

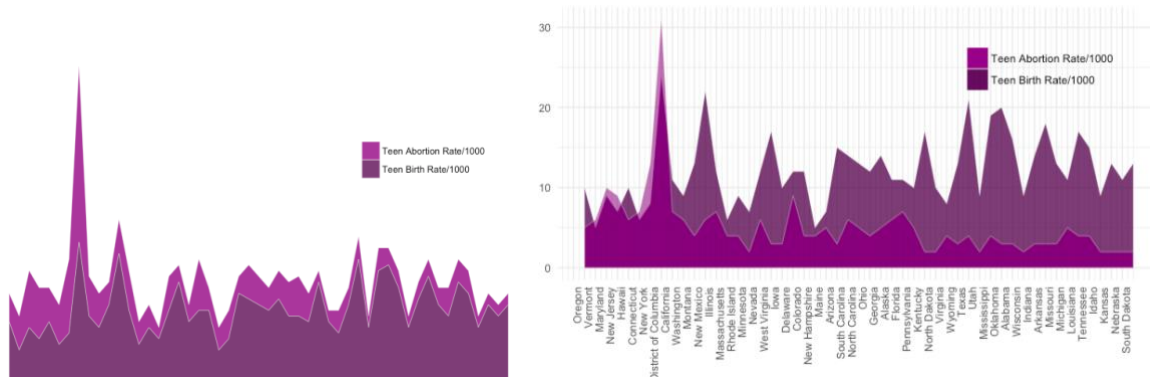
## Process Book

**Data:** Our primary datasets were from Guttmacher, but because we were interested in more than just statistics on abortion, we used data from “safeplaceproject.com” for information about driving times to and from abortion clinics and text data scraped from website platforms for pro choice (Planned Parenthood, Center for Reproductive Rights, etc.) and pro life organizations (National Right to Life, March for Life, etc.). Please see cited sources in the Shiny app.

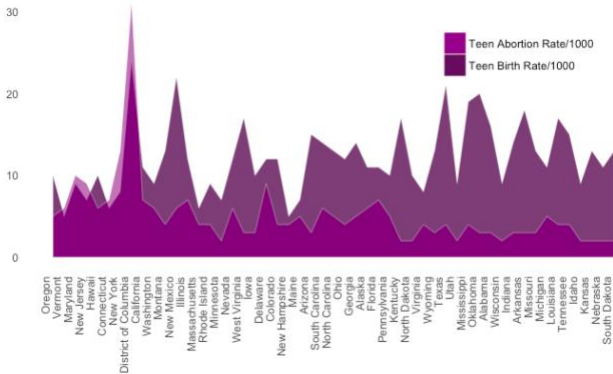
**Sophie** created visualizations about teens (Guttmacher defined teens as age 15-19) and their rates of abortions and births in each state. Using additional information about a comprehensive score on women’s health from the [Status of Women dataset](#), she created three charts to visualize the choices teens made in different states with varying women’s health grades. Unsurprisingly, in states with higher [composite scores](#) (“grades,”) more teens had abortions whereas in states with lower “grades,” more teen pregnancies ended in birth.

First, two bar charts were made -- one for teen abortion rate and the other showing teen birth rates. In Shiny, these two bar charts were overlaid such that a user could toggle between the two.

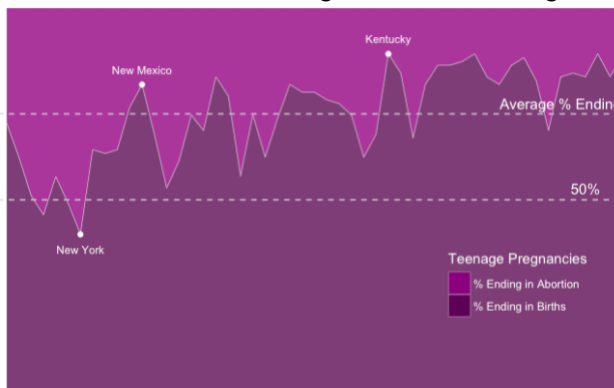
An area chart was made as a second way to view the same data. Previous versions of this area chart were stacked on top of each other rather than laid down one after another; as a group, we decided the stacked area chart was hard to understand as it presented seemingly misleading information. The two visualizations can be seen side by side below. The one to the *right* was chosen as a better representation of teen abortion & birth rates/1000 people.



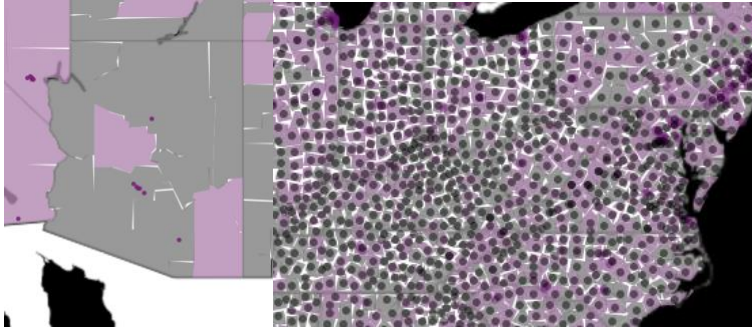
After selected, all grids were taken away to produce the following cleaner image:



The final visualization she created was a slightly altered filled stacked area chart that shows the *percent* of teen pregnancies ending in abortion per state. The initial static image highlighted a few states with either markedly high abortion rates or markedly low abortion rates (high birth rates). In the Shiny App, a user can click on a state of interest and the state, along with information about its women's health "grade" appears as a dot on the visualization. Text is displayed that tells the user what percentage of teen pregnancies end in births or abortions in that state. The static image is below; changes made can be seen in the [Shiny App](#).



Lastly, a waffle chart representing the difference between **Nicole** worked on creating the driving times interactive map. Using inspiration from [The Pudding](#), she sought out to improve the first glance understanding of the map. After scraping data from [The Safe Place Project](#), Nicole used a Google API key to find the longitude and latitude of each clinic in the US. In order to map the clinics as point data, a shapefile of US counties was collected, and centroids were calculated using QGIS software. Next, she used the "dist" function in the geosphere package to calculate the distance between the centroids and the clinics, then saved the minimum distance between centroids and clinics. Because of the nature of working with country data, it appears that some of the counties that visually look like they would be within a certain driving distance to a clinic are not shaded as such because the location of the centroid is further away. The size of a county is not uniform across states, which makes it appear that there are fewer counties in the Midwest that have access to abortion clinics. While this may be true politically, it is important to recognize that some lack of access can be attributable to differences in distance. This might explain why the Pudding used honeycombs to represent space in the US rather than county data. Although I received some suggestions to plot the point data to better illustrate the distance between centroids, the map is too difficult to read when counties are small, and it skews the points of clinics.



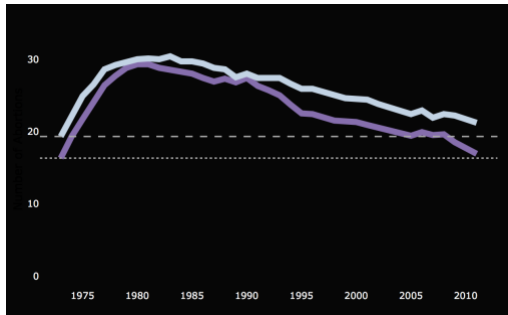
Because of some feedback about confusing the colors, I added some additional comments to the shiny app to explain how to interpret the graph.

After spending some time recoding/cleaning the data, we realized that we were losing some of the intricacies of the abortion data and decided to implement a data table to show how state differ. After some suggestions from classmates, Nicole cleaned up the data table to include things like “Not reported” and “Yes/No” instead of Xs.

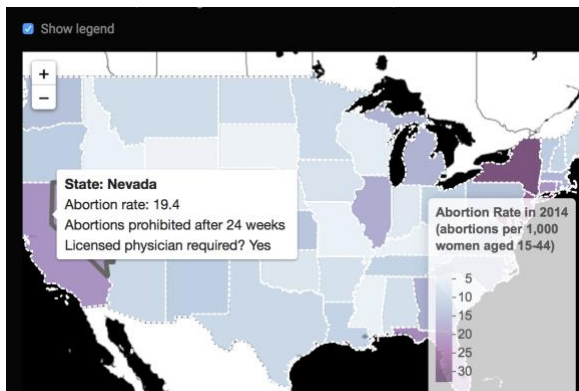
Nicole also scraped the data from pro choice and pro life organizational websites. To do this, she looked at the “About,” “Resources,” and “FAQ” sections of websites in an attempt to obtain similar information from both sides of the arguments. As mentioned in the presentation, a lot of the pro life organizations had stories, which may explain the lower frequencies of words.

Although similar to the wordcloud that Joo used for text analysis, Nicole made a pyramid plot that displays the top 25 words used by the two different organizations. In order to make the best use of the interactivity of shiny, Nicole chose to calculate the top 25 words for each organization plotted against the other in order to show how the words differ. However, it seems that the wordcloud and pyramid plot are representing the same information, and it may be better to condense this information into one visualization.

**Dana** created the time series plot of abortions since Roe v. Wade. She obtained the national abortion rates from 1983 - 2008 from this [2008 paper](#), published in *Perspectives on Sexual and Reproductive Health*, and the 2009 - 2011 data from a [Guttmacher Institute report](#). The data seemed to have consistent values for the years when it overlapped, which made it acceptable to combine the two sources. The time-series plot showed that the rate and ratio of abortions remained spiked and never quite returned to their original levels following the Supreme Court decision, until 2011, the last year in our time-series data. The abortion rate dipped below Roe levels for the first time in 2014, but we did not comfortable combining disparate datasets that had different measurements of abortion in the intervening years. The absence of pre-1973 data, which would have been useful for a before-and-after comparison, speaks to data limitations inherent to this topic: abortions certainly occurred prior, but without the legal legitimacy afforded by Roe, the information was not systematically collected in every state. It also would have been interesting to compare fluxuations in abortion rates across states and over time, particularly to look for any trends in policies affecting outcomes.



Dana was also responsible for the Leaflet choropleth map of abortion rates across the different states, with a hover tooltip showing that state's abortion rate, when abortion access is restricted (if at all), and whether a licensed physician is required to perform abortions. This map is useful for revealing spatial trends in abortion rates, which perhaps unsurprisingly, and like political allegiances, follow coastal patterns. Ideally, it would have been possible to include a `checkboxGroupInput` option allowing the user to toggle between the different restrictions, but after much research, there did not seem to be a way to add those features to filter data in a choropleth map in Shiny.



**Joo:**

## Waffle chart –

Joo created waffle charts to visualize abortion rates across race. The Guttmacher Institute provides data on the number of abortions per one-thousand Black, Hispanic, White, and other women, although this statistic is available for only 30 states. After concluding that visualizing in context is key for this particular question, she combined the waffle charts, so abortion rates of all four races are simultaneously displayed. Toggling would allow the user to gauge variation across states.

As the numbers of abortions ranged from 2 to 54 per thousand women and are thus relatively low, the white color denoting the number of pregnancies that did not end in abortion came to dominate the waffle charts, making it somewhat difficult to truly gauge the variation across race. To overcome this challenge, we added some text reporting the exact percentage of the abortion rate per race.

## **Word cloud -**

Abortion is a fiercely contested subject today and has cultivated the use of a particular language from its advocates as well as its critics. We wanted to illustrate the difference in words used by the two opposing sides through wordclouds and a pyramid chart. We chose to visualize words used by 7 pro-choice organizations and 7 pro-life organizations.

Again, we believed visualizing within context is key to ensuring that readers are able to immediately gauge the difference in language use, so we created a static graph that placed the two wordclouds side by side. They demonstrate subtle yet clear differences. On one hand, language employed by pro-choice supporters include words such as “law” and “states” that we presume are utilized to explain federal and state laws on abortion and how they may affect women’s getting an abortion should they desire the procedure. On the other hand, pro-life supporters tend to invoke emotion by using words like “fetus” and “baby”, likely in an attempt to propagate the notion that life begins at conception and persuade women to seek other options.