

Homework 8

Jeremy Ulfohn // jau392

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")
)

# first, join US_counties and US_census with FIPS. call new df 'joined'
joined <- left_join(US_counties, US_census, by = 'FIPS')
```

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.

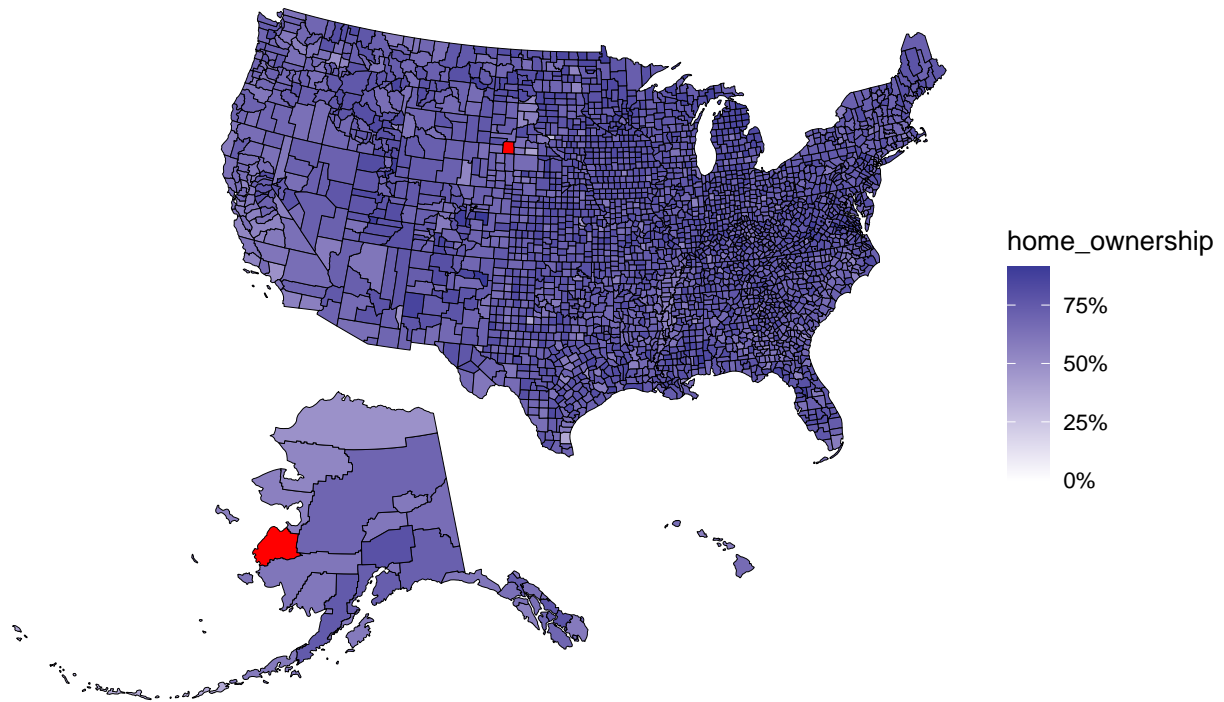
Hints:

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.

```
# make Choropleth of joined data
ggplot(joined) +
  geom_sf(
    aes(fill = home_ownership),
    color = "black", size = 0.1
  ) +
  scale_fill_gradient2(
    label = scales::label_percent(scale = 1),
    na.value = "red" # counties with data unavailable will have fill of RED
```

```
) +  
theme_void()
```



```
# find missing county in Alaska  
missing <- joined %>%  
  filter(is.na(home_ownership)) %>%  
  filter(name.x == "Alaska") %>%  
  select(name.x, NAME)  
missing
```

```
## Simple feature collection with 1 feature and 2 fields  
## geometry type:  MULTIPOLYGON  
## dimension:      XY  
## bbox:           xmin: -2156888 ymin: -1990730 xmax: -1839113 ymax: -1738003  
## CRS:            NA  
##   name.x      NAME                                geometry  
## 1 Alaska Kusilvak MULTIPOLYGON (((-2156888 -1...
```

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

```
# data wrangle for: foreign_born, name.x == "Texas", filter !is.na(foreign.born)  
mod <- joined %>%  
  filter(name.x == "Texas")  
  
ggplot(mod) +  
  geom_sf(
```

```

aes(fill = foreign_born),
color = "black", size = 0.1
) +
scale_fill_gradientn(
  colors = topo.colors(7),
  label = scales::label_percent(scale = 1),
  na.value = "white" # irrelevant for Texas
) +
theme_light(11)

```

