Final Project Part 2

Hannah Rosenbaum

Background

Fifty years after the passage of the Title IX Amendment, collegial sports equity has shown relatively minimal change. Allocations of sports budgets often highlight pay discrepancies in participants being "spent [on] $4,285 per men's participant versus $2,588 per women's participant." (Feinberg, D., & Hunzinger, E) With these vast differences in individual spending by gender, we see this phenomenon only heightened in the NCAA with women's basketball. Women's basketball not only fares having lower budgets from the NCAA but also, per an ESPN report, "is underpaying the NCAA for the tournament rights for 29 championships causing the association to lose out on substantial and crucial revenue... denoting that the current budget of $81 million to $112 million multiples more than what the network currently gives." (Zimmbalist) Thus, there is not only a discrepancy in budget allocations among the participants by gender but also amongst large broadcast networks.

Significant systemic issues occur within the gendered branding of 'March Madness.' This can be seen with differentiated treatment of male versus female brackets due to the lack of general awareness of when the women's bracket games even occur. Largely the inequity of the 'March Madness' tournament derives from a differentiation from the NCAA in "distribution agreements, corporate sponsorships, distribution of revenue, organizational structure and culture all to prioritize Division I men's basketball over everything else... to perpetuate gender inequities." (Blinder) Likewise, this institutional creation of a high investment in TV rights for men's basketball and minimal airtime for the women's bracket has led to smaller budgeting and fewer avenues to earn revenue. This has led women's teams to be "starved of a starring role in the national discourse." (Blinder) Thus, it creates a circular effect in women's basketball, deriding fewer resources even within facilities at the NCAA tournament in 2021 and in general awareness of TV times.

I am primarily interested in discussing sports equity in women's basketball due to my own personal experience at UF of wanting to watch NCAA basketball for women but having no general knowledge of when women play. I believe that the discussion of equity in sports for women is essential because of the common dismissal of watching women's sports as a pastime.

Research Questions

1. What is the relationship between female students' post-secondary education enrollment compared to the ratio of female athletes at those institutions?
2. How does the expenditure of those sports programs impact the percentage of females in university sports?
3. How does the revenue allocate to university sports reflect the percentage of females in university sports?

Hypothesis Testing

1. There is a relationship between a higher percentage of female students in post-secondary education and the rate of female athletes.
2. There is a relationship between expenditure on university sports programs and the percentage of females in university sports programs.
3. There is a relationship between revenue from university sports programs and the percentage of females in sports programs.

Descriptive Statistics

The Equity in Athletics Disclosures Act requires the full financial disclosure of total expenditures, revenue, staffing, and recruiting efforts by men's and women's athletic programs (Mock, J.T.). Data provided by the Equity in Sports project is from all postsecondary programs that receive government funding from Title IV funding and is an online database of funding expenses from 2015-2019.

There are 132,327 rows and a total of 28 columns.

---- ANSWERING 3. ——

To measure female participation, I will create a model with `sum_partic_women` as the dependent variable and `ef_female_count` as the explanatory variable.

-------- ANSWERING 6. —

The null data in the `data` matrix exist because a given entry has no male or female participation. The columns with null data are `rev_men`, `rev_women`, `exp_men`, `exp_women`.

Read in Sports Equity data-set

```r
sports <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/m
```

```
Rows: 132327 Columns: 28
-- Column specification --------------------------------------------------------
Delimiter: ","
chr  (8): institution_name, city_txt, state_cd, zip_text, classification_nam...
dbl (20): year, unitid, classification_code, ef_male_count, ef_female_count,...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
library(dplyr)
```

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```r
library(wesanderson)
library(ggplot2)
```

Removing 'Ottawa University-Pheonix' due to having zero total male and female attendance

```r
sports = filter(sports, institution_name != "Ottawa University-Phoenix")
```

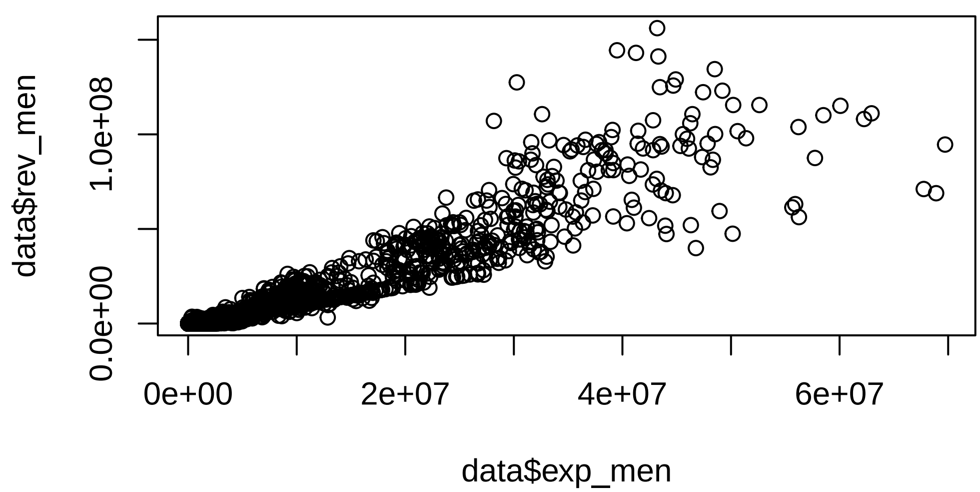
Create data-frames: Critical dimensions, Attendance specific, Basketball specific

```
data <- as.data.frame(sports[, c("year", "institution_name", "sports", "ef_male_count", "e

attendance_data <- data[,c("institution_name", "sports", "ef_male_count", "ef_female_count

basketball <- as.data.frame(sports[, c("year", "institution_name", "sports", "ef_male_coun
basketball <- filter(basketball, sports=='Basketball')
```

```
institute_lbl <- distinct(as.data.frame(data[, c("institution_name")]))
sport_lbl <- distinct(as.data.frame(data[, c("sports")]))
```

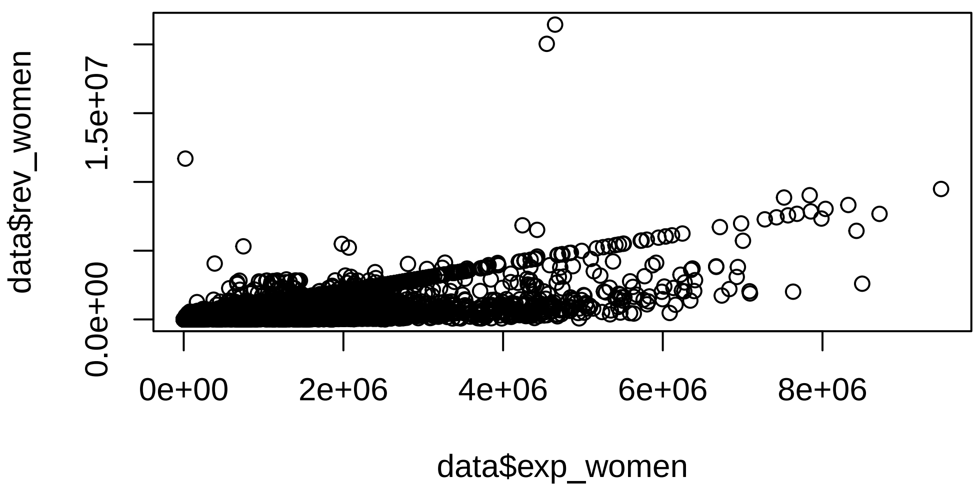
Scatter plots comparing Expenditures against Revenue by Gender

```
#data[is.na(data)] <- 0
plot(data$exp_men, data$rev_men)
```



```
#ggplot(data = data, aes(x=exp_men, y=rev_men), fill = institute_lbl) +
#geom_point() +
#scale_fill_manual(values = wes_palette(length(institute_lbl), name = "GrandBudapest1", ty
```

```
plot(data$exp_women, data$rev_women)
```



Descriptive Statistics

```
glimpse(data)
```

```
Rows: 132,317
Columns: 11
$ year              <dbl> 2015, 2015, 2015, 2015, 2015, 2015, 2015, 2015, 2015,~
$ institution_name <chr> "Alabama A & M University", "Alabama A & M University~
$ sports            <chr> "Baseball", "Basketball", "All Track Combined", "Foot~
$ ef_male_count     <dbl> 1923, 1923, 1923, 1923, 1923, 1923, 1923, 1923, 1923,~
$ ef_female_count   <dbl> 2300, 2300, 2300, 2300, 2300, 2300, 2300, 2300, 2300,~
$ sum_partic_men    <dbl> 31, 19, 61, 99, 9, 0, 0, 7, 0, 0, 32, 13, 0, 10, 2, 3~
$ sum_partic_women <dbl> 0, 16, 46, 0, 0, 21, 25, 10, 16, 9, 0, 20, 68, 7, 10,~
$ rev_men           <dbl> 345592, 1211095, 183333, 2808949, 78270, NA, NA, 7827~
```

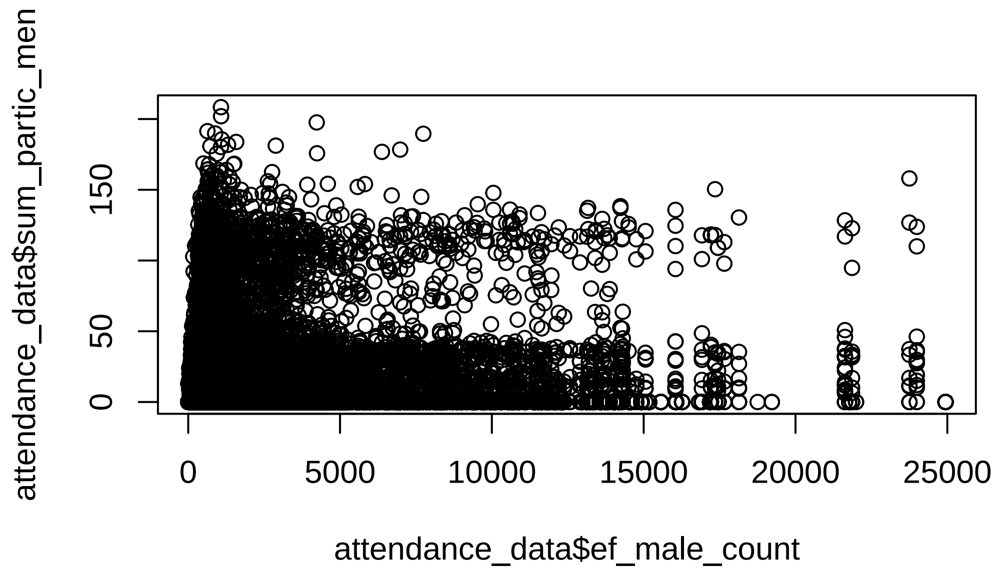
```
$ rev_women        <dbl> NA, 748833, 315574, NA, NA, 410717, 298164, 131145, 3~
$ exp_men          <dbl> 397818, 817868, 246949, 3059353, 83913, NA, NA, 99612~
$ exp_women        <dbl> NA, 742460, 251184, NA, NA, 432648, 340259, 113886, 3~
```

```r
summary(data)
```

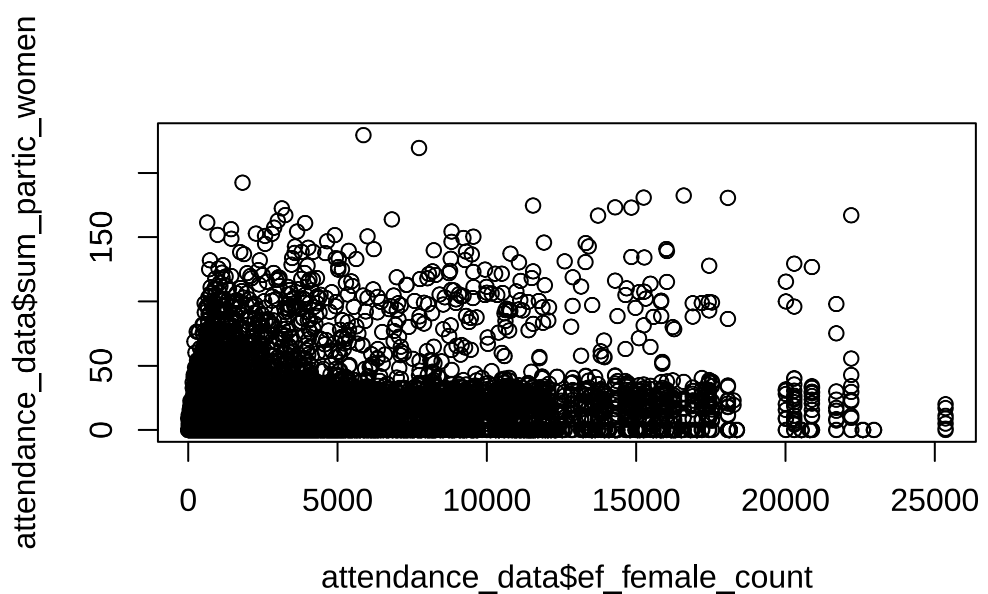
```
      year      institution_name      sports          ef_male_count
 Min.   :2015   Length:132317      Length:132317      Min.   :    0
 1st Qu.:2016   Class :character   Class :character   1st Qu.:  514
 Median :2018   Mode  :character   Mode  :character   Median :  986
 Mean   :2018                                         Mean   : 2126
 3rd Qu.:2019                                         3rd Qu.: 2385
 Max.   :2019                                         Max.   :35954

 ef_female_count sum_partic_men   sum_partic_women    rev_men
 Min.   :    0   Min.   :  0.00   Min.   :  0.00   Min.   :       65
 1st Qu.:  652   1st Qu.:  0.00   1st Qu.:  0.00   1st Qu.:    63406
 Median : 1249   Median :  0.00   Median :  6.00   Median :   158069
 Mean   : 2496   Mean   : 14.49   Mean   : 10.86   Mean   :   809028
 3rd Qu.: 2860   3rd Qu.: 20.00   3rd Qu.: 17.00   3rd Qu.:   400383
 Max.   :30325   Max.   :331.00   Max.   :327.00   Max.   :156147208
                                                   NA's   :70460
   rev_women           exp_men             exp_women
 Min.   :       0   Min.   :       65   Min.   :     65
 1st Qu.:   58742   1st Qu.:   63049   1st Qu.:  59294
 Median :  138292   Median :  159649   Median : 141780
 Mean   :  279332   Mean   :  662384   Mean   : 331585
 3rd Qu.:  331034   3rd Qu.:  423980   3rd Qu.: 361817
 Max.   :21440365   Max.   :69718059   Max.   :9485162
 NA's   :63441      NA's   :70460      NA's   :63439
```

Scatter plots comparing Institution Attendance against Participation by Gender

```r
plot(attendance_data$ef_male_count, attendance_data$sum_partic_men)
```

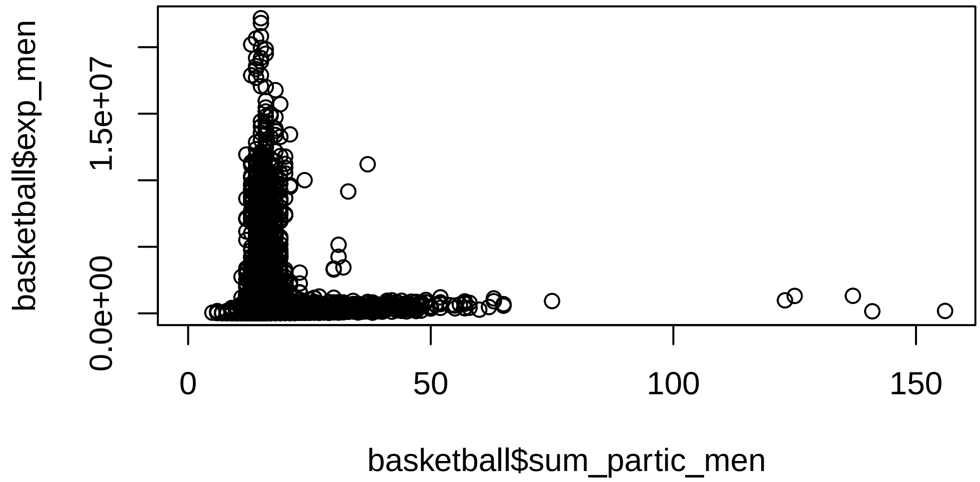



```
plot(attendance_data$ef_female_count, attendance_data$sum_partic_women)
```

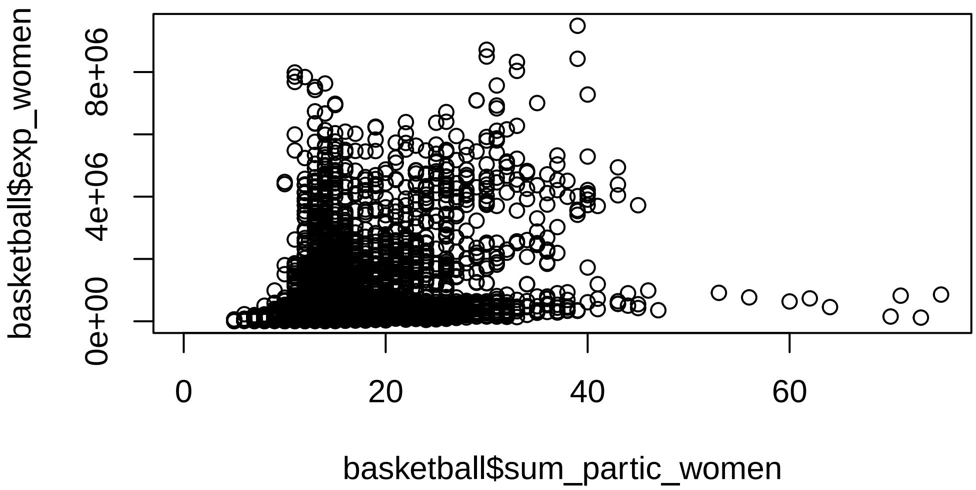


Scatter plots comparing basketball Participation against Expenditures by Gender

```
plot(basketball$sum_partic_men, basketball$exp_men)
```




```
plot(basketball$sum_partic_women, basketball$exp_women)
```




```
For the dataset, I could extrapolate my variables of interest as seen here: <https://github.c
```

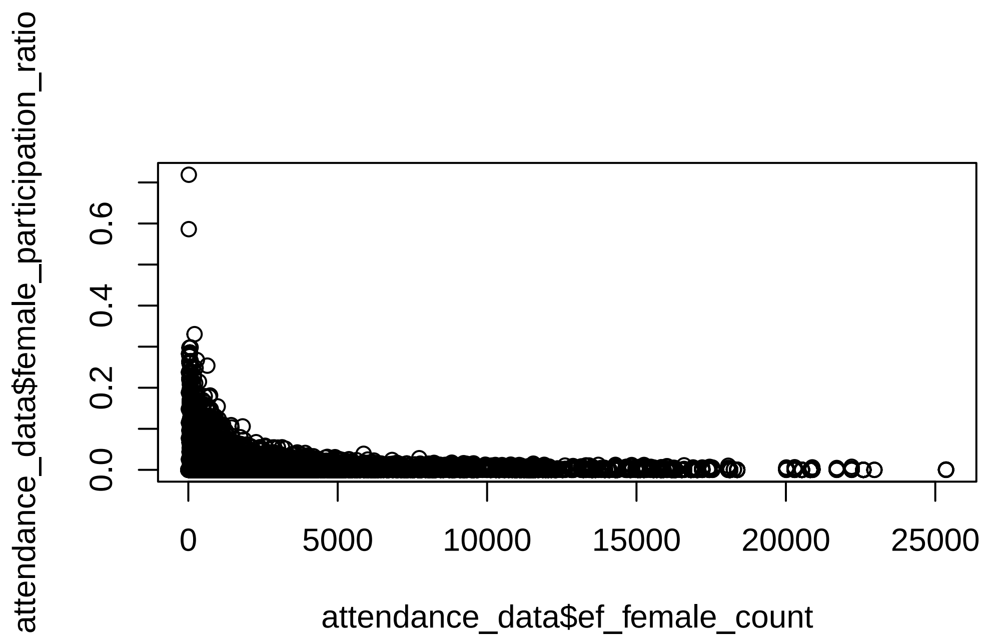
Hypothesis Test 1

Response variable: sum_partic_women

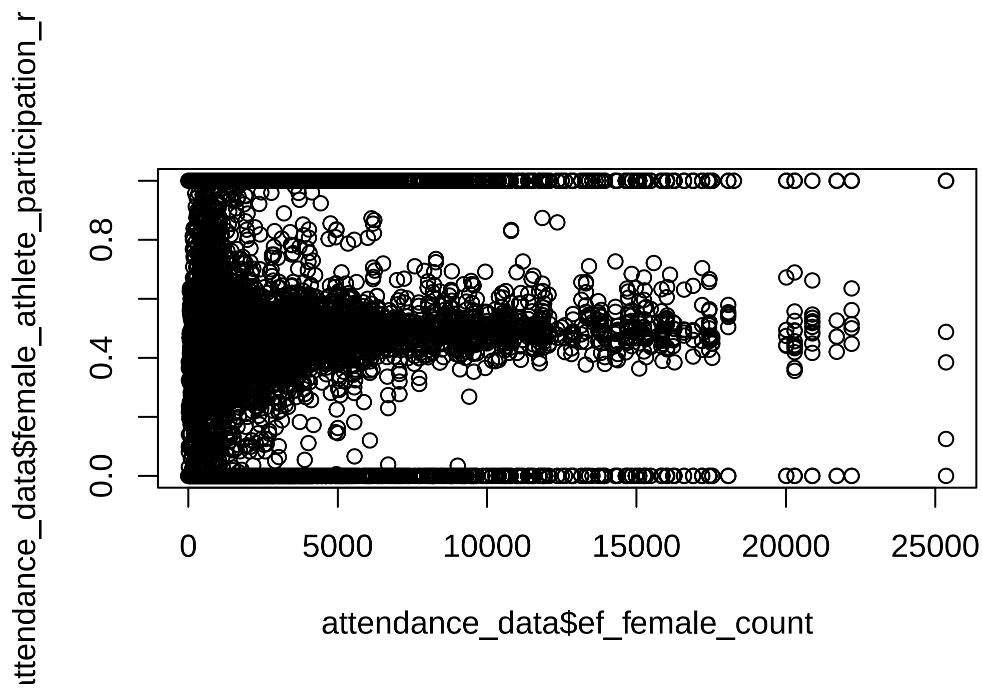
Explanatory variable: sum_partic_women / ef_female_count

Control variable: sum_partic_men / ef_male_count

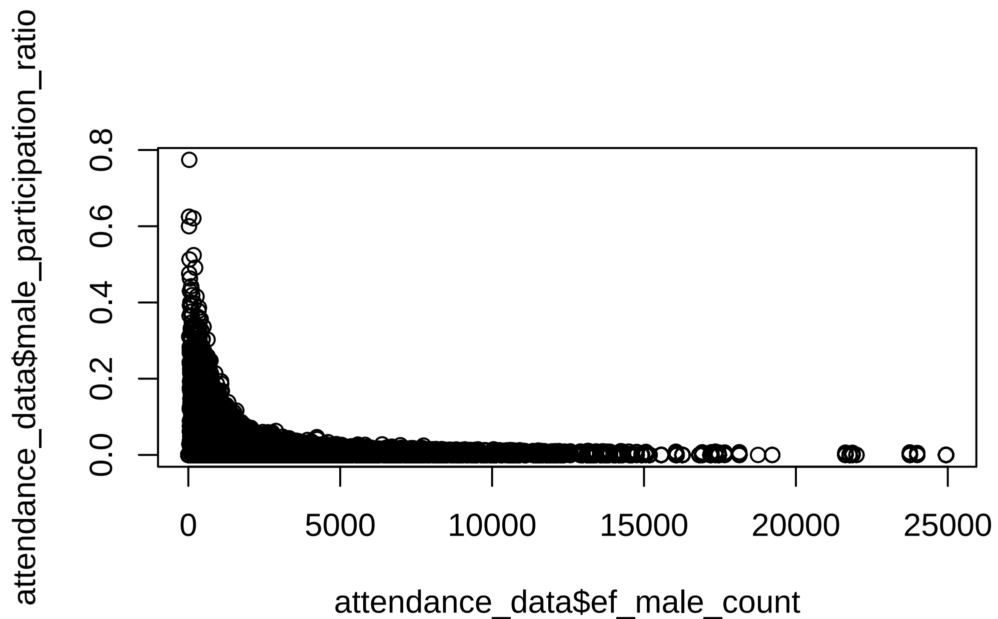
```
attendance_data$female_participation_ratio <- attendance_data$sum_partic_women / attendanc
attendance_data$female_athlete_participation_ratio <- attendance_data$sum_partic_women / (
attendance_data$male_participation_ratio <- attendance_data$sum_partic_men / attendance_da

#ggplot(data = attendance_data, aes(x=ef_female_count, y=female_participation_ratio)) + ge
plot(attendance_data$ef_female_count, attendance_data$female_participation_ratio)
```




```r
plot(attendance_data$ef_female_count, attendance_data$female_athlete_participation_ratio)
```



```r
plot(attendance_data$ef_male_count, attendance_data$male_participation_ratio)
```




```r
hyp_1_fit_1 <- lm(female_participation_ratio ~ ef_female_count, data = filter(attendance_d
hyp_1_fit_2 <- lm(female_participation_ratio ~ ef_female_count, data = filter(attendance_d
hyp_1_fit_3 <- lm(female_athlete_participation_ratio ~ ef_female_count, data = filter(atte
summary(hyp_1_fit_1)
```

```
Call:
lm(formula = female_participation_ratio ~ ef_female_count, data = filter(attendance_data,
    female_participation_ratio != Inf))

Residuals:
     Min       1Q   Median       3Q      Max
-0.00673 -0.00621 -0.00547 -0.00002  0.71203

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      6.728e-03  7.947e-05   84.66   <2e-16 ***
ef_female_count -6.975e-07  2.044e-08  -34.13   <2e-16 ***
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01605 on 63166 degrees of freedom
Multiple R-squared:  0.01811,   Adjusted R-squared:  0.01809
F-statistic:  1165 on 1 and 63166 DF,  p-value: < 2.2e-16
```

```r
summary(hyp_1_fit_2)
```

```
Call:
lm(formula = female_participation_ratio ~ ef_female_count, data = filter(attendance_data,
    female_participation_ratio != Inf & sum_partic_women > 0))

Residuals:
     Min       1Q   Median       3Q      Max
-0.02575 -0.01313 -0.00782  0.00427  0.69150

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      2.729e-02  2.515e-04  108.52   <2e-16 ***
ef_female_count -2.868e-06  6.068e-08  -47.26   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02531 on 16114 degrees of freedom
Multiple R-squared:  0.1218,    Adjusted R-squared:  0.1217
F-statistic:  2234 on 1 and 16114 DF,  p-value: < 2.2e-16
```

```r
summary(hyp_1_fit_3)
```

```
Call:
lm(formula = female_athlete_participation_ratio ~ ef_female_count,
    data = filter(attendance_data, female_athlete_participation_ratio !=
        Inf & sum_partic_women > 0))

Residuals:
    Min      1Q  Median      3Q     Max
-0.7416 -0.1988 -0.1326  0.3447  0.3774
```

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     6.226e-01  2.643e-03  235.54   <2e-16 ***
ef_female_count 9.622e-06  6.378e-07   15.09   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.266 on 16115 degrees of freedom
Multiple R-squared:  0.01393,   Adjusted R-squared:  0.01387
F-statistic: 227.6 on 1 and 16115 DF,  p-value: < 2.2e-16
```

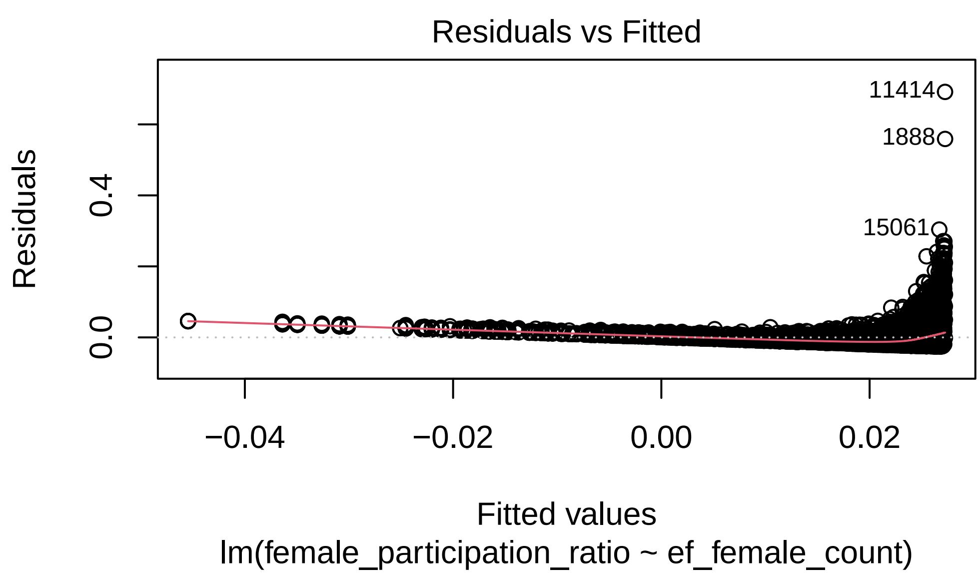
```r
AIC(hyp_1_fit_2)
```

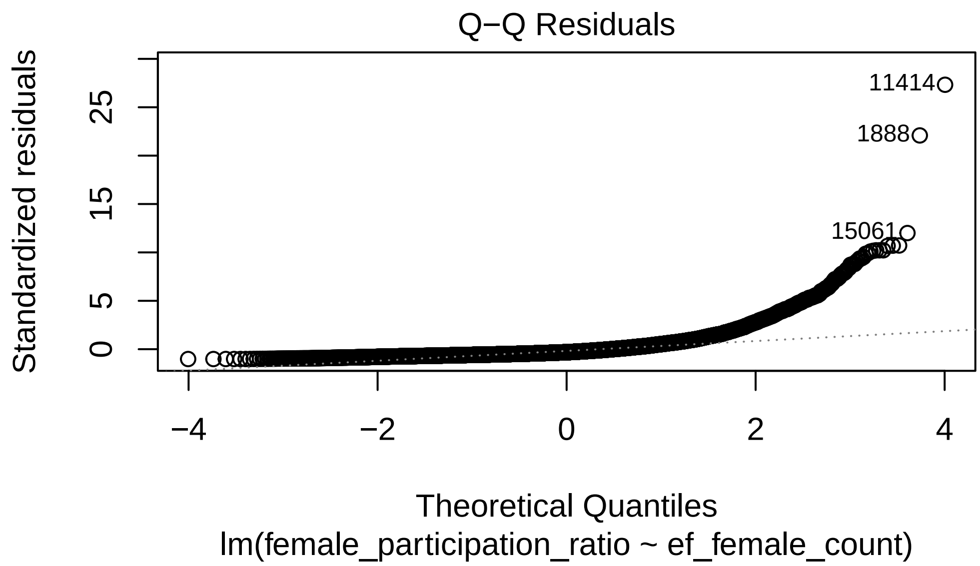
```
[1] -72766.08
```

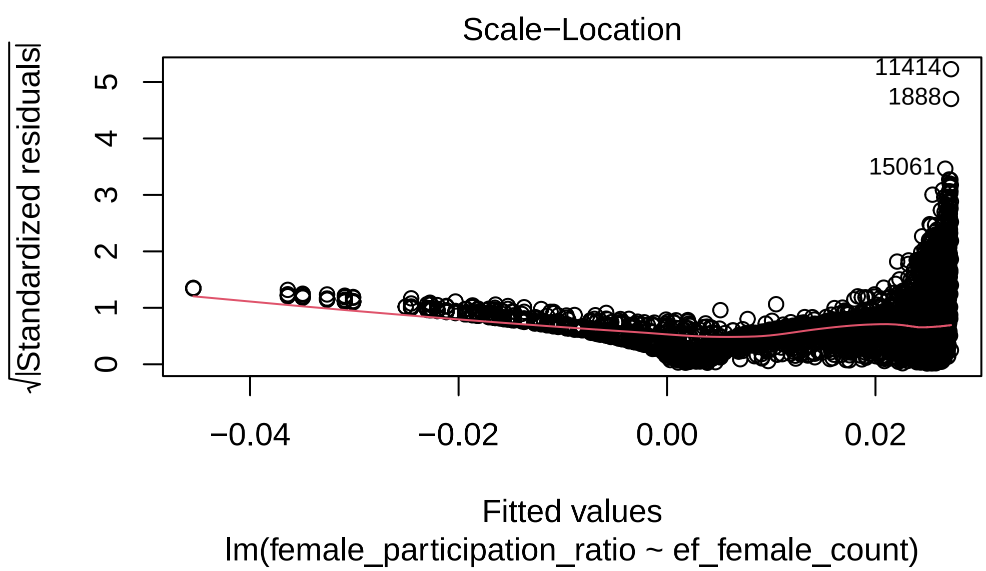
```r
BIC(hyp_1_fit_2)
```

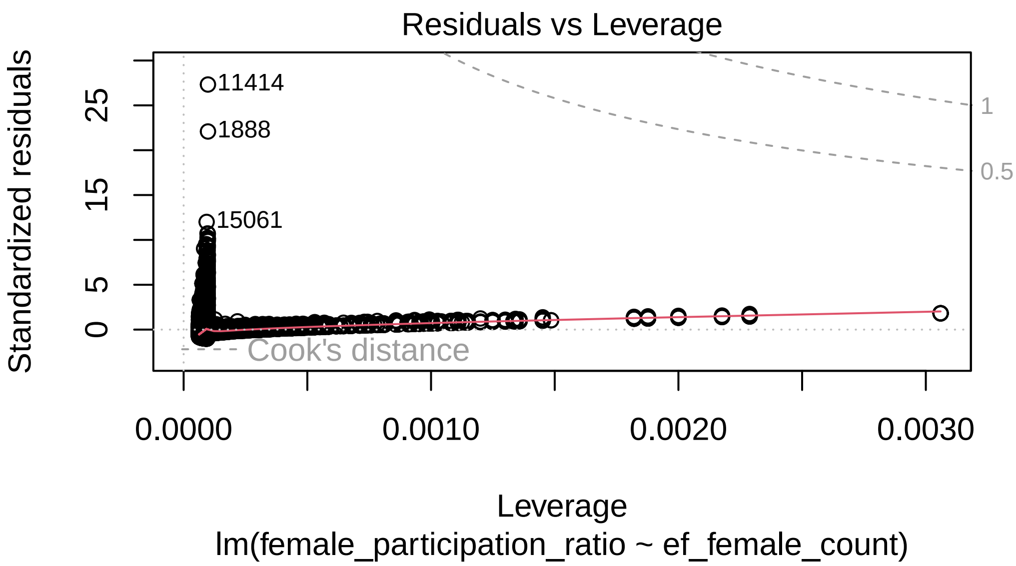
```
[1] -72743.02
```

```r
plot(hyp_1_fit_2)
```

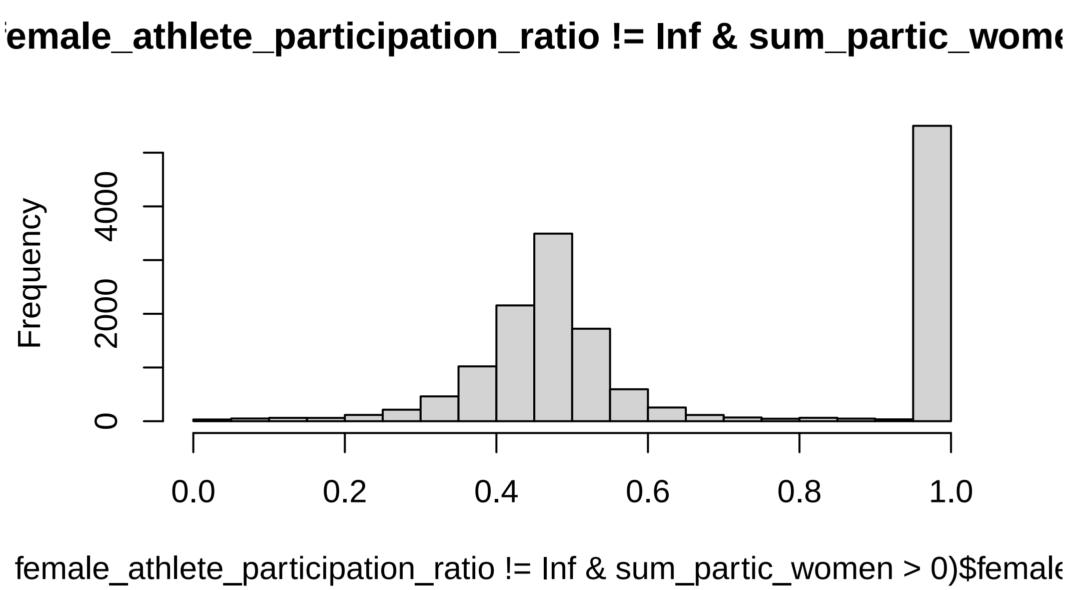



Q−Q Residuals
Standardized residuals
0
5
15
25
−4
−2
0
2
4
11414
1888
15061
Theoretical Quantiles
lm(female_participation_ratio ~ ef_female_count)


Scale−Location
√|Standardized residuals|
0
1
2
3
4
5
−0.04
−0.02
0.00
0.02
11414
1888
15061
Fitted values
lm(female_participation_ratio ~ ef_female_count)




```
hist(filter(attendance_data, female_athlete_participation_ratio != Inf & sum_partic_women
```



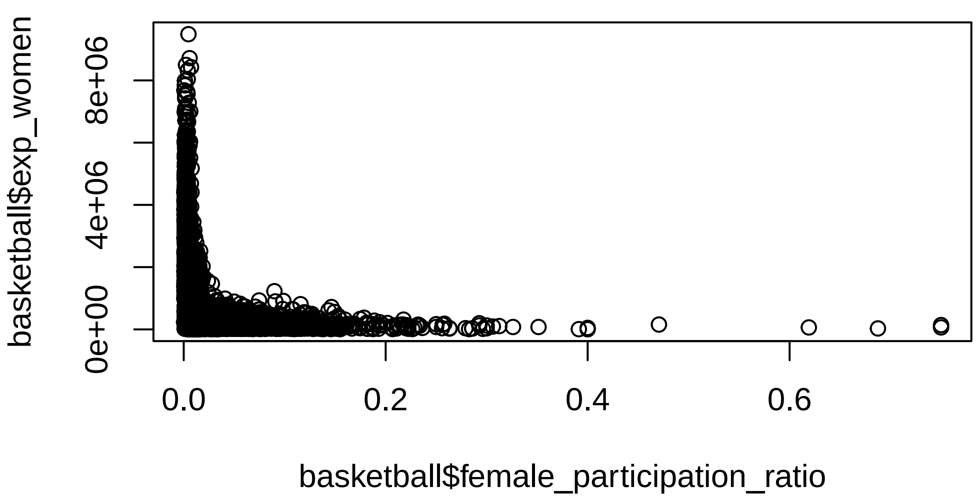
Hypothesis Test 2

```
basketball$female_participation_ratio <- basketball$sum_partic_women / basketball$ef_femal
basketball$female_athlete_participation_ratio <- basketball$sum_partic_women / (basketball

basketball$male_participation_ratio <- basketball$sum_partic_men / basketball$ef_male_coun
```

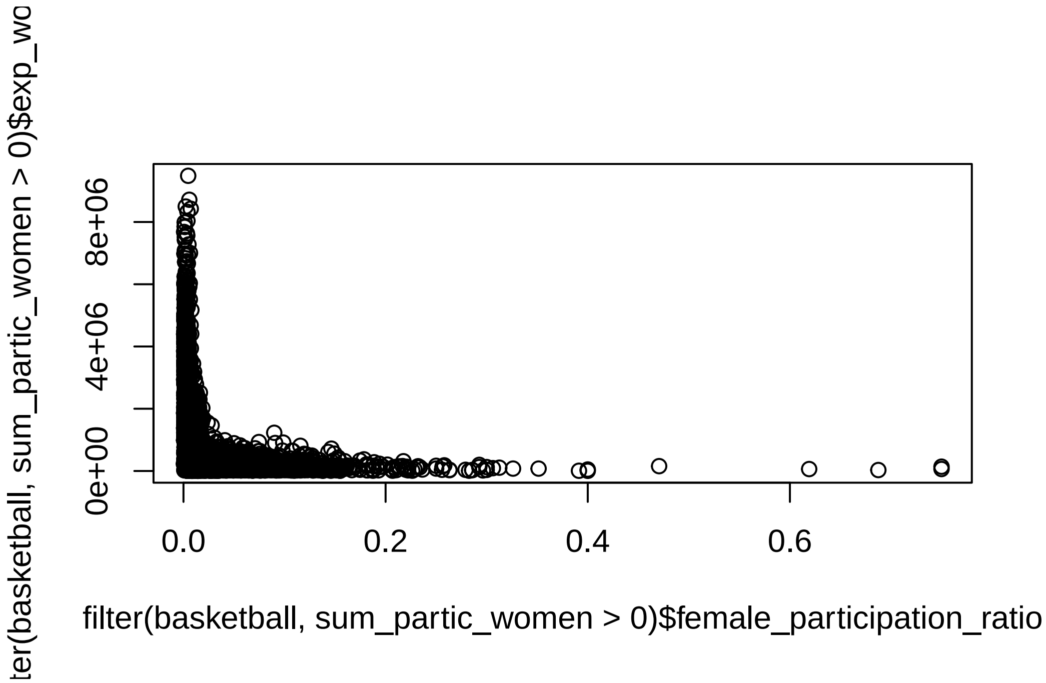
Transform basketball table to separate men and women by column

```
female <- as.data.frame(basketball[, c("year", "institution_name", "sports", "ef_female_co
female$gender <- "Female"
female <- female %>% rename("ef_count"="ef_female_count", "sum_partic"="sum_partic_women",

male <- as.data.frame(basketball[, c("year", "institution_name", "sports", "ef_male_count"
male$gender <- "Male"
male <- male %>% rename("ef_count"="ef_male_count", "sum_partic"="sum_partic_men", "rev"="
basketball_hist <- rbind(male, female)
basketball_hist <- filter(basketball_hist, sum_partic > 0)
```

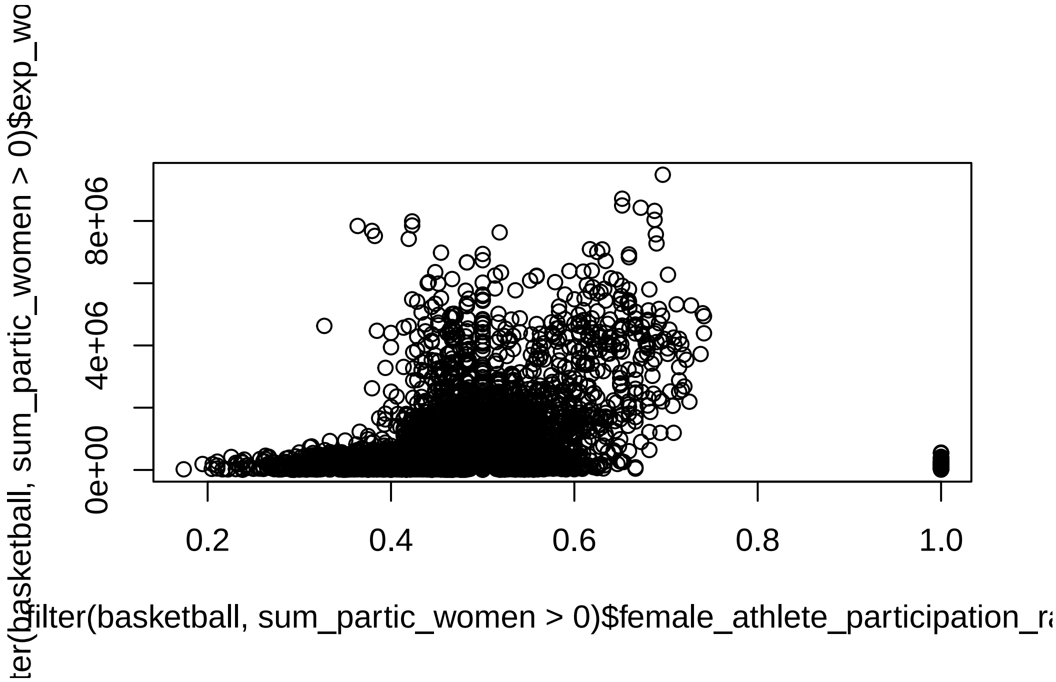
```
plot(basketball$female_participation_ratio, basketball$exp_women)
```



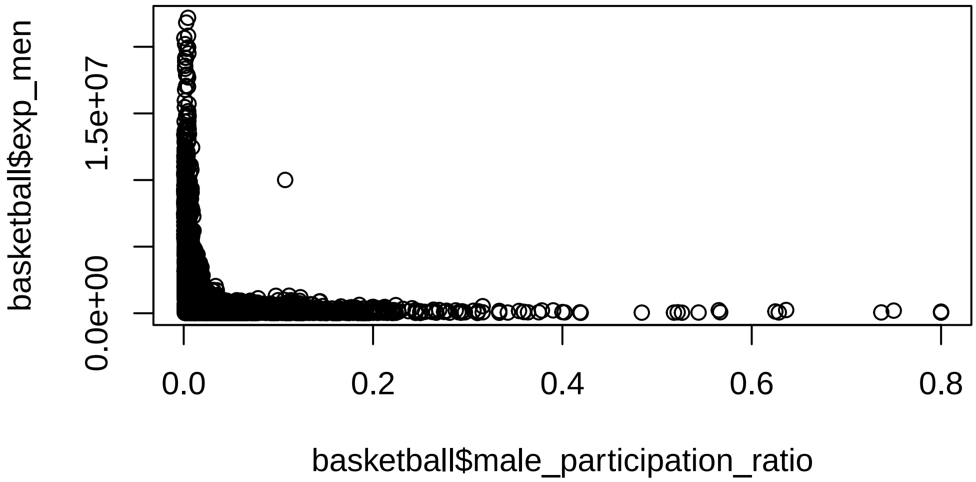
```
plot(filter(basketball, sum_partic_women > 0)$female_participation_ratio, filter(basketbal
```




```
plot(filter(basketball, sum_partic_women > 0)$female_athlete_participation_ratio, filter(b
```




```r
plot(basketball$male_participation_ratio, basketball$exp_men)
```



```r
hyp_2_fit_1 <- lm(exp_women ~ female_participation_ratio, data = filter(basketball, female
hyp_2_fit_2 <- lm(exp_women ~ female_participation_ratio, data = filter(basketball, female
hyp_2_fit_3 <- lm(exp_women ~ female_athlete_participation_ratio, data = filter(basketball
summary(hyp_2_fit_1)
```

```

Call:
lm(formula = exp_women ~ female_participation_ratio, data = filter(basketball,
    female_participation_ratio != Inf))

Residuals:
    Min      1Q  Median      3Q     Max
-669197 -493134 -320286   10208 8815838

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                  697788      11619   60.06   <2e-16 ***
female_participation_ratio -6077440     300720  -20.21   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 958400 on 9552 degrees of freedom
  (439 observations deleted due to missingness)
Multiple R-squared:  0.04101,   Adjusted R-squared:  0.0409
F-statistic: 408.4 on 1 and 9552 DF,  p-value: < 2.2e-16
```

```r
summary(hyp_2_fit_2)
```

```

Call:
lm(formula = exp_women ~ female_participation_ratio, data = filter(basketball,
    female_participation_ratio != Inf & sum_partic_women > 0))

Residuals:
    Min      1Q  Median      3Q     Max
-669197 -493134 -320286   10208 8815838

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                  697788      11619   60.06   <2e-16 ***
```

```
female_participation_ratio -6077440     300720  -20.21   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 958400 on 9552 degrees of freedom
Multiple R-squared:  0.04101,   Adjusted R-squared:  0.0409
F-statistic: 408.4 on 1 and 9552 DF,  p-value: < 2.2e-16
```

```r
summary(hyp_2_fit_3)
```

```
Call:
lm(formula = exp_women ~ female_athlete_participation_ratio,
    data = filter(basketball, female_participation_ratio != Inf &
        sum_partic_women > 0))

Residuals:
     Min       1Q   Median       3Q      Max
-2056112  -437273  -268066    15346  8284247

Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                         -785373      53101  -14.79   <2e-16 ***
female_athlete_participation_ratio  2852106     109722   25.99   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 945800 on 9552 degrees of freedom
Multiple R-squared:  0.06606,   Adjusted R-squared:  0.06597
F-statistic: 675.7 on 1 and 9552 DF,  p-value: < 2.2e-16
```

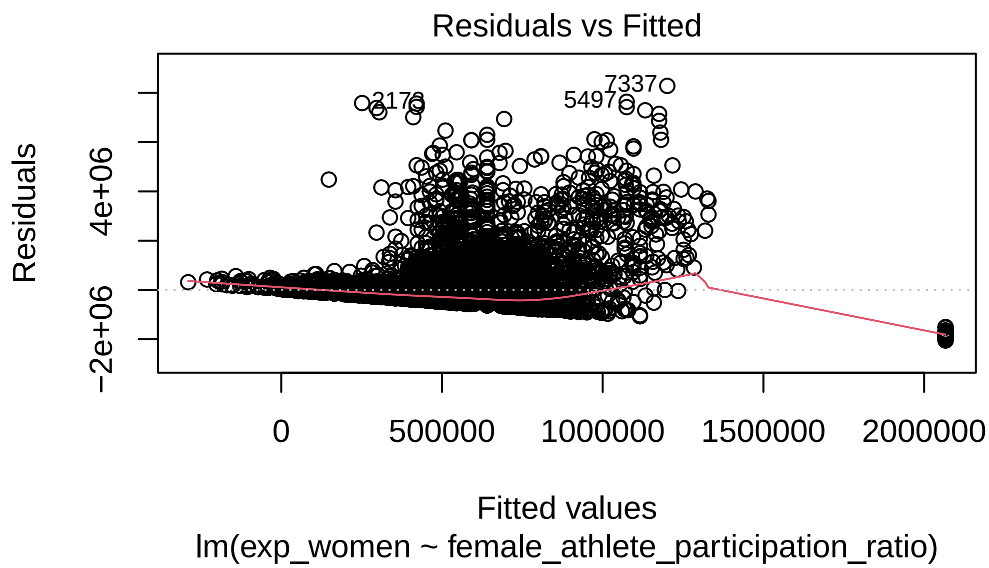
```r
AIC(hyp_2_fit_3)
```

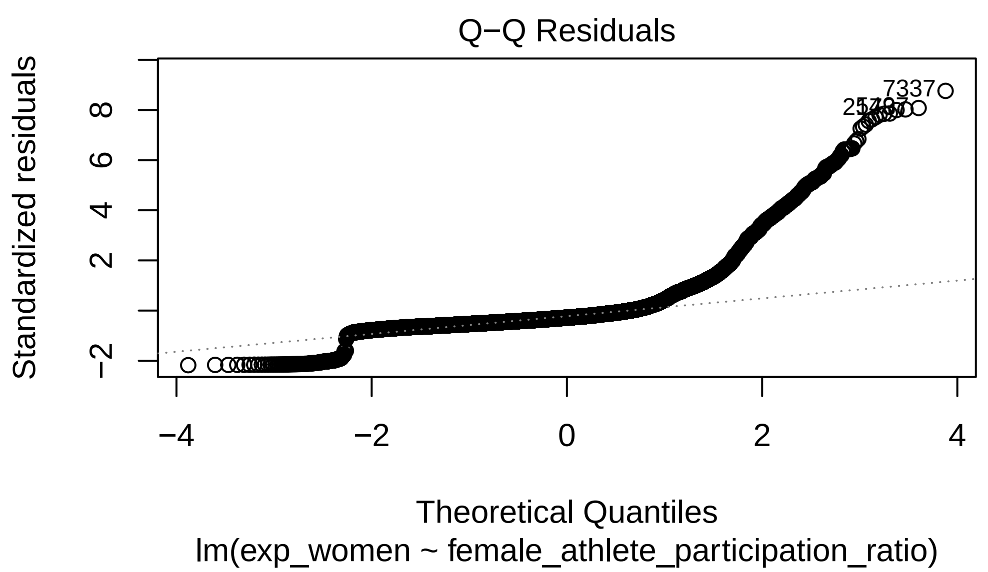
```
[1] 290039
```

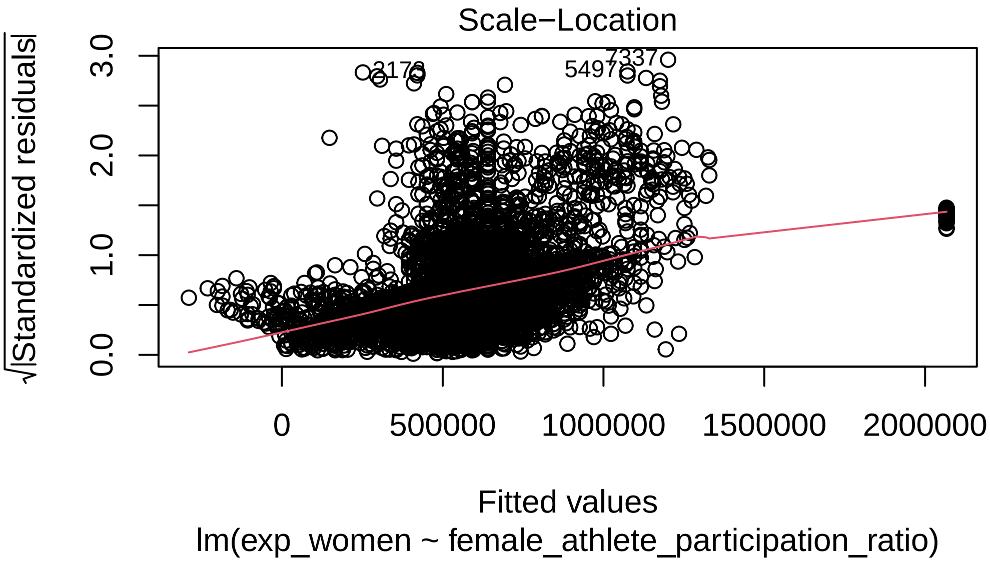
```r
BIC(hyp_2_fit_3)
```

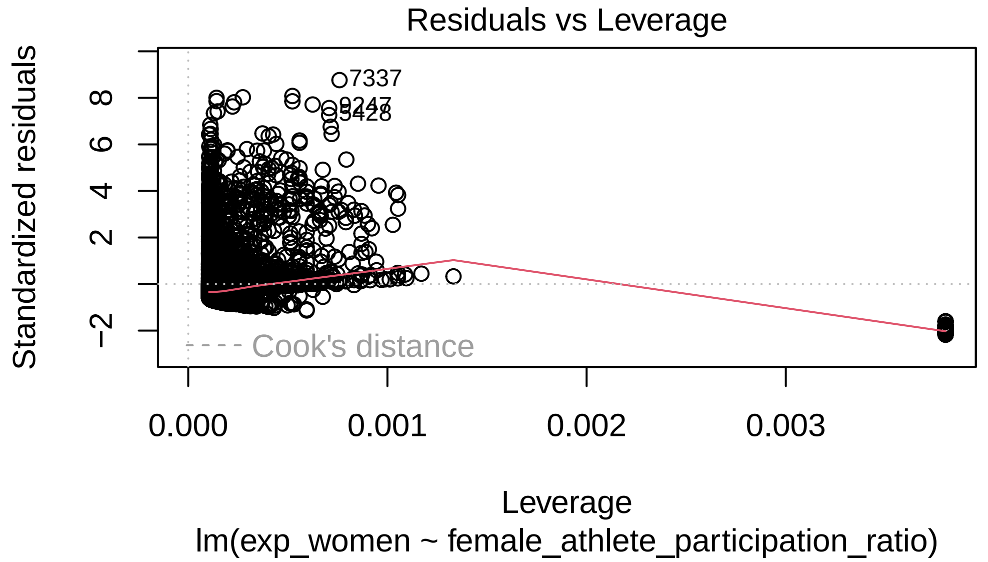
```
[1] 290060.5
```

```r
plot(hyp_2_fit_3)
```

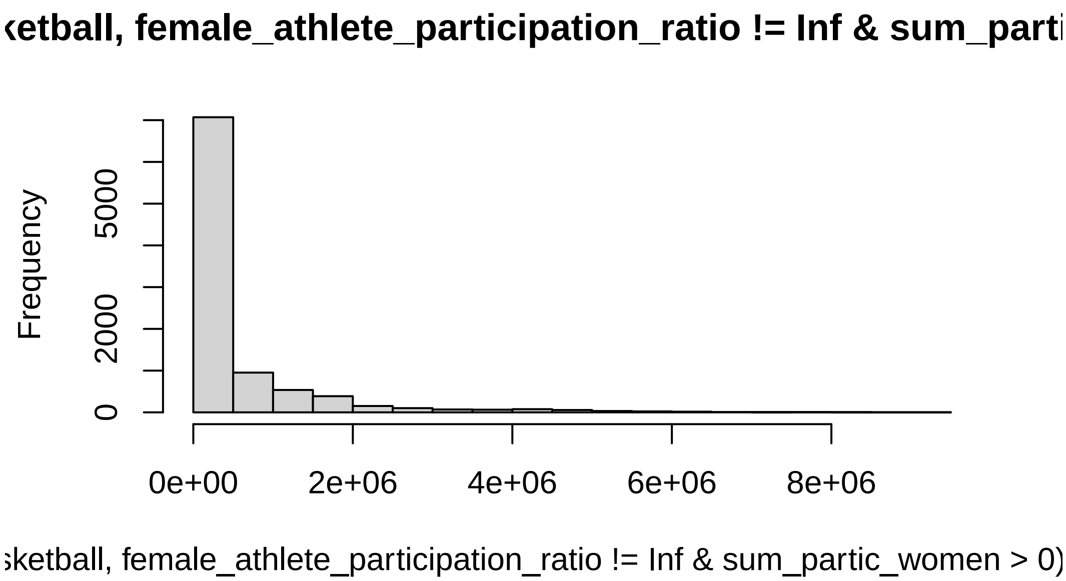




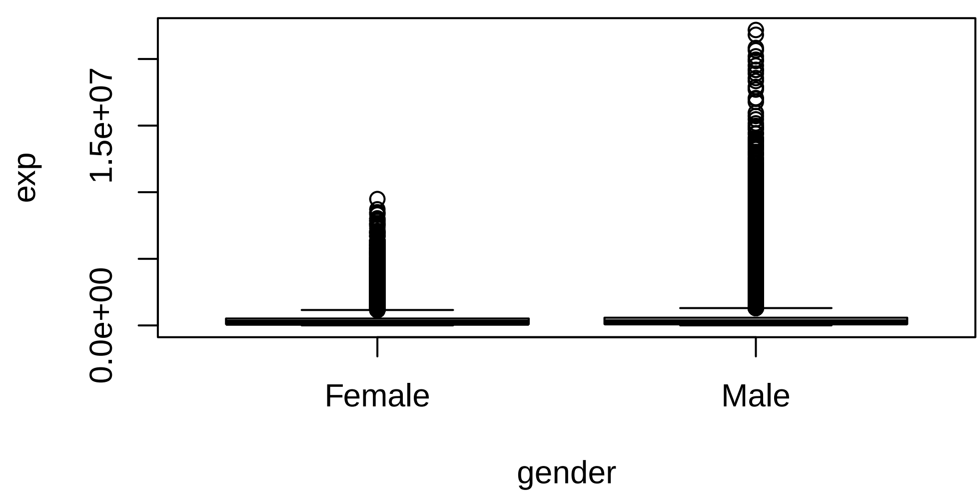





```
hist(filter(basketball, female_athlete_participation_ratio != Inf & sum_partic_women > 0)$
```

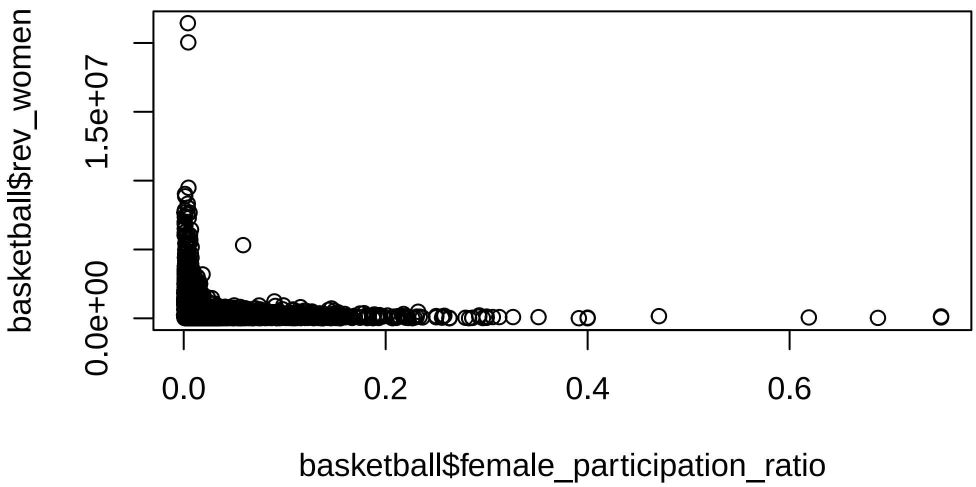



```
boxplot(exp ~ gender, data=basketball_hist)
```

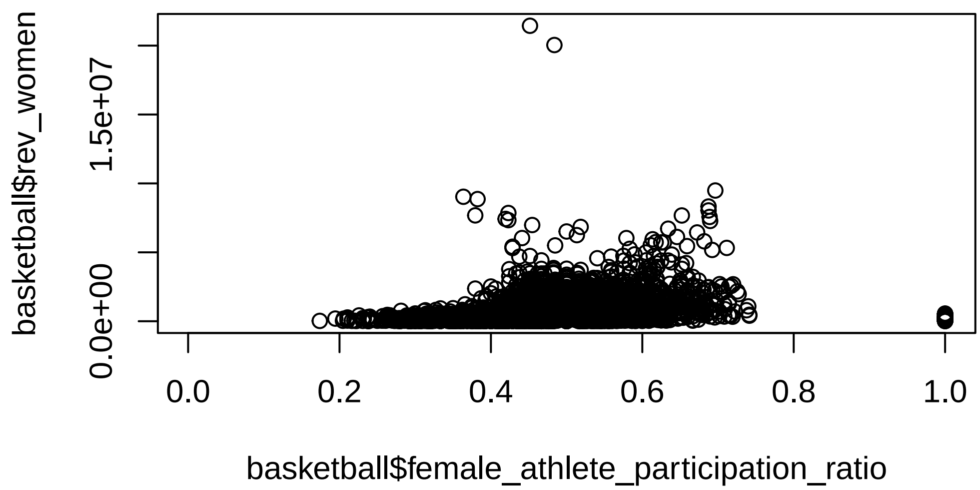


Hypothesis Test 3

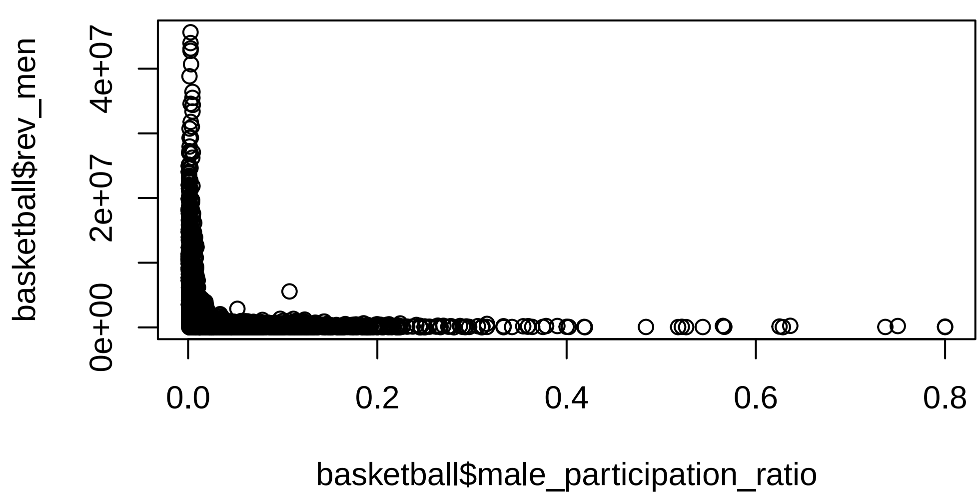
```r
plot(basketball$female_participation_ratio, basketball$rev_women)
```




```r
plot(basketball$female_athlete_participation_ratio, basketball$rev_women)
```




```r
plot(basketball$male_participation_ratio, basketball$rev_men)
```



```r
hyp_3_fit_1 <- lm(rev_women ~ female_participation_ratio, data = filter(basketball, female
hyp_3_fit_2 <- lm(rev_women ~ female_participation_ratio, data = filter(basketball, female
hyp_3_fit_3 <- lm(rev_women ~ female_athlete_participation_ratio, data = filter(basketball
summary(hyp_3_fit_1)
```

```
Call:
lm(formula = rev_women ~ female_participation_ratio, data = filter(basketball,
    female_participation_ratio != Inf))

Residuals:
     Min       1Q   Median       3Q      Max
 -546362  -388437  -240863    59195 20887546

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                  570845       9308   61.33   <2e-16 ***
female_participation_ratio -4393071     240902  -18.24   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 767800 on 9552 degrees of freedom
  (439 observations deleted due to missingness)
Multiple R-squared:  0.03364,   Adjusted R-squared:  0.03354
F-statistic: 332.5 on 1 and 9552 DF,  p-value: < 2.2e-16
```

```r
summary(hyp_3_fit_2)
```

```
Call:
lm(formula = rev_women ~ female_participation_ratio, data = filter(basketball,
    female_participation_ratio != Inf & sum_partic_women > 0))

Residuals:
     Min       1Q   Median       3Q      Max
 -546362  -388437  -240863    59195 20887546

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                  570845       9308   61.33   <2e-16 ***
```

```
female_participation_ratio -4393071     240902  -18.24   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 767800 on 9552 degrees of freedom
Multiple R-squared:  0.03364,   Adjusted R-squared:  0.03354
F-statistic: 332.5 on 1 and 9552 DF,  p-value: < 2.2e-16
```

```r
summary(hyp_3_fit_3)
```

```
Call:
lm(formula = rev_women ~ female_athlete_participation_ratio,
    data = filter(basketball, female_participation_ratio != Inf))

Residuals:
     Min       1Q   Median       3Q      Max
-1407754  -348702  -220354    43044 21003989

Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                         -372330      42943   -8.67   <2e-16 ***
female_athlete_participation_ratio  1790705      88733   20.18   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 764900 on 9552 degrees of freedom
  (439 observations deleted due to missingness)
Multiple R-squared:  0.04089,   Adjusted R-squared:  0.04079
F-statistic: 407.3 on 1 and 9552 DF,  p-value: < 2.2e-16
```

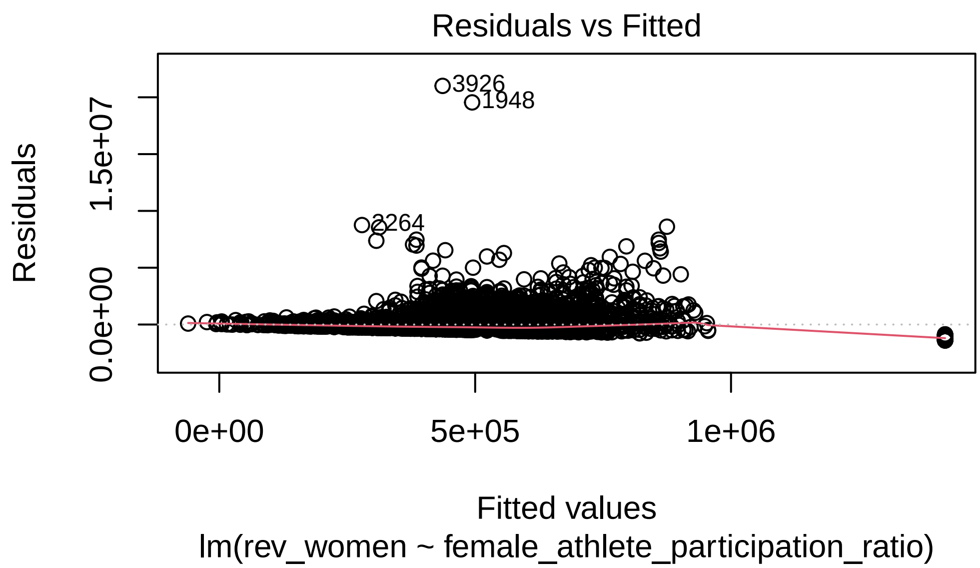
```r
AIC(hyp_3_fit_3)
```

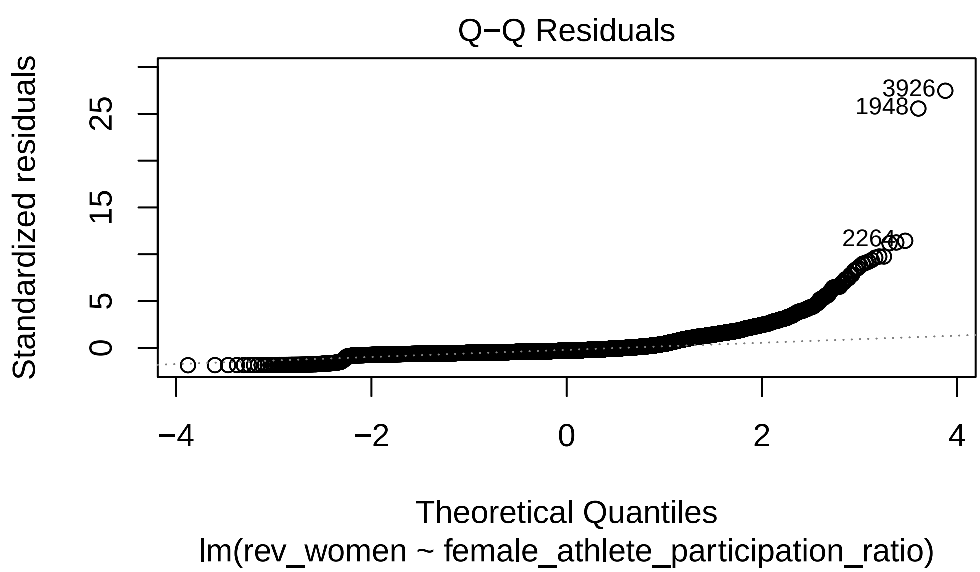
```
[1] 285982.1
```

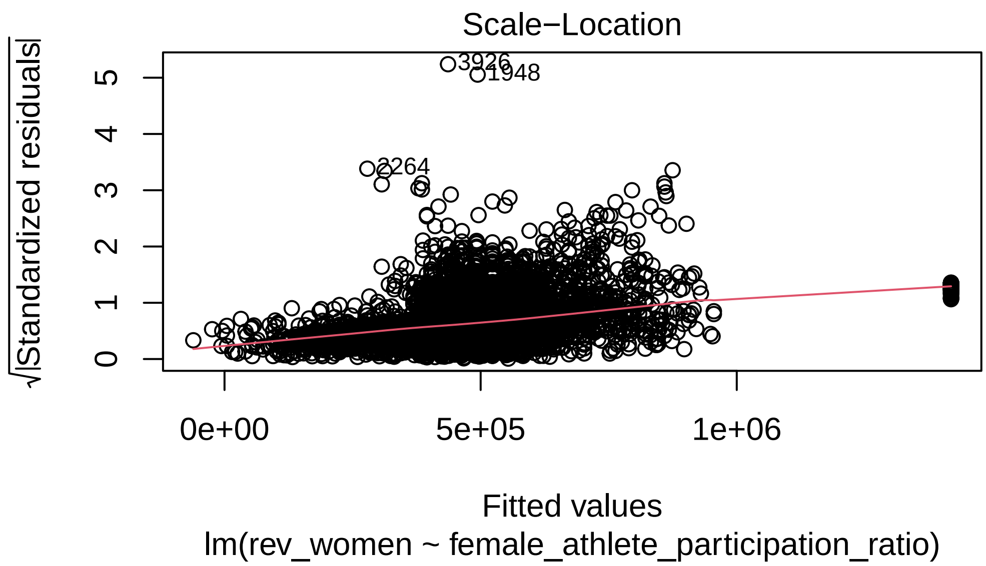
```r
BIC(hyp_3_fit_3)
```

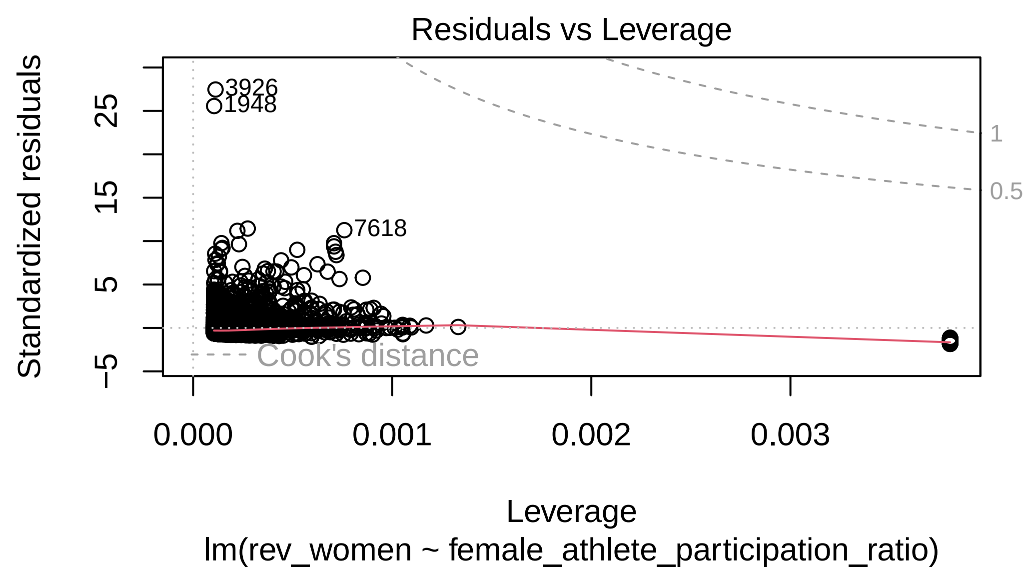
```
[1] 286003.6
```

```r
plot(hyp_3_fit_3)
```

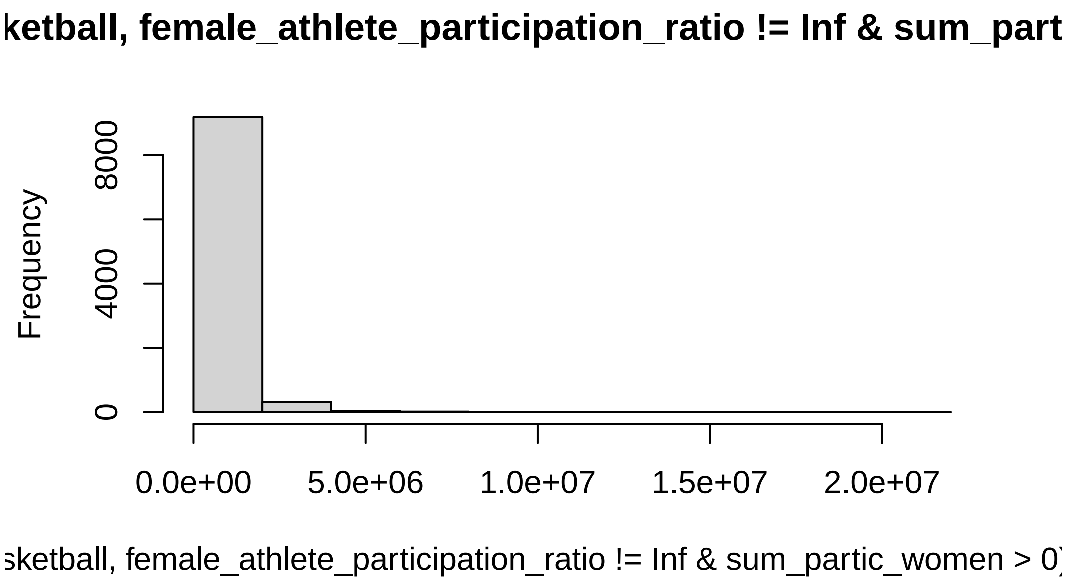




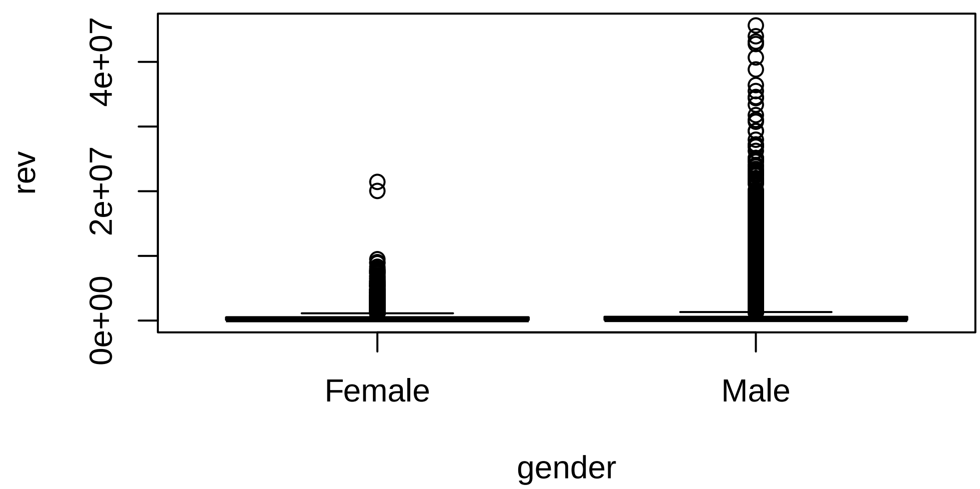





```
hist(filter(basketball, female_athlete_participation_ratio != Inf & sum_partic_women > 0)$
```




```
boxplot(rev ~ gender, data=basketball_hist)
```



My critical variables of interest are the following items:

- year: Period year
- institution name: School name
- sports: Sport name
- ef_male_count: Total male population
- ef_female_count: Total female population
- sum_partic_men: Total male participation
- sum_partic_women: Total female participation
- rev_men: Revenue in USD for men
- rev_women: Revenue in USD for women
- exp_men: Expenditures in USD for men
- exp_women: Expenditures in USD for women

Analysis:

For hypothesis 1, I added these new columns to the `attendance_data` data set:

1. female_participation_ratio
2. female_athlete_participation_ratio
3. male_participation_ratio

I used these metrics to test different approaches to measuring female participation at the collegial level to compare against males.

For hypotheses 2 & 3, I transformed the `basketball` data set to separate men and women by a new column `gender`, and also de-gendered the metrics to accommodate. The main reason was to use a histogram to better view data and compare gendered differences.

Model Comparisons and Diagnostics

Hypothesis 1 Models:

a. The first model wusednthe female participation ratio as the dependent and effective female count as the explanatory variable. The regression yielded .01809 for an R-Squared, denoting a low correlation between female participation to effective female count, thus indicating a failed hypothesis test.

b. The second model filters female participation greater thparticipationan 0. The R-Squared is at .1217, ; this a slight performance improvement but still is statistically insignificant. Thus, the hypothesis still fails on this test. However, in comparision to .013807 and .01809 the best performing model is in the second test and is what is chosen to represent the data set.

c. The third model is female athlete participation ratio (female participation divided by female and male participation) explained by ef_female_count. The third hypothesis 1 model shows slightly better performance at .013807 but still fails the hypothesis test.

Hypothesis 2 Models:

a. The second model measures the expenditure as a dependent and female participation as an explanatory. The R-Squared is .0409.

b. We see the same R-Squared in a and b due to the filter not removing the used observations.

c. I then use expenditures by the female athlete participation ratio ; we the R-Squared at .0657. Due to .0657 still being higher than the other R-Squared, , we use this as the model comparisons. However, we still reject this hypothesis.

Hypothesis 3 Models:

a. The third model measures the revenue as a dependent and female participation as an explanatory. The R-Squared is 0.03354.

b. We see the same R-Squared in a and b due to the filter not removing the used observations.

c. I then use revenue by the female athlete participation ratio; we the R-Squared at 0.04079. Due to 0.04079 still being higher than the other R-Squared, we use this as the model comparison. However, we still reject this hypothesis.

Interesting Plot Takeaways

For the boxplot comparing gender to revenue, we see that at the maximum, women make a quarter of the revenue. For the boxplot comparing gender to expenditures, we see that women are given half as much in funding for basketball.

Future Points of Project

I will add aes features to the plots and in the colors of basketball.

References

- Blinder, A. (2021, August 3). *Report: N.C.A.A. Prioritized Men's Basketball 'Over Everything Else.'* The New York Times. Retrieved April 12, 2023, from https://www.nytimes.com/2021/08/03/sports/ncaabasketball/ncaa-gender-equity-investigation.html?partner=slack&smid=sl-share.
- Feinberg, D., & Hunzinger, E. (2021, October 26). *Second NCAA Gender Equity Report Shows Spending Disparities.* US News. Retrieved April 11, 2023, from https://www.usnews.com/news/sports/articles/2021-10-26/second-ncaa-gender-equity-report-shows-spending-disparities#:~:text=The%20NCAA%20spent%20%244%2C285%20per,championsh
- Mock, J. T. (2022, March 29). *rfordatascience/tidytuesday.* GitHub. Retrieved April 10, 2023, from https://github.com/rfordatascience/tidytuesday/blob/master/data/2022/2022-03-29/readme.md.
- Zimbalist, A. (2022, October 12). *Female Athletes Are Undervalued, In Both Money And Media Terms.* Forbes. Retrieved April 12, 2023, from https://www.forbes.com/sites/andrewzimbalist/2019 athletes-are-undervalued-in-both-money-and-media-terms/?sh=5006015513ed.