

# Health Trends & Behaviors in the United States: Insights from Vaccination Rates, Medical Discourse, and Risk-Factors. (2018-2022)

[Code ▾](#)

## Module Project:

PSY6422 Data Management & Visualization

## INTRODUCTION

Preventative healthcare plays a crucial role in safeguarding public health. Today, it focuses on various strategies, treatments and interventions on at-risk populations to prevent the onset and progression on a variety of diseases. With developments in medicine and research, the past few decades have witnessed a rise in methods to identify risk factors and secure populations to promote over-all well being. Successfully tackling these healthcare outcomes and risk factors signifies the sustainability of healthcare systems for future generations. This project aims to explore specific questions pertaining to the state of healthcare services today – in understanding various facets of such interventions and its prevalence amongst the population. Due to considerations in data-availability, the project focus narrows down on the trends in such healthcare utilizations within the United States of America, from the year 2018 to 2022. The following sections explore specific elements and patterns pertaining to preventative medicine that provides insights into geographical and temporal trends, socio-demographic disparities into utilization and examining patterns of discourse pertaining to such interventions.

## DATA SOURCE

The data for this study was obtained from the health surveys conducted by the Behavioral Risk-Factor Surveillance System (BRFSS). Established by the Centers for Disease Control & Prevention (CDC), the BRFSS collects data on the prevalence of chronic health conditions, health promotion & risk behaviors and utilization rates of preventative medicine services. This is done through regular telephone interviews with the residents of the 50 states & territories of the United States. The BRFSS collects a wide variety of information – **socio-demographics**, **chronic healthcare conditions** (such as heart disease, cancer, diabetes), **health behaviors** (such as alcohol consumption, smoking, diet, physical activity), **preventative health services** (cancer screenings, flu vaccinations, blood pressure checks) and **access to healthcare** (health insurance coverage, utilization of healthcare resources). More information about the data collection methods can be found on [CDC BRFSS Website](https://www.cdc.gov/brfss/index.html) (<https://www.cdc.gov/brfss/index.html>)

For this project, I analyzed the datasets from 2018 to 2022 (latest down to 5 years), to explore my questions within a relatively recent timeframe. The links for obtaining the dataset are provided in the reference section.

The [Repository](https://github.com/purni24/PSY6422Project/tree/main) (<https://github.com/purni24/PSY6422Project/tree/main>) consists of the raw folders containing the data for the specific questions I have asked. This is discussed more in the Data Preparation Section. The libraries used are presented in the code chunk.

[Show](#)

Alternatively, I have added a `renv.lock` file - a package management tool that stores all information about the used libraries.

## RESEARCH QUESTIONS

The overarching aim of this project is to utilize the BRFSS datasets to identify socio-demographic and geographic trends in early detection and screening services utilized by individuals within different parts of the USA.

1. **Adult Flu Shot Rate:** What is the percentage of people who have taken their adult flu shot in the last 12 months across various states in the USA? (2022)
2. **Urban-Rural Disparities in Contraceptive Utilization:** What are the differences in utilization of various contraceptives in urban and rural areas? (2022)
3. **Medical Discourse on Prostate Screening:** What is the nature of the debriefing given to inform people about a blood test to screen for Prostate Cancer? (2018-2022)
4. **HIV Risk-Behaviors by Age:** What are the age-related trends in individuals who demonstrated HIV-Risk behaviors? (2018-2022)

## DATA PREPARATION

The original BRFSS files on the website are present in two types of file forms - `.xpt` and `.ascii`. I downloaded the `.XPT` file, and converted it into a `.csv` file format to examine the file's contents. The data for this step as well as the original BRFSS files for 2018 to 2022 are not uploaded on the repository due to file size limitations, but here is a syntax to convert the file format from `.XPT` to `.CSV` that I used:

[Show](#)

After an examination of the variables on the BRFSS 2018-2022 sheets, I decided what variables I was going to explore through this project. I retained variables of interest through this code. (syntax provided)

[Show](#)

With this, I selected the variables of interest. The list of variables selected are available in the "codebook.txt" file within the repository. I obtained 5 datasets from each of the Raw files obtained from the BRFSS website - each dataset representing files from 2018 to 2022. The chunk contains the code as well as the first few rows of a dataset - BRFSS2022.

[Show](#)

##	X_STATE	FMONTH	FLUSH07	HIVRISK5	PSATEST1	PCSTALK1	X_URBSTAT	X_AGE5YR
## 1	1	1	1	2	NA	NA	1	13
## 2	1	1	2	2	NA	NA	1	13
## 3	1	1	2	2	NA	NA	1	8
## 4	1	1	1	2	NA	NA	1	14
## 5	1	1	2	2	NA	NA	1	5
## 6	1	1	2	2	NA	NA	1	13
##	TYPNTR9							
## 1	NA							
## 2	NA							
## 3	NA							
## 4	NA							
## 5	NA							
## 6	NA							

Within the table, is a column called X\_STATE, representing numeric values as a key associated with every state. The key is available in the codebook on the repository as well as the BRFSS website. For ease, I altered the values of the column to their corresponding states through the following means.

Show

VISUALIZATION 1:

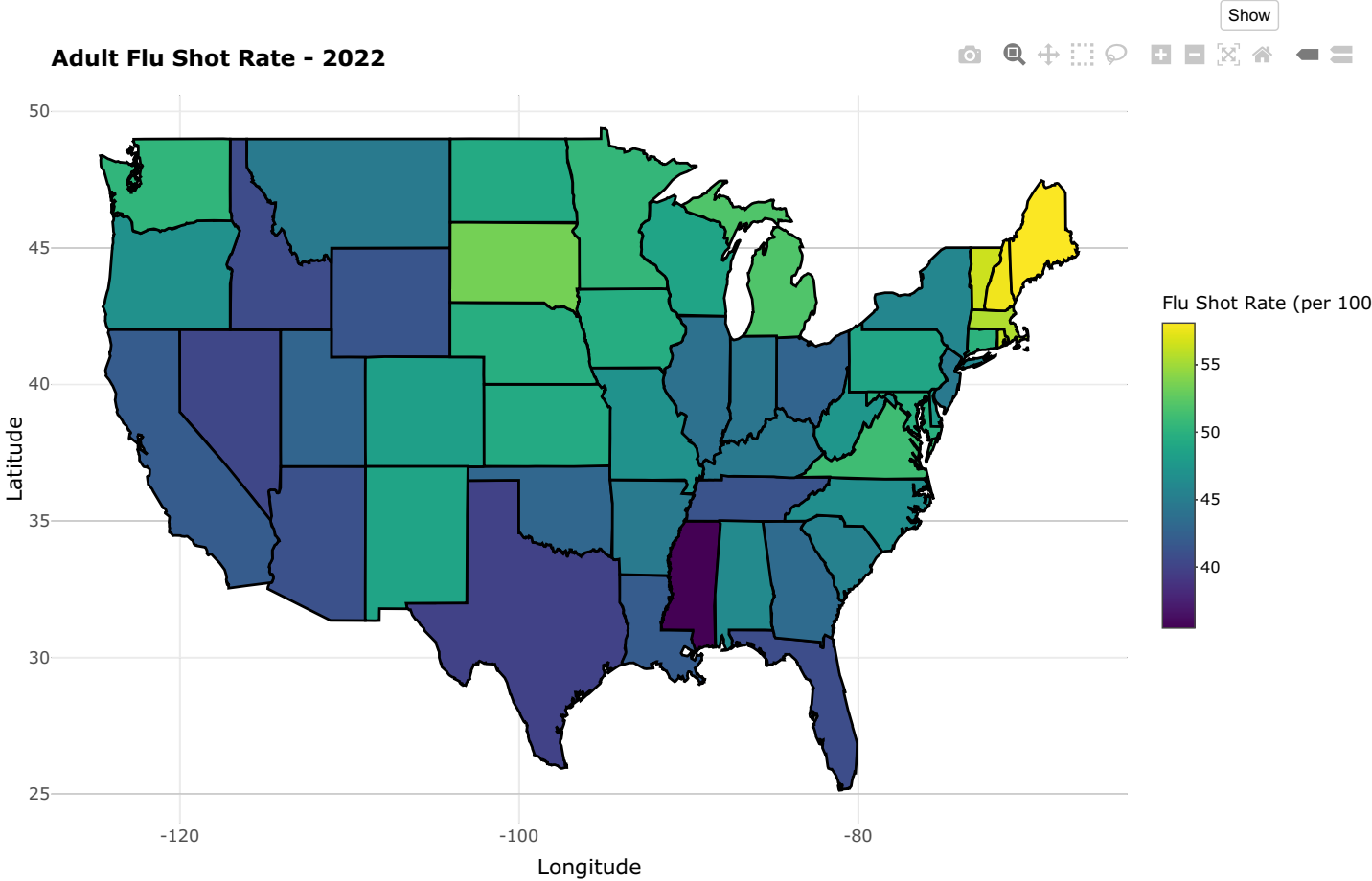
**Adult Flu Shot Rate:** What is the percentage of people who have taken their adult flu shot in the last 12 months across various states in the USA? (2022)

This visualization examines the rates of individuals across various states in the USA who reported of having received the flu-shot vaccine within the last 12 months.

Cleaning Data

Show

Visualizing Data



This choropleth map highlights the geographic trends in adult-flu shot rate across various states of the USA, in the year 2022. While it may be seen that regions in the far-east, such as Maine & New Hampshire report the highest rates of flu-shots taken and Mississippi the lowest, the overall range lies on an average between 35 to 58% of people amongst the entire population. In the preparation of the chart, I wanted to improve the user's ease in identifying the state and the flu-shot rate. To gain the most out of this map, please hover over the various states on the map of USA to gain more insights into the flushot rates per 100 per region.

VISUALIZATION 2:

**Urban-Rural Disparities in Contraceptive Utilization:** What are the differences in utilization of various contraceptives in urban and rural areas? (2022)

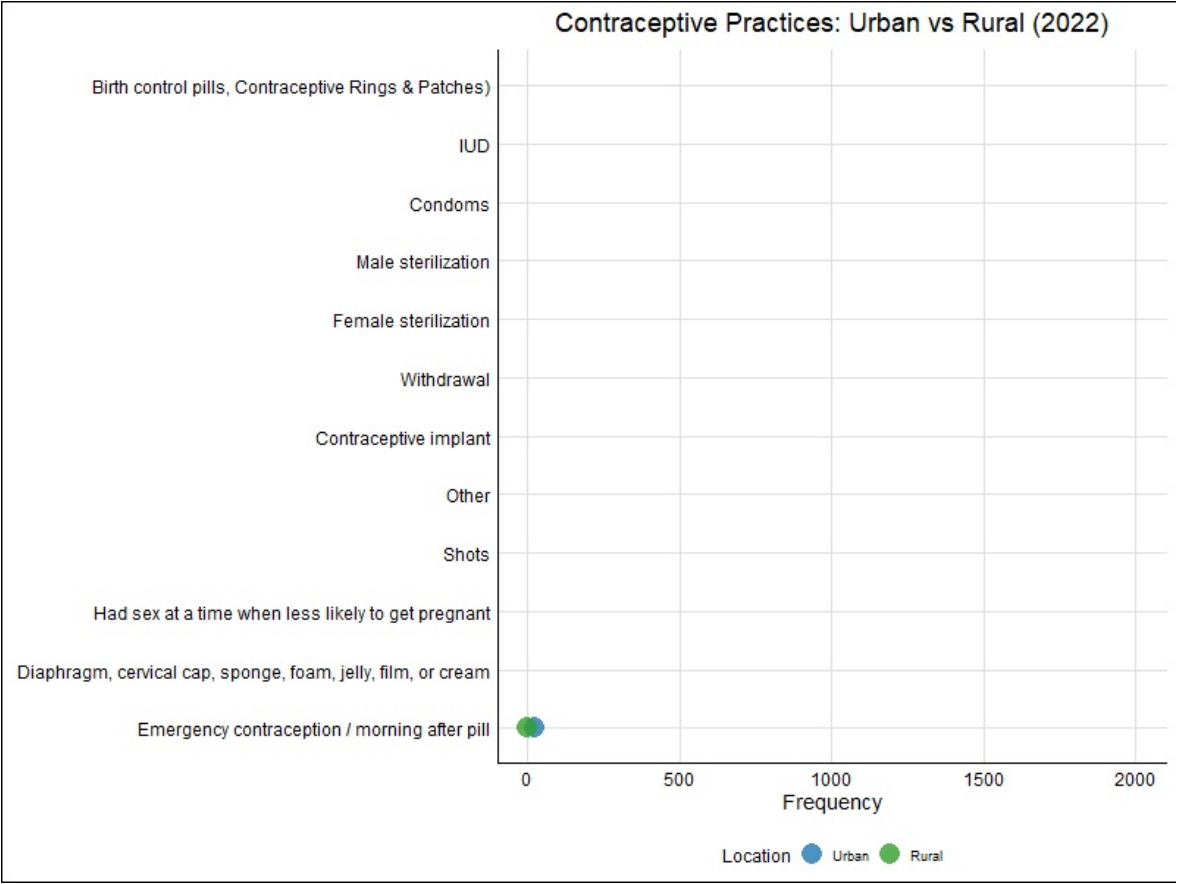
This visualization explores differences in contraceptive use amongst women residing in urban and rural areas across the USA.

Cleaning Data

Show

Visualizing Data

Show



As observed in this dataset, individuals within the rural areas tend of utilize contraceptive methods to a minimum, consistent across all types. Interestingly, the frequencies of the urban and rural population widen as the methods move from the bottom to the top of the Y-Axis. The widening discrepancy of the preferred utilization methods raises questions about geographic trends in the knowledge or availability of such methods, which warrants further analysis.

**VISUALIZATION 3:**

**Medical Discourse on Prostate Screening:** What is the nature of the debriefing given to inform people about a blood test to screen for Prostate Cancer? (2018-2022)

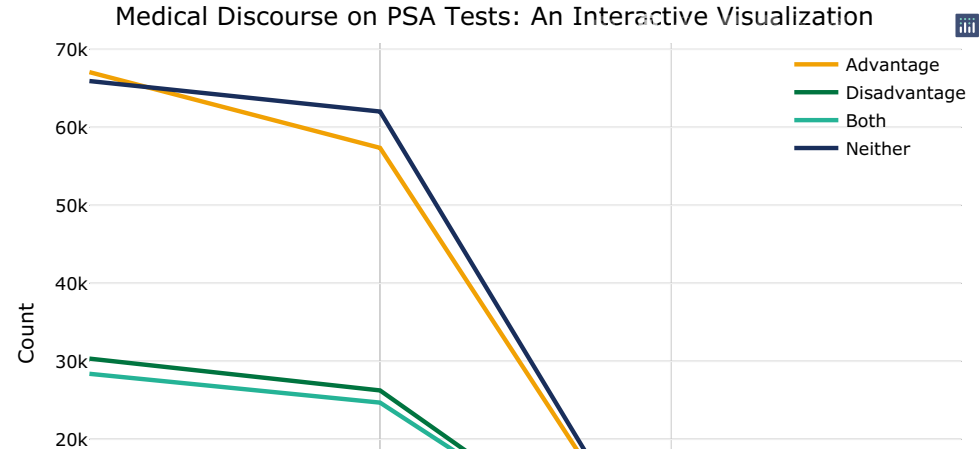
This dataset analyses the frequencies and ratios of individuals who were informed about the merits and demerits of undertaking a blood test to detect Prostate Cancer - measured across "Advantages", "Disadvantages", "Both" and "Neither".

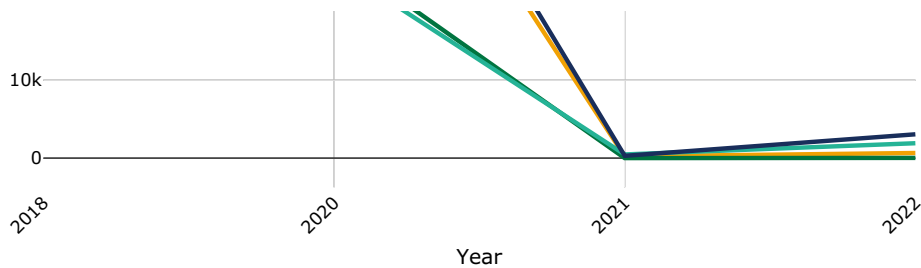
Cleaning Data

Show

Visualizing Data

Show





This interactive plot displays information on the number of people who were told about the advantages, disadvantages, both, and neither, on PSA tests. This chart allows individuals to click on the legend elements to select and deselect the items they wish to inspect, as well as hover over tips to display the exact number of individuals. The data was not available for the year 2019.

As you can see, there is a significant drop in all lines from 2020 to 2021. This could be misinterpreted as a drop in discussions about PSA. But, the actual drop is attributed to the reduced number of people who were asked this question in the 2021 and 2022 surveys, as is reported in the BRFSS codebook. I encountered a few challenges while making this visualization, which led me to make certain stylistic choices to improve clarity.

**Problem#1 :** The overlapping lines on the 2021 datapoint. **Solution:** As lines as overlapping for 2021 and 2022, I decided to include hover over tips where viewers could observe the number and nature of responses for each line. Each of the legends can also be selected, deselected, and zoomed into, to gain more clarity.

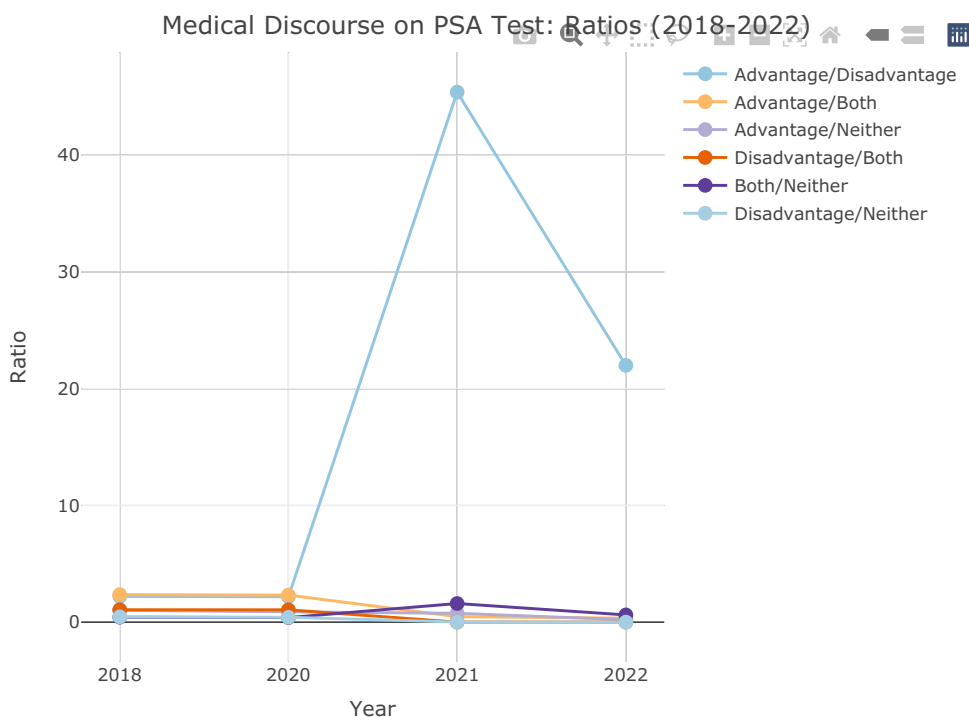
**Problem#2:** Keeping in mind issues about data integrity and disparities in individuals questioned about this topic, combining datasets in this case may lead to inappropriate representation of temporal trends.

**Solution:** Hence, I added another visualization that looked at the ratios of advantages, disadvantages, both, and neither discussed. This ensured the numbers for each of the types of discussions were relative to each year, thereby reducing speculation about the variations in numbers.

Show

## Visualizing Ratios

Show



The value of 0.43 as a ratio between Both and Neither suggest that there are a fewer number of people being informed about both advantages and disadvantages, as opposed to neither. Note the spike between 2020 and 2021, and the fall extending to the year 2022.

**Pro tip:** Deselect the data for Ratio between advantages and disadvantages (by clicking on the legend), to gain some interesting insights into the values lying close to the X-Axis.

## VISUALIZATION 4:

**HIV Risk-Behaviors by Age:** What are the age-related trends in individuals who demonstrated HIV-Risk behaviors? (2018-2022)

How many people reported of having participated in activities that were determined as contributing to the increased risk of HIV Behaviour? Note: This data is missing for the year 2021.

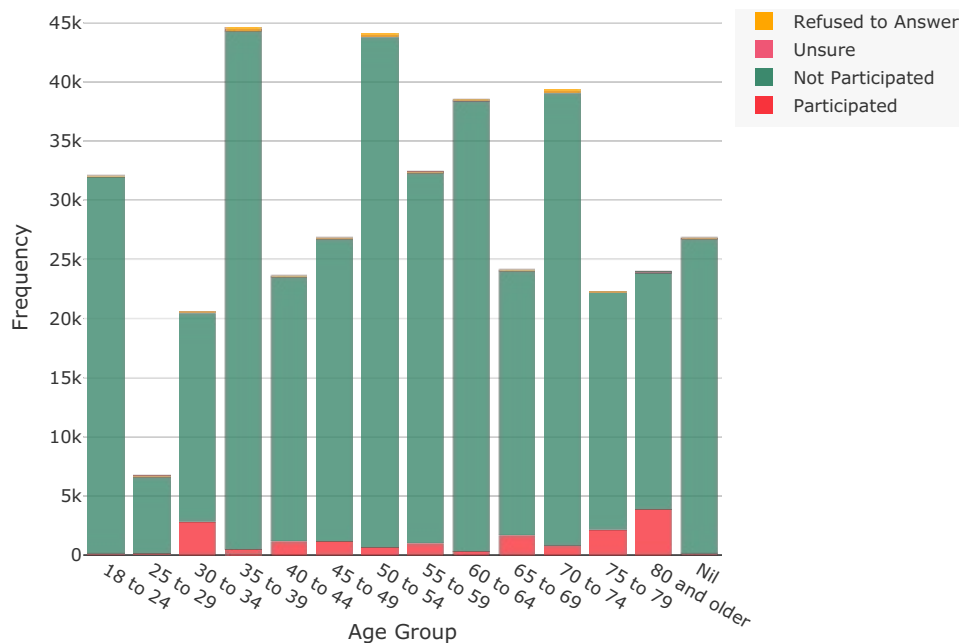
### Cleaning Data

Show

### Visualizing Data

Show

### HIV Risk-Behaviors: Age-based Differences (2018-2022)



Ending on an optimistic note, this visualization provided a lot of interesting insights - not just about the current trends, but also about the nature of data reporting. In a landscape consisting of data and statistical reporting of ongoing challenges in healthcare, it is certainly refreshing to see data pointing towards tangible efforts being made to reduce undesirable health outcomes. This data shows an overwhelming number of individuals across all age groups as not participating towards HIV-Risk Behaviours, aggregated between 2018 to 2022. Such positive trends highlight the room for informed and proactive health behaviours, as well as effectiveness for intervention strategies for the same.

## CONCLUSION

### Summary & Takeaways

This project aimed to investigate the nature of preventative healthcare services in the USA, over a period of 2018 to 2022, through data collected from the BRFSS Health Surveys. Due to project considerations, analysis and visualizations focused on the rate of flu-vaccinations, urban-rural disparities in contraceptive usage, patterns of medical discourse and knowledge disseminated in Prostate Cancer Screening, and HIV Risk Behavior patterns among various age groups. Through actively engaging with the data, I gained invaluable insights into the role of socio-demographic variables on healthcare utilization and disparities pertaining to various geographic regions. Pertaining to data management, I learnt a lot about the most effective methods to manage and manipulate data, keeping in mind considerations on data integrity and truthful communication of results. Most importantly, I learnt how to tweak visualizations to convey key findings to individuals, both in an aesthetic as well as practical sense.

Lastly, I enjoyed working and discovering various charts and tools that may best convey a story I wished to narrate.

### Future Directions

Carrying forward the idea of preventative healthcare, I would like to explore the following questions through this dataset. (1) With the variety of data available in the dataset across decades, I would explore changes in health outcomes longitudinally - such as diabetes, cardiovascular conditions, and cancers. (2) I would focus on understanding health behaviors and its impact on outcomes, utilizing statistical analysis. (3) With information available for specific regions, I would like to explore the reported data in tandem with programs and interventions introduced to assess its role in reported behaviours.

## REFERENCES

1. 2018 dataset CDC - 2018 BRFSS Survey Data and Documentation. [https://www.cdc.gov/brfss/annual\\_data/annual\\_2018.html](https://www.cdc.gov/brfss/annual_data/annual_2018.html) ([https://www.cdc.gov/brfss/annual\\_data/annual\\_2018.html](https://www.cdc.gov/brfss/annual_data/annual_2018.html))
2. 2019 dataset CDC - 2019 BRFSS Survey Data and Documentation. [https://www.cdc.gov/brfss/annual\\_data/annual\\_2019.html](https://www.cdc.gov/brfss/annual_data/annual_2019.html) ([https://www.cdc.gov/brfss/annual\\_data/annual\\_2019.html](https://www.cdc.gov/brfss/annual_data/annual_2019.html))
3. 2020 dataset CDC - 2020 BRFSS Survey Data and Documentation. [https://www.cdc.gov/brfss/annual\\_data/annual\\_2020.html](https://www.cdc.gov/brfss/annual_data/annual_2020.html) ([https://www.cdc.gov/brfss/annual\\_data/annual\\_2020.html](https://www.cdc.gov/brfss/annual_data/annual_2020.html))
4. 2021 dataset CDC - 2021 BRFSS Survey Data and Documentation. [https://www.cdc.gov/brfss/annual\\_data/annual\\_2021.html](https://www.cdc.gov/brfss/annual_data/annual_2021.html) ([https://www.cdc.gov/brfss/annual\\_data/annual\\_2021.html](https://www.cdc.gov/brfss/annual_data/annual_2021.html))
5. 2022 dataset CDC - 2022 BRFSS Survey Data and Documentation. [https://www.cdc.gov/brfss/annual\\_data/annual\\_2022.html](https://www.cdc.gov/brfss/annual_data/annual_2022.html) ([https://www.cdc.gov/brfss/annual\\_data/annual\\_2022.html](https://www.cdc.gov/brfss/annual_data/annual_2022.html))