

Effect of Covid 19 on Voting Patterns in the 2020 US Presidential Election

Amir Hamilton

11/19/2020

Introduction

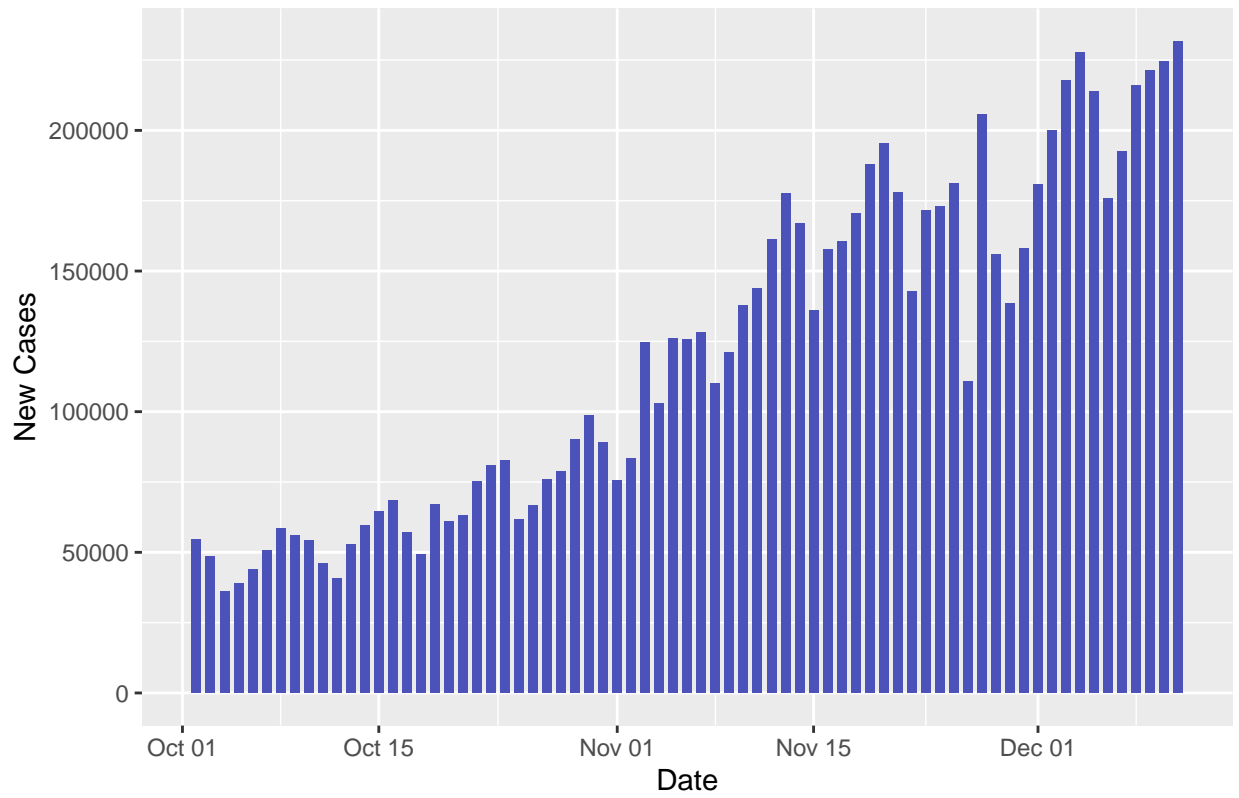
One way that social scientists are able to interact with a lay audience is by discussing the factors behind elections in the US. Some of the proposed factors include: the unequal distribution of voting machines by state, community, and race; a candidate's charisma and the perception of their character; the strength of the economy; etc. In general, these factors fall into two categories: a candidate's traits or societal trends.

This paper will look at the influence of the Covid-19 pandemic on the results of the 2020 Presidential Election. Specifically, it will examine whether there is a relationship between Covid numbers and the percentage of Trump or Biden votes on a county level.

Covid-19 has defined the world in 2020, with millions of deaths worldwide and widespread economic chaos. In the US, there have been three main waves of infection: Spring (March and April), Summer (July), and Winter (October - Present).

To begin, a bar plot of new covid cases per day in the winter wave, which is currently ongoing, is shown on the next page. Approximately one month of data preceding election day (November 4, 2020) is shown, as are the roughly one and a half months that have passed between then and now (December 14, 2020). Data has been read in from Our World in Data's (OWID) daily updated dataset.

Total New Cases in the US by Date From October 1 to December 11



Reproduction Number (R)

R , the reproduction number, is a measure of the infectious potential of a disease. If R equals 1, then the number of cases will remain constant. If R is less than 1, then the number of cases are decreasing, and thus if R is greater than 1 the number of cases are increasing. Determining R is valuable for public health officials and lawmakers because it tells you what proportion of new infections you need to prevent in order to go from increasing cases to constant or decreasing cases.

At a particular time in the pandemic, R_t is written to signify the reproduction number at time t .

There are several ways to calculate R , but I'm just using the data from OWID. The formula for R is considered to be different by different epidemiological schools, and can contain numbers such as:

T – the generation time – which is the time between infection events in an infector-infectee pair of individuals;

s – the serial interval – which is the average time between symptoms of infection in the transmitter to when the person they infect develops symptoms;

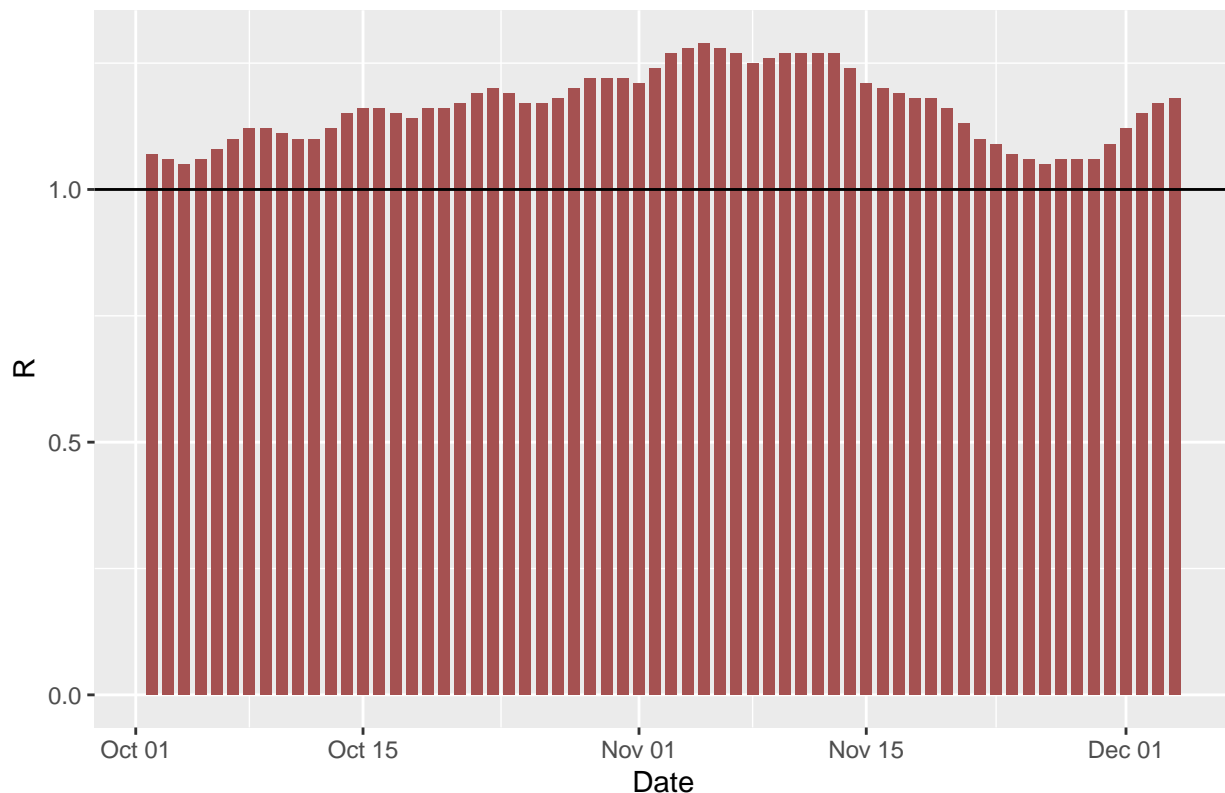
r – the epidemic growth rate – which is the rate at which new cases are occurring;

dt – doubling time – the amount of time which leads to a doubling of cases;

as well as the number of susceptible individuals, proportion of asymptomatic individuals, incubation time, disparate effects by age, duration from symptom onset to hospitalization or death, and the infectiousness of asymptomatic individuals. I'm leaving this calculation up to the experts, especially due to the fact that information changes frequently about these values.

Below is shown a box plot of R from the beginning of October to December 4, the latest day for which OWID has calculated R . A horizontal line of $R = 1$ shows that during the time period covered in this study, the severity of the pandemic has been increasing every day.

US Reproduction Number (R) Over Time



Presidential Election

To conduct my analysis, I had to get the data for election results. In order to achieve a sufficient level of granularity I gathered data on the county level. I could not find a dataset with all the election results for president by county, unfortunately, so I manually scraped the data from Politico and NBCNews. Another advantage of this method is that the data would be in the same format, since each state released their elections data in different formats.

Here, I read in the data from the various excel spreadsheets I created (minus Alaska and DC). Alaska was excluded because election data was unavailable on a county level and DC was excluded because there is no county there. I then merged the state/county level data to create a national dataframe of election results, which I would be able to work with.

The dataset is too large to show here in its entirety, but I've included a preview below:

```
## # A tibble: 10 x 7
##   county trump.votes trump.percent biden.votes biden.percent state dispID
##   <chr>      <dbl>      <dbl>      <dbl>      <dbl> <chr> <chr>
## 1 Autauga      19838      0.716      7503      0.271 Alaba~ Autauga, ~
## 2 Baldwin     83544      0.764     24578      0.225 Alaba~ Baldwin, ~
## 3 Barbour      5622      0.535      4816      0.458 Alaba~ Barbour, ~
## 4 Bibb         7525      0.785      1986      0.207 Alaba~ Bibb, Ala~
## 5 Blount      24711      0.897      2640      0.096 Alaba~ Blount, A~
## 6 Bullock      1146      0.249      3446      0.747 Alaba~ Bullock, ~
## 7 Butler       5458      0.576      3965      0.418 Alaba~ Butler, A~
## 8 Calhoun     35101      0.69      15216      0.299 Alaba~ Calhoun, ~
## 9 Chambe~      8753      0.574      6365      0.417 Alaba~ Chambers,~
## 10 Cherok~    10583      0.861      1624      0.132 Alaba~ Cherokee,~
```

Interactions

I decided to look at the relationship between the Percentage of Trump votes in a county or the Percentage of Biden votes in a county to the R at five different dates: Nov 4, Oct 28, Oct 21, Oct 14, and Oct 7. I chose dates that were multiples of one week prior to election day, as well as election day itself. So, I created five new dataframes created by merging the election results with the covid data on those four days.

To assist with understanding the national distribution of R on those five days, I have created a sample density plot for November 4 shown below. The median R is above zero.

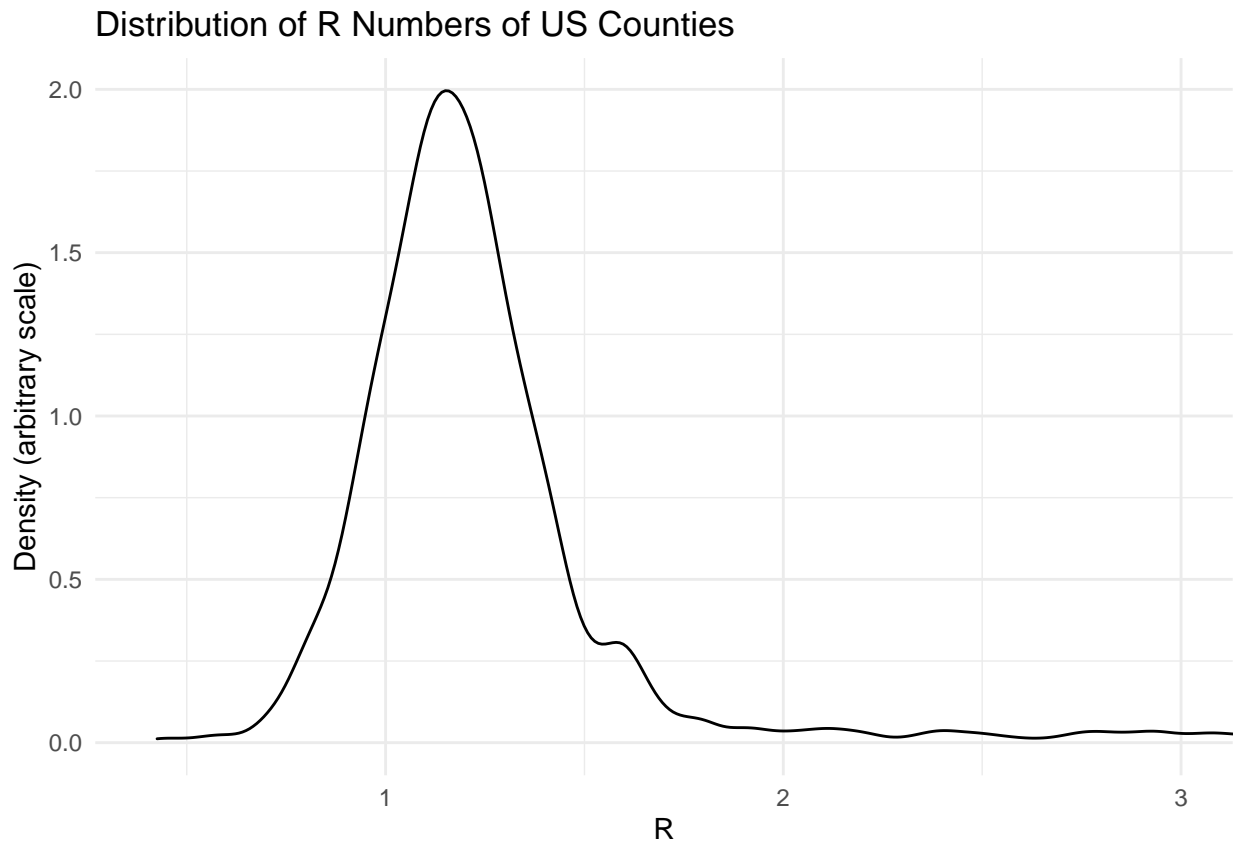


Table 1: Regression Table between Percentage of Trump Votes and R per County

	Nov 4	Oct 28	Oct 21	Oct 14	Oct 7
(Intercept)	0.572*** (0.011)	0.554*** (0.026)	0.449*** (0.022)	0.475*** (0.025)	0.455*** (0.024)
cc_nov_4\$Rt_plot	0.023*** (0.008)				
cc_oct_28\$Rt_plot		0.033 (0.022)			
cc_oct_21\$Rt_plot			-0.044** (0.019)		
cc_oct_14\$Rt_plot				0.088*** (0.021)	
cc_oct_7\$Rt_plot					0.099*** (0.020)
Num.Obs.	1343	1149	1042	951	856
R2	0.006	0.002	0.005	0.019	0.027
R2 Adj.	0.005	0.001	0.004	0.018	0.026
AIC	-1105.4	-920.1	-837.4	-760.8	-684.4
BIC	-1089.8	-905.0	-822.6	-746.3	-670.1
Log.Lik.	555.685	463.067	421.723	383.414	345.180
F	8.366	2.188	5.393	18.275	23.610

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Differences in Means for Percentage of Trump Votes on R for Nov 4, Oct 28, Oct 21, Oct 14, and Oct 7

When percentage of Trump votes per county (trump.percent) was predicted it was found that Rt_plot on election day (Nov 4) was a significant predictor ($t = 2.892$, $p = 0.00388$). The overall model fit is $R^2 = 0.0062$. The difference in means is 0.022859.

Because the p-value is lower than the standard threshold of $\alpha = .05$ we can reject the null hypothesis that the relationship of trump.percent to Rt_plot on Election Day is zero. This does not necessarily imply a causal relationship, however. It could be that in a county with high proportions of Trump support people are less likely to take Covid-19 seriously, or it could be that living in counties with high Rt_plot would make people more likely to vote for Trump (vote for the chaotic candidate in a chaotic circumstance).

Table 2: Regression Table between Percentage of Biden Votes and R

	Nov 4	Oct 28	Oct 21	Oct 14	Oct 7
(Intercept)	0.408*** (0.011)	0.428*** (0.026)	0.449*** (0.022)	0.508*** (0.024)	0.525*** (0.024)
cc_nov_4\$Rt_plot	-0.021*** (0.008)				
cc_oct_28\$Rt_plot		-0.032 (0.022)			
cc_oct_21\$Rt_plot			-0.044** (0.019)		
cc_oct_14\$Rt_plot				-0.089*** (0.021)	
cc_oct_7\$Rt_plot					-0.096*** (0.020)
Num.Obs.	1343	1149	1042	951	856
R2	0.005	0.002	0.005	0.019	0.026
R2 Adj.	0.004	0.001	0.004	0.018	0.025
AIC	-1122.5	-934.7	-837.4	-769.1	-692.6
BIC	-1106.9	-919.6	-822.6	-754.5	-678.3
Log.Lik.	564.242	470.360	421.723	387.557	349.296
F	6.870	2.056	5.393	18.540	22.770

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Differences in Means for Percentage of Biden Votes on R for Nov 4, Oct 28, Oct 21, Oct 14, and Oct 7

When percentage of Biden votes per county (biden.percent) was predicted it was found that R (Rt_plot) on election day (Nov 4) was a significant negative/inverse predictor ($t = -2.621$, $p = 0.00887$). The overall model fit is $R^2 = 0.005097$, and the difference in means is -0.020582 .

Because the p-value is smaller than the standard threshold of $\alpha = .05$ we can reject the null hypothesis that the relationship of biden.percent to R on Election Day is zero. In comparison to the relationship between trump.percent and R , biden.percent is less strongly related to R .

R on a County Level

To achieve a more granular perspective on R on November 4 (Election Day), a map is shown below, where darker shades indicate higher R. Some counties are missing due to the incompleteness of the dataset. OWID did not calculate R in counties that did not have a sufficiently high case count.

Sources and References

<https://royalsociety.org/news/2020/09/set-c-covid-r-rate/>

Chongsuvivatwong, Virasakdi. Analysis of epidemiological data using R and Epicalc. 2008, Hat Yai, Thailand: Chammuang Press.

Boelle, Pierre-Yves, Thomas Obadia. “Estimation of R_0 and Real-Time Reproduction Number from Epidemics.”

<https://ourworldindata.org/coronavirus>

https://github.com/lin-lab/COVID19-Viz/tree/master/clean_data

County data was manually scraped from nbcnews.com and politico.com on 2020-11-19, 2020-11-20, and 2020-12-11. Alaska was excluded due to discrepancies in the formatting of the data.

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. The Lancet infectious diseases, 20(5), 533-534.

For reproduction rate by county, the dataset was too large for github so I manually reduced the size to exclude data from before November 15 and to exclude counties/dates that had too small of a sample size to produce a reproduction rate, as determined by the study.