Regression

Justin Coles,Jessy Garcia, Joseph Salamante

2025-06-30

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.4.3
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.4.1
```

```
##
## Attaching package: 'dplyr'
```

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

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

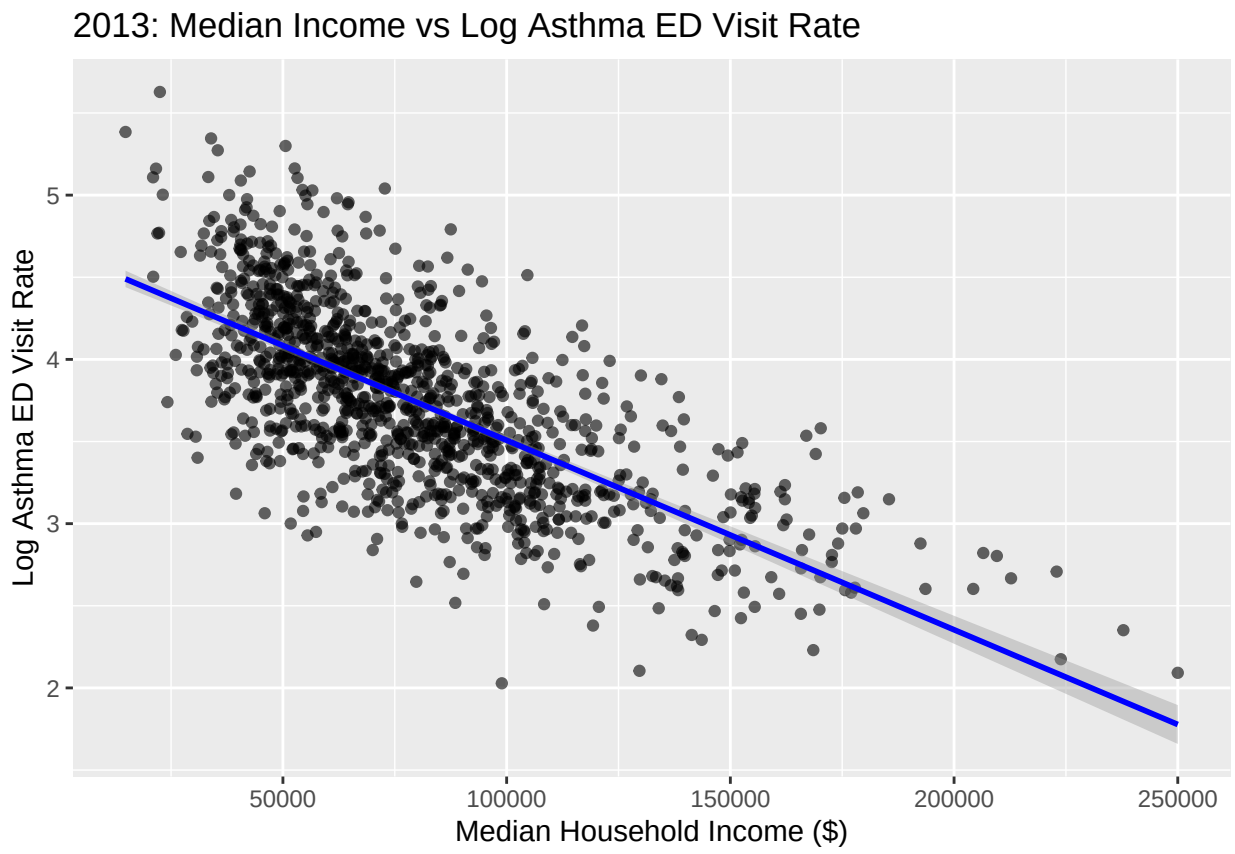
```
getwd()
```

```
## [1] "C:/Users/LAs Best/Desktop/EV-Proj_LAsBest"
```

```
EV_data <- read.csv("./Data/data_ZEV_asthmaED_2013_2022.csv")
# Income vs. Log Asthma Rate Plot
# For 2013
EV_data$nZEV1000pop <- EV_data$nZEV/EV_data$pop *1000
EV_data$log_AgeAdj_RoA_ED_Visit_Rate <- log(EV_data$Age_Adjusted_Rate_of_Asthma_ED_Visit_Rate)
EV_data_2013 <- EV_data %>% filter( yr == 2013)
EV_data_2022 <- EV_data %>% filter( yr == 2022)
```

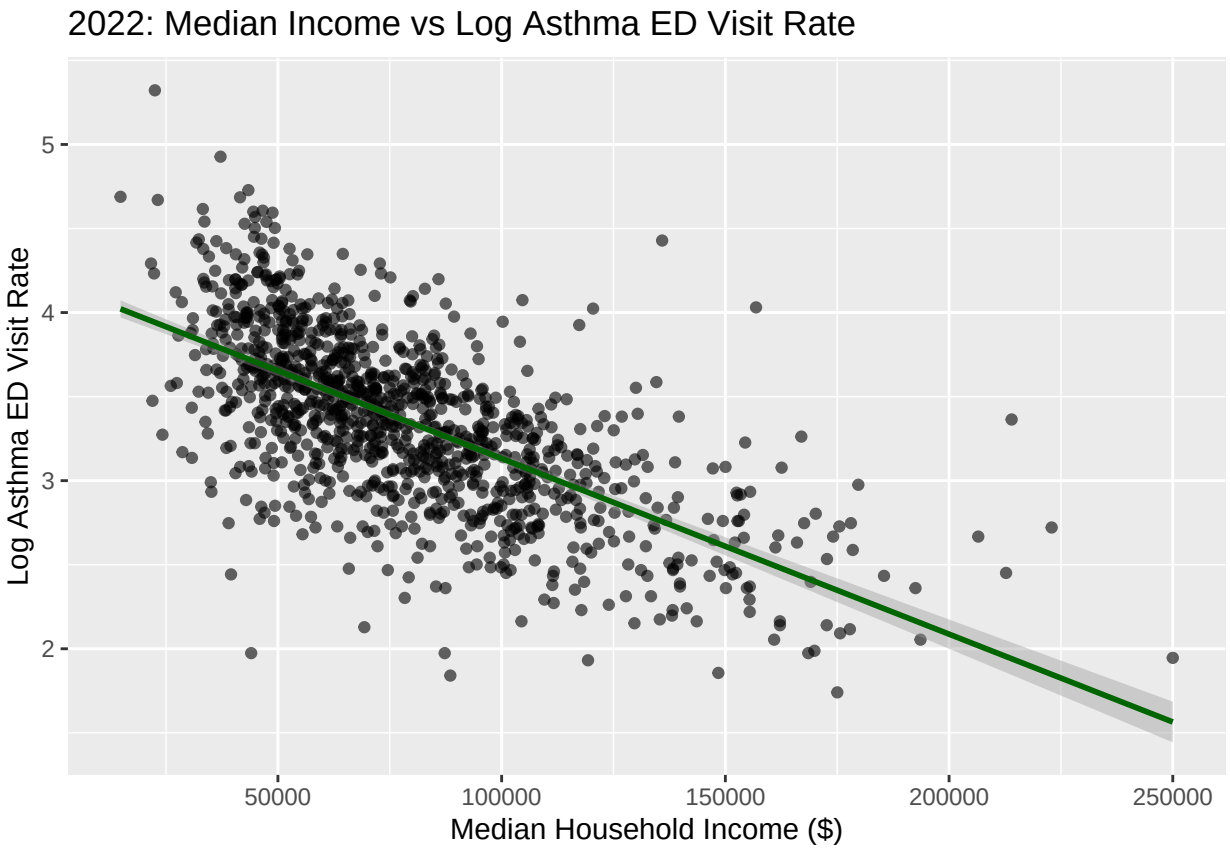
```
ggplot(EV_data_2013, aes(x = HHincomeMedian, y = log_AgeAdj_RoA_ED_Visit_Rate)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = TRUE, color = "blue") +
  labs(
    title = "2013: Median Income vs Log Asthma ED Visit Rate",
    x = "Median Household Income ($)",
    y = "Log Asthma ED Visit Rate"
  )
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
# For 2022
ggplot(EV_data_2022, aes(x = HHincomeMedian, y = log_AgeAdj_RoA_ED_Visit_Rate)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = TRUE, color = "darkgreen") +
  labs(
    title = "2022: Median Income vs Log Asthma ED Visit Rate",
    x = "Median Household Income ($)",
    y = "Log Asthma ED Visit Rate"
  )
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
lm(EV_data_2013$log_AgeAdj_RoA_ED_Visit_Rate ~ EV_data_2013$percPoverty    , data = EV_data)
```

```
##
## Call:
## lm(formula = EV_data_2013$log_AgeAdj_RoA_ED_Visit_Rate ~ EV_data_2013$percPoverty,
##     data = EV_data)
##
## Coefficients:
##               (Intercept)  EV_data_2013$percPoverty
##                   3.19704                   0.03981
```

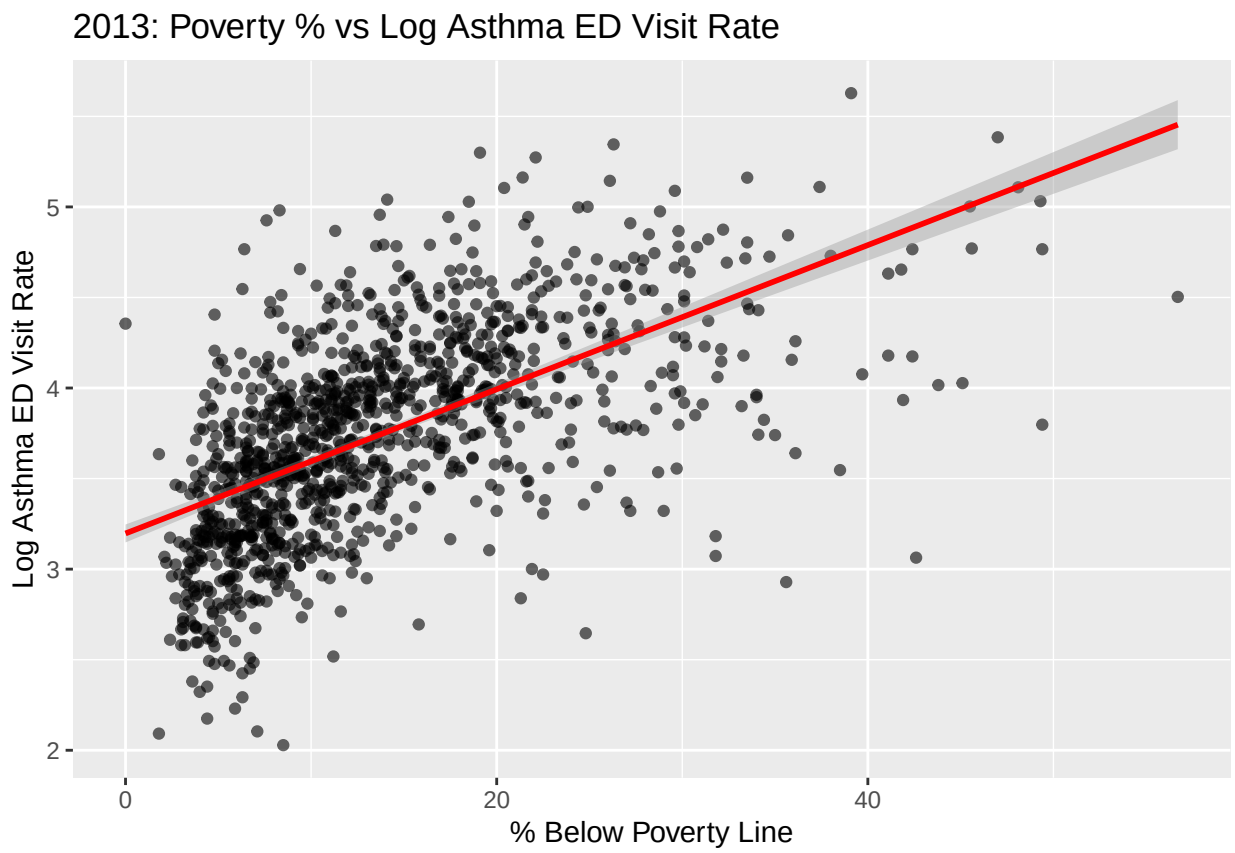
```
lm(EV_data_2022$log_AgeAdj_RoA_ED_Visit_Rate ~ EV_data_2022$percPoverty, data = EV_data)
```

```
##
## Call:
## lm(formula = EV_data_2022$log_AgeAdj_RoA_ED_Visit_Rate ~ EV_data_2022$percPoverty,
##     data = EV_data)
##
## Coefficients:
##               (Intercept)  EV_data_2022$percPoverty
##                   2.87082                   0.03478
```

```
#Poverty Percentage vs Log Asthma Rate
# 2013
```

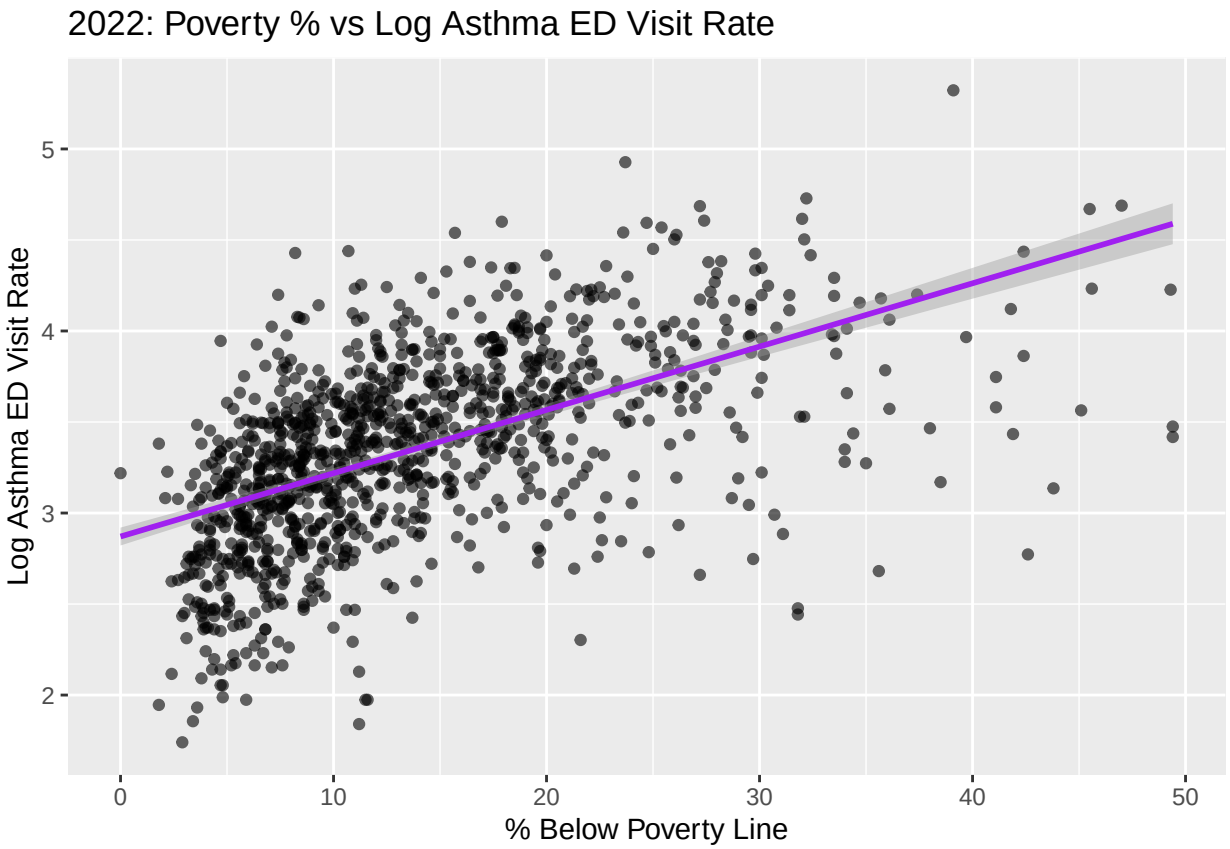
```
ggplot(EV_data_2013, aes(x = percPoverty, y = log_AgeAdj_RoA_ED_Visit_Rate)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", color = "red") +
  labs(
    title = "2013: Poverty % vs Log Asthma ED Visit Rate",
    x = "% Below Poverty Line",
    y = "Log Asthma ED Visit Rate"
  )
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
# 2022
ggplot(EV_data_2022, aes(x = percPoverty, y = log_AgeAdj_RoA_ED_Visit_Rate)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", color = "purple") +
  labs(
    title = "2022: Poverty % vs Log Asthma ED Visit Rate",
    x = "% Below Poverty Line",
    y = "Log Asthma ED Visit Rate"
  )
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
EV_data_2013 <- EV_data %>%
  filter( yr == 2013)
lm(EV_data_2013$log_AgeAdj_RoA_ED_Visit_Rate ~ EV_data_2013$EDUCpercBAplus    , data = EV_data)
```

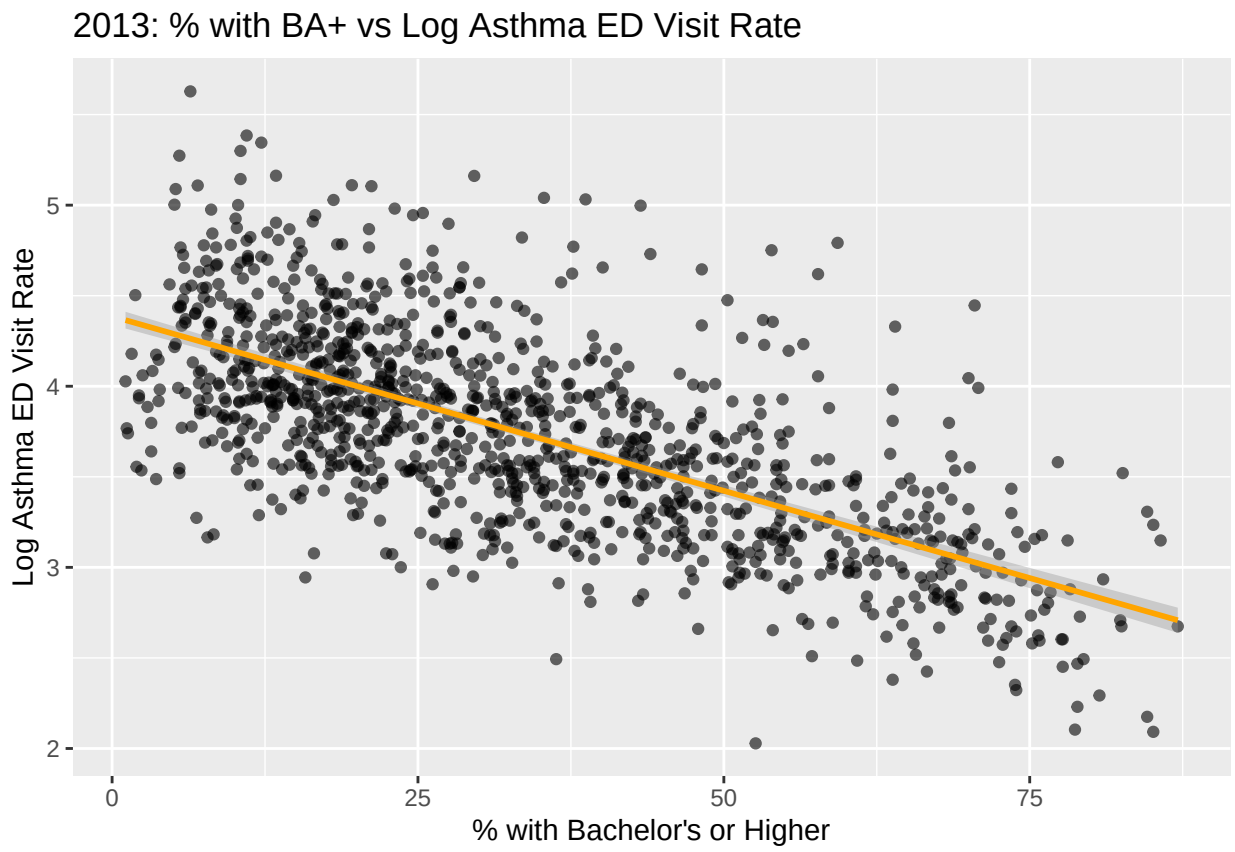
```
##
## Call:
## lm(formula = EV_data_2013$log_AgeAdj_RoA_ED_Visit_Rate ~ EV_data_2013$EDUCpercBAplus,
##     data = EV_data)
##
## Coefficients:
##                  (Intercept)  EV_data_2013$EDUCpercBAplus
##                      4.38614                      -0.01927
```

```
EV_data_2022 <- EV_data %>%
  filter( yr == 2022)
lm(EV_data_2022$log_AgeAdj_RoA_ED_Visit_Rate ~ EV_data_2022$EDUCpercBAplus, data = EV_data)
```

```
##
## Call:
## lm(formula = EV_data_2022$log_AgeAdj_RoA_ED_Visit_Rate ~ EV_data_2022$EDUCpercBAplus,
##     data = EV_data)
##
## Coefficients:
##                  (Intercept)  EV_data_2022$EDUCpercBAplus
##                      3.97003                      -0.01872
```

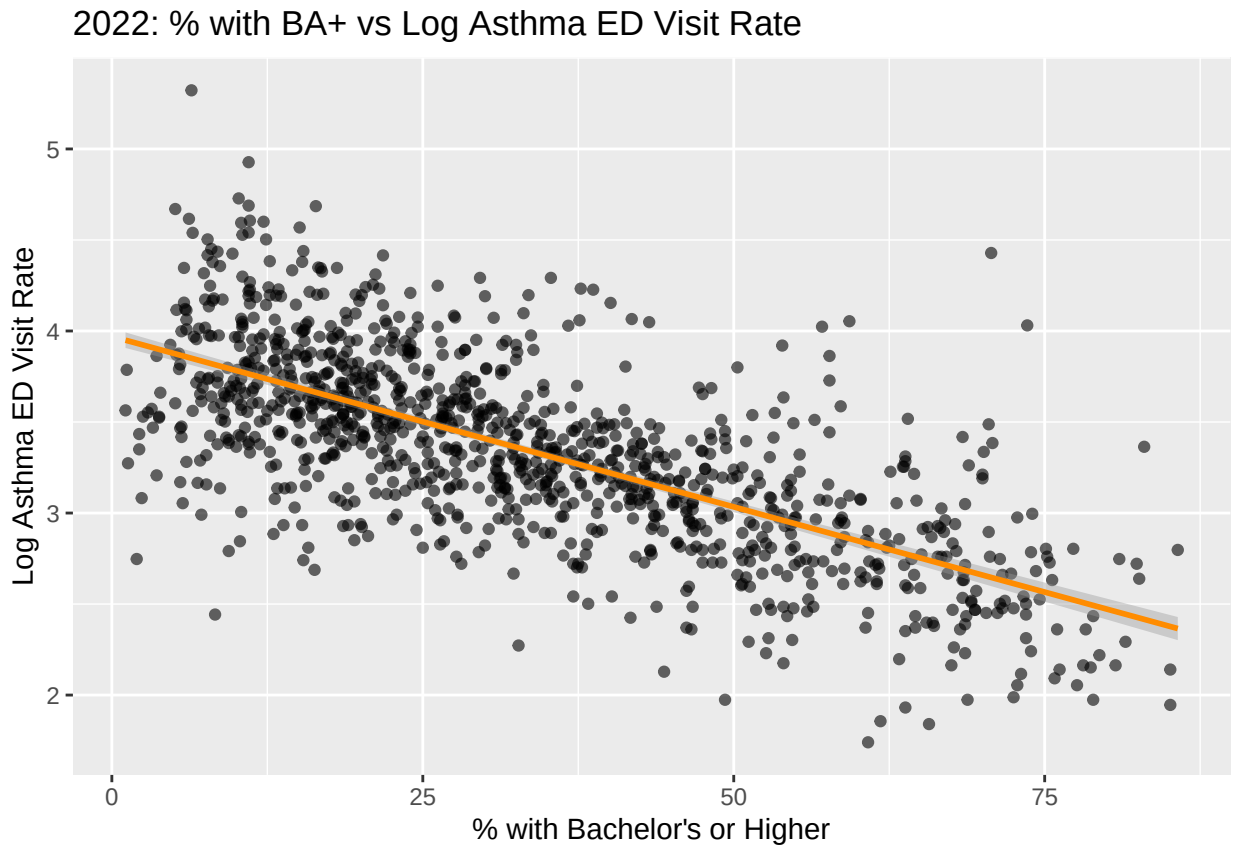
```r
#BA Education vs Log Asthma Rate
# 2013
ggplot(EV_data_2013, aes(x = EDUCpercBAplus, y = log_AgeAdj_RoA_ED_Visit_Rate)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", color = "orange") +
  labs(
    title = "2013: % with BA+ vs Log Asthma ED Visit Rate",
    x = "% with Bachelor's or Higher",
    y = "Log Asthma ED Visit Rate"
  )
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```r
# 2022
ggplot(EV_data_2022, aes(x = EDUCpercBAplus, y = log_AgeAdj_RoA_ED_Visit_Rate)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", color = "darkorange") +
  labs(
    title = "2022: % with BA+ vs Log Asthma ED Visit Rate",
    x = "% with Bachelor's or Higher",
    y = "Log Asthma ED Visit Rate"
  )
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
EV_data_2013 <- EV_data %>%
  filter( yr == 2013)
lm(EV_data_2013$log_AgeAdj_RoA_ED_Visit_Rate ~ EV_data_2013$nZEV1000pop    , data = EV_data)
```

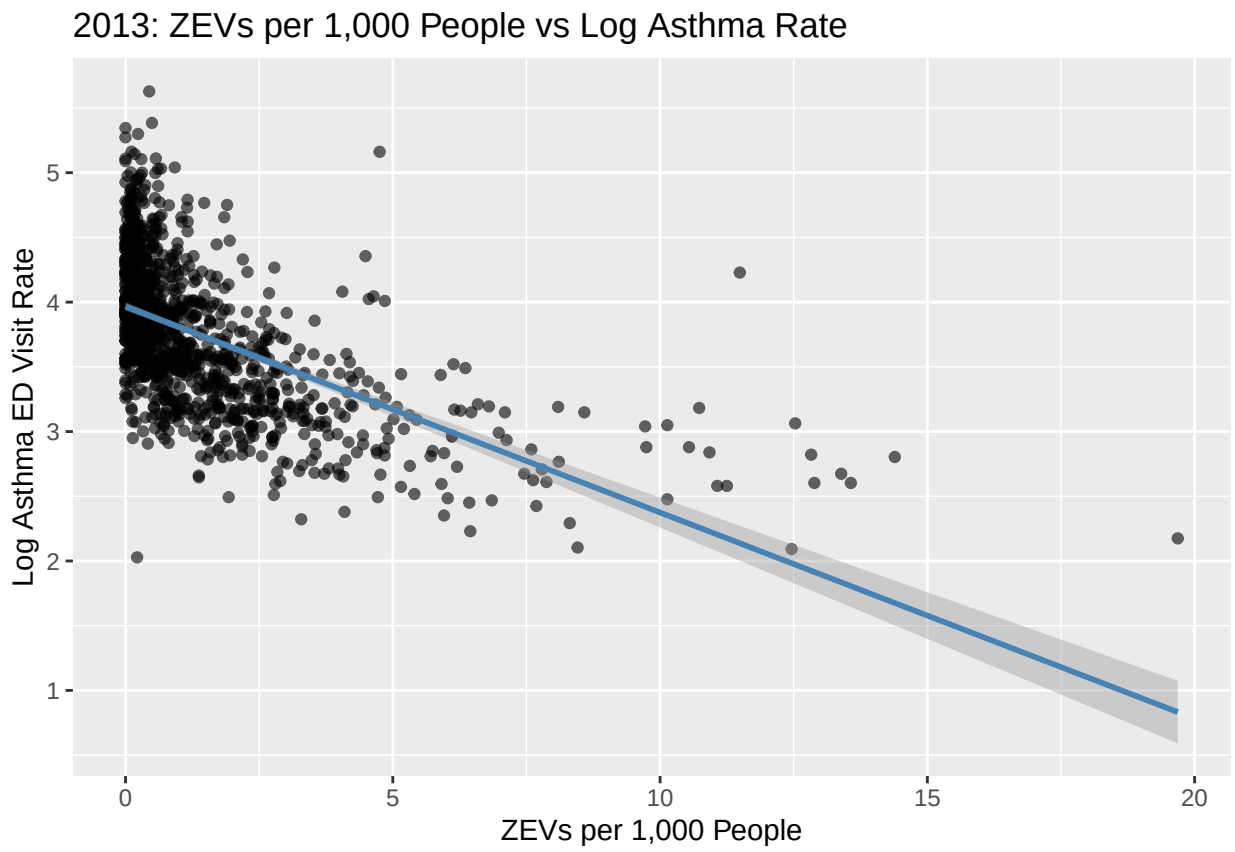
```
##
## Call:
## lm(formula = EV_data_2013$log_AgeAdj_RoA_ED_Visit_Rate ~ EV_data_2013$nZEV1000pop,
##     data = EV_data)
##
## Coefficients:
##               (Intercept)  EV_data_2013$nZEV1000pop
##                    3.9667                   -0.1592
```

```
EV_data_2022 <- EV_data %>%
  filter( yr == 2022)
lm(EV_data_2022$log_AgeAdj_RoA_ED_Visit_Rate ~ EV_data_2022$nZEV1000pop , data = EV_data)
```

```
##
## Call:
## lm(formula = EV_data_2022$log_AgeAdj_RoA_ED_Visit_Rate ~ EV_data_2022$nZEV1000pop,
##     data = EV_data)
##
## Coefficients:
##               (Intercept)  EV_data_2022$nZEV1000pop
##                   3.71291                  -0.01284
```

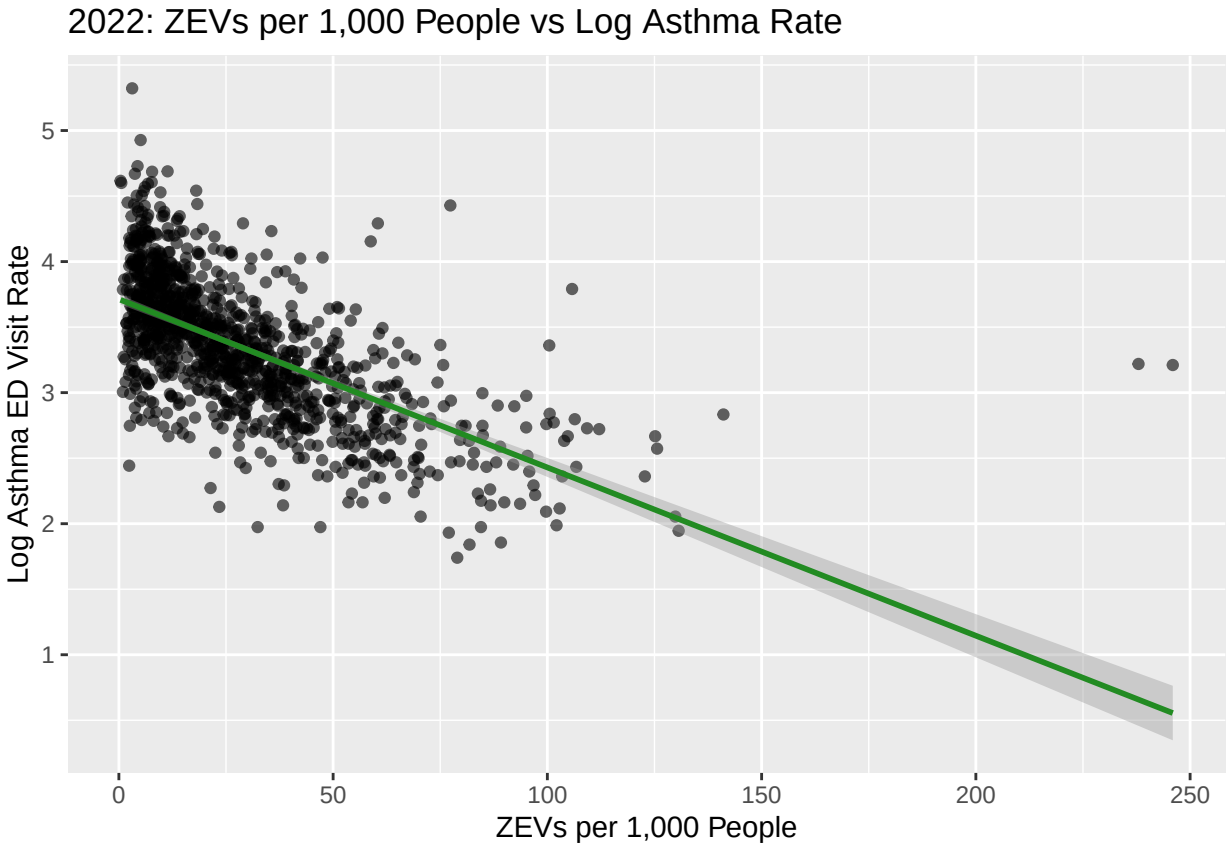
```
# ZEVs per 1,000 vs Log Asthma
# 2013
ggplot(EV_data_2013, aes(x = nZEV1000pop, y = log_AgeAdj_RoA_ED_Visit_Rate)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", color = "steelblue") +
  labs(
    title = "2013: ZEVs per 1,000 People vs Log Asthma Rate",
    x = "ZEVs per 1,000 People",
    y = "Log Asthma ED Visit Rate"
  )
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
# 2022
ggplot(EV_data_2022, aes(x = nZEV1000pop, y = log_AgeAdj_RoA_ED_Visit_Rate)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", color = "forestgreen") +
  labs(
    title = "2022: ZEVs per 1,000 People vs Log Asthma Rate",
    x = "ZEVs per 1,000 People",
    y = "Log Asthma ED Visit Rate"
  )
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
pov <-lm(log_AgeAdj_RoA_ED_Visit_Rate ~ I(nZEV1000pop/10) + percPoverty, data= EV_data_2022)

HH <- lm(log_AgeAdj_RoA_ED_Visit_Rate ~I(nZEV1000pop/10) + HHincomeMedian, data= EV_data_2022)

EDUC <- lm(log_AgeAdj_RoA_ED_Visit_Rate ~ I(nZEV1000pop/10) +EDUCpercBAplus, data= EV_data_2022)

summary(pov)
```

```
##
## Call:
## lm(formula = log_AgeAdj_RoA_ED_Visit_Rate ~ I(nZEV1000pop/10) +
##     percPoverty, data = EV_data_2022)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.49785 -0.23010  0.00285  0.22755  2.10010
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         3.350678   0.035752   93.72   <2e-16 ***
## I(nZEV1000pop/10) -0.093791   0.005461  -17.17   <2e-16 ***
## percPoverty         0.019247   0.001658   11.61   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.3881 on 1114 degrees of freedom
## Multiple R-squared:  0.4527, Adjusted R-squared:  0.4517
## F-statistic: 460.8 on 2 and 1114 DF,  p-value: < 2.2e-16
```

```
summary(HH)
```

```
##
## Call:
## lm(formula = log_AgeAdj_RoA_ED_Visit_Rate ~ I(nZEV1000pop/10) +
##     HHincomeMedian, data = EV_data_2022)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.60296 -0.22855 -0.00498  0.23181  1.75733
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         4.072e+00  3.260e-02 124.899  < 2e-16 ***
## I(nZEV1000pop/10) -5.608e-02  7.174e-03  -7.817 1.24e-14 ***
## HHincomeMedian    -7.117e-06  5.487e-07 -12.972  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.383 on 1114 degrees of freedom
## Multiple R-squared:  0.467,  Adjusted R-squared:  0.4661
## F-statistic: 488.1 on 2 and 1114 DF,  p-value: < 2.2e-16
```

```
summary(EDUC)
```

```
##
## Call:
## lm(formula = log_AgeAdj_RoA_ED_Visit_Rate ~ I(nZEV1000pop/10) +
##     EDUCpercBAplus, data = EV_data_2022)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.37149 -0.22033 -0.01322  0.20592  1.81520
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         3.9470905  0.0226121  174.56  < 2e-16 ***
## I(nZEV1000pop/10) -0.0349378  0.0074818   -4.67 3.38e-06 ***
## EDUCpercBAplus    -0.0150435  0.0009744  -15.44  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3729 on 1114 degrees of freedom
## Multiple R-squared:  0.4946, Adjusted R-squared:  0.4937
## F-statistic: 545.2 on 2 and 1114 DF,  p-value: < 2.2e-16
```

```
exp(-5.608e-02)
```

```
## [1] 0.9454635
```

```r
exp(-0.0349378)
```

```
## [1] 0.9656655
```

```r
1-0.9454635
```

```
## [1] 0.0545365
```

```r
1-0.9656655
```

```
## [1] 0.0343345
```

```r
0.0545365 *100
```

```
## [1] 5.45365
```

```r
0.0343345 *100
```

```
## [1] 3.43345
```

```r
#Mutli Linear Regression Model 2022

multi_model_2022_poverty <- lm(log_AgeAdj_RoA_ED_Visit_Rate ~ I(nZEV1000pop/10)   +  percPoverty, data=

summary(multi_model_2022_poverty)
```

```
##
## Call:
## lm(formula = log_AgeAdj_RoA_ED_Visit_Rate ~ I(nZEV1000pop/10) +
##     percPoverty, data = EV_data_2022)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.49785 -0.23010  0.00285  0.22755  2.10010
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        3.350678   0.035752   93.72   <2e-16 ***
## I(nZEV1000pop/10) -0.093791   0.005461  -17.17   <2e-16 ***
## percPoverty        0.019247   0.001658   11.61   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3881 on 1114 degrees of freedom
## Multiple R-squared:  0.4527, Adjusted R-squared:  0.4517
## F-statistic: 460.8 on 2 and 1114 DF,  p-value: < 2.2e-16
```

```r
exp(-0.093791)
```

```
## [1] 0.910473
```

```r
(1- 0.910473) *100
```

```
## [1] 8.9527
```

```r
#Mutli Linear Regression Model 2022 No SES

multi_model_2022 <- lm(log_AgeAdj_RoA_ED_Visit_Rate ~ I(nZEV1000pop/10), data= EV_data_2022)

summary(multi_model_2022)
```

```
##
## Call:
## lm(formula = log_AgeAdj_RoA_ED_Visit_Rate ~ I(nZEV1000pop/10),
##     data = EV_data_2022)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.32297 -0.24895 -0.01654  0.23441  2.65556
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         3.712907   0.018469   201.0   <2e-16 ***
## I(nZEV1000pop/10) -0.128381   0.004844   -26.5   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4107 on 1115 degrees of freedom
## Multiple R-squared:  0.3865, Adjusted R-squared:  0.386
## F-statistic: 702.5 on 1 and 1115 DF,  p-value: < 2.2e-16
```

```r
exp(-0.128381)
```

```
## [1] 0.8795182
```

```r
(1 - 0.8795182) * 100
```

```
## [1] 12.04818
```

‘‘‘