

Smoking in the U.S.

Lesi He, Wanxin Qi, Tsang Sze Pui, Ke Xu, Ziyue Yang
12/11/2021

Motivation

It is known that smoking causes a lot of health problems, but according to the Centers for Disease Control and Prevention (CDC), there were about 34 million U.S. adults smoke cigarettes in 2019, and 58 million nonsmoking people were exposed to secondhand smoke that severely negatively affected their health. Especially since the first laboratory-confirmed case of Covid-19 appeared in the United States on January 20, 2020, bad addictions such as smoking have been highlighted by health organizations looking at their relationship with Covid-19. The Centers for Disease Control and Prevention pointed out, “Being a current or former cigarette smoker can make you more likely to get severely ill from COVID-19. If you currently smoke, quit. If you used to smoke, don’t start again. If you’ve never smoked, don’t start.” Besides, people with certain medical conditions are more likely to get severe Covid-19, and smoking could make their situation worse, because smoking harms the immune system and can make the body less successful at fighting disease. With this information, we examined the trend of smoking in the U.S. in recent years.

Our main objective is to provide information to health organizations and healthcare facilities about the characteristics of smokers and people who quit smoking in the United States, as well as the specific populations which need more attention.

Initial Questions

1. What is the overall smoking trend in the years between 2017 and 2020? What is the distribution of smoker among different age group, race, and sex?
2. Which states has the highest and the lowest smoking proportion? Which states has the largest increase/ decrease in smoking proportion over the past 4 years?
3. What is the distribution of smokers among minority groups compared to majority groups, during the Covid-19 pandemic?
4. Is there a decrease in current smoking rates among people with specific medical conditions during the Covid-19 pandemic compared to the previous year?
5. What is the situation of quitting smoking in 2020?

Data Source

- *National Health Interview Survey (NHIS)*
The datasets of 2016 to 2020 were downloaded from the websites as mentioned above and were imported. We selected several potential variables for the project and created a YEAR variable to combine the datasets for analysis. The value of the data was given in number indicating the specific response of the participants, so we mutated and recoded the variables based on the codebook. For example, the disability variable was renamed from the `disab3_a` variable in the original dataset, and we transferred “1”, “2”, “9”, to “yes”, “no”, “don’t know”, respectively, indicating if the respondent has disability. Then

we mutated the variables into factors, except for variables like year and age which are continuous. There were several variables containing missing value, and we filled them with “missing”. Lastly, we filtered the dataset based on the smoke variable that we kept only current every day smoker, current some day smoker, former smoker, and never smoker, excluding those with unknown status.

The data can be accessed at <https://www.cdc.gov/nchs/nhis/2020nhis.htm>

- *Behavioral Risk Factor Surveillance System (BRFSS)*

The dataset was downloaded from BRFSS. It contains the tobacco use information of participants from all age, race, sex, and states. Since we are focusing years between 2017 to 2020, filter was applied to filtered out year equals to 2017-2020. There are 27 columns in the data, we only selected few columns useful to our analysis, which were locationabbr (State Name), geo_location(latitude and longitude for map), sample_size (sample size), topic(only question related to current smoking status is selected). Then we summed up the number of smokers in each state and calculated the smoking proportion with respect to their states.

The data can be accessed at https://www.cdc.gov/brfss/annual_data/annual_2020.html

- *CDC STATE System Tobacco Legislation – Tax*

Tax data was obtained from CDC open data website. We have three legislation categories (Combustible Tobacco Tax, Non-Combustible Tobacco, and Stamp Tax. Cigarette tax is the most common tax among 58 regions and takes most available data (>90%) in this dataset (Some region did not specify tax for less common types of combustible and non-combustible tobacco). In this case, we selected cigarette tax value (\$ per pack) under Combustible Tobacco category in 2021 and drop out meaningless columns as well as NA values. After that, we ranked 58 regions by cigarette tax value.

The data can be accessed at <https://chronicdata.cdc.gov/Legislation/CDC-STATE-System-Tobacco-Legislation-Tax/2dwv-vfam>

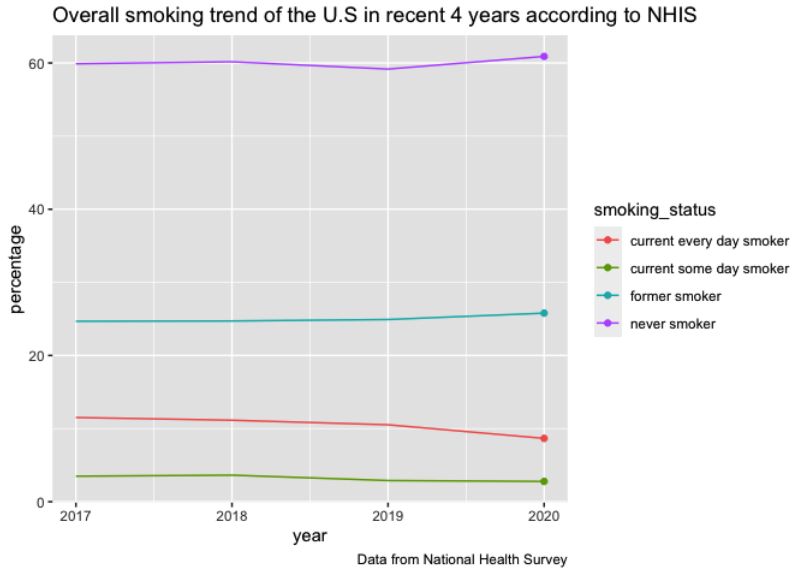
- *Quitline – Service Utilization – 2010 To Present*

The dataset was downloaded from the CDC open data website. It was filtered by locations (states in the U.S.) because it contains data from other regions. Also, it was filtered by the categorical variable Counseling and/or Medications for studying smokers who seeks for help in counseling and/or medication. Then, we filtered the data by year because we want to focus on smokers in 2020. Since the original data value was the number of calls per 1000 tobacco users, we divided the value by 1000 for better readability. Finally, the dataset is ready to be visualized.

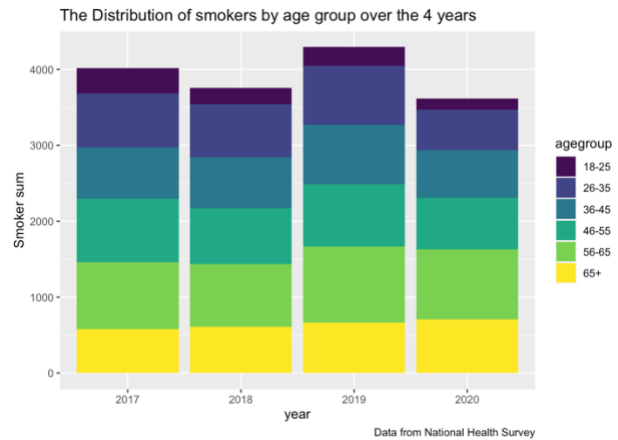
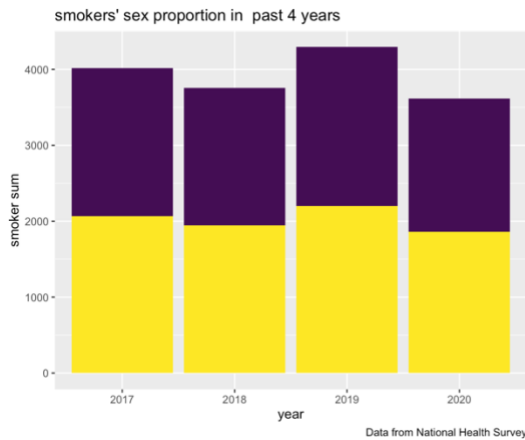
The data can be accessed at <https://chronicdata.cdc.gov/Quitline/Quitline-Service-Utilization-2010-To-Present/equ4-92qe>

Exploratory Analysis

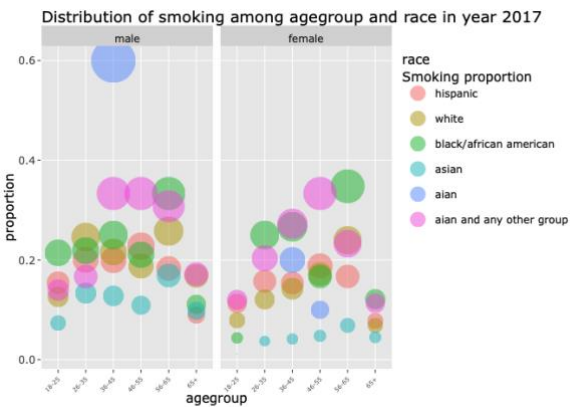
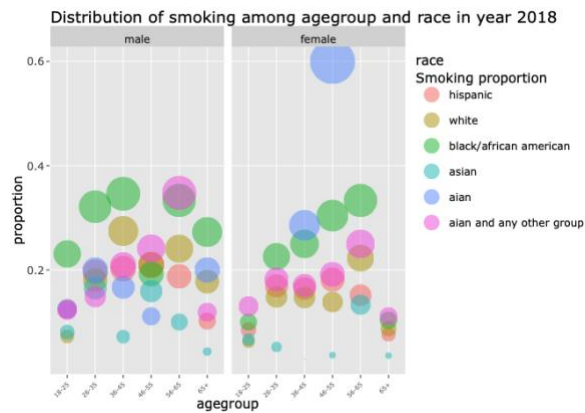
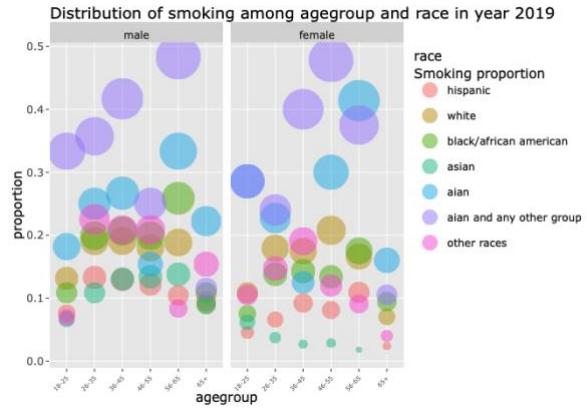
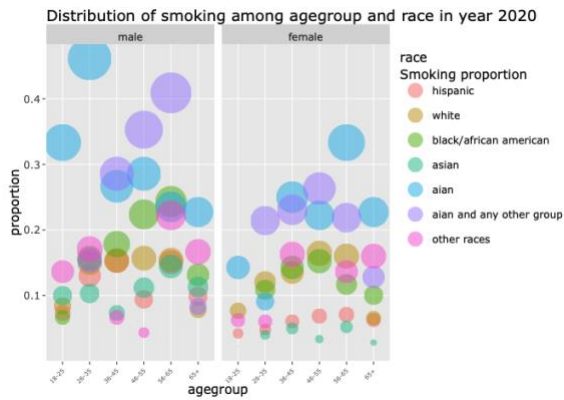
Overall Smoking Trend



There is a slightly increase in the percentage of former smoker and never smoker from 2017 to 2020. The percentage of former smoker increased from 24.6% to 25.8%. The percentage of never smoker increased from 59.9% to 60.9%. Meanwhile, the percentage of current every day smoker and current some day smoker were slightly decreased by 2.0% and 0.7% respectively.

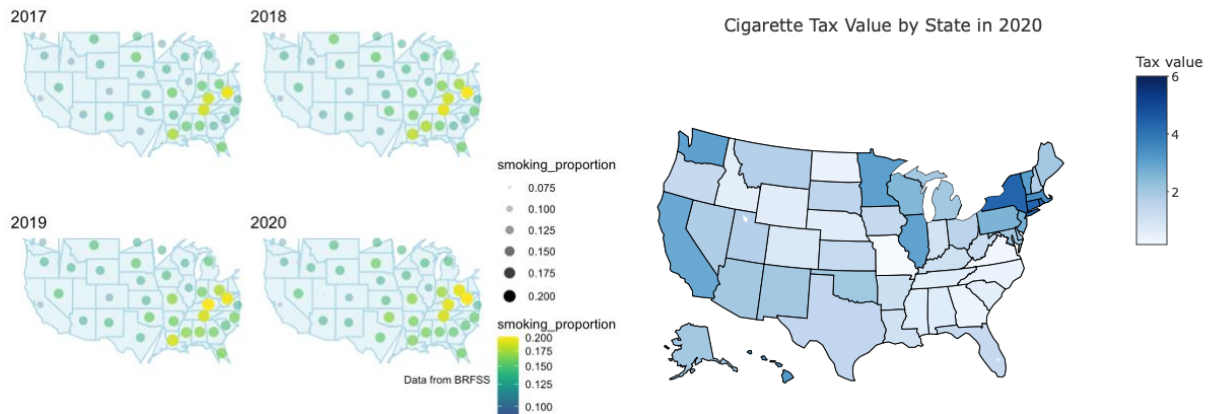


For people with different sex, the plot shows that there were more male smokers than female smokers every year of the past four years. Besides, the age group 56-65 has the most smokers every year in the past four years. The age group 36-45 and 46-55 are slightly lower than 56-55 but worth noting as well. The age group 18-25 has the least smokers. There was little difference in the age distribution of smokers from 2017 to 2020. Smokers in the 18-25 age group showed a decrease after 2017. This is a good sign.



We compared the smoker proportion of each year among different age, sex, and race groups. The smoker proportion is obtained by dividing the number of smokers to the total sum of people in the respective group. In 2020, there is a high proportion of smokers in AIAN/AIAN and any other group. Meanwhile, Asian group, especially female, has a comparatively low smokers' proportion. In 2019, the smoking proportion of male AIAN and other group is the highest in all age group except 65+. The smoker proportion in male AIAN and other group aged 56-65 is especially prominent, which reached to 48.3%, meaning that nearly half of the group are smokers. Meanwhile, Asian female has the lowest smoking proportion among all age groups. In 2018, the smoker proportion in female AIAN aged 46-55 is especially prominent, which reached to 60%, meaning that more than half of the group are smokers. There is a slightly increasing trend in smokers' proportion in female Black/African American along with the increase of age group between 18-25 to 56-65. In 2017, there is a comparatively low smoking proportion in both female and male Asian group among all age group. Overall, the smoking proportion of Asians, especially females is the least among age group.

Overall Smoking Trend by States



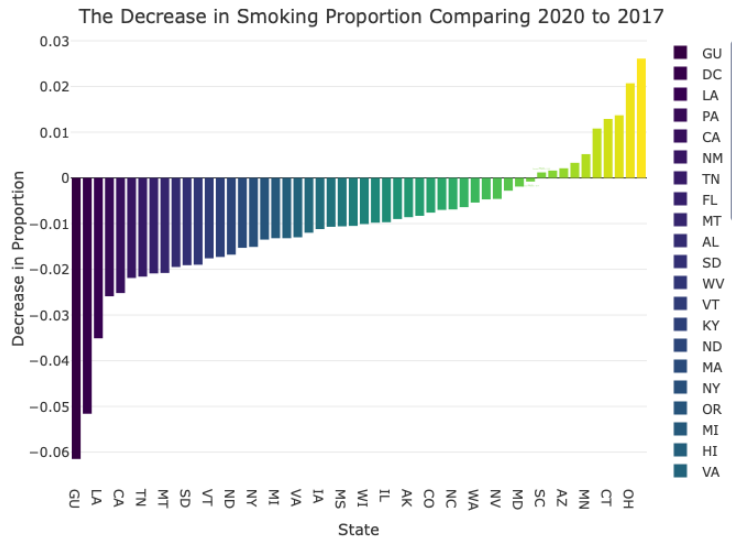
The more yellowish and bigger the dot, the higher the smoker's proportion. Conversely, the bluer and smaller the dot, the lower the smoker's proportion of the state. This graph gives an overview of the distribution of smoking proportion in the U.S. States in the southeast had higher smoking rates in the past 4 years. It may be explained by the states' cigarette tax. By comparing the state's tax and the smoking rates, it can be observed that southeast has a comparatively low cigarette tax when compared to other areas.

2017	2018	2019	2020
Guam, 0.226	West Virginia, 0.212	Kentucky, 0.206	West Virginia, 0.202
West Virginia, 0.221	Tennessee, 0.205	West Virginia, 0.201	Ohio, 0.196
Tennessee, 0.212	Kentucky, 0.199	Tennessee, 0.193	Kentucky, 0.194

The table shows the three highest smoking proportion with the states in each year. West Virginia, Kentucky, and Tennessee are the most seen in this table. West Virginia had 2nd, 1st, 2nd, 1st proportion of smoking in 2017 (22.1%), 2018 (21.2%), 2019 (20.1%), and 2020 (20.2%), respectively. Tennessee had the 3rd, 2nd, 3rd proportion of smoking in 2017 (21.2%), 2018 (20.5%), 2019 (19.3%), respectively. It shows a slightly decreasing trend with time, which is a good sign. Kentucky had the 3rd, 1st, and 3rd proportion of smoking in 2018 (19.9%), 2019 (20.6%), and 2020 (19.4%), respectively. In addition, Guam only had the highest smoking proportion in 2017 (22.6%), and not appeared in top 3 later. The state government might have paid attention to the smoking problem. However, Ohio became the 2nd highest smoking proportion in 2020 (19.6%).

2017	2018	2019	2020
Utah, 0.082	Utah, 0.082	Utah, 0.071	Utah, 0.07
Puerto Rico, 0.093	Puerto Rico, 0.09	Puerto Rico, 0.085	California, 0.085
Connecticut, 0.106	Washington, 0.104	California, 0.101	Puerto Rico, 0.089

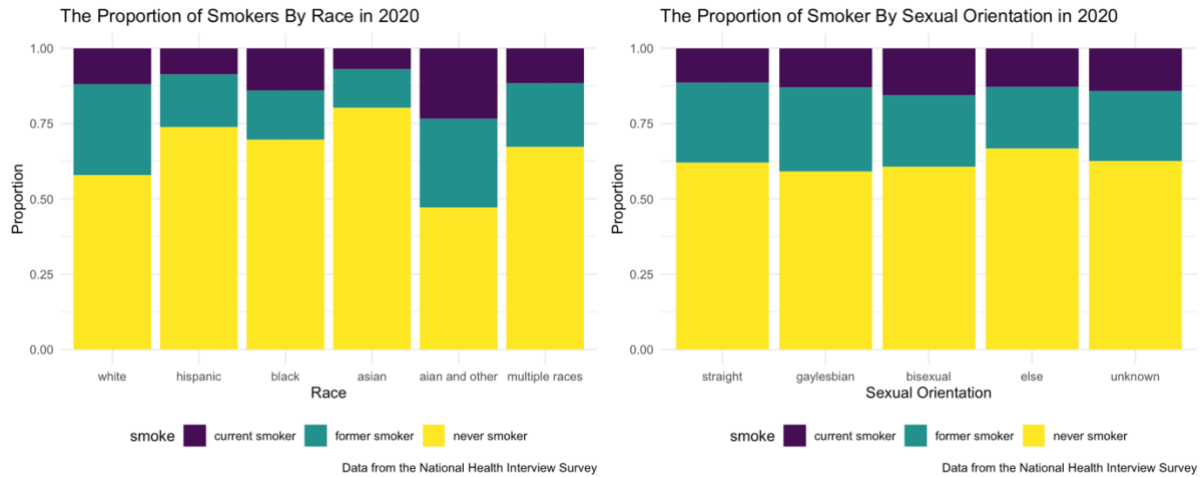
The table shows the three lowest smoking proportion with the states in each year. Among all the states, Utah has the lowest smoking proportion in the four years, which were 8.2% (2017), 8.2% (2018), 7.1% (2019), 7.0% (2020). Puerto Rico has the 2nd lowest smoking proportion in 2017 (9.3%), 2018 (9%), and 2019 (8.5%), and the 3rd in 2020. California also did well in 2019 (10.1%) and 2020 (8.5%). They even show a slightly decreasing trend with years. However, Connecticut and Washington only appeared once in 2017 (10.6%) and 2018 (10.4%) respectively.



Guam has the highest decrease in the proportion of smoking comparing 2020 to 2017 with a value of 6.15%. District of Columbia has the second large decrease of smoking proportion with a value of 5.16%. On the contrary, Texas has the highest increase in the proportion of smoking with a value of 2.61%. Ohio has the second highest increase in the proportion of smoking with a value of 2.07%.

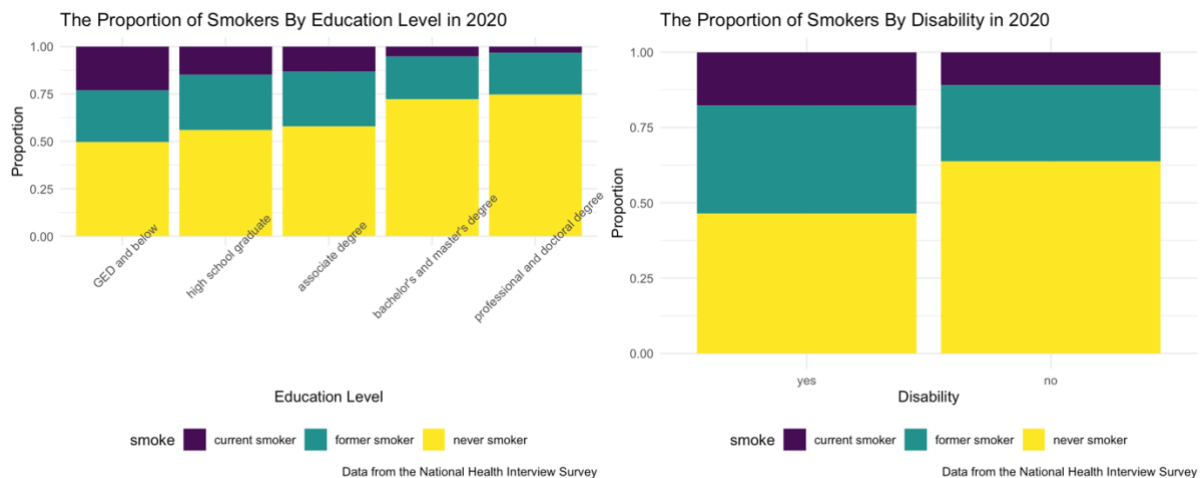
Smoking During the Covid 19 Pandemic

We aim to look at the distribution of current smoker, former smoker, and never smoker among different subgroups defined by demographic characteristics, and if the proportion of current smoker with certain medical conditions decreases in 2020 comparing to 2019.



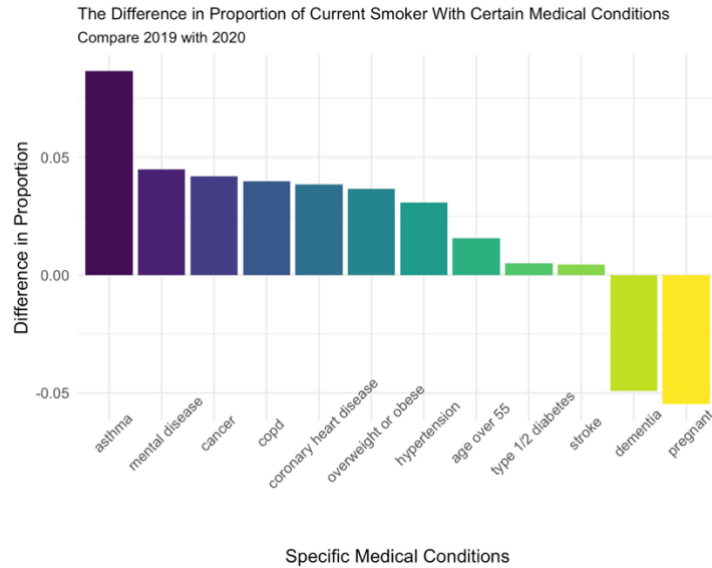
White people are the majority group in the U.S. AIAN and Multiple Races only had a very small number of people. Based on the graph, AIAN has the highest smoking rate combining current smoker and former smoker, and the current smoker rate is worth noting. Asian has the lowest proportion of current smoker and former smoker.

People with straight sexual orientation is considered as majority group, and all other sexual orientations are considered as minority groups. Based on the graph, the proportion of smoker (current + former) within the gay/lesbian group was larger than straight; Bisexual people has the largest current smoker rate, which can be clearly seen in the graph.



Based on the graph of smokers by education level, the proportion of either current smoker or former smoker decreased with the increase of education level.

People with disability is considered as a minority group in the U.S. Based on the graph, the proportion of either current smoker or former smoker are significantly larger than people without disability.

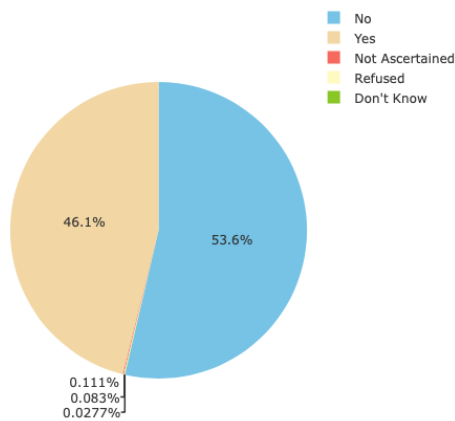


We selected the variables based on the instruction by CDC. People with these medical conditions are more likely to get severe Covid-19. Since smoking would weaken their immune system, making them harder to defend diseases, the proportion of current smoker should have decreased in 2020. The graph shows the difference in proportion of current smoker with certain medical conditions comparing 2019 to 2020. Specifically, we subtracted the proportion in 2020 by 2019, thus, we expected to see a graph with positive values. People with asthma has the highest decrease in proportion of current smoker; People with diabetes and stroke has very few changes in the proportion of current smoker; The proportion of current smoker with dementia and pregnancy increased in 2020, which worth noting.

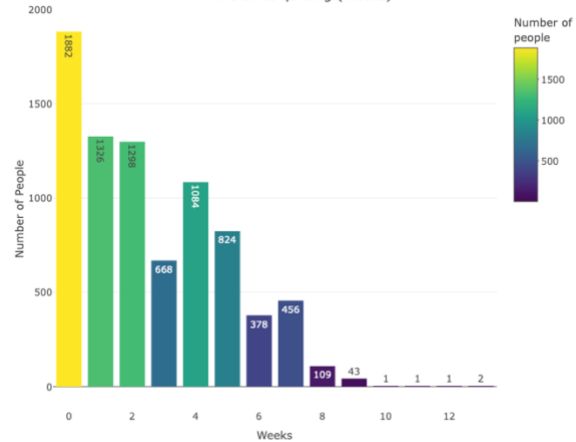
Quit Smoking

According to CDC, fewer than one in ten adult cigarette smokers succeed in quitting smoking each year. In 2018, only 7.5% of adult smokers successfully quit smoking in the past year. Thus, we aim to learn more about the situations of the group of people who are willing or trying to quit smoking in 2020.

Have you tried to stop smoking in the past year?



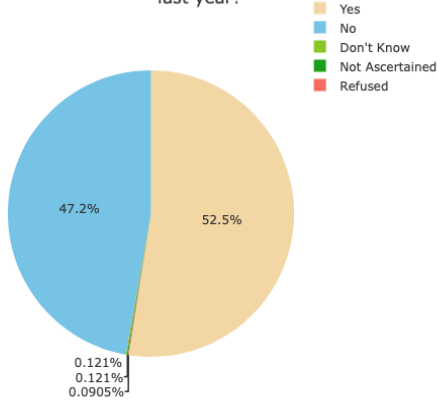
Time since quitting (Weeks)



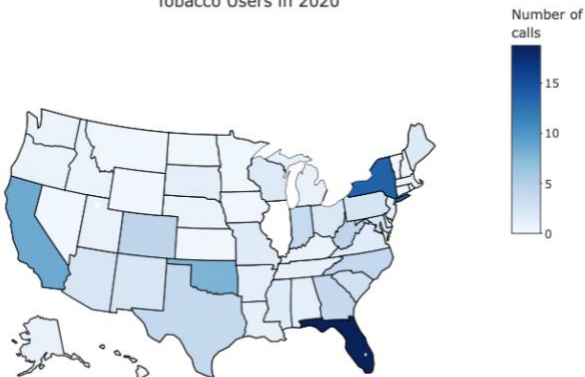
The pie chart illustrates the willingness to quit smoking of the smokers who answered the survey. Less than half, 46.1%, of the people have tried to stop smoking in 2020. Interested in this outcome, we investigate the number of days that these people persist in stop smoking, which is presented in the bar chart. It can be seemed that only a few people have quitted smoking for more than 10 weeks when they answered this survey. However, it usually takes around 12 weeks to withdraw all nicotine symptoms, despite heavy or light smokers. Many of them tried to stop smoking for less than 3 weeks. Due to limited data points, we were unable to examine whether these people quit smoking successfully in 2021. Therefore, the number of people who quit smoking in 2020 may increase.

We kept exploring two factors that may be related to quit smoking: Doctor's advice and tobacco tax. The plots listed below are current situations related to these factors.

Has a doctor advised you about stop smoking last year?

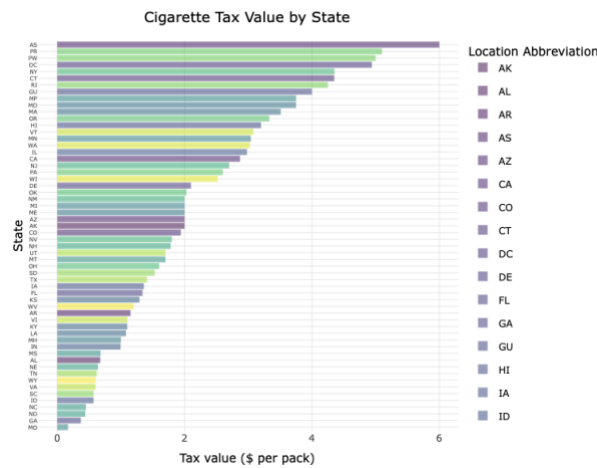


Number of Quitline Calls for Counseling/Medication per Tobacco Users in 2020



From the NISH dataset, we plotted the variable related to the question: Has a doctor, dentist, or other health professional advised you about ways to stop smoking or prescribed medication to help you quit? The pie chart shows that 52.5% of respondents said “Yes” to this question. To further explore this situation, we study the dataset related to the quit line calls received per tobacco users. Through the map, we can see that the greatest number of quit line calls received is in Florida, followed by New York state, Oklahoma, and California. This observation suggests

that a lot of smokers in these states cares about ways to stop smoking, and they are willing to ask for doctors' advice or take medication to quit smoking.



Legislation on Tobacco includes three categories:

- Combustible Tobacco Tax
- Non-Combustible Tobacco
- Stamp Tax

Combustible Tobacco covers a majority number of smokers in the country and among five sub-groups in Combustible Tobacco, cigarette plays dominate role (more than 95%), so we mainly focus the cigarette tax among regions in the U.S. The graph shows the amount of tax (\$ per pack) for 58 regions in the US (50 states and 8 territories that belongs to the U.S.). American Samoa (AS) has the highest cigarette tax, with 6 dollars per pack, followed by Puerto Rico (PR), Palau (PW), and District of Columbia (DC). According to our previous findings, these four regions have a higher decrease of people who smoke, and DC has the highest decrease. Among 50 states in the country, New York (NY) has the highest tax, with 4.35 dollars per pack. As for the states with low cigarette tax, 35 regions have cigarette tax lower than 2 dollars and 14 regions lower than 1 dollar. Ohio and Texas have the largest increase in smokers from 2017 to 2020, both of which have cigarette lower than 2 dollars.

Reflection

After we combined all the parts, we found that cooperation caused a lot of problems. Everyone has a different idea of how each plot should be made, and everyone's coding habits are different. It can now be seen that some of the information told by graphs overlap in the "Overall Trend of Smoking" and "Smoking and COVID-19" sections. However, after thinking thoroughly about the implications of these charts, we decided to keep some of them. There are still some plots with problems, such as that we should have used the proportion/percentage of smoker rather than the number of smokers to make plots, but we also need to use the number of smokers to illustrate some issues, and too many tables with plots would make the report and web page too complicated. At this point, we may need to think more and express what we wanted to illustrate in a better format.

Statistical Analysis

Statistical analysis is not applicable to our project, because the circumstances that lead a person to smoke or quit are complex. Smoking or quitting is a personal choice that may be influenced by factors such as education level, sexual orientation, and social networking, but such analyses are too large. It is also difficult to extract a few or dozens of factors from hundreds of questions in the health surveys to produce a statistically significant model of smoking. If to analyze and predict smoking rates in the United States, more datasets are needed to calculate the smoking rate in the previous year, which is a huge amount of processing, and the variables can change from year to year, especially since Covid-19 started. Therefore, we just used the data to show the recent situation of smoking in the United States.

Conclusion

There is a slightly decreasing smoking trend from 2017 to 2020, which meets our expectation. Men and people aged 56 to 65 consistently make up the majority of smokers. Smoking rates have dropped among 18-to-25-year-olds. Through comprehensive comparison of smoking rates, AIAN and other groups have the highest smoking rate, regardless of male or female. Smoking rates are lowest among Asians, especially women. The southeast has a higher smoking rate from 2017 to 2020 and based on the lower cigarette taxes in the southeast in 2020, we infer that there is some association. If there are more suitable data sets, we can do further investigation. West Virginia, Kentucky, and Tennessee kept the top three smoking rates for four years out of more than 50 states in the U.S. Utah and Puerto Rico kept the top three lowest smoking rates for four years, and California has done well in the past two years. Guam has the largest decrease and Texas has the largest increase of proportion of smoking comparing 2020 to 2017.

In 2020, minority groups defined by race, sexual orientation, and disability, and people with lower education levels still had a severe problem of high smoking rate. People with diabetes and stroke, who should quit smoking by 2020, did not pay much attention to this problem. The proportion of smokers increased among pregnant women and dementia patients. These are what we were not expected. Health organizations and medical facilities should focus about these patients and vigorously publicize the dangers of smoking to their fight against disease.

We observed that there are less than half of smokers had make attempts to quit smoking in 2020. Among those who tried to quit smoking, more than 70% of them had stop for smoking for only one month or less. However, more than half of them responded that a doctor, dentist, or other health professional have advised them to stop smoking in 2020. Across the U.S., many smokers have called the quit line to for counseling and/or medication related to quit smoking, especially in the Florida, New York, and California. This is a positive outcome suggesting that many smokers are willing to quit smoking. Tobacco tax remains steady throughout these years, and we consider tobacco tax only has a slight influence on proportion of quitting smoking. Even though several regions with higher taxes have a higher decrease of smokers, more influencing factors are related to people themselves as well as environment.

Discussion

The strength of our study is that people can get a sense of trends in smoking and quitting in recent years. We also looked at the smoking situation in the context of the COVID-19 pandemic. These could provide information to health organizations and healthcare facilities about which populations should be highlighted and noticed.

Our study has several limitations. First, as discussed in the statistical analysis section, we find it difficult to come up with a statistically significant model due to the complexity of variables. These survey data may be biased because respondents respond to the questionnaire based on their own situation and feelings. Questions such as “smoking frequency” were answered by self-assessment, which may vary from person to person. In the study, there was no clear experimental data to reasonably justify these manifestations. It leads to the second limitation, as there were no strict rules to define a smoker. Some people may smoke only occasionally for social purposes, while others may smoke packs or even packs of cigarettes a day. We can only categorize roughly by “current every day smoker”, “current some day smoker”, “former smoker”, and “never smoker”. Third, due to limited data, we focused only on cigarettes, while there are many other forms of tobacco smoking, such as small cigarettes, cigars, and other non-combustible smoking, such as chewing tobacco and e-cigarettes.

All in all, smoking is not a healthy behavior for the smokers as well as the people around the smoker because smoking will cause many health issues. According to the CDC, quitting smoking has health benefits at any age, no matter how long or how much you have smoked. We hope that more policies can be enforced to encourage current smokers to quit smoking and further reduce smoking rates in the United States.