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

Connor Brown

April 28, 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 * addmore

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.* background knowledge here

Methodology | Hypothesis

My hypothesis is that there will be a positive relationship between a nation's income and its population's risk of infection and between income and the population's proportions of both isolation and medical facilitation (separately). *explain why Using multiple regressions, both univariate and bivariate, I will explore these aspects of the pandemic and try to understand their association with income inequality. The independent variables; GDP, GDPC, and GINI, as described further below; will be used as predictors for three specific areas of interest: PROP_Risk, PROP_Medical, and PROP_Isolated. As hypothesized, these independent variables, in conjunction, should produce results of positive associations.

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 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.

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.

POP is the total population for each country, as of 2017. Of the 197 countries, only 30 are used. World Bank sourced this information from UN Population Division, including census reports and information from the US Census Bureau. The population is used to calculate GDP per Capita

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".

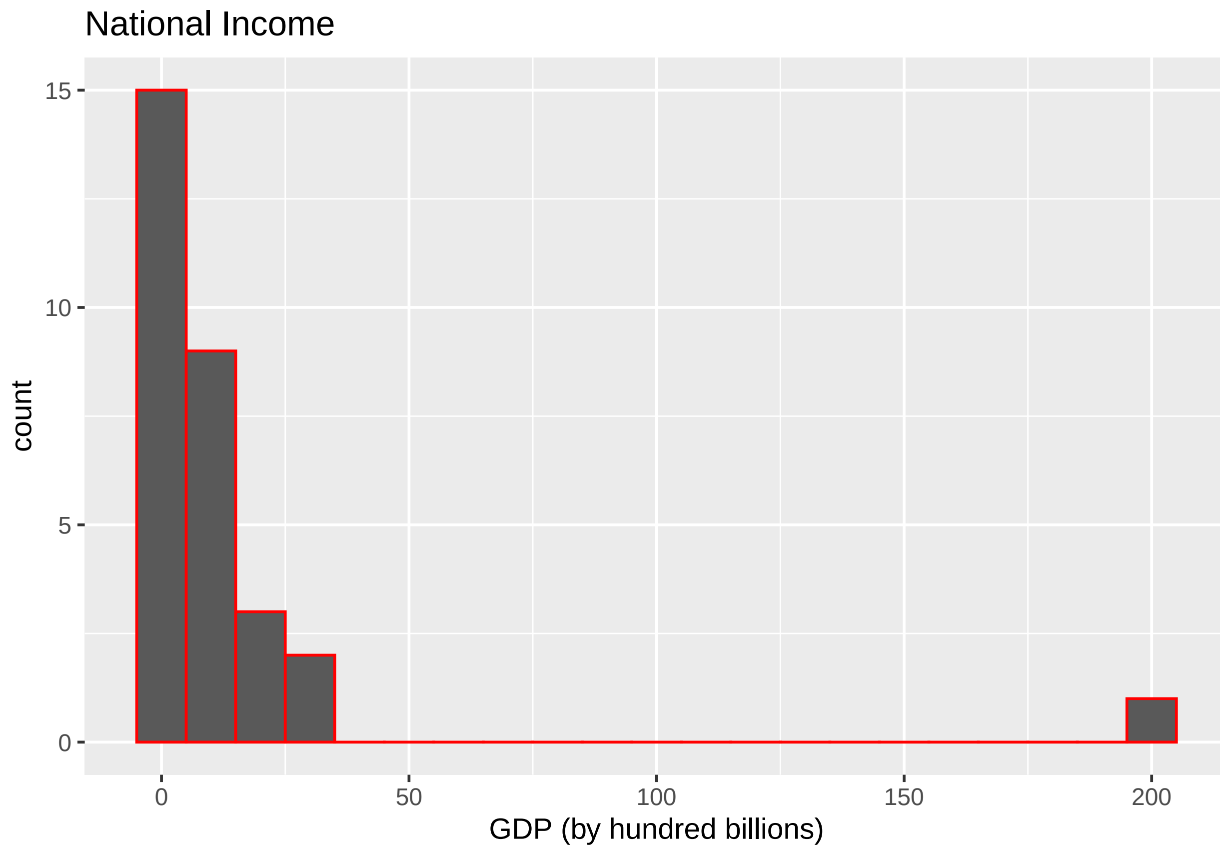
GDPC is the GDP per Capita, calculated as the GDP column divided by the Population column. This is used as a more normalized representation of household income by country.

Listed below are the summaries and statistics of these variables.

```
##       COUNTRY         GDP                   GINI             POP
##  Argentina: 1   Min.   :7.246e+09   Min.   :24.20   Min.   :  1791003
##  Austria  : 1   1st Qu.:2.056e+11   1st Qu.:29.02   1st Qu.:  7034630
##  Belgium  : 1   Median :4.600e+11   Median :34.00   Median : 10674558
##  Brazil   : 1   Mean   :1.333e+12   Mean   :34.84   Mean   : 46310040
##  Bulgaria : 1   3rd Qu.:9.765e+11   3rd Qu.:39.77   3rd Qu.: 44438634
##  Canada   : 1   Max.   :1.952e+13   Max.   :53.30   Max.   :324985539
##  (Other)  :24
##    PROP_Risk       PROP_Medical      PROP_Isolated         GDPC
##  Min.   :0.4210   Min.   :0.000000   Min.   :0.1870   Min.   : 3838
##  1st Qu.:0.6385   1st Qu.:0.000250   1st Qu.:0.3180   1st Qu.:13555
##  Median :0.6925   Median :0.001000   Median :0.4105   Median :21063
##  Mean   :0.6932   Mean   :0.001533   Mean   :0.4388   Mean   :29995
##  3rd Qu.:0.7552   3rd Qu.:0.002000   3rd Qu.:0.5773   3rd Qu.:46040
```

```
##  Max.   :0.9000   Max.   :0.007000   Max.   :0.6990   Max.   :80450
##
```

Figures 1 & 2


National Income
15
10
5
0
count
0
50
100
150
200
GDP (by hundred billions)

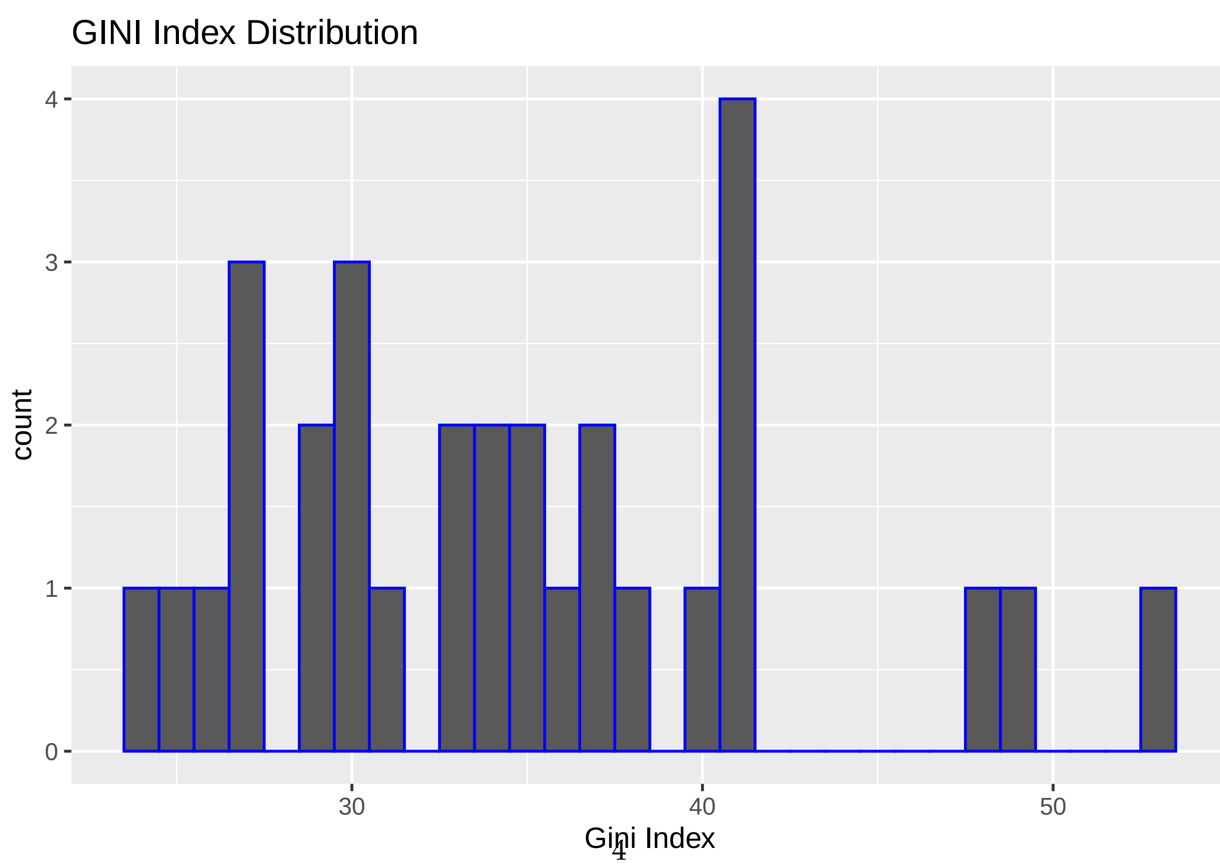

GINI Index Distribution
4
3
2
1
0
count
30
40
50
Gini Index

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.

Figure 3

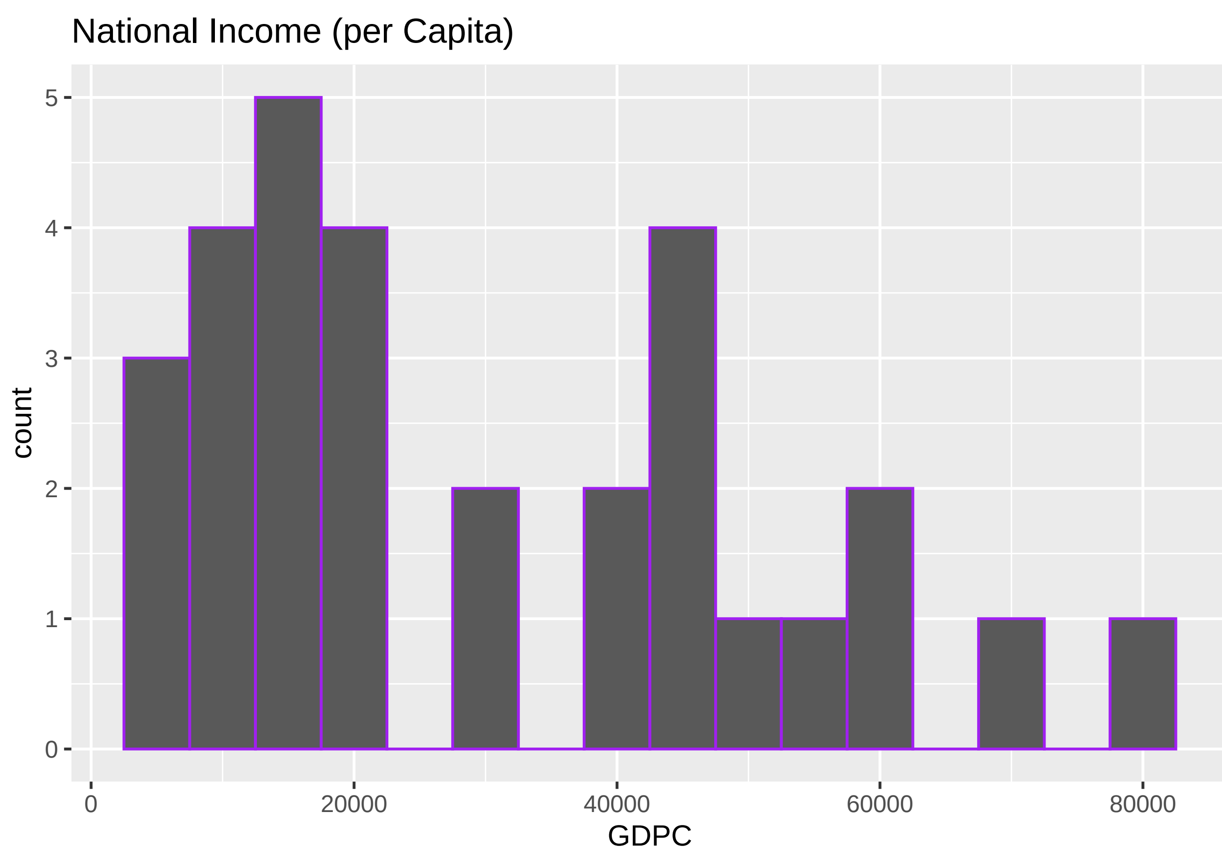


Figure 3 shows the distribution of GDP per Capita, which is far more normalized and easier to follow than that of Figure 1. My hope is that this measure will prove to be easier to understand in future use.

Results

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

term	estimate	std.error	statistic	p.value
(Intercept)	0.6877934	0.0299908	22.9334426	0.0000000
GDPC	0.0000002	0.0000008	0.2181977	0.8288576

```
##
## Call:
## lm(formula = PROP_Risk ~ GDPC, data = data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.267518 -0.058834  0.000335  0.057540  0.210429
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.878e-01  2.999e-02  22.933   <2e-16 ***
## GDPC        1.791e-07  8.210e-07   0.218    0.829
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09376 on 28 degrees of freedom
## Multiple R-squared:  0.001697,   Adjusted R-squared:  -0.03396
## F-statistic: 0.04761 on 1 and 28 DF,  p-value: 0.8289
```

With a sample size of just 30 countries, the data at hand is quite picky and will be difficult to model. As GDPC in a nation increases by 1 dollar, there is an associated increase by .00000018 in the proportion of the population that is at risk. However, the very large p-value is far from statistically significant, with an incredibly low standard error. To justify this, the fact that this model accounts for a miniscule .17% of the variance says a lot about its applicability.

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

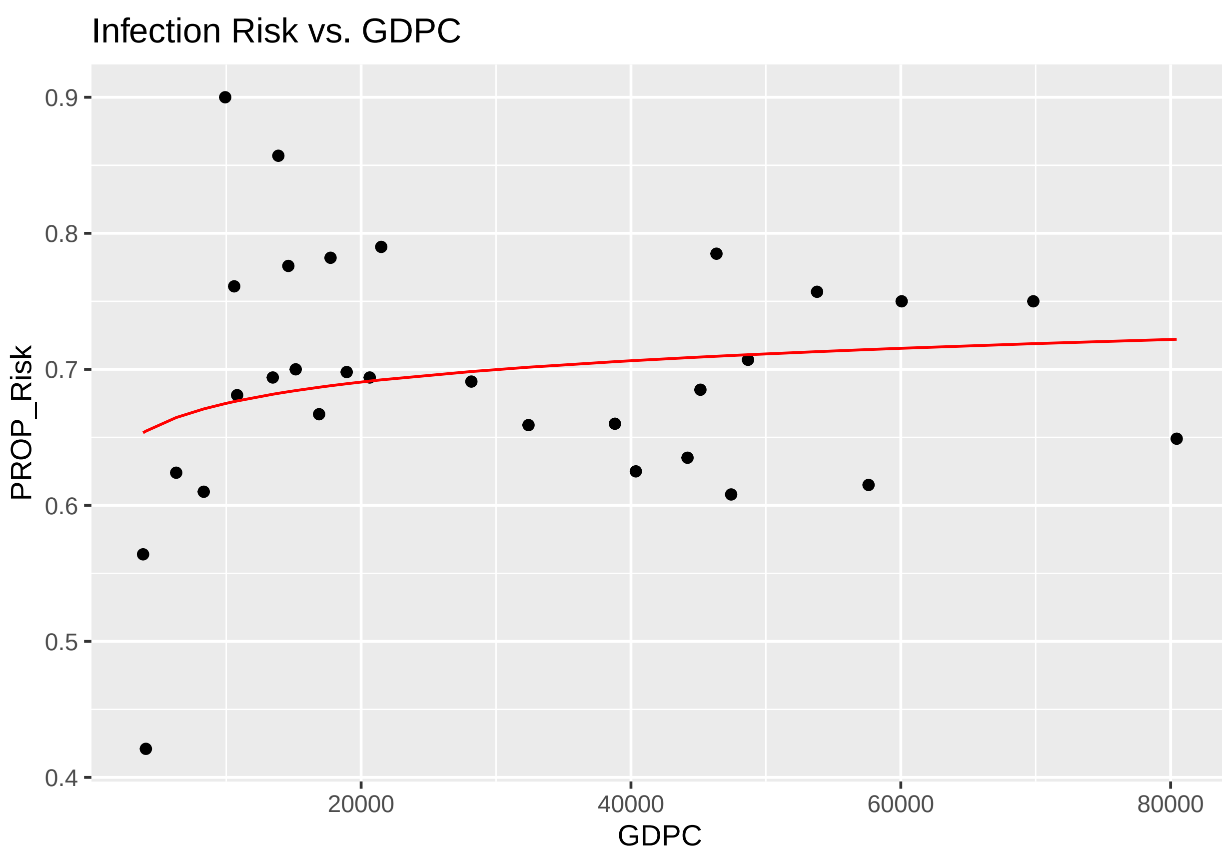
term	estimate	std.error	statistic	p.value
(Intercept)	0.4670418	0.2041911	2.287278	0.0299407
log(GDPC)	0.0225782	0.0203193	1.111170	0.2759459

```
##
## Call:
## lm(formula = PROP_Risk ~ log(GDPC), data = data)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.233563 -0.057064 -0.000539  0.041644  0.225174
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.46704    0.20419   2.287   0.0299 *
## log(GDPC)    0.02258    0.02032   1.111   0.2759
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09184 on 28 degrees of freedom
```

```
## Multiple R-squared:  0.04223,    Adjusted R-squared:  0.008028
## F-statistic: 1.235 on 1 and 28 DF,  p-value: 0.2759
```

Logging GDP has a strong effect on the model, indicating that a 1% increase in GDPC is associated with a .0226 increase in the percentage of those who have relatives at risk or risk themselves, with a standard error of ~.02. This model accounts for just 4.2% of the variance; still not a great model, but it is definitely improved.

Figure 4



When graphed, this model shows a model that struggles to fit the data, purely based on the spread of poorer countries and their different risks.

Table 3, $PROP\_Risk = \beta_0 + \beta_1 log(GDP) + \epsilon_i$

term	estimate	std.error	statistic	p.value
(Intercept)	0.1158804	0.2726905	0.424952	0.6741196
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)  0.11588    0.27269   0.425   0.6741
## 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
```

For purposes of trying to improve these models, I took a look at the overall national income that was skewed based on population. Oddly enough, using the non-normalized GDP actually created a better model. This model states that a 1% increase in GDP will result in a .022 increase in risk, at just half the standard error. It also has a far smaller p-value and is able to account for almost 14% of the variance in our data. Because of this, I will be utilizing GDP as my measure of national income from now on. This improvement is also reflected in Figure 5, below.

Figure 5

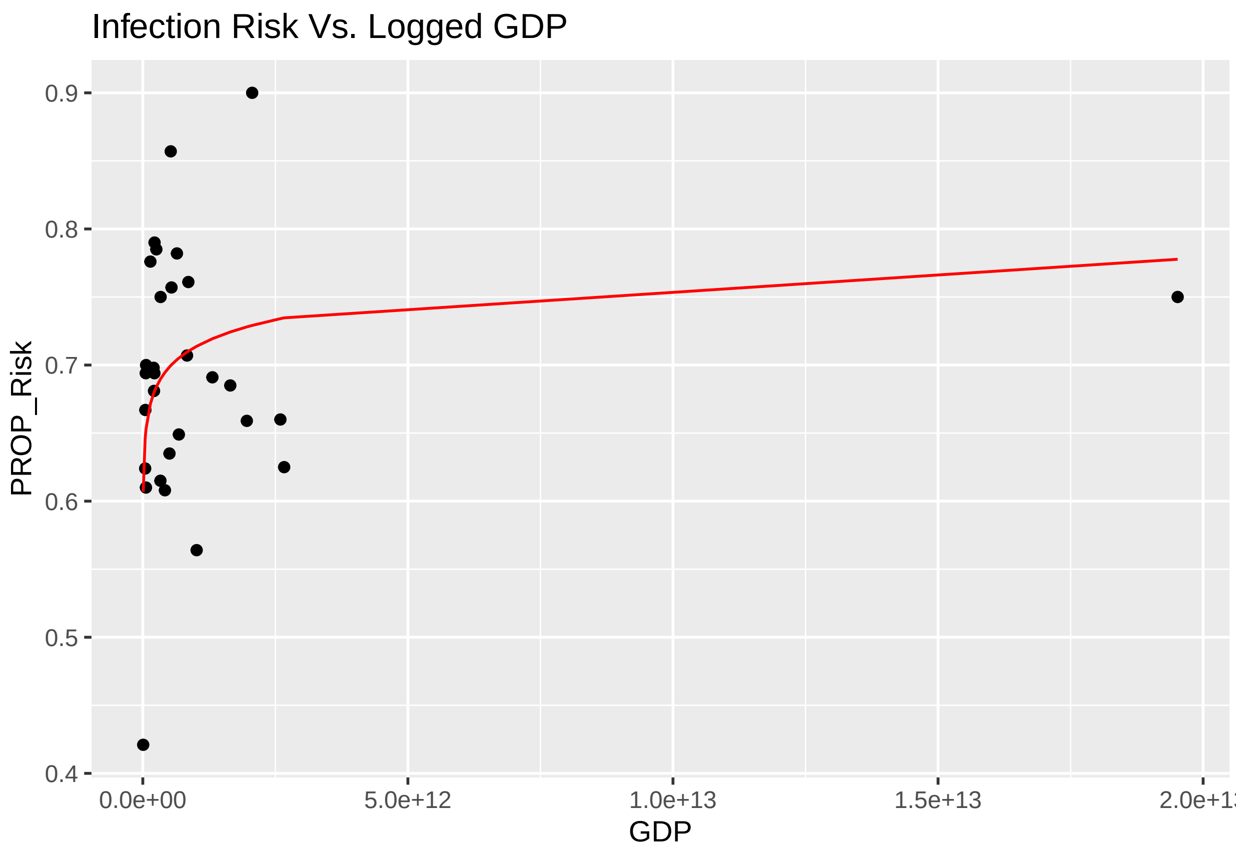


Figure 5 shows a graph of the newly created, logged GDP model. The United States is comically outlying and appears to skew this model, however it may be necessary to its understanding.

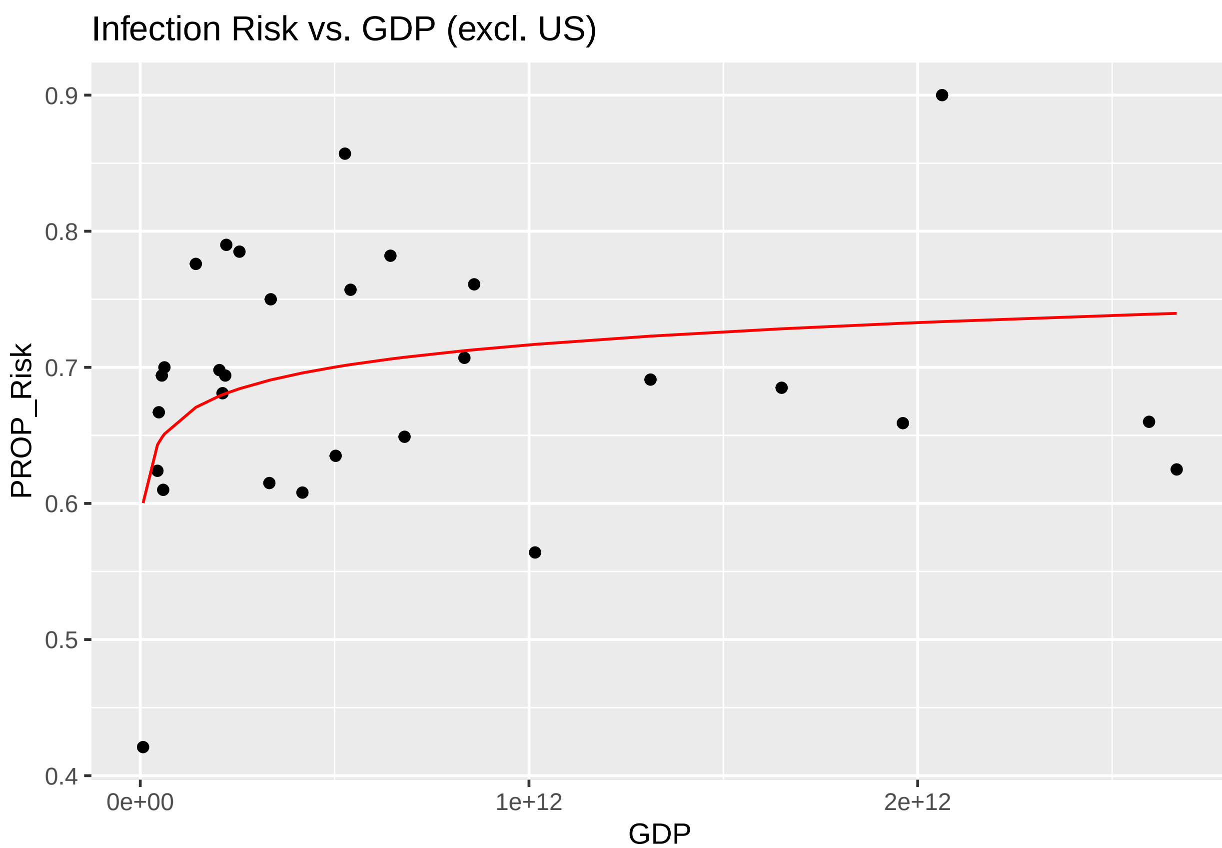
Table 4, $PROP\_Risk = \beta_0 + \beta_1 log(GDP) + \epsilon_i$ without United States

term	estimate	std.error	statistic	p.value
(Intercept)	0.0647117	0.3113530	0.2078402	0.8369135
log(GDP)	0.0235898	0.0117072	2.0149786	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)  0.06471    0.31135   0.208    0.837
```

```
## 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
```

Figure 6



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

term	estimate	std.error	statistic	p.value
(Intercept)	-0.0251616	0.2321690	-0.1083763	0.9144986
log(GDP)	0.0182375	0.0086091	2.1183965	0.0434895
GINI	0.0066452	0.0018592	3.5743342	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) -0.025162   0.232169  -0.108  0.91450
## 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 5 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 6, $PROP_Medical = \beta_0 + \beta_1 log(GDPC) + \beta_2 GINI + \epsilon_i$

term	estimate	std.error	statistic	p.value
(Intercept)	0.0062949	0.0040826	1.541871	0.1347451
log(GDPC)	-0.0006279	0.0003666	-1.712687	0.0982332
GINI	0.0000438	0.0000419	1.046010	0.3048309

```
##
## Call:
## lm(formula = PROP_Medical ~ log(GDPC) + GINI, data = data)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0020933 -0.0010640 -0.0004004  0.0005544  0.0047148
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  6.295e-03  4.083e-03   1.542   0.1347
## log(GDPC)   -6.279e-04  3.666e-04  -1.713   0.0982 .
## GINI         4.382e-05  4.189e-05   1.046   0.3048
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.00165 on 27 degrees of freedom
## Multiple R-squared:  0.1397, Adjusted R-squared:  0.07602
## F-statistic: 2.193 on 2 and 27 DF,  p-value: 0.1311
```

Table 6 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. Interestingly, GDPC actually serves to be a better indicator for Medical facilitation, accounting for 4% more of the variance than that of basic GDP. This model indicates that for every increase in 1% of GDPC, the proportion is expected to decrease by .00063, and for every 1 unit increase in the Gini index, it will increase by .000044. This model only accounts for a little under 14% 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 7, $PROP\_Isolated = \beta_0 + \beta_1 log(GDP) + \beta_2 log(GINI) + \epsilon_i$

term	estimate	std.error	statistic	p.value
(Intercept)	-1.2069225	0.5957623	-2.025846	0.0527670
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) -1.20692    0.59576  -2.026   0.0528 .
## 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 7 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 income inequality (measured through GDP and the Gini index) would have a positive relationship on all of the following: infection risk to self or relatives, medical isolation status, and general isolation status. After analysis, I've found that only two of the three analyses have conclusive results.

For the case of the proportion of medical isolation, the data I have for income inequality was not able to produce statistically significant explanation for the variance in medical isolation, with a p-value greater than .05 and a somewhat low R-squared value. However, this was not the case for my other two areas of interest. Using logged GDP and the Gini index, I found a statistically significant model to conclude that there is a positive relationship between a nation's income inequality and the proportion of that nation's population that is either at risk themselves, or has close family/friends who are at risk of infection. This model accounts for almost 42% of the data's variance, as well as having a p-value of .0007. In addition, I was also able to conlude that there is a positive relationship between income inequality and the proportion of isolation, using a model logging both GDP and the Gini index that accounts for ~23% of variance and has a p-value of .0296. While the correlation is only moderate, the conclusion itself is very thought-provoking. These findings defend two out of my three hypotheses, and also makes sense given other sources understandings of the connection between inequality and the COVID-19 pandemic.

Given more data, which I expect to come forward in the next few months with the pandemic slowing, I feel I could create much more accurate models to portray these relationships. I would hope to explore this problem more in-depth, hopefully utilizing other areas of risk within the pandemic, such as deaths, cases over time, etc. I also feel as though income inequality is not the only factor worth considering when it comes to the impact on this pandemic. Racial inequality is another huge problem across the globe that I believe plays a large part in reaction, especially now with vaccinations and freedoms becoming more open-ended. While the observations I found are on the smaller side, aI do believe that these findings are essential to understanding COVID as well as income inequality as a whole. Contemplating our past and our present is an important step in preparing the world for the future to come.

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