The Impact of Inequality on Global Populations during the COVID-19 Pandemic

Connor Brown

April 20, 2021

Abstract

This paper uses a global COVID-19 response survey in association with international inequality statistics to investigate how distribution of wealth has impacted the globe's reaction to a distressful period of loss, confusion and fear. Multivariate regressions between data collected from March 30 to May 20 of 2020

Introduction

The Coronavirus pandemic that traumatized the planet has had an immense effect on individuals and groups alike. Despite its havoc, worldwide issues, such as income inequality, still play a major role in the day-to-day lives of every nation's population.**

Methodology | Hypothesis

My hypothesis is that there will be a negative relationship between a nation's income and its population's risk of infection, and a positive relationship between income and its population's proportions of isolation and medical facilitation.

Data

Data for this analysis are taken from two sources: The World Bank's indicators database, and the 2020 COVIDiSTRESS Global Survey from the COVIDiSTRESS consortium. Indicators from the World Bank each have their own spreadsheets, while the COVIDiSTRESS survey provided proportional data for each country in tabular format. Given the difficulty in acquiring inequality statistics for 2020, I am using GINI indices and GDP per Capita measurements from 2017. In addition, given the smaller number of respondent countries to the COVIDiSTRESS survey, only 31 countries total have valid data associated with these variables.

This analysis uses the following variables:

GDP is GDP per Capita at the current value of USD. Data was for 197 countries, only 30 of which are used for this analysis. World Bank describes GDP as "the sum of gross value added by all resident producers in the economy, plus any product taxes and minus any subsidies not included in the value of the products." It is a conversion of domestic currencies to a value in current U.S. dollars. It is sourced from the World Bank's country departments and government statistical agencies. For ease of reading, the values have been reduced by a factor of one hundred million.

GINI is the Gini index of each country. Once again, data was for 197 countries, only 30 of which are used for this analysis. World Bank describes the Gini index as a measure of "the extent to which the distribution (or, in some cases, consumption expenditure) among individuals or households within an economy deviates from a perfectly equal distribution…0 represents perfect equality, while an index of 100 implies perfect inequality." It is also sourced from the World Bank's country departments and government statistical agencies.

PROP_Risk is the proportion of a country's respondents to the COVIDiSTRESS survey that responded "yes" to the question of whether or not they or their family members were at high risk. 39 countries responded with a total of 173,426 participants; only 30 countries' data was used in this analysis. This information is validated through several methods, including studies of extraversion, neuroticism, openness, etc.

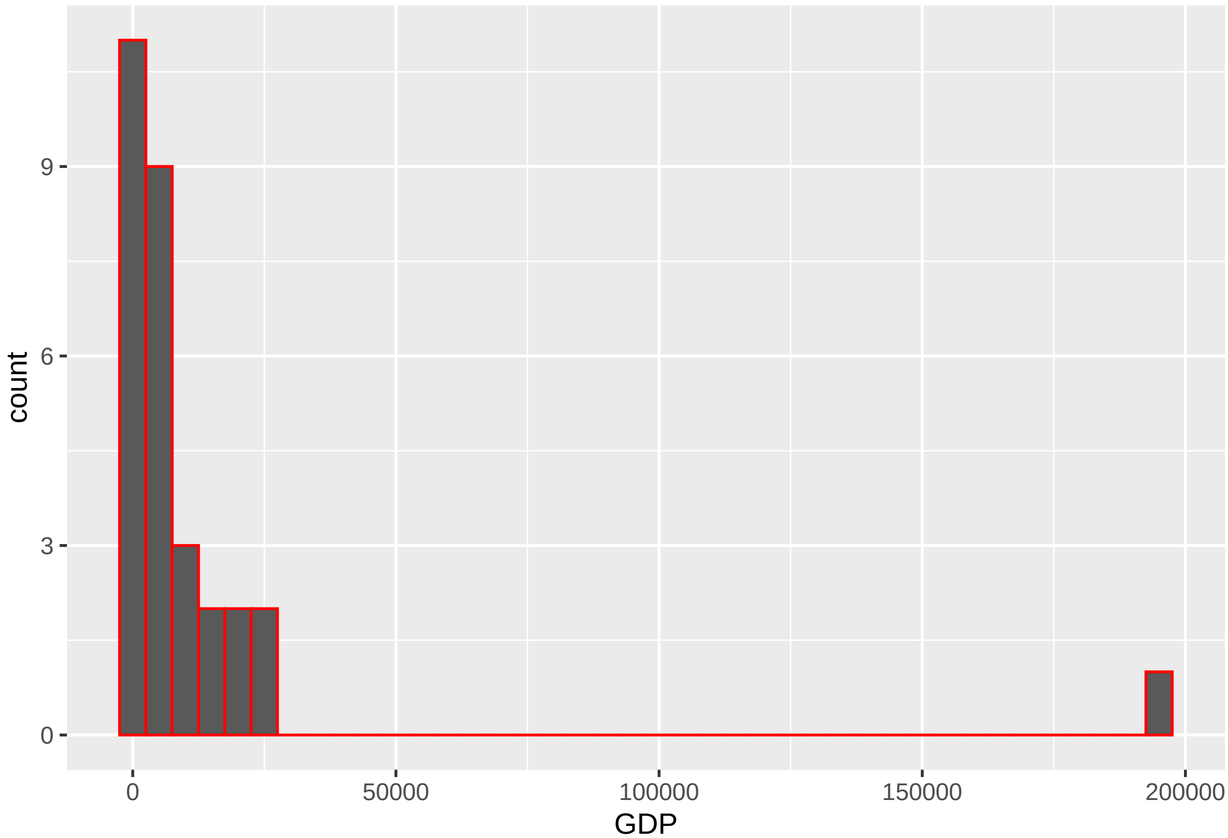
PROP_Medical is the proportion of a country's respondents to the survey that responded to the question of current isolation status as being "isolated in a medical facility or similar location."

PROP_Isolated is the proportion of a country's respondents to the survey that responded to the aforementioned question of current isolated status as being "isolated".

Listed below are summaries and statistics of these variables.

```
##       COUNTRY        GDP                 GINI          PROP_Risk
##  Argentina: 1   Min.   :    72.46   Min.   :24.20   Min.   :0.4210
##  Austria  : 1   1st Qu.:  2056.15   1st Qu.:29.02   1st Qu.:0.6385
##  Belgium  : 1   Median :  4599.68   Median :34.00   Median :0.6925
##  Brazil   : 1   Mean   : 13328.13   Mean   :34.84   Mean   :0.6932
##  Bulgaria : 1   3rd Qu.:  9764.61   3rd Qu.:39.77   3rd Qu.:0.7552
##  Canada   : 1   Max.   :195193.54   Max.   :53.30   Max.   :0.9000
##  (Other)  :24
##   PROP_Medical       PROP_Isolated
##  Min.   :0.000000   Min.   :0.1870
##  1st Qu.:0.000250   1st Qu.:0.3180
##  Median :0.001000   Median :0.4105
##  Mean   :0.001533   Mean   :0.4388
##  3rd Qu.:0.002000   3rd Qu.:0.5773
##  Max.   :0.007000   Max.   :0.6990
##
```

Figure 1 [left] | Figure 2 [right]



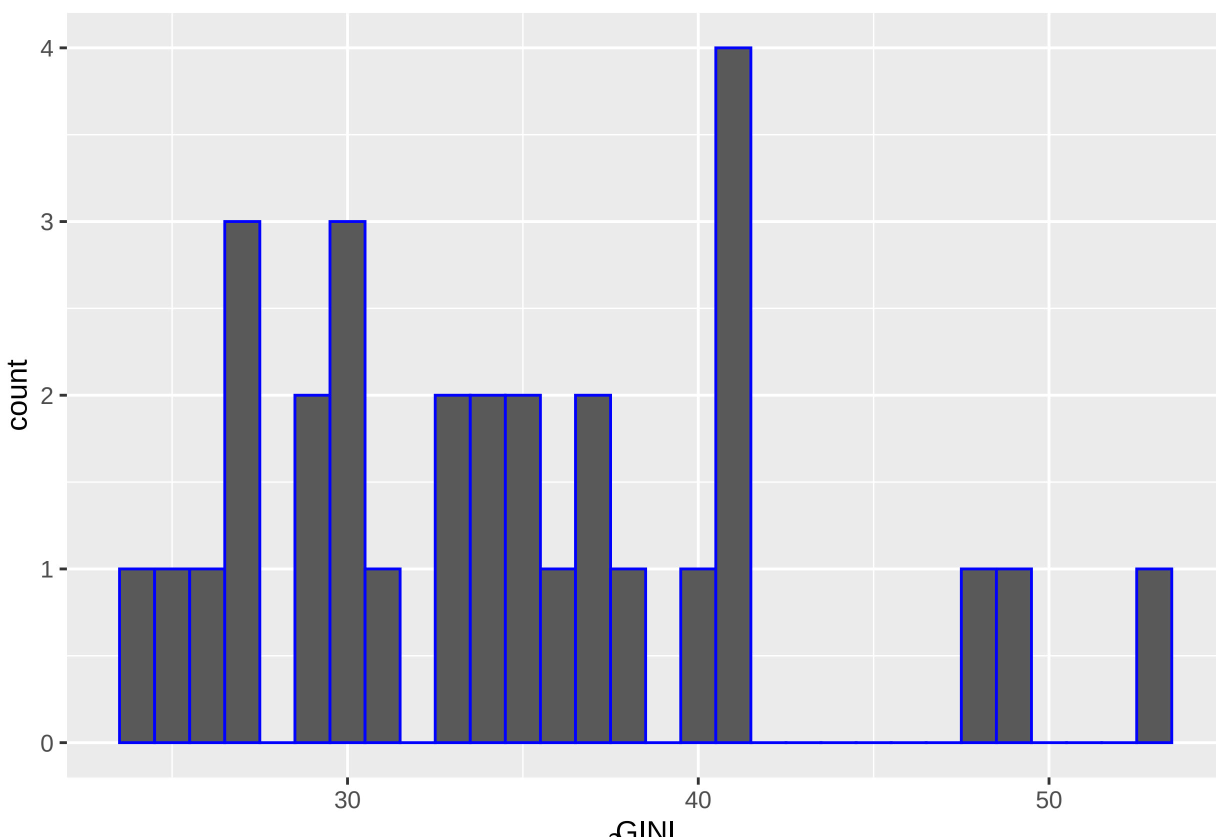


Figure 1 shows the distribution of GDP, showing the major differences between the countries analyzed. Figure 2 shows the distribution of the GINI index, showing a non-normal distribution for the variable. While Figure 2 is left skewed, showing a stronger distribution towards less inequality in some countries, Figure 1 shows a potential limitation: an extreme outlier. This just so happens to be the United States, which has a far higher national income than the other 30 countries; this is a problem that is addressed within my results.

Results

Table 1, $PROP\_Risk = \beta_0 + \beta_1 GDP + \epsilon_i$

term	estimate	std.error	statistic	p.value
(Intercept)	0.6885463	0.0181990	37.8343977	0.000000
GDP	0.0000003	0.0000005	0.7061133	0.485952

```
##
## Call:
## lm(formula = PROP_Risk ~ GDP, data = data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.267571 -0.051943 -0.004154  0.065006  0.204303
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.885e-01  1.820e-02  37.834   <2e-16 ***
## GDP         3.467e-07  4.909e-07   0.706    0.486
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09301 on 28 degrees of freedom
## Multiple R-squared:  0.0175, Adjusted R-squared:  -0.01759
## F-statistic: 0.4986 on 1 and 28 DF,  p-value: 0.486
```

With a sample size of just 30 countries, the data at hand is quite picky and will be difficult to model. As GDP per capita in a nation increases by 1 dollar, there is an associated increase by .00000035 in the proportion of the population that is at risk. However, the very large p-value is far from statistically significant, with a standard error of .0000005. To justify this, the fact that this model accounts for a miniscule 1.8% of the variance says a lot about its applicability.

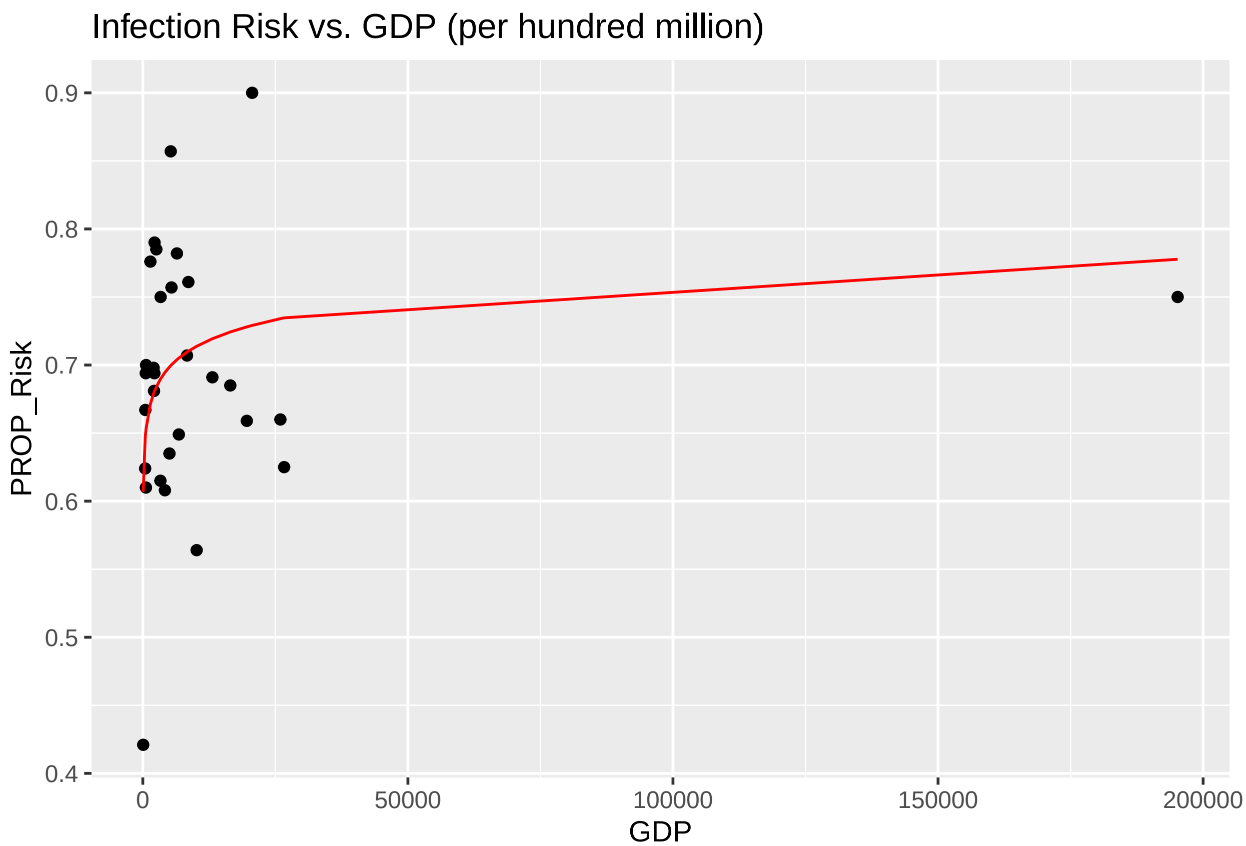
Table 2, $PROP\_Risk = \beta_0 + \beta_1 log(GDP) + \epsilon_i$

term	estimate	std.error	statistic	p.value
(Intercept)	0.5142674	0.0858479	5.990445	0.0000019
log(GDP)	0.0216272	0.0101986	2.120611	0.0429458

```
##
## Call:
## lm(formula = PROP_Risk ~ log(GDP), data = data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.185896 -0.061720 -0.000707  0.055325  0.170879
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.51427    0.08585   5.990 1.88e-06 ***
## log(GDP)     0.02163    0.01020   2.121   0.0429 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0871 on 28 degrees of freedom
## Multiple R-squared:  0.1384, Adjusted R-squared:  0.1076
## F-statistic: 4.497 on 1 and 28 DF,  p-value: 0.04295
```

Logging GDP has a strong effect on the model, indicating that a 1% increase in GDP is associated with a .022 increase in the percentage of those who have relatives at risk or risk themselves, with a standard error of ~.01. This model accounts for 13.8% of the variance; still not a great model, but it is definitely improved.

Figure 3



When graphed, this model quickly shows that the aforementioned outlier skews any model, both visually and statistically.

Table 3, $PROP\_Risk = \beta_0 + \beta_1 log(GDP) + \epsilon_i$ without United States

term	estimate	std.error	statistic	p.value
(Intercept)	0.4992527	0.0966704	5.164482	0.0000196
log(GDP)	0.0235898	0.0117072	2.014979	0.0539637

```
##
## Call:
## lm(formula = PROP_Risk ~ log(GDP), data = data_NoUS)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.179288 -0.065299  0.001102  0.054968  0.166396
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.49925    0.09667   5.164 1.96e-05 ***
```

```
## log(GDP)     0.02359    0.01171   2.015    0.054 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08849 on 27 degrees of freedom
## Multiple R-squared:  0.1307, Adjusted R-squared:  0.09852
## F-statistic:  4.06 on 1 and 27 DF,  p-value: 0.05396
```

Out of curiosity, I looked to see if voiding the United States from the dataset would allow a more defined relationship to appear; conversely, it actually lowered the accountability of variation as well as raised the p-value. It is clear that the data is easier to explain with the outlier, and further work on the model can be done.

Table 4, $PROP\_Risk = \beta_0 + \beta_1 log(GDP) + \beta_2 GINI + \epsilon_i$

term	estimate	std.error	statistic	p.value
(Intercept)	0.3107860	0.0918087	3.385149	0.0021918
log(GDP)	0.0182375	0.0086091	2.118396	0.0434895
GINI	0.0066452	0.0018592	3.574334	0.0013486

```
##
## Call:
## lm(formula = PROP_Risk ~ log(GDP) + GINI, data = data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.150642 -0.040539 -0.008726  0.040527  0.165330
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.310786   0.091809   3.385  0.00219 **
## log(GDP)    0.018238   0.008609   2.118  0.04349 *
## GINI        0.006645   0.001859   3.574  0.00135 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.07308 on 27 degrees of freedom
## Multiple R-squared:  0.4151, Adjusted R-squared:  0.3718
## F-statistic: 9.582 on 2 and 27 DF,  p-value: 0.0007166
```

Table 4 shows a much more moderately useful model, which incorporates Gini index as a second factor into our predictor. By including this index, I've reduced the p-value to a much lower and far more significant value of ~.0007, and can use this model to account for more than 41% of the variance. This model shows a positive relationship between income inequality and the proportion of those at risk of infection. For every 1% increase in GDP, proportion is expected to increased by .018, while the a 1 unit increase in the Gini index increase infection risk by ~.007. Both of these coefficients are statistically significant, indicating that as income inequality increases (richer countries tending to have more), citizen's general risk for themselves and their families to infection of COVID-19 also increases.

Table 5, $PROP_Medical = \beta_0 + \beta_1 log(GDP) + \beta_2 GINI + \epsilon_i$

term	estimate	std.error	statistic	p.value
(Intercept)	0.0014877	0.0021314	0.6980054	0.4911419
log(GDP)	-0.0002293	0.0001999	-1.1470792	0.2614124
GINI	0.0000557	0.0000432	1.2914655	0.2074895

```
##
## Call:
## lm(formula = PROP_Medical ~ log(GDP) + GINI, data = data)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0019007 -0.0012122 -0.0003356  0.0008021  0.0049613
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.488e-03  2.131e-03   0.698    0.491
## log(GDP)    -2.293e-04  1.999e-04  -1.147    0.261
## GINI         5.574e-05  4.316e-05   1.291    0.207
##
## Residual standard error: 0.001697 on 27 degrees of freedom
## Multiple R-squared:  0.0906, Adjusted R-squared:  0.02324
## F-statistic: 1.345 on 2 and 27 DF,  p-value: 0.2774
```

Table 5 shows an attempt at exploring another aspect of the survey: the proportions of nation's populations that are medically isolated, as explained by income inequality. This model indicates that for every increase in 1% of GDP, the proportion is expected to decrease by .00023, and for every 1 unit increase in the Gini index, it will increase by .000056. This model only accounts for a little over 9% of the variance. In this specific instance, with the data I have, it is clear that the variance in the proportion's cannot be explained in a statistically significant fashion.

Table 6, $PROP\_Isolated = \beta_0 + \beta_1 log(GDP) + \beta_2 GINI + \epsilon_i$

term	estimate	std.error	statistic	p.value
(Intercept)	-0.8047193	0.4610879	-1.745262	0.0923122
log(GDP)	0.0218343	0.0166433	1.311898	0.2005999
log(GINI)	0.3010889	0.1295547	2.324030	0.0278929

```
##
## Call:
## lm(formula = PROP_Isolated ~ log(GDP) + log(GINI), data = data)
##
```

```
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.226442 -0.117970  0.009513  0.112067  0.238338
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.80472    0.46109  -1.745   0.0923 .
## log(GDP)     0.02183    0.01664   1.312   0.2006
## log(GINI)    0.30109    0.12955   2.324   0.0279 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.141 on 27 degrees of freedom
## Multiple R-squared:  0.2294, Adjusted R-squared:  0.1724
## F-statistic:  4.02 on 2 and 27 DF,  p-value: 0.02964
```

Finally, Table 6 shows another attempt at exploring further. I tried to use these statistics of income inequality to model the proportion of a population that was isolated; the best model of which ended up being that of logging both GDP and the Gini index. The model indicates that for every increase in 1% in both GDP and the Gini index, the proportion will increase by ~.32, with a standard error of ~.15. Despite this, the model itself is barely statistically significant, and only accounts for ~22.9% of the variance in the data.

Discussion | Conclusion

This analysis attempted to cover three different major areas within COVID reactions. To start, I hypothesized that

References

World Bank, World Development Indicators. (2014). "GDP per capita (current US$)" [Data File]. Retrieved from https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?view=chart

World Bank, World Development Indicators. (2014). "GINI index (World Bank estimate)" [Data File]. Retrieved from https://data.worldbank.org/indicator/SI.POV.GINI

Yamada, Y., Ćepulić, DB., Coll-Martín, T. et al. COVIDiSTRESS Global Survey dataset on psychological and behavioural consequences of the COVID-19 outbreak. Sci Data 8, 3 (2021). https://doi.org/10.1038/s41597-020-00784-9