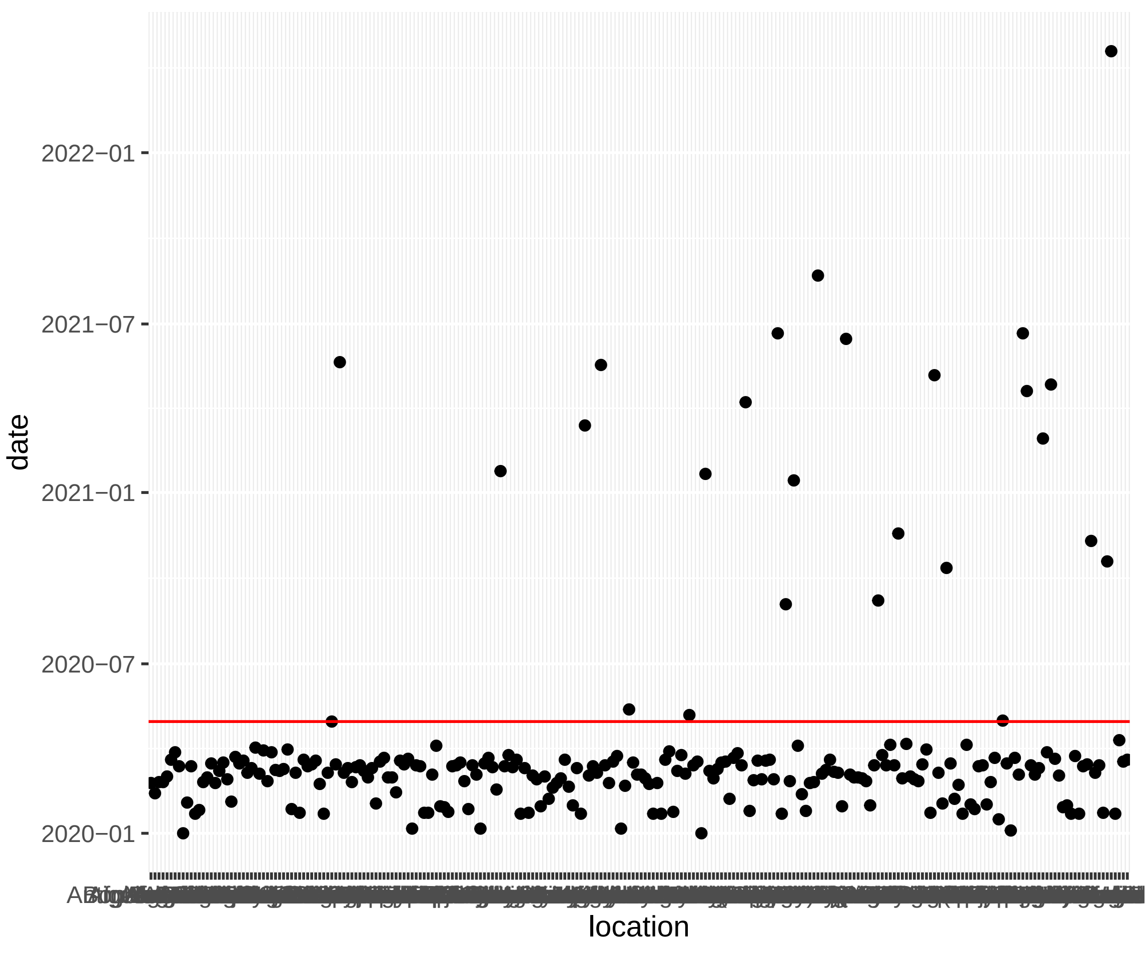
Cross Section Assignment

Jacques Rossouw[a]

[a] *Stellebosch, Western Cape, South Africa*




```
## # A tibble: 206 x 1
## # Groups:   location [206]
##    location
```

Email address: `gerardrossouw@gmail.com` (Jacques Rossouw)

```
##    <chr>
##  1 Afghanistan
##  2 Albania
##  3 Algeria
##  4 Andorra
##  5 Angola
##  6 Anguilla
##  7 Antigua and Barbuda
##  8 Argentina
##  9 Armenia
## 10 Aruba
## # ... with 196 more rows
```

Want to compare the US, with Europe, Africa (excluding SA), and SA (and maybe Asia, or specifically china)

```
## # A tibble: 7 x 1
##   continent
##   <chr>
## 1 Asia
## 2 <NA>
## 3 Europe
## 4 Africa
## 5 North America
## 6 South America
## 7 Oceania
```

[Share of woman vs men](#) in SA about equal

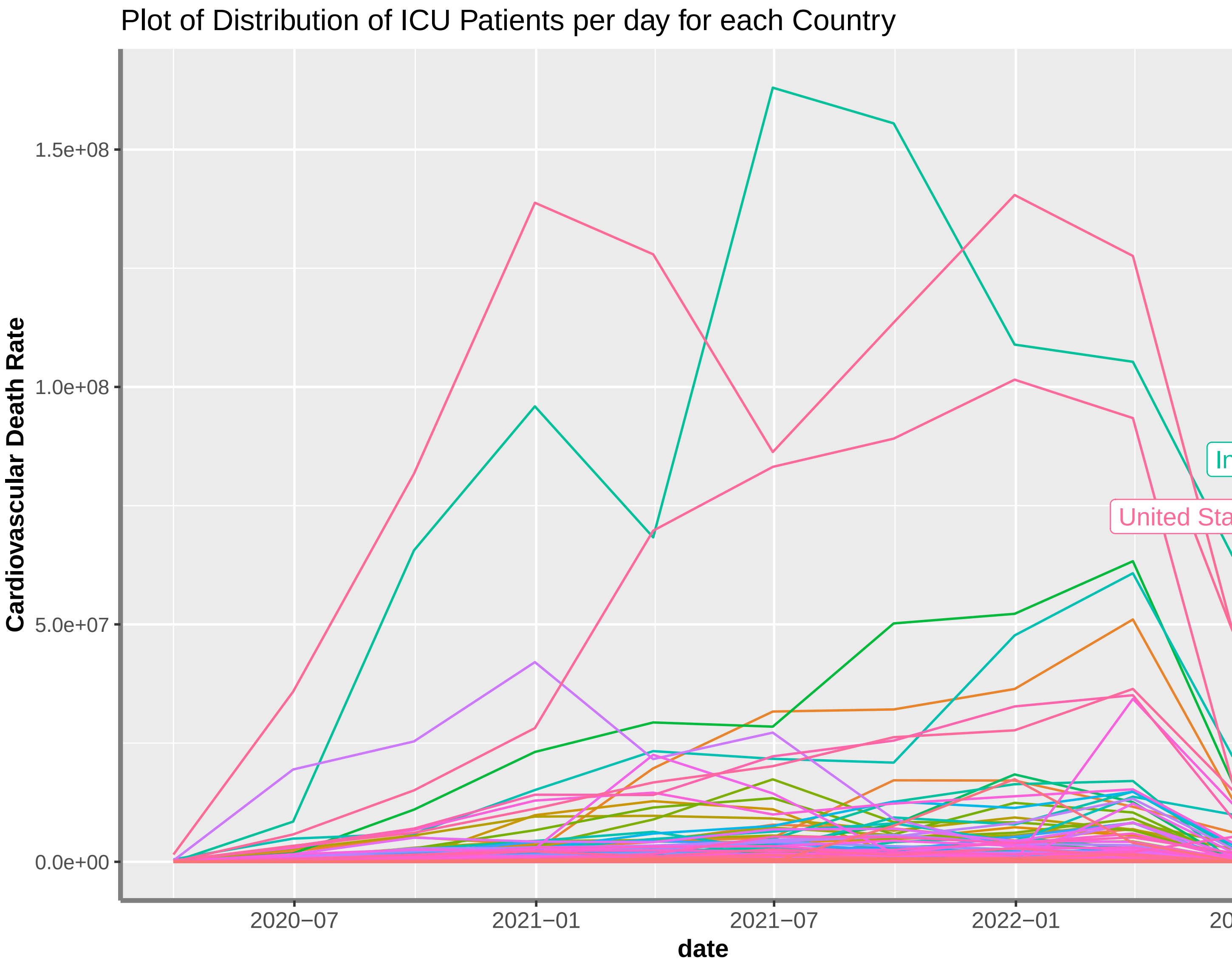
Towards estimating the fixed effects, let's see which variables do not change over the time period we have.

Checking the plots of all the features:

We need to have a look at the correlation between the different variables to see what type of relationships exist. Furthermore, it is important to

```
##                            mean           sd    min           max          range
```

```
## death_rate                1.88        3.09  0.00          37.50          37.50
## reproduction_rate         0.77        0.44 -0.01           2.06           2.08
## icu_patients          10972.02    83404.37  0.00     1605018.00     1605018.00
## hosp_patients         55286.41   368499.78  0.00     6894400.00     6894400.00
## new_tests           2434948.51 11629111.49  0.00   163020639.00   163020639.00
## new_vaccinations    4859361.59 43022342.39  0.00 1124854000.00 1124854000.00
## stringency_index         44.81       25.15  0.00          99.06          99.06
```



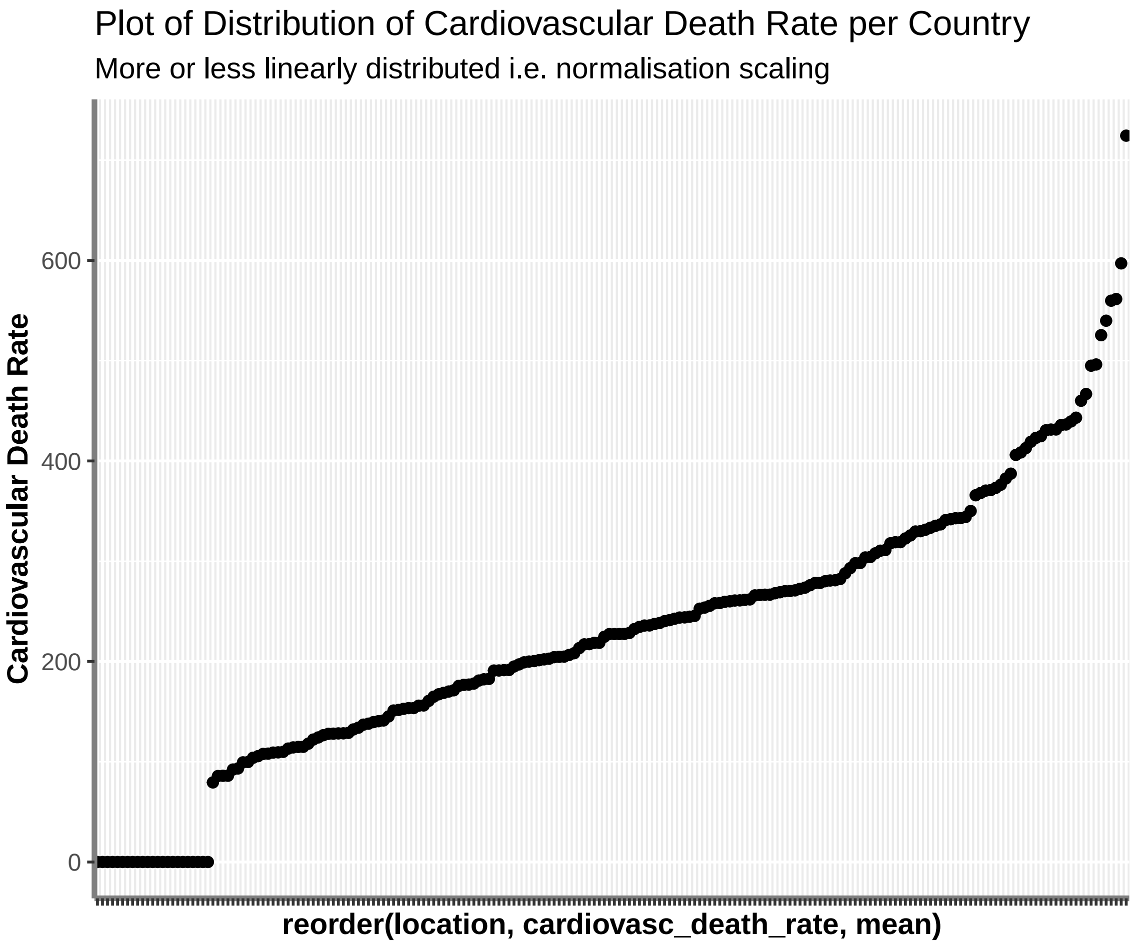
Thus, want to scale: `icu_patients, hosp_patients, new_test, new_vaccinations`

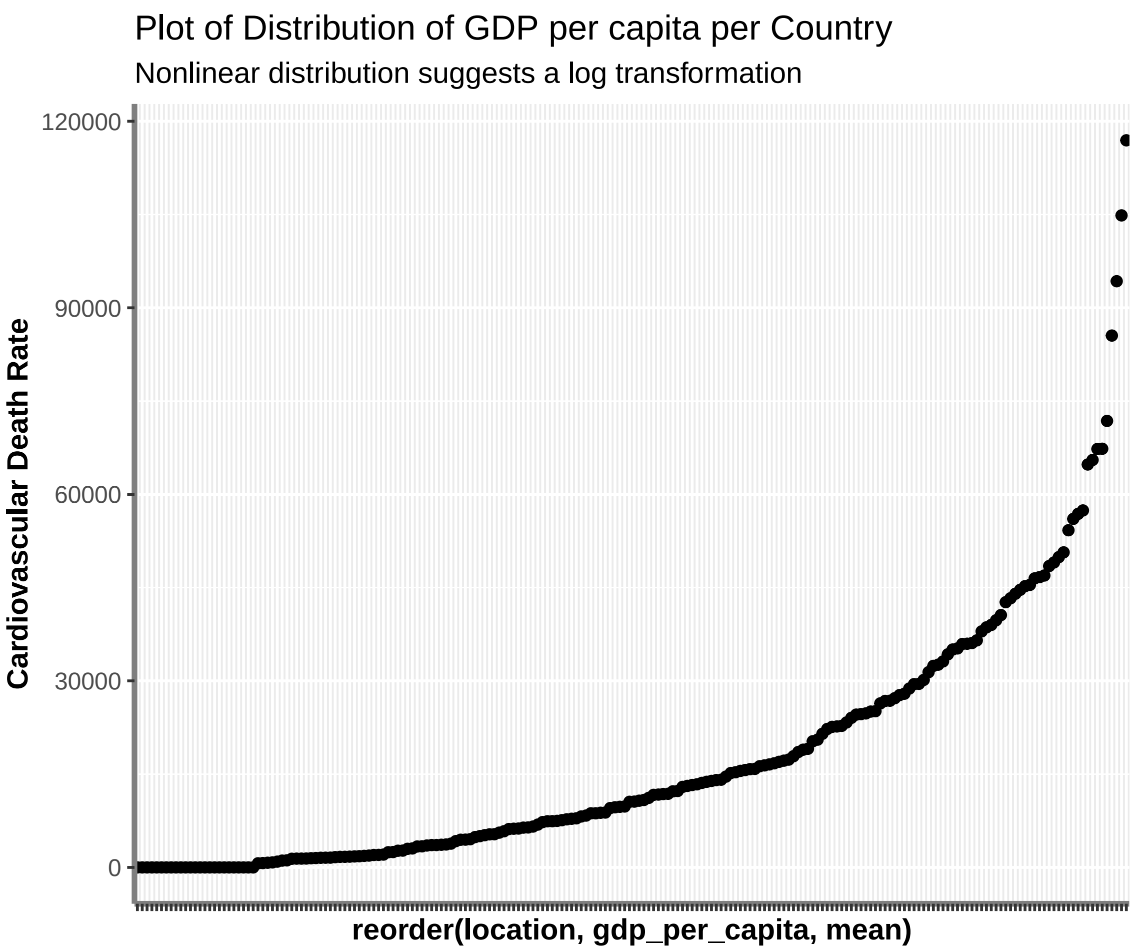
Now, to check the scales of the features that remain constant per country:

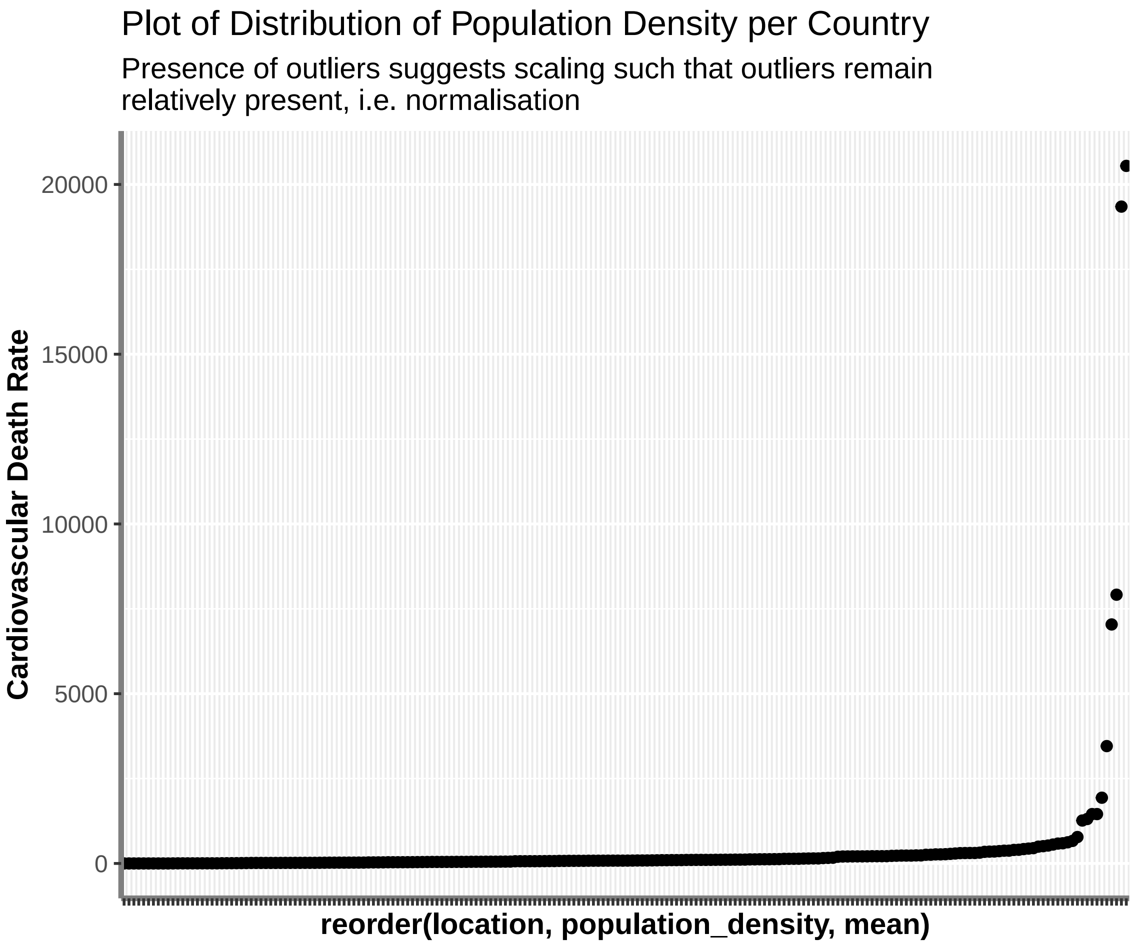
```
##                              mean       sd min       max     range
## gdp_per_capita           17697.35 20539.28   0 116935.60 116935.60
## population_density         444.44  2094.60   0  20546.77  20546.77
## median_age                  27.58    12.79   0     48.20     48.20
## aged_65_older                7.90     6.48   0     27.05     27.05
## extreme_poverty              7.83    16.76   0     77.60     77.60
## cardiovasc_death_rate      226.19   135.19   0    724.42    724.42
## diabetes_prevalence          7.52     4.59   0     23.36     23.36
## handwashing_facilities      21.89    32.72   0     99.00     99.00
## hosp_beds_1k                 2.38     2.51   0     13.80     13.80
## life_expectancy             73.36     9.08   0     86.75     86.75
## human_development_index      0.63     0.28   0      0.96      0.96
## smokers                     14.38    12.75   0     45.95     45.95
```

Additional features that need to be scaled are this

- `gdp_per_capita`
- `population_density`
- `cardiovasc_death_rate`


Plot of Distribution of Cardiovascular Death Rate per Country
More or less linearly distributed i.e. normalisation scaling
Cardiovascular Death Rate
600
400
200
0
reorder(location, cardiovasc_death_rate, mean)


Plot of Distribution of GDP per capita per Country
Nonlinear distribution suggests a log transformation
Cardiovascular Death Rate
120000
90000
60000
30000
0
reorder(location, gdp_per_capita, mean)



Now we can check all the descriptive stats for all the columns

```
##                          mean    sd   min   max range
## death_rate               1.88  3.09  0.00 37.50 37.50
## reproduction_rate        0.77  0.44 -0.01  2.06  2.08
## icu_patients             1.79  4.09  0.00 16.01 16.01
## hosp_patients            2.13  4.75  0.00 17.44 17.44
## new_tests                8.07  7.16  0.00 20.62 20.62
## new_vaccinations         6.20  7.11  0.00 21.92 21.92
## stringency_index        44.81 25.15  0.00 99.06 99.06
## population_density       0.00  1.00 -0.21  9.60  9.81
## median_age              27.58 12.79  0.00 48.20 48.20
## aged_65_older            7.90  6.48  0.00 27.05 27.05
## gdp_per_capita           8.21  3.22  0.00 11.67 11.67
## extreme_poverty          7.83 16.76  0.00 77.60 77.60
```

```
## cardiovasc_death_rate    0.00  1.00 -1.67  3.69  5.36
## diabetes_prevalence      7.52  4.59  0.00 23.36 23.36
## handwashing_facilities  21.89 32.72  0.00 99.00 99.00
## hosp_beds_1k             2.38  2.51  0.00 13.80 13.80
## life_expectancy         73.36  9.08  0.00 86.75 86.75
## human_development_index  0.63  0.28  0.00  0.96  0.96
## smokers                 14.38 12.75  0.00 45.95 45.95
```

```
## Pooling Model
##
## Call:
## plm(formula = world_df$death_rate ~ world_df$stringency_index +
##     world_df$aged_65_older + world_df$population_density + world_df$gdp_per_capita +
##     world_df$extreme_poverty + world_df$diabetes_prevalence +
##     world_df$smokers + world_df$handwashing_facilities + world_df$life_expectancy +
##     world_df$human_development_index, data = world_df, model = "pooling",
##     index = c("location", "date"))
##
## Unbalanced Panel: n = 206, T = 7-10, N = 2050
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -3.34154 -1.36040 -0.81011  0.26304 35.51716
##
## Coefficients:
##                                    Estimate Std. Error t-value  Pr(>|t|)
## (Intercept)                       3.1144564  0.7446268  4.1826 3.004e-05 ***
## world_df$stringency_index         0.0140255  0.0029613  4.7362 2.327e-06 ***
## world_df$aged_65_older            0.0200532  0.0153248  1.3085   0.19084
## world_df$population_density      -0.1279423  0.0728308 -1.7567   0.07912 .
## world_df$gdp_per_capita          -0.0263453  0.0330498 -0.7971   0.42546
## world_df$extreme_poverty         -0.0025897  0.0047104 -0.5498   0.58253
## world_df$diabetes_prevalence     -0.0197411  0.0160446 -1.2304   0.21869
## world_df$smokers                 -0.0082696  0.0070628 -1.1709   0.24179
## world_df$handwashing_facilities   0.0098525  0.0021872  4.5046 7.026e-06 ***
## world_df$life_expectancy         -0.0261921  0.0098585 -2.6568   0.00795 **
## world_df$human_development_index  0.2977904  0.4255525  0.6998   0.48415
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    19611
## Residual Sum of Squares: 18889
## R-Squared:      0.036848
## Adj. R-Squared: 0.032124
## F-statistic: 7.80073 on 10 and 2039 DF, p-value: 2.2171e-12
```