Exploratory Analysis of Socio-economic factors of Demographics in England (2021)

1 Introduction

This is a report of my investigation of a modified snapshot of data collected during a household census conducted in England in 2021. I performed exploratory data analysis of the dataset to derive insights about socio-economic conditions of the data subjects based on gender, income, level of education, ethnicity, marital status, and housing conditions.

1.1 The Data

The data had 9 columns (`ID`, `Person_ID`, `Age`, `Mar_Stat`, `INC`, `Female`, `H8`, `Eth`, `Highest Ed`) and 27,410 entries with each entry representing an individual. Thanks to Wickham et al. (2019) and Arnold (2021), I was able to perform analysis of this dataset in RStudio using packages in the R tidyverse and ggthemes libraries.

1.2 Data Transformation

I performed some data transformations on the dataset including renaming columns, transforming existing columns and creating new columns because the dataset did not follow the three interrelated rules that make a dataset tidy. Wickham (2014)

2 Analysis and Insights

I performed analysis of the dataset majorly by Ethnicity and Gender demographics, on 4 metrics: Income distribution, Level of Education, Marital Status, and Housing conditions

2.1 Income Distribution

I am interested in the distribution of income among the working class (above 18 years).I am particularly interested in the `Income` distribution by `Ethnicity` and `Gender`.
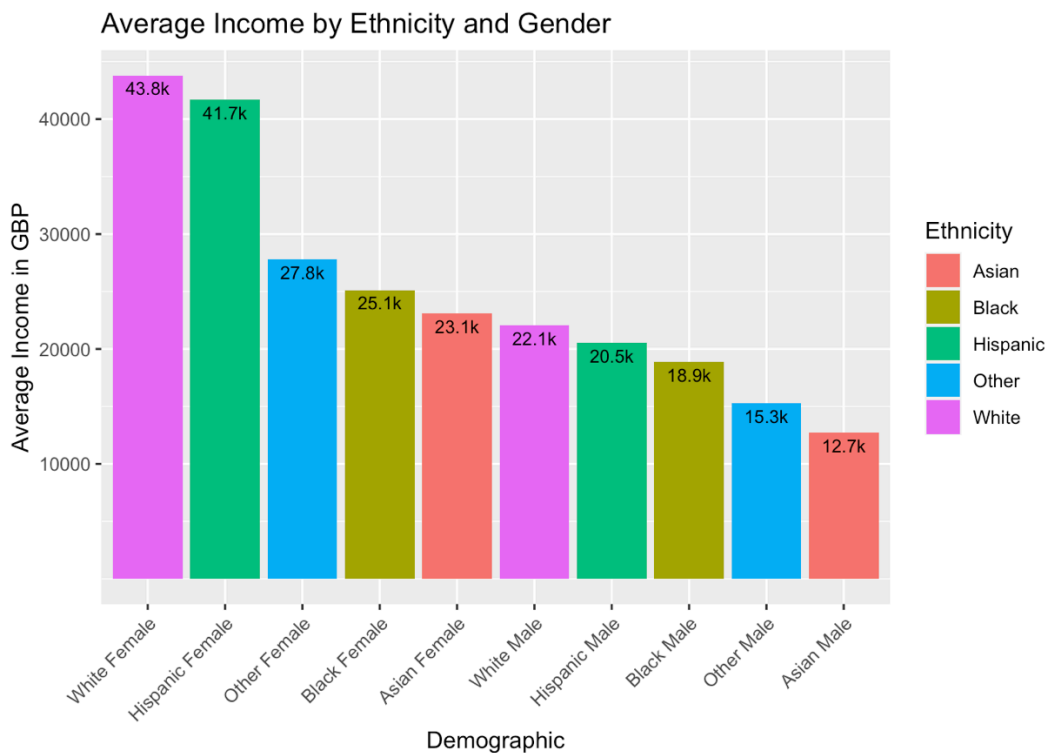

Figure 1: Average Income by Ethnicity & Gender

[Figure 1](#) shows the average income of demographies in England by their `Gender` and `Ethnicity`. It shows that the highest earning demographic are White females with an average annual income of 43,806 GBP while the lowest earning demographic are Asian Males with an average annual income of 12,701 GBP. It also appears that females earn more on average than Males, irrespective of their ethnicities. White Males, the highest earning male demographic earn an average of 22.1k GBP, almost 50% lower than the highest earning female demographic. Huge disparity in Income can be observed here.

2.2 Level of Education

Let us examine the Income distribution by level of education, and the

highest level of education of these demographics to see if any correlation can be observed between level of education and Income. This can help in explaining the income distribution observed by Ethnicity and Gender.
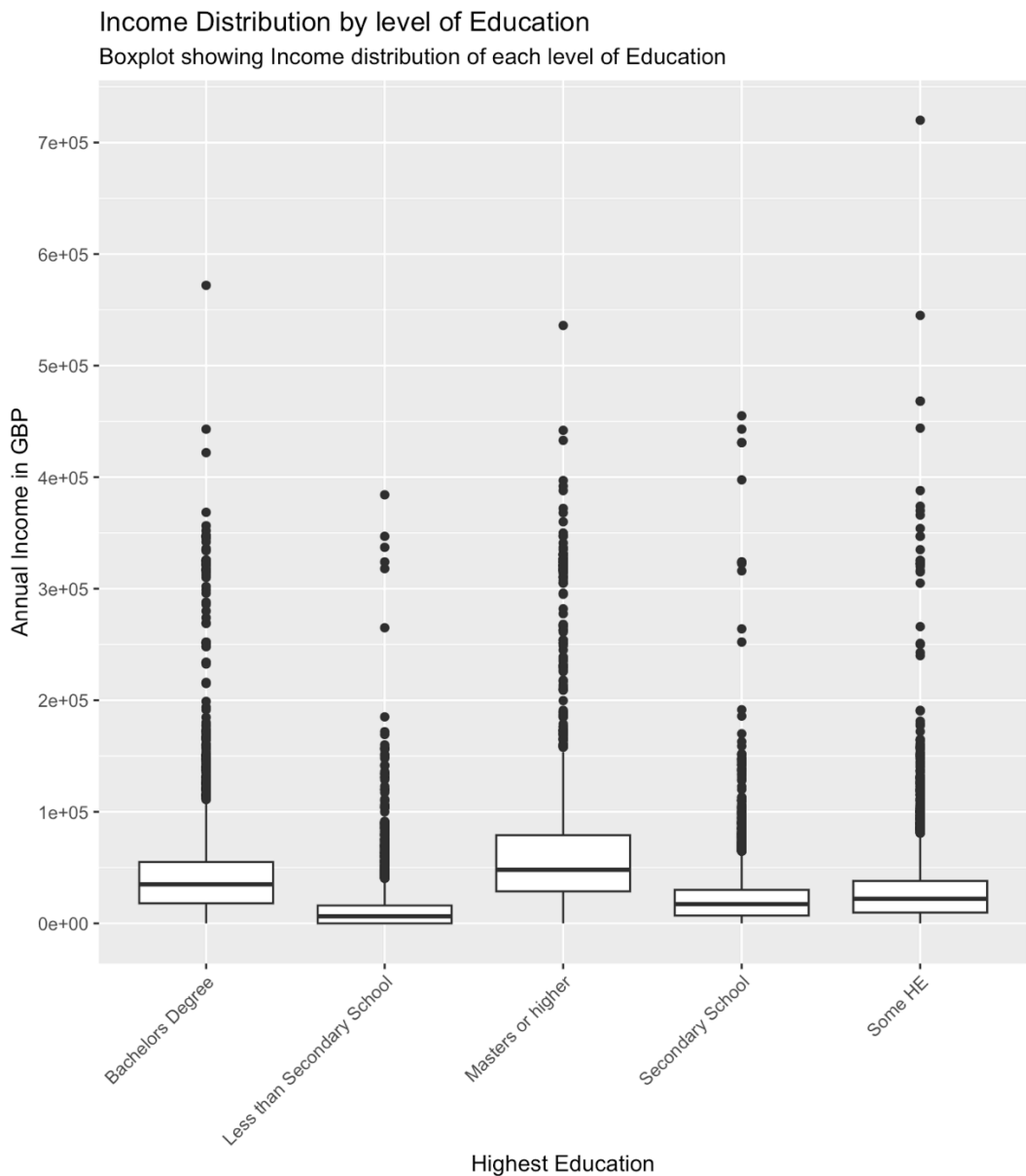

Figure 2: Income Distribution by level of Education

```
# A tibble: 10 × 3
# Groups:   Ethnicity [5]
   Ethnicity Gender Education
   <chr>     <chr>  <chr>
 1 Asian     Female Less than Secondary School
 2 Asian     Male   Less than Secondary School
 3 Black     Female Secondary School
 4 Black     Male   Less than Secondary School
```

```
 5 Hispanic  Female Some HE
 6 Hispanic  Male   Less than Secondary School
 7 Other     Female Secondary School
 8 Other     Male   Some HE
 9 White     Female Some HE
10 White     Male   Secondary School
```

Clearly, there appears to be some correlation between level of education and average income. The top 3 earners have majority of their demographic having some higher education (`Some HE`) while the bottom 3 earners have majority with `Less than Secondary School` education. `Other Female` and `Other Male` being outliers in both observations. The boxplot in [Figure 2](#) shows as expected, that individuals of working age with a `Masters or higher` earn more on average than others. Infact, the 50th percentile (top 50%) of holders of `Masters or higher` outearn every other individuals of lower levels of education, except the 75th percentile (top 25%) of holders of a `Bachelors Degree` . On the other hand, individuals with `Less than Secondary School` earn significantly low with at least 50% of them earning lower than every other individual of working age. As expected all categories of individuals have outliers who have high incomes. These outliers may skew the average, as observed in [Figure 1](#).

2.3 Divorce and Separation Rates

I would now like to examine the marital status and divorce/separation rates of the english populace. According to Karney ([2021](#)), lower socio-economic status couples are more likely to get divorced than higher-SES counterparts.
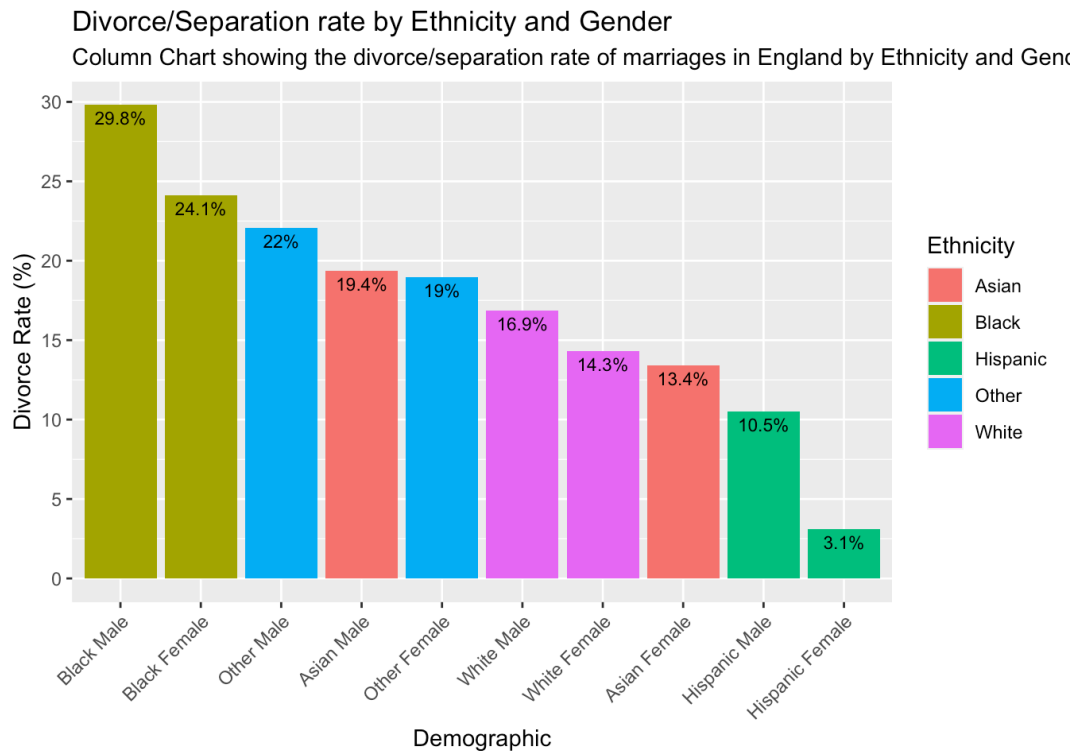
Divorce/Separation rate by Ethnicity and Gender
Column Chart showing the divorce/separation rate of marriages in England by Ethnicity and Gend

Figure 3: Divorce/Separation rates by ethnicity and gender

[Figure 3](#) supports the findings of Karney ([2021](#)). Demographics of lower socio-economic status in terms of level of education and average income appear to have a higher separation rate. Black males have the highest divorce/separation rate of all demographics in England (2021) at 29.8%. They are closely followed by Black females at 24.1%. Other Males and Asian Males follow closely at 22% and 19.4% respectively. Hispanic Males and Females have the lowest separation rates at 10.5% and 3.1% respectively.

Copen et al. ([2012](#)) proved from analysis that marriages of women with a college degree last longer than marriages of women with a high school diploma or less. This is supported by results from my analysis showing Hispanic Females and White females with majority having some higher education having significantly less separation rates than their counterparts with less education, with an exception of Asian females.

2.4 Housing Condition

Lastly, I am interested in the housing and living conditions of demographics in England. The question H8 gives insights into the type of

housing unit the subjects live in (i.e, whether its a `Self-Contained` Unit or a `Non-Self-Contained` Unit).

2.4.1 Housing Unit Type

Are you also wondering what type of housing units the majority of each Ethnicity live in?
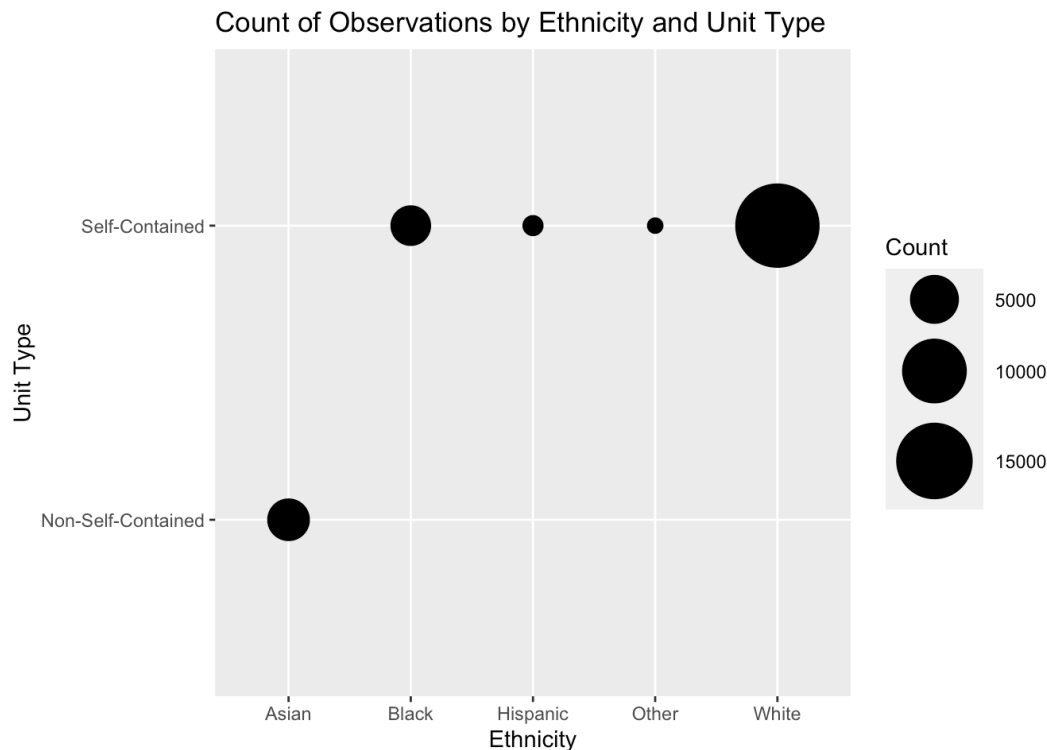

Figure 4: Count of Observations by Ethnicity and Unit Type

The circles in [Figure 4](#) represent the mapping of the count of observations at each Ethnicity – Unit type location. The area of each circle indicates the count of each observation. Asians, who happen to be the least educated demographic and arguably the least earning demographic are the only ethnicity with majority of them living in accommodation where all the rooms, including the kitchen, bathroom and toilet are NOT behind a door that only each household can use. In fact, ALL non-self-contained units in the dataset are occupied by Asians and ALL Asians live in non-self-contained units.

On the other hand, ALL Blacks live in Self-contained units, likewise Hispanics, Whites and Others. There is a huge count of whites as observed from the plot above due to a higher White population (18,929).

Just as the plot area for Hispanics and Others is really small due to the small population of both demographics in the observation (936 and 853 respectively).

2.4.2 Average Number of Occupants in a household

I was also wondering which ethnicities have the highest and lowest average number of occupants within each household.

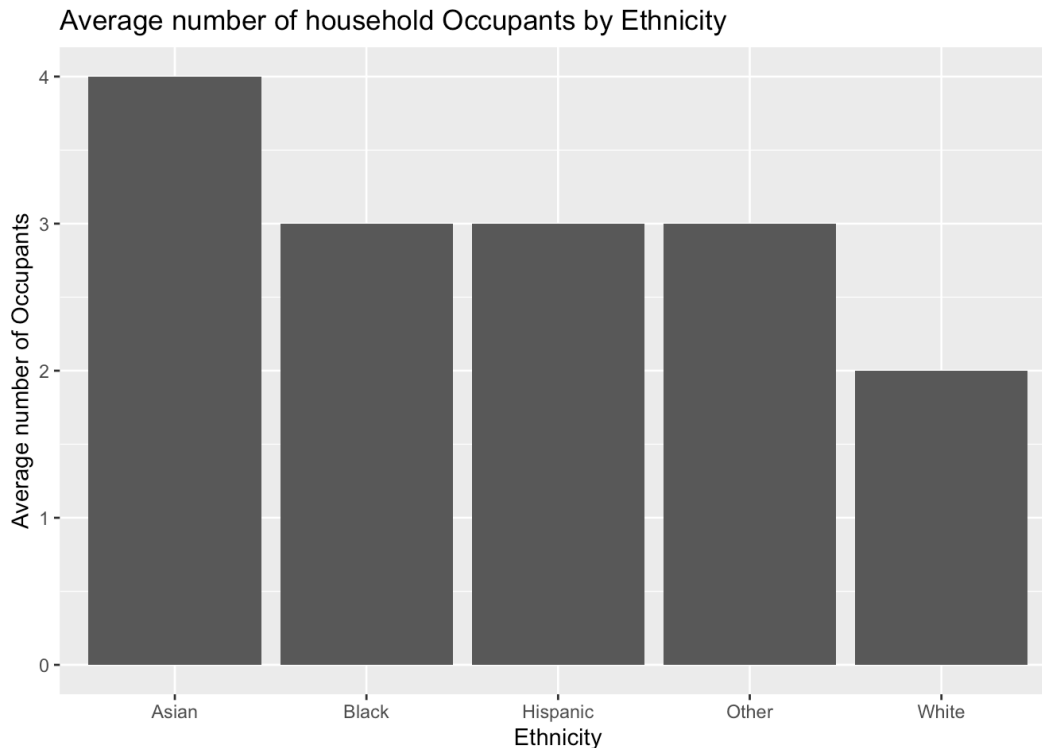Average number of household Occupants by Ethnicity

Figure 5: Average number of household Occupants by Ethnicity

It turns out that the Asians have the highest average number of occupants in a household, despite also having the least favorable kind of housing unit. I have now established that the Asians have the worst housing/living condition of all ethnicities in England, and the Whites have the most favorable living condition, considering that they have the least average number of occupants per household, and relatively good housing units

2.4.3 Average Number of Children per household

Additionally, I would like to know the ethnicity which has the highest average number of kids per household. The output below shows that the

Asians have the highest number of children on average than all other ethnicities, with the Whites having the least average number of children per household.

```
# A tibble: 5 × 2
  Ethnicity average_no_of_children
  <chr>                      <dbl>
1 Asian                       1.20
2 Other                       1.17
3 Black                      0.810
4 Hispanic                   0.753
5 White                      0.581
```

3 Conclusion

I have been able to derive some interesting insights about the socio-economic status of demographics in England from the hypothetical census dataset. It has been established that White Females earn higher on average than every other demographic in England while Asian males earn the least. On average, females earn higher than males irrespective of their ethnicities.

I have also established that Education always pays because a Master's degree or higher earns you the highest average income in England. Individuals with less than secondary education earn the least.

Divorce/Separation Rates are also higher in Blacks than other ethnicities, with black males having the highest divorce rate. Hispanics on the other hand have the lowest divorce rates with the hispanic females having the lowest divorce rate of any demographic in England.

Lastly, Asians have the worst housing and living conditions, because they live in non-self-contained units, have the highest average number of occupants, and highest average number of children in each household.

---

4 References

Arnold, Jeffrey B. 2021. "Ggthemes: Extra Themes, Scales and Geoms for 'Ggplot2.'" https://CRAN.R-project.org/package=ggthemes.

Copen, Casey E., Kimberly Daniels, Jonathan Vespa, and William D. Mosher.

2012. "First Marriages in the United States: Data from the 2006-2010 National Survey of Family Growth." *National Health Statistics Reports*, no. 49: 1–22.

Karney, Benjamin R. 2021. "Socioeconomic Status and Intimate Relationships." *Annual Review of Psychology* 72 (1): 391–414. https://doi.org/10.1146/annurev-psych-051920-013658.

Wickham, Hadley. 2014. "Tidy Data." *Journal of Statistical Software* 59 (10). https://doi.org/10.18637/jss.v059.i10.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the Tidyverse" 4: 1686. https://doi.org/10.21105/joss.01686.