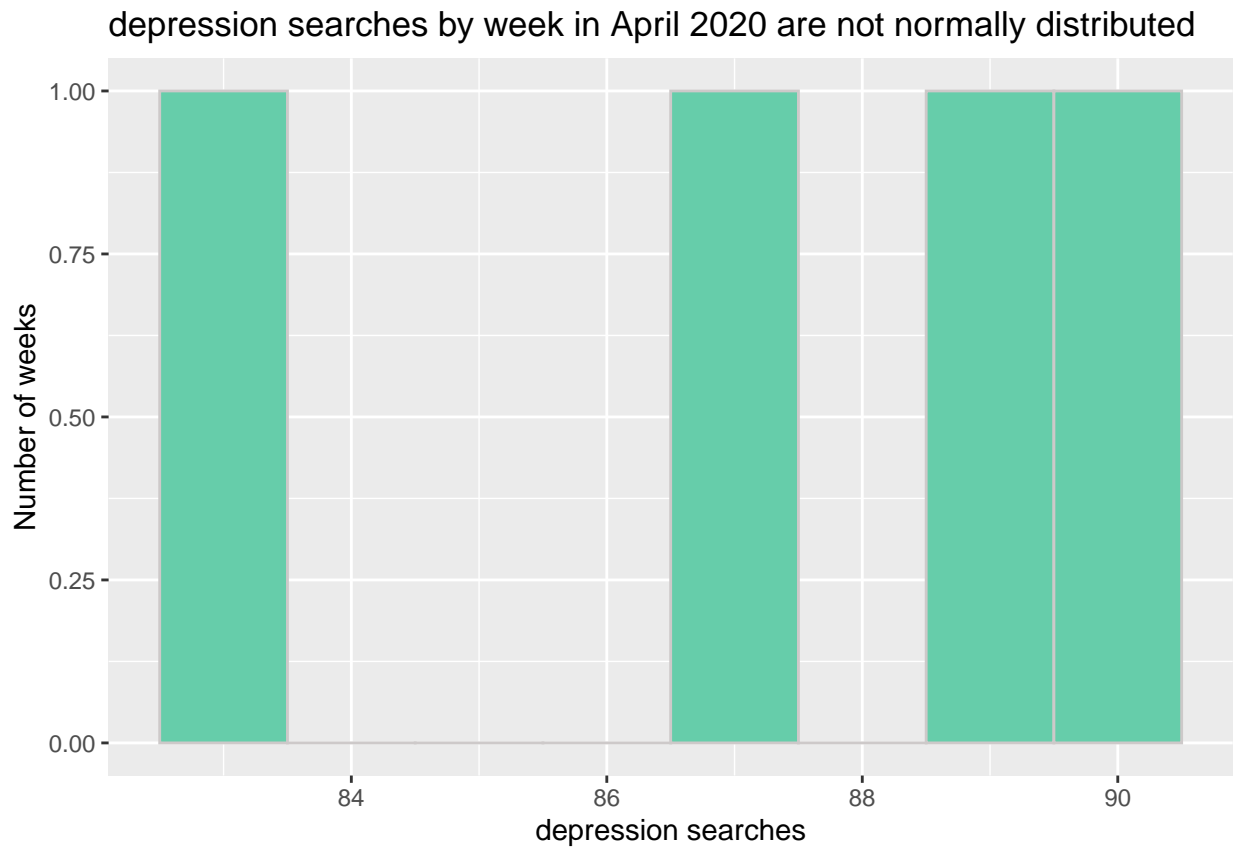# Final Project
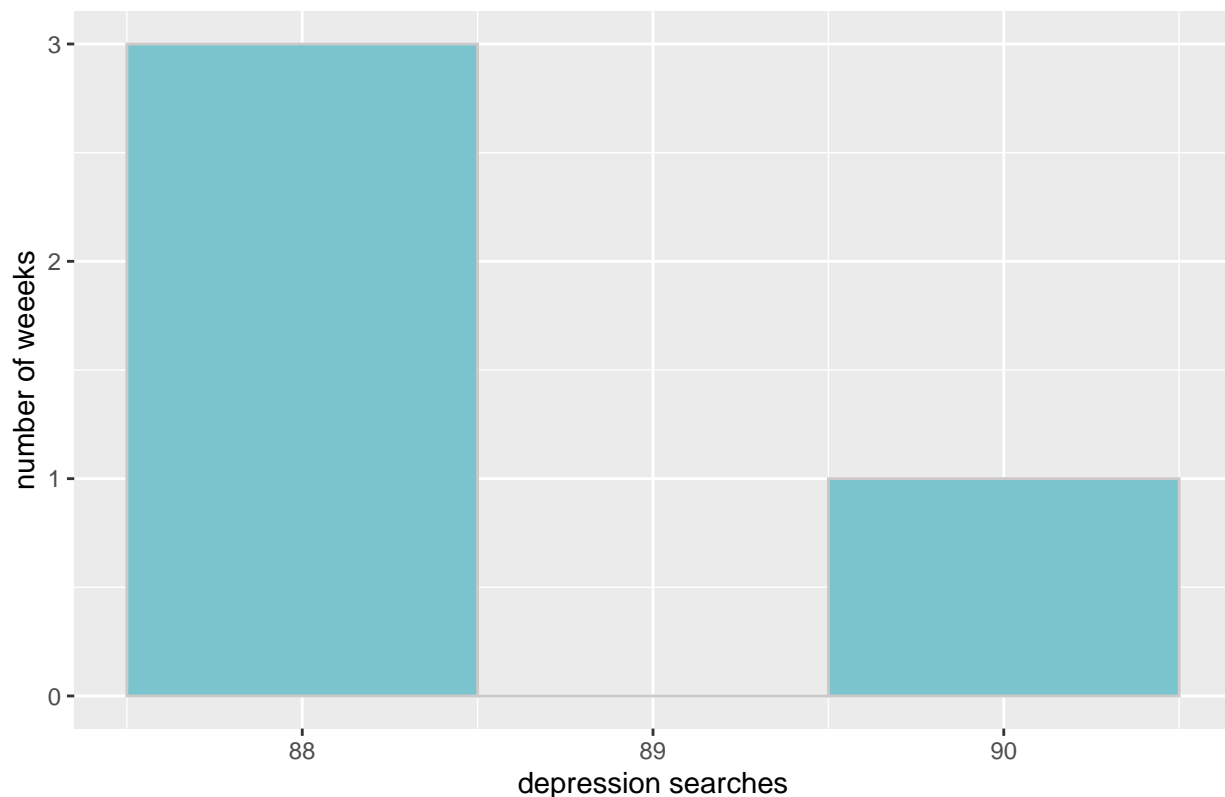
Brenda Yang, Charlie Bonetti, Nour Kanaan

7/19/2020

Comparing depression trends between April 2020 and April 2018

n<30 and not normal distribution: assumption for t-test not satisfied

depression searches by week in April 2020 are not normally distributed

n<30 and not normal distribution: assumption for t-test not satisfied

```
##
##  Paired t-test
##
## data:  d2020 and d2018
## t = 0.62017, df = 3, p-value = 0.5791
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.164426  7.664426
## sample estimates:
## mean of the differences
##                    1.25
```

The null hypothesis is that there is no difference in the mean amount of depression searches in the US between the times of April 2020 and April 2018. The alternate hypothesis is that there is a difference between the two means. Assuming that the null hypothesis is true, the model follows a t-distribution. The t-statistic is 0.755 and the df = 29. This corresponds to a p-value of 0.4561. We cannot reject the null at the alpha = 0.05 level. We do not have enough evidence to claim that there is a difference in the mean amount of depression searches in the US between the times of April 2020 and April 2018.

COVID cases vs. depression rate

```
##           State Confirmed.COVID.cases.as.of.4.01.20
## 1        Alaska                                 143
## 2       Arizona                                1413
## 3      Arkansas                                 624
## 4        Hawaii                                 258
```

```
## 5            Idaho                              669
## 6          Kentucky                            674
## 7            Maine                             344
## 8         Minnesota                            689
## 9           Montana                            219
## 10 North Carolina                             1591
## 11        Oklahoma                             719
## 12          Oregon                             737
## 13           Texas                            4402
## 14         Vermont                             321
## 15  West Virginia                              191
## 16       Wisconsin                            1552
## 17         Wyoming                             138
##    Confirmed.COVID.cases.as.of.4.31.20
## 1                                  353
## 2                                 7648
## 3                                 3281
## 4                                  609
## 5                                 2016
## 6                                 4708
## 7                                 1095
## 8                                 5136
## 9                                  452
## 10                               10507
## 11                                3618
## 12                                2510
## 13                               29072
## 14                                 866
## 15                                1126
## 16                                6854
## 17                                 559
##    Confirmed.new.COVID.cases.in.month.of.April.2020
## 1                                               210
## 2                                              6235
## 3                                              2657
## 4                                               351
## 5                                              1347
## 6                                              4034
## 7                                               751
## 8                                              4447
## 9                                               233
## 10                                             8916
## 11                                             2899
## 12                                             1773
## 13                                            24670
## 14                                              545
## 15                                              935
## 16                                             5302
## 17                                              421
##    X2020.State.Population..https...worldpopulationreview.com.states.
## 1                                                           734,002
## 2                                                         7,378,490
## 3                                                         3,039,000
## 4                                                         1,412,690
```

```
## 5                                                  1,826,160
## 6                                                  4,499,690
## 7                                                  1,345,790
## 8                                                  5,700,670
## 9                                                  1,086,760
## 10                                                10,611,900
## 11                                                 3,954,820
## 12                                                 4,301,090
## 13                                                29,472,300
## 14                                                   628,061
## 15                                                 1,778,070
## 16                                                 5,851,750
## 17                                                   567,025
##    New.COVID.cases.per.100.000.in.April depression anxiety restriction
## 1                             28.61028         79      77          21
## 2                             84.50238         88      78          32
## 3                             87.43008         90      76          12
## 4                             24.84622         84      81          51
## 5                             73.76134         87      73           5
## 6                             89.65062         87      89          24
## 7                             55.80365         94     100          26
## 8                             78.00837        100      79          21
## 9                             21.43988         82      91           7
## 10                            84.01888         92      78          28
## 11                            73.30296         91      80          15
## 12                            41.22211         80      77           3
## 13                            83.70572         90      74           8
## 14                            86.77501         87      97          39
## 15                            52.58511         93      79          23
## 16                            90.60537         82      82          34
## 17                            74.24717         69      65          18
##          party case_cat
## 1  republican      low
## 2  republican      low
## 3  republican      low
## 4    democrat      low
## 5  republican      low
## 6  republican      low
## 7    democrat      low
## 8  republican      low
## 9  republican      low
## 10 republican      low
## 11 republican      low
## 12   democrat      low
## 13 republican      low
## 14   democrat      low
## 15 republican      low
## 16 republican      low
## 17 republican      low

##             Df Sum Sq Mean Sq F value Pr(>F)
## case_cat     2  121.5   60.73   1.334  0.273
## Residuals   48 2185.2   45.52
```

The null is that there is no significant difference between the mean depression trends of states with low

COVID cases, medium COVID cases, and high COVID cases. The alternate hypothesis is that there exists at least one mean that is different. Assuming the null hypothesis is true, the model follows an F distribution with a df of 2. The F-statistic is 1.334, and the corresponding p-value is 0.273. Therefore, we can not reject the null under the alpha = 0.05 significance level. There is not enough evidence to suggest that there is at least one difference in mean depression trends of states with low, medium, and high COVID cases.

Restrictions vs. depression

```
##                  Df Sum Sq Mean Sq F value Pr(>F)
## restriction_cat  2    48.5   24.25   0.516    0.6
## Residuals        48 2258.1   47.04
```

The null is that there is no significant difference between the mean depression trends of states with low restrictions, medium restrictions, and high restrictions. The alternate hypothesis is that there exists at least one mean that is different. Assuming the null hypothesis is true, the model follows an F distribution with a df of 2. The F-statistic is 0.516, and the corresponding p-value is 0.6. Therefore, we can not reject the null under the alpha = 0.05 significance level. There is not enough evidence to suggest that there is at least one difference in mean depression trends of states with low, medium, and high restrictions.

Comparing anxiety trends in April 2020 to April 2018

```
ggplot(data = atrends2018, mapping = aes(x = anxiety)) +
  geom_histogram(color = "snow3", fill = "cornflowerblue", binwidth = 1)+
  labs(title = "anxiety searches by week in April 2020 are normally distributed",
       x = "anxiety searches",
       y = "number of weeeks")
```

anxiety searches by week in April 2020 are normally distributed
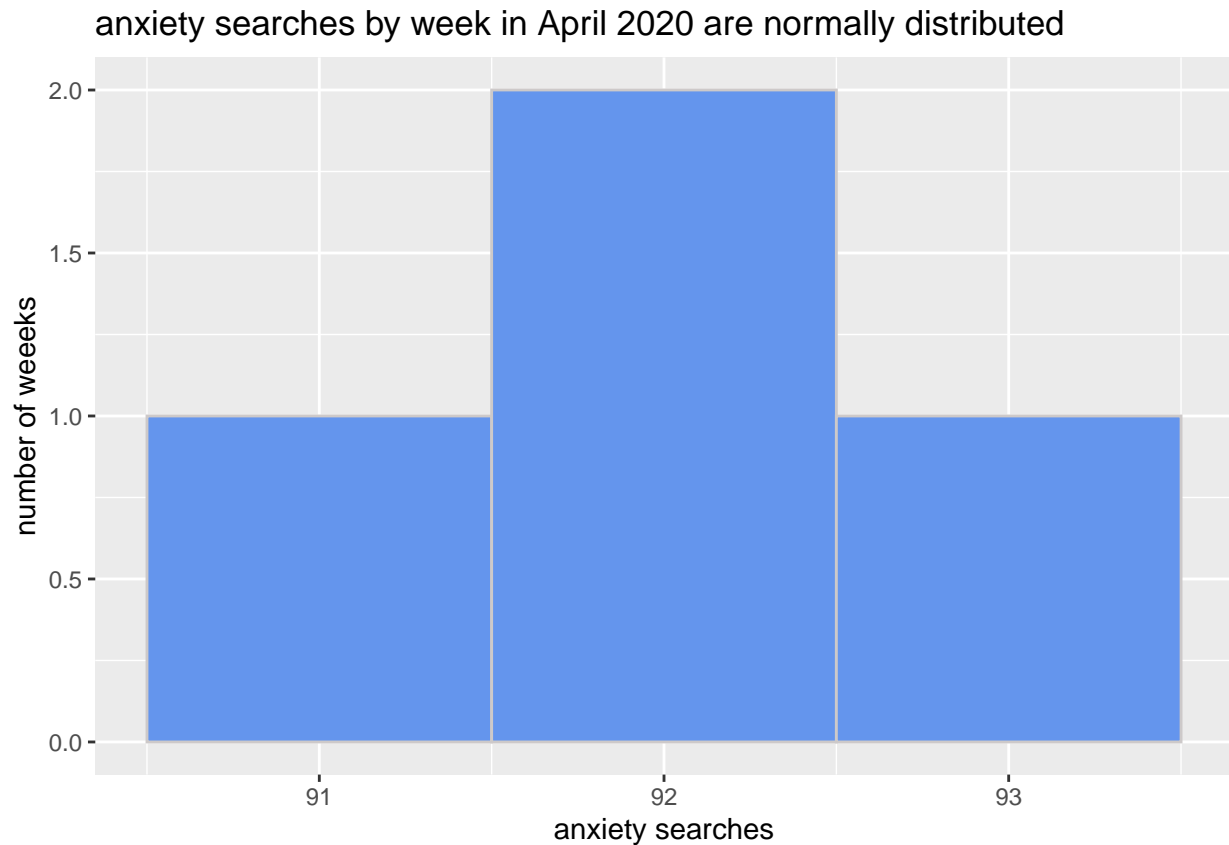
n<30, but has a normal distribution?? : assumption satisfied

```
ggplot(data = atrends2020, mapping = aes(x = anxiety)) +
  geom_histogram(color = "snow3", fill = "lightsteelblue", binwidth = 1)+
  labs(title = "anxiety searches by week in April 2020 are not normally distributed",
       x = "anxiety searches",
       y = "number of weeeks")
```

anxiety searches by week in April 2020 are not normally distributed
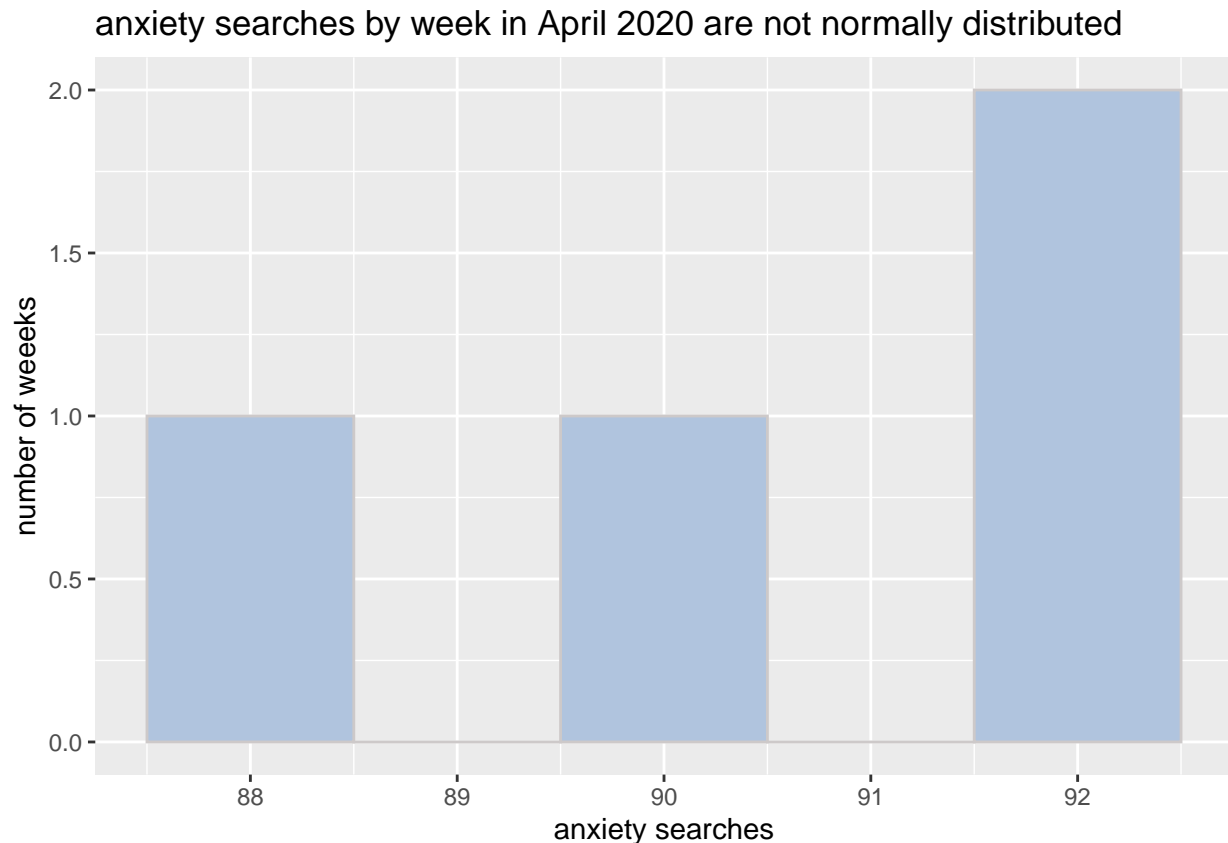
n<30 and not normal distribution: assumption not satisfied

```
##
##  Paired t-test
##
## data:  a2020 and a2018
## t = -1.2603, df = 3, p-value = 0.2967
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.287869  2.287869
## sample estimates:
## mean of the differences
##                    -1.5
```

The null hypothesis is that there is no difference in the mean amount of anxiety searches in the US between the times of April 2020 and April 2018. The alternate hypothesis is that there is a difference between the two means. Assuming that the null hypothesis is true, the model follows a t-distribution. The t-statistic is 1.66 and the df = 29. This corresponds to a p-value of 0.1086. We cannot reject the null at the alpha = 0.05 level. We do not have enough evidence to claim that there is a difference in the mean amount of anxiety searches in the US between the times of April 2020 and April 2018.

COVID cases vs. anxiety rate

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## case_cat     2  384.2  192.08   3.826 0.0287 *
## Residuals   48 2410.0   50.21
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The null is that there is no significant difference between the mean anxiety trends of states with low COVID cases, medium COVID cases, and high COVID cases. The alternate hypothesis is that there exists at least one mean that is different. Assuming the null hypothesis is true, the model follows an F distribution with a df of 2. The F-statistic is 3.826, and the corresponding p-value is 0.0287. Therefore, we can reject the null under the alpha = 0.05 significance level. There is enough evidence to suggest that there is at least one difference in mean anxiety trends of states with low, medium, and high COVID cases.

Restrictions vs. anxiety

```
##                   Df Sum Sq Mean Sq F value  Pr(>F)
## restriction_cat   2  612.9  306.43   6.743 0.00262 **
## Residuals        48 2181.3   45.44
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The null is that there is no significant difference between the mean anxiety trends of states with low restrictions, medium restrictions, and high restrictions. The alternate hypothesis is that there exists at least one mean that is different. Assuming the null hypothesis is true, the model follows an F distribution with a df of 2. The F-statistic is 6.746, and the corresponding p-value is 0.00262. Therefore, we reject the null under the alpha = 0.05 significance level. There is enough evidence to suggest that there is at least one difference in mean anxiety trends of states with low, medium, and high restrictions.
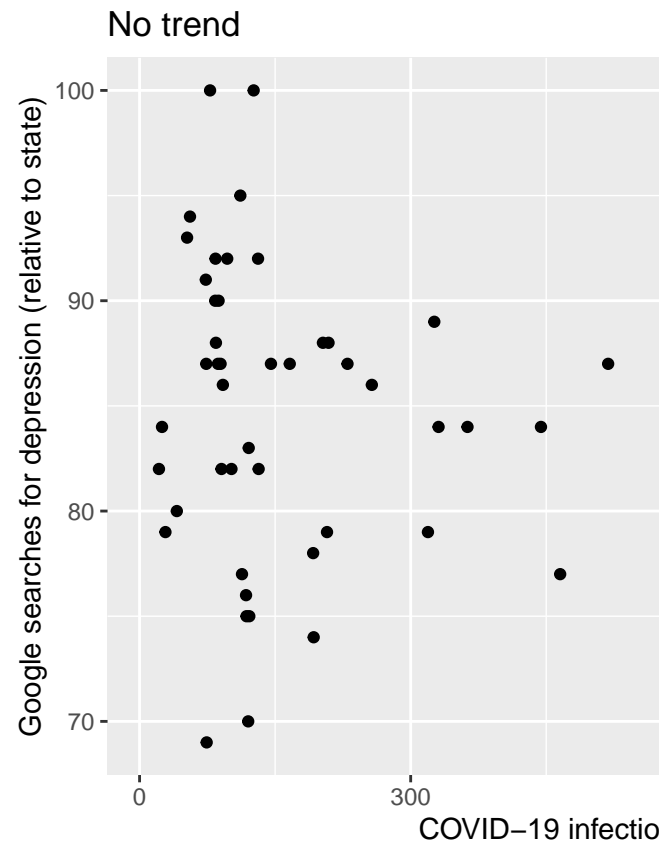
Depression rate in each state map

```
## Warning: Use of `map_df$x` is discouraged. Use `x` instead.

## Warning: Use of `map_df$y` is discouraged. Use `y` instead.

## Warning: Use of `map_df$group` is discouraged. Use `group` instead.
```
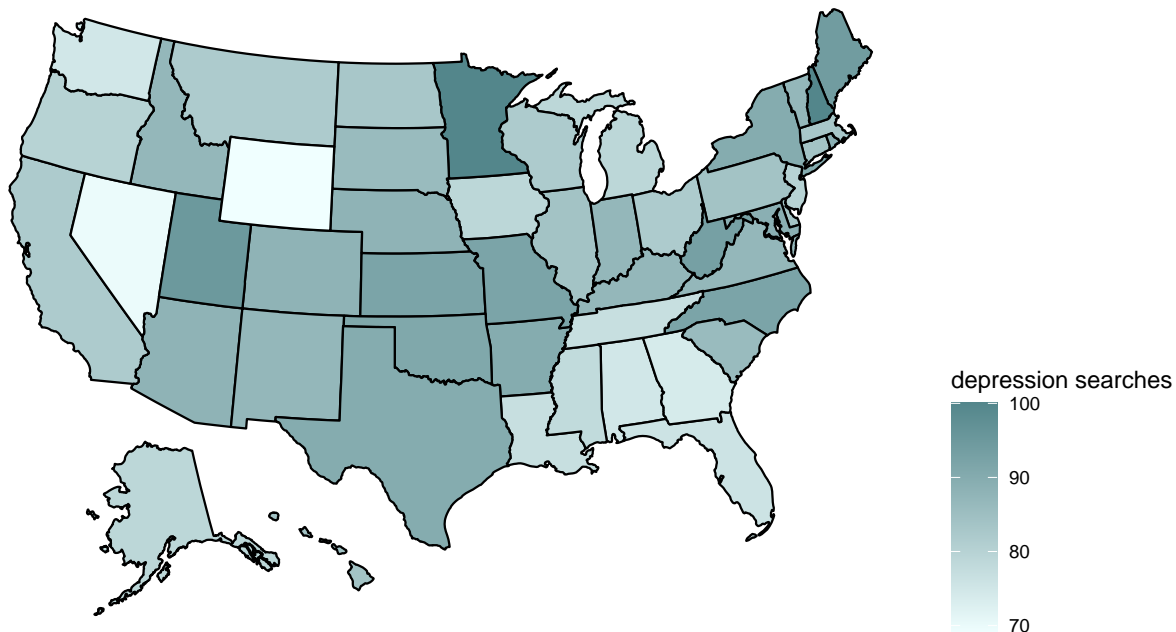

Nour Visuals

```
ggplot(data = covidrate, mapping = aes(x = New.COVID.cases.per.100.000.in.April, y = depression)) +
  geom_point() +
  labs(title = "No trend",
       x = "COVID-19 infection rate (per 100,000) ",
       y = "Google searches for depression (relative to state)" )
```


No trend

Effect of COVID on depression/anxiety rates in each State

```
ggplot(data = covidrate, mapping = aes(x = New.COVID.cases.per.100.000.in.April, y = anxiety)) +
  geom_point() +
  labs(title = "Curvilinear?",
       x = "COVID-19 infection rate (per 100,000) ",
       y = "Google searches for anxiety (relative to state)" )
```

Curvilinear?

COVID−19 infection rate (per 100,000)

Google searches for anxiety (relative to state)
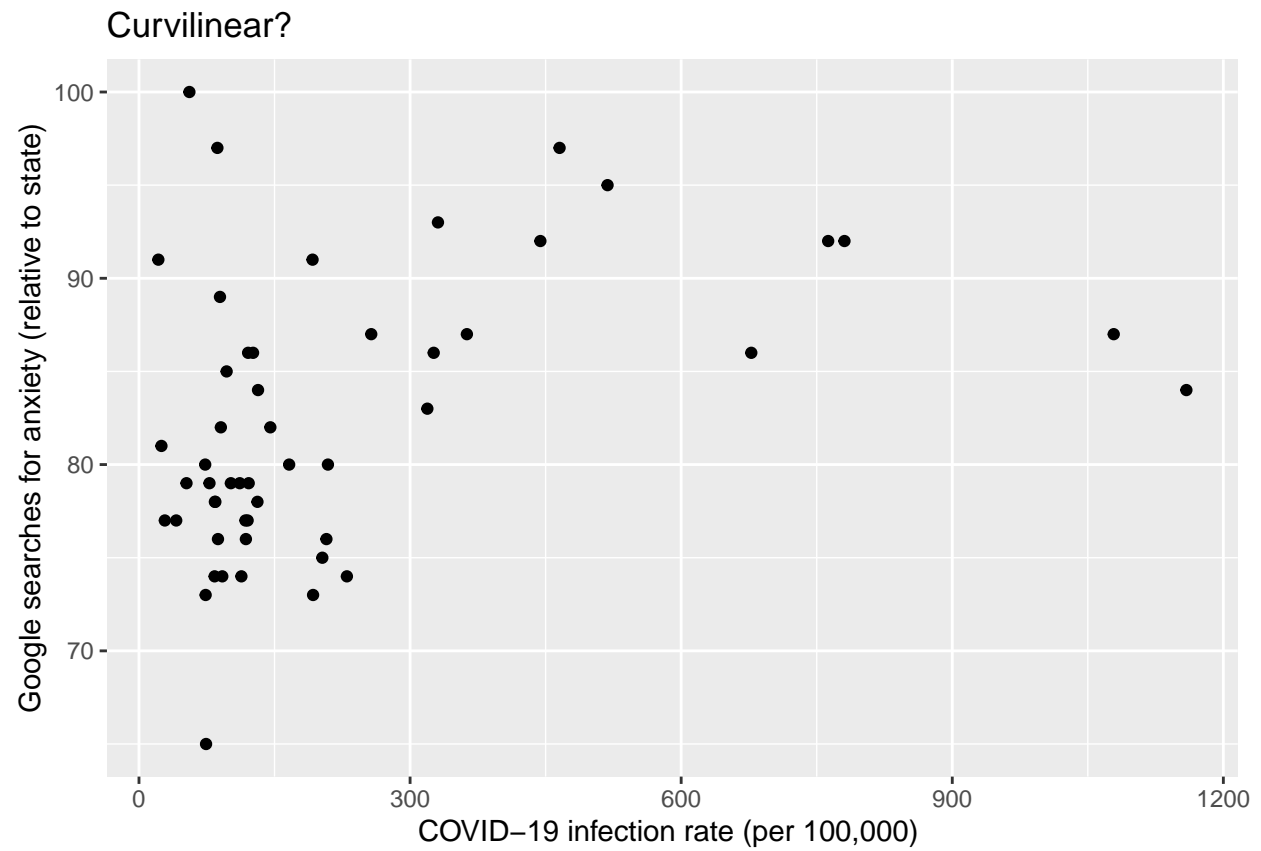
Effect of severity of restrictions on depression/anxiety rates in each State