

Top 10 hotels in the USA

A data visualization story of sentiment, crime, and weather

Jinxi Luo



Table of Contents

Dashboard (Top 10 Hotels in the USA)	1
Introduction	2
A look at all the hotels	2
Finding high rated and reviewed hotels	3
The safest hotels	7
Date and weather	8
Places to visit.....	9
Conclusion	12
References	12
Appendix: Data Sources	13

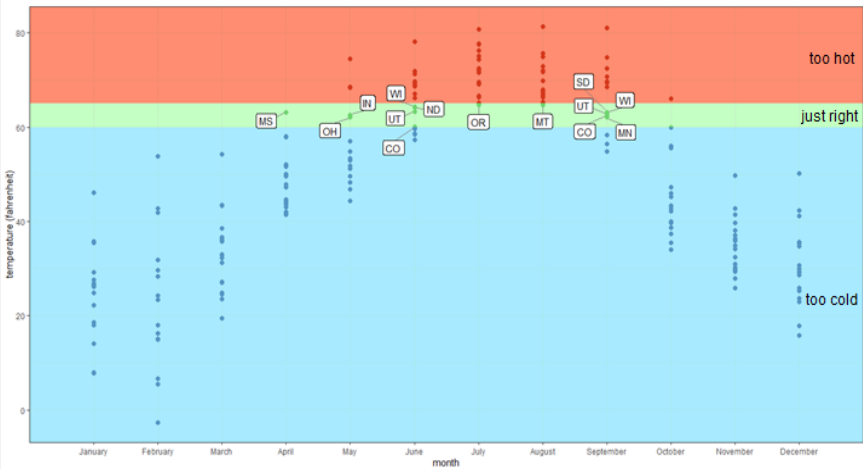
Dashboard (Top 10 Hotels in the USA)



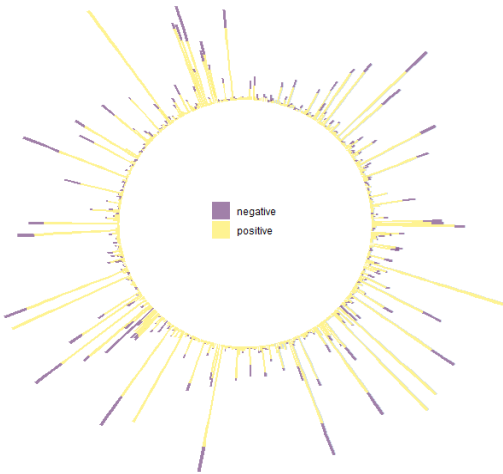
Top 10 Hotels Nearest to Cities

#	Hotel Name	Address
1	Quality Inn	20150 W Bluemound Rd
2	Homewood Suites By Hilton Jackson-ridgeland	853 Centre St
3	Best Western Plus Bandana Square	1010 Bandana Blvd W
4	Best Western Plus Eau Claire Conference Center	3340 Mondovi Rd
5	Tmi Hospitality	5108 Poplar Ave
6	Eagle's Nest Resort	3421 N 14th St
7	Sleep Inn Suites-marshall	10 Nw Minnesota Ave
8	Broadway Hotel	1300 Nwakama St
9	Oxford Hotel	7555 S Union Park Ave
10	Maple Leaf Motel	1000 3rd St Nw

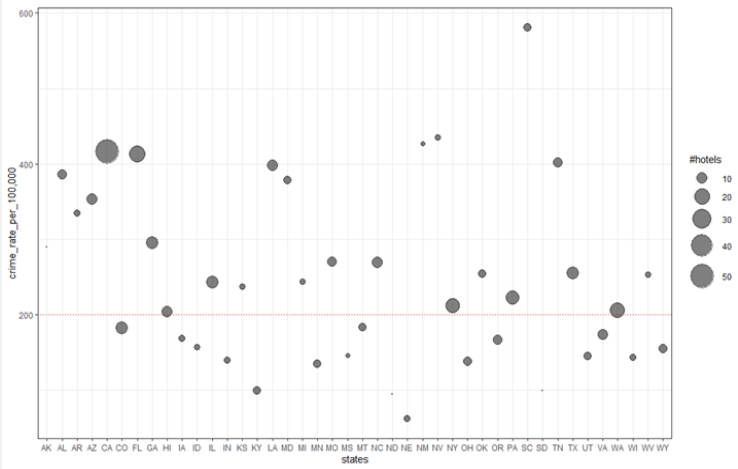
States with the Best Weather Throughout the Year(60-65°F)



Sentiment of Best Reviewed Hotels



Safest States with Crime Rate below 200 per 100,000



Introduction

The goal behind this project is to bring a new perspective into how data can be used to take finding a hotel to stay at to different level beyond common site filters. Instead of scrolling through hundreds of hotel listings, asking around in forums and comparing multiple sources of information at a time with pen and paper. Visualizing all the data you are interested in at a time can help you better make your own informed decisions and help you explore a wider avenue of great hotels you didn't know existed. While the scope of this project is limited by the range of data sources collected. We can still explore the potential of implementing the idea investigating which hotels have the best reviews, lowest regional crime rates, good weather and having plentiful places to visit.

This project will use a list of hotels from the Datafini's Business Database which provides the hotel information of 1433 hotels across the USA and 10,000 of their reviews [1]. View the appendix for more information on how the various sources of data used in this project is connected.

Let us begin by seeing all the hotels we could be going to ↓

A look at all the hotels

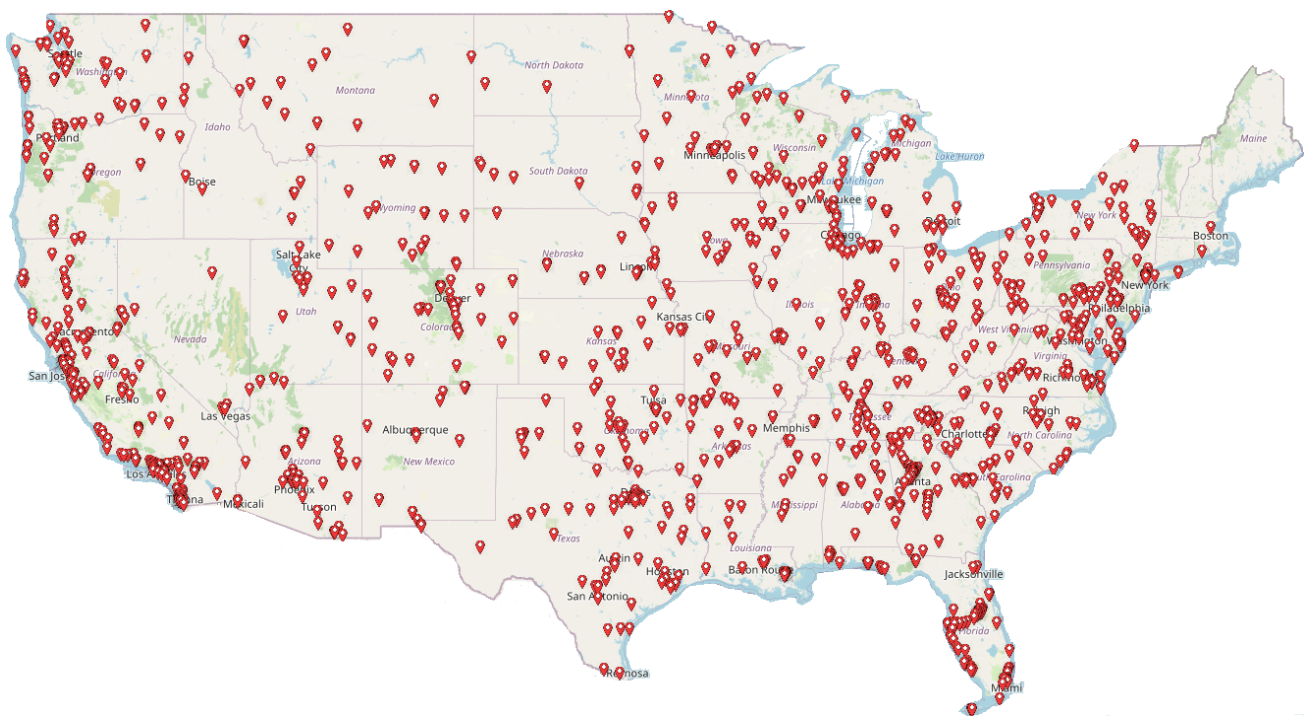


Figure 1: Map of all hotels (marked red) of interest located in the USA, excluding Alaska and Hawaii.

That is a lot to look at! Around a total of 1433 hotels are shown in the figure above, that is a bit too many to consider. Since I trust what others say let us try and filter out some hotels that have received bad ratings first.

Key Take Away!

The majority of hotels have high ratings of 4 to 5 but only a small number of reviews.

At this point, the main concern is of hotels with high ratings and a reasonable number of reviews.

Finding high rated and reviewed hotels

It can be seen in Figure 2, that on average, the hotels have very high ratings with a lot of people giving 4 and 5 stars. Shown by a brighter shift in colour.

There are also hotels that have gotten hundreds of reviews but for the majority this is not true. Figure 3 better shows the skew within the data with most hotels only getting around two reviews.

Given the information seen, all hotels with a rating below four are removed, leaving 956 hotels. That is still a lot, so hotels with two or more reviews are also removed bringing the remaining hotels to 485.

Let's see how the hotels look on the map now ↓

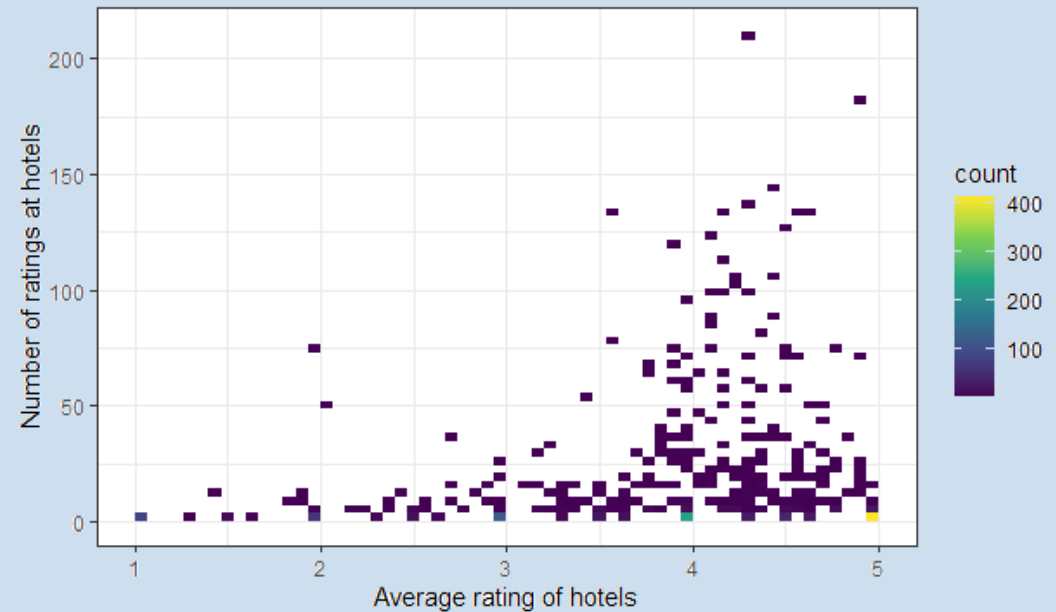


Figure 2: Scatter plot of average ratings for all 1433 hotels and their number of reviews.

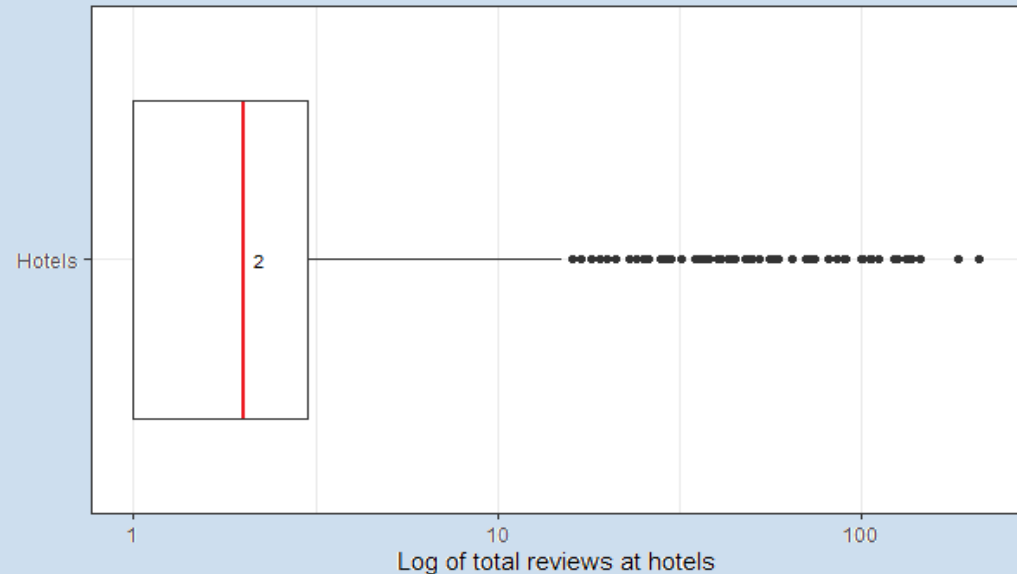


Figure 3: Log transformed boxplot to show skew within number of reviews at hotels.

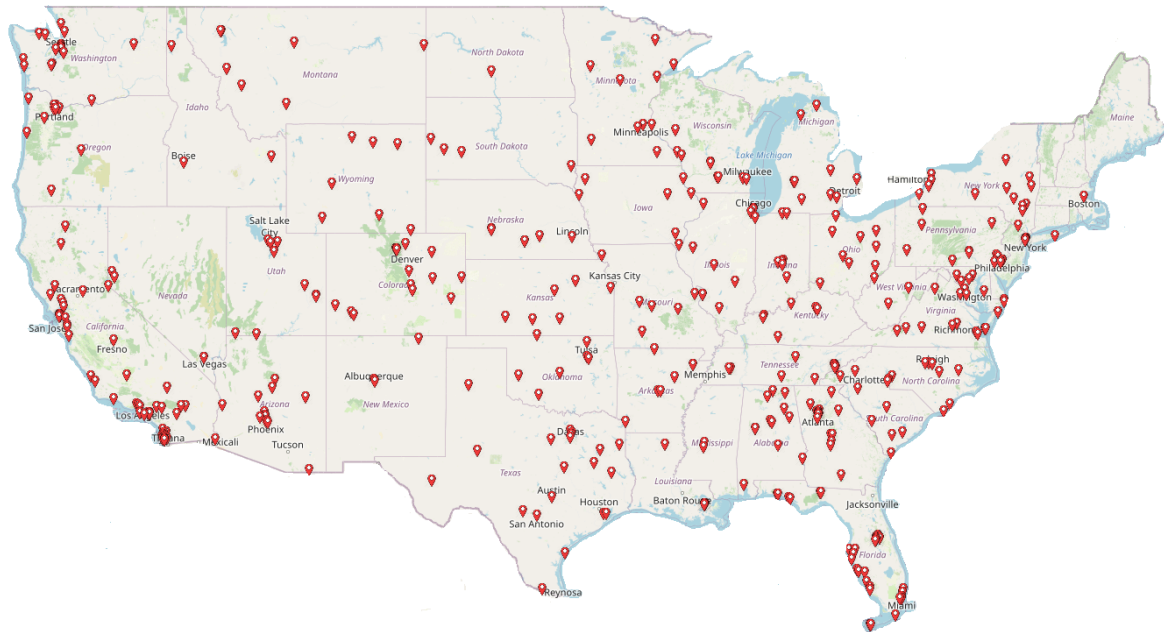


Figure 4: Map of remaining hotels after filtering bad rated and low reviewed hotels.

The map looks much cleaner, but you might have been wondering. Isn't a minimum of two reviews a bit low and why didn't we just only use ratings and reviews numbers to filter 10 best hotels with high ratings and reviews and be done with it?

An answer lies in how much information we are missing out on when filtering. Not everyone that goes to a hotel will leave a review on the website and a popular hotel with lots of reviews might not even be a nice place to stay at. Furthermore there may be a range of other factors that affect how one interprets a good hotel and so it isn't reasonable to put all your eggs in one basket per se, hoping that only ratings and reviews numbers will lead to a "good" hotel.

So, we continue and see if sentiment of the reviews will further help, which has been divided into a positive and negative category. Another data set containing labelled positive and negative reviews of hotels across Europe [2] was used to train a sentiment classifier for our data set. In the stacked bar plot below, the trend of sentiment can be examined amongst the hotels, with positive and negative reviews coloured yellow and purple, respectively.

Most hotels have over 70% positive reviews.

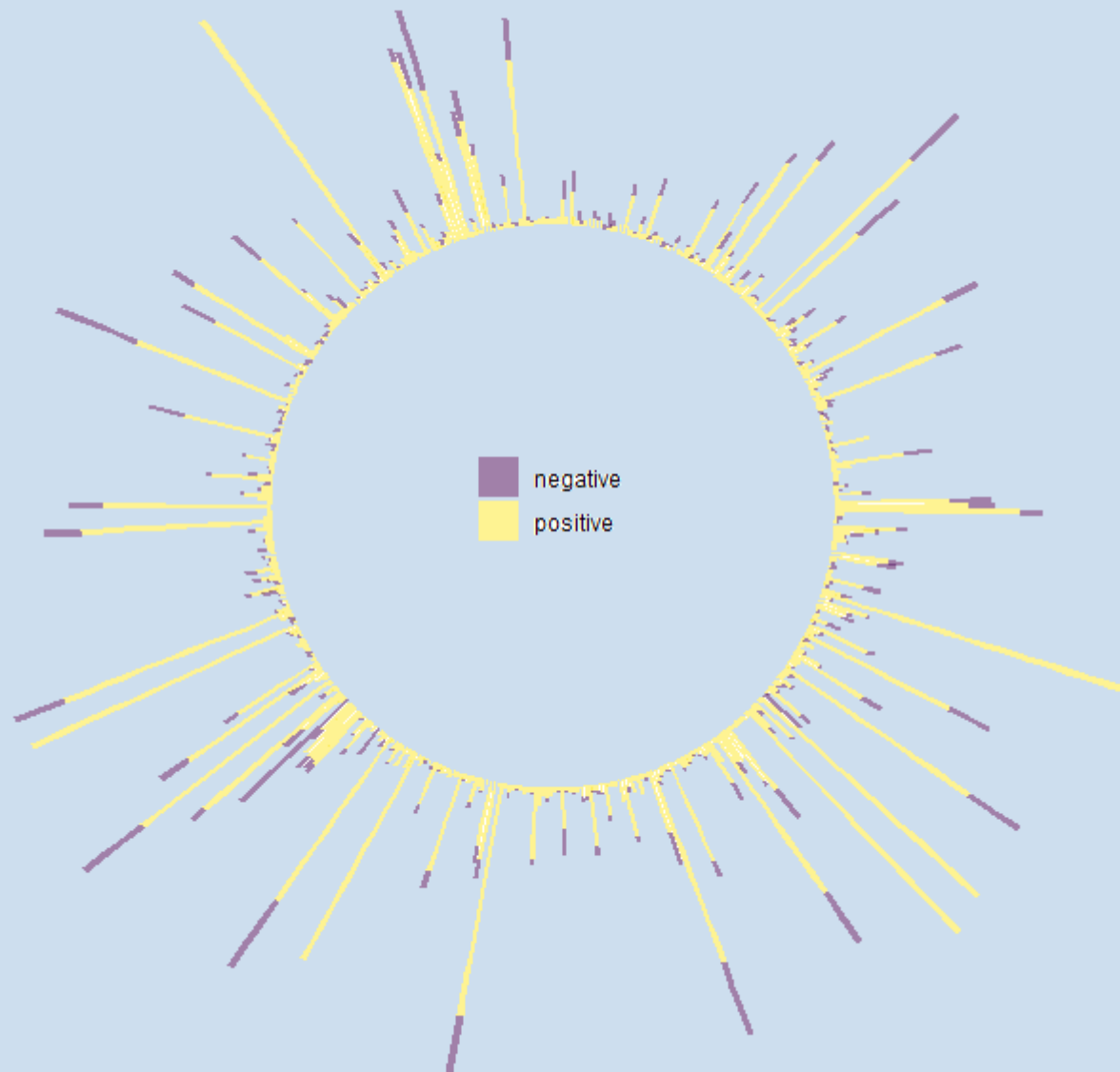


Figure 5: Circular stacked bar plot of positive and negative sentiment for the 485 hotels.

It is good to see that a lot of the remaining hotels are viewed positively, and eye balling the graphs it appears that around 70% of the reviews for the hotels are positive.

Using this division as a heuristic, 151 hotels were removed that had lower than 70% positive reviews. A word cloud of sentiment in Figures 6 and 7 below provides a comparison of general customer concerns.

It is noted that:

- There is little difference in concerns between sentiments in removed and kept hotels.
- Sentiment of remaining hotels seems to contain more positive descriptive words overall.

As the context behind the words are not known, it could be assumed that sentiments for the removed hotels were complaints around a lacking of services and experiences found in the remaining hotels that customers had praised about.



Figure 6: Word cloud of reviews for hotels with over 70% positive sentiment.



Figure 7: Word cloud of reviews for hotels with less than 70% positive sentiment.

The safest hotels

We then consider the next important thing which is safety. The data used in this section contains the crime statistics of 3,137 counties across the USA in 2016 [3]. I will be using the mean (total) crime rate per 100,000 people grouped by states as a statistic to determine the safest hotels to stay at.

From the previous filtering done, we were left with 334 hotels distributed across 42 states. For the crime rate safety threshold, I will pick states that have a crime rate below 200 resulting in 82 remaining hotels. As the 2019 national average for violent crime rate is 370 per 100,000 people, which would mostly be the only type of crime we are worried about considering we are renting [4]. A larger margin should help keep us extra safe, especially when considering the rate of other crimes combined.

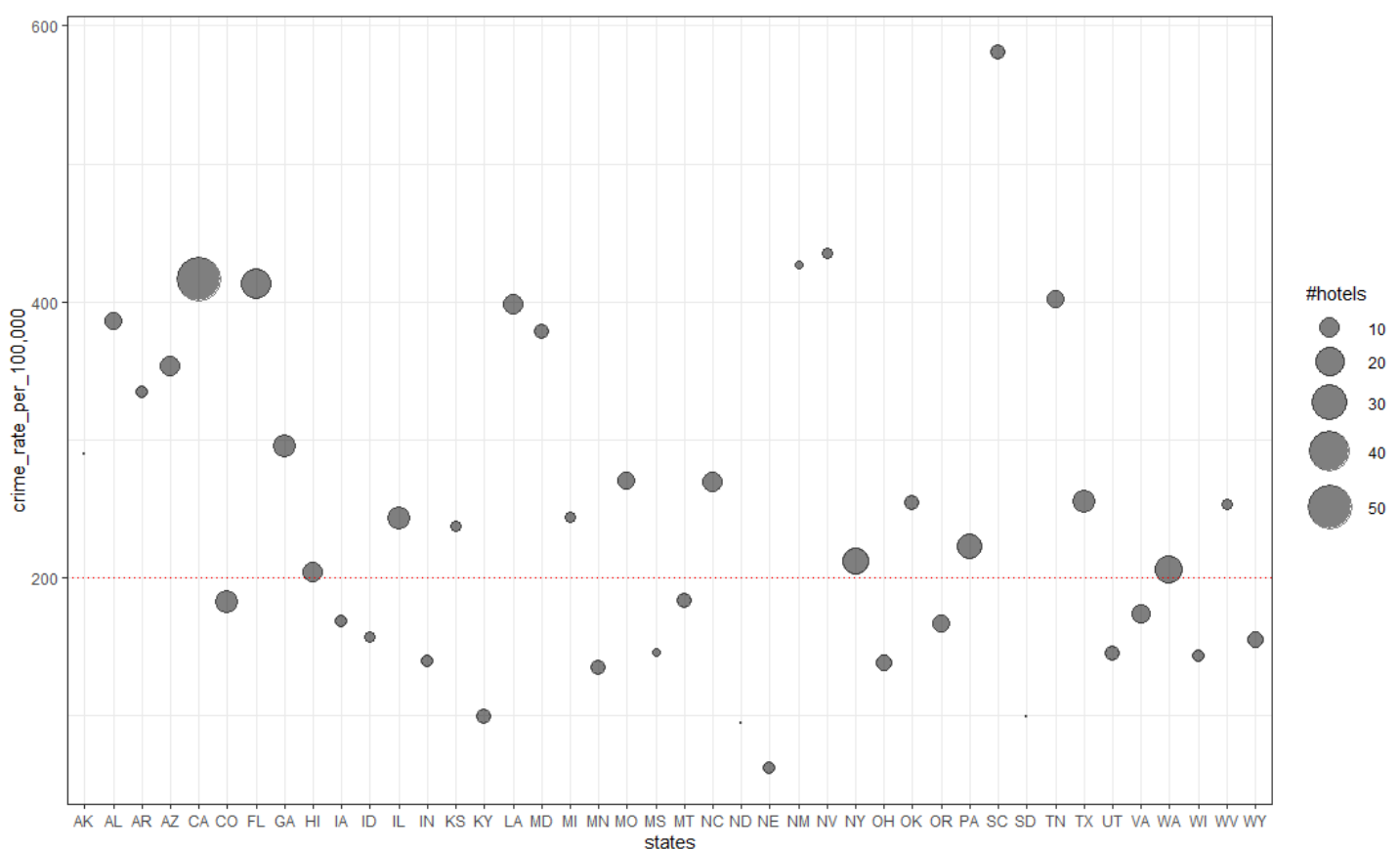


Figure 8: Bubble plot of crime rate and number of hotels in each state.

Figure 8 shows that this is an appropriate choice as there are still a variety of hotels in many states to go to. It can also be seen that generally the higher the crime rates the more hotels there are. This could be due to those states being larger or containing more populated cities resulting in higher crime rates. This is especially true for California (CA) being the third largest state by area and containing the city of Los Angeles that is one of the top 3 most populated cities in the US.

Now that the safest, best rated, and reviewed hotels have been found, we have more freedom around finding a hotel based on personal preferences. The next section will look at the best hotels based on weather as a preference.

Date and weather

I am most comfortable when the temperature is just a bit below room temperature so that it is cool being around 60°F to 65°F (15-18°C). This is so I will not get too hot when I am moving about. Thus, when looking for a hotel, I should also consider the temperatures in the region and if it varies by time that I will be there.

A dataset from the National Centers for Environmental Information will be used for this filtering task and contains the average temperatures measured across 48 states in 2019 each month [5]. Figure 9 below shows a plot of the data for the 17 safest states and it can be seen that only 11 states from April to September satisfy the desired temperatures. These months roughly correspond to the spring and fall seasons, so it will not be a bad idea to prepare hay fever medication and an umbrella in the case of light rain.

This leaves us with only 52 hotels which will be further filtered down to 10 based on another preference which is the number of places to visit around the hotels.

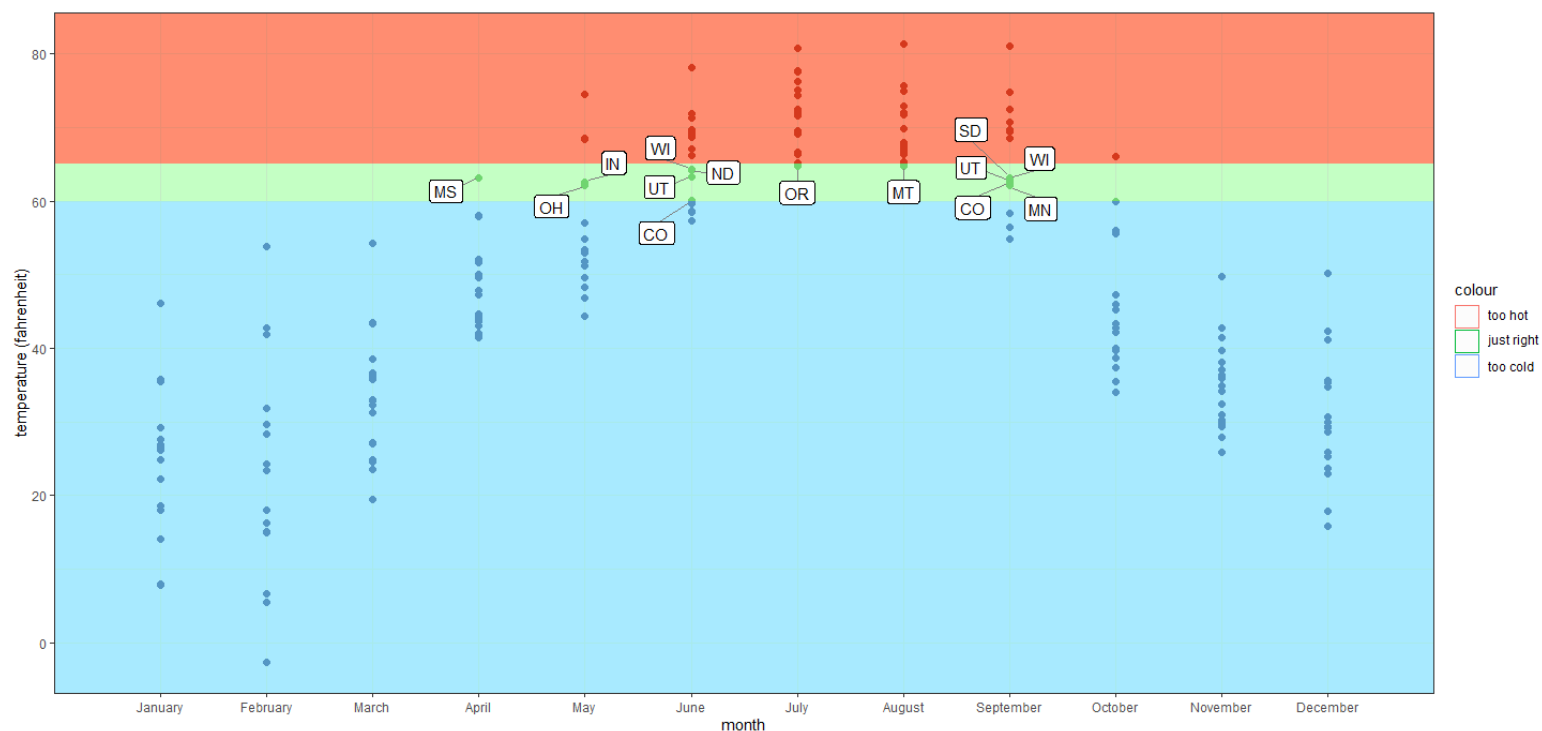


Figure 9: Scatter plot of average monthly temperatures for 17 states.

Places to visit

To recap from the last section, the following map shows the states with the most favorable temperatures. When looking for places to visit, we will only be concerned of locations within those states.

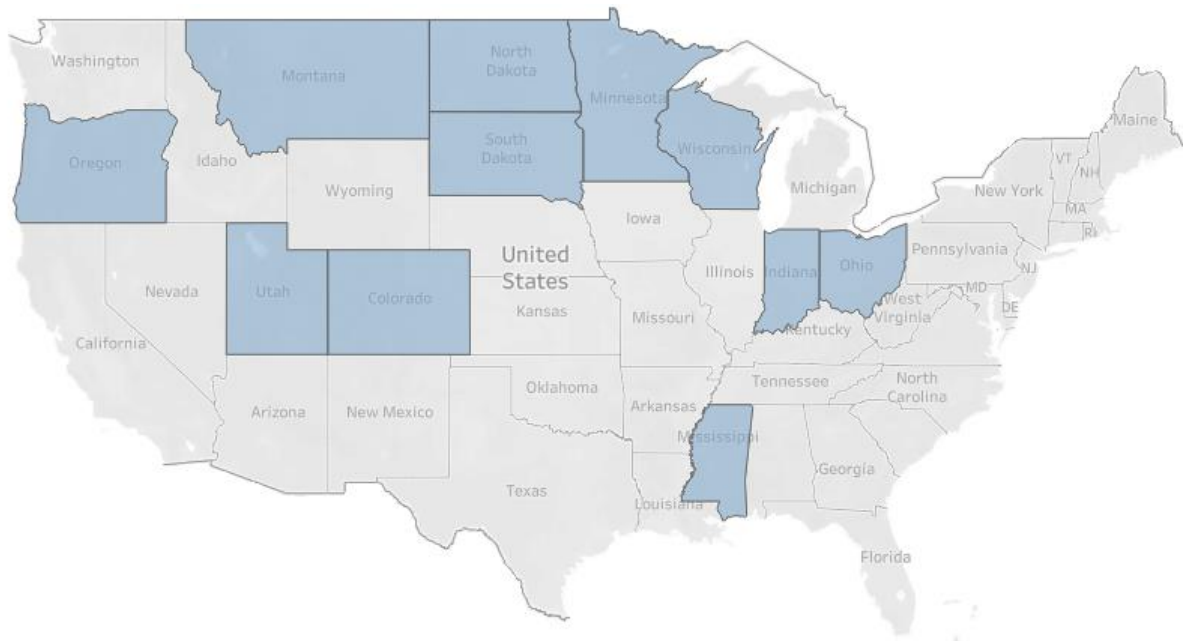


Figure 10: Map of states with favourable temperatures throughout the year.

When going outside to sightsee, I usually think of cities to be places with the most things to do and see. So, I will be seeing which hotels has the most cities near it to explore. The data used will be from Kaggle which contains the location of 769 cities across the USA [6]. Figure 11 below provides an expectation of which states might contain hotels that have the most cities to visit.

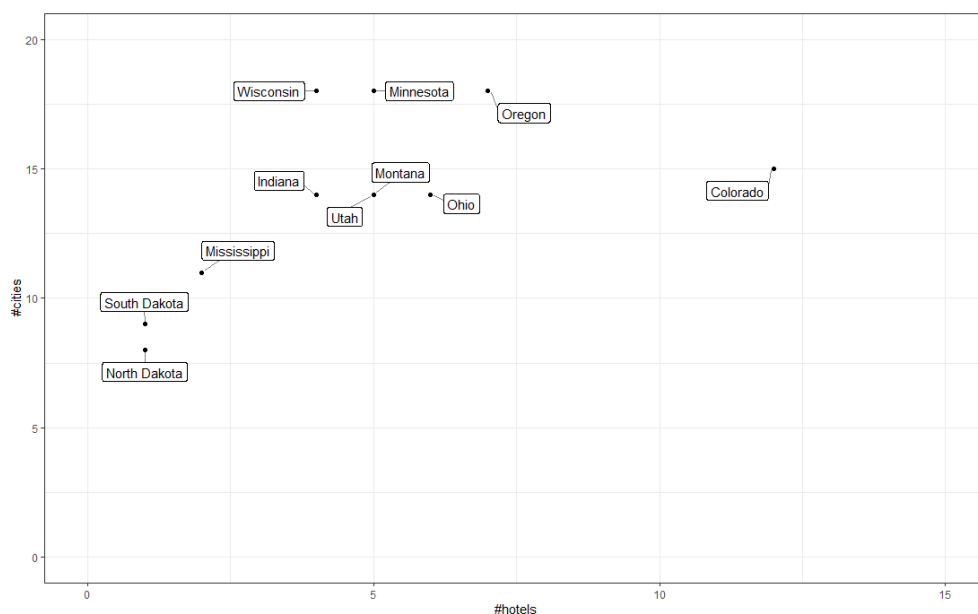


Figure 11: Scatter plot of the number of hotels and cities in each state.

It can be noted that:

- South Dakota, North Dakota and Mississippi all have a small number of hotels and cities
- Indiana, Utah, Wisconsin, Minnesota, Montana, Ohio and Oregon all have a medium number of hotels and larger number of cities.
- Colorado has a large number of hotels and cities.

So, it seems like Colorado hotels might be the best ones to go to.

To exactly find out which hotels have the most cities to visit, the distance between each city and hotel is calculated using the haversine formula, which finds distance between points on globe. This allows us to approximately (ignoring roads etc.) determine which hotels are closest to which city and order them based on how many cities they are closest to compared to other hotels. Figure 12 below shows the result of this and surprisingly a hotel in Wisconsin instead of Colorado took the top spot being the closest to 10 cities. Table 1 below provides the list of the hotels including their address and closest cities.

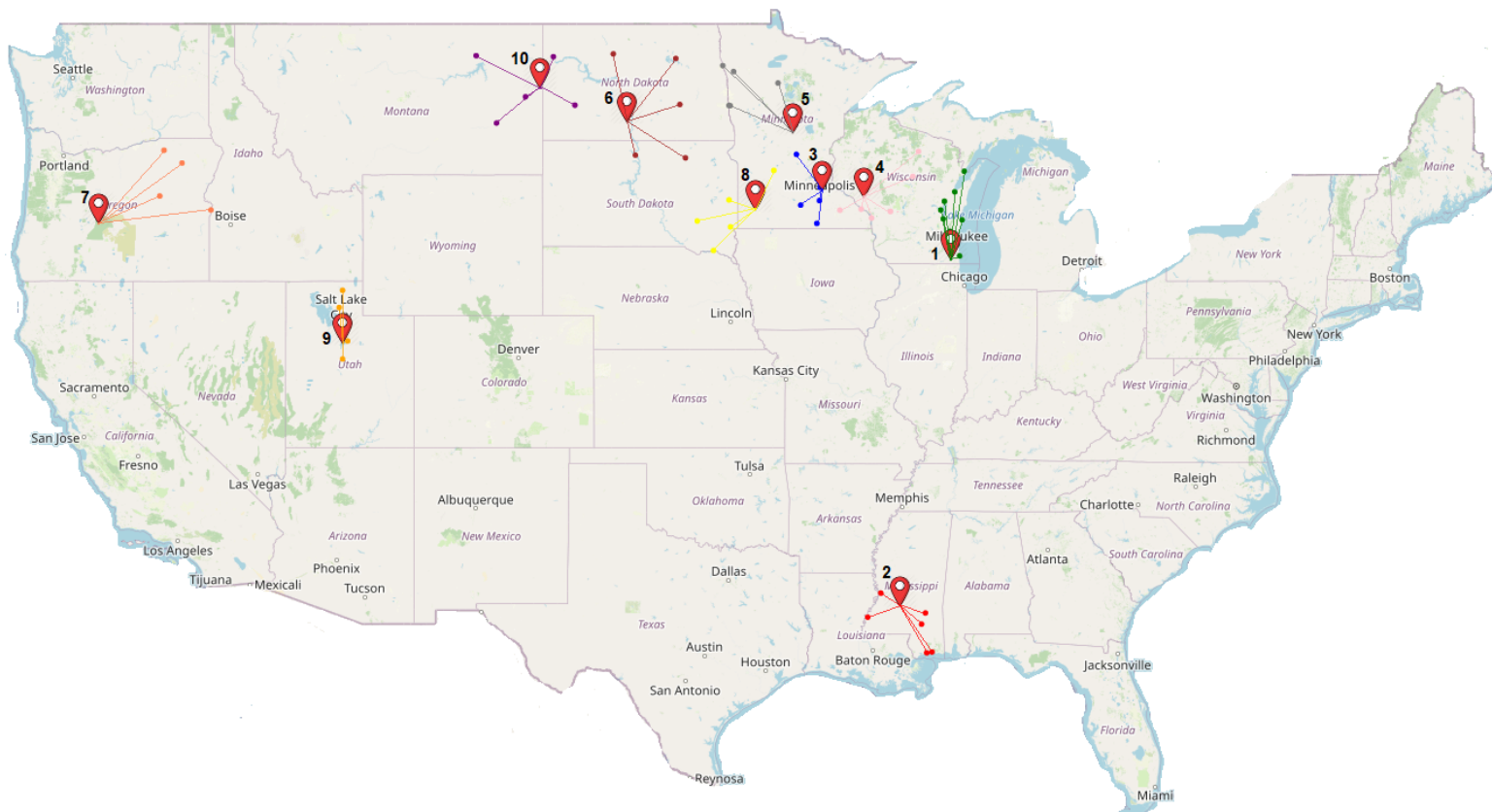


Figure 12: Map of the top hotels connected to their nearest cities. Cities are marked by coloured dot; location marker shows hotel position. (Note that not all cities may not be visible due to scale and being covered by location marker)

Additionally, we can examine the population of the cities (Figure 13) to get an idea of the places we could be visiting for a given hotel. It is seen that most of the top 4 hotels also have some of the largest populated cities to visit so they are good places to go to if you like crowded places. While the other hotels, except for Oxford, contain less populated cities.

Table 1: Top 10 best hotels based on ratings, reviews, crime rate, weather and number of places to visit. (Sorted by number of cities to visit)

#	Hotel Name	Address	Cities to visit
1	Quality Inn	20150 W Bluemound Rd	Marinette, Sheboygan, Racine, Milwaukee, Green Bay, West Bend, Waukesha, Appleton, Fond du Lac, Oshkosh
2	Homewood Suites By Hilton Jackson-ridgeland	853 Centre St	Biloxi, Gulfport, Laurel, Hattiesburg, Jackson, Vicksburg, Natchez
3	Best Western Plus Bandana Square	1010 Bandana Blvd W	St. Paul, Lakeville, Minneapolis, Faribault, Albert Lea, Mankato, St. Cloud
4	Best Western Plus Eau Claire Conference Center	3340 Mondovi Rd	Rhineland, Wausau, Tomah, La Crosse, Eau Claire, Winona, Rochester
5	Tmi Hospitality	5108 Poplar Ave	Brainerd, Bemidji, Crookston, Moorhead, Fargo, Grand Forks
6	Eagle's Nest Resort	3421 N 14th St	Aberdeen, Jamestown, Devils Lake, Mobridge, Bismarck, Minot
7	Sleep Inn Suites-marshall	10 Nw Minnesota Ave	Ontario, La Grande, Pendleton, John Day, Bend
8	Broadway Hotel	1300 Nwakama St	Willmar, Sioux Falls, Brookings, Yankton, Mitchell
9	Oxford Hotel	7555 S Union Park Ave	Provo, Logan, Nephi, Salt Lake City, Ogden
10	Maple Leaf Motel	1000 3rd St Nw	Dickinson, Williston, Glendive, Miles City, Glasgow

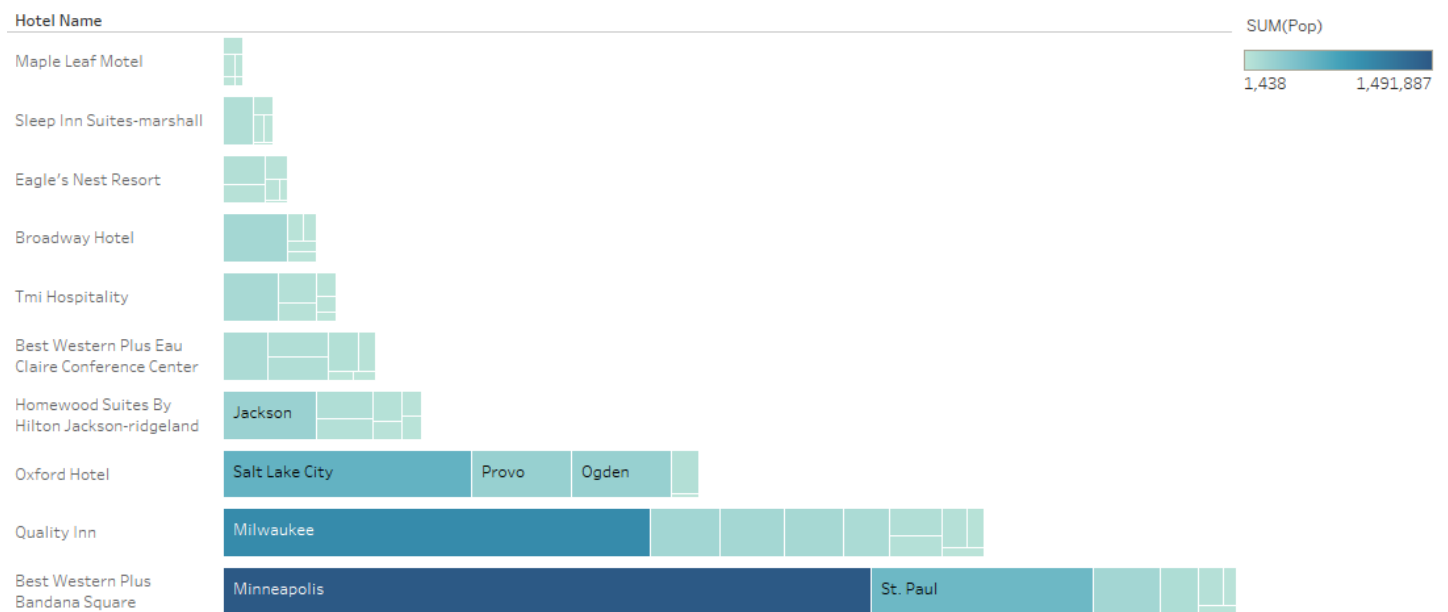


Figure 13: Tree map of city population for each hotel.

Conclusion

First filtering out low rated hotels and those with a small number of reviews, we obtain only hotels with an average rating of 4 and above which contain 2 or more reviews. Further using review sentiment, only hotels with over 70% positive reviews were selected, resulting in the elimination of 1099 unsuitable hotels out of 1433 based on the poor experience from other customers.

Then hotels were removed based on the state that they were in if the crime rate per 100,000 was over 200. Resulting in 82 hotels that were located in relatively safe states. Average temperatures in each state was then used to find hotels located in states that had a comfortable temperature between 60 – 65°F through the year. With 11 states meeting the criteria during the spring and fall seasons. Recommending that travel plans to the hotels to be made around the months April to September.

Lastly the top 10 best hotels were selected based on the previous satisfied criteria's and of their distance to the nearest cities. Resulting in hotels that are highly rated and reviewed, located in relatively safe states, have comfortable weather during certain times of the year and have plenty of places to visit. It is found that the Quality Inn hotel has the most nearest cities to visit and that like the other top 3 hotels, much of the cities are relatively populated. While the cities to visit for the bottom 7 hotels are less populated.

References

1. Definite 2019, Hotel Reviews, Kaggle, viewed 27 September 2020, <<https://www.kaggle.com/datafiniti/hotel-reviews>>
2. Liu, Jiashen 2017, 515k Hotel Reviews Data in Europe, Kaggle, viewed 27 September 2020, <<https://www.kaggle.com/jiashenliu/515k-hotel-reviews-data-in-europe>>
3. Jr, Mike Johnson 2016, United States crime rates by county, Kaggle, viewed 27 September 2020, <<https://www.kaggle.com/mikejohnsonjr/united-states-crime-rates-by-county>>
4. Edwards, R., 2020. '100 Safest Cities In America | Safewise'. *SafeWise*, <<https://www.safewise.com/safest-cities-america/>>
5. Oceanic and Atmospheric Administration (NOAA) 2020, Climate at a Glance: Divisional Mapping, viewed 27 September 2020, <<https://www.ncdc.noaa.gov/cag/divisional/mapping/110/tavg/201912/1/value>>
6. Mahbub 2018, USA Cities, Kaggle, viewed 27 September 2020, <<https://www.kaggle.com/mahbubrob/usa-cities>>

Appendix: Data Sources

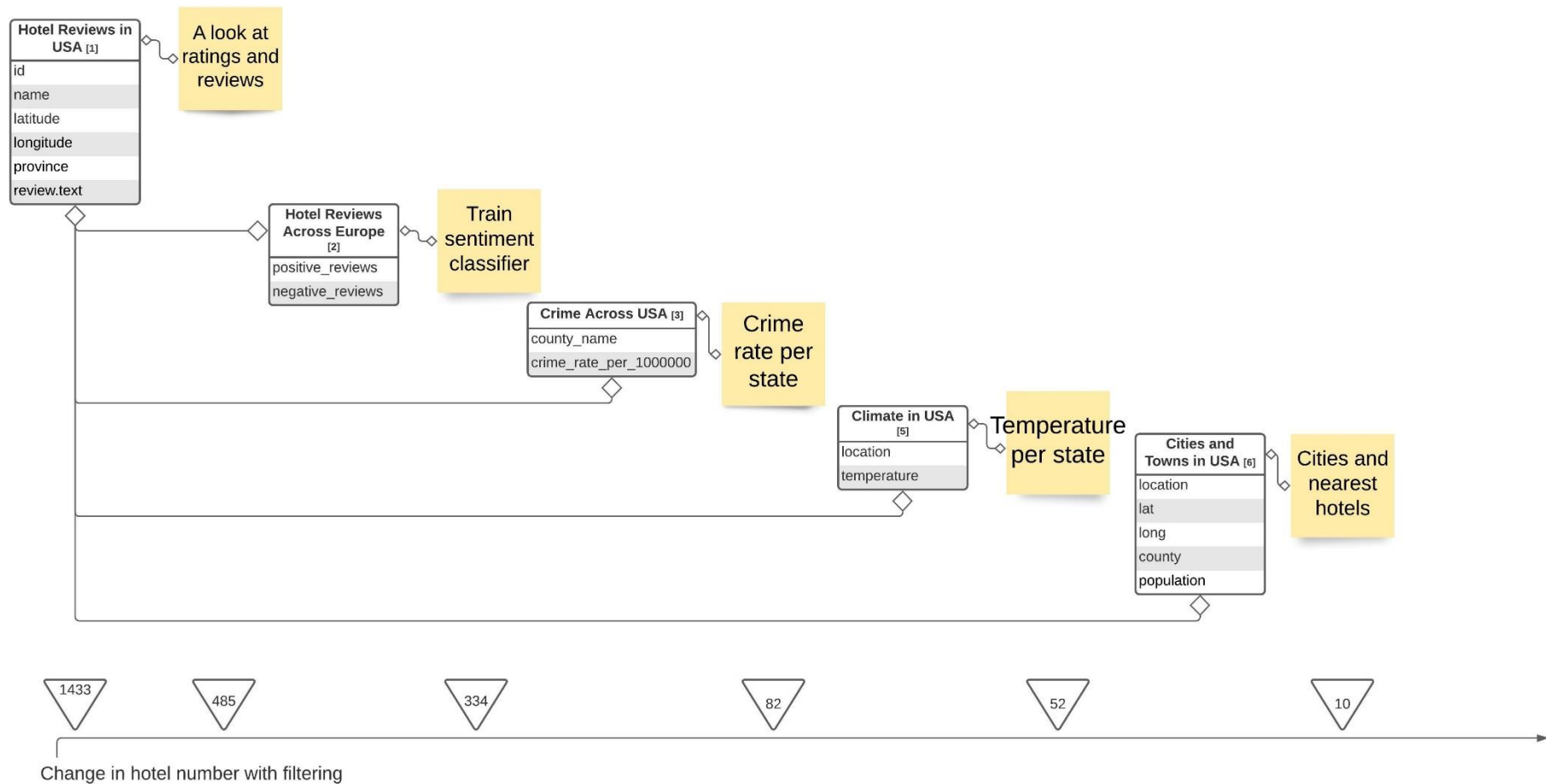


Figure 14: Simplified diagram of the data, variables, connections between data sources and their uses for filtering hotels. Python, R, Tableau and Photoshop are the techniques used in this project for visualizations and editing.