Vaccination Rates LA County

Oscar Monroy

10/30/2021

Step 1: Importing and Fixing the Data

Variable name on R vs. Variable name on Excel:

X12..with.1..Dose = 12+ with 1+ Dose

Population..12.. = Population (12+)

X12..Pop..Vaccinated.... = 12+ Pop. Vaccinated (%)

```
la_c <- read.csv("LA-County-COVID-19-Vaccine-Dashboard-Community-2021-10-24.csv")
head(la_c) # Looks messy, for R standards
```

```
##         City.Community X12..with.1..Dose Population..12..
## 1 City of Agoura Hills            15,212           18,920
## 2     City of Alhambra            64,216           76,362
## 3      City of Arcadia            45,160           53,504
## 4      City of Artesia            12,420           15,043
## 5       City of Avalon               760            3,266
## 6        City of Azusa            30,563           43,954
##   X12..Pop..Vaccinated....
## 1                    80.4%
## 2                    84.1%
## 3                    84.4%
## 4                    82.6%
## 5                    23.3%
## 6                    69.5%
```

```
dim(la_c) # Dimeensions of our data frame
```

```
## [1] 345   4
```

```
la_c$City.Community <- as.character(la_c$City.Community)
# For the next 3 lines, we'll change the class of
# the last 3 variables to numeric to make things easier.
la_c$X12..with.1..Dose <- as.numeric(gsub("[\\,]", "", la_c$X12..with.1..Dose))
```

```
## Warning: NAs introduced by coercion
```

```r
la_c$Population..12.. <- as.numeric(gsub("[\\,]", "", la_c$Population..12..))
```

```
## Warning: NAs introduced by coercion
```

```r
la_c$X12..Pop..Vaccinated.... <- as.numeric(gsub("[\\%,]", "", la_c$X12..Pop..Vaccinated....))
```

```
## Warning: NAs introduced by coercion
```

```r
head(la_c) # Looks much better now
```

```
##         City.Community X12..with.1..Dose Population..12..
## 1 City of Agoura Hills             15212            18920
## 2     City of Alhambra             64216            76362
## 3      City of Arcadia             45160            53504
## 4      City of Artesia             12420            15043
## 5       City of Avalon               760             3266
## 6        City of Azusa             30563            43954
##   X12..Pop..Vaccinated....
## 1                     80.4
## 2                     84.1
## 3                     84.4
## 4                     82.6
## 5                     23.3
## 6                     69.5
```

Step 2: Cleaning the Data

```r
summary(la_c)
```

```
##  City.Community     X12..with.1..Dose Population..12.. X12..Pop..Vaccinated....
##  Length:345         Min.   :    10    Min.   :    11   Min.   :17.60
##  Class :character   1st Qu.:  2659    1st Qu.:  3233   1st Qu.:70.58
##  Mode  :character   Median : 12120    Median : 13422   Median :76.30
##                     Mean   : 20641    Mean   : 25965   Mean   :74.32
##                     3rd Qu.: 28727    3rd Qu.: 35430   3rd Qu.:81.40
##                     Max.   :300966    Max.   :404580   Max.   :99.50
##                     NA's   :21        NA's   :5        NA's   :21
```

```r
# Looks about right considering this is LA County,
# however, those NA's need to be taken care of.
nr <- which(is.na(la_c$X12..with.1..Dose) == TRUE) # Checks index of NA's
nr
```

```
##  [1]  36  82 100 115 152 175 177 219 227 238 241 250 265 268 298 308 314 325 330
## [20] 334 343
```

```r
which(is.na(la_c$Population..12..) == TRUE)
```

```
## [1] 227 238 268 325 330
```

```r
which(is.na(la_c$X12..Pop..Vaccinated....) == TRUE)
```

```
##  [1]  36  82 100 115 152 175 177 219 227 238 241 250 265 268 298 308 314 325 330
## [20] 334 343
```

```r
# Appears that all our NA's happen on the same observations...
la_c2 <- la_c[-nr, ]
summary(la_c2) # No NA's; success
```

```
##  City.Community     X12..with.1..Dose Population..12.. X12..Pop..Vaccinated....
##  Length:324         Min.   :    10    Min.   :    20   Min.   :17.60
##  Class :character   1st Qu.:  2659    1st Qu.:  3649   1st Qu.:70.58
##  Mode  :character   Median : 12120    Median : 15070   Median :76.30
##                     Mean   : 20641    Mean   : 26944   Mean   :74.32
##                     3rd Qu.: 28727    3rd Qu.: 36333   3rd Qu.:81.40
##                     Max.   :300966    Max.   :404580   Max.   :99.50
```

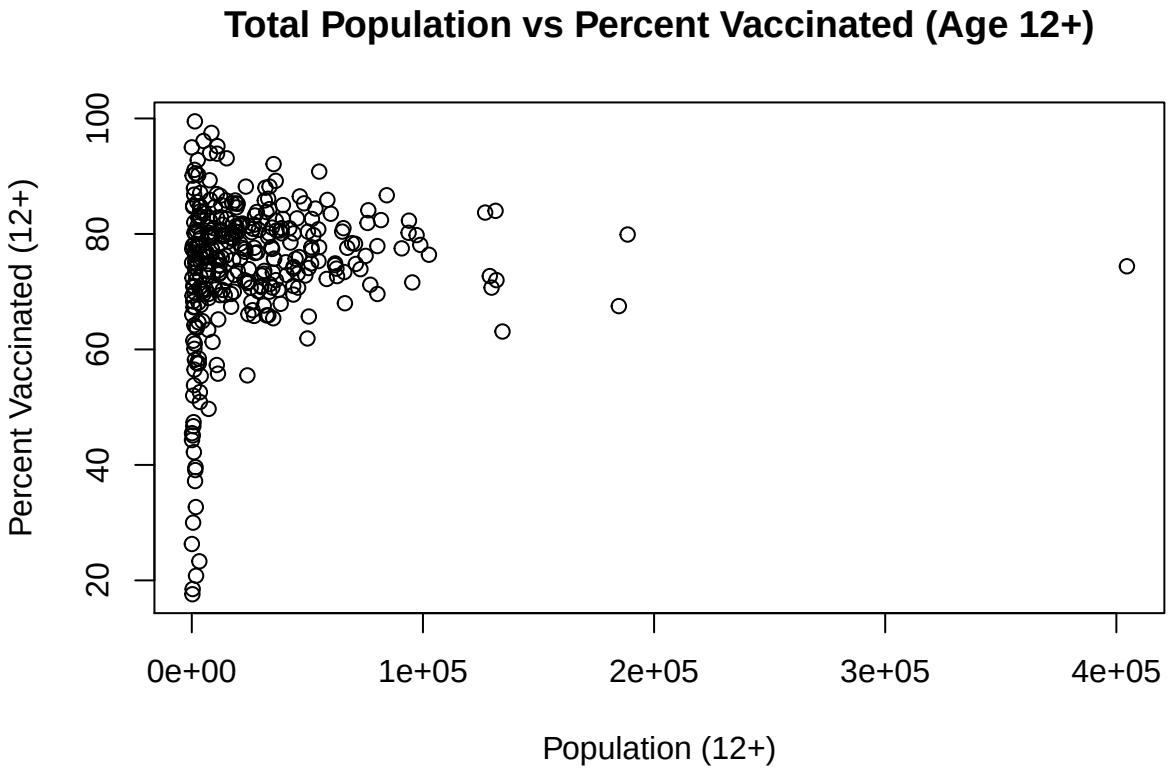
```r
# Now we check for duplicate observations
which(duplicated(la_c2) == TRUE) # No duplicates
```

```
## integer(0)
```

```r
# We are done with cleaning the data
```

Step 3: Visualizations and Analysis

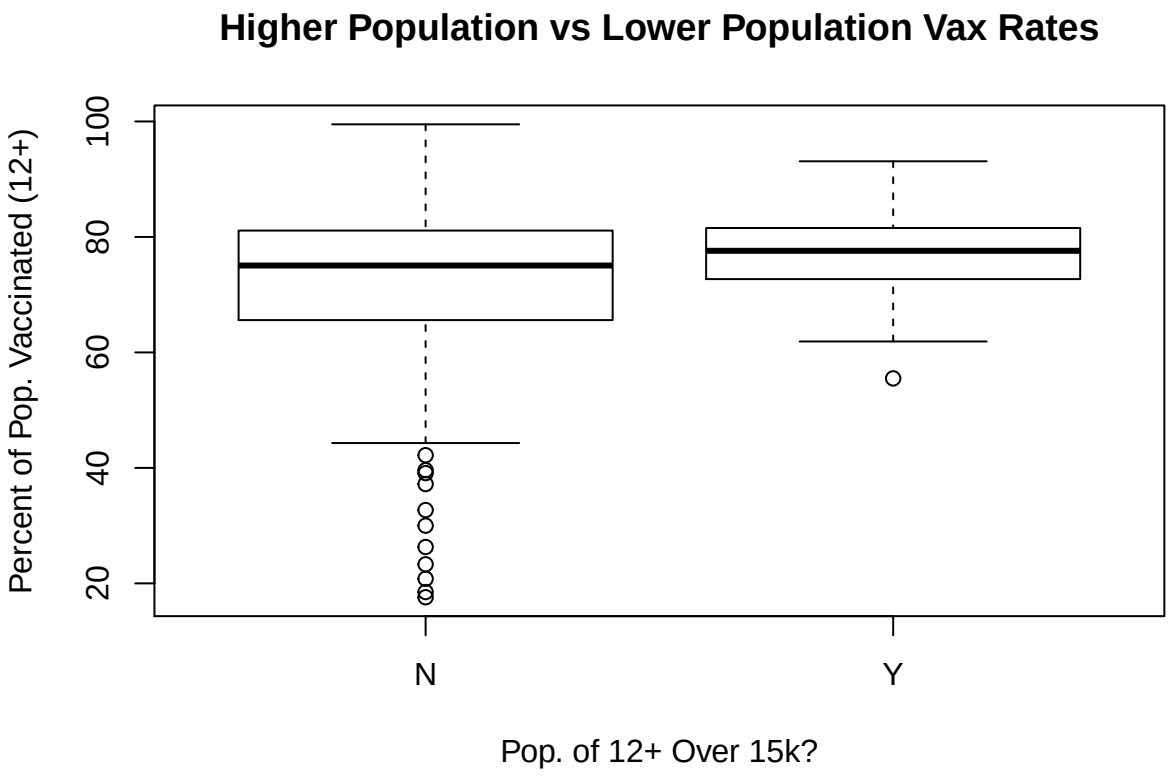
```r
la_c3 <- la_c2
# Let's compare the population to the percent of that
# population that vaccinated.
plot(la_c3$X12..Pop..Vaccinated.... ~ la_c3$Population..12..,
     xlab = "Population (12+)",
     ylab = "Percent Vaccinated (12+)",
     main = "Total Population vs Percent Vaccinated (Age 12+)") # Doesn't show a pattern...
```



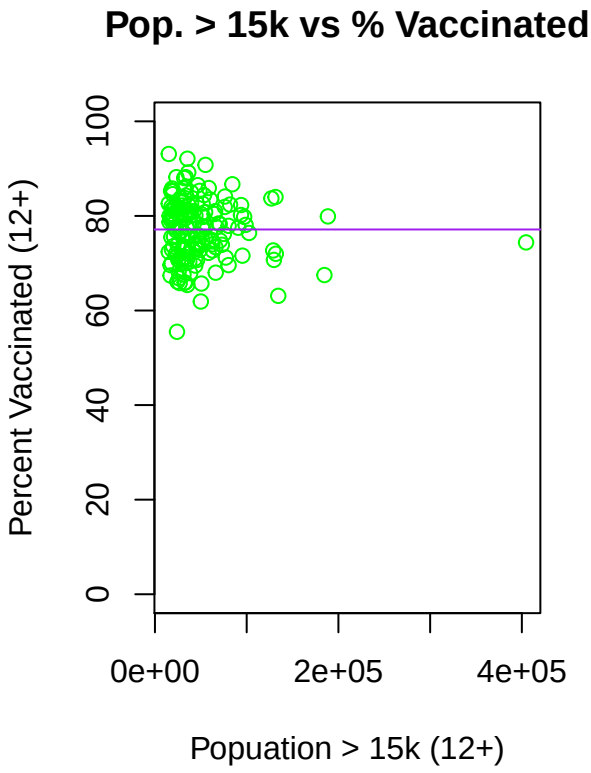

```
# Let's see if the more populated areas of
# LA County have a higher vaccination rate.
# We'll use a population of 15k as our arbitrary cutoff point
# between sparsely populated and highly populated.
tp <- which(la_c3$Population..12.. >= 15000)
over10k <- rep("N", 324)
over10k[tp] <- "Y"
la_c3 <- cbind(la_c3, over10k)
table(la_c3$over10k) # Amount of areas over/under pop. of 15k
```

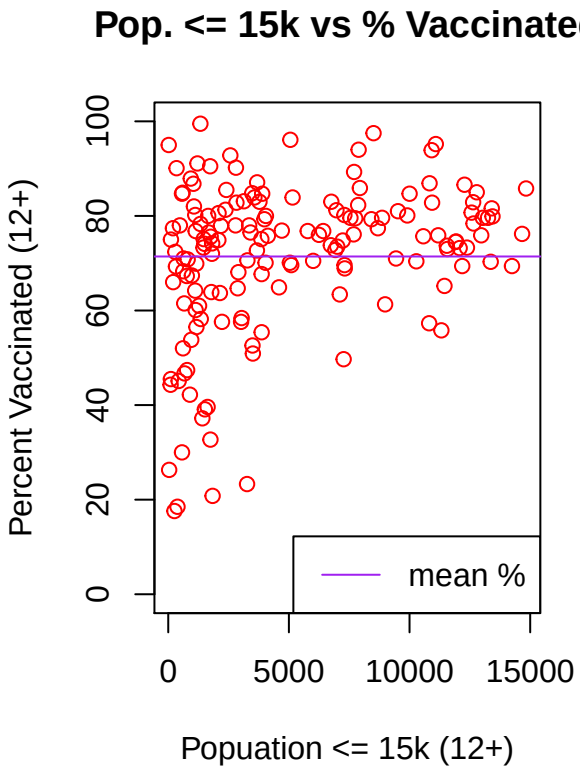
```
##
##   N   Y
## 160 164
```

```
plot(la_c3$over10k, la_c3$X12..Pop..Vaccinated....,
     xlab = "Pop. of 12+ Over 15k?",
     ylab = "Percent of Pop. Vaccinated (12+)",
     main = "Higher Population vs Lower Population Vax Rates")
```




```
par(mfrow = c(1, 2))
plot(la_c3$X12..Pop..Vaccinated....[tp] ~ la_c3$Population..12..[tp],
     ylim = c(0, 100), xlab = "Popuation > 15k (12+)",
     ylab = "Percent Vaccinated (12+)", col = "green",
     main = "Pop. > 15k vs % Vaccinated")
abline(h = mean(la_c3$X12..Pop..Vaccinated....[tp]), col = "purple")
plot(la_c3$X12..Pop..Vaccinated....[-tp] ~ la_c3$Population..12..[-tp],
     ylim = c(0, 100), xlab = "Popuation <= 15k (12+)",
     ylab = "Percent Vaccinated (12+)", col = "red",
     main = "Pop. <= 15k vs % Vaccinated")
abline(h = mean(la_c3$X12..Pop..Vaccinated....[-tp]), col = "purple")
legend("bottomright", legend = c("mean %"), col = c("purple"), lty = 1)
```





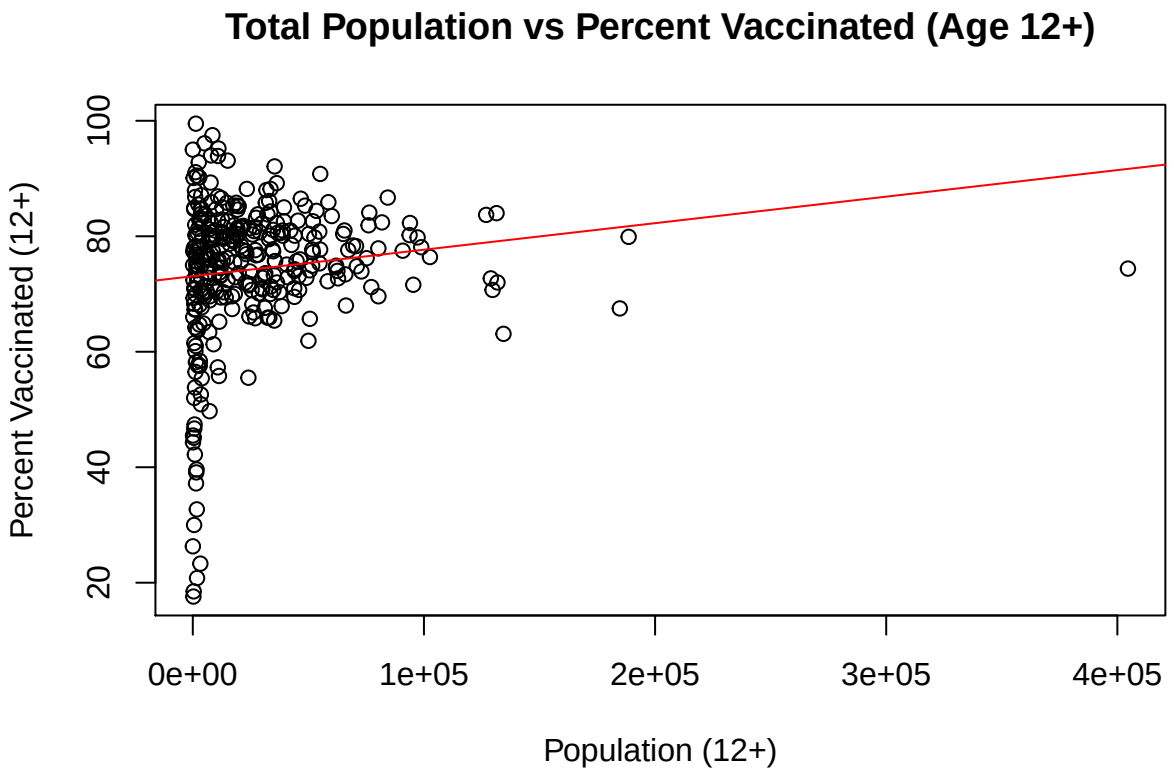
We do see that the more populated areas have a higher vaccination rate on average as we can see from the scatterplot where the higher pop. areas have a high concentration between 60% and 90%, whereas the lower pop. areas seems to show a more varied spread of the percentage vaccinated and a slightly lower average in general.

```
# Now let's a try a linear regression model
la_m <- lm(la_c3$X12..Pop..Vaccinated.... ~ la_c3$Population..12..)
summary(la_m)
```

```
##
## Call:
## lm(formula = la_c3$X12..Pop..Vaccinated.... ~ la_c3$Population..12..)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -55.490  -3.839   2.033   7.132  26.360
##
## Coefficients:
```

```
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)            7.308e+01  8.511e-01  85.867   <2e-16 ***
## la_c3$Population..12.. 4.595e-05  1.870e-05   2.457   0.0146 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.35 on 322 degrees of freedom
## Multiple R-squared:  0.0184, Adjusted R-squared:  0.01535
## F-statistic: 6.035 on 1 and 322 DF,  p-value: 0.01455
```

```r
plot(la_c3$X12..Pop..Vaccinated.... ~ la_c3$Population..12..,
     xlab = "Population (12+)",
     ylab = "Percent Vaccinated (12+)",
     main = "Total Population vs Percent Vaccinated (Age 12+)")
abline(la_m, col = "red")
```



While the summary of the model seems to be promising with how it shows a positive trend in vax rate as pop. increases, adding a regression line to the first scatterplot doesn't really work, although it could probably be useful to predict the vaccination rates of other districts not listed here. Unfortunately, this data set is fairly barebones, so we can't really do much else.
I'd love to try and add political alignments for each districts, but unfortunately, there is no current data set with such info and if it does, it groups all unincorporated regions into one.
For now, we do see that higher population areas do make a difference in vaccination rates, at least in the LA County.