Stat 480 - Homework #7

Vahid Azizi

3/5/2020

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. The data this week comes from The Wallstreet Journal (https://github.com/WSJ/measles-data). The data set includes immunization rate data for schools across the U.S. The accompanying article is published here (https://www.wsj.com/graphics/school-measles-rate-map/).

```
library(dplyr)
library(ggplot2)
library(readr)
measles <- read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-02-25/measles.csv')
```

3. For how many schools do we have data? How many of these schools recorded their Measles, Mumps, and Rubella (MMR) vaccination rate? Use the variable `mmr` to answer this question. Only consider schools with a rate > 0 for the remainder of the homework.

```
nrow(measles)
```

```
## [1] 66113
```

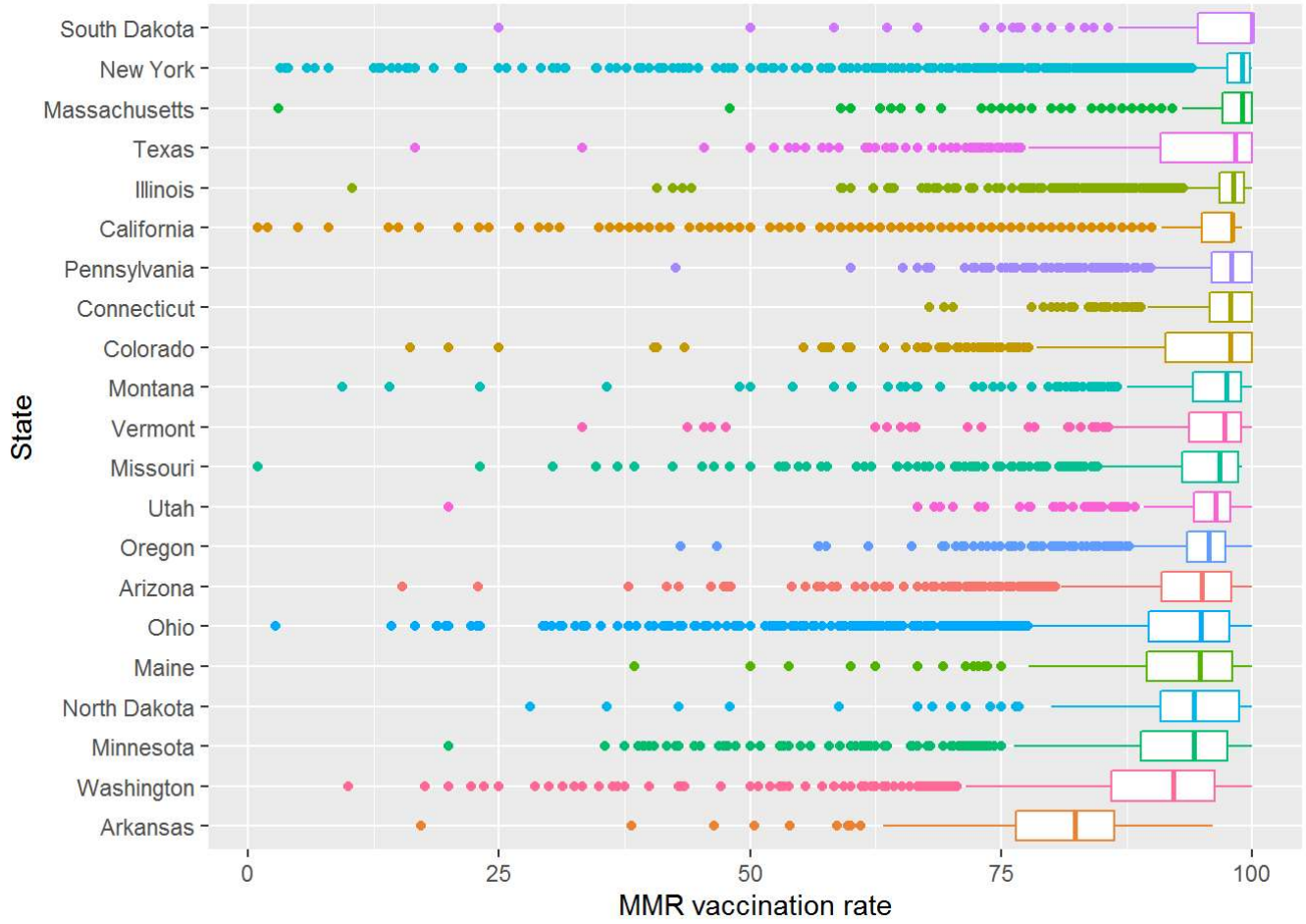
```
measles %>% filter(mmr> 0) %>% summarise(count = n())
```

```
## # A tibble: 1 x 1
##   count
##   <int>
## 1 44157
```

```
measles2<-measles %>% filter(mmr> 0)
```

4. Using `mutate()`, reorder the levels of the variable `state` according to the median MMR vaccination rate. Then “pipe” your results into ggplot and create box plots of MMR vaccination rates for each state. Map the variable `state` to `color`, include the parameter `show.legend = FALSE` within `geom_boxplot()`, and flip the coordinates. Interpret.

```
library(forcats)
measles2 %>% group_by(state) %>% mutate(med=median(mmr)) %>% arrange(desc(med)) %>%
  ggplot(aes(x=fct_reorder(state,med),y=mmr,color=state))+
  geom_boxplot(show.legend = FALSE)+
  coord_flip()+
  labs(x ="State", y = "MMR vaccination rate")
```



According to this plot all states have mmr vaccination rate's median larger than 75 percent. The best state in terms of mmr vaciination rate is South Dakota and Arkansas has the least rate of vaccination rate. California, New York, and ohio have the most outliers with mmr vaccination rate less than first quantile.

5. According to the CDC, 95% of a population needs to be vaccinated to stop the spread of measles and preserve herd immunity. Using `mutate()` and `case_when()`, introduce a new variable into the data set `mmr_threshold` where the value is "above" when `mmr` is greater than 95 and "below" otherwise. Is there a relationship between the type of school and the proportion of schools that did not reach that threshold? For each type of school, calculate the mean MMR vaccination rate. On how many responses are the averages based? Show these numbers together with the averages. Additionally, calculate the percentage of schools that did not reach that threshold. Arrange your results from greatest percentage to lowest. Comment on your results.

```
measles2 <- measles2 %>% mutate(mmr_threshold=as.factor(case_when(mmr>95 ~ "above",TRUE ~ "belo
w")))

measles2 %>% group_by(type)  %>% summarise(Total_number=n(),proportion_below=sum(mmr_threshold==
"below")/n(),mean_rate=mean(mmr)) %>% slice(1:6) %>% arrange(desc(proportion_below))
```

```
## # A tibble: 6 x 4
##   type         Total_number proportion_below mean_rate
##   <chr>               <int>            <dbl>     <dbl>
## 1 Charter               217           0.733       88.0
## 2 Private              4597           0.462       93.3
## 3 Kindergarten         1486           0.380       94.2
## 4 Nonpublic              18           0.278       94.4
## 5 Public              19762           0.239       96.2
## 6 BOCES                  47           0.0426      98.8
```

Charter, private, and kindergarten schools are the top three schools that their mmr vaccination rate is below the threshold. The number of public schools available is by far greater than the other types of schools. However, there is a relatively small proportion of schools in this category that their mmr vaccination rate is less than the threshold.

6. Use `dplyr` functions to:

- only include observations with enrollment greater than 0 and exclude the school “West Valley School Prek-6” (there is an issue with that observation)
- filter for rows that have a unique combination of the variables `year`, `city`, `state`, name, `type`, `enroll`, and `mmr` (there are duplicates in the data)
- Inside `mutate()` use `weighted.mean()` to calculate the mean MMR vaccination rates weighted by the enrollment. Name this new variable `state_avg`.

```
results<- measles2 %>% filter(enroll>0,!name=="West Valley School Prek-6") %>% distinct(year,cit
y,state,name,type,enroll,mmr) %>% group_by(city,state) %>% mutate(state_avg=weighted.mean(mmr,en
roll))
```

- For each city and state combination, calculate the mean MMR vaccination rate weighted by enrollment, the total number of students enrolled, and the mean of the state average calculated in the previous step.
- only consider rows where the total enrollment is more than 250 and less than 100,000.

```
results %>% filter(!is.na(city),!is.na(state))  %>% group_by(city,state) %>% summarise(mean_mmr=
weighted.mean(mmr,enroll),number_enroll=n(),mean_state_avg=mean(state_avg)) %>% filter(number_en
roll>250,number_enroll<1000)
```

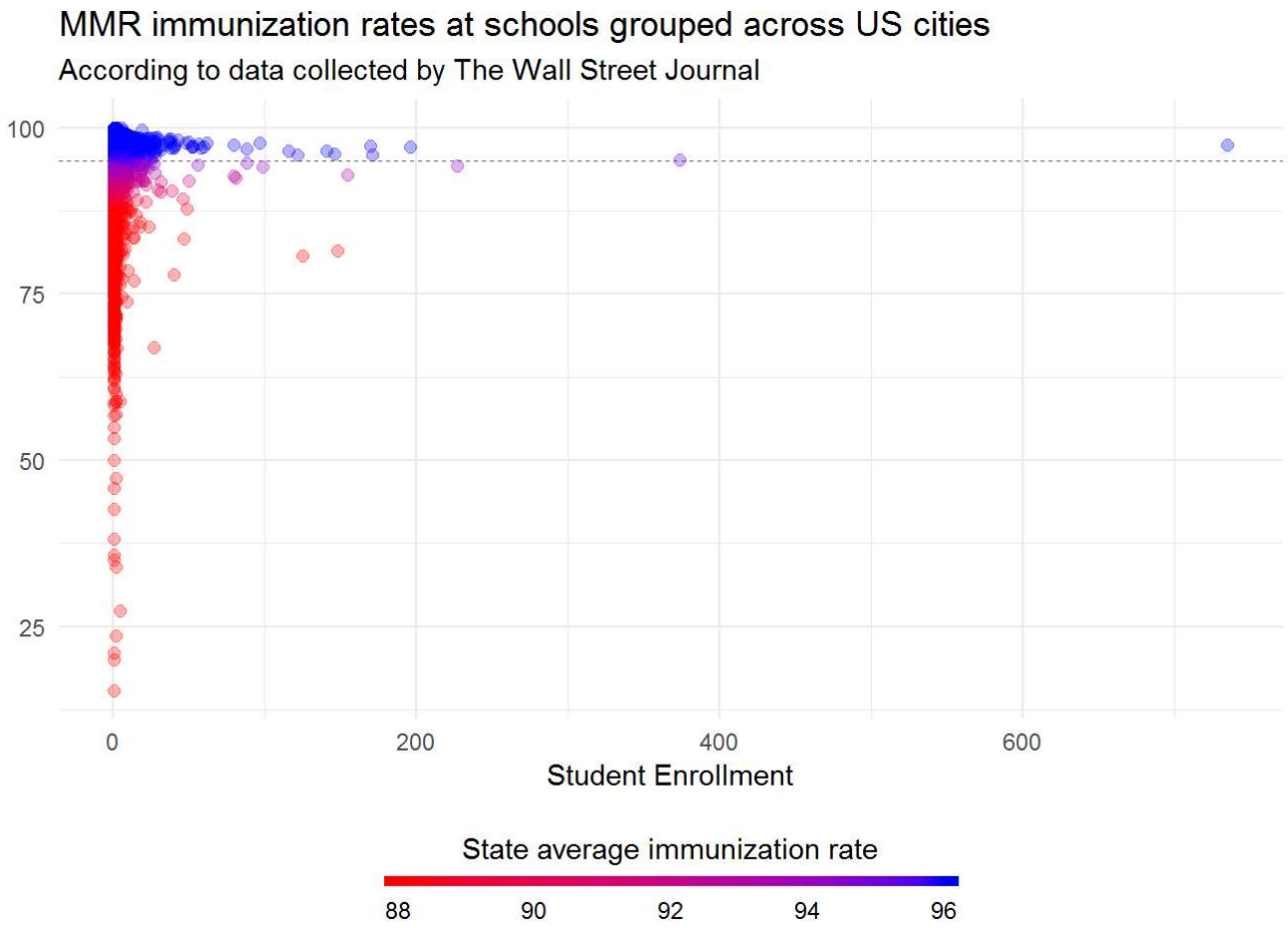
```
## # A tibble: 2 x 5
## # Groups:   city [2]
##   city        state      mean_mmr number_enroll mean_state_avg
##   <chr>       <chr>         <dbl>         <int>          <dbl>
## 1 Chicago     Illinois       97.4           735           97.4
## 2 Los Angeles California     95.1           374           95.1
```

```
question6_data <- results %>% filter(!is.na(city),!is.na(state))  %>% group_by(city,state) %>% s
ummarise(mean_mmr=weighted.mean(mmr,enroll),number_enroll=n(),mean_state_avg=mean(state_avg))
```

7. Now use the previous data set to draw a scatter plot with the mean MMR rate for each city on the y-axis and the student enrollment on the x-axis and color by the state mean MMR rate. Use the code below as your starting point and add in the necessary aesthetic mappings within `ggplot(aes( ))`. Describe and summarise the chart.

```
question6_data %>%
  ggplot(aes( )) +
  geom_hline(yintercept = 95, linetype = "dashed", size = 0.25, color = "grey40") +
  geom_point(size = 2, alpha = .3) +
  scale_color_gradient(low = "red", high = "blue", limits=c(88, 96), oob = scales::squish,
  guide = guide_colorbar(direction = "horizontal", title.position = "top",
                         title = "State average immunization rate", barwidth = 15, barheight = 0.
25,
                         ticks = FALSE, title.hjust = 0.5)) +
  theme_minimal() +
  theme(legend.position = "bottom") +
  ggtitle("MMR immunization rates at schools grouped across US cities") +
  labs(subtitle="According to data collected by The Wall Street Journal",
       x = "Student Enrollment", y = "") +
  scale_x_continuous(labels = scales::comma)
```

```
question6_data %>%
  ggplot(aes(y= mean_mmr, x=number_enroll, color=mean_state_avg)) +
  geom_hline(yintercept = 95, linetype = "dashed", size = 0.25, color = "grey40") +
  geom_point(size = 2, alpha = .3) +
  scale_color_gradient(low = "red", high = "blue", limits=c(88, 96), oob = scales::squish,
  guide = guide_colorbar(direction = "horizontal", title.position = "top",
                         title = "State average immunization rate", barwidth = 15, barheight = 0.
25,
                         ticks = FALSE, title.hjust = 0.5)) +
  theme_minimal() +
  theme(legend.position = "bottom") +
  ggtitle("MMR immunization rates at schools grouped across US cities") +
  labs(subtitle="According to data collected by The Wall Street Journal",
       x = "Student Enrollment", y = "") +
  scale_x_continuous(labels = scales::comma)
```



As can be seen in this plot, most of the cities has less than 200 students (just 3 of them are larger than 200). Apparently, there are more cities with mean mmr vaccination rate less than threshold but one cannot say it for sure based on this plot. There are small number of cities with mean mmr vaccination rate less than 50 percent.