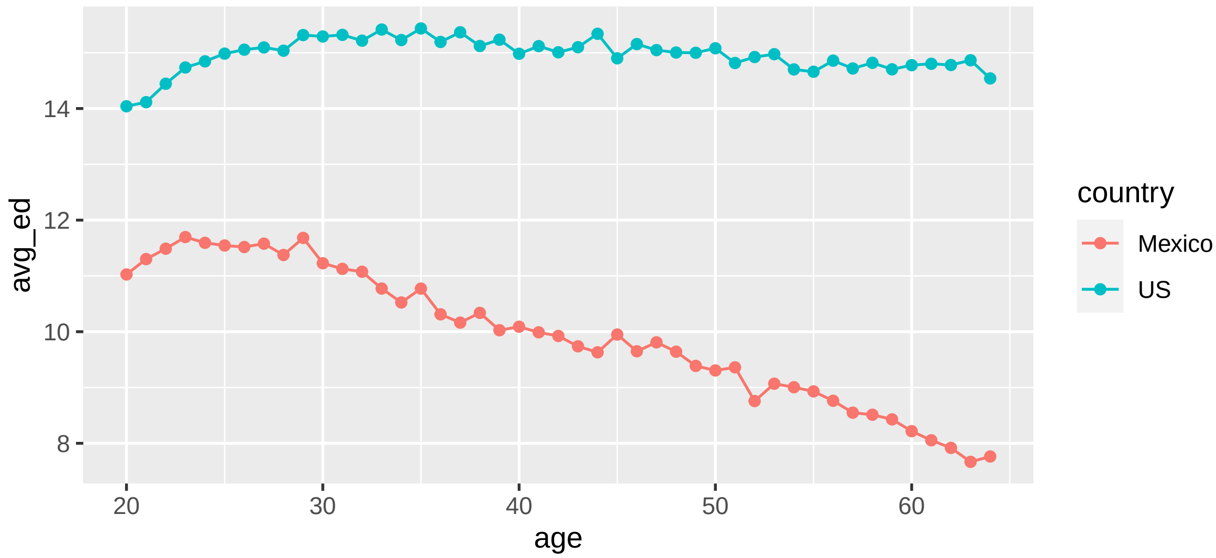
ECON 125 Problem Set 5 Solution

Question 1

Using data from only 2019 in both datasets, compute average years of education at each age. Plot the average on the vertical axis against age on the horizontal axis. At which ages do you see patterns that would most plausibly be described as age effects? At which ages do you see patterns that would most plausibly be described as cohort effects? For the latter age range, redraw the graph with birth year rather than age on the horizontal axis. What can you say about cross-cohort trends in educational attainment in the US and Mexico?

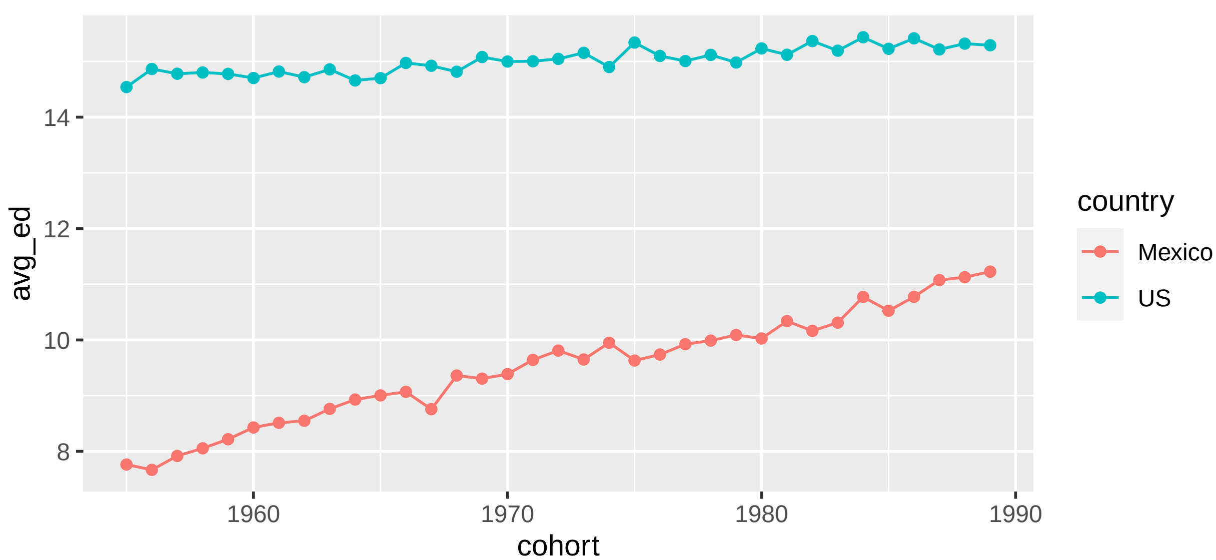
Answer: I create a table of age-specific means for the US in 2019, and the same for Mexico in 2019. I combine the US and Mexico tables so I can plot them together, but this is not necessary. The increasing pattern in the 20s is consistent with continuing accumulation of education as individuals get older, an age effect. Past age 30, educational attainment declines with age in both countries, consistent with cross increases in educational attainment, a cohort effect. (You might have correctly noted that the cohort effect starts dominating in the mid-20s in Mexico.)

```r
us <-
  cps |>
  filter(year==2019) |>
  group_by(age) |>
  summarise(avg_ed = mean(edyears)) |>
  mutate(country = "US")

mex <-
  enoe |>
  filter(year==2019) |>
  group_by(age) |>
  summarise(avg_ed = mean(edyears)) |>
  mutate(country = "Mexico")

combined <- bind_rows(us, mex)

ggplot(combined, aes(x=age, y=avg_ed, color=country)) +
  geom_point() +
  geom_line()
```



For ages 30+, I plot average educational attainment against birth year. Education trends upward across cohorts in both countries, but more steeply in Mexico.

```
combined <-
  combined |>
  filter(age>=30) |>
  mutate(cohort = 2019-age)

ggplot(combined, aes(x=cohort, y=avg_ed, color=country)) +
  geom_point() +
  geom_line()
```

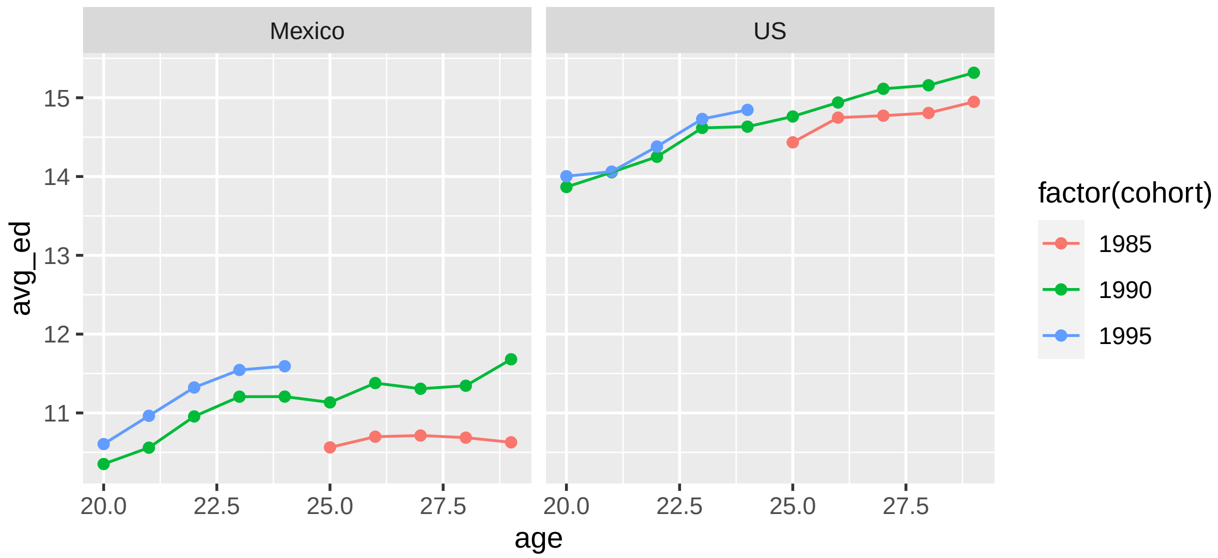


Question 2

Now focus on the ages you left out of the cohort figure. Using data from all years in both datasets, plot cohort age profiles of average educational attainment over this age range for birth years ending in 0 or 5. Do the age profiles have positive or negative slopes? Comparing two cohorts at a given age, does the later-born cohort have higher or lower average education than the earlier-born cohort? How do these patterns influence the cross-sectional plot for 2019 in Question 1?

Answer: For ages 20-29 in the 1985, 1990, and 1995 cohorts, I plot the path of educational attainment by age and cohort. The age profiles in Mexico have a positive slope in the early 20s, but then flatten out. The age profiles in the US have a positive slope throughout the age range. At the same age, later-born cohorts have more education than earlier-born cohorts in both countries. The age effects make the slope positive at early ages in Question 1, but the cohort effects work in the opposite direction, weakening the positive slope.

```
us <-
  cps |>
  filter(age<30) |>
  group_by(age, year) |>
  summarise(avg_ed = mean(edyears)) |>
  mutate(country = "US", cohort = year-age)

mex <-
  enoe |>
  filter(age<30) |>
  group_by(age, year) |>
  summarise(avg_ed = mean(edyears)) |>
  mutate(country = "Mexico", cohort = year-age)

combined <- bind_rows(us, mex) |> filter(cohort%%5==0)

ggplot(combined, aes(x=age, y=avg_ed, color=factor(cohort))) +
  geom_point() +
  geom_line() +
  facet_wrap(~country)
```

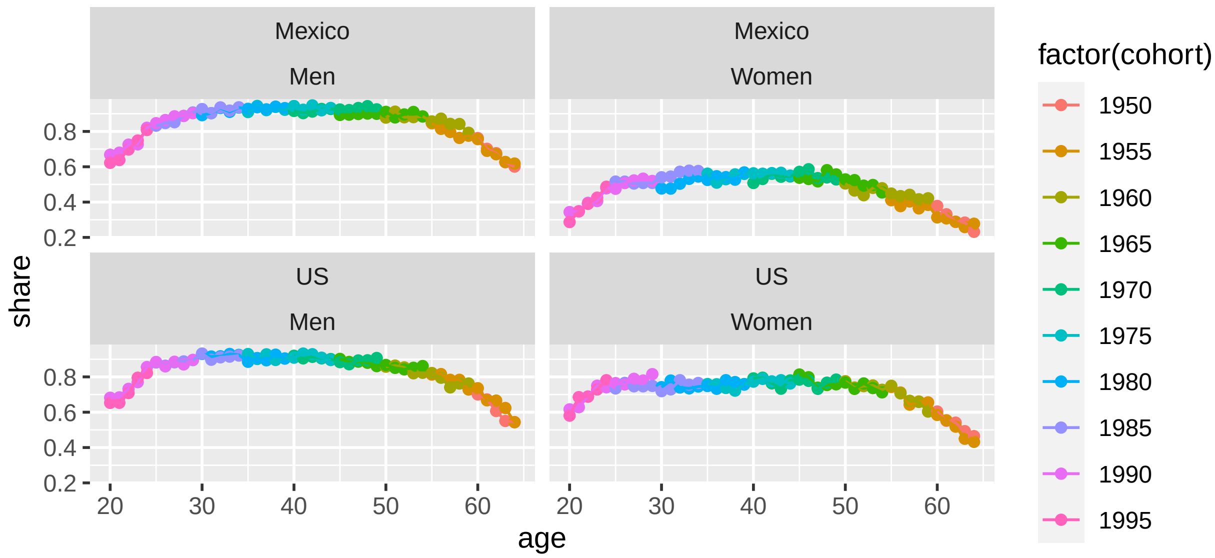


Question 3

For men and women in each country separately, draw cohort age profiles of the share employed for birth years ending in 0 or 5. Is there evidence of age effects? Cohort effects? At any given age, are men more or less likely to work than women?

Answer: I generate a dummy variable for working based in `classwk`. For each country, I create a table of the share working by age, year, and sex; I then define `cohort` as `year-age` and filter to cohorts ending in 0 or 5. I combine the country tables and plot by cohort. All plots show a clear hump-shaped age pattern, with no systematic level differences between cohorts. Thus, they show evidence of age effects but not cohort effects. Men are more likely to work than women at every age.

```
us <-
  cps |>
  mutate(work = if_else(classwk!="Not working", 1, 0),
         sex = if_else(male==1, "Men", "Women")) |>
  group_by(age, year, sex) |>
  summarise(share = mean(work)) |>
  mutate(country = "US", cohort = year-age)

mex <-
  enoe |>
  mutate(work = if_else(classwk!="Not working", 1, 0),
         sex = if_else(male==1, "Men", "Women")) |>
  group_by(age, year, sex) |>
  summarise(share = mean(work)) |>
  mutate(country = "Mexico", cohort = year-age)

combined <- bind_rows(us, mex) |> filter(cohort%%5==0)

ggplot(combined, aes(x=age, y=share, color=factor(cohort))) +
  geom_point() +
  geom_line() +
  facet_wrap(~country + sex)
```



Question 4

For the rest of the problem set, restrict both samples to men who work as wage/salary workers for at least 30 hours a week (or 1560 hours a year) with positive earnings. Generate an hourly earnings variable in both datasets and call it `wage`. Adjust the `wage` variable so it is denominated in 2010 currency units. In 2010, 10 Mexican pesos bought 1 US dollar's worth of goods and services. Convert the `wage` variable in Mexico so it is denominated in dollars. What is each country's average wage in 2010 dollars?

Answer: I restrict the sample as instructed and calculate the hourly wage. In the Mexican data, income is per month, while hours are per week, so I multiply hours by 4.3. I adjust for inflation and then divide Mexican wages by 10 to convert to dollars. The average American wage is 26 dollars per hour, while the average Mexican wage is 2.6 dollars per hour.

```
cps <- cps |>
  filter(classwk=="Wage/salary worker" & hours>=1560 & male==1 & incwage>0) |>
  mutate(wage = incwage/hours) |>
  mutate(wage = case_when(year==2011 ~ wage/1.03,
                          year==2012 ~ wage/1.05,
                          year==2013 ~ wage/1.07,
                          year==2014 ~ wage/1.09,
                          year==2015 ~ wage/1.09,
                          year==2016 ~ wage/1.10,
                          year==2017 ~ wage/1.12,
                          year==2018 ~ wage/1.15,
                          year==2019 ~ wage/1.17,
                          TRUE ~ wage))

enoe <- enoe |>
  filter(classwk=="Wage/salary worker" & hours>=30 & male==1 & incwage>0) |>
  mutate(wage = incwage/(4.3*hours)) |>
  mutate(wage = case_when(year==2011 ~ wage/1.03,
                          year==2012 ~ wage/1.08,
                          year==2013 ~ wage/1.12,
                          year==2014 ~ wage/1.16,
                          year==2015 ~ wage/1.19,
                          year==2016 ~ wage/1.23,
                          year==2017 ~ wage/1.30,
                          year==2018 ~ wage/1.37,
                          year==2019 ~ wage/1.42,
                          TRUE ~ wage)) |>
  mutate(wage = wage/10)

cps |> summarise(mean_wage = mean(wage))
```

```
## # A tibble: 1 x 1
##   mean_wage
##       <dbl>
## 1      26.1
```

```
enoe |> summarise(mean_wage = mean(wage))
```

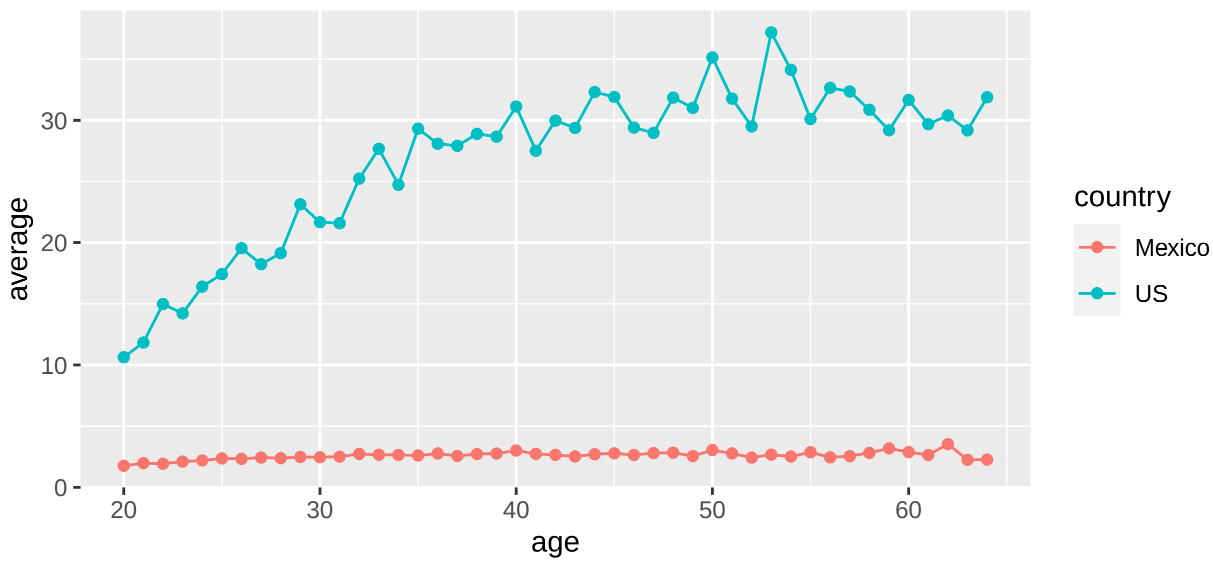
```
## # A tibble: 1 x 1
##   mean_wage
##       <dbl>
## 1      2.61
```

Question 5

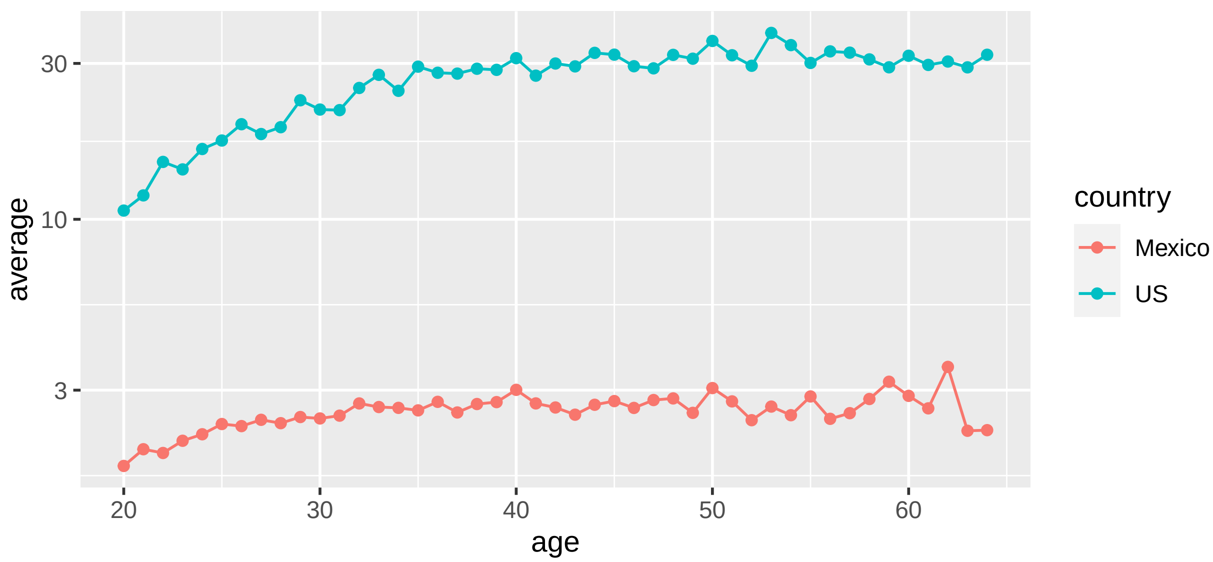
Draw the cross-sectional age profile of average wages for each country in 2019. Redraw the graph with a log scale. Which version is more informative? Why?

Answer: I calculate average wages by age for each country in 2019 and then plot them. The version with a log scale is more informative because it shows more detail for the Mexican plot. In the first graph, without a log scale, the Mexican plot appears flat because the vertical axis covers such a wide range. In the second graph, with a log scale, it becomes clear that the Mexican plot is increasing and concave, just like the American plot.

```r
us <-
  cps |>
  filter(year==2019) |>
  group_by(age) |>
  summarise(average = mean(wage)) |>
  mutate(country = "US")

mex <-
  enoe |>
  filter(year==2019) |>
  group_by(age) |>
  summarise(average = mean(wage)) |>
  mutate(country = "Mexico")

combined <- bind_rows(us, mex)

ggplot(combined, aes(x=age, y=average, color=country)) +
  geom_point() +
  geom_line()
```




```r
ggplot(combined, aes(x=age, y=average, color=country)) +
  geom_point() +
  geom_line() +
  scale_y_log10()
```


30
10
3
average
20
30
40
50
60
age
country
Mexico
US

Question 6

In 2019, what was the ratio of the average US wage to the average Mexico wage at age 20? At age 50? Provide one age-based explanation and one cohort-based explanation for the larger ratio at age 50.

Answer: I approach this problem by renaming variables in my existing US and Mexico tables, joining them by age, and then computing the ratio. At age 20, the average US wage is 6 times the average US wage. At age 50, the average US wage is 12 times the average US wage. An age-based explanation is that the return to work experience is higher in the United States. A cohort-based explanation is that the US-Mexico gap in education is larger for the 1969 cohort than for the 1999 cohort.

```
us <- us |> select(age, average) |> rename(average_us = average)
mex <- mex |> select(age, average) |> rename(average_mex = average)

left_join(us, mex, by = "age") |>
  mutate(ratio = average_us/average_mex) |>
  filter(age==20|age==50)
```

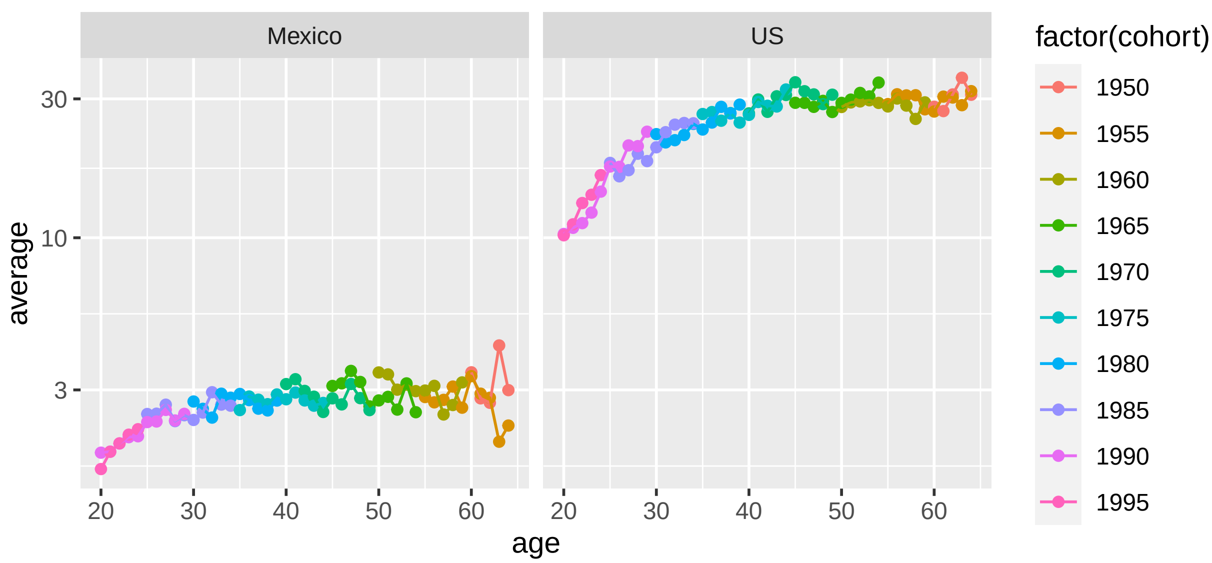
```
## # A tibble: 2 x 4
##     age average_us average_mex ratio
##   <dbl>      <dbl>       <dbl> <dbl>
## 1    20       10.6        1.76  6.05
## 2    50       35.1        3.04 11.5
```

Question 7

For each country, draw cohort age profiles of average wages for birth years ending in 0 or 5. Use a log scale. Do the cohort age profiles look similar to or different from the cross-sectional age profiles in Question 5? Do you see evidence of large cohort effects? Which of your explanations in Question 6 do you think is more likely to be correct?

Answer: I create tables by age and cohort for each country, and I then combine the country tables as before. I then draw the plots with a different color for each cohort, and I use `facet_wrap()` to create separate panels for each country. The resulting cohort age profiles look similar to the cross-sectional age profiles from Question 5. There are no noticeable level gaps between cohorts, implying the absence of large chort effects. This suggests that the age-based explanation in Question 6 is more likely to be correct.

```
us <-
  cps |>
  mutate(cohort = year - age) |>
  group_by(age, cohort) |>
  summarise(average = mean(wage)) |>
  mutate(country = "US")

mex <-
  enoe |>
  mutate(cohort = year - age) |>
  group_by(age, cohort) |>
  summarise(average = mean(wage)) |>
  mutate(country = "Mexico")

combined <- bind_rows(us, mex) |> filter(cohort%%5==0)

ggplot(combined, aes(x=age, y=average, color=factor(cohort))) +
  geom_point() +
  geom_line() +
  scale_y_log10() +
  facet_wrap(~country)
```

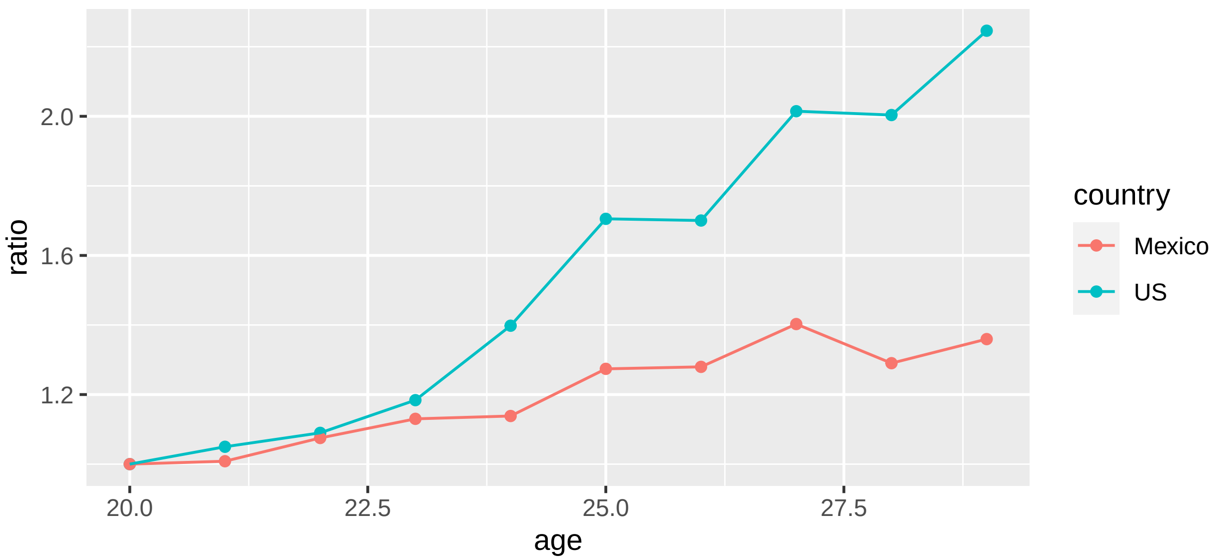


Question 8

Now zoom in on a single cohort, born in 1990. For this cohort only, compute the ratio of the average wage at each age to the average wage at age 20. Draw a graph that has this ratio on the vertical axis and age on the horizontal axis, with one plot for the US and one plot for Mexico. Do average wages rise more steeply over the life-cycle in the United States or Mexico?

Answer: I combine the US and Mexico tables from Question 7, filter to the 1990 cohort, and then compute the ratio as in the Week 9 Methods slides. Average wages grow more steeply over the life-cycle in the United States.

```
combined <-
  bind_rows(us, mex) |>
  filter(cohort==1990) |>
  group_by(country) |>
  arrange(age) |>
  mutate(ratio = average / first(average)) |>
  ungroup()

ggplot(combined, aes(x=age, y=ratio, color=country)) +
  geom_point() +
  geom_line()
```

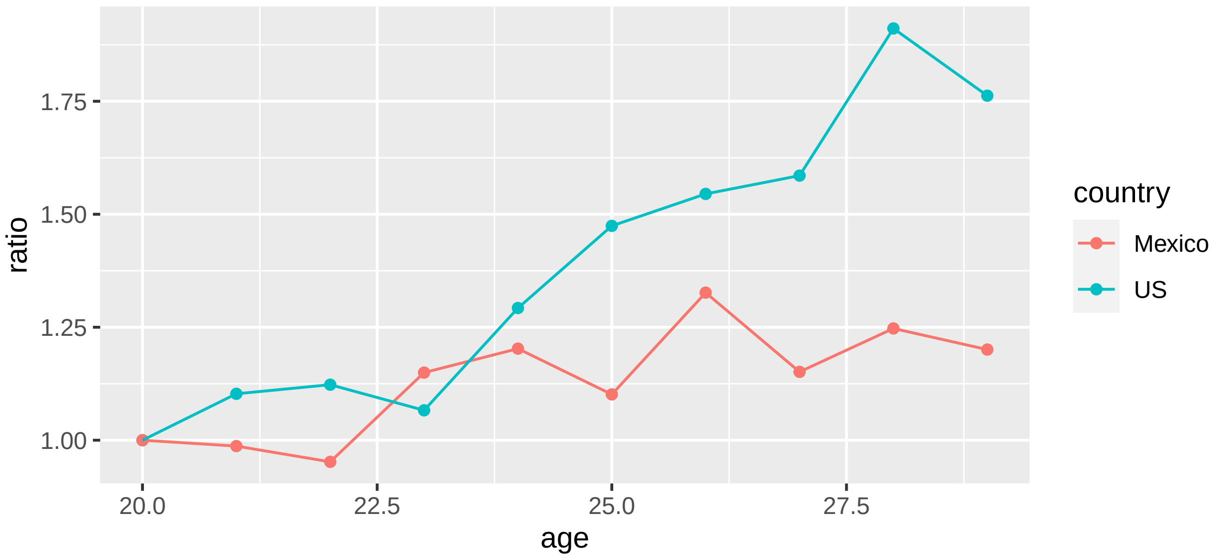


Question 9

Economists are inclined to interpret the results in Question 8 as returns to potential work experience ("potential" because we do not know whether a worker worked in every year). One critique of this interpretation is that workers of the same age can have very different amounts of potential work experience depending on when they left school. Another critique is that college graduates enter the sample after age 20, so the sample composition at age 29 may be different from the sample composition at age 20. To address these critiques, redraw the graph from Question 8 using only high school graduates. In the US, you should use workers with exactly 13 years of education (kindergarten plus 12 grades). In Mexico, you should use workers with exactly 12 years of education. Do your results suggest that the critiques are valid?

Answer: I recreate the figure from Question 8, but with an additional `filter()` to keep only men with exactly 13 (US) or 12 (Mexico) years of education. The qualitative conclusions are similar to before, with steeper wage growth in the US. But the quantitative magnitude of wage growth in this restricted sample is smaller. For example, in Question 8, average wages at age 29 were roughly 2.25 times larger than average wages at age 20 in the US. Now, the ratio is 1.75. A plausible explanation for this is that the average for 29 year-olds in Question 8 included high-earning college graduates, but the average for 20 year-olds did not. The entry of college graduates in the mid-20s makes wage growth appear larger than it is in reality.

```r
us <-
  cps |>
  filter(edyears==13) |>
  mutate(cohort = year - age) |>
  group_by(age, cohort) |>
  summarise(average = mean(wage)) |>
  mutate(country = "US")

mex <-
  enoe |>
  filter(edyears==12) |>
  mutate(cohort = year - age) |>
  group_by(age, cohort) |>
  summarise(average = mean(wage)) |>
  mutate(country = "Mexico")

combined <-
  bind_rows(us, mex) |>
  filter(cohort==1990) |>
  group_by(country) |>
  arrange(age) |>
  mutate(ratio = average / first(average)) |>
  ungroup()

ggplot(combined, aes(x=age, y=ratio, color=country)) +
  geom_point() +
  geom_line()
```


ratio
1.75
1.50
1.25
1.00
20.0
22.5
25.0
27.5
age
country
Mexico
US

Question 10

For high school graduates born in 1990, what is the proportional return to 9 years of potential work experience in the two countries? What do you think might drive the cross-country difference in experience returns? (The second question is open-ended, and a wide range of answers would be valid.)

Answer: The original question asked about 10 years of potential experience, but some of you noted that we can only answer for 9. We were lenient in grading if there was confusion on this point. The age 29 to age 20 radio is 1.76 in the US and 1.20 in Mexico, implying a 76% US return and a 20 percent Mexico return. One possibility is that because of greater access to technology, there is more on-the-job learning in the US.

```r
combined |>
  filter(age==29) |>
  select(age, country, ratio)
```

```
## # A tibble: 2 x 3
##     age country ratio
##   <dbl> <chr>   <dbl>
## 1    29 US       1.76
## 2    29 Mexico   1.20
```