P.P.A.P: A Pioneering Interactive Platform to Visualize Population Distribution and Demographics for Public Awareness and Enhanced Planning/Policy

Choy Yu Min Justin | Jiang Xi | Ng Jun Hong

Singapore Management University

Abstract — With Singapore's growing population and limited resources, she faces many pressing challenges for progressive development and economic growth. These challenges span across housing affordability, rising healthcare, aging population, education/income inequality, and low birth rates. To encourage open source problem solving, the government has made large pools of socially relevant data public. As such, we were inspired to use population data to build a visual analytics application to enhance the exploration of policy and planning. This visual analytics platform will largely be based on the 5 yearly Singapore population census data and uses several time series and map-based charts to manifest the different population trends. By using charts such as choropleth map, ternary plot, and horizontal network chart, we will enable the user to interactively explore large amounts of data in a much more interpretative manner. These charts have allowed us to better stand the population distribution better and gave us deeper insights into matters such as demographics and how they differ between planning areas. These insights are valuable in inspiring new ideas and approaches to building a more holistic and equal society – accross education, income, and age.

Key Words — Choropleth Map, Ternary Plot, Alluvium Chart (Horizontal Network Chart), Time Series Visualisation, Interactive Visualisation.

1 Introduction

In recent years Singapore has grown to house 5.9 million Residents in its 55 planning areas, which include two water-catchment areas. The largest region in terms of area is the West Region with 201.3 km2 (77.7 sq mi), while the Central Region is the most populous with an estimated population of 922,980 inhabitants in the area in 2019. Given its limited resources and land space, it is key for policies and planning to ensure that these residents are distributed evenly for holistic and reasonably equal communities throughout this little red dot.

However, although Singapore seems to enjoy consistent prosperity and harmony, it is evident that there are several developing social trends that need to be further analysed and controlled – trends such as growing income inequality [1], low birth rates, and equal opportunities for education across planning areas. As such, insights on educational qualification, resident distribution, and demographics are key in ensuring better awareness of these dangerous trends and will help educate and inspire both the public and policy makers to build a holistic Singapore together. Given the seriousness of this matter, the Government has pushed strongly to provide relevant data [2] and has also outlined similar objectives in its 2020 budget summary [3].

2 Motivation and objective

With the government's strong support and push for open source innovation, we felt that this is a key area that we could utilise our skills in bringing value to society through informing the public and assisting decision makers with planning. Furthermore, with the government's push towards a smart nation, there are increasing data sets available with reasonably high dimension that can allow us to get insights if visualised properly. In particular, there is a good amount of population data from Singstat.gov.sg that we have used for this project.

The key objectives we strive to achieve in this project consists of providing insights for in two main dimensions - Time series analysis of Resident Distribution & Deeper social demographic analysis using Population Census data. For each of these areas we have targeted to achieve the following:

Time series analysis of Resident Distribution:

- Provide an animated map view of Resident distribution for resident data (2011-2019)
- Provide complementary charts that give more specific views to observe trends over time series
- Provide interactive charts to allow users to set up simple filters and click into specific Planning Areas for further analysis.
- Provide chart views with multiple dimensions to allow for richer analysis (i.e. Ternary charts, Network charts)

Deeper social demographic analysis using Population Census data:

- Provide high level trend charts that map different social demographics and sentiments (2000, 2005, 2010, 2015)
- Provide interactive charts with relevant filters and customizable views to allow users to uncover deeper insights
- Provide chart views with multiple dimensions to allow for richer analysis (i.e. Ternary charts, Network charts)

By building these charts, we targeted to assist users to discover deeper insights in order to spur on more creative planning and policy making.

3 Data Selection

After our initial exploratory search, we recognised that our data sets of interest were collected across two different studies – namely the "Residential Time series data (2011-2019)", and "Population Census data (2010,2015)" These data sets will support the two main sets of dashboards to provide analysis for granular time series resident data, and higher dimensional population demographic data using the Population census data set. This is a brief outline of each of our data sets:

Resident by planning Area, subgroup, age Group, sex and dwelling (2011 - 2019, June) [4]**:** This dataset covers a good time series from 2011-2019 and the breakdown by subzone/planning area allows it to serve as the base platform to integrating with other population data sets that are grouped by subzone/planning area.

Singapore General Household Survey & Census of population (2010, 2015) [5]**:** This data adds a very rich level of dimensionality on top of the residential data as mentioned above. However, it only covers limited points in time and so we intend to use this data separately for more deep time static analysis. The data contains key attributes such as gender and Income levels, Educational Qualification, and Housing Type – all by planning area. For this project, we only processed data on Income, Education, and Housing types into a pivoted master data table. This was due to the complex nature of the raw data that made it difficult to combine many features into a single data frame. Furthermore, Planning Areas were different 2010 onwards, as such, we decided to drop the census data for 2000 and 2005. This was also because the format that 2005's census data was largely different from that of the rest, making it difficult to wrangle all the data together over the 4 data sets.

4 Approach

Having completed our data exploration, we proceeded to do further research on similar visualisation work in the field of visualising population demographics. We drew inspiration from 3 main sources: Singstat.gov, ggplot2 tidyverse gallery, and broad searching across the web. We repeated 3 rounds of searching, brainstorming, and critique before coming down to 5 main charts we decided were of value implementing in the application. This process will be explained in detail below.

4.1 Background Survey

We initially sourced out a majority of interactive time series maps [6] that allowed the users to evaluate how common resources were distributed over planning areas (i.e. hospitals, parks etc). However, after considering the nature of the data set, we later moved into charts that better showed trends of the population demographics and time series. Some charts that remained useful through our rounds of critique were: Bricks Map, Ternary Chart, and Network Diagram. We intended to use these charts to help users visualise trends such as: Distribution of aeging population, Income Inequality, and education levels.

4.2 Brainstorm design consideration

This following section covers they key charts we decided to implement, what we would use them to visualise, why they are useful, and how we sought to improve them in our own version of implementation.

4.2.1 Distribution Diagram

Figure 1: Distribution Interactive Area Chart

With the help of filters, this will allow users to discover all the subzone/ planning area income distribution in the same chart with a generic reference line of average income of the Singapore or world poverty line. However, the stacked area charts could serve to be more confusing and prevent the user from interpreting the trend with accuracy and ease. As such, in our implementation, we intend to use line charts instead.

4.2.2 Bricks Map

Figure 2: Bricks Map

The bricks map provides a quick overview of an indication in a geographical chart. Furthermore, it allows immediate distribution comparison in units measurement of a variable in the planning areas. This could serve to be useful to give a geological flavour to a clear and quick comparison of residential distribution. However, we do note that this chart is still not as clear as line or bar charts that immediately allow us to compare planning areas and their demographic more clearly. We also noted that there are currently no readily available packages that support this chart on R (only found support on Qlik sense); as such we will attempt to implement this using a bee swarm plot and translate it into a tmap. Nonetheless, the implementation of this chart is challenging and of high uncertainty. A possible substitute would be a simple time-series lollipop chart by planning areas complemented with a Choropleth map to provide the geographical view.

4.2.3 Ternary Point Chart

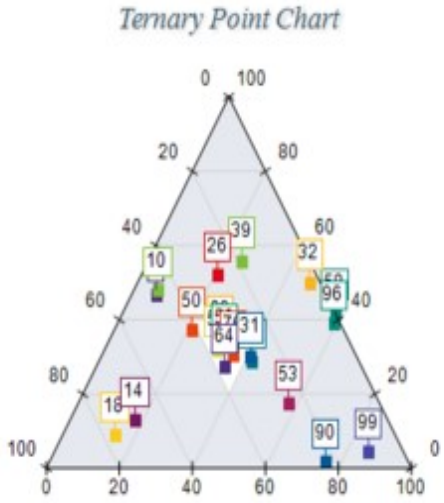

Figure 3: Ternary Point Chart

This chart helps viewers to find the population relationship between 3 variables. The colours are also used to distinguish the different data point groups based on Planning Area or other useful categories, and the grid lines assist the reader to match the points to their corresponding axis values. This chart will be useful when implemented with time series animation, as it would allow us to see the distribution of the population in 3 concurrent dimensions i.e. housing type, age group etc. This chart is also noted to have support and should be implementable.

4.2.4 Network Chart

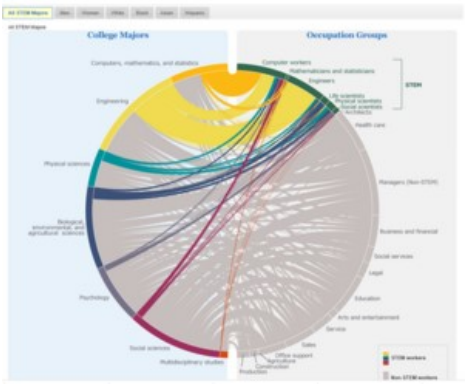

Figure 4: Network Chart

This chart helps us to make observations regarding different relationships between varying demographics of the population in a single view. This chart would help users gather deeper insights and allow them to trace interesting trends to corresponding ones based on the different network patterns. However, the spherical nature of this chart does not serve much purpose in making the visualisation clearer. Also, as we are only going to explore education, income and housing type, we will build a horizontal version of this chart to utilize the space and make it easier for users to draw observations.

4.2.5 Bar Charts

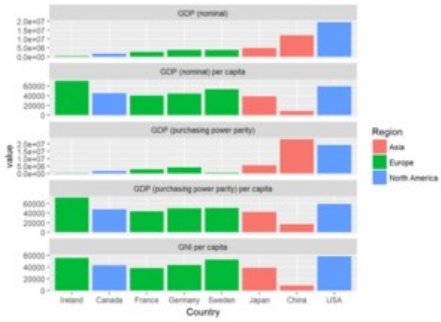

Figure 5: Bar Charts with Facets

This chart uses multiple grouped small graphs to allow for easy comparison of the same metrics for multiple planning areas. However, as we are using a small number of features for our analysis, we might not be using many facets and only use 1-2 relevant and important categorical groupings.

4.2.6 Overall Design

For the application, we decided to build a simple storyboard with 3 main views:

- **Population (Summary View):** A high level landing page that shows the user a time series set of charts with simple filters for quick analysis. These charts will be based on the Residential Time Series Data.
- **Population (Trend Finder):** A time series graph showing more specific breakdowns of time series data by set filters. This view will allow users to identify more specific population trends based on the Residential Time Series Data.
- **Demographic Analysis:** This view will primarily be using the Network Chart to allow users to explore the population demographics of education, income, and housing type by planning areas. The data will be taken from Population Census Data Source.

5 VISUALIZATION

The team followed through to build the application and attempted to improve the listed charts with improvements. Due to time limitations and lack of open source support for several charts, the team was forced to substitute two of the proposed graphs out. These substitutions were namely: Lollipop Chart & Choropleth in place of Bricks Map, and Line graphs in place of the Stacked Area Line Chart. The following section will cover the charts implemented in the final application and highlight interesting findings and uses that the team has explored with them.

5.1.1 Population (Summary View)

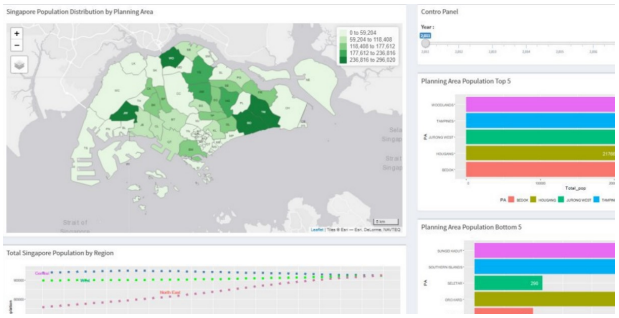

Figure 6: Population Summary View Dashboard

This first view gives users a quick overview and awareness of how Singapore's residents are distributed across the different planning areas of the country. The dashboard also has a timeline control slider to allow users to explore Singapore's population distribution through the years from 2011-2019. From our brief usage of this dashboard, we have found this chart proving useful to show us how Singapore's population density increased in certain areas up till a point in time that new housing estates were set up for new families i.e. Sengkang/Punggol in recent years and these areas quickly increased in population density. It also became clear that certain planning areas remained relatively less dense. This was likely because they were industrial/central business district areas and thus, held less residents. More detailed usage of each graph and their insights will be covered below.

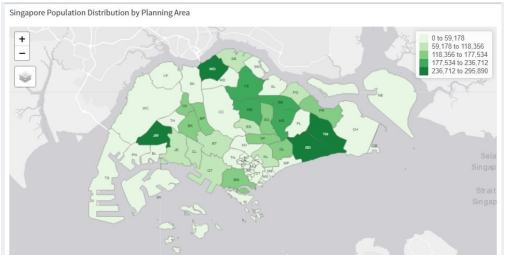

Figure 7: Choropleth Chart

This chart replaces the role of Bricks Map and uses colour intensity to display Singapore's population density by planning area. From exploring this chart with the time series control, we have found that most of the planning areas do not have much increase in population density except for those in the north region, especially Punggol. This is related to the many HDB projects in that area for new families in the recent years. Given this insight we are now keen to find out the demographics of this planning area to see if the community is balanced in this sector etc.

Figure 8: Bar Charts

This chart helps users narrow down on the top 5 and bottom 5 planning areas according to population density. This helps focus their attention on the biggest outliers of the planning area. From exploring this chart with the time series control, we have found that Hougang dropped out of the Top 5 and Sengkang has become 4th. This shows that the chart serves its purpose as this information is not obvious just by looking at the choropleth chart. Over the years we also see how the Top 5 and bottom show a gradual increase in population count with exceptional growth for some planning areas.

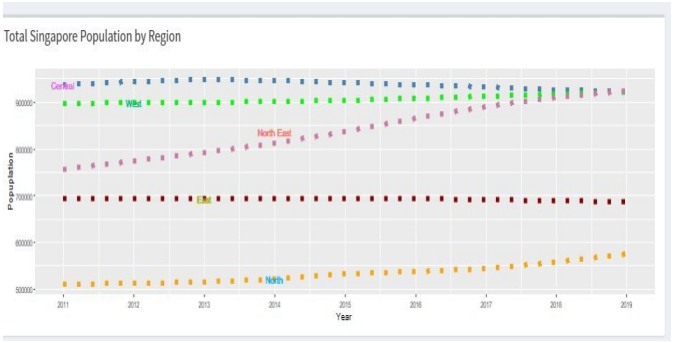

Figure 9: Singapore Population by Region Dotted Line Chart

This chart takes the role of our stacked area line chart to show the general overall trend of Singapore's Population by Region over the whole time period available. This chart is useful in allowing users to always maintain the big picture even as they scroll through the specific years for analysis. This chart has helped us recognise that regions of Singapore are likely to increase in population until they hit a saturation point (as seen from the top 3 regions). It also suggests to us that East and North might be the next two regions for housing development. Given this overview, we were also able to complement our insights drawn from specific yearly population distributions and get a general idea to explore demographics of perhaps the top 3 regions and see if they differ from the others.

5.1.2 Population (Trend Finder)

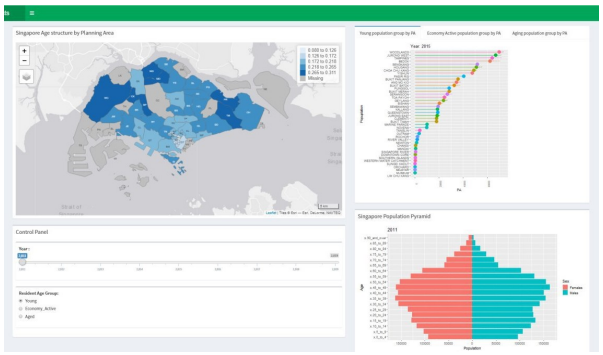

Figure 10: Population Trend Finder View

This dashboard revolves around allowing the user to explore the Population distribution by different age groups from 2011 to 2019. Using all the charts, and radio buttons it was made very clear that over time there is a notable increase in percentage of young working residents fanning outwards towards the north and west of Singapore. Likewise, we also note a gradual increase in the proportion of elderly in the central and eastern parts of Singapore. Knowing this might incentivize policy makers to intensify housing policies that promote cross generational families to live closer together. This would promote richer family units and a stronger Singapore community overall.

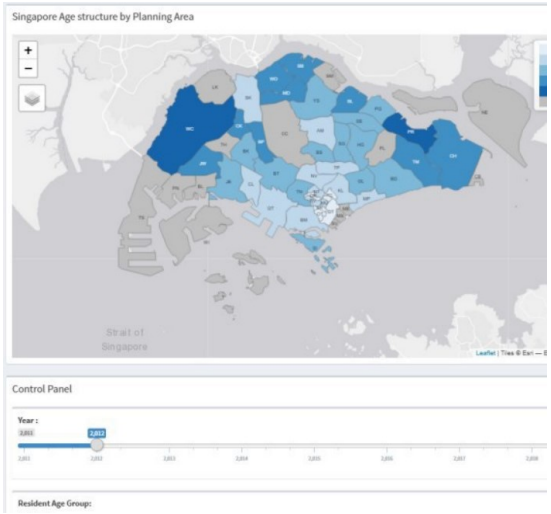

Figure 11: Choropleth Mapping (by age group)

This chart is similar to the one used in summary dashboard except that it shows the Percentage of the population that follows the radio button selected (Young, Economically Active, and Aged). This chart is controllable using a year-based timeline slider together with the radio button.

Figure 12: Lollipop Chart by Planning Area

This is a time series chart that plays through the data points from 2011 to 2019. It allows the user to quickly catch the population distribution trend over the years by planning area.

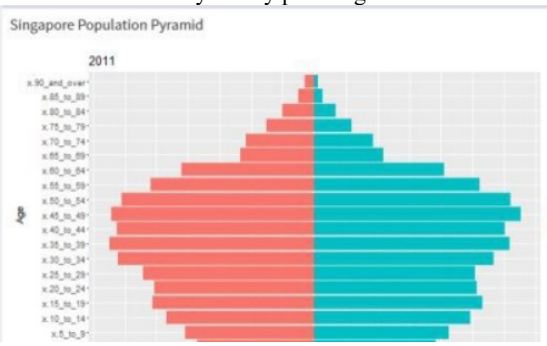

Figure 13: Age Sex Pyramid Chart

This chart helps us get a better picture of the population breakdown by gender. This chart is a substitute for the ternary char to still allow us to interpret the population in the dimension of gender.

5.1.3 Demographic Analysis

Figure 14: Population Demographic Analysis Dashboard

This dashboard consists of a single network chart showing the proportions of education qualifications to housing type and income bracket. This chart has 2 filters: one radio button to choose between 2010 and 2015, and a dropdown list to view the specific demographic for the selected planning area. This chart is for the last level of in-depth analysis to allow the user to gain deeper insights on specific planning areas they were following in the previous two dashboards. We have explored Punggol and Sengkang using this chart following our interest from above that these were the two planning areas that had the biggest and fastest population growth. For these two planning areas between 2010 and 2015, majority of their residents stay in HDBs and we have found that both Punngol and Sengkang have seen an significant increase in proportion of residents with university qualification. We also noted that majority of them also fell into the income bracket of 5000 and below. This suggests that the majority of the new residents taking up residency in the new HDBs here are young working adults and families.

6 Limitation

Overall, the application turned out useful and we were able to implement most of our intended charts. However, complex data, limited open source support, and package compatibility issues caused us to be unable to implement all our initial interactive elements and charts.

Complex data was one of the main issues that limited us from using all the charts according to our original plan – this was especially so for the population census data. Originally, we had intended to link the many dimensions of population demographics into a master data frame to allow the user to visualise and explore many correlations in the population demographics. However, after building our initial charts, we realised that the data formats of population census data differed between the 5-year intervals rather greatly, especially for 2005. Because of this, we had to narrow down to using only 2010 and 2015 Population Census Data and had to manually combine and rename several columns before finally pivoting the necessary fields to achieve a clean data set. Moreover, although the Residential Time series data (2011-2019) was a lot cleaner and almost ready to use, it lacked dimensionality to allow for further analysis.

The limited open source support and compatibility issues also hampered us from being able to develop our Bricks Map chart and ternary chart. Nonetheless, the group had already planned for alternatives during our thorough background survey and brainstorming sessions, and this allowed us to finish the project on time and up to standard.

7 Conclusion

This pioneering effort has already shown us that we can draw useful insights about Singapore's population and its demographics through useful visualisations. Although the insights drawn are just scratching the surface it serves already as a good platform for the public to gain awareness and understand Singapore better. Moreover, this application has already inspired us to explore planning and policies in general directions by helping us get a general picture of the dense/uncrowded planning areas of Singapore, their demographics, and how we might want to set up policies and plans to develop the communities in these planning areas to be more holistic.

As such, future work to improve this application would revolve around the following:

- Explore other population demographic factors such as employment, marital status and etc for more meaningful analysis.
- Work on more sophisticated and applicable visualisation charts such as brick map and etc.
- Create optimization model for future residential planning.

Acknowledgments

This project was only made successful under the supervision of Dr Kam Tin Seong, the guiding publications from Edward Tufte, and the many open source authors' works on the ggplot2 tidyverse gallery.

References

[1] M. o. F. Singapore, "2020 Budget Summary," 2020. [Online]. Available: https://www.singaporebudget.gov.sg/budget_2020/budget-measures/budget-summary.

[2] S. N. Singapore, 2020. [Online]. Available: https://www.smartnation.gov.sg/.

[3] T. S. Times, May 2018. [Online]. Available: https://www.straitstimes.com/politics/parliament-inequality-not-so-bad-in-singapore-but-cracks-showing-in-social-mobility-ong-ye.

[4] Singstat.gov, 2020. [Online]. Available: https://www.singstat.gov.sg/find-data/search-by-theme/population/geographic-distribution/latest-data.

[5] Singstat.gov, 2020. [Online]. Available: https://www.singstat.gov.sg/find-data/search-by-theme/population/geographic-distribution/latest-data.

[6] Morphcode. [Online]. Available: https://morphocode.com/the-5-minute-walk/.