Homework 8

Wenxuan Zhou wz4388

This homework is due on April 19, 2021 at 11:00pm. Please submit as a pdf file on Canvas.

In this homework, we will work with two datasets, `US_counties` and `US_census`. The dataset `US_counties` contains the geometry of each county in the US and thus can be used for drawing maps. The dataset `US_census` contains numerous pieces of information about US counties obtained from the US census. Both datasets have a column `FIPS` which can be used to uniquely identify each county in each dataset.

```
# data preparation
US_counties <- readRDS(url("https://wilkelab.org/SDS375/datasets/US_counties.rds")) %>%
  rename(FIPS = GEOID)

# workaround for missing rgdal on edupod
st_crs(US_counties) <- NA

US_census <- read_csv(
  "https://wilkelab.org/SDS375/datasets/US_census.csv",
  col_types = cols(FIPS = "c")
)
```

Problem 1: (6 pts) Make a choropleth map of the percent home-ownership (column `home_ownership` in `US_census`) for all counties in the US. Choose an appropriate color scale and design for this plot. You may notice that there is one county in Alaska for which home-ownership data is not available. Write data analysis code to identify this county.

Hints:

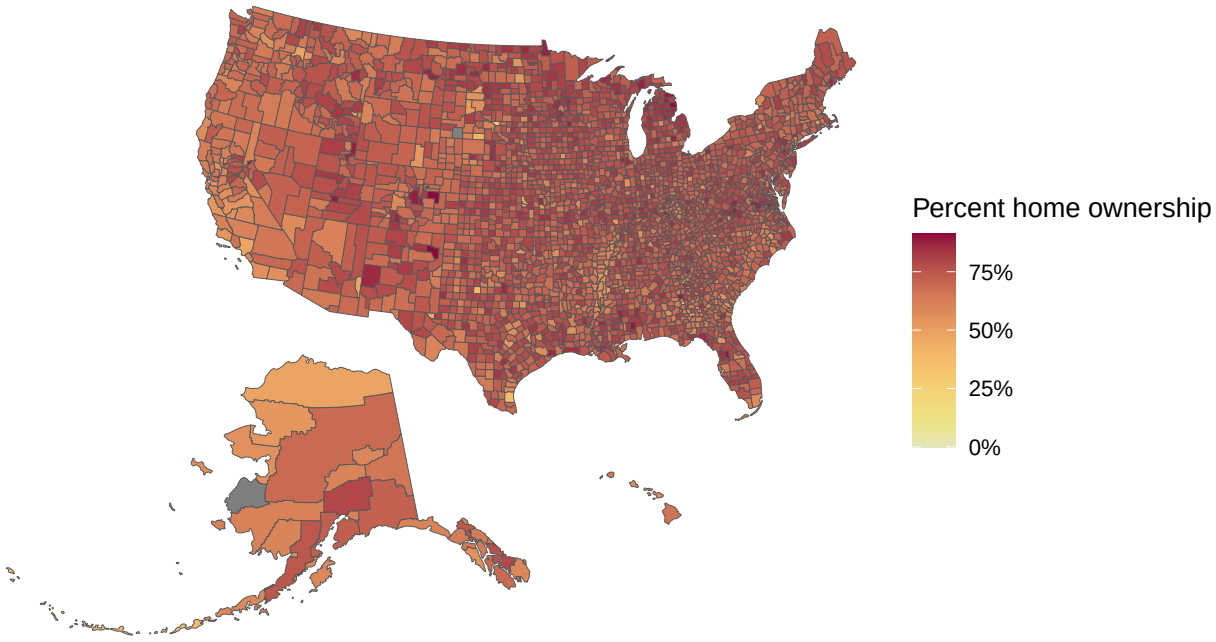
1. Use `theme_void()` as your theme
2. You will have to join `US_counties` and `US_census`. Join them by the `FIPS` column.
3. To make nice percent labels, you can use `label = scales::label_percent(scale = 1)` in your color scale function.
4. To find rows with missing data, you may want to use the function `is.na()`.

Grade breakdown: 2pt for the plot, 2pt for the plot design, and 2pt for identifying the county in Alaska for which home ownership data is not available.

```
US_census <- US_census %>% select(-name) # remove column 'name'

US_full <- left_join(US_counties, US_census, by = 'FIPS') # join the datasets

ggplot(US_full) +
  geom_sf(aes(fill = home_ownership), size = 0.1) +
  scale_fill_continuous_sequential(palette = "Heat",
                                   name = 'Percent home ownership',
                                   label = scales::label_percent(scale = 1)) +
  theme_void()
```




```
US_full$NAME[which(is.na(US_full$home_ownership))[2]]
```

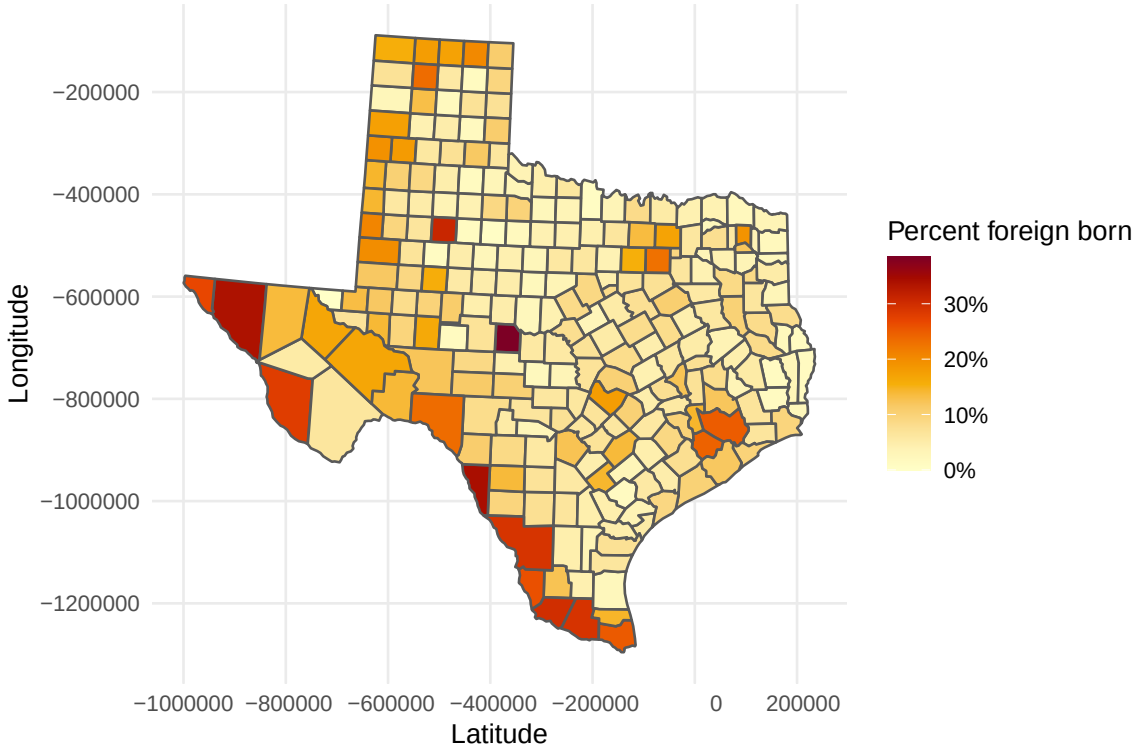
```
## [1] Kusilvak
## 1910 Levels: Abbeville Acadia Accomack Ada Adair Adams Addison ... Ziebach
```

The home-ownership data is not available for Kusilvak county in Alaska.

Problem 2: (4 pts) Make a choropleth map of the percent foreign born (column `foreign_born` in `US_census`) for the counties in Texas only. Use a different color scale than you used for Problem 1 and use a theme that shows longitude and latitude (nearly any theme other than `theme_void()` will work).

Grade breakdown: 3pt for the plot, 1pt for the plot design

```
US_full %>% filter(state %in% 'Texas') %>% ggplot() +
  geom_sf(aes(fill = foreign_born)) +
  scale_fill_continuous_sequential(palette = "YlOrRd",
                                   name = 'Percent foreign born',
                                   label = scales::label_percent(scale = 1)) +
  labs(x = 'Latitude', y = 'Longitude') +
  theme_minimal()
```


Longitude
−200000
−400000
−600000
−800000
−1000000
−1200000
−1000000
−800000
−600000
−400000
−200000
0
200000
Latitude
Percent foreign born
30%
20%
10%
0%