2B. Evidence of Discrimination?

January 22, 2019

1 Evidence of Discrimination?

The Department of Developmental Services (DDS) in California is responsible for allocating funds to support over 250,000 developmentally-disabled residents. The data set `https://raw.githubusercontent.com/dlsun/data-science-book/master/data/ca_dds_expenditures.csv` contains data about 1,000 of these residents. The data comes from a discrimination lawsuit which alleged that California's Department of Developmental Services (DDS) privileged white (non-Hispanic) residents over Hispanic residents in allocating funds. We will focus on comparing the allocation of funds (i.e., expenditures) for these two ethnicities only, although there are other ethnicities in this data set.

There are 6 variables in this data set:

- Id: 5-digit, unique identification code for each consumer (similar to a social security number and used for identification purposes)
- Age Cohort: Binned age variable represented as six age cohorts (0-5, 6-12, 13-17, 18-21, 22-50, and 51+)
- Age: Unbinned age variable
- Gender: Male or Female
- Expenditures: Dollar amount of annual expenditures spent on each consumer
- Ethnicity: Eight ethnic groups (American Indian, Asian, Black, Hispanic, Multi-race, Native Hawaiian, Other, and White non-Hispanic)

2 Question 1

Read in the data set. Make a graphic that compares the *average* expenditures by the DDS on Hispanic residents and white (non-Hispanic) residents. Comment on what you see.

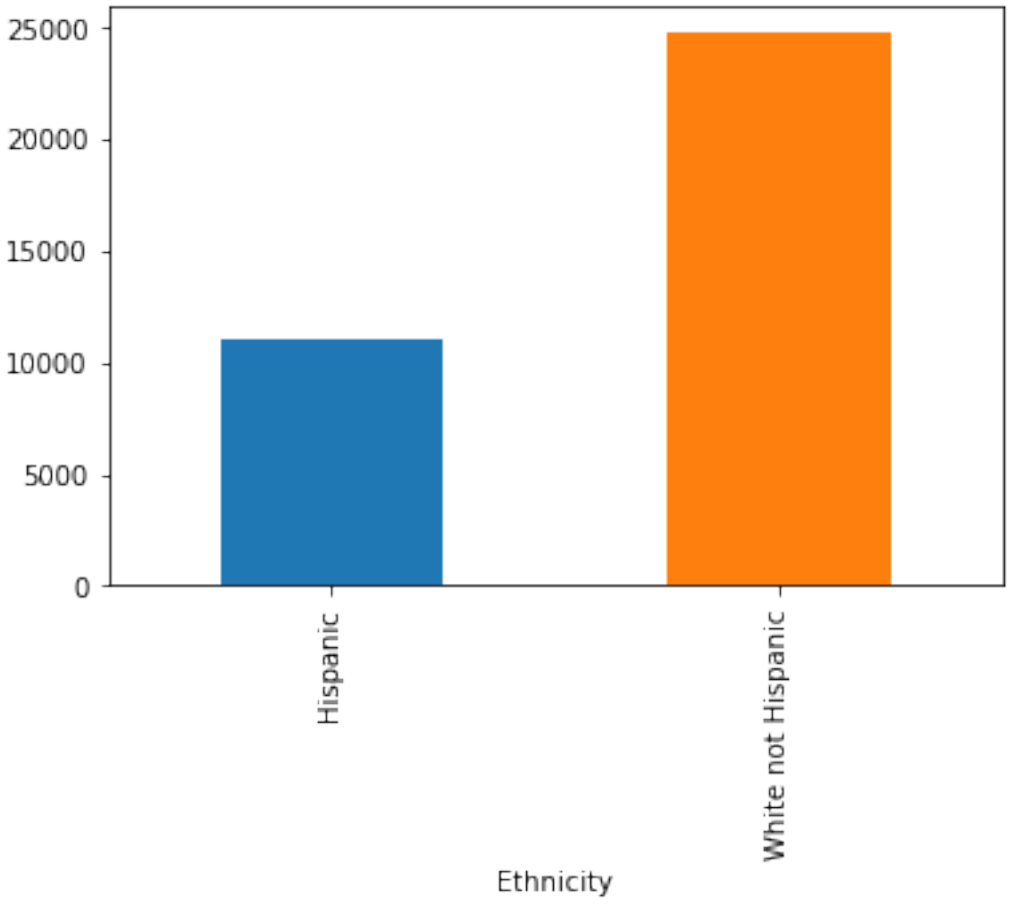
```
In [1]: import pandas as pd
        %matplotlib inline

        dds = pd.read_csv("https://raw.githubusercontent.com/dlsun/data-science-book/master/dat

        white_and_hispanic_only = dds[(dds.Ethnicity ==
                        "White not Hispanic") | (dds.Ethnicity ==
                                                 "Hispanic")]

        white_and_hispanic_only.groupby("Ethnicity")["Expenditures"].mean().plot.bar()
```

```
Out[1]: <matplotlib.axes._subplots.AxesSubplot at 0x7f2512d61be0>
```



According to the bar graph, we see that the average expenditures for White residents is 2x greater than the average expenditures for Hispanice residents.

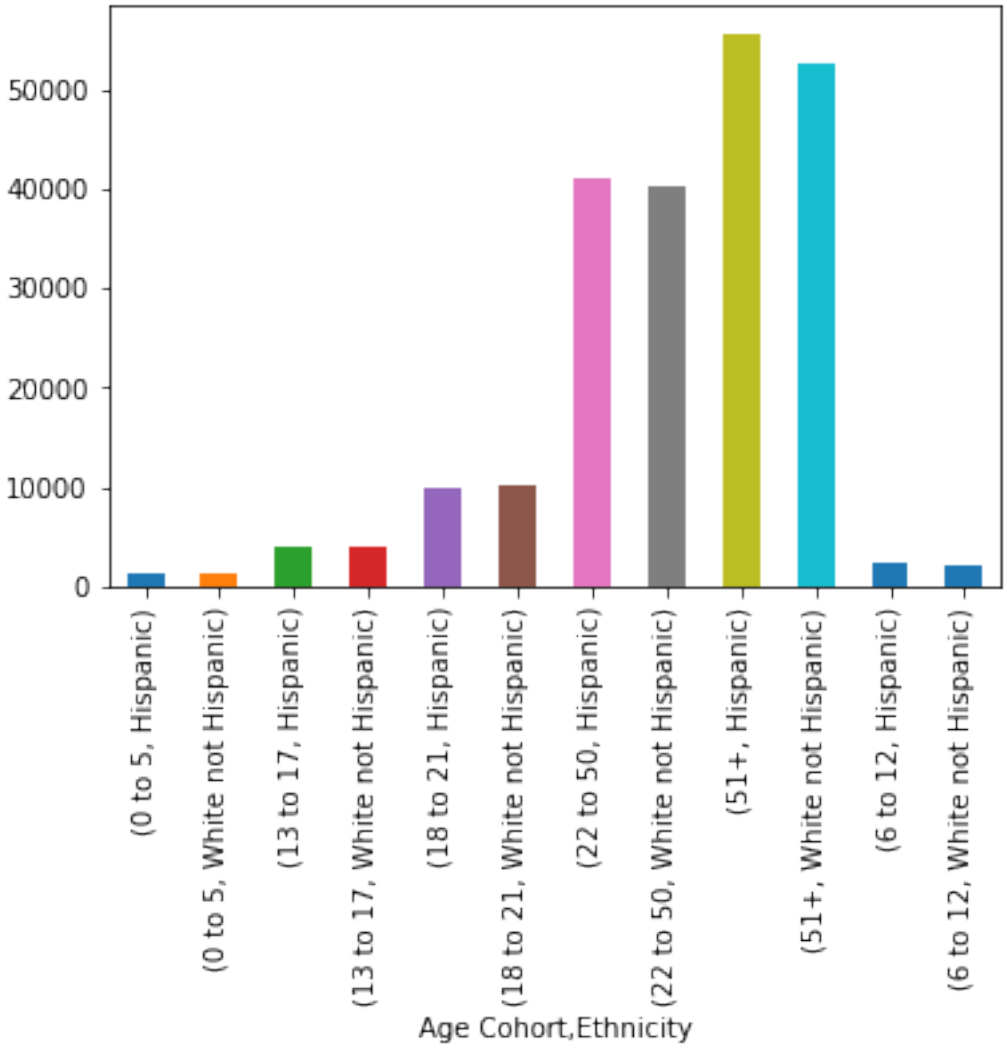
3 Question 2

Now, calculate the average expenditures by ethnicity and age cohort. Make a graphic that compares the average expenditure on Hispanic residents and white (non-Hispanic) residents, *within each age cohort*.

Comment on what you see. How do these results appear to contradict the results you obtained in Question 1?

```
In [2]: white_and_hispanic_only.groupby(["Age Cohort",
                        "Ethnicity"])["Expenditures"].mean().plot.bar()
```

```
Out[2]: <matplotlib.axes._subplots.AxesSubplot at 0x7f25108b20f0>
```



The average expenditures by age cohort and ethnicity is pracitically even at each age group. These results contradict the results obtained by Question 1, because it appears here that funding is evenly allocated. However, this distribution is analyzing the average amount of funding allocated and not the how many people are receiving funding.

4 Question 3

Can you explain the discrepancy between the two analyses you conducted above (i.e., Questions 1 and 2)? Try to tell a complete story that interweaves tables, graphics, and explanation.

Hint: You might want to consider looking at:

- the distributions of ages of Hispanics and whites
- the average expenditure as a function of age

Because there is a discrepancy in the 2 plots above, let's first take a look at the count of people receiving expenditures for white and Hispanic residents.

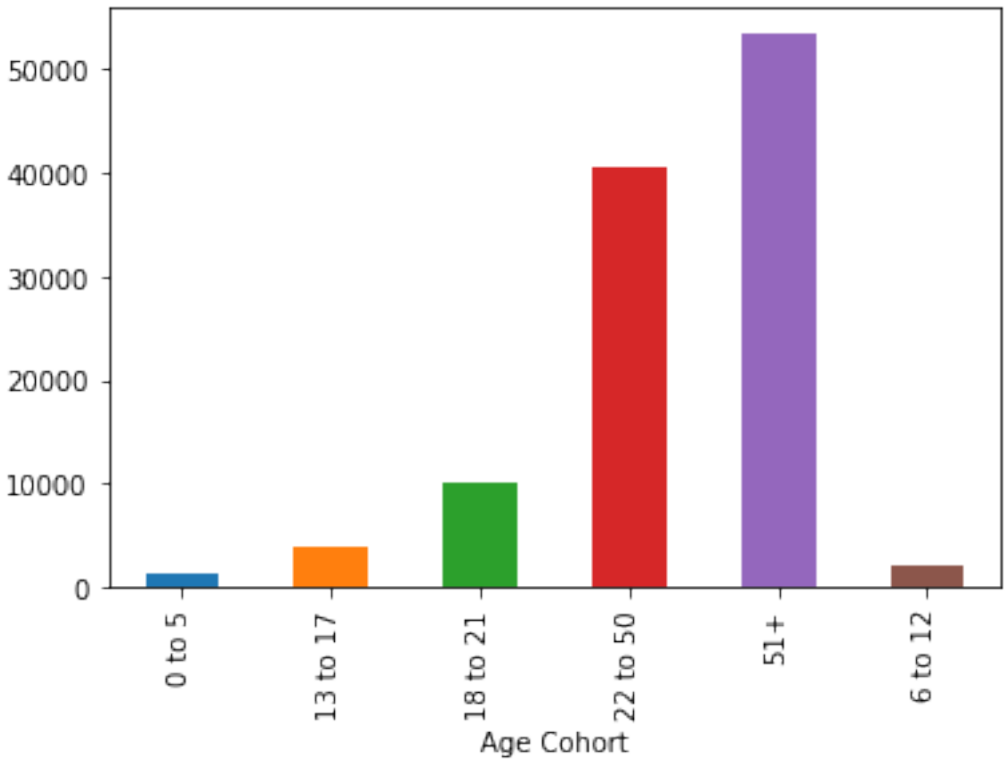
```
In [3]: white_and_hispanic_only.Ethnicity.value_counts()

Out[3]: White not Hispanic    401
        Hispanic              376
        Name: Ethnicity, dtype: int64
```

It turns out that nearly the same number of Hispanic people and White people are receiving expenditures from the DDS. This implies that perhaps the DDS is not discriminating against Hispanic people. So, we will next take a look at the distribution of expenditures across the age cohorts to see if different age groups of people are receiving expenditures more than other age groups.

```
In [4]: white_and_hispanic_only.groupby("Age Cohort")["Expenditures"].mean().plot.bar()

Out[4]: <matplotlib.axes._subplots.AxesSubplot at 0x7f25107ba7f0>
```

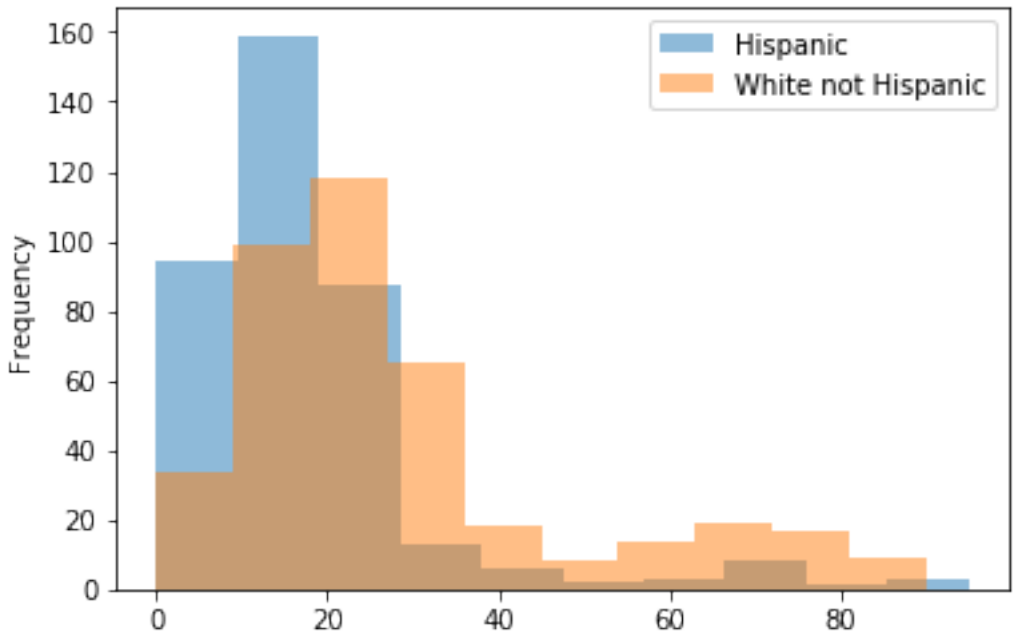


According to the bar plot, we see that the "22 to 50" and "51+" age cohorts receive more expenditures than any other age groups. This leads us to question whether or not White residents are receiving more expenditures in these 2 particular age cohorts in comparison to Hispanic residents.

```
In [5]: white_and_hispanic_only.groupby("Ethnicity")["Age"].plot.hist(alpha = 0.5,
                                                                      legend = True)

        white_and_hispanic_only.groupby(["Age Cohort","Ethnicity"])["Expenditures"].count()
```

```
Out[5]: Age Cohort  Ethnicity
        0 to 5      Hispanic               44
                    White not Hispanic     20
        13 to 17    Hispanic              103
                    White not Hispanic     67
        18 to 21    Hispanic               78
                    White not Hispanic     69
        22 to 50    Hispanic               43
                    White not Hispanic    133
        51+         Hispanic               17
                    White not Hispanic     66
        6 to 12     Hispanic               91
                    White not Hispanic     46
        Name: Expenditures, dtype: int64
```



According to the distribution, we do see that from ages 20 and onward there are many more White residents receiving expenditures than Hispanic residents. As a matter of fact, there are approximately 3x more White residents receiving expenditures from the DDS in the "22 to 50" cohort than Hispanic residents. White residents are also receiving expenditures in the "51+" cohort 4x more than Hispanic residents. Therefore, it does appear that the DDS allocated more funds to White residents than Hispanic residents because for the 2 age cohorts that received the most expenditures, White residents dominated the allocation of funds by 4x more and 3x more. However, we need to ask, "Are the White residents in these 2 age cohorts receiving more money because they need the expenditures more, or is there an equal amount of Hispanic residents in these 2 age cohorts that need the expenditures?"

4.1 Submission Instructions

Once you are finished, follow these steps:

1. Restart the kernel and re-run this notebook from beginning to end by going to `Kernel > Restart Kernel and Run All Cells`.
2. If this process stops halfway through, that means there was an error. Correct the error and repeat Step 1 until the notebook runs from beginning to end.
3. Double check that there is a number next to each code cell and that these numbers are in order.

Then, submit your lab as follows:

1. Go to `File > Export Notebook As > PDF`.
2. Double check that the entire notebook, from beginning to end, is in this PDF file. (If the notebook is cut off, try first exporting the notebook to HTML and printing to PDF.)
3. Upload the PDF to PolyLearn.