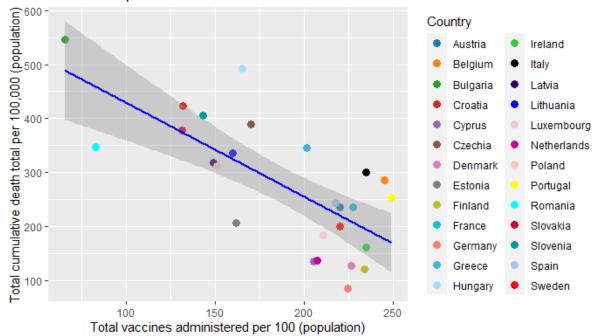## Relationship between number of vaccine doses administered and death


```
Call:

lm(formula = Deaths...cumulative.total.per.100.000.population ~

    Total.vaccine.doses.administered.per.100.population, data =
dataframe)

Residuals:

     Min        1Q    Median        3Q       Max

-126.713   -69.516     7.429    54.965   176.449

Coefficients:

                                                    Estimate Std. Error
t value Pr(>|t|)

(Intercept)                                         603.4869    65.5729
9.203 2.42e-09 ***

Total.vaccine.doses.administered.per.100.population  -1.7389     0.3386
-5.136 2.95e-05 ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 84.08 on 24 degrees of freedom

Multiple R-squared:  0.5236,  Adjusted R-squared:  0.5037

F-statistic: 26.38 on 1 and 24 DF,  p-value: 2.948e-05
```

The residuals in the linear regression model have a range of values from -126.713 to 176.449, which indicates that some of the model's predictions were far off from the actual data. The mean of the residuals is 1.78, which means that the model's predictions differ from the actual data by 1.78 units, indicating some variability not accounted for by the predictor variable.

The y-intercept of the coefficient is statistically irrelevant as it represents a predicted value of the response variable for a hypothetical scenario that does not make sense in the context of the model. The coefficient for "Total vaccine doses administered per 100 population" is -1.7389, indicating that for every unit increase in the total number of vaccine doses administered per 100 population, the cumulative total deaths per 100,000 population is expected to decrease by 1.7389. The p-value associated with the coefficient is <0.001, indicating strong evidence of a relationship between the independent and dependent variables.

The residual standard error (RSE) is 34.99, representing the average distance between observed and predicted values of the dependent variable. A lower RSE suggests a better predictive power of the model. The F-statistic tests the significance of the model, and the p-value is <0.001, indicating strong evidence of a relationship between the variables. The R-squared value of 0.5236 indicates that 52.36% of the variability in the dependent variable is explained by the independent variable.

Overall, the results suggest that increasing vaccine coverage could be an effective strategy in reducing COVID-19 deaths. However, the high value of the RSE indicates that other variables besides the vaccine may have influenced the number of deaths, so correlation does not necessarily imply causation.

Low vaccination rate	Deaths - cumulative total per 100,000 population	High vaccination rate	Deaths - cumulative total per 100,000 population
Bulgaria	545.73	Cyprus	135.13
Romania	347.75	Netherlands	137.31
Slovakia	378.37	Luxembourg	183.83
Croatia	423.84	Spain	243.47
Slovenia	405.04	Austria	236.48
Latvia	318.19	Sweden	200.95
Poland	311.38	Germany	86.39
Lithuania	336.57	Denmark	127.43
Estonia	207.45	France	236.71
Hungary	492.46	Finland	122
Czechia	389.77	Ireland	162.48
Greece	346.16	Italy	300.86
		Belgium	286.04
		Portugal	254.28

```
> mean(Low_vaccine_rate)

[1] 375.2258

> sd(Low_vaccine_rate)

[1] 87.83315
```
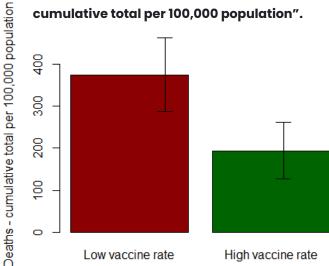
```
> mean(High_vaccine_rate)

[1] 193.8114

> sd(High_vaccine_rate)

[1] 67.15065
```

Mean with Standard Deviation of "Deaths – cumulative total per 100,000 population".

```
ANOVA of "Deaths - cumulative total per 100,000 population" between groups.

Df Sum Sq Mean Sq F value   Pr(>F)
group        1 212657  212657   35.57 3.72e-06 ***
Residuals   24 143481    5978
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The F-statistic for the group variable is 35.57, it has a small p-value of 3.72e-06, suggesting a significant difference in the means of the deaths between low and high vaccine groups.

The p-value is less than 0.05, suggesting that there is strong evidence to reject the null hypothesis that there is no difference in vaccine rates between the two groups.

This means that there is a statistically significant difference in the mean cumulative deaths per 100,000 population between the low and high vaccine rate groups.

The mean sum of squares for Residuals is much smaller than that of group, indicating that most of the variation in the model is explained by the group variable.

This model does not say specifically where the significance lies.

Overall, these results suggest that there is a significant difference in vaccine rates between the two groups of countries with low and high vaccine rates, highlighting the need to address the low vaccine rates in certain countries for public health and policy reasons.

Moreover, the conclusions of these data sets would demonstrate there is a need for vaccination. As the ANOVA shows significance in difference between the groups: high vaccination and low vaccination. Then the bar graph comparing means and standard deviation between deaths of low and high vaccination shows that even when accounting for standard deviation, the high vaccination countries had a lot less deaths. This is also supported by the linear regression model which shows a negative correlation between deaths and vaccination rates. These tests show when there is a relationship between number of vaccinations and the number of deaths. So, it would be up to researcher to demonstrate causality between the variables via other methods.