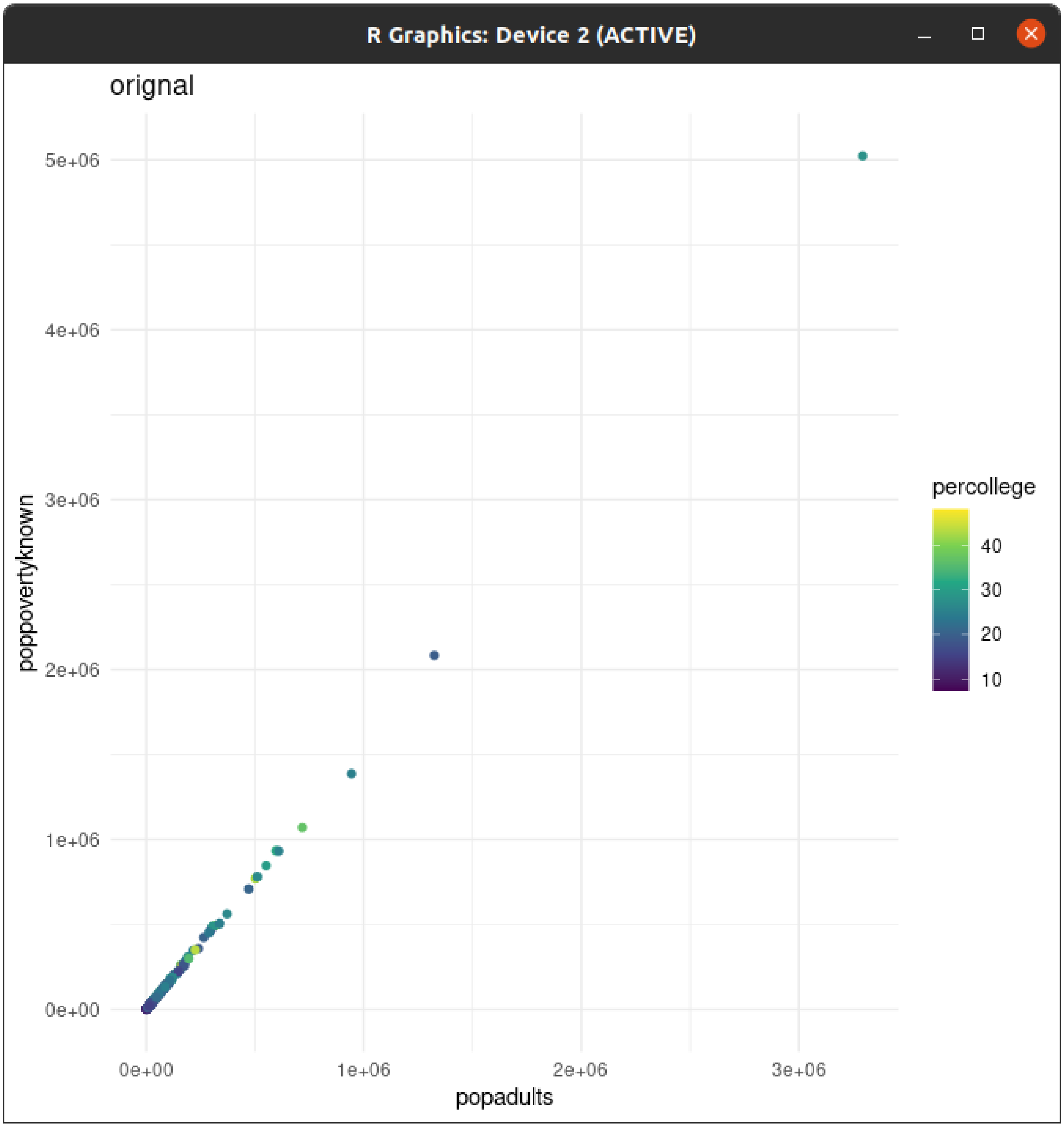
Group Project 2 MidWest

Anthony Hessler, Cole Ruppel, Jake Gadaleta

Set 1 (Jake)

Initial Graph



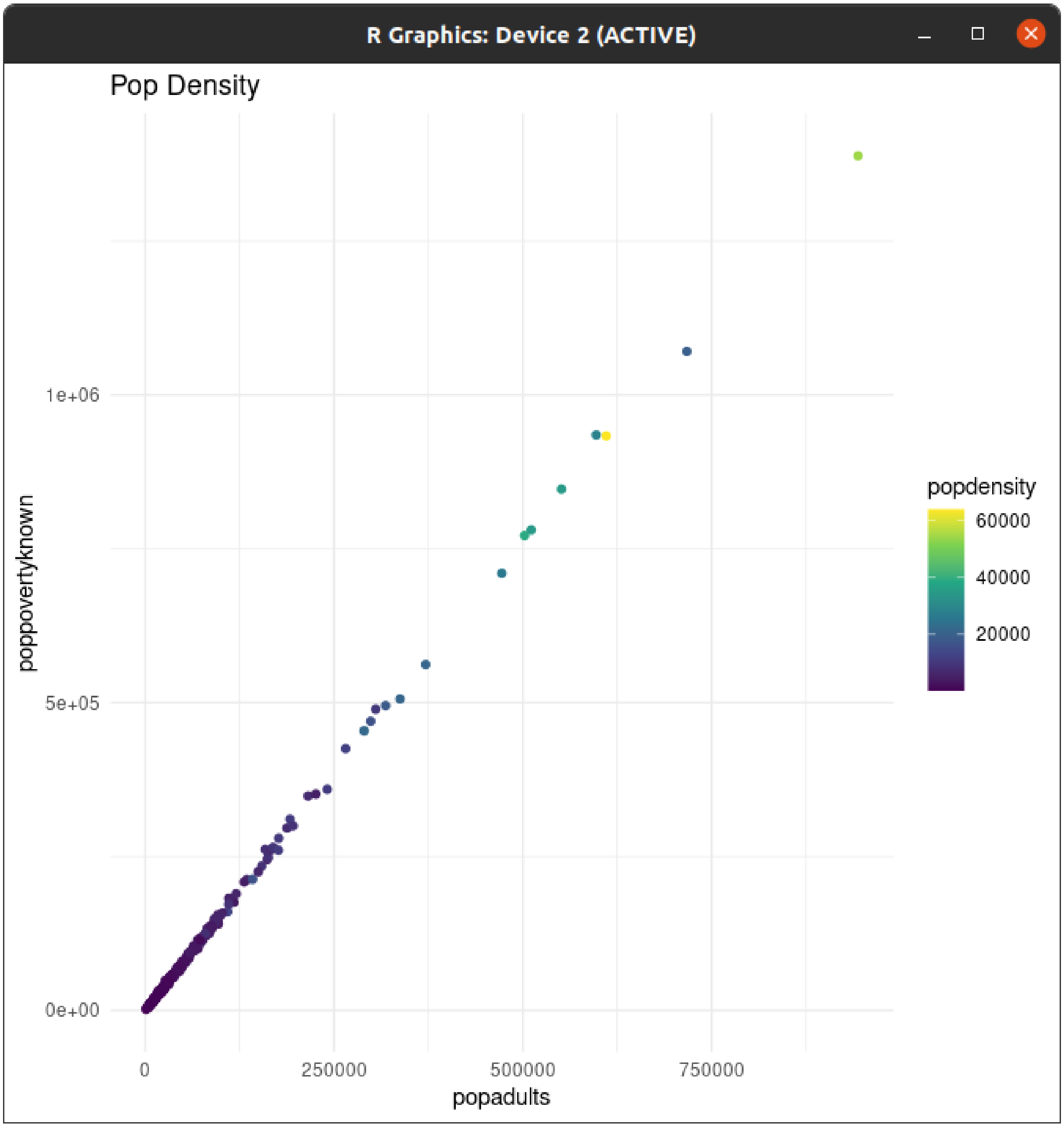
Here we look through the poverty population next to the population of adults colored by the percentage that went to college, shockingly the more people in general you have the more poverty you also have (who would've thunk)

Logic

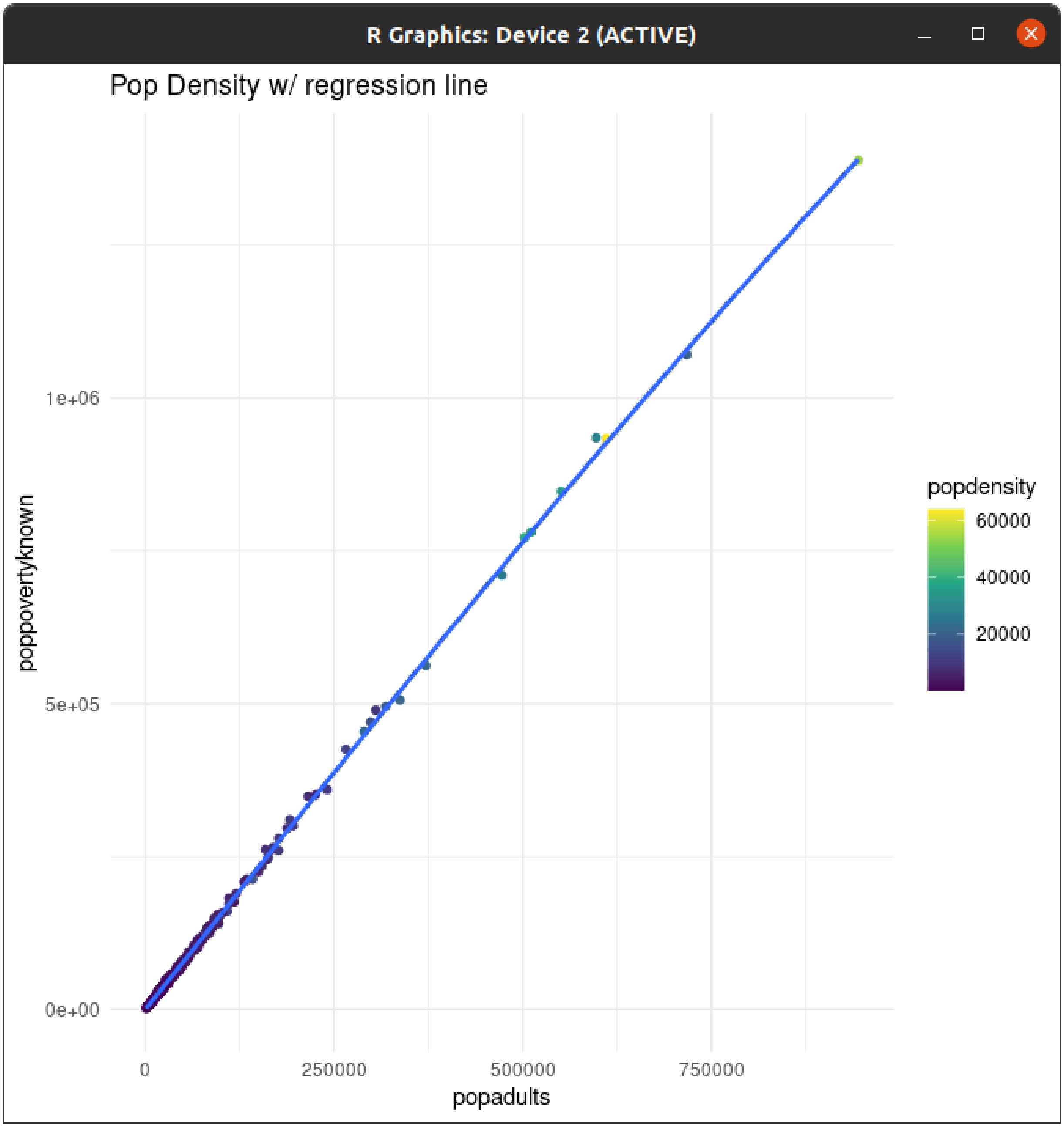
I then decided to take some time to see if I could parse out that city at the top with the large just to see what is up there. To do this I ran the following script.

```
midwest %>%
filter(popadults >= 3e+6) %>%
arrange(desc(popdensity))
```

This gave me the result of Cook IL as this outlier I then had the idea of parsing this outlier out of the code and then testing the following with regression.

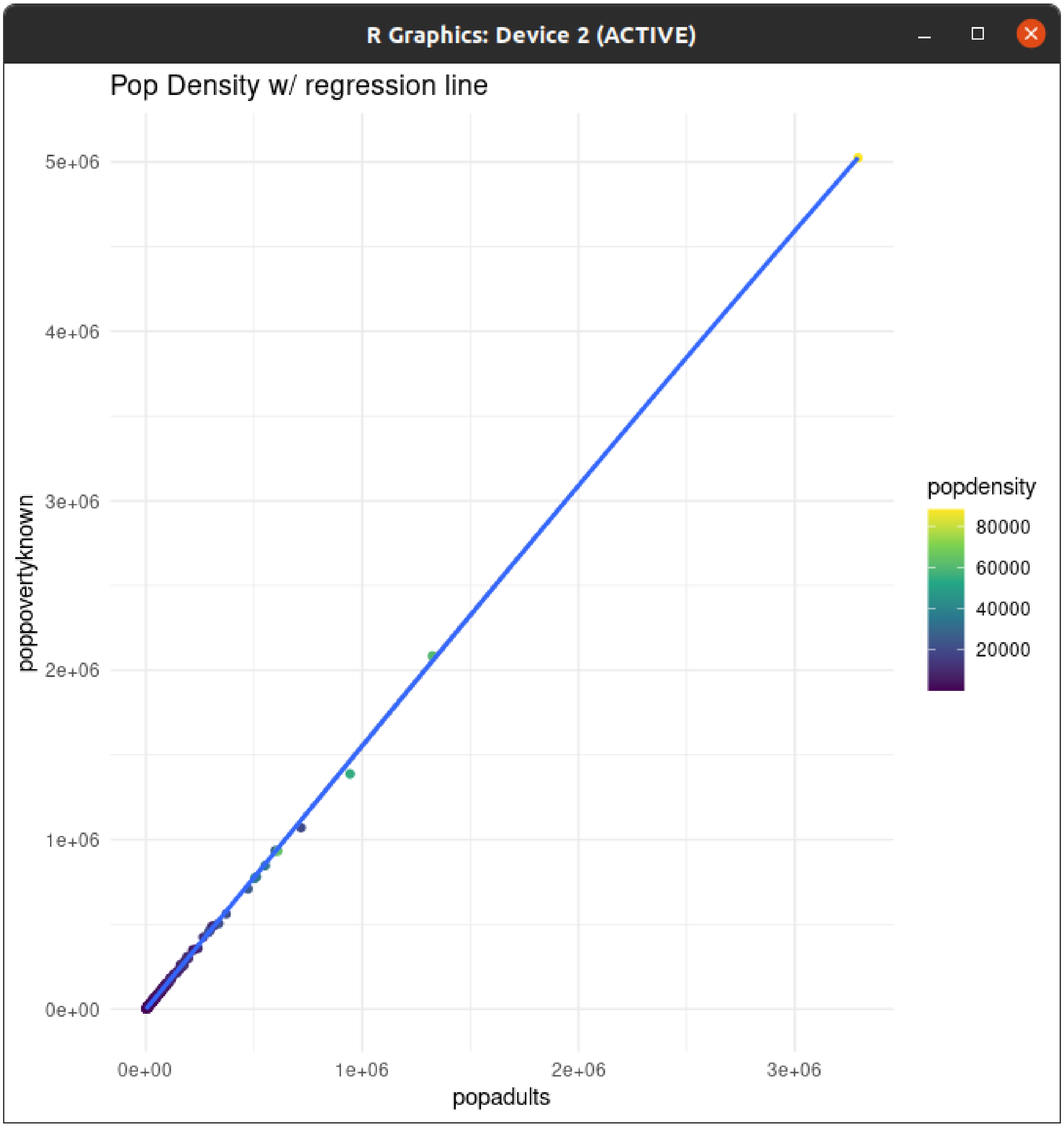


R Graphics: Device 2 (ACTIVE)
Pop Density
1e+06
5e+05
0e+00
poppovertyknown
0
250000
500000
750000
popadults
popdensity
60000
40000
20000



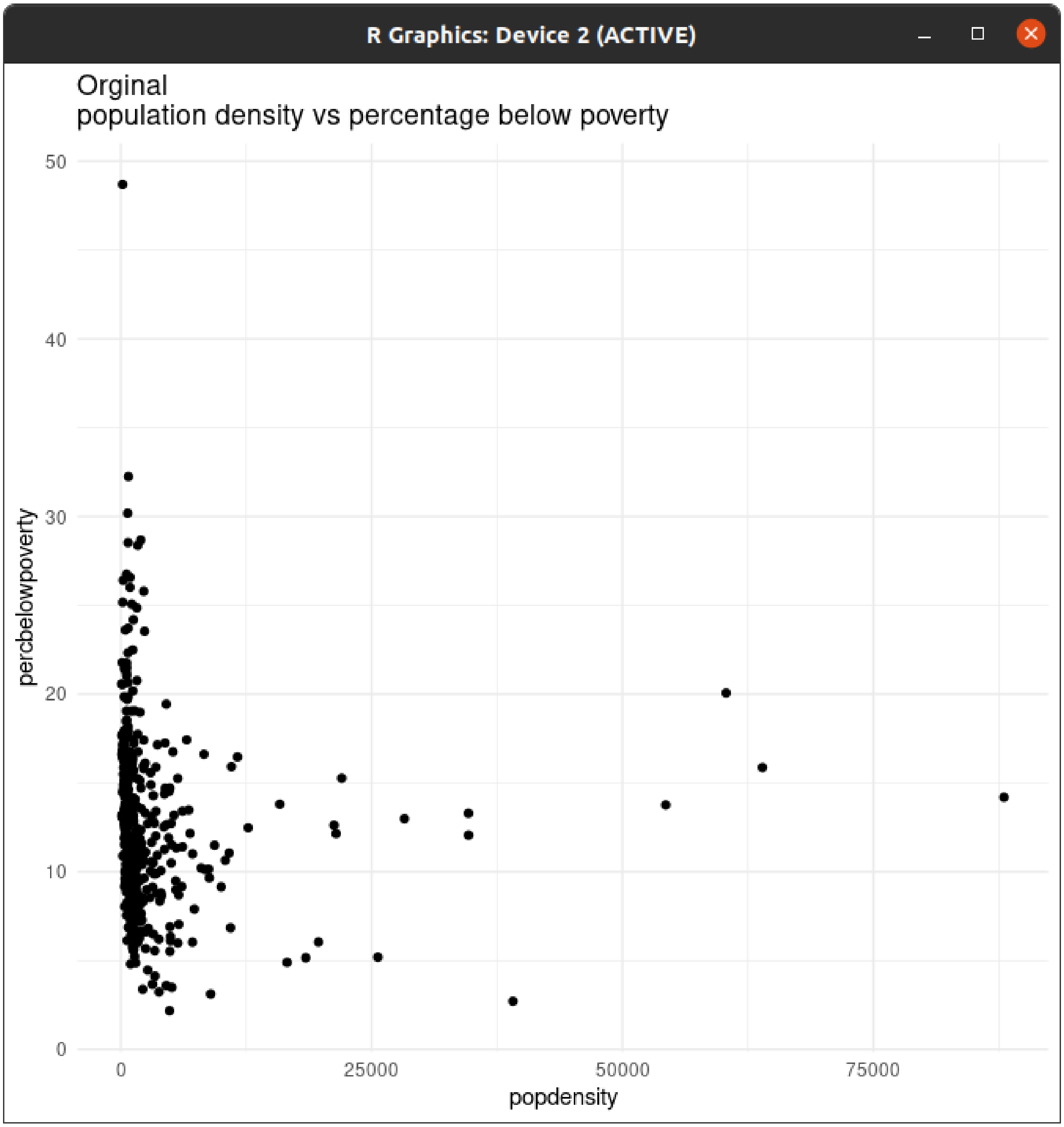
I then thought to myself, is this really necessary and decided to take back up one step and see if the outlier would have a negative impact on this line of regression and my final graph ended up looking like

Final Graph

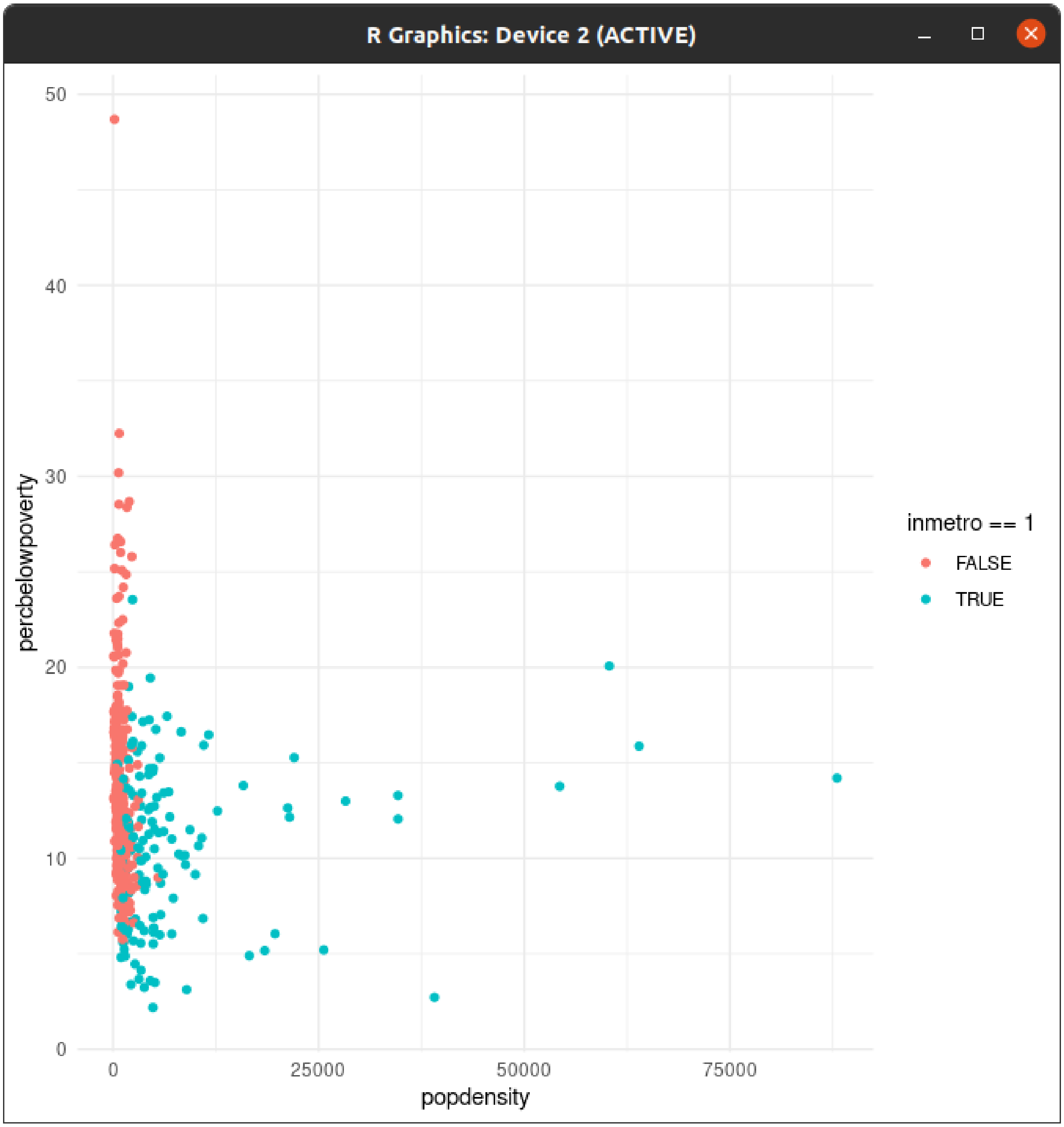

R Graphics: Device 2 (ACTIVE)
Pop Density w/ regression line
5e+06
4e+06
3e+06
2e+06
1e+06
0e+00
poppovertyknown
0e+00
1e+06
2e+06
3e+06
popadults
popdensity
80000
60000
40000
20000

Set 2 (Jake)

Initial Graph



Any sane person is gonna have a hard time unpacking this graph which means that it is filtering time, before I moved into filtering I wanted to add some color and see if the counties were impacted by the fact that they are cities or not.



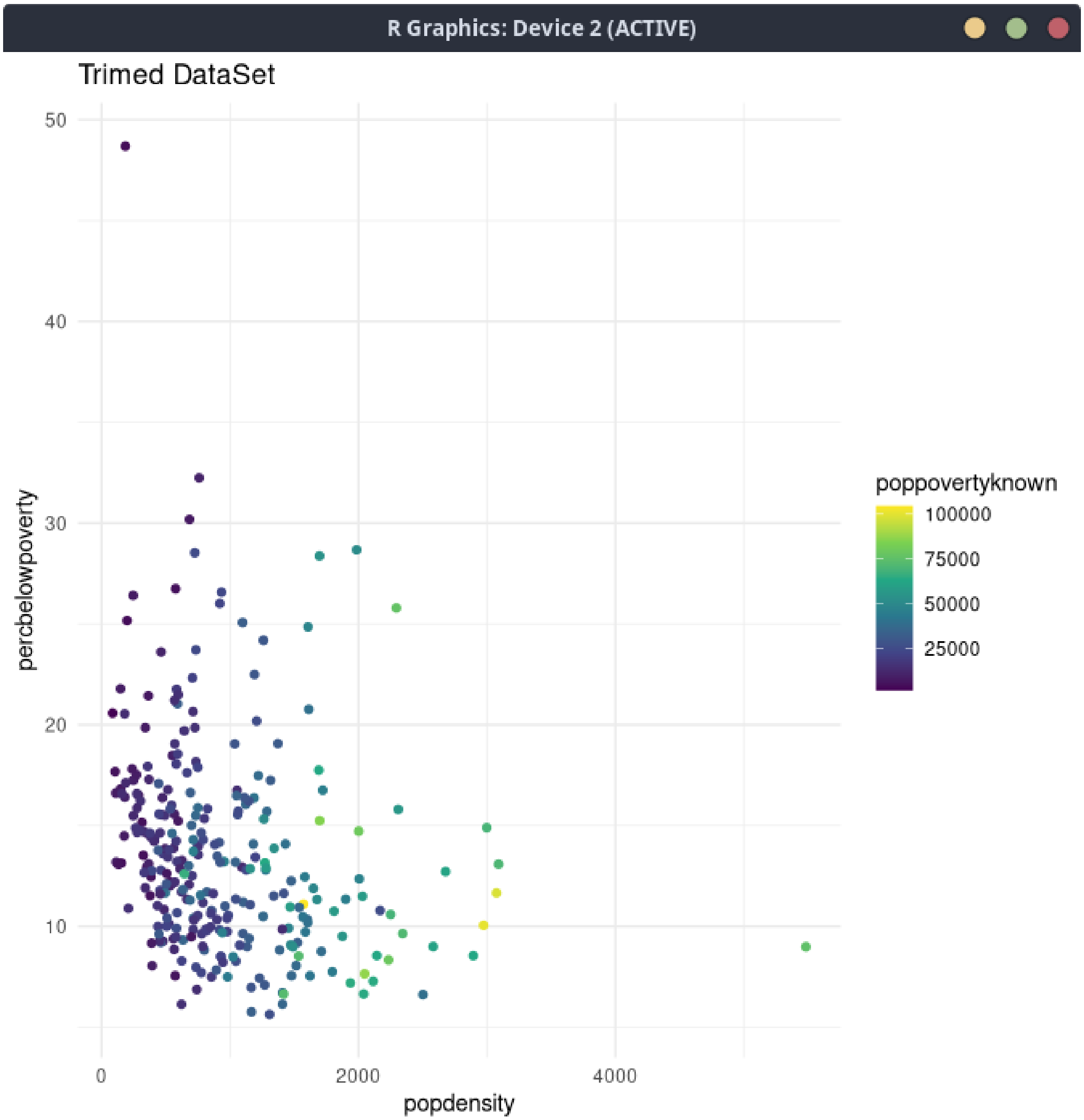
This lead to the shocking realization that a higher percentage of poverty is happening in rural areas, which I took into consideration for my filters.

Logic

I first made a new data set that threw out all of the metros so I could really drill down into why this huge percentage in so the first thing I've done is removed removed those pesky cities.

```
j2 <- midwest %>%
filter(inmetro != 1)
```

I then decided to plot this again.

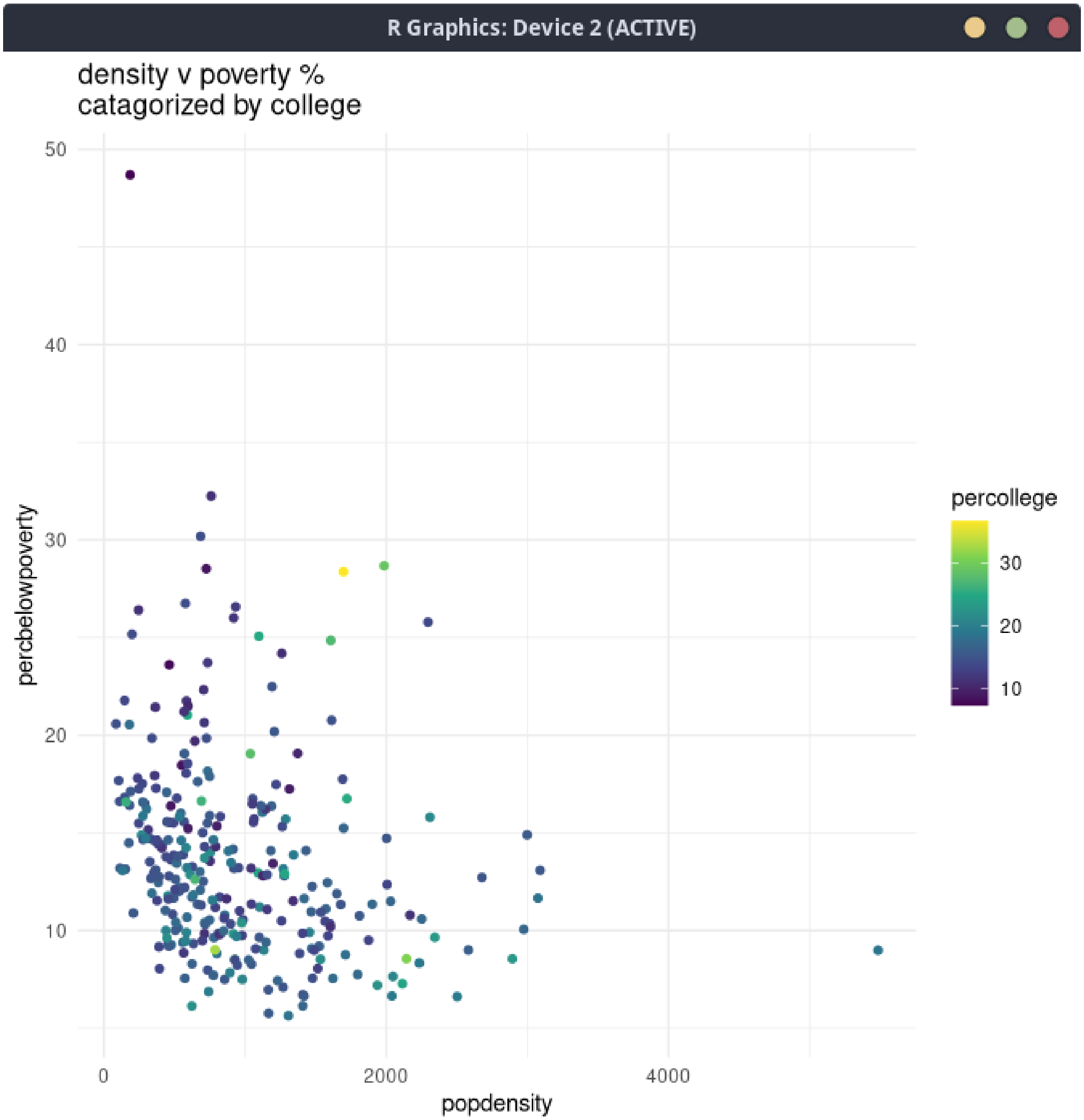


I decided to further trim the data set to focus in on those higher percentage populations by using the following script.

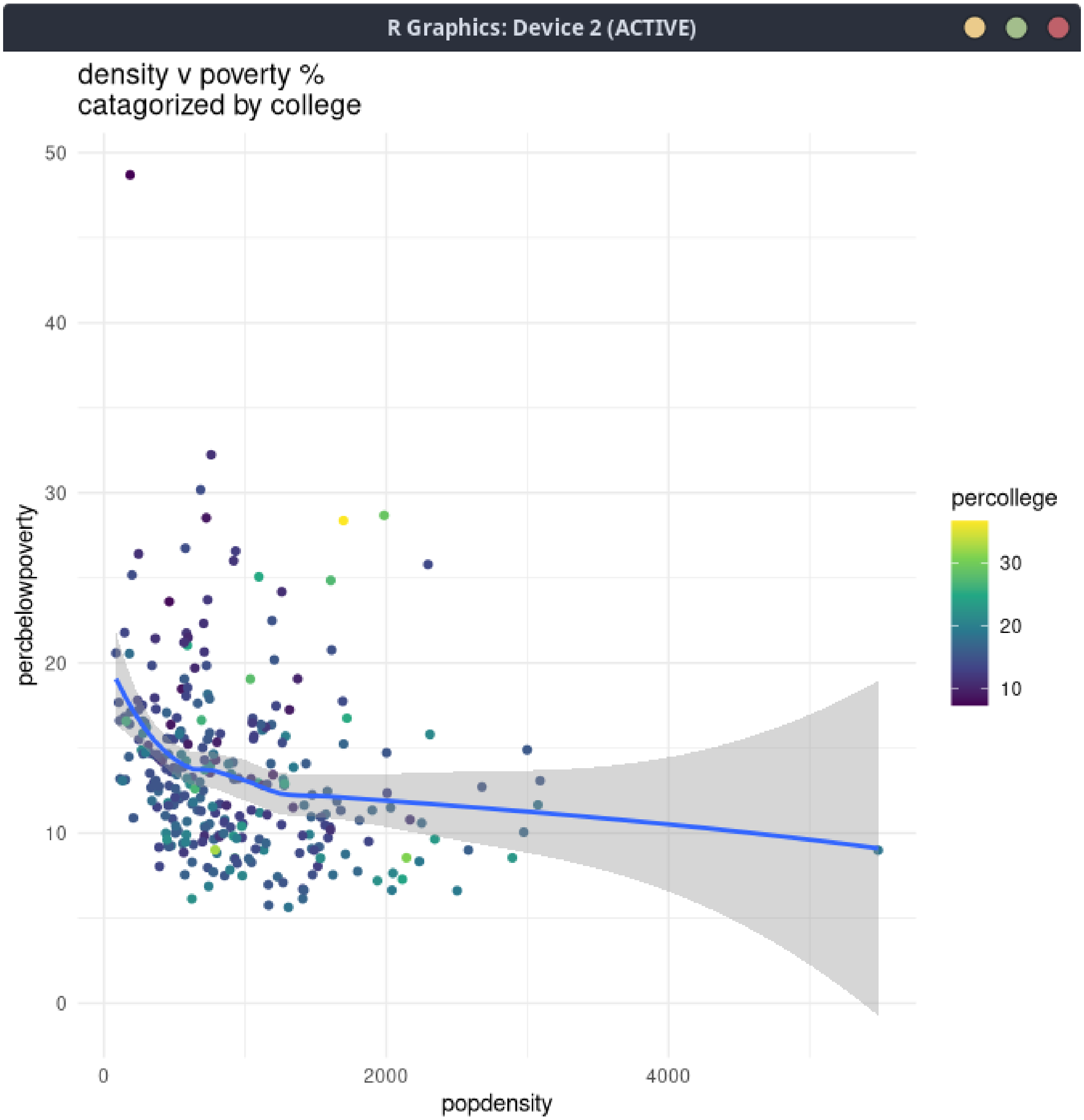
```
j2 <- midwest %>%
filter(inmetro != 1) %>%
filter(percbelowpoverty > 12.5)
```

This way we can reduce the amount of places with lower poverty percentages.

Final Graph



What I see is super shocking about all of this is that college percentage seems to have a shockingly low impact on the poverty percentage. Which really makes you wonder what we are all really doing here. We do need to take into account that one super outlier up there with a 50% poverty rate, which does in fact have an extremely low college percentage. This is the same county that we have zeroed in on in the last section. Because the data is so spread out I didn't think that adding a line of regression would really impact things but I did anyway.

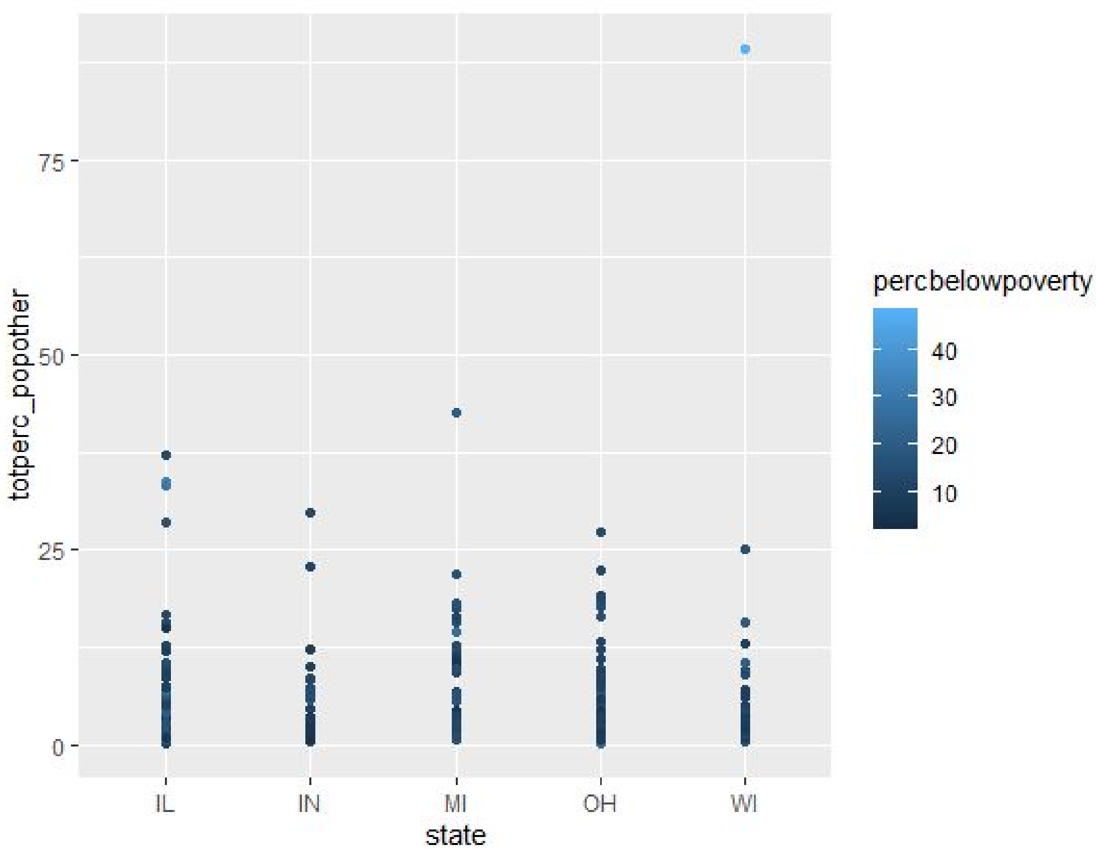


That huge spread really shows that this is pretty much shows that there really is no correlation between these 2 things.

Set 3 (Cole)

Initial Graph

```
     PID county state  area poptotal popdensity popwhite popblack popamerindian popasian
   <int> <chr>  <chr> <dbl>    <int>      <dbl>    <int>    <int>         <int>    <int>
 1   561 ADAMS  IL    0.052    66090      1271.    63917     1702            98      249
 2   562 ALEXA~ IL    0.014    10626       759      7054     3496            19       48
 3   563 BOND   IL    0.022    14991       681.    14477      429            35       16
 4   564 BOONE  IL    0.017    30806      1812.    29344      127            46      150
 5   565 BROWN  IL    0.018     5836       324.     5264      547            14        5
 6   566 BUREAU IL    0.05     35688       714.    35157       50            65      195
 7   567 CALHO~ IL    0.017     5322       313.     5298        1             8       15
 8   568 CARRO~ IL    0.027    16805       622.    16519      111            30       61
 9   569 CASS   IL    0.024    13437       560.    13384       16             8       23
10   570 CHAMP~ IL    0.058   173025      2983.   146506    16559           331     8033
# ... with 427 more rows, and 18 more variables: popother <int>, percwhite <dbl>,
#   percblack <dbl>, percamerindan <dbl>, percasian <dbl>, percother <dbl>,
#   popadults <int>, perchsd <dbl>, percollege <dbl>, percprof <dbl>,
#   poppovertyknown <int>, percpovertyknown <dbl>, percbelowpoverty <dbl>,
#   percchildbelowpovert <dbl>, percadultpoverty <dbl>, percelderlypoverty <dbl>,
#   inmetro <int>, category <chr>
>
```



Logic

Looking at the data I noticed that it gives percentages for the different races, i.e. black, native american, asian, and other. Those are typically referred to as minorities in the United States, so I used the mutate function to create a total percentage of all the minorities. Generally, from my experience the midwest population is white, so i wanted to look for the number of minorities by state. I also threw in the percent below poverty as color to spice up the graph as well. Looking at this graph I noticed an outlying point in Wisconsin that had over 75% minorities so I mutated that data again to separate that point so I could see it in the table.

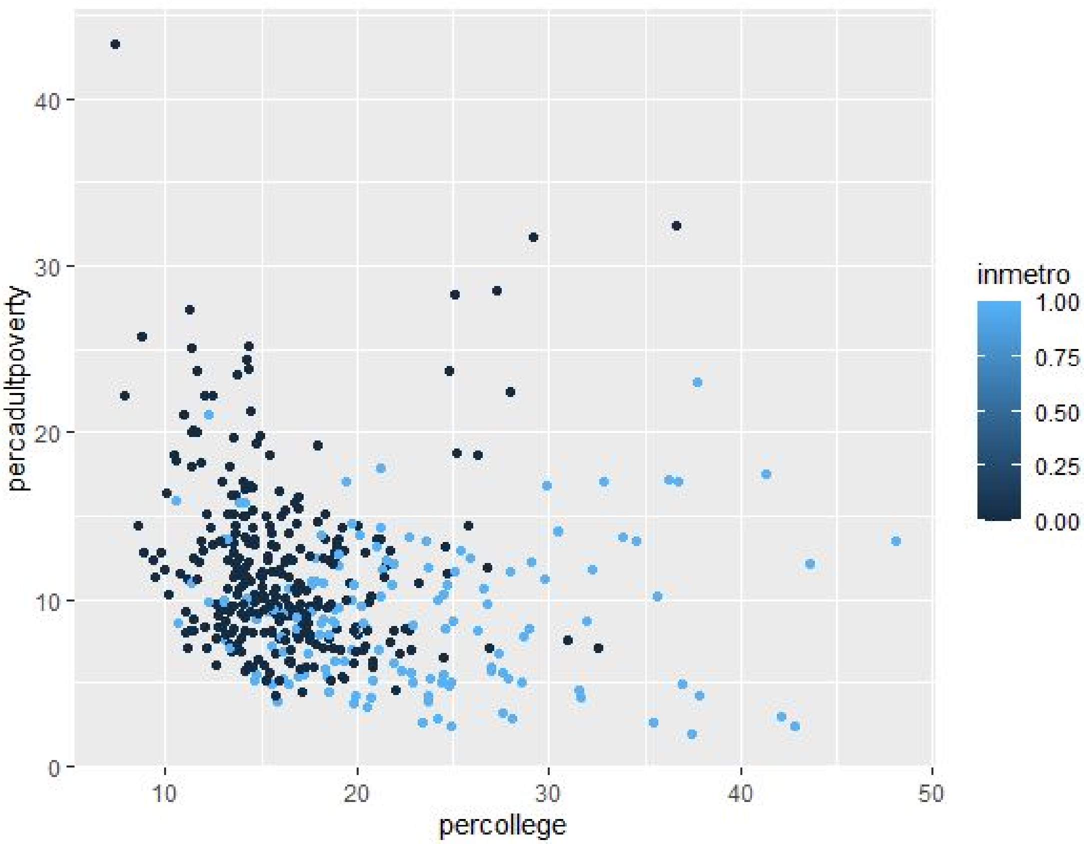
Final Graph

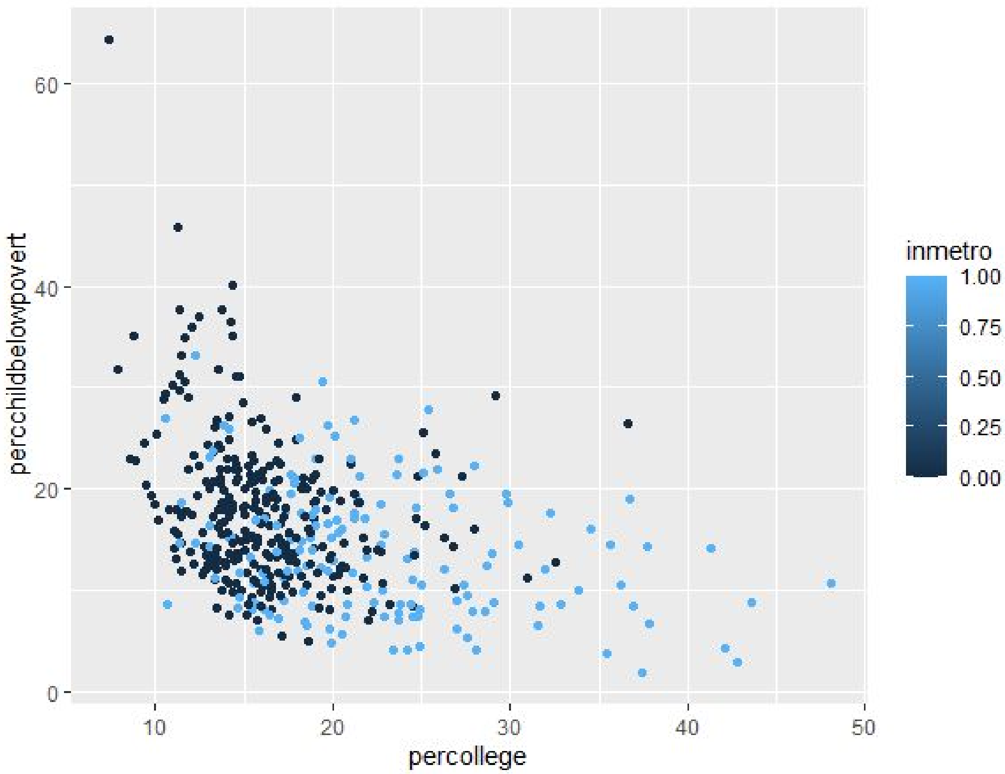
```
> c <- mutate(midwest, totperc_popother = percblack + percamerindan + percasian + percothe
r)
> ggplot(data=c) + geom_point(mapping=aes(x = state, y = totperc_popother,color = percbelo
wpoverty))
> c1 <- filter(c, state == "WI", totperc_popother >= 75)
> c1
# A tibble: 1 x 29
    PID county state  area poptotal popdensity popwhite popblack popamerindian popasian
  <int> <chr>  <chr> <dbl>    <int>      <dbl>    <int>    <int>         <int>    <int>
1  3020 MENOM~ WI    0.021     3890       185.      416        0          3469        0
# ... with 19 more variables: popother <int>, percwhite <dbl>, percblack <dbl>,
#   percamerindan <dbl>, percasian <dbl>, percother <dbl>, popadults <int>,
#   perchsd <dbl>, percollege <dbl>, percprof <dbl>, poppovertyknown <int>,
#   percpovertyknown <dbl>, percbelowpoverty <dbl>, percchildbelowpovert <dbl>,
#   percadultpoverty <dbl>, percelderlypoverty <dbl>, inmetro <int>, category <chr>,
#   totperc_popother <dbl>
> |
```

With this information I was able to see that the population was made up mostly by Native Americans as well as the name of the county so I did a quick google search to find the town was called Menomonee Falls in Wisconsin.

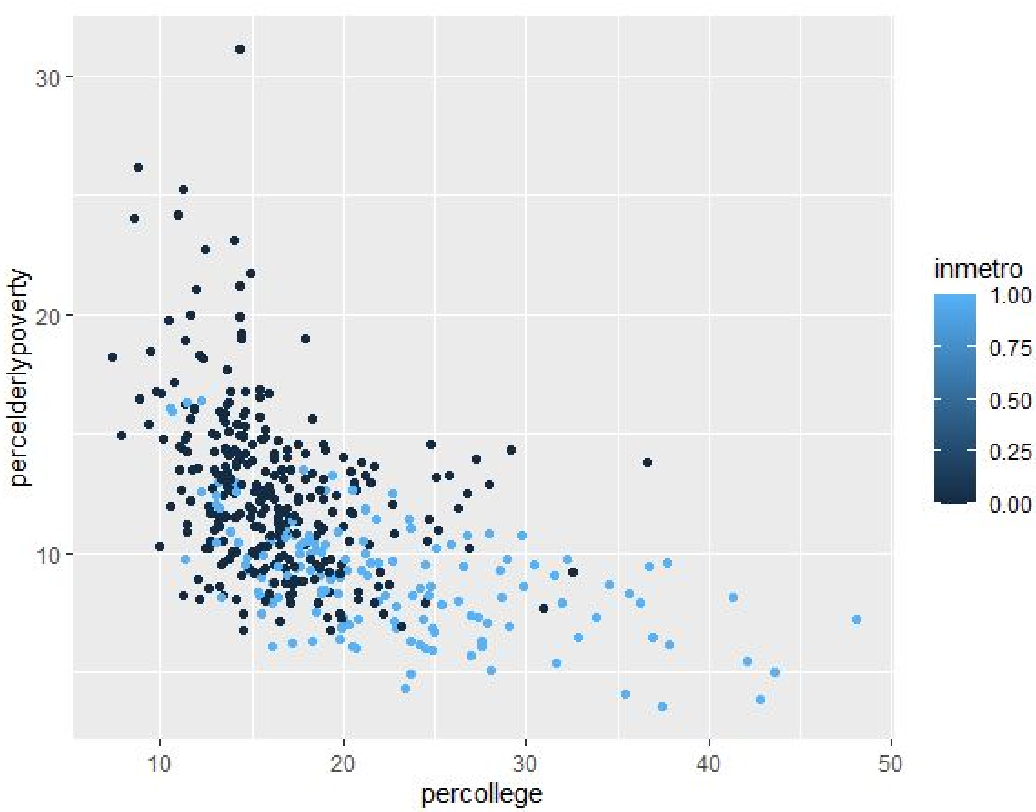
Set 4 (Cole)

Initial Graph





60
40
20
0
10
20
30
40
50
percchildbelowpovert
percollege
inmetro
1.00
0.75
0.50
0.25
0.00

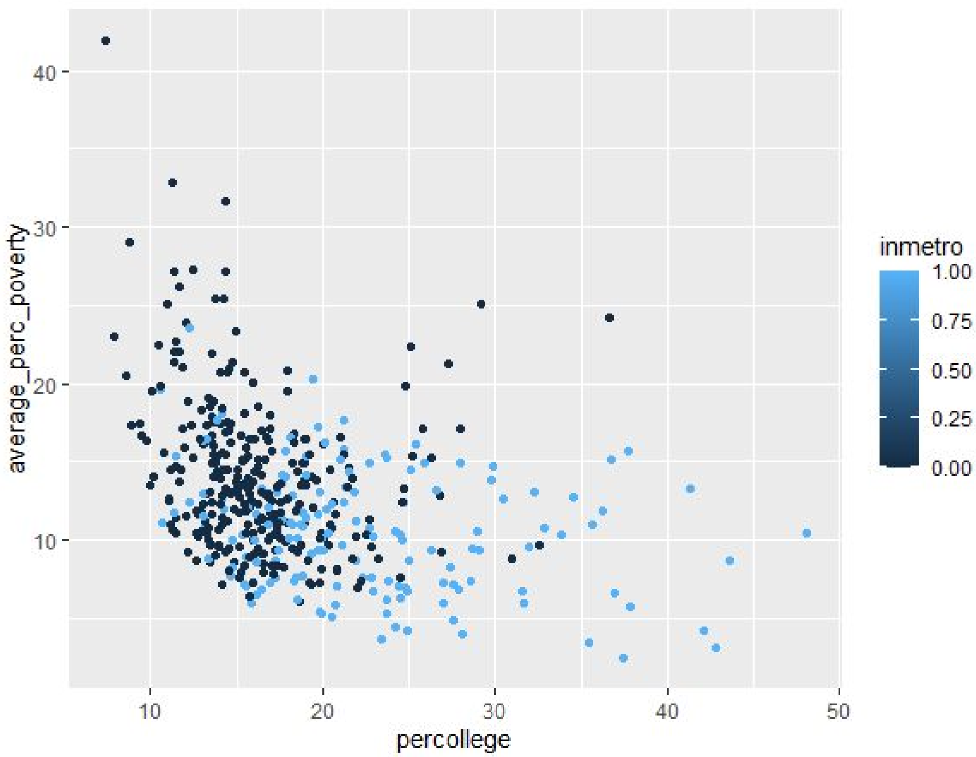


30
20
10
10
20
30
40
50
percelderlypoverty
percollege
inmetro
1.00
0.75
0.50
0.25
0.00

Logic

I wanted to see the relationship between percentage of college graduates to the percentage of poverty between the ages so I made the 3 graphs. I also colored it by city versus suburbs. I wanted to just look at one graph so I took the average of the 3 age groups and made another graph. I also wanted to see a table so I used the select function to remove everything but the percentage of poverty to compare them. I also arranged them in ascending order so they would be easy to look at.

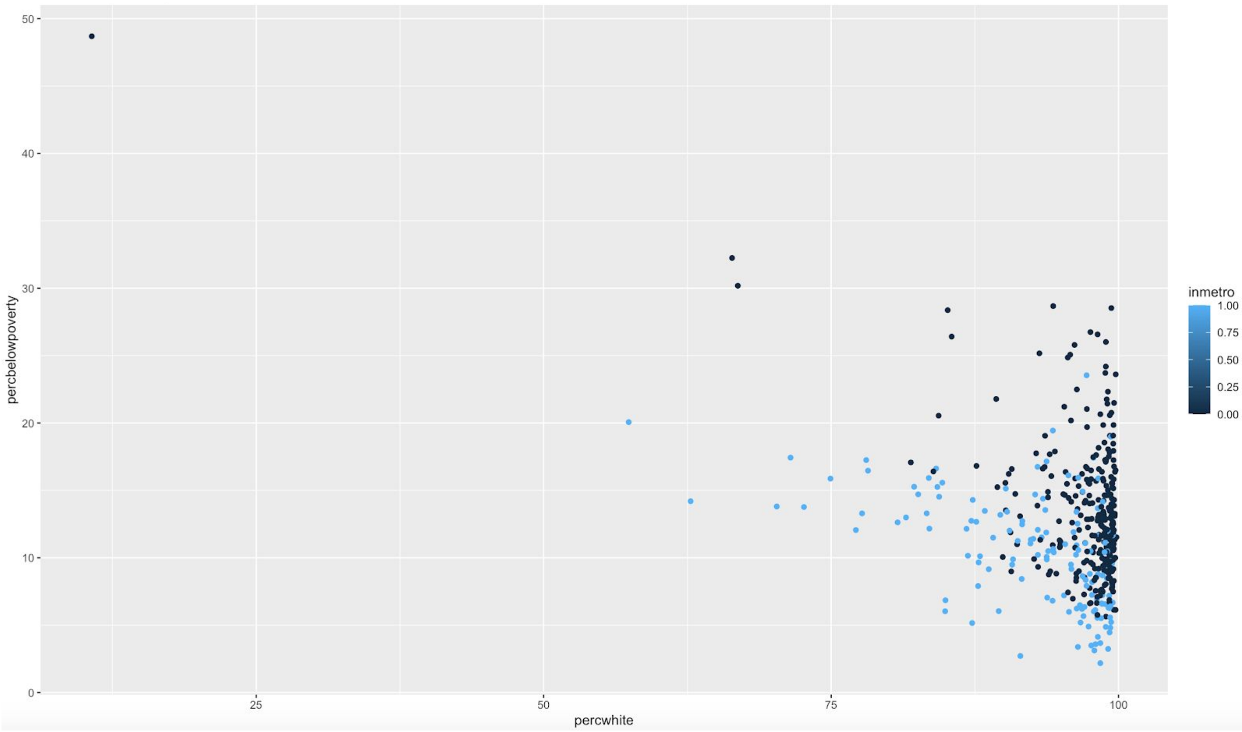
Final Graph




```
> c4 <- select(c3, contains("poverty") & -"poppovertyknown")
> arrange(c4, average_perc_poverty)
# A tibble: 437 x 5
   percpovertyknown percbelowpoverty percadultpoverty percelderlypover~ average_perc_po~
              <dbl>            <dbl>            <dbl>             <dbl>            <dbl>
 1             98.3             2.18             1.94              3.55             2.47
 2             98.7             2.71             2.40              3.84             3.06
 3             98.4             3.12             2.59              4.09             3.49
 4             98.8             3.24             2.58              4.28             3.65
 5             99.1             3.49             2.87              5.06             4.00
 6             99.1             3.59             2.93              5.45             4.21
 7             99.5             3.39             2.36              5.88             4.22
 8             96.7             3.67             2.84              6.14             4.40
 9             98.7             4.13             3.23              6.05             4.86
10             98.8             4.46             3.50              6.08             5.07
# ... with 427 more rows
> |
```

Set 5 (Anthony)

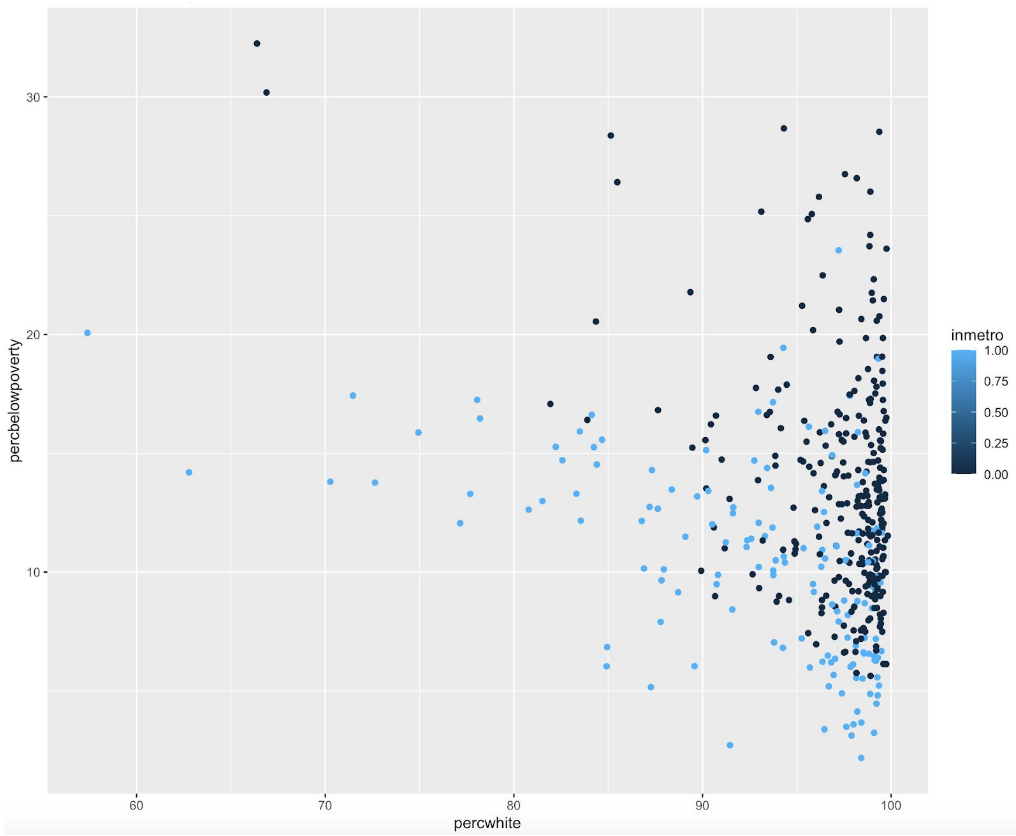
Initial Graph

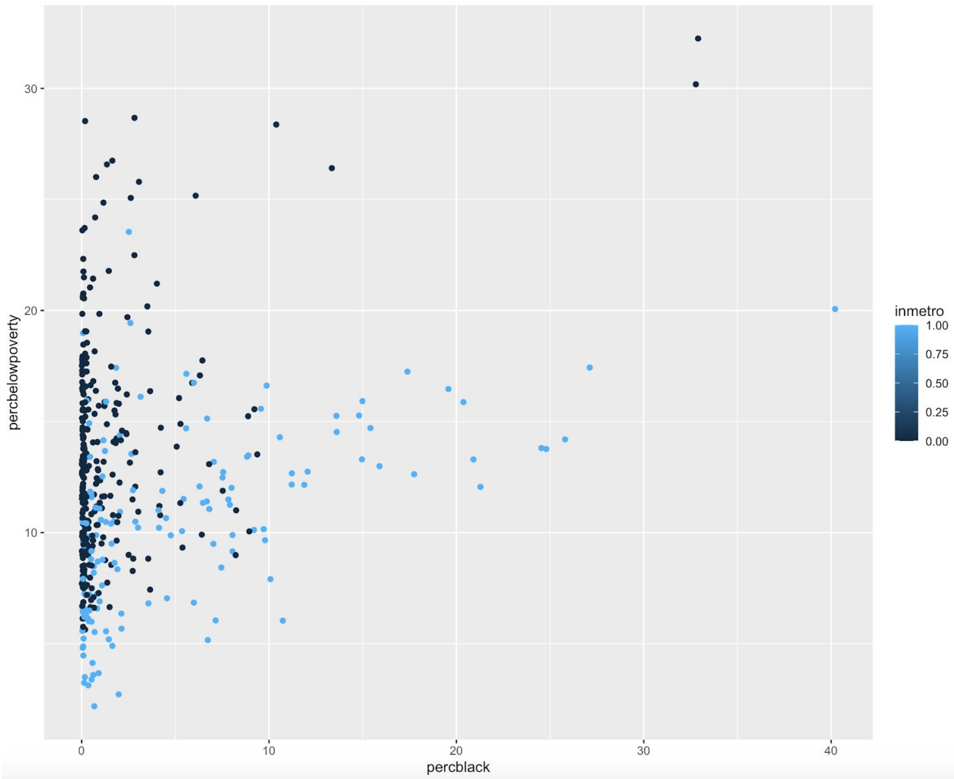


Logic

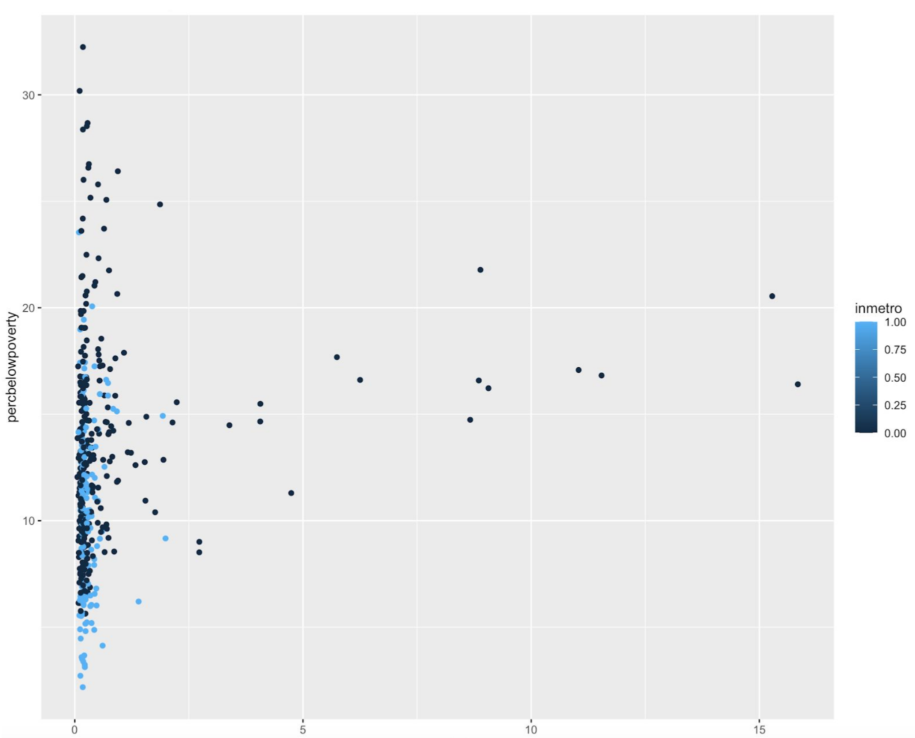
The first thing I noticed in this graph is the point at the top left, a huge outlier in both percent white and percent below poverty. As mentioned above, this county was easily identifiable as Menominee County, Wisconsin, which happens to be a Native Reserve. That explains why it's such an outlier for both statistics. With that out of the way, I wanted to look at the rest of the data more closely to see if there is a relationship between race and poverty, or if it is more due to metro status. I made graphs for each race, and interestingly found out that no matter what the race was, the worst poverty was always in rural areas. The other interesting observation I made is that there is an area of cities with lower white population and higher black and Asian population that has worse rates of poverty than most of the other cities, but not as bad as the rural areas. It generally seems like race has more of an effect on poverty in metro areas than it does in non-metro areas, but there does seem to be a (very unfortunate) relationship in which whiter areas are generally better off than less white areas.

Final Graphs

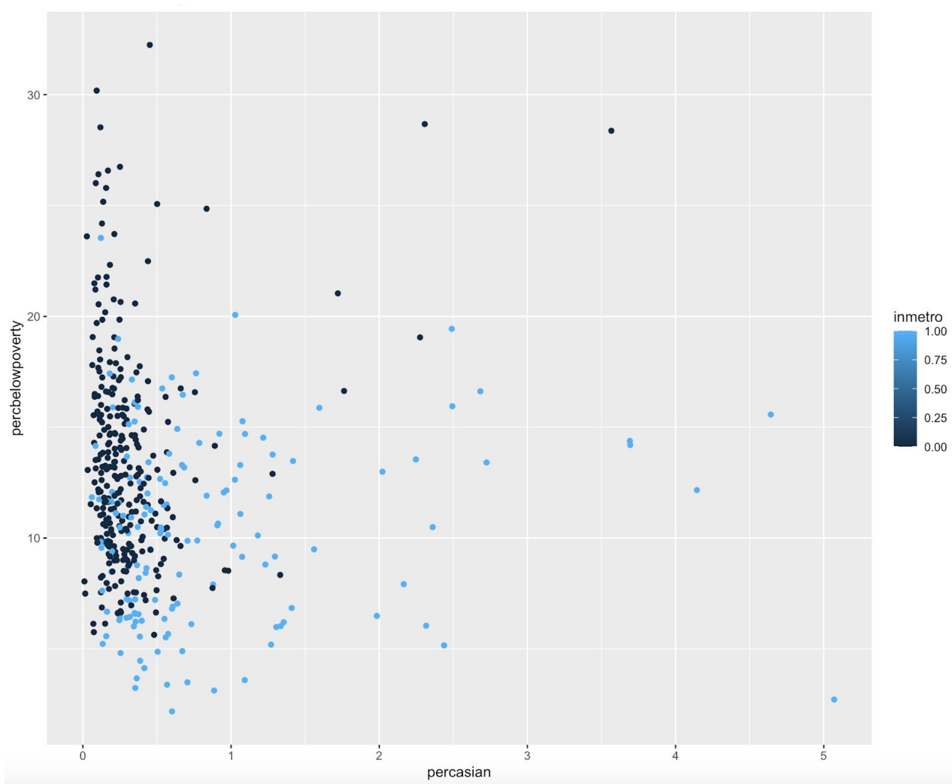


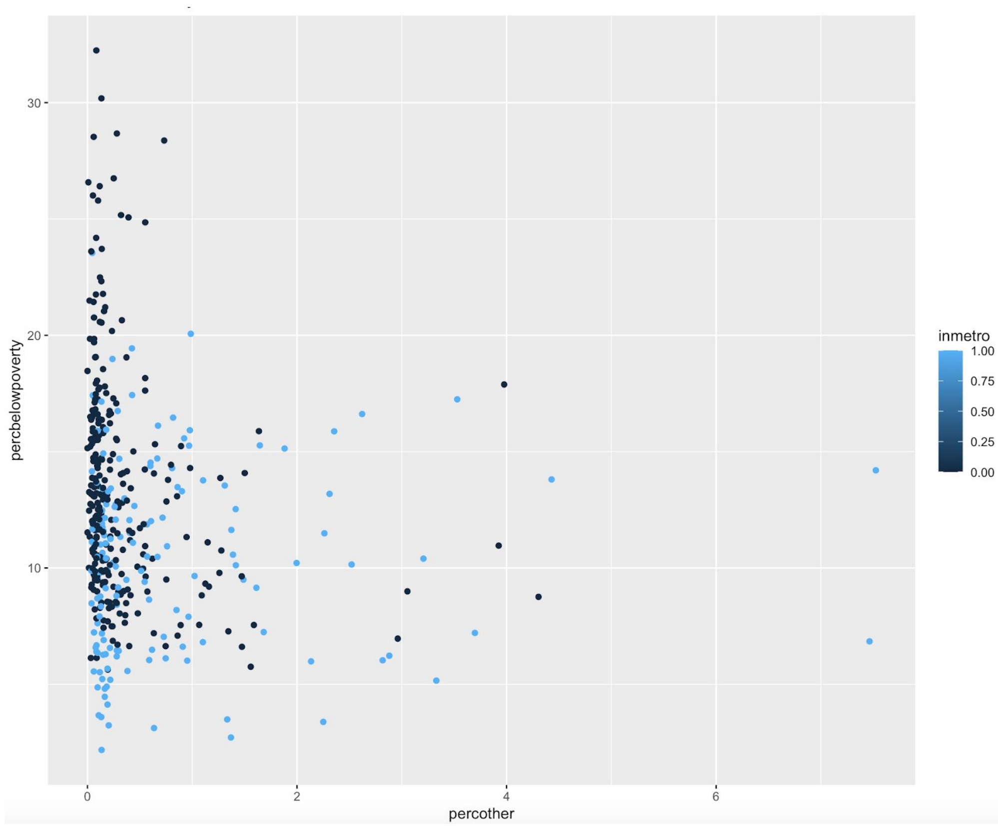


percbelowpoverty
percblack
inmetro
1.00
0.75
0.50
0.25
0.00



percbelowpoverty
percamerindan
inmetro
1.00
0.75
0.50
0.25
0.00


percbelowpoverty
percasian
inmetro
1.00
0.75
0.50
0.25
0.00
0
1
2
3
4
5
10
20
30


percbelowpoverty
percother
inmetro
1.00
0.75
0.50
0.25
0.00
0
2
4
6
10
20
30