Introduction

In the wake of recent high profile airline incidents involving equipment failures, there has been renewed discussion involving airline safety. The airline industry is amongst the most highly regulated, with many rules being written after catastrophic events leading to loss of life or significant airliner damage. Although many of these recent incidents have involved human negligence, there is another source of plane failure, caused by animals. Bird strikes involving aircraft are an important topic for study due to their potential for substantial economic impact as well as their high safety risk in the event of a serious strike (Sodhi 2002)[1]. Although called bird strikes, these events encompass any animal collision of an aircraft, which does occasionally include land mammals on runways.

Hypothesis

We suspect that factors including time of day, mass of the animal species, number of engines, and phase of the flightpath are significant predictors of whether an aircraft will sustain damage. Manipulating these variables could help to reduce the amount and severity of impacts, with the potential to save money and improve passenger safety. Furthermore, we postulate flights at night may have reduced potential for strikes, owing to reduced animal nighttime activity. Finally, in the event of a strike reported as damaging, we expect the typical weight of the species to be highly correlated with repair costs.

Method

Data obtained from the FAA (Federal Aviation Administration), a subset of 28,300 reports of bird strikes, was used for analysis (FAA Wildlife Strike Database)[2]. The subset includes only data where a bird strike was reported, regardless of severity. Incidents included in our dataset range from January 2000 to May 2015. A table of the top fifty species responsible for collisions was also drawn from our main FAA data. We manually matched typical adult weights for each species to the table from available online data (University of Michigan Animal Diversity Web Database) [3].

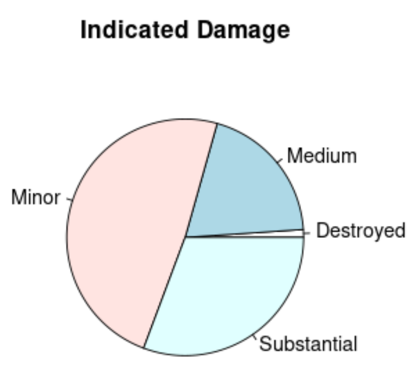

Figure 1. Breakdown of repair cost classifications when there has been a damaging strike.

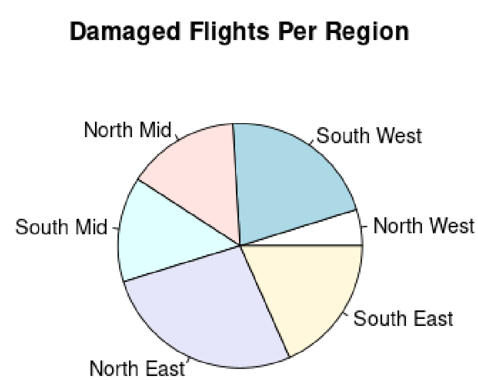

Figure 2. Breakdown of strikes by region, indicating some bias but a generally balanced distribution.

Exploratory analysis indicated that our class breakdown consists of 3111 damaging events and 25187 non-damaging strikes. The majority of damaging events fall into the minor category, with the remainder split between medium and substantial [Fig. 1]. Of particular interest are the 33 reports of ‘Destroyed’, where costs can climb into the multi-millions. We can also see that the data is relatively normally distributed in terms of regional incidents [Fig. 2]. We assume the relative variance in regional events is caused by a combination of higher aircraft volume and migratory bird flight patterns.

Our initial model was a multiple logistic regression to predict the probability of a damaging strike. Random forest is well known to be particularly sensitive to class imbalance, which was a substantial factor in this dataset. Rather than apply complex transformations, we opted for a simpler model. Based on available literature, EPV (Events per Variable) is a workable metric to determine how small a category of data can be while still allowing for robust regression (Austin & Steyerberg 2017) [4]. In our case, the smallest category is n = 33, which exceeds the researcher’s n = 20 guideline for all variables. An assumption in this model is that there is little error in the FAA records of each event, and that the fifteen year sample is generalizable to the population of flights throughout the United States.

Originally, our secondary model was intended to be a multiple linear regression to investigate the samples where damage occurred, using damage cost as the dependent variable. With this approach, we could estimate both the probability of damage occurring and an estimated cost if it does, given specific event parameters. Unfortunately, the cost total column indicating repair costs is very skewed. So skewed, that even with logarithmic transformation the data does not follow a normal distribution. Since the cost data violates assumptions of linear regression, we instead opted for a second logarithmic model, to classify damaging events between minor and substantial.

Results

```
Coefficients:
                              Estimate Std. Error z value Pr(>|z|)    
(Intercept)                  1.418e+00  1.879e-01   7.545 4.54e-14 ***
Aircraft..Number.of.engines  6.348e-01  8.718e-02   7.281 3.32e-13 ***
timeofday                   -2.112e-01  7.548e-02  -2.798  0.00514 ** 
Typical.Weight              -2.213e-03  1.552e-04 -14.255  < 2e-16 ***
Feet.above.ground           -4.697e-04  2.754e-05 -17.055  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 9366.9  on 13351  degrees of freedom
Residual deviance: 8315.9  on 13347  degrees of freedom
  (2610 observations deleted due to missingness)
AIC: 8325.9
```

Figure 3. Initial regression model, incorporating key explanatory variables. Note timeofday is significant here but will not be in the subsequent model.

```
Coefficients:
                              Estimate Std. Error z value Pr(>|z|)    
(Intercept)                  1.079e+00  1.952e-01   5.528 3.25e-08 ***
Aircraft..Number.of.engines  6.236e-01  8.768e-02   7.112 1.15e-12 ***
timeofday                   -1.253e-01  7.600e-02  -1.649  0.09910 .  
Typical.Weight              -2.148e-03  1.456e-04 -14.751  < 2e-16 ***
Feet.above.ground           -4.410e-04  2.728e-05 -16.166  < 2e-16 ***
season2                      6.912e-01  8.269e-02   8.358  < 2e-16 ***
season3                      2.020e-01  7.943e-02   2.543  0.01100 *  
season4                     -2.414e-01  9.101e-02  -2.653  0.00798 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 9366.9  on 13351  degrees of freedom
Residual deviance: 8190.8  on 13344  degrees of freedom
  (2610 observations deleted due to missingness)
AIC: 8206.8
```

Figure. 4 A refined model with additional significant parameters, here we can see timeofday is no longer significant, which is probably a result of collinearity between seasons and timeofday. The length of day or night corresponds with the season, as well as migratory patterns. Overall seasons are a better predictor than timeofday, which becomes clear when they are added to a model together.

```
               Reference
Prediction       Caused damage No damage
  Caused damage            923      2773
  No damage                573      9083
                                          
               Accuracy : 0.7494          
                 95% CI : (0.742, 0.7567)
    No Information Rate : 0.888           
    P-Value [Acc > NIR] : 1               
                                          
                  Kappa : 0.2332          
                                          
 Mcnemar's Test P-Value : <2e-16          
                                          
            Sensitivity : 0.7661          
            Specificity : 0.6170          
         Pos Pred Value : 0.9407          
         Neg Pred Value : 0.2497          
             Prevalence : 0.8880          
         Detection Rate : 0.6803          
   Detection Prevalence : 0.7232          
      Balanced Accuracy : 0.6915          
                                          
       'Positive' Class : No damage       
```

Figure 5. Our confusion matrix shows moderate accuracy, with some issues in specificity. The model struggles with the class imbalance to some degree.

Discussion

The limitations of our dataset became more apparent as the model was constructed and refined. Our initial model, figure 3, shows that we correctly hypothesized the species' average weight, and the engine count of an aircraft are statistically significant variables in predicting whether an animal collision will cause damage. Naturally, an aircraft with more engines has more points of failure, furthermore animal mass can increase the severity of an impact. This is well illustrated in figures 6 & 7 below, where we can see despite accounting for fewer strikes, the typical repair cost sustained is much higher for species such as geese and deer.

Feetaboveground was a proxy variable for the phase of flight, where a value of zero would indicate 'at the airport'. We see an inverse relationship, which indicates damage is most likely to occur near the airport, such as during landing or takeoff. This is corroborated by industry statistics about the most dangerous phase of flight (Armstrong 2024) [5].

The hypothesis that a nighttime flight would have less chance of a damaging strike was proven incorrect, as the timeofday explanatory variable was not very significant. We decided to refine the model by adding other significant variables like seasons as shown in figure 4 above. Once seasonality was added as an explanatory variable, we could see a reduction in the importance of timeofday. There is likely some degree of collinearity, but we believe seasons are more useful data points for explaining probability of a damaging animal strike.

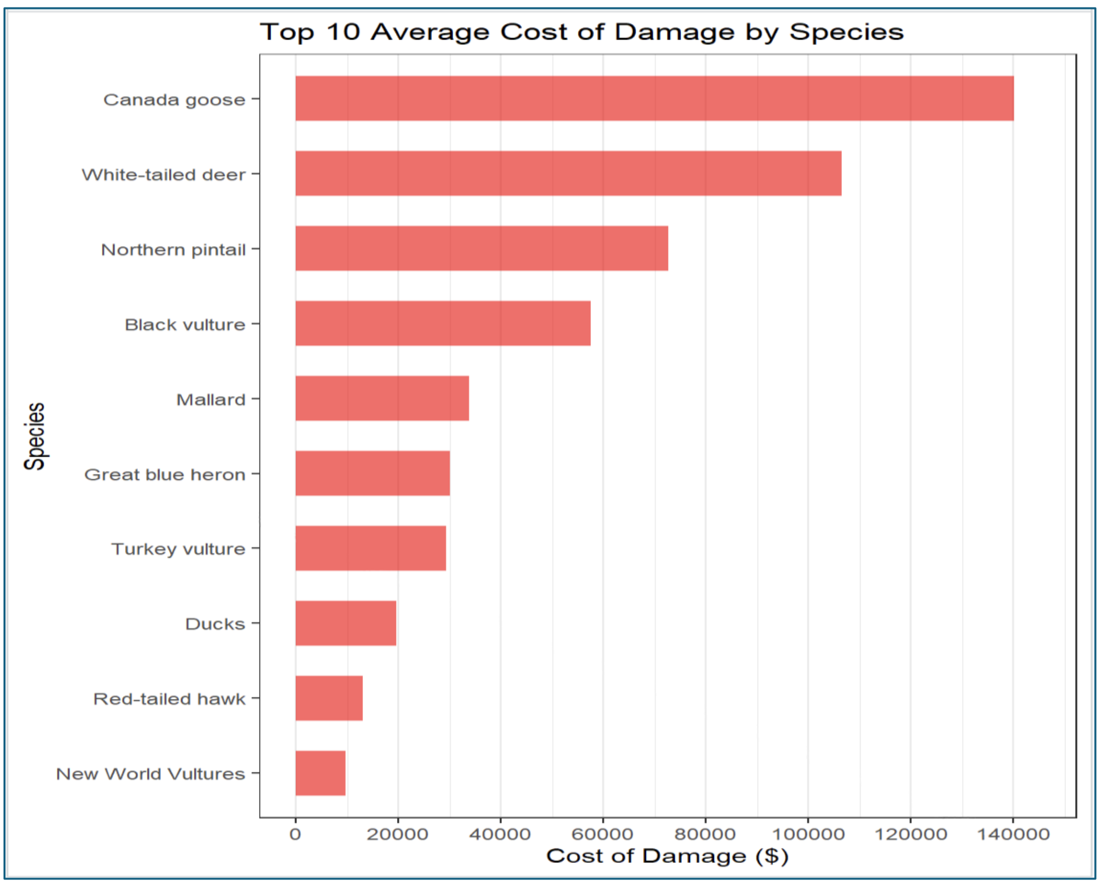

Figure 6. Shows how the severely damaging strikes tend to be concentrated into just a few species, which tend to be larger. The northern pintail is the exception and tends to fly in large flocks which may help to explain this result.

We adjusted the base threshold of .5 to .89 to improve the specificity, seen in figure X. We found this breakpoint based on an AUC curve. The model's predictive accuracy of 74.9% isn't ideal, but we attribute this to systemic issues in the dataset. For instance, our attempt to make a second model classifying damage between minor and substantial was fruitless due to key issues in the data. The Effect..Amount.of.damage column which was taken directly from the FAA and should classify repair costs into minor/medium/substantial/destroyed was relatively arbitrary. There was a large degree of overlap between these categories, which made using it as the dependent variable unviable since the model would be unable to accurately categorize data with such indistinguishable classifications.

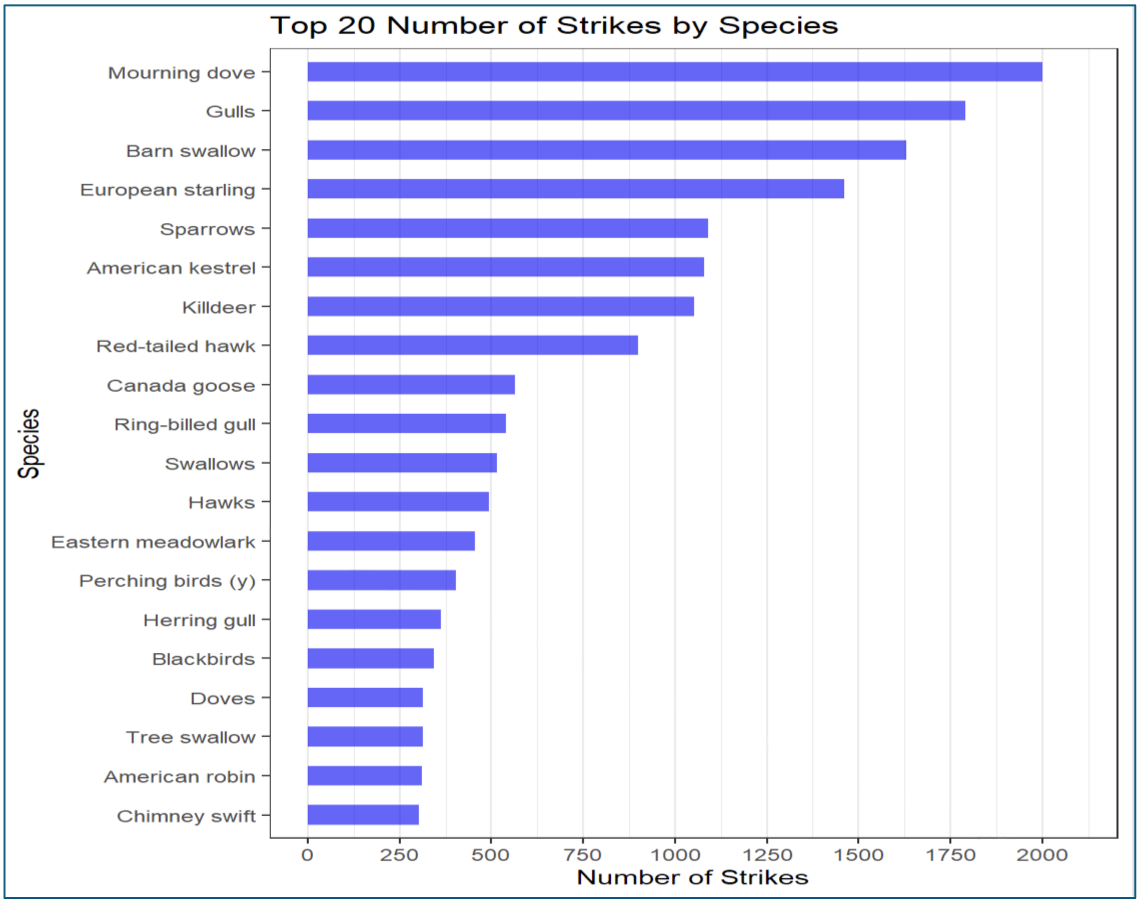

Figure 7. This plot shows the species that most commonly strikes aircraft, regardless of whether it was a damaging event. Unsurprisingly, small domestic birds are the most common offenders.

Conclusion

Ultimately, although our logistic model had somewhat limited predictive validity, in the process we discovered key gaps in data collection methodology, as well as potential avenues for collision reduction. Additional data such as aircraft heading and number of strikes, in the case of a flock, could potentially improve the model accuracy. Randomness is certainly a factor in animal collisions, nevertheless there are likely more predictive variables that could be incorporated, given additional data collection.

Figure 6 illustrates the disproportionately high impact mammal strikes have on damage costs, despite their low incidence compared with bird strikes. There may be a misalignment of incentive between airline companies which wish to avoid damage to their aircraft, and the airports which must spend their budget on runway perimeter security. Some spending on fence reinforcement or inspection during the fall months when white-tailed deer are most active could reduce the likelihood of finding them in the path of a landing aircraft.

Class imbalance and skewness of key metrics such as repair costs are significant hurdles that were partially mitigated but, consequently, limited our ability to effectively model the data. Perhaps drawing data from a larger timespan could reduce these limitations. The nature of investigating a rare event like bird strikes means outliers are not data noise to be removed, but the very object of study, difficult though it may be.

Further Research

Due to the infrequent occurrence of animal strikes we would like to see a larger data set. This could potentially allow for a better understanding of the likelihood of an animal strike and could potentially also lead to tracking trends over time, i.e. are animal strikes increasing or decreasing, is there a change in the frequency of strike for each animal species, are some locations improving their strike rate or cost over time.

In addition to expanding the period of data collection, we would also like to see a wider range of geographic data collected. It could be interesting to compare animal strike data collected globally, with analysis focused on variations between geographic regions. This would be a large undertaking, but it could identify which regions have the lowest strike numbers and potentially a better understanding of why those regions tend to have fewer strikes. In the end, other airports could adopt these mitigation strategies to reduce their own strike numbers, making flights safer and with less disruption to travelers.

Citations

1. Sodhi, N. (2002). Competition in the Air: Birds Versus Aircraft. *The Auk*, *119*(3), 587–595. https://academic.oup.com/auk/article/119/3/587/5561795?login=false

2. United States Department of Transportation: Federal Aviation Administration, *FAA Wildlife Strike Database*, accessed March 2023 https://wildlife.faa.gov/home

3. University of Michigan, Department of Ecology and Evolutionary Biology: Animal Diversity Web, *ADW Online Database*, accessed April 2024 https://animaldiversity.org/

4. Austin, P. C., & Steyerberg, E. W. (2017). Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. *Stat Methods Med Res*, *26*(2), 796–808. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5394463/

5. Armstrong, M. (2024). Most Airplane Accidents Happen During Landing. *Statista*, https://www.statista.com/chart/31529/most-airplane-accidents-happen-during-landing/

R Code

```
library(DAAG)
library(party)
library(rpart)
library(rpart.plot)
library(mlbench)
library(pROC)
library(tree)
library(caret)
library(e1071)
library(lime)
library(ggplot2)
library(dplyr)
library(tidyr)
library(forcats)

setwd("D:/STAT E109/Week 13/")

birds <- read.csv("faa_data_subset.csv")



top50 <- birds %>%
  group_by(Species=Wildlife..Species) %>%
  summarize(Count=n()) %>%
  arrange(desc(Count)) %>%
  head(50)
top50

damagedflights <- birds %>%
  group_by(birds$Effect..Indicated.Damage) %>%
  summarize(Count=n()) %>%
  arrange(desc(Count))
damagedflights
```

```
# Cleaning
birds$When..Time.of.day <- ifelse(nchar(birds$When..Time.of.day)==0, "Night", birds$When..Time.of.day)
birds$When..Time.of.day[birds$When..Time.of.day == "Dawn"] <- "Day"
birds$When..Time.of.day[birds$When..Time.of.day == "Dusk"] <- "Night"

birds$Feet.above.ground[birds$Feet.above.ground == ""] <- 0

birds$Collision.Date.and.Time <- as.Date(birds$Collision.Date.and.Time, format="%m/%d/%Y")
birds$season[format(birds$Collision.Date.and.Time, '%m') %in% c('03','04','05')] <- 1
birds$season[format(birds$Collision.Date.and.Time, '%m') %in% c('06','07','08')] <- 2
birds$season[format(birds$Collision.Date.and.Time, '%m') %in% c('09','10','11')] <- 3
birds$season[format(birds$Collision.Date.and.Time, '%m') %in% c('12','01','02')] <- 4
birds$season <- factor(birds$season)

birds$degreeofdamage[birds$Effect..Amount.of.damage..detailed. %in% c('Minor','Medium')] <- 0
birds$degreeofdamage[birds$Effect..Amount.of.damage..detailed. %in% c('Substantial','Destroyed')] <- 1
birds$degreeofdamage <- factor(birds$degreeofdamage)

# factor time of day
birds$timeofday[birds$When..Time.of.day == "Day"] <- 1
birds$timeofday[birds$When..Time.of.day == "Night"] <- 2


# Remove NAs
birds[!(is.na(birds$Typical.Weight) | birds$Typical.Weight==""), ]
birds.truncated <- birds[!(is.na(birds$Typical.Weight) | birds$Typical.Weight==""), ]


# Split data
set.seed(4321)
ind <- sample(2, nrow(birds.truncated), replace = T, prob = c(0.7, 0.3))
train <- birds.truncated[ind == 1,]
```

```
test <- birds.truncated[ind == 2,]


# One-Hot encoding for categorical data ? may not need

# initial logistic model, improved later
logisticm <- glm(as.factor(Effect..Indicated.Damage)~Aircraft..Number.of.engines + timeofday + Typical.Weight + Feet.above.ground, data = train, family = 'binomial')
summary(logisticm)

linearm <- lm(Cost..Total ~ Aircraft..Number.of.engines + timeofday + Typical.Weight + Feet.above.ground - 1, data = train)
summary(linearm)

###DY 05-01-24

#Logistic Regression model with all significant predictors by replacing timeofday with season
logisticm2 <- glm(factor(Effect..Indicated.Damage) ~ Aircraft..Number.of.engines + timeofday+Typical.Weight + Feet.above.ground + season, data=train, family='binomial')
summary(logisticm2)

cat('Baseline Rate : ', sum(train$Effect..Indicated.Damage=='No damage') / nrow(train), '\n')
p1 = predict(logisticm2, train, type='response')
pred1 = ifelse(p1>0.5, 'No damage', 'Caused damage')
confusionMatrix(factor(pred1), factor(train$Effect..Indicated.Damage), positive='No damage')


#rerun model with .89 threshold
pred2 = ifelse(p1>0.89, 'No damage', 'Caused damage')
confusionMatrix(factor(pred2), factor(train$Effect..Indicated.Damage), positive='No damage')

#r = multiclass.roc(factor(train$Effect..Indicated.Damage), p1, percent=T)
#roc = r[['rocs']]
#r1 = roc[[1]]
```

```
#plot.roc(r1, print.auc=T, print.thres=T, col='blue')
#auc(r1)
#coords(r1, 'best', ret=c('threshold', 'accuracy', 'sensitivity', 'specificity'))

par(mfrow=c(1,1))
ggplot(birds_vs_mammals, aes(x=Typical.Weight, y=log(Cost..Total), col=Wildlife..Animal.Category)) +
  geom_point() +
  geom_smooth(method='lm', se=0) +
  labs(title='Total Cost of Damage vs. Typical Weight, per Animal Category',
       y='Total Cost (Log)', x='Animal Weight (oz)')

###DY 05-02-24

#Re-classify amount of damage into two categories: 0=Minor, 1=Major (append chunk to Cleaning)
birds$degreeofdamage[birds$Effect..Amount.of.damage..detailed. %in% c('Minor','Medium')] <- 0
birds$degreeofdamage[birds$Effect..Amount.of.damage..detailed. %in% c('Substantial','Destroyed')] <- 1
birds$degreeofdamage <- factor(birds$degreeofdamage)

#Logistic Regression model to predict degree of damage
train2 <- train %>% filter(!is.na(degreeofdamage))
test2 <- test %>% filter(!is.na(degreeofdamage))

logisticm3 <- glm(degreeofdamage ~ Aircraft..Number.of.engines + Typical.Weight, data=train2, family='binomial')
summary(logisticm3)

cat('Baseline Rate : ', sum(train2$degreeofdamage==0) / nrow(train2), '\n')
p2 = predict(logisticm3, train2, type='response')
pred2 = ifelse(p2>0.5, 1, 0)
confusionMatrix(factor(pred2), train2$degreeofdamage, positive='1')



###DY 05-03-24
```

```
top50B <- birds.truncated %>%
  group_by(Species=Wildlife..Species) %>%
  summarise(Count=n(), Mean=mean(Cost..Total)) %>%
  arrange(desc(Count))

#Get top 20 total number of strikes by species
ggplot(top50B[1:20, ], aes(x=reorder(Species, Count), y=Count)) +
  geom_bar(stat='identity', fill='blue', alpha=.6, width=.5) +
  coord_flip() +
  labs(title='Top 20 Number of Strikes by Species', x='Species', y='Number of Strikes') +
  scale_y_continuous(limits=c(0,2100), n.breaks=10) +
  theme_bw() +
  theme(panel.grid.major.y=element_blank())

#Get top 10 average total cost of damage by species
top50C <- top50B %>% arrange(desc(Mean))
ggplot(top50C[1:10, ], aes(x=reorder(Species, Mean), y=Mean)) +
  geom_bar(stat='identity', fill='red', alpha=.6, width=.6) +
  coord_flip() +
  labs(title='Top 10 Average Cost of Damage by Species', x='Species', y='Cost of Damage ($)') +
  scale_y_continuous(limits=c(0,145000), n.breaks=10) +
  theme_bw() +
  theme(panel.grid.major.y=element_blank())
```