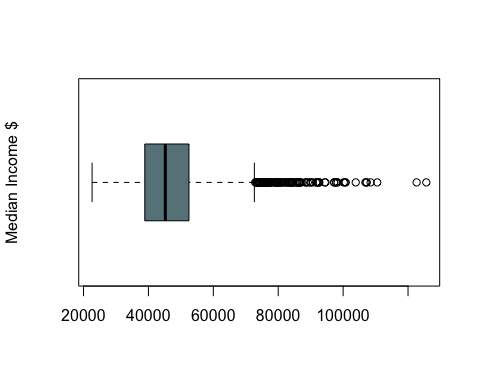
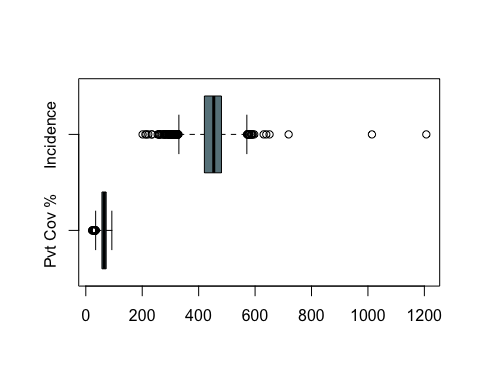
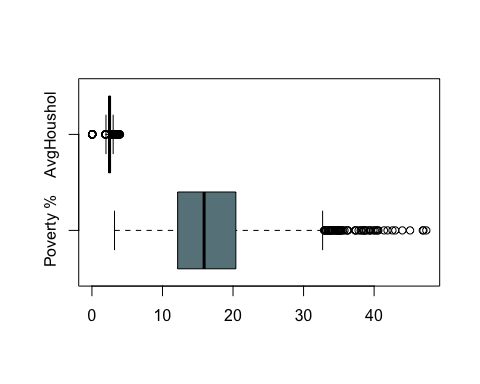
Group Project

field1, field2, field3

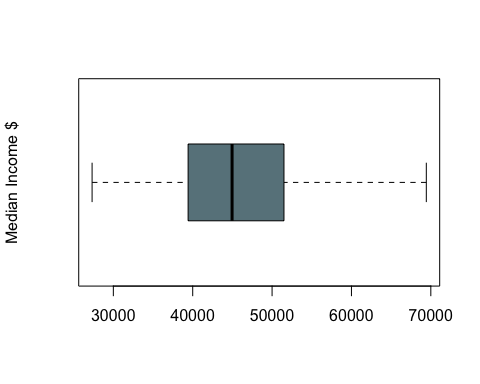
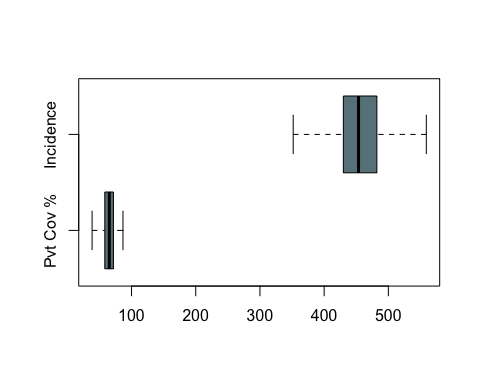
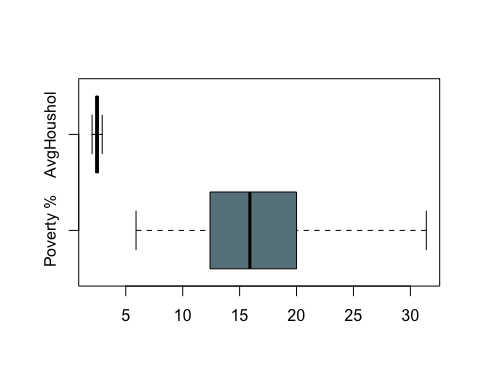
10/6/2021

# Project



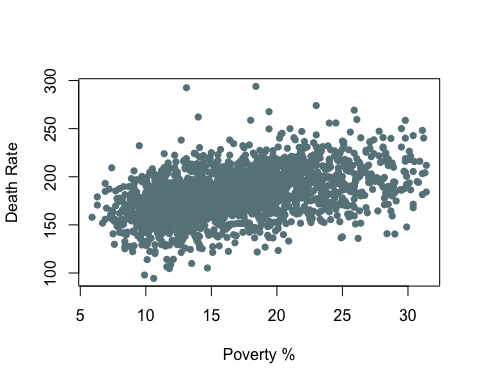
### Write something about it

We are going to remove outliers

 ### We removed all of the 447 outliers from dataset!

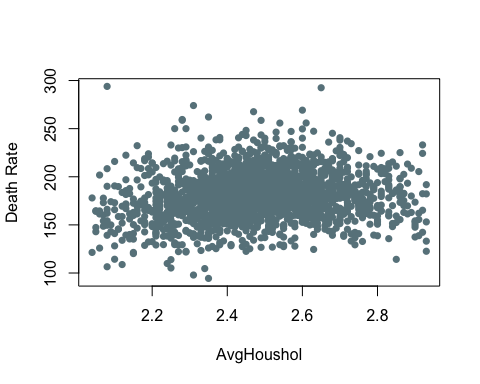
# Plots between predictors and death rate with their respective correlation coefficients

We need to write something underneath each of the plots



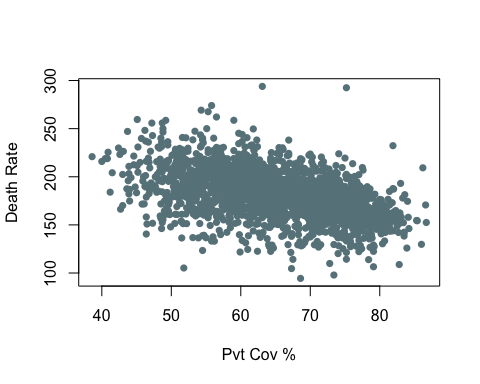
## [1] 0.4530695

### Write something here about this



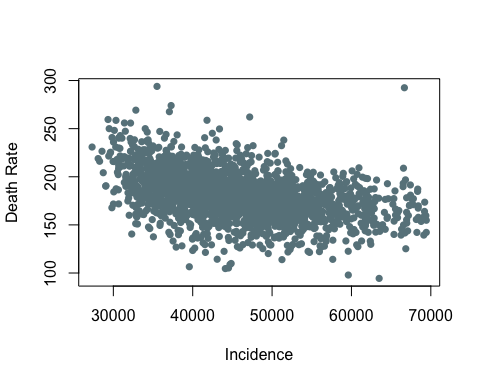
## [1] 0.155252

### Write something here about this



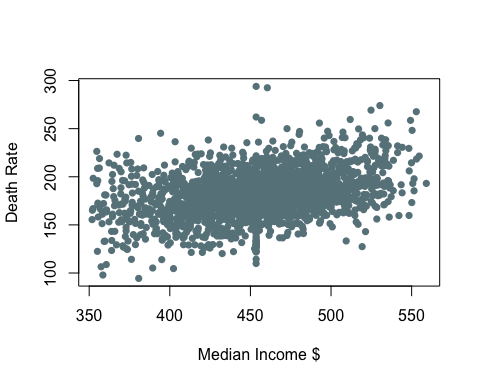
## [1] -0.4569882

### Write something here about this



## [1] -0.4461167

### Write something here about this



## [1] 0.3519352

### Write something here about this

### There are 2154 rows in our dataframe. We have randomly select 80% of rows from our dataframe for a total of 1723 data. Those data are now parts of a Train dataset we will use to perform regression analysis. At the same time we used the remaining 431 data for testing.

### Created two .txt files with data formatted as table. Uploaded on GitHub. This RMarkdown reads data directly from Github repository

## Now we are creating a multiple linear regression model to predict cancer death rate using our predictors

##   
## Call:  
## lm(formula = Traincancer\_data$TARGET\_deathRate ~ Traincancer\_data$povertyPercent +   
## Traincancer\_data$AvgHouseholdSize + Traincancer\_data$PctPrivateCoverage +   
## Traincancer\_data$incidenceRate + Traincancer\_data$medIncome +   
## as.factor(Traincancer\_data$Region.CAT.), data = Traincancer\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -71.103 -10.872 -0.368 10.724 107.361   
##   
## Coefficients:  
## Estimate Std. Error t value  
## (Intercept) 1.427e+02 1.191e+01 11.985  
## Traincancer\_data$povertyPercent -9.510e-02 1.822e-01 -0.522  
## Traincancer\_data$AvgHouseholdSize 9.557e+00 3.208e+00 2.979  
## Traincancer\_data$PctPrivateCoverage -7.079e-01 9.819e-02 -7.209  
## Traincancer\_data$incidenceRate 2.104e-01 1.152e-02 18.273  
## Traincancer\_data$medIncome -7.072e-04 1.204e-04 -5.872  
## as.factor(Traincancer\_data$Region.CAT.)Northeast -6.373e+00 1.747e+00 -3.649  
## as.factor(Traincancer\_data$Region.CAT.)South 2.120e+00 1.257e+00 1.686  
## as.factor(Traincancer\_data$Region.CAT.)West -1.575e+01 1.777e+00 -8.861  
## Pr(>|t|)   
## (Intercept) < 2e-16 \*\*\*  
## Traincancer\_data$povertyPercent 0.601789   
## Traincancer\_data$AvgHouseholdSize 0.002935 \*\*   
## Traincancer\_data$PctPrivateCoverage 8.40e-13 \*\*\*  
## Traincancer\_data$incidenceRate < 2e-16 \*\*\*  
## Traincancer\_data$medIncome 5.16e-09 \*\*\*  
## as.factor(Traincancer\_data$Region.CAT.)Northeast 0.000271 \*\*\*  
## as.factor(Traincancer\_data$Region.CAT.)South 0.091912 .   
## as.factor(Traincancer\_data$Region.CAT.)West < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 18.36 on 1714 degrees of freedom  
## Multiple R-squared: 0.4256, Adjusted R-squared: 0.4229   
## F-statistic: 158.8 on 8 and 1714 DF, p-value: < 2.2e-16