S&DS 363 Homework 4

Evan Collins, Kelly Farley, Ken Stier

30 March 2021

```
## ---
## biotools version 4.0
```

Contributors

Evan Collins (evan.collins@yale.edu (mailto:evan.collins@yale.edu))

Kelly Farley (kelly.farley@yale.edu (mailto:kelly.farley@yale.edu))

Ken Stier (ken.stier@yale.edu (mailto:ken.stier@yale.edu))

The Dataset

Raw dataset: COVID-19 infection and death statistics from U.S. counties (sourced from NYT), combined with economic, education, and population data (sourced from various government agencies) and also survey responses about mask-wearing frequencies (sourced from NYT). 3141 complete observations on 19 metric variables and 6 categorical variables. To avoid any outliers due to population size differences between counties, all variables are scaled as a percentage of population. Variable descriptions can be found here (http://evancollins.com/variable_descriptions.html).

Data of relevance for this pset:

Categorical Predictor 1 (rural_urban_code): The Rural-Urban Codes are numbered 1-9 according to descriptions provided by the USDA. We will regroup codes 1 through 9 as into three groups: (1) “Urban” for codes 1-3, (2) “Suburban” for codes 4-6, and (3) “Rural” for codes 7-9.

Categorical Predictor 2 (region): Each county is associated with a state name, which we will group into regions as defined by the U.S. Census Bureau: Northeast, Midwest, South, and West.

3 Continuous Response Variables: A NYT survey about masking behaviors asked people whether they wear a mask in public when they expect to be within 6 feet of another person and calculated the responses for each county for never, rarely, sometimes, frequently, and always mask. We choose to look at the extremes and will examine 3 continuous response variables: never mask, sometimes mask, and always mask.

```
raw <- readr::read_csv("https://evancollins.com/covid_and_demographics.csv")
```

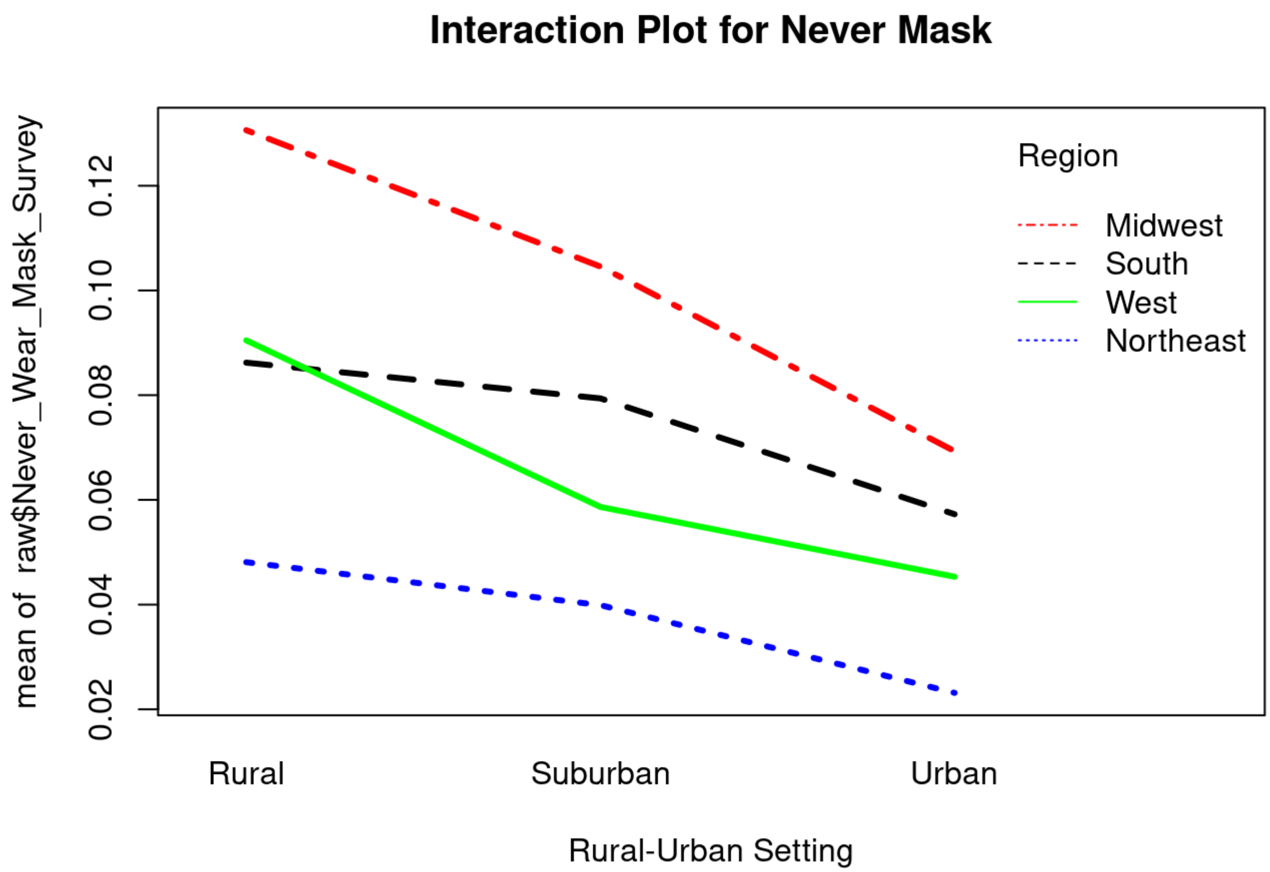
```
## Warning: Missing column names filled in: 'X1' [1]
```

```
##
## ── Column specification ────────────────────────────────────────────────────────
## cols(
##   .default = col_double(),
##   X1 = col_character(),
##   County_Name = col_character(),
##   State_Name = col_character(),
##   FIPS = col_character()
## )
## ℹ Use `spec()` for the full column specifications.
```

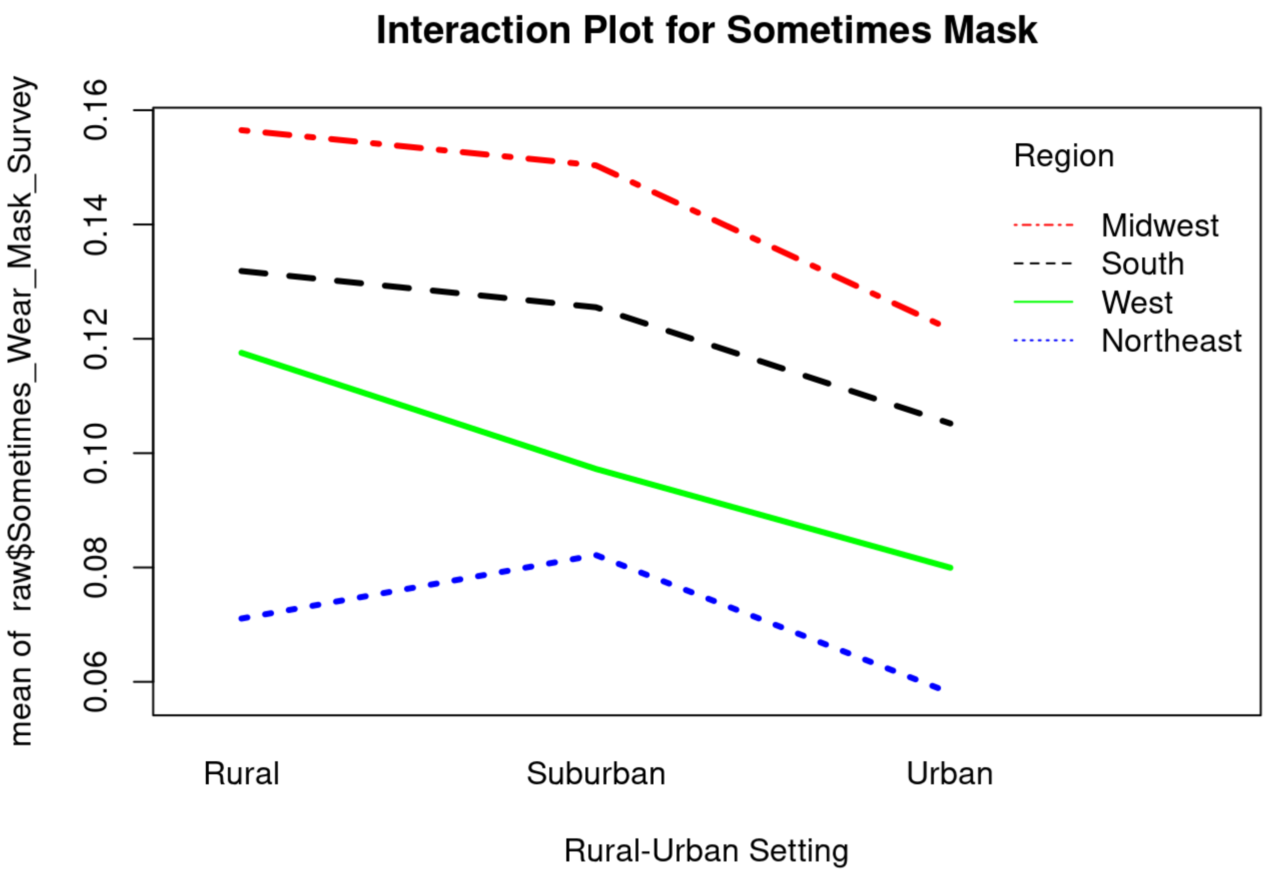
```
# create categorical variables: rural-urban code (3 levels), region (4 variables)
raw <-
  raw %>%
    mutate(region = case_when(
      State_Name %in% c("Washington", "Oregon", "California", "Nevada", "Idaho", "M
ontana", "Utah", "Arizona", "Wyoming", "Colorado", "New Mexico", "Alaska", "Hawaii"
) ~ "West",
      State_Name %in% c("North Dakota", "South Dakota", "Nebraska", "Kansas", "Minn
esota", "Iowa", "Missouri", "Wisconsin", "Illinois", "Michigan", "Indiana", "Ohio")
~ "Midwest",
      State_Name %in% c("Texas", "Oklahoma", "Arkansas", "Louisiana", "Mississippi"
, "Tennessee", "Kentucky", "Alabama", "Georgia", "Florida", "South Carolina", "Nort
h Carolina", "Virginia", "West Virginia", "District of Columbia", "Delaware", "Mary
land") ~ "South",
      State_Name %in% c("Pennsylvania", "New Jersey", "Connecticut", "Rhode Island"
, "Massachusetts", "New Hampshire", "Vermont", "Maine", "New York") ~ "Northeast"),
      rural_urban_code = case_when(
        Rural_Urban_Code_2013 %in% c(1, 2, 3) ~ "Urban",
        Rural_Urban_Code_2013 %in% c(4, 5, 6) ~ "Suburban",
        Rural_Urban_Code_2013 %in% c(7, 8, 9) ~ "Rural")
      )
raw$rural_urban_code <- as.factor(raw$rural_urban_code) # Rural is reference
```

1: Interactions Plot

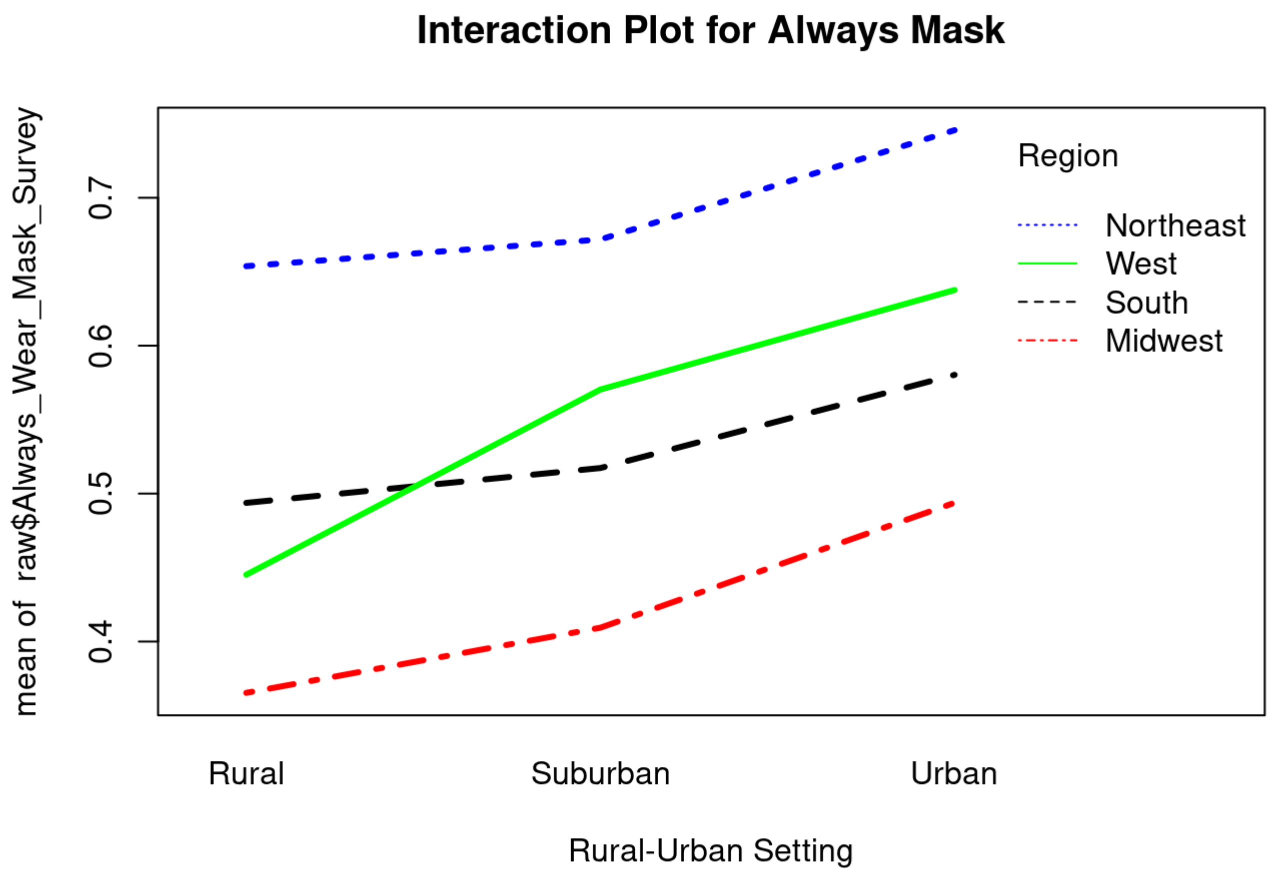
```
interaction.plot(raw$rural_urban_code, raw$region, raw$Never_Wear_Mask_Survey,
  lwd = 3, col = c("red", "blue", "black", "green"), trace.label = "Region",
  xlab = "Rural-Urban Setting", main = "Interaction Plot for Never Mask")
```




```
interaction.plot(raw$rural_urban_code, raw$region, raw$Sometimes_Wear_Mask_Survey,
  lwd = 3, col = c("red", "blue", "black", "green"), trace.label = "Region",
  xlab = "Rural-Urban Setting", main = "Interaction Plot for Sometimes Mask")
```




```
interaction.plot(raw$rural_urban_code, raw$region, raw$Always_Wear_Mask_Survey,
  lwd = 3, col = c("red", "blue", "black", "green"), trace.label = "Region",
  xlab = "Rural-Urban Setting", main = "Interaction Plot for Always Mask")
```



There does appear to be interaction between the rural-urban setting and the region, as evidenced by the non-parallel lines on the interaction plots for never masking, sometimes masking, and always masking. In particular, the West region seems to have behaviors that most contradict those of other regions, particularly in the Western suburbs.

2: Two-Way MANOVA

Univariate:

```
Anova(lm(Never_Wear_Mask_Survey ~ region*rural_urban_code, data = raw), type = 3)
```

```
## Anova Table (Type III tests)
##
## Response: Never_Wear_Mask_Survey
##                         Sum Sq   Df  F value    Pr(>F)
## (Intercept)             7.5571    1 2849.276 < 2.2e-16 ***
## region                  0.5692    3   71.535 < 2.2e-16 ***
## rural_urban_code        0.6754    2  127.318 < 2.2e-16 ***
## region:rural_urban_code 0.1324    6    8.321  5.75e-09 ***
## Residuals               8.2990 3129
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova(lm(Sometimes_Wear_Mask_Survey ~ region*rural_urban_code, data = raw), type =
3)
```

```
## Anova Table (Type III tests)
##
## Response: Sometimes_Wear_Mask_Survey
##                            Sum Sq   Df   F value  Pr(>F)
## (Intercept)               10.8499    1 3966.9689 < 2e-16 ***
## region                     0.3800    3   46.3072 < 2e-16 ***
## rural_urban_code           0.2320    2   42.4113 < 2e-16 ***
## region:rural_urban_code    0.0294    6    1.7926 0.09662 .
## Residuals                  8.5580 3129
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Anova(lm(Always_Wear_Mask_Survey ~ region*rural_urban_code, data = raw), type = 3)
```

```
## Anova Table (Type III tests)
##
## Response: Always_Wear_Mask_Survey
##                           Sum Sq   Df  F value    Pr(>F)
## (Intercept)               59.119    1 4181.798 < 2.2e-16 ***
## region                     4.959    3  116.919 < 2.2e-16 ***
## rural_urban_code           2.979    2  105.361 < 2.2e-16 ***
## region:rural_urban_code    0.934    6   11.017 3.485e-12 ***
## Residuals                 44.236 3129
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Univariate:

The region, rural-urban code, and their interaction are significant in all 3 univariate models for masking behaviors, with p-values << alpha = 0.05. Based on coefficients, region is a more important predictor of sometimes masking, then always masking, then never masking; rural-urban code is a more important predictor of never masking, then sometimes masking, then always masking. but overall a less important predictor than region. Their interaction is most important in never masking, then always masking, then sometimes masking.

Multivariate:

Overall, there are differences between region and rural-urban codes (all multivariate statistics are significant). All of the multivariate tests suggest there is an interaction effect between region and rural-urban code.

```
multimod <- lm(cbind(Never_Wear_Mask_Survey, Sometimes_Wear_Mask_Survey, Always_Wea
r_Mask_Survey) ~ region*rural_urban_code, data = raw)
summary(Anova(multimod), univariate=T)
```

```
##
## Type II MANOVA Tests:
##
## Sum of squares and products for error:
##                            Never_Wear_Mask_Survey Sometimes_Wear_Mask_Survey
## Never_Wear_Mask_Survey                   8.298969                   1.335098
## Sometimes_Wear_Mask_Survey               1.335098                   8.558012
## Always_Wear_Mask_Survey                -11.042686                 -11.266079
##                            Always_Wear_Mask_Survey
## Never_Wear_Mask_Survey                   -11.04269
## Sometimes_Wear_Mask_Survey               -11.26608
## Always_Wear_Mask_Survey                   44.23570
##
## ------------------------------------------
##
## Term: region
##
## Sum of squares and products for the hypothesis:
##                            Never_Wear_Mask_Survey Sometimes_Wear_Mask_Survey
## Never_Wear_Mask_Survey                  0.9675336                   1.031265
## Sometimes_Wear_Mask_Survey              1.0312653                   1.172121
## Always_Wear_Mask_Survey                -3.7941207                  -4.016547
##                            Always_Wear_Mask_Survey
## Never_Wear_Mask_Survey                   -3.794121
## Sometimes_Wear_Mask_Survey               -4.016547
## Always_Wear_Mask_Survey                  15.197097
##
## Multivariate Tests: region
##                  Df test stat approx F num Df   den Df     Pr(>F)
## Pillai            3 0.2828710 108.5832      9 9387.000 < 2.22e-16 ***
## Wilks             3 0.7240752 119.9592      9 7610.447 < 2.22e-16 ***
## Hotelling-Lawley  3 0.3714955 129.0190      9 9377.000 < 2.22e-16 ***
## Roy               3 0.3438340 358.6189      3 3129.000 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: rural_urban_code
##
## Sum of squares and products for the hypothesis:
##                            Never_Wear_Mask_Survey Sometimes_Wear_Mask_Survey
## Never_Wear_Mask_Survey                  0.9798586                  0.7168383
## Sometimes_Wear_Mask_Survey              0.7168383                  0.5430791
## Always_Wear_Mask_Survey                -2.7350157                 -2.0139890
##                            Always_Wear_Mask_Survey
## Never_Wear_Mask_Survey                   -2.735016
## Sometimes_Wear_Mask_Survey               -2.013989
## Always_Wear_Mask_Survey                   7.643303
##
## Multivariate Tests: rural_urban_code
##                  Df test stat  approx F num Df den Df     Pr(>F)
## Pillai            2 0.1625339  92.22958      6   6256 < 2.22e-16 ***
```

```
## Wilks             2 0.8378157  96.42712      6   6254 < 2.22e-16 ***
## Hotelling-Lawley  2 0.1931626 100.63771      6   6252 < 2.22e-16 ***
## Roy               2 0.1909774 199.12574      3   3128 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: region:rural_urban_code
##
## Sum of squares and products for the hypothesis:
##                            Never_Wear_Mask_Survey Sometimes_Wear_Mask_Survey
## Never_Wear_Mask_Survey                 0.13241775                 0.04391835
## Sometimes_Wear_Mask_Survey             0.04391835                 0.02941714
## Always_Wear_Mask_Survey               -0.21615727                -0.13410491
##                            Always_Wear_Mask_Survey
## Never_Wear_Mask_Survey                  -0.2161573
## Sometimes_Wear_Mask_Survey              -0.1341049
## Always_Wear_Mask_Survey                  0.9344807
##
## Multivariate Tests: region:rural_urban_code
##                  Df test stat  approx F num Df   den Df     Pr(>F)
## Pillai            6 0.0418657  7.380650     18 9387.000 < 2.22e-16 ***
## Wilks             6 0.9585755  7.405305     18 8844.977 < 2.22e-16 ***
## Hotelling-Lawley  6 0.0427549  7.424303     18 9377.000 < 2.22e-16 ***
## Roy               6 0.0254642 13.279579      6 3129.000 6.5347e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Type II Sums of Squares
##                           df Never_Wear_Mask_Survey Sometimes_Wear_Mask_Survey
## region                     3                0.96753                   1.172121
## rural_urban_code           2                0.97986                   0.543079
## region:rural_urban_code    6                0.13242                   0.029417
## residuals               3129                8.29897                   8.558012
##                         Always_Wear_Mask_Survey
## region                                 15.19710
## rural_urban_code                        7.64330
## region:rural_urban_code                 0.93448
## residuals                              44.23570
##
##  F-tests
##                         Never_Wear_Mask_Survey Sometimes_Wear_Mask_Survey
## region                                  121.60                     214.28
## rural_urban_code                        123.15                      99.28
## region:rural_urban_code                  16.64                       5.38
##                         Always_Wear_Mask_Survey
## region                                  179.16
## rural_urban_code                         90.11
## region:rural_urban_code                  11.02
##
##  p-values
##                         Never_Wear_Mask_Survey Sometimes_Wear_Mask_Survey
## region                  < 2.22e-16             < 2.22e-16
```

```
## rural_urban_code        < 2.22e-16              < 2.22e-16
## region:rural_urban_code 1.0011e-10              0.0046608
##                         Always_Wear_Mask_Survey
## region                  < 2.22e-16
## rural_urban_code        < 2.22e-16
## region:rural_urban_code 3.4855e-12
```

3 Contrasts (Univariate and Multivariate)

Let's do the following comparisons:

- Multivariate - Rural vs. Suburban
- Multivariate - Rural vs. Urban
- Univariate - Rural vs. Urban
- Multivariate - Interaction between Rural and Urban between regions Northeast and South
- Univariate - Interaction between Rural and Urban between regions Northeast and South

First, let's make a variable `catcomb` that combines both categorical variables.

```
raw$catcomb <- paste(raw$rural_urban_code, raw$region, sep = "")
table(raw$catcomb)
```

```
##
##      RuralMidwest    RuralNortheast        RuralSouth         RuralWest
##               443                29               406               199
##   SuburbanMidwest SuburbanNortheast     SuburbanSouth      SuburbanWest
##               310                58               424               107
##      UrbanMidwest    UrbanNortheast        UrbanSouth         UrbanWest
##               302               130               592               141
```

```
options(contrasts = c("contr.treatment", "contr.poly"))

# Make catcomb a factor
raw$catcomb <- as.factor(raw$catcomb) # RuralMidwest is reference level

# Multivariate - Fit one way MANOVA model
multimod2 <- lm(cbind(Never_Wear_Mask_Survey, Sometimes_Wear_Mask_Survey, Always_We
ar_Mask_Survey) ~ catcomb, data = raw)
# Univariate - Fit one way ANOVA model just for Never_Wear_Mask_Survey
multimodNever <- lm(Never_Wear_Mask_Survey ~ catcomb, data = raw)

contrasts(raw$catcomb)
```

```
##                    RuralNortheast RuralSouth RuralWest SuburbanMidwest
## RuralMidwest                    0          0         0               0
## RuralNortheast                  1          0         0               0
## RuralSouth                      0          1         0               0
## RuralWest                       0          0         1               0
## SuburbanMidwest                 0          0         0               1
## SuburbanNortheast               0          0         0               0
## SuburbanSouth                   0          0         0               0
## SuburbanWest                    0          0         0               0
## UrbanMidwest                    0          0         0               0
## UrbanNortheast                  0          0         0               0
## UrbanSouth                      0          0         0               0
## UrbanWest                       0          0         0               0
##                    SuburbanNortheast SuburbanSouth SuburbanWest UrbanMidwest
## RuralMidwest                       0             0            0            0
## RuralNortheast                     0             0            0            0
## RuralSouth                         0             0            0            0
## RuralWest                          0             0            0            0
## SuburbanMidwest                    0             0            0            0
## SuburbanNortheast                  1             0            0            0
## SuburbanSouth                      0             1            0            0
## SuburbanWest                       0             0            1            0
## UrbanMidwest                       0             0            0            1
## UrbanNortheast                     0             0            0            0
## UrbanSouth                         0             0            0            0
## UrbanWest                          0             0            0            0
##                    UrbanNortheast UrbanSouth UrbanWest
## RuralMidwest                    0          0         0
## RuralNortheast                  0          0         0
## RuralSouth                      0          0         0
## RuralWest                       0          0         0
## SuburbanMidwest                 0          0         0
## SuburbanNortheast               0          0         0
## SuburbanSouth                   0          0         0
## SuburbanWest                    0          0         0
## UrbanMidwest                    0          0         0
## UrbanNortheast                  1          0         0
## UrbanSouth                      0          1         0
## UrbanWest                       0          0         1
```

```
levels(raw$catcomb)
```

```
##  [1] "RuralMidwest"      "RuralNortheast"    "RuralSouth"
##  [4] "RuralWest"         "SuburbanMidwest"   "SuburbanNortheast"
##  [7] "SuburbanSouth"     "SuburbanWest"      "UrbanMidwest"
## [10] "UrbanNortheast"    "UrbanSouth"        "UrbanWest"
```

```
# Get multivariate contrast for Rural vs. Suburban
linearHypothesis(multimod2, "catcombRuralNortheast + catcombRuralSouth + catcombRur
alWest - catcombSuburbanMidwest - catcombSuburbanNortheast - catcombSuburbanSouth -
catcombSuburbanWest = 0")
```

```
##
## Sum of squares and products for the hypothesis:
##                           Never_Wear_Mask_Survey Sometimes_Wear_Mask_Survey
## Never_Wear_Mask_Survey                0.06956023                0.020719275
## Sometimes_Wear_Mask_Survey            0.02071928                0.006171462
## Always_Wear_Mask_Survey              -0.20126412               -0.059948715
##                           Always_Wear_Mask_Survey
## Never_Wear_Mask_Survey                -0.20126412
## Sometimes_Wear_Mask_Survey            -0.05994872
## Always_Wear_Mask_Survey                0.58233336
##
## Sum of squares and products for error:
##                           Never_Wear_Mask_Survey Sometimes_Wear_Mask_Survey
## Never_Wear_Mask_Survey                  8.298969                   1.335098
## Sometimes_Wear_Mask_Survey              1.335098                   8.558012
## Always_Wear_Mask_Survey               -11.042686                 -11.266079
##                           Always_Wear_Mask_Survey
## Never_Wear_Mask_Survey                  -11.04269
## Sometimes_Wear_Mask_Survey              -11.26608
## Always_Wear_Mask_Survey                  44.23570
##
## Multivariate Tests:
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1 0.0156343 16.55496      3   3127 1.1358e-10 ***
## Wilks             1 0.9843657 16.55496      3   3127 1.1358e-10 ***
## Hotelling-Lawley  1 0.0158826 16.55496      3   3127 1.1358e-10 ***
## Roy               1 0.0158826 16.55496      3   3127 1.1358e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

For the multivariate tests between Rural and Suburban shown above, we can see that the Pillai, Wilks, Hotelling-Lawley, and Roy values all have p=1.1358e-10 < 0.05. Thus, we reject the null hypothesis and conclude Rural and Suburban are significantly different.

```
# Get multivariate contrast for Rural vs. Urban
linearHypothesis(multimod2, "catcombRuralNortheast + catcombRuralSouth + catcombRur
alWest - catcombUrbanMidwest - catcombUrbanNortheast - catcombUrbanSouth - catcombU
rbanWest = 0")
```

```
##
## Sum of squares and products for the hypothesis:
##                           Never_Wear_Mask_Survey Sometimes_Wear_Mask_Survey
## Never_Wear_Mask_Survey                 0.4017247                  0.2808438
## Sometimes_Wear_Mask_Survey             0.2808438                  0.1963366
## Always_Wear_Mask_Survey               -1.2512379                 -0.8747344
##                           Always_Wear_Mask_Survey
## Never_Wear_Mask_Survey                 -1.2512379
## Sometimes_Wear_Mask_Survey             -0.8747344
## Always_Wear_Mask_Survey                 3.8971867
##
## Sum of squares and products for error:
##                           Never_Wear_Mask_Survey Sometimes_Wear_Mask_Survey
## Never_Wear_Mask_Survey                  8.298969                   1.335098
## Sometimes_Wear_Mask_Survey              1.335098                   8.558012
## Always_Wear_Mask_Survey               -11.042686                 -11.266079
##                           Always_Wear_Mask_Survey
## Never_Wear_Mask_Survey                  -11.04269
## Sometimes_Wear_Mask_Survey              -11.26608
## Always_Wear_Mask_Survey                  44.23570
##
## Multivariate Tests:
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1 0.0840648 95.66564      3   3127 < 2.22e-16 ***
## Wilks             1 0.9159352 95.66564      3   3127 < 2.22e-16 ***
## Hotelling-Lawley  1 0.0917803 95.66564      3   3127 < 2.22e-16 ***
## Roy               1 0.0917803 95.66564      3   3127 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

For the multivariate tests between Rural and Urban shown above, we can see that the Pillai, Wilks, Hotelling-Lawley, and Roy values all have $p=2.22e-16 < 0.05$. Thus, we reject the null hypothesis and conclude Rural and Urban are significantly different.

```
# Get univariate contrast for Rural vs. Urban
linearHypothesis(multimodNever, "catcombRuralNortheast + catcombRuralSouth + catcom
bRuralWest - catcombUrbanMidwest - catcombUrbanNortheast - catcombUrbanSouth - catc
ombUrbanWest = 0")
```

```
## Linear hypothesis test
##
## Hypothesis:
## catcombRuralNortheast  + catcombRuralSouth  + catcombRuralWest - catcombUrbanMid
west - catcombUrbanNortheast - catcombUrbanSouth - catcombUrbanWest = 0
##
## Model 1: restricted model
## Model 2: Never_Wear_Mask_Survey ~ catcomb
##
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1   3130 8.7007
## 2   3129 8.2990  1   0.40172 151.46 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

For the univariate F-test between Rural and Urban shown above, we can see that the $p<2.2e-16 < 0.05$. Thus, we reject the null hypothesis and conclude Rural and Urban are significantly different for `Never_Wear_Mask_Survey`.

```
#Get multivariate contrast for Northeast,South and Rural,Urban interaction
linearHypothesis(multimod2, "catcombRuralNortheast - catcombUrbanNortheast - catcom
bRuralSouth + catcombUrbanSouth = 0")
```

```
##
## Sum of squares and products for the hypothesis:
##                        Never_Wear_Mask_Survey Sometimes_Wear_Mask_Survey
## Never_Wear_Mask_Survey           0.0003439765                0.001177750
## Sometimes_Wear_Mask_Survey       0.0011777504                0.004032532
## Always_Wear_Mask_Survey          0.0004625952                0.001583892
##                        Always_Wear_Mask_Survey
## Never_Wear_Mask_Survey            0.0004625952
## Sometimes_Wear_Mask_Survey        0.0015838923
## Always_Wear_Mask_Survey           0.0006221191
##
## Sum of squares and products for error:
##                        Never_Wear_Mask_Survey Sometimes_Wear_Mask_Survey
## Never_Wear_Mask_Survey               8.298969                   1.335098
## Sometimes_Wear_Mask_Survey           1.335098                   8.558012
## Always_Wear_Mask_Survey            -11.042686                 -11.266079
##                        Always_Wear_Mask_Survey
## Never_Wear_Mask_Survey                -11.04269
## Sometimes_Wear_Mask_Survey            -11.26608
## Always_Wear_Mask_Survey                44.23570
##
## Multivariate Tests:
##                  Df test stat approx F num Df den Df  Pr(>F)
## Pillai            1 0.0012273 1.280839      3   3127 0.27918
## Wilks             1 0.9987727 1.280839      3   3127 0.27918
## Hotelling-Lawley  1 0.0012288 1.280839      3   3127 0.27918
## Roy               1 0.0012288 1.280839      3   3127 0.27918
```

For the multivariate test above evaluating the interaction between Rural and Urban between regions Northeast and South, the difference is not shown to be significantly different, as the p value (0.27918) of the multivariate tests is greater than 0.05.

```
#Get univariate contrast for Northeast,South and Rural,Urban interaction
linearHypothesis(multimodNever, "catcombRuralNortheast - catcombUrbanNortheast - ca
tcombRuralSouth + catcombUrbanSouth = 0")
```

```
## Linear hypothesis test
##
## Hypothesis:
## catcombRuralNortheast - catcombRuralSouth - catcombUrbanNortheast  + catcombUrba
nSouth = 0
##
## Model 1: restricted model
## Model 2: Never_Wear_Mask_Survey ~ catcomb
##
##   Res.Df    RSS Df  Sum of Sq      F Pr(>F)
## 1   3130 8.2993
## 2   3129 8.2990  1 0.00034398 0.1297 0.7188
```

For the univariate test above evaluating the interaction between Rural and Urban between regions Northeast and South, the difference in `Never_Mask_Survey` is not shown to be significantly different, as the p value (0.7188) is greater than 0.05.

4 Multiple-Response Linear Model

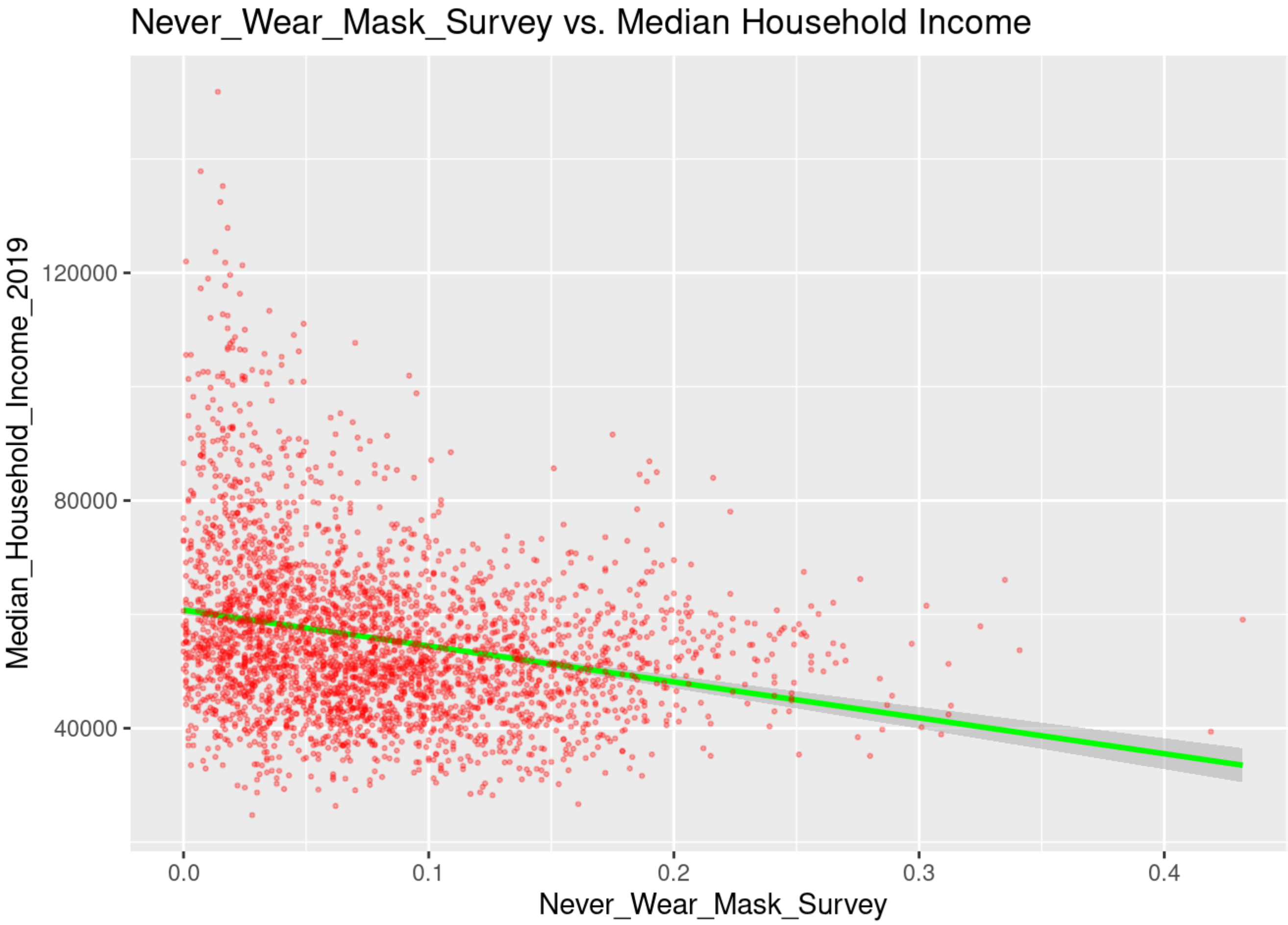
Let's add two other continuous variables as covariates to the model and fit as a multiple-response linear model. We will include ` and Percent_Adults_Bachelors_or_Higher` as covariates.

Let's first plot the relationships between the covariates and the three response variables.

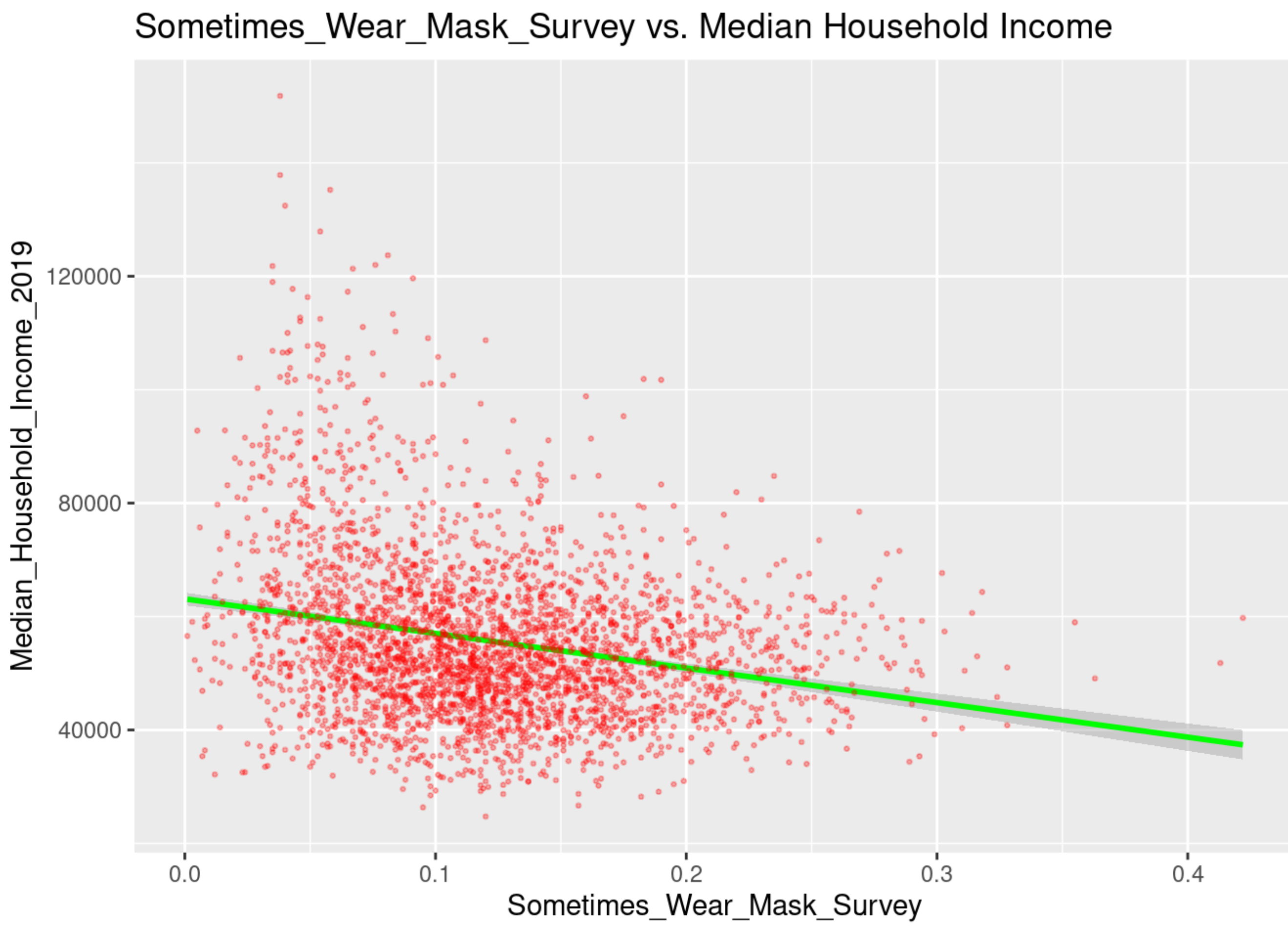
```
names(raw)
```

```
##  [1] "X1"
##  [2] "County_Name"
##  [3] "State_Name"
##  [4] "FIPS"
##  [5] "Never_Wear_Mask_Survey"
##  [6] "Rarely_Wear_Mask_Survey"
##  [7] "Sometimes_Wear_Mask_Survey"
##  [8] "Frequently_Wear_Mask_Survey"
##  [9] "Always_Wear_Mask_Survey"
## [10] "Unemployment_Rate_2019"
## [11] "Median_Household_Income_2019"
## [12] "Median_Household_Income_Percent_of_State_Total_2019"
## [13] "Percent_Poverty_2019"
## [14] "Percent_Adults_Less_Than_HS"
## [15] "Percent_Adults_Bachelors_or_Higher"
## [16] "Population_2019"
## [17] "Net_Migration_Rate_2019"
## [18] "Death_Rate_2019"
## [19] "Birth_Rate_2019"
## [20] "Rural_Urban_Code_2013"
## [21] "Economic_Typology_2015"
## [22] "Covid_Confirmed_Cases_as_pct"
## [23] "Covid_Deaths_as_pct"
## [24] "Covid_New_Cases_as_pct"
## [25] "Civilian_Labor_Force_2019_as_pct"
## [26] "region"
## [27] "rural_urban_code"
## [28] "catcomb"
```

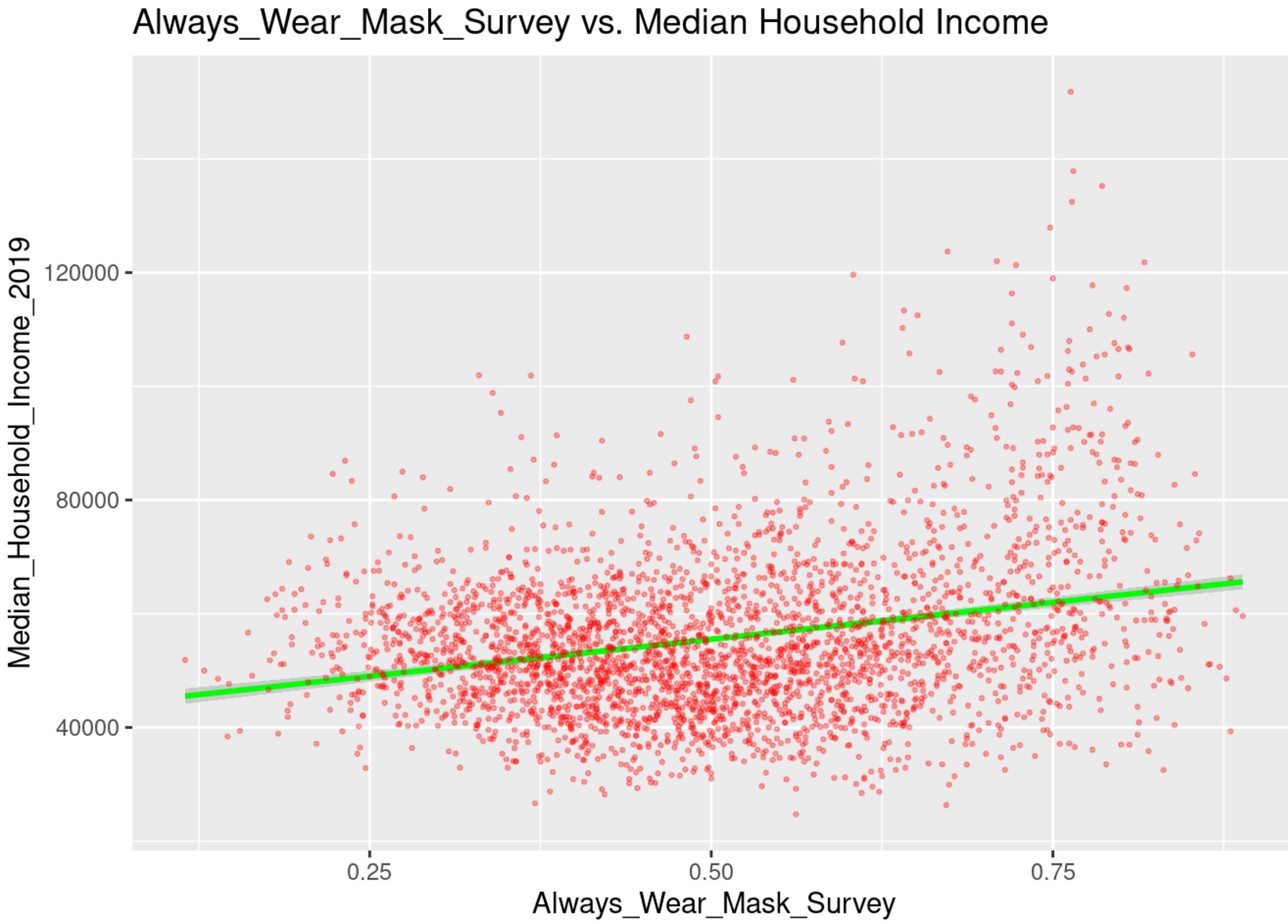
```
# For Median_Household_Income_2019
ggplot(raw, aes(x=Never_Wear_Mask_Survey, y=Median_Household_Income_2019)) + geom_s
mooth(method = lm, color = "green") + geom_point(color = "red", cex=0.5, alpha=0.3)
+ labs(title="Never_Wear_Mask_Survey vs. Median Household Income")
```



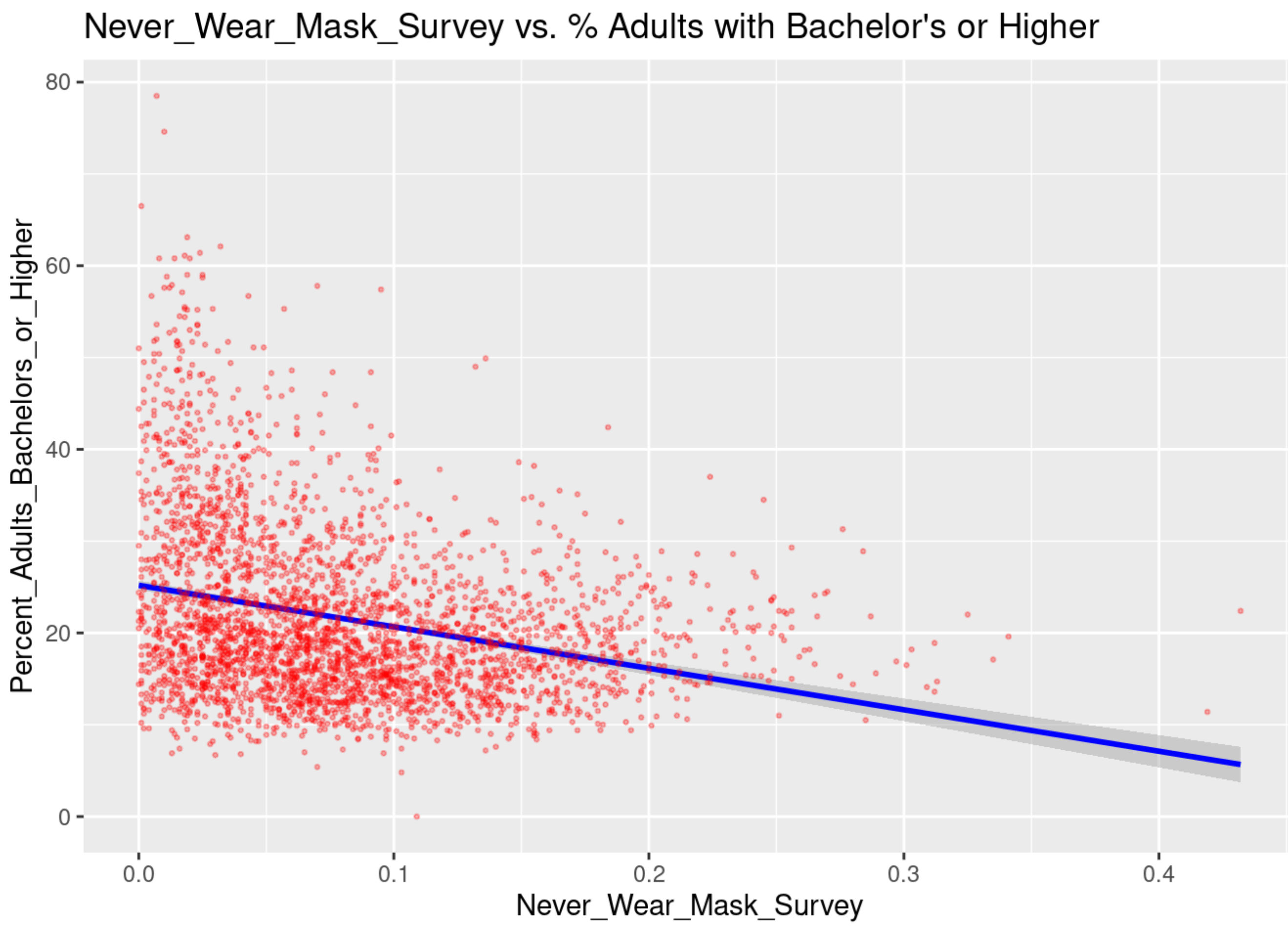

```
ggplot(raw, aes(x=Sometimes_Wear_Mask_Survey, y=Median_Household_Income_2019)) + ge
om_smooth(method = lm, color = "green") + geom_point(color = "red", cex=0.5, alpha=
0.3) + labs(title="Sometimes_Wear_Mask_Survey vs. Median Household Income")
```



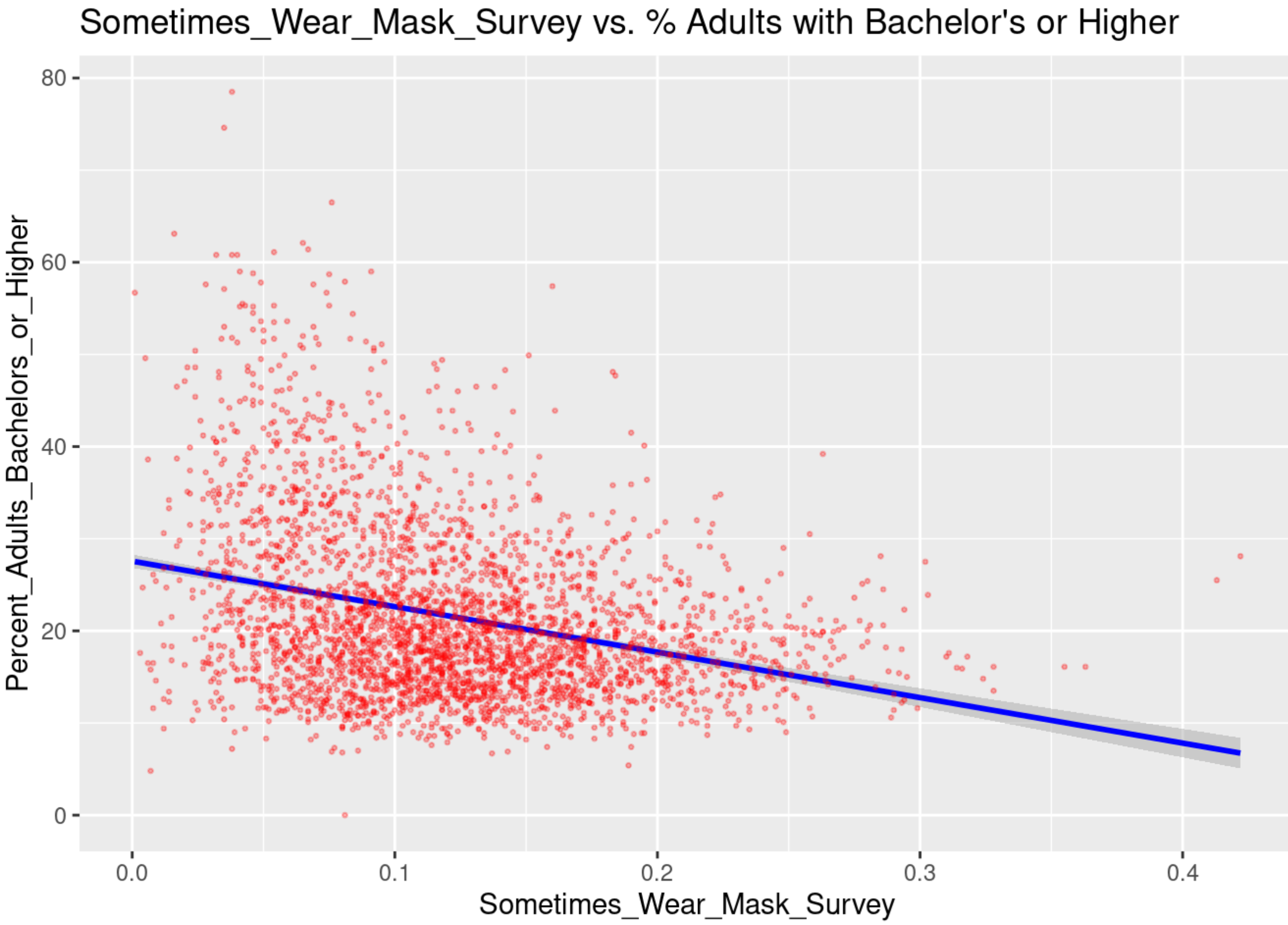

```
ggplot(raw, aes(x=Always_Wear_Mask_Survey, y=Median_Household_Income_2019)) + geom_
smooth(method = lm, color = "green") + geom_point(color = "red", cex=0.5, alpha=0.3
) + labs(title="Always_Wear_Mask_Survey vs. Median Household Income")
```



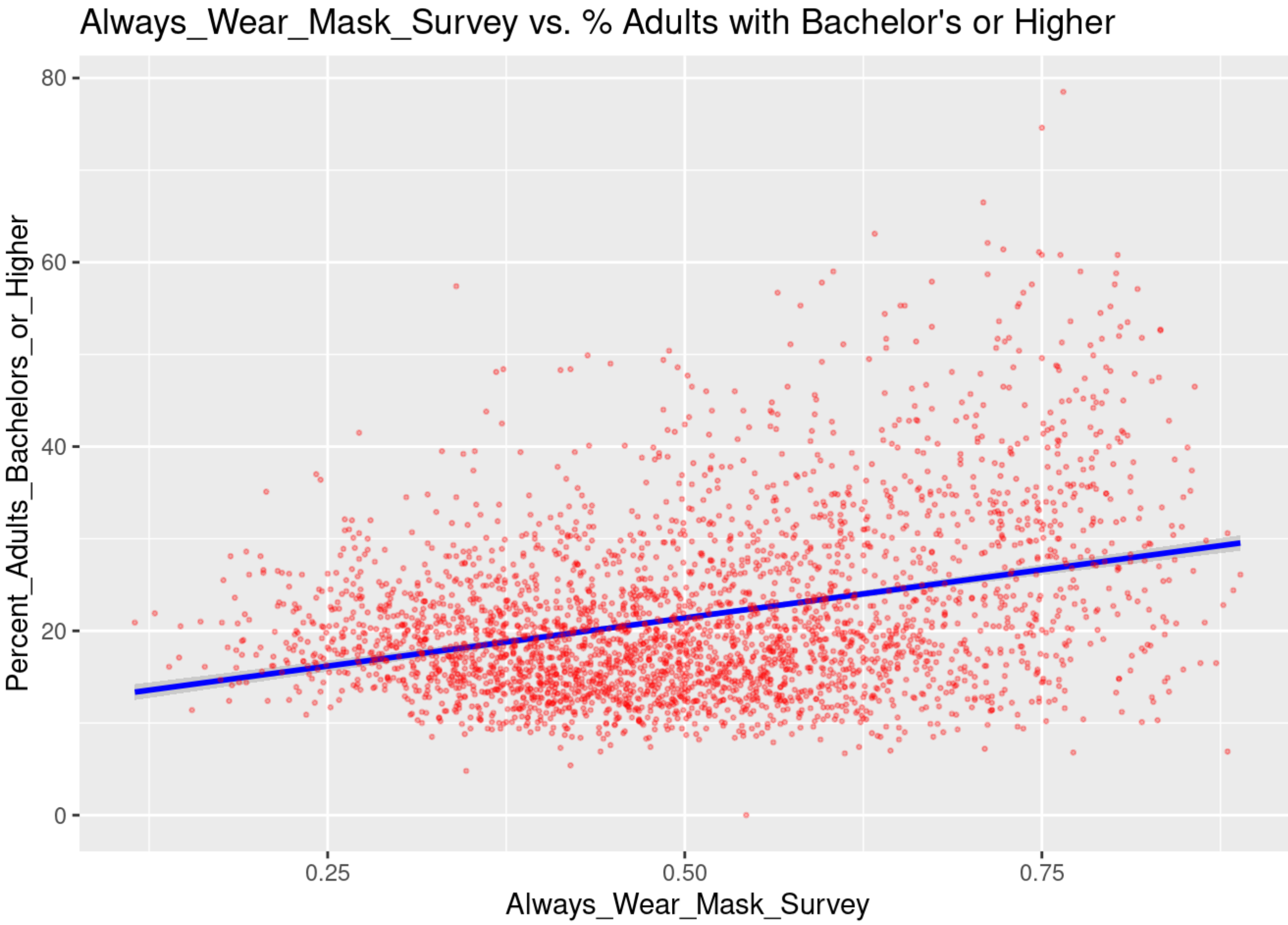

```
# For Percent_Adults_Bachelors_or_Higher
ggplot(raw, aes(x=Never_Wear_Mask_Survey, y=Percent_Adults_Bachelors_or_Higher)) +
 geom_smooth(method = lm, color = "blue") + geom_point(color = "red", cex=0.5, alph
a=0.3) + labs(title="Never_Wear_Mask_Survey vs. % Adults with Bachelor's or Higher"
)
```




```
ggplot(raw, aes(x=Sometimes_Wear_Mask_Survey, y=Percent_Adults_Bachelors_or_Highe
r)) + geom_smooth(method = lm, color = "blue") + geom_point(color = "red", cex=0.5,
alpha=0.3) + labs(title="Sometimes_Wear_Mask_Survey vs. % Adults with Bachelor's or
Higher")
```




```
ggplot(raw, aes(x=Always_Wear_Mask_Survey, y=Percent_Adults_Bachelors_or_Higher)) +
geom_smooth(method = lm, color = "blue") + geom_point(color = "red", cex=0.5, alpha
=0.3) + labs(title="Always_Wear_Mask_Survey vs. % Adults with Bachelor's or Higher"
)
```




```
options(contrasts = c("contr.sum", "contr.poly"))

multimod3 <- lm(cbind(Never_Wear_Mask_Survey, Sometimes_Wear_Mask_Survey, Always_We
ar_Mask_Survey) ~ region*rural_urban_code + Median_Household_Income_2019 + Percent_
Adults_Bachelors_or_Higher, data = raw)

#Multivariate results and univariate results with with type 3 Sum of squares
summary(Anova(multimod3, type = 3), univariate = T)
```

```
##
## Type III MANOVA Tests:
##
## Sum of squares and products for error:
##                            Never_Wear_Mask_Survey Sometimes_Wear_Mask_Survey
## Never_Wear_Mask_Survey                   8.024013                   1.052061
## Sometimes_Wear_Mask_Survey               1.052061                   8.248288
## Always_Wear_Mask_Survey                -10.274843                 -10.429628
##                            Always_Wear_Mask_Survey
## Never_Wear_Mask_Survey                   -10.27484
## Sometimes_Wear_Mask_Survey               -10.42963
## Always_Wear_Mask_Survey                   41.97602
##
## ------------------------------------------
##
## Term: (Intercept)
##
## Sum of squares and products for the hypothesis:
##                            Never_Wear_Mask_Survey Sometimes_Wear_Mask_Survey
## Never_Wear_Mask_Survey                   1.465860                   1.927422
## Sometimes_Wear_Mask_Survey               1.927422                   2.534319
## Always_Wear_Mask_Survey                  6.249751                   8.217641
##                            Always_Wear_Mask_Survey
## Never_Wear_Mask_Survey                    6.249751
## Sometimes_Wear_Mask_Survey                8.217641
## Always_Wear_Mask_Survey                  26.646065
##
## Multivariate Tests: (Intercept)
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1  0.805661  4318.38      3   3125 < 2.22e-16 ***
## Wilks             1  0.194339  4318.38      3   3125 < 2.22e-16 ***
## Hotelling-Lawley  1  4.145645  4318.38      3   3125 < 2.22e-16 ***
## Roy               1  4.145645  4318.38      3   3125 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: region
##
## Sum of squares and products for the hypothesis:
##                            Never_Wear_Mask_Survey Sometimes_Wear_Mask_Survey
## Never_Wear_Mask_Survey                  0.8891676                  0.8878287
## Sometimes_Wear_Mask_Survey              0.8878287                  0.9500496
## Always_Wear_Mask_Survey                -3.4477626                 -3.4579297
##                            Always_Wear_Mask_Survey
## Never_Wear_Mask_Survey                   -3.447763
## Sometimes_Wear_Mask_Survey               -3.457930
## Always_Wear_Mask_Survey                  13.494766
##
## Multivariate Tests: region
##                  Df test stat approx F num Df   den Df     Pr(>F)
## Pillai            3 0.2577033  97.9518      9 9381.000 < 2.22e-16 ***
```

```
## Wilks              3 0.7457420 108.2628      9 7605.579 < 2.22e-16 ***
## Hotelling-Lawley  3 0.3363322 116.7322      9 9371.000 < 2.22e-16 ***
## Roy               3 0.3220783 335.7130      3 3127.000 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: rural_urban_code
##
## Sum of squares and products for the hypothesis:
##                            Never_Wear_Mask_Survey Sometimes_Wear_Mask_Survey
## Never_Wear_Mask_Survey                 0.1804586                 0.10959621
## Sometimes_Wear_Mask_Survey             0.1095962                 0.07673975
## Always_Wear_Mask_Survey               -0.6026318                -0.37800526
##                            Always_Wear_Mask_Survey
## Never_Wear_Mask_Survey                 -0.6026318
## Sometimes_Wear_Mask_Survey             -0.3780053
## Always_Wear_Mask_Survey                 2.0266368
##
## Multivariate Tests: rural_urban_code
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            2 0.0489154 26.12387      6   6252 < 2.22e-16 ***
## Wilks             2 0.9511373 26.42162      6   6250 < 2.22e-16 ***
## Hotelling-Lawley  2 0.0513174 26.71925      6   6248 < 2.22e-16 ***
## Roy               2 0.0502123 52.32121      3   3126 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Median_Household_Income_2019
##
## Sum of squares and products for the hypothesis:
##                            Never_Wear_Mask_Survey Sometimes_Wear_Mask_Survey
## Never_Wear_Mask_Survey                0.03812532                0.014684402
## Sometimes_Wear_Mask_Survey            0.01468440                0.005655864
## Always_Wear_Mask_Survey              -0.04490704               -0.017296456
##                            Always_Wear_Mask_Survey
## Never_Wear_Mask_Survey                -0.04490704
## Sometimes_Wear_Mask_Survey            -0.01729646
## Always_Wear_Mask_Survey                0.05289508
##
## Multivariate Tests: Median_Household_Income_2019
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1 0.0053326 5.584584      3   3125 0.00081009 ***
## Wilks             1 0.9946674 5.584584      3   3125 0.00081009 ***
## Hotelling-Lawley  1 0.0053612 5.584584      3   3125 0.00081009 ***
## Roy               1 0.0053612 5.584584      3   3125 0.00081009 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
```

```
## Term: Percent_Adults_Bachelors_or_Higher
##
## Sum of squares and products for the hypothesis:
##                           Never_Wear_Mask_Survey Sometimes_Wear_Mask_Survey
## Never_Wear_Mask_Survey                0.07095388                  0.1041369
## Sometimes_Wear_Mask_Survey            0.10413686                  0.1528385
## Always_Wear_Mask_Survey              -0.27609096                 -0.4052103
##                           Always_Wear_Mask_Survey
## Never_Wear_Mask_Survey                 -0.2760910
## Sometimes_Wear_Mask_Survey             -0.4052103
## Always_Wear_Mask_Survey                 1.0743065
##
## Multivariate Tests: Percent_Adults_Bachelors_or_Higher
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1 0.0283637 30.40804      3   3125 < 2.22e-16 ***
## Wilks             1 0.9716363 30.40804      3   3125 < 2.22e-16 ***
## Hotelling-Lawley  1 0.0291917 30.40804      3   3125 < 2.22e-16 ***
## Roy               1 0.0291917 30.40804      3   3125 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: region:rural_urban_code
##
## Sum of squares and products for the hypothesis:
##                           Never_Wear_Mask_Survey Sometimes_Wear_Mask_Survey
## Never_Wear_Mask_Survey                0.15414500                 0.05640962
## Sometimes_Wear_Mask_Survey            0.05640962                 0.03422600
## Always_Wear_Mask_Survey              -0.25568153                -0.15534870
##                           Always_Wear_Mask_Survey
## Never_Wear_Mask_Survey                 -0.2556815
## Sometimes_Wear_Mask_Survey             -0.1553487
## Always_Wear_Mask_Survey                 1.0127658
##
## Multivariate Tests: region:rural_urban_code
##                  Df test stat  approx F num Df  den Df     Pr(>F)
## Pillai            6 0.0460407  8.122949     18 9381.00 < 2.22e-16 ***
## Wilks             6 0.9544997  8.152075     18 8839.32 < 2.22e-16 ***
## Hotelling-Lawley  6 0.0471036  8.174216     18 9371.00 < 2.22e-16 ***
## Roy               6 0.0266350 13.881261      6 3127.00 1.2243e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Type III Sums of Squares
##                                      df Never_Wear_Mask_Survey
## (Intercept)                           1               1.465860
## region                                3               0.889168
## rural_urban_code                      2               0.180459
## Median_Household_Income_2019          1               0.038125
## Percent_Adults_Bachelors_or_Higher    1               0.070954
## region:rural_urban_code               6               0.154145
## residuals                          3127               8.024013
##                                    Sometimes_Wear_Mask_Survey
```

```
## (Intercept)                                         2.5343190
## region                                              0.9500496
## rural_urban_code                                    0.0767398
## Median_Household_Income_2019                        0.0056559
## Percent_Adults_Bachelors_or_Higher                  0.1528385
## region:rural_urban_code                             0.0342260
## residuals                                           8.2482885
##                                    Always_Wear_Mask_Survey
## (Intercept)                                     26.646065
## region                                          13.494766
## rural_urban_code                                 2.026637
## Median_Household_Income_2019                     0.052895
## Percent_Adults_Bachelors_or_Higher               1.074307
## region:rural_urban_code                          1.012766
## residuals                                       41.976018
##
##  F-tests
##                                    Never_Wear_Mask_Survey
## (Intercept)                                        571.25
## region                                             346.51
## rural_urban_code                                    70.33
## Median_Household_Income_2019                        14.86
## Percent_Adults_Bachelors_or_Higher                  27.65
## region:rural_urban_code                             60.07
##                                    Sometimes_Wear_Mask_Survey
## (Intercept)                                            320.26
## region                                                 360.17
## rural_urban_code                                         9.70
## Median_Household_Income_2019                             2.14
## Percent_Adults_Bachelors_or_Higher                      19.31
## region:rural_urban_code                                 12.98
##                                    Always_Wear_Mask_Survey
## (Intercept)                                         992.50
## region                                              167.55
## rural_urban_code                                     75.49
## Median_Household_Income_2019                          0.66
## Percent_Adults_Bachelors_or_Higher                   40.02
## region:rural_urban_code                              12.57
##
##  p-values
##                                    Never_Wear_Mask_Survey
## (Intercept)                        < 2.22e-16
## region                             < 2.22e-16
## rural_urban_code                   < 2.22e-16
## Median_Household_Income_2019       0.00011827
## Percent_Adults_Bachelors_or_Higher 1.5506e-07
## region:rural_urban_code            1.2280e-14
##                                    Sometimes_Wear_Mask_Survey
## (Intercept)                        < 2.22e-16
## region                             < 2.22e-16
## rural_urban_code                   2.2800e-06
## Median_Household_Income_2019       0.14321118
## Percent_Adults_Bachelors_or_Higher 2.0891e-12
## region:rural_urban_code            0.00032053
```

```
##                                       Always_Wear_Mask_Survey
## (Intercept)                           < 2.22e-16
## region                                < 2.22e-16
## rural_urban_code                      < 2.22e-16
## Median_Household_Income_2019          0.68473479
## Percent_Adults_Bachelors_or_Higher < 2.22e-16
## region:rural_urban_code               4.6446e-14
```

In the above output chunk labeled “Term: region”, we can see region is a significant (p<2.22e-16) multivariate predictor.

In the above output chunk labeled “Term: rural_urban_code”, we can see rural-urban code is a significant (p<2.22e-16) multivariate predictor.

In the above output chunk labeled “Term: Median_Household_Income_2019”, we can see median household income is a significant (p=0.00081009) multivariate predictor.

In the above output chunk labeled “Term: Percent_Adults_Bachelors_or_Higher”, we can see median household income is a significant (p<2.22e-16) multivariate predictor.

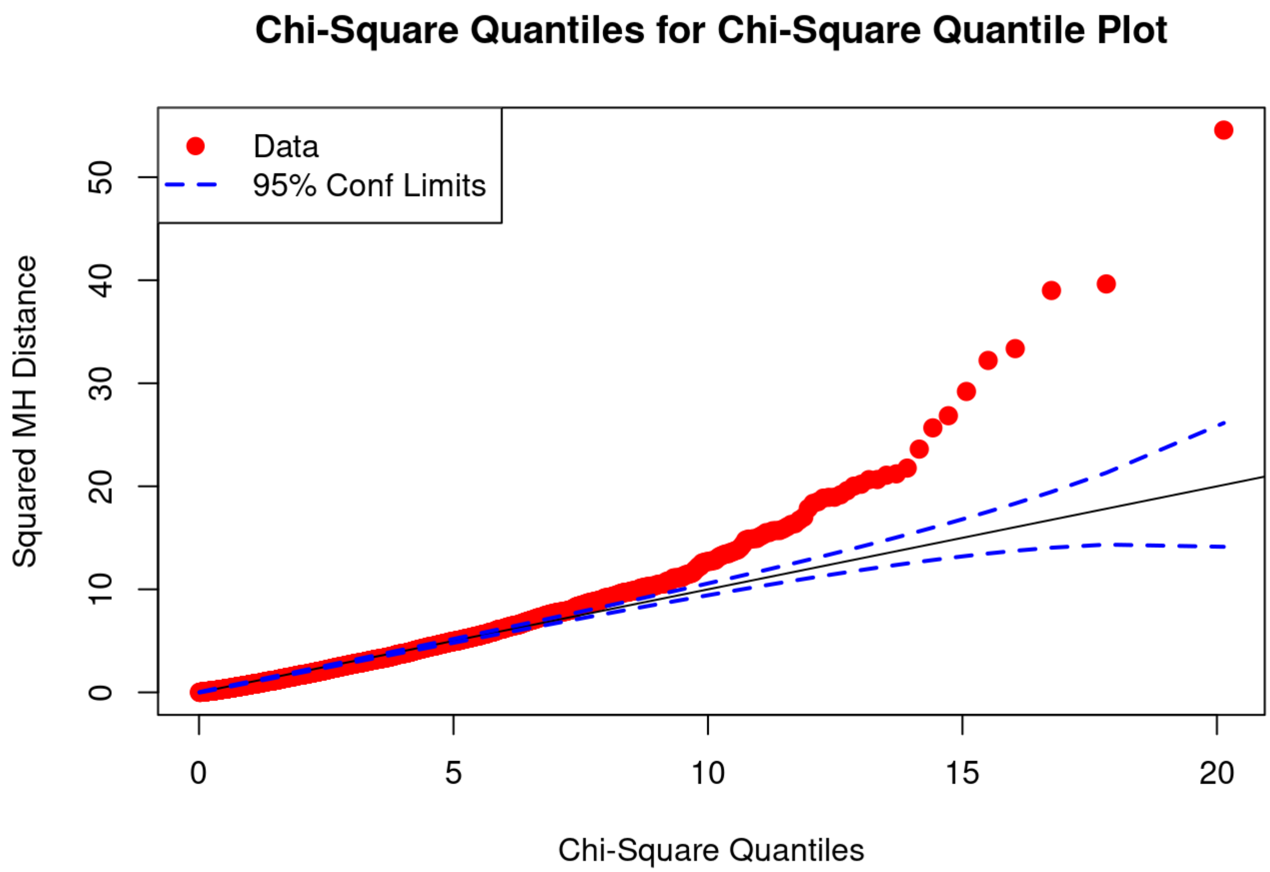
In the above output chunk labeled “Term: region:rural_urban_code”, we can see the interaction between rural and rural-urban code is a significant (p<2.22e-16) multivariate predictor.

In the bottom chunk labeled “Type III Sums of Squares”, for each response variables, we can see the type III sum of squares, type III F-tests, and associated p values. From these univariate results, we can see that region is significant for each of the three response variables (Never_Wear_Mask_Survey, Sometimes_Wear_Mask_Survey, Always_Wear_Mask_Survey). Moreover, rural-urban code is significant for each of the three response variables. Median household income is significant just for Never_Wear_Mask_Survey. Percent of adults with a bachelor’s degree of higher is significant for all three response variables. And the interaction between region and rural-urban code is significant for each of the three response variables.

5 Chi-Square Quantile Plots

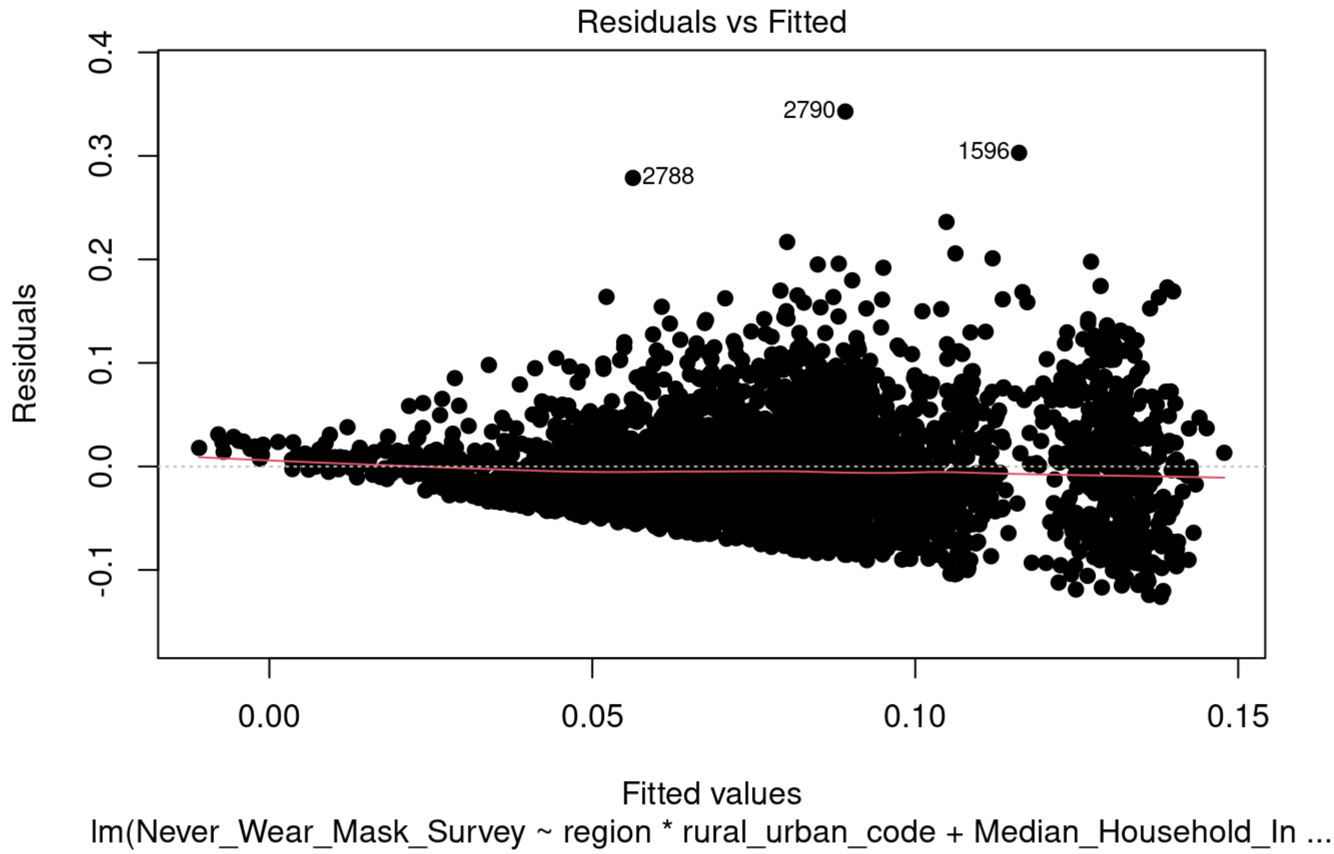
Looking at the CSQ plot for the existing model…

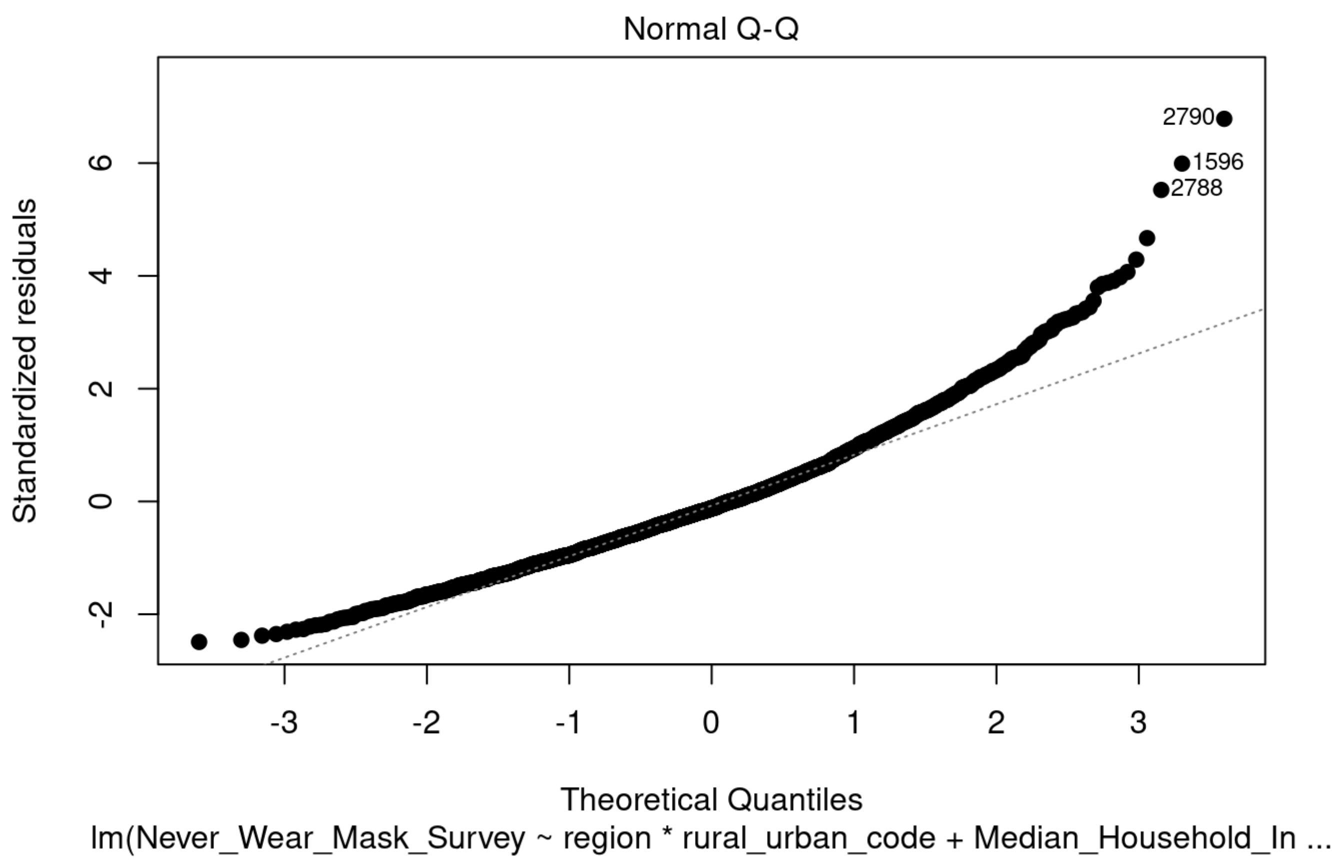
```
CSQPlot(multimod3$residuals)
```



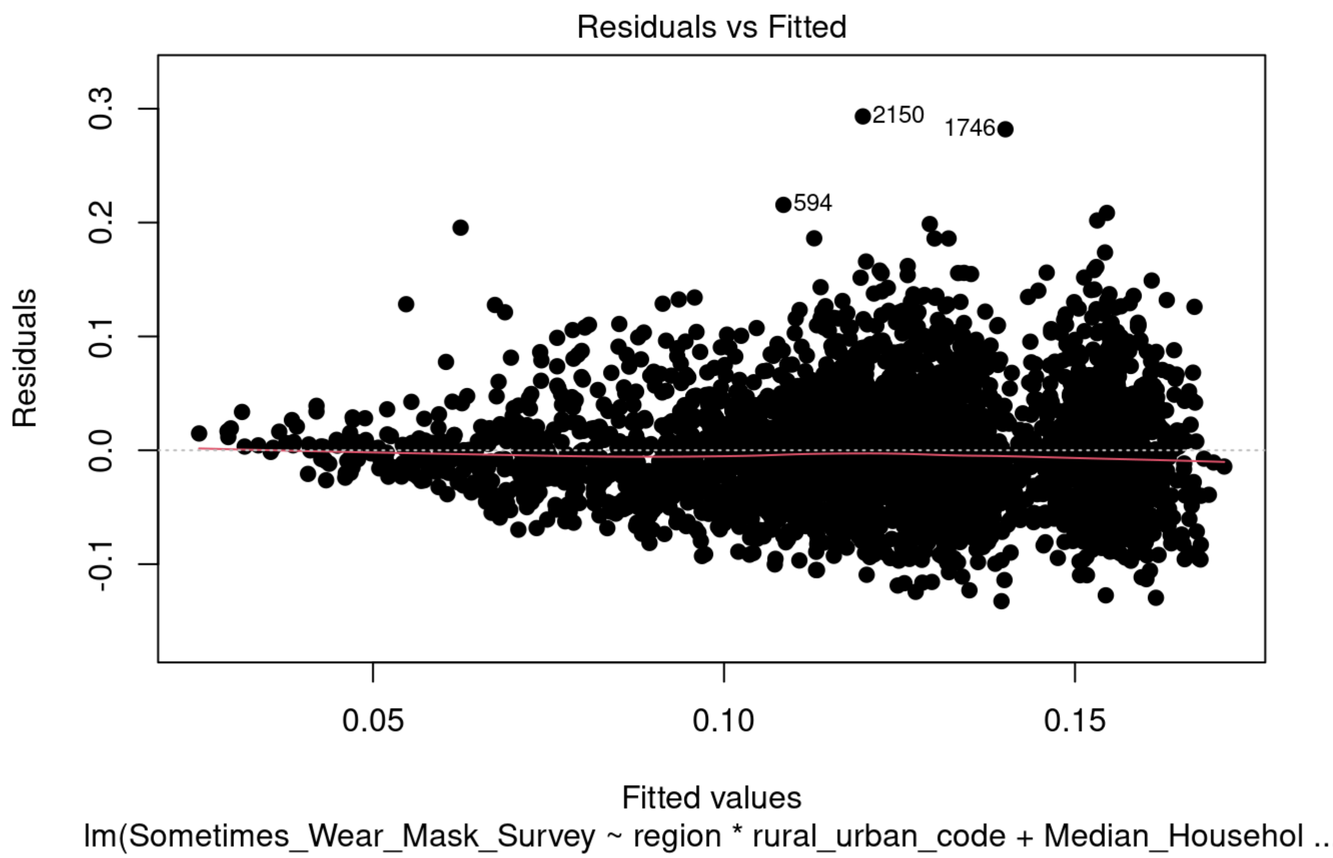
Ooh, I don't love the look of that chi-square quantile plot. Let's take a closer look at the linear model features.

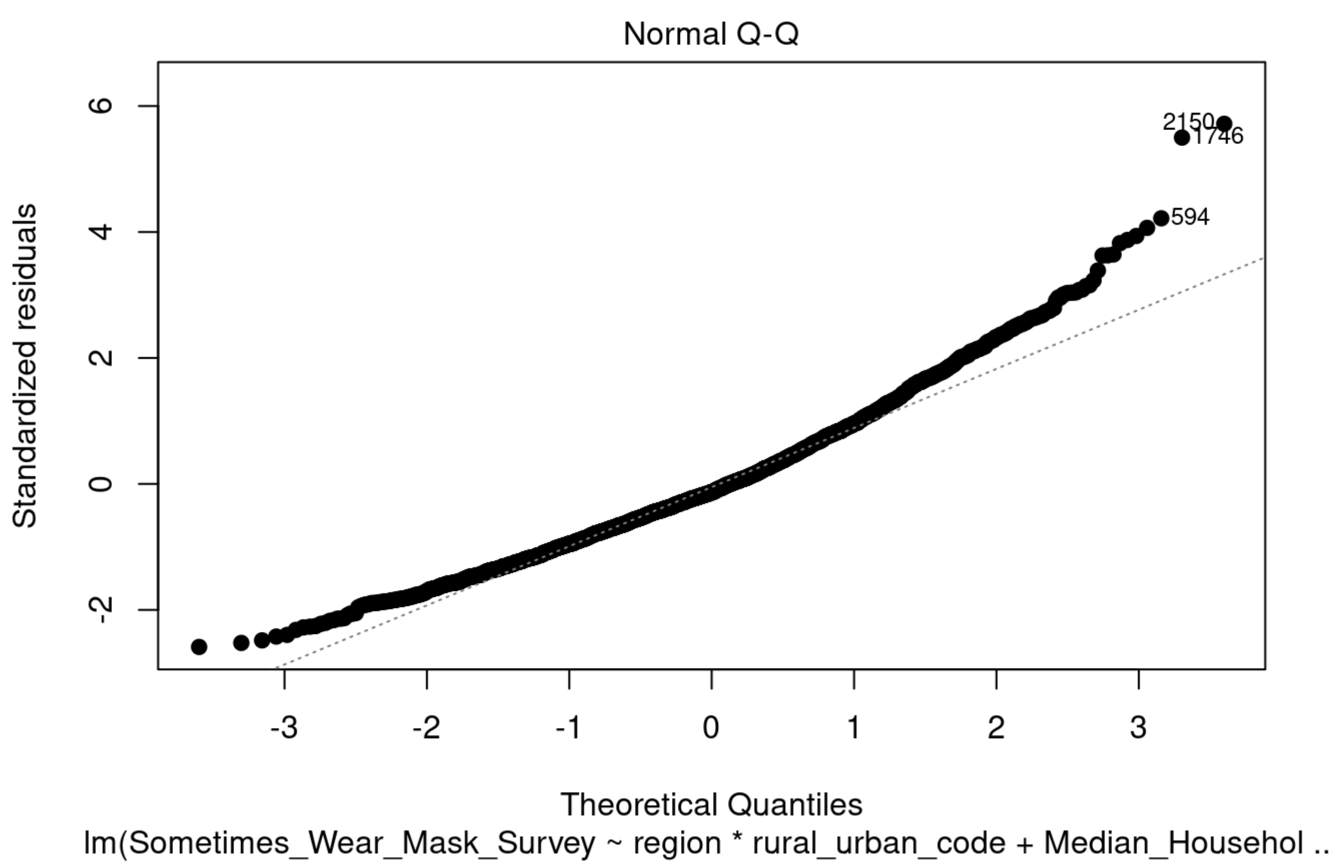
```
modnever <- lm(Never_Wear_Mask_Survey ~ region*rural_urban_code + Median_Household_
Income_2019 + Percent_Adults_Bachelors_or_Higher, data = raw)
modsometimes <- lm(Sometimes_Wear_Mask_Survey ~ region*rural_urban_code + Median_Ho
usehold_Income_2019 + Percent_Adults_Bachelors_or_Higher, data = raw)
modalways <- lm(Always_Wear_Mask_Survey ~ region*rural_urban_code + Median_Househol
d_Income_2019 + Percent_Adults_Bachelors_or_Higher, data = raw)
plot(modnever, which = c(1,2), pch=19)
```


Residuals vs Fitted
Residuals
Fitted values
lm(Never_Wear_Mask_Survey ~ region * rural_urban_code + Median_Household_In ...
2790
1596
2788
0.4
0.3
0.2
0.1
0.0
-0.1
0.00
0.05
0.10
0.15

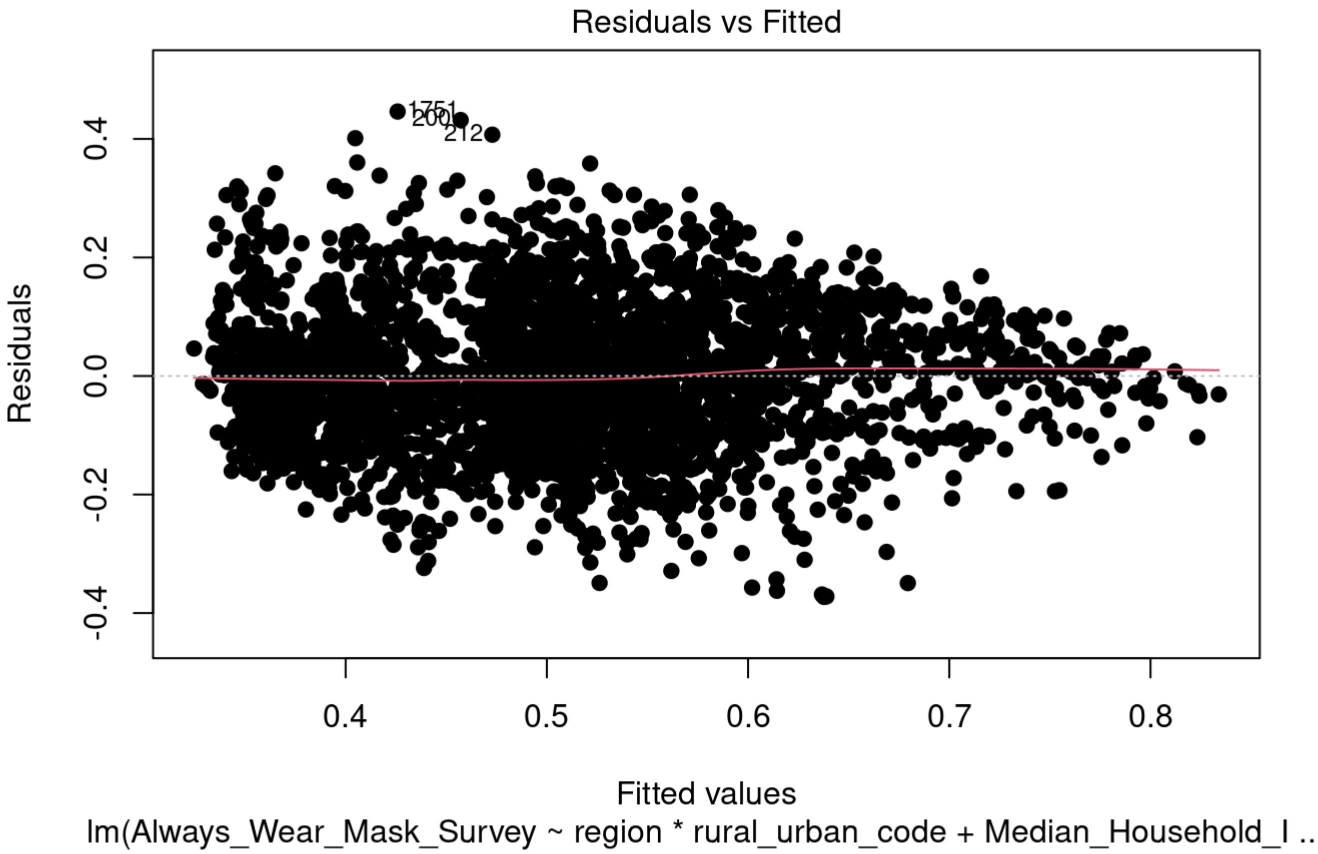

Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(Never_Wear_Mask_Survey ~ region * rural_urban_code + Median_Household_In ...
2790
1596
2788
6
4
2
0
-2
-3
-2
-1
0
1
2
3

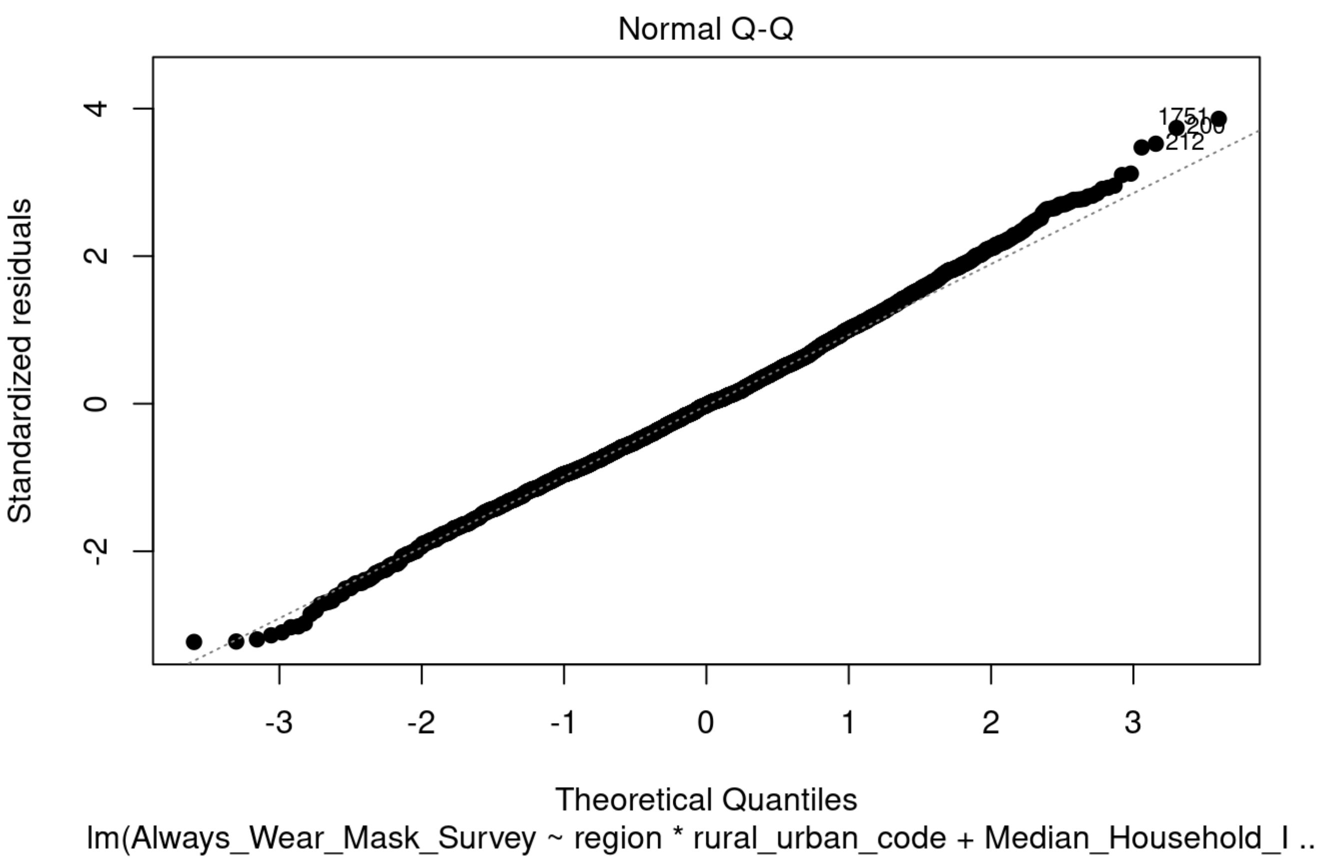
```
plot(modsometimes, which = c(1,2), pch=19)
```


Residuals vs Fitted
Residuals
0.3
0.2
0.1
0.0
-0.1
2150
1746
594
0.05
0.10
0.15
Fitted values
lm(Sometimes_Wear_Mask_Survey ~ region * rural_urban_code + Median_Househol ...


Normal Q-Q
Standardized residuals
6
4
2
0
-2
2150
1746
594
-3
-2
-1
0
1
2
3
Theoretical Quantiles
lm(Sometimes_Wear_Mask_Survey ~ region * rural_urban_code + Median_Househol ...

```
plot(modalways, which = c(1,2), pch=19)
```

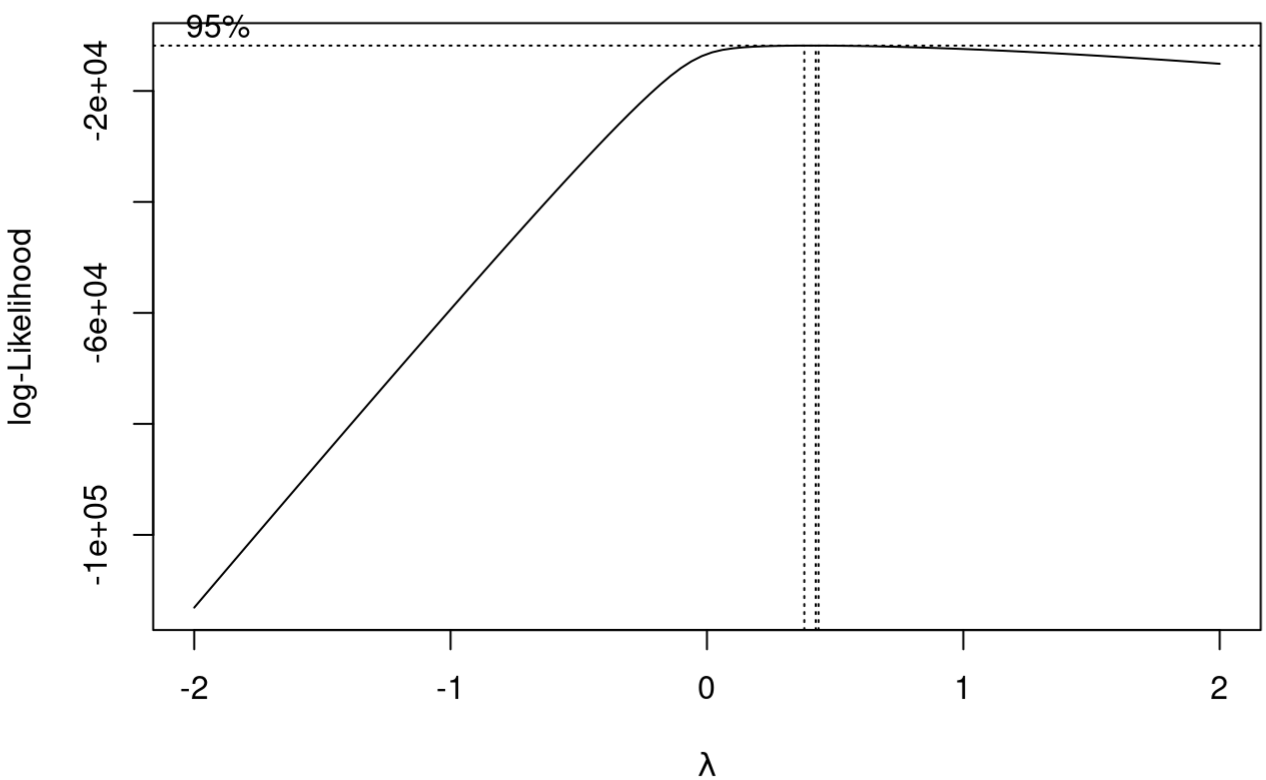

Residuals vs Fitted
Residuals
Fitted values
lm(Always_Wear_Mask_Survey ~ region * rural_urban_code + Median_Household_I ...
0.4
0.2
0.0
-0.2
-0.4
0.4
0.5
0.6
0.7
0.8
1751
200
212


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(Always_Wear_Mask_Survey ~ region * rural_urban_code + Median_Household_I ...
4
2
0
-2
-3
-2
-1
0
1
2
3
1751
200
212

That could do it. There's considerable heteroskedasticity in the data. It looks like we're going to have to try a boxcox transformation.

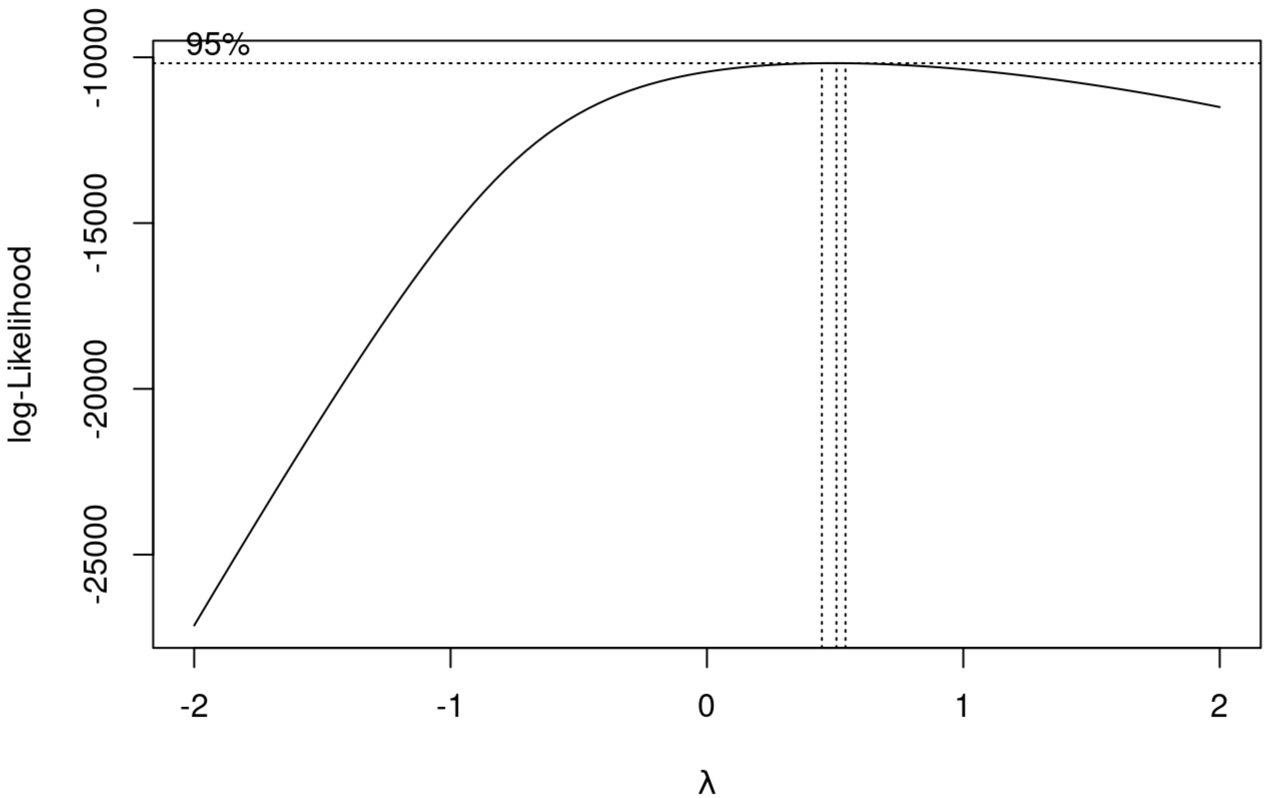
```
bcnever <- MASS::boxcox(Never_Wear_Mask_Survey+1/1000000000 ~ region*rural_urban_co
de + Median_Household_Income_2019 + Percent_Adults_Bachelors_or_Higher, data = raw,
optimize=TRUE)
```

```
## Warning: In lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...) :
##  extra argument 'optimize' will be disregarded
```



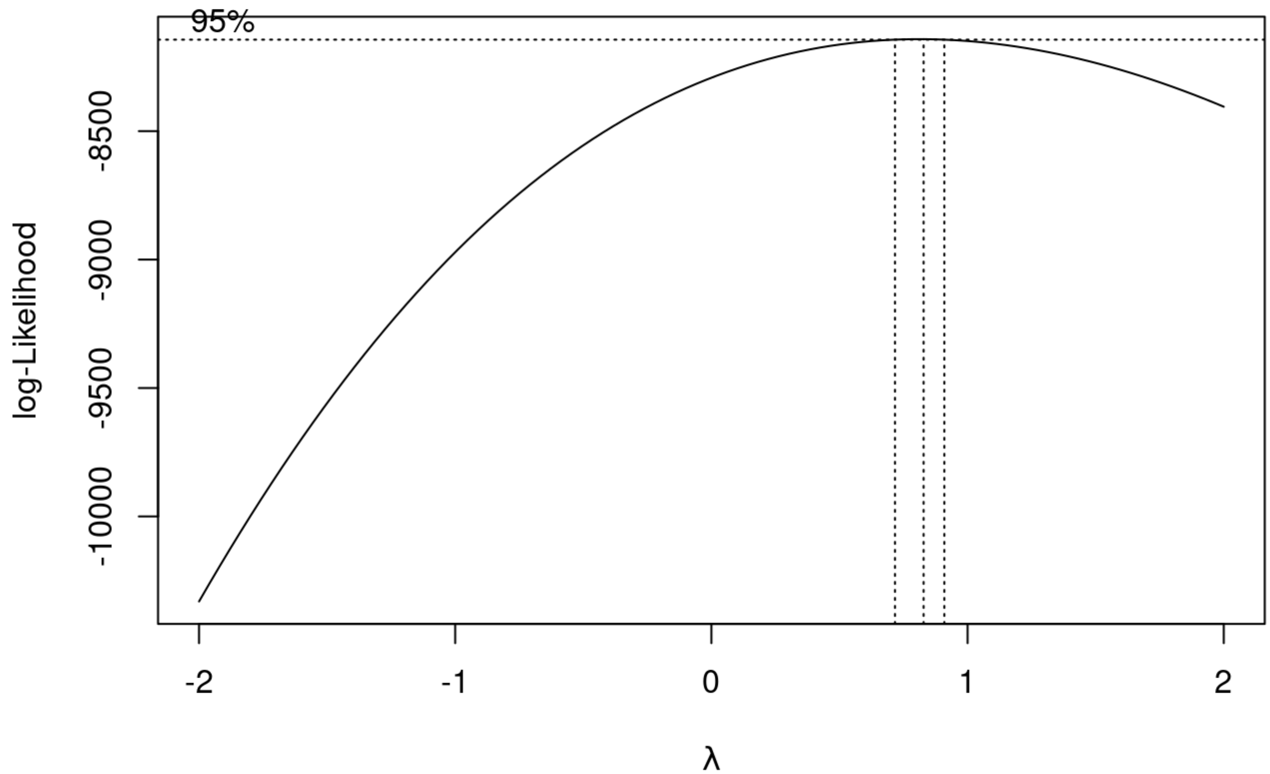

```
lambdanever <- bcnever$x[which.max(bcnever$y)]
bcsometimes <- MASS::boxcox(Sometimes_Wear_Mask_Survey+1/1000000000 ~ region*rural_
urban_code + Median_Household_Income_2019 + Percent_Adults_Bachelors_or_Higher, dat
a = raw, optimize=TRUE)
```

```
## Warning: In lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...) :
##  extra argument 'optimize' will be disregarded
```




```
lambdasometimes <- bcsometimes$x[which.max(bcsometimes$y)]
bcalways <- MASS::boxcox(Always_Wear_Mask_Survey+1/1000000000 ~ region*rural_urban_
code + Median_Household_Income_2019 + Percent_Adults_Bachelors_or_Higher, data = ra
w, optimize=TRUE)
```

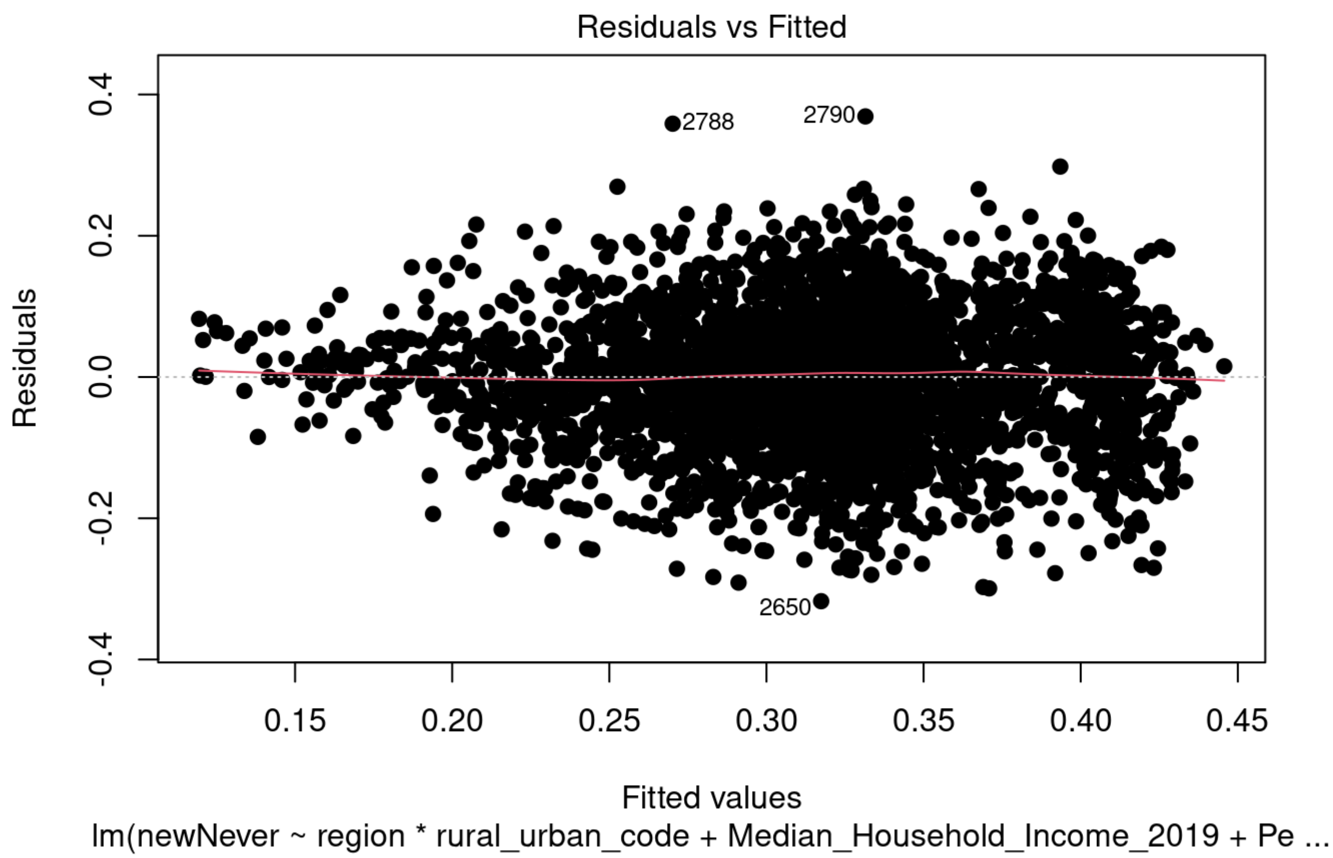
```
## Warning: In lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...) :
##  extra argument 'optimize' will be disregarded
```



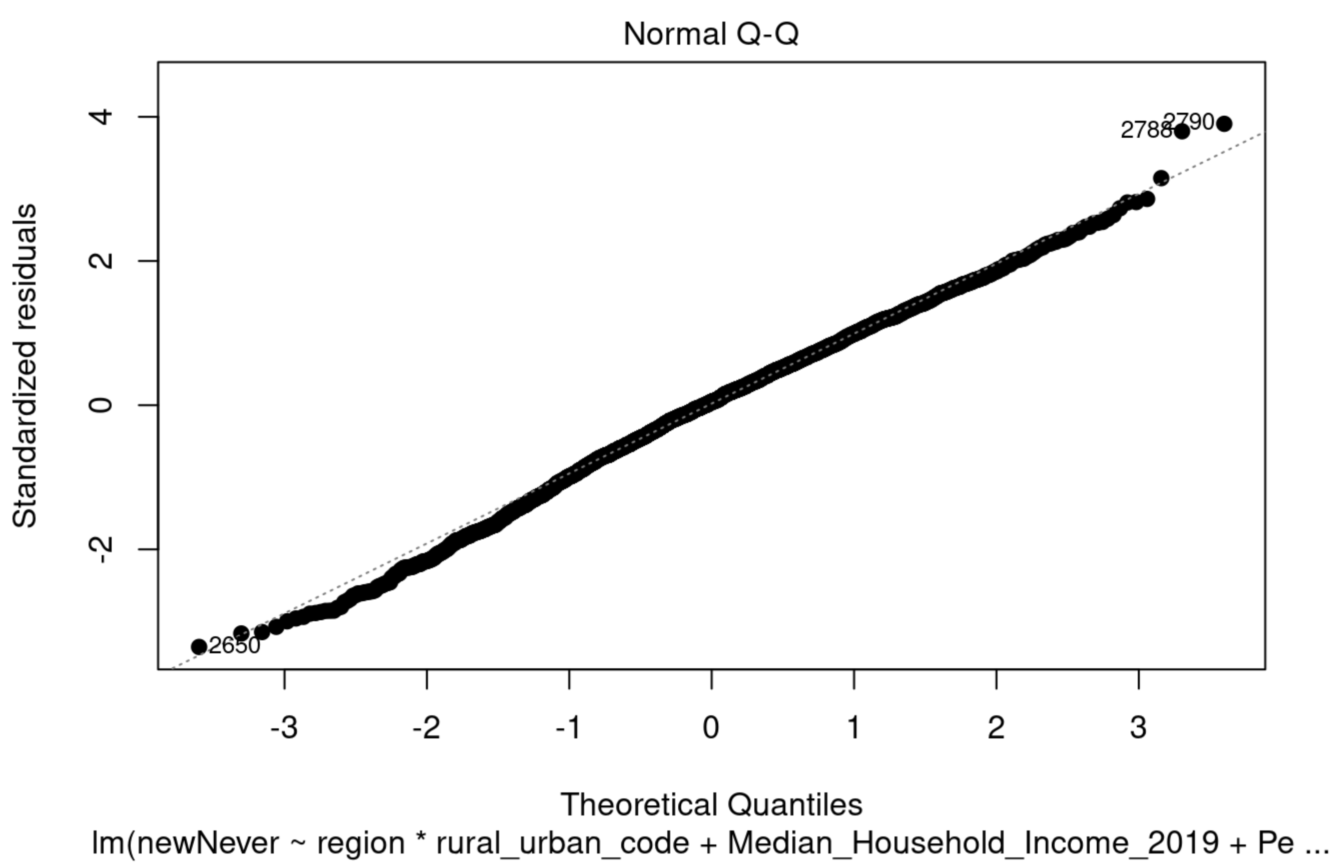

```
lambdaalways <- bcalways$x[which.max(bcalways$y)]
```

We have our lambdas locked and loaded. Note, we added a minuscule amount to the survey values. That's just because the boxcox function requires positive values, and there were some Aileens in the mix.

```
raw$newNever <- (raw$Never_Wear_Mask_Survey)^lambdanever
raw$newSometimes <- (raw$Sometimes_Wear_Mask_Survey)^lambdasometimes
raw$newAlways <- (raw$Always_Wear_Mask_Survey)^lambdaalways

modnever2 <- lm(newNever ~ region*rural_urban_code + Median_Household_Income_2019 +
Percent_Adults_Bachelors_or_Higher, data = raw)
modsometimes2 <- lm(newSometimes ~ region*rural_urban_code + Median_Household_Incom
e_2019 + Percent_Adults_Bachelors_or_Higher, data = raw)
modalways2 <- lm(newAlways ~ region*rural_urban_code + Median_Household_Income_2019
+ Percent_Adults_Bachelors_or_Higher, data = raw)
plot(modnever2, which = c(1,2), pch=19)
```

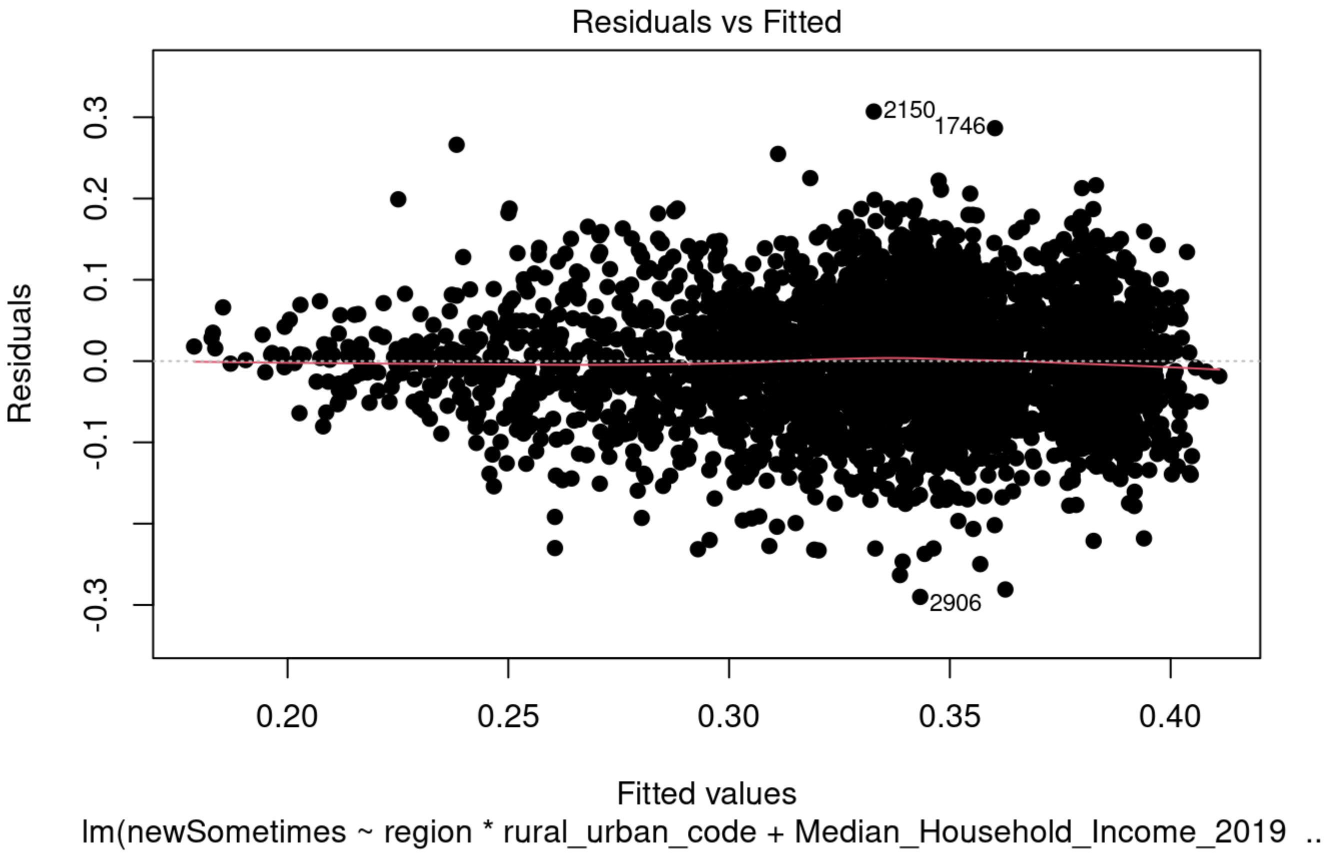


Residuals vs Fitted
2788
2790
2650
Residuals
Fitted values
lm(newNever ~ region * rural_urban_code + Median_Household_Income_2019 + Pe ...

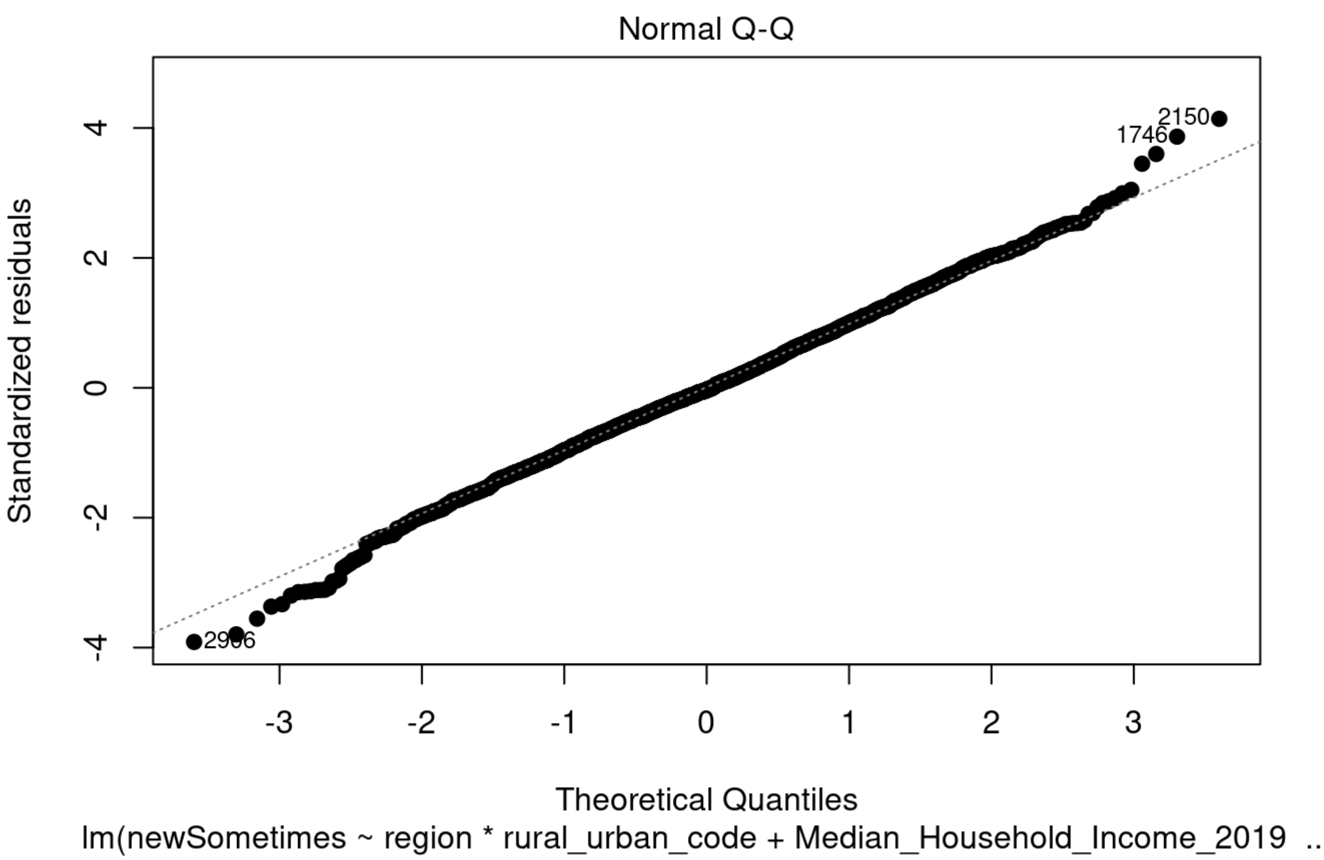


Normal Q-Q
2788
2790
2650
Standardized residuals
Theoretical Quantiles
lm(newNever ~ region * rural_urban_code + Median_Household_Income_2019 + Pe ...

```
plot(modsometimes2, which = c(1,2), pch=19)
```

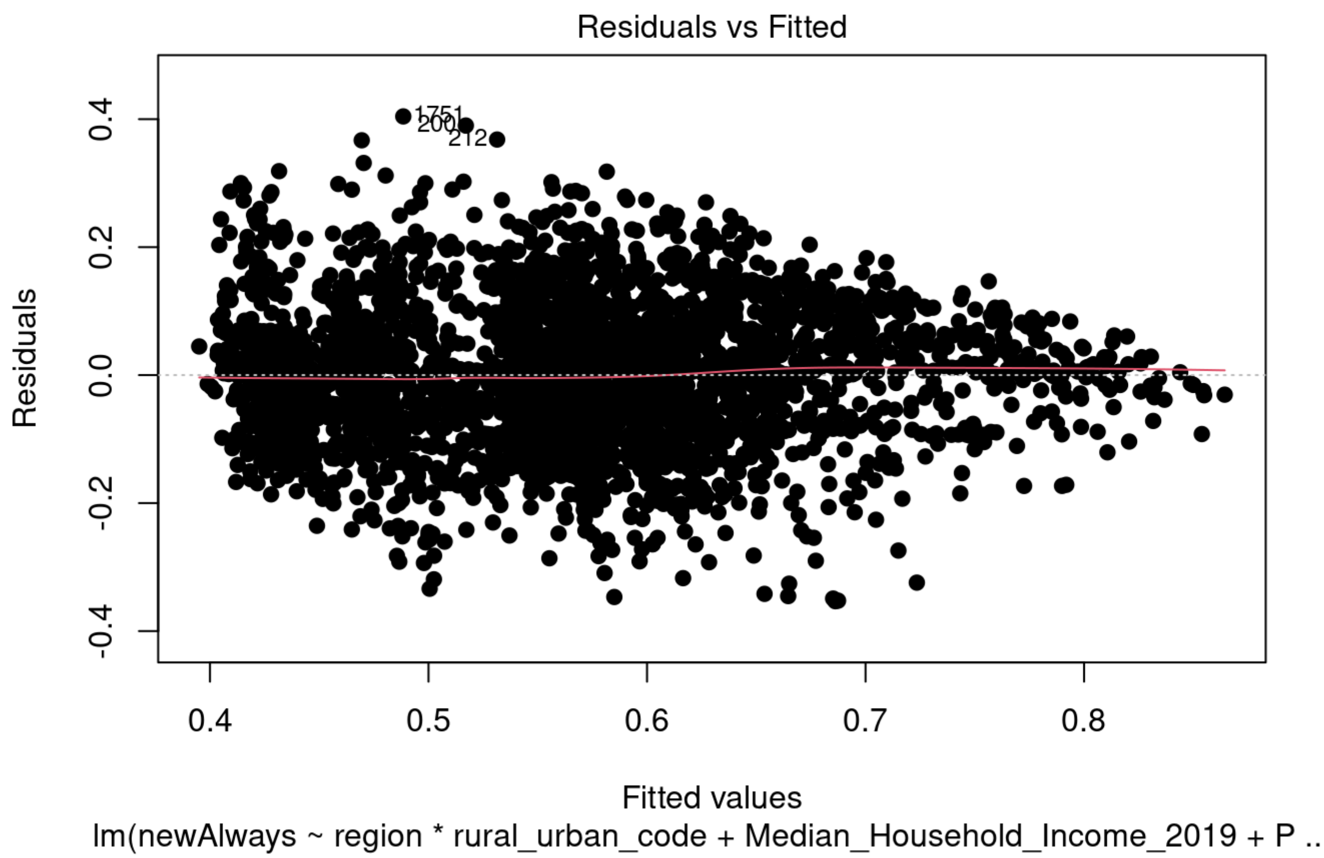


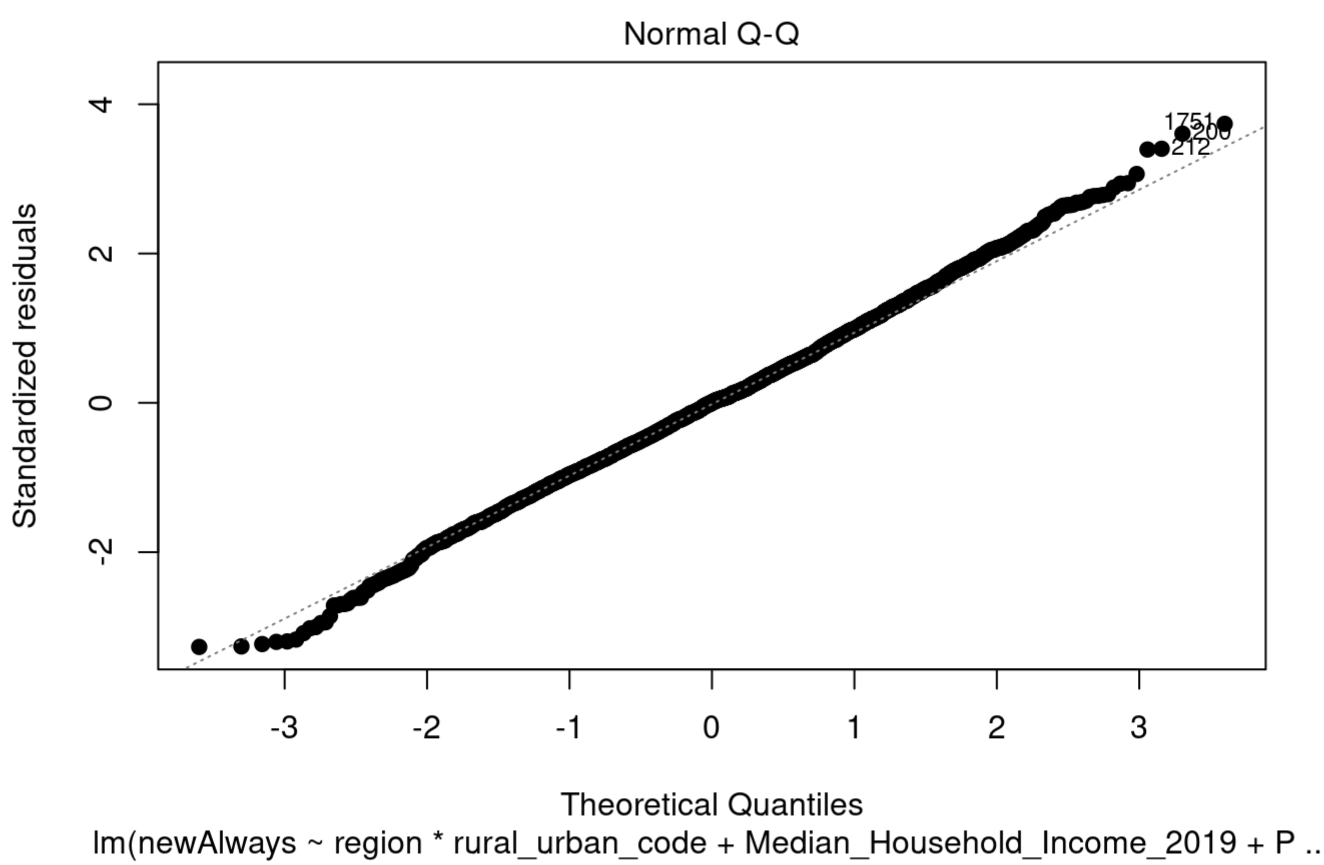
Residuals vs Fitted
Residuals
Fitted values
lm(newSometimes ~ region * rural_urban_code + Median_Household_Income_2019 ...
2150
1746
2906



Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(newSometimes ~ region * rural_urban_code + Median_Household_Income_2019 ...
2150
1746
2906

```
plot(modalways2, which = c(1,2), pch=19)
```


Residuals vs Fitted
Residuals
Fitted values
lm(newAlways ~ region * rural_urban_code + Median_Household_Income_2019 + P ...
1751
200
212
0.4
0.2
0.0
-0.2
-0.4
0.5
0.6
0.7
0.8


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(newAlways ~ region * rural_urban_code + Median_Household_Income_2019 + P ...
1751
200
212
4
2
0
-2
-3
-2
-1
0
1
2
3

After transformation, the heteroskedasticity looks to be fairly diminished. Good job, boxcox! Let's check out the chi square quantile plot for our linear model.

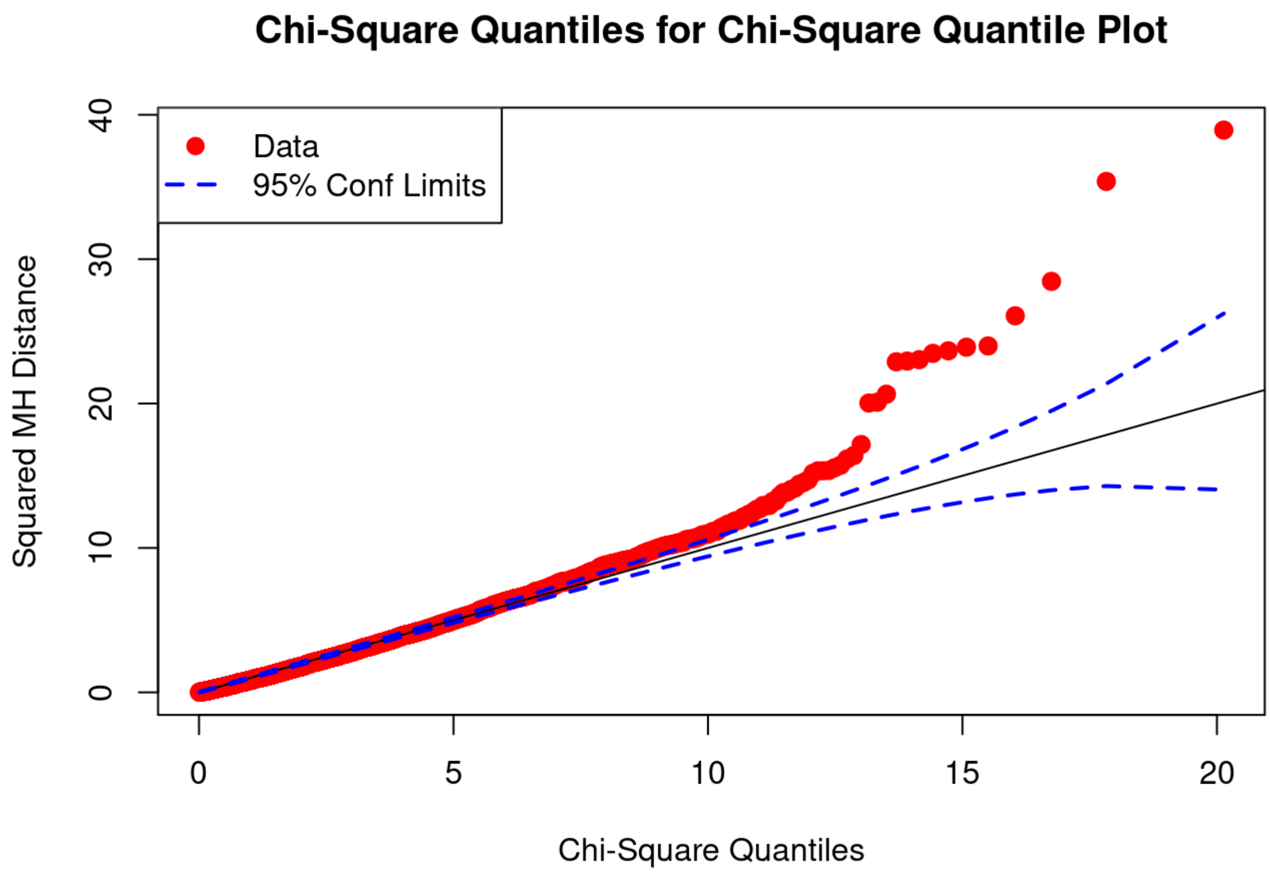
```
multimod4 <- lm(cbind(newNever, newSometimes, newAlways) ~ region*rural_urban_code
+ Median_Household_Income_2019 + Percent_Adults_Bachelors_or_Higher, data = raw)

summary(Anova(multimod4, type = 3), univariate = T)
```

```
##
## Type III MANOVA Tests:
##
## Sum of squares and products for error:
##                newNever newSometimes newAlways
## newNever      28.085717     4.419848 -19.59807
## newSometimes   4.419848    17.235154 -14.57270
## newAlways    -19.598073   -14.572703  36.78109
##
## ------------------------------------------
##
## Term: (Intercept)
##
## Sum of squares and products for the hypothesis:
##              newNever newSometimes newAlways
## newNever     18.53698     17.88614  25.34249
## newSometimes 17.88614     17.25814  24.45270
## newAlways    25.34249     24.45270  34.64652
##
## Multivariate Tests: (Intercept)
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1  0.907757 10250.99      3   3125 < 2.22e-16 ***
## Wilks             1  0.092243 10250.99      3   3125 < 2.22e-16 ***
## Hotelling-Lawley  1  9.840953 10250.99      3   3125 < 2.22e-16 ***
## Roy               1  9.840953 10250.99      3   3125 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: region
##
## Sum of squares and products for the hypothesis:
##               newNever newSometimes newAlways
## newNever      3.455842     2.787809 -6.307450
## newSometimes  2.787809     2.355970 -5.008809
## newAlways    -6.307450    -5.008809 11.693871
##
## Multivariate Tests: region
##                  Df test stat approx F num Df   den Df     Pr(>F)
## Pillai            3 0.2669884 101.8257      9 9381.000 < 2.22e-16 ***
## Wilks             3 0.7390734 111.7878      9 7605.579 < 2.22e-16 ***
## Hotelling-Lawley  3 0.3448502 119.6886      9 9371.000 < 2.22e-16 ***
## Roy               3 0.3192642 332.7797      3 3127.000 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: rural_urban_code
##
## Sum of squares and products for the hypothesis:
##                newNever newSometimes newAlways
```

```
## newNever      0.6100708    0.2695302 -1.027832
## newSometimes  0.2695302    0.1374388 -0.482690
## newAlways    -1.0278315   -0.4826900  1.776192
##
## Multivariate Tests: rural_urban_code
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            2 0.0494034 26.39108      6   6252 < 2.22e-16 ***
## Wilks             2 0.9506508 26.69492      6   6250 < 2.22e-16 ***
## Hotelling-Lawley  2 0.0518540 26.99866      6   6248 < 2.22e-16 ***
## Roy               2 0.0507307 52.86136      3   3126 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Median_Household_Income_2019
##
## Sum of squares and products for the hypothesis:
##                 newNever newSometimes   newAlways
## newNever      0.18233827   0.07310395 -0.08410323
## newSometimes  0.07310395   0.02930919 -0.03371908
## newAlways    -0.08410323  -0.03371908  0.03879248
##
## Multivariate Tests: Median_Household_Income_2019
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1 0.0090479 9.510968      3   3125 2.9819e-06 ***
## Wilks             1 0.9909521 9.510968      3   3125 2.9819e-06 ***
## Hotelling-Lawley  1 0.0091305 9.510968      3   3125 2.9819e-06 ***
## Roy               1 0.0091305 9.510968      3   3125 2.9819e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Percent_Adults_Bachelors_or_Higher
##
## Sum of squares and products for the hypothesis:
##                newNever newSometimes  newAlways
## newNever      0.4151073    0.3784088 -0.6155878
## newSometimes  0.3784088    0.3449547 -0.5611654
## newAlways    -0.6155878   -0.5611654  0.9128926
##
## Multivariate Tests: Percent_Adults_Bachelors_or_Higher
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1 0.0300024 32.21917      3   3125 < 2.22e-16 ***
## Wilks             1 0.9699976 32.21917      3   3125 < 2.22e-16 ***
## Hotelling-Lawley  1 0.0309304 32.21917      3   3125 < 2.22e-16 ***
## Roy               1 0.0309304 32.21917      3   3125 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: region:rural_urban_code
```

```
##
## Sum of squares and products for the hypothesis:
##                newNever newSometimes  newAlways
## newNever      0.3626872    0.1179123 -0.4677310
## newSometimes  0.1179123    0.0685082 -0.2239319
## newAlways    -0.4677310   -0.2239319  0.9235213
##
## Multivariate Tests: region:rural_urban_code
##                  Df test stat  approx F num Df  den Df     Pr(>F)
## Pillai            6 0.0355136  6.243409     18 9381.00 1.4478e-15 ***
## Wilks             6 0.9647280  6.274311     18 8839.32 1.1602e-15 ***
## Hotelling-Lawley  6 0.0363115  6.301391     18 9371.00 9.2897e-16 ***
## Roy               6 0.0275445 14.355291      6 3127.00 3.2681e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Type III Sums of Squares
##                                     df newNever newSometimes newAlways
## (Intercept)                          1 18.53698    17.258144 34.646522
## region                               3  3.45584     2.355970 11.693871
## rural_urban_code                     2  0.61007     0.137439  1.776192
## Median_Household_Income_2019         1  0.18234     0.029309  0.038792
## Percent_Adults_Bachelors_or_Higher   1  0.41511     0.344955  0.912893
## region:rural_urban_code              6  0.36269     0.068508  0.923521
## residuals                         3127 28.08572    17.235154 36.781086
##
##  F-tests
##                                   newNever newSometimes newAlways
## (Intercept)                        2063.87      1043.72   1472.76
## region                              384.77       427.45    165.70
## rural_urban_code                     67.92         8.31     75.50
## Median_Household_Income_2019         20.30         5.32      0.55
## Percent_Adults_Bachelors_or_Higher   46.22        20.86     38.81
## region:rural_urban_code              40.38        12.43     13.09
##
##  p-values
##                                   newNever   newSometimes newAlways
## (Intercept)                       < 2.22e-16 < 2.22e-16   < 2.22e-16
## region                            < 2.22e-16 < 2.22e-16   < 2.22e-16
## rural_urban_code                  2.4707e-16 1.6670e-05   < 2.22e-16
## Median_Household_Income_2019      6.8586e-06 0.02117604   0.77057309
## Percent_Adults_Bachelors_or_Higher 1.2627e-11 2.2216e-13   < 2.22e-16
## region:rural_urban_code           2.3934e-10 0.00042866   1.1207e-14
```

```
CSQPlot(multimod4$residuals)
```

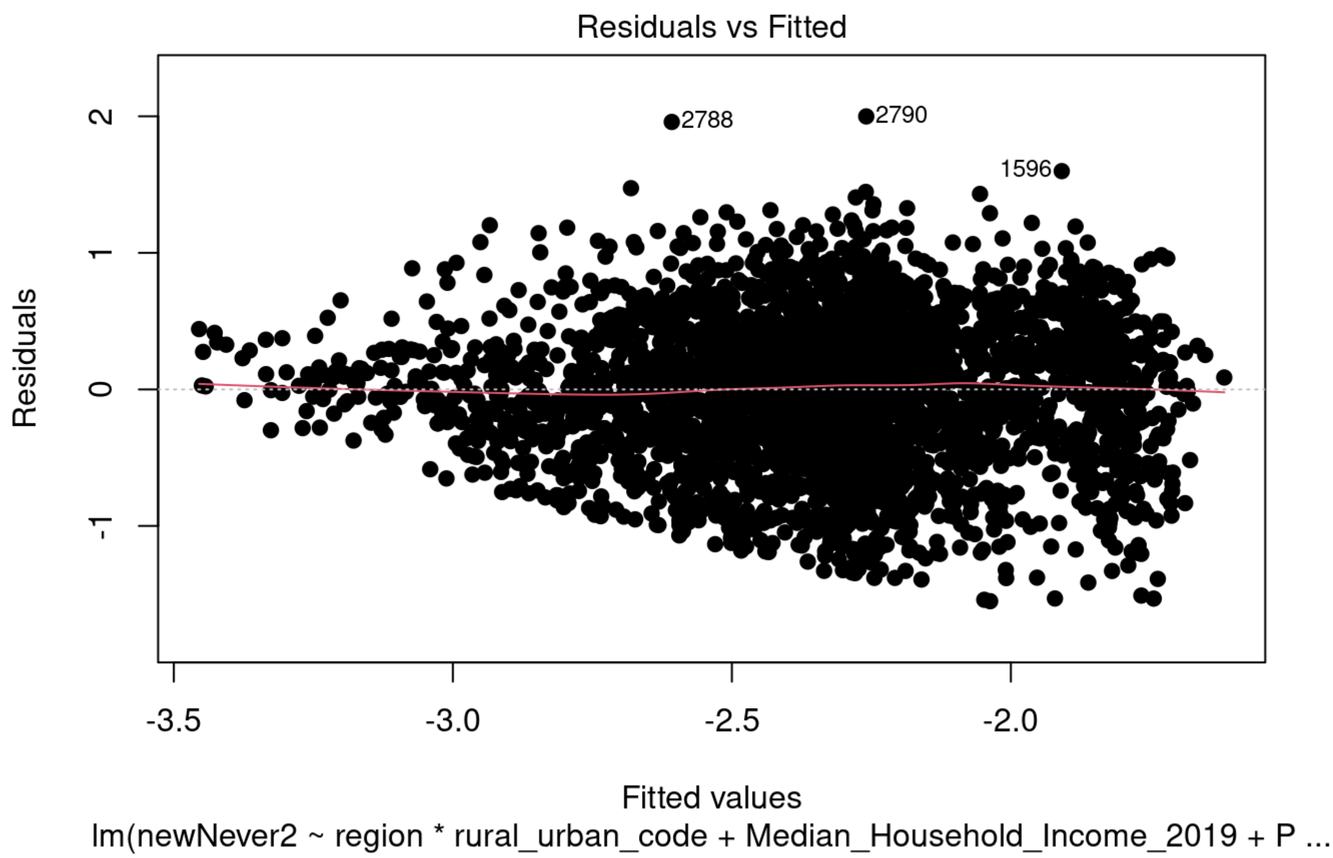


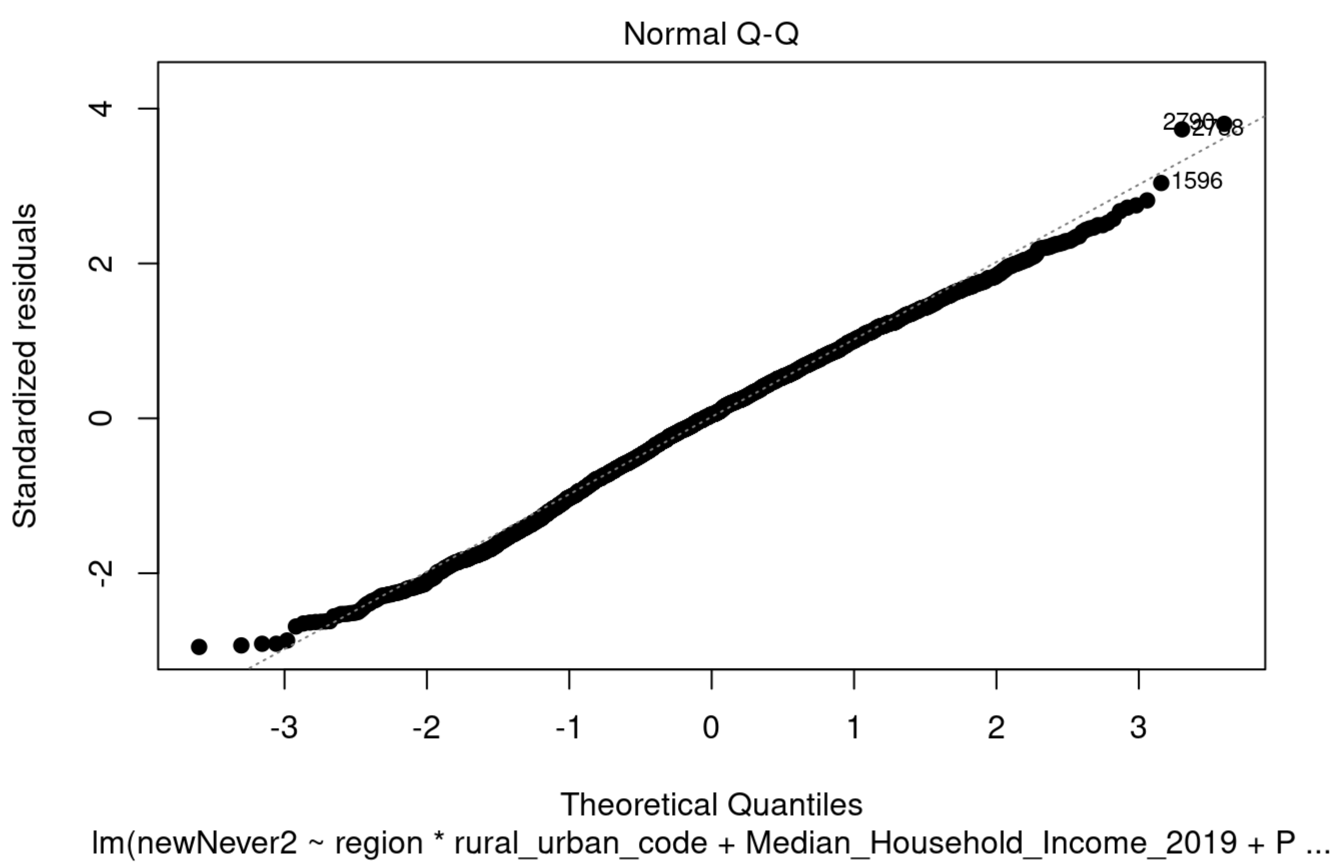
That looks a lot better, especially at the low values. At higher values, the points start to stray outside the 95% CI. Perhaps boxcox isn't really the way to go here. At JDRS's recommendation, let's go for Logit instead.

```
raw$newNever2 <- logit(raw$Never_Wear_Mask_Survey)
```

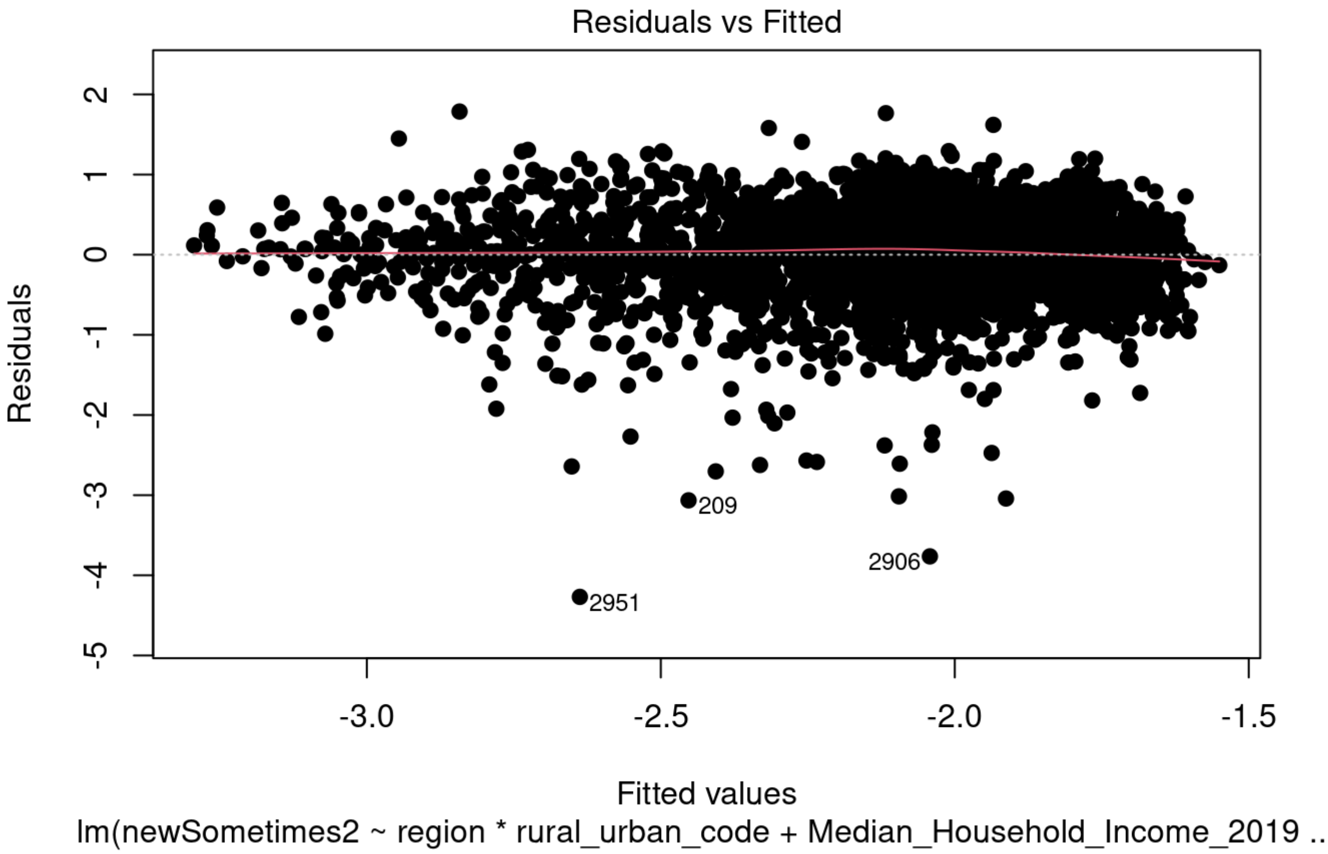
```
## Warning in logit(raw$Never_Wear_Mask_Survey): proportions remapped to (0.025,
## 0.975)
```

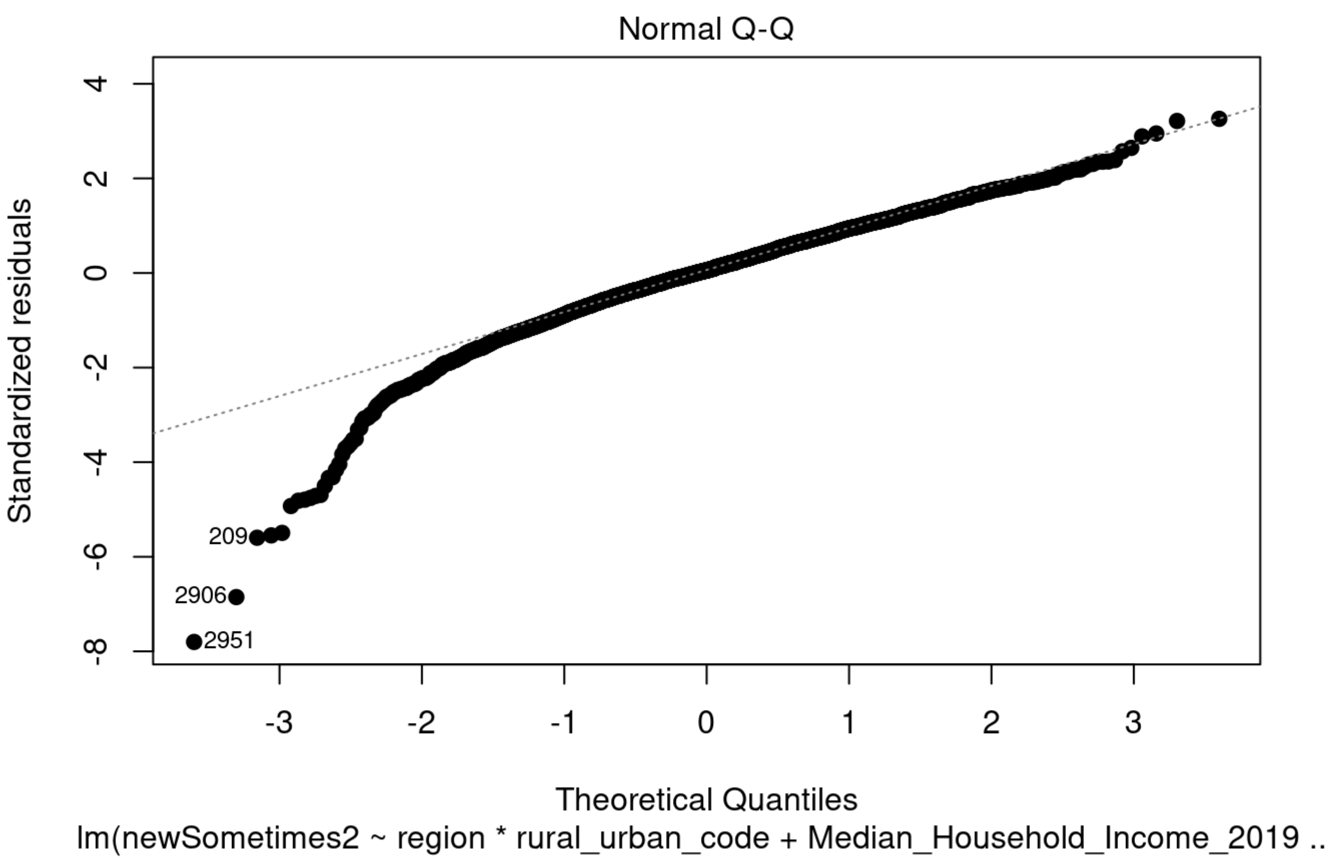
```
raw$newSometimes2 <- logit(raw$Sometimes_Wear_Mask_Survey)
raw$newAlways2 <- logit(raw$Always_Wear_Mask_Survey)

modnever3 <- lm(newNever2 ~ region*rural_urban_code + Median_Household_Income_2019
+ Percent_Adults_Bachelors_or_Higher, data = raw)
modsometimes3 <- lm(newSometimes2 ~ region*rural_urban_code + Median_Household_Inco
me_2019 + Percent_Adults_Bachelors_or_Higher, data = raw)
modalways3 <- lm(newAlways2 ~ region*rural_urban_code + Median_Household_Income_201
9 + Percent_Adults_Bachelors_or_Higher, data = raw)
plot(modnever3, which = c(1,2), pch=19)
```


Residuals vs Fitted
Residuals
Fitted values
lm(newNever2 ~ region * rural_urban_code + Median_Household_Income_2019 + P ...
2788
2790
1596
-3.5
-3.0
-2.5
-2.0
2
1
0
-1

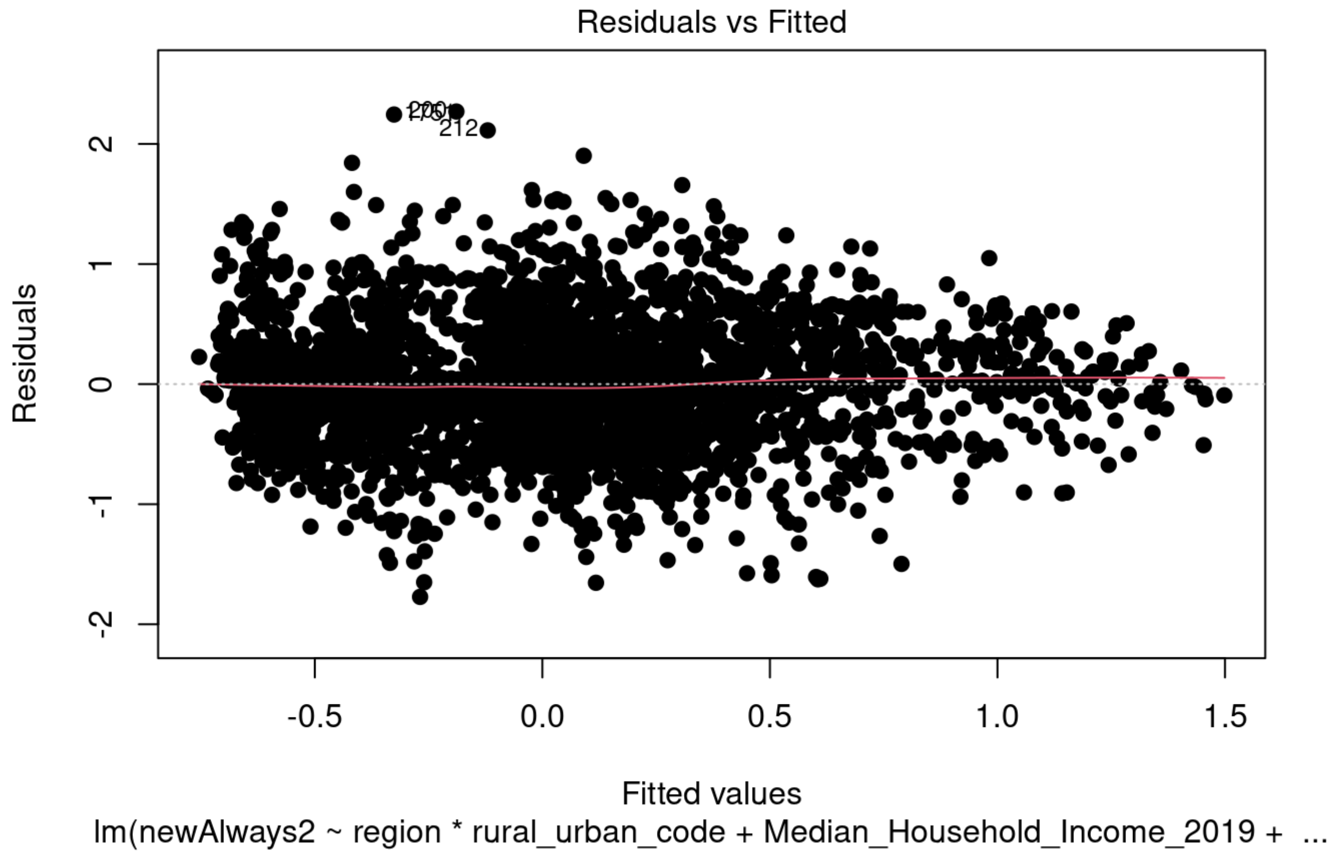

Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(newNever2 ~ region * rural_urban_code + Median_Household_Income_2019 + P ...
2790
2788
1596
-3
-2
-1
0
1
2
3
4
2
0
-2

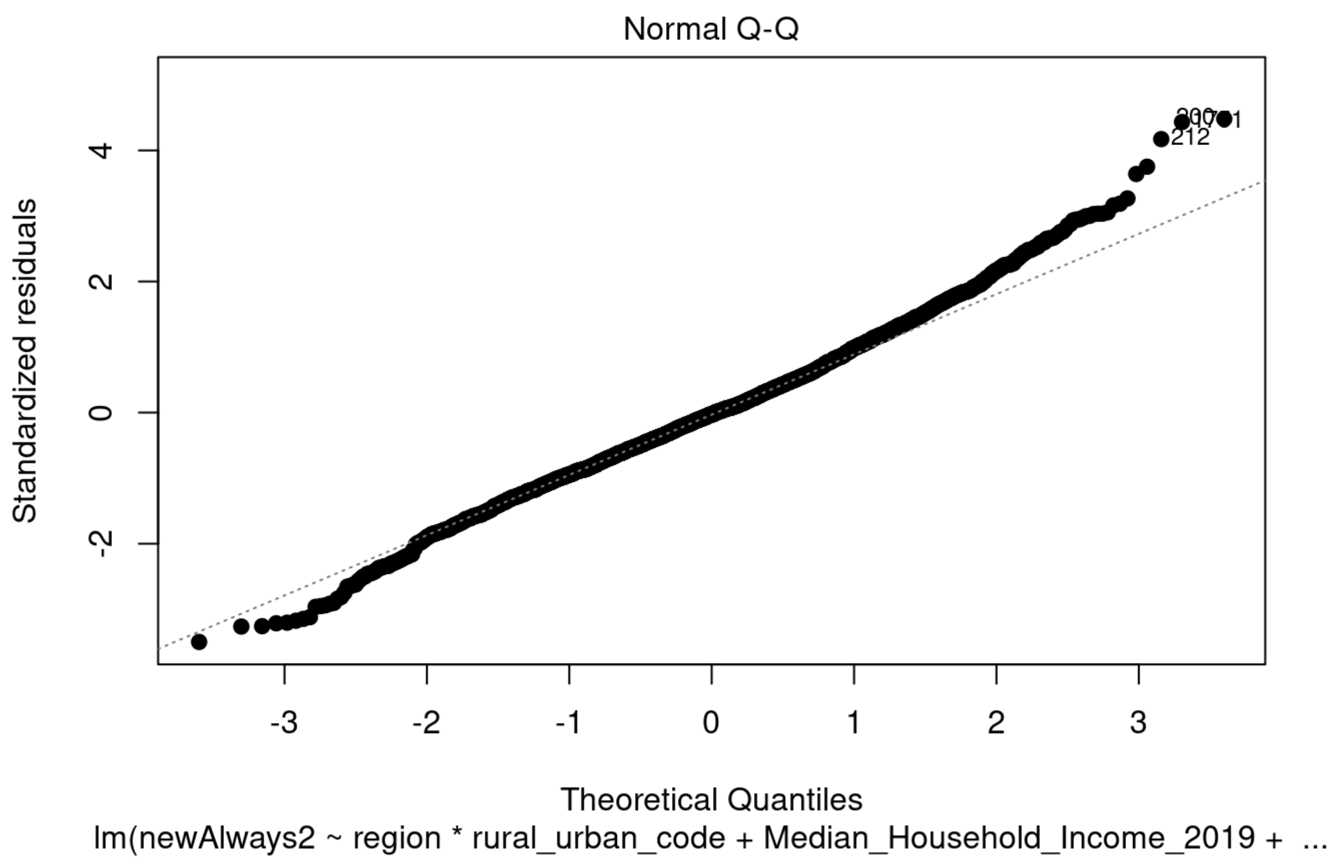
```
plot(modsometimes3, which = c(1,2), pch=19)
```


Residuals vs Fitted
Residuals
Fitted values
lm(newSometimes2 ~ region * rural_urban_code + Median_Household_Income_2019 ...
209
2906
2951


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(newSometimes2 ~ region * rural_urban_code + Median_Household_Income_2019 ...
209
2906
2951

```
plot(modalways3, which = c(1,2), pch=19)
```


Residuals vs Fitted
Residuals
Fitted values
lm(newAlways2 ~ region * rural_urban_code + Median_Household_Income_2019 + ...
212


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(newAlways2 ~ region * rural_urban_code + Median_Household_Income_2019 + ...
212

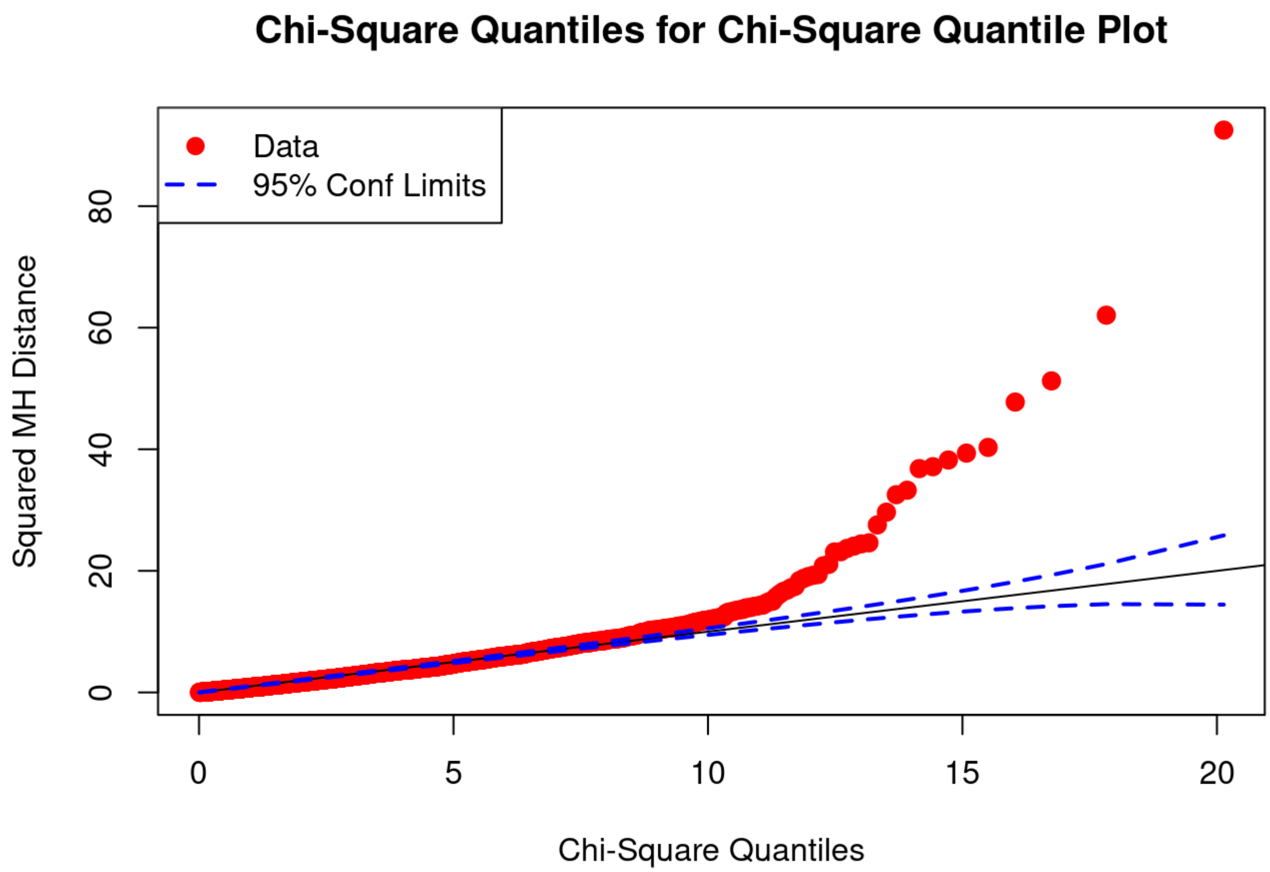
```
multimod5 <- lm(cbind(newNever2, newSometimes2, newAlways2) ~ region*rural_urban_co
de + Median_Household_Income_2019 + Percent_Adults_Bachelors_or_Higher, data = raw)

summary(Anova(multimod5, type = 3), univariate = T)
```

```
##
## Type III MANOVA Tests:
##
## Sum of squares and products for error:
##               newNever2 newSometimes2 newAlways2
## newNever2      868.5341      193.6919  -507.2176
## newSometimes2  193.6919      946.2551  -502.6998
## newAlways2    -507.2176     -502.6998   805.6647
##
## ------------------------------------------
##
## Term: (Intercept)
##
## Sum of squares and products for the hypothesis:
##               newNever2 newSometimes2 newAlways2
## newNever2     494.15809     462.96071  45.003127
## newSometimes2 462.96071     433.73290  42.161972
## newAlways2     45.00313      42.16197   4.098448
##
## Multivariate Tests: (Intercept)
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1 0.6919851 2340.204      3   3125 < 2.22e-16 ***
## Wilks             1 0.3080149 2340.204      3   3125 < 2.22e-16 ***
## Hotelling-Lawley  1 2.2465962 2340.204      3   3125 < 2.22e-16 ***
## Roy               1 2.2465962 2340.204      3   3125 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: region
##
## Sum of squares and products for the hypothesis:
##               newNever2 newSometimes2 newAlways2
## newNever2      107.4053      116.6820  -164.1904
## newSometimes2  116.6820      135.0029  -176.2496
## newAlways2    -164.1904     -176.2496   254.5546
##
## Multivariate Tests: region
##                  Df test stat approx F num Df   den Df     Pr(>F)
## Pillai            3 0.2672215 101.9233      9 9381.000 < 2.22e-16 ***
## Wilks             3 0.7390077 111.8227      9 7605.579 < 2.22e-16 ***
## Hotelling-Lawley  3 0.3447466 119.6526      9 9371.000 < 2.22e-16 ***
## Roy               3 0.3184367 331.9172      3 3127.000 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: rural_urban_code
##
## Sum of squares and products for the hypothesis:
##                newNever2 newSometimes2 newAlways2
```

```
## newNever2      19.177750      9.777144  -27.59785
## newSometimes2   9.777144      5.163270  -14.37589
## newAlways2    -27.597851    -14.375886   40.23889
##
## Multivariate Tests: rural_urban_code
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            2 0.0521158 27.87882      6   6252 < 2.22e-16 ***
## Wilks             2 0.9479077 28.23964      6   6250 < 2.22e-16 ***
## Hotelling-Lawley  2 0.0549303 28.60035      6   6248 < 2.22e-16 ***
## Roy               2 0.0544748 56.76275      3   3126 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Median_Household_Income_2019
##
## Sum of squares and products for the hypothesis:
##               newNever2 newSometimes2 newAlways2
## newNever2      5.816115      3.544362  -2.476064
## newSometimes2  3.544362      2.159948  -1.508923
## newAlways2    -2.476064     -1.508923   1.054122
##
## Multivariate Tests: Median_Household_Income_2019
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1 0.0093685 9.851159      3   3125 1.8281e-06 ***
## Wilks             1 0.9906315 9.851159      3   3125 1.8281e-06 ***
## Hotelling-Lawley  1 0.0094571 9.851159      3   3125 1.8281e-06 ***
## Roy               1 0.0094571 9.851159      3   3125 1.8281e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Percent_Adults_Bachelors_or_Higher
##
## Sum of squares and products for the hypothesis:
##               newNever2 newSometimes2 newAlways2
## newNever2      13.18602      15.75405  -16.40943
## newSometimes2  15.75405      18.82221  -19.60523
## newAlways2    -16.40943     -19.60523   20.42082
##
## Multivariate Tests: Percent_Adults_Bachelors_or_Higher
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1 0.0302729 32.51866      3   3125 < 2.22e-16 ***
## Wilks             1 0.9697271 32.51866      3   3125 < 2.22e-16 ***
## Hotelling-Lawley  1 0.0312179 32.51866      3   3125 < 2.22e-16 ***
## Roy               1 0.0312179 32.51866      3   3125 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: region:rural_urban_code
```

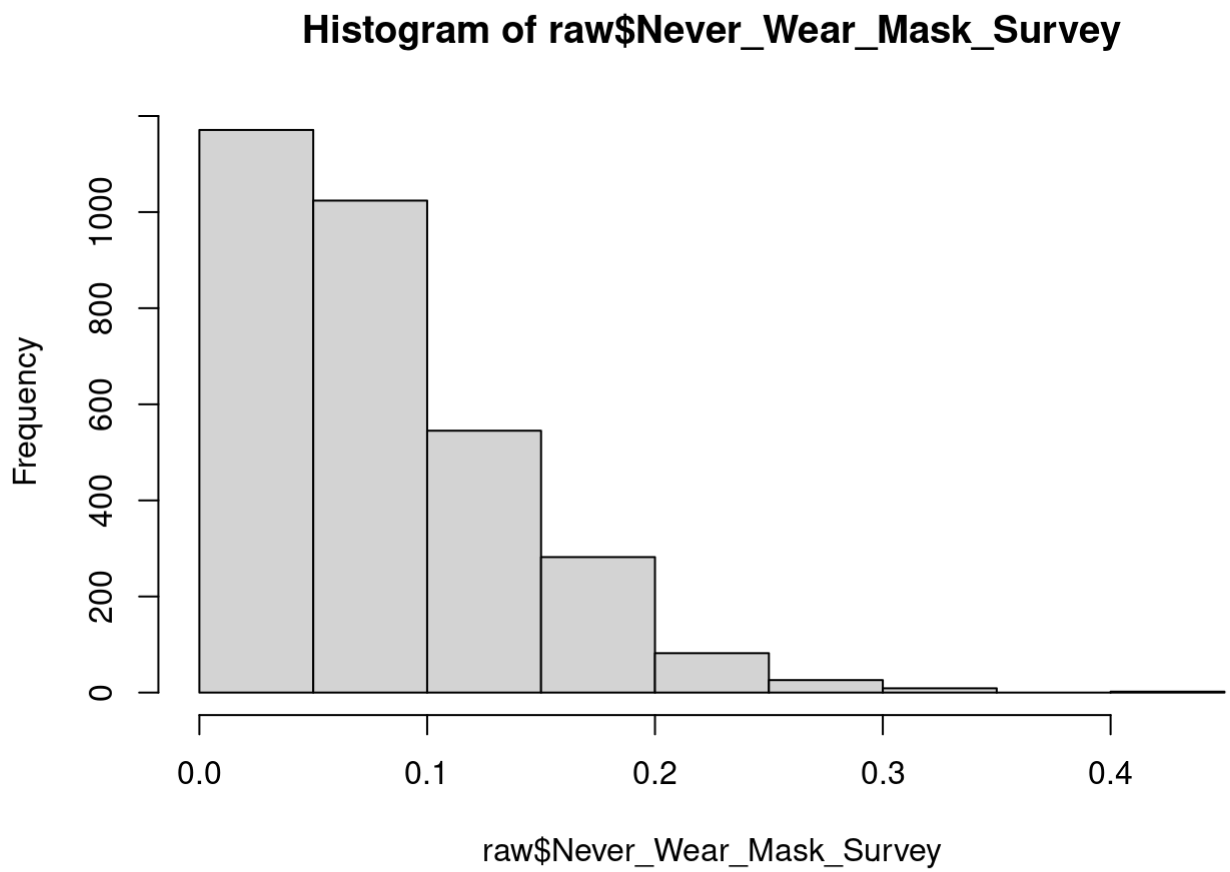
```
##
## Sum of squares and products for the hypothesis:
##              newNever2 newSometimes2 newAlways2
## newNever2      10.60991      3.979610 -11.162680
## newSometimes2   3.97961      3.349167  -6.862209
## newAlways2    -11.16268     -6.862209  19.707359
##
## Multivariate Tests: region:rural_urban_code
##                  Df test stat  approx F num Df  den Df     Pr(>F)
## Pillai            6 0.0363394  6.390371     18 9381.00 4.6931e-16 ***
## Wilks             6 0.9639232  6.421082     18 8839.32 3.7665e-16 ***
## Hotelling-Lawley  6 0.0371550  6.447768     18 9371.00 3.0204e-16 ***
## Roy               6 0.0277545 14.464728      6 3127.00 2.4089e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Type III Sums of Squares
##                                     df newNever2 newSometimes2 newAlways2
## (Intercept)                          1  494.1581      433.7329     4.0984
## region                               3  107.4053      135.0029   254.5546
## rural_urban_code                     2   19.1777        5.1633    40.2389
## Median_Household_Income_2019         1    5.8161        2.1599     1.0541
## Percent_Adults_Bachelors_or_Higher   1   13.1860       18.8222    20.4208
## region:rural_urban_code              6   10.6099        3.3492    19.7074
## residuals                         3127  868.5341      946.2551   805.6647
##
##  F-tests
##                                   newNever2 newSometimes2 newAlways2
## (Intercept)                         1779.13        477.77       7.95
## region                               386.69        446.13     164.67
## rural_urban_code                      69.05          5.69      78.09
## Median_Household_Income_2019          20.94          7.14       0.68
## Percent_Adults_Bachelors_or_Higher    47.47         20.73      39.63
## region:rural_urban_code               38.20         11.07      12.75
##
##  p-values
##                                   newNever2  newSometimes2 newAlways2
## (Intercept)                       < 2.22e-16 < 2.22e-16    0.00035856
## region                            < 2.22e-16 < 2.22e-16    < 2.22e-16
## rural_urban_code                  < 2.22e-16 0.00070044    < 2.22e-16
## Median_Household_Income_2019      4.9237e-06 0.00758681    0.66432668
## Percent_Adults_Bachelors_or_Higher 6.7127e-12 2.6761e-13    < 2.22e-16
## region:rural_urban_code           7.2152e-10 0.00088861    2.8645e-14
```

```
CSQPlot(multimod5$residuals)
```

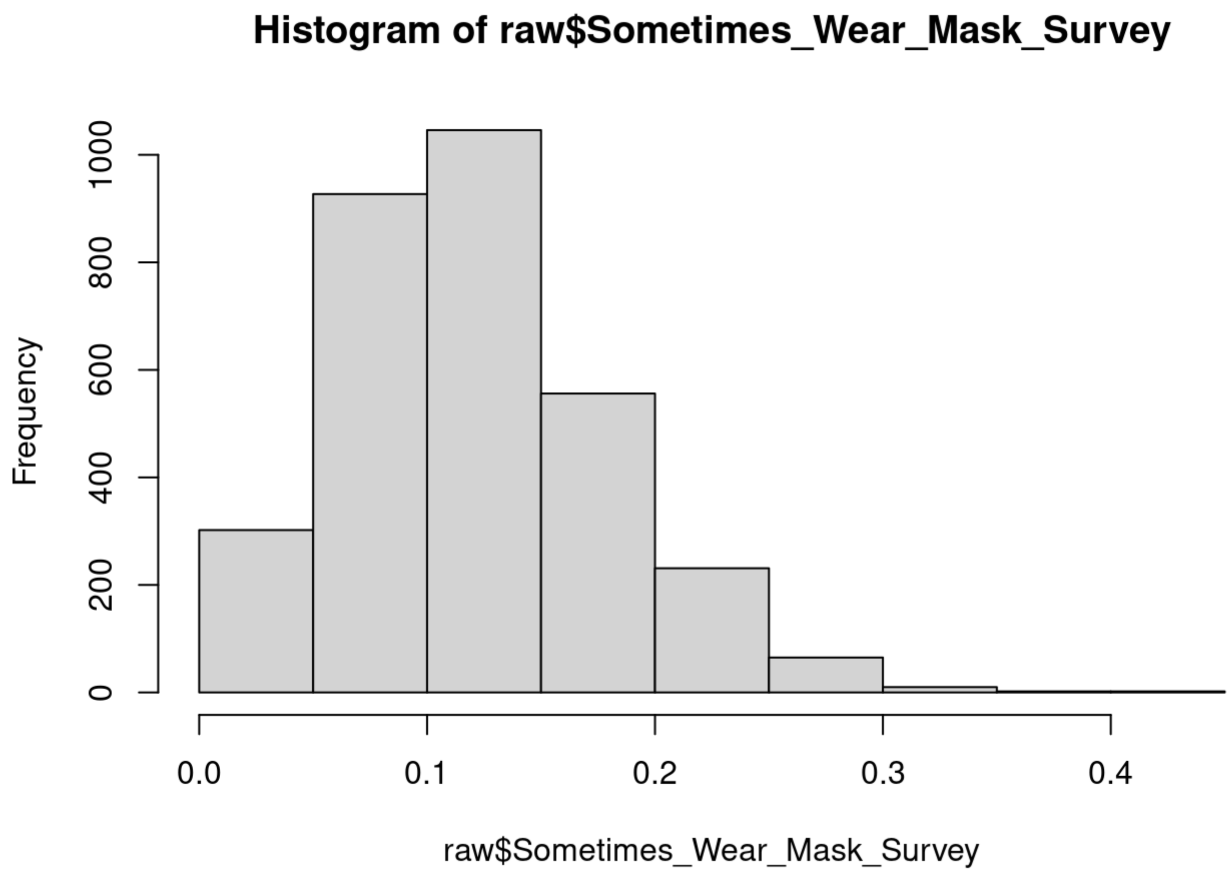


Not terribly happy with how that came out. Let's draw some histograms to see why.

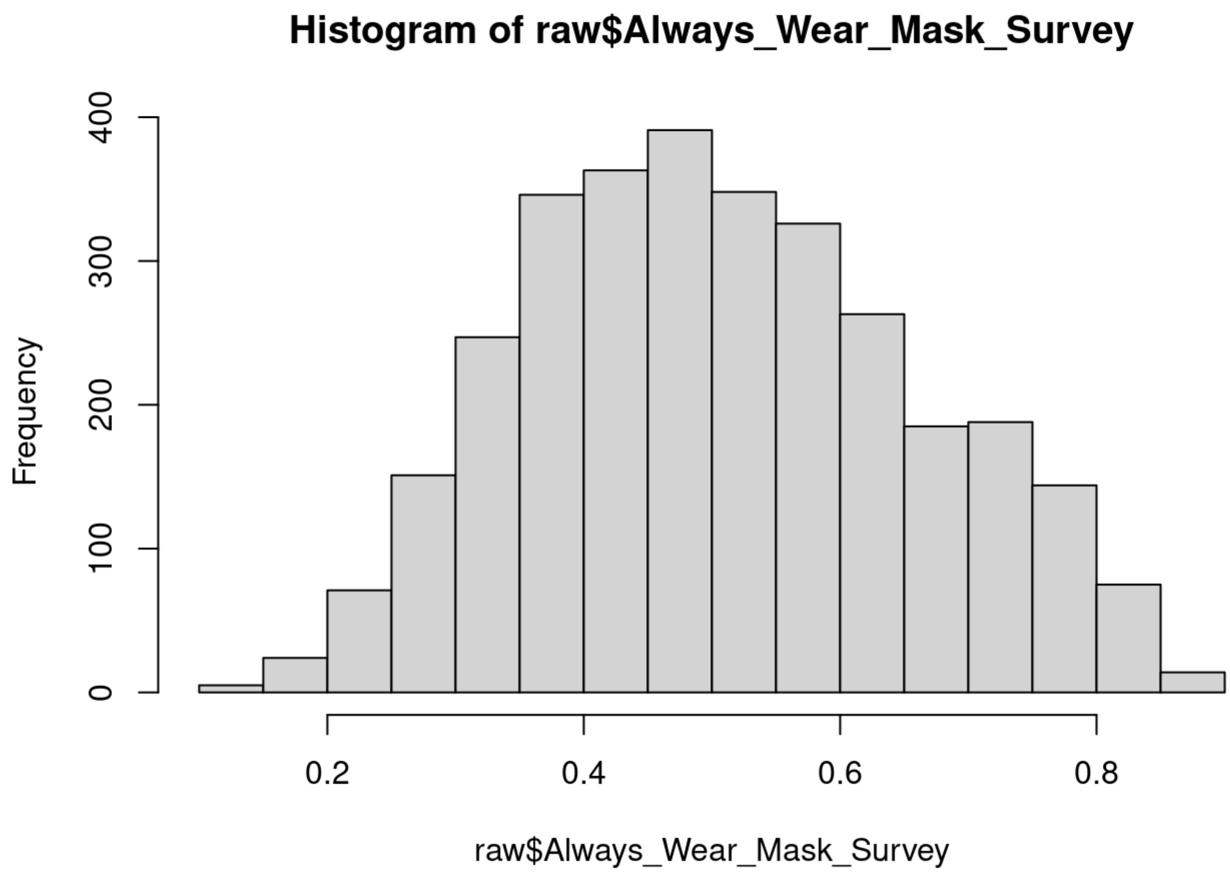
```
hist(raw$Never_Wear_Mask_Survey)
```




```
hist(raw$Sometimes_Wear_Mask_Survey)
```

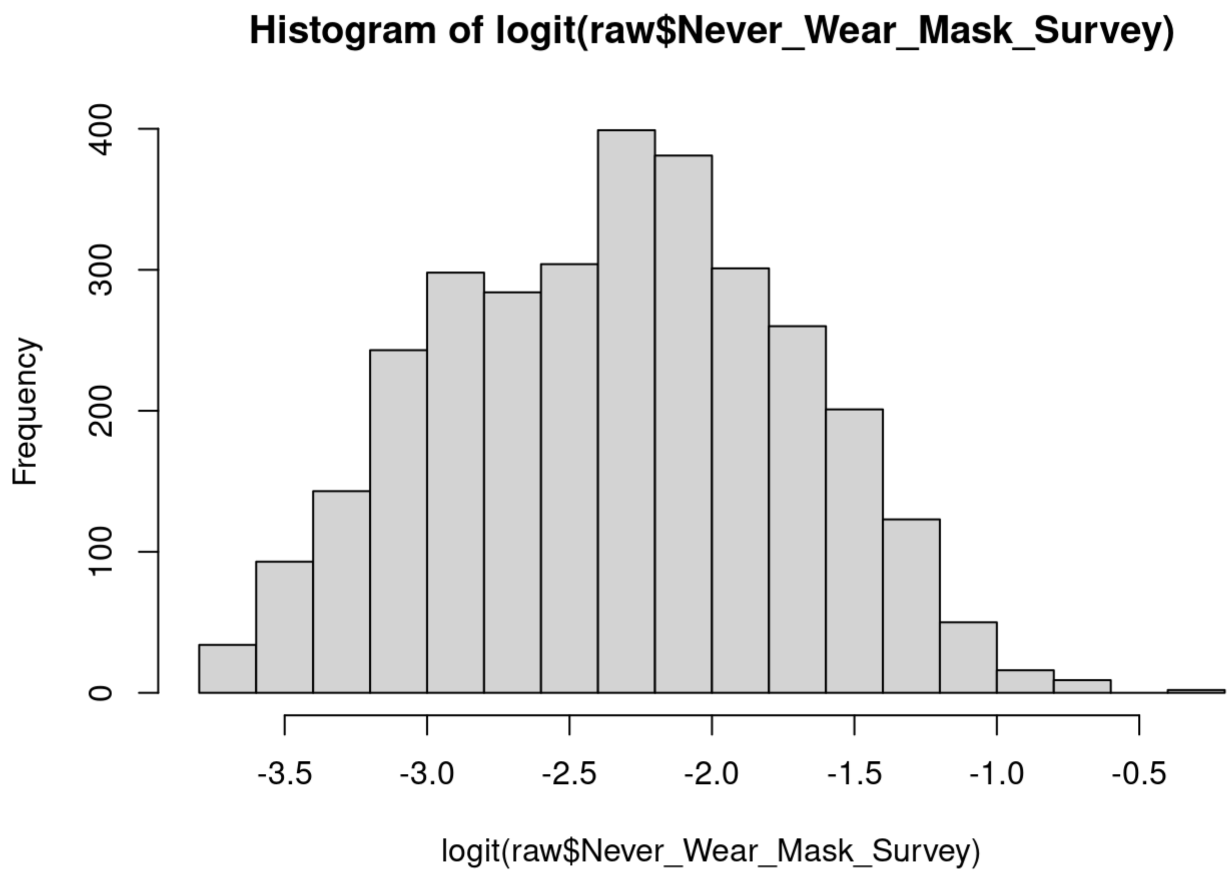



```
hist(raw$Always_Wear_Mask_Survey)
```

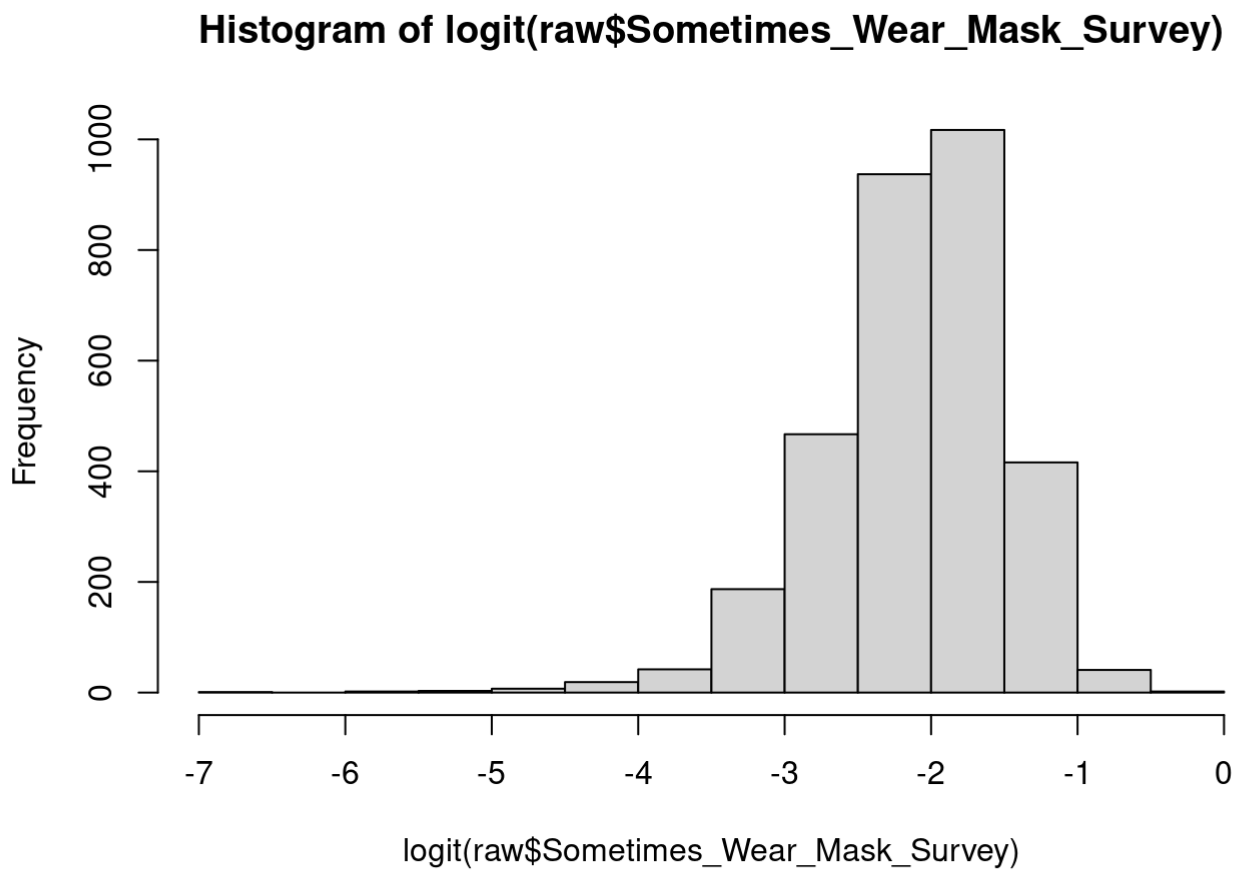



```
hist(logit(raw$Never_Wear_Mask_Survey))
```

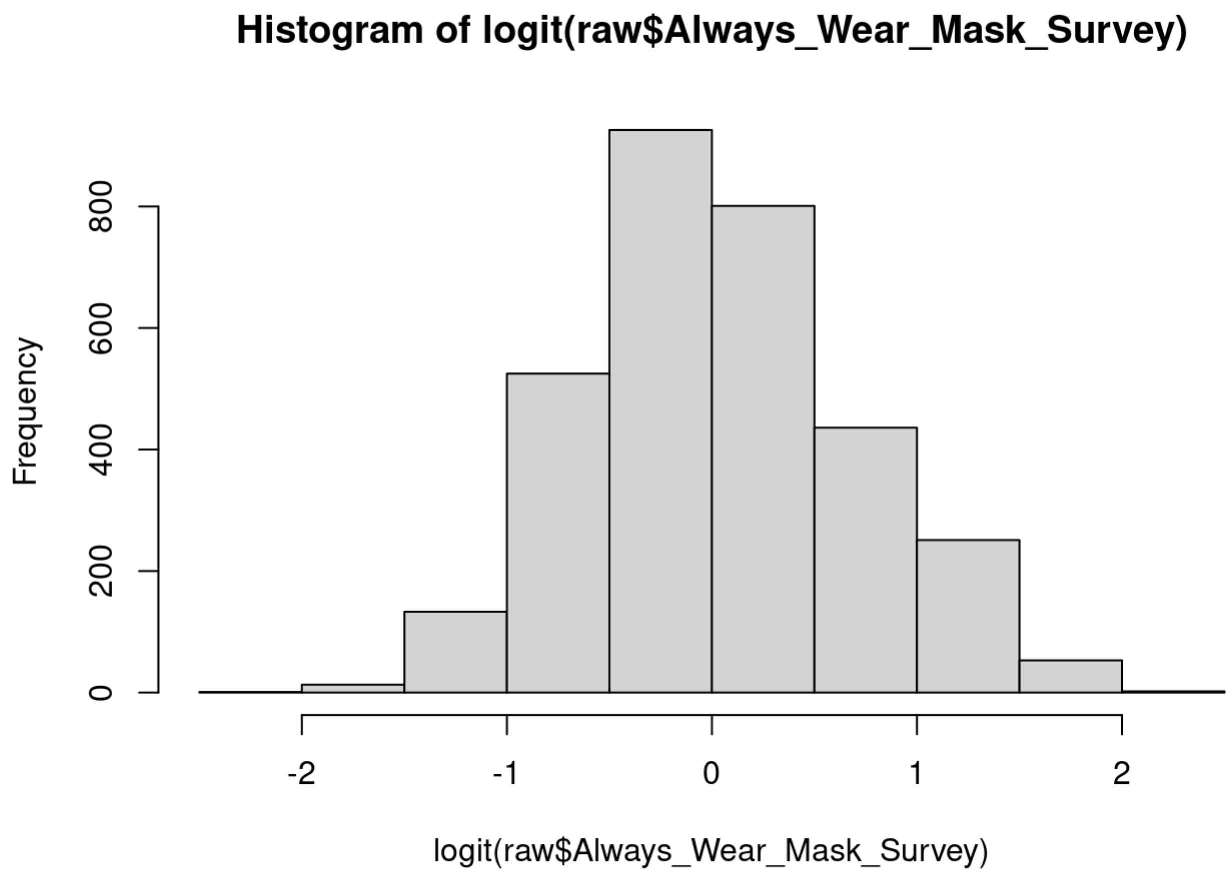
```
## Warning in logit(raw$Never_Wear_Mask_Survey): proportions remapped to (0.025,
## 0.975)
```




```
hist(logit(raw$Sometimes_Wear_Mask_Survey))
```




```
hist(logit(raw$Always_Wear_Mask_Survey))
```



That adds up. Partly literally – some of the issue is that the variables are necessarily interrelated and interdependent, in that the five mask survey response variables all fall on a single multidimensional curve by the nature of how they're determined. It's probably best to leave them untransformed and say that the data just have an unusual quality, per JDRS's recommendation, or alternatively to merge the variables into a single metric and add different response variables, which JDRS suggested but said is not necessarily warranted for the purposes of this pset. Before giving up on transformation, let's try applying logit to just the "never" variable, since it's the only one that is horribly skewed and butting up against 0.

```
multimod6 <- lm(cbind(newNever2, Sometimes_Wear_Mask_Survey, Always_Wear_Mask_Surve
y) ~ region*rural_urban_code + Median_Household_Income_2019 + Percent_Adults_Bachel
ors_or_Higher, data = raw)

summary(Anova(multimod6, type = 3), univariate = T)
```

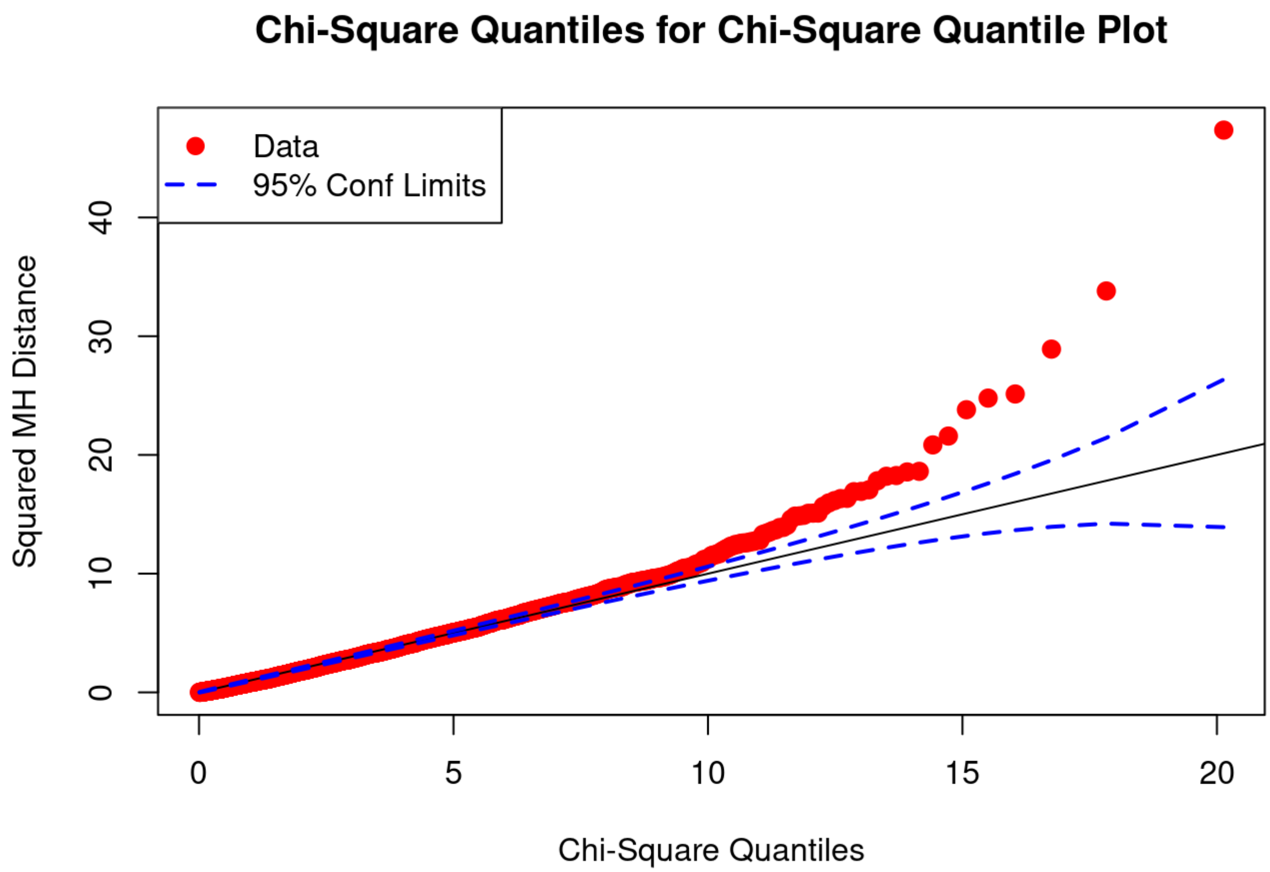
```
##
## Type III MANOVA Tests:
##
## Sum of squares and products for error:
##                             newNever2 Sometimes_Wear_Mask_Survey
## newNever2                   868.53413                  14.739272
## Sometimes_Wear_Mask_Survey   14.73927                   8.248288
## Always_Wear_Mask_Survey    -116.97881                 -10.429628
##                            Always_Wear_Mask_Survey
## newNever2                               -116.97881
## Sometimes_Wear_Mask_Survey               -10.42963
## Always_Wear_Mask_Survey                   41.97602
##
## ------------------------------------------
##
## Term: (Intercept)
##
## Sum of squares and products for the hypothesis:
##                             newNever2 Sometimes_Wear_Mask_Survey
## newNever2                   494.15809                 -35.388618
## Sometimes_Wear_Mask_Survey  -35.38862                   2.534319
## Always_Wear_Mask_Survey    -114.74916                   8.217641
##                            Always_Wear_Mask_Survey
## newNever2                              -114.749156
## Sometimes_Wear_Mask_Survey                8.217641
## Always_Wear_Mask_Survey                  26.646065
##
## Multivariate Tests: (Intercept)
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            1 0.6770143  2183.45      3   3125 < 2.22e-16 ***
## Wilks             1 0.3229857  2183.45      3   3125 < 2.22e-16 ***
## Hotelling-Lawley  1 2.0961119  2183.45      3   3125 < 2.22e-16 ***
## Roy               1 2.0961119  2183.45      3   3125 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: region
##
## Sum of squares and products for the hypothesis:
##                             newNever2 Sometimes_Wear_Mask_Survey
## newNever2                  107.405301                  9.9375884
## Sometimes_Wear_Mask_Survey   9.937588                  0.9500496
## Always_Wear_Mask_Survey    -37.830707                 -3.4579297
##                            Always_Wear_Mask_Survey
## newNever2                                -37.83071
## Sometimes_Wear_Mask_Survey                -3.45793
## Always_Wear_Mask_Survey                   13.49477
##
## Multivariate Tests: region
##                  Df test stat approx F num Df   den Df     Pr(>F)
## Pillai            3 0.2603331  99.0463      9 9381.000 < 2.22e-16 ***
```

```
## Wilks             3 0.7437988 109.2853      9 7605.579 < 2.22e-16 ***
## Hotelling-Lawley  3 0.3388975 117.6225      9 9371.000 < 2.22e-16 ***
## Roy               3 0.3216764 335.2940      3 3127.000 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: rural_urban_code
##
## Sum of squares and products for the hypothesis:
##                           newNever2 Sometimes_Wear_Mask_Survey
## newNever2                 19.177750                 1.11352371
## Sometimes_Wear_Mask_Survey  1.113524                 0.07673975
## Always_Wear_Mask_Survey    -6.190180                -0.37800526
##                           Always_Wear_Mask_Survey
## newNever2                              -6.1901803
## Sometimes_Wear_Mask_Survey             -0.3780053
## Always_Wear_Mask_Survey                 2.0266368
##
## Multivariate Tests: rural_urban_code
##                  Df test stat approx F num Df den Df     Pr(>F)
## Pillai            2 0.0486506 25.97893      6   6252 < 2.22e-16 ***
## Wilks             2 0.9514094 26.26892      6   6250 < 2.22e-16 ***
## Hotelling-Lawley  2 0.0510092 26.55880      6   6248 < 2.22e-16 ***
## Roy               2 0.0497413 51.83043      3   3126 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: Median_Household_Income_2019
##
## Sum of squares and products for the hypothesis:
##                            newNever2 Sometimes_Wear_Mask_Survey
## newNever2                  5.8161145                0.181370216
## Sometimes_Wear_Mask_Survey  0.1813702                0.005655864
## Always_Wear_Mask_Survey    -0.5546565               -0.017296456
##                           Always_Wear_Mask_Survey
## newNever2                             -0.55465650
## Sometimes_Wear_Mask_Survey            -0.01729646
## Always_Wear_Mask_Survey                0.05289508
##
## Multivariate Tests: Median_Household_Income_2019
##                  Df test stat approx F num Df den Df    Pr(>F)
## Pillai            1 0.0077744 8.161807      3   3125 2.067e-05 ***
## Wilks             1 0.9922256 8.161807      3   3125 2.067e-05 ***
## Hotelling-Lawley  1 0.0078353 8.161807      3   3125 2.067e-05 ***
## Roy               1 0.0078353 8.161807      3   3125 2.067e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
```

```
## Term: Percent_Adults_Bachelors_or_Higher
##
## Sum of squares and products for the hypothesis:
##                           newNever2 Sometimes_Wear_Mask_Survey
## newNever2                 13.186018                  1.4196237
## Sometimes_Wear_Mask_Survey  1.419624                  0.1528385
## Always_Wear_Mask_Survey    -3.763751                 -0.4052103
##                           Always_Wear_Mask_Survey
## newNever2                              -3.7637514
## Sometimes_Wear_Mask_Survey             -0.4052103
## Always_Wear_Mask_Survey                 1.0743065
##
## Multivariate Tests: Percent_Adults_Bachelors_or_Higher
##                  Df test stat  approx F num Df den Df     Pr(>F)
## Pillai            1 0.0301116  32.3401      3   3125 < 2.22e-16 ***
## Wilks             1 0.9698884  32.3401      3   3125 < 2.22e-16 ***
## Hotelling-Lawley  1 0.0310465  32.3401      3   3125 < 2.22e-16 ***
## Roy               1 0.0310465  32.3401      3   3125 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ------------------------------------------
##
## Term: region:rural_urban_code
##
## Sum of squares and products for the hypothesis:
##                            newNever2 Sometimes_Wear_Mask_Survey
## newNever2                 10.6099103                  0.4970945
## Sometimes_Wear_Mask_Survey  0.4970945                  0.0342260
## Always_Wear_Mask_Survey    -2.5930134                 -0.1553487
##                           Always_Wear_Mask_Survey
## newNever2                              -2.5930134
## Sometimes_Wear_Mask_Survey             -0.1553487
## Always_Wear_Mask_Survey                 1.0127658
##
## Multivariate Tests: region:rural_urban_code
##                  Df test stat  approx F num Df  den Df     Pr(>F)
## Pillai            6 0.0362071  6.366827     18 9381.00 5.6228e-16 ***
## Wilks             6 0.9640663  6.394968     18 8839.32 4.6025e-16 ***
## Hotelling-Lawley  6 0.0369897  6.419087     18 9371.00 3.7652e-16 ***
## Roy               6 0.0267923 13.963246      6 3127.00 9.7438e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Type III Sums of Squares
##                                      df newNever2 Sometimes_Wear_Mask_Survey
## (Intercept)                           1  494.1581                  2.5343190
## region                                3  107.4053                  0.9500496
## rural_urban_code                      2   19.1777                  0.0767398
## Median_Household_Income_2019          1    5.8161                  0.0056559
## Percent_Adults_Bachelors_or_Higher    1   13.1860                  0.1528385
## region:rural_urban_code               6   10.6099                  0.0342260
## residuals                          3127  868.5341                  8.2482885
##                                    Always_Wear_Mask_Survey
```

```
## (Intercept)                                    26.646065
## region                                         13.494766
## rural_urban_code                                2.026637
## Median_Household_Income_2019                    0.052895
## Percent_Adults_Bachelors_or_Higher              1.074307
## region:rural_urban_code                         1.012766
## residuals                                      41.976018
##
##  F-tests
##                                    newNever2 Sometimes_Wear_Mask_Survey
## (Intercept)                          1779.13                     320.26
## region                                386.69                     360.17
## rural_urban_code                       69.05                       9.70
## Median_Household_Income_2019           20.94                       2.14
## Percent_Adults_Bachelors_or_Higher     47.47                      19.31
## region:rural_urban_code                38.20                      12.98
##                                    Always_Wear_Mask_Survey
## (Intercept)                                          992.50
## region                                               167.55
## rural_urban_code                                      75.49
## Median_Household_Income_2019                           0.66
## Percent_Adults_Bachelors_or_Higher                    40.02
## region:rural_urban_code                               12.57
##
##  p-values
##                                    newNever2  Sometimes_Wear_Mask_Survey
## (Intercept)                        < 2.22e-16 < 2.22e-16
## region                             < 2.22e-16 < 2.22e-16
## rural_urban_code                   < 2.22e-16 2.2800e-06
## Median_Household_Income_2019       4.9237e-06 0.14321118
## Percent_Adults_Bachelors_or_Higher 6.7127e-12 2.0891e-12
## region:rural_urban_code            7.2152e-10 0.00032053
##                                    Always_Wear_Mask_Survey
## (Intercept)                        < 2.22e-16
## region                             < 2.22e-16
## rural_urban_code                   < 2.22e-16
## Median_Household_Income_2019       0.68473479
## Percent_Adults_Bachelors_or_Higher < 2.22e-16
## region:rural_urban_code            4.6446e-14
```

```
CSQPlot(multimod6$residuals)
```



Our best options appear to be either applying a logit transformation to the never-wears-mask values, or to just not transform at all. Given the nature of the data, no transformation is probably the way to go. At least, that's what JDRS advised, and it makes sense.

6 MRPP Test

Our data set is not Overwhelmingly Large™, so there is a reasonable expectation that MRPP will be able to provide a satisfactorily reliable p-value. Let's give it a go.

```
(mrppout <- mrpp(raw[,c("Never_Wear_Mask_Survey", "Sometimes_Wear_Mask_Survey", "Al
ways_Wear_Mask_Survey")], raw$rural_urban_code))
```

Now, RStudio's knitting function seems to have some difficulty processing MRPP. That's not much of a surprise; it takes my computer long enough to run as is, and we've had our share of issues with things running fine but not knitting properly. In any case, we had to disable evaluation of the function in order to knit, and the would-be output is copied below.

```
Dissimilarity index: euclidean
Weights for groups:  n

Class means and counts:

      Rural  Suburban Urban
delta 0.2015 0.1966   0.1838
n     1077   899      1165

Chance corrected within-group agreement A: 0.07795
Based on observed delta 0.1935 and expected delta 0.2099

Significance of delta: 0.001
Permutation: free
Number of permutations: 999
```

And those results look pretty good! The p-value is 0.001 according to the above output, so MRPP tells us that the multivariate means of the groups are significantly different.