island_confounding

Analysis of island status as a confounding variable

Start by loading the six month data:

```
data <- read.csv(file = '../prepped_data/sixmonth.csv')
```

Regressions on island nation status

Run regressions with is_island as the sole explanatory variable:

```
summary(lm(formula = casepc ~ factor(is_island), data = data))
```

```
##
## Call:
## lm(formula = casepc ~ factor(is_island), data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.111 -3.368 -2.404  1.147 37.460
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)            4.1141     0.4882   8.428 1.16e-14 ***
## factor(is_island)TRUE -0.9532     1.0767  -0.885    0.377
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.838 on 178 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.004384,   Adjusted R-squared:  -0.00121
## F-statistic: 0.7837 on 1 and 178 DF,  p-value: 0.3772
```

```
plot(data$is_island, data$casepc, pch=19)
```
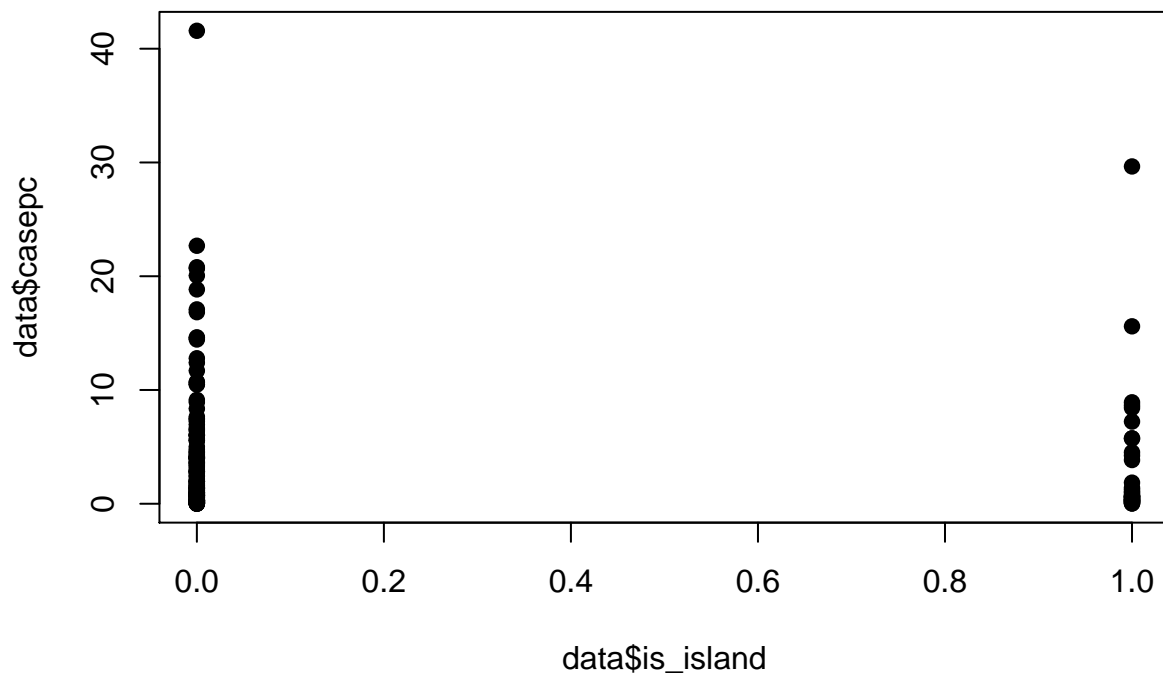
```r
summary(lm(formula = deathpc ~ factor(is_island), data = data))
```

```
##
## Call:
## lm(formula = deathpc ~ factor(is_island), data = data)
##
## Residuals:
##      Min      1Q   Median       3Q      Max
## -0.12256 -0.10717 -0.05387 -0.00094  1.11784
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            0.12256    0.01557   7.873 3.27e-13 ***
## factor(is_island)TRUE -0.07097    0.03434  -2.067   0.0402 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1862 on 178 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.02344,    Adjusted R-squared:  0.01796
## F-statistic: 4.273 on 1 and 178 DF,  p-value: 0.04018
```
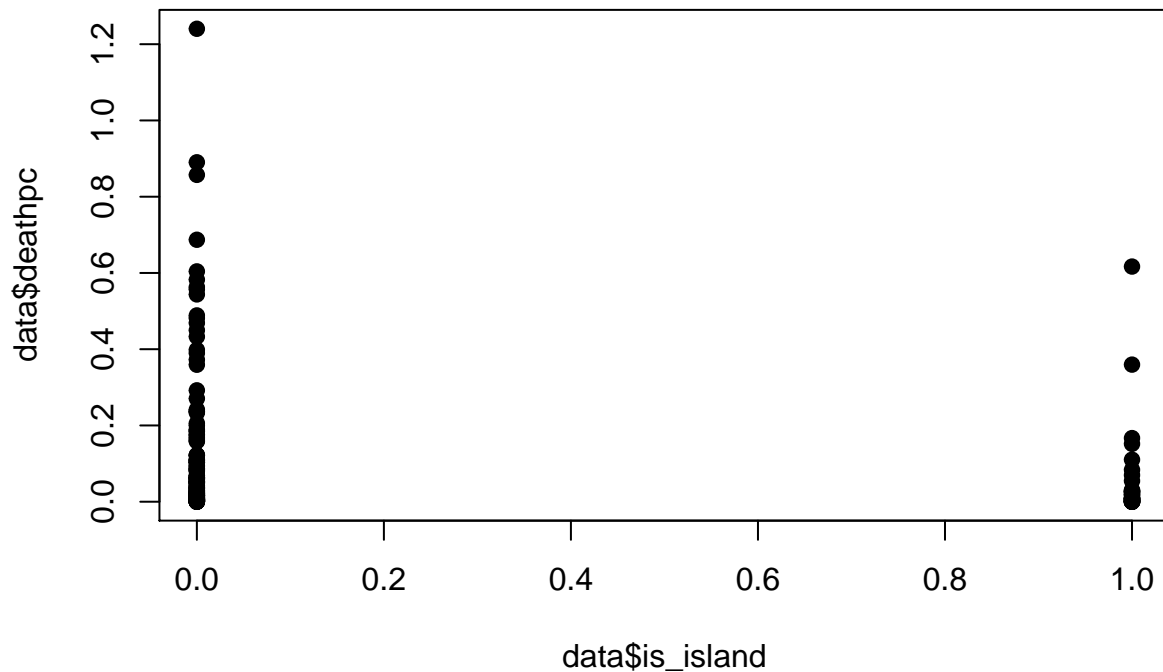
```r
plot(data$is_island, data$deathpc, pch=19)
```

```r
summary(lm(formula = cfratio ~ factor(is_island), data = data))
```

```
##
## Call:
## lm(formula = cfratio ~ factor(is_island), data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.0843 -1.7024 -0.8132  0.5641 25.8849
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)             3.0843     0.2694  11.450   <2e-16 ***
## factor(is_island)TRUE  -1.1317     0.5911  -1.914   0.0571 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.244 on 181 degrees of freedom
##   (9 observations deleted due to missingness)
## Multiple R-squared:  0.01985,    Adjusted R-squared:  0.01443
## F-statistic: 3.665 on 1 and 181 DF,  p-value: 0.05714
```

```r
plot(data$is_island, data$cfratio, pch=19)
```
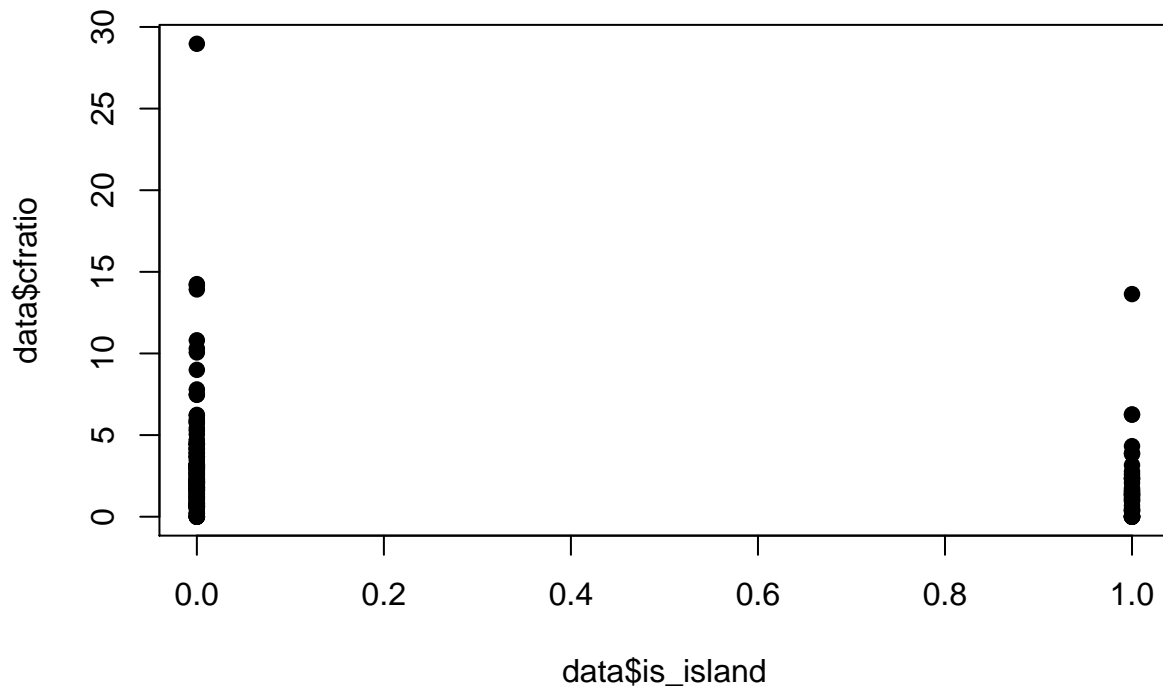
Overall findings are that the island nation status alone is not a great predictor of cases-per-capita, deaths-per-capita, and case fatality ratio, but from the plots you can see that island nations are on average lower on all three of these measures.

Regressions on GHSI overall score and island nation status

```
summary(lm(formula = casepc ~ overall, data = data))
```

```
##
## Call:
## lm(formula = casepc ~ overall, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.881 -3.102 -2.174  1.181 37.684
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.26832    1.34301   0.944   0.3463
## overall      0.06365    0.03055   2.083   0.0387 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.78 on 178 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.0238, Adjusted R-squared:  0.01832
```

```
## F-statistic:  4.34 on 1 and 178 DF,  p-value: 0.03866
summary(lm(formula = casepc ~ overall + factor(is_island), data = data))

##
## Call:
## lm(formula = casepc ~ overall + factor(is_island), data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.954  -3.086  -2.185   1.189  37.537
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         1.51048    1.39411   1.083    0.280
## overall             0.06134    0.03080   1.992    0.048 *
## factor(is_island)TRUE -0.71060    1.07478  -0.661    0.509
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.79 on 177 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.0262, Adjusted R-squared:  0.0152
## F-statistic: 2.382 on 2 and 177 DF,  p-value: 0.09537
plot(data$overall, data$casepc, pch=19, col=as.factor(data$is_island))
```
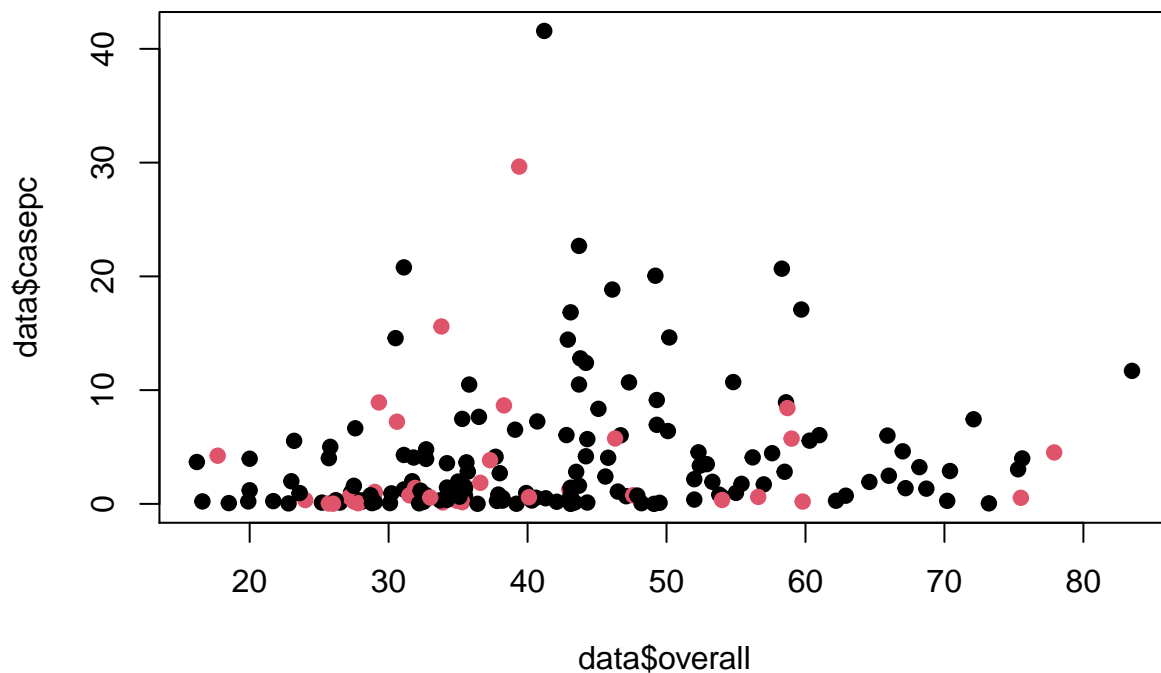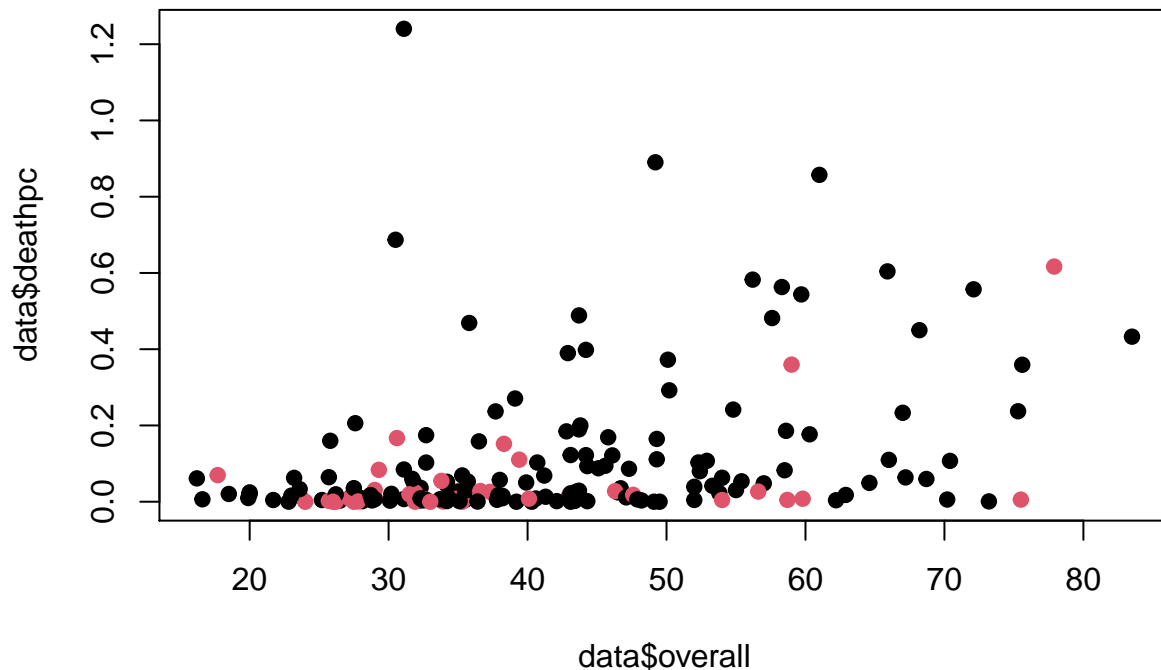
```
summary(lm(formula = deathpc ~ overall, data = data))

##
## Call:
## lm(formula = deathpc ~ overall, data = data)
##
## Residuals:
##       Min       1Q   Median       3Q      Max
## -0.25839 -0.08905 -0.03871  0.01284  1.18090
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.0836539  0.0410562  -2.038   0.0431 *
## overall      0.0046029  0.0009341   4.928 1.89e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1767 on 178 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:   0.12,  Adjusted R-squared:  0.1151
## F-statistic: 24.28 on 1 and 178 DF,  p-value: 1.893e-06
```

```
summary(lm(formula = deathpc ~ overall + factor(is_island), data = data))

##
## Call:
## lm(formula = deathpc ~ overall + factor(is_island), data = data)
##
## Residuals:
##       Min       1Q   Median       3Q      Max
## -0.25795 -0.09006 -0.04047  0.01182  1.16809
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -0.0654370  0.0423515  -1.545    0.124
## overall               0.0044293  0.0009357   4.734  4.5e-06 ***
## factor(is_island)TRUE -0.0534562  0.0326504  -1.637    0.103
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1759 on 177 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.1332, Adjusted R-squared:  0.1234
## F-statistic:  13.6 on 2 and 177 DF,  p-value: 3.213e-06
```

```
plot(data$overall, data$deathpc, pch=19, col=as.factor(data$is_island))
```

```r
summary(lm(formula = cfratio ~ overall, data = data))
```

```
##
## Call:
## lm(formula = cfratio ~ overall, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7871 -1.6501 -0.7537  0.5766 27.3988
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.52444    0.73506   0.713 0.476482
## overall      0.05654    0.01676   3.374 0.000908 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.186 on 179 degrees of freedom
##   (11 observations deleted due to missingness)
## Multiple R-squared:  0.05979,    Adjusted R-squared:  0.05454
## F-statistic: 11.38 on 1 and 179 DF,  p-value: 0.0009083
```

```r
summary(lm(formula = cfratio ~ overall + factor(is_island), data = data))
```

```
##
## Call:
## lm(formula = cfratio ~ overall + factor(is_island), data = data)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4916 -1.6218 -0.7256  0.7636 27.1401
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.83708    0.75774   1.105  0.27078
## overall              0.05363    0.01679   3.195  0.00166 **
## factor(is_island)TRUE -0.93769    0.58829  -1.594  0.11273
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.173 on 178 degrees of freedom
##   (11 observations deleted due to missingness)
## Multiple R-squared:  0.07302,    Adjusted R-squared:  0.06261
## F-statistic: 7.011 on 2 and 178 DF,  p-value: 0.001172
```

```
plot(data$overall, data$cfratio, pch=19, col=as.factor(data$is_island))
```
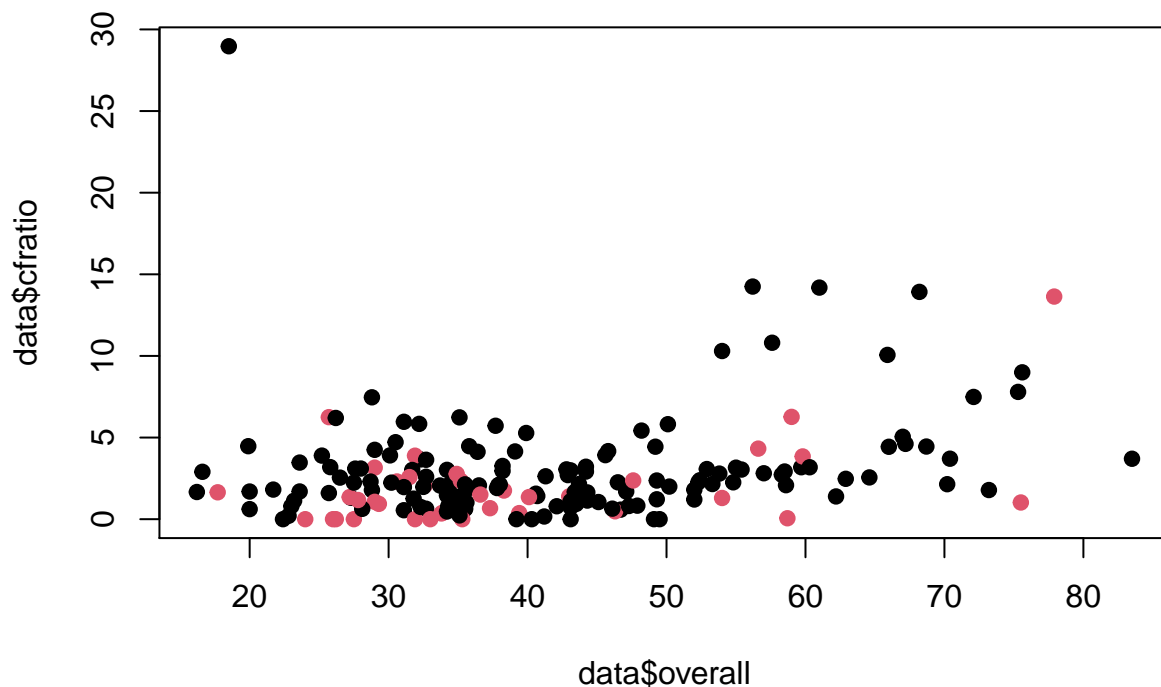

In all three of these cases, adding is_island to the regression does not meaningfully increase the R-squared measure or decrease the residual standard error. So adding island nation status doesn't help explain changes in cases or deaths relative to the GHSI scores.

Regressions on GHSI subcomponent scores and island nation status

```r
summary(lm(formula = casepc ~ prev_emergence_pathogens + early_detection + rapid_response + robust_heal
```

```
##
## Call:
## lm(formula = casepc ~ prev_emergence_pathogens + early_detection +
##     rapid_response + robust_health_sector + commitments + risk_environment,
##     data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.572 -3.179 -1.061  1.273 32.684
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)               3.746745   2.194852   1.707  0.08961 .
## prev_emergence_pathogens  0.009702   0.050357   0.193  0.84744
## early_detection          -0.001960   0.028613  -0.068  0.94547
## rapid_response            0.078963   0.046369   1.703  0.09038 .
## robust_health_sector      0.019410   0.051746   0.375  0.70805
## commitments              -0.174885   0.043168  -4.051 7.68e-05 ***
## risk_environment          0.089156   0.032277   2.762  0.00636 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.41 on 173 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.1691, Adjusted R-squared:  0.1402
## F-statistic: 5.866 on 6 and 173 DF,  p-value: 1.367e-05
```

```r
summary(lm(formula = casepc ~ prev_emergence_pathogens + early_detection + rapid_response + robust_heal
```

```
##
## Call:
## lm(formula = casepc ~ prev_emergence_pathogens + early_detection +
##     rapid_response + robust_health_sector + commitments + risk_environment +
##     factor(is_island), data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.944 -2.919 -1.165  1.179 32.090
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)               3.262258   2.199440   1.483 0.139846
## prev_emergence_pathogens -0.003181   0.050600  -0.063 0.949953
## early_detection          -0.006853   0.028582  -0.240 0.810783
## rapid_response            0.089533   0.046490   1.926 0.055773 .
## robust_health_sector      0.003746   0.052216   0.072 0.942897
## commitments              -0.168681   0.043060  -3.917 0.000129 ***
## risk_environment          0.112129   0.034672   3.234 0.001464 **
## factor(is_island)TRUE    -1.924259   1.100503  -1.749 0.082157 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## 
## Residual standard error: 5.378 on 172 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.1836, Adjusted R-squared:  0.1503
## F-statistic: 5.525 on 7 and 172 DF,  p-value: 9.492e-06
```

In this case adding the is_island variable improves the R-squared by about 10%, but the overall R-squared
and errors are pretty bad. It also should be noted that is_island has a p-value of 0.08 which isn't statistically
significant, but is the 4th most significant of the 7 variables in the regression.