Regression

Justin Coles,Jessy Garcia, Joseph Salamante

2025-06-30

```
library(ggplot2)
library(dplyr)
```

```
##
## 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
```

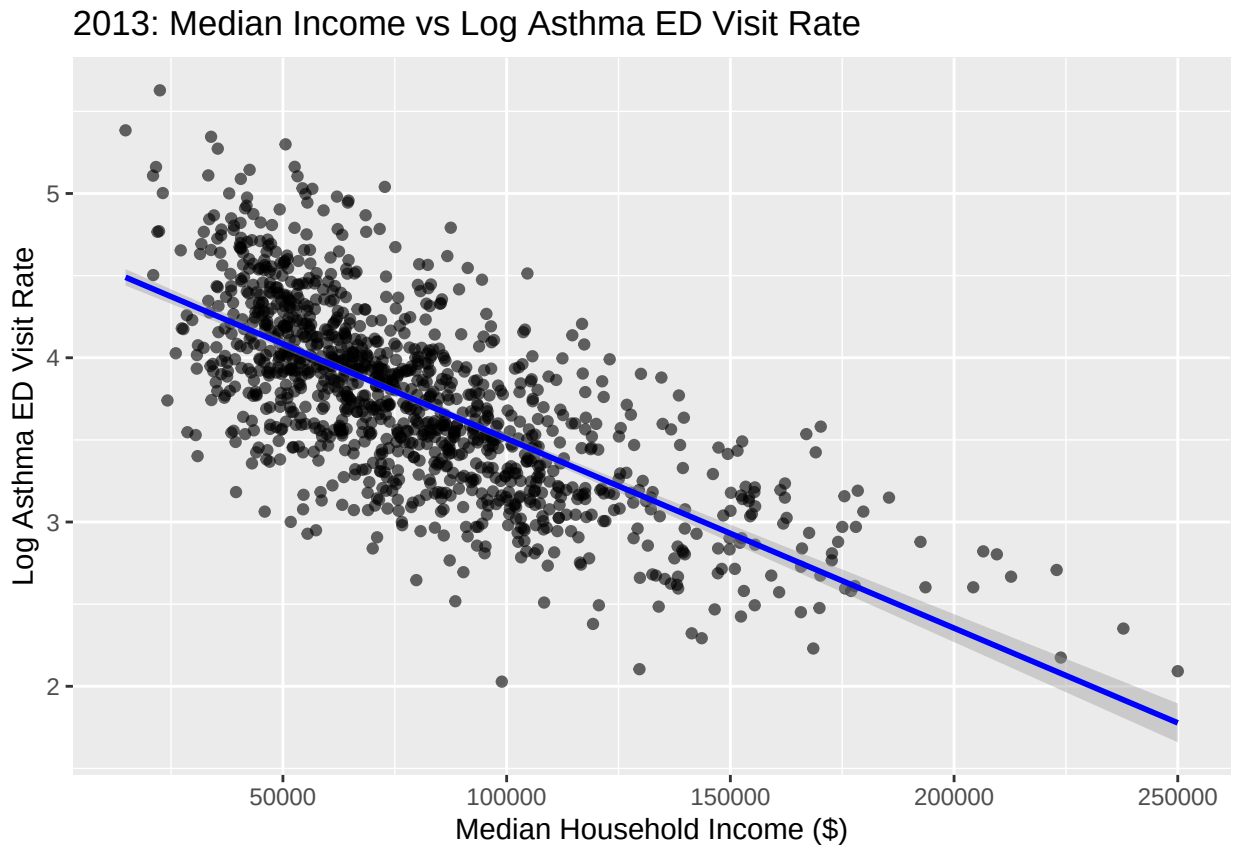
```
getwd()
```

```
## [1] "/Users/jessygarcia/LA BEST EV GROUP PROJECT/EV PROJECT"
```

```
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)
```

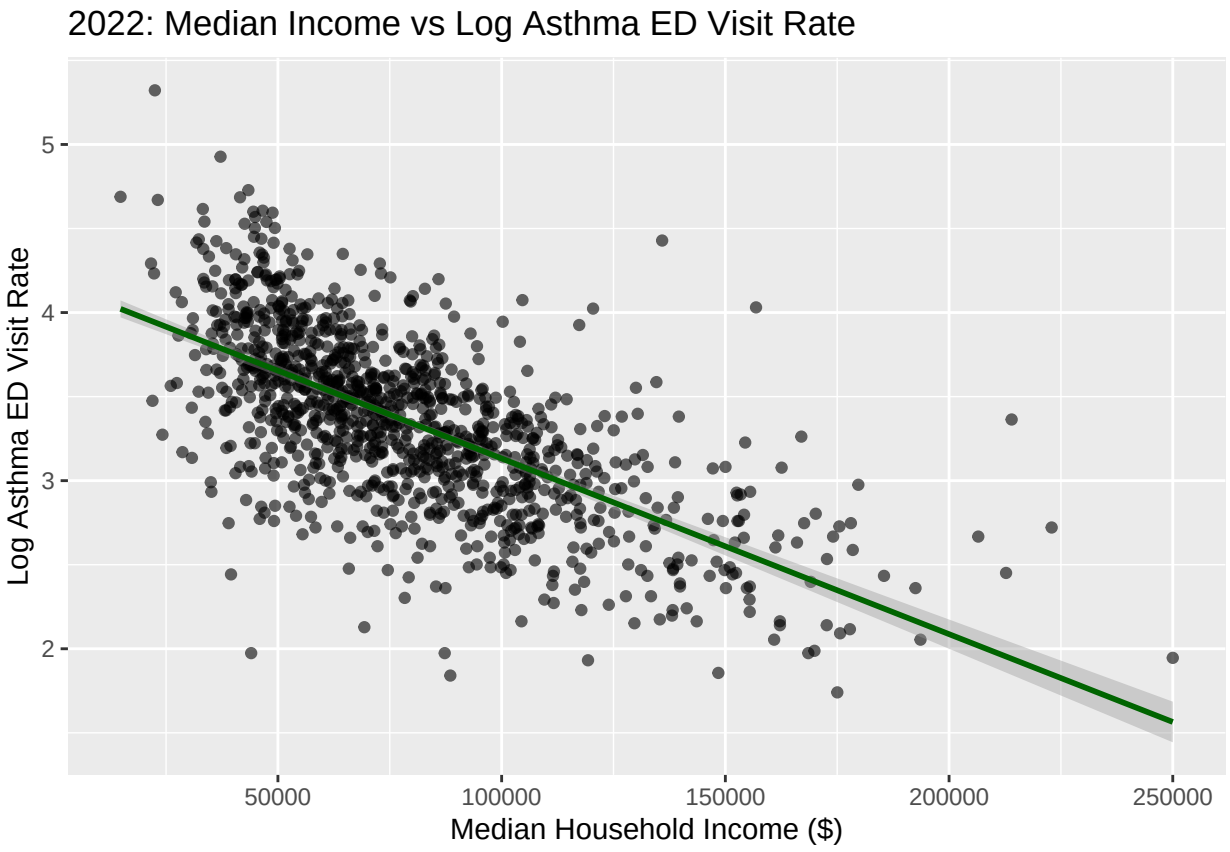
```
ggplot(EV_data_2013, aes(x = HHincomeMedian, y = log_AgeAdj_RoA_ED_Visit_Rate)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = TRUE, color = "blue") +
  labs(
    title = "2013: Median Income vs Log Asthma ED Visit Rate",
    x = "Median Household Income ($)",
    y = "Log Asthma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

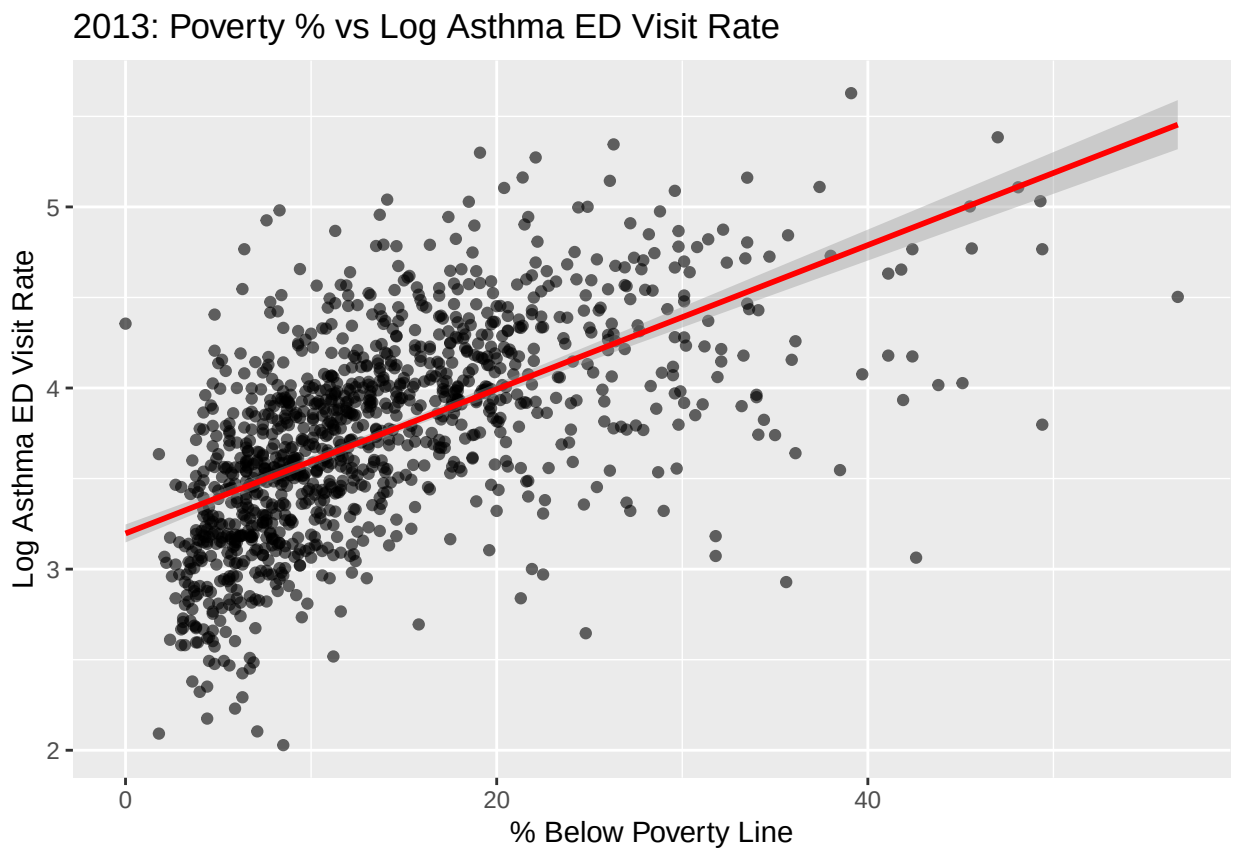
```
lm(EV_data_2022$log_AgeAdj_RoA_ED_Visit_Rate ~ EV_data_2022$percPoverty, data = EV_data)
```

```
##
## 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
```

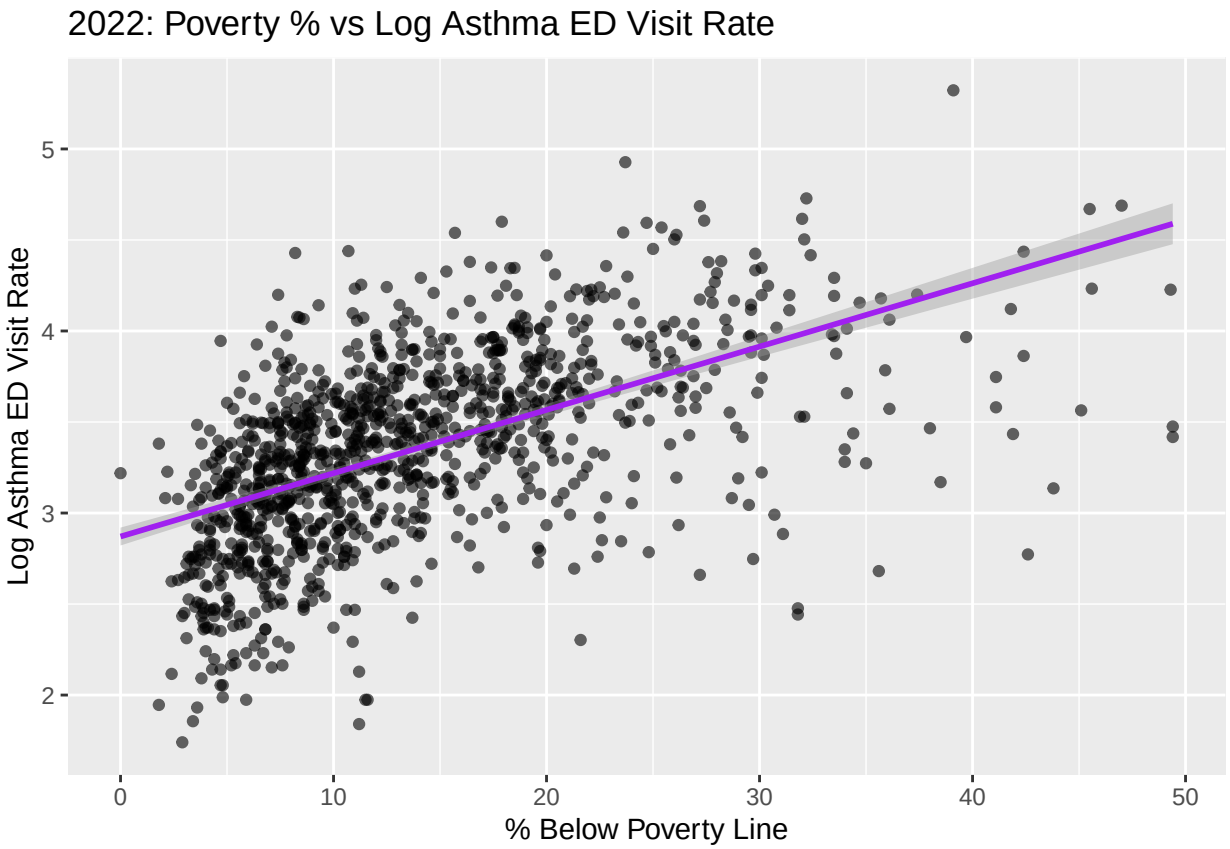
```
ggplot(EV_data_2013, aes(x = percPoverty, y = log_AgeAdj_RoA_ED_Visit_Rate)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", color = "red") +
  labs(
    title = "2013: Poverty % vs Log Asthma ED Visit Rate",
    x = "% Below Poverty Line",
    y = "Log Asthma ED Visit Rate"
  )
```

```
## `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)
```

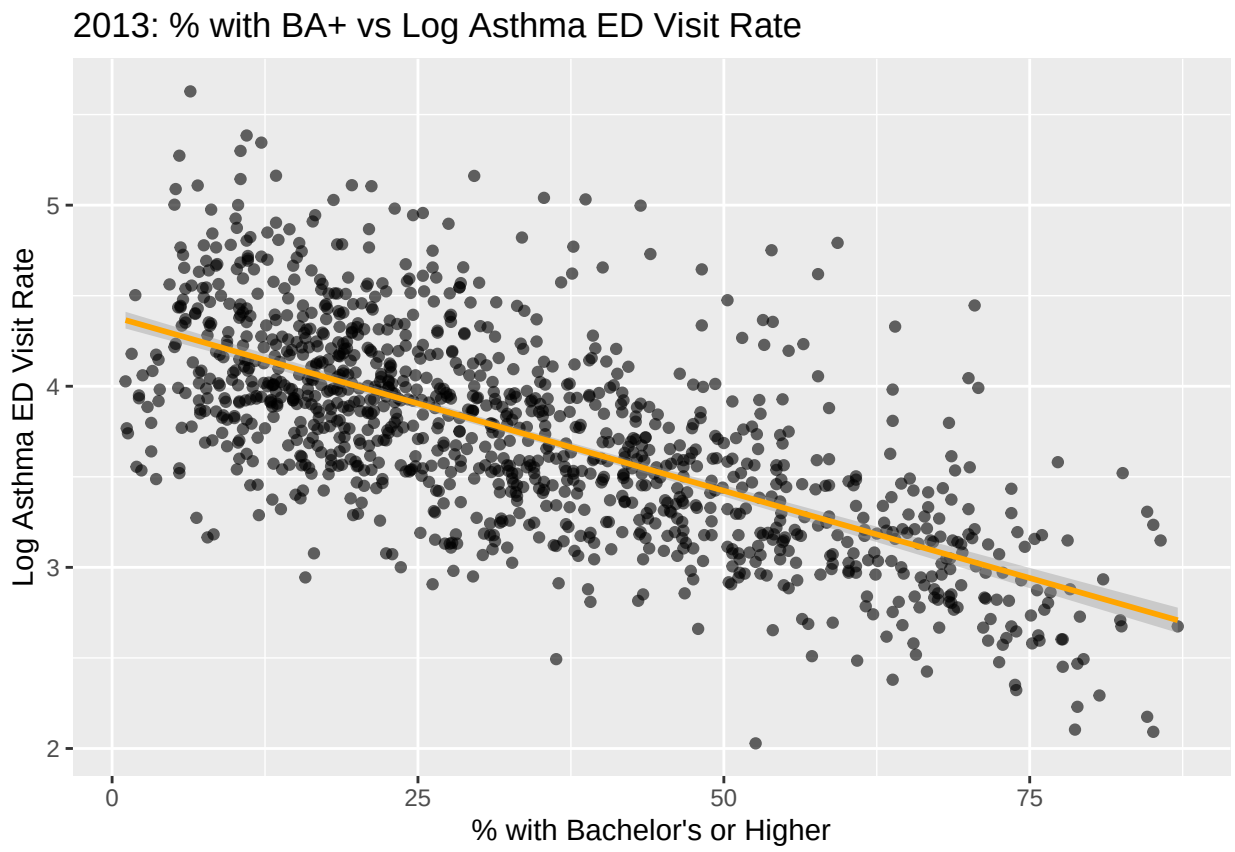
```
##
## Call:
## lm(formula = EV_data_2013$log_AgeAdj_RoA_ED_Visit_Rate ~ EV_data_2013$EDUCpercBAplus,
##     data = EV_data)
##
## Coefficients:
##                  (Intercept)  EV_data_2013$EDUCpercBAplus
##                      4.38614                      -0.01927
```

```
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
```

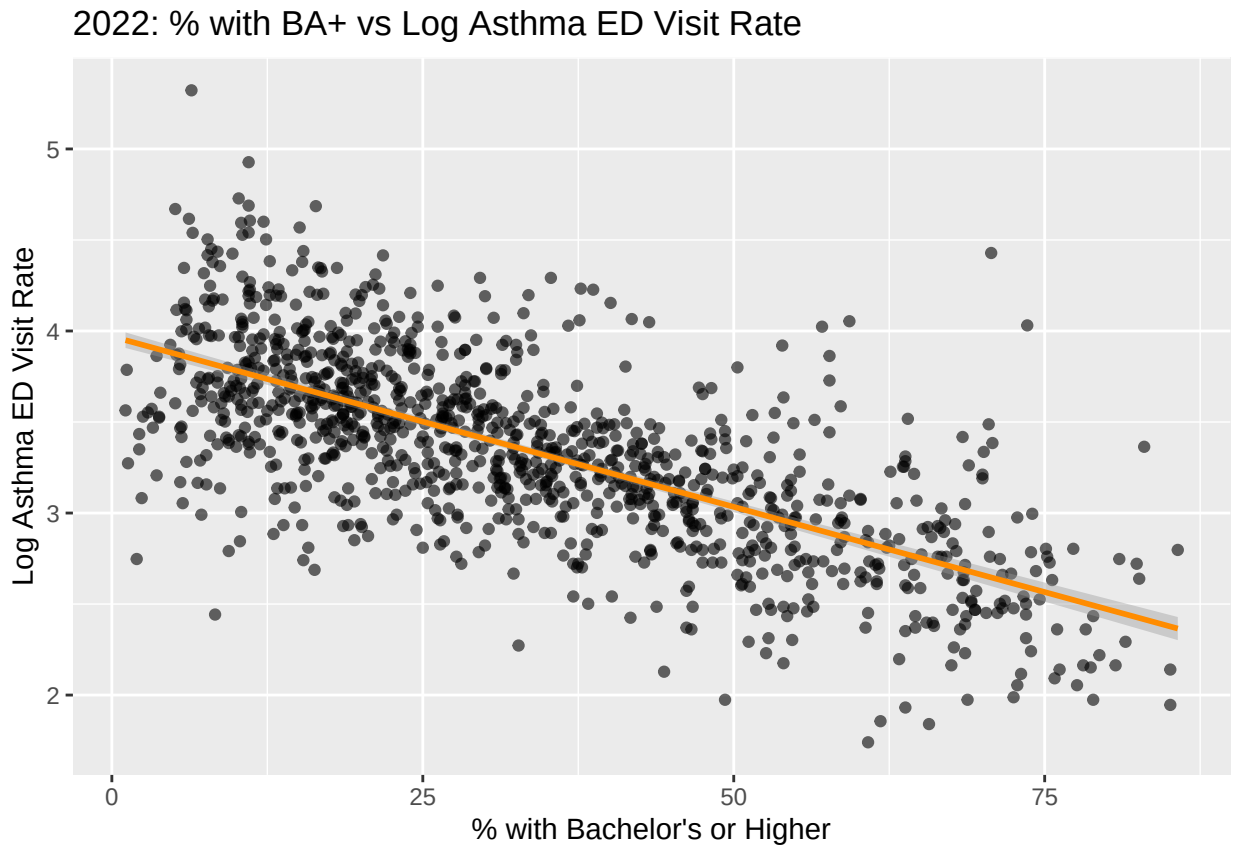
```r
#BA Education vs Log Asthma Rate
# 2013
ggplot(EV_data_2013, aes(x = EDUCpercBAplus, y = log_AgeAdj_RoA_ED_Visit_Rate)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", color = "orange") +
  labs(
    title = "2013: % with BA+ vs Log Asthma ED Visit Rate",
    x = "% with Bachelor's or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

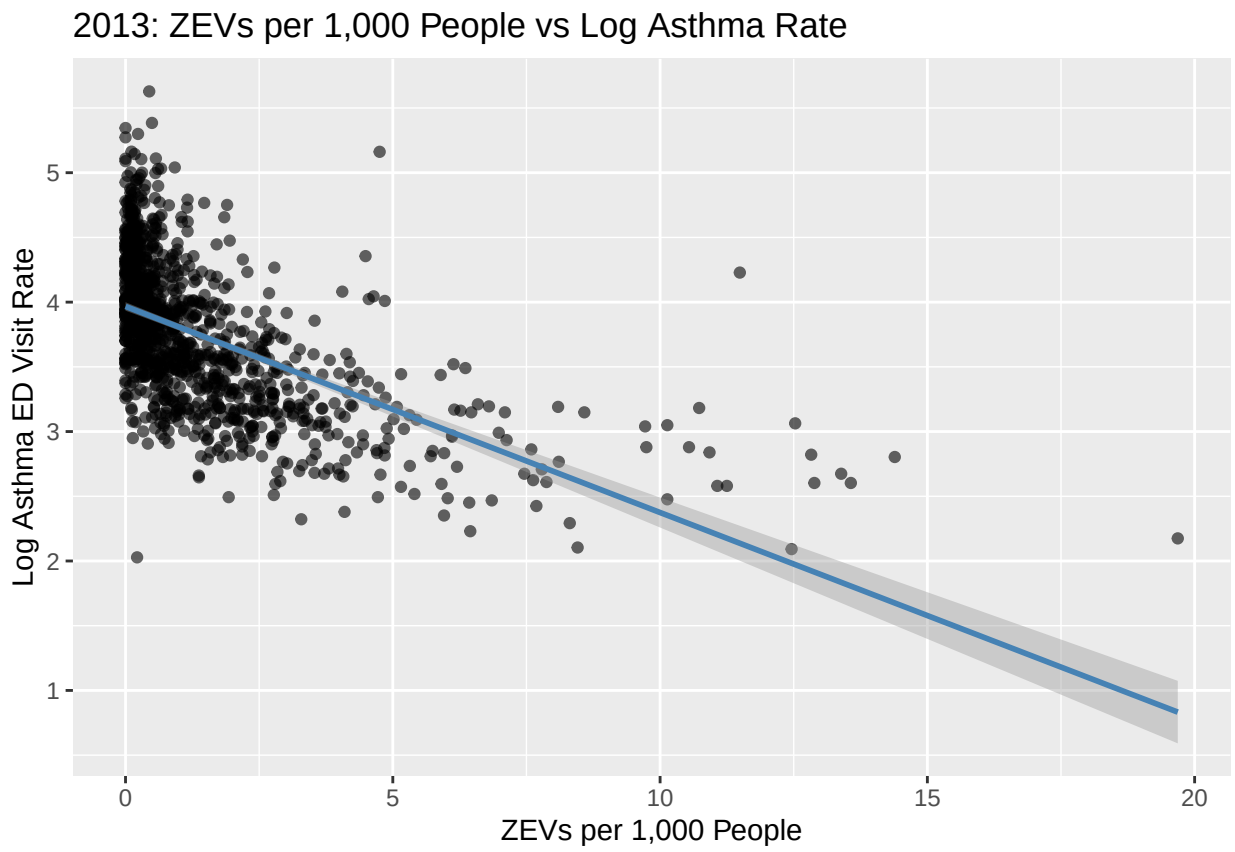
```
##
## Call:
## lm(formula = EV_data_2013$log_AgeAdj_RoA_ED_Visit_Rate ~ EV_data_2013$nZEV1000pop,
##     data = EV_data)
##
## Coefficients:
##               (Intercept)  EV_data_2013$nZEV1000pop
##                    3.9667                   -0.1592
```

```
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
```

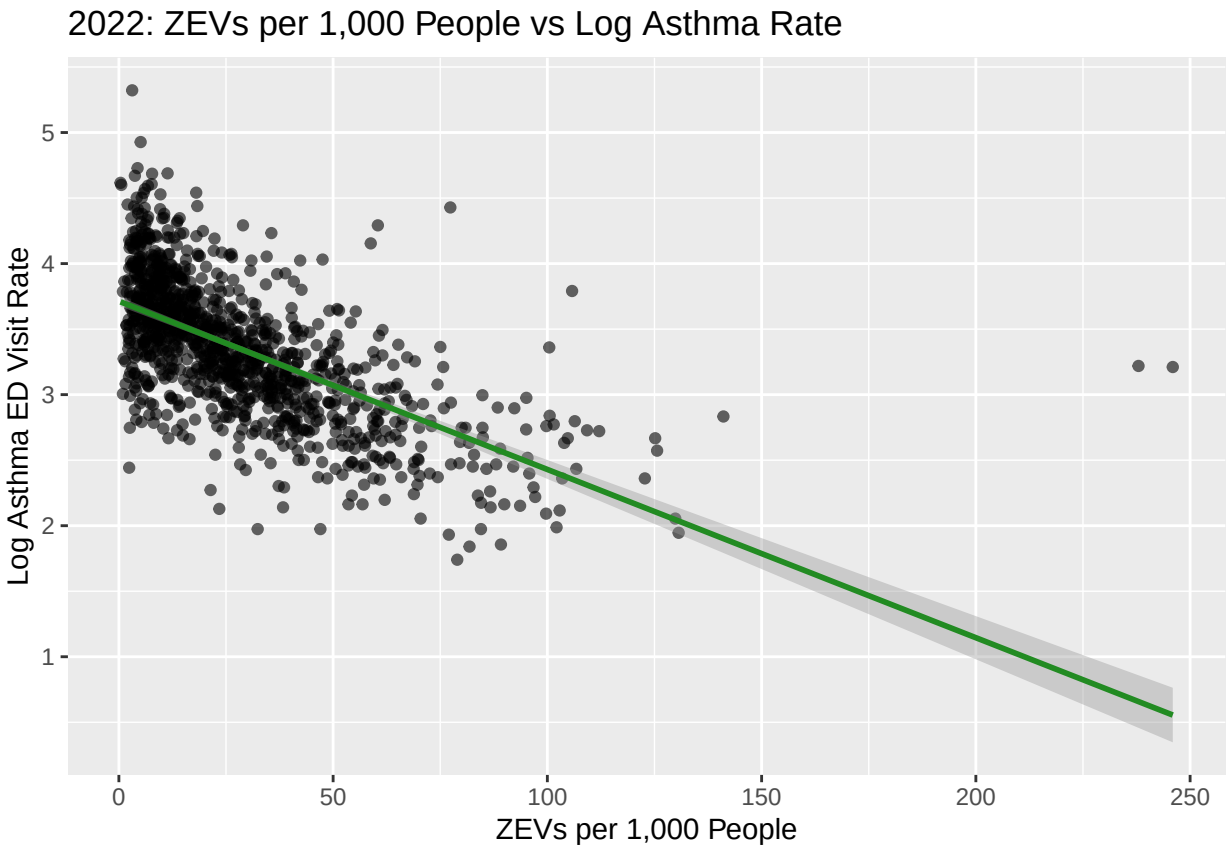
```
# ZEVs per 1,000 vs Log Asthma
# 2013
ggplot(EV_data_2013, aes(x = nZEV1000pop, y = log_AgeAdj_RoA_ED_Visit_Rate)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", color = "steelblue") +
  labs(
    title = "2013: ZEVs per 1,000 People vs Log Asthma Rate",
    x = "ZEVs per 1,000 People",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




```
#Mutli Linear Regression Model 2022

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

exp(-0.093791)
```

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

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

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

```
#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
```

```
exp(-0.128381)
```

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

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

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

```
# more data processing
EV_data$yrC <- EV_data$yr - 2013 # create a year variable centered at the first study year, 2013
EV_data$yrC2 <- EV_data$yrC^2 # squared centered year term to allow for nonlinear trends in time
EV_data$yrC3 <- EV_data$yrC^3 # cubic centered year term to allow for nonlinear trends in time (you mig

# code for random intercept longitudinal model (I'll draw this on the board for you)

# load nlme package to fit random effects (longitudinal) model
library(nlme)
```

```
##
## Attaching package: 'nlme'

## The following object is masked from 'package:dplyr':
##
##     collapse
```

```
# requires dataset to be non-missing in included variables
d_complete <- subset(EV_data,!is.na(nZEV1000pop) & !is.na(percPoverty))

# fit model, with random intercept for zip
fit1 <- lme(log_AgeAdj_RoA_ED_Visit_Rate ~ nZEV1000pop + yrC + yrC2 + EDUCpercBAplus,
                            random=list(~1|zip),# random intercept at zip code level
                            data=d_complete,
                            method="REML")
summary(fit1)
```

```
## Linear mixed-effects model fit by REML
##   Data: d_complete
```

```
##        AIC      BIC    logLik
##   5087.011 5138.392 -2536.505
##
## Random effects:
##  Formula: ~1 | zip
##         (Intercept)  Residual
## StdDev:   0.3802859 0.2559741
##
## Fixed effects:  log_AgeAdj_RoA_ED_Visit_Rate ~ nZEV1000pop + yrC + yrC2 + EDUCpercBAplus
##                    Value   Std.Error    DF   t-value p-value
## (Intercept)     4.473164 0.022678678 10147 197.24095       0
## nZEV1000pop     0.001325 0.000317465 10147   4.17316       0
## yrC            -0.016708 0.003083633 10147  -5.41835       0
## yrC2           -0.006897 0.000338483 10147 -20.37565       0
## EDUCpercBAplus -0.020239 0.000573577  1240 -35.28477       0
##  Correlation:
##                (Intr) nZEV10 yrC    yrC2
## nZEV1000pop     0.190
## yrC            -0.225 -0.073
## yrC2            0.127 -0.221 -0.919
## EDUCpercBAplus -0.839 -0.256  0.019  0.057
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -5.46830309 -0.56318312  0.07079687  0.61322066  4.64386351
##
## Number of Observations: 11392
## Number of Groups: 1242
```

```
intervals(fit1)
```

```
## Approximate 95% confidence intervals
##
##  Fixed effects:
##                         lower         est.        upper
## (Intercept)     4.4287092645  4.473163959  4.517618653
## nZEV1000pop     0.0007025361  0.001324830  0.001947123
## yrC            -0.0227527208 -0.016708190 -0.010663659
## yrC2           -0.0075603165 -0.006896822 -0.006233327
## EDUCpercBAplus -0.0213638207 -0.020238532 -0.019113244
##
##  Random Effects:
##   Level: zip
##                       lower      est.     upper
## sd((Intercept)) 0.3647338 0.3802859 0.3965011
##
##  Within-group standard error:
##     lower      est.     upper
## 0.2524755 0.2559741 0.2595210
```

```
#class(d_complete$zip)
#factor(d_complete$zip)
```

```r
# fit another model, with also a random slope on year
fit2 <- lme(log_AgeAdj_RoA_ED_Visit_Rate ~ nZEV1000pop + yrC + yrC2 + EDUCpercBAplus,
    #control=c(maxIter=5000),
                            random=list(~yrC|zip),# random intercept at zip code level
                            data=d_complete,
                            method="REML")
summary(fit2)
```

```
## Linear mixed-effects model fit by REML
##   Data: d_complete
##        AIC      BIC    logLik
##   4881.836 4947.898 -2431.918
##
## Random effects:
##  Formula: ~yrC | zip
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 0.4179915 (Intr)
## yrC         0.0250175 -0.437
## Residual    0.2448858
##
## Fixed effects:  log_AgeAdj_RoA_ED_Visit_Rate ~ nZEV1000pop + yrC + yrC2 + EDUCpercBAplus
##                    Value   Std.Error    DF   t-value p-value
## (Intercept)     4.496935 0.023726383 10147 189.53309       0
## nZEV1000pop     0.002268 0.000380348 10147   5.96231       0
## yrC            -0.017318 0.003046032 10147  -5.68552       0
## yrC2           -0.007075 0.000329181 10147 -21.49208       0
## EDUCpercBAplus -0.021023 0.000597763  1240 -35.16889       0
##  Correlation:
##                (Intr) nZEV10 yrC    yrC2
## nZEV1000pop     0.286
## yrC            -0.264 -0.088
## yrC2            0.073 -0.274 -0.870
## EDUCpercBAplus -0.831 -0.375  0.032  0.104
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -5.15786814 -0.54237302  0.06284579  0.59483162  4.55476946
##
## Number of Observations: 11392
## Number of Groups: 1242
```

```r
intervals(fit2)
```

```
## Approximate 95% confidence intervals
##
##  Fixed effects:
##                        lower         est.        upper
## (Intercept)     4.450426421  4.496934825  4.543443229
## nZEV1000pop     0.001522192  0.002267749  0.003013305
## yrC            -0.023289115 -0.017318289 -0.011347463
## yrC2           -0.007720042 -0.007074782 -0.006429523
## EDUCpercBAplus -0.022195387 -0.021022649 -0.019849911
```

```
##
##  Random Effects:
##   Level: zip
##                             lower       est.      upper
## sd((Intercept))       0.39932700  0.4179915  0.4375283
## sd(yrC)               0.02266712  0.0250175  0.0276116
## cor((Intercept),yrC) -0.50788075 -0.4365252 -0.3592201
##
##  Within-group standard error:
##     lower      est.     upper
## 0.2412882 0.2448858 0.2485371
```

```
exp(0.002267749)
```

```
## [1] 1.00227
```

```
(1.00227-1) * 100
```

```
## [1] 0.227
```