````
---
title: "raptor"
output:
  word_document: default
  pdf_document: default
  html_document: default
---

```{r}
library(tidyverse)
library(readxl)
library(grid)
library(glmmTMB)
library(mgcv)
library(lme4)
library(matrixStats)
library(mice)
library(gridExtra)

```
import data
```{r}
# Note the data is not the orginal data, I have changed some names in the excel,such as Wind
#spd to wind_spd, wind Dir to wind_Dir, please upload the correct data, and I have uploaded
# my data to the net: https://github.com/S1877256/raptor
HMworking <- read_excel("HMworkingdatav2.xlsx")
View(HMworking)
```
missing counter id = 100
```{r}
# impute NA value for counter
hm <- filter(HMworking , Observer != 0) %>%
      mutate(Counter = (ifelse(Counter == "NA", 100 , Counter)))

df1 <- select(hm, Counter,Observer1,Observer2.new,Observer3,Observer4)

 # build a little counter function
counter <- function(strings, input) {
  return(sapply(strings, function(i) 5-sum(grepl(i, input))))
}

# get the counts
myCounts <- sapply(1:nrow(df1), function(i) counter(strings=c("NA"), df1[i,]))

hm <- cbind(hm, myCounts)
hm1<- mutate(hm, amount = as.integer(Observer/Duration))%>%
      mutate(error = amount - myCounts,
            Observer4 = ifelse(error > 0, 99, Observer4))
raptor <-   separate(hm1, Date, into = c("year", "month","days"), sep = "-")%>%
            separate(days, into = c("day", "time"), sep = " ")%>%
            separate(Flight_HT, into = c("height", "name"), sep = ":") %>%
            mutate(height=factor(as.numeric(height)))%>%
            select( -myCounts,-amount,-error,-year,-month,-time,-Wind_Spd,-Precipitation,-name)



```


exploratory data analysis
```{r}
month.count <- select(raptor,Year, Month,TOTAL) %>%
  group_by(Year, Month) %>%
  summarise(n = sum(TOTAL)) %>%
  unite(date,Year, Month, sep = '/') %>%
  separate(date, into = c("1", "dates"), sep = "20") %>%
  select(-"1")
p1 =  ggplot(as.data.frame(month.count), mapping = aes(x = dates, y = n)) +
    geom_bar(stat = "identity", color='blue',fill='blue',alpha=0.5) +
````
````

```
    geom_text(mapping = aes(label= n),size = 3.5,vjust = -0.5) +
    ggtitle("the number of raptors in every months")

wind <- select(raptor,Wind_Dir,TOTAL,Wind_Spd2) %>%
          group_by(Wind_Dir,Wind_Spd2) %>%
          summarise(n = sum(TOTAL))
p2= ggplot(data = wind) +
  geom_bar(mapping = aes(x = Wind_Spd2, y = n),stat = "identity",
color='blue',fill='blue',alpha=0.5) +
  facet_wrap(~ Wind_Dir, nrow = 2)


p3= ggplot(data = raptor) +
  geom_point(mapping = aes(x = Temp, y = TOTAL),stat = "identity",
color='blue',fill='blue',alpha=0.5) +
  facet_wrap(~ Precipitation2 , nrow = 2)

p4= ggplot(data = raptor) +
  geom_point(mapping = aes(x = Cloud_Cover, y = TOTAL),stat = "identity",
color='blue',fill='blue',alpha=0.5)

 p5= ggplot(data = raptor) +
  geom_point(mapping = aes(x =  Humidity, y = TOTAL),stat = "identity",
color='blue',fill='blue',alpha=0.5)


p6 =  ggplot(data = raptor) +
  geom_point(mapping = aes(x =BARO, y = TOTAL),stat = "identity",
color='blue',fill='blue',alpha=0.5)

p7 =  ggplot(data = raptor) +
  geom_point(mapping = aes(x = as.numeric(Visibility), y = TOTAL),stat = "identity",
color='blue',fill='blue',alpha=0.5)+
  xlab("Visibility")
p1
p2
p3
grid.arrange(p4, p5,p6,p7, ncol=2)


```

```{r}
df.count <- rowwise(raptor) %>%
              mutate(eagles =sum(BE+GE+UE),
                     hawks =sum(OS+NH+SS+CH+NG+RS+BW+RT+RL+UA),
                     falcons =sum(AK+ML+PG+UF),
                     buzzards =sum(BV+TV+UB),
                     unknown.raptor = UR) %>%
              select(Temp, Precipitation2, TOTAL,eagles, hawks, falcons,
buzzards,unknown.raptor,Month)
   temp.count <- group_by(df.count,Temp) %>%
                summarise(count = sum(TOTAL),
                          eagle = sum(eagles),
                          hawk = sum(hawks),
                          falcon = sum(falcons),
                          buzzard = sum(buzzards),
                          unk_raptor=sum(unknown.raptor)) %>%
                pivot_longer(c('eagle', 'hawk', 'falcon', 'buzzard','unk_raptor'), names_to =
"species", values_to = "number")

C <- c('#ff8080','#c5944e','#458B00','#55a0fb','#8470FF')
p1<-ggplot(as.data.frame(temp.count), mapping = aes(x = Temp, y = number)) +
      geom_line(mapping = aes(colour = species),size=0.5) +
    scale_colour_manual(values = C)+
    ggtitle("the number of different raptors in temperature")

c2=c('#ff8080','#c5944e','#458B00','#8470FF')
 p2 <- filter(temp.count,species != "hawk") %>%
```

```
  ggplot(mapping = aes(x = Temp, y = number)) +
      geom_smooth(mapping = aes(colour = species),show.legend = F)+
   scale_colour_manual(values = c2)+
   theme(
  plot.title = element_blank(),
  axis.title.x = element_blank(),
  axis.title.y = element_text(size=8))+
  ylab("trend")

  p1
  sub <- viewport(width = 0.4,height = 0.55,x = 0.3,y = 0.6)
  print(p2,vp = sub)
```

```{r}

```

impute missing value

```{r}
# missing value
missing <- raptor %>% filter(Wind_Spd2 == 'NA' | Wind_Dir == 'NA' | Temp =='NA'| Precipitation2
=='NA' | Humidity =="NA" | BARO == 'NA'| Cloud_Cover =="NA"| Visibility =="NA")

# visibilty impute NA
set.seed(11)
indhdm1 <- which(raptor$Visibility != 'NA' & raptor$Cloud_Cover == 100 & raptor$Temp < 0 &
raptor$Precipitation2 == 1 & raptor$Wind_Spd2 == 2)
indhdm1
donor1 <- sample(indhdm1, 1, replace = TRUE)
donor1

indhdm2 <- which(raptor$Visibility != 'NA' & raptor$Cloud_Cover == 0 & raptor$Temp < 0 &
raptor$Wind_Spd2 >= 2 & raptor$Precipitation2 >= 1 )
indhdm2
donor2 <- sample(indhdm2, 1, replace = TRUE)
donor2
raptor$visibility <-
c(raptor$Visibility[raptor$Visibility!='NA'],raptor$Visibility[donor1],raptor$Visibility[donor2])


# impute baro and humidity 0
missing2 <- raptor %>% filter( Humidity == 0 | BARO == 0| BARO >= 40)

humidity <- select(raptor,Humidity,Precipitation2, Temp,Cloud_Cover,BARO) %>%
  mutate(humidity=ifelse(Humidity==0,NA,Humidity),
         Precipitation=ifelse( Precipitation2=='NA',NA, Precipitation2),
         Precipitation=factor(Precipitation),
         baro = ifelse(BARO == 0,NA,BARO))%>%
  select(-Humidity,-Precipitation2,-BARO)
imp <- mice(humidity, m = 5, method = "pmm")
data_imp <- complete(imp)
humidity2 = data_imp$humidity
precipitation2 = data_imp$Precipitation
baro2 = data_imp$baro


raptor<- cbind(raptor,humidity2,precipitation2,baro2)


H<- select(raptor,height,Temp,precipitation2,humidity2,Visibility)
imp2 <- mice(H, m = 5, method = "pmm")
data_imp2 <- complete(imp2)
height2 = data_imp2$height
raptor<- cbind(raptor,height2)

# obserbver represent
raptor_ <- raptor %>% select(Observer1,Observer2.new,Observer3,Observer4,TOTAL)%>%
                      pivot_longer(c(Observer1,Observer2.new,Observer3,Observer4),names_to =
"obs", values_to = "id")%>%
```

```
                      group_by(id)%>%
                      summarise(obs_total=sum(TOTAL),
                                n=n())%>%
                      arrange(n,obs_total)%>%
                      filter(id!='NA')

#factor(,order=TRUE,levels=c("poor","improved","excellent")




raptor2 <- select(raptor ,-Visibility) %>%
           filter(Wind_Dir != 'NA' | Humidity != 0) %>%
           filter(Humidity != 127) %>%
           mutate(Counter = factor(Counter),
                  observer1=factor(Observer1, order=TRUE,levels=c('NA',raptor_$id)),
                  observer2=factor(Observer2.new, order=TRUE,levels=c('NA',raptor_$id)),
                  observer3=factor(Observer3, order=TRUE,levels=c('NA',raptor_$id)),
                  observer4=factor(Observer4, order=TRUE,levels=c('NA',raptor_$id)))%>%
                  mutate(observer =do.call('pmax',
                       data.frame(observer1,observer2,observer3,observer4))
                          )%>%
           rowwise()%>%
           mutate(eagles =sum(BE+GE+UE),
                  hawks =sum(OS+NH+SS+CH+NG+RS+BW+RT+RL+UA),
                  falcons =sum(AK+ML+PG+UF),
                  buzzards =sum(BV+TV+UB),
                  unknown.raptor = UR)%>%
           select(-Observer1,-Observer2.new,-Observer3,-Observer4,-BV:-UR,-Humidity,-
Precipitation2,-BARO,-height)
```

```

```{r}
str(raptor2)
```

glmmtmb
```{r}
tmb1 <- glmmTMB(TOTAL ~ log(Duration) + factor(Wind_Spd2) + factor(Wind_Dir) + Temp +
precipitation2 + humidity2 + baro2 + as.numeric(visibility) + (1|Counter)+(1|observer)+
(1|height2) ,  data = raptor2 , family = nbinom2)
summary(tmb1)

```

```{r}
tmb2 <- glmmTMB(TOTAL ~  log(Duration) + factor(Wind_Spd2) + factor(Wind_Dir) + Temp +
precipitation2 + humidity2 + baro2 + as.numeric(visibility) + (1|Counter)+(1|observer)+
(1|height2), data = raptor2, family=poisson)
summary(tmb2) # drop perticipation
```

gmcv
```{r}
gam1 <- gam( TOTAL ~ log(Duration) + factor(Wind_Spd2) + factor(Wind_Dir) + Temp +
precipitation2 + baro2 + Cloud_Cover + humidity2 + as.numeric(visibility) + s(height2,bs =
"re")+ s(Counter, bs = "re")+ s(observer, bs = "re") , data = raptor2, family = poisson, method
= "REML")
summary(gam1)
AIC(gam1)
```
```

```{r}
par(mfrow=c(2,2))
plot(predict(gam1,type="response"),residuals(gam1), main="a")
plot(predict(gam1,type="response"),gam1$y, main="b");abline(0,1,col=2)
plot(gam1$linear.predictors,gam1$y, main="c")
qq.gam(gam1,rep=20,level=1, main="d")

```

```{r}
nzeros<-numeric()
for (i in 1:100){
  simcounts<-rpois(n=length(gam1$residuals),lambda=predict(gam1,type="response")) # simulate
from poisson distribution using fitted values
  nzeros<-c(nzeros, sum(simcounts==0))
}
hist(nzeros); mean(nzeros); sum(raptor2$TOTAL== 0)
abline(v=sum(raptor2$TOTAL==0),col=2)
```

```{r}
gam2 <- gam(TOTAL ~ log(Duration) + factor(Wind_Spd2) + factor(Wind_Dir) + Temp +
precipitation2 + baro2 + Cloud_Cover + humidity2 + as.numeric(visibility) + s(height2,bs =
"re")+ s(Counter, bs = "re")+ s(observer, bs = "re") , data = raptor2, family = nb, method =
"REML",select = TRUE)
summary(gam2)
AIC(gam2)
```

```{r}
par(mfrow=c(2,2))
plot(predict(gam2,type="response"),residuals(gam2), main="a")
plot(predict(gam2,type="response"),gam2$y, main="b");abline(0,1,col=2)
plot(gam2$linear.predictors,gam2$y, main="c")
qq.gam(gam2,rep=20,level=1, main="d")
```

```{r}
thb <- gam2$family$getTheta(TRUE) ## extract final theta estimate
nzeros<-numeric()
for (i in 1:100){
  simcounts<-rnbinom(size=thb, n=length(gam2$residuals),mu=predict(gam2,type="response"))
  nzeros<-c(nzeros, sum(simcounts==0))
}
hist(nzeros); mean(nzeros);sum(raptor2$TOTAL==0)
abline(v=sum(raptor2$TOTAL==0),col=2)
```

```{r}
modPred <- predict.gam(gam2, se.fit=TRUE,type="response")
summary(modPred$fit)
summary(raptor2$TOTAL)
```

```{r}
par(mfrow=c(2,2))
gam.check(gam2)
```

```{r}
gam3 <- gam( TOTAL ~  log(Duration) +  factor(Wind_Spd2) + factor(Wind_Dir) + Temp +
precipitation2 + humidity2 + baro2 + as.numeric(visibility) + s(Counter, bs = "re")+
s(observer, bs = "re"), data = raptor2, family = nb, method = "REML",select = TRUE)
summary(gam3)
AIC(gam3)
```

```{r}
par(mfrow=c(2,2))
plot(predict(gam3,type="response"),residuals(gam3), main="a")
plot(predict(gam3,type="response"),gam3$y, main="b");abline(0,1,col=2)
plot(gam3$linear.predictors,gam3$y, main="c",xlim = c(-4,7))
qq.gam(gam3,rep=20,level=1, main="d")
```

```{r}

thb <- gam3$family$getTheta(TRUE) ## extract final theta estimate
nzeros<-numeric()
for (i in 1:100){
  simcounts<-rnbinom(size=thb, n=length(gam3$residuals),mu=predict(gam3,type="response"))
  nzeros<-c(nzeros, sum(simcounts==0))
}
hist(nzeros); mean(nzeros);sum(raptor2$TOTAL==0)
abline(v=sum(raptor2$TOTAL==0),col=2)
```

```{r}
coefficient <-exp(gam3$coefficients)
coefficient
as.data.frame(coefficient)
```

```{r}
anova(gam2)
anova(gam3)
```

```{r}
mod1<-  gam(TOTAL ~log(Duration) + factor(Wind_Spd2) + factor(Wind_Dir) + Temp + precipitation2
+ baro2 + Cloud_Cover + humidity2 + as.numeric(visibility) + s(Counter, bs = "re")+ s(observer,
bs = "re"), data = raptor2, family = nb, method = "REML",select = TRUE)


mod2<-  gam( TOTAL ~   factor(Wind_Spd2) + factor(Wind_Dir) + Temp + precipitation2 + humidity2
+ baro2 + Cloud_Cover +  as.numeric(visibility) + s(Counter, bs = "re")+ s(observer, bs =
"re"), data = raptor2, family = nb, method = "REML")  # drop Duration

mod3<-  gam( TOTAL ~ log(Duration) + factor(Wind_Dir) + Temp + precipitation2 + humidity2 +
baro2 + Cloud_Cover +  as.numeric(visibility) + s(Counter, bs = "re")+ s(observer, bs = "re"),
data = raptor2, family = nb, method = "REML")# drop Wind_Spd2

mod4<-  gam( TOTAL ~ log(Duration) + factor(Wind_Spd2) + Temp +precipitation2 + humidity2 +
baro2 + Cloud_Cover +  as.numeric(visibility) + s(Counter, bs = "re")+ s(observer, bs = "re"),
data = raptor2, family = nb, method = "REML") # drop Wind_Dir

mod5<-  gam( TOTAL ~ log(Duration) + factor(Wind_Spd2) + factor(Wind_Dir) + precipitation2 +
humidity2 + baro2 + Cloud_Cover +  as.numeric(visibility) + s(Counter, bs = "re")+ s(observer,
bs = "re"), data = raptor2, family = nb, method = "REML") # drop Temp

mod6<-  gam( TOTAL ~  log(Duration) + factor(Wind_Spd2) + factor(Wind_Dir) + Temp + humidity2 +
baro2+ Cloud_Cover +  as.numeric(visibility) + s(Counter, bs = "re")+ s(observer, bs = "re"),
data = raptor2, family = nb, method = "REML")# drop Precipitation2

mod7<-  gam( TOTAL ~  log(Duration) +  factor(Wind_Spd2) + factor(Wind_Dir) + Temp +
precipitation2 + baro2+ Cloud_Cover +  as.numeric(visibility) + s(Counter, bs = "re")+
s(observer, bs = "re"), data = raptor2, family = nb, method = "REML") # drop Humidity

mod8<-  gam( TOTAL ~  log(Duration) +  factor(Wind_Spd2) + factor(Wind_Dir) + Temp +
precipitation2 + humidity2 + Cloud_Cover +  as.numeric(visibility) + s(Counter, bs = "re")+
s(observer, bs = "re"), data = raptor2, family = nb, method = "REML") # drop BARO

mod9<-  gam( TOTAL ~  log(Duration) +  factor(Wind_Spd2) + factor(Wind_Dir) + Temp +
precipitation2 + humidity2 + baro2 +as.numeric(visibility) + s(Counter, bs = "re")+ s(observer,
bs = "re"), data = raptor2, family = nb, method = "REML") # drop Cloud_Cover

mod10<-  gam( TOTAL ~  log(Duration) + Wind_Spd2 + factor(Wind_Dir) + Temp + precipitation2 +
humidity2 + baro2 + Cloud_Cover + s(Counter, bs = "re")+ s(observer, bs = "re"), data =
```

```
raptor2, family = nb, method = "REML")    # drop visibility

aic.vec<-(c(AIC(mod1), AIC(mod2),AIC(mod3), AIC(mod4), AIC(mod5),AIC(mod6),
AIC(mod7),AIC(mod8),AIC(mod9),AIC(mod10)))
plot(aic.vec, ylab="AIC", xlab="model")


```

```{r}
#eagle
gam8 <- gam(eagles ~  log(Duration) + factor(Wind_Dir) + Temp + precipitation2 + humidity2  +
as.numeric(visibility) + s(Counter, bs = "re")+ s(observer, bs = "re"), data = raptor2, family
= nb, method = "REML",select = TRUE)
summary(gam8)
AIC(gam8)

```

```{r}
#hawk
gam4 <- gam( hawks ~ log(Duration) +  factor(Wind_Dir) + Temp + precipitation2 +
as.numeric(visibility)+  humidity2 + s(Counter, bs = "re")+ s(observer, bs = "re"), data =
raptor2, family = nb, method = "REML",select = TRUE)
summary(gam4)
AIC(gam4)
```

```{r}
# falcon
gam5 <- gam( falcons ~  log(Duration) +  factor(Wind_Dir) + Temp + precipitation2 +
as.numeric(visibility)+  humidity2 + s(Counter, bs = "re")+ s(observer, bs = "re"), data =
raptor2, family = nb, method = "REML",select = TRUE)
summary(gam5)
AIC(gam5)
```

```{r}
#buzzard
gam6 <- gam(buzzards ~ log(Duration) +  factor(Wind_Dir) + Temp + precipitation2 +
as.numeric(visibility)+  humidity2 + s(Counter, bs = "re")+ s(observer, bs = "re"), data =
raptor2, family = nb, method = "REML")
summary(gam6)
AIC(gam6)

```

```{r}
#unknown.raptor
gam7 <- gam( unknown.raptor ~ log(Duration) +  factor(Wind_Dir) + Temp + precipitation2 +
as.numeric(visibility)+  humidity2 + s(Counter, bs = "re")+ s(observer, bs = "re"), data =
raptor2, family = nb, method = "REML")
summary(gam7)
AIC(gam7)
```