ECON1000 - Using Big Data to Solve Economic and Social Problems
Economics and Watson
Brown University
Prof. John Friedman

ECON 1000

Empirical Exercise #1 Submission Template

Due TUESDAY 9/19/2023, by 10am EDT on Gradescope

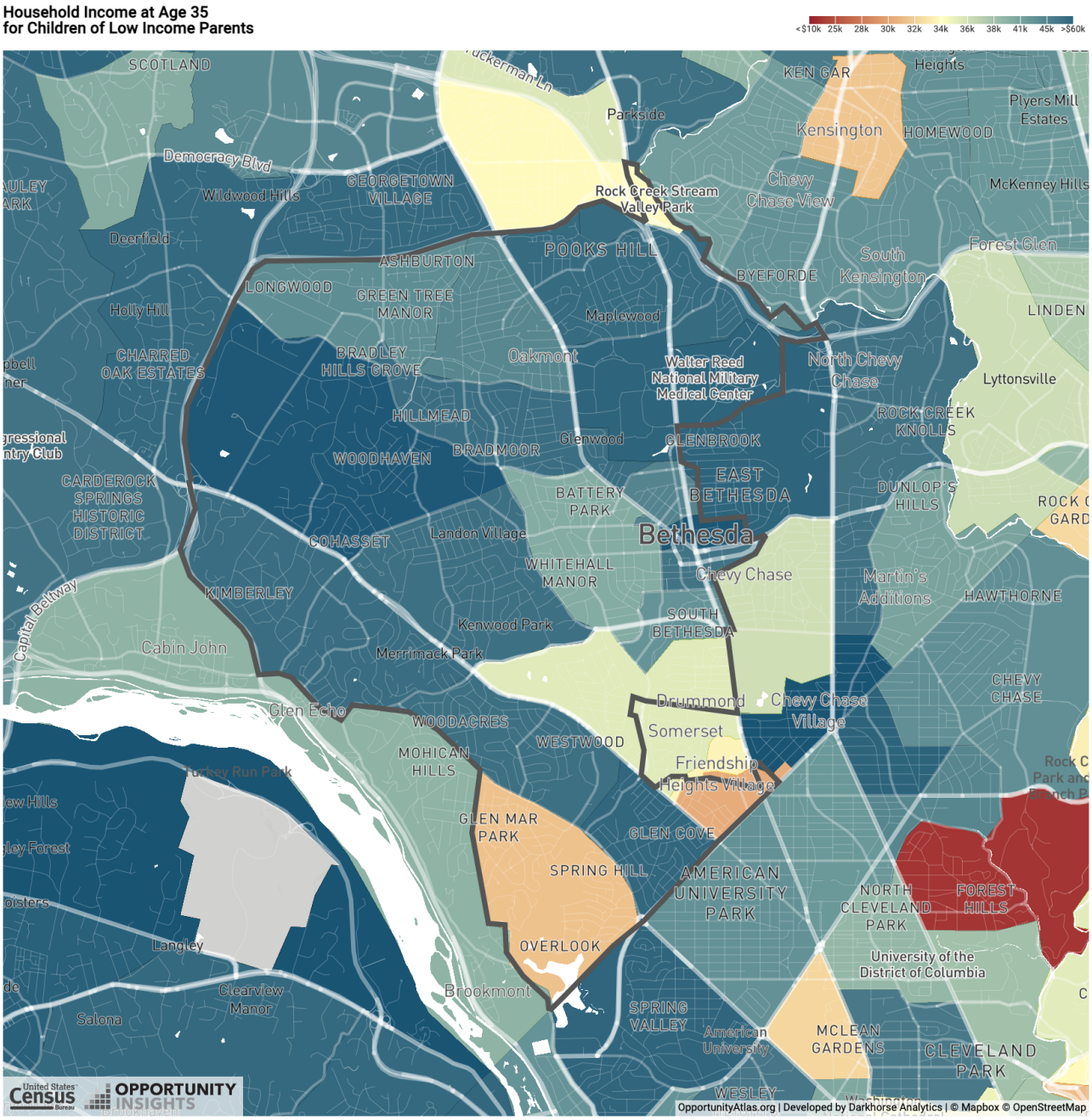
Gradescope Course Link: https://www.gradescope.com/courses/565275

Name: Sonya Rashkovan

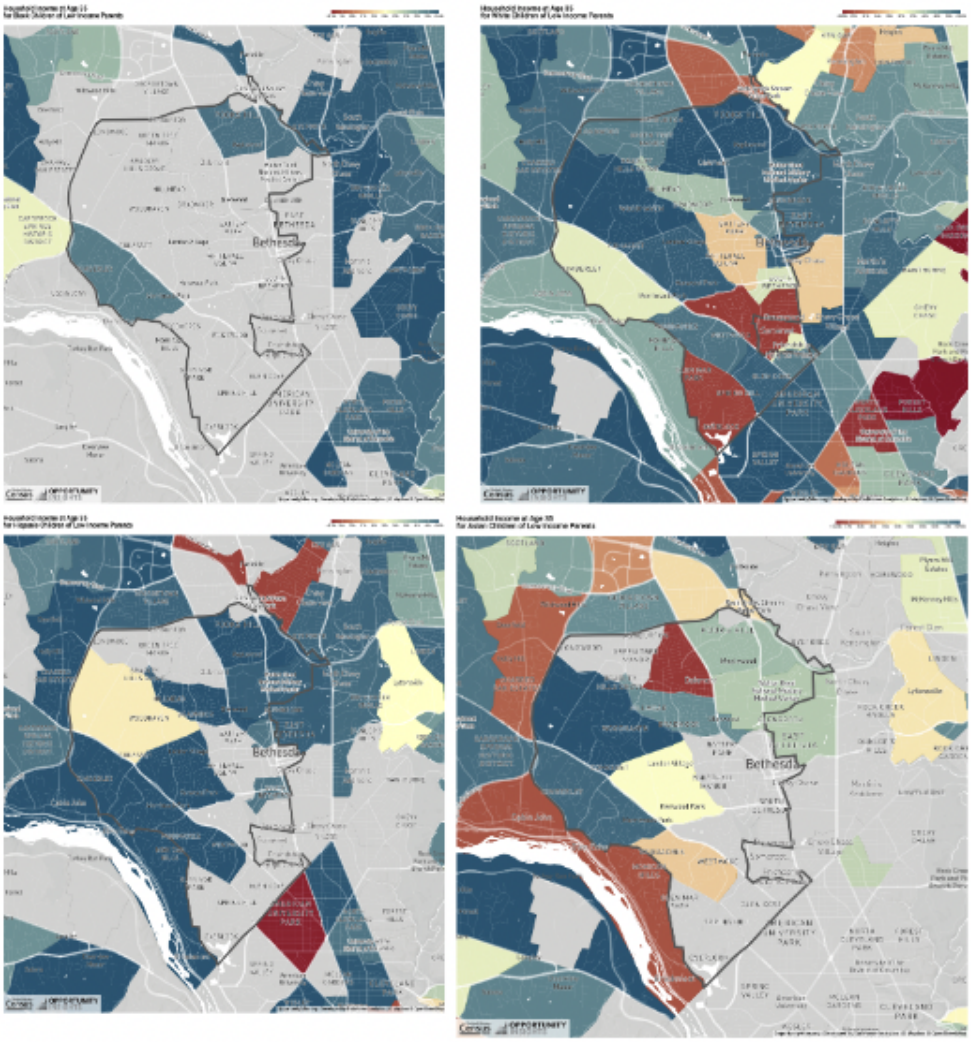
Group members with whom you worked[1]: N/A

[1] In this class we encourage working in groups because you will learn a great deal from your peers. At the end of the day, however, it is important that you write up your own analysis. Concretely, this means you should choose your own city or county for the full exercise, as well as your own new variable to study in question 7. Please then list all group members with whom you worked. If you have any questions, please ask.

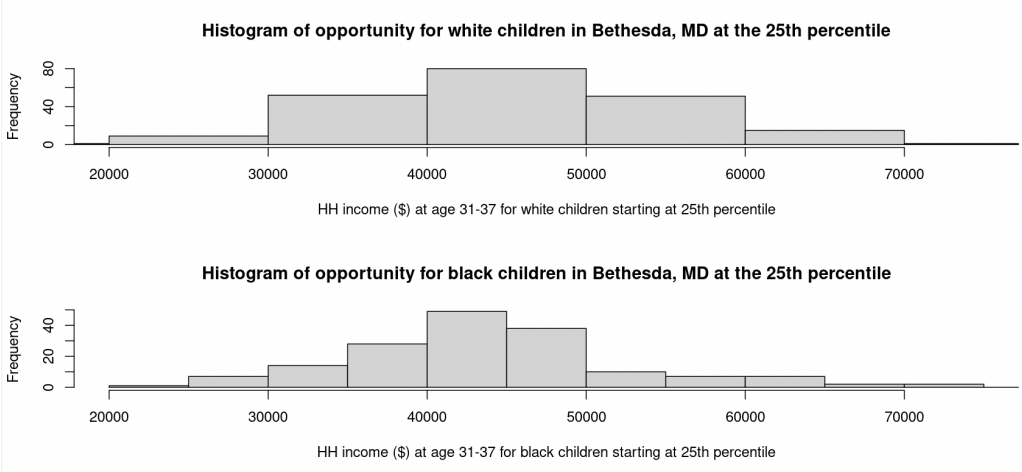
1. Produce a map of opportunity for low-income children in your chosen city.



2. Qualitatively, how does this picture differ by race and ethnicity?



3. Produce a dual-histogram showing the distribution of opportunity for low-income children from at least two different races or ethnicities in your county.



4. Calculate the weighted mean of opportunity for low-income children in your chosen county for each race.

White: $46916

Black: $33697

Asian: $62934

Hispanic: $43338

American Indian: N/A

Interpretation: Weighted mean helps us assess if the opportunities are distributed equitably among low-income children of different races, accounting for the relative contribution. In the case of Bethesda, MD , there is a clearly inequitable distribution of wealth among low-incoming children of different races with Asian children coming in first with an average $62934, followed by White and Hispanic children with $46916 and $43338, respectively, and Black children significantly lagging behind with average $33697. Native American children are not represented in the data enough to gauge their wealth outcomes. This goes in accordance with the histogram in #3 that shows a more even distribution of wealth among white children while Black children's wealth distribution is more scattred to the left towards lower income.

5. Calculate the weighted standard deviation of opportunity for low-income children

in your chosen county for each race.

White: $9445

Black: $5581

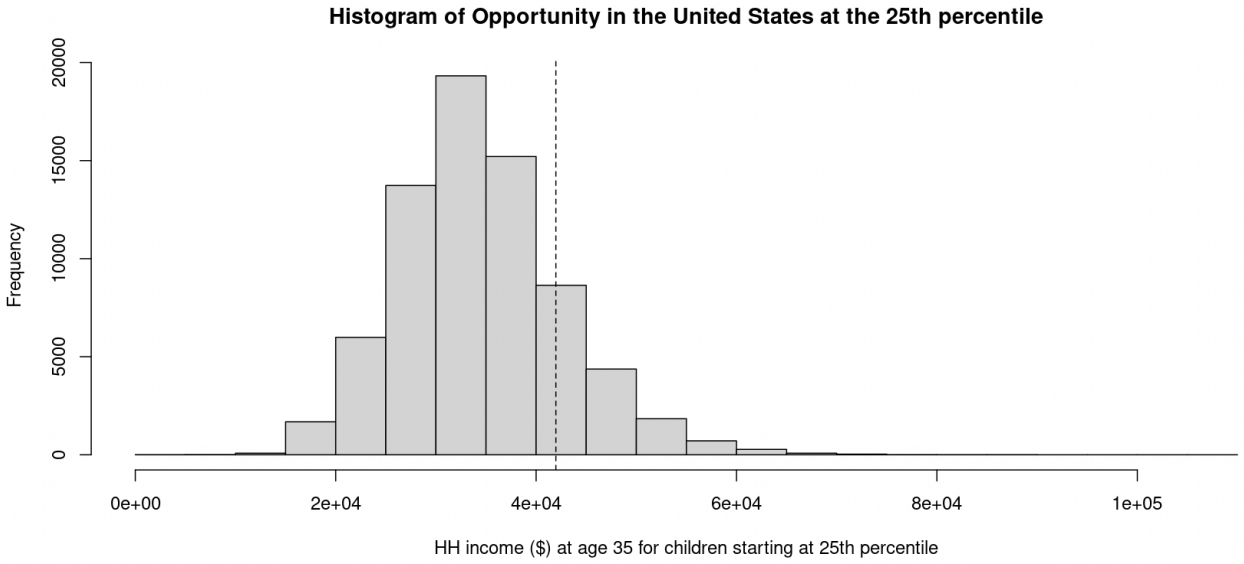
Asian: $11542

Hispanic: $7656

American Indian: N/A

Interpretation: The standard deviation shows the variation of outcomes for low-income children across races. In this case, we see a significantly higher SD for Asian and white children, $11542 and $9445 respectively, which suggests a higher level variance of outcomes while for Black children the SD is significantly lower with just $5581 which suggests more evenly distributed outcomes. Native American children are not represented in the data enough to gauge their wealth outcomes. This goes in accordance with the histogram in #3 that shows a less "diverse" or variant distribution of wealth in Black children and a bigger variance – it's also important to note that the mean for Black children is on average lower than white children, so the variance appears in different income brackets.

6. Using data for the entire USA, plot the distribution of opportunity for low-income kids. Where in this distribution is the mean for your chosen county?

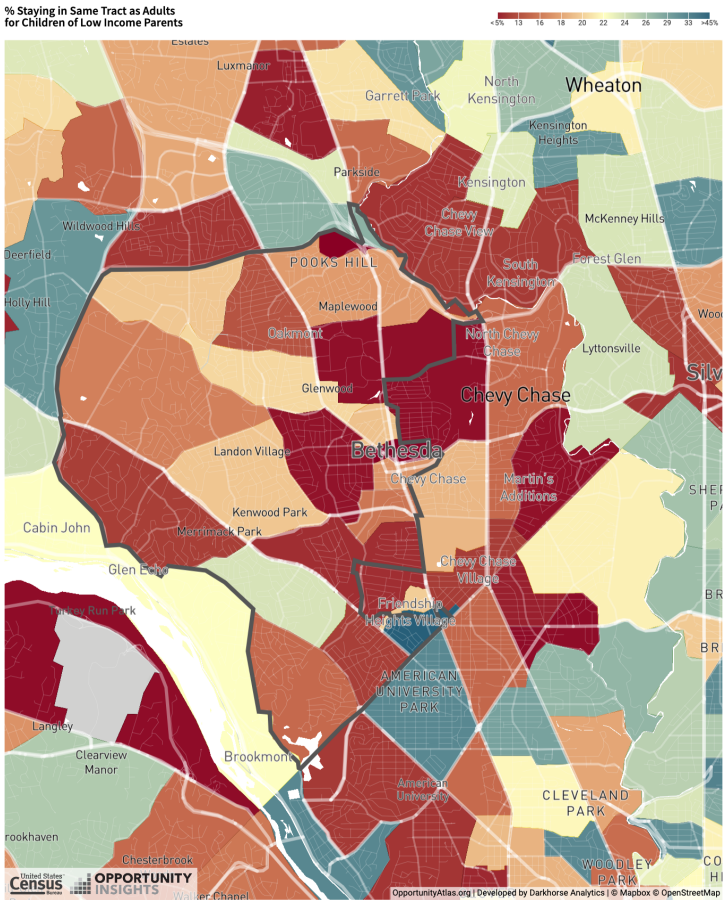


---

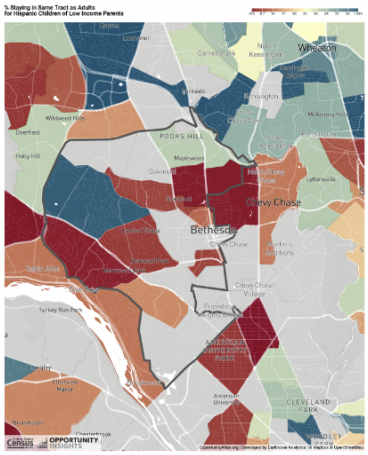
7. Repeat steps 1-6 looking at a different tract-level outcome in the Opportunity Atlas. You may choose from the following list: incarceration, teenbirth, fraction

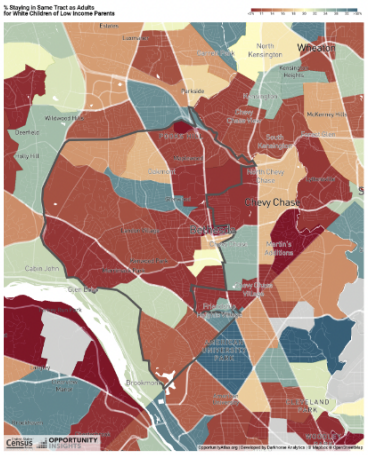
married, spouse income, fraction top 20%, % staying in the same tract, or % staying in the same commuting zone.

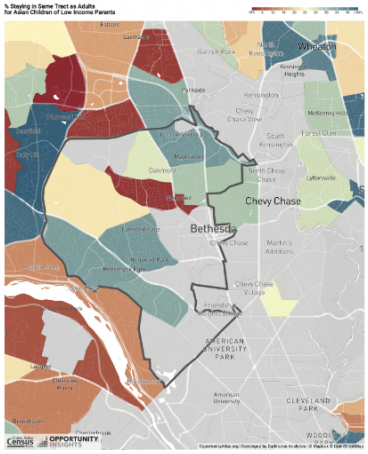
1. Produce a map of your chosen variable for low-income children in your chosen city.

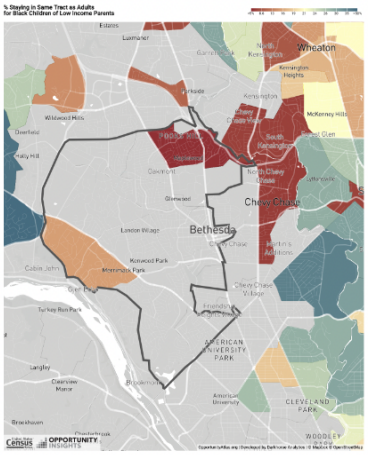


2. Qualitatively, how does your picture in 7.1 differ by race and ethnicity?

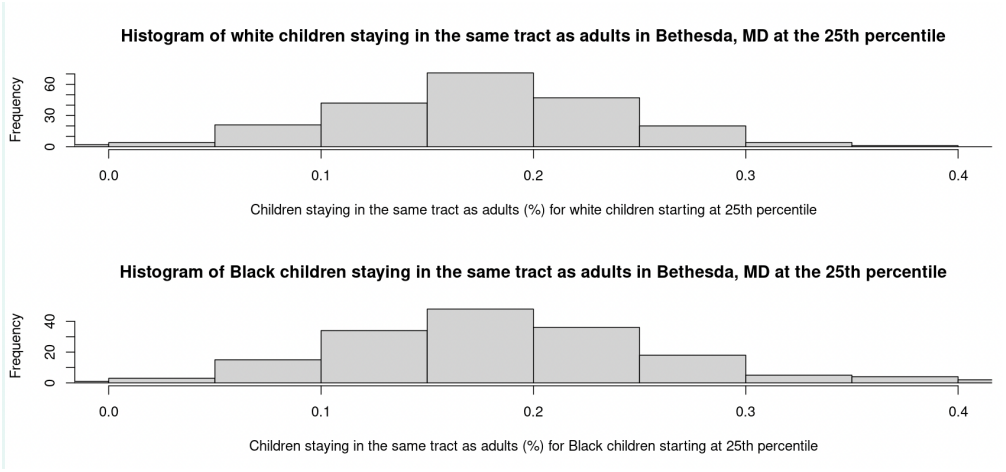








3. Produce a dual-histogram showing the distribution of your variable for low-income children from at least two different races or ethnicities in your county.



4. Calculate the weighted mean of your variable for low-income children in your chosen county for each race.

White: 0.174%

Black: 0.191%

Asian: 0.237%

Hispanic: 0.251%

American Indian: N/A

Interpretation: Weighted mean helps us assess if low-income students of different races have the same likelihood of staying on the same economic level. In the case of Bethesda, MD, that probability doesn't significantly differ across races with Asian and Hispanic children being more likely to stay in the same tract as their parents – however, for Asian children, it's important to note that they are more likely to stay in the already higher average income. For Black and white children, the probability of that saying in the same economic bracket is very low in Bethesda, MD.

5. Calculate the weighted standard deviation of incarceration for low-income children in your chosen county for each race.

White: 0.0588

Black: 0.0630

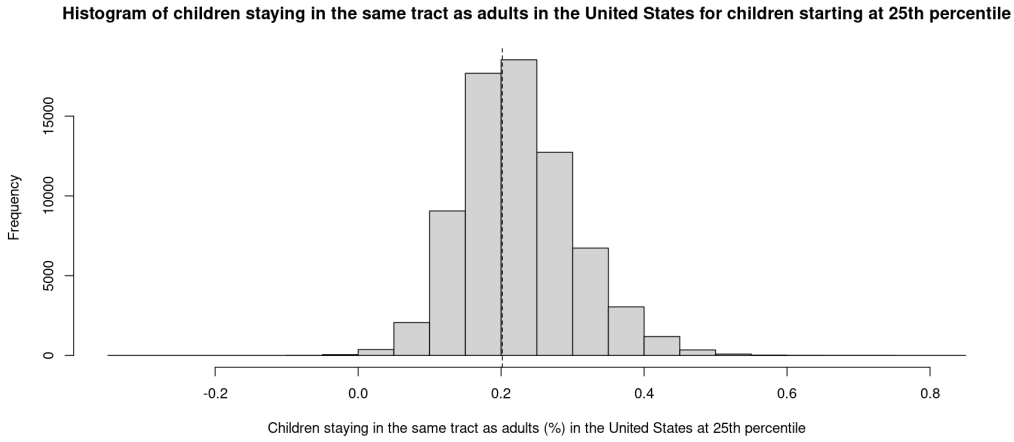
Asian: 0.105

Hispanic: 0.0802

American Indian: N/A

Interpretation: The standard deviation shows the variation of the probability of staying in the same economic tract for low-income children across races. In this case, we see a significantly higher SD for Asian children – who are more likely to stay in the same, but higher!, economic tract, while it's statistically insignificant for the other races. This goes in accordance with the histogram in #3 that shows a not "diverse" or variant distribution of wealth among Black children; Asian children aren't included in the histogram.

6. Using data for the entire USA, plot the distribution of your variable for low-income kids.



Please include your code report here. The simplest way of doing this is to open the html

file (generated from the Download Code button at the end of the HW - see below) and copy all of that text here.

Report - Exercise 1: HW

Name: Sonya Rashkovan

Explore the Data

Box 1:

Question 3

Box 1:

```
Bethesda <- OI_atlas_data_full %>% filter(state == 24 & county == 031)
```

Box 2:

```
layout(matrix(c(1,2), 2, 1, byrow = TRUE)) # this changes the layout of the plot
hist(Bethesda$kfr_white_p25, xlab = "HH income ($) at age 31-37 for white children
starting at 25th percentile", xlim = range(20000:75000), main = "Histogram of
opportunity for white children in Bethesda, MD at the 25th percentile")
hist(Bethesda$kfr_black_p75, xlab = "HH income ($) at age 31-37 for black children
starting at 25th percentile", xlim = range(20000:75000), main = "Histogram of
opportunity for black children in Bethesda, MD at the 25th percentile")
```

Question 4

Box 1:

```
Bethesda %>% summarise(mean = wtd.mean(kfr_white_p25, weights = count_white, na.rm =
TRUE))
```

Box 2:

```
Bethesda %>% summarise(mean = wtd.mean(kfr_black_p25, weights = count_black, na.rm =
TRUE))
```

Box 3:

```
Bethesda %>% summarise(mean = wtd.mean(kfr_asian_p25, weights = count_asian, na.rm =
TRUE))
```

```
Bethesda %>% summarise(mean = wtd.mean(kfr_hisp_p25, weights = count_hisp, na.rm =
```

```
TRUE))


Bethesda %>% summarise(mean = wtd.mean(kfr_natam_p25, weights = count_natam, na.rm =
TRUE))
```

Question 5

Box 1:

```
Bethesda %>% summarise(sd = sqrt(wtd.var(kfr_white_p25, count_white, na.rm = TRUE)))
```

Box 2:

```
Bethesda %>% summarise(sd = sqrt(wtd.var(kfr_black_p25, count_black, na.rm = TRUE)))
```

Box 3:

```
Bethesda %>% summarise(sd = sqrt(wtd.var(kfr_asian_p25, count_asian, na.rm = TRUE)))


Bethesda %>% summarise(sd = sqrt(wtd.var(kfr_hisp_p25, count_hisp, na.rm = TRUE)))


Bethesda %>% summarise(sd = sqrt(wtd.var(kfr_natam_p25, count_natam, na.rm = TRUE)))
```

Question 6

Box 1:

```
hist(OI_atlas_data_full$kfr_pooled_p25, xlab = "HH income ($) at age 35 for children
starting at 25th percentile", main = "Histogram of Opportunity in the United States
at the 25th percentile") abline(v = mean(Bethesda$kfr_pooled_p25, na.rm = TRUE), lty
= 2)
```

Question 7.3

Box 1:

```
Bethesda <- OI_atlas_data_full %>% filter(state == 24 & county == 031)
```

Box 2:

```
layout(matrix(c(1,2), 2, 1, byrow = TRUE)) # this changes the layout of the plot
hist(Bethesda$staytract_white_pooled_p25, xlab = "Children staying in the same tract
as adults (%) for white children starting at 25th percentile", xlim =
range((0:4)/10), main = "Histogram of white children staying in the same tract as
adults in Bethesda, MD at the 25th percentile")
hist(Bethesda$staytract_black_pooled_p25, xlab = "Children staying in the same tract
as adults (%) for Black children starting at 25th percentile", xlim =
range((0:4)/10), main = "Histogram of Black children staying in the same tract as
adults in Bethesda, MD at the 25th percentile")
```

Question 7.4

Box 1:

```
Bethesda %>% summarise(mean = wtd.mean(staytract_white_pooled_p25, weights =
count_white, na.rm = TRUE))
```

Box 2:

```
Bethesda %>% summarise(mean = wtd.mean(staytract_black_pooled_p25, weights =
count_black, na.rm = TRUE))
```

Box 3:

```
Bethesda %>% summarise(mean = wtd.mean(staytract_asian_pooled_p25, weights =
count_asian, na.rm = TRUE))
```

```
Bethesda %>% summarise(mean = wtd.mean(staytract_hisp_pooled_p25, weights =
count_hisp, na.rm = TRUE))
```

```
Bethesda %>% summarise(mean = wtd.mean(staytract_natam_pooled_p25, weights =
count_natam, na.rm = TRUE))
```

Question 7.5

Box 1:

```
Bethesda %>% summarise(sd = sqrt(wtd.var(staytract_white_pooled_p25, count_white,
na.rm = TRUE)))
```

Box 2:

```
Bethesda %>% summarise(sd = sqrt(wtd.var(staytract_black_pooled_p25, count_black,
na.rm = TRUE)))
```

Box 3:

```
Bethesda %>% summarise(sd = sqrt(wtd.var(staytract_asian_pooled_p25, count_asian,
na.rm = TRUE)))
```

```
Bethesda %>% summarise(sd = sqrt(wtd.var(staytract_natam_pooled_p25, count_natam,
na.rm = TRUE)))
```

```
Bethesda %>% summarise(sd = sqrt(wtd.var(staytract_hisp_pooled_p25, count_hisp, na.rm
= TRUE)))
```

Question 7.6

Box 1:

```
hist(OI_atlas_data_full$staytract_pooled_pooled_p25, xlab = "Children staying in the
same tract as adults (%) in the United States at 25th percentile", main = "Histogram
of children staying in the same tract as adults in the United States for children
starting at 25th percentile") abline(v = mean(Bethesda$staytract_pooled_pooled_p25,
na.rm = TRUE), lty = 2)
```