

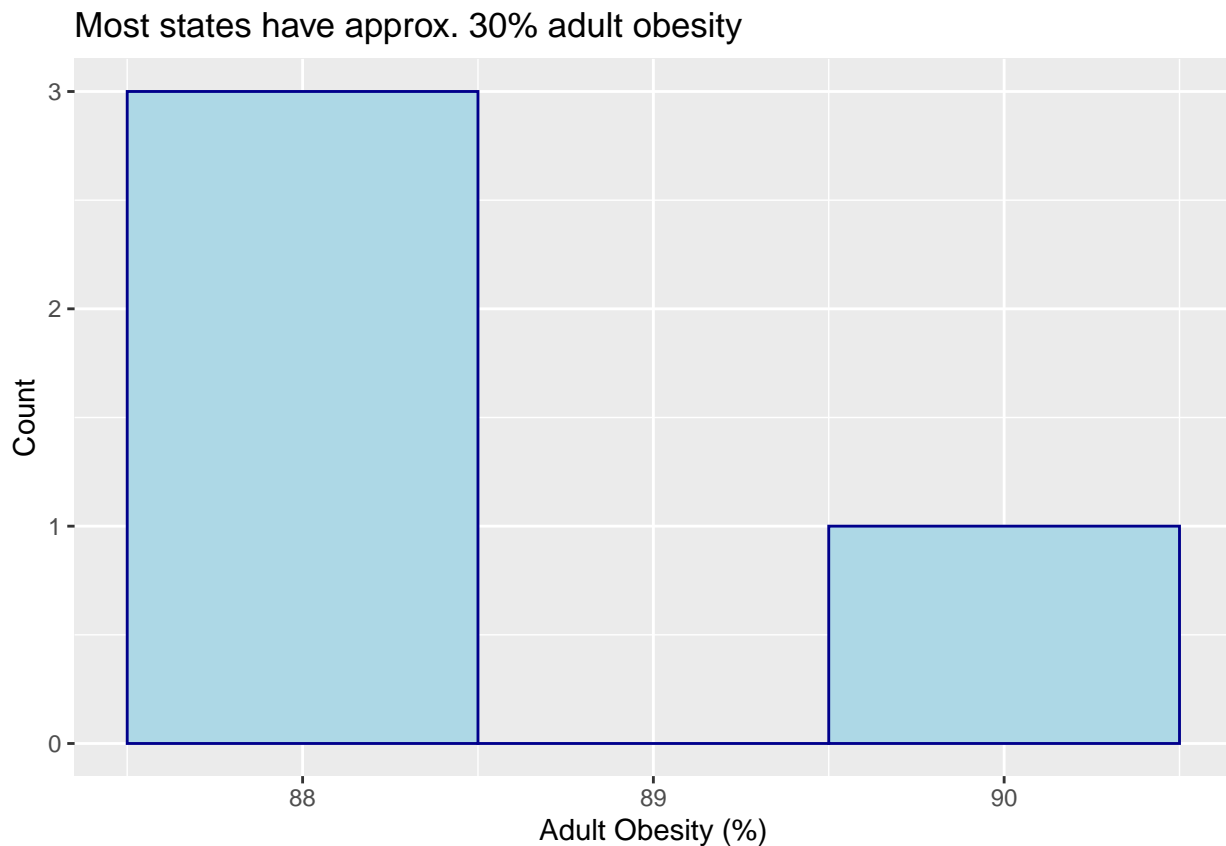
Final Project

Brenda Yang, Charlie Bonetti, Nour Kanaan

7/19/2020

Comparing depression trends between April 2020 and April 2018

```
ggplot(data = dtrends2020, mapping = aes(x = depression)) +  
  geom_histogram(color = "darkblue", fill = "lightblue", binwidth = 1)+  
  labs(title = "Most states have approx. 30% adult obesity",  
       x = "Adult Obesity (%)",  
       y = "Count")
```



```
##  
## 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

```
restriction_by_case <- aov(`New COVID cases per 100,000 in April` ~ case_cat, data = covidrate)
summary(restriction_by_case)
```

```
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## case_cat    2 1776921  888460    28.9 5.77e-09 ***
## Residuals  48 1475461   30739
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

interpretation for ANOVA here

Restriction vs. depression rate

Comparing anxiety trends in April 2020 to April 2018

```
a2020 <- atrends2020 %>%
  select(anxiety) %>%
  pull()

a2018 <- atrends2018 %>%
  select(anxiety) %>%
  pull()

t.test(a2020, a2018, mu = 0,
       alternative = "two.sided",
       paired = TRUE,
       var.equal = FALSE,
       conf.level = 0.95)
```

```
##
## 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.

Depression rate in each state map

```
library(usmap)
library(ggplot2)

plot_usmap(data = dstate, values = "depression", color = "red") +
  scale_fill_continuous(name = "depression searches", label = scales::comma) +
  theme(legend.position = "right")
```

```
## 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.
```

