Cahier d'avancement Projet plastique et biodiversité

Jonathan Pattin Cottet Eyosyas Abebe Simon Verdu Florian Gaspard

17/11/2020 : Choix d'un sujet lié à l'environnement

Nous avons choisi de sensibiliser les consommateurs dans leurs achats d'objets plastiques non réutilisables et de montrer l'impact qu'ils peuvent avoir sur la faune et la flore marine.

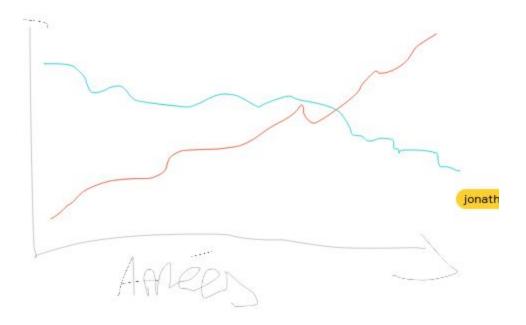


20/11/2020 : Brainstorm premières esquisses

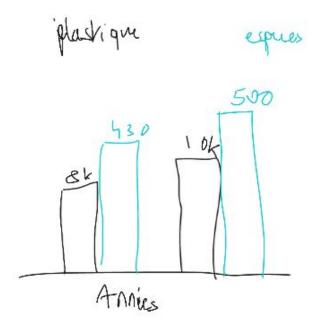
Nous avons créé 10 esquisses dans un premier brainstorm pour pouvoir réunir nos idées et voter pour la meilleure visualisation.



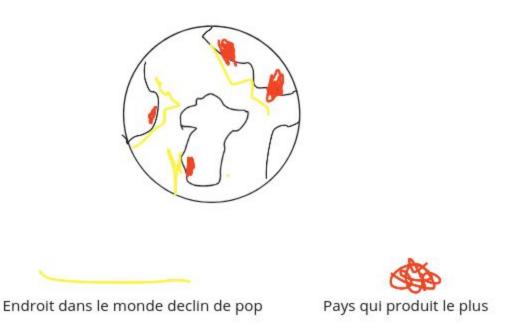
- Quantité des déchets plastiques proche des plages selon la grandeur des bulles



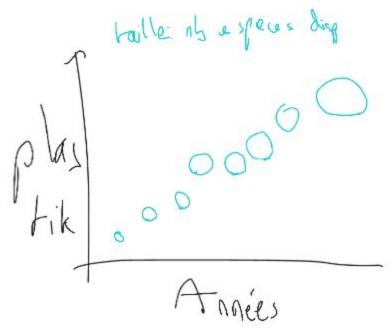
- Croissance des déchets plastiques et de la biodiversité marine en fonction du temps.



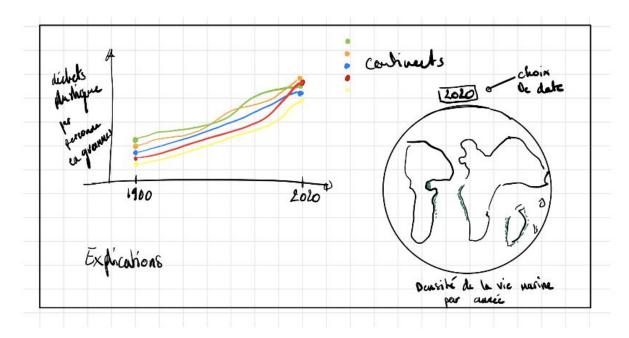
- Pour une zone maritime précise, observation de l'évolution de la quantité plastique et des espèces martimes.



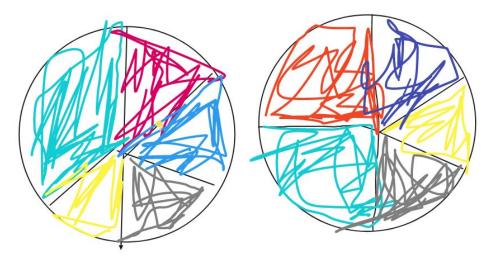
Globe avec les pays produisant le plus de plastique et les zones où la population marine est en déclin.



- Évolution de la quantité de plastique chaque année, et nombre d'espèces marines disparues représentées par la taille des bulles (bubble chart)



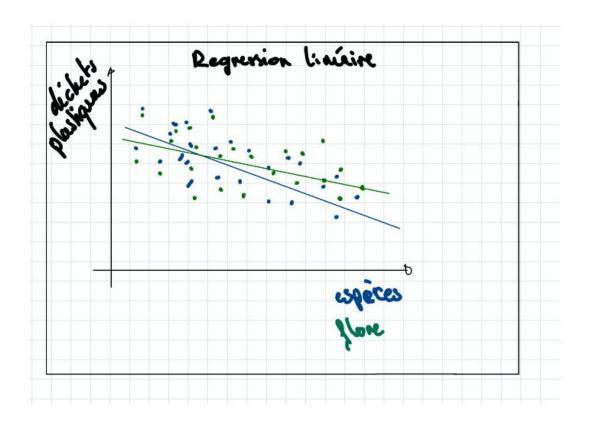
- Croissance de la consommation de plastique par personne en grammes ainsi qu'une représentation sur une carte de la densité de la vie marine au fil du temps.



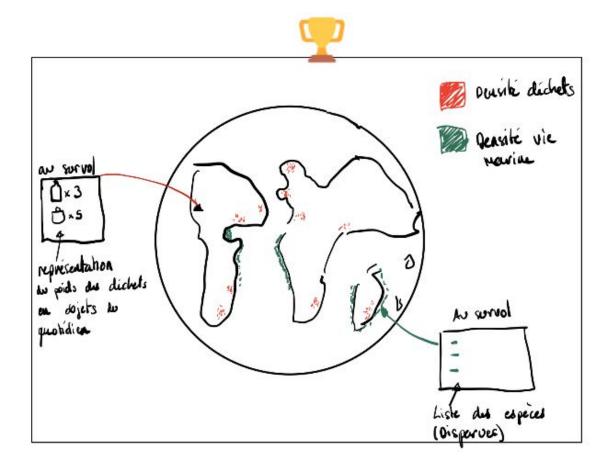
Les pays qui polluent le plus

La mortalité des especes proches de ces pays

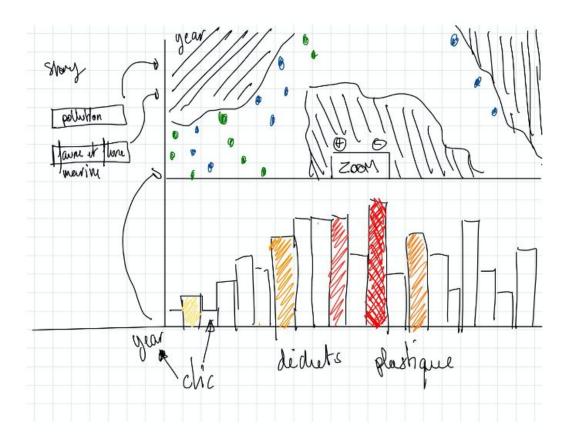
- Diagrammes circulaires: un premier pour les pays les plus pollueurs de plastiques et un second pour le taux de mortalité des espèces marines pour ces pays.



 Régression linéaire afin d'observer si une corrélation existe entre l'évolution de la quantité mondiale de déchets plastiques avec le nombre global d'espèces maritimes.



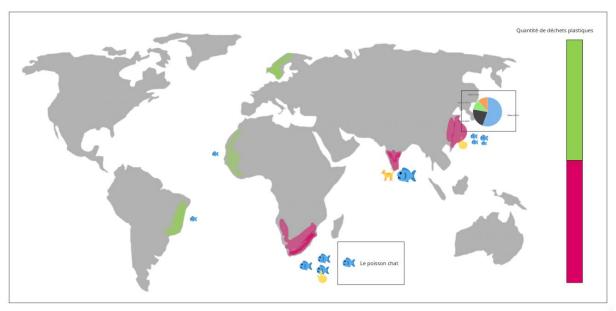
- Une carte avec des zones représentant la densité de déchets plastique et des zones représentant la densité de la vie marine. Au survol de la densité de plastique, on représente le poids par des objets plastiques du quotidien (bouteilles plastiques, sachets plastiques...). Au survol de la vie marine, on affiche la liste des espèces disparues.
- Cela permet une meilleure sensibilisation du public cible : les consommateurs.



- Une visualisation des données faite en deux temps. Quand on zoom sur une partie de la carte on aura un diagramme de bâton qui montrera les déchets plastiques dans cette zone en fonction du temps. Et sur la carte on montera la densité de plastique ainsi que la faune et flore marine.

24/11/2020: Brainstorm d'esquisses finales

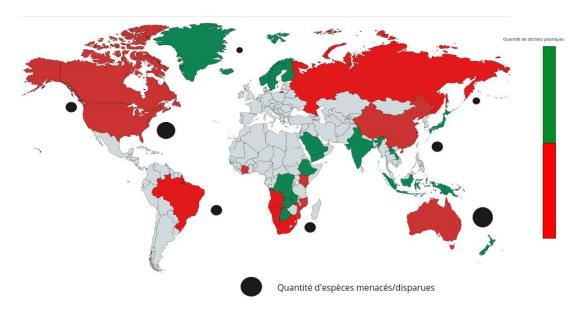
Nous avons fait un brainstorm à partir de l'esquisse choisi précédemment



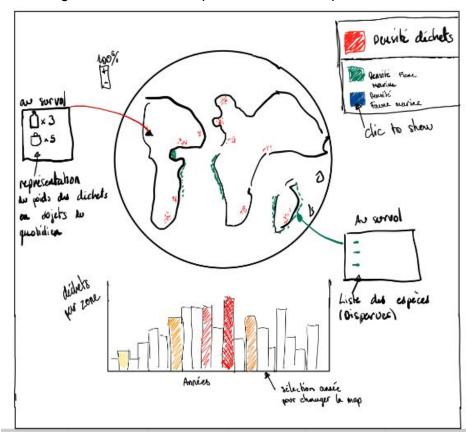
- Heat map permettant d'analyser la quantité des déchets plastiques dans le monde.
- Bubble map (en forme de poisson) afin de montrer la quantité d'espèces marines à un endroit précis.
- Diagramme circulaire lors du survol afin de montrer les principales espèces.



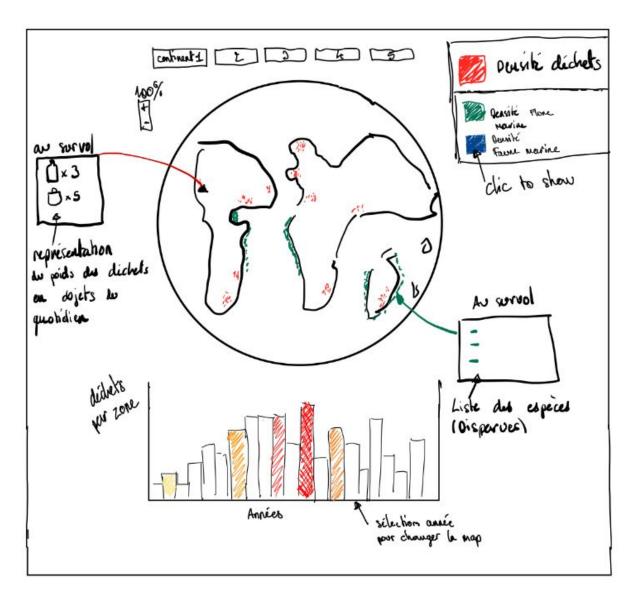
 Représenter la quantité pollution plastique par des bulles rouges et le niveau d'effet sur la population marine par des bulles vertes



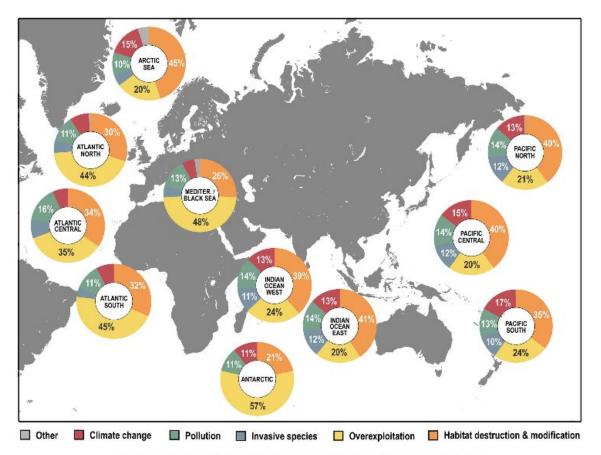
- Représenter la quantité de déchets plastiques pour chaque pays avec un dégradé de vert à rouge. Des bulles noires pour le nombre d'espèces menacées et disparues.



- Reprendre l'idée de base et ajouter un histogramme en dessous de la carte pour représenter par années la quantité de déchets plastique, avec la possibilité de cliquer sur une barre de l'histogramme pour adapter la carte à l'année sélectionné.



- Ajouter à l'idée précédente des boutons au dessus de la carte pour pouvoir zoomer sur un continent précis

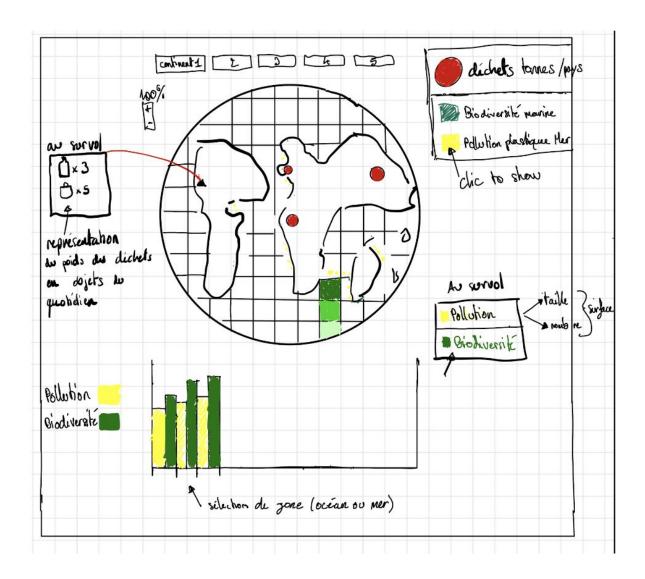


Status of Marine Biodiversity in the Anthropocene

- Représentation des conditions de vie de la biodiversité marine dans les mers et océans en forme de diagramme circulaire sur une carte.

01/12/2020:

- Travaux réalisés :
 - Recherches datasets
 - Mise en place structure page html.
 - Création de l'esquisse finale
- Travaux restants :
 - Nettoyage des données.
 - [bar chart] Affectation à chaque relevé d'un secteur géographique (mer ou océan).
 - Design page html
 - o [bar chart] Construire Visualisation



Mutualisation des 4 esquisses sélectionnées lors de la dernière séance.

- La première visualisation (globe) sera consacrée aux données terrestres et marines afin de voir la consommation et la pollution plastique
- La deuxième sera consacrée aux données marines afin de conclure sur la situation des mers/océans: biodiversité et pollution.

Nous avons dû supprimer l'évolution par année, car nous n'avons pas réussi à trouver des données sur une large période. Nous avons donc sélectionné l'année 2010, car nous avons trouvé les 3 datasets correspondants.

03/12/2020:

Travaux réalisés :

- [bar chart]: Nettoyage des données.
- o Ajout des centroïds (coordonnées géographiques représentant le centre de chaque pays en latitude et longitude) aux données des déchets plastiques par pays afin de pouvoir placer les bulles de la bubble map.
- [bar chart] Requetage API Google pour affecter à chaque prélevé un secteur géographique (résultats insuffisants)
- o [bar chart] Réflexion sur un nouvel angle d'approche du problème d'affectation des secteurs géographiques.
- Recherche de datasets:

Données biodiversité: https://data.unep-wcmc.org/datasets/17

Données sur la consommation de plastique par pays:

https://hub.arcgis.com/datasets/CESJ::estimate-of-plastic-pollution-in-the-worlds-oceans-km2-2 00-mm

Données sur la pollution des océans :

https://ourworldindata.org/plastic-pollution#total-plastic-waste-by-country



o Etudes des données disponibles:



• Travaux restants:

 [bar chart] Affectation à chaque relevé d'un secteur géographique (mer ou océan).

14/12/2020:

• Travaux réalisés :

- [bar chart] Conception d'un outil permettant de récupérer des prélèvements par zone géographique.
- [bar chart] Réflexion sur la manière la plus pertinente de regrouper les deux jeux de données.

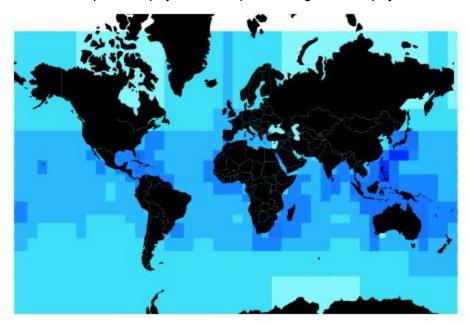
• Travaux restants:

- o [bar chart] Construire Visualisation
- o [bar chart] Grouper les deux jeux de données.

04/01/2021:

• Travaux réalisés :

- Affichage des données de la biodiversité: grille de couleur de différentes intensité de bleus.
- Affichage des données de consommation de plastique par personne en gramme pour les pays côtiers : points rouge sur les pays de tailles différentes.





07/01/2021:

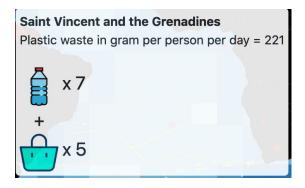
- Travaux réalisés :
 - o Possibilité de zoomer et de bouger la carte
 - Création des boutons pour chaque continent pour pouvoir zoomer sur un continent précis. Nous avons dû récupérer les coordonnées géographiques de chaque continent.



 Possibilité de hover sur les représentations pour avoir plus d'informations sur des tooltips. Correspondance de la quantité plastique dans les mers/océans en nombre de bouteilles ou en court de tennis pouvant contenir 8953 bouteilles.



Et correspondance de la quantité déchets plastiques en nombre de bouteilles plus en nombre de sac plastiques.



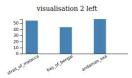
- [bar chart] Groupement des données de biodiversité et de quantitée de plastique par application de l'algorithme NearestNeighbors.
- [bar chart] Création des premiers jeux de données exploitables en .csv pour la visualisation.

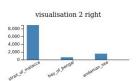
Travaux restants :

o [bar chart] Produire la visualisation.

11/01/2021:

- Travaux réalisés :
 - Rédaction des storylines
 - Finalisation du css.
 - o [bar chart] Première visualisation.





 [bar chart] Réflexion nouvelle mise en forme pour visualisation plus agréable à lire en conservant des échelles pertinentes.

The lives of our single use plastics within the marine flora and fauna in 2010

We are currently consuming far too many plastic disposables.

90% of non reusable plastics end up in the natural environment

12/01/2021:

- Travaux réalisés :
 - Création des légendes.
 - o [bar chart] Adaptation des données aux vues de la visualisation du dessus.
 - [bar chart] Mise en place de la visualisation finale. Afin de mieux comparer les deux données pour une même mer/océan, nous avons changé le type visualisation de 2 diagrammes bâtons à 1 diagramme pyramide.
 - ([bar chart] Si vous souhaitez avoir un aperçu plus détaillé de l'ensemble du travail réalisé sur les données le notebook utilisé est disponible en annexe 1.)

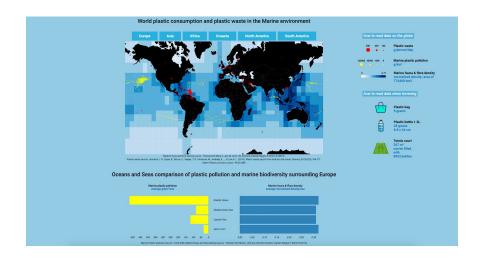


13/01/2021:

- Dernières modifications
 - o Harmoniser les deux titres des visualisations
 - o Déplacer la légende de la deuxième visualisation en dessous du tit







Annexe 1 (23 Pages):

data_preprocessing

January 13, 2021

0.1 1) Netoyage et fusion des données

0.1.1 1.1) Netoyage des données

/home/verdu/.local/lib/python3.6/site-packages/ipykernel_launcher.py:6:
SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame. Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
[2]: ocean_biodiv_source = pd.io.parsers.

→read_csv("WCMC-019-PatternsBiodiversity2010-AcrossTaxa.csv")

#print(ocean_biodiv_source.shape)

# GRIDCODE,N,10,0 X_COORD,N,18,5 Y_COORD,N,18,5 AllTaxa,N,19,11

→AllNorm,N,19,11 CoastNorm,N,19,11 OceanNorm,N,19,

ocean_biodiv = ocean_biodiv_source.iloc[:,[1,2,4]]

ocean_biodiv.columns= ['LONGITUDE','LATITUDE',"VALUE"]

#print(ocean_plastic.shape)

#ocean_plastic.to_csv(r'ocean_biodiversity.csv', index = False, header=True)
```

```
[3]: import pandas as pd
    ocean_plastic = pd.io.parsers.read_csv("ocean_plastic.csv")
    ocean_biodiv = pd.io.parsers.read_csv("ocean_biodiversity.csv")
```

```
[4]: import requests
     import smtplib
     import time # 100 requetes par secondes MAX
     nb_elem = 300
     def geocode (df):
         list_loc = []
         not_found = 0
         start_time = time.time()
         for index, row in df.iterrows():
            lat = row["LATITUDE"]
            long = row["LONGITUDE"]
            url = 'https://maps.googleapis.com/maps/api/geocode/json?latlng=' +u
      ⇒str(lat) +','+ str(long)
      →+'&result_type=natural_feature&key=AIzaSyBNJy6FzFAgIniCXx0WXubaBFBoNxxr8FU'
            r = requests.get(url)
            res = r.json()
             if res["status"] == 'OK' :
                 pass
                 list_loc.
      →append(res["results"][0]['address_components'][0]["long_name"])
                 #print(res["results"][0]['address components'][0]["long name"])
             else :
                 not_found+=1
                 list_loc.append(res["status"])
         elapsed_time = time.time() - start_time
         print("{0:.1f} % localisations non trouvées ! ".format((not_found/nb_elem)_
      →* 100))
     geocode(ocean_plastic.head(nb_elem))
     # GOOGLE MAPS NE NOUS AIDE PAS VRAIMENT!
```

53.7 % localisations non trouvées !

L'api google ne nous donne pas de résultats satisfaisants, il faut donc prendre le problème sous un autre angle !

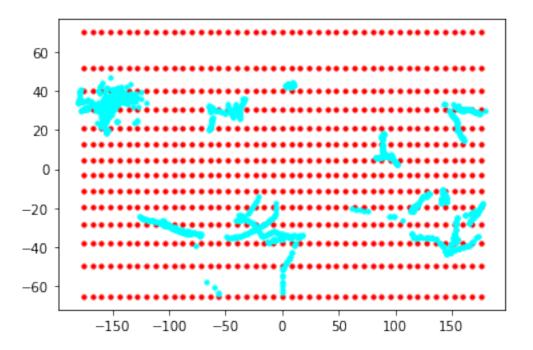
0.1.2 Fusion des données

```
[5]: import matplotlib.pyplot as plt
               plt.scatter(ocean_biodiv['LONGITUDE'], ocean_biodiv['LATITUDE'], s = 10, c = 1
                  →'red', marker = 'o')
               plt.scatter(ocean_plastic["LONGITUDE"],ocean_plastic["LATITUDE"], s = 10, c = 
                  In /home/verdu/.local/lib/python3.6/site-packages/matplotlib/mpl-
             data/stylelib/ classic test.mplstyle:
             The text.latex.preview rcparam was deprecated in Matplotlib 3.3 and will be
             removed two minor releases later.
             In /home/verdu/.local/lib/python3.6/site-packages/matplotlib/mpl-
             data/stylelib/_classic_test.mplstyle:
             The mathtext.fallback_to_cm rcparam was deprecated in Matplotlib 3.3 and will be
             removed two minor releases later.
             In /home/verdu/.local/lib/python3.6/site-packages/matplotlib/mpl-
             data/stylelib/_classic_test.mplstyle: Support for setting the
              'mathtext.fallback_to_cm' rcParam is deprecated since 3.3 and will be removed
             two minor releases later; use 'mathtext.fallback : 'cm' instead.
             In /home/verdu/.local/lib/python3.6/site-packages/matplotlib/mpl-
             data/stylelib/_classic_test.mplstyle:
             The validate bool maybe none function was deprecated in Matplotlib 3.3 and will
             be removed two minor releases later.
             In /home/verdu/.local/lib/python3.6/site-packages/matplotlib/mpl-
             data/stylelib/_classic_test.mplstyle:
             The savefig.jpeg_quality rcparam was deprecated in Matplotlib 3.3 and will be
             removed two minor releases later.
             In /home/verdu/.local/lib/python3.6/site-packages/matplotlib/mpl-
             data/stylelib/_classic_test.mplstyle:
             The keymap.all_axes rcparam was deprecated in Matplotlib 3.3 and will be removed
             two minor releases later.
             In /home/verdu/.local/lib/python3.6/site-packages/matplotlib/mpl-
             data/stylelib/_classic_test.mplstyle:
             The animation.avconv_path rcparam was deprecated in Matplotlib 3.3 and will be
             removed two minor releases later.
             In /home/verdu/.local/lib/python3.6/site-packages/matplotlib/mpl-
             data/stylelib/_classic_test.mplstyle:
```

[5]: <matplotlib.collections.PathCollection at 0x7f5e0271c780>

removed two minor releases later.

The animation.avconv_args rcparam was deprecated in Matplotlib 3.3 and will be



```
[6]: from sklearn.neighbors import NearestNeighbors
     import numpy as np
     def distance(p1, p2):
         lon1, lat1 = p1
         lon2, lat2 = p2
         # convert decimal degrees to radians
         lon1, lat1, lon2, lat2 = map(np.radians, [lon1, lat1, lon2, lat2])
         # haversine formula
         dlon = lon2 - lon1
         dlat = lat2 - lat1
         a = np.sin(dlat/2)**2 + np.cos(lat1) * np.cos(lat2) * np.sin(dlon/2)**2
         c = 2 * np.arcsin(np.sqrt(a))
         km = 6367 * c
         return km
     biodiv_array = []
     biodiv_lat = ocean_biodiv["LATITUDE"]
     biodiv_long = ocean_biodiv["LONGITUDE"]
     for i in range(len(biodiv_lat)) :
         biodiv_array.append([biodiv_long[i],biodiv_lat[i]])
         biodiv_2D = np.array(biodiv_array)
     tmp_array = []
     polution = []
     biodiv_tab = []
```

```
plastic_lat = ocean_plastic["LATITUDE"]
plastic_long = ocean_plastic["LONGITUDE"]
type(plastic_long)
for i in range(len(plastic_lat)) :
    tmp = [plastic_long[i], plastic_lat[i]]
    #print(tmp)
    tmp_array = np.append(biodiv_2D,[tmp],axis=0)
    nbrs = NearestNeighbors(n_neighbors=2, metric=distance).fit(tmp_array)
    distances, indices = nbrs.kneighbors(tmp_array)
    #print("id : "+ str(i) + " voisin : " + str(indices[-1][1]) + " distance :
    '' + str(distances[-1][1]))
    biodiv_tab.append(ocean_biodiv['VALUE'][indices[-1][1]])
```

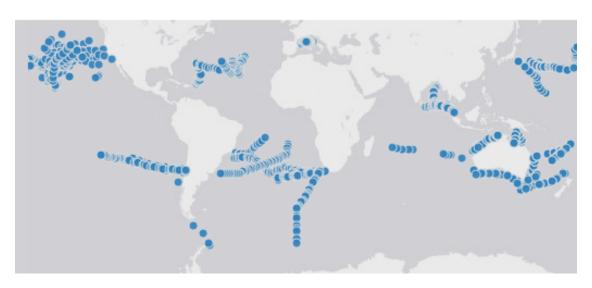
```
[7]: global_dataset = pd.io.parsers.read_csv("ocean_plastic.csv")
global_dataset["BIODIVERSITY"] = biodiv_tab
global_dataset.columns= ['LATITUDE', 'LONGITUDE', "PLASTIC", "BIODIVERSITY"]
#print(global_dataset.head())
#global_dataset.to_csv(r'global_dataset.csv', index = False, header=True)
```

0.2 2) Définition des espaces maritimes

```
[8]: global_dataset = pd.io.parsers.read_csv("global_dataset.csv")
```

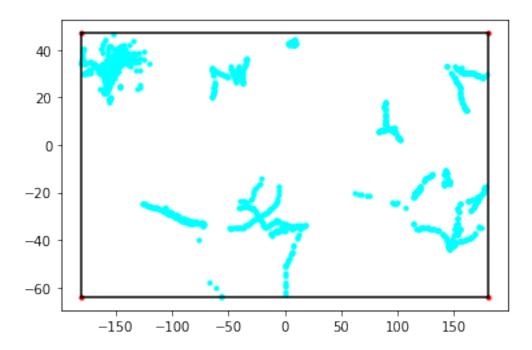
```
[9]: from scipy.spatial import ConvexHull
     from scipy.spatial import Delaunay
     import matplotlib.pyplot as plt
     import numpy as np
     import pandas as pd
     import statistics
     def in_hull(p, hull):
         if not isinstance(hull,Delaunay):
             hull = Delaunay(hull)
         return hull.find_simplex(p)>=0
     def zoom_graph(long_limit, lat_limit, dataset) :
         hull list = []
         for i in range(len(long_limit)) :
             hull_list.append([long_limit[i],lat_limit[i]])
         array_hull = np.array(hull_list)
         hull = ConvexHull(array_hull)
         include_lat = []
         include_long = []
         include_plastic = []
```

```
include_biodiv = []
  points_array = []
  for i in range(len(dataset["LONGITUDE"])) :
      points_array.append([dataset["LONGITUDE"][i], dataset["LATITUDE"][i]])
   i = 0
  include_plastic = 0
   include_biodiv = 0
  for id ,val in enumerate(points_array):
       if in_hull(val,array_hull) :
           include_long.append(val[1])
           include_lat.append(val[0])
           include_plastic += dataset["PLASTIC"][id]
           include_biodiv += dataset["BIODIVERSITY"][id]
  print(i)
  if(i == 0):
      print('no data !')
      return
  avg_biodiv = round((include_biodiv * 100 / i),2)
  avg_plastic = round((include_plastic / i),2)
  print(" {0:.1f} % de biodiversité, {1:.1f} de plastique ???".
→format((avg_biodiv),(avg_plastic)))
  plt.scatter(include_lat, include_long, s = 10, c = 'cyan', marker = 'o')
  plt.scatter(long_limit, lat_limit, s = 10, c = 'red', marker = 'o')
  for simplex in hull.simplices:
      plt.plot(array_hull[simplex, 0], array_hull[simplex, 1], 'k-')
  return([avg_biodiv, avg_plastic])
```

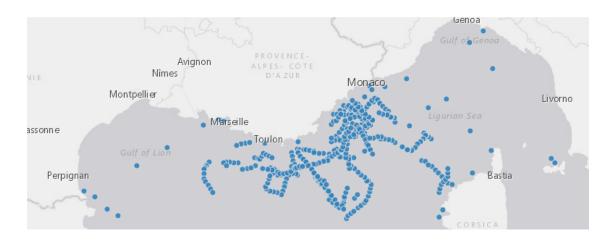


```
[10]: world_data = zoom_graph([-180,-180,180], [-64,47,47,-64],global_dataset) print(world_data)
```

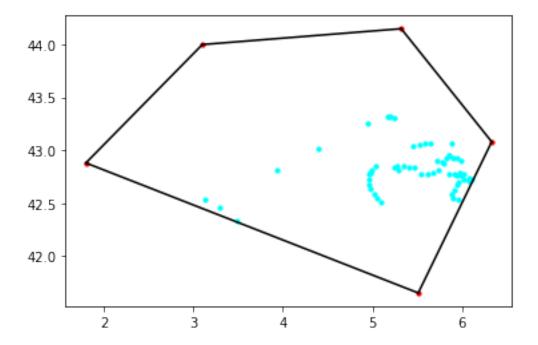
1571 34.0 % de biodiversité, 3084.7 de plastique ??? [33.96, 3084.65]



0.2.1 2.1) Europe



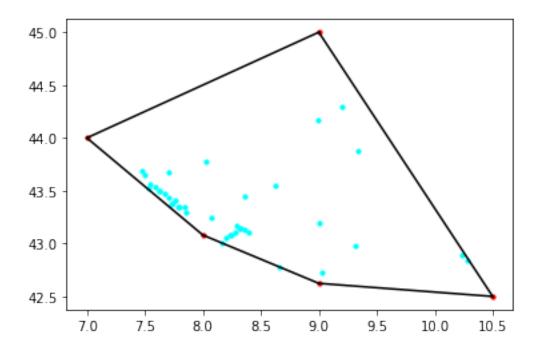
31.6 % de biodiversité, 30.6 de plastique $\ref{10}$?? [31.6, 30.56]



```
[12]: ligurian_sea = zoom_graph([8,9,9,10.5,7],[43.080925,45, 42.623855,42.

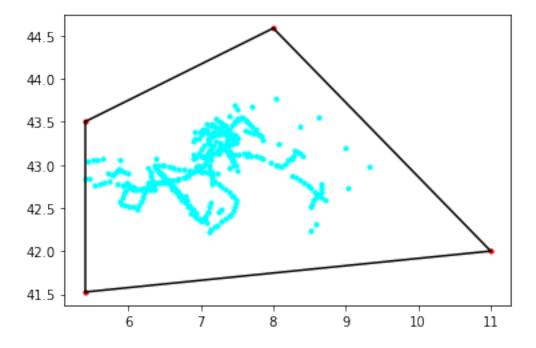
$\infty$5,44],global_dataset)
print(ligurian_sea)
```

31.2 % de biodiversité, 117.5 de plastique ??? [31.19, 117.51]



[13]: mediterranean_sea = zoom_graph([8,5.4,11,5.4],[44.586555,41.525030,42,43. →5],global_dataset)
print(mediterranean_sea)

383 31.2 % de biodiversité, 81.5 de plastique ??? [31.19, 81.46]



```
[14]: d = {'gulf_of_lion' : gulf_of_lion, 'ligurian_sea' : ligurian_sea, □

→'mediterranean_sea' : mediterranean_sea }

df = pd.DataFrame(d)

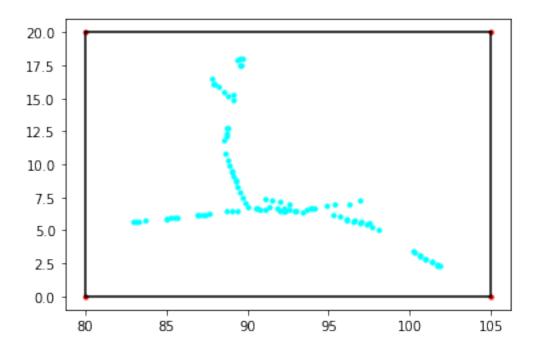
df.to_csv (r'europe.csv', index = False, header=True) # biodiv, plastic
```

0.2.2 2.3) Asie



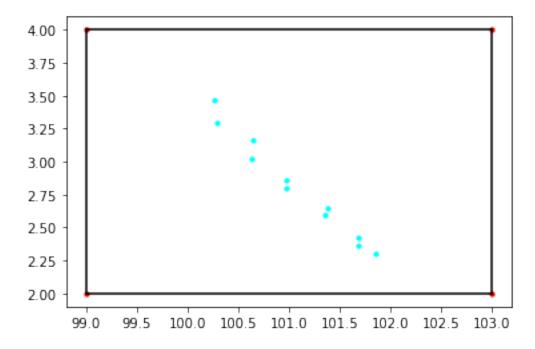
```
[15]: asie = zoom_graph([105,105,80,80] ,[20,0,20,0],global_dataset) print(asie)
```

98 47.3 % de biodiversité, 2332.2 de plastique ??? [47.3, 2332.25]



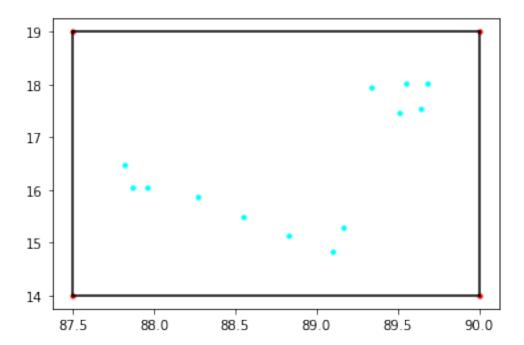
[16]: strait_of_malacca = zoom_graph([99,99,103,103],[2,4,2,4],global_dataset)
print(strait_of_malacca)

11 54.1 % de biodiversité, 8928.8 de plastique ??? [54.09, 8928.78]

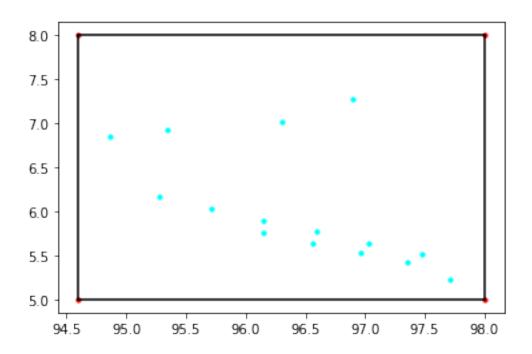


```
[17]: bay_of_bengal = zoom_graph([87.5,87.5,90,90],[14,19,14,19],global_dataset) print(bay_of_bengal)
```

13 43.4 % de biodiversité, 633.5 de plastique ??? [43.43, 633.51]



15 56.7 % de biodiversité, 1564.8 de plastique ??? [56.66, 1564.8]

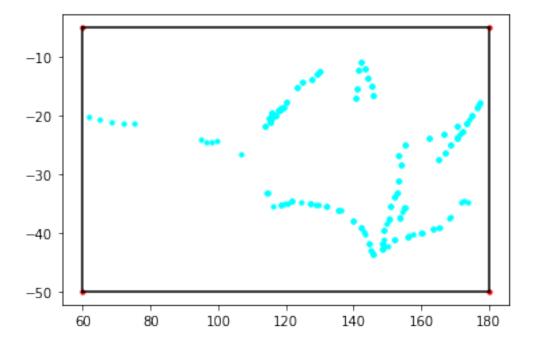


0.2.3 2.3) Oceanie



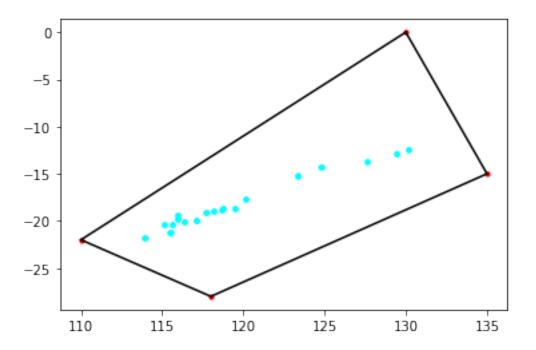
[20]: oceanie = zoom_graph([60,60,180,180],[-50,-5,-50,-5],global_dataset) print(oceanie)

241 48.6 % de biodiversité, 69.8 de plastique ??? [48.62, 69.83]



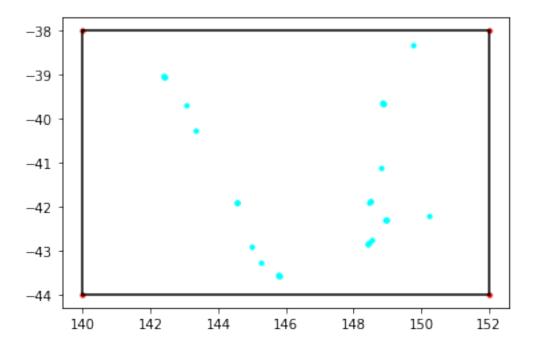
[21]: timar_sea = zoom_graph([118,110,135,130],[-28,-22,-15,0],global_dataset) print(timar_sea)

59 55.0 % de biodiversité, 0.0 de plastique ??? [55.04, 0.0]



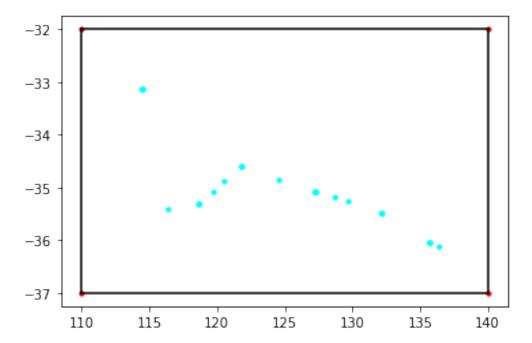
[22]: bass_strait = zoom_graph([140,140,152,152],[-44,-38,-44,-38],global_dataset) print(bass_strait)

44.1 % de biodiversité, 168.4 de plastique ??? [44.09, 168.41]



```
[23]: great_australian_bay = com_graph([110,110,140,140],[-37,-32,-37,-32],global_dataset)
print(great_australian_bay)
```

25 37.9 % de biodiversité, 8.2 de plastique ??? [37.91, 8.22]



```
[24]: d = {'timar_sea' : timar_sea, 'bass_strait' : bass_strait, □

□'great_australian_bay' : great_australian_bay }

df = pd.DataFrame(d)

df.to_csv (r'oceanie.csv', index = False, header=True) # biodiv, plastic
```

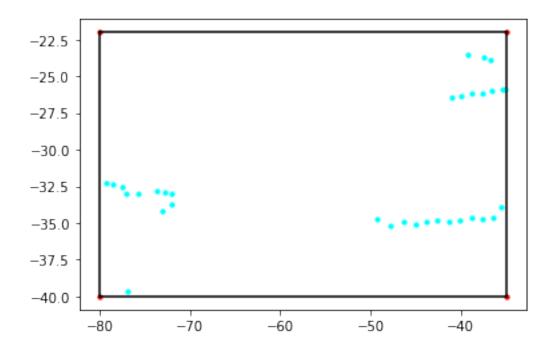
0.2.4 2.4) Amerique latine



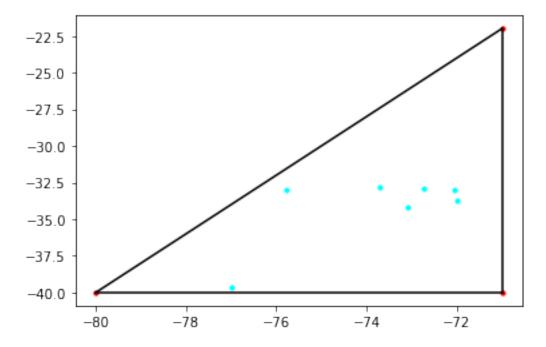
[25]: latine_america = zoom_graph([-80,-80,-35,-35], [-40,-22,-22,-40],global_dataset) print(latine_america)

33

30.3 % de biodiversité, 204.5 de plastique $\ref{200}$ [30.29, 204.53]



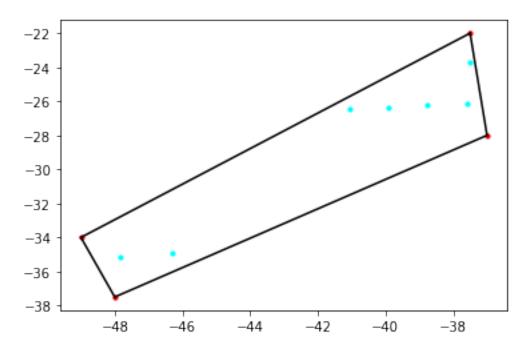
7 25.0 % de biodiversité, 439.9 de plastique ??? [24.98, 439.95]



```
[27]: brasilia_seaboar = zoom_graph([-48,-49,-37.5,-37], [-37. 

$\iff 5,-34,-22,-28],global_dataset)$
print(brasilia_seaboar)
```

37.2 % de biodiversité, 88.4 de plastique ??? [37.23, 88.4]

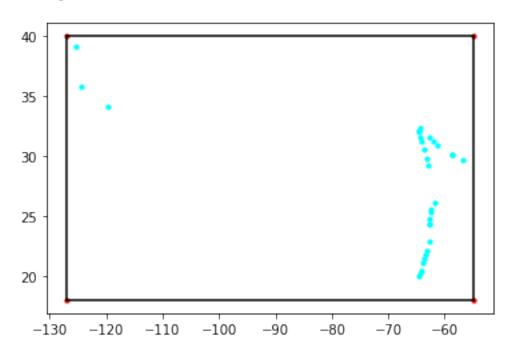


0.2.5 2.5) Amerique du nord



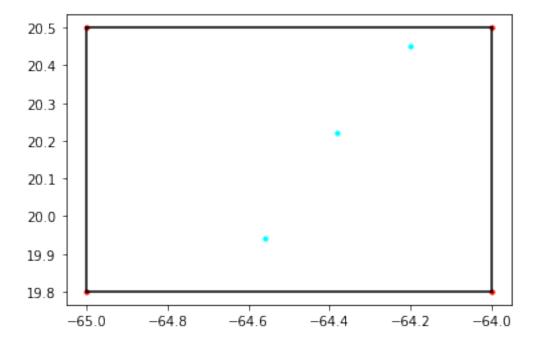
[28]: north_america = zoom_graph([-127,-127,-55,-55], [18,40,40,18],global_dataset) print(north_america)

31 38.1 % de biodiversité, 1438.6 de plastique ??? [38.11, 1438.59]



print(puerto_rico_bay)

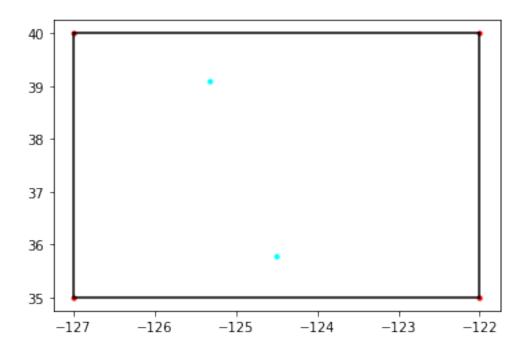
3 41.5 % de biodiversité, 1218.2 de plastique ??? [41.55, 1218.16]



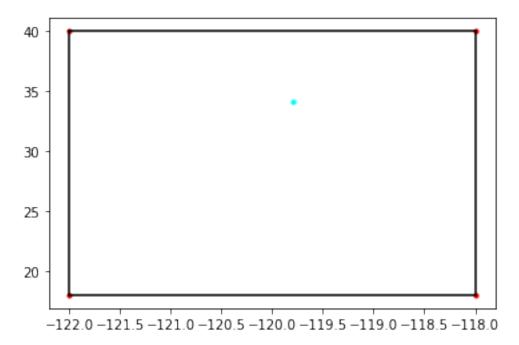
```
[30]: san_francisco_bay = zoom_graph([-127,-127,-122,-122], 

→ [35,40,40,35],global_dataset)
print(san_francisco_bay)
```

2 37.0 % de biodiversité, 2131.7 de plastique ??? [36.99, 2131.7]



41.7 % de biodiversité, 0.0 de plastique ??? [41.72, 0.0]



```
[32]: d = {'puerto_rico_bay' : puerto_rico_bay, 'san_francisco_bay' : □

⇒san_francisco_bay, 'los_angeles_bay' : los_angeles_bay }

df = pd.DataFrame(d)

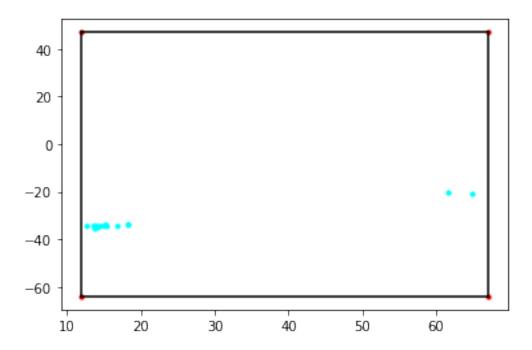
df.to_csv (r'north_america.csv', index = False, header=True) # biodiv, plastic
```

0.2.6 2.5) Afrique

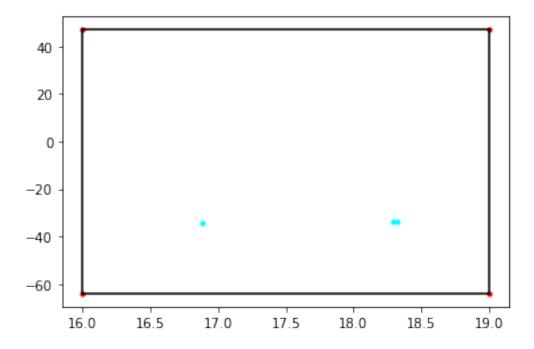


```
[33]: africa = zoom_graph([12,12,67,67], [-64,47,47,-64],global_dataset) print(africa)
```

15 45.9 % de biodiversité, 733.5 de plastique ??? [45.86, 733.51]

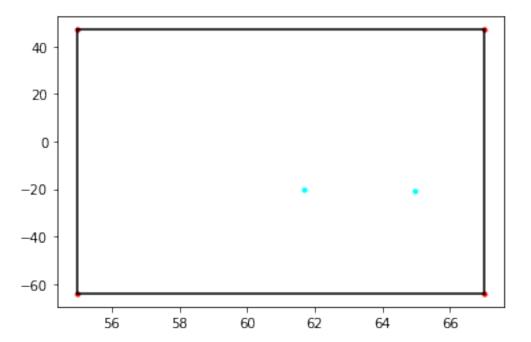


47.5 % de biodiversité, 433.8 de plastique ??? [47.46, 433.77]



```
[35]: mauritus_bay = zoom_graph([55,55,67,67], [-64,47,47,-64],global_dataset) print(mauritus_bay)
```

2 35.5 % de biodiversité, 3872.8 de plastique ??? [35.48, 3872.79]



```
[36]: d = {'cape_town_bay' : cape_town_bay, 'mauritus_bay' : mauritus_bay}
df = pd.DataFrame(d)
df.to_csv (r'africa.csv', index = False, header=True) # biodiv, plastic
```

[]: