CSCM27

Assignment One (A1): Information Visualisation

2031545

Maciej Legas

1 DATASET DESCRIPTION

Name of the chosen dataset: Contraceptive Method Choice Data Set.

This dataset contains 1473 instances and 10 attributes, where the latter are as follows:

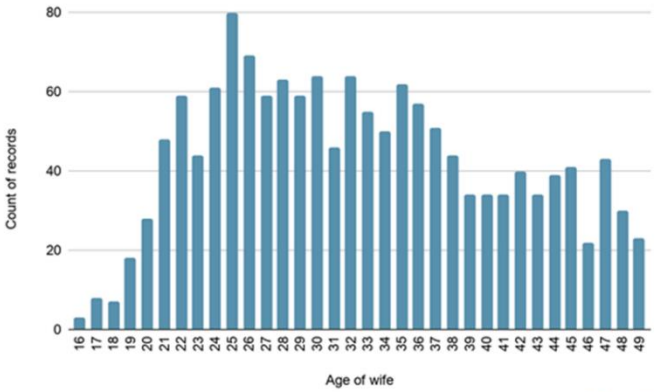
Attribute No.	Name	Value range limit, if applicable	Explanation of values, if applicable	Classification
1	Wife's age	–	–	Continuous
2	Wife's education	1, 2, 3, 4	1 – Low … 4 – High	Ordered
3	Husband's education	1, 2, 3 ,4	1 – Low … 4 – High	Ordered
4	Number of children ever born	–	–	Continuous
5	Wife's religion	0, 1	0 – Non-Islam 1 – Islam	Boolean
6	Wife's now working?	0, 1	0 – Yes 1 – No	Boolean
7	Husband's occupation	1, 2, 3, 4	1 – Low … 4 – High	Ordered
8	Standard-of-living index	1, 2, 3, 4	1 – Low … 4 – High	Ordered
9	Media exposure	1, 2, 3, 4	0 – Good 1 – Not good	Boolean
10	Contraceptive method used	1, 2, 3	1 – No use 2 – Long-term 3 – Short-term	Categorical

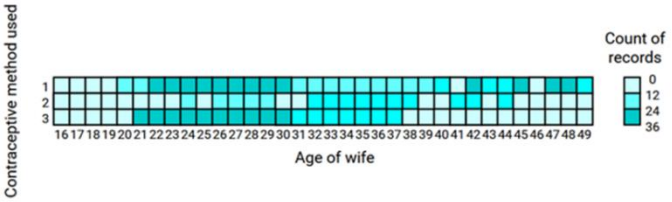
The user using this data set should be able to predict the contraceptive method used by a married woman from Indonesia, being given her social, demographic and economic circumstances.

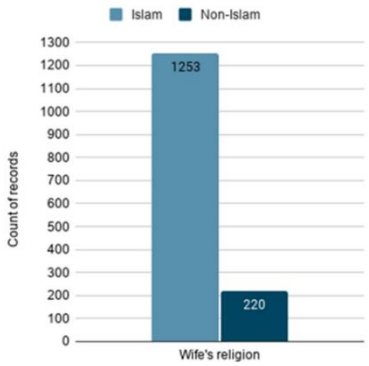
Data set provider: https://archive.ics.uci.edu/ml/datasets/Contraceptive+Method+Choice

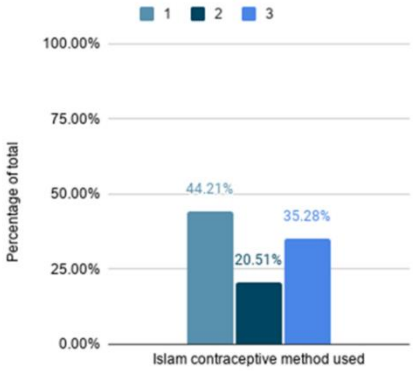
2 DESIGNS

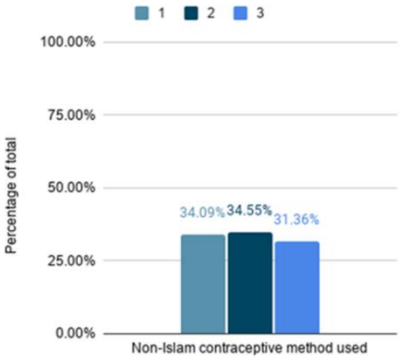
2.1 DESIGN 1

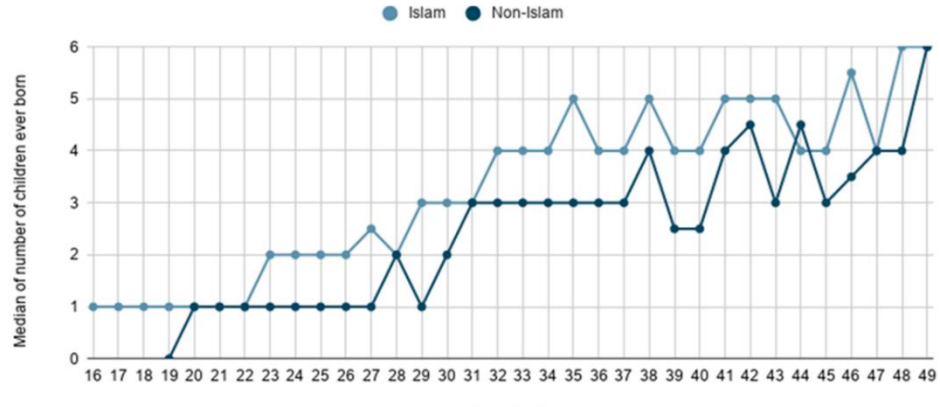

Count of records
Age of wife

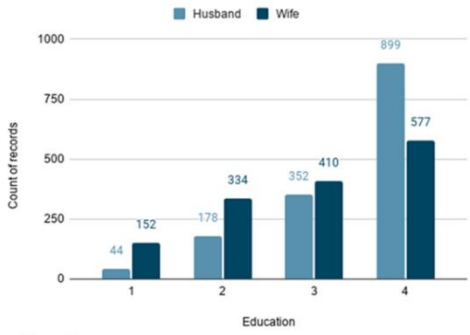

Contraceptive method used
Count of records
Age of wife

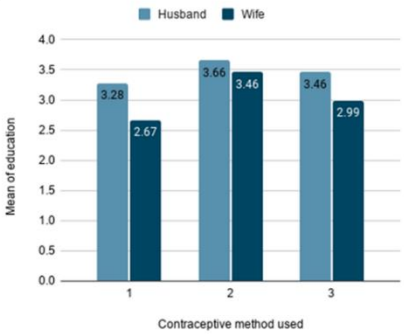

Islam
Non-Islam
Count of records
Wife's religion

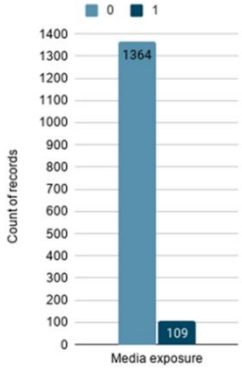

Percentage of total
Islam contraceptive method used

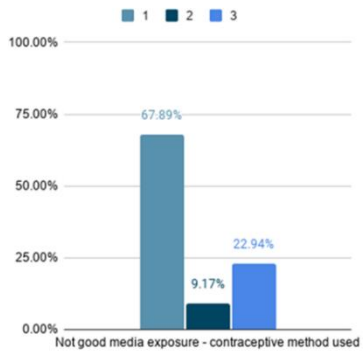

Percentage of total
Non-Islam contraceptive method used

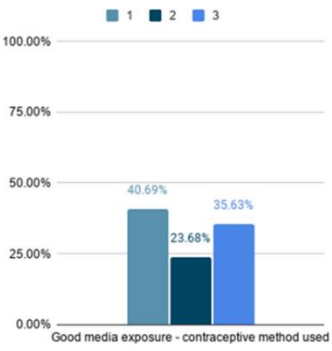

Islam
Non-Islam
Median of number of children ever born
Age of wife

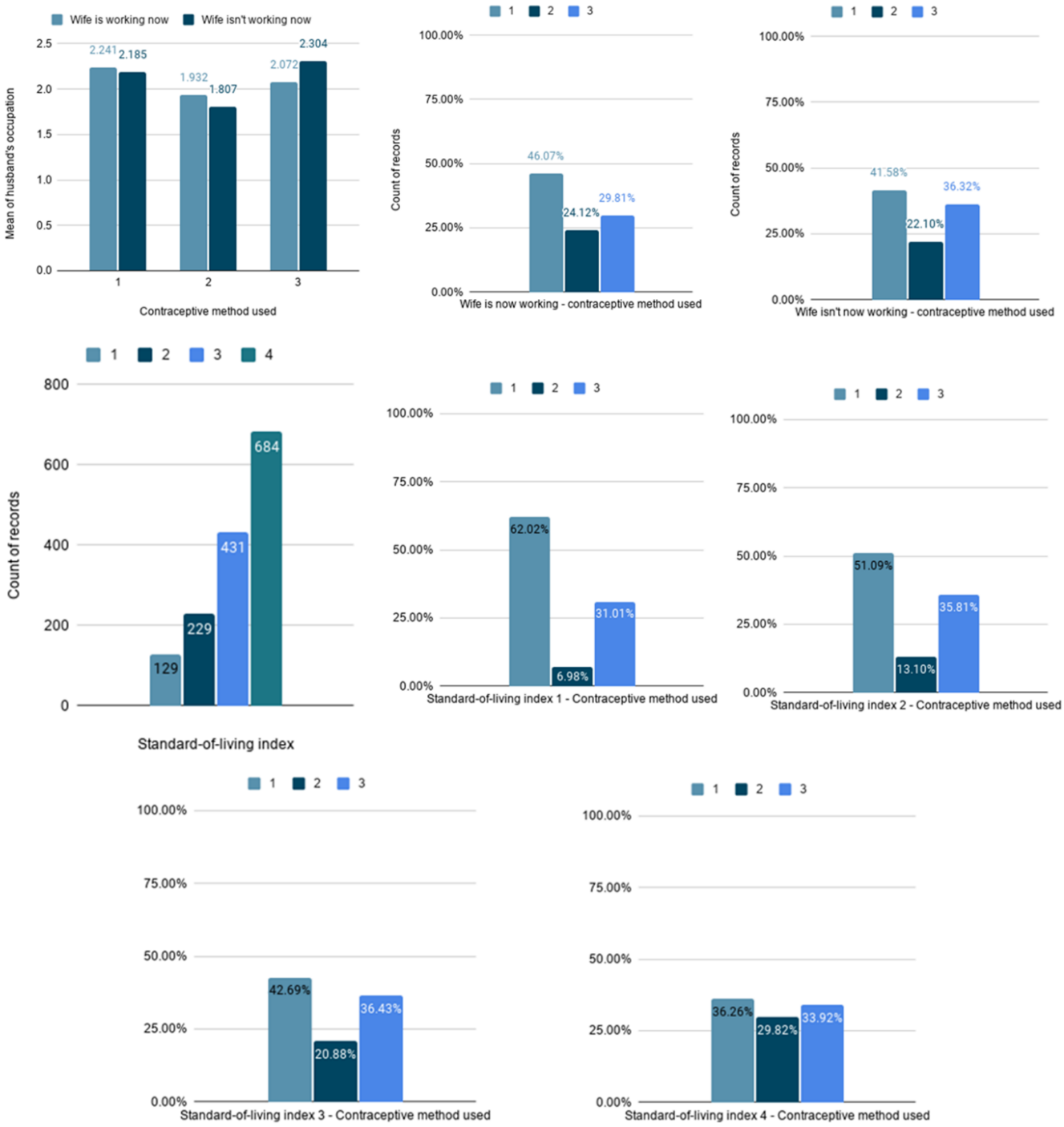

Husband
Wife
Count of records
Education


Husband
Wife
Mean of education
Contraceptive method used


Count of records
Media exposure


Not good media exposure - contraceptive method used


Good media exposure - contraceptive method used



This design has been created with an implicit split into categories: Demography, religion, education, media exposure, employment and standard of living, in order to allow the user to have a quick transition from an overview to a detailed comparison of the data.

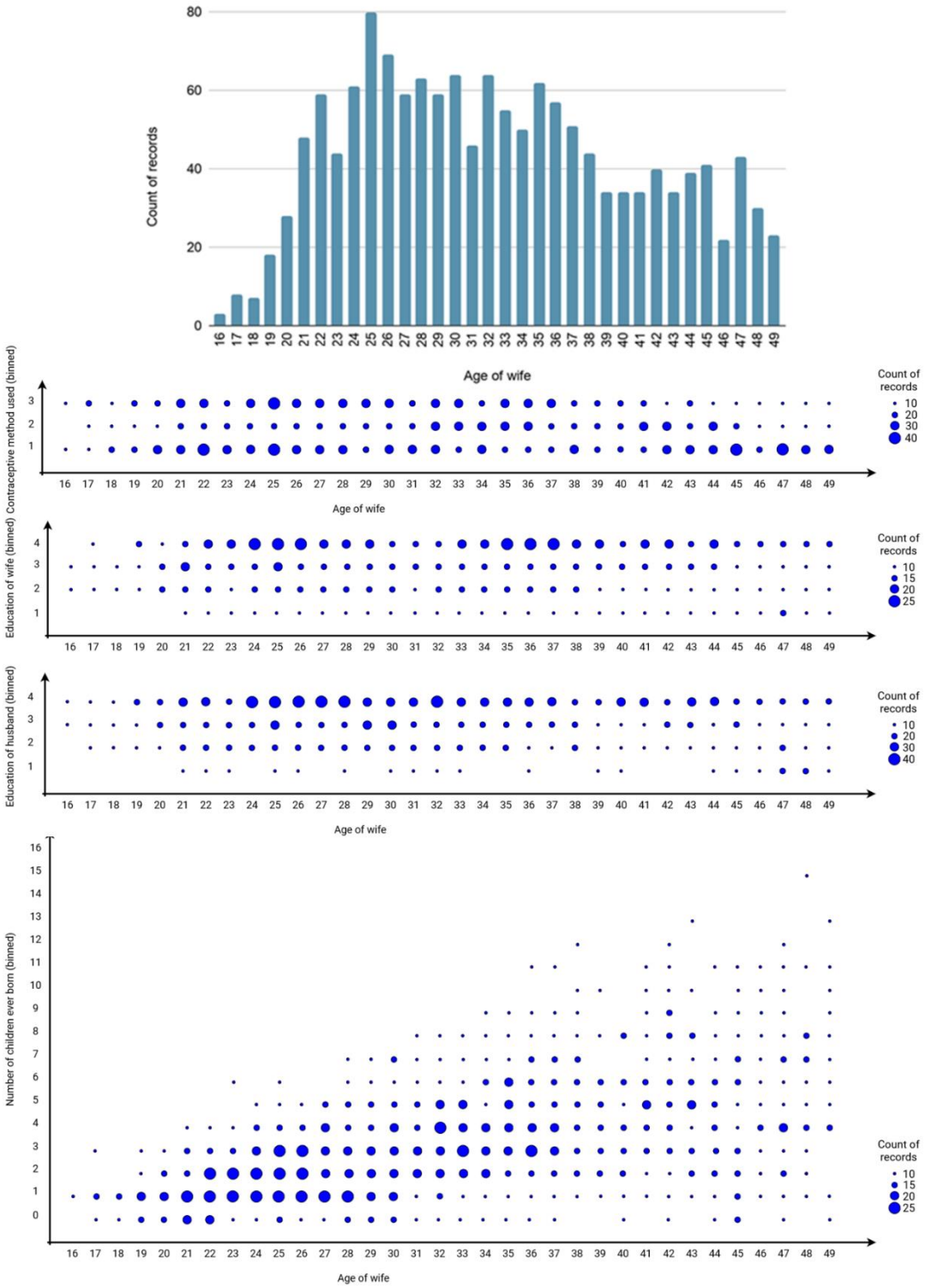
As this data set contains quite a bit of categorical and Boolean data, but a low number of categories instead, a decision has been made to base most of the charts as bar charts, as perceiving length and position channels is easy for an user [1], allowing him to focus mostly on the comparison between charts.

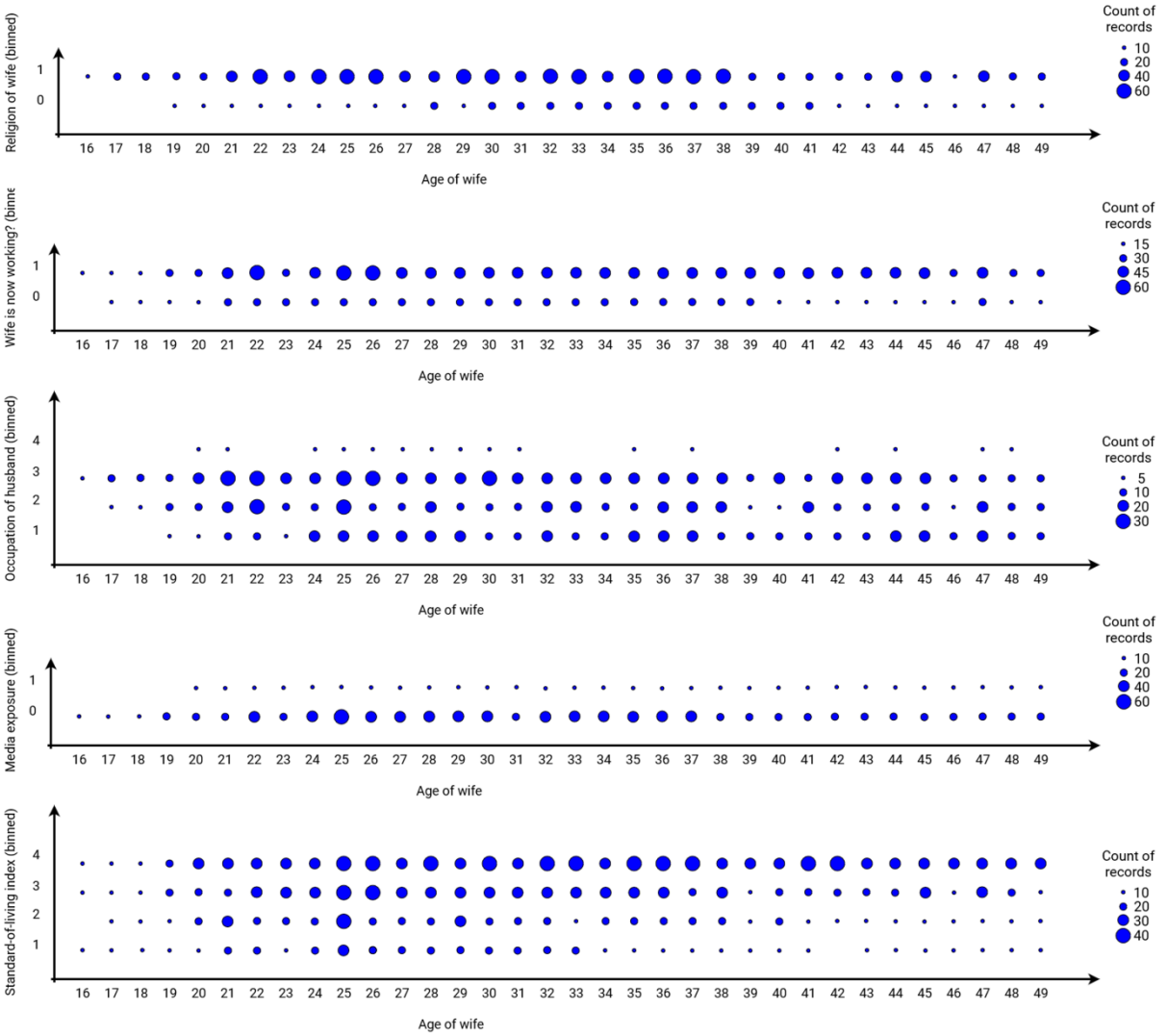
The comparison of data in this design is based on the use of a view composition, having similar charts grouped together. As a typical user usually has trouble to use the recent memory of a chart for comparison [2], related charts will be stacked side-by-side for easy comparison.

A heatmap is used to showcase the distribution of used contraceptive methods per age of wife. While it is known that hue and saturation are ranked worse than area and position [3], as this chart is being one of the most important charts in this design a heatmap has been chosen to easily contrast with the rest of the charts, especially with the line chart which could be not the case with a binned point chart.

Since this data set is mostly not uniformly distributed, percentage bar charts are used to help with prediction alongside a count of records.

2.2 DESIGN 2





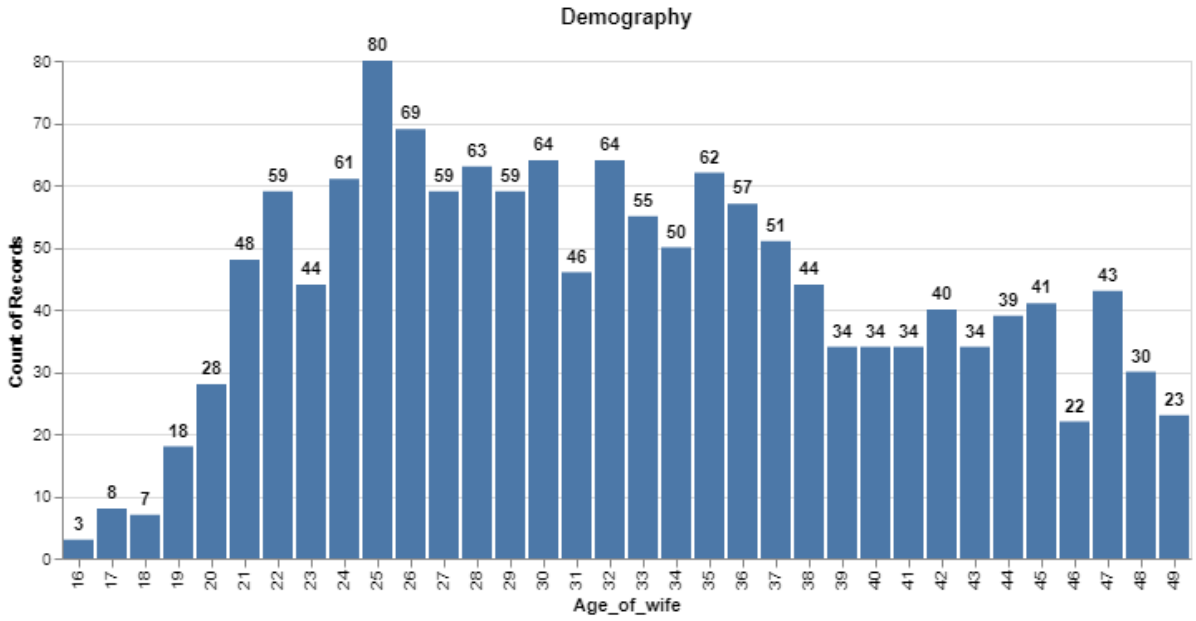
This design focuses on plotting each of the attributes from the dataset in sequential order against the age to find general trends in the female Indonesian part of the society.

While it does not go into a detailed approach for each category, this design is better for chunk memory and possibly for quick prediction, containing just one histogram and one style of binned point chart being plot against one constant axis. A user can scroll easily throughout the design and compare each chart with another.

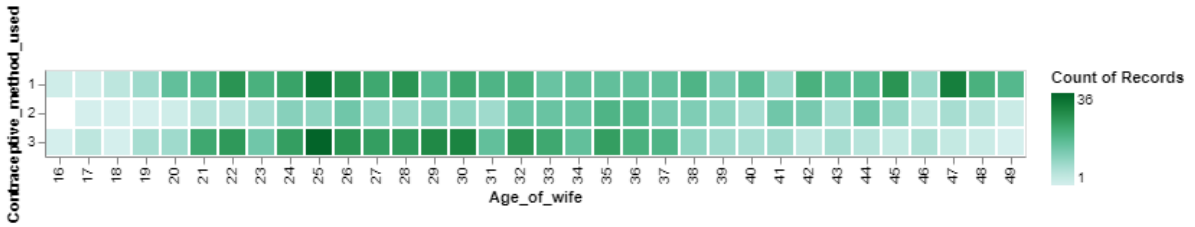
Heatmaps could be used instead of binned point charts, but binned point charts use area and position, while heatmaps only base on hue and saturation, which is ranked worse for visualisation [3].

3 DISCOVERED FEATURES AND IMPLEMENTATION

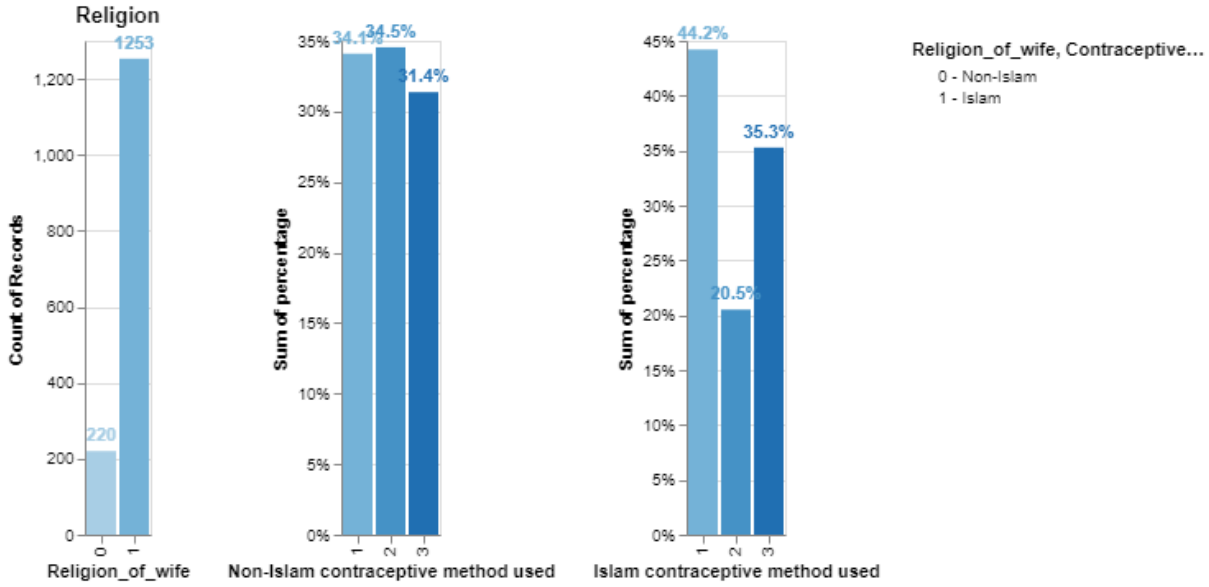
The first design has been chosen for implementation.



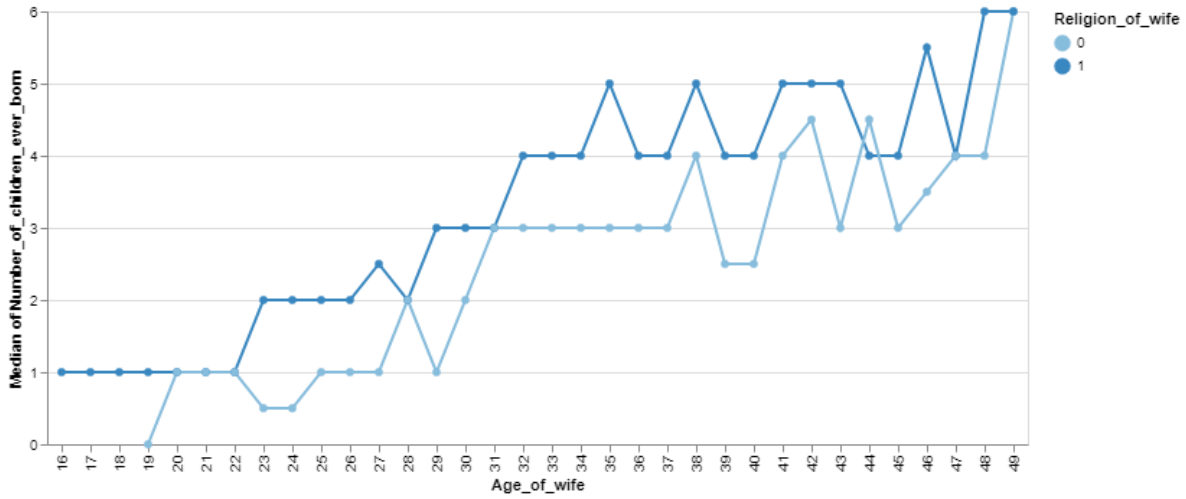
The dataset is positively skewed, barely containing any samples for the younger population (16-18), which therefore may not be accurately represented by any data analysis of this dataset.



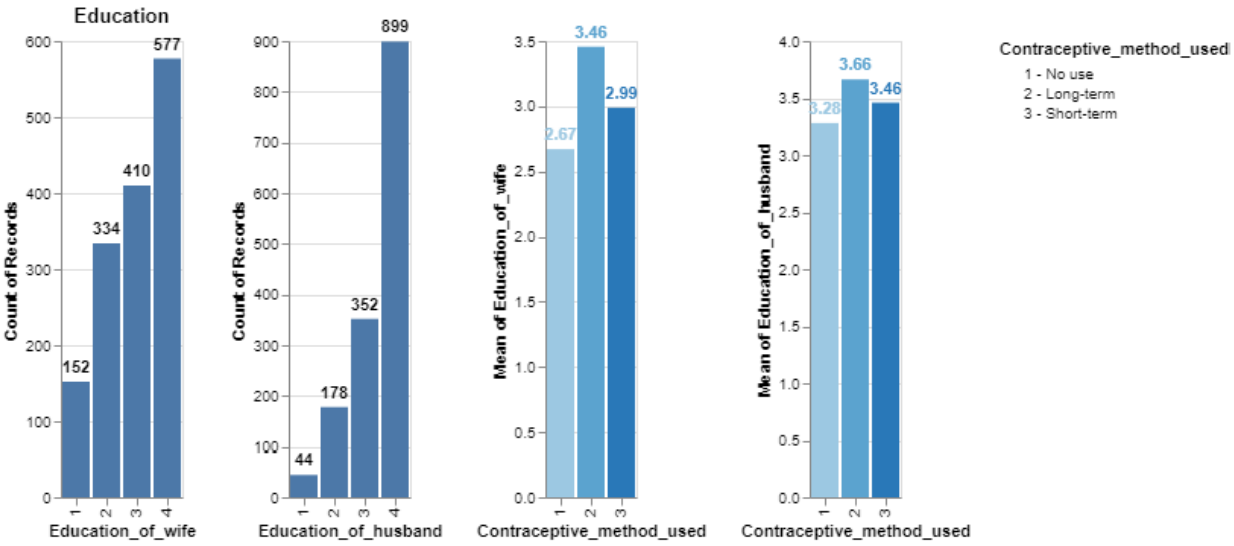
The heatmap above shows that Indonesian wives mostly either use no contraceptives at all or just short-term ones, with the long-term methods having most use by women between 32 and 37 years of age. Furthermore, the data suggests that the older a married woman is, the more likely she is to not use any contraceptives at all.



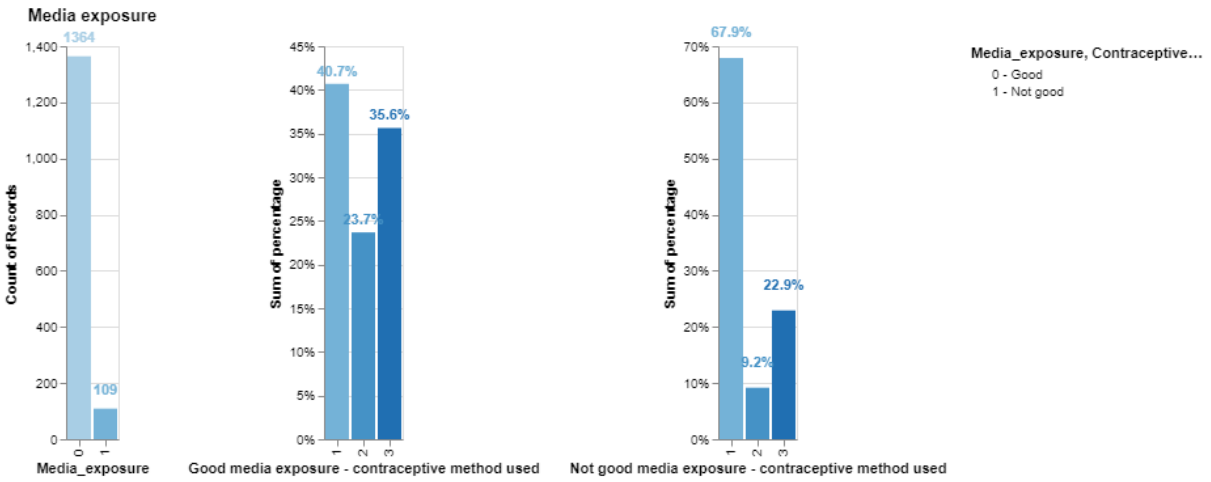
Most of the Indonesian women in this dataset are Muslim. Comparing the non-Islam chart, which has almost an equal division between all contraceptive methods, and the Islam chart, there is a visible trend for Muslim wives to not use any contraception at all or use short-term methods instead of long-term ones.



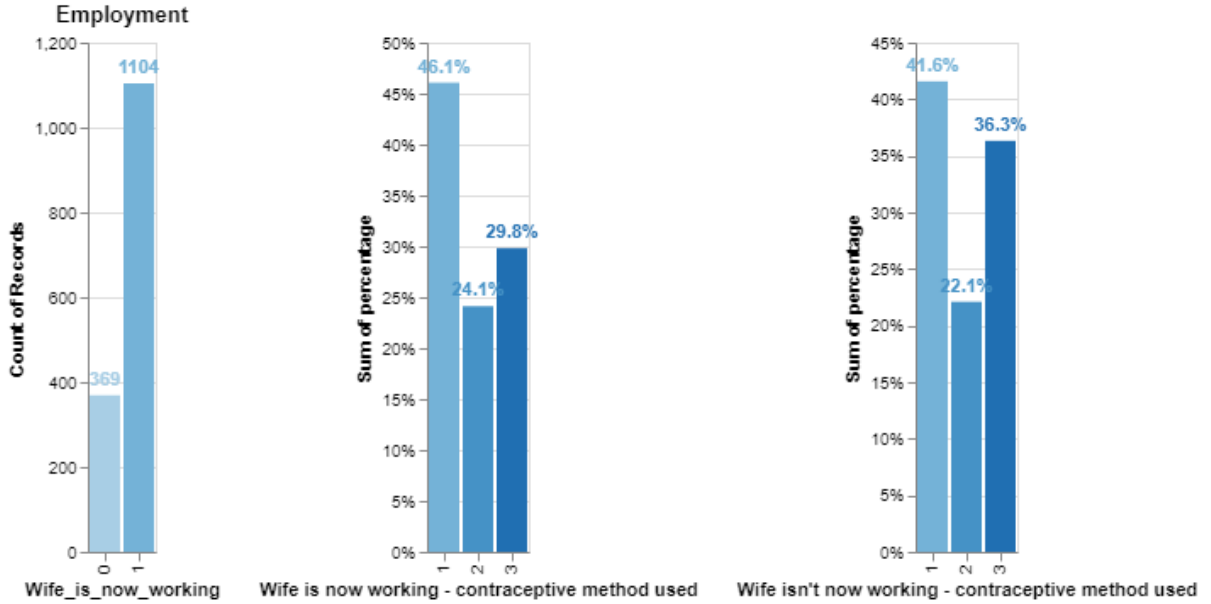
Besides any potential health reasons, the majority of older women not using any contraception at all is possibly related with the number of children ever born. The line chart above shows that per median, a Muslim family has at least one child more than a non-Muslim one. Compared with the percentage bar chart of contraceptive methods used by Muslim wives, and the heatmap with contraceptive methods used per age, it shows that the older a wife is, and the more children she has, the lesser chance of her using any contraceptive is.



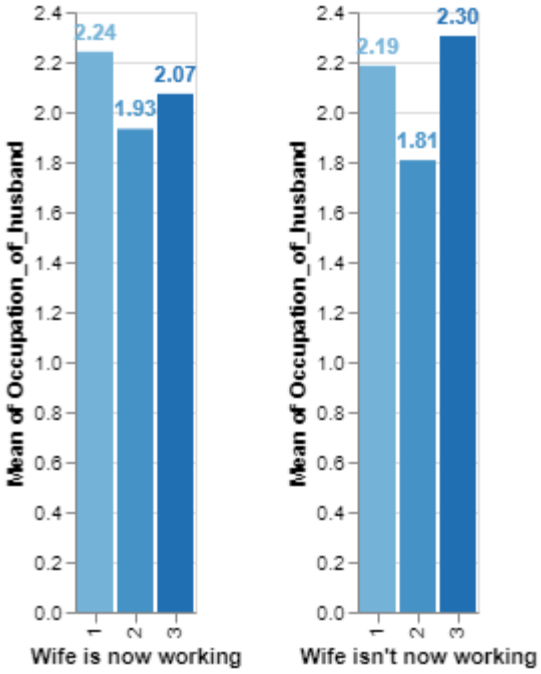
The resulting graphs of means for both husbands and wives predict that the lower the education is, the wife is most likely not using any contraceptives at all, and the higher the education of both the wife and the husband, the more possible that the wife is using a long-term contraceptive method, with the short-term methods being in between.



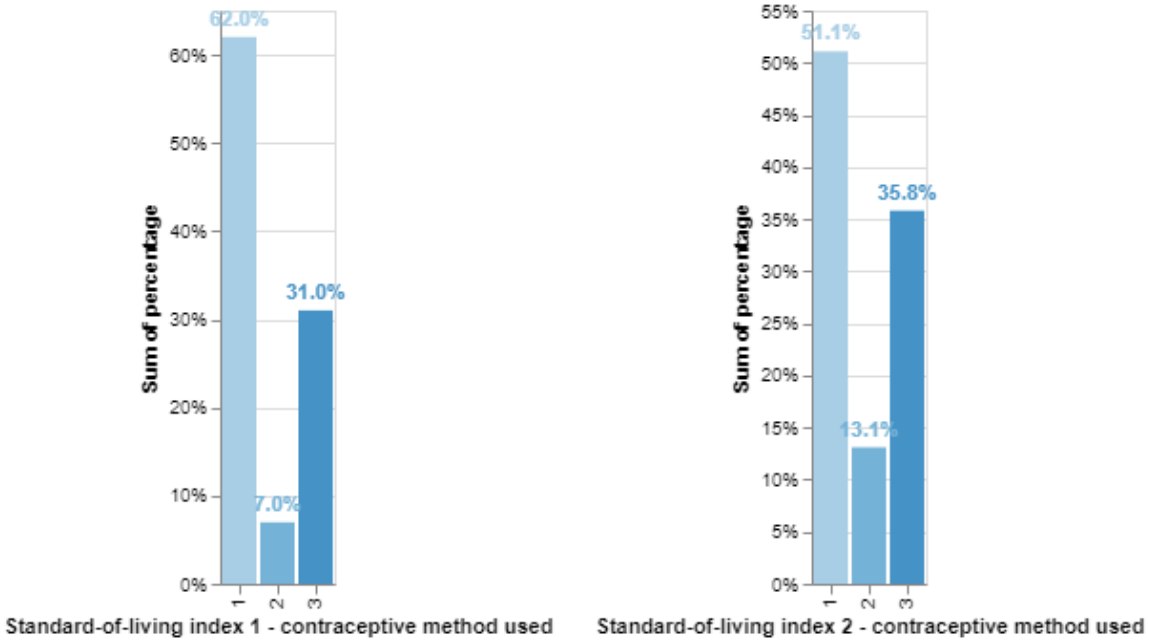
Women that are not well exposed to media use no contraception at all more often than wives that are, showing that media are a critical role factor in expanding a woman's awareness about possible contraceptives.

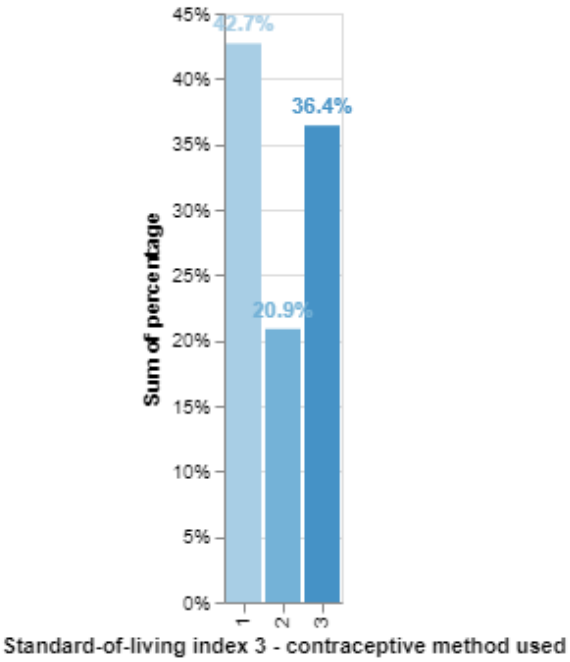


Wives that are not working tend to use more often short-term contraceptives than wives that are working, which have a slightly higher chance of not using any contraceptives at all.

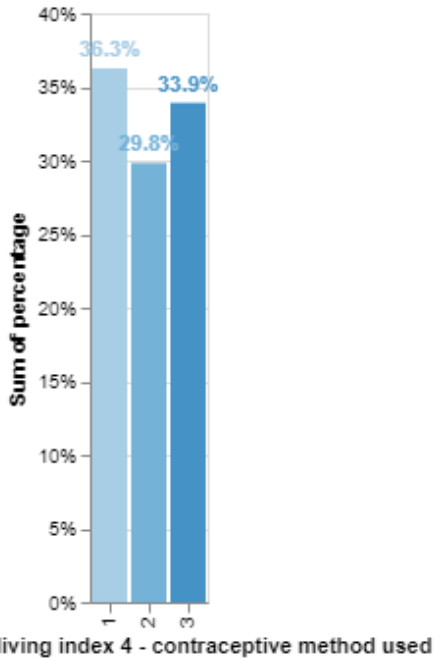


When comparing both employed and unemployed wives, it seems that wives using long-term contraceptive methods have husbands having lower earning jobs than wives using other methods. It is interesting to note the difference between means of occupation of husband for short-term contraceptives, as the data suggests that wives that are not working and use short-term contraceptives have higher earning husbands than those that are working and use the same contraceptives.





Standard-of-living index 3 - contraceptive method used



Standard-of-living index 4 - contraceptive method used

With the standard-of-living index rising, long-term contraceptive methods are jumping up in use, with the no use criteria significantly dropping. It is interesting to notice that short-term methods are staying more or less the same between each standard-of-living index.

4 BIBLIOGRAPHY

[1] Christa Kelleher, Thorsten Wagener, Ten guidelines for effective data visualization in scientific publications, Environmental Modelling & Software, Volume 26, Issue 6, 2011, Page 824, ISSN 1364-8152, https://doi.org/10.1016/j.envsoft.2010.12.006

[2] Daniel Archambault, Human-Centred Visual Analytics: Visualisation Rules of Thumb, Page 17

[3] Daniel Archambault, Human-Centred Visual Analytics: Data and Encodings, Page 12