1. Project Statement

Motivation:

Gender pay gap is a hot issue and we have examined it closely more than once already.

We found out that women need to take a step back to support their children therefore less working hour for them. Also, women tend to choose and stay on jobs that pay lower like health-care aides, receptionists, cashiers and food servers. There are other causes that I can’t list them all here but all this discussion got me thinking, what about race? Does race have anything to do with our paycheck? Furthermore, as a woman myself, is it being minority worsen my situation or improve it? So, I want to add the ‘race’ card into the play so that women or minority group can have an understanding where they should stand.

Objective:

By explore the factor ‘race’, I shall present how ‘race’ have decreased or widen the pay gap.

Another objective is trying to purposefully distort the data and deceive the audience.  If this objective is successfully achieved, we will examine future visualization objectively and take great caution when developing our own visualization product.

Data

I found a spread sheet on the Internet titled ‘Salary By Race and Sex’

Link to data source

<http://maamodt.asp.radford.edu/HR%20Statistics/Salary%20by%20Sex%20and%20Race.htm>

Project Plan:

Search data for related visualization topic ‘Racial Pay Gap’

Get the data and develop the intermediate deceptive claim and the opposite claim

Data Wrangling

Final Visualization for deceptive claim and the opposite

2. “Making-of” documentation (Details of your development process, data wrangling steps, your reasoning,

detours, literature, etc.)

Development Process:

The data I collected is from 2002 to 2011 US Median Salary by SEX and Race.

First I extracted the data I needed in excel and add a new row ‘All Minority’ for every year.

It looks like this.







Second step:

Load this data into Tableau

Third step:

Data Wrangling in Jupyter Notebook

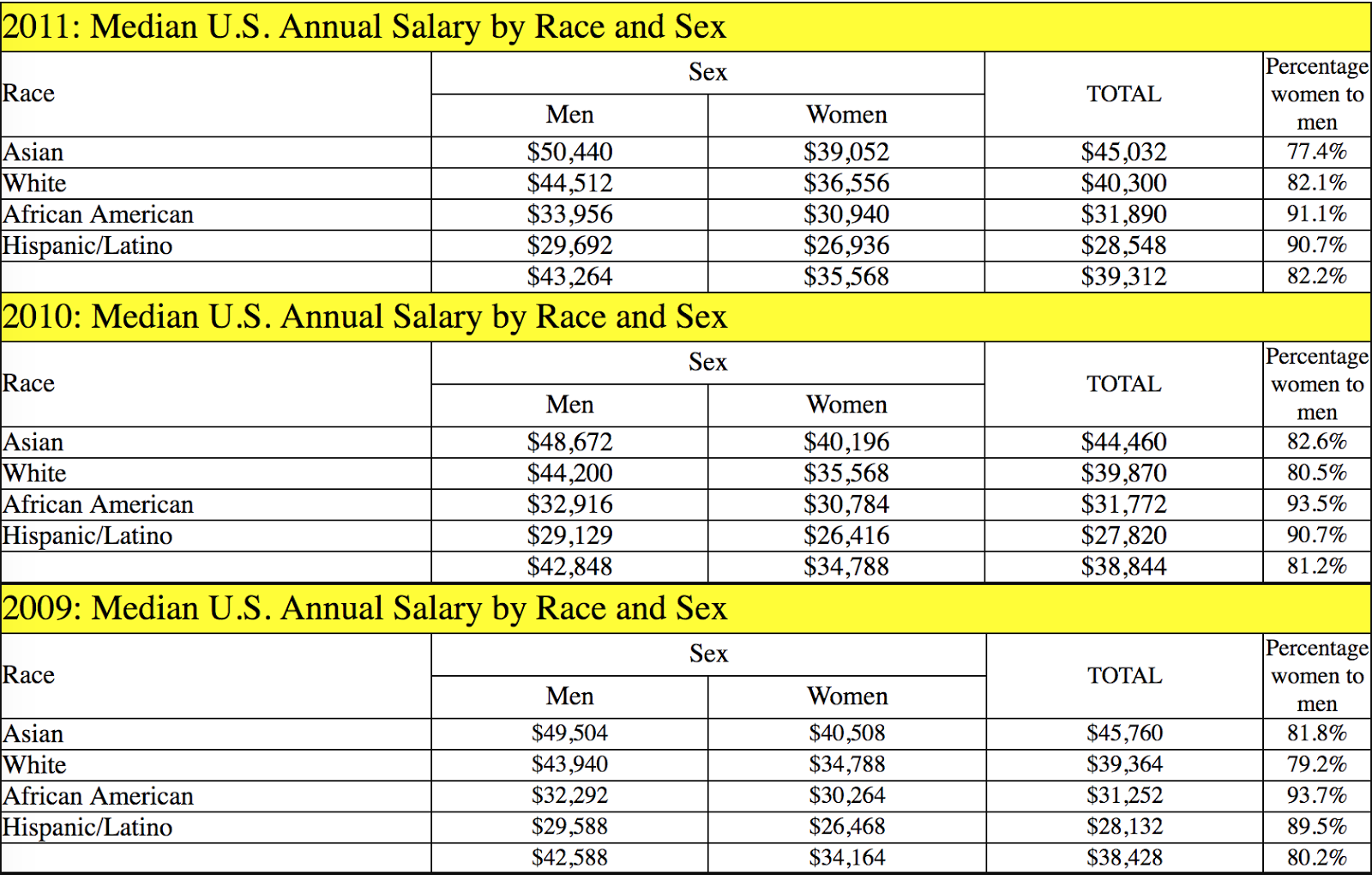
Fourth step:

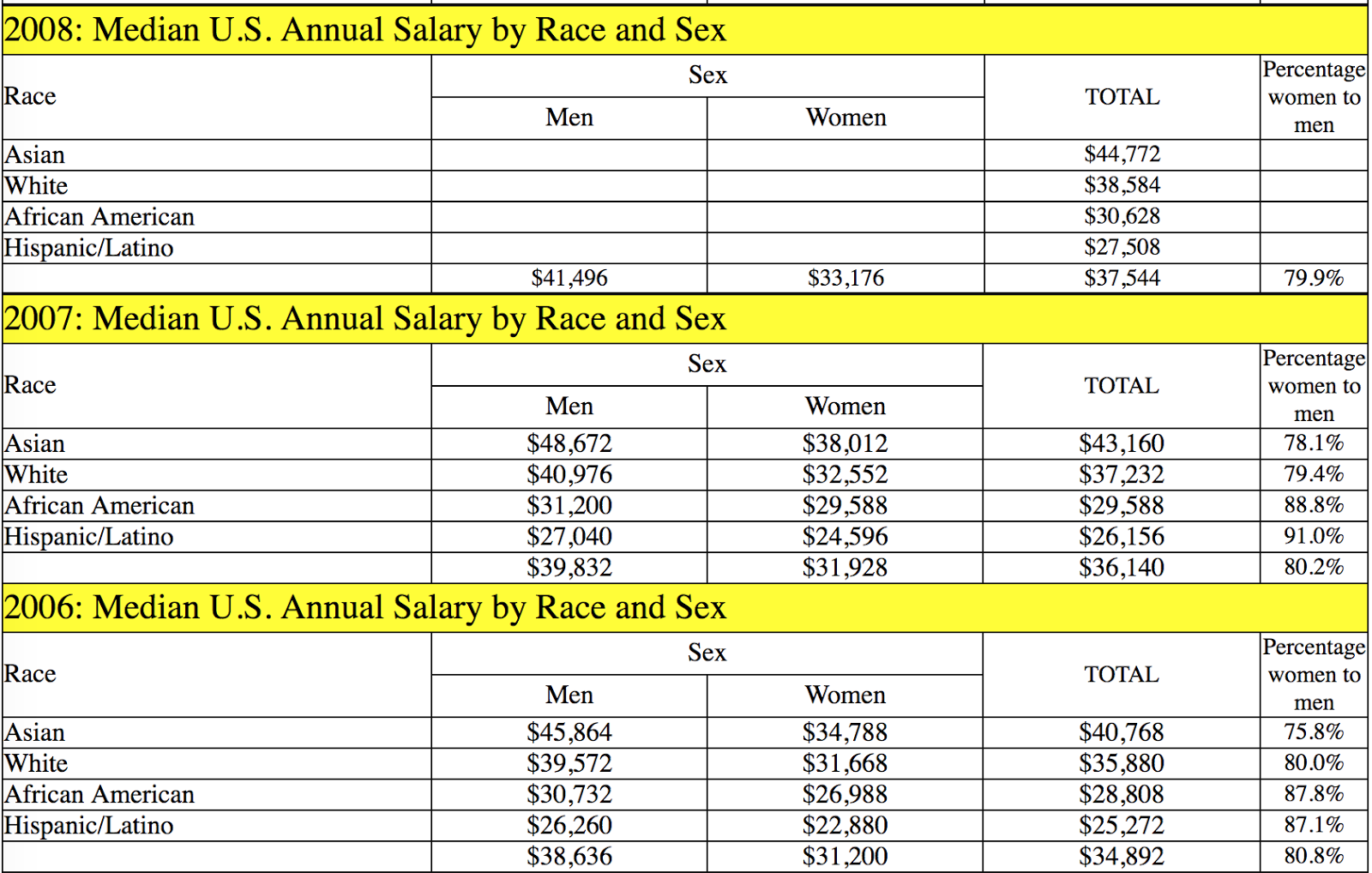
Reload saved csv file into Tableau and refine the visualization

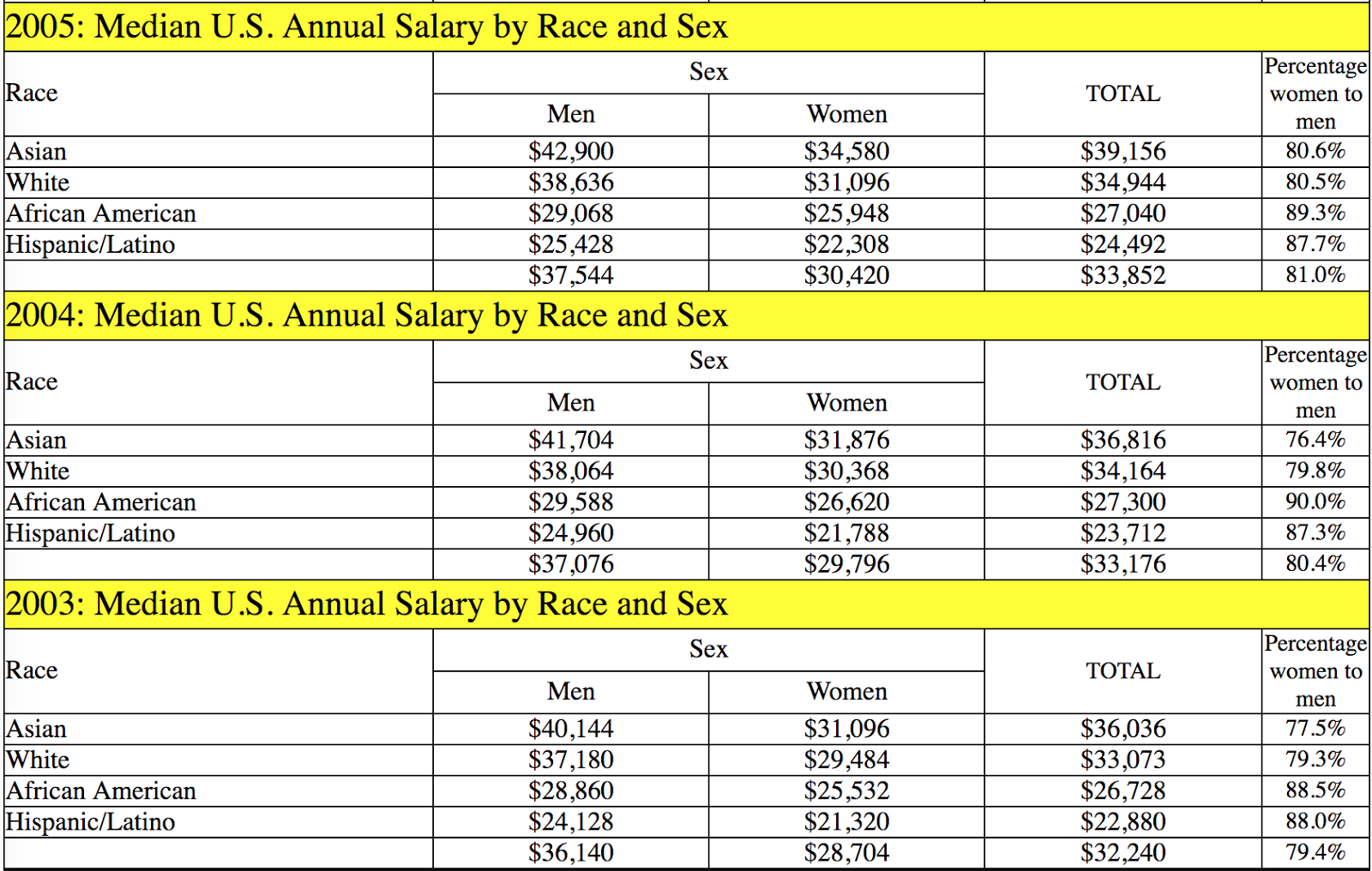
Data Wrangling Steps:

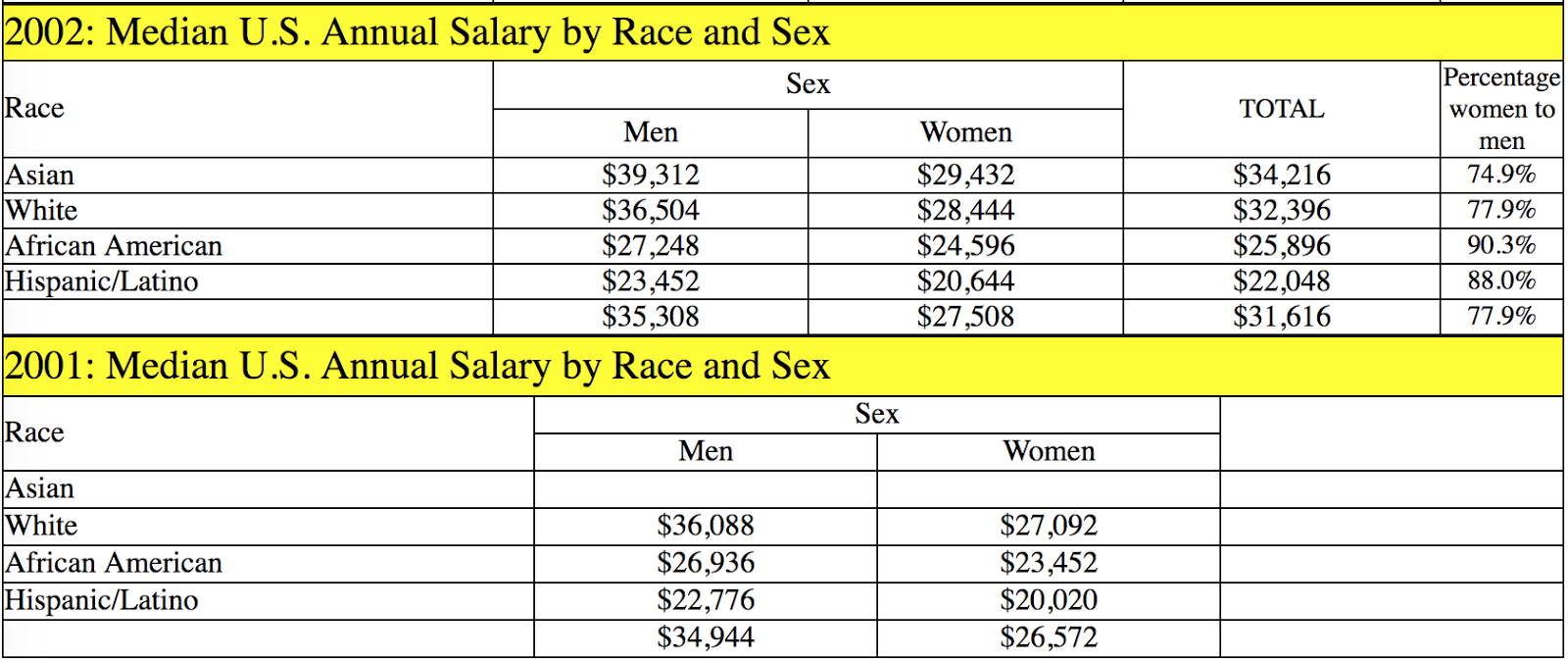
After searching ‘Race Pay Gap’ on google, I found this spreadsheet on this website

http://maamodt.asp.radford.edu/HR%20Statistics/Salary%20by%20Sex%20and%20Race.htm.





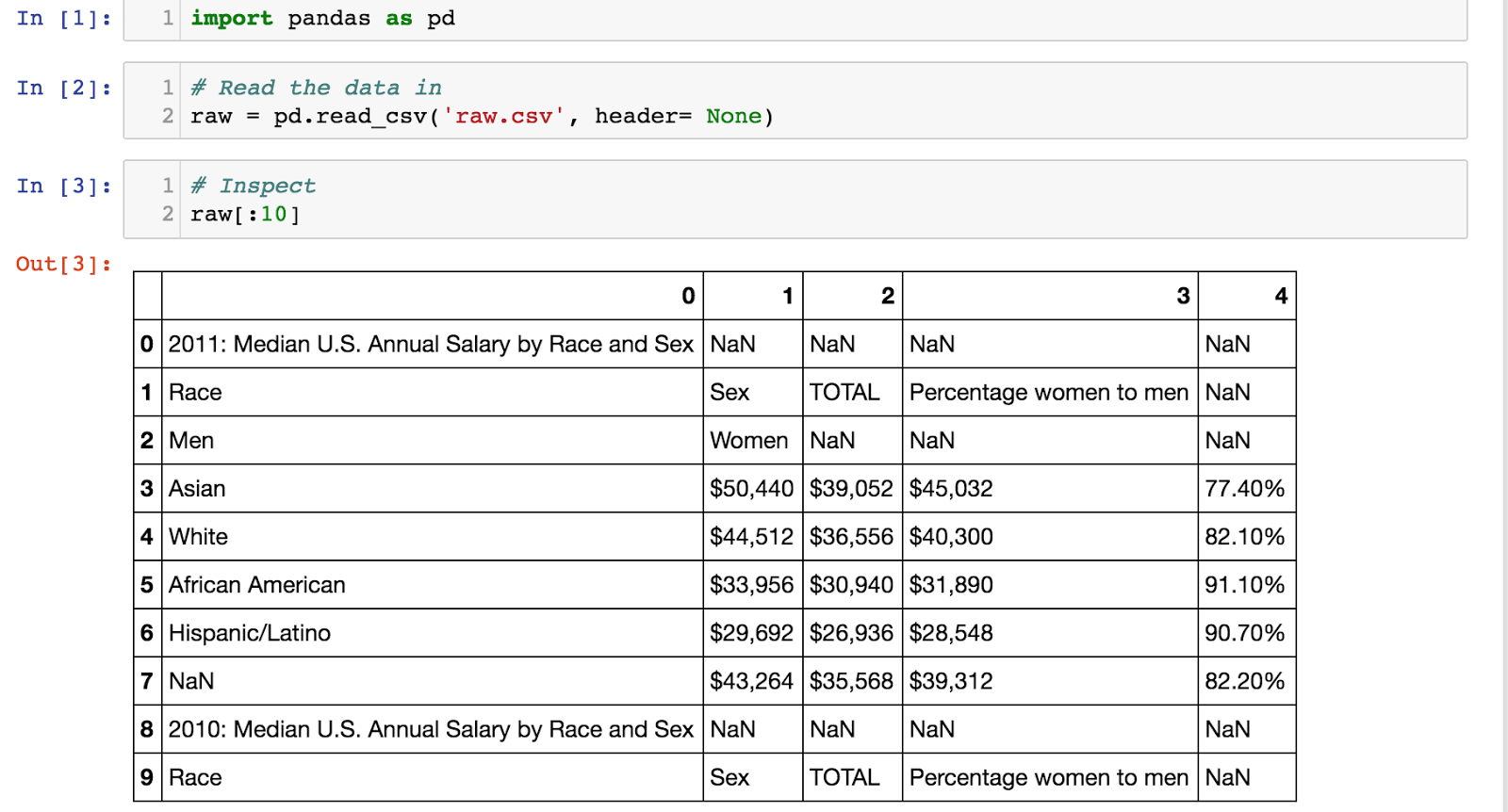
****

****

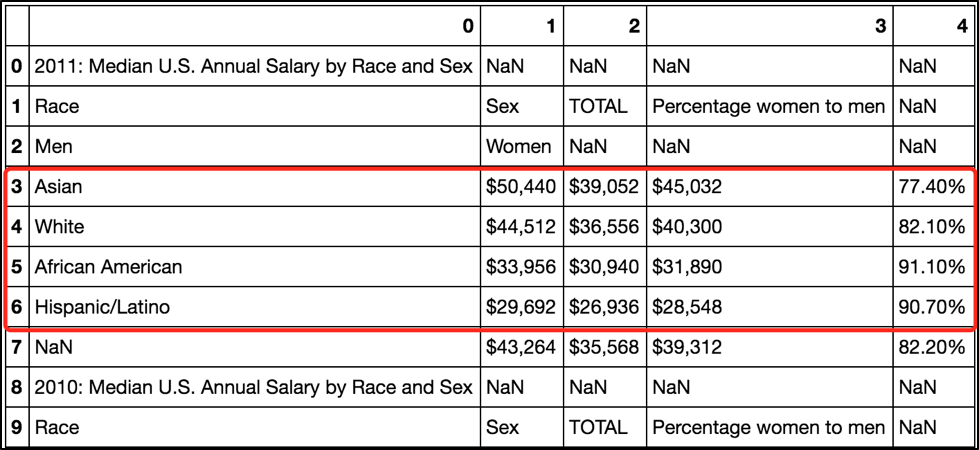
Since this is a HTML page, web crawling on a HTML table is not much of help. So, I just copied and pasted this table into an excel spreadsheet.

Then load it into a Jupyter notebook

Read the data in:



From the first 10 rows, we can see that we only need those without 'NaN' data

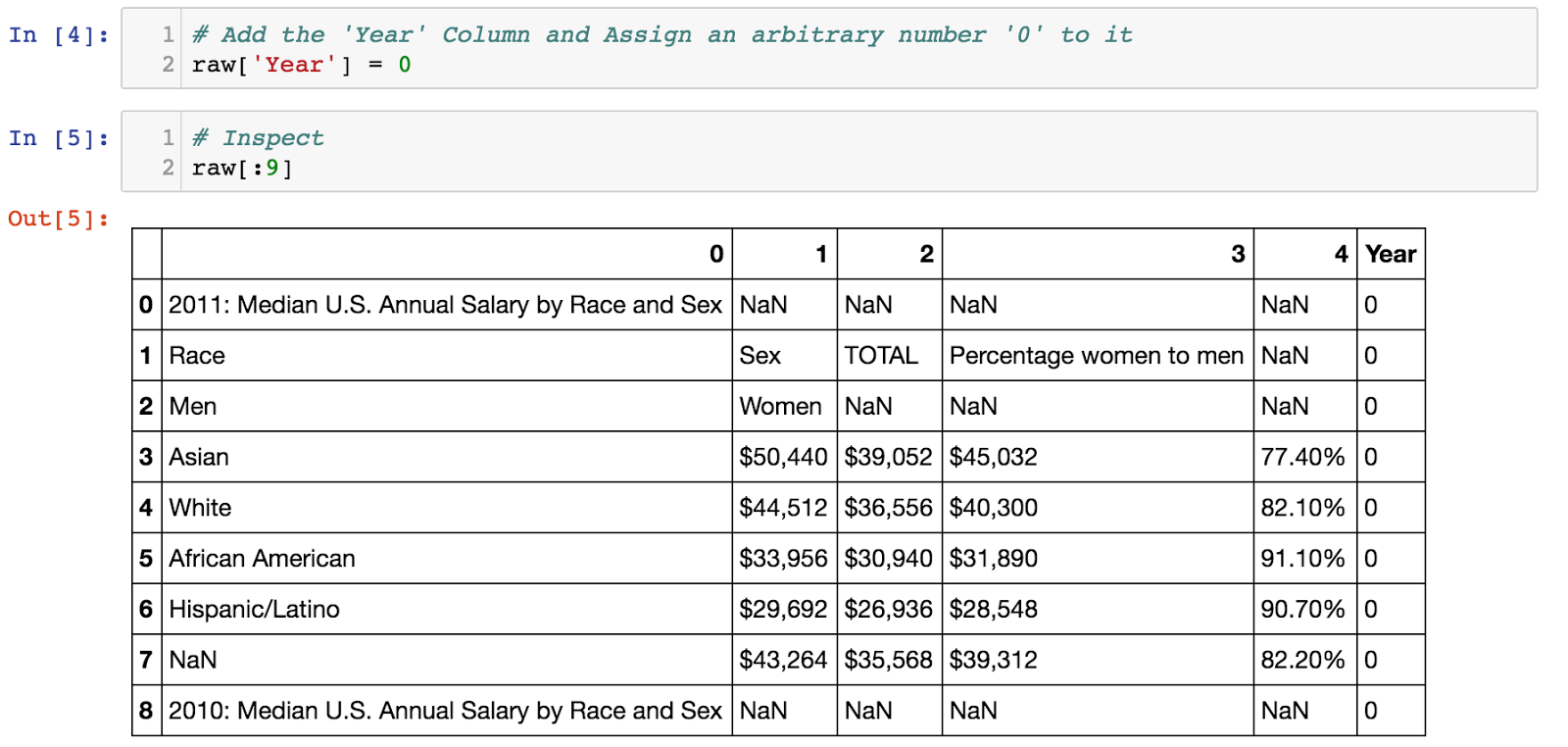


Row 3 to Row 6 are the four races' income in 2011.

So, all we need is to extract these four lines for every year.

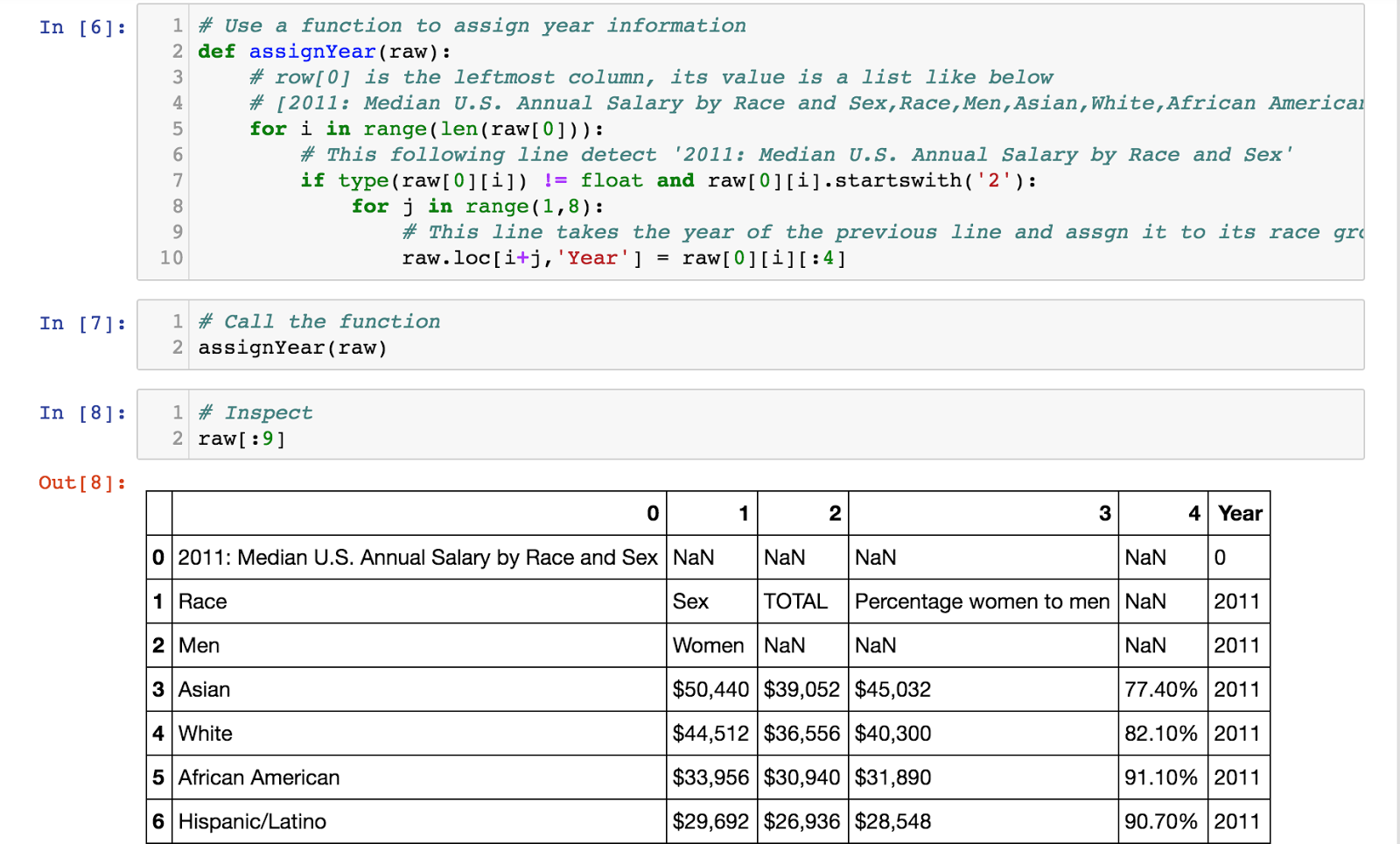
We can drop all NaN rows now, but if we drop them all, the 'Year' information will be lost, i.e., we wouldn't be able to know these four lines are which year's number.

So, before we can remove all 'NaN', I will add ‘Year’ for each group.

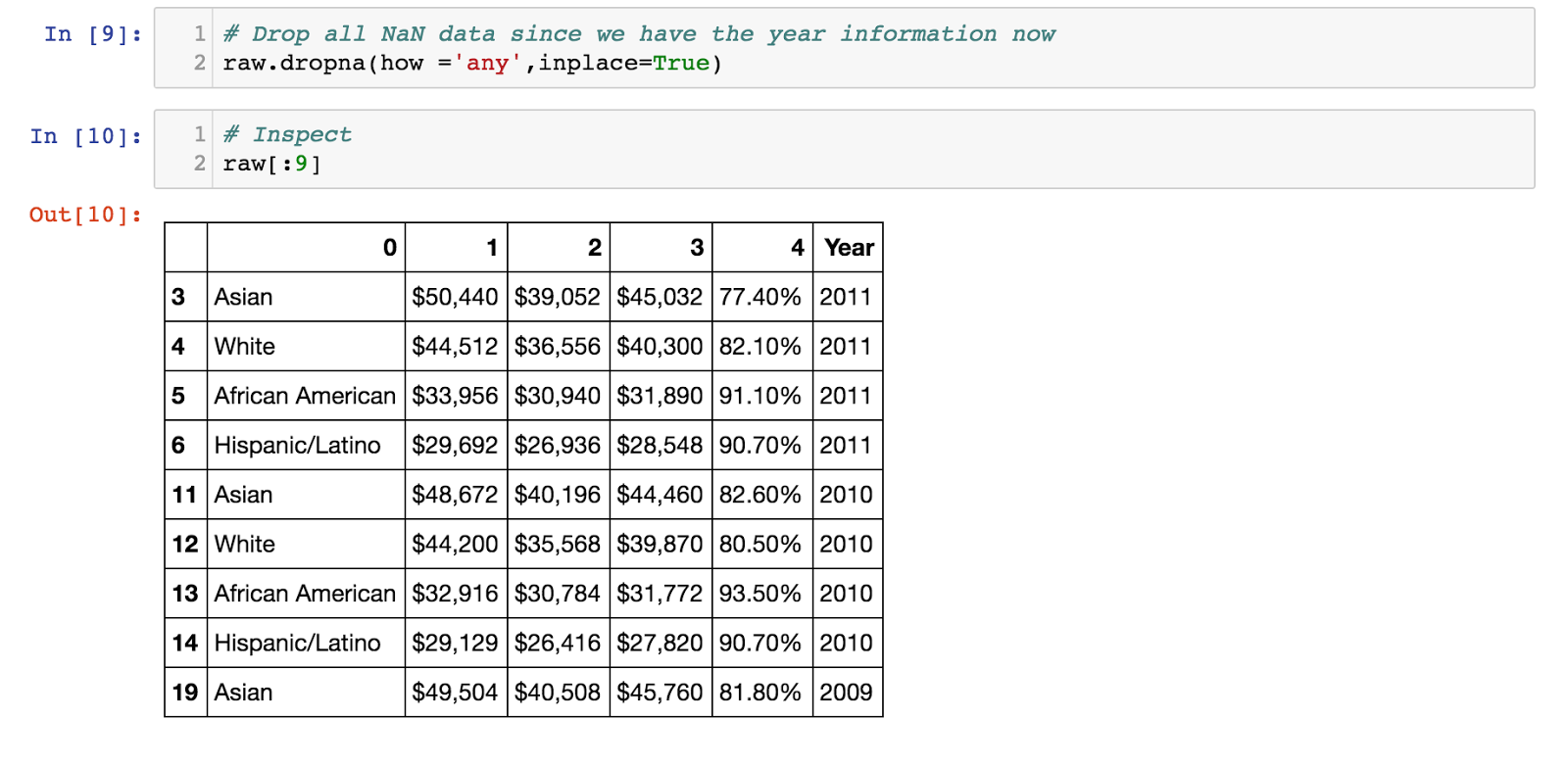


To add the correct 'Year' for each group, we noticed that two rows after every Year-included-row such as '2011: Median U.S. Annual Salary by Race and Sex', it will be the start of four group 'Asian’.

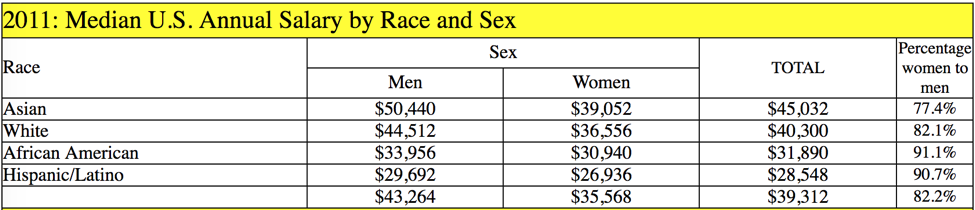
So, we use this pattern to add 'Year' information for each group.



Drop any rows that has ‘NaN’ in it

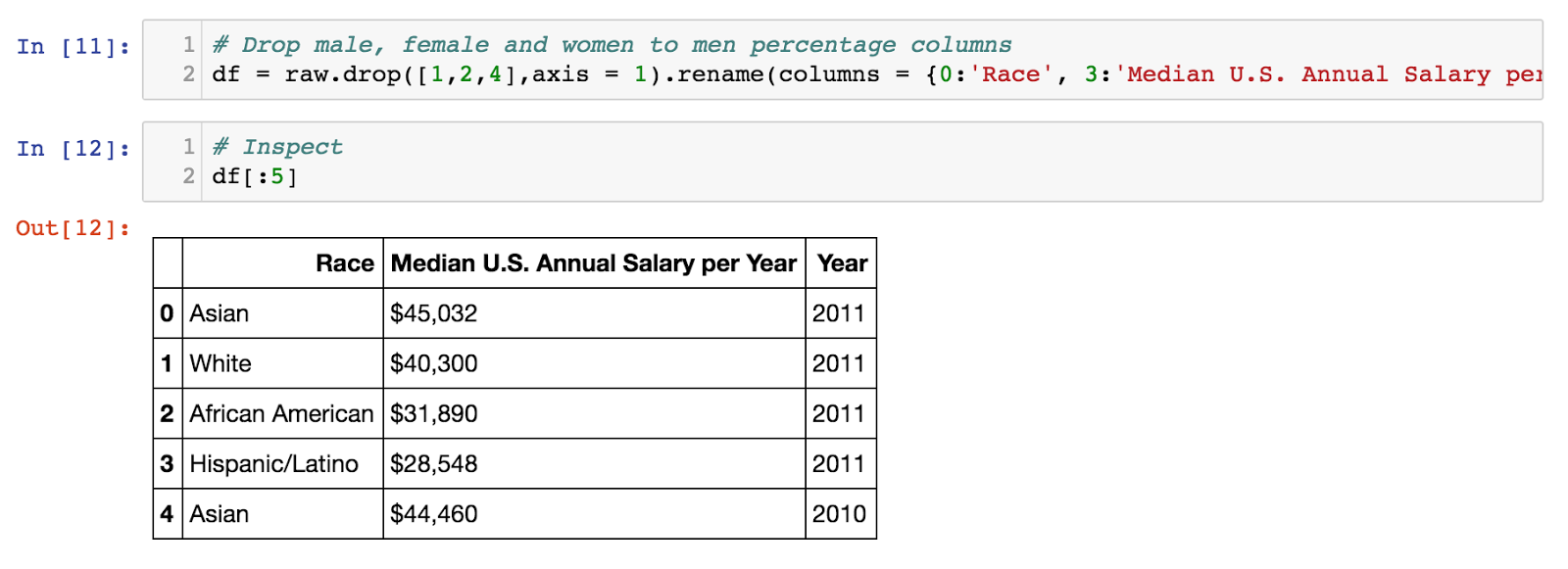


What the original table looks like:

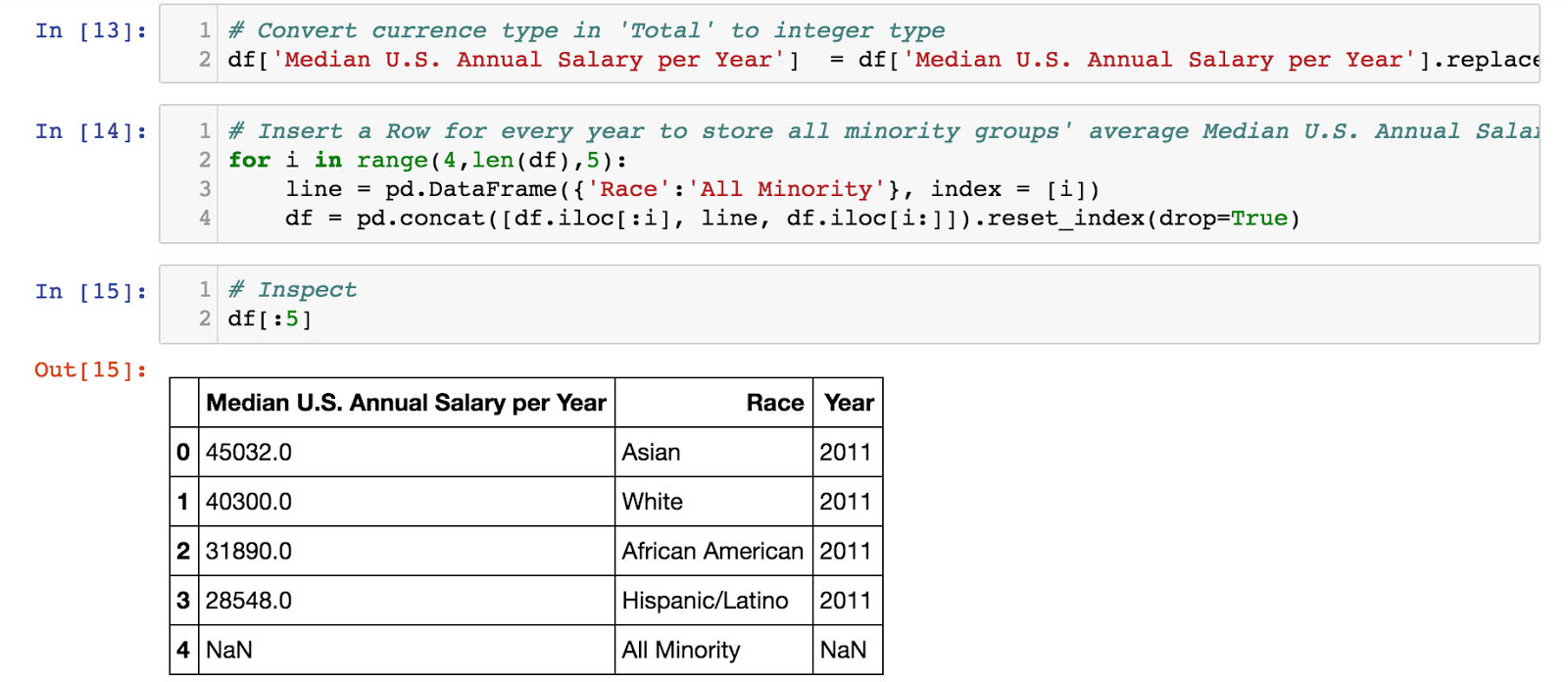


The first number column is for men, the second column is for women, and the third one is for all.

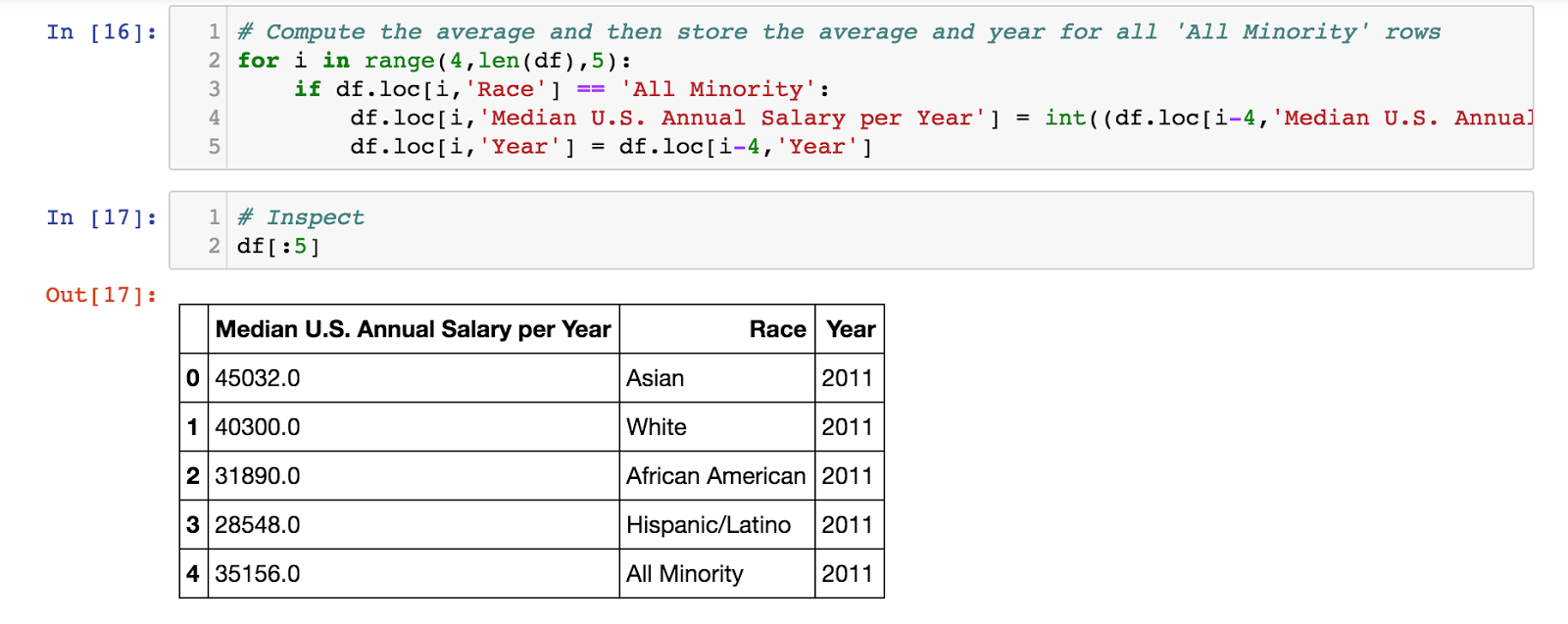
All we need is the third column. So, drop first, second number and the percentage columns.



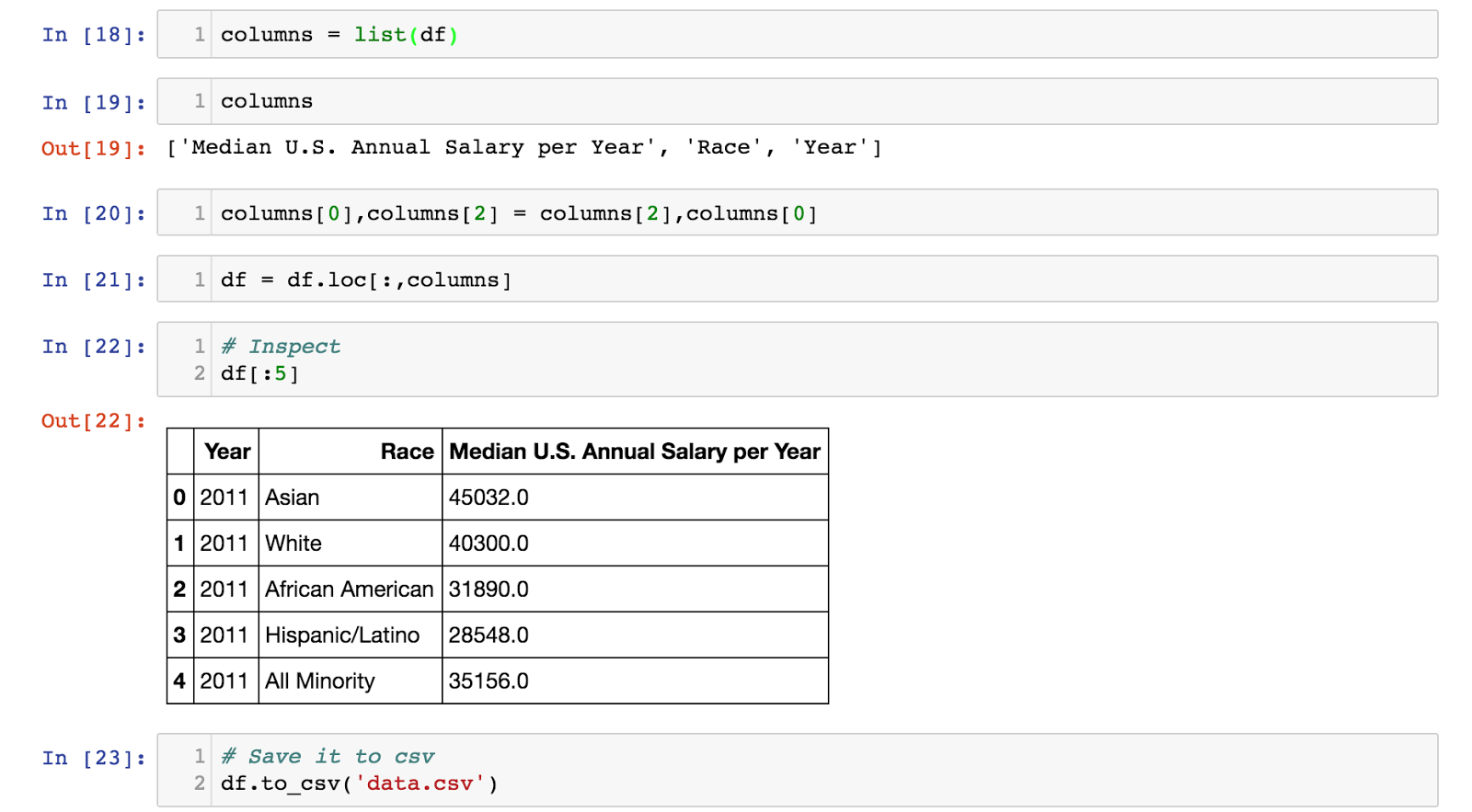
Insert an ‘All Minority’ row for every year for later tableau use



Fill the Median Salary and Year information for ‘All Minority’ rows.

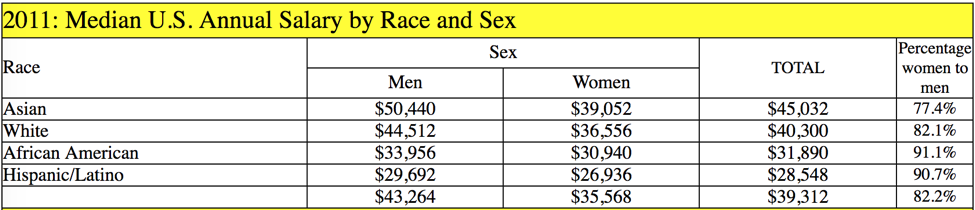


Rearrange the columns



Reasoning

Reading the raw data only, I found that the Asian American earn the most. Then White, African American, Hispanic/Latino follow for all the years that I have collected data for.



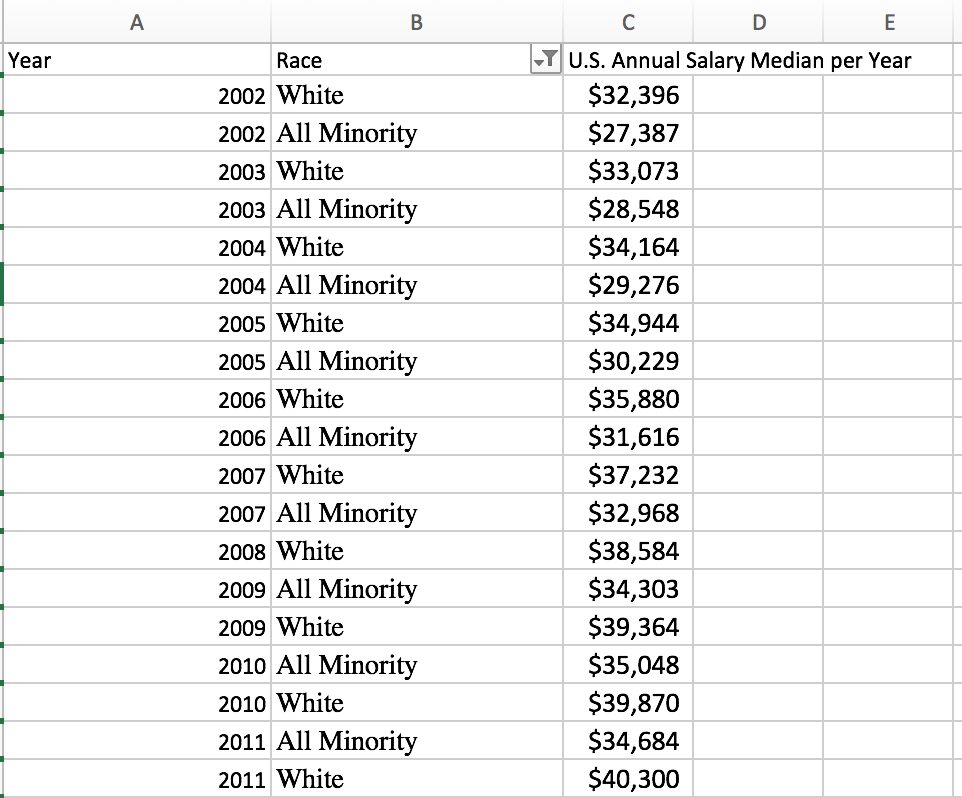
It appears to me that there must be some reason behind the differences and the reason is not discrimination because clearly Asian as one of the minority groups out earn White American.

And it didn’t happen for just one year but (at least) for the years I’ve collected data for.

So, the deceptive/opposite claim would be ‘There is discrimination among races when it comes to your paycheck.

To achieve this purpose, I would average all three minority groups’ salary and compare it with ‘White’.

Like this table:



Now White has more salary than all minority groups.

The other side of claim would be what the original data suggested ‘There is no discrimination among races when it comes to your paycheck.

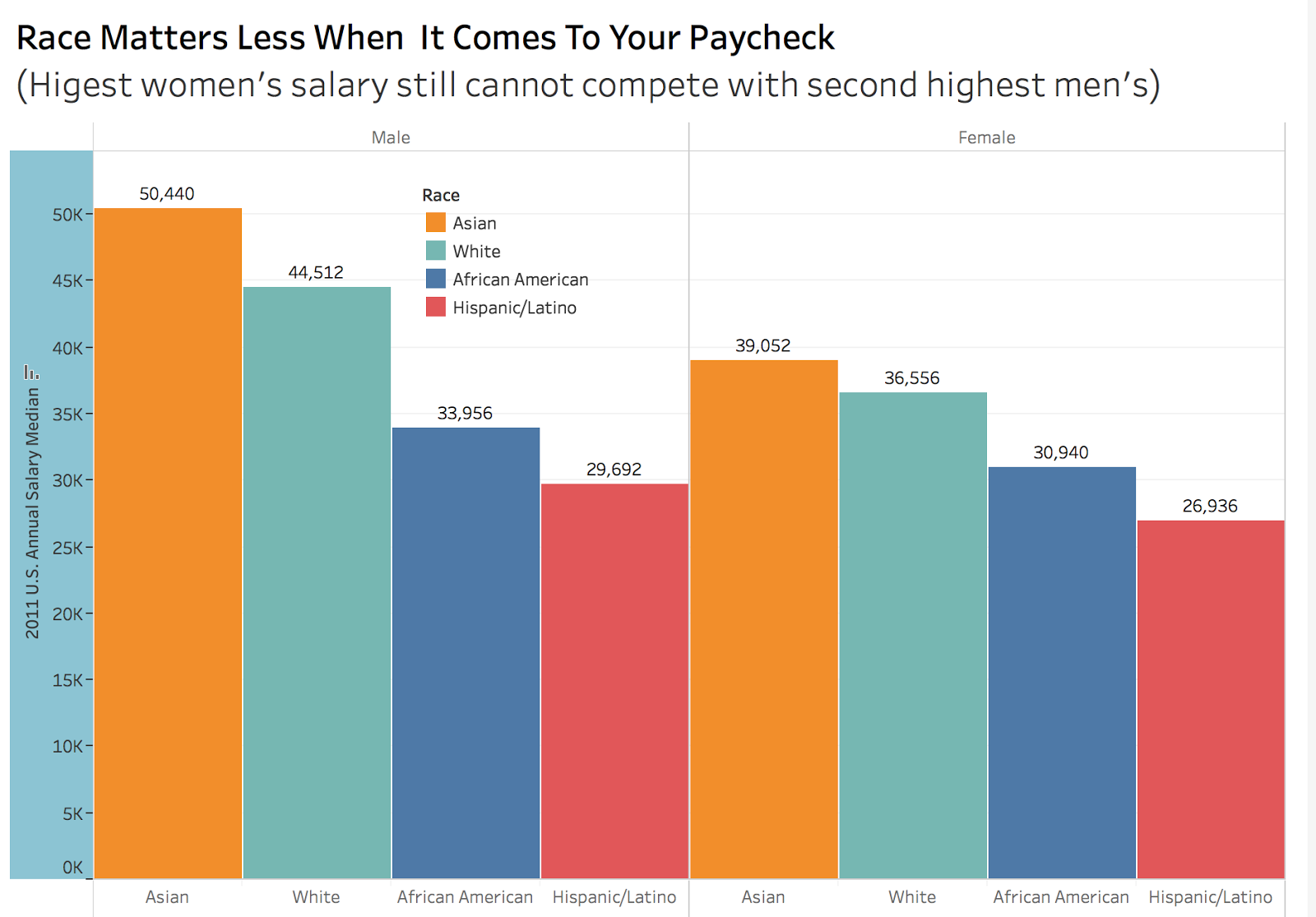
Detours

I felt that I was not confident enough to say ‘There is absolutely no discrimination’ because even though Asian out earn White, there are still two minority groups make less money than White. What if the discrimination lies just in these two groups?

I dug a little deeper and found out some article saying that because Asian are more likely to stay at Management job whereas the other two stayed at relatively low-paying jobs. The reason behind this is probably education.

I tried to incorporate this data in but didn’t succeed.

So, I pivoted the focus to ‘Gender’.  After rearranging data Tableau, I was able to detect that ‘Race matters less when it comes to your paycheck’



Later I realized it still says ‘Race matters’!  Hence, discrimination still exists.

So, this is removed from the final visualization.

3. Several intermediate visualization prototypes

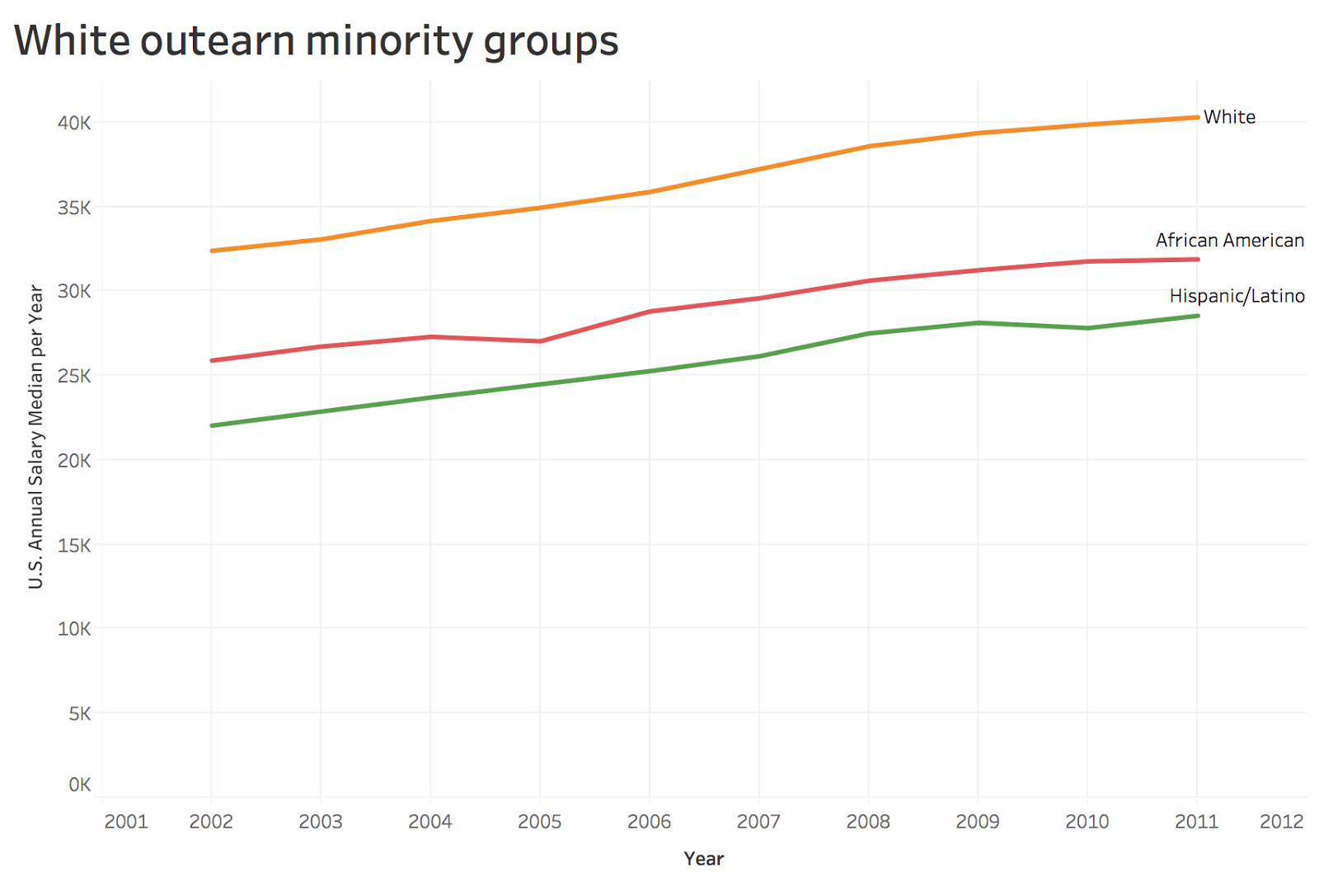
Deceptive Claim: ‘RACE MATTERS A LOT WHEN IT COMES TO YOUR PAYCHECK

(White out earn minority groups)’

The visualization itself is not difficult to achieve.

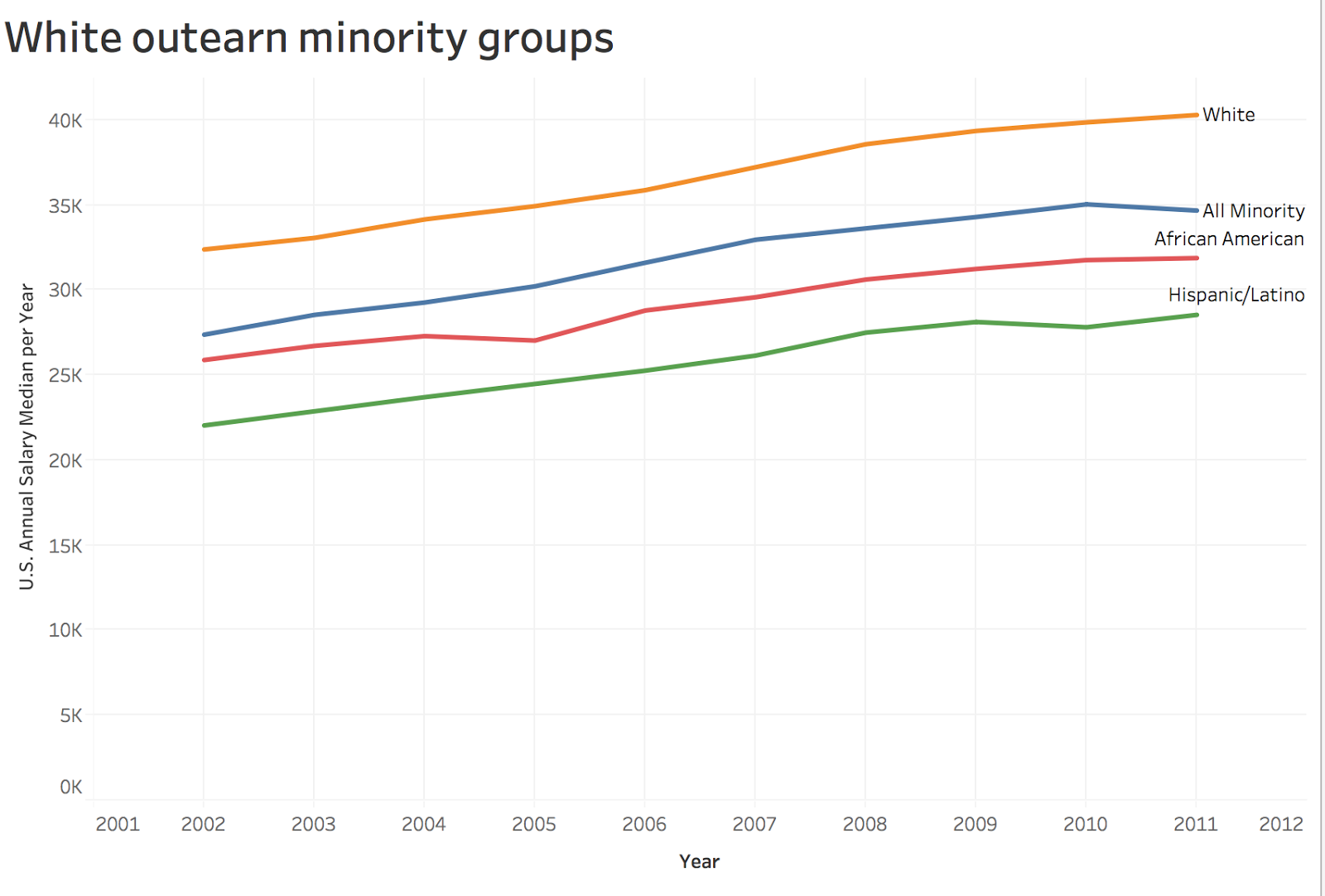
I put year on the X axis and Salary on the Y axis and plot lines for the groups.

At the very first stage, I thought about removing Asian completely from the chart.



Then I thought the audience would’ve asked the question ‘Where did the Asian go’?

So, I averaged these three minority groups.

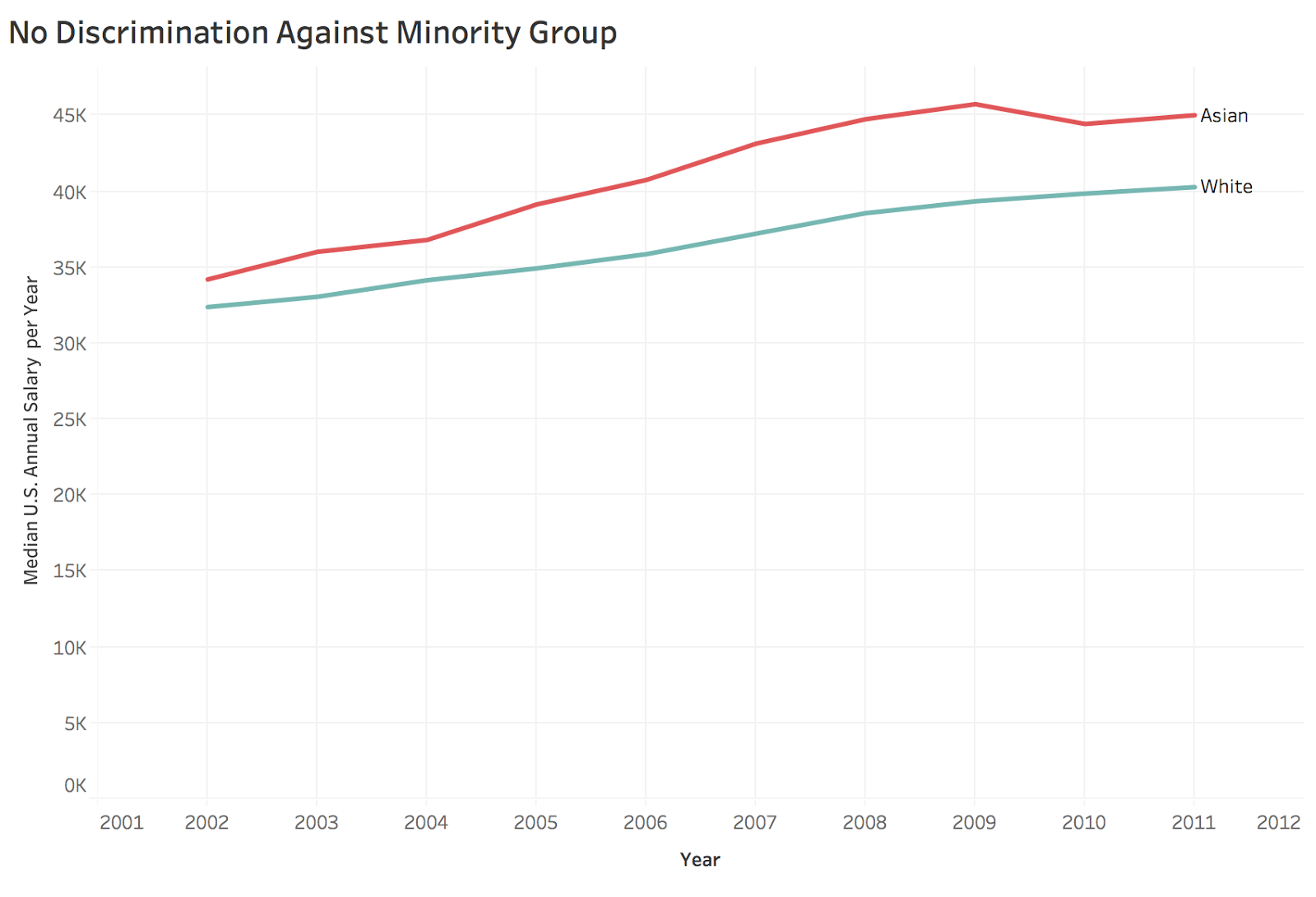


Then I realized that showing ‘All Minority’ with ‘African American’ and ‘Hispanic/Latino’ would have the audience suspect these must be a minority earn a higher income than these two, otherwise the line would be higher than those two. I wonder what other groups are behind this line chart?

So, for the final Visualization, I removed ‘African American’ and ‘Hispanic/Latino’ too and left ‘White’ and ‘All Minority only’.

Opposite Claim: There is no discrimination against Minority Groups when it comes to your paycheck per se because White group doesn’t out earn all minority groups/

Corresponding with the deceptive claim, my initial for this claim would be leaving only Asian and White on the chart.



The same question applies here. Where are the African American and Latino group?

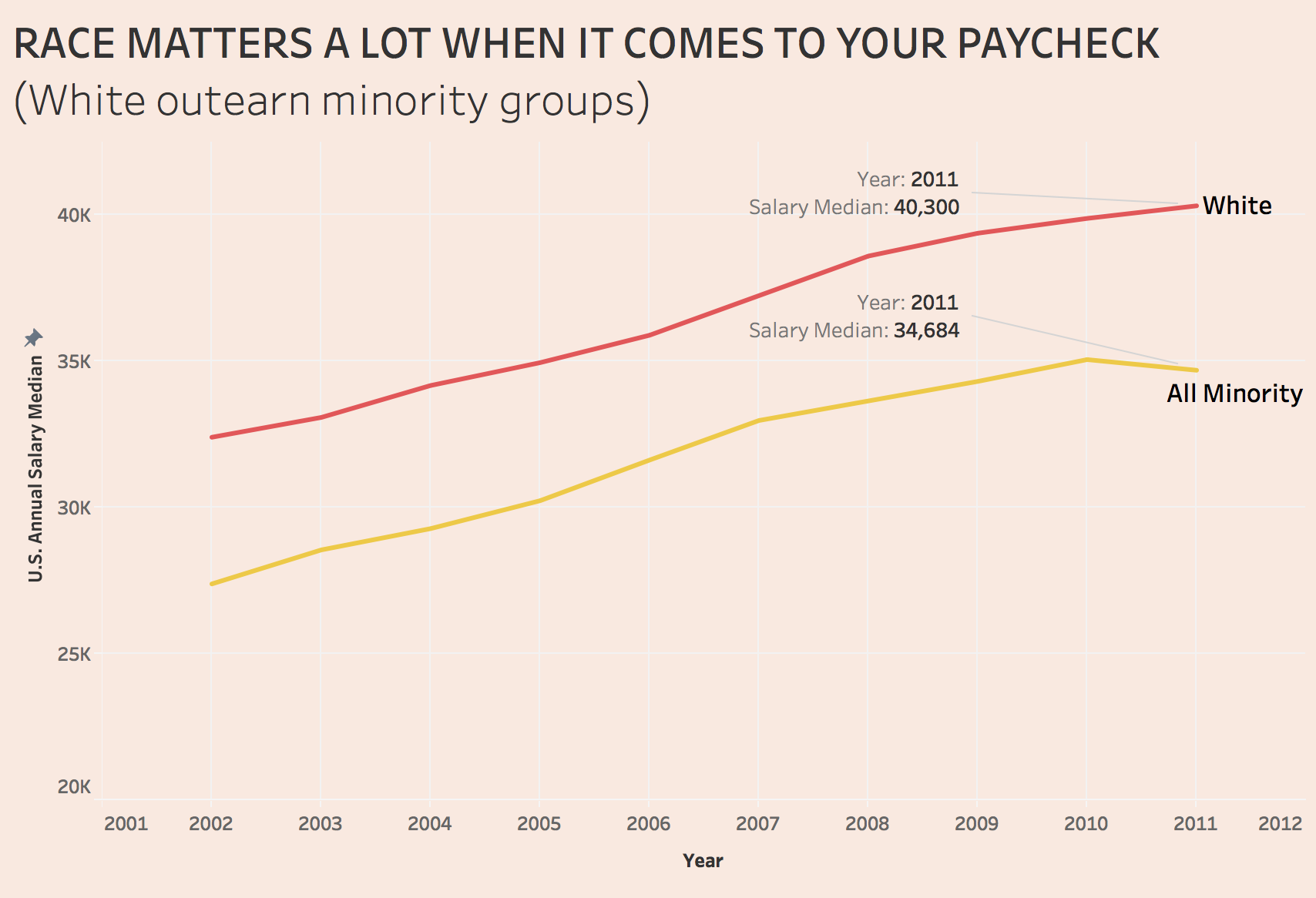
So, for the final visualization, I put all groups on the chart and used a very long and descriptive title ‘There is no discrimination against Minority Groups when it comes to your paycheck per se because White group doesn’t out earn all minority groups.’

4. Final data visualization

The deceptive claim: ‘RACE MATTERS A LOT WHEN IT COMES TO YOUR PAYCHECK

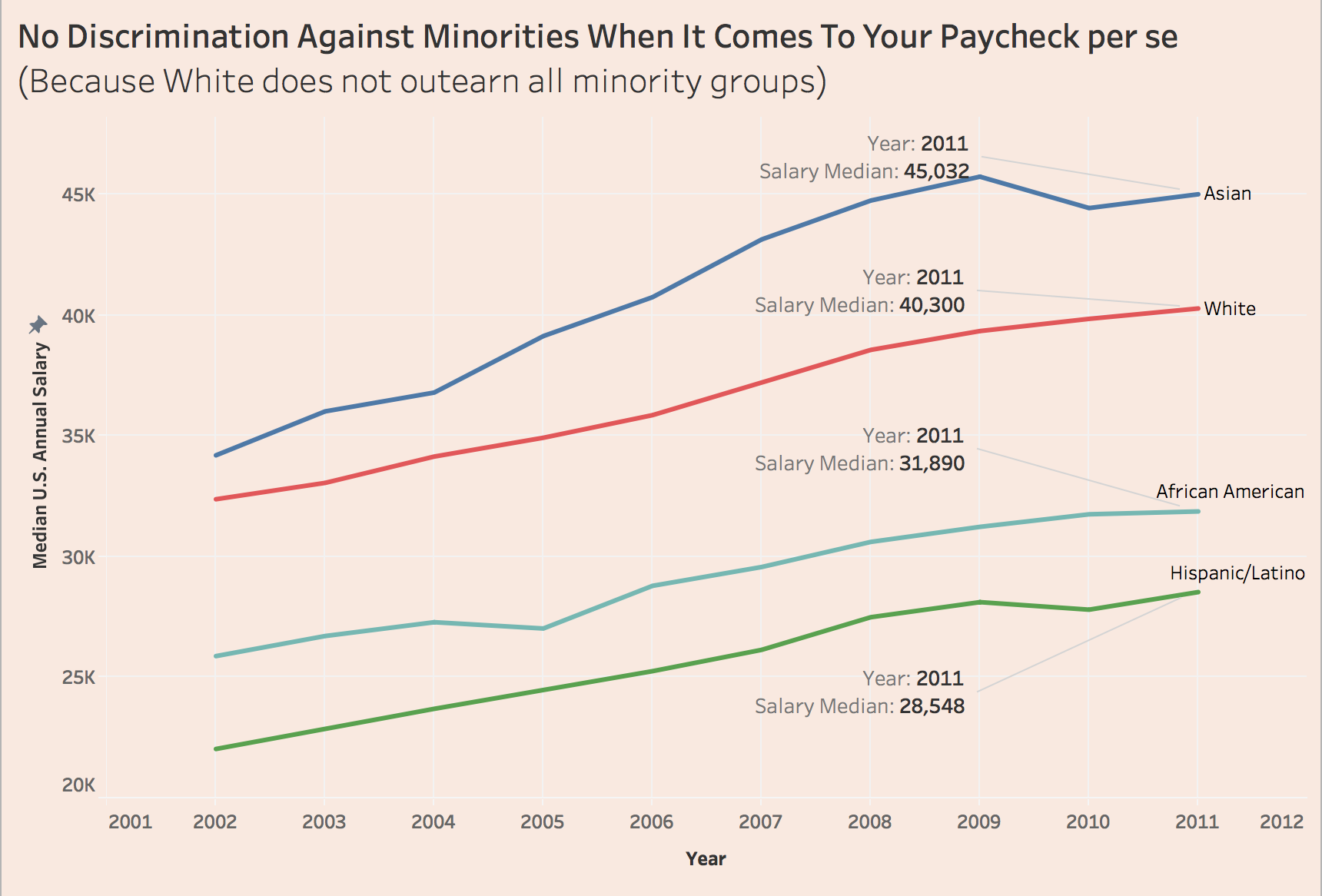
(White out earn minority groups)

’



The rebuttal claim:

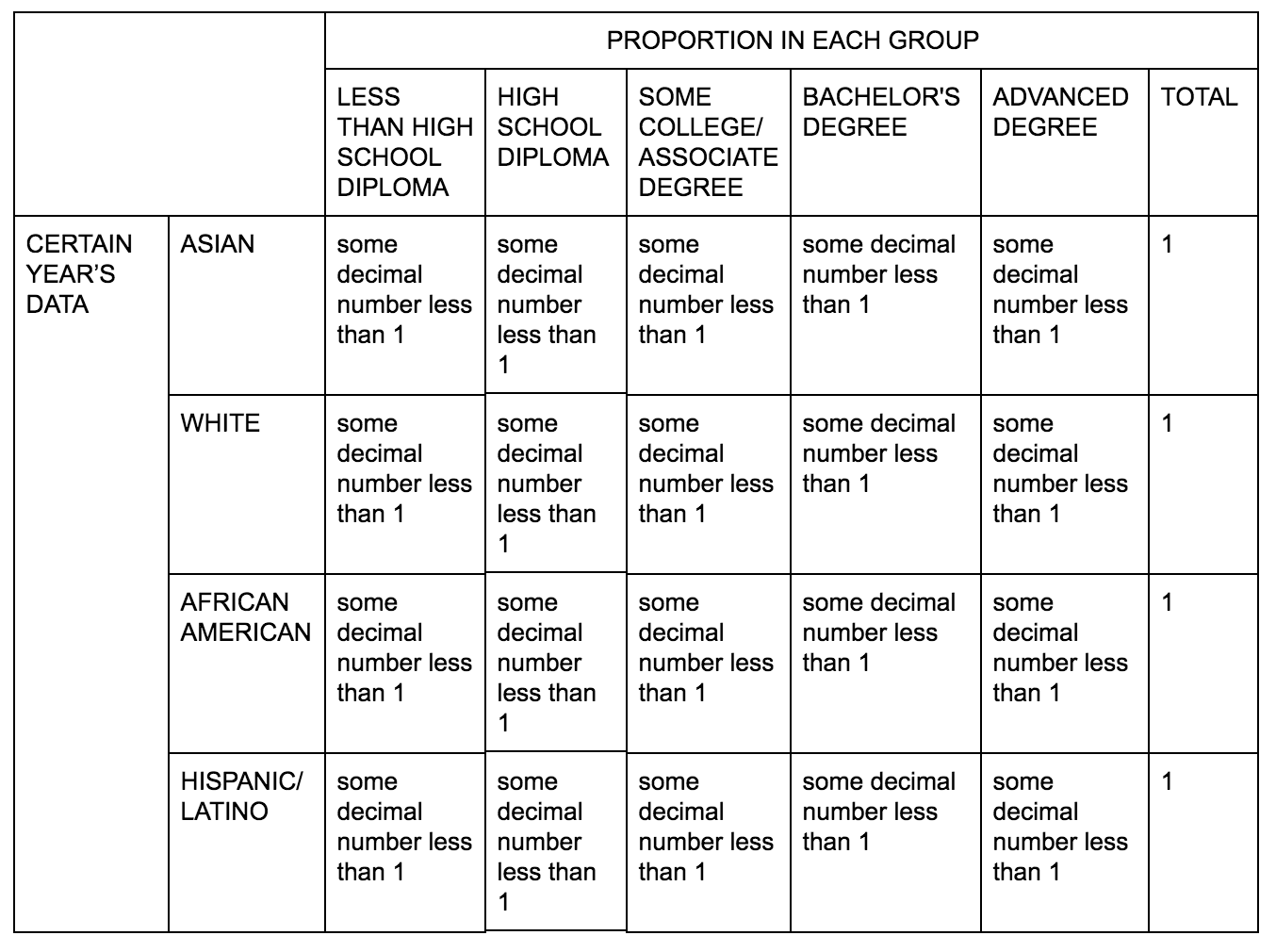
‘There is no discrimination against Minority Groups when it comes to your paycheck per se because White group doesn’t out earn all minority groups.’



5. Road map with future features/enhancements/features

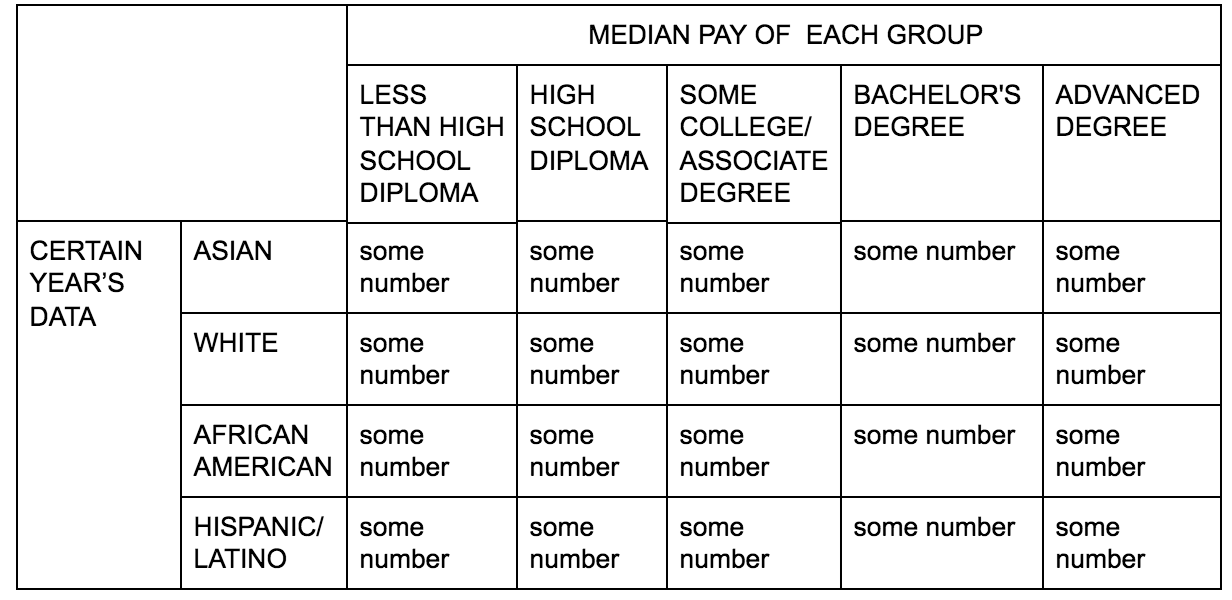
For the opposite claim, I will collect the education background for each group and present the findings.

Ideally, I would be able to create this table for each group:



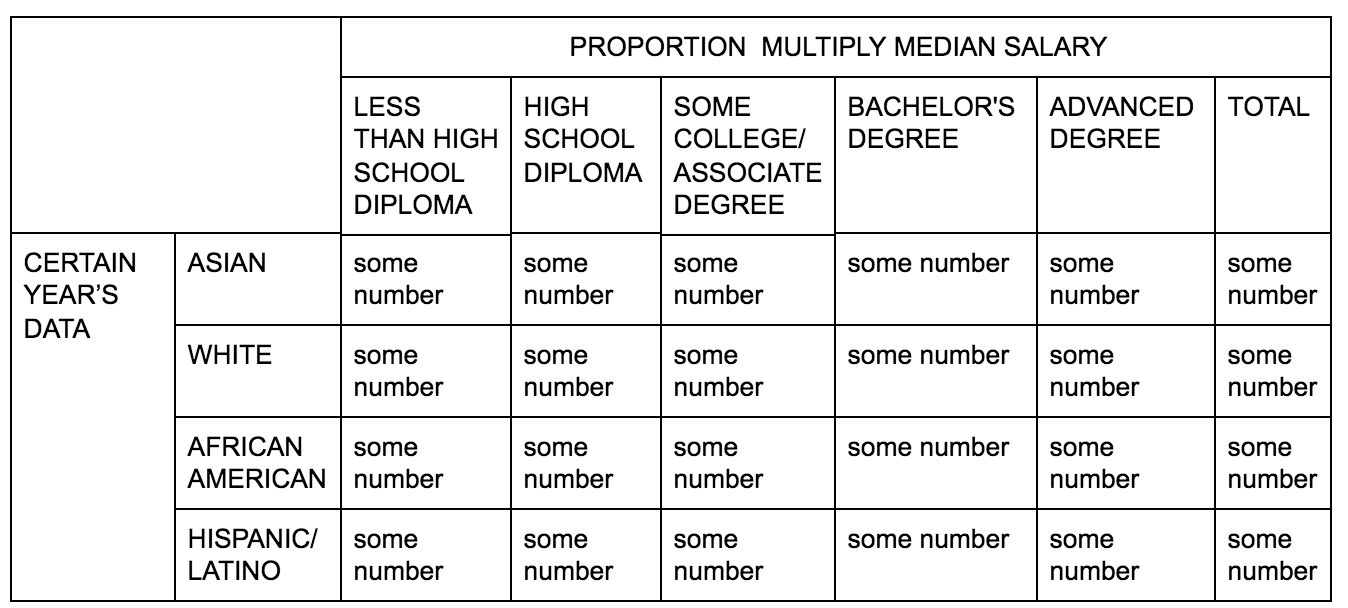
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | PROPORTION IN EACH GROUP | | | | | |
| LESS THAN HIGH SCHOOL  DIPLOMA | HIGH SCHOOL DIPLOMA | SOME COLLEGE/ASSOCIATE DEGREE | BACHELOR'S DEGREE | ADVANCED DEGREE | TOTAL |
| CERTAIN YEAR’S DATA | ASIAN | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 | 1 |
| WHITE | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 | 1 |
| AFRICAN AMERICAN | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 | 1 |
| HISPANIC/LATINO | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 | 1 |

Then I should list the median salary for each group:



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | MEDIAN PAY OF EACH GROUP | | | | |
| LESS THAN HIGH SCHOOL  DIPLOMA | HIGH SCHOOL DIPLOMA | SOME COLLEGE/ASSOCIATE DEGREE | BACHELOR'S DEGREE | ADVANCED DEGREE |
| CERTAIN YEAR’S DATA | ASIAN | some number | some number | some number | some number | some number |
| WHITE | some number | some number | some number | some number | some number |
| AFRICAN AMERICAN | some number | some number | some number | some number | some number |
| HISPANIC/LATINO | some number | some number | some number | some number | some number |

Then multiply these two tables



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | PROPORTION MULTIPLY MEDIAN SALARY | | | | | |
| LESS THAN HIGH SCHOOL  DIPLOMA | HIGH SCHOOL DIPLOMA | SOME COLLEGE/ASSOCIATE DEGREE | BACHELOR'S DEGREE | ADVANCED DEGREE | TOTAL |
| CERTAIN YEAR’S DATA | ASIAN | some number | some number | some number | some number | some number | some number |
| WHITE | some number | some number | some number | some number | some number | some number |
| AFRICAN AMERICAN | some number | some number | some number | some number | some number | some number |
| HISPANIC/LATINO | some number | some number | some number | some number | some number | some number |

And finally, use the number for each group and divide the total median pay:



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | PROPORTION OF MEDIAN PAY OF EACH GROUP | | | | |
| LESS THAN HIGH SCHOOL  DIPLOMA | HIGH SCHOOL DIPLOMA | SOME COLLEGE/ASSOCIATE DEGREE | BACHELOR'S DEGREE | ADVANCED DEGREE |
| CERTAIN YEAR’S DATA | ASIAN | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 |
| WHITE | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 |
| AFRICAN AMERICAN | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 |
| HISPANIC/LATINO | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 | some decimal number less than 1 |

And do the above for year 2002 to 2001.

If the proportion of Asian’s and White’s at Bachelor’s Degree and Advanced Degree are higher than African American and Hispanic/Latino, then we can confidently see there is no racial discrimination when it comes to your paycheck. The difference is caused by education background and occupation selection which is also majorly caused by education.

6. Showcase video (<90 sec)