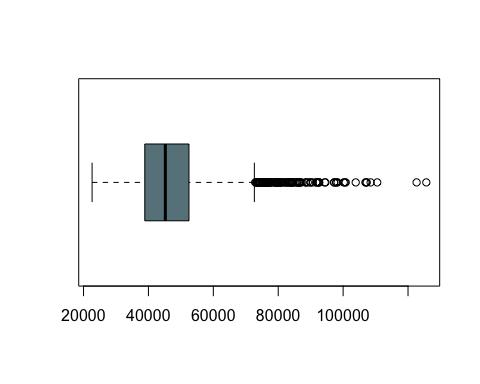
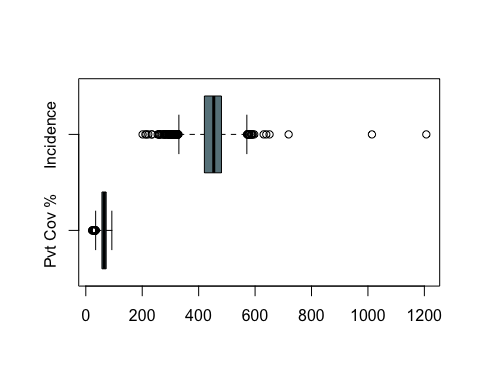
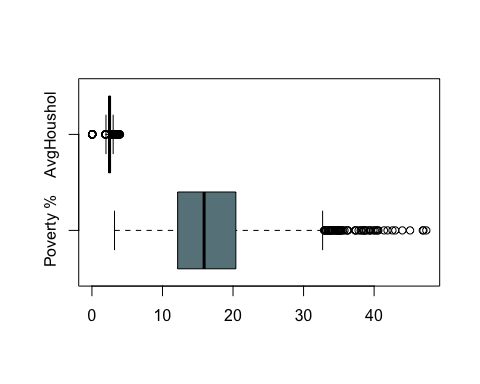
Group Project

field1, field2, field3

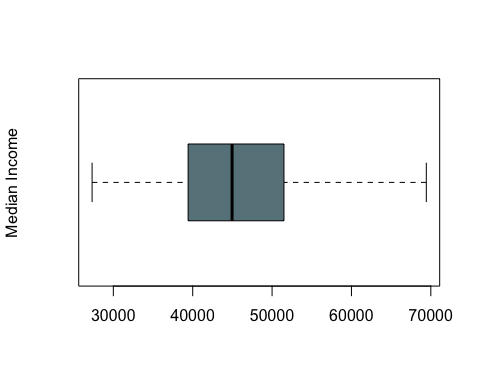
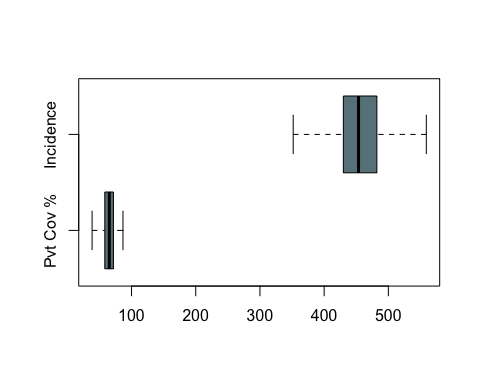
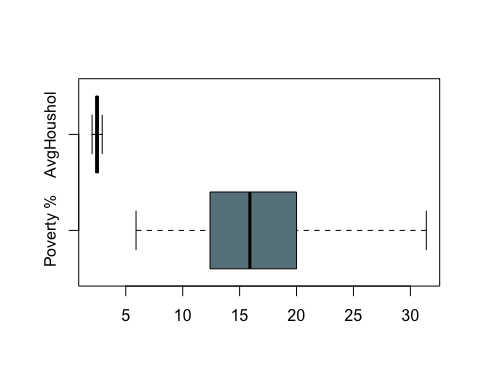
10/6/2021

# Project



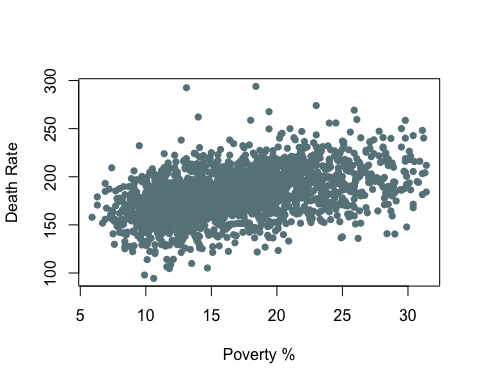
### Write something about it

We are going to remove outliers

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

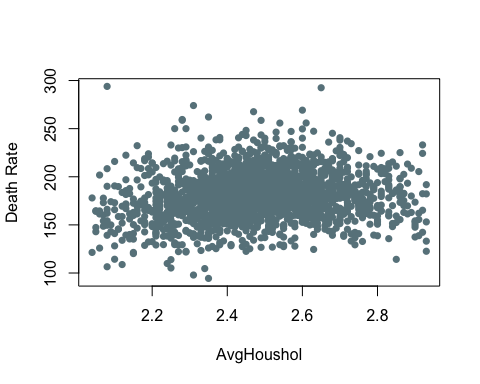
# Plots between predictors and death rate with their respective correlation coefficients

We need to write something underneath each of the plots



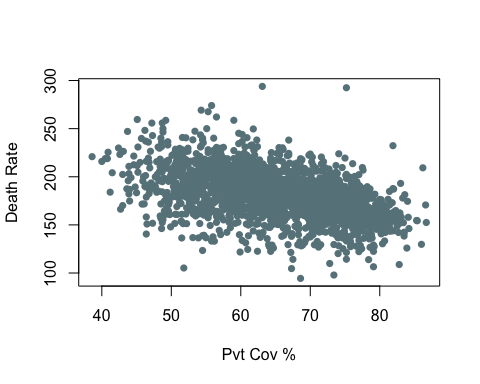
## [1] 0.4530695

### Write something here about this



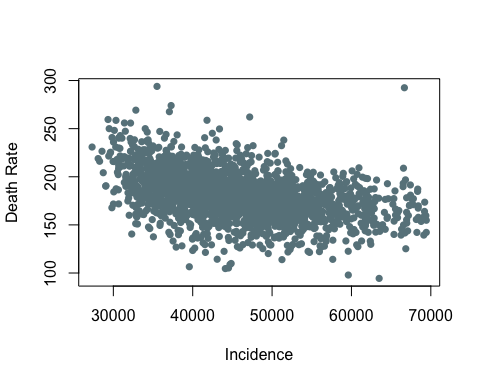
## [1] 0.4530695

### Write something here about this



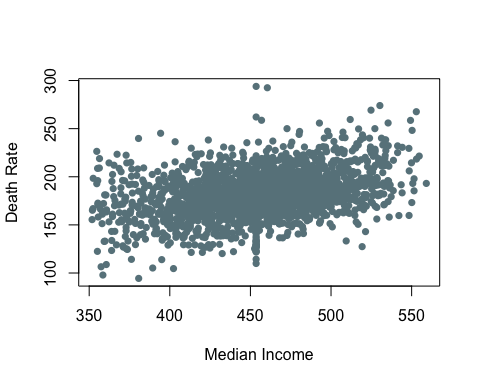
## [1] -0.4569882

### Write something here about this



## [1] -0.4461167

### Write something here about this



## [1] 0.3519352

### Write something here about this

### We have randomly 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,   
## data = Traincancer\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -69.722 -11.439 -0.331 11.178 131.949   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.154e+02 1.220e+01 9.457 < 2e-16 \*\*\*  
## Traincancer\_data$povertyPercent 9.171e-03 1.854e-01 0.049 0.961   
## Traincancer\_data$AvgHouseholdSize 1.641e+01 3.263e+00 5.031 5.40e-07 \*\*\*  
## Traincancer\_data$PctPrivateCoverage -5.745e-01 9.522e-02 -6.034 1.96e-09 \*\*\*  
## Traincancer\_data$incidenceRate 2.256e-01 1.169e-02 19.298 < 2e-16 \*\*\*  
## Traincancer\_data$medIncome -8.872e-04 1.200e-04 -7.391 2.26e-13 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 19.16 on 1717 degrees of freedom  
## Multiple R-squared: 0.3819, Adjusted R-squared: 0.3801   
## F-statistic: 212.2 on 5 and 1717 DF, p-value: < 2.2e-16