

Process Book

Willemijn Barens - Data Processing 2022

Sidenote: Several pages from Stackoverflow were used every time I worked on the project. However, I didn't think it was necessary to list all these pages.

Date of entry: 6-12-2022

What I've worked on: Setting up my research question, cleaning my data, and learning the first steps of the Bokeh library.

What problems I encountered: I had some problems with data that was missing and how to deal with that in my DataFrame. I decided to change all the empty values into 0.

What I learned: To make my first Bokeh line plot, to add tooltips to my plot.

Which resources did I use: https://docs.bokeh.org/en/latest/docs/first_steps.html, https://www.youtube.com/watch?v=2TR_6VaVSOs

Date of entry: 8-12-2022

What I've worked on: Trying out different things for the map and finding a dataset with location coordinates. I'm aiming to merge this dataset with my shark attack data.

What problems I encountered: To use Gmap from Bokeh I need a Google API key which costs money.

What I learned: Different ways to create a map with Python, one of which is the Folium package. I decided to create my map with that library instead.

Which resources did I use: https://www.youtube.com/watch?v=2TR_6VaVSOs, <https://python-visualization.github.io/folium/quickstart.html#Getting-Started>, <https://www.kaggle.com/datasets/juanmah/world-cities>

Date of entry: 9-12-2022

What I've worked on: finishing up merging the location coordinates data with the shark attack data. Creating my first world map using Folium.

What problems I encountered: How to properly display the map and how to make the circles bigger based on fatal sharks attacks in that area. For the shark attacks that didn't include area data, I decided to join the datasets on country instead, so that these attacks wouldn't be excluded from the map. This caused for some extra work in calculating fatalities per area/country and representing this in the map.

What I learned: Learning about the different markers and possibilities within Folium. And how to use HTML within these markers and tooltips.

Which resources did I use: <https://python-visualization.github.io/folium/quickstart.html#GeoJSON/TopoJSON-Overlays>

Date of entry: 13-12-2022

What I've worked on: Creating a plot which displays the amount of shark attacks and fatalities per year.

What problems I encountered: Figuring out which plot from the Bokeh library would be the best to use.

What I learned: How to create multiple plots in one file and how to properly display them.

Which resources did I use: https://docs.bokeh.org/en/latest/docs/first_steps/first_steps_1.html, <https://docs.bokeh.org/en/latest/docs/gallery.html>

Date of entry: 16-12-2022

What I've worked on: Updating my Folium map. Instead of having a list of all activities people were doing while getting attacked by a shark I only included the top 5 most common activities.

What problems I encountered: How to deal with areas that didn't have enough data to form a top 5.

What I learned: How to use html in popups.

Which resources did I use: <https://towardsdatascience.com/folium-map-how-to-create-a-table-style-pop-up-with-html-code-76903706b88a>

Date of entry: 18-12-2022

What I've worked on: Finishing up with the last few changes and proof reading my code.

What problems I encountered: None

What I learned: Not anything new

Which resources did I use: None