

Poverty vs. Diabetes

Bret Harvestine

11/14/2016

R Markdown

Is poverty becoming a rising cause in the prevalence of Type II Diabetes?

```
#Remove any existing variables and setwd

rm(list=ls())
setwd('/Users/bretharvestine/Desktop/333')

#require packages

require(bls scrapeR) # get data

## Loading required package: blscrapeR

require(choroplethr) # make maps

## Loading required package: choroplethr

## Loading required package: acs

## Loading required package: stringr

## Loading required package: plyr

## Loading required package: XML

##
## Attaching package: 'acs'

## The following object is masked from 'package:base':
##      apply

require(dplyr) # wrangling

## Loading required package: dplyr

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:acs':
##      combine
```

```

## The following objects are masked from 'package:plyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

```

Load in Data

```

library("readxl")

#load in income data
median_house_income = read_excel("MedianHouseholdIncome.xlsx")
i = median_house_income #save as variable i

#load uninsured data
uninsured = read.csv("County_Data_2016.csv", header = TRUE)
u = uninsured #save as variable u

#load poverty rate data
poverty = read.csv("Poverty.csv", header = TRUE)

#load in diabetes data
diabetes2013 = read_excel("diabetes2014.xlsx")
d = diabetes2013 #save as variable d

#load access to grocery data
access = read_excel("access.xlsx")
a = access

#load leisure time spent physically inactive
leisure = read_excel("leisure_inactivity_prevalence.xlsx")
l = leisure

#load health sample data
health.sample.data =read_excel(
  "County_Ranked_healthdata.xlsx")
h= health.sample.data

#cbind data in order to merge
mydataI = cbind(region = as.numeric(i$FIPS_Code), value = i$Median_Household_Income_2014)

mydataD = cbind(region = as.numeric(d$FIPS_Code), value = as.numeric(d$Percent))

## Warning in cbind(region = as.numeric(d$FIPS_Code), value = as.numeric(d
## $Percent)): NAs introduced by coercion

```

```
mydataP = cbind(region = as.numeric(poverty$FIPS), value = poverty$Percent.Poverty)
head(mydataP)
```

```
##      region value
## [1,]    1000    91
## [2,]    1001    37
## [3,]    1003    44
## [4,]    1005   182
## [5,]    1007   133
## [6,]    1009    74
```

```
mydataA = cbind(region = as.numeric(a$FIPS), value = a$PCT_LACCESS_POP10)
```

```
mydataL = cbind(region = as.numeric(l$FIPS), value = as.numeric(l$percent))
```

```
## Warning in cbind(region = as.numeric(l$FIPS), value = as.numeric(l
## $percent)): NAs introduced by coercion
```

```
head(mydataL)
```

```
##      region value
## [1,]    1001  28.6
## [2,]    1003  22.3
## [3,]    1005  31.8
## [4,]    1007  33.9
## [5,]    1009  28.0
## [6,]    1011  31.7
```

```
mydataF = cbind(region = as.numeric(h$FIPS), value = h$Percent_Fast_Foods)
```

```
mydataU13 = cbind(region = as.numeric(u$county_fips), value = u$X2013.uninsured.rate)
```

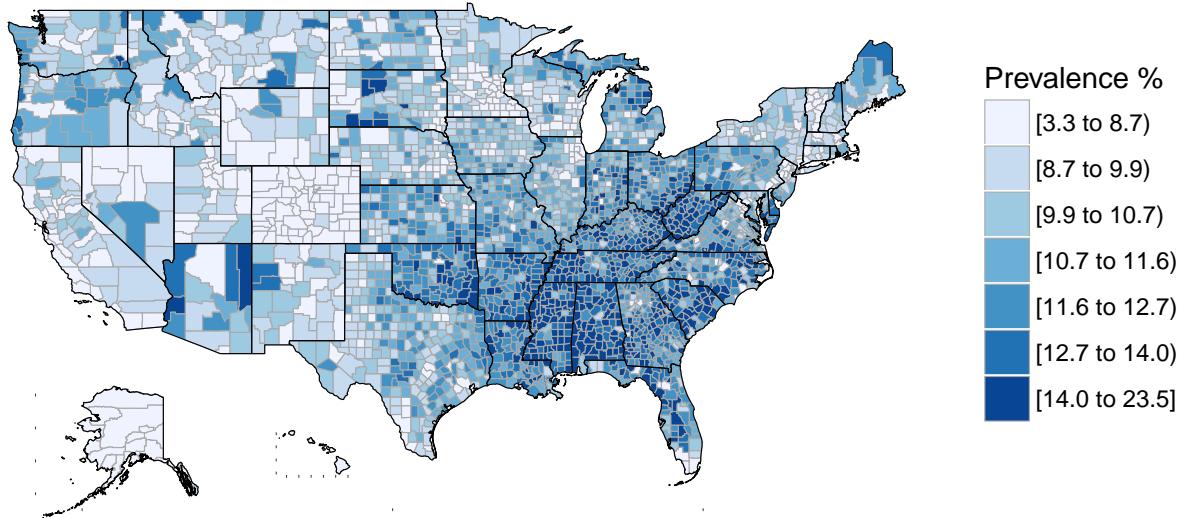
```
mydataU16 = cbind(region = as.numeric(u$county_fips), value = u$X2016.uninsured.rate)
```

Map Data by County

```
## Warning in super$initialize(map.df, user.df): Your data.frame contains the
## following regions which are not mappable: 2201, 2232, 2280, 72001, 72003,
## 72005, 72007, 72009, 72011, 72013, 72015, 72017, 72019, 72021, 72023,
## 72025, 72027, 72029, 72031, 72033, 72035, 72037, 72039, 72041, 72043,
## 72045, 72047, 72049, 72051, 72053, 72054, 72055, 72057, 72059, 72061,
## 72063, 72065, 72067, 72069, 72071, 72073, 72075, 72077, 72079, 72081,
## 72083, 72085, 72087, 72089, 72091, 72093, 72095, 72097, 72099, 72101,
## 72103, 72105, 72107, 72109, 72111, 72113, 72115, 72117, 72119, 72121,
## 72123, 72125, 72127, 72129, 72131, 72133, 72135, 72137, 72139, 72141,
## 72143, 72145, 72147, 72149, 72151, 72153
```

```
## Warning in self$bind(): The following regions were missing and are being
## set to NA: 51515
```

Prevalence of Diabetes by County 2013

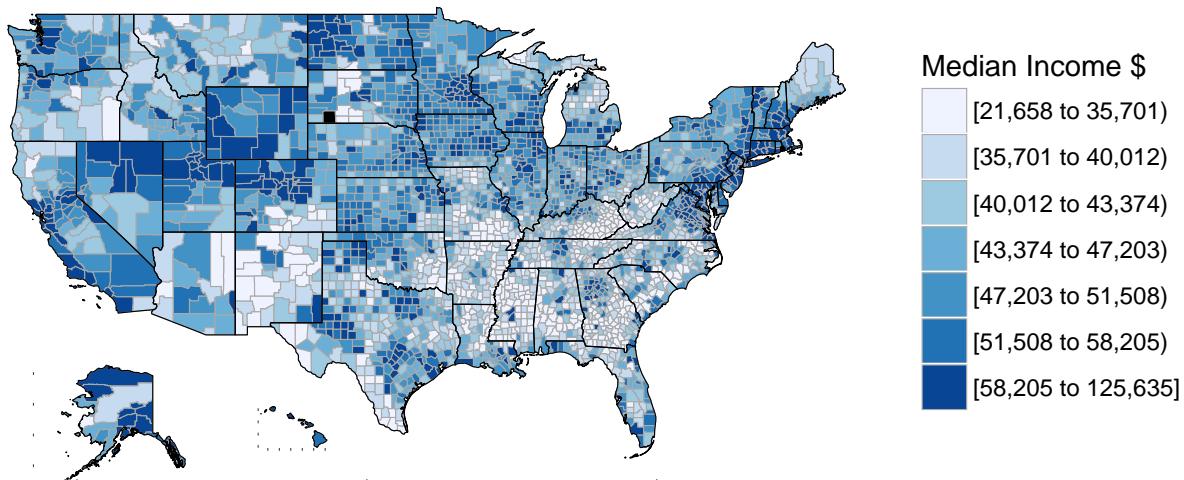


```
## Warning in super$initialize(map.df, user.df): Your data.frame contains the
## following regions which are not mappable: 1000, 2000, 2158, 2201, 2232,
## 2280, 4000, 5000, 6000, 8000, 9000, 10000, 11000, 12000, 13000, 15000,
## 16000, 17000, 18000, 19000, 20000, 21000, 22000, 23000, 24000, 25000,
## 26000, 27000, 28000, 29000, 30000, 31000, 32000, 33000, 34000, 35000,
## 36000, 37000, 38000, 39000, 40000, 41000, 42000, 44000, 45000, 46000,
## 46102, 47000, 48000, 49000, 50000, 51000, 53000, 54000, 55000, 56000, NA
```

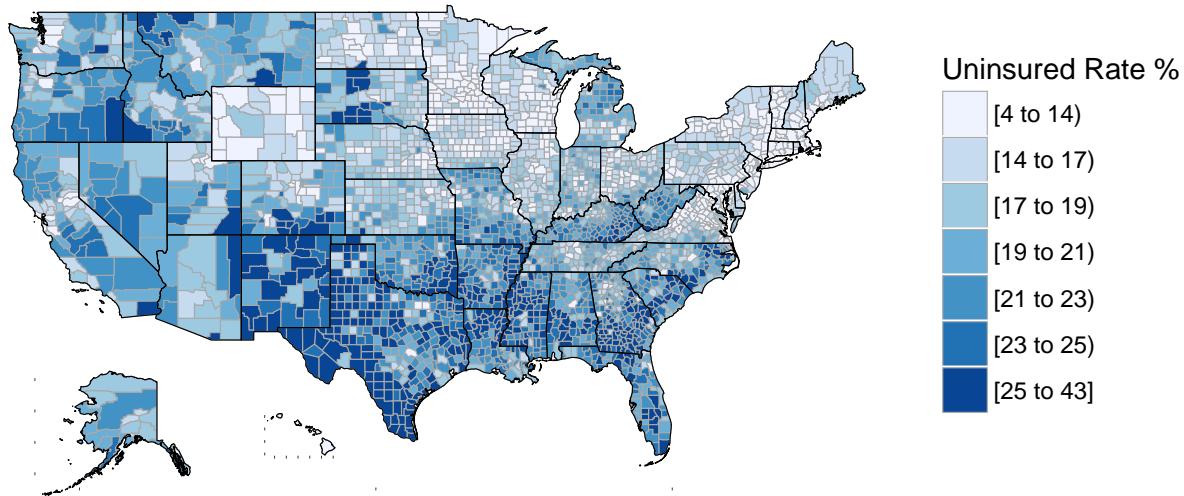


```
## Warning in self$bind(): The following regions were missing and are being
## set to NA: 46113, 15005, 51515
```

Median Household Income by County 2013

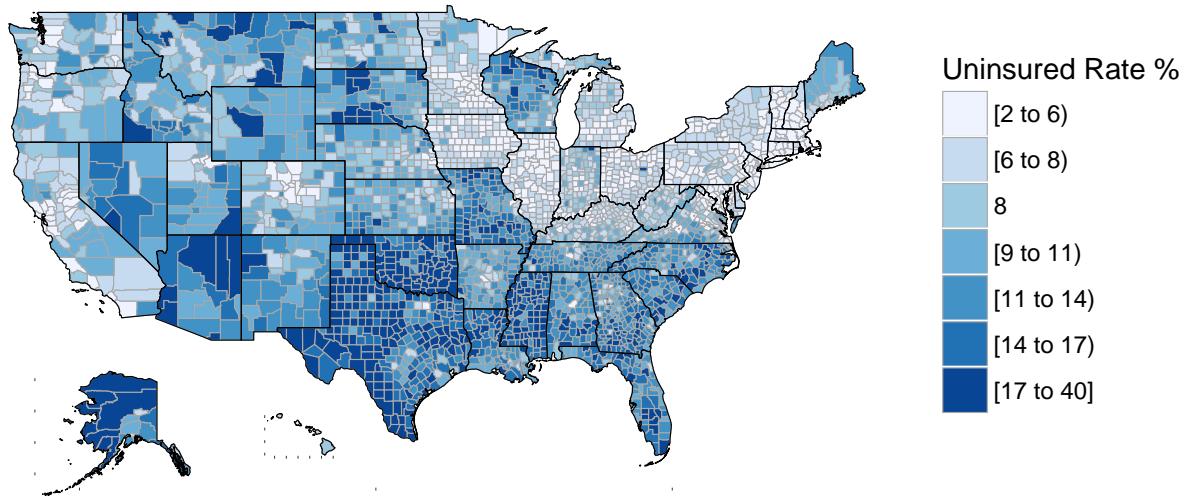


Uninsured Rate by County 2013



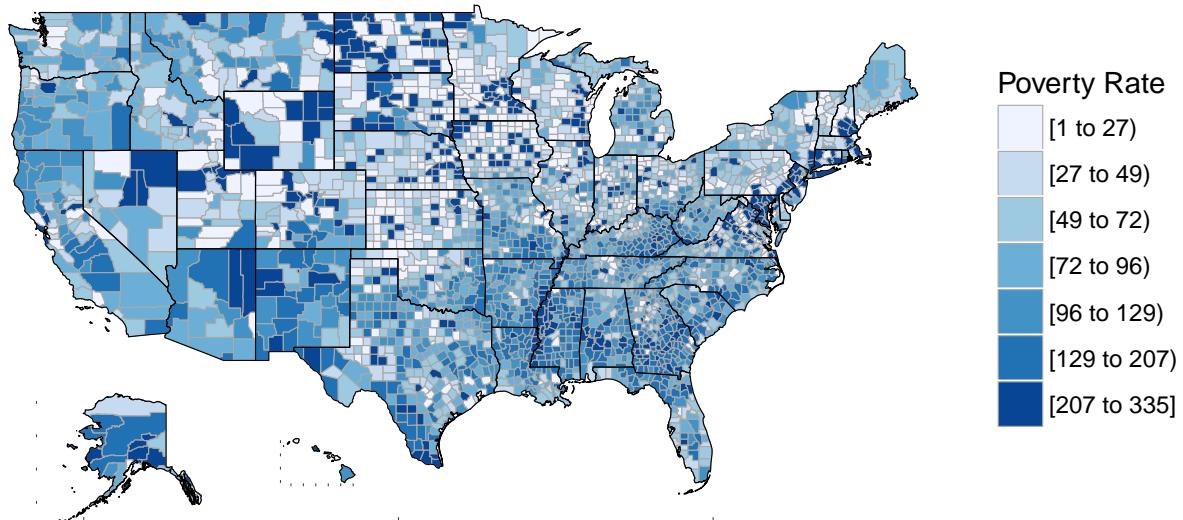
```
## Warning in self$bind(): The following regions were missing and are being
## set to NA: 51515
```

Uninsured Rate by County 2016



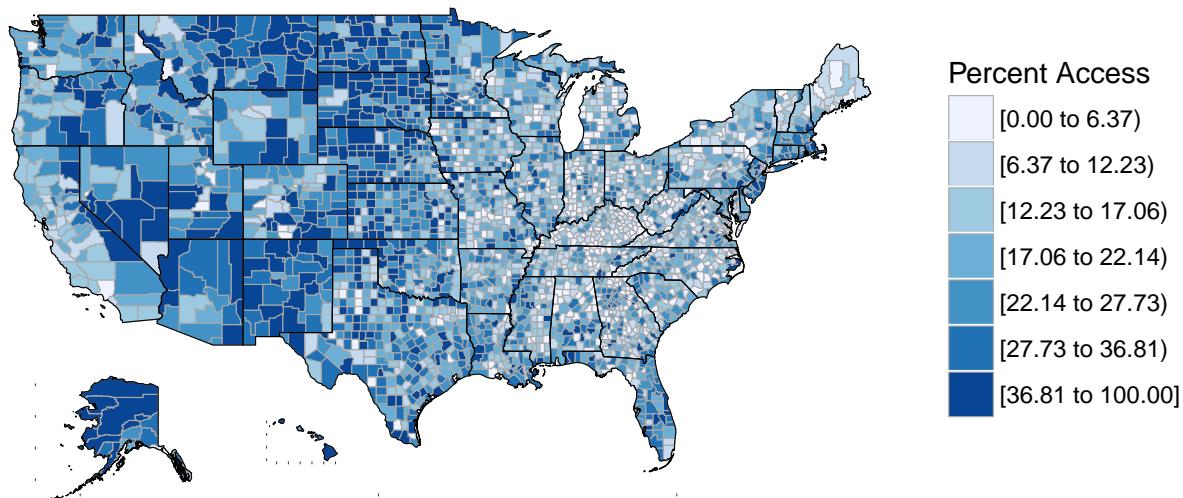
```
## Warning in super$initialize(map.df, user.df): Your data.frame contains the
## following regions which are not mappable: 1000, 2000, 4000, 5000, 6000,
## 8000, 9000, 10000, 11000, 12000, 13000, 15000, 16000, 17000, 18000, 19000,
## 20000, 21000, 22000, 23000, 24000, 25000, 26000, 27000, 28000, 29000,
## 30000, 31000, 32000, 33000, 34000, 35000, 36000, 37000, 38000, 39000,
## 40000, 41000, 42000, 44000, 45000, 46000, 47000, 48000, 49000, 50000,
## 51000, 53000, 54000, 55000, 56000
```

Poverty Rate by County 2013



```
## Warning in super$initialize(map.df, user.df): Your data.frame contains the
## following regions which are not mappable: 2010, 2201, 2231, 2232, 2280,
## 12025, 15901, 30113, 51560, 51780, 51901, 51903, 51907, 51909, 51911,
## 51913, 51918, 51919, 51921, 51923, 51929, 51931, 51933, 51939, 51941,
## 51942, 51944, 51945, 51947, 51949, 51951, 51953, 51955, 51958, 72001,
## 72003, 72005, 72007, 72009, 72011, 72013, 72015, 72017, 72019, 72021,
## 72023, 72025, 72027, 72029, 72031, 72033, 72035, 72037, 72039, 72041,
## 72043, 72045, 72047, 72049, 72051, 72053, 72054, 72055, 72057, 72059,
## 72061, 72063, 72065, 72067, 72069, 72071, 72073, 72075, 72077, 72079,
## 72081, 72083, 72085, 72087, 72089, 72091, 72093, 72095, 72097, 72099,
## 72101, 72103, 72105, 72107, 72109, 72111, 72113, 72115, 72117, 72119,
## 72121, 72123, 72125, 72127, 72129, 72131, 72133, 72135, 72137, 72139,
## 72141, 72143, 72145, 72147, 72149, 72151, 72153
```

Access to Grocery Stores by County 2013



```
## Warning in super$initialize(map.df, user.df): Your data.frame contains the
```

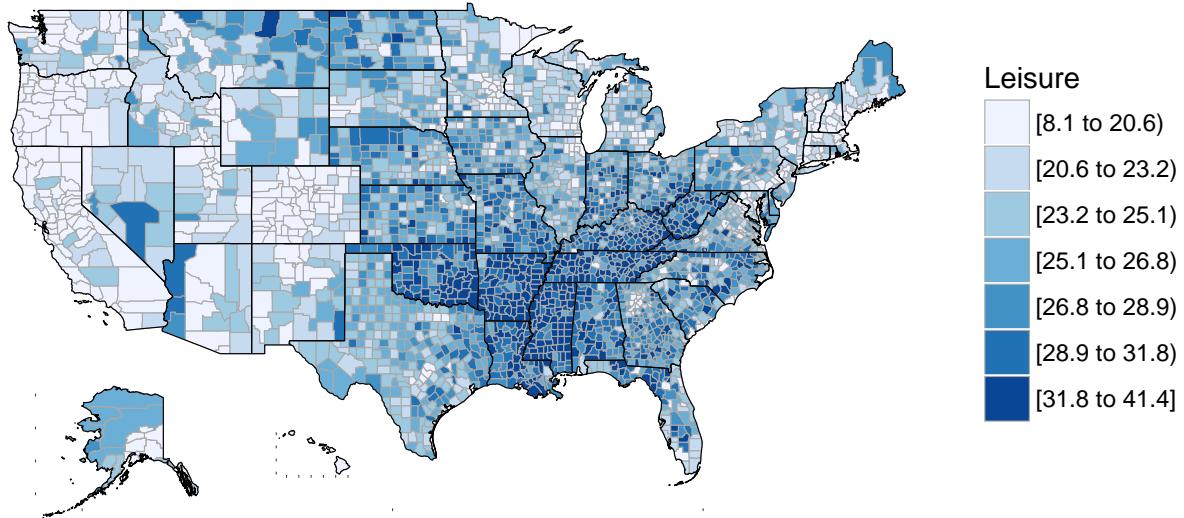
```

## following regions which are not mappable: 2201, 2232, 2280, 72001, 72003,
## 72005, 72007, 72009, 72011, 72013, 72015, 72017, 72019, 72021, 72023,
## 72025, 72027, 72029, 72031, 72033, 72035, 72037, 72039, 72041, 72043,
## 72045, 72047, 72049, 72051, 72053, 72054, 72055, 72057, 72059, 72061,
## 72063, 72065, 72067, 72069, 72071, 72073, 72075, 72077, 72079, 72081,
## 72083, 72085, 72087, 72089, 72091, 72093, 72095, 72097, 72099, 72101,
## 72103, 72105, 72107, 72109, 72111, 72113, 72115, 72117, 72119, 72121,
## 72123, 72125, 72127, 72129, 72131, 72133, 72135, 72137, 72139, 72141,
## 72143, 72145, 72147, 72149, 72151, 72153

## Warning in self$bind(): The following regions were missing and are being
## set to NA: 51515

```

Leisure Time Spent Physically Inactive by County 2013



compare... maps seem to follow a similar trend

let's take a closer look

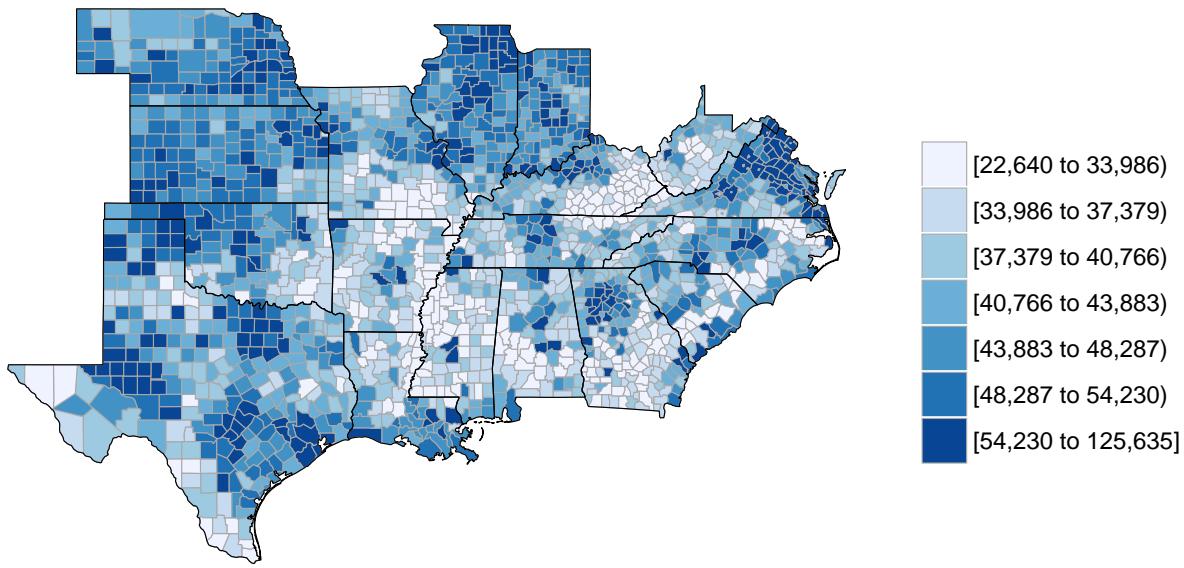
```

## Warning in super$initialize(map.df, user.df): Your data.frame contains the
## following regions which are not mappable: 1000, 2000, 2158, 2201, 2232,
## 2280, 4000, 5000, 6000, 8000, 9000, 10000, 11000, 12000, 13000, 15000,
## 16000, 17000, 18000, 19000, 20000, 21000, 22000, 23000, 24000, 25000,
## 26000, 27000, 28000, 29000, 30000, 31000, 32000, 33000, 34000, 35000,
## 36000, 37000, 38000, 39000, 40000, 41000, 42000, 44000, 45000, 46000,
## 46102, 47000, 48000, 49000, 50000, 51000, 53000, 54000, 55000, 56000, NA

## Warning in self$bind(): The following regions were missing and are being
## set to NA: 51515

```

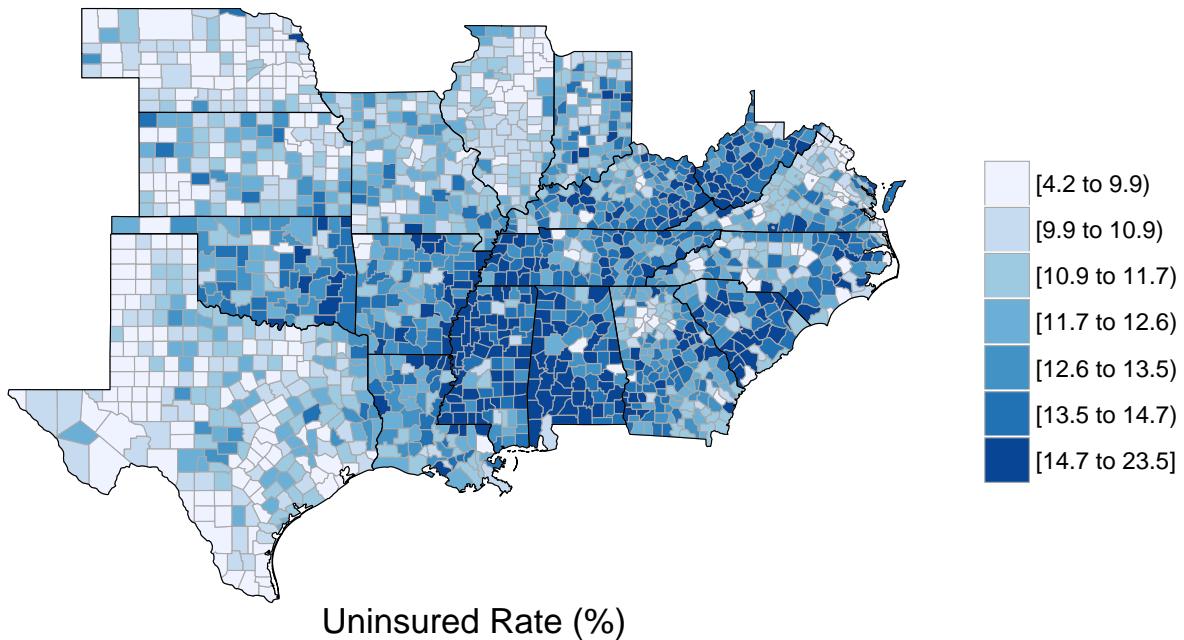
Low Median Household Income



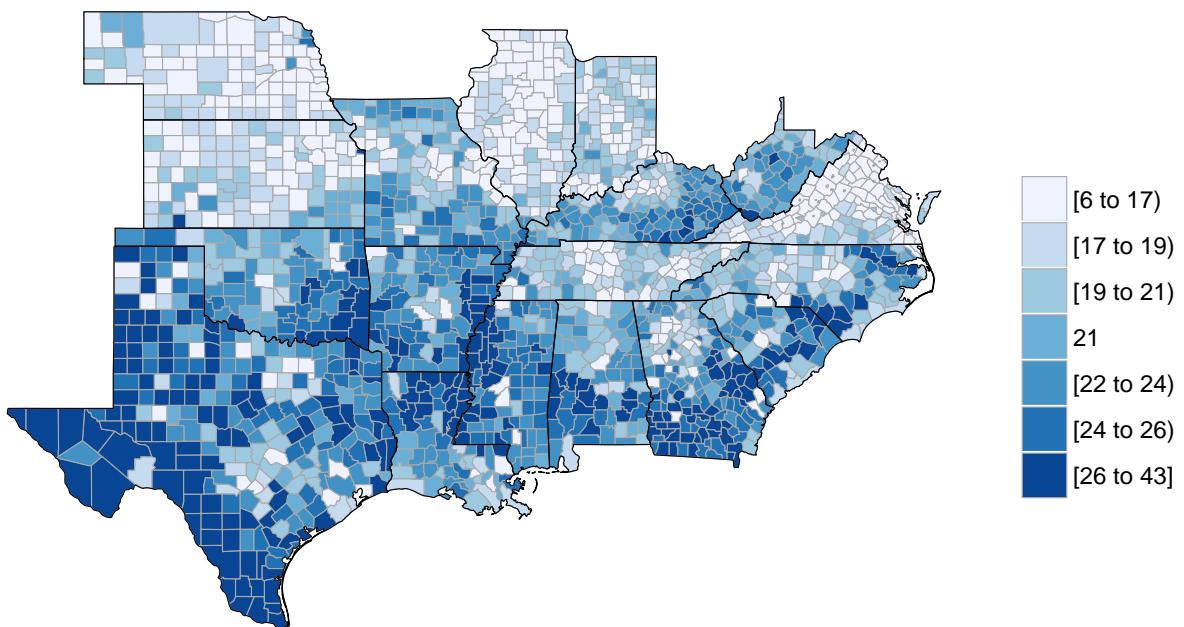
```
## Warning in super$initialize(map.df, user.df): Your data.frame contains the
## following regions which are not mappable: 2201, 2232, 2280, 72001, 72003,
## 72005, 72007, 72009, 72011, 72013, 72015, 72017, 72019, 72021, 72023,
## 72025, 72027, 72029, 72031, 72033, 72035, 72037, 72039, 72041, 72043,
## 72045, 72047, 72049, 72051, 72053, 72054, 72055, 72057, 72059, 72061,
## 72063, 72065, 72067, 72069, 72071, 72073, 72075, 72077, 72079, 72081,
## 72083, 72085, 72087, 72089, 72091, 72093, 72095, 72097, 72099, 72101,
## 72103, 72105, 72107, 72109, 72111, 72113, 72115, 72117, 72119, 72121,
## 72123, 72125, 72127, 72129, 72131, 72133, 72135, 72137, 72139, 72141,
## 72143, 72145, 72147, 72149, 72151, 72153

## Warning in super$initialize(map.df, user.df): The following regions were
## missing and are being set to NA: 51515
```

Diabetes Prevalence (%)

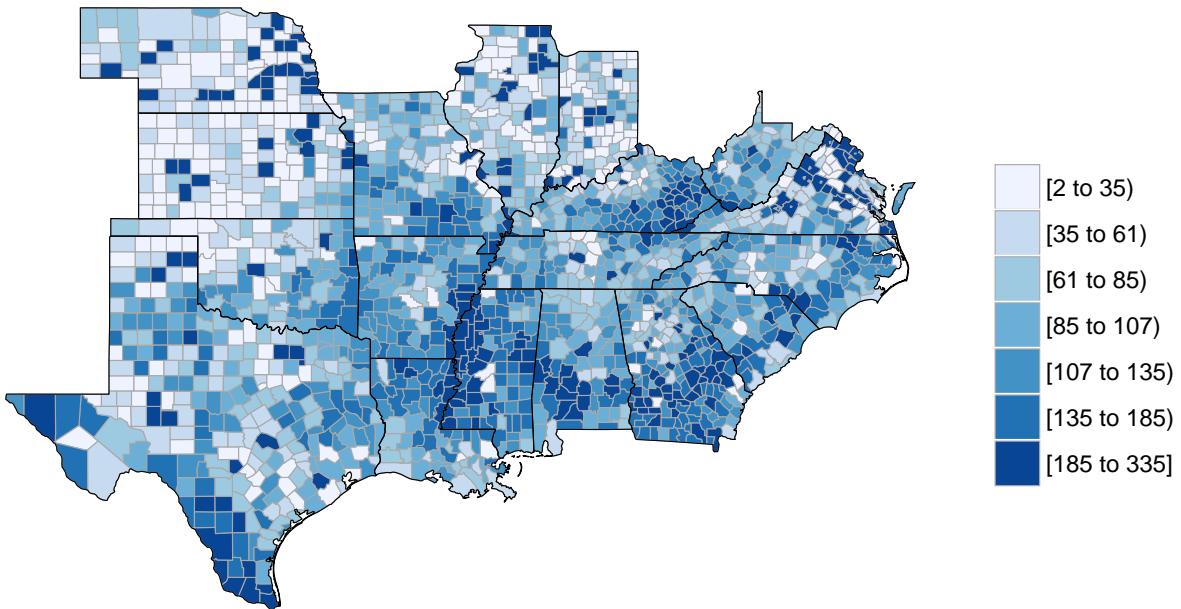


Uninsured Rate (%)

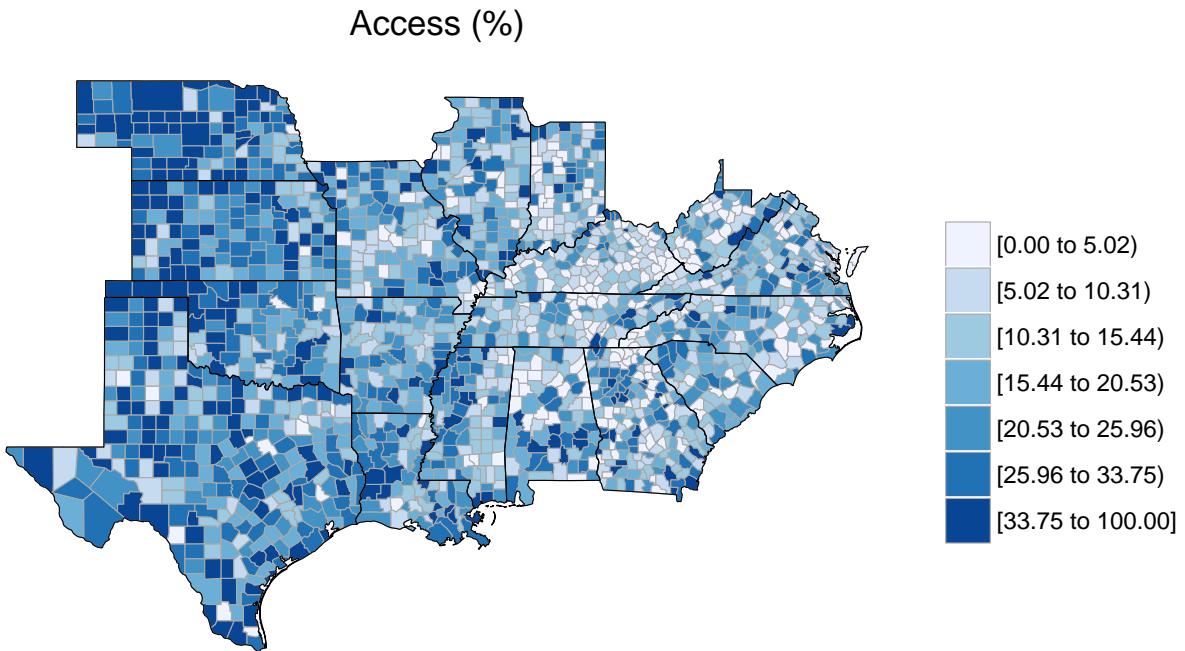


```
## Warning in super$initialize(map.df, user.df): Your data.frame contains the
## following regions which are not mappable: 1000, 2000, 4000, 5000, 6000,
## 8000, 9000, 10000, 11000, 12000, 13000, 15000, 16000, 17000, 18000, 19000,
## 20000, 21000, 22000, 23000, 24000, 25000, 26000, 27000, 28000, 29000,
## 30000, 31000, 32000, 33000, 34000, 35000, 36000, 37000, 38000, 39000,
## 40000, 41000, 42000, 44000, 45000, 46000, 47000, 48000, 49000, 50000,
## 51000, 53000, 54000, 55000, 56000
```

Uninsured Rate (%)



```
## Warning in super$initialize(map.df, user.df): Your data.frame contains the
## following regions which are not mappable: 2010, 2201, 2231, 2232, 2280,
## 12025, 15901, 30113, 51560, 51780, 51901, 51903, 51907, 51909, 51911,
## 51913, 51918, 51919, 51921, 51923, 51929, 51931, 51933, 51939, 51941,
## 51942, 51944, 51945, 51947, 51949, 51951, 51953, 51955, 51958, 72001,
## 72003, 72005, 72007, 72009, 72011, 72013, 72015, 72017, 72019, 72021,
## 72023, 72025, 72027, 72029, 72031, 72033, 72035, 72037, 72039, 72041,
## 72043, 72045, 72047, 72049, 72051, 72053, 72054, 72055, 72057, 72059,
## 72061, 72063, 72065, 72067, 72069, 72071, 72073, 72075, 72077, 72079,
## 72081, 72083, 72085, 72087, 72089, 72091, 72093, 72095, 72097, 72099,
## 72101, 72103, 72105, 72107, 72109, 72111, 72113, 72115, 72117, 72119,
## 72121, 72123, 72125, 72127, 72129, 72131, 72133, 72135, 72137, 72139,
## 72141, 72143, 72145, 72147, 72149, 72151, 72153
```

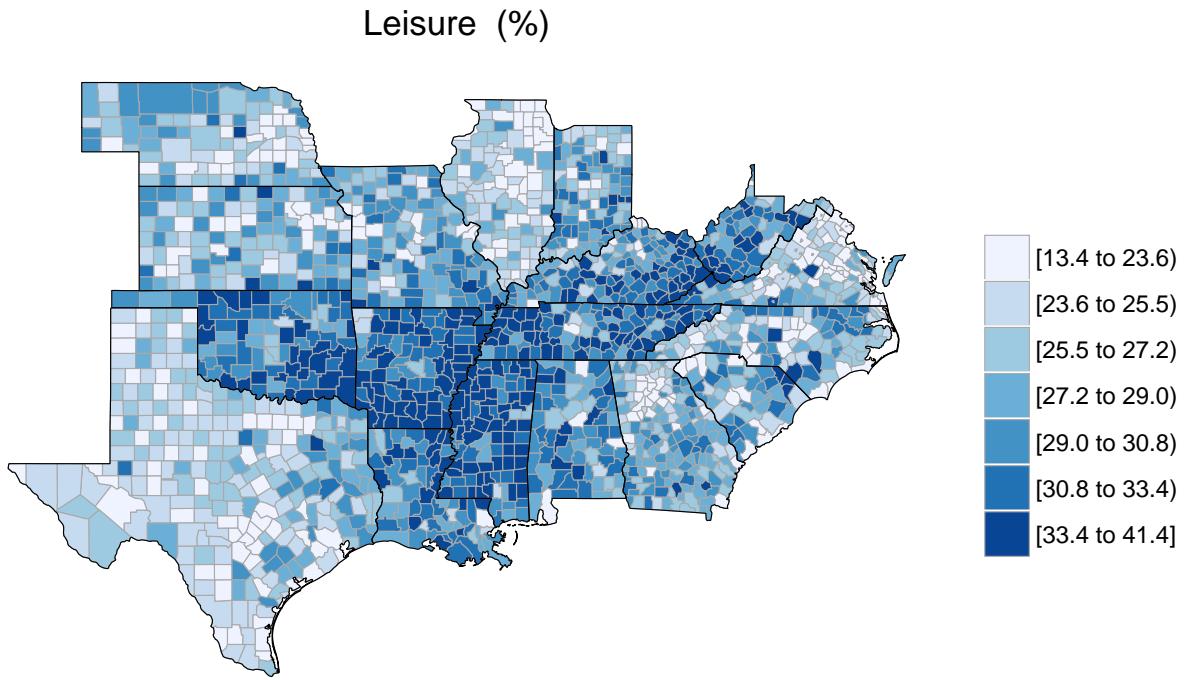


```

## Warning in super$initialize(map.df, user.df): Your data.frame contains the
## following regions which are not mappable: 2201, 2232, 2280, 72001, 72003,
## 72005, 72007, 72009, 72011, 72013, 72015, 72017, 72019, 72021, 72023,
## 72025, 72027, 72029, 72031, 72033, 72035, 72037, 72039, 72041, 72043,
## 72045, 72047, 72049, 72051, 72053, 72054, 72055, 72057, 72059, 72061,
## 72063, 72065, 72067, 72069, 72071, 72073, 72075, 72077, 72079, 72081,
## 72083, 72085, 72087, 72089, 72091, 72093, 72095, 72097, 72099, 72101,
## 72103, 72105, 72107, 72109, 72111, 72113, 72115, 72117, 72119, 72121,
## 72123, 72125, 72127, 72129, 72131, 72133, 72135, 72137, 72139, 72141,
## 72143, 72145, 72147, 72149, 72151, 72153

## Warning in super$initialize(map.df, user.df): The following regions were
## missing and are being set to NA: 51515

```



Notice the switch?

This region has both low median household income and high percent of diabetes prevalence.

Merge Data and Prepare for Linear Regression

```
##   FIPS uninsured13 uninsured16
## 1 1001        17         8
## 2 1003        17         8
## 3 1005        24        15
## 4 1007        20        10
## 5 1009        19        10
## 6 1011        28        17

##   FIPS uninsured13 uninsured16 prevalence
## 1 1001        17         8      13.0
## 2 1003        17         8      10.4
## 3 1005        24        15     18.4
## 4 1007        20        10     14.8
## 5 1009        19        10     14.1
## 6 1011        28        17     19.6

##   FIPS uninsured13 uninsured16 prevalence income poverty    access leisure
## 1 1001        17         8      13.0  54366       37 33.769657    28.6
## 2 1003        17         8      10.4  49626       44 19.318473    22.3
## 3 1005        24        15     18.4  34971      182 20.840972    31.8
## 4 1007        20        10     14.8  39546      133 4.559753    33.9
## 5 1009        19        10     14.1  45567       74 2.700840    28.0
## 6 1011        28        17     19.6  26580      237 37.474652    31.7
```

Linear Regressions:

Response: Diabetes Prevalence (13'&16')

Treatment: Poverty Rate; Uninsured Rate; Median Household Income; Access to Grocery Stores; Leisure Time spent Physically Inactive

```
#simple linear models
lm.p = lm(prevalence ~ poverty, data = final)
summary(lm.p)

##
## Call:
## lm(formula = prevalence ~ poverty, data = final)
##
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -8.1195 -1.6950 -0.1215  1.6373 12.1500 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 1.115e+01  6.902e-02 161.477   <2e-16 ***
## poverty     8.277e-04  4.896e-04   1.691    0.091 .  
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 2.484 on 3148 degrees of freedom
## Multiple R-squared:  0.0009071, Adjusted R-squared:  0.0005897 
## F-statistic: 2.858 on 1 and 3148 DF,  p-value: 0.09101

lm.i = lm(prevalence ~ income, data = final)
summary(lm.i)

##
## Call:
## lm(formula = prevalence ~ income, data = final)
##
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -8.5359 -1.2927 -0.0093  1.3177 10.3405 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 1.675e+01  1.462e-01 114.53   <2e-16 ***
## income      -1.169e-04  3.005e-06 -38.92   <2e-16 *** 
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 2.042 on 3148 degrees of freedom
## Multiple R-squared:  0.3249, Adjusted R-squared:  0.3247 
## F-statistic: 1515 on 1 and 3148 DF,  p-value: < 2.2e-16
```

```

lm.un13 = lm(prevalence ~ uninsured13, data = final)
summary(lm.un13)

## 
## Call:
## lm(formula = prevalence ~ uninsured13, data = final)
## 
## Residuals:
##    Min     1Q Median     3Q    Max 
## -7.998 -1.314  0.060  1.432 10.131 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 7.301853  0.150837  48.41   <2e-16 ***
## uninsured13 0.209212  0.007737  27.04   <2e-16 ***  
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 2.239 on 3148 degrees of freedom
## Multiple R-squared:  0.1885, Adjusted R-squared:  0.1882 
## F-statistic: 731.1 on 1 and 3148 DF,  p-value: < 2.2e-16

lm.a = lm(prevalence ~ access, data = final)
summary(lm.a)

## 
## Call:
## lm(formula = prevalence ~ access, data = final)
## 
## Residuals:
##    Min     1Q Median     3Q    Max 
## -8.1234 -1.6607 -0.1111  1.5770 12.8247 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 11.764938  0.066523 176.85   <2e-16 *** 
## access      -0.022373  0.002125 -10.53   <2e-16 ***  
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 2.443 on 3148 degrees of freedom
## Multiple R-squared:  0.03403, Adjusted R-squared:  0.03372 
## F-statistic: 110.9 on 1 and 3148 DF,  p-value: < 2.2e-16

lm.l = lm(prevalence ~ leisure, data = final)
summary(lm.l)

## 
## Call:
## lm(formula = prevalence ~ leisure, data = final)
## 
## Residuals:

```

```

##      Min     1Q Median     3Q    Max
## -7.1450 -1.1067 -0.0678  0.9465 10.8701
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.04701   0.15118 13.54 <2e-16 ***
## leisure     0.35394   0.00571 61.98 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.668 on 3148 degrees of freedom
## Multiple R-squared:  0.5496, Adjusted R-squared:  0.5495
## F-statistic:  3842 on 1 and 3148 DF,  p-value: < 2.2e-16

#multiple linear regression

mult.lm = lm(prevalence ~ uninsured13 + income + poverty + leisure + access, data = final)
summary(mult.lm)

##
## Call:
## lm(formula = prevalence ~ uninsured13 + income + poverty + leisure +
##     access, data = final)
##
## Residuals:
##      Min     1Q Median     3Q    Max
## -7.6549 -1.0694 -0.0511  0.9658 10.3842
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.452e+00 3.804e-01 16.962 < 2e-16 ***
## uninsured13 -8.873e-03 8.229e-03 -1.078 0.281
## income      -5.125e-05 3.867e-06 -13.252 < 2e-16 ***
## poverty      2.029e-03 3.227e-04  6.287 3.68e-10 ***
## leisure      2.854e-01 6.609e-03 43.183 < 2e-16 ***
## access       -1.113e-02 1.385e-03 -8.036 1.30e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.579 on 3144 degrees of freedom
## Multiple R-squared:  0.5968, Adjusted R-squared:  0.5961
## F-statistic: 930.6 on 5 and 3144 DF,  p-value: < 2.2e-16

#remove uninsured13 because not significant

new.lm = lm(prevalence ~ income + poverty + access +leisure, data = final)
summary(new.lm)

##
## Call:
## lm(formula = prevalence ~ income + poverty + access + leisure,
##     data = final)
##

```

```

## Residuals:
##      Min     1Q Median     3Q    Max
## -7.5999 -1.0657 -0.0426  0.9705 10.3516
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.167e+00 2.738e-01 22.527 < 2e-16 ***
## income      -4.844e-05 2.859e-06 -16.944 < 2e-16 ***
## poverty     1.959e-03 3.162e-04   6.197 6.52e-10 ***
## access      -1.118e-02 1.384e-03  -8.077 9.37e-16 ***
## leisure     2.852e-01 6.607e-03  43.170 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.579 on 3145 degrees of freedom
## Multiple R-squared:  0.5966, Adjusted R-squared:  0.5961
## F-statistic: 1163 on 4 and 3145 DF, p-value: < 2.2e-16

```

```

new.lm.noleisure = lm(prevalence ~ income + poverty + access, data = final)
summary(new.lm.noleisure)

```

```

##
## Call:
## lm(formula = prevalence ~ income + poverty + access, data = final)
##
## Residuals:
##      Min     1Q Median     3Q    Max
## -8.2417 -1.2365  0.0342  1.2759 10.2909
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.686e+01 1.474e-01 114.327 < 2e-16 ***
## income      -1.182e-04 2.975e-06 -39.742 < 2e-16 ***
## poverty     3.236e-03 3.972e-04   8.148 5.3e-16 ***
## access      -1.682e-02 1.739e-03  -9.673 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.993 on 3146 degrees of freedom
## Multiple R-squared:  0.3576, Adjusted R-squared:  0.357
## F-statistic: 583.7 on 3 and 3146 DF, p-value: < 2.2e-16

```

```

#add interactions (interact income*access and poverty*leisure)
#look to interact uninsured13 for significance

```

```

inter = lm(prevalence ~ income*access + poverty*leisure, final) #interact median income with access to ...
summary(inter)

```

```

##
## Call:
## lm(formula = prevalence ~ income * access + poverty * leisure,
##      data = final)
## 
```

```

## Residuals:
##      Min     1Q Median     3Q    Max
## -7.4551 -1.0490 -0.0277  0.9383 10.3186
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)            6.987e+00  3.406e-01 20.513 < 2e-16 ***
## income                -4.974e-05 4.579e-06 -10.862 < 2e-16 ***
## access                -2.317e-02 6.427e-03 -3.605 0.000317 *** 
## poverty               -4.883e-03 1.662e-03 -2.939 0.003321 **  
## leisure               2.537e-01 9.690e-03 26.181 < 2e-16 ***
## income:access         2.568e-07 1.337e-07  1.921 0.054851 .  
## poverty:leisure       2.824e-04 6.799e-05  4.154 3.35e-05 *** 
## ---                
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.574 on 3143 degrees of freedom
## Multiple R-squared:  0.5995, Adjusted R-squared:  0.5987 
## F-statistic: 784.1 on 6 and 3143 DF,  p-value: < 2.2e-16
```

interaction = lm(prevalence ~ poverty*leisure +access +income, final)
summary(interaction)

```

## 
## Call:
## lm(formula = prevalence ~ poverty * leisure + access + income,
##      data = final)
##
## Residuals:
##      Min     1Q Median     3Q    Max
## -7.6447 -1.0553 -0.0261  0.9446 10.1907
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)            6.662e+00  2.958e-01 22.523 < 2e-16 ***
## poverty                -5.114e-03 1.658e-03 -3.084 0.00206 ** 
## leisure                2.543e-01 9.688e-03 26.253 < 2e-16 *** 
## access                 -1.111e-02 1.381e-03 -8.049 1.17e-15 *** 
## income                 -4.325e-05 3.091e-06 -13.990 < 2e-16 *** 
## poverty:leisure        2.943e-04 6.774e-05  4.345 1.44e-05 *** 
## ---                
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.575 on 3144 degrees of freedom
## Multiple R-squared:  0.599, Adjusted R-squared:  0.5984 
## F-statistic: 939.4 on 5 and 3144 DF,  p-value: < 2.2e-16
```

#try to interact income with uninsured and see if there is an affect.

#this will set up the linear model to use for estimating prevalence of diabetes with new 2016 uninsured
estimate.lm = lm(prevalence ~ income*uninsured13 + leisure*poverty + access, final)
summary(estimate.lm)

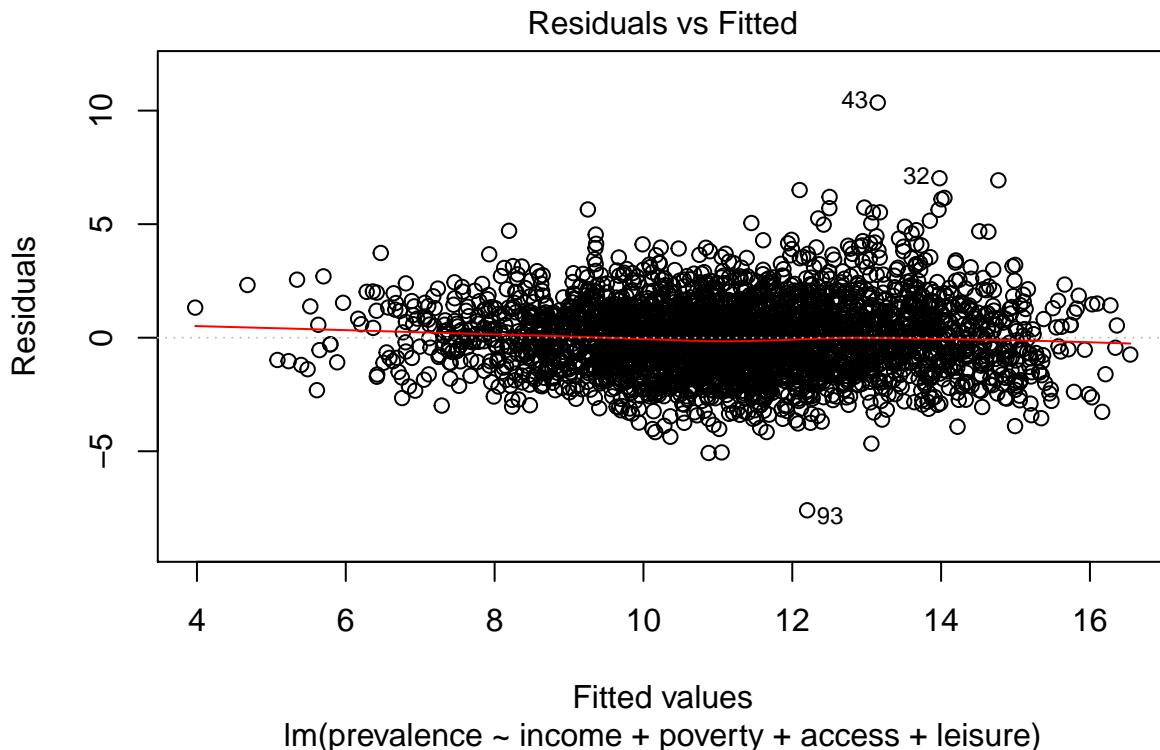
##

```

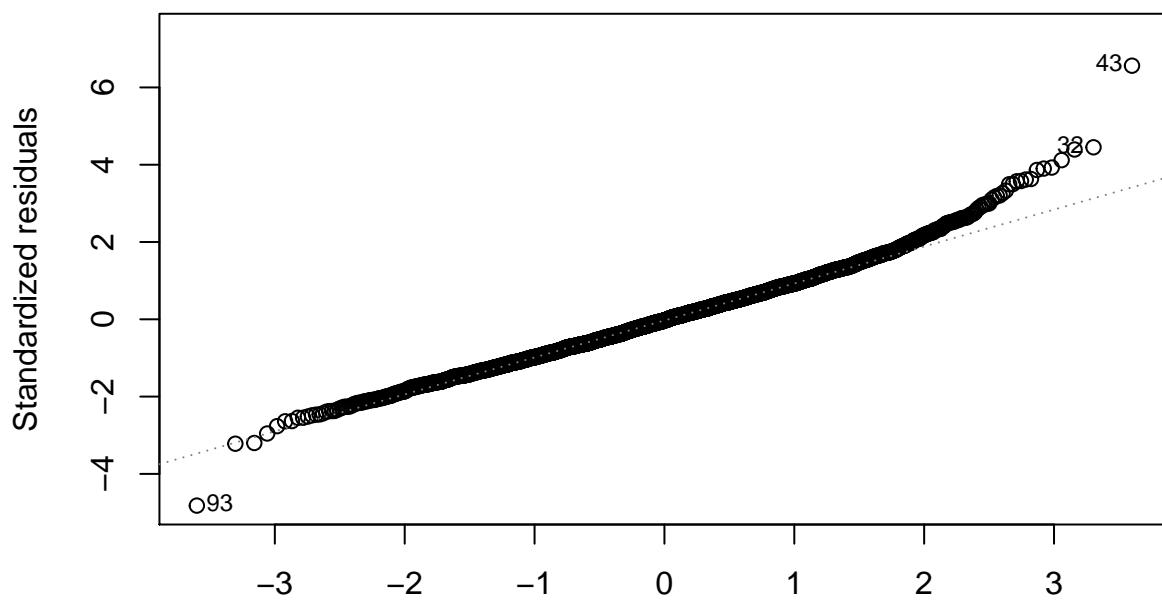
## Call:
## lm(formula = prevalence ~ income * uninsured13 + leisure * poverty +
##      access, data = final)
##
## Residuals:
##    Min     1Q Median     3Q    Max 
## -7.6217 -1.0322 -0.0311  0.9328  9.8374 
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 4.336e+00 4.747e-01   9.134 < 2e-16 ***
## income      1.629e-05 7.366e-06   2.212 0.027048 *  
## uninsured13 1.797e-01 2.052e-02   8.757 < 2e-16 ***  
## leisure     2.549e-01 9.537e-03  26.724 < 2e-16 ***  
## poverty     -5.807e-03 1.633e-03  -3.555 0.000384 ***  
## access      -9.905e-03 1.364e-03  -7.260 4.85e-13 ***  
## income:uninsured13 -4.395e-06 4.353e-07 -10.096 < 2e-16 ***  
## leisure:poverty  2.465e-04 6.695e-05   3.682 0.000236 ***  
## ---    
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
##
## Residual standard error: 1.55 on 3142 degrees of freedom
## Multiple R-squared:  0.6118, Adjusted R-squared:  0.611 
## F-statistic: 707.5 on 7 and 3142 DF,  p-value: < 2.2e-16

```

```
plot(new.lm)
```



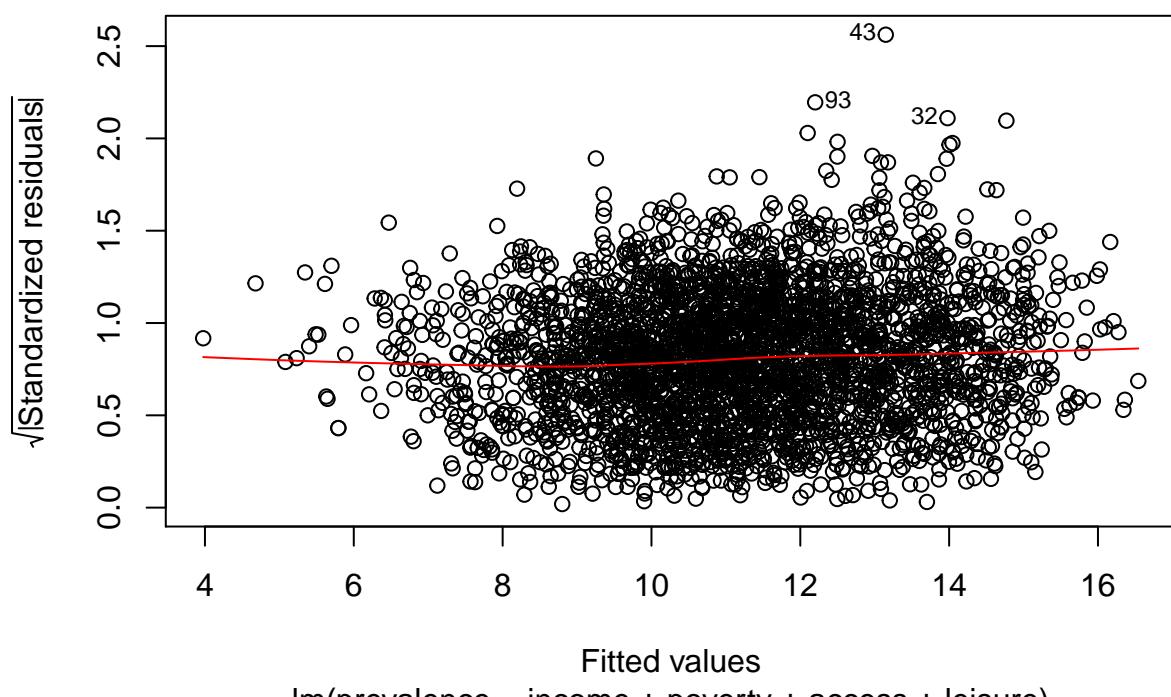
Normal Q–Q



Theoretical Quantiles

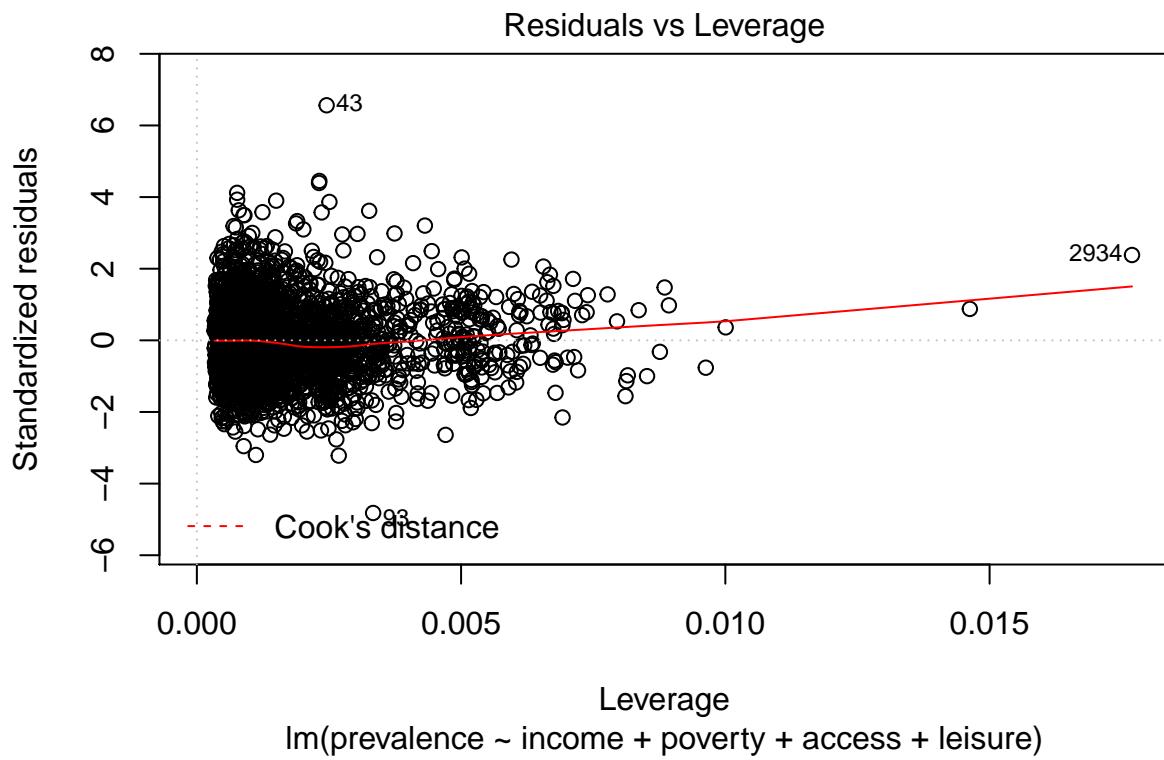
lm(prevalence ~ income + poverty + access + leisure)

Scale–Location



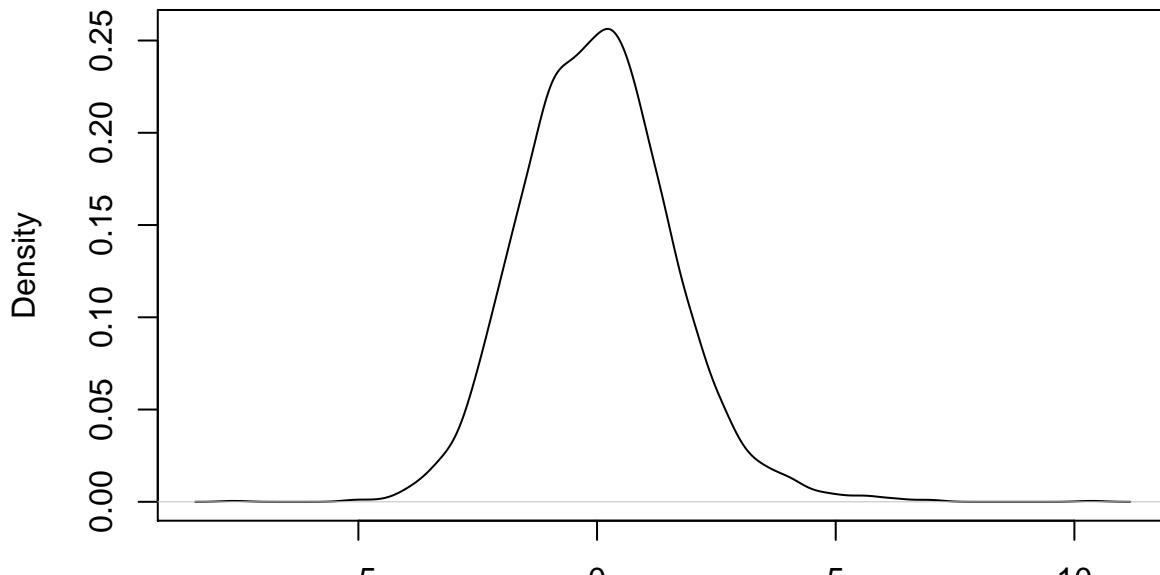
Fitted values

lm(prevalence ~ income + poverty + access + leisure)



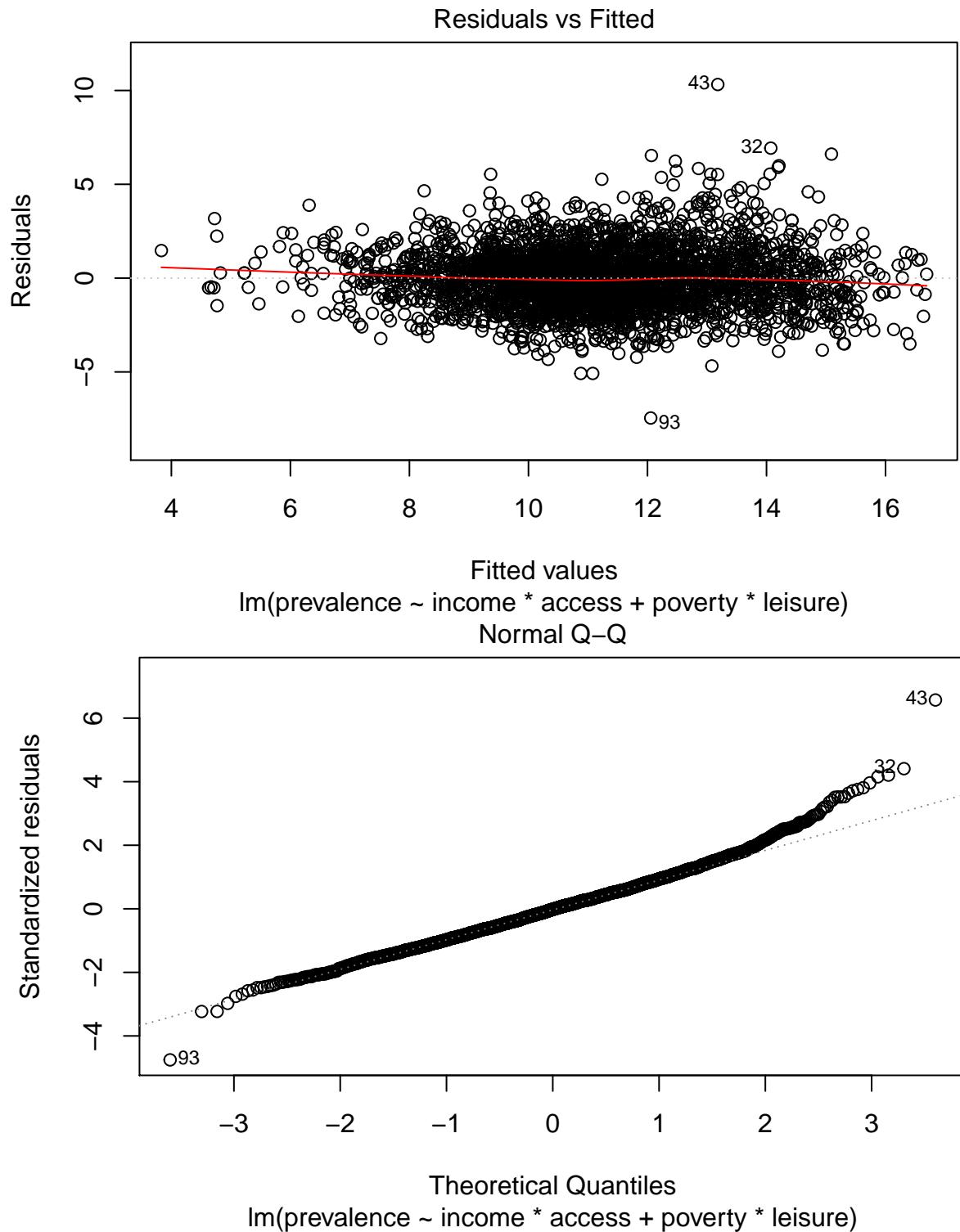
```
plot(density(resid(new.lm)))
```

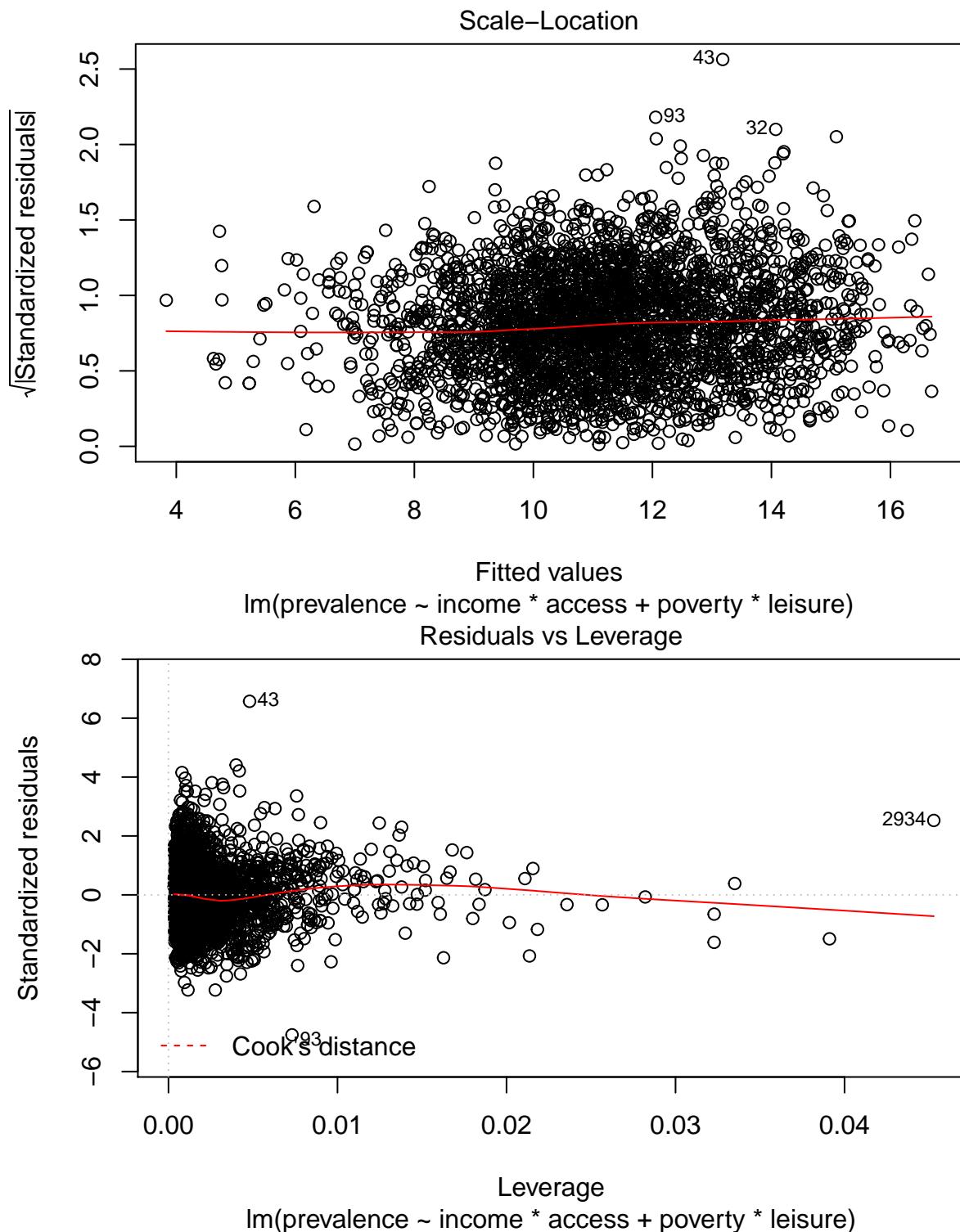
density.default(x = resid(new.lm))



N = 3150 Bandwidth = 0.2731

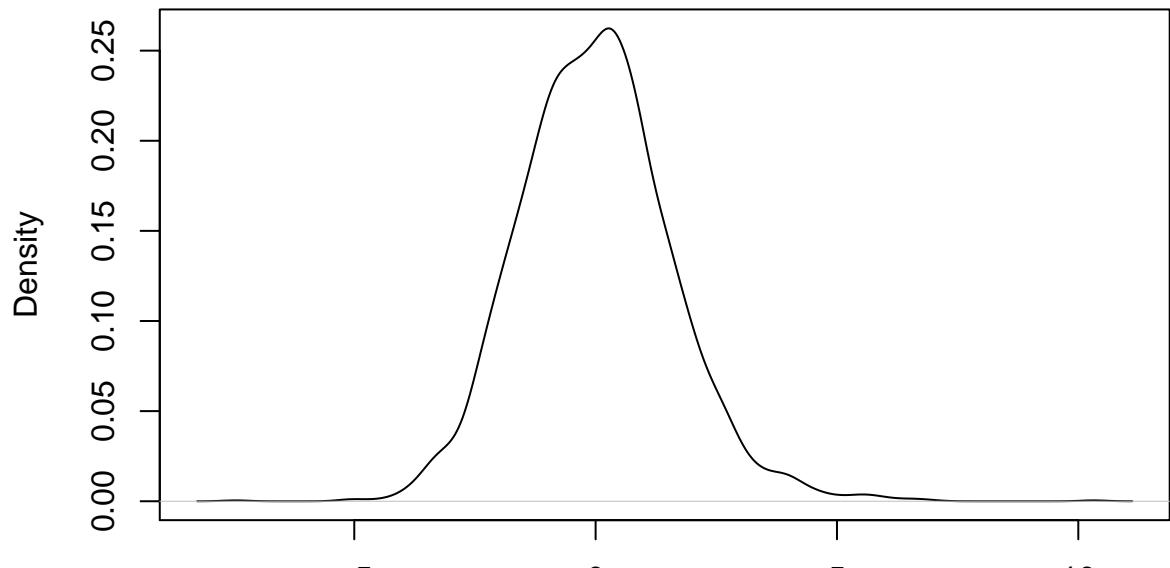
```
plot(inter)
```





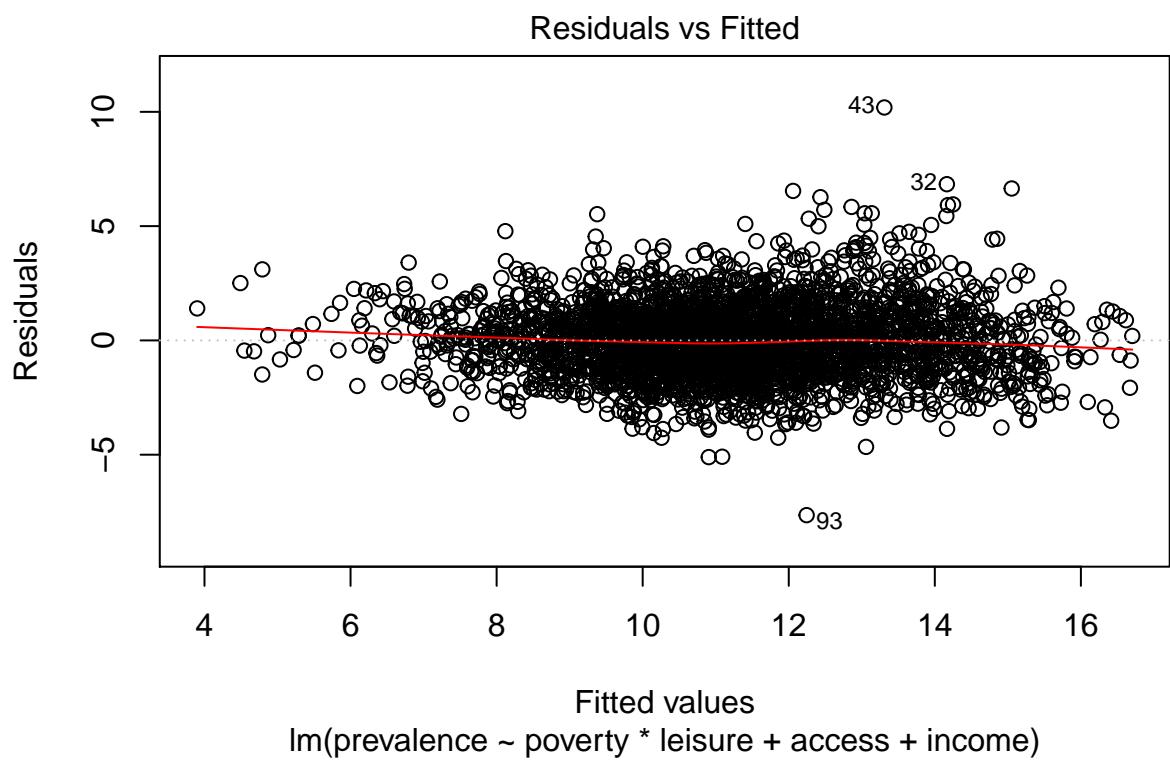
```
plot(density(resid(inter)))
```

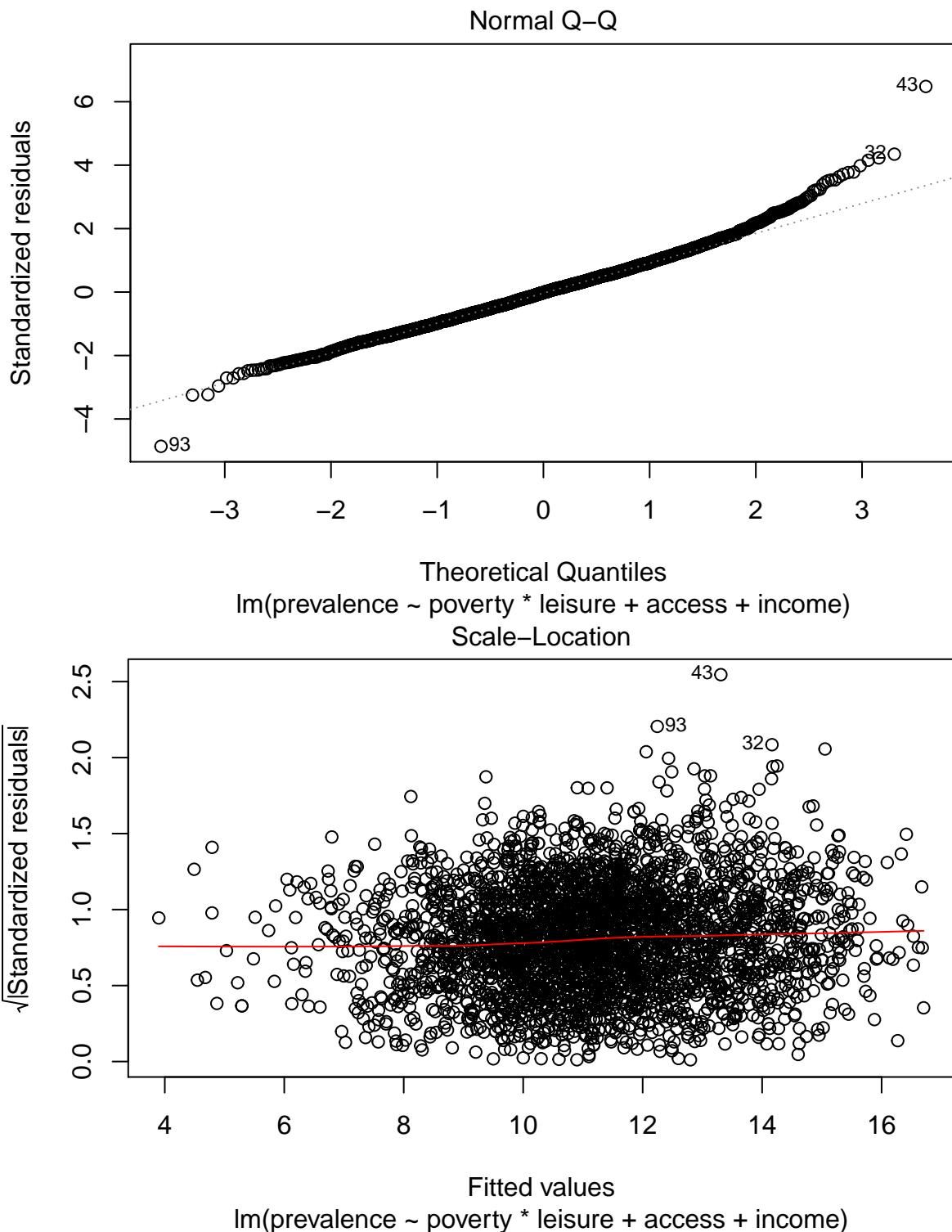
density.default(x = resid(inter))



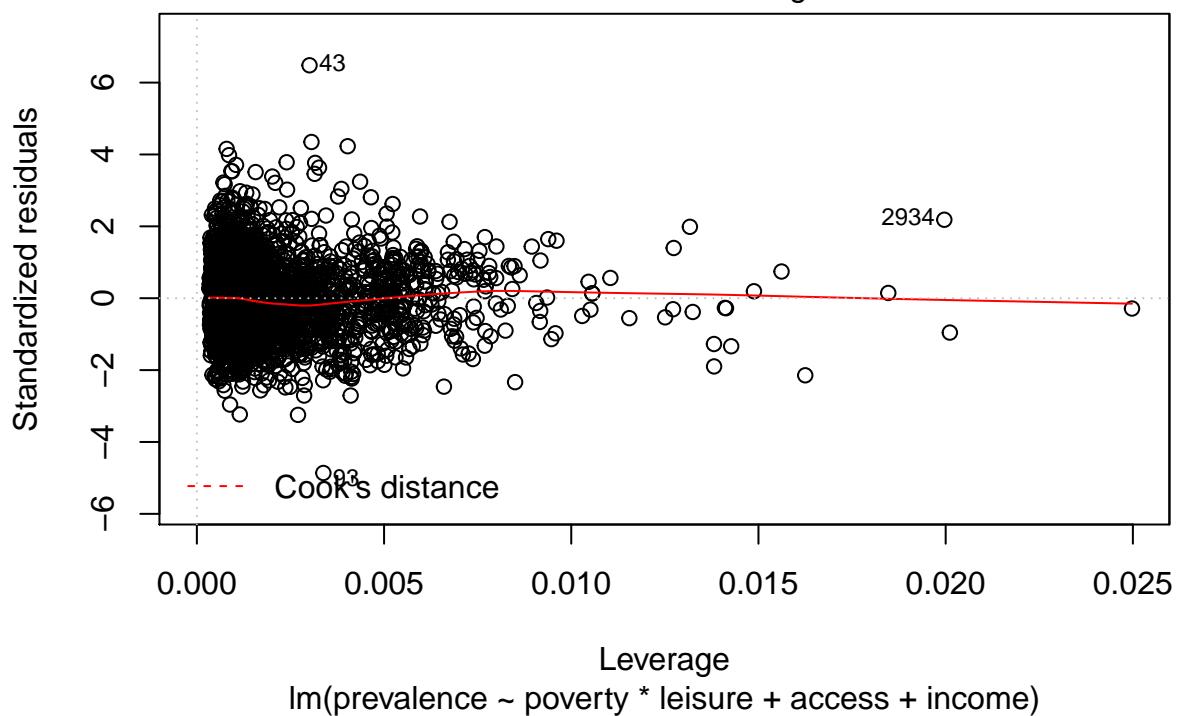
N = 3150 Bandwidth = 0.2665

```
plot(interaction)
```



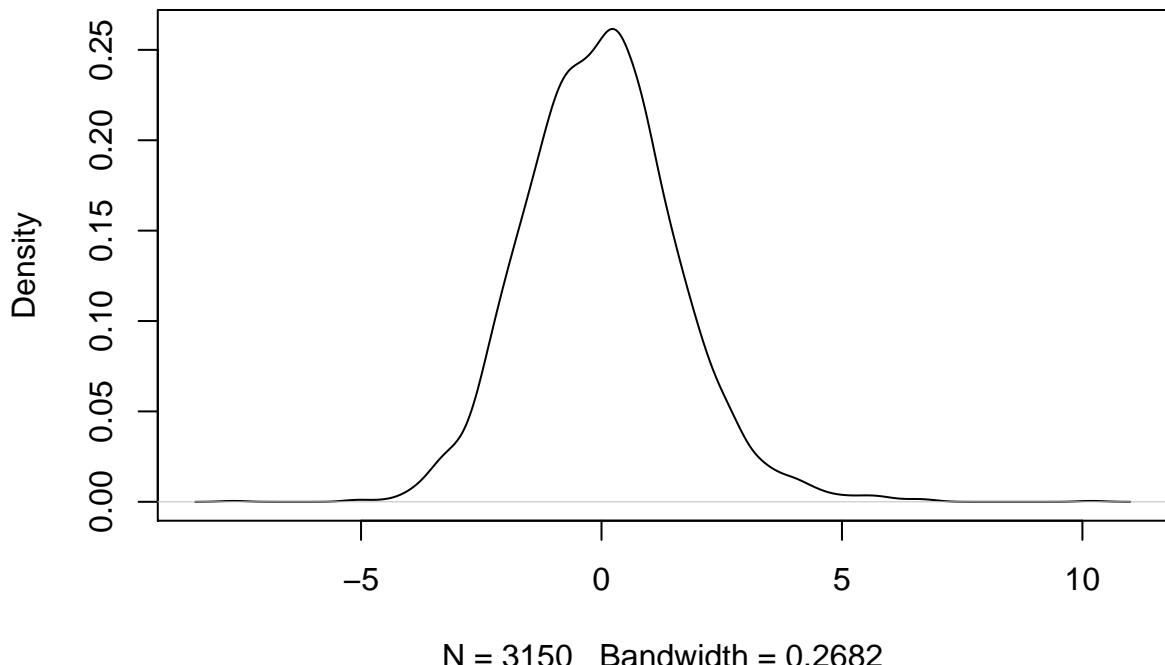


Residuals vs Leverage

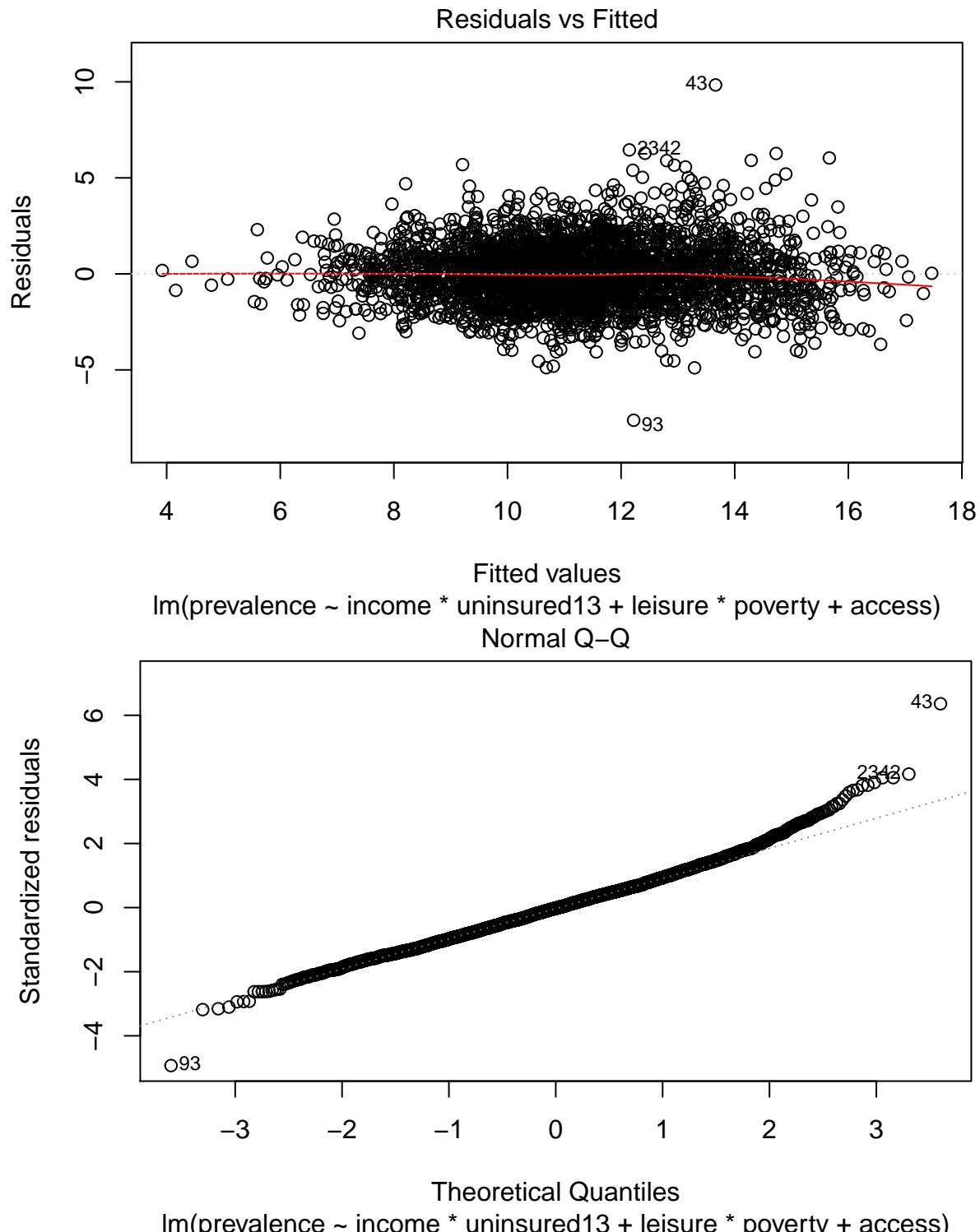


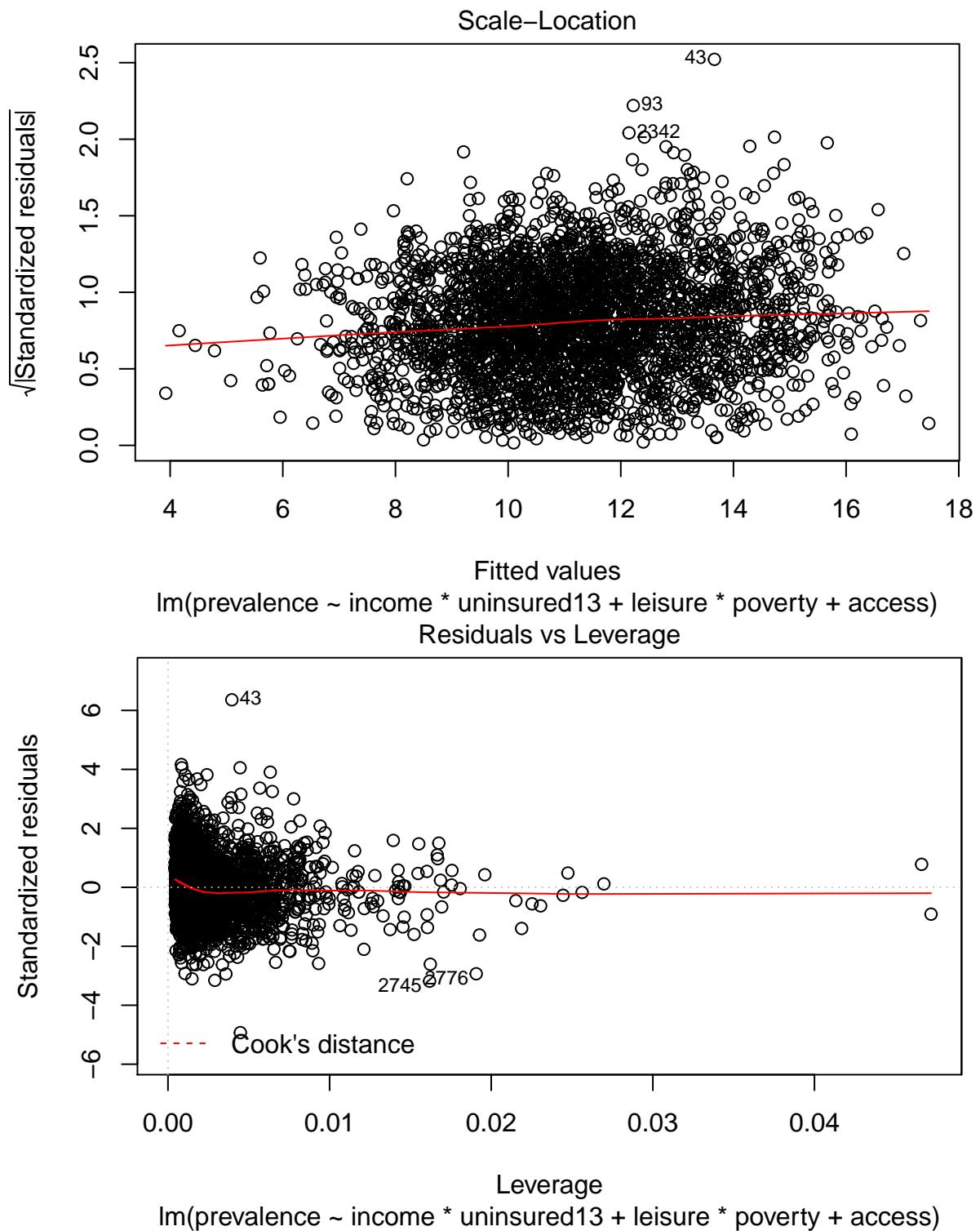
```
plot(density(resid(interaction)))
```

density.default(x = resid(interaction))



```
plot(estimate.lm)
```





Bootstrap

```
library(boot)
n = length(final$prevalence)
```

```
B = 1000
result = rep(NA, B)
for (i in 1:B) {
  boot.sample = sample(n, replace = TRUE)
  result[i] = mean(final$prevalence[boot.sample])
}
with(final, mean(prevalence) + c(-1,1) *2 *sd(result))

## [1] 11.14791 11.32231
```