island_confounding

Analysis of island status as a confounding variable

Start by loading the six month data:

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

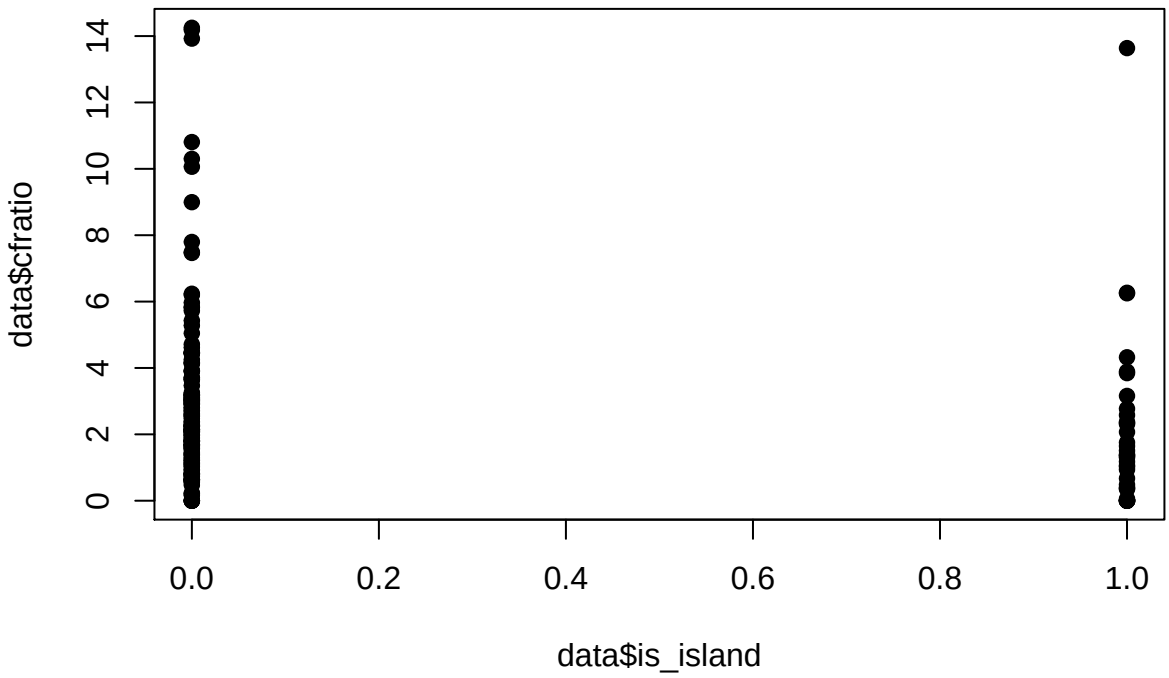
Regressions on island nation status

Run regressions with is_island as the sole explanatory variable:

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

```
##
## Call:
## lm(formula = cfratio ~ factor(is_island), data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9455 -1.5829 -0.6883  0.6541 11.6738
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)             2.9455     0.2196  13.411   <2e-16 ***
## factor(is_island)TRUE  -0.9820     0.4831  -2.033   0.0436 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.617 on 177 degrees of freedom
## Multiple R-squared:  0.02282,    Adjusted R-squared:  0.01729
## F-statistic: 4.133 on 1 and 177 DF,  p-value: 0.04356
```

```
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

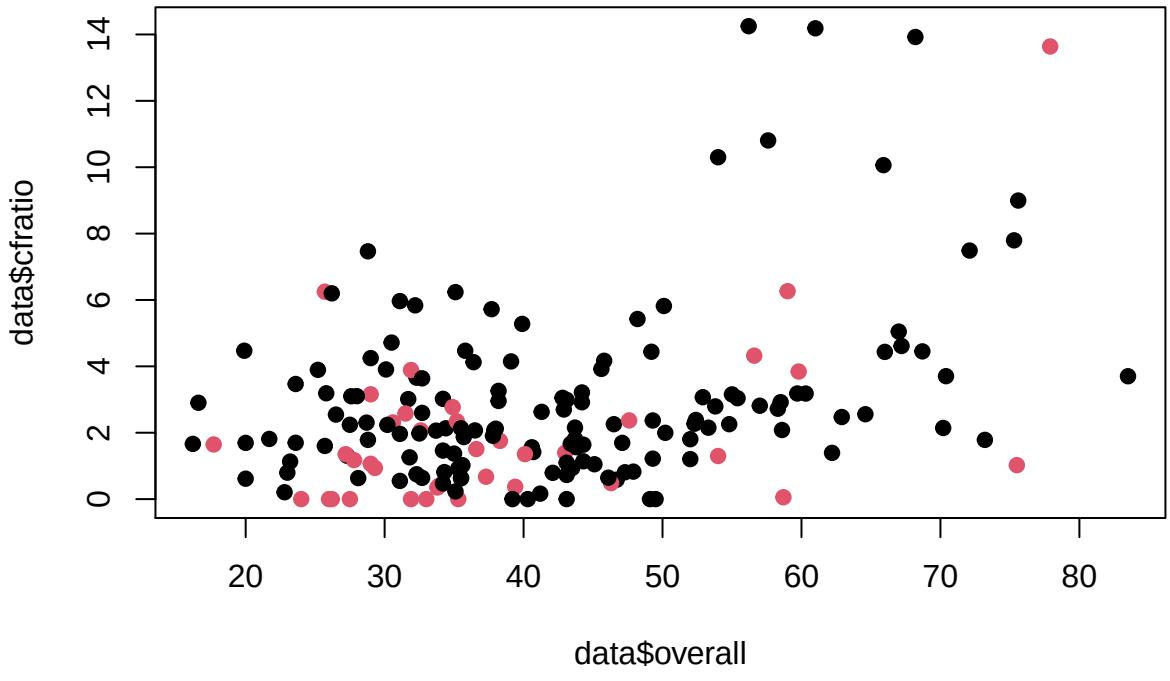
```
summary(lm(formula = cfratio ~ overall, data = data))
```

```
##
## Call:
## lm(formula = cfratio ~ overall, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2044 -1.4438 -0.5437  0.7918 10.4436
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.33252    0.57139  -0.582    0.561
## overall      0.07363    0.01297   5.677 5.51e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.435 on 177 degrees of freedom
## Multiple R-squared:  0.154,  Adjusted R-squared:  0.1493
## F-statistic: 32.23 on 1 and 177 DF,  p-value: 5.51e-08
```

```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
## -3.5788 -1.4916 -0.5250  0.8769 10.3355
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           -0.09161    0.59075  -0.155    0.877
## overall                0.07127    0.01301   5.477 1.48e-07 ***
## factor(is_island)TRUE -0.68816    0.45100  -1.526    0.129
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.426 on 176 degrees of freedom
## Multiple R-squared:  0.1651, Adjusted R-squared:  0.1556
## F-statistic:  17.4 on 2 and 176 DF,  p-value: 1.271e-07
```

```r
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

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

```
##
## Call:
## lm(formula = cfratio ~ prev_emergence_pathogens + early_detection +
##     rapid_response + robust_health_sector + commitments + risk_environment,
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1533 -1.5581 -0.5003  0.8940 10.6471
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               0.3509193  1.0011137   0.351    0.726
## prev_emergence_pathogens  0.0412611  0.0227935   1.810    0.072 .
## early_detection           0.0049748  0.0129809   0.383    0.702
## rapid_response           -0.0007135  0.0210045  -0.034    0.973
## robust_health_sector      0.0206773  0.0234523   0.882    0.379
## commitments               0.0107225  0.0195402   0.549    0.584
## risk_environment         -0.0074782  0.0147077  -0.508    0.612
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.448 on 172 degrees of freedom
## Multiple R-squared:  0.169,  Adjusted R-squared:   0.14
## F-statistic: 5.829 on 6 and 172 DF,  p-value: 1.494e-05
```

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

```
##
## Call:
## lm(formula = cfratio ~ prev_emergence_pathogens + early_detection +
##     rapid_response + robust_health_sector + commitments + risk_environment +
##     factor(is_island), data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7214 -1.5270 -0.5069  0.8730 10.5529
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               0.2157153  1.0082382   0.214    0.831
## prev_emergence_pathogens  0.0376126  0.0230259   1.633    0.104
## early_detection           0.0035716  0.0130376   0.274    0.784
## rapid_response            0.0022688  0.0211708   0.107    0.915
## robust_health_sector      0.0162571  0.0237882   0.683    0.495
## commitments               0.0124866  0.0195966   0.637    0.525
## risk_environment         -0.0009942  0.0158590  -0.063    0.950
## factor(is_island)TRUE    -0.5456180  0.5008412  -1.089    0.278
##
## Residual standard error: 2.447 on 171 degrees of freedom
## Multiple R-squared:  0.1747, Adjusted R-squared:  0.1409
```

```
## F-statistic: 5.171 on 7 and 171 DF,  p-value: 2.323e-05
```

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.