R assignment for Ben

Monday, August 04, 2014

Do this assignment. If you have questions, google them. If you spend more than 2 minutes trying to figure out 1 question, email me.

1. Download immunization rate data and read it into R, naming the dataframe ir.

You can find it here: https://docs.google.com/spreadsheets/d/1KpUK8IX0E5L9MhGvdS5jbtqDh2GB0PGx4iiwux89Nqk/export?&format=csv

You can download it and then use the `read.csv()` function like this.

```
ir <- read.csv("pathway/to/where/you/put/the/file.csv")
```

Or, you can use the below script to read it in directly from the internet into R:

```
#install.packages("RCurl") # only run this line once
require(RCurl)
```

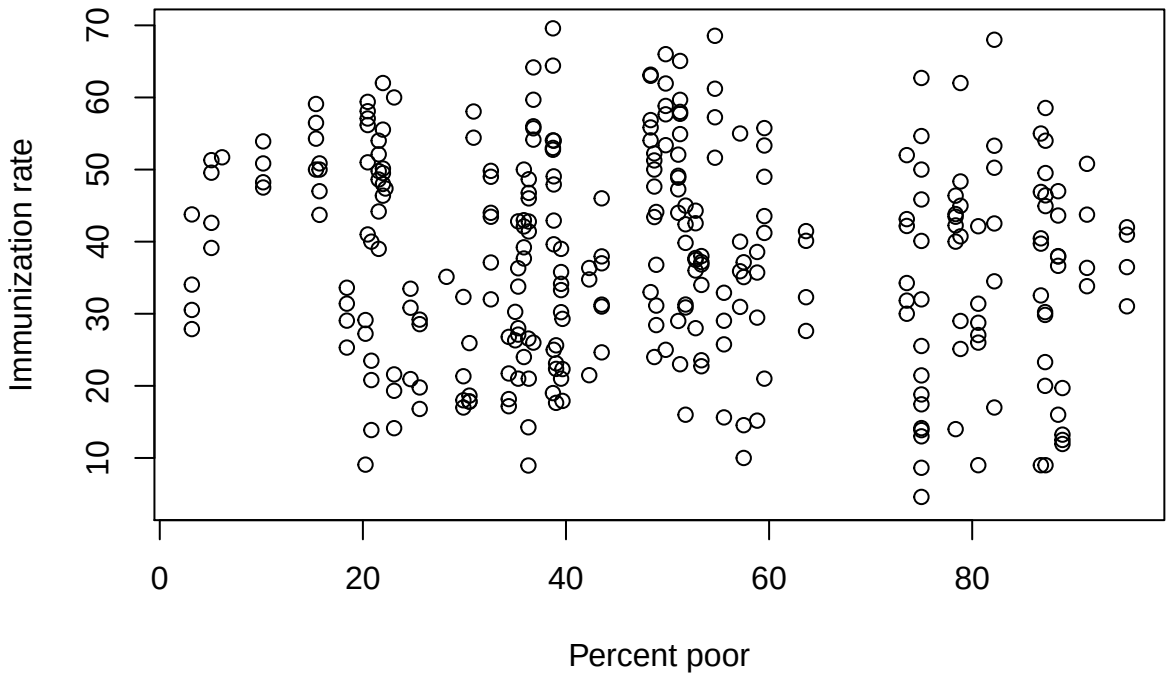
```
## Loading required package: RCurl
## Loading required package: bitops
```

```
options(RCurlOptions = list(cainfo = system.file("CurlSSL", "cacert.pem", package = "RCurl")))

myLink <- paste0("https://docs.google.com/spreadsheets/d/1KpUK8IX0E5L9MhGvdS5jbtqDh2GB0PGx4ii",
                 "wux89Nqk/export?&format=csv")

myCsv <- getURL(myLink)
ir <- read.csv(textConnection(myCsv))
```

2. Use the `View()` function to examine the data a bit.

You'll notice 10 columns. "doses" is the number of kids immunized, "totMem", is the "total membership" (number of kids at school), "p.vfc" is the percentage of kids who qualify for the "Vaccines for Children" program (basicall, it means poverty).

3. Make a plot that shows the relationship between p.vfc (poor kids) and immnunization rate. I should look like this



4. What is the mean immunization rate for all schools (all years)?

Use the `mean()` function.

5 What is the weighted mean immunization rate for all schools (all years)?

You don't want to give each school the same weight, since getting 100 immunization at a school of 10 people is very different from doing so at a school of 1000 people. Use the `weighted.mean()` function.

6. How many total students has the program immunized?

Use the `sum()` function.

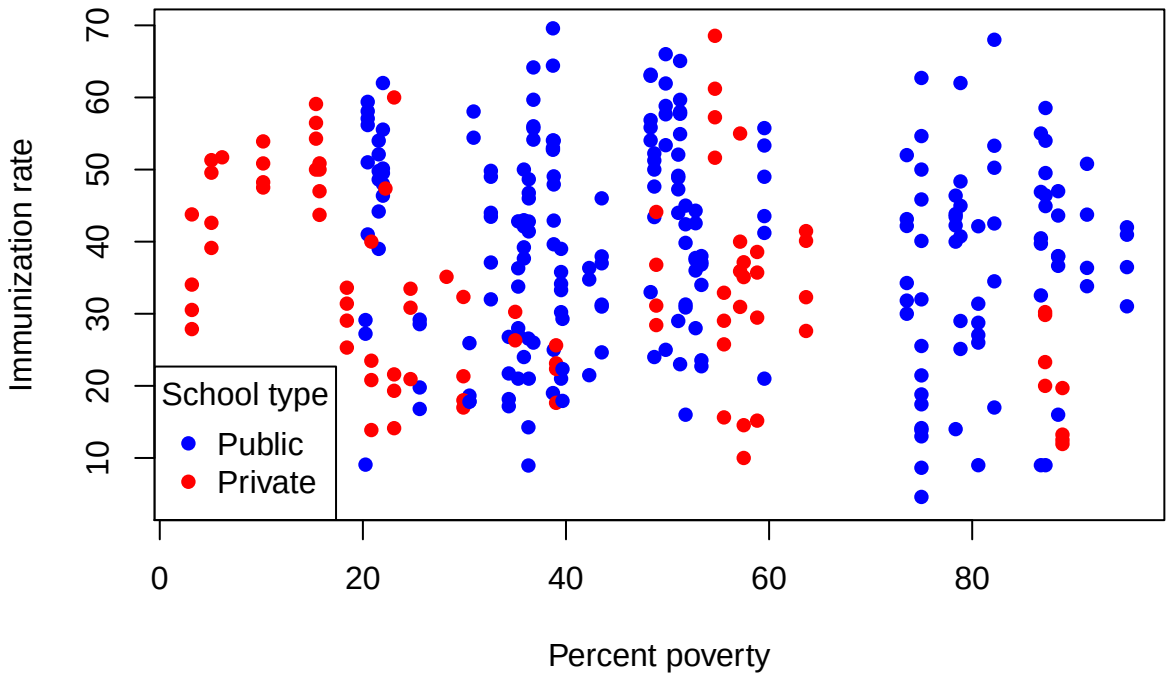
7. How many students did we immunize in 2013?

Use the `sum()` function, as well as the `which()` function. You can either subset the data first (to make a dataframe of only 2013 schools), or you can through the which statement within the sum function.

8. Which had higher weighted immunization rates in 2013 - private or public schools?

It's probably best to subset first into two dataframes. Name them priv13 and pub13. Then get the `weighted.mean()` function to calculate each.

9. Create a column in ir called color. Use the `ifelse()` function to assign to that column "blue" if it's a public school and "red" if it's a private school. Then make the following plot.



10. Create a linear model using the `lm()` function in which the outcome is immunization rate, and the independent variable is percentage of kids who qualify for VFC. Name the model lm1. Then examine it using `summary(lm1)`. What does this mean?

11. Extra credit: Add other relevant variables (such as school type and whether the school is public or private). Name your new model lm2. What is the effect of poverty on immunization rates when you adjust for pubPriv and age of kids (type)? What about the effect of public vs. private school on immunization rates when you adjust for poverty and age?