Thoughts on Storks - Model selection and averaging

Lizzie Jones

26/07/2021

Quantitative analysis/modelling

Quantitative data analysis: Descriptive and statistical - to understand variation in respondent's awareness, knowledge and attitudes towards white storks and their reintroduction.

Methods plan

- GLM approach + model selection and averaging
    - Anderson, D. and Burnham, K., 2004. Model selection and multi-model inference. Second. NY: Springer-Verlag, 63(2020), p.10.
    - Burnham, K.P., Anderson, D.R. and Huyvaert, K.P., 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behavioral ecology and sociobiology, 65(1), pp.23-35.
- Compare OverallAttitudeScore to "Q15. Do you support WS reintro (yes/no)"

Possible predictor variables

Factor variables

- Age (collapse further?)
- Gender (female / male)
- Urban / suburban / rural
- Highest education (collapse – e.g. degree; below degree)
- Occupation (use? If so, would need to collapse! unemployed; retired; potentially pool responses except for those who answered "environment, nature & wildlife")
- Visited Knepp (yes / no)
- Time spent in nature
- Member of conservation/environmental organisation (quite a few people listed RSPB)
- Awareness
- Heard of white stork before taking this survey?
- Heard of white stork project / reintroduction effort?

Numeric variables

- Contact and connection with nature; general environmental attitudes and behaviour
- Nature Connection Index (composite score)
- Environmental concern (composite score)
- General attitude towards birds (composite score)

```
# First, rename that REALLY long column (Q22)
names(final_data)[names(final_data) == "Q22....Are.you.a.member.of.any.environmental..wildlife.
or.conservation.organisations."] <- "Q22_env_org_member"

final_data <- final_data %>% mutate(Q8_Seen =
                     case_when(Q8_wild_seen == 1L ~ "Wild",
                               Q8_wild_seen == 1L & Q8_captivity_seen == 1L ~ "Wild",
                               Q8_wild_seen == 0L & Q8_captivity_seen == 1L ~ "Captivity",
                               Q8_wild_seen == 0L & Q8_captivity_seen == 1L & Q8_pictures_video
== 1L ~ "Captivity",
                               Q8_wild_seen == 1L & Q8_captivity_seen == 1L & Q8_No == 1L ~ "N
o/Not sure",
                               Q8_wild_seen == 0L & Q8_captivity_seen == 1L & Q8_No == 1L ~ "N
o/Not sure",
                               Q8_wild_seen == 1L & Q8_captivity_seen == 1L & Q8_NotSure == 1L
 ~ "No/Not sure",
                               Q8_wild_seen == 0L & Q8_captivity_seen == 1L & Q8_NotSure == 1L
 ~ "No/Not sure",
                               Q8_wild_seen == 0L & Q8_captivity_seen == 1L & Q8_NotSure == 1L
 & Q8_NotSure == 1L ~ "No/Not sure",
                               Q8_wild_seen == 0L & Q8_captivity_seen == 0L & Q8_pictures_video
== 1L ~ "No/Not sure",
                               Q8_wild_seen == 0L & Q8_captivity_seen == 0L & Q8_pictures_video
== 0L ~ "No/Not sure"))

# Select all possible predictor vars
model_data <- final_data %>%
  dplyr::select(UniqueID_all, OverallAttitudeScore, SiteLocal, SurveyType,
         Age_short, Gender, Area_type, Education_short, Occupation_short_clean,
         Q14.5_agreement_score, Q8_Seen,
         Q27_Knepp_visit, Q18_exp_nature,
         Q1_aware_stork, Q9_heard, KnowledgeScore, Q22_env_org_member,
         NCI, ProCoBS, BirdInterestScore, EnvConcern.score)
nrow(model_data)
```

```
## [1] 3531
```

```
model_clean <- model_data[!is.na(model_data$OverallAttitudeScore), ]
nrow(model_clean) ## Dropped ~1100 rows due to NA in AttitudeScore
```

```
## [1] 2492
```

```
# Select numeric variables
model_clean1 <- model_clean %>%
  drop_na()
nrow(model_clean1)
```

```
## [1] 2445
```

```
# Check that empties have been dropped
summary(model_clean1$Education_short)
```

```
##        No formal quals.                    Other Secondary school/College
##                      46                       50                      861
##     University graduate
##                    1488
```

```
### Clean factor predictors for modelling
# Gender
model_clean2 <- model_clean1[model_clean1$Gender!= "N/A", ]
# Age
model_clean2 <- model_clean2[model_clean2$Age_short!= "N/A", ]
# Occupation
model_clean2 <- model_clean2[model_clean2$Occupation_short!= "Prefer not to say", ]

# Education
model_clean2 <- model_clean2[model_clean2$Education_short!= "Other", ]

# Drop empty factor levels
model_clean2 <- droplevels(model_clean2)
# Check n
nrow(model_clean2)
```

```
## [1] 2330
```

```
# Check that empties have been dropped
summary(model_clean2$Occupation_short)
```

```
##         Environment/Nature Natural resource management
##                        275                          66
##                      Other                     Retired
##                       1422                         465
##                 Unemployed
##                        102
```

```
### Mnaually categorised Occupation here to form new column = Occupation_short_clean

#### Also need to rename/ shorten some variable names
model_clean2 <- model_clean2 %>%
  dplyr::rename(Age = Age_short,
                Locality = SiteLocal,
                Area.type = Area_type,
                Survey.type = SurveyType,
                Occupation = Occupation_short_clean,
                Education = Education_short,
                Aware.of.storks = Q1_aware_stork,
                Heard.of.WSP = Q9_heard,
                Support.reintroductions = Q14.5_agreement_score,
                Seen.in.Wild.Captivity = Q8_Seen,
                Frequency.exp.nature = Q18_exp_nature,
                Visited.Knepp = Q27_Knepp_visit,
                Member.of.Environmental.Organisation = Q22_env_org_member,
                Knowledge.Score = KnowledgeScore,
                Bird.Interest.Score = BirdInterestScore,
                Environmental.Concern.Score = EnvConcern.score)
colnames(model_clean2)
```

```
##  [1] "UniqueID_all"
##  [2] "OverallAttitudeScore"
##  [3] "Locality"
##  [4] "Survey.type"
##  [5] "Age"
##  [6] "Gender"
##  [7] "Area.type"
##  [8] "Education"
##  [9] "Occupation"
## [10] "Support.reintroductions"
## [11] "Seen.in.Wild.Captivity"
## [12] "Visited.Knepp"
## [13] "Frequency.exp.nature"
## [14] "Aware.of.storks"
## [15] "Heard.of.WSP"
## [16] "Knowledge.Score"
## [17] "Member.of.Environmental.Organisation"
## [18] "NCI"
## [19] "ProCoBS"
## [20] "Bird.Interest.Score"
## [21] "Environmental.Concern.Score"
```

```
### Relevel Occupation factor
model_clean2$Occupation <- relevel(model_clean2$Occupation,"Environment/Nature")
### Relevel Education factor
model_clean2$Education <- relevel(model_clean2$Education,"No formal quals.")
### Relevel Seen(Wild/Captivity) factor
model_clean2$Seen.in.Wild.Captivity <- factor(model_clean2$Seen.in.Wild.Captivity, levels = c(
"Wild", "Captivity", "No/Not sure"))
### Relevel Freq_exp_nature factor
model_clean2$Frequency.exp.nature <- relevel(model_clean2$Frequency.exp.nature, "None")
```

Exploring the response variable

Response variable = Attitudes to WS reintroduction (Composite score)

Now that the data has been cleaned we can explore the distribution of the response variable. We can see from the histogram (density plot) that the data is left-skewed as the reponses are geenrally towards the upper end of the response scale (0-5). A Shpiro test indicates that the distribution of non-normal, and QQ plots show that Squaring the response variable does the best job of normalising the distribution, but it's still non-normal.

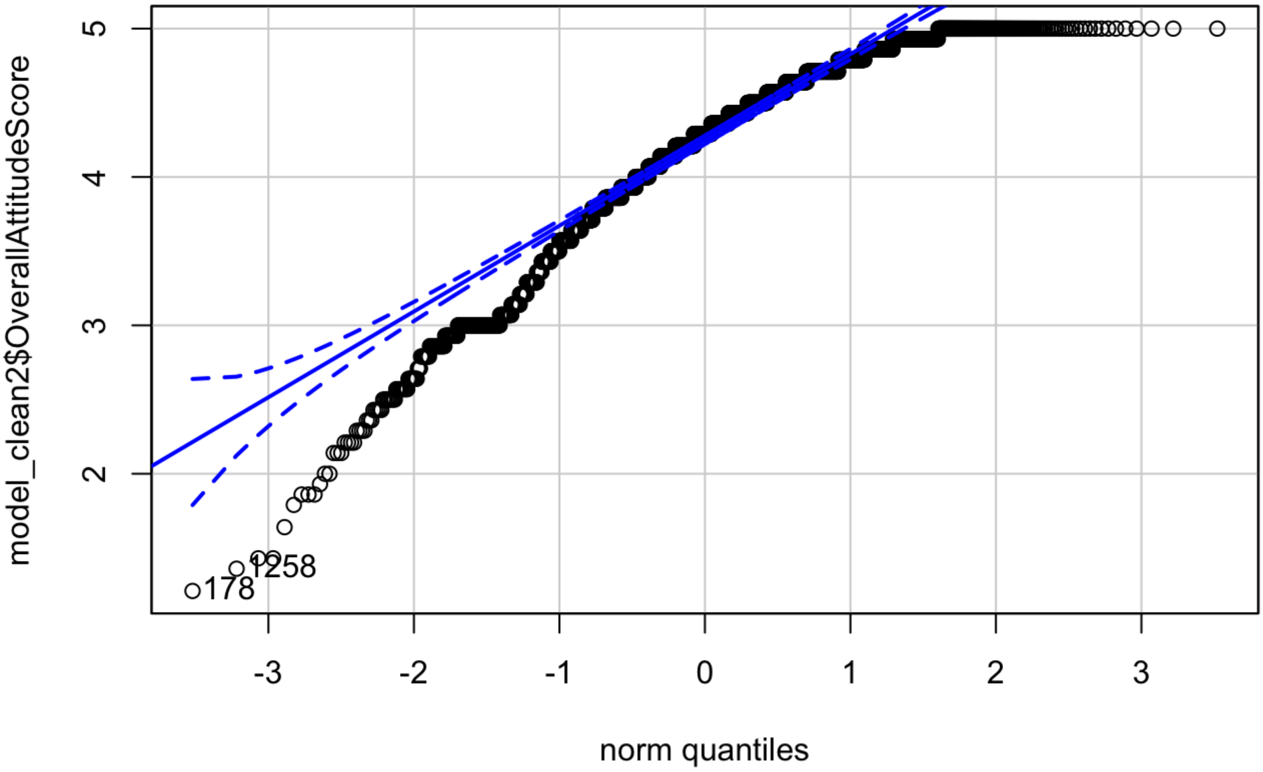
```
str(model_clean2$OverallAttitudeScore) # check it's a numeric column
```

```
##  num [1:2330] 4.93 4.64 4.86 4.57 4.86 3.93 4.71 4.36 3.79 4.71 ...
```

```
model_clean2 %>%
     group_by(Survey.type) %>%
     summarise(sum(!is.na(OverallAttitudeScore))) ## Counting sample size (Non-NA) values per s
urvey type
```

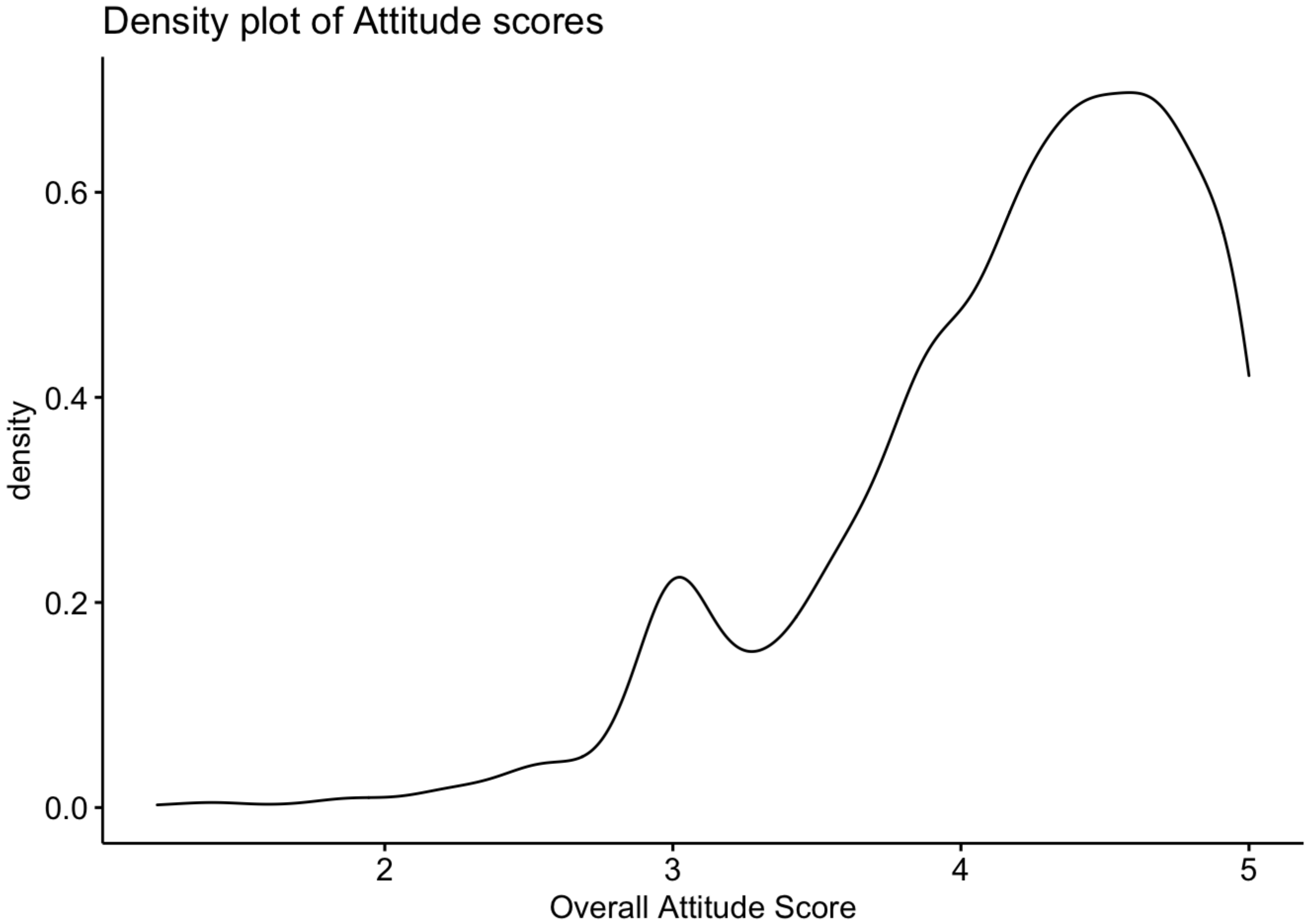
```
## # A tibble: 2 x 2
##   Survey.type `sum(!is.na(OverallAttitudeScore))`
##   <fct>                                     <int>
## 1 NatRep                                      706
## 2 Proactive                                  1624
```

```
## UNTRANSFORMED DATA
# QQ plot shows non-normality with skew towards the right
qqPlot(model_clean2$OverallAttitudeScore)
```




```
## [1]  178 1258
```

```
# Also seen ina  density plot
ggdensity(model_clean2$OverallAttitudeScore,
          main = "Density plot of Attitude scores",
          xlab = "Overall Attitude Score")
```




```
mean(model_clean2$OverallAttitudeScore)
```

```
## [1] 4.168373
```

```
median(model_clean2$OverallAttitudeScore)
```

```
## [1] 4.29
```

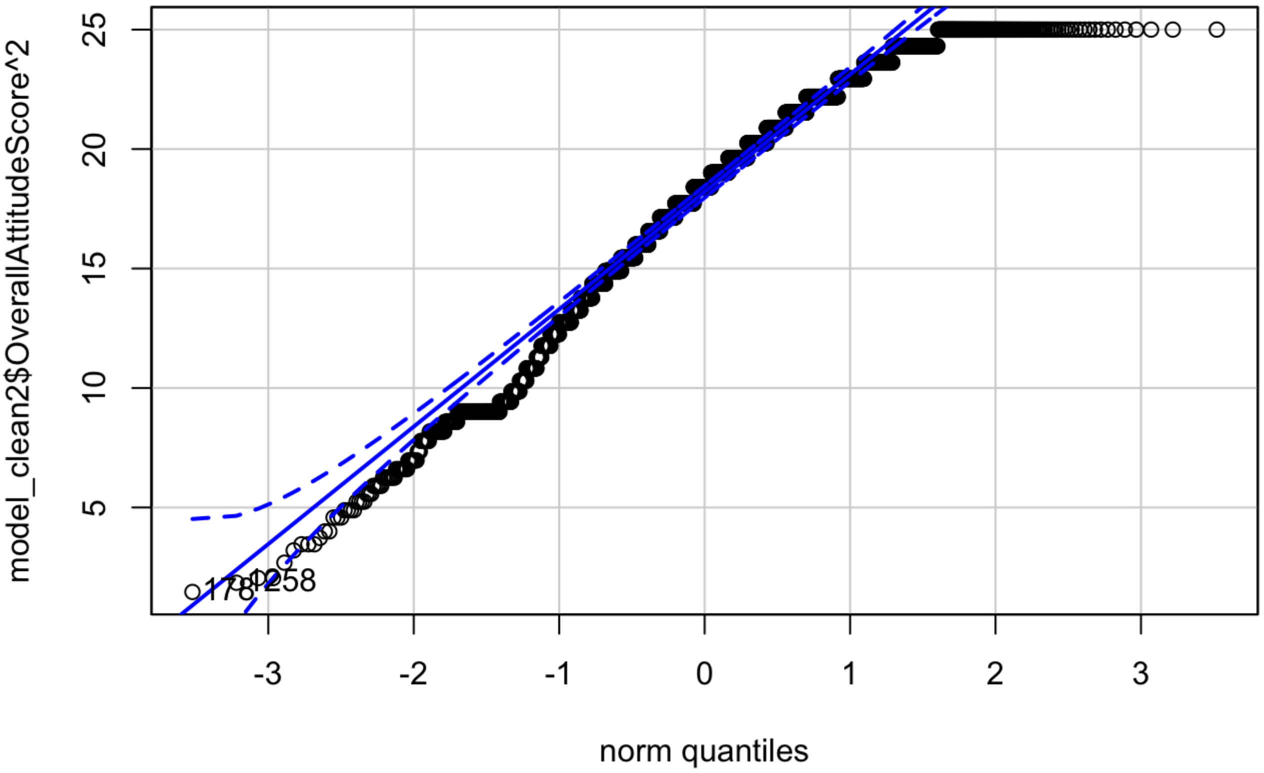
```
# Select variables
model_clean2 %>%
  dplyr::select(Survey.type, Locality, OverallAttitudeScore) %>%
  group_by()
```

```
## # A tibble: 2,330 x 3
##    Survey.type Locality  OverallAttitudeScore
##    <fct>       <fct>                    <dbl>
##  1 Proactive   Not local                 4.93
##  2 Proactive   Not local                 4.64
##  3 Proactive   Local                     4.86
##  4 Proactive   Not local                 4.57
##  5 Proactive   Not local                 4.86
##  6 Proactive   Not local                 3.93
##  7 Proactive   Not local                 4.71
##  8 Proactive   Not local                 4.36
##  9 Proactive   Not local                 3.79
## 10 Proactive   Not local                 4.71
## # … with 2,320 more rows
```

```
# Running a Shapiro test to make sure
shapiro.test(model_clean2$OverallAttitudeScore)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  model_clean2$OverallAttitudeScore
## W = 0.92637, p-value < 2.2e-16
```

```
### TRANSFORMATIONS
# Square
qqPlot(model_clean2$OverallAttitudeScore^2)
```

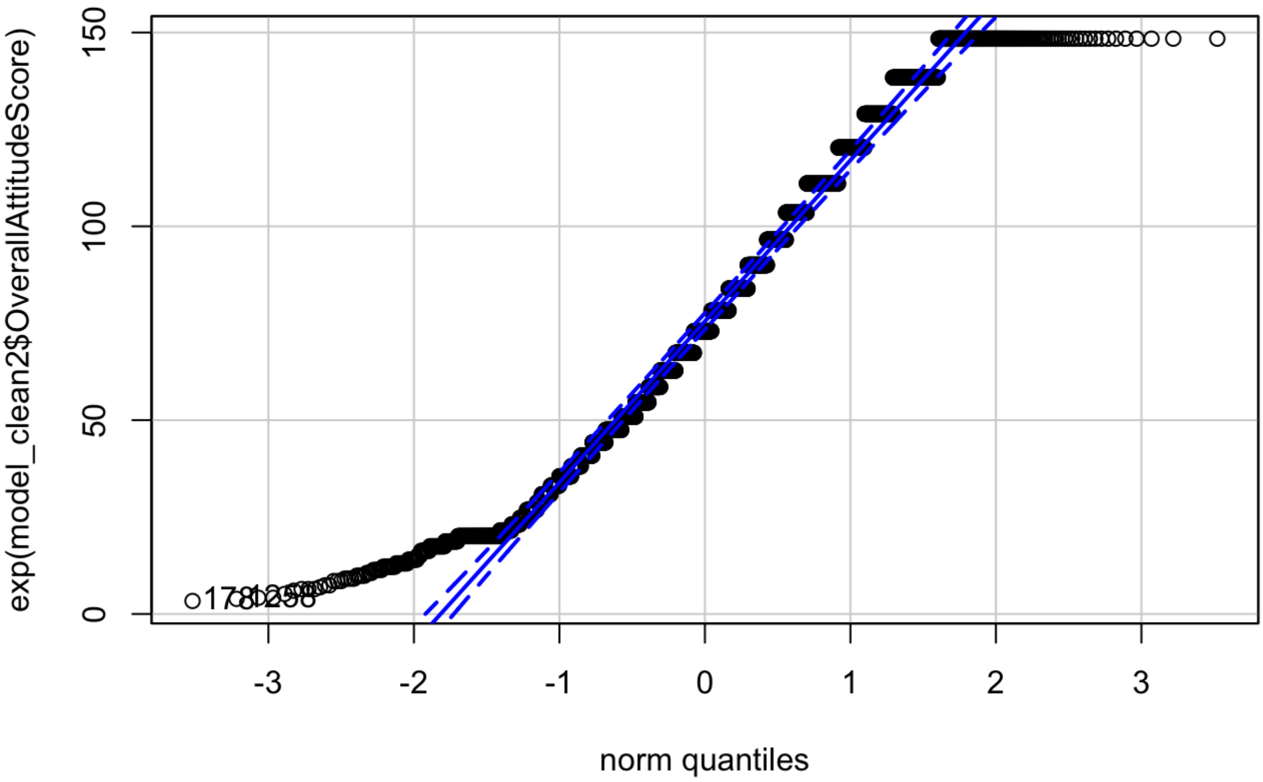



```
## [1]  178 1258
```

```
shapiro.test((model_clean2$OverallAttitudeScore)^2)
```

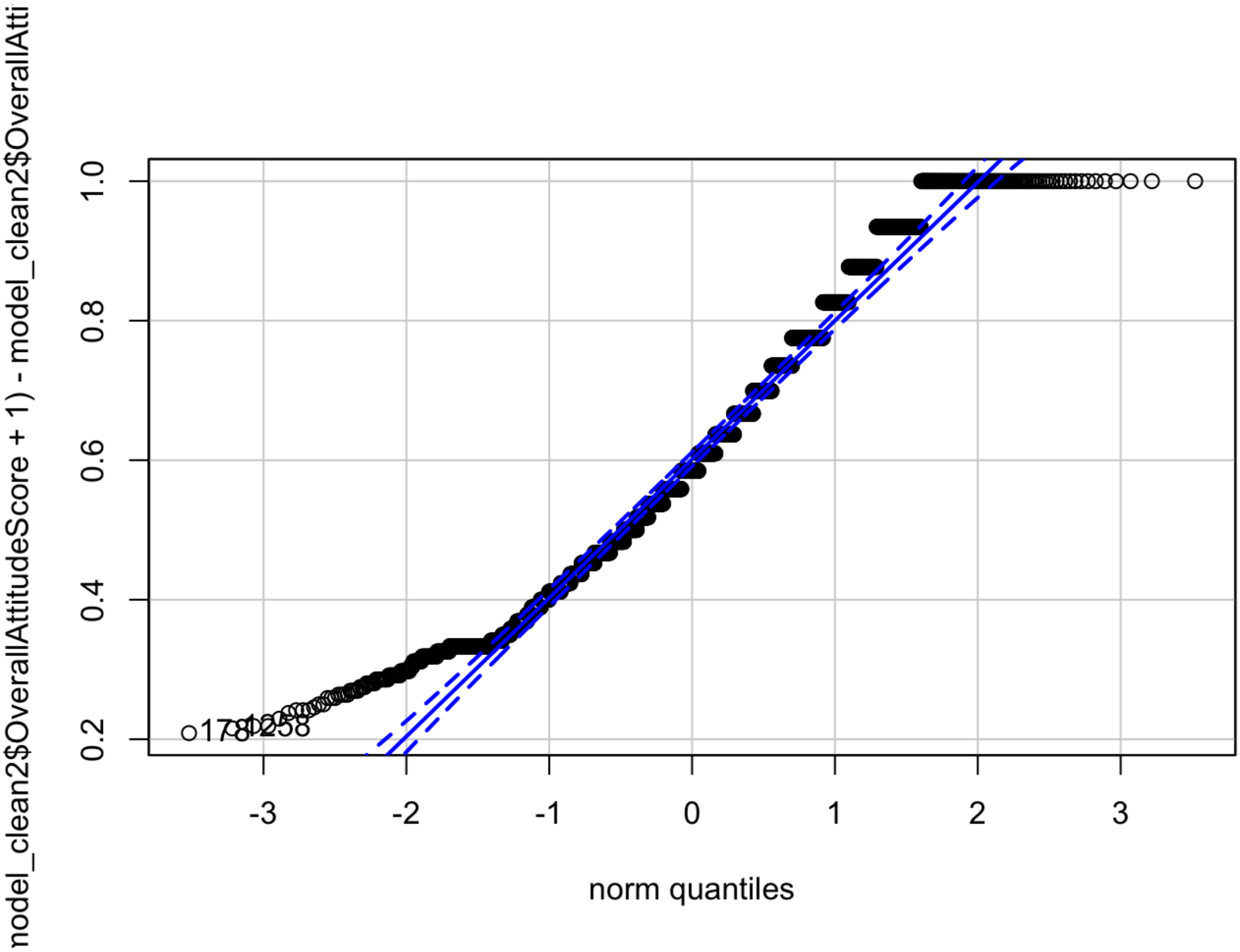
```
##
##  Shapiro-Wilk normality test
##
## data:  (model_clean2$OverallAttitudeScore)^2
## W = 0.95764, p-value < 2.2e-16
```

```
# Exponential
qqPlot(exp(model_clean2$OverallAttitudeScore))
```




```
## [1]  178 1258
```

```
# Inverse transformation for severse skew
qqPlot(1/(max(model_clean2$OverallAttitudeScore+1) - model_clean2$OverallAttitudeScore))
```



```
## [1]  178 1258
```

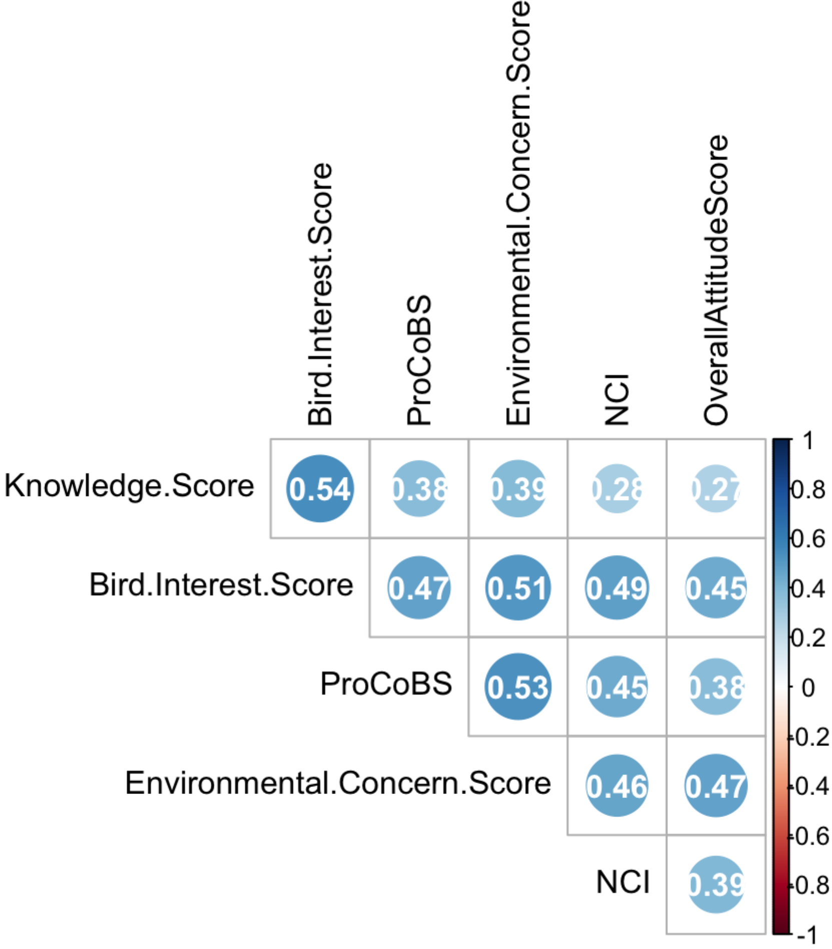
Predictor correlation matrix of numeric variables

Moderate but insignificant correlation seen between all of the numeric score-based variables/predictors. BirdInterestScore tends to show strongest correlation with other predictors so this might be the most effective to remove from the model if the VIF score is also high.

```
# Select numeric variables
model_numeric <- model_clean2 %>%
  dplyr::select_if(., is.numeric) %>%
  dplyr::select(., -UniqueID_all, -Support.reintroductions) %>%
  drop_na()
head(model_numeric)
```

```
##   OverallAttitudeScore Knowledge.Score NCI ProCoBS Bird.Interest.Score
## 1                 4.93             6.1  43      19                  17
## 2                 4.64             5.6 100      23                  19
## 3                 4.86             3.0  59      17                  17
## 4                 4.57             2.7  59      16                  17
## 5                 4.86             2.0  59      18                  19
## 6                 3.93             5.7  45      22                  15
##   Environmental.Concern.Score
## 1                          10
## 2                          10
## 3                          10
## 4                          10
## 5                          10
## 6                          10
```

```
# Create corrlation matrix
model.cor = cor(model_numeric, method = c("spearman"))
res1 <- cor.mtest(model_numeric, conf.level = .95)
# Create corrplot
corrplot::corrplot(model.cor, p.mat = res1$p, method = "circle", type = "upper", insig='blank',
tl.col = "black",
         addCoef.col ='white', order = "AOE", diag=FALSE) ### All variables moderately correlat
ed but not significant
```



Global model generation

We now generate the global model. This is a saturated model with all of the fixed effects and their interesting interactions. There are no random effects in this model so we use a linear model.

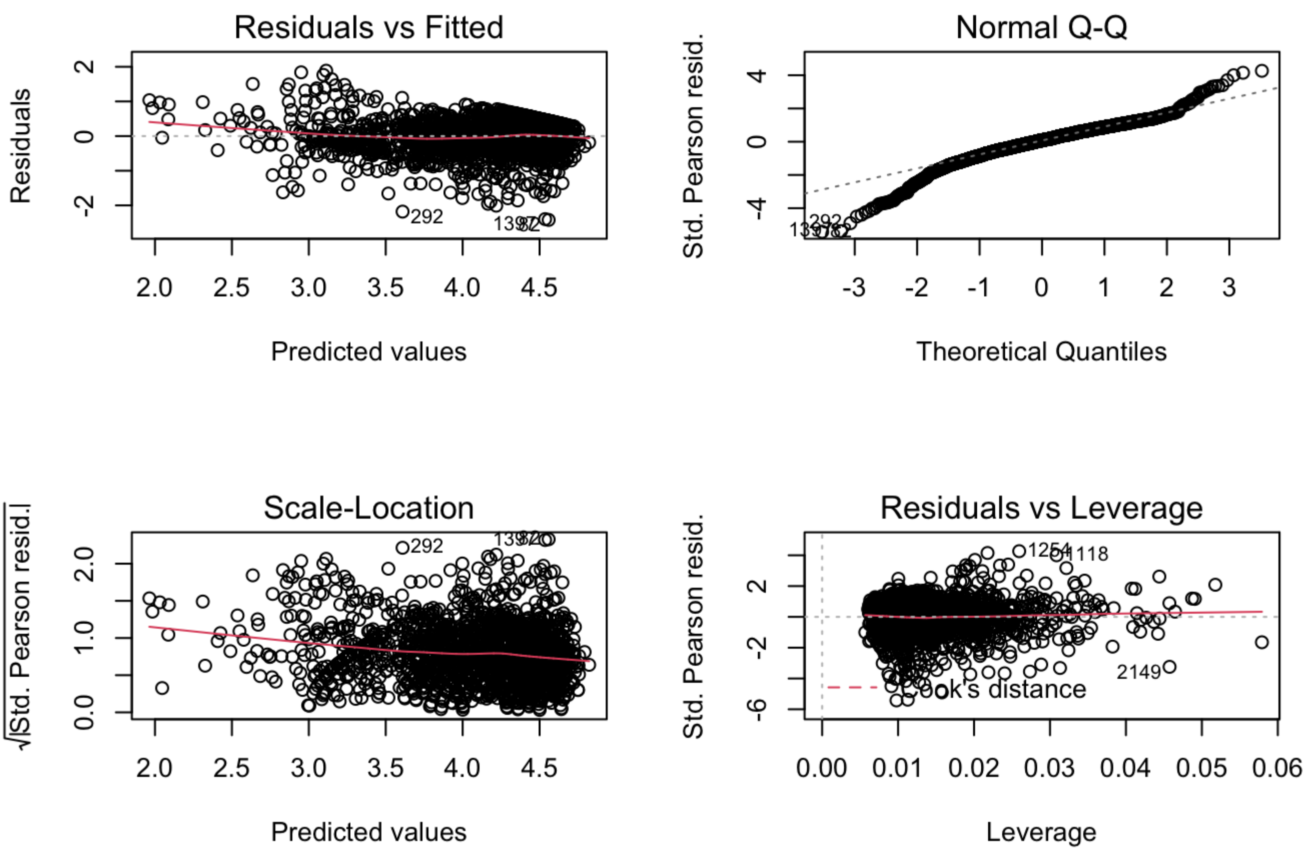
```
# Create saturated model (potentially SurveyType or locality as a random effect?)
global_model <- glm(OverallAttitudeScore ~ Locality + Survey.type +
         Age + Gender + Area.type + Education + Occupation + Aware.of.storks +
         Support.reintroductions + Seen.in.Wild.Captivity +
         Visited.Knepp + # Have you visited Knepp?
         Frequency.exp.nature + # Likert - experience of nature
         Heard.of.WSP + # Had you ever heard of the WSP and reintroduction of WS to southern En
gland?
         Member.of.Environmental.Organisation +
         Knowledge.Score + NCI + ProCoBS + Bird.Interest.Score + Environmental.Concern.Score,
         data = model_clean2)

summary(global_model)
```

```
##
## Call:
## glm(formula = OverallAttitudeScore ~ Locality + Survey.type +
##     Age + Gender + Area.type + Education + Occupation + Aware.of.storks +
##     Support.reintroductions + Seen.in.Wild.Captivity + Visited.Knepp +
##     Frequency.exp.nature + Heard.of.WSP + Member.of.Environmental.Organisation +
##     Knowledge.Score + NCI + ProCoBS + Bird.Interest.Score + Environmental.Concern.Score,
##     data = model_clean2)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.41894  -0.21892   0.03618   0.28483   1.88721
##
## Coefficients:
##                                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                             1.1606575  0.1203543   9.644  < 2e-16
## LocalityNot local                       0.0395021  0.0255163   1.548  0.12173
## Survey.typeProactive                    0.0795818  0.0323729   2.458  0.01403
## Age45-64                                0.0028644  0.0225077   0.127  0.89874
## Age65+                                  0.0685899  0.0344370   1.992  0.04652
## GenderMale                             -0.0796871  0.0198049  -4.024 5.92e-05
## Area.typeSub-urban                     -0.0374864  0.0228421  -1.641  0.10091
## Area.typeUrban                         -0.0592257  0.0273908  -2.162  0.03070
## EducationSecondary school/College      -0.0563596  0.0706863  -0.797  0.42535
## EducationUniversity graduate           -0.0663831  0.0708651  -0.937  0.34898
## OccupationNatural resource management   0.1183916  0.0628814   1.883  0.05986
## OccupationOther                         0.1760229  0.0326043   5.399 7.40e-08
## OccupationRetired                       0.1199579  0.0423241   2.834  0.00463
## OccupationUnemployed                    0.1100589  0.0556927   1.976  0.04825
## Aware.of.storksYes                      0.0548575  0.0316244   1.735  0.08294
## Support.reintroductions                 0.3641312  0.0123364  29.517  < 2e-16
## Seen.in.Wild.CaptivityCaptivity         0.0228649  0.0380255   0.601  0.54770
## Seen.in.Wild.CaptivityNo/Not sure       0.0482295  0.0245850   1.962  0.04991
## Visited.KneppYes                        0.0135869  0.0261324   0.520  0.60316
## Frequency.exp.nature1-2 days           -0.0524714  0.0481147  -1.091  0.27559
## Frequency.exp.nature3-4 days           -0.0729885  0.0499208  -1.462  0.14385
## Frequency.exp.nature5-6 days           -0.0641347  0.0516824  -1.241  0.21475
## Frequency.exp.natureEvery day, 7 days  -0.0696971  0.0518038  -1.345  0.17863
## Heard.of.WSPNot sure                   -0.0665145  0.0515982  -1.289  0.19750
## Heard.of.WSPYes                         0.0367588  0.0276944   1.327  0.18454
## Member.of.Environmental.OrganisationYes -0.0527248  0.0243428  -2.166  0.03042
## Knowledge.Score                        -0.0064880  0.0079117  -0.820  0.41227
## NCI                                     0.0019100  0.0004503   4.242 2.31e-05
## ProCoBS                                 0.0043602  0.0022504   1.938  0.05280
## Bird.Interest.Score                     0.0280821  0.0046533   6.035 1.85e-09
## Environmental.Concern.Score             0.0675764  0.0100331   6.735 2.06e-11
##
## (Intercept)                            ***
## LocalityNot local
## Survey.typeProactive                   *
## Age45-64
## Age65+                                 *
## GenderMale                             ***
## Area.typeSub-urban
## Area.typeUrban                         *
## EducationSecondary school/College
## EducationUniversity graduate
## OccupationNatural resource management  .
## OccupationOther                        ***
## OccupationRetired                      **
## OccupationUnemployed                   *
```

```
## Aware.of.storksYes                         .
## Support.reintroductions                    ***
## Seen.in.Wild.CaptivityCaptivity
## Seen.in.Wild.CaptivityNo/Not sure          *
## Visited.KneppYes
## Frequency.exp.nature1-2 days
## Frequency.exp.nature3-4 days
## Frequency.exp.nature5-6 days
## Frequency.exp.natureEvery day, 7 days
## Heard.of.WSPNot sure
## Heard.of.WSPYes
## Member.of.Environmental.OrganisationYes *
## Knowledge.Score
## NCI                                        ***
## ProCoBS                                    .
## Bird.Interest.Score                        ***
## Environmental.Concern.Score                ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2008046)
##
##     Null deviance: 945.65  on 2329  degrees of freedom
## Residual deviance: 461.65  on 2299  degrees of freedom
## AIC: 2904.4
##
## Number of Fisher Scoring iterations: 2
```

```
par(mfrow = c(2, 2))
plot(global_model)
```



```r
# http://www.sthda.com/english/articles/39-regression-model-diagnostics/161-linear-regression-a
ssumptions-and-diagnostics-in-r-essentials/

# Check for variance inflation factors (VIF > 2 is worth removing and rechecking)
vif(global_model)
```

```
##                                         GVIF Df GVIF^(1/(2*Df))
## Locality                            1.542174  1        1.241843
## Survey.type                         2.568176  1        1.602553
## Age                                 1.948750  2        1.181514
## Gender                              1.128561  1        1.062338
## Area.type                           1.309027  2        1.069639
## Education                           1.236705  2        1.054548
## Occupation                          2.338344  4        1.112022
## Aware.of.storks                     1.717921  1        1.310695
## Support.reintroductions             1.340245  1        1.157689
## Seen.in.Wild.Captivity              1.669406  2        1.136686
## Visited.Knepp                       1.547447  1        1.243964
## Frequency.exp.nature                1.542393  4        1.055661
## Heard.of.WSP                        2.164533  2        1.212945
## Member.of.Environmental.Organisation 1.716715  1        1.310235
## Knowledge.Score                     2.179805  1        1.476416
## NCI                                 1.513783  1        1.230359
## ProCoBS                             1.709734  1        1.307568
## Bird.Interest.Score                 2.255971  1        1.501989
## Environmental.Concern.Score         1.892637  1        1.375731
```

```r
with(summary(global_model), 1 - deviance/null.deviance)
```

```
## [1] 0.511817
```

```r
### Create new model remoing Bird Interest Score and Heard of WSP
global_model1 <- glm(OverallAttitudeScore ~ Locality + Survey.type +
         Age + Gender + Area.type + Education + Occupation + Aware.of.storks +
         Support.reintroductions +
         Visited.Knepp + # Have you visited Knepp?
         Frequency.exp.nature + # Likert - experience of nature
         Member.of.Environmental.Organisation +
         Knowledge.Score + NCI + ProCoBS + Bird.Interest.Score + Environmental.Concern.Score, d
ata = model_clean2)
# Models summaries
summary(global_model1)
```

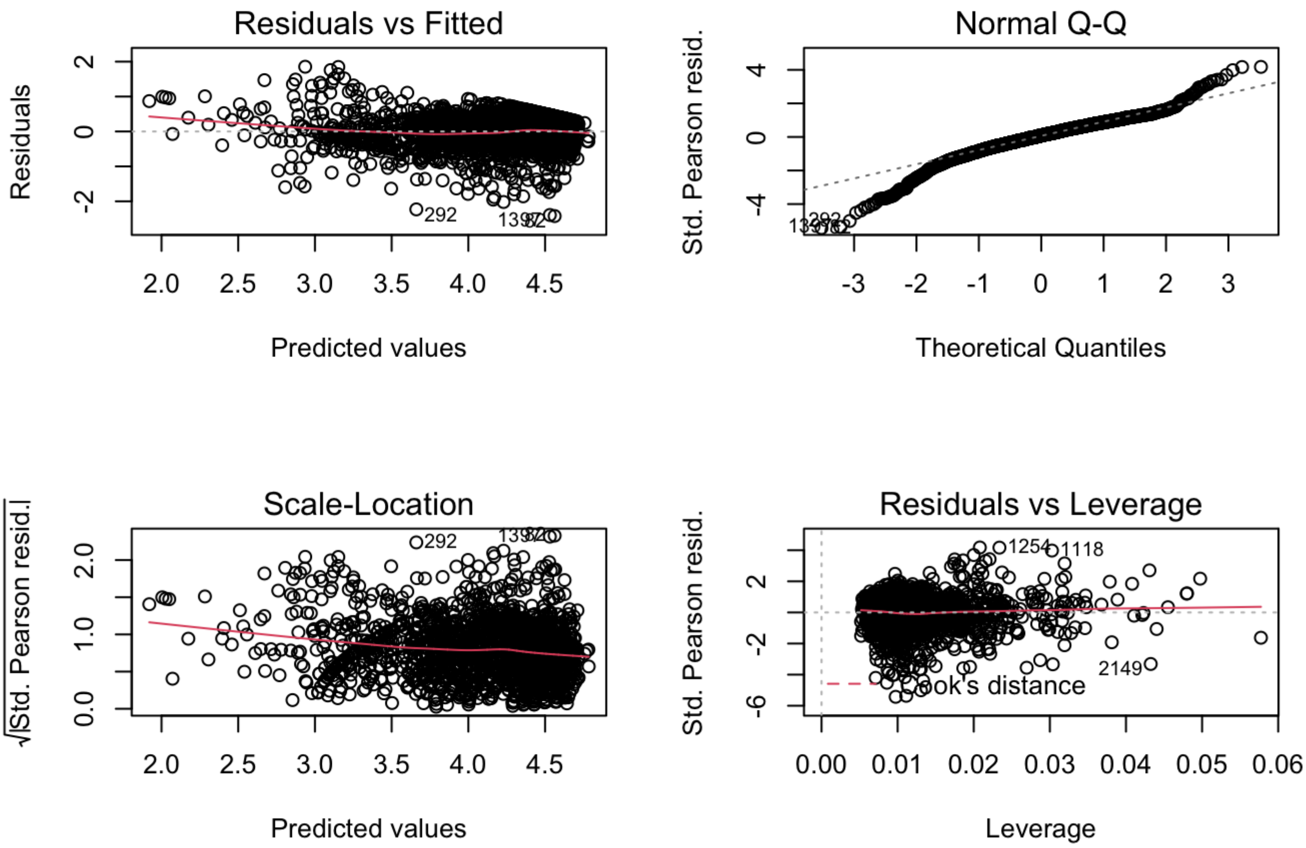
```
##
## Call:
## glm(formula = OverallAttitudeScore ~ Locality + Survey.type +
##     Age + Gender + Area.type + Education + Occupation + Aware.of.storks +
##     Support.reintroductions + Visited.Knepp + Frequency.exp.nature +
##     Member.of.Environmental.Organisation + Knowledge.Score +
##     NCI + ProCoBS + Bird.Interest.Score + Environmental.Concern.Score,
##     data = model_clean2)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.42238  -0.22565   0.04065   0.28324   1.85352
##
## Coefficients:
##                                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                             1.1813024  0.1171166  10.087  < 2e-16
## LocalityNot local                       0.0405340  0.0254770   1.591  0.11175
## Survey.typeProactive                    0.0894412  0.0311837   2.868  0.00417
## Age45-64                                0.0055097  0.0224555   0.245  0.80620
## Age65+                                  0.0734936  0.0343551   2.139  0.03252
## GenderMale                             -0.0790244  0.0198043  -3.990 6.81e-05
## Area.typeSub-urban                     -0.0339710  0.0228210  -1.489  0.13673
## Area.typeUrban                         -0.0580530  0.0274010  -2.119  0.03423
## EducationSecondary school/College      -0.0539131  0.0706747  -0.763  0.44564
## EducationUniversity graduate           -0.0670609  0.0708549  -0.946  0.34402
## OccupationNatural resource management   0.1196718  0.0629265   1.902  0.05733
## OccupationOther                         0.1778896  0.0326081   5.455 5.41e-08
## OccupationRetired                       0.1199462  0.0423321   2.833  0.00464
## OccupationUnemployed                    0.1104765  0.0556680   1.985  0.04731
## Aware.of.storksYes                      0.0542842  0.0309974   1.751  0.08004
## Support.reintroductions                 0.3652960  0.0123283  29.631  < 2e-16
## Visited.KneppYes                        0.0097004  0.0243175   0.399  0.69000
## Frequency.exp.nature1-2 days           -0.0559709  0.0481097  -1.163  0.24479
## Frequency.exp.nature3-4 days           -0.0747323  0.0499179  -1.497  0.13450
## Frequency.exp.nature5-6 days           -0.0677640  0.0516664  -1.312  0.18980
## Frequency.exp.natureEvery day, 7 days  -0.0699642  0.0518121  -1.350  0.17704
## Member.of.Environmental.OrganisationYes -0.0491476  0.0241044  -2.039  0.04157
## Knowledge.Score                        -0.0090446  0.0074973  -1.206  0.22779
## NCI                                     0.0019254  0.0004501   4.277 1.97e-05
## ProCoBS                                 0.0043376  0.0022504   1.928  0.05404
## Bird.Interest.Score                     0.0276994  0.0046316   5.981 2.57e-09
## Environmental.Concern.Score             0.0698409  0.0099940   6.988 3.63e-12
##
## (Intercept)                            ***
## LocalityNot local
## Survey.typeProactive                   **
## Age45-64
## Age65+                                 *
## GenderMale                             ***
## Area.typeSub-urban
## Area.typeUrban                         *
## EducationSecondary school/College
## EducationUniversity graduate
## OccupationNatural resource management  .
## OccupationOther                        ***
## OccupationRetired                      **
## OccupationUnemployed                   *
## Aware.of.storksYes                     .
## Support.reintroductions                ***
## Visited.KneppYes
## Frequency.exp.nature1-2 days
```

```
## Frequency.exp.nature3-4 days
## Frequency.exp.nature5-6 days
## Frequency.exp.natureEvery day, 7 days
## Member.of.Environmental.OrganisationYes *
## Knowledge.Score
## NCI                                     ***
## ProCoBS                                 .
## Bird.Interest.Score                     ***
## Environmental.Concern.Score             ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2011664)
##
##     Null deviance: 945.65  on 2329  degrees of freedom
## Residual deviance: 463.29  on 2303  degrees of freedom
## AIC: 2904.7
##
## Number of Fisher Scoring iterations: 2
```

```
vif(global_model1)
```

```
##                                         GVIF Df GVIF^(1/(2*Df))
## Locality                            1.534660  1        1.238814
## Survey.type                         2.378676  1        1.542296
## Age                                 1.931476  2        1.178887
## Gender                              1.126458  1        1.061347
## Area.type                           1.303015  2        1.068409
## Education                           1.229890  2        1.053093
## Occupation                          2.326391  4        1.111310
## Aware.of.storks                     1.647502  1        1.283551
## Support.reintroductions             1.336075  1        1.155887
## Visited.Knepp                       1.337559  1        1.156529
## Frequency.exp.nature                1.532538  4        1.054815
## Member.of.Environmental.Organisation 1.680234  1        1.296238
## Knowledge.Score                     1.953909  1        1.397823
## NCI                                 1.509849  1        1.228759
## ProCoBS                             1.706601  1        1.306370
## Bird.Interest.Score                 2.230890  1        1.493617
## Environmental.Concern.Score         1.874533  1        1.369136
```

```
# Check model residuals
par(mfrow = c(2, 2))
plot(global_model1)
```




```
# Calculate r2
with(summary(global_model1), 1 - deviance/null.deviance)
```

```
## [1] 0.5100863
```

```
## Sample sizes for the global model/dataset
c(table(model_clean2$SurveyType))
```

```
## integer(0)
```

```
c(table(model_clean2$SiteLocal))
```

```
## integer(0)
```

```
### Not in the model averaging = Freq.experience, SeenWildCaptivity, Heard of WSP, KnowledgeSco
re and BIS
```

Model selection

Caveats to model selection

- Depends on the models included in the candidate set.
- The parameter estimates and predictions arising from the “best” model or set of best models should be biologically meaningful.
- Need to decide whether to use model selection or common inferential statistics (e.g. based on P-values). Techniques that rely on both approaches are possible (e.g. backward variable selection followed by averaging of top models), such as the example provided above.

```r
#### MODEL SELECTION USING MUMIN PACKAGE
options(na.action = na.fail)

# Check VIF assumptions
vif(global_model1)
```

```
##                                          GVIF Df GVIF^(1/(2*Df))
## Locality                             1.534660  1        1.238814
## Survey.type                          2.378676  1        1.542296
## Age                                  1.931476  2        1.178887
## Gender                               1.126458  1        1.061347
## Area.type                            1.303015  2        1.068409
## Education                            1.229890  2        1.053093
## Occupation                           2.326391  4        1.111310
## Aware.of.storks                      1.647502  1        1.283551
## Support.reintroductions              1.336075  1        1.155887
## Visited.Knepp                        1.337559  1        1.156529
## Frequency.exp.nature                 1.532538  4        1.054815
## Member.of.Environmental.Organisation 1.680234  1        1.296238
## Knowledge.Score                      1.953909  1        1.397823
## NCI                                  1.509849  1        1.228759
## ProCoBS                              1.706601  1        1.306370
## Bird.Interest.Score                  2.230890  1        1.493617
## Environmental.Concern.Score          1.874533  1        1.369136
```

```r
# Dredge all possible models (model selection step)
attitude_dredge <- dredge(global_model1)
```

```
## Fixed term is "(Intercept)"
```

```r
# Summarise the top model
summary(get.models(attitude_dredge, 1)[[1]])
```

```
##
## Call:
## glm(formula = OverallAttitudeScore ~ Age + Aware.of.storks +
##     Bird.Interest.Score + Environmental.Concern.Score + Gender +
##     Member.of.Environmental.Organisation + NCI + Occupation +
##     ProCoBS + Support.reintroductions + Survey.type + 1, data = model_clean2)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.43297  -0.22891   0.03835   0.28536   1.87709
##
## Coefficients:
##                                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)                           1.0785740  0.0891888  12.093  < 2e-16
## Age45-64                              0.0115558  0.0219744   0.526  0.59902
## Age65+                                0.0802038  0.0336759   2.382  0.01732
## Aware.of.storksYes                    0.0427079  0.0297891   1.434  0.15180
## Bird.Interest.Score                   0.0265641  0.0044798   5.930 3.49e-09
## Environmental.Concern.Score           0.0696420  0.0099400   7.006 3.20e-12
## GenderMale                           -0.0811876  0.0193005  -4.207 2.69e-05
## Member.of.Environmental.OrganisationYes -0.0529033  0.0233036  -2.270  0.02329
## NCI                                   0.0018436  0.0004459   4.135 3.68e-05
## OccupationNatural resource management   0.1298176  0.0623229   2.083  0.03736
## OccupationOther                       0.1755924  0.0317352   5.533 3.50e-08
## OccupationRetired                     0.1220165  0.0419751   2.907  0.00369
## OccupationUnemployed                  0.1177999  0.0549034   2.146  0.03201
## ProCoBS                               0.0041515  0.0022156   1.874  0.06110
## Support.reintroductions               0.3652849  0.0122734  29.762  < 2e-16
## Survey.typeProactive                  0.0678547  0.0275117   2.466  0.01372
##
## (Intercept)                          ***
## Age45-64
## Age65+                               *
## Aware.of.storksYes
## Bird.Interest.Score                  ***
## Environmental.Concern.Score          ***
## GenderMale                           ***
## Member.of.Environmental.OrganisationYes *
## NCI                                  ***
## OccupationNatural resource management   *
## OccupationOther                      ***
## OccupationRetired                    **
## OccupationUnemployed                 *
## ProCoBS                              .
## Support.reintroductions              ***
## Survey.typeProactive                 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2011393)
##
##     Null deviance: 945.65  on 2329  degrees of freedom
## Residual deviance: 465.44  on 2314  degrees of freedom
## AIC: 2893.4
##
## Number of Fisher Scoring iterations: 2
```

```
#### Use a all-subsets model subsetting approach find a confidence set of models, recalculating weights each time:
attitude_del <- subset(attitude_dredge, delta <= 2, recalc.weights = TRUE) # delta(AIC) cutoff
# Save results table as a CSV file
# write.csv(as.data.frame(attitude_del), "Attitude_model_selection.csv")
```

Model averaging

As we have so many predictors in the global model it's unlikely that only one model explains all the variation in the data. From the dredge output we can see there is little difference in the AIC and weights of the first few models.

But how do we decide which model(s) to interpret? It's agreed that models with delta AIC (or other criterion) less than 2 are considered to be just as good as the top model, so therefore shouldn't be discounted. Additionally, we could use the weights: if a model has weight greater or equal to 95% then it is likely to be the top model. Otherwise we can generate a "credibility" set consisting of all models whose cumulative sum of AIC weights is 0.95.

In any case, the point is that we have no good reason to exclude models other than the top one when the next models after it are likely to be just as good. Therefore, model averaging (AKA multi-model inference), is used to average the parameter estimates across multiple models and avoids the issue of model uncertainty. See below for the code and results of model averaging on this dataset for all models with a delta AIC <= 2.

Key references

- Harrison XA, Donaldson L, Correa-Cano ME, Evans J, Fisher DN, Goodwin CED, Robinson BS, Hodgson DJ, Inger R. 2018. A brief introduction to mixed effects modelling and multi-model inference in ecology. PeerJ 6:e4794 https://doi.org/10.7717/peerj.4794 (https://doi.org/10.7717/peerj.4794)

```
### Model averaging the top 10 models according to the delta AIC value
attitude_aves <- model.avg(get.models(attitude_del, subset = delta < 2))
summary(attitude_aves)
```

```
##
## Call:
## model.avg(object = get.models(attitude_del, subset = delta <
##     2))
##
## Component model call:
## glm(formula = OverallAttitudeScore ~ <19 unique rhs>, data =
##      model_clean2)
##
## Component models:
##                                df   logLik    AICc delta weight
## 1/3/4/5/6/9/10/11/12/13/14       17 -1429.72 2893.71  0.00   0.09
## 1/4/5/6/9/10/11/12/13/14         16 -1430.76 2893.75  0.04   0.09
## 1/3/4/5/6/7/9/10/11/12/13/14     18 -1429.00 2894.30  0.59   0.07
## 1/2/4/5/6/8/9/10/11/12/13/14     19 -1428.02 2894.37  0.67   0.06
## 1/2/3/4/5/6/8/9/10/11/12/13/14   20 -1427.08 2894.52  0.81   0.06
## 1/2/4/5/6/9/10/11/12/13/14       18 -1429.12 2894.54  0.84   0.06
## 1/3/4/5/6/8/9/10/11/12/13/14     18 -1429.17 2894.63  0.92   0.06
## 1/2/3/4/5/6/9/10/11/12/13/14     19 -1428.18 2894.69  0.99   0.05
## 1/4/5/6/8/9/10/11/12/13/14       17 -1430.22 2894.70  1.00   0.05
## 1/2/3/4/5/6/7/8/9/10/11/12/13/14 21 -1426.21 2894.82  1.12   0.05
## 1/3/4/5/6/7/8/9/10/11/12/13/14   19 -1428.41 2895.15  1.44   0.04
## 1/2/3/4/5/6/7/9/10/11/12/13/14   20 -1427.39 2895.15  1.45   0.04
## 1/4/5/6/7/9/10/11/12/13/14       17 -1430.45 2895.16  1.45   0.04
## 1/3/4/5/6/9/10/11/13/14          16 -1431.49 2895.21  1.50   0.04
## 1/4/5/6/9/10/11/13/14            15 -1432.59 2895.38  1.67   0.04
## 1/2/4/5/6/8/9/10/11/13/14        18 -1429.56 2895.42  1.71   0.04
## 1/2/3/4/5/6/8/9/10/11/13/14      19 -1428.56 2895.45  1.74   0.04
## 1/2/4/5/6/7/8/9/10/11/12/13/14   20 -1427.60 2895.57  1.86   0.04
## 1/3/4/5/6/9/10/11/12/13/14/15    18 -1429.68 2895.65  1.94   0.03
##
## Term codes:
##                                Age                          Area.type
##                                  1                                  2
##                    Aware.of.storks                Bird.Interest.Score
##                                  3                                  4
##         Environmental.Concern.Score                             Gender
##                                  5                                  6
##                    Knowledge.Score                           Locality
##                                  7                                  8
## Member.of.Environmental.Organisation                                NCI
##                                  9                                 10
##                         Occupation                            ProCoBS
##                                 11                                 12
##             Support.reintroductions                        Survey.type
##                                 13                                 14
##                      Visited.Knepp
##                                 15
##
## Model-averaged coefficients:
## (full average)
##                                       Estimate Std. Error Adjusted SE
## (Intercept)                          1.0830097  0.0927118   0.0927587
## Age45-64                             0.0097791  0.0222881   0.0222995
## Age65+                               0.0790709  0.0340733   0.0340910
## Aware.of.storksYes                   0.0264293  0.0323625   0.0323710
## Bird.Interest.Score                  0.0274171  0.0045442   0.0045465
## Environmental.Concern.Score          0.0696528  0.0099925   0.0099977
## GenderMale                          -0.0793417  0.0195638   0.0195739
## Member.of.Environmental.OrganisationYes -0.0502763  0.0236232   0.0236353
## NCI                                  0.0018881  0.0004489   0.0004492
```

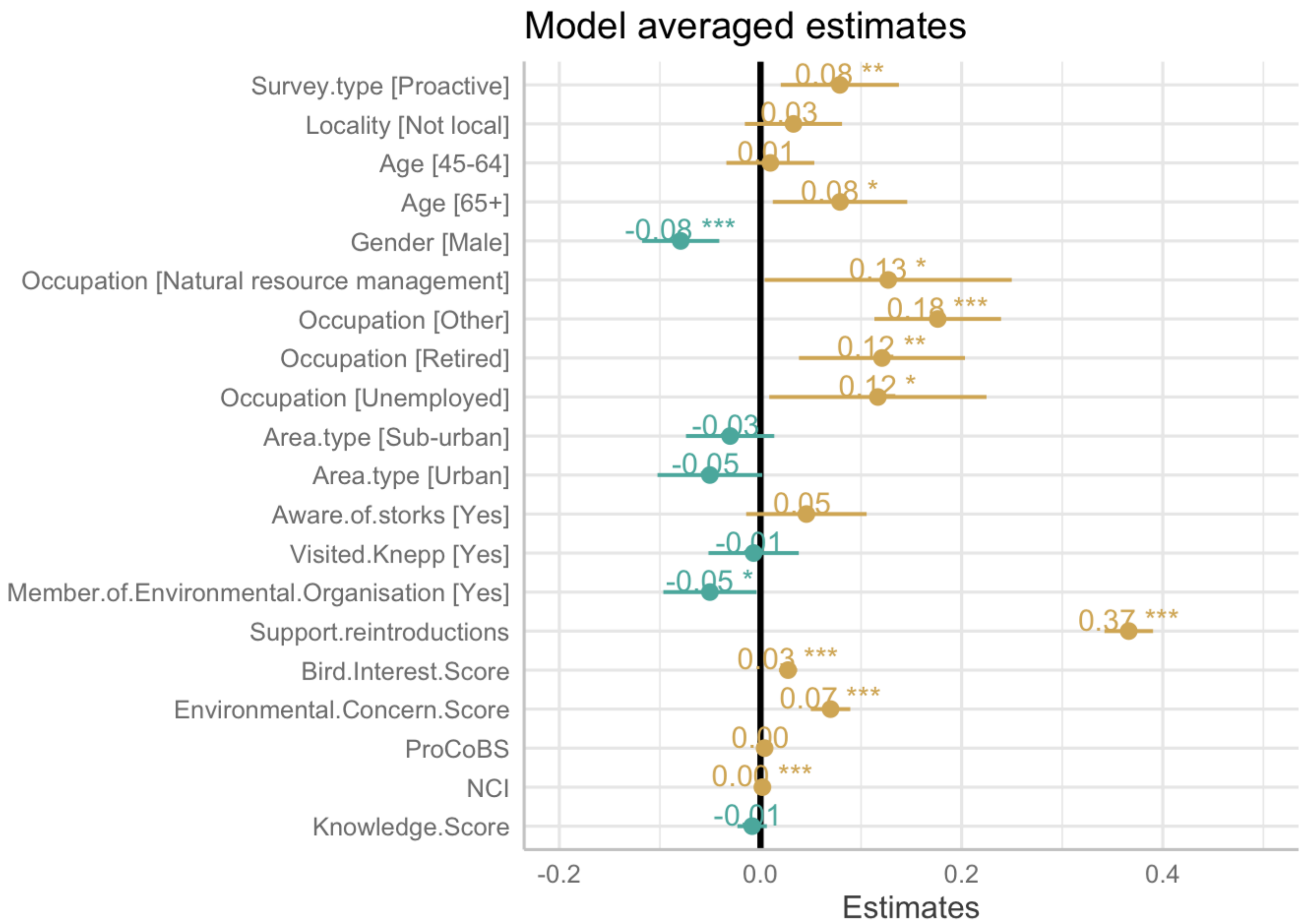
```
## OccupationNatural resource management    0.1270019  0.0626367   0.0626693
## OccupationOther                          0.1762143  0.0321020   0.0321186
## OccupationRetired                        0.1207787  0.0420784   0.0421003
## OccupationUnemployed                     0.1167043  0.0550769   0.0551056
## ProCoBS                                  0.0034515  0.0025288   0.0025296
## Support.reintroductions                  0.3661471  0.0123119   0.0123183
## Survey.typeProactive                     0.0788863  0.0299502   0.0299644
## Knowledge.Score                         -0.0023548  0.0054633   0.0054647
## Area.typeSub-urban                      -0.0133834  0.0211372   0.0211426
## Area.typeUrban                          -0.0223157  0.0306206   0.0306258
## LocalityNot local                        0.0144386  0.0230631   0.0230688
## Visited.KneppYes                        -0.0002313  0.0043909   0.0043930
##                                         z value Pr(>|z|)
## (Intercept)                              11.676  < 2e-16 ***
## Age45-64                                  0.439  0.66100
## Age65+                                    2.319  0.02037 *
## Aware.of.storksYes                        0.816  0.41424
## Bird.Interest.Score                       6.030  < 2e-16 ***
## Environmental.Concern.Score               6.967  < 2e-16 ***
## GenderMale                                4.053 5.05e-05 ***
## Member.of.Environmental.OrganisationYes   2.127  0.03341 *
## NCI                                       4.204 2.63e-05 ***
## OccupationNatural resource management     2.027  0.04271 *
## OccupationOther                           5.486  < 2e-16 ***
## OccupationRetired                         2.869  0.00412 **
## OccupationUnemployed                      2.118  0.03419 *
## ProCoBS                                   1.364  0.17243
## Support.reintroductions                  29.724  < 2e-16 ***
## Survey.typeProactive                      2.633  0.00847 **
## Knowledge.Score                           0.431  0.66654
## Area.typeSub-urban                        0.633  0.52673
## Area.typeUrban                            0.729  0.46621
## LocalityNot local                         0.626  0.53139
## Visited.KneppYes                          0.053  0.95802
##
## (conditional average)
##                                          Estimate Std. Error Adjusted SE
## (Intercept)                              1.0830097  0.0927118   0.0927587
## Age45-64                                 0.0097791  0.0222881   0.0222995
## Age65+                                   0.0790709  0.0340733   0.0340910
## Aware.of.storksYes                       0.0456503  0.0305217   0.0305373
## Bird.Interest.Score                      0.0274171  0.0045442   0.0045465
## Environmental.Concern.Score              0.0696528  0.0099925   0.0099977
## GenderMale                              -0.0793417  0.0195638   0.0195739
## Member.of.Environmental.OrganisationYes -0.0502763  0.0236232   0.0236353
## NCI                                      0.0018881  0.0004489   0.0004492
## OccupationNatural resource management    0.1270019  0.0626367   0.0626693
## OccupationOther                          0.1762143  0.0321020   0.0321186
## OccupationRetired                        0.1207787  0.0420784   0.0421003
## OccupationUnemployed                     0.1167043  0.0550769   0.0551056
## ProCoBS                                  0.0040931  0.0022265   0.0022276
## Support.reintroductions                  0.3661471  0.0123119   0.0123183
## Survey.typeProactive                     0.0788863  0.0299502   0.0299644
## Knowledge.Score                         -0.0083061  0.0074733   0.0074771
## Area.typeSub-urban                      -0.0301902  0.0223708   0.0223823
## Area.typeUrban                          -0.0503398  0.0265397   0.0265532
## LocalityNot local                        0.0327694  0.0246275   0.0246396
## Visited.KneppYes                        -0.0068174  0.0228799   0.0228919
##                                         z value Pr(>|z|)
## (Intercept)                              11.676  < 2e-16 ***
## Age45-64                                  0.439  0.66100
## Age65+                                    2.319  0.02037 *
```

```
## Aware.of.storksYes                         1.495  0.13494
## Bird.Interest.Score                        6.030  < 2e-16 ***
## Environmental.Concern.Score                6.967  < 2e-16 ***
## GenderMale                                 4.053 5.05e-05 ***
## Member.of.Environmental.OrganisationYes   2.127  0.03341 *
## NCI                                        4.204 2.63e-05 ***
## OccupationNatural resource management     2.027  0.04271 *
## OccupationOther                            5.486  < 2e-16 ***
## OccupationRetired                          2.869  0.00412 **
## OccupationUnemployed                       2.118  0.03419 *
## ProCoBS                                    1.837  0.06615 .
## Support.reintroductions                   29.724  < 2e-16 ***
## Survey.typeProactive                       2.633  0.00847 **
## Knowledge.Score                            1.111  0.26662
## Area.typeSub-urban                         1.349  0.17739
## Area.typeUrban                             1.896  0.05799 .
## LocalityNot local                          1.330  0.18354
## Visited.KneppYes                           0.298  0.76585
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# View the model estimates
sjPlot::plot_model(attitude_aves, type = "est", vline.color = "black", order.terms = c(15,19,1,
2,6,9,10,11,12,17,18,3,20,7,14,4,5,13,8,16),
                   show.values = TRUE, value.offset = .3, title = "Model averaged estimates") +
ylim(-.2, 0.5) +
  scale_color_wsp("likert") + theme_sjplot()
```

```
## Scale for 'y' is already present. Adding another scale for 'y', which will
## replace the existing scale.
```

```
## Scale for 'colour' is already present. Adding another scale for 'colour',
## which will replace the existing scale.
```




```
# ggplot2::ggsave(filename = "Attitude_averaging_table.png", width = 7, height = 8, dpi = 300)


# Print model averaged coefficients in table
attitude_confint <-as.table(round(confint(attitude_aves), 3))
attitude_exp <- as.table(exp(coefficients(attitude_aves)))
attitude_ave <- round(summary(attitude_aves)$coefmat.subset, 3)
# stargazer(attitude_ave, digits=3, title="Model averaged results", type = "html",
#           out="Attitude_averages_table.doc")

# Create tab_df table of model averaged estimates
attitude_ave1 <- tibble::rownames_to_column(as.data.frame(attitude_ave), "Predictor")
sjPlot::tab_df(attitude_ave1, title = "Model averaged results", alternate.rows = TRUE, digits=3
, use.viewer = TRUE)
```

Model averaged results

Predictor	*Estimate*	*Std..Error*	*Adjusted.SE*	*z.value*	*Pr...z..*
(Intercept)	1.083	0.093	0.093	11.676	0.000
Age45-64	0.010	0.022	0.022	0.439	0.661
Age65+	0.079	0.034	0.034	2.319	0.020
Aware.of.storksYes	0.046	0.031	0.031	1.495	0.135
Bird.Interest.Score	0.027	0.005	0.005	6.030	0.000
Environmental.Concern.Score	0.070	0.010	0.010	6.967	0.000
GenderMale	-0.079	0.020	0.020	4.053	0.000
Member.of.Environmental.OrganisationYes	-0.050	0.024	0.024	2.127	0.033

NCI	0.002	0.000	0.000	4.204	0.000
OccupationNatural resource management	0.127	0.063	0.063	2.027	0.043
OccupationOther	0.176	0.032	0.032	5.486	0.000
OccupationRetired	0.121	0.042	0.042	2.869	0.004
OccupationUnemployed	0.117	0.055	0.055	2.118	0.034
ProCoBS	0.004	0.002	0.002	1.837	0.066
Support.reintroductions	0.366	0.012	0.012	29.724	0.000
Survey.typeProactive	0.079	0.030	0.030	2.633	0.008
Knowledge.Score	-0.008	0.007	0.007	1.111	0.267
Area.typeSub-urban	-0.030	0.022	0.022	1.349	0.177
Area.typeUrban	-0.050	0.027	0.027	1.896	0.058
LocalityNot local	0.033	0.025	0.025	1.330	0.184
Visited.KneppYes	-0.007	0.023	0.023	0.298	0.766

Exploring the relationship between attitudes and Q15. “Do you support the reintroduction of WS to southern England?”

Agreement with each of the 14 Likert item attitude statements (14 statements used to create the attitude composite score - Appendix 4 and Figure S4) did not vary significantly between survey samples, which is reflected by a high level of support for the reintroduction of white storks to southern England across all samples (Proactive = 91.2%, Nat.rep. = 74.8% - Table S8). Significant positive relationships were found between respondent’s overall attitude scores and those who selected ‘Yes’ (1.406 ± 0.054, $p < 0.001$) and ‘Not sure’ (0.393 ± 0.066, $p < 0.001$) when asked if they support the reintroduction of white storks to southern England, compared to those who selected ‘No’. A positive interaction was also found between support for reintroductions and locality, with local people who were either supportive (0.422 ± 0.120, $p < 0.001$) or not sure (0.460 ± 0.162, $p < 0.001$) of reintroductions, having higher attitude scores than unsupportive non-local respondents. However, overall non-local respondents were more supportive than local respondents (-0.332 ± 0.118, $p < 0.01$).

```r
# Select variables and run model within a pipe %>%
model_support <- final_data %>%
  dplyr::select(SurveyType, SiteProximity, OverallAttitudeScore, Q15_WSP_support) %>%
  drop_na() %>%  # N = 2,492
  lm(OverallAttitudeScore ~ Q15_WSP_support*SiteProximity, data = .)

# View model summary
summary(model_support)
```

```
##
## Call:
## lm(formula = OverallAttitudeScore ~ Q15_WSP_support * SiteProximity,
##     data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.91325 -0.27325  0.06677  0.36675  1.59955
##
## Coefficients:
##                                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                               2.86700    0.05241  54.705  < 2e-16
## Q15_WSP_supportNot sure                   0.39345    0.06574   5.985 2.47e-09
## Q15_WSP_supportYes                        1.40625    0.05394  26.071  < 2e-16
## SiteProximityYes                         -0.33155    0.11825  -2.804 0.005090
## Q15_WSP_supportNot sure:SiteProximityYes  0.46023    0.16219   2.838 0.004582
## Q15_WSP_supportYes:SiteProximityYes       0.42152    0.12045   3.499 0.000474
##
## (Intercept)                              ***
## Q15_WSP_supportNot sure                  ***
## Q15_WSP_supportYes                       ***
## SiteProximityYes                         **
## Q15_WSP_supportNot sure:SiteProximityYes **
## Q15_WSP_supportYes:SiteProximityYes      ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4972 on 2486 degrees of freedom
## Multiple R-squared:  0.3924, Adjusted R-squared:  0.3912
## F-statistic: 321.1 on 5 and 2486 DF,  p-value: < 2.2e-16
```