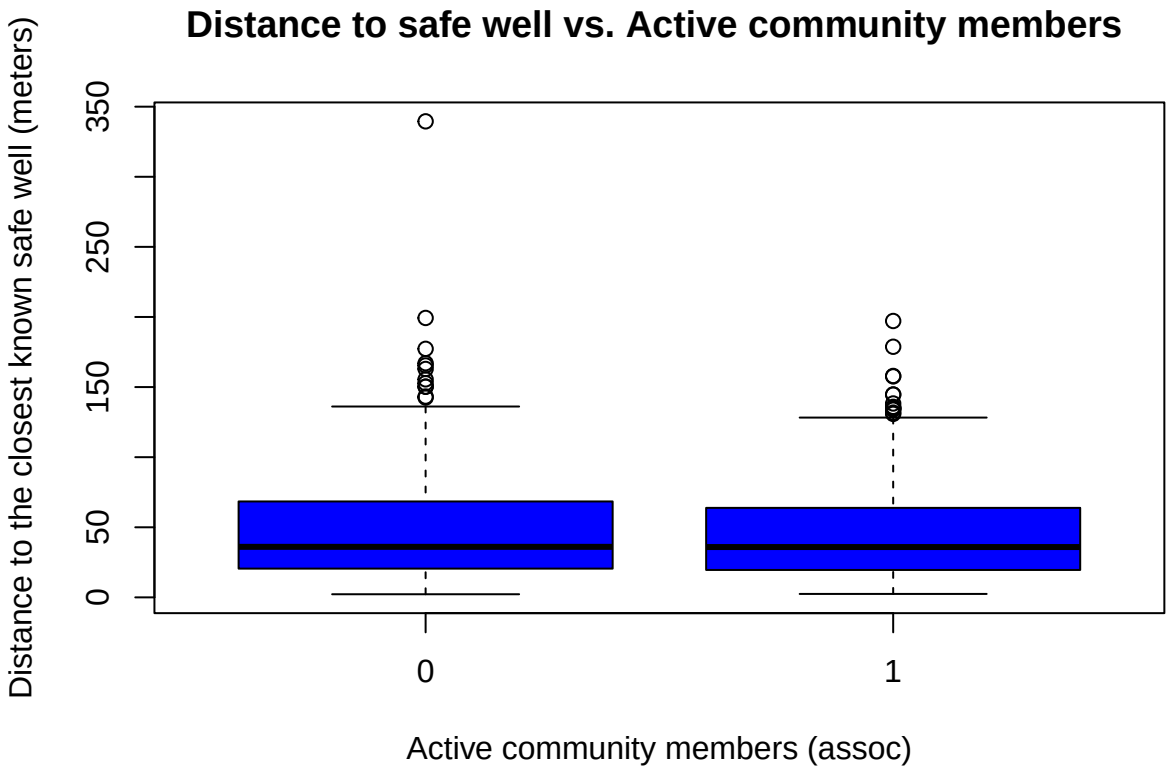
# Assignment 1

```
ars_data <- read.table("arsenic-2.txt", header = TRUE)
head(ars_data)
```

```
##      switch arsenic    dist assoc educ
## 2480      1    2.16 114.147     1   12
## 2600      1    1.30  20.387     0   16
## 2695      0    0.95  54.894     1    4
## 2367      1    0.81  24.004     1   10
## 2912      0    4.39  60.708     1    6
## 2018      0    0.74  48.806     0    5
```

a) Producing plots

```
#plot(ars_data$assoc, ars_data$dist, ylab = "Dist (in meters) to safest well", xlab = "Active community
boxplot( dist~assoc,
         data=ars_data,
         main="Distance to safe well vs. Active community members ",
         xlab="Active community members (assoc)",
         ylab="Distance to the closest known safe well (meters)",
         col="blue",
         border="black"
       )
```




We notice when fitting distance to the closest known safe well and whether members of the household are active community members, that there isn't much difference. Citizens who aren't active community members travel more of a distance compared to those who are but not by a significant margin. Both groups in general seem to travel around the same distance indicated by the median. The only concern is the outlier existing in the group of citizens that aren't active community members. This individual traveled almost 350 meters.

b)

Logistic regression model with all additive predictors of arsenic, distance traveled in meters, whether members of the household were an active community member and education.

```
ars_predicts <- glm(formula = switch ~ arsenic + dist + assoc + educ ,
                    family = binomial(link = "logit"), data = ars_data)
ars_predicts
```

```
##
## Call:  glm(formula = switch ~ arsenic + dist + assoc + educ, family = binomial(link = "logit"),
##     data = ars_data)
##
## Coefficients:
## (Intercept)      arsenic         dist        assoc         educ
##    -0.27523      0.54627     -0.01032     -0.13054      0.05895
##
## Degrees of Freedom: 499 Total (i.e. Null);  495 Residual
## Null Deviance:       680.9
## Residual Deviance: 636   AIC: 646
```

Logistic regression model with all additive predictors and an interaction between distance traveled in meters and whether members of the house hold was an active community member.

```
ars_interact <- glm(formula = switch ~ arsenic + dist + assoc + educ
                    + dist*assoc,
                    family = binomial(link = "logit"), data = ars_data)
ars_interact
```

```
##
## Call:  glm(formula = switch ~ arsenic + dist + assoc + educ + dist *
##     assoc, family = binomial(link = "logit"), data = ars_data)
##
## Coefficients:
## (Intercept)      arsenic         dist        assoc         educ   dist:assoc
##   -0.413015     0.548363    -0.007689     0.172078     0.060017    -0.006367
##
## Degrees of Freedom: 499 Total (i.e. Null);  494 Residual
## Null Deviance:       680.9
## Residual Deviance: 634.5     AIC: 646.5
```

We computer the brier score for the original data for both models

```
predicted_predicts.model <- predict(ars_predicts, ars_data, type = "response")
brierscore_predicts.model <- (mean((predicted_predicts.model - ars_predicts$residuals)^2))/nrow(ars_data
brierscore_predicts.model
```

```
## [1] 0.01072395
```

```
predicted_interact.model <- predict(ars_interact, ars_data, type = "response")
brierscore_interact.model <- (mean((predicted_interact.model - ars_interact$residuals)^2))/nrow(ars_data
brierscore_interact.model
```

```
## [1] 0.01073027
```

The model with all additive predictors no interactions has the better Brier score but not by that much. The additive predictors model has a lower Brier score of 0.01072395 in comparison to 0.01073027 of the interaction model of distance and active community member

#Question-C.

We generate a random sample that has the same number of rows from our orginal data which is about 500. which essientially gives us the index.train data from the validate function.

```
set.seed(1002625448)
random_sample <- sample(x=1:nrow(ars_data), size = nrow(ars_data), replace = TRUE)
arsenic_sample <- ars_data[random_sample,]
```

We will now refit a model with all additive predictors and a model with interactions agian using this sample.

```
ars_random.predicts <- glm(formula = switch ~ arsenic + dist + assoc + educ ,
                    family = binomial(link = "logit"), data = arsenic_sample)
ars_random.predicts
```

```
##
## Call:  glm(formula = switch ~ arsenic + dist + assoc + educ, family = binomial(link = "logit"),
##     data = arsenic_sample)
##
## Coefficients:
## (Intercept)      arsenic         dist        assoc         educ
##    -0.33090      0.56358     -0.01110     -0.08272      0.05715
##
## Degrees of Freedom: 499 Total (i.e. Null);  495 Residual
## Null Deviance:       683.9
## Residual Deviance: 634.2     AIC: 644.2
```

Same thing fro the interaction model

```
ars_random.interact <- glm(formula = switch ~ arsenic + dist + assoc + educ
                    + dist*assoc,
                    family = binomial(link = "logit"), data = arsenic_sample)
ars_random.interact
```

```
##
## Call:  glm(formula = switch ~ arsenic + dist + assoc + educ + dist *
##     assoc, family = binomial(link = "logit"), data = arsenic_sample)
##
## Coefficients:
## (Intercept)      arsenic         dist        assoc         educ   dist:assoc
##    -0.59299      0.56747     -0.00651      0.43687      0.06256     -0.01040
##
## Degrees of Freedom: 499 Total (i.e. Null);  494 Residual
## Null Deviance:       683.9
## Residual Deviance: 630   AIC: 642
```

Brier Score computations

```
predicted_predicts.ran <- predict(ars_random.predicts, arsenic_sample, type = "response")
brierscore_predicts.new.model <- (mean((predicted_predicts.ran - ars_random.predicts$residuals)^2))/nrow
brierscore_predicts.new.model
```

```
## [1] 0.01060896
```

```
predicted_interact.ran <- predict(ars_random.interact, arsenic_sample, type = "response")
brierscore_interact.new.model <- (mean((predicted_interact.ran - ars_random.interact$residuals)^2))/nrow
brierscore_interact.new.model
```

```
## [1] 0.01069681
```

#Question-D. We now run the same computations of this sample data on the original data, which essientially gives us the index.text data from the validate function.

```
predicted_predicts.ran2 <- predict(ars_random.predicts, ars_data, type = "response")
brierscore_predicts.new.model2 <- (mean((predicted_predicts.ran2 - ars_random.predicts$residuals)^2))/n
brierscore_predicts.new.model2
```

```
## [1] 0.0105905
```

```
predicted_interact.ran2 <- predict(ars_random.interact, ars_data, type = "response")
brierscore_interact.new.model2 <- (mean((predicted_interact.ran2 - ars_random.interact$residuals)^2))/n
brierscore_interact.new.model2
```

```
## [1] 0.01067548
```

Overall, we find that the model with additive predictors always produces the better brier score than the interaction model. Thus, the additive predictors model is the best model.

#Question 2 a)

loading and looking at data

```
hiv_data <- read.table("hiv-1.txt", header = TRUE)
head(hiv_data)
```

```
##   sex bs_hiv bupacts fupacts     intervention
## 1   0      0      24       9   no counselling
## 2   0      0       2       2  both counselled
## 3   0      0      15       4  both counselled
## 4   0      1       9       2  both counselled
## 5   0      1       9       1 woman counselled
## 6   0      1       2       0   no counselling
```

After viewing we convert sex from integer to a factor

```
hiv_data$sex <- as.factor(hiv_data$sex)
unique(hiv_data$sex)
```

```
## [1] 0 1
## Levels: 0 1
```

We also convert intervention into a factor format as well

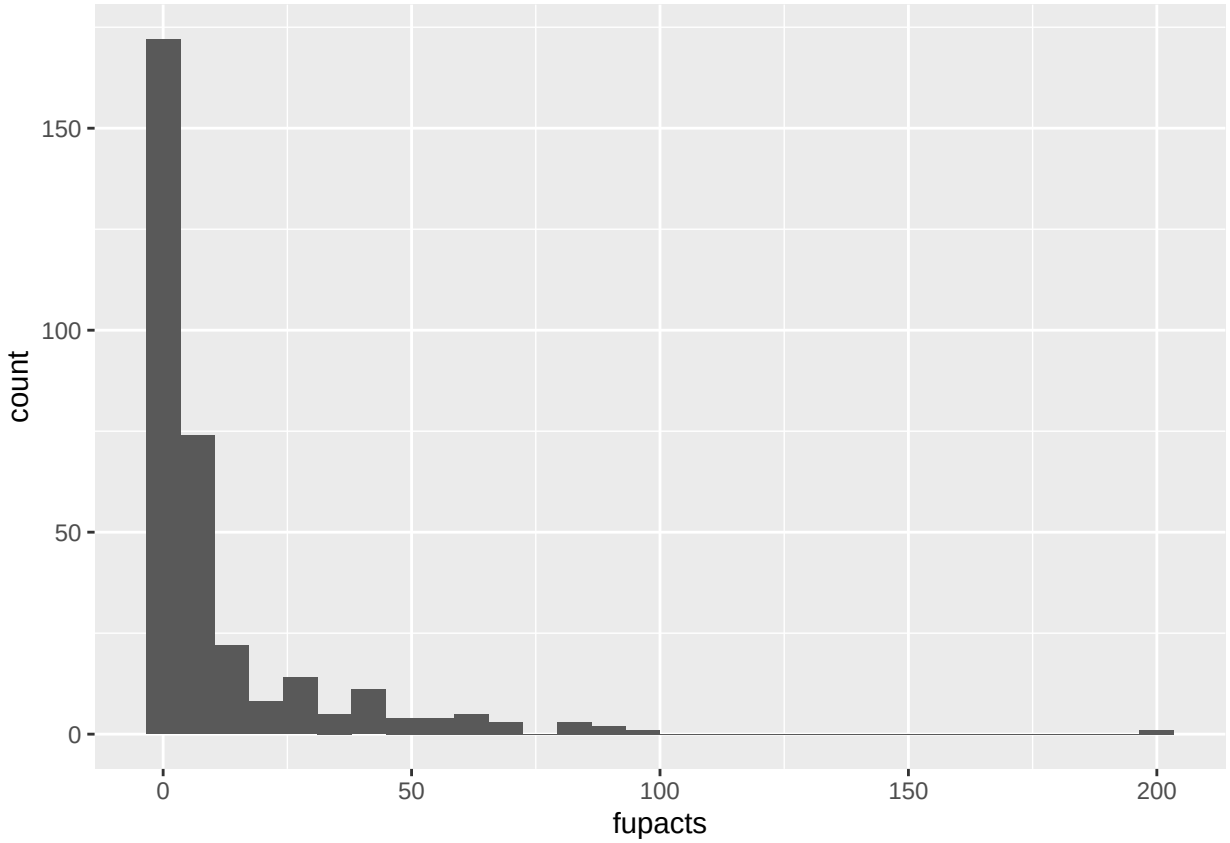
```
hiv_data$intervention <- as.factor(hiv_data$intervention)
unique(hiv_data$intervention)
```

```
## [1] no counselling   both counselled  woman counselled
## Levels: both counselled no counselling woman counselled
```

We want to do some exploratory analysis here by getting a better understanding of our outcome variable; fupacts. Fupacts is the number of unprotected sexual acts after the intervention

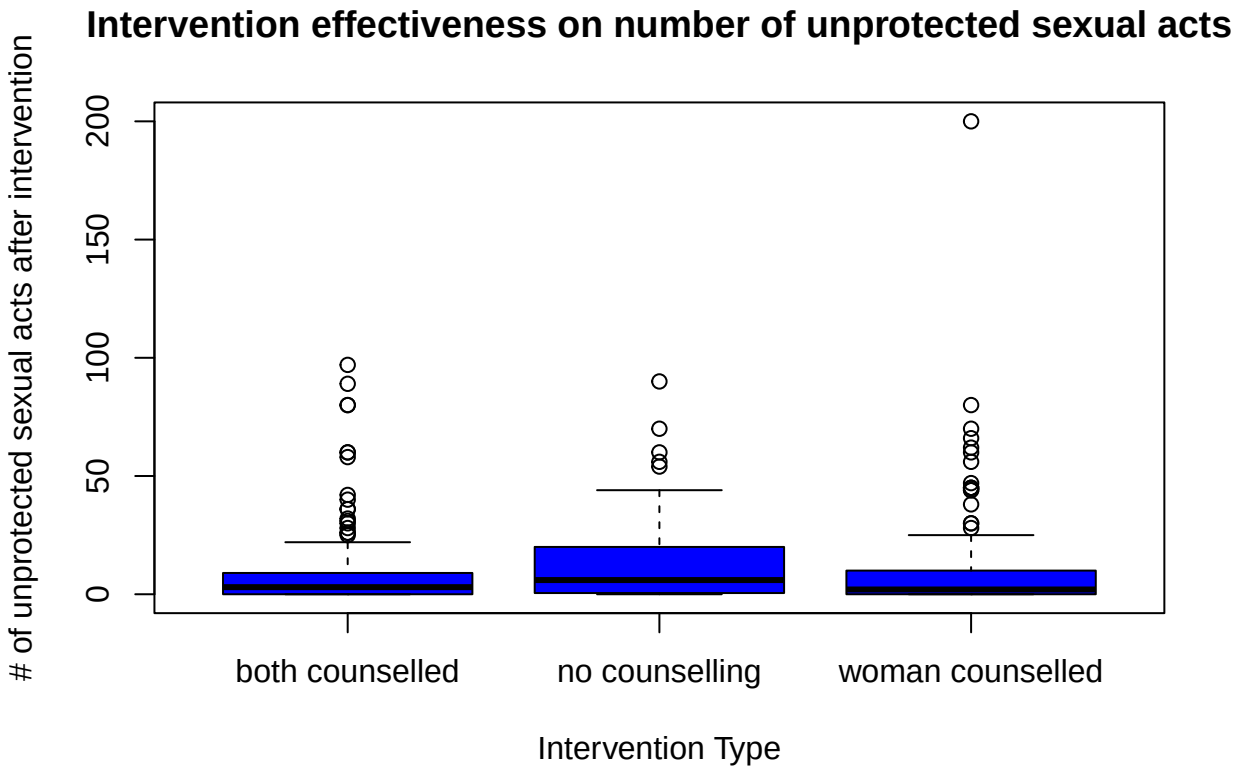
```
hiv_data %>% ggplot(aes(x=fupacts))+geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




Based off this histogram we notice we're going to need a discrete distribution where values are defined at positive values. The poisson distribution fits this criteria

```
boxplot( fupacts ~ intervention ,
         data=hiv_data,
         main="Intervention effectiveness on number of unprotected sexual acts",
         xlab="Intervention Type",
         ylab="# of unprotected sexual acts after intervention",
         col="blue",
         border="black"
       )
```




Exploring this data with number of unprotected sexual acts after the intervention vs the intervention type seems to tell us that having both partners counselled and having the woman counselled reduces the number of unprotected sexual acts. Particularly, having the woman counselled produces the fewest amount as we can see by the median line on the boxplot.

```
hiv_data_clean <- hiv_data %>% slice(-c(140))
hiv_data_clean
```

```
##     sex bs_hiv bupacts fupacts     intervention
## 1     0      0      24       9   no counselling
## 2     0      0       2       2  both counselled
## 3     0      0      15       4  both counselled
## 4     0      1       9       2  both counselled
## 5     0      1       9       1 woman counselled
## 6     0      1       2       0   no counselling
## 7     0      1      40       0 woman counselled
## 8     0      1      15       2 woman counselled
## 9     0      0      60      30 woman counselled
## 10    0      0      10       9 woman counselled
## 11    0      0      80       3  both counselled
## 12    0      1      24       4 woman counselled
## 13    0      0      15       6   no counselling
## 14    0      0      46      40   no counselling
## 15    0      0       6       0 woman counselled
## 16    0      1       5       0   no counselling
## 17    0      0      18      10   no counselling
```

```
## 18    0      0      35       0 woman counselled
## 19    0      0       4       3  both counselled
## 20    0      0      13       2 woman counselled
## 21    0      0       9       3  both counselled
## 22    0      0      10       9  both counselled
## 23    0      0      20       0 woman counselled
## 24    0      0      50      56 woman counselled
## 25    0      0      17       6   no counselling
## 26    0      0       6       2 woman counselled
## 27    0      0      45      27   no counselling
## 28    0      1      85      15  both counselled
## 29    0      1      28       2 woman counselled
## 30    0      1      60       5  both counselled
## 31    0      1      56      35   no counselling
## 32    0      1      34       0  both counselled
## 33    0      0      15       7 woman counselled
## 34    0      0      45      32   no counselling
## 35    0      0      18       0 woman counselled
## 36    0      0      20       2   no counselling
## 37    0      0      30       0 woman counselled
## 38    0      0      20       8   no counselling
## 39    0      0      30       0  both counselled
## 40    0      1      17       1 woman counselled
## 41    0      0      10       0  both counselled
## 42    0      0      64      44 woman counselled
## 43    0      0      15       7  both counselled
## 44    0      0      12      10   no counselling
## 45    0      0      60      66 woman counselled
## 46    0      0      17       0   no counselling
## 47    0      1      20      20   no counselling
## 48    0      1       3       0 woman counselled
## 49    0      0       4       5   no counselling
## 50    0      0      36       3   no counselling
## 51    0      0      60      44 woman counselled
## 52    0      0       2       0 woman counselled
## 53    0      1       8       0  both counselled
## 54    0      0       9       4 woman counselled
## 55    0      0      10      12   no counselling
## 56    0      0       4       2   no counselling
## 57    0      1       1       0 woman counselled
## 58    0      0      35       0  both counselled
## 59    0      1       8       2 woman counselled
## 60    0      0      60      62 woman counselled
## 61    0      1      28       4  both counselled
## 62    0      0      60       6  both counselled
## 63    0      1       4       0  both counselled
## 64    0      0      78      80  both counselled
## 65    0      0       4       2 woman counselled
## 66    0      0      12      13  both counselled
## 67    0      1      40      15   no counselling
## 68    0      1      50      14  both counselled
## 69    0      1       5       0 woman counselled
## 70    0      1      40       0 woman counselled
## 71    0      1       2       2  both counselled
```

```
## 72    0      0       6       7  both counselled
## 73    0      0      36       7 woman counselled
## 74    0      0       1       0  both counselled
## 75    0      0      60      36  both counselled
## 76    0      0      70      56   no counselling
## 77    0      0      24      26   no counselling
## 78    0      0       6       3  both counselled
## 79    0      0      30       0  both counselled
## 80    0      1       3       0   no counselling
## 81    0      0       4       0  both counselled
## 82    0      1      28       0  both counselled
## 83    0      0       5       0 woman counselled
## 84    0      0      17       0 woman counselled
## 85    0      0      20       1   no counselling
## 86    0      0      90      60  both counselled
## 87    0      0      50       6   no counselling
## 88    0      1      30       0  both counselled
## 89    0      0      15      15  both counselled
## 90    0      0       4       5   no counselling
## 91    0      0      40      31  both counselled
## 92    0      0      10       4 woman counselled
## 93    0      0       2       0  both counselled
## 94    0      0      24       0 woman counselled
## 95    0      0      60      60  both counselled
## 96    0      0      30       0 woman counselled
## 97    0      1       1       0 woman counselled
## 98    0      0      10       1  both counselled
## 99    0      1       8       3   no counselling
## 100   0      0     100      28  both counselled
## 101   0      0       5       4  both counselled
## 102   0      0      28      26  both counselled
## 103   0      0      40       4   no counselling
## 104   0      1      12       0 woman counselled
## 105   0      0      16       5 woman counselled
## 106   0      0      12       3  both counselled
## 107   0      1       3       0   no counselling
## 108   0      0      15       0 woman counselled
## 109   0      0      36      42   no counselling
## 110   0      0      20      14 woman counselled
## 111   0      0      12       0   no counselling
## 112   0      0      60       3 woman counselled
## 113   0      1      15      11 woman counselled
## 114   0      0       2       0 woman counselled
## 115   0      0      90       0  both counselled
## 116   0      1      11       0  both counselled
## 117   0      1       5       0 woman counselled
## 118   0      0      36       6 woman counselled
## 119   0      0       1       0   no counselling
## 120   0      1      10       0  both counselled
## 121   0      0      30       8  both counselled
## 122   0      0      48       8  both counselled
## 123   0      0      10       2 woman counselled
## 124   0      0      90       0  both counselled
## 125   0      0       2       0   no counselling
```

```
## 126   0      0      60      54   no counselling
## 127   0      0       9       2 woman counselled
## 128   0      0       4       1  both counselled
## 129   0      0       3       3 woman counselled
## 130   0      1      25       0 woman counselled
## 131   0      0      25       0 woman counselled
## 132   0      0       9       0   no counselling
## 133   0      1      45      42  both counselled
## 134   0      0      90      60   no counselling
## 135   0      1       8       0  both counselled
## 136   0      0       4       0  both counselled
## 137   0      0      45       0   no counselling
## 138   0      0      40       0 woman counselled
## 139   0      0      50      60 woman counselled
## 141   0      0      28       0 woman counselled
## 142   0      0       2       0  both counselled
## 143   0      0       8       0  both counselled
## 144   0      0       1       0  both counselled
## 145   0      0       3       0 woman counselled
## 146   0      1      15       0  both counselled
## 147   0      0       2       1  both counselled
## 148   0      0       9       0  both counselled
## 149   0      0      20       2  both counselled
## 150   0      1      90       0  both counselled
## 151   0      0       1       0  both counselled
## 152   0      0      50       0 woman counselled
## 153   0      0      80      89  both counselled
## 154   0      1       2       2  both counselled
## 155   0      0      90       0 woman counselled
## 156   0      0      26       0 woman counselled
## 157   0      0      30      10  both counselled
## 158   0      0       8       4  both counselled
## 159   0      0      18       2  both counselled
## 160   0      0      34      10  both counselled
## 161   0      0      40       0  both counselled
## 162   0      0      36      13  both counselled
## 163   0      0      36       3   no counselling
## 164   0      0      20       4   no counselling
## 165   0      0      33       2   no counselling
## 166   0      0      20       7   no counselling
## 167   1      0      10       6   no counselling
## 168   1      0       1       0  both counselled
## 169   1      1       7       0  both counselled
## 170   1      1      20       5  both counselled
## 171   1      1      10       0 woman counselled
## 172   1      1       5       0   no counselling
## 173   1      1      50       0 woman counselled
## 174   1      0       7       2 woman counselled
## 175   1      0      33      12 woman counselled
## 176   1      0      90       0 woman counselled
## 177   1      0      12       1 woman counselled
## 178   1      0      15       3  both counselled
## 179   1      0      87       6   no counselling
## 180   1      1       8       5 woman counselled
```

```
## 181   1      0      45      16   no counselling
## 182   1      0      47      24   no counselling
## 183   1      1       1       0 woman counselled
## 184   1      0      70       0   no counselling
## 185   1      1      19       2  both counselled
## 186   1      0      36       7 woman counselled
## 187   1      1       4       0 woman counselled
## 188   1      0      22      20  both counselled
## 189   1      0       5       0 woman counselled
## 190   1      0      17       0 woman counselled
## 191   1      0      30       2 woman counselled
## 192   1      0      50      40   no counselling
## 193   1      0      20      20   no counselling
## 194   1      0      70      45 woman counselled
## 195   1      0       8      10 woman counselled
## 196   1      0      12      15 woman counselled
## 197   1      0       4       5   no counselling
## 198   1      1      70      10  both counselled
## 199   1      1      36       0 woman counselled
## 200   1      0      16       3 woman counselled
## 201   1      1      90       0  both counselled
## 202   1      0       3       1   no counselling
## 203   1      1      37       0  both counselled
## 204   1      0       2       1 woman counselled
## 205   1      0      24       6  both counselled
## 206   1      0      45      30  both counselled
## 207   1      0      90      39   no counselling
## 208   1      0      60      45 woman counselled
## 209   1      0      60      44   no counselling
## 210   1      0      60      24 woman counselled
## 211   1      0      60      25   no counselling
## 212   1      0       4       2   no counselling
## 213   1      0      90       0  both counselled
## 214   1      1      90       0 woman counselled
## 215   1      0      90      80 woman counselled
## 216   1      0      16       6  both counselled
## 217   1      0      70      15   no counselling
## 218   1      0      30      10   no counselling
## 219   1      0      60      47 woman counselled
## 220   1      0       1       0   no counselling
## 221   1      0       6       0 woman counselled
## 222   1      0      20      15   no counselling
## 223   1      0      45      40   no counselling
## 224   1      0      20      20   no counselling
## 225   1      1       4       3  both counselled
## 226   1      0      12      10 woman counselled
## 227   1      0      60      70   no counselling
## 228   1      0      70      22  both counselled
## 229   1      1      50      15  both counselled
## 230   1      1      10       0 woman counselled
## 231   1      0      70      25 woman counselled
## 232   1      1      20      16  both counselled
## 233   1      0      35       5  both counselled
## 234   1      1       6       0  both counselled
```

```
## 235   1      0      15       4  both counselled
## 236   1      0      15      15 woman counselled
## 237   1      1       8       7  both counselled
## 238   1      1      90      80  both counselled
## 239   1      0       8       5  both counselled
## 240   1      0      80      90   no counselling
## 241   1      1      90      40  both counselled
## 242   1      1      30      10 woman counselled
## 243   1      0       6       0 woman counselled
## 244   1      0       4       0  both counselled
## 245   1      0      50      10   no counselling
## 246   1      0      15       0  both counselled
## 247   1      0       9       1  both counselled
## 248   1      0      70       8  both counselled
## 249   1      0     228      97  both counselled
## 250   1      0      60       4   no counselling
## 251   1      0       4       5   no counselling
## 252   1      1      47      32  both counselled
## 253   1      0       4       4  both counselled
## 254   1      1       6       0  both counselled
## 255   1      0      50      16 woman counselled
## 256   1      0      36      22 woman counselled
## 257   1      0       7       9 woman counselled
## 258   1      0      10       1   no counselling
## 259   1      0      20       5  both counselled
## 260   1      0       4       0   no counselling
## 261   1      0       3       0  both counselled
## 262   1      0      25       3   no counselling
## 263   1      0      36      26  both counselled
## 264   1      0      50       9 woman counselled
## 265   1      0       1       0  both counselled
## 266   1      1       5       3 woman counselled
## 267   1      1      12       0   no counselling
## 268   1      0      36       0   no counselling
## 269   1      0      24      10 woman counselled
## 270   1      0      10       2  both counselled
## 271   1      0      50      25  both counselled
## 272   1      0       5       1  both counselled
## 273   1      0      36      13  both counselled
## 274   1      0      36      30 woman counselled
## 275   1      0      12       0   no counselling
## 276   1      0       3       2 woman counselled
## 277   1      1       6       0  both counselled
## 278   1      1       2       0   no counselling
## 279   1      0      20       3 woman counselled
## 280   1      0      10       5  both counselled
## 281   1      0      30       3 woman counselled
## 282   1      0      12      10   no counselling
## 283   1      1      50      28 woman counselled
## 284   1      0       3       0 woman counselled
## 285   1      0       3       1 woman counselled
## 286   1      0      80      14  both counselled
## 287   1      0       2       0  both counselled
## 288   1      0      12       0 woman counselled
```

```
## 289   1      0      35       8 woman counselled
## 290   1      0      90       0  both counselled
## 291   1      0      12       6 woman counselled
## 292   1      0       3       0 woman counselled
## 293   1      0      15       0   no counselling
## 294   1      0      10      10  both counselled
## 295   1      0      40       4  both counselled
## 296   1      0       3       0  both counselled
## 297   1      0      35      11   no counselling
## 298   1      0      20      10 woman counselled
## 299   1      0      24      10 woman counselled
## 300   1      0      15       0 woman counselled
## 301   1      0      20       0   no counselling
## 302   1      0     270       5  both counselled
## 303   1      0      10       0  both counselled
## 304   1      0       5       5  both counselled
## 305   1      1      40      16   no counselling
## 306   1      1       1       0  both counselled
## 307   1      0      36      38 woman counselled
## 308   1      0      99      45 woman counselled
## 309   1      0      90      70 woman counselled
## 310   1      0      12       0 woman counselled
## 311   1      1       2       0  both counselled
## 312   1      0      20       0  both counselled
## 313   1      0      16       8  both counselled
## 314   1      0       3       3 woman counselled
## 315   1      1       4       3  both counselled
## 316   1      0      12       7  both counselled
## 317   1      0      10       0  both counselled
## 318   1      0      24       0  both counselled
## 319   1      1      90      58  both counselled
## 320   1      1       1       0  both counselled
## 321   1      0       4       5 woman counselled
## 322   1      1      17       0  both counselled
## 323   1      0      90       0 woman counselled
## 324   1      0      36      36  both counselled
## 325   1      0      24       4  both counselled
## 326   1      0      18       0  both counselled
## 327   1      0      10       5  both counselled
## 328   1      0      24      27   no counselling
## 329   1      0      24       1   no counselling
```

Noticed the outlier on the women counselled boxplot where about 200 sexual acts were committed seems unreasonable so we will remove this.

b)

As previously stated, we're going to need a discrete distribution where values are defined at positive values since this is a count outcome. Also, remember our outcome variable fupacts is an unrestricted count.

c) Fit a Poisson regression model

```
hiv_poireg <- glm(data = hiv_data_clean, formula = fupacts ~ sex + bs_hiv
                  + intervention,
                  offset = log(bupacts), family = poisson)
summary(hiv_poireg)
```

```
##
## Call:
## glm(formula = fupacts ~ sex + bs_hiv + intervention, family = poisson,
##     data = hiv_data_clean, offset = log(bupacts))
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -12.0481   -2.6126   -0.8647    1.4343    9.5228
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -1.05975    0.03497 -30.307  < 2e-16 ***
## sex1                         -0.01527    0.03379  -0.452  0.65130
## bs_hiv                       -0.33484    0.04861  -6.889 5.63e-12 ***
## interventionno counselling    0.34336    0.04188   8.198 2.44e-16 ***
## interventionwoman counselled  0.12400    0.04083   3.037  0.00239 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 3879.4  on 327  degrees of freedom
## Residual deviance: 3732.6  on 323  degrees of freedom
## AIC: 4592.3
##
## Number of Fisher Scoring iterations: 5
```

We fit a poisson model with number of unprotected sexual acts after the intervention (fupacts) as the outcome variable. Sex, HIV status, and intervention as predictors. We use the number of unprotected sexual acts before the intervention (bupacts) as the offset as we want to know the number of unprotected sexuals acts after the intervention in proportion to the number of how much they were doing before. We use log to make sure R doesn't complain. We notice an interaction between having hiv and a woman being counselled

```
summary(hiv_poireg)
```

```
##
## Call:
## glm(formula = fupacts ~ sex + bs_hiv + intervention, family = poisson,
##     data = hiv_data_clean, offset = log(bupacts))
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -12.0481   -2.6126   -0.8647    1.4343    9.5228
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -1.05975    0.03497 -30.307  < 2e-16 ***
```

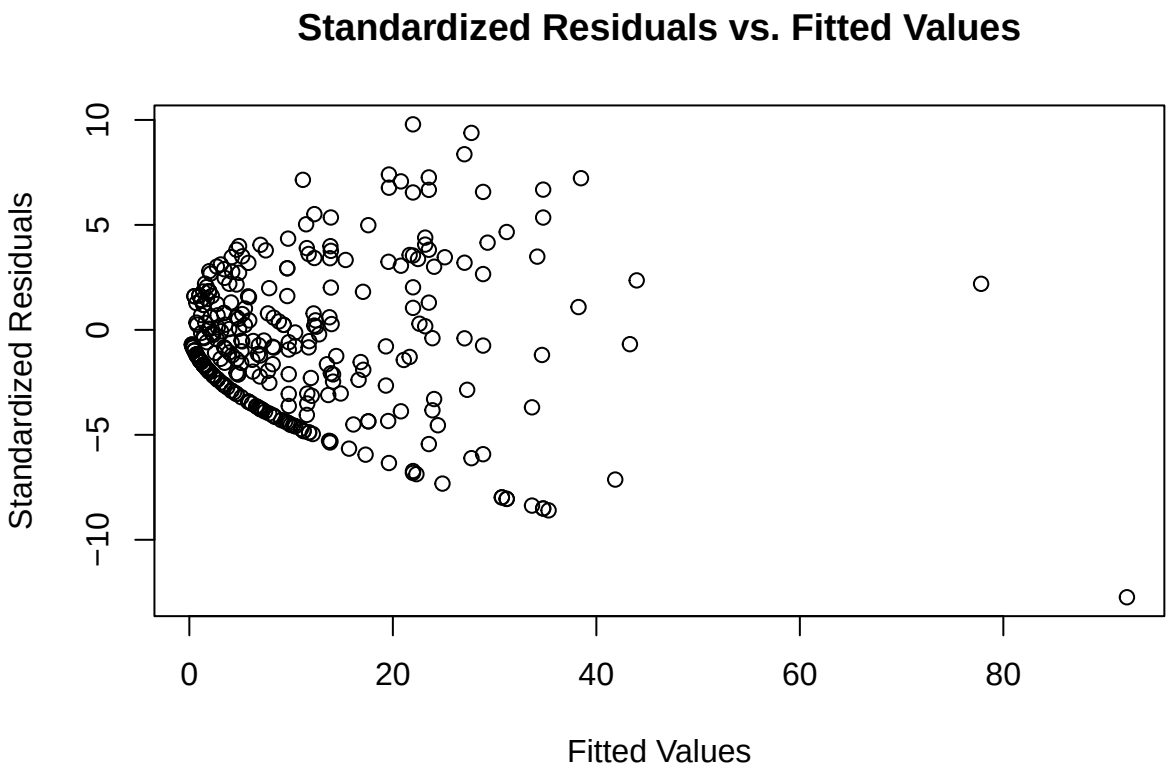
```
## sex1                          -0.01527    0.03379  -0.452  0.65130
## bs_hiv                        -0.33484    0.04861  -6.889 5.63e-12 ***
## interventionno counselling     0.34336    0.04188   8.198 2.44e-16 ***
## interventionwoman counselled   0.12400    0.04083   3.037  0.00239 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 3879.4  on 327  degrees of freedom
## Residual deviance: 3732.6  on 323  degrees of freedom
## AIC: 4592.3
##
## Number of Fisher Scoring iterations: 5
```

d)

We notice a cone/fanning out shape when graphically displays the residuals, which is problematic. This indicates evidence of overdispersion.

```
hiv_stres <- rstandard(hiv_poireg)
test_stat <- sum(hiv_stres^2)

plot(x= hiv_poireg$fitted.values, y= hiv_stres,
xlab = "Fitted Values",
ylab = "Standardized Residuals",
main = "Standardized Residuals vs. Fitted Values")
```




Numerically assessing for overdispersion confirms this.

```
degrees_freedom_hiv = nrow(hiv_data_clean) - length(hiv_poireg$coefficients)

1-pchisq(test_stat, degrees_freedom_hiv)
```

```
## [1] 0
```

Our test_stat is so low that it renders as 0. This indicates evidence of overdispersio as it's below a p-value of 0.05

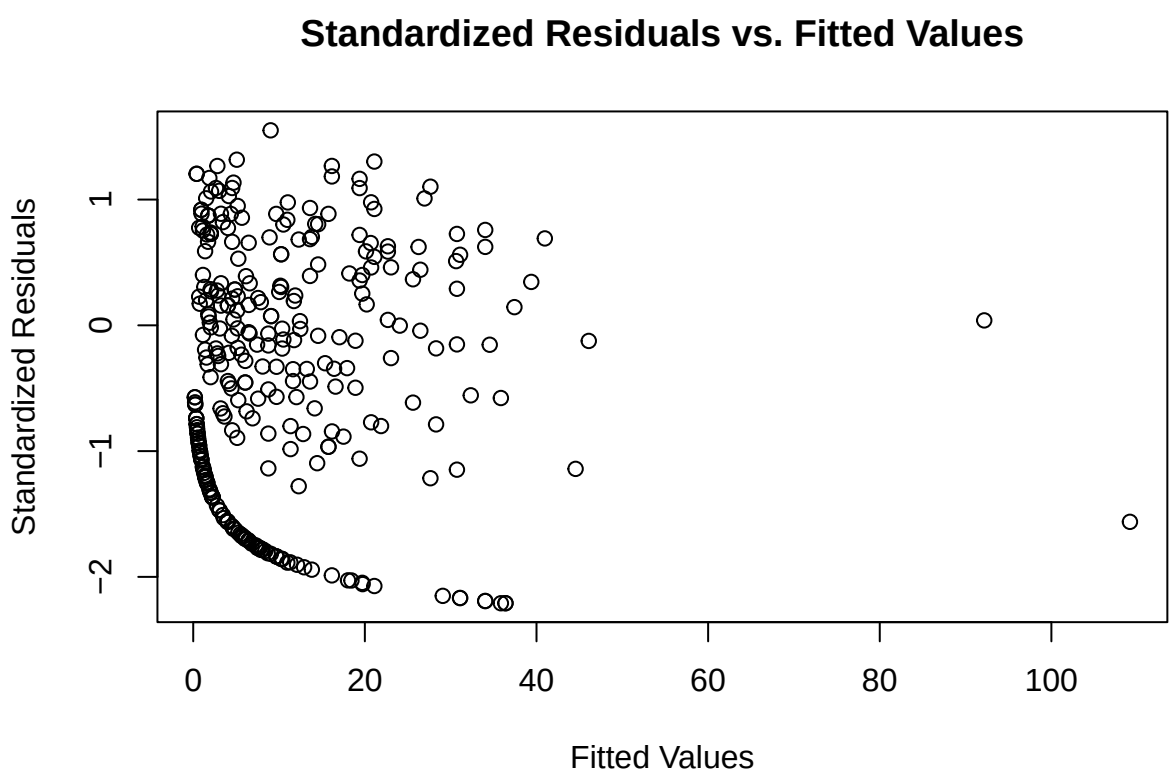
e)

We run a negative binomial model to account for this overdispersion as this has a different form than the poisson. We also notice our estimates are close as our possion model but not the same.

```
nb_model <- glm.nb(fupacts ~ sex + bs_hiv + intervention
                   + offset(log(bupacts)), data = hiv_data_clean)

hiv_nbstres <- rstandard(nb_model)
test_stat <- sum(hiv_nbstres^2)

plot(x= nb_model$fitted.values, y= hiv_nbstres,
xlab = "Fitted Values",
ylab = "Standardized Residuals",
main = "Standardized Residuals vs. Fitted Values")
```




```
summary(nb_model)
```

```
##
## Call:
## glm.nb(formula = fupacts ~ sex + bs_hiv + intervention + offset(log(bupacts)),
##     data = hiv_data_clean, init.theta = 0.578964512, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1938  -1.1947  -0.3392   0.3692   1.5334
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   -1.06276    0.15842  -6.709 1.97e-11 ***
## sex1                           0.15711    0.15740   0.998  0.31821
## bs_hiv                        -0.54461    0.19317  -2.819  0.00481 **
## interventionno counselling     0.23649    0.20224   1.169  0.24226
## interventionwoman counselled -0.06704    0.18310  -0.366  0.71424
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(0.579) family taken to be 1)
##
##     Null deviance: 364.69  on 327  degrees of freedom
## Residual deviance: 353.35  on 323  degrees of freedom
```

```
## AIC: 1858.3
##
## Number of Fisher Scoring iterations: 1
##
##
##                Theta:  0.5790
##            Std. Err.:  0.0582
##
##  2 x log-likelihood:  -1846.2610
```

f)

It seems as if have both counselled proves to be the most effective as no counselling gives a p-value of 0.24 and woman counselled gives a p-value of 0.71424.