

ECON 125 Problem Set 4 Solution

Question 1

What share of the sample is foreign-born? Draw a graph that shows how this share varies by age and sex. Describe your results. Based on your results, how do you think the age structure of the US adult population would change if the US had no immigrants?

Answer: To assess the foreign-born share overall, we can take the mean of `foreign_born`. We find that 17 percent of the sample is foreign-born.

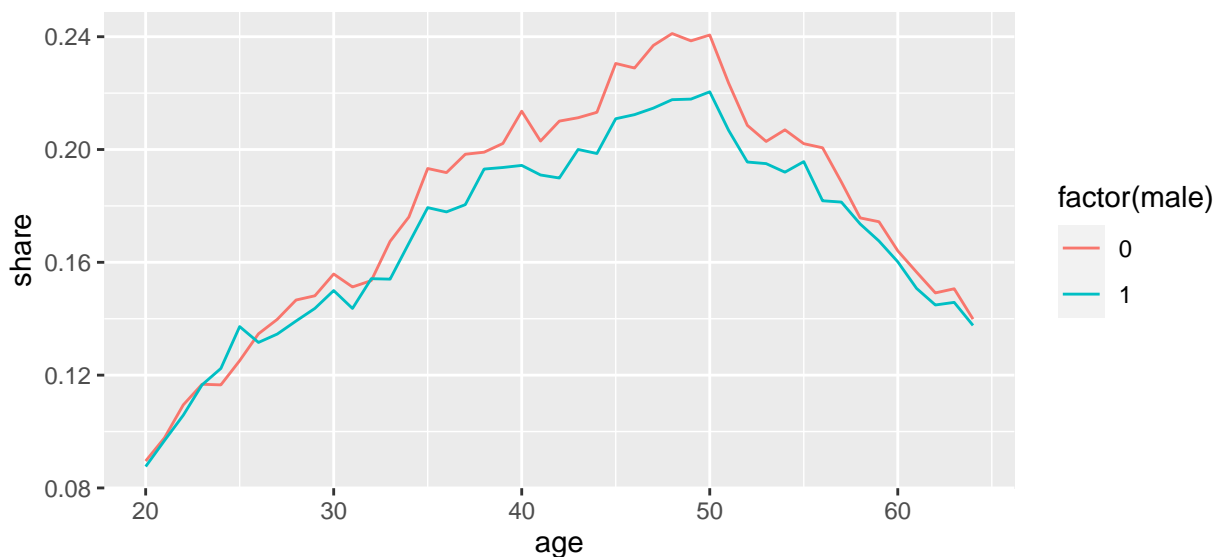
```
acs2023 |> summarise(share = mean(foreign_born))
```

```
## # A tibble: 1 x 1
##   share
##   <dbl>
## 1 0.173
```

To assess how this share varies by age and sex, we create a table using `group_by()` and then draw a graph using `ggplot()`. I chose to put two age profiles in the same plot using `color`. You could have also drawn separate plots using `facet_wrap()`. The foreign-born share rises from 8 percent at age 20 up to more than 20 percent at age 50, after which it steadily falls. The peak for women (24 percent) is higher than the peak for men (22 percent).

```
table <-
  acs2023 |>
  group_by(age, male) |>
  summarise(share = mean(foreign_born))

ggplot(table, aes(x=age, y=share, color=factor(male))) +
  geom_line()
```



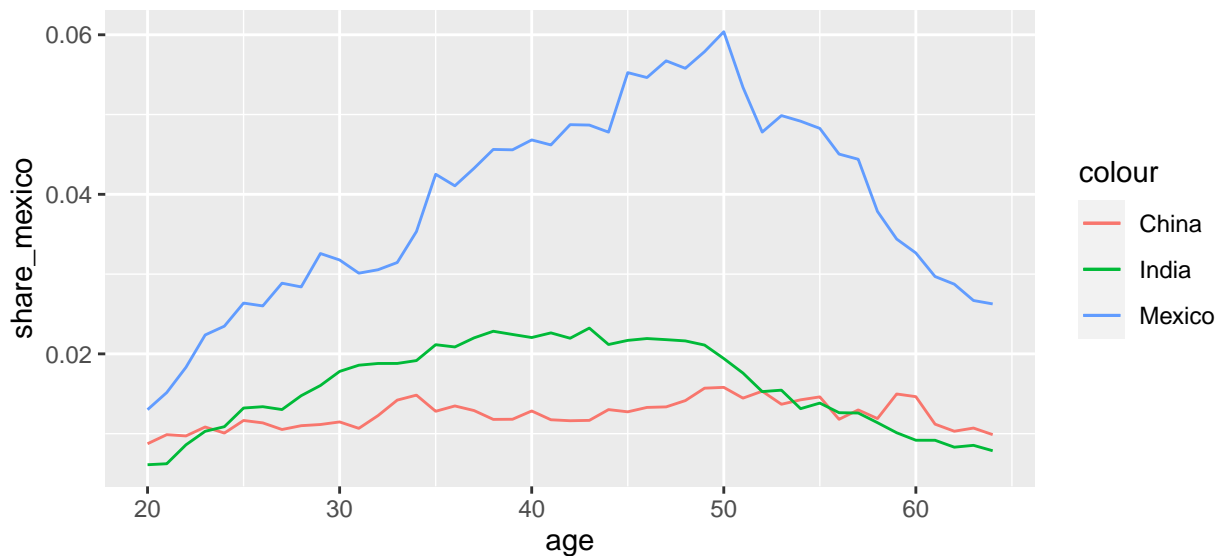
The results imply that the US adult population would have a smaller share of middle aged individuals.

Question 2

Does the age composition of the foreign-born vary by national origin? Plot the share foreign-born by age separately for the three origin countries that account for the most immigrants: Mexico, China, and India. Describe your results. Which country's age pattern is most similar to your graph in Question 1?

Answer: As in Question 1, we create a table and then plot it using `ggplot()`. The syntax is a little different because we will have three y-variables in the graph, one for each of the origin countries. The results show a nearly flat age profile for China, a slightly hump-shaped profile for India, and a steep rise and fall for Mexico. The pattern for Mexico is qualitatively similar to the sample-wide pattern. Note that the wording of this question was a bit confusing, so we accepted other reasonable interpretations of what it was asking.

```
table <-  
  acs2023 |>  
  group_by(age) |>  
  summarise(share_mexico = mean(mexico_born),  
            share_china = mean(china_born),  
            share_india = mean(india_born),)  
  
ggplot(table, aes(x=age)) +  
  geom_line(aes(y=share_mexico, color="Mexico")) +  
  geom_line(aes(y=share_china, color="China")) +  
  geom_line(aes(y=share_india, color="India"))
```



Question 3

Now consider the racial and ethnic composition of the native- and foreign-born. Create a table that reports the share of native-borns in each racial category and the share of foreign-borns in each racial category. Create another table that reports the share of each group in each Hispanic category. Describe your results. Based on your results, how do you think the racial and ethnic composition of the US adult population would change if the US had no immigrants?

Answer: There are many valid ways to approach the problem, but I chose to create one table for natives and one table for immigrants, and then to join them. This approach leads to a nice format with two side-by-side columns.

```
# race table
table_native <- acs2023 |>
  filter(foreign_born==0) |> count(race) |> mutate(native = n/sum(n)) |> select(race, native)
table_foreign <- acs2023 |>
  filter(foreign_born==1) |> count(race) |> mutate(foreign = n/sum(n)) |> select(race, foreign)
left_join(table_native, table_foreign, by = "race")

## # A tibble: 9 x 3
##   race                                native foreign
##   <fct>                                <dbl>   <dbl>
## 1 White                                0.737   0.220
## 2 Black/African American               0.0950  0.0713
## 3 American Indian or Alaska Native    0.0134  0.0139
## 4 Chinese                             0.00535 0.0745
## 5 Japanese                             0.00145 0.00730
## 6 Other Asian or Pacific Islander      0.0163  0.222
## 7 Other race, nec                      0.0395  0.187
## 8 Two major races                     0.0856  0.195
## 9 Three or more major races            0.00680 0.00839

# hispanic table
table_native <- acs2023 |>
  filter(foreign_born==0) |> count(hispan) |> mutate(native = n/sum(n)) |> select(hispan, native)
table_foreign <- acs2023 |>
  filter(foreign_born==1) |> count(hispan) |> mutate(foreign = n/sum(n)) |> select(hispan, foreign)
left_join(table_native, table_foreign, by = "hispan")

## # A tibble: 5 x 3
##   hispan                                native foreign
##   <fct>                                <dbl>   <dbl>
## 1 Not Hispanic 0.883   0.588
## 2 Mexican     0.0721  0.222
## 3 Puerto Rican 0.0166  0.00345
## 4 Cuban        0.00364 0.0255
## 5 Other        0.0243  0.161
```

The results show that 74 percent of native-borns identify as White, compared to 22 percent of foreign born. Meanwhile, 2 percent of native-borns identify as Asian, compared with 30 percent among foreign-borns. Additionally, 12 percent of native-borns identify as Hispanic, compared with 41 percent among foreign-borns. The share identifying as Black is slightly higher among native-borns. Overall, without the foreign-born, the US adult population would be more White and less Asian and Hispanic.

Question 4

Are the foreign-born more or less educated than the native-born? First, compute average years of education for both groups. Second, plot the histogram of years of education for each group, with shares rather than counts on the vertical axis. What do your results tell you about educational differences between immigrants and natives?

Answer: We can compute average years of education for the two groups using `group_by()` and `summarise()`.

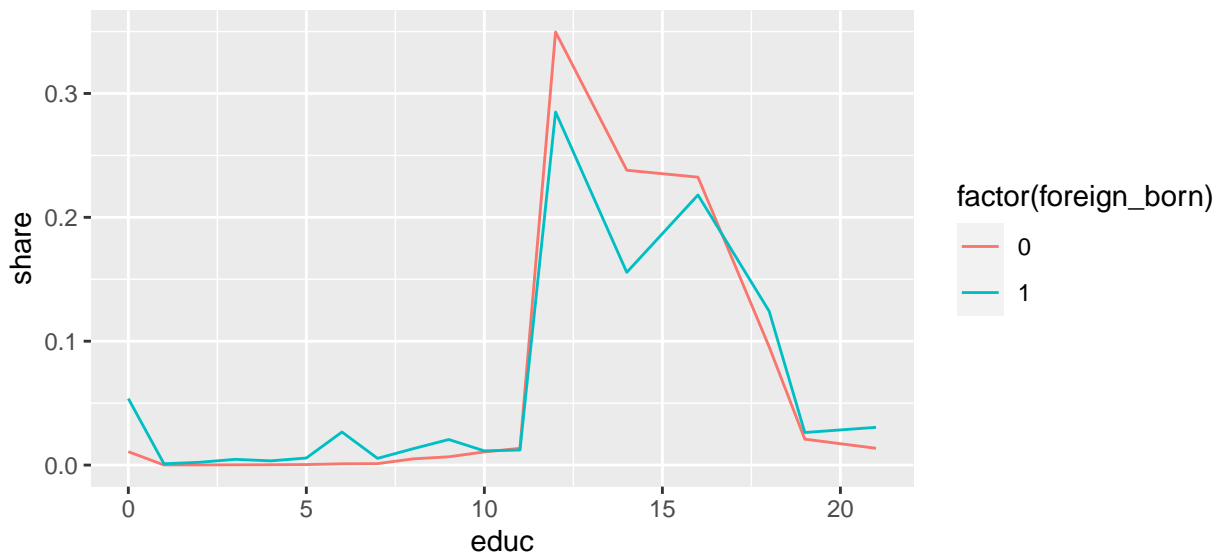
```
acs2023 |> group_by(foreign_born) |> summarise(avg_educ = mean(educ))
```

```
## # A tibble: 2 x 2
##   foreign_born avg_educ
##         <dbl>   <dbl>
## 1             0    14.0
## 2             1    13.3
```

For the histogram, we could use the same approach I used in Question 3, but here is another elegant way using `group_by()`.

```
table <- acs2023 |>
  count(foreign_born, educ) |>
  group_by(foreign_born) |>
  mutate(share = n / sum(n))

ggplot(table, aes(x=educ, y=share, color=factor(foreign_born))) +
  geom_line()
```



The means comparison shows that foreign-born are less educated on average, by three-quarters of a year. The histogram shows that this statement covers over nuanced differences in the distribution of education. The foreign-born are less likely to have high school or college education and more likely to have less than high school **or** more than college.

Question 5

Now investigate the distribution of education levels among immigrants by national origin. Generate dummy variables for the following categories of highest level completed: less than primary (< 5 years), primary (5-11 years), secondary (12-15 years), college (≥ 16 years). Create a table that reports the share of the adult population with each level of education among Mexican immigrants, Chinese immigrants, Indian immigrants, and immigrants from all other countries. What differences do you notice among Mexican, Chinese, and Indian immigrants? How does the distribution of education levels among Mexican immigrants compare with the distribution we saw in the 2020 Mexican census during the methods lecture? Are Mexican immigrants positively or negatively selected?

Answer: I generate the dummy variables specified in the question, and I also create a new `origin` variable for easy tabulation using `group_by()`. The `origin` variable was not necessary but makes it easier to create the table.

```
acs2023 <-  
  acs2023 |>  
  mutate(lt_primary = if_else(educ<5, 1, 0),  
         primary = if_else(educ>=5&educ<12, 1, 0),  
         secondary = if_else(educ>=12&educ<16, 1, 0),  
         college = if_else(educ>=16, 1, 0),  
         origin = case_when(mexico_born==1 ~ "Mexico",  
                           china_born==1 ~ "China",  
                           india_born==1 ~ "India",  
                           TRUE ~ "Other"))  
  
acs2023 |>  
  filter(foreign_born==1) |>  
  group_by(origin) |>  
  summarise(share_lt_primary = mean(lt_primary),  
            share_primary = mean(primary),  
            share_secondary = mean(secondary),  
            share_college = mean(college))
```

```
## # A tibble: 4 x 5  
##   origin share_lt_primary share_primary share_secondary share_college  
##   <chr>         <dbl>         <dbl>         <dbl>         <dbl>  
## 1 China          0.0435          0.0380          0.294          0.624  
## 2 India          0.0338          0.0224          0.198          0.746  
## 3 Mexico         0.131           0.250          0.504          0.115  
## 4 Other          0.0483          0.0567          0.471          0.424
```

Chinese and Indian immigrants are more educated than Mexican immigrants. 62% of Chinese-born individuals and 75% of Indian-born individuals finished college, compared to 12% of Mexican-born individuals. Compared to Mexican-born individuals living in Mexico, Mexican-born individuals in the United States are more likely to have secondary as the highest completed level and less likely to have primary as the highest completed level. Mexican immigrants to the United States are neither positively nor negatively selected; instead, they are drawn from the upper-middle area of the education distribution. (This conclusion is consistent with what we read in the Abramitzky and Boustan article.) Your conclusion would be a bit different if you relied on the first iteration of Week 7 Methods, so we were very flexible in grading the selection part of the question.

Question 6

One sometimes hears the stereotype that immigrants are lazy. On average, are the foreign-born more or less likely to work? To investigate this issue further, plot the share employed by age for native- vs. foreign-born men, and also the share employed by age for native- vs. foreign-born women. Are employment differences between native-borns and foreign-borns similar for different ages and between men and women?

Answer: To answer the first question, we can use `group_by()`. The foreign-born are very slightly more likely to work than the native-born: 76% versus 74%.

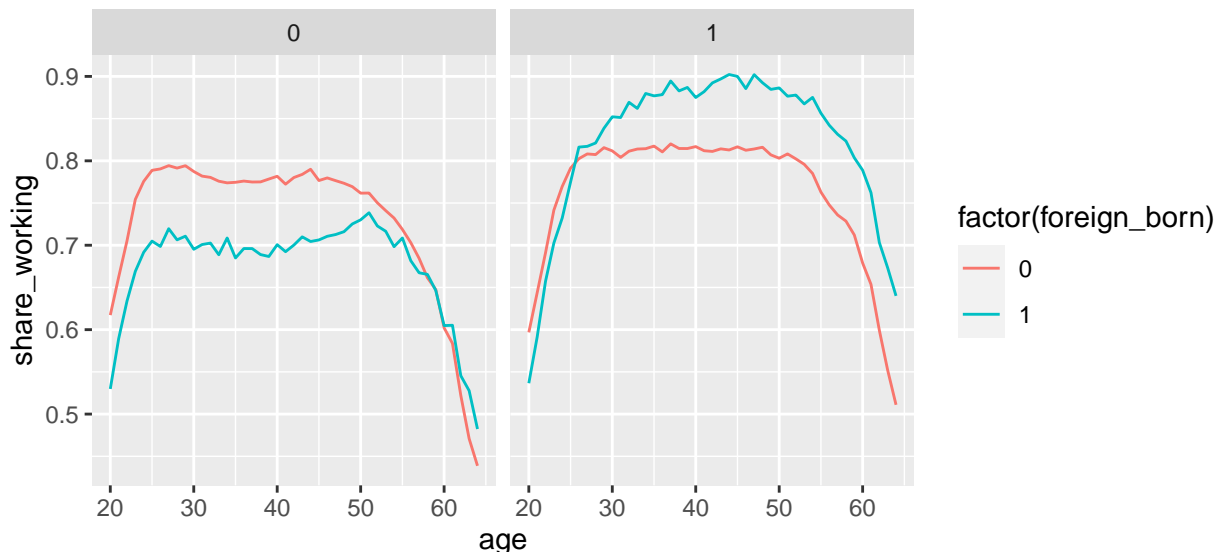
```
acs2023 |>
  group_by(foreign_born) |>
  summarise(share_working = mean(employed))
```

```
## # A tibble: 2 x 2
##   foreign_born share_working
##         <dbl>         <dbl>
## 1           0           0.739
## 2           1           0.755
```

To answer the second question, we can use `group_by()` and then `ggplot()`. Employment gaps differ by sex. At most ages, foreign-born men are **more** likely to work than native-born men, while foreign-born women are **less** likely to work than native-born women. These gaps are smaller at ages 20-29 and 55-64 than at ages in between.

```
table <-
  acs2023 |>
  group_by(age, male, foreign_born) |>
  summarise(share_working = mean(employed))

ggplot(table, aes(x=age, y=share_working, color=factor(foreign_born))) +
  geom_line() +
  facet_wrap(~male)
```



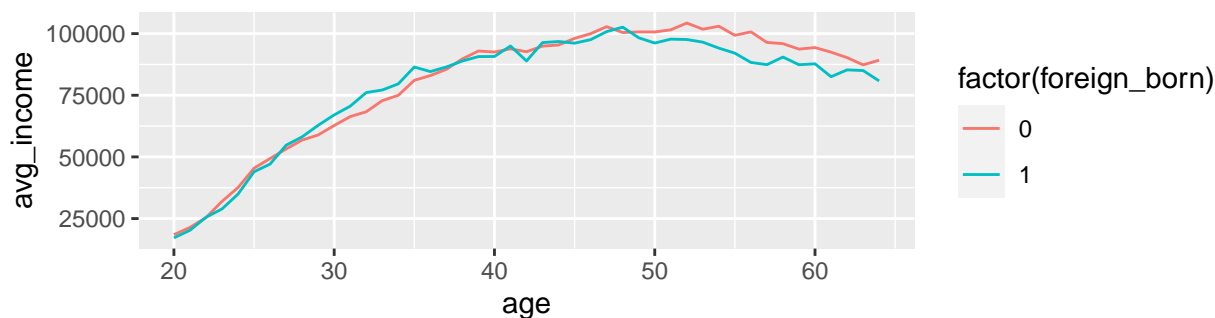
Question 7

Do foreign-born workers earn more or less than native-born workers over the life-cycle? To answer this question, focus on men who have positive earnings. In a first graph, plot average earnings by age for native-borns vs. foreign-borns. Are there large earnings gaps at any ages? In a second graph, repeat the same exercise, but with separate panels for men with and without college degrees. Do any new insights emerge?

Answer: For the first graph, we `filter()` to men with positive earnings and then `group_by()` age and foreign-born status. We then use `ggplot()` to graph the two age profiles. Earnings are similar at every age: slightly higher for immigrants in the 30s and somewhat higher for natives in the 50s and 60s.

```
# pooling all male workers
table <-
  acs2023 |>
  filter(income>0&male==1) |>
  group_by(age, foreign_born) |>
  summarise(avg_income = mean(income))

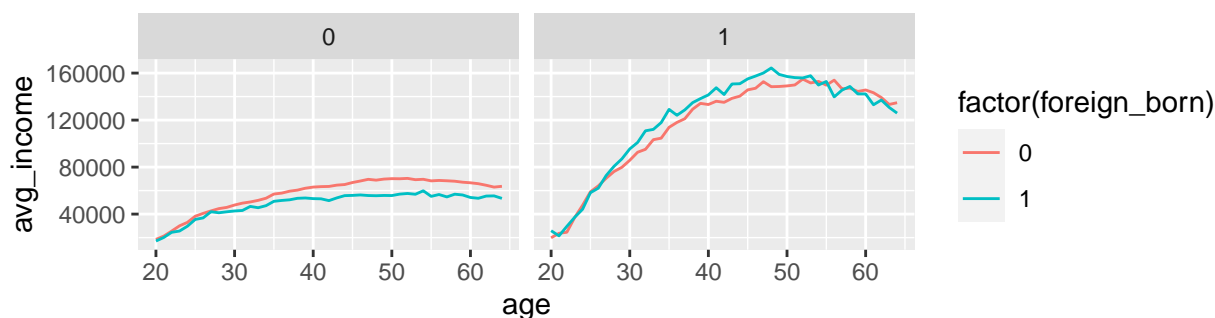
ggplot(table, aes(x=age, y=avg_income, color=factor(foreign_born))) +
  geom_line()
```



For the second graph, we add `college` to `group_by()`. Now we see a very different story. College-educated immigrant men outearn college-educated native men at most ages. However, among male workers without college, natives outearn immigrants at every age.

```
# separating college and non-college workers
table <-
  acs2023 |>
  filter(income>0, male==1) |>
  group_by(age, foreign_born, college) |>
  summarise(avg_income = mean(income))

ggplot(table, aes(x=age, y=avg_income, color=factor(foreign_born))) +
  geom_line() +
  facet_wrap(~college)
```



Question 8

Another stereotype posits that immigrants rely on welfare benefits more than natives. Is the share receiving positive welfare income higher among native-borns or foreign-borns? Is this evidence consistent with the stereotype?

Answer: We generate a dummy variable for any welfare income and then use `group_by()`. Foreign-born adults are **less** likely than native-born adults to receive welfare income. This result is inconsistent with the stereotype.

```
acs2023 |>
  mutate(welfare = if_else(incwelfr>0, 1, 0)) |>
  group_by(foreign_born) |>
  summarise(share = mean(welfare))
```

```
## # A tibble: 2 x 2
##   foreign_born share
##         <dbl> <dbl>
## 1           0 0.0143
## 2           1 0.0119
```


Question 9

Many people believe that although immigrants have low earnings when they arrive in the US, they catch up with natives eventually. How do average earnings compare between immigrants who arrived in the last year and immigrants who arrived earlier? Do these differences tell us whether immigrant earnings catch up after sufficient time in the US?

Answer: To compare average earnings among recent and earlier arrivals, we `group_by()` the variable `mig1yr`.

```
# create table
acs2023 |>
  filter(foreign_born==1) |>
  group_by(mig1yr) |>
  summarise(avg_earnings = mean(incearn))
```

```
## # A tibble: 2 x 2
##   mig1yr avg_earnings
##   <dbl>      <dbl>
## 1     0      58193.
## 2     1      26874.
```

We find that immigrants who arrived in the last year average less than half the earnings of immigrants who arrived earlier. However, this difference does not tell us about catch-up patterns because it could be driven by cohort effects rather than time-since-migration effects. You could have also noted that to study catch-up, we would also need to use data on natives.

Question 10

How would the characteristics of the US population change if all foreign-born people left the country? Compute the share Black, share Hispanic, share Asian, share college-educated, and share employed for the full sample and the native-born subsample. Which characteristics, if any, would change substantively?

Answer: I start by generating dummy variables for Black, Hispanic, and Asian. This is not strictly necessary, but it makes the question easier to answer using `group_by()`.

```
# generate dummy variables for Black, Hispanic, Asian
acs2023 <-
  acs2023 |>
  mutate(black = if_else(race=="Black/African American", 1, 0),
         hispanic = if_else(hispan!="Not Hispanic", 1, 0),
         asian = if_else(race=="Chinese" |
                        race=="Japanese" |
                        race=="Other Asian or Pacific Islander",
                        1, 0))
```

Now, I generate two tables of means, one for the full sample and one excluding immigrants. The results show that without immigrants, the population would have substantially smaller shares of Hispanic and Asian individuals, but the share Black as well as the college and employment shares would not change much.

```
# with immigrants
acs2023 |>
  summarise(share_black = mean(black),
            share_hispanic = mean(hispanic),
            share_asian = mean(asian),
            share_college = mean(college),
            share_working = mean(employed))
```

```
## # A tibble: 1 x 5
##   share_black share_hispanic share_asian share_college share_working
##   <dbl>         <dbl>         <dbl>         <dbl>         <dbl>
## 1      0.0909      0.168      0.0719      0.369      0.742
```

```
# without immigrants
acs2023 |>
  filter(foreign_born==0) |>
  summarise(share_black = mean(black),
            share_hispanic = mean(hispanic),
            share_asian = mean(asian),
            share_college = mean(college),
            share_working = mean(employed))
```

```
## # A tibble: 1 x 5
##   share_black share_hispanic share_asian share_college share_working
##   <dbl>         <dbl>         <dbl>         <dbl>         <dbl>
## 1      0.0950      0.117      0.0231      0.362      0.739
```