

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

Analyzing shark attacks data from [Global Shark Attack File](https://www.sharkattackfile.net/)

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 shark attack incidents dataset from Kaggle:

<https://www.kaggle.com/teajay/global-shark-attacks>. This data was compiled by the Global Shark Attack File <http://www.sharkattackfile.net/>.

Summary of Data

Prior to analyzing the data, I have done some extensive data cleaning. The data consists of attacks descriptions with missing information or free flow text.

	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

I have applied various techniques for missing and zero values like data imputation. I used categorization and NLP techniques on textual data.

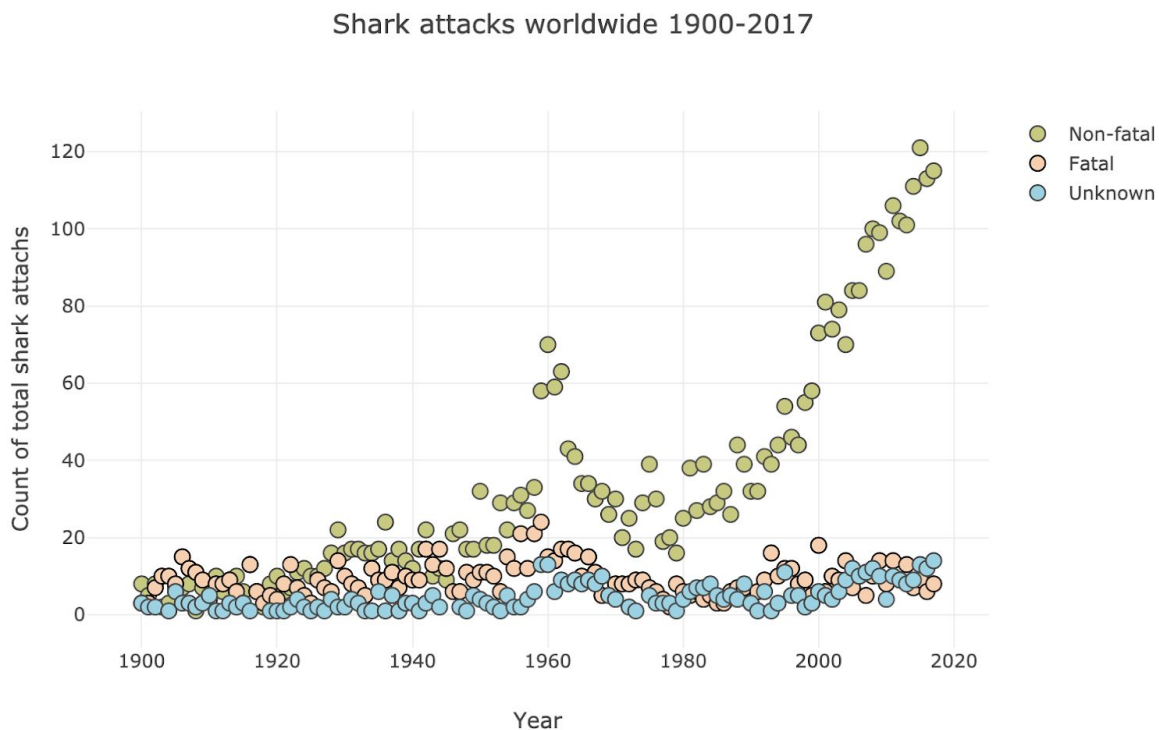
I have decided to focus on data between 1900 and 2018, mostly because the population of sharks is on the decrease and as well as people are going more often into the water to enjoy a variety of water activities. Moreover, the data set included some of the attacks described in the ancient or medieval literature, so I wanted to exclude it.

As part of the requirement for this final project, I am going to present 10 specific plots below and give you a short description. 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 attacks from 1900 to 2017.



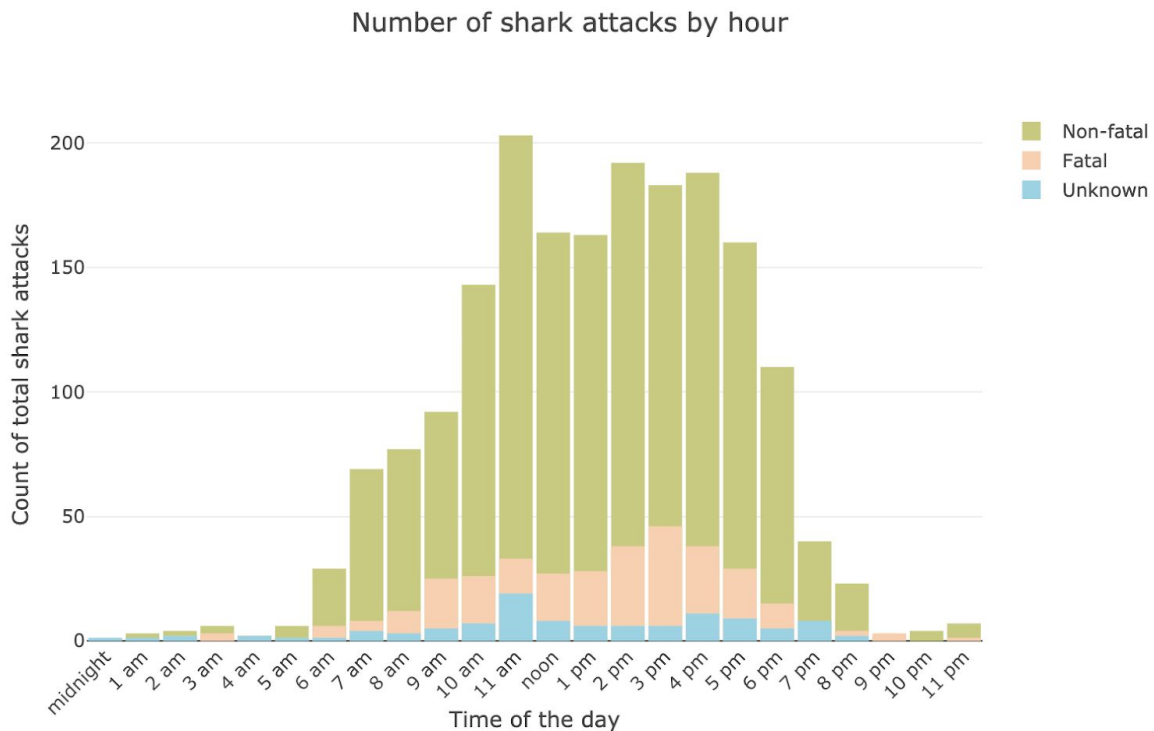
Generally, shark attacks are increasing, though the fatal attacks are staying at the same level throughout the century, at about average of 9.2 number of fatal attacks globally per year. This number doesn't change much by decade either.

If you look at the shark attacks development, especially non-fatal ones, you can see a peak around 1960. The peak happens in the USA and Australia, driven by surfing starting to get popular in the 60s.

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

Histogram

Sharks hunt at dawn and dusk to benefit from darkness by using not only their highly evolved sense of smell but also detecting electricity and vibrations in the water. If sharks prey on humans, we should see a peak of attacks around dusk and dawn.

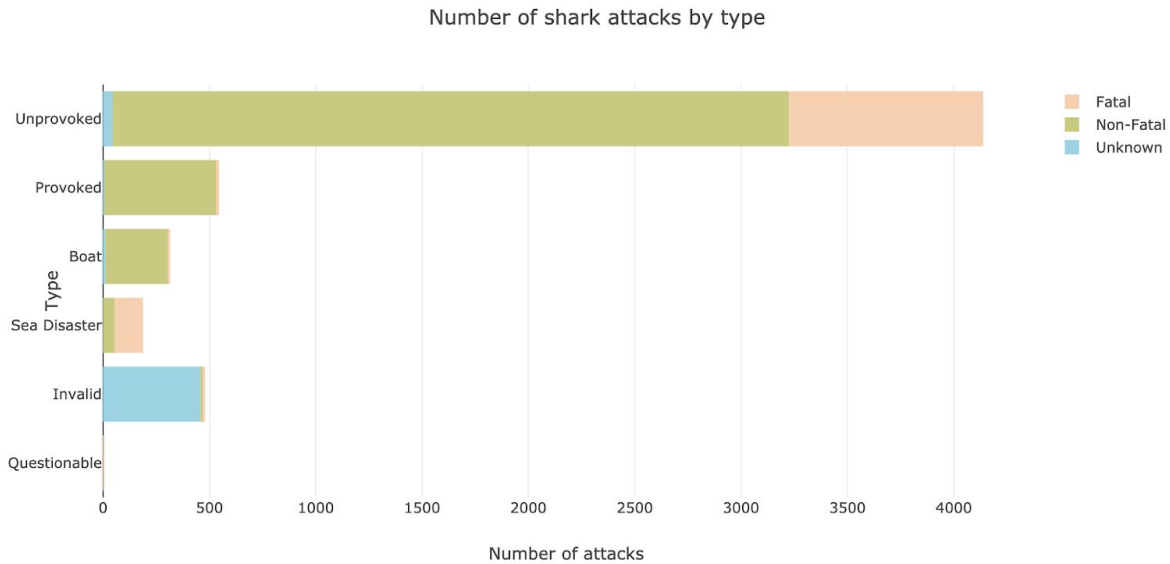


Not true! Obviously, humans are not part of sharks' diet and they are not targetting human, so the attacks happen when more people are in the water, between around 10 am and 5 pm.

Barplot

The myth of the shark attack hunting people is scary. Let's look at the data to find out how sharks attack.

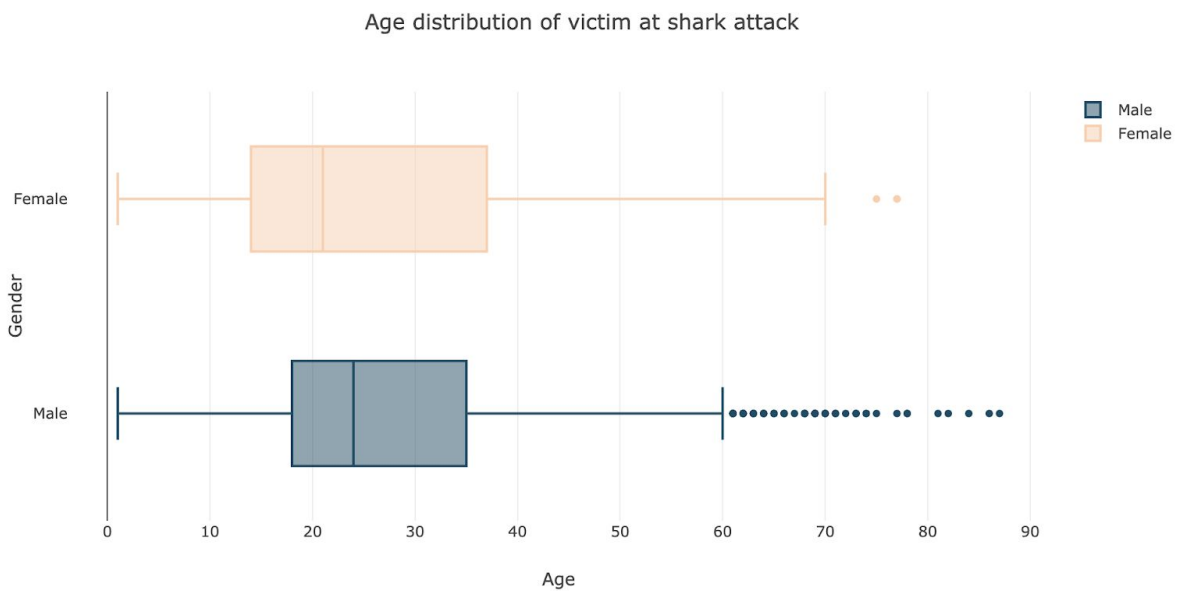
From this plot, you can see that most of the time, the attack is unprovoked. Though, I believe the attacks happen when a shark mistakes people for their prey, as we will see in later charts showing the number of attacks per activity. For example, a surfer sitting on their surfboard waiting for a wave looks and behave like a hurt Seal or Sea Lion from the bottom.



We can see that about $\frac{1}{3}$ of all attacks are either provoked or happening while people are on the Boat or at Sea Disaster or when the type is Invalid/Unknown. The boat attacks in the data set are usually when the shark bumps the boat. Sea Disaster is usually suspected that the shark was feasting on cadaver after people were drawn. This shows us that only in $\frac{2}{3}$ of the cases there is an unprovoked attack.

Boxplot

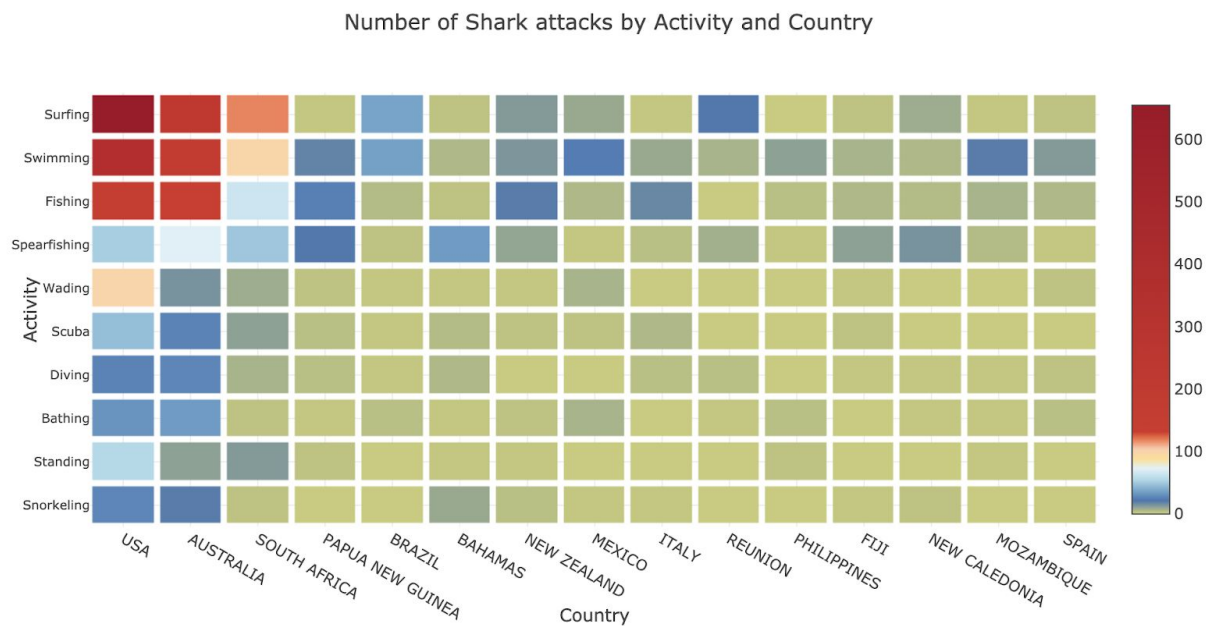
Let's take a look at the age distribution of shark attack victims.



Seems that female victim age distribution is wider than the male one and the medium age is a bit lower. Overall, I think this chart reflects the age distribution of people who are active in water sports earlier in their life.

Heat map

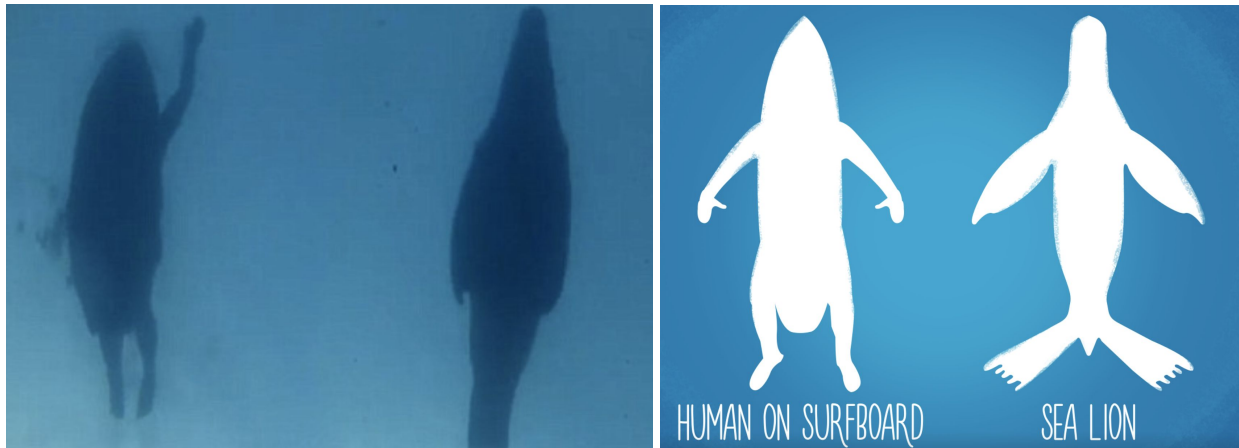
We have seen a peak in the shark attacks during the '60s and I assumed that it strongly correlates with the increasing popularity of surfing. Let's analyze if it is true and what other activities people are engaged in when they are attacked by a shark. I also wanted to see a breakdown by country, thus I analyzed the top 15 countries by the count of shark attacks.



It seems that surfing in the USA and Australia is by far the most dangerous activity.

Sharks have been known to attack humans when they are confused or curious. Sharks often mistake surfers or a human splashing in the water for prey. They get curious and may try to investigate.

Here is the picture of how surfer on the board looks similar to shark's prey:



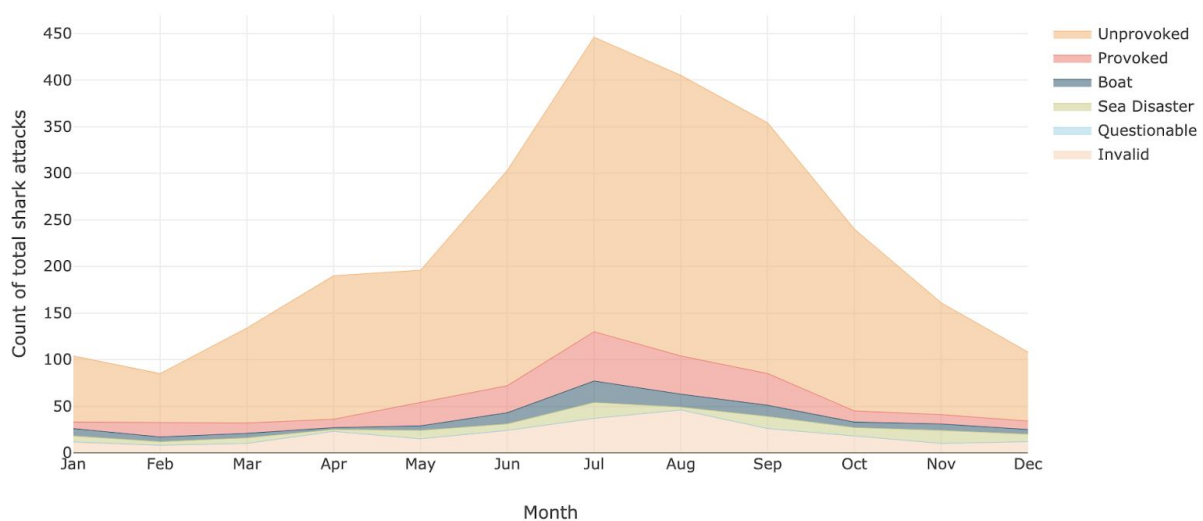
Another interesting fact is that sharks are getting confused and overwhelmed by the presence of dead fish, thus a large number of shark attacks are happening when people are spearfishing, wading or fishing in the sharks' territories. Scuba and snorkeling are quite safe, from a shark attack perspective.

Stacked area graph

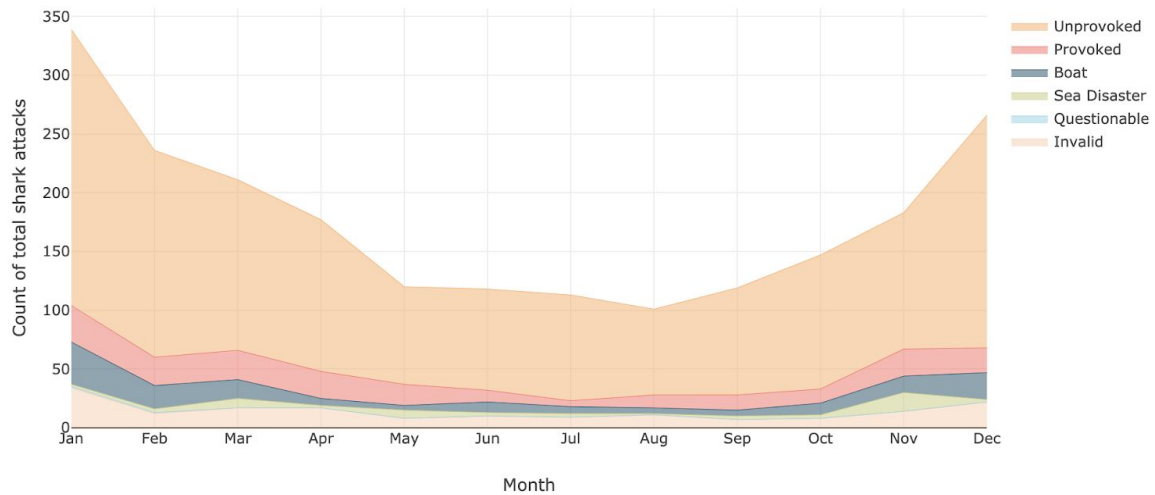
Another common misconception is that shark attacks increase when sharks are migrating or pupping. For example, shark pupping season along SoCal coast is known to be around April and May and shark migration from SoCal to Central and South America around winter time.

I wanted to know if it is true that the attacks are increasing due to migration and pupping, so I grouped attacks by months and separated countries into Northern and Southern Hemisphere as the distribution for these two groups would be different by months.

Number of shark attacks by months in Northern Hemisphere 1900-2017



Number of shark attacks by months in Southern Hemisphere 1900-2017



You can see that most attacks happen when people are on vacation and enjoying summer months in both Northern and Southern hemispheres and has very less to do with shark migration and pupping seasons.

Following graphs were generated in Tableau:

Bubble Map

We know that the shark's population is decreasing, thus the patterns in the attacks might be different now than a couple of decades ago. Therefore, in a couple of following graphs, I focused on the shark attacks during the last decade.

This bubble map shows the number of fatal shark 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

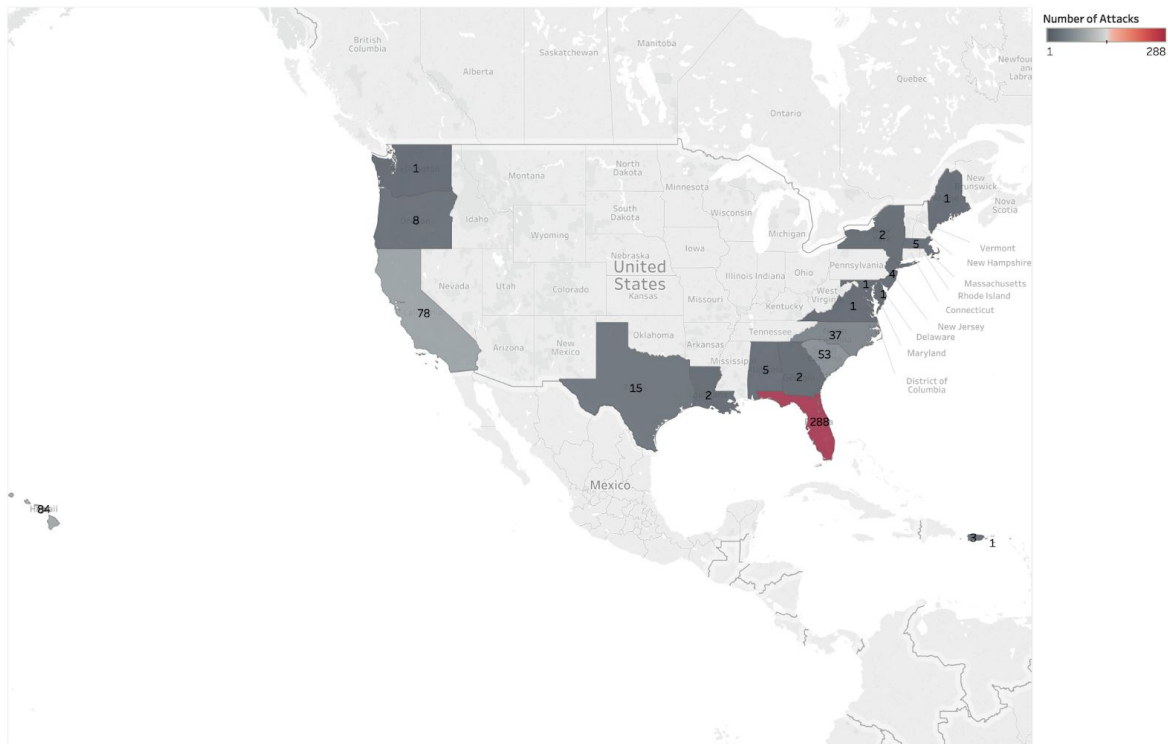


Map based on Longitude (generated) and Latitude (generated). Size shows sum of Number of Records. Details are shown for Country. The data is filtered on Year and Fatal (Y/N). The Year filter ranges from 2008 to 2018. The Fatal (Y/N) filter keeps Y. The view is filtered on Latitude (generated) and Longitude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

Choropleth Map

To dig a bit deeper on the shark attacks in the US during 2008-2018, I plotted state location and the number of attacks on the choropleth map. Here all attacks are included: fatal, non-fatal, and unknown fatality. The number of attacks is pretty low. The attacks happen in places where people tend to enjoy the sea the most: Florida and California.

Shark Attacks in the US during 2008-2018



Treemapping

Let's take a look, which species of sharks have a higher likelihood to attack humans. Most of the attacks don't have identified species; thus they are excluded from this analysis.

Number of shark attacks by species 2008 - 2018



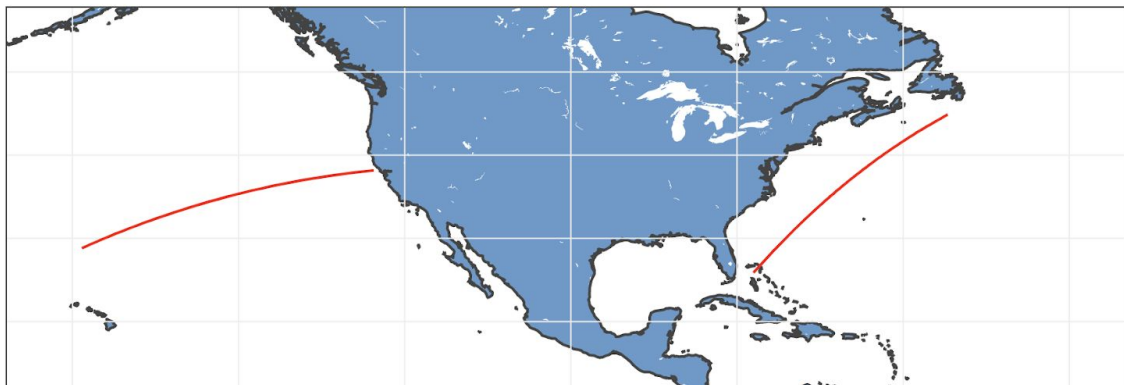
You can see that from global shark attacks during the last decade, the white shark is leading followed by a bull shark. Interestingly, that third place is when shark involvement is unconfirmed. This is mostly the cases where the people drown and sharks ate the cadaver or there was some other animal attacking such as pinniped.

Just to give you a perspective, about 30-50 dies in the US yearly as a consequence of dog attack. Which is about 300-500 deaths in the US only for 2008-2018. This is more than double than the white shark attack worldwide, which includes non-fatal and unknown outcome.

Connection Map

Since my data doesn't really have any basis for a connection map, I decided to plot the major migration routes for sharks. These routes are based on my knowledge and internet research and do not represent any scientific point of view. There is a 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.

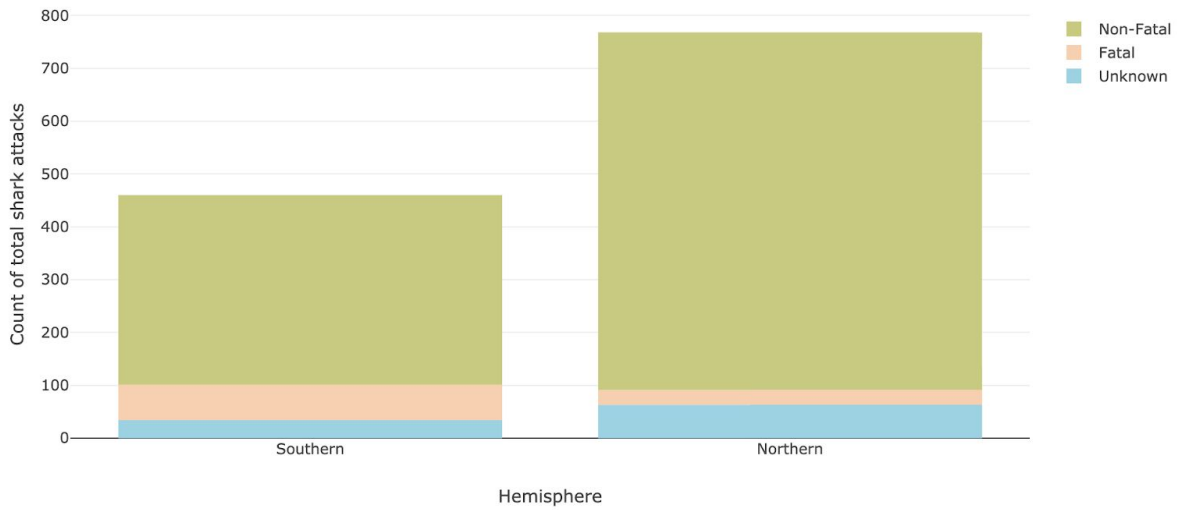
Approximate shark migration routes in Pacific and Atlantic side of the US



Story

While I had various interesting plots, I would like to focus on a stacked area graph, which supports my insight the most. I am showing the shark attack count by month and fatality. The fatality numbers are broken down by activity: surfing and other activity. To account for the difference in seasons, I grouped countries into Northern and Southern hemispheres. For the time period, I decided to explore a period of 10 years. But first, I would like to start with supporting chart, bar plot by hemisphere, which shows that there are far more shark attacks in Northern hemisphere than in Southern, although deadly attacks are twice as likely in the Southern hemisphere than in the Northern.

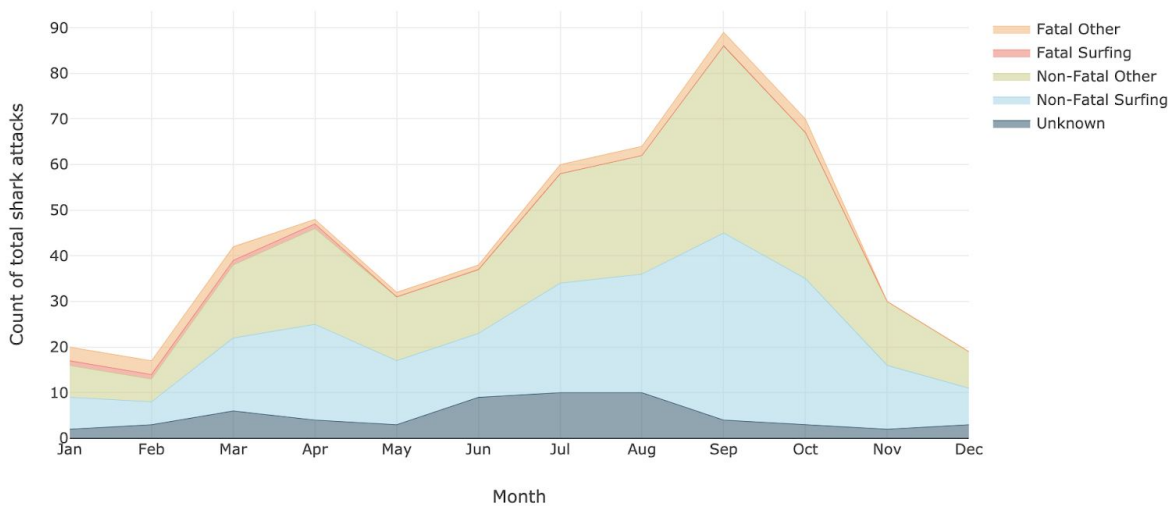
Shark attacks by hemisphere 2007-2017



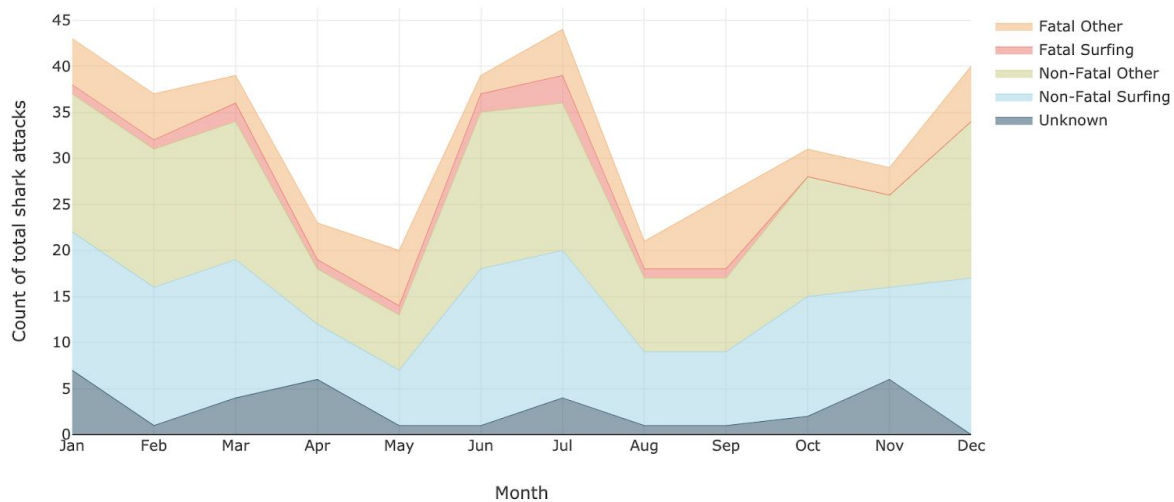
From the two stacked plots below, you can see that generally, fatal shark attacks are very rare, especially in the Northern Hemisphere, although the overall there are more attacks in the Northern Hemisphere.

The fatality rate is higher in the Southern Hemisphere, the number of attacks is higher in total for Northern Hemisphere. Nearly half of the attacks happen while people are surfing, the situation where the shark mistake surfers for their favorite prey - seals.

Number of shark attacks by months in Northern Hemisphere 2007-2017



Number of shark attacks by months in Southern Hemisphere 2007-2017



Summary

While a shark attack is a frightening event, we need to understand that it is very rare and happens by mistake. Most of the shark accidents occur when the shark feels threatened or confused by either the presence of many dead fishes, murky water or mistakes people for food.

Surfers, swimmers, and fishers in the USA and Australia are the most common case for the shark attacks. Though in 84.4 % the attack is not fatal.

Many of the attacks are very minor cases recorded such as when shark bumps the boat. There are also many cases where shark attack is not confirmed, either the attack is by other animal or the shark attack happened post-mortem in case of drowning.

I hope this data and my visualizations could convince you that sharks are not the killer machines waiting for you to enter the waters, but just curious animals which need to be treated with respect.

Code and story on my GitHub:

https://github.com/katjawittfoth/Data_Viz/blob/master/Notebooks/Shark%20Attacks.ipynb

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

Sources:

<https://www.projectaware.org/sharks>

<https://www.sharkwater.com/>

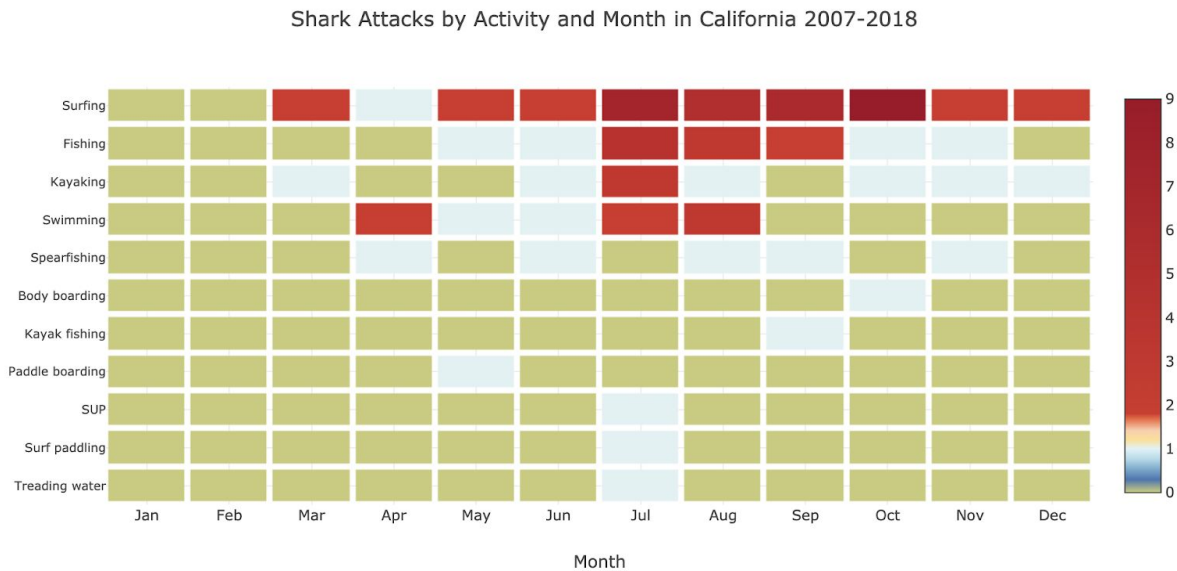
https://en.m.wikipedia.org/wiki/Fatal_dog_attacks_in_the_United_States

<https://www.dailymail.co.uk/news/Researchers-examine-sharks-mistake-surfers-intended-prey>

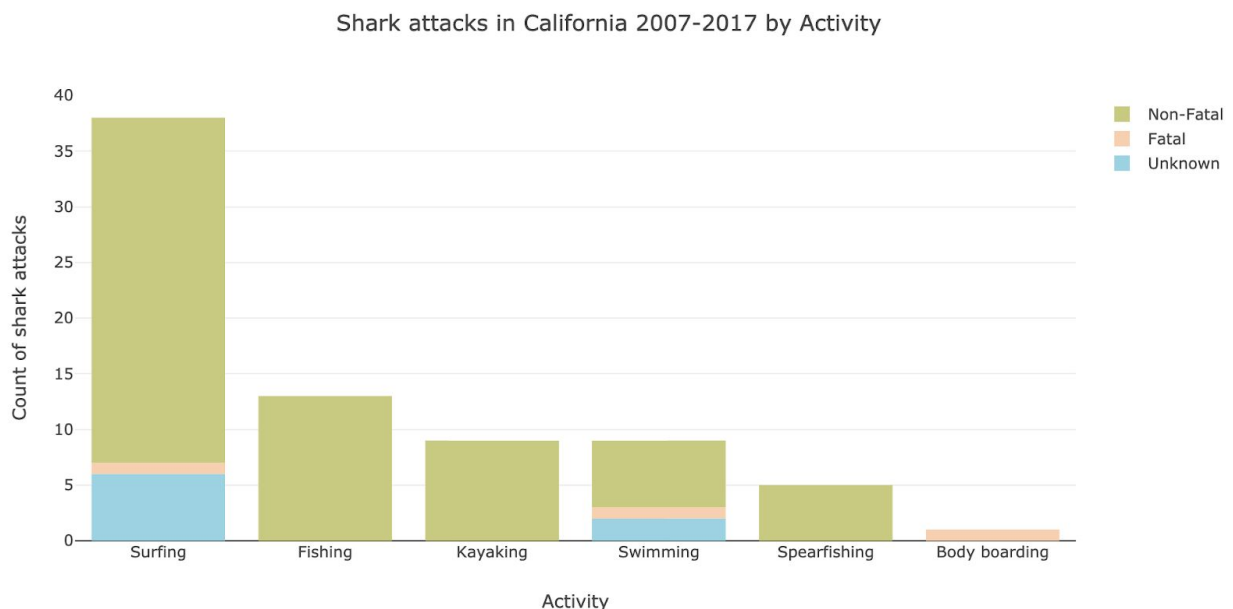
Bonus: California shark attacks 2007 - 2017

As a scuba diver in California, I wanted to know what is the state statistics of shark attacks for the last 10 years.

There are a total of 86 attacks during 2007 -2017 and there were no attacks on scuba divers recorded. Most attacks involve surfers, fishers, and kayakers.



Let's investigate the fatality and injury.



You can see that there is only 3 fatal case in California for the last 10 years.

Many attacks have no injury at all, this is where shark only bumps a kayak or a surfboard. If I exclude those 'no injury' attacks we can see that only 48.8% of all 86 shark attacks in CA in the past 10 years led to injury and only 3.4% of them were deadly.