Aircraft Wildlife collisions

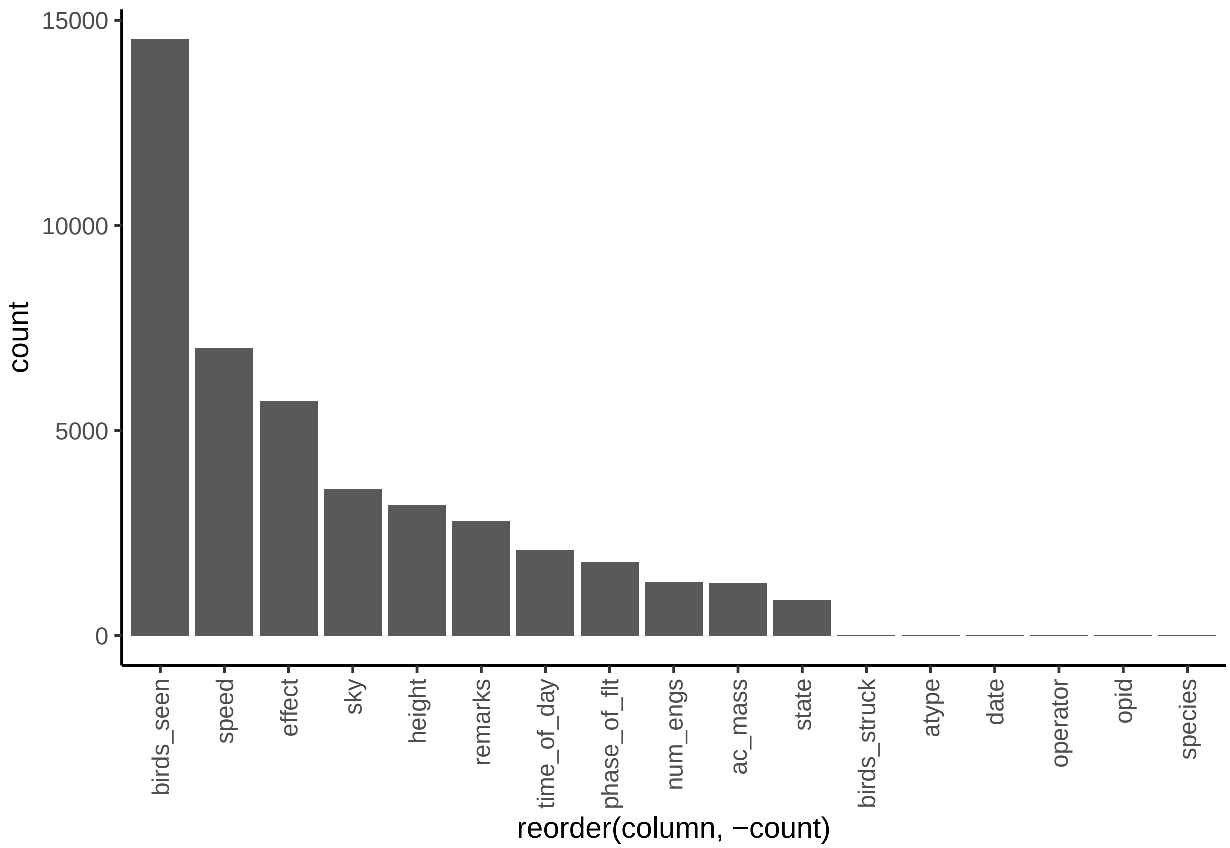
Alexander Flick & Vy Nguyen

21/07/2021

Data preparation

```
## Loading required package: airports
## Loading required package: cherryblossom
## Loading required package: usdata
## [1] 9412
## [1] "2-10"   "11-100"
```

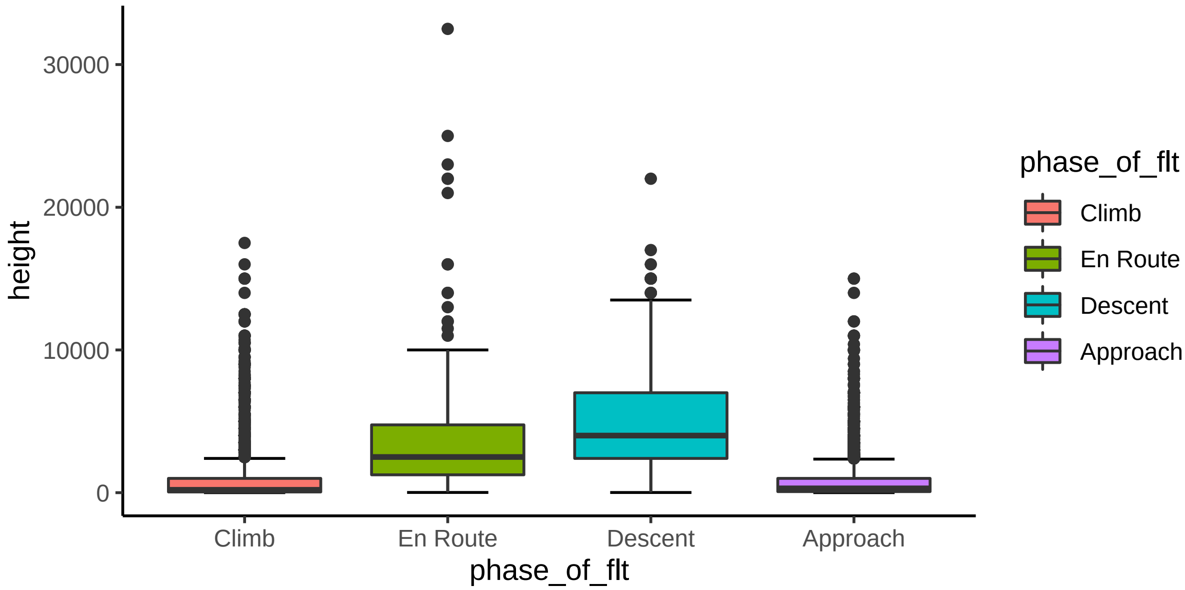
Handling missing values

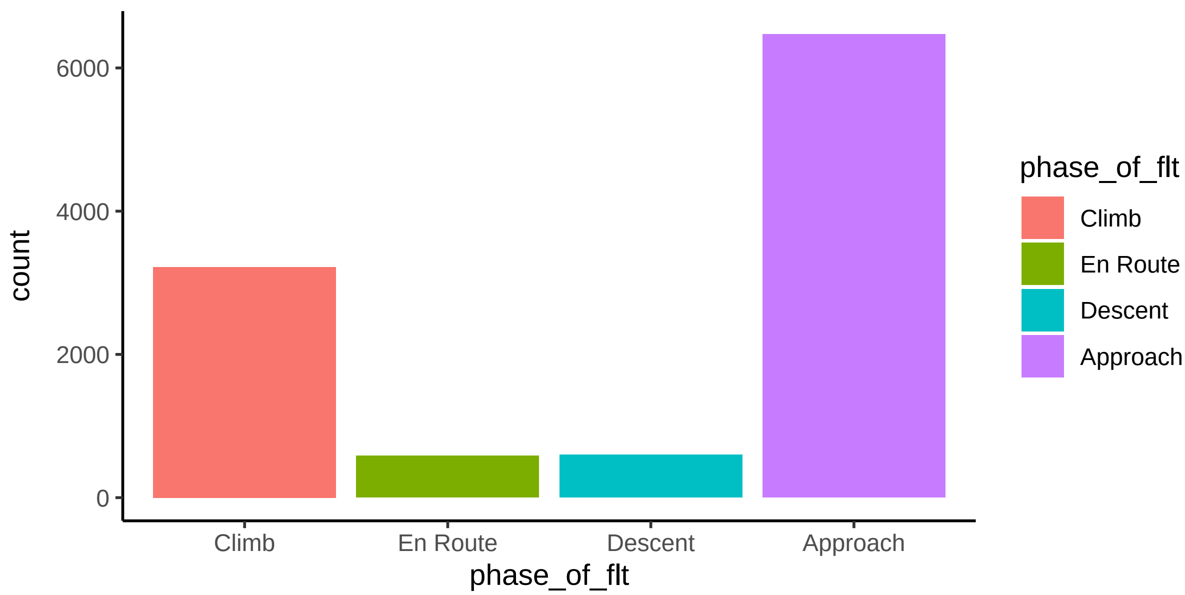


Analysis of strikes (birds_struck)

```
## Warning: Removed 1446 rows containing non-finite values (stat_boxplot).

## Warning: Removed 1446 rows containing non-finite values (stat_boxplot).
```

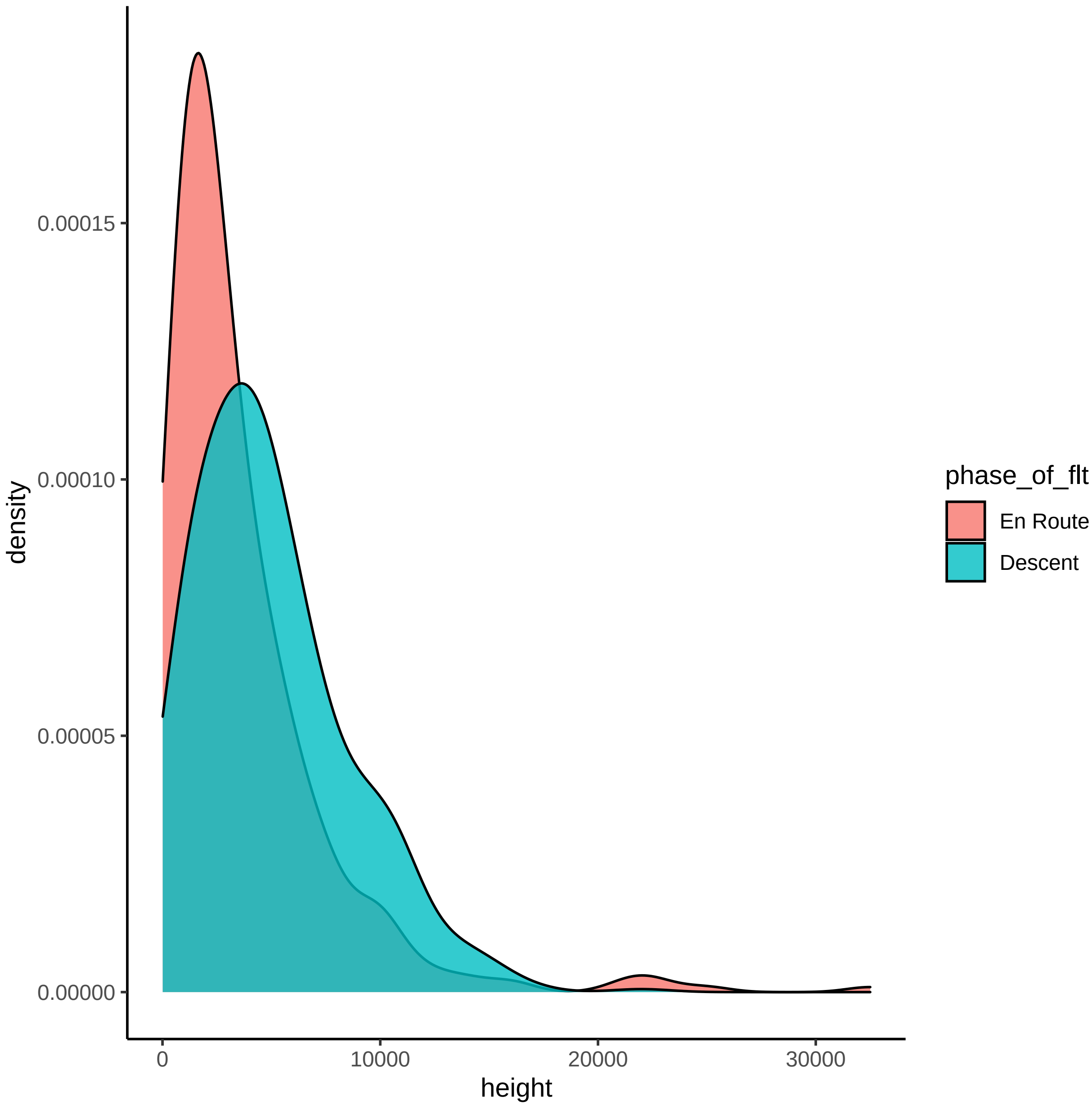




Closer look on En Rout and Descent

```
#plot height density for Descent and EnRoute
ggplot(birds[birds$phase_of_flt== "Descent" | birds$phase_of_flt=="En Route",]) +
  aes(x = height, group=phase_of_flt,fill=phase_of_flt ) +
  geom_density(adjust=1.5,alpha=.8)
```

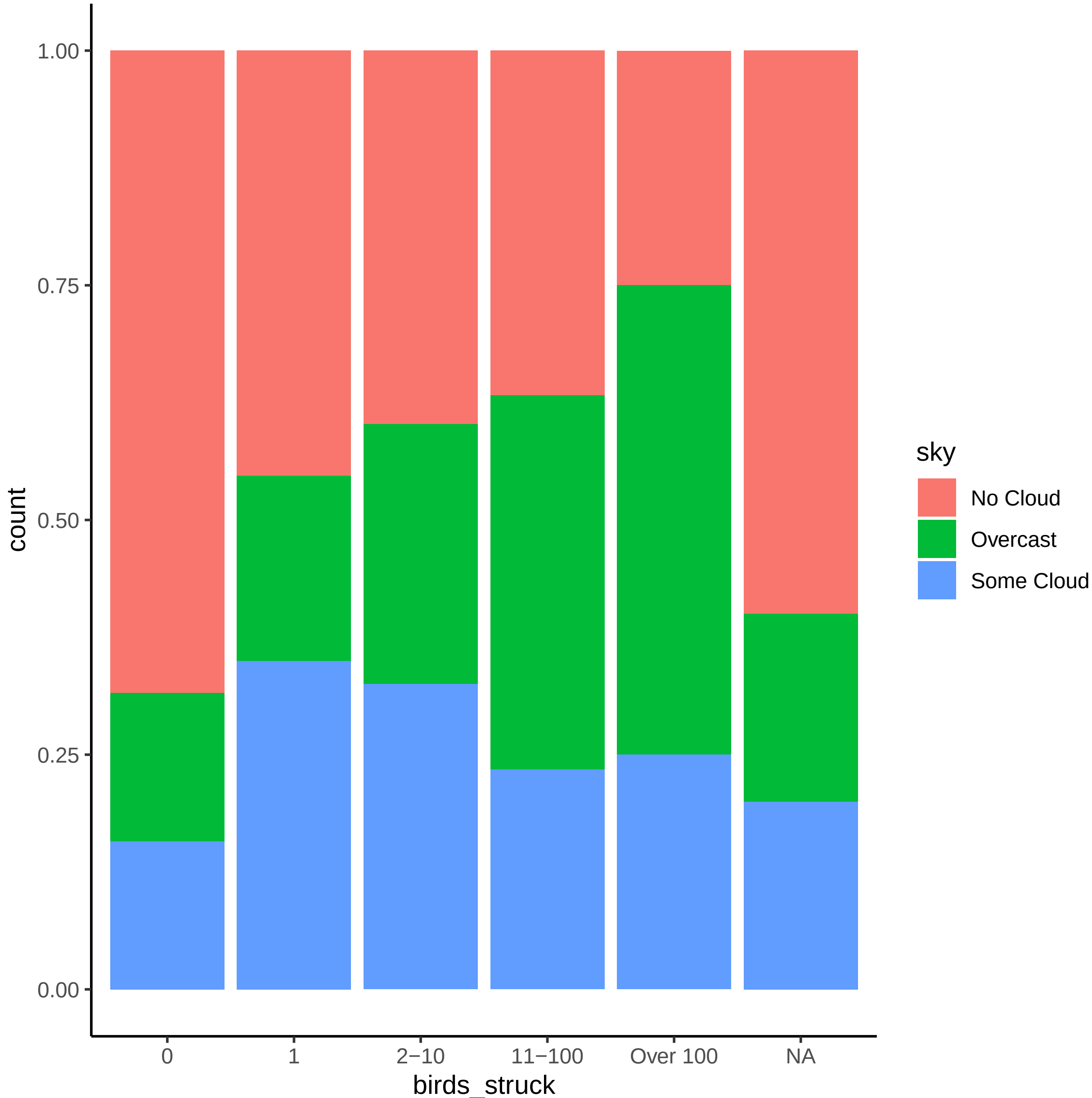
```
## Warning: Removed 2072 rows containing non-finite values (stat_density).
```




```
#Two sample t.test for height dependent on phase of flight ("Descent" and "En Route")
t.test(height~phase_of_flt, data = birds[as.integer(birds$phase_of_flt)>4 & as.integer(birds$phase_of_fl
```

```
##
##  Welch Two Sample t-test
##
## data:  height by phase_of_flt
## t = -5.4321, df = 764.3, p-value = 7.491e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1871.7879  -878.0419
## sample estimates:
## mean in group En Route  mean in group Descent
##               3609.116               4984.031
##
```

```
##                          0            1         2-10        11-100     Over 100
##   No Cloud   0.0024514426 0.8053931737 0.1806524609 0.0109372054 0.0005657175
##   Overcast   0.0011542901 0.7156598692 0.2566371681 0.0242400923 0.0023085802
##   Some Cloud 0.0007276255 0.7996604414 0.1899102595 0.0089740480 0.0007276255
```



For 0 birs struck we have highest percentage of no clouds, so clouds might have an affect on avoiding striking birds. Its about 20% higher compared to 1 bird struck. for birdsstruck by sky we can see that with increasing amount of birds the sky tends to be more more overcast, as no clouds and some clouds decreases

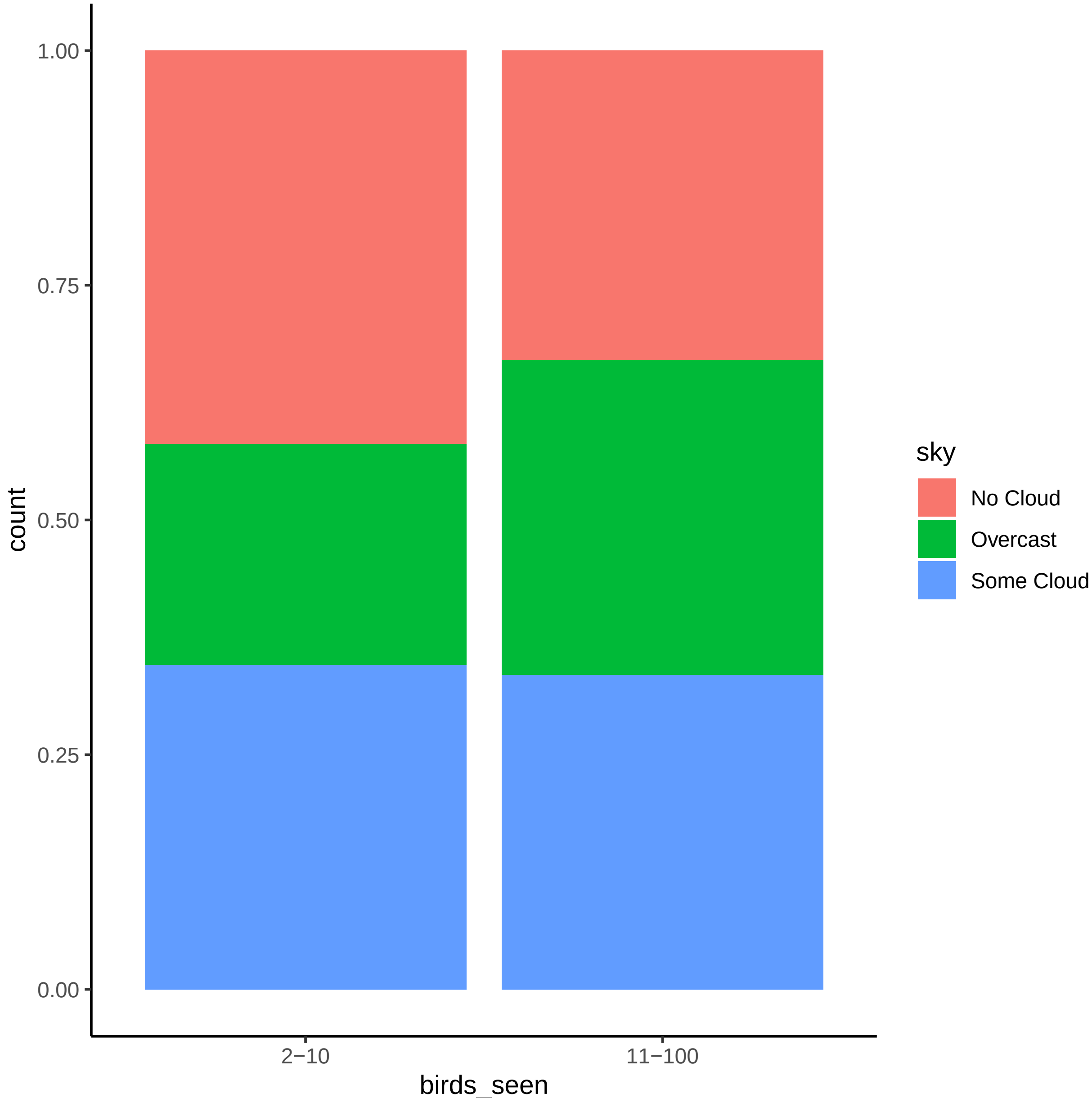
#Clouds with birds seen

there are only 2 levels in the data available for birds seen

```
##
##   2-10 11-100
##   3775    988
```

```
##
```

```
##                  2-10    11-100
##   No Cloud   0.8002646 0.1997354
##   Overcast   0.6889564 0.3110436
##   Some Cloud 0.7647510 0.2352490
```

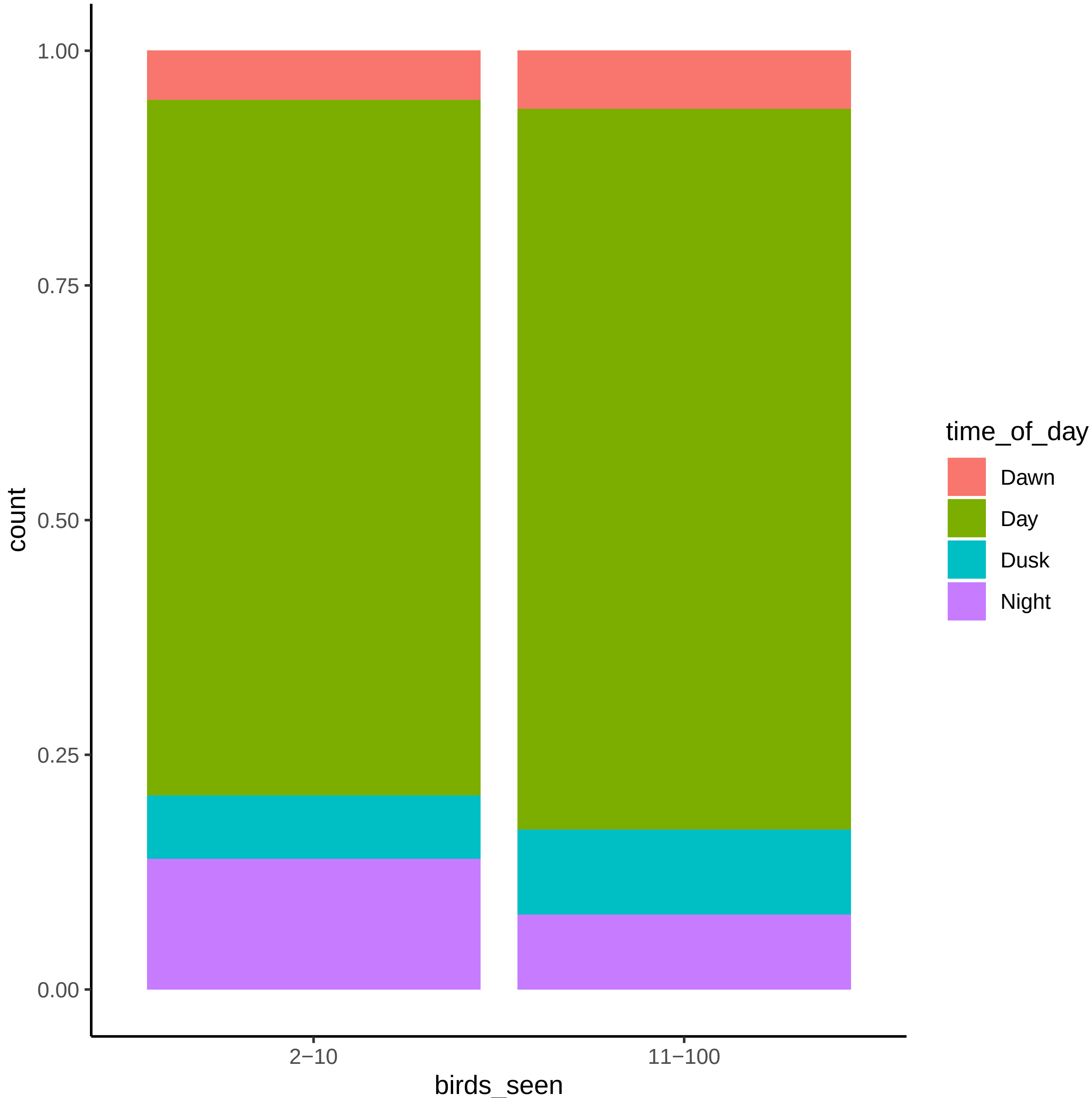


the proportion of some clouds is nearly the same, but pilots do more often see bigger groups of birds (11-100) when its overcast compared to small groups

```
##
##   2-10 11-100
##   3775    988
```

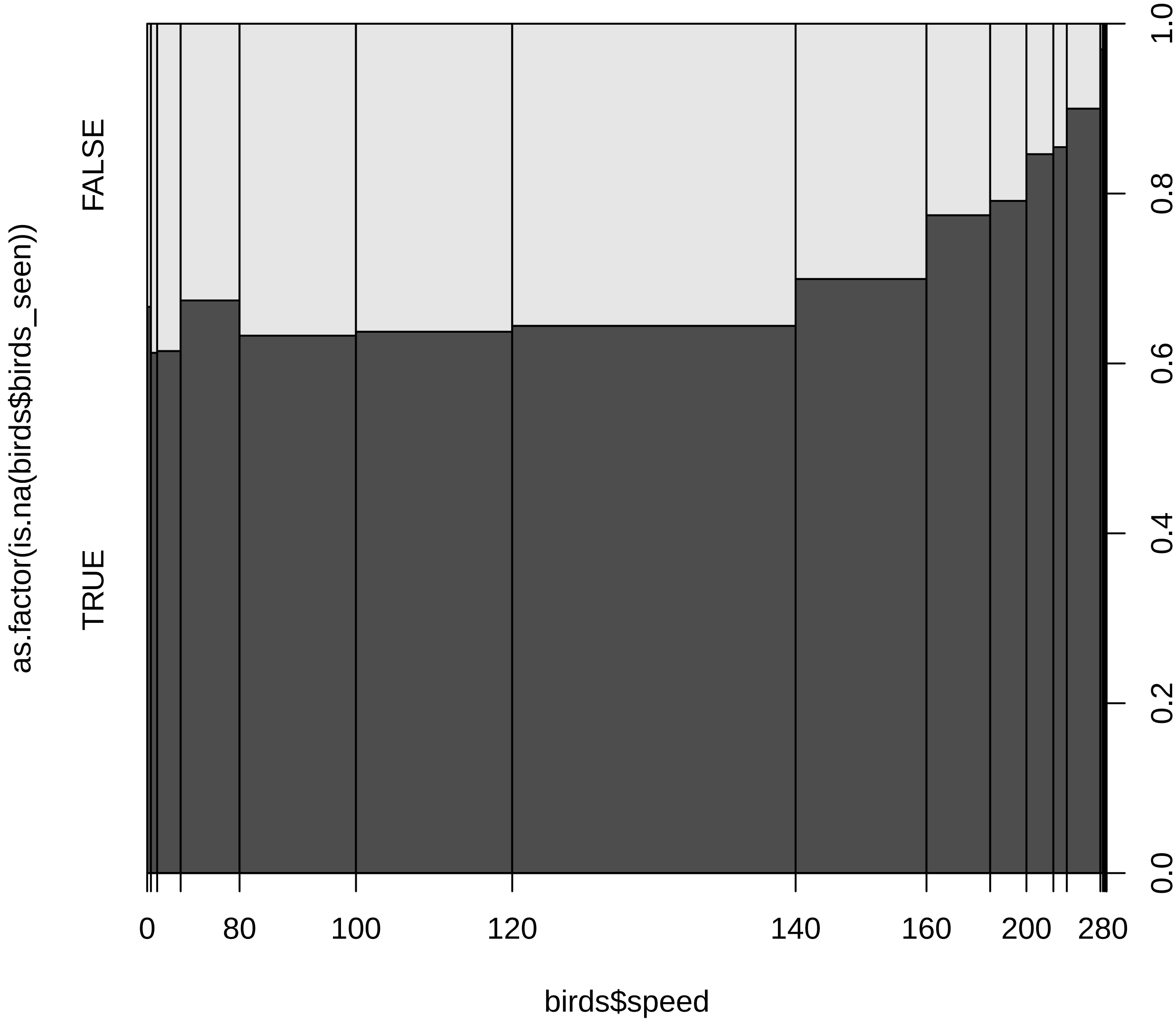
```
##
##             2-10    11-100
##   Dawn  0.7281553 0.2718447
##   Day   0.7533712 0.2466288
```

```
##    Dusk  0.7028986 0.2971014
##    Night 0.8468085 0.1531915
```



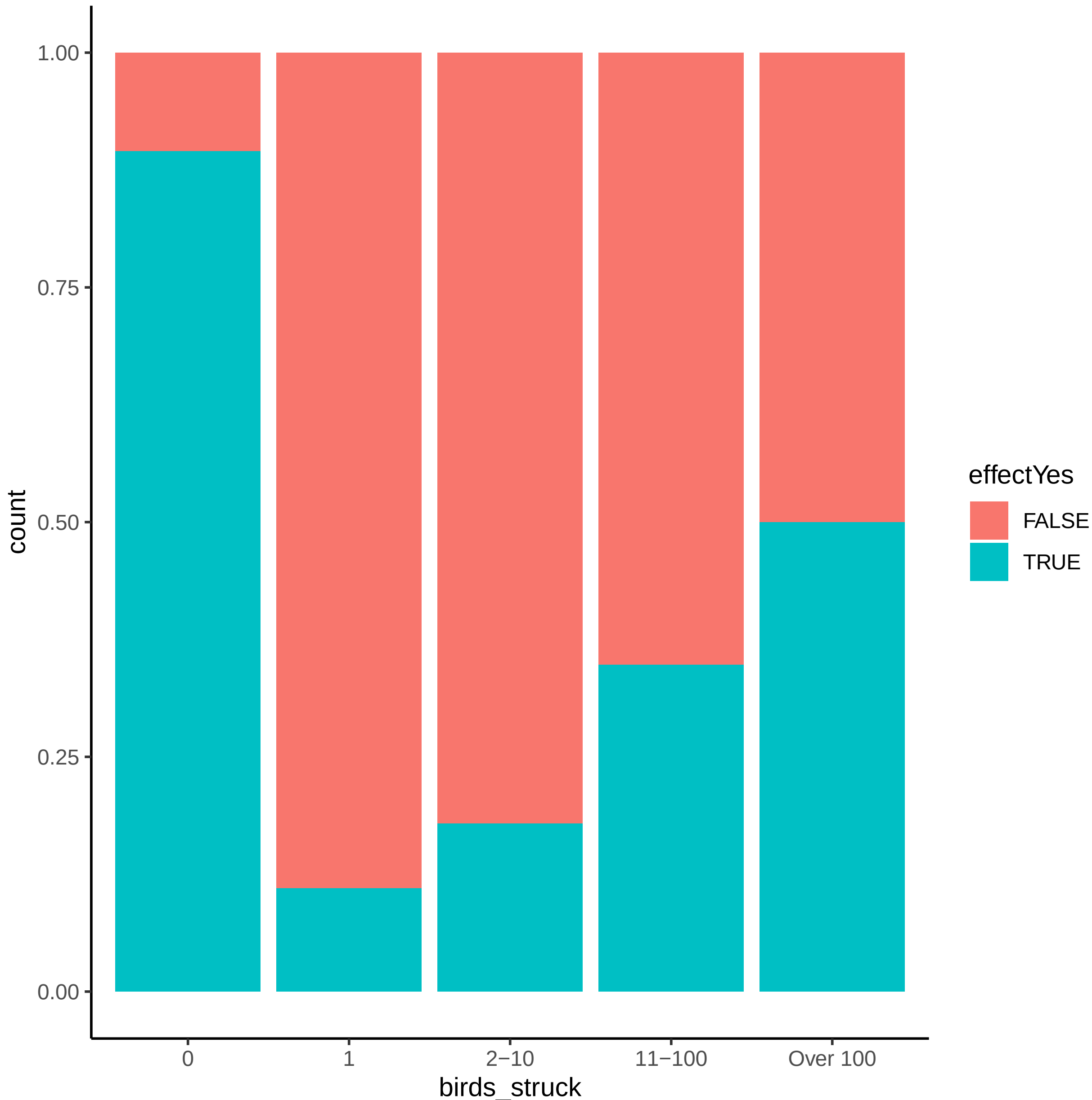
Birds struck out of birds seen

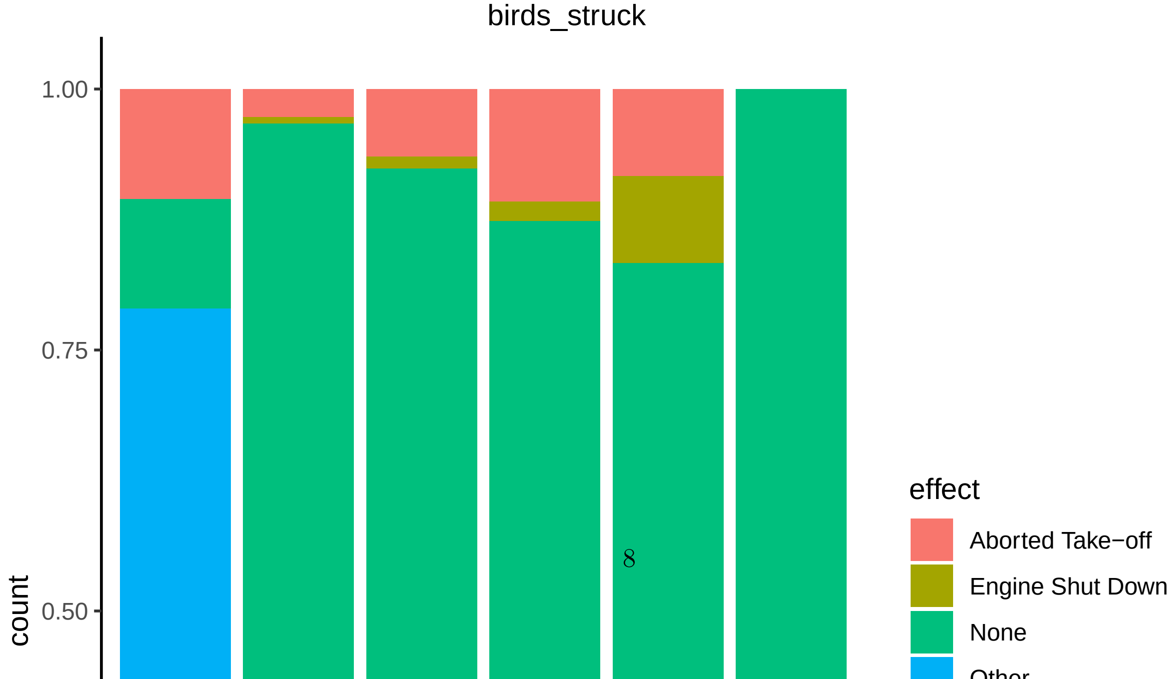
check speed

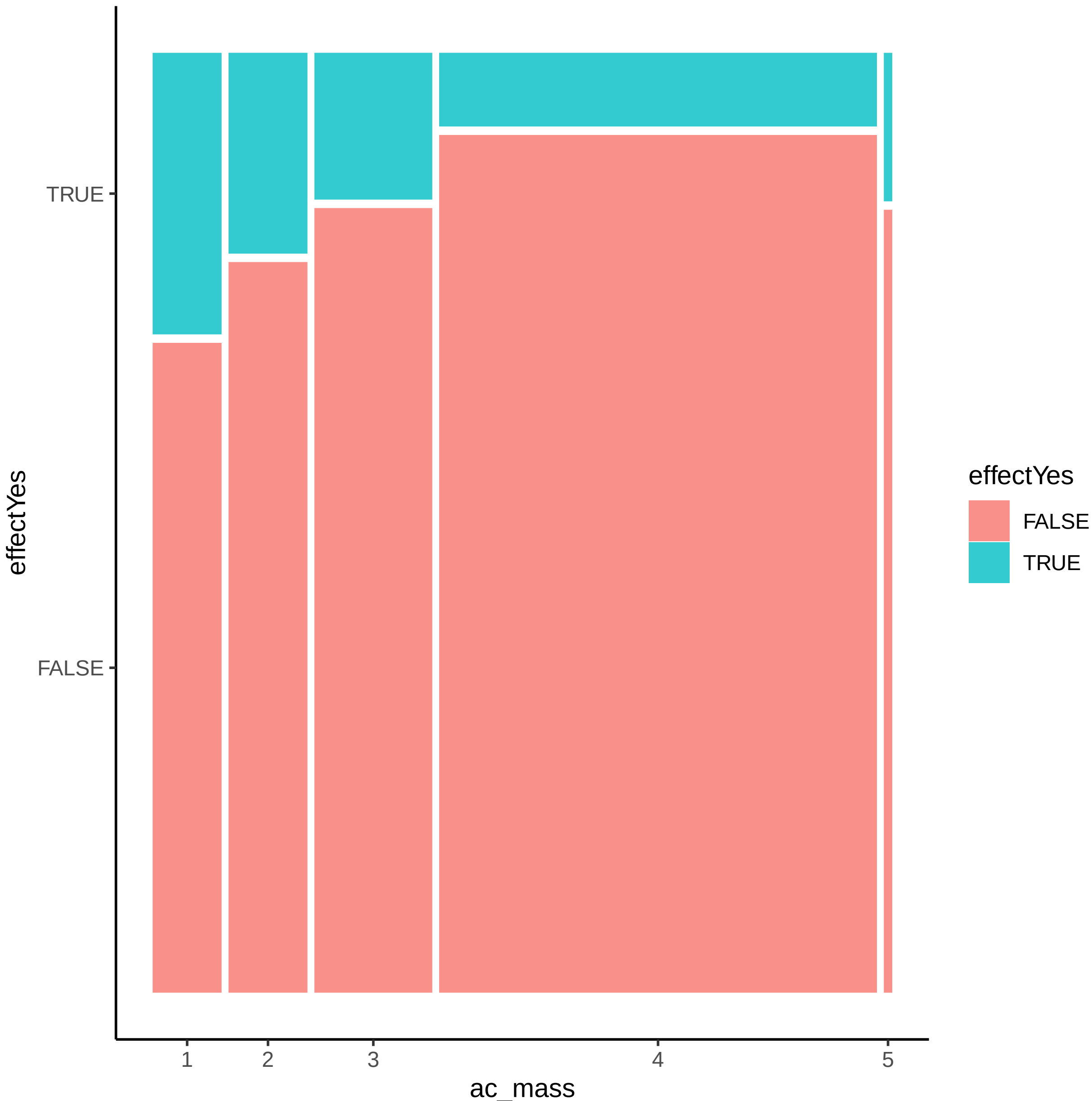


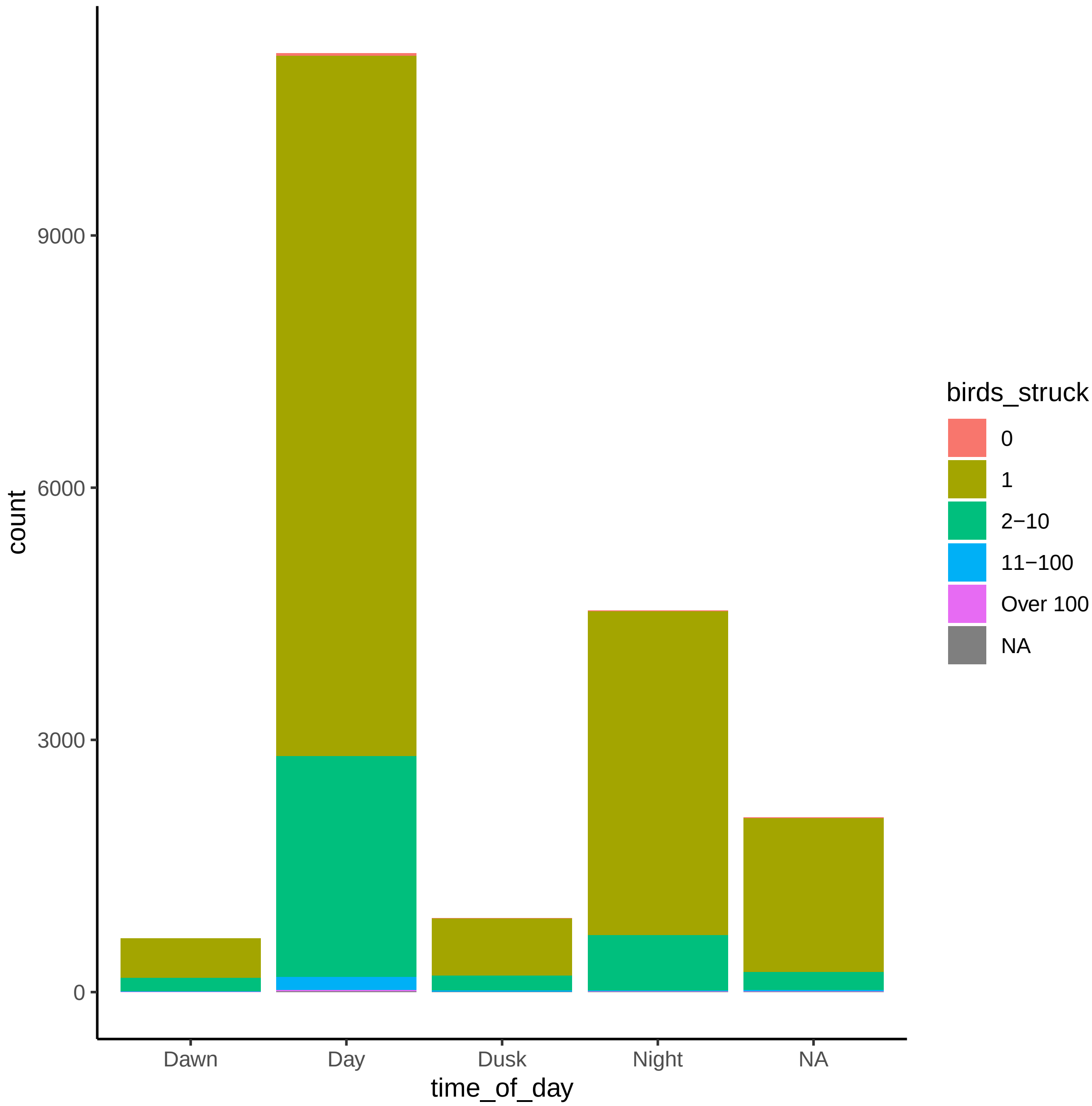
The higher the speed, the less likely to see a bird

Analysis of effect variable



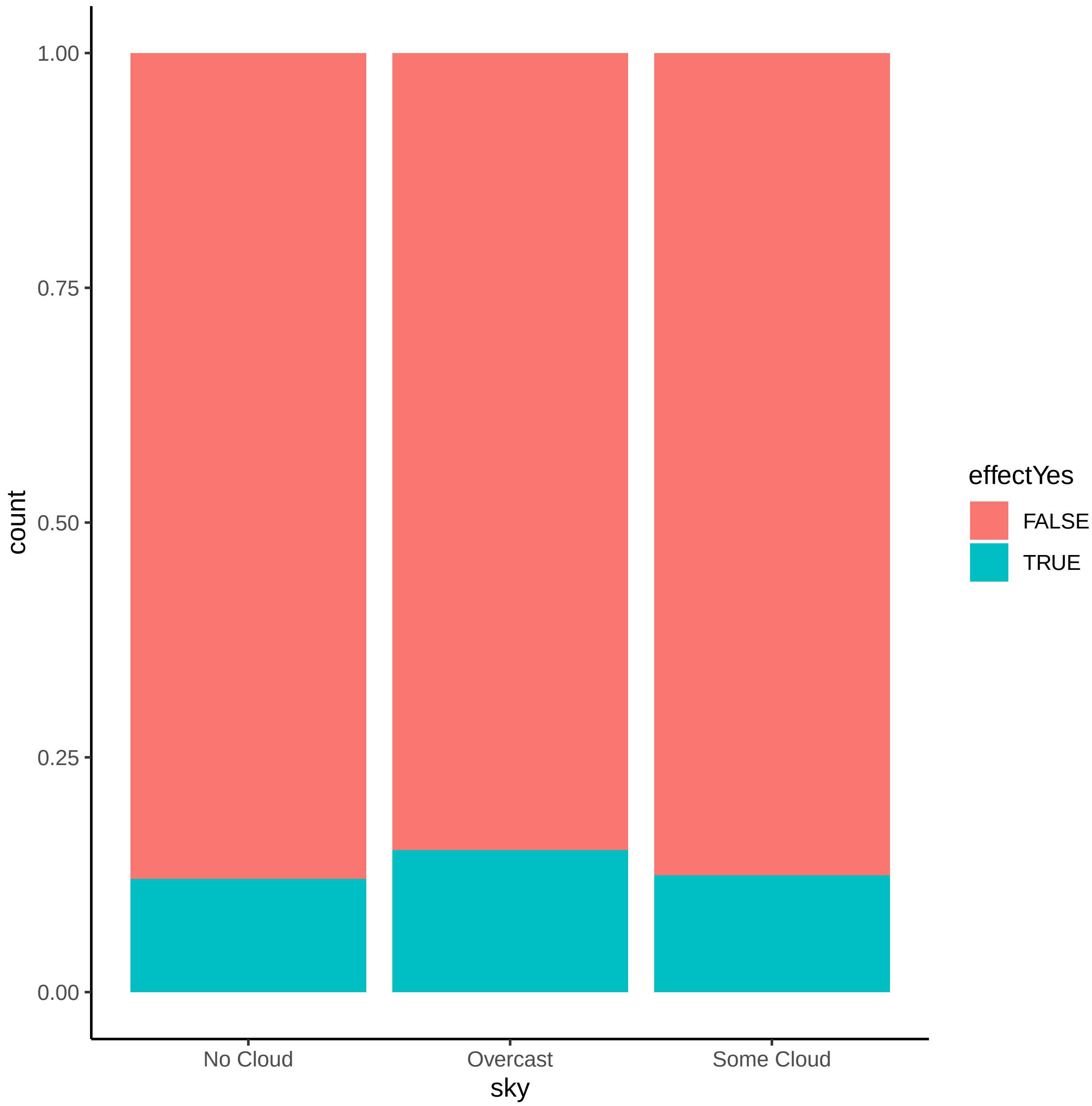



effectYes
TRUE
FALSE
1
2
3
4
5
ac_mass
effectYes
FALSE
TRUE



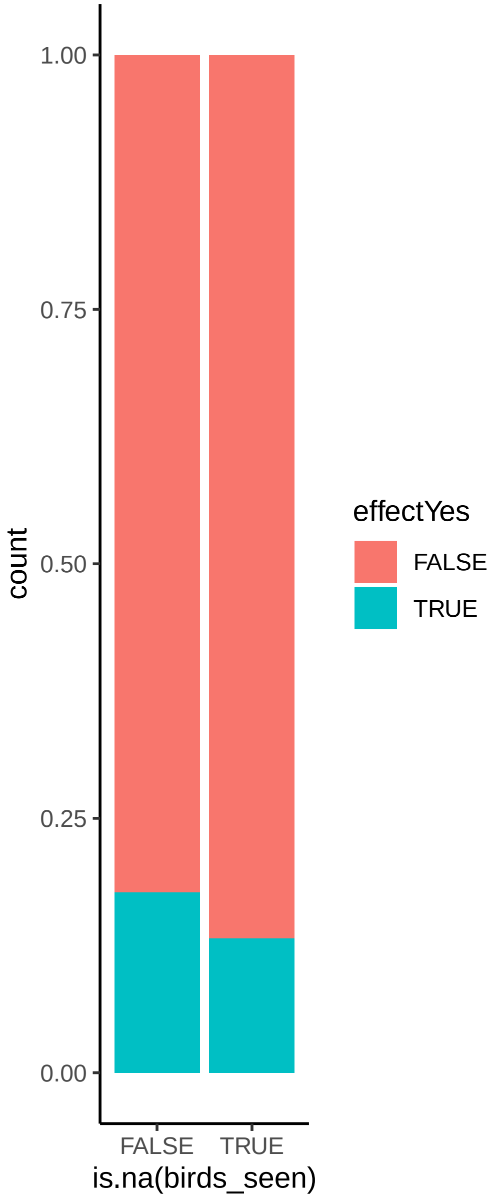
check if clouds have an influence on effect hyptohesis: clouds lead to less vision therefore the pilot might not see the birds

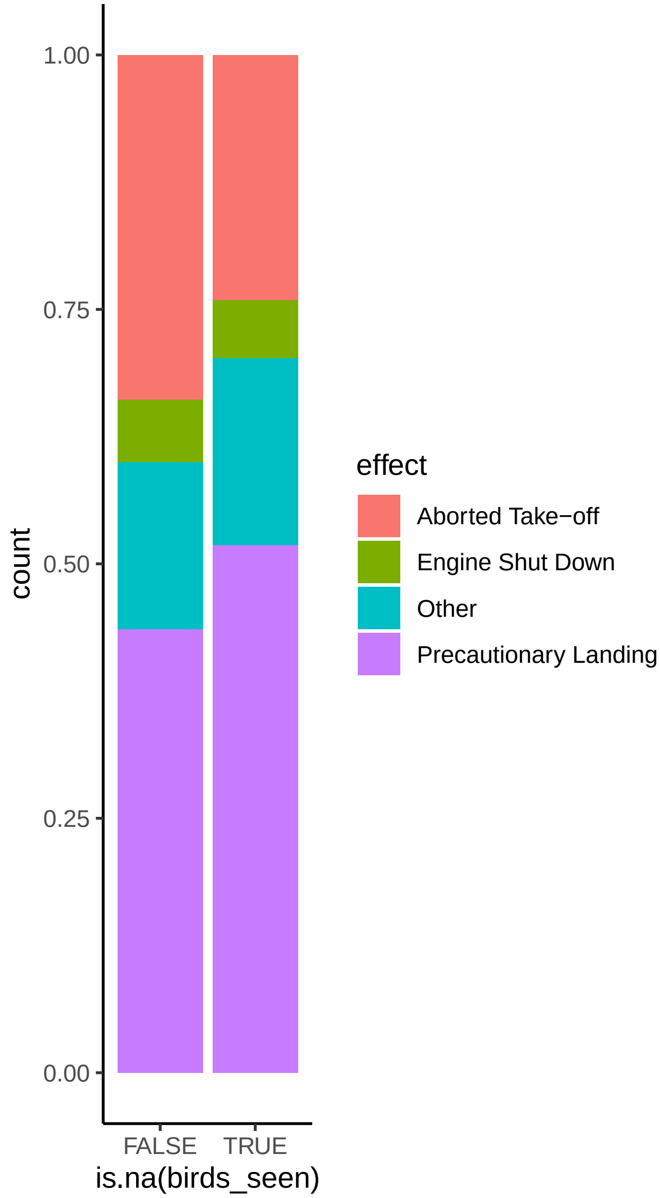
```
##
##                  FALSE      TRUE
##   No Cloud   0.8794500 0.1205500
##   Overcast   0.8489043 0.1510957
##   Some Cloud 0.8758788 0.1241212
```



results show, that clouds do not have an impact on effect variable, the proportion of effect=true is a little higher for some clouds and overcast

check influence of birds seen on effect variable, we assume NA values as if no birds have been seen before the strike occured





proportion of effect is a little higher when birds were seen

dp<-ggplot(iris) + aes(x=Petal.Width, color=Species, fill=Species) + geom_density(alpha=0.25) + theme(legend.position = "none") + scale_color_manual(values = color_pal) + scale_fill_manual(values = color_pal)

bp<-ggplot(iris) + aes(x=Petal.Width, y=Species, fill=Species, color=Species) + stat_boxplot(geom="errorbar",width=0.2) + geom_boxplot(varwidth = TRUE, alpha=0.2) + geom_jitter(alpha = 0.25, width = 0.2) + theme(legend.position = "none") + scale_color_manual(values = color_pal) + scale_fill_manual(values = color_pal)

h<-ggplot(iris) + aes(x=Petal.Width, color=Species, fill=Species) + geom_histogram(bins=30, alpha=0.25) + scale_color_manual(values = color_pal) + scale_fill_manual(values = color_pal)

dpl<-ggplot(iris) + aes(x=Petal.Length, color=Species, fill=Species) + geom_density(alpha=0.25) + theme(legend.position = "none") + scale_color_manual(values = color_pal) + scale_fill_manual(values = color_pal) bpl<-ggplot(iris) + aes(y=Petal.Length, x=Species, fill=Species, color=Species) + stat_boxplot(geom="errorbar",width=0.2) + geom_boxplot(varwidth = TRUE, alpha=0.2) + theme(legend.position = "none") + scale_color_manual(values = color_pal) + scale_fill_manual(values = color_pal)

sp<-ggplot(iris) + aes(x = Petal.Length, y = Petal.Width, shape = Species, color=Species, fill=Species) + geom_point() + facet_wrap(~Species) + scale_color_manual(values = color_pal) + scale_fill_manual(values = color_pal) + labs(title = "Petal Width and Length of 150 flowers of Iris", subtitle = "In relation to their species", x = "Petal Length in cm", y = "Petal Width in cm") + theme.title

dps<-ggplot(iris) + aes(x=Sepal.Length, color=Species, fill=Species) + geom_density(alpha=0.25) + theme(legend.position = "none") + scale_color_manual(values = color_pal) + scale_fill_manual(values = color_pal)

hs<-ggplot(iris) + aes(x=Sepal.Length, color=Species, fill=Species) + geom_histogram(bins=30, alpha=0.25) + scale_color_manual(values = color_pal) + scale_fill_manual(values = color_pal)

bps<-ggplot(iris) + aes(x=Sepal.Length, y=Species, fill=Species, color=Species) + stat_boxplot(geom="errorbar",width=0.2) + geom_boxplot(varwidth = TRUE, alpha=0.2) + geom_jitter(alpha = 0.25, width = 0.2) + theme(legend.position = "none") + scale_color_manual(values = color_pal) + scale_fill_manual(values = color_pal)

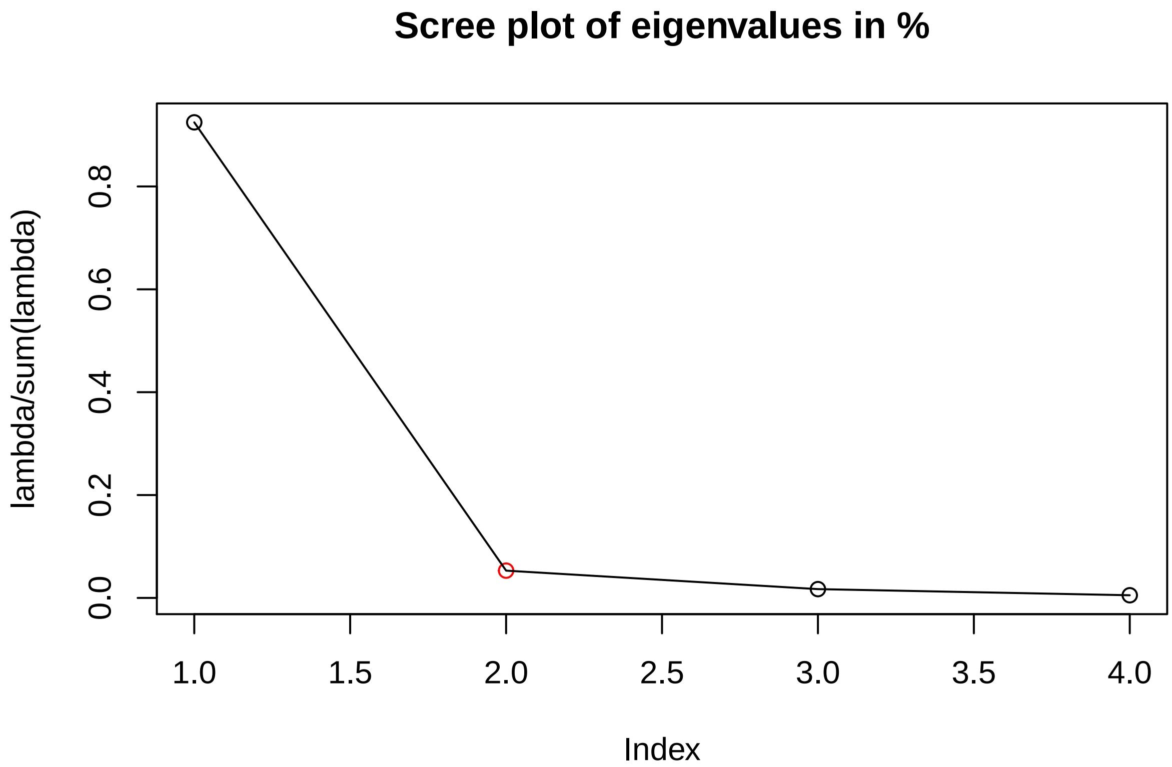
sps<-ggplot(iris) + aes(x = Sepal.Width, y = Sepal.Length, shape = Species, color=Species, fill=Species) + geom_point() + facet_wrap(~Species) + scale_color_manual(values = color_pal) + scale_fill_manual(values = color_pal) + labs(title = "Sepal Width and Length of 150 flowers of Iris", subtitle = "In relation to their species", x = "Sepal Width in cm", y = "Sepal Length in cm") + theme.title

(dp + h) / bp / (dpl + bpl) / sp sps/ (dps + hs) / bps

PCA for Iris data

To do a Principal Component Analysis we will use *prcomp()* and plot the eigenvalues in a scree plot. We see that there's an elbow at 2, so we choose 2 dimensions. *iris.pca$x* has the same dimensions as *X*.

```r
X<-iris[,1:4]               ## exclude the Species column
S <- cov(X)                 ## compute variance-covariance matrix
iris.pca <- prcomp(X)       ## perform PCA
lambda <- eigen(S)$values ## calculate eigenvalues
```


Scree plot of eigenvalues in %
lambda/sum(lambda)
0.0
0.2
0.4
0.6
0.8
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Index