Binf 6210 Assignment 1

Introduction:

Natural environmental products have always provided a starting framework for new drug developments. The phylum of Bryozoa, found in marine habitats, has demonstrated potential for anticancer drug development and potential benefits to human health (Figuerola and Avila, 2019), (Ciavatta *et al,* 2020). Bryozoa are an aquatic organism, also known as moss animals, that consist of tiny modules, called zooids (Ciavatta, 2020). Found in greater abundance in Canada and United States. Among the marine invertebrates, Bryozoa are an important bioactive compound (Ciavatta, 2020). Currently, two-hundred and thirty compounds have been isolated from twenty-six species of Bryozoa (Ciavatta, 2020). These bioactive compounds, have demonstrated, illustrated in Ciavatta, 2020, how they can complement and outline short comings of existing cancer treatments (Ciavatta, 2020).

The significance of this organism is growing with its potential promise for enhancing cancer treatment, this project poses a research question, how many species in the phylum of Bryozoa, have been barcoded in Canada vs United States, which have contributed to enhancing cancer treatments. According to the BOLD database, bryozoan species are found in greater abundance in North America in comparison to other continents, it is predicated that bryozoan species whom have been found effective for cancer treatment will be found in higher quantities in Canada and United States in comparison to other parts of the world.
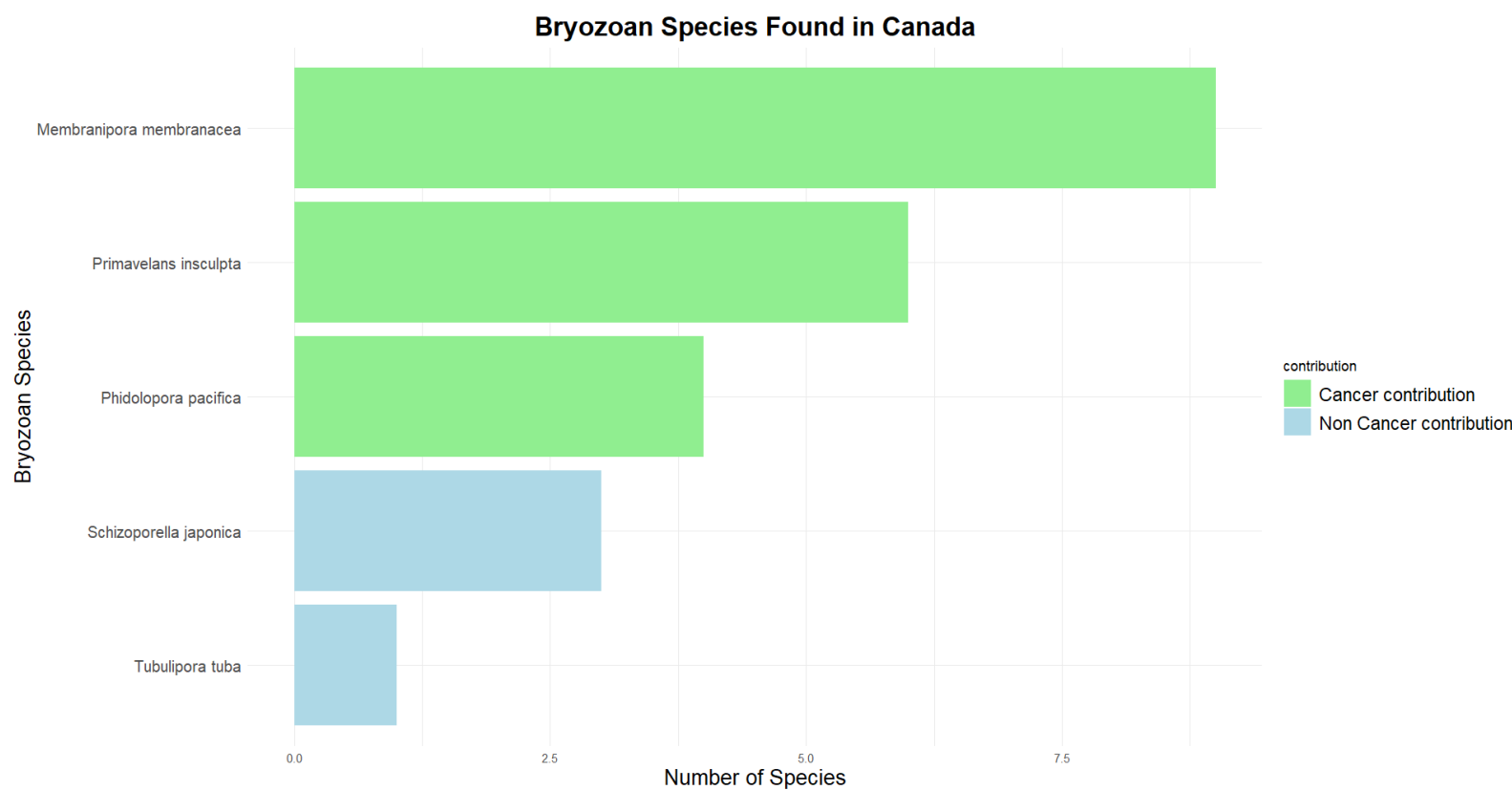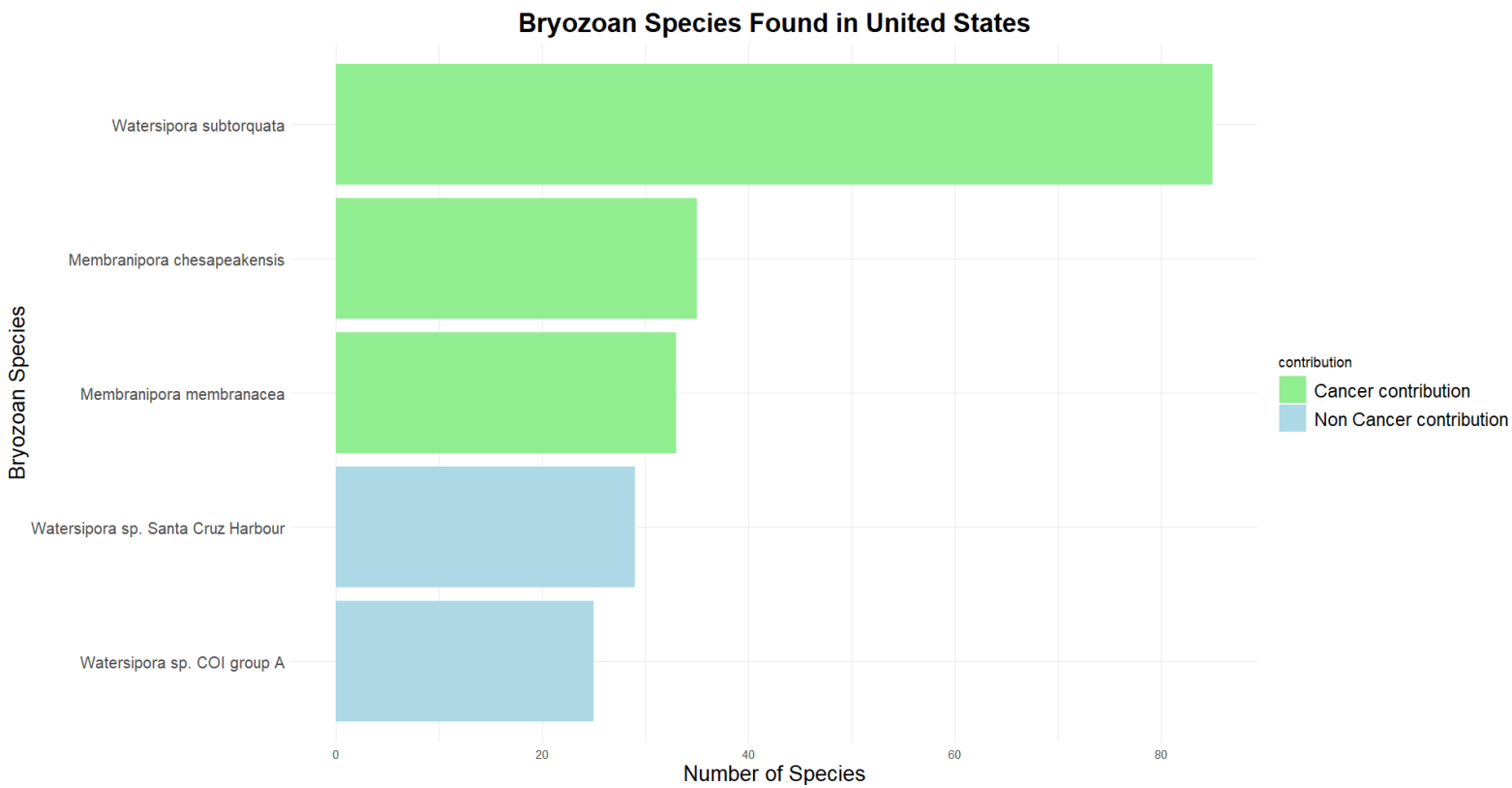
Figures:

Figure 1:

Figure 2:
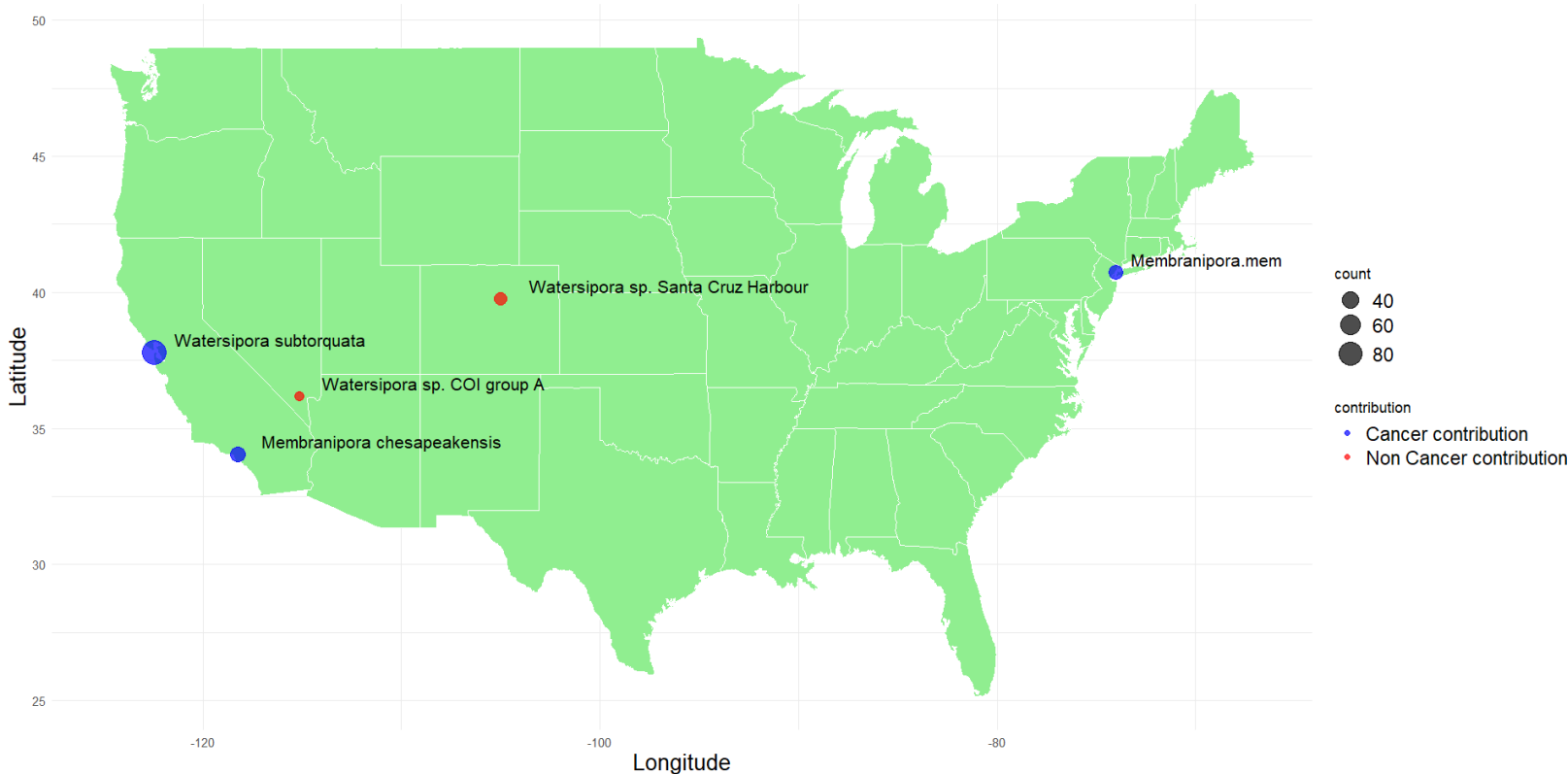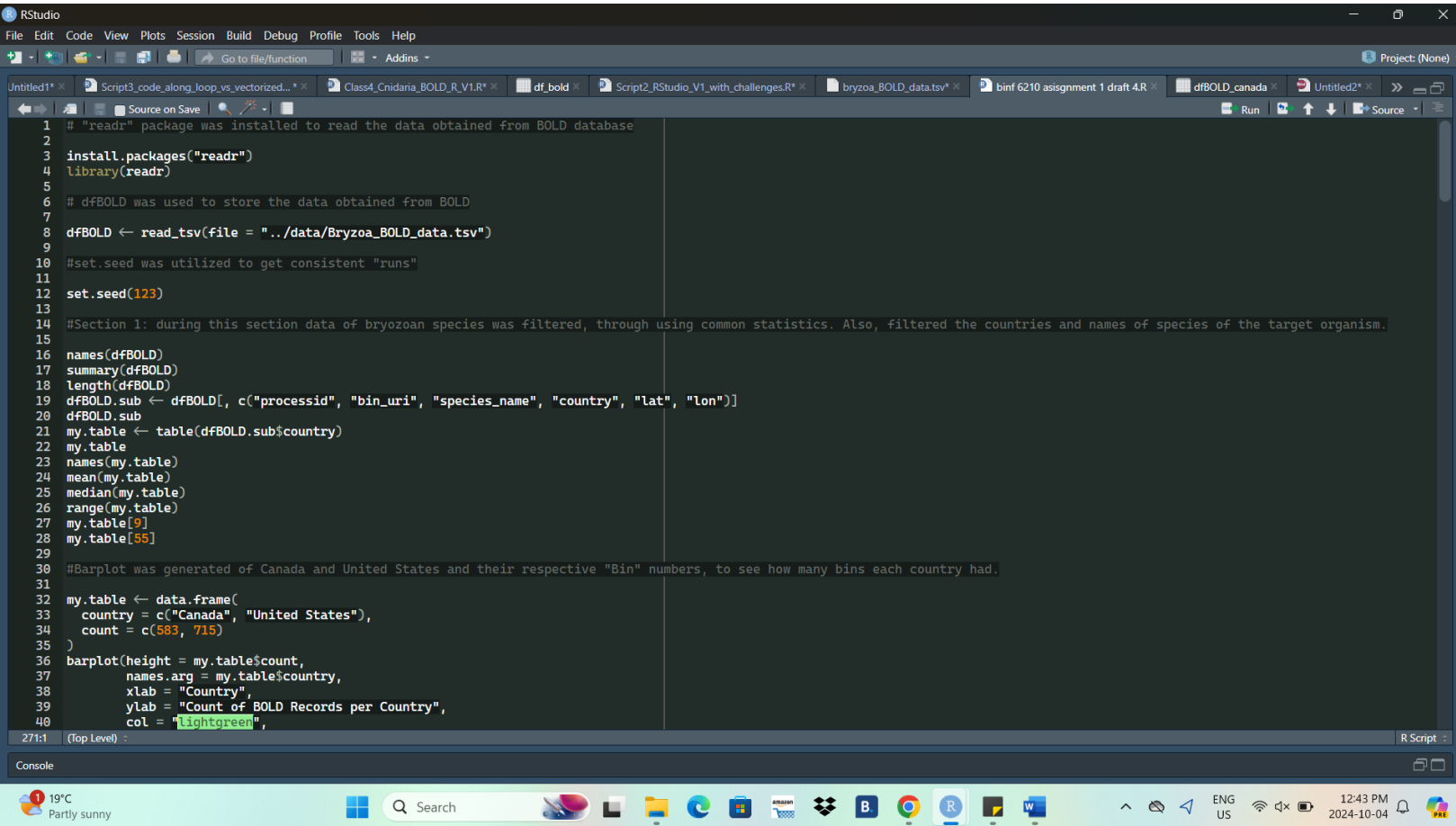
Figure 3:

Coding Section:


```r
# "readr" package was installed to read the data obtained from BOLD database

install.packages("readr")
library(readr)

# dfBOLD was used to store the data obtained from BOLD

dfBOLD <- read_tsv(file = "../data/Bryzoa_BOLD_data.tsv")

#set.seed was utilized to get consistent "runs"

set.seed(123)

#Section 1: during this section data of bryozoan species was filtered, through using common statistics. Also, filtered the countries and names of species of the target organism.

names(dfBOLD)
summary(dfBOLD)
length(dfBOLD)
dfBOLD.sub <- dfBOLD[, c("processid", "bin_uri", "species_name", "country", "lat", "lon")]
dfBOLD.sub
my.table <- table(dfBOLD.sub$country)
my.table
names(my.table)
mean(my.table)
median(my.table)
range(my.table)
my.table[9]
my.table[55]

#Barplot was generated of Canada and United States and their respective "Bin" numbers, to see how many bins each country had.

my.table <- data.frame(
  country = c("Canada", "United States"),
  count = c(583, 715)
)
barplot(height = my.table$count,
        names.arg = my.table$country,
        xlab = "Country",
        ylab = "Count of BOLD Records per Country",
        col = "lightgreen",
```

```r
        col = "lightgreen",
        main = "BOLD Records by Country")



#The Bryozoan species in obtained from the BOLD data base was filtered out and barplot was generated to see the names and amount of particular species in the BOLD data base. Basic statistics were also conducted.

my.table <- table(dfBOLD.sub$species_name)
my.table
barplot(my.table,
        main = "Bold Record of Bryozoan Species",
        xlab = "Bryozoan Species",
        ylab = "Number of Bold Records")
names(my.table)
mean(my.table)
median(my.table)
range(my.table)

#Section 2: During this section latitude and longitude of the data was generated,so we can filter the lat and long of Canada and United States.


my.table <- table(dfBOLD.sub$bin_uri)
my.table
my.table <-table(dfBOLD.sub$lat)
my.table
my.table <- table(dfBOLD.sub$lon)
my.table
library(dplyr)

dfBOLD_United_States <- dfBOLD.sub %>%
  filter(country == "United States")

table_lat <- table(dfBOLD_United_States$lat)

table_lat
library(dplyr)

dfBOLD_United_States <- dfBOLD.sub %>%
  filter(country == "United States")
```

R RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Untitled1* | Script3_code_along_loop_vs_vectorized... * | Class4_Cnidaria_BOLD_R_V1.R* | df_bold | Script2_RStudio_V1_with_challenges.R* | bryzoa_BOLD_data.tsv* | binf 6210 asisgnment 1 draft 4.R | dfBOLD_canada | Untitled2*

Source on Save

```
 79   table_lon <- table(dfBOLD_United_States$lon)
 80
 81   table_lon
 82
 83   #The BOLD records of Bryozoan species per country, who had high abundace was sorted from highest to lowest, to get a good idea of which country had high abundance. I had figured it
       was United States and Canada, however I wanted to make sure.
 84
 85   plot(sort(table(dfBOLD.sub$country), decreasing = TRUE)[1:5],
 86        main = "Bold Records of Bryozoan Species per Country",
 87        xlab = "Countries",
 88        ylab = "Number of Bold Records")
 89
 90   # A table was generated in effort to filter the Bryozoan species with their respective conutries to further filter out the data to answer my research question.
 91
 92   y <- table(dfBOLD.sub$country,dfBOLD.sub$species_name)
 93   y
 94   dfBOLD_canada <- dfBOLD.sub[dfBOLD.sub$country =="Canada",]
 95   library(dplyr)
 96   dfBOLD_canada <- dfBOLD.sub %>% filter(country == "Canada")
 97   ls()
 98   y <- table(dfBOLD_canada$species_name)
 99   y
100
101   #The Abundance of North American Bryozoan spcies were graphed.
102
103   barplot(y,
104          main = "The Abundance of Bryozoan Species",
105          xlab = "Bryozoan Species",
106          ylab = "The Abundance")
107   counts <- c(9, 6, 4, 3, 1)
108   names <- c("Membranipora membranacea", "Primavelans insculpta", "Phidolopora pacifica", "Schizoporella japonica", "Tubulipora tuba")
109
110   # Section 3: During this section Bryozoan species found in Canada were plotted using ggplot. This plot highlights the species found in highest abundance in Canada, and also, the
       species that contained bioactive compounds,which contribute to cancer. The first three species are highlighted as contribution to cancer treatment.
111
112
113   library(ggplot2)
114   df <- data.frame(
115       species = c("Membranipora membranacea", "Primavelans insculpta", "Phidolopora pacifica", "Schizoporella japonica", "Tubulipora tuba"),
116
```

271:1 | (Top Level) | R Script

Console

```r
116        count = c(9,6,4,3,1),
117        contribution = c("Cancer contribution", "Cancer contribution", "Cancer contribution", "Non Cancer contribution", "Non Cancer contribution")
118
119  )
120  ggplot(df,aes(reorder(species, count), y = count, fill = contribution)) +
121    geom_bar(stat = "identity") +
122    labs(
123      title = "Bryozoan Species Found in Canada",
124      x = "Bryozoan Species",
125      y = "Number of Species"
126    ) +
127    coord_flip() +
128    theme_minimal() +
129    theme(
130      plot.title = element_text(size = 20,face = "bold", hjust = 0.5),
131      axis.title.x = element_text(size = 16),
132      axis.title.y = element_text(size = 16),
133      axis.text.y = element_text(size = 12),
134      legend.text = element_text(size = 14),
135      legend.key.size = unit(1.5,"lines")
136    ) +
137    scale_fill_manual(values = c("lightgreen", "lightblue"))
138
139
140  #  The Countries were filtered out and then particular data pertaining to United States, such as the species found and their abundance was filtered out. Then barplot was generated to
     see the highest abundance of the target species found in United States and also highlighting the species that contribute to cancer treatment. The first three species were highlighted.
141
142
143  unique(dfBOLD$country)
144  dfBOLD_United_States <- dfBOLD.sub[dfBOLD.sub$country =="United States",]
145  library(dplyr)
146  dfBOLD_United_States <- dfBOLD.sub %>% filter(country == "United States")
147  ls()
148  y <- table(dfBOLD_United_States$species_name)
149  y
150  counts <- c(85,35,33,29,25)
151  names <- c("Watersipora subtorquata","Membranipora chesapeakensis","Membranipora membranacea","Watersipora sp. COI group A","Watersipora sp. Santa Cruz Harbour ")
152
153  library(ggplot2)
154  df <- data.frame(
```

```r
154  df <- data.frame(
155    species = c("Watersipora subtorquata","Membranipora chesapeakensis","Membranipora membranacea","Watersipora sp. COI group A","Watersipora sp. Santa Cruz Harbour"),
156
157    count = c(85,35,33,25,29),
158    contribution = c("Cancer contribution", "Cancer contribution", "Cancer contribution", "Non Cancer contribution", "Non Cancer contribution")
159  )
160  ggplot(df,aes(reorder(species, count), y = count, fill = contribution)) +
161    geom_bar(stat = "identity") +
162    labs(
163      title = "Bryozoan Species Found in United States",
164      x = "Bryozoan Species",
165      y = "Number of Species"
166    ) +
167    coord_flip() +
168    theme_minimal() +
169    theme(
170      plot.title = element_text(size = 20,face = "bold", hjust = 0.5),
171      axis.title.x = element_text(size = 16),
172      axis.title.y = element_text(size = 16),
173      axis.text.y = element_text(size = 12),
174      legend.text = element_text(size = 14),
175      legend.key.size = unit(1.5,"lines")
176    ) +
177    scale_fill_manual(values = c("lightgreen", "lightblue"))
178
179
180    #Section 4: During this section the Bryozoan species from Canada and United States, which also contribute to cancer treatment as found was the previous graphs plotted, were graphed
         together. This was done in effort to compare between the two countries the species in abundance and to answer my research question in a more clearer way. The particular abundance of
         target species found in United states and Canada can now be clearly intreperated.
181
182
183  dfBOLD.sub <- data.frame(
184    species_name = c("Membranipora membranacea","Membranipora membranacea","Watersipora subtorquata","Primavelans insculpta"),
185    country = c("United States", "Canada", "United States","Canada"),
186    contribution = c(35,9,85,6))
187
188  library(dplyr)
189  df_combined <- dfBOLD.sub %>%
190    filter(country %in% c("United States", "Canada")) %>%
191    group_by(species_name, country) %>%
```

```r
191    group_by(species_name, country) %>%
192    summarise(contribution = sum(contribution)) %>%
193    ungroup()
194
195  library(ggplot2)
196
197  ggplot(df_combined, aes(x = "species name",y = contribution,fill = country)) +
198    geom_bar(stat = "identity", position = position_dodge(width = 0.8),width = 0.4) +
199    labs(title = "The Abundance of Membranipora membranacea Species Found in United States and Canada",
200         x = "Membranipora membranacea",
201         y = "Number of species") +
202    theme_minimal() +
203    theme(
204      plot.title = element_text(size = 20, face = "bold", hjust = 0.5),
205      axis.title.x = element_text(size = 16),
206      axis.title.y = element_text(size = 16),
207    )
208
209    # A stacked bar graph was generated to make the intreperation even more precice and clear, as the graph indicated United States has a higher abundance of target species, which
       contribute to cancer treatment.
210
211  ggplot(df_combined, aes(x = species_name, y = contribution, fill = country)) +
212    geom_bar(stat = "identity") +
213    labs(
214      title = "Cancer Contributing Bryozoan Species Found in United States and Canada",
215      x = "Bryozoan Species",
216      y = "Number of Species"
217    ) +
218    theme_minimal() +
219    theme(
220      plot.title = element_text(size = 20, face = "bold", hjust = 0.5),
221      axis.title.x = element_text(size = 16),
222      axis.title.y = element_text(size = 16),
223      axis.text.y = element_text(size = 14),
224      legend.text = element_text(size = 14),
225      legend.key.size = unit(1.5,"lines")
226    ) +
227    coord_flip()
228
229
```

```r
#Section 5: Since we found that United States has a higher abundance in comparison to Canada, map plot was generated to see which part of the country the species are found and in what
abundance.


library(ggplot2)
library(maps)

us_map <- map_data("state")

df <- data.frame(
  species = c("Watersipora subtorquata", "Membranipora chesapeakensis",
              "Membranipora.mem", "Watersipora sp. COI group A",
              "Watersipora sp. Santa Cruz Harbour"),
  count = c(85, 35, 33, 25, 29),
  contribution = c("Cancer contribution", "Cancer contribution",
                   "Cancer contribution", "Non Cancer contribution",
                   "Non Cancer contribution"),
  lat = c(37.7749, 34.0522, 40.7128, 36.1699, 39.7392),
  lon = c(-122.4194, -118.2437, -74.0060, -115.1398, -104.9903)
)

ggplot() +
  geom_polygon(data = us_map, aes(x = long, y = lat, group = group), fill = "lightgreen", color = "white") +
  geom_point(data = df, aes(x = lon, y = lat, color = contribution, size = count), alpha = 0.7) +
geom_text(data = df, aes(x =lon,y = lat, label = species), hjust = -0.1, vjust = -0.5, size = 4.5) +
  scale_size(range = c(3, 8)) +
  labs(
    title = "Distribution of Bryozoan Species in the United States",
    x = "Longitude",
    y = "Latitude"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 20, face = "bold", hjust = 0.5),
    axis.title.x = element_text(size = 16),
    axis.title.y = element_text(size = 16),
    legend.text = element_text(size = 14),
  ) +
  scale_color_manual(values = c("blue", "red"))
```

Results and Discussion:

In the process of answering the research question posed above, it was found many species of Bryozoa have been barcoded in United States and Canada. In figure 1, Bryozoan species, which were found in high abundance in Canada were graphed. The figure depicts abundance of the Membranipora membrancea, Primavelans insculpta, and Phidolopora pacifica. The Bryozoan species mentioned, according to Figuerola and Avila, 2019, have shown compounds and cytotoxic activity against cancer. Similarly, figure 2 illustrates the Bryozoan species found in high abundance in United states. Species such as, Watersipora subtorquata, Membranipora chespeakensis, and Membranipora membrancea were found in high abundance and based on Figuerola and Avila, 2019, these organisms contain compounds and cytotoxic activity against cancer. The results of this project indicates that many Bryozoan species are barcoded in United States and Canada, which have contributed to enhancing cancer treatment, however pertaining to the second question asked in the introduction, relatively high abundance of species were found in United States, in contrast, Canada did not have high abundance of species, which contributed to cancer.

The data obtained from BOLD data base indicates that there may have been sampling bias, as higher abundance of Byrozoan species were found in North America (US-715, Can-583), in comparison to other continents of the world. The reason this is a sampling bias is because countries like Japan (356), New Zealand (224), and Australia (390) also have high bin quantities of the species. To further progress in this project, other countries, which showcase high abundance, should be further sampled and the surrounding countries to limit sampling bias and potentially advance in cancer treatment using Bryozoan species and its compounds.

Acknowledgments:

I would like to thank and acknowledge my classmates who have either helped or gave me either, which completing this assignment. I would like to give a special thank you to the TA, Brittany, for helping me throughout this assignment and give us encourage

References:

1. Ciavatta, M. L., Lefranc, F., Vieira, L. M., Kiss, R., Carbone, M., van Otterlo, W. A. L., Lopanik, N. B., & Waeschenbach, A. (2020). The Phylum Bryozoa: From Biology to Biomedical Potential. *Marine drugs*, *18*(4), 200. https://doi.org/10.3390/md18040200

2. Figuerola, B., & Avila, C. (2019). The Phylum Bryozoa as a Promising Source of Anticancer Drugs. *Marine drugs*, *17*(8), 477. https://doi.org/10.3390/md17080477