crime&community

Tiana_Noll_Walker

12/13/2021

#The Process

When I transferred the original dataset to Excel I was able to easily transport the values into CSV format because they were already separated by commas. However, I didn't have the labels. I could have entered the labels by copy and pasting, but the labels weren't in my desired format so I chose to manually enter the labels. There's also dozens of entries, so I figured going through and manually entering the labels was a good way for me to really go through and examine each of the columns and think about how they each related to the dataset as I entered them.

The first thing I did after transferring my csv to R was running the summary statistics. Although it seems very simple, this was a very necessary and helpful part of the EDA process. It allowed me to go through each attribute and see the basic stats, the most important being the mean. Once I had the mean, I could really think about some of the bigger problems associated with the data. For example, race, gender, and age statistics, the mean amount of violent crime, etc. This was the essential first step in establishing my process and choosing what specific analyses I wanted to run. However, obviously most of the values need to be pulled and ran in regression models in relation to other specific variables. There is a lot of data here, and they aren't all relative to each other.

I decided I wanted to do all my EDA on mostly WA, so I extracted all the WA info and transferred it to a different CSV, which I transferred to R. There is a lot of missing data for different cities, so some of the plots and analysis I only had the option to run data on about 5 different cities for some of the stats that I wanted regarding LEMAS, the police, and violent crime. I labeled and explained most of the prcocesses and plots as I go through, but some of them are pretty self explanatory.

#Big Question and Dataset Choice

I chose to do the crime and community dataset. I am very interested in this one, because I am curious to know what the statistics are for myself. In issues involving crime and community, I feel like the media often gives narratives siding with one extreme or the other, and was curious to conduct my own analysis and find out for myself what the real numbers actually are. My "big question" of interest relating to the analysis in a broader sense was originally "is the media telling the truth" about crime and community stats, and seeing what story the data itself tells. The results are important and relevant to myself, and to the community at large.

What I discovered though is exactly what we covered when we were studying EDA. You start out with one question, you perform EDA, and then you are left with a lot of other questions that make you rephrase the original question. Doing PCA and figuring out the biggest predictors of violent crimes really made me question a lot of things. I explain more below, but it seems as though no matter what, the percentage of black people in a population is always an important predictor. Well, in some ways this confirms the "big question", because that is often a hot topic in the media, on all sides of the political spectrum. However, I realize now after doing the EDA that the bigger question is why. I need more tools and time to figure out the answer to that fully, but this was a good place to start. As I went through, I started asking different questions and the biggest one ended up being in relation to race and trying to figure out what the biggest predictors of violent crime were and why. I also did some Mutual Information Analysis in Python to look further into some of the variables, but I didn't know how to do it in R so I didn't include it here. I didn't

really use that information directly, but it did give me a better idea of what variables might be the most helpful in relation to this dataset. I made plots and did further EDA in this markdown that was found in that MIA to see if the results were reproducible and relevant. #Methodology

There were some issues. At first the original dataset was too big, and then I ended up narrowing it down to five cities and that was too small. I eventually just decided to go with WA, which was in the middle. At times, I do feel like this dataset was still too small. The biggest difficulty I had was definitely running PCA, but mostly it was just time consuming. PCA was helpful because it was able to take 130 variables and narrow them down to the 4 most important and relevant ones. Once I found those, I made a bunch of plots and visualizations and kept asking more questions.

#Conclusions Like I mentioned above, my original big question was "is the media telling the truth about crime and community?" The media often links crime rates to black and brown population, and income. Based on this analysis, these stipulations seem correct. However, I think there is a lot more analysis that needs to be done and nuance that needs to be asserted when talking about these subjects. Some things that I found interesting that I feel like isn't really talked about as much in the media when it comes to crime is age, as well as the amount of homes with more than two parents. But again, these might be because of systemic or historical issues that need more context. There are a lot of correlations in a lot of these variables, but that doesn't necessarily mean that they are causing the rate of violent crimes.

```
cc <- read.csv("community_crime.csv")
summary(cc)
```

```
##   state_numeric   county_numeric   community_numeric  community_name
##   Min.   : 1.00   Length:1994      Length:1994        Length:1994
##   1st Qu.:12.00   Class :character Class :character   Class :character
##   Median :34.00   Mode  :character Mode  :character   Mode  :character
##   Mean   :28.68
##   3rd Qu.:42.00
##   Max.   :56.00
##       fold          population       householdsize     racepctblack
##   Min.   : 1.000   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000
##   1st Qu.: 3.000   1st Qu.:0.01000   1st Qu.:0.3500   1st Qu.:0.0200
##   Median : 5.000   Median :0.02000   Median :0.4400   Median :0.0600
##   Mean   : 5.494   Mean   :0.05759   Mean   :0.4634   Mean   :0.1796
##   3rd Qu.: 8.000   3rd Qu.:0.05000   3rd Qu.:0.5400   3rd Qu.:0.2300
##   Max.   :10.000   Max.   :1.00000   Max.   :1.0000   Max.   :1.0000
##   racepctwhite     racepctasian     racepcthisp      agepct12t21
##   Min.   :0.0000   Min.   :0.0000   Min.   :0.000   Min.   :0.0000
##   1st Qu.:0.6300   1st Qu.:0.0400   1st Qu.:0.010   1st Qu.:0.3400
##   Median :0.8500   Median :0.0700   Median :0.040   Median :0.4000
##   Mean   :0.7537   Mean   :0.1537   Mean   :0.144   Mean   :0.4242
##   3rd Qu.:0.9400   3rd Qu.:0.1700   3rd Qu.:0.160   3rd Qu.:0.4700
##   Max.   :1.0000   Max.   :1.0000   Max.   :1.000   Max.   :1.0000
##   agepct12t29      agepct16t24      agepct65up        numburban
##   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
##   1st Qu.:0.4100   1st Qu.:0.2500   1st Qu.:0.3000   1st Qu.:0.00000
##   Median :0.4800   Median :0.2900   Median :0.4200   Median :0.03000
##   Mean   :0.4939   Mean   :0.3363   Mean   :0.4232   Mean   :0.06407
##   3rd Qu.:0.5400   3rd Qu.:0.3600   3rd Qu.:0.5300   3rd Qu.:0.07000
##   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.00000
##    pcturban         medIncome        pctWWage        pctWFarmSelf
##   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##   1st Qu.:0.0000   1st Qu.:0.2000   1st Qu.:0.4400   1st Qu.:0.1600
##   Median :1.0000   Median :0.3200   Median :0.5600   Median :0.2300
```

2

```
##   Mean    :0.6963   Mean    :0.3611   Mean    :0.5582   Mean    :0.2916
##   3rd Qu.:1.0000   3rd Qu.:0.4900   3rd Qu.:0.6900   3rd Qu.:0.3700
##   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000
##     pctWInvInc        pctWSocSec        pctWPubAsst        pctWRetire
##   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
##   1st Qu.:0.3700   1st Qu.:0.3500   1st Qu.:0.1425   1st Qu.:0.3600
##   Median :0.4800   Median :0.4750   Median :0.2600   Median :0.4700
##   Mean    :0.4957   Mean    :0.4711   Mean    :0.3178   Mean    :0.4792
##   3rd Qu.:0.6200   3rd Qu.:0.5800   3rd Qu.:0.4400   3rd Qu.:0.5800
##   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000
##     medFamInc         perCapInc         whitePerCap       blackPerCap
##   Min.    :0.0000   Min.    :0.0000   Min.    :0.000   Min.    :0.0000
##   1st Qu.:0.2300   1st Qu.:0.2200   1st Qu.:0.240   1st Qu.:0.1725
##   Median :0.3300   Median :0.3000   Median :0.320   Median :0.2500
##   Mean    :0.3757   Mean    :0.3503   Mean    :0.368   Mean    :0.2911
##   3rd Qu.:0.4800   3rd Qu.:0.4300   3rd Qu.:0.440   3rd Qu.:0.3800
##   Max.    :1.0000   Max.    :1.0000   Max.    :1.000   Max.    :1.0000
##     indianPercap      AsianPerCap       OtherPerCap        HisPerCap
##   Min.    :0.0000   Min.    :0.0000   Length:1994       Min.    :0.0000
##   1st Qu.:0.1100   1st Qu.:0.1900   Class :character   1st Qu.:0.2600
##   Median :0.1700   Median :0.2800   Mode  :character   Median :0.3450
##   Mean    :0.2035   Mean    :0.3224                     Mean    :0.3863
##   3rd Qu.:0.2500   3rd Qu.:0.4000                       3rd Qu.:0.4800
##   Max.    :1.0000   Max.    :1.0000                     Max.    :1.0000
##     NumUnderPov       pctPopUnderPov    pctLess9thGrade   PctNotHSGrad
##   Min.    :0.00000   Min.    :0.000    Min.    :0.0000   Min.    :0.0000
##   1st Qu.:0.01000   1st Qu.:0.110    1st Qu.:0.1600   1st Qu.:0.2300
##   Median :0.02000   Median :0.250    Median :0.2700   Median :0.3600
##   Mean    :0.05551   Mean    :0.303    Mean    :0.3158   Mean    :0.3833
##   3rd Qu.:0.05000   3rd Qu.:0.450    3rd Qu.:0.4200   3rd Qu.:0.5100
##   Max.    :1.00000   Max.    :1.000    Max.    :1.0000   Max.    :1.0000
##     PctBSorMore       PctUnemployed      PctEmploy        PctEmpolManu
##   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
##   1st Qu.:0.2100   1st Qu.:0.2200   1st Qu.:0.3800   1st Qu.:0.2500
##   Median :0.3100   Median :0.3200   Median :0.5100   Median :0.3700
##   Mean    :0.3617   Mean    :0.3635   Mean    :0.5011   Mean    :0.3964
##   3rd Qu.:0.4600   3rd Qu.:0.4800   3rd Qu.:0.6275   3rd Qu.:0.5200
##   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000
##   PctEmpProfServ    PctOccupManu     PctOccupMgmtProf MalepctDivorce
##   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
##   1st Qu.:0.3200   1st Qu.:0.2400   1st Qu.:0.3100   1st Qu.:0.3300
##   Median :0.4100   Median :0.3700   Median :0.4000   Median :0.4700
##   Mean    :0.4406   Mean    :0.3912   Mean    :0.4413   Mean    :0.4612
##   3rd Qu.:0.5300   3rd Qu.:0.5100   3rd Qu.:0.5400   3rd Qu.:0.5900
##   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000
##   MalePctNevMarr    FemalePctDiv      TotalPctDiv       PersPerFam
##   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000   Min.    :0.0000
##   1st Qu.:0.3100   1st Qu.:0.3600   1st Qu.:0.3600   1st Qu.:0.4000
##   Median :0.4000   Median :0.5000   Median :0.5000   Median :0.4700
##   Mean    :0.4345   Mean    :0.4876   Mean    :0.4943   Mean    :0.4877
##   3rd Qu.:0.5000   3rd Qu.:0.6200   3rd Qu.:0.6300   3rd Qu.:0.5600
##   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000   Max.    :1.0000
##     PercFam2Par       PctKids2Par      PctYoungKids2Par  PctTeen2Par
##   Min.    :0.0000   Min.    :0.0000   Min.    :0.000   Min.    :0.0000
```

```
##   1st Qu.:0.4900    1st Qu.:0.4900    1st Qu.:0.530    1st Qu.:0.4800
##   Median :0.6300    Median :0.6400    Median :0.700    Median :0.6100
##   Mean   :0.6109    Mean   :0.6207    Mean   :0.664    Mean   :0.5829
##   3rd Qu.:0.7600    3rd Qu.:0.7800    3rd Qu.:0.840    3rd Qu.:0.7200
##   Max.   :1.0000    Max.   :1.0000    Max.   :1.000    Max.   :1.0000
##   PctWorkMomYoungKids   PctWorkMom      NumIllegal        PctIllegal
##   Min.   :0.0000        Min.   :0.0000  Min.   :0.00000   Min.   :0.00
##   1st Qu.:0.3900        1st Qu.:0.4200  1st Qu.:0.00000   1st Qu.:0.09
##   Median :0.5100        Median :0.5400  Median :0.01000   Median :0.17
##   Mean   :0.5014        Mean   :0.5267  Mean   :0.03629   Mean   :0.25
##   3rd Qu.:0.6200        3rd Qu.:0.6500  3rd Qu.:0.02000   3rd Qu.:0.32
##   Max.   :1.0000        Max.   :1.0000  Max.   :1.00000   Max.   :1.00
##    NumImmigrant      PctImmigrantRct  PctImmigrantRec5  PctImmgRec8
##   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000    Min.   :0.0000
##   1st Qu.:0.00000   1st Qu.:0.1600   1st Qu.:0.2000    1st Qu.:0.2500
##   Median :0.01000   Median :0.2900   Median :0.3400    Median :0.3900
##   Mean   :0.03006   Mean   :0.3202   Mean   :0.3606    Mean   :0.3991
##   3rd Qu.:0.02000   3rd Qu.:0.4300   3rd Qu.:0.4800    3rd Qu.:0.5300
##   Max.   :1.00000   Max.   :1.0000   Max.   :1.0000    Max.   :1.0000
##   PctImmigRec10     PctRecentImmig    PctRecImmig5      PctRecImmig8
##   Min.   :0.0000    Min.   :0.0000    Min.   :0.0000    Min.   :0.0000
##   1st Qu.:0.2800    1st Qu.:0.0300    1st Qu.:0.0300    1st Qu.:0.0300
##   Median :0.4300    Median :0.0900    Median :0.0800    Median :0.0900
##   Mean   :0.4279    Mean   :0.1814    Mean   :0.1821    Mean   :0.1848
##   3rd Qu.:0.5600    3rd Qu.:0.2300    3rd Qu.:0.2300    3rd Qu.:0.2300
##   Max.   :1.0000    Max.   :1.0000    Max.   :1.0000    Max.   :1.0000
##   PctRecImmig10     PctSpeakEnglOnly PctNotSpeakEnglWell PctLargeHouse
##   Min.   :0.0000    Min.   :0.0000   Min.   :0.0000      Min.   :0.0000
##   1st Qu.:0.0300    1st Qu.:0.7300   1st Qu.:0.0300      1st Qu.:0.1500
##   Median :0.0900    Median :0.8700   Median :0.0600      Median :0.2000
##   Mean   :0.1829    Mean   :0.7859   Mean   :0.1506      Mean   :0.2676
##   3rd Qu.:0.2300    3rd Qu.:0.9400   3rd Qu.:0.1600      3rd Qu.:0.3100
##   Max.   :1.0000    Max.   :1.0000   Max.   :1.0000      Max.   :1.0000
##   PctLargHouseOccup PersPerOccupHouse PersPerOwnOccHouse PersPerRentOccHouse
##   Min.   :0.0000    Min.   :0.0000    Min.   :0.0000     Min.   :0.0000
##   1st Qu.:0.1400    1st Qu.:0.3400    1st Qu.:0.3900     1st Qu.:0.2700
##   Median :0.1900    Median :0.4400    Median :0.4800     Median :0.3600
##   Mean   :0.2519    Mean   :0.4621    Mean   :0.4944     Mean   :0.4041
##   3rd Qu.:0.2900    3rd Qu.:0.5500    3rd Qu.:0.5800     3rd Qu.:0.4900
##   Max.   :1.0000    Max.   :1.0000    Max.   :1.0000     Max.   :1.0000
##   PctPersOwnOccup   PctPersDenseHouse PctHouseLess3BR    MedNumBR
##   Min.   :0.0000    Min.   :0.0000    Min.   :0.0000     Min.   :0.0000
##   1st Qu.:0.4400    1st Qu.:0.0600    1st Qu.:0.4000     1st Qu.:0.0000
##   Median :0.5600    Median :0.1100    Median :0.5100     Median :0.5000
##   Mean   :0.5626    Mean   :0.1863    Mean   :0.4952     Mean   :0.3147
##   3rd Qu.:0.7000    3rd Qu.:0.2200    3rd Qu.:0.6000     3rd Qu.:0.5000
##   Max.   :1.0000    Max.   :1.0000    Max.   :1.0000     Max.   :1.0000
##    HouseVacant       PctHouseOccup     PctHouseOwnOccu    PctVacantBoarded
##   Min.   :0.00000   Min.   :0.0000    Min.   :0.0000     Min.   :0.0000
##   1st Qu.:0.01000   1st Qu.:0.6300    1st Qu.:0.4300     1st Qu.:0.0600
##   Median :0.03000   Median :0.7700    Median :0.5400     Median :0.1300
##   Mean   :0.07682   Mean   :0.7195    Mean   :0.5487     Mean   :0.2045
##   3rd Qu.:0.07000   3rd Qu.:0.8600    3rd Qu.:0.6700     3rd Qu.:0.2700
##   Max.   :1.00000   Max.   :1.0000    Max.   :1.0000     Max.   :1.0000
```

4

```
##   PctVacMore6Mos  MedYrHouseBuilt  PctHouseNoPhone  PctWOFullPlumb
##   Min.   :0.0000  Min.   :0.0000   Min.   :0.0000   Min.    :0.0000
##   1st Qu.:0.2900  1st Qu.:0.3500   1st Qu.:0.0600   1st Qu.:0.1000
##   Median :0.4200  Median :0.5200   Median :0.1850   Median :0.1900
##   Mean   :0.4333  Mean   :0.4942   Mean   :0.2645   Mean    :0.2431
##   3rd Qu.:0.5600  3rd Qu.:0.6700   3rd Qu.:0.4200   3rd Qu.:0.3300
##   Max.   :1.0000  Max.   :1.0000   Max.   :1.0000   Max.    :1.0000
##   OwnOccLowQuart   OwnOccMedVal     OwnOccHiQuart       RentLowQ
##   Min.   :0.0000  Min.   :0.0000   Min.   :0.0000   Min.    :0.0000
##   1st Qu.:0.0900  1st Qu.:0.0900   1st Qu.:0.0900   1st Qu.:0.1700
##   Median :0.1800  Median :0.1700   Median :0.1800   Median :0.3100
##   Mean   :0.2647  Mean   :0.2635   Mean   :0.2689   Mean    :0.3464
##   3rd Qu.:0.4000  3rd Qu.:0.3900   3rd Qu.:0.3800   3rd Qu.:0.4900
##   Max.   :1.0000  Max.   :1.0000   Max.   :1.0000   Max.    :1.0000
##     RentMedian       RentHighQ         MedRent       MedRentPctHouseInc
##   Min.   :0.0000  Min.   :0.000    Min.   :0.0000   Min.    :0.0000
##   1st Qu.:0.2000  1st Qu.:0.220    1st Qu.:0.2100   1st Qu.:0.3700
##   Median :0.3300  Median :0.370    Median :0.3400   Median :0.4800
##   Mean   :0.3725  Mean   :0.423    Mean   :0.3841   Mean    :0.4901
##   3rd Qu.:0.5200  3rd Qu.:0.590    3rd Qu.:0.5300   3rd Qu.:0.5900
##   Max.   :1.0000  Max.   :1.000    Max.   :1.0000   Max.    :1.0000
##   MedOwnCostPctInc MedOwnCostPctIncNoMtg NumInShelters       NumStreet
##   Min.   :0.0000  Min.   :0.0000        Min.   :0.00000   Min.    :0.00000
##   1st Qu.:0.3200  1st Qu.:0.2500        1st Qu.:0.00000   1st Qu.:0.00000
##   Median :0.4500  Median :0.3700        Median :0.00000   Median :0.00000
##   Mean   :0.4498  Mean   :0.4038        Mean   :0.02944   Mean    :0.02278
##   3rd Qu.:0.5800  3rd Qu.:0.5100        3rd Qu.:0.01000   3rd Qu.:0.00000
##   Max.   :1.0000  Max.   :1.0000        Max.   :1.00000   Max.    :1.00000
##   PctForeignBorn   PctBornSameState PctSameHouse85    PctSameCity85
##   Min.   :0.0000  Min.   :0.0000   Min.   :0.0000   Min.    :0.0000
##   1st Qu.:0.0600  1st Qu.:0.4700   1st Qu.:0.4200   1st Qu.:0.5200
##   Median :0.1300  Median :0.6300   Median :0.5400   Median :0.6700
##   Mean   :0.2156  Mean   :0.6089   Mean   :0.5351   Mean    :0.6264
##   3rd Qu.:0.2800  3rd Qu.:0.7775   3rd Qu.:0.6600   3rd Qu.:0.7700
##   Max.   :1.0000  Max.   :1.0000   Max.   :1.0000   Max.    :1.0000
##   PctSameState85  LemasSwornFT     LemasSwFTPerPop    LemasSwFTFieldOps
##   Min.   :0.0000  Length:1994      Length:1994       Length:1994
##   1st Qu.:0.5600  Class :character Class :character  Class :character
##   Median :0.7000  Mode  :character Mode  :character  Mode  :character
##   Mean   :0.6515
##   3rd Qu.:0.7900
##   Max.   :1.0000
##   LemasSwFTFieldPerPop LemasTotalReq    LemasTotReqPerPop  PoliceReqPerOffic
##   Length:1994          Length:1994      Length:1994        Length:1994
##   Class :character     Class :character Class :character   Class :character
##   Mode  :character     Mode  :character Mode  :character   Mode  :character
##
##
##
##   PolicePerPop      RacialMatchCommPol PctPoliceWhite     PctPoliceBlack
##   Length:1994       Length:1994        Length:1994        Length:1994
##   Class :character  Class :character   Class :character   Class :character
##   Mode  :character  Mode  :character   Mode  :character   Mode  :character
##
```

```
##
##
## PctPoliceHisp     PctPoliceAsian    PctPoliceMinor    OfficAssgnDrugUnits
## Length:1994       Length:1994       Length:1994       Length:1994
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
## NumKindsDrugsSeiz  PoliceAvgOTWorked    LandArea          PopDens
## Length:1994       Length:1994       Min.   :0.00000   Min.   :0.0000
## Class :character   Class :character   1st Qu.:0.02000   1st Qu.:0.1000
## Mode  :character   Mode  :character   Median :0.04000   Median :0.1700
##                                       Mean   :0.06523   Mean   :0.2329
##                                       3rd Qu.:0.07000   3rd Qu.:0.2800
##                                       Max.   :1.00000   Max.   :1.0000
## PctUsePubTrans     PoliceCars        PoliceOperBudg    LemasPctPoliceOnPatrol
## Min.   :0.0000    Length:1994       Length:1994       Length:1994
## 1st Qu.:0.0200    Class :character   Class :character   Class :character
## Median :0.0700    Mode  :character   Mode  :character   Mode  :character
## Mean   :0.1617
## 3rd Qu.:0.1900
## Max.   :1.0000
## LemasGangUnitDeploy LemasPctOfficDrugUn PliceBudgPerPop   ViolentCrimesPerPop
## Length:1994       Min.   :0.00000   Length:1994       Min.   :0.000
## Class :character   1st Qu.:0.00000   Class :character   1st Qu.:0.070
## Mode  :character   Median :0.00000   Mode  :character   Median :0.150
##                    Mean   :0.09405                      Mean   :0.238
##                    3rd Qu.:0.00000                      3rd Qu.:0.330
##                    Max.   :1.00000                      Max.   :1.000
```

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr   0.3.4
## v tibble  3.1.4      v dplyr   1.0.7
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   2.0.1      v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(ggplot2)
```

```r
wa5<- read.csv("wa5counties.csv")
summary(wa5)
```

```
##  state_numeric county_numeric    community_numeric  community_name
## Min.   :53     Length:5          Length:5           Length:5
## 1st Qu.:53     Class :character   Class :character   Class :character
```

```
##   Median :53     Mode  :character   Mode  :character   Mode  :character
##   Mean   :53
##   3rd Qu.:53
##   Max.   :53
##        fold        population    householdsize    racepctblack    racepctwhite
##   Min.   : 2.0   Min.   :0.100   Min.   :0.19   Min.   :0.030   Min.   :0.630
##   1st Qu.: 3.0   1st Qu.:0.120   1st Qu.:0.29   1st Qu.:0.040   1st Qu.:0.670
##   Median : 3.0   Median :0.270   Median :0.32   Median :0.040   Median :0.800
##   Mean   : 5.2   Mean   :0.314   Mean   :0.30   Mean   :0.106   Mean   :0.776
##   3rd Qu.: 8.0   3rd Qu.:0.270   3rd Qu.:0.33   3rd Qu.:0.200   3rd Qu.:0.880
##   Max.   :10.0   Max.   :0.810   Max.   :0.37   Max.   :0.220   Max.   :0.900
##    racepctasian    racepcthisp    agepct12t21    agepct12t29    agepct16t24
##   Min.   :0.130   Min.   :0.040   Min.   :0.270   Min.   :0.47   Min.   :0.290
##   1st Qu.:0.240   1st Qu.:0.040   1st Qu.:0.340   1st Qu.:0.47   1st Qu.:0.320
##   Median :0.420   Median :0.050   Median :0.340   Median :0.49   Median :0.320
##   Mean   :0.424   Mean   :0.052   Mean   :0.346   Mean   :0.49   Mean   :0.318
##   3rd Qu.:0.610   3rd Qu.:0.060   3rd Qu.:0.390   3rd Qu.:0.50   3rd Qu.:0.330
##   Max.   :0.720   Max.   :0.070   Max.   :0.390   Max.   :0.52   Max.   :0.330
##    agepct65up     numburban      medincome      medIncome
##   Min.   :0.33   Min.   :0.110   Min.   :0.17   Length:5
##   1st Qu.:0.40   1st Qu.:0.140   1st Qu.:0.23   Class :character
##   Median :0.42   Median :0.280   Median :0.28   Mode  :character
##   Mean   :0.43   Mean   :0.326   Mean   :0.30
##   3rd Qu.:0.48   3rd Qu.:0.280   3rd Qu.:0.29
##   Max.   :0.52   Max.   :0.820   Max.   :0.53
##     pctWWage      pctWFarmSelf     pctWInvInc      pctWSocSec     pctWPubAsst
##   Min.   :0.370   Min.   :0.150   Min.   :0.410   Min.   :0.30   Min.   :0.120
##   1st Qu.:0.430   1st Qu.:0.210   1st Qu.:0.420   1st Qu.:0.44   1st Qu.:0.300
##   Median :0.520   Median :0.220   Median :0.460   Median :0.44   Median :0.520
##   Mean   :0.514   Mean   :0.218   Mean   :0.522   Mean   :0.44   Mean   :0.404
##   3rd Qu.:0.530   3rd Qu.:0.240   3rd Qu.:0.610   3rd Qu.:0.47   3rd Qu.:0.530
##   Max.   :0.720   Max.   :0.270   Max.   :0.710   Max.   :0.55   Max.   :0.550
##    pctWRetire     medFamInc       perCapInc      whitePerCap
##   Min.   :0.430   Min.   :0.220   Min.   :0.240   Min.   :0.250
##   1st Qu.:0.470   1st Qu.:0.260   1st Qu.:0.250   1st Qu.:0.270
##   Median :0.500   Median :0.290   Median :0.300   Median :0.300
##   Mean   :0.514   Mean   :0.348   Mean   :0.376   Mean   :0.398
##   3rd Qu.:0.560   3rd Qu.:0.380   3rd Qu.:0.450   3rd Qu.:0.510
##   Max.   :0.610   Max.   :0.590   Max.   :0.640   Max.   :0.660
##    blackPerCap    indianPercap     AsianPerCap     OtherPerCap      HisPerCap
##   Min.   :0.220   Min.   :0.100   Min.   :0.160   Min.   :0.240   Min.   :0.27
##   1st Qu.:0.230   1st Qu.:0.120   1st Qu.:0.170   1st Qu.:0.250   1st Qu.:0.28
##   Median :0.230   Median :0.120   Median :0.200   Median :0.300   Median :0.37
##   Mean   :0.264   Mean   :0.162   Mean   :0.234   Mean   :0.302   Mean   :0.40
##   3rd Qu.:0.270   3rd Qu.:0.180   3rd Qu.:0.270   3rd Qu.:0.320   3rd Qu.:0.52
##   Max.   :0.370   Max.   :0.290   Max.   :0.370   Max.   :0.400   Max.   :0.56
##    NumUnderPov    pctPopUnderPov  pctLess9thGrade  PctNotHSGrad     PctBSorMore
##   Min.   :0.04   Min.   :0.130   Min.   :0.06   Min.   :0.070   Min.   :0.21
##   1st Qu.:0.07   1st Qu.:0.310   1st Qu.:0.17   1st Qu.:0.210   1st Qu.:0.24
##   Median :0.26   Median :0.320   Median :0.19   Median :0.270   Median :0.33
##   Mean   :0.24   Mean   :0.332   Mean   :0.17   Mean   :0.244   Mean   :0.43
##   3rd Qu.:0.27   3rd Qu.:0.440   3rd Qu.:0.20   3rd Qu.:0.320   3rd Qu.:0.62
##   Max.   :0.56   Max.   :0.460   Max.   :0.23   Max.   :0.350   Max.   :0.75
##   PctUnemployed     PctEmploy      PctEmpolManu    PctEmpProfServ
```

```
##  Min.   :0.130   Min.   :0.350   Min.   :0.220   Min.   :0.250
##  1st Qu.:0.280   1st Qu.:0.400   1st Qu.:0.280   1st Qu.:0.370
##  Median :0.390   Median :0.480   Median :0.350   Median :0.450
##  Mean   :0.358   Mean   :0.498   Mean   :0.384   Mean   :0.444
##  3rd Qu.:0.460   3rd Qu.:0.570   3rd Qu.:0.380   3rd Qu.:0.560
##  Max.   :0.530   Max.   :0.690   Max.   :0.690   Max.   :0.590
##   PctOccupManu    PctOccupMgmtProf MalepctDivorce  MalePctNevMarr
##  Min.   :0.150   Min.   :0.300    Min.   :0.45   Min.   :0.440
##  1st Qu.:0.260   1st Qu.:0.340    1st Qu.:0.70   1st Qu.:0.440
##  Median :0.350   Median :0.420    Median :0.76   Median :0.450
##  Mean   :0.338   Mean   :0.468    Mean   :0.71   Mean   :0.504
##  3rd Qu.:0.450   3rd Qu.:0.600    3rd Qu.:0.76   3rd Qu.:0.480
##  Max.   :0.480   Max.   :0.680    Max.   :0.88   Max.   :0.710
##   FemalePctDiv     TotalPctDiv      PersPerFam      PercFam2Par      PctKids2Par
##  Min.   :0.580   Min.   :0.540   Min.   :0.29   Min.   :0.410   Min.   :0.440
##  1st Qu.:0.670   1st Qu.:0.710   1st Qu.:0.36   1st Qu.:0.420   1st Qu.:0.470
##  Median :0.770   Median :0.800   Median :0.36   Median :0.420   Median :0.470
##  Mean   :0.724   Mean   :0.748   Mean   :0.37   Mean   :0.488   Mean   :0.524
##  3rd Qu.:0.770   3rd Qu.:0.800   3rd Qu.:0.38   3rd Qu.:0.470   3rd Qu.:0.480
##  Max.   :0.830   Max.   :0.890   Max.   :0.46   Max.   :0.720   Max.   :0.760
##  PctYoungKids2Par PctTeen2Par   PctWorkMomYoungKids  PctWorkMom
##  Min.   :0.49    Min.   :0.39   Min.   :0.270     Min.   :0.30
##  1st Qu.:0.49    1st Qu.:0.42   1st Qu.:0.360     1st Qu.:0.41
##  Median :0.50    Median :0.42   Median :0.380     Median :0.44
##  Mean   :0.57    Mean   :0.47   Mean   :0.406     Mean   :0.44
##  3rd Qu.:0.61    3rd Qu.:0.44   3rd Qu.:0.430     3rd Qu.:0.48
##  Max.   :0.76    Max.   :0.68   Max.   :0.590     Max.   :0.57
##    NumIllegal      PctIllegal      NumImmigrant   PctImmigrantRct PctImmigrantRec5
##  Min.   :0.010   Min.   :0.09   Min.   :0.03   Min.   :0.230   Min.   :0.320
##  1st Qu.:0.040   1st Qu.:0.24   1st Qu.:0.05   1st Qu.:0.270   1st Qu.:0.340
##  Median :0.090   Median :0.30   Median :0.08   Median :0.330   Median :0.350
##  Mean   :0.106   Mean   :0.25   Mean   :0.14   Mean   :0.372   Mean   :0.404
##  3rd Qu.:0.120   3rd Qu.:0.31   3rd Qu.:0.09   3rd Qu.:0.450   3rd Qu.:0.480
##  Max.   :0.270   Max.   :0.31   Max.   :0.45   Max.   :0.580   Max.   :0.530
##   PctImmgRec8     PctImmigRec10   PctRecentImmig  PctRecImmig5    PctRecImmig8
##  Min.   :0.36   Min.   :0.350   Min.   :0.09   Min.   :0.09   Min.   :0.080
##  1st Qu.:0.45   1st Qu.:0.510   1st Qu.:0.10   1st Qu.:0.13   1st Qu.:0.160
##  Median :0.48   Median :0.510   Median :0.15   Median :0.16   Median :0.210
##  Mean   :0.48   Mean   :0.504   Mean   :0.26   Mean   :0.24   Mean   :0.252
##  3rd Qu.:0.54   3rd Qu.:0.570   3rd Qu.:0.42   3rd Qu.:0.39   3rd Qu.:0.390
##  Max.   :0.57   Max.   :0.580   Max.   :0.54   Max.   :0.43   Max.   :0.420
##  PctRecImmig10   PctSpeakEnglOnly PctNotSpeakEnglWell PctLargeHouse
##  Min.   :0.070   Min.   :0.740   Min.   :0.050    Min.   :0.170
##  1st Qu.:0.170   1st Qu.:0.780   1st Qu.:0.110    1st Qu.:0.200
##  Median :0.200   Median :0.830   Median :0.150    Median :0.200
##  Mean   :0.244   Mean   :0.828   Mean   :0.138    Mean   :0.218
##  3rd Qu.:0.380   3rd Qu.:0.870   3rd Qu.:0.180    3rd Qu.:0.200
##  Max.   :0.400   Max.   :0.920   Max.   :0.200    Max.   :0.320
##  PctLargHouseOccup PersPerOccupHouse PersPerOwnOccHouse PersPerRentOccHouse
##  Min.   :0.130    Min.   :0.17    Min.   :0.290    Min.   :0.130
##  1st Qu.:0.150    1st Qu.:0.28    1st Qu.:0.350    1st Qu.:0.240
##  Median :0.150    Median :0.33    Median :0.370    Median :0.250
##  Mean   :0.172    Mean   :0.30    Mean   :0.374    Mean   :0.268
##  3rd Qu.:0.170    3rd Qu.:0.35    3rd Qu.:0.410    3rd Qu.:0.360
```

```
##   Max.   :0.260    Max.   :0.37     Max.   :0.450    Max.   :0.360
##  PctPersOwnOccup PctPersDenseHouse PctHouseLess3BR    MedNumBR
##  Min.   :0.310   Min.   :0.11      Min.   :0.460    Min.   :0.0
##  1st Qu.:0.410   1st Qu.:0.11      1st Qu.:0.590    1st Qu.:0.0
##  Median :0.420   Median :0.18      Median :0.590    Median :0.0
##  Mean   :0.438   Mean   :0.17      Mean   :0.608    Mean   :0.1
##  3rd Qu.:0.510   3rd Qu.:0.22      3rd Qu.:0.690    3rd Qu.:0.0
##  Max.   :0.540   Max.   :0.23      Max.   :0.710    Max.   :0.5
##   HouseVacant     PctHouseOccup   PctHouseOwnOccu PctVacantBoarded
##  Min.   :0.090   Min.   :0.730   Min.   :0.320   Min.   :0.030
##  1st Qu.:0.110   1st Qu.:0.730   1st Qu.:0.360   1st Qu.:0.190
##  Median :0.250   Median :0.770   Median :0.420   Median :0.250
##  Mean   :0.278   Mean   :0.778   Mean   :0.414   Mean   :0.302
##  3rd Qu.:0.280   3rd Qu.:0.820   3rd Qu.:0.480   3rd Qu.:0.420
##  Max.   :0.660   Max.   :0.840   Max.   :0.490   Max.   :0.620
##  PctVacMore6Mos  MedYrHouseBuilt PctHouseNoPhone PctWOFullPlumb
##  Min.   :0.080   Min.   :0.21    Min.   :0.05    Min.   :0.100
##  1st Qu.:0.240   1st Qu.:0.27    1st Qu.:0.16    1st Qu.:0.200
##  Median :0.260   Median :0.33    Median :0.27    Median :0.280
##  Mean   :0.276   Mean   :0.40    Mean   :0.21    Mean   :0.282
##  3rd Qu.:0.370   3rd Qu.:0.52    3rd Qu.:0.27    3rd Qu.:0.370
##  Max.   :0.430   Max.   :0.67    Max.   :0.30    Max.   :0.460
##  OwnOccLowQuart   OwnOccMedVal    OwnOccHiQuart     RentLowQ       RentMedian
##  Min.   :0.080   Min.   :0.070   Min.   :0.08    Min.   :0.170   Min.   :0.190
##  1st Qu.:0.130   1st Qu.:0.120   1st Qu.:0.12    1st Qu.:0.250   1st Qu.:0.280
##  Median :0.220   Median :0.210   Median :0.24    Median :0.350   Median :0.360
##  Mean   :0.236   Mean   :0.246   Mean   :0.28    Mean   :0.332   Mean   :0.342
##  3rd Qu.:0.280   3rd Qu.:0.330   3rd Qu.:0.39    3rd Qu.:0.350   3rd Qu.:0.370
##  Max.   :0.470   Max.   :0.500   Max.   :0.57    Max.   :0.540   Max.   :0.510
##    RentHighQ        MedRent      MedRentPctHouseInc MedOwnCostPctInc
##  Min.   :0.200   Min.   :0.190   Min.   :0.45     Min.   :0.33
##  1st Qu.:0.310   1st Qu.:0.270   1st Qu.:0.52     1st Qu.:0.33
##  Median :0.390   Median :0.330   Median :0.54     Median :0.44
##  Mean   :0.388   Mean   :0.334   Mean   :0.54     Mean   :0.40
##  3rd Qu.:0.450   3rd Qu.:0.360   3rd Qu.:0.59     3rd Qu.:0.44
##  Max.   :0.590   Max.   :0.520   Max.   :0.60     Max.   :0.46
##  MedOwnCostPctIncNoMtg NumInShelters      NumStreet      PctForeignBorn
##  Min.   :0.180         Min.   :0.010   Min.   :0.020   Min.   :0.110
##  1st Qu.:0.220         1st Qu.:0.130   1st Qu.:0.030   1st Qu.:0.190
##  Median :0.350         Median :0.150   Median :0.040   Median :0.230
##  Mean   :0.322         Mean   :0.304   Mean   :0.214   Mean   :0.262
##  3rd Qu.:0.430         3rd Qu.:0.230   3rd Qu.:0.130   3rd Qu.:0.390
##  Max.   :0.430         Max.   :1.000   Max.   :0.850   Max.   :0.390
##  PctBornSameState PctSameHouse85  PctSameCity85    PctSameState85
##  Min.   :0.330    Min.   :0.340   Min.   :0.560   Min.   :0.42
##  1st Qu.:0.390    1st Qu.:0.410   1st Qu.:0.560   1st Qu.:0.49
##  Median :0.470    Median :0.410   Median :0.600   Median :0.55
##  Mean   :0.446    Mean   :0.412   Mean   :0.604   Mean   :0.53
##  3rd Qu.:0.510    3rd Qu.:0.450   3rd Qu.:0.630   3rd Qu.:0.58
##  Max.   :0.530    Max.   :0.450   Max.   :0.670   Max.   :0.61
##   LemasSwornFT    LemasSwFTPerPop LemasSwFTFieldOps LemasSwFTFieldPerPop
##  Min.   :0.010   Min.   :0.09    Min.   :0.870     Min.   :0.110
##  1st Qu.:0.020   1st Qu.:0.10    1st Qu.:0.950     1st Qu.:0.130
##  Median :0.040   Median :0.10    Median :0.950     Median :0.130
```

```
## Mean   :0.044    Mean   :0.12    Mean   :0.942    Mean    :0.152
## 3rd Qu.:0.040    3rd Qu.:0.13    3rd Qu.:0.970    3rd Qu.:0.170
## Max.   :0.110    Max.   :0.18    Max.   :0.970    Max.    :0.220
## LemasTotalReq    LemasTotReqPerPop  PoliceReqPerOffic  PolicePerPop
## Min.   :0.030    Min.   :0.120     Min.   :0.29      Min.    :0.09
## 1st Qu.:0.060    1st Qu.:0.130     1st Qu.:0.38      1st Qu.:0.10
## Median :0.070    Median :0.130     Median :0.40      Median :0.10
## Mean   :0.084    Mean   :0.172     Mean   :0.43      Mean    :0.12
## 3rd Qu.:0.090    3rd Qu.:0.190     3rd Qu.:0.54      3rd Qu.:0.13
## Max.   :0.170    Max.   :0.290     Max.   :0.54      Max.    :0.18
## RacialMatchCommPol PctPoliceWhite  PctPoliceBlack  PctPoliceHisp
## Min.   :0.660      Min.   :0.820   Min.   :0.040   Min.    :0.000
## 1st Qu.:0.690      1st Qu.:0.830   1st Qu.:0.060   1st Qu.:0.000
## Median :0.890      Median :0.870   Median :0.070   Median :0.030
## Mean   :0.818      Mean   :0.868   Mean   :0.078   Mean    :0.032
## 3rd Qu.:0.920      3rd Qu.:0.900   3rd Qu.:0.100   3rd Qu.:0.040
## Max.   :0.930      Max.   :0.920   Max.   :0.120   Max.    :0.090
## PctPoliceAsian  PctPoliceMinor  OfficAssgnDrugUnits  NumKindsDrugsSeiz
## Min.   :0.000   Min.   :0.070   Min.   :0.030       Min.    :0.430
## 1st Qu.:0.080   1st Qu.:0.110   1st Qu.:0.040       1st Qu.:0.500
## Median :0.800   Median :0.110   Median :0.040       Median :0.570
## Mean   :0.538   Mean   :0.128   Mean   :0.048       Mean    :0.614
## 3rd Qu.:0.810   3rd Qu.:0.170   3rd Qu.:0.060       3rd Qu.:0.640
## Max.   :1.000   Max.   :0.180   Max.   :0.070       Max.    :0.930
## PoliceAvgOTWorked   LandArea      PopDens      PctUsePubTrans    PoliceCars
## Min.   :0.180     Min.   :0.07   Min.   :0.20   Min.   :0.200   Min.   :0.06
## 1st Qu.:0.240     1st Qu.:0.09   1st Qu.:0.26   1st Qu.:0.220   1st Qu.:0.07
## Median :0.320     Median :0.14   Median :0.27   Median :0.250   Median :0.15
## Mean   :0.304     Mean   :0.14   Mean   :0.31   Mean   :0.394   Mean   :0.23
## 3rd Qu.:0.350     3rd Qu.:0.16   3rd Qu.:0.31   3rd Qu.:0.380   3rd Qu.:0.25
## Max.   :0.430     Max.   :0.24   Max.   :0.51   Max.   :0.920   Max.   :0.62
## PoliceOperBudg  LemasPctPoliceOnPatrol LemasGangUnitDeploy LemasPctOfficDrugUn
## Min.   :0.030   Min.   :0.790          Min.   :0.5         Min.   :0.360
## 1st Qu.:0.040   1st Qu.:0.790          1st Qu.:0.5         1st Qu.:0.450
## Median :0.050   Median :0.800          Median :1.0         Median :0.630
## Mean   :0.068   Mean   :0.816          Mean   :0.8         Mean   :0.592
## 3rd Qu.:0.090   3rd Qu.:0.840          3rd Qu.:1.0         3rd Qu.:0.690
## Max.   :0.130   Max.   :0.860          Max.   :1.0         Max.   :0.830
## PliceBudgPerPop ViolentCrimesPerPop
## Min.   :0.090   Min.   :0.090
## 1st Qu.:0.110   1st Qu.:0.220
## Median :0.160   Median :0.330
## Mean   :0.164   Mean   :0.396
## 3rd Qu.:0.210   3rd Qu.:0.580
## Max.   :0.250   Max.   :0.760
```

Does income seem to make a difference in the amount of police in the community? There isn't enough variables here to make an informed decision but it seems like there tends to be more police for less income. With the exception of Bellevue, which is basically twice as rich as everyone else yet has the same amount of police. For some reason, Seattle which has a way bigger population than any of the other cities, has the average income between 5 cities and yet has the least police. Why would this be?

```
library(ggplot2)
ggplot(data=wa5, mapping = aes(x= LemasSwFTFieldOps
, y= medIncome, color= community_name)) +geom_point(size=2)+ggtitle("Avg Income vs How Many Police?")+la
```

Avg Income vs How Many Police?
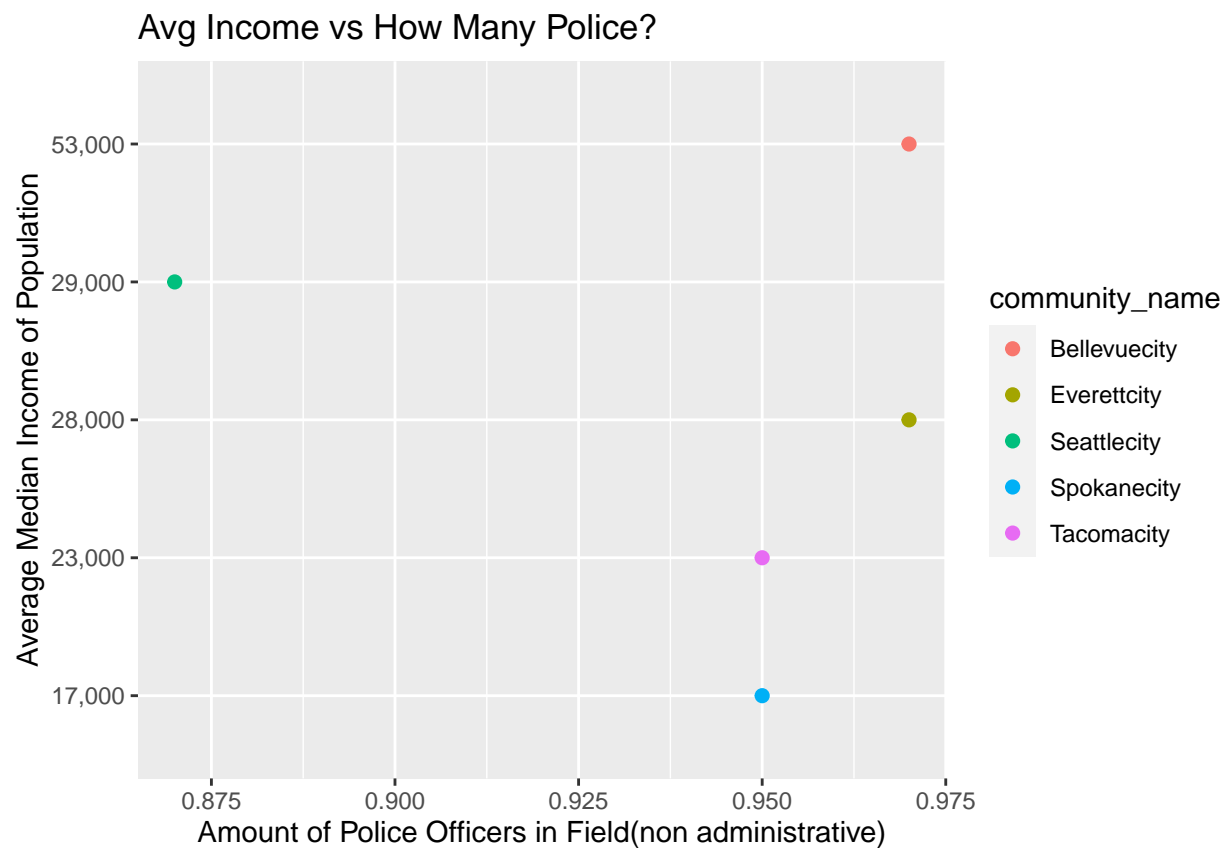

```
ggplot(data=wa5, mapping = aes(x= LemasSwFTFieldOps
, y= community_name, color = community_name)) +geom_point(size=2) +geom_point(size=2)+ggtitle("Amount o
```

Amount of In−Field Police for Every City
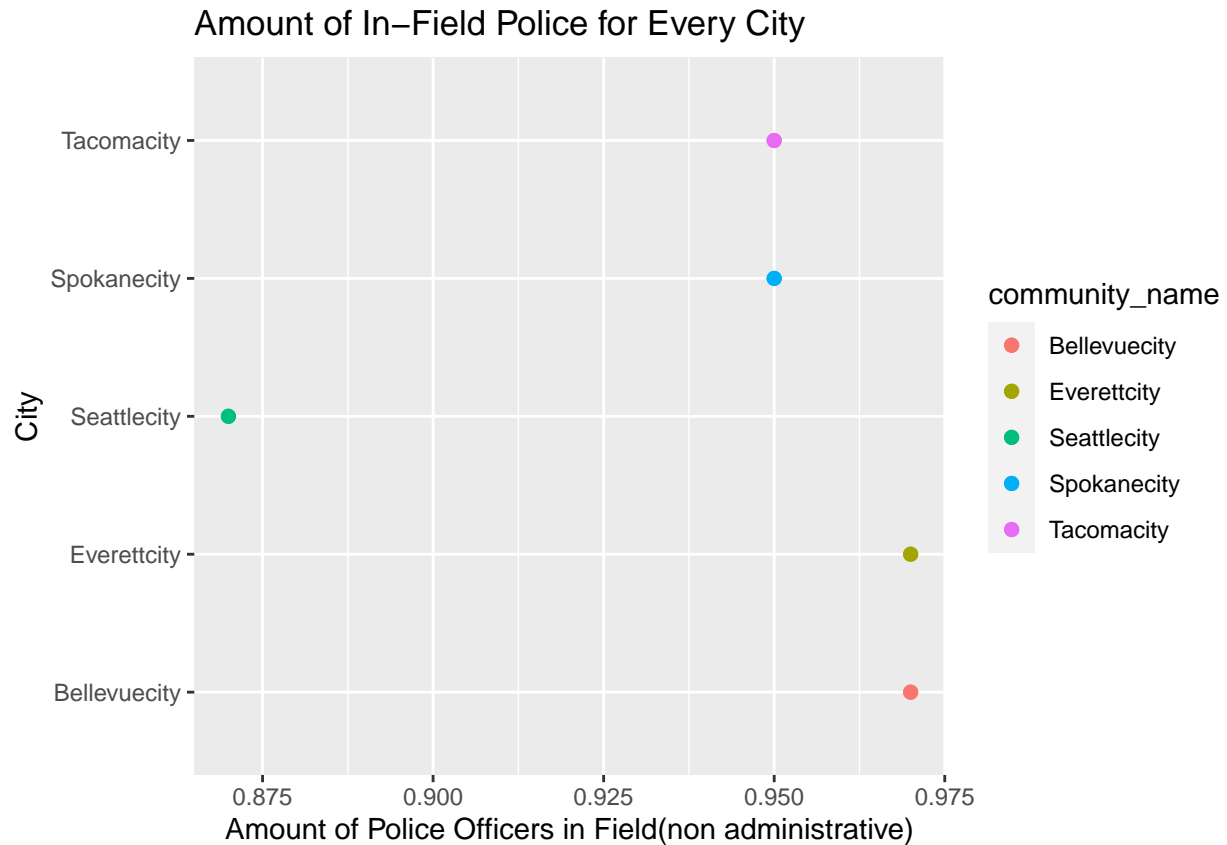
Why do Everett and Bellevue, the least populated cities (by like 70%), have the most in-field police? Perhaps Bellevue is richer so they can pay to have more police? Maybe Everett has more violent crime so they need more police? I'm going to now look at the amount of violent crimes in each city to see if the amount of police makes sense. Also going to look at how many calls the police get for each city. Maybe if there are a lot of police calls in Bellevue, it makes sense for them to have more police?

```
ggplot(data=wa5, mapping = aes(x= ViolentCrimesPerPop
, y=community_name, color= community_name)) +geom_point(size=2) +geom_point(size=2)+ggtitle("Violent Cri
```

Violent Crimes Per 100k for Each City
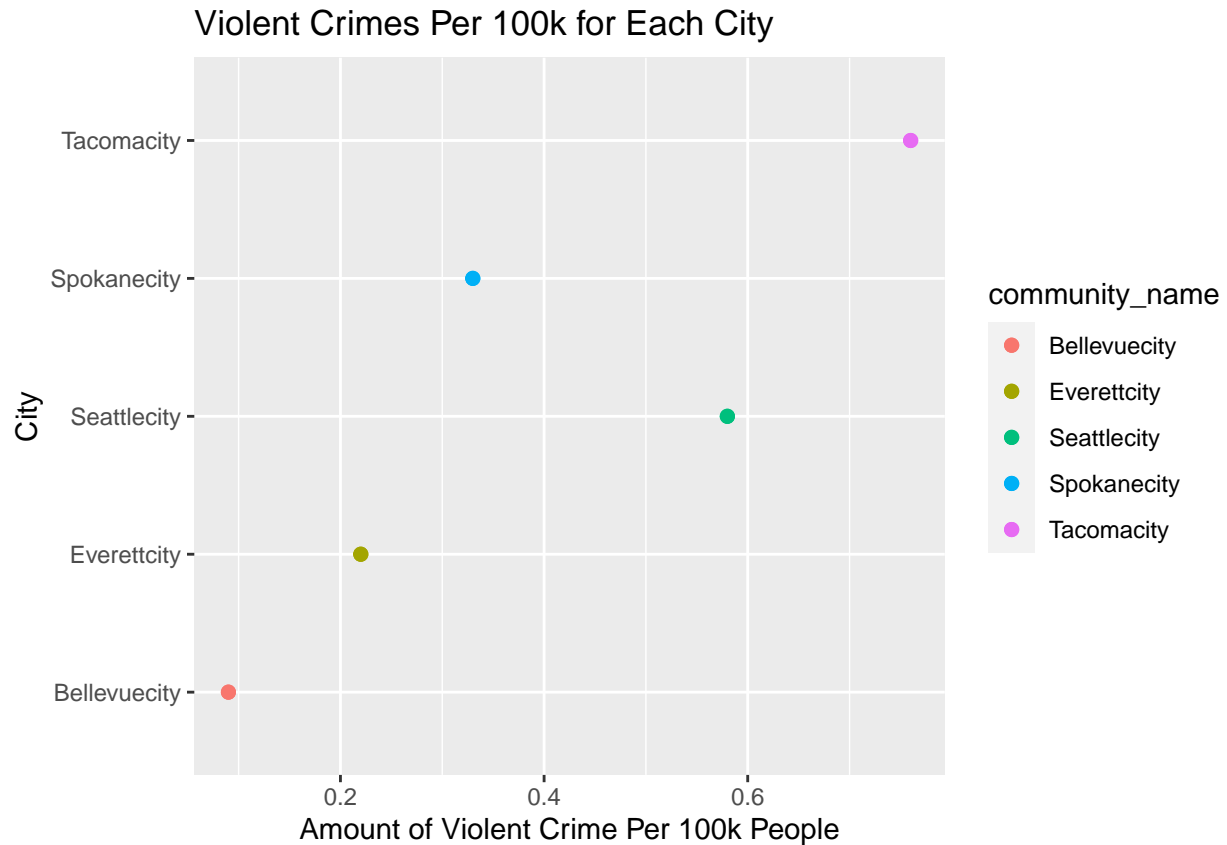
Ok so this is interesting. Everett and Bellevue actually have the least amount of violent crimes per 100k people, and yet they have the most amount of police. Everett's median income amount is almost half of what Bellevue's is. The most violent crimes are in Tacoma, which had the second biggest police force. So that makes sense. But Seattle has the largest population and the second largest amount of violent crime, although it has the smallest amount of police.

```
ggplot(aes(x=LemasTotReqPerPop,y=community_name, color= community_name),data=wa5)+
  geom_point() +geom_point(size=2)+ggtitle("Amount of Police Requests For Each City")+labs(y= "City", x
```

Amount of Police Requests For Each City
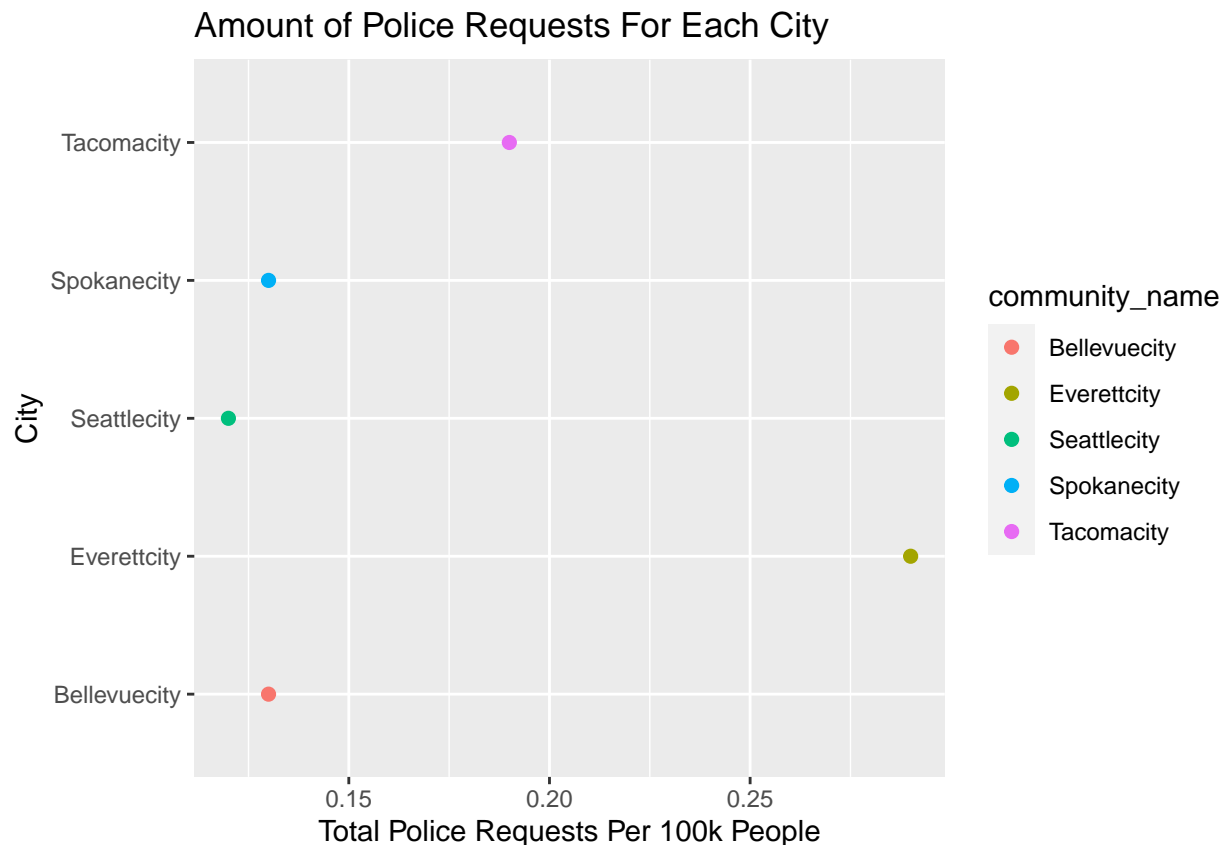
Bellevue and Seattle make sense. There isn't much crime, so there isn't that many requests for police. I wonder why then Bellevue has so many police? Maybe because they are a richer city so they can just afford it. Everett has the most requests for police, even though it has a comparatively low violent crime rate. However, they do have a lot of field officers so it makes sense. Maybe there is just a lot of petty crime rather than violent crime in Everett.
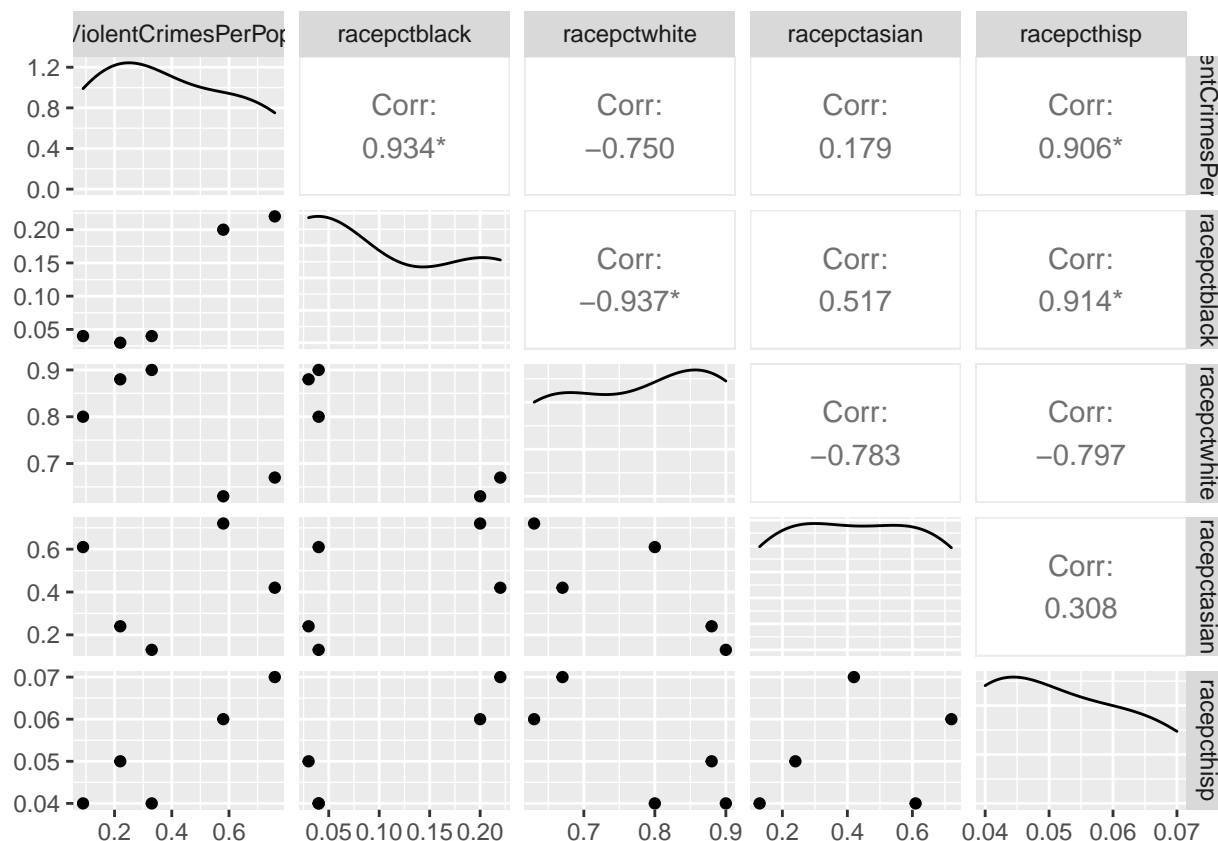
```
library(corrr)
```

```
## Warning: package 'corrr' was built under R version 4.1.2
```

```
library(dplyr)
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
numonly<-read.csv("numericonly.csv")
ggpairs(wa5[, c(128,8,9,10,11)])
```

This correlation table shows there are some pretty big correlations between white, black, and hispanic races and violent crimes. However, there is no correlation between Asians and violent crime. This is only for the five cities though, Seattle, Spokane, Bellevue, Everett and Tacoma.

```
wacrime<- read.csv("wacrime.csv")
pwa<-(wacrime[c(6,8,9,10,11,12,13,17,18,26,27,28,29,30,35,36,37,96)])
d<-dist(wacrime[c(6,8,9,10,11,12,13,17,18,26, 27,28,29,30,35,36,37,96)])
fit <- cmdscale(d, eig=TRUE, k=2)

head(pwa)
```
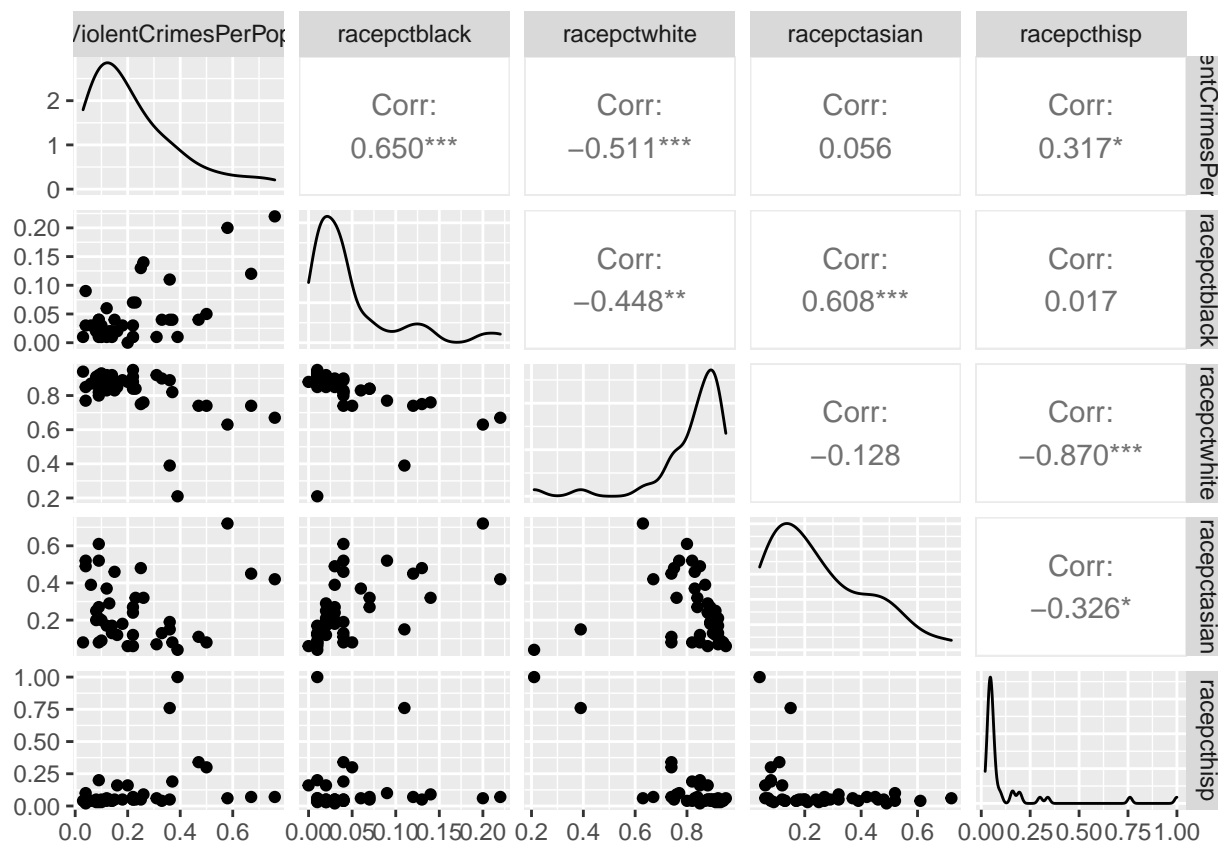
```
##   population racepctblack racepctwhite racepctasian racepcthisp agepct12t21
## 1       0.00         0.12         0.74         0.45        0.07        0.26
## 2       0.00         0.01         0.95         0.06        0.06        0.40
## 3       0.01         0.01         0.93         0.09        0.03        0.36
## 4       0.02         0.00         0.88         0.06        0.16        0.36
## 5       0.04         0.03         0.89         0.18        0.05        0.34
## 6       0.06         0.04         0.89         0.19        0.05        0.34
##   agepct12t29 pcturban medIncome perCapInc whitePerCap blackPerCap indianPercap
## 1        0.59        1      0.31      0.37        0.38        0.33         0.16
## 2        0.44        0      0.17      0.19        0.18        0.16         0.09
## 3        0.36        0      0.22      0.27        0.24        0.16         0.48
## 4        0.40        0      0.18      0.24        0.25        0.46         0.13
## 5        0.50        1      0.30      0.30        0.30        0.32         0.14
## 6        0.49        1      0.16      0.25        0.25        0.19         0.18
##   AsianPerCap pctLess9thGrade PctNotHSGrad PctBSorMore NumStreet
```

```
## 1        0.30        0.14        0.24        0.30        0.00
## 2        0.35        0.32        0.46        0.15        0.00
## 3        0.17        0.20        0.34        0.22        0.00
## 4        0.21        0.34        0.43        0.26        0.01
## 5        0.23        0.18        0.30        0.19        0.01
## 6        0.18        0.25        0.33        0.23        0.05
```

```
ggpairs(pwa)
```


This correlation table is not that helpful, because it is much more difficult to read. I'm going to look at the same few variables that we looked at above but for all of WA.
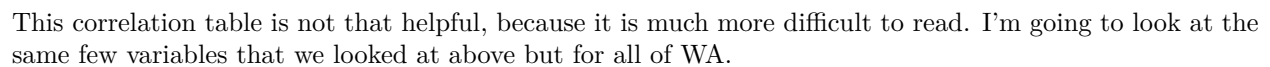
```
ggpairs(wacrime[, c(128,8,9,10,11)])
```

Here we can see that throughout WA, there was the most correlation between violent crime, and the percentage of both white and black people.

Next I'm going to get the summary of the primary components so I can run PCA.

```
wacrime<- read.csv("wacrime.csv")
wapca<-prcomp(pwa, center = TRUE, scale = TRUE)
summary(wapca)
```

```
## Importance of components:
##                           PC1     PC2    PC3    PC4     PC5     PC6     PC7
## Standard deviation     2.7136  1.7431 1.5981 1.3644 1.02924 0.79225 0.67244
## Proportion of Variance 0.4091  0.1688 0.1419 0.1034 0.05885 0.03487 0.02512
## Cumulative Proportion  0.4091  0.5779 0.7198 0.8232 0.88206 0.91693 0.94206
##                           PC8     PC9   PC10    PC11    PC12    PC13   PC14
## Standard deviation     0.58836 0.47504 0.39363 0.35442 0.28567 0.21108 0.1943
## Proportion of Variance 0.01923 0.01254 0.00861 0.00698 0.00453 0.00248 0.0021
## Cumulative Proportion  0.96129 0.97382 0.98243 0.98941 0.99394 0.99642 0.9985
##                          PC15    PC16    PC17    PC18
## Standard deviation     0.12247 0.10095 0.03551 0.01596
## Proportion of Variance 0.00083 0.00057 0.00007 0.00001
## Cumulative Proportion  0.99935 0.99992 0.99999 1.00000
```

Looking at the principal component analysis of the 18 chosen variables.

```
names(wapca)
```

```
## [1] "sdev"     "rotation" "center"   "scale"    "x"
```
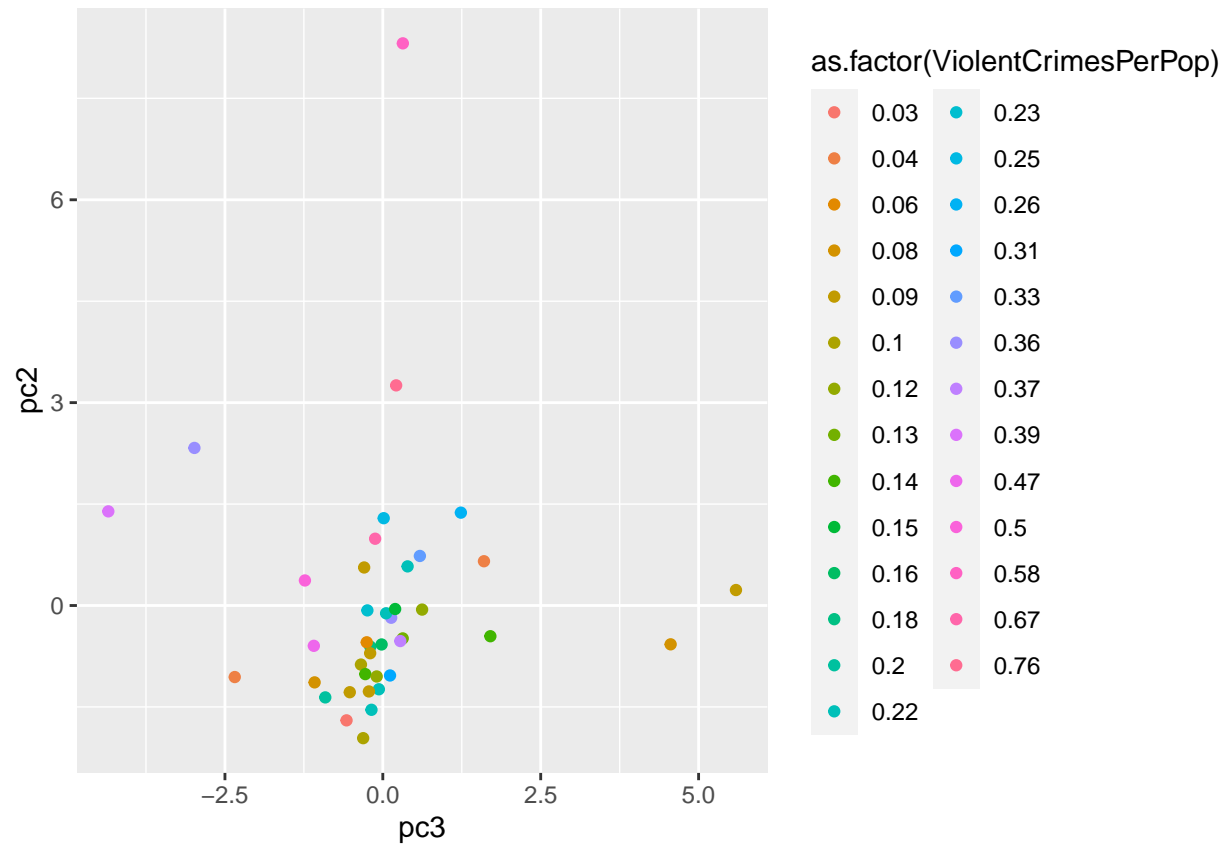
```
round(wapca$rotation,2)
```

```
##                   PC1   PC2   PC3   PC4   PC5   PC6   PC7   PC8   PC9  PC10
## population       0.07  0.48  0.04 -0.20  0.32 -0.24  0.10  0.08  0.02  0.21
## racepctblack     0.03  0.48  0.04 -0.07 -0.25  0.46 -0.07 -0.13 -0.28  0.45
## racepctwhite     0.15 -0.35  0.28 -0.36  0.12 -0.11  0.06 -0.01 -0.12  0.04
## racepctasian     0.22  0.33  0.18  0.15 -0.13  0.33 -0.16  0.11  0.21 -0.26
## racepcthisp     -0.22  0.12 -0.35  0.36 -0.03 -0.09 -0.03  0.02  0.08 -0.07
## agepct12t21     -0.17 -0.02  0.37  0.43  0.03 -0.28  0.10 -0.08 -0.07  0.34
## agepct12t29     -0.12  0.13  0.45  0.32 -0.25 -0.11 -0.03 -0.08 -0.12 -0.51
## pcturban         0.19  0.15 -0.09 -0.16 -0.57 -0.33  0.62 -0.22  0.01 -0.03
## medIncome        0.34 -0.07 -0.20  0.09 -0.05  0.03 -0.01  0.00  0.17 -0.06
## perCapInc        0.35 -0.02 -0.15  0.06  0.07  0.00 -0.01 -0.14  0.20 -0.07
## whitePerCap      0.34  0.04 -0.19  0.11  0.07 -0.01 -0.02 -0.14  0.27 -0.07
## blackPerCap      0.27 -0.04 -0.16  0.18 -0.23 -0.18 -0.03  0.79 -0.34  0.09
## indianPercap     0.21 -0.12 -0.09  0.33  0.38  0.35  0.49 -0.15 -0.50 -0.12
## AsianPerCap      0.29 -0.04 -0.13  0.04 -0.11 -0.31 -0.56 -0.44 -0.46  0.03
## pctLess9thGrade -0.28  0.09 -0.35  0.17  0.03 -0.11  0.04 -0.07  0.02  0.04
## PctNotHSGrad    -0.31  0.05 -0.32  0.02  0.03 -0.08  0.01 -0.07 -0.07 -0.01
## PctBSorMore      0.26  0.05  0.20  0.38  0.18 -0.20  0.02 -0.04  0.29  0.38
## NumStreet        0.06  0.46  0.03 -0.17  0.40 -0.30  0.00  0.08 -0.18 -0.35
##                  PC11  PC12  PC13  PC14  PC15  PC16  PC17  PC18
## population       0.04 -0.09  0.14 -0.66 -0.13 -0.13  0.01 -0.02
## racepctblack     0.30  0.13  0.00  0.23 -0.06 -0.08  0.11  0.00
## racepctwhite     0.09 -0.16  0.11  0.20 -0.15 -0.53  0.46  0.08
## racepctasian    -0.46 -0.48 -0.02  0.01  0.09 -0.19  0.14  0.00
## racepcthisp     -0.09  0.40 -0.17 -0.11  0.14 -0.36  0.56 -0.03
## agepct12t21      0.18 -0.37 -0.37  0.01  0.35 -0.07 -0.03  0.00
## agepct12t29      0.35  0.13  0.27 -0.13 -0.28 -0.03  0.01  0.00
## pcturban        -0.19 -0.01 -0.02  0.05  0.04  0.03  0.02  0.00
## medIncome        0.29 -0.14 -0.61 -0.10 -0.56  0.06  0.06  0.02
## perCapInc        0.33 -0.05  0.19  0.06  0.26 -0.13 -0.04 -0.74
## whitePerCap      0.36 -0.03  0.23  0.04  0.32 -0.04 -0.02  0.67
## blackPerCap      0.08 -0.05  0.14  0.03  0.04 -0.01 -0.03  0.00
## indianPercap    -0.12 -0.02  0.05 -0.10 -0.03  0.00  0.00  0.03
## AsianPerCap     -0.23 -0.04  0.00 -0.12  0.00  0.00 -0.02  0.02
## pctLess9thGrade -0.03 -0.23  0.16  0.25 -0.32 -0.52 -0.47  0.03
## PctNotHSGrad     0.14 -0.53  0.28  0.09 -0.12  0.43  0.45 -0.04
## PctBSorMore     -0.26  0.21  0.28  0.29 -0.34  0.20  0.12 -0.02
## NumStreet        0.04  0.08 -0.26  0.50  0.07  0.09  0.00  0.00
```
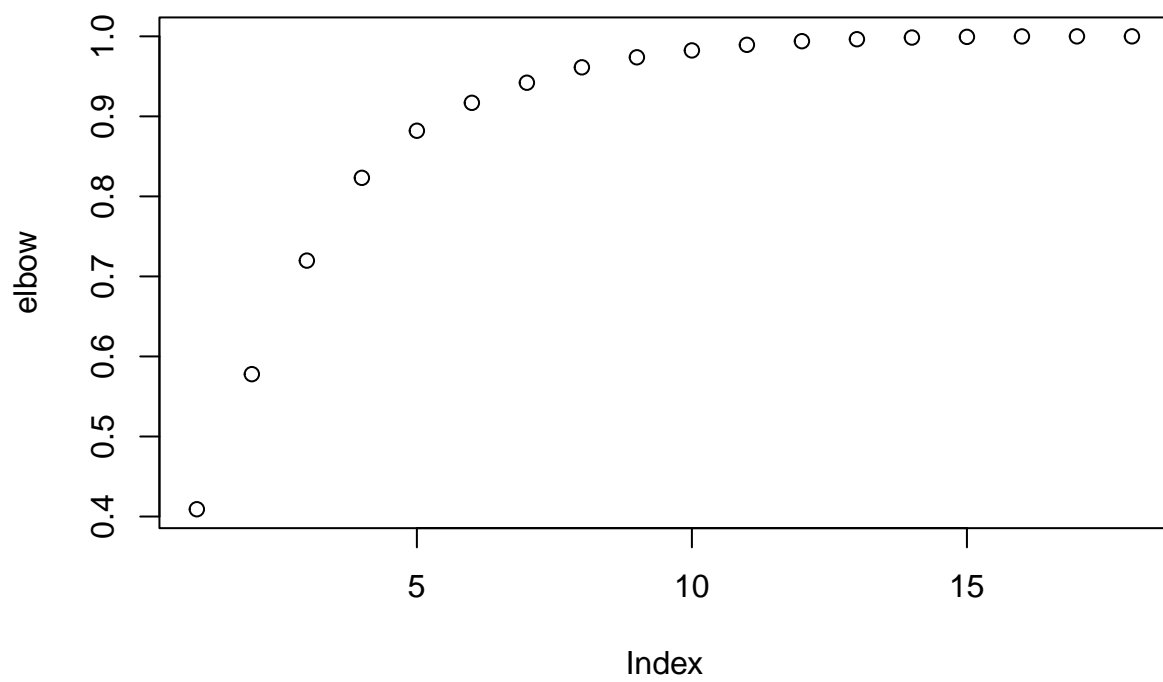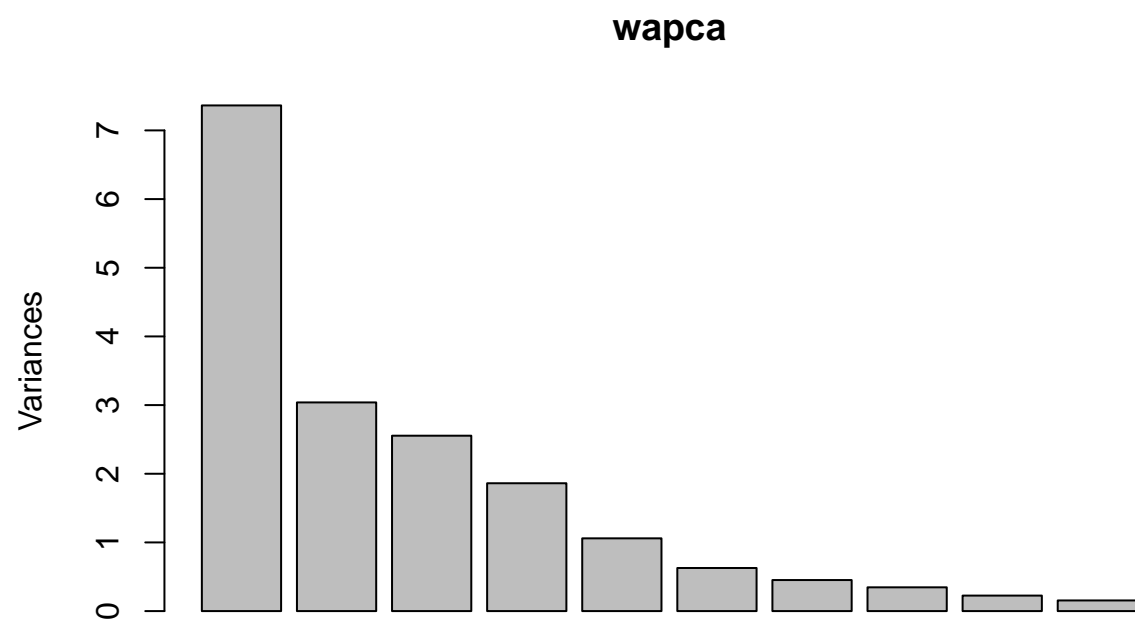
```
sum(wapca$rotation[,1]**2)
```

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

```
wacrime$pc1 <- wapca$x[,1]
wacrime$pc2 <- wapca$x[,2]
wacrime$pc3<- wapca$x[,3]
ggplot(wacrime, aes(x=pc3, y =pc2, color=as.factor(ViolentCrimesPerPop))) +geom_point()
```
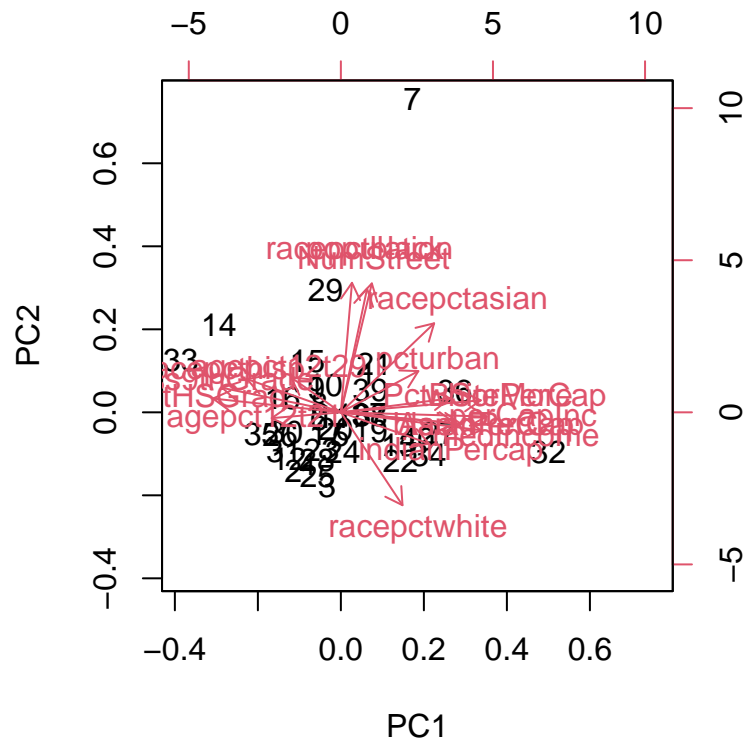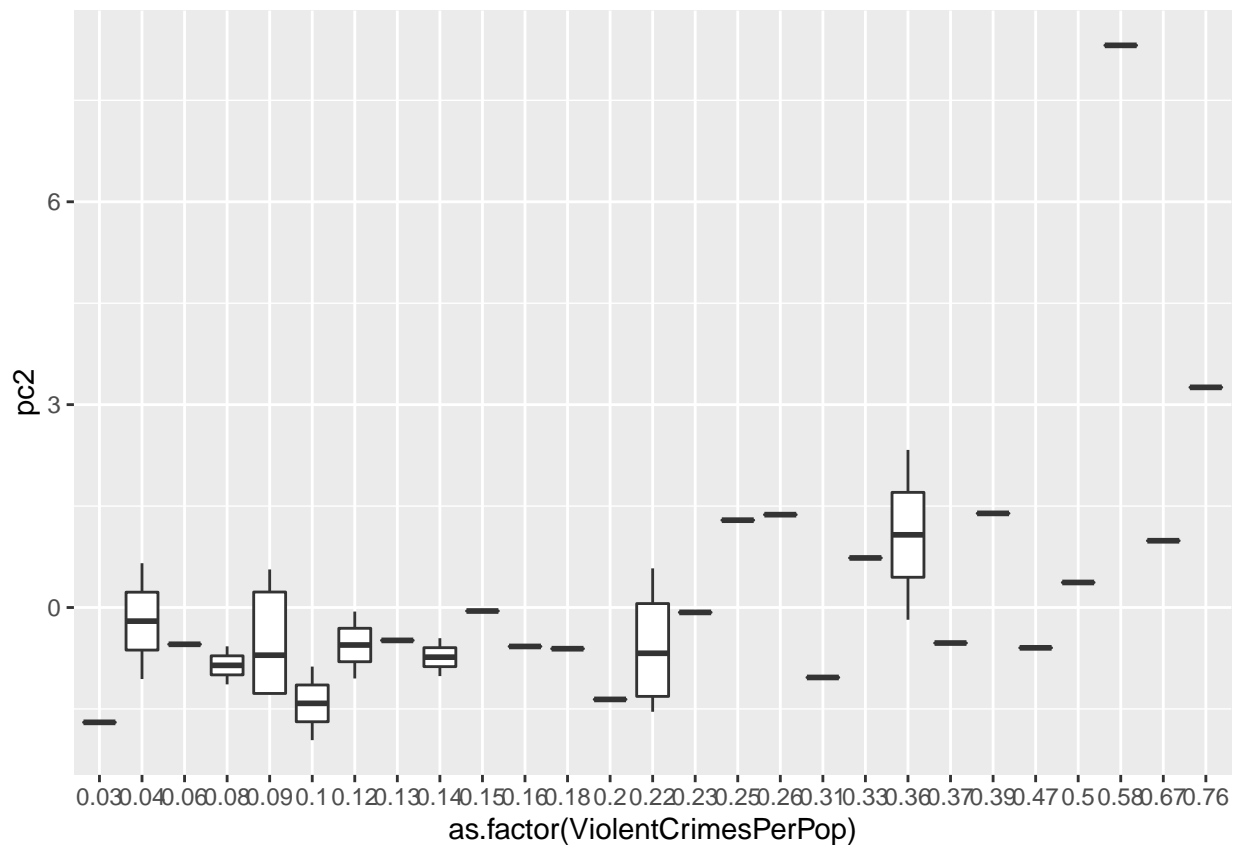
as.factor(ViolentCrimesPerPop)

●	0.03	●	0.23
●	0.04	●	0.25
●	0.06	●	0.26
●	0.08	●	0.31
●	0.09	●	0.33
●	0.1	●	0.36
●	0.12	●	0.37
●	0.13	●	0.39
●	0.14	●	0.47
●	0.15	●	0.5
●	0.16	●	0.58
●	0.18	●	0.67
●	0.2	●	0.76
●	0.22		

```
elbow<-cumsum(wapca$sdev^2/sum(wapca$sdev^2))
plot(elbow)
```

```
screeplot(wapca)
```

wapca


```
biplot(wapca)
```

```
ggplot(wacrime,aes(y=pc3, x= as.factor(ViolentCrimesPerPop)))+geom_boxplot()
```
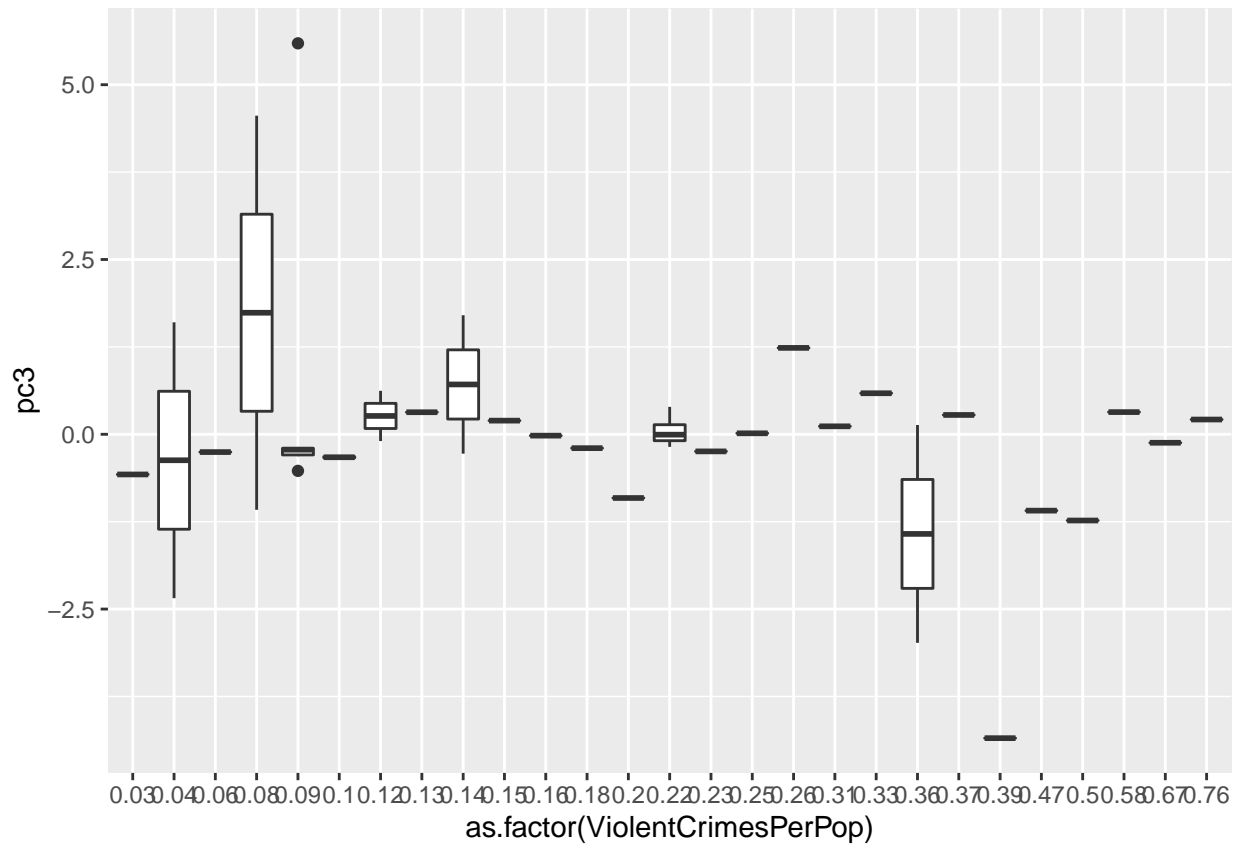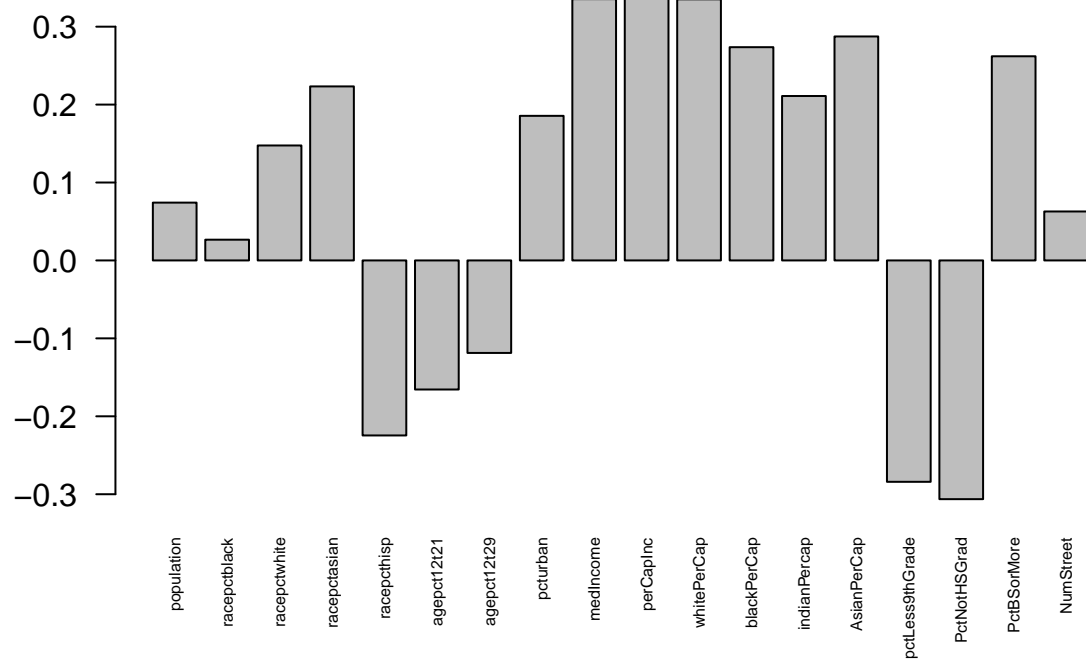
```
ggplot(wacrime,aes(y=pc2, x=as.factor(ViolentCrimesPerPop)))+geom_boxplot()
```
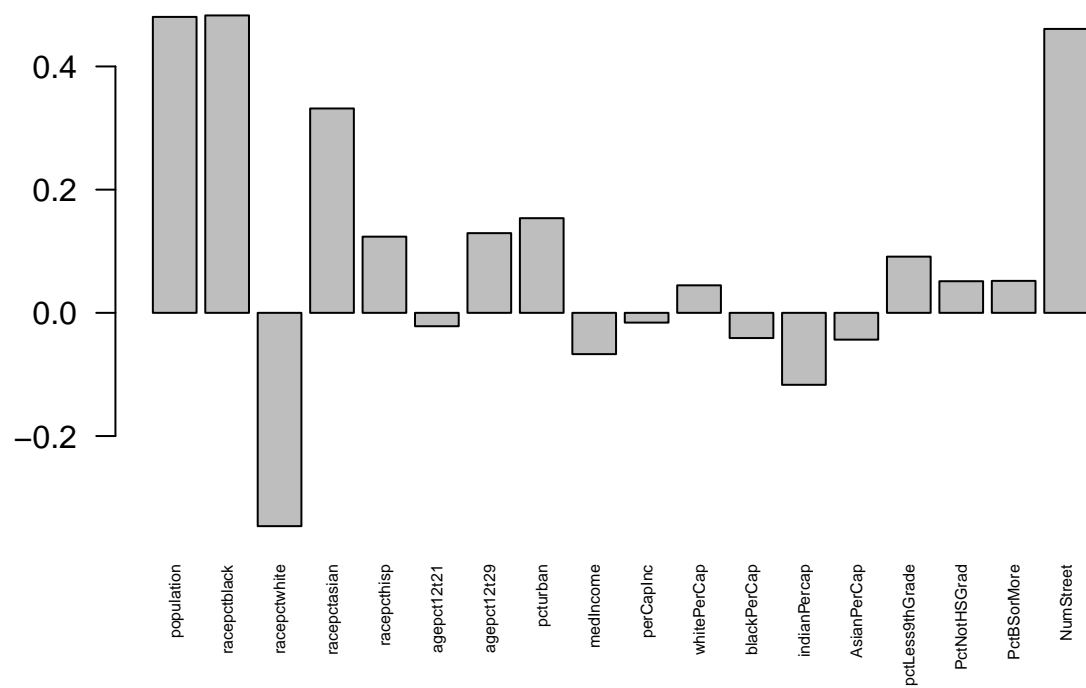
We can see from running Principal component analysis that the top 3 components capture 72% of the variance in the data. The highest ranking components are age12-29(45%), population(48%), racepctblack(48%), and numStreet(46%)(this is the number of homeless people).
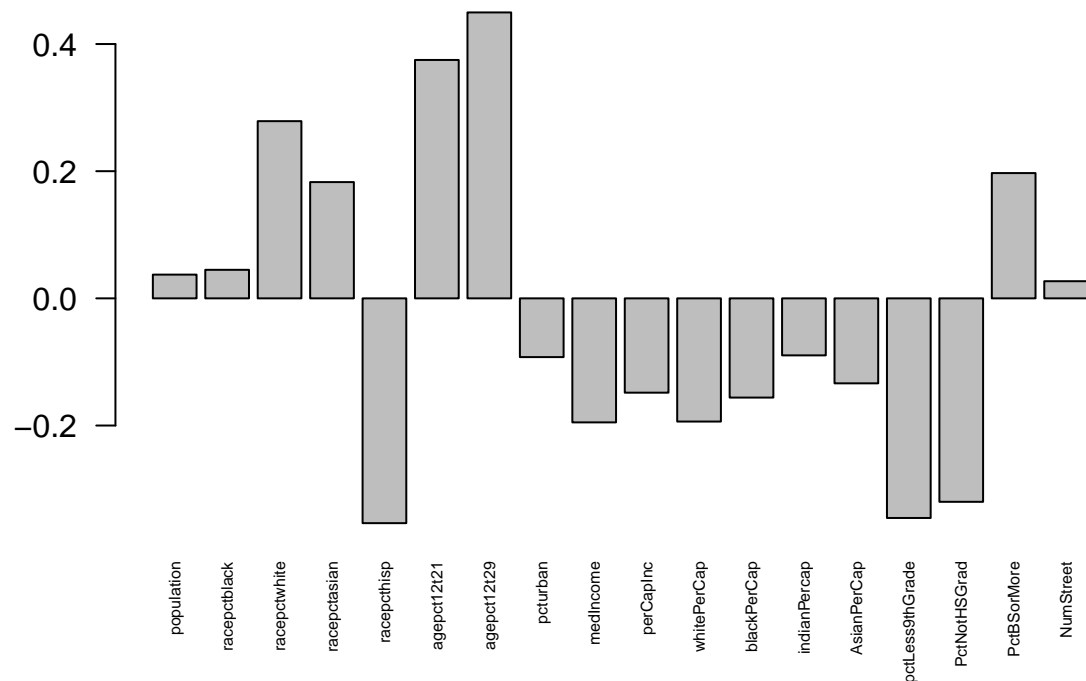
```
barplot(wapca$rotation[,1], las=2, cex.names=.5)
```

```
barplot(wapca$rotation[,2], las=2, cex.names = .5)
```

```
barplot(wapca$rotation[,3], las=2, cex.names = .5)
```

Next, I am going to apply K means clustering to the numeric wacrime columns with 2 clusters. I didn't use pc1 because there wasn't much variance or largely relevant values in pc1. I separate them by city.
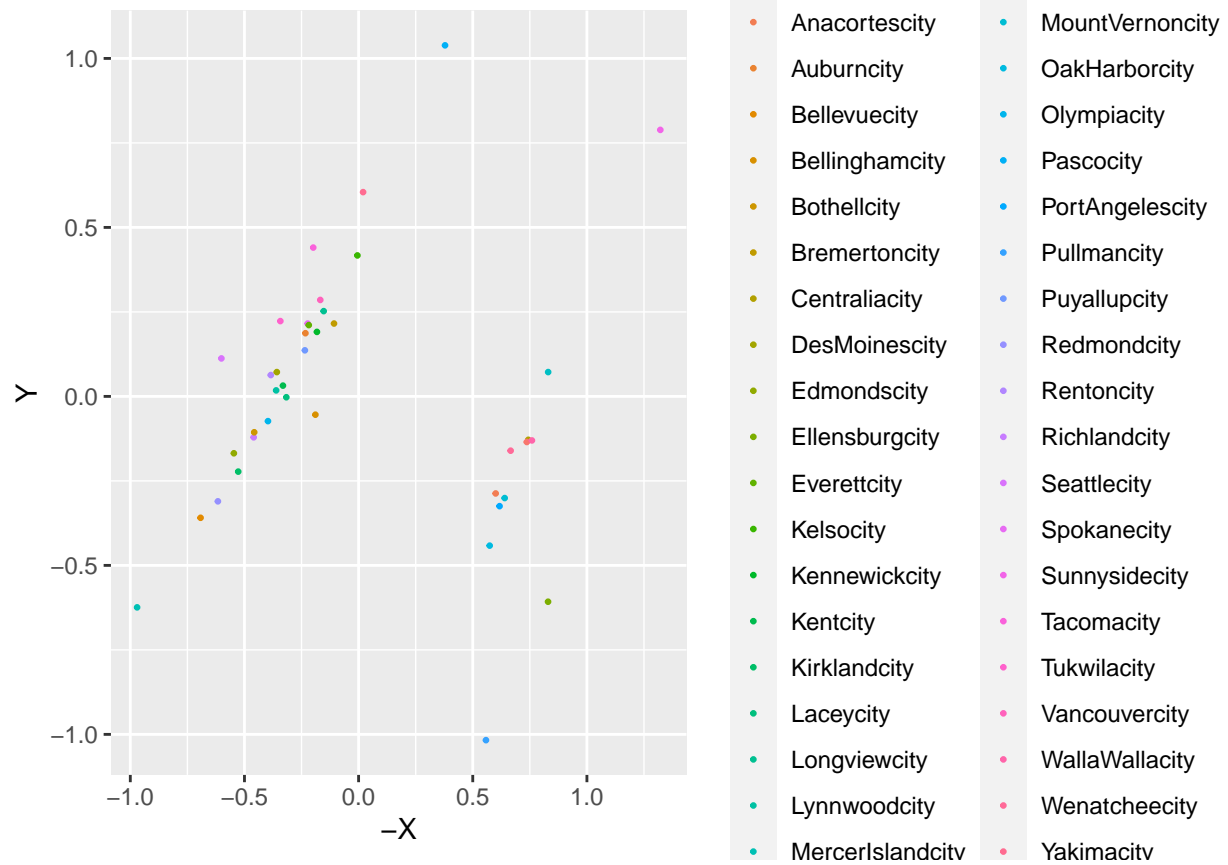
```
pwa<-wacrime[c(4,6,8,9,10,11,12,13,14,17,26,27,28,29,30,35,36,37,96,128)]
d<-dist(wacrime[c(4,6,8,9,10,11,12,13,14,17,26,27,28,29,30,35,36,37,96,128)])
```

```
## Warning in dist(wacrime[c(4, 6, 8, 9, 10, 11, 12, 13, 14, 17, 26, 27, 28, : NAs
## introduced by coercion
```
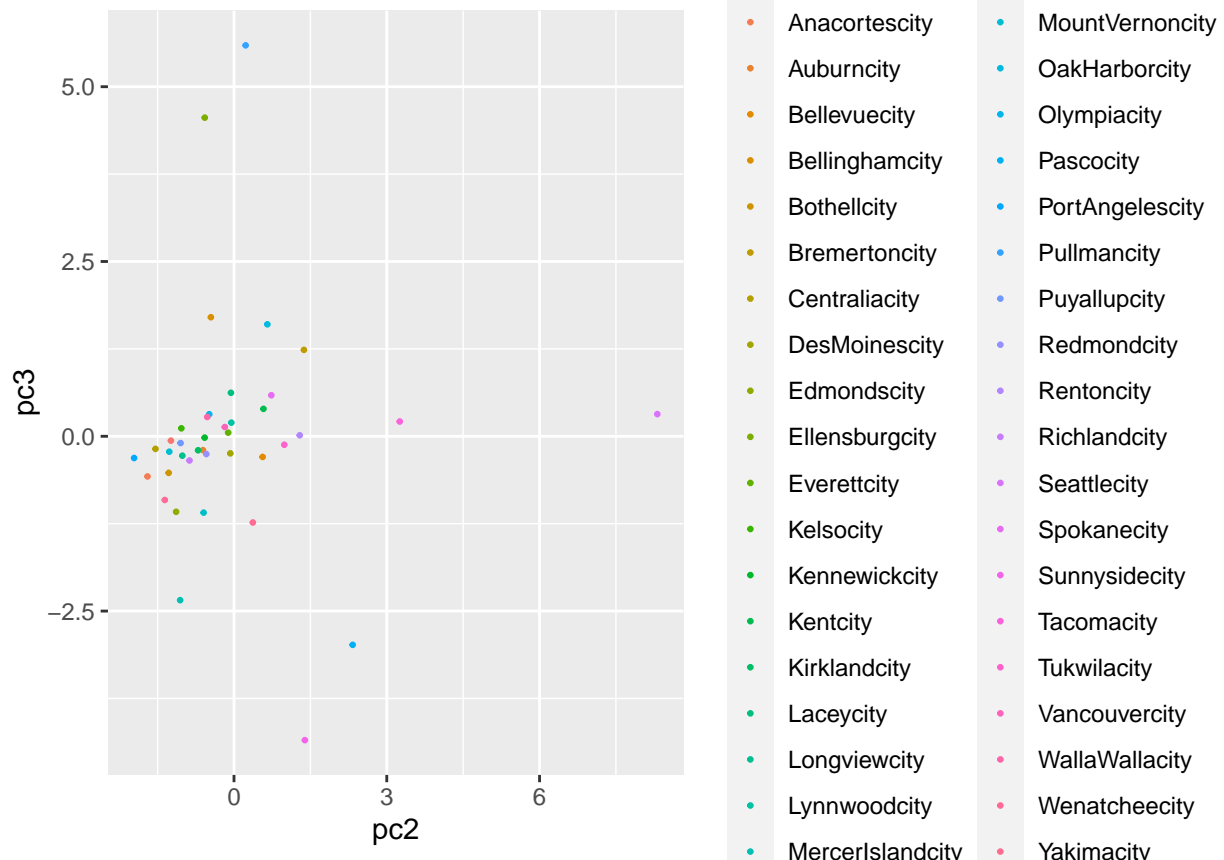
```
fit <- cmdscale(d, eig=TRUE, k=2)

mdsdf<-data.frame(X=fit$points[,1],Y=fit$points[,2], ViolentCrimesPerPop=as.factor(wacrime$community_na

ggplot(mdsdf, aes(x=-X, y = Y, color=as.factor(wacrime$community_name)))+geom_point(size=.5)
```

```
ggplot(wacrime, aes(x=pc2, y = pc3, color=as.factor(community_name)))+geom_point(size=.5)
```

Here I apply k means clustering to the second and third principal components with two clusters.

```
wacrimek2<-kmeans(wacrime[c(6,8,9,10,11,12,13,14,17,26,27,28,29,30,35,36,37,96)],2)
wacrimekw3<-kmeans(wacrime[,129:130],2)
wacrimekw3$tot.withinss
```

```
## [1] 232.7613
```

```
wapca$tot.withinss
```

```
## NULL
```
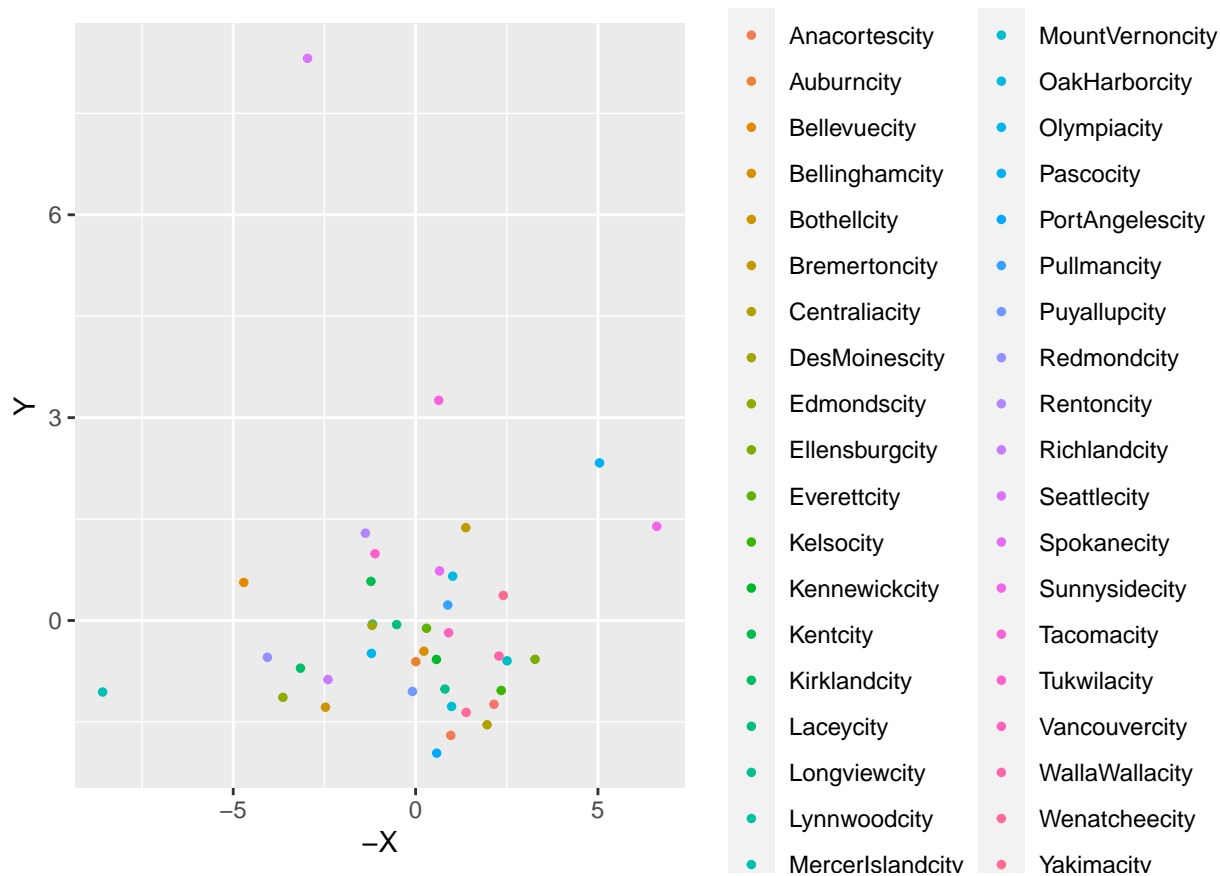
```
wacrime$cluster2 <-wacrimekw3$cluster
wapca$cluster2<-wapca$cluster

ggplot(wacrime,aes(x= pc2, y = pc3, color=as.factor(cluster2)))+geom_point()
```
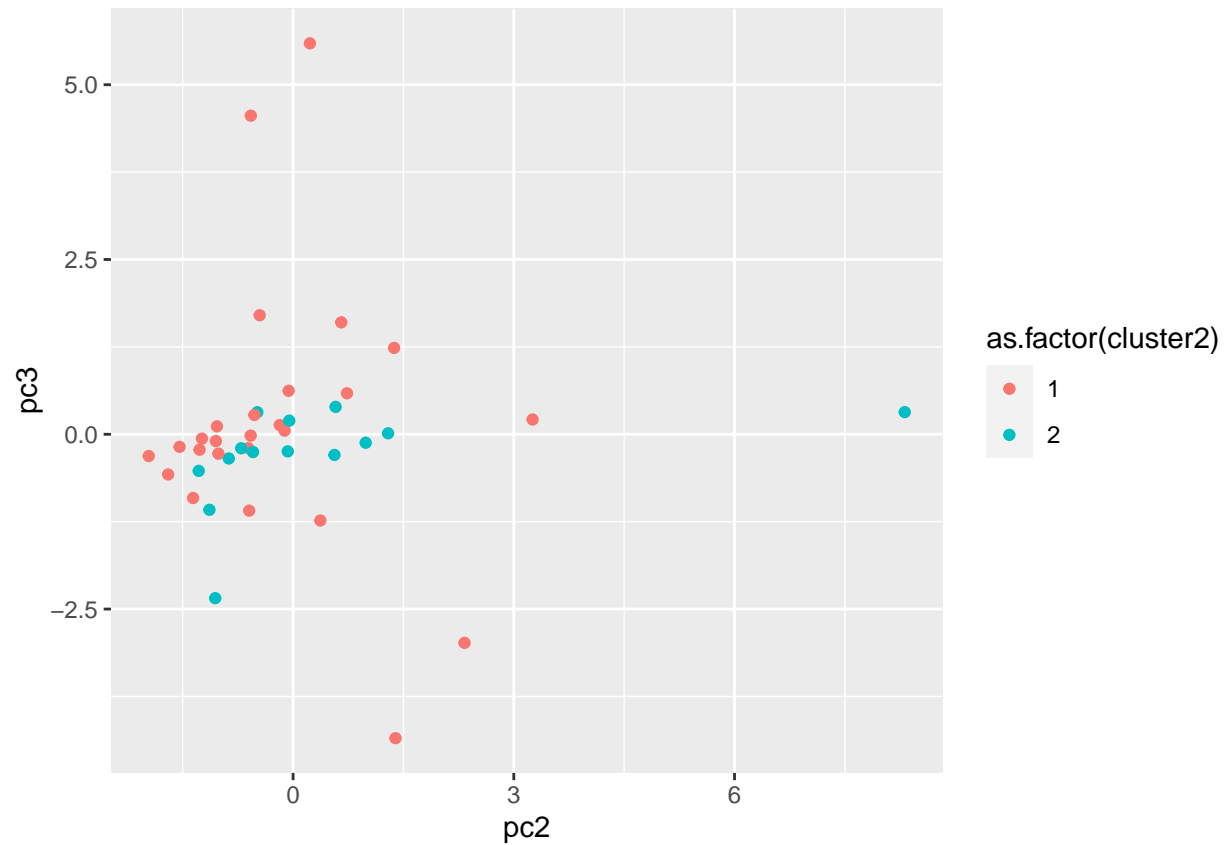
```
d<- dist(wacrime[,129:130])
fit1 <-cmdscale(d, eig=TRUE, k=2)
mdsdf<-data.frame(X=fit1$points[,1], Y= fit1$points[,2], type=as.factor(wacrime$community_name))

ggplot(mdsdf, aes(x=-X, y = Y, color=type))+geom_point(size=1)
```

The legend lists the following cities:

Anacortescity, Auburncity, Bellevuecity, Bellinghamcity, Bothellcity, Bremertoncity, Centraliacity, DesMoinescity, Edmondscity, Ellensburgcity, Everettcity, Kelsocity, Kennewickcity, Kentcity, Kirklandcity, Laceycity, Longviewcity, Lynnwoodcity, MercerIslandcity, MountVernoncity, OakHarborcity, Olympiacity, Pascocity, PortAngelescity, Pullmancity, Puyallupcity, Redmondcity, Rentoncity, Richlandcity, Seattlecity, Spokanecity, Sunnysidecity, Tacomacity, Tukwilacity, Vancouvercity, WallaWallacity, Wenatcheecity, Yakimacity

This is all the principal component analysis. I clustered in two dimensions to see how the data would fit.

From this PCA, we found out which were the biggest variables involving violent crime in WA cities. Honestly, some of the results made sense and some didn't. Population, of course makes sense in relation to the amount of violent crime. The amount of people in a city correlates to how much potential there is for crime. Age(12-29) is a big predictor, and that makes sense. I don't think a lot of people over the age of 30 are committing violent crimes, at least as much as younger people are. The only predictor that didn't really make sense to me was the percentage of black people. This doesn't make sense because it seems like black people throughout WA always have a very small amount of the population, and it seems like the stats across variables are pretty similar to the amount of Hispanics, yet Hispanics aren't as high of components.

So after finding out the biggest variables relating to violent crime, I went back to compare Everett and Bellevue to see if there was a reason why even though Bellevue and Everett have so many police, even though the amount of violent crime is the lowest there. Even though Everett has a low amount of violent crime, they have the highest amount of people aged 12-29. So maybe since that is such a relevant factor, they predict that they need more police? I did this for the percentage of black people as well and it didn't really have a difference in Everett. Tacoma however, had the largest amount of black people and also had the largest rate of violent crime. So maybe there is a correlation there.

```
ggplot(aes(x=agepct12t29,y=community_name, color=community_name),data=wa5)+
  geom_point() +geom_point(size=2)+ggtitle("Avg Income vs How Many Police?")+labs(y= "City", x = "Percer
```

Avg Income vs How Many Police?
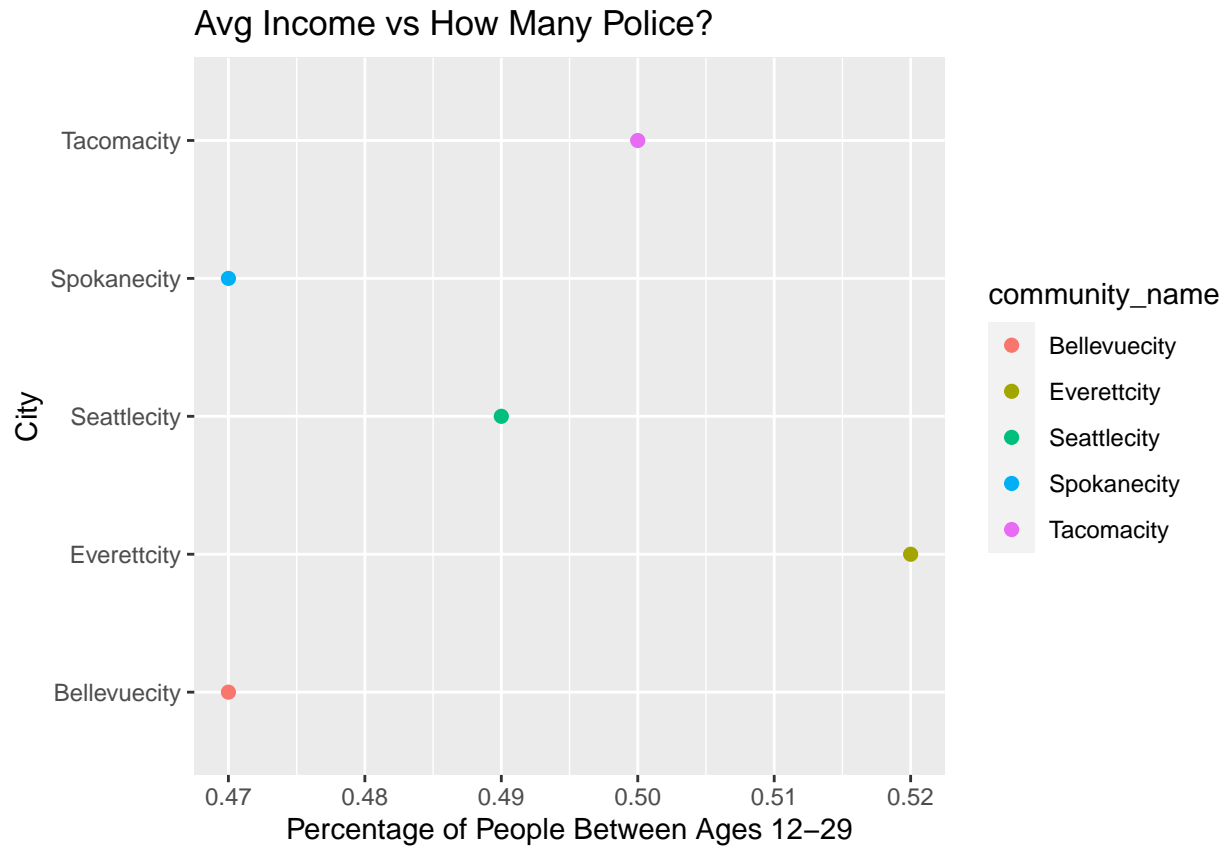

```
ggplot(aes(x=racepctblack,y=community_name, color= community_name),data=wa5)+
  geom_point() +geom_point(size=2)+ggtitle("Percentage of Black People In Each City")+labs(y= "City", x
```

Percentage of Black People In Each City
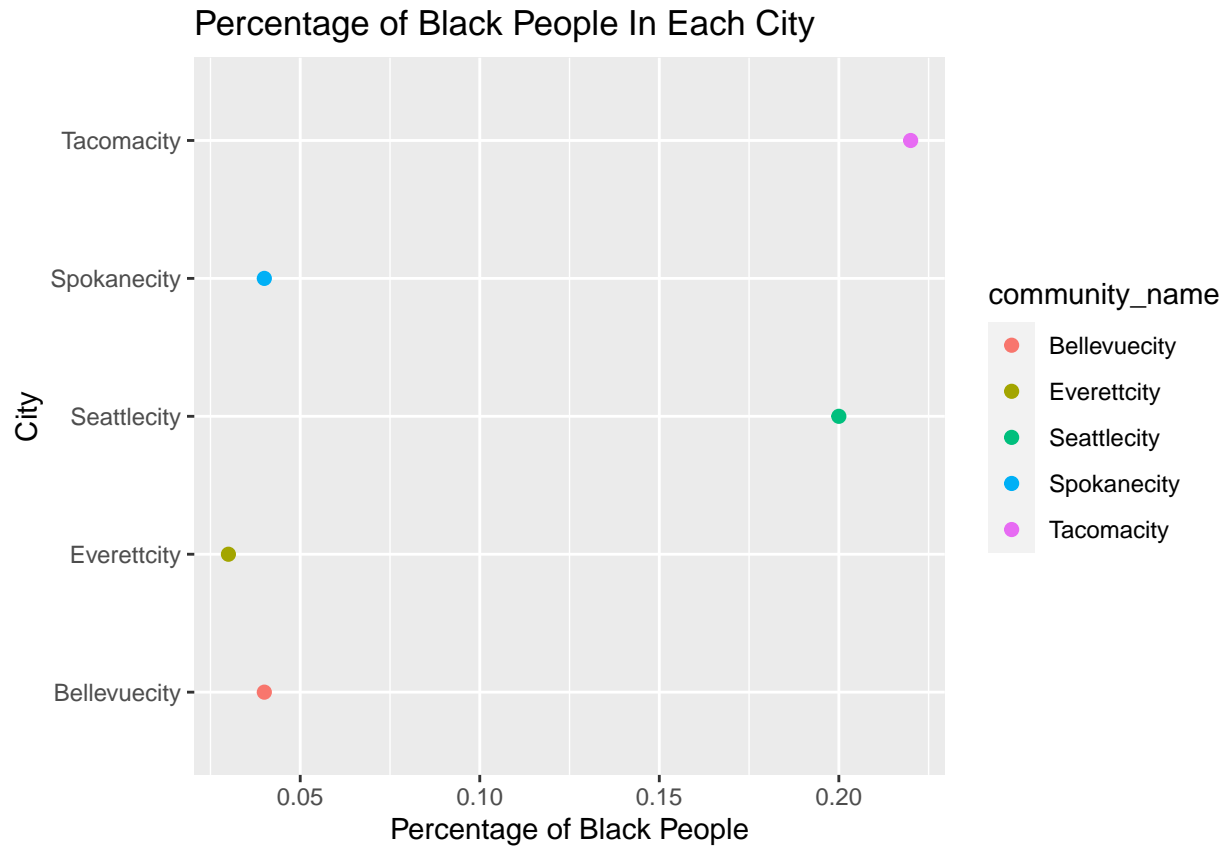
Below, I look at the average median rent and how it relates to violent crime. It shows that the less people are paying for rent(I'm assuming poorer neighborhoods), the higher the crime rate is.

```
ggplot(aes(x=MedRent, y=ViolentCrimesPerPop, color=community_name), data=wacrime)+geom_point() +geom_po
```

Median Rent vs. Violent Crimes

Legend:
Aberdeencity, Anacortescity, Auburncity, Bellevuecity, Bellinghamcity, Bothellcity, Bremertoncity, Centraliacity, DesMoinescity, Edmondscity, Ellensburgcity, Everettcity, K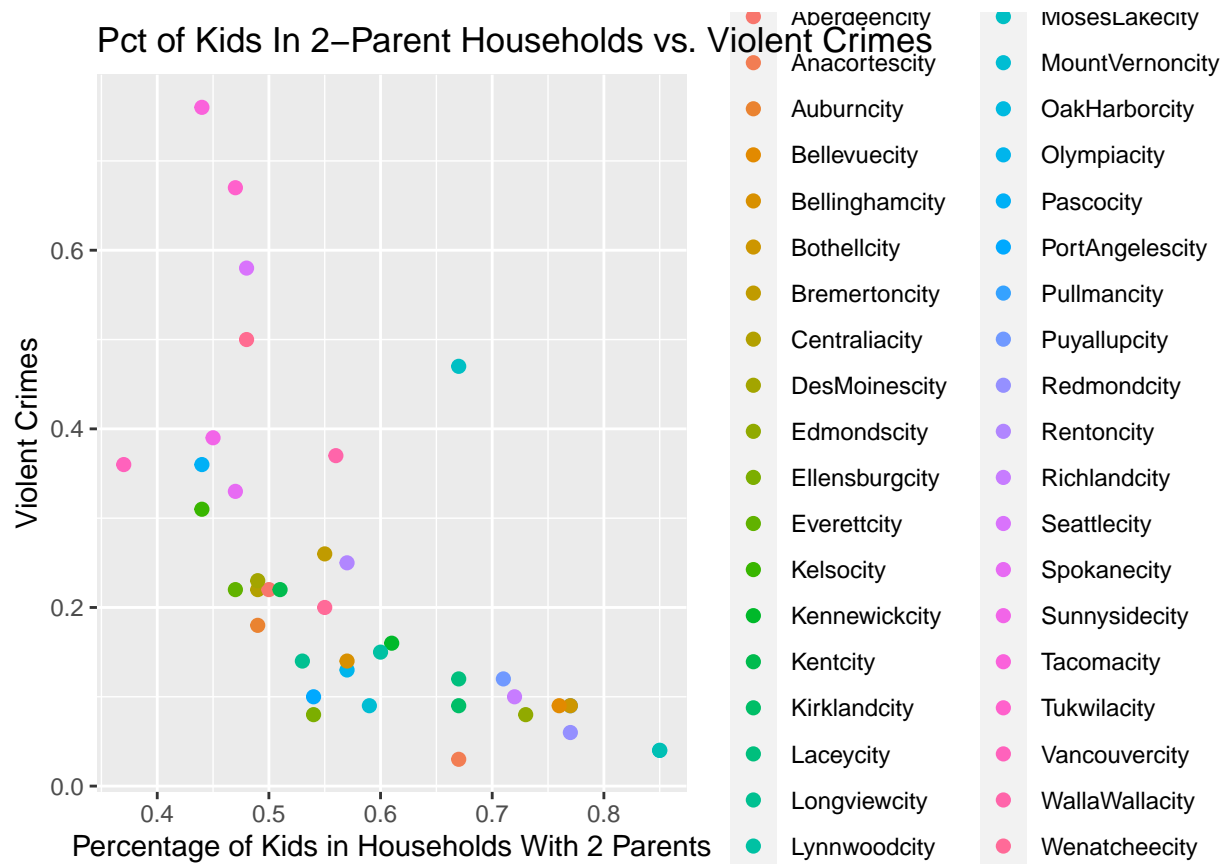elsocity, Kennewickcity, Kentcity, Kirklandcity, Laceycity, Longviewcity, Lynnwoodcity, MosesLakecity, MountVernoncity, OakHarborcity, Olympiacity, Pascocity, PortAngelescity, Pullmancity, Puyallupcity, Redmondcity, Rentoncity, Richlandcity, Seattlecity, Spokanecity, Sunnysidecity, Tacomacity, Tukwilacity, Vancouvercity, WallaWallacity, Wenatcheecity

This was an interesting metric that I was curious about. Does the number of parents in the household (ie not single parent households) make a difference in violent crime? Is there a pattern there? I feel it definitely shows a correlation between more violent crime and single parent homes. The cities with the highest percentage of 2-parent homes have a very low amount of violent crime.

```
ggplot(aes(x=PctKids2Par, y=ViolentCrimesPerPop, color=community_name), data=wacrime)+geom_point() +geom
```

Pct of Kids In 2–Parent Households vs. Violent Crimes

Violent Crimes (y-axis)
Percentage of Kids in Households With 2 Parents (x-axis)

Legend:
- Aberdeencity
- Anacortescity
- Auburncity
- Bellevuecity
- Bellinghamcity
- Bothellcity
- Bremertoncity
- Centraliacity
- DesMoinescity
- Edmondscity
- Ellensburgcity
- Everettcity
- Kelsocity
- Kennewickcity
- Kentcity
- Kirklandcity
- Laceycity
- Longviewcity
- Lynnwoodcity
- MosesLakecity
- MountVernoncity
- OakHarborcity
- Olympiacity
- Pascocity
- PortAngelescity
- Pullmancity
- Puyallupcity
- Redmondcity
- Rentoncity
- Richlandcity
- Seattlecity
- Spokanecity
- Sunnysidecity
- Tacomacity
- Tukwilacity
- Vancouvercity
- WallaWallacity
- Wenatcheecity

These are some plots I did just for fun because I was curious about the results. Evaluating ages with median household income, separated by city. Interestingly enough the lowest and the highest income both had the smallest amount of people aged 12-29.

```
ggplot(wa5, aes(x = agepct12t29, y=medincome, color= community_name)) + geom_point(size=2) +geom_point(
```

Median Income vs Pct of Ppl Age 12–29
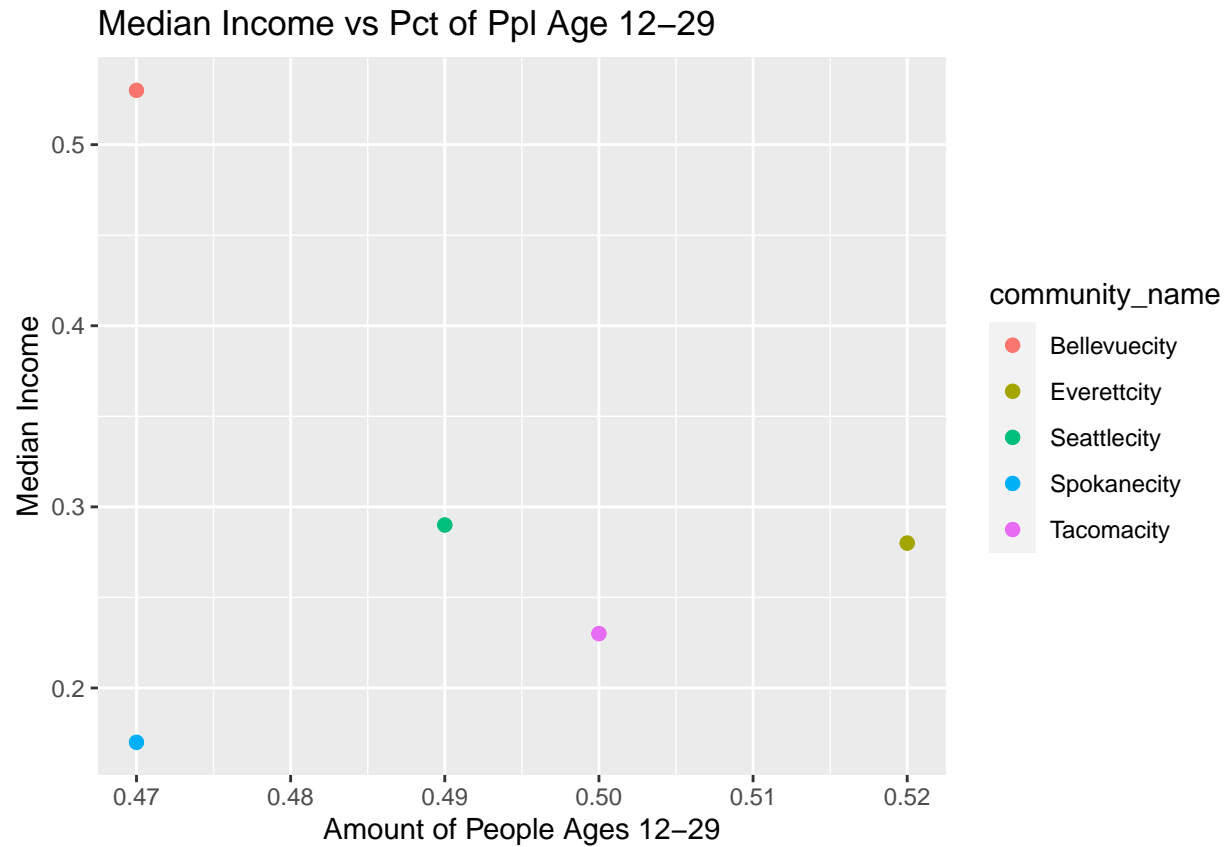
Here I looked at divorce and whether or not the amount of divorced people in a population has any relation to violent crime. It seems that the higher the amount of divorced people in a city, the higher the violent crime rate.

```
ggplot(wacrime, aes(x = TotalPctDiv, y=ViolentCrimesPerPop, color= community_name)) + geom_point(size=2)
```

Divorced vs Violent Crime