Project 1

Jon Moreno-Medina

Question 1

Start by looking up the city where you grew up on the Opportunity Atlas. Zoom in to the Census tracts around your home. Figure 1 in your narrative should be a map of the Census tracts in your hometown from the Opportunity Atlas. Examples for Milwaukee, WI (where Professor Chetty grew up) and Los Angeles, CA (discussed in Lecture 1) are shown on the next page. The text of your narrative should describe what you see, and what data are being visualized. Examine the patterns for a number of different groups (e.g., lowest income children, high income children) and outcomes (e.g., earnings in adulthood, incarceration rates). Only choose one or two of these to include in your narrative.

Answer

LEAVE UP TO YOU.

Question 2

(To answer this question, read the Opportunity Atlas manuscript) What period do the data you are analyzing come from? Are you concerned that the neighborhoods you are studying may have changed for kids now growing up there? What evidence do Chetty et al. (2018) provide suggesting that such changes are or are not important? What type of data could you use to test whether your neighborhood has changed in recent years?

Answer

Look at Section IV.D and Figure IX.a and b. ;)

Question 3

Now turn to the `atlas.rds` data set. How does average upward mobility, pooling races and genders, for children with parents at the 25th percentile (`kfr_pooled_p25`) in your home Census tract compare to mean (population-weighted, using `count_pooled`) upward mobility in your state and in the U.S. overall? Do kids where you grew up have better or worse chances of climbing the income ladder than the average child in America?

Answer

First, load packages:

```
library(ggplot2)
library(dplyr)
library(Hmisc)
```

Load the data:

```
atlas <- readRDS(gzcon(url("https://raw.githubusercontent.com/jrm87/ECO3253_fall2023/maste
```

Create data for tract, county and state. My tract is: - st = "48" - county = "029" - tract = "181820"

```
atlas_texas<-atlas%>%
  filter(state==48)

atlas_bexar<-atlas%>%
  filter(state==48, county==029)

atlas_utsa<-atlas%>%
  filter(state==48, county==029, tract==181820)
```

Let's calculate the mobility averages for different places:

Mobility at UTSA tract:

```
atlas_utsa$kfr_pooled_p25
```

```
[1] 51314.05
attr(,"label")
[1] "Household income ($) for children with parents at 25 percentile"
attr(,"format.stata")
[1] "%9.0g"
```

Mean mobility for children in the US:

```
wtd.mean(atlas$kfr_pooled_p25, atlas$count_pooled)
```

```
[1] 34311.68
```

The average mobility across Texas is:

```
wtd.mean(atlas_texas$kfr_pooled_p25, atlas_texas$count_pooled)
```

```
[1] 34728.31
```

The average mobility across Bexar county is:

```
wtd.mean(atlas_bexar$kfr_pooled_p25, atlas_bexar$count_pooled)
```

```
[1] 32731.71
```

Question 4

What is the standard deviation of upward mobility (population-weighted) in your home county? Is it larger or smaller than the standard deviation across tracts in your state? Across tracts in the country? What do you learn from these comparisons?

Answer

The standard deviation gives us a measure of inequality of opportunity.

SD for the US:

```
sqrt(wtd.var(atlas$kfr_pooled_p25, atlas$count_pooled))
```

```
[1] 7899.531
```

SD for Texas:

```
sqrt(wtd.var(atlas_texas$kfr_pooled_p25, atlas_texas$count_pooled))
```

```
[1] 6703.9
```

SD for Bexar County:

```
sqrt(wtd.var(atlas_bexar$kfr_pooled_p25, atlas_bexar$count_pooled))
```

```
[1] 5274.712
```

We can see that inequality of economic mobility for low income families is larger across the US, somewhat lower within Texas, and lower still within Bexar County.

Question 5

Now let's turn to downward mobility: repeat questions (3) and (4) looking at children who start with parents at the 75th and 100th percentiles. How do the patterns differ?

Answer

Same as above, but with `kfr_pooled_p75` and `kfr_pooled_p100`. Here, you should point how your neighborhood, county, and state compare with each other, and with the overall mobility for p75. Do you find the same patterns as you did with `kfr_pooled_p25`?

Question 6

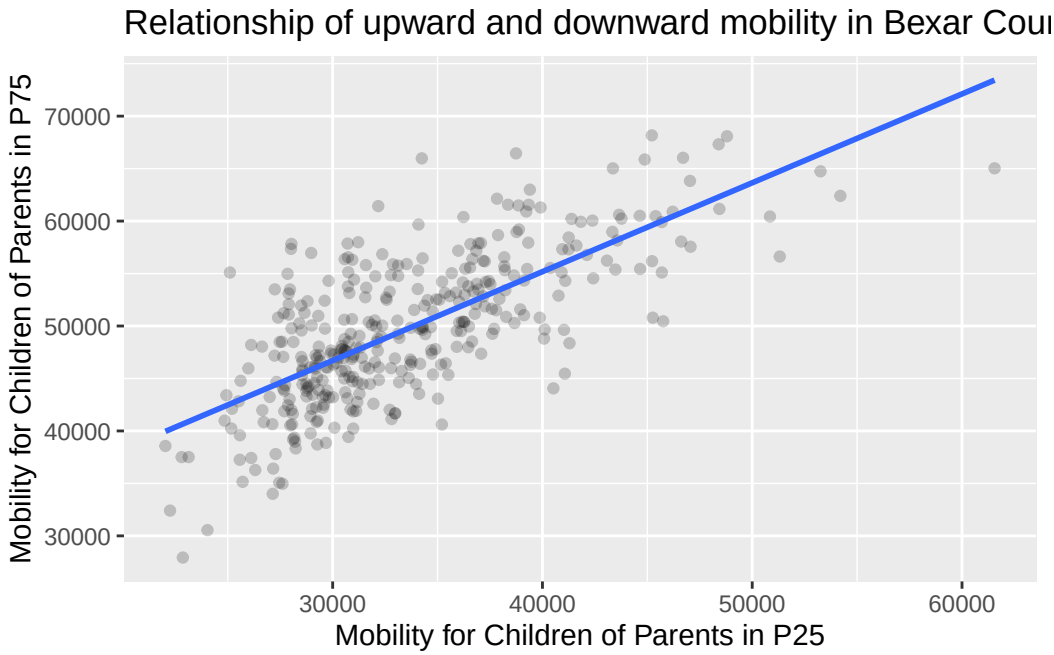
Using a linear regression, estimate the relationship between outcomes of children at the 25th and 75th percentile for the Census tracts in your home county. Generate a scatter plot to visualize this regression. Do areas where children from low-income families do well generally have better outcomes for those from high-income families, too?

Answer:

Just look at the Section on plotting the regression line in Chapter 14 ;)

Let's plot the correlation between kfr_pooled_p25 and kfr_pooled_p75 across the the county.

```
ggplot(atlas_bexar, aes(x = kfr_pooled_p25, y = kfr_pooled_p75)) +
  geom_point(alpha = 0.2) +
  labs(x = "Mobility for Children of Parents in P25", y = "Mobility for Children of Parent
       title = "Relationship of upward and downward mobility in Bexar County") +
  geom_smooth(method = "lm", se = FALSE)
```



It shows a positive slope, which implies a positive correlation between the variables. In this particular case, it shows that as tracts tend to have higher mobility for low income children,

they also have higher mobility for high income children. Thus, there does not seem to be a trade offs between improving mobility for both groups in principle.

How big is this relationship? We can run a simple linear regression to find out the exact number.

We run the following command, that creates a 'statistical model' of that linear regression. We will save the model in an object called `model1`, where we will take `kfr_pooled_p75` in the left hand side (explained variable), and `kfr_pooled_p25` in the right hand side (explanatory variable):

```
model1<-lm(kfr_pooled_p75~kfr_pooled_p25, data=atlas_bexar)
```

Let's see what is inside this 'model':

```
summary(model1)
```

```
Call:
lm(formula = kfr_pooled_p75 ~ kfr_pooled_p25, data = atlas_bexar)

Residuals:
     Min       1Q   Median       3Q      Max
-12728.5  -3360.3   -195.9   3128.7  15665.2

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.130e+04  1.473e+03   14.47   <2e-16 ***
kfr_pooled_p25 8.469e-01  4.321e-02   19.60   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4853 on 357 degrees of freedom
  (5 observations deleted due to missingness)
Multiple R-squared:  0.5183,    Adjusted R-squared:  0.517
F-statistic: 384.1 on 1 and 357 DF,  p-value: < 2.2e-16
```

How do we read this? Just focus on the table called 'Coefficients' [just ignore the rest for now]. The number in front of 'kfr_pooled_p25' is the estimated correlation (not exactly the correlation, as correlations are numbers between -1 and +1, but similar enough as we will see in a second) with the mobility variable. This number is `8.469e-01`, which is 0.847. You read this as increasing the variable `kfr_pooled_p25` (income mobility for kids with parents

in the 25th percentile) by 1 (dollar), predicts an increment of 0.85 dollars in `kfr_pooled_p75` (income mobility for kids with parents in the 75th percentile). By the way, that is the slope you see in the Figure above.

Is this relationship precisely estimated, or is it noisy? To answer this, we read the column `Pr(>|t|)` - called the p-value. If the number is below 0.05, we say the relationship is quite precisely estimated. In this case, the number is `<2e-16`, which says that the p-value is below `0.00000000000000002`, so yes, it is lower than 0.05.

Question 7

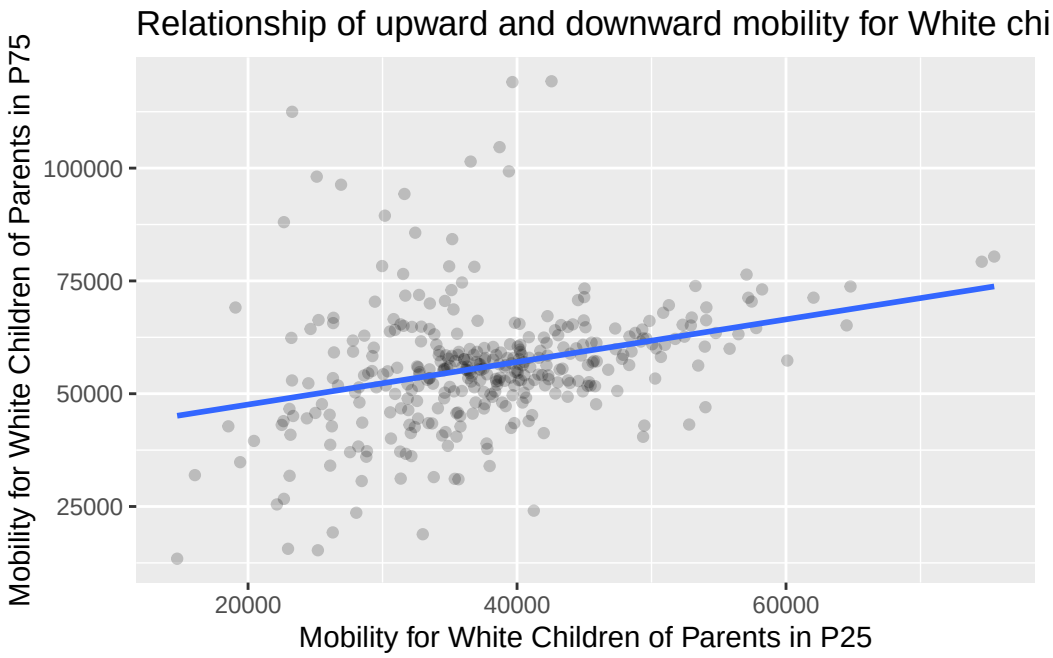
Next, examine whether the patterns you have looked at above are similar by race. If there is not enough racial heterogeneity in the area of interest (i.e., data is missing for most racial groups), then choose a different area to examine.

Answer

I'll simplify the analysis by considering what is happening with Black, Hispanic and White children. You can, of course, extend the same analysis to Asian and Native American children, so far as the as data allows. To do this, we explore what is happening with the variables `kfr_white_p25`, `kfr_black_p25` and `kfr_hisp_p25`.

White Children

```
ggplot(atlas_bexar, aes(x = kfr_white_p25, y = kfr_white_p75)) +
  geom_point(alpha = 0.2) +
  labs(x = "Mobility for White Children of Parents in P25", y = "Mobility for White Childr
       title = "Relationship of upward and downward mobility for White children in Bexar C
  geom_smooth(method = "lm", se = FALSE)
```



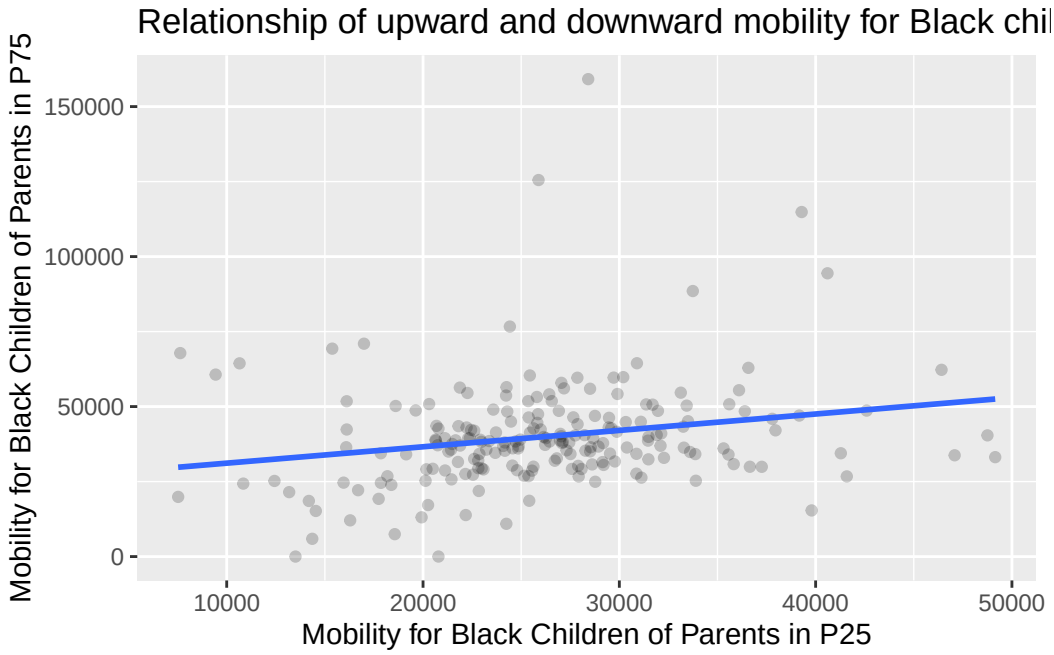

```r
mod1_w <- lm(kfr_white_p75 ~kfr_white_p25 , data = atlas_bexar)
summary(mod1_w)
```

```
Call:
lm(formula = kfr_white_p75 ~ kfr_white_p25, data = atlas_bexar)

Residuals:
   Min     1Q Median     3Q    Max
-34876  -6207   -561   3878  63338

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.816e+04  3.119e+03   12.23  < 2e-16 ***
kfr_white_p25 4.719e-01  8.012e-02    5.89 9.24e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13440 on 342 degrees of freedom
  (20 observations deleted due to missingness)
Multiple R-squared:  0.0921,    Adjusted R-squared:  0.08944
F-statistic: 34.69 on 1 and 342 DF,  p-value: 9.239e-09
```

Again, across Bexar County, we see that among white children, we see a positive slope between the mobility for low and high income children. An increment of 1 dollar in the income of low income White children is associated with an increase of 0.47 dollars for high income White children. This relationship also is precisely estimated.

Black Children

```
ggplot(atlas_bexar, aes(x = kfr_black_p25, y = kfr_black_p75)) +
  geom_point(alpha = 0.2) +
  labs(x = "Mobility for Black Children of Parents in P25", y = "Mobility for Black Childr
       title = "Relationship of upward and downward mobility for Black children in Bexar C
  geom_smooth(method = "lm", se = FALSE)
```




```
mod1_b <- lm(kfr_black_p75 ~kfr_black_p25 , data = atlas_bexar)
summary(mod1_b)
```

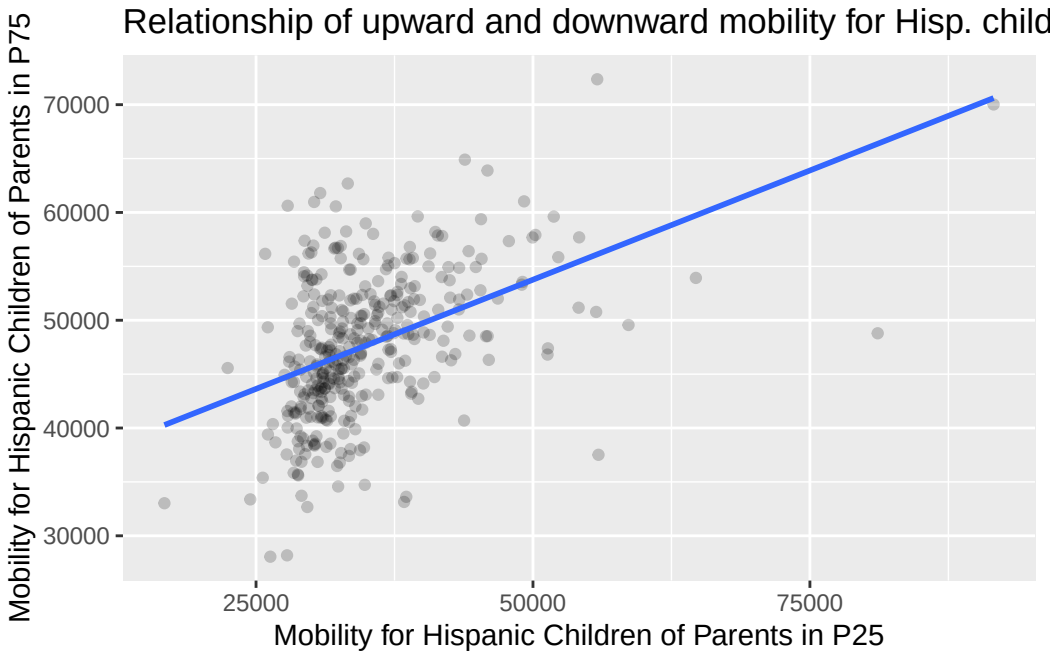
```
Call:
lm(formula = kfr_black_p75 ~ kfr_black_p25, data = atlas_bexar)

Residuals:
```

```
   Min     1Q Median     3Q    Max
-37018  -9951  -1953   5575 117959

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.568e+04  4.692e+03   5.472 1.31e-07 ***
kfr_black_p25 5.460e-01  1.727e-01   3.162  0.00181 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17600 on 201 degrees of freedom
  (161 observations deleted due to missingness)
Multiple R-squared:  0.04737,   Adjusted R-squared:  0.04263
F-statistic: 9.995 on 1 and 201 DF,  p-value: 0.001812
```

Again, we see that among Black children, we also see a positive slope between the mobility for low and high income children, somewhat larger in magnitude that the slope among White children. An increment of 1 dollar in the income of low income Black children is associated with an increase of 0.54 dollars for high income Black children.

Hispanic Children

```r
ggplot(atlas_bexar, aes(x = kfr_hisp_p25, y = kfr_hisp_p75)) +
  geom_point(alpha = 0.2) +
  labs(x = "Mobility for Hispanic Children of Parents in P25", y = "Mobility for Hispanic
       title = "Relationship of upward and downward mobility for Hisp. children in Bexar C
  geom_smooth(method = "lm", se = FALSE)
```




```
mod1_h <- lm(kfr_hisp_p75 ~kfr_hisp_p25 , data = atlas_bexar)
summary(mod1_h)
```

```
Call:
lm(formula = kfr_hisp_p75 ~ kfr_hisp_p25, data = atlas_bexar)

Residuals:
     Min       1Q   Median       3Q      Max
-18644.3  -3768.5    -80.2   3727.0  16243.9

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.350e+04  1.538e+03  21.786   <2e-16 ***
kfr_hisp_p25 4.053e-01  4.314e-02   9.396   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5964 on 356 degrees of freedom
  (6 observations deleted due to missingness)
Multiple R-squared:  0.1987,    Adjusted R-squared:  0.1965
F-statistic: 88.28 on 1 and 356 DF,  p-value: < 2.2e-16
```

Lastly, we see a similar pattern, now among Hispanic children, although a bit lower than the average slope. An increment of 1 dollar in the income of low income Hispanic children is associated with an increase of 0.41 dollars for high income Hispanic children.

Question 8

Using the Census tracts in your home county, can you identify any covariates which help explain some of the patterns you have identified above? Some examples of covariates you might examine include housing prices, income inequality, fraction of children with single parents, job density, etc. For 2 or 3 of these, report estimated correlation coefficients along with their 95% confidence intervals.

Answer

Here you should have been creative with your choice of variables to explore. I'll choose poverty rates and rental prices, and see how do they correlate with mobility for low income children. You could have chosen other variables, of course.

This analysis allows us to answer thw following question: What if we not only want to consider the association between mobility and poverty, say, but also want to explore another variable at the same time?

If we want to explore changes in mobility, with changes in poverty AND rental prices (the median rental price for a two bedroom housing unit in 2015, to be precise). We can run the following regression:

```
model3<-lm(kfr_pooled_p25~poor_share2010+rent_twobed2015, data=atlas) # notice the '+'
summary(model3)
```

```
Call:
lm(formula = kfr_pooled_p25 ~ poor_share2010 + rent_twobed2015,
    data = atlas)

Residuals:
   Min     1Q Median     3Q    Max
-39020  -4132   -583   3409  70002

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      3.491e+04  9.255e+01  377.19   <2e-16 ***
```

```
poor_share2010  -3.115e+04  2.398e+02 -129.94   <2e-16 ***
rent_twobed2015  3.997e+00  7.258e-02   55.07   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6380 on 56124 degrees of freedom
  (17151 observations deleted due to missingness)
Multiple R-squared:  0.3276,    Adjusted R-squared:  0.3276
F-statistic: 1.367e+04 on 2 and 56124 DF,  p-value: < 2.2e-16
```

This tells us that, across the US, an increment of 1% in the poverty rate of the tract, predicts 311 dollars less in income for children of low income parents IF we fix also the rental prices. Similarly, FIXING poverty rates, a one dollar increase in rental prices is associated with an increase of $3.99 in mobility.

To see the confidence intervals, you can use the function `confint` and put the model in the argument, like this:

```
confint(model3)
```

```
                      2.5 %      97.5 %
(Intercept)      34729.323649  35092.1356
poor_share2010  -31624.499157 -30684.6514
rent_twobed2015      3.854669      4.1392
```

It shows that with a different sample, 95% of all coefficients for the poverty rate will be between -316 and -306, while the equivalent interval for the rental is 3.85 and 4.14.

Question 9

Open question: formulate a hypothesis for why you see the variation in upward mobility for children who grew up in the Census tracts near your home and provide correlation evidence testing that hypothesis.

Answer

Here, again, you should be creative with your hypothesis, and you should show all the evidence to support it as you can muster with this dataset and analysis.