Final Project — Milestone 1

ALY 6010 Probability Theory and Statistics

RAHUL AVINASH JADHAV

Northeastern University

College of Professional Studies, Northeastern University, Boston, MA 02115

Contact: jadhav.ra@northeastern.edu

Submitted to Professor: Prof. Roy Wada

Date of submission: 11/15/2021

Introduction

The Given data contains information of individual born between 1957-1964. The data was collected in the year 1994. The provided data set consists of 8847 records and 37 attributes. The Data provides information such as individuals born year, their country of birth, which race they belong to, their gender, education qualification, income, marital status etc. Using this dataset, we can know how much specific degree holder can earn, what impact does size of family has on savings, if individual would buy health plans, Incomes based on various factors etc.

Analysis

Table 1: Summary table of the provided dataset

	count	mean	standard_deviation	min	max	standard_error
GENDER*	8847	1.5	0.500008	1	2	0.00531593
YEAR_OF_BIRTH	8847	1960.6	2.207806	1957	1964	0.023472693
Age	8847	33.4	2.207806	30	37	0.023472693
RACE*	8847	2.2	0.876907	1	3	0.009323
HAVING_HEALTHPLAN*	8847	2.6	0.801076	1	3	0.008516791
REGION_*	8847	2.8	1.363219	1	5	0.014493313
URBAN_RURAL_*	8847	2.6	0.78943	1	3	0.008392964
MARSTAT_KEY_*	8847	2.3	0.796232	1	6	0.00846529
WKSUEMP_PCY_	8683	2.6	7.740386	0	52	0.083066847
EDU_DEGREE*	8847	4	1.569133	1	8	0.016682521
MAJOR_1_*	8847	17.9	8.923551	1	26	0.094872361
INCOME_	8847	18677.7	18953.83	0	101653	201.5110459
NET_WORTH_	8847	53514.8	137001.9	0	946749	1456.560434

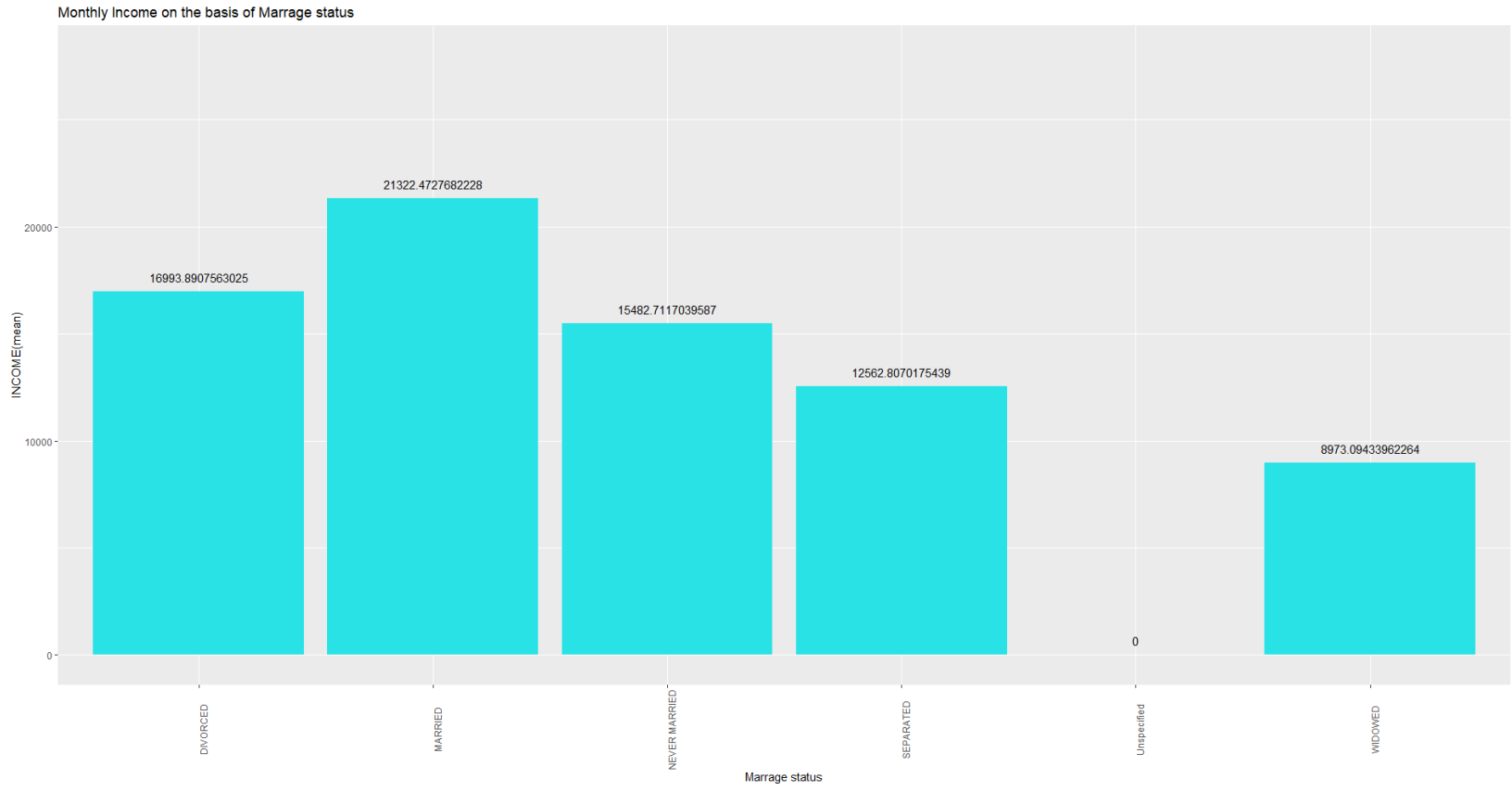
The above table provides count , mean, standard deviation, minimum value, maximum value and standard error of the variables from our dataset. Using this table we can know how much variations are there among the variables , their mean count, what is the minimum and maximum value of the variables and their standard error.

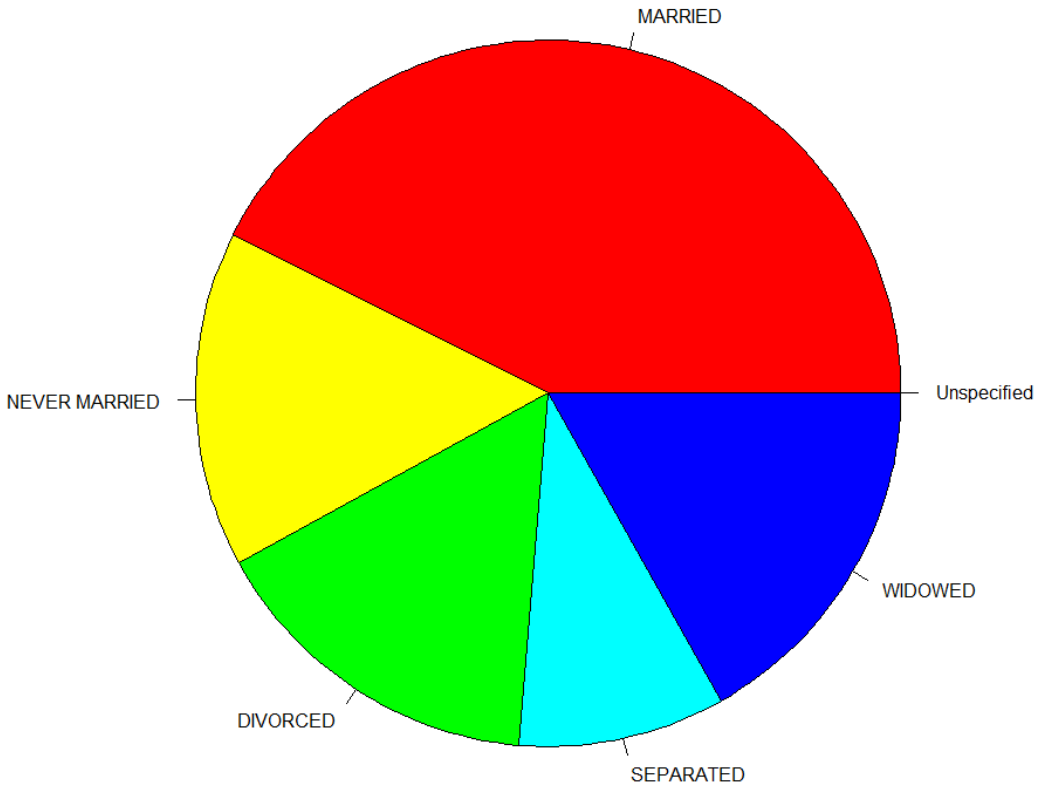
- **INCOME BASED ON MARITAL STATUS**

Table 2: Summary of income based on marital status

MARSTAT_KEY_	mean(INCOME_)	mean(NET_WORTH_)	count	proportion
MARRIED	21322.47277	75724.85708	4884	0.552
NEVER MARRIED	15482.7117	27312.46816	2324	0.263
DIVORCED	16993.89076	27887.20635	1071	0.121
SEPARATED	12562.80702	16916.64912	513	0.058
WIDOWED	8973.09434	29910.62264	53	0.006
Unspecified	0	0	2	0

Bar plot 1: Monthly income based on marital status





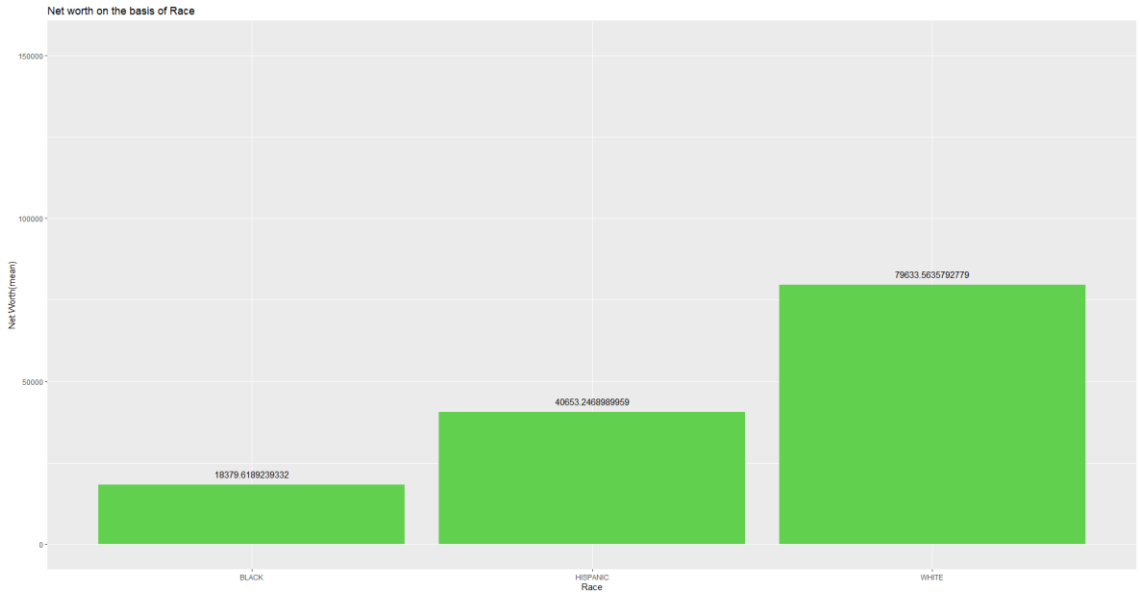
From table one we can see that the proportion of income data is more of married person. If closely observed we can also say that married couple have more net worth and income than rest marital status where as the income and net worth of Separated people are the least.

- **Income based on race**

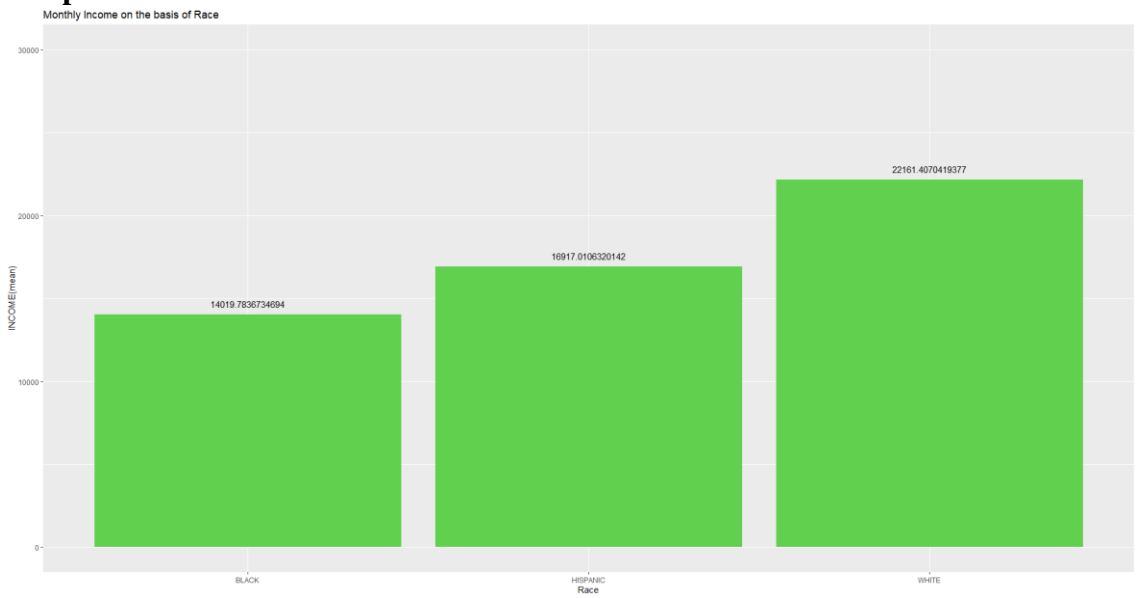
Table 3: Summary of income based on Race

	RACE	mean(INCOME_)	mean(NET_WORTH_)	count	proportion
1	WHITE	22161.41	79633.56	4459	0.504
2	BLACK	14019.78	18379.62	2695	0.305
3	HISPANIC	16917.01	40653.25	1693	0.191

Bar plot 2: Net worth based on Race



Bar plot 3: Income based on Race



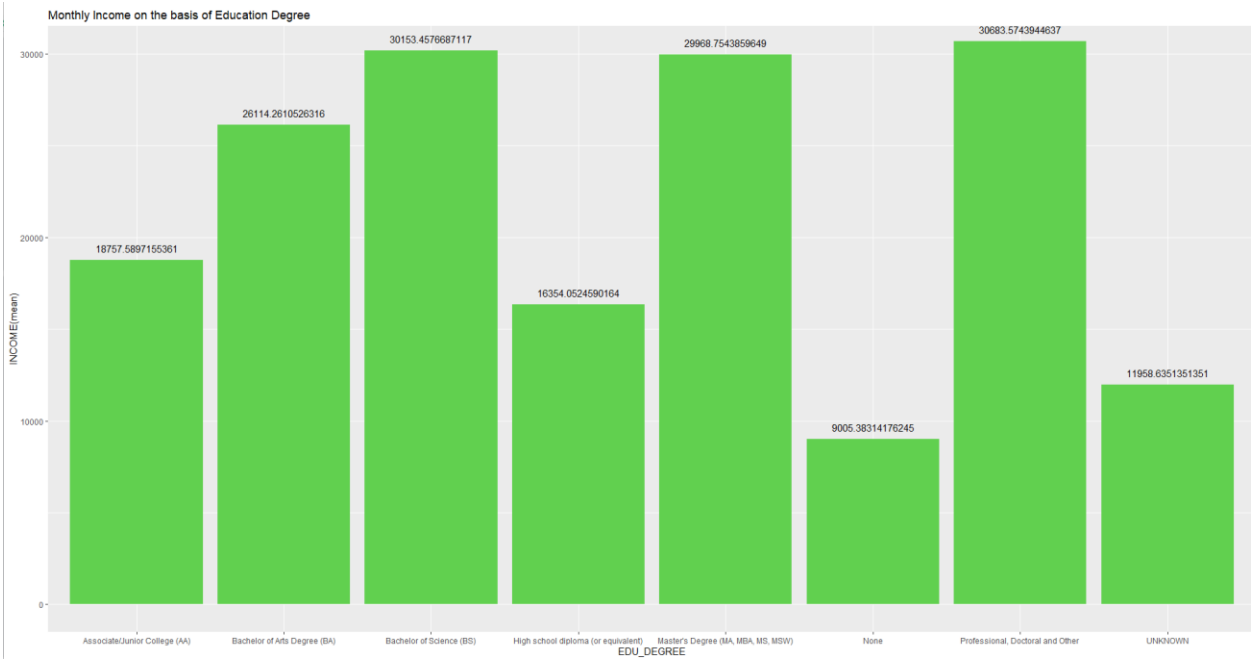
From the above table and bar plot we can say that people of race white have the highest income and net worth whereas people of race black have the least.

- **INCOME BASED ON EDUCATION QUALIFICATION**

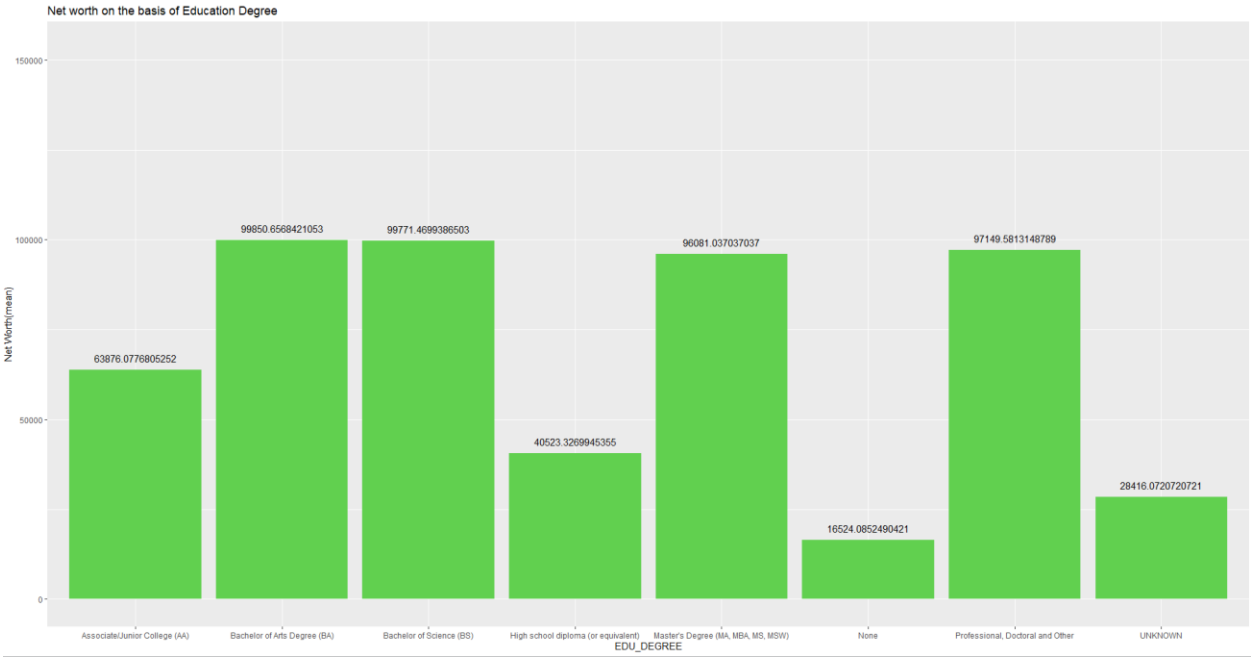
Table 4: Summary of income based on educational qualification

	EDU_DEGREE	mean(INCOME_)	mean(NET_WORTH_)	count	proportion
1	High school diploma (or equivalent)	16354.05	40523.33	4575	0.517
2	None	9005.383	16524.09	1044	0.118
3	Associate/Junior College (AA)	18757.59	63876.08	914	0.103
4	Bachelor of Science (BS)	30153.46	99771.47	815	0.092
5	Master's Degree (MA, MBA, MS, MSW)	29968.75	96081.04	513	0.058
6	Bachelor of Arts Degree (BA)	26114.26	99850.66	475	0.054
7	Professional, Doctoral and Other	30683.57	97149.58	289	0.033
8	UNKNOWN	11958.64	28416.07	222	0.025

Bar plot 4: Income based on Educational Qualification



Bar plot 5: Net worth based on Educational Qualification



From the above plots of income based on educational qualification we can see that individual with professional Doctoral or other degree have the highest income where as individual with no degree have the least income. But when we check the net worth plot, we can see that individual with bachelor of arts degree have the highest net worth where as the individual with no degree have the least income.

Conclusion

- As per the marital status, income of the individual with marital status as married have the highest income and net worth.
- As per the Race, income of the individual of race white has the highest income and net worth.
- As per Qualification, income of individual of qualification of doctoral has highest income.
- From this project I learnt that the individual who has highest education qualification can earn more than the rest.

Bibliography

- Kabacoff, R. (2011). *R in action: Data analysis and graphics with R*. Manning.
- Holtz, Y. (n.d.). *Barchart | the R Graph Gallery*. The R Graph Gallery. https://www.r-graph-gallery.com/barplot.html
- *Multiple condition if-else using dplyr, custom function, or purr*. (2018, August 26). Stack Overflow. https://stackoverflow.com/questions/52028764/multiple-condition-if-else-using-dplyr-custom-function-or-purr