How dangerous are sharks and is all the fear people have justified?

Analyzing shark attacks data from

Why

Most people are afraid of sharks, and there's a misconception that if you are in the water with them you are in deep trouble. The idea that sharks are killing machines preying on humans is a myth that I learned to overcome after I became an avid scuba diver, but I've only had two encounters with sharks so far. I'm often diving in the murky waters of Northern California, which is famous for its large population of White Sharks, so I wanted to learn more about the attacks and real risks.

Dataset

I am using a dataset from Kaggle:
https://www.kaggle.com/teajay/global-shark-attacks

The shark attack incidents data was compiled by the Global Shark Attack File http://www.sharkattackfile.net/. Please see their website for more details on where this data comes from.

Summary of Data

Prior to analyzing the data, I have done some extensive cleaning. The data consists of attacks descriptions, some of them have missing information or with descriptive text fields:

	Date	Year	Type	Country	Area	Location	Activity	Sex	Age	Injury	Fatal (Y/N)	Time	Species
0	25-Jun-2018	2018.0	Boating	USA	California	Oceanside, San Diego County	Paddling	F	57	No injury to occupant, outrigger canoe and pad...	N	18h00	White shark
1	18-Jun-2018	2018.0	Unprovoked	USA	Georgia	St. Simon Island, Glynn County	Standing	F	11	Minor injury to left thigh	N	14h00 -15h00	NaN
2	09-Jun-2018	2018.0	Invalid	USA	Hawaii	Habush, Oahu	Surfing	M	48	Injury to left lower leg from surfboard skeg	N	07h45	NaN
3	08-Jun-2018	2018.0	Unprovoked	AUSTRALIA	New South Wales	Arrawarra Headland	Surfing	M	NaN	Minor injury to lower leg	N	NaN	2 m shark

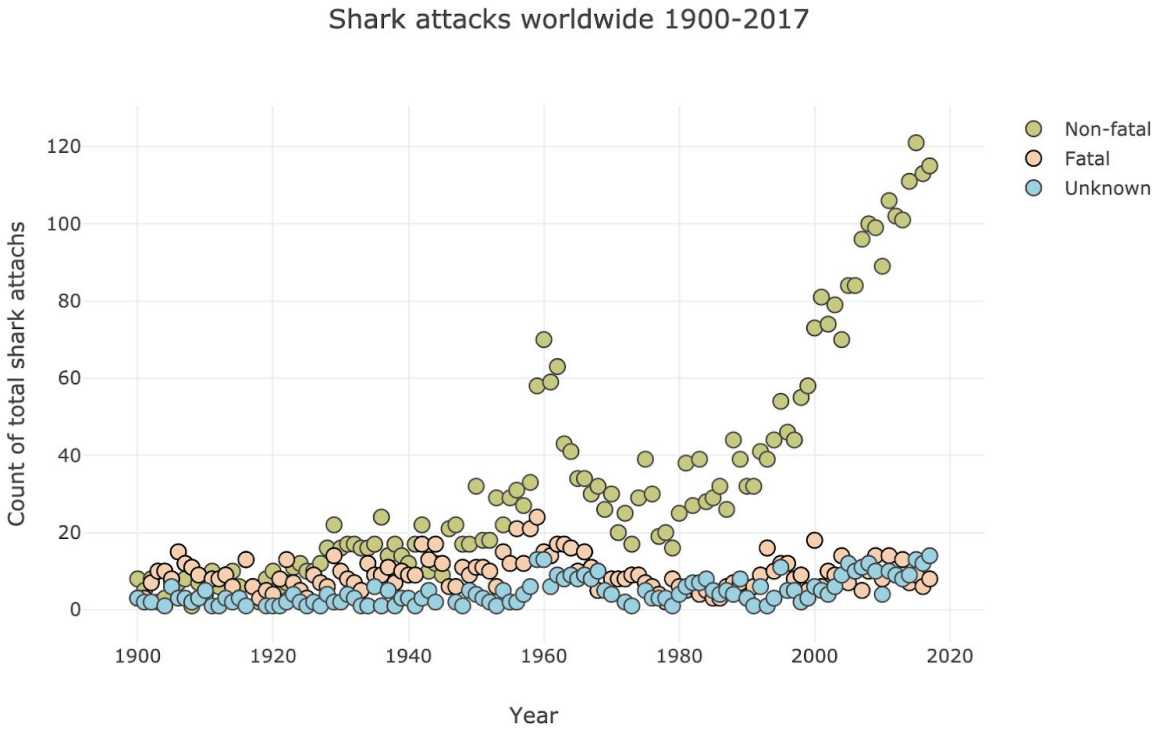
After accounting for missing and zero values, imputation of the data, categorization and some NLP, I added columns for the month of attack and hemisphere information. Additionally, I decided to focus on data between 1900 and 2018, for a couple of reason, some of them is that the population of sharks is on the decrease and as well as people are going more often into the water to enjoy a variety of activities such as surfing, swimming, scuba diving, etc. Moreover, the data set included some of the attacks described in the ancient literature (in years B.C.) as the basis for their recordings.

As part of the requirement for this homework, I was asked to make 10 various plots, which I am going to present below with a short description of the information. Most of the plots were created using interactive visualization tool Plotly in Python. Some of the geographical plots were created in Tableau.

Charts

Scatterplot

Let’s take a look at the general development of shark attack from 1900 till 2017 (2018 data has only half of the year until June 2018 thus not included on this chart).

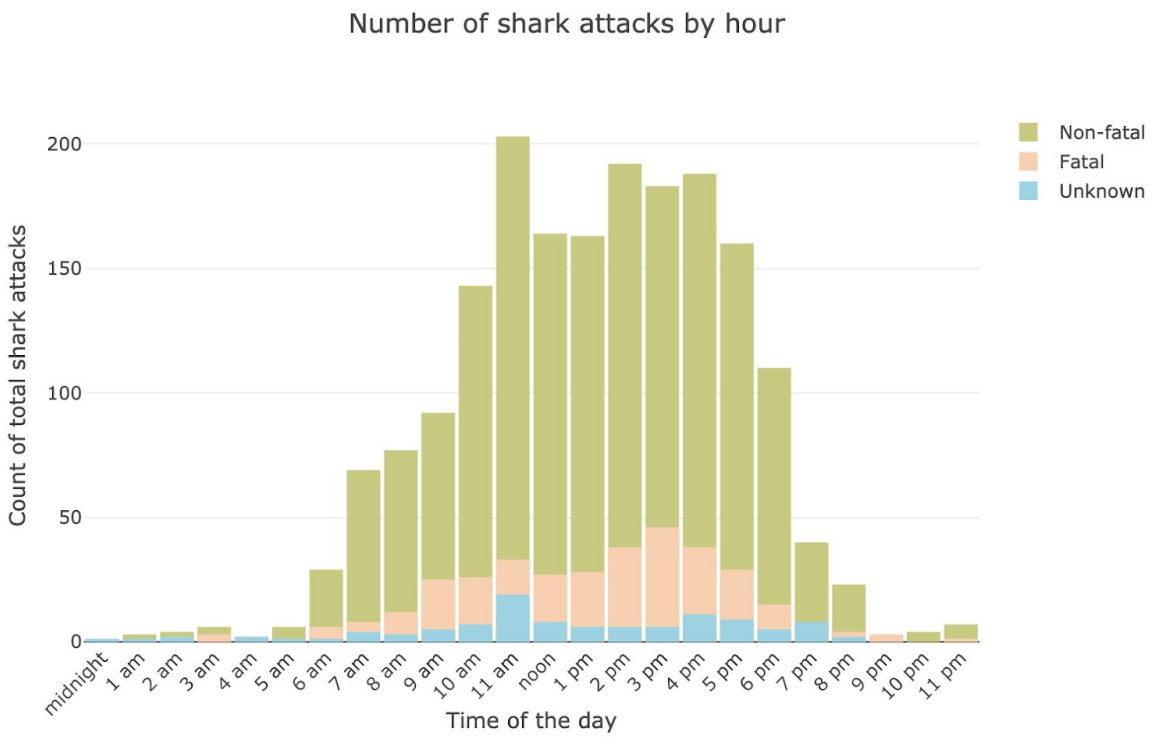


We can see that shark attacks are increasing, though the fatal attacks are staying at the same level throughout the century at about 9.2 average number of fatal shark attacks globally per year. This number doesn't change much by decade either.
If you look at the shark attacks development, especially non-fatal ones, you can see a pick around 1960. The pick is coming from the USA and Australia and is driven by the increase of popularity of surfing in 60's

Disclaimer: all following plots are using the data from 1900 to 2018 unless otherwise commented.

Histogram

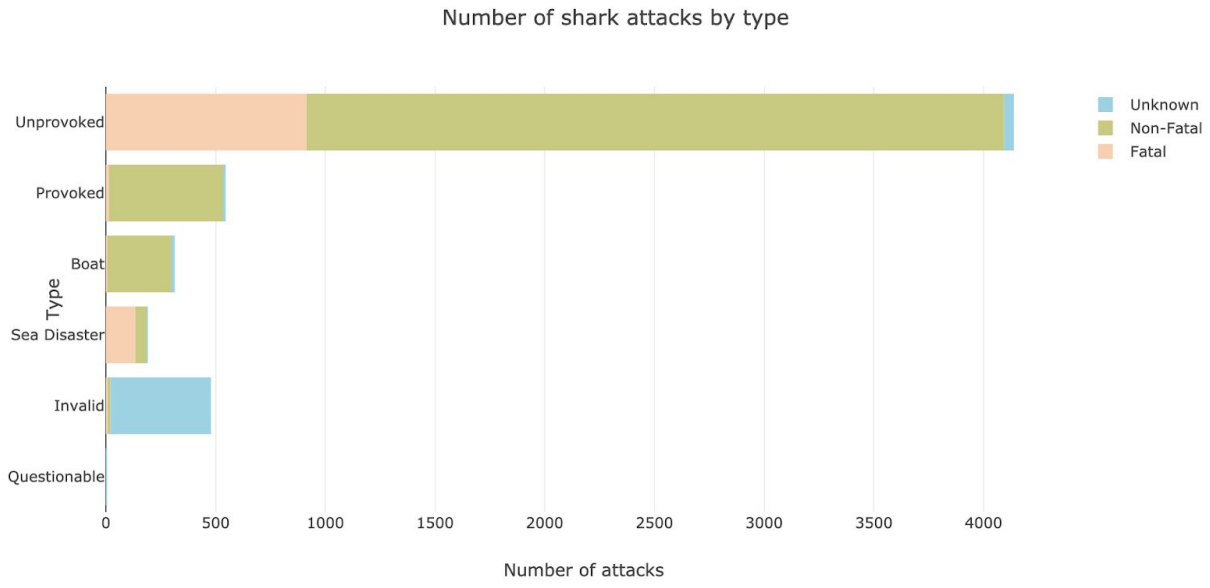
It is commonly known that sharks hunt for prey at dawn and dusk. If humans are on their menu, we would clearly see it on on the histogram plot of shark attacks by the time of the day.



As you can see that is not true. Obviously, humans are not part of sharks diet and the attacks happen when people are enjoying the beach and water time between around 10 am and 5 pm.

Barplot

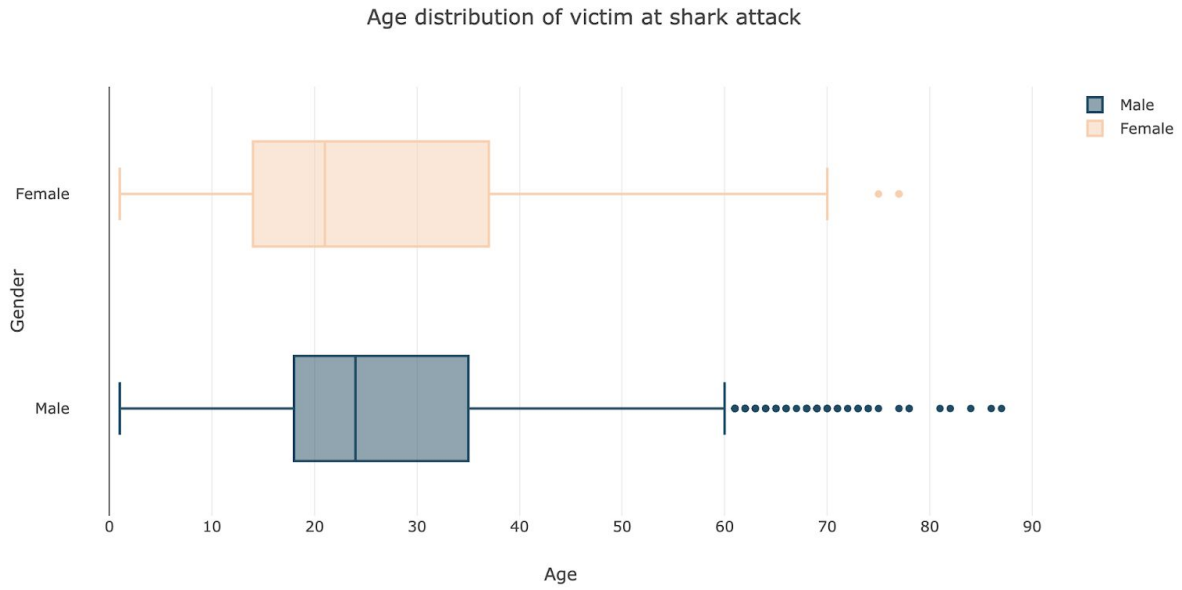
The myth of the shark attack hunting people is scary. Let's look at the data to find out if all attacks are unprovoked or not



From this plot you can see that, most of the time the attacks are unprovoked, though usually, the shark mistakes people as seals surfers on their board waiting for the wave looks for a shark as their favorite prey - seal. You can see though that about half of the attacks are provoked or happening from the Boat, Sea Disaster or are unknown. The boat attacks in the data set are usually when the shark bumps the boat. Shipwreck is usually suspected that the shark was feasting on cadaver after people were drawn.

Boxplot

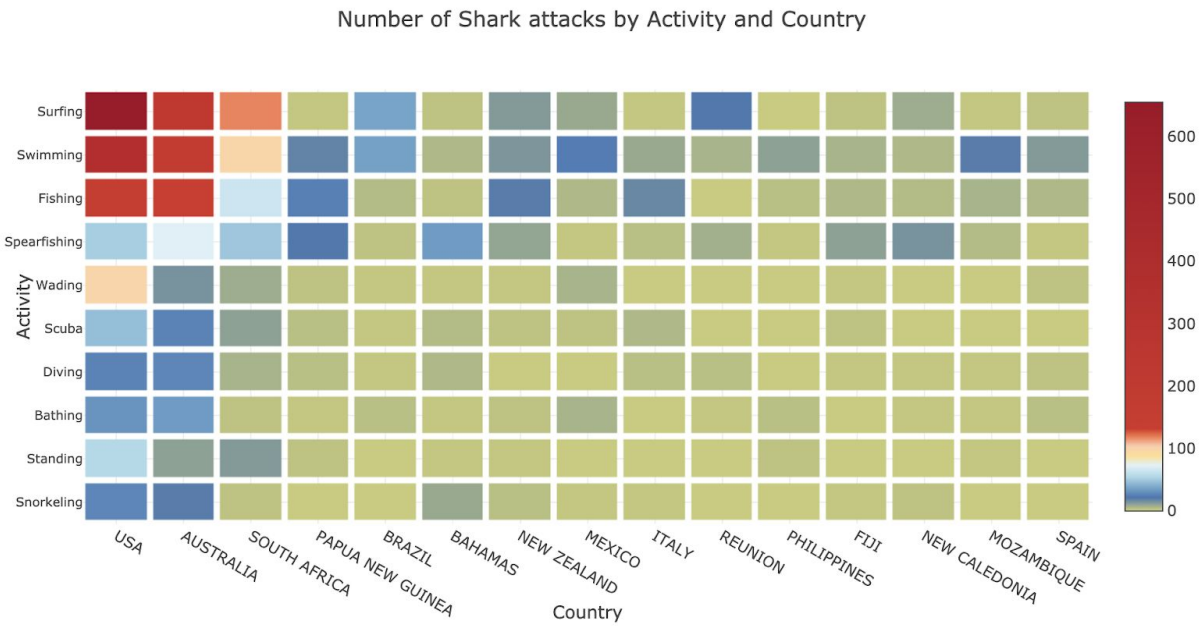
For the distribution plot, I was wondering what is the age distribution of shark attack victims. Seems that female victim age distribution is wider than male one and the medium age is lower for female than male victims.



Overall, I think this chart shows the general pattern of people who are active in water sports at a younger age

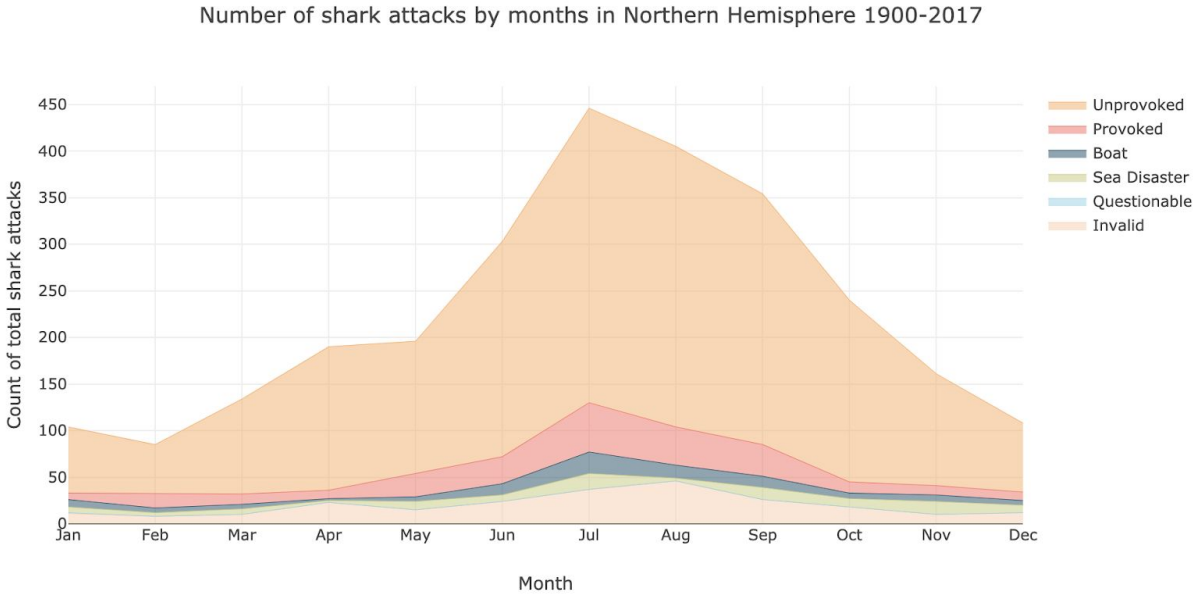
Heat map

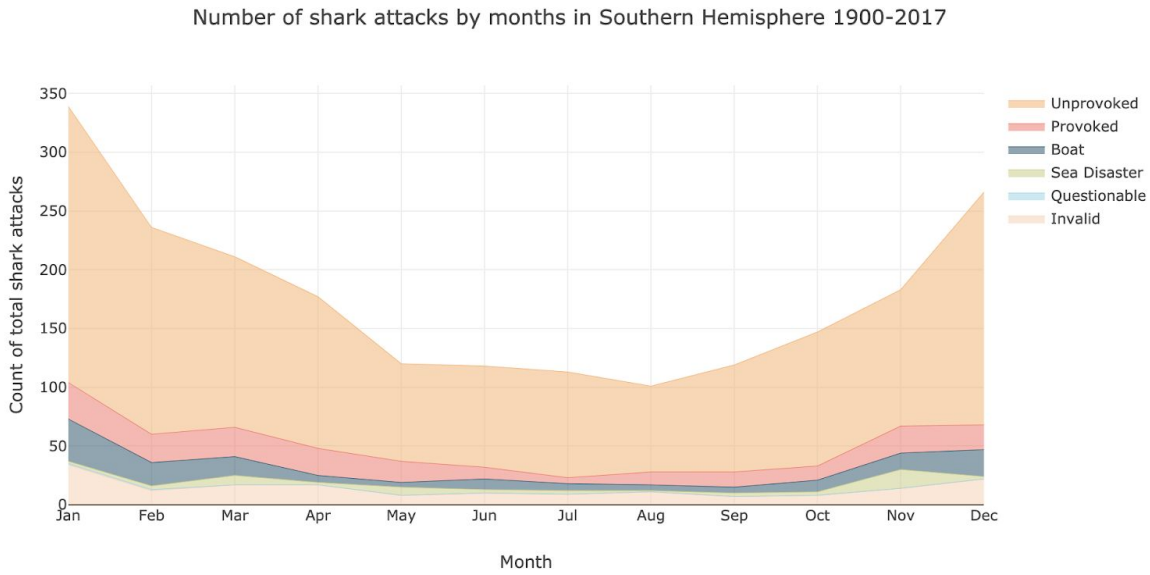
Now as I was guessing about all the different activities people increasing enjoying the water, I wanted to analyze if it is true. For heatmap, I was interested to know what are the most dangerous activity and location to enjoy the water. It seems that Surfing in the USA and Australia is by far the most case where people are attacked by sharks. Again, shark experts think the attacks are happening by mistake where the shark thinks it is preying on seals.
Another interesting fact is that sharks are getting confused and overwhelmed by a present of dead fish, thus another chunk of shark attacks are happening when people spearfish, wade or fish in the sharks' territories.



Stacked area graph

Next, I wanted to know which time of the year most of the attacks are happening and I wanted to see the types of the attacks. For that, I have tagged countries as Northern Hemisphere and Sourthern Hemisphere as the distribution for these two groups would be different.





You can see that most attacks happen when people are on vacation and enjoying summer months in both Northern and Southern hemispheres.

Following graphs were generated in Tableau:

Bubble Map

I wanted to see how the shark attacks look like in the last decade, thus I analyzed 2008-2018. Below on the bubble map, you can see the number of fatal shark attacks, where the size of the bubble is the number of attacks. From this graph, we can clearly see that the attacks are very rare. For example, in the US there were only 9 fatal shark attacks during the last 10 years.

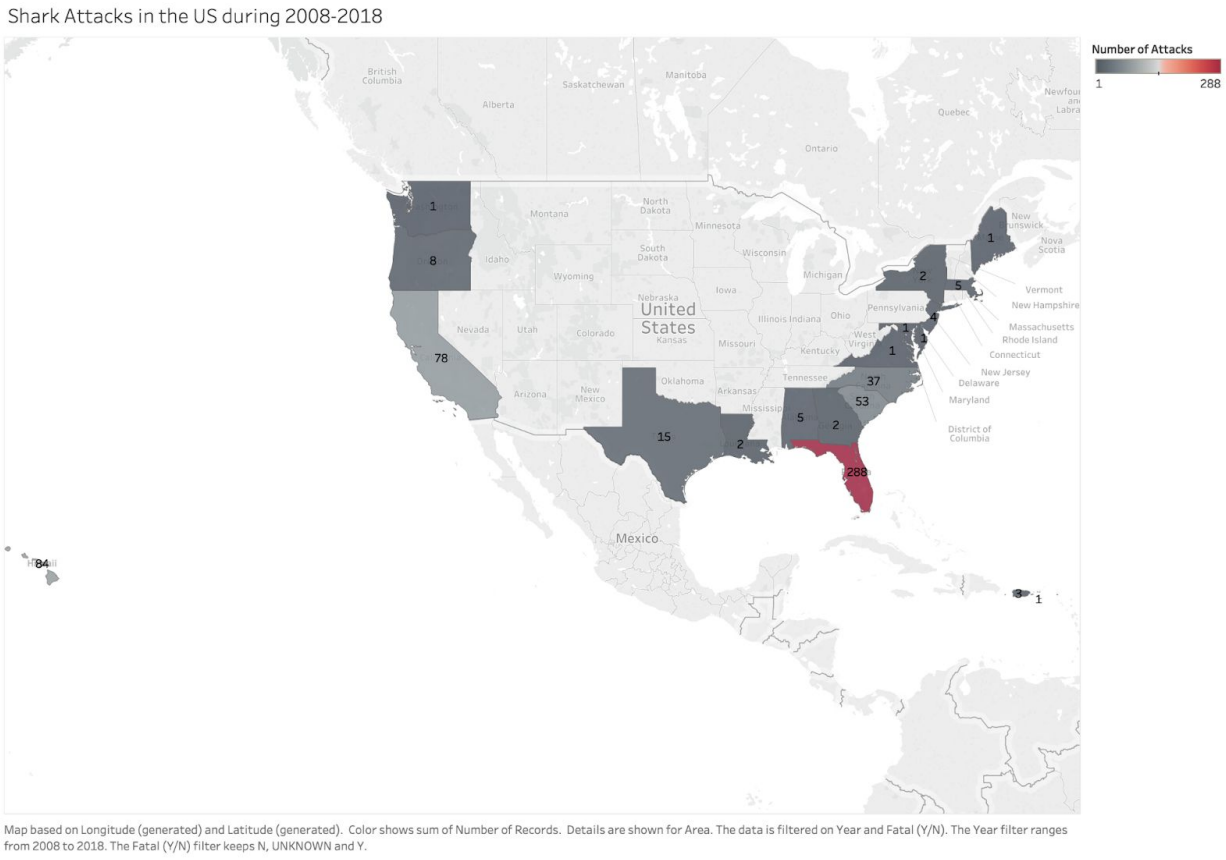
Fatal shark attack in 2008-2018

Map based on Longitude (generated) and Latitude (generated). Size shows sum of Number of Records. Details are shown for Country. The data is filtered on Year and Fatal (Y/N). The Year filter ranges from 2008 to 2018. The Fatal (Y/N) filter keeps Y. The view is filtered on Latitude (generated) and Longitude (generated). The Latitude (generated) filter keeps non-Null values only. The Longitude (generated) filter keeps non-Null values only.

Choropleth Map

On the choropleth, you can see in which state where those attacks happened in the US. I have included all attacks: fatal, non-fatal and with unknown fatality. Again the message is similar, The number of attacks is pretty low. The attacks happen on the coast where people tend to enjoy the

water, i.e. in California and Florida.

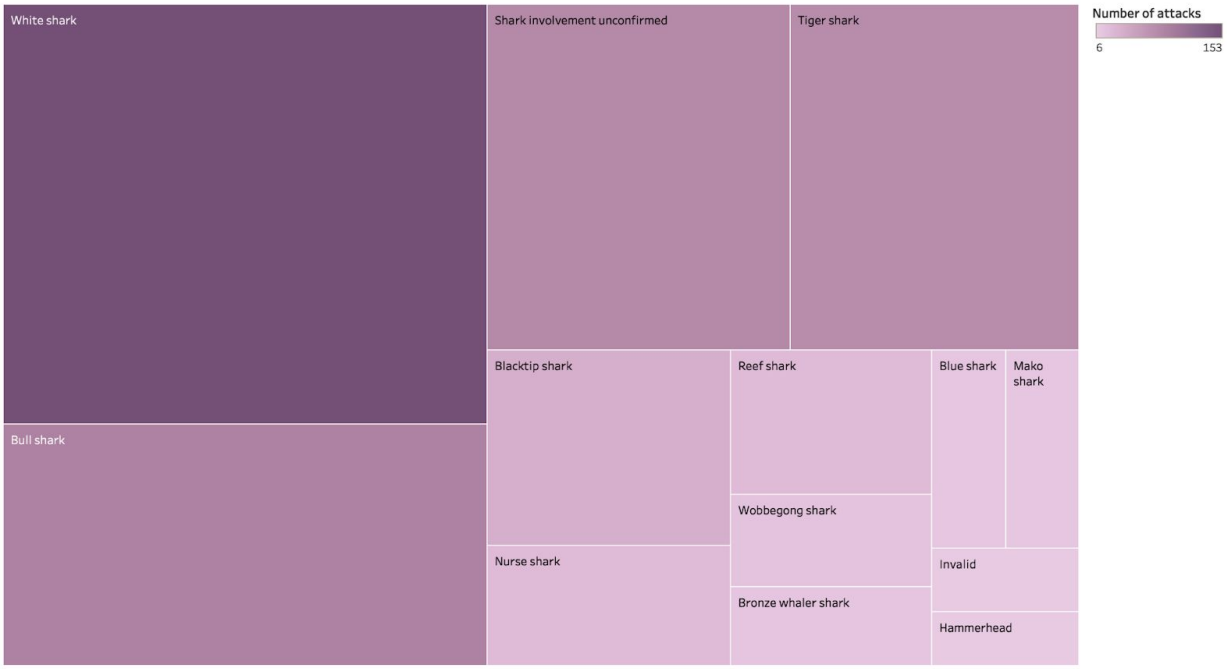


Map based on Longitude (generated) and Latitude (generated). Color shows sum of Number of Records. Details are shown for Area. The data is filtered on Year and Fatal (Y/N). The Year filter ranges from 2008 to 2018. The Fatal (Y/N) filter keeps N, UNKNOWN and Y.

Treemapping

For treemap, I wanted to know the species of the shark attacks. Mostly, it is white sharks, followed by bull sharks. But interestingly in the third place, the shark involvement is unconfirmed. This is mostly the cases where the people drown and sharks ate the cadaver or there was some other animal attacking.

Number of shark attacks by species 2008 - 2018

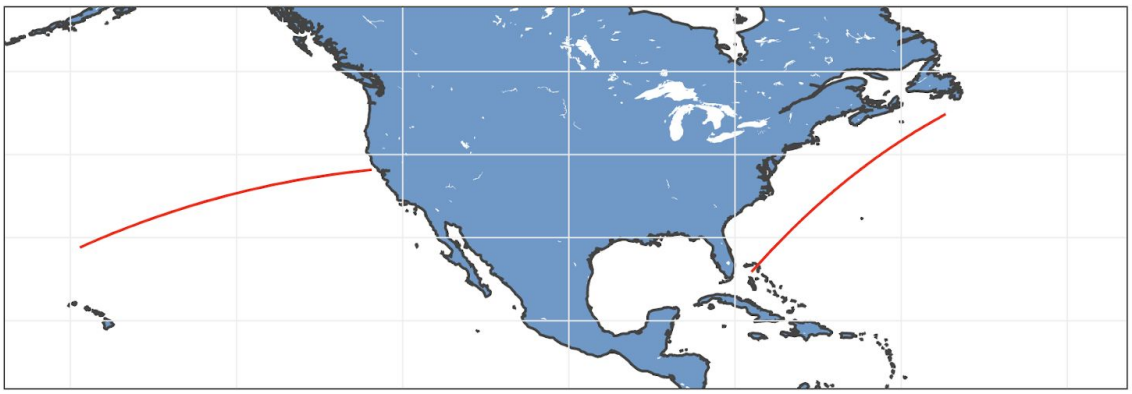


Species. Color shows sum of Number of Records. Size shows sum of Number of Records. The marks are labeled by Species. The data is filtered on Exclusions (Species,Species (group)) and Year. The Exclusions (Species,Species (group)) filter keeps 16 members. The Year filter ranges from 2008 to 2018. The view is filtered on Species, which excludes Null, Questionable and Raggedtooth.

Connection Map

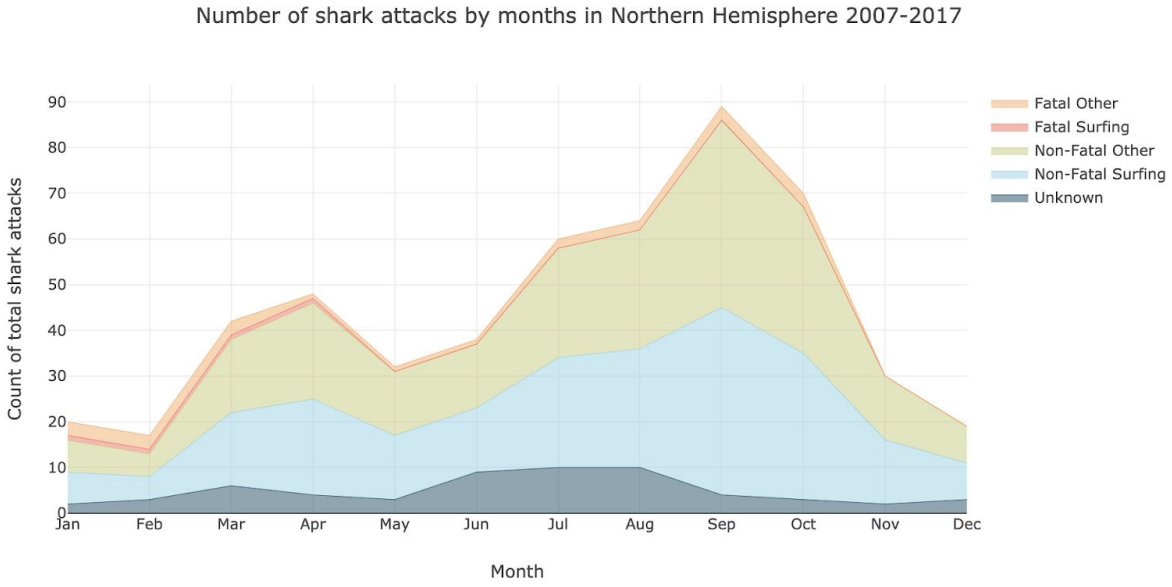
Since my data doesn't really have any foundation for connection map, I decided to plot the approximate major migration routes for sharks in Plotly. Below routes are based on my knowledge and internet research and does not represent any scientific point of view. There is pacific migration, happening from central California coast to other feeding grounds far away in the Pacific Ocean. Another major migration route is along the East coast, where sharks are migrating from the north to the warmer waters of Florida.

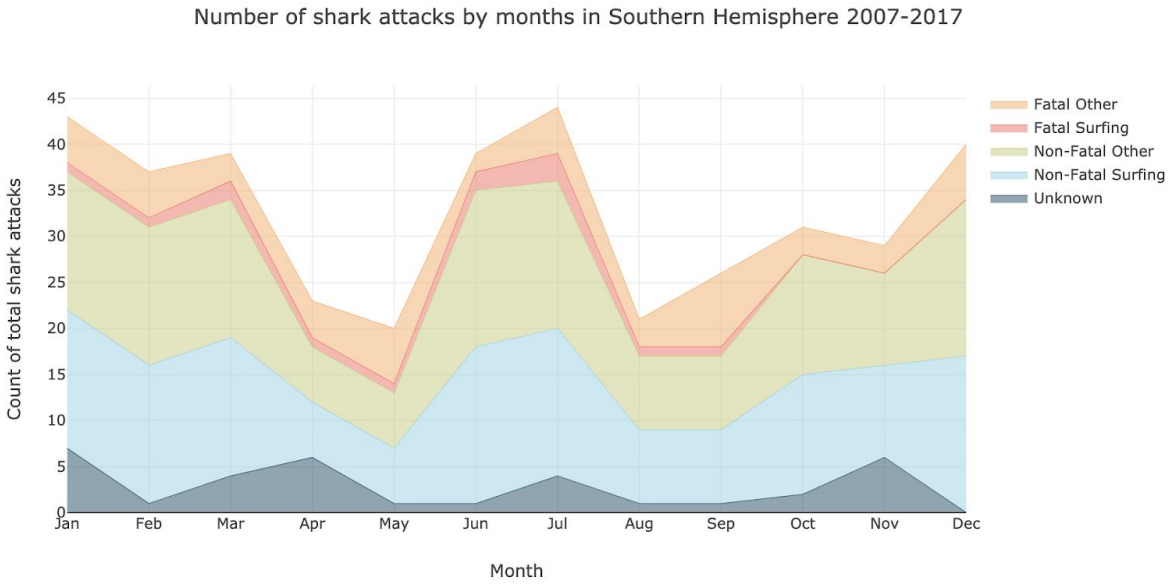
Approximate shark migration routes in Pacific and Atlantic side of the US



Story

While I had many different interesting plots, I would like to focus on my story on these stacked plots which show shark attack by months and fatality and broken down by activity (Surfing and Other) as well as by the Northern or Southern hemisphere. I decided to explore 10 last years.





From these two plots, you can see that generally, fatal shark attacks are very rare. The fatality rate is higher in the Southern Hemisphere, the number of attacks is higher in total for Northern Hemisphere. As you can see the most of the attacks happen while people are surf, where shark mistake surfers for their favorite prey - seals.

Summary

Most of the accidents are motivated is occurred when sharks feel threatened along with the presence of many fishes or murky water. Yet each year, for every human killed by a shark, our species slaughter millions of sharks - about 73 million sharks last year. Sharks are vital to the ocean ecosystem. Without sharks, our planet's ocean would soon become a watery graveyard. I hope I could convince you that sharks are not killer machine waiting for you to enter the waters. Shark attacks are very rare and most of the time happens by mistake. Surfers, swimmers, and fishers in the USA and Australia are the most common case for the shark attacks. But any case as well as non-fatal and not dangerous. Some of the attacks are provoked by human or sharks are feasting on the human cadaver.

Appendix:

https://github.com/katjawittfoth/Data_Viz/blob/master/Project.ipynb

Link to my github:

https://github.com/katjawittfoth/Data_Viz/story.pdf

Sources:

https://www.projectaware.org/sharks
http://www.sharkattackfile.net
http://www.sharkattackdata.com/
https://www.sharkwater.com/