

# Lab 4: Does Prenatal Care Improve Infant Health?

*Carmen Easterwood, James Nguyen*

*April 23, 2017*

## Introduction

```
load("bwght_w203.Rdata")
library(lattice)
library(caret)

## Warning: package 'caret' was built under R version 3.3.3
## Loading required package: ggplot2
## Warning: package 'ggplot2' was built under R version 3.3.3
library(lmtest)

## Loading required package: zoo
## Warning: package 'zoo' was built under R version 3.3.3
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric
library(sandwich)

## Warning: package 'sandwich' was built under R version 3.3.3
library(corrplot)

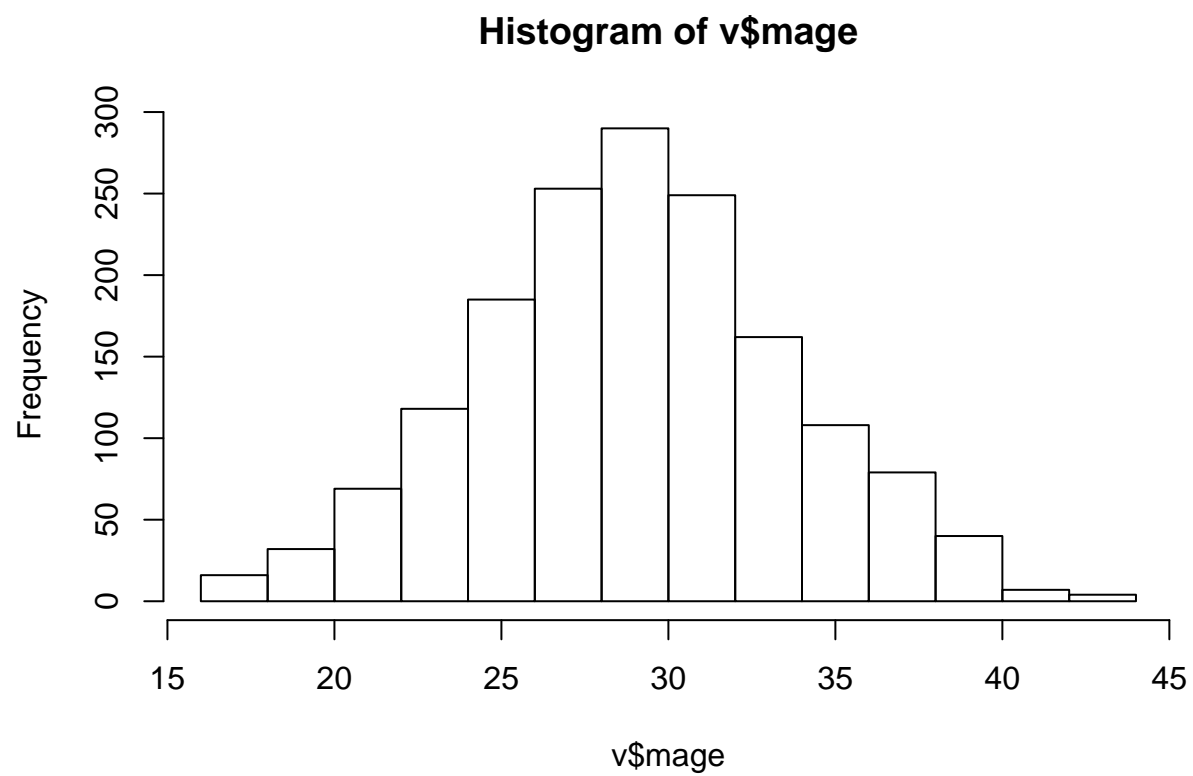
## Warning: package 'corrplot' was built under R version 3.3.3
```

## Exploratory Data Analysis

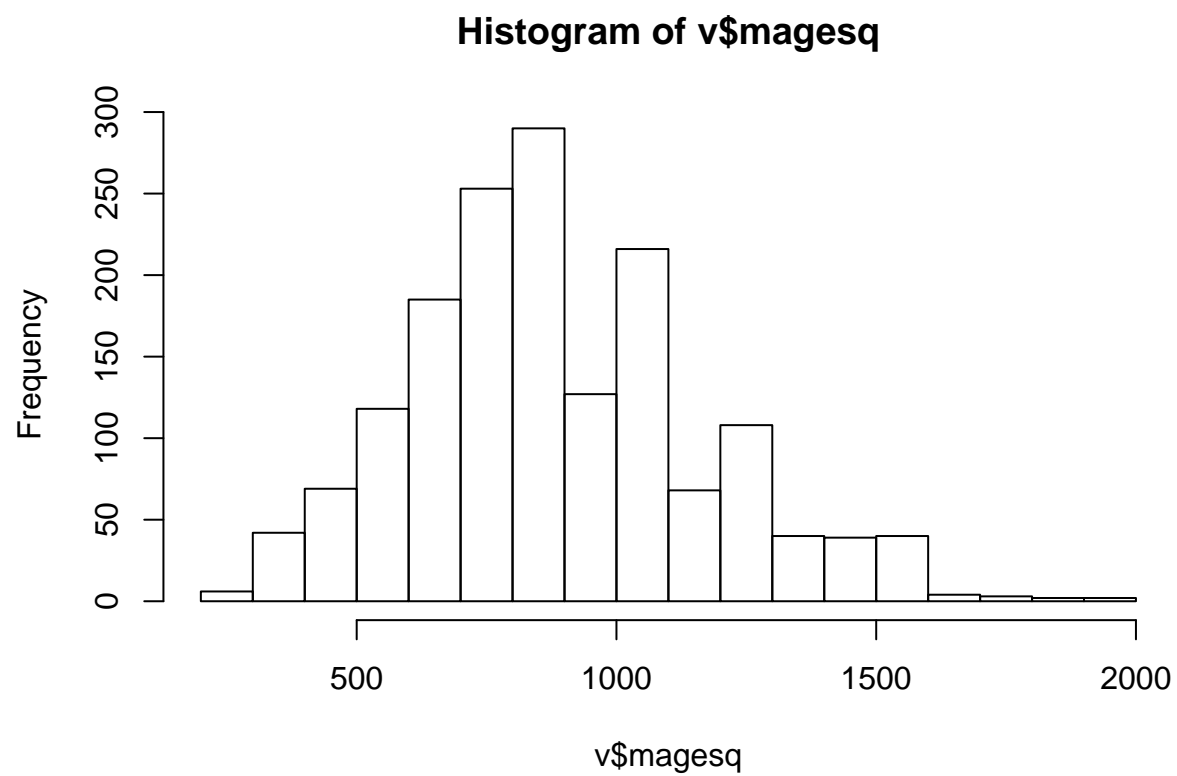
```
### Clean up NA values
v = na.exclude(data, complete.cases(v))

# Might want to beef up this section

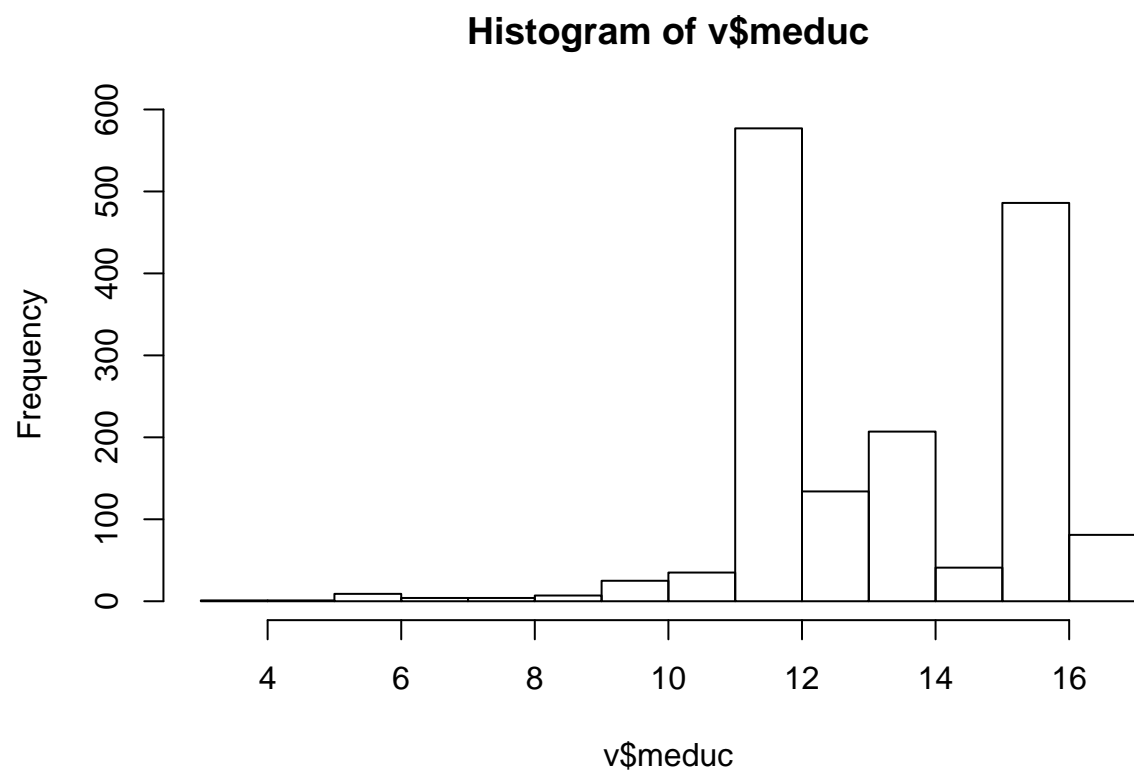
#Mom
hist(v$mage) #Very normal, avg late 20s
```



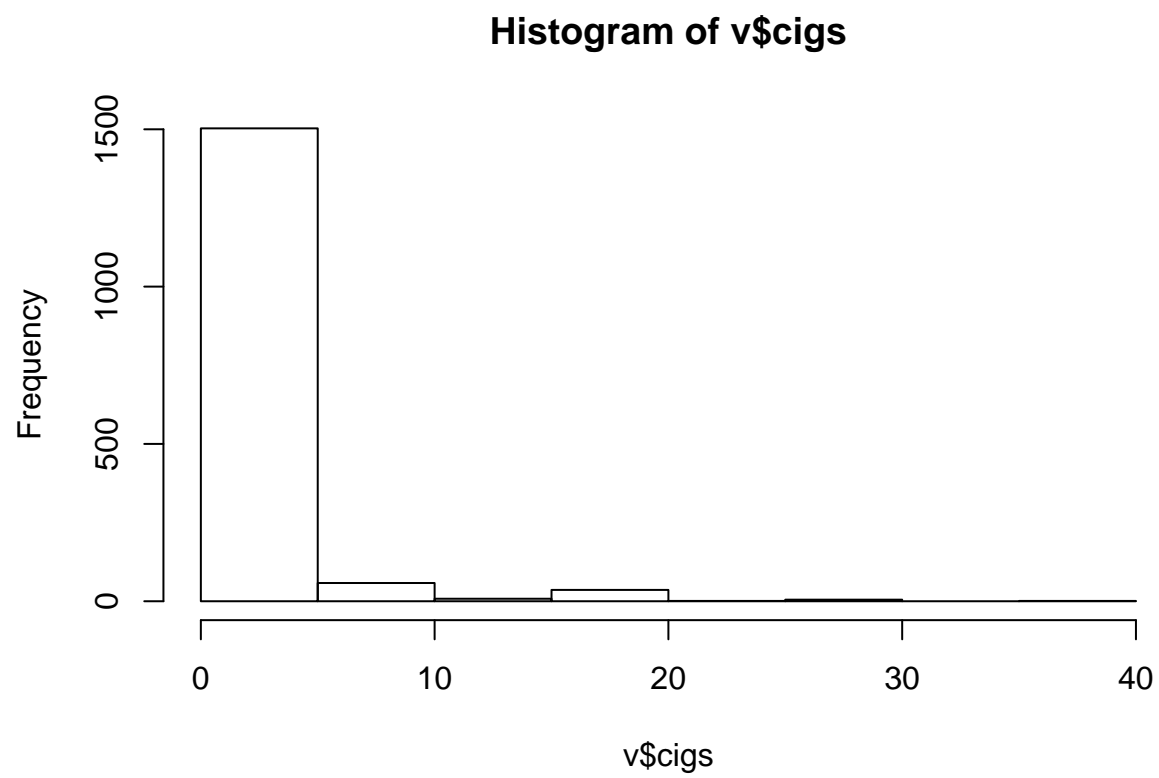
```
hist(v$magesq) #Slightly right skewed
```



```
hist(v$meduc) #Spikes at 12 and 16
```

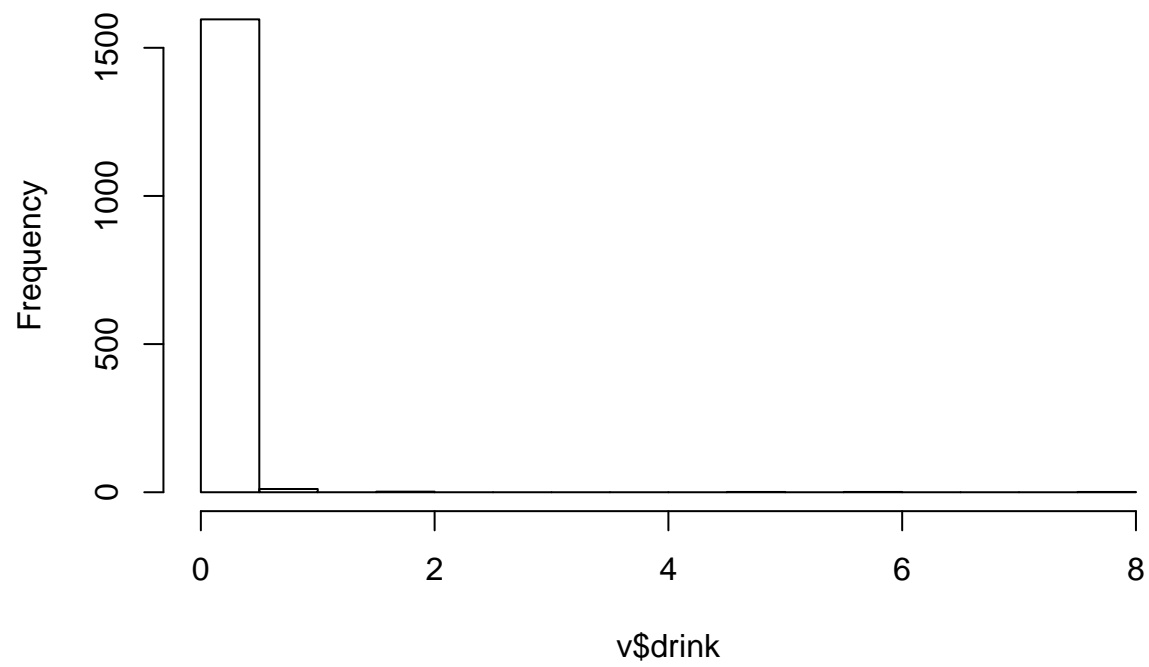


```
hist(v$cigs) #Big mass at 0. Right skewed. Above 20, corresponds to packs.
```

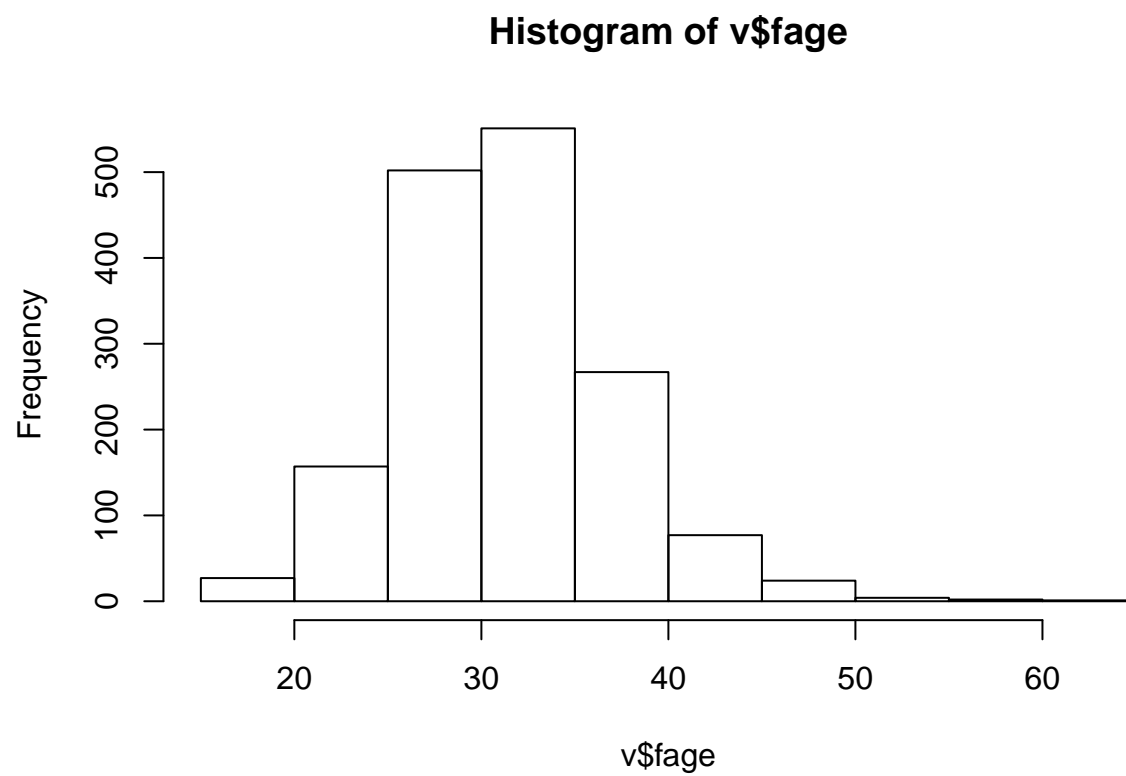


```
hist(v$drink) #Big mass at 0. Right skewed.
```

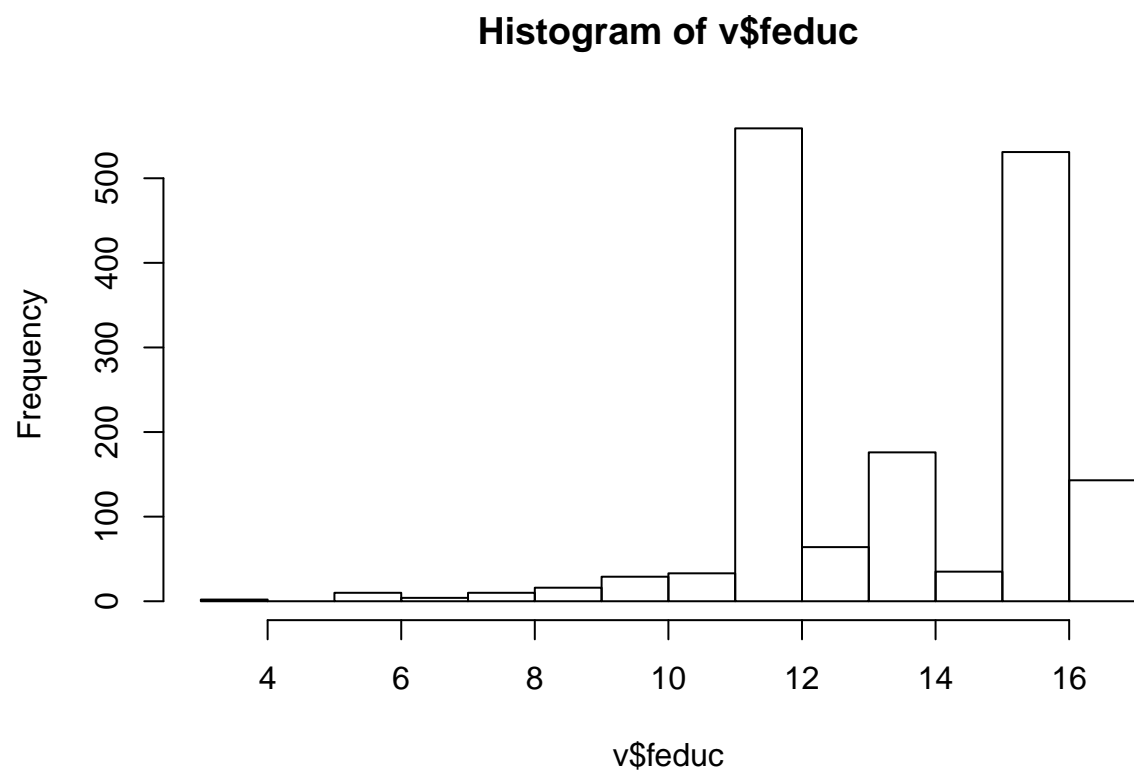
**Histogram of v\$drink**



```
#Dad  
hist(v$fage) #Slight right skew, avg early 30s
```



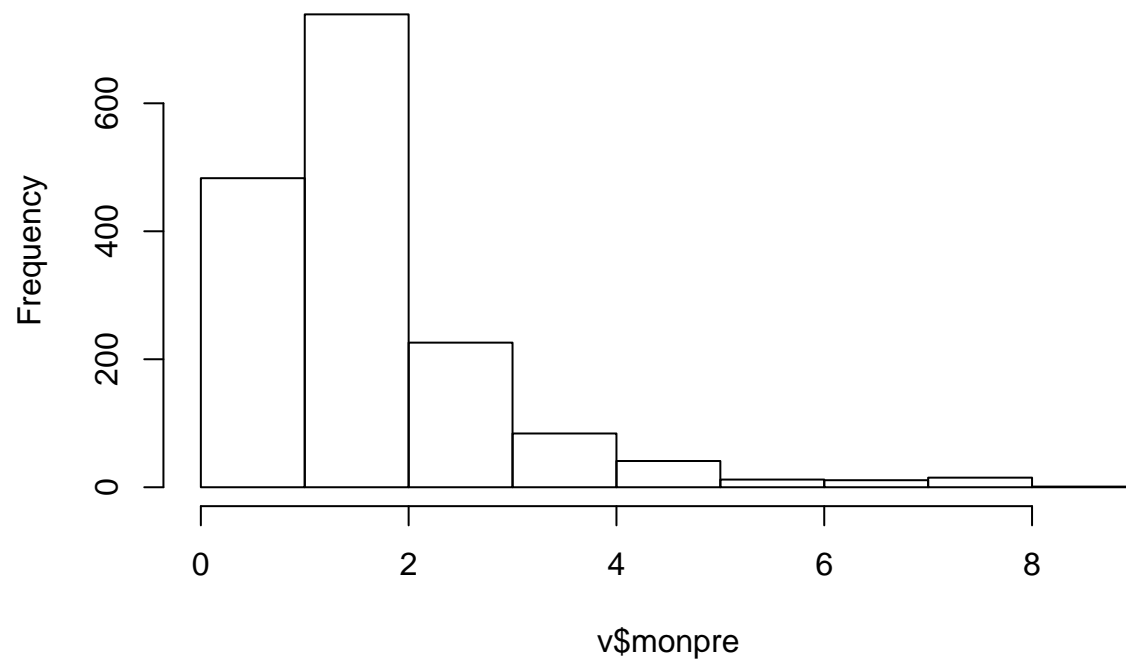
```
hist(v$feduc) #Spikes at 12 and 16
```



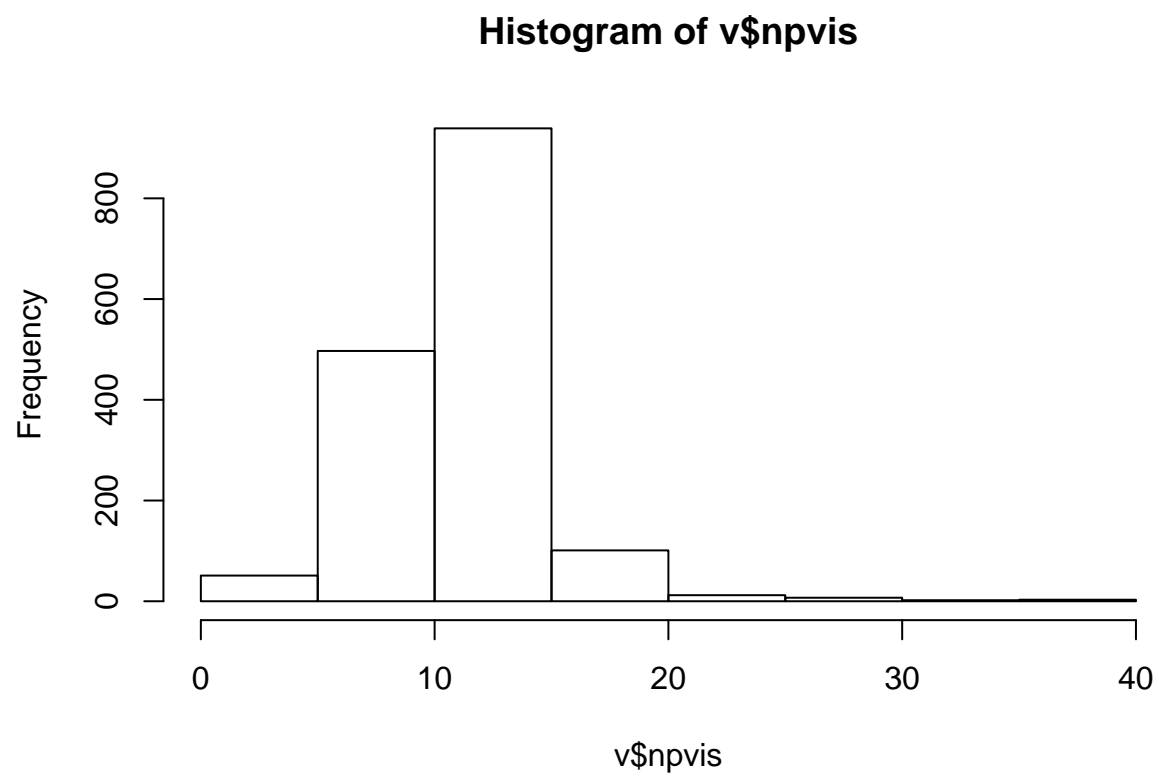
```
#Prenatal care  
hist(v$monpre) #Majority begin in first 2 months; right skewed
```



**Histogram of v\$monpre**

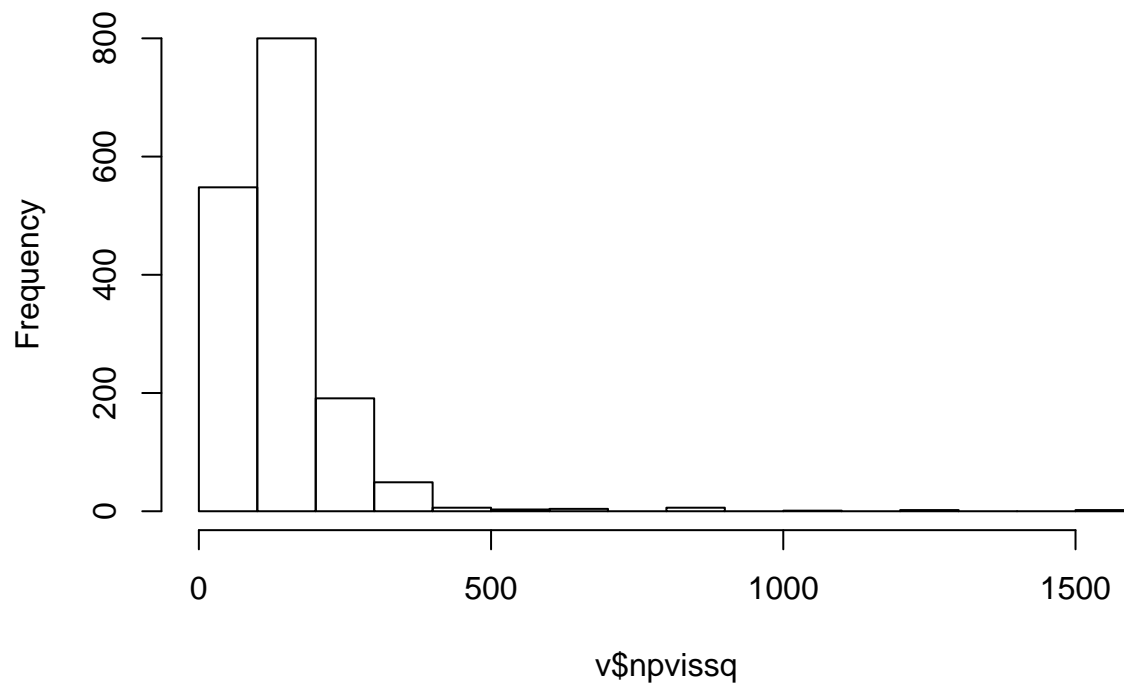


```
hist(v$npvis) #Spike at 10-15 visits
```



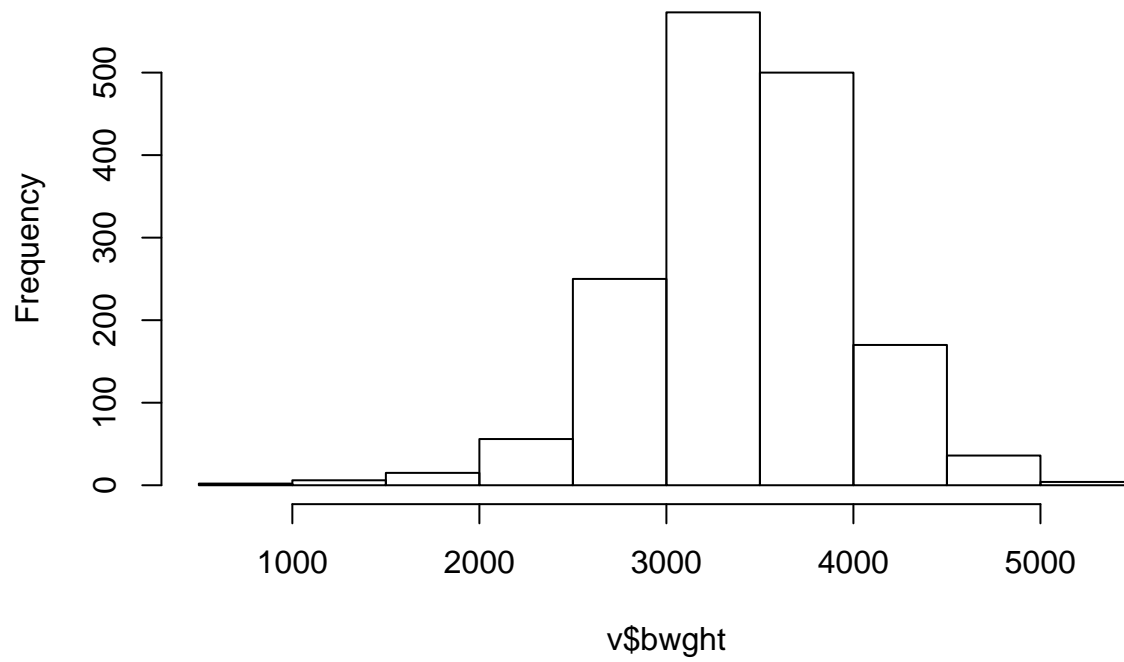
```
hist(v$npvissq) #Right skewed
```

**Histogram of v\$npvissq**



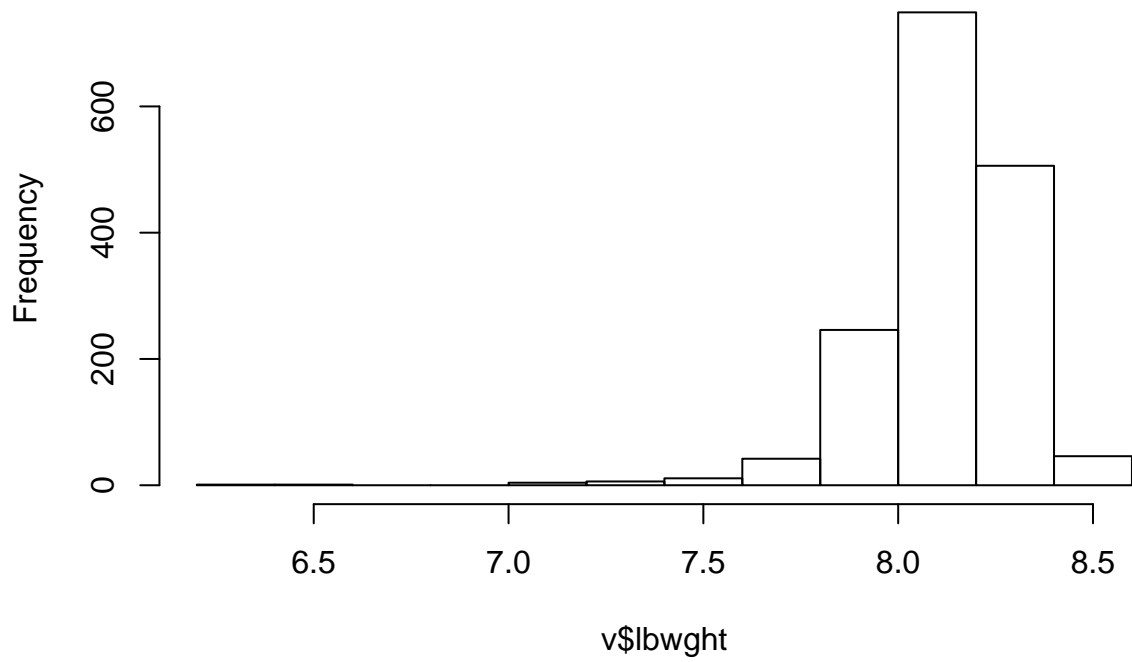
```
#Baby  
hist(v$bwght) #Normalish, avg 3000-3500 grams (6.6-7.7 lbs)
```

**Histogram of v\$bwght**



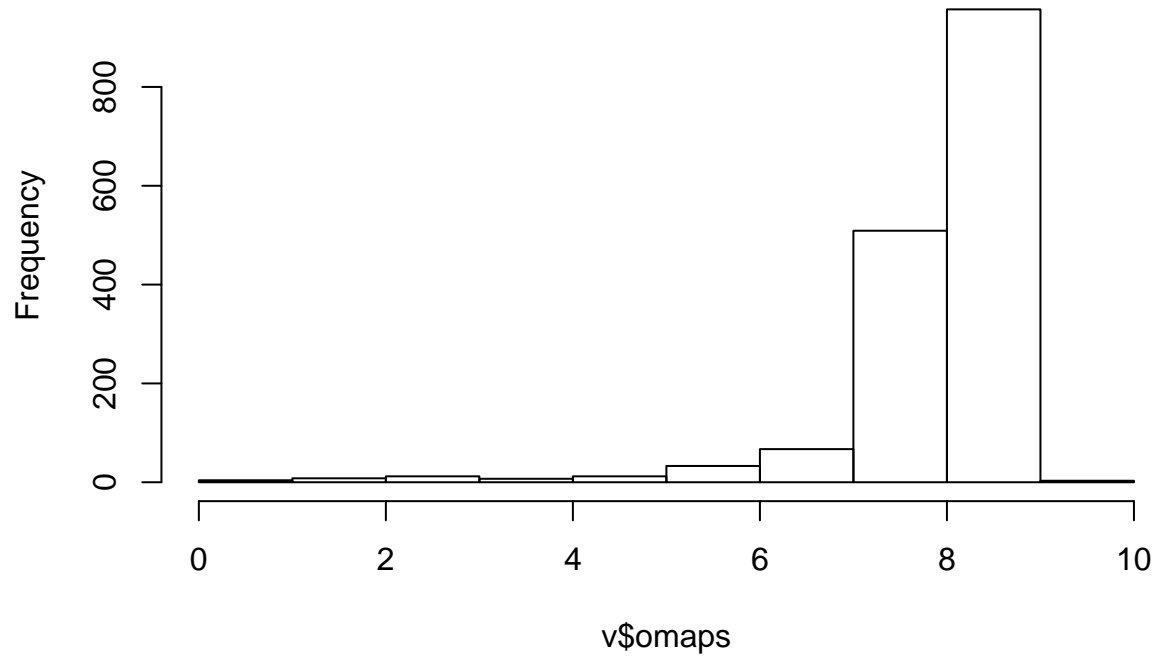
```
hist(v$lbwght) #Bit left skewed
```

**Histogram of v\$lbwght**



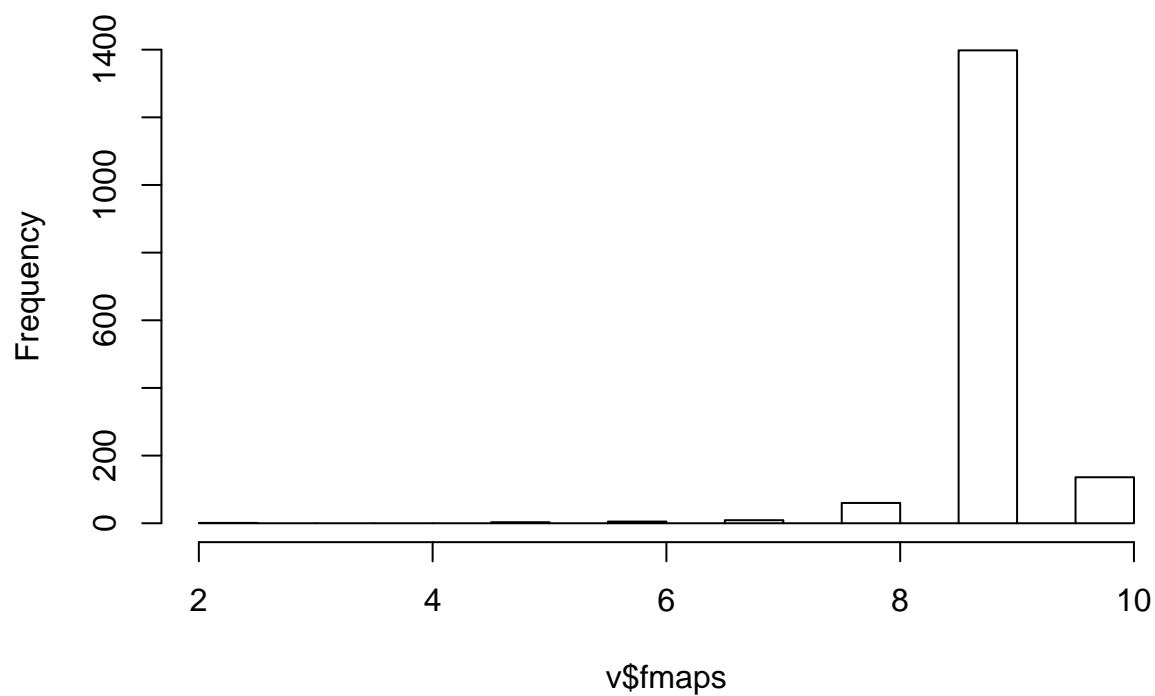
```
hist(v$omaps) #Left skewed, almost all 8 or 9
```

## Histogram of v\$omaps



```
hist(v$fmaps) #Left skewed, almost all 9
```

## Histogram of v\$fmaps



```
table(v$lbw) #Only 30 lbw (1.6%)
```

```
##
##    0    1
## 1589  23
```

```
table(v$vlbw) #Only 13 vlbw (0.7%)
```

```
##
##    0    1
## 1604    8
```

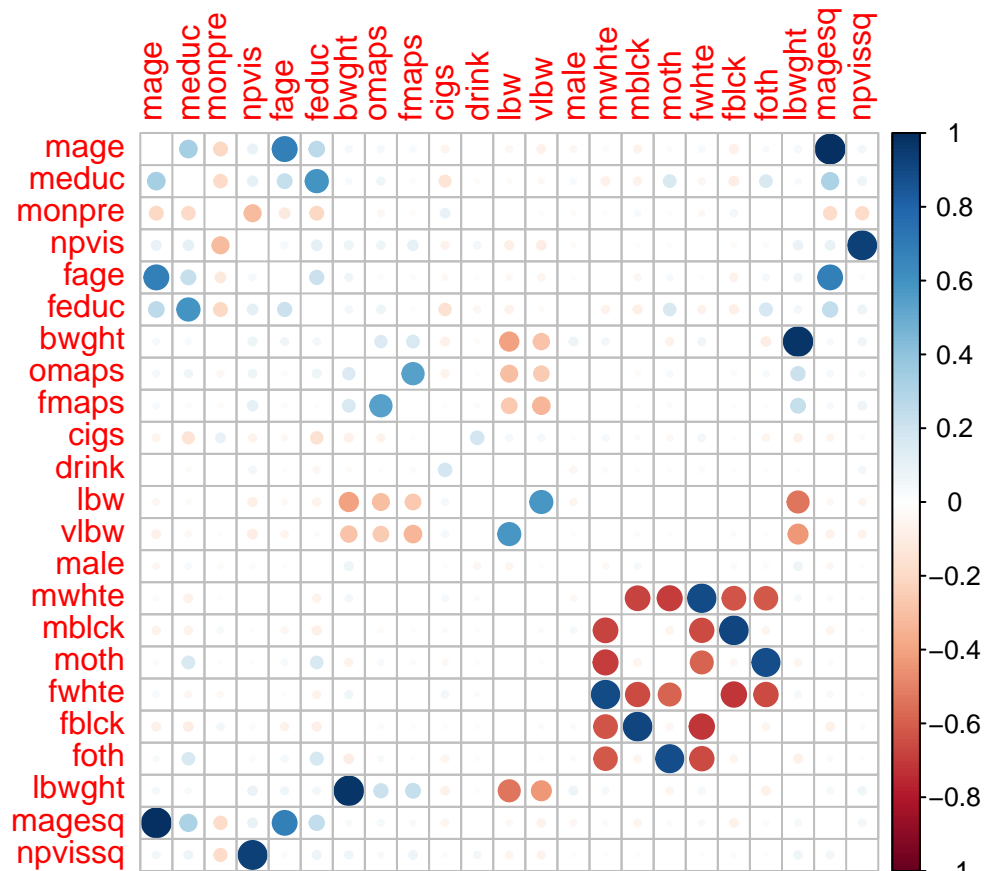
```
table(v$male) #941 male (51.4%)
```

```
##
##    0    1
## 784 828
```

## Correlation analysis

We build a correlation matrix to identify pairs of variables that have high correlation relationship.

```
cor=cor(v)
corrplot(cor, method="circle",diag = FALSE)
```



Looking at the correlation matrix below, we can see that father and mother tend to be in the same race, father and mother age have high correlation, father's education and mother education...

```
#Race Table
mrace <- as.factor(c())
levels(mrace) <- c("white", "black", "other")

for (i in 1:length(v$bwght)){
  if (v[i,]$mwhte == 1) {mrace[i] <- "white"}
  else if (v[i,]$mblick == 1) {mrace[i] <- "black"}
  else if (v[i,]$moth == 1) {mrace[i] <- "other"}
}

frace <- as.factor(c())
levels(frace) <- c("white", "black", "other")

for (i in 1:length(v$bwght)){

  if (v[i,]$fwhte == 1) {frace[i] <- "white"}
  else if (v[i,]$fblick == 1) {frace[i] <- "black"}
  else if (v[i,]$foth == 1) {frace[i] <- "other"}
}

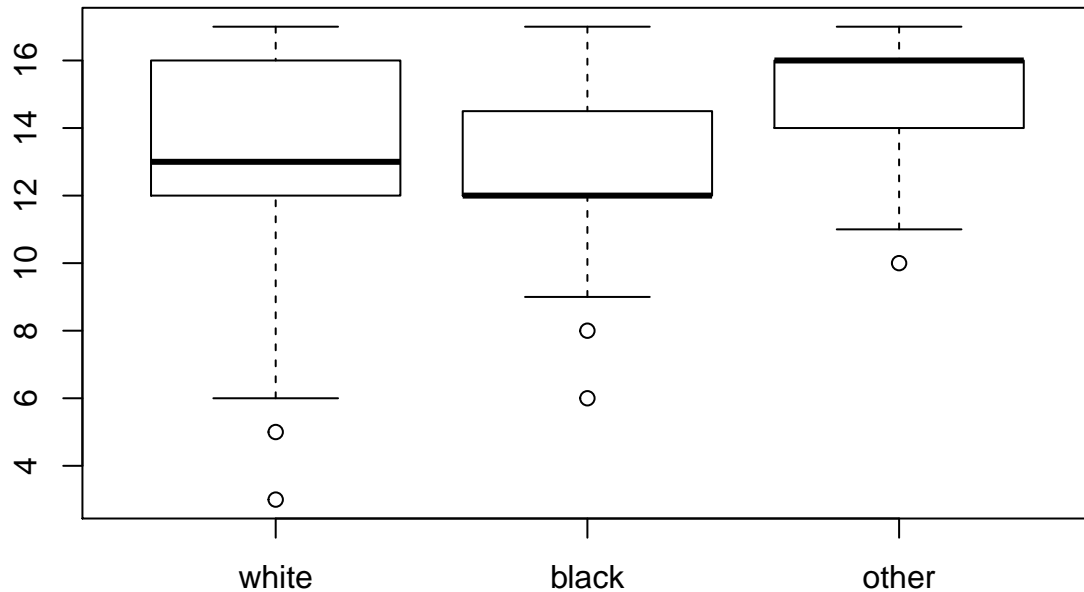
table(mrace, frace)

##          frace
## mrace  white black other
```



```
##   white 1420    8    3
##   black   5   83    0
##   other  16    1   76
```

```
# Could do a boxplot by race, e.g.:
boxplot(meduc ~ mrace, data = v)
```



```
# But black & other plots don't have a lot of data
```

## Model Specifications

Explain: - We are excluding lbw, vlbw - Using white as the baseline category for race because 90% of people are white (is data for other 2 races reliable?) - Using bwght as outcome and not omaps or fmaps

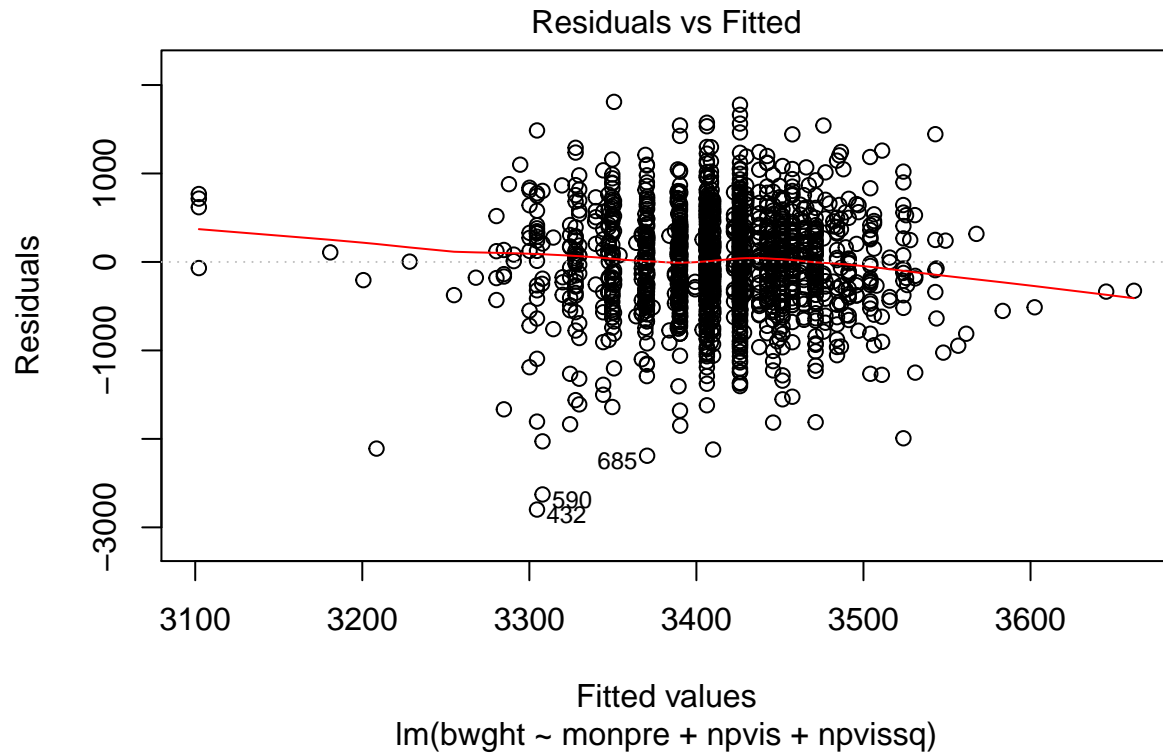
### Model 1

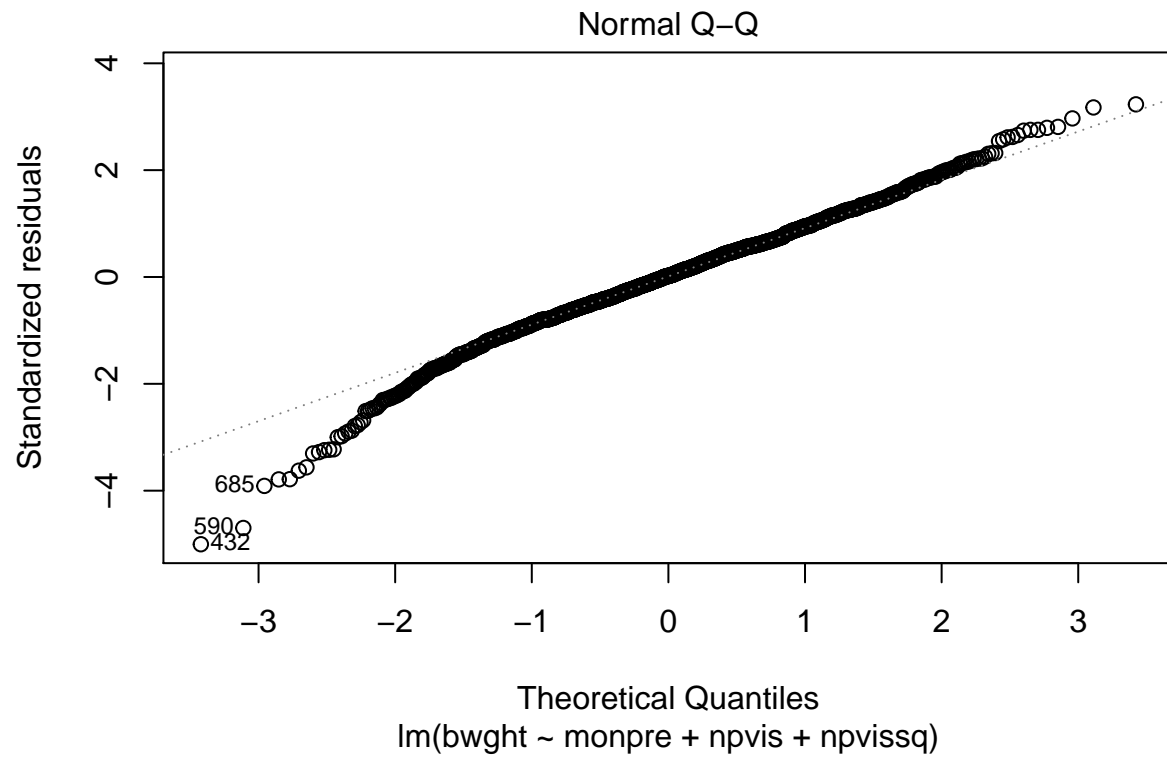
```
modell1 <- lm(bwght ~ monpre + npvis + npvissq, data = v)
coeftest(modell1, vcov = vcovHC) # All significant; concerned that monpre coefficient is positive
```

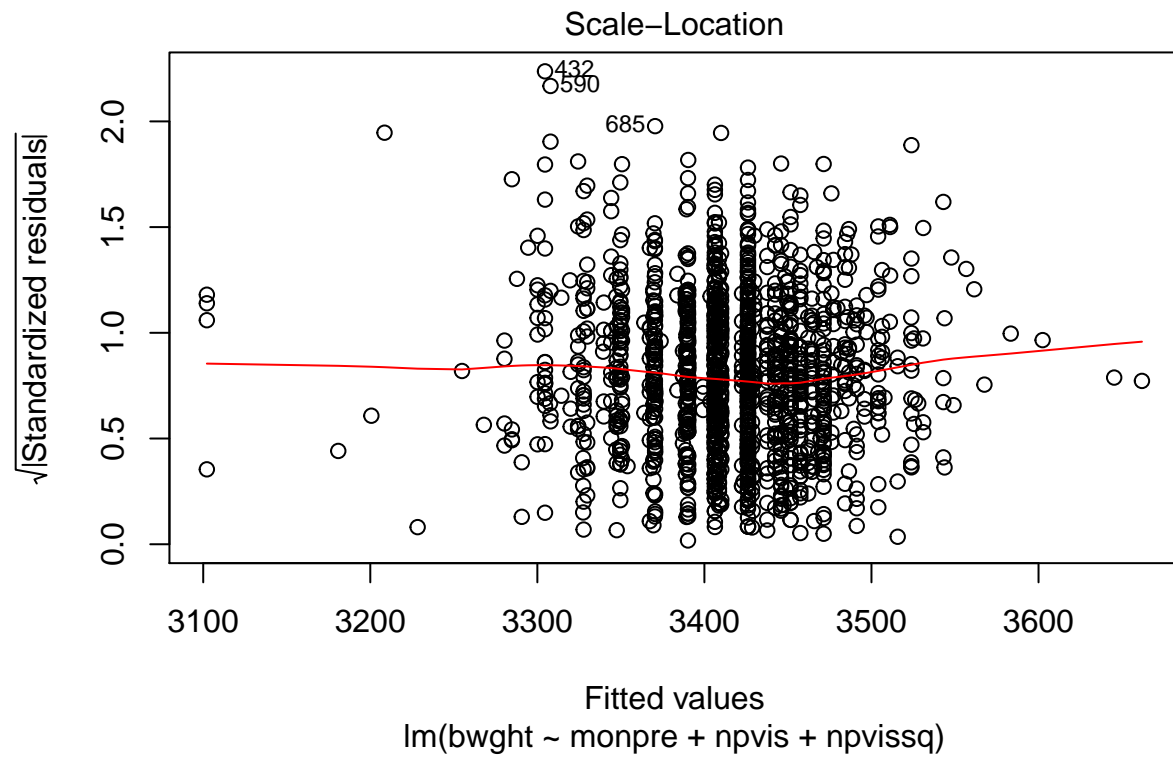
```
##
## t test of coefficients:
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3102.16475  126.10443  24.6000  < 2e-16 ***
```

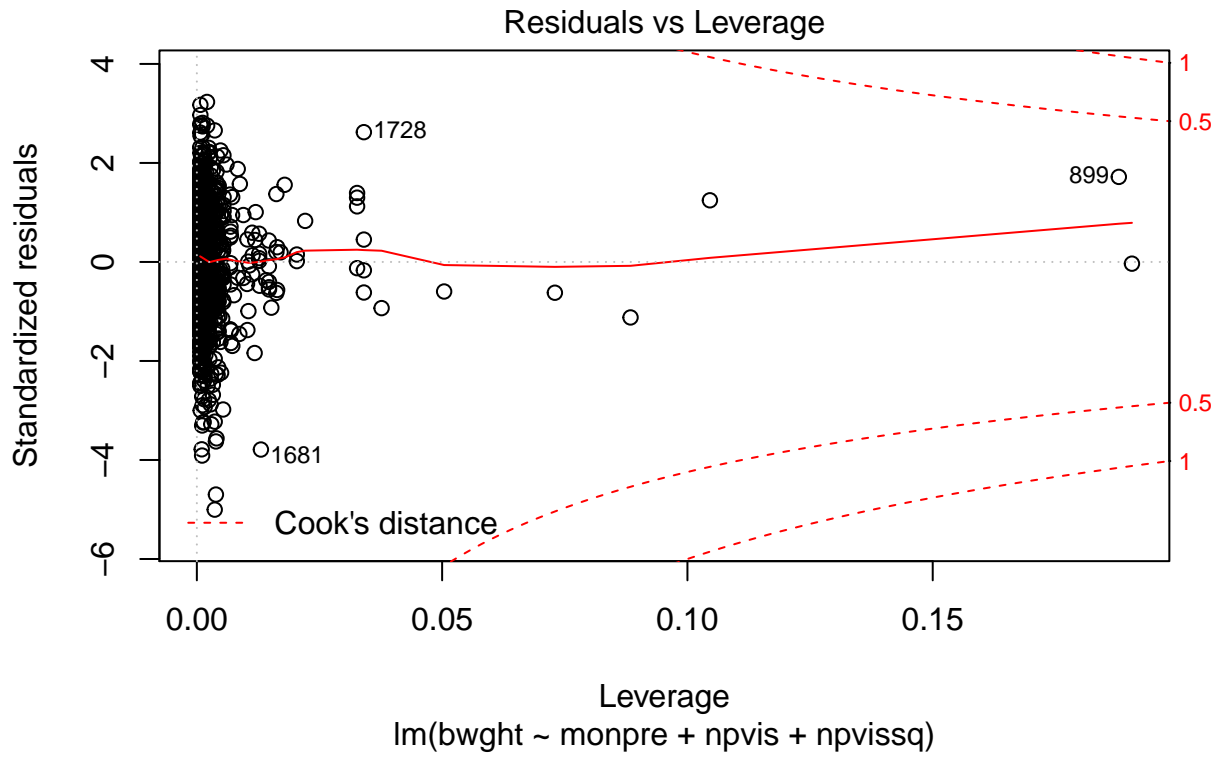
```
## monpre      19.81582    13.02410    1.5215    0.12834
## npvis       30.57030    14.22104    2.1497    0.03173 *
## npvissq     -0.57321     0.44954   -1.2751    0.20246
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(model1) # Meets assumptions relatively well
```









```
AIC(model1)
```

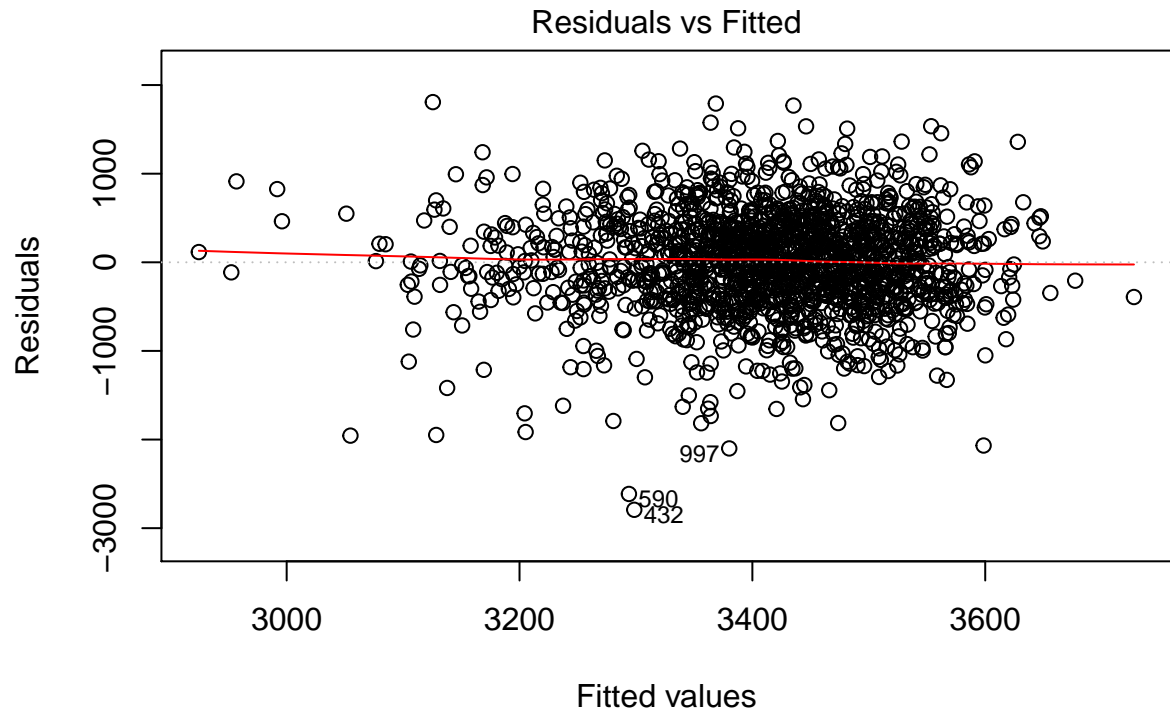
```
## [1] 24984.34
```

## Model 2

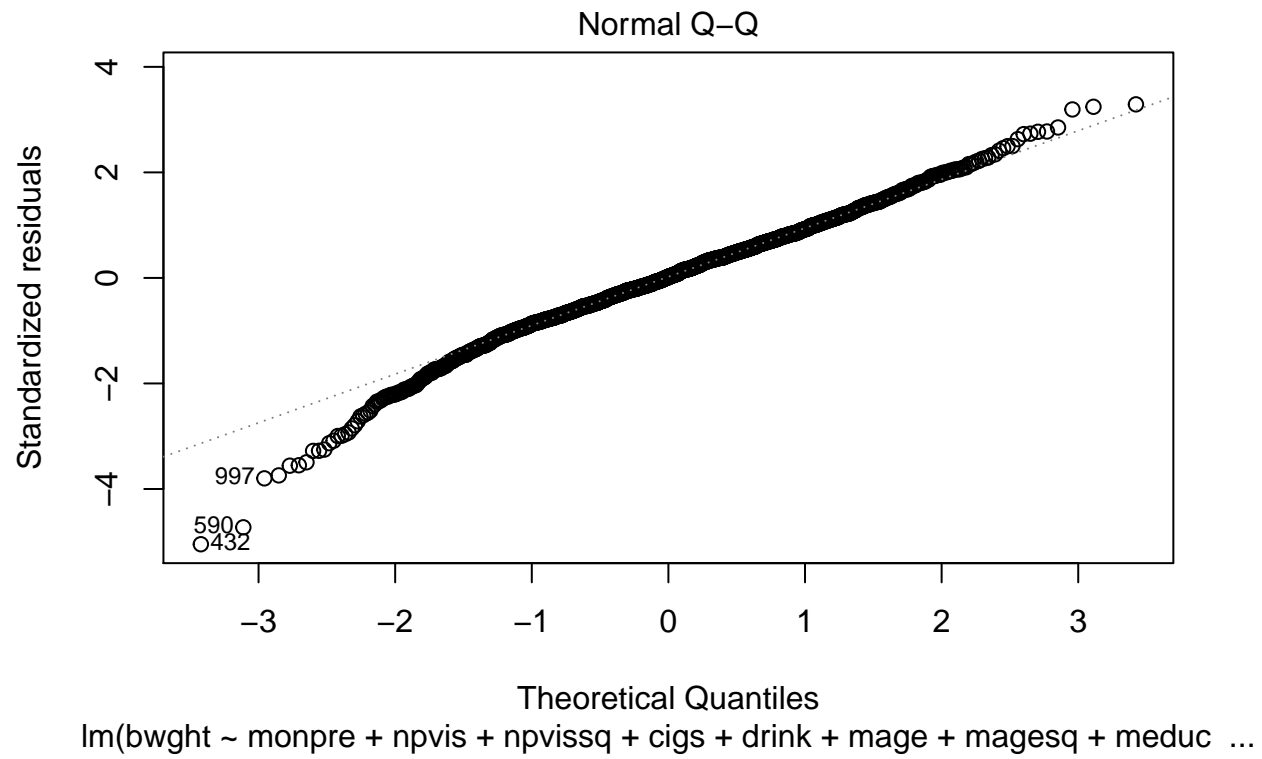
```
model2 <- lm(bwght ~ monpre + npvis + npvissq + cigs + drink + mage + magesq + meduc + fage + feduc + male + mblck)
coeftest(model2, vcov = vcovHC) # Significant vars: monpre (still positive), npvis, cigs, mage, male, m
```

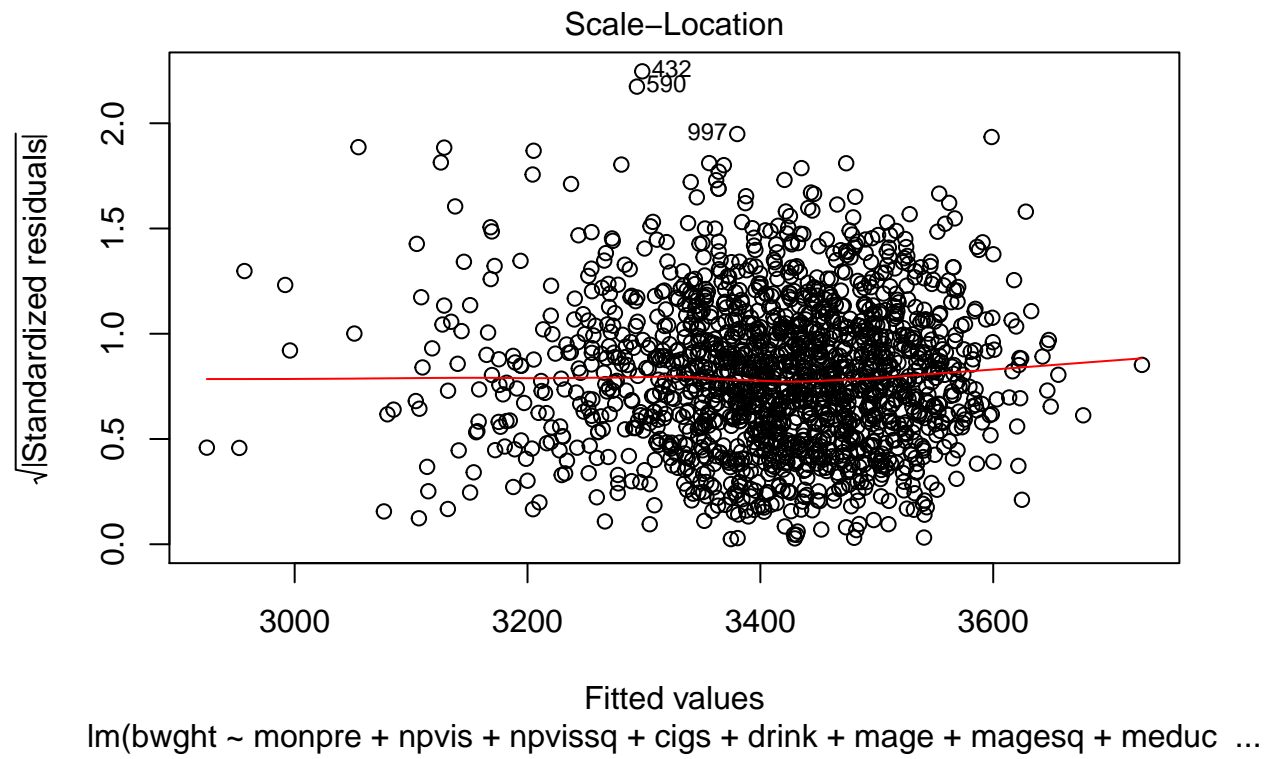
```
##
## t test of coefficients:
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1744.64168  472.98570  3.6886 0.0002331 ***
## monpre       33.72256   13.39644  2.5173 0.0119237 *
## npvis        30.68329   13.88498  2.2098 0.0272591 *
## npvissq      -0.57453    0.42602 -1.3486 0.1776568
## cigs         -9.03534    3.61764 -2.4976 0.0126043 *
## drink       -15.17192   30.82896 -0.4921 0.6226936
## mage         70.63267   30.34938  2.3273 0.0200727 *
## magesq       -1.18919    0.49269 -2.4137 0.0159050 *
## meduc        -2.47179    8.69871 -0.2842 0.7763275
## fage          5.42568    3.52899  1.5375 0.1243789
## feduc        10.05383    8.17409  1.2300 0.2188922
## male         93.80395   28.13067  3.3346 0.0008738 ***
## mblck       -27.40350   57.09202 -0.4800 0.6313015
```

```
## moth          -184.31007    57.27474 -3.2180 0.0013168 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
plot(model2) # Meets assumptions relatively well
```

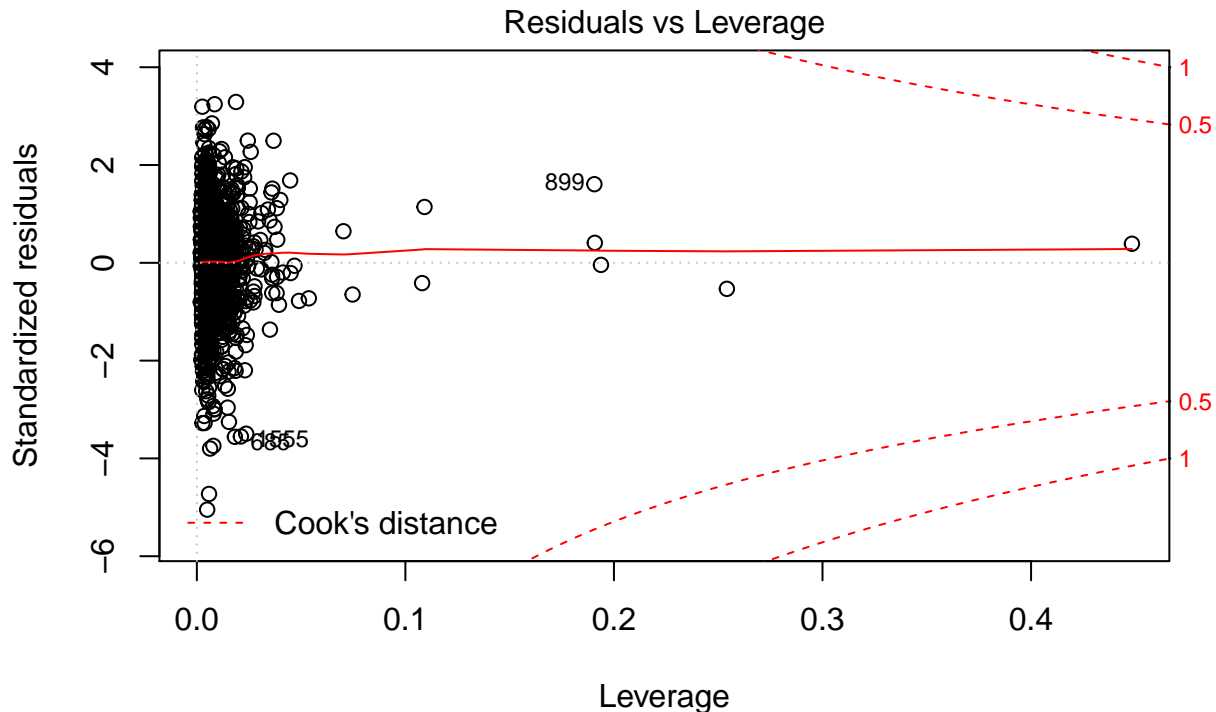


lm(bwght ~ monpre + npvis + npvissq + cigs + drink + mage + magesq + meduc ...









lm(bwght ~ monpre + npvis + npvissq + cigs + drink + mage + magesq + meduc ...

```
AIC(model2) # Lower than model 1
```

```
## [1] 24961.06
```

### Model 3

```
model3 <- lm(bwght ~ monpre + npvis + npvissq + cigs + drink + mage + magesq + meduc + fage + feduc + m
coefest(model3, vcov = vcovHC) # Same significant vars as model 2 except: mother race no longer signif
```

```
##
```

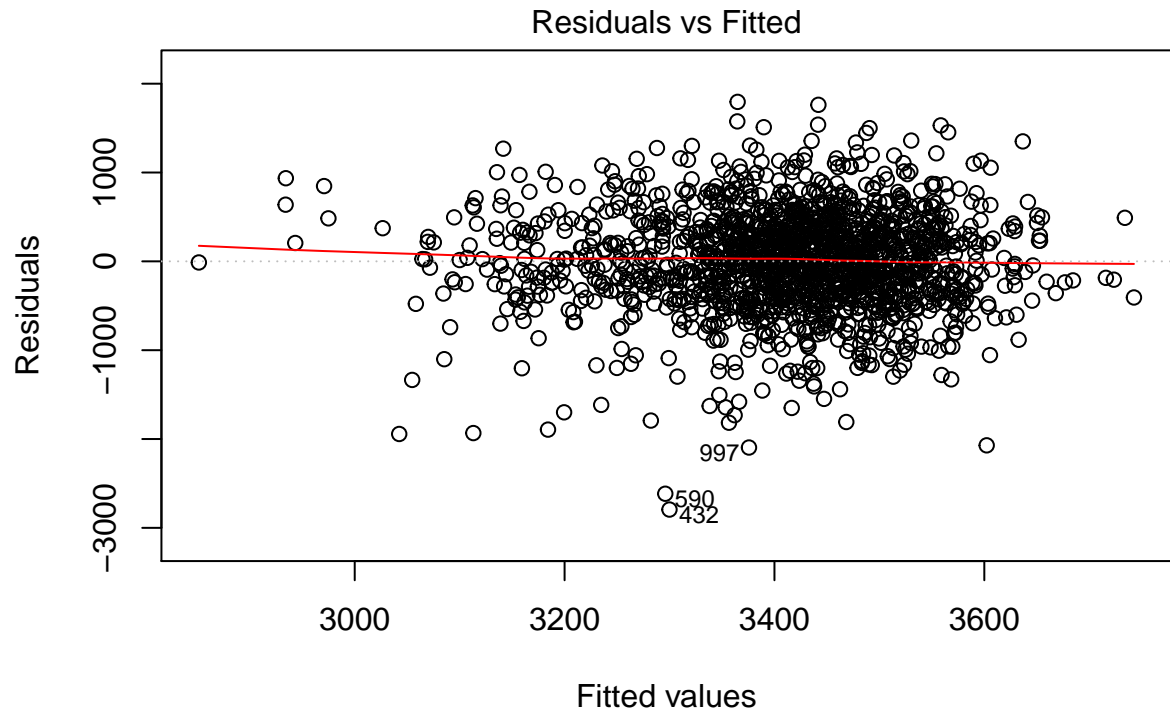
```
## t test of coefficients:
```

```
##
```

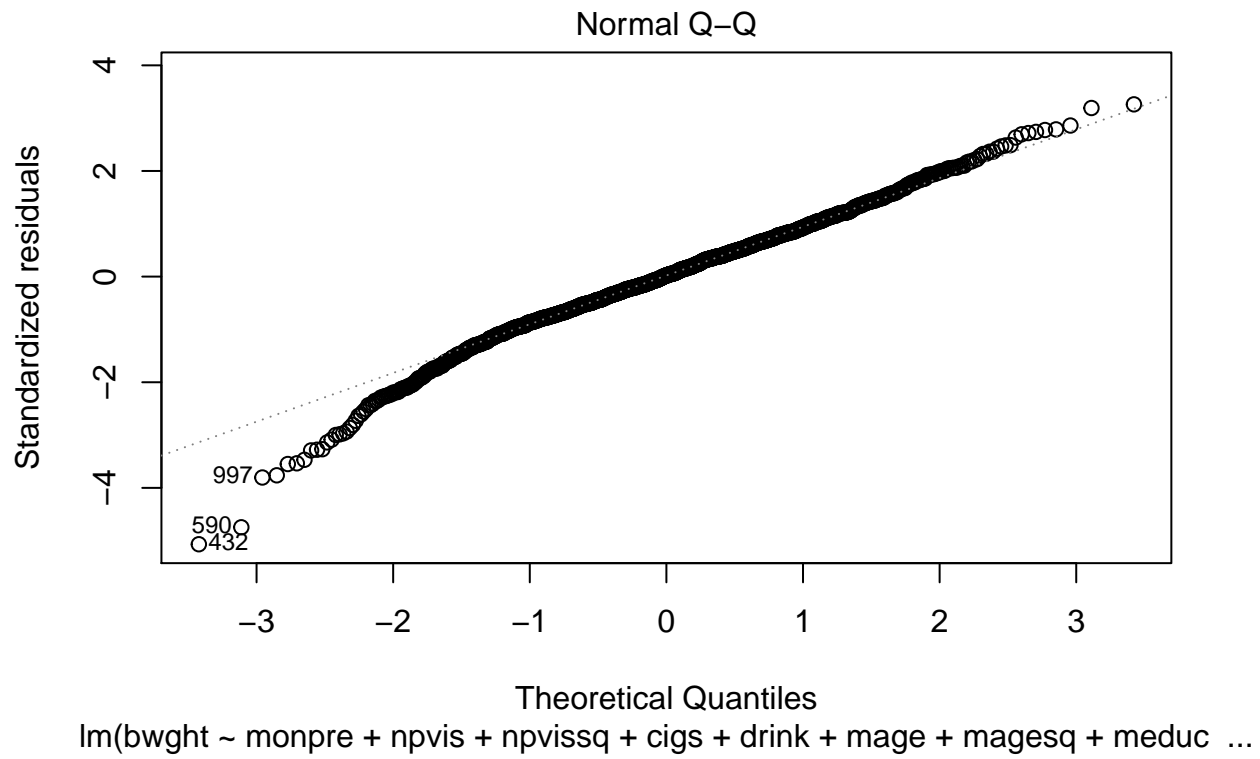
	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	1621.78301	475.83257	3.4083	0.0006699	***
monpre	35.10252	13.44988	2.6099	0.0091422	**
npvis	31.44915	13.87828	2.2661	0.0235805	*
npvissq	-0.58576	0.42517	-1.3777	0.1684916	
cigs	-9.51933	3.65746	-2.6027	0.0093346	**
drink	-13.78866	30.72249	-0.4488	0.6536274	
mage	76.60951	30.46449	2.5147	0.0120105	*
magesq	-1.29042	0.49446	-2.6098	0.0091451	**
meduc	-1.99109	8.67486	-0.2295	0.8184914	
fage	5.68216	3.51242	1.6177	0.1059179	
feduc	11.08433	8.12071	1.3649	0.1724626	
male	90.00177	28.19723	3.1919	0.0014411	**
mblck	-205.39479	141.77361	-1.4488	0.1476033	

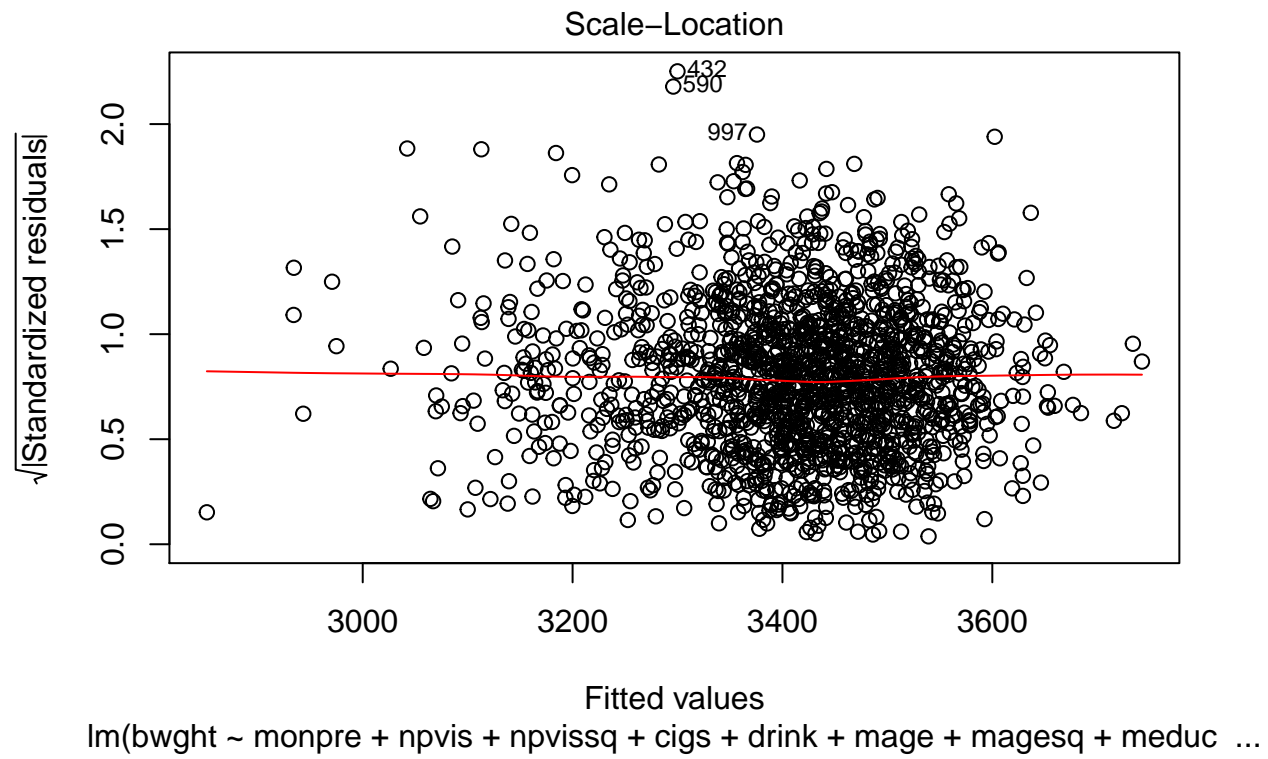
```
## moth      178.81985  108.05079   1.6550 0.0981290 .
## fblck     190.89194  135.38841   1.4100 0.1587470
## foth     -450.09894  116.82241  -3.8528 0.0001214 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

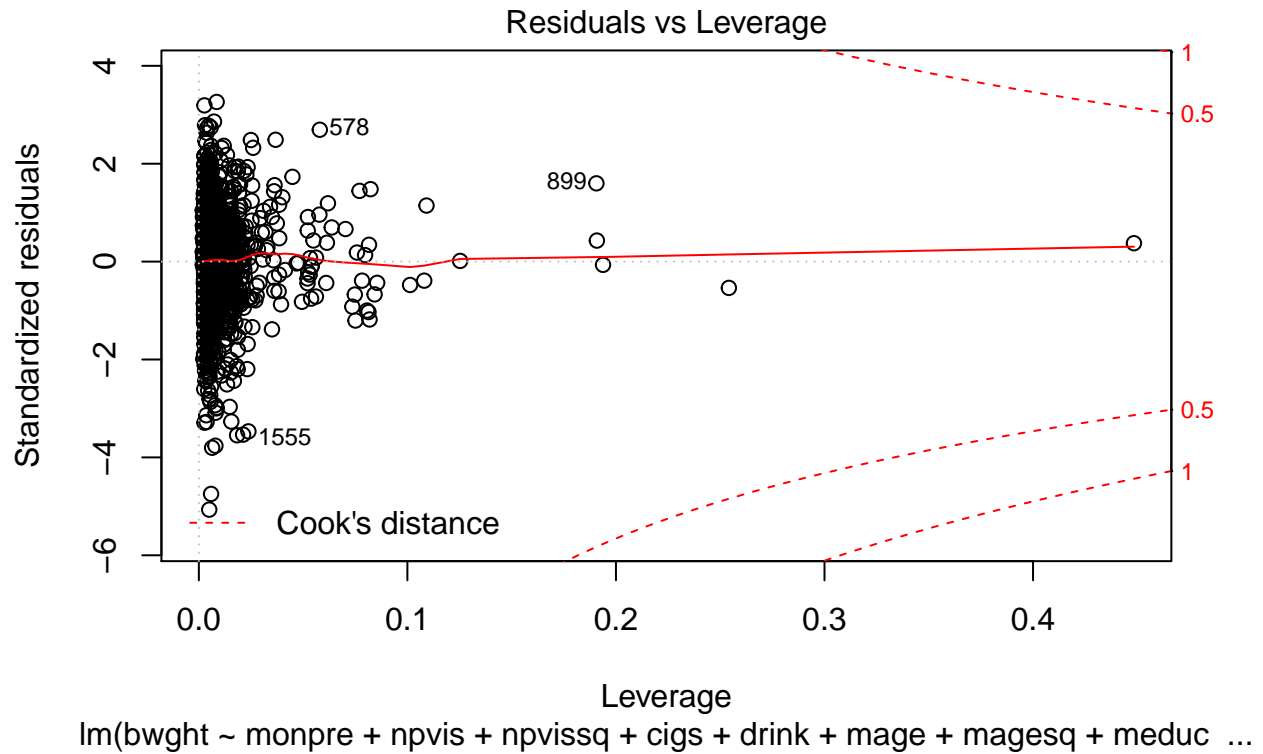
```
plot(model3) # Meets assumptions relatively well
```



lm(bwght ~ monpre + npvis + npvissq + cigs + drink + mage + magesq + meduc ...







```
AIC(model13) # Basically the same as model 2
```

```
## [1] 24951.87
```

## Assessment of CLM Assumptions

Model 1

Model 2

Model 3

## Summary of Results

## Causality

## Conclusion