

Enhanced.R

STEVE

2024-07-07

```
##loading the required libraries####
```

```
library(tidyverse)
```

```
#install.packages("hexbin")
```

```
library(hexbin)
```

```
##Import the dataset###
```

```
my_data<- read.csv("enhanced_parrotlet.csv", header = TRUE)
```

```
head(my_data)
```

##	Presence	Long	Lat	alt	mdr	wc2.1_10m_prec_02.tif
## 1	0	-92.24583	15.58750	1515	138	10
## 2	0	-96.11250	16.09583	1393	118	11
## 3	0	-97.74583	16.13750	21	148	5
## 4	0	-96.33750	16.21250	2846	119	11
## 5	0	-94.22083	16.32083	33	120	5
## 6	0	-97.99583	16.41250	503	153	2
##	wc2.1_10m_tavg_02.tif					
## 1		17.04075				
## 2		21.06850				
## 3		24.41080				
## 4		16.76275				
## 5		26.64079				
## 6		23.70225				

```
tail(my_data)
```

##	Presence	Long	Lat	alt	mdr	wc2.1_10m_prec_02.tif
## 383	0	-115.7292	31.54583	1042	152	39
## 384	0	-115.3042	31.72083	93	177	16
## 385	0	-112.5125	31.77083	709	169	17
## 386	0	-115.1125	31.97083	3	181	8
## 387	0	-114.3458	32.00417	81	162	7
## 388	0	-116.2875	32.28750	1214	148	49
##	wc2.1_10m_tavg_02.tif					
## 383		9.37725				
## 384		15.49850				
## 385		13.69600				
## 386		15.96200				
## 387		15.63075				
## 388		7.91600				

```
dim(my_data)
```

```
## [1] 388    7
```

##Data cleaning###

```
summary(is.na(my_data)) #checking for missing values
```

```
##   Presence           Long           Lat
##   Mode :logical      Mode :logical   Mode :logical
## FALSE:388          FALSE:388        FALSE:388
##    alt           mdr      wc2.1_10m_prec_02.tif
##   Mode :logical      Mode :logical   Mode :logical
## FALSE:388          FALSE:388        FALSE:388
## wc2.1_10m_tavg_02.tif
##   Mode :logical
## FALSE:388
```

```
my_data$Presence<-as.factor(my_data$Presence)
```

```
str(my_data)
```

```
## 'data.frame':    388 obs. of  7 variables:
## $ Presence      : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ Long          : num -92.2 -96.1 -97.7 -96.3 -94.2 ...
## $ Lat           : num 15.6 16.1 16.1 16.2 16.3 ...
## $ alt           : int 1515 1393 21 2846 33 503 646 722 284 703 ...
## $ mdr           : int 138 118 148 119 120 153 159 160 114 127 ...
## $ wc2.1_10m_prec_02.tif: num 10 11 5 11 5 2 7 7 8 15 ...
## $ wc2.1_10m_tavg_02.tif: num 17 21.1 24.4 16.8 26.6 ...
```

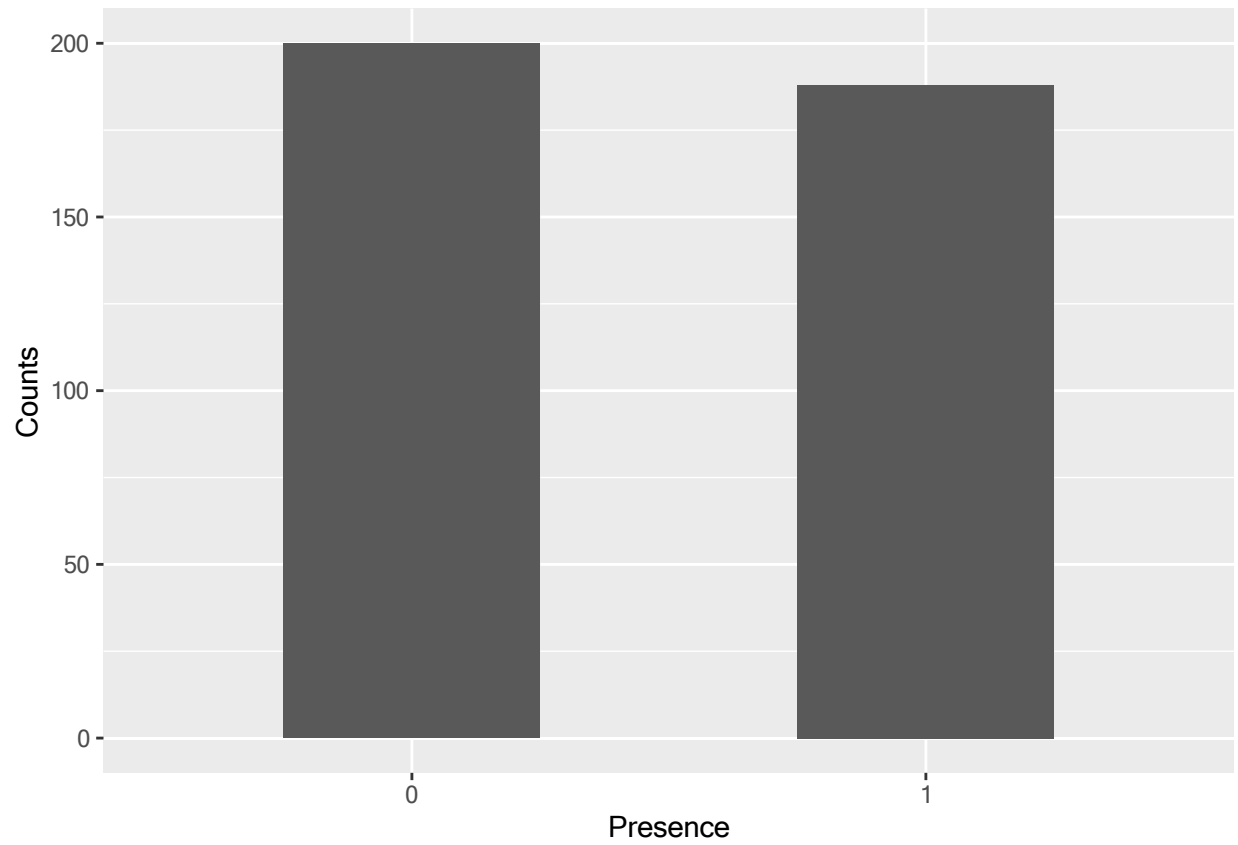
#EXploring the dataset###

```
summary(my_data)
```

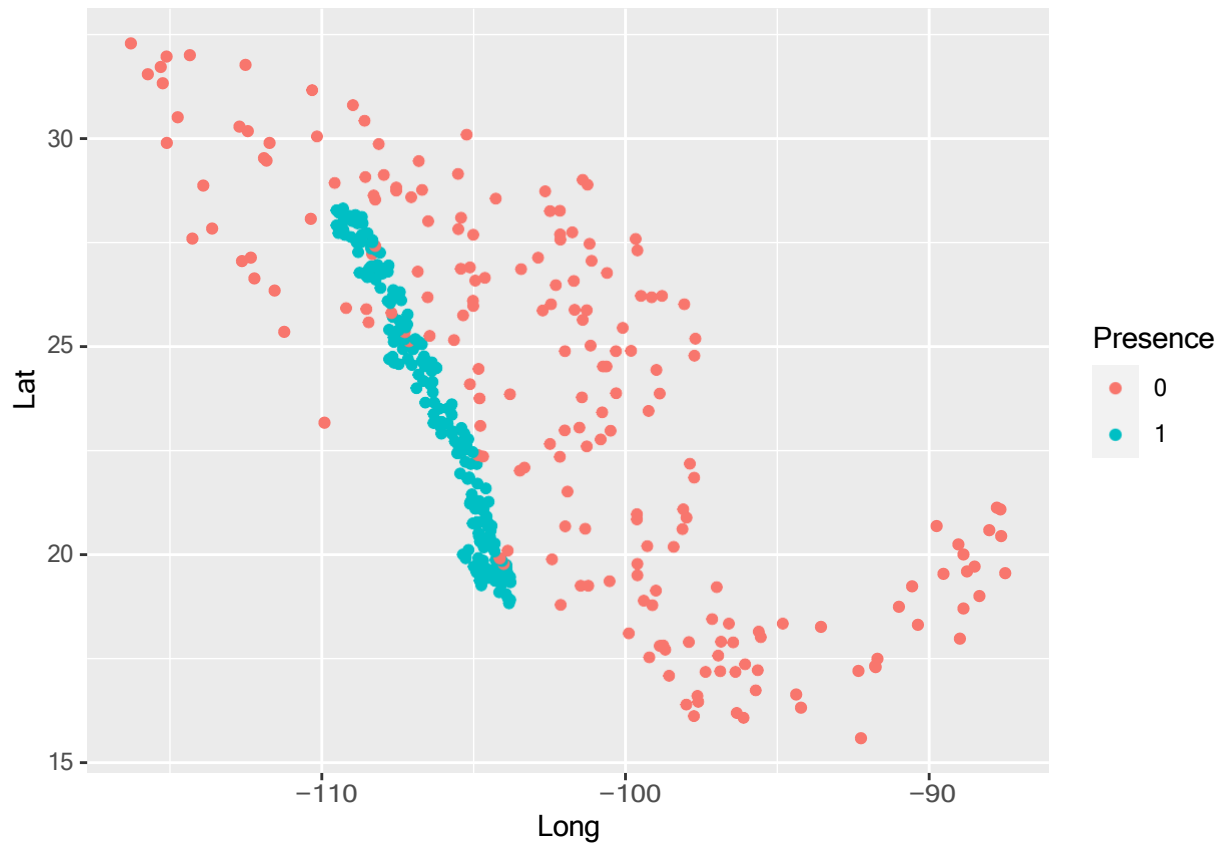
```
## Presence           Long           Lat           alt
## 0:200   Min.   :-116.3   Min.   :15.59   Min.   : 0.0
## 1:188   1st Qu.: -107.4   1st Qu.:20.07   1st Qu.: 263.5
##         Median :-104.9   Median :23.70   Median : 788.0
##         Mean   :-104.0   Mean   :23.52   Mean   : 960.5
##         3rd Qu.: -101.4   3rd Qu.:26.72   3rd Qu.:1530.5
##         Max.    : -87.5   Max.    :32.29   Max.    :2995.0
##      mdr      wc2.1_10m_prec_02.tif wc2.1_10m_tavg_02.tif
## Min.   : 85.0   Min.   : 2.00   Min.   : 5.928
## 1st Qu.:132.0   1st Qu.: 8.00   1st Qu.:14.260
## Median :149.0   Median :13.00   Median :17.374
## Mean   :147.8   Mean   :17.98   Mean   :17.260
## 3rd Qu.:166.0   3rd Qu.:24.00   3rd Qu.:20.585
## Max.   :193.0   Max.   :97.00   Max.   :26.641
```

visualizing the distribution

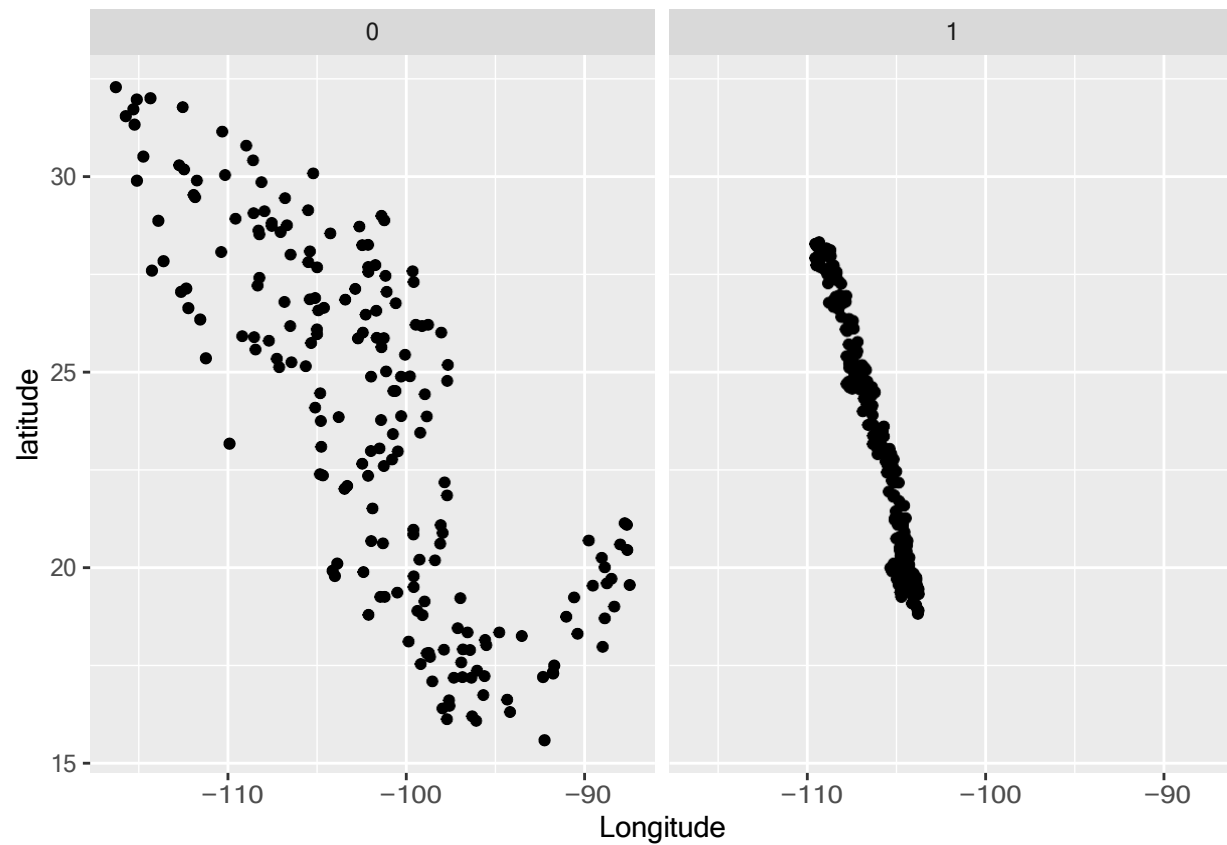
```
ggplot(my_data, aes(x=Presence))+
  geom_bar(width = 0.5)+
  labs(x="Presence", y="Counts")
```



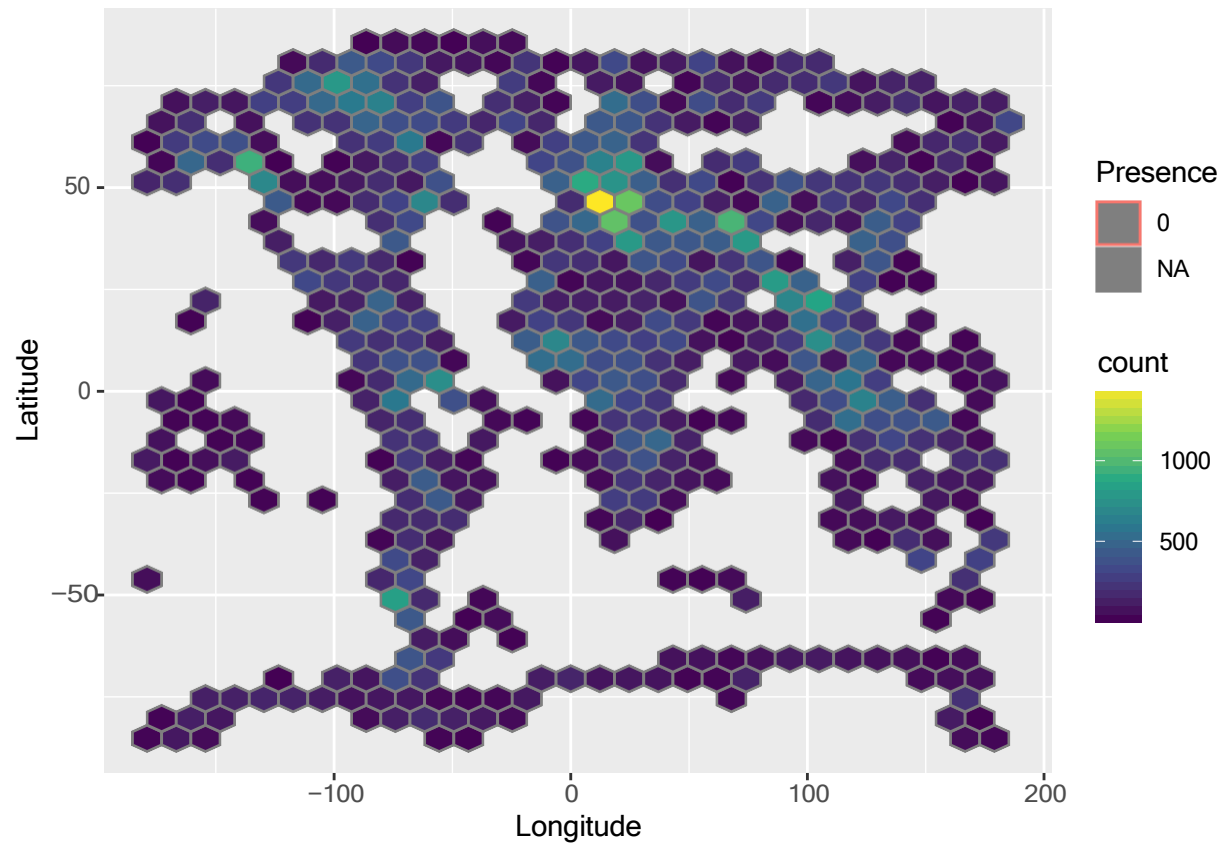
```
ggplot(my_data, aes(x=Long, y=Lat, color= Presence))+geom_point()
```



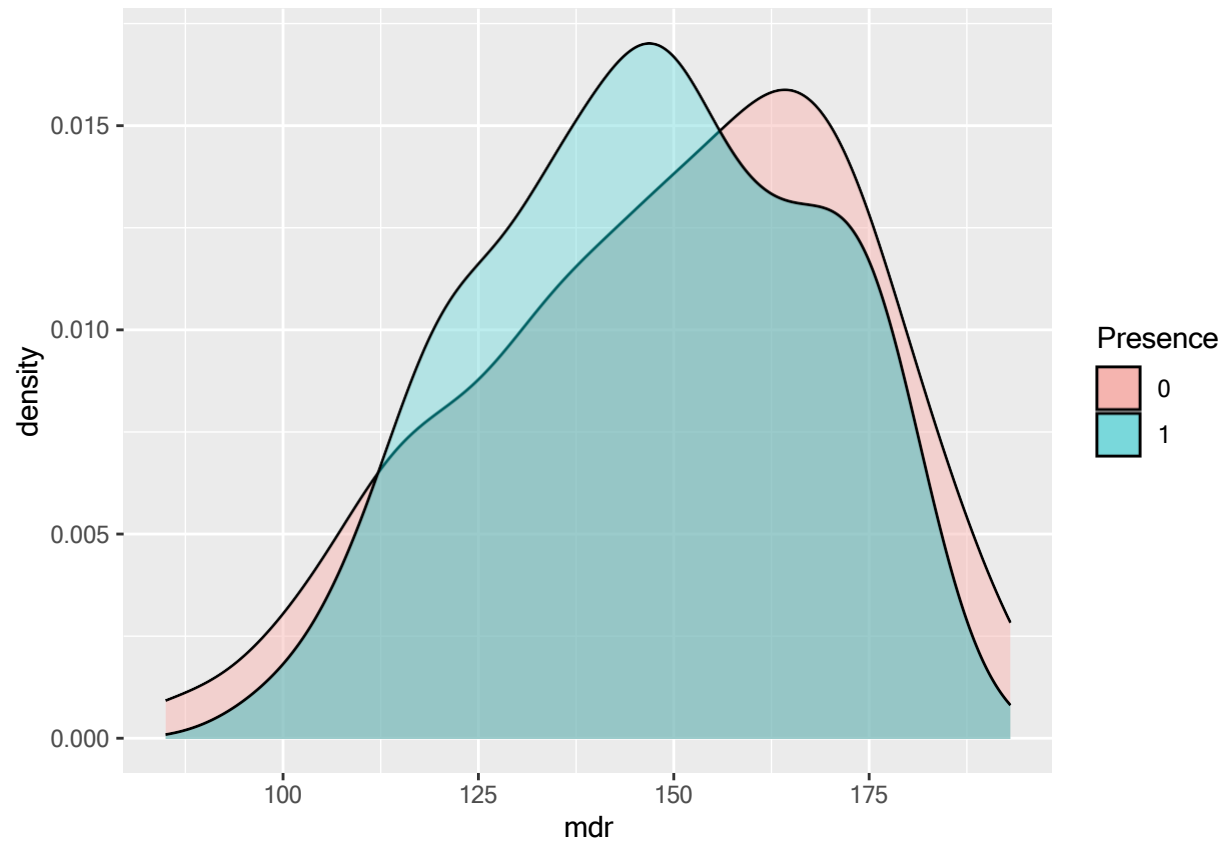
```
ggplot(my_data, aes(x=Long,y=Lat))+geom_point()+  
  facet_grid(~Presence)+  
  labs(x= "Longitude", y="latitude")
```



```
mapdata<-map_data("world")
View(mapdata)
mapdata1 <-left_join(mapdata,my_data, by=c("long"="Long"))
ggplot(mapdata1, aes(x=long, y=lat, color=Presence))+
  stat_bin_hex()+
  scale_fill_continuous(type="viridis")+
  labs(x="Longitude", y="Latitude")
```



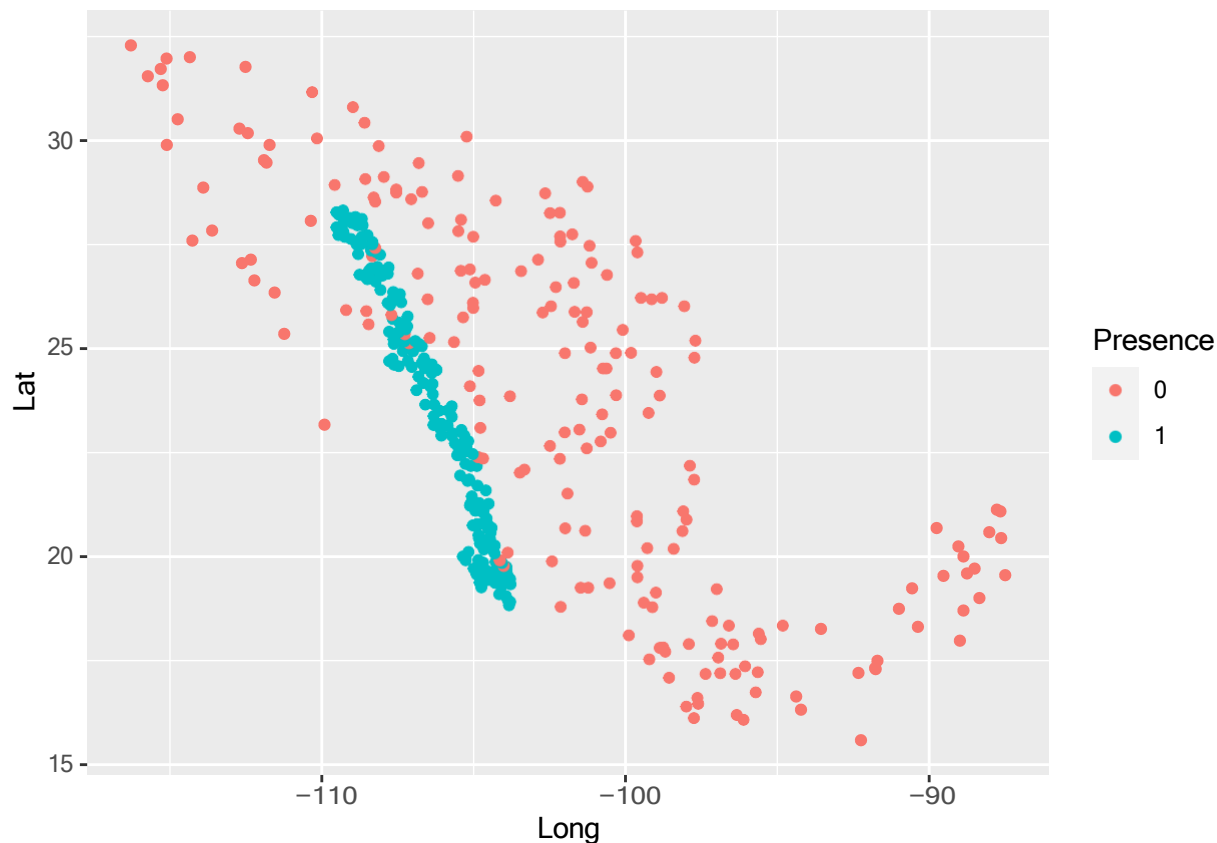
```
ggplot(my_data, aes(x=mdr, fill=Presence))+geom_density(alpha=0.5)
```



```
ggplot(my_data, aes(x=alt, fill=Presence))+geom_density(alpha=0.5)
```



```
ggplot(my_data, aes(x=Long, y=Lat, color=Presence))+geom_point()+  
labs(color="Presence")
```

###regression####

```
cor(my_data %>% select(alt, mdr, Long, Lat))
```

```
##           alt           mdr           Long           Lat
## alt      1.00000000  0.3804194 -0.03622941  0.0337363
## mdr      0.38041940  1.0000000 -0.52115011  0.5274415
## Long     -0.03622941 -0.5211501  1.00000000 -0.6769991
## Lat      0.03373630  0.5274415 -0.67699913  1.0000000
```

##Logistics###

```
model<-glm(Presence~alt+mdr, data=my_data, family = "binomial")
summary(model)
```

```
##
## Call:
## glm(formula = Presence ~ alt + mdr, family = "binomial", data = my_data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4954  -1.1043  -0.7263   1.1095   1.7309
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.0689433  0.7149404  -0.096   0.923
## alt         -0.0006761  0.0001551  -4.358 1.31e-05 ***
## mdr          0.0043821  0.0050985   0.859   0.390
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 537.51  on 387  degrees of freedom
## Residual deviance: 516.12  on 385  degrees of freedom
## AIC: 522.12
##
## Number of Fisher Scoring iterations: 4
```