

Ethic distrubution against unemployment

Visualising US Census Data

Quinten van der Leest (s3546583)

4 May 2018

```
library(tinytex)
library(ggplot2)
library(dplyr)
library(ggmap)
library(googleVis)
library(rgdal)
library(tidyverse)
library(tidyr)
library(broom)
library(RColorBrewer)
library(cowplot)
```

Data

Reading in the Data, and converting the shp file to a data.frame. Also removed all States in the US_map_df that were not part of 'mainland' USA to make a cleaner plot.

```
US_census <- read.csv('acs2015_county_data.csv')
US_map <- rgdal::readOGR('cb_2017_us_state_5m/cb_2017_us_state_5m.shp')

## OGR data source with driver: ESRI Shapefile
## Source: "cb_2017_us_state_5m/cb_2017_us_state_5m.shp", layer: "cb_2017_us_state_5m"
## with 56 features
## It has 9 fields
## Integer64 fields read as strings:  ALAND AWATER

US_map_df <- tidy(US_map, region='NAME')
US_map_centre <- read.csv('centre data - Sheet2.csv')

US_census <- US_census[!US_census$State=="Hawaii",]
US_census <- US_census[!US_census$State=="Puerto Rico",]
US_census <- US_census[!US_census$State=="Alaska",]

colnames(US_map_df)[7] <- "State"

US_map_df <- US_map_df[!US_map_df$State=="Hawaii",]
US_map_df <- US_map_df[!US_map_df$State=="Puerto Rico",]
US_map_df <- US_map_df[!US_map_df$State=="Alaska",]
US_map_df <- US_map_df[!US_map_df$State=="American Samoa",]
US_map_df <- US_map_df[!US_map_df$State=="Commonwealth of the Northern Mariana Islands",]
US_map_df <- US_map_df[!US_map_df$State=="Guam",]
US_map_df <- US_map_df[!US_map_df$State=="United States Virgin Islands",]
```

Averaged every county value by state on the columns I was interested in visualising. Merged them into a State data.frame and then merge the map coord data and the state centre coord data with the State data.frame to create the final data frame used for plotting.

```

state_income <- summarise(group_by(US_census, State),
                           Income = mean(Income, na.rm = TRUE))
state_hispanic <- summarise(group_by(US_census, State),
                             Hispanic = mean(Hispanic, na.rm = TRUE))
state_white <- summarise(group_by(US_census, State),
                           White = mean(White, na.rm = TRUE))
state_black <- summarise(group_by(US_census, State),
                           Black = mean(Black, na.rm = TRUE))
state_native <- summarise(group_by(US_census, State),
                            Native = mean(Native, na.rm = TRUE))
state_asian <- summarise(group_by(US_census, State),
                           Asian = mean(Asian, na.rm = TRUE))
state_pacific <- summarise(group_by(US_census, State),
                             Pacific = mean(Pacific, na.rm = TRUE))
state_unemployment <- summarise(group_by(US_census, State),
                                  Unemployment = mean(Unemployment, na.rm = TRUE))

state <- cbind(state_income, state_hispanic[,2])
state <- cbind(state, state_white[,2])
state <- cbind(state, state_black[,2])
state <- cbind(state, state_native[,2])
state <- cbind(state, state_asian[,2])
state <- cbind(state, state_pacific[,2])
state <- cbind(state, state_unemployment[,2])

state$State <- as.character(state$State)

state <- cbind(state, US_map_centre[,2])

state <- cbind(state, US_map_centre[,3])

colnames(state)[10] <- "Lat_centre"

colnames(state)[11] <- "Long_centre"

US_final <- merge(US_map_df, state, by="State", all.x=TRUE)

US_final <- US_final[order(US_final$order), ]

```

Visualisation

The four plots produce investigate the population's ethnic distribution and how it relates to unemployment.

Plot 1

```

p1 <- ggplot(data = US_final,
             aes(x = long, y = lat, group = group,
                 fill = White)) +
  geom_polygon(color = "black", size = 0.25) +
  coord_map() +
  scale_fill_distiller(name = element_blank(),

```

```

        guide = "legend",
        palette = "PuBuGn", direction = 1) +
theme_minimal() + theme(axis.title.x = element_blank(),
                        axis.title.y = element_blank(),
                        axis.text.x = element_blank(),
                        axis.text.y = element_blank(),
                        panel.grid = element_blank()) +
labs(title="White Population (%)")

```

Plot 2

```

p2 <- ggplot(data = US_final,
             aes(x = long, y = lat, group = group,
                 fill = Black)) +
geom_polygon(color = "black", size = 0.25) +
coord_map() +
scale_fill_distiller(name = element_blank(),
                    guide = "legend",
                    palette = "PuBuGn", direction = 1) +
theme_minimal() + theme(axis.title.x = element_blank(),
                        axis.title.y = element_blank(),
                        axis.text.x = element_blank(),
                        axis.text.y = element_blank(),
                        panel.grid = element_blank()) +
labs(title="Black Population (%)")

```

Plot 3

```

p3 <- ggplot(data = US_final,
             aes(x = long, y = lat, group = group,
                 fill = Hispanic)) +
geom_polygon(color = "black", size = 0.25) +
coord_map() +
scale_fill_distiller(name = element_blank(),
                    guide = "legend",
                    palette = "PuBuGn", direction = 1) +
theme_minimal() + theme(axis.title.x = element_blank(),
                        axis.title.y = element_blank(),
                        axis.text.x = element_blank(),
                        axis.text.y = element_blank(),
                        panel.grid = element_blank()) +
labs(title="Hispanic Population (%)")

```

Plot 4 and the Plot_grid

```

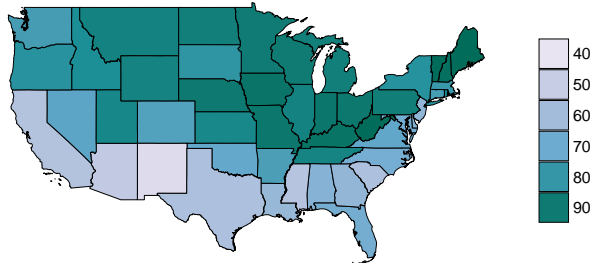
p4 <- ggplot(data = US_final,
             aes(x = long, y = lat, group = group,
                 fill = Unemployment)) +
geom_polygon(color = "black", size = 0.25) +
coord_map() +
scale_fill_distiller(name = element_blank(),
                    guide = "legend",
                    palette = "PuRd", direction = 1) +
theme_minimal() + theme(axis.title.x = element_blank(),
                        axis.title.y = element_blank(),
                        axis.text.x = element_blank(),

```

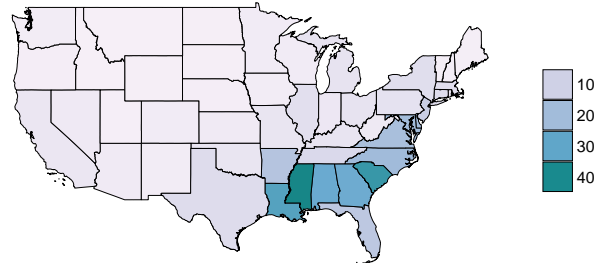
```
axis.text.y = element_blank(),
panel.grid = element_blank()) +
labs(title="Unemployed Population (%)")
```

```
plot_grid(p1,p2,p3,p4,ncol=2)
```

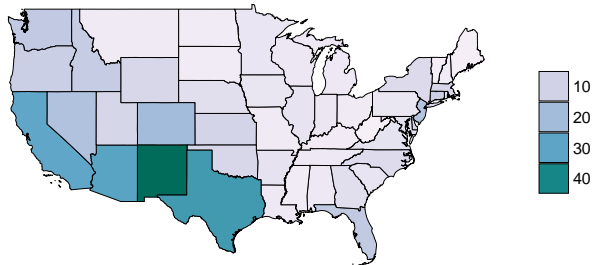
White Population (%)



Black Population (%)



Hispanic Population (%)



Unemployed Population (%)

