

Second Project Check-in

Labour Force Dataset

Team Members: Ali Arab,
Parnian Taghipour,
Bita Azari

Table of Contents

Dataset Overview	3
Sketchable-data subset	4
10 + 10 ideation	5
10 visualization	6
10 variation of our selected sketch	16
Discussion of choices and directions	26
Implementation	30
Insights learned from the data	33
Conclusion	34
Appendix	36

Dataset Overview

Our dataset is Labour Force Survey which records monthly information on employment and unemployment in Canada. It has more than 60 variables gathered from non-institutionalized individuals 15 years of age and over. The sample size is approximately 100,000 individuals. The data is available through a set of tables.

The data can be downloaded through the BC library Abacus Data server which is accessible to SFU students. Our dataset is very recent, so the extracted insights are very relevant. The large sample size, high number of variables, and the large time span covered are the three main factors that make the visualization of this dataset challenging.

In our previous check-in after getting feedback from Prof Sheelagh Carpendale, we changed 2 views that we already had. We reconsidered our dataset and questions, and tried to answer some new questions:

In the previous one, we tried to find an answer to the following questions:

- How does one's gender affect the salary?
- What is the relation between age and salary? Does gender play a role?
- Which occupations lead to higher salaries?
- Is there a difference between different provinces in Canada in terms of average wage?

In this step we added more questions:

- Is there a difference between different provinces in Canada in terms of unemployment rate?
- How is the unemployment rate changing through time in our dataset?
- In what job category are we observing more extra hours of work?
- How is gender equality in each category of job?
- How much work do participants in different job categories do and how is their salary?

In this phase, we use the following fields from the LFS dataset to answer the above questions (see details in appendix):

- Province
- Age group, in form of 5 years interval starting from 15-19, 20-24, to 70+

- Sex
- Educational attainment, ranging from 0 (middle school) to 6 (above bachelor's)
- Immigration status
- Occupation, one of 40 NOC occupation categories
- Marital status: married, single, and divorced
- Extra hours worked
- Industry of main job
- Occupation at main job
- Usual hours worked per week at main job
- Actual hours worked per week at main job
- Union status, employees only

Sketchable-data subset

In order to draw the prototypes, we created two subsets of the dataset that represents the whole population

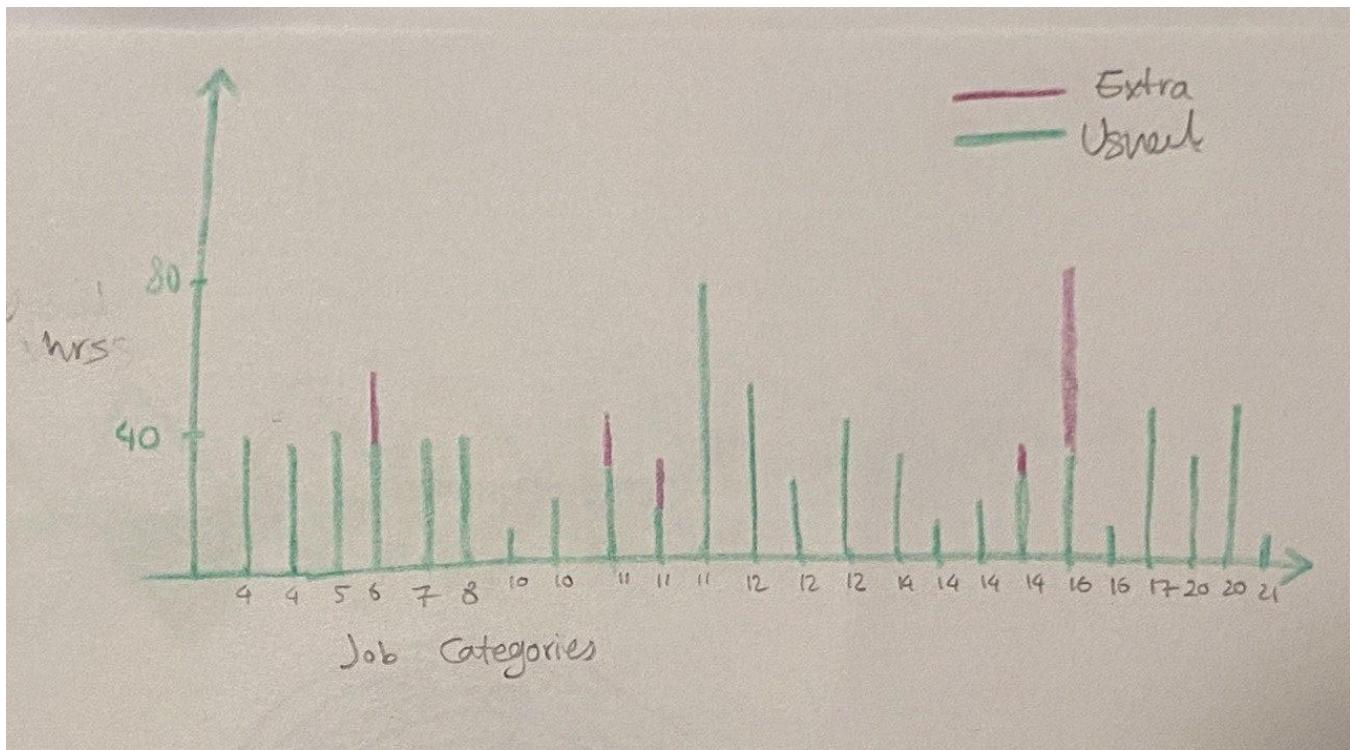
- The first data subset is drawn from only the last time point. The data subset includes 26 samples and was chosen to maximize the variation among key variables. To achieve that, in each iteration, we sorted the data based on the column, and selected several samples from both the extremities and also the samples around the mean. This process was repeated for each key column. The data subset has been attached as an excel file.
- The second dataset is drawn from 12 time points to reveal the temporal information in the data. To do so, 30 samples were selected randomly from the available data for each month. However, to make sure that the subset maintains the behaviour of the whole population, we made sure that data have the same distribution as the whole population in terms of unemployment rate.

10 + 10 ideation

We designed 10 sketches based on our new set of questions that we decided to focus on. We then started sketching and trying to answer the above mentioned questions, we were looking for the most effective and informative design.

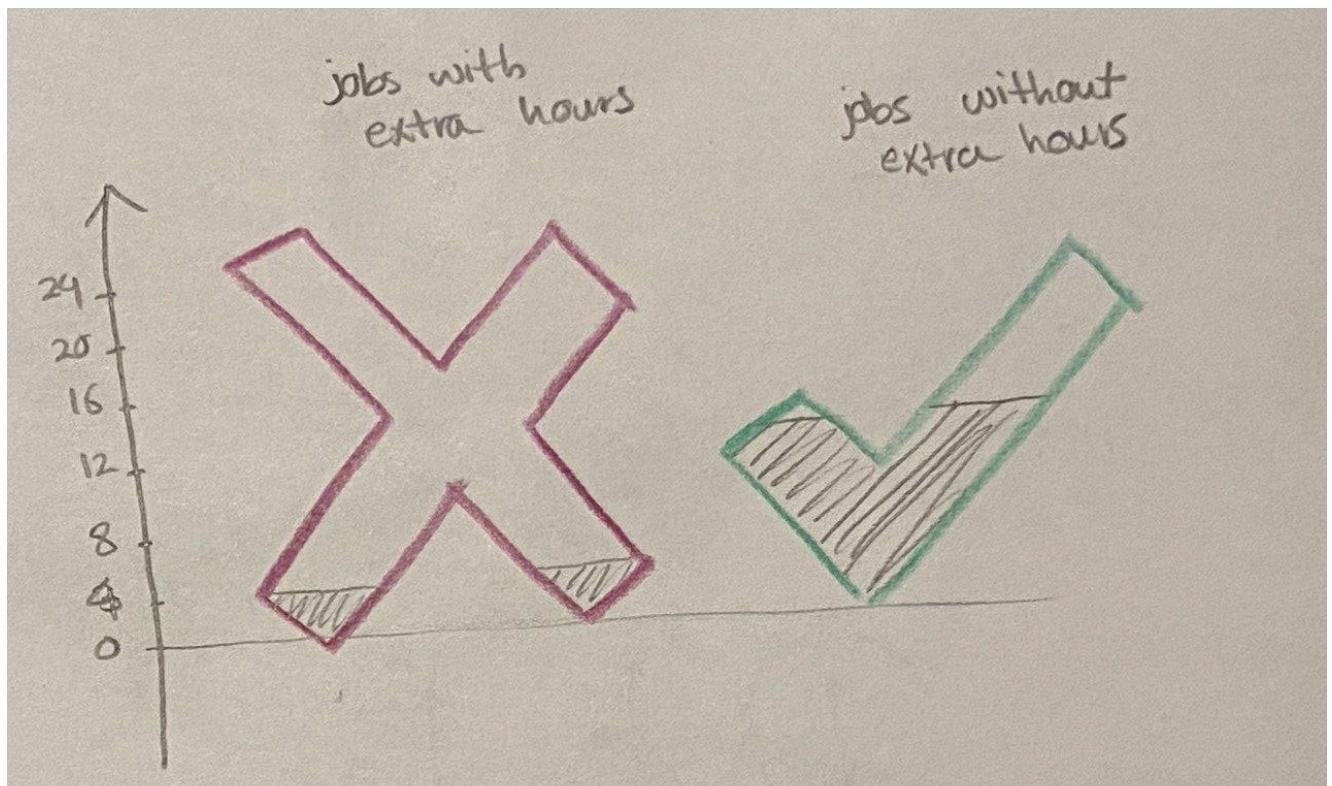
The 10 designs for our first chosen view were reported in our previous check-in report and the designs for the second one are shown below. Though for the third one which captures temporality and unemployment rate we used the same technique (10+10 ideation), the details and designs are not included in this report. The sketches are followed by their description.

10 visualization



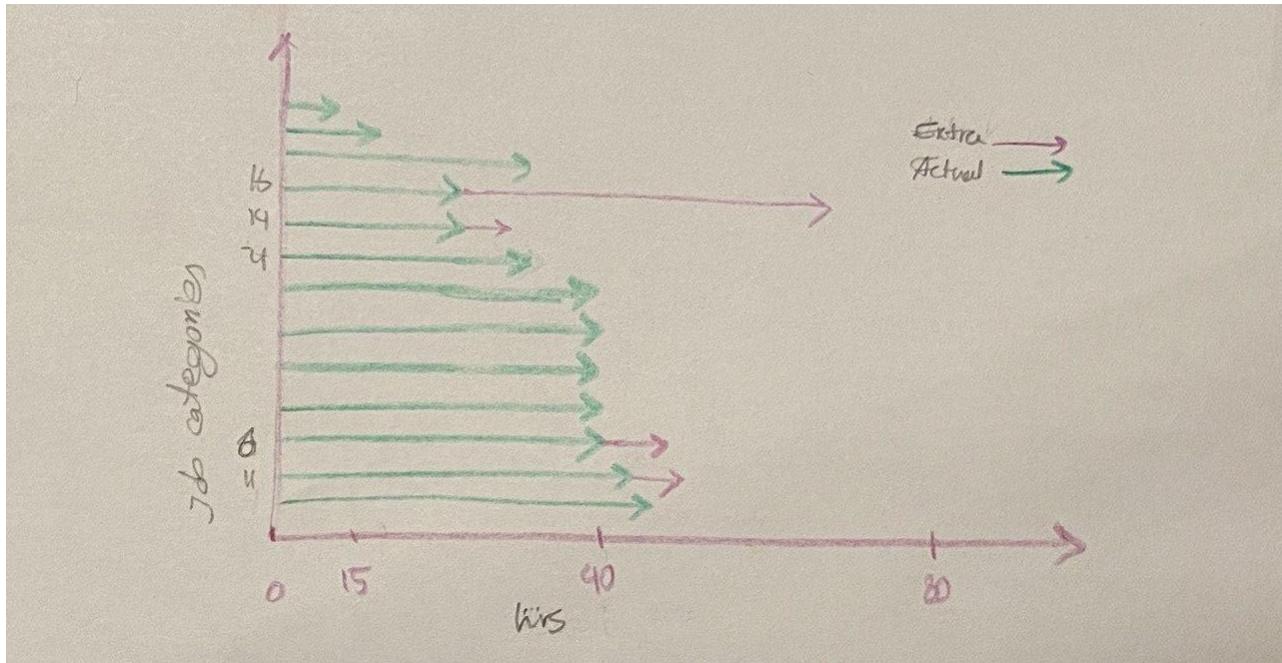
Sketch 1

This sketch shows the extra hours each person worked beyond their actual work hours.



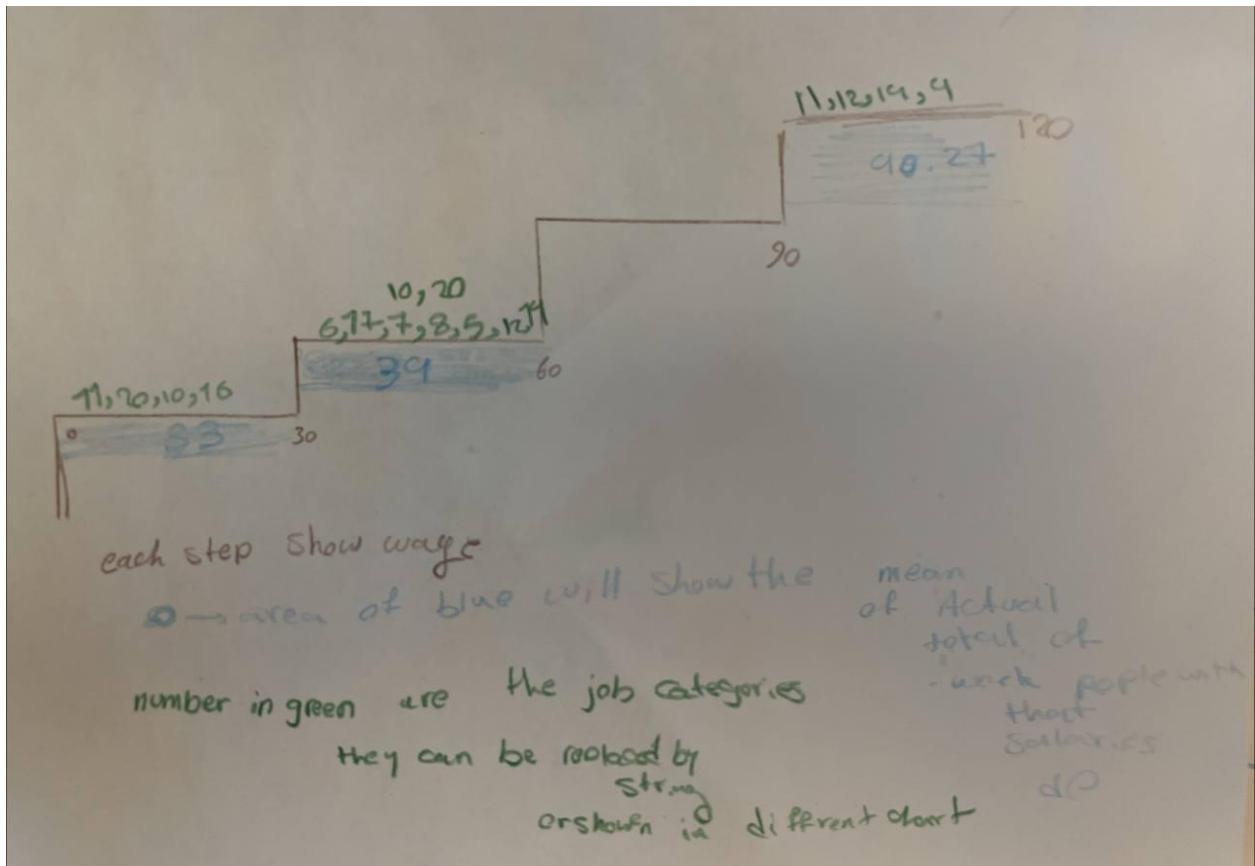
Sketch 2

Two symbols were shown to depict the fact how many job categories have employers that work extra hours. The vertical axis shows the number of data points and the colored bar shows the number of employers working extra hours vs employers working only in usual hours.



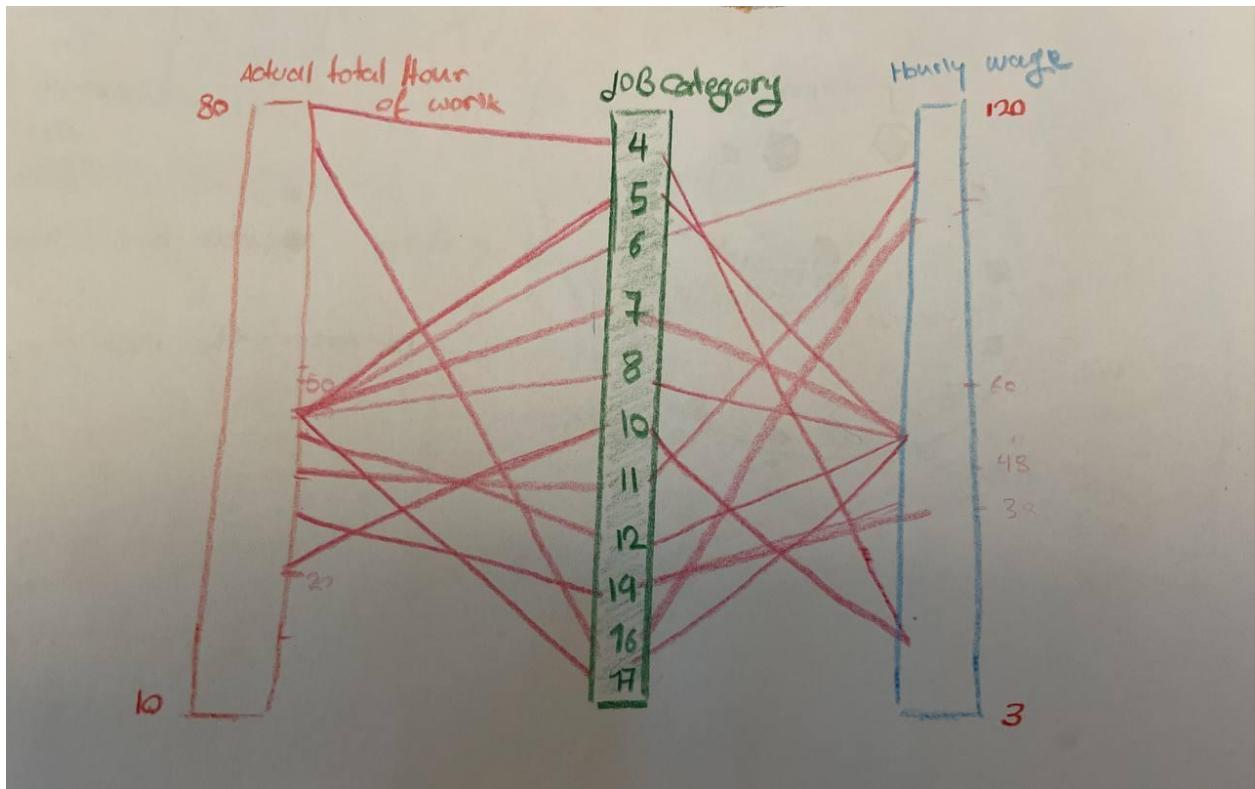
Sketch 3

In this one, we are showing the job categories and their actual time they have worked plus the red bar which is the extra hours they have worked. It is based on the average of actual hours and extra hours in each job category.



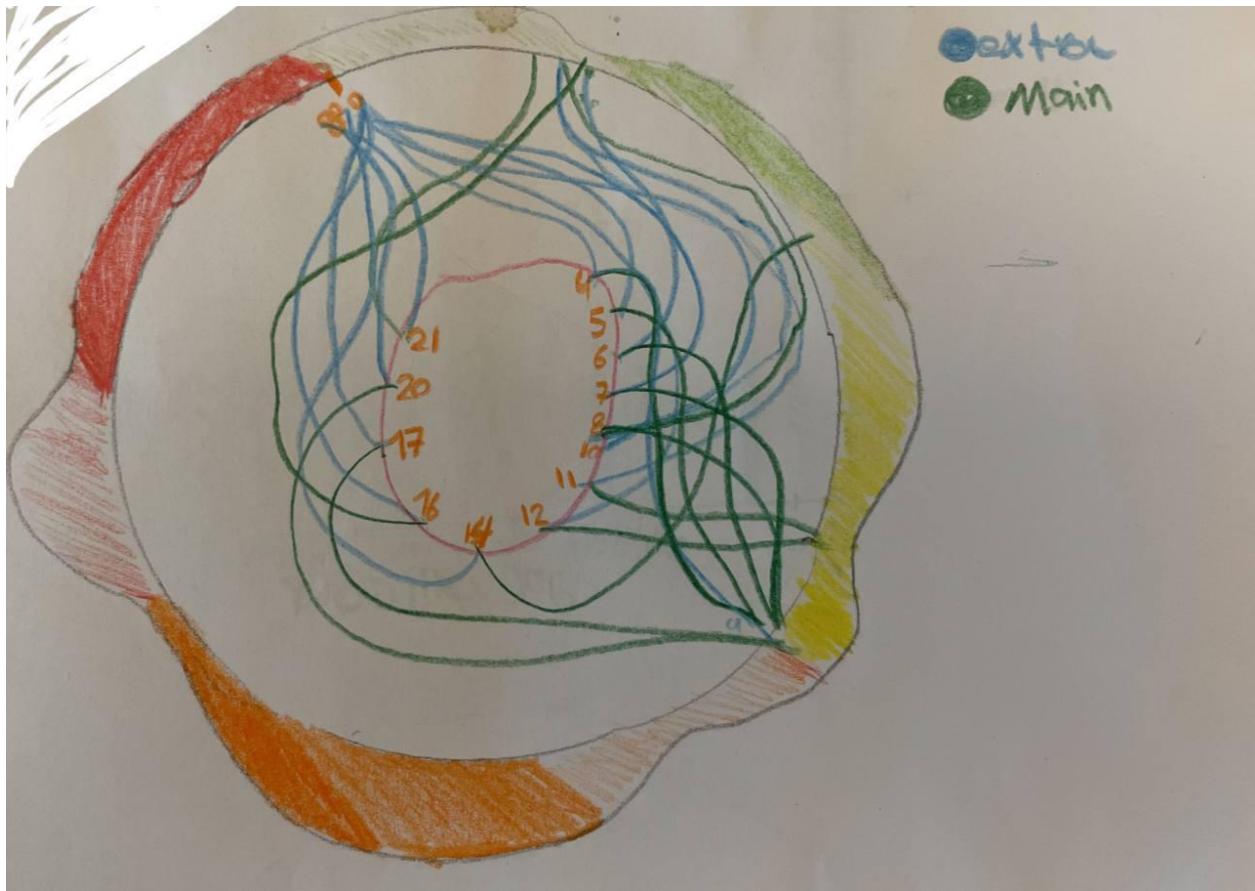
Sketch 4

Here each step will show the range of wage (number of steps can change in order to reach a balance between giving information and simplicity) and the area of the blue rectangles below them represent the Actual total work of job categories that have hourly wage in this range. Finally above the steps we will write the job categories in string form(currently it is written their numbers to save space and what each number represent can be written in a note below the visualization)



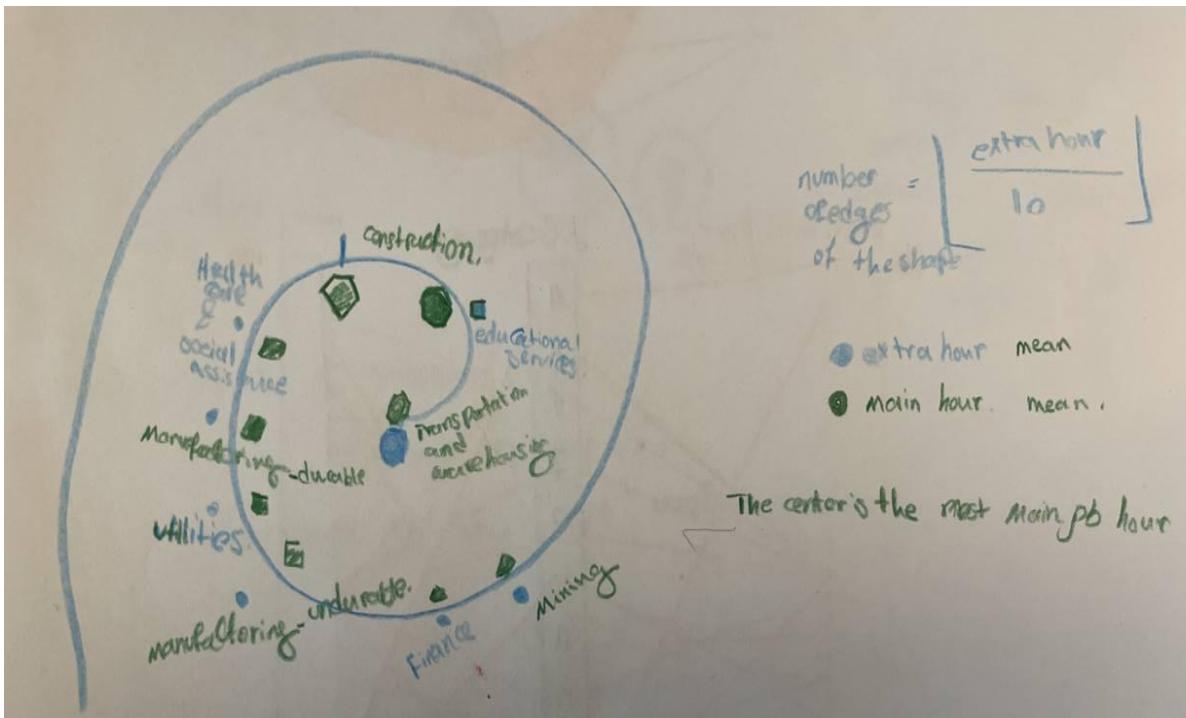
Sketch 5

In this visualization the right rectangle will show the actual total work hours and the right one will show the hourly wage for each category that is written in the middle rectangle.



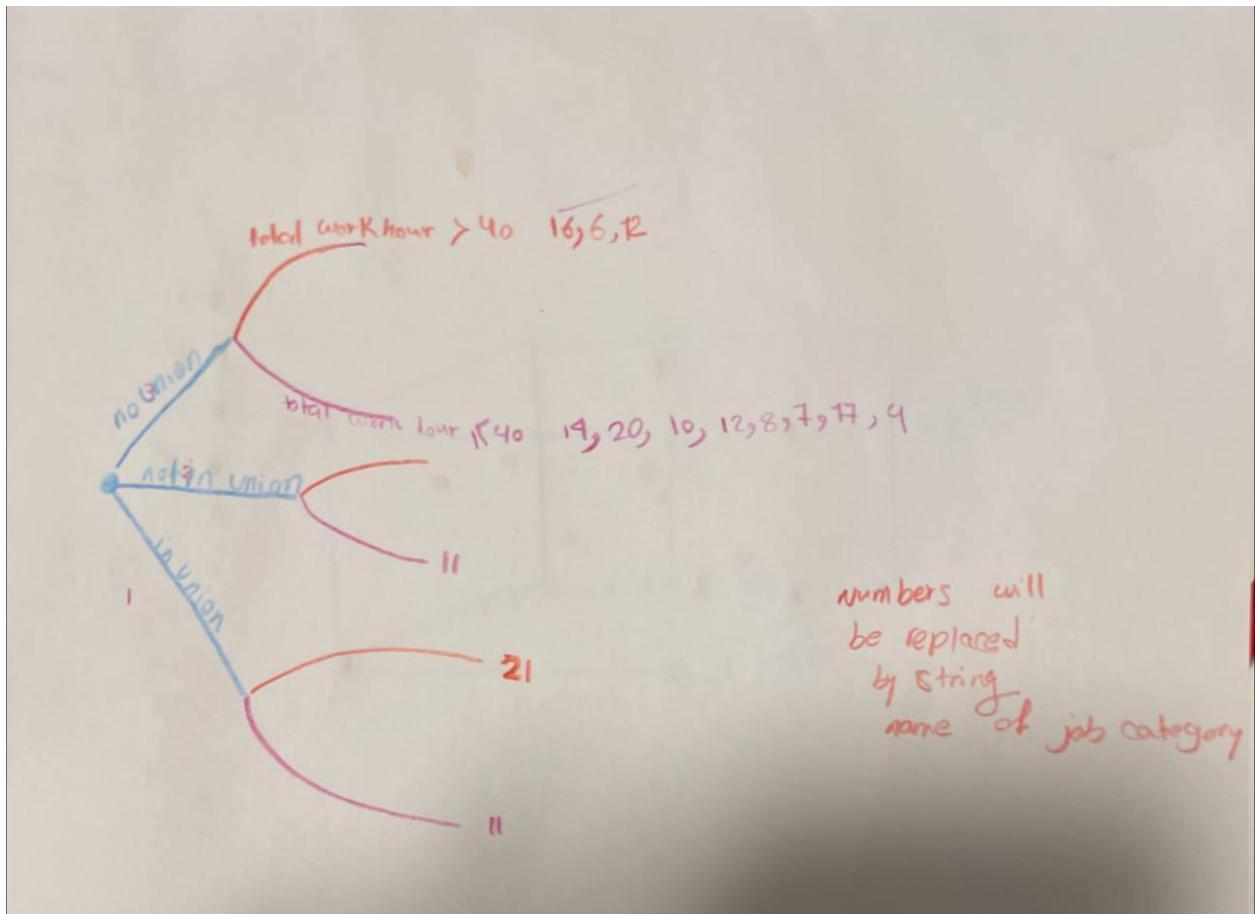
Sketch 6

The bigger circle is showing the possible hours and lines will match each category with the average of our work and the colour of the line will show which version of work hour we are working with(we can add different versions of recorded work here with different colours) . Also, it is possible to replace the numbers in the middle circle with the string format of job categories or right each category's string name separately below the visualization.



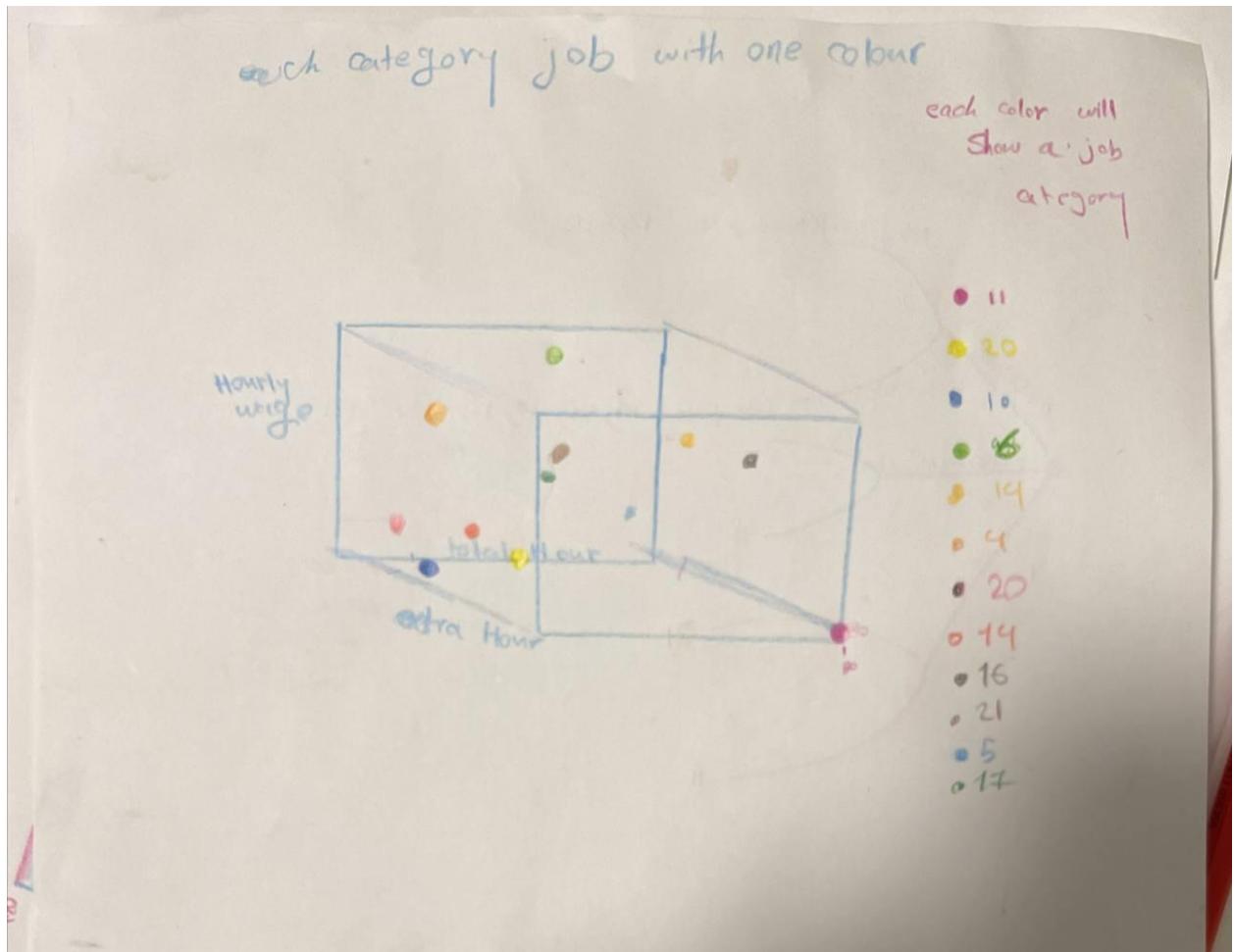
Sketch 7

The shapes on the spiral line are showing the hours of work for each job category and they are starting with the most summation of work hours for them. Also, it is possible to change their place in the way that their place on the spiral line shows their hourly wage. This way we can convey more information)



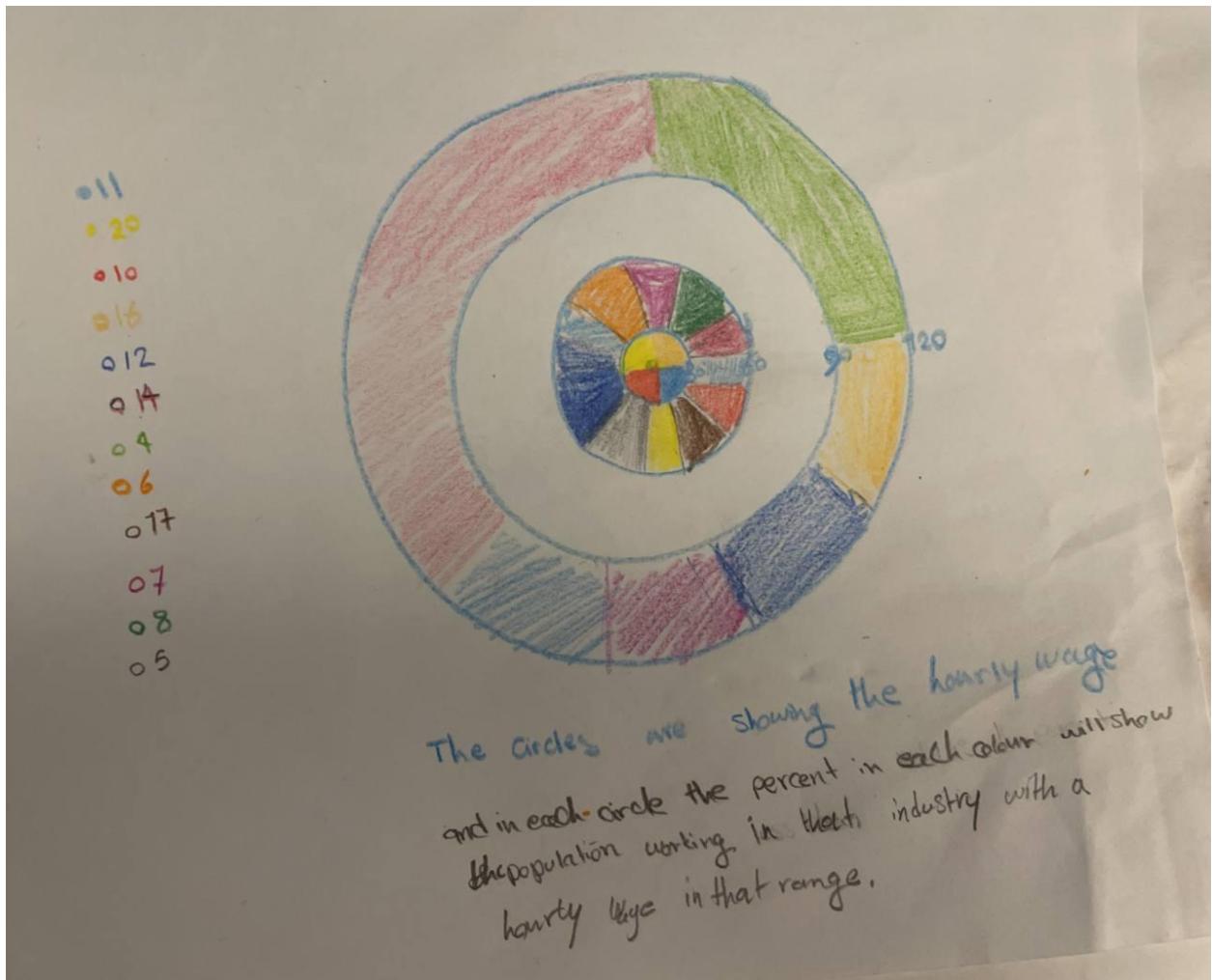
Sketch 8

This tree is at first divide categories based on having union or not(the participation in the union is also matter) and then with show which categories have how much total work hours(whether it is more than 40 or less than it)



Sketch 9

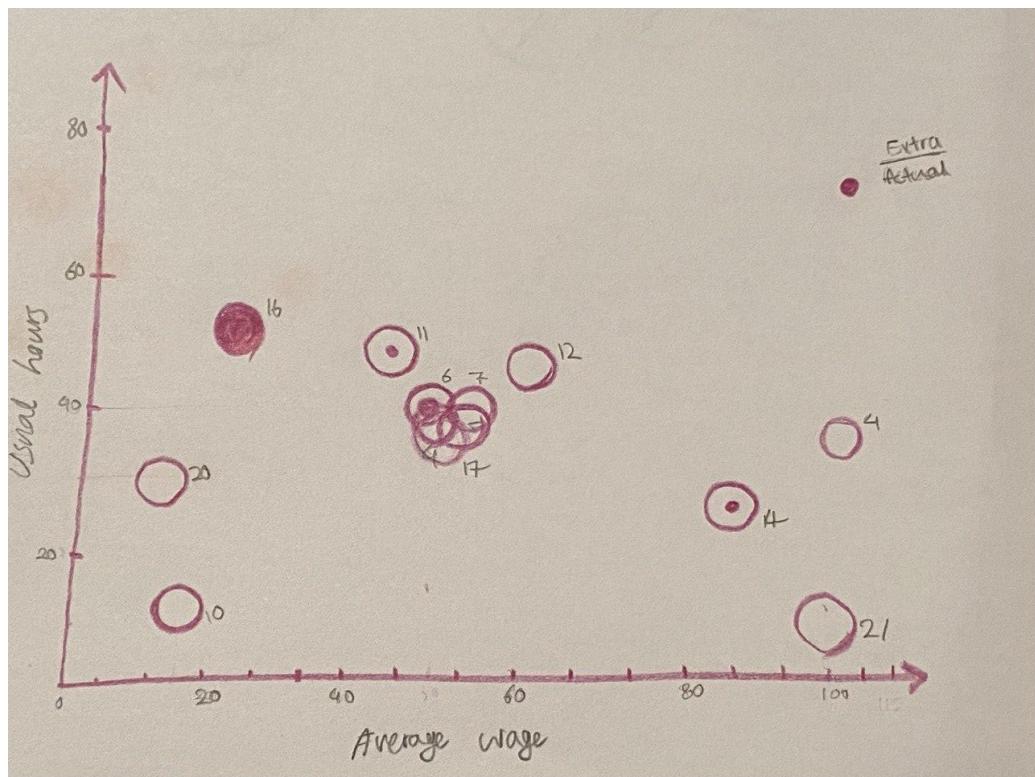
The colour of points is showing their category, and the 3 dimensions of the visualization is showing actual work hours, extra work hours and the hourly wage.



Sketch 10

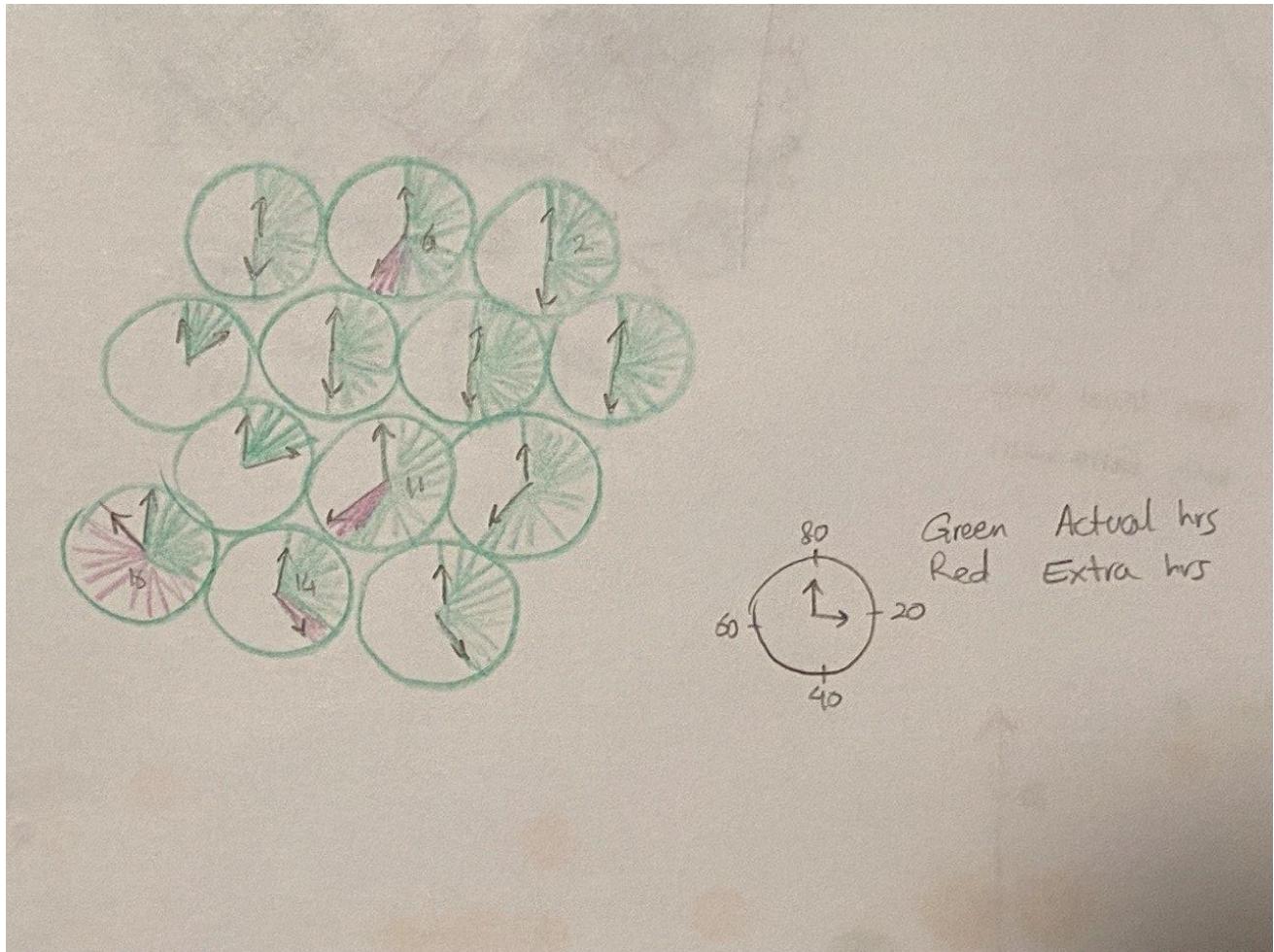
In this visualization circles are related to the hourly wage (smallest one is below 30, then 60, 90 and 120) and in each circle the colours are representing the job categories and they show what proportion of people with that wage are from those jobs. Here, we understand that a circle and using size to show measures is good for our purpose, and continue with these concepts.

10 variation of our selected sketch



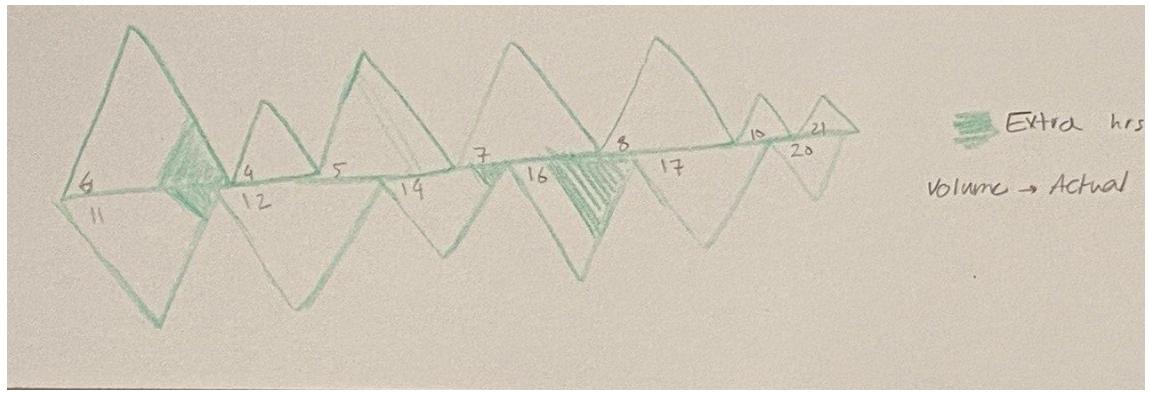
Sketch 1

Each circle is showing the relation between average wage and usual hours. The ratio of extra hours to actual hours is shown as the filled circles inside the circles. The numbers beside each circle indicate the job categories.



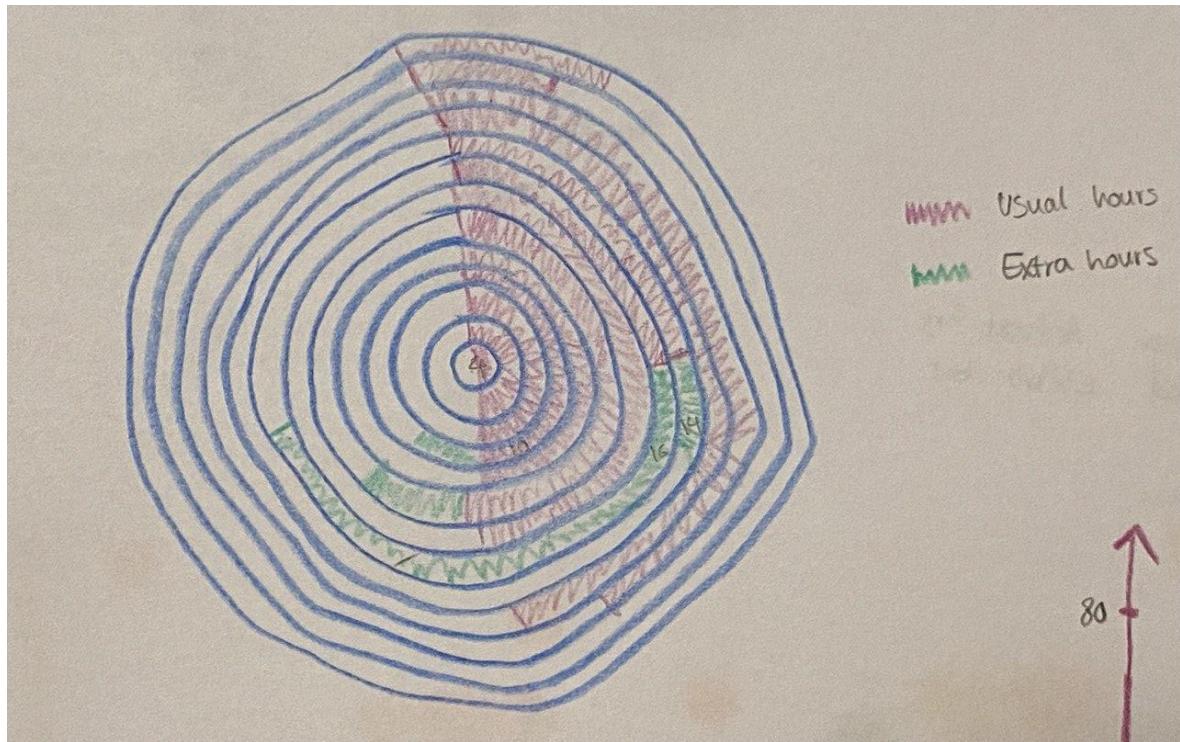
Sketch 2

In this diagram each job category is shown using a circle with 80 parts (ex. Half the clock is 40hrs) the green part shows the actual hours and the red parts are the extra hours. Numbers in each clock face are showing the job categories.



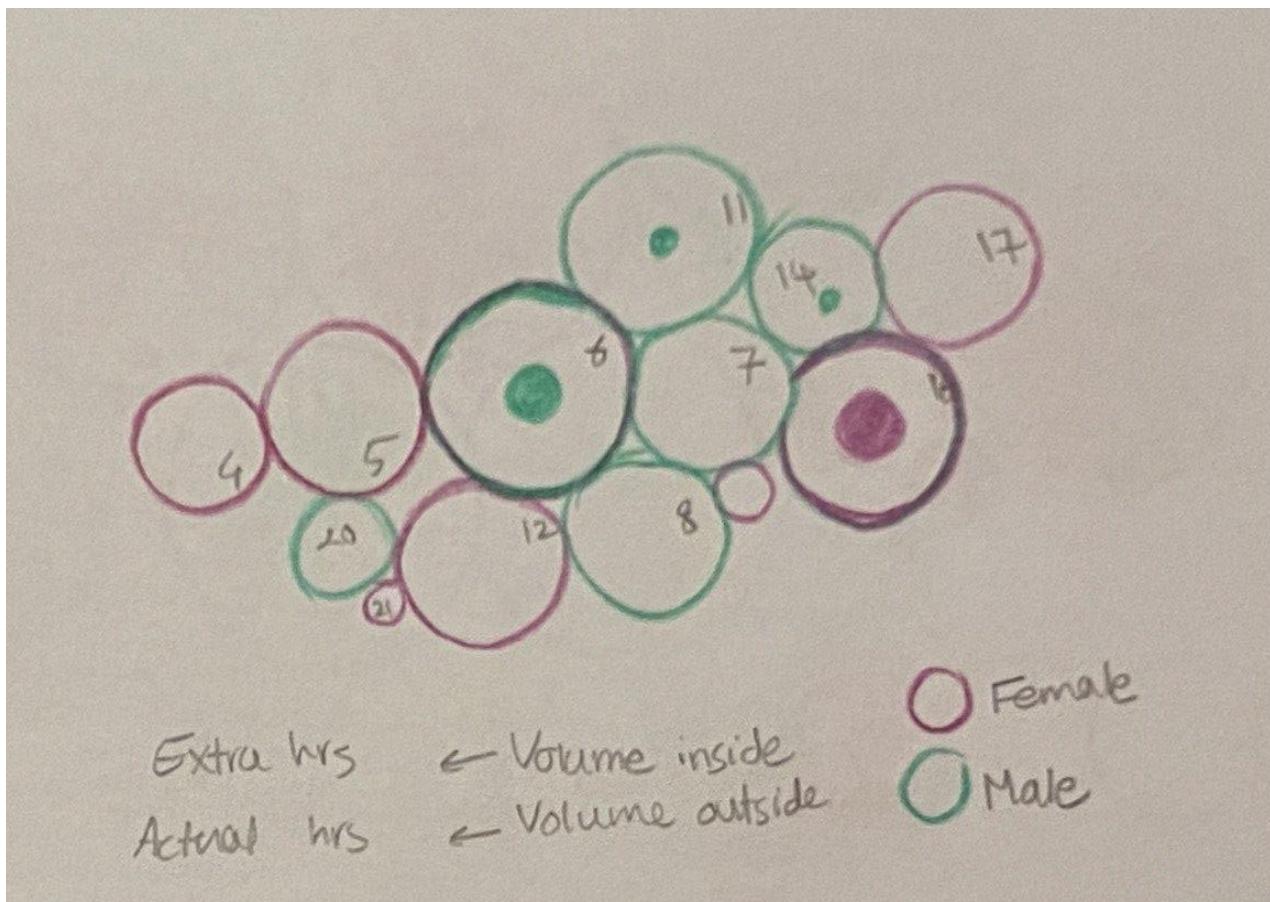
Sketch 3

In this visualization, each job category is a triangle. The size of a triangle shows the actual hours in each category. The parts in each triangle which is filled shows the extra hours in each category. The numbers in each triangle shows the category of jobs.



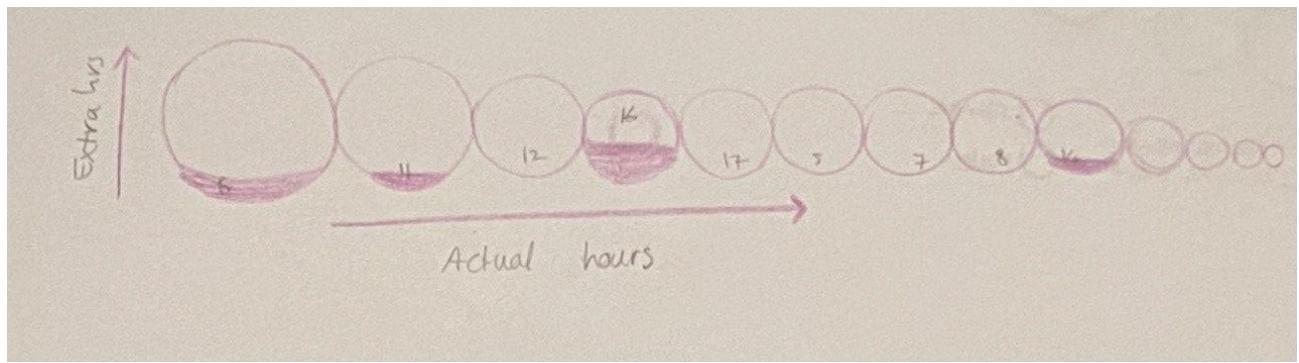
Sketch 4

In this chart each layer shows one category of jobs and the green color shows the actual hours and the red part shows the extra hours.



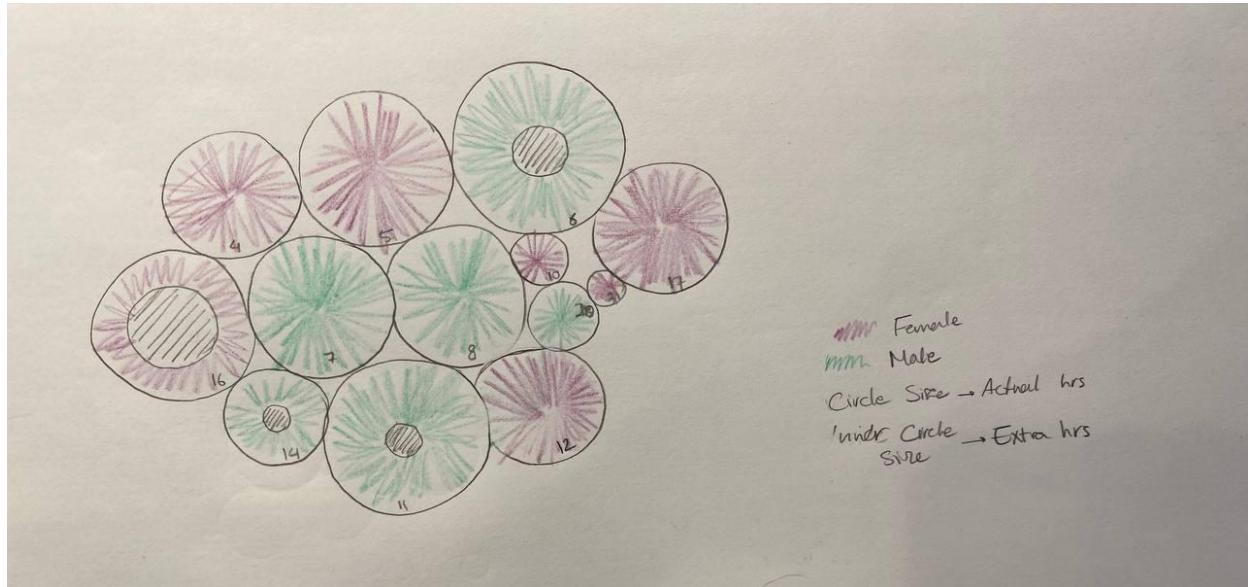
Sketch 5

In this chart, we decided to show the dominant gender in each category of jobs. Male being green and Female being red. Each circle shows the average of actual work hours in each category and the filled circles show the average of extra hours worked in each category. The numbers in each circle show the category of each job.



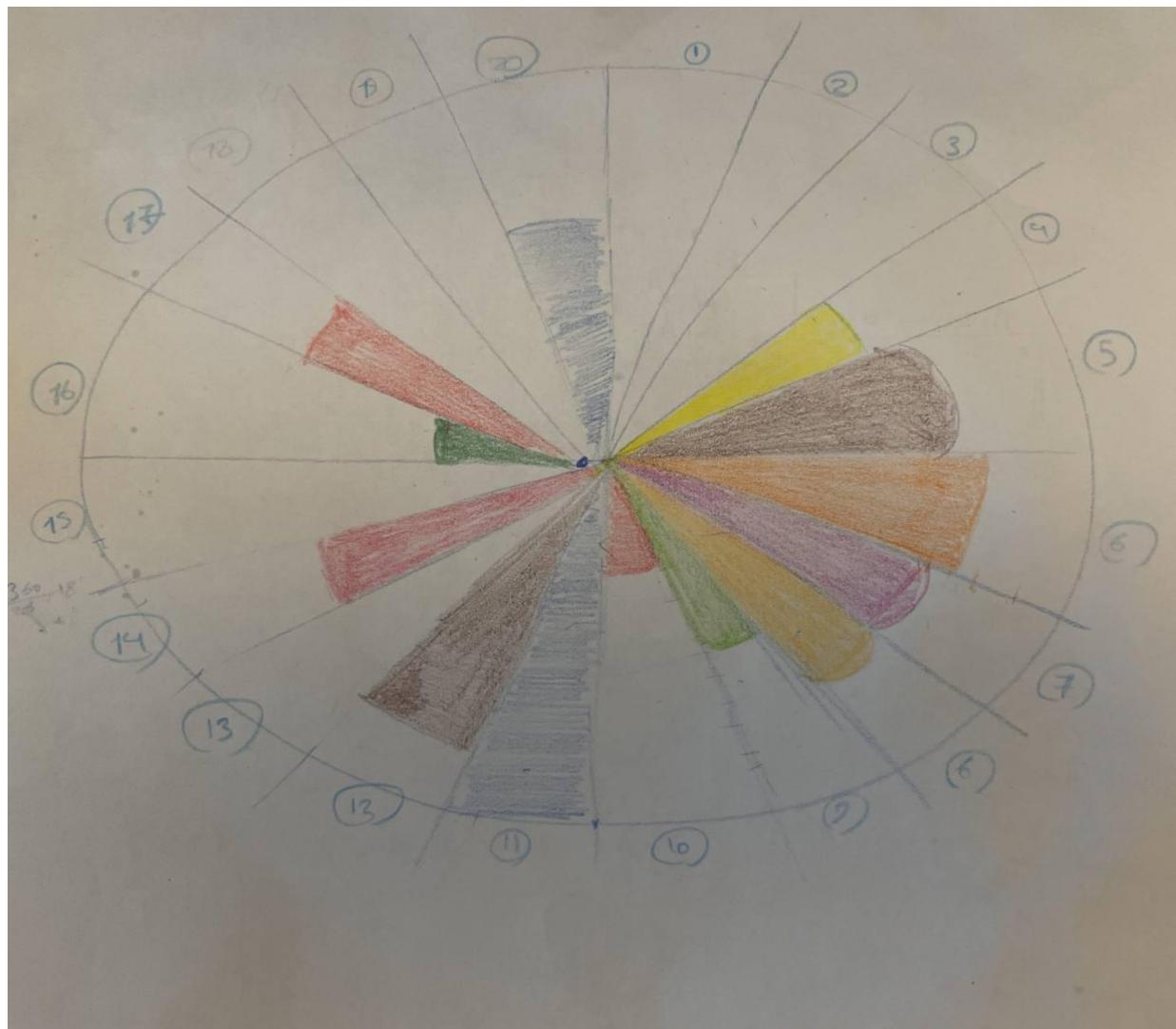
Sketch 6

In this sketch each job category is a circle they are organized based on the average of actual hours employers work in those job categories. The colored parts are showing the extra hours worked in each job category. The numbers shown in the circles indicate the job categories.



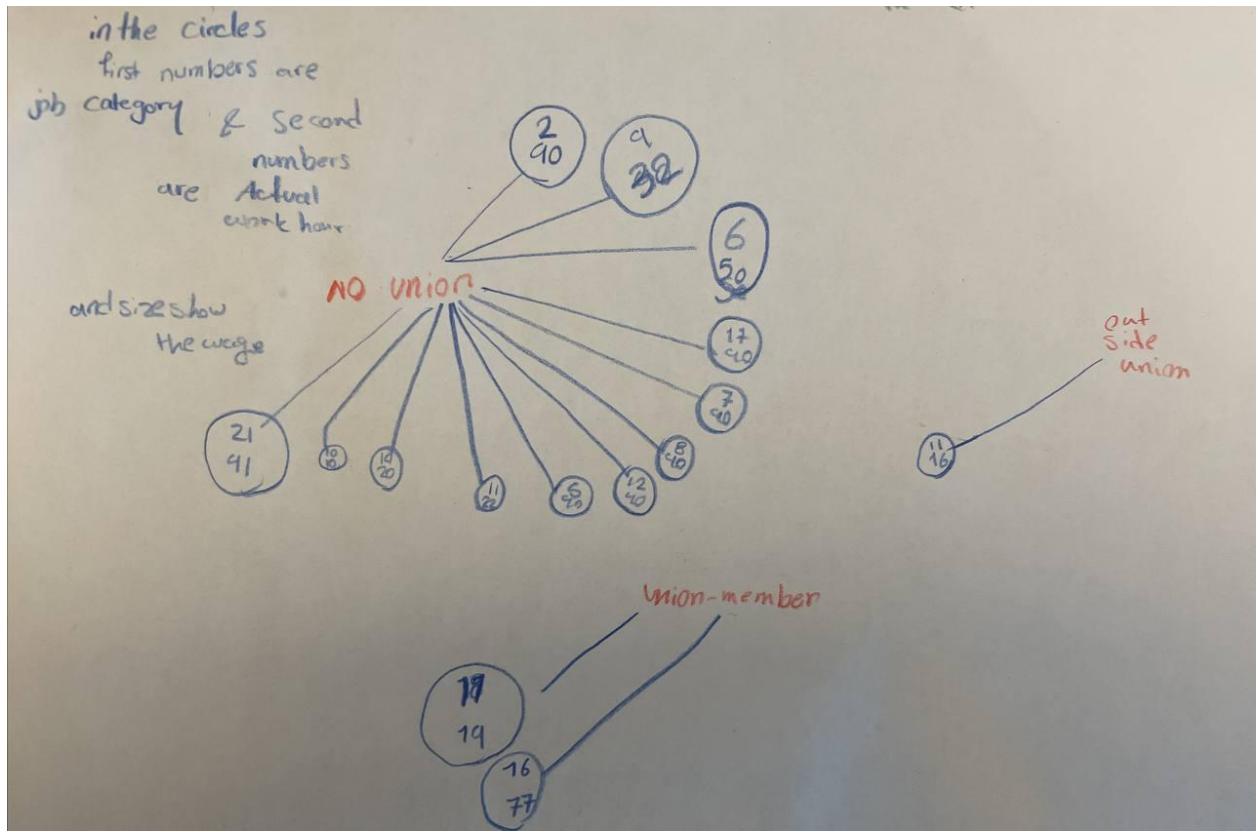
Sketch 7

The size of circles show the actual hours worked and the filled part is the extra hours. The numbers indicate the job categories. The colors are the dominant gender in each category.



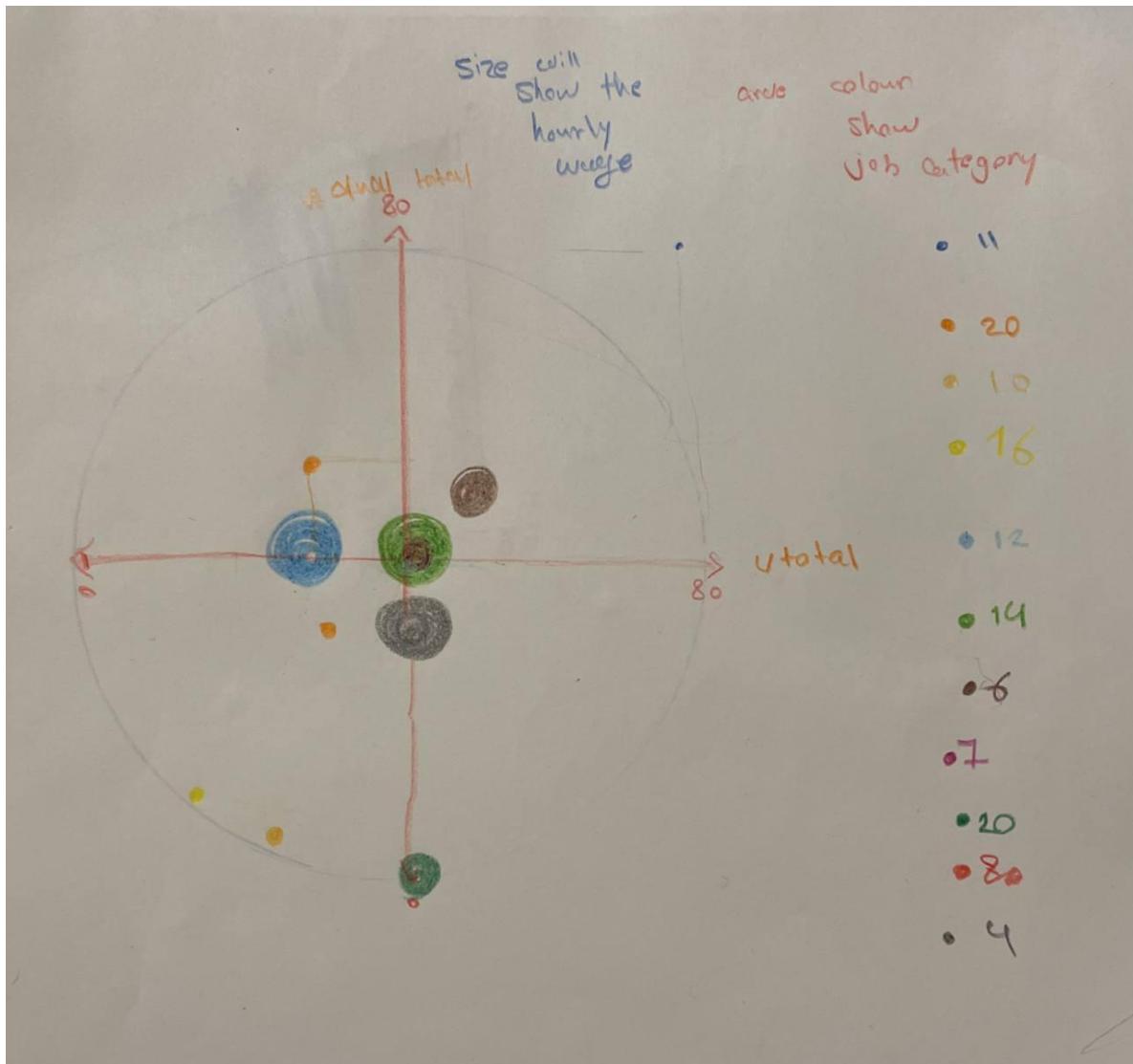
Sketch 8

Each number is showing one job category and the length of the coloured part in that section is the actual hour. The maximum colourd part is the radius of the circle.



Sketch 9

First separate the categories based on being in union or not or not having any union and then show the actual work hour and category inside circles and the size of circles are related to the wage.



Sketch 10

The axes are usually total and actual total working time and the middle point is 40 for each of them. The colour of circles are showing their category and the size is showing the wage. We can put the string version of categories instead of numbers.

Discussion of choices and directions

Among all the evaluated visualizations, we chose three views from our previous designing steps. From the 10+10 design process we end up with 3 views. The exact process of our 10+10 process is shown in this report and for two of the views. The final selected views are as follows:

- 1) A sunburst graph to visualize the relationship between wage category, sex, and education level. This type of graph shows hierarchies through a series of rings, which are fragmented based on category nodes. Each ring represents a level, with the central circle being the root node, and as we go outward, the depth increases. Sunburst graphs show how the outer rings relate to the inner rings and are most effective at showing how one ring is broken into its contributing pieces; so, they are ideal to visualize hierarchical structures in a single-screen snapshot.

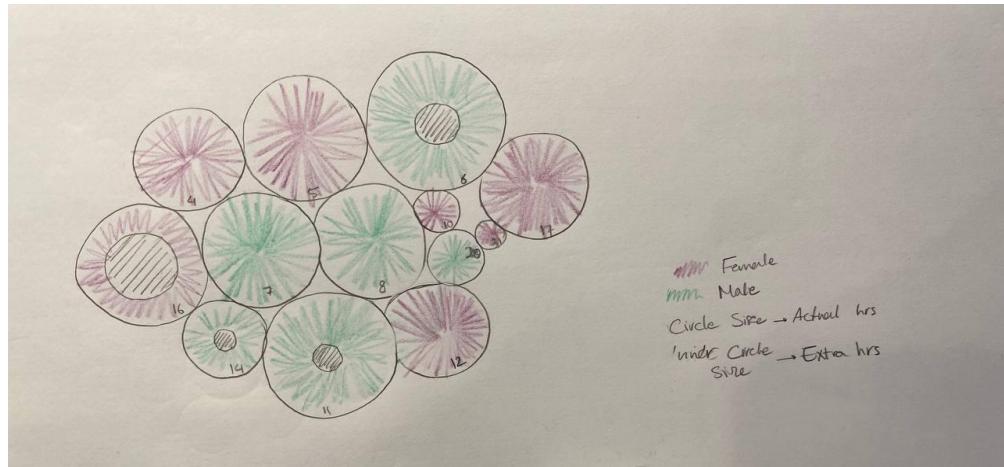
Interactive feature: when the user hovers the mouse over a segment, the sequence is highlighted and the percentage of samples matching with the sequence is shown.

- 2) The circular diagram which shows the categories of jobs (check appendix for details), the size of each circle shows average of actual hours employers worked in a month and the colored circles in each circle shows the average extra hours employers worked in a month in each category.

We chose this Bubble chart which allows us to visualize four different components at the same time (gender, categories, actual hrs and extra hrs). This visualization seems easy to read and it allows us to make critical associations between these components. It is not complex and at the same time it is showing a complex concept in our data subset. As we had two continuous variables and one discrete one to answer our question this visualization also let us save colour for pointing out other information that we used as our interactive feature.

In addition, By using a diameter proportional to the total hours worked and the extra hours worked, the user can process the visual very quickly, make a comparison among all job categories, and also can get a sense of the proportion of extra hours in each category. In addition, assigning the color based on dominance of gender to circles, makes it easier to reveal any association between gender and hours worked.

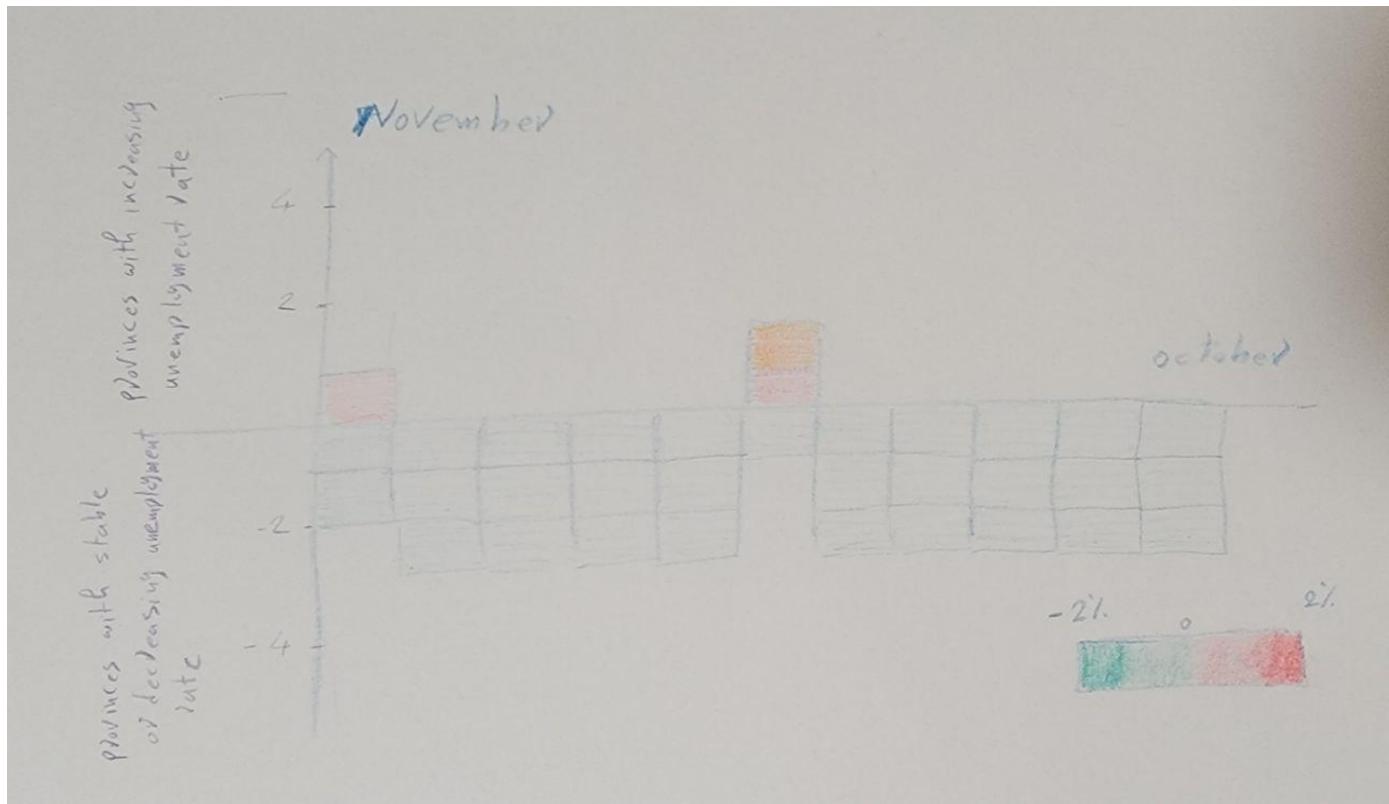
Interactive feature: The colors in each circle shows the dominant gender in each category of jobs, but not precisely. By hovering to each circle the viewer can see the exact distribution of women and men in the category of work as a pie chart.



- 3) Our third final view captures the progress of the unemployment rate for each province through time. Each province is represented by a box, and its location is specified by whether the corresponding change in employment rate is positive or negative. Provinces with increasing unemployment rate are located above the x-axis, and provinces with decreasing or stable unemployment rate are located under the x-axis. To enable the user to get a sense of relative increase or decrease, we have color coded the change in unemployment rate. Provinces with increasing unemployment rate are colored red, and provinces with decreasing unemployment rate are colored green, and the intensity is proportional to the absolute value of the change. This diagram is our chosen one because it shows the trend of employment and unemployment rate clearly. It is a good way of comparing each province according to the time of the year. The colors are a good way to show the average rates and viewers can see the trend easily.

Moreover, Similarity of this chart to the conventional bar chart creates a familiar look that the user may feel more comfortable with. Conjunction of the temporal representation, vertical positioning, and the color encoding, makes critical data available to users all at the same time, which reduces the demand on short term memory, and minimizes the cognitive load.

Interactive feature: Once the user clicks on one square (choose one province), all squares representing the same province will be highlighted.



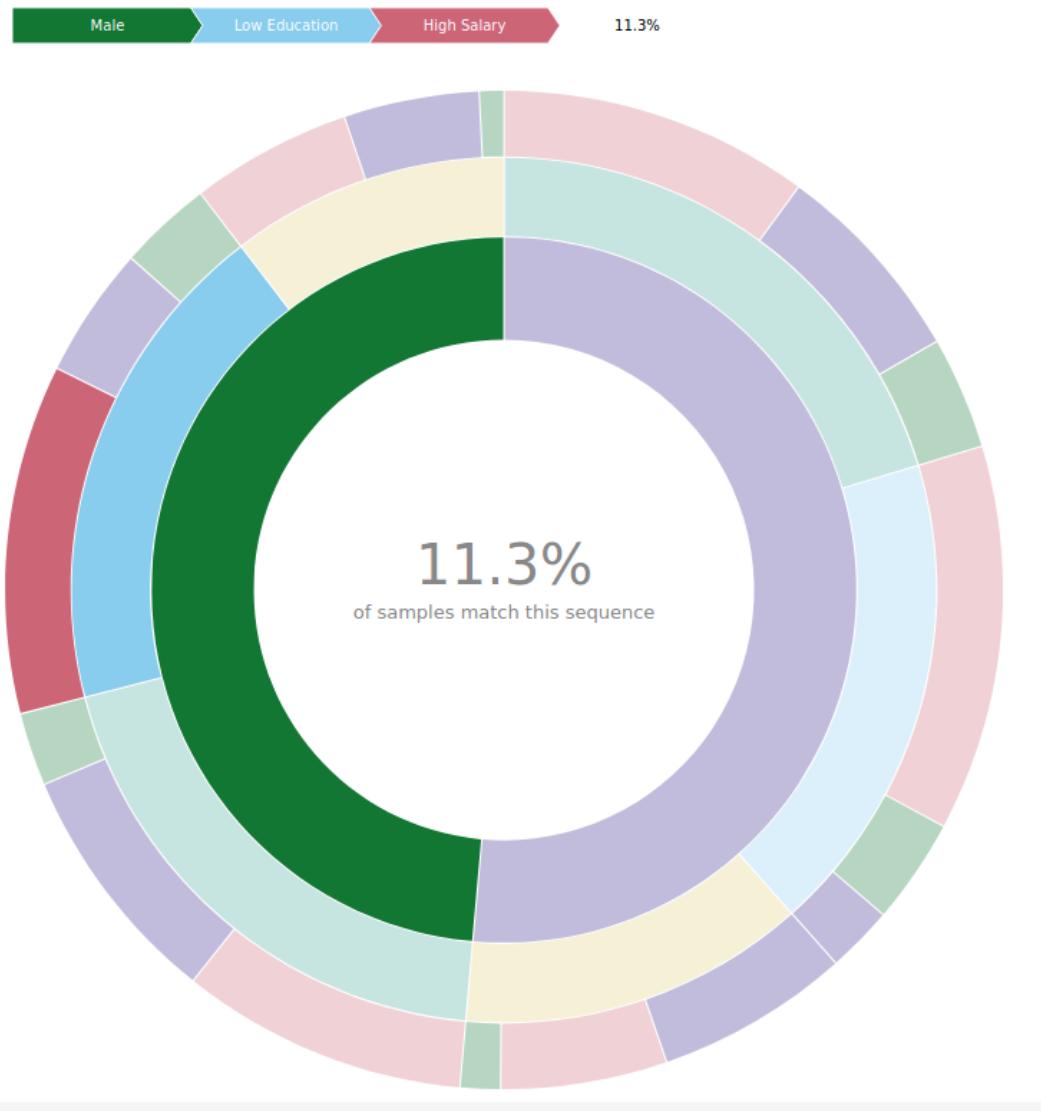
In designing all three views, we gave special attention to the selection of colours, so that people with colour blindness would have a lower chance of difficulty with using the visuals. In particular, we used a bright qualitative colour scheme proposed by Paul Tol. The colours in this schema are shown in the first column of the next figure. The columns on the left show how those colours are perceived for people with protanopia, deutanopia, or tritanopia colour blindness.

Color Palette



Implementation

We implemented the final views using the D3 javascript library. The following image shows a screenshot of the view. In this image, the sequence Male->Low Education->High salary is selected by hovering the mouse of the corresponding section, and the percentage of samples matching with the sequence is shown.



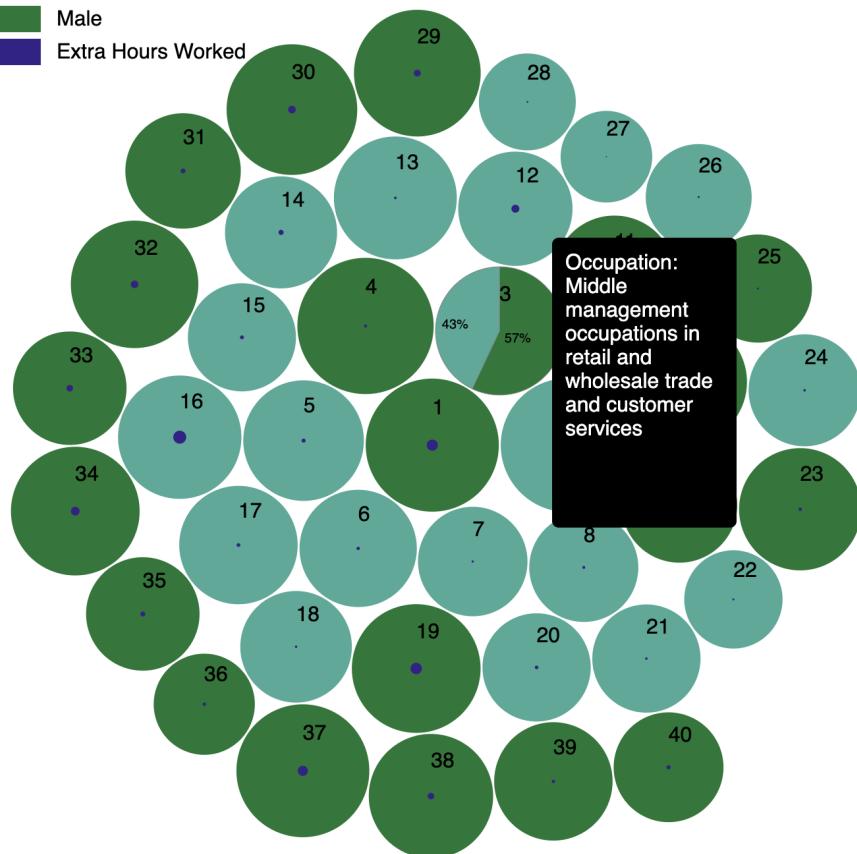
The following is a screenshot from the second view implemented using the D3 library, where the relationship between job categories, average actual hours worked, and average extra hours is portrayed.

Actual Hours and Male/Female distribution in Occupations.

Circles show the Actual Hours

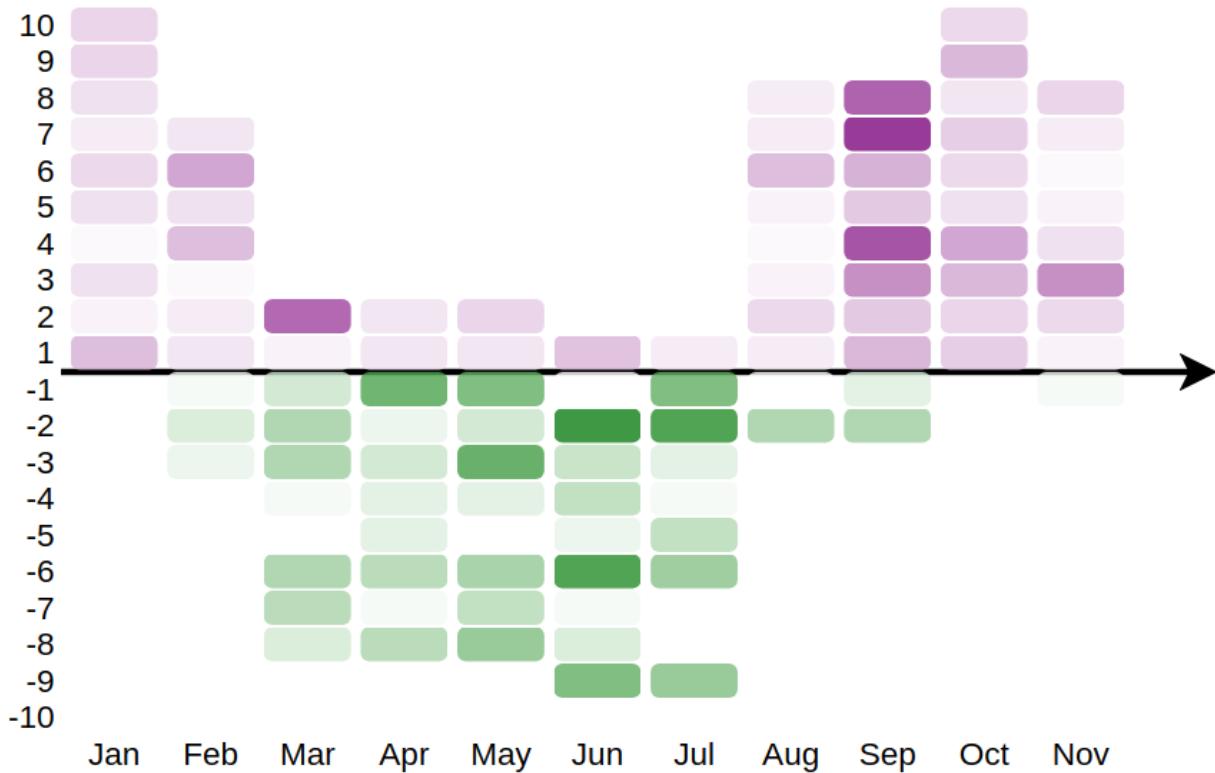
Numbers indicate job categories

- Female
- Male
- Extra Hours Worked



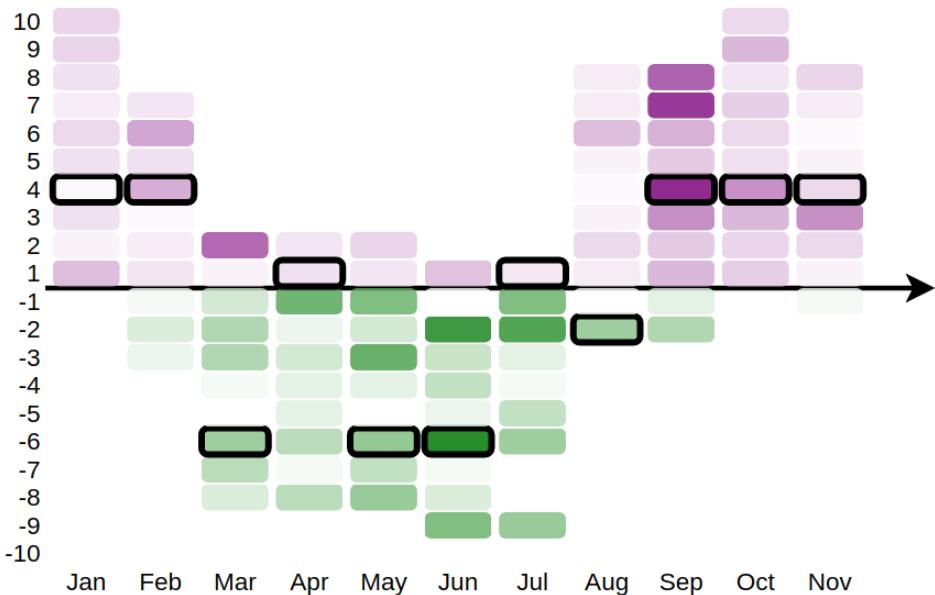
And the following graph shows a screenshot from the third view, where the temporal dynamic of the unemployment rate is highlighted.

Unemployment rate change in different provinces



And the following screenshot shows the interactive response of the graph when the mouse is hovered over one of the provinces. When a mouse is hovered over a province, all boxes corresponding to that province are highlighted in different time points. In addition, the exact value of the unemployment rate is and the selected province is shown in a tooltip below the graph.

Unemployment rate change in different provinces



The exact value of
this cell is: 0.005 and the Province is: Prince Edward Island

Insights learned from the data

Considering the portion of high salary in males and females with high education in the first view, we can see among highly educated people, males have a higher chance of having a high salary. And interestingly, the proportion of females with higher education is higher than men with higher education.

According to this bubble chart, some categories are mostly male dominated jobs such as “Maintenance and equipment operation trades”. Some higher rank positions like “senior management”, “middle management” are also male-dominated. But more health-related jobs and positions in most ranks are female-dominated. We can still see the traditional ranking and job stereotypes.

Based on this temporal graph, most provinces have a decreasing trend in employment rate from March to August, with New Brunswick having the highest decrease. For the rest of

the year, all provinces experience an increase in the unemployment rate, and interestingly, in September, New Brunswick is again at the top, however, this time at having the highest increase in the unemployment rate. This shows that this province suffers from an unstable economy. In addition, Alberta consistently has one of the lowest unemployment rates throughout the year, and on the other hand, Newfoundland has the highest unemployment rate.

Conclusion

In this project, we developed visual tools to analyze the Labor Force Survey dataset. Our project began by understanding the dataset and the challenges and preliminary statistical analysis to get a sense of the data, and proceed with extracting two subsets of the data for the purpose of data sketching. To decide on the most suitable views on the data, we used 10+10 ideation to create visuals that are informative while being easily accessible to the user, and we finally selected three views as the final set. Finally, we implemented the views using D3 javascript library, and the implementations revealed very interesting insights on the data. We learned that salary versus education is different for males and females. We also noticed a gender imbalance in high rank job, with men being the dominant gender. Furthermore, we got insights over the dynamic of unemployment rate for different provinces; in particular, we realized Alberta has one of the lowest unemployment rates, and Newfoundland suffers from the highest unemployment rate.

Appendix

The details of each column name and category numbers shown in each sketch ar as follows:

REC_NUM - Order of record in file

SURVYEAR - Survey year

SURVMNTH - Survey month

LFSSTAT - Labour force status

PROV - Province

AGE_12 - Five-year age group of respondent

SEX - Sex of respondent

MARSTAT - Marital status of respondent

EDUC - Highest educational attainment

IMMIG - Immigration status

NAICS_21 - Industry of main job

NOC_40 - Occupation at main job

UHRSMAN - Usual hours worked per week at main job

AHRSMAN - Actual hours worked per week at main job

HRSAWAY - Hours away from work, part-week absence only

YAWAY - Reason for part-week absence

PAIDOT - Paid overtime hours in reference week

UNPAIDOT - Unpaid overtime hours in reference week

XTRAHRS - Number of overtime or extra hours worked

HRLYEARN - Usual hourly wages, employees only

UNION - Union status, employees only

EFAMTYPE - Type of economic family

AGYOWNK - Age of youngest child

XTRAHRS - [AHRSMAN - UHRSMAN]

SURVMNTH - Survey month

Value	Label
8.	August
11.	November
5.	May
1.	January
3.	March
7.	July
6.	June
4.	April
12.	December
9.	September
10.	October
2.	February

LFSSTAT - Labour force status

Value	Label
2.	Employed, absent from work
3.	Unemployed
4.	Not in labour force
1.	Employed, at work

PROV - Province

Value	Label
48.	Alberta
10.	Newfoundland and Labrador
47.	Saskatchewan
12.	Nova Scotia
13.	New Brunswick
46.	Manitoba
11.	Prince Edward Island
35.	Ontario
24.	Quebec
59.	British Columbia

AGE_12 - Five-year age group of respondent

Value	Label
5.	35 to 39 years
7.	45 to 49 years
1.	15 to 19 years
9.	55 to 59 years
11.	65 to 69 years
12.	70 and over
6.	40 to 44 years
2.	20 to 24 years
8.	50 to 54 years
3.	25 to 29 years
10.	60 to 64 years
4.	30 to 34 years

SEX - Sex of respondent

Value	Label
1.	Male
2.	Female

MARSTAT - Marital status of respondent

Value	Label
6.	Single, never married
2.	Living in common-law
5.	Divorced
3.	Widowed
4.	Separated
1.	Married

EDUC - Highest educational attainment

Value	Label
1.	Some high school
3.	Some postsecondary
4.	Postsecondary certificate or diploma
6.	Above bachelor's degree
2.	High school graduate
0.	0 to 8 years
5.	Bachelor's degree

IMMIG - Immigration status

Value	Label
3.	Non-immigrant
2.	Immigrant, landed more than 10 years earlier
1.	Immigrant, landed 10 or less years earlier

NAICS_21 - Industry of main job

Value	Label
8.	Manufacturing - non-durable goods
10.	Retail trade
17.	Health care and social assistance
4.	Mining, quarrying, and oil and gas extraction
11.	Transportation and warehousing
14.	Professional, scientific and technical services
7.	Manufacturing - durable goods
1.	Agriculture
6.	Construction
2.	Forestry and logging and support activities for forestry
3.	Fishing, hunting and trapping
21.	Public administration
20.	Other services (except public administration)
5.	Utilities
18.	Information, culture and recreation
13.	Real estate and rental and leasing
15.	Business, building and other support services
16.	Educational services
9.	Wholesale trade
12.	Finance and insurance
19.	Accommodation and food services

NOC 40 - Occupation at main job

Value	Label
38.	Processing and manufacturing machine operators and related production workers
19.	Occupations in front-line public protection services
22.	Technical occupations in art, culture, recreation and sport
14.	Technical occupations in health
9.	Distribution, tracking and scheduling co-ordination occupations
12.	Professional occupations in nursing
10.	Professional occupations in natural and applied sciences
4.	Middle management occupations in trades, transportation, production and utilities
31.	Other installers, repairers and servicers and material handlers
23.	Retail sales supervisors and specialized sales occupations
8.	Office support occupations
1.	Senior management occupation
21.	Professional occupations in art and culture
39.	Assemblers in manufacturing
11.	Technical occupations related to natural and applied sciences
13.	Professional occupations in health (except nursing)
18.	Paraprofessional occupations in legal, social, community and education services
37.	Processing, manufacturing and utilities supervisors and central control operators
30.	Maintenance and equipment operation trades
33.	Trades helpers, construction labourers and related occupations
25.	Sales representatives and salespersons - wholesale and retail trade
5.	Professional occupations in business and finance
36.	Harvesting, landscaping and natural resources labourers
20.	Care providers and educational, legal and public protection support occupations
16.	Professional occupations in education services
2.	Specialized middle management occupations
26.	Service representatives and other customer and personal services occupations
34.	Supervisors and technical occupations in natural resources, agriculture and related production
15.	Assisting occupations in support of health services
27.	Sales support occupations
24.	Service supervisors and specialized service occupations
3.	Middle management occupations in retail and wholesale trade and customer services
6.	Administrative and financial supervisors and administrative occupations
29.	Industrial, electrical and construction trades
28.	Service support and other service occupations, n.e.c.
32.	Transport and heavy equipment operation and related maintenance occupations
17.	Professional occupations in law and social, community and government services
40.	Labourers in processing, manufacturing and utilities
7.	Finance, insurance and related business administrative occupations
35.	Workers in natural resources, agriculture and related production

UNION - Union status, employees only

Value	Label
3.	Non-unionized
2.	Not a member but covered by a union contract or collective agreement
1.	Union member

EFAMTYPE - Type of economic family

Value	Label
1.	Person not in an economic family
2.	Dual-earner couple, no children or none under 25
10.	Single-earner couple, female employed, youngest child 18 to 24
12.	Non-earner couple, youngest child 0 to 17
7.	Single-earner couple, male employed, youngest child 18 to 24
17.	Lone-parent family, parent not employed, youngest child 18 to 24
5.	Single-earner couple, male employed, no children or none under 25
13.	Non-earner couple, youngest child 18 to 24
14.	Lone-parent family, parent employed, youngest child 0 to 17
4.	Dual-earner couple, youngest child 18 to 24
3.	Dual-earner couple, youngest child 0 to 17
11.	Non-earner couple, no children or none under 25
18.	Other families
9.	Husband-wife, single earner couple, wife employed, youngest child 0 to 17
6.	Single-earner couple, male employed, youngest child 0 to 17
8.	Single-earner couple, female employed, no children or none under 25
15.	Lone-parent family, parent employed, youngest child 18 to 24
16.	Lone-parent family, parent not employed, youngest child 0 to 17

AGYOWNK - Age of youngest child

Value	Label
1.	Youngest child less than 6 years
2.	Youngest child 6 to 12 years
3.	Youngest child 13 to 17 years
4.	Youngest child 18 to 24 years