# Section 5: Outpatient Visits Model

We built a similar model to the emergency department visits model to model the relationship between outpatient visits per 1,000 beneficiaries and Medicaid Expansion status. We added potential covariates one by one and evaluated the fit using likelihood ratio tests, with the intention of getting the best fit possible with this data. We did not identify any of these covariates as potentially confounding the relationship between the outpatient visit rate and Medicaid expansion status.

Check distribution of outcome

```
medicare %>%
  ggplot(aes(x= op_visits_per_1000_beneficiaries)) +
  geom_histogram(color = "black") +
  scale_x_continuous(labels = scales::comma) +
  scale_y_continuous(labels = scales::comma) +
  xlab("Outpatient Visits per 1,000 Beneficiaries") +
  ylab("Count") +
  ggtitle("Distribution of Outpatient Visit Rates (2007-2018)") +
  theme_bw()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 21 rows containing non-finite values (stat_bin).
```

Distribution of Outpatient Visit Rates (2007–2018)
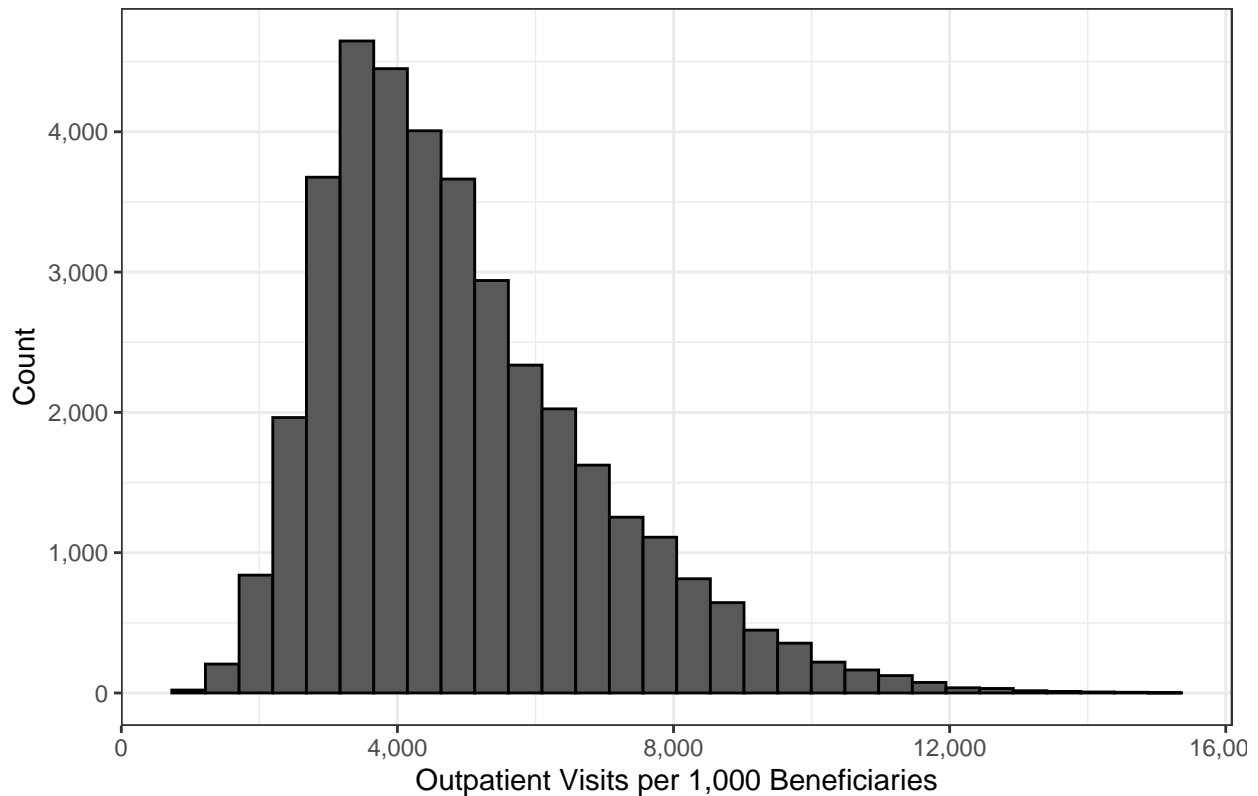
The outcome is pretty right skewed.

We next check missingness in the outcome

```r
medicare %>%
  select(op_visits_per_1000_beneficiaries) %>%
  summarize_all(.funs = function(x)
    return(paste0(round(sum(is.na(x))/length(x), 3)*100, "%"))) %>%
  t() %>% data.frame()
```

```
##                                       .
## op_visits_per_1000_beneficiaries 0.1%
```

Model Fitting

First we fit the base model:

```r
mod.nb_op<- MASS::glm.nb( op_visits_per_1000_beneficiaries ~ percent_eligible_for_medicaid +   offset(log

summary(mod.nb_op)
```

```
##
## Call:
## MASS::glm.nb(formula = op_visits_per_1000_beneficiaries ~ percent_eligible_for_medicaid +
##     offset(log(year)), data = medicare, init.theta = 6.954913505,
##     link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.6566  -0.8550  -0.2011   0.5338   3.7530
```

```
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 1.021573   0.005111   199.9   <2e-16 ***
## percent_eligible_for_medicaid -0.604220   0.021816   -27.7   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(6.9549) family taken to be 1)
##
##     Null deviance: 39206  on 37543  degrees of freedom
## Residual deviance: 38439  on 37542  degrees of freedom
##   (192 observations deleted due to missingness)
## AIC: 668152
##
## Number of Fisher Scoring iterations: 1
##
##
##              Theta:  6.9549
##          Std. Err.:  0.0497
##
##  2 x log-likelihood:  -668145.7500
```

```r
# Model diagnostics table
model_diagnostics <- data.frame(
  Model = "OP Visits ~ Exp",
  AIC = AIC(mod.nb_op),
  BIC = BIC(mod.nb_op)
)
```

Improve Model Fit Using Other Covariates   The covariates percent_male, percent_eligible_for_medicaid, average_age, do not satisfy the criteria for confounding the relationship between Medicaid Expansion status and Outpatient Visit rates, so we will add each of these covariates one by one to determine if they improve the model fit using Likelihood Ratio tests.

Percent Male

```r
mod.nb_op_c1 <- MASS::glm.nb(op_visits_per_1000_beneficiaries ~ expansion_status + percent_male + offset
summary(mod.nb_op_c1)
```

```
##
## Call:
## MASS::glm.nb(formula = op_visits_per_1000_beneficiaries ~ expansion_status +
##     percent_male + offset(log(year)), data = medicare, init.theta = 7.194866745,
##     link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.4204  -0.8343  -0.2310   0.5041   3.9482
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       1.351387   0.036210  37.321   < 2e-16 ***
## expansion_status2 -0.025227   0.008222  -3.068  0.00215 **
## expansion_status3  0.089442   0.009134   9.793   < 2e-16 ***
```

```
## expansion_status4 -0.149473   0.007876 -18.977  < 2e-16 ***
## percent_male       -0.848699   0.076371 -11.113  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(7.1949) family taken to be 1)
##
##       Null deviance: 40547  on 37662  degrees of freedom
## Residual deviance: 38531  on 37658  degrees of freedom
##    (73 observations deleted due to missingness)
## AIC: 668943
##
## Number of Fisher Scoring iterations: 1
##
##
##              Theta:  7.1949
##          Std. Err.:  0.0513
##
##  2 x log-likelihood:  -668930.8930
```

```r
# Use Likelihood Ratio test to determine if confounder is necessary
anova(mod.nb_op, mod.nb_op_c1) # Yes
```

```
## Likelihood ratio tests of Negative Binomial Models
##
## Response: op_visits_per_1000_beneficiaries
##                                                  Model    theta Resid. df
## 1 expansion_status + percent_male + offset(log(year)) 7.194867     37658
## 2   percent_eligible_for_medicaid + offset(log(year)) 6.954914     37542
##      2 x log-lik.   Test    df LR stat. Pr(Chi)
## 1        -668930.9
## 2        -668145.7 1 vs 2   116 785.1427       0
```

```r
# # Model diagnostics table
model_diagnostics <- model_diagnostics %>%
  bind_rows(
    data.frame(
    Model = "OP Visits ~ Exp + Male",
    AIC = AIC(mod.nb_op_c1),
    BIC = BIC(mod.nb_op_c1)
    )
  )
```

By the LRT, we add percent male into the model.

Percent Eligible for Medicaid

```r
mod.nb_op_c2 <- MASS::glm.nb(op_visits_per_1000_beneficiaries ~ expansion_status + percent_male + percen
summary(mod.nb_op_c2)
```

```
##
## Call:
## MASS::glm.nb(formula = op_visits_per_1000_beneficiaries ~ expansion_status +
##     percent_male + percent_eligible_for_medicaid + offset(log(year)),
##     data = medicare, init.theta = 7.338953384, link = log)
##
## Deviance Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -3.5564  -0.8251  -0.2149   0.5061   3.8062
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   1.334902   0.036280  36.794  < 2e-16 ***
## expansion_status2            -0.012869   0.008148  -1.579    0.114
## expansion_status3             0.101837   0.009052  11.250  < 2e-16 ***
## expansion_status4            -0.135903   0.007812 -17.397  < 2e-16 ***
## percent_male                 -0.558468   0.076757  -7.276 3.44e-13 ***
## percent_eligible_for_medicaid -0.598658   0.021379 -28.002  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(7.339) family taken to be 1)
##
##     Null deviance: 41162  on 37503  degrees of freedom
## Residual deviance: 38351  on 37498  degrees of freedom
##   (232 observations deleted due to missingness)
## AIC: 665407
##
## Number of Fisher Scoring iterations: 1
##
##
##              Theta:  7.3390
##          Std. Err.:  0.0525
##
##  2 x log-likelihood:  -665392.5350
# Use Anova to determine if confounder is necessary
anova(mod.nb_op_c1, mod.nb_op_c2) # Yes

## Likelihood ratio tests of Negative Binomial Models
##
## Response: op_visits_per_1000_beneficiaries
##                                                                         Model
## 1                            expansion_status + percent_male + offset(log(year))
## 2 expansion_status + percent_male + percent_eligible_for_medicaid + offset(log(year))
##      theta Resid. df   2 x log-lik.   Test    df LR stat. Pr(Chi)
## 1 7.194867     37658     -668930.9
## 2 7.338953     37498     -665392.5 1 vs 2   160 3538.357       0
# # Model diagnostics table
model_diagnostics <- model_diagnostics %>%
  bind_rows(
    data.frame(
    Model = "OP Visits ~ Exp + Male + % Elig",
    AIC = AIC(mod.nb_op_c2),
    BIC = BIC(mod.nb_op_c2)
    )
  )
```

By the LRT, we add percent eligible for medicaid to the model.

Average Age

```
mod.nb_op_c3 <- MASS::glm.nb(op_visits_per_1000_beneficiaries ~ expansion_status + percent_male + percen
summary(mod.nb_op_c3)
```

```
##
## Call:
## MASS::glm.nb(formula = op_visits_per_1000_beneficiaries ~ expansion_status +
##     percent_male + percent_eligible_for_medicaid + average_age +
##     offset(log(year)), data = medicare, init.theta = 8.140083142,
##     link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.0958  -0.7759  -0.1720   0.5095   4.0356
##
## Coefficients:
##                                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)                    -5.992178   0.115512 -51.875   <2e-16 ***
## expansion_status2               0.008968   0.007749   1.157    0.247
## expansion_status3               0.128680   0.008605  14.955   <2e-16 ***
## expansion_status4              -0.118146   0.007428 -15.905   <2e-16 ***
## percent_male                    1.235517   0.077243  15.995   <2e-16 ***
## percent_eligible_for_medicaid   0.722657   0.027552  26.229   <2e-16 ***
## average_age                     0.086813   0.001320  65.790   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(8.1401) family taken to be 1)
##
##     Null deviance: 45647  on 37503  degrees of freedom
## Residual deviance: 38268  on 37497  degrees of freedom
##   (232 observations deleted due to missingness)
## AIC: 661363
##
## Number of Fisher Scoring iterations: 1
##
##
##              Theta:  8.1401
##          Std. Err.:  0.0584
##
##  2 x log-likelihood:  -661347.0920
```

```
# Use Likelihood ratio test to determine if covariate is necessary
anova(mod.nb_op_c2, mod.nb_op_c3) # Yes
```

```
## Likelihood ratio tests of Negative Binomial Models
##
## Response: op_visits_per_1000_beneficiaries
##                                                                              Model
## 1                 expansion_status + percent_male + percent_eligible_for_medicaid + offset(log(year))
## 2 expansion_status + percent_male + percent_eligible_for_medicaid + average_age + offset(log(year))
##      theta Resid. df   2 x log-lik.    Test    df LR stat. Pr(Chi)
## 1 7.338953     37498        -665392.5
## 2 8.140083     37497        -661347.1 1 vs 2     1 4045.443       0
```

```r
# # Model diagnostics table
model_diagnostics <- model_diagnostics %>%
  bind_rows(
    data.frame(
    Model = "OP Visits ~ Exp + Male + % Elig + Age",
    AIC = AIC(mod.nb_op_c3),
    BIC = BIC(mod.nb_op_c3)
    )
  )
```

By the LRT, we add average age to the model.

```r
mod.nb_op_1 <- MASS::glm.nb(op_visits_per_1000_beneficiaries ~ expansion_status + percent_male + percen

summary(mod.nb_op_1)
```

Check for quadratic effects of `percent_eligible_for_medicaid`

```
##
## Call:
## MASS::glm.nb(formula = op_visits_per_1000_beneficiaries ~ expansion_status +
##     percent_male + percent_eligible_for_medicaid + average_age +
##     percent_eligible_for_medicaid_2 + offset(log(year)), data = medicare %>%
##     mutate(percent_eligible_for_medicaid_2 = percent_eligible_for_medicaid^2),
##     init.theta = 8.157521428, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.9647  -0.7789  -0.1763   0.5090   3.9912
##
## Coefficients:
##                                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)                     -6.405498   0.124798 -51.327   <2e-16 ***
## expansion_status2                0.010467   0.007742   1.352    0.176
## expansion_status3                0.129356   0.008596  15.049   <2e-16 ***
## expansion_status4               -0.114854   0.007427 -15.465   <2e-16 ***
## percent_male                     1.367662   0.078633  17.393   <2e-16 ***
## percent_eligible_for_medicaid    1.528583   0.089633  17.054   <2e-16 ***
## average_age                      0.090389   0.001385  65.285   <2e-16 ***
## percent_eligible_for_medicaid_2 -1.442403   0.151279  -9.535   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(8.1575) family taken to be 1)
##
##     Null deviance: 45745  on 37503  degrees of freedom
## Residual deviance: 38266  on 37496  degrees of freedom
##   (232 observations deleted due to missingness)
## AIC: 661282
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  8.1575
```

```
##          Std. Err.:   0.0585
##
##  2 x log-likelihood:  -661263.7960
```

```
anova(mod.nb_op_c3, mod.nb_op_1) # Yes
```

```
## Likelihood ratio tests of Negative Binomial Models
##
## Response: op_visits_per_1000_beneficiaries
##
## 1                                  expansion_status + percent_male + percent_eligible_for_medicaid
## 2 expansion_status + percent_male + percent_eligible_for_medicaid + average_age + percent_eligible_f
##       theta Resid. df    2 x log-lik.   Test    df LR stat. Pr(Chi)
## 1 8.140083     37497        -661347.1
## 2 8.157521     37496        -661263.8 1 vs 2     1 83.29644       0
```

```
model_diagnostics <- model_diagnostics %>%
  bind_rows(
    data.frame(
    Model = "OP Visits ~ Exp + Male + % Elig + Age + % Elig^2",
    AIC = AIC(mod.nb_op_1),
    BIC = BIC(mod.nb_op_1)
    )
  )
```

Effect Modification   We think there may be effect modification between % eligible for medicaid and
expansion status, since in general Medicaid expansion means a higher percentage of the population qualifies
for Medicaid.

```
mod.nb_op_2 <- MASS::glm.nb(op_visits_per_1000_beneficiaries ~ expansion_status + percent_male + percen
```

```
summary(mod.nb_op_2)
```

```
##
## Call:
## MASS::glm.nb(formula = op_visits_per_1000_beneficiaries ~ expansion_status +
##     percent_male + percent_eligible_for_medicaid + average_age +
##     percent_eligible_for_medicaid_2 + percent_eligible_for_medicaid *
##     expansion_status + offset(log(year)), data = medicare %>%
##     mutate(percent_eligible_for_medicaid_2 = percent_eligible_for_medicaid^2),
##     init.theta = 8.209500433, link = log)
##
## Deviance Residuals:
##     Min      1Q   Median       3Q      Max
## -3.9103  -0.7766  -0.1715   0.5051   4.0269
##
## Coefficients:
##                                     Estimate Std. Error z value
## (Intercept)                        -6.312747   0.124998 -50.503
## expansion_status2                  -0.264026   0.021031 -12.554
## expansion_status3                  -0.059563   0.023529  -2.531
## expansion_status4                  -0.264591   0.020027 -13.212
## percent_male                        1.306633   0.078867  16.568
## percent_eligible_for_medicaid       0.667399   0.118588   5.628
## average_age                         0.091921   0.001389  66.189
## percent_eligible_for_medicaid_2    -1.443392   0.151651  -9.518
```

```
## expansion_status2:percent_eligible_for_medicaid  1.324995    0.094930   13.958
## expansion_status3:percent_eligible_for_medicaid  0.932099    0.105085    8.870
## expansion_status4:percent_eligible_for_medicaid  0.748069    0.090782    8.240
##                                                  Pr(>|z|)
## (Intercept)                                       < 2e-16 ***
## expansion_status2                                 < 2e-16 ***
## expansion_status3                                  0.0114 *
## expansion_status4                                 < 2e-16 ***
## percent_male                                      < 2e-16 ***
## percent_eligible_for_medicaid                    1.82e-08 ***
## average_age                                       < 2e-16 ***
## percent_eligible_for_medicaid_2                   < 2e-16 ***
## expansion_status2:percent_eligible_for_medicaid  < 2e-16 ***
## expansion_status3:percent_eligible_for_medicaid  < 2e-16 ***
## expansion_status4:percent_eligible_for_medicaid  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(8.2095) family taken to be 1)
##
##     Null deviance: 46036  on 37503  degrees of freedom
## Residual deviance: 38262  on 37493  degrees of freedom
##   (232 observations deleted due to missingness)
## AIC: 661040
##
## Number of Fisher Scoring iterations: 1
##
##
##              Theta:  8.2095
##          Std. Err.:  0.0589
##
##  2 x log-likelihood:  -661016.4490
```

```r
anova(mod.nb_op_1, mod.nb_op_2)
```

```
## Likelihood ratio tests of Negative Binomial Models
##
## Response: op_visits_per_1000_beneficiaries
##
## 1                                                 expansion_status + percent_male + percent_eligi
## 2 expansion_status + percent_male + percent_eligible_for_medicaid + average_age + percent_eligible_f
##      theta Resid. df    2 x log-lik.   Test    df LR stat. Pr(Chi)
## 1 8.157521     37496       -661263.8
## 2 8.209500     37493       -661016.4 1 vs 2     3 247.3464       0
```

```r
model_diagnostics <- model_diagnostics %>%
  bind_rows(
    data.frame(
    Model = "OP Visits ~ Exp + Male + % Elig + Age + % Elig^2 + % Elig*Exp",
    AIC = AIC(mod.nb_op_2),
    BIC = BIC(mod.nb_op_2)
    )
  )
```

By the LRT, we do add the interaction between Medicaid expansion status and percent_eligible_for_medicaid.

```
model_diagnostics
```

Model Diagnostics

```
##                                                                 Model      AIC
## 1                                                     OP Visits ~ Exp 668151.7
## 2                                              OP Visits ~ Exp + Male 668942.9
## 3                                     OP Visits ~ Exp + Male + % Elig 665406.5
## 4                               OP Visits ~ Exp + Male + % Elig + Age 661363.1
## 5                    OP Visits ~ Exp + Male + % Elig + Age + % Elig^2 661281.8
## 6 OP Visits ~ Exp + Male + % Elig + Age + % Elig^2 + % Elig*Exp 661040.4
##        BIC
## 1 668177.3
## 2 668994.1
## 3 665466.3
## 4 661431.3
## 5 661358.6
## 6 661142.8
```

We have the smallest AIC and BIC with the full model.

```
mod.pois_ed <- glm(op_visits_per_1000_beneficiaries ~ expansion_status + percent_male + percent_eligibl

# Evaluate overdispersion
deviance(mod.pois_ed)/mod.pois_ed$df.residual # 604.9794
```

Check Poisson model

```
## [1] 604.9794
```

```
pearson.stat_ed <- sum((na.omit(medicare %>% select(percent_male, percent_eligible_for_medicaid, averag
pearson.stat_ed/mod.pois_ed$df.residual #  633.6993
```

```
## [1] 633.6993
```

```
# Evidence of overdispersion
mean(medicare$op_visits_per_1000_beneficiaries, na.rm = T)
```

```
## [1] 4907.535
```

```
var(medicare$op_visits_per_1000_beneficiaries, na.rm = T)
```

```
## [1] 3821163
```

Due to the evidence of dispersion, the negative binomial model seems to be the appropriate appropriate model.

```
exp(coef(mod.nb_op_2))
```

Interpretation

```
##                               (Intercept)
##                               0.001813046
##                          expansion_status2
##                               0.767953921
##                          expansion_status3
```

```
##                                                 0.942176216
##                              expansion_status4
##                                                 0.767519980
##                                    percent_male
##                                                 3.693716819
##                     percent_eligible_for_medicaid
##                                                 1.949161834
##                                     average_age
##                                                 1.096278527
##                  percent_eligible_for_medicaid_2
##                                                 0.236125358
## expansion_status2:percent_eligible_for_medicaid
##                                                 3.762167259
## expansion_status3:percent_eligible_for_medicaid
##                                                 2.539833966
## expansion_status4:percent_eligible_for_medicaid
##                                                 2.112915583
```

```r
exp(confint(mod.nb_op_2))
```

```
## Waiting for profiling to be done...
##                                                   2.5 %       97.5 %
## (Intercept)                                  0.001412489 0.002327059
## expansion_status2                            0.736181320 0.801089333
## expansion_status3                            0.898595842 0.987874404
## expansion_status4                            0.737227774 0.799052378
## percent_male                                 3.144681343 4.339131603
## percent_eligible_for_medicaid                1.531841237 2.480052929
## average_age                                  1.093283405 1.099282332
## percent_eligible_for_medicaid_2              0.173750454 0.321258592
## expansion_status2:percent_eligible_for_medicaid 3.105943746 4.553878822
## expansion_status3:percent_eligible_for_medicaid 2.053248823 3.140143153
## expansion_status4:percent_eligible_for_medicaid 1.758023257 2.537349885
```