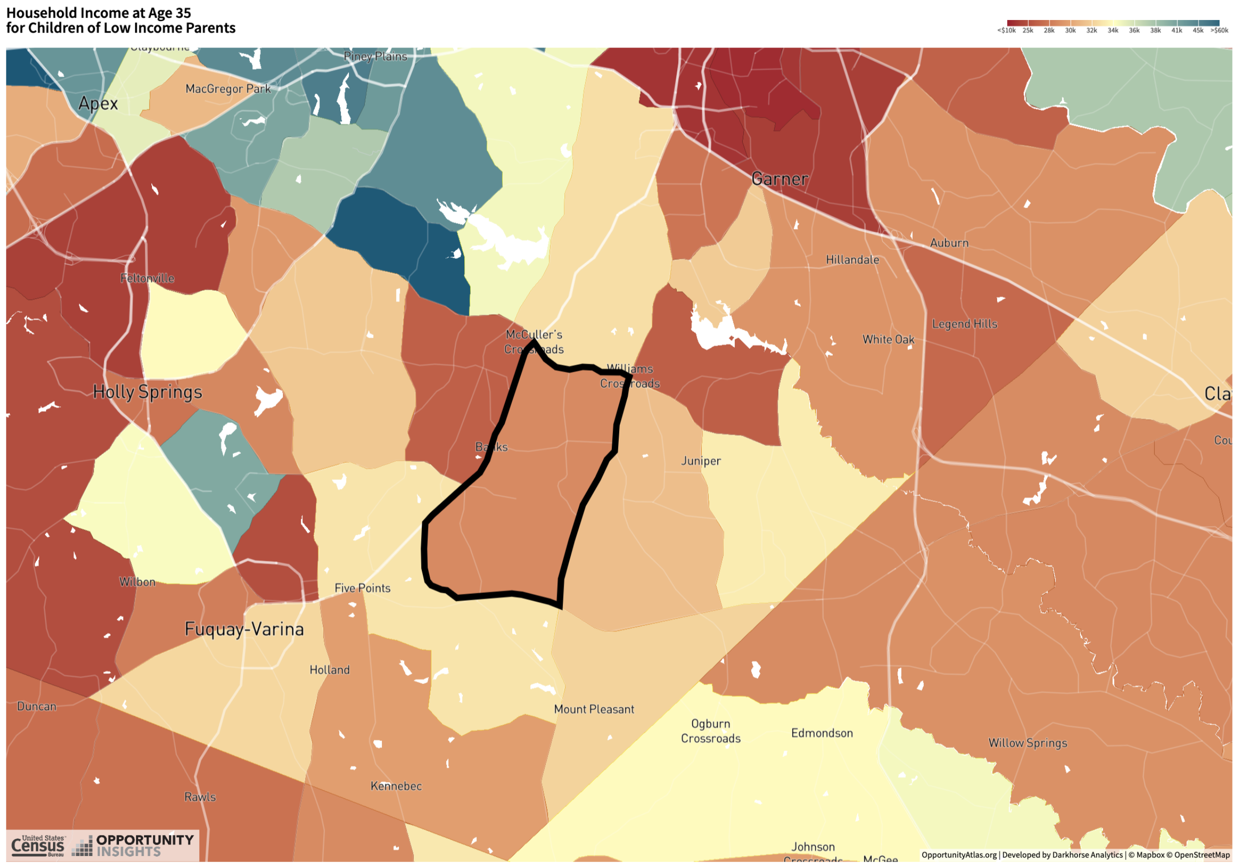
Project_Pt1

2023-02-25

Question 1 My name is Jose and I am from Willow Spring, North Carolina. My community is on the southern edge of Wake County. The map below shows adult household income for children born to low income parents. Compared to similar children nationwide, my neighborhood has below average outcomes.




```r
#Question 2: Summary statistics of all variables, including recording all missing values.
summary(atlas)
```

```
##      state           county           tract          tract_name
##  Min.   : 1.00   Min.   :  1.00   Min.   :   100   Length:73199
##  1st Qu.:13.00   1st Qu.: 29.00   1st Qu.: 10300   Class :character
##  Median :28.00   Median : 63.00   Median : 44702   Mode  :character
##  Mean   :28.18   Mean   : 85.95   Mean   :255238
##  3rd Qu.:42.00   3rd Qu.:109.00   3rd Qu.:458650
##  Max.   :72.00   Max.   :840.00   Max.   :989200
##
##        cz           czname          kfr_pooled_pooled_p25 kfr_natam_pooled_p25
##  Min.   :  100   Length:73199       Min.   : -3.286       Min.   : -9.99
```

```
##  1st Qu.:11600   Class :character   1st Qu.: 38.070      1st Qu.: 27.80
##  Median :19902   Mode  :character   Median : 42.520      Median : 33.41
##  Mean   :21974                      Mean   : 42.858      Mean   : 34.45
##  3rd Qu.:32000                      3rd Qu.: 47.350      3rd Qu.: 40.08
##  Max.   :90000                      Max.   :103.349      Max.   :107.68
##                                     NA's   :1189         NA's   :71464
##  kfr_asian_pooled_p25 kfr_black_pooled_p25 kfr_hisp_pooled_p25
##  Min.   :-13.32       Min.   :-48.47       Min.   :-23.97
##  1st Qu.: 51.50       1st Qu.: 29.97       1st Qu.: 39.11
##  Median : 57.96       Median : 33.03       Median : 43.27
##  Mean   : 57.98       Mean   : 33.99       Mean   : 43.70
##  3rd Qu.: 64.32       3rd Qu.: 37.01       3rd Qu.: 47.65
##  Max.   :113.61       Max.   : 99.52       Max.   :125.98
##  NA's   :57759        NA's   :39111        NA's   :35581
##  kfr_white_pooled_p25 kir_pooled_female_p25 kir_pooled_male_p25
##  Min.   :-13.92       Min.   :-42.61        Min.   :-21.42
##  1st Qu.: 41.43       1st Qu.: 37.38        1st Qu.: 42.47
##  Median : 45.74       Median : 40.98        Median : 46.81
##  Mean   : 46.30       Mean   : 41.86        Mean   : 47.10
##  3rd Qu.: 50.72       3rd Qu.: 45.56        3rd Qu.: 51.44
##  Max.   :103.18       Max.   : 87.90        Max.   :105.74
##  NA's   :5221         NA's   :1554          NA's   :1512
##  kir_natam_female_p25 kir_asian_female_p25 kir_black_female_p25
##  Min.   :17.21        Min.   :-43.99       Min.   : 4.04
##  1st Qu.:30.80        1st Qu.: 51.04       1st Qu.:38.09
##  Median :34.69        Median : 57.60       Median :41.35
##  Mean   :35.12        Mean   : 57.49       Mean   :42.02
##  3rd Qu.:38.91        3rd Qu.: 63.87       3rd Qu.:45.32
##  Max.   :76.50        Max.   :166.61       Max.   :89.37
##  NA's   :72350        NA's   :65467        NA's   :47861
##  kir_hisp_female_p25 kir_white_female_p25 kir_natam_male_p25 kir_asian_male_p25
##  Min.   :-15.03      Min.   :-74.19       Min.   :15.66      Min.   :-124.87
##  1st Qu.: 38.64      1st Qu.: 36.45       1st Qu.:32.85      1st Qu.:  52.81
##  Median : 42.74      Median : 40.83       Median :38.54      Median :  58.69
##  Mean   : 43.12      Mean   : 41.83       Mean   :39.21      Mean   :  58.86
##  3rd Qu.: 47.20      3rd Qu.: 46.19       3rd Qu.:44.97      3rd Qu.:  65.02
##  Max.   : 92.85      Max.   :109.12       Max.   :71.51      Max.   : 114.93
##  NA's   :47319       NA's   :7983         NA's   :72347      NA's   :65112
##  kir_black_male_p25 kir_hisp_male_p25 kir_white_male_p25 jail_pooled_pooled_p25
##  Min.   :-20.40     Min.   :-24.03    Min.   :-100.99    Min.   :-0.2145
##  1st Qu.: 35.17     1st Qu.: 45.23    1st Qu.:  44.82    1st Qu.: 0.0070
##  Median : 39.05     Median : 49.62    Median :  49.38    Median : 0.0180
##  Mean   : 39.57     Mean   : 49.80    Mean   :  49.84    Mean   : 0.0216
##  3rd Qu.: 43.42     3rd Qu.: 54.04    3rd Qu.:  54.48    3rd Qu.: 0.0325
##  Max.   : 85.17     Max.   :119.73    Max.   : 111.89    Max.   : 0.6909
##  NA's   :48140      NA's   :47796     NA's   :7683       NA's   :1322
##  jail_natam_pooled_p25 jail_asian_pooled_p25 jail_black_pooled_p25
##  Min.   :-0.11         Min.   :-0.82         Min.   :-0.55
##  1st Qu.: 0.00         1st Qu.:-0.02         1st Qu.: 0.02
##  Median : 0.03         Median : 0.00         Median : 0.05
##  Mean   : 0.04         Mean   : 0.00         Mean   : 0.05
##  3rd Qu.: 0.06         3rd Qu.: 0.02         3rd Qu.: 0.08
##  Max.   : 0.43         Max.   : 0.79         Max.   : 0.56
##  NA's   :71790         NA's   :59738         NA's   :42014
```

```
##  jail_hisp_pooled_p25 jail_white_pooled_p25 jail_pooled_female_p25
##  Min.   :-0.40        Min.   :-0.636        Min.   :-0.5948
##  1st Qu.: 0.00        1st Qu.: 0.002        1st Qu.:-0.0036
##  Median : 0.01        Median : 0.013        Median : 0.0029
##  Mean   : 0.02        Mean   : 0.017        Mean   : 0.0038
##  3rd Qu.: 0.03        3rd Qu.: 0.027        3rd Qu.: 0.0105
##  Max.   : 0.83        Max.   : 0.820        Max.   : 0.3893
##  NA's   :38468        NA's   :5816          NA's   :1723
##  jail_pooled_male_p25 jail_natam_female_p25 jail_asian_female_p25
##  Min.   :-0.5659      Min.   :-0.11         Min.   :-0.46
##  1st Qu.: 0.0123      1st Qu.:-0.01         1st Qu.:-0.02
##  Median : 0.0333      Median : 0.00         Median : 0.00
##  Mean   : 0.0409      Mean   : 0.01         Mean   : 0.00
##  3rd Qu.: 0.0618      3rd Qu.: 0.02         3rd Qu.: 0.02
##  Max.   : 1.1290      Max.   : 0.16         Max.   : 0.34
##  NA's   :1725         NA's   :72505         NA's   :66571
##  jail_black_female_p25 jail_hisp_female_p25 jail_white_female_p25
##  Min.   :-0.16         Min.   :-0.19        Min.   :-1.300
##  1st Qu.:-0.01         1st Qu.:-0.01        1st Qu.:-0.008
##  Median : 0.00         Median : 0.00        Median : 0.003
##  Mean   : 0.01         Mean   : 0.00        Mean   : 0.004
##  3rd Qu.: 0.02         3rd Qu.: 0.01        3rd Qu.: 0.015
##  Max.   : 0.32         Max.   : 0.31        Max.   : 1.355
##  NA's   :49929         NA's   :49276        NA's   :8717
##  jail_natam_male_p25 jail_asian_male_p25 jail_black_male_p25 jail_hisp_male_p25
##  Min.   :-0.06       Min.   :-2.04       Min.   :-0.45       Min.   :-1.27
##  1st Qu.: 0.01       1st Qu.:-0.02       1st Qu.: 0.06       1st Qu.: 0.00
##  Median : 0.05       Median : 0.00       Median : 0.10       Median : 0.02
##  Mean   : 0.06       Mean   : 0.01       Mean   : 0.11       Mean   : 0.03
##  3rd Qu.: 0.10       3rd Qu.: 0.03       3rd Qu.: 0.15       3rd Qu.: 0.05
##  Max.   : 0.42       Max.   : 1.77       Max.   : 0.71       Max.   : 0.84
##  NA's   :72572       NA's   :66368       NA's   :51527       NA's   :50725
##  jail_white_male_p25     HOLC_A          HOLC_B          HOLC_C
##  Min.   :-1.044      Min.   :0.00    Min.   :0.00    Min.   :0.00
##  1st Qu.: 0.004      1st Qu.:0.00    1st Qu.:0.00    1st Qu.:0.00
##  Median : 0.023      Median :0.00    Median :0.00    Median :0.17
##  Mean   : 0.029      Mean   :0.03    Mean   :0.13    Mean   :0.33
##  3rd Qu.: 0.048      3rd Qu.:0.00    3rd Qu.:0.11    3rd Qu.:0.62
##  Max.   : 1.212      Max.   :1.00    Max.   :1.00    Max.   :1.00
##  NA's   :8623        NA's   :63923   NA's   :63923   NA's   :63923
##      HOLC_D        pm25_1982       pm25_1990        pm25_2000
##  Min.   :0.00    Min.   : 1.97   Min.   : 1.501   Min.   : 1.64
##  1st Qu.:0.00    1st Qu.:16.68   1st Qu.:13.494   1st Qu.:10.21
##  Median :0.00    Median :20.85   Median :17.094   Median :12.71
##  Mean   :0.23    Mean   :20.41   Mean   :16.844   Mean   :12.50
##  3rd Qu.:0.38    3rd Qu.:24.28   3rd Qu.:20.106   3rd Qu.:14.79
##  Max.   :1.00    Max.   :35.92   Max.   :32.744   Max.   :25.37
##  NA's   :63923   NA's   :1296    NA's   :1296     NA's   :1296
##    pm25_2010        vegetation       extreme_heat        developed
##  Min.   : 1.531   Min.   :-0.6337   Min.   :-14.2736   Min.   :0.0000
##  1st Qu.: 7.887   1st Qu.:-0.1392   1st Qu.:  0.2996   1st Qu.:0.1233
##  Median : 9.506   Median :-0.0695   Median :  2.6096   Median :0.6734
##  Mean   : 9.286   Mean   :-0.0838   Mean   :  2.6758   Mean   :0.5751
##  3rd Qu.:10.799   3rd Qu.:-0.0119   3rd Qu.:  5.2012   3rd Qu.:0.9983
```

```
##  Max.   :16.958   Max.   : 0.3957   Max.   : 14.6032   Max.   :1.0000  
##  NA's   :1296     NA's   :2531      NA's   :2531       NA's   :2531    
##  hhinc_mean2000   mean_commutetime2000 frac_coll_plus2000 frac_coll_plus2010
##  Min.   :  7240   Min.   : 2.50        Min.   :0.0000     Min.   :0.0000    
##  1st Qu.: 57377   1st Qu.:22.04        1st Qu.:0.1108     1st Qu.:0.1310    
##  Median : 71721   Median :26.17        Median :0.1870     Median :0.2186    
##  Mean   : 80335   Mean   :26.95        Mean   :0.2378     Mean   :0.2694    
##  3rd Qu.: 94675   3rd Qu.:31.05        3rd Qu.:0.3254     3rd Qu.:0.3709    
##  Max.   :330042   Max.   :80.03        Max.   :1.0000     Max.   :1.0000    
##  NA's   :893      NA's   :882          NA's   :852        NA's   :202       
##  foreign_share2010 med_hhinc1990    med_hhinc2016    popdensity2000    
##  Min.   :0.0000    Min.   :  4999   Min.   :  3250   Min.   :     0.00  
##  1st Qu.:0.0230    1st Qu.: 22261   1st Qu.: 38809   1st Qu.:    97.78  
##  Median :0.0671    Median : 29810   Median : 52333   Median :   760.47  
##  Mean   :0.1209    Mean   : 32179   Mean   : 58898   Mean   :  1987.44  
##  3rd Qu.:0.1689    3rd Qu.: 39409   3rd Qu.: 71935   3rd Qu.:  2005.30  
##  Max.   :1.0000    Max.   :150001   Max.   :250001   Max.   :205382.19  
##  NA's   :916       NA's   :882      NA's   :436      NA's   :726        
##  poor_share2010    poor_share2000   poor_share1990   share_white2010 
##  Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000  
##  1st Qu.:0.05911   1st Qu.:0.0502   1st Qu.:0.0488   1st Qu.:0.4156  
##  Median :0.11581   Median :0.0956   Median :0.0955   Median :0.7355  
##  Mean   :0.15092   Mean   :0.1284   Mean   :0.1322   Mean   :0.6327  
##  3rd Qu.:0.20493   3rd Qu.:0.1715   3rd Qu.:0.1752   3rd Qu.:0.8922  
##  Max.   :1.00000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000  
##  NA's   :262       NA's   :880      NA's   :872      NA's   :84      
##  share_black2010   share_hisp2010    share_asian2010  share_black2000 
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :0.0000  
##  1st Qu.:0.01336   1st Qu.:0.02443   1st Qu.:0.0039   1st Qu.:0.0090  
##  Median :0.04345   Median :0.06308   Median :0.0117   Median :0.0328  
##  Mean   :0.14089   Mean   :0.16099   Mean   :0.0377   Mean   :0.1316  
##  3rd Qu.:0.15185   3rd Qu.:0.18790   3rd Qu.:0.0350   3rd Qu.:0.1297  
##  Max.   :1.00000   Max.   :1.00000   Max.   :0.8918   Max.   :1.0000  
##  NA's   :84        NA's   :84        NA's   :1250     NA's   :827     
##  share_white2000  share_hisp2000   share_asian2000  gsmn_math_g3_2013
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :-2.706   
##  1st Qu.:0.5331   1st Qu.:0.0132   1st Qu.:0.0031   1st Qu.: 2.608   
##  Median :0.8132   Median :0.0361   Median :0.0092   Median : 3.220   
##  Mean   :0.6943   Mean   :0.1182   Mean   :0.0306   Mean   : 3.188   
##  3rd Qu.:0.9302   3rd Qu.:0.1215   3rd Qu.:0.0280   3rd Qu.: 3.761   
##  Max.   :1.0000   Max.   :1.0000   Max.   :0.9000   Max.   : 6.878   
##  NA's   :827      NA's   :827      NA's   :2146     NA's   :1109     
##  rent_twobed2015  singleparent_share2010 singleparent_share2000
##  Min.   :  99.0   Min.   :0.0000         Min.   :0.0000        
##  1st Qu.: 682.0   1st Qu.:0.1851         1st Qu.:0.1825        
##  Median : 853.0   Median :0.2970         Median :0.2588        
##  Mean   : 951.2   Mean   :0.3322         Mean   :0.2935        
##  3rd Qu.:1125.0   3rd Qu.:0.4435         3rd Qu.:0.3663        
##  Max.   :3501.0   Max.   :1.0000         Max.   :1.0000        
##  NA's   :16592    NA's   :631            NA's   :910           
##  singleparent_share1990 traveltime15_2010    emp2000       mail_return_rate2010
##  Min.   :0.0000         Min.   :0.0000    Min.   :0.0000   Min.   :  0.00      
##  1st Qu.:0.1242         1st Qu.:0.1842    1st Qu.:0.5331   1st Qu.: 74.30      
##  Median :0.1869         Median :0.2655    Median :0.6080   Median : 79.70      
```

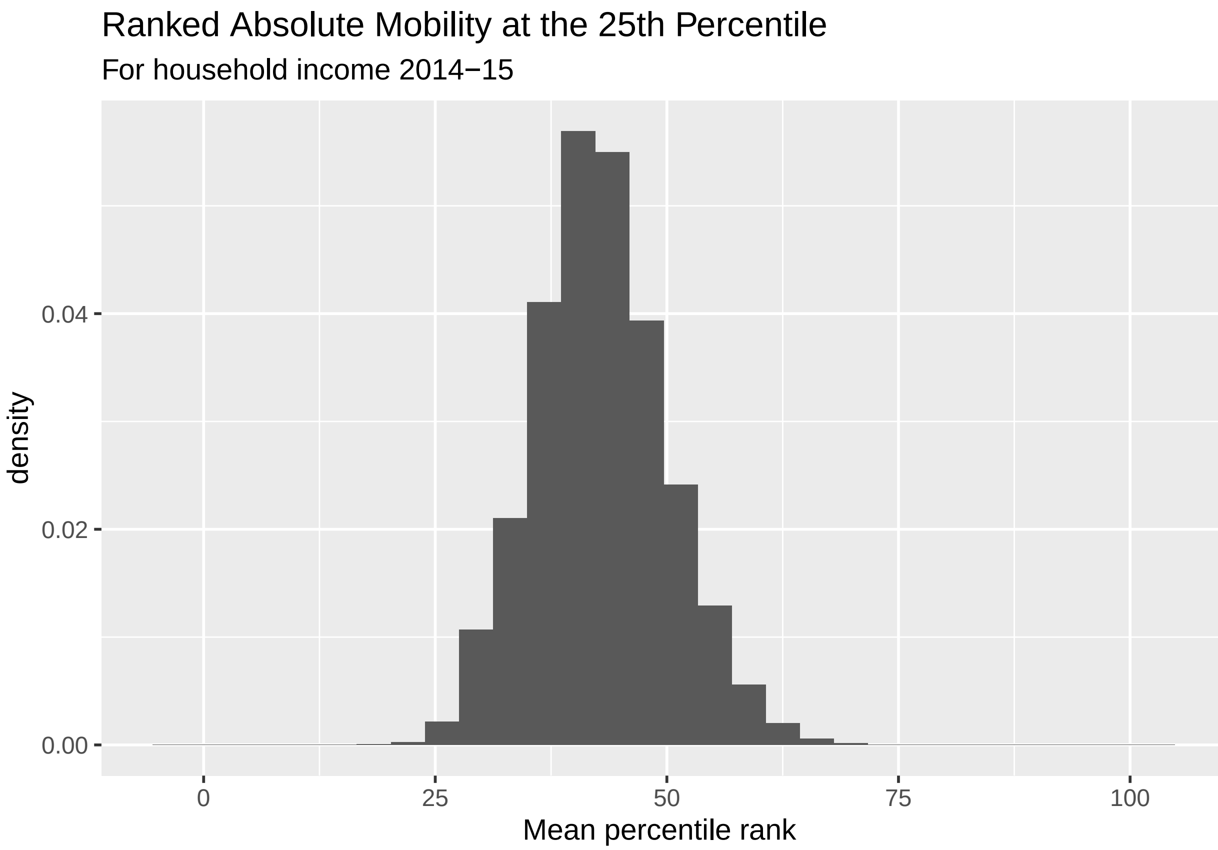
```
##  Mean   :0.2288         Mean   :0.2950    Mean   :0.5945   Mean   : 78.73
##  3rd Qu.:0.2857         3rd Qu.:0.3762    3rd Qu.:0.6709   3rd Qu.: 84.30
##  Max.   :1.0000         Max.   :1.0000    Max.   :1.0000   Max.   :100.00
##  NA's   :999            NA's   :256       NA's   :851      NA's   :652
##  ln_wage_growth_hs_grad jobs_total_5mi_2015 jobs_highpay_5mi_2015
##  Min.   :-3.199         Min.   :      0     Min.   :      0
##  1st Qu.:-0.137         1st Qu.:   6522     1st Qu.:   2304
##  Median : 0.046         Median :  42385     Median :  16981
##  Mean   : 0.042         Mean   : 111996     Mean   :  58566
##  3rd Qu.: 0.224         3rd Qu.: 122916     3rd Qu.:  57828
##  Max.   : 3.075         Max.   :2826437     Max.   :1794186
##  NA's   :21563          NA's   :888         NA's   :888
##  popdensity2010     ann_avg_job_growth_2004_2013 job_density_2013
##  Min.   :     0.0   Min.   :-0.6067              Min.   :      0.0
##  1st Qu.:   319.4   1st Qu.:-0.0189              1st Qu.:     56.7
##  Median :  2193.8   Median : 0.0085              Median :    412.6
##  Mean   :  5236.2   Mean   : 0.0153              Mean   :   2157.0
##  3rd Qu.:  5277.5   3rd Qu.: 0.0410              3rd Qu.:   1371.1
##  Max.   :543333.3   Max.   : 1.3365              Max.   :2905290.2
##  NA's   :5          NA's   :2531                 NA's   :736
```

Most variables have at least some number of missing values. Many of them lack values for almost all tracts, especially the variables that pool data at the household level for different racial groups. This may be because many racial tracts do not have enough people belonging to that racial group to collect enough aggregate data for that given variable.

Question 3 The `kfr_pooled_pooled_p25` variable calculates the mean percentile rank for household income for people born to parents at the 25th percentile of the national distribution of household income. In other words, the higher the value for this variable, the higher absolute mobility. This is calculated with using a linear model to capture the effect that being born to parents at the 25th percentile of household income will have on income outcomes as adults, pooled at the census tract level.

```r
#Question 4: Histogram of Absolute Mobility at the 25th percentile

absmob_histo <- atlas |>
                ggplot() +
                geom_histogram(aes(x = kfr_pooled_pooled_p25, y = after_stat(density))) +
                labs(title = "Ranked Absolute Mobility at the 25th Percentile",
                     subtitle = "For household income 2014-15",
                     x = "Mean percentile rank ")
absmob_histo
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 1189 rows containing non-finite values (`stat_bin()`).
```




```r
ggsave("absmob_histo.png")
```

```
## Saving 6.5 x 4.5 in image
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## Warning: Removed 1189 rows containing non-finite values (`stat_bin()`).
```

The histogram shows an approximately normal distribution of mean percentile rank of absolute mobility at the 25th percentile, slightly left skewed. The majority of pooled census tract absolute mobility ranks are between 25 and 50.

```r
#Question 5: Summary statistics for kfr_pooled_pooled_p25

sumstats <- summary(atlas$kfr_pooled_pooled_p25, na.rm = TRUE)
sdstats <- sd(atlas$kfr_pooled_pooled_p25, na.rm = TRUE)
#creating a table of those values
sumstat <- data.frame(c("Min", "1st Qu.", "Median", "Mean", "3rd Qu.", "Max", "SD", "NAs"),
                       c(-3.286, 38.070, 42.520, 42.858, 47.350, 103.349, 7.126422, 1189))
names(sumstat)[1] <- "Summary Stats"
names(sumstat)[2] <- "Values"
sumstat <- sumstat |>
          mutate(Values = round(Values, 2))
sumstat
```

```
##   Summary Stats  Values
## 1           Min   -3.29
## 2       1st Qu.   38.07
## 3        Median   42.52
## 4          Mean   42.86
```

```
## 5       3rd Qu.   47.35
## 6           Max  103.35
## 7            SD    7.13
## 8           NAs 1189.00
```

Question 6 `kfr_pooled_pooled_p25` can be negative or above 100 in these data because of the limitations of a simple linear model. The model does not know that we are trying to create a standardized percentile rank variable that (logically) starts at 0 and ends at 100. It simply receives a variable (in this case, parents' percentile rank of national household income pooled at the census tract) and uses it to predict the values of a dependent variable (kid percentile rank of household income). Because we are working with a large dataset, it is natural that the model, even when using standardized percentile values, will report values that are below 0 and above 100.

```
#Question 7: comparing absolute mobiltiy at the 25th percentile in my census tract, state, and across t

usabsmob <- mean(atlas$kfr_pooled_pooled_p25, na.rm = TRUE) #mean of absolute mobility nationwide
ncabsmob <- mean(atlas$kfr_pooled_pooled_p25[atlas$state == "37"], na.rm = TRUE) #mean of absolute mobi
myhood <- atlas %>% subset(state == "37" & county == "183" & tract == "53110") #creating data frame of
myhoodabsmob<- mean(myhood$kfr_pooled_pooled_p25, na.rm = TRUE) #mean of absolute mobility in my neighb

abscomp <- data.frame(c("My Neighborhood", "North Carolina", "United States"),
                      c(myhoodabsmob, ncabsmob, usabsmob))
names(abscomp)[1] <- "Level"
names(abscomp)[2] <- "Absolute Mobility at the 25th Percentile"
abscomp
```

```
##             Level Absolute Mobility at the 25th Percentile
## 1 My Neighborhood                                 38.18818
## 2  North Carolina                                 37.89704
## 3   United States                                 42.85813
```

My neighborhood/census tract has a slightly higher level of absolute mobility (at the 25th percentile) compared to my home state of North Carolina. Both my census tract and North Carolina have a lower level of absolute mobility than the national average.

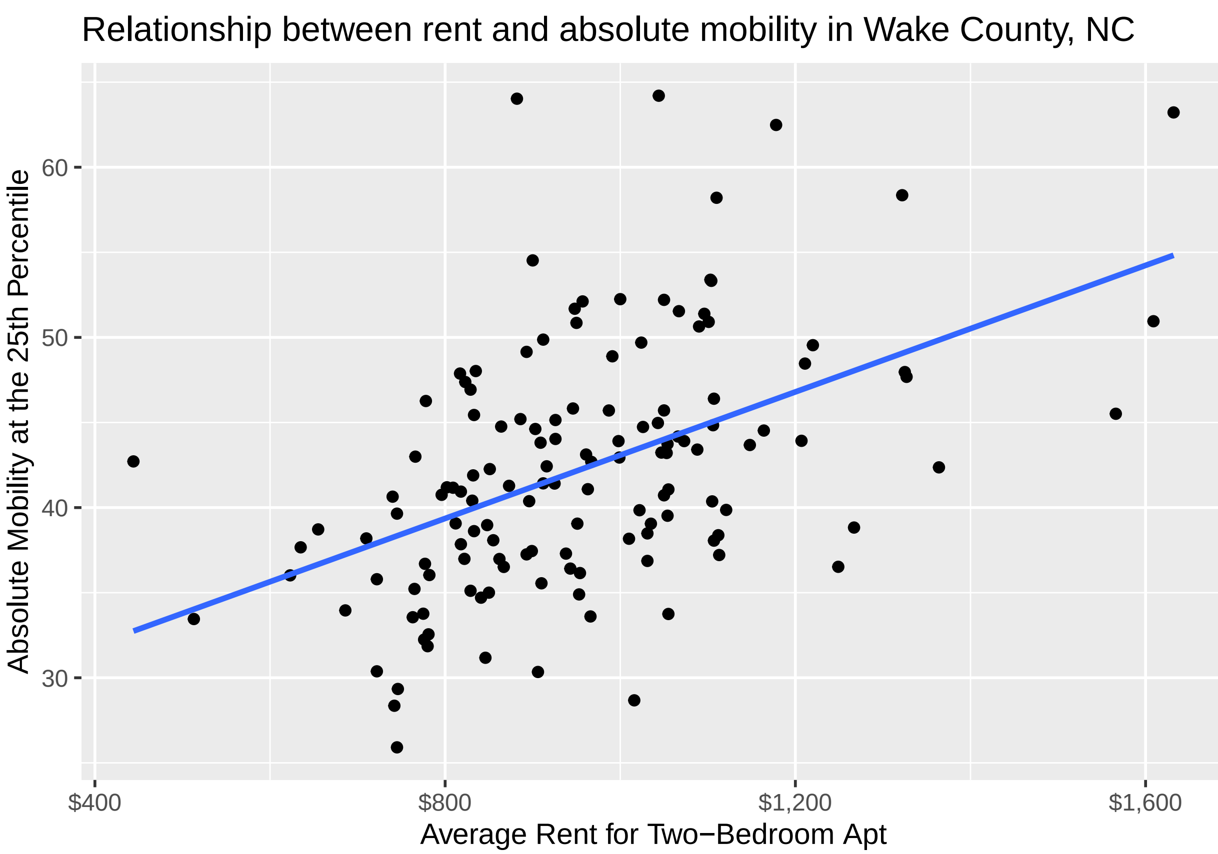
```
#Question 8: calculating and comparing standard deviations of absolute mobility in my home county, stat

usasd <- sd(atlas$kfr_pooled_pooled_p25, na.rm = TRUE) #sd of absolute mobility nationwide
ncsd <- sd(atlas$kfr_pooled_pooled_p25[atlas$state == "37"], na.rm = TRUE) #sd of absolute mobility in
wake_co <- atlas %>% subset(state == "37" & county == "183") #creating data frame of just my tract
countysd<- sd(wake_co$kfr_pooled_pooled_p25, na.rm = TRUE) #sd of absolute mobility in my county

sdcomp <- data.frame(c("Wake County", "North Carolina", "United States"),
                     c(countysd, ncsd, usasd))
names(sdcomp)[1] <- "Level"
names(sdcomp)[2] <- "Std Dev of Absolute Mobility at the 25th Percentile"
sdcomp
```

```
##            Level Std Dev of Absolute Mobility at the 25th Percentile
## 1    Wake County                                            7.692496
## 2 North Carolina                                            5.756876
## 3  United States                                            7.126422
```

The standard deviation of mobility outcomes for my Wake County, NC (7.69) is slightly higher than the standard deviation at the national level. North Carolina's mobility outcome standard deviation of 5.75 is lower than the amount of spread at the county or national level. This gives us a sense of how results may vary, which makes sense because of the increasing sample size for each successive level of analysis.

```r
#Question 9: plotting the relationship between upward mobility and rent in my neighborhood

#9a: Scatter plot of my home county

scatterrent <- atlas |>
           filter(state == 37,
                  county == 183) |>
           ggplot() +
           geom_point(aes(x = rent_twobed2015, y = kfr_pooled_pooled_p25)) +
           geom_smooth(aes(x = rent_twobed2015, y = kfr_pooled_pooled_p25), method = "lm", se = F) +
           labs(x = "Average Rent for Two-Bedroom Apt",
                y = "Absolute Mobility at the 25th Percentile",
                title = "Relationship between rent and absolute mobility in Wake County, NC") +
          scale_x_continuous(labels = scales::dollar_format())
scatterrent
```

```
## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 50 rows containing non-finite values (`stat_smooth()`).

## Warning: Removed 50 rows containing missing values (`geom_point()`).
```




```r
ggsave("scatterrent.png")
```

```
## Saving 6.5 x 4.5 in image
## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 50 rows containing non-finite values (`stat_smooth()`).
## Removed 50 rows containing missing values (`geom_point()`).
```

Question 9B There is an apparent positive correlation between rent and absolute mobility at the 25th percentile. The line of best fit suggests that as rent increases in my home county, the percentile rank of absolute mobility increases as well. This suggests that low income children living in higher-priced tracts experience higher mobility than low income children living in cheaper (and, presumably, more economicallly and socially disadvantaged) tracts.

Question 9C Opportunity bargains are neighborhoods with cheaper than average rents and above average absolute mobility outcomes. For this question, I am defining a given neighborhood an opportunity bargain if it has under $1500 in median two-bedroom rent and the highest percentile rank for absolute mobility at the 25th percentile for my given county.

```
#Question 9c: Defining Opportunity Bargains
myhood <- atlas %>% subset(state == "37" & county == "183" & tract == "53110") #creating data frame of j
myhoodabsmob
```

```
## [1] 38.18818
```

```
myhoodrent <- mean(myhood$rent_twobed2015, na.rm = TRUE) #myneighborhood rent
myhoodrent
```

```
## [1] 710
```

With an absolute mobility rank (at the 25th percentile) of 38 and a median two-bedroom apartment rent of $710, my neighborhood can perhaps be considered an opportunity tract, but we would need to compare it to other tracts in my county to make a more practical judgment. What about other census tracts in my home county of Wake County?

```
#Finding other opportunity bargains in my home county
wake_co <- atlas %>% subset(state == "37" & county == "183") #creating data frame of just my tract
wake_co$bargain_index <- wake_co$kfr_pooled_pooled_p25/wake_co$rent_twobed2015
wake_bargains <- wake_co |>
                 select(tract_name, tract, kfr_pooled_pooled_p25, rent_twobed2015, bargain_index) |>
                 arrange(desc(bargain_index))
print(wake_bargains)
```

```
## # A tibble: 186 x 5
##    tract_name                     tract kfr_pooled_pooled_p25 rent_two~1 barga~2
##    <chr>                          <dbl>                 <dbl>      <dbl>   <dbl>
##  1 Holly Springs, NC              53206                  42.7        444  0.0962
##  2 Brookgreen Forest, Cary, NC    53423                  64.0        882  0.0726
##  3 Wendell, NC                    54403                  33.5        513  0.0652
##  4 Umstead, Raleigh, NC           53717                  64.2       1044  0.0615
##  5 North Raleigh, Raleigh, NC     53715                  54.5        900  0.0606
##  6 Wake Forest, NC                54204                  46.3        778  0.0595
##  7 Wendell, NC                    54402                  37.7        635  0.0593
##  8 Wendell, NC                    54111                  38.7        655  0.0591
##  9 Southwest Raleigh, Raleigh, NC 52404                  47.9        817  0.0586
## 10 Zebulon, NC                    54301                  36.0        623  0.0578
## # ... with 176 more rows, and abbreviated variable names 1: rent_twobed2015,
## #   2: bargain_index
```

Using the opportunity bargain index method derived by dividing the absolute mobility rank by the median two-bedroom rent, we see that my census tract is #21 out of the 186 tracts in my home county. Furthermore, we see that my tract's bargain index variable is influenced by the very cheap median rent in my neighborhood - but cheap rent alone does not make an opportunity bargain. Therefore, I would not classify my neighborhood/census tract as an opportunity bargain. I will go ahead and determine opportunity bargains those census tracts with rents under $1500 and an absolute mobility rank of over 50. Let's see where they are, if any.

```r
#Finding opportunity bargains in Wake County
top_bargains <- wake_bargains |>
                filter(rent_twobed2015 <= 1500,
                       kfr_pooled_pooled_p25 >= 50)
print(top_bargains)
```

```
## # A tibble: 17 x 5
##    tract_name                       tract kfr_pooled_pooled_p25 rent_two~1 barga~2
##    <chr>                            <dbl>                 <dbl>      <dbl>   <dbl>
##  1 Brookgreen Forest, Cary, NC      53423                  64.0        882  0.0726
##  2 Umstead, Raleigh, NC             53717                  64.2       1044  0.0615
##  3 North Raleigh, Raleigh, NC       53715                  54.5        900  0.0606
##  4 Northwest Raleigh, Raleigh, NC 53712                    51.7        948  0.0545
##  5 Dutchess Village, Cary, NC       53524                  52.1        957  0.0545
##  6 North Hills, Raleigh, NC         51501                  50.9        950  0.0535
##  7 Cary, NC                         53603                  62.5       1178  0.0530
##  8 Breckenridge, Raleigh, NC        53725                  58.2       1110  0.0524
##  9 North Raleigh, Raleigh, NC       54011                  52.2       1000  0.0522
## 10 North Raleigh, Raleigh, NC       54016                  52.2       1050  0.0497
## 11 Northwest Raleigh, Raleigh, NC 53724                    53.4       1103  0.0484
## 12 Southwest Raleigh, Raleigh, NC 52301                    51.5       1067  0.0483
## 13 Cary, NC                         53425                  53.3       1104  0.0483
## 14 Lochmere, Cary, NC               53004                  51.4       1096  0.0469
## 15 Hillsdale Forest, Cary, NC       53505                  50.6       1090  0.0465
## 16 Stonebridge, Raleigh, NC         53807                  50.9       1101  0.0462
## 17 Medfield Estates, Raleigh, NC  53521                    58.4       1322  0.0441
## # ... with abbreviated variable names 1: rent_twobed2015, 2: bargain_index
```

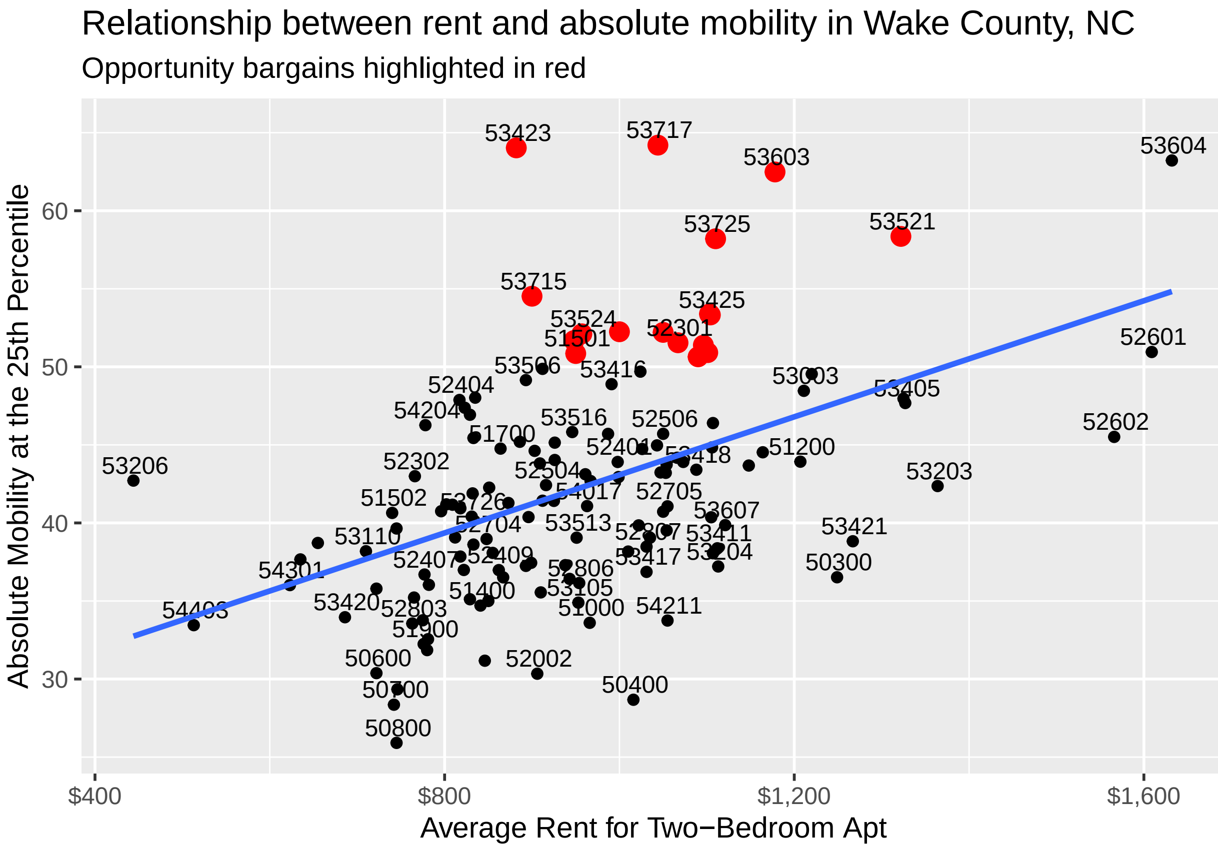
So now we see a list of 17 census tracts - from my experience, most of them clustered in the neighborhoods of upper middle-class North Raleigh and rapidly growing Cary - that we can reasonably say are opportunity bargains. Let's highlight them on the scatterplot of rent and absolute mobility to place them in the context of other census tracts. Do they stand out?

```r
#Highlighting opportunity bargains in Wake County, NC

scatterbargain <- atlas |>
           filter(state == 37,
                  county == 183) |>
           ggplot() +
           geom_point(aes(x = rent_twobed2015, y = kfr_pooled_pooled_p25)) +
           geom_point(data = top_bargains, aes(x = rent_twobed2015, y = kfr_pooled_pooled_p25),
                      color ="red", size = 3) +
           geom_text(aes(x = rent_twobed2015, y = kfr_pooled_pooled_p25, label = tract), check_overlap =
                     size = 3, nudge_x = 2, nudge_y = 1) +
           geom_smooth(aes(x = rent_twobed2015, y = kfr_pooled_pooled_p25), method = "lm", se = F) +
           labs(x = "Average Rent for Two-Bedroom Apt",
                y = "Absolute Mobility at the 25th Percentile",
                title = "Relationship between rent and absolute mobility in Wake County, NC",
                subtitle = "Opportunity bargains highlighted in red") +
          scale_x_continuous(labels = scales::dollar_format())
scatterbargain
```

```
## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 50 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 50 rows containing missing values (`geom_point()`).
```

```
## Warning: Removed 50 rows containing missing values (`geom_text()`).
```




```r
ggsave("scatterbargain.png")
```

```
## Saving 6.5 x 4.5 in image
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 50 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 50 rows containing missing values (`geom_point()`).
```

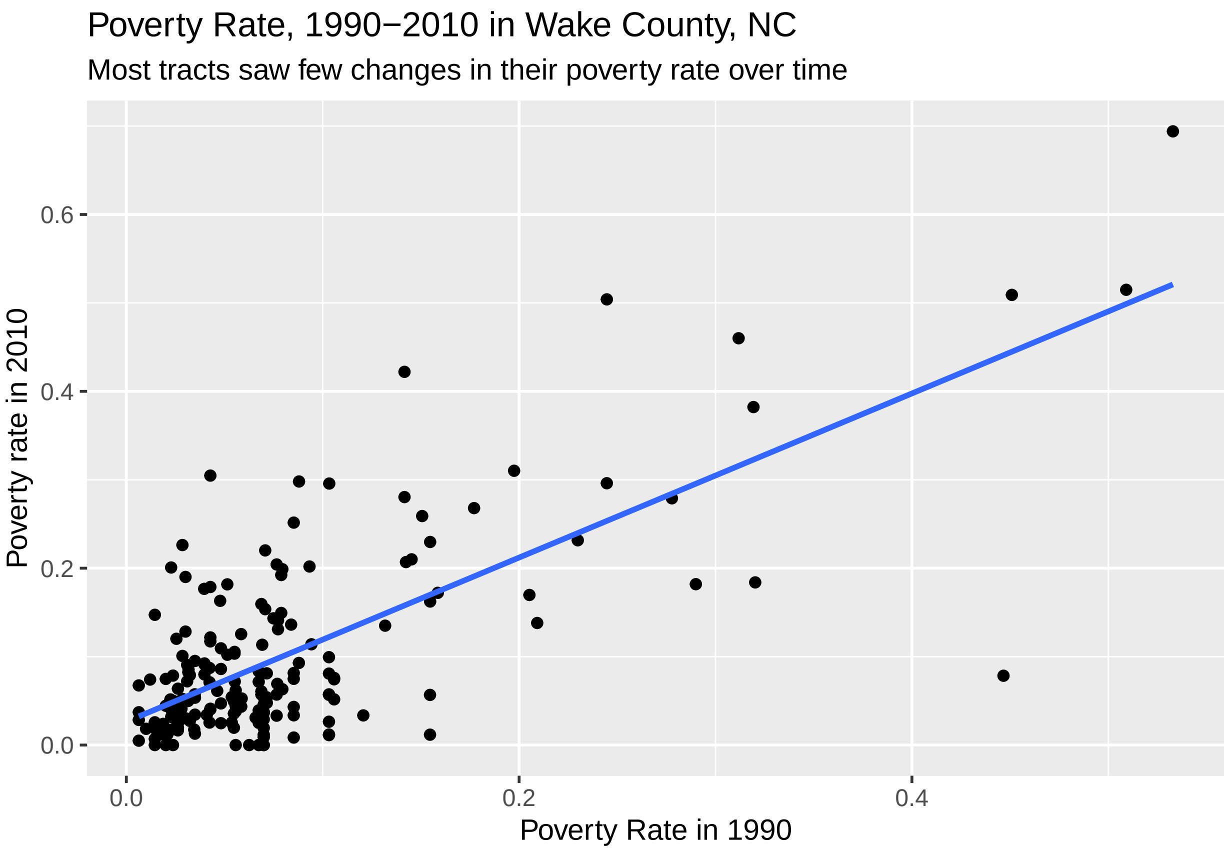
```
## Warning: Removed 50 rows containing missing values (`geom_text()`).
```

And there they are! These are the bargain tracts with cheap rent and above average mobility outcomes in Wake County, NC.

```r
#Question 10: Tracking changes in Wake County over the last 20 years

#10A: scatter plot of poverty rate in 1990 and 2010 for Wake County

wakepoverty <- atlas |>
           filter(state == 37,
                  county == 183) |>
           ggplot(aes(x = poor_share1990, y = poor_share2010)) +
           geom_point(aes(x = poor_share1990, y = poor_share2010)) +
           geom_smooth(method = "lm", se = F) +
           labs(x = "Poverty Rate in 1990",
                y = "Poverty rate in 2010",
                title = "Poverty Rate, 1990-2010 in Wake County, NC",
```

```
                   subtitle = "Most tracts saw few changes in their poverty rate over time")
wakepoverty
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

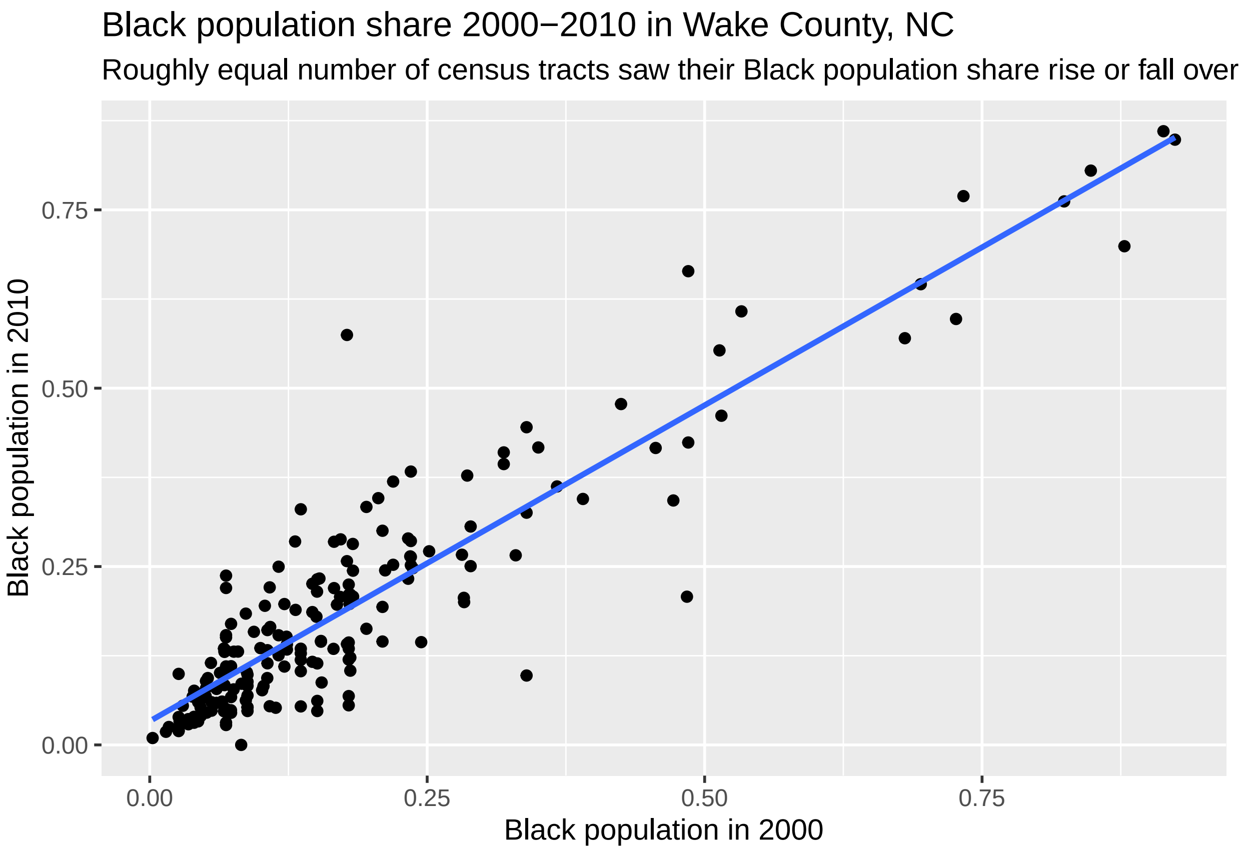



```
ggsave("wakepoverty.png")
```

```
## Saving 6.5 x 4.5 in image
## `geom_smooth()` using formula = 'y ~ x'
```

```
#10A: scatter plot of Black population in 1990 and 2010 for Wake County

wakeblack <- atlas |>
           filter(state == 37,
                  county == 183) |>
           ggplot(aes(x = share_black2000, y = share_black2010)) +
           geom_point() +
           geom_smooth(method = "lm", se = F) +
           labs(x = "Black population in 2000",
                y = "Black population in 2010",
                title = "Black population share 2000-2010 in Wake County, NC",
                subtitle = "Roughly equal number of census tracts saw their Black population share rise
wakeblack
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

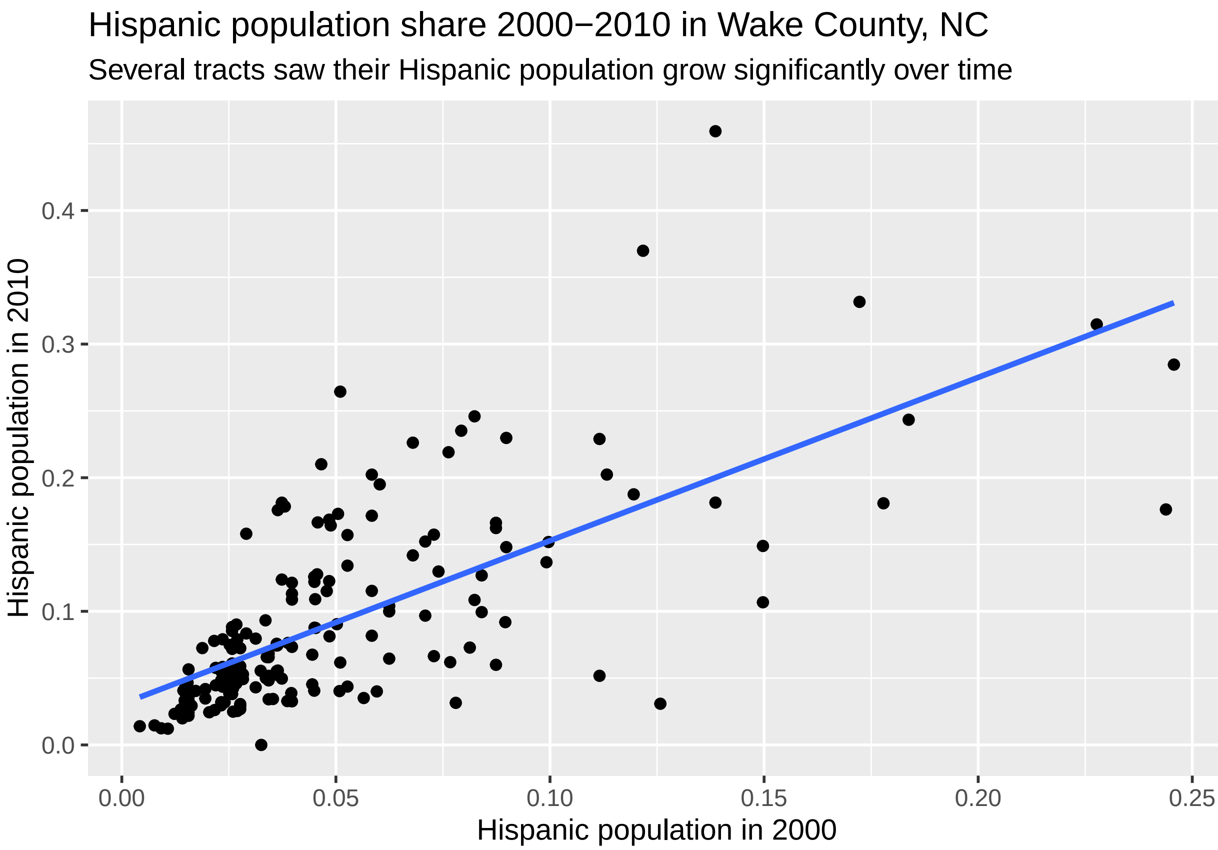



```
ggsave("wakeblack.png")
```

```
## Saving 6.5 x 4.5 in image
## `geom_smooth()` using formula = 'y ~ x'
```

```
#10A: scatter plot of Hispanic population in 1990 and 2010 for Wake County

wakehisp <- atlas |>
           filter(state == 37,
                  county == 183) |>
           ggplot(aes(x = share_hisp2000, y = share_hisp2010)) +
           geom_point() +
           geom_smooth(method = "lm", se = F) +
           labs(x = "Hispanic population in 2000",
                y = "Hispanic population in 2010",
                title = "Hispanic population share 2000-2010 in Wake County, NC",
                subtitle = "Several tracts saw their Hispanic population grow significantly over time")
wakehisp
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
ggsave("wakehisp.png")
```

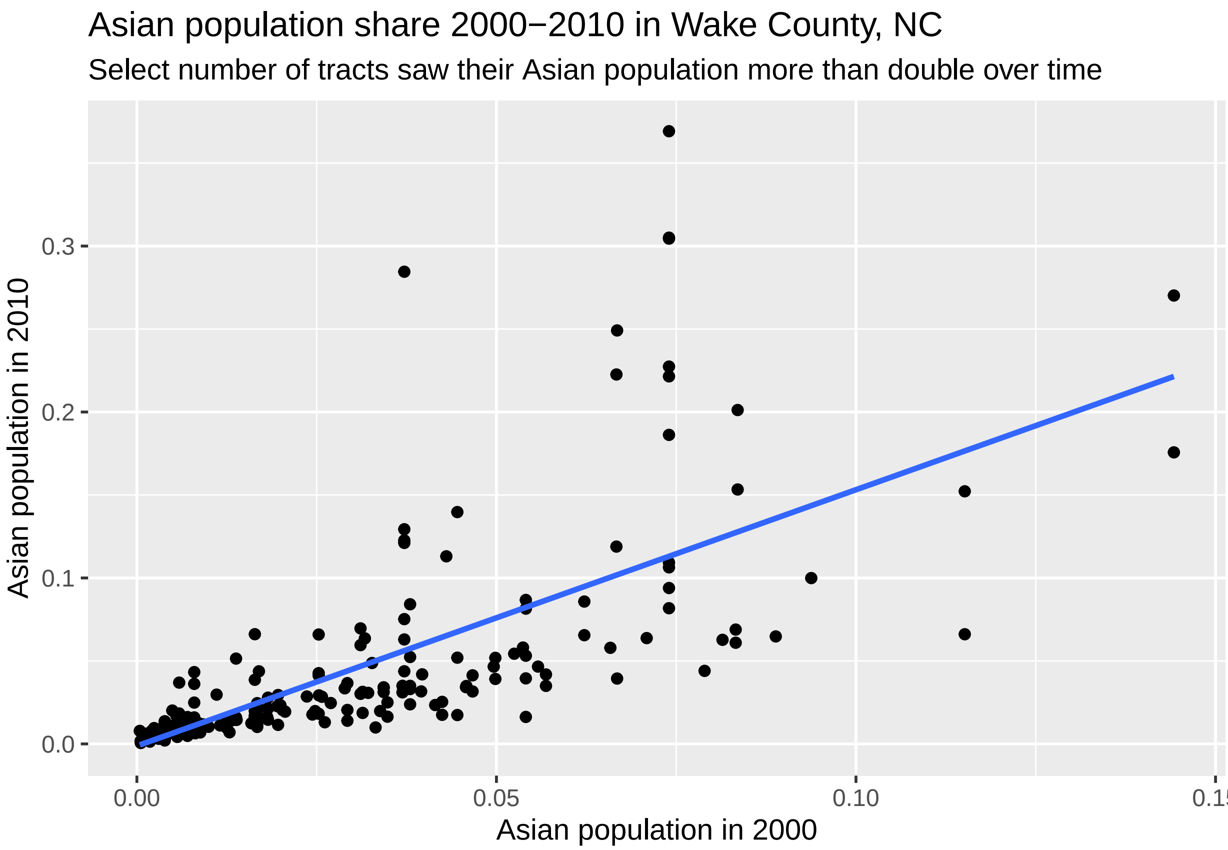
```
## Saving 6.5 x 4.5 in image
## `geom_smooth()` using formula = 'y ~ x'
```

```
#10A: scatter plot of Asian population in 1990 and 2010 for Wake County

wakeasian <- atlas |>
           filter(state == 37,
                  county == 183) |>
           ggplot(aes(x = share_asian2000, y = share_asian2010)) +
           geom_point() +
           geom_smooth(method = "lm", se = F) +
           labs(x = "Asian population in 2000",
                y = "Asian population in 2010",
                title = "Asian population share 2000-2010 in Wake County, NC",
                subtitle = "Select number of tracts saw their Asian population more than double over ti
wakeasian
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```



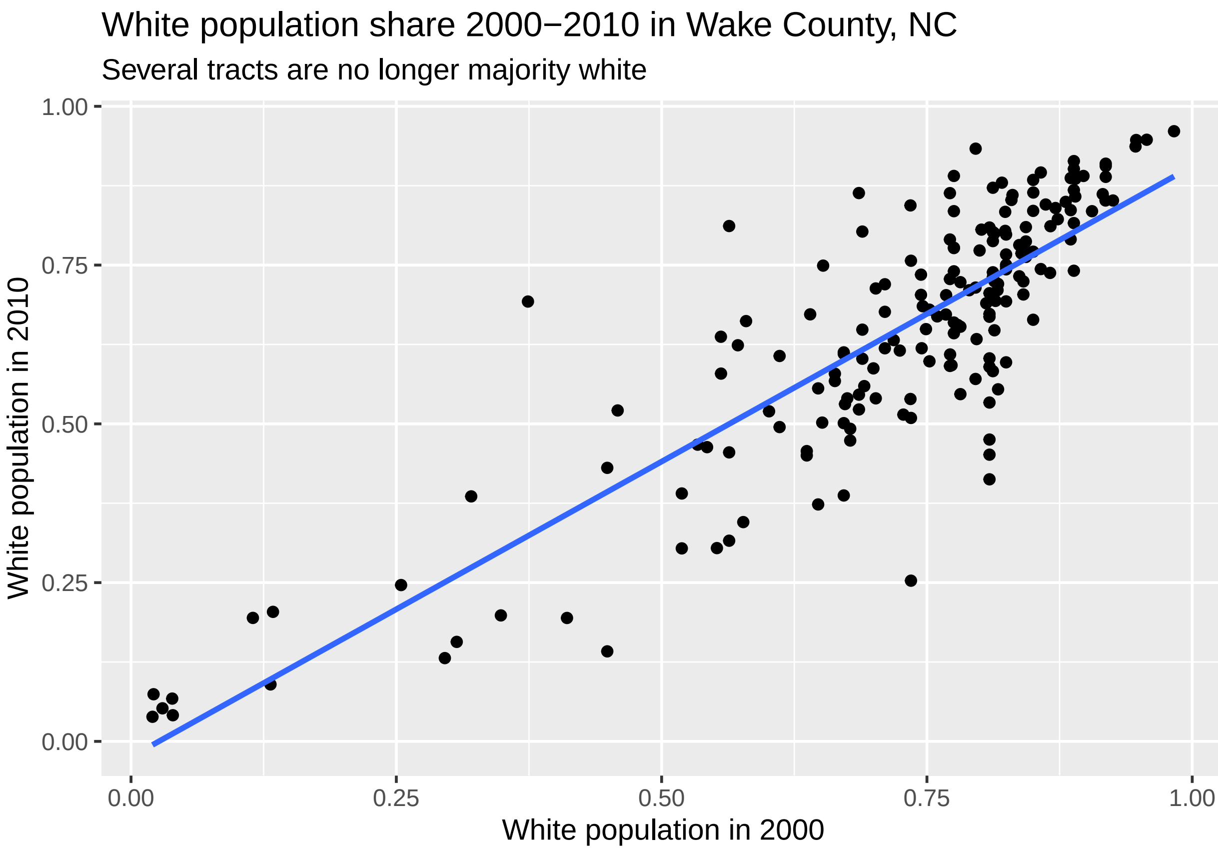

```
ggsave("wakeasian.png")
```

```
## Saving 6.5 x 4.5 in image
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).
## Removed 2 rows containing missing values (`geom_point()`).
```

```
#10A: scatter plot of White population in 1990 and 2010 for Wake County

wakewhite <- atlas |>
           filter(state == 37,
                  county == 183) |>
           ggplot(aes(x = share_white2000, y = share_white2010)) +
           geom_point() +
           geom_smooth(method = "lm", se = F) +
           labs(x = "White population in 2000",
                y = "White population in 2010",
                title = "White population share 2000-2010 in Wake County, NC",
                subtitle = "Several tracts are no longer majority white")
wakewhite
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```r
ggsave("wakewhite.png")
```

```
## Saving 6.5 x 4.5 in image
## `geom_smooth()` using formula = 'y ~ x'
```

Question 10B Between 1990, 2000, and 2010, Wake County saw relatively minor changes to its poverty rate, but significant changes to its demographic composition. Most tracts saw fairly little change in their poverty rate between 1990 and 2010. However, several tracts saw significant increases in their share of Asian and Hispanic populations between 2000 and 2010. At the same time, many tracts saw their share of the White population fall under 50%. This highlights the growing share of nonwhite populations in Wake County over the last few decades.

```r
#Question 11a: Average absolute mobility rate for neighborhoods grouped by HOLC grade (A, B, C, and D)

holcmobil <- atlas |>
             mutate(grade = case_when(HOLC_A > 0.5 & !is.na(HOLC_A) ~ 'A',
                                      HOLC_B > 0.5 & !is.na(HOLC_B) ~ 'B',
                                      HOLC_C > 0.5 & !is.na(HOLC_C) ~ 'C',
                                      HOLC_D > 0.5 & !is.na(HOLC_D) ~ 'D')) |>
            group_by(grade) |>
            summarize(mean_absolute_mobility = mean(kfr_pooled_pooled_p25, na.rm = TRUE))

holcmobil
```

```
## # A tibble: 5 x 2
##   grade mean_absolute_mobility
##   <chr>                  <dbl>
## 1 A                       44.0
## 2 B                       42.5
```

```
## 3 C                          39.9
## 4 D                          36.2
## 5 <NA>                       43.2
```

There is a clear pattern of increasing upward mobility for children born in the 1980s in Census tracts with higher 1930s HOLC grades. For tracts where a majority of its area was graded A, the average mobility rank at the 25th percentile is roughly 44. As we start to look at tracts with majority of their areas graded B through D, we see a consistent pattern of a decrease of 2-3 mobility percentile ranks. This suggests that practices such as redlining may have had an adverse effect on mobility outcomes in various neighborhoods that persists to this day.

```
#Q 11b: Black population share in 2000 for neighborhoods grouped by HOLC grade

holcblack <- atlas |>
             mutate(grade = case_when(HOLC_A > 0.5 & !is.na(HOLC_A) ~ 'A',
                                      HOLC_B > 0.5 & !is.na(HOLC_B) ~ 'B',
                                      HOLC_C > 0.5 & !is.na(HOLC_C) ~ 'C',
                                      HOLC_D > 0.5 & !is.na(HOLC_D) ~ 'D')) |>
            group_by(grade) |>
            summarize(mean_blackpop = mean(share_black2000, na.rm = TRUE))

holcblack
```

```
## # A tibble: 5 x 2
##   grade mean_blackpop
##   <chr>         <dbl>
## 1 A             0.201
## 2 B             0.289
## 3 C             0.333
## 4 D             0.466
## 5 <NA>          0.111
```

It is clear that when looking at the racial composition of HOLC graded neighborhoods, we see that neighborhoods graded A have less than half of Black population share as we see in D rated neighborhoods. There is a clear pattern of increasing Black population share as we move across neighborhoods by their HOLC grade. We cannot rule out race as a confounding variable in driving mobility outcomes.

```
#Q11C: Average of Black and White pooled absolute mobility for neighborhoods grouped by HOLC grade

holcracefixed <- atlas |>
             mutate(grade = case_when(HOLC_A > 0.5 & !is.na(HOLC_A) ~ 'A',
                                      HOLC_B > 0.5 & !is.na(HOLC_B) ~ 'B',
                                      HOLC_C > 0.5 & !is.na(HOLC_C) ~ 'C',
                                      HOLC_D > 0.5 & !is.na(HOLC_D) ~ 'D')) |>
            group_by(grade) |>
            summarize(mean_blackmobil = mean(kfr_black_pooled_p25, na.rm = TRUE),
                      mean_whitemobil = mean(kfr_white_pooled_p25, na.rm = TRUE))

holcracefixed
```

```
## # A tibble: 5 x 3
##   grade mean_blackmobil mean_whitemobil
##   <chr>           <dbl>           <dbl>
## 1 A                34.4            50.4
## 2 B                34.5            48.8
## 3 C                33.2            46.3
## 4 D                31.6            44.1
```

```
## 5 <NA>              34.1            46.3
```

When we look at absolute mobility rates by HOLC neighborhood grade and race, racial composition cannot be a confounder in this analysis because we are isolating racial variables. When we look solely at Black mobility rates or white mobility rates, we are looking solely at Black child outcomes, so the potential effect that former HOLC neighborhood grade may have on mobility outcomes should be more clear.

Looking at Black and white pooled mobility rates, it's evident that mobility percentile ranks decrease as neighborhood HOLC grades decrease. There is a decrease of roughly 3 percentile ranks for Black children's mobility outcomes between grade A and D neighborhoods. Similarly, there is a decrease of roughly 6 percentile ranks for White children's outcomes between A and D neighborhoods. What's notable is that the average mobility rank for white children living in D neighborhoods remains higher (44 vs. 34) than the average mobility rank for Black children who lived in A neighborhoods. Clearly, HOLC grade and race have separate but similar impacts on mobility outcomes.

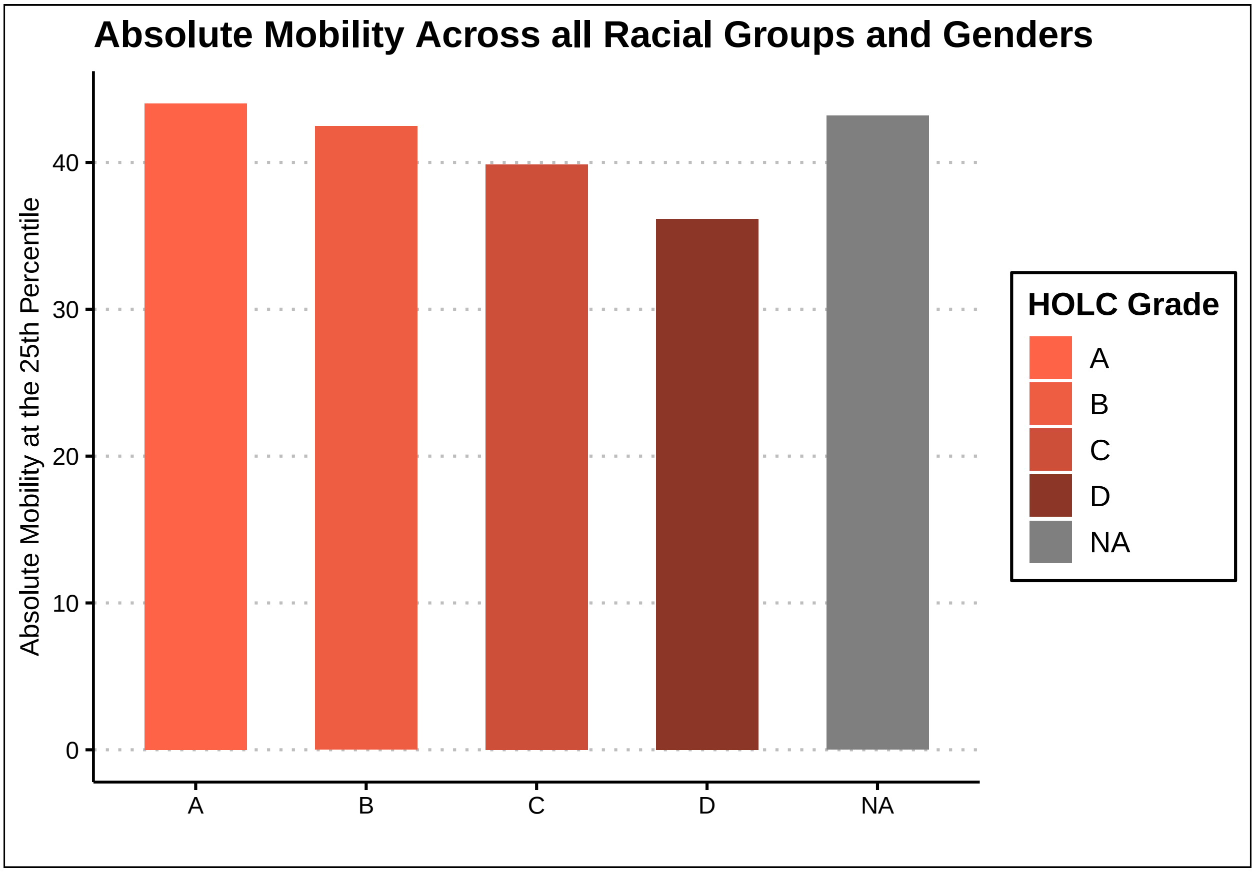
```
#Q11D: Analyzing potential causal relationship between HOLC neighborhood grade and vegetation, extreme 

holccausal <- atlas |>
             mutate(grade = case_when(HOLC_A > 0.5 & !is.na(HOLC_A) ~ 'A',
                                      HOLC_B > 0.5 & !is.na(HOLC_B) ~ 'B',
                                      HOLC_C > 0.5 & !is.na(HOLC_C) ~ 'C',
                                      HOLC_D > 0.5 & !is.na(HOLC_D) ~ 'D')) |>
            group_by(grade) |>
            summarize(mean_veg = mean(vegetation, na.rm = TRUE),
                      mean_heat = mean(extreme_heat, na.rm = TRUE),
                      mean_developed = mean(developed, na.rm = TRUE))

holccausal
```

```
## # A tibble: 5 x 4
##   grade mean_veg mean_heat mean_developed
##   <chr>    <dbl>     <dbl>          <dbl>
## 1 A      -0.0802      4.22          0.892
## 2 B      -0.205       5.73          0.940
## 3 C      -0.211       5.76          0.943
## 4 D      -0.263       6.39          0.945
## 5 <NA>   -0.0721      2.40          0.544
```

The analysis shows that neighborhoods with higher HOLC grades have more vegetation, lower extreme heat temperatures, and slightly less developed land (all three variables relative to a baseline). While all graded neighborhood levels have lower levels of vegetation than the baseline (which makes sense, since graded neighborhoods were more likely to be in urban areas), the level of vegetation gets lower with each decreasing HOLC grade. As HOLC grades decrease, there is a rise in extreme land surface heat during summer. Finally, while all graded tracts are fairly highly developed, lower graded tracts have higher rates of land development. This suggests that factors like less green space and investment in public parks are a legacy of redlining, and potentially contributing to lower mobility outcomes today.

```
#Q11E: Visualizing variables means affected by HOLC neighborhood grades

#Absolute mobility pooled for all races and genders, grouped by HOLC neighborhood grade

mobilgraph <- atlas |>
             mutate(grade = case_when(HOLC_A > 0.5 & !is.na(HOLC_A) ~ 'A',
                                      HOLC_B > 0.5 & !is.na(HOLC_B) ~ 'B',
                                      HOLC_C > 0.5 & !is.na(HOLC_C) ~ 'C',
                                      HOLC_D > 0.5 & !is.na(HOLC_D) ~ 'D')) |>
            group_by(grade) |>
```

```
            summarize(mean_absolute_mobility = mean(kfr_pooled_pooled_p25, na.rm = TRUE)) |>
            ggplot(aes(x = grade, y = mean_absolute_mobility, fill = grade)) +
            geom_bar(stat = "identity", show.legend = TRUE, width = 0.6) +
            scale_fill_manual(values = c("tomato1", "tomato2", "tomato3", "tomato4")) +
            labs(x = "",
                 y = "Absolute Mobility at the 25th Percentile",
                 title = "Absolute Mobility Across all Racial Groups and Genders",
                 fill = "HOLC Grade") +
            theme_clean()

mobilgraph
```

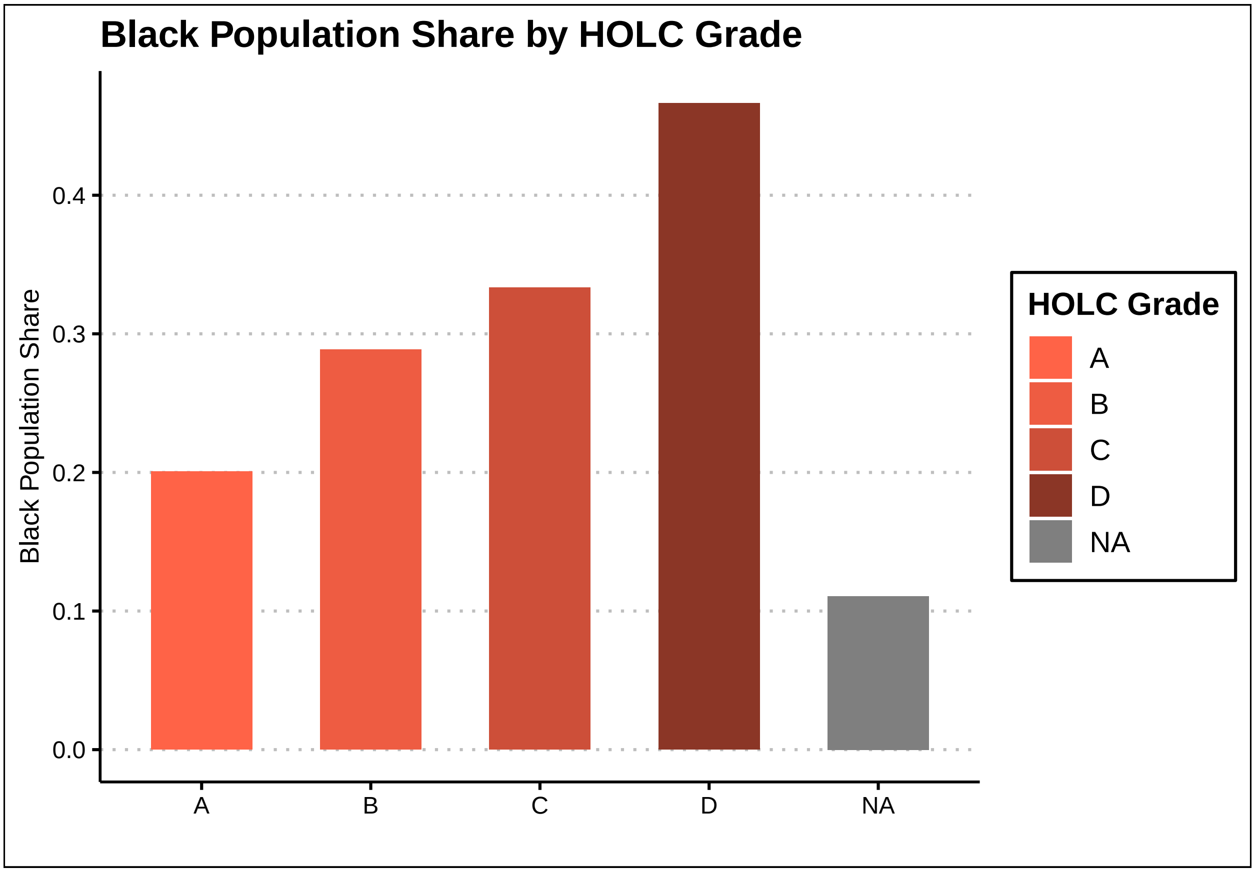



```
ggsave("mobilgraph.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
#Visualizing Black population share across census tracts grouped by HOLC neighborhood grade

holcblackgraph <- atlas |>
             mutate(grade = case_when(HOLC_A > 0.5 & !is.na(HOLC_A) ~ 'A',
                                      HOLC_B > 0.5 & !is.na(HOLC_B) ~ 'B',
                                      HOLC_C > 0.5 & !is.na(HOLC_C) ~ 'C',
                                      HOLC_D > 0.5 & !is.na(HOLC_D) ~ 'D')) |>
            group_by(grade) |>
            summarize(mean_blackpop = mean(share_black2000, na.rm = TRUE))|>
            ggplot(aes(x = grade, y = mean_blackpop, fill = grade)) +
            geom_bar(stat = "identity", show.legend = TRUE, width = 0.6) +
```

```
          scale_fill_manual(values = c("tomato1", "tomato2", "tomato3", "tomato4")) +
          labs(x = "",
               y = "Black Population Share",
               title = "Black Population Share by HOLC Grade",
               fill = "HOLC Grade") +
          theme_clean()

holcblackgraph
```

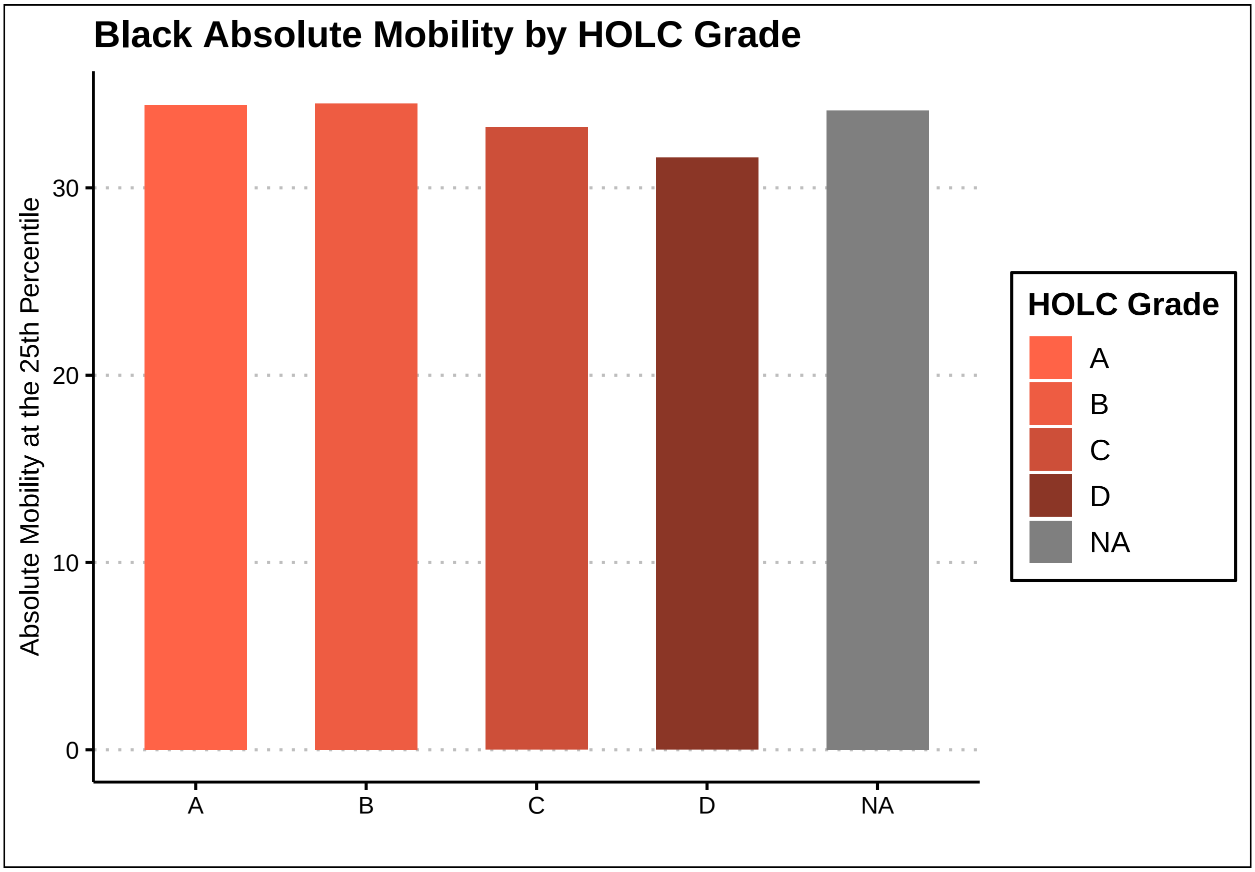



```
ggsave("holcblackgraph.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
#Visualizing Black and White pooled absolute mobility for neighborhoods grouped by HOLC grade

#Black mobility by HOLC graded neighborhood
blackmobilgraph <- atlas |>
           mutate(grade = case_when(HOLC_A > 0.5 & !is.na(HOLC_A) ~ 'A',
                                    HOLC_B > 0.5 & !is.na(HOLC_B) ~ 'B',
                                    HOLC_C > 0.5 & !is.na(HOLC_C) ~ 'C',
                                    HOLC_D > 0.5 & !is.na(HOLC_D) ~ 'D')) |>
          group_by(grade) |>
          summarize(mean_blackmobil = mean(kfr_black_pooled_p25, na.rm = TRUE),
                    mean_whitemobil = mean(kfr_white_pooled_p25, na.rm = TRUE)) |>
          ggplot() +
          geom_bar(aes(x = grade, y = mean_blackmobil, fill = grade), stat = "identity", show.legend =
          scale_fill_manual(values = c("tomato1", "tomato2", "tomato3", "tomato4")) +
```

```
          labs(x = "",
               y = "Absolute Mobility at the 25th Percentile",
               title = "Black Absolute Mobility by HOLC Grade",
               fill = "HOLC Grade") +
          theme_clean()

blackmobilgraph
```

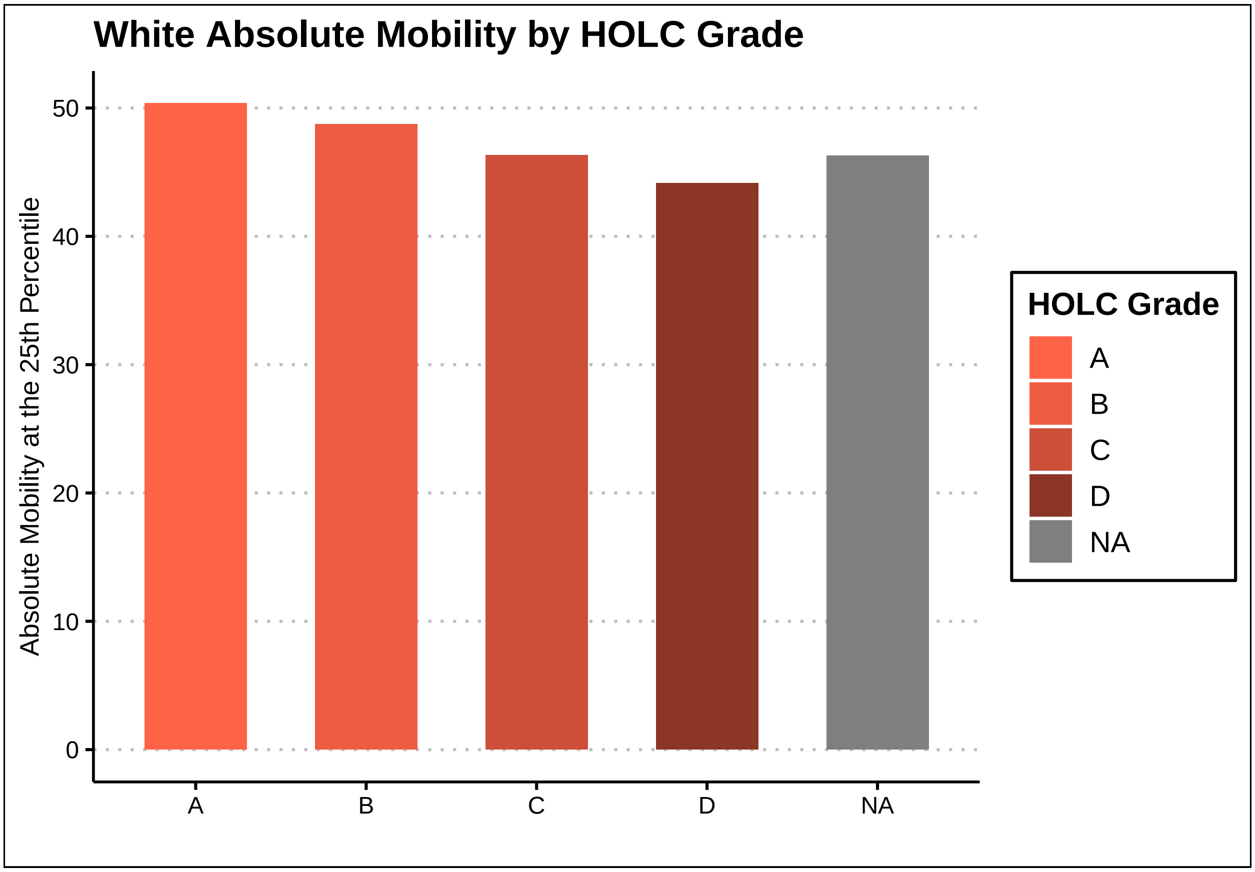



```
ggsave("blackmobilgraph.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
#White mobility by HOLC graded neighborhood
whitemobilgraph <- atlas |>
             mutate(grade = case_when(HOLC_A > 0.5 & !is.na(HOLC_A) ~ 'A',
                                      HOLC_B > 0.5 & !is.na(HOLC_B) ~ 'B',
                                      HOLC_C > 0.5 & !is.na(HOLC_C) ~ 'C',
                                      HOLC_D > 0.5 & !is.na(HOLC_D) ~ 'D')) |>
            group_by(grade) |>
            summarize(mean_blackmobil = mean(kfr_black_pooled_p25, na.rm = TRUE),
                      mean_whitemobil = mean(kfr_white_pooled_p25, na.rm = TRUE)) |>
            ggplot(aes(x = grade, y = mean_whitemobil, fill = grade)) +
            geom_bar(stat = "identity", show.legend = TRUE, width = 0.6) +
            scale_fill_manual(values = c("tomato1", "tomato2", "tomato3", "tomato4")) +
            labs(x = "",
                 y = "Absolute Mobility at the 25th Percentile",
                 title = "White Absolute Mobility by HOLC Grade",
```

```
             fill = "HOLC Grade") +
          theme_clean()

whitemobilgraph
```

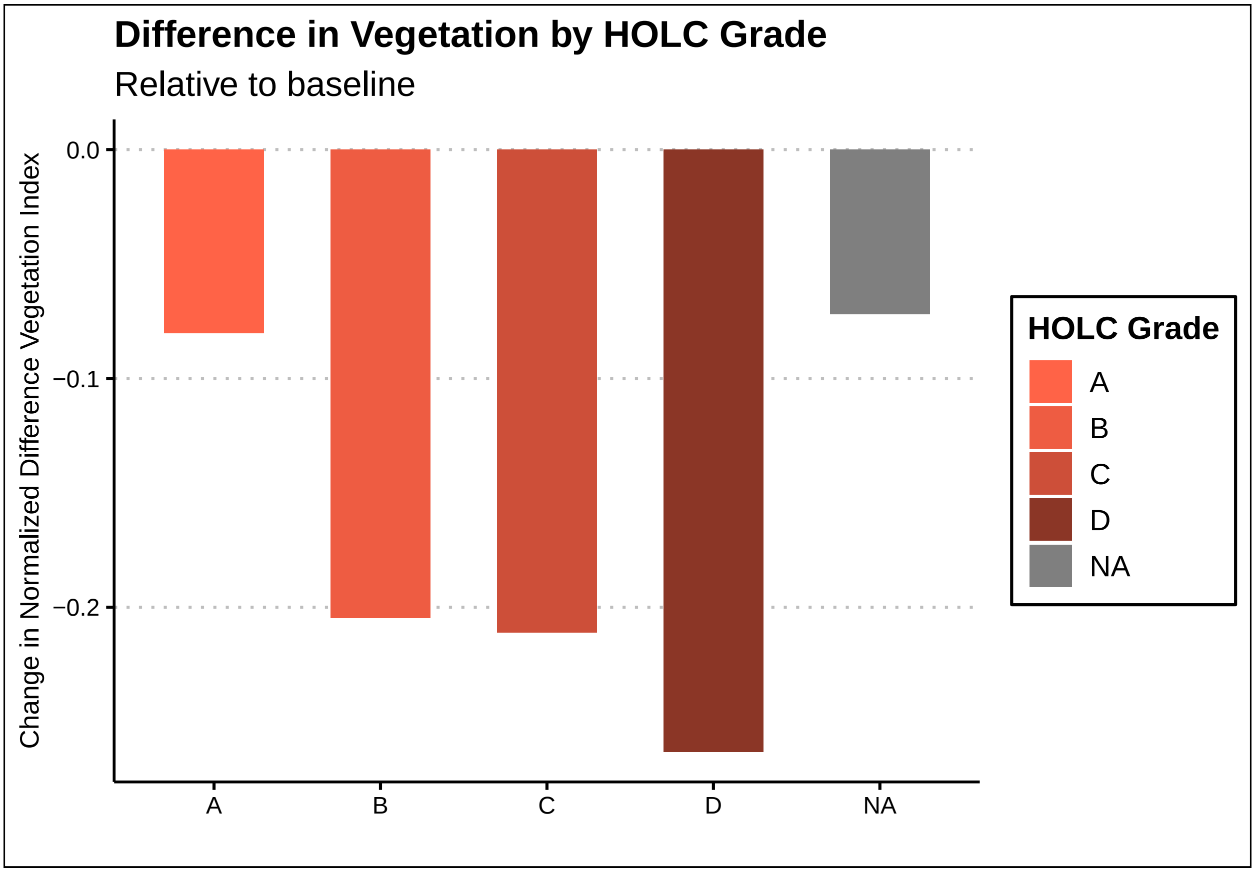



```
ggsave("whitemobilgraph.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
#Visualizing vegetation by HOLC neighborhood grade

veggraph <- atlas |>
            mutate(grade = case_when(HOLC_A > 0.5 & !is.na(HOLC_A) ~ 'A',
                                     HOLC_B > 0.5 & !is.na(HOLC_B) ~ 'B',
                                     HOLC_C > 0.5 & !is.na(HOLC_C) ~ 'C',
                                     HOLC_D > 0.5 & !is.na(HOLC_D) ~ 'D')) |>
          group_by(grade) |>
          summarize(mean_veg = mean(vegetation, na.rm = TRUE),
                    mean_heat = mean(extreme_heat, na.rm = TRUE),
                    mean_developed = mean(developed, na.rm = TRUE)) |>
          ggplot(aes(x = grade, y = mean_veg, fill = grade)) +
          geom_bar(stat = "identity", show.legend = TRUE, width = 0.6) +
          scale_fill_manual(values = c("tomato1", "tomato2", "tomato3", "tomato4")) +
          labs(x = "",
               y = "Change in Normalized Difference Vegetation Index",
               title = "Difference in Vegetation by HOLC Grade",
               subtitle = "Relative to baseline",
```

```
              fill = "HOLC Grade") +
            theme_clean()

veggraph
```

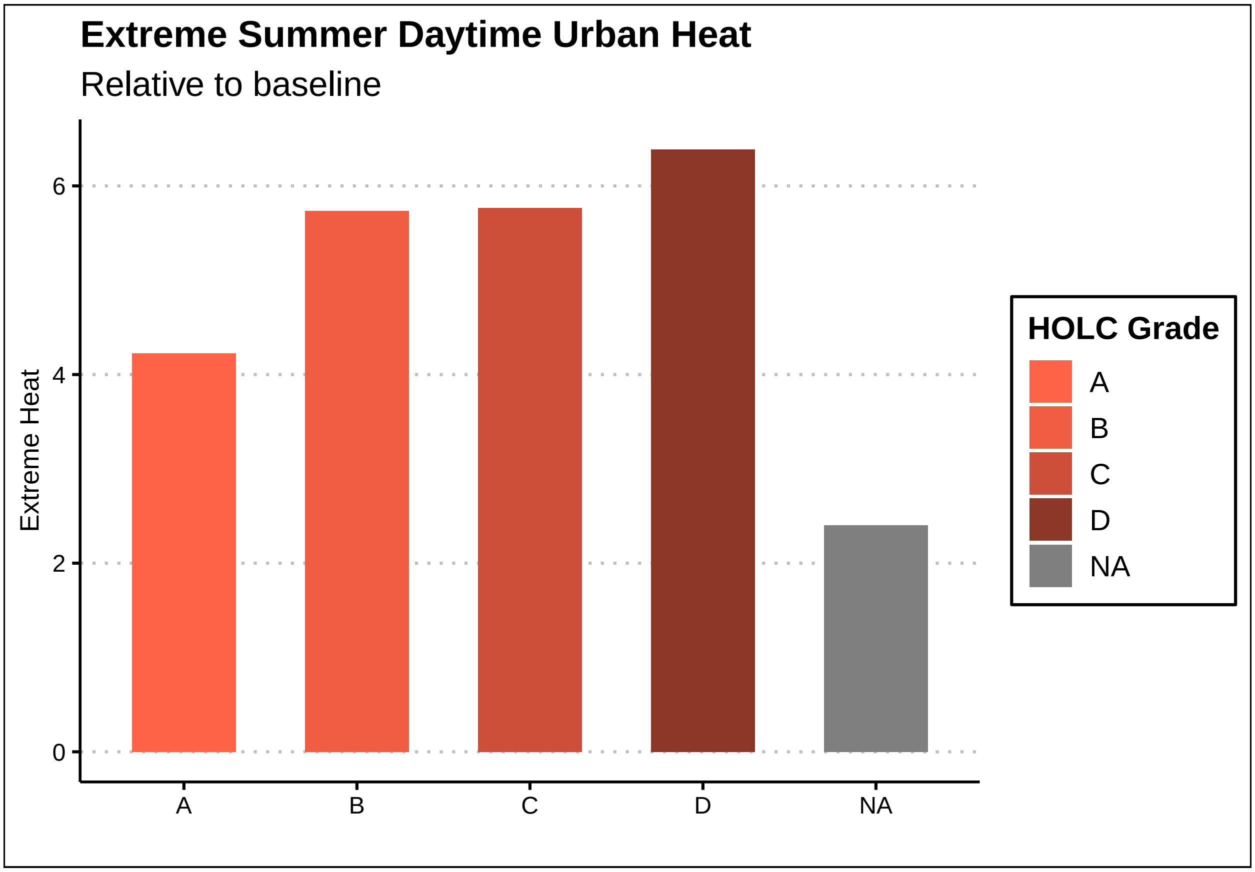



```
ggsave("veggraph.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
#Visualizing extreme summer daytime heat by HOLC grade

heatgraph <- atlas |>
             mutate(grade = case_when(HOLC_A > 0.5 & !is.na(HOLC_A) ~ 'A',
                                      HOLC_B > 0.5 & !is.na(HOLC_B) ~ 'B',
                                      HOLC_C > 0.5 & !is.na(HOLC_C) ~ 'C',
                                      HOLC_D > 0.5 & !is.na(HOLC_D) ~ 'D')) |>
            group_by(grade) |>
            summarize(mean_veg = mean(vegetation, na.rm = TRUE),
                      mean_heat = mean(extreme_heat, na.rm = TRUE),
                      mean_developed = mean(developed, na.rm = TRUE)) |>
            ggplot(aes(x = grade, y = mean_heat, fill = grade)) +
            geom_bar(stat = "identity", show.legend = TRUE, width = 0.6) +
            scale_fill_manual(values = c("tomato1", "tomato2", "tomato3", "tomato4")) +
            labs(x = "",
                 y = "Extreme Heat",
                 title = "Extreme Summer Daytime Urban Heat",
                 subtitle = "Relative to baseline",
```

```
             fill = "HOLC Grade") +
        theme_clean()

heatgraph
```




```
ggsave("heatgraph.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
#Visualizing land development by HOLC grade

devgraph <- atlas |>
             mutate(grade = case_when(HOLC_A > 0.5 & !is.na(HOLC_A) ~ 'A',
                                      HOLC_B > 0.5 & !is.na(HOLC_B) ~ 'B',
                                      HOLC_C > 0.5 & !is.na(HOLC_C) ~ 'C',
                                      HOLC_D > 0.5 & !is.na(HOLC_D) ~ 'D')) |>
            group_by(grade) |>
            summarize(mean_veg = mean(vegetation, na.rm = TRUE),
                      mean_heat = mean(extreme_heat, na.rm = TRUE),
                      mean_developed = mean(developed, na.rm = TRUE)) |>
            ggplot(aes(x = grade, y = mean_developed, fill = grade)) +
            geom_bar(stat = "identity", show.legend = TRUE, width = 0.6) +
            scale_fill_manual(values = c("tomato1", "tomato2", "tomato3", "tomato4")) +
            labs(x = "",
                 y = "Fraction of Land Area Developed",
                 title = "Land Development by HOLC Grade",
                 fill = "HOLC Grade")
```

```r
devgraph
```


```r
ggsave("devgraph.png")
```

```
## Saving 6.5 x 4.5 in image
```

```r
#Question 12: Tracking air pollution via particulate matter changes over time

pm82 <- mean(atlas$pm25_1982, na.rm = TRUE)
pm90 <- mean(atlas$pm25_1990, na.rm = TRUE)
pm00 <- mean(atlas$pm25_2000, na.rm = TRUE)
pm10 <- mean(atlas$pm25_2010, na.rm = TRUE)

airpollution <- data.frame(c("1982", "1990", "2000", "2010"),
                           c(pm82, pm90, pm00, pm10))
names(airpollution)[1] <- "Year"
names(airpollution)[2] <- "Particulate Matter Level"
airpollution
```

```
##   Year Particulate Matter Level
## 1 1982                20.410793
## 2 1990                16.844180
## 3 2000                12.500640
## 4 2010                 9.286397
```

We see that air pollution, as measured by the particulate matter level data, has steadily decreased in the United States since 1982, with an average decrease of about 4 units fine particulate matter every 10 years.

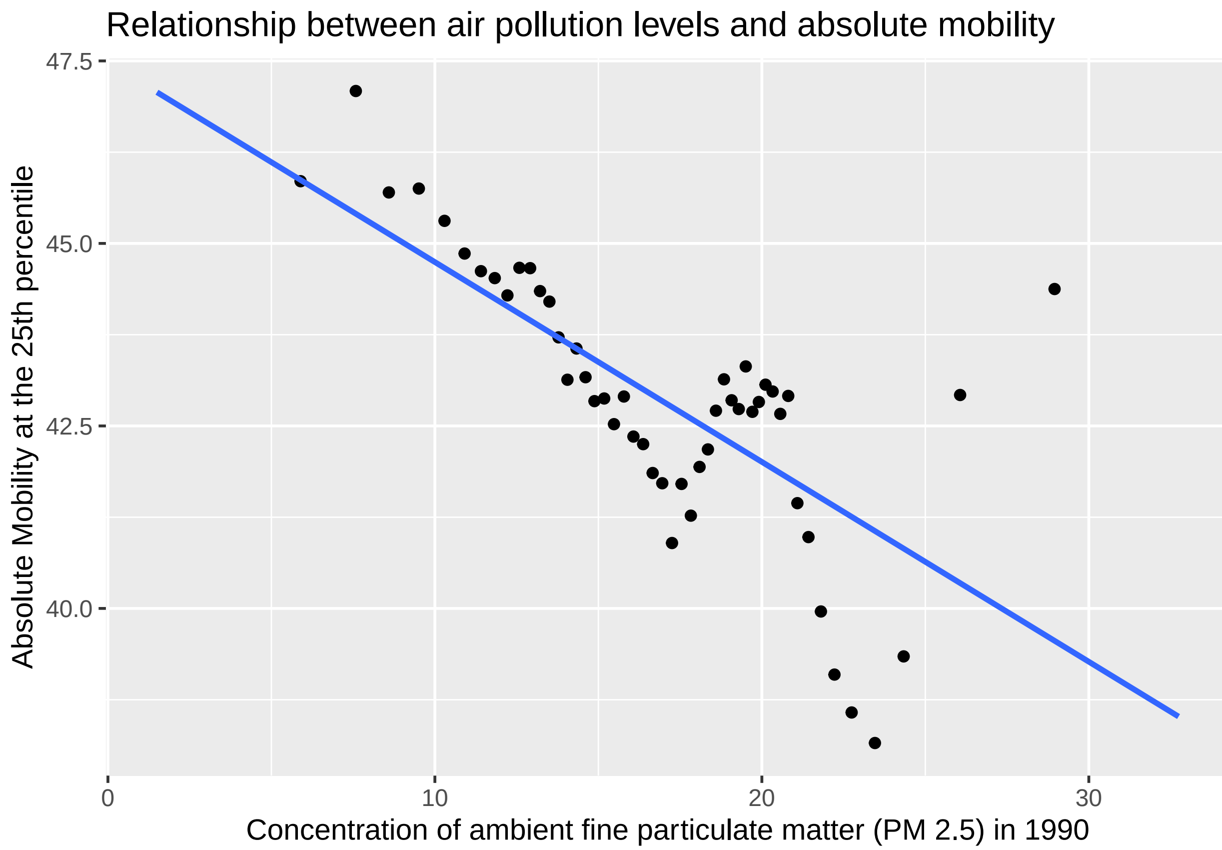
```r
#Q12B: Comparing air pollution in my home Census tract, state, and nationwide

myhood <- atlas %>% subset(state == "37" & county == "183" & tract == "53110") #creating data frame of 
myhoodap <- mean(myhood$pm25_1990, na.rm = TRUE)
ncap <- mean(atlas$pm25_1990[atlas$state == "37"], na.rm = TRUE)
usap <- mean(atlas$pm25_1990, na.rm = TRUE)

apcomp <- data.frame(c("My Census Tract", "North Carolina", "United States"),
                     c(myhoodap, ncap, usap))
names(apcomp)[1] <- "Level"
names(apcomp)[2] <- "PM Concentration"
apcomp
```

```
##             Level PM Concentration
## 1 My Census Tract         17.37184
## 2  North Carolina         18.58069
## 3   United States         16.84418
```

We see that the national air pollution average in 1990 was approximately 16.84 PM units, with higher concentrations of particulate matter for my home tract (17.37), and even higher for my home state of North Carolina (18.58).

```r
#Q12C: Visualizing the relationship between absolute mobility and air pollution in 1990

scatterap <- atlas |>
             ggplot(aes(x = pm25_1990, y = kfr_pooled_pooled_p25)) +
             stat_binmean(n = 50) +
             stat_smooth(method = "lm", se = FALSE) +
             labs(title = "Relationship between air pollution levels and absolute mobility",
                  x = "Concentration of ambient fine particulate matter (PM 2.5) in 1990",
                  y = "Absolute Mobility at the 25th percentile")

scatterap
```

```
## Warning: Removed 1736 rows containing non-finite values (`stat_binmean()`).

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 1736 rows containing non-finite values (`stat_smooth()`).
```




```r
ggsave("scatterap.png")
```

```
## Saving 6.5 x 4.5 in image
## Warning: Removed 1736 rows containing non-finite values (`stat_binmean()`).
## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 1736 rows containing non-finite values (`stat_smooth()`).
```

```r
#Q12D: Deriving the correlation coefficient between absolute mobility and air pollution in 1990

cor(atlas$kfr_pooled_pooled_p25, atlas$pm25_1990, use = "complete.obs")
```

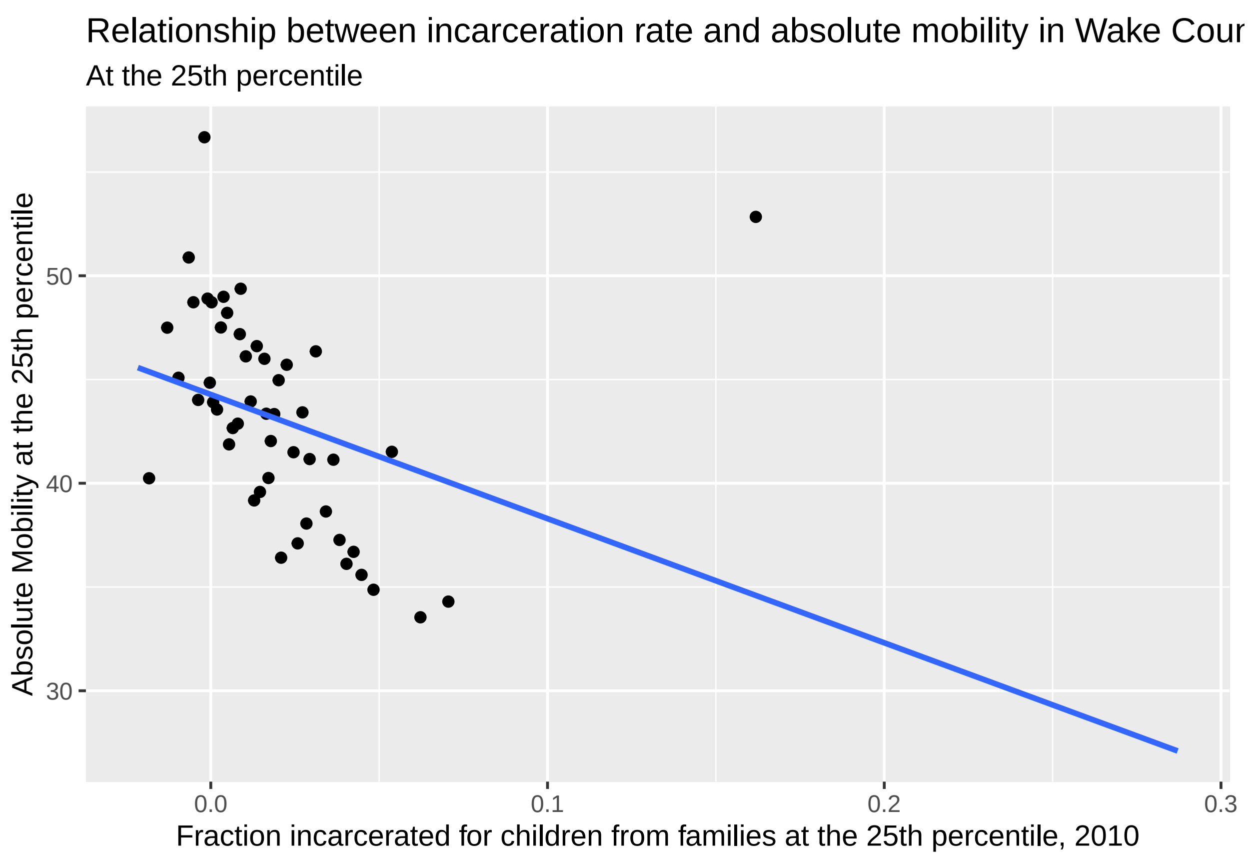
```
## [1] -0.1837617
```

The correlation coefficient between absolute mobility at the 25th percentile and air pollution levels in 1990 is approximately -0.18. This is smaller than the correlation at the county level. This makes sense because data grouped at the tract level leads to a larger sample size than data grouped at the county level. With a larger sampel size, variability increases, leading to a reduced correlation coefficient.

```r
#Q13A: Visualizing the relationship between strongly related covariates and absolute mobility

#Covariate 1: Incarceration Rate
wakeco <- atlas |>
          filter(state == 37,
                 county == 183) #creating data frame of just my home county

scatterjail <- wakeco |>
             ggplot(aes(x = jail_pooled_pooled_p25, y = kfr_pooled_pooled_p25)) +
```

```
            stat_binmean(n = 50) +
            stat_smooth(method = "lm", se = FALSE) +
            labs(title = "Relationship between incarceration rate and absolute mobility in Wake County
                 subtitle = "At the 25th percentile",
                 x = "Fraction incarcerated for children from families at the 25th percentile, 2010",
                 y = "Absolute Mobility at the 25th percentile")

scatterjail
```

```
## Warning: Removed 2 rows containing non-finite values (`stat_binmean()`).
## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).
```



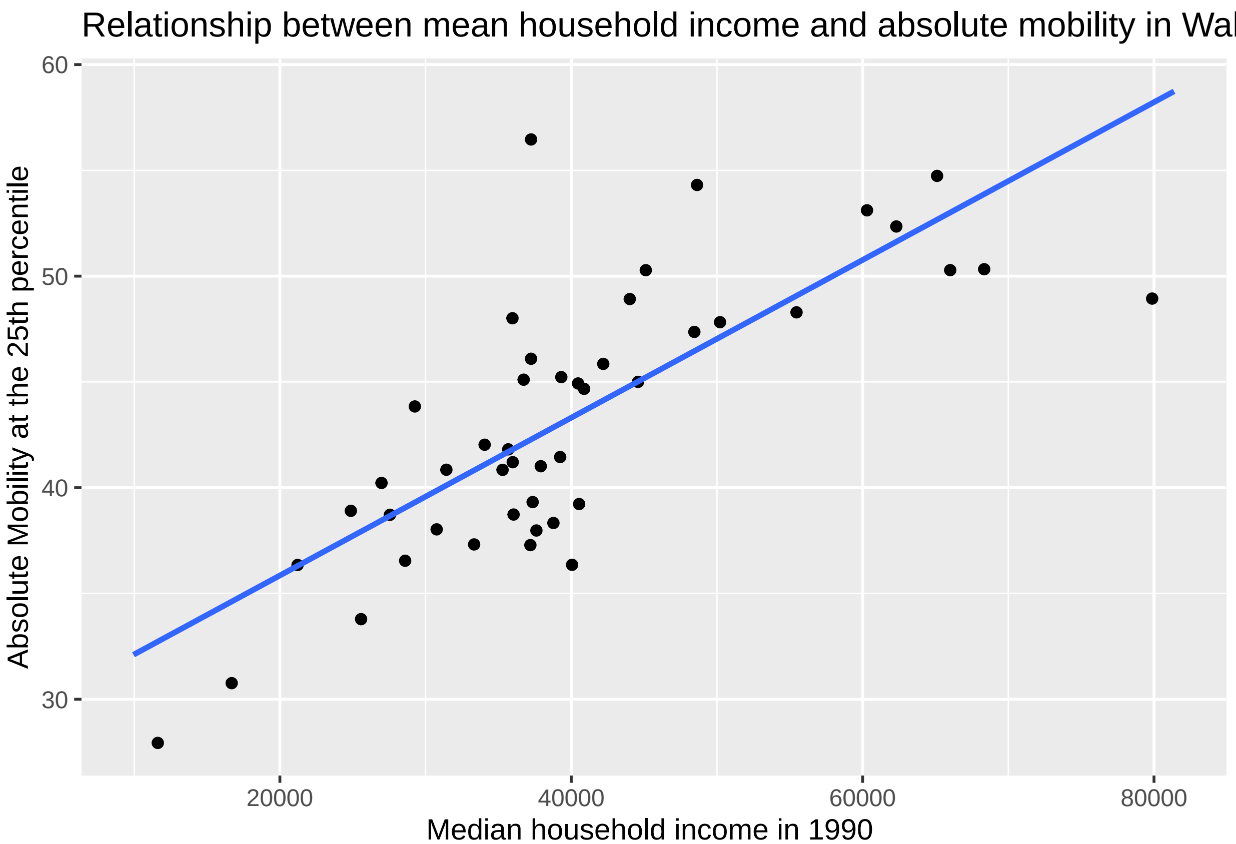

```
ggsave("scatterjail.png")
```

```
## Saving 6.5 x 4.5 in image
## Warning: Removed 2 rows containing non-finite values (`stat_binmean()`).
## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).
```

```
#Covariate 2: Median household income

scatterincome <- wakeco |>
            ggplot(aes(x = med_hhinc1990, y = kfr_pooled_pooled_p25)) +
            stat_binmean(n = 50) +
```

```
            stat_smooth(method = "lm", se = FALSE) +
            labs(title = "Relationship between mean household income and absolute mobility in Wake Cou
                 x = "Median household income in 1990 ",
                 y = "Absolute Mobility at the 25th percentile")

scatterincome
```

```
## Warning: Removed 2 rows containing non-finite values (`stat_binmean()`).
## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).
```



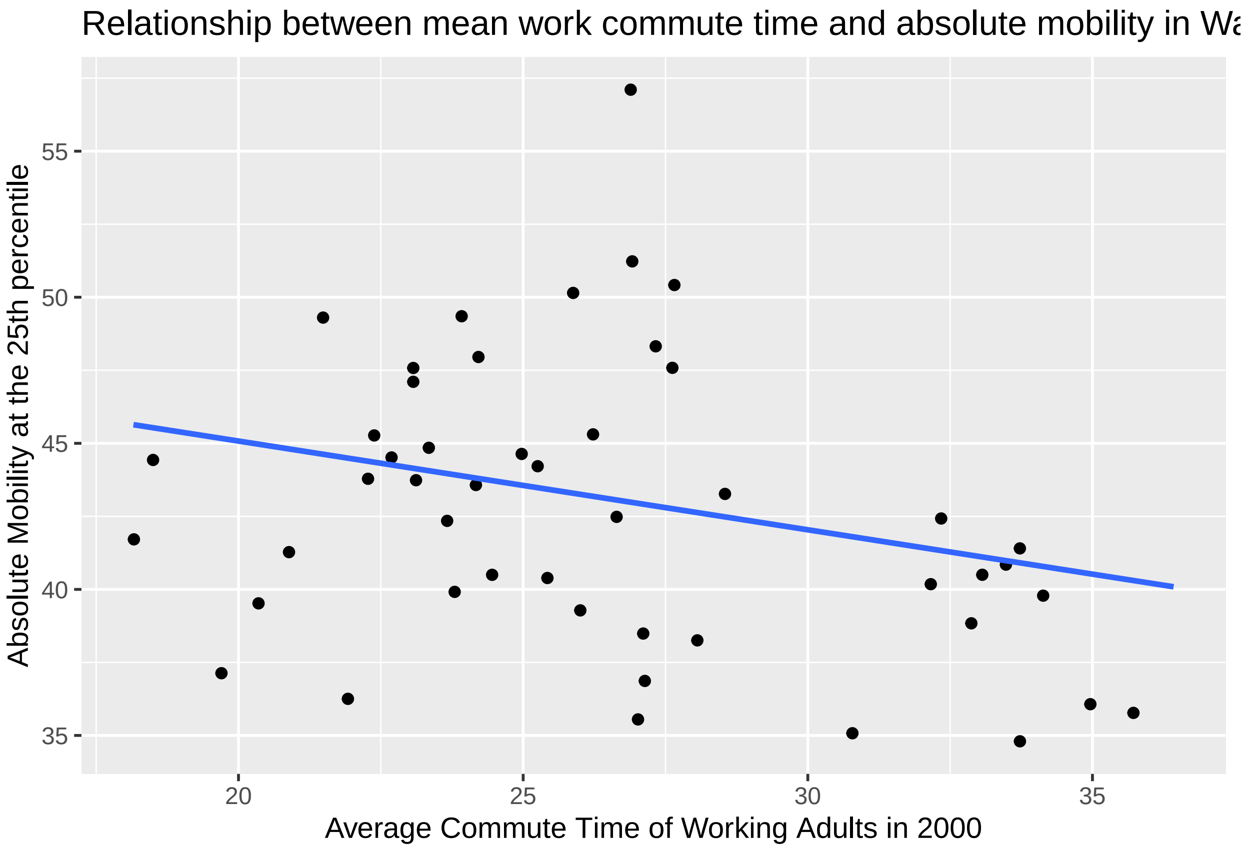

```
ggsave("scatterincome.png")
```

```
## Saving 6.5 x 4.5 in image
## Warning: Removed 2 rows containing non-finite values (`stat_binmean()`).
## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).
```

```
#Covariate 3: Average Commute Time in 2000

scattercommute <- wakeco |>
            ggplot(aes(x = mean_commutetime2000, y = kfr_pooled_pooled_p25)) +
            stat_binmean(n = 50) +
            stat_smooth(method = "lm", se = FALSE) +
            labs(title = "Relationship between mean work commute time and absolute mobility in Wake Co
                 x = "Average Commute Time of Working Adults in 2000",
```

```
                  y = "Absolute Mobility at the 25th percentile")
scattercommute
```

```
## Warning: Removed 2 rows containing non-finite values (`stat_binmean()`).
## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).
```




```
ggsave("scattercommute.png")
```

```
## Saving 6.5 x 4.5 in image
## Warning: Removed 2 rows containing non-finite values (`stat_binmean()`).
## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).
```

```
#Q13B: Reporting correlation coefficients of covariates with absolute mobility

jailcorcoef <- cor(wakeco$kfr_pooled_pooled_p25, wakeco$jail_pooled_pooled_p25, use = "complete.obs")
inccorcoef <- cor(wakeco$kfr_pooled_pooled_p25, wakeco$med_hhinc1990, use = "complete.obs")
commutecorcoef <- cor(wakeco$kfr_pooled_pooled_p25, wakeco$mean_commutetime2000, use = "complete.obs")

coefcomp <- data.frame(c("Incarceration Rate 2010", "Median Household Income 1990", "Mean Commute Time 2
                       c(jailcorcoef, inccorcoef, commutecorcoef))
names(coefcomp)[1] <- "Covariate"
names(coefcomp)[2] <- "Correlation Coefficient with absolute mobility"
```

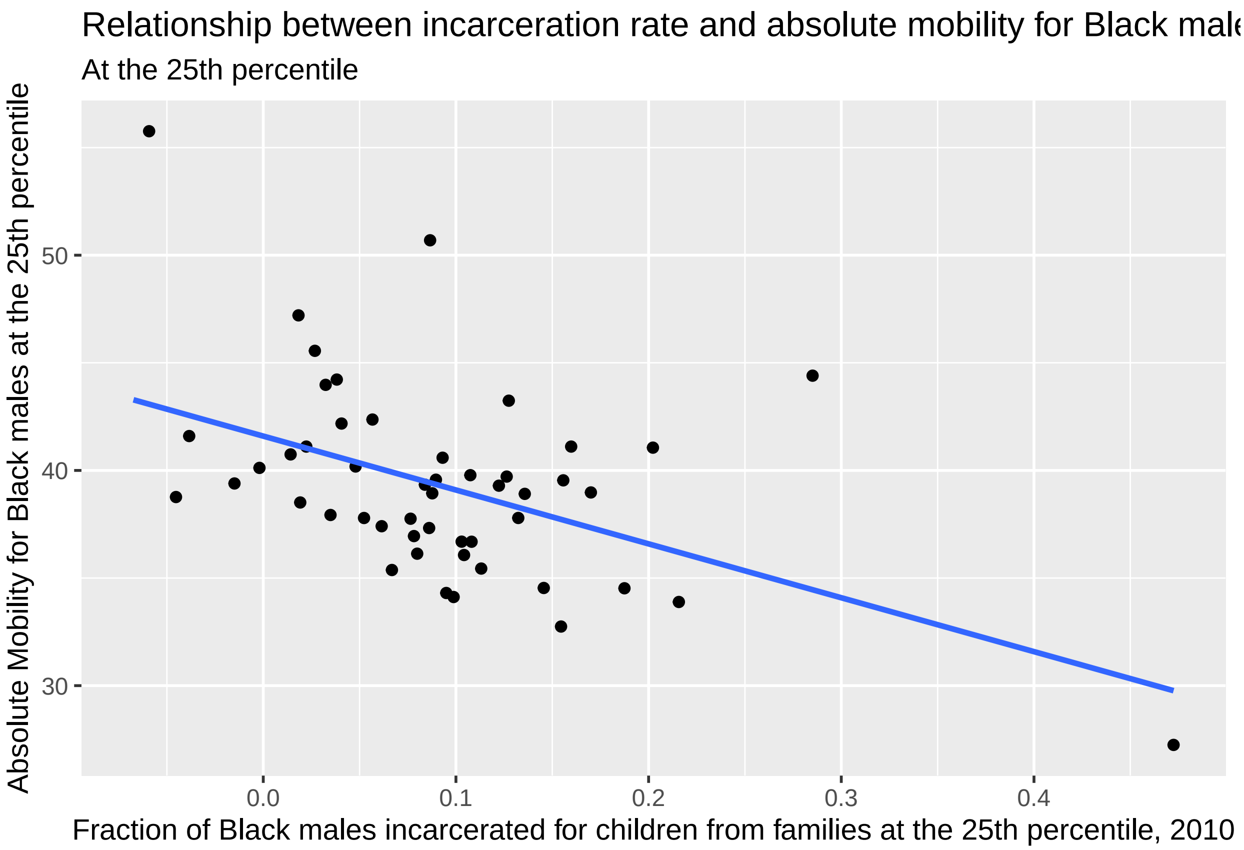
```r
coefcomp
```

```
##                     Covariate Correlation Coefficient with absolute mobility
## 1      Incarceration Rate 2010                                    -0.2266184
## 2 Median Household Income 1990                                     0.6268116
## 3       Mean Commute Time 2000                                    -0.1786249
```

We see that there are negative correlations between the incarceration rate and the mean commute time and absolute mobility, respectively, and a fairly strong positive correlation between median household income (in 1990) and absolute mobility.

```r
#Q14: Examining the relationship between covariates and absolute mobility by race and gender

#Incarceration Rate: We will look at incarceration rates for Black males, Hispanic males, and White mal

#Black Males

jailblack <- wakeco |>
             ggplot(aes(x = jail_black_male_p25, y = kir_black_male_p25)) +
             stat_binmean(n = 50) +
             stat_smooth(method = "lm", se = FALSE) +
             labs(title = "Relationship between incarceration rate and absolute mobility for Black males
                  subtitle = "At the 25th percentile",
                  x = "Fraction of Black males incarcerated for children from families at the 25th perc
                  y = "Absolute Mobility for Black males at the 25th percentile")
jailblack
```

```
## Warning: Removed 88 rows containing non-finite values (`stat_binmean()`).

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 88 rows containing non-finite values (`stat_smooth()`).
```

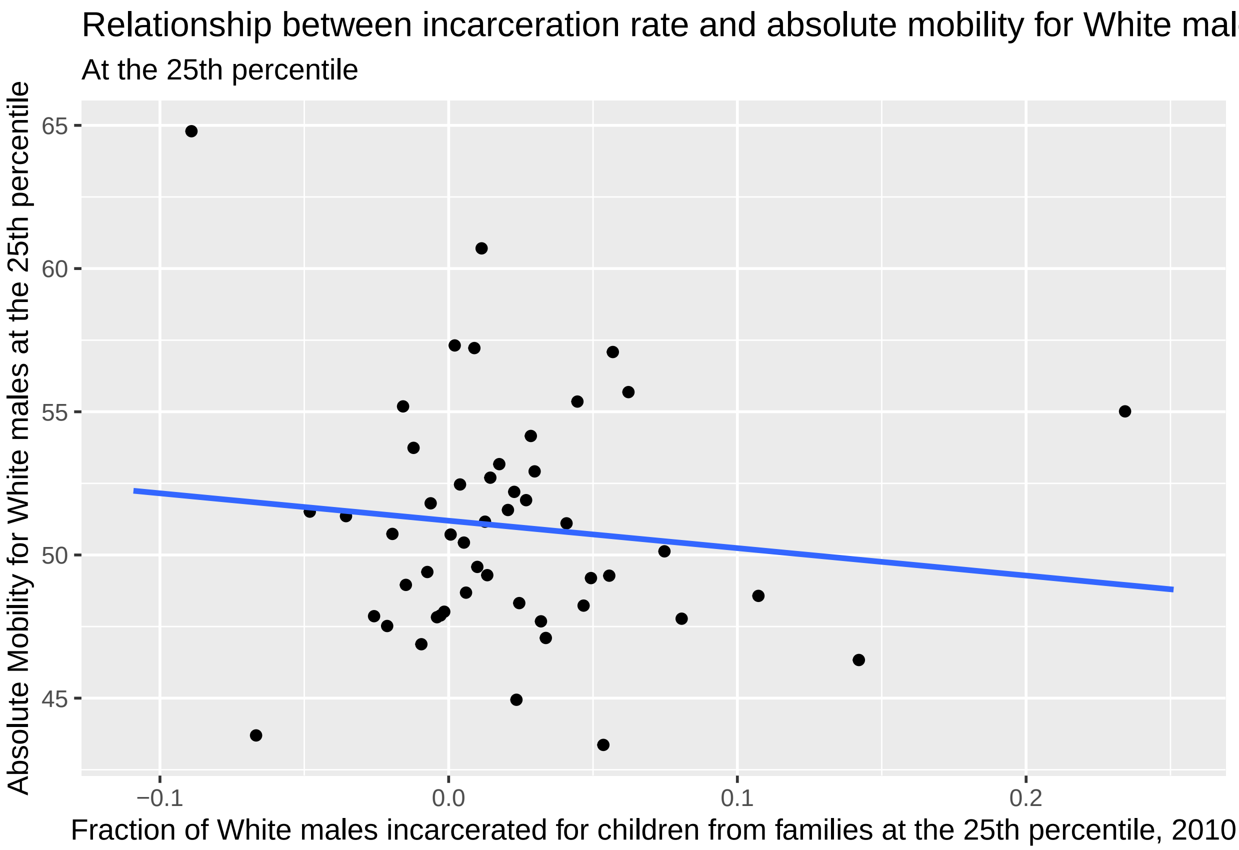



```
ggsave("jailblack.png")
```

```
## Saving 6.5 x 4.5 in image
## Warning: Removed 88 rows containing non-finite values (`stat_binmean()`).
## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 88 rows containing non-finite values (`stat_smooth()`).
```

```
#White Males

jailwhite <- wakeco |>
             ggplot(aes(x = jail_white_male_p25, y = kir_white_male_p25)) +
             stat_binmean(n = 50) +
             stat_smooth(method = "lm", se = FALSE) +
             labs(title = "Relationship between incarceration rate and absolute mobility for White males
                  subtitle = "At the 25th percentile",
                  x = "Fraction of White males incarcerated for children from families at the 25th perce
                  y = "Absolute Mobility for White males at the 25th percentile")
jailwhite
```

```
## Warning: Removed 14 rows containing non-finite values (`stat_binmean()`).
## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 14 rows containing non-finite values (`stat_smooth()`).
```

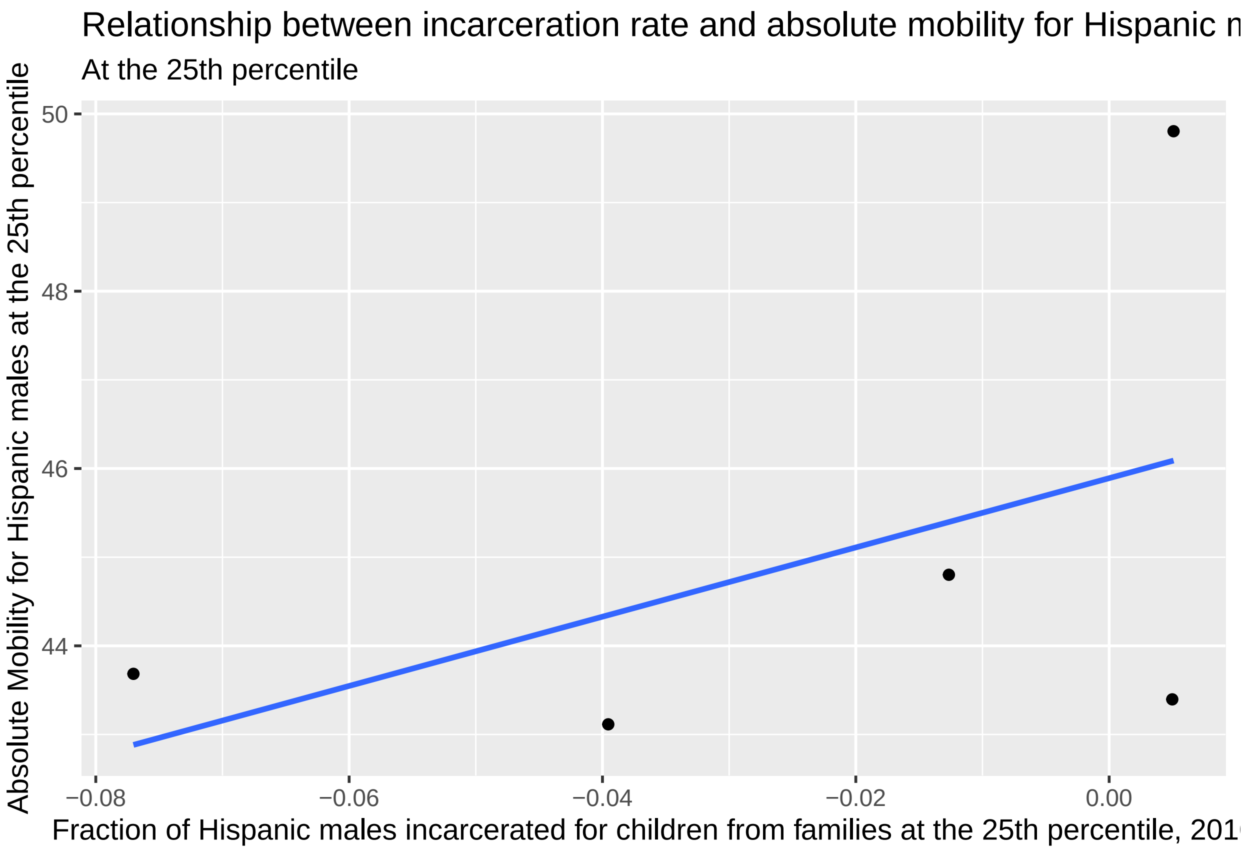



```
ggsave("jailwhite.png")
```

```
## Saving 6.5 x 4.5 in image
## Warning: Removed 14 rows containing non-finite values (`stat_binmean()`).
## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 14 rows containing non-finite values (`stat_smooth()`).
```

```
#Hispanic Males

jailhisp <- wakeco |>
            ggplot(aes(x = jail_hisp_male_p25, y = kir_hisp_male_p25)) +
            stat_binmean(n = 50) +
            stat_smooth(method = "lm", se = FALSE) +
            labs(title = "Relationship between incarceration rate and absolute mobility for Hispanic ma
                 subtitle = "At the 25th percentile",
                 x = "Fraction of Hispanic males incarcerated for children from families at the 25th pe
                 y = "Absolute Mobility for Hispanic males at the 25th percentile")
jailhisp
```

```
## Warning: Removed 181 rows containing non-finite values (`stat_binmean()`).
## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 181 rows containing non-finite values (`stat_smooth()`).
```

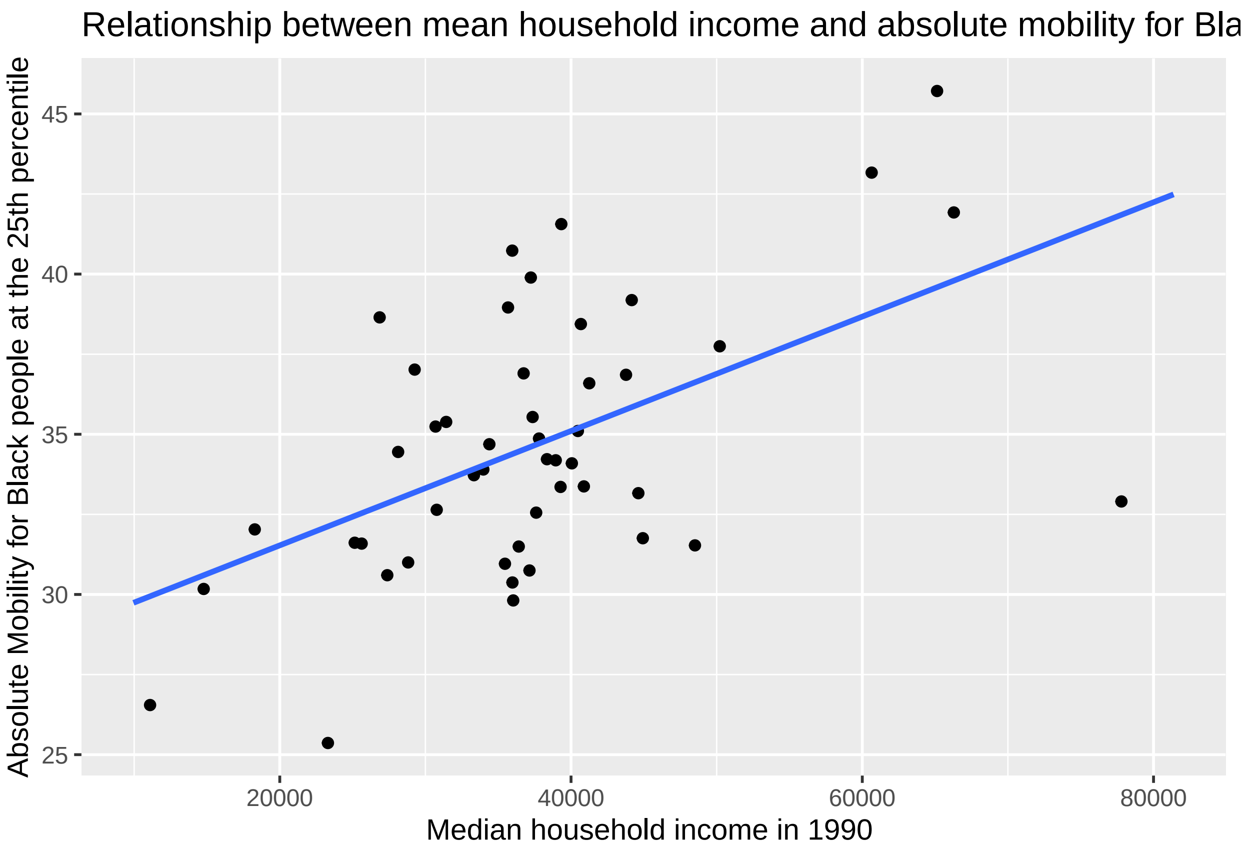



```
ggsave("jailhisp.png")
```

```
## Saving 6.5 x 4.5 in image

## Warning: Removed 181 rows containing non-finite values (`stat_binmean()`).

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 181 rows containing non-finite values (`stat_smooth()`).
```

```
#Median Household Income: Next, we will look at median household income in 1990 and absolute mobility f

#Black Children
blackincome <- wakeco |>
             ggplot(aes(x = med_hhinc1990, y = kfr_black_pooled_p25)) +
             stat_binmean(n = 50) +
             stat_smooth(method = "lm", se = FALSE) +
             labs(title = "Relationship between mean household income and absolute mobility for Black pe
                  x = "Median household income in 1990 ",
                  y = "Absolute Mobility for Black people at the 25th percentile")

blackincome
```

```
## Warning: Removed 36 rows containing non-finite values (`stat_binmean()`).

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 36 rows containing non-finite values (`stat_smooth()`).
```




```
ggsave("blackincome.png")
```

```
## Saving 6.5 x 4.5 in image
## Warning: Removed 36 rows containing non-finite values (`stat_binmean()`).
## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 36 rows containing non-finite values (`stat_smooth()`).
```

```
#White Children
whiteincome <- wakeco |>
             ggplot(aes(x = med_hhinc1990, y = kfr_white_pooled_p25)) +
             stat_binmean(n = 50) +
             stat_smooth(method = "lm", se = FALSE) +
             labs(title = "Relationship between mean household income and absolute mobility for White pe
                  x = "Median household income in 1990 ",
                  y = "Absolute Mobility for White people at the 25th percentile")

whiteincome
```

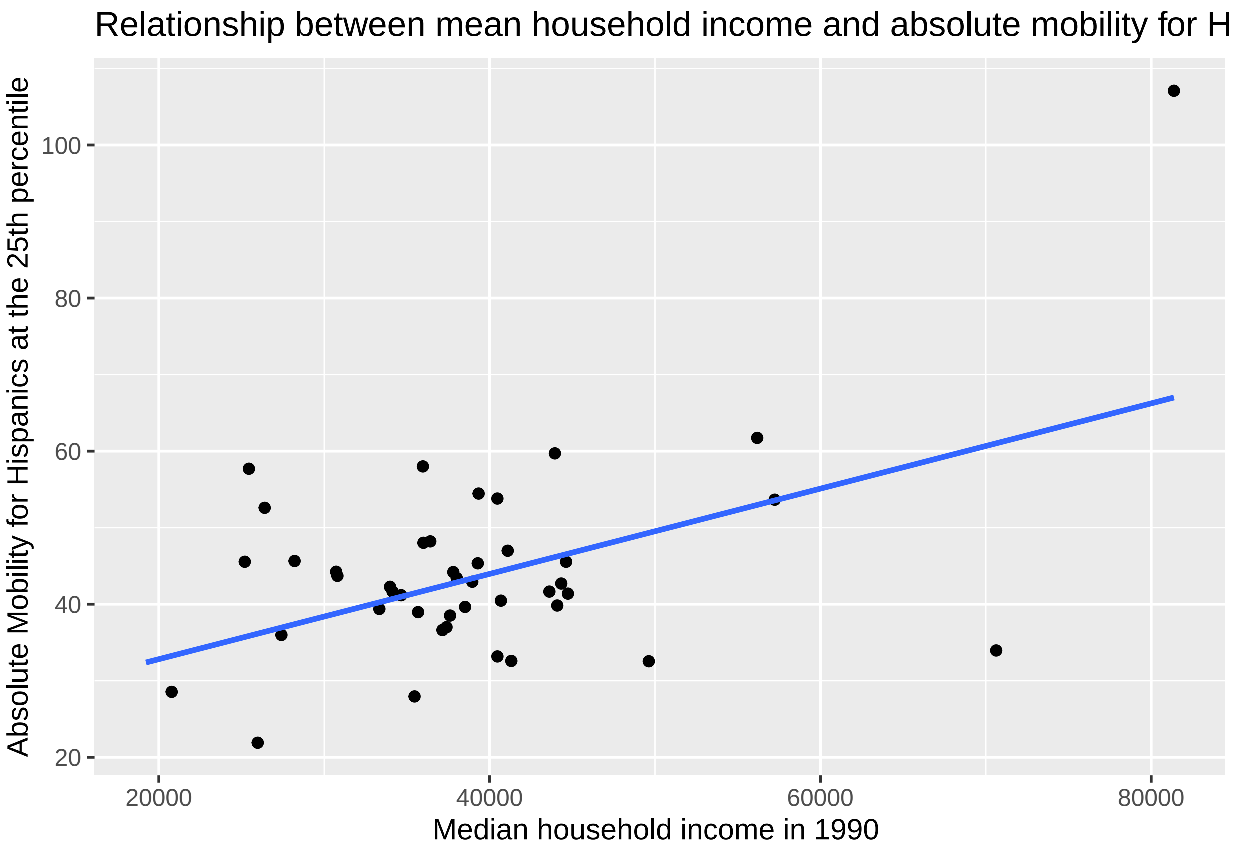
```
## Warning: Removed 8 rows containing non-finite values (`stat_binmean()`).
## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 8 rows containing non-finite values (`stat_smooth()`).
```


```
ggsave("whiteincome.png")
```

```
## Saving 6.5 x 4.5 in image
## Warning: Removed 8 rows containing non-finite values (`stat_binmean()`).
## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 8 rows containing non-finite values (`stat_smooth()`).
```

```
#Hispanic Children
hispincome <- wakeco |>
             ggplot(aes(x = med_hhinc1990, y = kfr_hisp_pooled_p25)) +
             stat_binmean(n = 50) +
             stat_smooth(method = "lm", se = FALSE) +
             labs(title = "Relationship between mean household income and absolute mobility for Hispanic
                  x = "Median household income in 1990 ",
                  y = "Absolute Mobility for Hispanics at the 25th percentile")

hispincome
```

```
## Warning: Removed 127 rows containing non-finite values (`stat_binmean()`).
## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 127 rows containing non-finite values (`stat_smooth()`).
```

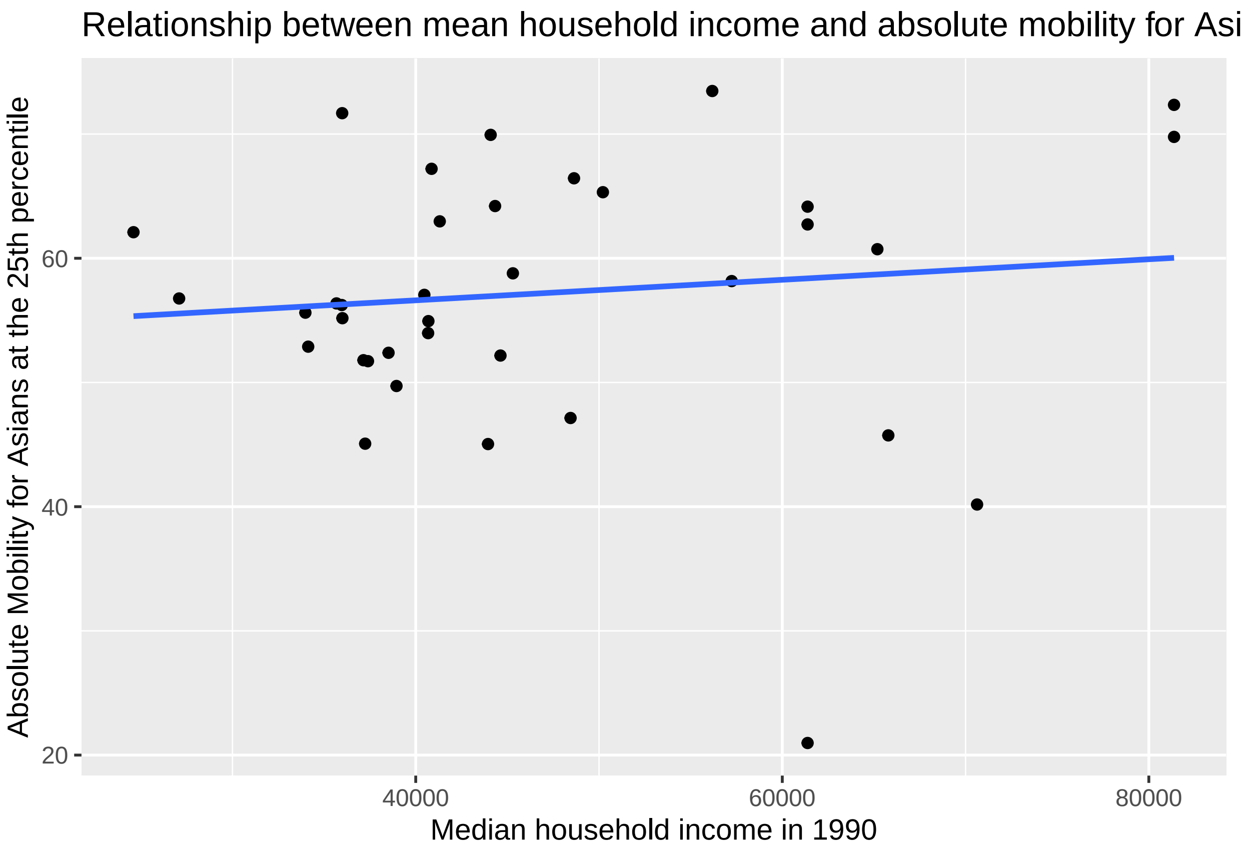



```
ggsave("hispincome.png")
```

```
## Saving 6.5 x 4.5 in image

## Warning: Removed 127 rows containing non-finite values (`stat_binmean()`).

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 127 rows containing non-finite values (`stat_smooth()`).
```

```
#Asian Children
asianincome <- wakeco |>
             ggplot(aes(x = med_hhinc1990, y = kfr_asian_pooled_p25)) +
             stat_binmean(n = 50) +
             stat_smooth(method = "lm", se = FALSE) +
             labs(title = "Relationship between mean household income and absolute mobility for Asians i
                  x = "Median household income in 1990",
                  y = "Absolute Mobility for Asians at the 25th percentile")

asianincome
```

```
## Warning: Removed 140 rows containing non-finite values (`stat_binmean()`).

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 140 rows containing non-finite values (`stat_smooth()`).
```




```r
ggsave("asianincome.png")
```

```
## Saving 6.5 x 4.5 in image

## Warning: Removed 140 rows containing non-finite values (`stat_binmean()`).

## `geom_smooth()` using formula = 'y ~ x'

## Warning: Removed 140 rows containing non-finite values (`stat_smooth()`).
```

```r
#Q14 part 2: Reporting correlation coefficients of covariates with absolute mobility across race and ge

jailblackcorcoef <- cor(wakeco$kir_black_male_p25, wakeco$jail_black_male_p25, use = "complete.obs")
jailwhitecorcoef <- cor(wakeco$kir_white_male_p25, wakeco$jail_white_male_p25, use = "complete.obs")
jailhispcorcoef <- cor(wakeco$kir_hisp_male_p25, wakeco$jail_hisp_male_p25, use = "complete.obs")
incblackcorcoef <- cor(wakeco$kfr_black_pooled_p25, wakeco$med_hhinc1990, use = "complete.obs")
incwhitecorcoef <- cor(wakeco$kfr_white_pooled_p25, wakeco$med_hhinc1990, use = "complete.obs")
inchispcorcoef <- cor(wakeco$kfr_hisp_pooled_p25, wakeco$med_hhinc1990, use = "complete.obs")
incasiancorcoef <- cor(wakeco$kfr_asian_pooled_p25, wakeco$med_hhinc1990, use = "complete.obs")

coefcomp_incarcerate <- data.frame(c("Incarceration Rate for Black Males", "Incarceration Rate for Whit
                       c(jailblackcorcoef, jailwhitecorcoef, jailhispcorcoef))
names(coefcomp_incarcerate)[1] <- "Incarceration Rate for Racial Group"
names(coefcomp_incarcerate)[2] <- "Correlation Coefficient with Absolute Mobility for that Racial Group
coefcomp_incarcerate
```

```
##     Incarceration Rate for Racial Group
## 1    Incarceration Rate for Black Males
## 2    Incarceration Rate for White Males
## 3 Incarceration Rate for Hispanic Males
```

```
##   Correlation Coefficient with Absolute Mobility for that Racial Group
## 1                                                          -0.33926569
## 2                                                          -0.05986192
## 3                                                           0.48964213
```

```
coefcomp_medinc <- data.frame(c("Median Household Income 1990", "Median Household Income 1990", "Median
                              c("Black Absolute Mobility", "White Absolute Mobility", "Hispanic Absolute
                              c(incblackcorcoef, incwhitecorcoef, inchispcorcoef, incasiancorcoef))
names(coefcomp_medinc)[1] <- "Covariate"
names(coefcomp_medinc)[2] <- "Absolute Mobility Group"
names(coefcomp_medinc)[3] <- "Correlation Coefficient with Absolute Mobility for that Racial Group"
coefcomp_medinc
```

```
##                      Covariate    Absolute Mobility Group
## 1 Median Household Income 1990    Black Absolute Mobility
## 2 Median Household Income 1990    White Absolute Mobility
## 3 Median Household Income 1990 Hispanic Absolute Mobility
## 4 Median Household Income 1990    Asian Absolute Mobility
##   Correlation Coefficient with Absolute Mobility for that Racial Group
## 1                                                           0.31632544
## 2                                                           0.45236858
## 3                                                           0.45840067
## 4                                                           0.08887583
```

When examining the relationship between incarceration rates and absolute mobility across race and gender, we see a fairly strong negative correlation coefficient between the black male incarceration rate and the absolute mobility for black males (-0.33). We see a weaker negative correlation coefficient between said covariates for white males (-0.05), and the sample size for hispanic males in Wake County is too small for us to consider the correlation coefficient for that racial group (0.48).

There is a clear pattern of fairly strong positive correlation coefficients between the median household income in 1990 and absolute mobility ranks for Blacks, Whites, Hispanics, and Asians. However, we should discount the correlation coefficient for Asian absolute mobility due to its small sample size. We should also note the difference between the correlation coefficient for Blacks' absolute mobility and the median household income (0.31) and that for Whites and Hispanics (0.452 and 0.458 respectively) - suggesting that, at least for Black children growing up in Wake County, there isn't as strong of a relationship between median household and their absolute mobility outcomes.