Анализ пиратских атак на морские судна

Импортируем библиотеки

```
library(readr)
library(ggmap)
```

```
## Loading required package: ggplot2
```

```
## The legacy packages maptools, rgdal, and rgeos, underpinning the sp package,
## which was just loaded, were retired in October 2023.
## Please refer to R-spatial evolution reports for details, especially
## https://r-spatial.org/r/2023/05/15/evolution4.html.
## It may be desirable to make the sf package available;
## package maintainers should consider adding sf to Suggests:.
```

```
## ℹ Google's Terms of Service: <https://mapsplatform.google.com>
## ℹ Please cite ggmap if you use it! Use `citation("ggmap")` for details.
```

```
library(ggplot2)
library(tidyr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'
##
## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```
library(data.table)
```

```
##
## Attaching package: 'data.table'
##
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
print_df <- function(df)
{
  df |>
    kable(format = "html") |>
    kable_styling() |>
    kableExtra::scroll_box(width = "100%", height = "100%")
}
```

```
df <- read.csv(
  file = "/Users/georgymilyuskhov/pirate_attacks.csv",stringsAsFactors=FALSE,fileEncoding="latin1"
)
```

Читаем данные

```
df |> head(10) |> print_df()
```

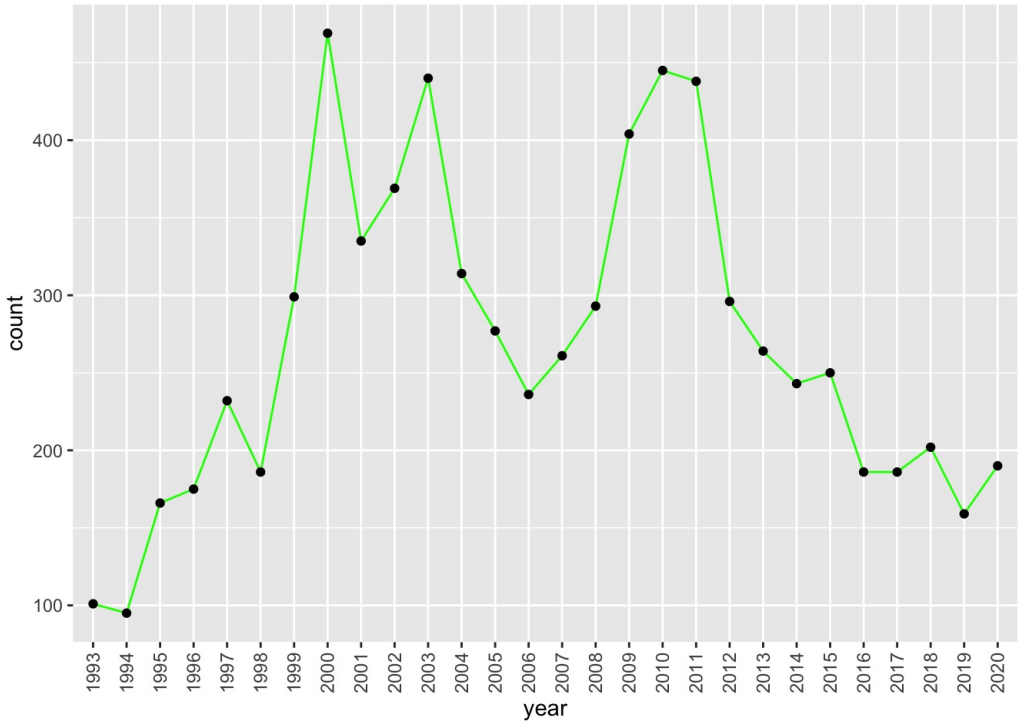
date	time	longitude	latitude	attack_type	location_description	nearest_country	eez_country	shore_distance	shore_
1993-01-02	NA	116.9667	19.700000	NA	Hong Kong - Luzon - Hainan	CHN	TWN	357.5023726	
1993-01-04	NA	116.0000	22.350000	NA	Hong Kong - Luzon - Hainan	CHN	CHN	47.4315725	
1993-01-06	NA	115.2500	19.670000	NA	Hong Kong - Luzon - Hainan	CHN	TWN	280.8118709	
1993-01-08	NA	124.5833	29.900000	NA	East China Sea	CHN	CHN	209.9233962	
1993-01-12	NA	120.2667	18.133333	NA	Hong Kong - Luzon - Hainan	PHL	PHL	22.0273321	
1993-01-13	NA	101.8500	9.717300	NA	Gulf of Thailand. South China Sea	KHM	KHM	184.7966407	
1993-01-25	NA	106.8667	8.566667	NA	South China Sea	VNM	VNM	27.4373438	
1993-01-26	NA	123.5000	29.000000	NA	East China Sea	CHN	CHN	141.5155643	
1993-01-26	NA	13.2500	-8.800000	NA	Luanda, Angola	AGO	AGO	0.5092157	
1993-01-26	NA	119.4000	4.433333	NA	South China Sea	PHL	PHL	25.3785876	

Проверяем когда атаки чаще всего происходили

```
df1<-df
df1$date<-format(as.Date(df1$date),"%Y")
```

```
year_count <- df1 %>%
  group_by(year=(df1$date)) %>%
  summarise(count = n()) %>%
arrange(desc(count))
```

```
ggplot(data=year_count, aes(x=year, y=count, group=1)) +
  geom_line(color="green")+
  geom_point()+ theme(axis.text.x = element_text(angle = 90, vjust = 0.5, size = 9))
```



Нужно отметить, что здесь не учитывается рост судоходства с 1993. Заметны новые методы по борьбе с пиратами в 2010.

Смотрим где данные атаки происходили

Рассматриваем ближайшую страну, а не исключительную экономическую зону.

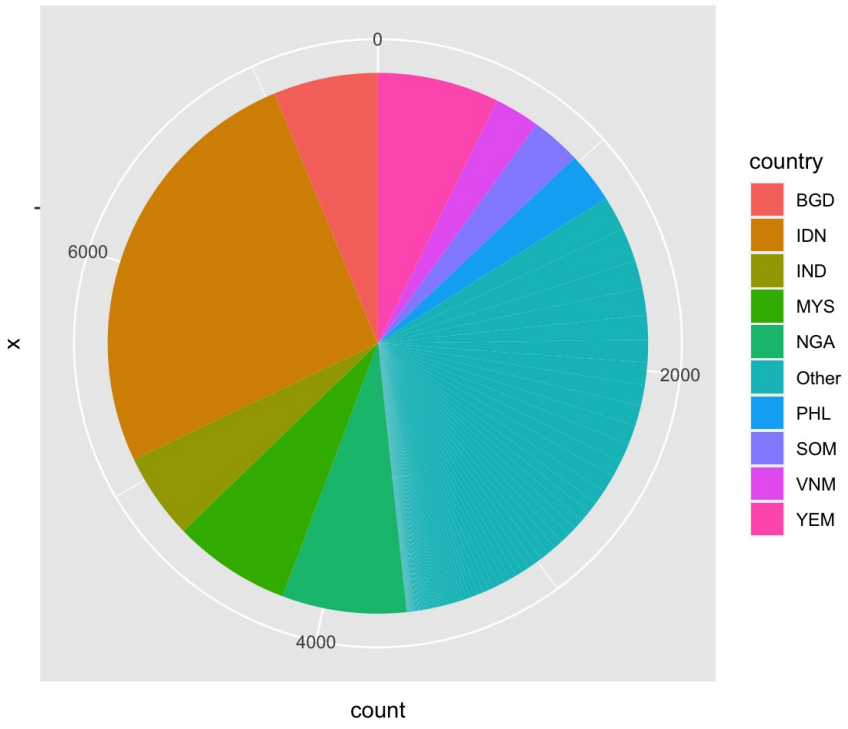
```
nearest_country_count <- df1%>%
  group_by(country=(df1$nearest_country)) %>%
  summarise(count = n()) %>%
  arrange(desc(count))
nearest_country_count |> head(15) |> print_df()
```

country	count
IDN	1939
NGA	559
YEM	546
MYS	529
BGD	470
IND	386
PHL	227
SOM	224
VNM	204
CHN	154
BRA	153
OMN	123
PER	121
TZA	110
GHA	101

```
plotting_data <- nearest_country_count %>%
  mutate(rank = rank(-count),
         country = ifelse(rank <= 9, country, 'Other'))
```

Нужно иметь ввиду количество островов в Индонезии.

```
ggplot(plotting_data, aes(x="", y=count, fill=country)) +
  geom_bar(stat="identity", width=1) +
  coord_polar("y", start=0)
```



Рассматриваем максимальную и минимальную дистанцию атак пиратов до берега.

```
df[which.max(df$shore_distance),]|> print_df()
```

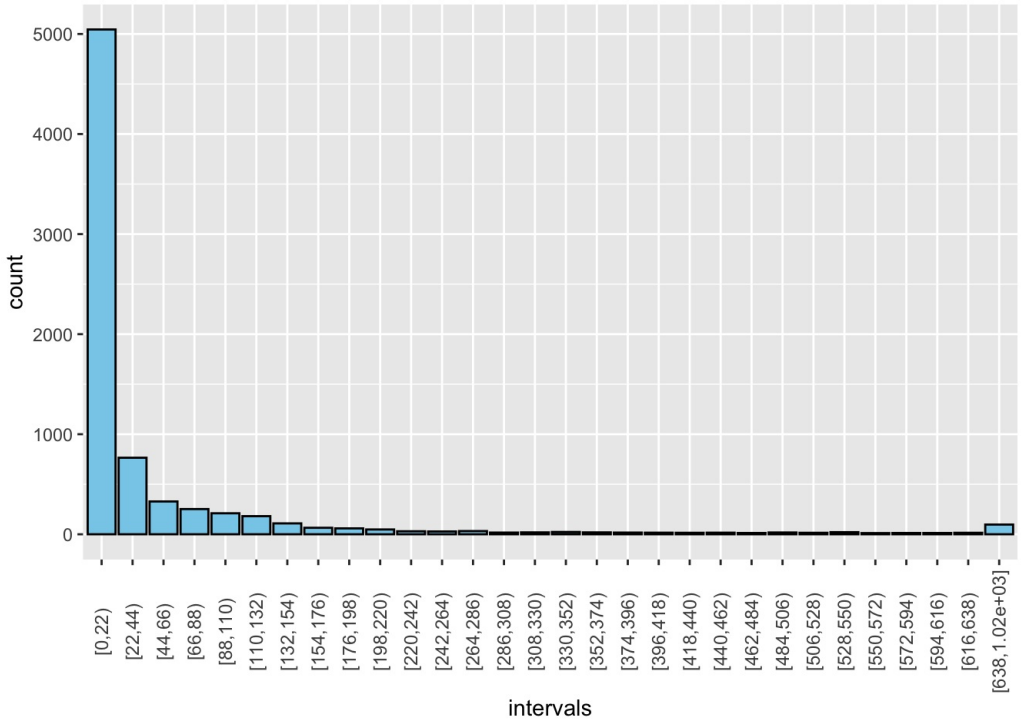
	date	time	longitude	latitude	attack_type	location_description	nearest_country	eez_country	shore_distance	sh
4949	2010-09-29	NA	61.85	6.7917	Attempted	(Indian Ocean), Off Somalia	YEM	NA	1024.029	

```
df[which.min(df$shore_distance),]|> print_df()
```

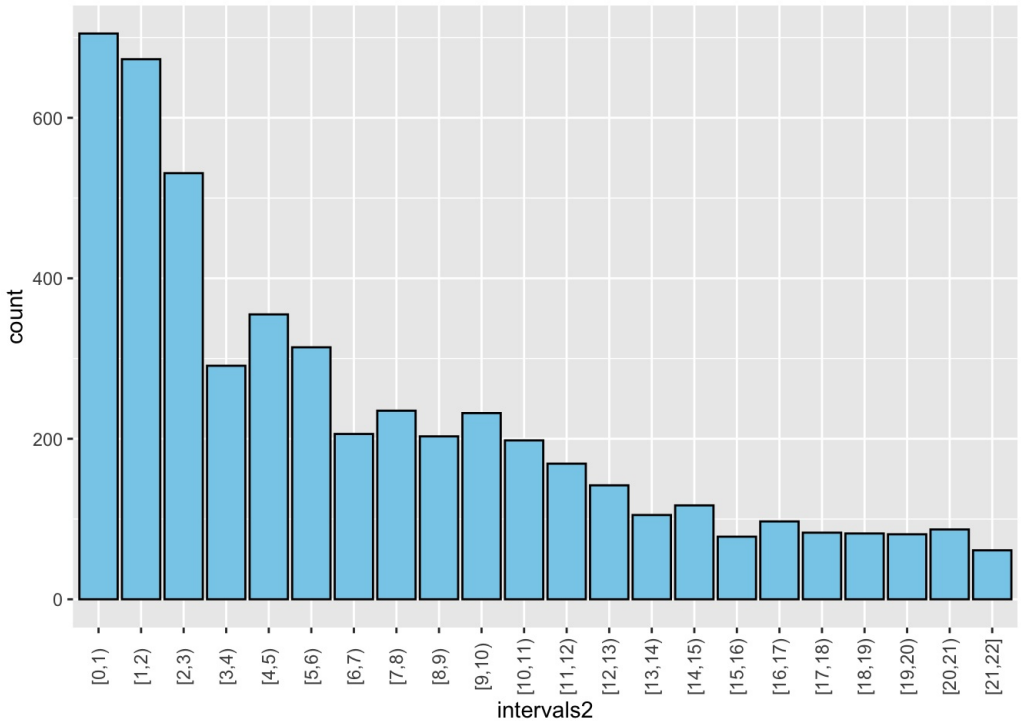
	date	time	longitude	latitude	attack_type	location_description	nearest_country	eez_country	shore_distance	sh
6040	2013-10-19	NA	-58.17	6.8133	Boarded	Georgetown, Guyana	GUY	GUY	0.0391893	

Построим гистограмму, чтобы посмотреть, как далеко от берега происходят атаки.

```
intervals <- c(0, 22, 44, 66, 88, 110, 132, 154, 176, 198, 220, 242, 264, 286, 308, 330, 352, 374, 396, 418, 440,
462, 484, 506, 528, 550, 572, 594, 616, 638, 1025)
df2<-df
df2$intervals<-cut(df2$shore_distance, breaks = intervals, include.lowest = TRUE, right = FALSE)
counts <- table(df2$intervals)

ggplot(df2, aes(x = intervals)) +
geom_bar(fill = "skyblue", color = "black")+ theme(axis.text.x = element_text(angle = 90, vjust = 0.5, size = 9))
```




```
step <- 1
breaks <- seq(0, 22, by = step)
df2$intervals2 <- cut(df2$shore_distance, breaks = breaks, include.lowest = TRUE, right = FALSE)
counts2 <- table(df2$intervals2)

ggplot(data=subset(df2, !is.na(intervals2)), aes(x = intervals2)) +
geom_bar(fill = "skyblue", color = "black")+ theme(axis.text.x = element_text(angle = 90, vjust = 0.5, size = 9))
```



Заметно, что атаки происходят наиболее часто в тереториальных водах страны.

Интерессно посмотреть, есть ли судна на которых нападали несколько раз.

```
ship_name <- df %>%
  group_by(vessel_name=df1$vessel_name) %>%
  summarise(count = n())%>%
  filter(!is.na(vessel_name))|>
  arrange(desc(count))
ship_name|>head(10)|>print_df()
```

vessel_name	count
Name Withheld	61
Unspecified	18
Agate	8
Maersk Alabama	8
Actuaria	7
Thor Falcon	6
Petrobulk Racer	5
Shamrock	5
Border	4
Darya Rani	4

Нужно иметь ввиду, что тут есть популярные название для пароходов.

Рассмотрим на карте где атаки происходили наиболее часто

```
world <- get_stamenmap(
    bbox = c(left = -180, bottom = -57, right = 179, top = 82.1),
    maptype = "toner",
    zoom = 3)
```

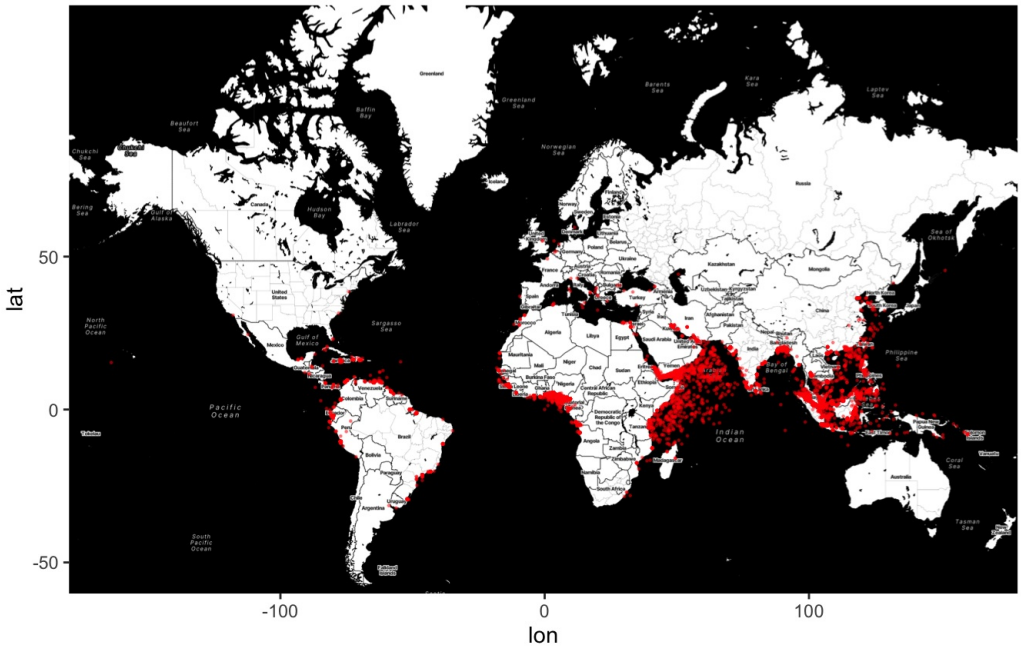
```
## i Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL.
```

```
## i 48 tiles needed, this may take a while (try a smaller zoom?)
```

```
ggmap(world) +
  geom_point(data = df,
             aes(x = df$longitude, y = df$latitude),
             alpha = .4,
             size = .2,
             color="red")
```

```
## Warning: Use of `df$longitude` is discouraged.
## i Use `longitude` instead.
```

```
## Warning: Use of `df$latitude` is discouraged.
## i Use `latitude` instead.
```



На какие типы пароходов было больше всего нападений.

```
ship_type <- df|>
  group_by(vessel_type)|>
  summarise(count=n())|>
  filter(!is.na(vessel_type))|>
  arrange(desc(count))

ship_type|>head(14)|>print_df()
```

vessel_type	count
Bulk Carrier	322
Product Tanker	252
Container	119
General Cargo	67
Chemical Tanker	53
Tug	45
Crude Oil Tanker	40
LPG Tanker	40
Tanker	30
Fishing Vessel	27
Crude Tanker	23
Refrigerated Cargo	16
Offshore Supply Ship	15
LNG Tanker	9

Логично, что большие пароходы не входят в топ.

Посмотрим в каком сотоянии они находились при нападении.

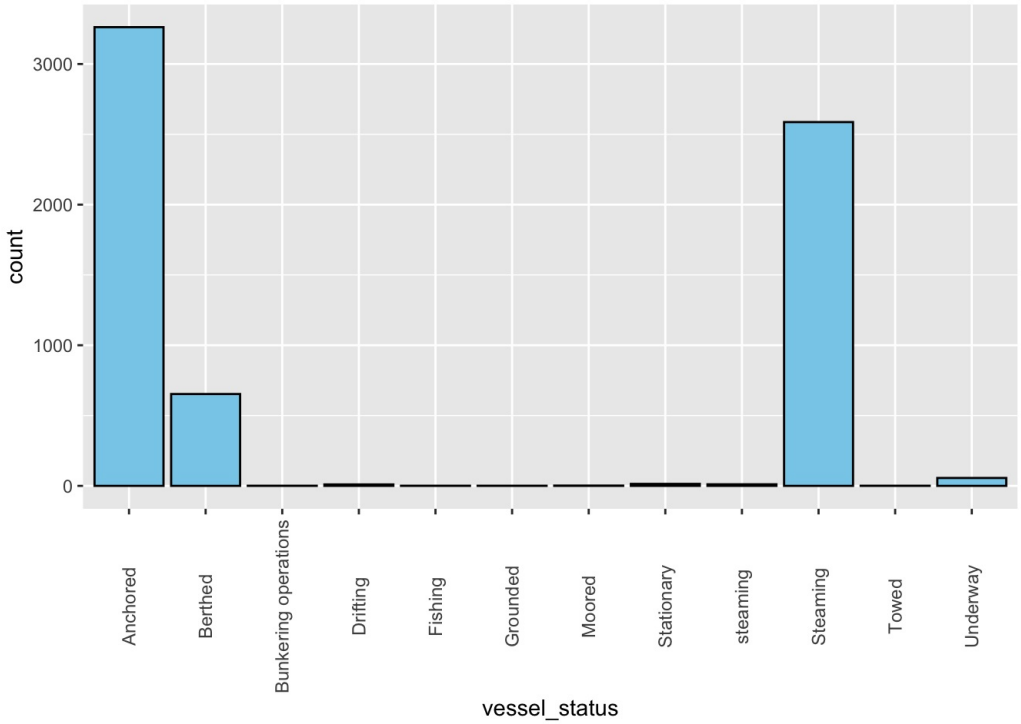
```
ship_status <- df|>
  group_by(vessel_status)|>
  summarise(count=n())|>
  filter(!is.na(vessel_status))|>
  arrange(desc(count))

ship_status|>print_df()
```

vessel_status	count
Anchored	3262
Steaming	2587
Berthed	653
Underway	56
Stationary	14
steaming	11
Drifting	10
Moored	2
Bunkering operations	1
Fishing	1
Grounded	1
Towed	1

```
ship_status <- df|>
  group_by(vessel_status)|>
  summarise(count=n())|>
  filter(!is.na(vessel_status))|>
  arrange(desc(count))
ship_status|>print_df()
```

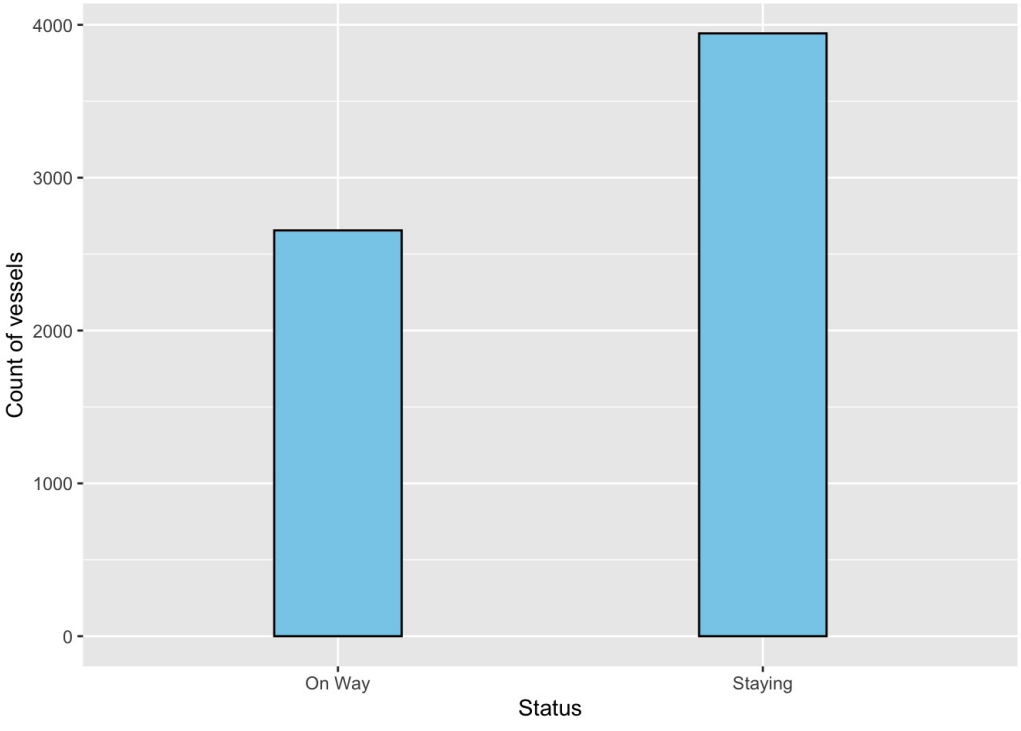
vessel_status	count
Anchored	3262
Steaming	2587
Berthed	653
Underway	56
Stationary	14
steaming	11
Drifting	10
Moored	2
Bunkering operations	1
Fishing	1
Grounded	1
Towed	1

```
ggplot(data=subset(df2, !is.na(vessel_status)), aes(x = vessel_status)) +
geom_bar(fill = "skyblue", color = "black")+ theme(axis.text.x = element_text(angle = 90, vjust = 0.5, size = 9))
```



Важно отметить, что это состояние парохода на момент получения данных.

```
OnWay <-ship_status$count[2]+ship_status$count[4]+ship_status$count[6]+ship_status$count[10]
Staying <- ship_status$count[1]+ship_status$count[3]+ship_status$count[5]+ship_status$count[7]+ship_status$count[8]+ship_status$count[9]+ship_status$count[11]+ship_status$count[12]


dat <- data.frame(
names = c("Staying", "On Way"),
values = c(Staying, OnWay)
)

ggplot(dat, aes(x = names, y = values)) +
geom_bar(stat = "identity", width = 0.3, fill="skyblue", color = "black") +
ylab("Count of vessels") +
xlab("Status")
```



Какого типа была атака?

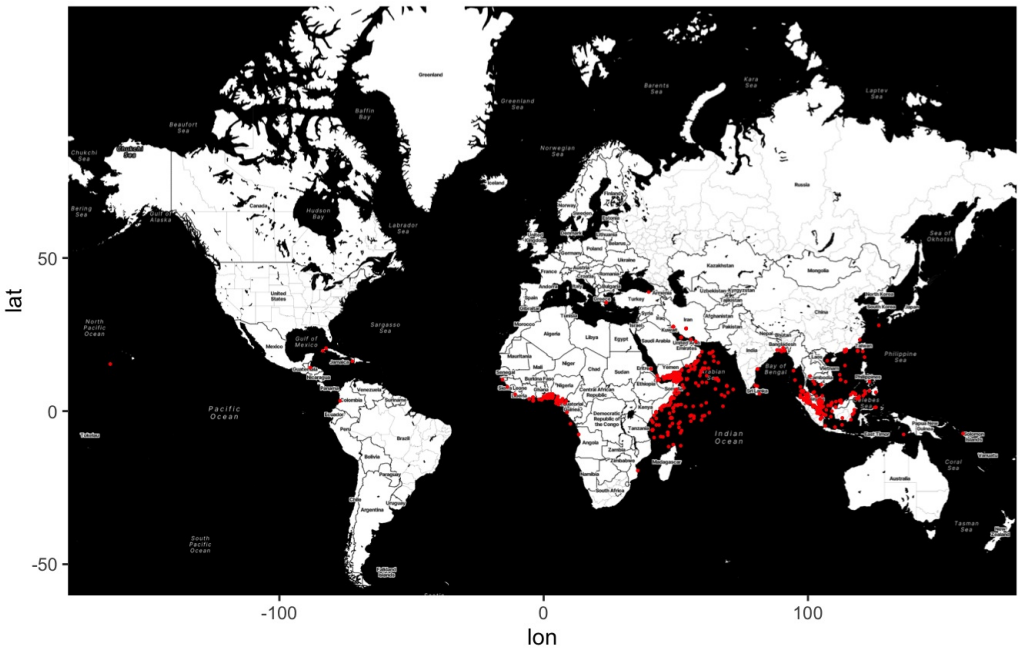
```
vessel_attacks <- df|>
  group_by(attack_type)|>
  summarise(count=n())|>
  filter(!is.na(attack_type))|>
  arrange(desc(count))

vessel_attacks|>print_df()
```

attack_type	count
Boarded	3421
Attempted	1999
Boarding	1367
Hijacked	511
Fired Upon	73
Suspicious	16
Explosion	3
Detained	1

Рассмотрим где пароход был наиболее часто угнан.

```
df3 <- subset(df, df$attack_type == "Hijacked")
```

```
ggmap(world) +
  geom_point(data = df3,
             aes(x = df3$longitude, y = df3$latitude),
             alpha = .9,
             size = .2,
             color="red")
```



```
ship_status2 <- df3|>
  group_by(vessel_status)|>
  summarise(count=n())|>
  filter(!is.na(vessel_status))|>
  arrange(desc(count))

OnWay1 <-ship_status$count[1]+ship_status$count[3]+ship_status$count[6]
Staying1 <- ship_status$count[2]+ship_status$count[4]+ship_status$count[5]+ship_status$count[7]

dat1 <- data.frame(
names = c("Staying", "On Way"),
values = c(Staying1, OnWay1)
)

ggplot(dat1, aes(x = names, y = values)) +
geom_bar(stat = "identity", width = 0.3, fill="skyblue", color = "black") +
ylab("Count of vessels") +
xlab("Status")
```