Name: Sultan Banabila

Course: INFO 281

Project Plan

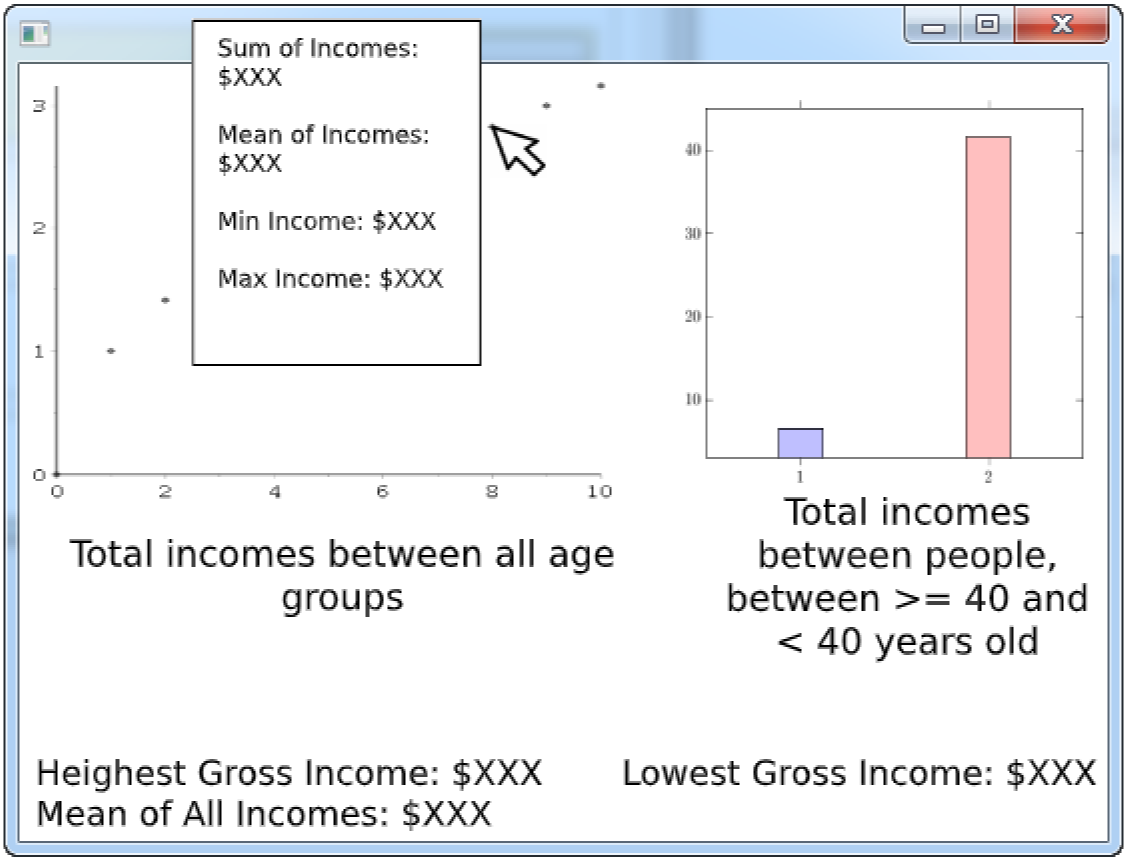
The aspect of inequality I am intending to explore is about income distribution within New Zealand; that is, by comparing people's incomes in order to find out who earns the most amount of incomes based on the given dataset with respect to their age groups. That is, by dividing the people in the dataset into two different groups, the first of which includes people who are forty years old or older, and the second group consists of the people who are younger than forty years old; then, plotting the total incomes of both groups. In addition, I will be aiming to display a chart to describe the total incomes of individuals based on all the provided age groups in the dataset. That is, to find out if there is a certain group of people, that have something in common, who owns more wealth than others. On the other side of the spectrum, I will also be exploring if there is a certain group of people, who have some common attributes that might be earning less weekly income than the rest of the population.

The data that will be used in the project is taken from the old Stats NZ website. Up to this point, I have been able to find one artificial dataset that reflects real life statistics without giving away any real information about the participants, which is based on the results of multiple income surveys, that were conducted in New Zealand in the year of 2011. The dataset is the best and most recent one that is available at the moment, as there does not seem to be any newer versions available at the time of writing this report, that have the same format as the spreadsheet that I was able to find; furthermore, after looking through the new Stats NZ website, it seems that they have decided to post all the upcoming data, which relate to income inequality to a new site called “Well Being Indicators”. Moreover, when looking through the aforementioned website, it seems that there will be a newer version of the dataset, however, it will not be released until early next year. Hence, I will not have access to it during the duration of this course.

Moving on, the categorical data in the available CSV file is described using numeric codes, the translations for which are available on another spreadsheet, which means that the dataset has to be refactored into a readable format, in order to make it easier to understand. Hence, I wrote an R script, that goes through the categorical data and translates the numeric codes into what they refer to in words instead. After that, I stored the translated data set into a data frame in R, which I then wrote into a CSV file in order to make it faster to load, and easily accessible from other R script files.

As for the content of dataset itself, it is divided into multiple columns describing the status of each of the participants, such as, which age group does each individual fall in, where the age groups start from 15-20 to 65+ years old. Furthermore, the results of the surveys also indicate the person’s sex, ethnicity, which region of the country they are based in, their highest qualification, what the participant’s occupation is at the time of the recording, how many hours they work per week excluding personal employment. Therefore, it makes sense to see that the person who owns the most amount of income per week is shown as if they work for 0 hours according to the dataset; even though, that is probably not the case at all, and lastly, how much money does each participant earn per week in gross income from all sources.

Wireframe


3
2
1
0
0
2
4
6
8
10
Total incomes between all age groups
40
30
20
10
1
2
Total incomes between people, between >= 40 and < 40 years old

Sum of Incomes: $XXX

Mean of Incomes: $XXX

Min Income: $XXX

Max Income: $XXX

Heighest Gross Income: $XXX
Mean of All Incomes: $XXX

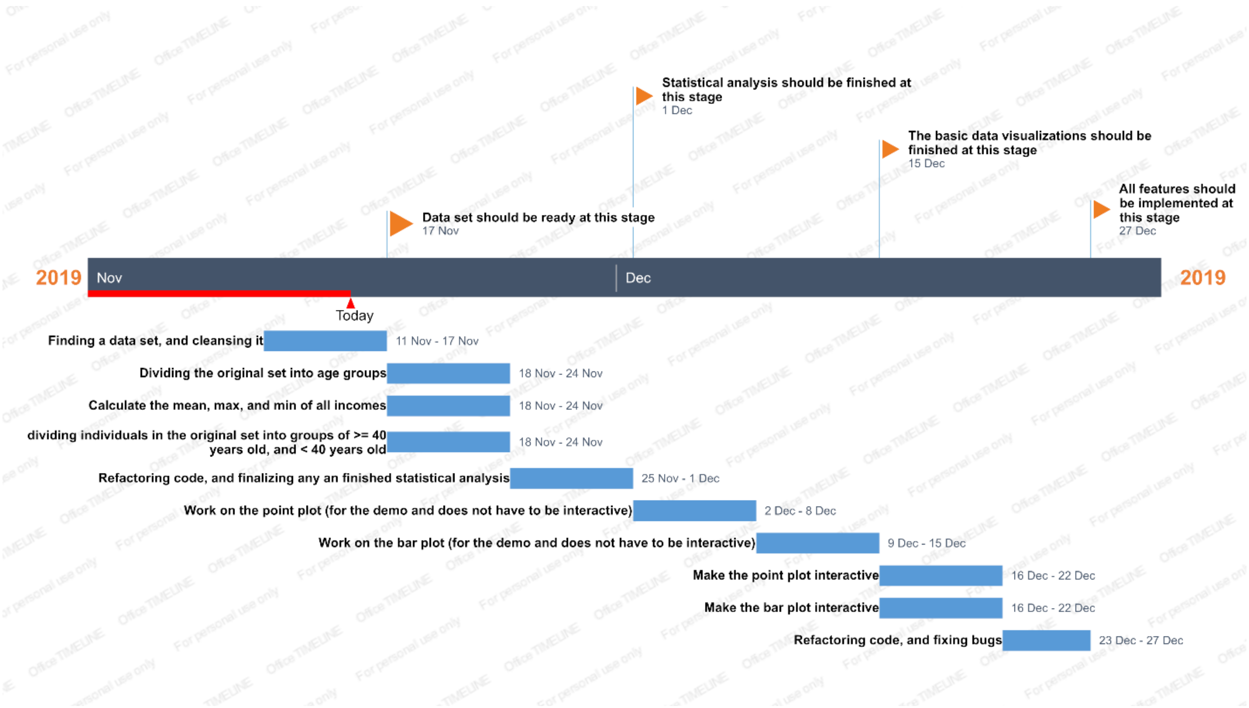
Lowest Gross Income: $XXX

The wireframe describes what the project's appearance should look like. As seen above, there are two different charts, the one on the top left describes the total incomes of all the participants in the aforementioned survey based on their age groups, furthermore, that chart should be interactive; such that, if the user clicks on one of the points on the plot, it should display a small popup message indicating specific data about that point.

On the other hand, the second plot should display total incomes by dividing the survey’s participants into two separate groups, the first of which includes people who are forty years old or older, and the second group should contains everyone who is younger than forty years old.

Finally, the bottom part of the wireframe displays statistical analysis of the entire dataset, which will be used to assess whether there is an imbalance in income distribution within the given dataset or not.

Workflow



As can be seen from the project plan, I intend on having a clean data set by the end of the first week, in order to be able to start working on it on the week after that. Furthermore, on the second and third week I will be applying statistical analysis on the data to find out income statistics, which is needed to visualize aspects of inequality in income distribution; for instance, I will be calculating the maximum, and minimum incomes of the entire dataset, as well as, dividing individuals into age groups in order to figure out the total of how much each group earns on a weekly base. Finally, I intend on using the analysed data to make two different plots, the first of which, indicates the total incomes based on age groups, and the second one should display the same information based but for people who are forty years old, or older, and individuals who are younger than forty years old. Finally, in the second to last week, I will be working on making both charts interactive; such that, users should be able to see more specific information when clicking on certain point in the plots.

Citations

- “Income Inequality.” Income Inequality | Ngā Tūtohu Aotearoa – Indicators Aotearoa New Zealand, https://wellbeingindicators.stats.govt.nz/en/income-inequality/.“New Zealand Income Survey 2011 CART SURF.

- ” New Zealand Income Survey 2011 CART SURF, Stats NZ, http://archive.stats.govt.nz/tools_and_services/microdata-access/nzis-2011-cart-surf.aspx.