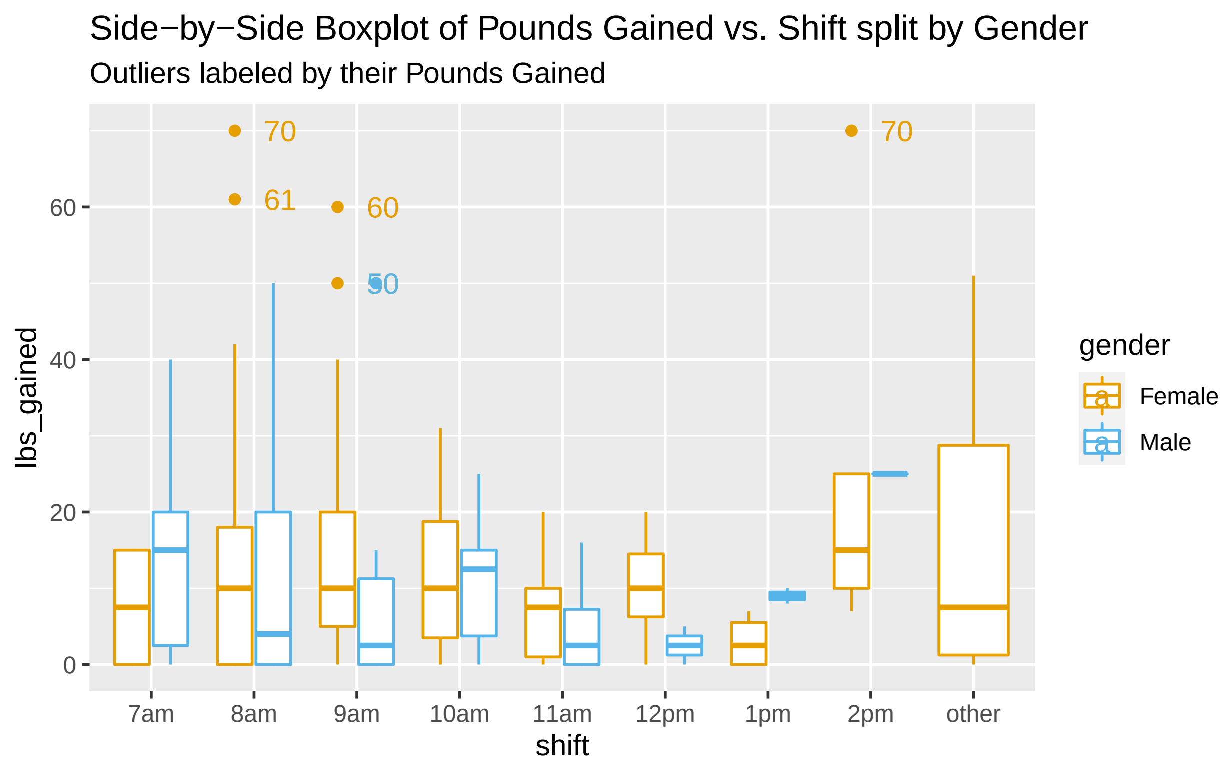
Effects of Work Shift and Total Excersize Time on Weight Gain

Paul Hunt, Brandon Kill, Brendan Winters

9/25/2021

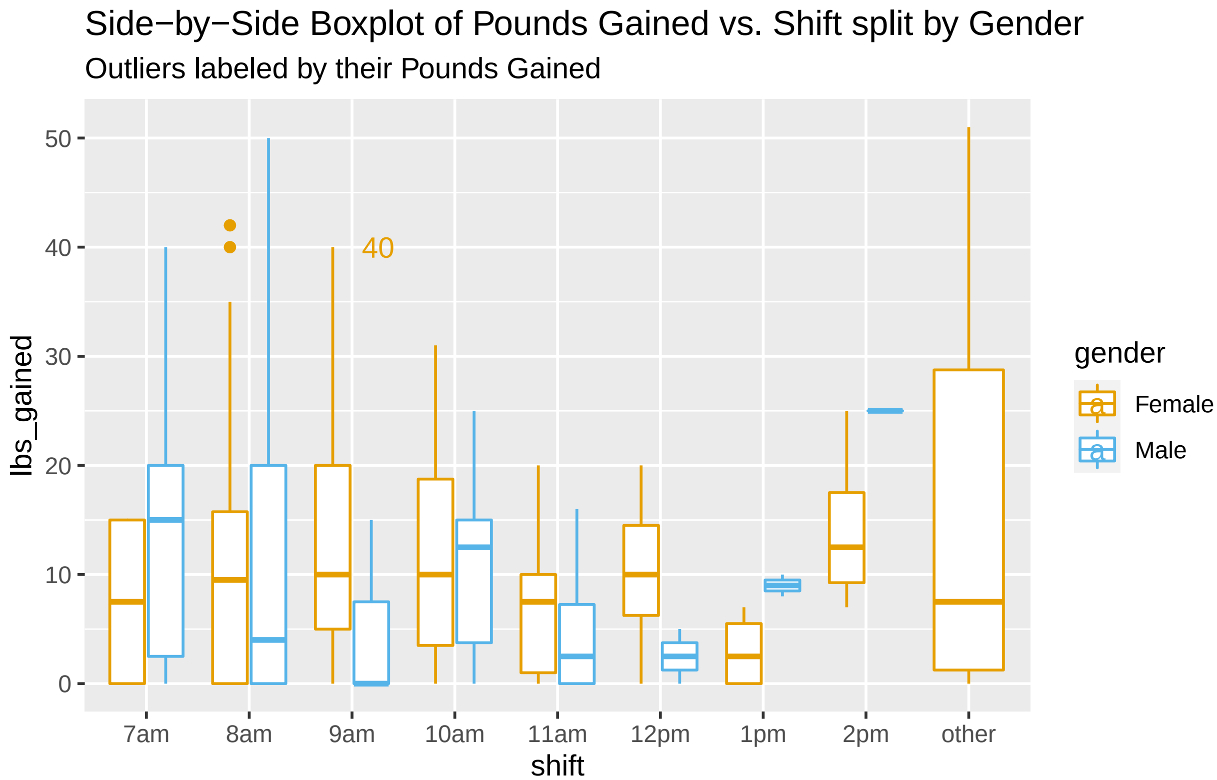
Removing Outliers

In a data set this large, there are bound to be some extreme cases. It is more useful to try and explain how weight gain behaves for the *vast majority* of people than it is to try and explain every single observation. So, for good practice, we will ignore any observations that we deem too extreme.



We see here 6 observations that count as outliers in their group. Each of the values would have been outliers in nearly any other group as well, reassuring us that removing them is an appropriate step to take.

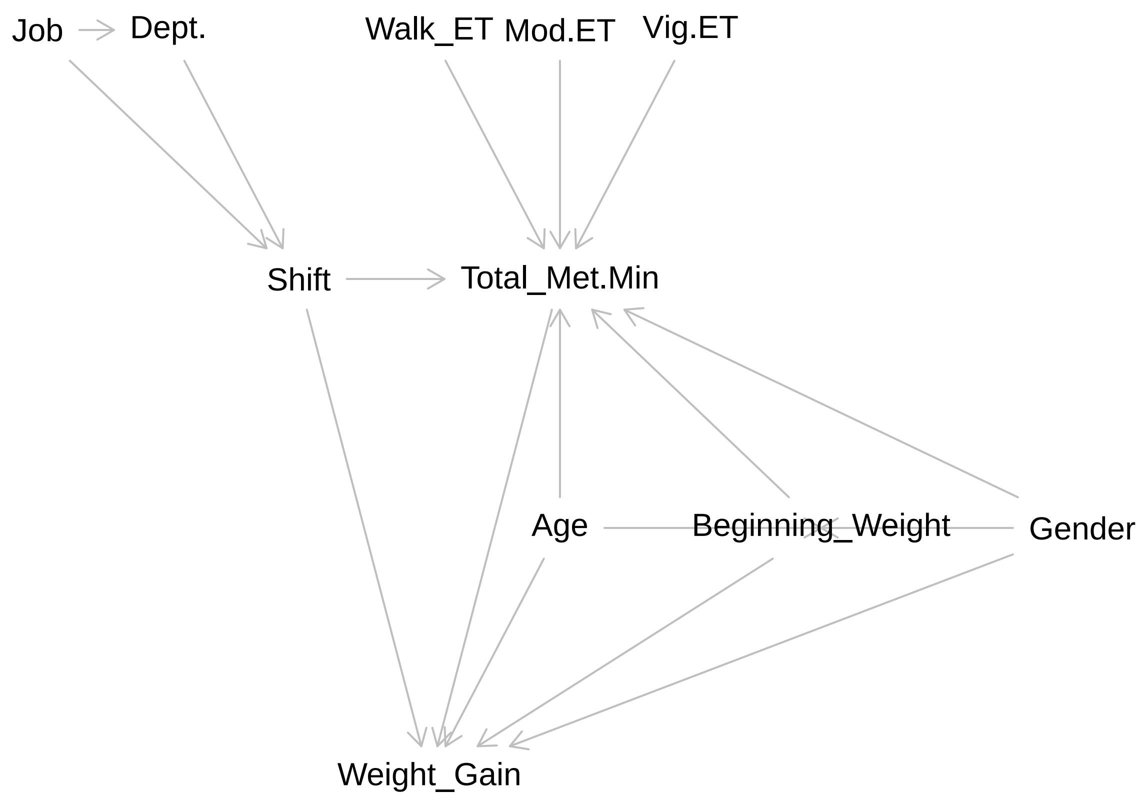
As shift and gender are both included in the model,



Only two outliers remain after removing the first six. However, these would not be outliers if they fell into the category shift = `other`, enticing us to leave them in.

After the six extreme cases are removed, there are now two observations that qualify as outliers by the IQR rule. However, these two values (females gaining 40 and 42 pounds) would be well within the range of the IQR rule if they were in the "other" shift category. Because of this, removing these values feels like too aggressive of a maneuver at this juncture, so we will leave those two observations in.

Causal Inference



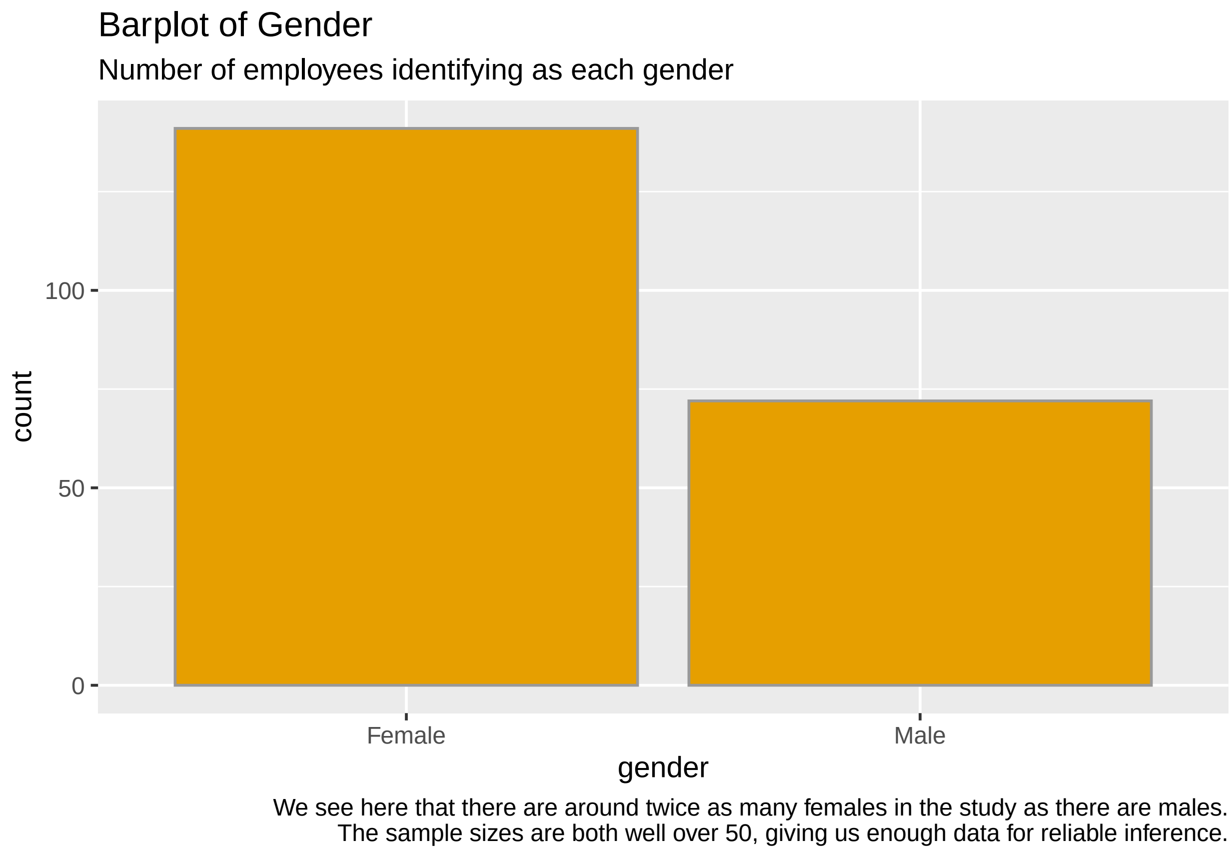
What follows is a list of the predictors under consideration and their hypothesized effects on weight gain and hypothesized interactions with other predictors.

- **Total Metabolic Minutes: This variable represents total exercise time for respondents in an average week, weighted by the intensity of the exercise. We expect exercise to have a direct effect on weigh gain, most likely with an inverse relationship.**
- **Shift: This variable indicates the shift in which respondents begin their work day. For ease of interpretation, we have chosen to model it as an ordinal variable, allowing us to analyze the effect of starting work earlier or later. We expect shift to have an effect on total metabolic minutes, since workers may struggle to fit regular exercise into their daily routines depending on their work schedules. Additionally, we are interested in modeling any potential direct effect of shift on weight gain to see if the call center could implement some change to their schedules to facilitate better health.**
- **Age: We expect the respondent's age to have some effect on total metabolic minutes, since people may find more vigorous exercise more difficult as they age. It may also affect the respondent's beginning weight. Additionally, we expect some direct effect of age on weight gain as metabolism changes with age.**
- **Gender: The respondent's gender likely affects total metabolic minutes as men and women might tend to engage in different types of exercise on average, and we can expect beginning weight to be affected since men are heavier than women on average. It may also have an effect on weight gain, assuming that the is some difference between the metabolic processes of mean and women on average.**

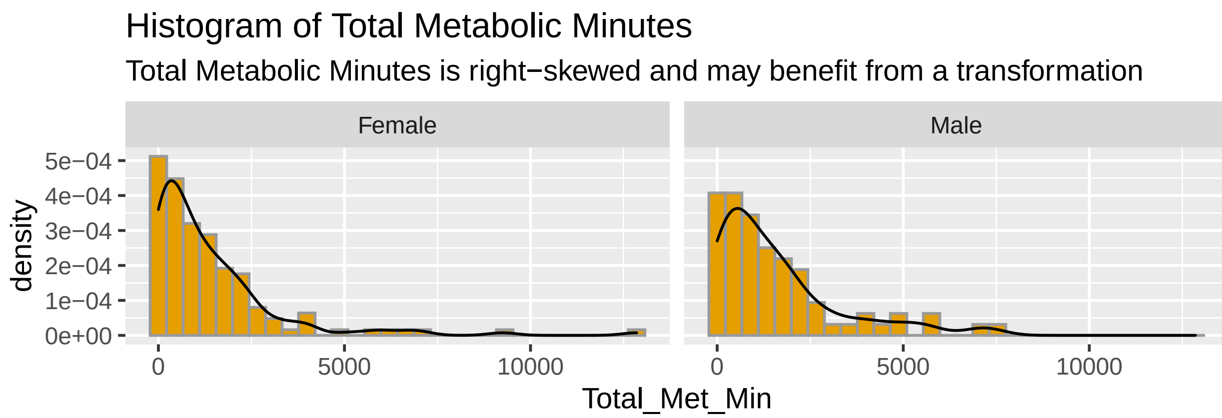
- **Beginning Weight: We can expect some influence of initial weight on total metabolic minutes, since the respondent's weight may inform us to their propensity to exercise. It may also impact weight gain, assuming that it serves as a proxy for overall health or innate metabolic levels.**

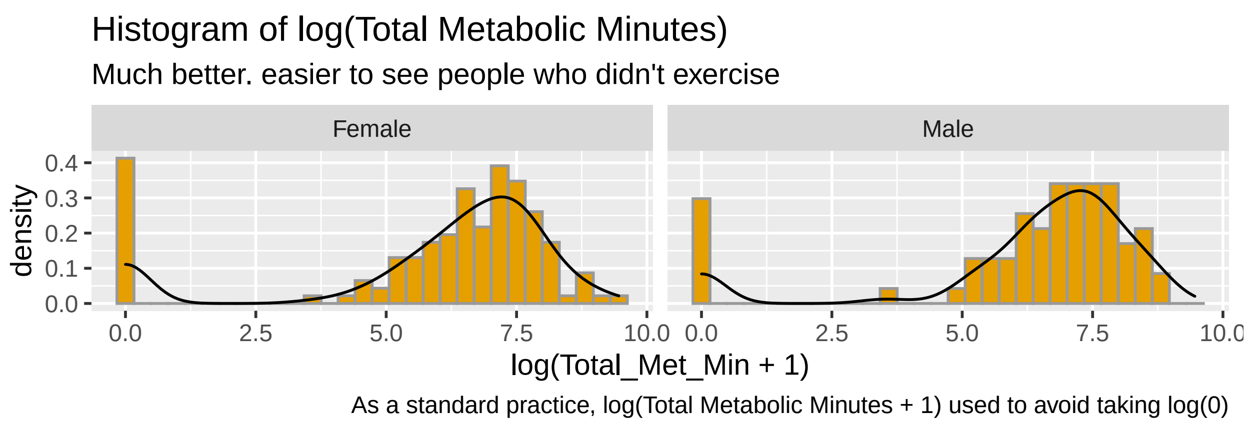
Thinking about Transformation - Exploratory Data Analysis

Gender



Total Metabolic Minutes

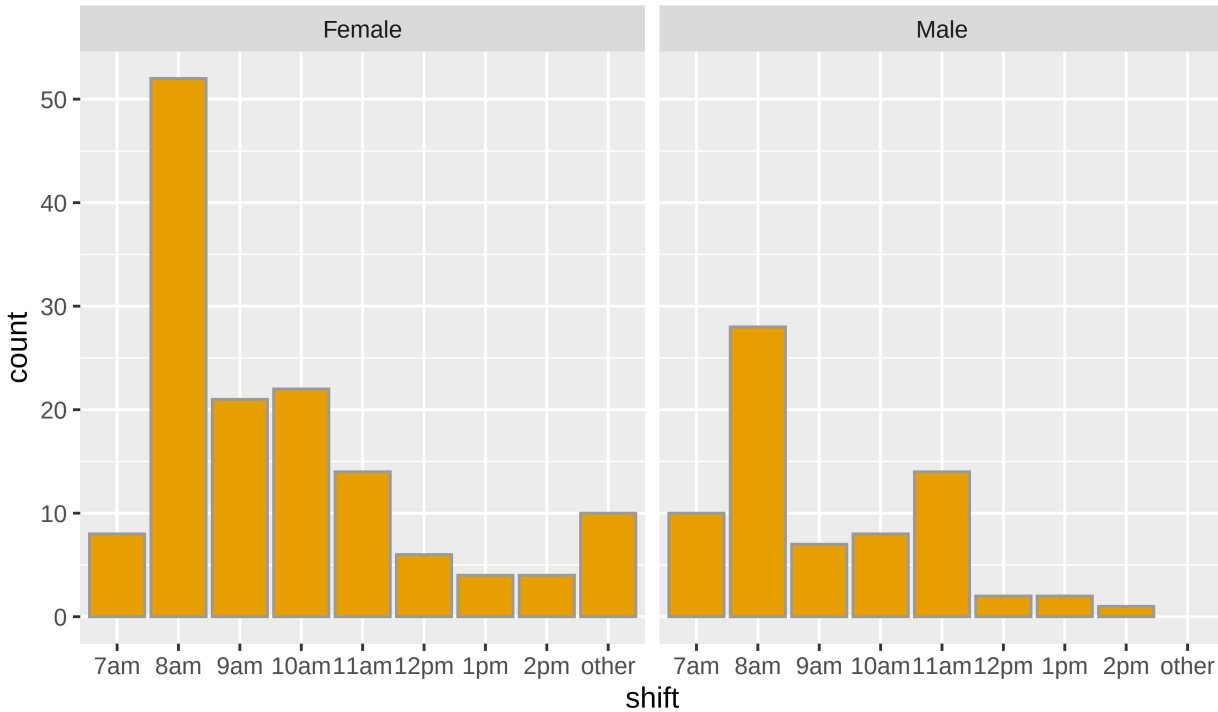




Shift

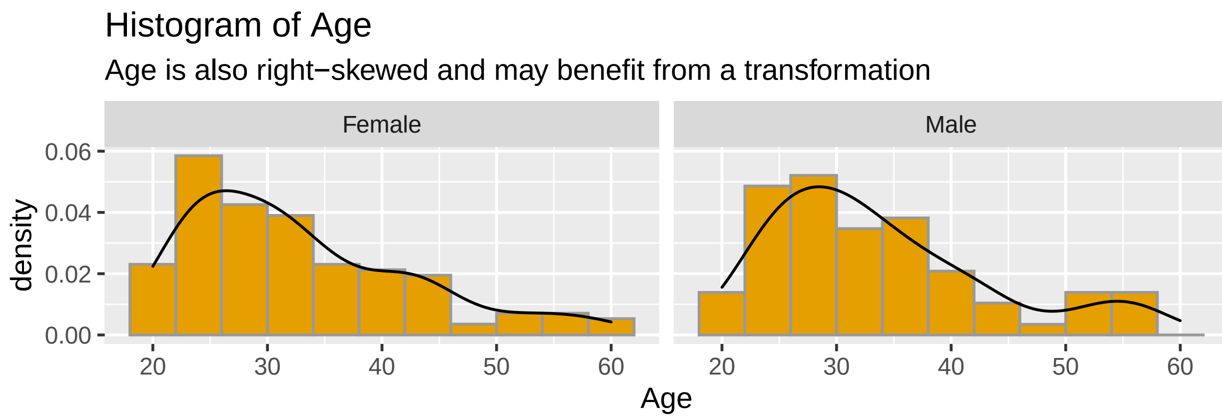
Barplot of Shift

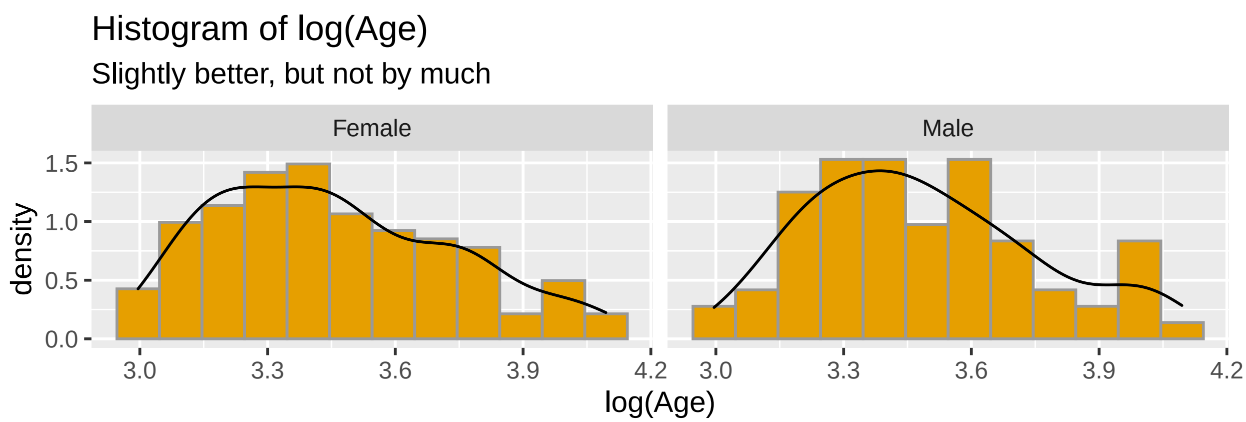
Number of employees in each shift displayed



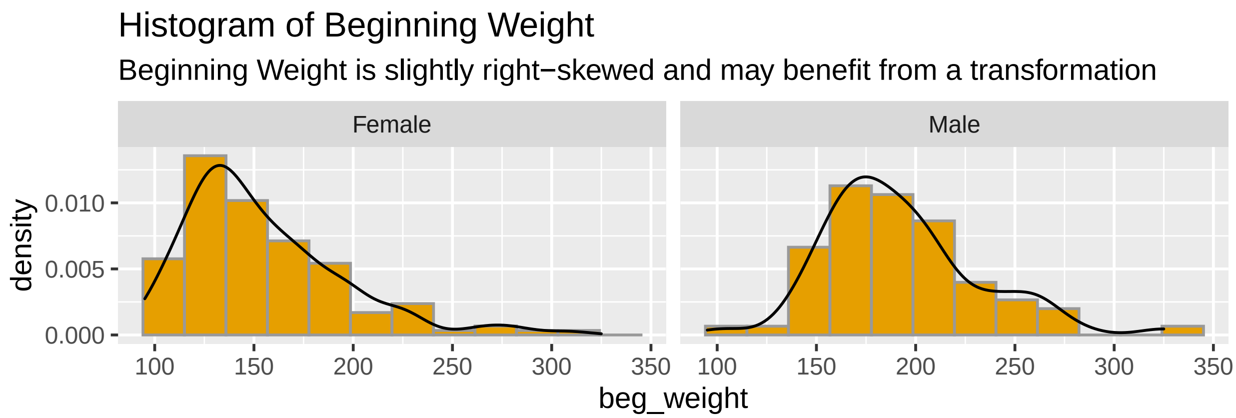
We may want to make shift into an ordinal variable, as 7am is objectively earlier than 8 am, etc.

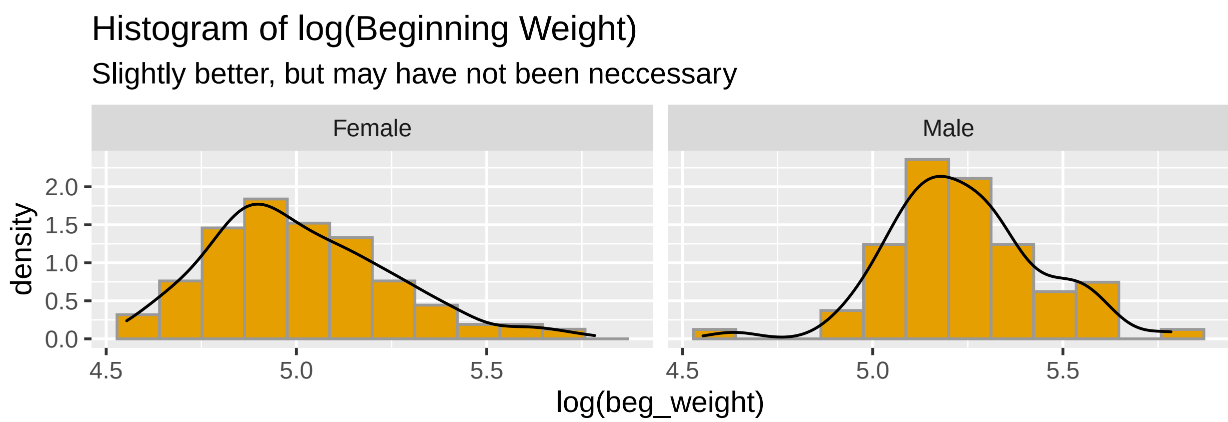
Age


Histogram of Age
Age is also right−skewed and may benefit from a transformation
Female
Male
density
0.06
0.04
0.02
0.00
20
30
40
50
60
Age


Histogram of log(Age)
Slightly better, but not by much
Female
Male
density
1.5
1.0
0.5
0.0
3.0
3.3
3.6
3.9
4.2
log(Age)

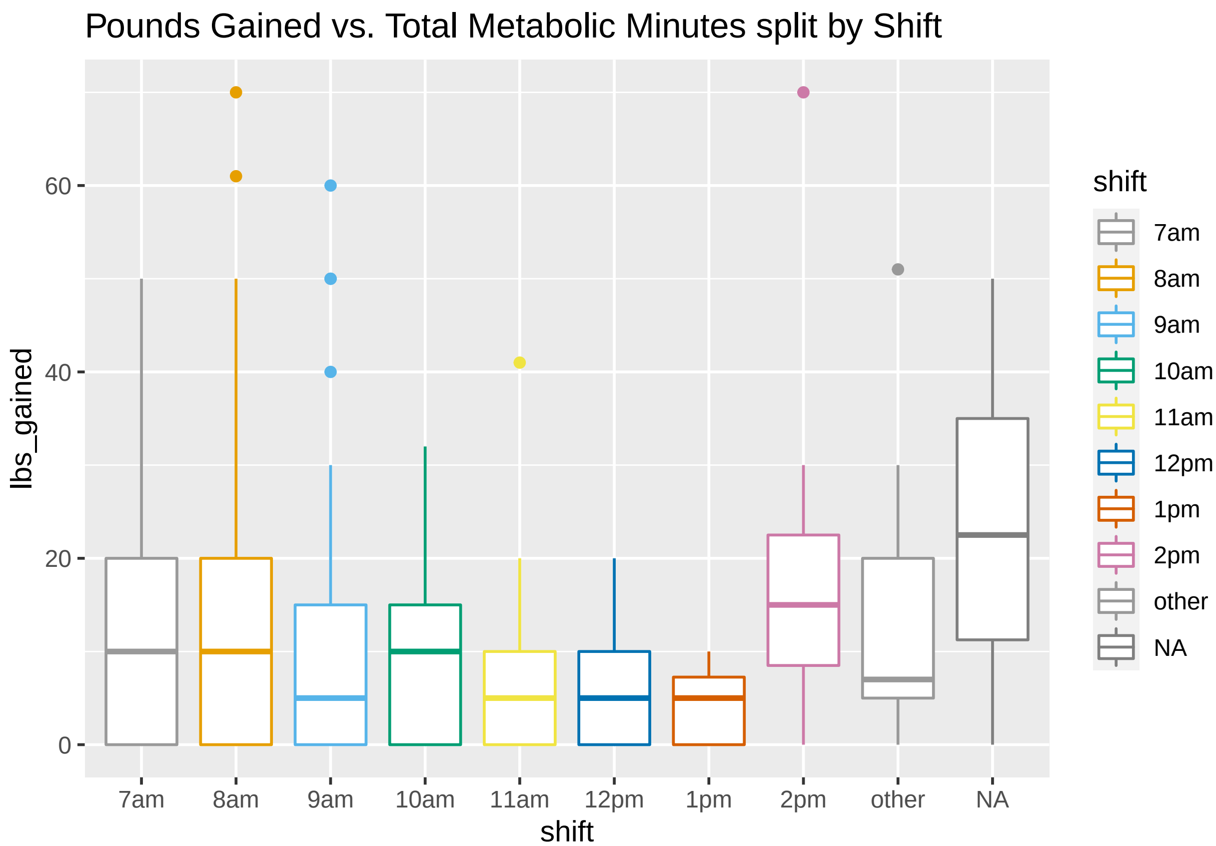
Beginning Weight

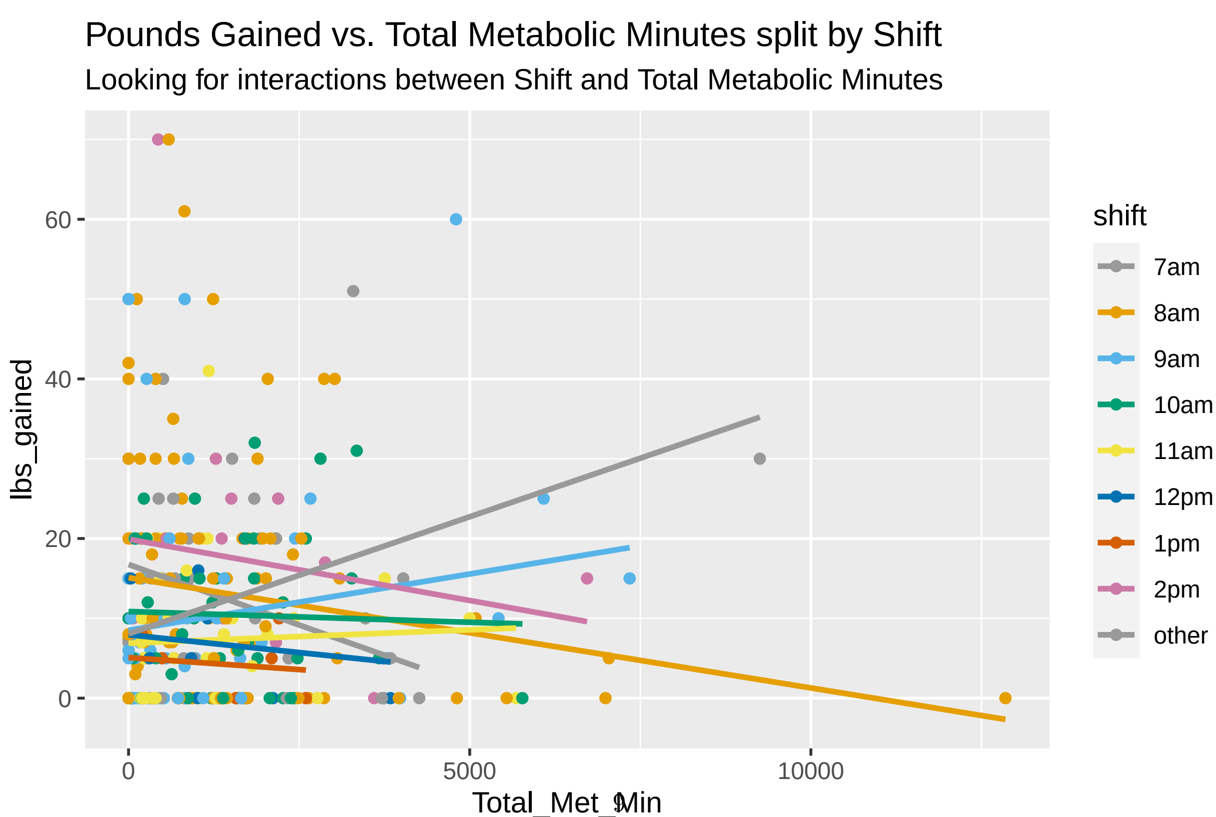




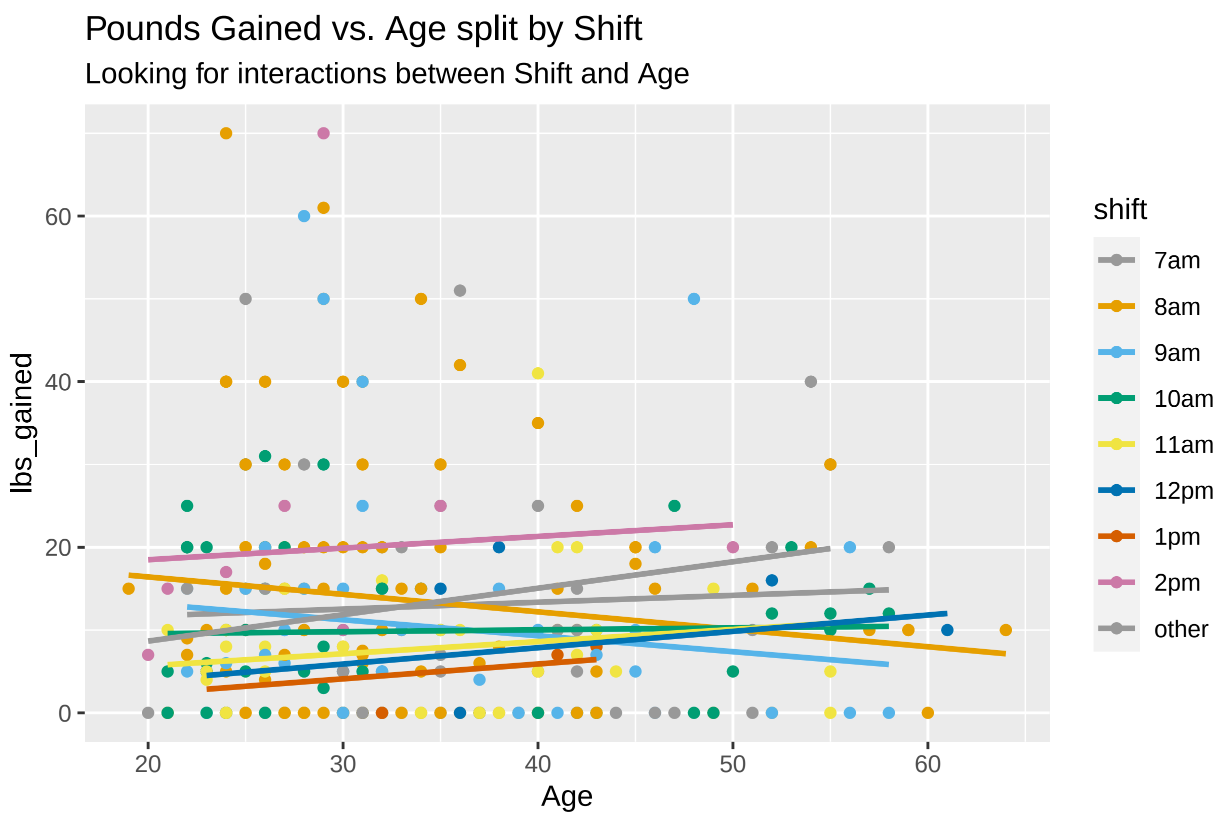
Checking for Interactions

Interaction between Predictors and Shift

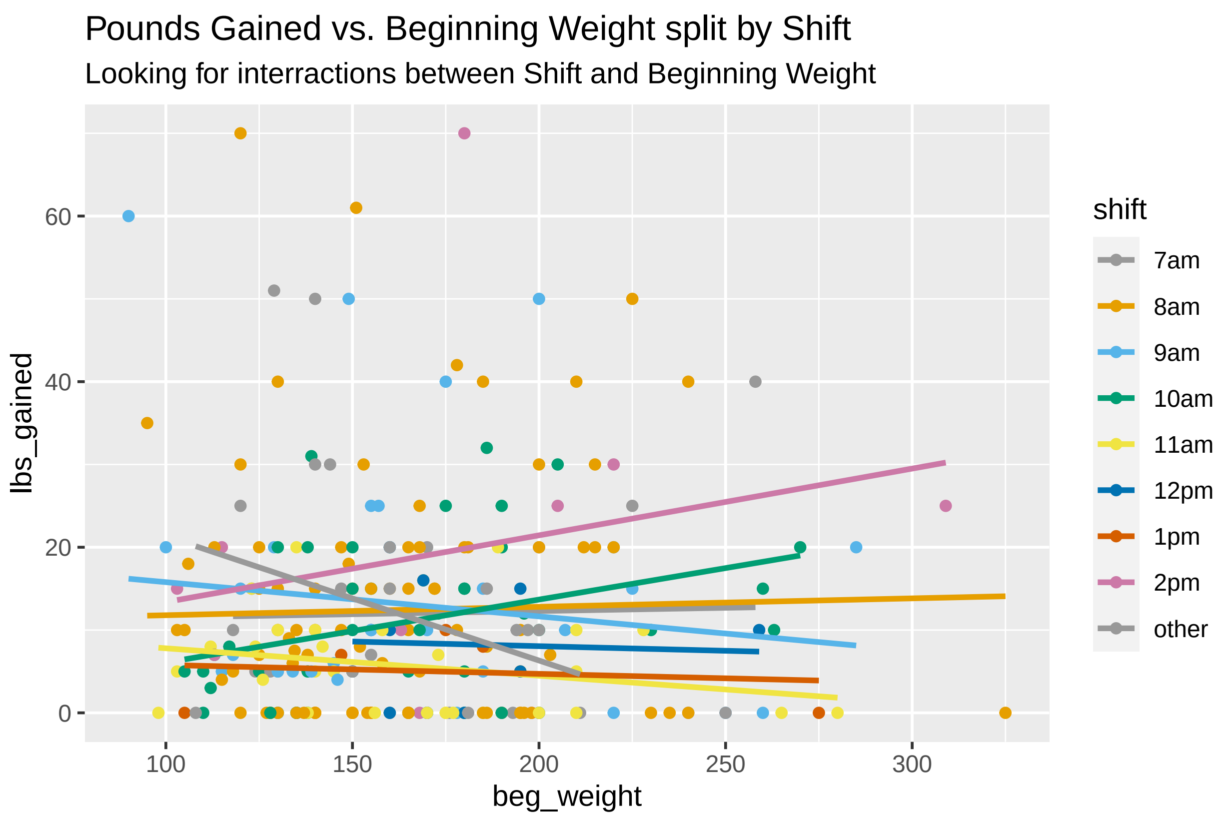




EXPLAIN WHAT WE'RE SEEING THOROUGHLY HERE

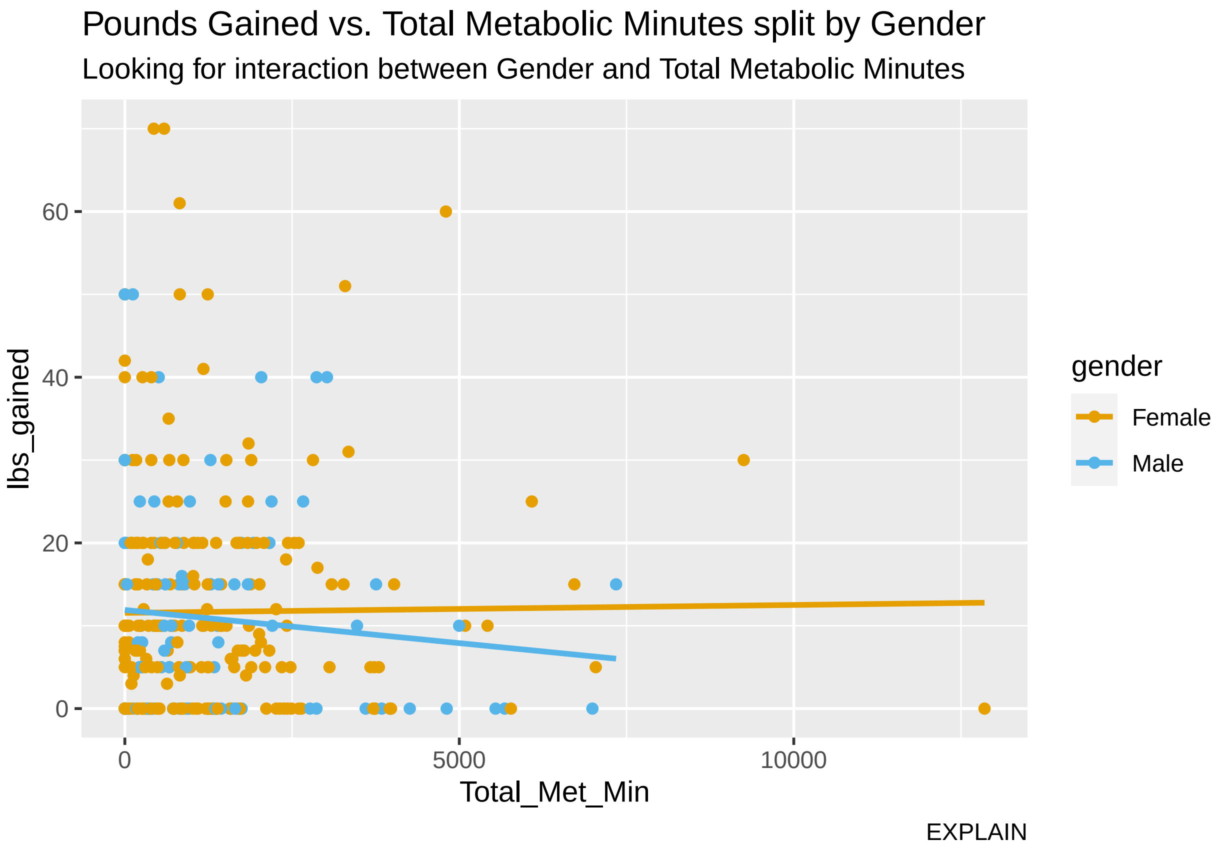


EXPLAIN THOROUGHLY HERE

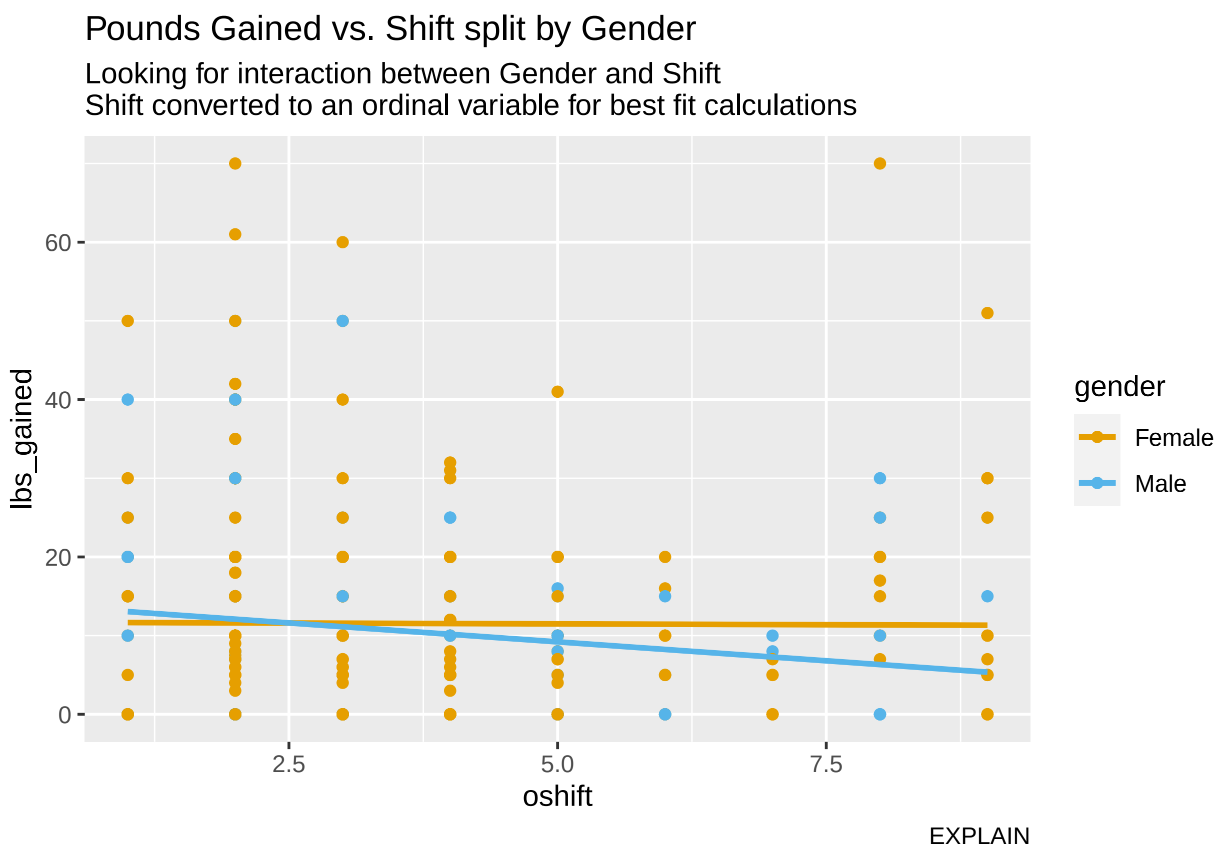


EXPLAIN THOROUGHLY HERE

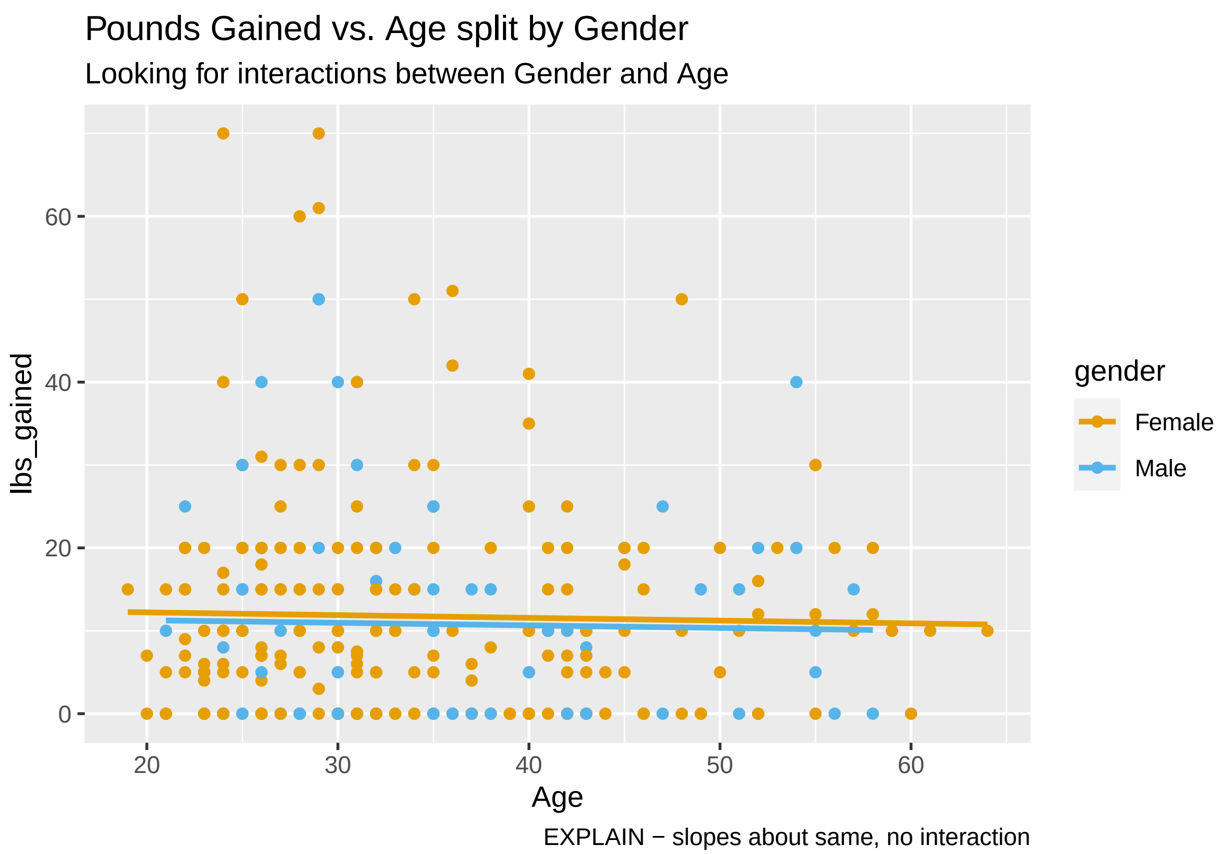
Interactions between Predictors and Gender

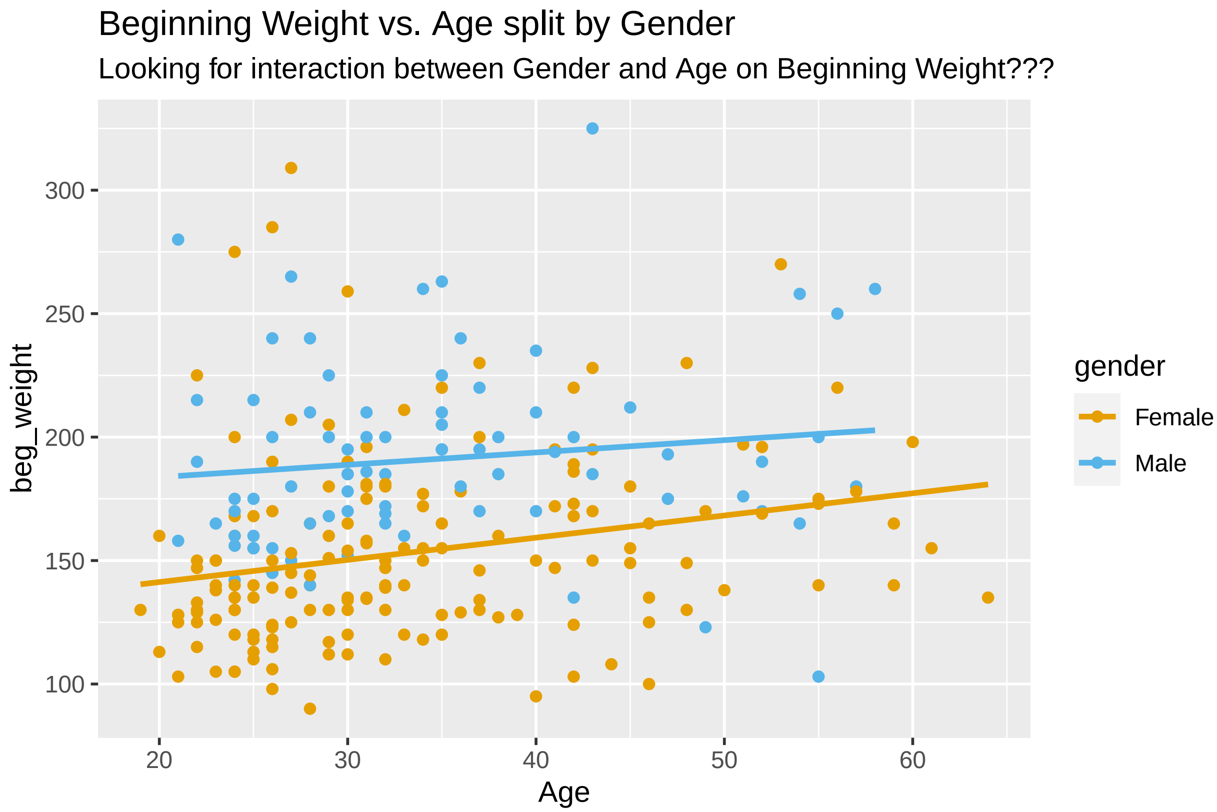


These plots



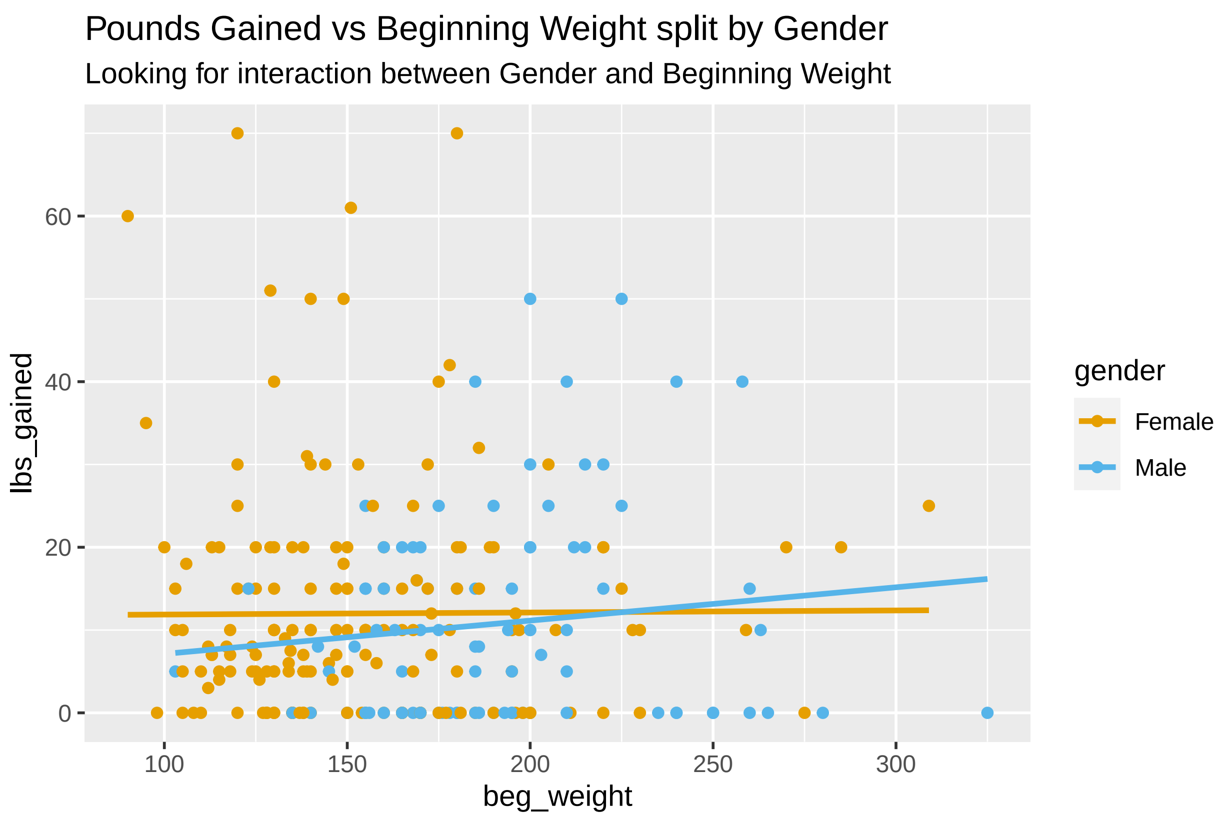
Pounds Gained vs. Shift split by Gender
Looking for interaction between Gender and Shift
Shift converted to an ordinal variable for best fit calculations
lbs_gained
60
40
20
0
2.5
5.0
7.5
oshift
gender
Female
Male
EXPLAIN


Pounds Gained vs. Age split by Gender
Looking for interactions between Gender and Age
lbs_gained
60
40
20
0
20
30
40
50
60
Age
gender
Female
Male
EXPLAIN − slopes about same, no interaction



Beginning Weight vs. Age split by Gender
Looking for interaction between Gender and Age on Beginning Weight???
beg_weight
300
250
200
150
100
20
30
40
50
60
Age
gender
Female
Male

EXPLAIN



EXPLAIN

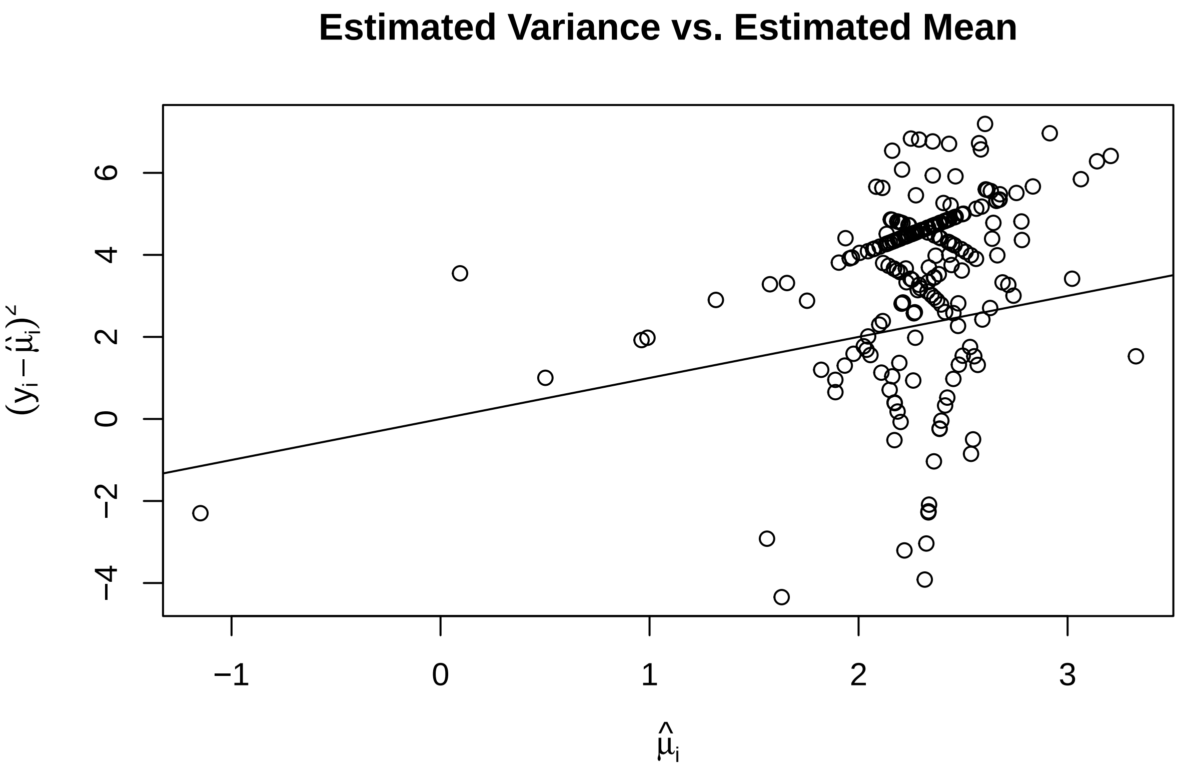
ZIP Model

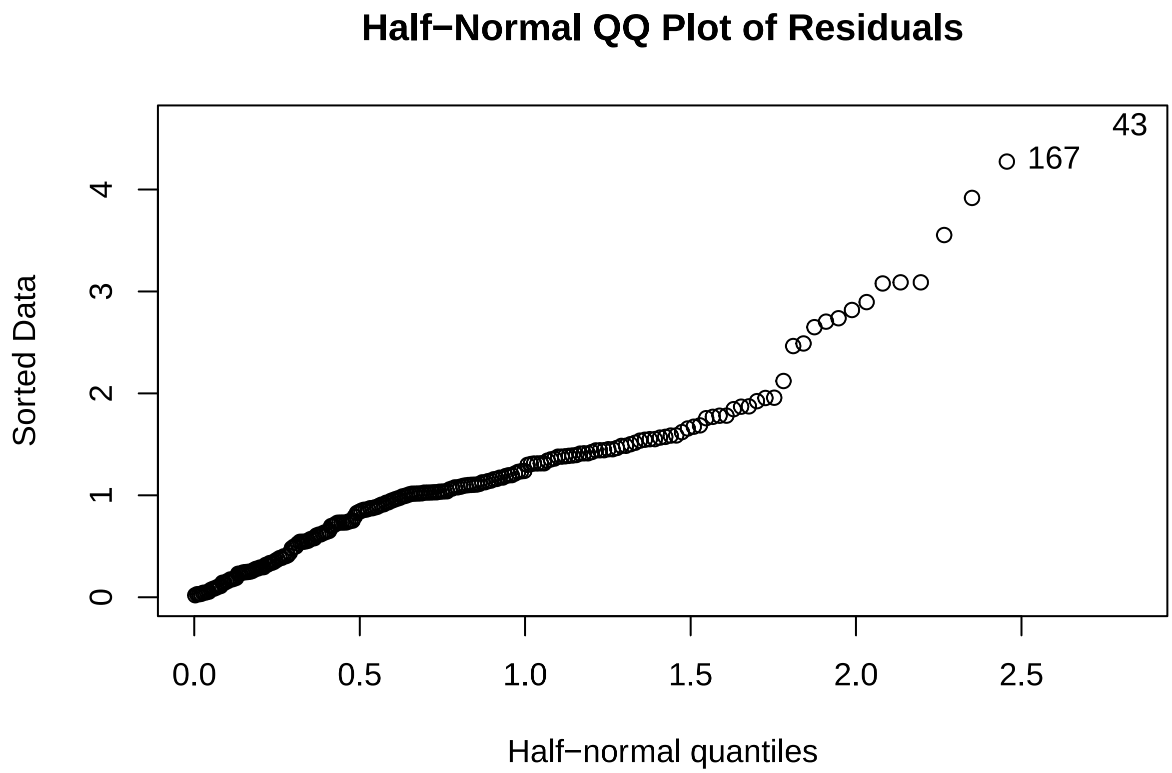
```
##
## Call:
## zeroinfl(formula = round(lbs_gained, 0) ~ oshift:log(Total_Met_Min +
##     1) + oshift:beg_weight + oshift * gender + log(Total_Met_Min + 1) *
##     gender + beg_weight * gender + Age, data = p1.Data2)
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -1.9574 -1.0669 -0.2508  0.8569  4.6394
##
## Count model coefficients (poisson with log link):
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                    3.1466721  0.3290802   9.562  < 2e-16 ***
## oshift                        -0.3560522  0.0716007  -4.973 6.60e-07 ***
## genderMale                    -0.4319834  0.3186407  -1.356  0.17519
## log(Total_Met_Min + 1)        -0.1445012  0.0245422  -5.888 3.91e-09 ***
## beg_weight                     0.0036250  0.0017298   2.096  0.03612 *
## Age                           -0.0009427  0.0022889  -0.412  0.68043
## oshift:log(Total_Met_Min + 1)  0.0549928  0.0075299   7.303 2.81e-13 ***
## oshift:beg_weight             -0.0001834  0.0002251  -0.815  0.41526
## oshift:genderMale             -0.1328542  0.0275504  -4.822 1.42e-06 ***
## log(Total_Met_Min + 1):genderMale  0.0145246  0.0208338   0.697  0.48570
## beg_weight:genderMale          0.0041399  0.0015286   2.708  0.00676 **
```

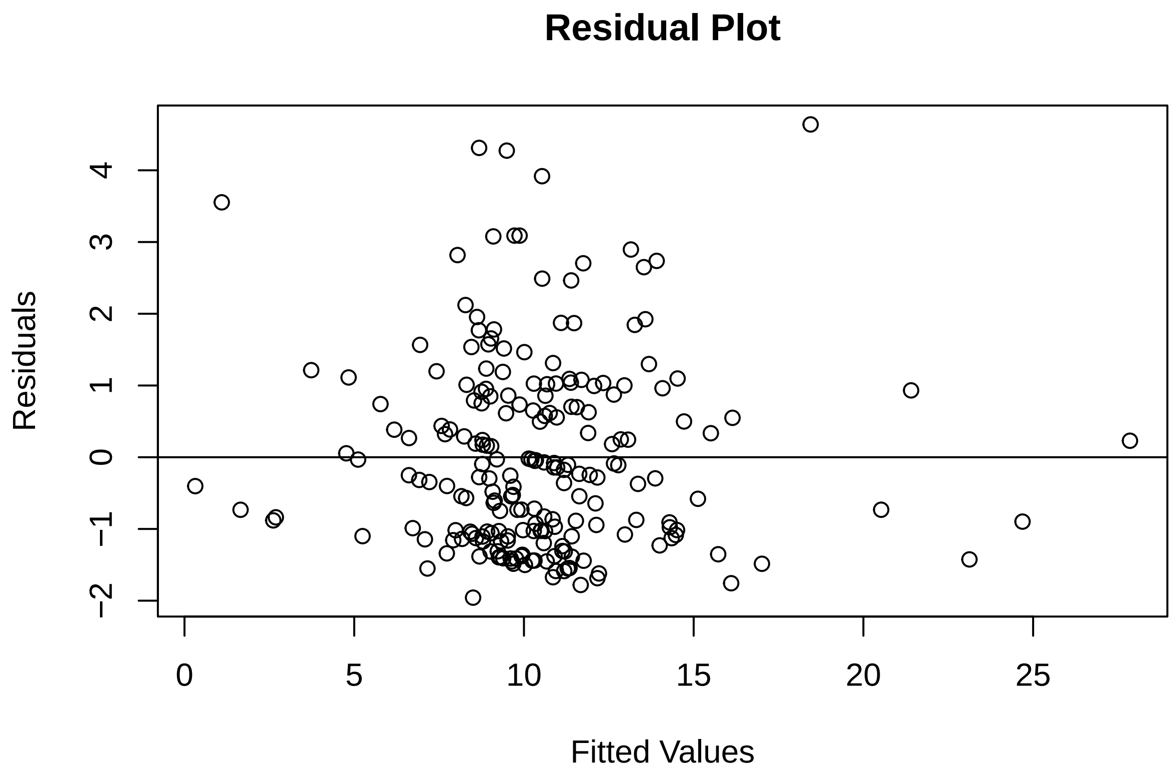
```
##
## Zero-inflation model coefficients (binomial with logit link):
##                                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)                     -0.269053   2.127681  -0.126    0.899
## oshift                           0.027947   0.487485   0.057    0.954
## genderMale                      -0.371729   1.767541  -0.210    0.833
## log(Total_Met_Min + 1)           0.056925   0.140277   0.406    0.685
## beg_weight                      -0.002888   0.013103  -0.220    0.826
## Age                             -0.011617   0.016914  -0.687    0.492
## oshift:log(Total_Met_Min + 1)   -0.048241   0.037619  -1.282    0.200
## oshift:beg_weight                0.001259   0.003086   0.408    0.683
## oshift:genderMale               -0.033052   0.196047  -0.169    0.866
## genderMale:log(Total_Met_Min + 1)  0.061117   0.130364   0.469    0.639
## genderMale:beg_weight            0.004120   0.008617   0.478    0.633
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 27
## Log-likelihood: -767.7 on 22 Df


##
## Call:
## hurdle(formula = round(lbs_gained, 0) ~ oshift:log(Total_Met_Min + 1) +
##     oshift:beg_weight + oshift * gender + log(Total_Met_Min + 1) * gender +
##     beg_weight * gender + Age, data = p1.Data2)
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -1.9486 -1.0660 -0.2483  0.8587  4.6690
##
## Count model coefficients (truncated poisson with log link):
##                                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)                       3.1459149  0.3294914   9.548  < 2e-16 ***
## oshift                           -0.3556610  0.0719069  -4.946 7.57e-07 ***
## genderMale                       -0.4303434  0.3184554  -1.351  0.17658
## log(Total_Met_Min + 1)           -0.1443403  0.0245756  -5.873 4.27e-09 ***
## beg_weight                        0.0036275  0.0017282   2.099  0.03582 *
## Age                              -0.0009647  0.0022889  -0.421  0.67341
## oshift:log(Total_Met_Min + 1)     0.0549448  0.0075531   7.274 3.48e-13 ***
## oshift:beg_weight                -0.0001837  0.0002251  -0.816  0.41429
## oshift:genderMale                -0.1328146  0.0275222  -4.826 1.39e-06 ***
## log(Total_Met_Min + 1):genderMale  0.0143453  0.0208002   0.690  0.49040
## beg_weight:genderMale             0.0041368  0.0015277   2.708  0.00677 **
## Zero hurdle model coefficients (binomial with logit link):
##                                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)                      0.324665   2.137422   0.152    0.879
## oshift                          -0.053886   0.492860  -0.109    0.913
## genderMale                       0.359873   1.765705   0.204    0.838
## log(Total_Met_Min + 1)          -0.067032   0.136962  -0.489    0.625
## beg_weight                       0.002935   0.012948   0.227    0.821
## Age                              0.011919   0.016925   0.704    0.481
## oshift:log(Total_Met_Min + 1)    0.052193   0.035440   1.473    0.141
## oshift:beg_weight               -0.001272   0.003040  -0.418    0.676
## oshift:genderMale                0.034638   0.194818   0.178    0.859
```

```
## genderMale:log(Total_Met_Min + 1) -0.059565   0.129975  -0.458    0.647
## genderMale:beg_weight              -0.004150   0.008589  -0.483    0.629
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 18
## Log-likelihood: -767.7 on 22 Df
```

Regression Results:




Half−Normal QQ Plot of Residuals
43
167
Sorted Data
0
1
2
3
4
0.0
0.5
1.0
1.5
2.0
2.5
Half−normal quantiles


Residual Plot
Residuals
Fitted Values
4
3
2
1
0
−1
−2
0
5
10
15
20
25