Group Project R Script and Outputs

Emma Beyer & Ayoung Kim

2024-04-13

```
#loading packages
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.5.0     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggplot2)
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
library(rmarkdown)

#setting WD
setwd("/home/guest/Stat_Modeling_Lab/ENV710Ayoung_Emma")

#read in original data
original_data <- read.csv("dolphins.csv")

#read in data with unneeded variables
strandings <- read.csv("dolphins_cleaned.csv")
#attach data
attach(strandings)
```

```
#remove unneeded covariates in cleaned data
cleaned_strandings <- subset(strandings, select =
                               -c(Shot, Fishery.Interaction, Boat.Collision, Weight))

#remove NAs from chosen variables
cleaned_strandings <- na.omit(cleaned_strandings)
```

```r
#removing blanks in Age.Class variable
cleaned_strandings <-
  cleaned_strandings[!grepl("^\\s*$", cleaned_strandings$Age.Class), ]
#removing unknowns in Age.Class
cleaned_strandings <-
  cleaned_strandings[!cleaned_strandings$Age.Class %in% c("UNKNOWN"), ]

#removing unknowns in Sex variable
cleaned_strandings <-
  cleaned_strandings[!cleaned_strandings$Sex %in% c("UNKNOWN"), ]

#create binary covariate for states with and without offshore wind
cleaned_strandings <- cleaned_strandings %>%
  mutate(turbine_presence = if_else(State %in% c("VA", "NY", "RI", "MA"), 1, 0))

#create subset of only states with offshore wind
turbine_data <-
  cleaned_strandings[cleaned_strandings$State %in%
                             c("VA", "NY", "RI", "MA"), ]

#use mutate to create state binary variables for fit_4
turbine_data <- turbine_data %>%
  mutate(
    VA = ifelse(State == "VA", 1, 0),
    NY = ifelse(State == "NY", 1, 0),
    RI = ifelse(State == "RI", 1, 0),
    MA = ifelse(State == "MA", 1, 0))
```

```r
#count of strandings in wind farm states = 80
sum(cleaned_strandings$turbine_presence)
```

```
## [1] 80
```

```r
#count of total number of stranding = 1419
nrow(cleaned_strandings)
```

```
## [1] 1419
```

```r
#summary statistics of all variables
summary_stats <- summary(cleaned_strandings)
summary_stats
```

```
##     State           Year.of.Observation     Sex              Age.Class
##  Length:1419        Min.   :2017        Length:1419        Length:1419
##  Class :character   1st Qu.:2017        Class :character   Class :character
##  Mode  :character   Median :2018        Mode  :character   Mode  :character
##                     Mean   :2018
##                     3rd Qu.:2019
##                     Max.   :2019
##      Length      turbine_presence
##  Min.   :  0.0   Min.   :0.00000
```

```
##   1st Qu.:134.0   1st Qu.:0.00000
##   Median :209.0   Median :0.00000
##   Mean   :194.2   Mean   :0.05638
##   3rd Qu.:248.0   3rd Qu.:0.00000
##   Max.   :366.0   Max.   :1.00000
```

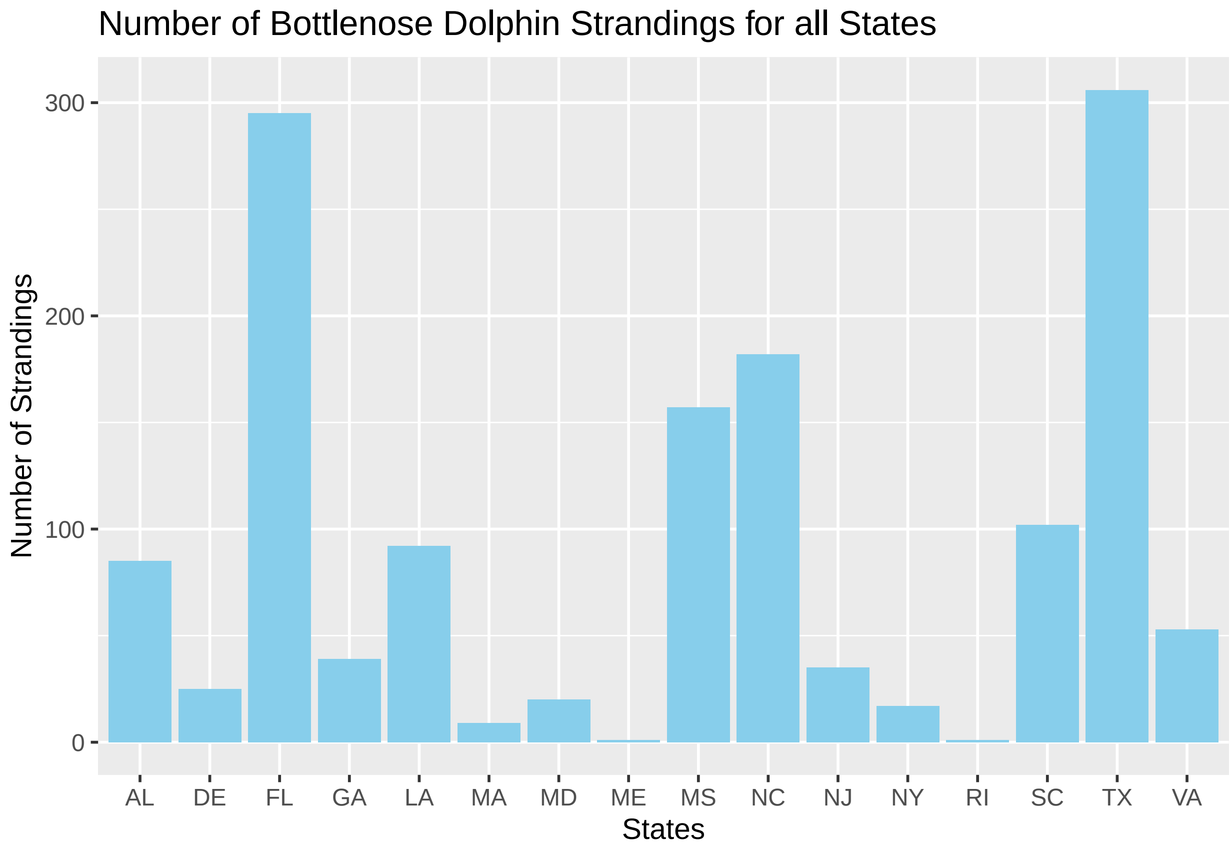
```r
#reordering age classes so they are from youngest to oldest
age_class_order <- c("PUP/CALF", "YEARLING", "SUBADULT", "ADULT")

#States info
table(cleaned_strandings$State)
```

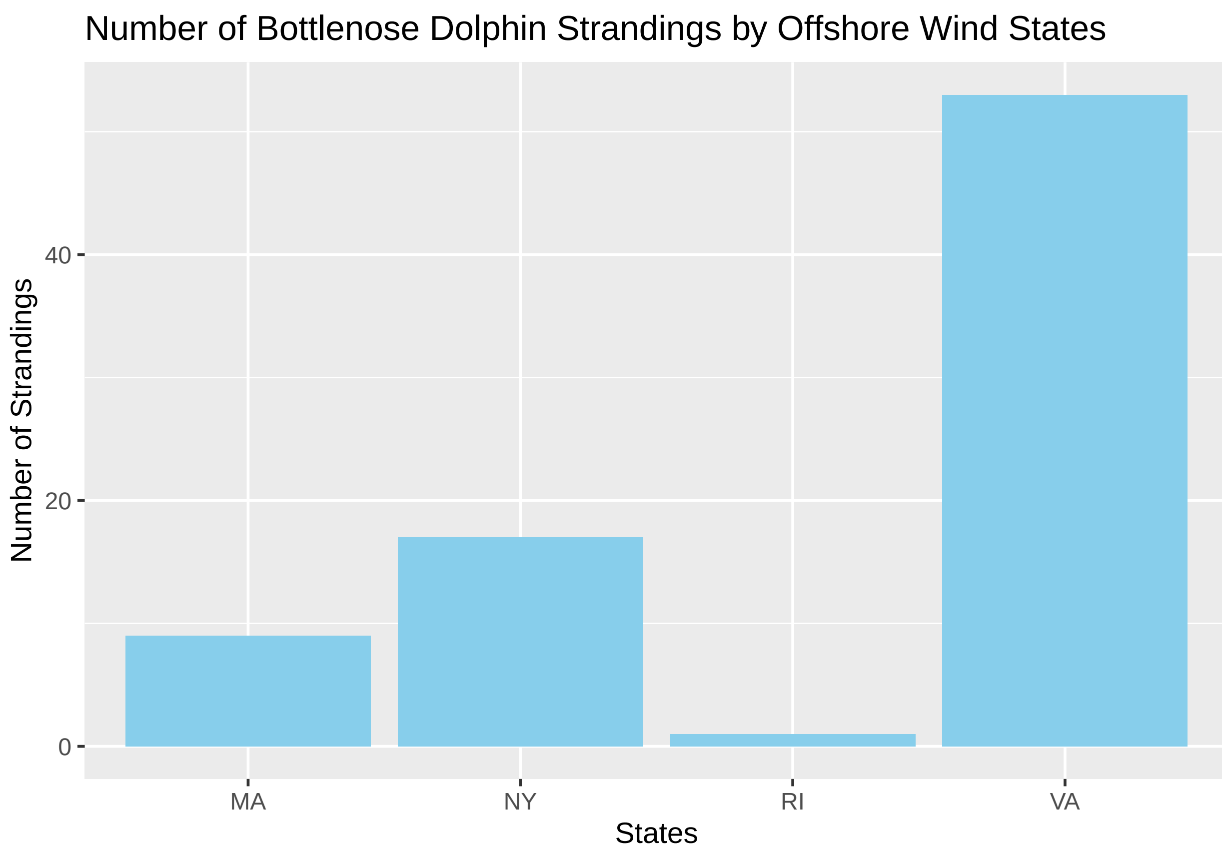
```
##
##  AL  DE  FL  GA  LA  MA  MD  ME  MS  NC  NJ  NY  RI  SC  TX  VA
##  85  25 295  39  92   9  20   1 157 182  35  17   1 102 306  53
```

```r
# 15 states included
# Most strandings were in Florida (295) and Texas (306)
# Least were in Maine (1) and Rhode Island (1)
# States with offshore wind: Virginia, New York, Rhode Island, and Massachusetts

#plot of number of strandings for all states
standings_state <- ggplot(cleaned_strandings, aes(x = State)) +
  geom_bar(fill = "skyblue") +
  labs(x = "States", y = "Number of Strandings") +
  ggtitle("Number of Bottlenose Dolphin Strandings for all States")
standings_state
```




```r
#plot of number of strandings for each offshore wind state
standings_wf_state <- ggplot(turbine_data, aes(x = State)) +
  geom_bar(fill = "skyblue") +
  labs(x = "States", y = "Number of Strandings") +
  ggtitle("Number of Bottlenose Dolphin Strandings by Offshore Wind States")
standings_wf_state
```




```
#Length info
# average overall length 194.2091
mean(cleaned_strandings$Length)
```

```
## [1] 194.2091
```

```
# average offshore wind length = 198.175
mean(turbine_data$Length)
```

```
## [1] 198.175
```

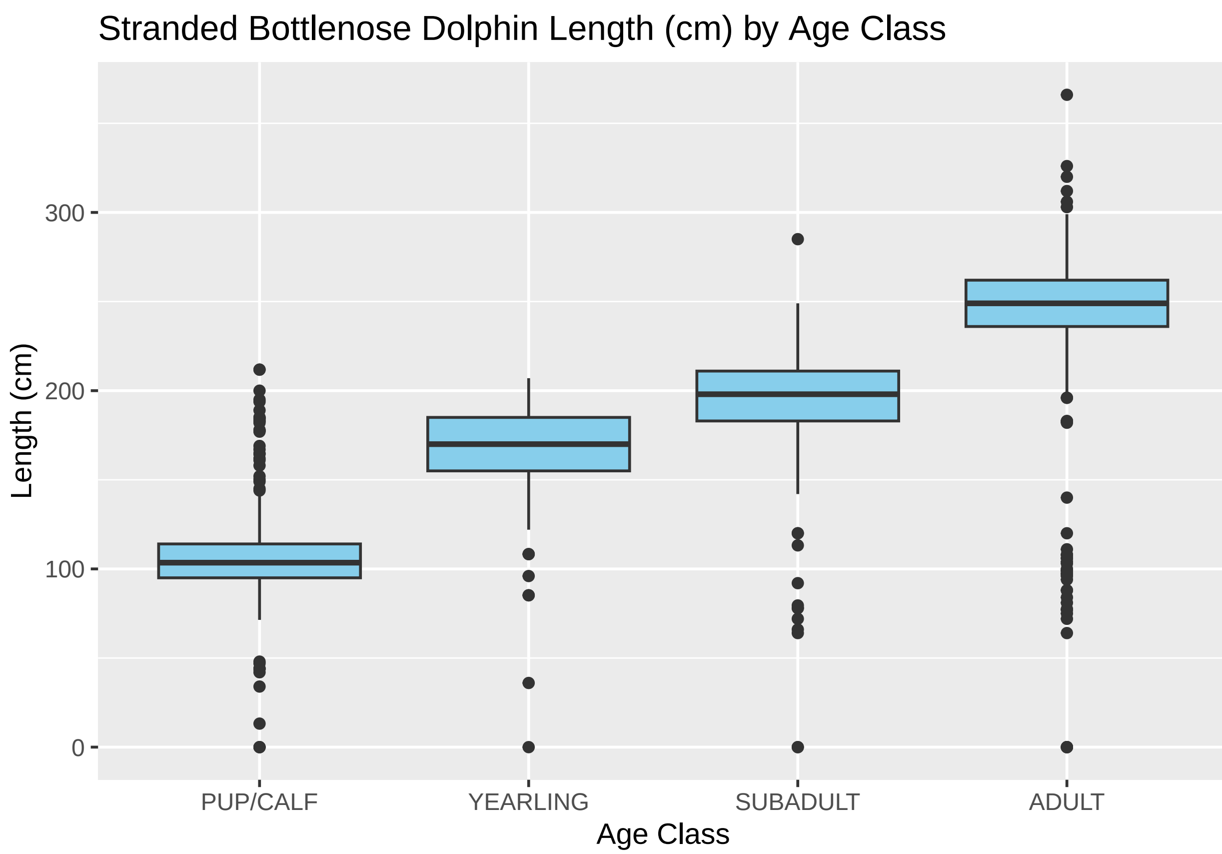
```
# variance of all lengths = 4270.962 so 65.37 cm deviation from the mean
variance_all <- var(cleaned_strandings$Length)
variance_all
```

```
## [1] 4270.962
```

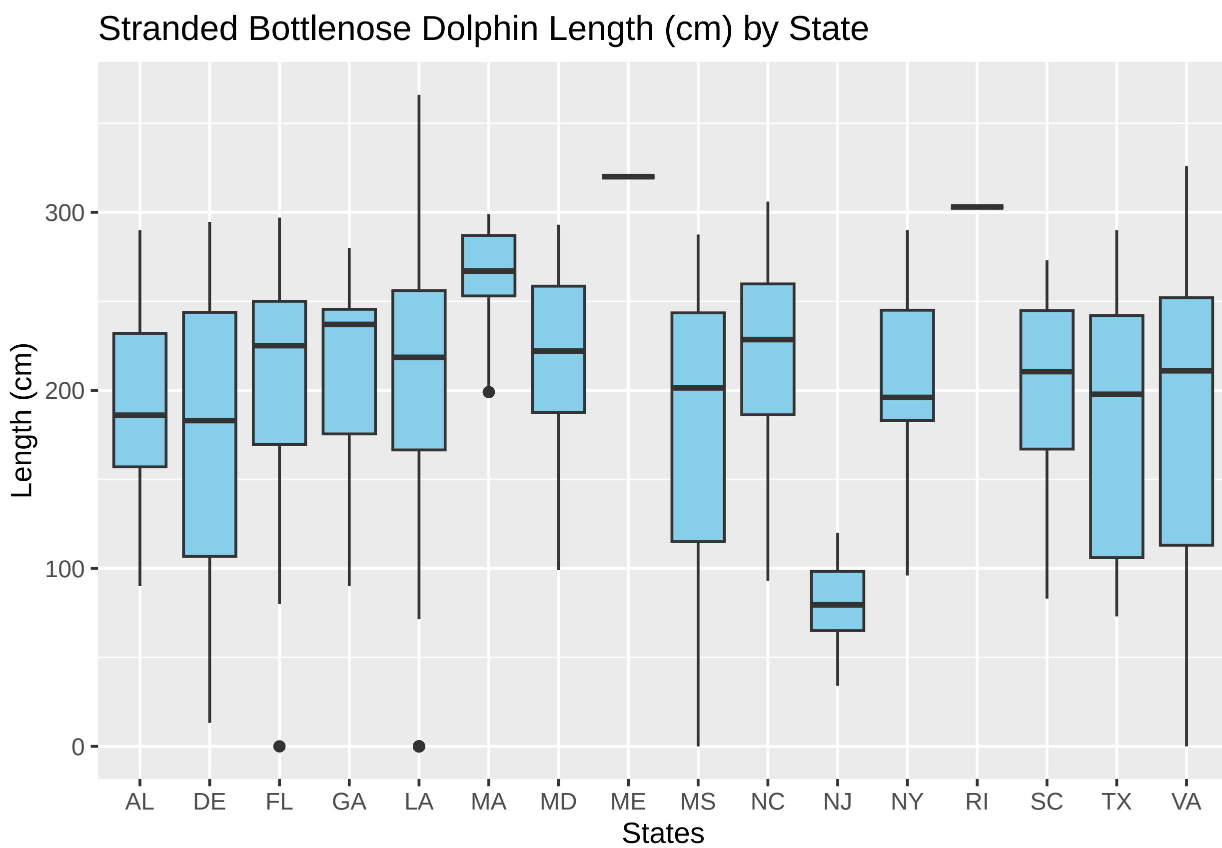
```
# variance of offshore wind lengths = 6133.309 so 78.36 cm deviation from the mean
variance_osw <- var(turbine_data$Length)
variance_osw
```

```
## [1] 6133.309
```

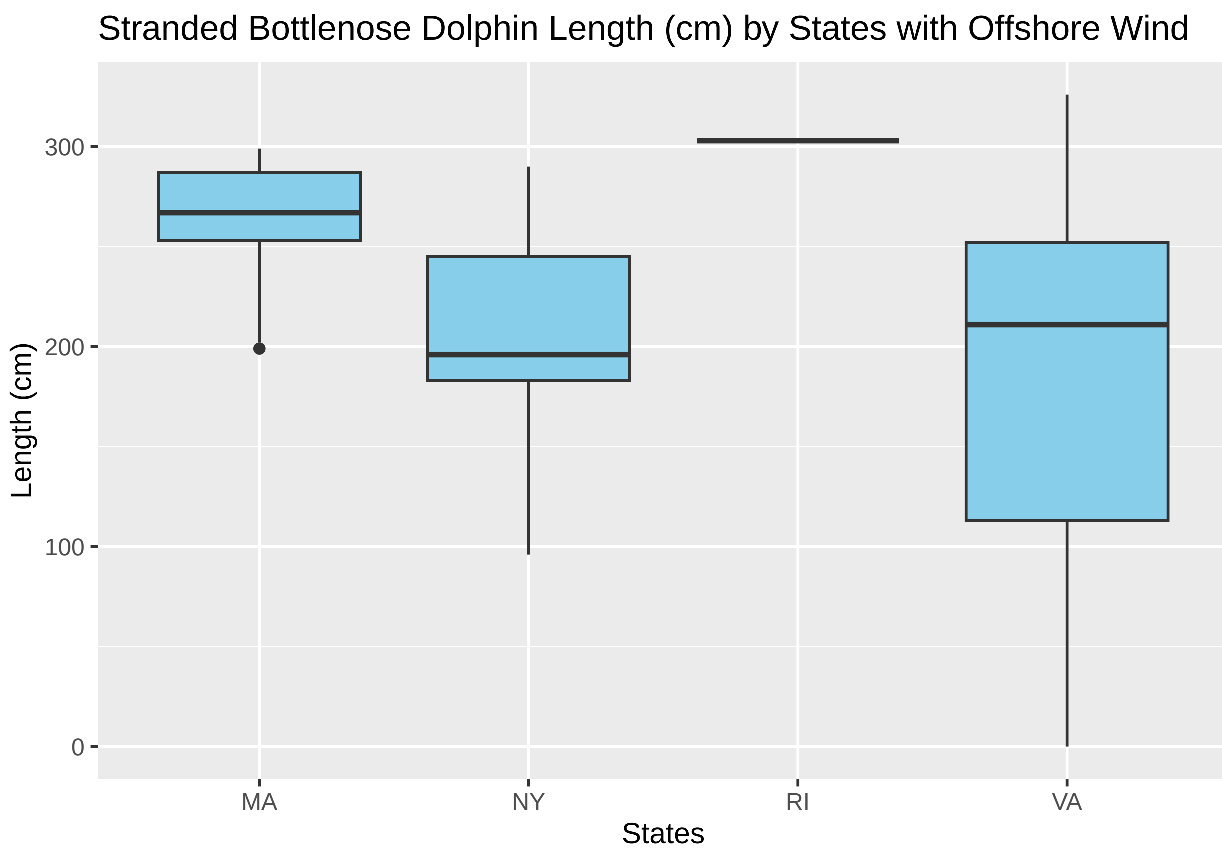
```r
#summary graph of Age Class by Length
length_ageclass <- ggplot(cleaned_strandings,
                          aes(x = factor(Age.Class, levels = age_class_order),
                              y = Length)) +
  geom_boxplot(fill="skyblue") +
  labs(x = "Age Class", y = "Length (cm)") +
  ggtitle("Stranded Bottlenose Dolphin Length (cm) by Age Class")
length_ageclass
```




```r
#plot of Lengths in each state
state_length <- ggplot(cleaned_strandings, aes(x = State, y = Length)) +
  geom_boxplot(fill="skyblue") +
  labs(x = "States", y = "Length (cm)") +
  ggtitle("Stranded Bottlenose Dolphin Length (cm) by State")
state_length
```




```r
#plot of Lengths in each offshore wind state
turbinestate_length <- ggplot(turbine_data, aes(x = State, y = Length)) +
  geom_boxplot(fill="skyblue") +
  labs(x = "States", y = "Length (cm)") +
  ggtitle("Stranded Bottlenose Dolphin Length (cm) by States with Offshore Wind")
turbinestate_length
```




```r
#Age Class Info
#finding ave length per age class: PUP/CALF 107.1709, YEARLING 163.3439, SUBADULT   194.0060, ADULT 242
ave_length_ageclass <-
  aggregate(Length ~ Age.Class, data = cleaned_strandings, FUN = mean)
ave_length_ageclass
```

```
##   Age.Class   Length
## 1     ADULT 242.8233
## 2  PUP/CALF 107.1709
## 3  SUBADULT 194.0060
## 4  YEARLING 163.3439
```

```r
#count of strandings in all states by age class ADULT   669, PUP/CALF   335, SUBADULT   308, YEARLING
table(cleaned_strandings$Age.Class)
```

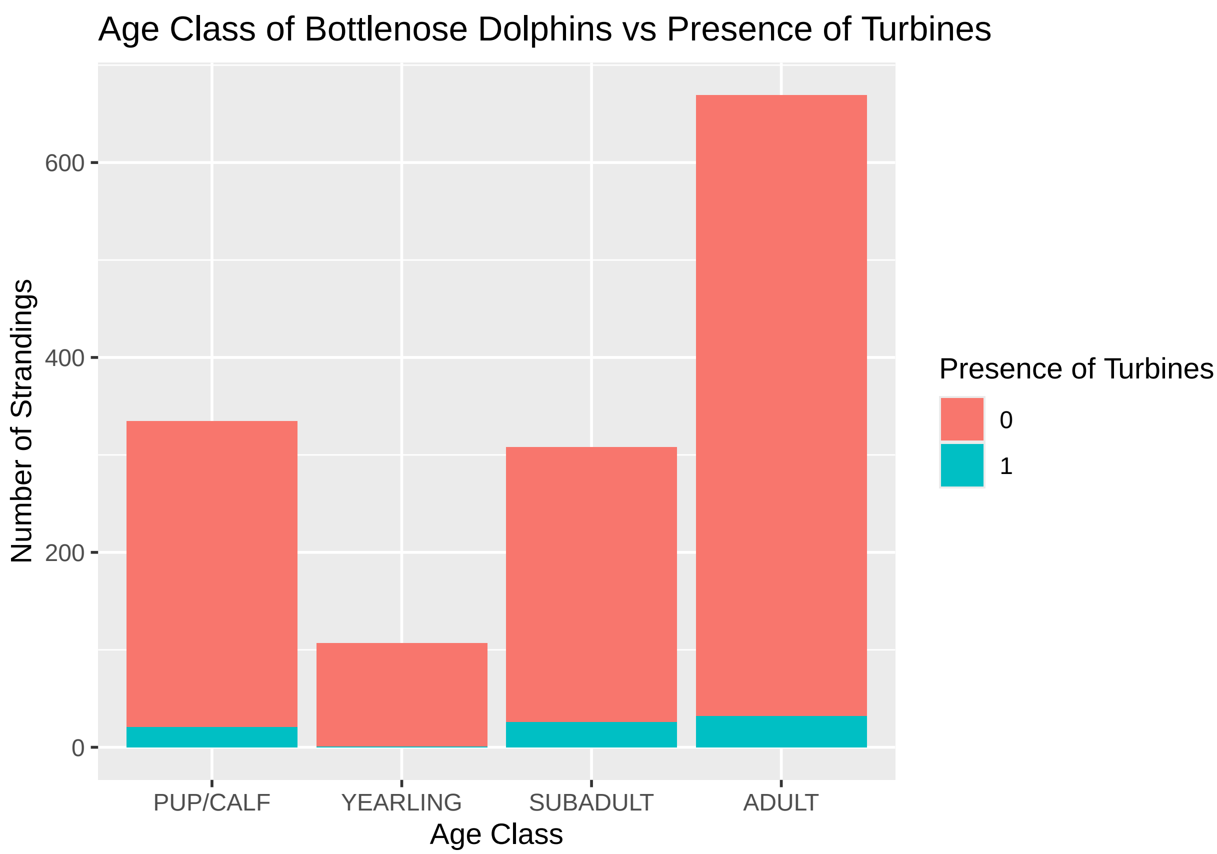
```
##
##    ADULT PUP/CALF SUBADULT YEARLING
##      669      335      308      107
```

```r
#count of strandings in offshore wind states by age class ADULT 32, PUP/CALF    21, SUBADULT    26, YEA
count_ageclass <- aggregate(turbine_presence ~ Age.Class, data = cleaned_strandings, FUN = sum)
count_ageclass
```

```
##   Age.Class turbine_presence
```

```
## 1     ADULT                32
## 2  PUP/CALF                21
## 3  SUBADULT                26
## 4  YEARLING                 1
```

```
#plot of the number of strandings in each age class
strandings_ageclass <- ggplot(cleaned_strandings,
                              aes(x = factor(Age.Class, levels = age_class_order),
                                  fill = factor(turbine_presence))) +
  geom_bar(position = "stack") +
  labs(x = "Age Class", y = "Number of Strandings", fill = "Presence of Turbines") +
  ggtitle("Age Class of Bottlenose Dolphins vs Presence of Turbines")
strandings_ageclass
```




```
#Sex Info
#finding ave length per sex: FEMALE 191.8816, MALE  195.8511
ave_length_sex <-
  aggregate(Length ~ Sex, data = cleaned_strandings, FUN = mean)
ave_length_sex
```

```
##      Sex   Length
## 1 FEMALE 191.8816
## 2   MALE 195.8511
```

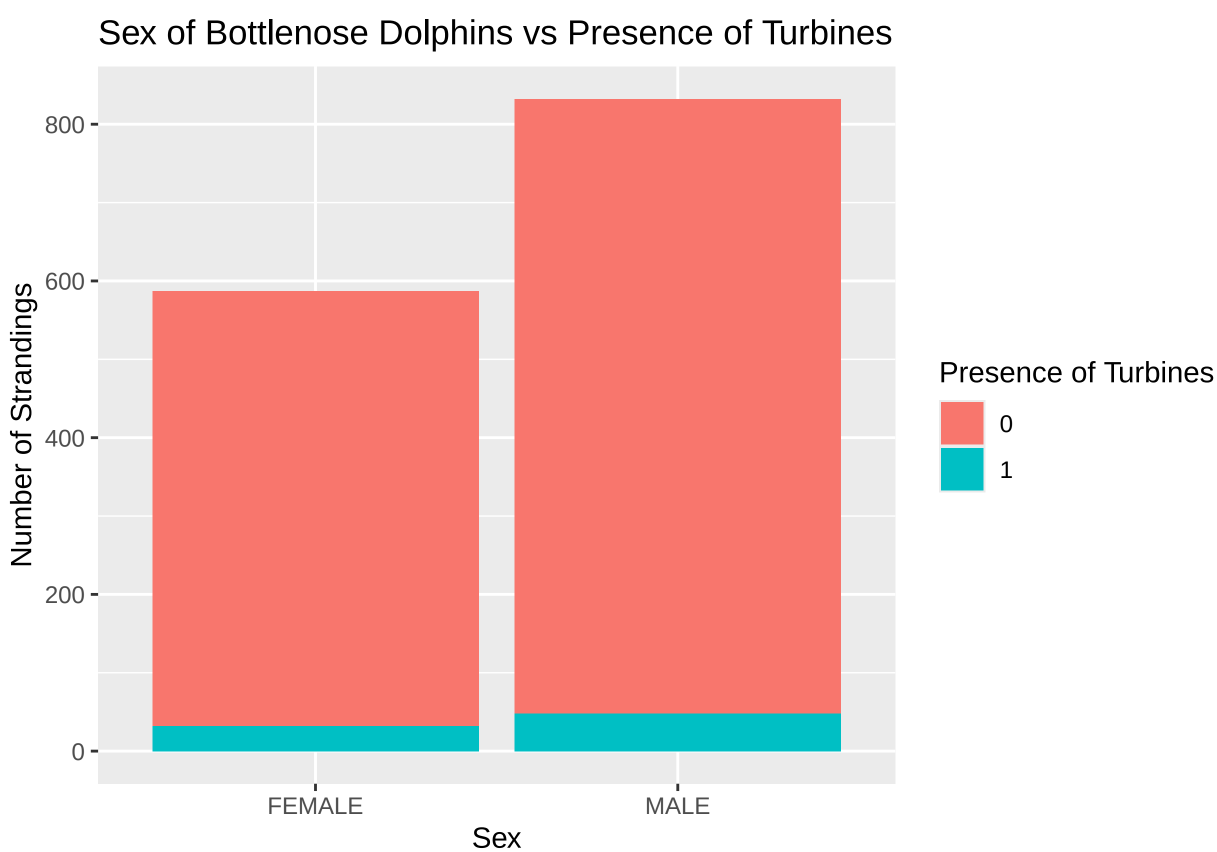
```r
#count of strandings in all states by sex FEMALE 587, MALE 832
table(cleaned_strandings$Sex)
```

```
##
## FEMALE   MALE
##    587    832
```

```r
#count of strandings in offshore states by sex FEMALE   32, MALE    48
count_sex <-
  aggregate(turbine_presence ~ Sex, data = cleaned_strandings, FUN = sum)
count_sex
```

```
##      Sex turbine_presence
## 1 FEMALE               32
## 2   MALE               48
```

```r
#plot of the number of strandings in each sex
strandings_sex <- ggplot(cleaned_strandings,
                         aes(x = factor(Sex), fill = factor(turbine_presence))) +
  geom_bar(position = "stack") +
  labs(x = "Sex", y = "Number of Strandings", fill = "Presence of Turbines") +
  ggtitle("Sex of Bottlenose Dolphins vs Presence of Turbines")
strandings_sex
```



```r
#Fitting a regression model (Turbine Presence)
fit_1 <- glm(turbine_presence~1, data = cleaned_strandings)
#Summary of the regression model fit_1
summary(fit_1)
```

```
##
## Call:
## glm(formula = turbine_presence ~ 1, data = cleaned_strandings)
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.056378   0.006125   9.204   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.0532368)
##
##     Null deviance: 75.49  on 1418  degrees of freedom
## Residual deviance: 75.49  on 1418  degrees of freedom
## AIC: -131.99
##
## Number of Fisher Scoring iterations: 2
```

```r
#Fitting a regression model
fit_2 <- glm(turbine_presence~Age.Class,family='binomial', data = cleaned_strandings)
#Summary of the regression model fit_2
summary(fit_2)
```

```
##
## Call:
## glm(formula = turbine_presence ~ Age.Class, family = "binomial",
##     data = cleaned_strandings)
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -2.9910     0.1812 -16.510   <2e-16 ***
## Age.ClassPUP/CALF   0.2862     0.2892   0.990   0.3224
## Age.ClassSUBADULT   0.6072     0.2735   2.220   0.0264 *
## Age.ClassYEARLING  -1.6724     1.0209  -1.638   0.1014
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 615.51  on 1418  degrees of freedom
## Residual deviance: 603.61  on 1415  degrees of freedom
## AIC: 611.61
##
## Number of Fisher Scoring iterations: 7
```

```r
#Calculations of odds (probability) of each age class
exp(-2.99)
```

```
## [1] 0.05028744
```

```r
#calculations of odd of each age class
exp(-2.9910+0.2862)
```

```
## [1] 0.0668837
```

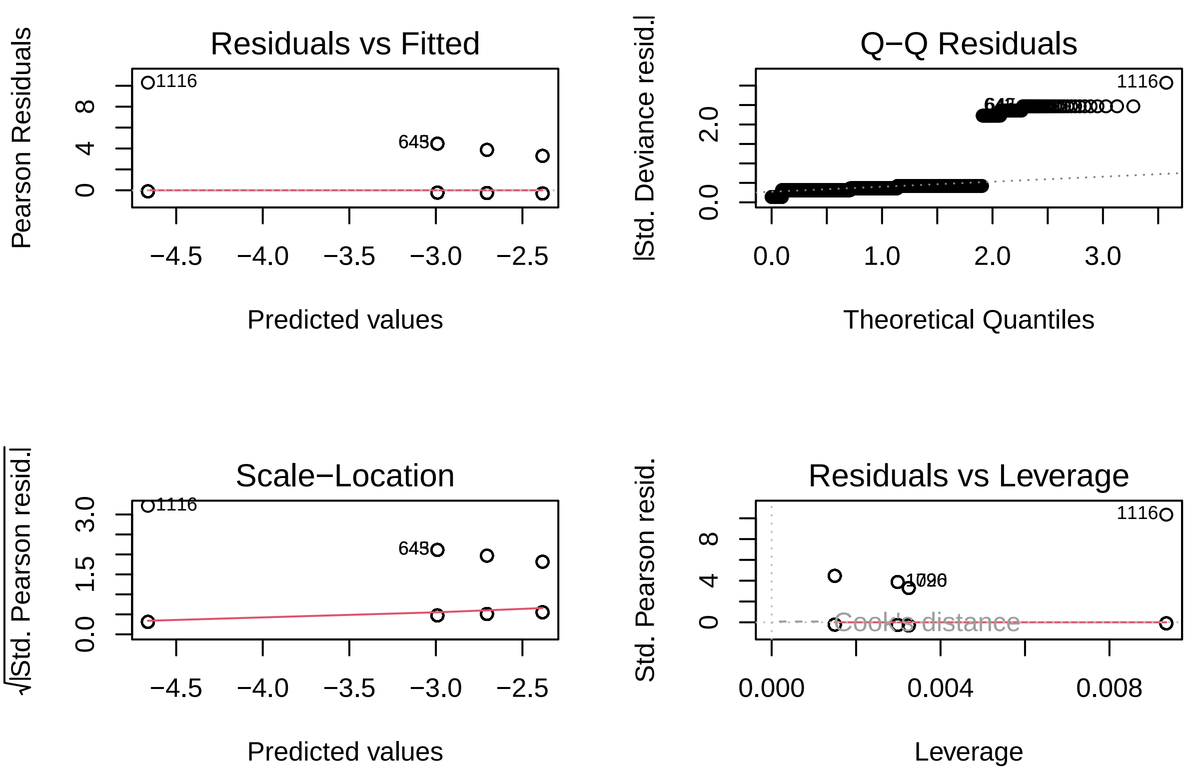
```r
exp(-2.9910+0.6072)
```

```
## [1] 0.09219955
```

```r
exp(-2.9910-1.6724)
```

```
## [1] 0.009434331
```

```r
#qq plots for fit_2
par(mfrow=c(2,2))
plot(fit_2)
```

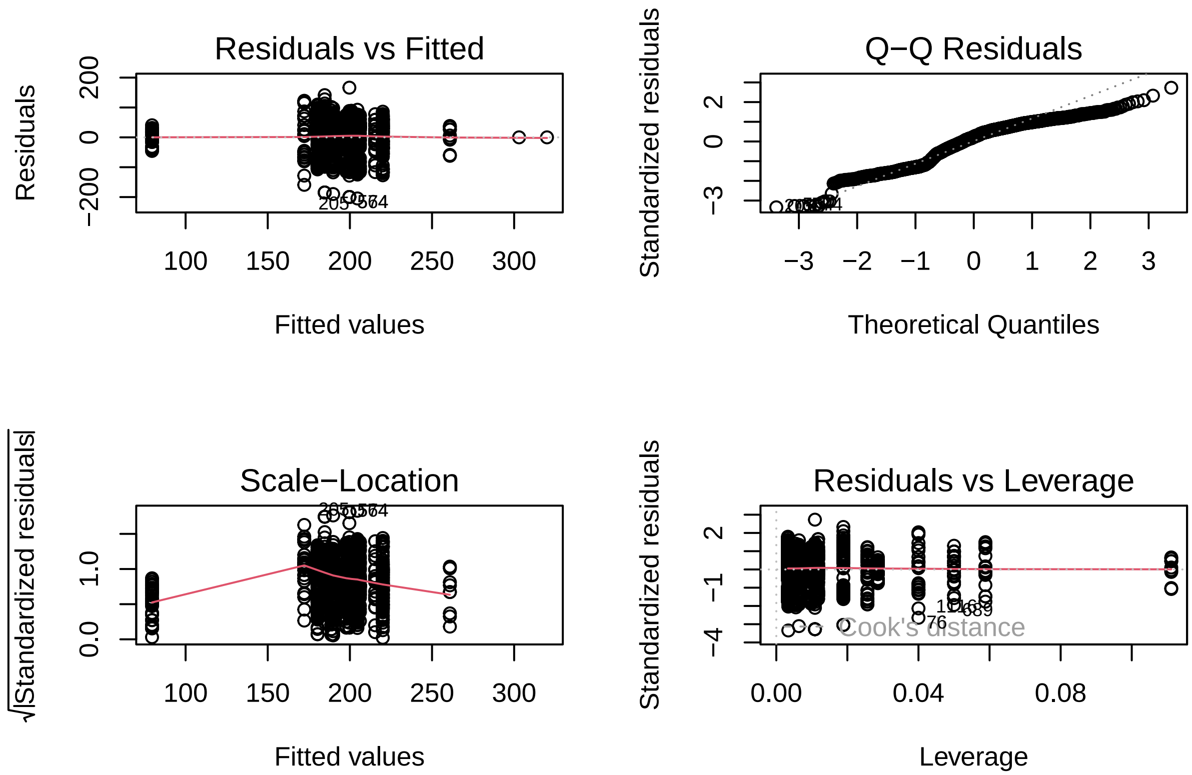



```r
# use to find difference in length in states
fit_3 <- lm(Length ~ State, data = cleaned_strandings)
#summary of linear regression
summary(fit_3)
```

```
##
## Call:
## lm(formula = Length ~ State, data = cleaned_strandings)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -204.59  -45.68   13.90   48.64  166.30
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  188.747      6.650  28.381  < 2e-16 ***
## StateDE      -16.521     13.950  -1.184 0.236508
## StateFL       15.847      7.548   2.099 0.035952 *
## StateGA       17.356     11.858   1.464 0.143541
## StateLA       10.951      9.225   1.187 0.235375
## StateMA       72.142     21.493   3.357 0.000810 ***
## StateMD       26.638     15.238   1.748 0.080663 .
## StateME      131.253     61.674   2.128 0.033496 *
## StateMS        1.064      8.257   0.129 0.897506
## StateNC       31.281      8.055   3.883 0.000108 ***
## StateNJ     -109.191     12.314  -8.867  < 2e-16 ***
## StateNY       12.159     16.290   0.746 0.455558
## StateRI      114.253     61.674   1.853 0.064159 .
## StateSC        9.081      9.005   1.009 0.313384
## StateTX       -8.427      7.518  -1.121 0.262504
## StateVA       -4.075     10.731  -0.380 0.704178
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 61.31 on 1403 degrees of freedom
## Multiple R-squared:  0.1291, Adjusted R-squared:  0.1198
## F-statistic: 13.86 on 15 and 1403 DF,  p-value: < 2.2e-16
```

```r
#set up a multi-panel plot layout
par(mfrow = c(2, 2))
plot(fit_3)
```

```
## Warning: not plotting observations with leverage one:
##   369, 670
```

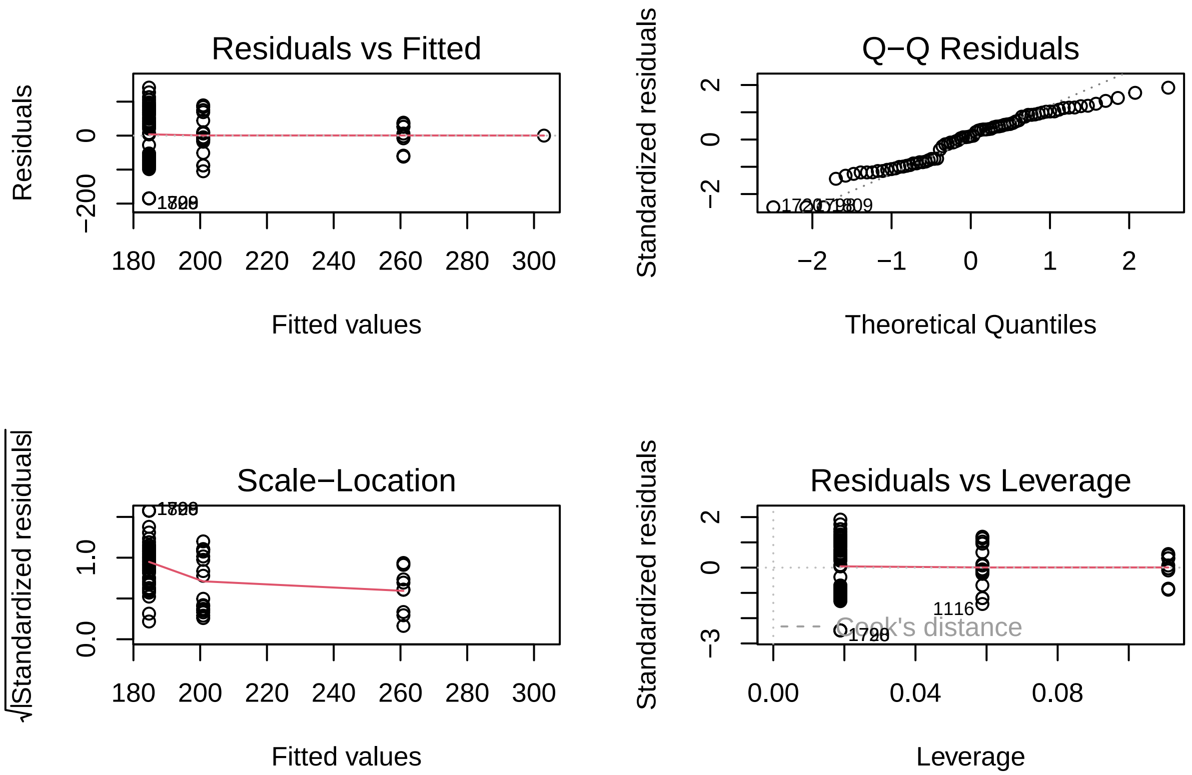



```
#finding significance of dolphin length across offshore wind states
fit_4 <- lm(Length ~ VA + NY + RI + MA, data = turbine_data)
#summary of linear regression
summary(fit_4)
```

```
##
## Call:
## lm(formula = Length ~ VA + NY + RI + MA, data = turbine_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -184.672  -61.876    8.211   62.328  141.328
##
## Coefficients: (1 not defined because of singularities)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   260.89      25.02  10.425 2.62e-16 ***
## VA            -76.22      27.07  -2.816  0.00619 **
## NY            -59.98      30.95  -1.938  0.05632 .
## RI             42.11      79.14   0.532  0.59618
## MA                NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 75.07 on 76 degrees of freedom
## Multiple R-squared:  0.1159, Adjusted R-squared:  0.08104
## F-statistic: 3.322 on 3 and 76 DF,  p-value: 0.02413
```

```r
#set up a multi-panel plot layout
par(mfrow = c(2, 2))
plot(fit_4)
```

```
## Warning: not plotting observations with leverage one:
##   27
```




```r
#finding the mean of length within VA
mean_length_VA <- mean(turbine_data$Length[turbine_data$VA == "1"])
mean_length_VA
```

```
## [1] 184.6717
```

```r
#finding the mean of length within VA
mean_length_NY <- mean(turbine_data$Length[turbine_data$NY == "1"])
mean_length_NY
```

```
## [1] 200.9059
```

```r
#finding the mean of length within VA
mean_length_RI <- mean(turbine_data$Length[turbine_data$RI == "1"])
mean_length_RI
```

```
## [1] 303
```

```r
#finding the mean of length within VA
mean_length_MA <- mean(turbine_data$Length[turbine_data$MA == "1"])
mean_length_MA
```

```
## [1] 260.8889
```

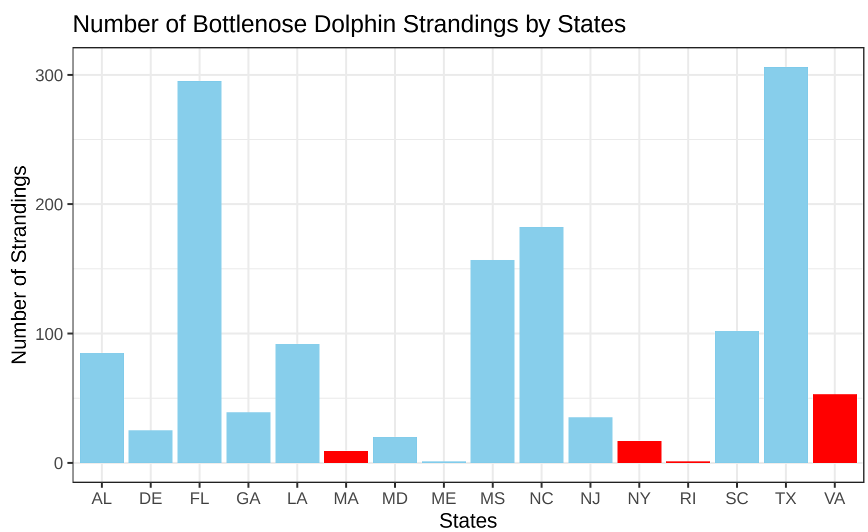


Figure 1: The number of stranded dolphins reported in each state across the East Coast. States with active offshore wind projects are labeled in red.

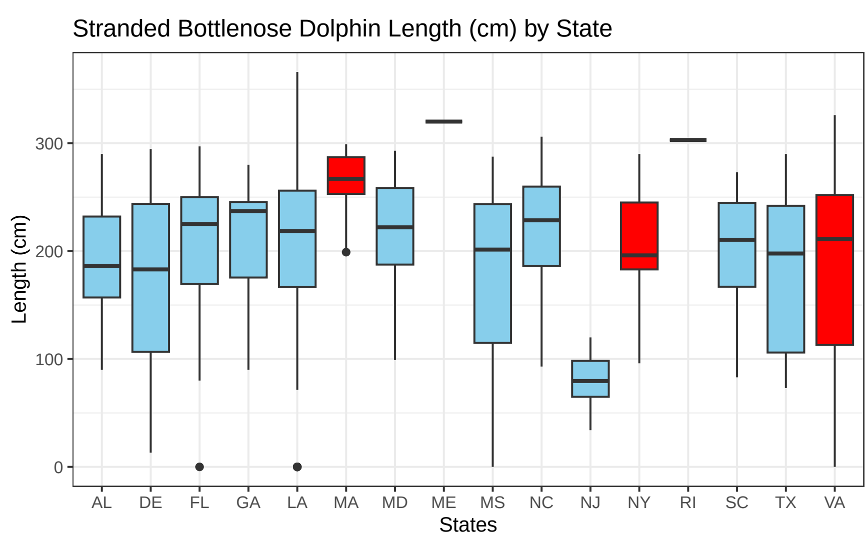


Figure 2: The distribution of length (cm) of stranded dolphins reported in each state across the East Coast. States with active offshore wind projects are labeled in red.

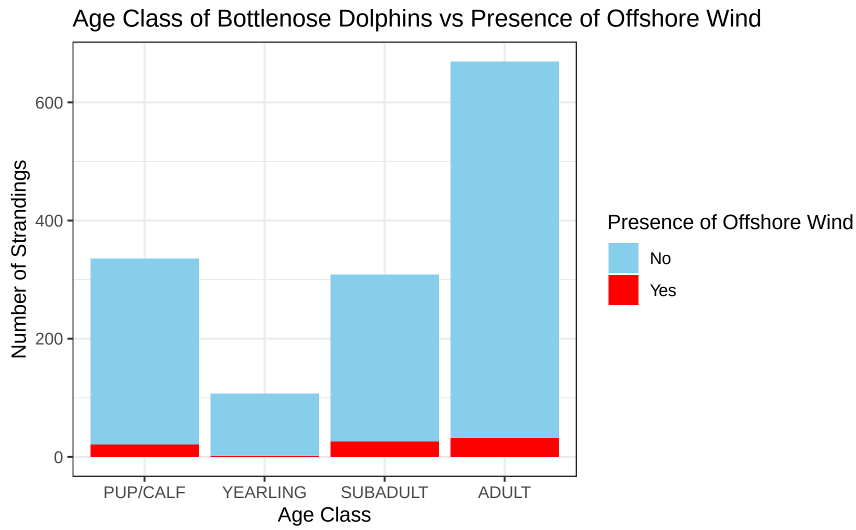


Figure 3: The number of strandings for each age class of bottlenose dolphin and whether or not there is a presence of offshore wind.

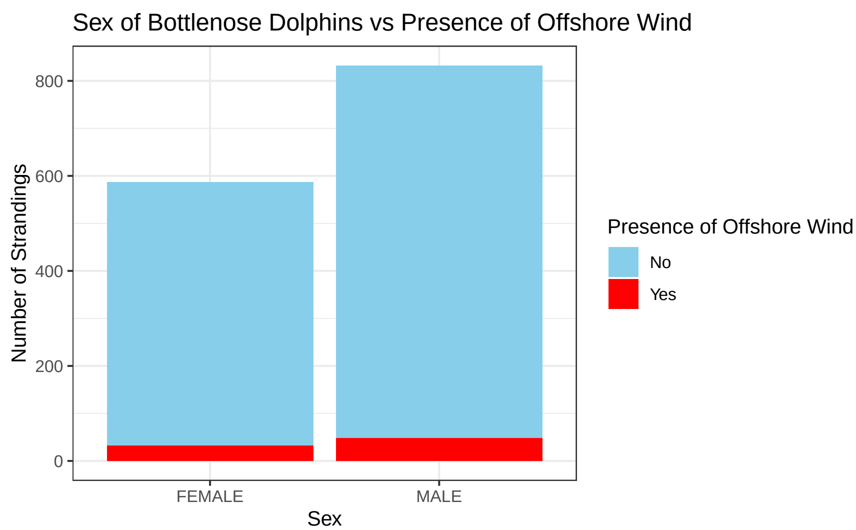


Figure 4: The number of strandings for each sex and whether or not there is a presence of offshore wind.

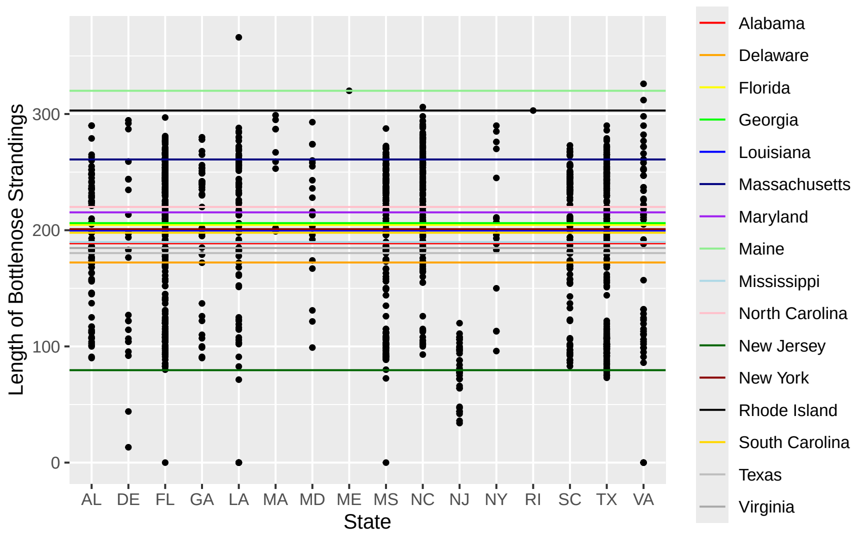


Figure 5: The distribution of lengths (cm) of stranded bottlenose dolphins across all states.

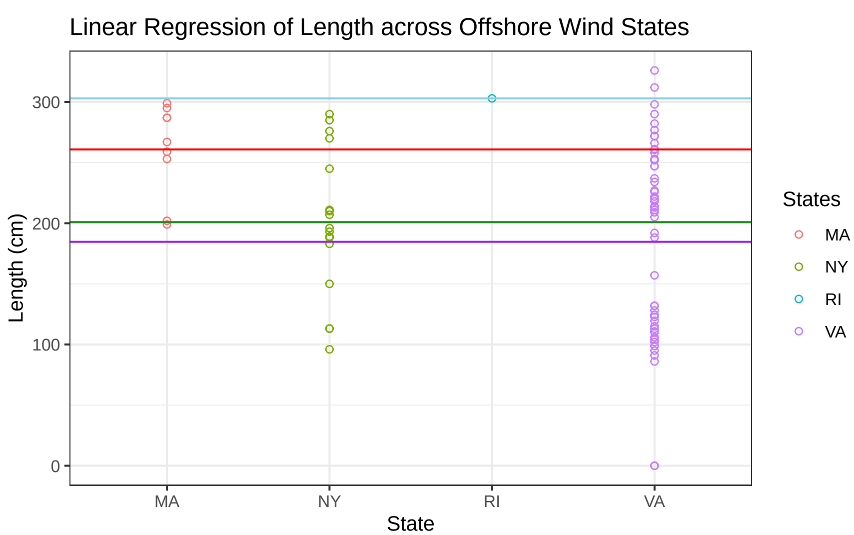


Figure 6: Linear regression of the lengths (cm) of stranded dolphin across states with offshore wind projects.