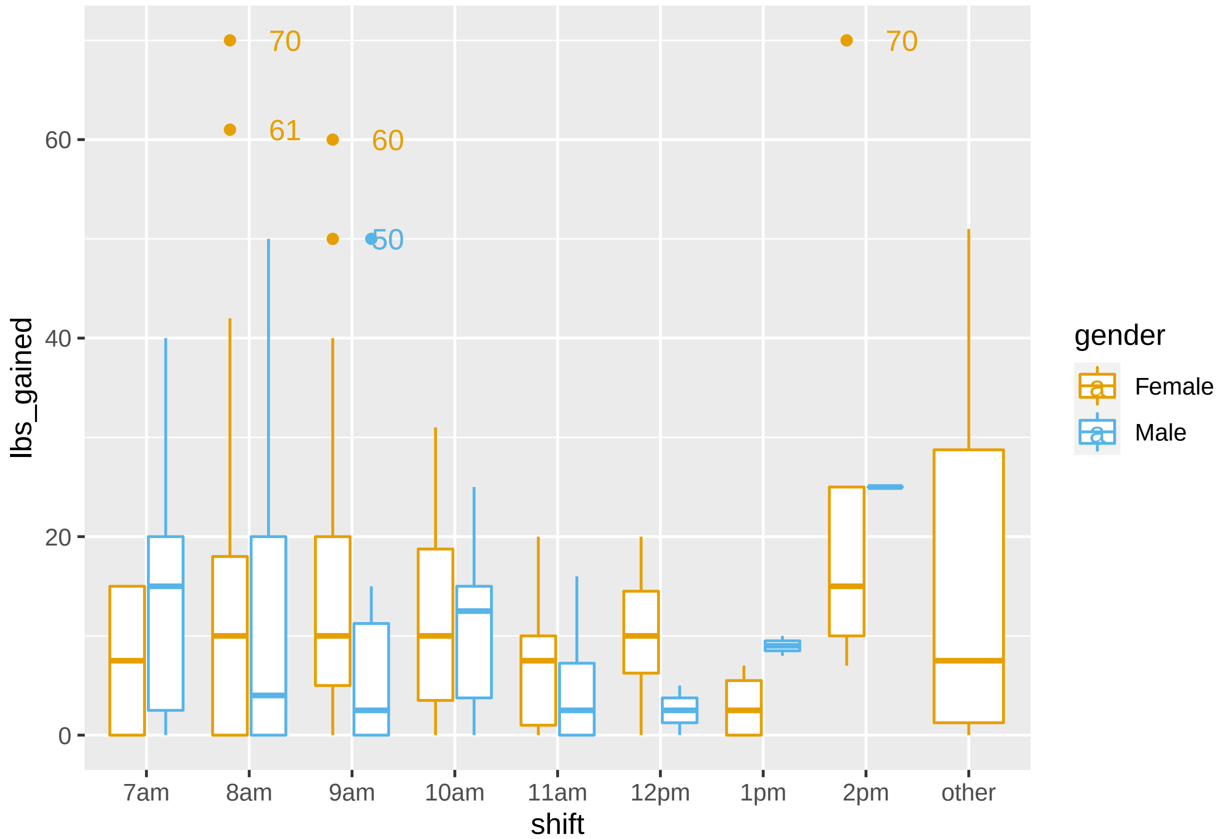
Effects of Work Shift and Total Excersize Time on Weight Gain

Paul Hunt, Brandon Kill, Brendan Winters

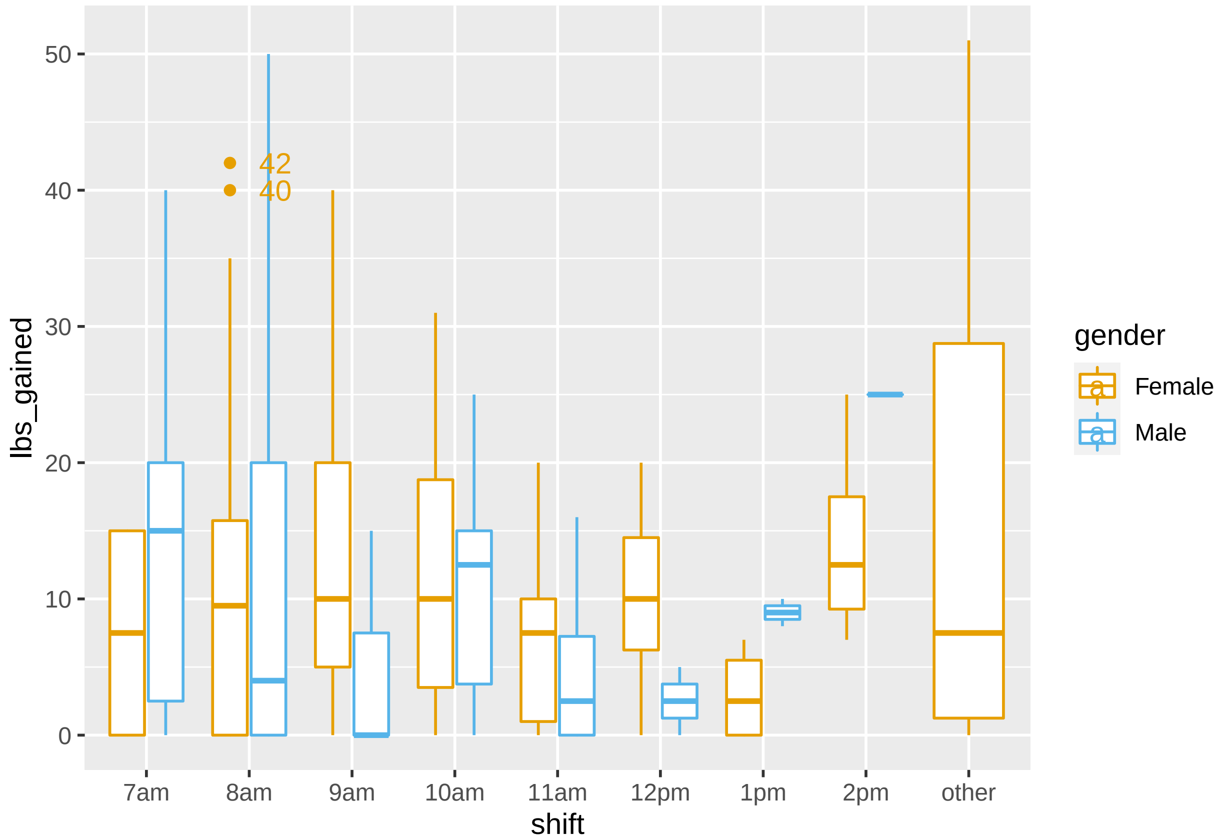
9/25/2021

Identify and remove outliers

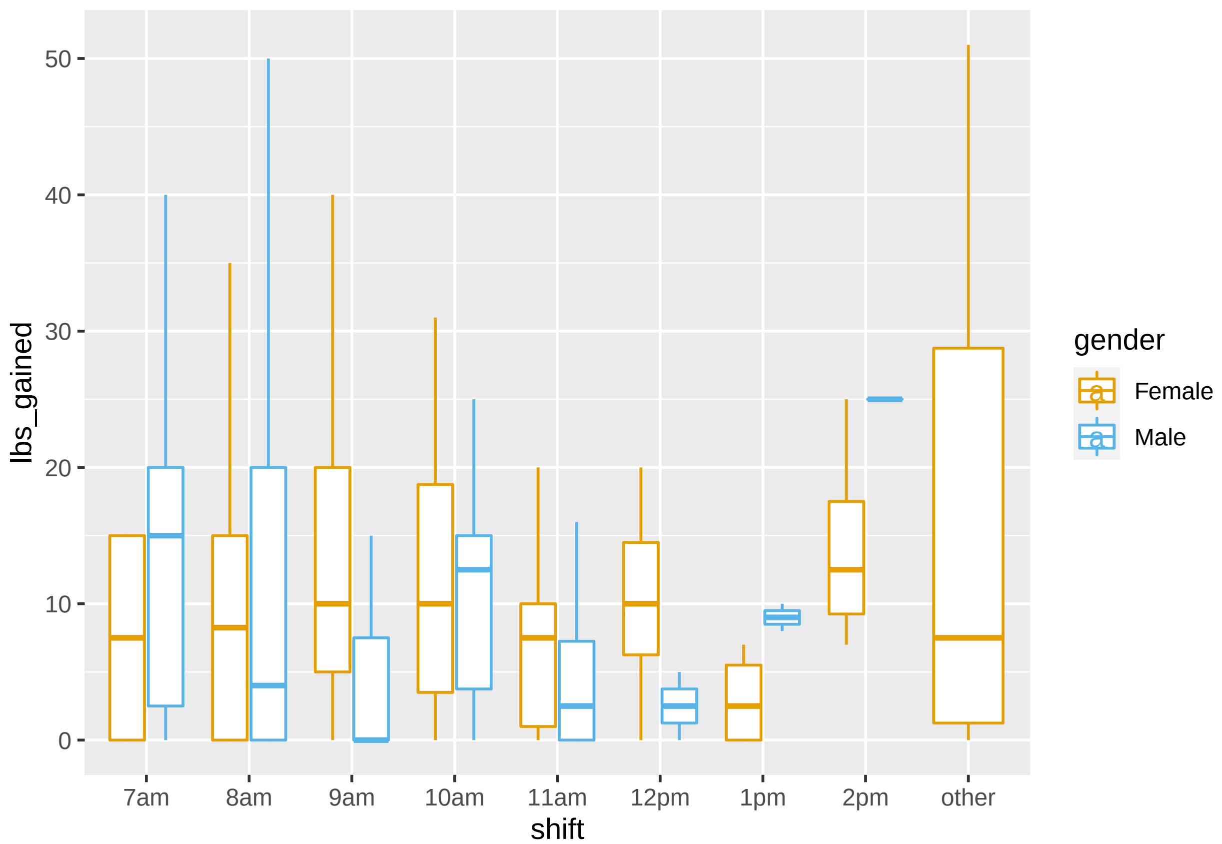
```
library(tidyverse)

# color blind pallet - we were told to use it in EDA
cb_palette = c("#999999", "#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00", "#CC79A7")

p1.Data1 <- na.omit(p1.Data)


is_outlier <- function(x) {
  return(x < quantile(x, 0.25) - 1.5 * IQR(x) | x > quantile(x, 0.75) + 1.5 * IQR(x))
}

p1.Data1 %>%
  group_by(shift, gender) %>%
  mutate(outlier = ifelse(is_outlier(lbs_gained),
                          lbs_gained,
                          as.numeric(NA))) %>%
  ggplot(., aes(x = shift, y = lbs_gained, color = gender)) +
  geom_boxplot() +
  geom_text(aes(label = outlier), na.rm = TRUE, hjust = -0.6) +
  scale_color_manual(values  =  cb_palette[2:3])
```




```
p1.Data.out <- p1.Data1 %>%
  group_by(shift) %>%
  mutate(outlier = ifelse(is_outlier(lbs_gained), "Yes", "No"))

p1.Data2 <- p1.Data.out %>%
  filter(outlier == "No")



p1.Data2 <- p1.Data2[,-ncol(p1.Data2)]

p1.Data2 %>%
  group_by(shift, gender) %>%
  mutate(outlier = ifelse(is_outlier(lbs_gained),
                          lbs_gained,
                          as.numeric(NA))) %>%
  ggplot(., aes(x = shift, y = lbs_gained, color = gender)) +
  geom_boxplot() +
  geom_text(aes(label = outlier, na.rm = TRUE, hjust = -0.3)) +
  scale_color_manual(values  =  cb_palette[2:3])
```




```
p1.Data.out2 <- p1.Data2 %>%
  group_by(shift, gender) %>%
  mutate(outlier = ifelse(is_outlier(lbs_gained), "Yes", "No"))

p1.Data3 <- p1.Data.out2 %>% filter(outlier == "No")

p1.Data3 %>%
  group_by(shift, gender) %>%
  mutate(outlier = ifelse(is_outlier(lbs_gained), lbs_gained, as.numeric(NA))) %>%
  ggplot(., aes(x = shift, y = lbs_gained, color = gender)) +
  geom_boxplot() +
  geom_text(aes(label = outlier), na.rm = TRUE, hjust = -0.3) +
  scale_color_manual(values  =  cb_palette[2:3])
```




```r
zipmod <- zeroinfl(round(lbs_gained,0)~oshift+I(Total_Met_Min/100)+
                     beg_weight*gender+Age*gender, data = p1.Data1)
zipmod2 <- zeroinfl(round(lbs_gained,0)~oshift+I(Total_Met_Min/100)+
                      beg_weight*gender+Age*gender, data = p1.Data2)
zipmod3 <- zeroinfl(round(lbs_gained,0)~oshift+I(Total_Met_Min/100)+
                      beg_weight*gender+Age*gender, data = p1.Data3)
summary(zipmod)
```

```
##
## Call:
## zeroinfl(formula = round(lbs_gained, 0) ~ oshift + I(Total_Met_Min/100) +
##     beg_weight * gender + Age * gender, data = p1.Data1)
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -2.0006 -1.0965 -0.3502  0.8093  8.2128
##
## Count model coefficients (poisson with log link):
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)            2.7354082  0.1151809  23.749  < 2e-16 ***
## oshift                -0.0249417  0.0098754  -2.526  0.01155 *
## I(Total_Met_Min/100)   0.0032972  0.0012573   2.622  0.00873 **
## beg_weight             0.0009657  0.0006057   1.594  0.11083
## genderMale            -1.2553529  0.2656986  -4.725 2.30e-06 ***
## Age                   -0.0005247  0.0024812  -0.211  0.83252
```

```
## beg_weight:genderMale  0.0080407  0.0012410   6.479 9.21e-11 ***
## genderMale:Age        -0.0075550  0.0047364  -1.595  0.11070
##
## Zero-inflation model coefficients (binomial with logit link):
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -1.0076986  0.9414633  -1.070    0.284
## oshift                -0.0726360  0.0771848  -0.941    0.347
## I(Total_Met_Min/100)   0.0082111  0.0081909   1.002    0.316
## beg_weight             0.0037651  0.0048715   0.773    0.440
## genderMale             0.1591692  1.6564239   0.096    0.923
## Age                   -0.0134470  0.0208867  -0.644    0.520
## beg_weight:genderMale  0.0004044  0.0080646   0.050    0.960
## genderMale:Age         0.0085492  0.0331502   0.258    0.796
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 1
## Log-likelihood: -1057 on 16 Df
```

```
summary(zipmod2)
```

```
##
## Call:
## zeroinfl(formula = round(lbs_gained, 0) ~ oshift + I(Total_Met_Min/100) +
##     beg_weight * gender + Age * gender, data = p1.Data2)
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -1.8770 -1.0835 -0.2884  0.8226  6.6653
##
## Count model coefficients (poisson with log link):
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)            2.2946377  0.1249330  18.367  < 2e-16 ***
## oshift                -0.0298849  0.0107025  -2.792 0.005233 **
## I(Total_Met_Min/100)   0.0046750  0.0013339   3.505 0.000457 ***
## beg_weight             0.0022252  0.0006386   3.484 0.000493 ***
## genderMale            -0.9297303  0.2747511  -3.384 0.000715 ***
## Age                    0.0024574  0.0026600   0.924 0.355577
## beg_weight:genderMale  0.0067029  0.0012765   5.251 1.51e-07 ***
## genderMale:Age        -0.0079964  0.0048759  -1.640 0.101011
##
## Zero-inflation model coefficients (binomial with logit link):
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -0.850833   0.948842  -0.897    0.370
## oshift                -0.073298   0.077416  -0.947    0.344
## I(Total_Met_Min/100)   0.007763   0.008253   0.941    0.347
## beg_weight             0.003168   0.004902   0.646    0.518
## genderMale             0.053574   1.658576   0.032    0.974
## Age                   -0.013678   0.020949  -0.653    0.514
## beg_weight:genderMale  0.001187   0.008085   0.147    0.883
## genderMale:Age         0.007174   0.033240   0.216    0.829
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Number of iterations in BFGS optimization: 1
## Log-likelihood: -814.1 on 16 Df
```

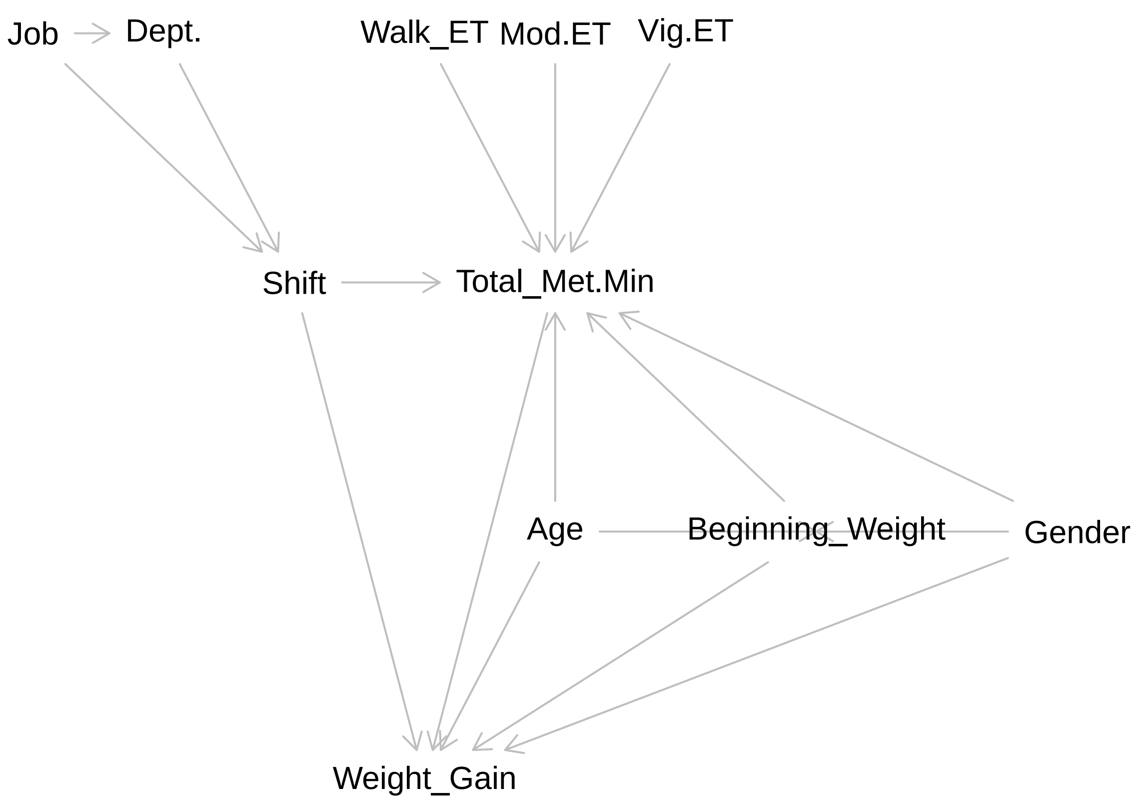
```
summary(zipmod3)
```

```
##
## Call:
## zeroinfl(formula = round(lbs_gained, 0) ~ oshift + I(Total_Met_Min/100) +
##     beg_weight * gender + Age * gender, data = p1.Data3)
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -1.8937 -1.0858 -0.2679  0.8250  6.4087
##
## Count model coefficients (poisson with log link):
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)            2.1532575  0.1283238  16.780  < 2e-16 ***
## oshift                -0.0205723  0.0107424  -1.915  0.05549 .
## I(Total_Met_Min/100)   0.0058220  0.0013267   4.388 1.14e-05 ***
## beg_weight             0.0021868  0.0006516   3.356  0.00079 ***
## genderMale            -0.8750168  0.2758314  -3.172  0.00151 **
## Age                    0.0041885  0.0027054   1.548  0.12157
## beg_weight:genderMale  0.0068701  0.0012823   5.358 8.43e-08 ***
## genderMale:Age        -0.0092460  0.0049043  -1.885  0.05939 .
##
## Zero-inflation model coefficients (binomial with logit link):
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -0.780301   0.951887  -0.820    0.412
## oshift                -0.078759   0.077655  -1.014    0.310
## I(Total_Met_Min/100)   0.007259   0.008269   0.878    0.380
## beg_weight             0.003199   0.004894   0.654    0.513
## genderMale             0.019309   1.659209   0.012    0.991
## Age                   -0.014524   0.020955  -0.693    0.488
## beg_weight:genderMale  0.001163   0.008080   0.144    0.886
## genderMale:Age         0.007642   0.033238   0.230    0.818
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 1
## Log-likelihood:  -773 on 16 Df
```

Causal Inference

```
CI.graph <- dagitty('dag {
Job                [pos="0,1"]
Dept.              [pos="1,1"]
Walk_ET            [pos="3,1"]
Mod.ET             [pos="4,1"]
Vig.ET             [pos="5,1"]
Shift              [pos="2,2"]
Total_Met.Min      [pos="4,2"]
```

```
Age              [pos="4,3"]
Gender           [pos="8,3"]
Beginning_Weight[pos="6,3"]
Weight_Gain      [pos="3,4"]

Job -> Dept. -> Shift
Job -> Shift -> Total_Met.Min
Shift -> Weight_Gain
Walk_ET -> Total_Met.Min
Mod.ET -> Total_Met.Min
Vig.ET -> Total_Met.Min -> Weight_Gain
Age -> Total_Met.Min
Age -> Weight_Gain
Age -> Beginning_Weight
Gender -> Total_Met.Min
Gender -> Weight_Gain
Gender -> Beginning_Weight
Beginning_Weight -> Weight_Gain
Beginning_Weight -> Total_Met.Min
}')

plot(CI.graph)
```



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

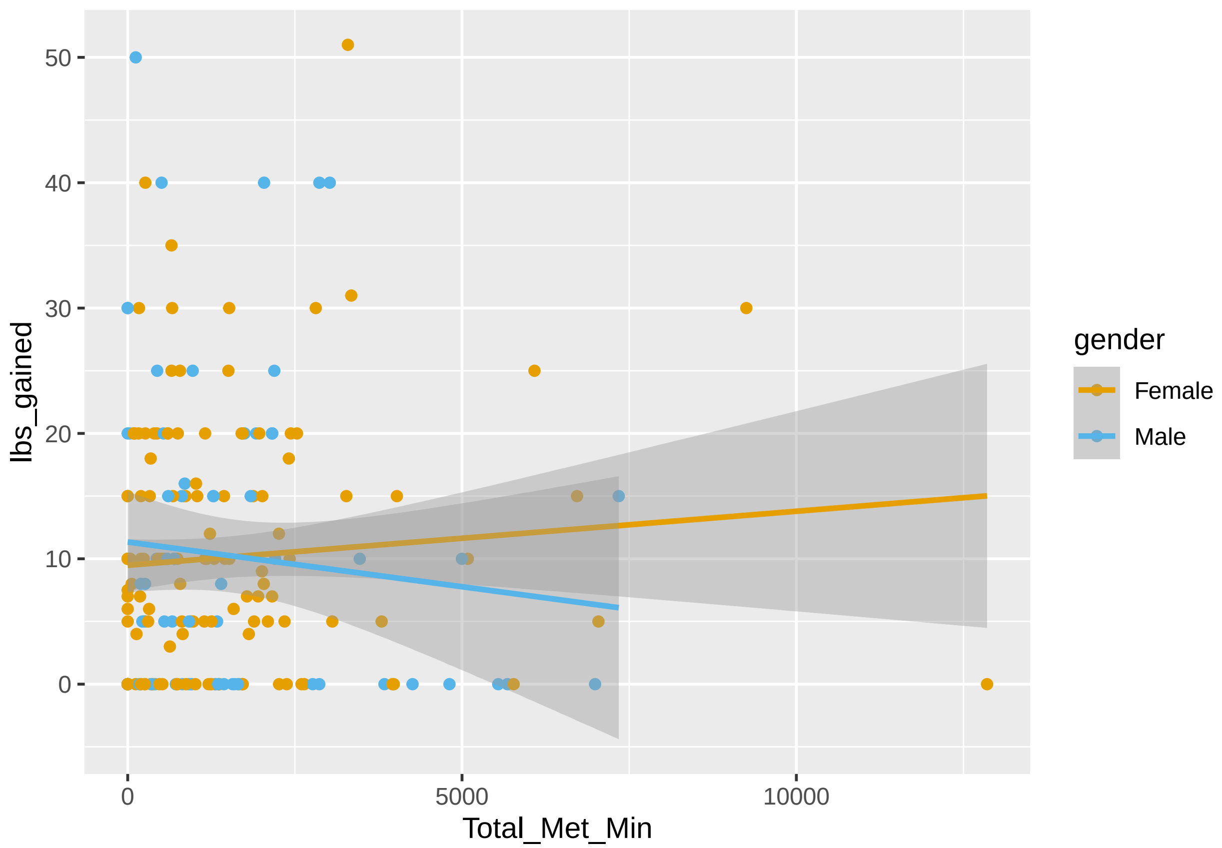
- **Total Metabolic Minutes: This variable represents total exercise time for respondents in an average week, weighted by the intensity of the exercise. We expect exercise to have a direct effect on weigh gain, most likely with an inverse relationship.**
- **Shift: This variable indicates the shift in 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.**

Plotting the Predictors on Weight Gain

These plots

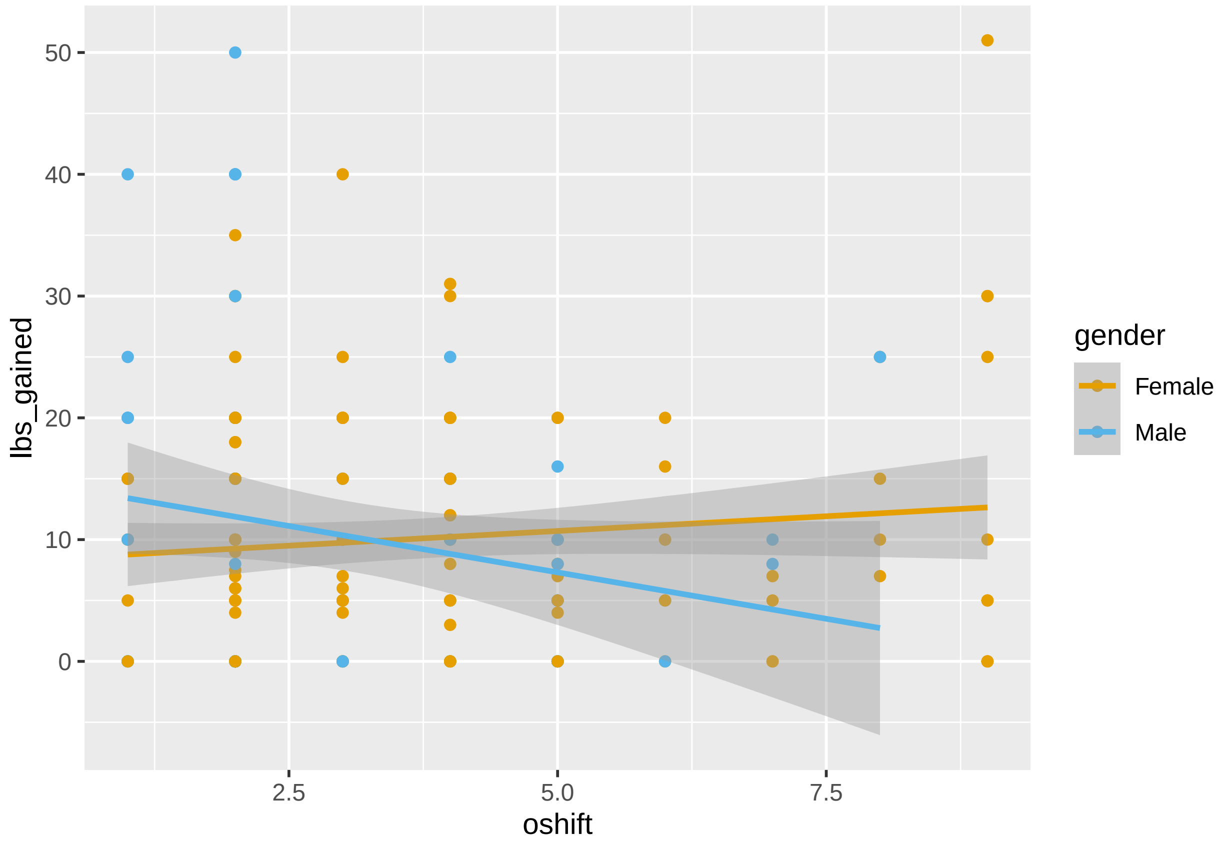
```
ggplot(p1.Data3, aes(x=Total_Met_Min, y=lbs_gained, color=gender)) +geom_point()+
  geom_smooth(method="lm") +
  scale_color_manual(values  =  cb_palette[2:3])
```

```
## `geom_smooth()` using formula 'y ~ x'
```



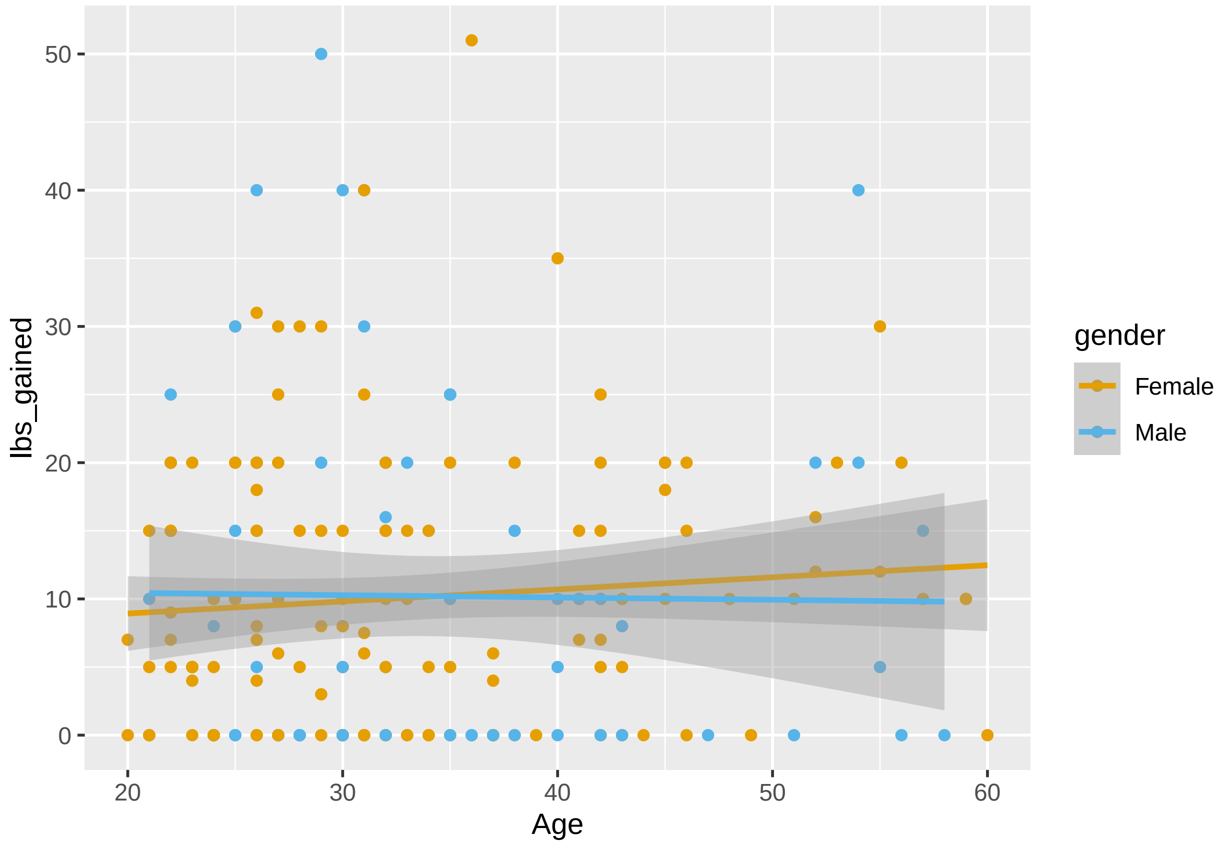

```
ggplot(p1.Data3, aes(x=oshift, y=lbs_gained, color=gender)) +geom_point()+
  geom_smooth(method="lm") +
  scale_color_manual(values  =  cb_palette[2:3])
```

```
## `geom_smooth()` using formula 'y ~ x'
```



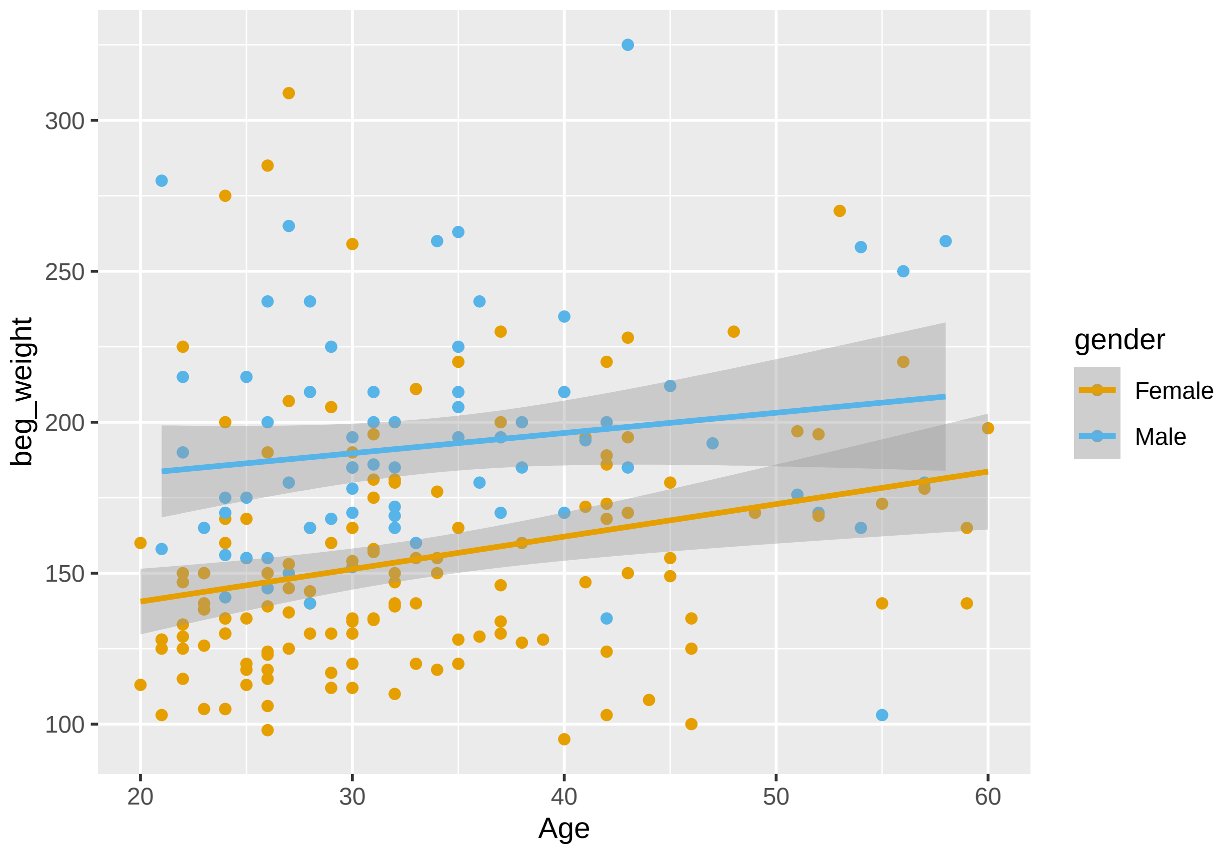

```
ggplot(p1.Data3, aes(x=Age, y=lbs_gained, color=gender)) +geom_point()+
  geom_smooth(method="lm") +
  scale_color_manual(values  =  cb_palette[2:3])
```

```
## `geom_smooth()` using formula 'y ~ x'
```



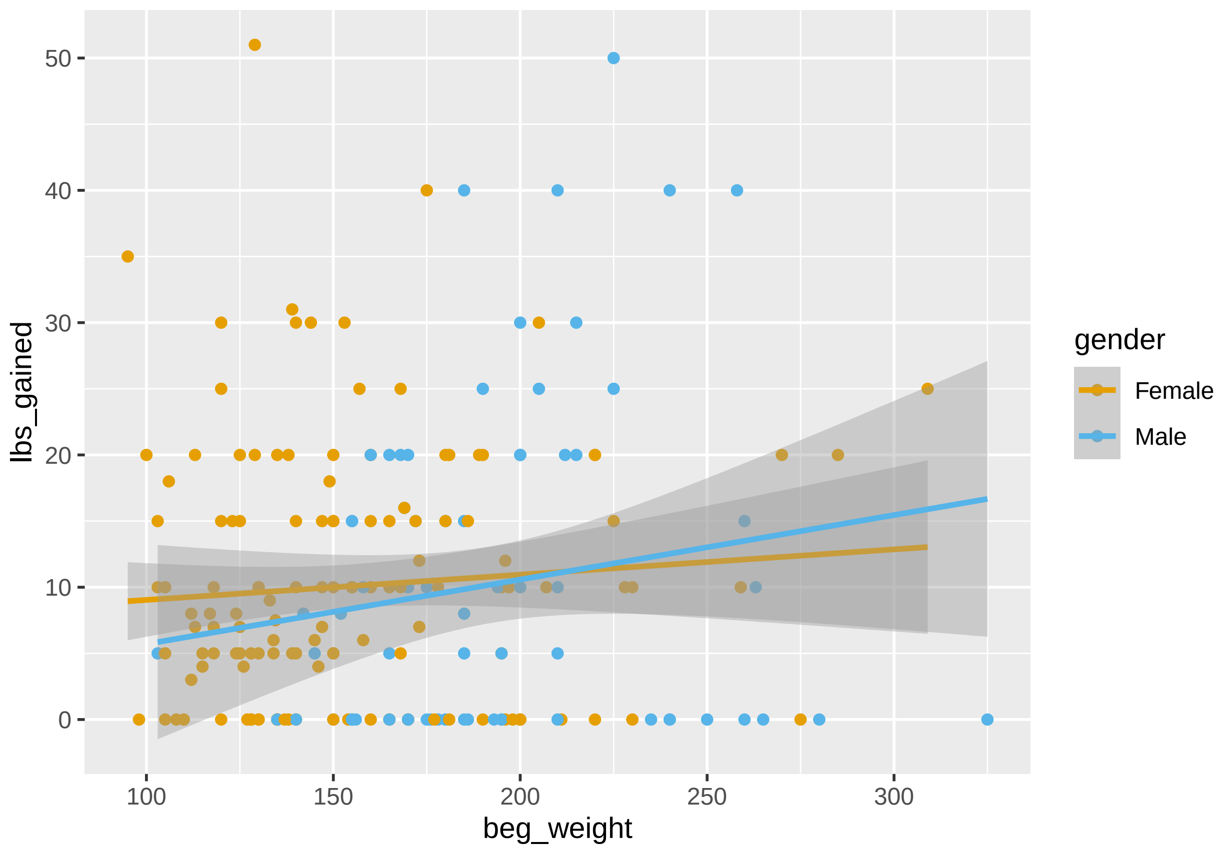

```
ggplot(p1.Data3, aes(x=Age, y=beg_weight, color=gender)) +geom_point()+
  geom_smooth(method="lm") +
  scale_color_manual(values  =  cb_palette[2:3])
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
ggplot(p1.Data3, aes(x=beg_weight, y=lbs_gained, color=gender)) +geom_point()+
  geom_smooth(method="lm") +
  scale_color_manual(values  =  cb_palette[2:3])
```

```
## `geom_smooth()` using formula 'y ~ x'
```



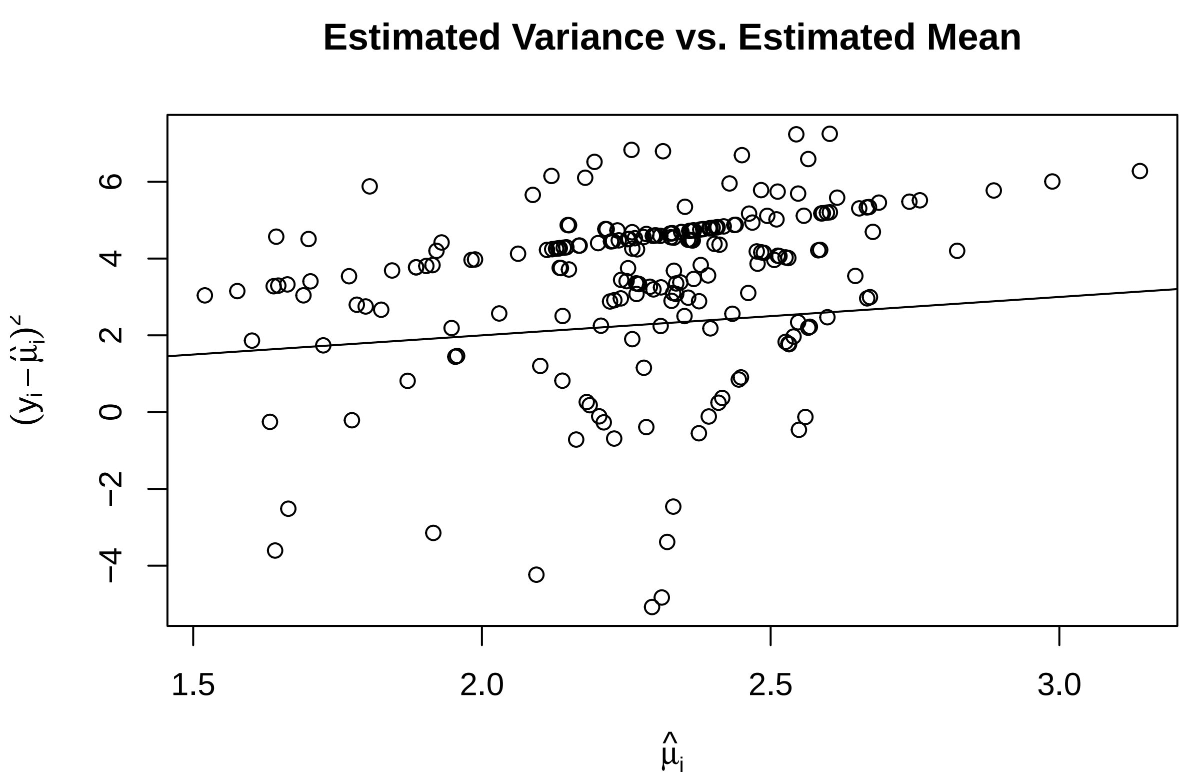
ZIP Model

```
zipmod <- zeroinfl(round(lbs_gained,0)~oshift*gender+log(Total_Met_Min+1)*gender+beg_weight*gender+Age,
summary(zipmod)
```

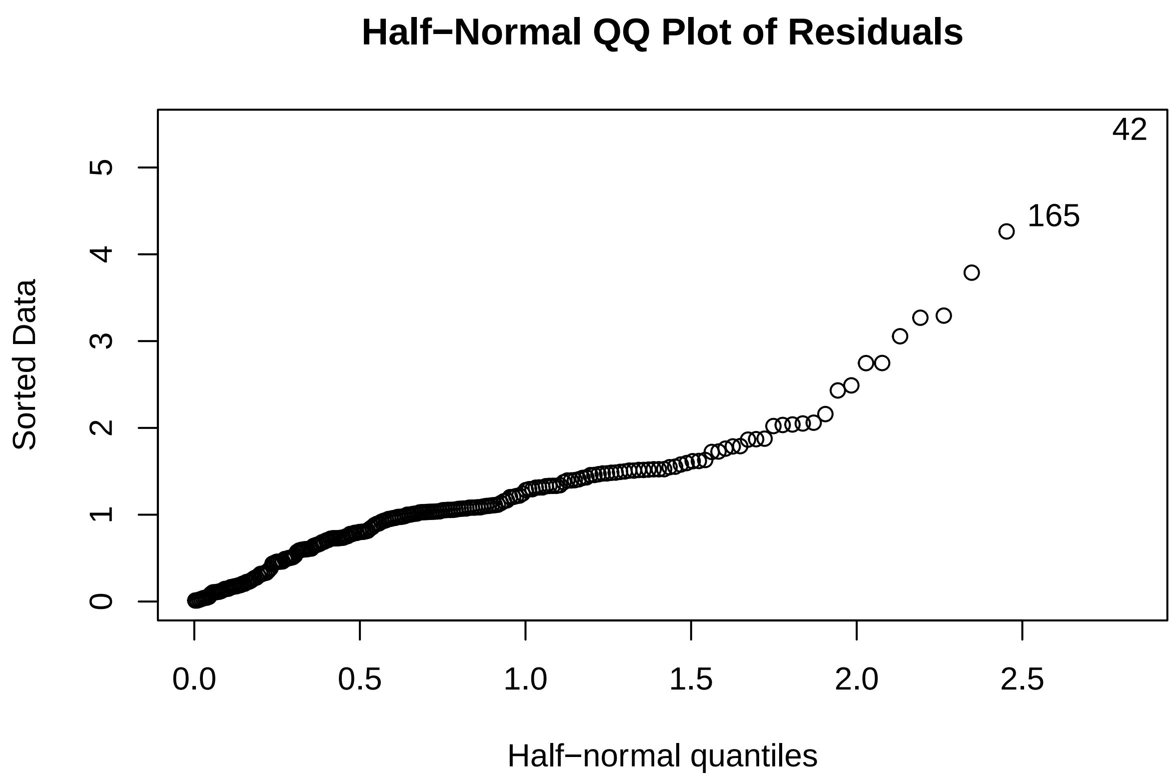
```
##
## Call:
## zeroinfl(formula = round(lbs_gained, 0) ~ oshift * gender + log(Total_Met_Min +
##     1) * gender + beg_weight * gender + Age, data = p1.Data3)
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -1.8760 -1.0553 -0.2049  0.7903  5.4479
##
## Count model coefficients (poisson with log link):
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   1.8428070  0.1564563  11.778  < 2e-16 ***
## oshift                        0.0196194  0.0120182   1.632 0.102579
## genderMale                    0.0929988  0.3036942   0.306 0.759433
## log(Total_Met_Min + 1)        0.0536279  0.0137059   3.913 9.12e-05 ***
## beg_weight                    0.0024341  0.0006264   3.886 0.000102 ***
## Age                           0.0001468  0.0022502   0.065 0.947967
## oshift:genderMale            -0.1626671  0.0261504  -6.220 4.96e-10 ***
```

```
## genderMale:log(Total_Met_Min + 1) -0.0678243  0.0211855  -3.201 0.001367 **
## genderMale:beg_weight              0.0050286  0.0012804   3.927 8.59e-05 ***
##
## Zero-inflation model coefficients (binomial with logit link):
##                                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)                        0.112529   1.020149   0.110   0.9122
## oshift                            -0.067222   0.092951  -0.723   0.4696
## genderMale                        -0.496499   1.745203  -0.284   0.7760
## log(Total_Met_Min + 1)            -0.117648   0.071511  -1.645   0.0999 .
## beg_weight                         0.002389   0.004881   0.489   0.6245
## Age                               -0.015117   0.016680  -0.906   0.3648
## oshift:genderMale                 -0.019459   0.171806  -0.113   0.9098
## genderMale:log(Total_Met_Min + 1)  0.110090   0.126446   0.871   0.3839
## genderMale:beg_weight              0.002438   0.007919   0.308   0.7582
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 1
## Log-likelihood: -750.6 on 18 Df
```

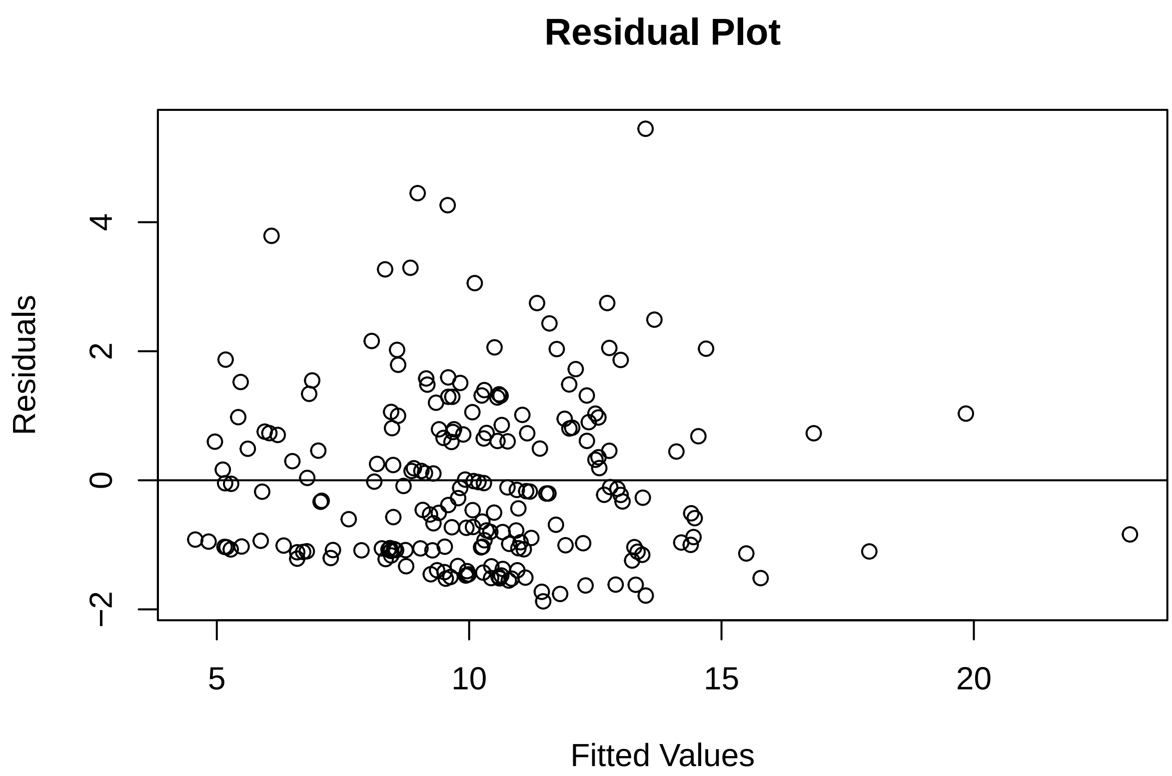
```r
plot(log(fitted(zipmod)), log((p1.Data3$lbs_gained-fitted(zipmod))^2),
     xlab=expression(hat(mu)[i]),
     ylab=expression((y[i]-hat(mu)[i])^2),
     main = "Estimated Variance vs. Estimated Mean")
abline(0,1)
```



```r
halfnorm(residuals(zipmod), main = "Half-Normal QQ Plot of Residuals")
```




```r
plot(fitted(zipmod), resid(zipmod),
     ylab="Residuals",
     xlab="Fitted Values",
     main = "Residual Plot")
abline(0,0)
```


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