project_pt2

2023-04-13

```
#Initial data wrangling
#Subsetting for Wake County
wake_co <- atlas %>% subset(state == "37" & county == "183")
view(wake_co)
#How many census tracts in Wake County?
length(unique(wake_co$tract))
```

```
## [1] 186
```

```
#Subsetting for my neighborhood (Fuquay-Varina, Holly Springs, and Willow Spring)
fuquay <- wake_co |> subset(str_detect(tract_name, "Fuquay Varina") | str_detect(tract_name, "Holly Spri
view(fuquay)

#Subsetting for upper-middle class neighborhood (Cary)
cary <- wake_co |> subset(str_detect(tract_name, "Cary"))
view(cary)
```

```
#Upward mobility in Wake County, Fuquay-Varina, and Cary
wake_upmobility <- mean(wake_co$kfr_pooled_pooled_p25, na.rm = TRUE)
fuquay_upmobility <- mean(fuquay$kfr_pooled_pooled_p25, na.rm = TRUE)
cary_upmobility <- mean(cary$kfr_pooled_pooled_p25, na.rm = TRUE)


abscomp <- data.frame(c("Wake County", "Fuquay-Varina Tracts", "Cary Tracts"),
                      c(wake_upmobility, fuquay_upmobility, cary_upmobility))
names(abscomp)[1] <- "Level"
names(abscomp)[2] <- "Absolute Mobility at the 25th Percentile"
abscomp
```

```
##                 Level Absolute Mobility at the 25th Percentile
## 1          Wake County                                 43.10832
## 2 Fuquay-Varina Tracts                                 39.71029
## 3          Cary Tracts                                 48.77694
```

```
#College Degree attainment in Wake County, Fuquay, and Cary
wake_college <- mean(wake_co$frac_coll_plus2010, na.rm = TRUE)
fuquay_college <- mean(fuquay$frac_coll_plus2010, na.rm = TRUE)
cary_college <- mean(cary$frac_coll_plus2010, na.rm = TRUE)

#Median household income in Wake County, Fuquay, and Cary
wake_medinc <- mean(wake_co$med_hhinc2016, na.rm = TRUE)
fuquay_medinc <- mean(fuquay$med_hhinc2016, na.rm = TRUE)
cary_medinc <- mean(cary$med_hhinc2016, na.rm = TRUE)

#Poverty rate in Wake County, Fuquay, and Cary
wake_poor <- mean(wake_co$poor_share2010, na.rm = TRUE)
fuquay_poor <- mean(fuquay$poor_share2010, na.rm = TRUE)
```

```r
cary_poor <- mean(cary$poor_share2010, na.rm = TRUE)

indicators <- data.frame(c("Wake County", "Fuquay-Varina Tracts", "Cary Tracts"),
                         c(wake_college, fuquay_college, cary_college),
                         c(wake_medinc, fuquay_medinc, cary_medinc),
                         c(wake_poor, fuquay_poor, cary_poor))

names(indicators)[1] <- "Region"
names(indicators)[2] <- "2010 College Degree Attainment Rate"
names(indicators)[3] <- "2016 Median Household Income"
names(indicators)[4] <- "2006-2010 Poverty Rate"

indicators
```

```
##                 Region 2010 College Degree Attainment Rate
## 1          Wake County                           0.4871529
## 2 Fuquay-Varina Tracts                           0.4073116
## 3          Cary Tracts                           0.6235936
##   2016 Median Household Income 2006-2010 Poverty Rate
## 1                     77560.94             0.10126243
## 2                     78324.08             0.05733536
## 3                     98308.97             0.05075697
```

```r
#White population share in Wake County, Fuquay, and Cary
wake_white <- mean(wake_co$share_white2010, na.rm = TRUE)
fuquay_white <- mean(fuquay$share_white2010, na.rm = TRUE)
cary_white <- mean(cary$share_white2010, na.rm = TRUE)

#Black population share in Wake County, Fuquay, and Cary
wake_black <- mean(wake_co$share_black2010, na.rm = TRUE)
fuquay_black <- mean(fuquay$share_black2010, na.rm = TRUE)
cary_black <- mean(cary$share_black2010, na.rm = TRUE)

#Hispanic population share in Wake County, Fuquay, and Cary
wake_hispanic <- mean(wake_co$share_hisp2010, na.rm = TRUE)
fuquay_hispanic <- mean(fuquay$share_hisp2010, na.rm = TRUE)
cary_hispanic <- mean(cary$share_hisp2010, na.rm = TRUE)

demos <- data.frame(c("Wake County", "Fuquay-Varina Tracts", "Cary Tracts"),
                         c(wake_white, fuquay_white, cary_white),
                         c(wake_black, fuquay_black, cary_black),
                         c(wake_hispanic, fuquay_hispanic, cary_hispanic))

names(demos)[1] <- "Region"
names(demos)[2] <- "2010 White Population Share"
names(demos)[3] <- "2010 Black Population Share"
names(demos)[4] <- "2010 Hispanic Population Share"

demos
```

```
##                 Region 2010 White Population Share 2010 Black Population Share
## 1          Wake County                   0.6445501                   0.19562181
## 2 Fuquay-Varina Tracts                   0.7381529                   0.15185540
## 3          Cary Tracts                   0.7044272                   0.08447644
##   2010 Hispanic Population Share
```

```
## 1                     0.09075263
## 2                     0.07848141
## 3                     0.07485468
```

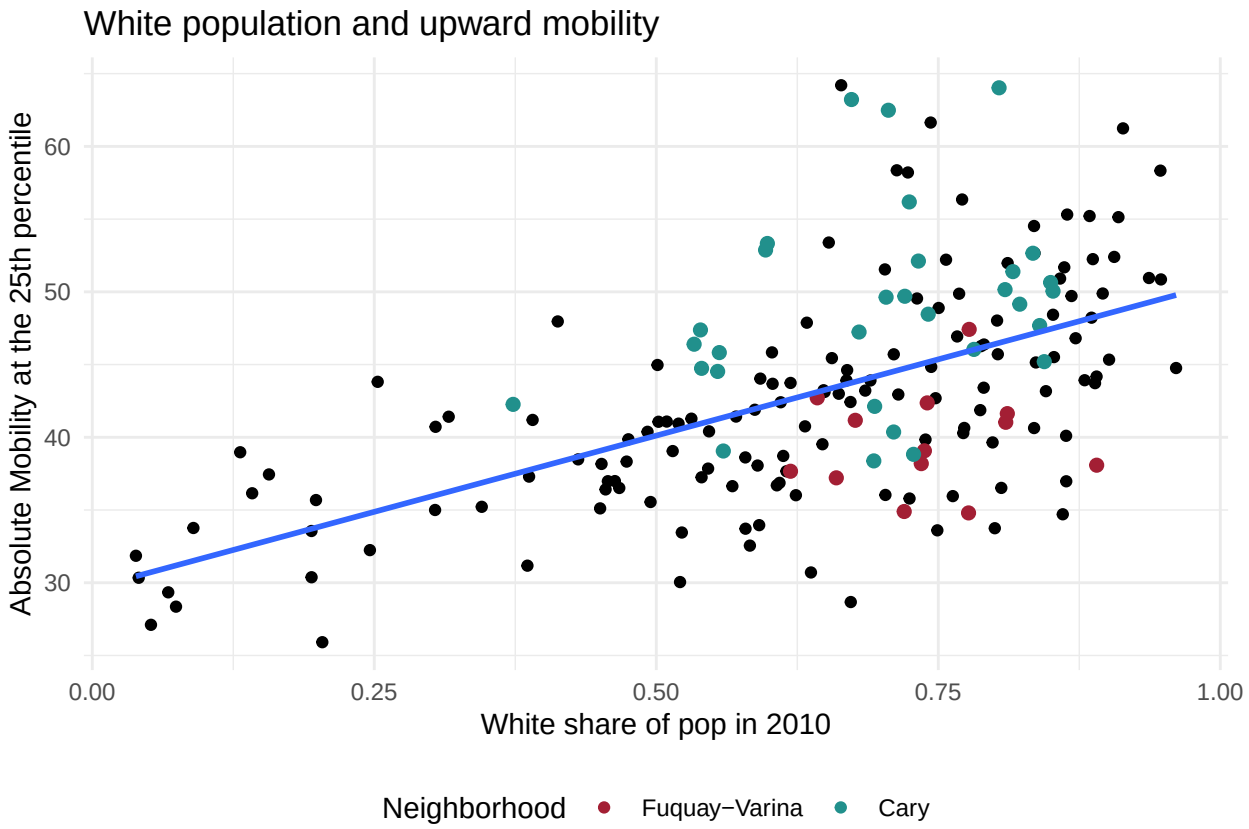
```
#Descriptive relationships between present-day statistics and upward mobility in Wake County
#Creating custom columns for graphs
fuquay <- fuquay %>% mutate(neighborhood = "A")
cary <- cary %>% mutate(neighborhood = "B")

#Combine the datasets
combined_data <- rbind(fuquay, cary)
view(combined_data)

#White share of population
white <- wake_co |>
        ggplot(aes(x = share_white2010, y = kfr_pooled_pooled_p25)) +
           geom_point(aes(x = share_white2010, y = kfr_pooled_pooled_p25)) +
           geom_point(data = combined_data,
             aes(color = neighborhood)) +
          geom_point(data = fuquay, aes(x = share_white2010, y = kfr_pooled_pooled_p25),
                     color = "#A31F34", size = 2) +
           geom_point(data = cary, aes(x = share_white2010, y = kfr_pooled_pooled_p25),
                      color = "#21908CFF", size = 2) +
          scale_color_manual(name = "Neighborhood",
                     values = c("A" = "#A31F34", "B" =  "#21908CFF"),
                     labels = c("Fuquay-Varina", "Cary")) +
           geom_smooth(method = "lm", se = F) +
           labs(x = "White share of pop in 2010",
                y = "Absolute Mobility at the 25th percentile",
                title = "White population and upward mobility") +
          theme_minimal() +
          theme(legend.position = "bottom")
white
```

```
## `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()`).
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```




```
ggsave("currentwhite.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()`).
```

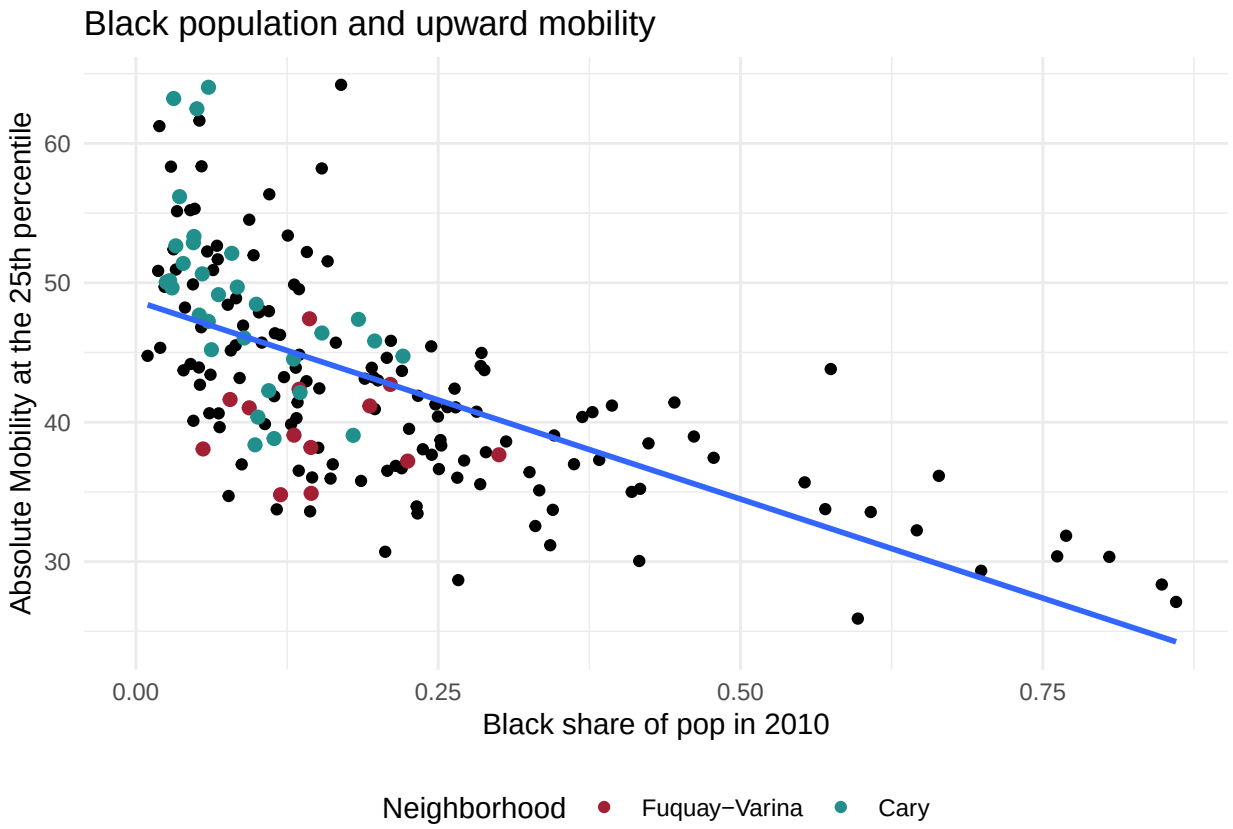
```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```

```
#Black share of population
black <- wake_co |>
        ggplot(aes(x = share_black2010, y = kfr_pooled_pooled_p25)) +
           geom_point(aes(x = share_black2010, y = kfr_pooled_pooled_p25)) +
    geom_point(data = combined_data,
               aes(color = neighborhood)) +
          geom_point(data = fuquay, aes(x = share_black2010, y = kfr_pooled_pooled_p25),
                     color = "#A31F34", size = 2) +
           geom_point(data = cary, aes(x = share_black2010, y = kfr_pooled_pooled_p25),
                     color = "#21908CFF", size = 2) +
          scale_color_manual(name = "Neighborhood",
                     values = c("A" = "#A31F34", "B" =  "#21908CFF"),
                     labels = c("Fuquay-Varina", "Cary")) +
           geom_smooth(method = "lm", se = F) +
           labs(x = "Black share of pop in 2010",
                y = "Absolute Mobility at the 25th percentile",
                title = "Black population and upward mobility") +
          theme_minimal() +
```

```
          theme(legend.position = "bottom")
black
```

```
## `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()`).
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```




```
ggsave("currentblack.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()`).
## Warning: Removed 2 rows containing missing values (`geom_point()`).
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```

```
#Hispanic share of population
hisp <- wake_co |>
        ggplot(aes(x = share_hisp2010, y = kfr_pooled_pooled_p25)) +
           geom_point(aes(x = share_hisp2010, y = kfr_pooled_pooled_p25)) +
          geom_point(data = combined_data,
             aes(color = neighborhood)) +
          geom_point(data = fuquay, aes(x = share_hisp2010, y = kfr_pooled_pooled_p25),
                     color = "#A31F34", size = 2) +
```

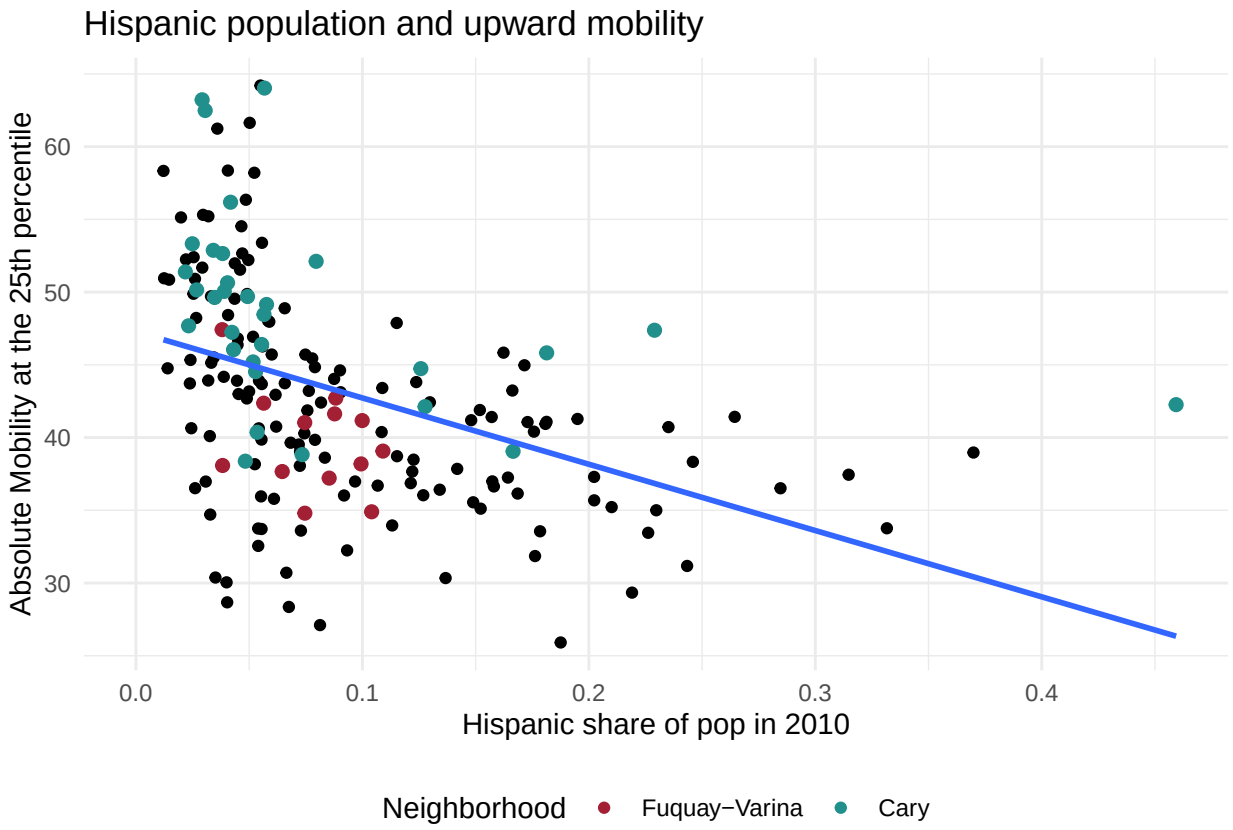
```
            geom_point(data = cary, aes(x = share_hisp2010, y = kfr_pooled_pooled_p25),
                       color = "#21908CFF", size = 2) +
          scale_color_manual(name = "Neighborhood",
                             values = c("A" = "#A31F34", "B" =  "#21908CFF"),
                             labels = c("Fuquay-Varina", "Cary")) +
           geom_smooth(method = "lm", se = F) +
           labs(x = "Hispanic share of pop in 2010",
                y = "Absolute Mobility at the 25th percentile",
                title = "Hispanic population and upward mobility") +
          theme_minimal() +
          theme(legend.position = "bottom")
hisp
```

```
## `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()`).
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```




```
ggsave("currenthisp.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()`).
```

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

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```

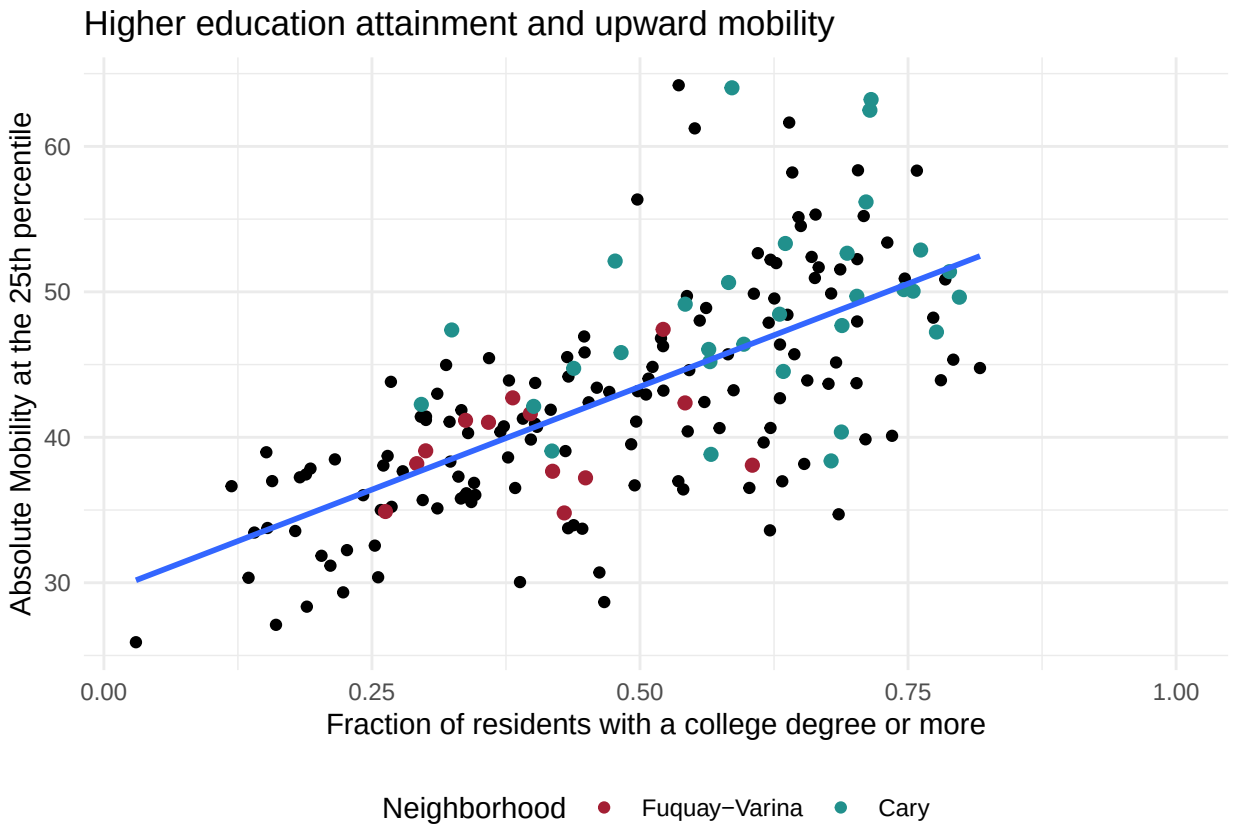
```
#College attainment
college <- wake_co |>
        ggplot(aes(x = frac_coll_plus2010, y = kfr_pooled_pooled_p25)) +
           geom_point(aes(x = frac_coll_plus2010, y = kfr_pooled_pooled_p25)) +
    geom_point(data = combined_data,
             aes(color = neighborhood)) +
          geom_point(data = fuquay, aes(x = frac_coll_plus2010, y = kfr_pooled_pooled_p25),
                     color = "#A31F34", size = 2) +
           geom_point(data = cary, aes(x = frac_coll_plus2010, y = kfr_pooled_pooled_p25),
                      color = "#21908CFF", size = 2) +
          scale_color_manual(name = "Neighborhood",
                     values = c("A" = "#A31F34", "B" =  "#21908CFF"),
                     labels = c("Fuquay-Varina", "Cary")) +
           geom_smooth(method = "lm", se = F) +
           labs(x = "Fraction of residents with a college degree or more",
                y = "Absolute Mobility at the 25th percentile",
                title = "Higher education attainment and upward mobility") +
          theme_minimal() +
          theme(legend.position = "bottom")
college
```

```
## `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()`).
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```



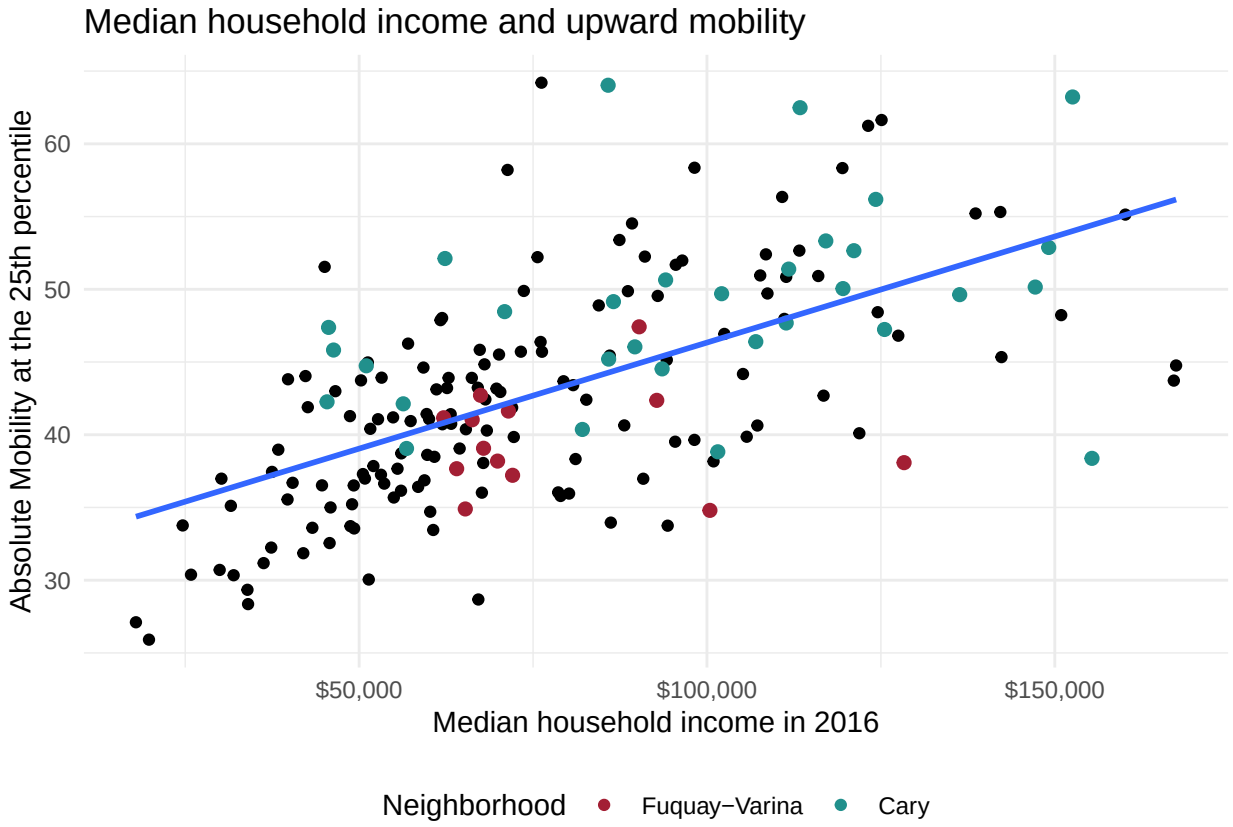

```
ggsave("currentcol.png")
```

```
## Saving 6.5 x 4.5 in image
## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).
## Warning: Removed 2 rows containing missing values (`geom_point()`).
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```

```
#Median household income
medincome <- wake_co |>
        ggplot(aes(x = med_hhinc2016, y = kfr_pooled_pooled_p25)) +
           geom_point(aes(x = med_hhinc2016, y = kfr_pooled_pooled_p25)) +
           geom_point(data = combined_data,
             aes(color = neighborhood)) +
          geom_point(data = fuquay, aes(x = med_hhinc2016, y = kfr_pooled_pooled_p25),
                     color = "#A31F34", size = 2) +
           geom_point(data = cary, aes(x = med_hhinc2016, y = kfr_pooled_pooled_p25),
                     color = "#21908CFF", size = 2) +
          scale_color_manual(name = "Neighborhood",
                     values = c("A" = "#A31F34", "B" =  "#21908CFF"),
                     labels = c("Fuquay-Varina", "Cary")) +
           geom_smooth(method = "lm", se = F) +
           labs(x = "Median household income in 2016",
                y = "Absolute Mobility at the 25th percentile",
                title = "Median household income and upward mobility") +
          scale_x_continuous(labels = scales::dollar_format()) +
```

```
          theme_minimal() +
          theme(legend.position = "bottom")
medincome
```

```
## `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()`).
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```



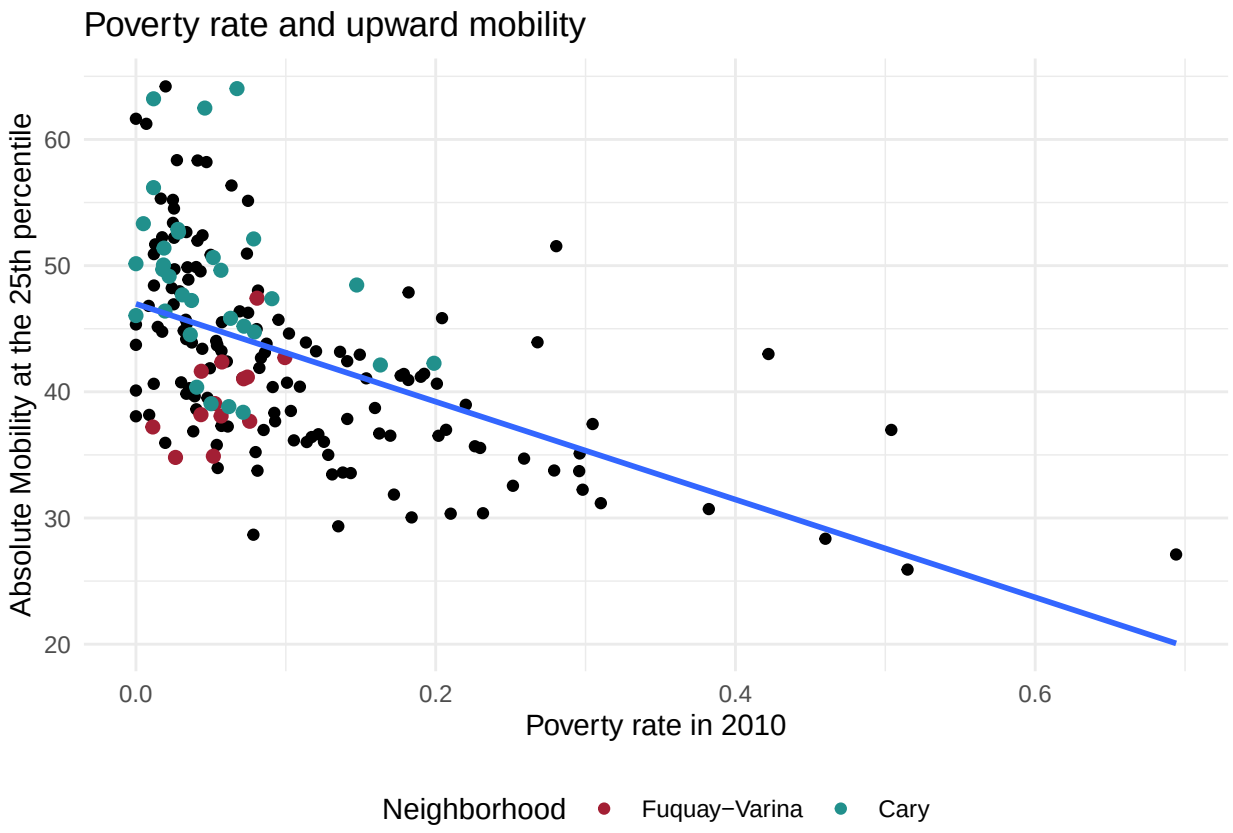

```
ggsave("currentmedinc.png")
```

```
## Saving 6.5 x 4.5 in image
## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).
## Warning: Removed 2 rows containing missing values (`geom_point()`).
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```

```
#Poverty rate
poverty <- wake_co |>
        ggplot(aes(x = poor_share2010, y = kfr_pooled_pooled_p25)) +
           geom_point(aes(x = poor_share2010, y = kfr_pooled_pooled_p25)) +
           geom_point(data = combined_data,
             aes(color = neighborhood)) +
          geom_point(data = fuquay, aes(x = poor_share2010, y = kfr_pooled_pooled_p25),
```

```
                    color = "#A31F34", size = 2) +
          geom_point(data = cary, aes(x = poor_share2010, y = kfr_pooled_pooled_p25),
                     color = "#21908CFF", size = 2) +
          scale_color_manual(name = "Neighborhood",
                     values = c("A" = "#A31F34", "B" =  "#21908CFF"),
                     labels = c("Fuquay-Varina", "Cary")) +
          geom_smooth(method = "lm", se = F) +
          labs(x = "Poverty rate in 2010",
               y = "Absolute Mobility at the 25th percentile",
               title = "Poverty rate and upward mobility") +
          theme_minimal() +
          theme(legend.position = "bottom")
poverty
```

```
## `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()`).
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```




```
ggsave("currentpoverty.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()`).
```

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

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```

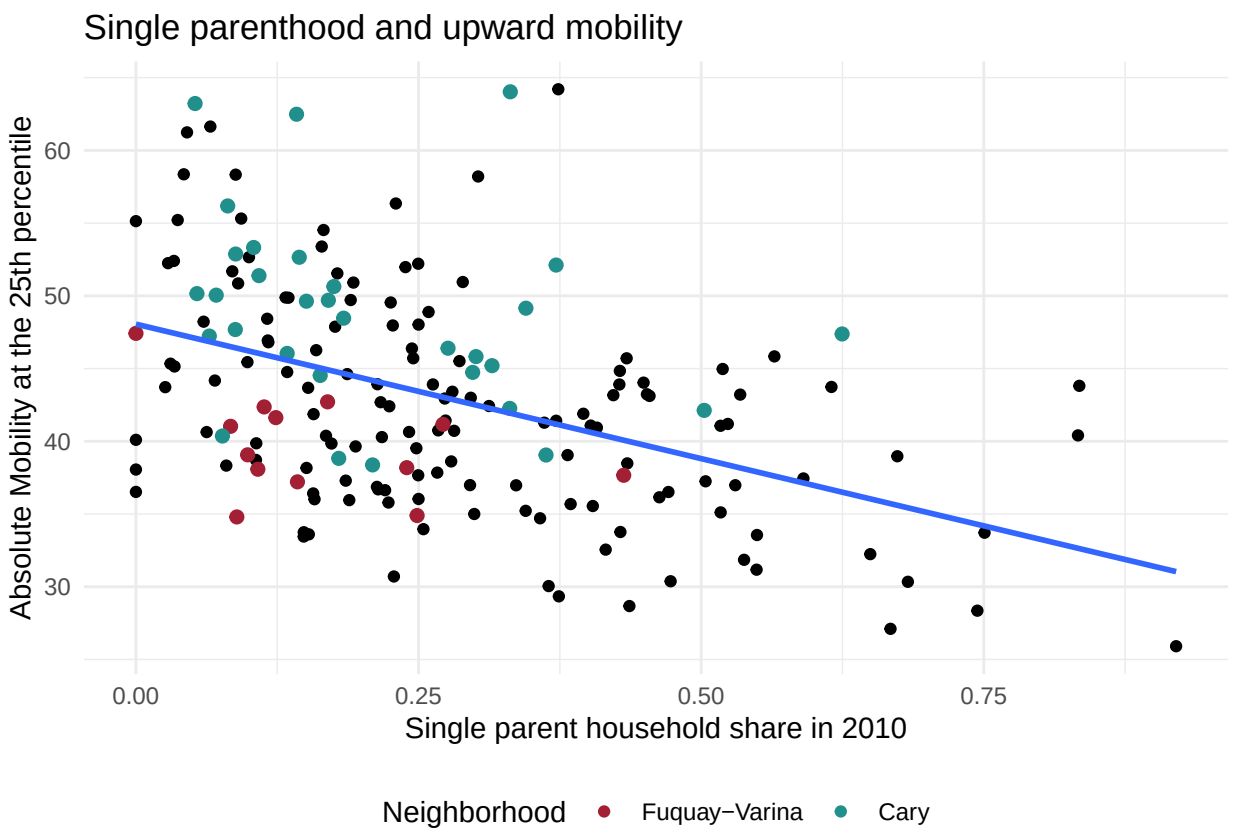
```
#Single parent share
singleparent <- wake_co |>
        ggplot(aes(x = singleparent_share2010, y = kfr_pooled_pooled_p25)) +
           geom_point(aes(x = singleparent_share2010, y = kfr_pooled_pooled_p25)) +
           geom_point(data = combined_data,
             aes(color = neighborhood)) +
          geom_point(data = fuquay, aes(x = singleparent_share2010, y = kfr_pooled_pooled_p25),
                      color = "#A31F34", size = 2) +
           geom_point(data = cary, aes(x = singleparent_share2010, y = kfr_pooled_pooled_p25),
                      color = "#21908CFF", size = 2) +
          scale_color_manual(name = "Neighborhood",
                     values = c("A" = "#A31F34", "B" =  "#21908CFF"),
                     labels = c("Fuquay-Varina", "Cary")) +
           geom_smooth(method = "lm", se = F) +
           labs(x = "Single parent household share in 2010",
                y = "Absolute Mobility at the 25th percentile",
                title = "Single parenthood and upward mobility") +
          theme_minimal() +
          theme(legend.position = "bottom")
singleparent
```

```
## `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()`).
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```




```
ggsave("currentsingle.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()`).
```

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

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```

```
#Incarceration Rate for low-income children
jailall <- wake_co |>
        ggplot(aes(x = jail_pooled_pooled_p25, y = kfr_pooled_pooled_p25)) +
           geom_point(aes(x = jail_pooled_pooled_p25, y = kfr_pooled_pooled_p25)) +
           geom_point(data = combined_data,
             aes(color = neighborhood)) +
          geom_point(data = fuquay, aes(x = jail_pooled_pooled_p25, y = kfr_pooled_pooled_p25),
                     color = "#A31F34", size = 2) +
           geom_point(data = cary, aes(x = jail_pooled_pooled_p25, y = kfr_pooled_pooled_p25),
                     color = "#21908CFF", size = 2) +
          scale_color_manual(name = "Neighborhood",
                     values = c("A" = "#A31F34", "B" =  "#21908CFF"),
                     labels = c("Fuquay-Varina", "Cary")) +
           geom_smooth(method = "lm", se = F) +
           labs(x = "Share of low-income children incarcerated in 2010",
                y = "Absolute Mobility at the 25th percentile",
                title = "Incarceration rate and upward mobility") +
          theme_minimal() +
```

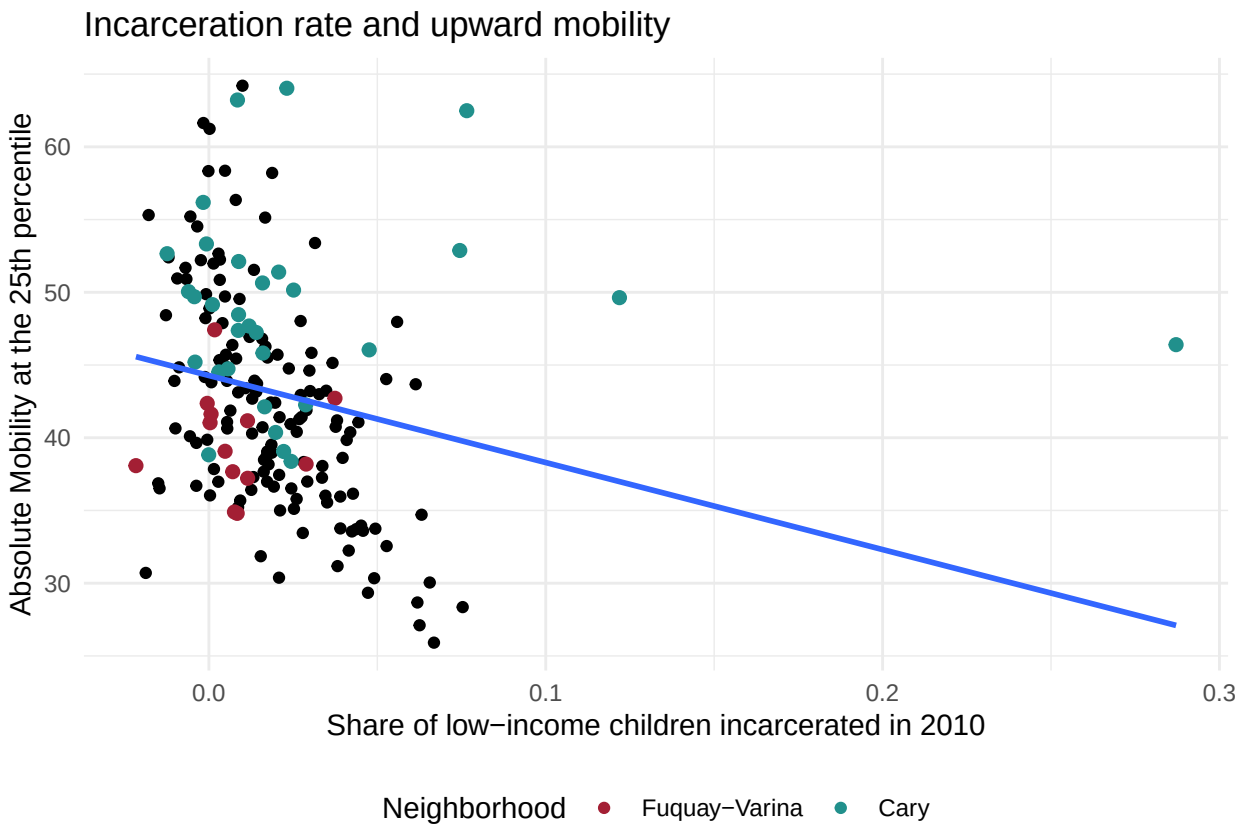
```
          theme(legend.position = "bottom")
jailall
```

```
## `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()`).
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```




```
ggsave("currentjail.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()`).
```

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

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```

```
#Correlation coefficient table between present-day statistics and upward mobility

#Derive correlation coefficients
white_wake <- cor(wake_co$kfr_pooled_pooled_p25, wake_co$share_white2010, use = "complete.obs")
white_fuquay <- cor(fuquay$kfr_pooled_pooled_p25, fuquay$share_white2010, use = "complete.obs")
white_cary <- cor(cary$kfr_pooled_pooled_p25, cary$share_white2010, use = "complete.obs")
```

```r
black_wake <- cor(wake_co$kfr_pooled_pooled_p25, wake_co$share_black2010, use = "complete.obs")
black_fuquay <- cor(fuquay$kfr_pooled_pooled_p25, fuquay$share_black2010, use = "complete.obs")
black_cary <- cor(cary$kfr_pooled_pooled_p25, cary$share_black2010, use = "complete.obs")

hisp_wake <- cor(wake_co$kfr_pooled_pooled_p25, wake_co$share_hisp2010, use = "complete.obs")
hisp_fuquay <- cor(fuquay$kfr_pooled_pooled_p25, fuquay$share_hisp2010, use = "complete.obs")
hisp_cary <- cor(cary$kfr_pooled_pooled_p25, cary$share_hisp2010, use = "complete.obs")

col_wake <- cor(wake_co$kfr_pooled_pooled_p25, wake_co$frac_coll_plus2010, use = "complete.obs")
col_fuquay <- cor(fuquay$kfr_pooled_pooled_p25, fuquay$frac_coll_plus2010, use = "complete.obs")
col_cary <- cor(cary$kfr_pooled_pooled_p25, cary$frac_coll_plus2010, use = "complete.obs")

medinc_wake <- cor(wake_co$kfr_pooled_pooled_p25, wake_co$med_hhinc2016, use = "complete.obs")
medinc_fuquay <- cor(fuquay$kfr_pooled_pooled_p25, fuquay$med_hhinc2016, use = "complete.obs")
medinc_cary <- cor(cary$kfr_pooled_pooled_p25, cary$med_hhinc2016, use = "complete.obs")

poor_wake <- cor(wake_co$kfr_pooled_pooled_p25, wake_co$poor_share2010, use = "complete.obs")
poor_fuquay <- cor(fuquay$kfr_pooled_pooled_p25, fuquay$poor_share2010, use = "complete.obs")
poor_cary <- cor(cary$kfr_pooled_pooled_p25, cary$poor_share2010, use = "complete.obs")

single_wake <- cor(wake_co$kfr_pooled_pooled_p25, wake_co$singleparent_share2010, use = "complete.obs")
single_fuquay <- cor(fuquay$kfr_pooled_pooled_p25, fuquay$singleparent_share2010, use = "complete.obs")
single_cary <- cor(cary$kfr_pooled_pooled_p25, cary$singleparent_share2010, use = "complete.obs")

jail_wake <- cor(wake_co$kfr_pooled_pooled_p25, wake_co$jail_pooled_pooled_p25, use = "complete.obs")
jail_fuquay <- cor(fuquay$kfr_pooled_pooled_p25, fuquay$jail_pooled_pooled_p25, use = "complete.obs")
jail_cary <- cor(cary$kfr_pooled_pooled_p25, cary$jail_pooled_pooled_p25, use = "complete.obs")

#Create correlation coefficient dataframe
coefcomp <- data.frame(c("White Population Share", "White Population Share", "White Population Share",
                         "Black Population Share", "Black Population Share", "Black Population Share",
                         "Hispanic Population Share", "Hispanic Population Share", "Hispanic Population
                         "College Educated Population", "College Educated Population", "College Educate
                         "Median Household Income", "Median Household Income", "Median Household Income
                         "Poverty Rate", "Poverty Rate", "Poverty Rate",
                         "Single Parent Share", "Single Parent Share", "Single Parent Share",
                         "Low-Income Child Incarceration Rate", "Low-Income Child Incarceration Rate",
                       c("Wake County", "Fuquay-Varina", "Cary",
                         "Wake County", "Fuquay-Varina", "Cary",
                         "Wake County", "Fuquay-Varina", "Cary",
                         "Wake County", "Fuquay-Varina", "Cary",
                         "Wake County", "Fuquay-Varina", "Cary",
                         "Wake County", "Fuquay-Varina", "Cary",
                         "Wake County", "Fuquay-Varina", "Cary",
                         "Wake County", "Fuquay-Varina", "Cary"),
                       c(white_wake , white_fuquay, white_cary,
                         black_wake, black_fuquay, black_cary,
                         hisp_wake, hisp_fuquay, hisp_cary,
                         col_wake, col_fuquay, col_cary,
                         medinc_wake, medinc_fuquay, medinc_cary,
                         poor_wake, poor_fuquay, poor_cary,
                         single_wake, single_fuquay, single_cary,
                         jail_wake, jail_fuquay, jail_cary))
```

```r
#Name columns
names(coefcomp)[1] <- "Covariate"
names(coefcomp)[2] <- "Region"
names(coefcomp)[3] <- "Correlation Coefficient"

#check
coefcomp
```

```
##                             Covariate        Region Correlation Coefficient
## 1               White Population Share   Wake County               0.57118658
## 2               White Population Share Fuquay-Varina               0.08165056
## 3               White Population Share          Cary               0.28912931
## 4               Black Population Share   Wake County              -0.64482230
## 5               Black Population Share Fuquay-Varina              -0.07358894
## 6               Black Population Share          Cary              -0.57115137
## 7            Hispanic Population Share   Wake County              -0.43211550
## 8            Hispanic Population Share Fuquay-Varina              -0.38032489
## 9            Hispanic Population Share          Cary              -0.36720634
## 10         College Educated Population   Wake County               0.66764405
## 11         College Educated Population Fuquay-Varina               0.31632545
## 12         College Educated Population          Cary               0.39585855
## 13             Median Household Income   Wake County               0.60792816
## 14             Median Household Income Fuquay-Varina              -0.02183167
## 15             Median Household Income          Cary               0.36477098
## 16                        Poverty Rate   Wake County              -0.53150295
## 17                        Poverty Rate Fuquay-Varina               0.61358028
## 18                        Poverty Rate          Cary              -0.33944326
## 19                Single Parent Share   Wake County              -0.44495672
## 20                Single Parent Share Fuquay-Varina              -0.41541650
## 21                Single Parent Share          Cary              -0.27249873
## 22 Low-Income Child Incarceration Rate   Wake County              -0.22661839
## 23 Low-Income Child Incarceration Rate Fuquay-Varina               0.01674997
## 24 Low-Income Child Incarceration Rate          Cary               0.01304451
```

```r
#create table
table_presentday <- coefcomp |>
                    kable("html",
                          caption = "Correlations Between Present Day Demographic Variables and Upward M
                          align = "lcr",
                          col.names = c("Covariate", "Region", "Correlation Coefficient")) |>
                    kable_styling("striped",  # Table style: "striped", "bordered", "hover", "condensed
                   full_width = FALSE) |>
                   save_kable(file = "presentday.png",
                              zoom = 1)
#check
table_presentday
```

```
## [1] "/Users/Jose/Documents/R/EC50/Project1/presentday.png"
## attr(,"info")
## # A tibble: 1 x 7
##   format width height colorspace matte filesize density
##   <chr>  <int>  <int> <chr>      <lgl>    <int> <chr>
## 1 PNG      506    890 sRGB       TRUE         0 28x28
```

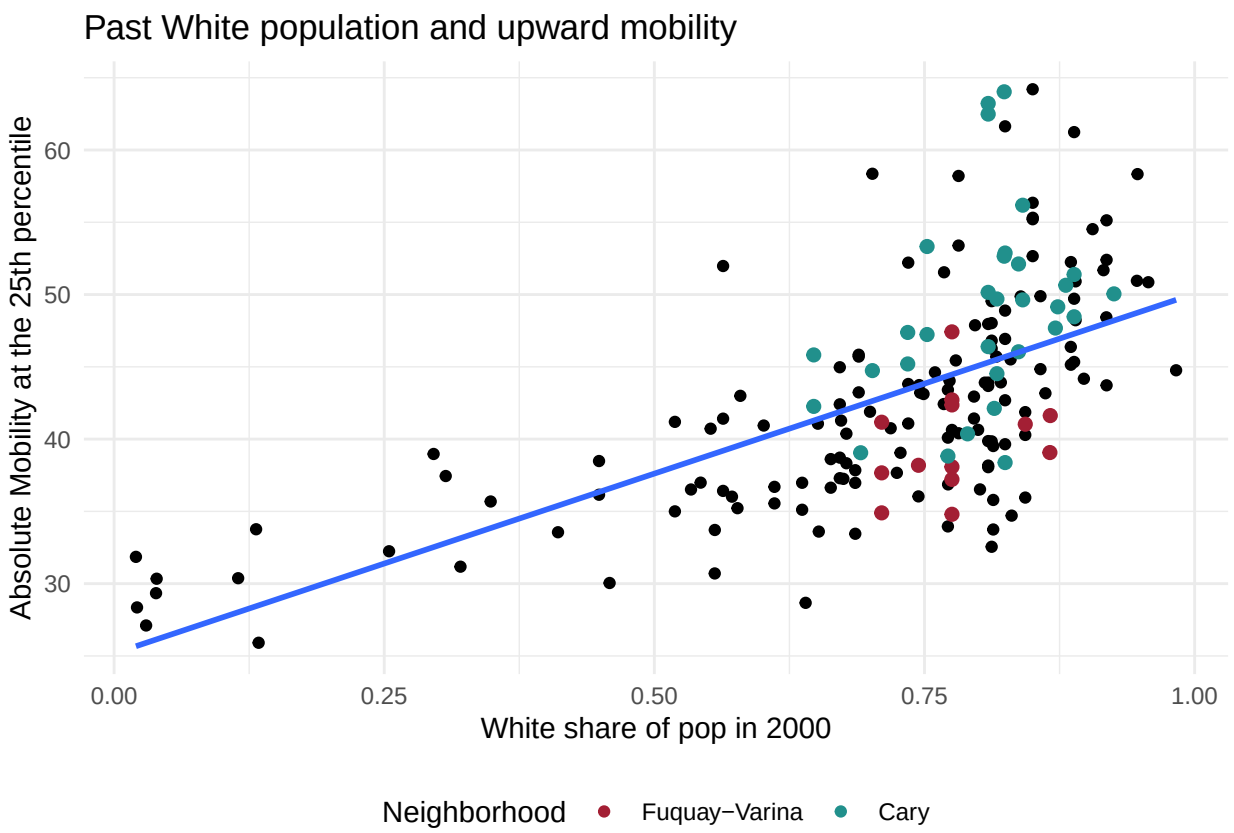
```r
#Historical covariates and their causal relationship to upward mobility

#White share of population
white2000 <- wake_co |>
        ggplot(aes(x = share_white2000, y = kfr_pooled_pooled_p25)) +
           geom_point(aes(x = share_white2000, y = kfr_pooled_pooled_p25)) +
           geom_point(data = combined_data,
             aes(color = neighborhood)) +
          geom_point(data = fuquay, aes(x = share_white2000, y = kfr_pooled_pooled_p25),
                     color = "#A31F34", size = 2) +
           geom_point(data = cary, aes(x = share_white2000, y = kfr_pooled_pooled_p25),
                      color = "#21908CFF", size = 2) +
          scale_color_manual(name = "Neighborhood",
                     values = c("A" = "#A31F34", "B" =  "#21908CFF"),
                     labels = c("Fuquay-Varina", "Cary")) +
           geom_smooth(method = "lm", se = F) +
           labs(x = "White share of pop in 2000",
                y = "Absolute Mobility at the 25th percentile",
                title = "Past White population and upward mobility") +
          theme_minimal() +
          theme(legend.position = "bottom")
white2000
```

```
## `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()`).
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```




```
ggsave("2000white.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()`).
```

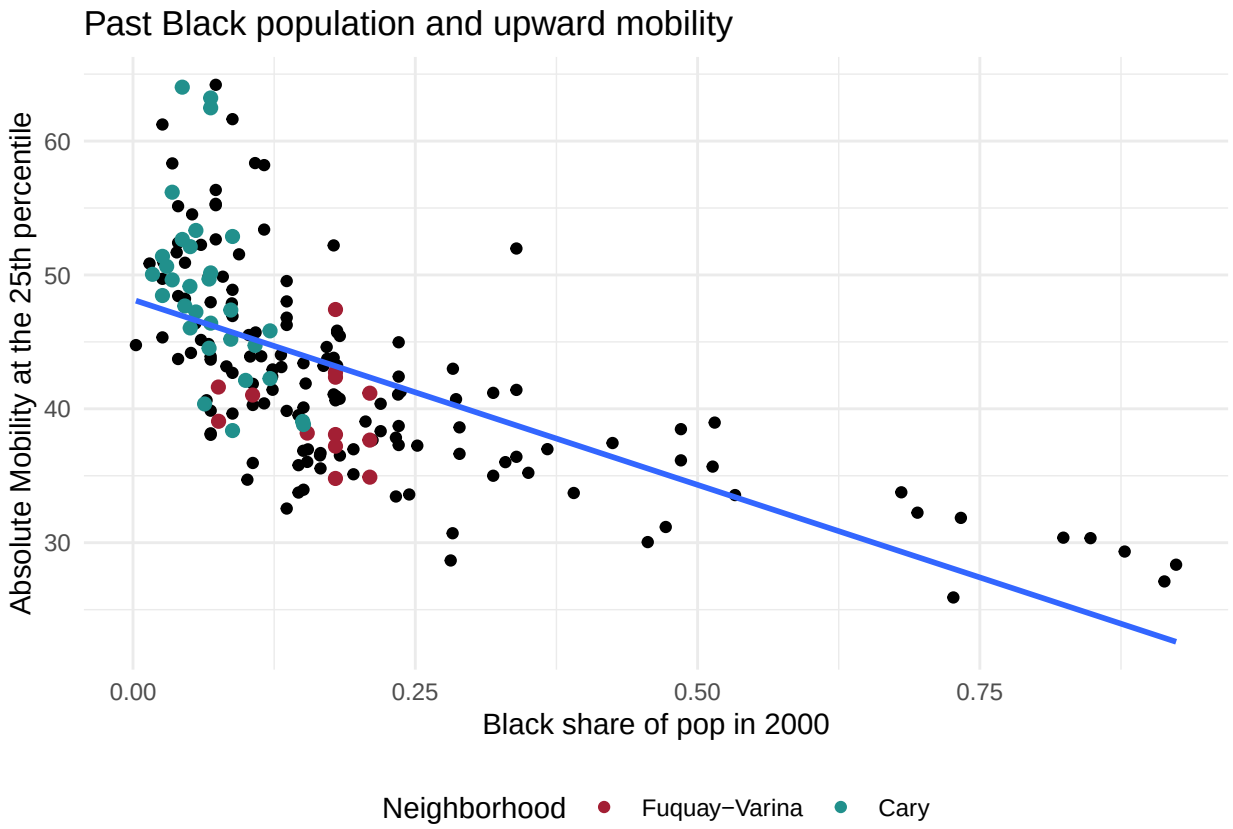
```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```

```
#Black share of population
black2000 <- wake_co |>
        ggplot(aes(x = share_black2000, y = kfr_pooled_pooled_p25)) +
           geom_point(aes(x = share_black2000, y = kfr_pooled_pooled_p25)) +
    geom_point(data = combined_data,
               aes(color = neighborhood)) +
          geom_point(data = fuquay, aes(x = share_black2000, y = kfr_pooled_pooled_p25),
                     color = "#A31F34", size = 2) +
           geom_point(data = cary, aes(x = share_black2000, y = kfr_pooled_pooled_p25),
                     color = "#21908CFF", size = 2) +
          scale_color_manual(name = "Neighborhood",
                     values = c("A" = "#A31F34", "B" =  "#21908CFF"),
                     labels = c("Fuquay-Varina", "Cary")) +
           geom_smooth(method = "lm", se = F) +
           labs(x = "Black share of pop in 2000",
                y = "Absolute Mobility at the 25th percentile",
                title = "Past Black population and upward mobility") +
          theme_minimal() +
```

```
          theme(legend.position = "bottom")
black2000
```

```
## `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()`).
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```




```
ggsave("2000black.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()`).
## Warning: Removed 2 rows containing missing values (`geom_point()`).
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```

```
#Hispanic share of population
hisp2000 <- wake_co |>
        ggplot(aes(x = share_hisp2000, y = kfr_pooled_pooled_p25)) +
           geom_point(aes(x = share_hisp2000, y = kfr_pooled_pooled_p25)) +
          geom_point(data = combined_data,
             aes(color = neighborhood)) +
          geom_point(data = fuquay, aes(x = share_hisp2000, y = kfr_pooled_pooled_p25),
                     color = "#A31F34", size = 2) +
```

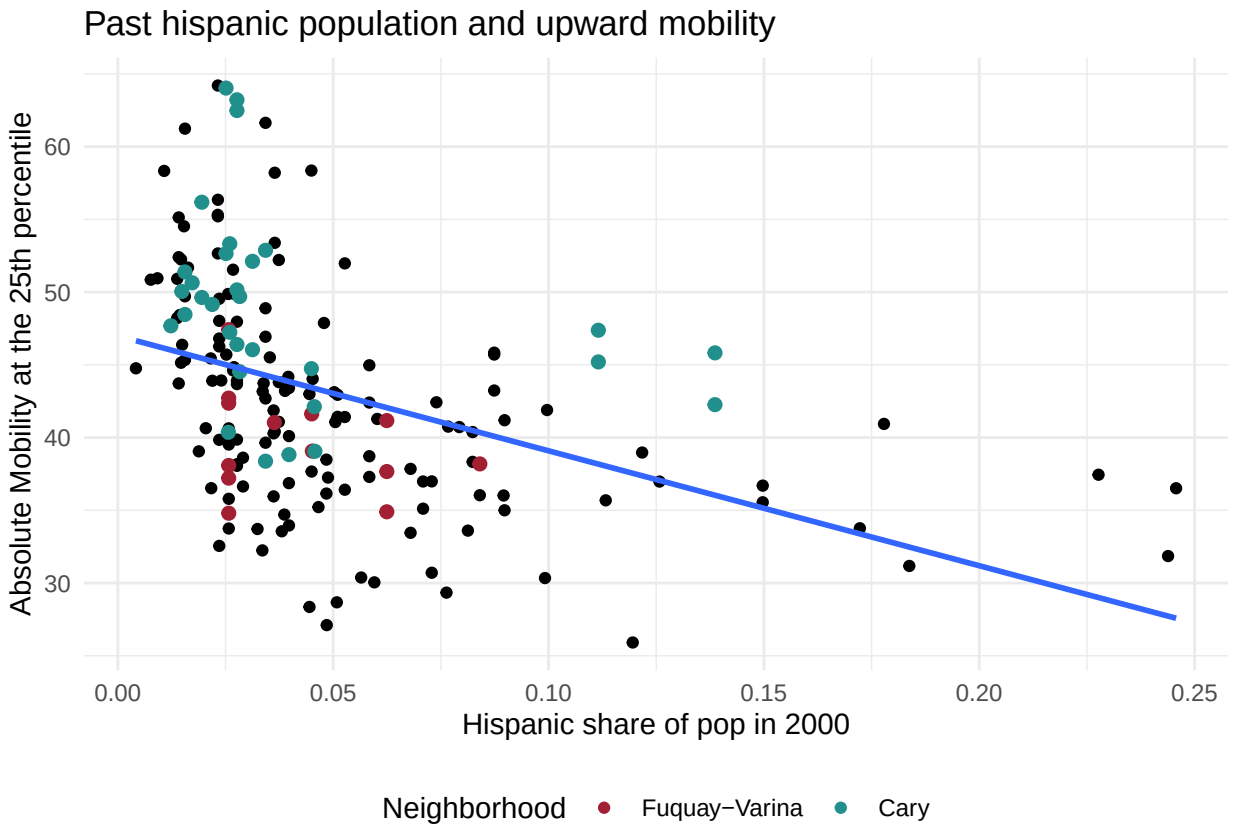
```
           geom_point(data = cary, aes(x = share_hisp2000, y = kfr_pooled_pooled_p25),
                      color = "#21908CFF", size = 2) +
          scale_color_manual(name = "Neighborhood",
                     values = c("A" = "#A31F34", "B" =  "#21908CFF"),
                     labels = c("Fuquay-Varina", "Cary")) +
           geom_smooth(method = "lm", se = F) +
           labs(x = "Hispanic share of pop in 2000",
                y = "Absolute Mobility at the 25th percentile",
                title = "Past hispanic population and upward mobility") +
          theme_minimal() +
          theme(legend.position = "bottom")
hisp2000
```

```
## `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()`).
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```




```
ggsave("2000hisp.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()`).
```

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

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```

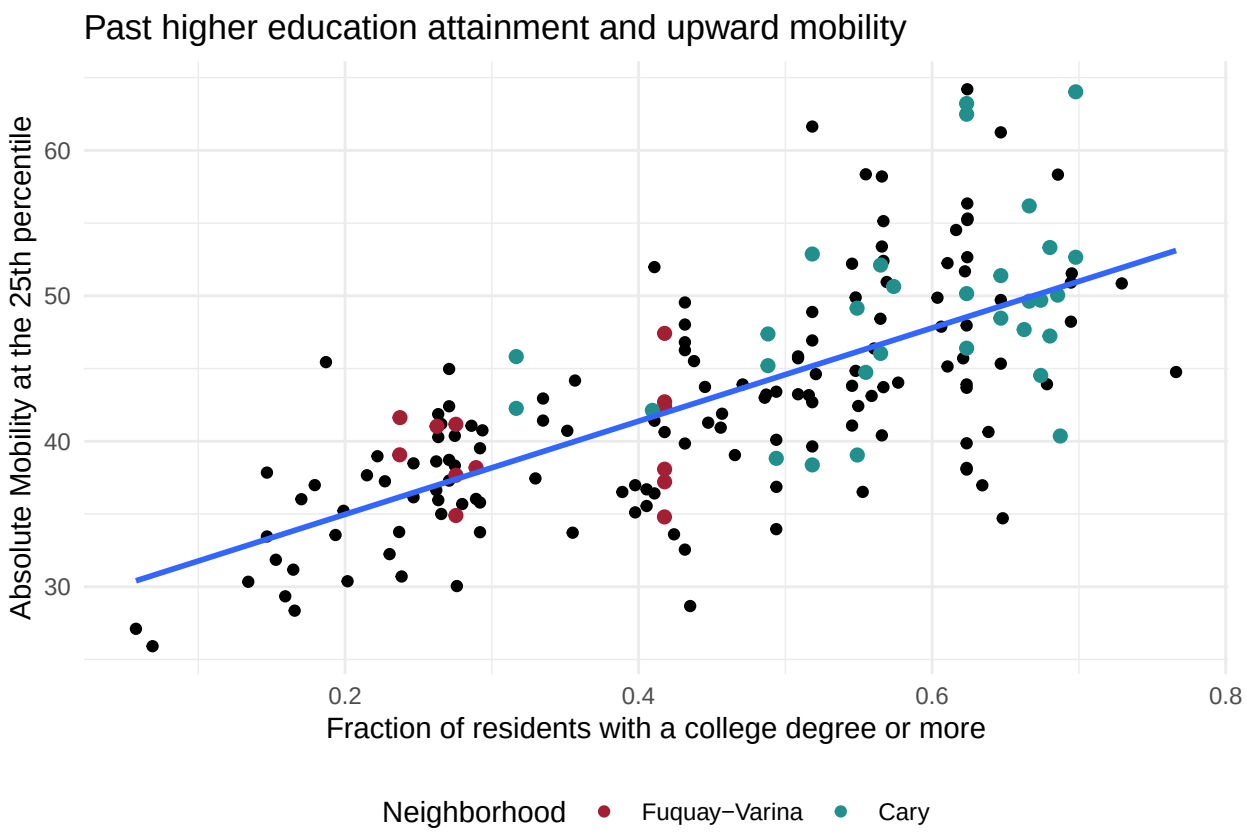
```
#College attainment
college2000 <- wake_co |>
        ggplot(aes(x = frac_coll_plus2000, y = kfr_pooled_pooled_p25)) +
           geom_point(aes(x = frac_coll_plus2000, y = kfr_pooled_pooled_p25)) +
    geom_point(data = combined_data,
              aes(color = neighborhood)) +
          geom_point(data = fuquay, aes(x = frac_coll_plus2000, y = kfr_pooled_pooled_p25),
                     color = "#A31F34", size = 2) +
           geom_point(data = cary, aes(x = frac_coll_plus2000, y = kfr_pooled_pooled_p25),
                      color = "#21908CFF", size = 2) +
          scale_color_manual(name = "Neighborhood",
                     values = c("A" = "#A31F34", "B" =  "#21908CFF"),
                     labels = c("Fuquay-Varina", "Cary")) +
           geom_smooth(method = "lm", se = F) +
           labs(x = "Fraction of residents with a college degree or more",
                y = "Absolute Mobility at the 25th percentile",
                title = "Past higher education attainment and upward mobility") +
          theme_minimal() +
          theme(legend.position = "bottom")
college2000
```

```
## `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()`).
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```



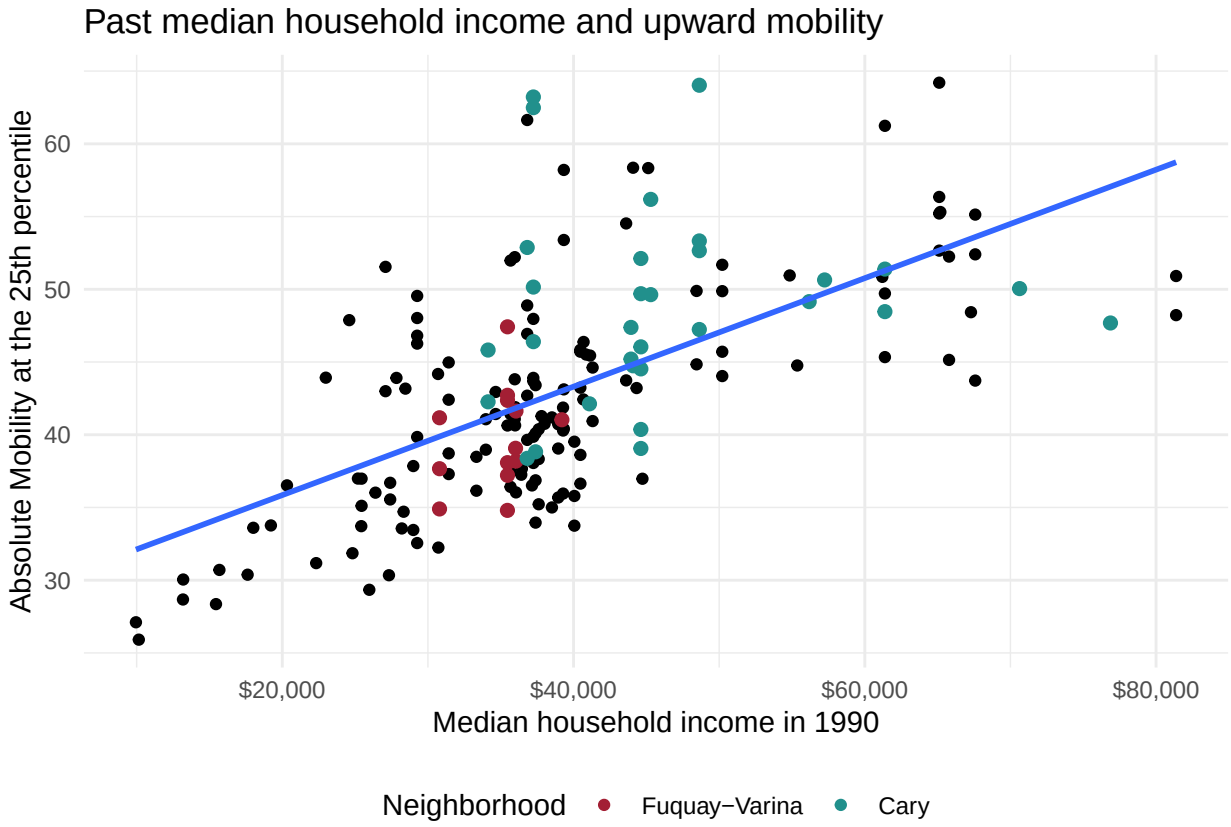

```
ggsave("2000col.png")
```

```
## Saving 6.5 x 4.5 in image
## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 2 rows containing non-finite values (`stat_smooth()`).
## Warning: Removed 2 rows containing missing values (`geom_point()`).
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```

```
#Median household income
med1990 <- wake_co |>
        ggplot(aes(x = med_hhinc1990, y = kfr_pooled_pooled_p25)) +
           geom_point(aes(x = med_hhinc1990, y = kfr_pooled_pooled_p25)) +
           geom_point(data = combined_data,
             aes(color = neighborhood)) +
          geom_point(data = fuquay, aes(x = med_hhinc1990, y = kfr_pooled_pooled_p25),
                     color = "#A31F34", size = 2) +
           geom_point(data = cary, aes(x = med_hhinc1990, y = kfr_pooled_pooled_p25),
                     color = "#21908CFF", size = 2) +
          scale_color_manual(name = "Neighborhood",
                     values = c("A" = "#A31F34", "B" =  "#21908CFF"),
                     labels = c("Fuquay-Varina", "Cary")) +
           geom_smooth(method = "lm", se = F) +
           labs(x = "Median household income in 1990",
                y = "Absolute Mobility at the 25th percentile",
                title = "Past median household income and upward mobility") +
          scale_x_continuous(labels = scales::dollar_format()) +
```

```
          theme_minimal() +
          theme(legend.position = "bottom")
med1990
```

```
## `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()`).
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```




```
ggsave("1990medinc.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()`).
## Warning: Removed 2 rows containing missing values (`geom_point()`).
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```

```
#Poverty rate
poverty1990 <- wake_co |>
        ggplot(aes(x = poor_share1990, y = kfr_pooled_pooled_p25)) +
           geom_point(aes(x = poor_share1990, y = kfr_pooled_pooled_p25)) +
           geom_point(data = combined_data,
             aes(color = neighborhood)) +
          geom_point(data = fuquay, aes(x = poor_share1990, y = kfr_pooled_pooled_p25),
```

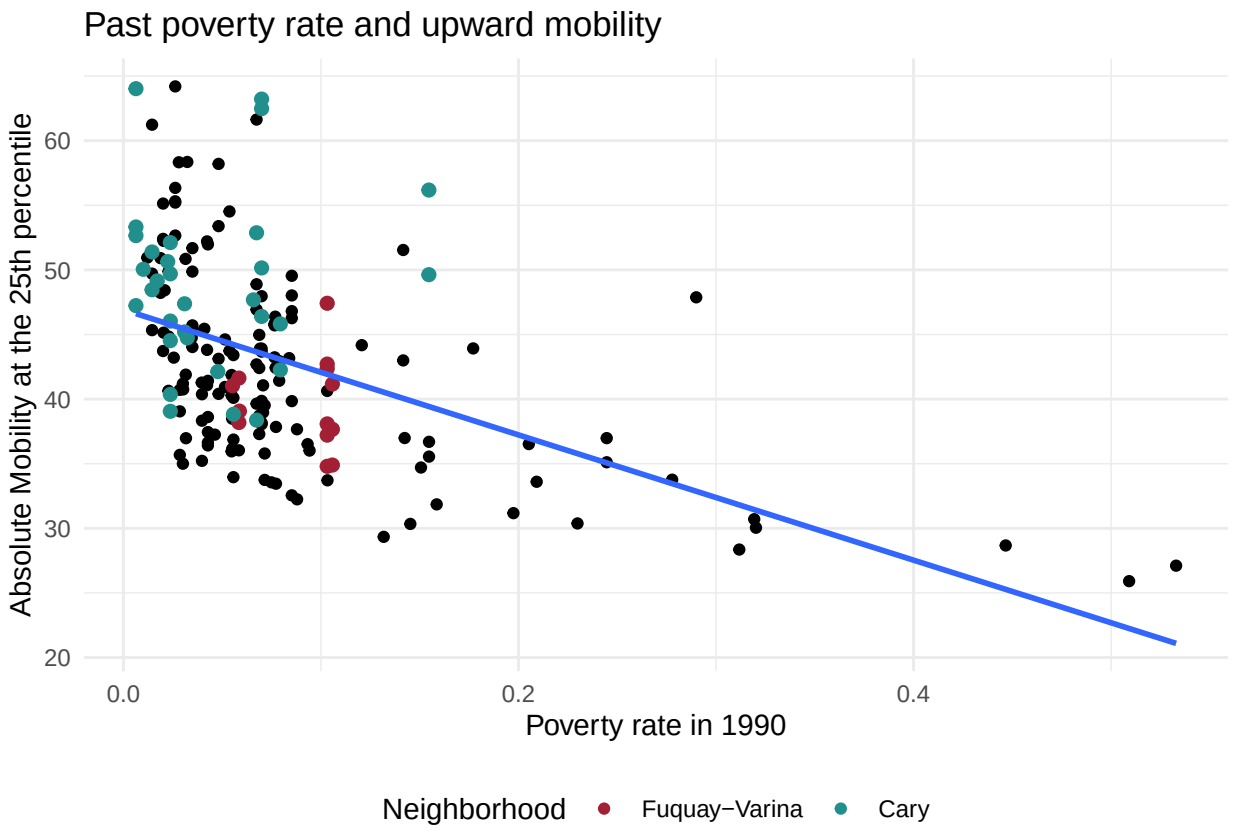
```
                    color = "#A31F34", size = 2) +
          geom_point(data = cary, aes(x = poor_share1990, y = kfr_pooled_pooled_p25),
                     color = "#21908CFF", size = 2) +
          scale_color_manual(name = "Neighborhood",
                     values = c("A" = "#A31F34", "B" =  "#21908CFF"),
                     labels = c("Fuquay-Varina", "Cary")) +
          geom_smooth(method = "lm", se = F) +
          labs(x = "Poverty rate in 1990",
               y = "Absolute Mobility at the 25th percentile",
               title = "Past poverty rate and upward mobility") +
          theme_minimal() +
          theme(legend.position = "bottom")
poverty1990
```

```
## `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()`).
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```




```
ggsave("1990poverty.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()`).
```

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

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```

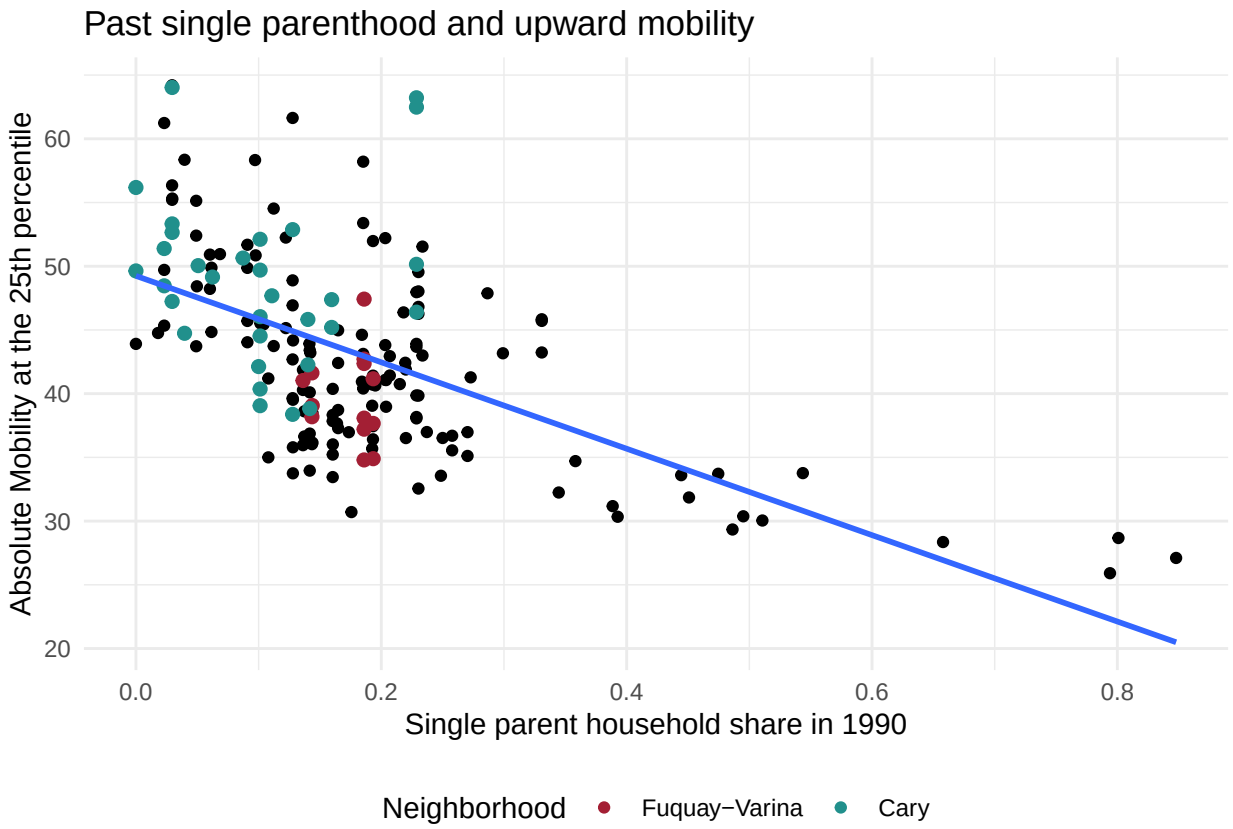
```
#Single parent share
singleparent1990 <- wake_co |>
        ggplot(aes(x = singleparent_share1990, y = kfr_pooled_pooled_p25)) +
           geom_point(aes(x = singleparent_share1990, y = kfr_pooled_pooled_p25)) +
           geom_point(data = combined_data,
             aes(color = neighborhood)) +
          geom_point(data = fuquay, aes(x = singleparent_share1990, y = kfr_pooled_pooled_p25),
                     color = "#A31F34", size = 2) +
           geom_point(data = cary, aes(x = singleparent_share1990, y = kfr_pooled_pooled_p25),
                     color = "#21908CFF", size = 2) +
          scale_color_manual(name = "Neighborhood",
                     values = c("A" = "#A31F34", "B" =  "#21908CFF"),
                     labels = c("Fuquay-Varina", "Cary")) +
           geom_smooth(method = "lm", se = F) +
           labs(x = "Single parent household share in 1990",
                y = "Absolute Mobility at the 25th percentile",
                title = "Past single parenthood and upward mobility") +
          theme_minimal() +
          theme(legend.position = "bottom")
singleparent1990
```

```
## `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()`).
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```




```
ggsave("1990single.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()`).
```

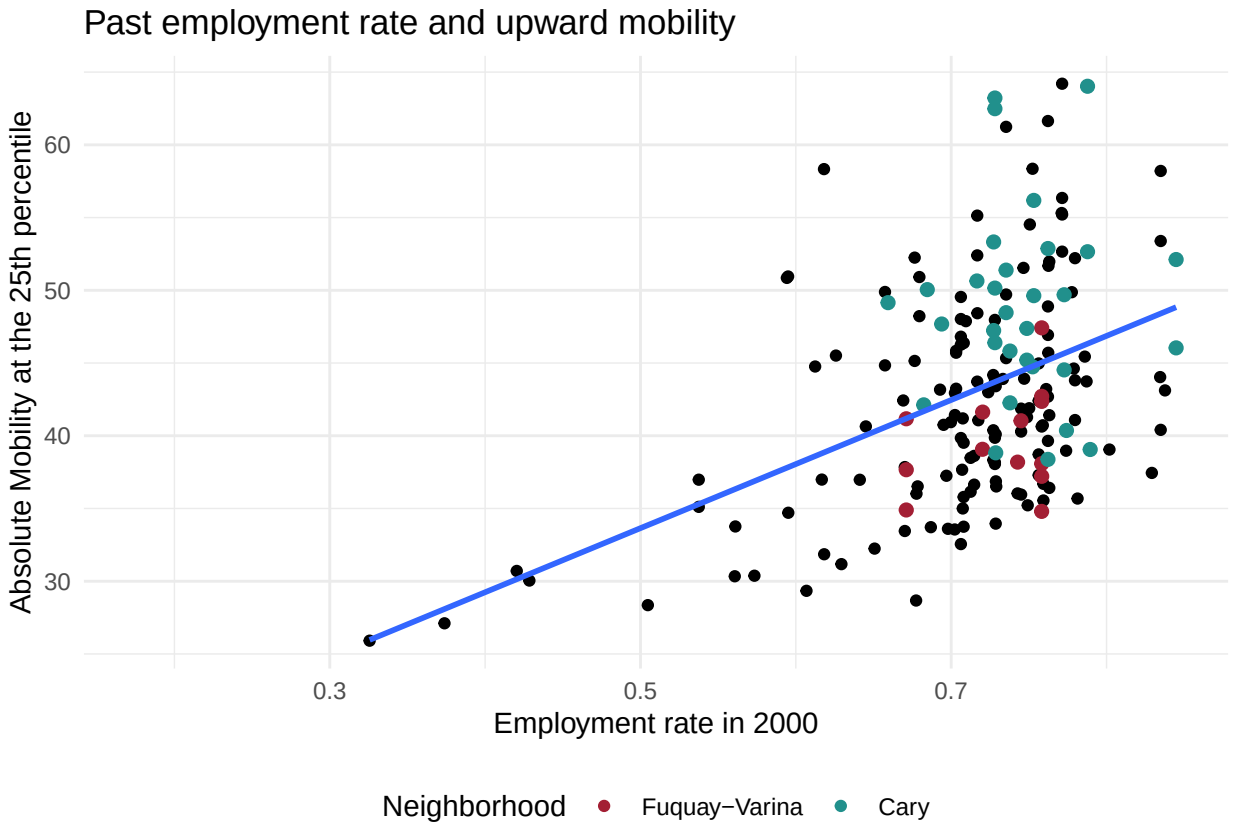
```
## Warning: Removed 2 rows containing missing values (`geom_point()`).
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```

```
#Employment rate
employ2000 <- wake_co |>
        ggplot(aes(x = emp2000, y = kfr_pooled_pooled_p25)) +
           geom_point(aes(x = emp2000, y = kfr_pooled_pooled_p25)) +
           geom_point(data = combined_data,
             aes(color = neighborhood)) +
          geom_point(data = fuquay, aes(x = emp2000, y = kfr_pooled_pooled_p25),
                     color = "#A31F34", size = 2) +
           geom_point(data = cary, aes(x = emp2000, y = kfr_pooled_pooled_p25),
                     color = "#21908CFF", size = 2) +
          scale_color_manual(name = "Neighborhood",
                     values = c("A" = "#A31F34", "B" =  "#21908CFF"),
                     labels = c("Fuquay-Varina", "Cary")) +
           geom_smooth(method = "lm", se = F) +
           labs(x = "Employment rate in 2000",
                y = "Absolute Mobility at the 25th percentile",
                title = "Past employment rate and upward mobility") +
          theme_minimal() +
```

```
          theme(legend.position = "bottom")
employ2000
```

```
## `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()`).
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```




```
ggsave("employ2000.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()`).
## Warning: Removed 2 rows containing missing values (`geom_point()`).
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```

```
#Population density
popden2000 <- wake_co |>
        ggplot(aes(x = popdensity2000, y = kfr_pooled_pooled_p25)) +
           geom_point(aes(x = popdensity2000, y = kfr_pooled_pooled_p25)) +
           geom_point(data = combined_data,
             aes(color = neighborhood)) +
          geom_point(data = fuquay, aes(x = popdensity2000, y = kfr_pooled_pooled_p25),
```

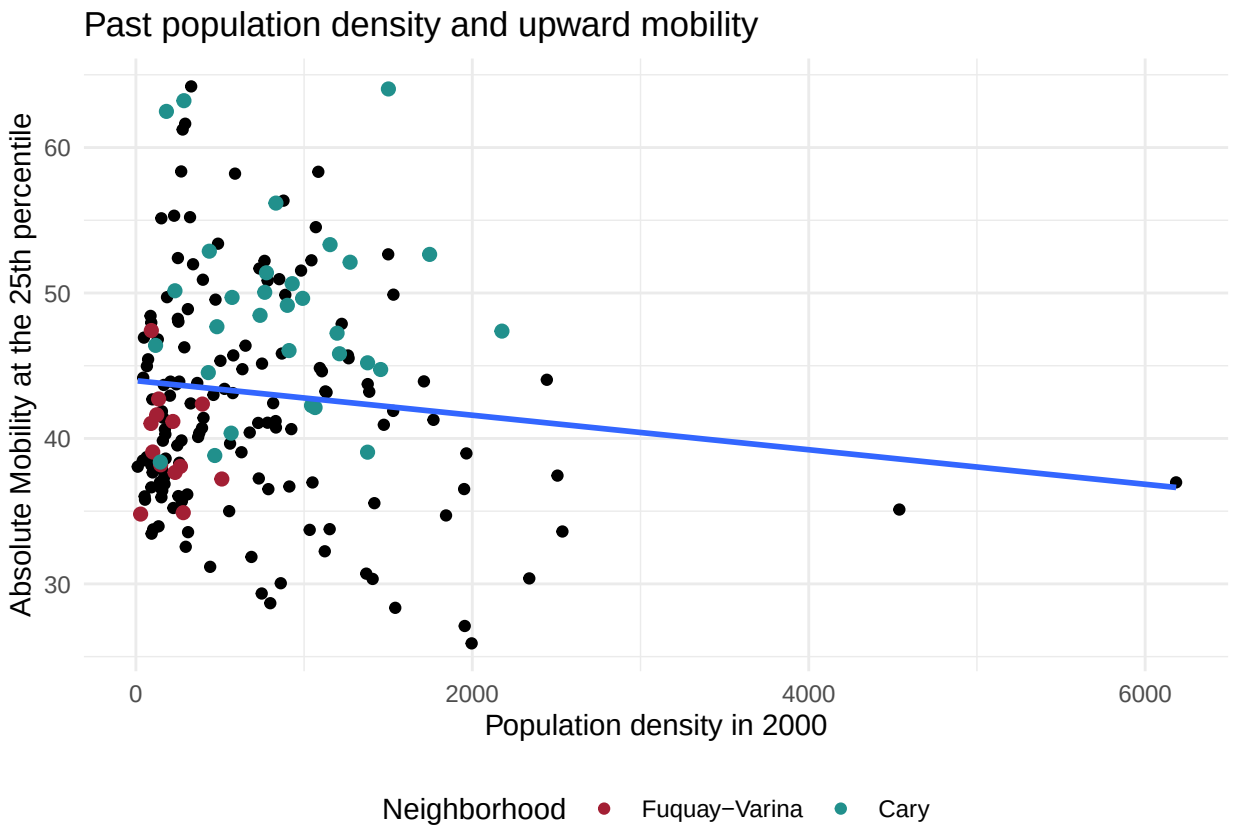
```
                    color = "#A31F34", size = 2) +
          geom_point(data = cary, aes(x = popdensity2000, y = kfr_pooled_pooled_p25),
                     color = "#21908CFF", size = 2) +
          scale_color_manual(name = "Neighborhood",
                     values = c("A" = "#A31F34", "B" =  "#21908CFF"),
                     labels = c("Fuquay-Varina", "Cary")) +
          geom_smooth(method = "lm", se = F) +
          labs(x = "Population density in 2000",
               y = "Absolute Mobility at the 25th percentile",
               title = "Past population density and upward mobility") +
          theme_minimal() +
          theme(legend.position = "bottom")
popden2000
```

```
## `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()`).
```

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```




```
ggsave("popden2000.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()`).
```

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

```
## Warning: Removed 1 rows containing missing values (`geom_point()`).
## Removed 1 rows containing missing values (`geom_point()`).
```

```
#Correlation coefficient table between historical covariates and upward mobility

#Derive correlation coefficients
white_wake2000 <- cor(wake_co$kfr_pooled_pooled_p25, wake_co$share_white2000, use = "complete.obs")
white_fuquay2000 <- cor(fuquay$kfr_pooled_pooled_p25, fuquay$share_white2000, use = "complete.obs")
white_cary2000 <- cor(cary$kfr_pooled_pooled_p25, cary$share_white2000, use = "complete.obs")

black_wake2000 <- cor(wake_co$kfr_pooled_pooled_p25, wake_co$share_black2000, use = "complete.obs")
black_fuquay2000 <- cor(fuquay$kfr_pooled_pooled_p25, fuquay$share_black2000, use = "complete.obs")
black_cary2000 <- cor(cary$kfr_pooled_pooled_p25, cary$share_black2000, use = "complete.obs")

hisp_wake2000 <- cor(wake_co$kfr_pooled_pooled_p25, wake_co$share_hisp2000, use = "complete.obs")
hisp_fuquay2000 <- cor(fuquay$kfr_pooled_pooled_p25, fuquay$share_hisp2000, use = "complete.obs")
hisp_cary2000 <- cor(cary$kfr_pooled_pooled_p25, cary$share_hisp2000, use = "complete.obs")

col_wake2000 <- cor(wake_co$kfr_pooled_pooled_p25, wake_co$frac_coll_plus2000, use = "complete.obs")
col_fuquay2000 <- cor(fuquay$kfr_pooled_pooled_p25, fuquay$frac_coll_plus2000, use = "complete.obs")
col_cary2000 <- cor(cary$kfr_pooled_pooled_p25, cary$frac_coll_plus2000, use = "complete.obs")

medinc_wake1990 <- cor(wake_co$kfr_pooled_pooled_p25, wake_co$med_hhinc1990, use = "complete.obs")
medinc_fuquay1990 <- cor(fuquay$kfr_pooled_pooled_p25, fuquay$med_hhinc1990, use = "complete.obs")
medinc_cary1990 <- cor(cary$kfr_pooled_pooled_p25, cary$med_hhinc1990, use = "complete.obs")

poor_wake1990 <- cor(wake_co$kfr_pooled_pooled_p25, wake_co$poor_share1990, use = "complete.obs")
poor_fuquay1990 <- cor(fuquay$kfr_pooled_pooled_p25, fuquay$poor_share1990, use = "complete.obs")
poor_cary1990 <- cor(cary$kfr_pooled_pooled_p25, cary$poor_share1990, use = "complete.obs")

single_wake1990 <- cor(wake_co$kfr_pooled_pooled_p25, wake_co$singleparent_share1990, use = "complete.ob
single_fuquay1990 <- cor(fuquay$kfr_pooled_pooled_p25, fuquay$singleparent_share1990, use = "complete.ob
single_cary1990 <- cor(cary$kfr_pooled_pooled_p25, cary$singleparent_share1990, use = "complete.obs")

emp_wake <- cor(wake_co$kfr_pooled_pooled_p25, wake_co$emp2000, use = "complete.obs")
emp_fuquay <- cor(fuquay$kfr_pooled_pooled_p25, fuquay$emp2000, use = "complete.obs")
emp_cary <- cor(cary$kfr_pooled_pooled_p25, cary$emp2000, use = "complete.obs")

popden_wake <- cor(wake_co$kfr_pooled_pooled_p25, wake_co$popdensity2000, use = "complete.obs")
popden_fuquay <- cor(fuquay$kfr_pooled_pooled_p25, fuquay$popdensity2000, use = "complete.obs")
popden_cary <- cor(cary$kfr_pooled_pooled_p25, cary$popdensity2000, use = "complete.obs")

#Create correlation coefficient dataframe
coefcomp_historic <- data.frame(c("2000 White Population Share", "2000 White Population Share", "2000 Wh
                          "2000 Black Population Share", "2000 Black Population Share", "2000 Black Popul
                          "2000 Hispanic Population Share", "2000 Hispanic Population Share", "2000 Hispa
                          "2000 College Educated Population", "2000 College Educated Population", "2000 C
                          "1990 Median Household Income", "1990 Median Household Income", "1990 Median Ho
                          "1990 Poverty Rate", "1990 Poverty Rate", "1990 Poverty Rate",
                          "1990 Single Parent Share", "1990 Single Parent Share", "1990 Single Parent Sha
                          "2000 Employment Rate", "2000 Employment Rate", "2000 Employment Rate",
                          "2000 Pop. Density", "2000 Pop. Density", "2000 Pop. Density"),
                        c("Wake County", "Fuquay-Varina", "Cary",
```

```
                        "Wake County", "Fuquay-Varina", "Cary",
                        "Wake County", "Fuquay-Varina", "Cary",
                        "Wake County", "Fuquay-Varina", "Cary",
                        "Wake County", "Fuquay-Varina", "Cary",
                        "Wake County", "Fuquay-Varina", "Cary",
                        "Wake County", "Fuquay-Varina", "Cary",
                        "Wake County", "Fuquay-Varina", "Cary",
                        "Wake County", "Fuquay-Varina", "Cary"),
                      c(white_wake2000, white_fuquay2000, white_cary2000,
                        black_wake2000, black_fuquay2000, black_cary2000,
                        hisp_wake2000, hisp_fuquay2000, hisp_cary2000,
                        col_wake2000, col_fuquay2000, col_cary2000,
                        medinc_wake1990, medinc_fuquay1990, medinc_cary1990,
                        poor_wake1990, poor_fuquay1990, poor_cary1990,
                        single_wake1990, single_fuquay1990, single_cary1990,
                        emp_wake, emp_fuquay, emp_cary,
                        popden_wake, popden_fuquay, popden_cary))

#Name columns
names(coefcomp_historic)[1] <- "Covariate"
names(coefcomp_historic)[2] <- "Region"
names(coefcomp_historic)[3] <- "Correlation Coefficient"

#check
coefcomp_historic
```

```
##                          Covariate        Region Correlation Coefficient
## 1       2000 White Population Share   Wake County              0.620324316
## 2       2000 White Population Share Fuquay-Varina              0.268415468
## 3       2000 White Population Share          Cary              0.346345912
## 4       2000 Black Population Share   Wake County             -0.642523474
## 5       2000 Black Population Share Fuquay-Varina             -0.179968811
## 6       2000 Black Population Share          Cary             -0.535303036
## 7    2000 Hispanic Population Share   Wake County             -0.427331697
## 8    2000 Hispanic Population Share Fuquay-Varina             -0.301371936
## 9    2000 Hispanic Population Share          Cary             -0.313481381
## 10 2000 College Educated Population   Wake County              0.693882710
## 11 2000 College Educated Population Fuquay-Varina              0.156721242
## 12 2000 College Educated Population          Cary              0.442685547
## 13     1990 Median Household Income   Wake County              0.626811630
## 14     1990 Median Household Income Fuquay-Varina              0.287674950
## 15     1990 Median Household Income          Cary              0.102431202
## 16                1990 Poverty Rate   Wake County             -0.508409431
## 17                1990 Poverty Rate Fuquay-Varina             -0.071432044
## 18                1990 Poverty Rate          Cary              0.067460892
## 19         1990 Single Parent Share   Wake County             -0.593419864
## 20         1990 Single Parent Share Fuquay-Varina             -0.100160966
## 21         1990 Single Parent Share          Cary             -0.017097962
## 22             2000 Employment Rate   Wake County              0.436586051
## 23             2000 Employment Rate Fuquay-Varina              0.282339723
## 24             2000 Employment Rate          Cary              0.003285609
## 25                2000 Pop. Density   Wake County             -0.116940431
## 26                2000 Pop. Density Fuquay-Varina             -0.197430259
## 27                2000 Pop. Density          Cary             -0.023841395
```

```r
#create table
table_historic <- coefcomp_historic |>
                   kable("html",
                         caption = "Correlations Between Historical Covariates and Upward Mobility",
                         align = "lcr",
                         col.names = c("Covariate", "Region", "Correlation Coefficient")) |>
                   kable_styling("striped",  # Table style: "striped", "bordered", "hover", "condensed
                  full_width = FALSE) |>
                  save_kable(file = "historic.png",
                             zoom = 1)
#check
table_historic
```

```
## [1] "/Users/Jose/Documents/R/EC50/Project1/historic.png"
## attr(,"info")
## # A tibble: 1 x 7
##   format width height colorspace matte filesize density
##   <chr>  <int>  <int> <chr>      <lgl>    <int> <chr>
## 1 PNG      491   1003 sRGB       TRUE         0 28x28
```

```r
#Multivariate regression of historical covariates and upward mobility

#Wake County
wakereg <- lm(kfr_pooled_pooled_p25 ~ share_white2000 + share_black2000 + share_hisp2000 +
                                          frac_coll_plus2000 + med_hhinc1990 + poor_share1990 +
                                          singleparent_share1990 + emp2000 + popdensity2000,
                                        data = wake_co)

summary(wakereg)
```

```
##
## Call:
## lm(formula = kfr_pooled_pooled_p25 ~ share_white2000 + share_black2000 +
##     share_hisp2000 + frac_coll_plus2000 + med_hhinc1990 + poor_share1990 +
##     singleparent_share1990 + emp2000 + popdensity2000, data = wake_co)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.3641  -3.1826  -0.3067   2.9183  16.3136
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)              4.438e+01  1.680e+01   2.642  0.00899 **
## share_white2000         -2.087e+01  1.513e+01  -1.380  0.16948
## share_black2000         -2.372e+01  1.609e+01  -1.474  0.14227
## share_hisp2000          -3.398e+01  2.029e+01  -1.675  0.09579 .
## frac_coll_plus2000       1.565e+01  4.596e+00   3.405  0.00082 ***
## med_hhinc1990            1.367e-04  5.184e-05   2.637  0.00913 **
## poor_share1990           6.069e+00  9.767e+00   0.621  0.53514
## singleparent_share1990 -8.217e+00  6.370e+00  -1.290  0.19883
## emp2000                  1.193e+01  7.301e+00   1.635  0.10394
## popdensity2000          -3.452e-04  5.868e-04  -0.588  0.55709
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 5.004 on 174 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.5977, Adjusted R-squared:  0.5769
## F-statistic: 28.73 on 9 and 174 DF,  p-value: < 2.2e-16
```

```
#Fuquay-Varina tracts
fuquayreg <- lm(kfr_pooled_pooled_p25 ~ share_white2000 + share_black2000 + share_hisp2000 +
                                        frac_coll_plus2000 + med_hhinc1990 + poor_share1990 +
                                        singleparent_share1990 + emp2000 + popdensity2000,
                                       data = fuquay)

summary(fuquayreg)
```

```
##
## Call:
## lm(formula = kfr_pooled_pooled_p25 ~ share_white2000 + share_black2000 +
##     share_hisp2000 + frac_coll_plus2000 + med_hhinc1990 + poor_share1990 +
##     singleparent_share1990 + emp2000 + popdensity2000, data = fuquay)
##
## Residuals:
##          1          2          3          4          5          6          7
## -1.469e-06 -3.153e+00  1.873e+00  1.280e+00 -1.543e-04  1.547e-04  4.314e-07
##          8          9         10         11         12         13
## -2.767e+00  2.173e+00  1.362e-01  2.028e+00 -1.571e+00 -1.907e-04
## attr(,"label")
## [1] "Mean pctile rank in the national distribution of household income in 2014-2015"
## attr(,"format.stata")
## [1] "%9.0g"
##
## Coefficients: (1 not defined because of singularities)
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)            -4.726e+07  4.315e+07  -1.095    0.335
## share_white2000         4.739e+07  4.333e+07   1.094    0.336
## share_black2000         4.842e+07  4.428e+07   1.094    0.336
## share_hisp2000          4.900e+07  4.479e+07   1.094    0.335
## frac_coll_plus2000      1.080e+06  9.930e+05   1.087    0.338
## med_hhinc1990           6.082e+00  4.833e+00   1.258    0.277
## poor_share1990          2.990e+06  3.084e+06   0.969    0.387
## singleparent_share1990 -2.158e+06  2.395e+06  -0.901    0.419
## emp2000                         NA         NA      NA       NA
## popdensity2000         -2.644e-02  1.100e-02  -2.404    0.074 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.918 on 4 degrees of freedom
## Multiple R-squared:  0.7669, Adjusted R-squared:  0.3006
## F-statistic: 1.645 on 8 and 4 DF,  p-value: 0.3317
```

```
#Cary Tracts
caryreg <- lm(kfr_pooled_pooled_p25 ~ share_white2000 + share_black2000 + share_hisp2000 +
                                      frac_coll_plus2000 + med_hhinc1990 + poor_share1990 +
                                      singleparent_share1990 + emp2000 + popdensity2000,
                                     data = cary)

summary(caryreg)
```

```
##
## Call:
## lm(formula = kfr_pooled_pooled_p25 ~ share_white2000 + share_black2000 +
##     share_hisp2000 + frac_coll_plus2000 + med_hhinc1990 + poor_share1990 +
##     singleparent_share1990 + emp2000 + popdensity2000, data = cary)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.660 -3.277  0.146  2.140 11.514
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)             8.534e+01  5.381e+01   1.586   0.1277
## share_white2000        -1.459e+01  4.534e+01  -0.322   0.7508
## share_black2000        -1.509e+02  7.316e+01  -2.063   0.0517 .
## share_hisp2000         -6.149e+01  9.273e+01  -0.663   0.5145
## frac_coll_plus2000      9.773e+00  2.511e+01   0.389   0.7010
## med_hhinc1990          -2.189e-04  1.843e-04  -1.188   0.2482
## poor_share1990          2.498e+01  3.231e+01   0.773   0.4481
## singleparent_share1990  4.143e+01  2.439e+01   1.698   0.1042
## emp2000                -2.416e+01  3.104e+01  -0.778   0.4450
## popdensity2000          5.860e-03  3.480e-03   1.684   0.1070
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.552 on 21 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.5009, Adjusted R-squared:  0.287
## F-statistic: 2.342 on 9 and 21 DF,  p-value: 0.05205
```

```
# Create a regression table using the stargazer() function
stargazer(wakereg, type = "text",  # Change 'type' to "latex" or "html" for different output formats
          title = "Regression Results",
          align = TRUE,
          column.labels = c("Wake County"),
          covariate.labels = c("2000 White pop. share", "2000 Black pop. share", "2000 Hisp. pop. share
                               "2000 College degree plus share", "1990 Median household income", "1990 p
                               "1990 Single parent share", "2000 Employment rate", "2000 Population dens
          dep.var.caption = "Dependent variable: Absolute mobility at the 25th percentile",
          dep.var.labels.include = TRUE,
          digits = 2,
          intercept.bottom = TRUE,
          model.numbers = TRUE,
          no.space = TRUE,
          omit.stat = c("f", "ser"),
          single.row = TRUE)
```

```
##
## Regression Results
## ================================================================================================
##                                  Dependent variable: Absolute mobility at the 25th percentile
##                                  ------------------------------------------------------------
##                                                  kfr_pooled_pooled_p25
##                                                       Wake County
## ------------------------------------------------------------------------------------------------
```

```
## 2000 White pop. share                                -20.87 (15.13)
## 2000 Black pop. share                                -23.72 (16.09)
## 2000 Hisp. pop. share                               -33.98* (20.29)
## 2000 College degree plus share                      15.65*** (4.60)
## 1990 Median household income                      0.0001*** (0.0001)
## 1990 poverty share                                   6.07 (9.77)
## 1990 Single parent share                             -8.22 (6.37)
## 2000 Employment rate                                 11.93 (7.30)
## 2000 Population density                             -0.0003 (0.001)
## Constant                                            44.38*** (16.80)
## ------------------------------------------------------------------------------------------
## Observations                                               184
## R2                                                         0.60
## Adjusted R2                                                0.58
## ==========================================================================================
## Note:                                                *p<0.1; **p<0.05; ***p<0.01
```