DataDeficiency\_analysis

Alex

8/29/2019

### This is an R Markdown file that will show the annotated analysis of our data on factors contributing to data deficiency in all known shark species.

library(xtable); library(MASS); library(tidyverse);library(lme4);library(sjPlot)

## Warning: package 'xtable' was built under R version 3.5.2

## Warning: package 'MASS' was built under R version 3.5.2

## ── Attaching packages ────────────────────────────────────────────────────── tidyverse 1.2.1 ──

## ✔ ggplot2 3.2.0 ✔ purrr 0.2.4  
## ✔ tibble 1.4.2 ✔ dplyr 0.7.8  
## ✔ tidyr 0.8.0 ✔ stringr 1.3.1  
## ✔ readr 1.1.1 ✔ forcats 0.3.0

## Warning: package 'ggplot2' was built under R version 3.5.2

## ── Conflicts ───────────────────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ✖ dplyr::select() masks MASS::select()

## Loading required package: Matrix

##   
## Attaching package: 'Matrix'

## The following object is masked from 'package:tidyr':  
##   
## expand

## Warning in checkMatrixPackageVersion(): Package version inconsistency detected.  
## TMB was built with Matrix version 1.2.15  
## Current Matrix version is 1.2.14  
## Please re-install 'TMB' from source using install.packages('TMB', type = 'source') or ask CRAN for a binary version of 'TMB' matching CRAN's 'Matrix' package

# Read in data  
dd <- read.csv("DD\_analysis\_datafile\_4.2.csv", stringsAsFactors = FALSE)  
dd = dd[dd$Order != "",]  
head(dd)

## Order Family Genus species  
## 2 Hexanchiformes Chlamydoselachidae Chlamydoselachus africana  
## 4 Hexanchiformes Chlamydoselachidae Chlamydoselachus anguineus  
## 6 Hexanchiformes Hexanchidae Heptranchias perlo  
## 8 Hexanchiformes Hexanchidae Hexanchus griseus  
## 10 Hexanchiformes Hexanchidae Hexanchus nakamurai  
## 12 Hexanchiformes Hexanchidae Notorynchus cepedianus  
## Common.name Data.Deficient Deepwater Coastal Pelagic  
## 2 southern african frilled shark 1 0 1 0  
## 4 frilled shark 0 1 0 0  
## 6 sharpnose sevengill shark 0 1 0 0  
## 8 bluntnose sixgill shark 0 1 0 0  
## 10 bigeye sixgill shark 1 1 0 0  
## 12 broadnose sevengill shark 1 1 0 0  
## Benthic BrackishFreshwater Tropical Temperate Pacific Atlantic Indian  
## 2 1 0 0 1 1 0 0  
## 4 1 0 1 1 1 1 1  
## 6 0 0 1 1 1 1 1  
## 8 0 0 1 1 1 1 1  
## 10 1 0 1 1 1 1 1  
## 12 0 0 1 1 1 1 0  
## Arctic Antartic Depth X0.100 X101.200 X201.300 X301.400 X401.500  
## 2 0 0 300-1400 0 0 1 1 1  
## 4 0 0 20-1500 1 1 1 1 1  
## 6 0 0 27-720 1 1 1 1 1  
## 8 0 0 200-1100 0 1 1 1 1  
## 10 0 0 90-621 1 1 1 1 1  
## 12 0 0 0-570 1 0 0 0 0  
## X501.600 X601.700 X701.800 X801.900 X901.1000 X1001.1100 X1101.1200  
## 2 1 1 1 1 1 1 1  
## 4 1 1 1 1 1 1 1  
## 6 1 1 1 0 0 0 0  
## 8 1 1 1 1 1 1 0  
## 10 1 1 0 0 0 0 0  
## 12 0 0 0 0 0 0 0  
## X1201.1300 X1301.1400 X1401.1500 X1501.1600 X1601.1700 X1701.1800  
## 2 1 1 0 0 0 0  
## 4 1 1 1 0 0 0  
## 6 0 0 0 0 0 0  
## 8 0 0 0 0 0 0  
## 10 0 0 0 0 0 0  
## 12 0 0 0 0 0 0  
## X1801.1900 X1901.2000 X2001.2100 X2101.2200 X2201.2300 X2301.2400  
## 2 0 0 0 0 0 0  
## 4 0 0 0 0 0 0  
## 6 0 0 0 0 0 0  
## 8 0 0 0 0 0 0  
## 10 0 0 0 0 0 0  
## 12 0 0 0 0 0 0  
## X2401.2500 X2501.2600 X2601.2700 X2701.2800 X2801.2900 X2901.3000  
## 2 0 0 0 0 0 0  
## 4 0 0 0 0 0 0  
## 6 0 0 0 0 0 0  
## 8 0 0 0 0 0 0  
## 10 0 0 0 0 0 0  
## 12 0 0 0 0 0 0  
## X3001.3100 X3101.3200 X3201.3300 X3301.3400 X3401.3500 X3501.3600  
## 2 0 0 0 0 0 0  
## 4 0 0 0 0 0 0  
## 6 0 0 0 0 0 0  
## 8 0 0 0 0 0 0  
## 10 0 0 0 0 0 0  
## 12 0 0 0 0 0 0  
## X3601.3700 X3701.3800 X3801.3900 X3901.4000 X4001.4100 X4101.4200  
## 2 0 0 0 0 0 0  
## 4 0 0 0 0 0 0  
## 6 0 0 0 0 0 0  
## 8 0 0 0 0 0 0  
## 10 0 0 0 0 0 0  
## 12 0 0 0 0 0 0  
## X4201.4300 X4301.4400 X4401.4500 X0.200 X201.1000 X1000. Epipelagic  
## 2 0 0 0 0 1 1 0  
## 4 0 0 0 1 1 1 1  
## 6 0 0 0 1 1 0 1  
## 8 0 0 0 0 1 1 0  
## 10 0 0 0 1 1 0 1  
## 12 0 0 0 1 0 0 1  
## Mesopelagic Bathypelagic Abyssopelagic Fisheries Gamefish Aquarium Size  
## 2 1 1 0 NA NA NA 117  
## 4 1 1 0 1 NA NA 196  
## 6 1 0 0 1 NA NA 139  
## 8 1 1 0 1 1 NA 482  
## 10 1 0 0 1 NA NA 180  
## 12 0 0 0 1 1 1 296  
## X.Reproductive.Strategy  
## 2   
## 4 Ovoviviparous  
## 6 Ovoviviparous  
## 8 Ovoviviparous  
## 10 Ovoviviparous  
## 12 Ovoviviparous

nrow(dd)

## [1] 501

# Set NAs to 0  
dd[is.na(dd)] = 0  
table(dd$Data.Deficient)

##   
## 0 1   
## 313 188

head(dd)

## Order Family Genus species  
## 2 Hexanchiformes Chlamydoselachidae Chlamydoselachus africana  
## 4 Hexanchiformes Chlamydoselachidae Chlamydoselachus anguineus  
## 6 Hexanchiformes Hexanchidae Heptranchias perlo  
## 8 Hexanchiformes Hexanchidae Hexanchus griseus  
## 10 Hexanchiformes Hexanchidae Hexanchus nakamurai  
## 12 Hexanchiformes Hexanchidae Notorynchus cepedianus  
## Common.name Data.Deficient Deepwater Coastal Pelagic  
## 2 southern african frilled shark 1 0 1 0  
## 4 frilled shark 0 1 0 0  
## 6 sharpnose sevengill shark 0 1 0 0  
## 8 bluntnose sixgill shark 0 1 0 0  
## 10 bigeye sixgill shark 1 1 0 0  
## 12 broadnose sevengill shark 1 1 0 0  
## Benthic BrackishFreshwater Tropical Temperate Pacific Atlantic Indian  
## 2 1 0 0 1 1 0 0  
## 4 1 0 1 1 1 1 1  
## 6 0 0 1 1 1 1 1  
## 8 0 0 1 1 1 1 1  
## 10 1 0 1 1 1 1 1  
## 12 0 0 1 1 1 1 0  
## Arctic Antartic Depth X0.100 X101.200 X201.300 X301.400 X401.500  
## 2 0 0 300-1400 0 0 1 1 1  
## 4 0 0 20-1500 1 1 1 1 1  
## 6 0 0 27-720 1 1 1 1 1  
## 8 0 0 200-1100 0 1 1 1 1  
## 10 0 0 90-621 1 1 1 1 1  
## 12 0 0 0-570 1 0 0 0 0  
## X501.600 X601.700 X701.800 X801.900 X901.1000 X1001.1100 X1101.1200  
## 2 1 1 1 1 1 1 1  
## 4 1 1 1 1 1 1 1  
## 6 1 1 1 0 0 0 0  
## 8 1 1 1 1 1 1 0  
## 10 1 1 0 0 0 0 0  
## 12 0 0 0 0 0 0 0  
## X1201.1300 X1301.1400 X1401.1500 X1501.1600 X1601.1700 X1701.1800  
## 2 1 1 0 0 0 0  
## 4 1 1 1 0 0 0  
## 6 0 0 0 0 0 0  
## 8 0 0 0 0 0 0  
## 10 0 0 0 0 0 0  
## 12 0 0 0 0 0 0  
## X1801.1900 X1901.2000 X2001.2100 X2101.2200 X2201.2300 X2301.2400  
## 2 0 0 0 0 0 0  
## 4 0 0 0 0 0 0  
## 6 0 0 0 0 0 0  
## 8 0 0 0 0 0 0  
## 10 0 0 0 0 0 0  
## 12 0 0 0 0 0 0  
## X2401.2500 X2501.2600 X2601.2700 X2701.2800 X2801.2900 X2901.3000  
## 2 0 0 0 0 0 0  
## 4 0 0 0 0 0 0  
## 6 0 0 0 0 0 0  
## 8 0 0 0 0 0 0  
## 10 0 0 0 0 0 0  
## 12 0 0 0 0 0 0  
## X3001.3100 X3101.3200 X3201.3300 X3301.3400 X3401.3500 X3501.3600  
## 2 0 0 0 0 0 0  
## 4 0 0 0 0 0 0  
## 6 0 0 0 0 0 0  
## 8 0 0 0 0 0 0  
## 10 0 0 0 0 0 0  
## 12 0 0 0 0 0 0  
## X3601.3700 X3701.3800 X3801.3900 X3901.4000 X4001.4100 X4101.4200  
## 2 0 0 0 0 0 0  
## 4 0 0 0 0 0 0  
## 6 0 0 0 0 0 0  
## 8 0 0 0 0 0 0  
## 10 0 0 0 0 0 0  
## 12 0 0 0 0 0 0  
## X4201.4300 X4301.4400 X4401.4500 X0.200 X201.1000 X1000. Epipelagic  
## 2 0 0 0 0 1 1 0  
## 4 0 0 0 1 1 1 1  
## 6 0 0 0 1 1 0 1  
## 8 0 0 0 0 1 1 0  
## 10 0 0 0 1 1 0 1  
## 12 0 0 0 1 0 0 1  
## Mesopelagic Bathypelagic Abyssopelagic Fisheries Gamefish Aquarium Size  
## 2 1 1 0 0 0 0 117  
## 4 1 1 0 1 0 0 196  
## 6 1 0 0 1 0 0 139  
## 8 1 1 0 1 1 0 482  
## 10 1 0 0 1 0 0 180  
## 12 0 0 0 1 1 1 296  
## X.Reproductive.Strategy  
## 2   
## 4 Ovoviviparous  
## 6 Ovoviviparous  
## 8 Ovoviviparous  
## 10 Ovoviviparous  
## 12 Ovoviviparous

summary(dd)

## Order Family Genus   
## Length:501 Length:501 Length:501   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
## species Common.name Data.Deficient Deepwater   
## Length:501 Length:501 Min. :0.0000 Min. :0.0000   
## Class :character Class :character 1st Qu.:0.0000 1st Qu.:0.0000   
## Mode :character Mode :character Median :0.0000 Median :1.0000   
## Mean :0.3752 Mean :0.6607   
## 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :1.0000 Max. :1.0000   
## Coastal Pelagic Benthic BrackishFreshwater  
## Min. :0.0000 Min. :0.00000 Length:501 Min. :0.00000   
## 1st Qu.:0.0000 1st Qu.:0.00000 Class :character 1st Qu.:0.00000   
## Median :1.0000 Median :0.00000 Mode :character Median :0.00000   
## Mean :0.5908 Mean :0.05788 Mean :0.08383   
## 3rd Qu.:1.0000 3rd Qu.:0.00000 3rd Qu.:0.00000   
## Max. :1.0000 Max. :1.00000 Max. :1.00000   
## Tropical Temperate Pacific Atlantic   
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :1.0000 Median :1.0000 Median :1.0000 Median :0.0000   
## Mean :0.6188 Mean :0.5569 Mean :0.5908 Mean :0.3373   
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000   
## Indian Arctic Antartic Depth   
## Min. :0.0000 Min. :0.000000 Min. :0 Length:501   
## 1st Qu.:0.0000 1st Qu.:0.000000 1st Qu.:0 Class :character   
## Median :0.0000 Median :0.000000 Median :0 Mode :character   
## Mean :0.3792 Mean :0.003992 Mean :0   
## 3rd Qu.:1.0000 3rd Qu.:0.000000 3rd Qu.:0   
## Max. :1.0000 Max. :1.000000 Max. :0   
## X0.100 X101.200 X201.300 X301.400   
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :0.0000 Median :0.0000 Median :0.0000 Median :0.0000   
## Mean :0.4551 Mean :0.4032 Mean :0.4012 Mean :0.3912   
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000   
## X401.500 X501.600 X601.700 X701.800   
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :0.0000 Median :0.0000 Median :0.0000 Median :0.0000   
## Mean :0.3633 Mean :0.3234 Mean :0.3054 Mean :0.2615   
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000   
## X801.900 X901.1000 X1001.1100 X1101.1200   
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :0.0000 Median :0.0000 Median :0.0000 Median :0.0000   
## Mean :0.2176 Mean :0.2096 Mean :0.1597 Mean :0.1277   
## 3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:0.0000   
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000   
## X1201.1300 X1301.1400 X1401.1500 X1501.1600   
## Min. :0.0000 Min. :0.00000 Min. :0.00000 Min. :0.00000   
## 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:0.00000 1st Qu.:0.00000   
## Median :0.0000 Median :0.00000 Median :0.00000 Median :0.00000   
## Mean :0.1118 Mean :0.09182 Mean :0.07585 Mean :0.05389   
## 3rd Qu.:0.0000 3rd Qu.:0.00000 3rd Qu.:0.00000 3rd Qu.:0.00000   
## Max. :1.0000 Max. :1.00000 Max. :1.00000 Max. :1.00000   
## X1601.1700 X1701.1800 X1801.1900 X1901.2000   
## Min. :0.00000 Min. :0.00000 Min. :0.00000 Min. :0.00000   
## 1st Qu.:0.00000 1st Qu.:0.00000 1st Qu.:0.00000 1st Qu.:0.00000   
## Median :0.00000 Median :0.00000 Median :0.00000 Median :0.00000   
## Mean :0.04391 Mean :0.04192 Mean :0.03593 Mean :0.02595   
## 3rd Qu.:0.00000 3rd Qu.:0.00000 3rd Qu.:0.00000 3rd Qu.:0.00000   
## Max. :1.00000 Max. :1.00000 Max. :1.00000 Max. :1.00000   
## X2001.2100 X2101.2200 X2201.2300 X2301.2400   
## Min. :0.00000 Min. :0.00000 Min. :0.00000 Min. :0.000000   
## 1st Qu.:0.00000 1st Qu.:0.00000 1st Qu.:0.00000 1st Qu.:0.000000   
## Median :0.00000 Median :0.00000 Median :0.00000 Median :0.000000   
## Mean :0.01796 Mean :0.01397 Mean :0.00998 Mean :0.007984   
## 3rd Qu.:0.00000 3rd Qu.:0.00000 3rd Qu.:0.00000 3rd Qu.:0.000000   
## Max. :1.00000 Max. :1.00000 Max. :1.00000 Max. :1.000000   
## X2401.2500 X2501.2600 X2601.2700   
## Min. :0.000000 Min. :0.000000 Min. :0.000000   
## 1st Qu.:0.000000 1st Qu.:0.000000 1st Qu.:0.000000   
## Median :0.000000 Median :0.000000 Median :0.000000   
## Mean :0.005988 Mean :0.005988 Mean :0.005988   
## 3rd Qu.:0.000000 3rd Qu.:0.000000 3rd Qu.:0.000000   
## Max. :1.000000 Max. :1.000000 Max. :1.000000   
## X2701.2800 X2801.2900 X2901.3000   
## Min. :0.000000 Min. :0.000000 Min. :0.000000   
## 1st Qu.:0.000000 1st Qu.:0.000000 1st Qu.:0.000000   
## Median :0.000000 Median :0.000000 Median :0.000000   
## Mean :0.005988 Mean :0.005988 Mean :0.005988   
## 3rd Qu.:0.000000 3rd Qu.:0.000000 3rd Qu.:0.000000   
## Max. :1.000000 Max. :1.000000 Max. :1.000000   
## X3001.3100 X3101.3200 X3201.3300   
## Min. :0.000000 Min. :0.000000 Min. :0.000000   
## 1st Qu.:0.000000 1st Qu.:0.000000 1st Qu.:0.000000   
## Median :0.000000 Median :0.000000 Median :0.000000   
## Mean :0.005988 Mean :0.005988 Mean :0.005988   
## 3rd Qu.:0.000000 3rd Qu.:0.000000 3rd Qu.:0.000000   
## Max. :1.000000 Max. :1.000000 Max. :1.000000   
## X3301.3400 X3401.3500 X3501.3600   
## Min. :0.000000 Min. :0.000000 Min. :0.000000   
## 1st Qu.:0.000000 1st Qu.:0.000000 1st Qu.:0.000000   
## Median :0.000000 Median :0.000000 Median :0.000000   
## Mean :0.005988 Mean :0.005988 Mean :0.003992   
## 3rd Qu.:0.000000 3rd Qu.:0.000000 3rd Qu.:0.000000   
## Max. :1.000000 Max. :1.000000 Max. :1.000000   
## X3601.3700 X3701.3800 X3801.3900   
## Min. :0.000000 Min. :0.000000 Min. :0.000000   
## 1st Qu.:0.000000 1st Qu.:0.000000 1st Qu.:0.000000   
## Median :0.000000 Median :0.000000 Median :0.000000   
## Mean :0.003992 Mean :0.001996 Mean :0.001996   
## 3rd Qu.:0.000000 3rd Qu.:0.000000 3rd Qu.:0.000000   
## Max. :1.000000 Max. :1.000000 Max. :1.000000   
## X3901.4000 X4001.4100 X4101.4200   
## Min. :0.000000 Min. :0.000000 Min. :0.000000   
## 1st Qu.:0.000000 1st Qu.:0.000000 1st Qu.:0.000000   
## Median :0.000000 Median :0.000000 Median :0.000000   
## Mean :0.001996 Mean :0.001996 Mean :0.001996   
## 3rd Qu.:0.000000 3rd Qu.:0.000000 3rd Qu.:0.000000   
## Max. :1.000000 Max. :1.000000 Max. :1.000000   
## X4201.4300 X4301.4400 X4401.4500 X0.200   
## Min. :0.000000 Min. :0.000000 Min. :0.000000 Min. :0.0000   
## 1st Qu.:0.000000 1st Qu.:0.000000 1st Qu.:0.000000 1st Qu.:0.0000   
## Median :0.000000 Median :0.000000 Median :0.000000 Median :1.0000   
## Mean :0.001996 Mean :0.001996 Mean :0.001996 Mean :0.5529   
## 3rd Qu.:0.000000 3rd Qu.:0.000000 3rd Qu.:0.000000 3rd Qu.:1.0000   
## Max. :1.000000 Max. :1.000000 Max. :1.000000 Max. :1.0000   
## X201.1000 X1000. Epipelagic Mesopelagic   
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :1.0000 Median :0.0000 Median :1.0000 Median :1.0000   
## Mean :0.5988 Mean :0.1577 Mean :0.5529 Mean :0.5988   
## 3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000   
## Bathypelagic Abyssopelagic Fisheries Gamefish   
## Min. :0.0000 Min. :0.000000 Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.000000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :0.0000 Median :0.000000 Median :0.0000 Median :0.0000   
## Mean :0.1577 Mean :0.001996 Mean :0.4132 Mean :0.1377   
## 3rd Qu.:0.0000 3rd Qu.:0.000000 3rd Qu.:1.0000 3rd Qu.:0.0000   
## Max. :1.0000 Max. :1.000000 Max. :1.0000 Max. :1.0000   
## Aquarium Size X.Reproductive.Strategy  
## Min. :0.00000 Min. : 16.0 Length:501   
## 1st Qu.:0.00000 1st Qu.: 51.0 Class :character   
## Median :0.00000 Median : 80.0 Mode :character   
## Mean :0.03593 Mean : 119.2   
## 3rd Qu.:0.00000 3rd Qu.: 130.0   
## Max. :1.00000 Max. :1900.0

# Rename some columns  
dd = dd %>% rename("Rep\_Strategy" = "X.Reproductive.Strategy",  
 "Antarctic" = "Antartic")  
  
class(dd$Epipelagic)

## [1] "numeric"

#Cleaning up depth data  
# assigning values to the 5 layers of the ocean  
# epipelagic - 0-200 m  
# mesopelagic - 200 - 1000 m  
# bathypelagic - 1,000 - 4,000 m   
# abyssopelagic - 4,000-6,000 m   
  
dd[dd$Epipelagic == ""] <- "0"  
dd$Epipelagic <- factor(dd$Epipelagic, levels = c("0", "1"))  
class(dd$Epipelagic)

## [1] "factor"

dd[dd$Mesopelagic == ""] <- "0"  
dd$Mesopelagic <- factor(dd$Mesopelagic, levels = c("0", "1"))  
class(dd$Mesopelagic)

## [1] "factor"

dd[dd$Bathypelagic == ""] <- "0"  
dd$Bathypelagic <- factor(dd$Bathypelagic, levels = c("0", "1"))  
class(dd$Bathypelagic)

## [1] "factor"

dd[dd$Abyssopelagic == ""] <- "0"  
dd$Abyssopelagic <- factor(dd$Abyssopelagic, levels = c("0", "1"))  
class(dd$Abyssopelagic)

## [1] "factor"

# Cleaning up fisheries data  
dd$Fisheries[dd$Fisheries == "NA"] <- "Unknown"  
dd$Fisheries[dd$Fisheries == ""] <- "Unknown"  
dd$Fisheries <- factor(dd$Fisheries, levels = c("Unknown", "0", "1"))  
table(dd$Fisheries)

##   
## Unknown 0 1   
## 0 294 207

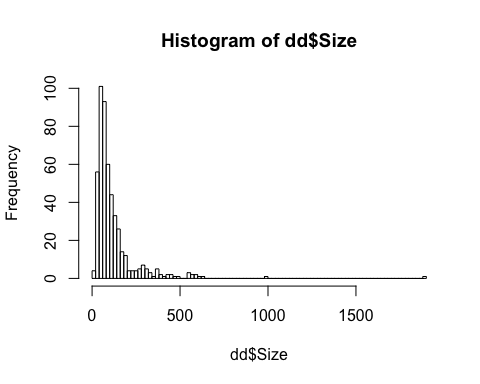
# Cleaning up reproductive strategy data  
dd$Rep\_Strategy[dd$Rep\_Strategy == ""] = "Unknown"  
dd$Rep\_Strategy = factor(dd$Rep\_Strategy, levels = c("Unknown", "Ovoviviparous", "Viviparous", "Oviparous"))  
table(dd$Rep\_Strategy)

##   
## Unknown Ovoviviparous Viviparous Oviparous   
## 126 171 83 121

# Removing columns unused in data  
dd$Benthic <- as.numeric(dd$Benthic) %>% replace\_na(0)

## Warning in eval(lhs, parent, parent): NAs introduced by coercion

## Check normality of size variable; needs to be log transformed  
hist(dd$Size,100)



dd$log\_size<- log(dd$Size) #added a column with log transformed values, which we will use in the GLM  
summary(dd$log\_size)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 2.773 3.932 4.382 4.464 4.868 7.550

#simple look:   
#Out of 501 shark species assessed: 37.5% are classified as DD, with ~ 4% not evaluated  
sum(dd$Data.Deficient==1, na.rm=T)/nrow(dd) #.375

## [1] 0.3752495

sum(dd$Data.Deficient==0, na.rm=T)/nrow(dd) # .585

## [1] 0.6247505

### After we have our superficial values describing the data, time to do some stats and visualization. What potential predictive factors are we interested in?

*Biology*  
Size  
Reproductive strategy

*Human Use*  
Human use in general (any use)  
Type of use (fisheries, gamefish, aquaculture) Fisheries (any fishery)

*Ecology*  
Geographic range (oceans, trans-oceanic) Latitude/temperature (tropical/temperate)  
Habitat (benthic, deepwater, pelagic, coastal, freshwater)  
Depth range (epipelagic, mesopelagic, bathypelagic, abyssopelagic)

# Potential models (both first- and second-order interactions included)

#First, we will look at the coefficients of every single first order variable and extract p-values to see what we are dealing with  
firstorder.mod <- glm(Data.Deficient ~ Deepwater + Coastal + Pelagic + Benthic + BrackishFreshwater +  
 Tropical\*Temperate +   
 Pacific + Atlantic + Indian + Fisheries+  
 Gamefish + Aquarium + Epipelagic + Mesopelagic + Bathypelagic +  
 log\_size + Rep\_Strategy, data = dd, family = "binomial")  
   
firstorder.mod

##   
## Call: glm(formula = Data.Deficient ~ Deepwater + Coastal + Pelagic +   
## Benthic + BrackishFreshwater + Tropical \* Temperate + Pacific +   
## Atlantic + Indian + Fisheries + Gamefish + Aquarium + Epipelagic +   
## Mesopelagic + Bathypelagic + log\_size + Rep\_Strategy, family = "binomial",   
## data = dd)  
##   
## Coefficients:  
## (Intercept) Deepwater   
## 1.79473 0.22436   
## Coastal Pelagic   
## -0.46632 -0.46412   
## Benthic BrackishFreshwater   
## 0.56906 -1.54869   
## Tropical Temperate   
## 0.91286 0.68344   
## Pacific Atlantic   
## 0.27927 0.45216   
## Indian Fisheries1   
## -0.74141 -0.33739   
## Gamefish Aquarium   
## -0.43982 0.98588   
## Epipelagic1 Mesopelagic1   
## -0.50291 -0.34206   
## Bathypelagic1 log\_size   
## -0.81606 -0.60830   
## Rep\_StrategyOvoviviparous Rep\_StrategyViviparous   
## 0.04194 0.44226   
## Rep\_StrategyOviparous Tropical:Temperate   
## 0.02486 -0.85348   
##   
## Degrees of Freedom: 500 Total (i.e. Null); 479 Residual  
## Null Deviance: 663   
## Residual Deviance: 560.2 AIC: 604.2

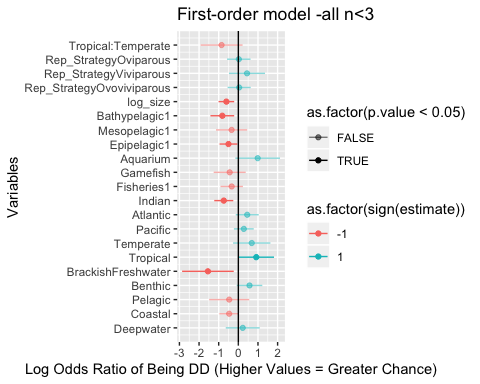
### final model  
secondorder.mod6 <- glm(Data.Deficient ~   
 Coastal\*Benthic\*Fisheries + Deepwater\*Benthic\*Fisheries + BrackishFreshwater+ ## no interaction term for brackishfreshwater due to small n  
Pelagic\*Fisheries+  
 Tropical\*Temperate +   
 Pacific + Atlantic + Indian +   
 Fisheries + Gamefish + Aquarium +   
 Epipelagic + Mesopelagic + Bathypelagic +  
 log\_size + Rep\_Strategy, data = dd, family = "binomial")  
  
secondorder.mod6

##   
## Call: glm(formula = Data.Deficient ~ Coastal \* Benthic \* Fisheries +   
## Deepwater \* Benthic \* Fisheries + BrackishFreshwater + Pelagic \*   
## Fisheries + Tropical \* Temperate + Pacific + Atlantic + Indian +   
## Fisheries + Gamefish + Aquarium + Epipelagic + Mesopelagic +   
## Bathypelagic + log\_size + Rep\_Strategy, family = "binomial",   
## data = dd)  
##   
## Coefficients:  
## (Intercept) Coastal   
## -0.57011 2.06960   
## Benthic Fisheries1   
## 3.71972 2.78048   
## Deepwater BrackishFreshwater   
## 1.89787 -1.52914   
## Pelagic Tropical   
## -0.09195 0.91963   
## Temperate Pacific   
## 0.68483 0.22841   
## Atlantic Indian   
## 0.34023 -0.90258   
## Gamefish Aquarium   
## -0.42300 1.17743   
## Epipelagic1 Mesopelagic1   
## -0.50070 -0.34106   
## Bathypelagic1 log\_size   
## -0.77113 -0.68176   
## Rep\_StrategyOvoviviparous Rep\_StrategyViviparous   
## -0.02626 0.49283   
## Rep\_StrategyOviparous Coastal:Benthic   
## 0.04485 -2.48501   
## Coastal:Fisheries1 Benthic:Fisheries1   
## -2.61819 -3.19172   
## Benthic:Deepwater Fisheries1:Deepwater   
## -2.18636 -1.79160   
## Fisheries1:Pelagic Tropical:Temperate   
## -0.64644 -0.85026   
## Coastal:Benthic:Fisheries1 Benthic:Fisheries1:Deepwater   
## 1.66091 2.83593   
##   
## Degrees of Freedom: 500 Total (i.e. Null); 471 Residual  
## Null Deviance: 663   
## Residual Deviance: 542 AIC: 602

### Now to visualize a coefficient plots to see what is significant!

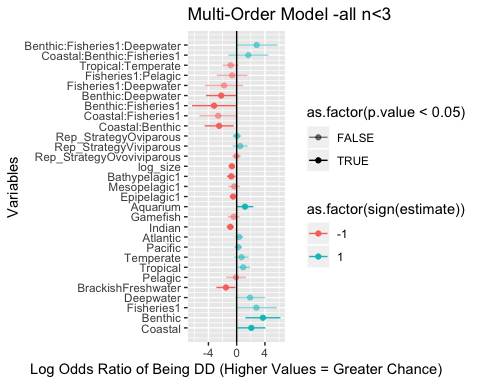
# # Pull out plot data  
coefs1 <- plot\_model(firstorder.mod, transform = NULL)   
coefs1$data %>%  
 mutate(term = factor(term, levels = term)) %>%  
 ggplot(aes(x = estimate,  
 y = as.numeric(term),  
 color = as.factor(sign(estimate)),  
 alpha = as.factor(p.value < 0.05))) +  
 geom\_point() +   
 geom\_segment(aes(x = conf.low,  
 xend = conf.high,  
 yend = as.numeric(term))) +  
 geom\_vline(xintercept = 0) +  
 scale\_y\_continuous(labels = as.character(coefs1$data$term),  
 breaks = seq(1, 21, 1)) +  
 scale\_alpha\_discrete(range = c(.5, 1)) +  
 xlab("Log Odds Ratio of Being DD (Higher Values = Greater Chance)") +  
 ylab("Variables")+ ggtitle("First-order model -all n<3")

## Warning: Using alpha for a discrete variable is not advised.



##### Chosen model  
coefs2.6 <- plot\_model(secondorder.mod6, transform = NULL)   
coefs2.6$data %>%  
 mutate(term = factor(term, levels = term)) %>%  
 ggplot(aes(x = estimate,  
 y = as.numeric(term),  
 color = as.factor(sign(estimate)),  
 alpha = as.factor(p.value < 0.05))) +  
 geom\_point() +   
 geom\_segment(aes(x = conf.low,  
 xend = conf.high,  
 yend = as.numeric(term))) +  
 geom\_vline(xintercept = 0) +  
 scale\_y\_continuous(labels = as.character(coefs2.6$data$term),  
 breaks = seq(1, 29, 1)) +  
 scale\_alpha\_discrete(range = c(.5, 1)) +  
 xlab("Log Odds Ratio of Being DD (Higher Values = Greater Chance)") +  
 ylab("Variables")+ ggtitle("Multi-Order Model -all n<3")

## Warning: Using alpha for a discrete variable is not advised.



### Let’s visualize some of our our significant results

Start with making a table of the emmeans, which give us our predicted, backtransformed values of probability of DD based on the glm

## subset coefficients by category  
fullcoeffs <- summary(secondorder.mod6)$coefficients %>% as.data.frame(.) %>% rownames\_to\_column()  
coef\_conf <- confint(secondorder.mod6) %>% as.data.frame(.) %>% rownames\_to\_column() #error here

## Waiting for profiling to be done...

coef\_table <- left\_join(fullcoeffs, coef\_conf) %>%  
 rename("Coef" = "rowname")

## Joining, by = "rowname"

########### Generating Estimated Marginal Mean ##########  
# estimates the average value for that combination of parameters, and predicts the resulting mean for a set of specified values #  
  
library(emmeans)  
  
# size  
logsize\_means <- emmeans(firstorder.mod, ~ log\_size,   
 var = "log\_size",   
 at = list(log\_size = seq(2, 8, by = .25)),  
 type = "response")  
  
#habitat  
Brackish\_means <- emmeans(firstorder.mod, ~ BrackishFreshwater,   
 var = "BrackishFreshwater",   
 at = list(BrackishFreshwater = c(0,1),  
 type = "response"))  
  
DW\_means <- emmeans(firstorder.mod, ~ Deepwater,   
 var = "Deepwater",   
 at = list(Deepwater = c(0,1),  
 type = "response")) #giving probability that the animal is data deficient if benthic = 0,1 and fished  
  
  
coastalfish\_means <- emmeans(firstorder.mod, ~ Coastal,   
 var = "Coastal",   
 at = list(Coastal = c(0,1),  
 type = "response"))  
  
pelagicfish\_means <- emmeans(firstorder.mod, ~ Pelagic,   
 var = "Pelagic",   
 at = list(Pelagic = c(0,1)),  
 type = "response")  
  
  
#oceans  
Atlantic\_means <- emmeans(firstorder.mod, ~ Atlantic,   
 var = "Atlantic",   
 at = list(Atlantic = c(0,1)),  
 type = "response")  
  
Pacific\_means <- emmeans(firstorder.mod, ~ Pacific,   
 var = "Pacific",   
 at = list(Pacific = c(0,1)),  
 type = "response")  
  
Indian\_means <- emmeans(firstorder.mod, ~ Indian,   
 var = "Indian",   
 at = list(Indian = c(0,1)),  
 type = "response")  
  
# depth  
bathy\_means <- emmeans(firstorder.mod, ~ Bathypelagic,   
 var = "Bathypelagic",   
 at = list(Bathypelagic = factor(c(0,1))),  
 type = "response")  
  
meso\_means <- emmeans(firstorder.mod, ~ Mesopelagic,   
 var = "Mesopelagic",   
 at = list(Mesopelagic = factor(c(0,1))),  
 type = "response")  
  
epipelagic\_means <- emmeans(firstorder.mod, ~ Epipelagic,   
 var = "Epipelagic",   
 at = list(Epipelagic = factor(c(0,1))),  
 type = "response")  
  
  
# tropical/temperate  
temptrop\_means <- emmeans(firstorder.mod, ~ Tropical\*Temperate,   
 var = "Tropical\*Temperate",   
 at = list(Tropical = c(0,1),  
 Temperate = c(0,1)),  
 type = "response")  
  
# coastal and deepwater alone (NOTE: all other variables are averaged; this is not affected by any other variable)  
benthic\_alone\_means <- emmeans(firstorder.mod, ~ Benthic,   
 var = "Benthic",   
 at = list(Benthic = c(0,1)),  
 type = "response")  
  
fisheries\_means <- emmeans(firstorder.mod, ~ Fisheries,   
 var = "Fisheries",   
 at = list(Fisheries = factor(c(0,1))),  
 type = "response")  
  
gamefish\_means <- emmeans(firstorder.mod, ~ Gamefish,   
 var = "Gamefish",   
 at = list(Gamefish = c(0,1)),  
 type = "response")  
  
aquarium\_means <- emmeans(firstorder.mod, ~ Aquarium,   
 var = "Aquarium",   
 at = list(Aquarium = c(0,1)),  
 type = "response")  
  
  
logsize\_means

## log\_size prob SE df asymp.LCL asymp.UCL  
## 2.00 0.6284249 0.12182602 Inf 0.3782241 0.8246282  
## 2.25 0.5922782 0.11452159 Inf 0.3644501 0.7863199  
## 2.50 0.5551045 0.10557499 Inf 0.3505659 0.7425350  
## 2.75 0.5173037 0.09532379 Inf 0.3364894 0.6936976  
## 3.00 0.4793039 0.08421284 Inf 0.3220905 0.6407263  
## 3.25 0.4415420 0.07278200 Inf 0.3071634 0.5850652  
## 3.50 0.4044444 0.06165587 Inf 0.2913828 0.5286477  
## 3.75 0.3684075 0.05153894 Inf 0.2742465 0.4737924  
## 4.00 0.3337819 0.04319936 Inf 0.2550448 0.4230235  
## 4.25 0.3008605 0.03736552 Inf 0.2330135 0.3787083  
## 4.50 0.2698708 0.03441808 Inf 0.2078993 0.3423316  
## 4.75 0.2409730 0.03402485 Inf 0.1806494 0.3137274  
## 5.00 0.2142615 0.03524889 Inf 0.1531906 0.2913037  
## 5.25 0.1897707 0.03709628 Inf 0.1273750 0.2731634  
## 5.50 0.1674825 0.03887082 Inf 0.1043337 0.2578492  
## 5.75 0.1473360 0.04019435 Inf 0.0844988 0.2444260  
## 6.00 0.1292368 0.04090952 Inf 0.0678480 0.2323263  
## 6.25 0.1130660 0.04099050 Inf 0.0541187 0.2212047  
## 6.50 0.0986894 0.04048357 Inf 0.0429463 0.2108453  
## 6.75 0.0859635 0.03947068 Inf 0.0339439 0.2011084  
## 7.00 0.0747426 0.03804753 Inf 0.0267441 0.1919000  
## 7.25 0.0648824 0.03631070 Inf 0.0210193 0.1831547  
## 7.50 0.0562439 0.03435023 Inf 0.0164875 0.1748250  
## 7.75 0.0486956 0.03224588 Inf 0.0129126 0.1668754  
## 8.00 0.0421152 0.03006552 Inf 0.0101004 0.1592784  
##   
## Results are averaged over the levels of: Fisheries, Epipelagic, Mesopelagic, Bathypelagic, Rep\_Strategy   
## Confidence level used: 0.95   
## Intervals are back-transformed from the logit scale

Brackish\_means

## BrackishFreshwater emmean SE df asymp.LCL asymp.UCL  
## 0 -0.8435952 0.1753329 Inf -1.187241 -0.4999491  
## 1 -2.3922874 0.6619408 Inf -3.689667 -1.0949073  
##   
## Results are averaged over the levels of: Fisheries, Epipelagic, Mesopelagic, Bathypelagic, Rep\_Strategy   
## Results are given on the logit (not the response) scale.   
## Confidence level used: 0.95

DW\_means

## Deepwater emmean SE df asymp.LCL asymp.UCL  
## 0 -1.1216551 0.3080005 Inf -1.725325 -0.5179851  
## 1 -0.8972958 0.2518912 Inf -1.390993 -0.4035980  
##   
## Results are averaged over the levels of: Fisheries, Epipelagic, Mesopelagic, Bathypelagic, Rep\_Strategy   
## Results are given on the logit (not the response) scale.   
## Confidence level used: 0.95

coastalfish\_means

## Coastal emmean SE df asymp.LCL asymp.UCL  
## 0 -0.6979171 0.2113887 Inf -1.112231 -0.2836027  
## 1 -1.1642340 0.2168857 Inf -1.589322 -0.7391459  
##   
## Results are averaged over the levels of: Fisheries, Epipelagic, Mesopelagic, Bathypelagic, Rep\_Strategy   
## Results are given on the logit (not the response) scale.   
## Confidence level used: 0.95

pelagicfish\_means

## Pelagic prob SE df asymp.LCL asymp.UCL  
## 0 0.2795771 0.03517427 Inf 0.21604880 0.3533655  
## 1 0.1961267 0.08406800 Inf 0.07901745 0.4096073  
##   
## Results are averaged over the levels of: Fisheries, Epipelagic, Mesopelagic, Bathypelagic, Rep\_Strategy   
## Confidence level used: 0.95   
## Intervals are back-transformed from the logit scale

temptrop\_means

## Tropical Temperate prob SE df asymp.LCL asymp.UCL  
## 0 0 0.1645476 0.06307730 Inf 0.07418724 0.3261904  
## 1 0 0.3291770 0.04916848 Inf 0.24079565 0.4315581  
## 0 1 0.2806302 0.04583167 Inf 0.19999847 0.3783938  
## 1 1 0.2927720 0.05435110 Inf 0.19838202 0.4091500  
##   
## Results are averaged over the levels of: Fisheries, Epipelagic, Mesopelagic, Bathypelagic, Rep\_Strategy   
## Confidence level used: 0.95   
## Intervals are back-transformed from the logit scale

Atlantic\_means

## Atlantic prob SE df asymp.LCL asymp.UCL  
## 0 0.2449090 0.03838246 Inf 0.1775921 0.3275790  
## 1 0.3376488 0.05678927 Inf 0.2365873 0.4560891  
##   
## Results are averaged over the levels of: Fisheries, Epipelagic, Mesopelagic, Bathypelagic, Rep\_Strategy   
## Confidence level used: 0.95   
## Intervals are back-transformed from the logit scale

Pacific\_means

## Pacific prob SE df asymp.LCL asymp.UCL  
## 0 0.2426095 0.04322876 Inf 0.1680482 0.3368579  
## 1 0.2975166 0.04190565 Inf 0.2223309 0.3855234  
##   
## Results are averaged over the levels of: Fisheries, Epipelagic, Mesopelagic, Bathypelagic, Rep\_Strategy   
## Confidence level used: 0.95   
## Intervals are back-transformed from the logit scale

Indian\_means

## Indian prob SE df asymp.LCL asymp.UCL  
## 0 0.3335324 0.04318876 Inf 0.2548224 0.4227615  
## 1 0.1925285 0.03685532 Inf 0.1302995 0.2750772  
##   
## Results are averaged over the levels of: Fisheries, Epipelagic, Mesopelagic, Bathypelagic, Rep\_Strategy   
## Confidence level used: 0.95   
## Intervals are back-transformed from the logit scale

bathy\_means

## Bathypelagic prob SE df asymp.LCL asymp.UCL  
## 0 0.3622997 0.03203128 Inf 0.3021403 0.4271065  
## 1 0.2007755 0.04788453 Inf 0.1227835 0.3107579  
##   
## Results are averaged over the levels of: Fisheries, Epipelagic, Mesopelagic, Rep\_Strategy   
## Confidence level used: 0.95   
## Intervals are back-transformed from the logit scale

meso\_means

## Mesopelagic prob SE df asymp.LCL asymp.UCL  
## 0 0.3095136 0.06314576 Inf 0.2007683 0.4444085  
## 1 0.2415030 0.04315359 Inf 0.1671186 0.3356531  
##   
## Results are averaged over the levels of: Fisheries, Epipelagic, Bathypelagic, Rep\_Strategy   
## Confidence level used: 0.95   
## Intervals are back-transformed from the logit scale

epipelagic\_means

## Epipelagic prob SE df asymp.LCL asymp.UCL  
## 0 0.3269590 0.04520935 Inf 0.2451527 0.4208442  
## 1 0.2270788 0.03732725 Inf 0.1622321 0.3083069  
##   
## Results are averaged over the levels of: Fisheries, Mesopelagic, Bathypelagic, Rep\_Strategy   
## Confidence level used: 0.95   
## Intervals are back-transformed from the logit scale

benthic\_alone\_means

## Benthic prob SE df asymp.LCL asymp.UCL  
## 0 0.1955833 0.04956100 Inf 0.1159306 0.3107278  
## 1 0.3004668 0.03946863 Inf 0.2291502 0.3829524  
##   
## Results are averaged over the levels of: Fisheries, Epipelagic, Mesopelagic, Bathypelagic, Rep\_Strategy   
## Confidence level used: 0.95   
## Intervals are back-transformed from the logit scale

fisheries\_means

## Fisheries prob SE df asymp.LCL asymp.UCL  
## 0 0.3090149 0.04628970 Inf 0.2262468 0.4061676  
## 1 0.2419309 0.04350312 Inf 0.1670011 0.3368836  
##   
## Results are averaged over the levels of: Epipelagic, Mesopelagic, Bathypelagic, Rep\_Strategy   
## Confidence level used: 0.95   
## Intervals are back-transformed from the logit scale

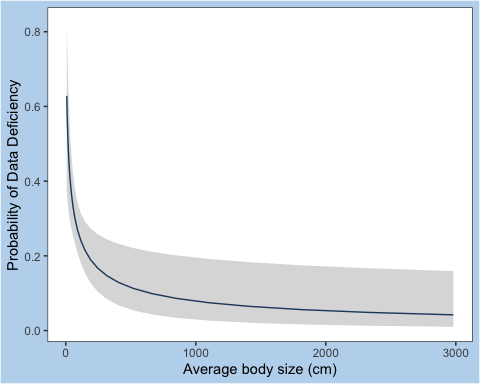
gamefish\_means

## Gamefish prob SE df asymp.LCL asymp.UCL  
## 0 0.2864167 0.03675775 Inf 0.2200547 0.3634652  
## 1 0.2054338 0.06673787 Inf 0.1039566 0.3655560  
##   
## Results are averaged over the levels of: Fisheries, Epipelagic, Mesopelagic, Bathypelagic, Rep\_Strategy   
## Confidence level used: 0.95   
## Intervals are back-transformed from the logit scale

aquarium\_means

## Aquarium prob SE df asymp.LCL asymp.UCL  
## 0 0.2672057 0.03440513 Inf 0.2053427 0.339738  
## 1 0.4942580 0.14579393 Inf 0.2375564 0.754023  
##   
## Results are averaged over the levels of: Fisheries, Epipelagic, Mesopelagic, Bathypelagic, Rep\_Strategy   
## Confidence level used: 0.95   
## Intervals are back-transformed from the logit scale

# Plotting the size figure  
data.frame(logsize\_means) %>%  
 ggplot(aes(x = exp(log\_size), # Exponentiating log size  
 y = prob)) + geom\_line(color="#1F4E79")+  
 geom\_ribbon(aes(ymin = asymp.LCL,  
 ymax = asymp.UCL),  
 alpha = .2) + theme\_bw() + theme(panel.grid.major = element\_blank(), panel.grid.minor = element\_blank()) + theme(plot.background=element\_rect(fill = "#BDD7EE"),  
 panel.background = element\_rect(fill = "white"))+  
xlab("Average body size (cm)") +  
 ylab("Probability of Data Deficiency")



### try with new color  
data.frame(logsize\_means) %>%  
 ggplot(aes(x = exp(log\_size), # Exponentiating log size  
 y = prob)) + geom\_line(color="#1F4E79")+  
 geom\_ribbon(aes(ymin = asymp.LCL,  
 ymax = asymp.UCL),  
 alpha = .2) + theme\_bw() + theme(panel.grid.major = element\_blank(), panel.grid.minor = element\_blank()) + theme(plot.background=element\_rect(fill = "#EDD58B"),  
 panel.background = element\_rect(fill = "white"))+   
xlab(" ") +  
 ylab(" ")

