Data Description

Barry Beckers extracted the chosen dataset from the 1994 census database, which was then submitted to UCI as an Adult dataset. The dataset consists of the following 15 features and is a survey of about 50,000 people.

1. Age
2. Workclass
3. Fnlwgt
4. Education
5. Education-num
6. Marital-status
7. Occupation
8. Relationship
9. Race
10. Sex
11. Capital-gain
12. Capital-loss
13. Hours-per-week
14. Native-country
15. Income (Target Variable)

OBJECTIVE

The major objective is to determine how factors like age, education, marital status, and others will affect a person's income prediction in order to comprehend the total positive and bad effects on finances, which could then have an impact on the growth of the nation. We are attempting to comprehend the relationships between the features and the impact they have on our target variable as a whole.

Understanding the relationship between marital status and various age groups is one of the topics we will attempt to answer.
• How does a person's salary depend on their level of education?
• The impact of gender on income levels
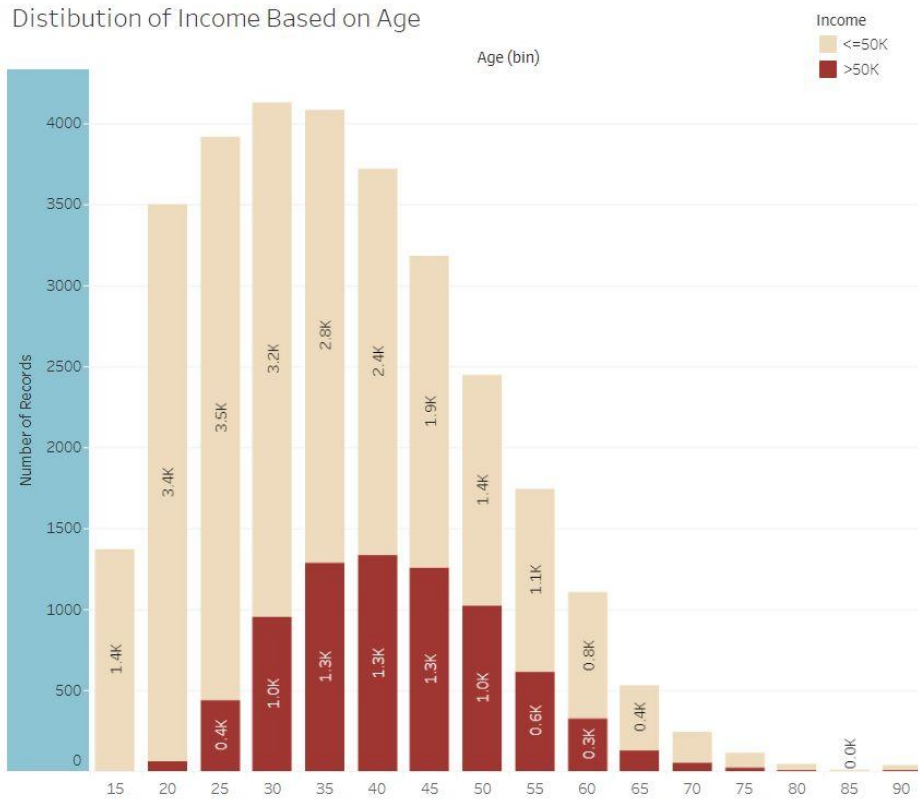• The effect of occupation and weekly labor hours on income levels

Distribution based on different countries

The dataset is highly biased towards the United States and there are lesser instances for other countries. The US accounts for approximately 92% of the surveyed population.

Distribution by Different Countries

Country Percentage

Choose Nat..
Other Coun.. 8.81%
United-Stat.. 91.19%

Income Distribution

Income

Country Count

19M
5,145M
3M
5M
8M
12M
11M
177M 17M
17M 11M 8M
8M

© 2022 Mapbox © OpenStreetMap

Map based on Longitude (generated) and Latitude (generated). Color shows sum of Fnlwgt. The marks are labeled by sum of Fnlwgt. Details are shown for Native-Country. The view is filtered on Native-Country, which keeps 41 of 41 members.

Count of Surveyed People
0M 5,145M

Number of Records
20K
15K
10K
5K
0K

22.7K
7.5K

<=50K >50K

Count of adult_c&v.csv for each Income. Color shows details about Income. The marks are labeled by count of adult_c&v.csv.

Income
<=50K
>50K

Distribution based on Age Groups w.r.t. Sex

1. The majority of survey participants are between the ages of 25 and 40 and have annual incomes under $50,000.

2.) Users earning over $50,000, however, have the biggest capital gains when they are between the ages of 60 and 75.

3. There is no information available to account for the age range of 85 to 90 years.

4. The age range of ladies with the highest earnings (> $50,000) is 20 to 25 years, but for men, it is beyond 90 years.

Different Age Groups w.r.t. Sex

Distibution of Income Based on Age

Age vs Capital Gain

Income
<=50K
>50K

Age (bin)	Income	
	<=50K	>50K
15	82	0
20	61	7,583
25	147	2,421
30	132	3,109
35	177	3,356
40	156	3,597
45	169	4,240
50	157	4,537
55	186	4,493
60	268	3,811
65	397	9,491
70	373	9,544

Average of Capital-Gain broken down by Income vs. Age (bin). Color shows average of Capital-Gain. The marks are labeled by average of Capital-Gain. The data is filtered on Sex, which keeps Female and Male.

Avg. Capital-Gain
0 — 9,544

Count of adult_c&v.csv for each Age (bin). Color shows details about Income. The marks are labeled by count of adult_c&v.csv. The data is filtered on Sex, which keeps Female and Male.

Distribution based on Education Level

Income under $50,000
The group of students with the lowest income levels, those with high school degrees, earn the most.

Income > $50,000
The biggest capital increase is experienced by male students in bachelor's programs, followed by those in professional programs.

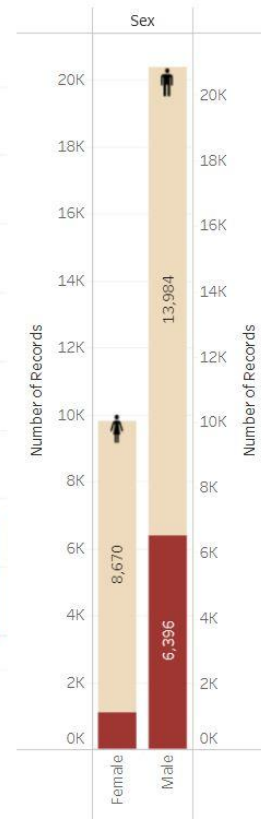For the female population, graduates from high school have the greatest capital increase.

Capital gain w.r.t. Education Level
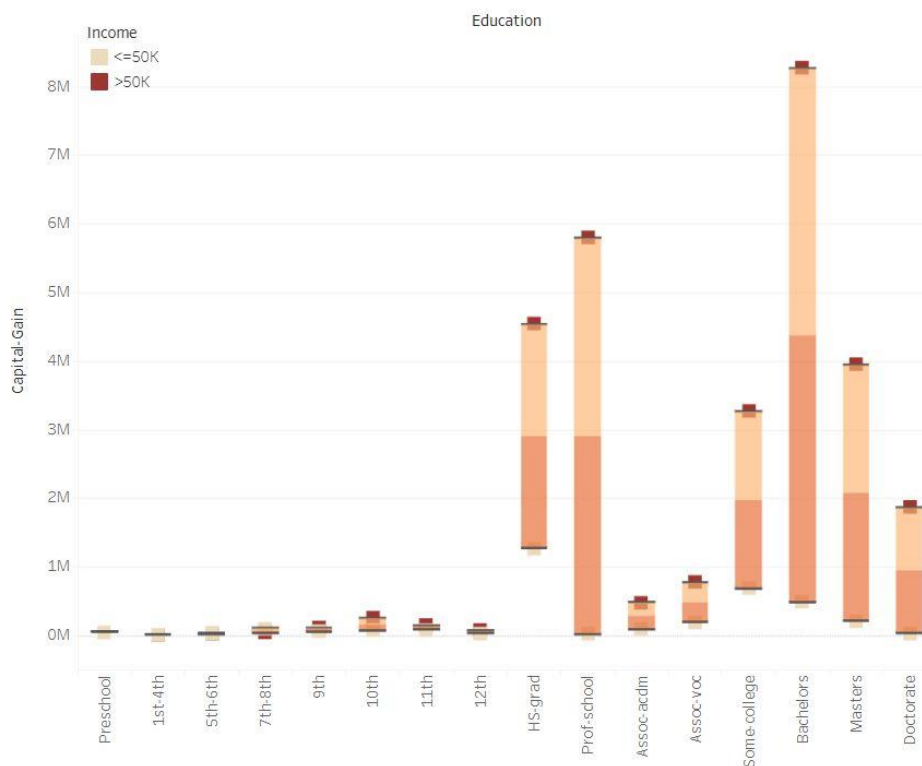
Sum of Capital-Gain for each Education. Color shows details about Income. The data is filtered on Sex, which keeps Female and Male.

Capital gain and Loss w.r.t. Marital Status and Relationship

With respect to marital status, it is observed that husbands have the highest capital gain as well as the highest capital loss followed by people who were never married.

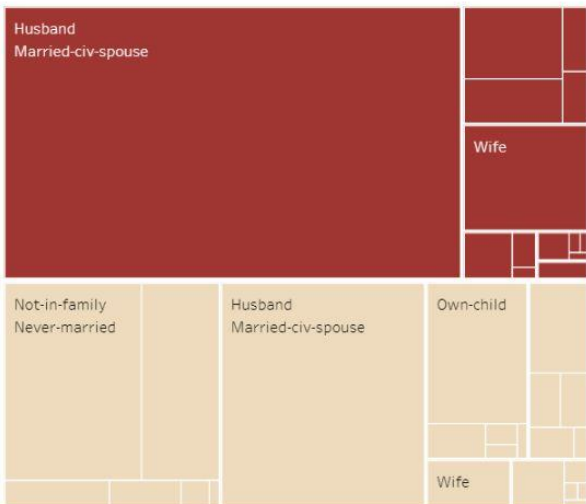Capital Gain and Loss w.r.t. Marital Status and Relationsip

Capital Gain and Loss w.r.t. Occupation and Workclass

The heatmap shows that Craft Repair Workers, Executive Managers, and Specialty Professions (those between the ages of 30 and 45) have the most hours worked.

However, the Sales department employees have the greatest overall capital gain.

However, individuals who serve in the armed forces experience the greatest capital loss in their incomes while working approximately 40 hours per week.
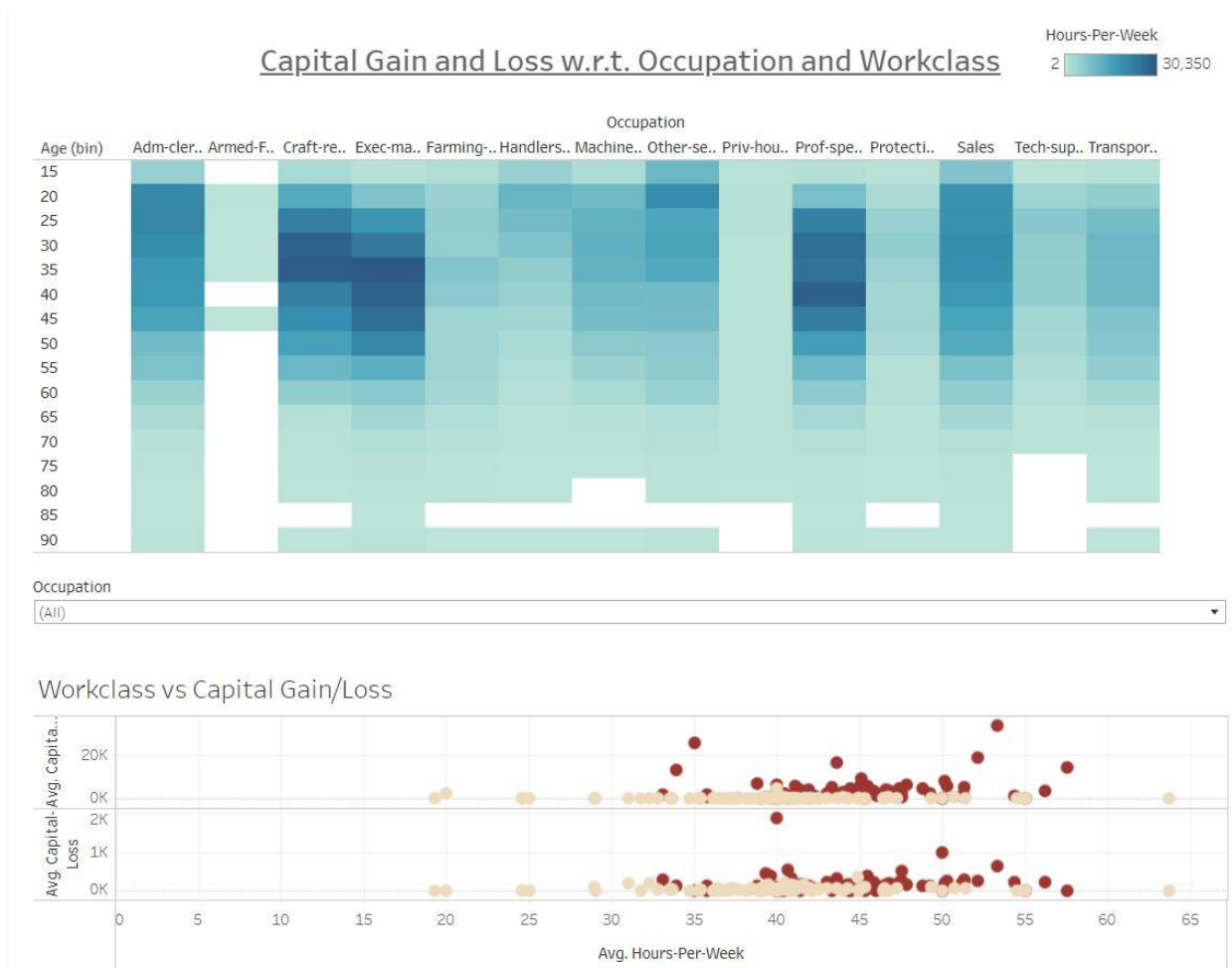
Distribution based on the Race

Since the survey is biased for the US region, there is a larger proportion of White people in both the income categories of less than 50k as well as greater than 50k.

About 60% of the white poulation fall under the income category of less than 50k.

Different Races and their Distribution
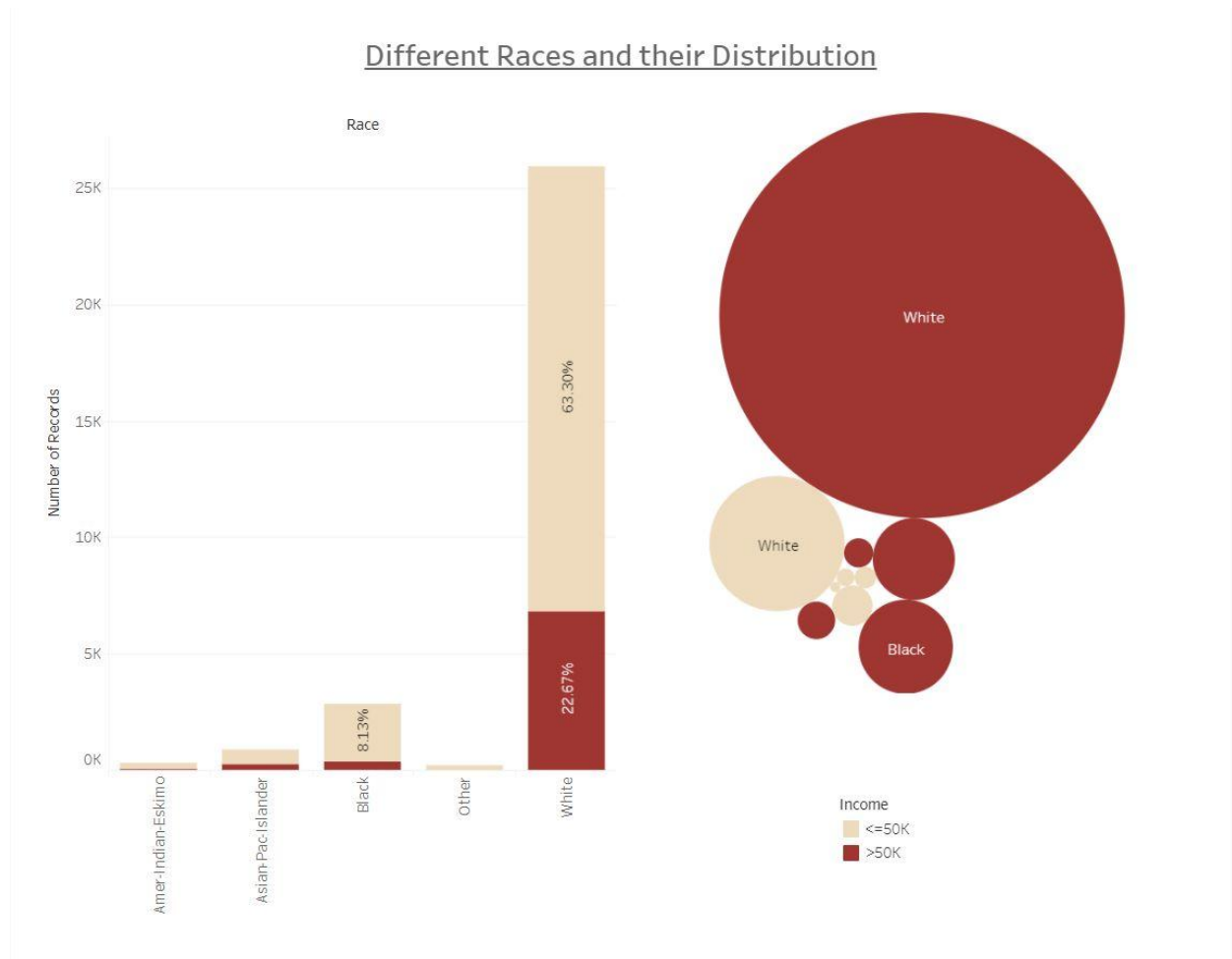
CONCLUSION

1.) People over 65 have the biggest capital gains, and their income often exceeds $50,000. This could be a result of the fact that at this age, people have settled their lives, have no debt, and start to qualify for tax exemptions.

2.) The majority of people begin working as soon as they graduate from college or even high school. As a result, there are fewer students with higher degree levels, which contributes to the higher salary levels among those with bachelor's degrees.

3.) Employees in the executive, management, and sales fields have the best-earning potential and put in the most hours each week. Although craft repair demands long hours of effort, the capital gain is not very significant for these workers.

4.) Divorced individuals without children and married individuals with civilian wives have the highest average financial gains.