Week 5 Lab Report

E

3/8/2021

Interacting with Migrants

Goal: Recommend habitat purchases of either anthro or natural sites for migrant birds.

Write a brief summary describing findings and recommendations describing your findings and recommendations (no more than 2 pages of single spaced writing, not including code or figures). Summary should include your process and results, including 1-2 supporting figures for model interpretation/prediction, and 1-2 figures for evaluating model assumptions.

Summary

Upon review of the data available, we suggest that purchase should be for natural sites to attract migratory birds. To achieve this result, we conducted a two way analysis answering the question, *is there a significant difference in number of birds based on the type of bird (migratory or resident) and type of site (anthro influenced or natural sites)*? We followed the modeling process (detailed below in notes). After evaluating models, the best model was the full model, which included the influence of both site and bird type. We achieved this through AIC model selection. We then evaluated the model to meet the assumptions for an ANOVA model.

These figures below represent the validation of assumptions for the model. In the `Residuals vs Fitted` plot, we see slight variation in residuals. Our `Q-Q plot` shows that our data is normally distributed. Based on these figures, we can move forward with saying our data meets our model's assumptions

```
## hat values (leverages) are all = 0.03571429
##  and there are no factor predictors; no plot no. 5
```
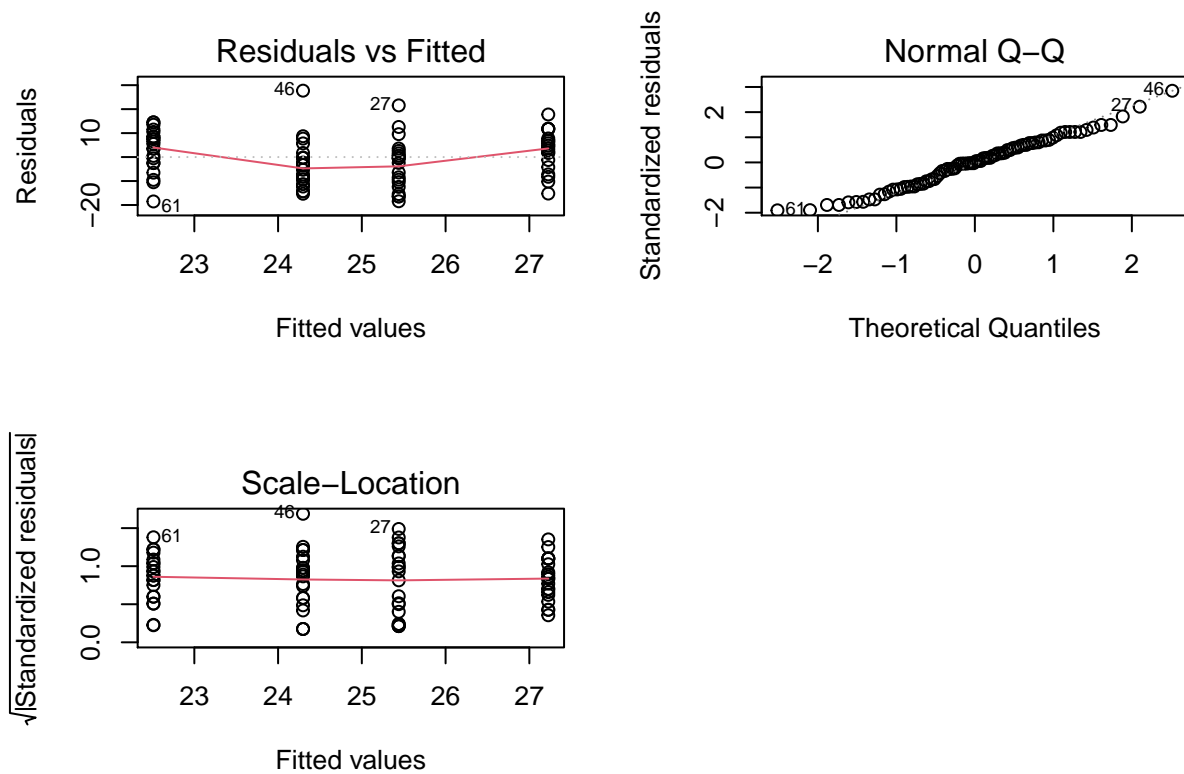
To answer our question regarding which plot to purchase we can look at our model outputs. The coefficients output of the model is as follows:

```
##   (Intercept) urb.catnatural    migstatresi
##     25.440476      -2.928571       1.785714
```

The intercept represents our reference group, migratory at anthro sites. The group labeled `urb.catnatural` represents the difference between migratory birds at natural sites and anthro sites. The coefficient labeled `migstatresi` is difference of resident birds from migratory birds at anthro sites
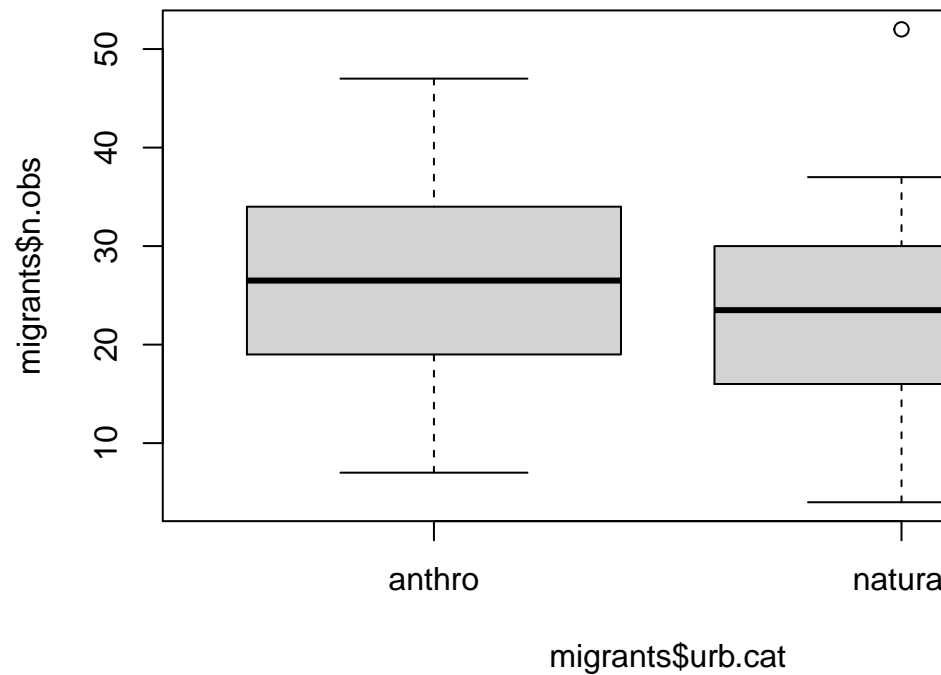
Below is a graph that help decide our results
When examining difference between anthro and natural sites, we can be see that anthro sites have large
bird observations. However, when we consider our model and look within those observations, we see that
the bird makeup within those sites vary.

```
## `summarise()` has grouped output by 'urb.cat'. You can override using the `.groups` argument.
```

3

Number of Birds per site
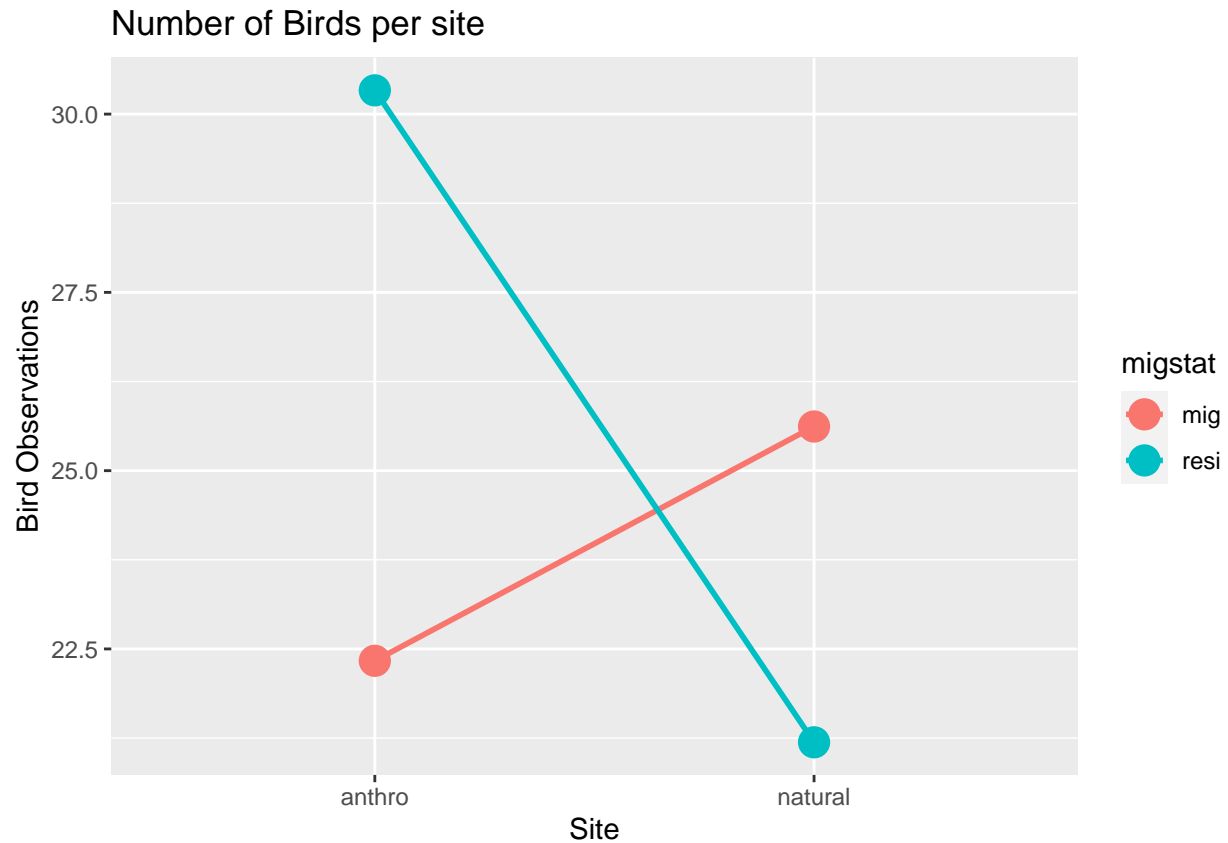
With the goal of providing habitat for as many migrants birds as possible, we suggest investing in natural sites. We notice a slight increase in the presence of birds in natural sites however, our model does not conclude that it was a significant difference (p-value = 0.179).

WorkSpace

```
migrants <- read.csv(file = "Data/migrants.csv", header = TRUE)
str(migrants)
```

```
## 'data.frame':    84 obs. of  4 variables:
##  $ site   : chr  "01.Gill" "01.Gill" "02.Colrain" "02.Colrain" ...
##  $ migstat: chr  "mig" "resi" "mig" "resi" ...
##  $ urb.cat: chr  "anthro" "anthro" "natural" "natural" ...
##  $ n.obs  : int  25 17 31 17 34 15 28 29 17 12 ...
```

```
head(migrants)
```

```
##            site migstat urb.cat n.obs
## 1      01.Gill     mig  anthro    25
## 2      01.Gill    resi  anthro    17
## 3   02.Colrain     mig natural    31
```

```
## 4   02.Colrain    resi natural    17
## 5 03.Green GTD     mig natural    34
## 6 03.Green GTD    resi natural    15
```

```
# We notice our data has both migrant and resident birds
table(migrants$migstat)
```

```
##
##  mig resi
##   42   42
```

```
summary(migrants)
```

```
##      site              migstat             urb.cat              n.obs
##  Length:84          Length:84          Length:84          Min.   : 4.00
##  Class :character   Class :character   Class :character   1st Qu.:17.00
##  Mode  :character   Mode  :character   Mode  :character   Median :25.50
##                                                           Mean   :24.87
##                                                           3rd Qu.:32.00
##                                                           Max.   :52.00
```
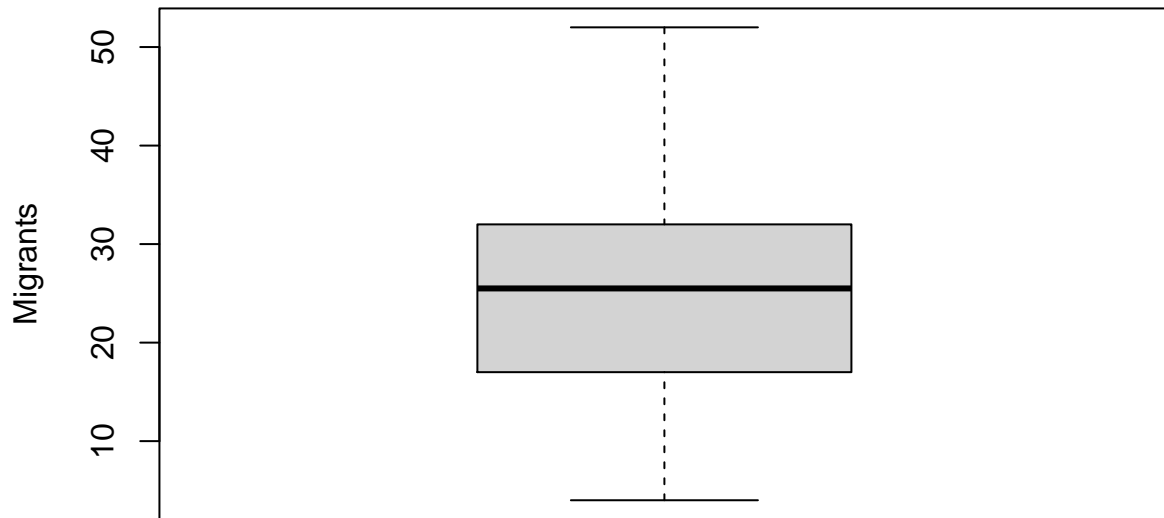
STEPS

Develop a set of hypotheses (null and alternative) based on project goals
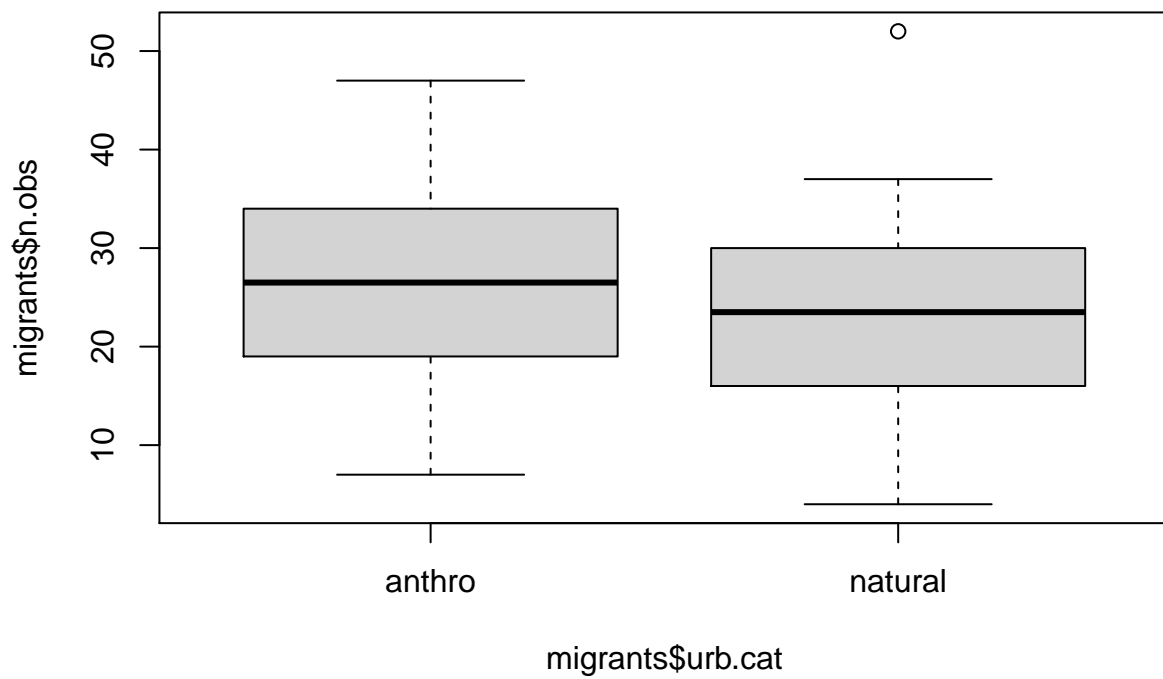
1. State the question
   - Is there a significant difference in number of birds among antho and natural sites between migrant and resident birds?
   - Response:
     - Number of birds observed
   - Explanatory:
     - sites - `anthro` or `nat`
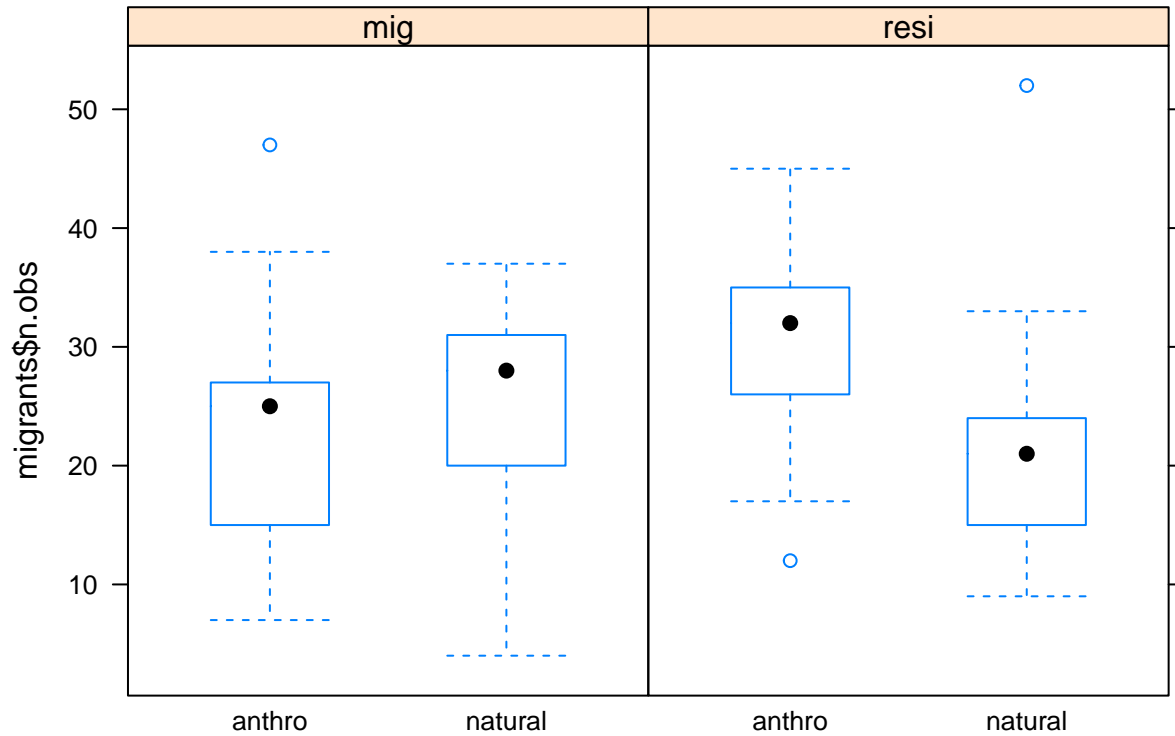   - migstat - `mig` or `resi`
2. Data Exploration

```
boxplot(migrants$n.obs, ylab = "Migrants")
```

```
boxplot(migrants$n.obs ~ migrants$urb.cat)
```

```
bwplot(migrants$n.obs ~ migrants$urb.cat|migrants$migstat)
```

```
tapply(migrants$n.obs, migrants$urb.cat, mean)
```

```
##    anthro   natural
## 26.33333  23.40476
```

```
tapply(migrants$n.obs, migrants$migstat, mean)# Raw means
```

```
##      mig     resi
## 23.97619 25.76190
```

Looks like there may be a difference between migratory and resident birds.

Describe all relevant statisical model(s) in words and algebra

3. Describe the model:

In words: Is there a difference between the habitat means of migrants observed between migrants and residents? * H_0: There is no difference

- In Mathematical form
  - $y_i = beta_0 + beta_{1(g)}Site_{1i(g)} + beta_{2(g)}Migstat + 2i(g) + e_i$
  - y_i is number of birds observed
- Model assumptions are:
  - Residuals are normally distributed

- Contant variance (homogeneity)
- Observations are independent
- Predictors measured without error (fixed X)

Fit candidate models and evaluate using AIC to select best candidate model

4. Fit the model

```
mSiteMig <- lm(n.obs ~ migstat + urb.cat, data = migrants)
summary(mSiteMig)
```

```
##
## Call:
## lm(formula = n.obs ~ migstat + urb.cat, data = migrants)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -18.5119  -8.2440   0.0952   6.9524  27.7024
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      25.440      1.871  13.598   <2e-16 ***
## migstatresi       1.786      2.160   0.827    0.411
## urb.catnatural   -2.929      2.160  -1.356    0.179
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.9 on 81 degrees of freedom
## Multiple R-squared:  0.03018,    Adjusted R-squared:  0.006238
## F-statistic:  1.26 on 2 and 81 DF,  p-value: 0.289
```

```
coef(mSiteMig)
```

```
##    (Intercept)    migstatresi urb.catnatural
##      25.440476       1.785714      -2.928571
```

```
tapply(migrants$n.obs, list(migrants$migstat, migrants$urb.cat), mean, na.rm = T)
```

```
##          anthro   natural
## mig   22.33333 25.61905
## resi  30.33333 21.19048
```

Here are our coefficients along with the computed means for each group.

Our first estimate of the intercept is the estimate of the mean number of birds observed for migrants at the anthro site. The difference between > Evaluate and validate the top model(s)

5. Evaluate the model - Model validation and model selection

9

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

```
## hat values (leverages) are all = 0.03571429
##  and there are no factor predictors; no plot no. 5
```
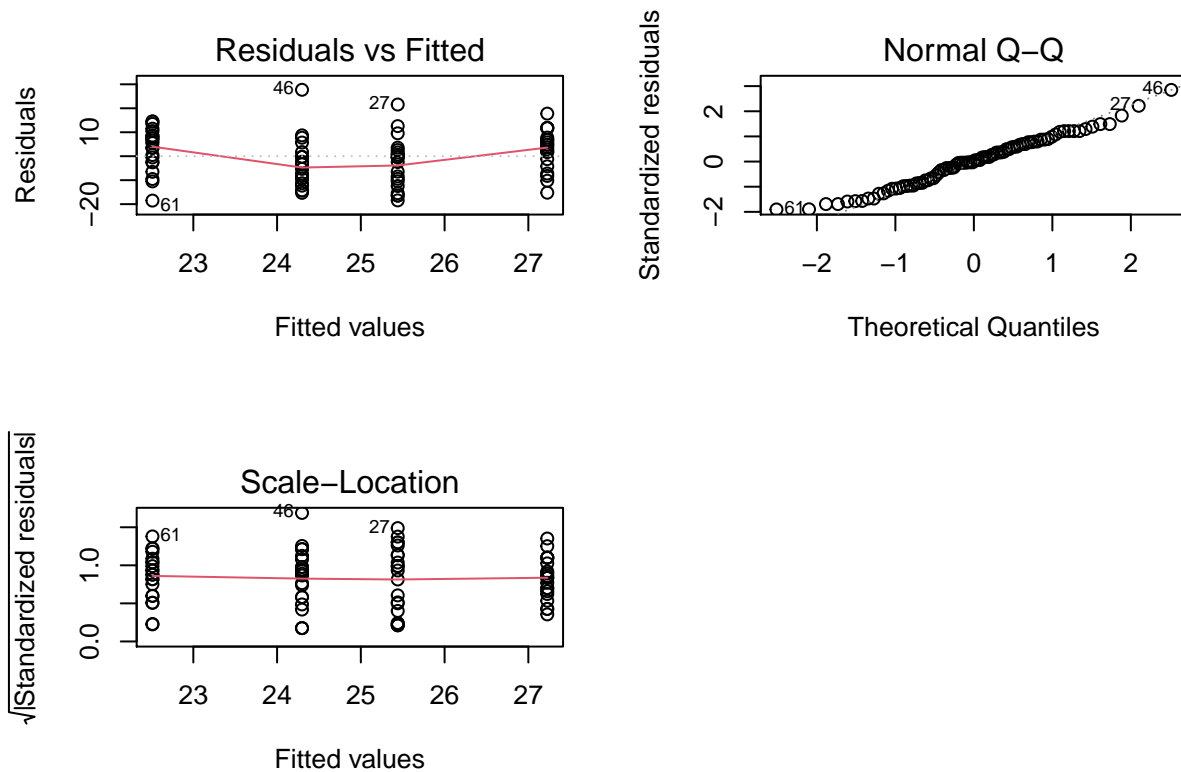

The results look to meet our assumptions.

```
# Model Selection
#Full Model
#Site only
# Migratory status only
# Null model

modlist <- list()
modlist[["mSiteMig"]] <- lm(n.obs ~ migstat + urb.cat, data = migrants)
modlist[["mSite"]] <- lm(n.obs ~ urb.cat, data = migrants)
modlist[["mMig"]] <- lm(n.obs ~ migstat, data = migrants)
modlist[["mNull"]] <- lm(n.obs ~ 1, data = migrants)
# AIC
aictab(modlist)
```

```
##
## Model selection based on AICc:
##
```

```
##           K   AICc Delta_AICc AICcWt Cum.Wt      LL
## mNull    2 627.19       0.00   0.36   0.36 -311.52
## mSite    3 627.47       0.28   0.31   0.68 -310.59
## mMig     3 628.65       1.46   0.17   0.85 -311.18
## mSiteMig 4 628.98       1.78   0.15   1.00 -310.23
```

Hmmm... We see that our null best describes our model but we can't get much info from that. All four models fall between 2 deltaAICs. Lets keep the full model to interpret our data

> Interpret results, including description of all model parameters and what estimates mean including graphical and verbal summaries of the model predictions

6. Interpret Results

```
mSiteMig <- lm(n.obs~ urb.cat + migstat, data = migrants)
summary(mSiteMig)
```

```
##
## Call:
## lm(formula = n.obs ~ urb.cat + migstat, data = migrants)
##
## Residuals:
##      Min      1Q   Median      3Q     Max
## -18.5119  -8.2440   0.0952   6.9524  27.7024
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      25.440      1.871  13.598   <2e-16 ***
## urb.catnatural   -2.929      2.160  -1.356    0.179
## migstatresi       1.786      2.160   0.827    0.411
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.9 on 81 degrees of freedom
## Multiple R-squared:  0.03018,    Adjusted R-squared:  0.006238
## F-statistic: 1.26 on 2 and 81 DF,  p-value: 0.289
```

```
coef(mSiteMig)
```

```
##    (Intercept) urb.catnatural    migstatresi
##      25.440476      -2.928571       1.785714
```
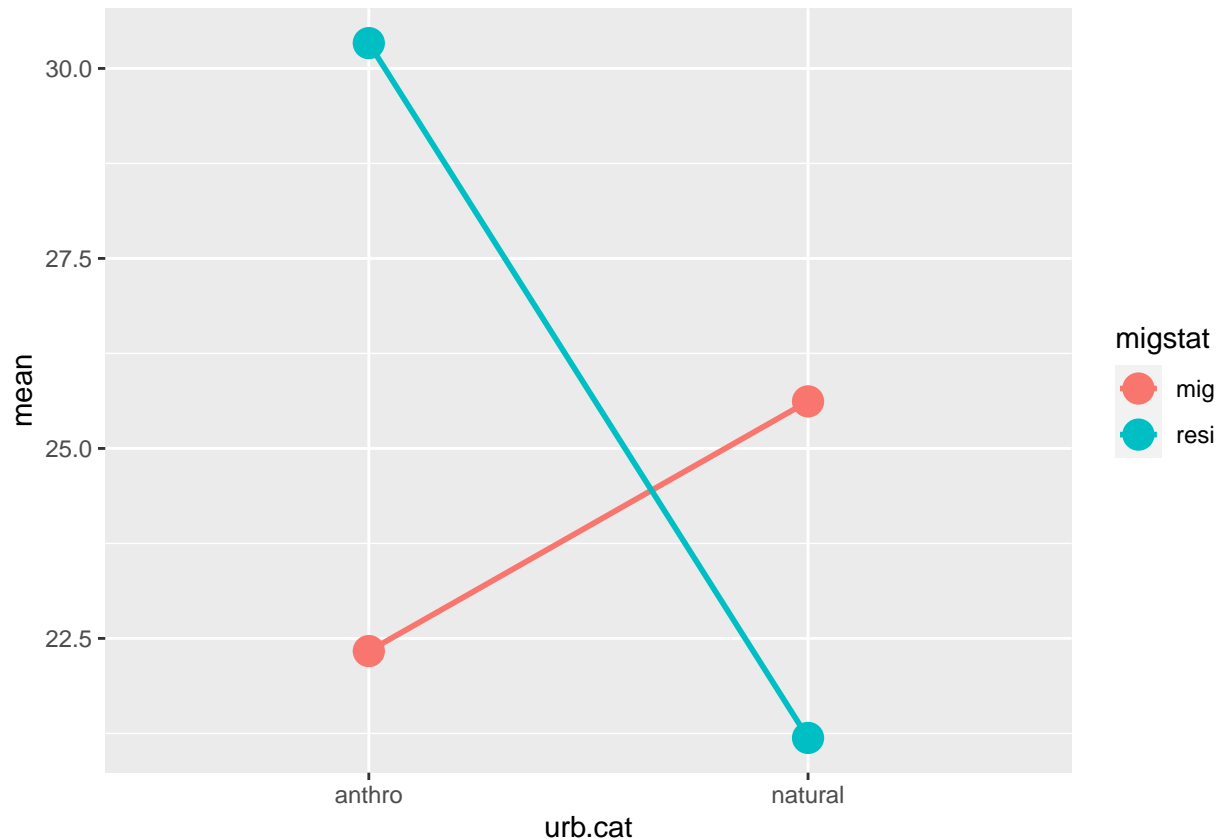
```
tapply(migrants$n.obs, list(migrants$migstat, migrants$urb.cat), mean, na.rm = T)
```

```
##         anthro   natural
## mig  22.33333  25.61905
## resi 30.33333  21.19048
```
```

```
plotdata <- migrants %>%
  group_by(urb.cat, migstat) %>%
  summarize(n = n(),
            mean = mean(n.obs),
            sd = sd(n.obs),
            se = sd/sqrt(n),
            ci = qt(0.975, df = n-1 * sd/ sqrt(n)))
```

`summarise()` has grouped output by 'urb.cat'. You can override using the `.groups` argument.

```
par(mfrow=c(1,1))
ggplot(plotdata,
       aes(x = urb.cat,
           y = mean,
       group= migstat,
       color = migstat))+
  geom_point(size = 5) +
  geom_line(size =1)
```


Lets explain the output

- intercept = migratory at anthro sites
- urb.catnatural = difference between migratory at natural sites of anthro sites
- migstatresi = difference of residents from migratory at anthro sites

  Include an annotated R script or do this as an R Markdowon file