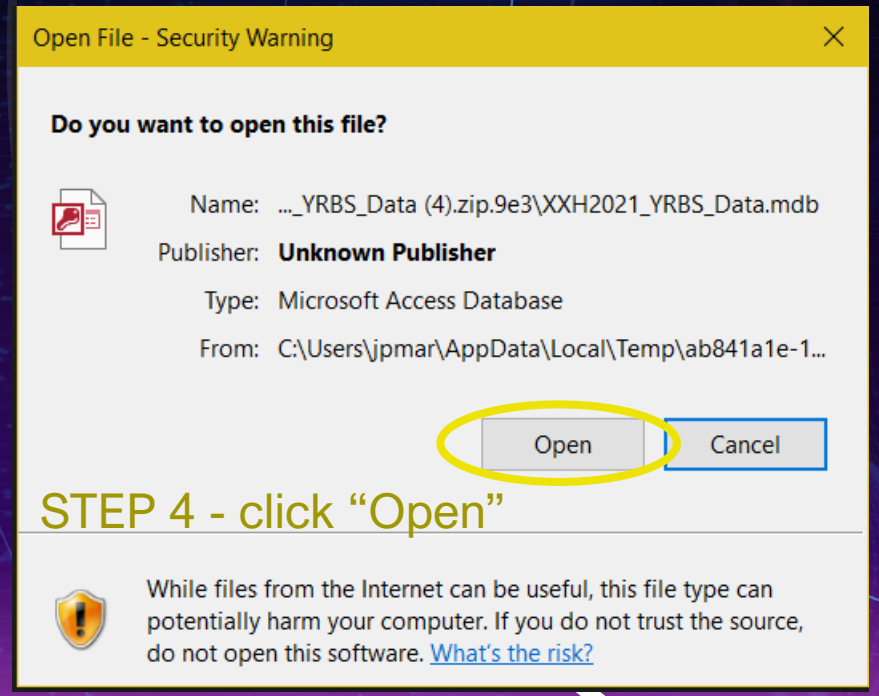




Downloading the Dataset

- Step 3 (blue): A compressed file folder will pop-up as a Microsoft Access Database -- click on this to now open your dataset
- Step 4 (yellow): Select “Open” to access the data
- **NOTE:** You do not need to worry about the Security Warning(s). This occurs because the dataset was created on another computer.

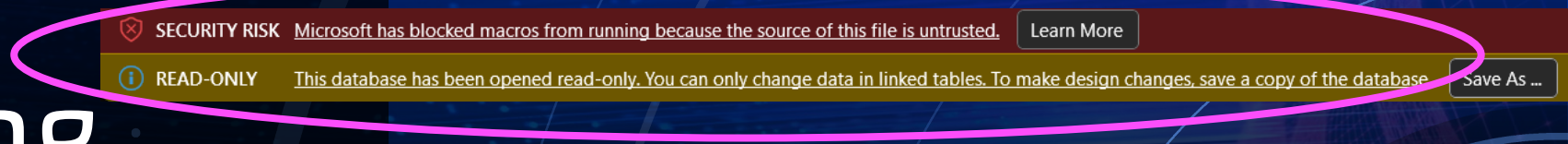


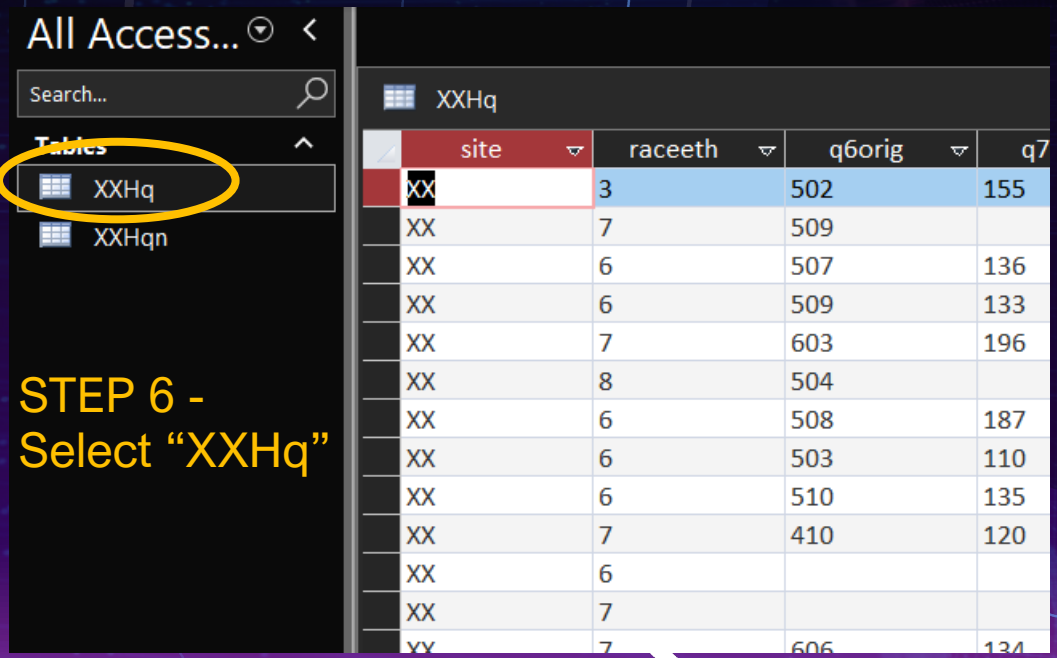


Downloading the Dataset (Continued)

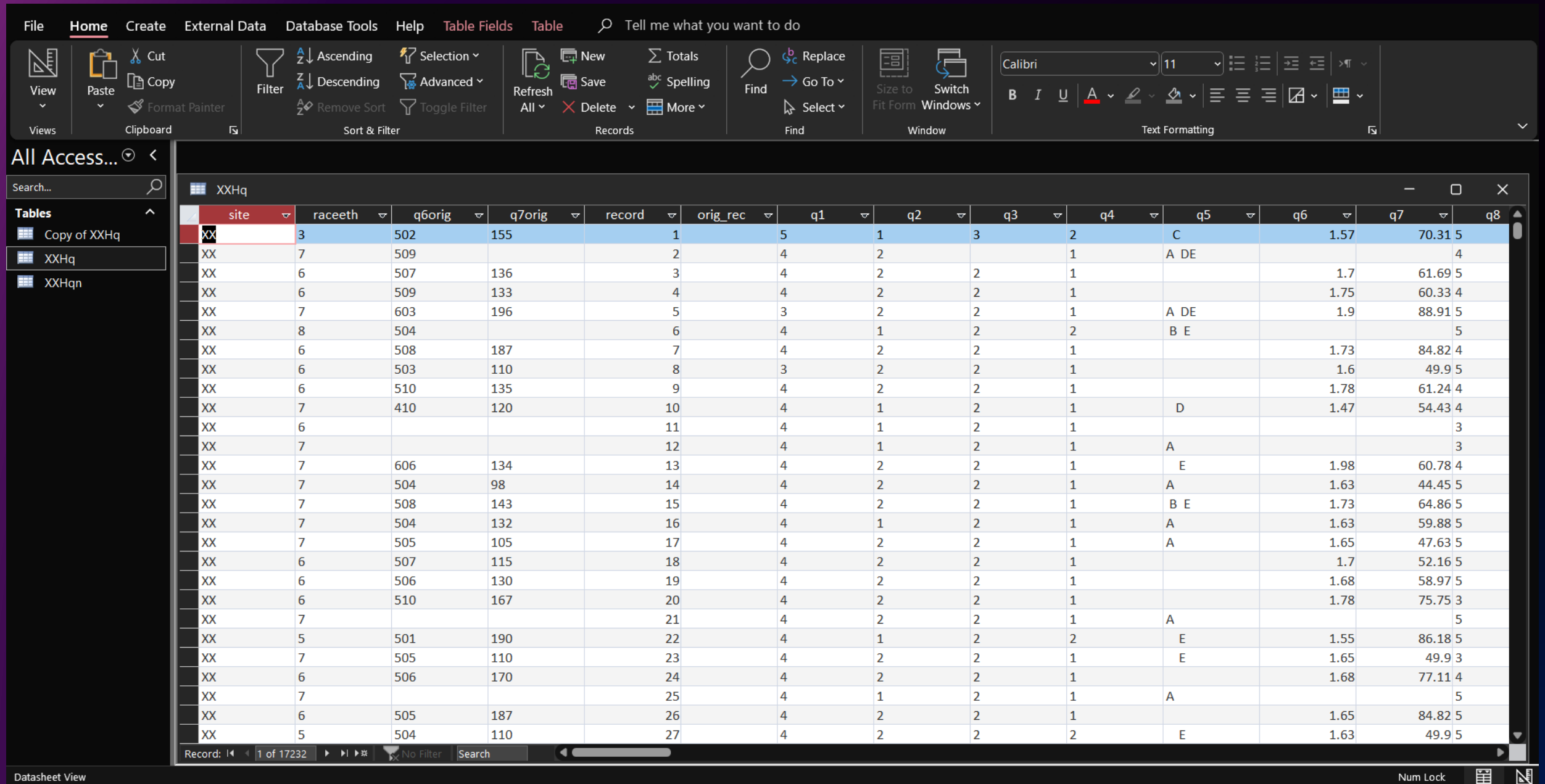
- Step 5 (purple): Microsoft Access Database will open; you will need to save the file and enable the database by pressing the X on the Security Risk pop-up
- Step 6 (orange): Select “XXHq” and you now have complete access to your database (as seen in next slide)

STEP 3 - Press X on security risk and save the file





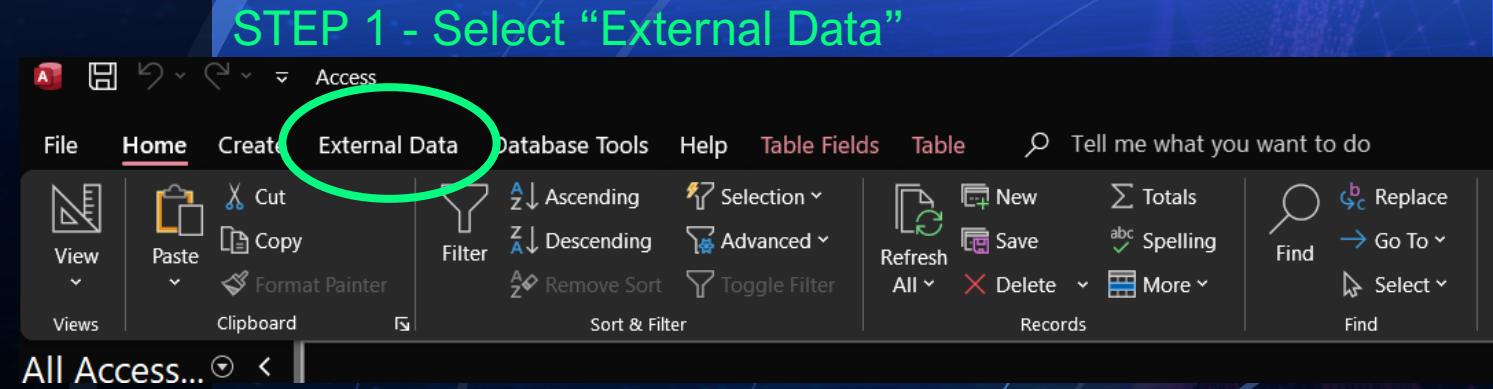
This is the dataset you get access to on Microsoft Access Database. Check the next slide to learn how to convert to an Excel file for easier review of the data!

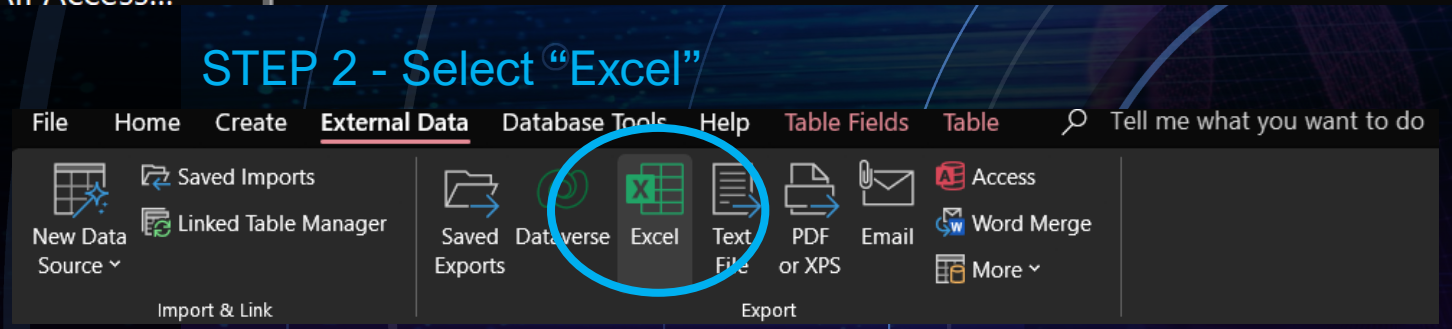


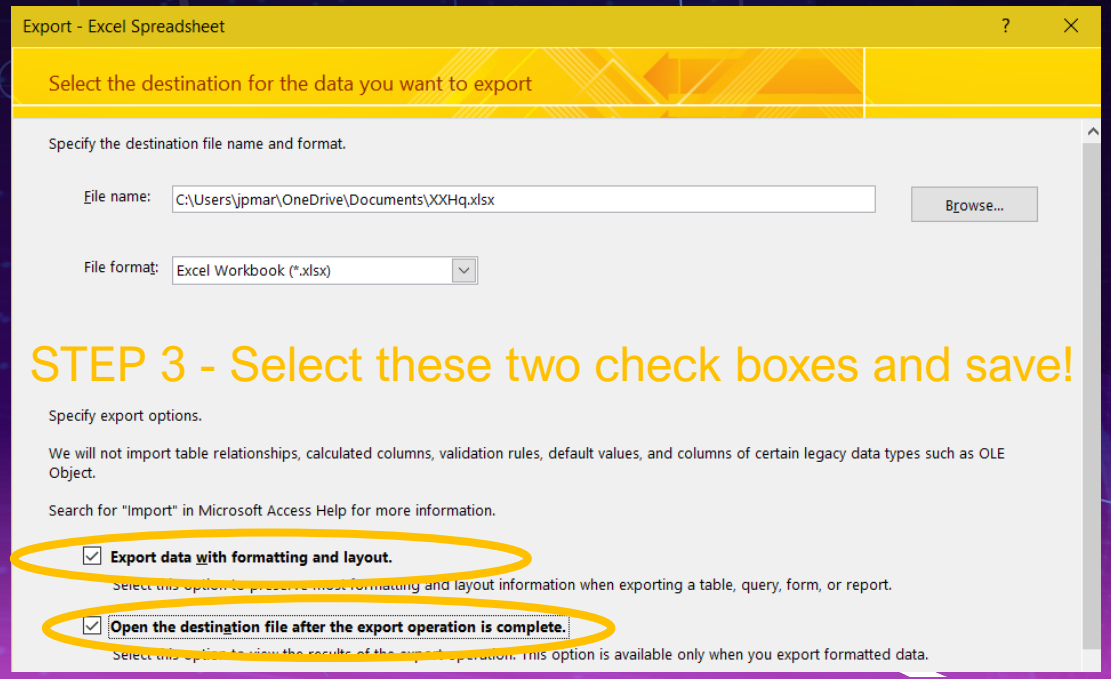
site	raceeth	q6orig	q7orig	record	orig_rec	q1	q2	q3	q4	q5	q6	q7	q8
XX	3	502	155	1		5	1	3	2	C	1.57	70.31	5
XX	7	509		2		4	2		1	A DE			4
XX	6	507	136	3		4	2	2	1		1.7	61.69	5
XX	6	509	133	4		4	2	2	1		1.75	60.33	4
XX	7	603	196	5		3	2	2	1	A DE	1.9	88.91	5
XX	8	504		6		4	1	2	2	B E			5
XX	6	508	187	7		4	2	2	1		1.73	84.82	4
XX	6	503	110	8		3	2	2	1		1.6	49.9	5
XX	6	510	135	9		4	2	2	1		1.78	61.24	4
XX	7	410	120	10		4	1	2	1	D	1.47	54.43	4
XX	6			11		4	1	2	1				3
XX	7			12		4	1	2	1	A			3
XX	7	606	134	13		4	2	2	1	E	1.98	60.78	4
XX	7	504	98	14		4	2	2	1	A	1.63	44.45	5
XX	7	508	143	15		4	2	2	1	B E	1.73	64.86	5
XX	7	504	132	16		4	1	2	1	A	1.63	59.88	5
XX	7	505	105	17		4	2	2	1	A	1.65	47.63	5
XX	6	507	115	18		4	2	2	1		1.7	52.16	5
XX	6	506	130	19		4	2	2	1		1.68	58.97	5
XX	6	510	167	20		4	2	2	1		1.78	75.75	3
XX	7			21		4	2	2	1	A			5
XX	5	501	190	22		4	1	2	2	E	1.55	86.18	5
XX	7	505	110	23		4	2	2	1	E	1.65	49.9	3
XX	6	506	170	24		4	2	2	1		1.68	77.11	4
XX	7			25		4	1	2	1	A			5
XX	6	505	187	26		4	2	2	1		1.65	84.82	5
XX	5	504	110	27		4	2	2	2	E	1.63	49.9	5

Making the dataset into an Excel File

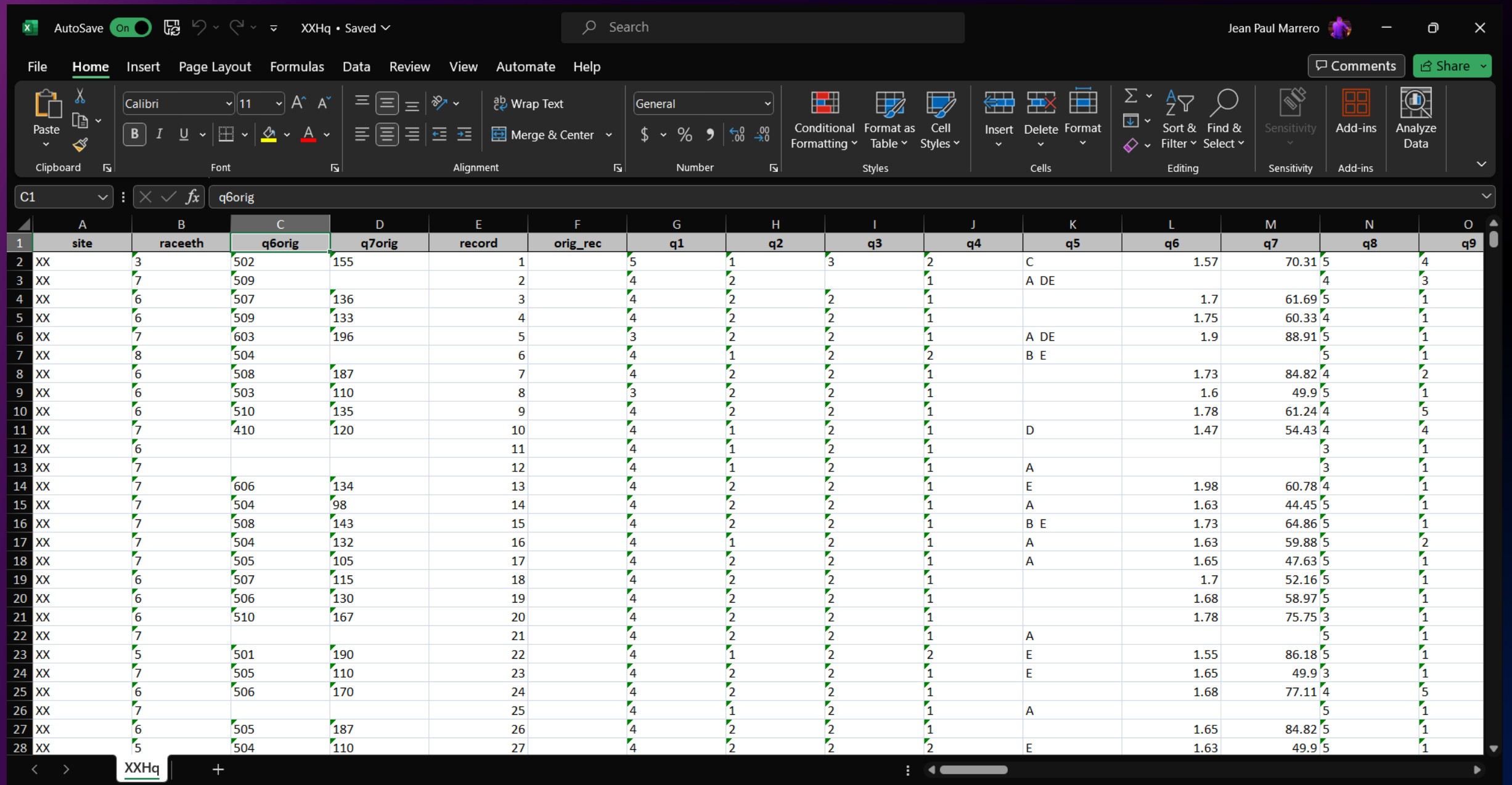
- Step 1 (green): Select “External Data” at the top of the screen
- Step 2 (blue): Click on the “Excel” tab
- Step 3 (orange): Give a file name, then select both “Export data with formatting and layout” and “Open the destination file after export operation is complete” and then hit “OK” to save the file
- This will create an Excel file that automatically will open (see next page)







This is the Excel file you have created. Both Microsoft Excel and Access Database are user-friendly, but the choice in which database you use is up to you!



AutoSave On | XXHq • Saved | Search | Jean Paul Marrero

File Home Insert Page Layout Formulas Data Review View Automate Help | Comments | Share

Paste | Calibri 11 | Wrap Text | Merge & Center | General | Conditional Formatting | Format as Table | Cell Styles | Insert | Delete | Format | Sort & Filter | Find & Select | Sensitivity | Add-ins | Analyze Data

Clipboard | Font | Alignment | Number | Styles | Cells | Editing | Sensitivity | Add-ins

C1 | q6orig

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	site	raceeth	q6orig	q7orig	record	orig_rec	q1	q2	q3	q4	q5	q6	q7	q8	q9
2	XX	3	502	155	1		5	1	3	2	C	1.57	70.31	5	4
3	XX	7	509		2		4	2		1	A DE			4	3
4	XX	6	507	136	3		4	2	2	1		1.7	61.69	5	1
5	XX	6	509	133	4		4	2	2	1		1.75	60.33	4	1
6	XX	7	603	196	5		3	2	2	1	A DE	1.9	88.91	5	1
7	XX	8	504		6		4	1	2	2	B E			5	1
8	XX	6	508	187	7		4	2	2	1		1.73	84.82	4	2
9	XX	6	503	110	8		3	2	2	1		1.6	49.9	5	1
10	XX	6	510	135	9		4	2	2	1		1.78	61.24	4	5
11	XX	7	410	120	10		4	1	2	1	D	1.47	54.43	4	4
12	XX	6			11		4	1	2	1				3	1
13	XX	7			12		4	1	2	1	A			3	1
14	XX	7	606	134	13		4	2	2	1	E	1.98	60.78	4	1
15	XX	7	504	98	14		4	2	2	1	A	1.63	44.45	5	1
16	XX	7	508	143	15		4	2	2	1	B E	1.73	64.86	5	1
17	XX	7	504	132	16		4	1	2	1	A	1.63	59.88	5	2
18	XX	7	505	105	17		4	2	2	1	A	1.65	47.63	5	1
19	XX	6	507	115	18		4	2	2	1		1.7	52.16	5	1
20	XX	6	506	130	19		4	2	2	1		1.68	58.97	5	1
21	XX	6	510	167	20		4	2	2	1		1.78	75.75	3	1
22	XX	7			21		4	2	2	1	A			5	1
23	XX	5	501	190	22		4	1	2	2	E	1.55	86.18	5	1
24	XX	7	505	110	23		4	2	2	1	E	1.65	49.9	3	1
25	XX	6	506	170	24		4	2	2	1		1.68	77.11	4	5
26	XX	7			25		4	1	2	1	A			5	1
27	XX	6	505	187	26		4	2	2	1		1.65	84.82	5	1
28	XX	5	504	110	27		4	2	2	2	E	1.63	49.9	5	1

XXHq

Navigating the Dataset

- Because the data is already defined and categorized, you will need to download the Data User's Guide to determine what variable(s) you are using
- You can also click on the link to take you directly to the guide!

Click here!

Navigating the Dataset (Continued)

- Once you have downloaded the Data User's Guide, you can review the questions that you think would be beneficial for you and your team's project
- See some of the examples here (Q77 and Q78) for some of the variables and their meanings!

Q77. During the past 7 days, on how many days were you physically active for a total of **at least 60 minutes per day**? (Add up all the time you spent in any kind of physical activity that increased your heart rate and made you breathe hard some of the time.)

A. 0 days
B. 1 day
C. 2 days
D. 3 days
E. 4 days
F. 5 days
G. 6 days
H. 7 days

Variable label: Physical activity >= 5 days

QN77:
Numerator: Students who answered F, G, or H for Q77
Denominator: Students who answered A, B, C, D, E, F, G, or H for Q77
Summary text: Percentage of students who were physically active at least 60 minutes per day on 5 or more days (in any kind of physical activity that increased their heart rate and made them breathe hard some of the time during the 7 days before the survey)
QN variable label: Were physically active at least 60 minutes per day on 5 or more days
Dependence: Required by QNPA0DAY and QNPA7DAY

Q78. On an average school day, how many hours do you spend in front of a TV, computer, smart phone, or other electronic device watching shows or videos, playing games, accessing the Internet, or using social media (also called "screen time")? (Do **not** count time spent doing schoolwork.)

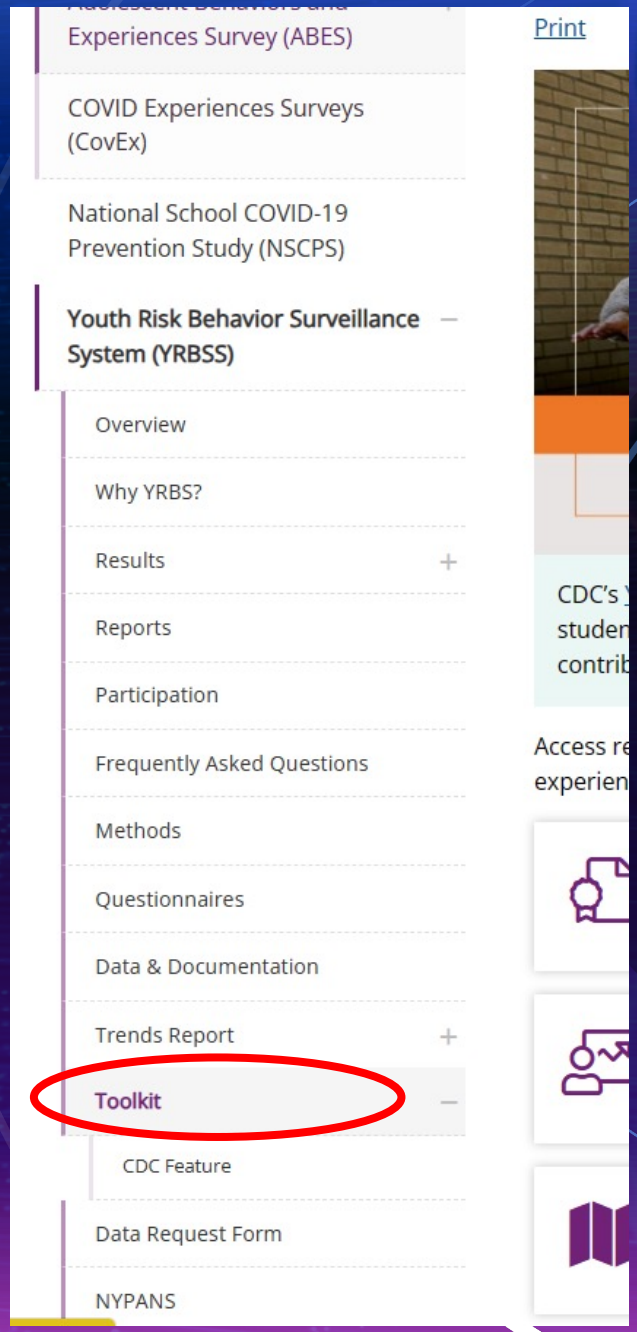
A. Less than 1 hour per day
B. 1 hour per day
C. 2 hours per day
D. 3 hours per day
E. 4 hours per day
F. 5 or more hours per day

Variable label: Screen time

QN78:
Numerator: Students who answered D, E, or F for Q78
Denominator: Students who answered A, B, C, D, E, or F for Q78
Summary text: Percentage of students who spent 3 or more hours per day on screen time (in front of a TV, computer, smart phone, or other electronic device watching shows or videos, playing games, accessing the Internet, or using social media, not counting time spent doing schoolwork, on an average school day)
QN variable label: Spent 3 or more hours per day on screen time

Utilizing the YRBSS' Analysis Tool

- In addition to accessing the full dataset, you can utilize an analysis tool that the CDC has created to analyze the YRBSS dataset
- This is accessed directly from the home page under the Toolkit tab as seen pictured here!



Utilizing the YRBSS' Analysis Tool (Continued)

- After selecting the Toolkit tab, you are taken to this page -- as seen in the screenshot
- Click on the Youth Online Data Analysis Tool!

Click here!

CDC's Youth Risk Behavior Surveillance System (YRBSS) monitors health behaviors and experiences among high school students across the country. The Youth Risk Behavior Survey (YRBS) results help in understanding the factors that contribute to the leading causes of illness, death and disability among youth.

Access resources below to inform partners and community members about adolescent health risk behaviors and experiences.

- YRBS Feature
- Data & Documentation
- Youth Online Data Analysis Tool
- YRBS Explorer
- YRBS Reports and Publications
- YRBSS Results

Utilizing the YRBSS' Analysis Tool (Continued)

- Once you access the Youth Online Data Analysis Tool tab, you can do various forms of statistical analyses which provide you with quick data on any variable(s) that you are interested in
- Go to the next slide to see what the analysis tool main page looks like!

Analysis Tool Main Page

https://nccd.cdc.gov/Youthonline/App/Default.aspx

Youth Online

Contact Us

Help

Youth Online lets you analyze national, state, and local Youth Risk Behavior Surveillance System (YRBSS) data from high school and middle school surveys conducted during 1991 – 2021 and Global School-based Student Health Survey (GSHS) data from students aged 13-17 conducted during 2003 – 2015. YRBSS data can be filtered and sorted on the basis of race/ethnicity, sex, academic grade, and sexual orientation while GSHS data can be filtered and sorted on sex and age. Additionally you can filter and sort both surveys by location and year, create customized tables, maps, graphs, and perform statistical tests by location and health topic. Please see Youth Online Help for more information about using all Youth Online capabilities.

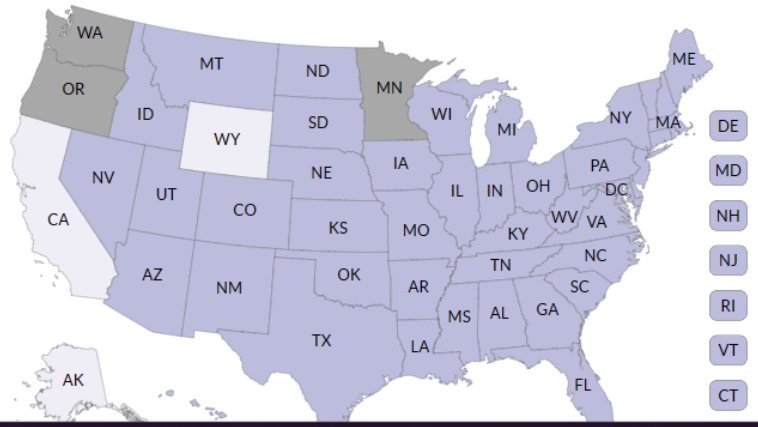
View data from: ◉ High School YRBS ○ Middle School YRBS ○ GSHS

View all responses for one location

View survey results for the United States, a state, a territory or freely associated state, or selected location. To view U.S. data, click the View U.S. Data link. To choose a specific location, make a selection in the dropdown menu or click a location on the map.

Select State, Local Site, Territory or Freely Associated State.

Select One ▾ GO View U.S. Data



View one question for all locations

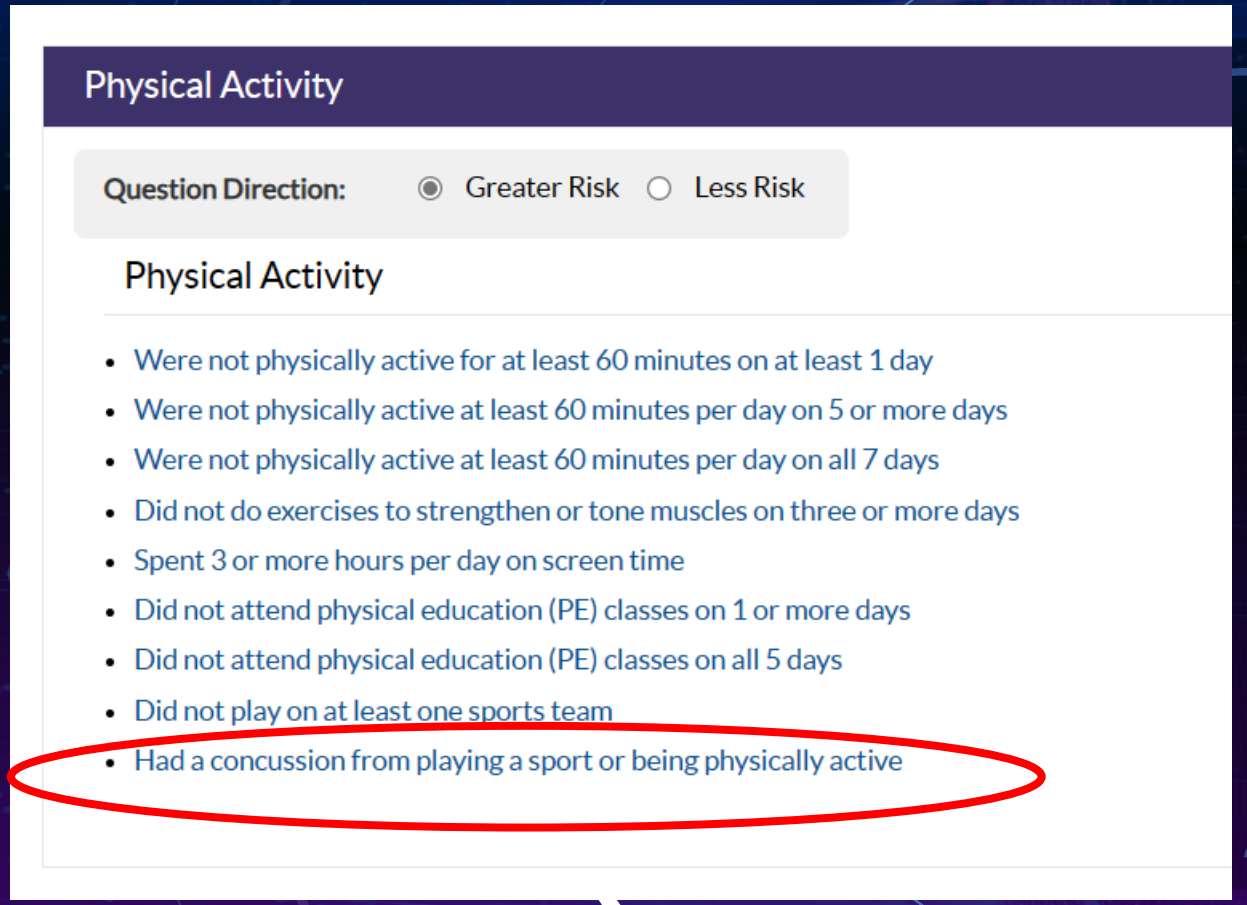
View results for one question for all states and locations that have data available. To choose a question, first select a Health Topic.

Select a Health Topic:

- Unintentional Injuries and Violence
- Tobacco Use
- Alcohol and Other Drug Use
- Sexual Behaviors
- Dietary Behaviors
- Physical Activity
- Obesity, Overweight, and Weight Control
- Other Health Topics

Let's Look at Some Examples!

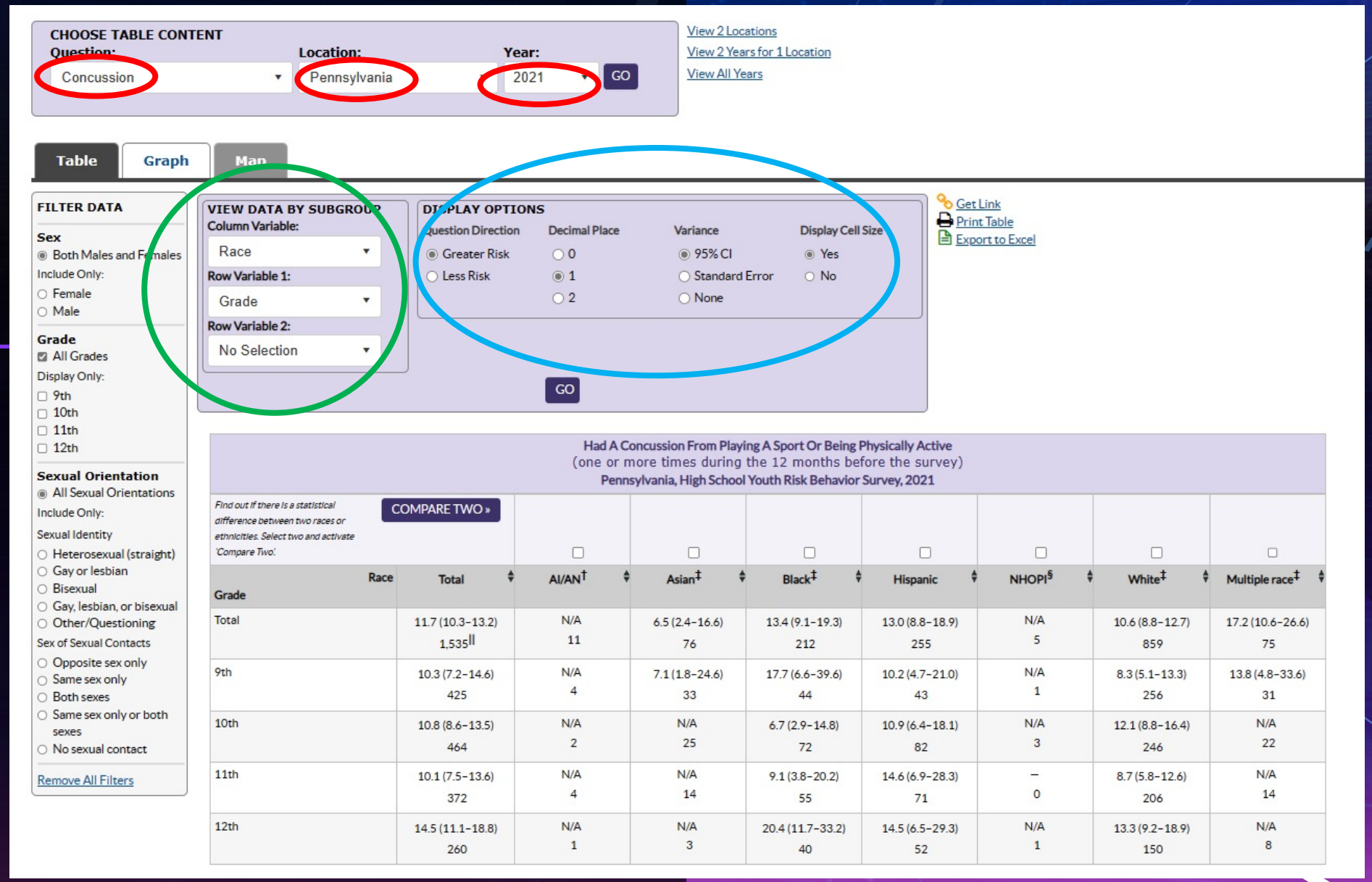
- Research Question 1 (RQ1): Are there differences in concussion rates when considering grade-level and race for US high school students in Pennsylvania?
- Step 1: Go to the Analysis Tool Main Page, select "Physical Activity," and then select the last bullet (circled in the picture here)



RQ1 Next Steps

- Step 2: Select your variables! Determine your location, year of data, race/ethnicity, academic grade year, sexual orientation and more. These are vital to addressing your overall question(s) as you review the data and should align with what has been asked. Our question involves concussion rates when considering 1) academic year and 2) ethnicity/race
- Therefore, let's select the following:
    - Table Content (red circles): Question = Concussion; Location = Pennsylvania; Year = 2021
    - Data by Subgroup (green circle): Column Variable = Race; Row 1 Variable = Grade
    - Display Options (blue circle): Direction = Greater Risk; Decimal Place = 1; Variance = 95% CI; Display Cell Size = Yes
- Once you have selected all of these, press "Go" at the top to run your analysis!
- Check the next slide to see how this will look!

RQ1 Variable Selection



CHOOSE TABLE CONTENT

Question: Concussion | Location: Pennsylvania | Year: 2021 | GO

View 2 Locations
View 2 Years for 1 Location
View All Years

Table | Graph | Map

FILTER DATA

Sex
- Both Males and Females
Include Only:
- Female
- Male

Grade
- All Grades
Display Only:
- 9th
- 10th
- 11th
- 12th

Sexual Orientation
- All Sexual Orientations
Include Only:
Sexual Identity
- Heterosexual (straight)
- Gay or lesbian
- Bisexual
- Gay, lesbian, or bisexual
- Other/Questioning
Sex of Sexual Contacts
- Opposite sex only
- Same sex only
- Both sexes
- Same sex only or both sexes
- No sexual contact

Remove All Filters

VIEW DATA BY SUBGROUP
Column Variable: Race
Row Variable 1: Grade
Row Variable 2: No Selection

DISPLAY OPTIONS

Question Direction	Decimal Place	Variance	Display Cell Size
Greater Risk	0	95% CI	Yes
Less Risk	1	Standard Error	No
	2	None	

GO

Get Link
Print Table
Export to Excel

Had A Concussion From Playing A Sport Or Being Physically Active
(one or more times during the 12 months before the survey)
Pennsylvania, High School Youth Risk Behavior Survey, 2021

Find out if there is a statistical difference between two races or ethnicities. Select two and activate 'Compare Two'. COMPARE TWO »

Grade \ Race	Total	AI/AN†	Asian‡	Black‡	Hispanic	NHOPI§	White‡	Multiple race‡
Total	11.7 (10.3–13.2) 1,535‖	N/A 11	6.5 (2.4–16.6) 76	13.4 (9.1–19.3) 212	13.0 (8.8–18.9) 255	N/A 5	10.6 (8.8–12.7) 859	17.2 (10.6–26.6) 75
9th	10.3 (7.2–14.6) 425	N/A 4	7.1 (1.8–24.6) 33	17.7 (6.6–39.6) 44	10.2 (4.7–21.0) 43	N/A 1	8.3 (5.1–13.3) 256	13.8 (4.8–33.6) 31
10th	10.8 (8.6–13.5) 464	N/A 2	N/A 25	6.7 (2.9–14.8) 72	10.9 (6.4–18.1) 82	N/A 3	12.1 (8.8–16.4) 246	N/A 22
11th	10.1 (7.5–13.6) 372	N/A 4	N/A 14	9.1 (3.8–20.2) 55	14.6 (6.9–28.3) 71	– 0	8.7 (5.8–12.6) 206	N/A 14
12th	14.5 (11.1–18.8) 260	N/A 1	N/A 3	20.4 (11.7–33.2) 40	14.5 (6.5–29.3) 52	N/A 1	13.3 (9.2–18.9) 150	N/A 8

What You'll Get!

- Once you press "Go," you will get some data output (as seen in the bottom table of the previous slide)
- From here, you can do many things with the data. Let's do a comparison between Hispanic and White Pennsylvanian HS students to see if there are any significant differences in concussion rates at any specific academic year for these two ethnic/racial groups
    - This is done by selecting the check box at the top of the table above the ethnic/racial groups we are assessing (circled in blue in the next slide)
    - After doing this, press the "Compare Two" button (circled in red in the next slide) to run the analysis
        - NOTE: p-values less than 0.05 mean you have a statistically significant finding!
- Go to Slide 26 for the final outcome(s) of our data analysis

RQ1 Data Output

Had A Concussion From Playing A Sport Or Being Physically Active
(one or more times during the 12 months before the survey)
Pennsylvania, High School Youth Risk Behavior Survey, 2021

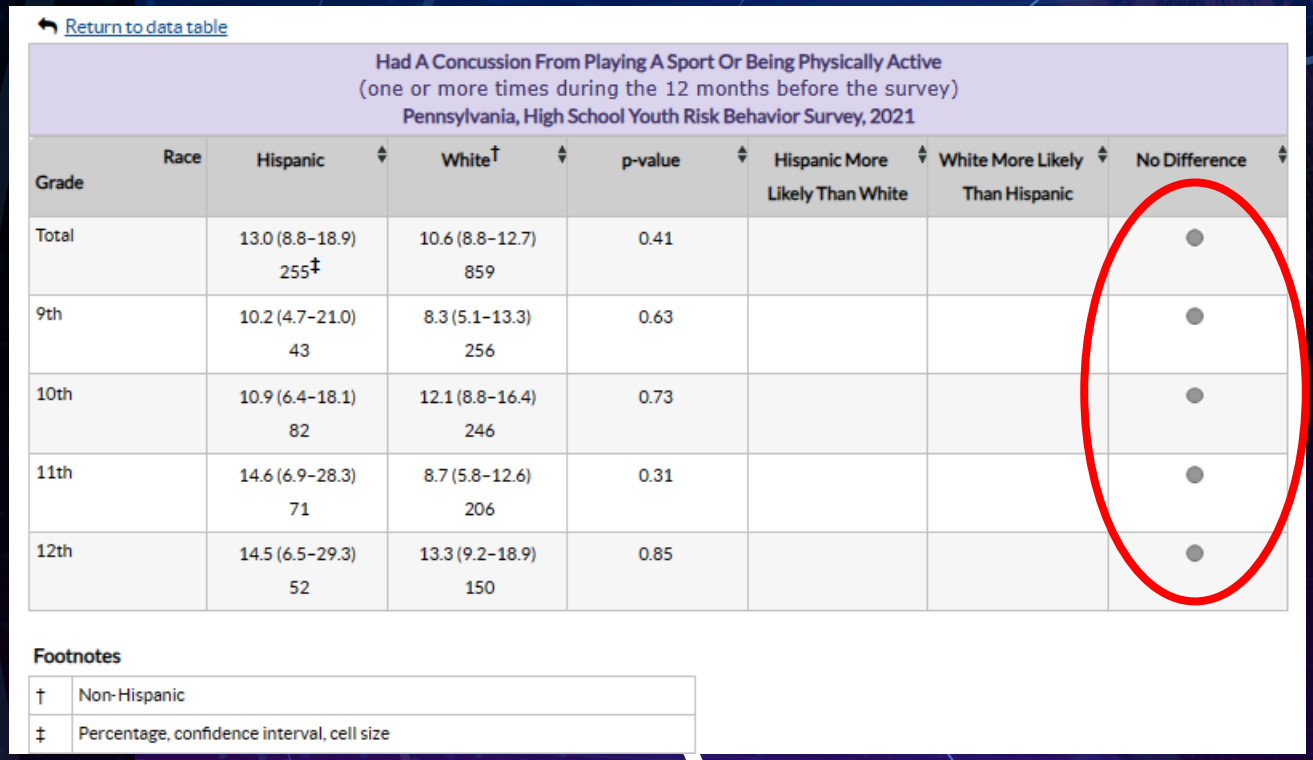
Find out if there is a statistical difference between two races or ethnicities. Select two and activate 'Compare Two'.

COMPARE TWO »

Grade \ Race	Total	AI/AN†	Asian‡	Black‡	Hispanic	NHOPI§	White‡	Multiple race‡
Total	11.7 (10.3–13.2) 1,535‖	N/A 11	6.5 (2.4–16.6) 76	13.4 (9.1–19.3) 212	13.0 (8.8–18.9) 255	N/A 5	10.6 (8.8–12.7) 859	17.2 (10.6–26.6) 75
9th	10.3 (7.2–14.6) 425	N/A 4	7.1 (1.8–24.6) 33	17.7 (6.6–39.6) 44	10.2 (4.7–21.0) 43	N/A 1	8.3 (5.1–13.3) 256	13.8 (4.8–33.6) 31
10th	10.8 (8.6–13.5) 464	N/A 2	N/A 25	6.7 (2.9–14.8) 72	10.9 (6.4–18.1) 82	N/A 3	12.1 (8.8–16.4) 246	N/A 22
11th	10.1 (7.5–13.6) 372	N/A 4	N/A 14	9.1 (3.8–20.2) 55	14.6 (6.9–28.3) 71	– 0	8.7 (5.8–12.6) 206	N/A 14
12th	14.5 (11.1–18.8) 260	N/A 1	N/A 3	20.4 (11.7–33.2) 40	14.5 (6.5–29.3) 52	N/A 1	13.3 (9.2–18.9) 150	N/A 8

RQ1 Analysis Results

- After running our analysis with the CDC tool provided, we see that there appears to be no difference (circled in red) in concussion rate across all academic years when comparing Hispanic and White high school students in Pennsylvania because our p-values are all greater than 0.05



Return to data table

Had A Concussion From Playing A Sport Or Being Physically Active
(one or more times during the 12 months before the survey)
Pennsylvania, High School Youth Risk Behavior Survey, 2021

Grade / Race	Hispanic	White†	p-value	Hispanic More Likely Than White	White More Likely Than Hispanic	No Difference
Total	13.0 (8.8–18.9) 255‡	10.6 (8.8–12.7) 859	0.41			
9th	10.2 (4.7–21.0) 43	8.3 (5.1–13.3) 256	0.63			
10th	10.9 (6.4–18.1) 82	12.1 (8.8–16.4) 246	0.73			
11th	14.6 (6.9–28.3) 71	8.7 (5.8–12.6) 206	0.31			
12th	14.5 (6.5–29.3) 52	13.3 (9.2–18.9) 150	0.85			

Footnotes

†	Non-Hispanic
‡	Percentage, confidence interval, cell size

Let's do another example!

- Following the steps previous addressed, let's select our variables to ensure that we address the question adequately
- Research Question 2 (RQ2): Are there differences in alcohol consumption when considering sex and race for high school students in the United States?
- Use the variables below in the Analysis Tool to find out!
    - Table Content: Question = Current alcohol use; Location = United States; Year = 2021
    - Data by Subgroup: Column Variable = Race; Row 1 Variable = Sex
    - Display Options: Direction = Greater Risk; Decimal Place = 1; Variance = 95% CI; Display Cell Size = Yes
- The next slide will provide you with the results that you should see!

RQ2 Data Output

Currently Drank Alcohol
(at least one drink of alcohol, on at least 1 day during the 30 days before the survey)
United States, High School Youth Risk Behavior Survey, 2021

Find out if there is a statistical difference between two races or ethnicities. Select two and activate 'Compare Two'.

COMPARE TWO »

Sex \ Race	Total	AI/AN†	Asian‡	Black‡	Hispanic	NHOPI§	White‡	Multiple race‡
Total	22.7 (21.1–24.4) 16,234‖	32.0 (23.1–42.5) 140	10.6 (9.1–12.4) 809	13.2 (11.3–15.4) 2,123	22.9 (19.8–26.2) 2,993	21.7 (13.1–33.6) 79	25.9 (24.2–27.7) 8,757	23.5 (20.1–27.4) 948
Female	26.8 (24.6–29.0) 7,693	41.2 (28.8–54.9) 63	12.8 (10.0–16.2) 400	18.0 (15.0–21.4) 1,032	27.4 (23.6–31.5) 1,453	42.7 (22.7–65.3) 37	29.9 (27.6–32.2) 4,095	27.1 (21.5–33.6) 495
Male	18.8 (17.2–20.5) 8,297	24.0 (14.3–37.3) 77	8.2 (6.2–10.8) 401	8.7 (6.4–11.7) 1,080	17.6 (15.3–20.1) 1,502	6.4 (1.4–24.7) 42	22.7 (20.4–25.2) 4,560	18.6 (15.4–22.3) 447

What if we compare alcohol consumption between Black and White students? Check the next slide for the data output!

Currently Drank Alcohol (at least one drink of alcohol, on at least 1 day during the 30 days before the survey) United States, High School Youth Risk Behavior Survey, 2021						
Sex / Race	Black†	White†	p-value	Black More Likely Than White	White More Likely Than Black	No Difference
Total	13.2 (11.3–15.4) 2,123‡	25.9 (24.2–27.7) 8,757	0.00		●	
Female	18.0 (15.0–21.4) 1,032	29.9 (27.6–32.2) 4,095	0.00		●	
Male	8.7 (6.4–11.7) 1,080	22.7 (20.4–25.2) 4,560	0.00		●	

Footnotes

†	Non-Hispanic
‡	Percentage, confidence interval, cell size

We have statistically significant data! A p value of 0.00 means that our p values are all less than 0.001!

RQ2 Results Analysis

- All of our comparisons that were conducted resulted in statistically significant results. This ultimately means that we found that:
    - White high school students are significantly more like to consume alcohol when compared to Black high school students
    - White female high school students are significantly more likely to consume alcohol when compared to Black female high school students
    - White male high school students are significantly more likely to consume alcohol when compared to Black male high school students
        - This agrees with data from previous years of the YRBSS as seen here!

Contact Information for Questions

JP Marrero-Rivera, MS

Graduate Student Researcher

University of Pittsburgh

Email: jpm190@pitt.edu

Work Phone: 412-383-4048