

GROUP 1 FINAL PAPER

Anthony Hu

6/22/2021

Data Import and Merging

```
Chicago_census <- read.csv(url("https://datahub.cmap.illinois.gov/dataset/1d2dd970-f0a6-4736-96a1-3caeb"))
Chicago_socionomic <- read.csv(url("https://data.cityofchicago.org/api/views/kn9c-c2s2/rows.csv"))
Chicago_schools <- read.csv(url("https://data.cityofchicago.org/api/views/9xs2-f89t/rows.csv"))

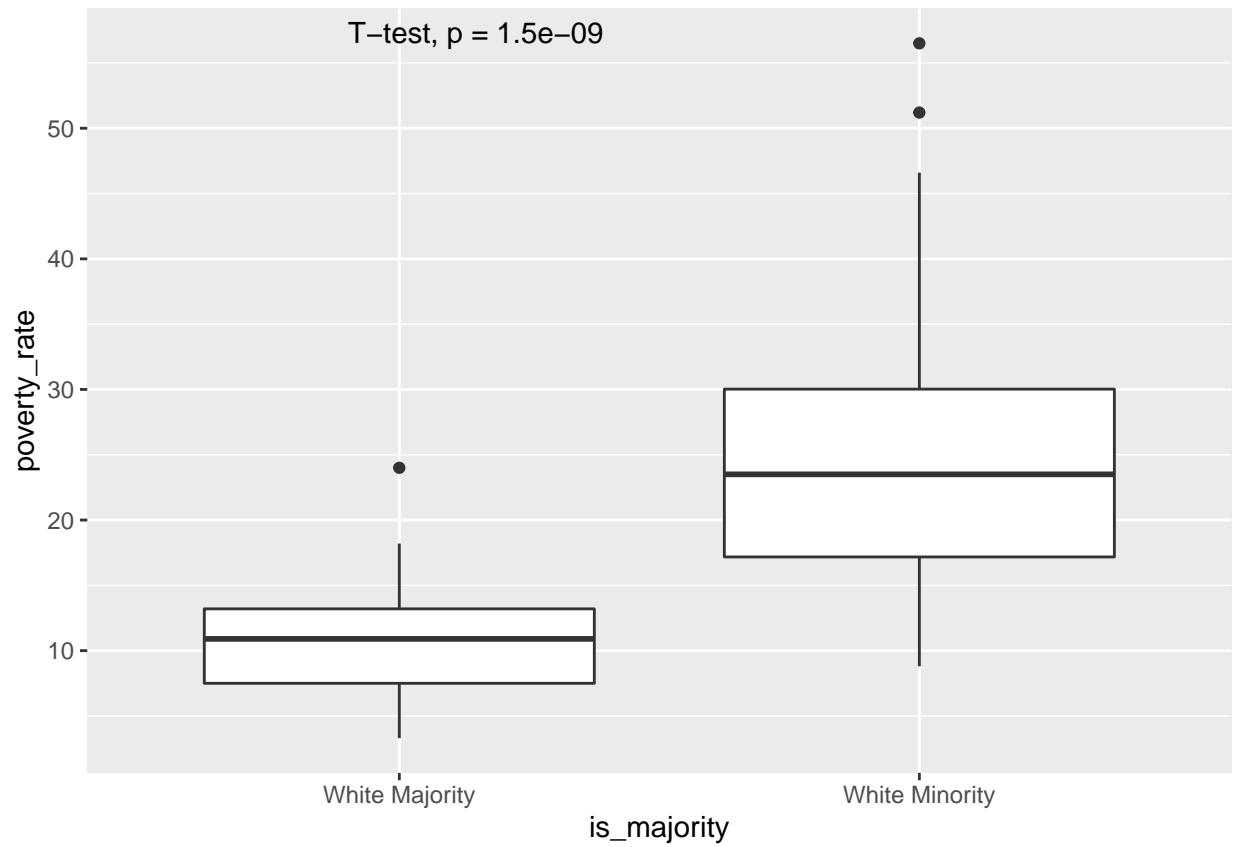
Chicago_combined <- Chicago_socionomic %>%
  rename(GEOG = "COMMUNITY.AREA.NAME") %>%
  left_join(Chicago_census, by = "GEOG")
```

Relationship between ethnicity and poverty rates

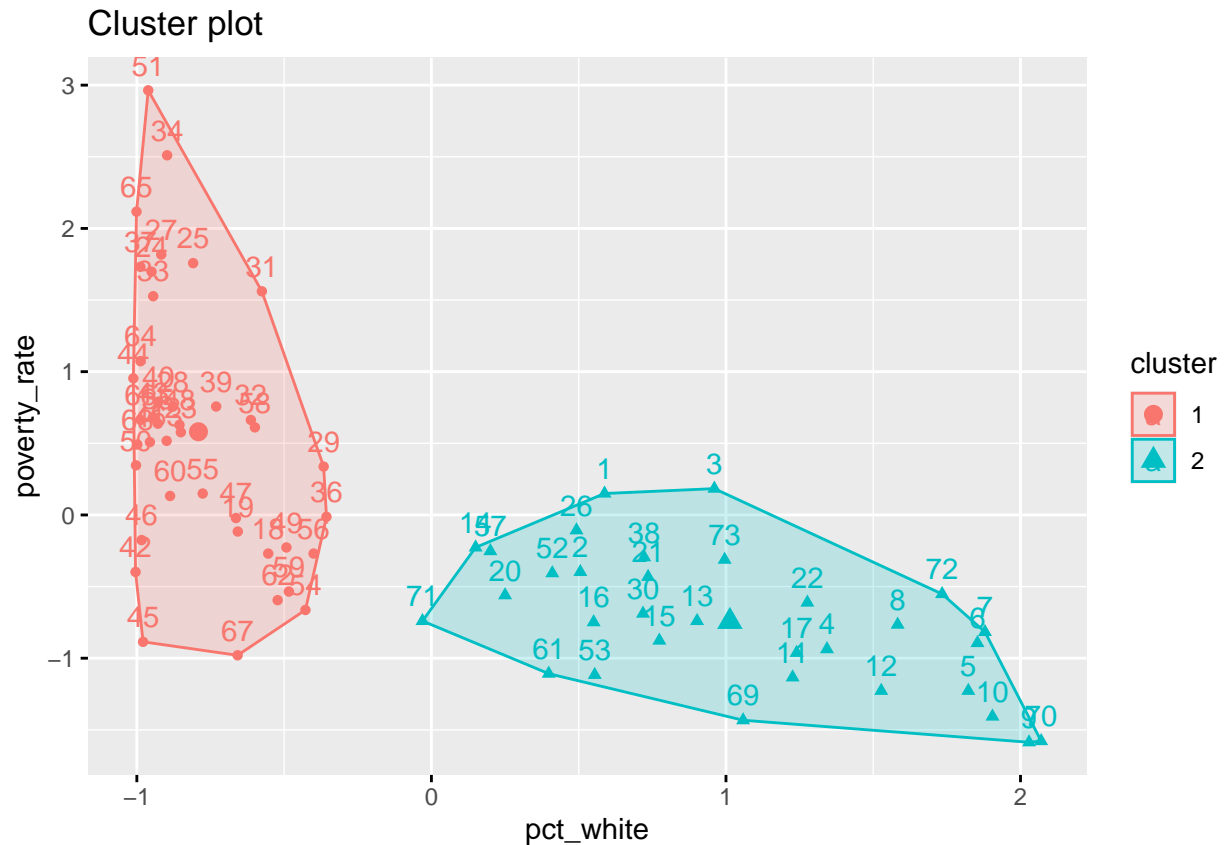
```
poverty <- Chicago_combined %>%
  mutate(pct_white = WHITE / TOT_POP) %>%
  rename(poverty_rate = PERCENT.HOUSEHOLDS.BELOW.POVERTY) %>%
  filter(pct_white > 0) %>%
  mutate(is_majority = ifelse(pct_white < 0.5, "White Minority", "White Majority")) %>%
  select(GEOG, poverty_rate, pct_white, is_majority)

a <- ggplot(data = poverty, mapping = aes(x = is_majority, y = poverty_rate)) +
  geom_boxplot()

a + stat_compare_means(method = "t.test")
```



```
A <- poverty%>%  
  select(pct_white, poverty_rate) %>%  
  na.omit()  
A1 <- scale(A)  
A2 <- kmeans(A1, centers = 2, nstart = 50)  
fviz_cluster(A2, data = A)
```

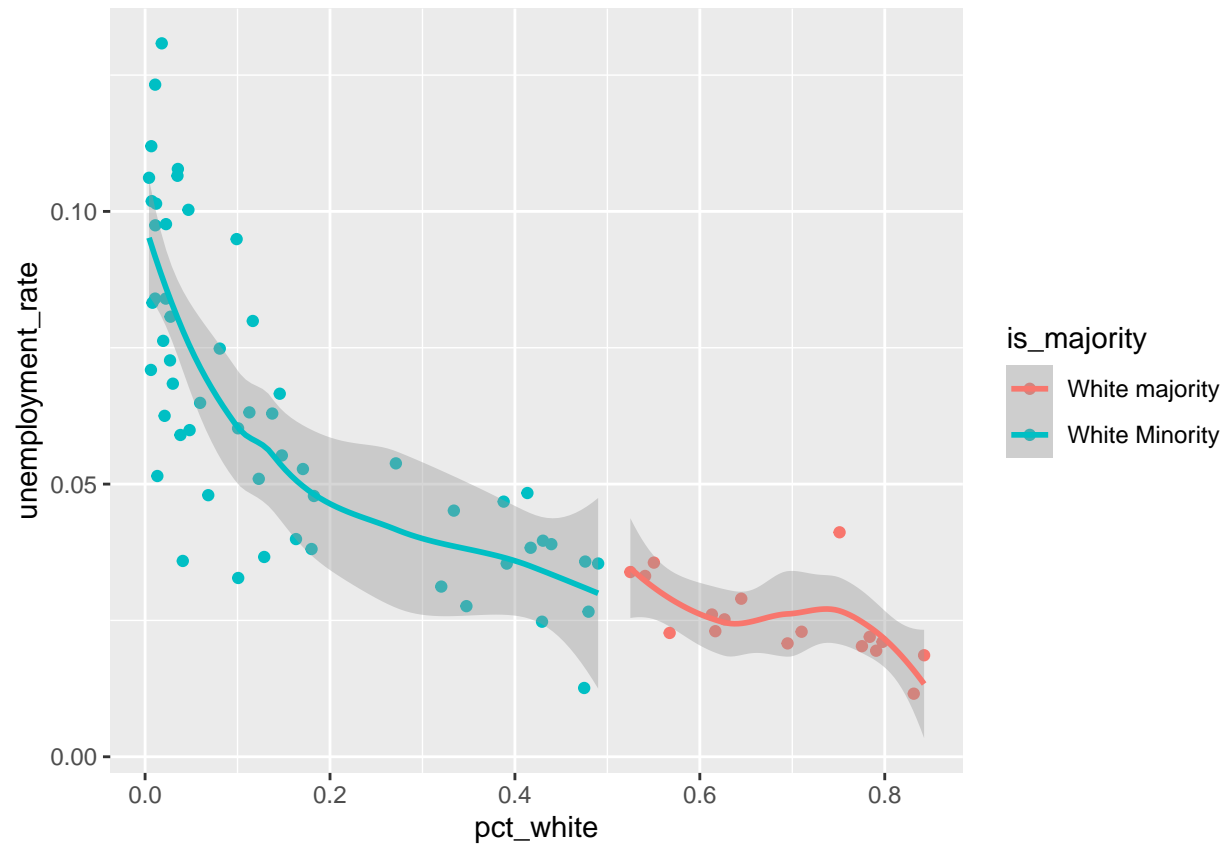


Relationship between ethnicity and unemployment rates

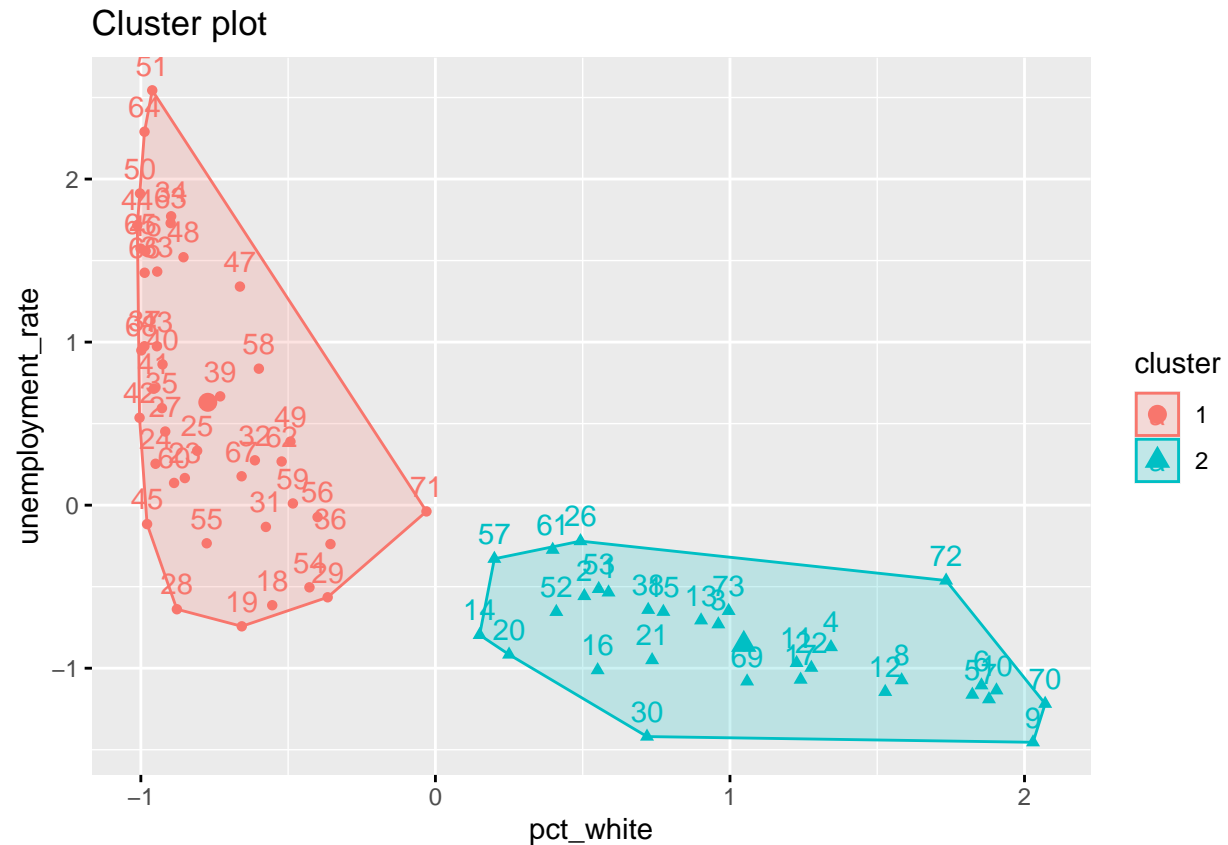
```
unemployment <- Chicago_combined %>%
  mutate(pct_white = WHITE / TOT_POP) %>%
  mutate(unemployment_rate = UNEMP / TOT_POP) %>%
  filter(pct_white > 0) %>%
  mutate(is_majority = ifelse(pct_white < 0.5, "White Minority", "White majority")) %>%
  select(GEOG, unemployment_rate, pct_white, is_majority)
```

```
ggplot(data = unemployment, mapping = aes(x = pct_white, y = unemployment_rate, color = is_majority)) +
  geom_point() +
  geom_smooth(method = "loess")
```

```
## 'geom_smooth()' using formula 'y ~ x'
```



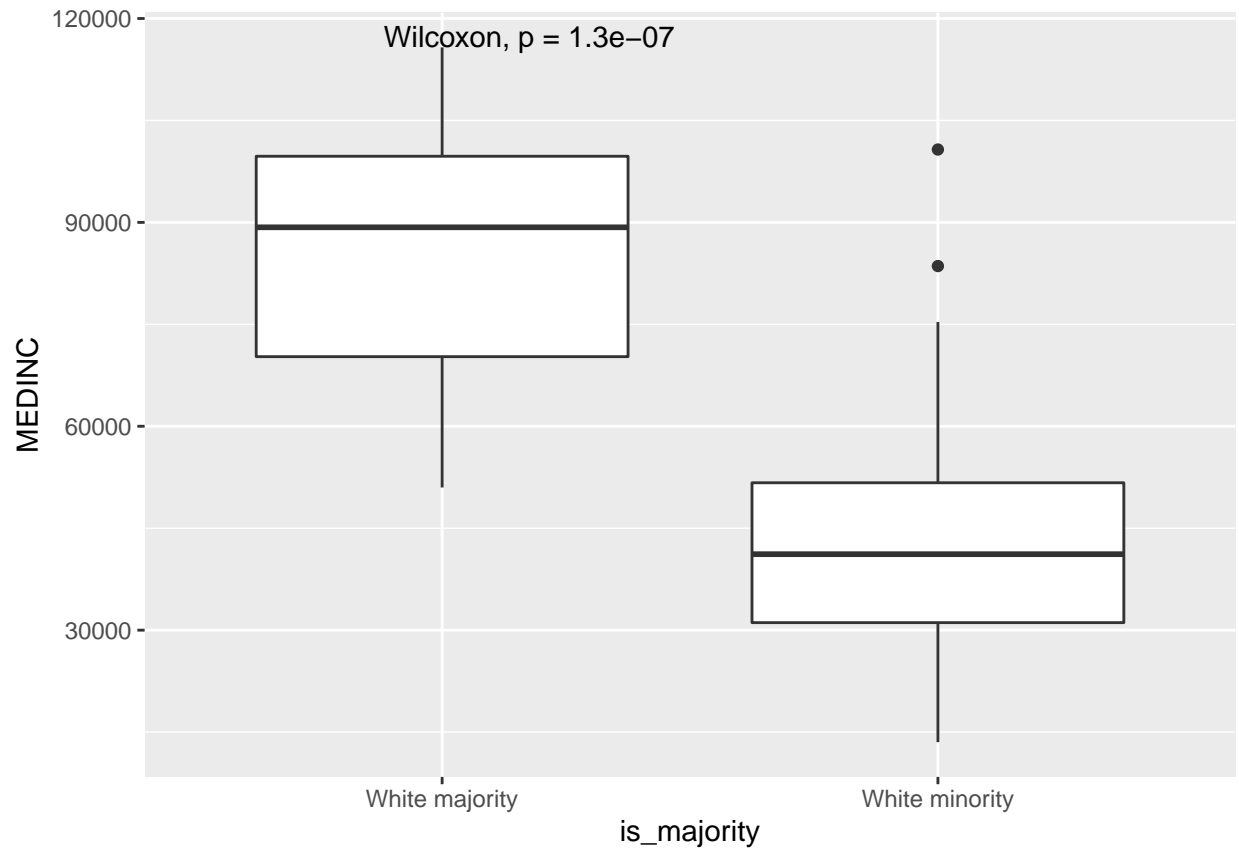
```
c <- unemployment %>%
  select(pct_white, unemployment_rate) %>%
  na.omit()
c1 <- scale(c)
c2 <- kmeans(c1, centers = 2, nstart = 50)
fviz_cluster(c2, data = c)
```



Relationship between ethnicity and median income.

```
income <- Chicago_combined %>%
  filter(WHITE > 0) %>%
  filter(TOT_POP > 0) %>%
  mutate(white_pct = WHITE/TOT_POP) %>%
  mutate(is_majority = ifelse(white_pct < 0.5, "White minority", "White majority")) %>%
  select(is_majority, everything())

a <- ggplot(data = income, mapping = aes(x = is_majority, y = MEDINC)) +
  geom_boxplot()
a + stat_compare_means()
```



```
a + stat_compare_means(method = "t.test")
```

