How to Get Away With [Bird] Murder

emma hermacinski and Andrew Robertson

28/10/2021

When given 4 data sets, our group had to decide between 4 options, one of which immediately stood out. There was a data set which the US Federal Aviation Administration compiled from 1990 to 1997, detailing Aircraft-Wildlife Collisions.

Birds hitting planes.

We quickly realized it was our civic duty to study this data set, in order to ascertain one certain bit of information—what are the best conditions to eliminate as many birds as possible?

To the uninformed, this goal may seem cruel or brutal, but it has been claimed and clearly shown to be true that these so-called "birds" are *nothing more than government spies.* This work we have done here is incredibly important in protecting the average civilian from tyranny.

```
its.a.bird <- read.csv("birds.csv")
head(its.a.bird)
```

```
##   opid                operator      atype               remarks phase_of_flt
## 1  AAL       AMERICAN AIRLINES      MD-80             NO DAMAGE      Descent
## 2  USA               US AIRWAYS FK-28-4000    2 BIRDS, NO DAMAGE.        Climb
## 3  AAL       AMERICAN AIRLINES  B-727-200                  <NA>     Approach
## 4  AAL       AMERICAN AIRLINES      MD-82                  <NA>        Climb
## 5  AAL       AMERICAN AIRLINES      MD-82             NO DAMAGE        Climb
## 6  GFT GULFSTREAM INTL AIRLINES      BE-99 FLT 714.  TIME = 1951Z Landing Roll
##   ac_mass num_engs              date time_of_day state height speed effect
## 1       4        2  9/30/1990 0:00:00       Night    IL   7000   250   <NA>
## 2       4        2 11/29/1993 0:00:00         Day    MD     10   140   None
## 3       4        3  8/13/1993 0:00:00         Day    TN    400   140   None
## 4       4        2  10/7/1993 0:00:00         Day    VA    100   200   None
## 5       4        2  9/25/1993 0:00:00         Day    SC     50   170   None
## 6       2        2  9/20/1993 0:00:00         Day    FL      0    40   None
##          sky                species birds_seen birds_struck
## 1   No Cloud UNKNOWN BIRD - MEDIUM       <NA>            1
## 2   No Cloud UNKNOWN BIRD - MEDIUM       2-10         2-10
## 3 Some Cloud  UNKNOWN BIRD - SMALL       2-10            1
## 4   Overcast  UNKNOWN BIRD - SMALL       <NA>            1
## 5 Some Cloud  UNKNOWN BIRD - SMALL       2-10            1
## 6 Some Cloud                 HAWKS       <NA>            1
```

After reading in the data frame, we investigated the first six rows of a data set with more than 19,000 rows and noted the vector types of each of the 17 variables. In order to make the most effective graphs, we decided to change a number of vectors from integer and character vectors into factor vectors. We made this choice because factor vectors allow us to label and sort variables quite easily, which we'd do a great deal throughout our project.

```r
its.a.bird$operator <- as.factor(its.a.bird$operator)
its.a.bird$time_of_day <- as.factor(its.a.bird$time_of_day)
its.a.bird$state <- as.factor(its.a.bird$state)
its.a.bird$atype <- as.factor(its.a.bird$atype)
its.a.bird$sky <- as.factor(its.a.bird$sky)
its.a.bird$phase_of_flt <- as.factor(its.a.bird$phase_of_flt)
its.a.bird$birds_seen <- as.factor(its.a.bird$birds_seen)
its.a.bird$birds_struck <- as.factor(its.a.bird$birds_struck)
its.a.bird$effect <- as.factor(its.a.bird$effect)

head(its.a.bird)
```

```
##   opid                 operator      atype               remarks phase_of_flt
## 1  AAL        AMERICAN AIRLINES      MD-80             NO DAMAGE      Descent
## 2  USA               US AIRWAYS FK-28-4000    2 BIRDS, NO DAMAGE.        Climb
## 3  AAL        AMERICAN AIRLINES  B-727-200                  <NA>     Approach
## 4  AAL        AMERICAN AIRLINES      MD-82                  <NA>        Climb
## 5  AAL        AMERICAN AIRLINES      MD-82             NO DAMAGE        Climb
## 6  GFT GULFSTREAM INTL AIRLINES      BE-99 FLT 714.  TIME = 1951Z Landing Roll
##   ac_mass num_engs               date time_of_day state height speed effect
## 1       4        2  9/30/1990 0:00:00       Night    IL   7000   250   <NA>
## 2       4        2 11/29/1993 0:00:00         Day    MD     10   140   None
## 3       4        3  8/13/1993 0:00:00         Day    TN    400   140   None
## 4       4        2  10/7/1993 0:00:00         Day    VA    100   200   None
## 5       4        2  9/25/1993 0:00:00         Day    SC     50   170   None
## 6       2        2  9/20/1993 0:00:00         Day    FL      0    40   None
##          sky               species birds_seen birds_struck
## 1   No Cloud UNKNOWN BIRD - MEDIUM       <NA>            1
## 2   No Cloud UNKNOWN BIRD - MEDIUM       2-10         2-10
## 3 Some Cloud  UNKNOWN BIRD - SMALL       2-10            1
## 4   Overcast  UNKNOWN BIRD - SMALL       <NA>            1
## 5 Some Cloud  UNKNOWN BIRD - SMALL       2-10            1
## 6 Some Cloud                 HAWKS       <NA>            1
```

Here's why we made each of the preceding character and integer vectors into factor vectors:

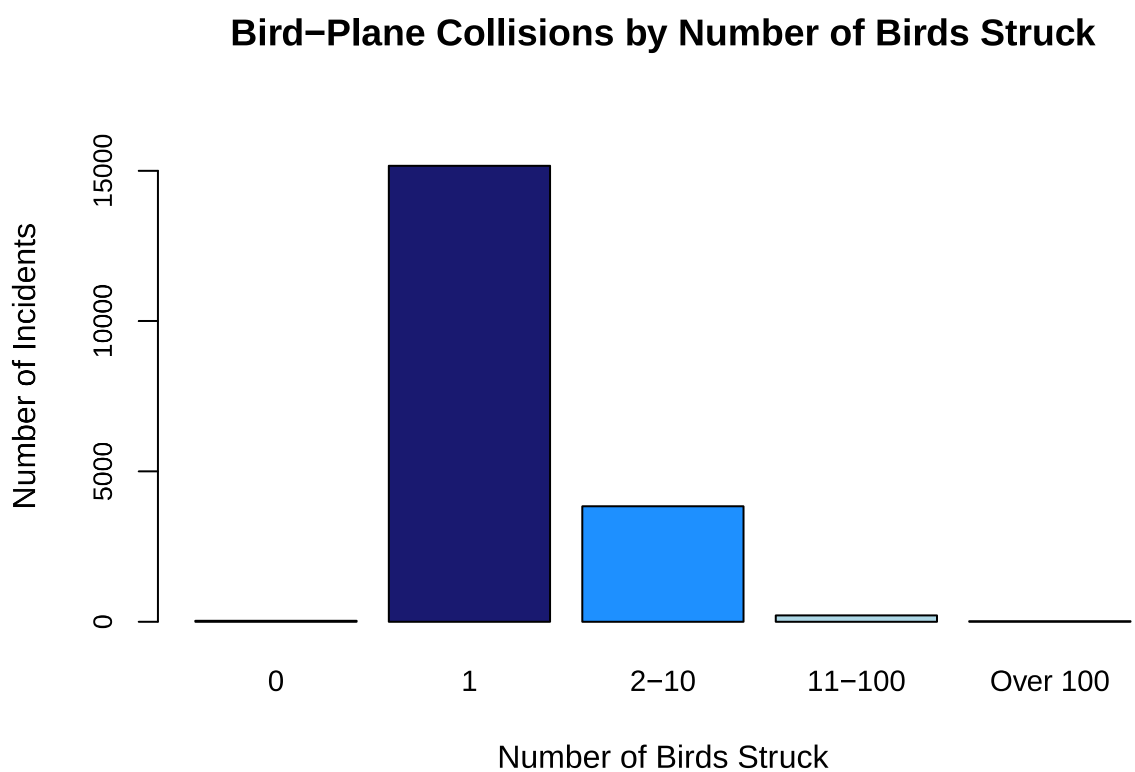
- **Operator:** There is a finite number of aircraft operators and accidents could be meaningfully grouped by operator.
- **Time_of_day:** The data set only gives four different times of day, meaning we could use this information to draw conclusions by formally categorizing it.
- **State:** Similar to "Operator," there are only 50 U.S. States, meaning we could group accidents by their location in a meaningful way.
- **Atype:** Similar to "Operator" and "State," there is a finite number of aircraft models, meaning we could utilize this data in factor form to figure out which planes have the most collisions.
- **Sky:** Similar to "Time_of_Day," there are only four sky conditions, meaning we could easily draw conclusions with this vector in factor form.
- **Birds_Seen:** Though this vector does contain numbers, it starts a character vector due to its providing of ranges—we can turn it into a factor vector to interpret these ranges.
- **Birds_Struck:** See Birds_Seen.
- **Effect** Similar to "Time_of_Day," there are only five levels of effect, so it's best that we use a factor vector to recognize and sort those levels.

```r
ouch.plane.levels <- c("0", "1", "2-10", "11-100", "Over 100")

its.a.bird$birds_struck <- factor(its.a.bird$birds_struck,
                                  levels = ouch.plane.levels,
                                  ordered = TRUE)

head(its.a.bird$birds_struck)
```

```
## [1] 1    2-10 1    1    1    1
## Levels: 0 < 1 < 2-10 < 11-100 < Over 100
```

While making preliminary graphs for this data set, we saw the birds_struck vector, detailing the number of birds struck in each collision. This, being exactly what we were looking for in order to get rid of government drones, was the best thing we could ask for. However, 11-100 was before 2-10, which was difficult to understand. In order to create understandable graphs, we sorted by the number of birds struck, which was up above.

```r
barplot(table(its.a.bird$birds_struck),
         main = "Bird-Plane Collisions by Number of Birds Struck",
        ylim = c(0,17000),
         xlab = "Number of Birds Struck",
         ylab = "Number of Incidents",
         cex.axis = 0.8,
         col = c("darkslateblue", "midnightblue", "dodgerblue", "lightblue", "powderblue"),
         cex.names = .9,)
```



Now, upon seeing this graph, there is one common reaction. Why do so many planes hit so few? Many planes hit none, despite the very parameters of the data set, many planes just one, and plenty just 2-10. In order to properly understand the ideal conditions for destroying these flying, tyrannical panopticons, we decided to simply isolate the patriotic pilots who struck down over 11 birds in one fell swoop. So, we created a new data set with just the incidents that destroyed 11 or more drones.

```r
bird.battery <- its.a.bird[its.a.bird$birds_struck == "11-100" | its.a.bird$birds_struck == "Over 100",]
bird.battery <- bird.battery[!is.na(bird.battery$birds_struck),]
head(bird.battery)
```

```
##     opid                operator      atype
## 29   UAL         UNITED AIRLINES  B-727-200
## 38   AAL       AMERICAN AIRLINES    DC-9-80
## 146  SDU         WESTAIR COMMUTER     BA-146
## 196  NWA      NORTHWEST AIRLINES   DC-10-40
## 282  NAE          NASHVILLE EAGLE SHORTS 360
## 474  SPA SIERRA PACIFIC AIRLINES  B-737-200
##
## 29
## 38  TWR REPTD THAT PREVIOUSLY LANDING AIRCRAFT HAD SEEN A FLOCK  OF BIRDS IN THE VICINITY OF RWY 24R
## 146
## 196
## 282
## 474                                                       DENT IN L INBOARD WING LE, IN LE OF L INBOA
##     phase_of_flt ac_mass num_engs              date time_of_day state height
## 29         Climb       4        3 5/12/1990 0:00:00         Day    NE     35
## 38      Approach       4        2 9/10/1991 0:00:00         Day    OH    100
## 146     Approach       4        4  9/5/1990 0:00:00         Day    CA     10
## 196 Landing Roll       4        3 9/13/1992 0:00:00         Day    CA      0
## 282 Take-off run       3        2  7/8/1991 0:00:00        Dusk    NY      0
## 474     Approach       4        2  9/6/1992 0:00:00       Night    UT   2000
##     speed effect        sky              species birds_seen birds_struck
## 29    140   None   Overcast           BLACKBIRDS       <NA>       11-100
## 38    140   None   No Cloud     EUROPEAN STARLING       <NA>       11-100
## 146   110   <NA>   No Cloud          ROCK PIGEON     11-100       11-100
## 196   120   None   No Cloud UNKNOWN BIRD - MEDIUM       <NA>     Over 100
## 282    NA   None Some Cloud  UNKNOWN BIRD - SMALL       <NA>       11-100
## 474   190   None   No Cloud                DUCKS       <NA>       11-100
```

This new data set only contains the planes that hit anywhere over 11 birds, regardless of any other factors. This number can range only from 11 to hundreds. This is what we were looking for exactly. Now we can use this data set to find out the conditions in which 11+ birds were hit. We do need to organize it once again though, so we'll do that right down below. It's good to have it cleanly ordered.

```r
ouch.plane.level2 <- c("11-100", "Over 100")

bird.battery$birds_struck <- factor(bird.battery$birds_struck,
                                    levels = ouch.plane.level2,
                                    ordered = TRUE)
```

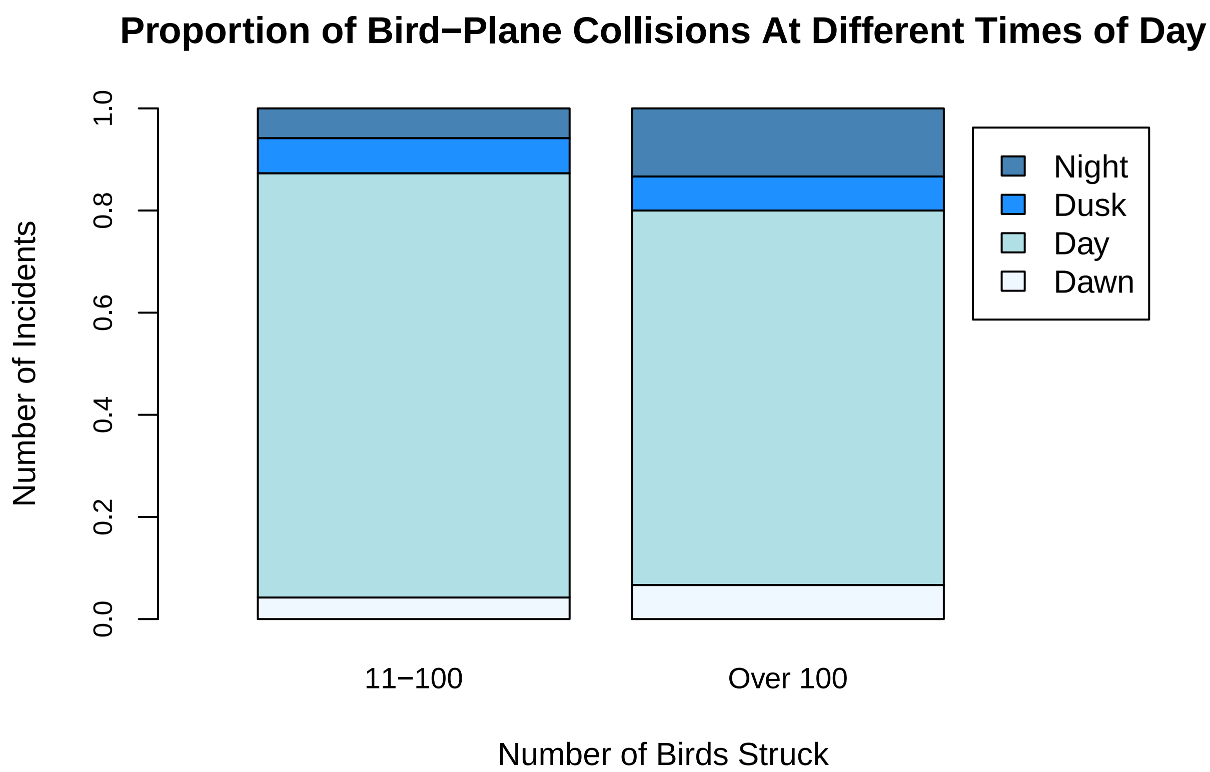
```r
levels(bird.battery$time_of_day)
```

```
## [1] "Dawn"  "Day"   "Dusk"  "Night"
```

After a bit more organizing in order to create the legend, we are able to make our first graph.

```r
barplot(prop.table(table(bird.battery$time_of_day, bird.battery$birds_struck),2),
        main = "Proportion of Bird-Plane Collisions At Different Times of Day",
        xlim = c(0,3),
        ylim = c(0,1),
        xlab = "Number of Birds Struck",
        ylab = "Number of Incidents",
```

```
        legend.text = c("Dawn", "Day", "Dusk", "Night"),
        cex.axis = 0.8,
        col = c("aliceblue", "powderblue", "dodgerblue", "steelblue"),
        cex.names = .9,)
```



This is an interesting graph. On both sides, the vast majority of collisions happen during the day. It appears to be around 80% on the 11-100 side, and around 65% on the 100+ side. On both sides, especially when you consider than there are 3 other categories, day is so strongly the majority. Yet, one more interesting thing to note is that the lower percentage on the 100+ means a higher percentage of other categories - there are more night and dawn, with around the same number of dusk.

```
bird.battery$ac_mass <- as.factor(bird.battery$ac_mass)
levels(bird.battery$ac_mass) <- c(1,2,3,4,5)
sort(bird.battery$ac_mass)
```

```
##   [1] 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3
##  [38] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4 4 4 4
##  [75] 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4
## [112] 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4
## [149] 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4
## [186] 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 5 5 5
## Levels: 1 2 3 4 5
```

```
levels(bird.battery$ac_mass) <- c("Tiny", "Small", "Medium", "Large", "Huge")
levels(bird.battery$effect)
```

```
## [1] "Aborted Take-off"      "Engine Shut Down"      "None"
## [4] "Other"                 "Precautionary Landing"
```

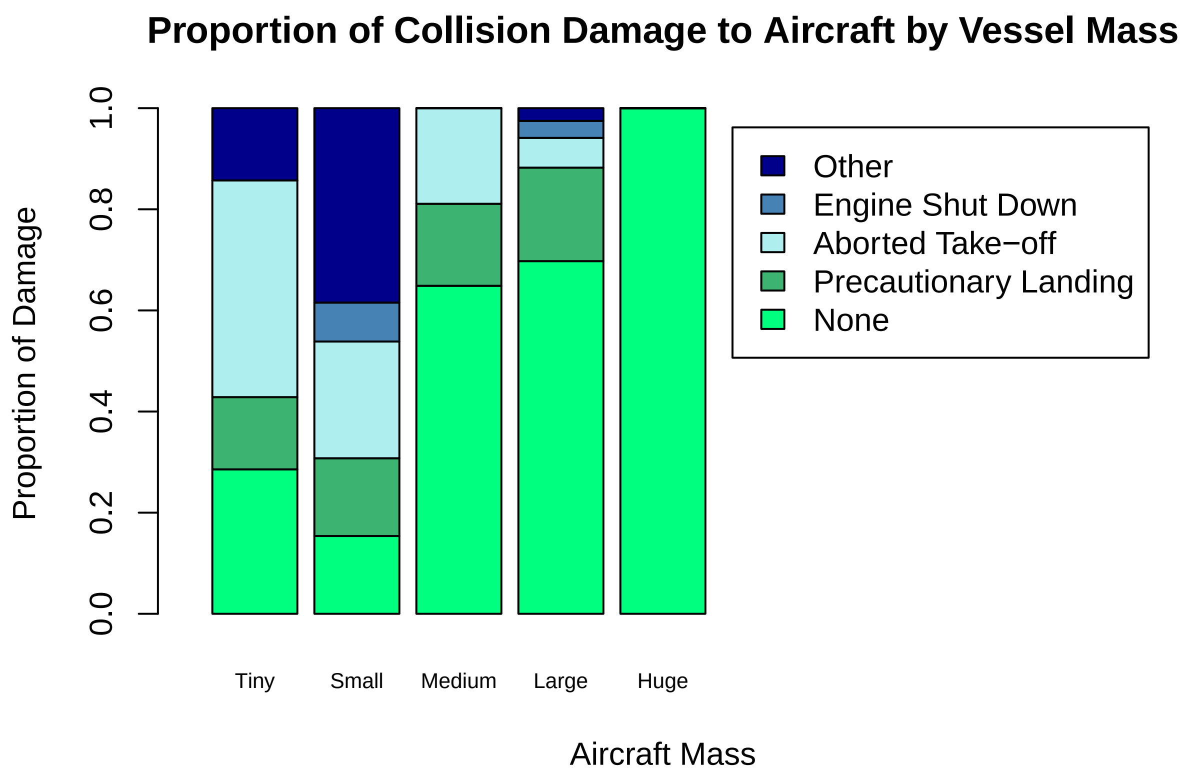
```
plane.effect <- c("None","Precautionary Landing","Aborted Take-off", "Engine Shut Down", "Other")
bird.battery$effect <- factor( bird.battery$effect,
```

```r
                              levels = plane.effect,
                              ordered = TRUE)
levels(bird.battery$effect)
```

```
## [1] "None"                  "Precautionary Landing" "Aborted Take-off"
## [4] "Engine Shut Down"      "Other"
```

In preparation for our next bar graph, we reorganize the labels for the mass vector to change the numbered groups into understandable ones labelled with words. We also reorder the levels of the effect vector from least to greatest amount of damage.

```r
barplot(prop.table(table(bird.battery$effect, bird.battery$ac_mass),2),
        main = "Proportion of Collision Damage to Aircraft by Vessel Mass",
        xlab = "Aircraft Mass",
        ylab = "Proportion of Damage",
        col = c("springgreen", "mediumseagreen", "paleturquoise", "steelblue", "darkblue"),
        legend.text = c("None","Precautionary Landing","Aborted Take-off", "Engine Shut Down", "Other")
        cex.names = .7,
        xlim = c(0,11))
```



This segmented bar graph shows the proportion of aircrafts damaged to various degrees in columns grouping the crafts by their size. With this graph, we can clearly see the differences between small and large aircraft in terms of the proportion of aircrafts severely damaged. Only around 25% of tiny aircraft and ~15% of small aircraft sustained no damage in bird-plane collisions, as opposed to ~65% of medium aircraft, ~70% of large aircraft, and 100% of huge aircraft. It is important to note that there are only 3 huge aircraft in this data set, so conclusions cannot be drawn from the biggest aircraft alone, but the high proportion of undamaged medium and large aircraft supports the claim that larger vessels are less likely to sustain damage in bird-plane collisions.

In terms of less safe collision outcomes, it is important to note that while the proportion of precautionary landings remains approximately the same across tiny, small, medium, and large craft, the proportion of aborted take-offs decreases as aircraft size increases.

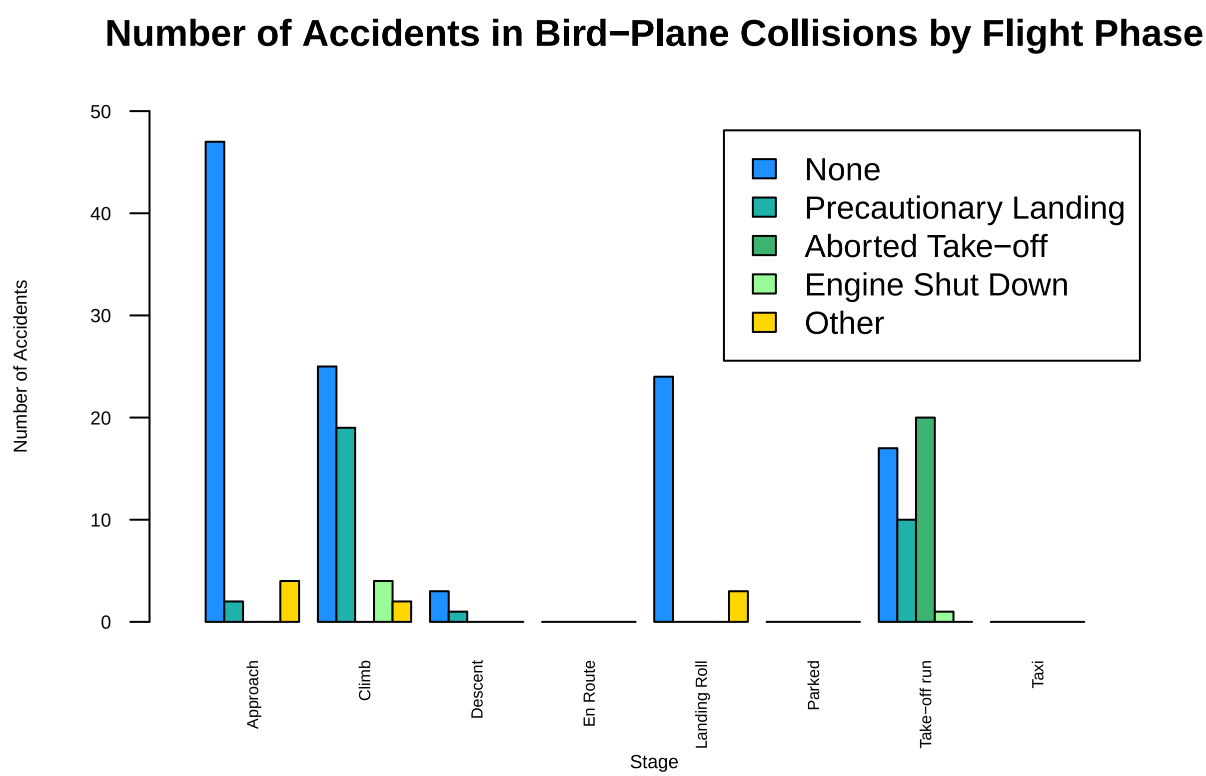
Additionally, it is important to note that around 40% of small aircraft sustained damage denoted "other." We would like to know what "other" denotes to obtain a better understanding of the way small vessels suffer damage in bird-plane collisions.

```
levels(bird.battery$phase_of_flt)
```

```
## [1] "Approach"     "Climb"        "Descent"      "En Route"     "Landing Roll"
## [6] "Parked"       "Take-off run" "Taxi"
```

Now, we print the levels of the factor vector bird.battery$phase_of_flt in preparation for the legend of our next bar graph.

```
barplot(table(bird.battery$effect, bird.battery$phase_of_flt),
        beside = TRUE,
        main = "Number of Accidents in Bird-Plane Collisions by Flight Phase",
        xlab = "Stage",
        ylab = "Number of Accidents",
        col = c("dodgerblue", "lightseagreen", "mediumseagreen", "palegreen", "gold"),
        legend.text = c("None","Precautionary Landing","Aborted Take-off", "Engine Shut Down", "Other")
        las = 2,
        cex.names = 0.5,
        cex.axis = 0.6,
        cex.lab = 0.6,
        xlim = c(0,50),
        ylim = c(0,50))
```

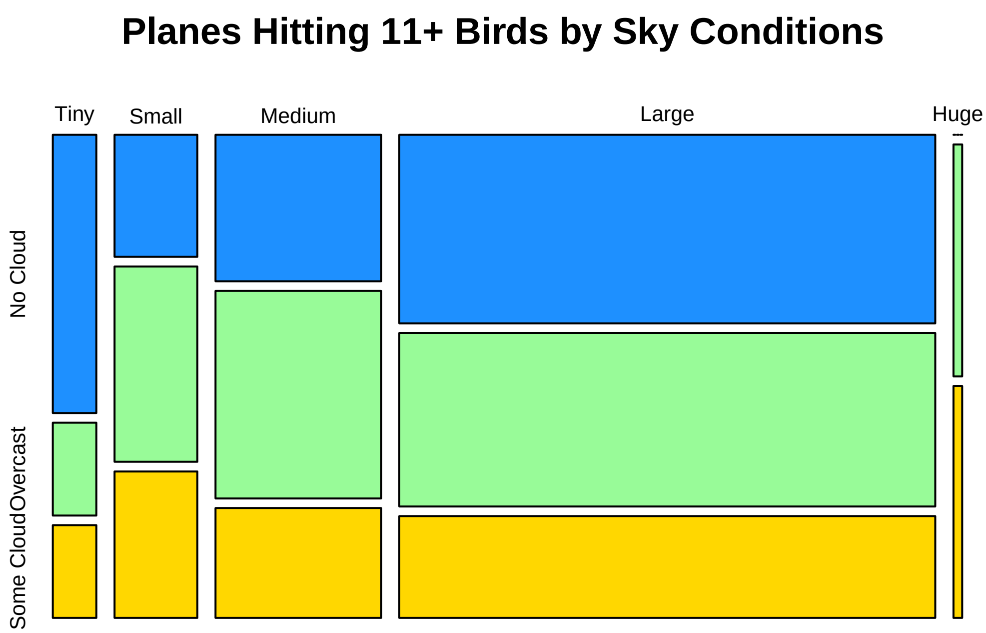


This bar graph shows us the bird-plane collisions of various magnitudes that occurred in each phase of the aircraft's flight. With the exception of a few minor instances in the descent phase, bird-plane collisions occur in the take-off run, climb, approach, and landing roll phases of flight. Initially, we were confused by the approach phase having so many more collisions than descent, but we discovered that approach is defined as the phase of flight between a plane's travel below 5,000 meters and its touching the runway, as opposed to descent which occurs higher in the air.

It seems that planes on their way down, or in the approach and landing roll phases, suffer far fewer consequences from collisions than planes taking off or climbing. Of course, the only time in which aborted take-offs occur is in the take-off phase, but precautionary landings are also relatively common in these initial phases. The climb and take-off phases are also the only ones that saw instances of engine shut down.

While the early phases of flight had the greatest rates of consequence, the approach and landing roll phases had the greatest rates of accidents that did not impact the plane. The approach phase saw a staggering 45+ instances of collision without consequence. It is notable that the approach and landing roll phases [as well as a bit in the climb phase] suffered damage denoted "other," which we once again cannot comment on. Overall, it seems that these final phases of flight are the best time to strike.

```
mosaicplot(bird.battery$ac_mass ~ bird.battery$sky,
           ylab = "",
           main = "Planes Hitting 11+ Birds by Sky Conditions ",
           col = c("dodgerblue", "palegreen", "gold"),
    xlab = ""
      )
```



This mosaic plot details the weather conditions during which incidents occurred. It details that, while there are some mild changes between sizes of aircraft, it does not make a significant difference what weather you fly in. It displays that, while oftentimes sunny weather is better, there isn't a huge percent difference. Similarly, we don't know what total percent of flights occur during day. If it's 99% of flights surveyed that happen during the day, but only this percent hit birds, well, you're far better off flying at night, but based off of the data set we can not draw a proper conclusion, simply that, from what we can tell, the time of the day does not matter much.

In conclusion, the optimal way to achieve our stated goal of eliminating as many birds as possible without suffering the consequences can be summed up in these Sacred Precepts:

The Sacred Precepts of American Freedom

or, how to remove a lot of birds from existence with no consequences.

1. Get a big plane. Smaller planes are restricted by cloud cover and more likely to suffer serious damage in a collision.

2. Time of day isn't particularly important. Fly any time of the day you can - a moment on the ground is a moment being watched.
3. Don't get greedy on the way up. Takeoff and climb phases of flight are by far the most dangerous phases of flight for your precious, fine-tuned counter-surveillance plane.
4. At the end of your flight, celebrate your successful voyage by going ham! The approach and descent phases of your journey are great times to assert yourself as an apex predator of the skies without resistance.

Follow these precepts and you're statistically likely to ensure the freedom of our nation.