Final Project

Brenda Yang, Charlie Bonetti, Nour Kanaan

7/19/2020

Comparing depression trends between April 2020 and April 2018

```
##
##  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

```
##             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.

Comparing anxiety trends in April 2020 to April 2018

```
##
##  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.

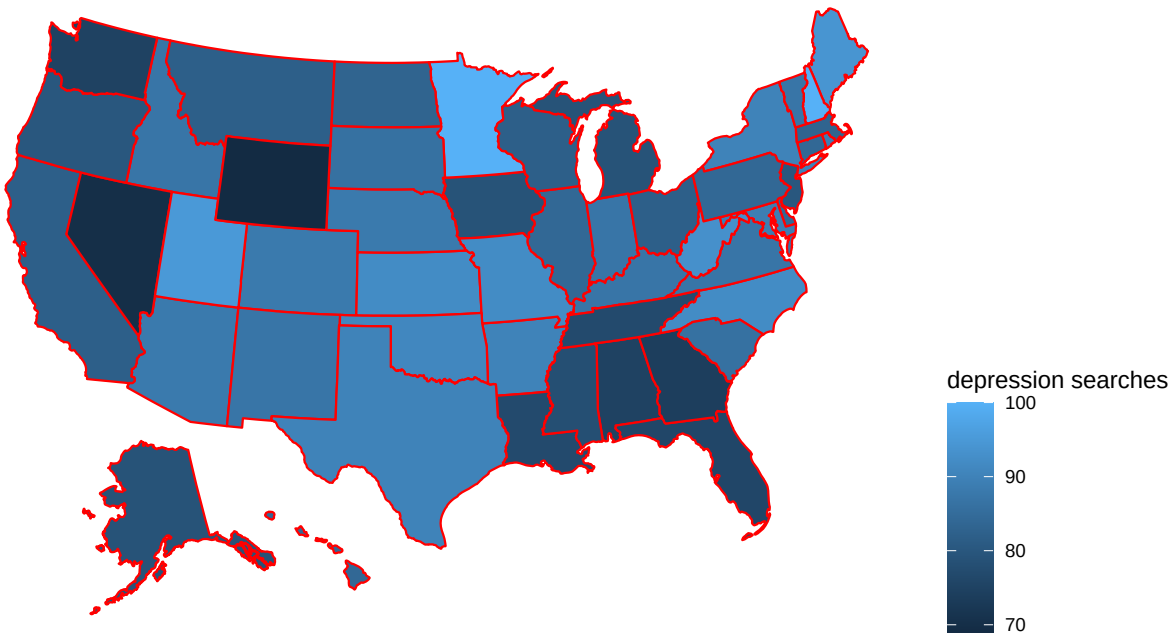
COVID cases vs. anxiety rate

```
##             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.

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 State_trends1 <- merge(astate, dstate, by= "state") State_trends3 <- merge(restriction_rankings, stateCovid, by= "state") State_Data <- merge(State_trends1,State_trends3, by= "state") State_final <- State_Data %>% select(state, anxiety, depression, '5-May', `New COVID cases per 100,000 in April`) covidrate_rank <- State_final %>% select(`New COVID cases per 100,000 in April`) %>% mutate(case_cat = case_when(`New COVID cases per 100,000 in April` <= 91 ~ "low", `New COVID cases per 100,000 in April` > 91 & `New COVID cases per 100,000 in April` < 200 ~ "medium", `New COVID cases per 100,000 in April` >= 200 ~ "high"))