Q4

Hawkins_Kelsey

2024-12-09

~~~~ Question 4 ~~~~

Based on the flight phase ("Climb", "Landing Roll", "Approach", "Take-off run"), which phase is correlated with the largest average number of bird strikes? Is there a casual relationship between flight phase and bird strikes?

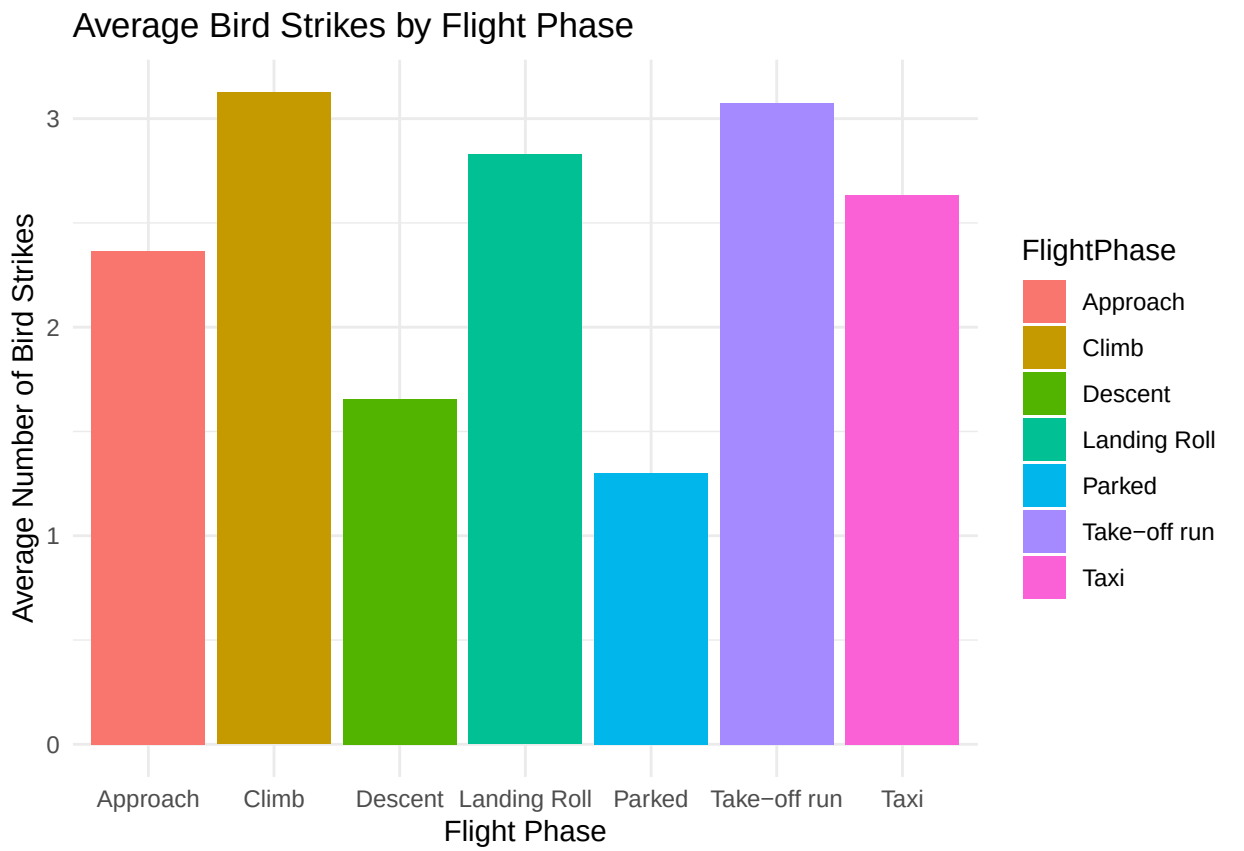
~~~~~ ANALYSIS ~~~~~

To answer this question, I plan to utilize a bayesian Generalized Linear Model (GLM). I plan to use this to answer the correlation and causal relationship portion of the questions. To begin, I will utilize basic EDA approaches by plotting summary statistics about the data I plan to analyze. By grabbing the average bird strikes for each flight phase, I plan to look at the differences between the averages in each flight phase to gain insights on the data to set my priors for the GLM. I originally plan to utilize normal informative priors with a poisson family with a log link function to account for the non-linear relationship with the data. Once the EDA is complete, I plan to build multiple GLM models with different familys and priors to test the different effects on the model, while utilizing prior predictive checks and looking at the ESS and posterior intervals and distributions.

I chose this analysis because I am able to use a GLM to answer the correlation question as well as the causal relationship inference portion. I plan to be able to answer both questions with one model as we can utilize all the model summary and information drawn from the the outputs to answer each portion of the questions. I could have taken a frequentist approach to answer the questions, however a bayesian approach will allow me to deeper assess and tune the model to answer the question as best as possible. I plan to use variables that I believe are correlated with bird strikes in the sense that they are connected to the flight, flight information, and geographical information. To use the DAG in this sense to be able to control for confounders will help make my model and analysis stronger.

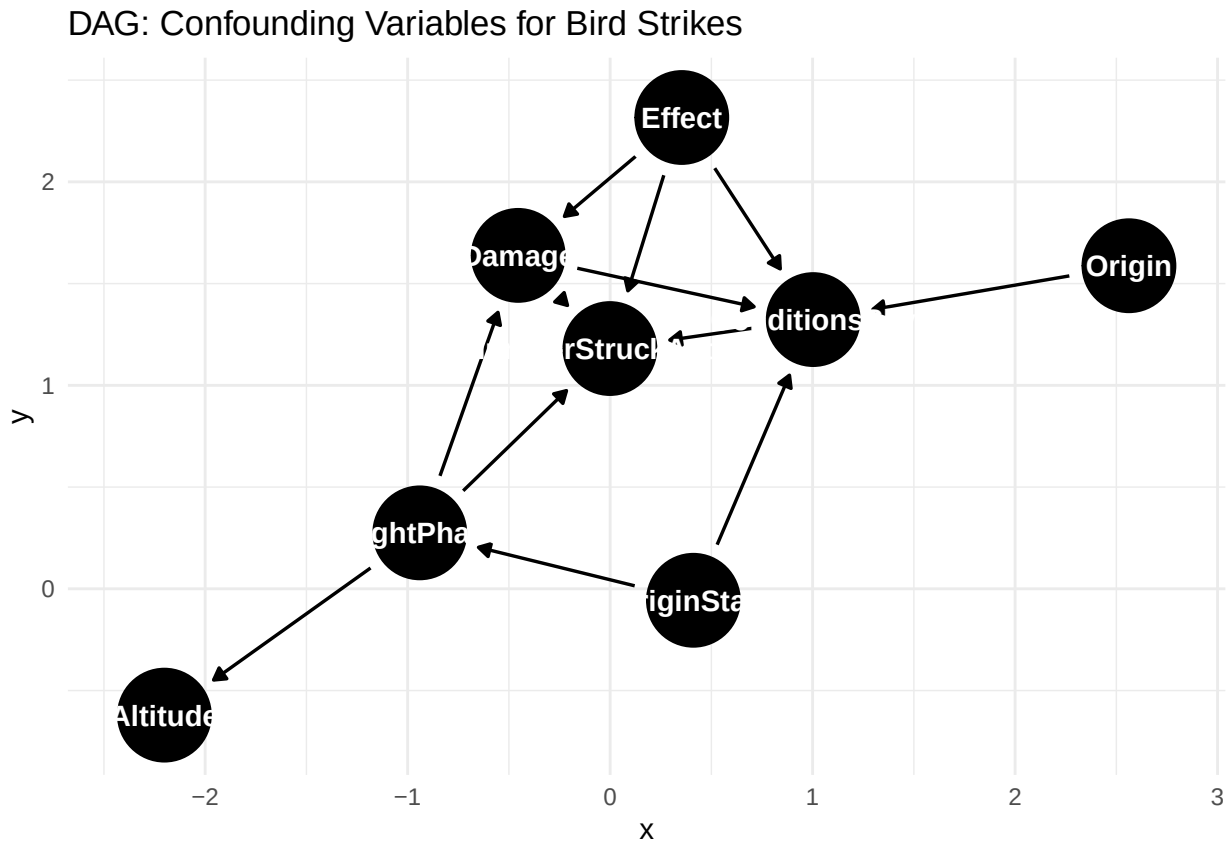
EDA

```
## # A tibble: 7 x 4
##   FlightPhase  MeanStrikes VarianceStrikes Count
##   <fct>              <dbl>           <dbl> <int>
## 1 Approach            2.37           139.  10382
## 2 Climb               3.13           386.   4429
## 3 Descent             1.66            19.9   776
## 4 Landing Roll        2.83            86.5  5047
## 5 Parked              1.3              0.9    10
## 6 Take-off run        3.07           121.   4711
## 7 Taxi                2.64           114.     74
```



DAG Creation & Quantification

In this DAG, we use four variables to test for a correlation between the number of birds struck. The variables are included: Effect, ConditionsSky, FlightPhase, OriginState, and Damage with FlightPhase being the exposure variable and the other three are confounders. We also take the previously made bar plot to see which flight phase had the most average bird strikes to gain information into the DAG. To gain insight into the relationships between variables, we use a linear regression model due to the variables being categorical. An ANOVA test could have been done, but I wanted to quantify the variables utilizing a 0.05 alpha to test for the P_Values being under the threshold. If the variable is under the threshold, we can conclude that there is a statistically significant relationship between the variable and NumberStruckActal. This information will be taken into the Bayesian parameter estimation model to estimate a causal relationship between flightPhase and numberStruckActual including the confounders found here.



Bayesian Parameter Estimation + Causal Inference

For the bayesian parameter estimation portion, we test multiple different priors to see which ones are the best fit to the dataset. After running the models, the one I use is as follows. The priors that were chosen are weakly informative priors of a normal distribution, with a mean of 0 and variation of 2.5. The family is a negative binomial regression model with a log link function. The response variable is NumberStruckActual that is count data showing the number of bird strikes. The predictors are flight phase, conditionsSky, Effect, Damage, and OriginState. The link function is a log function as we want a log-linear function of the predictors to ensure values are positive. The predictor coefficients, or the prior, is a normal prior for the coefficients, assuming that most predictors will have little or no influence on the outcome but with moderately large variation of 2.5. The prior intercept is a normal prior with a mean of 0 and variation of 5 so the model can have a larger baseline counts of bird strikes which shows more uncertainty. The model has 4 chains with 8000 iterations so that we can have a larger posterior sample size and having more robust posterior estimaties. The negative binomial family handles overdispersion where variance exceeds mean and response variable as bird strikes have very high variability across the dataset.

```
##
## Model Info:
##  function:     stan_glm
##  family:       neg_binomial_2 [log]
##  formula:      NumberStruckActual ~ FlightPhase + ConditionsSky + Effect + Damage +
##     OriginState
##  algorithm:    sampling
##  sample:       16000 (posterior sample size)
##  priors:       see help('prior_summary')
##  observations: 25429
```

```
##  predictors:   74
##
## Estimates:
##                                      mean   sd   10%   50%   90%
## (Intercept)                           1.7    0.1  1.6   1.7   1.8
## FlightPhaseClimb                      0.0    0.0  0.0   0.0   0.0
## FlightPhaseDescent                   -0.3    0.0 -0.4  -0.3  -0.3
## FlightPhaseLanding Roll               0.2    0.0  0.2   0.2   0.3
## FlightPhaseParked                    -1.0    0.4 -1.6  -1.0  -0.5
## FlightPhaseTake-off run               0.2    0.0  0.2   0.2   0.2
## FlightPhaseTaxi                      -0.1    0.1 -0.3  -0.1   0.1
## ConditionsSkyOvercast                 0.2    0.0  0.1   0.2   0.2
## ConditionsSkySome Cloud               0.0    0.0 -0.1   0.0   0.0
## EffectEngine Shut Down                1.2    0.1  1.1   1.2   1.4
## EffectNone                           -0.4    0.1 -0.5  -0.4  -0.3
## EffectOther                           0.3    0.1  0.2   0.3   0.4
## EffectPrecautionary Landing           0.4    0.1  0.3   0.4   0.5
## DamageNo damage                      -0.4    0.0 -0.4  -0.4  -0.3
## OriginStateAlabama                   -0.2    0.1 -0.3  -0.2  -0.1
## OriginStateAlaska                    -0.2    0.1 -0.4  -0.2  -0.1
## OriginStateAlberta                   -1.1    0.8 -2.1  -1.1   0.0
## OriginStateArizona                   -0.7    0.1 -0.9  -0.7  -0.6
## OriginStateArkansas                   0.2    0.1  0.1   0.2   0.4
## OriginStateBritish Columbia           0.5    0.3  0.1   0.5   0.9
## OriginStateCalifornia                -0.1    0.1 -0.2  -0.1  -0.1
## OriginStateColorado                  -0.4    0.1 -0.5  -0.4  -0.3
## OriginStateConnecticut                0.1    0.1 -0.1   0.1   0.2
## OriginStateDC                         0.0    0.1 -0.1   0.0   0.1
## OriginStateDelaware                   0.9    0.2  0.6   0.9   1.1
## OriginStateFlorida                   -0.5    0.1 -0.5  -0.5  -0.4
## OriginStateGeorgia                   -0.1    0.1 -0.2  -0.1   0.0
## OriginStateHawaii                     0.0    0.1 -0.1   0.0   0.1
## OriginStateIdaho                     -0.1    0.1 -0.3  -0.1   0.1
## OriginStateIllinois                  -0.2    0.1 -0.3  -0.2  -0.1
## OriginStateIndiana                   -0.1    0.1 -0.2  -0.1   0.0
## OriginStateIowa                      -0.4    0.1 -0.5  -0.4  -0.3
## OriginStateKansas                    -0.3    0.1 -0.5  -0.3  -0.2
## OriginStateKentucky                  -0.2    0.1 -0.3  -0.2  -0.1
## OriginStateLouisiana                 -0.1    0.1 -0.2  -0.1   0.0
## OriginStateMaine                     -0.4    0.2 -0.6  -0.4  -0.1
## OriginStateMaryland                  -0.2    0.1 -0.3  -0.2  -0.1
## OriginStateMassachusetts              0.0    0.1 -0.1   0.0   0.1
## OriginStateMichigan                  -0.5    0.1 -0.6  -0.5  -0.4
## OriginStateMinnesota                 -0.4    0.1 -0.5  -0.4  -0.2
## OriginStateMississippi               -0.1    0.1 -0.2  -0.1   0.1
## OriginStateMissouri                  -0.4    0.1 -0.5  -0.4  -0.3
## OriginStateMontana                   -0.6    0.2 -0.8  -0.6  -0.3
## OriginStateNebraska                  -0.3    0.1 -0.4  -0.3  -0.2
## OriginStateNevada                    -0.6    0.1 -0.7  -0.6  -0.5
## OriginStateNew Hampshire             -0.1    0.1 -0.3  -0.1   0.0
## OriginStateNew Jersey                -0.2    0.1 -0.3  -0.2  -0.1
## OriginStateNew Mexico                -0.5    0.1 -0.6  -0.5  -0.3
## OriginStateNew York                  -0.1    0.1 -0.2  -0.1   0.0
## OriginStateNewfoundland and Labrador -0.9    0.9 -2.0  -0.9   0.3
```

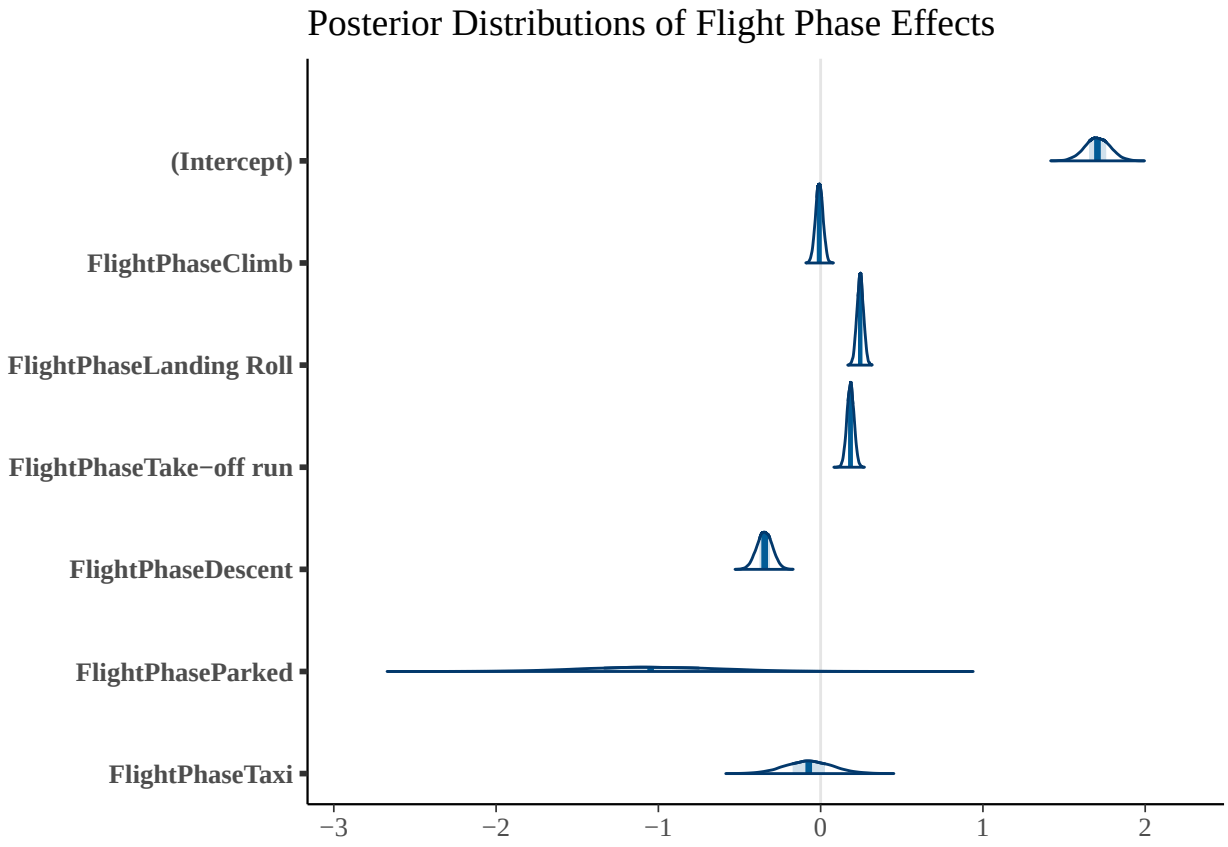
```
## OriginStateNorth Carolina           -0.2    0.1 -0.3  -0.2  -0.1
## OriginStateNorth Dakota             -0.4    0.1 -0.6  -0.4  -0.3
## OriginStateOhio                     -0.2    0.1 -0.3  -0.2  -0.1
## OriginStateOklahoma                 -0.6    0.1 -0.7  -0.6  -0.4
## OriginStateOntario                  -0.1    0.3 -0.5  -0.1   0.3
## OriginStateOregon                   -0.7    0.1 -0.8  -0.7  -0.6
## OriginStatePennsylvania              0.2    0.1  0.1   0.2   0.2
## OriginStatePrince Edward Island     -0.5    0.1 -0.7  -0.5  -0.3
## OriginStatePuerto Rico              -0.1    0.2 -0.3  -0.1   0.1
## OriginStateQuebec                    0.1    0.4 -0.4   0.0   0.6
## OriginStateRhode Island              0.1    0.1 -0.1   0.1   0.2
## OriginStateSaskatchewan             -0.6    1.4 -2.4  -0.7   1.1
## OriginStateSouth Carolina           -0.2    0.1 -0.3  -0.2  -0.1
## OriginStateSouth Dakota             -0.4    0.2 -0.6  -0.4  -0.2
## OriginStateTennessee                -0.1    0.1 -0.2  -0.1   0.0
## OriginStateTexas                    -0.2    0.1 -0.3  -0.2  -0.2
## OriginStateUtah                     -0.1    0.1 -0.2  -0.1   0.0
## OriginStateVermont                  -0.5    0.2 -0.7  -0.5  -0.2
## OriginStateVirgin Islands           -0.6    0.2 -0.9  -0.6  -0.3
## OriginStateVirginia                  0.8    0.1  0.7   0.8   0.9
## OriginStateWashington                0.3    0.1  0.2   0.3   0.4
## OriginStateWest Virginia            -0.8    0.2 -1.0  -0.8  -0.6
## OriginStateWisconsin                -0.4    0.1 -0.5  -0.4  -0.3
## OriginStateWyoming                  -0.3    0.2 -0.6  -0.3   0.0
## reciprocal_dispersion                1.0    0.0  1.0   1.0   1.1
##
## Fit Diagnostics:
##            mean   sd   10%   50%   90%
## mean_PPD 2.7    0.0  2.7   2.7   2.7
##
## The mean_ppd is the sample average posterior predictive distribution of the outcome variable (for de
##
## MCMC diagnostics
##                                    mcse Rhat n_eff
## (Intercept)                        0.0  1.0   1603
## FlightPhaseClimb                   0.0  1.0  11940
## FlightPhaseDescent                 0.0  1.0  13996
## FlightPhaseLanding Roll            0.0  1.0  11694
## FlightPhaseParked                  0.0  1.0  15181
## FlightPhaseTake-off run            0.0  1.0  10066
## FlightPhaseTaxi                    0.0  1.0  13530
## ConditionsSkyOvercast              0.0  1.0  13949
## ConditionsSkySome Cloud            0.0  1.0  13694
## EffectEngine Shut Down             0.0  1.0   9441
## EffectNone                         0.0  1.0   5007
## EffectOther                        0.0  1.0   6802
## EffectPrecautionary Landing        0.0  1.0   5619
## DamageNo damage                    0.0  1.0  14168
## OriginStateAlabama                 0.0  1.0   2111
## OriginStateAlaska                  0.0  1.0   2660
## OriginStateAlberta                 0.0  1.0  13773
## OriginStateArizona                 0.0  1.0   1906
## OriginStateArkansas                0.0  1.0   3528
## OriginStateBritish Columbia        0.0  1.0  10933
```

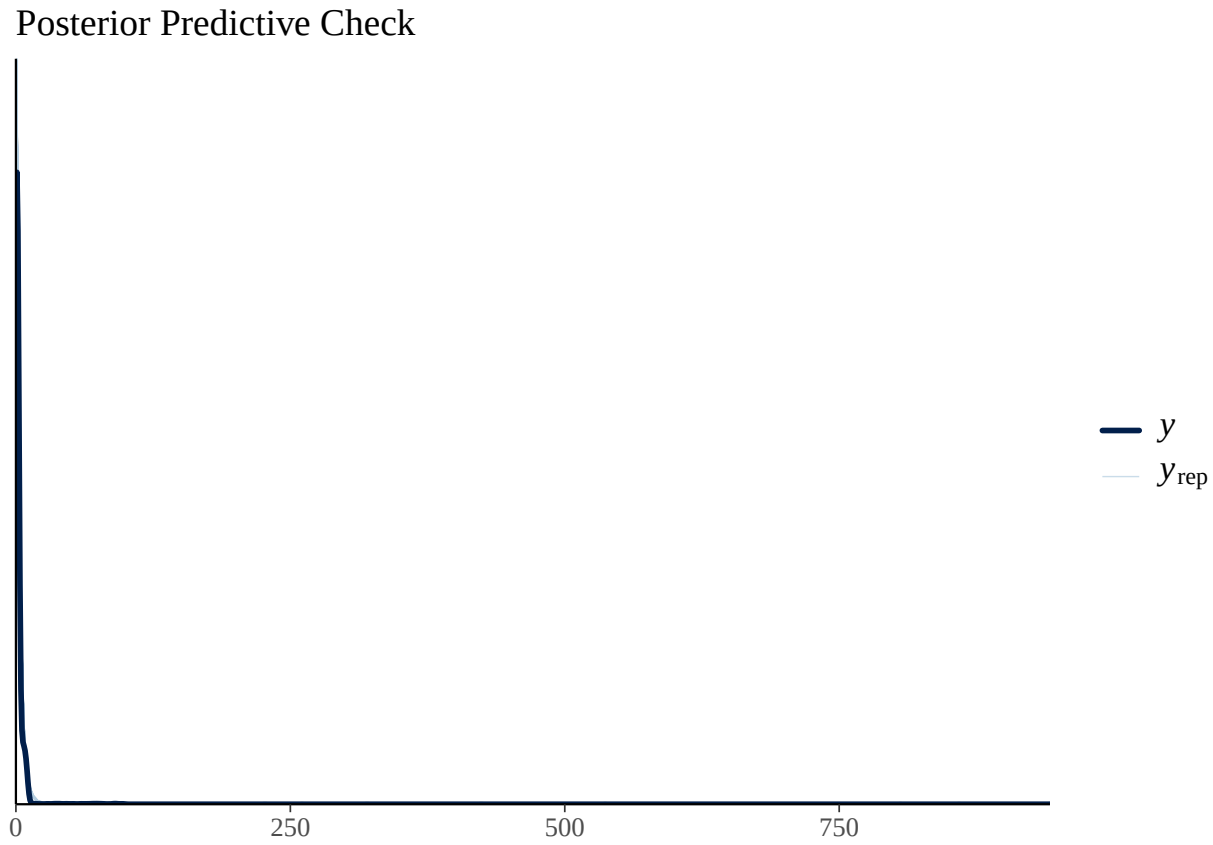
```
## OriginStateCalifornia                0.0  1.0    976
## OriginStateColorado                  0.0  1.0   1449
## OriginStateConnecticut               0.0  1.0   2082
## OriginStateDC                        0.0  1.0   1541
## OriginStateDelaware                  0.0  1.0   8054
## OriginStateFlorida                   0.0  1.0    958
## OriginStateGeorgia                   0.0  1.0   1525
## OriginStateHawaii                    0.0  1.0   1230
## OriginStateIdaho                     0.0  1.0   4913
## OriginStateIllinois                  0.0  1.0   1188
## OriginStateIndiana                   0.0  1.0   1987
## OriginStateIowa                      0.0  1.0   2519
## OriginStateKansas                    0.0  1.0   4206
## OriginStateKentucky                  0.0  1.0   1214
## OriginStateLouisiana                 0.0  1.0   1741
## OriginStateMaine                     0.0  1.0   7466
## OriginStateMaryland                  0.0  1.0   1719
## OriginStateMassachusetts             0.0  1.0   1763
## OriginStateMichigan                  0.0  1.0   1322
## OriginStateMinnesota                 0.0  1.0   1882
## OriginStateMississippi               0.0  1.0   3187
## OriginStateMissouri                  0.0  1.0   1194
## OriginStateMontana                   0.0  1.0   6463
## OriginStateNebraska                  0.0  1.0   1859
## OriginStateNevada                    0.0  1.0   3110
## OriginStateNew Hampshire             0.0  1.0   3555
## OriginStateNew Jersey                0.0  1.0   1448
## OriginStateNew Mexico                0.0  1.0   4058
## OriginStateNew York                  0.0  1.0   1088
## OriginStateNewfoundland and Labrador 0.0  1.0  12259
## OriginStateNorth Carolina            0.0  1.0   1251
## OriginStateNorth Dakota              0.0  1.0   4233
## OriginStateOhio                      0.0  1.0   1250
## OriginStateOklahoma                  0.0  1.0   2667
## OriginStateOntario                   0.0  1.0  11423
## OriginStateOregon                    0.0  1.0   1919
## OriginStatePennsylvania              0.0  1.0   1109
## OriginStatePrince Edward Island      0.0  1.0   4814
## OriginStatePuerto Rico               0.0  1.0   4818
## OriginStateQuebec                    0.0  1.0  13171
## OriginStateRhode Island              0.0  1.0   3389
## OriginStateSaskatchewan              0.0  1.0  14953
## OriginStateSouth Carolina            0.0  1.0   2884
## OriginStateSouth Dakota              0.0  1.0   5689
## OriginStateTennessee                 0.0  1.0   1287
## OriginStateTexas                     0.0  1.0    925
## OriginStateUtah                      0.0  1.0   1557
## OriginStateVermont                   0.0  1.0   8257
## OriginStateVirgin Islands            0.0  1.0   8491
## OriginStateVirginia                  0.0  1.0   1674
## OriginStateWashington                0.0  1.0   1822
## OriginStateWest Virginia             0.0  1.0   5171
## OriginStateWisconsin                 0.0  1.0   2165
## OriginStateWyoming                   0.0  1.0   7868
```

```
## reciprocal_dispersion               0.0  1.0  15648
## mean_PPD                            0.0  1.0  16896
## log-posterior                       0.1  1.0   6627
##
## For each parameter, mcse is Monte Carlo standard error, n_eff is a crude measure of effective sample
```

```
##                                          2.5%       97.5%
## (Intercept)                         1.55228424  1.860352815
## FlightPhaseClimb                   -0.05298594  0.035482423
## FlightPhaseDescent                 -0.43961288 -0.251205473
## FlightPhaseLanding Roll             0.20407949  0.283905427
## FlightPhaseParked                  -1.88352351 -0.169334699
## FlightPhaseTake-off run             0.14072733  0.227023222
## FlightPhaseTaxi                    -0.35987507  0.219872521
## ConditionsSkyOvercast               0.13109665  0.213387788
## ConditionsSkySome Cloud            -0.06352814  0.002815879
## EffectEngine Shut Down              0.97597448  1.455811170
## EffectNone                         -0.50236539 -0.290550048
## EffectOther                         0.10372412  0.408641533
## EffectPrecautionary Landing         0.28957067  0.536027636
## DamageNo damage                    -0.42733137 -0.327077114
## OriginStateAlabama                 -0.33153141  0.007033489
## OriginStateAlaska                  -0.42839349 -0.031930797
## OriginStateAlberta                 -2.69566028  0.613399527
## OriginStateArizona                 -0.91392244 -0.581230234
## OriginStateArkansas                 0.01747314  0.464542440
## OriginStateBritish Columbia        -0.07533192  1.168108496
## OriginStateCalifornia              -0.23486503 -0.009126559
## OriginStateColorado                -0.57570845 -0.298693681
## OriginStateConnecticut             -0.11077641  0.239203138
## OriginStateDC                      -0.15922889  0.139085895
## OriginStateDelaware                 0.47912369  1.279605006
## OriginStateFlorida                 -0.58027410 -0.347546622
## OriginStateGeorgia                 -0.24299729  0.040704799
## OriginStateHawaii                  -0.17020147  0.094477621
## OriginStateIdaho                   -0.40395678  0.178714697
## OriginStateIllinois                -0.35741242 -0.104879107
## OriginStateIndiana                 -0.27686967  0.039038399
## OriginStateIowa                    -0.58706183 -0.214700896
## OriginStateKansas                  -0.59610789 -0.081589336
## OriginStateKentucky                -0.29415908 -0.034980334
## OriginStateLouisiana               -0.25455032  0.057018166
## OriginStateMaine                   -0.73123479  0.031440247
## OriginStateMaryland                -0.30937915 -0.006950665
## OriginStateMassachusetts           -0.12933324  0.184470942
## OriginStateMichigan                -0.62900582 -0.356112182
## OriginStateMinnesota               -0.52260433 -0.194160075
## OriginStateMississippi             -0.32364877  0.140344588
## OriginStateMissouri                -0.55393594 -0.297593282
## OriginStateMontana                 -0.94853343 -0.189357969
## OriginStateNebraska                -0.42250664 -0.102715281
## OriginStateNevada                  -0.82298900 -0.375808794
## OriginStateNew Hampshire           -0.34679552  0.116381828
## OriginStateNew Jersey              -0.35642481 -0.075692291
```

```
## OriginStateNew Mexico                -0.73613653 -0.188933915
## OriginStateNew York                  -0.22168243  0.019244329
## OriginStateNewfoundland and Labrador -2.48611379  1.021679557
## OriginStateNorth Carolina            -0.31997824 -0.048641273
## OriginStateNorth Dakota              -0.71175410 -0.158931521
## OriginStateOhio                      -0.36593488 -0.102486194
## OriginStateOklahoma                  -0.76584024 -0.373725796
## OriginStateOntario                   -0.72483444  0.531721268
## OriginStateOregon                    -0.84926960 -0.530019184
## OriginStatePennsylvania               0.04638110  0.291082834
## OriginStatePrince Edward Island      -0.80785286 -0.228961414
## OriginStatePuerto Rico               -0.39494934  0.204261589
## OriginStateQuebec                    -0.67975676  0.849327040
## OriginStateRhode Island              -0.15799352  0.290047848
## OriginStateSaskatchewan              -3.36292491  2.131311443
## OriginStateSouth Carolina            -0.38948758  0.016175754
## OriginStateSouth Dakota              -0.69763992 -0.068966788
## OriginStateTennessee                 -0.24166019  0.030451812
## OriginStateTexas                     -0.35935379 -0.131546805
## OriginStateUtah                      -0.24997621  0.038831312
## OriginStateVermont                   -0.85492873 -0.041889899
## OriginStateVirgin Islands            -1.03439775 -0.124949873
## OriginStateVirginia                   0.61055915  0.933149525
## OriginStateWashington                 0.10480352  0.424033463
## OriginStateWest Virginia             -1.08220129 -0.487230537
## OriginStateWisconsin                 -0.57296488 -0.223289677
## OriginStateWyoming                   -0.67997065  0.120216867
## reciprocal_dispersion                 1.01890562  1.061931512
```


Posterior Distributions of Flight Phase Effects
(Intercept)
FlightPhaseClimb
FlightPhaseLanding Roll
FlightPhaseTake−off run
FlightPhaseDescent
FlightPhaseParked
FlightPhaseTaxi
−3
−2
−1
0
1
2



~~~~~ RESULTS ~~~~~

From the results of the model, we can reach a conclusion and answer our research question. Overall, the model was able to find a relationship between flight phase and number of bird strikes while accounting for weather conditions, aircraft effects, damage levels, and origin states as possible confounders. Overall, the key findings were as follows. Climb has a posterior mean of near 0 which means there is very little evidence of this phase being associated with more bird strikes than the baseline. The landing roll has a posterior mean of 0.2 which means there is a significant positive association with bird strikes. The Take-off run has a posterior mean of 0.2 as there is also a significant increase in bird strikes during this phase. The last phase descent has a posterior mean is -0.3 which means there is a reduced risk of bird strikes relative to the baseline. The overcast conditions also have a positive association with bird strikes. The some cloud conditions have a posterior mean of 0 which means there is no significant association with bird strikes. The engine shut down and precautionary landing has a posterior mean of 1.2 which is a strong asosciation with increased bird strike counts. The origin state of delaware and montana both have a positive strong association with bird strike counts.

When we analyze the posterior predictive check graph, it seems like the data is extremely right skewed. Unfortunately a lot of the bird strikes are very small or 0 because most flights dont hit many birds. The prior family of a negative binomial should capture the overdispersion but it seems like the plot is very skewed. There seems to have a long tail due to extreme values occurring in the dataset. However, the peak at 0 means that the model captures the proportion of flights with zero or very few bird strikes. Even though the graph is not great, it does show that the model is well fit to the data that is present, even though its skewed.

Based on the prior distributions of landing roll and take off run, these are the flight phases most strongly associated with higher bird strike counts. These findings show that these phases are greater risks for bird strikes, alongside the confounders that can be taken into account to make data driven actions. For example,

environmental conditions and flight context influence likelihood of bird strikes. Weather conditions such as overcast and engine shutdowns contribute to more bird strikes.

To answer the causal portion of the question, the model will be utilized to assess causal effects. Climb and take-off run have positive coefficients which indicate a higher number of bird strikes relative to the baseline. The credible intervals do not cross zero which suggest a strong evidence of a causal relationship. Landing roll is still positive but has a smaller effect than the other two phases. Parked and Taxi have negative coefficients which means there are fewer bird strikes compared to baseline which mean these phases are less likely to contribute to bird strikes causally, since their credible intervals cross zero. From the DAG, we control the weather conditions, geographical location, damage and effects. When we add them into our model, we control for them to isolate the causal impact of the flight phase on number of bird strikes. Since the narrow credible intervals of climb (0.1 - 0.2) and Take-off run (0.14 - 0.22), these are key contributers to bird strikes. The bayesian model not only answered the correlation portion, but also led us to a causal inference conclusion.

~~~~~ DISCUSSION ~~~~~

The findings that were discussed in the results section covered the numerical analysis results, but there are certain limitations to this analysis. Not only can we not actually control for variables in flights as the weather can not be controlled, we have to accept that there is a lot of uncertainty in this analysis. We can quantify our causal analysis and correlations to our numerical analysis with uncertainty factored in, but in reality we can't control for all of those counfounders. The outputs, however, were in line with what I had previously assumed that has the most bird strikes. By confirming it with posterior predictive checks, we can ensure that the priors were fit for the model.