VLAD.FOSTER

Vlad

2022-11-06

Set up Chunk

import the data

```
foster <- read.csv("/Users/monikavlad/Desktop/dataset/Fc2020v1.csv")
```

wrangling

cleaning up caretaker family structure

```
foster <- foster%>%
  filter(foster$ctkfamst %in% c("Married couple", "Unmarried couple", "Single female", "Single male"))
tab(foster, ctkfamst)
```

```
        ctkfamst |      Freq.   Percent      Cum.
-----------------+---------------------------------
  Married couple |    106863     18.22     18.22
   Single female |    291552     49.72     67.94
     Single male |     34852      5.94     73.88
Unmarried couple |    153176     26.12    100.00
```

```
group_by(foster, ctkfamst) %>%
  summarise(
    count = n(),
    mean = mean(lifelos, na.rm = TRUE),
    sd = sd(lifelos, na.rm = TRUE),
    median = median(lifelos, na.rm = TRUE),
    min = min(lifelos, na.rm = TRUE),
    max = max(lifelos, na.rm = TRUE)
  )
```

```
# A tibble: 4 × 7
  ctkfamst          count  mean    sd median   min   max
  <chr>             <int> <dbl> <dbl>  <dbl> <int> <int>
1 Married couple   106863  703.  709.    504     0  7666
2 Single female    291552  710.  703.    528     0  7670
3 Single male       34852  713.  688.    527     0  6678
4 Unmarried couple 153176  659.  652.    496     0  7666
```

race variable

```
foster <- foster%>%
  filter(foster$race %in% c("White", "Black or African American", "American
Indian or Alaska Native", "Asian", "Hawaiian or Other Pacific Islander",
"More Than One Race"))
tab(foster, race)
```

race	Freq.	Percent	Cum.
American Indian or Alaska Native	15743	2.78	2.78
Asian	3231	0.57	3.35
Black or African American	136170	24.04	27.39
Hawaiian or Other Pacific Islander	1898	0.34	27.73
More Than One Race	54412	9.61	37.33
White	354932	62.67	100.00

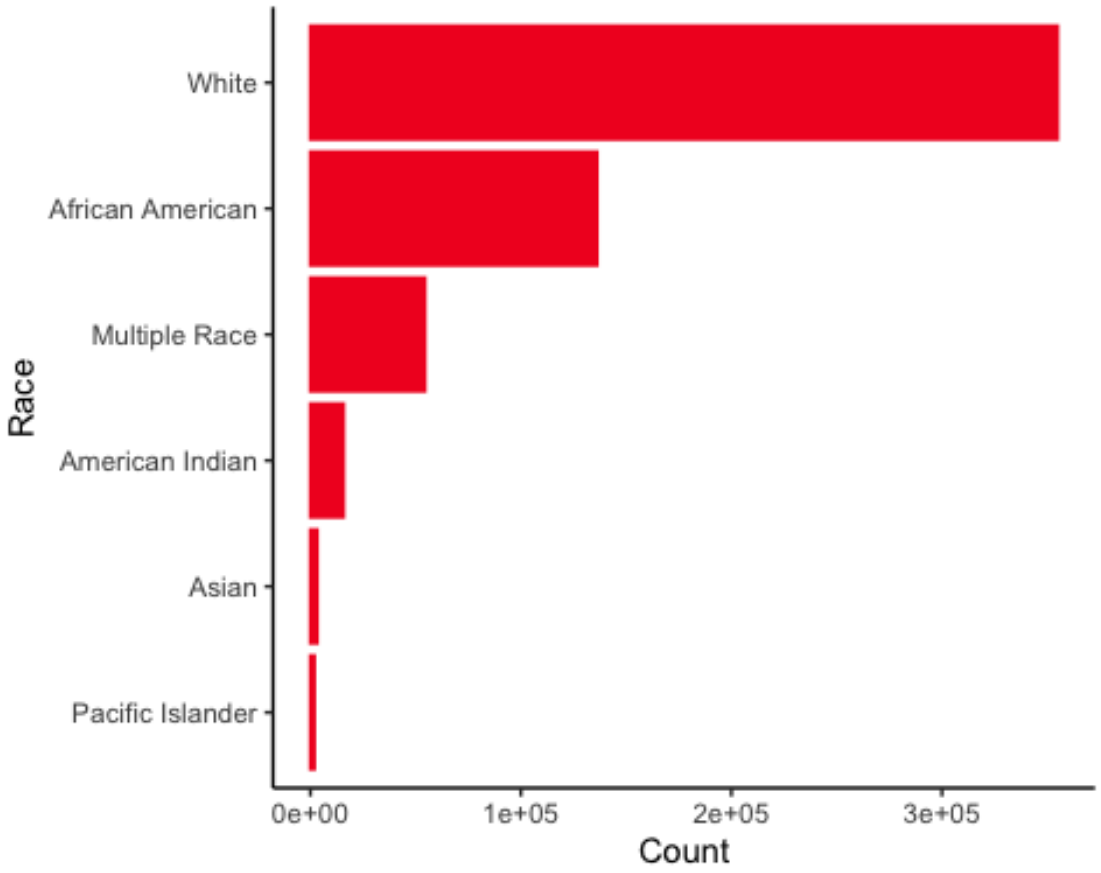
renaming names for race categories to preserve space on the poster

```
foster$race <- recode_factor(foster$race, "White"="White", "Black or African
American" = "African American", "American Indian or Alaska Native" =
"American Indian", "Asian" = "Asian", "Hawaiian or Other Pacific Islander" =
"Pacific Islander", "More Than One Race" = "Multiple Race")
levels(foster$race)
```

```
[1] "White"            "African American" "American Indian"  "Asian"
[5] "Pacific Islander" "Multiple Race"
```

plot for race

```
reorder_size <- function(race) {
  factor(race, levels = names(sort(table(race), decreasing = FALSE)))
}
ggplot(foster, aes(x = reorder_size(race))) +
  geom_bar(color = "#F21D25", fill= "#F21D25") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  coord_flip() +
  labs(y = "Count", x = "Race") +
  theme_classic()+
  theme(legend.position = "none")
```



cleaning up physically disabled variable

```
foster <- foster %>%
  filter(foster$phydis %in% c("Yes", "No"))
tab(foster, phydis)
```

```
phydis |      Freq.   Percent      Cum.
-------+--------------------------------
    No |    553699     99.01     99.01
   Yes |      5548      0.99    100.00
```

additional info about phydis

```
group_by(foster, phydis) %>%
  summarise(
    count = n(),
    mean = mean(lifelos, na.rm = TRUE),
    sd = sd(lifelos, na.rm = TRUE),
    median = median(lifelos, na.rm = TRUE),
    min = min(lifelos, na.rm = TRUE),
    max = max(lifelos, na.rm = TRUE)
  )
```

```
# A tibble: 2 × 7
  phydis  count  mean    sd median   min   max
  <chr>   <int> <dbl> <dbl>  <dbl> <int> <int>
1 No     553699  697.  682.    518     0  7670
2 Yes      5548 1149. 1119.    821     0  7480
```

cleaning up gender

```
foster <- foster %>%
  filter(foster$sex %in% c("Female", "Male"))
tab(foster, sex)


   sex |      Freq.   Percent      Cum.
-------+----------------------------------
Female |    272189     48.68     48.68
  Male |    286986     51.32    100.00
```

additional info about gender

```
group_by(foster, sex) %>%
  summarise(
    count = n(),
    mean = mean(lifelos, na.rm = TRUE),
    sd = sd(lifelos, na.rm = TRUE),
    median = median(lifelos, na.rm = TRUE),
    min = min(lifelos, na.rm = TRUE),
    max = max(lifelos, na.rm = TRUE)
  )

# A tibble: 2 × 7
  sex     count  mean    sd median   min   max
  <chr>   <int> <dbl> <dbl>  <dbl> <int> <int>
1 Female 272189  685.  662.    516     0  7670
2 Male   286986  716.  713.    526     0  7669
```

cleaning up clindis

```
foster <- foster %>%
  filter(foster$clindis %in% c("Yes", "No"))
tab(foster, clindis)


clindis |      Freq.   Percent      Cum.
--------+----------------------------------
     No |    353457     73.11     73.11
    Yes |    129987     26.89    100.00
```

additional info about clindis

```
group_by(foster, clindis) %>%
  summarise(
```

```
    count = n(),
    mean = mean(lifelos, na.rm = TRUE),
    sd = sd(lifelos, na.rm = TRUE),
    median = median(lifelos, na.rm = TRUE),
    min = min(lifelos, na.rm = TRUE),
    max = max(lifelos, na.rm = TRUE)
  )
```

```
# A tibble: 2 × 7
  clindis  count  mean    sd median   min   max
  <chr>    <int> <dbl> <dbl>  <dbl> <int> <int>
1 No      353457  634.  591.    489     0  7669
2 Yes     129987  991.  906.    741     0  7670
```

cleaning up mr

```
foster <- foster %>%
  filter(foster$mr %in% c("Yes", "No"))
tab(foster, mr)
```

```
 mr |      Freq.   Percent      Cum.
----+------------------------------
 No |    467022     96.60     96.60
Yes |     16422      3.40    100.00
```

additional info about mr

```
group_by(foster, mr) %>%
  summarise(
    count = n(),
    mean = mean(lifelos, na.rm = TRUE),
    sd = sd(lifelos, na.rm = TRUE),
    median = median(lifelos, na.rm = TRUE),
    min = min(lifelos, na.rm = TRUE),
    max = max(lifelos, na.rm = TRUE)
  )
```

```
# A tibble: 2 × 7
  mr      count  mean    sd median   min   max
  <chr>   <int> <dbl> <dbl>  <dbl> <int> <int>
1 No     467022  709.  677.    532     0  7670
2 Yes     16422 1242. 1135.    911     0  7663
```

cleaning up vishear

```
foster <- foster %>%
  filter(foster$vishear %in% c("Yes", "No"))
tab(foster, vishear)
```

```
vishear |      Freq.   Percent      Cum.
```

```
----------+-----------------------------------
       No |    470859    97.40    97.40
      Yes |     12584     2.60   100.00
```

additional info about vishear

```
group_by(foster, daparent) %>%
  summarise(
    count = n(),
    mean = mean(lifelos, na.rm = TRUE),
    sd = sd(lifelos, na.rm = TRUE),
    median = median(lifelos, na.rm = TRUE),
    min = min(lifelos, na.rm = TRUE),
    max = max(lifelos, na.rm = TRUE)
  )
```

```
# A tibble: 3 × 7
  daparent  count  mean    sd median   min   max
  <chr>     <int> <dbl> <dbl>  <dbl> <int> <int>
1 ""          975  770.  780.    510     0  5316
2 "No"     306873  777.  769.    561     0  7668
3 "Yes"    175595  641.  568.    511     0  7670
```

cleaning up PhyAbuse

```
foster <- foster %>%
  filter(foster$phyabuse %in% c("Yes", "No"))
tab(foster, phyabuse)
```

```
phyabuse |      Freq.  Percent     Cum.
---------+-----------------------------
      No |    419091    86.86    86.86
     Yes |     63377    13.14   100.00
```

cleaning up neglect

```
foster <- foster %>%
  filter(foster$neglect %in% c("Yes", "No"))
tab(foster, neglect)
```

```
neglect |      Freq.  Percent     Cum.
--------+-----------------------------
     No |    168006    34.82    34.82
    Yes |    314462    65.18   100.00
```

cleaning up manner of removal

```
foster <- foster %>%
  filter(foster$manrem %in% c("Court ordered", "Voluntary"))
tab(foster, manrem)
```

```
       manrem |      Freq.     Percent        Cum.
--------------+-----------------------------------
Court ordered |     467743       97.77       97.77
    Voluntary |      10674        2.23      100.00
```

view sexabuse

```
tab(foster, sexabuse)
```

```
sexabuse |      Freq.     Percent        Cum.
---------+-----------------------------------
      No |     458158       95.77       95.77
     Yes |      20259        4.23      100.00
```

view aaparent

```
tab(foster, aaparent)
```

```
aaparent |      Freq.     Percent        Cum.
---------+-----------------------------------
      No |     451809       94.44       94.44
     Yes |      26608        5.56      100.00
```

view daparent

```
tab(foster, daparent)
```

```
daparent |      Freq.     Percent        Cum.
---------+-----------------------------------
      No |     303679       63.48       63.48
     Yes |     174738       36.52      100.00
```

```
group_by(foster, daparent) %>%
  summarise(
    count = n(),
    mean = mean(lifelos, na.rm = TRUE),
    sd = sd(lifelos, na.rm = TRUE),
    median = median(lifelos, na.rm = TRUE),
    min = min(lifelos, na.rm = TRUE),
    max = max(lifelos, na.rm = TRUE)
  )
```

```
# A tibble: 2 × 7
  daparent  count  mean    sd median   min   max
  <chr>     <int> <dbl> <dbl>  <dbl> <int> <int>
1 No       303679  782.  770.    566     0  7668
2 Yes      174738  642.  569.    512     0  7670
```

view aachild

```
tab(foster, aachild)


aachild |      Freq.   Percent     Cum.
--------+-----------------------------
     No |    476802     99.66    99.66
    Yes |      1615      0.34   100.00
```

view dachild

```
tab(foster, dachild)


dachild |      Freq.   Percent     Cum.
--------+-----------------------------
     No |    469063     98.04    98.04
    Yes |      9354      1.96   100.00
```

view childis

```
tab(foster, childis)


childis |      Freq.   Percent     Cum.
--------+-----------------------------
     No |    469955     98.23    98.23
    Yes |      8462      1.77   100.00
```

```
group_by(foster, childis) %>%
  summarise(
    count = n(),
    mean = mean(lifelos, na.rm = TRUE),
    sd = sd(lifelos, na.rm = TRUE),
    median = median(lifelos, na.rm = TRUE),
    min = min(lifelos, na.rm = TRUE),
    max = max(lifelos, na.rm = TRUE)
  )
```

```
# A tibble: 2 × 7
  childis  count  mean    sd median   min   max
  <chr>    <int> <dbl> <dbl>  <int> <int> <int>
1 No      469955  726.  698.    542     0  7670
2 Yes       8462  969. 1010.    632     0  7077
```

view chbehprb

```
tab(foster, chbehprb)


chbehprb |      Freq.   Percent     Cum.
```

```
----------+-----------------------------------
       No |    442579    92.51    92.51
      Yes |     35838     7.49   100.00
```

view prtsjail

```
tab(foster, prtsjail)


prtsjail |      Freq.  Percent     Cum.
---------+-----------------------------------
       No |    447100    93.45    93.45
      Yes |     31317     6.55   100.00
```

view abandmnt

```
tab(foster, abandmnt)


abandmnt |      Freq.  Percent     Cum.
---------+-----------------------------------
       No |    454431    94.99    94.99
      Yes |     23986     5.01   100.00
```

view housing

```
tab(foster, housing)


housing |      Freq.  Percent     Cum.
--------+-----------------------------------
      No |    423779    88.58    88.58
     Yes |     54638    11.42   100.00
```

view LifeLOS

```
sum_up(foster, lifelos)


Variable |       Obs  Missing     Mean   StdDev       Min       Max
---------+-----------------------------------------------------------
 lifelos |    456278    22139  730.216  705.173         0      7670
```

Regression wrangling

creating a subset with only the variables we are using for analysis subset name -> foster.regression

```
foster.regression.variables <- c("lifelos", "phyabuse", "sexabuse",
"neglect", "aaparent", "daparent", "aachild", "dachild", "childis",
"chbehprb", "prtsjail", "abandmnt", "housing")
foster.regression <- foster[foster.regression.variables]
```

cleaning lifelos

```
foster.regression$lifelos <- as.numeric(foster.regression$lifelos)
foster.regression = foster.regression[!foster.regression$lifelos <=0,]
```

Multiple Linear Regression #Predicting the Length of Stay in FC for a child based on reasons of removal from the caretakers home #Dependent variable = lifelos= total number of days the child has been in foster care #Independent variables in the original model: #phyabuse: alleged or substantiated physical abuse, injury or maltreatment of the child by a person responsible for the childs welfare #sexabuse:alleged or substantiated sexual abuse or exploitation of a child by a person who is responsible for the childs welfare #neglect: alleged or substantiated negligent treatment or maltreatment, including failure to provide adequate food, clothing, shelter or care #AAparent: the principal caretaker`s compulsive use of alcohol that is not of temporary nature #DAparent: the principal caretaker's compulsive use of drugs that is not of temporary nature #AAchild: the child's compulsive use of or need for alcohol. Includes infants addicted at birth #DAChild: the child's used of drugs that is not of a temporary nature. Includes infants exposed to drugs during pregnancy #ChilDis: a clinical diagnosis from home and contact with the foster care system, a clinical diagnosis by a qualified professional of one or more of the following: mental retardation; emotional disturbance, specific learnign disability, hearing, speech or sight impairmnet; physical disability; or other clinically diagnosed handicap #ChiBehPrb: child's behavior in the schooland/or community that adversely affects socialization, learning, growth and moral development. #PrtsJail: temporary or permanent placement of a parent or caretaker in jail that adversely affects care of the child #Abandmnt: the child has been left alone or with others; caretaker did not return or make whereabouts known #Housing housing facilities were substandard, overcrowded, unsafe or otherwise inadequate resulting in their not being appropriate for the parents and child to resid etoegther. Also includes homelessness

regression model

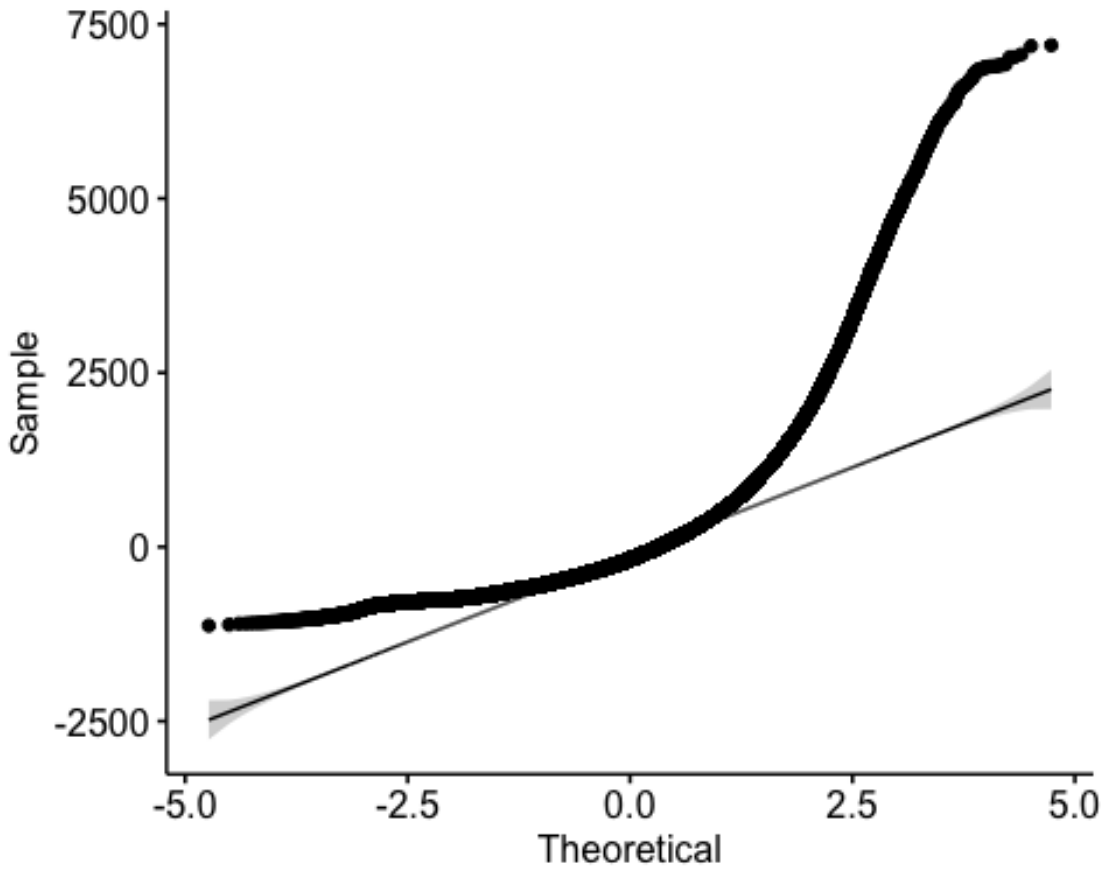
```
foster_model_1 <- lm(foster.regression$lifelos ~ foster.regression$phyabuse +
foster.regression$sexabuse + foster.regression$neglect  +
foster.regression$childis + foster.regression$aachild +
foster.regression$chbehprb  + foster.regression$abandmnt
+foster.regression$housing +  foster.regression$prtsjail +
foster.regression$daparent + foster.regression$aaparent  +
foster.regression$dachild )
summary(foster_model_1)


Call:
lm(formula = foster.regression$lifelos ~ foster.regression$phyabuse +
    foster.regression$sexabuse + foster.regression$neglect +
    foster.regression$childis + foster.regression$aachild +
foster.regression$chbehprb +
    foster.regression$abandmnt + foster.regression$housing +
    foster.regression$prtsjail + foster.regression$daparent +
    foster.regression$aaparent + foster.regression$dachild)
```

```
Residuals:
    Min      1Q  Median      3Q     Max 
-1128.6  -447.7  -176.2   227.8  7197.9 

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)    
(Intercept)                   740.506      2.347 315.539  < 2e-16 ***
foster.regression$phyabuseYes  25.245      3.115   8.104 5.34e-16 ***
foster.regression$sexabuseYes  47.476      5.175   9.174  < 2e-16 ***
foster.regression$neglectYes   30.686      2.278  13.470  < 2e-16 ***
foster.regression$childisYes  244.834      8.170  29.966  < 2e-16 ***
foster.regression$aachildYes   17.851     18.524   0.964   0.3352    
foster.regression$chbehprbYes  63.093      4.328  14.577  < 2e-16 ***
foster.regression$abandmntYes  46.740      4.820   9.696  < 2e-16 ***
foster.regression$housingYes   46.447      3.271  14.200  < 2e-16 ***
foster.regression$prtsjailYes   9.718      4.200   2.314   0.0207 *  
foster.regression$daparentYes -130.299      2.233 -58.357  < 2e-16 ***
foster.regression$aaparentYes   -8.009      4.542  -1.763   0.0779 .  
foster.regression$dachildYes  -160.110      7.768 -20.613  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 700.4 on 455830 degrees of freedom
  (22139 observations deleted due to missingness)
Multiple R-squared:  0.01345,   Adjusted R-squared:  0.01342 
F-statistic: 517.8 on 12 and 455830 DF,  p-value: < 2.2e-16
```

Predicted Length = 813.902 + 16.268(phyabuseYes) + 23.809(neglectYes) + 269.254(childisYes) + 20.630(abandmntYes) + 41.423(housingYes) - 11.858(prtsjailYes) - 118.059(daparentYes) - 30.343(aaparentYes) -98.432(dachildYes)

check normality assumption

```
foster_model_1$residuals %>%
  ggpubr::ggqqplot()
```



vif

```
car::vif(foster_model_1)
```

```
foster.regression$phyabuse foster.regression$sexabuse
                  1.034334                   1.014080
 foster.regression$neglect  foster.regression$childis
                  1.086287                   1.036542
 foster.regression$aachild foster.regression$chbehprb
                  1.041580                   1.121508
foster.regression$abandmnt  foster.regression$housing
                  1.012867                   1.021818
foster.regression$prtsjail foster.regression$daparent
                  1.010352                   1.082568
foster.regression$aaparent  foster.regression$dachild
                  1.006587                   1.066268
```

low vif lowest vif is one

testing variances

```
res <- var.test(lifelos ~ aaparent, data=foster.regression)
res
```

```
    F test to compare two variances

data:  lifelos by aaparent
F = 0.9485, num df = 430498, denom df = 25343, p-value = 5.481e-09
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.9320600 0.9650893
sample estimates:
ratio of variances
         0.9485038
```

cannot assume equal variances We ran a global test to see which variables are significant

t-test for significant variables

phyabuse summary

```
by(foster.regression$lifelos, foster.regression$phyabuse, summary)

foster.regression$phyabuse: No
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.0   281.0   545.0   725.9   947.0  7670.0
------------------------------------------------------------
foster.regression$phyabuse: Yes
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.0   274.0   539.0   763.9   977.0  7666.0
```

t-test for phyabuse

```
t.test(foster.regression$lifelos ~ foster.regression$phyabuse,
var.equal=FALSE)


    Welch Two Sample t-test

data:  foster.regression$lifelos by foster.regression$phyabuse
t = -11.328, df = 75589, p-value < 2.2e-16
alternative hypothesis: true difference in means between group No and group
Yes is not equal to 0
95 percent confidence interval:
 -44.53758 -31.39915
sample estimates:
 mean in group No mean in group Yes
         725.8941          763.8625
```

I am 95% confident that when a child has experienced physical abuse from their caretaker, the child stay 22 to 36 days longer in foster care than a child who has not been physically abused by their caretaker.

additional info for phyabuse

```
group_by(foster, phyabuse) %>%
  summarise(
    count = n(),
    mean = mean(lifelos, na.rm = TRUE),
    sd = sd(lifelos, na.rm = TRUE),
    median = median(lifelos, na.rm = TRUE),
    min = min(lifelos, na.rm = TRUE),
    max = max(lifelos, na.rm = TRUE)
  )

# A tibble: 2 × 7
  phyabuse  count  mean    sd median   min   max
  <chr>     <int> <dbl> <dbl>  <int> <int> <int>
1 No       415538  725.  693.    544     0  7670
2 Yes       62879  763.  777.    539     0  7666
```

sexabuse summary

```
by(foster.regression$lifelos, foster.regression$sexabuse, summary)

foster.regression$sexabuse: No
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.0   280.0   543.0   727.8   946.0  7670.0
------------------------------------------------------------
foster.regression$sexabuse: Yes
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.0   283.0   575.0   801.2  1057.0  7366.0
```

t-test for sexabuse

```
sexabuse.t <- t.test(foster.regression$lifelos ~ foster.regression$sexabuse,
var.equal=FALSE)
```

additional info for sexabuse

```
group_by(foster, sexabuse) %>%
  summarise(
    count = n(),
    mean = mean(lifelos, na.rm = TRUE),
    sd = sd(lifelos, na.rm = TRUE),
    median = median(lifelos, na.rm = TRUE),
    min = min(lifelos, na.rm = TRUE),
    max = max(lifelos, na.rm = TRUE)
  )

# A tibble: 2 × 7
  sexabuse  count  mean    sd median   min   max
  <chr>     <int> <dbl> <dbl>  <int> <int> <int>
1 No       458158  727.  701.    542     0  7670
2 Yes       20259  801.  781.    574     0  7366
```

I am 95% confident that when a child has experienced sexual abuse from their caretaker, the child stays 6 to 28 days longer in foster care than a child who has not been sexually abused by their caretaker

neglect summary

```
by(foster.regression$lifelos, foster.regression$neglect, summary)

foster.regression$neglect: No
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    1.0   264.0   531.0   713.7   940.0  7669.0 
------------------------------------------------------------ 
foster.regression$neglect: Yes
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    1.0   289.0   551.0   739.9   957.0  7670.0
```

t-test for neglect

```
neglect.t <- t.test(foster.regression$lifelos ~ foster.regression$neglect,
var.equal=FALSE)
```

additional info for neglect

```
group_by(foster, neglect) %>%
  summarise(
    count = n(),
    mean = mean(lifelos, na.rm = TRUE),
    sd = sd(lifelos, na.rm = TRUE),
    median = median(lifelos, na.rm = TRUE),
    min = min(lifelos, na.rm = TRUE),
    max = max(lifelos, na.rm = TRUE)
  )

# A tibble: 2 × 7
  neglect  count  mean    sd median   min   max
  <chr>    <int> <dbl> <dbl>  <int> <int> <int>
1 No      166724  713.  691.    531     0  7669
2 Yes     311693  739.  712.    550     0  7670
```

I am 95% confident that when a child has been removed from home due to neglect from the caretaker, the child stays 21 to 30 days longer in foster care than when the child has not been neglected

childis summary

```
by(foster.regression$lifelos, foster.regression$childis, summary)

foster.regression$childis: No
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    1.0   279.0   543.0   726.8   947.0  7670.0 
------------------------------------------------------------ 
foster.regression$childis: Yes
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    1.0   310.0   632.0   969.7  1253.0  7077.0 
```

t-test for childis

```
childis.t <- t.test(foster.regression$lifelos ~ foster.regression$childis, var.equal=FALSE)
```

I am 95% confident that a child who has been clinically diagnosed with a disability stays 246 to 296 days longer in foster care than a child who was diagnosed to not have any clinical disabilities additional info for childis

```
group_by(foster, daparent) %>%
  summarise(
    count = n(),
    mean = mean(lifelos, na.rm = TRUE),
    sd = sd(lifelos, na.rm = TRUE),
    median = median(lifelos, na.rm = TRUE),
    min = min(lifelos, na.rm = TRUE),
    max = max(lifelos, na.rm = TRUE)
  )

# A tibble: 2 × 7
  daparent  count  mean    sd median   min   max
  <chr>     <int> <dbl> <dbl>  <dbl> <int> <int>
1 No       303679  782.  770.    566     0  7668
2 Yes      174738  642.  569.    512     0  7670
```

chbehprb summary

```
by(foster.regression$lifelos, foster.regression$chbehprb, summary)

foster.regression$chbehprb: No
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    1.0   282.0   543.0   724.6   941.0  7670.0 
------------------------------------------------------------ 
foster.regression$chbehprb: Yes
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    1.0   267.0   569.0   815.9  1108.0  7668.0 
```

t-test for chbehprb

```
chbehprb.t <- t.test(foster.regression$lifelos ~ foster.regression$chbehprb, var.equal=FALSE)
```

I am 95% confident that when the child has behavior problems, the child stays 14 to 33 days longer in foster care than a child who does not have behavioral problems.

additional info for daparent

```
group_by(foster, daparent) %>%
  summarise(
```

```
    count = n(),
    mean = mean(lifelos, na.rm = TRUE),
    sd = sd(lifelos, na.rm = TRUE),
    median = median(lifelos, na.rm = TRUE),
    min = min(lifelos, na.rm = TRUE),
    max = max(lifelos, na.rm = TRUE)
  )

# A tibble: 2 × 7
  daparent  count  mean    sd median   min   max
  <chr>     <int> <dbl> <dbl>  <dbl> <int> <int>
1 No       303679  782.  770.    566     0  7668
2 Yes      174738  642.  569.    512     0  7670
```

abandmnt summary

```
by(foster.regression$lifelos, foster.regression$abandmnt, summary)

foster.regression$abandmnt: No
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.0   278.0   541.0   727.8   945.0  7670.0
------------------------------------------------------------
foster.regression$abandmnt: Yes
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.0   320.0   609.0   791.6  1055.0  7358.0
```

t-test for abandmnt

```
abandmnt.t <- t.test(foster.regression$lifelos ~ foster.regression$abandmnt,
var.equal=FALSE)
```

I am 95% confident that when the child has been abandoned by the caretaker, the child stays 27 to 47 days longer in foster care than the child whose caretaker did not abandon them.

additional info for abandmnt

```
group_by(foster, abandmnt) %>%
  summarise(
    count = n(),
    mean = mean(lifelos, na.rm = TRUE),
    sd = sd(lifelos, na.rm = TRUE),
    median = median(lifelos, na.rm = TRUE),
    min = min(lifelos, na.rm = TRUE),
    max = max(lifelos, na.rm = TRUE)
  )

# A tibble: 2 × 7
  abandmnt  count  mean    sd median   min   max
  <chr>     <int> <dbl> <dbl>  <dbl> <int> <int>
1 No       454431  727.  705.    540     0  7670
2 Yes       23986  791.  713.    608     0  7358
```

housing summary

```
by(foster.regression$lifelos, foster.regression$housing, summary)

foster.regression$housing: No
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    1.0   273.0   539.0   728.5   947.0  7670.0 
------------------------------------------------------------ 
foster.regression$housing: Yes
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    1.0   327.0   587.0   749.3   973.0  7546.0
```

t-test for housing

```
housing.t <- t.test(foster.regression$lifelos ~ foster.regression$housing, 
var.equal=FALSE)
```

I am 95% confident that when the child's caretaker failed to provide safe housing for the child, the child stays 13 to 27 days longer in foster care than when the carektaker provided safe housing for the child

additional info for housing

```
group_by(foster, housing) %>%
  summarise(
    count = n(),
    mean = mean(lifelos, na.rm = TRUE),
    sd = sd(lifelos, na.rm = TRUE),
    median = median(lifelos, na.rm = TRUE),
    min = min(lifelos, na.rm = TRUE),
    max = max(lifelos, na.rm = TRUE)
  )

# A tibble: 2 × 7
  housing  count  mean    sd median   min   max
  <chr>    <int> <dbl> <dbl>  <int> <int> <int>
1 No      423779  728.  710.    538     0  7670
2 Yes      54638  749.  665.    586     0  7546
```

prtsjail summary t-test for prtsjail

```
by(foster.regression$lifelos, foster.regression$prtsjail, summary)

foster.regression$prtsjail: No
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    1.0   278.0   544.0   731.6   952.0  7670.0 
------------------------------------------------------------ 
foster.regression$prtsjail: Yes
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
    1.0   303.0   558.0   721.6   937.0  7666.0
```

t-test for prtsjail

```
t.test(foster.regression$lifelos ~ foster.regression$prtsjail,
var.equal=FALSE)


    Welch Two Sample t-test

data:  foster.regression$lifelos by foster.regression$prtsjail
t = 2.5675, df = 35429, p-value = 0.01025
alternative hypothesis: true difference in means between group No and group
Yes is not equal to 0
95 percent confidence interval:
  2.352521 17.532420
sample estimates:
 mean in group No mean in group Yes
         731.5688          721.6263
```

I am 95% confident that when the child's caretaker has been placed in jail(temporary or permanently), the child stays 22 to 38 days less in foster care than when the caretaker has not been placed in jail. I am 95% confident that when child's caretaker has not been placed in jail(temporary or permanatly), the child stays 22 to 38 days longer in foster care than when the caretaker has been placed in jail. daparent summary

additional info for prtsjail

```
group_by(foster, prtsjail) %>%
  summarise(
    count = n(),
    mean = mean(lifelos, na.rm = TRUE),
    sd = sd(lifelos, na.rm = TRUE),
    median = median(lifelos, na.rm = TRUE),
    min = min(lifelos, na.rm = TRUE),
    max = max(lifelos, na.rm = TRUE)
  )

# A tibble: 2 × 7
  prtsjail  count  mean    sd median   min   max
  <chr>     <int> <dbl> <dbl>  <dbl> <int> <int>
1 No       447100  731.  709.    543     0  7670
2 Yes       31317  721.  645.    557     0  7666
```

t-test for daparent

```
by(foster.regression$lifelos, foster.regression$daparent, summary)

foster.regression$daparent: No
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
      1     283     566     783    1029    7668
------------------------------------------------------------
foster.regression$daparent: Yes
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.0   276.0   512.0   643.1   847.0  7670.0
```

t-test for daparent

```
daparent.t <- t.test(foster.regression$lifelos ~ foster.regression$daparent,
var.equal=FALSE)
```

I am 95% confident that when the child`s caretaker has a drug abuse problem, the child stays 115 to 123 days less in foster care than when the caretaker is not a drug abuser. I am 95% confident that whne the child's caretaker does not have a drug abuse problem, the child stays 115 to 123 days longer in foster care than when the caretaker is a drug abuser

additional info for daparent

```
group_by(foster, daparent) %>%
  summarise(
    count = n(),
    mean = mean(lifelos, na.rm = TRUE),
    sd = sd(lifelos, na.rm = TRUE),
    median = median(lifelos, na.rm = TRUE),
    min = min(lifelos, na.rm = TRUE),
    max = max(lifelos, na.rm = TRUE)
  )

# A tibble: 2 × 7
  daparent  count  mean    sd median   min   max
  <chr>     <int> <dbl> <dbl>  <dbl> <int> <int>
1 No       303679  782.  770.    566     0  7668
2 Yes      174738  642.  569.    512     0  7670
```

aaparent summary

```
by(foster.regression$lifelos, foster.regression$aaparent, summary)

foster.regression$aaparent: No
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.0   282.0   546.0   731.8   952.0  7670.0
------------------------------------------------------------
foster.regression$aaparent: Yes
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.0   259.0   509.0   716.5   924.0  7669.0
```

t-test for aaaparent

```
aaparent.t <- t.test(foster.regression$lifelos ~ foster.regression$aaparent,
var.equal=FALSE)
```

I am 95% confident that when the child's caretaker has an alcohol abuse problem, the child stays 27 to 47 days less in foster care than when the caretaker is not an alcohol abuser. I am 95% confident that when the child's caretaker does not have an alcohol abuse problem, the child stays 27 to 47 days longer in foster care than when the caretaker is an alcohol abuser

additional info for aaparent

```
group_by(foster, aaparent) %>%
  summarise(
    count = n(),
    mean = mean(lifelos, na.rm = TRUE),
    sd = sd(lifelos, na.rm = TRUE),
    median = median(lifelos, na.rm = TRUE),
    min = min(lifelos, na.rm = TRUE),
    max = max(lifelos, na.rm = TRUE)
  )

# A tibble: 2 × 7
  aaparent  count  mean    sd median   min   max
  <chr>     <int> <dbl> <dbl>  <int> <int> <int>
1 No       451809  731.  704.    545     0  7670
2 Yes       26608  716.  723.    508     0  7669
```

dachild summary t-test for dachild

```
by(foster.regression$lifelos, foster.regression$dachild, summary)

foster.regression$dachild: No
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.0   281.0   546.0   733.4   954.0  7670.0
------------------------------------------------------------
foster.regression$dachild: Yes
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.0   251.0   476.0   605.3   789.0  7668.0
```

t-test for dachild

```
dachild.t <- t.test(foster.regression$lifelos ~ foster.regression$dachild,
var.equal=FALSE)
```

I am 95% confident that when the child has a drug problem, the child stays 63 to 96 days less in foster care than a child who does not have a drug abuse problem. I am 95% confident that when the child does not have a drug problem, the child stays 63 to 96 days longer in foster care than a child who has a drug problem.

additional info for dachild

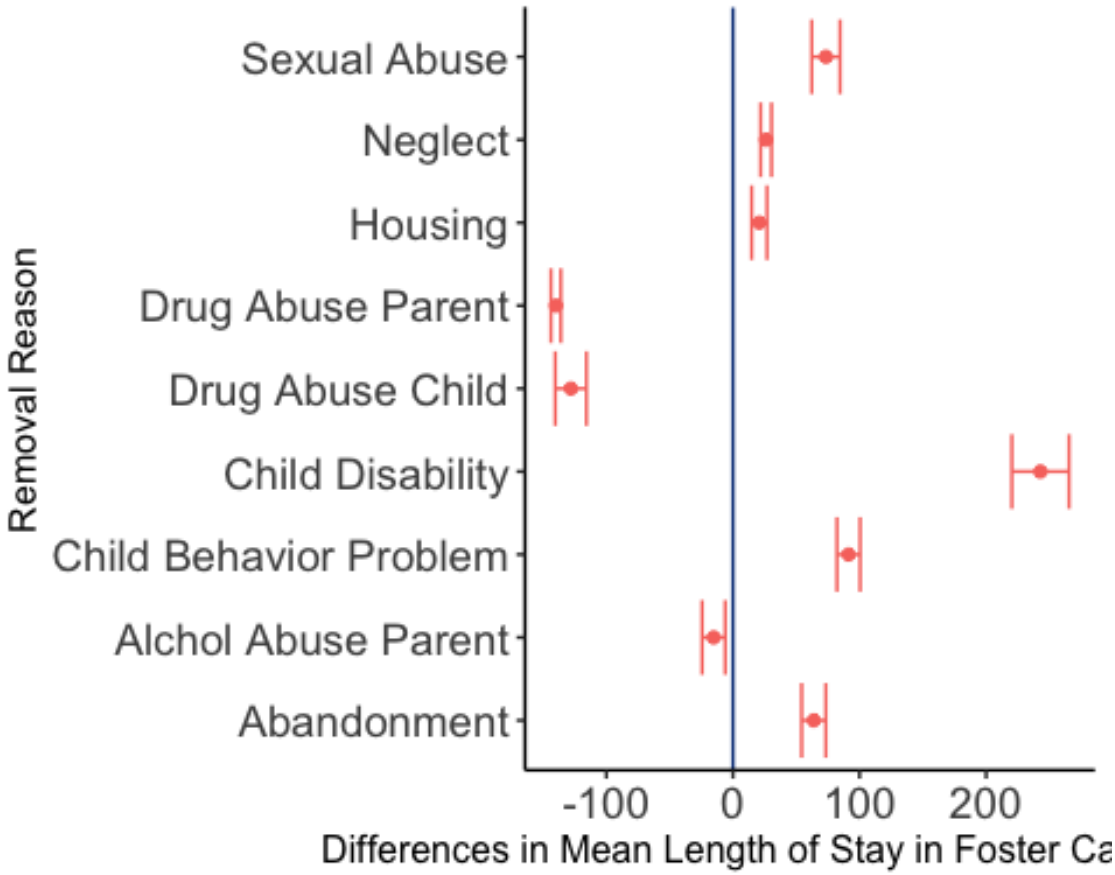
```
group_by(foster, dachild) %>%
  summarise(
    count = n(),
    mean = mean(lifelos, na.rm = TRUE),
    sd = sd(lifelos, na.rm = TRUE),
    median = median(lifelos, na.rm = TRUE),
    min = min(lifelos, na.rm = TRUE),
    max = max(lifelos, na.rm = TRUE)
  )

# A tibble: 2 × 7
  dachild  count  mean    sd median   min   max
```

```
  <chr>    <int> <dbl> <dbl>  <dbl> <int> <int>
1 No      469063  733.  707.    545     0  7670
2 Yes       9354  605.  583.    476     0  7668
```

relevel all categorical variables binding all ttests and creating a graphic

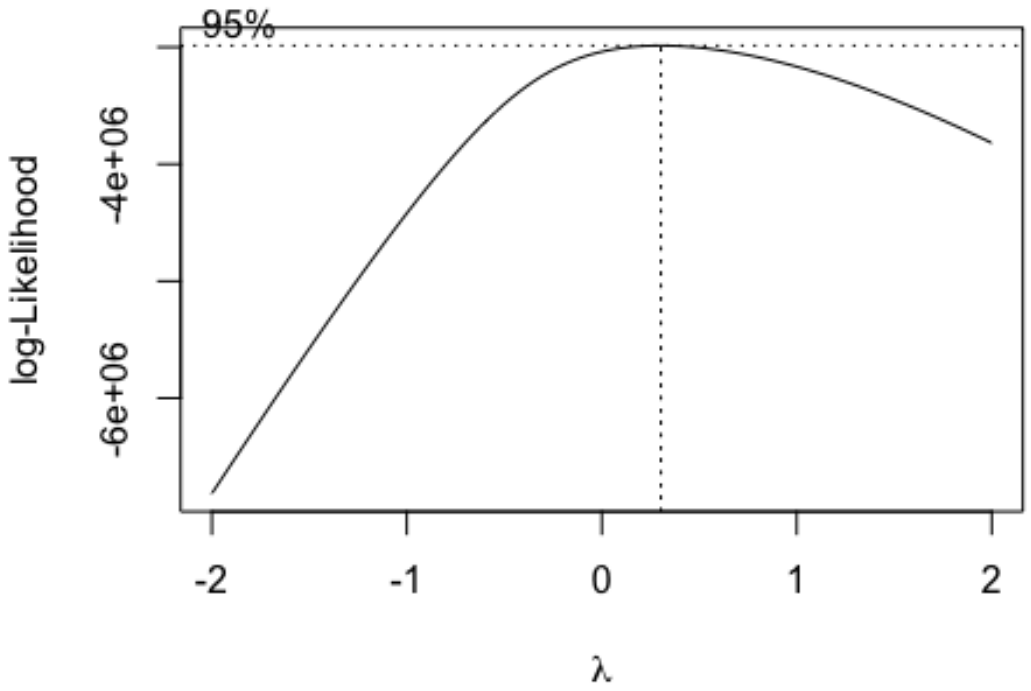
```
foster.regression <- foster.regression %>%
  mutate(childis = childis %>%
           fct_relevel("Yes"))
foster.regression$childis %>% levels()

[1] "Yes" "No"

foster.regression <- foster.regression %>%
  mutate(neglect = neglect %>%
           fct_relevel("Yes"))
foster.regression$neglect %>% levels()

[1] "Yes" "No"

foster.regression <- foster.regression %>%
  mutate(abandmnt = abandmnt %>%
           fct_relevel("Yes"))
foster.regression$abandmnt %>% levels()

[1] "Yes" "No"

foster.regression <- foster.regression %>%
  mutate(housing = housing %>%
           fct_relevel("Yes"))
foster.regression$housing %>% levels()

[1] "Yes" "No"

foster.regression <- foster.regression %>%
  mutate(sexabuse = sexabuse %>%
           fct_relevel("Yes"))
foster.regression$sexabuse %>% levels()

[1] "Yes" "No"

foster.regression <- foster.regression %>%
  mutate(chbehprb = chbehprb %>%
           fct_relevel("Yes"))
foster.regression$chbehprb %>% levels()

[1] "Yes" "No"

foster.regression <- foster.regression %>%
  mutate(aaparent = aaparent %>%
           fct_relevel("Yes"))
foster.regression$aaparent %>% levels()
```

```
[1] "Yes" "No"

foster.regression <- foster.regression %>%
  mutate(daparent = daparent %>%
           fct_relevel("Yes"))
foster.regression$daparent %>% levels()

[1] "Yes" "No"

foster.regression <- foster.regression %>%
  mutate(dachild = dachild %>%
           fct_relevel("Yes"))
foster.regression$dachild %>% levels()

[1] "Yes" "No"

foster.test <-
  dplyr::bind_rows(
  foster.regression %>%
rstatix::t_test(lifelos~neglect, detailed=T),
foster.regression %>%
  rstatix::t_test(lifelos~childis, detailed=T),
foster.regression %>%
  rstatix::t_test(lifelos~abandmnt, detailed=T),
foster.regression %>%
  rstatix::t_test(lifelos~housing, detailed=T),
foster.regression %>%
  rstatix::t_test(lifelos~sexabuse, detailed=T),
foster.regression %>%
  rstatix::t_test(lifelos~chbehprb, detailed=T),
foster.regression %>%
  rstatix::t_test(lifelos~aaparent, detailed=T),
foster.regression %>%
  rstatix::t_test(lifelos~daparent, detailed=T),
foster.regression %>%
  rstatix::t_test(lifelos~dachild, detailed=T)) %>%
  dplyr::mutate(`Removal Reason`=c("Neglect", "Child Disability",
"Abandonment", "Housing", "Sexual Abuse", "Child Behavior Problem", "Alchol
Abuse Parent", "Drug Abuse Parent", "Drug Abuse Child"))
 foster.test %>%
   ggplot(aes(x=`Removal Reason`, y = estimate, color = "#F21D25")) +
  geom_point() +
  geom_errorbar(aes(ymin=conf.low, ymax=conf.high)) +
  geom_hline(yintercept=0, color = "#083D87") +
   labs(y = "Differences in Mean Length of Stay in Foster Care (yes-no)") +
  coord_flip() +
   theme_classic() +
   theme(legend.position = "none", axis.text = element_text(size = 14),
axis.title.y = element_text(size = 12), axis.title.x = element_text(size =
12))
```

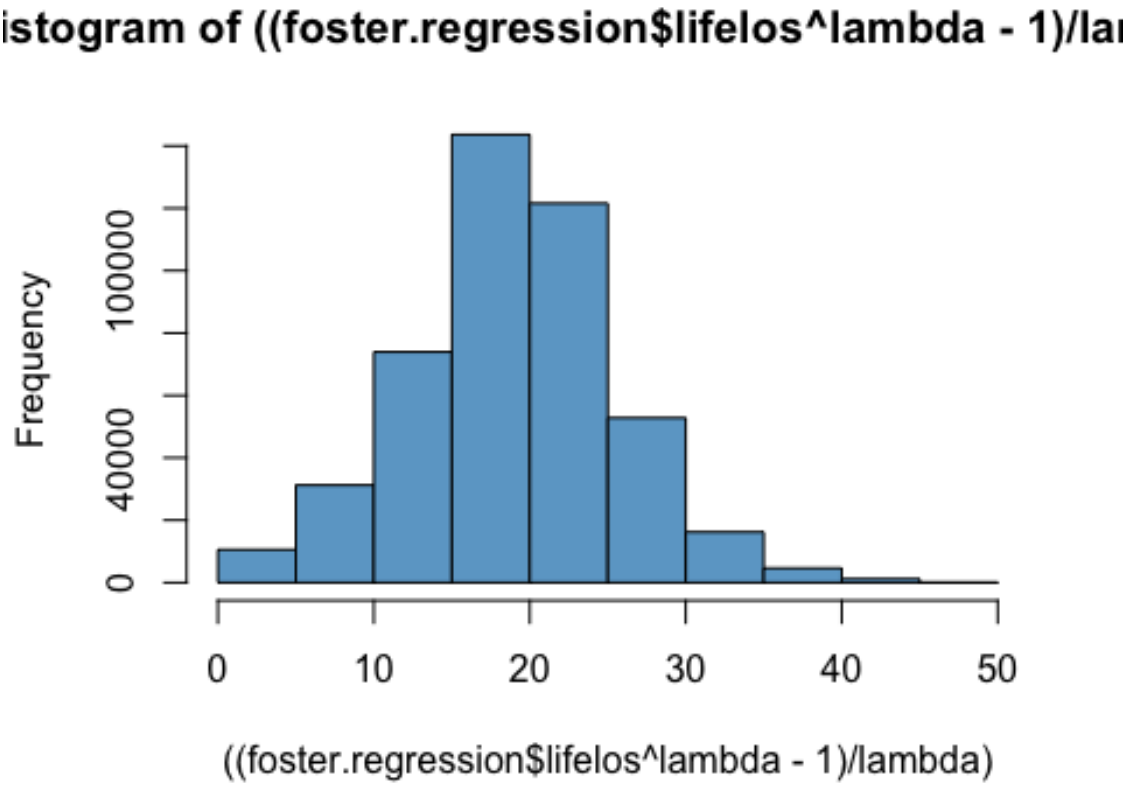


Transformation using boxcox

```
bc1 <- boxcox(foster.regression$lifelos ~ foster.regression$phyabuse +
foster.regression$sexabuse + foster.regression$neglect +
foster.regression$aaparent + foster.regression$daparent +
foster.regression$aachild + foster.regression$dachild +
foster.regression$childis + foster.regression$chbehprb +
foster.regression$prtsjail + foster.regression$abandmnt
+foster.regression$housing)
```




```
lambda <- bc1$x[which.max(bc1$y)]
print(lambda)

[1] 0.3030303

hist(((foster.regression$lifelos^lambda-1)/lambda),
     col = 'skyblue3', breaks = 10)
```




```
foster.regression$lifelos_transformed <- foster.regression$lifelos^lambda
foster_reg_transformed <- lm(foster.regression$lifelos_transformed ~
foster.regression$phyabuse + foster.regression$sexabuse +
foster.regression$neglect  + foster.regression$aachild  +
foster.regression$childis + foster.regression$chbehprb  +
foster.regression$abandmnt +foster.regression$housing +
foster.regression$aaparent + foster.regression$daparent +
foster.regression$dachild + foster.regression$prtsjail)
summary(foster_reg_transformed)


Call:
lm(formula = foster.regression$lifelos_transformed ~
foster.regression$phyabuse +
    foster.regression$sexabuse + foster.regression$neglect +
    foster.regression$aachild + foster.regression$childis +
foster.regression$chbehprb +
    foster.regression$abandmnt + foster.regression$housing +
    foster.regression$aaparent + foster.regression$daparent +
    foster.regression$dachild + foster.regression$prtsjail)

Residuals:
    Min      1Q  Median      3Q     Max
```

```
-6.4659 -1.2232  0.0146  1.2554  8.9714

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                   7.200032   0.039625 181.702  < 2e-16 ***
foster.regression$phyabuseYes  0.036716   0.008932   4.110 3.95e-05 ***
foster.regression$sexabuseNo  -0.100213   0.014839  -6.753 1.45e-11 ***
foster.regression$neglectNo   -0.109489   0.006532 -16.761  < 2e-16 ***
foster.regression$aachildYes  -0.042936   0.053115  -0.808    0.419
foster.regression$childisNo   -0.495051   0.023428 -21.131  < 2e-16 ***
foster.regression$chbehprbNo  -0.143261   0.012411 -11.543  < 2e-16 ***
foster.regression$abandmntNo  -0.182179   0.013822 -13.180  < 2e-16 ***
foster.regression$housingNo   -0.198034   0.009379 -21.115  < 2e-16 ***
foster.regression$aaparentNo   0.082378   0.013024   6.325 2.53e-10 ***
foster.regression$daparentNo   0.269949   0.006402  42.164  < 2e-16 ***
foster.regression$dachildNo    0.393930   0.022273  17.687  < 2e-16 ***
foster.regression$prtsjailYes  0.080727   0.012043   6.703 2.04e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.008 on 455830 degrees of freedom
  (22139 observations deleted due to missingness)
Multiple R-squared:  0.008557,  Adjusted R-squared:  0.008531
F-statistic: 327.9 on 12 and 455830 DF,  p-value: < 2.2e-16
```

```
f_model <- lm(foster.regression$lifelos_transformed ~
foster.regression$sexabuse + foster.regression$neglect  +
foster.regression$aachild  + foster.regression$childis +
foster.regression$chbehprb  + foster.regression$abandmnt
+foster.regression$housing + foster.regression$aaparent +
foster.regression$daparent + foster.regression$dachild +
foster.regression$prtsjail)
summary(f_model)
```

```
Call:
lm(formula = foster.regression$lifelos_transformed ~
foster.regression$sexabuse +
    foster.regression$neglect + foster.regression$aachild +
foster.regression$childis +
    foster.regression$chbehprb + foster.regression$abandmnt +
    foster.regression$housing + foster.regression$aaparent +
    foster.regression$daparent + foster.regression$dachild +
    foster.regression$prtsjail)

Residuals:
    Min      1Q  Median      3Q     Max
-6.4696 -1.2241  0.0154  1.2561  8.9662

Coefficients:
```

```
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                    7.197840   0.039623 181.660  < 2e-16 ***
foster.regression$sexabuseNo  -0.100588   0.014839  -6.778 1.22e-11 ***
foster.regression$neglectNo   -0.106351   0.006488 -16.393  < 2e-16 ***
foster.regression$aachildYes  -0.041242   0.053115  -0.776    0.437
foster.regression$childisNo   -0.498546   0.023413 -21.294  < 2e-16 ***
foster.regression$chbehprbNo  -0.138024   0.012345 -11.180  < 2e-16 ***
foster.regression$abandmntNo  -0.179529   0.013807 -13.002  < 2e-16 ***
foster.regression$housingNo   -0.196904   0.009375 -21.003  < 2e-16 ***
foster.regression$aaparentNo   0.081577   0.013023   6.264 3.75e-10 ***
foster.regression$daparentNo   0.273044   0.006358  42.945  < 2e-16 ***
foster.regression$dachildNo    0.394152   0.022273  17.696  < 2e-16 ***
foster.regression$prtsjailYes  0.080143   0.012042   6.655 2.84e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.008 on 455831 degrees of freedom
  (22139 observations deleted due to missingness)
Multiple R-squared:  0.008521,  Adjusted R-squared:  0.008497
F-statistic: 356.1 on 11 and 455831 DF,  p-value: < 2.2e-16
```

```
f_model_1 <- lm(foster.regression$lifelos_transformed ~
foster.regression$sexabuse + foster.regression$neglect  +
foster.regression$aachild  + foster.regression$childis +
foster.regression$chbehprb  + foster.regression$abandmnt
+foster.regression$housing + foster.regression$aaparent +
foster.regression$daparent + foster.regression$dachild)
summary(f_model_1)
```

```

Call:
lm(formula = foster.regression$lifelos_transformed ~
foster.regression$sexabuse +
    foster.regression$neglect + foster.regression$aachild +
foster.regression$childis +
    foster.regression$chbehprb + foster.regression$abandmnt +
    foster.regression$housing + foster.regression$aaparent +
    foster.regression$daparent + foster.regression$dachild)

Residuals:
    Min      1Q  Median      3Q     Max
-6.4697 -1.2236  0.0166  1.2560  8.9592

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                    7.206884   0.039601 181.987  < 2e-16 ***
foster.regression$sexabuseNo -0.100207   0.014840  -6.752 1.46e-11 ***
foster.regression$neglectNo  -0.105236   0.006486 -16.226  < 2e-16 ***
foster.regression$aachildYes -0.038971   0.053116  -0.734    0.463
foster.regression$childisNo  -0.497411   0.023413 -21.245  < 2e-16 ***
```

```
foster.regression$chbehprbNo -0.135495   0.012340 -10.980  < 2e-16 ***
foster.regression$abandmntNo -0.180044   0.013808 -13.039  < 2e-16 ***
foster.regression$housingNo  -0.199602   0.009367 -21.310  < 2e-16 ***
foster.regression$aaparentNo  0.077520   0.013009   5.959 2.54e-09 ***
foster.regression$daparentNo  0.270846   0.006350  42.655  < 2e-16 ***
foster.regression$dachildNo   0.394258   0.022274  17.700  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.008 on 455832 degrees of freedom
  (22139 observations deleted due to missingness)
Multiple R-squared:  0.008424,  Adjusted R-squared:  0.008403
F-statistic: 387.3 on 10 and 455832 DF,  p-value: < 2.2e-16
```

```
f_model_2 <- lm(foster.regression$lifelos_transformed ~
foster.regression$sexabuse + foster.regression$neglect +
foster.regression$childis + foster.regression$chbehprb  +
foster.regression$abandmnt +foster.regression$housing +
foster.regression$aaparent + foster.regression$daparent +
foster.regression$dachild)
summary(f_model_2)


Call:
lm(formula = foster.regression$lifelos_transformed ~
foster.regression$sexabuse +
    foster.regression$neglect + foster.regression$childis +
foster.regression$chbehprb +
    foster.regression$abandmnt + foster.regression$housing +
    foster.regression$aaparent + foster.regression$daparent +
    foster.regression$dachild)

Residuals:
    Min      1Q  Median      3Q     Max
-6.4687 -1.2234  0.0167  1.2560  8.9621

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                  7.202540   0.039156 183.945  < 2e-16 ***
foster.regression$sexabuseNo -0.100119   0.014839  -6.747 1.51e-11 ***
foster.regression$neglectNo  -0.105236   0.006486 -16.226  < 2e-16 ***
foster.regression$childisNo  -0.496881   0.023402 -21.232  < 2e-16 ***
foster.regression$chbehprbNo -0.135069   0.012326 -10.958  < 2e-16 ***
foster.regression$abandmntNo -0.179980   0.013808 -13.035  < 2e-16 ***
foster.regression$housingNo  -0.199572   0.009367 -21.307  < 2e-16 ***
foster.regression$aaparentNo  0.077841   0.013002   5.987 2.14e-09 ***
foster.regression$daparentNo  0.270845   0.006350  42.654  < 2e-16 ***
foster.regression$dachildNo   0.397139   0.021925  18.113  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

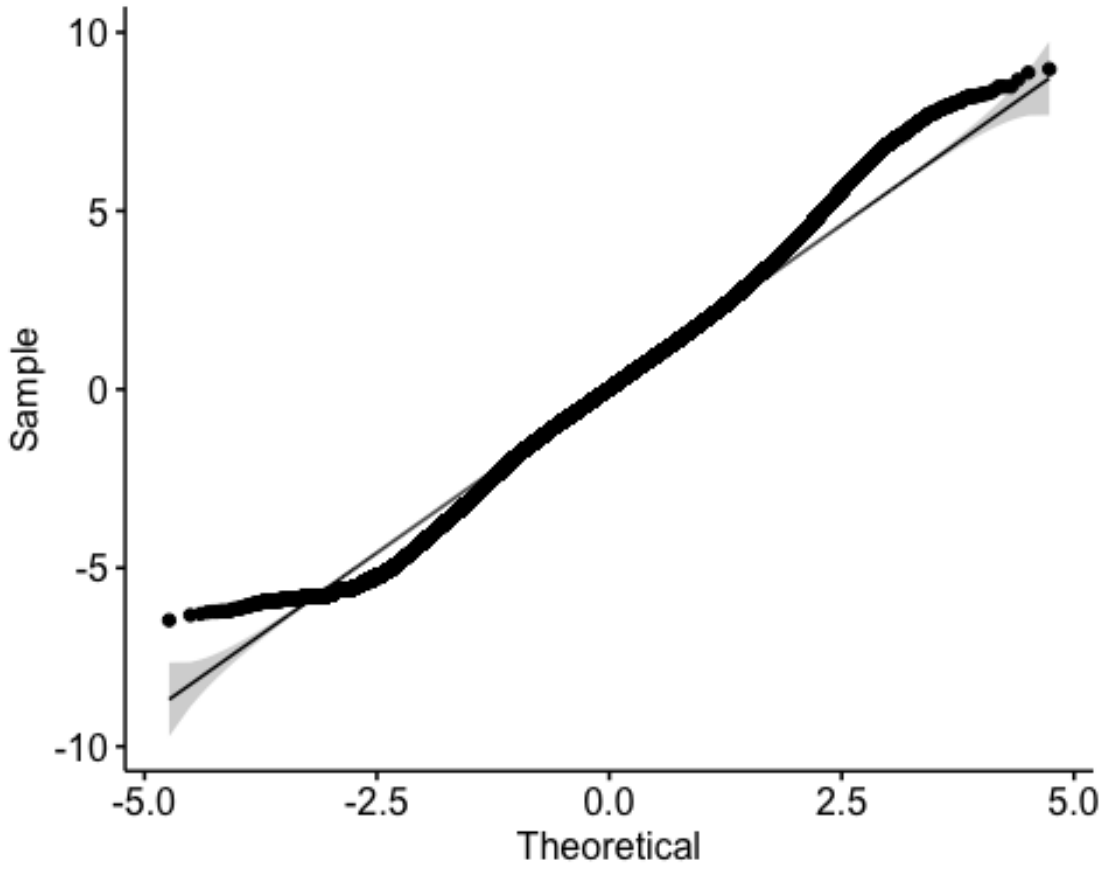
```
Residual standard error: 2.008 on 455833 degrees of freedom
  (22139 observations deleted due to missingness)
Multiple R-squared:  0.008423,  Adjusted R-squared:  0.008404
F-statistic: 430.2 on 9 and 455833 DF,  p-value: < 2.2e-16
```

```
f_model_3 <- lm(foster.regression$lifelos_transformed ~
foster.regression$neglect + foster.regression$childis +
foster.regression$abandmnt +foster.regression$housing +
foster.regression$aaparent + foster.regression$daparent +
foster.regression$dachild +foster.regression$sexabuse
+foster.regression$chbehprb)
summary(f_model_3)
```

```

Call:
lm(formula = foster.regression$lifelos_transformed ~
foster.regression$neglect +
    foster.regression$childis + foster.regression$abandmnt +
    foster.regression$housing + foster.regression$aaparent +
    foster.regression$daparent + foster.regression$dachild +
    foster.regression$sexabuse + foster.regression$chbehprb)

Residuals:
    Min      1Q  Median      3Q     Max
-6.4687 -1.2234  0.0167  1.2560  8.9621

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                   7.202540   0.039156 183.945  < 2e-16 ***
foster.regression$neglectNo  -0.105236   0.006486 -16.226  < 2e-16 ***
foster.regression$childisNo  -0.496881   0.023402 -21.232  < 2e-16 ***
foster.regression$abandmntNo -0.179980   0.013808 -13.035  < 2e-16 ***
foster.regression$housingNo  -0.199572   0.009367 -21.307  < 2e-16 ***
foster.regression$aaparentNo  0.077841   0.013002   5.987 2.14e-09 ***
foster.regression$daparentNo  0.270845   0.006350  42.654  < 2e-16 ***
foster.regression$dachildNo   0.397139   0.021925  18.113  < 2e-16 ***
foster.regression$sexabuseNo -0.100119   0.014839  -6.747 1.51e-11 ***
foster.regression$chbehprbNo -0.135069   0.012326 -10.958  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.008 on 455833 degrees of freedom
  (22139 observations deleted due to missingness)
Multiple R-squared:  0.008423,  Adjusted R-squared:  0.008404
F-statistic: 430.2 on 9 and 455833 DF,  p-value: < 2.2e-16
```

check normality assumption

```
foster_reg_transformed$residuals %>%
  ggpubr::ggqqplot()
```



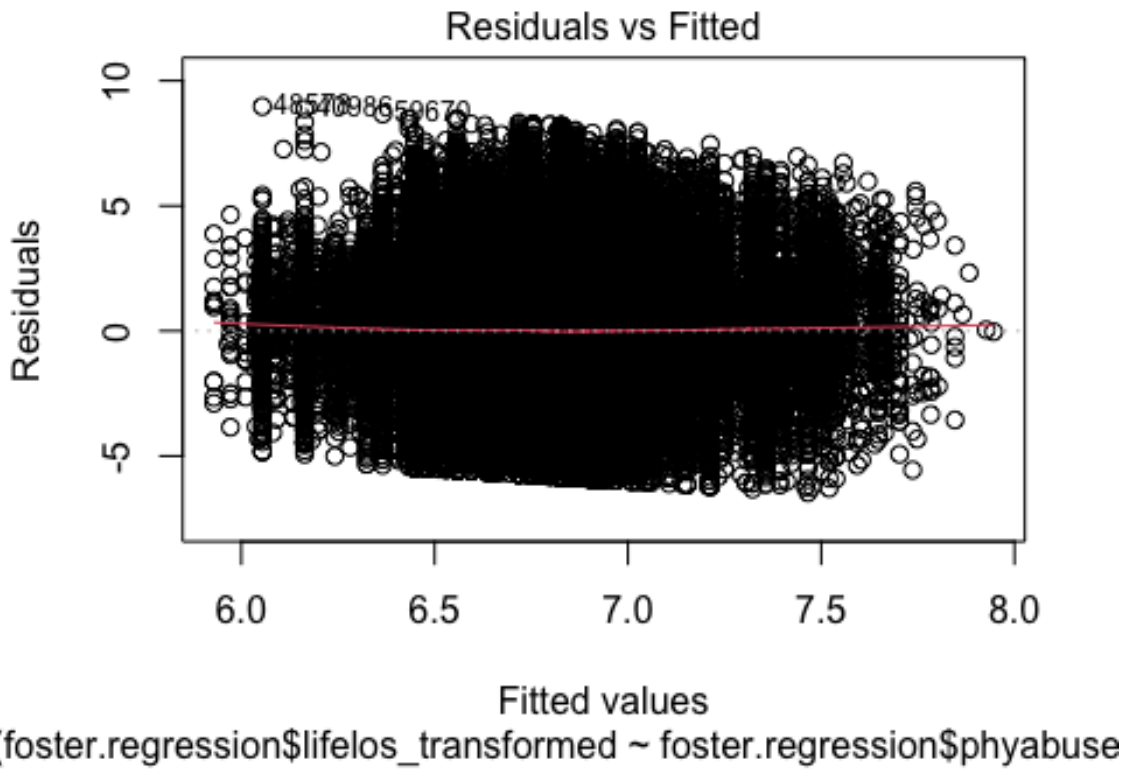
vif

```
car::vif(foster_reg_transformed)
```

```
foster.regression$phyabuse foster.regression$sexabuse
                  1.034334                   1.014080
 foster.regression$neglect  foster.regression$aachild
                  1.086287                   1.041580
 foster.regression$childis foster.regression$chbehprb
                  1.036542                   1.121508
foster.regression$abandmnt  foster.regression$housing
                  1.012867                   1.021818
foster.regression$aaparent foster.regression$daparent
                  1.006587                   1.082568
 foster.regression$dachild foster.regression$prtsjail
                  1.066268                   1.010352
```

vif are still low check for constant variance

```
plot(foster_reg_transformed,1)
```




```
summary(foster_reg_transformed)


Call:
lm(formula = foster.regression$lifelos_transformed ~
foster.regression$phyabuse +
    foster.regression$sexabuse + foster.regression$neglect +
    foster.regression$aachild + foster.regression$childis +
foster.regression$chbehprb +
    foster.regression$abandmnt + foster.regression$housing +
    foster.regression$aaparent + foster.regression$daparent +
    foster.regression$dachild + foster.regression$prtsjail)

Residuals:
    Min      1Q  Median      3Q     Max
-6.4659 -1.2232  0.0146  1.2554  8.9714

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                   7.200032   0.039625 181.702  < 2e-16 ***
foster.regression$phyabuseYes  0.036716   0.008932   4.110 3.95e-05 ***
foster.regression$sexabuseNo  -0.100213   0.014839  -6.753 1.45e-11 ***
foster.regression$neglectNo   -0.109489   0.006532 -16.761  < 2e-16 ***
```

```
foster.regression$aachildYes  -0.042936   0.053115  -0.808    0.419
foster.regression$childisNo   -0.495051   0.023428 -21.131  < 2e-16 ***
foster.regression$chbehprbNo  -0.143261   0.012411 -11.543  < 2e-16 ***
foster.regression$abandmntNo  -0.182179   0.013822 -13.180  < 2e-16 ***
foster.regression$housingNo   -0.198034   0.009379 -21.115  < 2e-16 ***
foster.regression$aaparentNo   0.082378   0.013024   6.325 2.53e-10 ***
foster.regression$daparentNo   0.269949   0.006402  42.164  < 2e-16 ***
foster.regression$dachildNo    0.393930   0.022273  17.687  < 2e-16 ***
foster.regression$prtsjailYes  0.080727   0.012043   6.703 2.04e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.008 on 455830 degrees of freedom
  (22139 observations deleted due to missingness)
Multiple R-squared:  0.008557,  Adjusted R-squared:  0.008531
F-statistic: 327.9 on 12 and 455830 DF,  p-value: < 2.2e-16
```

the presence of physical abuse, sex abuse, neglect, aachild, childis, chbehprb, abandmnt, housing predict longer length of of stay in foster care

CHI SQUARE

creating a subset for chi square

```
foster.chi.variables <- c("ctkfamst", "race")
foster.chi <- foster[foster.chi.variables]
```

table for observed count

```
tbl2 <- table(foster.chi$race, foster.chi$ctkfamst)
tbl2
```

```
                  Married couple Single female Single male Unmarried couple
  White                    64280         132517       19405            83634
  African American         11951          75397        5701            24206
  American Indian           1560           6629         736             3212
  Asian                      896           1081         172              610
  Pacific Islander           355            681          84              570
  Multiple Race             7107          23195        2548            11890
```

```
proportions(tbl2, 1)
```

```
                  Married couple Single female Single male Unmarried couple
  White               0.21438386    0.44196494  0.06471871       0.27893248
  African American    0.10192316    0.64301736  0.04862053       0.20643896
  American Indian     0.12853259    0.54618110  0.06064102       0.26464530
  Asian               0.32475535    0.39180863  0.06234143       0.22109460
  Pacific Islander    0.21005917    0.40295858  0.04970414       0.33727811
  Multiple Race       0.15885114    0.51843987  0.05695127       0.26575771
```

table for observed count

```
tbl3 <- table(foster.chi$ctkfamst, foster.chi$race)
tbl3
```

```
                   White African American American Indian  Asian
  Married couple    64280            11951            1560    896
  Single female    132517            75397            6629   1081
  Single male       19405             5701             736    172
  Unmarried couple  83634            24206            3212    610

                  Pacific Islander Multiple Race
  Married couple                355          7107
  Single female                 681         23195
  Single male                    84          2548
  Unmarried couple              570         11890
```

observed proportion

```
proportions(tbl2,1)
```

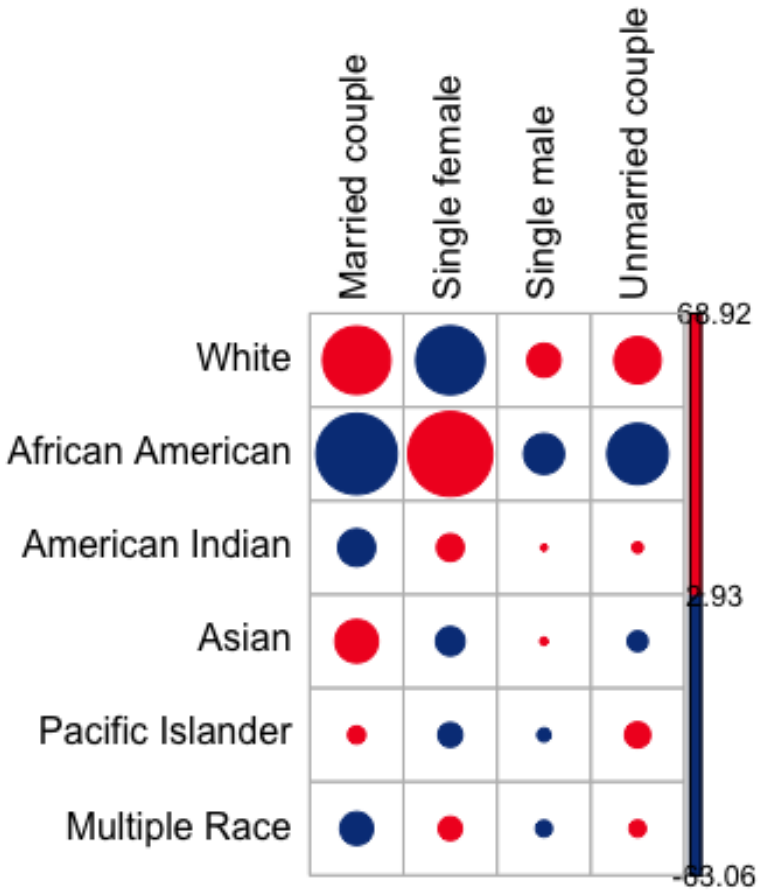
```
                  Married couple Single female Single male Unmarried couple
  White                0.21438386    0.44196494  0.06471871       0.27893248
  African American     0.10192316    0.64301736  0.04862053       0.20643896
  American Indian      0.12853259    0.54618110  0.06064102       0.26464530
  Asian                0.32475535    0.39180863  0.06234143       0.22109460
  Pacific Islander     0.21005917    0.40295858  0.04970414       0.33727811
  Multiple Race        0.15885114    0.51843987  0.05695127       0.26575771
```

chi square independence test + correlation plot

```
chisq2 <- chisq.test(x=tbl2, correct=FALSE)
colors6<-c("#664799" , "#895FCC", "#A976FC", "#0DDBB2","#0AAD8C", "#07846B")
colors2<- c("#083D87", "#F21D25")
names(chisq2)
```

```
[1] "statistic" "parameter" "p.value"   "method"    "data.name" "observed"
[7] "expected"  "residuals" "stdres"
```

```
corrplot(chisq2$residuals, is.cor = FALSE, col=colors2, tl.col="black")
```



SIMPLE LINEAR REGRESSION

cleaning up age

```
foster = foster[!foster$ageatstart == "DOB Missing",]
foster$ageatstart <- as.numeric(foster$ageatstart)
foster = foster[!foster$ageatstart <=0,]
```

creating a subset

```
foster.slr.variables <- c("ctkfamst", "race", "lifelos", "ageatstart", "sex")
foster.slr <- foster[foster.slr.variables]
sum_up(foster.slr, ageatstart)
```

Variable	Obs	Missing	Mean	StdDev	Min	Max
ageatstart	420158	0	8.47315	5.29597	1	28

cleaning up gender

```
foster.slr <- foster.slr %>%
  filter(foster.slr$sex %in% c("Female", "Male"))
tab(foster, sex)
```

```
   sex |      Freq.   Percent      Cum.
-------+-----------------------------------
Female |     204473     48.67     48.67
  Male |     215685     51.33    100.00
```

cleaning up caretaker family structure

```
foster.slr <- foster.slr%>%
  filter(foster.slr$ctkfamst %in% c("Married couple", "Unmarried couple",
"Single female", "Single male"))
tab(foster.slr, ctkfamst)
```

```
        ctkfamst |      Freq.   Percent      Cum.
-----------------+-----------------------------------
  Married couple |      79904     19.02     19.02
   Single female |     209725     49.92     68.93
     Single male |      27853      6.63     75.56
Unmarried couple |     102676     24.44    100.00
```

cleaning up race variable

```
foster.slr <- foster.slr%>%
  filter(foster.slr$race %in% c("White", "Black or African American",
"American Indian or Alaska Native", "Asian", "Hawaiian or Other Pacific
Islander", "More Than One Race"))
tab(foster.slr, race)
```

```
 race |      Freq.   Percent      Cum.
------+-----------------------------------
White |     261913     99.06     99.06
Asian |       2491      0.94    100.00
```

only keeping complete cases

```
foster.slr <- foster.slr[complete.cases(foster.slr),]
```

slr model

```
reg_model <- lm(lifelos ~ ageatstart, data=foster.slr)
summary(reg_model)
```

```

Call:
lm(formula = lifelos ~ ageatstart, data = foster.slr)

Residuals:
    Min      1Q  Median      3Q     Max
-1084.1  -423.0  -115.1   270.3  6535.9
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 466.2696     2.3888   195.2   <2e-16 ***
ageatstart   33.3901     0.2448   136.4   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 641.5 on 252005 degrees of freedom
Multiple R-squared:  0.06874,   Adjusted R-squared:  0.06873
F-statistic: 1.86e+04 on 1 and 252005 DF,  p-value: < 2.2e-16
```

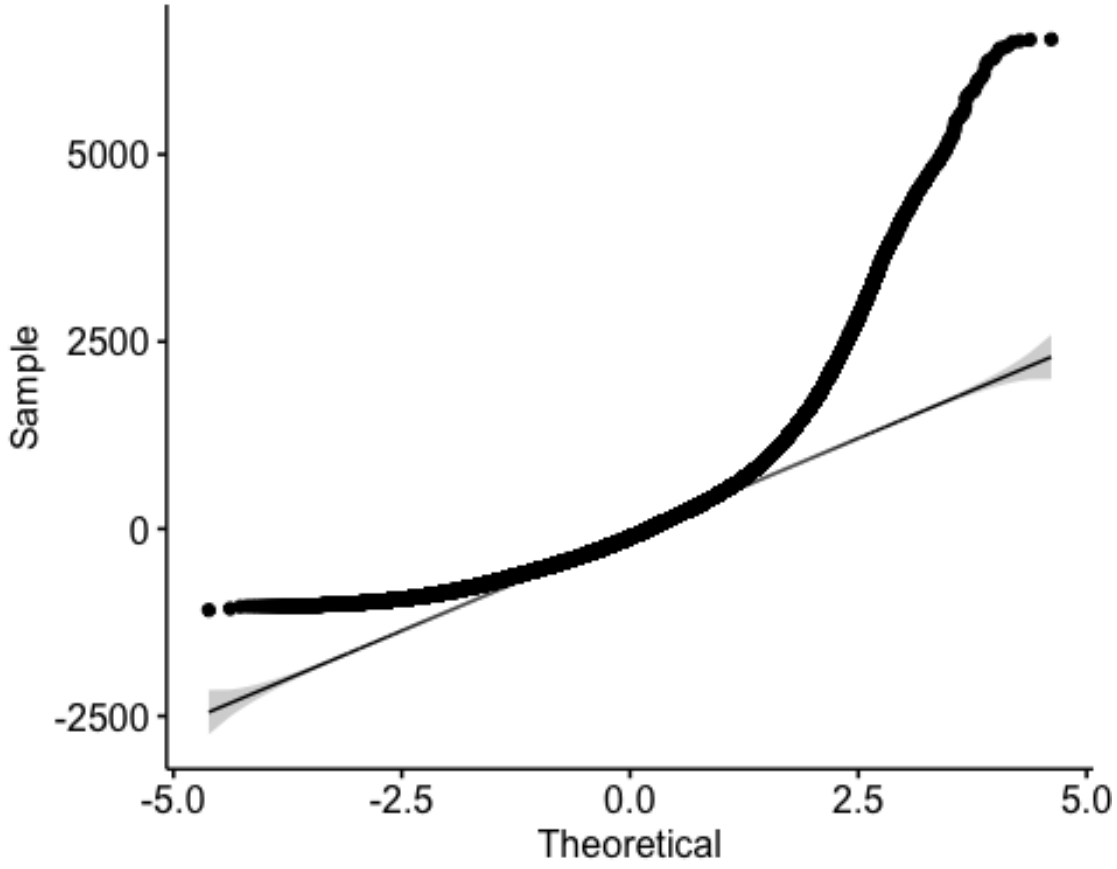
confidence interval

```
confint(reg_model, "ageatstart", level=0.95)
```

```
              2.5 %   97.5 %
ageatstart 32.91022 33.86992
```

check normality assumption

```
reg_model$residuals %>%
  ggpubr::ggqqplot()
```



histogram of the residuals

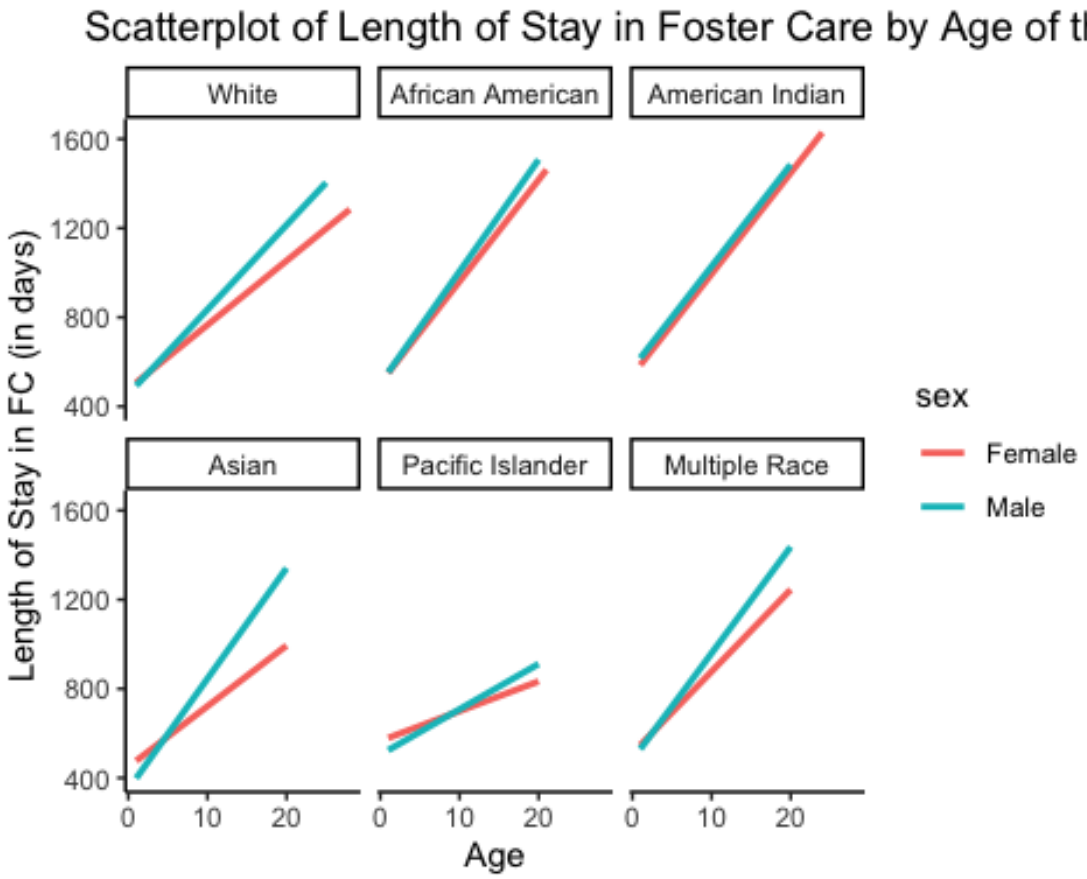
```
hist(reg_model$residuals)
```

check constant variance

#plot(reg_model, 1)

multi layer scatterplot

```
foster %>%
  dplyr::arrange(desc(ageatstart)) %>%
  ggplot(aes(as.numeric(ageatstart), as.numeric(lifelos))) +
  #geom_point(aes(size=.000001, color=sex), alpha=0.5) +
  geom_smooth(aes(color=sex), method='lm', se=F) +
  labs(x = "Age",
       y = "Length of Stay in FC (in days)",
       title = "Scatterplot of Length of Stay in Foster Care by Age of the
Child") +
  theme_classic() +
  theme(plot.title=element_text(hjust=0.1)) + facet_wrap(~race)
```




```
tab(foster.slr, race)
```

race	Freq.	Percent	Cum.
White	249610	99.05	99.05
Asian	2397	0.95	100.00

CHI SQUARE

creating a subset for chi square

```
foster.chi1.variables <- c("ctkfamst", "disreasn")
foster.chi1 <- foster[foster.chi1.variables]
```

cleaning up disreasn variable

```
foster.chi1 <- foster.chi1 %>%
  filter(foster.chi1$disreasn %in% c("Adoption", "Death of child",
"Emancipation", "Guardianchip", "Living with other relative(s)", "Reunified
with parent, primary caretaker", "Runaway"))
tab(foster.chi1, disreasn)
```

```
                                disreasn |      Freq.     Percent        Cum.
-----------------------------------------+-----------------------------------
                                Adoption |      44153       32.04       32.04
                           Death of child |        207        0.15       32.19
                             Emancipation |      15448       11.21       43.40
             Living with other relative(s) |       7404        5.37       48.77
Reunified with parent, primary caretaker |      70239       50.97       99.73
                                 Runaway |        367        0.27      100.00
```

changing names of categories for discharge reason to make space on poster

```
foster.chi1$disreasn <-  recode_factor(foster.chi1$disreasn, "Reunified with
parent, primary caretaker"= "Reunified with parent")
levels(foster.chi1$disreasn)

[1] "Reunified with parent"         "Emancipation"
[3] "Living with other relative(s)" "Runaway"
[5] "Adoption"                      "Death of child"
```

table for observed count

```
tbl1 <- table(foster.chi1$disreasn, foster.chi1$ctkfamst)
tbl1


                               Married couple Single female Single male
  Reunified with parent                 14701         33922        4355
  Emancipation                           3955          7723        1691
  Living with other relative(s)          1473          4059         502
  Runaway                                  64           225          45
  Adoption                               7016         22583        2055
  Death of child                           34           108          13

                               Unmarried couple
  Reunified with parent                 17261
  Emancipation                           2079
  Living with other relative(s)          1370
  Runaway                                  33
  Adoption                               12499
  Death of child                           52
```

table for observed count

```
tbl <- table(foster.chi1$ctkfamst, foster.chi1$disreasn)
tbl


                  Reunified with parent Emancipation
  Married couple                  14701         3955
  Single female                   33922         7723
```

```
  Single male                      4355         1691
  Unmarried couple                17261         2079

                   Living with other relative(s) Runaway Adoption
  Married couple                             1473      64     7016
  Single female                              4059     225    22583
  Single male                                 502      45     2055
  Unmarried couple                           1370      33    12499

                   Death of child
  Married couple                34
  Single female                108
  Single male                   13
  Unmarried couple              52
```

observed proportion

```
proportions(tbl,1)


                   Reunified with parent Emancipation
  Married couple             0.5396248578 0.1451749073
  Single female              0.4943456718 0.1125473623
  Single male                0.5028287727 0.1952430435
  Unmarried couple           0.5184417613 0.0624436835

                   Living with other relative(s)      Runaway     Adoption
  Married couple                     0.0540689351 0.0023492273 0.2575340454
  Single female                      0.0591518508 0.0032789274 0.3291023025
  Single male                        0.0579609745 0.0051957049 0.2372705230
  Unmarried couple                   0.0411485553 0.0009911696 0.3754129873

                   Death of child
  Married couple     0.0012480270
  Single female      0.0015738852
  Single male        0.0015009814
  Unmarried couple   0.0015618430
```

chi square independence test

```
chisq <- chisq.test(x=tbl1, correct=FALSE)
```

view disreasn

```
tab(foster, disreasn)

                              disreasn |      Freq.     Percent        Cum.
---------------------------------------+-----------------------------------
                              Adoption |      44153       10.51       10.51
                         Death of child |        207        0.05       10.56
```

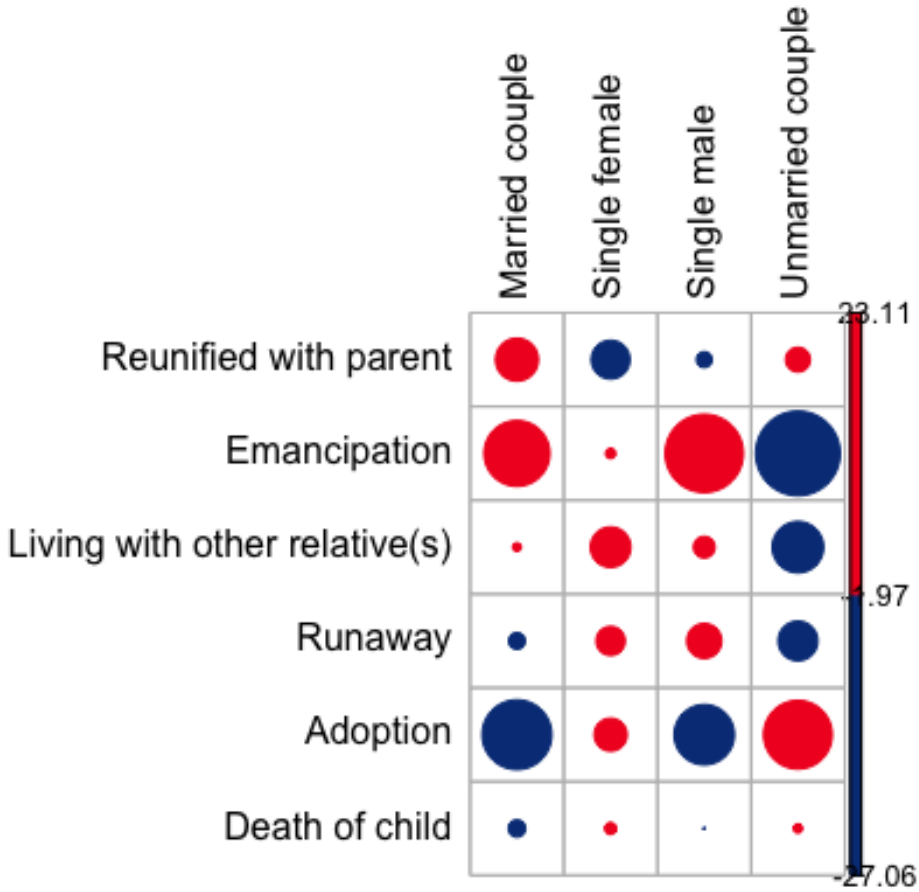
```
                              Emancipation |    15448     3.68    14.23
                              Guardianship |    18000     4.28    18.52
           Living with other relative(s) |     7404     1.76    20.28
                                   Missing |   262943    62.58    82.86
Reunified with parent, primary caretaker |    70239    16.72    99.58
                                   Runaway |      367     0.09    99.67
                Transfer to another agency |     1397     0.33   100.00
```

correlation plot

```
#tl.srt tilts the labels at a 45 degree angle
#tl.col specifies the color of the factor labels on the contingency table
colors6<-c("#664799" , "#895FCC", "#A976FC", "#0DDBB2","#0AAD8C", "#07846B")
colors2<- c("#083D87", "#F21D25")
names(chisq)
```

```
[1] "statistic" "parameter" "p.value"   "method"    "data.name" "observed"
[7] "expected"  "residuals" "stdres"
```

```
corrplot(chisq$residuals, is.cor = FALSE, col=colors2, tl.col="black")
```




```
#corrplot(chisq$stdres, is.cor = FALSE, col=colors2, tl.col="black",
tl.srt=45)
```

```
#corrplot::corrplot(chisq$residuals, is.cor = FALSE, col=colors2,
tl.col="black", tl.srt=45)
```

ONE WAY ANOVA

Does the race of the child have an effect on the length of childs stay in foster care H0: The race of the child has no affect on the length of childs stay in foster care Ha: Child of at least one race has a diiferent length of stay in foster care Race: White, American Indian or Alaska Native, Asian, Black or African American, Hawaiian or Other Pacific Islander, More than One Race How many days the child has spent in foster care Type of measure: days One categorical with 6 categories

creating a subset for one way anova

```
foster.one.way.variables <- c("lifelos", "race")
foster.one.way <- foster[foster.one.way.variables]
```

cleaning up race variable

```
foster.one.way <- foster.one.way%>%
  filter(foster.one.way$race %in% c("White", "African American", "American
Indian", "Asian", "Pacific Islander", "Multiple Race"))
tab(foster.one.way, race)
```

```
            race |      Freq.   Percent      Cum.
-----------------+---------------------------------
           White |    261913     62.34     62.34
African American |    104334     24.83     87.17
 American Indian |     10676      2.54     89.71
           Asian |      2491      0.59     90.30
Pacific Islander |      1488      0.35     90.66
   Multiple Race |     39256      9.34    100.00
```

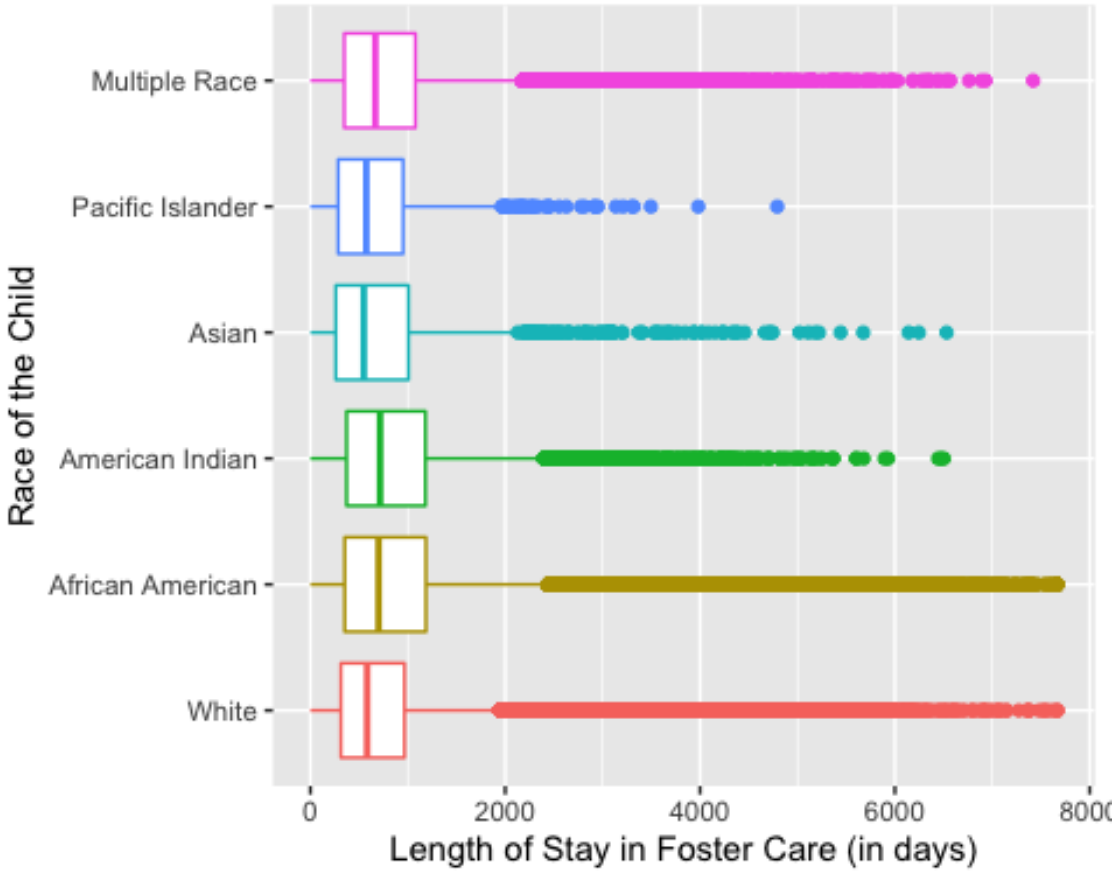
cleaning up lifelos

```
foster.one.way = foster.one.way[!foster.one.way$lifelos <=0,]
sum_up(foster.one.way, lifelos)
```

```
Variable |       Obs  Missing      Mean    StdDev       Min       Max
---------+--------------------------------------------------------------
 lifelos |    397810    22040   795.891   729.132         1      7670
```

HOMOGENEITY OF VARIANCES rotate (rename to make shorter)

```
plot <- ggplot(foster, aes(x=lifelos, y = race, color = race)) +
  labs(y = "Race of the Child", x = "Length of Stay in Foster Care (in
days)") +
  geom_boxplot()+theme(legend.position = "none")
plot
```



ONE WAY ANOVA

```
one.way1 <- aov(lifelos ~ race, data = foster.one.way)
summary(one.way1)

               Df    Sum Sq   Mean Sq F value Pr(>F)
race            5 2.272e+09 454387334     864 <2e-16 ***
Residuals   397804 2.092e+11    525928
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
22040 observations deleted due to missingness
```

summary statistics

```
by(foster.one.way$lifelos, foster.one.way$race, summary)

foster.one.way$race: White
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.0   314.0   582.0   741.9   966.0  7670.0
------------------------------------------------------------
foster.one.way$race: African American
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.0   353.0   700.0   914.3  1187.0  7669.0
------------------------------------------------------------
```

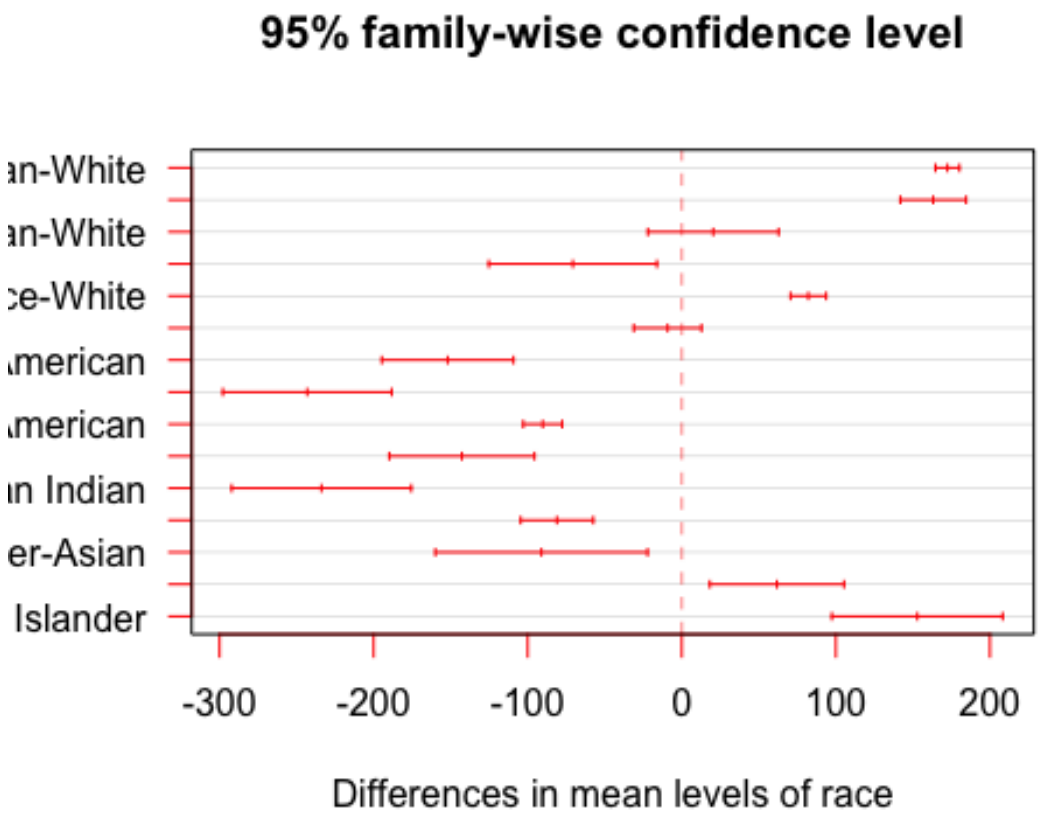
```
foster.one.way$race: American Indian
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.0   372.0   716.5   905.1  1181.0  6503.0
------------------------------------------------------------
foster.one.way$race: Asian
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.0   261.0   548.0   762.3  1007.0  6531.0
------------------------------------------------------------
foster.one.way$race: Pacific Islander
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.0   285.8   574.5   671.2   951.2  4792.0
------------------------------------------------------------
foster.one.way$race: Multiple Race
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    1.0   348.0   668.0   824.1  1078.0  7419.0

tab(foster.one.way, race)


            race |     Freq.  Percent     Cum.
-----------------+-----------------------------
           White |   249399    59.40    59.40
African American |    98076    23.36    82.76
 American Indian |     9748     2.32    85.08
           Asian |     2397     0.57    85.65
Pacific Islander |     1444     0.34    86.00
   Multiple Race |    36746     8.75    94.75
              NA |    22040     5.25   100.00
```

post-hoc test

```
TukeyHSD(one.way1, conf.level = 0.95)

  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = lifelos ~ race, data = foster.one.way)

$race
                                          diff        lwr        upr     p adj
African American-White              172.376300  164.58704  180.16556 0.0000000
American Indian-White               163.158711  141.82182  184.49561 0.0000000
Asian-White                          20.401784  -22.01193   62.81550 0.7446665
Pacific Islander-White              -70.716368 -125.25865  -16.17409 0.0030147
Multiple Race-White                  82.110562   70.56264   93.65849 0.0000000
American Indian-African American     -9.217589  -31.16493   12.72975 0.8385969
Asian-African American             -151.974517 -194.69858 -109.25045 0.0000000
Pacific Islander-African American  -243.092668 -297.87663 -188.30870 0.0000000
Multiple Race-African American      -90.265738 -102.90603  -77.62544 0.0000000
Asian-American Indian              -142.756928 -189.87312  -95.64073 0.0000000
Pacific Islander-American Indian   -233.875079 -292.14920 -175.60096 0.0000000
```

```
Multiple Race-American Indian      -81.048149 -104.59316  -57.50314 0.0000000
Pacific Islander-Asian             -91.118151 -159.96242  -22.27388 0.0022419
Multiple Race-Asian                 61.708778   18.14242  105.27514 0.0007704
Multiple Race-Pacific Islander     152.826930   97.38358  208.27028 0.0000000
```

```
plot(TukeyHSD(one.way1, conf.level = 0.95), las=1, col = "red")
```




```
tukey.plot.aov <- aov(lifelos ~ race, data=foster.one.way)
```

mean, sd

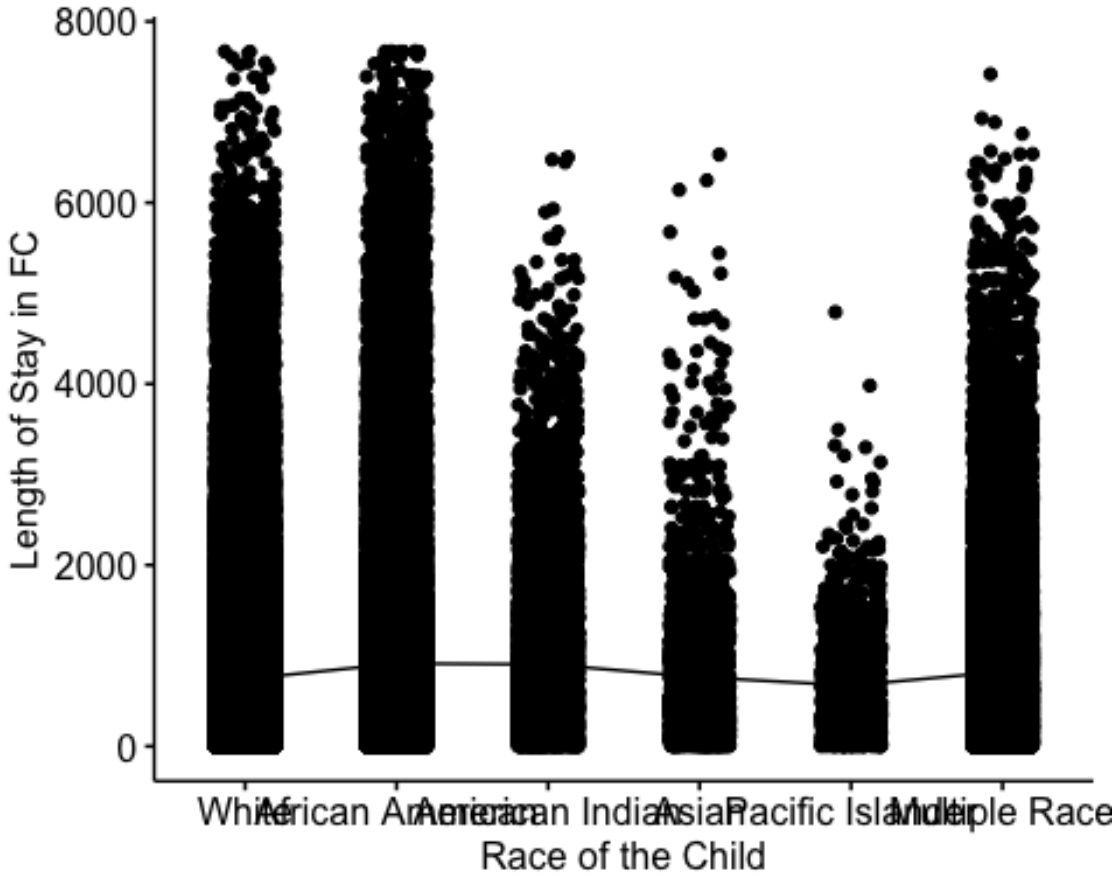
```
group_by(foster.one.way, race) %>%
  summarise(
    count = n(),
    mean = mean(lifelos, na.rm = TRUE),
    sd = sd(lifelos, na.rm = TRUE),
    median = median(lifelos, na.rm = TRUE),
    min = min(lifelos, na.rm = TRUE),
    max = max(lifelos, na.rm = TRUE)
  )
```

```
# A tibble: 7 × 7
  race              count  mean    sd median   min   max
  <fct>             <int> <dbl> <dbl>  <dbl> <dbl> <dbl>
```

```
1 White            249399  742.  664.   582      1  7670
2 African American  98076  914.  861.   700      1  7669
3 American Indian    9748  905.  786.   716.     1  6503
4 Asian              2397  762.  776.   548      1  6531
5 Pacific Islander   1444  671.  542.   574.     1  4792
6 Multiple Race     36746  824.  715.   668      1  7419
7 <NA>              22040  NaN    NA     NA    Inf  -Inf
```

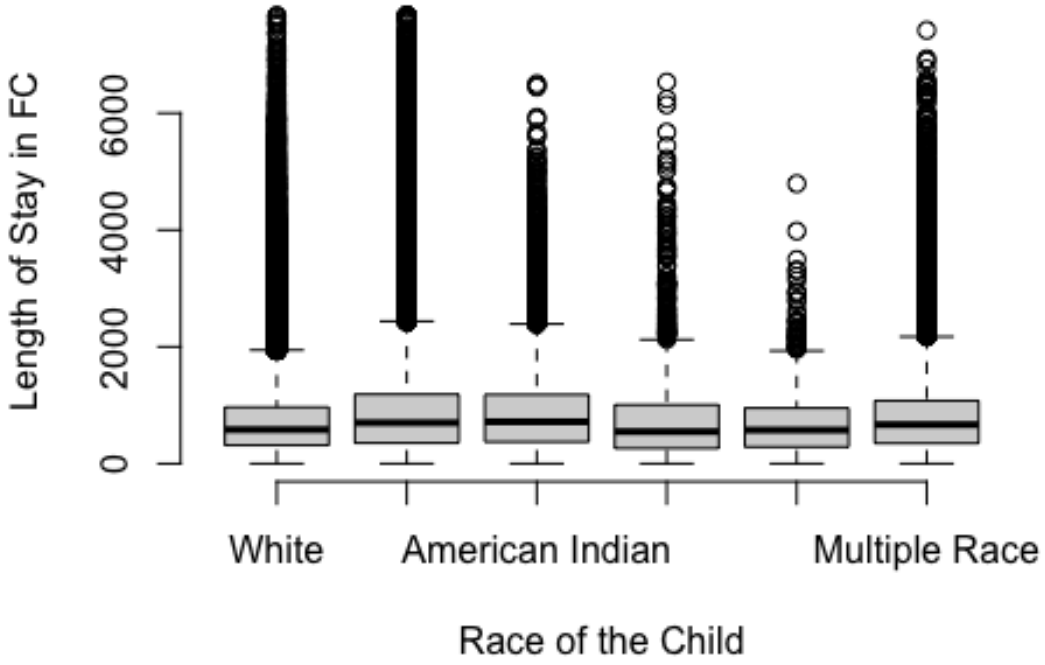
std 905/574 <2

```
ggline(foster, x = "race", y = "lifelos",
       add = c("mean_se", "jitter"),
       ylab = "Length of Stay in FC", xlab = "Race of the Child")
```



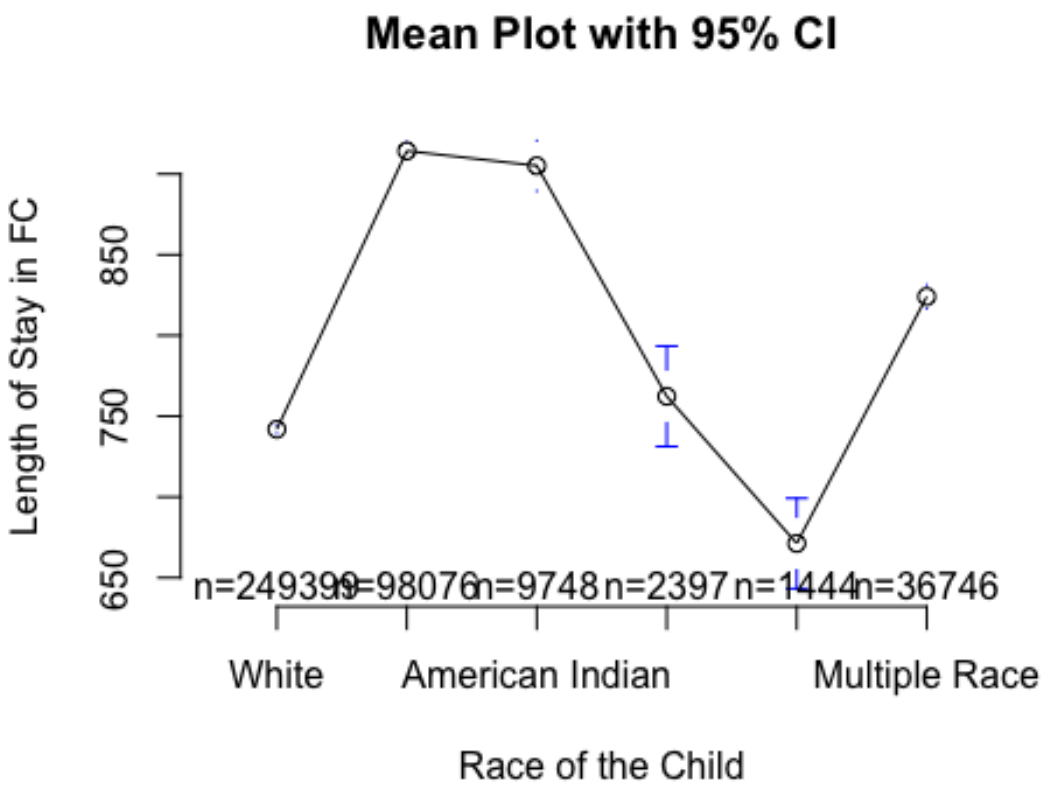
boxplot

```
boxplot(lifelos ~ race, data = foster.one.way,
        xlab = "Race of the Child", ylab = "Length of Stay in FC",
        frame = FALSE)
```



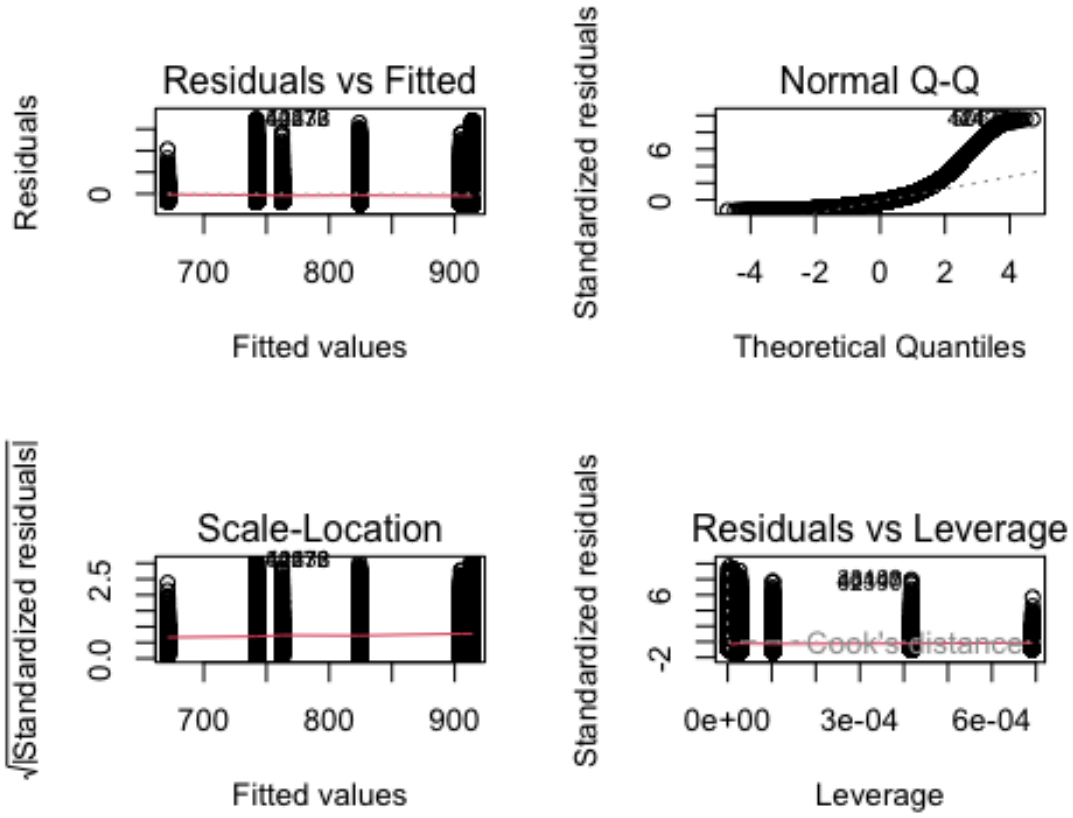
plotmeans

```
plotmeans(lifelos ~ race, data=foster.one.way, frame = FALSE,
          xlab = "Race of the Child", ylab = "Length of Stay in FC",
          main = "Mean Plot with 95% CI")
```



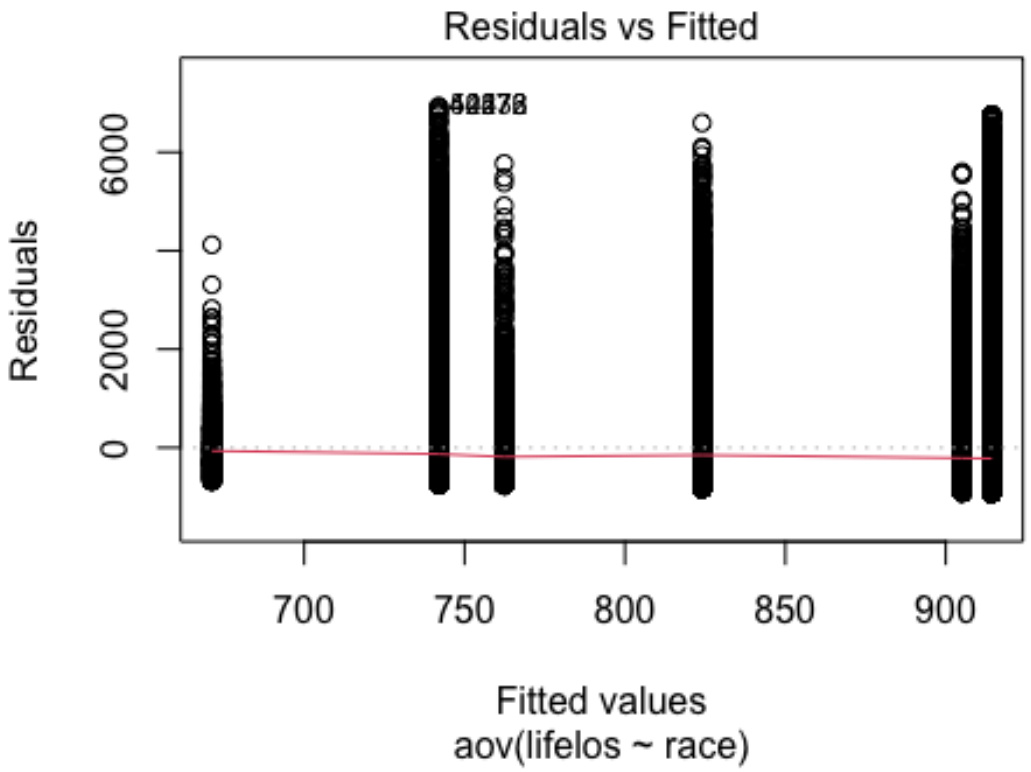
diagnostic

```
par(mfrow=c(2,2))
plot(one.way1)
```



residuals vs fitted

```
plot(one.way1, 1)
```



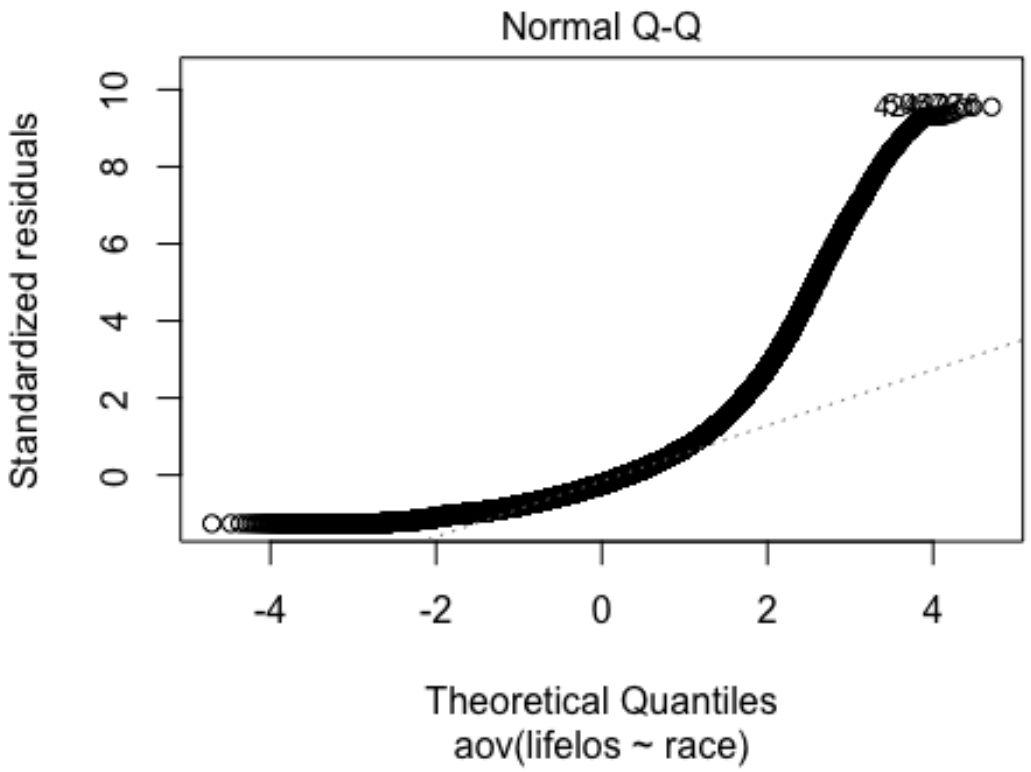
levenes test

```
leveneTest(lifelos ~ race, data = foster.one.way)
```

```
Levene's Test for Homogeneity of Variance (center = median)
          Df F value    Pr(>F)
group      5  699.57 < 2.2e-16 ***
      397804
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

normality

```
plot(one.way1, 2)
```



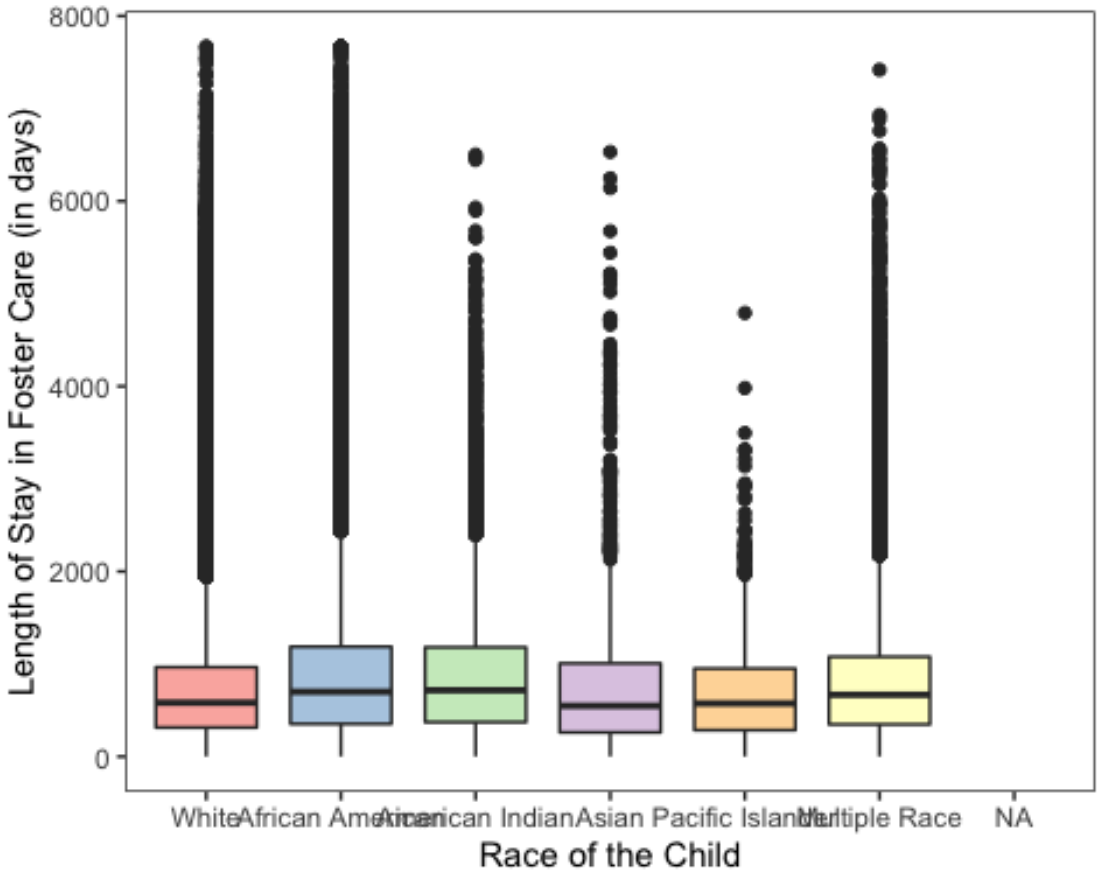
nonparametric

```
kruskal.test(lifelos ~ race, data = foster.one.way)
```

```
    Kruskal-Wallis rank sum test

data:  lifelos by race
Kruskal-Wallis chi-squared = 3102.7, df = 5, p-value < 2.2e-16
```

boxplot

```
ggplot(foster.one.way, aes(race, lifelos)) +
  geom_boxplot(aes(fill = race), show.legend = FALSE) +
  labs(x="Race of the Child", y="Length of Stay in Foster Care (in days)") +
  theme_bw() +
  theme(panel.grid.major = element_blank(), panel.grid.minor =
element_blank()) +
  #geom_text(data = foster.one.way, aes(x = race, y = lifelos), size = 3,
vjust=-1, hjust =-1) +
  scale_fill_brewer(palette = "Pastel1")
```


8000
6000
4000
2000
0
Length of Stay in Foster Care (in days)
White
African American
American Indian
Asian Pacific Islander
Multiple Race
NA
Race of the Child