Stat 480 - Homework #11 (extra credit)

Vahid Azizi

4/23/2020

Submission Details

Due date: This homework is for 15 points of extra credit to be added to the midterm exam points and is due before class on Thursday.

Submission process: submit both the R Markdown file and the corresponding html file on canvas. Please submit both the `.Rmd` and the `.html` files separately and do not zip the two files together.

Measles Vaccination Rates

1. Download the RMarkdown file with these homework instructions to use as a template for your work. Make sure to replace "Your Name" in the YAML with your name.

2. This week we will return to the measles data from homework 7.

```
library(tidyverse)
measles <- read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-02-25/measles.csv')
head(measles)
```

```
## # A tibble: 6 x 16
##   index state year  name  type  city  county district enroll   mmr overall xrel
##   <dbl> <chr> <chr> <chr> <chr> <chr> <chr>  <lgl>     <dbl> <dbl>   <dbl> <lgl>
## 1     1 Ariz~ 2018~ A J ~ Publ~ Noga~ Santa~ NA           51   100      -1 NA
## 2     2 Ariz~ 2018~ Acad~ Char~ Tucs~ Pima   NA           22   100      -1 NA
## 3     3 Ariz~ 2018~ Acad~ Char~ Tucs~ Pima   NA           85   100      -1 NA
## 4     4 Ariz~ 2018~ Acad~ Char~ Phoe~ Maric~ NA           60   100      -1 NA
## 5     5 Ariz~ 2018~ Accl~ Char~ Phoe~ Maric~ NA           43   100      -1 NA
## 6     6 Ariz~ 2018~ Alfr~ Publ~ Phoe~ Maric~ NA           36   100      -1 NA
## # ... with 4 more variables: xmed <dbl>, xper <dbl>, lat <dbl>, lng <dbl>
```

3. Data clean up:

- Remove duplicates by keeping the distinct combinations of the variables `state`, `name`, `county`, `overall`, `mmr`, `lat`, `lng`, and `enroll`.
- Filter out schools outside the continent (use `lng` < 0).
- Use filtering to keep observations where `mmr` >= 0 or `overall` >= 0.
- Using the function `ifelse()`, create a new variable `rate` where the value is equal to the value of `mmr` when `mmr` >= 0 is true and equal to `overall` when `mmr` >= 0 is false.
- Save this data as `measles_new`

```
measles_one<-measles %>% group_by(state,name,county,overall,mmr,lat,lng,enroll) %>% distinct() %>% ungroup() %>% filter(lng<
0) %>% mutate(rate=ifelse(mmr>=0,mmr,overall))
head(measles_one)
```

```
## # A tibble: 6 x 17
##   index state year  name  type  city  county district enroll   mmr overall xrel
##   <dbl> <chr> <chr> <chr> <chr> <chr> <chr>  <lgl>     <dbl> <dbl>   <dbl> <lgl>
## 1     1 Ariz~ 2018~ A J ~ Publ~ Noga~ Santa~ NA           51   100      -1 NA
## 2     2 Ariz~ 2018~ Acad~ Char~ Tucs~ Pima   NA           22   100      -1 NA
## 3     3 Ariz~ 2018~ Acad~ Char~ Tucs~ Pima   NA           85   100      -1 NA
## 4     4 Ariz~ 2018~ Acad~ Char~ Phoe~ Maric~ NA           60   100      -1 NA
## 5     5 Ariz~ 2018~ Accl~ Char~ Phoe~ Maric~ NA           43   100      -1 NA
## 6     6 Ariz~ 2018~ Alfr~ Publ~ Phoe~ Maric~ NA           36   100      -1 NA
## # ... with 5 more variables: xmed <dbl>, xper <dbl>, lat <dbl>, lng <dbl>,
## #   rate <dbl>
```

4. Load the county map data using the function `map_data()` and save as `us_counties`.

```
us_counties<-map_data("county")
```

5. Additional data cleaning and data joining:

- Begin with the `measle_one` data created in question 3.
- Use the function `str_to_lower()` from the `stringr` package, to create a variable `subregion` that is the lowercase version of `county`.
- Use the function `str_to_lower()` from the `stringr` package, to create a variable `region` that is the lowercase version of `state`.
- For each `region` - `subregion` pair, use `summarise()` to calculate the mean of the variable `rate`.
- Create a new variable `threshold` that is `TRUE` when `rate` > 0 and `FALSE` otherwise.
- Use `full_join()` to join in the `us_counties` data by `region` and `subregion`.
- Save this data as `measles_two`.

```
measles_two <- measles_one %>% mutate(subregion=str_to_lower(county)) %>% mutate(region=str_to_lower(state)) %>% group_by(re
gion,subregion) %>% summarise(rate=mean(rate)) %>% mutate(threshold=ifelse(rate> 0,TRUE,FALSE)) %>% full_join(us_counties,c(
'region','subregion'))
measles_two
```

```
## # A tibble: 87,980 x 8
## # Groups:    region [49]
##     region subregion   rate threshold   long    lat group order
##     <chr>  <chr>      <dbl> <lgl>      <dbl> <dbl> <dbl> <int>
##  1 arizona apache      50.6 TRUE       -109.  36.0    68  3160
##  2 arizona apache      50.6 TRUE       -109.  35.0    68  3161
##  3 arizona apache      50.6 TRUE       -109.  34.6    68  3162
##  4 arizona apache      50.6 TRUE       -109.  33.8    68  3163
##  5 arizona apache      50.6 TRUE       -109.  33.8    68  3164
##  6 arizona apache      50.6 TRUE       -109.  33.8    68  3165
##  7 arizona apache      50.6 TRUE       -109.  33.8    68  3166
##  8 arizona apache      50.6 TRUE       -109.  33.8    68  3167
##  9 arizona apache      50.6 TRUE       -109.  33.8    68  3168
## 10 arizona apache      50.6 TRUE       -109.  33.7    68  3169
## # ... with 87,970 more rows
```

6. Draw a map of the us counties and map the variable `threshold` to fill. Set the outline of the counties to the color "#f4f4f8" and a size of 0.1. Apply the palette "Dark2" from RColorBrewer, name the scale "Rate > 95%", and set an `na.value` of "grey80". Apply the theme `theme_minimal()`.

```
measles_two %>% ggplot(aes(x=long,y=lat))+
  geom_polygon(aes(group=group,fill= threshold), na.value="grey80",color="#f4f4f8",size=.1,show.legend = TRUE)+scale_colour_
brewer(palette = "Dark2")+theme_minimal()
```
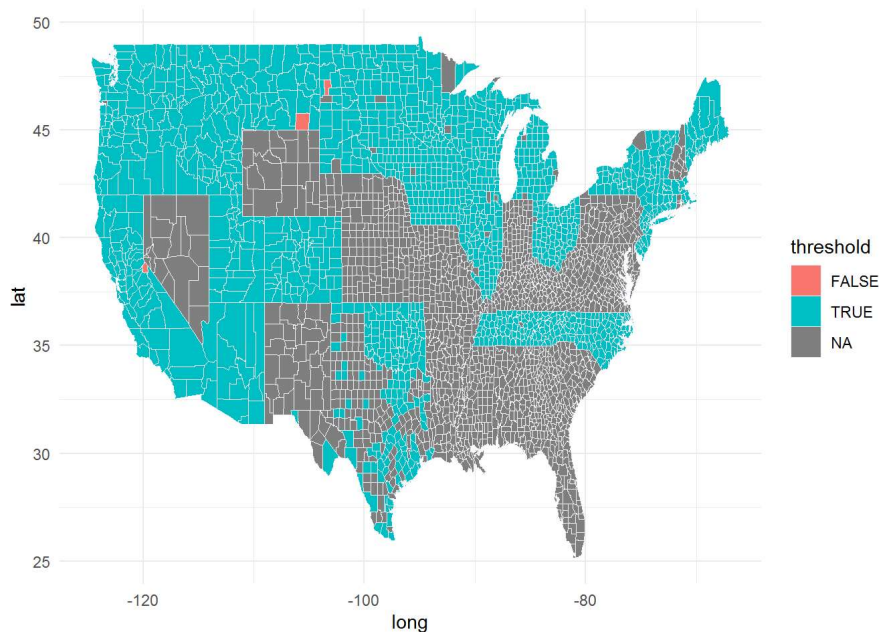

7. Use `geom_polygon()` to draw a map of the counties (don't use the `measles_one` data yet). Apply the following parameters (outside of `aes()`): `fill = "#f5f5f2"`, `color = "black"`, and `size = 0.1`. Save this plot as `plot1` and print.

```
plot1<-us_counties %>% ggplot(aes(x=long,y=lat))+
  geom_polygon(aes(group=group),fill= "#f5f5f2",color="black",size=.1,show.legend = FALSE)
plot1
```
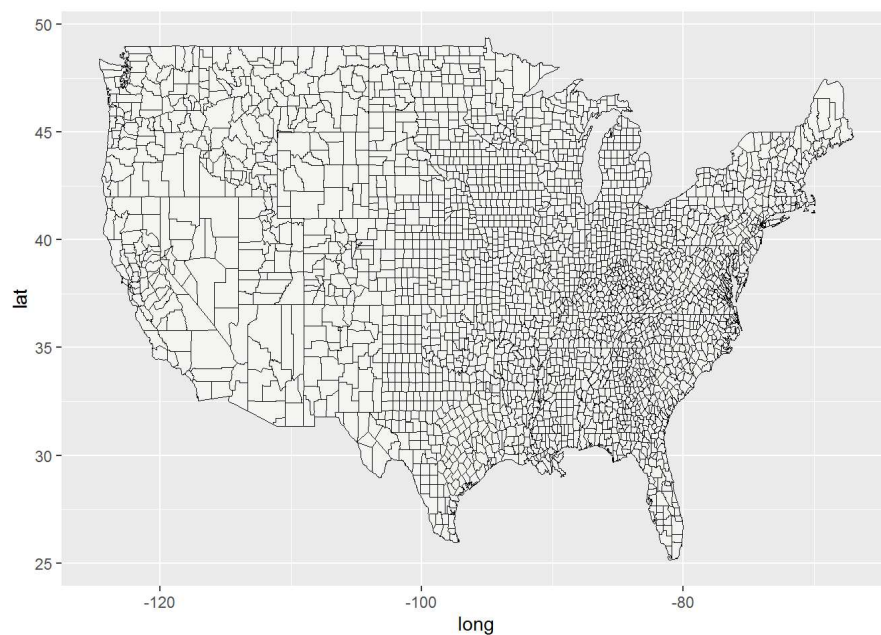
```
ggsave("plot1.png", plot = plot1)
```

8. Now we want to add on the `measles_one` data. Start with `plot1` and add an additional layer of points using the `measles_one` data created in question 3 with the points colored according the variable `rate`. Add an `alpha` of 0.5 and `size` of 0.5. Save this plot as `plot2` and print.

```
plot2 <- plot1 + geom_point(data=measles_one,aes(x=lng,y=lat,color=rate),alpha=0.5,size=0.5)
plot2
```
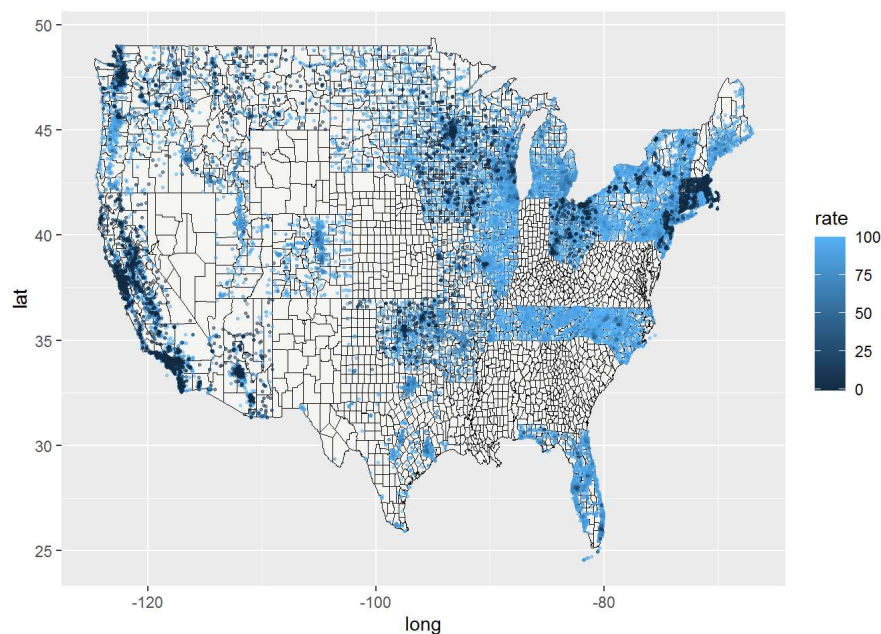


```
ggsave("plot2.png", plot = plot2)
```

9. Now we will add some labels. Start with `plot2` and add the title `"Schools' Vaccination Rate in 31 U.S. States"` and the caption
   `"Overall vaccination rate shown in \nFL, ID, IA, MI, NJ, NC, OK, RI, TN, WI. Elsewhere, the \nMeasles, Mumps, and Rubella (MMR) vaccination`
   Save this plot as `plot3` and print.

```
plot3 <- plot2 +
  labs(title = "Schools' Vaccination Rate in 31 U.S. States",
    caption = "Overall vaccination rate shown in \nFL, ID, IA, MI, NJ, NC, OK, RI, TN, WI. Elsewhere, the \nMeasles, Mumps,
 and Rubella (MMR) vaccination rate is shown.")

plot3
```

Schools' Vaccination Rate in 31 U.S. States

Overall vaccination rate shown in
FL, ID, IA, MI, NJ, NC, OK, RI, TN, WI. Elsewhere, the
Measles, Mumps, and Rubella (MMR) vaccination rate is shown.
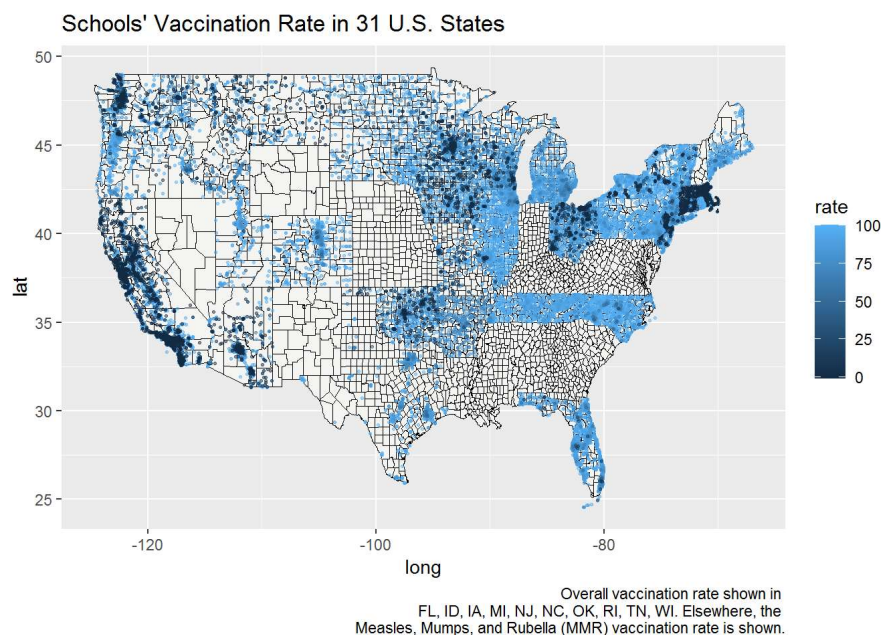
```
ggsave("plot3.png", plot = plot3)
```

10. Now we will modify the color scale. Add the color scale `scale_color_viridis_c()` with palette "plasma". Name the scale "Vaccination rate", reverse the direction (`direction = -1`), and add the argument
   `guide = guide_legend(direction = "horizontal", title.position = "top", title.hjust = 0.5)` . Save this plot as `plot4` and print.

```
plot4<-plot3 + scale_color_viridis_c(option = "plasma",direction = -1,guide = guide_legend(direction = "horizontal", title.p
osition = "top", title.hjust = 0.5))

plot4
```

Schools' Vaccination Rate in 31 U.S. States

Overall vaccination rate shown in
FL, ID, IA, MI, NJ, NC, OK, RI, TN, WI. Elsewhere, the
Measles, Mumps, and Rubella (MMR) vaccination rate is shown.
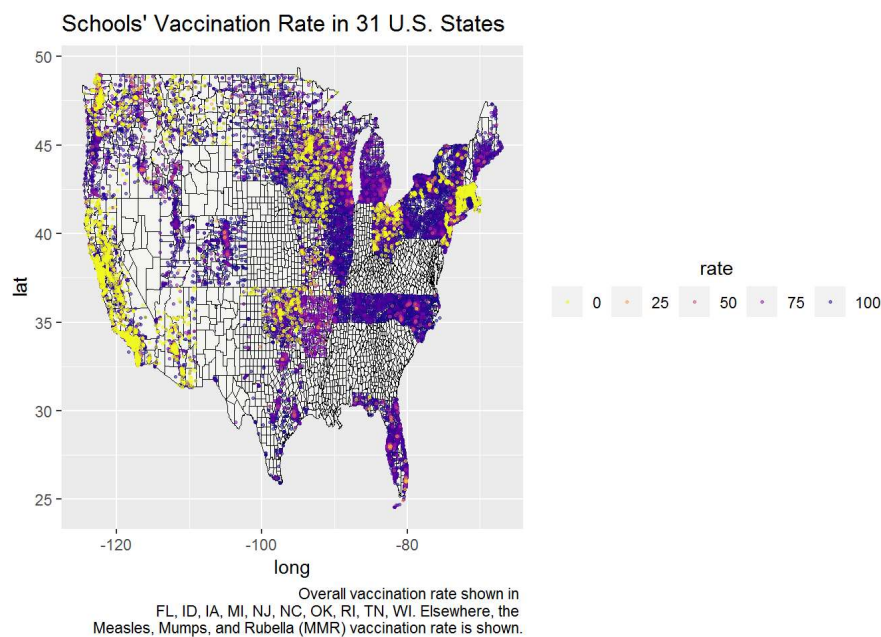
```
ggsave("plot4.png", plot = plot4)
```

11. The final step is to customize the theme. Let's first start with a blank slate by applying `theme_void()` to `plot4` . Next, add the following theme components with the function `theme()` and print your final map.

   - Move the legend to the bottom.
   - Move the plot title to the center by adding an `hjust` and `vjust` of 0.5.
   - Move the plot caption to the center by adding an `hjust` of 0.5.
   - Apply a plot background with a `fill = "#f5f5f2"` and `color = NA` .
   - Add the argument `plot.margin = margin(t = 1, r = 1, b = 1, l = 1, unit = "cm")` .
   - Add the argument `legend.margin = margin(t = 2, r = 2, b = 5, l = 2, unit = "mm")` .

```
plot4+theme_void()+theme(legend.position="bottom",plot.title=element_text(hjust=0.5,vjust=0.5), plot.caption = element_text
(hjust=0.5),plot.background = element_rect(fill = "#f5f5f2",color = NA),plot.margin = margin(t = 1, r = 1, b = 1, l = 1, uni
t = "cm"),legend.margin = margin(t = 2, r = 2, b = 5, l = 2, unit = "mm") )
```


Schools' Vaccination Rate in 31 U.S. States

rate
0   25   50   75   100

Overall vaccination rate shown in
FL, ID, IA, MI, NJ, NC, OK, RI, TN, WI. Elsewhere, the
Measles, Mumps, and Rubella (MMR) vaccination rate is shown.